論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Methodology of data efficient deep learning for newly built construction detection in bitemporal SAR images
著者(和文)	JATURAPITPORNCHAI
Author(English)	Raveerat Jaturapitpornchai
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11650号, 授与年月日:2020年9月25日, 学位の種別:課程博士, 審査員:松岡 昌志,中村 芳樹,室町 泰徳,那須 聖,淺輪 貴史
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11650号, Conferred date:2020/9/25, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Doctoral Dissertation

Methodology of data efficient deep learning for newly built construction detection in bitemporal SAR images

Raveerat Jaturapitpornchai

Department of Architecture and Building Engineering
School of Environment and Society
Tokyo Institute of Technology

September 2020

Abstract

While remote sensing data is suitable for monitoring urban expansion by detecting changes of new constructions from time-series images, the conventional methods by optical sensors are incapable of locating such a specific change, especially under cloudy weather condition. Thus, the aim of this study is to identify newly built constructions with a deep learning technique in bitemporal synthetic aperture radar (SAR) data. Although the training of a deep learning network usually requires a high amount of training data, the method proposed in this thesis is designed to utilize the limited quantity of the training data and still can generate an accurate detection result by taking an advantage of the having both ordinary and reverse chronological order of the time-series data as a training set. The results in this study show that the model trained by the proposed method is versatile as it can be used with multiple types of terrains, SAR images with different frequency bands and orbit direction, and images with speckle noise.

Table of contents

Chapter 1. Introduction

1.1 Background and urbanization monitoring

For the past century, the number of urban populations compared to the population in the rural areas has been significantly increasing. For instance, the number of people residing in urban area was 30% of the world population in 1950 while there was approximately 54% in 2014 [1]. Eventually, by 2050, about 66% of the world population would be residing in urban area or approximately adding 2.5 billion people to the world's urban population where nearly 90% of the increase are concentrated in Asia and Africa according to the estimation by United Nation [2]. Especially in case of Asian countries, the rapid urbanization, which can be considered as the key driving factor of development growth, is evidenced by the increasing of physical growth that extends beyond metropolitan and city boundaries [3-4]. Asian megacities have attracted people, investments, businesses, and organizations as they became a center of the rapid growth of economic. Since they facing urban expansion into their periphery areas, they bring both benefits and the problems of urbanization [5]. As the rapid urbanization has been predicted as a process that will continue in the coming years, the sustainable development challenges will have to be concentrated, particularly in the lower-middle-income countries where experiencing fastest urbanization.

The needs to recognize the expanding of urban areas are always exist due to many reasons as it can be associated with other important aspects such as economic, social, and environment. Based on UN, urban living is often associated with higher levels of literacy and education, better health condition, greater access to social and economic services, and enhanced opportunities for cultural and political participation. This means rapid and

unplanned urban growth as well as urban expansion can be a threat for sustainable development if the necessary infrastructure is not developed or when policies are not well implemented. Unplanned or inadequately managed urban expansion leads to rapid sprawl, pollution, and environmental degradation, together with unsustainable production and consumption patterns. The rapid urban growth, high population density and high consumption rate of residents in megacities has led to a wide range of local and global socioeconomic and environmental impacts which requires attention since it can significantly affect the global sustainability. Continuing urbanization or migration from rural to urban areas will eventually cause environmental deterioration, inadequate housing, traffic congestion, slums, the rising crime rate, homelessness and so forth [6-7]. Thus, these issues need to be brought to the attention to ensure the role of urban center as the engine of economic growth will be continuously maintained and enhanced [8]. For the purpose of preventing the consequences of poor urban planning management, the accurate monitoring of the urban area expansion needed to be delivered. Since Bangkok is one of the most strategic area in one of the developing countries in tropical region, research associated with the development of assessment methods or models to solve all issues and problems that occur in these areas should be very strategic to be conducted.

Generally, in order to make a construction in Thailand, a building permit must be obtained from a municipal office in the area which the site is located (except for the construction smaller than 100 m^2). The entrepreneurs have to submit an application for a permit with supporting documents attached, including a copy of construction blueprint [9]. However, since the procedure does not require a site inspection after the construction has been finished, in many cases, the actual construction can be different from those in the submitted documents. As a result, the database of the detail of constructions collected from these

construction permissions may not accurately represent the shape and size of the buildings. Thus, the automatic detection system can help verifying the records of the new constructions without having to send an officer to perform a field survey. Furthermore, since many buildings were constructed without any permission especially outside the capital cities where the inspecting of government is difficult to access [10-12], the automatic detection system can also help the government or local municipalities monitor the unauthorized buildings.

In the past, numerous of the studies to measure the expanding of the urban area by locating the exact position of newly built constructions for tackling these issues have been developed. Beside the field survey that consumes high amount of time and human resource, an effort to monitor and analyze the construction of buildings has been roughly divided into two categories. The first category is measuring the urban sprawl by using some various specific indicators proposed by each researcher [13-16], while the second category is utilizing the spatial and temporal technologies such as GIS and remote sensing with additional statistical techniques [17-29]. However, the mapping and monitoring of urban expansion and building changes using GIS and remote sensing techniques has attracted more interests and has largely proved to be effective and valuable tools for such objective over a time period [30-37].

1.2 Remote sensing data for change detection

Definition of remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object [38]. By mounted on satellites or aircrafts, remote sensing sensors collect data by detecting the energy reflected from the

Earth's surface and acquire massive amounts of data daily about the earth. This data has significant ground surface coverage and suitable time intervals between acquisition dates for effective temporal analysis of the ground surface. As a result, the interpretation of remotely sensed data can be used in effective and automatic detection of changes that have occurred between any bitemporal acquisitions of time-series data. Timely and accurate change detection results are important in many application fields. With this ability at disposal, the people can better manage land use planning—for example, when planning the expansion of a city [39–41].

The remote sensing can be separated into active sensor and passive sensor according to whether they generate and emit signals from themselves in the data collection process. A commonly used approach for this type of observation is the use of remote sensing data from optical sensors, which is one of the passive sensors, because such data allow the easy creation of maps as the image is illuminated with sunlight. Despite the excellence of optical data, some developing countries are located in tropical areas where clouds cover parts of the area all year round. Unfortunately, optical sensors cannot capture Earth's surface below these clouds. Because synthetic-aperture radar (SAR) captures images using microwave signals that can penetrate clouds, the use of SAR data is a secondary option to handle the problem. However, the difficulty in the interpretation of SAR images makes it harder to identify locations of new constructions. In the literature, many methods have been proposed for SAR image change detection with threshold method and clustering methods [42-46] with the post-classification comparison technique. This technique is used to compare images in which the urban area has been classified over different time periods. Although it has been widely used [47-50], a significant amount of pre-processing of these images is necessary due to differences between observing environments such as serious effects of atmospheric

disturbances, missing data at a desired date, and correction of inaccurately observed time spans [51]. On the other hand, some publications generate a difference image from the pixel information of two images before the processing of classification, from which it is difficult to identify one specific change, such as the appearance of new buildings, as any kind of change similar to the target change would be involved in the results. For instance, Y. Ban and O. Yousif [52] used the threshold-based method on a difference image in detecting urban change. Despite the good detection result, there is a possibility to detect falsely when the urban or non-urban area has unordinary intensity change behavior. In a real-life application of change detection, users usually want to see only the changes of interest while ignoring all others; thus, instead of detecting all of the changes that occurred, the application should only detect changes in a specific target, for example, detect changes in buildings while ignoring paddy field seasonal changes.

Traditionally, such specific detections have been made manually, based on the experience of human experts, but manual methods are expensive, time-consuming and error-prone. Besides, most existing algorithms are designed for dealing with data provided by the same sensor, with the same spectral range and number of spectral bands, and are sensitive to pre-processing and noise.

To overcome these difficulties and develop more reliable and automatic change detection methods from remotely sensed data, machine learning techniques have been extensively employed by researchers. Among all of the machine learning techniques, the deep learning is the most promising as it produced many state-of-the-art results in various fields of image processing tasks and it has become some of the most active research directions in remote sensing [53-56]. Although designing a general deep learning system for image

understanding is challenging, it is becoming more acceptable in the remote sensing and computer science community. In particular, among all types of changes occurring in the world, building changes have attracted considerable attention [57-59]. They are some of the most important object classes in the GIS, especially for urban planning, and they illustrate the extent to which humans have utilized and reformed the natural environment. Thus, this study is focusing on applying deep learning method on remote sensing data for detecting changes in building constructions.

1.3 Machine learning in remote sensing

For many years, machine learning algorithms have demonstrated their successful development in working with data intensive technical and scientific fields such as search engines, speech recognition, and robotics. Remote sensing applications also have to deal with mass data and complex pattern recognition tasks as the remote sensing sensors have wide range of the electromagnetic spectrum. Not to mention the capabilities of the sensors include single band images as well as multi- or hyperspectral data. Moreover, due to the fact that remote sensing applications are often monitoring tasks, long time series data are in the focus of image exploitation as well. This leads to the successful application of the machine learning algorithms for remote sensing application for decades. The used algorithms range from basic algorithms such as Principle Component Analysis (PCA) and K-Means to more complicate classification and regression frameworks such as Support Vector Machines (SVMs), decision trees, Random Forests, and artificial neural networks.

The impressive development in the last years is deep learning, especially with Convolutional Neural Networks (CNNs), a specific variant of artificial neural networks (ANNs). While

numerous traditional methods used the conditioning of intensity change in detecting the building of new constructions, deep leaning, however, does not need the setting of any condition to make it learn the change between two images. The deep learning is a kind of machine learning where the fundamental idea is to let the machine learn to solve a given problem by itself. As humans are exposed to myriad of sensory data received every second of the day and are somehow able to capture critical aspects of this data in a way that allows for its future use in a concise manner, the deep learning tries to mimic this working mechanism of human's brain in order to study the features to gain full knowledge of the perceived object [60].

The ideas for designing systems that represent information have been led by neuroscience findings which have provided insight into the principles governing information representation in the mammal brain. One of the key findings has been that the neocortex, which is associated with many cognitive abilities. It does not explicitly pre-process sensory signals, but rather allows them to propagate through a complex hierarchy [61] of modules that, over time, learn to represent observations based on the regularities they exhibit [62]. This discovery motivated the emergence of the subfield of deep machine learning as an attempt to exhibit the characteristics of the neocortex, which focuses on computational models for information representation by training such a hierarchical network on a large set of observations and later extract signals from this network to a relatively simple classification engine for the purpose of robust pattern recognition. By mimicking the human's cognitive procedures, the achieved deep learning would have a robustness that includes an ability to exhibit classification invariance to a diverse range of transformations and distortions, including noise, scale, rotation, various lighting conditions, displacement, etc.

The first successful deep learning approach where many layers of a hierarchy were successfully trained in a robust manner is a CNN. CNN has shown great promise for future work in the deep learning field. In more recent work, researchers have applied CNNs to various machine learning problems including face detection [63-64], document analysis [65], and speech detection [66]. CNNs have recently [67] been trained with a temporal coherence objective to leverage the frame-to-frame coherence found in videos, though this objective need not be specific to CNNs.

While many state-of-the-art deep learning methods based on CNN have been rapidly published for several recent years, they all are sharing the fact that the higher number of training results in higher detection accuracy, which leads to the issue of the requirement of sufficiently large and variable datasets. There are many problems of acquiring such amount of training data. One of the biggest problems is that such data may be laborious and expensive to acquire and label since annotations are usually performed manually in order to name a particular event, action, motion, object, shape, color etc. This is the crucial part for creating the training data and must be performed with care, otherwise the machine learning algorithm may not converge during the training phase due to the misnaming of annotations. There are some more several issues where, in many cases, data availability is limited due to privacy, copyright or intellectual property issues or there may be a cost in acquiring it [68]. Unfortunately, these issues also apply to the remote sensing data and thus the sufficient number of remote sensing data available as a training data for newly built constructions detection is difficult to be obtained. As a result, this study would like to focus on proposing a novel deep learning architecture where a high amount of training data is not required.

1.4 Objectives

Because the number of populations residing in urban area is increasing, there is always a high demand for the construction of new residential and business areas. The expanding of the urban area requires close attention since it can significantly affect the global sustainability. Thus, this issue needs to be addressed to prevent such consequences.

The remote sensing data is a valuable information that can be used for managing land use planning by monitoring the expansion of a city. Since many cities with a rapid urbanization rate are located in tropical area, such as Bangkok, Thailand, and the Earth surface can be covered by clouds, the SAR image which can be captured regardless the weather condition is selected to use in this study. Because of the manual detection based on the experience of human experts are expensive and time-consuming, the algorithms for automatic detection have been proposing for past several decades. However, the traditional methods where the fixed mathematical conditions are used are difficult to identify one specific change, such as the appearance of new buildings, as any kind of change similar to the target change would be involved in the results. As a result, the machine learning methods, specifically deep learning, have acquired the attention from many researchers because of its ability to generate an accurate result without relying on any fixed conditions by using the learning of the existing dataset instead. Although deep learning, can produce an accurate detection result when sufficient amount of training data is used, but because of the difficulty in acquiring training data for the newly built constructions detections purpose due to the publicity of data and labeling cost, the author decides to create our own ground truth to fulfill this purpose. However, the creation of the ground truth is involving a lot of manually labeling works, which is very tedious and resulting in only a moderate quantity with narrow study areas.

The aim of this study is to tackle the problem by proposing the deep learning architecture, which does not rely on certain conditions like in conventional methods, for detecting newly built constructions in bitemporal SAR images without a needing of high number of training data. The proposed method is designed to be able to use in under variety of geographical profile and not sticking with only a few datasets, making it can be used with images from different satellites as well as areas in different geological scenarios.

1.5 Structure of the thesis

This thesis is divided into seven chapters. The first chapter explained the background of the objective for this thesis, which is the detecting of the new constructions in remote sensing imagery. This chapter also describe a goal for this thesis which is to tackle the problem of the locating changes of constructions occurring between satellite images from two different times to help with urbanization managing by using the deep learning-based method as it has a significant ability to recognize features and changes which come from the mimicking of the human brain.

Chapter 2 provides some fundamental knowledge required for understanding the thesis including information of SAR images and deep learning. The literature review for the urban change detection in SAR images using deep learning techniques and their problems is addressed in this chapter.

Chapter 3 describes the information of the SAR images used in this study. The explanation of each study area along with the process of creating ground truth for using as training data and testing data are described. The problem of created ground truth in term of unbalance

amount between each class and the solution for this problem have been addressed in this chapter.

Chapter 4 introduces U-net and its architecture detail. This chapter includes the experiment of using U-net model with Bangkok testing area and its problem in applying to an image from other sensors and consuming large amount of training data.

Chapter 5 presents the solution to the data consuming problem in U-net by introducing a novel deep learning architecture that can utilize training data efficiently. This chapter describe the idea and architecture detail of the proposed method including showing an experimental result of detecting new buildings in Bangkok. The data efficiency of the proposed method has been compared with U-net in this chapter.

Chapter 6 demonstrates the versatility of the proposed method by showing the experimental results of using with other study areas including images from another sensor and acquisition conditions. This chapter also shows the robustness of the proposed method by testing with the noisy images.

Chapter 7 concludes all the main points of the thesis. The prospective future development is also discussed in this chapter.

Reference

[1] Andrea, Emma Pravitasari (2015) Study on Impact of Urbanization and Rapid Urban Expansion in Java and Jabodetabek Megacity, Indonesia. Kyoto University.

[2] United Nation. (2014). World Urbanization Prospects. The 2014 Revision. ISBN: 978-92-1-151517-6.

[3] Hugo, G. "Population development and the urban outlook for Southeast Asia." Marshall Cavendish International (Singapore) Private Limited, 2006.

[4] McGee, T. G., and Robinson, I. M. (eds) (1995) The mega-urban regions of southeast Asia. UBC Press, Vancouver

[5] Asian Development Bank. (2008). Managing Asian Cities. Sustainable and inclusive urban solutions. ISBN: 978-971- 561-698-0.

[6] Makinde, O.O. (2012), "Urbanization, housing and environment: Megacities of Africa", International Journal of Development and Sustainability, Vol. 1 No. 3, pp. 976-993.

[7] Brian, P. G. (2000). The governance of the city: A system at odds with itself. University of New York, pp. 5-6.

[8] Jusoh, H., and Rashid, A. A. (2008). Efficiency in Urban Governance towards sustainability and competitiveness of city: A case of Kuala Lumpur. International Journal of Social, Education, Economics and Management Engineering Vol. 2 (4).

[9] คู่มือประชาชน สำนักการโยธา (ปรับปรุง 7 มี.ค. 2562), http://203.155.220.238/dpw/index.php/yota-services/yota-menu-handbook/44-grantandservice-peoplehandbook-yota/222-grantandservice-peoplehandbook-yota-khor1

[10] Carl V. Patton & Costas M. Sophoutis (1989) UNAUTHORIZED SUBURBAN HOUSING PRODUCTION IN GREECE, Urban Geography, 10:2, 138-156, DOI: 10.2747/0272-3638.10.2.138

[11] Müller, Y., & Lješković, S. (2008). Illegal construction in Montenegro. Techika Chronika Scientific Journal TCG, 1, 1-2.

[12] Zegarac, Z. (1999). Illegal construction in Belgrade and the prospects for urban development planning. Cities, 16(5), 365-370.

[13] R.B. Peiser Density and urban sprawl Land Econo., 65 (3) (1989), pp. 193-204

[14] H. El Nasser, P. Overberg A comprehensive look at sprawl in America USA Today, 22 (2001), p. 1

[15] G. Galster, R. Hanson, M.R. Ratcliffe, H. Wolman, S. Coleman, J. Freihage, Wrestling sprawl to the ground: defining and measuring an elusive concept, Housing Policy Debate, 12 (4) (2001), pp. 681-717

[16] R. Ewing, R. Pendall, D. Chen, Measuring Sprawl and Its Impacts Smart Growth America, Washington, DC (2003)

[17] K.C. Clarke, L.J. Gaydos, Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore, Int. J. Geograph. Inform. Sci., 12 (7) (1998), pp. 699-714

[18] G. Galster, R. Hanson, M.R. Ratcliffe, H. Wolman, S. Coleman, J. Freihage, Wrestling sprawl to the ground: defining and measuring an elusive concept, Housing Policy Debate, 12 (4) (2001), pp. 681-717

[19] A.G.O. Yeh, L. Xia, Measurement and monitoring of urban sprawl in a rapidly growing region using entropy, Photogram. Eng. Remote Sens., 67 (1) (2001), pp. 83-90

[20] J.E. Hasse, R.G. Lathrop, Land resource impact indicators of urban sprawl, Appl. Geograph., 23 (2) (2003), pp. 159-175

[21] N. Thomas, C. Hendrix, R.G. Congalton, A Comparison of Urban Mapping Methods using High-resolution Digital Imagery, Photogram. Eng. Remote Sens., 69 (9) (2003), pp. 963-972

[22] W. Ji, J. Ma, R.W. Twibell, K. Underhill, Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics, Comput. Environ. Urban Syst., 30 (6) (2006), pp. 861-879

[23] M.K. Jat, P.K. Garg, D. Khare, Modelling of urban growth using spatial analysis techniques: a case study of Ajmer city (India), Int. J. Remote Sens., 29 (2) (2008), pp. 543-567

[24] A.M. Dewan, Y. Yamaguchi, Land use and land cover change in Greater Dhaka, Bangladesh: using remote sensing to promote sustainable urbanization, Appl. Geograph., 29 (3) (2009), pp. 390-401

[25] M.G. Tewolde, P. Cabral, Urban sprawl analysis and modeling in Asmara, Eritrea Remote Sens., 3 (10) (2011), pp. 2148-2165

[26] J.S. Rawat, V. Biswas, M. Kumar, Changes in land use/cover using geospatial techniques: a case study of Ramnagar town area, district Nainital, Uttarakhand, India, Egypt. J. Remote Sens. Space Sci., 16 (1) (2013), pp. 111-117

[27] S. Deep, A. Saklani, Urban sprawl modeling using cellular automata, Egypt. J. Remote Sens Space Sci., 17 (2) (2014), pp. 179-187

[28] D.D. Alexakis, M.G. Gryllakis, A.G. Koutroulis, A. Agapiou, K. Themistocleous, I.K. Tsanis, S. Michaelides, S. Pashiardis, C. Demetriou, K. Aristeidou, A. Retalis, F. Tymvios, D.G. Hadjimitsis, GIS and remote sensing techniques for the assessment of land use changes impact on flood hydrology: the case study of Yialias Basin in Cyprus, Nat. Hazard Earth Syst. Sci. Discuss., 14 (2014), pp. 413-426

[29] T. Liu, X. Yang, Monitoring land changes in an urban area using satellite imagery, GIS and landscape metrics, Appl. Geograph., 56 (2015), pp. 42-54

[30] A.G.O. Yeh, X. Li, An integrated remote sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the Pearl River Delta, China, Int. Plan. Stud., 2 (2) (1997), pp. 193-210

[31] I. Masser, Managing our urban future: the role of remote sensing and geographic information systems, Habit. Int., 25 (4) (2001), pp. 503-512

[32] M.K. Jat, P.K. Garg, D. Khare, Monitoring and modelling of urban sprawl using remote sensing and GIS techniques, Int. J. Appl. Earth Obs. Geoinf., 10 (1) (2008), pp. 26-43

[33] A.A. Belal, F.S. Moghanm, Detecting urban growth using remote sensing and GIS techniques in Al Gharbiya governorate, Egypt, Egypt. J. Remote Sens. Space Sci., 14 (2) (2011), pp. 73-79

[34] A. Butt, R. Shabbir, S.S. Ahmad, N. Aziz, Land use change mapping and analysis using Remote Sensing and GIS: a case study of Simly watershed, Islamabad, Pakistan, Egypt. J. Remote Sens Space Sci., 18 (2) (2015), pp. 251-259

[35] M. Dadras, H.Z. Shafri, N. Ahmad, B. Pradhan, S. Safarpour, Spatio-temporal analysis of urban growth from remote sensing data in Bandar Abbas city, Iran, Egypt. J. Remote Sens Space Sci., 18 (1) (2015), pp. 35-52

[36] J. Epsteln, K. Payne, E. Kramer, Techniques for mapping suburban sprawl, Photogram Eng. Remote Sens., 63 (2002), pp. 913-918

[37] B.N. Haack, A. Rafter, Urban growth analysis and modeling in the Kathmandu Valley, Nepal, Habit. Int., 30 (4) (2006), pp. 1056-1065

[38] Liu J.G., and Mason P.J. 2009. Essential Image Processing and GIS for Remote Sensing. UK, Wiley-Blackwell, ISBN: 978-0-470-51032-2.

[39] Rogan, J.; Chen, D. Remote sensing technology for mapping and monitoring land-cover and land-use change. Prog. Plan. 2004, 61, 301–325.

[40] Shalaby, A.; Tateishi, R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. Appl. Geogr. 2007, 27, 28–41.

[41] Dewan, A.M.; Yamaguchi, Y. Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. Appl. Geogr. 2009, 29, 390–401.

[42] Bazi, Y.; Bruzzone, L.; Melgani, F. Automatic Identification of the Number and Values of Decision Thresholds in the Log-Ratio Image for Change Detection in SAR Images. IEEE Geosci. Remote Sens. Lett. 2006, 3, 349–353, doi:10.1109/LGRS.2006.869973.

[43] Mu, C.; Li, C.; Liu, Y.; Sun, M.; Jiao, L.; Qu, R. Change detection in SAR images based on the salient map guidance and an accelerated genetic algorithm. In Proceedings of the 2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastian, Spain, 5–8 June 2017; pp. 1150–1157, doi:10.1109/CEC.2017.7969436.

[44] Liu, M.; Zhang, H.; Wang, C.; Wu, F. Change Detection of Multilook Polarimetric SAR Images Using Heterogeneous Clutter Models. IEEE Trans. Geosci. Remote Sens. 2014, 52, 7483–7494.

[45] Hu, H.; Ban, Y. Unsupervised Change Detection in Multitemporal SAR Images Over Large Urban Areas. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2014, 7, 3248–3261.

[46] Gong, M.; Zhou, Z.; Ma, J. Change Detection in Synthetic Aperture Radar Images based on Image Fusion and Fuzzy Clustering. IEEE Trans. Image Process. 2012, 21, 2141–2151.

[47] Alphan, H. (2003). Land-use change and urbanization of Adana, Turkey. Land Degradation & Development, 14(6), 575–586.

[48] Weber, C., & Puissant, a. (2003). Urbanization pressure and modeling of urban growth: example of the Tunis Metropolitan Area. Remote Sensing of Environment, 86(3), 341–352.

[49] Yu, X. J., & Ng, C. N. (2007). Spatial and temporal dynamics of urban sprawl along two urban–rural transects: A case study of Guangzhou, China. Landscape and Urban Planning, 79(1), 96–109.

[50] Jat, M. K., Garg, P. K., Khare, D., & Jat, M. (2008). Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. International Journal of Applied Earth Observation and Geoinformation, 10(1), 26–43.

[51] Patino, J. E., & Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. Computers, Environment and Urban Systems, 37, 1–17.

[52] Ban, Y.; Yousif, O. Multitemporal Spaceborne SAR Data for Urban Change Detection in China. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2012, 5, 1087–1094.

[53] Frate F.D., Schiavon G., and Solimini C. 2004. "Application of Neural Networks Algorithms to Quickbird Imagery for Classification and Change Detection of Urban Areas." In: IEEE International Geoscience and Remote Sensing Symposium, Alaska, USA, 2662-2664.

[54] Lunetta R.S., Knight J.F., Ediriwickrema J., Lyon J.G., and Worthy L.D. 2006. "Land-Cover Change Detection Using Multi-Temporal MODIS NDVI Data." Remote Sensing of Environment, 105(2), 142-154.

[55] Bazi Y., Melgani F., and Al-Sharari H.D. 2010. "Unsupervised Change Detection in Multispectral Remotely Sensed Imagery with Level Set Methods." IEEE Transactions on Geoscience and Remote Sensing, 48(8), 3178-3187.

[56] Hussain M., Chen D.M., Cheng A., Wei H., and Stanley D. 2013. "Change Detection From Remotely Sensed Images: From Pixel-Based to Object-Based Approaches." ISPRS Journal of Photogrammetry and Remote Sensing, 80, 91-106.

[57] Trinder J., and Salah M. 2012. "Aerial Images and Lidar Data Fusion for Disaster Change Detection." In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Melbourne, Australia, 227-232.

[58] Huang X., and Zhang L.P. 2012. “Morphological Building/Shadow Index for Building Extraction from High-Resolution Imagery over Urban Areas.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1), 161-172.

[59] Chen Y.M., Cheng L., Li M.C., Wang J.C., Tong L.H., and Yang K. 2014. “Multiscale Grid Method for Detection and Reconstruction of Building Roofs from Airborne LiDAR Data.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(10), 4081-4094.

[60] I. Arel, D. C. Rose and T. P. Karnowski, "Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]," in IEEE Computational Intelligence Magazine, vol. 5, no. 4, pp. 13-18, Nov. 2010, doi: 10.1109/MCI.2010.938364.

[61] T. Lee and D. Mumford, “Hierarchical Bayesian inference in the visual cortex,” J. Opt. Soc. Amer., vol. 20, pt. 7, pp. 1434–1448, 2003.

[62] T. Lee, D. Mumford, R. Romero, and V. Lamme, “The role of the primary visual cortex in higher level vision,” Vision Res., vol. 38, pp. 2429–2454, 1998.

[63] F. H. C. Tivive and A. Bouzerdoum, “A new class of convolutional neural networks (SICoNNets) and their application of face detection,” in Proc. Int. Joint Conf. Neural Networks, 2003, vol. 3, pp. 2157–2162.

[64] Y.-N. Chen, C.-C. Han, C.-T. Wang, B.-S. Jeng, and K.-C. Fan, “The application of a convolution neural network on face and license plate detection,” in Proc. 18th Int. Conf. Pattern Recognition (ICPR’06), 2006, pp. 552–555.

[65] P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in Proc. 7th Int. Conf. Document Analysis and Recognition, 2003, pp. 958–963.

[66] S. Sukittanon, A. C. Surendran, J. C. Platt, and C. J. C. Burges, “Convolutional networks for speech detection,” Interspeech, pp. 1077–1080, 2004.

[67] H. Mobahi, R. Collobert, and J. Weston, “Deep learning from temporal coherence in video,” in Proc. 26th Annu. Int. Conf. Machine Learning, 2009, pp. 737–744

[68] Mastorakis, G. (2018). Human-like machine learning: limitations and suggestions. arXiv preprint arXiv:1811.06052.

Chapter 2. Literature Review

2.1 Fundamental of SAR image

SAR image is completely different from ordinary optical image in term of acquisition mechanism. Optical image is captured with passive sensors which detect sunlight radiation reflected from the earth in the visible and infrared of the electromagnetic spectrum (right side of Figure 2.1). This kind of sensors do not emit their own radiation, but receive natural light from the earth's surface. In the other hand, the sensor used in capturing SAR images is active sensors that emit artificial radiation to monitor the earth surface (left side of Figure 2.1). Radar satellites use short pulses of electromagnetic radiation in the microwave spectral range, therefore they do not depend on daylight and are hardly affected by clouds, dust, fog, wind and bad weather conditions, allows them to be operated under any time or weather. They measure the radar pulses reflected from the ground, analyze the signal intensity in order to retrieve information on the structure of the earth surface as shown in Figure 2.2.

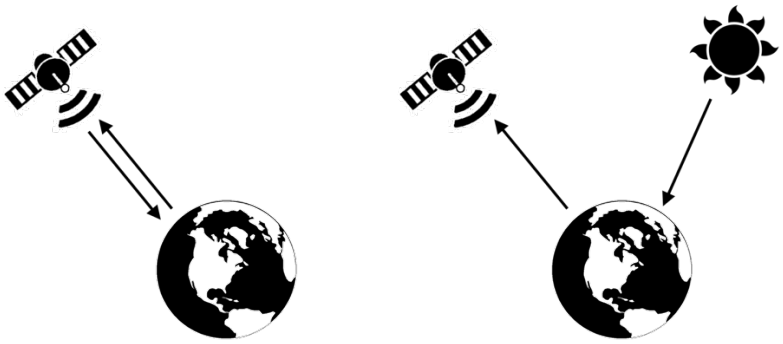


Figure 2.1. (Left) Active sensor, (Right) Passive sensor

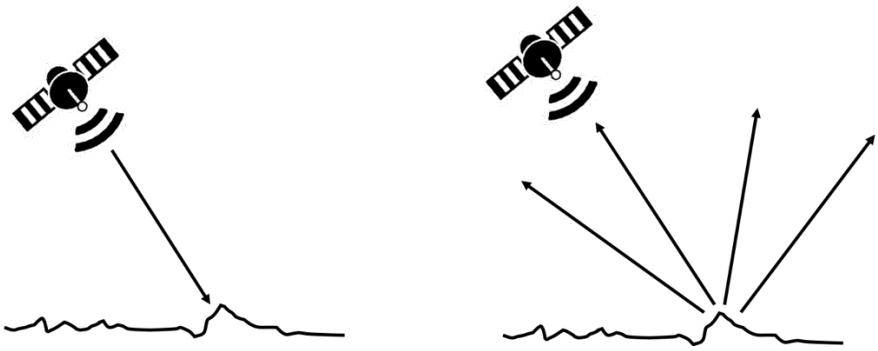


Figure 2.2. Echo of radar signal back to antenna

The wavelength frequency of the microwave can be varied from 300 KHz to 47 GHz (VHF to Ka-Band) depending on the satellite mission as shown in Table 2.1. For example, L-band is ideal for the study of wetlands because the signal penetrates trough the canopy and can sense if there is stranding water underneath. Generally, the longer wavelength, the stronger the penetration into the target. Another example can be seen in Figure 2.3 where the C-band radar is usually reflected at the canopy of forest while L-band, which has longer wavelength, can be reflected at the ground surface.

Table 2.1. SAR frequency bands and applications

Frequency band	Frequency range	Application example
VHF	300 Khz – 300 MHz	Foliage/Ground penetration, biomass
P-band	300 MHz – 1 GHz	Biomass, soil moisture, penetration
L-band	1 GHz – 2 GHz	Agriculture, forestry, soil moisture
C-band	4 GHz – 8 GHz	Ocean, agriculture
X-band	8 GHz – 12 GHz	Agriculture, ocean, high resolution radar
Ku-band	14 GHz – 18 GHz	Glaciology (snow cover mapping)
Ka-band	27 GHz – 47 GHz	High resolution radars

Other than wavelength frequency, there are several more parameters need to be considered for a study of backscatter of radar signal.

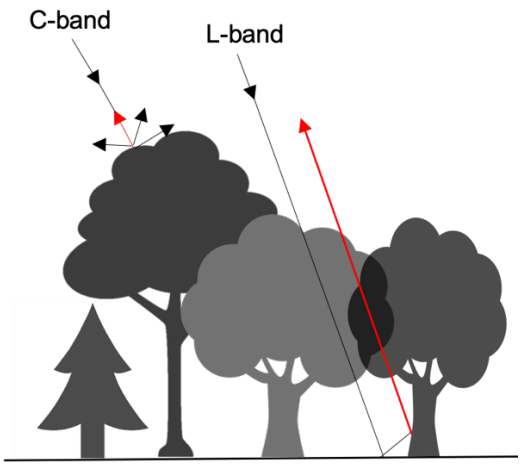


Figure 2.3. C-band and L-band radar reflectance on forest

The radar signal is polarized in horizontal and vertical, making it has four polarizations in total including, HH: horizontal transmit, horizontal receive, HV: horizontal transmit, vertical receive, VH: vertical transmit, horizontal receive, and VV: vertical transmit, vertical receive. Different polarization can determine physical properties of the object observed differently. In practice, HH and VV are the most benefitted from the double bounce scattering but the HH is more used in urban area since it is more likely to reflect onto building surface while VV is usually associated to the vegetation.

Incidence angle is also one of the factors affecting radar backscattering. It indicates the angle between the direction of illumination of the radar and the Earth's surface plane. The incidence angle will change depending on the height of the sensor, resulting in the different geometry from point to point in the range direction in an image, i.e. in the near range, the viewing geometry may be referred to as being steep, relative to the far range, where the viewing geometry is shallow.

The surface parameters including dielectric constant, surface roughness and structure and orientation of objects on the surface, also play an important role on the radar backscattering. The dielectric property of a material influences an ability to absorb microwave energy, and therefore critically affects the scattering of microwave energy. For example, the lower the dielectric constant, the more incident energy is absorbed, the darker the object will be on the image. We could simply say that the magnitude of the radar backscatter is proportional to the dielectric constant of the surface. This property can be benefit in some application such as measuring soil moisture or monitoring freeze/thaw transition of the land surface. Another thing affecting the bright/dark on the image is the surface roughness. The smooth surface causes a few signals to be reflected back to the radar's antenna while it is opposite for the rough surface. The structure and orientation of the object can also affect the backscattering by causing the double bounce or strong single bounce which can make the image appear in very bright color. Each scattering mechanisms can be summarized in Figure 2.4.

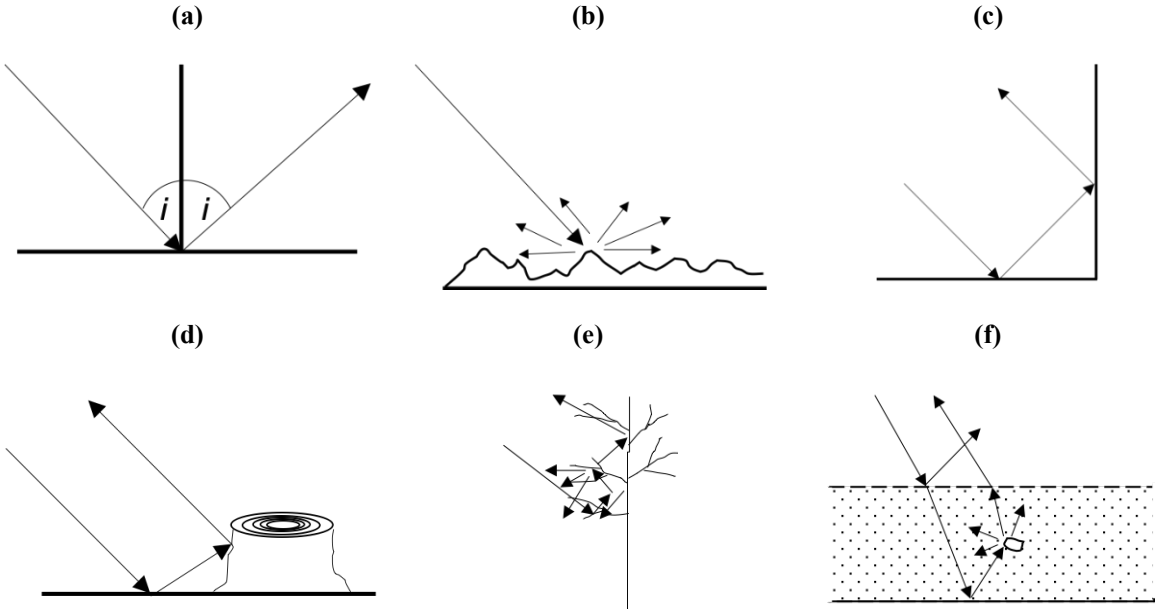


Figure 2.4. Scatter mechanic: (a) Reflection off a smooth surface, (b) Scattering off a rough surface, (c) Double bounce (man-made object), (d) Double bounce (natural), (e) Volumetric scattering (tree), (f) Volumetric scattering (layer of snow)

Because of its imaging mechanism, SAR images can suffer from geometric and radiometric distortion of the radar signal. In term of geometric distortion, since most of the satellites use side looking radar, it is inevitable to face layover, foreshortening and shadowing problems, as shown in Figure 2.5. Layover is the phenomenon happening at the very steep slope that cause the ordering of surface elements on the radar image is the reverse of the ordering on the ground. Foreshortening is usually occurred in mountainous areas which makes the mountains seem to lean towards the sensor. The shadow in SAR image can be happened when the height of the slope prevents the radar beam from reaching the surface at the far range side. In term of radiometric distortion, an image can be affected from the influence of topography on backscatter and the correction might eliminates high values in areas of complex topography. Since these phenomena are often occurred in mountain areas, it is important to consider these distortions when the interested object located at such terrains because the image intensity change in the object and the distortion might become similar.

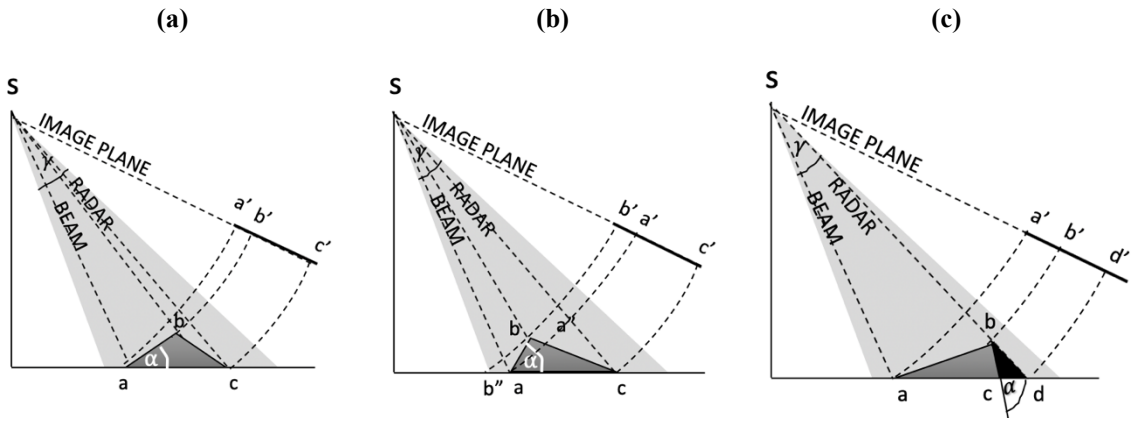


Figure 2.5. Geometric distortion: (a) Foreshortening, (b) Layover, (c) Shadowing

The SAR image always contains a granular noise called speckle noise that inherently exists in and degrades the quality of an image. The reduction of speckle can be done by multi-look processing or spatial filtering. Multi-look processing is the summing and averaging of the different looks obtained by dividing radar beam into several, narrower sub-beams. In the other hand, the spatial filtering applies a mathematical calculation on the pixel values through the moving of window over each pixel in the image.

For the SAR image interpretation, the very high intensity usually can be interpreted as a building or man-made structure because of the double bounce effect. Double bounce backscatter occurs in areas like cities where a large concentration of humanmade features exist. In urban areas, the radar signal is first reflected specularly as it encounters roads and sidewalks or surrounding ground area. The specularly reflected signal then bounces off the sides of buildings and is returned back to the SAR sensor. Double bounce reflection causes most of the radar signal to return to the sensor, resulting in high backscatter and bright areas in the SAR scene. The buildings can also cause the strong single bounce onto the roof which return the high backscatter similar to the double bounce, these phenomena are illustrated in Figure 2.6.

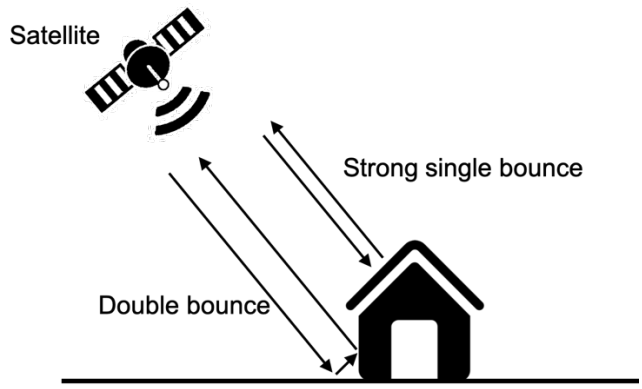


Figure 2.6. SAR signal backscatter on a building

In order to utilize the use of SAR data in urban area, the studies on the effect of the SAR properties against the building have been made. The study by Hussin [1] demonstrated the influence of polarization and incidence angle on the double-bounce effect, which showed that the corner reflector has, generally, a higher return in HH polarization. Instead, VV polarization is more sensitive to variations in the incidence angle. This analysis was conducted only on buildings that were parallel or perpendicular to the azimuth direction. In term of influence of both incidence and orientation angles on the scattering from urban environments using actual SAR airborne data, the study indicates that the buildings which are parallel to the azimuth direction have a stronger double-bounce contribution than the buildings facing away from the radar [2]. The studies in [3-4] also confirmed that the double-bounce effect gives a strong power signature to buildings with walls almost parallel to the SAR azimuth direction but decays rapidly in a narrow range of orientation angles.

2.2 Fundamental of deep learning methods

Deep learning has become one of the most important breakthroughs in artificial intelligence over the past decade. Deep learning contains a variety of methods to tackle with different problems, including neural networks, hierarchical probabilistic models, and many specific unsupervised and supervised feature-learning algorithms. The main concept of the deep learning is to let the machine learn to solve the problem itself instead of by specifying fixed conditions to the solution, which is more similar to how actual human's brain works and would allow the machine to be robust to the changing of environments of the problem. With this idea, the human work can be greatly reduced and therefore the popularity of this approach is significantly increasing. For instance, in visual recognition problem, the biggest difference between deep learning and the traditional methods is that deep learning methods

automatically learn features from a huge amount of data, rather than requiring engineering features by hand. Therefore, deep learning can conveniently learn good features for new specific tasks without much expertise and effort of designing features. Not only the simplicity of the idea, deep learning is also famous from its accurate result especially in image processing field as evidence in the Large Scale Visual Recognition Challenge in 2012 [5] where the deep learning was able to achieve the highest accuracy over numerous numbers of methods. From that time, the deep learning has been widely studied by many researchers from many fields, causing many state-of-the-art deep learning architectures such as Alex-net, VGG16, U-net, etc.

Deep learning is about learning hierarchical feature representations. Deep architectures with multiple levels attempt to learn hierarchical structures and seem promising in learning simple concepts first and then successfully building up more complex concepts by composing the simpler ones together. It accords with human's visual cognition of learning abstract concepts on top of less abstract ones. These high-level feature representations are more powerful and robust in typical visual tasks. For example, when dealing with 2D images, multiple layers of convolution are used to extract features from an image from low-level to high-level through each layer and when the last layer is processed, the network will calculate the loss with assigned loss function. There are several kinds of loss function which usage is depending on the characteristic of the dataset. The loss function determines how good the network has learned and the calculated loss will be back propagated to update the weight of the convolution filter in each layer to make it recognizes features better next time. These processes are repeated until the weight is stable and can recognize features efficiently or the testing accuracy is satisfied, then the trained model can eventually be used with any other data outside those in training set.

The most fundamental architecture in the field, that is later have been developed into many state-of-the-art deep learning architectures, is CNN. A CNN is an architecture that leverages spatial and temporal relationships to reduce the number of parameters which must be learned and thus improves upon general feed-forward back propagation training. CNNs were proposed as a deep learning framework that is motivated by minimal data preprocessing requirements. The general architecture of CNNs can be seen in Figure 2.7 where small portions of the image, as known as a local receptive field, are treated as inputs to the lowest layer of the hierarchical structure. They are then propagated through the different layers of the network and are applied with digital filtering at each layer in order to obtain salient features of the data observed. The method provides a level of invariance to shift, scale and rotation as the local receptive field allows the neuron or processing unit access to elementary features such as oriented edges or corners. The intimate relationship between the layers and spatial information in CNNs renders them well suited for image processing and understanding, and they generally perform well at autonomously extracting salient features from images.

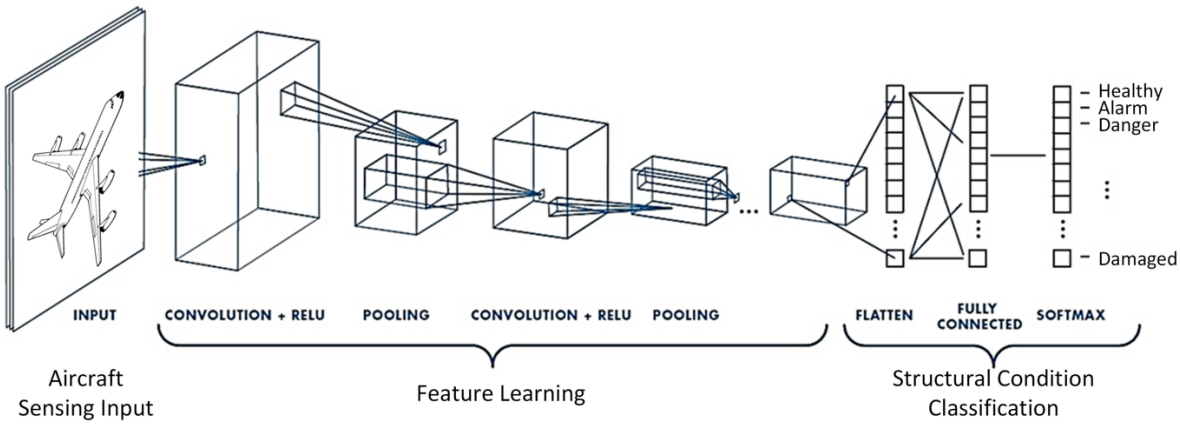


Figure 2.7. CNN (image from: https://doi.org/10.3390/s19224933)

The main focuses on deep learning in image processing are classification and segmentation tasks, which this study fall into latter category. The simplest architecture to use in segmentation is Fully convolutional network (FCN) which has proven its efficiency in many studies. Although it shows good performance in many studies, it reveals the problem when using with our objective where the image resolution is upsampled (unpooling in Figure 2.8) at the decoder part. The purpose of the decoder is to bring the image resolution back after the encoder have learned the features until the image gets very small. By trying to upsample with the deconvolution technique, it is inevitable for the loss of spatial resolution and checker board problem to occur [6]. To avoid this problem, the U-net which offer a very accurate segmentation result even in boundary areas has been proposed. U-net is one of the most used deep learning architectures that based on FCN as it has proven its ability to maintain image resolution especially at the boundary areas by the special layer called the skip connection. Skip connection stays between encoder and decoder part to pass the low-level features from each layer of encoder to corresponding layer of decoder. It helps decoders to generate prediction result more accurate as it learns the boundary information from the features that skip connection passed to it.

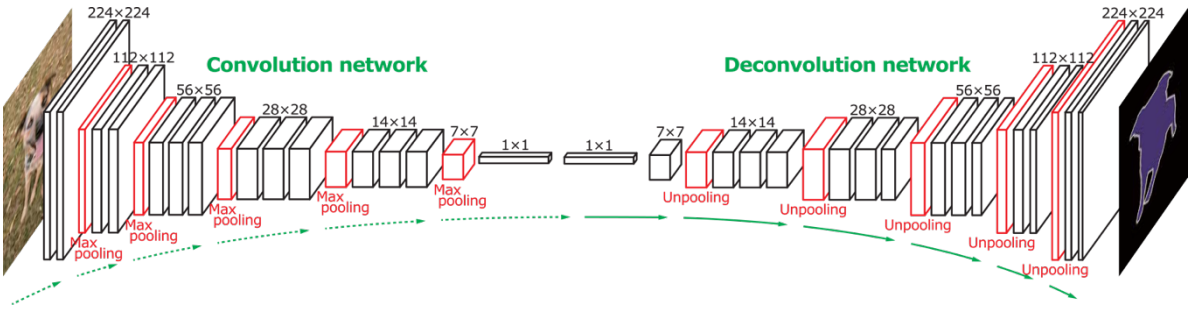


Figure 2.8. Fully convolutional network (image from:

http://cvlab.postech.ac.kr/research/deconvnet/)

In practice, remote sensing of images has been successfully applied in many fields, such as classification and change detection. Neural networks, the basis of deep learning algorithms, have been used in the remote sensing community for many years. The deep learning algorithms have achieved significant success in remote sensing community at many image analysis tasks including land use and land cover (LULC) classification, scene classification, and object detection [7-16].

For all the mentioned tasks, deep learning showed precise accuracy compared with traditional classifiers (e.g., Random forest and Support vector machine). However, by comparing with scene classification and object detection, the performance of deep learning in LULC classification is still inferior. Since the studies suggest that a supervised deep learning model (e.g., CNN) must be based on large quantities of training samples, this may be attributed to the frequent use of benchmark datasets in scene classification and object detection for previous studies. In practice, the acquisition cost of training samples is relatively high, and therefore some augmentation techniques are desirable for increasing the size or efficiency of training datasets, e.g., through transfer learning or active learning. Therefore, developments in deep learning in term of data efficiency are expected to be focused as it will have further application in practical remote-sensing images.

The deep learning models have also been adapted their implementation for non-standard image processing tasks, e.g., object-based image analysis and time-series analysis. For object-based image classification, a patch-based strategy is a generally accepted method [17-18], which integrates CNNs with Object-based Image Analysis. The critical issue with this approach is how to determine the values of the relevant parameters (e.g., patch size), because classification accuracy is largely affected by these parameter values. Time-series

analysis, a common processing task in remote sensing, has been the focus of very few studies involving deep learning. Therefore, it is necessary to further explore the application potential of deep learning in time series analysis (e.g., Landsat or Sentinel), in particular, as deep learning actually possesses some advantages for the processing of time series data.

2.3 SAR-based change detection for building

In remote sensing field, the publication on change detection especially in building detection has been very active. The most used data is a very-high-resolution optical image because it can easily tell the features of the building, allowing it to accurately detect their appearance. For example, Y. Xu et al. [19] were able to detect buildings with clear boundaries using U-net on very-high-resolution satellite imagery, these kinds of data using optical sensors unfortunately cannot capture Earth's surface when it is covered by clouds. The problem applies to some accurate building detection methods, which have a possibility to be extended for using in building change detection, that use the characteristic of the shape of building roofs from a very-high-resolution imagery [20-22]. As well as being hindered by cloud contamination, these images are also difficult to obtain especially when studying an event in the distant past.

As a result, many researchers chose to use SAR data instead to overcome the problem of cloud covering. Currently, many proposed methods in building change detection still use some kind of fixed mathematical conditions to generate the difference image (DI) between two images using mathematical operations, and then perform segmentation to extract the area of changes. Some publications use threshold-based methods for segmentation, such as Y. Ban and O. Yousif [23], who applied several thresholding criteria on the difference image

to obtain the urban change area. However, thresholding methods usually lead to false detections when the urban or non-urban area has an unordinary intensity in terms of its change behavior. With this approach, there is a possibility that the method cannot tell whether the change is appearing or disappearing of a building or even cannot tell the different between building change and change caused by seasons to vegetation fields. Another point is these methods are mostly limited the applicability to one specific acquisition type of SAR image, making it difficult to use with the data at disposal. Thus, the using of deep learning can be the perfect substitution to fulfill this objective as it is not restricted by any fixed condition and has more flexibility in using with variety of scenarios or complex scenes. As a result, methods using deep learning have been widely researched.

Publications of studies that used deep learning with SAR images [24–27] have reported excellent change detection results and prove that deep learning can be used with SAR images. S. Iino et al. [28] successfully used a convolutional neural network with an SAR image for land cover classification to find an urban distribution map for short-term change detection. However, their results included all of the changes that occurred on two occasions, as shown in Figure 2.9, regardless of the source of the changes, because they used only the information of the difference in intensities or the digital surface model. This can happen with any change detection approach that is based on the difference between images, even when using deep learning techniques such as in [29]. As a result, methods using deep learning while not having to depend on the DI must be introduced. While U-net can give a very accurate detection result [19], the publication involving using U-net on building change detection without generating of DI has not been published before. Therefore, the using of U-net for such purpose will be studied in this dissertation.

Although the deep learning has shown the promising results so far, in order to train a deep learning network efficiently, a large amount of data in both quantity and variation need to be used in the training process. However, as many satellites, especially satellites with SAR sensors, are orbiting around the earth, the images of the interested location might be captured by many different SAR sensors with various imaging setting and are difficult to pre-process or to select to use in making of time-series data for training the network to reach the maximum potential. Moreover, not only the SAR time-series images are required, the ground truths of the building of constructions are also needed in order to train a deep learning network. As the ground truths are usually created by humans, obtaining enough data for the training process is both costly and time-consuming. Thus, a new deep learning architecture that does not require high number of training data must be proposed.

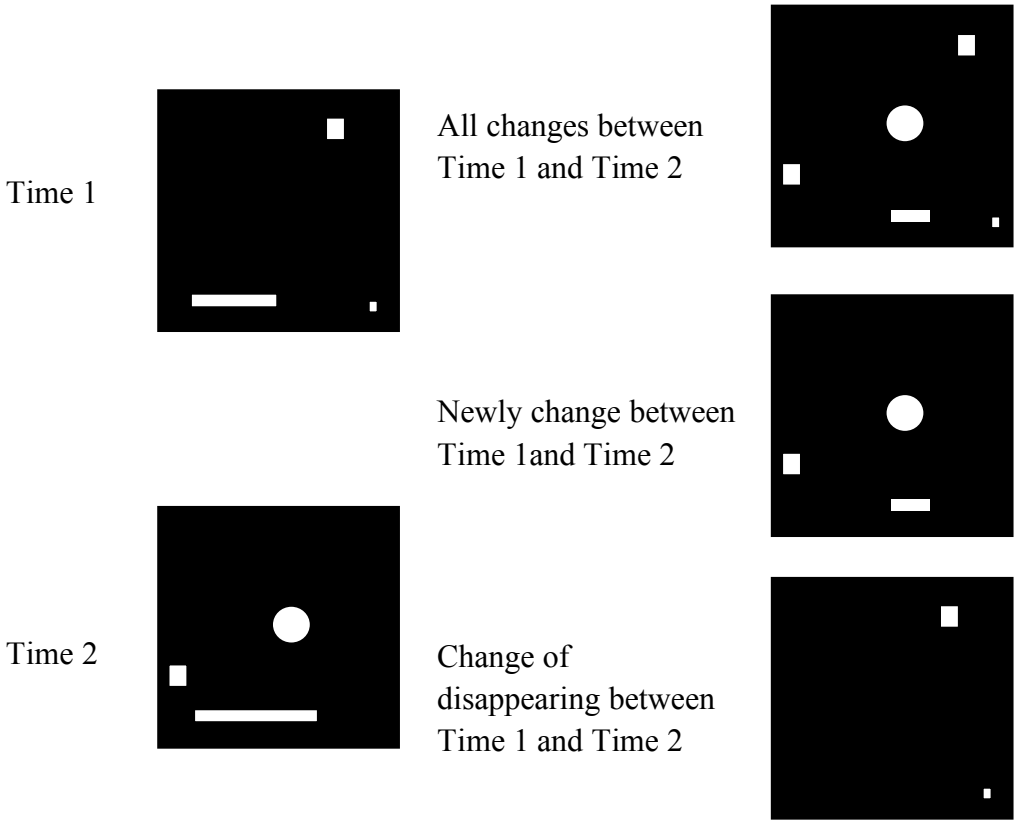


Figure 2.9. Types of changes (White part: building area; Black part: non-building area)

2.4 Summary

This chapter describe about the SAR data which is used to fulfil the purpose of this thesis. The properties and imaging mechanism including problems have been explained in this chapter. The related literatures of change detection methods including the deep learning based have been described and shown their disadvantages leads to the problems that need to be addressed.

The deep learning has shown its potential to use in the study in change detection in remote sensing image as the existing methods have been studied and summarized in this chapter. As many studies showed not enough accuracy due to the insufficient training data problem, the author need to design new deep learning architecture that can maximize the utilization of a limited number of available training data. Thus, the original problem for this thesis has been stated in this chapter that is to create the deep learning architecture that can be used to train the model for detecting newly built constructions in SAR images with only a low number of training data that robust to noisy images and can be used with different wavelength band from those in training data.

Reference

[1] Y. A. Hussin, "Effect of polarization and incidence angle on radar return from urban features using L-band aircraft radar data," in Proc. IEEE IGARSS, Florence, Italy, Jul. 10–14, 1995, pp. 178–180.

[2] Y. Dong, B. Forster, and C. Ticehurst, "Radar backscatter analysis for urban environments," Int. J. Remote Sens., vol. 18, no. 6, pp. 1351–1364, Apr. 1997.

[3] T. Kempf, M. Peichl, S. Dill, and H. Süß, “Microwave radar signature ac- quisition of urban structures,” in Proc. ITG WFMN, Chemnitz, Germany, Jul. 4–5, 2007, pp. 68–73.

[4] D. Brunner, L. Bruzzone, A. Ferro, J. Fortuny, and G. Lemoine, “Analysis of the double bounce scattering mechanism of buildings in VHR SAR data,” in Proc. SPIE Conf. Image Signal Process. Remote Sens. XIV, Cardiff, U.K., Sep. 15–18, 2008, vol. 7109, pp. 710 90Q–710 90Q–12.

[5] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[6] Aitken, A.P.; Ledig, C.; Theis, L.; Caballero, J.; Wang, Z.; Shi, W. Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize. arXiv 2017, arXiv:1707.02937.

[7] Chen, Yushi & Lin, Zhouhan & Zhao, Xing & Wang, Gang & Gu, Yanfeng. (2014). Deep Learning-Based Classification of Hyperspectral Data. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of. 7. 2094-2107. 10.1109/JSTARS.2014.2329330.

[8] X. Yu, X. Wu, C. Luo, Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework, GIScience Remote Sens., 54 (5) (2017), pp. 741-758

[9] A. Vetrivel, M. Gerke, N. Kerle, F. Nex, G. Vosselman, Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high

resolution oblique aerial images, and multiple-kernel-learning, ISPRS J. Photogramm. Remote Sens., 140 (2018), pp. 45-59

[10] A. Sharma, X. Liu, X. Yang, D. Shi, A patch-based convolutional neural network for remote sensing image classification, Neural Networks, 95 (2017), pp. 19-28

[11] A. Romero, C. Gatta, G. Camps-Valls, Unsupervised deep feature extraction for remote sensing image classification, IEEE Trans. Geosci. Remote Sens., 54 (3) (2016), pp. 1349-1362

[12] D. Marmanis, M. Datcu, T. Esch, U. Stilla, Deep learning earth observation classification using ImageNet pretrained networks, IEEE Geosci. Remote Sens. Lett., 13 (1) (2016), pp. 105-109

[13] N. Kussul, M. Lavreniuk, S. Skakun, A. Shelestov, Deep learning classification of land cover and crop types using remote sensing data, IEEE Geosci. Remote Sens. Lett., 14 (5) (2017), pp. 778-782

[14] Q. Zou, L. Ni, T. Zhang, Q. Wang, Deep learning based feature selection for remote sensing scene classification IEEE Geosci. Remote Sens. Lett., 12 (11) (2015), pp. 2321-2325

[15] Y. Chen, X. Zhao, X. Jia, Spectral-spatial classification of hyperspectral data based on deep belief network IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 8 (6SI) (2015), pp. 2381-2392

[16] G. Cheng, C. Yang, X. Yao, L. Guo, J. Han, When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs, IEEE Trans. Geosci. Remote Sens., 56 (5) (2018), pp. 2811-2821

[17] B. Huang, B. Zhao, Y. Song, Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery, Remote Sens. Environ., 214 (2018), pp. 73-86

[18] T. Fu, L. Ma, M. Li, B.A. Johnson, Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery, J. Appl. Remote Sens., 12 (2018), p. 0250102

[19] Xu, Y.; Wu, L.; Xie, Z.; Chen, Z. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. Remote Sens. 2018, 10, 144

[20] Benedek, C.; Descombes, X.; Zerubia, J. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 33–50.

[21] Shi, W.; Mao, Z.; Liu, J. Building area extraction from the high spatial resolution remote sensing imagery. Earth Sci. Inform. 2019, 12, 19–29.

[22] Konstantinidis, D.; Argyriou, V.; Stathaki, T.; Grammalidis, N. A modular CNN-based building detector for remote sensing images. Comput. Netw. 2020, 168, 107034.

[23] Ban, Y.; Yousif, O. Multitemporal Spaceborne SAR Data for Urban Change Detection in China. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2012, 5, 1087–1094.

[24] Gong, M.; Yang, H.; Zhang, P. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images. ISPRS J. Photogramm. Remote Sens. 2017, 129, 212–225.

[25] Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change Detection in Synthetic Aperture Radar Images Based on Deep Neural Networks. IEEE Trans. Neural. Netw. Learn. Syst. 2016, 27, 125–138, doi:10.1109/TNNLS.2015.2435783.

[26] Ajadi, O.A.; Meyer, F.J.; Webley, P.W. Change Detection in Synthetic Aperture Radar Images Using a Multiscale-Driven Approach. Remote Sens. 2016, 8, 482.

[27] Bazi, Y.; Bruzzone, L.; Melgani, F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. IEEE Trans. Geosci. Remote Sens. 2005, 43, 874–887.

[28] Iino, S.; Ito, R.; Doi, K.; Imaizumi, T.; Hikosaka, S. CNN-based generation of high-accuracy urban distribution maps utilising SAR satellite imagery for short-term change monitoring. Int. J. Image Data Fusion 2018, 9, 302–318, doi:10.1080/19479832.2018.1491897.

[29] Yang, M.; Jiao, L.; Liu, F.; Hou, B.; Yang, S. Transferred Deep Learning-Based Change Detection in Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 2019, 57, 6960–6973.

Chapter 3. Dataset developing for deep learning network

3.1 Introduction

The essential part of training any deep learning network is the training data. The network will learn to perform a task from the given training data. For example, if a user wants to train the network to detect an apple in the image, the large amount of apple images in various settings are required to train the network. In the other hand, if images used in training are not enough or lack of environment variation, for example, the training set only comprise of the images of an apple on a tree, the trained model might fail to detect the image of an apple on the table since it has never learned such environment before. Not to mention that it has a possibility to detect an orange as an apple if orange images are not included in the training set. As a result, it has been proven that more amount of the training data in term of quantity and diversity could create more efficient deep learning model. There is no difference in the case of using with remote sensing data. Any of the most successful models for objectives in remote sensing fields are used high amount of satellite imagery as a training data. The satellite images are expensive but they are not very hard to acquire, however, the deep learning also require the proper ground truth for each specific task, which is very hard to obtain. In most of the published researches, the ground truth is obtained from the field survey or created manually with human's eyes, which require so much times and money to complete. Generally, the popular dataset such as COCO [1] or ImageNet [2] for training successful deep learning model are comprised of more than 1 million images, this amount of data is nearly impossible to acquire for the satellite images as the satellites are orbiting around the earth all the time and it is hard to make it capture the earth surface of the specific place at the specific times. Not to mention about many delicate details of the acquisition

condition such as the incidence angle, look direction, radar wavelength band, etc. Especially for the specific objective, as the newly built constructions detection in our case, the extra requirements of the training data are that the images must be taken at the same place at least twice and they must contain the newly built construction in the images. Not only the satellite image that match the requirement are hard to obtain, the ground truth corresponding to them are even more difficult to create. This is the reason why the objective of this study is to create the deep learning network that does not require high amount of training data.

3.2 Dataset creation

3.2.1 Defining study areas

In this work, the author uses image from ALOS-PALSAR in HH polarization at a 15 m/pixel resolution in ascending orbit mode. These SAR images were acquired in the right-looking direction with an off-nadir angle of 34.3°. The images are from 3 area: Bangkok, Hanoi, and Xiamen. The Bangkok is the only area that has been used as both training and testing data, while the rest are used for only testing purpose. In addition, one more testing set is added which is Chiang Mai, Thailand, captured by the C-band Sentinel-1 satellite to see if the model trained with ALOS-PALSAR, which was captured in L-band, can detect new constructions in images from different satellites or not. While images from Sentinel-1 in this study were the same as ALOS-PALSAR in term of orbit mode and looking direction, other properties were different from those in the training data in many aspects; for instance, the resolution was 10 m/pixel and the polarization was VV. The acquisition date from each separate set is shown in Table 3.1. All dataset used in this study and their purpose are summarized and shown in Figure 3.1.

Table 3.1. Dataset used in this study

Purpose	Location	Acquisition Date of SAR Images (Time 1–Time 2)	Acquisition Satellite	Resolution (meters)
Training	Bangkok, Thailand	1 January 2008 – 15 January 2010	ALOS-PALSAR	15
		12 January 2009 – 15 January 2010	ALOS-PALSAR	15
		1 January 2008 – 12 January 2009	ALOS-PALSAR	15
Testing	Bangkok, Thailand	27 November 2008 – 15 January 2010	ALOS-PALSAR	15
		12 January 2009 – 21 November 2009	ALOS-PALSAR	15
	Hanoi, Vietnam	2 February 2007 – 13 February 2011	ALOS-PALSAR	15
	Xiamen, China	22 January 2007 – 2 November 2010	ALOS-PALSAR	15
	Chiang Mai, Thailand	9 December 2015 – 24 December 2017	Sentinel-1	10

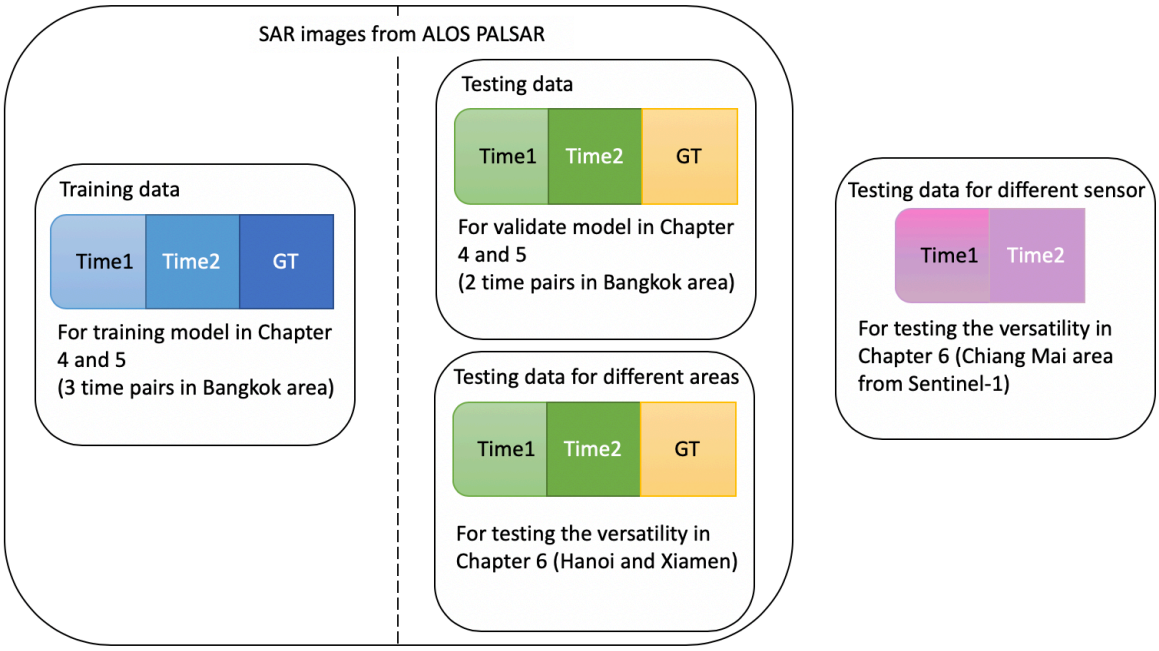


Figure 3.1. Summary of all dataset and their purpose

Bangkok was selected because it has a rapid development in the recent years. It is also located in tropical area where the optical image is difficult to use here since the cloud covering problem. The selected area is Rangsit which is located in the northern part of Bangkok. Most of the areas are rice field while the creation of villages can be seen sparsely all over the area. The author divided this area into 2 sets of data, the first set used as the training data, the second set is for testing the model. Testing areas consist of 2 specific areas. The first area, as shown in Figure 3.2a,b, is mostly rice fields areas with several large groups of villages scattered around the area. The second area has a similar characteristic with a lower number of villages but has a continuously developing, large temple (Figure 3.2c,d) as a landmark at the middle of the image.

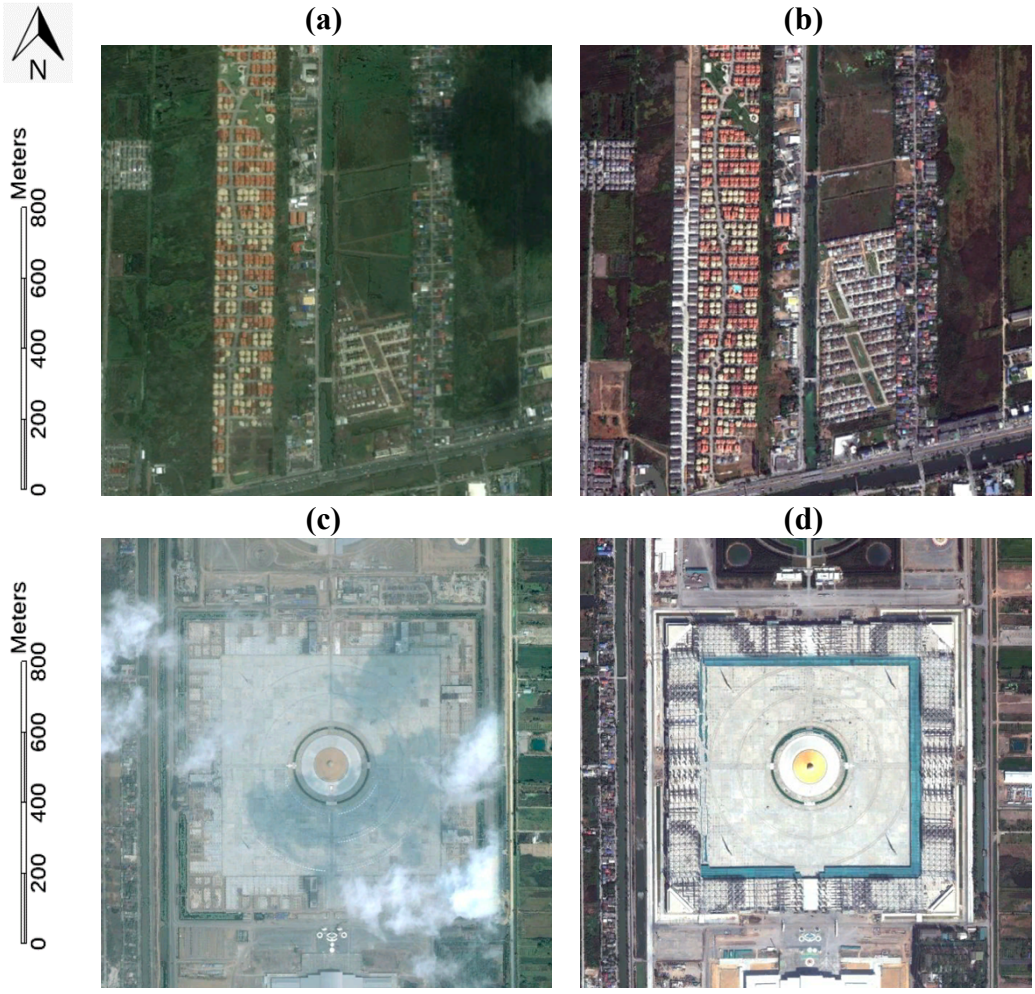


Figure 3.2. Optical image showing an example of new constructions of Bangkok testing area. The size of each image is 1.3 × 1.3 km. (a) Time 1 image of first testing area from 22 August 2008, (b) Time 2 image of first testing area from 18 December 2009, (c) Time 1 image of second testing area from 10 February 2005, (d) Time 2 image of second testing area from 18 December 2009.

The selected area is in the southern part of Hanoi in the Văn Điển town (Figure 3.3). Hanoi has a similar geological characteristic with Bangkok area where areas comprise of fields and sparse group of villages. Despite the similar characters, the buildings in this study area has much more complex shape than in Bangkok area where most of the change area is in rectangle shape.

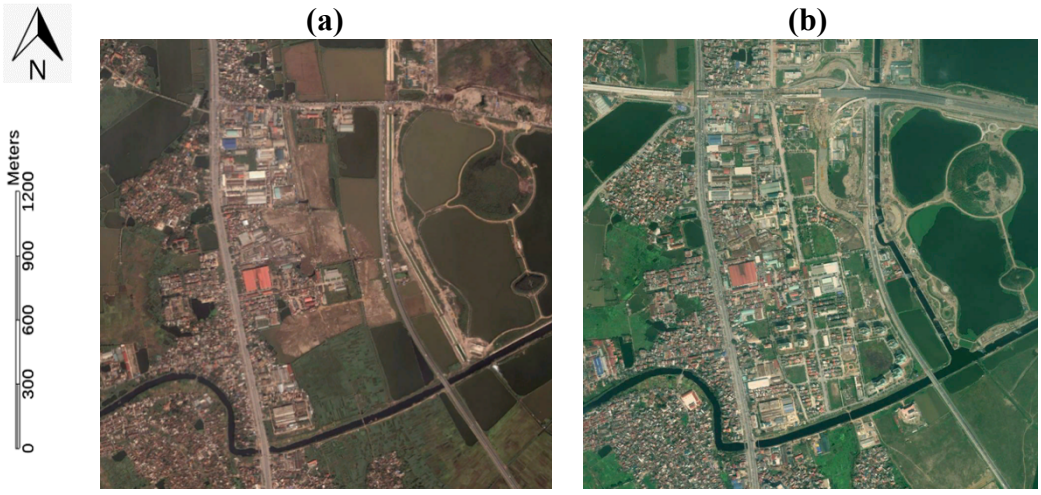


Figure 3.3. Optical image showing an example of new constructions of Hanoi testing area. The size of each image is 2 × 2 km. (a) Time 1 image from 15 November 2002, (b) Time 2 image from 9 February 2010.

Xiamen is also selected for testing, not only because it has been developing rapidly throughout the last decade, but because it also has a completely different terrain from those in Bangkok and Hanoi. Xiamen study area is an island which mean the area is surrounded by water, which is the thing that training area does not included. Some of the target change

in this area are the building that constructed on the water, making it has different intensity change behavior from a plain area to construction. With this distinct different of intensity change, the method that detection is based on the fixed condition has a high possibility to fail to correctly detect both of the building change types. Two areas were selected for the Xiamen area: one contains three bridges as a landmark of the area (Figure 3.4a,b), which the model is not supposed to detect, the another one contains the building changes on the water (Figure 3.4b), which the model should be able to detect.

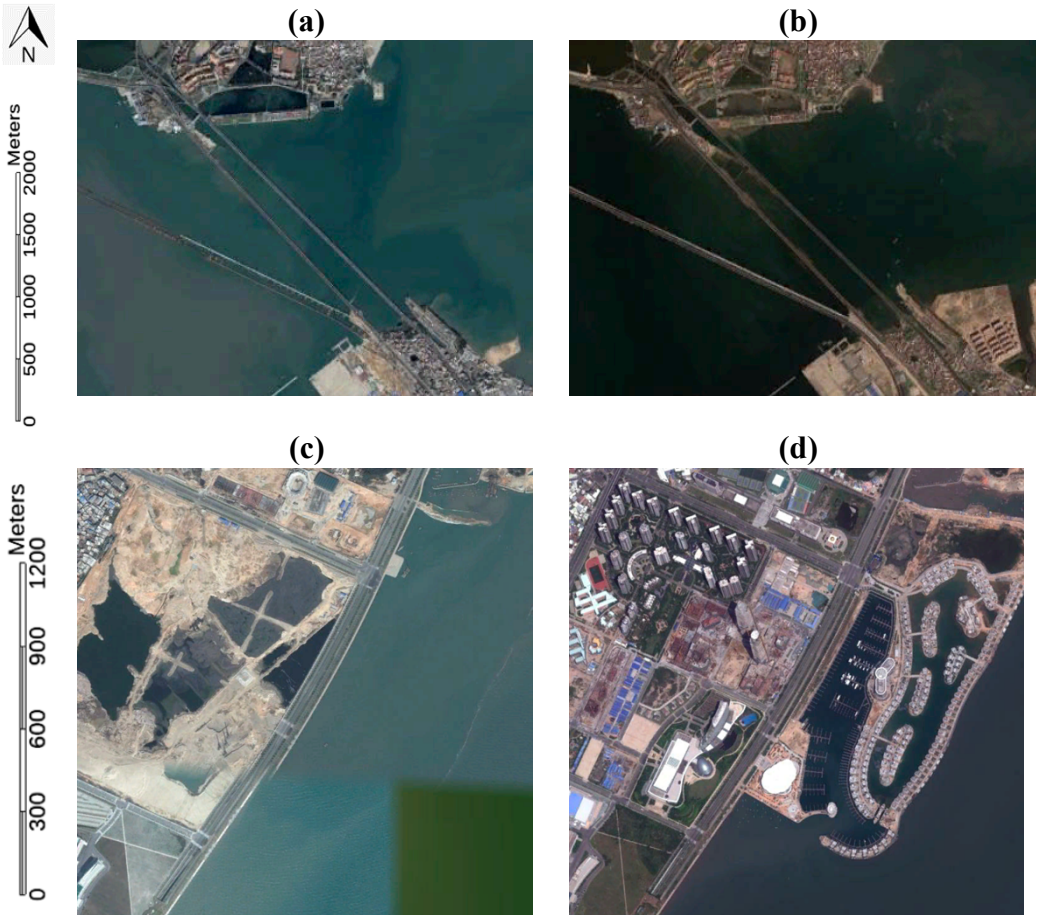


Figure 3.4. Optical image showing an example of new constructions of Xiamen testing area. The size of (a) and (b) image are 3.7 × 2.8 km and size of (c) and (d) image are 1.7 × 1.7 km. (a) Time 1 image of first testing area from 12 May 2006, (b) Time 2 image of first testing area from 29 October 2009, (c) Time 1 image of second testing area from 5 December 2006, (d) Time 2 image of second testing area from 17 September 2011.

For the further testing of the image from different sensor, the Chiang Mai, Thailand. The Chiang Mai testing area viewed from the Sentinel-1 satellite was selected to test the model on images from other satellite with other acquisition conditions, and also to test the applicability of the model on mountain area. The area in question is at Doi Suthep mountain; as seen in Figure 3.5, half of the image is mountain and another half is mainly scattered with small houses.

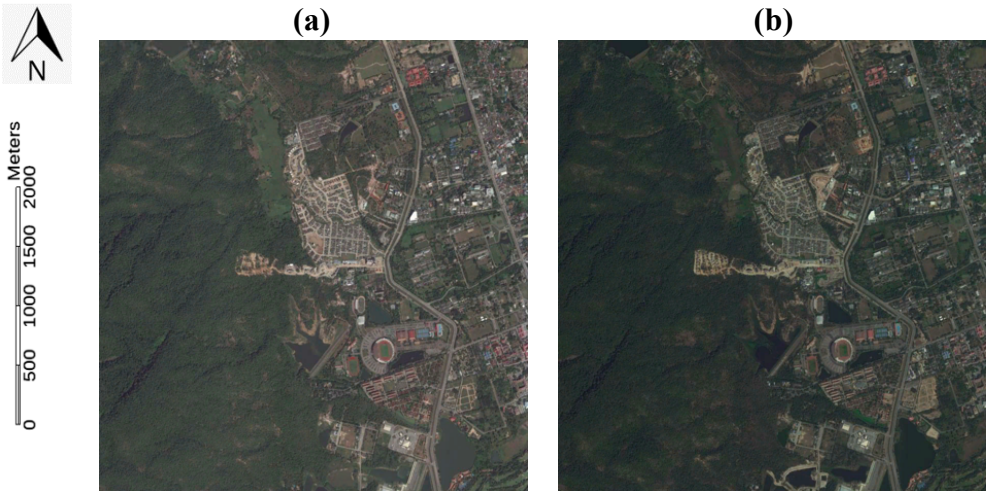


Figure 3.5. Optical image showing an example of the new constructions in the Chiang Mai testing area. The size of each image is 3.77 km × 3.97 km. (a) Time 1 image from 17 November 2015, (b) Time 2 image from 24 December 2017.

3.2.2 Ground truth creation

The ground truth data that correspond to our SAR data has been created. The process of creating the ground truth was entirely manual and done by the authors and summarized in Figure 3.6. All of the ground truths were created by drawing polygons (red objects in Figure 3.7) directly onto the optical images (examples shown in Figure 3.7a,b) available in Google Earth software after comparing the images of the same location from two different times. The criteria used for selecting the date of the optical images corresponding to Time 1 and

Time 2 of the SAR data is that the date must be as close as possible to the SAR data, while Time 1 of optical data must not exceed Time 1 of SAR data, and the Time 2 of optical data must not be before Time 2 of SAR data. Because the boundaries of our ground truths are large, the dates of the optical images from Google Earth vary depending on the area within the ground truth boundary, the lack of optical information, and the cloud cover problem; for example, the dates for the optical data selected for Time 1 of the SAR pair 1 January 2008/12 January 2009 are 18 December 2004 and 10 February 2005; for Time 2, the dates 18 December 2009, 11 April 2010, and 15 April 2010 were selected.

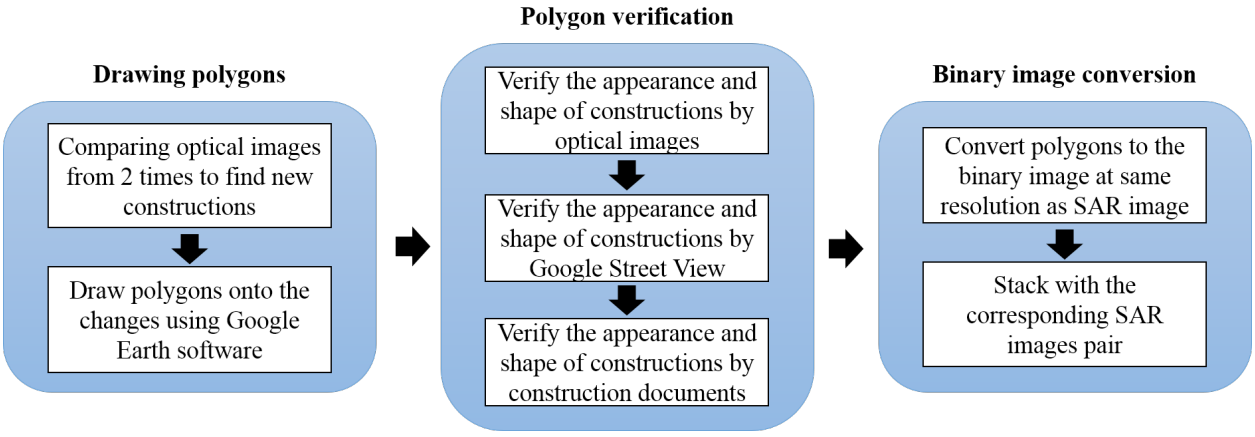


Figure 3.6. The overall process of ground truth preparation

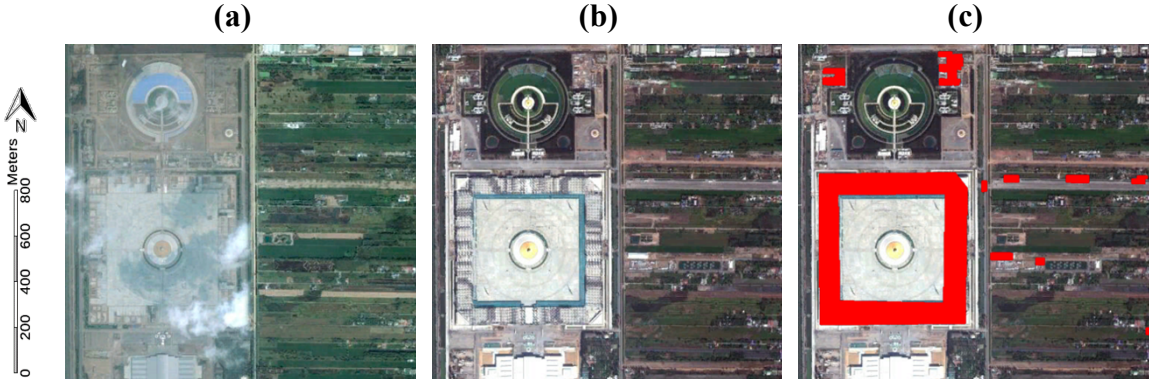


Figure 3.7. Examples of ground truth data (red polygons) overlaid on an optical image from Google Earth of a temple construction in Bangkok: (**a**) optical image from 10 February 2005, (**b**) optical image from 18 December 2009, (**c**) created ground truth.

In order to make the ground truth as precise as possible, the verification for confirming that every created polygon has a change of constructions has been made by checking it repeatedly between Time 1 and Time 2 of optical images along with the Google Street View to ensure about the appearance and shape of the constructed building. Furthermore, in an area that the appearance of the construction was unclear, the searching for documents of construction evident has been made to see the detail of the construction date and the total area and shape of constructions. Through all the process, the total time used to create the ground truth for all dataset is approximately 3 months which can be separated as 1 and a half month for the process of locating changes and drawing polygons, and another 1 and a half month for verifying and correcting shape of the created polygons. Although the strictly verification process, the ground truth sometimes contain a small error such as the date of the available optical image including Google Street View image are not exactly the same with the date of SAR images and a document of the construction cannot be found, it is possible that the shape of the construction of the created polygons are not the same with what has occurred in SAR images. These errors can affect the training of the deep learning network where it can learn the false building change features, but since the number of these errors is a very few and the training of the network is done in multiple epoch, the false of mislearning is getting lower until it barely effects the network training.

For the ground truth of training data, the average size of the created polygon is at the 22,524.87 m^2 where the minimum, maximum and standard deviation are at 882 m^2, 498,215 m^2 and 45,814.16 respectively. The number of polygons created for each dataset, including testing set, is shown in Table 3.2. Figure 3.8 shows that most of the polygons size are in the range of 6,401 – 12,800 m^2 while the range of 3,200 – 6,401 m^2 is the second most.

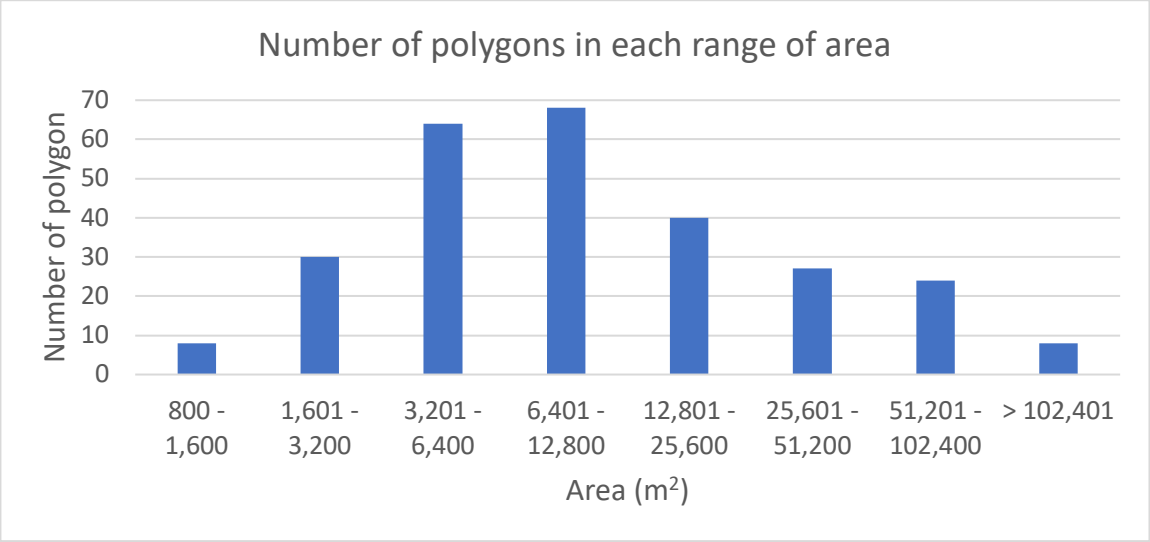


Figure 3.8. Number of polygons in each range of area in the ground truth for training data

Since the ground truth contains variety of construction sizes including small buildings from 882 m^2 to large buildings at 498,215 m^2, this ground truth is expected to make the deep learning network that trained on this dataset be able to detect wide range of size as much as possible. However, the model is expected to be able to detect building in size of 1,601 – 102,400 m^2 because this range is where the most of the created polygons is concentrated at. All the created polygons are then converted into a binary image at the same spatial resolution as the SAR image for using in the training of the deep learning network, the example can be seen in Figure 3.9 where the polygons for image pairs 12 January 2009 – 15 January 2010 has been converted to the binary image. Please note that in our case, the created ground truth has to be corresponded with the ALOS-PALSAR image which has a low resolution at 15 m/pixel, meaning the shape of smaller polygons would not be accurately represented when binarized to the same resolution as SAR image. For example, as one pixel of the ALOS-PALSAR image is 225 m^2, the smallest polygon in our ground truth at 882 m^2 has been converted to only 4 pixels in the binary ground truth image, which would be more difficult for the network to learn on compared to the larger constructions where the features

can be clearly recognized. As a result, the model trained with this dataset is not expected to be able to detect the change that is smaller than 4 pixels.

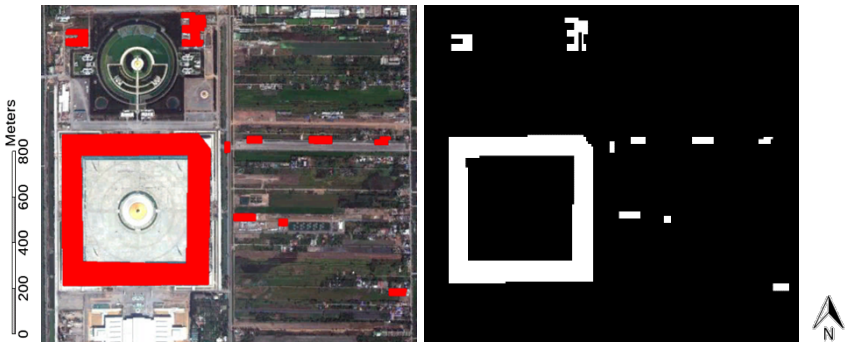


Figure 3.9. Examples of ground truth data conversion: (left) created polygons in red color overlaid on an optical image from Google Earth of a temple construction in Bangkok from 18 December 2009, (right) binary ground truth

Table 3.2. Number of polygons in each dataset.

Purpose	Location	Acquisition Date of SAR Images (Time 1–Time 2)	Number of Polygons in Ground Truth
Training	Bangkok, Thailand	1 January 2008–15 January 2010	164
		12 January 2009–15 January 2010	68
		1 January 2008–12 January 2009	38
Testing	Bangkok, Thailand	27 November 2008–15 January 2010	12
		12 January 2009–21 November 2009	16
	Hanoi, Vietnam	2 February 2007–13 February 2011	108
	Xiamen, China	22 January 2007–2 November 2010	68

3.2.3 Preparing the dataset

Before any further action, the author first reduced the speckle noise in the entire dataset using the Lee filter [3] with a filter size of 3 × 3 to prevent potential errors due to noisy

values from occurring during the training process. The author then normalized the intensity value of the data to a range of [−1, 1] to facilitate network training by avoiding inconsistent SAR intensities. To enable identification of the positions of new constructions that were built between two different times, the author selected data from the dataset acquired on different dates with the same data acquisition conditions and geolocations. The author then matched the selected data to form a pair of Time 1 and Time 2 SAR images. The images from Time 1 and Time 2 and their corresponding ground truth were then stacked and prepared for cutting into small patches for training the network. To cut the SAR images taken at two different times and the corresponding ground truth to use in network training (as Time 1, Time 2, and the ground truth) for loss calculation, the author used a sliding window with a sliding step of 50 pixels along the images to cut them into patches. Fifty was deemed the most suitable number of pixels for the sliding step because it results in a patch that is cut without skipping buildings, but is also not too repetitive. Only the patches containing at least one polygon according to the corresponding ground truth were selected for use in the training process. As a result, the dataset contains 2028 pairs after discarding patches that contained only negative pixels (please note that 10 percent of the patches from 2028 pairs were randomly selected for the validation of the model at the end of each training epoch). The patches with only negative pixels were removed because the author wants the network to learn from positive samples so that it can locate the construction of a building; also, to maintain a balance between positive and negative data during training as patches containing positive pixels also contain negative pixels. The patches cut for training the network were 256 × 256 pixels, which is a size that is suitable for detecting a building as it has the appropriate proportion of positive and negative pixels. Another thing to note is that the areas used for testing purposes in Bangkok, Hanoi, and Xiamen were manually selected

at 400 × 400 pixels, which differs from the training data and was chosen for the ease of inspection.

3.3 Unbalancing problem in training data

As stated earlier, the selecting of loss function is depending on the training set. As in some situation the network could encounter the unbalance in classes in the training data, which is when the training data contains huge different amount between each class. Imbalanced data means that the data used in training has an imbalanced distribution between the different classes. In many proposed publications, the using of balanced datasets in training show far more superior than those trained with imbalanced datasets in performance [4]. In practice, the available data is often imbalanced [5]. However, most machine learning algorithms assume a balanced distribution or the same distribution of classes in new, unlabeled data as in the known training data [6]. Such algorithms underperform if the training data does not have the same distribution as the unknown data that needs to be classified [7]. Furthermore, most machine learning algorithms aim to minimize the overall error rate which results in worse performance for the classes that are under-represented in the training data. This can have a very negative impact if the rare classes are of importance, for example in rare disease diagnostics. This can lead to the bias learning towards one specific class that is bigger than the others. However, imbalanced data has received a great deal of research interest and there are many successful methods of countering it. The solutions to this problem have been published throughout years [8-10], one of the simplest yet efficient to deal with unbalance classes is to use weighted-loss function. By weighting the loss function, the network has been told to learn more at the class that has smaller number which makes the network tends to learn features in each class more equally. The weight put in the loss function is only

applied to the training process and has nothing to do with the number of each class in the testing set.

3.4 Solution to unbalancing for training data

In the loss calculations, the loss function in our method is the cross-entropy, which normally can be calculated as

$$\boldsymbol{L = -\sum_{c=1}^{M} y_{o,c} \log\left(p_{o,c}\right)} \tag{1}$$

where M is the number of classes, y is the binary indicator (0 or 1) that represents whether class label c is the correct classification for observation o, and p is the predicted probability that observation o is of class c. However, in our case, M equals 2 because it is a binary classification (changed or unchanged). Thus, L from (1) can be derived as

$$\boldsymbol{L = -(y \log(p) + (1-y) \log(1-p))} \tag{2}$$

However, because there are far fewer positive pixels than there are negative pixels, the author considered applying class weight balancing to the loss function in order to prevent the network from excessive activations for negative parts and never for positive parts. The weighted loss function has proven its efficiency in handling imbalance class dataset [11], which is also applicable to our case. As a result, the calculation of the loss function becomes

$$\boldsymbol{L = -\left(y \log(p)\left(\omega_p\right) + (1-y) \log(1-p)\right)} \tag{3}$$

where

$$\boldsymbol{\omega_p = \frac{percentage\ of\ negative\ pixels\ in\ training\ set}{percentage\ of\ positive\ pixels\ in\ training\ set}} \tag{4}$$

This weight ω_p makes the network focus equally on how changes happened in positive areas and in negative areas. Although the negative area should be given a higher priority in training because, in most cases, the majority of the area is negative, the author want the model to be applicable to any situation regardless of the ratio of positive to negative area, so the author decided to use weights that result in the network learning both classes to an equal extent.

The value of ω_p can vary depending on the dataset used to train the network. In our case, the value is 181.5, which is the result of the rate of white pixels (new construction areas) = 0.548% and the rate of black pixels (non-changed areas) = 99.452%. The author did not use the ratio from ground truths corresponding to the whole SAR image because it contains too many black pixels in patches that were discarded (i.e., negative patches) and thus excluded from the network training process; thus, the ground truth ratio would not match the ratio received by the network from the training set.

3.5 Summary

The process of creating the training data for the study area in this thesis is described in this chapter. The study area is Bangkok which has a rapid development throughout years. The images of Bangkok were cut into patches and the ground truth were created by the manually hand drawing. Samples of training data can be seen in Appendix A. This chapter also describe the problem of the current training data, which is an unbalance classes problem, and the way to counter it.

Reference

[1] Lin TY. et al. (2014) Microsoft COCO: Common Objects in Context. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham

[2] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[3] Lee, J.S. Speckle analysis and smoothing of synthetic aperture radar images. *Comput. Graph. Image Process.* **1981**, *17*, 24–32.

[4] Masko, D. & Hensman, P. The impact of imbalanced training data for convolutional neural networks. Bachelor thesis, KTH, School of Computer Science and Communication (2015).

[5] Chao Chen, Andy Liaw, and Leo Breiman. Using Random Forest to Learn Imbalanced Data. Tech. rep. Department of Statistics, University of Berkeley, 2004.

[6] Slobodan Vucetic and Zoran Obradovic. “Classification on Data with Biased Class Distribution”. English. In: Machine Learning: ECML 2001. Ed. by Luc De Raedt and Peter Flach. Vol. 2167. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2001, pp. 527–538.

[7] Haibo He and Edwardo A. Garcia. “Learning from Imbalanced Data”. In: IEEE Transactions on Knowledge and Data Engineering 21.9 (2009), pp. 1263–1284.

[8] Khan, S. H., Hayat, M., Bennamoun, M., Sohel, F. A., & Togneri, R. (2017). Cost-sensitive learning of deep feature representations from imbalanced data. IEEE transactions on neural networks and learning systems, 29(8), 3573-3587.

[9] Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks, 106, 249-259.

[10] Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221-232.

[11] Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. Expert Syst. Appl. 2017, 73, 220–239.

Chapter 4. Newly built construction detection with U-net

4.1 Introduction

While conventional methods are unable to locate the position of the new buildings without including other kinds of changes especially when facing unordinary intensity change behavior, the using of deep learning can be the perfect substitution to fulfill the thesis's objective as it is not restricted by any fixed condition and has more flexibility in using with variety of scenarios or complex scenes. To date, deep learning has been highly effective, especially in the image processing field. One of the most successful deep learning networks that is to be considered using in this study is the U-net [1]. The U-net, proposed in 2016 for the purpose of medical image segmentation, was built on the basis of adding a skip connection to the fully convolutional network (FCN) [2] between the encoder part and decoder part. With the skip connection, the decoders can receive a low-level feature from the encoder and form the output without losing boundary information in the process. Because of its precisely predicted output at the boundary part of an image, it is now one of the most cited papers in the deep learning field. In our case, it is extremely important to preserve the boundary information because SAR data do not provide very clear information; this is because the observation mechanism of SAR is completely different from those of other sensors. As study on building change detection using U-net in the time-series SAR data without generating difference image still does not exist up to our knowledge, the author would like to test its performance in this study. Because the U-net is the deep learning architectures that based on FCN model, in order to express the reasons for its selection among all other models, the background of some other FCNs needs to be explained first.

The FCN is an architecture built only upon locally connected layers, such as the convolution, pooling, and upsampling layers. The network is usually divided into an encoder part and a decoder part. The encoder is responsible for gathering the information or features of objects in an input image, while the decoder is for recovering spatial information. One of the best examples of FCN architecture is SegNet [3], which was proposed for the semantic segmentation of an RGB image. The architecture consists of the same number of encoders and decoders, and each encoder applies convolution, batch normalization, ReLU, and max pooling to downsample the result. The decoder carries out almost the same procedure as the encoder, but without a ReLU step and with upsampling instead of downsampling. The output of the last decoder is then subjected to the Softmax function to generate the segmentation prediction result.

The architecture of the U-net is very similar to that of SegNet, but with an additional skip connection between each corresponding encoder and decoder. The skip connection makes a huge difference. Without a skip connection, the output prediction result lacks sharpness around the boundary areas, which is especially crucial for the SAR images in our case. Although comparing the result from the U-net with that from SegNet would be informative, it is impossible to generate the output using SegNet because features are too blurry to be identified. The result of using SegNet indicates that the skip connection is very important when dealing with images without significant sharpness, as is the case for our dataset.

Besides SegNet, any other FCN that involves a deconvolution layer [4] as an upsampling layer cannot generate a suitable result, as the checkerboard phenomenon will occur [5]. Therefore, it would be difficult to compare the results of these FCNs with those of the U-net.

The network created for this study is shown in Figure 4.1. Each encoder block consists of a convolution–BatchNorm–ReLU layer. The values of the number of channels, spatial filter size, and stride size of the convolution filters in each step are shown in Figure 4.2. As the modules are in the form of convolution–BatchNorm–ReLU [6], it is noted that the first layer in the encoder does not apply BatchNorm. As the author followed the method applied by Isola et al. [7], in the encoder, all ReLU functions are leaky with a slope of 0.2, while the ReLU functions in the decoder are not leaky. The dropout rate is 0.5. The skip connections in the U-net architecture were placed to concatenate activations between each layer i in the encoder and layer $n - i$ in the decoder, where n is the total number of layers. The concatenation leads to a change in the number of channels in the decoder. At the last layer in the decoder, a convolution function is applied to map the output, followed by a sigmoid function.

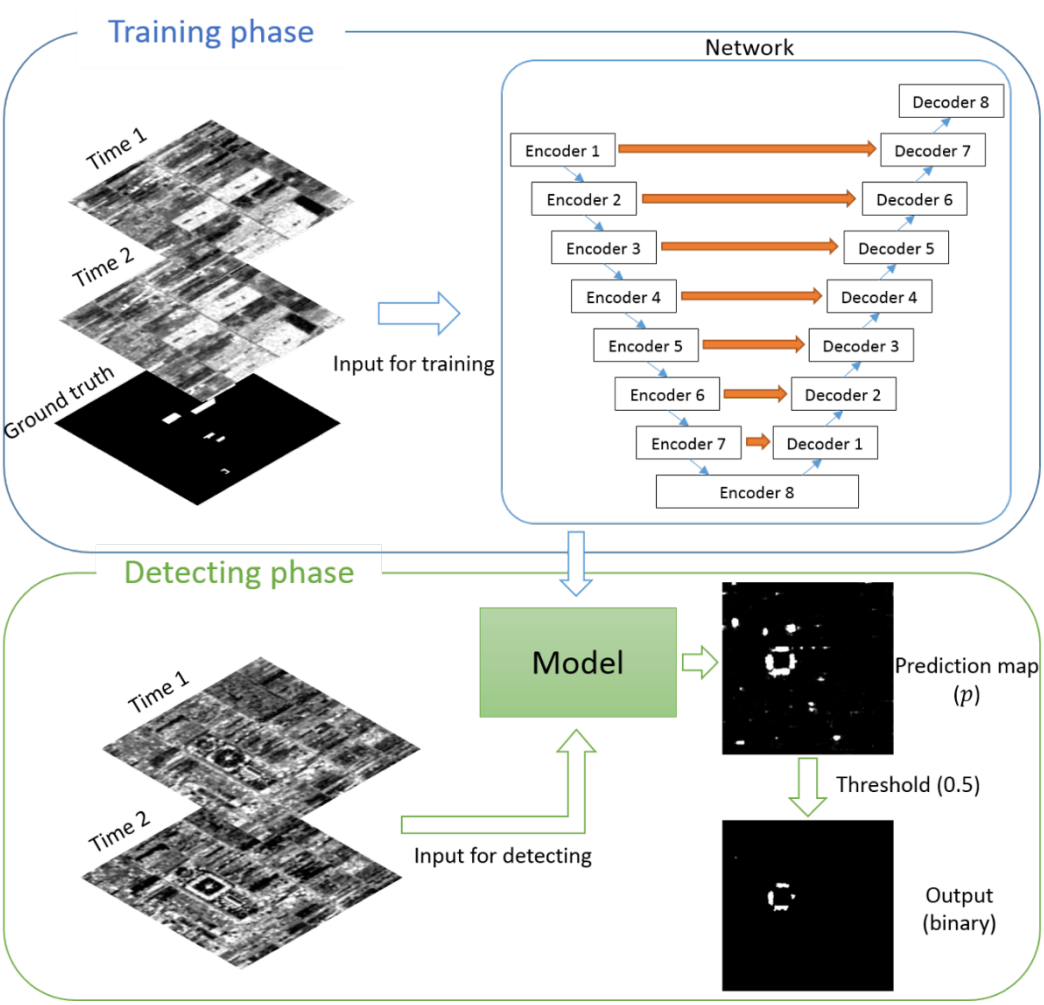


Figure 4.1. Process of detecting newly built constructions.

The simplest architecture to use in segmentation is FCN. But with using of the deconvolution technique, it is inevitable for the loss of spatial resolution and checker board problem to occur. U-net is developed based on the FCN and became one of the most used deep learning architectures as it has proven its ability to maintain image resolution especially at the boundary areas by the special layer called the skip connection. Skip connection stays between encoder and decoder part to pass the low-level features from each layer of encoder to corresponding layer of decoder. It helps decoders to generate prediction result more accurate as it learns the boundary information from the features that skip connection passed to it. The architecture of U-net is shown in Figure 4.2.

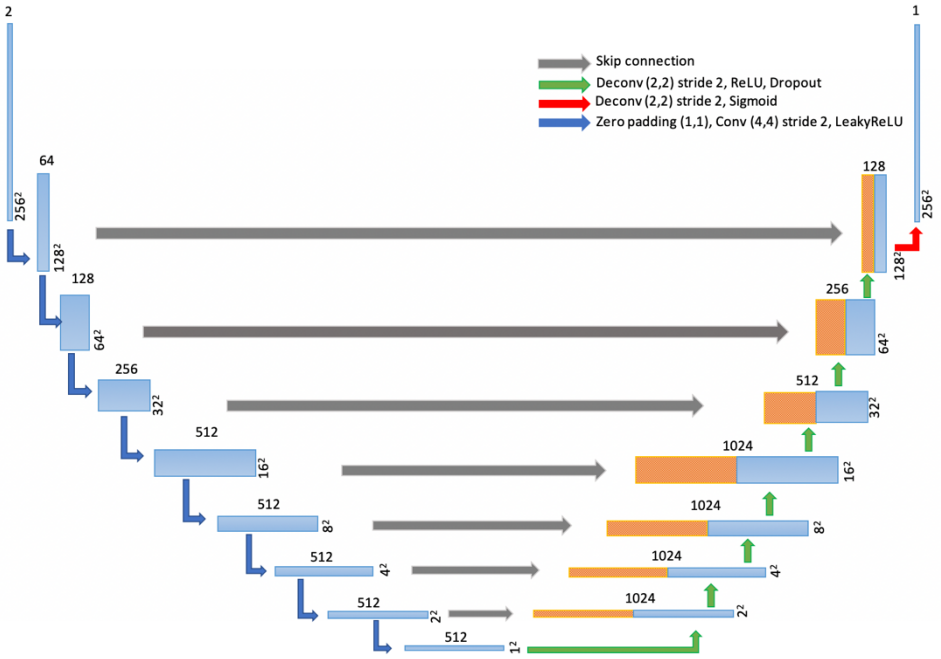


Figure 4.2. Detail of U-net architecture

The FCNs are not used in the comparison in this section because they cannot generate a decent detection result, as shown in Figure 4.3. Please note that in Figure 4.3, the range of prediction value of SegNet is [0.39 × 10−2, 1.01 × 10−2], while that of the network based on the U-net is [8.39 × 10−6, 0.99]. As the prediction range of SegNet is very small, it is difficult to generate the binary output map as the proper threshold value cannot be obtained.

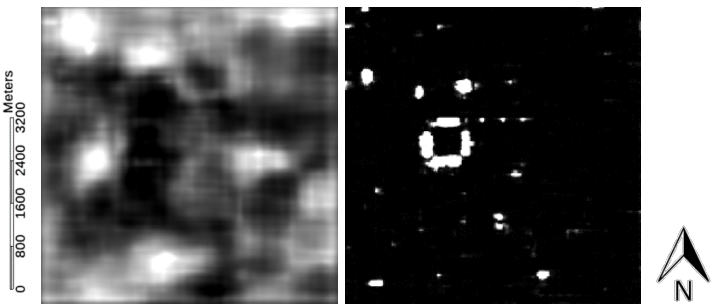


Figure 4.3. Comparison between prediction map of using SegNet (left) and U-net (right).

4.2 Experimental Result of Bangkok Testing Site

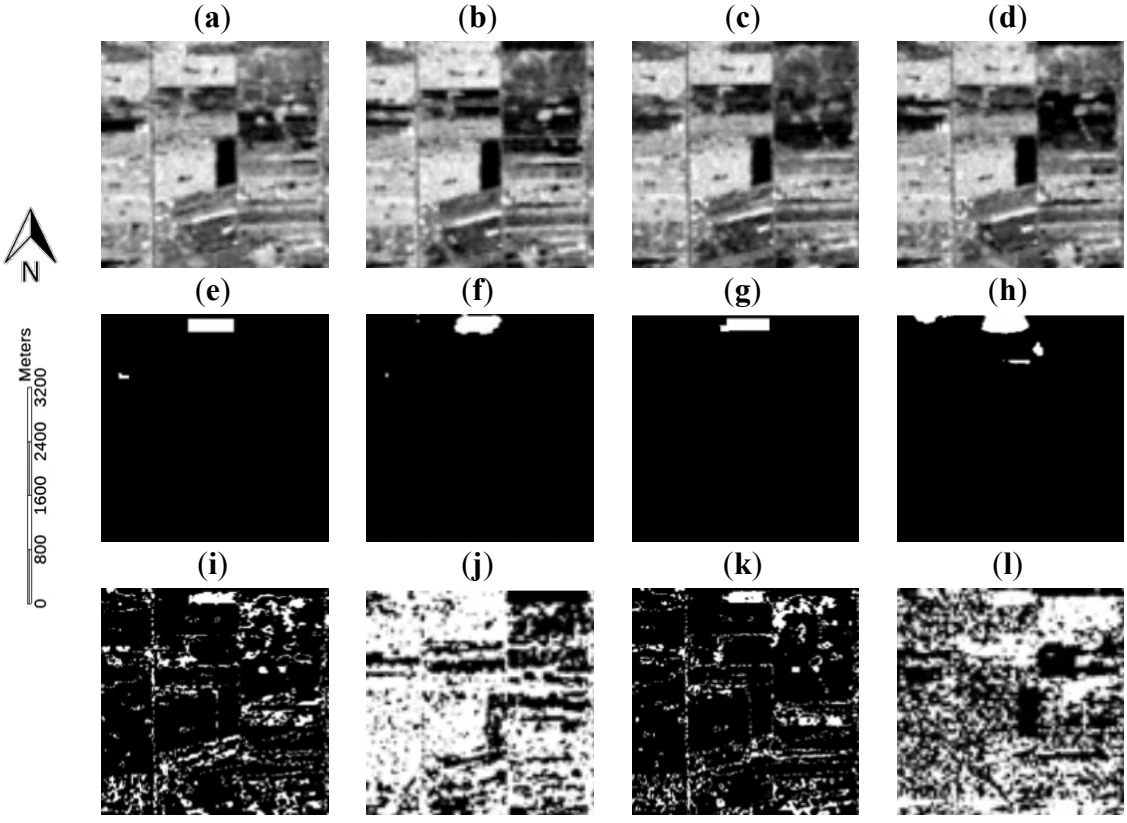
To test the ability of U-net in detecting new construction, the experiment on Bangkok testing area has been conducted. The U-net is trained by the created training set from chapter 3 in ordinary chronological order. The testing is on the Bangkok test site including two date pairs: 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009. The results of the model shown in this section are the binary maps obtained from prediction maps with a threshold of 0.5. The model used in this experiment was trained with a weight of = 181.5. The results of U-net model are shown for two different areas in Figures 4.4 and 4.5. The result of the U-net is compared with the results using fuzzy c-means (FCM) clustering [8] and Otsu thresholding [9].

The accuracy in this experiment, as well as the rest of the paper, was calculated in the form of overall accuracy, precision, recall, F measure, F1 measure, Kappa, intersect over union (IOU), false negative (FN) rate, and false positive (FP) rate. The false negative rate was obtained by the number of pixels that were in the ground truth, but not in U-net predicted result, multiplied by 100 and then divided by the total number of positive pixels in the ground truth. The false positive rate was the number of pixels that were not in the ground truth, but were in U-net predicted result, multiplied by 100 and then divided by the total number of negative pixels in the ground truth. The calculation of each validation method, excluding the false negative and false positive rates, is shown in Table 4.1. The TP in Table 4.1 stands for true positive, while TN stands for true negative. Please note that the β value of the F measure was 0.3. The accuracies that should be focused on are kappa and IOU. Kappa is used to measure on classifier performance, especially on imbalanced data set, which is suitable for our case, while IOU is used to tell how much the detection result is intersect with the ground truth. For the objective of detecting the change in Bangkok area,

which is the same area as in training data, the Kappa and IOU should be at least 0.4 and 0.3 respectively.

Table 4.1. The calculation of each validation method. IOU—intersect over union; TP–true positive; TN–true negative.

Validation Method	Calculation
Overall accuracy	$\boldsymbol{Overall\ accuracy} = \dfrac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}}$
Precision	$\boldsymbol{Precision} = \dfrac{\boldsymbol{TP}}{\boldsymbol{TP + FP}}$
Recall	$\boldsymbol{Recall} = \dfrac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$
F measure	$\boldsymbol{F_\beta} = (\mathbf{1} + \boldsymbol{\beta^2}) \cdot \dfrac{\boldsymbol{precision \cdot recall}}{(\boldsymbol{\beta^2 \cdot precision}) + \boldsymbol{recall}}$
F1 measure	$\boldsymbol{F_1} = \mathbf{2} \cdot \dfrac{\boldsymbol{precision \cdot recall}}{\boldsymbol{precision + recall}}$
Kappa	$\boldsymbol{Kappa} = \dfrac{\boldsymbol{Observed\ agreement - chance\ agreement}}{\mathbf{1} - \boldsymbol{chance\ agreement}}$
IOU	$\boldsymbol{IoU} = \dfrac{\boldsymbol{target} \cap \boldsymbol{prediction}}{\boldsymbol{target} \cup \boldsymbol{prediction}}$



SAR pair: 27 November 2008/15 January 2010

SAR pair: 12 January 2009/21 November 2009

Figure 4.4. Results of the Bangkok site in the first area. The resolution of each image is 3 km × 3 km (for SAR pairs 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009, respectively: (**a**,**c**) Time 1 SAR image, (**b**,**d**) Time 2 SAR image, (**e**,**g**) ground truth, (**f**,**h**) result of U-net, (**i**,**k**) result of fuzzy c-means (FCM), (**j**,**l**) result of Otsu).

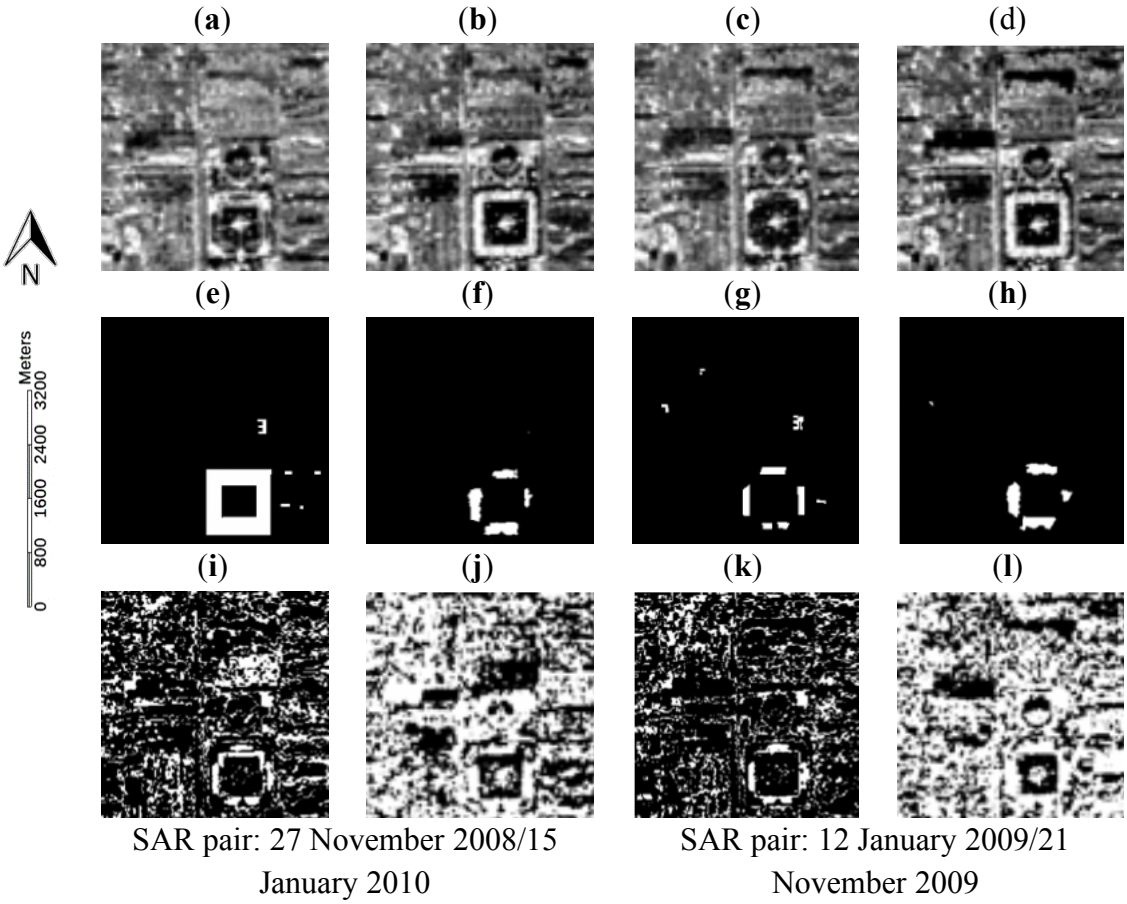


Figure 4.5. Results of the Bangkok site in the second area. The resolution of each image is 3 km × 3 km (for SAR pairs 27 November 2008/15 January 2010 and 12 January 2009/21 November 2009, respectively: (a,**c**) Time 1 SAR image, (**b**,**d**) Time 2 SAR image, (**e**,**g**) ground truth, (**f**,**h**) result of U-net, (**i**,**k**) result of FCM, (**j**,**l**) result of Otsu).

From the results of the first test area, in which paddy fields account for the majority of the area, the model can predict the construction of buildings while avoiding the changes in

paddy fields caused by seasonal effects. On the other hand, while both FCM and Otsu can capture most of the building changes, they fail to ignore the changes in other parts; this is especially the case for Otsu, which is very sensitive to intensity changes, resulting in about half of the image being detected as a building change.

Similar to the first area, the changes in the second area in paddy fields are ignored, while the construction of the temple (the big square object in Figure 4.5e) and surrounding constructions are detected. The results from the FCM and Otsu methods are similar to those for the first area—they fail to detect only the building changes.

The accuracy of each method for the Bangkok test site is shown in Table 4.2.

Table 4.2. Accuracy of each model in the Bangkok area. FCM—fuzzy c-means.

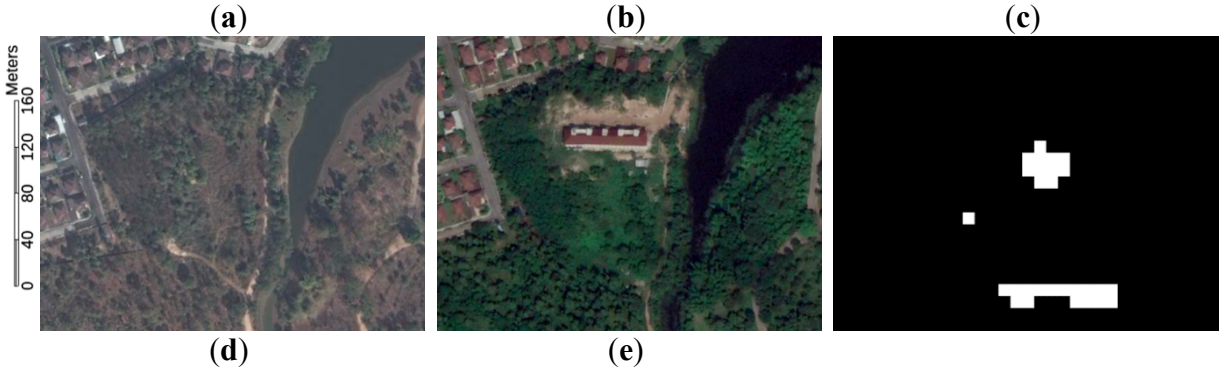
Validation Method	U-net	FCM	Otsu's Threshold
False negative	55.8006	51.4676	21.8357
False positive	0.4033	14.8646	58.3693
Overall accuracy	99.04%	84.77%	42.00%
Precision	0.5269	0.0321	0.0134
Recall	0.4420	0.4853	0.7816
F measure	0.5187	0.0348	0.0146
F1 measure	0.4807	0.0602	0.0264
Kappa	0.4759	0.0422	0.0068
IOU	0.3164	0.0311	0.0134

Despite the very high overall accuracy, U-net model has quite a high false negative rate, which means that it detects buildings as being smaller or in the wrong shape compared with those in the ground truth. However, the low false positive rate means that it has a very low chance of detecting other types of changes as a building change, and this is the target of our research. Other accuracies are not very high, but they are all at an acceptable level, especially when compared with the FCM and Otsu methods.

4.3 Problem of U-net versatility

Although U-net model can generate the prediction output that identifies building constructions in the Bangkok area, the training set for the model is also from the Bangkok area and they all are acquired from the same ALOS-PALSAR in L-band, so its applicability to other datasets can be questioned. Besides, in practice, the image of the desired area at the desired acquisition time may not be available, therefore another experiment has been conducted with the image that are from another satellite with completely different setting to test that whether U-net model can be used globally. It is tested with the image from Sentinel-1 which is captured in C-band. Moreover, the image was captured in VV polarization, which the applicable to urban area is slightly lower than HH polarization, as in training data images, since the reflectance on the building is weaker than in HH polarization.

The Chiang Mai area, Thailand, has been selected where the terrain is mountain, which are completely different from those in training data, to see if it can detect constructions as effectively as it does in the Bangkok area. Please note that since the area contains mountain part, there is a possibility that the image will be suffered from geometric distortions, such as foreshortening, layover and shadowing. The detection results from this experiment are shown in Figure 4.6-4.8.



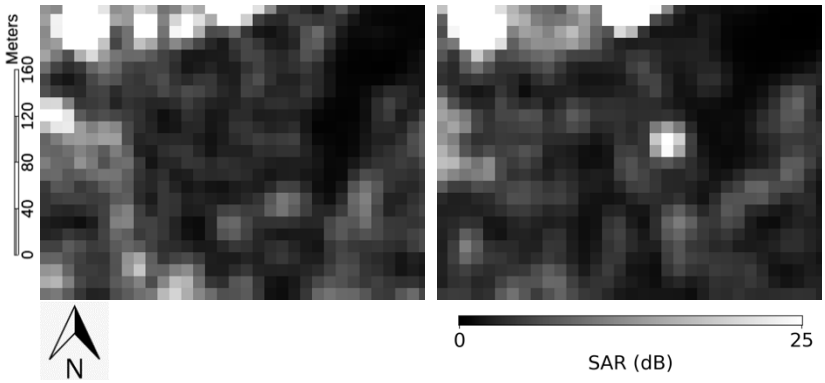


Figure 4.6. Detection results of the first area in Chiang Mai at 18°51’23.49”N 98°57’17.90”E. The size of each image is 0.32 × 0.23 km. (**a**) Time 1 optical data; (**b**) Time 2 optical data; (**c**) result of U-net; (**d**) Time 1 SAR data ‘Copernicus Sentinel data [2015]’; (**e**) Time 2 SAR data ‘Copernicus Sentinel data [2017]’.

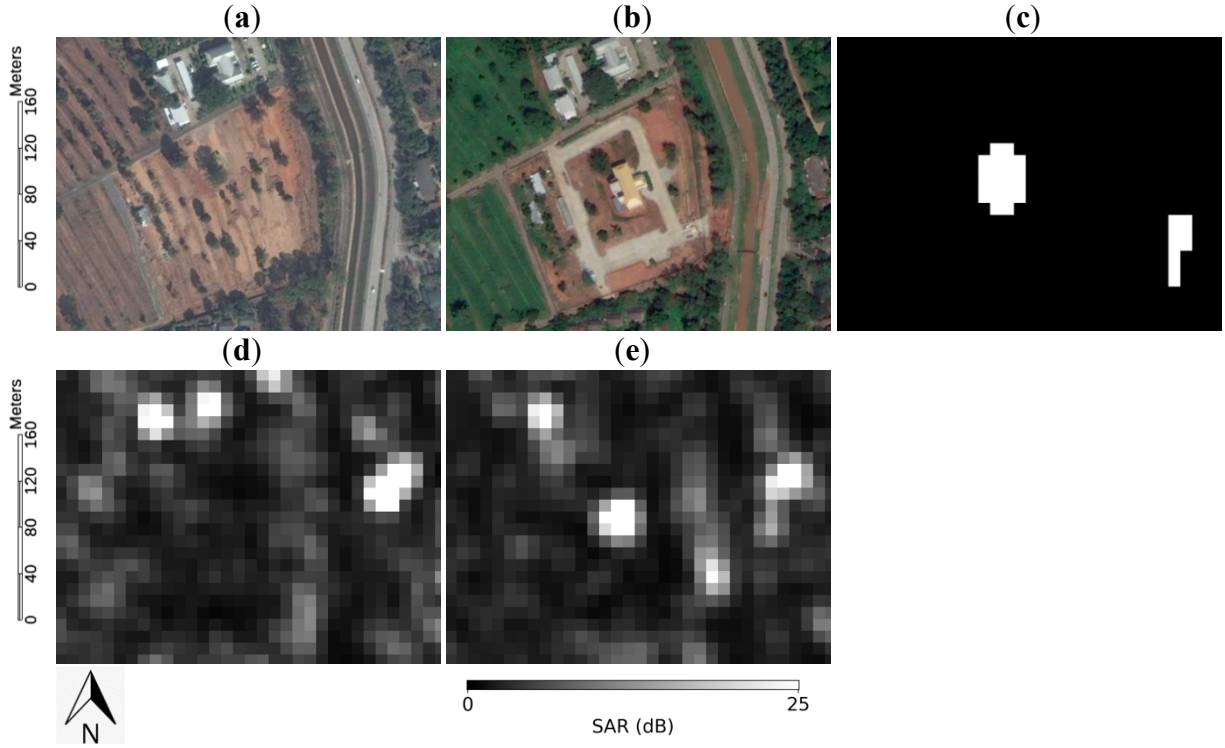


Figure 4.7. Detection results of the second area in Chiang Mai at 18°51’22.36”N 98°57’40.70”E. The size of each image is 0.32 × 0.23 km. (**a**) Time 1 optical data; (**b**) Time 2 optical data; (**c**) result of U-net; (**d**) Time 1 SAR data ‘Copernicus Sentinel data [2015]’; (**e**) Time 2 SAR data ‘Copernicus Sentinel data [2017]’.

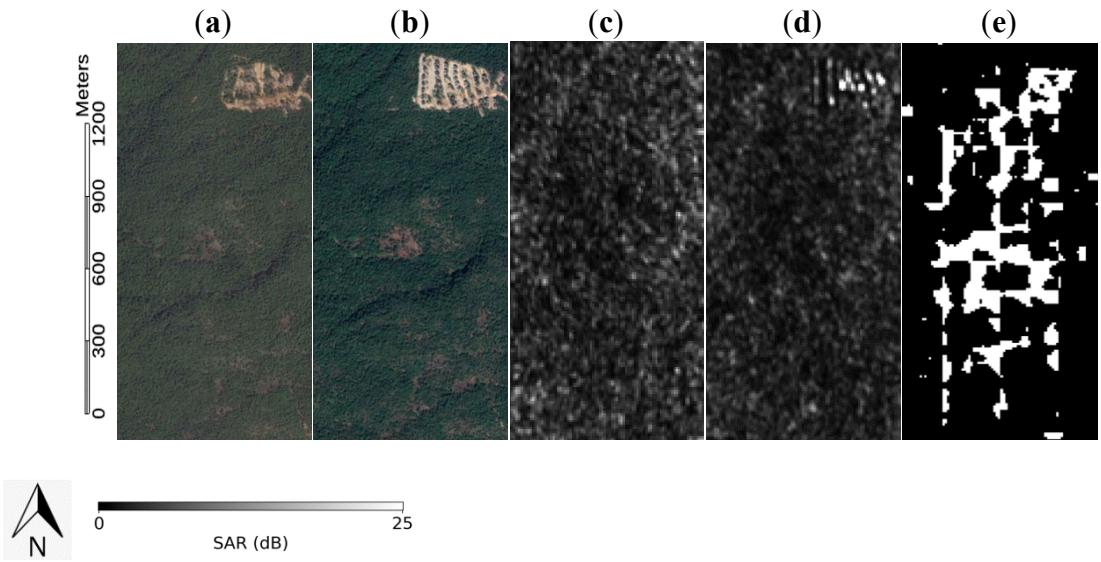


Figure 4.8. Detection results of the third area in Chiang Mai at 18°50'48.34"N 98°56'55.95"E. The size of each image is 0.8 × 1.75 km. (**a**) Time 1 optical data; (**b**) Time 2 optical data; (**c**) Time 1 SAR data 'Copernicus Sentinel data [2015]'; (**d**) Time 2 SAR data 'Copernicus Sentinel data [2017]'; (**e**) result of U-net.

While U-net can detect new constructions in first and second areas, the results also show that the U-net has generated false positive results. In the bottom part of Figure 4.6c, U-net mistakenly detected forest area as a building. Please note that the high intensity spot in the middle of Figure 4.6e is not the building. In Figure 4.7, U-net correctly detected a building in the center of the image. However, another object detected was an existing road in the right side of Figure 4.7c, which was a false detection. In Figure 4.8, even though U-net was also able to detect construction in Sentinel-1 data, as seen in the top right corner area of Figure 4.8e, it failed to handle data containing changes in mountain areas and ended up involving them in the detection result instead.

It is also worth mentioning that the author tried to randomly reduce the number of training sets from 2028 pairs to 1500 pairs and 500 pairs, respectively, to observe the learning capability of U-net in slightly lower training set situations and very low training set

situations. For each number of training sets, the networks were trained four times with four different randomly selected training pairs, and then tested with the Bangkok testing site. The results of this experiment are shown in Table 4.3 as the averages of four times the testing results. As expected, in the case of 1500 training pairs, the accuracies of U-net dropped from when trained with 2028 pairs (Table 4.2) but still is at the acceptable level. The use of 500 training pairs indicates that U-net cannot be trained with a very small dataset, as is reflected in the very low accuracies as can be seen in Figure 4.9 and 4.10. Although the accuracy of the training of 1500 and 2028 pair of training data met the expectation of 0.4 Kappa and 0.3 IOU accuracy, the accuracies of using 500 pairs training data are far from satisfy criteria where it only gets 0.15 Kappa and 0.092 IOU.

Table 4.3. Accuracies of the models in the different number of training data at the Bangkok site.

Validation Method	500 pairs	1500 pairs
False negative	27.512	47.232
False positive	17.970	0.683
Overall accuracy	81.934%	98.848%
Precision	0.100	0.492
Recall	0.725	0.528
F measure	0.107	0.488
F1 measure	0.165	0.485
Kappa	0.150	0.480
IOU	0.092	0.321

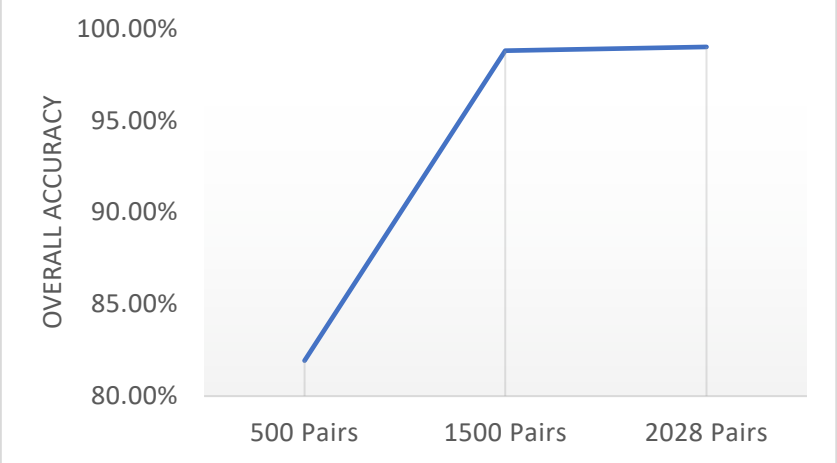


Figure 4.9. Overall accuracy in each number of training set

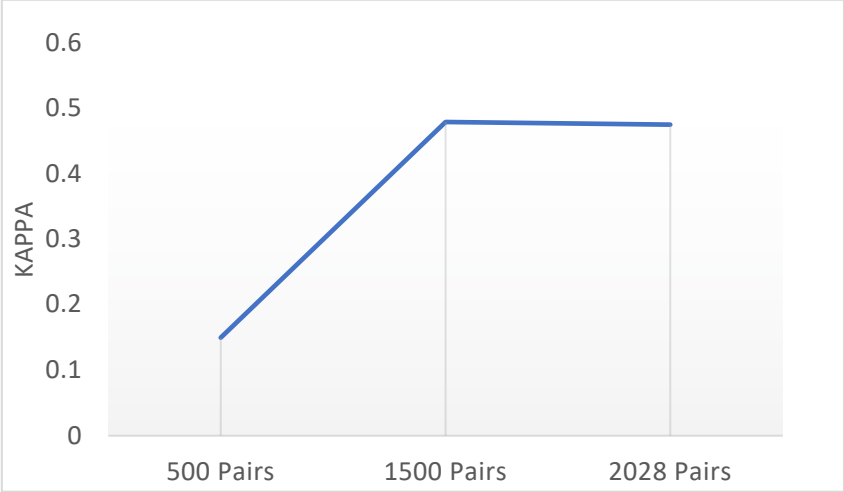


Figure 4.10. Overall accuracy in each number of training set

As the experiment on using with other terrain type and wavelength band SAR image were conducted, the results suggest that U-net cannot detect new constructions not accurate enough with the current training data. The increasing of the amount of training data could solve this problem but in practice the data is difficult to obtain, therefore new deep learning architecture need to be designed that can maximize the utilization of the data at disposal.

4.4 Summary

The experiment to demonstrate the performance of U-net in detecting newly built constructions has been conducted in this chapter. The result shows that U-net can accurately fulfil the objective when the testing area is Bangkok, which is the same city as in training set, but after the experiment on using with some other areas and C-band SAR image were conducted, the result turned out that U-net cannot detect new constructions not accurate enough. The increasing of the amount of training data could solve this problem but in

practice the data is difficult to obtain, therefore new deep learning architecture need to be designed that can maximize the utilization of the data at disposal.

Reference

[1] Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutionalnetworks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Inter-Vention, Munich, Germany, 5–9 October 2015; pp. 234–241.

[2] Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 640–651

[3] Badrinarayanan, V.; Alex, K.; Roberto, C. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv 2015, arXiv:1511.00561.

[4] Zeiler, M.D.; Taylor, G.W.; Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2018–2025.

[5] Aitken, A.P.; Ledig, C.; Theis, L.; Caballero, J.; Wang, Z.; Shi, W. Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize. arXiv 2017, arXiv:1707.02937.

[6] Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv 2015, arXiv:1502.03167.

[7] Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. arXiv 2016, arXiv:1611.07004.

[8] Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. Comput. Geosci. 1984, 10, 191–203.

[9] Otsu, N. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man 1979, 9, 62–66.

Chapter 5. Proposal of Chronological order reverse network

5.1 Introduction

Although the U-net can perform an accurate detection in Bangkok testing site, which the terrains are mostly plain, its accuracy is decrease when testing with other areas, i.e. Hanoi and Xiamen, with different environment where Hanoi has much more different building shapes and Xiamen area is surrounded by water. The U-net also failed at detecting new buildings in mountainous area in Chiang Mai which the images are acquired from other satellite. Moreover, the experiment when the number of training set is reduces shows that the U-net is unable to detect any result when the training data is not enough. While, the increasing of the amount of training data could solve this problem but in practice the data is difficult to obtain, therefore new deep learning architecture need to be designed that can maximize the utilization of the data at disposal.

As a result, the Chronological order reverse network (CORN) has been proposed which can deal with detecting newly built constructions very accurately. CORN adopts the general idea of U-net for learning the building change detection. CORN is designed to deal with the dataset and problems that the time-order has significant importance, such as detecting appearing or disappearing of a specific object from bitemporal images in a form of before-after pair, in case of this dissertation, a newly built constructions detection. Since the training data is difficult to obtain and the ground truth is hard to create, The author tackle this problem by making CORN tolerance to the low amount of training data. CORN is designed based on the assumption that change of the target object is in the same place and shape although the order of image pair is Time1-Time2 or Time2-Time1 which means it

can share the same ground truth data. The author adopts the concept of skip connection from U-net and applied it to CORN. With all these concepts, CORN can achieve a better accuracy than U-net at the same number of training data, or even lower.

5.2 Idea of Chronological order reverse network

While the conventional methods are inaccurate and cannot distinguish types of change, deep learning method like U-net was able to detect the target change with high accuracy. The major different between conventional methods and deep learning are that conventional method uses fixed mathematical condition in detecting change, but deep learning learns to find the change from the characteristics of changes through hierarchical layers of features recognition, which is much more flexible and robust when fully trained. Especially with U-net that contains skip connection which makes it able to retrieve image resolution from learned features, resulting in precise detected constructions boundaries. Unfortunately, the 2028 pairs of SAR images used as the training data is considered insufficient and lack of diversion, thus the result of detection of newly built constructions from two different time points using SAR images based on U-net architecture cannot achieve more accuracy and robustness. Until now, many data augmentation techniques have been introduced for the purpose of increasing number of training set. Most of them involving the image transformation such as flip, rotate, scale, crop, etc. However, in our case, these techniques only help increasing the training set in term of quantity, instead, they reduce the variation of the dataset by increasing repetitive images which lead the training of deep learning network to be risked from overfitting problem. While many researches published the deep learning study on time-series data, there is no publication that take the advantage of the chronological order for increasing the performance of the model.

As the past experiments indicate that the model trained with before–after (Time 1–Time 2) set of SAR bitemporal images on U-net can get high accuracy result, so the author had a hypothesis that if the network is trained with after–before (Time 2–Time 1) images set and tested it with the testing image in Time 2–Time 1 order, the result should be mostly the same with a little difference in detection detail. The experiment shows that even though both models showed mostly the same detection result and the accuracy were almost at the same level, the detected positions between these two models were not exactly in the same places where in some aspects of the detail of detected shapes were different. The differences are caused by the network learned to see the change differently where in Time 1–Time 2, the network will learn that the feature of intensity change of new buildings is mostly from low intensity to high intensity while it is reverse for the network that learned from Time 2–Time 1. There are also many more factors in detail such as the intensity or shape around the buildings. These are the reasons why the order of training data matters and why the detection results are slightly different. The evident can be found in the using of the U-net model trained with Time 1–Time 2 data that the chronological order of the testing data must be the same with the training data, that is it also have to be Time 1–Time 2. This statement also applies to the U-net model trained with Time 2–Time 1 data where it must be tested with Time 2–Time 1 data. This statement indicates that the weights in the filters in U-net of these two models are different, meaning that they have learned the features differently although they can produce the similar results. Therefore, there is an opportunity to utilize this advantage by creating the network that can learn features of building changes in both ways.

In order to utilize the training data, in this thesis, a new way to detect newly built constructions in SAR images has been introduced by proposing a network architecture called “Chronological Order Reverse Network” (CORN), which can learn to detect

constructions more efficiently when the same number of SAR time-series data and ground truths are used. CORN is based on the assumption that regardless of whether the changes are found from Time 1–Time 2 or Time 2–Time 1, even though the detection in Time 1–Time 2 result in the appearance of constructions and Time 2–Time 1 result in the disappearance of constructions, the changes are still at the same spots with the same shape. This means that both types can be correctly associated with the same ground truth data. While normally, the detection of new buildings is supposed to use the data in Time 1–Time 2 format, our proposed architecture takes both Time 1–Time 2 and Time 2–Time 1 formats of data to allow learning based on both of the changing features to make it more viable. This allows the network to be trained with a greater variation of data, and can result an increased detection accuracy without having to use more SAR data or create any additional ground truths. While U-net is powerful, it neither can be trained with both time formats at the same times nor adding the redundant data to the training set since it would be risk for the model to be suffered from overfitting problem. But because of the high performance of U-net, CORN adopts the two adaptations of U-net where each one learns from each chronological order set and share the features learned by these two U-net through the specially designed skip connection. Moreover, CORN has the potential to use SAR images from other satellites and other environments because the training back and forth causes the model to be more robust.

5.3 Design of Chronological order reverse network architecture

CORN is designed to deal with the dataset and problems that the time-order has significant importance, such as detecting appearing or disappearing of a specific object from bitemporal images in a form of before-after pair, in case of this dissertation, a newly built constructions

detection. Since the training data is difficult to obtain and the ground truth is hard to create, the author tackles this problem by making CORN tolerance to the low amount of training data. CORN is designed based on the assumption that change of the target object is in the same place and shape although the order of image pair is Time1-Time2 or Time2-Time1 which means it can share the same ground truth data. The author adopts the concept of skip connection from U-net and applied it to CORN. With all these concepts, CORN can achieve a better accuracy than U-net at the same number of training data, or even lower.

The proposed architecture mainly consists of two U-net networks. The first one (upper side of Figure 5.1) is for training the network to learn the features of change in the appearance of buildings from Time 1–Time 2 times-series SAR images. The another one (lower side of Figure 5.1) is for learning the change in the disappearance of buildings from Time 2–Time 1 image. Each encoder 1 receives pairs of training data in the form of Time 1–Time 2 for the upper side and Time 2–Time 1, which are generated by the reverse of the original data, for the lower side. By having two networks, the network is allowed to learn a greater variety of features of change in different perspectives, since it learns from different forms of the dataset. These two U-nets have mostly same architecture as the original one, except our modification at encoder 8 and the skip connection. Instead of using the ordinary encoder 8, which is obtained through seven repetitions of Zeropadding-convolution-BatchNorm-ReLU [1] layers from each encoder block (details shown in Table 5.1), this architecture lets encoder 8 from the Time 1–Time 2 side and from the Time 2–Time 1 side share the features they have learned by

$$Encoder8 = \alpha\left(Encoder8_{self}\right) + \beta(Encoder8_{opposite}) \quad (4)$$

This means encoder 8 from the Time 2–Time 1 side consists of α% of what it has learned by itself, and β% of what the Time 2–Time 1 side has learned; in the case of encoder 8 from

the Time 2–Time 1 side, this pattern would be reversed. The finding of the most suitable portion of $\alpha:\beta$ for our dataset that would give the encoder 8 on each side some features that cannot be learned by itself, while not too much of what it has learned is affected, is conducted in the latter chapter.

While the skip connection in the original U-net directly passes the features from each encoder to each corresponding decoder, which allows it to receive significant information regarding the edges and boundaries of the features, the author does the same in the proposed architecture, but in a different way. As encoder 8, which is the starting point for the decoder, is influenced by the information of the opposite side, using such a straightforward skip connection would result in the decoder failing to generate an output that includes features from both sides. Thus, the author solved this by adding the features from the encoders of both sides before passing it to the corresponding decoder. By following this approach, the decoder is able to generate an output with features learned by its own encoders, but with influence from the other side, while receiving all boundary information from both sides. However, for controlling the balance amount of features sharing, the specific design of the skip connection must be delivered. Thus, the experiment for designing the skip connection in CORN has also been conducted as well.

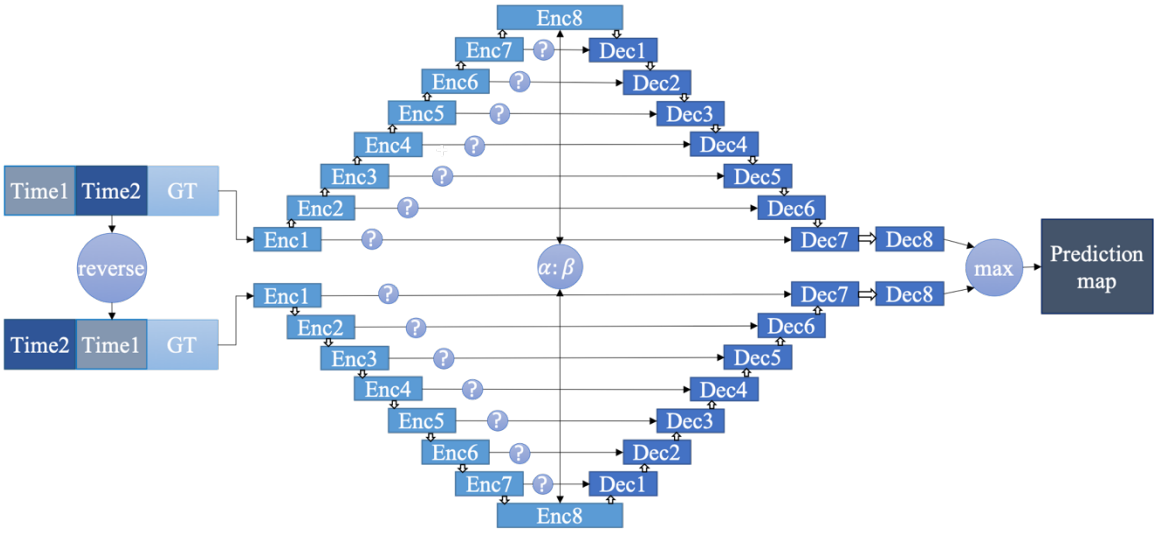


Figure 5.1. Conceptual design of architecture of the Chronological Order Reverse Network (CORN).

Lastly, applied the maximum operation between these two results to draw the best result out of each one. In training, loss was calculated with the weighted binary cross entropy function [2], as our dataset contained a lot of negative class pixels (non-changed areas), while the number of positive class pixels (new construction areas) was small. The weight of loss function is the division of the percentage of negative pixels in the training set by the percentage of positive pixels in the training set. In our case, the weight was 181.5, which is the result of the rate of white pixels (positive class pixels) = 0.548% and the rate of black pixels (negative class pixels) = 99.452%.

Table 5.1. Detail of the encoder and the decoder.

Encoder	**Decoder**
PCR (256,2,4,2)	CRD (1,512,2,2)
PCBR (128,64,4,2)	CRD (2,1024,2,2)
PCBR (64,128,4,2)	CRD (4,1024,2,2)
PCBR (32,256,4,2)	CRD (8,1024,2,2)
PCBR (16,512,4,2)	CRD (16,1024,2,2)
PCBR (8,512,4,2)	CRD (32,512,2,2)
PCBR (4,512,4,2)	CRD (64,256,2,2)
PCBR (2,512,4,2)	C (128,128,2,2)

In Table 5.1, P, C, B, R, and D represent the layers of zero padding (size 1,1), convolution (in the encoder) or deconvolution (in the decoder), batch normalization (0.2), ReLU, and dropout, respectively. From left to right, the numbers in parentheses indicate the input size^2, number of features, filter size, and stride amount of convolution filters, respectively. As the author followed the method applied by Isola et al. [3], all of the ReLUs in the encoder were leaky with a slope of 0.2, while the ReLU functions in the decoder were not leaky. The dropout rate was 0.5.

5.4 Chronological order reverse network parameters

As the proposed network is still in questioned in its detail, this chapter intends to clarify the decision of the creating by comparing the performance of CORN in multiple settings. In order to select the most appropriate setting for CORN, a number of experiments have been conducted to ensure the architecture works properly, which will be explained in this section. First, to support our assumption on the ratio of the ordinary input set to the reverse input set to be calculated in (1), the author conducted experiments to find the most suitable ratio for encoder 8 among the 6:4, 7:3, 8:2, and 9:1 ratios for the model to learn the shared features between the two input sets. Also, an experiment on skip connection was conducted, where the model trained with the architecture that has only an addition skip connection at one side is tested and compared the result with the model trained with the architecture with an additional skip connection to both decoder sides.

The feature sharing between two side is not made only by the skip connection, the author also made a sharing function between the encoder 8 from two sides by the proper ratio. In this chapter, the best ratio of feature sharing function has been clarified.

To compare the encoder 8 ratio, we tested each model with the Bangkok site. The results of the first testing area from the SAR pair of 12 January 2009/21 November 2009 are displayed in Figure 5.2, as it is the easiest with which to notice the difference. The buildings tend to be detected less when the influence from the opposite side of encoder 8 is smaller, as seen in Figure 5.2d, at a ratio of 9:1. In contrast, the 6:4 ratio in Figure 5.2a includes too many, too-large buildings in the detection result, since it experiences more influence from the opposite encoder 8. The 7:3 and 8:2 ratios have similar detection results, but 7:3 was chosen for our work since it works significantly better in reducing the false positive rate, as shown in Table 5.2.

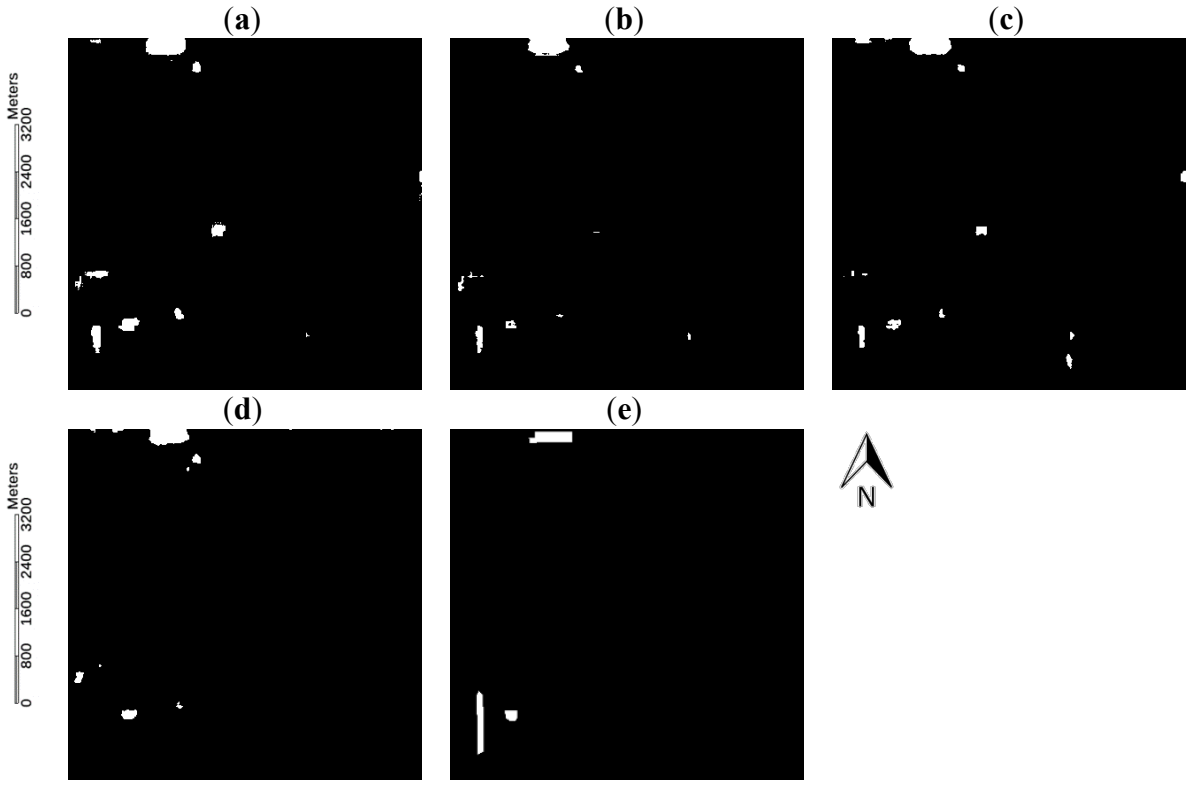


Figure 5.2. Results of the Bangkok site in the first area of the SAR pair of 12 January 2009/21 November 2009, where the size of each image is 6 × 6 km: (**a**) the encoder 8 portion is 6:4, (**b**) the encoder 8 portion is 7:3, (**c**) the encoder 8 portion is 8:2, (**d**) the encoder 8 portion is 9:1, (**e**) ground truth. 14°1'2.26"N 100°41'15.99"E.

Table 5.2. Accuracy of the model in the different encoder 8 portions at the Bangkok site.

Validation Method	6:4	7:3	8:2	9:1
False negative	47.368	54.667	55.739	59.621
False positive	0.601	0.210	0.391	0.285
Overall accuracy	98.928%	99.791%	99.614%	99.717%
Precision	0.471	0.687	0.535	0.590
Recall	0.526	0.453	0.442	0.404
F measure	0.475	0.659	0.526	0.568
F1 measure	0.497	0.546	0.484	0.479
Kappa	0.492	0.543	0.480	0.475
IOU	0.331	0.376	0.320	0.315

The skip connection in original U-net is placed between layers of encoder and decoder. However, the direct place of skip connection cannot help the network to share the features it has learned from each encoder side. The author instead adds the features from two sides before passing to decoders. In this part clarify what will happen if the skip connection has been made the other ways. So far, the author has proposed two type of skip connections for the architecture, the first one contains additional skip connection at one side of the network (Figure 5.3) and another architecture contains an additional skip connection to both decoder sides (Figure 5.4).

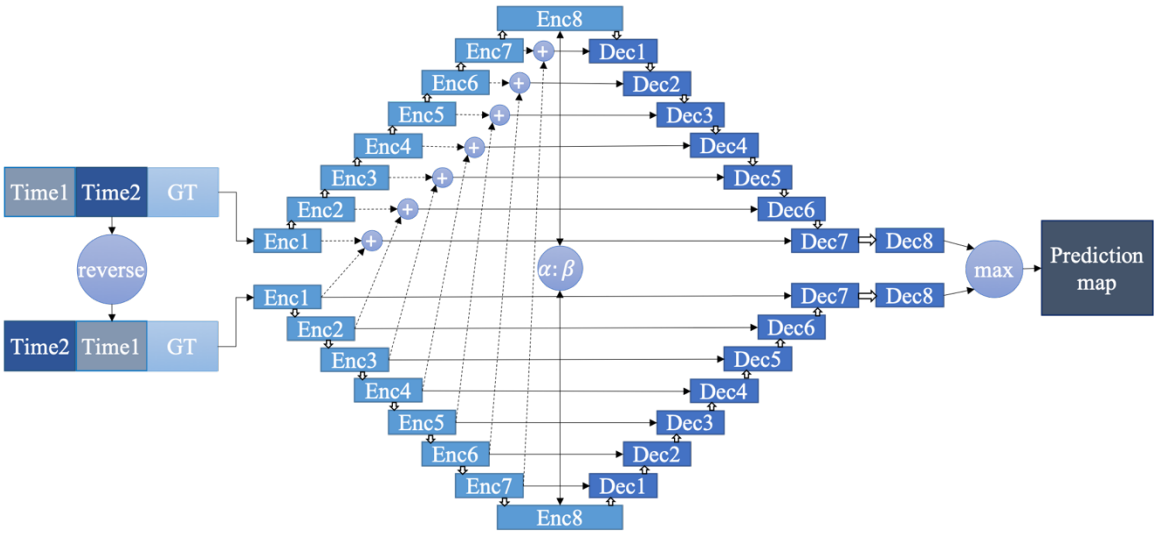


Figure 5.3. Design of CORN with an additional skip connection to one side decoder side

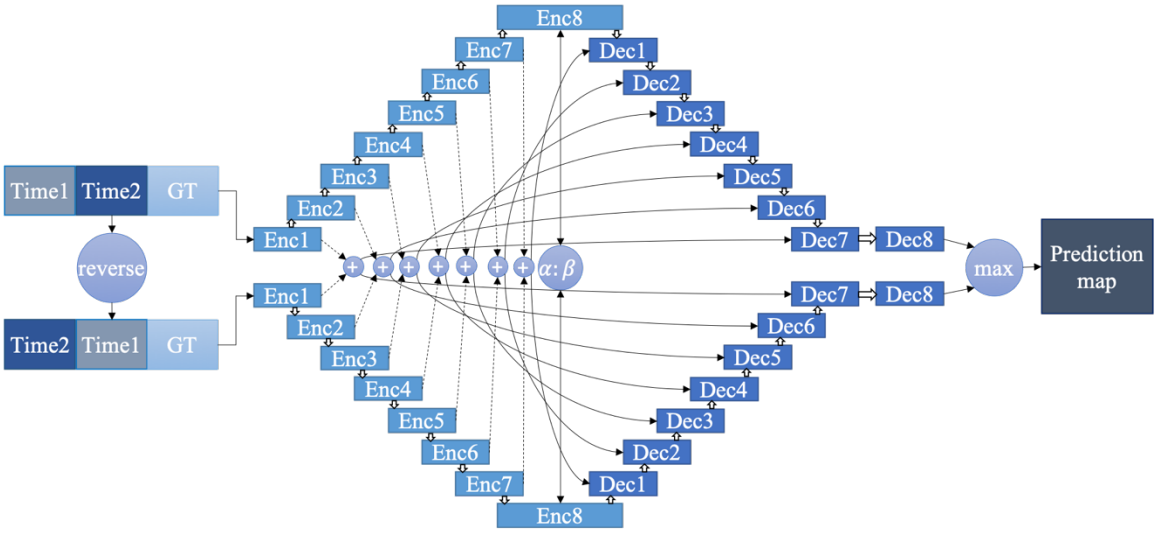


Figure 5.4. Design of CORN with an additional skip connection to both decoder sides

The second area of the Bangkok site from the SAR pair of 12 January 2009/21 November 2009 was used to display the difference between sending the added features with skip connection to one side of the decoder (Figure 5.3), and to both side of the decoders (Figure 5.4) in Figure 5.5.

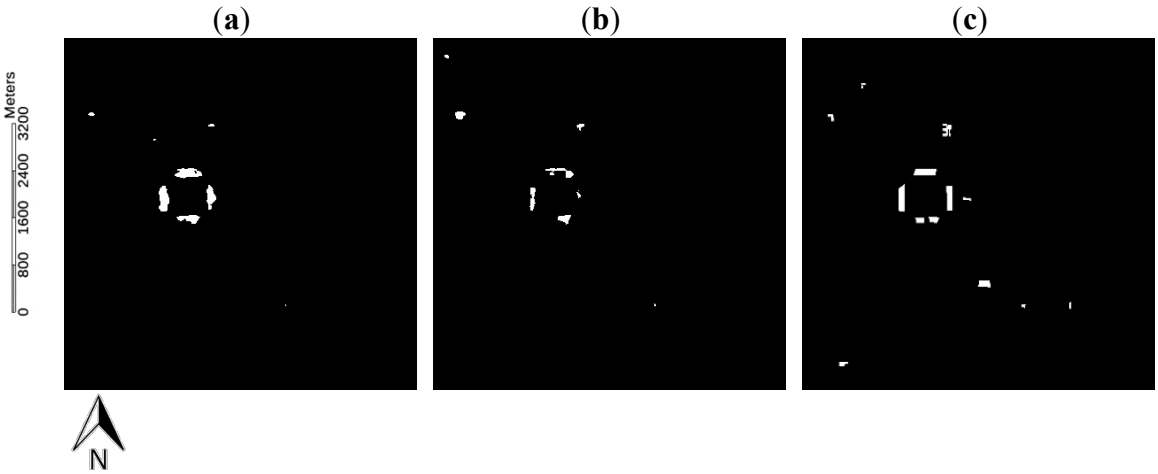


Figure 5.5. Results of the Bangkok site in the second area of the SAR pair of 12 January 2009/21 November 2009, where the size of each image is 6 × 6 km: (**a**) Additional skip connection on one side of the network; (**b**) Additional skip connection on both sides of the network, (**c**) ground truth.

As a result, the author only applied the addition skip connection with the Time 1–Time 2 side, while the Time 2–Time 1 side used the traditional direct skip connection (Figure 5.3). The reason is that when it is applied on both sides (Figure 5.4), the boundary information shared between them will be too much, and will lead to a limited result within these boundaries. The result in Figure 5.5 also support that the shape of a detected building is too limited to the boundary information sent by the addition of encoders when the addition of the skip connection is applied to both sides of the network, as can be observed by the square-shape-like building change at the center of Figure 5.5b. As a result, the false negative rate increases, as shown in Table 5.3.

Table 5.3. Accuracy of the model in the different skip connections in the architecture at the Bangkok site.

Validation Method	**Additional Skip Connection on One Side**	**Additional Skip Connection on Both Sides**
False negative	54.667	66.936
False positive	0.210	0.180
Overall accuracy	99.791%	99.149%
Precision	0.687	0.652
Recall	0.453	0.331
F measure	0.659	0.603
F1 measure	0.546	0.439
Kappa	0.543	0.435
IOU	0.376	0.281

After the author has got the proper setting for our architecture network, the experiment has been conducted to show the different between using weighted and non-weighted binary cross entropy loss functions. The accuracy of each model is shown in Table 5.4.

Table 5.4. The accuracy of non-weighted loss compared with that of weighted loss for the Bangkok testing site.

Validation Method	Non-Weighted Loss	Weighted Loss (ω_p = 181.5)
False negative	51.110	54.667
False positive	0.503	0.210
Overall accuracy	98.99%	99.791%
Precision	0.497	0.687
Recall	0.489	0.453
F measure	0.496	0.659
F1 measure	0.493	0.546
Kappa	0.488	0.543
IOU	0.327	0.376

Moreover, the experiment for comparing the accuracies resulting from using a patch size of 128 × 128 and 256 × 256 has been conducted. As stated earlier, the author aimed to use a patch size that yields the most suitable ratio between black pixels and white pixels that enables the network to learn the positions of building constructions from positive samples; in this research, the author used a patch size of 256 × 256 pixels. Before selecting this size, the author conducted an experiment to test which patch size—128 × 128 or 256 × 256 pixels—results in better accuracies. The ω_p of the 128 × 128 patch size is 80, and the ω_p of the 256 × 256 patch size is 181.5. ω_p = 80 is obtained when the proportion of white pixels = 1.23% and the proportion of black pixels = 98.77%, while ω_p = 181.5 is obtained when the proportion of white pixels = 0.55% and the proportion of black pixels = 99.45%. Please note that the total number of images in training data of 256 × 256 patch size is 2,028 pairs while in 128 × 128 patch size is 17,925 pairs since the cutting size is smaller while the size of the whole scene of the ground truth is the same. The results in Table 5.5 indicate that using a 256 × 256 patch size leads to better accuracies. As a result, the author decided to use a patch size of 256 × 256 in all experiments. A 256 × 256 patch size results in better accuracies because a 128 × 128 patch size is too small, causing the loss of features of some parts, such as paddy fields, and leading to the network's inability to fully learn the change pattern of these areas. Even though the negative part is not the focus of this study, it is

indispensable for network training, which can be more recognizable in a 256 × 256 patch size. Moreover, because the sliding step in patches cutting is smaller in a 128 × 128 patch size, there is a significant increasing of the number of the training set which causing the cut patches to have too many repetitive patterns of both positive and negative features, which can cause the model to be overfitted at an early stage.

Table 5.5. The accuracy of a 128 × 128 patch size compared with a 256 × 256 patch size on the Bangkok testing site.

Validation Method	128 × 128 Patch Size	256 × 256 Patch Size
False negative	86.923	54.667
False positive	0.132	0.210
Overall accuracy	98.99%	99.791%
Precision	0.502	0.687
Recall	0.131	0.453
F measure	0.407	0.659
F1 measure	0.207	0.546
Kappa	0.204	0.543
IOU	0.116	0.376

As all of the experiment for CORN architecture detail and parameters determining have been conducted so far, the author has got the best setting of the CORN architecture as a final proposed model for this thesis, as shown in Figure 5.6, and it will be used for the rest of the experiment.

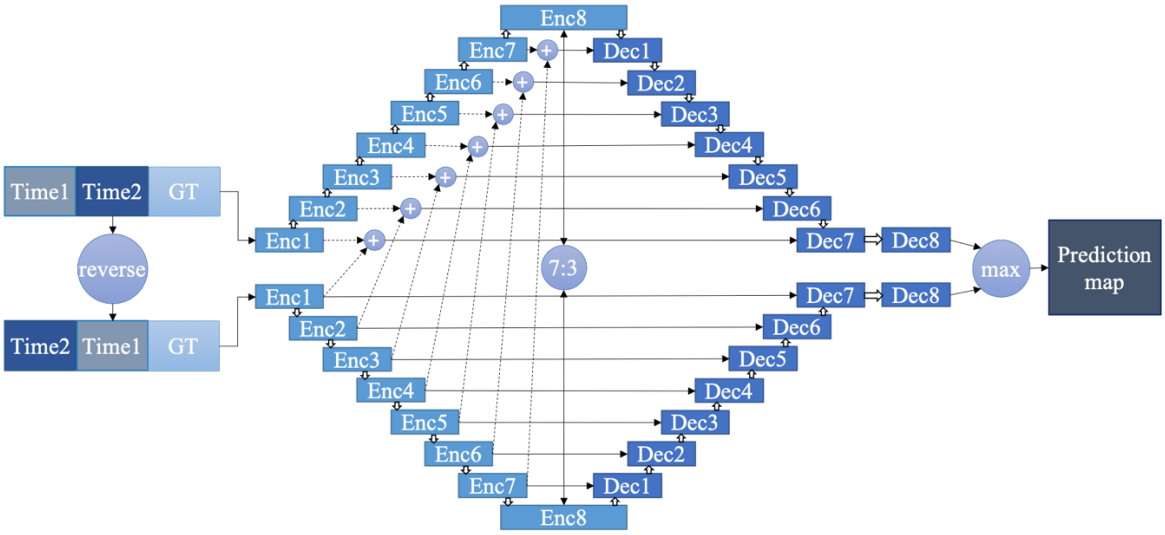


Figure 5.6. Proposed architecture of the Chronological Order Reverse Network (CORN).

5.5 Chronological order reverse network experiment & comparison

In this chapter, the performance of the model has been tested at the Bangkok area, which is the same area as in training set but acquired from different dates. This experiment also show that model of CORN can work even when trained with the small number of dataset by comparing with U-net.

The same area as in the U-net experiment was selected for testing the model in Bangkok, which was the same city chosen in the training data. The results are shown in Figures 5.7 and 5.8. Although multiple testing areas were tested, the author selected one area to show for ease of inspection. The pixel number of the testing area was 640,000, including 6439 positive pixels and 633,561 negative pixels in ground truths.

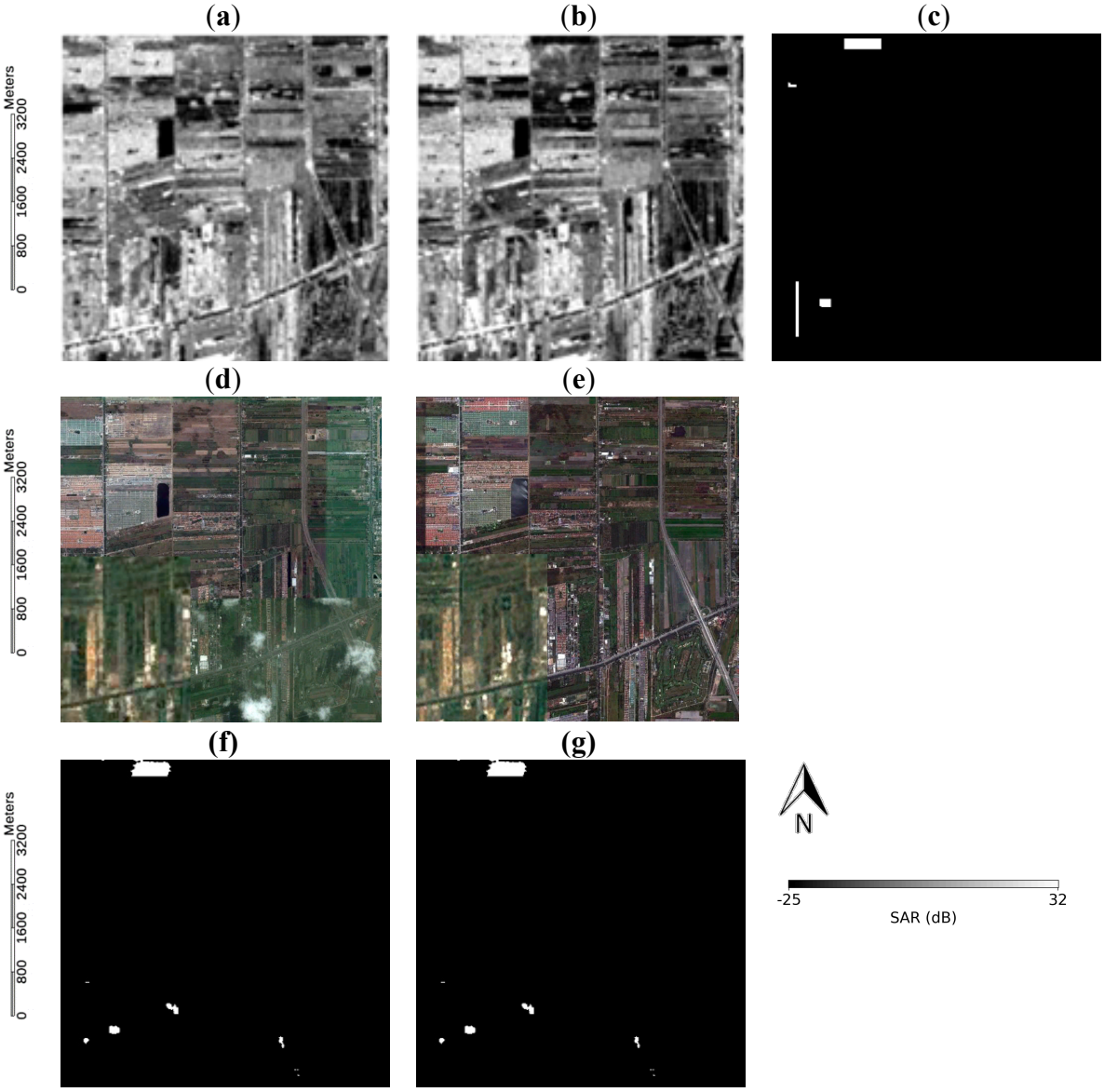


Figure 5.7. Results of the Bangkok site in the first area at 14°1’2.26”N 100°41’15.99”E. The size of each image is 6 × 6 km (for SAR pairs 27 November 2008/15 January 2010: (**a**) Time 1 SAR image; (**b**) Time 2 SAR image; (**c**) ground truth; (**d**) result of U-net; (**e**) proposed result).

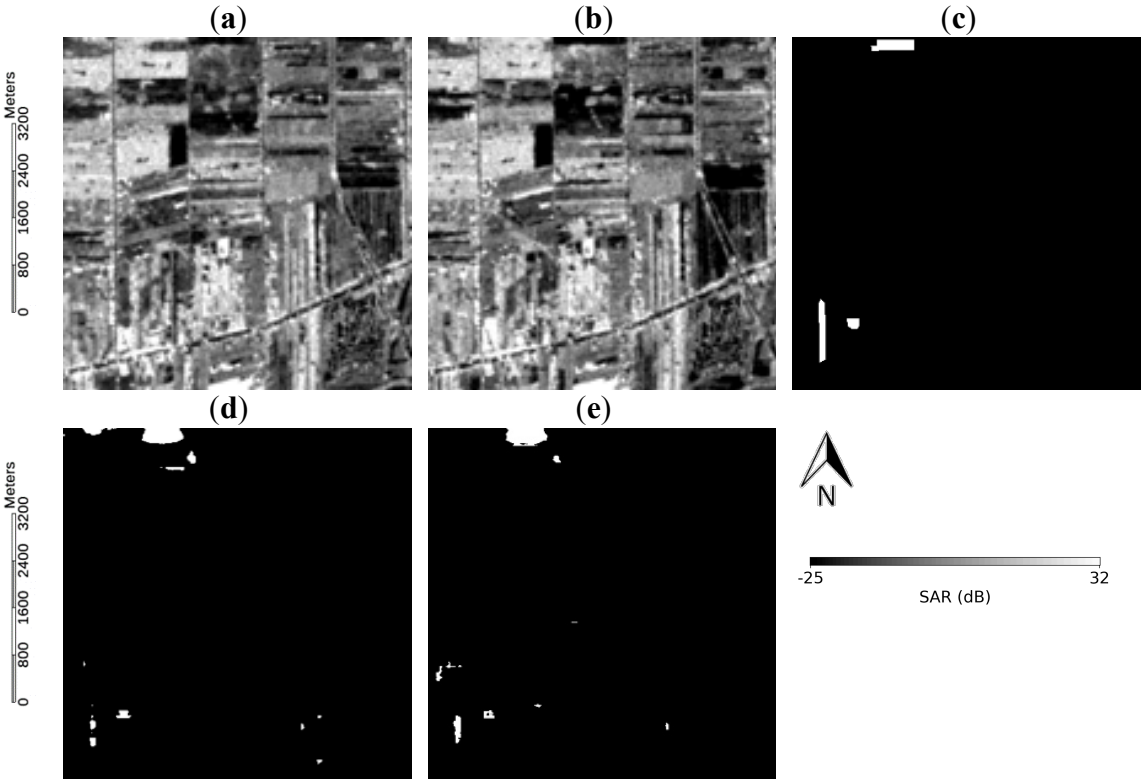


Figure 5.8. Results of the Bangkok site in the first area at 14°1'2.26"N 100°41'15.99"E. The size of each image is 6 × 6 km (for SAR pairs 12 January 2009/21 November 2009: (**a**) Time 1 SAR image; (**b**) Time 2 SAR image; (**c**) ground truth; (**d**) result of U-net; (**e**) proposed result).

The results were able to detect only the construction of buildings while avoiding the change caused by the season. However, the results from the proposed architecture are visually better than that of U-net, as it can detect more detailed buildings and provides more accurate shapes of buildings, thus reflecting lower false negative and false positive rate in Table 5.6. The new model can also detect rows of buildings at the lower left part of the image more accurately, even though it has a low intensity difference between Time 1 and Time 2 images.

With the advantage of training using both Time 1–Time 2 and Time 2–Time 1, the use of

CORN resulted in more precise detection at the edges of buildings, resulting in improved accuracies for the tested Bangkok dataset where can be seen in Figure 5.9. On visual inspection, it can be seen that the edges of the detected buildings were more similar to the ground truth when compare to U-net. The area in Figure 5.9 is a cropped version of the first test area of Bangkok to discuss whether the model can or cannot detect newly built constructions from SAR images. The model can precisely detect the construction if the change in SAR intensity is significant, as can be seen at the top-right corner in Figure 5.9d,e. Please note that the southern part of the mentioned area has high intensity in the Time 1 SAR image, but there is no house in the optical image because of a time gap between the available optical image and our SAR dataset. On the other hand, in some cases, the model is not able to detect constructions precisely if the difference in SAR intensity is too small, as in the case of the row of houses on the left side in Figure 5.9d,e. However, it is nearly impossible for any algorithm or even manual inspection to detect changes if the difference in intensity is very low even with U-net where the detected result is only a small area as shown in Figure 5.9f. The reason for the low intensity of the houses in the red rectangular area compared with the intensity of those in the blue rectangular area is the difference in the orientation of the houses. As they are constructed in different orientations, it is possible that they reflect the SAR signal differently. The high intensity in the village at the top-right corner is possibly the result of the double bounce on the houses' walls or a strong single bounce on the houses' roofs, but these phenomena do not happen with the houses on the left side because the orientation of the houses is different, and the latter may end up reflecting the SAR signal at their roof edge.

(a) **(b)** **(c)**

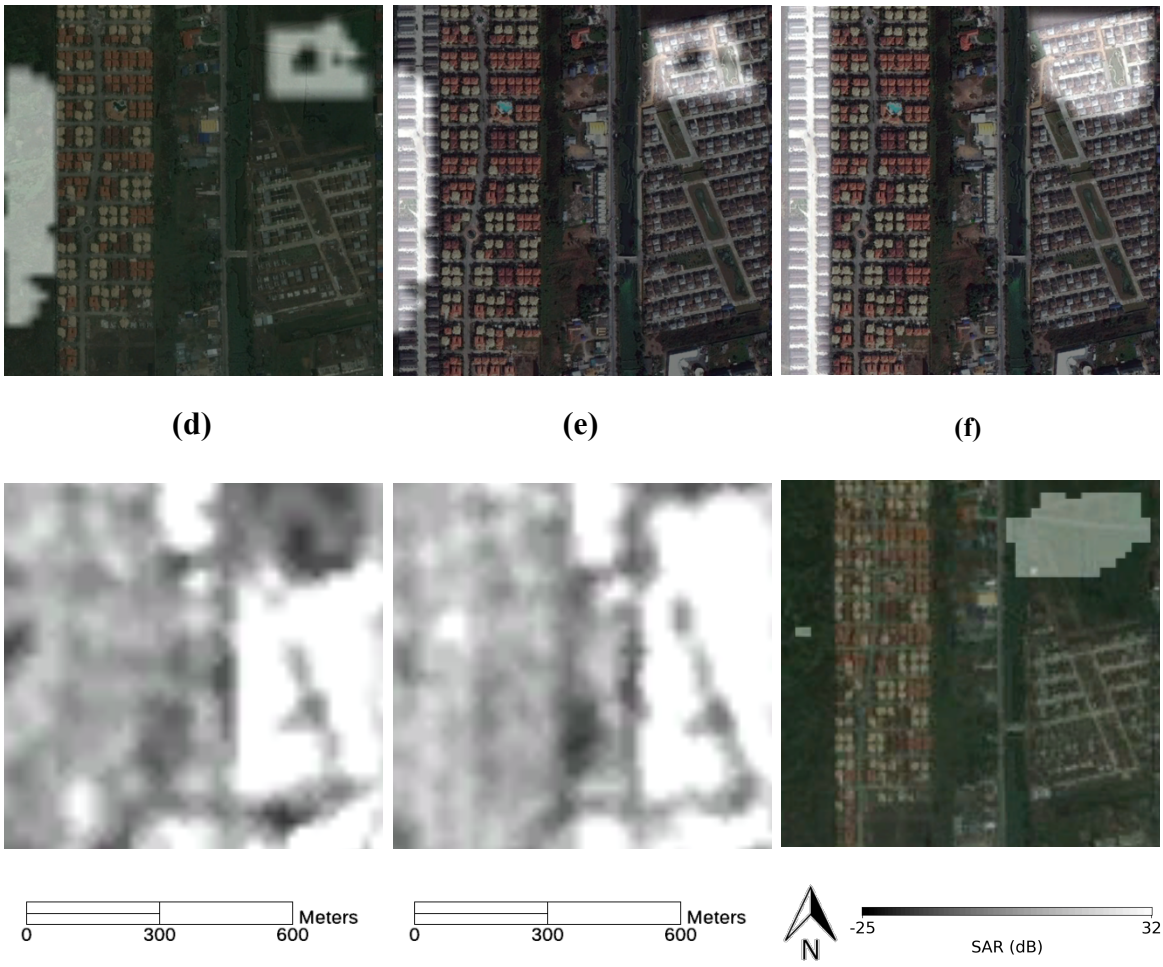


Figure 5.9. Example of detection results of CORN at Bangkok testing site for SAR pairs 12 January 2009/21 November 2009. (a) CORN result overlays on Time 1 optical image, (b) CORN result overlays on Time 2 optical image, (c) ground truth overlays on Time 2 optical image, (d) Time 1 SAR image, (e) Time 2 SAR image, (f) U-net result overlays on Time 1 optical image.

In addition, to test the learning capability of CORN in lower training set situations, the author tried to randomly reduce the number of training sets from 2028 pairs to 1500 pairs and 500 pairs, respectively. Similar to the experiment of U-net, for each number of training sets, the networks were trained four times with four different randomly selected training pairs, and then tested with the Bangkok testing site. The results of this experiment are shown in Table 5.6 as the averages of four times the testing results. As expected, in the case of

1500 training pairs, the accuracies of both CORN and U-net dropped from when trained with 2028 pairs, but CORN still surpassed U-net, except in false negatives and recall. The use of 500 training pairs indicates that U-net cannot be trained with a very small dataset, as is reflected in the very low accuracies. While the accuracies of CORN were relatively low, they were still in the acceptable range, which means that the network is able to learn even with a very small training set. This result supports our assumption that learning features from two formats of bitemporal data helps the network to become better at detecting newly built constructions where it can be seen from Figure 5.10 and 5.11 that both of overall accuracy and Kappa of CORN are higher than U-net in both training data numbers. The time taken by CORN in training 1500 pairs was 53 min, whereas for U-net, it was 48 min. For the single training of 500 pairs, CORN spent 18 min and U-net spent 15 min. This experiment has proven that CORN can solve the problem of U-net which is the failure in detection when the available training data is low. For the versatility testing of CORN where the author conducts an experiment of using CORN in detecting new buildings in Sentinel-1 image at Chiang Mai in order to compare with U-net, it will be discussed in chapter 6. Although the accuracy of 500 pairs training data cannot exceed the 0.4 Kappa and 0.3 IOU, it almost gets the satisfy result which resulting in an acceptable accuracy.

Table 5.6. Accuracies of the models in the different number of training data at the Bangkok site.

Validation Method	500 Pairs		1500 Pairs		2028 Pairs	
	CORN	**U-net**	**CORN**	**U-net**	**CORN**	**U-net**
False negative	40.860	27.512	53.254	47.232	54.667	55.801
False positive	1.467	17.970	0.383	0.683	0.210	0.403
Overall accuracy	98.136%	81.934%	99.085%	98.848%	99.791%	99.04%
Precision	0.310	0.100	0.590	0.492	0.687	0.527
Recall	0.591	0.725	0.467	0.528	0.453	0.442
F measure	0.321	0.107	0.571	0.488	0.659	0.519
F1 measure	0.400	0.165	0.507	0.485	0.546	0.481

Kappa	0.391	0.150	0.503	0.480	0.543	0.476
IOU	0.251	0.092	0.340	0.321	0.376	0.316

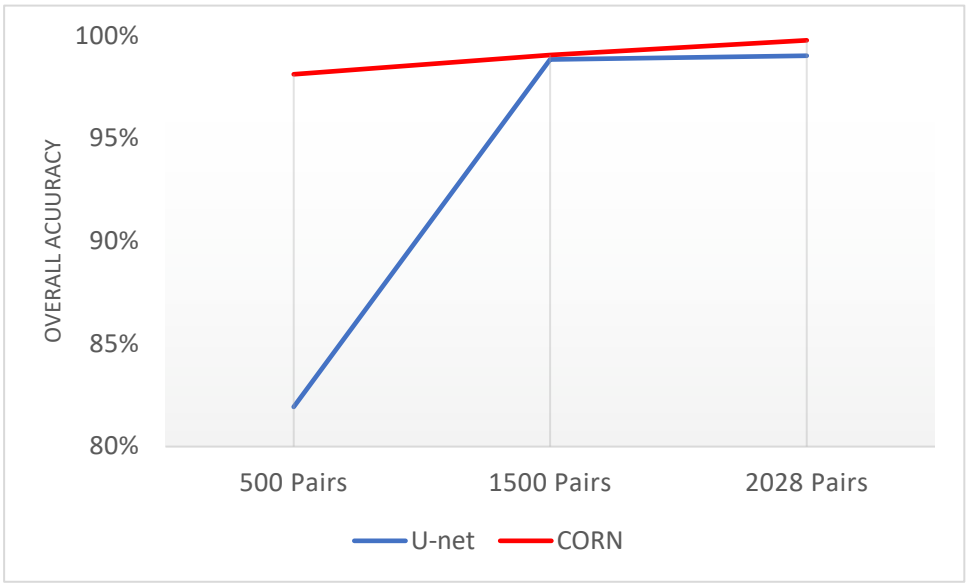


Figure 5.10. Overall accuracy of CORN comparing with U-net when the number of training data are 500 pairs and 1500 pairs

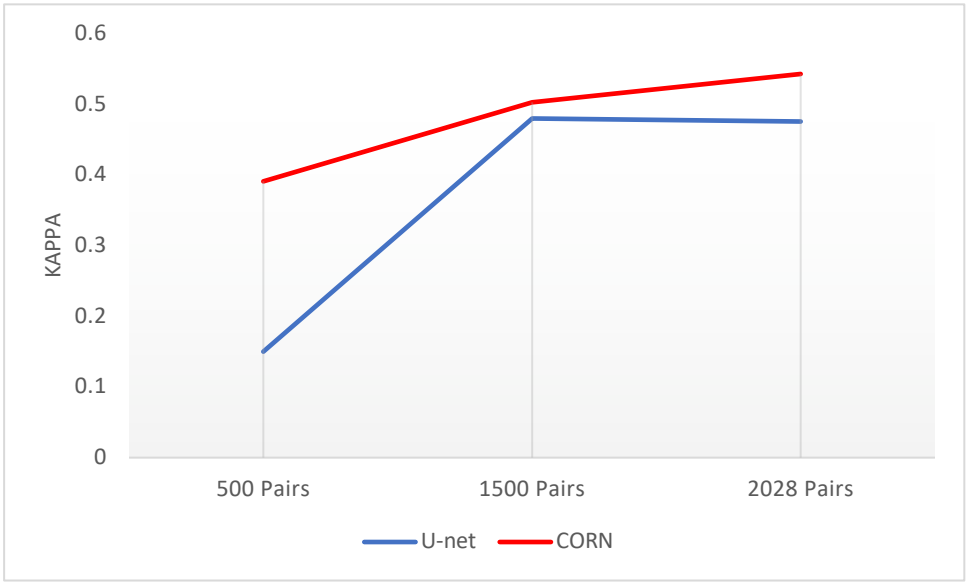


Figure 5.11. Kappa coefficient of CORN comparing with U-net when the number of training data are 500 pairs and 1500 pairs

5.6 Summary

The problem of the existing deep learning method which is U-net is described in this chapter. The experiment shows that U-net requires high number of training data to be able to make an accurate detection and thus the idea and architecture detail of the proposed method are explained in this chapter. The CORN is based on the idea of swapping the chronological order of training data which still can be associated with the same ground truth and use all of them to train the network. The model trained with CORN has been tested on the Bangkok area to find the new constructions and the result is higher than U-net. Moreover, the experiment of lowering training data has been conducted and the result shows that CORN can be used in such situation. The advantage of CORN can be summarized as shown in Table 5.7.

Table 5.7. Advantages of each method

Approach	**Accuracy**	**Robustness**	**Reasons**
Conventional methods	Low	Low	Use fixed condition which is difficult to distinguish types of change
U-net trained with Time 1 – Time 2	High	Low	Achieve high accuracy from having the skip connection but has low robustness from low and unvarying training data
U-net trained with Time 2 – Time 1	High	Low	Same reason as U-net trained with Time 1 – Time 2
CORN	High	High	Obtain the accuracy from the performance of U-net and robustness from learning from two diverse training set

Reference

[1] Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv 2015, arXiv:1502.03167.

[2] Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. Expert Syst. Appl. 2017, 73, 220–239.

[3] Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. arXiv 2016, arXiv:1611.07004.

Chapter 6. Versatility of Chronological order reverse network

6.1 Introduction

The past experiments have demonstrated the performance of U-net in detecting newly built constructions by detecting new buildings in Bangkok area. However, the past experiments also showed that U-net failed in using with some other areas with different terrain type and wavelength band from in training data and the increasing of the amount of training data is the significant way to solve this problem. While the proposed method in this thesis, CORN, which utilize the time-series data by training on both ordinary and reverse chronological order data, has surpassed the accuracy in detecting change of new buildings in U-net in Bangkok testing area without having to use any more data, its versatility in adapting to other study areas and image with different properties has not been clarified. Because of the model trained with CORN should have more robustness from the benefit of having the model trained by more variation of data, this chapter is intended to further investigate into the versatility of CORN in using with images with different terrains and environments, i.e. Hanoi and Xiamen, where Hanoi has a very different building shape from Bangkok while Xiamen is a land surrounded by water, which is also different from Bangkok area. The image of Chiang Mai from Sentinel-1 C-band sensor which U-net has failed to perform a detection is also tested with CORN in this chapter. Moreover, an experiment of using CORN model that has been trained with image capture in ascending direction with image of Bangkok that has been captured with the descending observation looking direction is conducted. Lastly, as the speckle noise can be occurred in any SAR image, the experiment of using the proposed model with noisy data has been conducted in this chapter.

6.2 Experiment on different areas

To investigate the ability to be used globally of the proposed model, the author tested it with the Hanoi and Xiamen areas, which are completely different from the training area, to see if it can detect constructions as effectively as it does in the Bangkok area. The comparison of the CORN and U-net results in the Hanoi testing site is shown in Figure 6.1 while that of the Xiamen site, which has two testing areas, is in Figures 6.2 and 6.3. The Hanoi testing set was, in total, 160,000 pixels, including 859 positive pixels and 159,141 negative pixels, and that of the Xiamen testing site was, in total, 320,000 pixels, including 4482 positive pixels and 325,518 negative pixels.

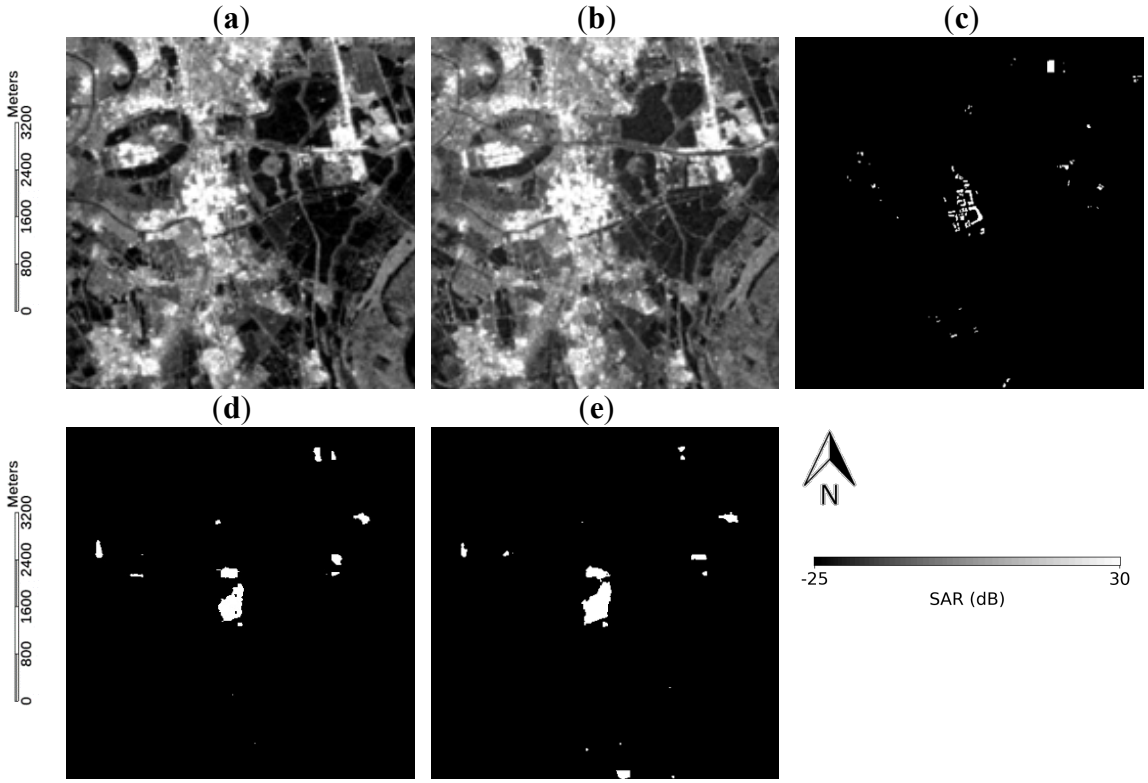


Figure 6.1. Result of the Hanoi site at 20°57′06.35”N 105°51′08.87”E. The size of each image is 6 × 6 km. (**a**) Time 1 SAR data; (**b**) Time 2 SAR data; (**c**) ground truth; (**d**) result of proposed model; (**e**) result of U-net.

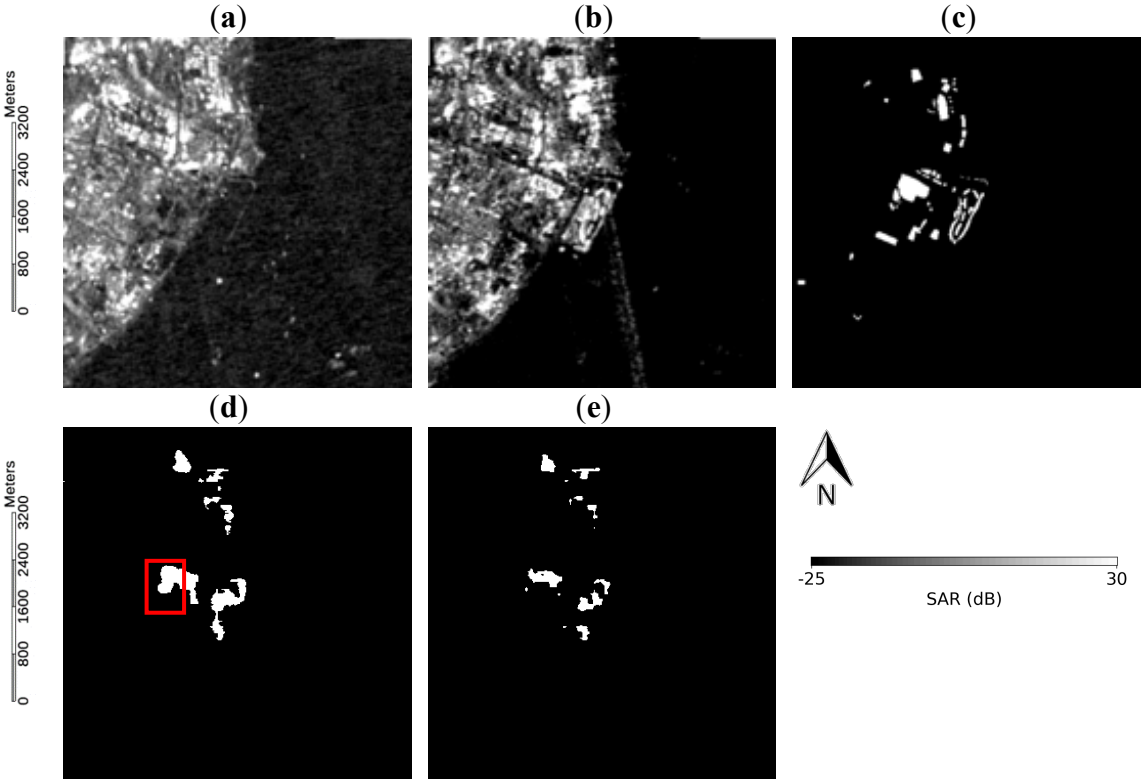


Figure 6.2. Result of the first Xiamen test site at 24°28′35.28”N 118°11′36.12”E. The size of each image is 6 × 6 km. (**a**) Time 1 SAR data; (**b**) Time 2 SAR data; (**c**) ground truth; (**d**) result of proposed model; (**e**) result of U-net.

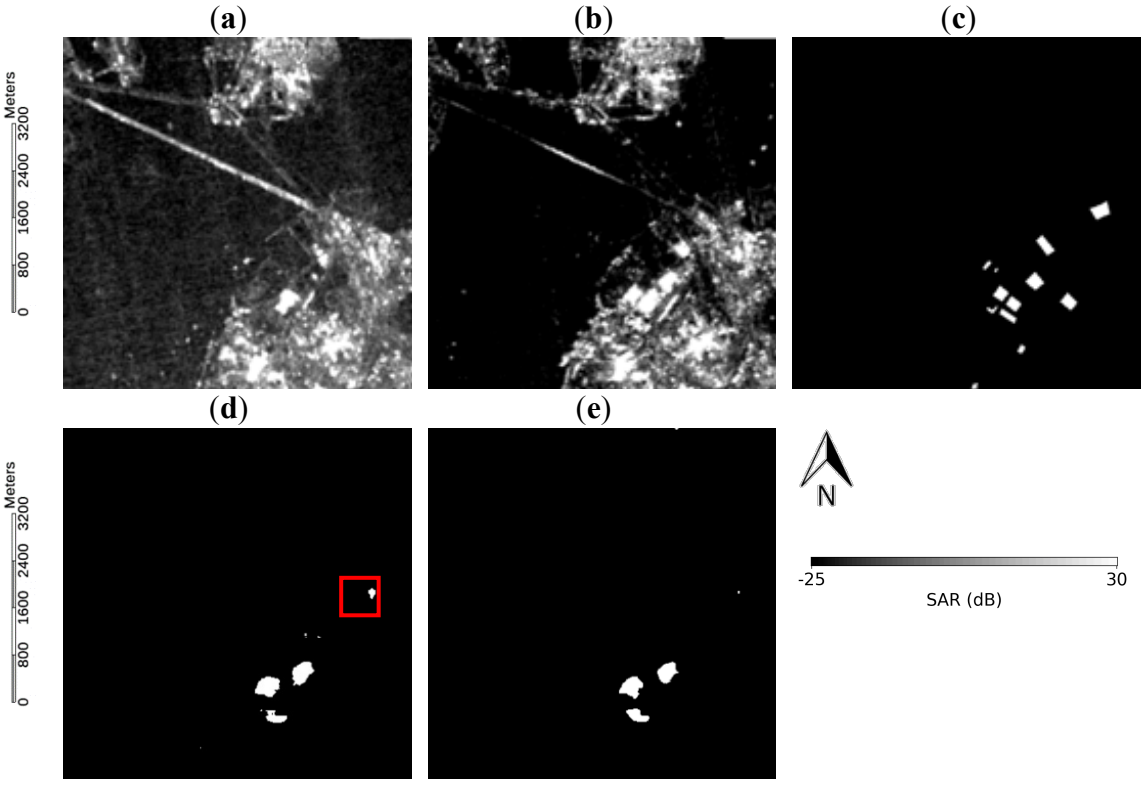


Figure 6.3. Result of the second Xiamen test site at 24°31′59.42”N 118°06′54.51”E. The size of each image is 6 × 6 km. (**a**) Time 1 SAR data; (**b**) Time 2 SAR data; (**c**) ground truth; (**d**) result of proposed model; (**e**) result of U-net.

For the Hanoi site, the accuracies of the results from the proposed network shown in Table 6.1 are very close to that of the U-net. Since the constructions that occurred in this dataset were mainly of small buildings as average size is only at 2,164.76 m^2, our model tried to generate the shapes of the changes as accurately as possible, which led to very small detection results in some areas—so much so that some detected objects appeared in very few pixels or were even omitted, as can be especially seen in the bottom half of Figure 6.1d. As a result, the false negative rate increased in our results, which made that of recall, F1 measure, Kappa, and IOU slightly lower than in U-net. In contrast, while several objects were detected by U-net in the bottom half of Figure 6.1e, they did not all correlate with those in the ground truths, causing U-net to have more false positive values than CORN. The lower false positive rate of our model resulted in higher overall accuracy, precision, and F measure rates than U-net.

Table 6.1. Accuracy of the model in the Hanoi area.

Validation Method	Proposed Network	U-net
False negative	62.980	58.324
False positive	0.782	0.922
Overall accuracy	99.522%	98.77%
Precision	0.204	0.196
Recall	0.370	0.417
F measure	0.211	0.205
F1 measure	0.263	0.267
Kappa	0.258	0.261
IOU	0.151	0.154

The accuracies of Xiamen are shown in Table 6.2.

Table 6.2. Accuracy of the model in the Xiamen area.

Validation Method	Proposed Network	U-net
False negative	68.652	77.577
False positive	0.861	0.508
Overall accuracy	98.189%	98.412%
Precision	0.341	0.385
Recall	0.313	0.224
F measure	0.338	0.364
F1 measure	0.327	0.283
Kappa	0.317	0.276
IOU	0.195	0.165

Xiamen is a city surrounded by water, which is an area type that the training data did not include. Unlike with Hanoi, the results for Xiamen from our model achieved better accuracies over U-net in recall, F1 measure, Kappa, and IOU, because of the reduction in the false negative rate due to the larger size of constructions at the average of 16,691.75 m^2. This reduction was a result of a better detection rate for constructions as the model can extract more features of changes, especially in building boundaries, as highlighted in the red rectangles in Figures 6.2d and 6.3d where the U-net can only detect as a small group of pixels. Please note that although CORN used information from both Time 1–Time 2 and Time 2–Time 1 formats, it can also avoid detecting a noise in the SAR image, displayed as a faded line from the center to the bottom in Figure 6.3b, compared to U-net that only uses change information in the Time 1–Time 2 format.

Despite the improved accuracy, some areas, especially the Hanoi site, still have a relatively high false negative rate. This is due to the fact that most of the constructions in the training data have a larger size than those in the Hanoi area, and the construction shapes are also way too different from each other, as can be seen in Figure 6.4, and this is also applicable to some construction in Xiamen as can be seen in Figure 6.5. As a result, the model failed

to detect some of the constructions and caused the high false negative rate in such areas. To tackle this problem, supplementing the training data with various sizes and shapes of buildings could be very helpful.

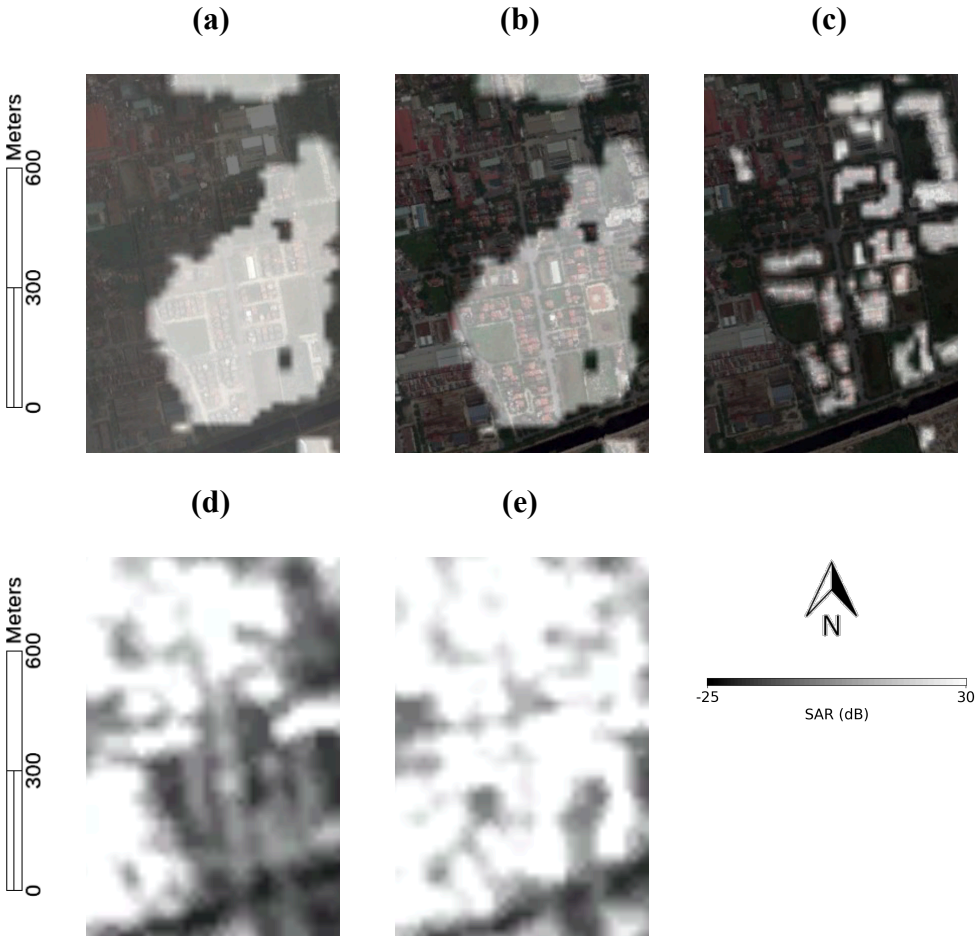


Figure 6.4. Example of detection results of CORN at Hanoi testing site. (a) CORN result overlays on Time 1 optical image, (b) CORN result overlays on Time 2 optical image, (c) ground truth overlays on Time 2 optical image, (d) Time 1 SAR image, (e) Time 2 SAR image.

(a) **(b)** **(c)**

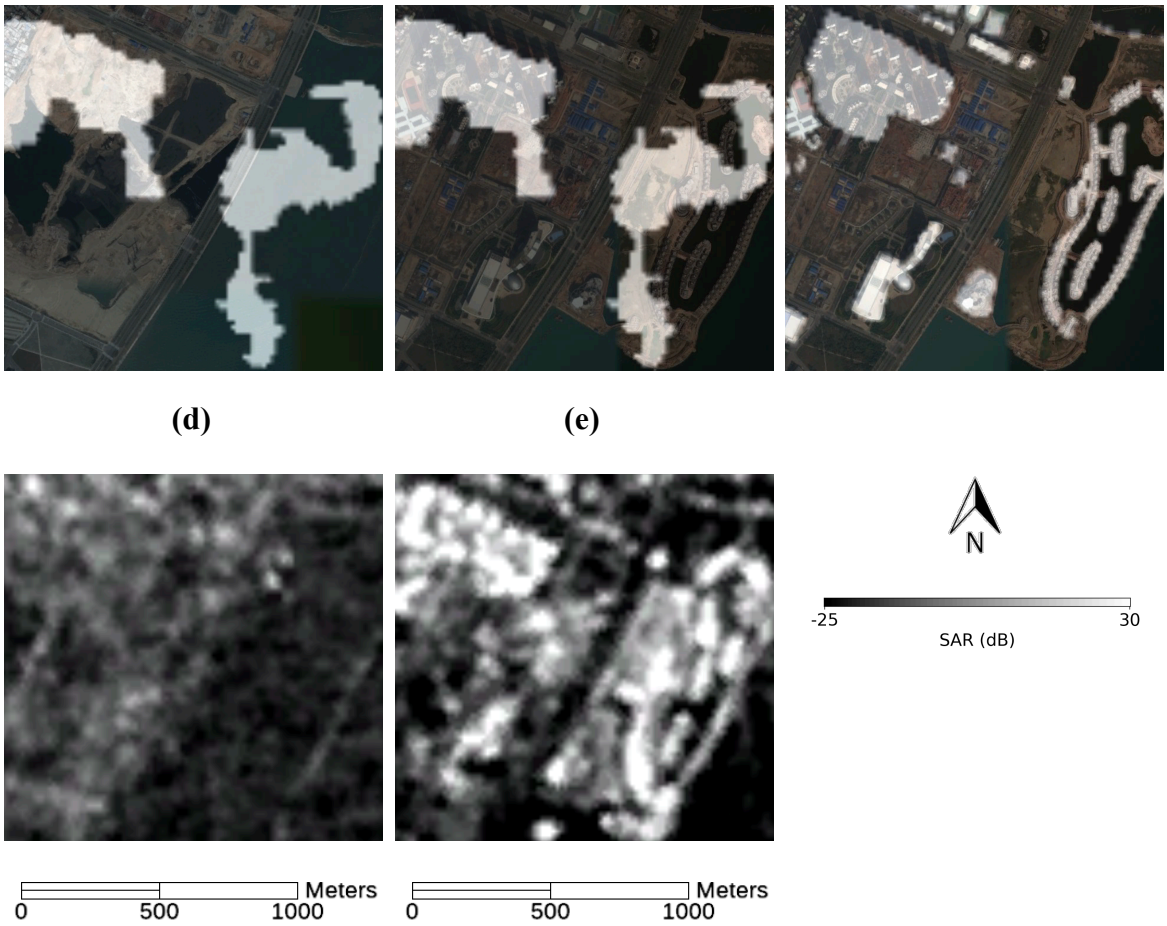


Figure 6.5. Example of detection results of CORN at Xiamen testing site. (a) CORN result overlays on Time 1 optical image, (b) CORN result overlays on Time 2 optical image, (c) ground truth overlays on Time 2 optical image, (d) Time 1 SAR image, (e) Time 2 SAR image.

This experiment on the Hanoi area shows that the CORN still has a limitation when dealing with an area with a wide range of building shape especially within the low spatial resolution image like ALOS-PALSAR image where the feature of the building cannot be clearly seen as evidenced by the low detection accuracy in Table 6.1, which means that the quality of the SAR images is partly a source of the error in detection result. However, the main part of the accuracy of detection is the deep learning model as shown in the higher accuracy of the CORN over U-net as the CORN can learn more various features causing effect of the low resolution of SAR image to be decreased.

6.3 Experiment on different observation looking direction

In addition to testing with other areas, the author further tested the model, which was trained with ascending SAR data, with descending SAR data. The result in Figure 6.6 shows that the proposed model can also be used with SAR data from another orbit. The reason for selecting the area shown in Figure 6.6 for testing is because it is the only area that available descending images has intersect with our ground truth, and thus the result can be calculated in term of accuracy. Please note that the ground truth is the same for the 12 January 2009/15 January 2010 SAR pair.

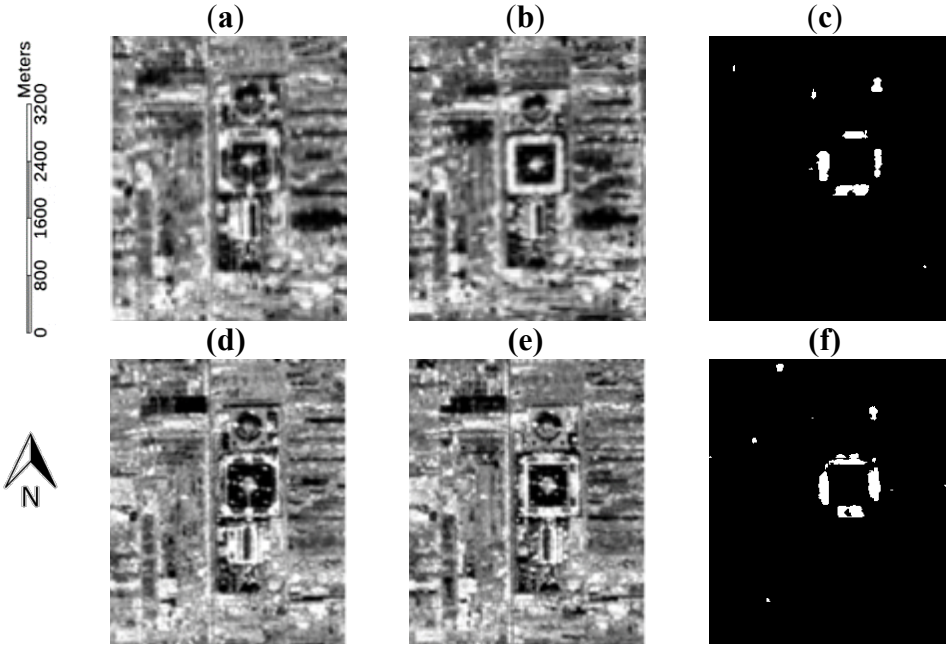


Figure 6.6. The result of the proposed method in ascending SAR image compare to descending SAR image data. The resolution of each image is 2.8 km × 3.5 km. (**a**) ascending SAR image from 27 November 2008; (**b**) ascending SAR image 15 January 2010; (**c**) the proposed result from ascending data; (**d**) descending SAR image from 18 September 2008; (**e**) descending SAR image from 9 August 2010; (**f**) the proposed result from descending data

The accuracy shown is in Table 6.3.

Table 6.3. Accuracy of the model with descending SAR data.

Validation Method	Ascending CORN	Descending CORN
False negative	71.741	44.615
False positive	0.071	0.727
Overall accuracy	98.475%	98.61%
Precision	0.891	0.540
Recall	0.283	0.554
F measure	0.757	0.541
F1 measure	0.429	0.546
Kappa	0.423	0.539
IOU	0.273	0.376

As there is no tall building constructed in between Time 1 and Time 2 images, the images are not suffered from geometric distortions, i.e. foreshortening and shadowing, which can affect the detection result since these distortions in ascending images will occur in the opposite direction in descending images. While the result shows that the model trained with ascending image can also be used with descending image, most of the accuracies resulting from using the model with ascending SAR images are slightly lower than those with descending SAR images of Bangkok (Table 6.3). The reason for a higher false positive rate in descending CORN model is because there are more spots with significant change in intensity values between a pair of descending images compared with ascending images, causing the model to detect construction more than in reality.

6.4 Experiment on different sensor

Past experiments show that the current model trained with images of Bangkok city can be used with other areas viewed from the same satellite. However, we wanted to show that it can also be used with SAR images from other satellites too. The author tested the model with a C-band SAR image from the Sentinel-1 satellite, while an image from the ALOS-

PALSAR training data was captured in L-band. Other properties were also different from those in the training data in many aspects; for instance, the resolution was 10 m/pixel and the polarization was VV. Please note that by using the image in VV polarization, the intensity reflected on the building is less than in HH polarization, meaning that the pixel intensity of the building would be lower than in HH image. The selected area was Chiang Mai in the northern part of Thailand, where most of the area is mountainous, while Bangkok, the city used in the training of the model, comprises mostly plain areas. Since the terrain is mostly mountain, it should be noted that the geometric distortions in SAR image, i.e. foreshortening, layover and shadowing could appear in an image and might affect the detection result of the model if the model's robustness is not high enough, similar to what happened with the experiment with U-net model in Chapter 4.3. Some parts of the detection results were cropped and are shown in Figures 6.7–6.9. As this area was an additional area to the previous work, the ground truth of this area was created, and thus, the validation was done by visual comparison with optical images, since accuracies cannot be calculated and shown in terms of numbers. The date of the Time 1 optical images in Figures 6.7 and 6.8 was 7 January 2016, while in Figure 6.9, it was 17 November 2015, due to the cloud cover problem. The Time 2 optical images in Figures 6.7 and 6.8 were from 29 October 2017; while in Figure 6.9, they were from 24 December 2017 due to the availability of the existing data. Please note that all of the optical images in this experiment were selected from Google Earth software, where images were captured by a variety of satellites and aircraft, meaning it is difficult to determine the image source. However, according to the rough data provided by the software, some images were captured with Landsat 7 at a 30 m/pixel resolution.

(**a**) (**b**) (**c**)

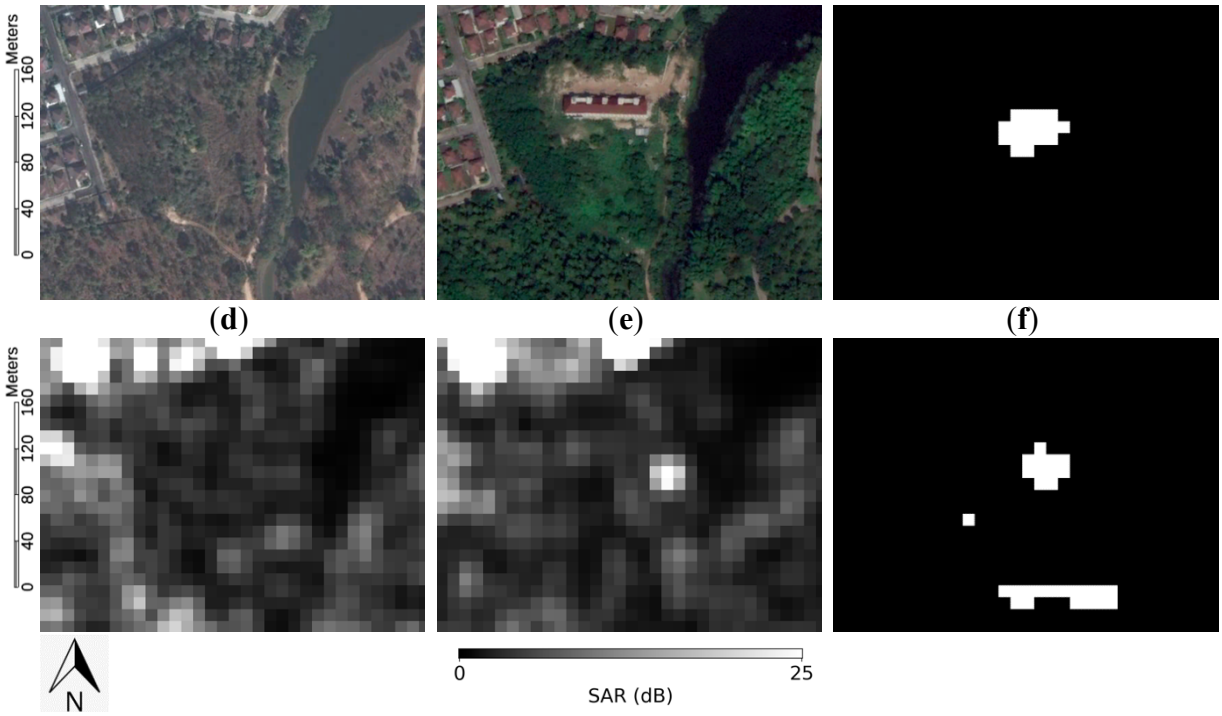


Figure 6.7. Detection results of the first area in Chiang Mai at 18°51'23.49"N 98°57'17.90"E. The size of each image is 0.32 × 0.23 km. (**a**) Time 1 optical data; (**b**) Time 2 optical data; (**c**) result of CORN; (**d**) Time 1 SAR data 'Copernicus Sentinel data [2015]'; (**e**) Time 2 SAR data 'Copernicus Sentinel data [2017]'; (**f**) result of U-net.

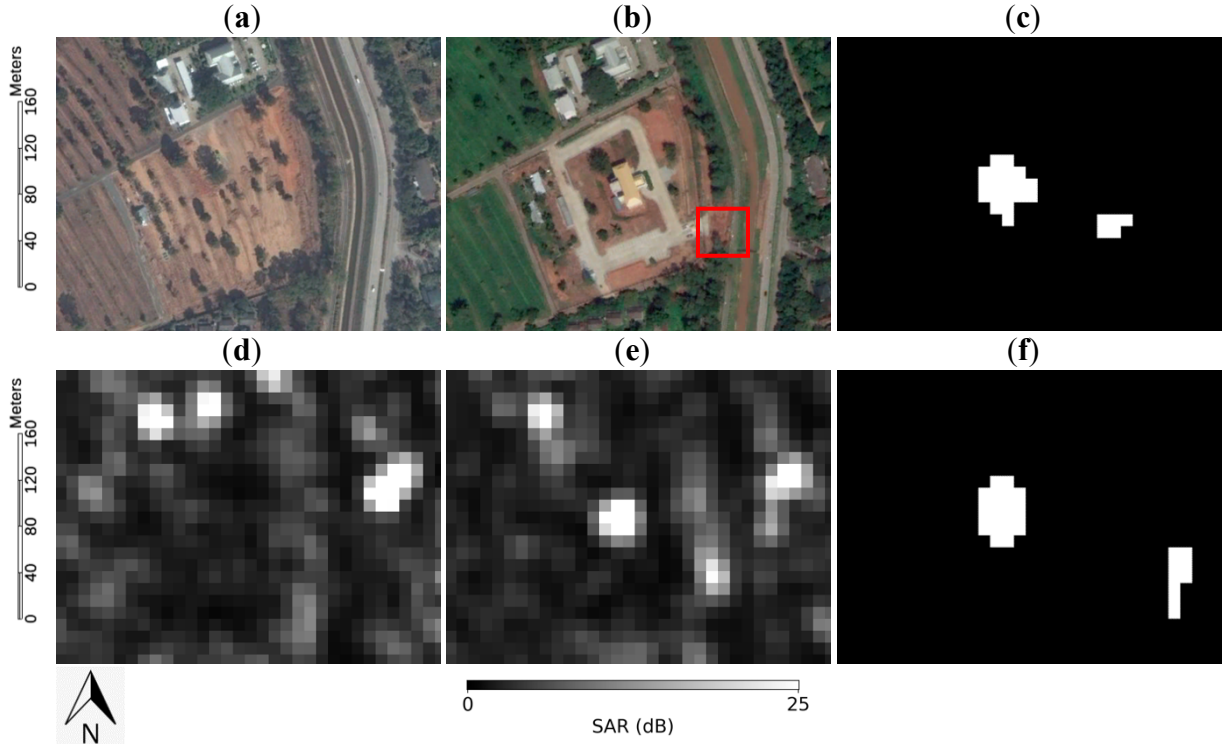


Figure 6.8. Detection results of the second area in Chiang Mai at 18°51'22.36"N 98°57'40.70"E. The size of each image is 0.32 × 0.23 km. (**a**) Time 1 optical data; (**b**) Time 2 optical data; (**c**) result of CORN; (**d**) Time 1 SAR data 'Copernicus Sentinel data [2015]'; (**e**) Time 2 SAR data 'Copernicus Sentinel data [2017]'; (**f**) result of U-net.

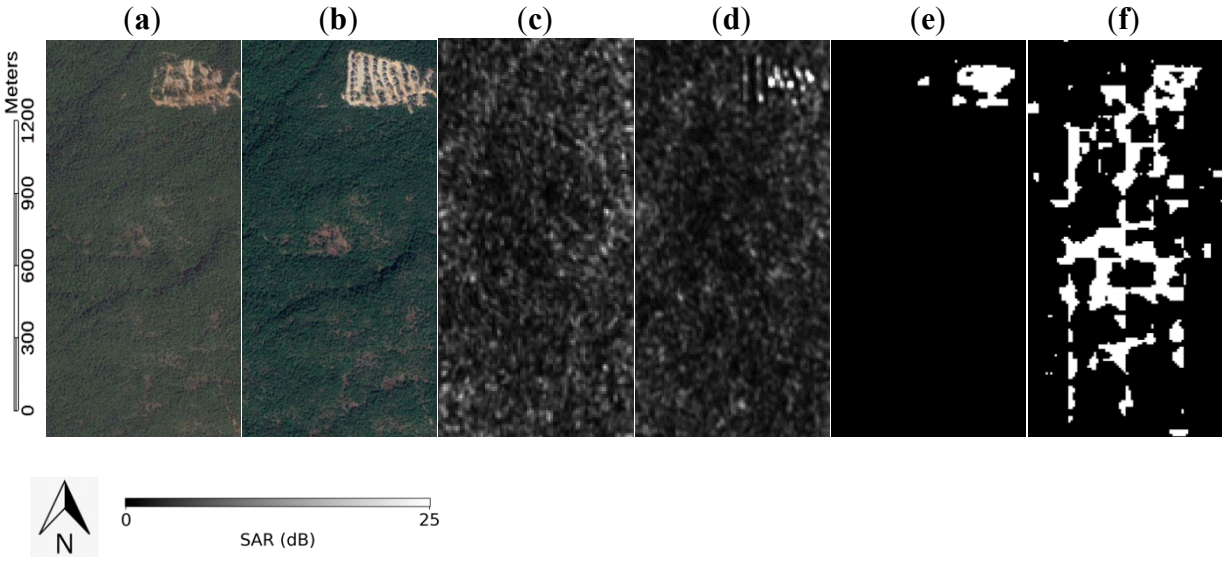


Figure 6.9. The result of the proposed model with an image from Sentinel-1 at 18°50'48.34"N 98°56'55.95"E. The size of each image is 0.8 × 1.75 km. (**a**) Time 1 optical data; (**b**) Time 2 optical data; (**c**) Time 1 SAR data 'Copernicus Sentinel data [2015]'; (**d**) Time 2 SAR data 'Copernicus Sentinel data [2017]'; (**e**) result of CORN; (**f**) result of U-net.

Although both our proposed network and U-net can detect new constructions, the results indicate that U-net generated more false positive results than CORN. In Figure 6.7, CORN correctly detected building change without any false detection, while U-net mistakenly detected forest area in the bottom part of Figure 6.7f. Please note that the high intensity spot in the middle of Figure 6.7e is not the building. In Figure 6.8, both CORN and U-net show two detected buildings in their results. They both correctly detected a building in the center of the image. However, another object that U-net detected was an existing road in the right

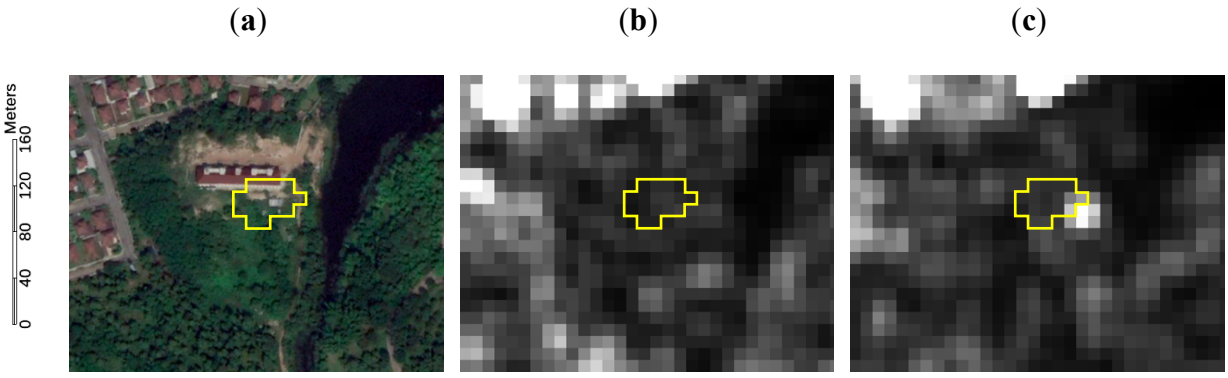
side of Figure 6.8f, which was a false detection, while the false detection in CORN is the area in the red rectangle in Figure 6.8b. From the selected Time 2 optical image from 29 October 2017, it is difficult to see what CORN detected in the area, but in the optical image from 3 March 2018 (Figure 6.10), there is a bridge placed next to CORN's detected area. Thus, it can be assumed that the other object detected by CORN was a bridge under construction, since our Time 2 SAR image was taken on 24 December 2017, which was around the middle of the dates that these two optical images were taken.

Figure 6.10. Optical image of the second area in Chiang Mai on 3 March 2018 at 18°51'22.36"N 98°57'40.70"E.

In Figure 6.9, even though U-net was also able to detect construction in Sentinel-1 data, as seen in the top right corner area, it failed to handle data containing changes in mountain areas and ended up involving them in the detection result instead. On the other hand, CORN was able to avoid the intensity change of mountain areas and detected only the building changes. This experiment suggests that by combining the training of ordinary and reverse time-series data, our model can eliminate a greater variety of false positives, which means it can be used even with images from other satellites. Please note that the building changes in Figure 6.7 and 6.8 are in sizes of approximately 700 m^2 and 300 m^2 respectively, which

is smaller than those in the created ground truth. This is an evident that the model can detect the single small building changes when the testing image has a high resolution.

Further evidence that shows the benefit of the more robust detection provided by our model is the result of using Sentiel-1 satellite. While the area in Figure 6.7, which was captured in a completely different setting to those in the training data, involved a lot of intensity change (as shown in Figure 6.11), CORN was able to detect the building correctly and avoided seasonal changes even though they have a similar intensity of change to buildings, especially the bright spot in the middle of the image which is brighter than the actual building, making it visually very similar to the building change. Although U-net was also able do the same thing, it still detected many incorrect changes in the image. Please note that in Figure 6.7, the detection results for both networks appear to shift to the south-east from the real building location in the optical image. This is probably because the intensity of the building in the SAR image was too low to display obvious features, and therefore it was difficult for the models to detect it in its exact position. The reason this building has a lower intensity than the surrounding buildings is because its roof shape and orientation are different from the other buildings. The high intensity of surrounding buildings is possibly the result of the double bounce on the walls or a strong single bounce on the roofs, which did not happen with the detected building due to the reason stated above.



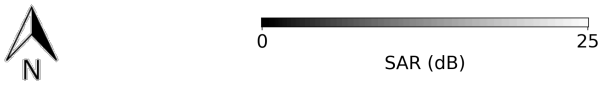
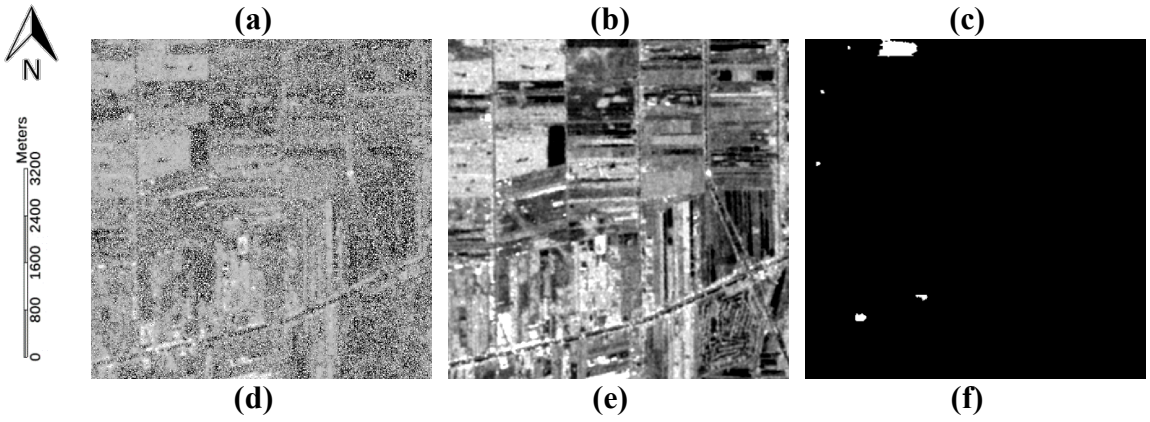


Figure 6.11. Detection results of CORN in the first area of Chiang Mai at 18°51'23.49"N 98°57'17.90"E in yellow hollowed polygon overlays on (**a**) a Time 2 optical images; (**b**) a Time 1 SAR image 'Copernicus Sentinel data [2015]' and (**c**) a Time 2 SAR image 'Copernicus Sentinel data [2017]'.

The result in this experiment support that the using of the higher spatial resolution SAR image gives higher accuracy detection especially on the small buildings. However, the limitation of CORN still persists in low resolution SAR images such as ALOS-PALSAR image.

6.5 Noise robustness

The model trained with CORN is robust against the noisy images, whether the noise is on either one of the images in images pair or both of the images in a pair is noisy. Speckle noise is a multiplicative noise [1]. Our noisy images generated from Gaussian distribution. The model can handle whether the case as shown in Figure 6.12 and 6.13.



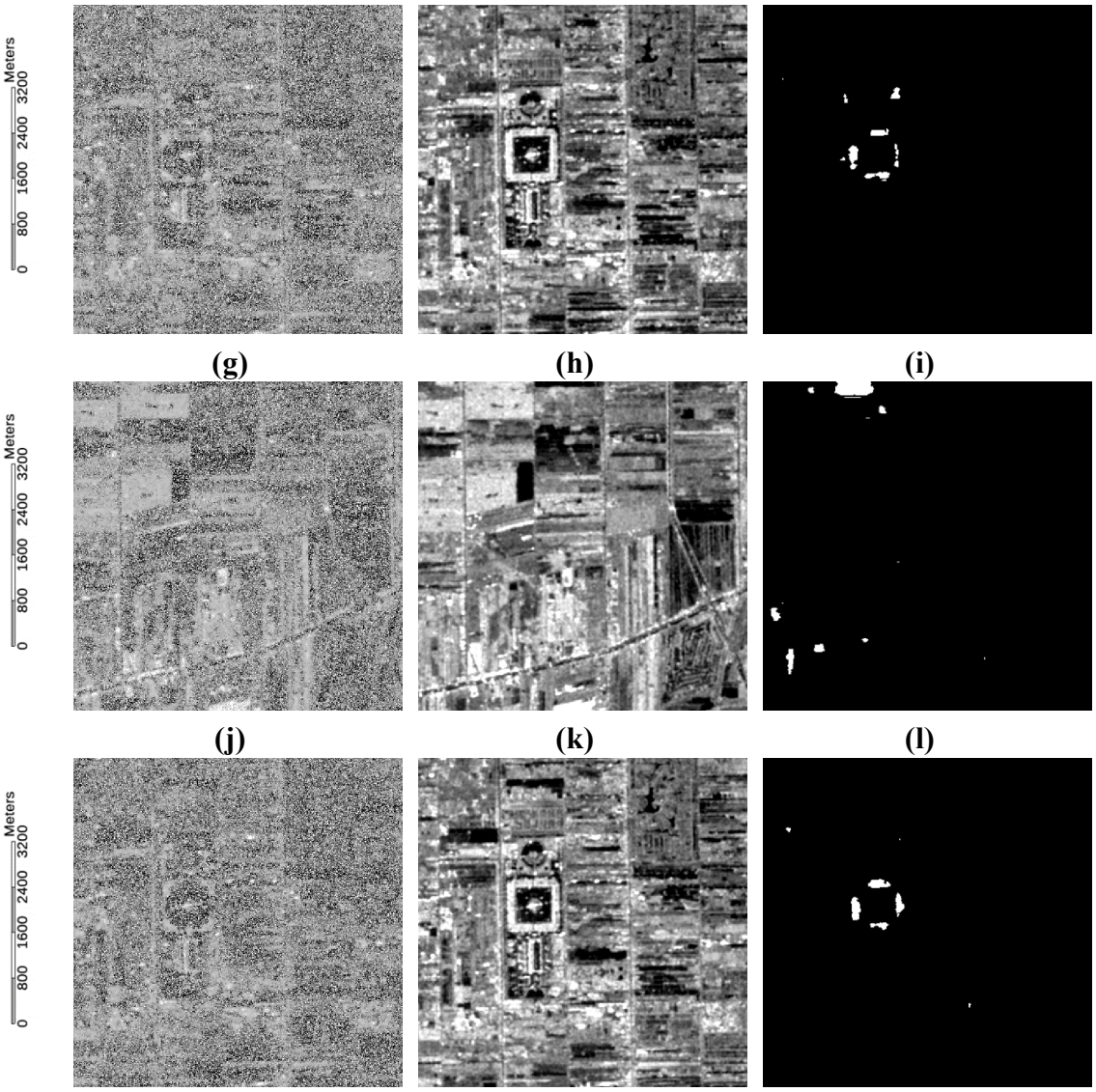


Figure 6.12. Detection result of when Time 1 image is full of noise. (a),(b),(c) show Time 1, Time 2, result of CORN in the first area of SAR pair 27 November 2008/15 January 2010 respectively; (d),(e),(f) show Time 1, Time 2, result of CORN in the second area of SAR pair 27 November 2008/15 January 2010 respectively; (g),(h),(i) show Time 1, Time 2, result of CORN in the first area of SAR pair 12 January 2009/21 November 2009 respectively; (j),(k),(l) show Time 1, Time 2, result of CORN in the first area of SAR pair 12 January 2009/21 November 2009 respectively

The result of when both images are noisy is shown as Figure 6.13.

(a) **(b)** **(c)**

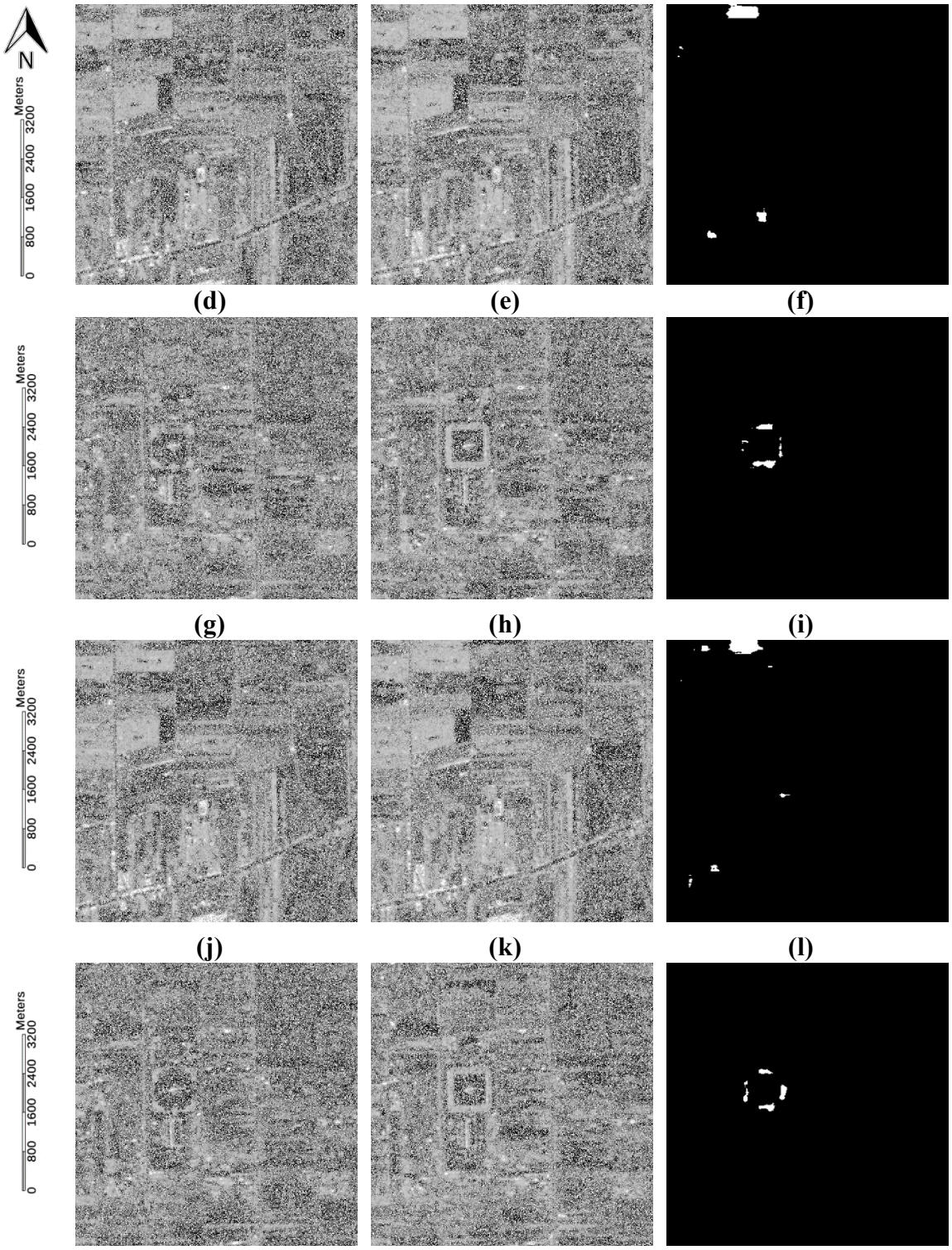


Figure 6.13. Detection result when both Time 1 and Time 2 are full of noise. (a),(b),(c) show Time 1, Time 2, result of CORN in the first area of SAR pair 27 November 2008/15 January 2010 respectively; (d),(e),(f) show Time 1, Time 2, result of CORN in the second area of SAR pair 27 November 2008/15 January 2010 respectively; (g),(h),(i) show Time 1, Time 2, result of CORN in the first area of SAR pair 12 January 2009/21 November

2009 respectively; (j),(k),(l) show Time 1, Time 2, result of CORN in the first area of SAR pair 12 January 2009/21 November 2009 respectively

The accuracy of when only Time 1 and both Time 1 and Time 2 are noisy is shown in Table 6.4.

Table 6.4. Accuracy of when both Time 1 and Time 2 are full of noise

Validation Method	Normal CORN	Noisy Time 1 CORN	Noisy Time1/Time2 CORN
False negative	54.667	60.025	68.256
False positive	0.210	0.193	0.147
Overall accuracy	99.791%	99.205%	99.168%
Precision	0.687	0.678	0.687
Recall	0.453	0.400	0.317
F measure	0.659	0.641	0.627
F1 measure	0.546	0.503	0.434
Kappa	0.543	0.499	0.431
IOU	0.376	0.336	0.277

The accuracies of both cases indicate that the model is robust against noise. In the case of when Time 1 image is noisy, the accuracy is only slightly lower than the noise-free image scenario and, in fact, even higher than the noise-free U-net model. However, when both Time 1 and Time 2 images are noisy, the accuracy drop is more significant but it still maintains at the acceptable rate where all the oblivious changes are detected.

6.6 Summary

This chapter demonstrate the versatility and robustness of the proposed algorithm. The model trained with Bangkok dataset has been used in detecting new constructions in Hanoi and Xiamen and still got the accurate result where the Kappa and IOU accuracies more than

0.25 and 0.15 respectively which are enough to satisfy the objective of detecting images from different areas as it shows the ability to detect some major changes. The model has been further tested with the image with completely different setting including the different orbit direction and different satellite and still able to detect the new constructions with decent accuracy and satisfied at more than 0.5 Kappa and 0.3 IOU. The model also can work under the noisy images where the accuracy is only slightly lower than the noise-free image where Kappa and IOU were satisfied ant more than 0.4 and 0.3 respectively. To summarize, the scenario that CORN is able to make an accurate detection is the high-resolution SAR image such as in Chiang Mai area in Sentinel-1 image, which is mountainous area with forest land cover type, while flat land with agriculture land cover type such as Bangkok is also able to get the accurate detection result. Xiamen area which is island area is also can be detected with high accuracy. However, the Hanoi area, which is plain area with complex building shape, is still not in a very high accuracy due to their shape of building change that are too different from those in ground truth in training data. This can be concluded as the type of land cover may not cause major effect on detection result as long as the features of the construction in term of intensity change and shape or size are still the same.

Reference

[1] Xie, H.; Pierce, L.E.; Ulaby, F.T. Statistical properties of logarithmically transformed speckle. IEEE Trans. Geosci. Remote Sens. 2002, 40, 721–727.

Chapter 7. Conclusions

The past studies show that the people residing in urban area is increasing and the number is expected to be even higher in the future, resulting in a high demand for the construction of new residential and business areas, especially in developing countries in Asia. Continuing urbanization or migration from rural to urban areas will eventually cause environmental deterioration, inadequate housing, traffic congestion, slums, the rising crime rate, homelessness. Thus, this issue needs to be addressed otherwise the population could face the consequences of poor urban planning management.

The remote sensing data has proven to be an effective information for monitoring urban expansion in many publications by employing techniques for automatically detecting changes that have occurred between any bitemporal acquisitions of time-series data. The using of synthetic-aperture radar (SAR) which captures images using microwave signals that can penetrate clouds in change detection has been actively studied. However, the properties of SAR image cause most of the conventional methods that based on a fixed mathematical condition are incapable since it is difficult to identify one specific change, such as the appearance of new buildings, as any kind of change similar to the target change would be involved in the results. Traditionally, such specific detections have been made manually, based on the experience of human experts, but manual methods are expensive, time-consuming and error-prone. Besides, most existing algorithms are designed for dealing with data provided by the same sensor, with the same spectral range and number of spectral bands, and are sensitive to pre-processing and noise.

Instead of limiting to the fixed condition, deep learning simulate the experience-based learning mechanism of the human brain using training data and ground truth data in the

same way that humans learn. This property makes the trained deep learning model robust against a diverse range of intensity change behaviors, geographical terrains and other physical properties of SAR images. However, to reach the mentioned robustness, the deep learning requires a large amount of training data and the corresponding ground truth in term of quantity and variation, which are, unfortunately, difficult to be obtained.

As a result, the author has defined our own dataset used in this study and create our own ground truth for training the deep learning network. The training data includes a Bangkok area from ALOS-PALSAR L-band sensor, while the testing area are Bangkok, Hanoi and Xiamen captured from the same sensor. The ground truth of these dataset used for network training and testing was created manually by drawing polygons on optical images obtained from Google Earth. Since the created ground truth has unbalanced class weights problem, the author tackles it by weighting the loss function with the ratio between the positive class percentage and negative class percentage. In addition, one more study area has been added which is captured by Sentinel-1 C-band sensor at Chiang Mai to use in testing the versatility of the trained model.

Despite the preparation of the dataset, the creation of the ground truth is involving a lot of manually labeling works, which is very tedious and resulting in only a moderate quantity with narrow study areas. As a result, when the author tested the U-net model trained with the prepared training set, the result suggests that the U-net model requires large quantity of data in order to reach the desired robustness as the experimental result is only has high accuracy when the testing area is Bangkok. The experiment also shows that the U-net will fail to perform a detection if the number of training data is very low.

Thus, this thesis proposes a novel deep learning architecture that can generate an accurate result under the circumstance of limited training data for the purpose of newly built

construction detection in SAR images so called “Chronological order reverse network” (CORN). While normally, the detection of new buildings is supposed to use the data in Time 1–Time 2 format, the proposed architecture takes both Time 1–Time 2 and Time 2–Time 1 formats of data as training data to allow learning based on both of the changing features to make it more viable. With this architecture, the training data of the network is appeared to be more variation while the actual used data were in the same quantity. The proposed architecture adopts the advantage of U-net which is skip connection that passing the low-level features from encoders to decoders to allow it to generate a result with more solid boundaries. In detail, the architecture consists of two U-net bases for the network to learn differences—both forward and backward—by training it using Time 1–Time 2 and Time 2–Time 1 data. The features between the two U-net adaptations are shared through encoder 8, and the addition of encoders before feeding to decoders via skip connection.

The model trained with dataset of SAR images of Bangkok area captured by ALOS-PALSAR has demonstrated the successful of the proposed architecture as it is able to detect the newly built constructions in multiple terrains including Bangkok, Hanoi and Xiamen area with higher accuracy than U-net, one of the state-of-the-art methods, at the same number of training set. Moreover, the model can perform the detection on SAR images at Chiang Mai, Thailand, which is capture by the Sentinel-1 satellite with mostly different properties, including radar wavelength, spatial resolution and polarization. The experiments in this thesis further show that the model trained with proposed architecture can be used with images taken from the descending orbit direction of a satellite and also robust against noisy image. The source code of the method proposed in this thesis is available at https://github.com/Raveerat-titech/CORN.

Future work

As the proposed method has been successful so far, there are some possibilities to further develop this architecture in the near future. One possibility is to directly improve the model to make it can handle more variety types of terrain with more accuracy by training with more data. The mentioned training data could be generated by the proposed model itself by using the model with many kinds of areas to use the detection result as a future training data. The more benefit from increasing training data in the future work is to deal with foreshortening that might occur in tall buildings as such object does not appear in the current dataset. In addition, the possibility of detecting change of small buildings in a low-resolution SAR image by creating ground truth of such changes to be used as a training set will be studied in the future. Moreover, the proposed method has a possibility to apply to other objective outside the building change detection, such as the earthquake damage detection, or it could be modified to be used in other fields of study such as using with medical images (X-ray, MRI, etc.).

Acknowledgement

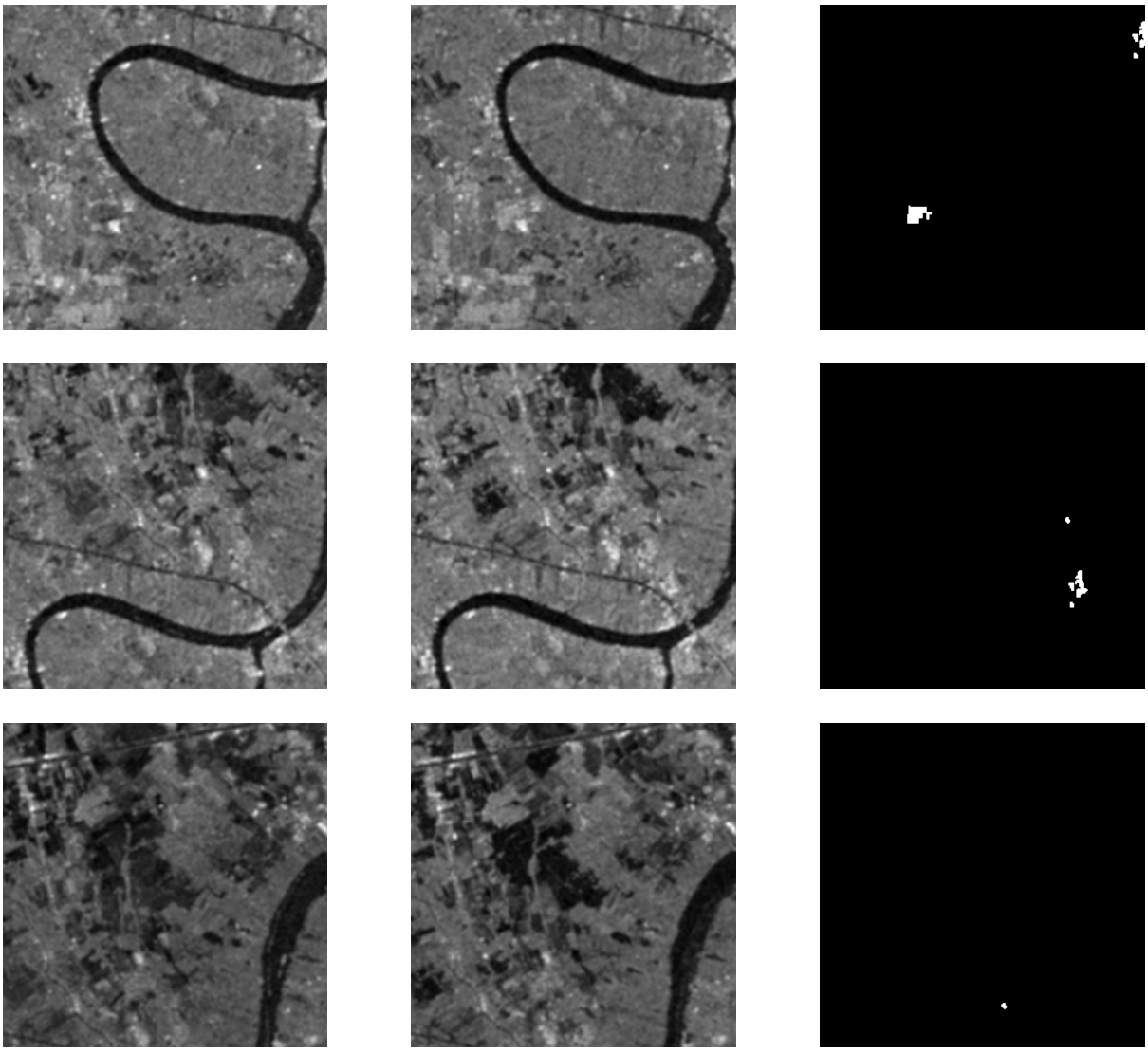
Through tough times of writing the thesis, there were many people selflessly offered their help which I would like to express my gratitude to. First, I would like to show my appreciation to the people from AI team at AIST for providing me many knowledges I have never known before. I would like to thank you to SpaceShift and Prof. Teerasit from Kasetsart University for exchanging ideas leading to this thesis. Most of all, thank you to Matsuoka-sensei for navigating me through this long journey and helping me with everything. I am proud to a part of his lab and will forever remain this feeling in my heart.

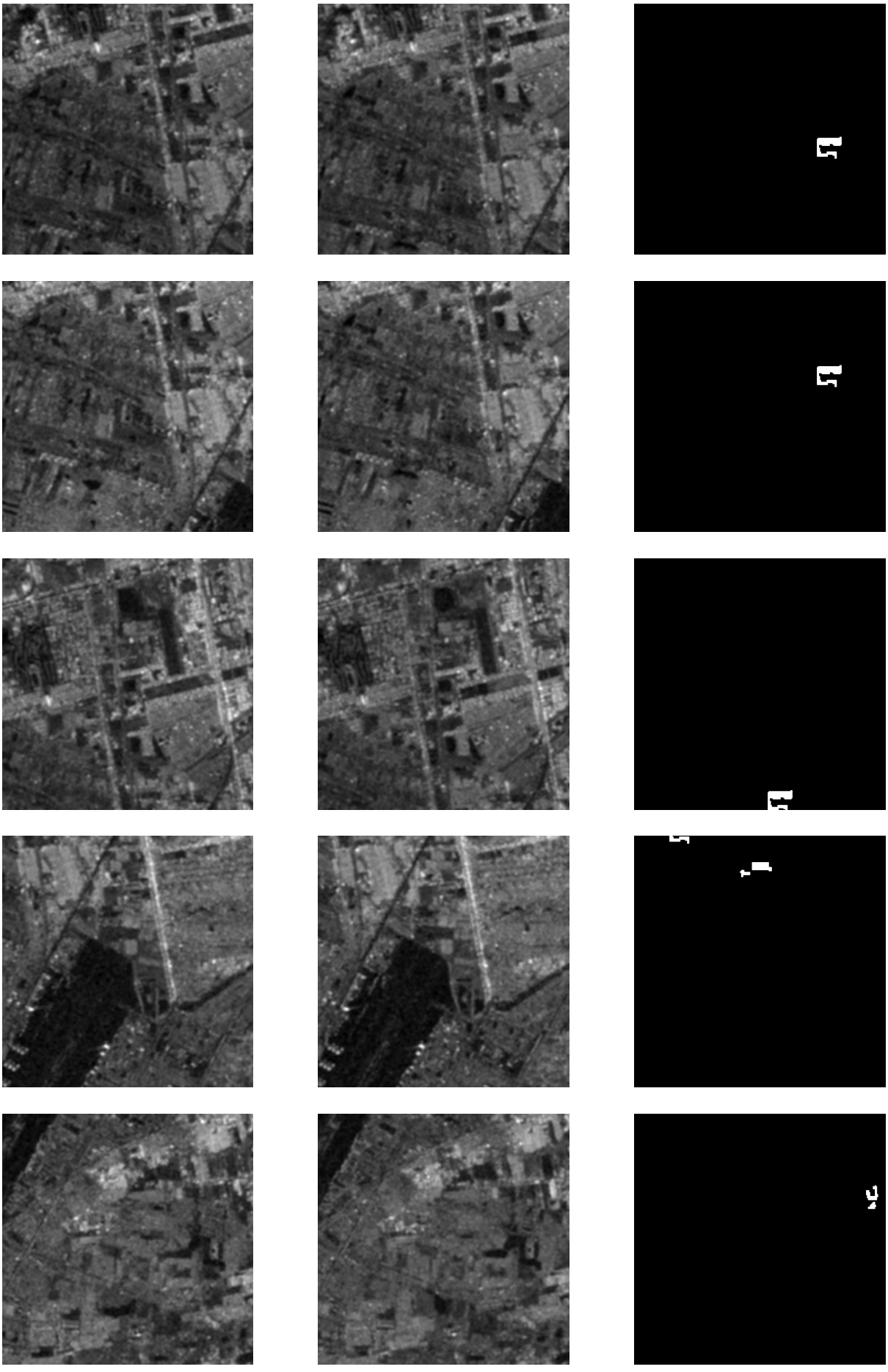
Appendix-A

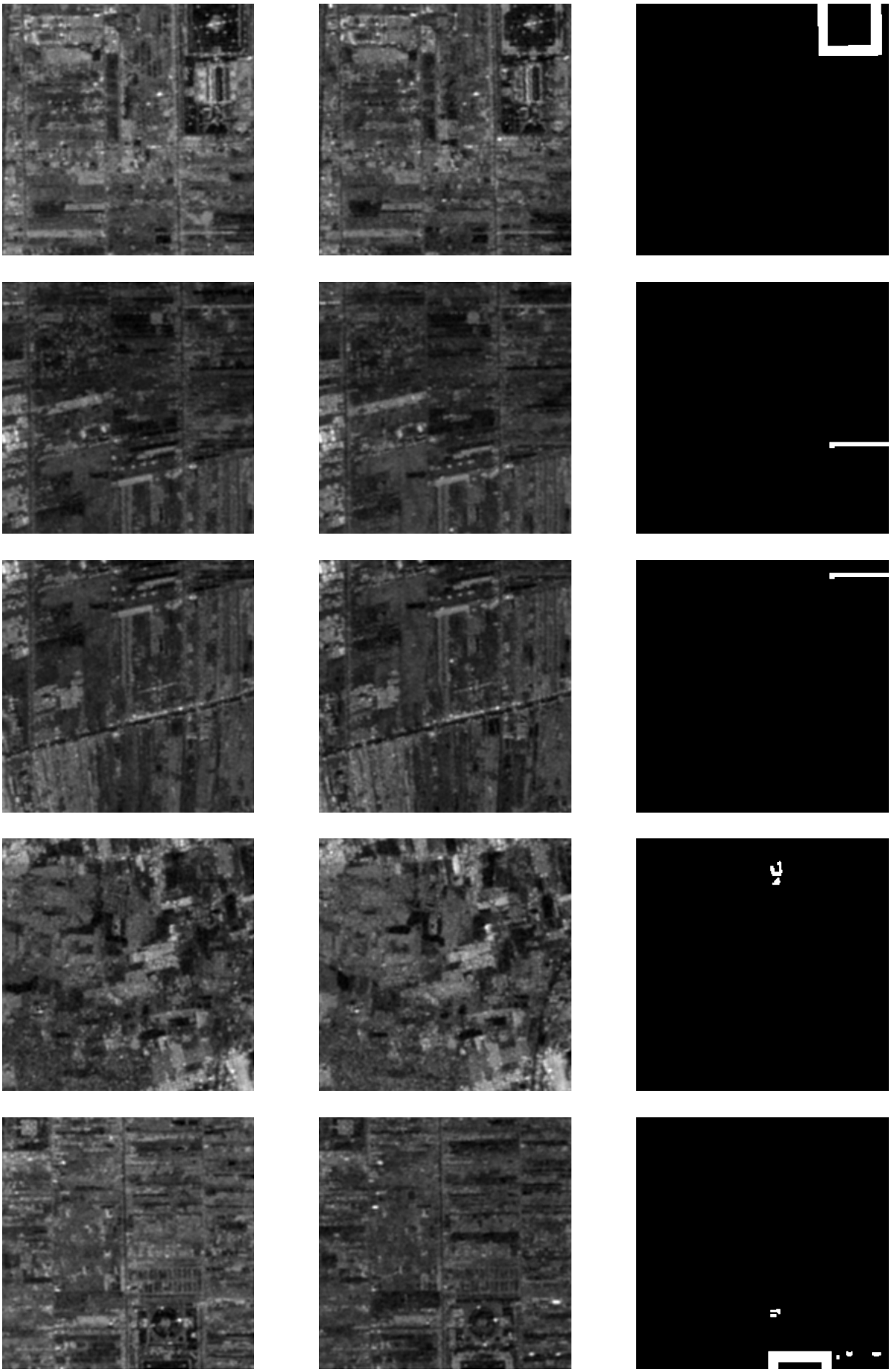
Training data

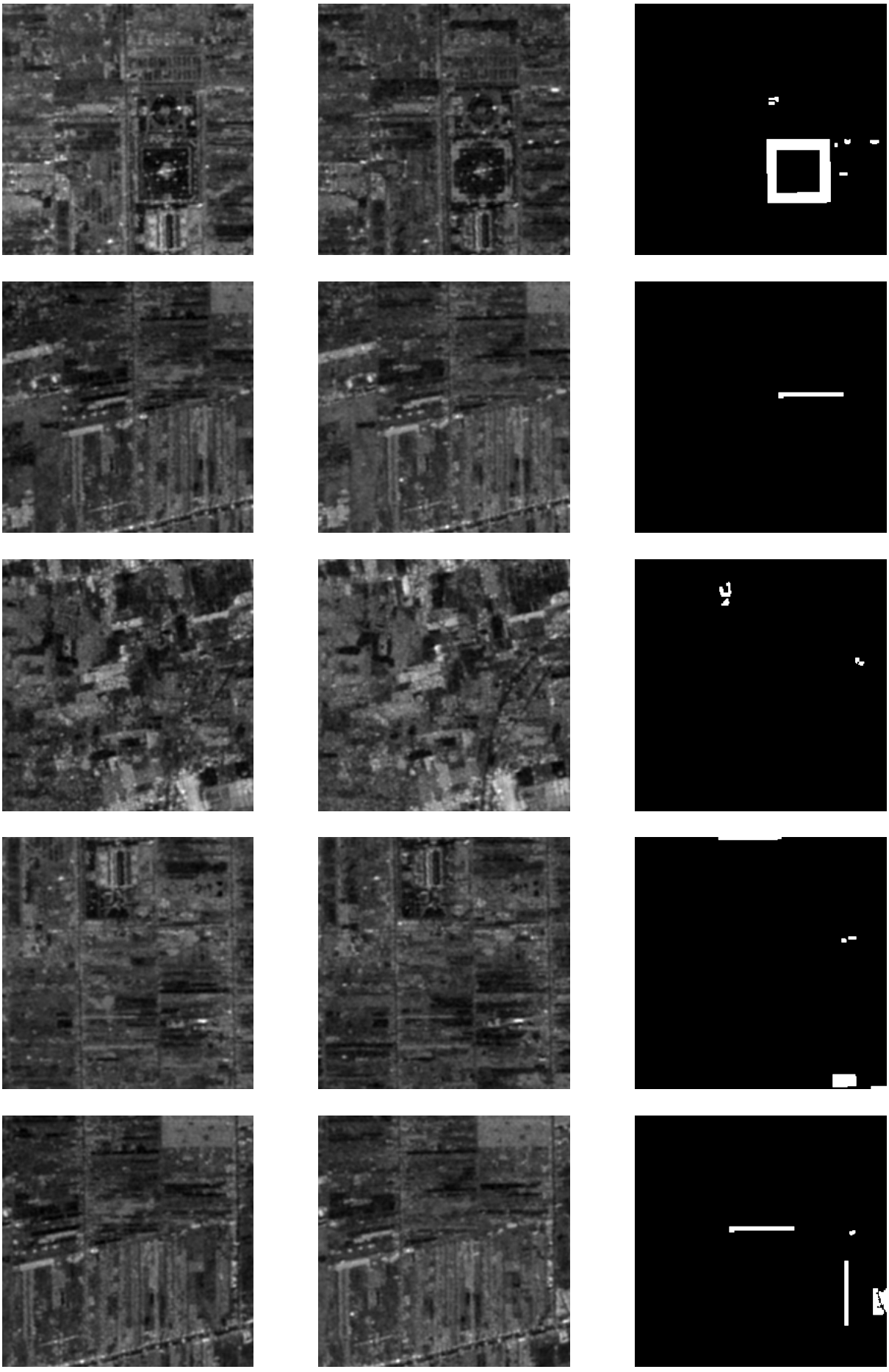
This section contains 200 examples of the SAR images used for training the network proposed in this thesis (CORN).

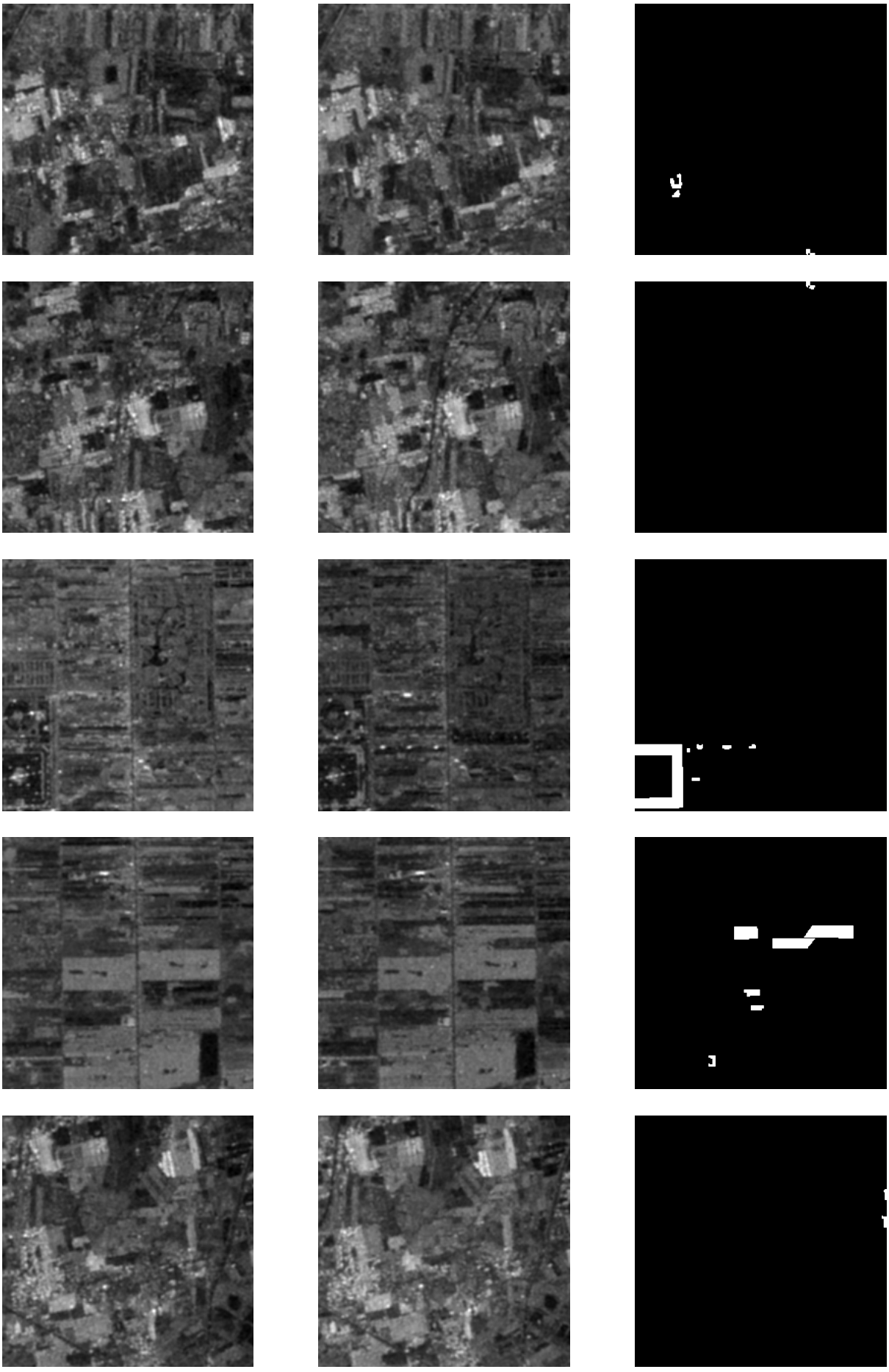
Time 1 | Time 2 | Ground truth

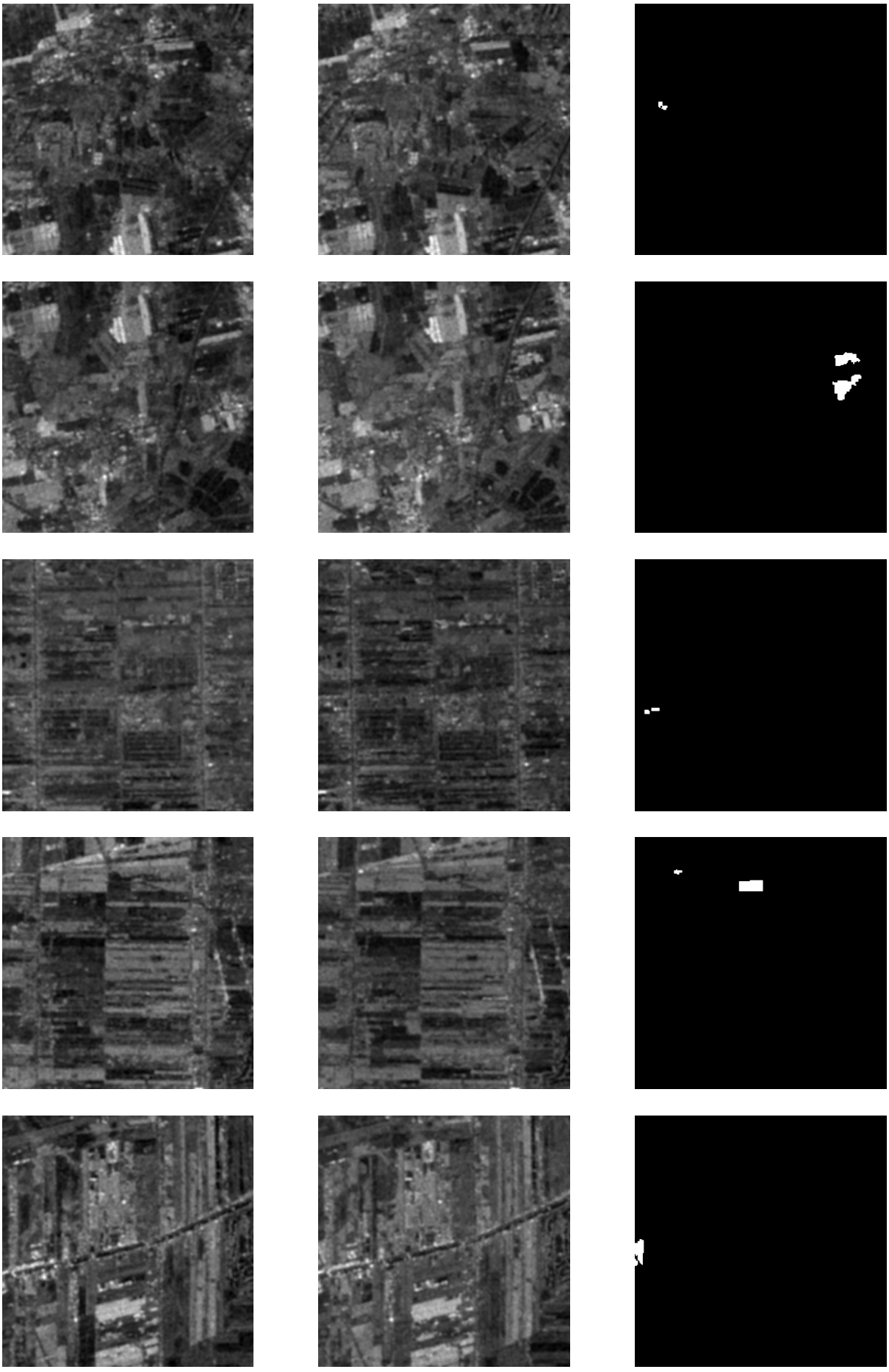


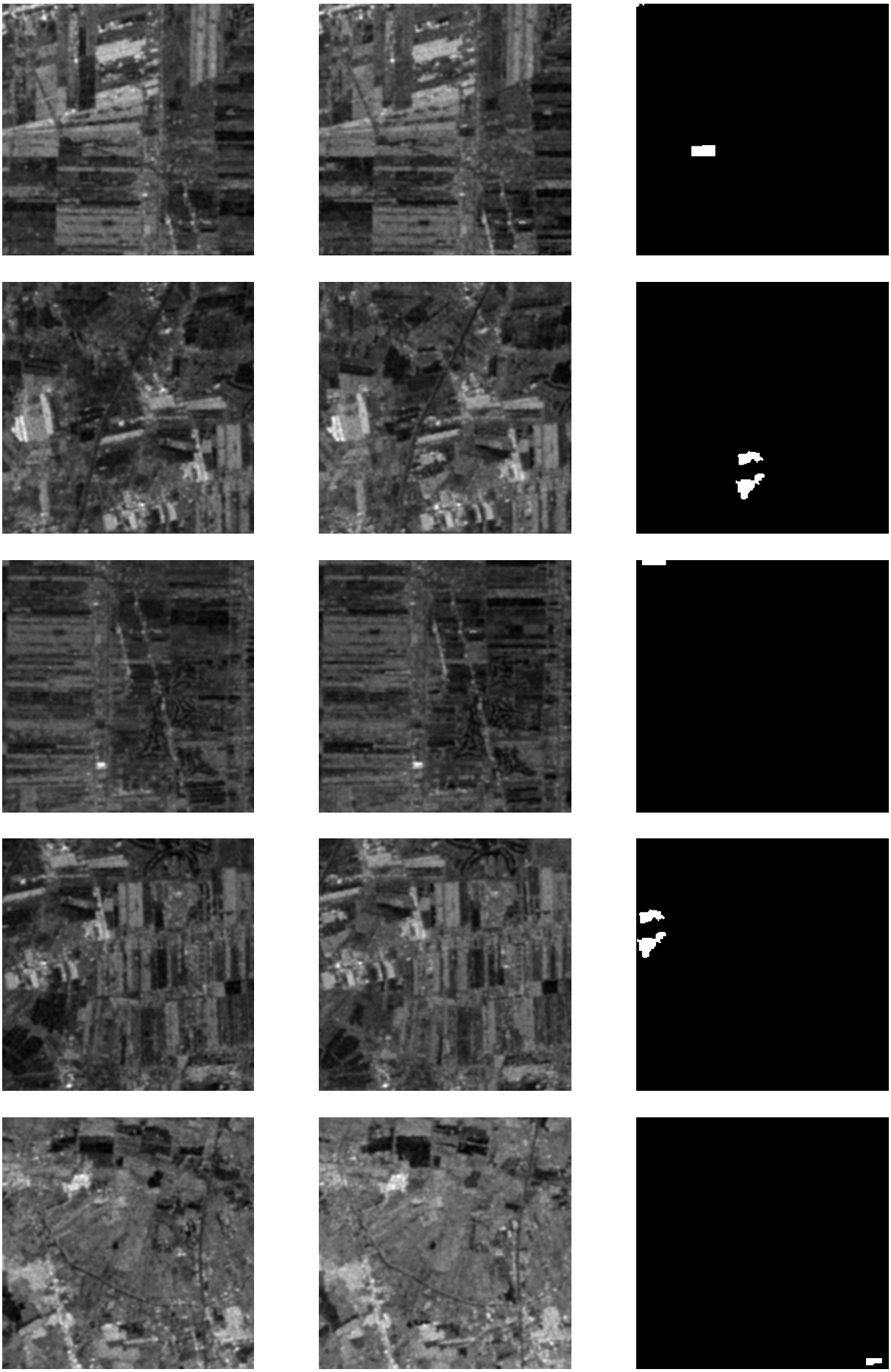


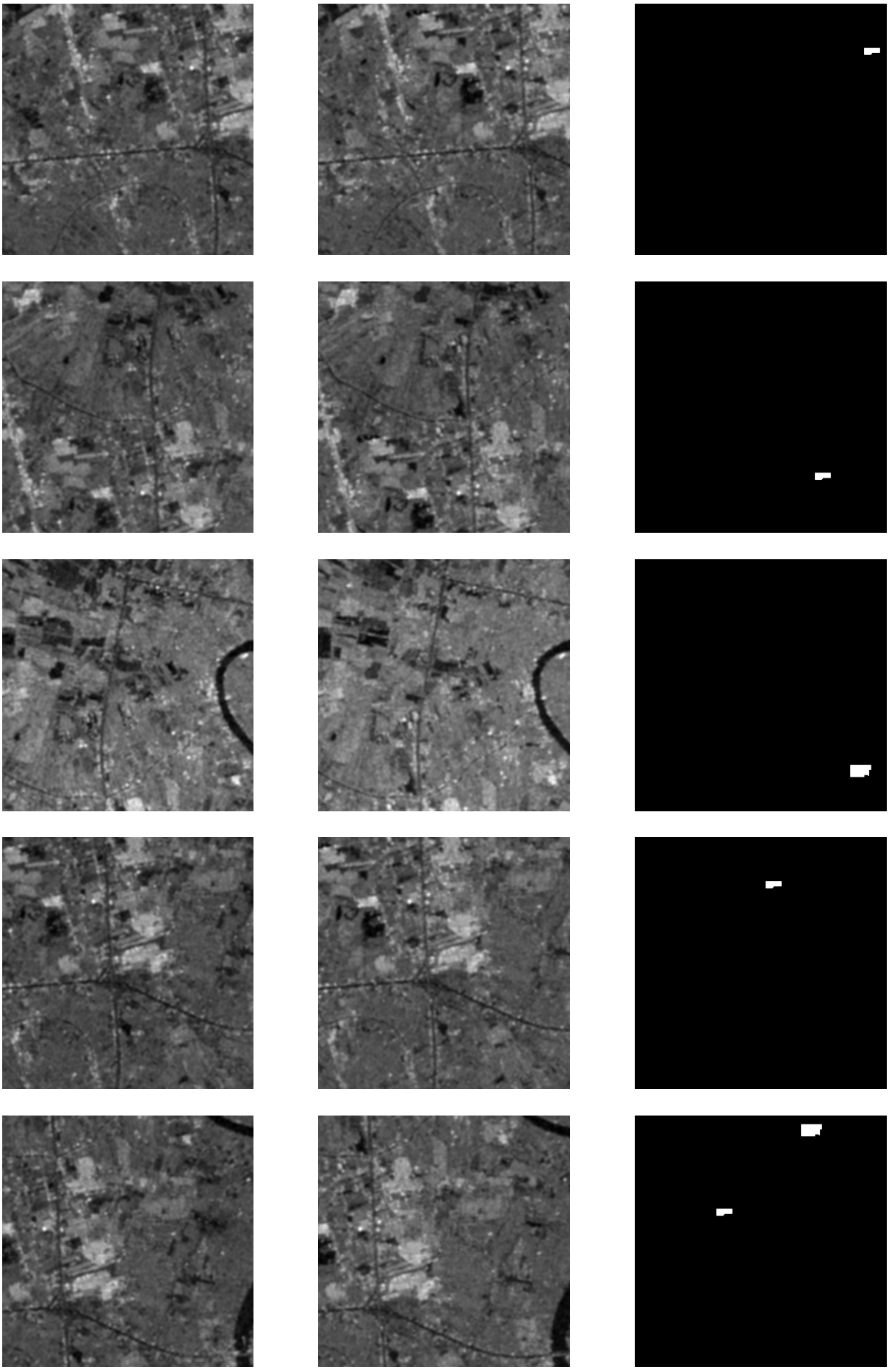


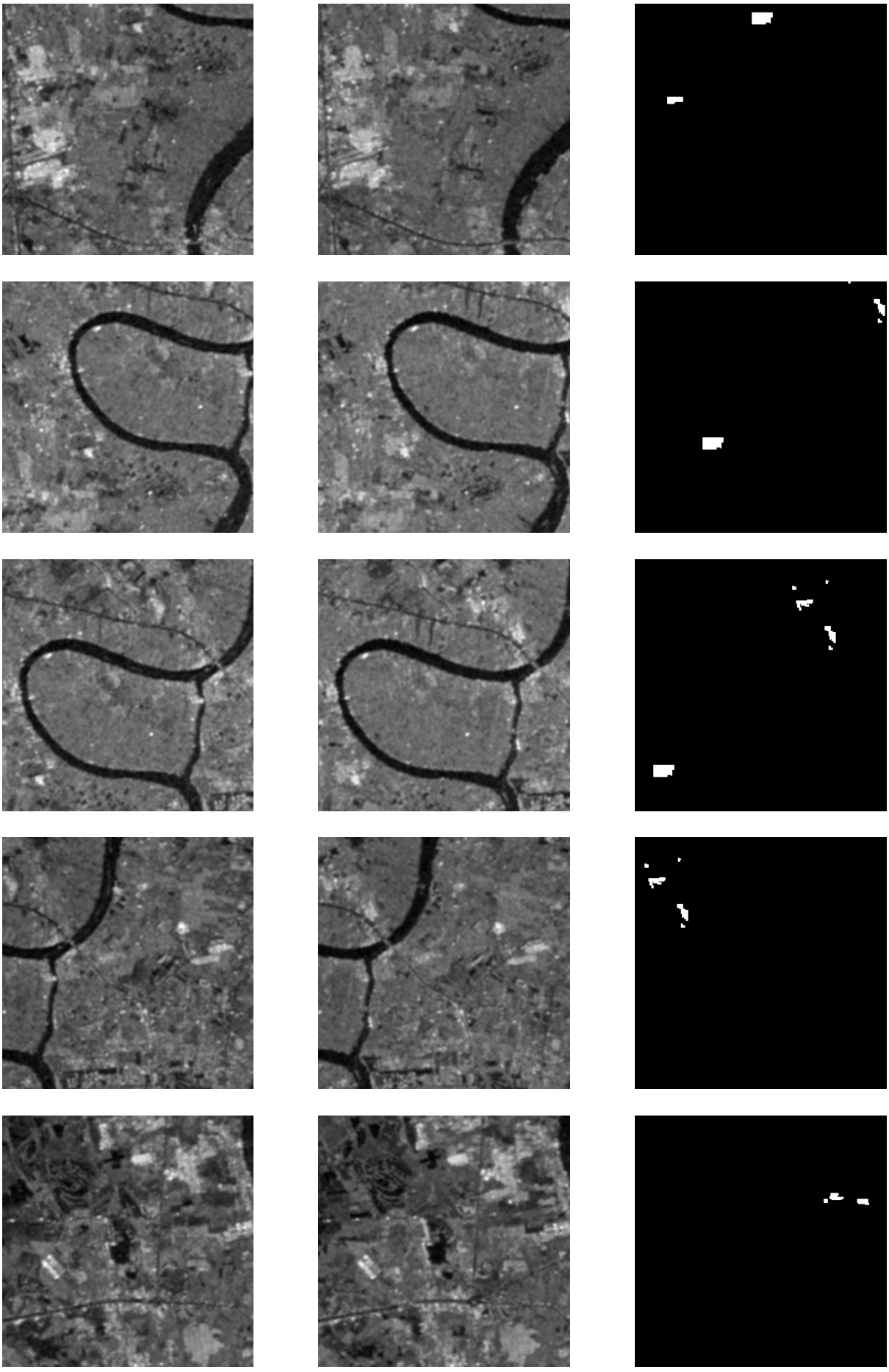


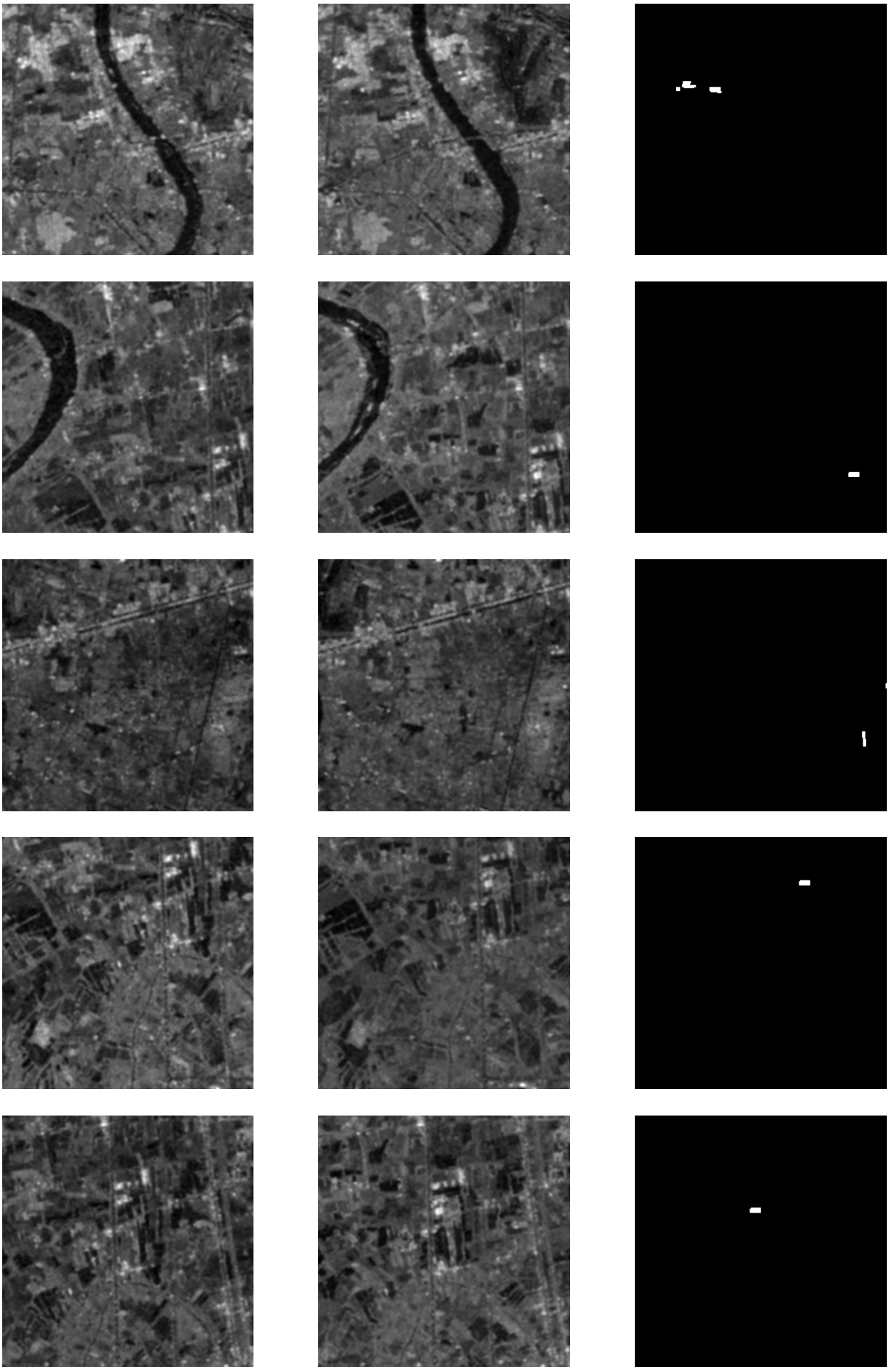


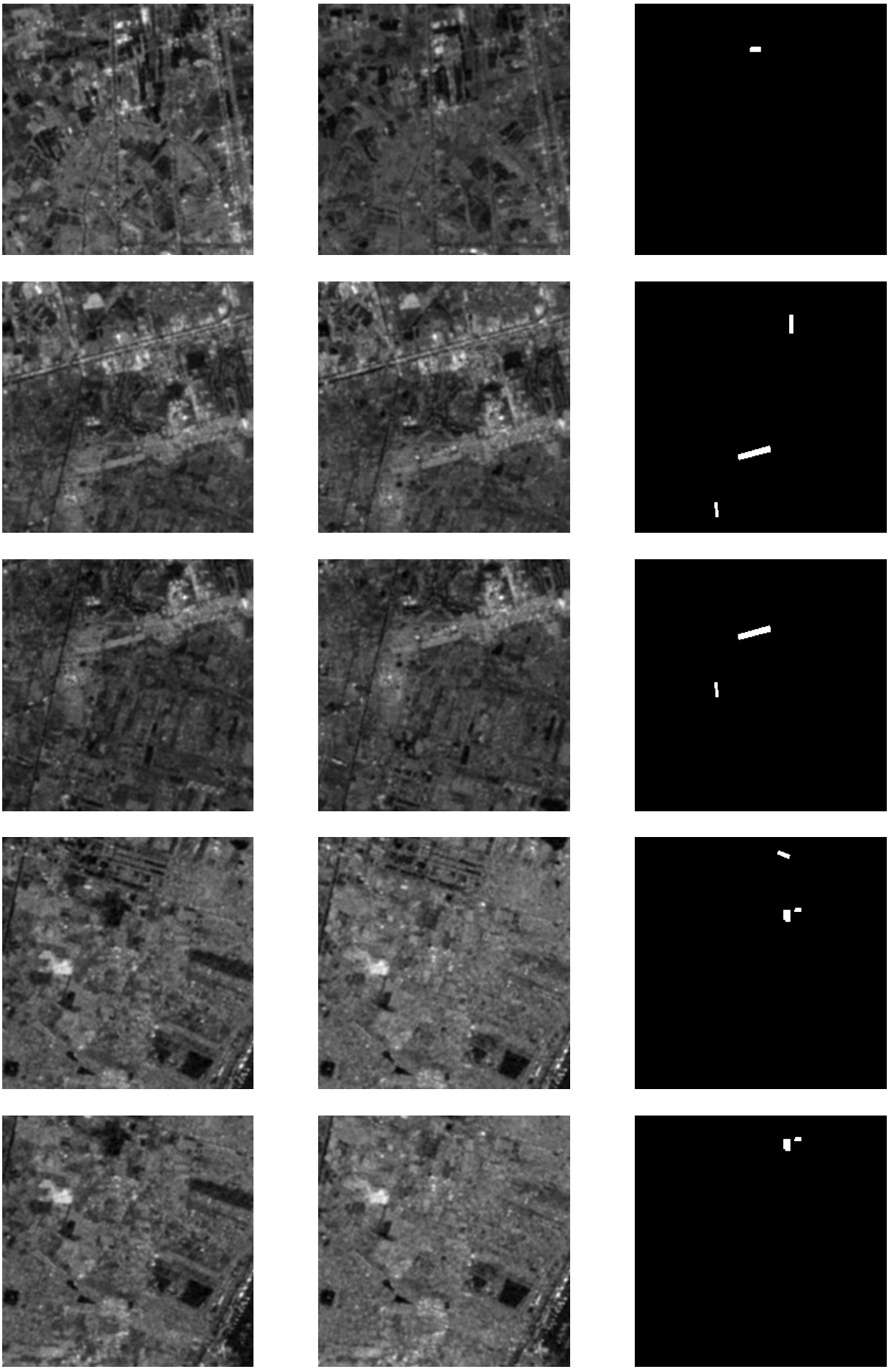


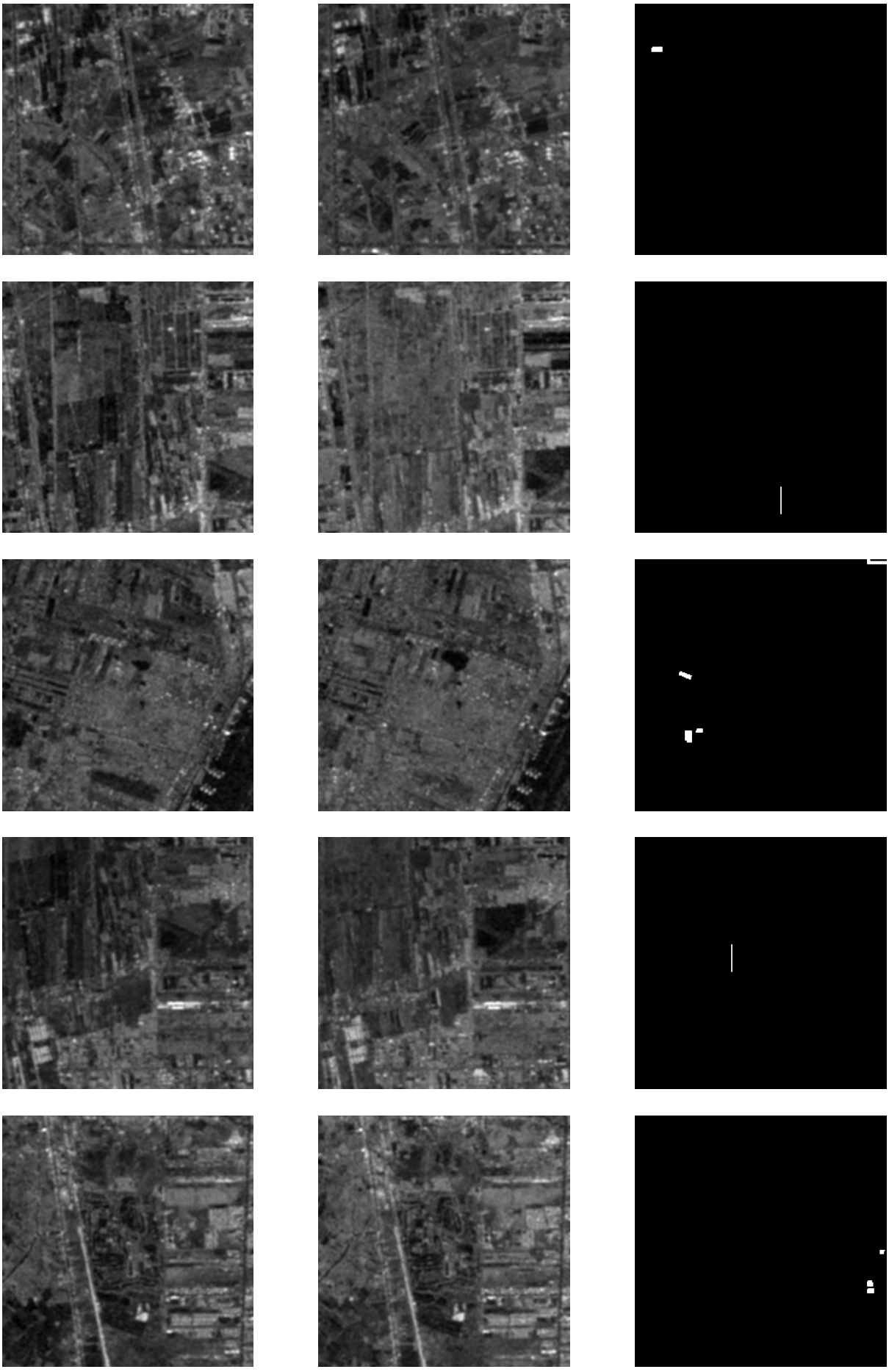


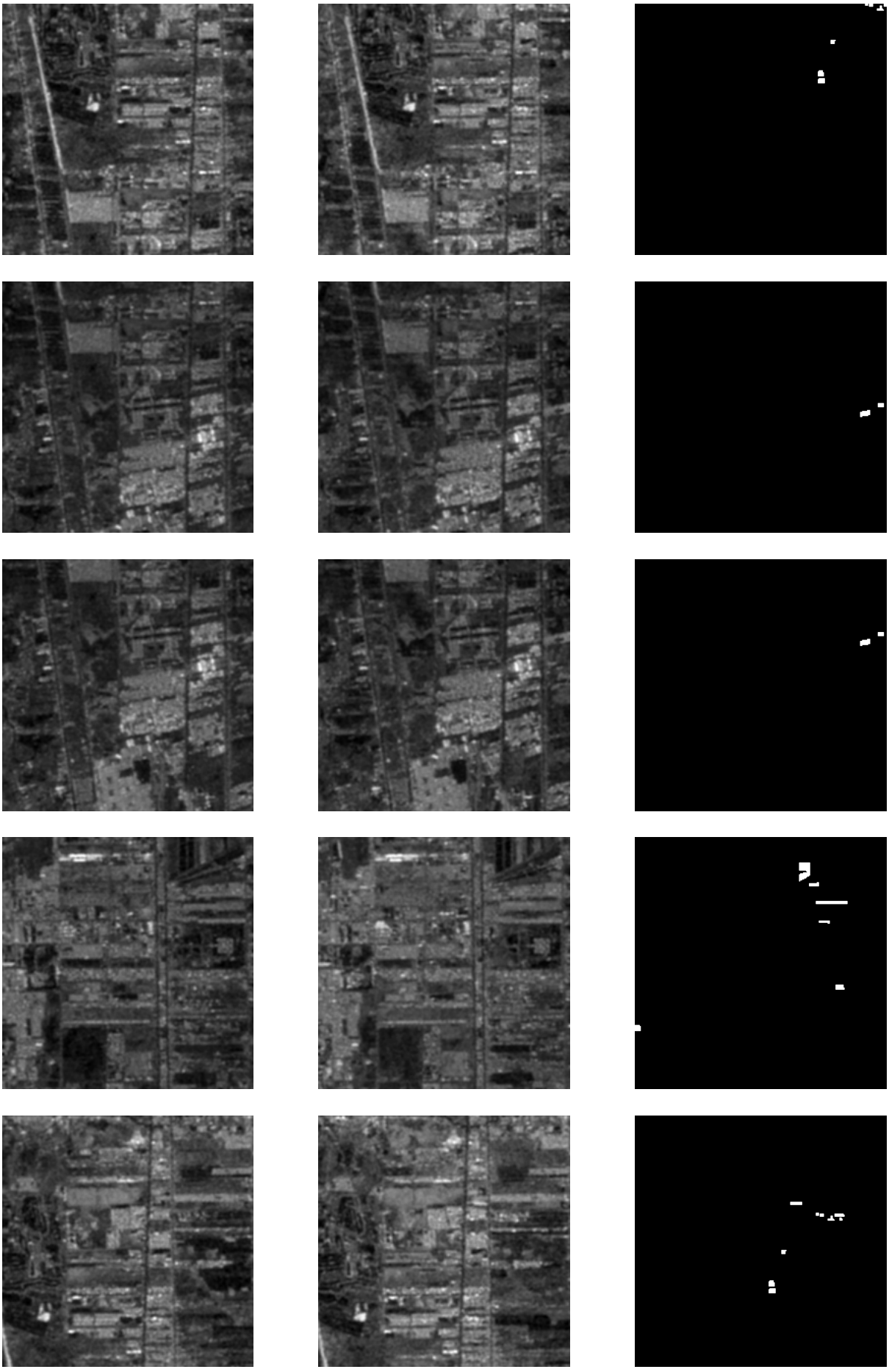


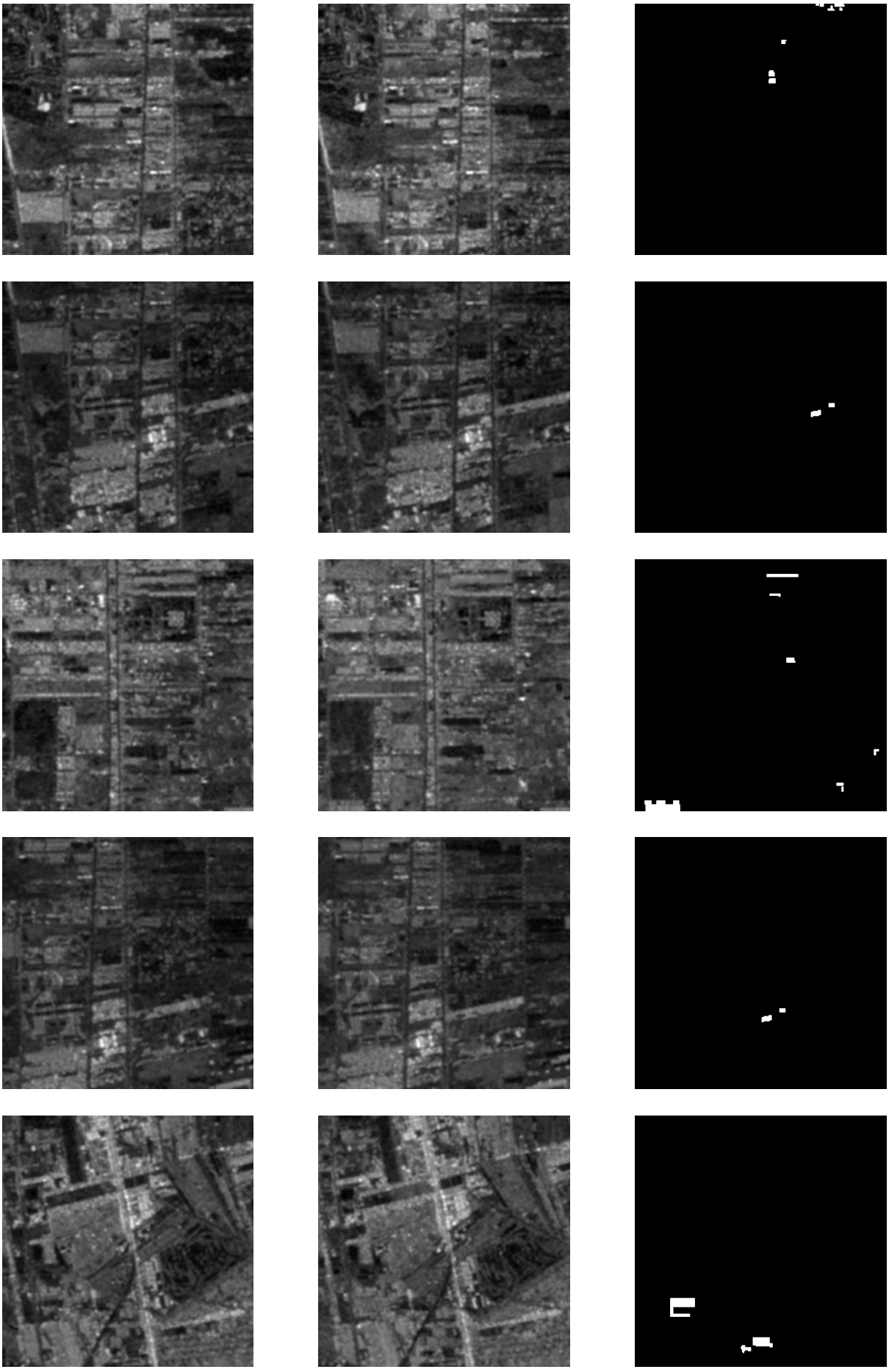


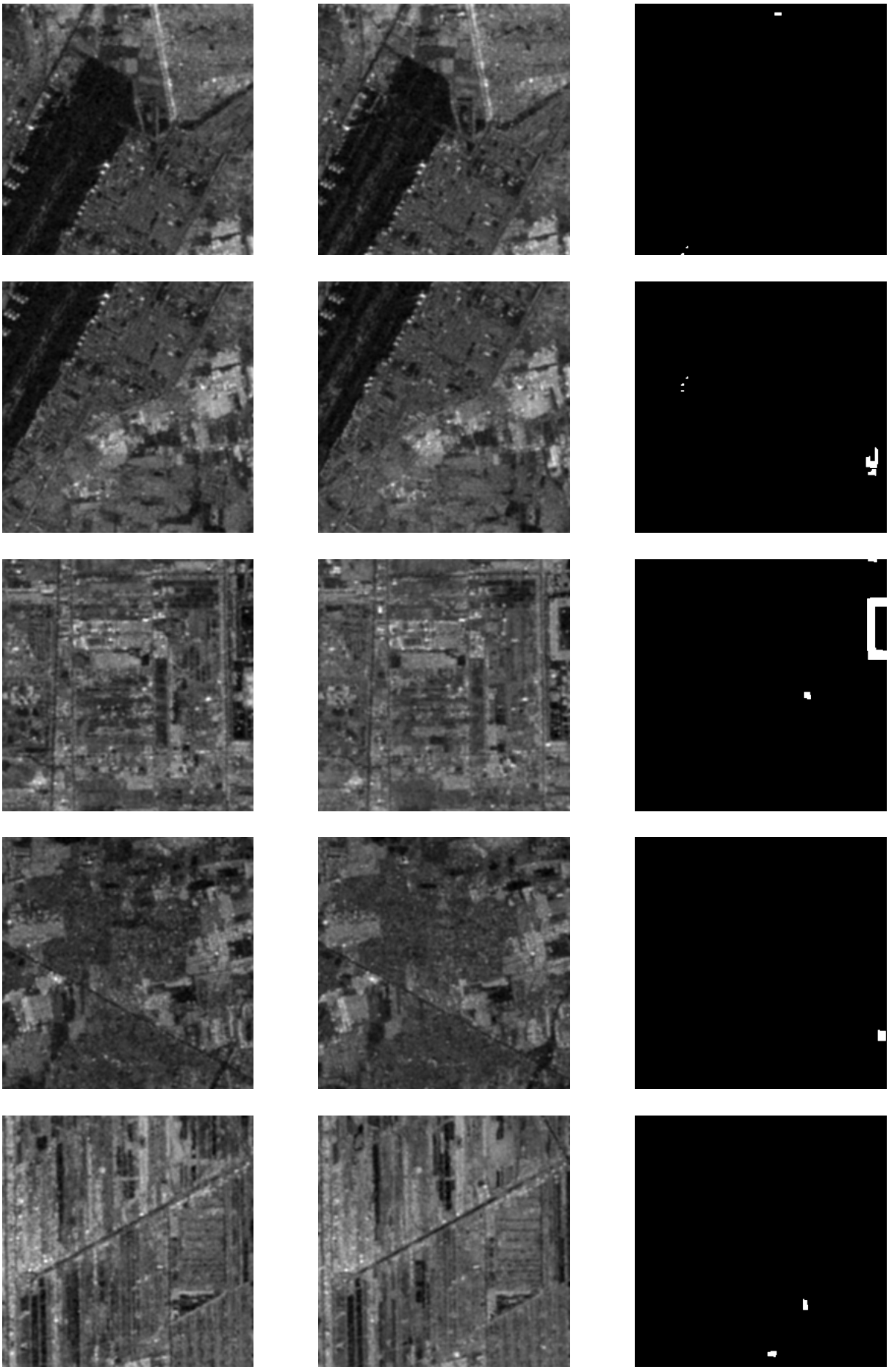


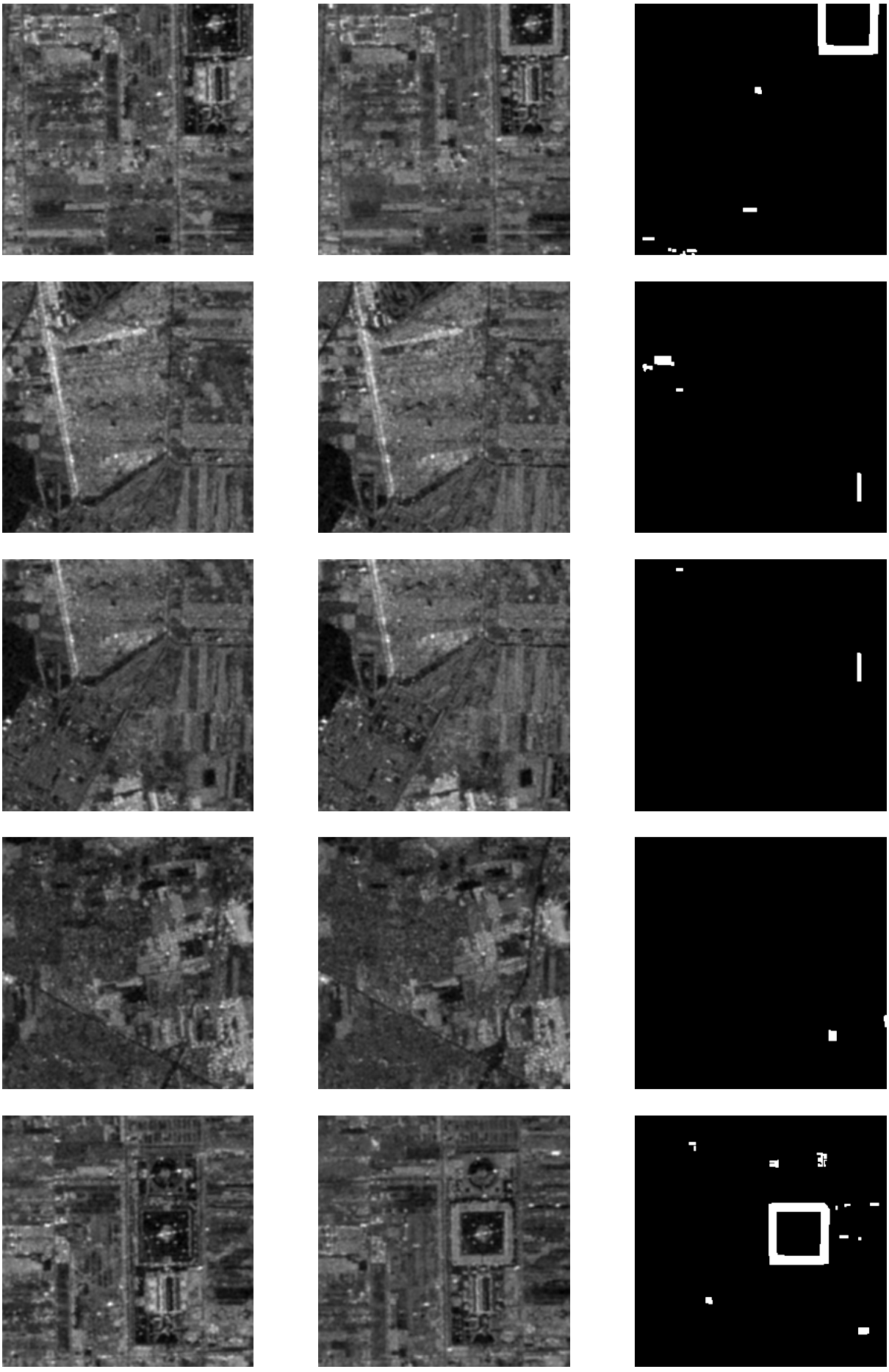


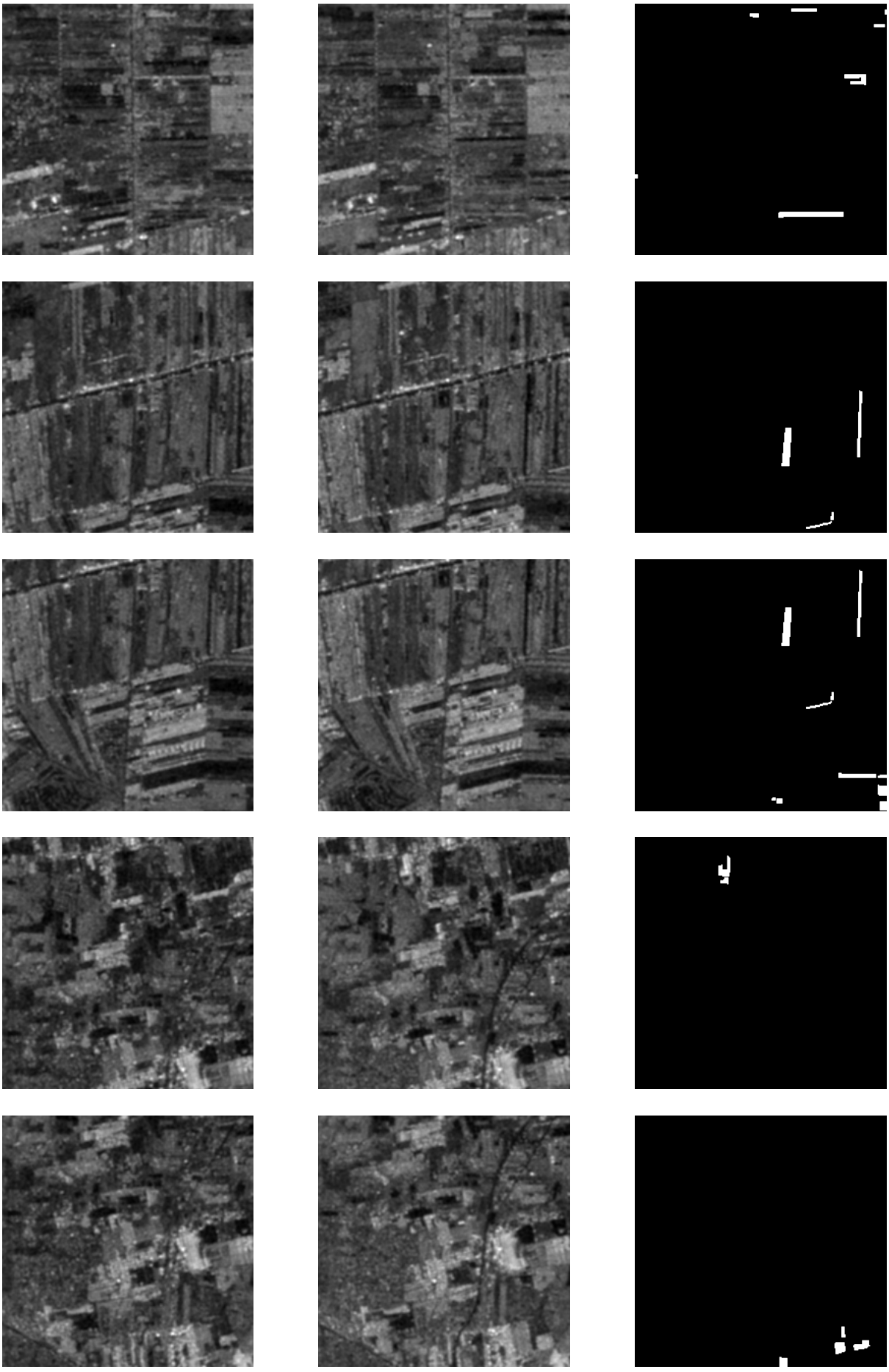


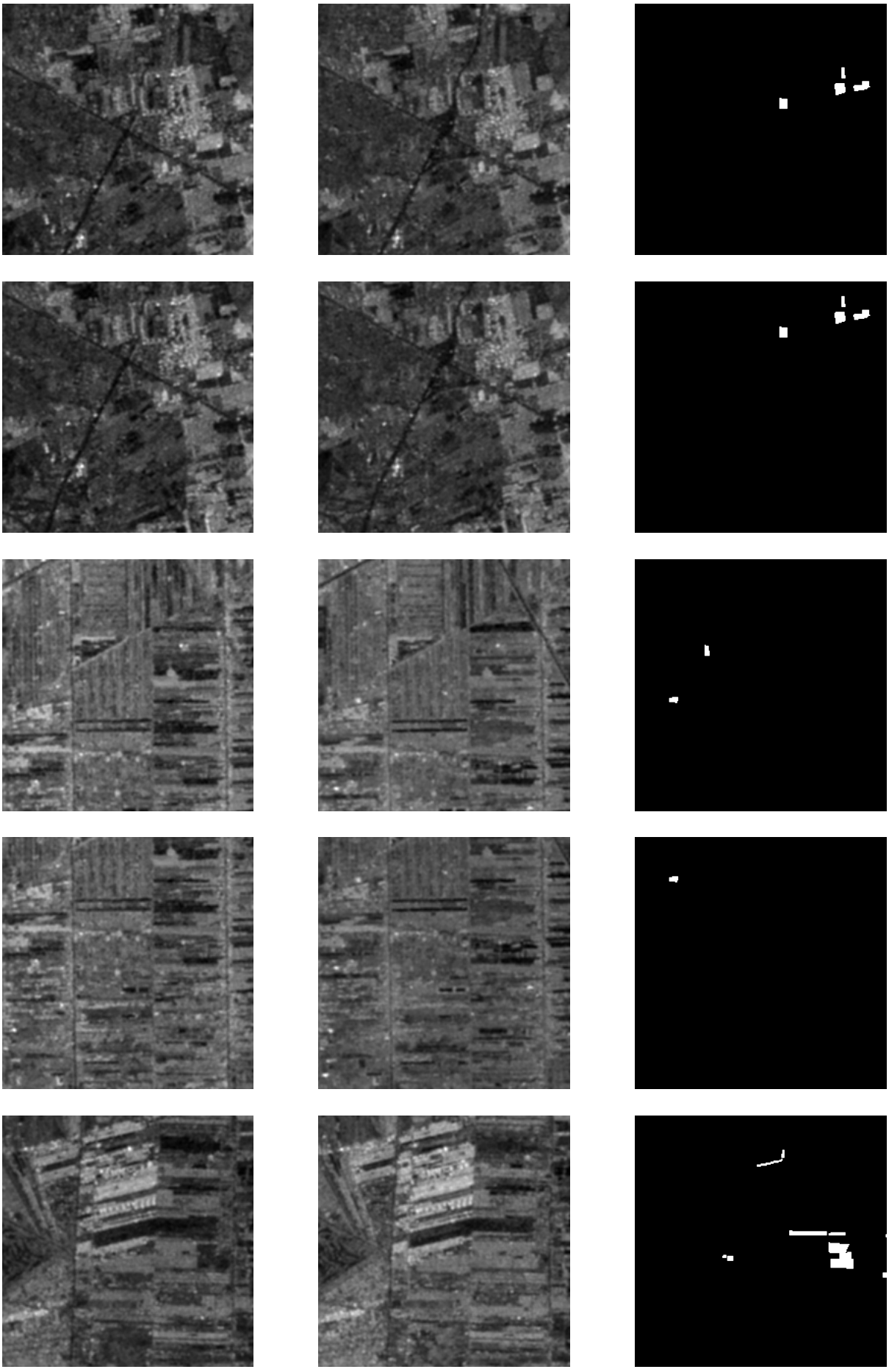


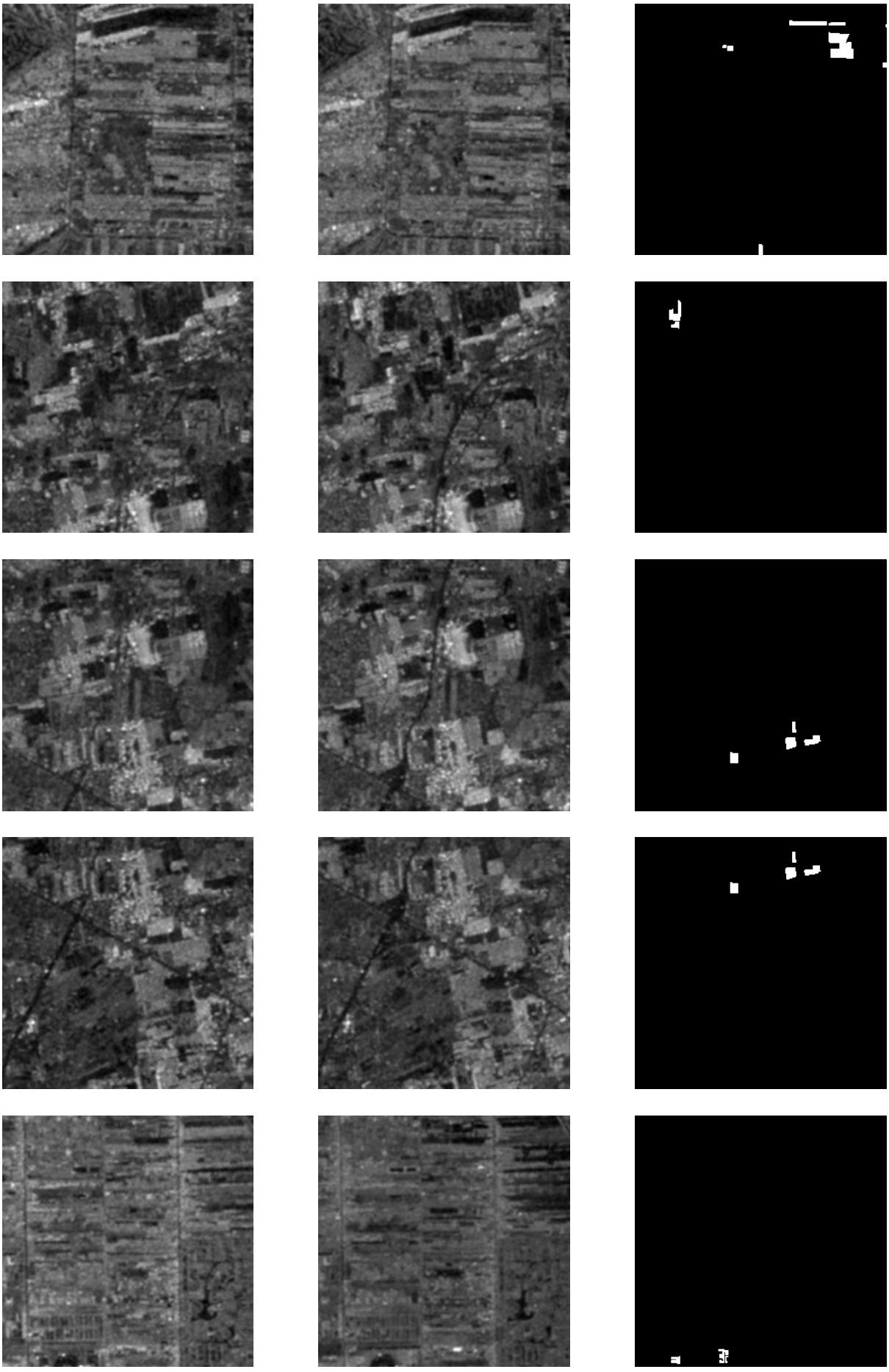


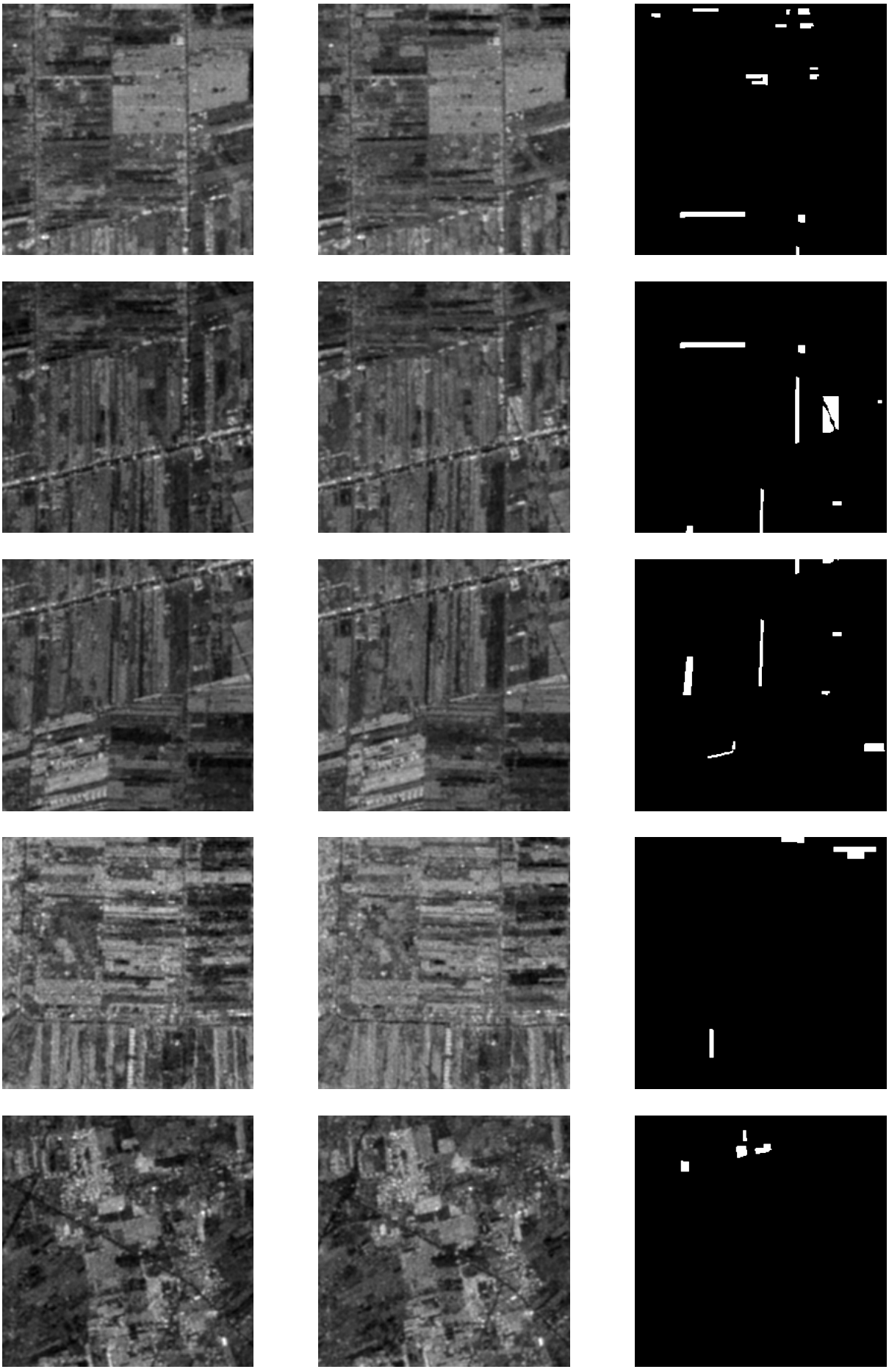


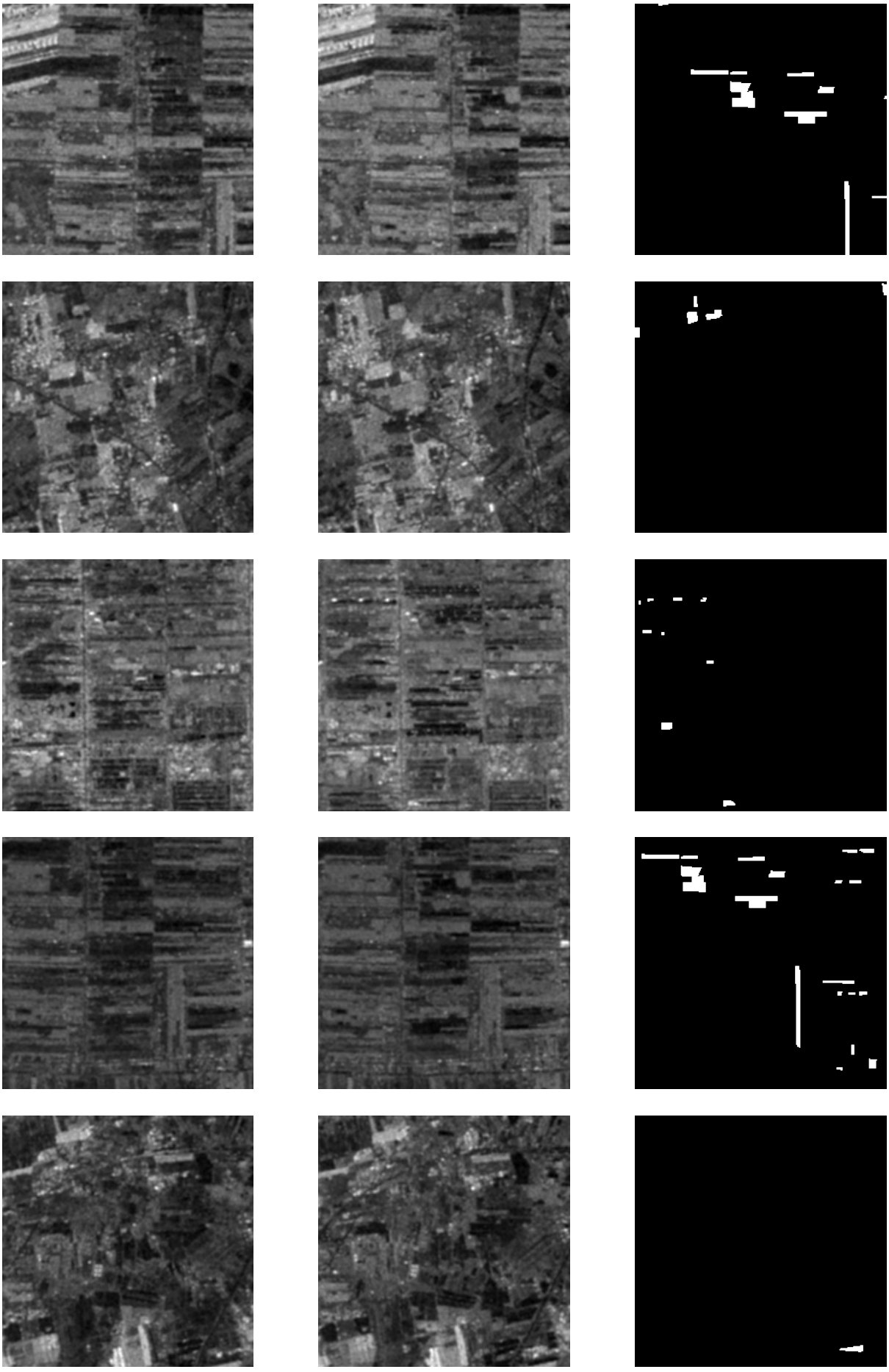


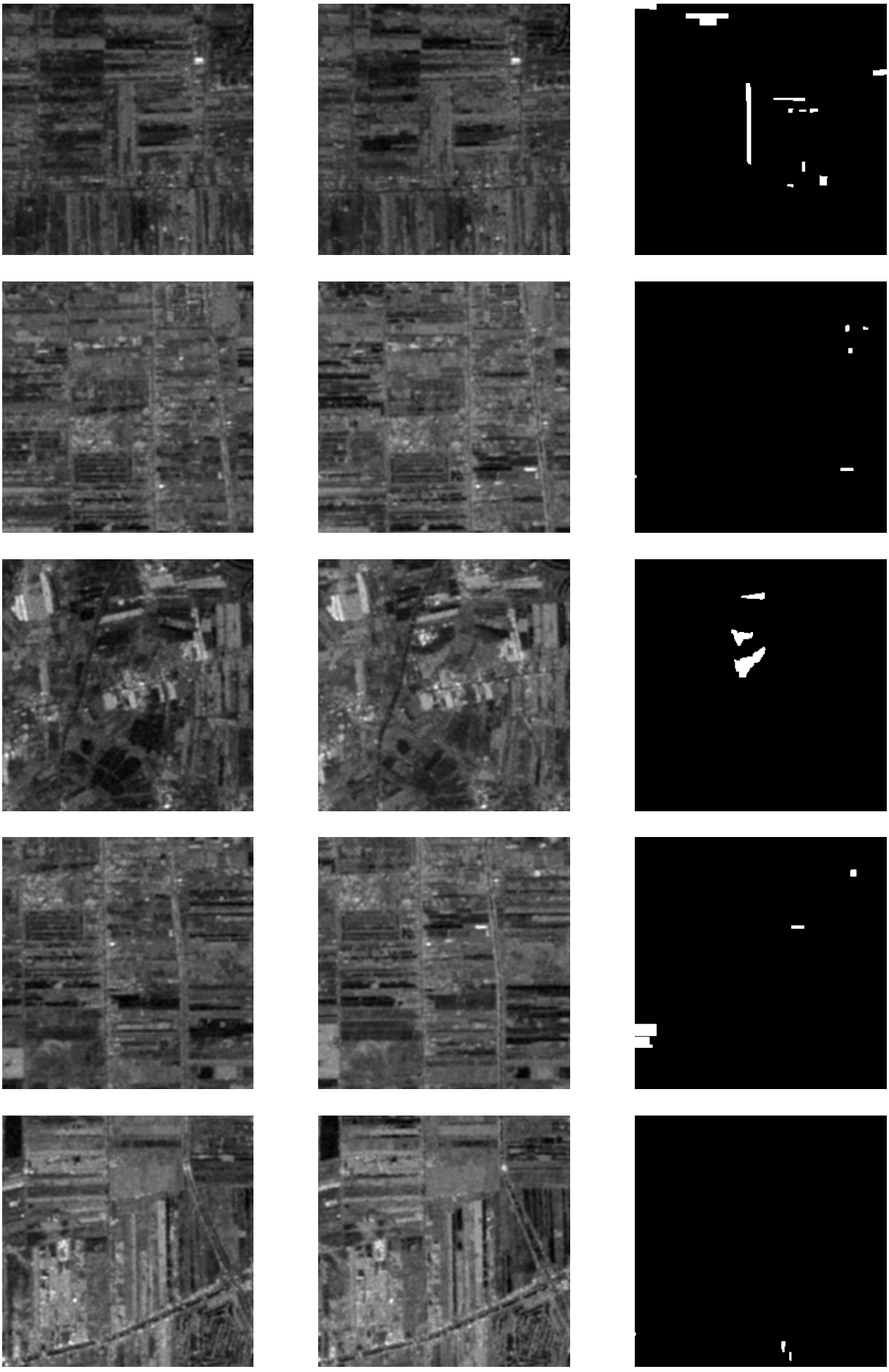


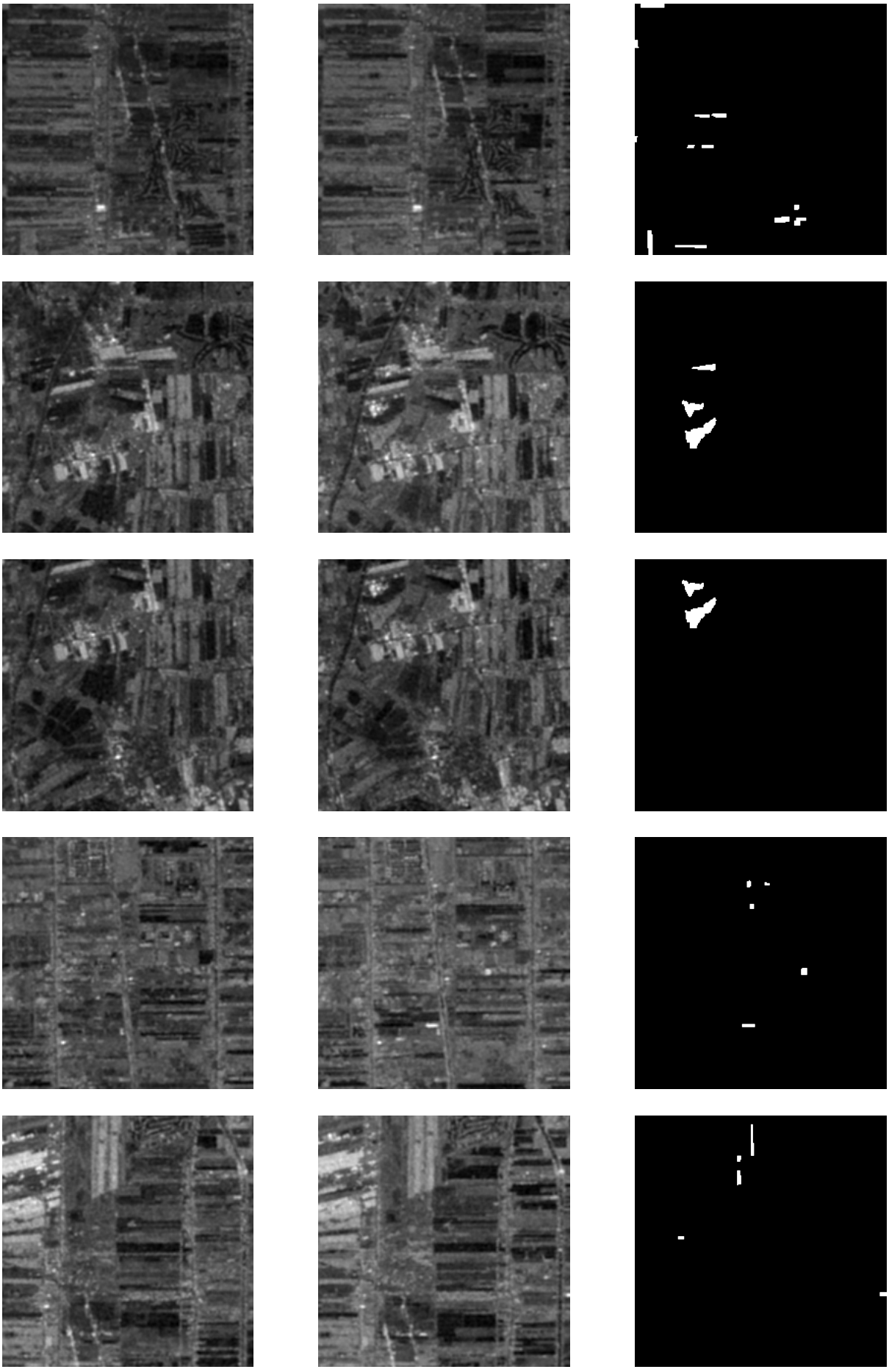


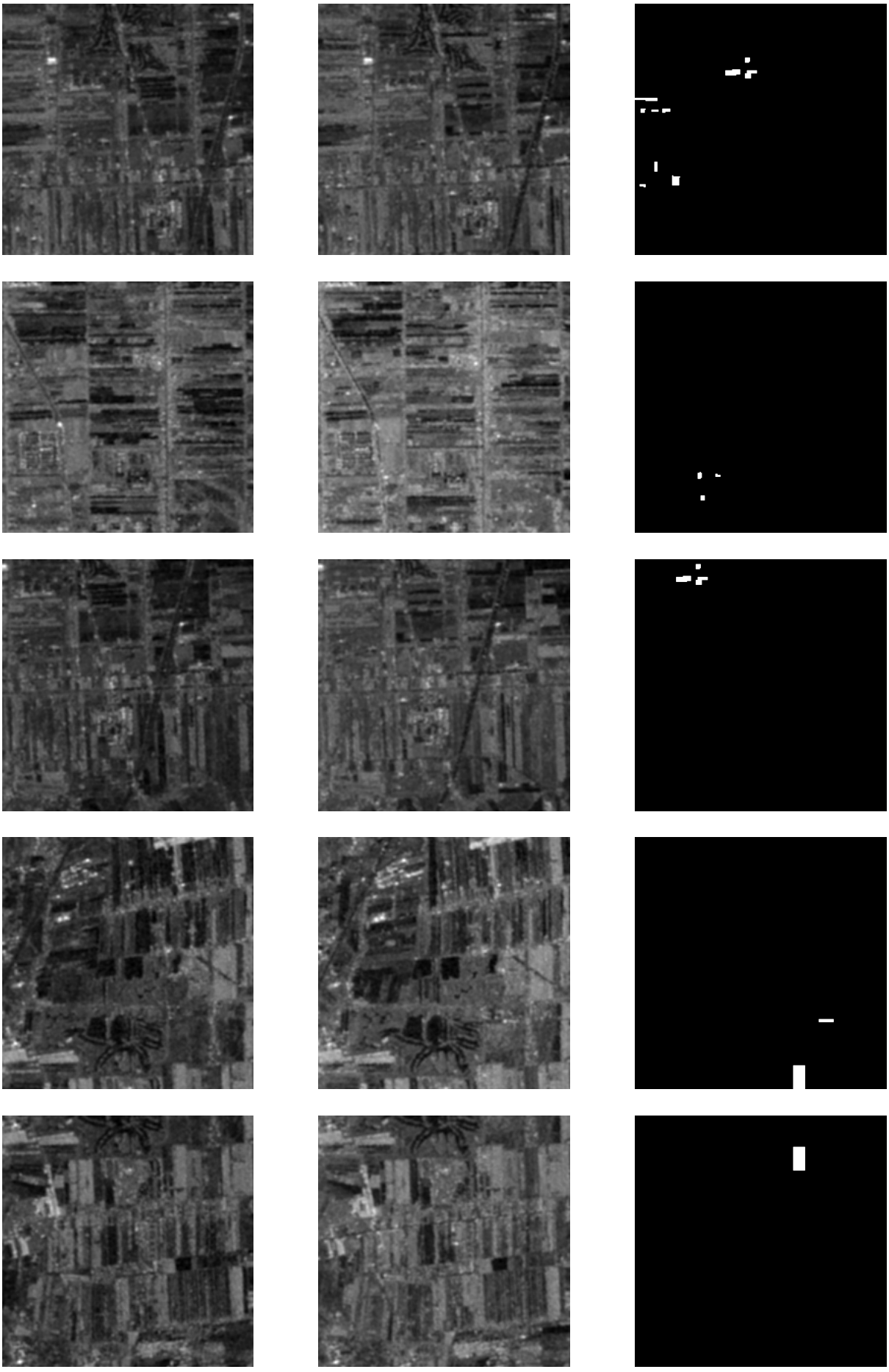


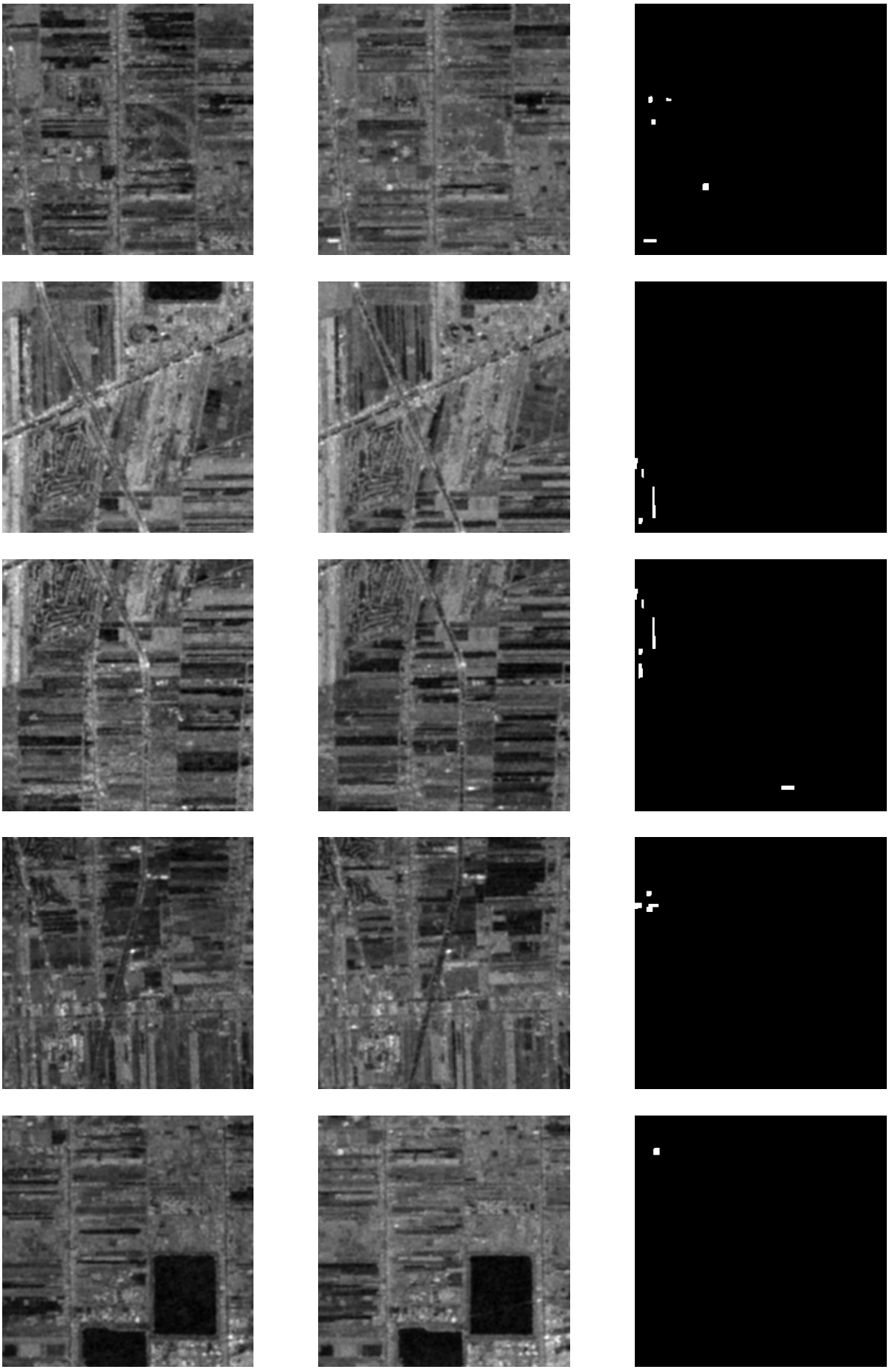


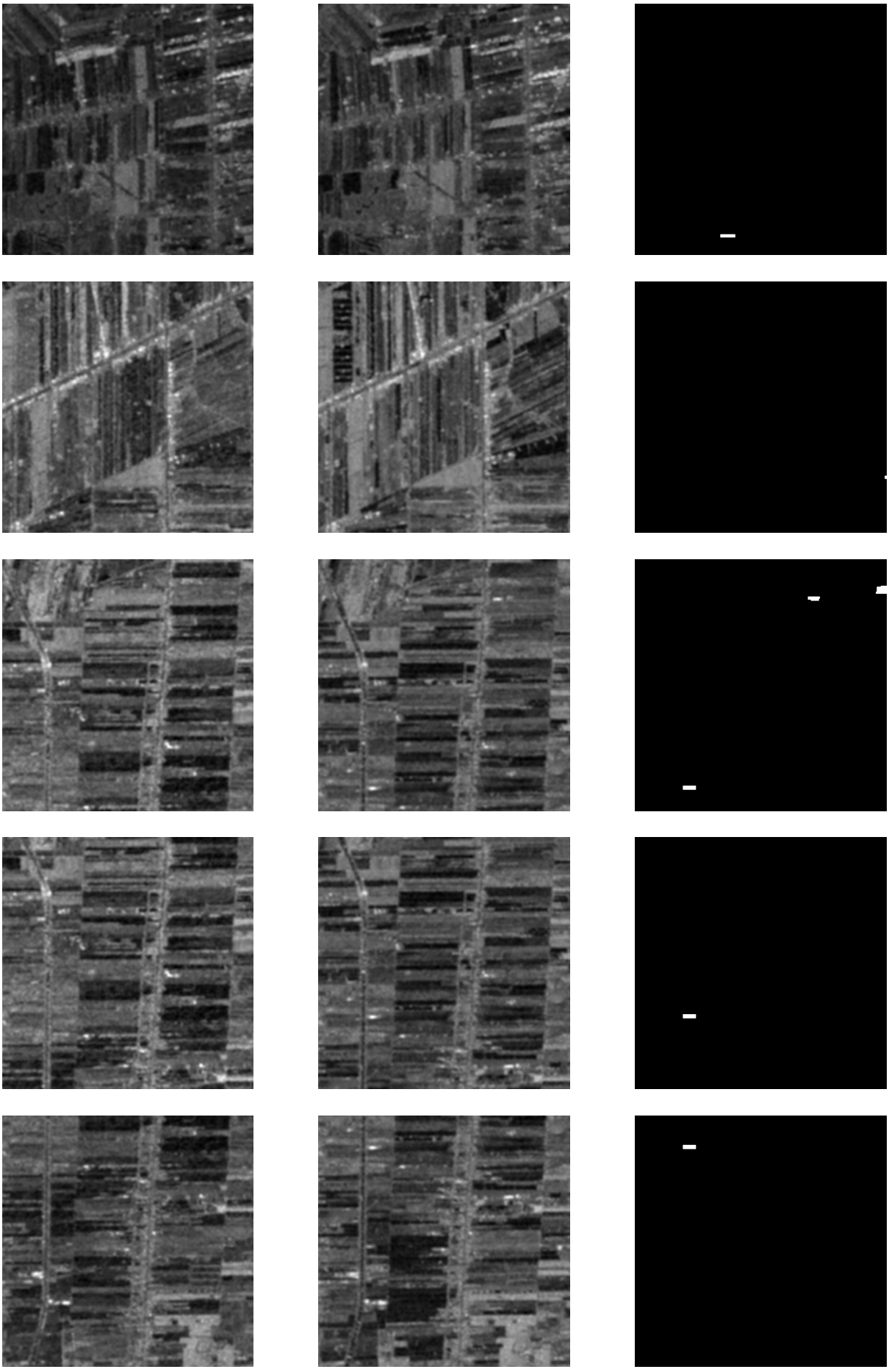


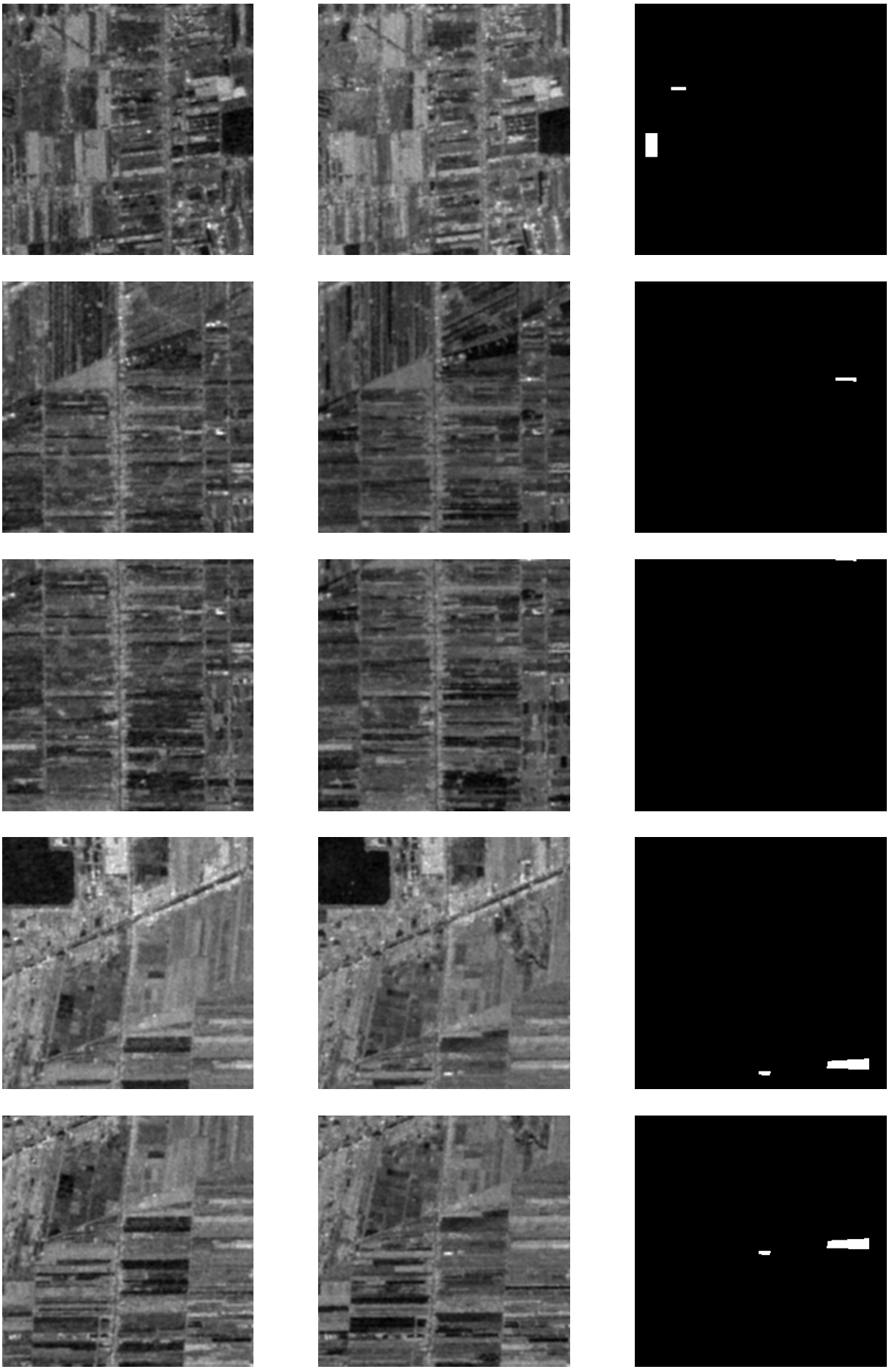


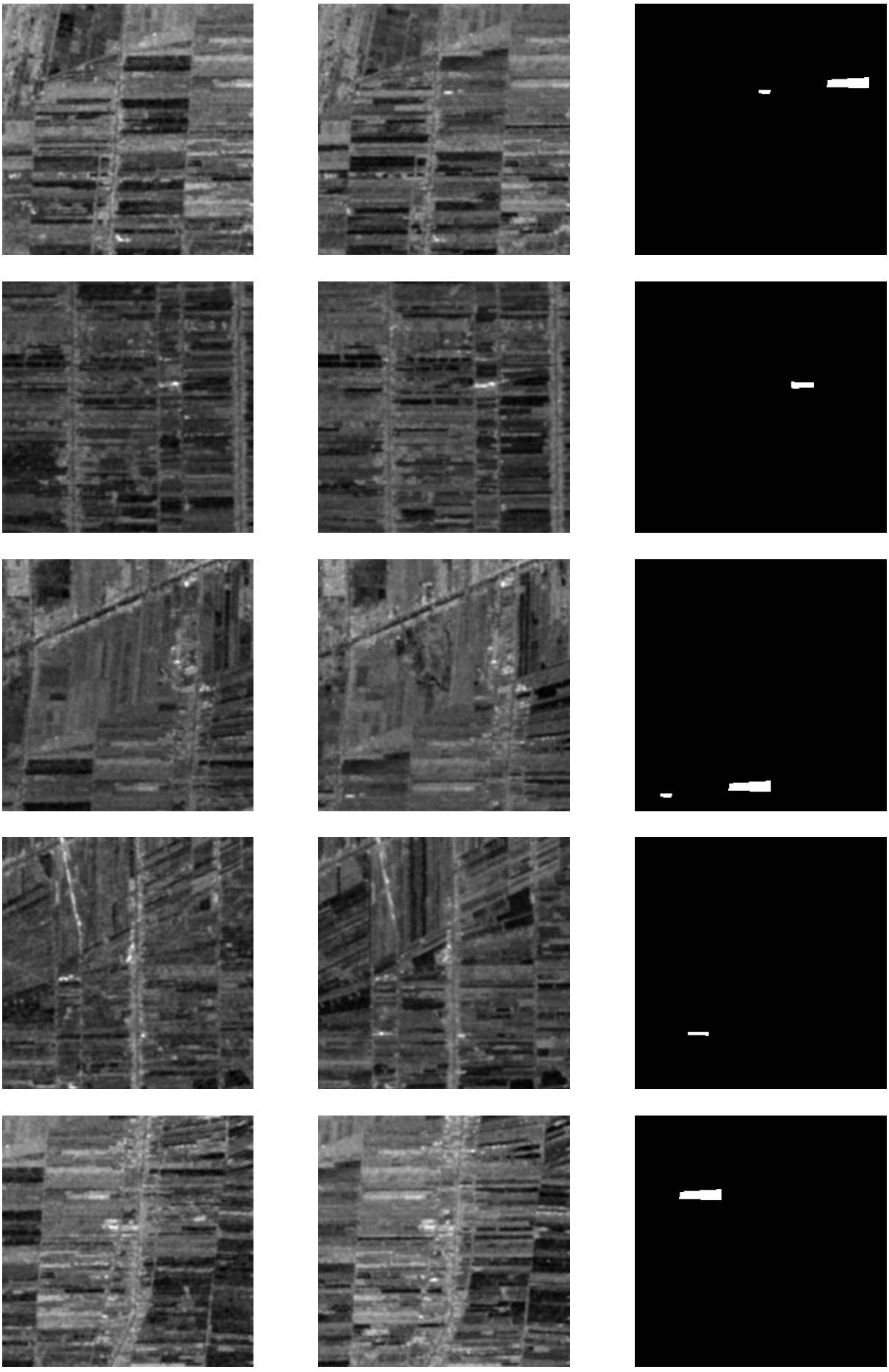


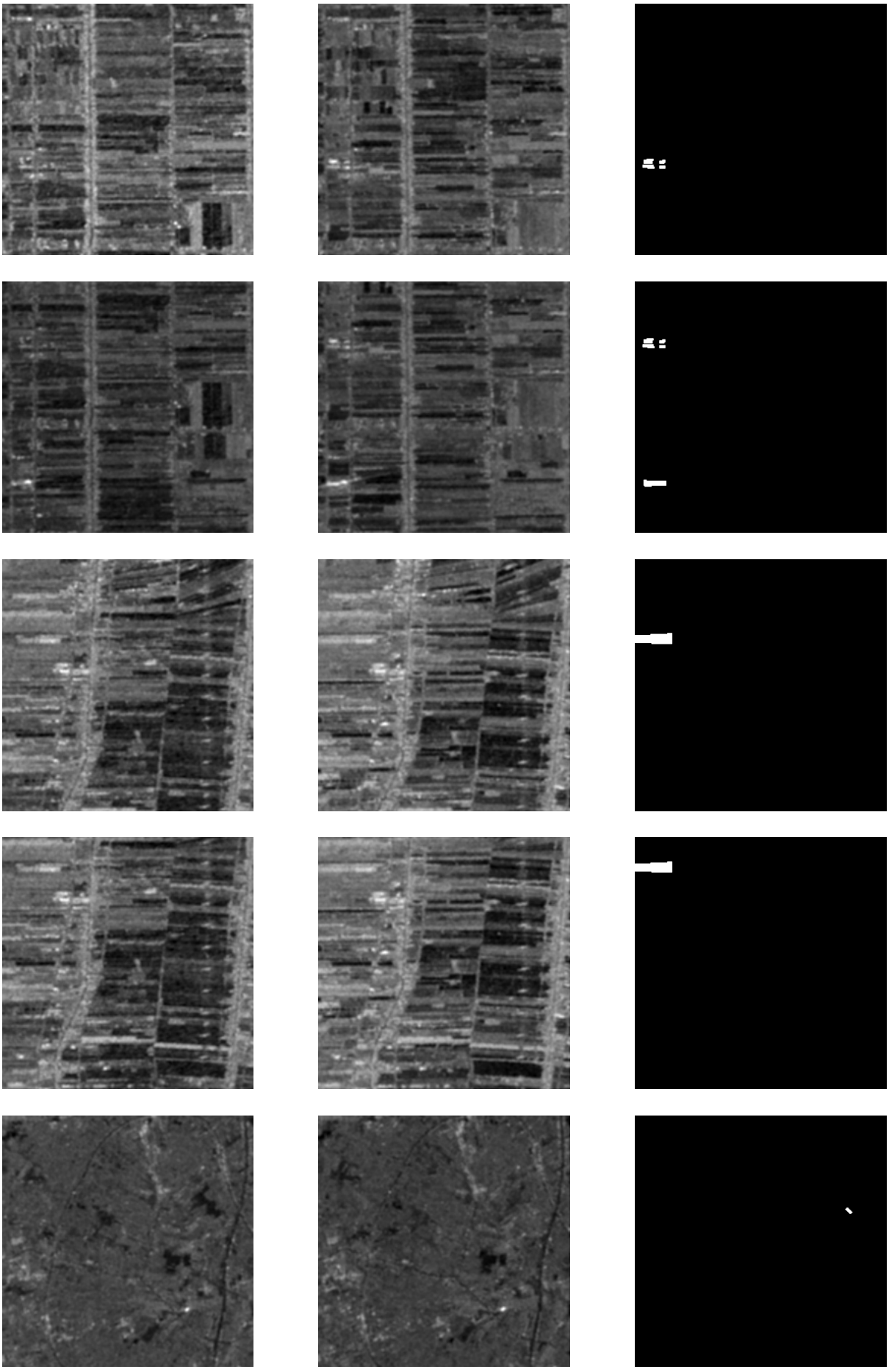


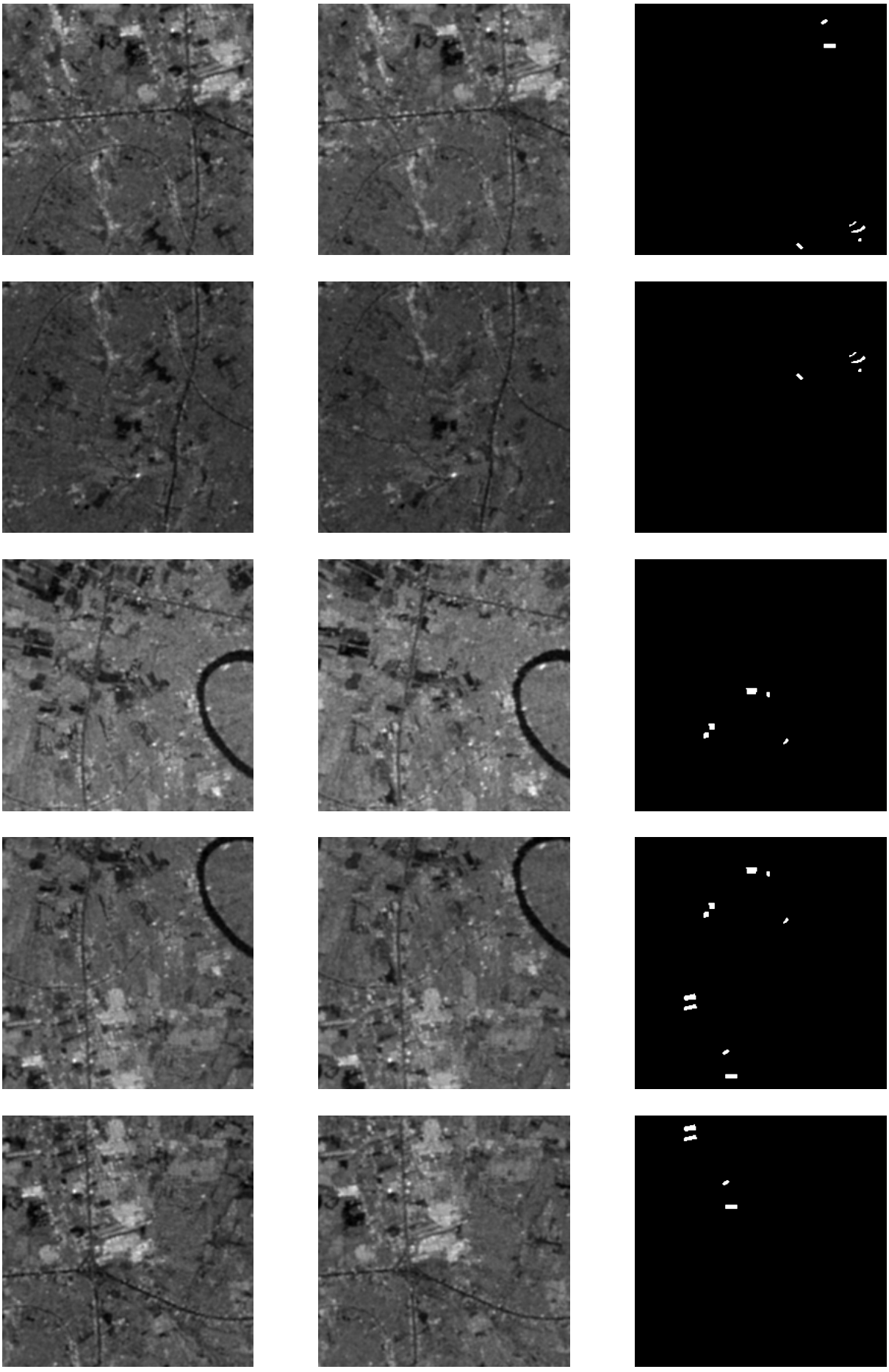


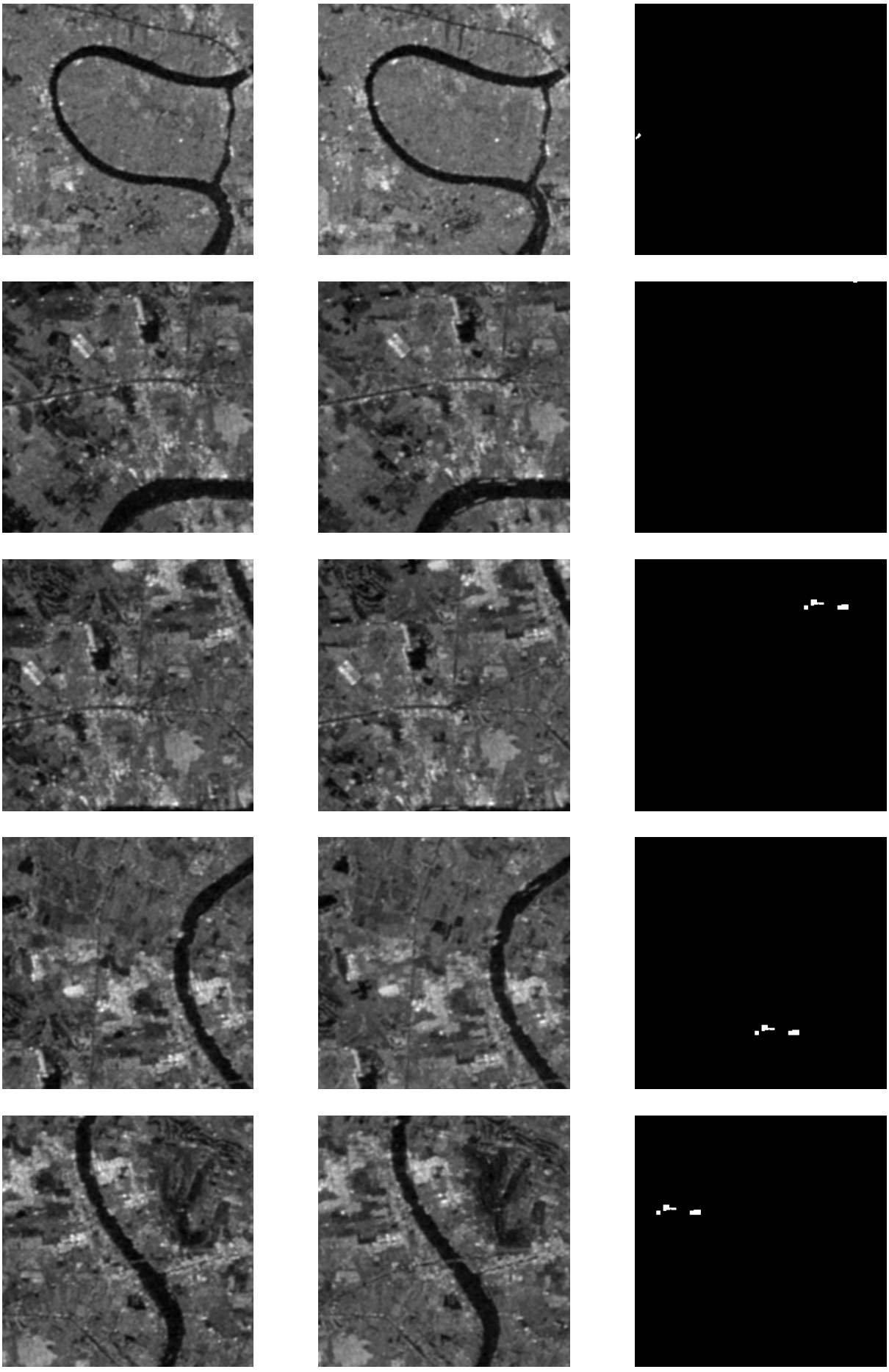


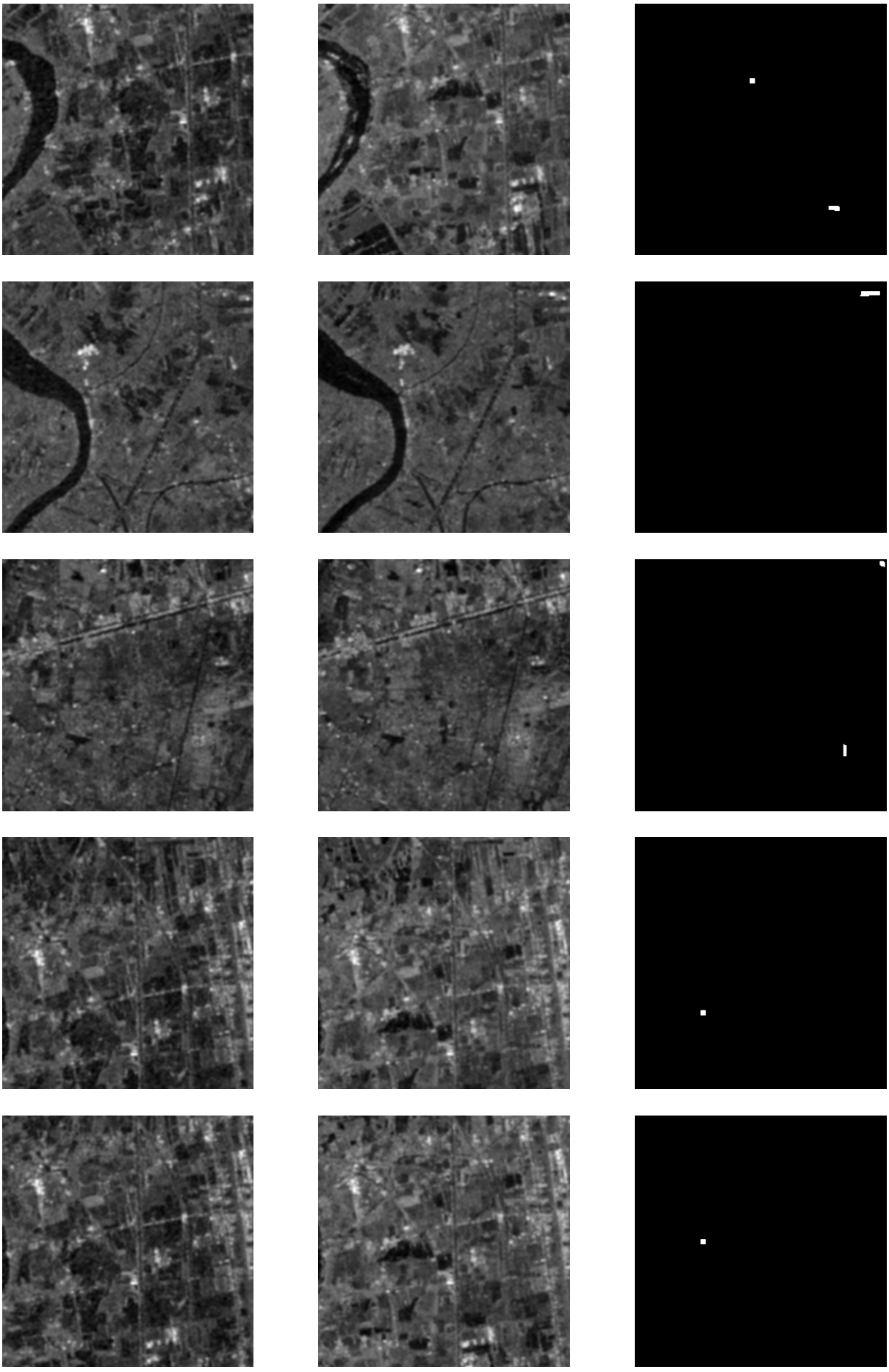


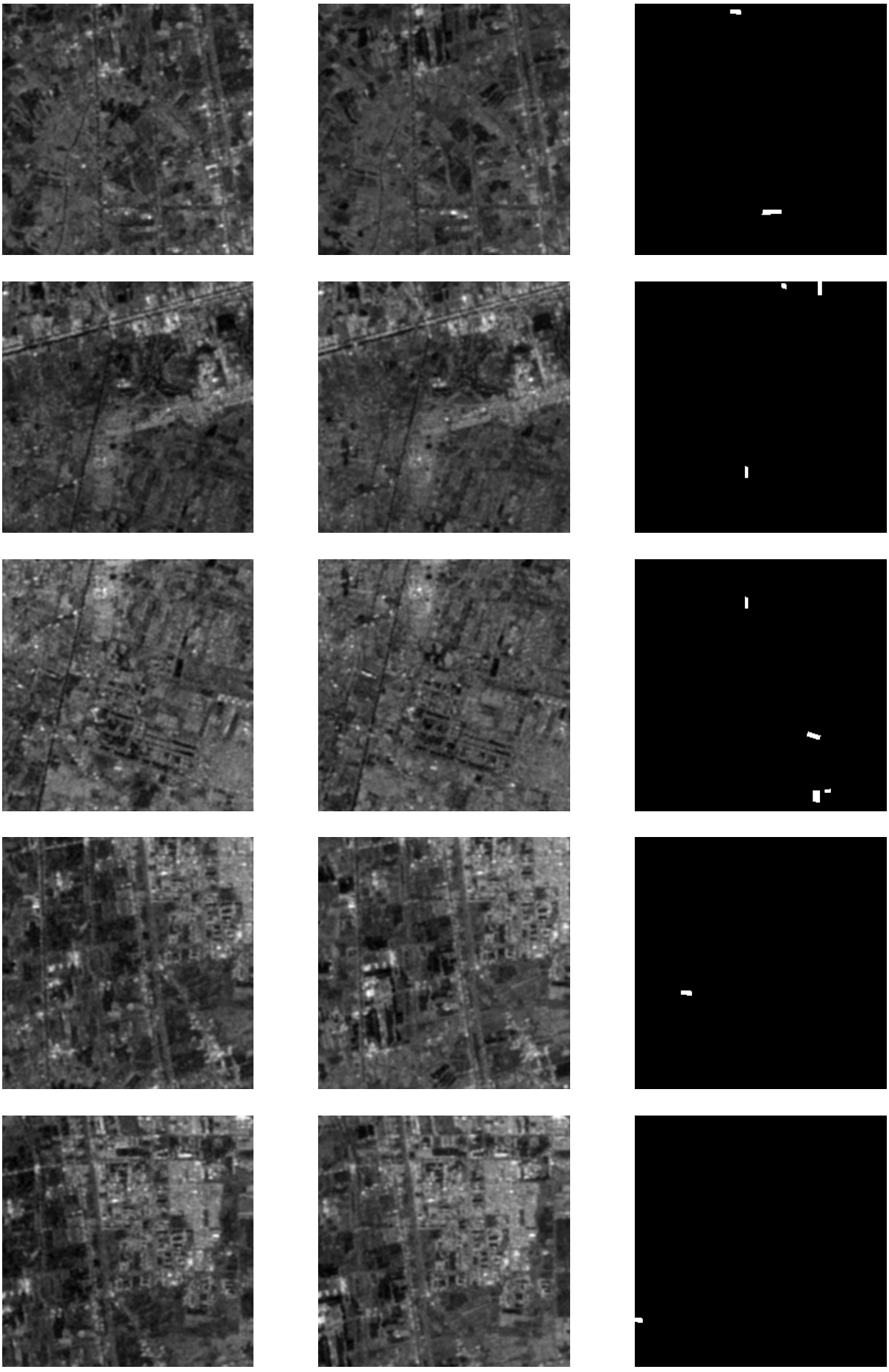


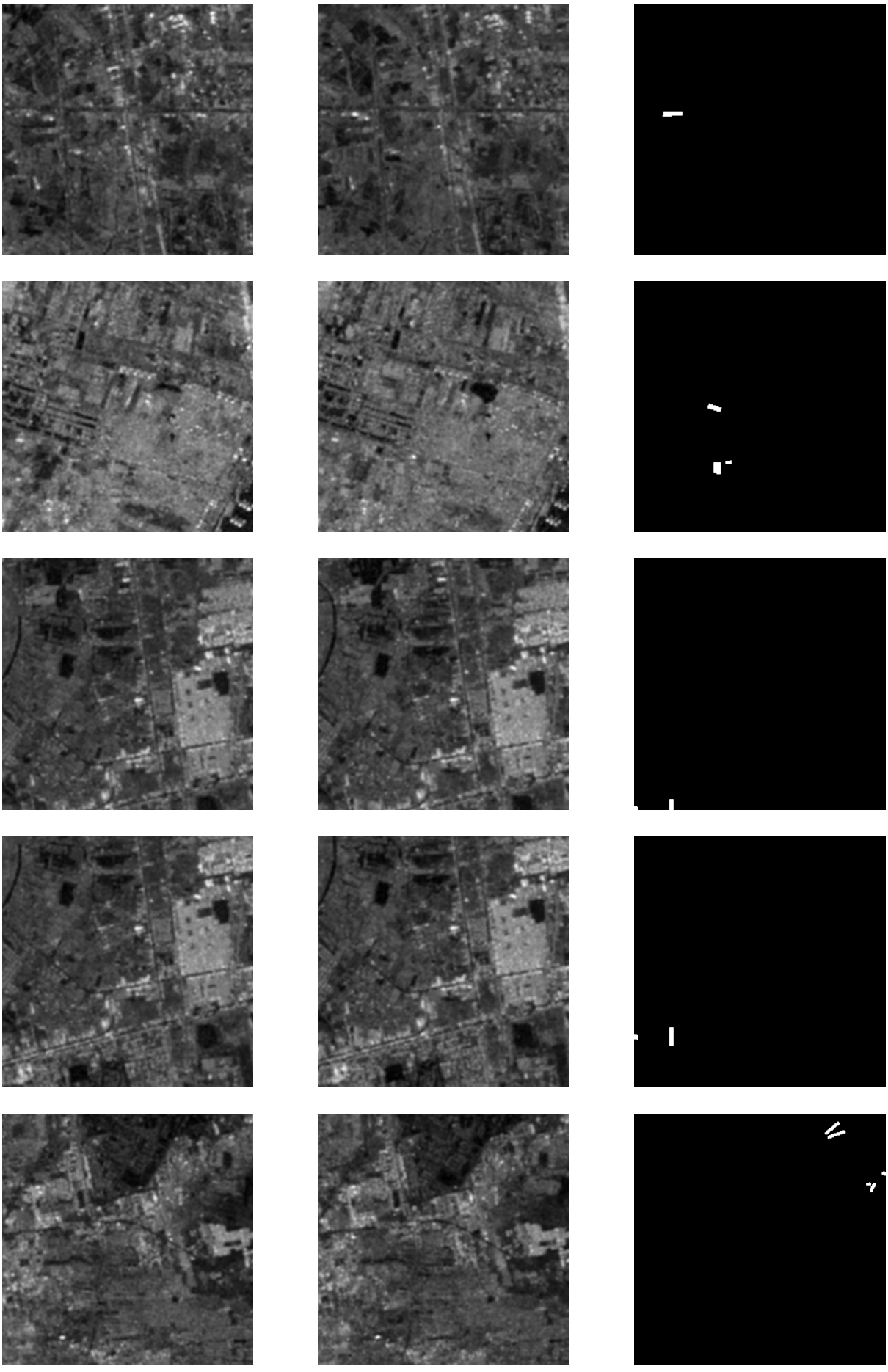


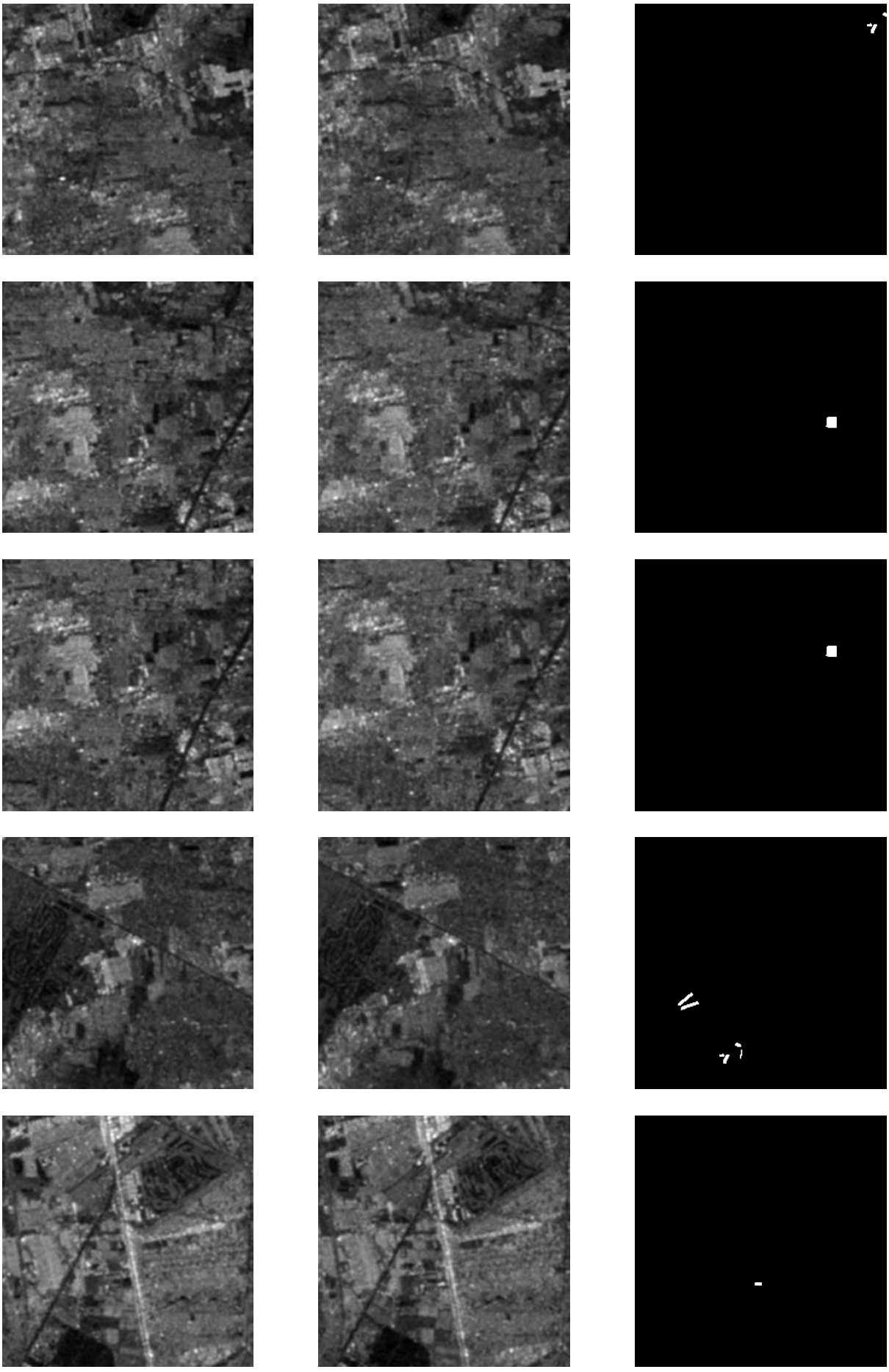


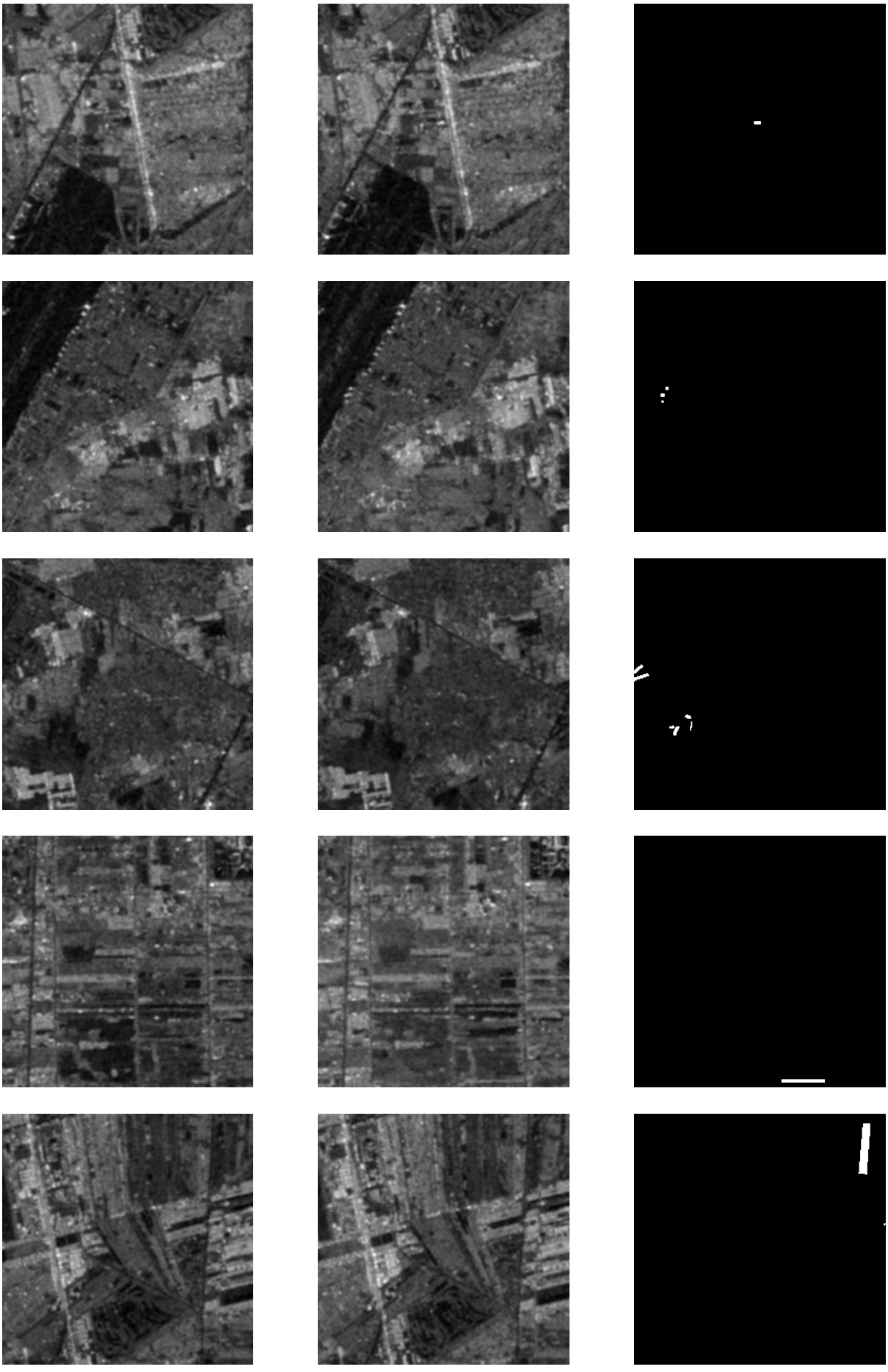


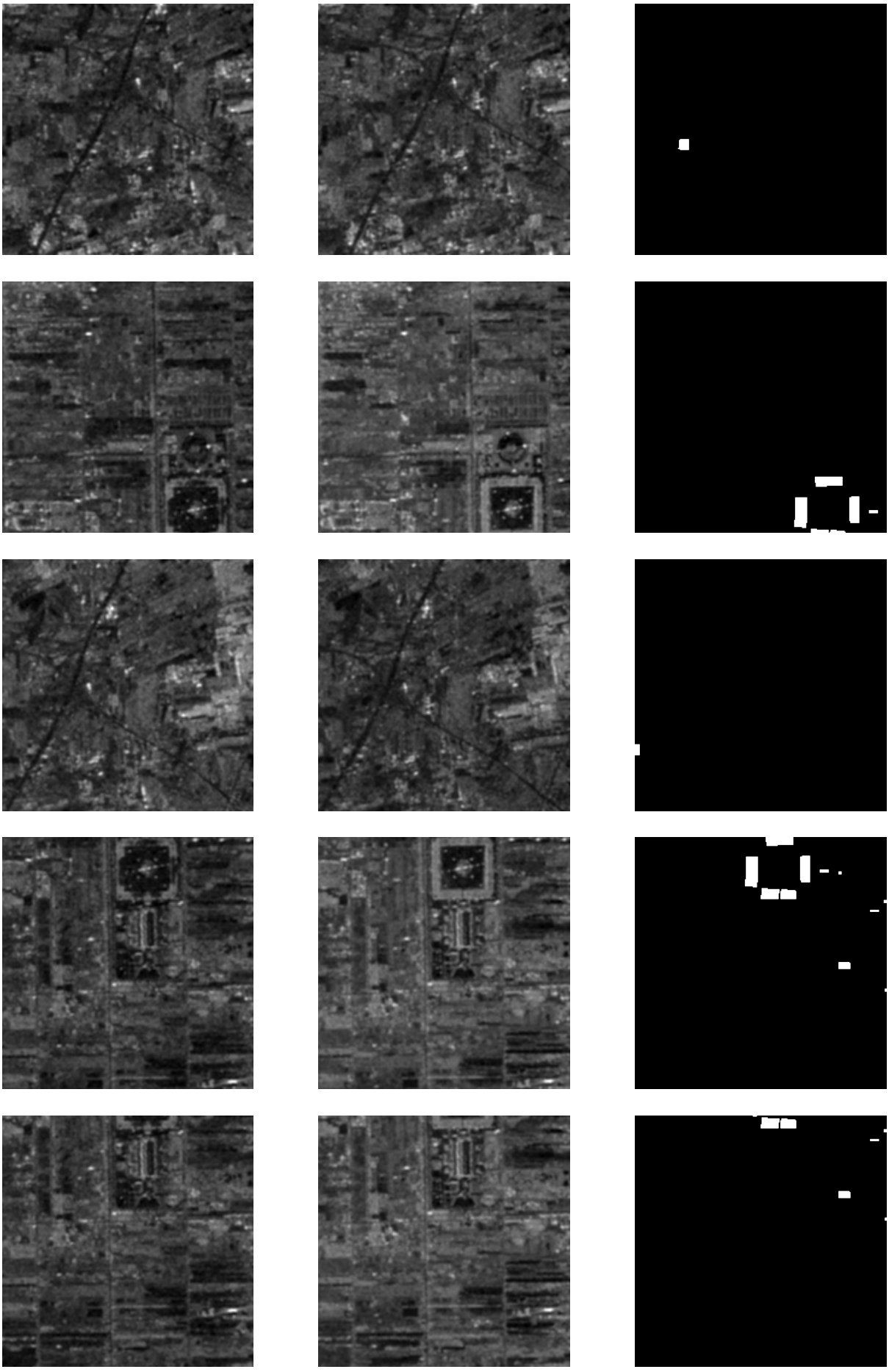


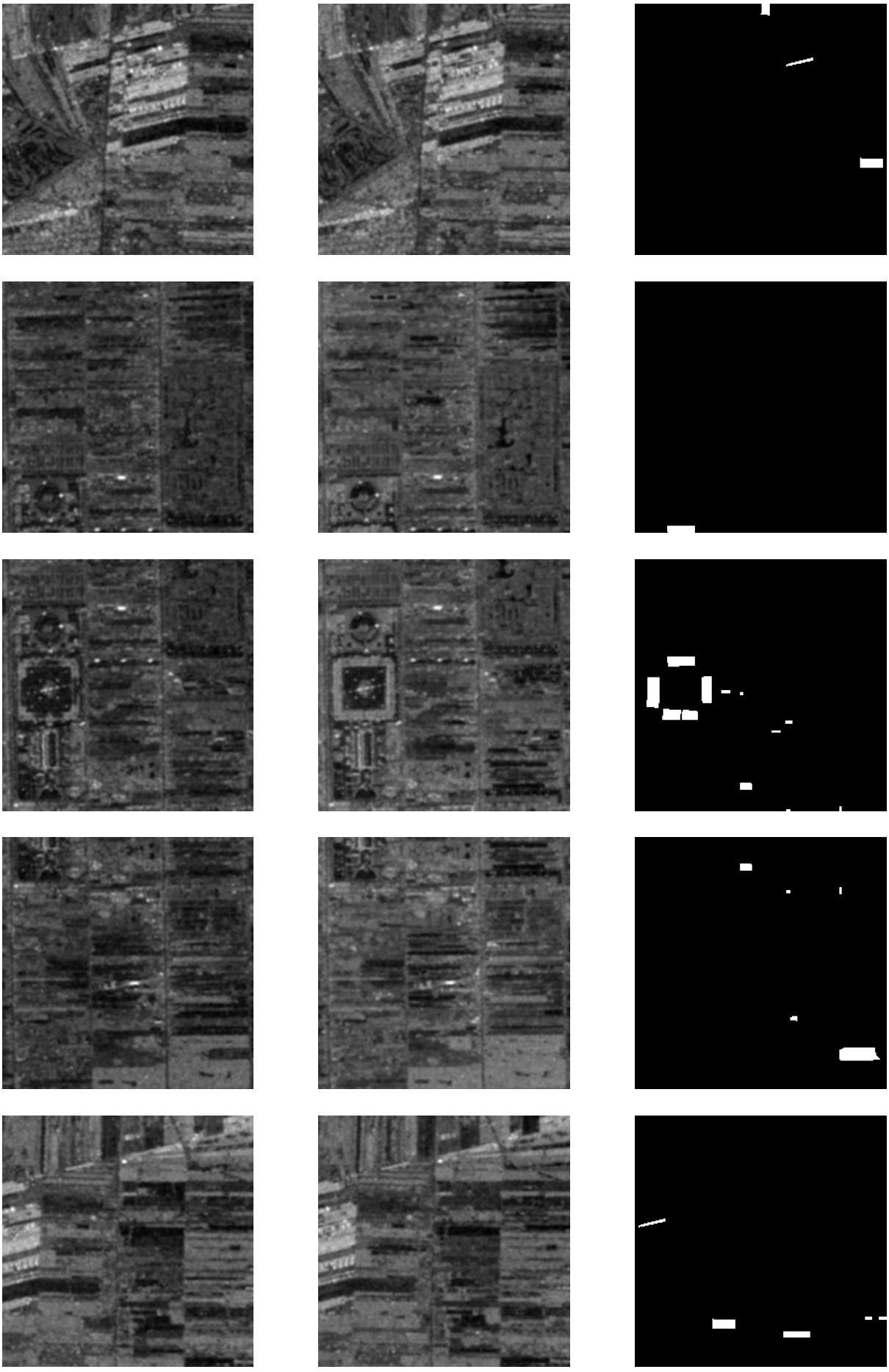


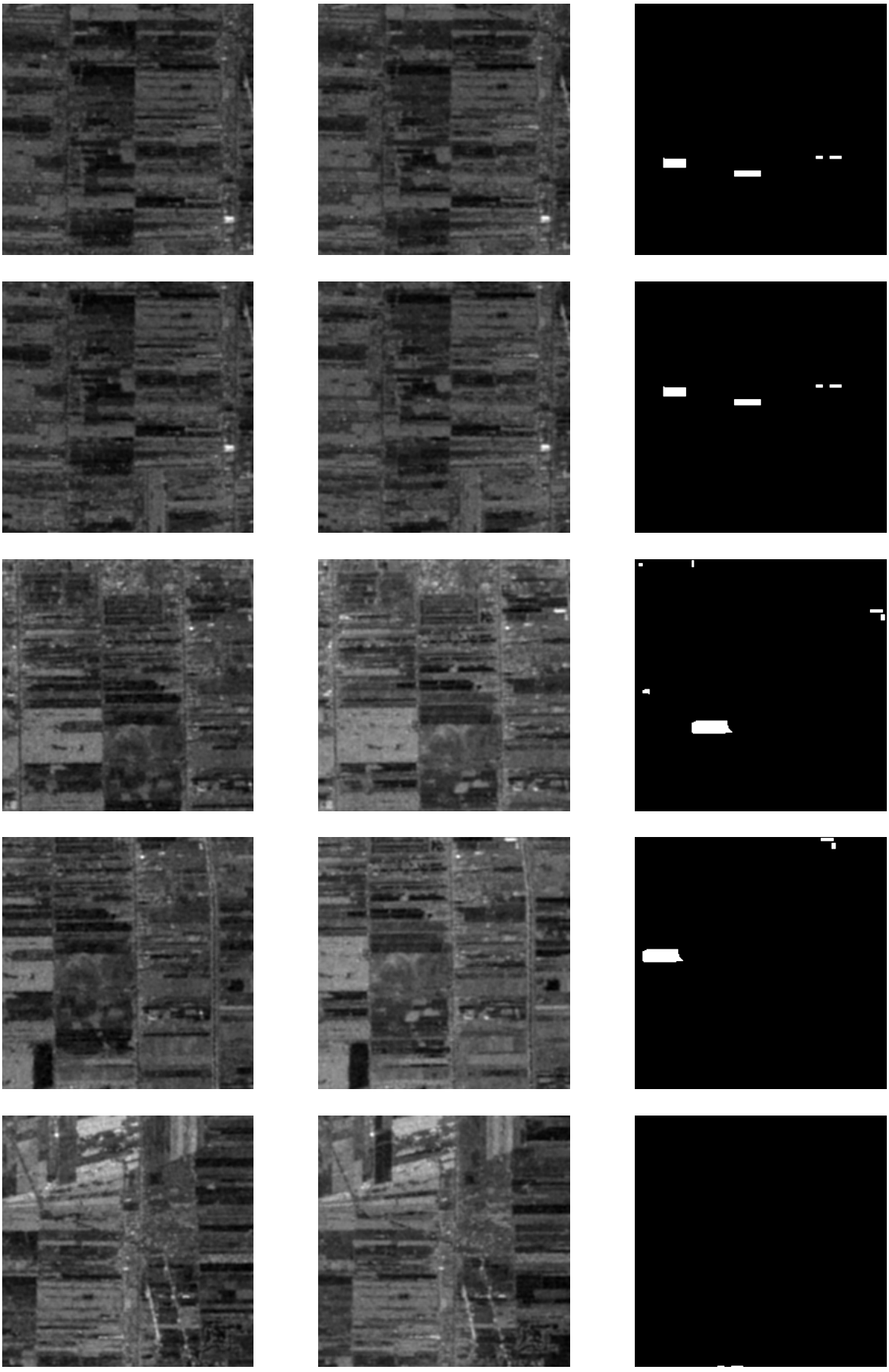


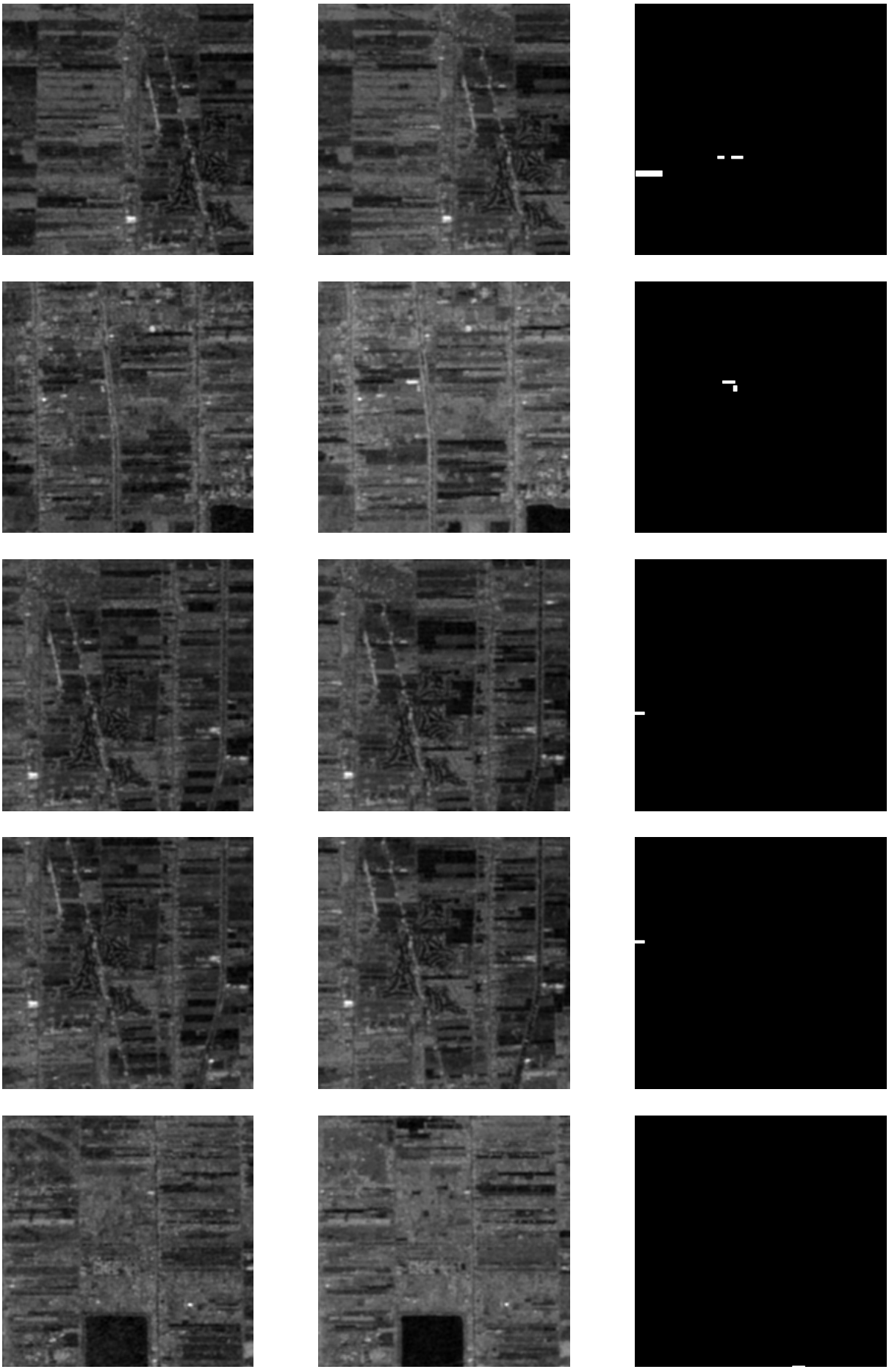


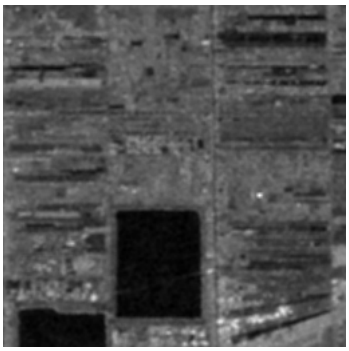
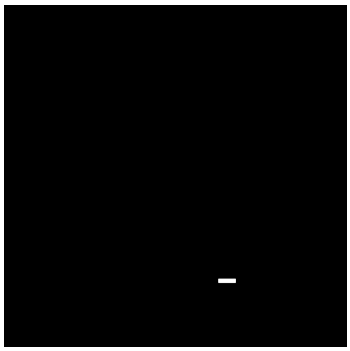


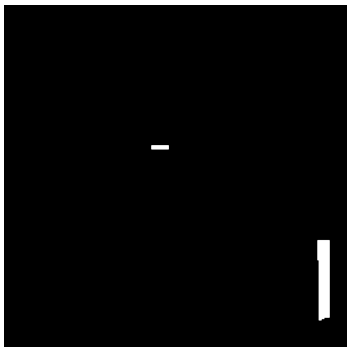








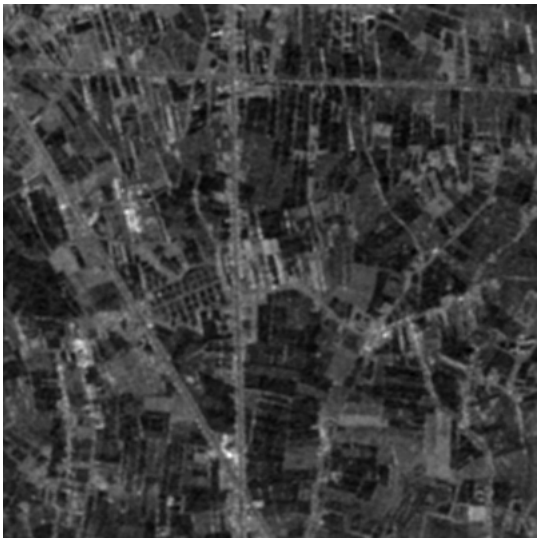
Appendix-B

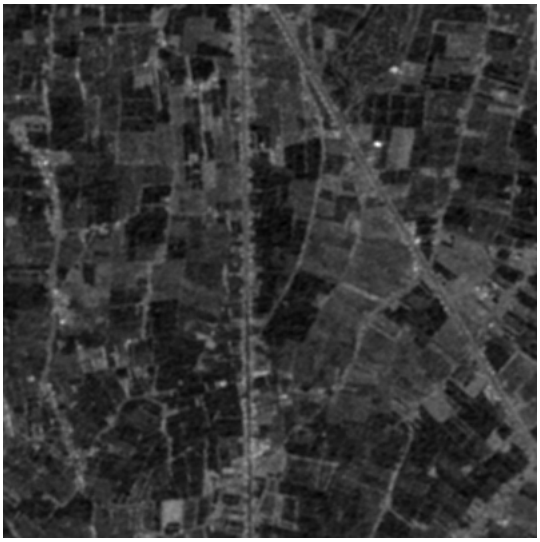
Testing area

This section contains all of the test areas used in this thesis.

Bangkok Area

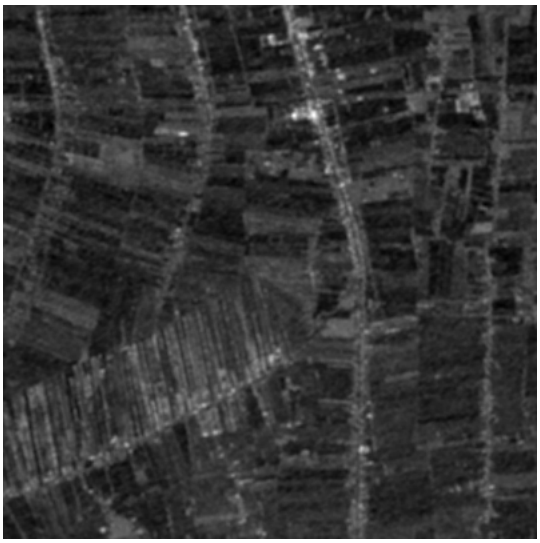
Time 1

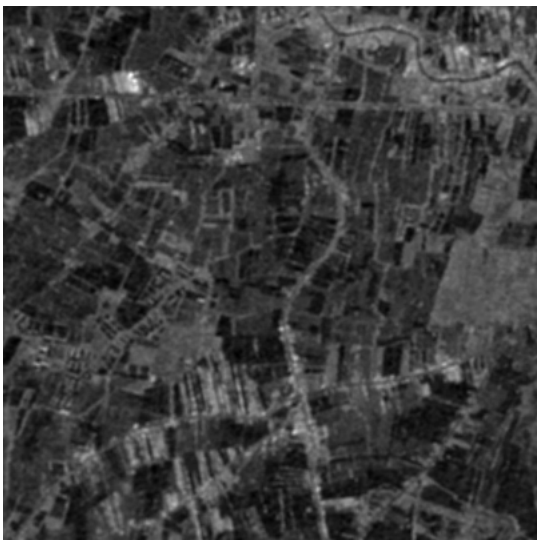
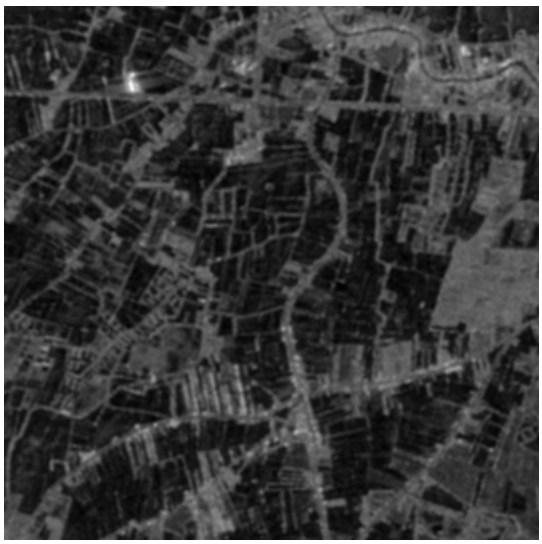


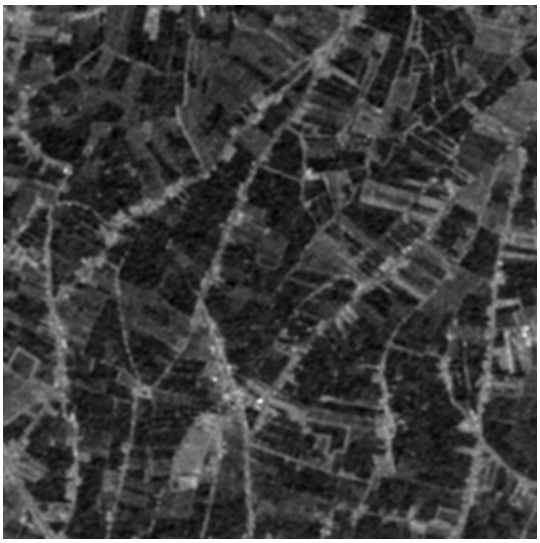


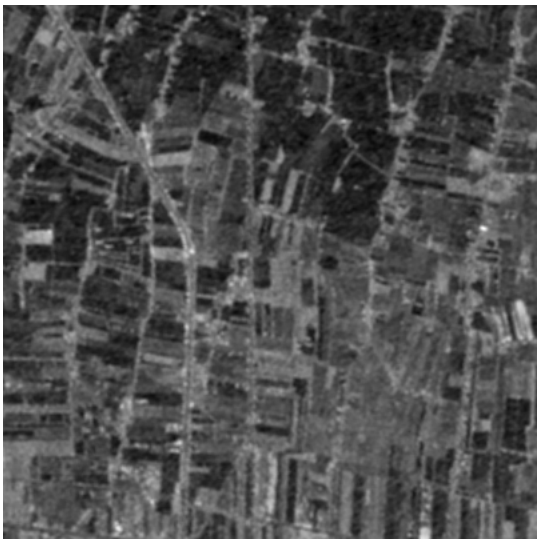
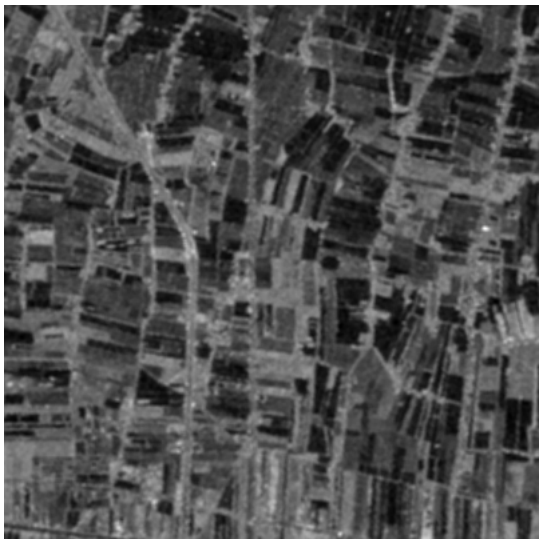
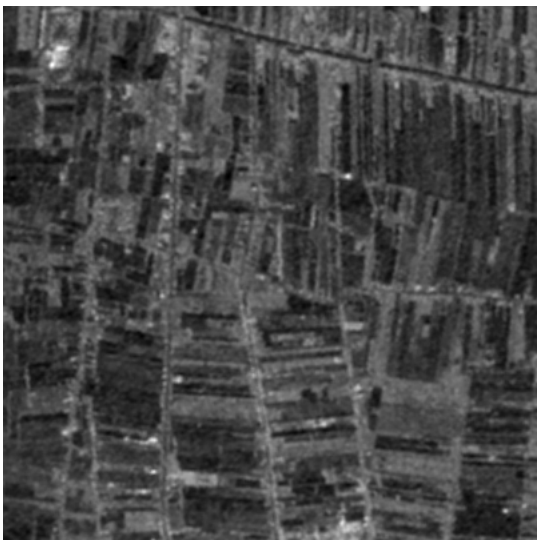
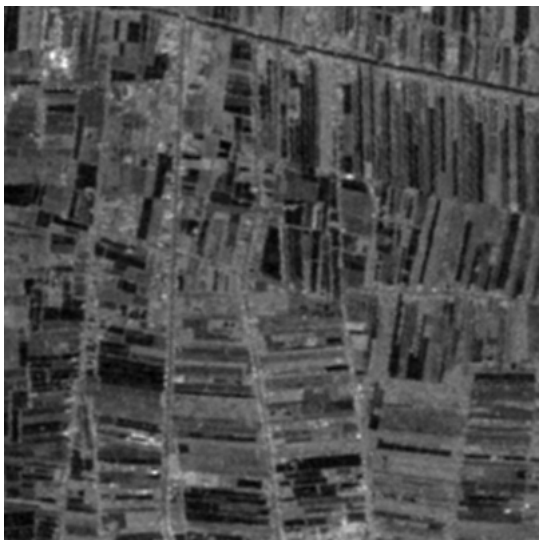
Time 2

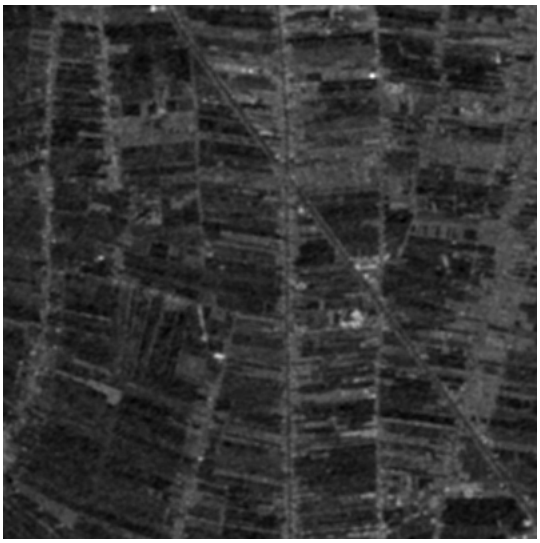
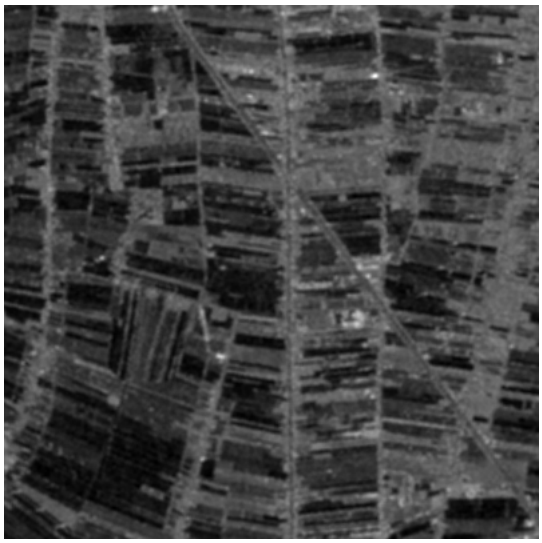
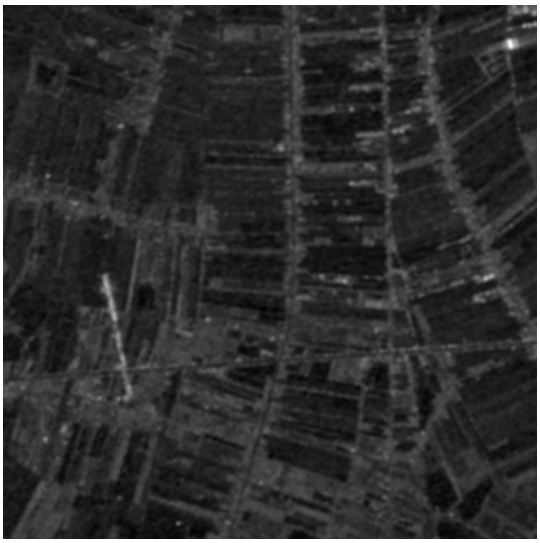
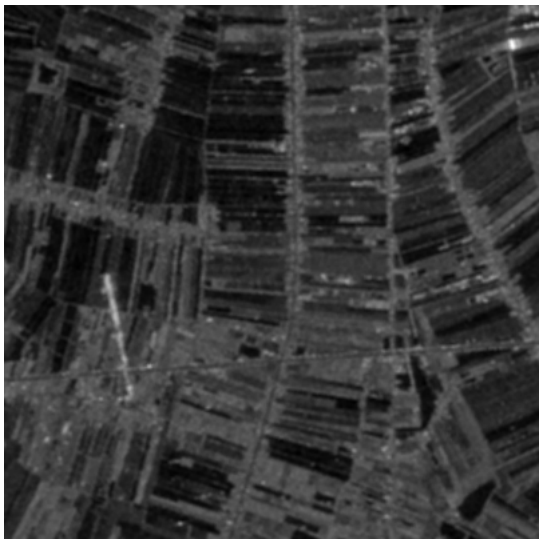
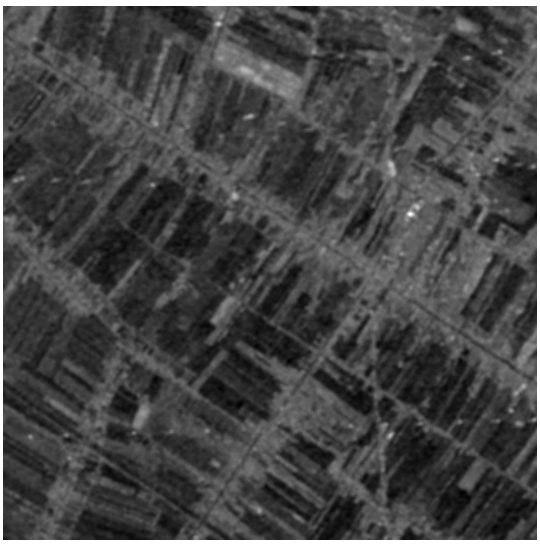
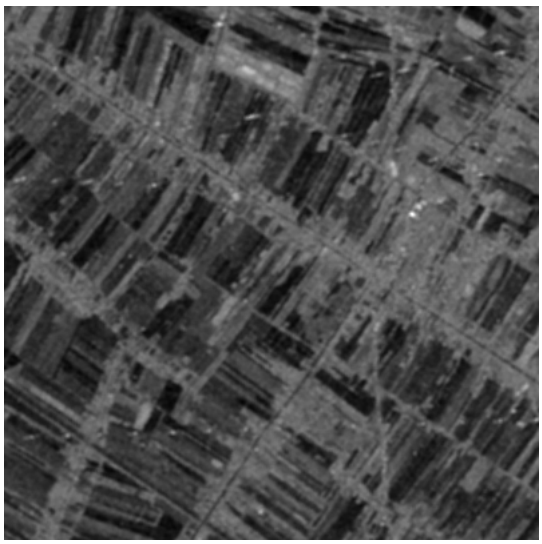


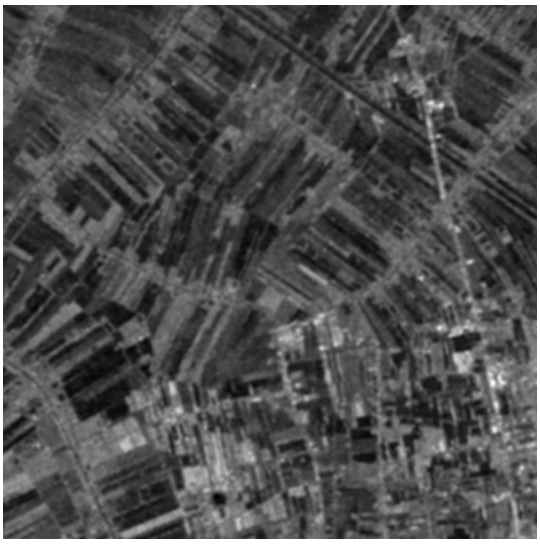
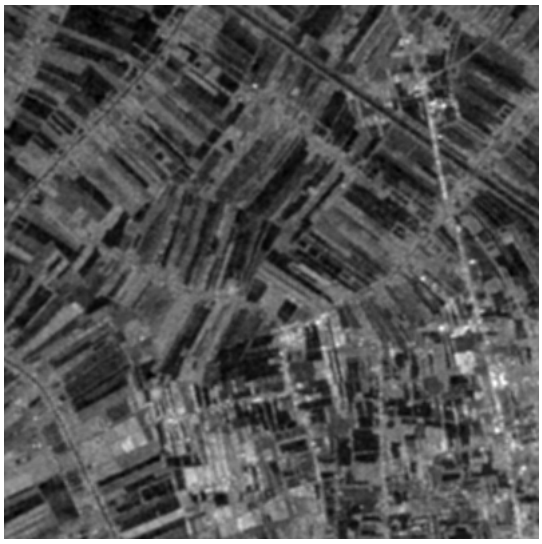
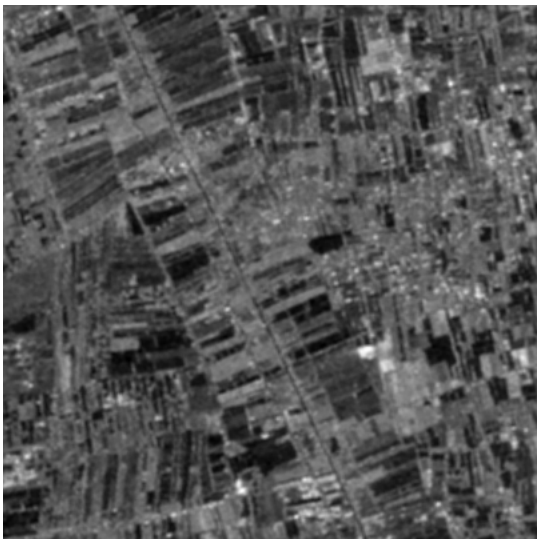
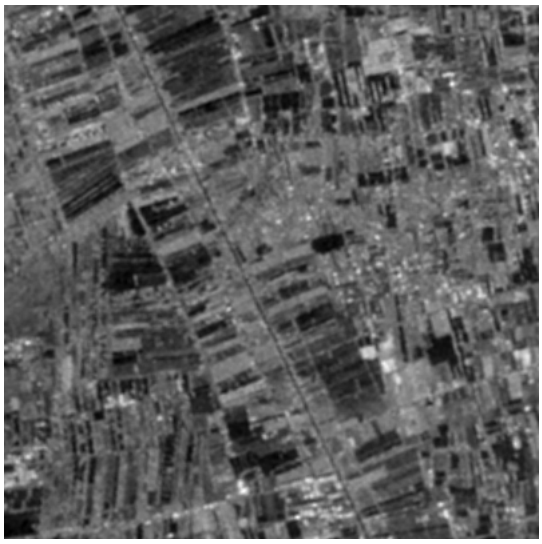


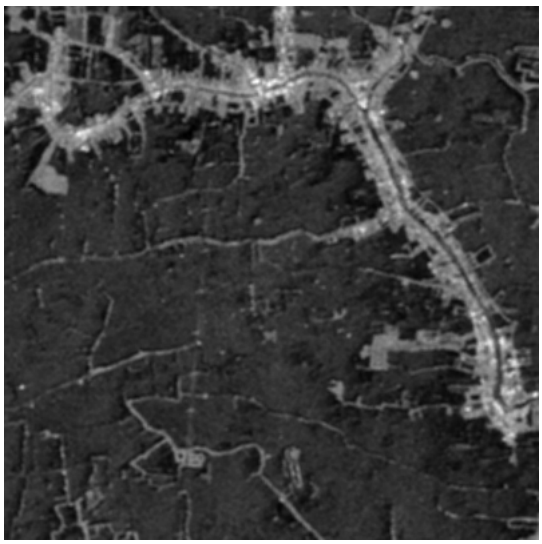



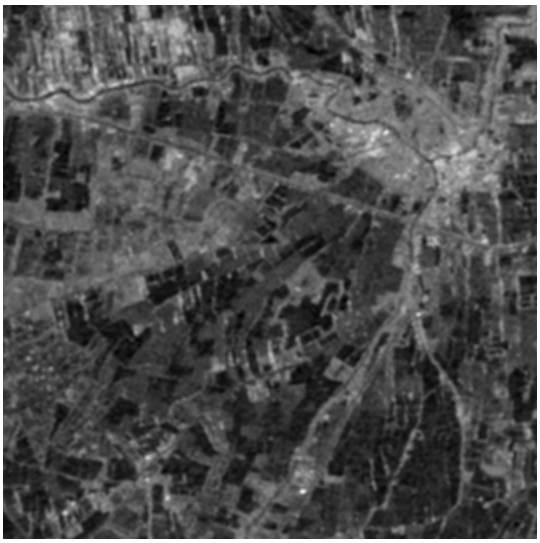
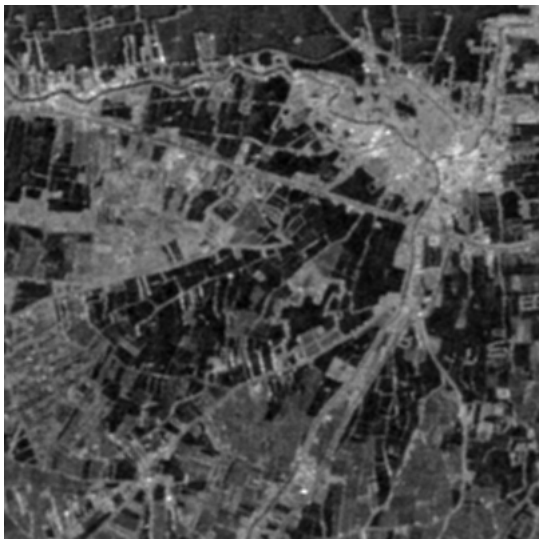
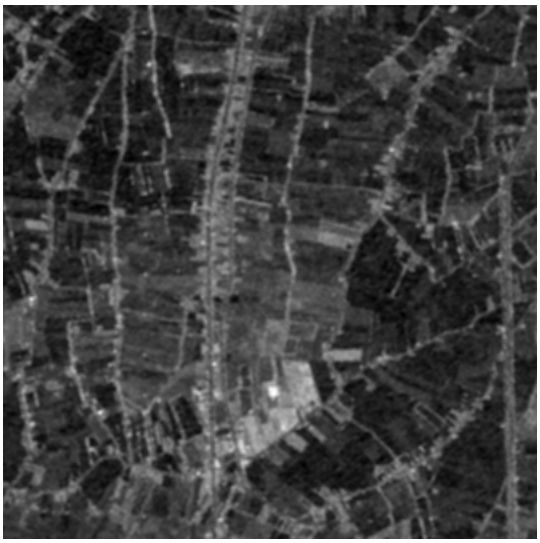
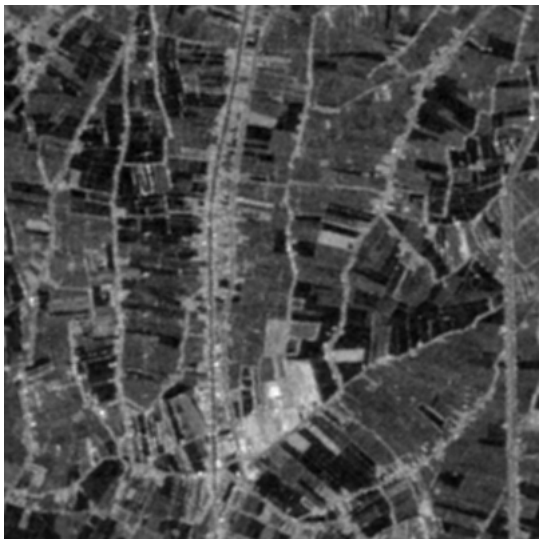
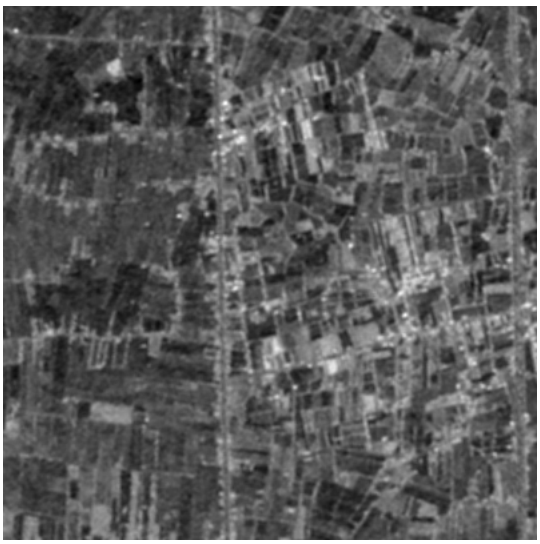
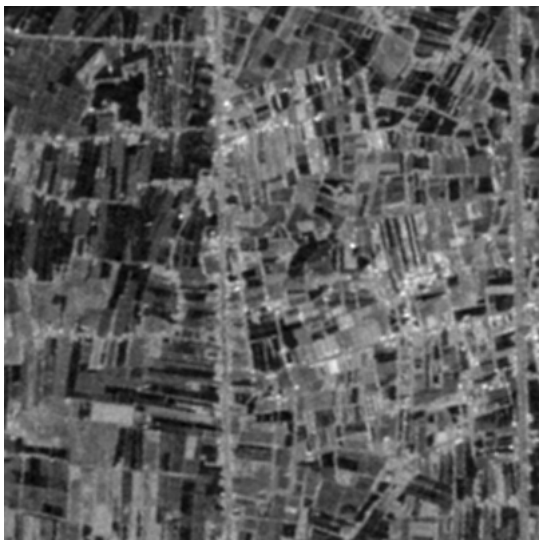


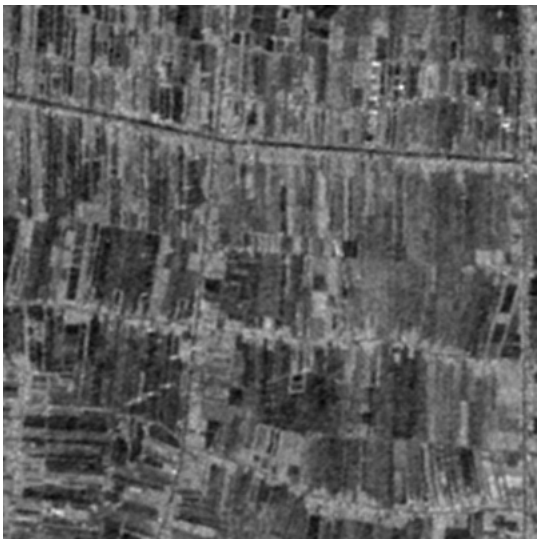
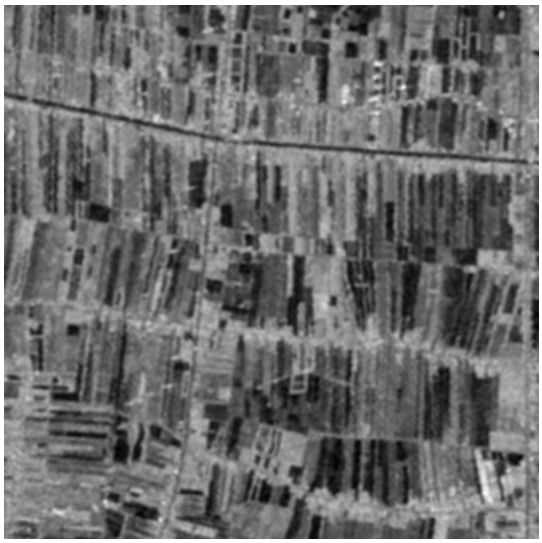
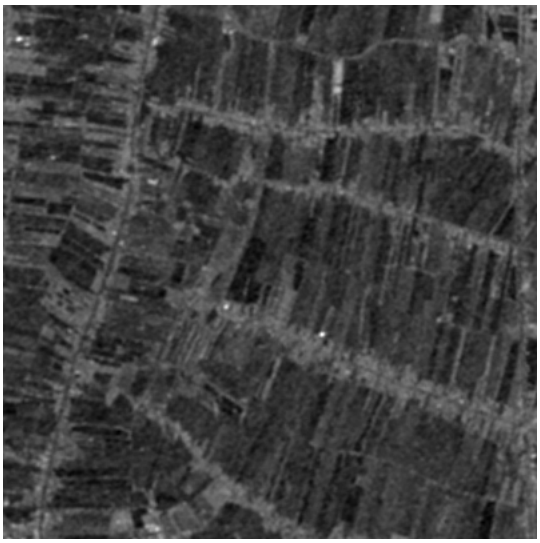
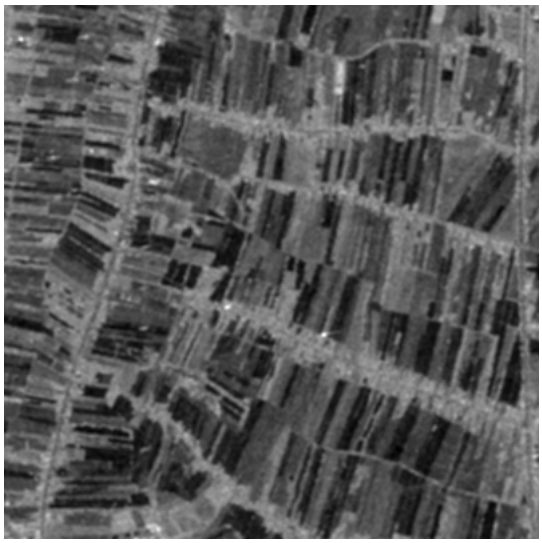
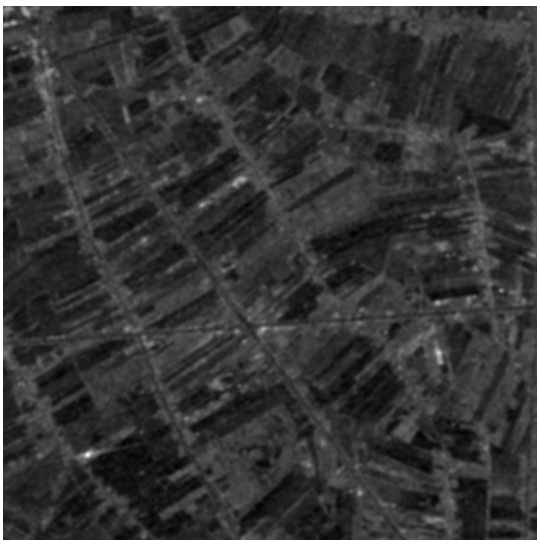
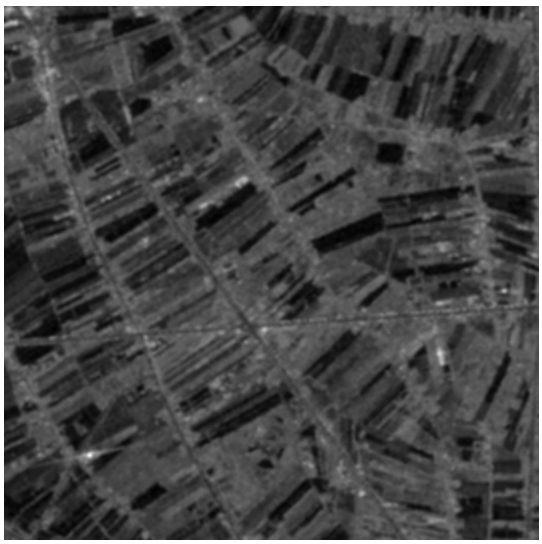





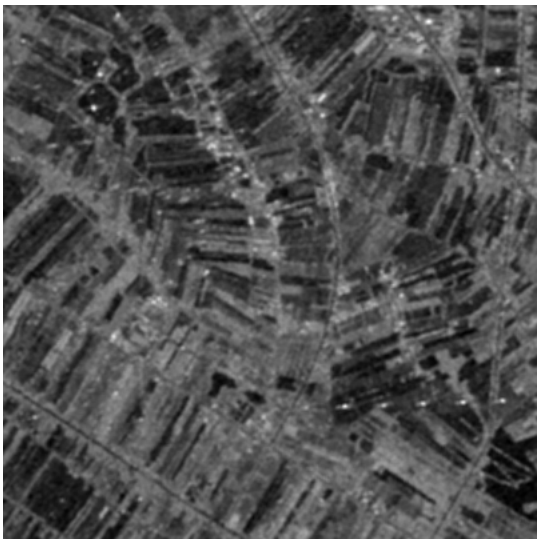
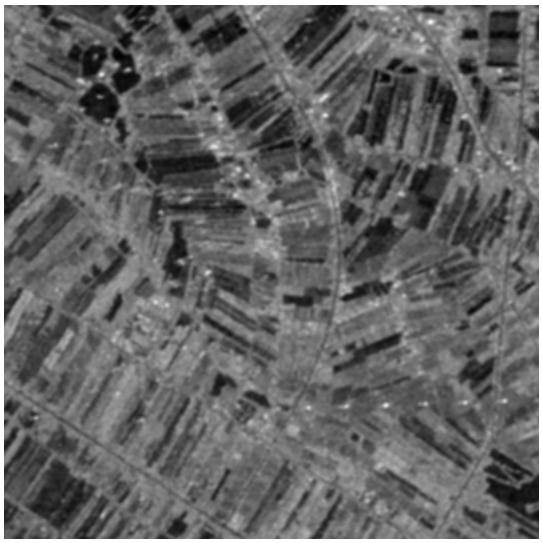
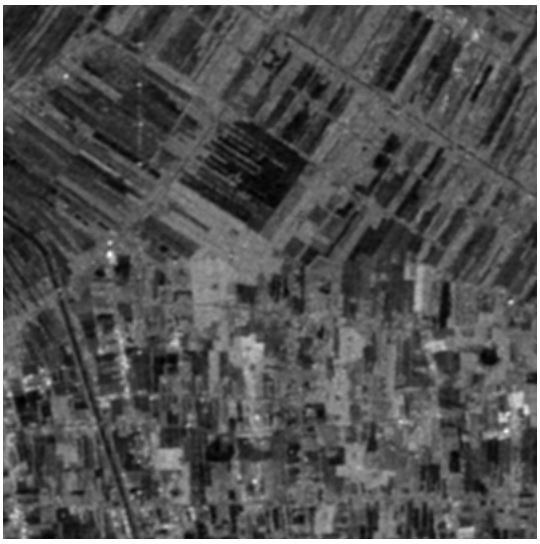
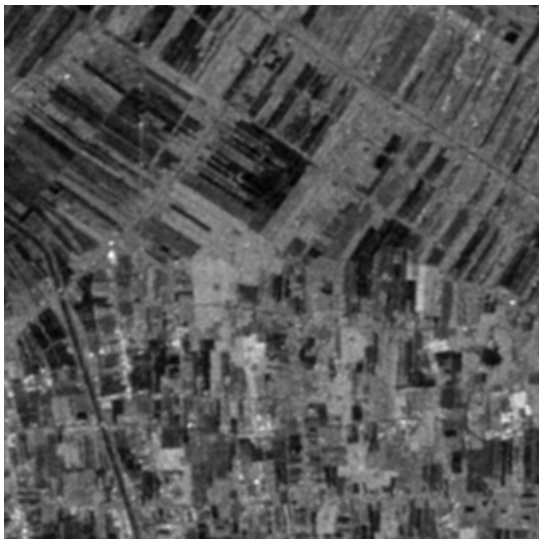
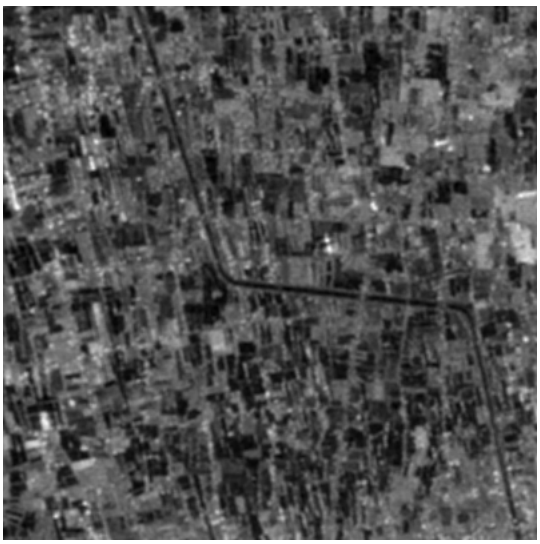
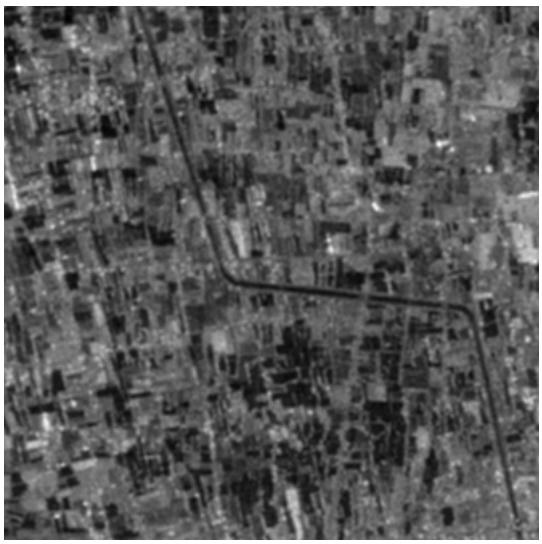







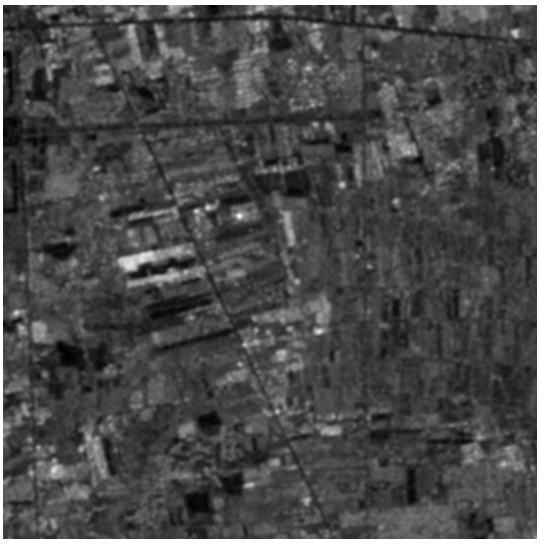
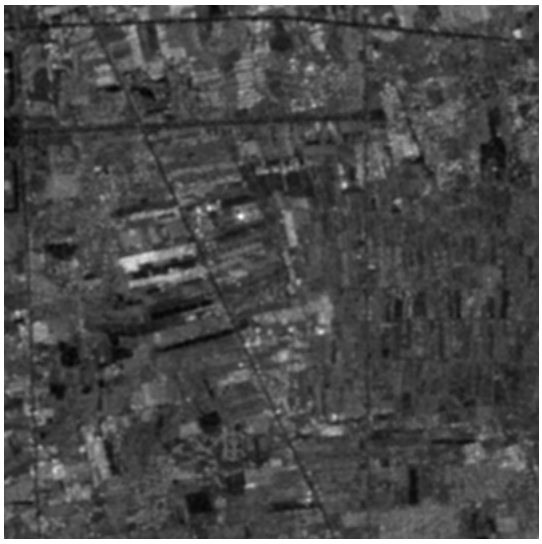


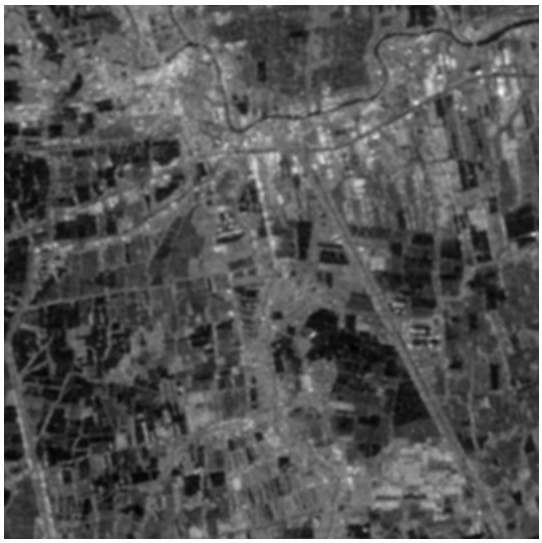
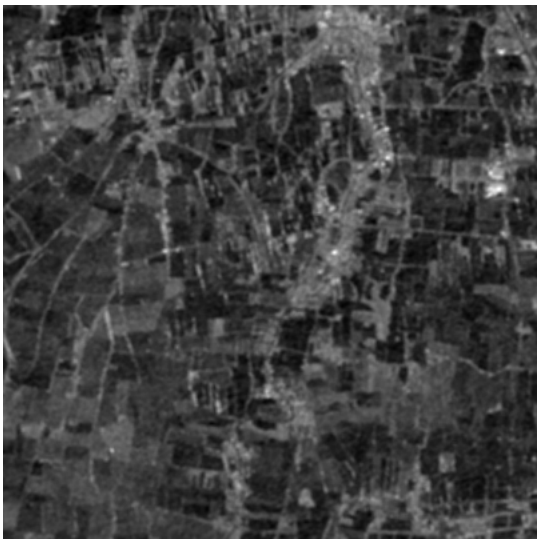
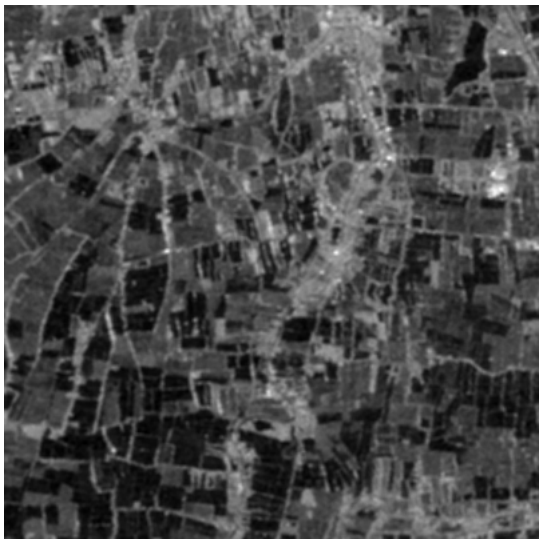
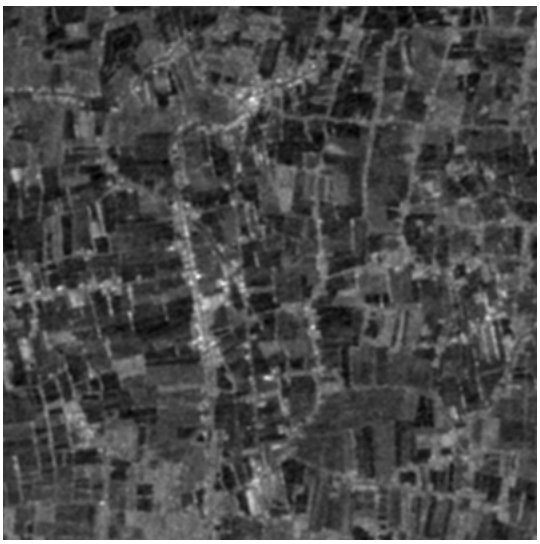
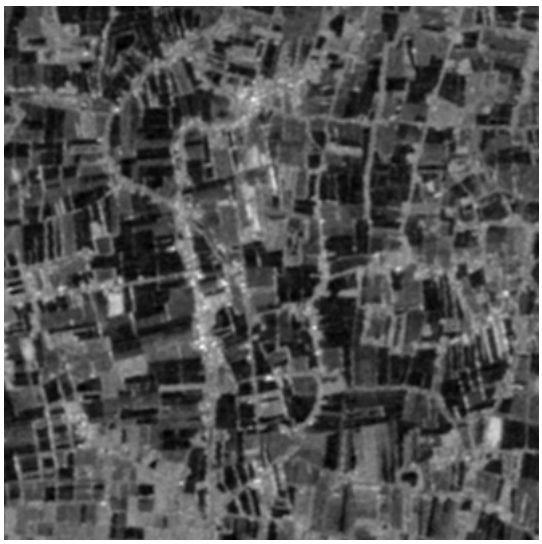







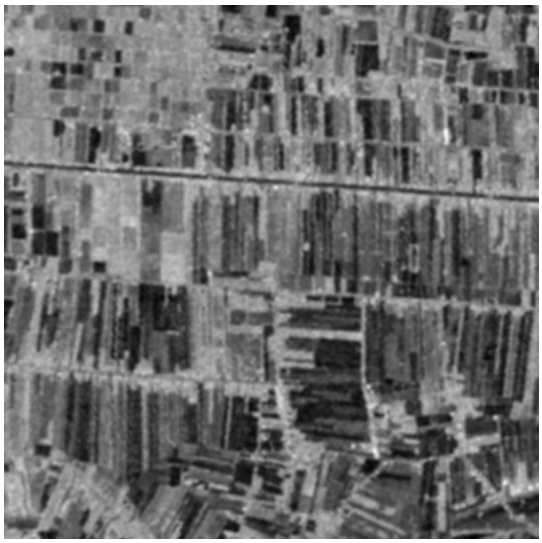
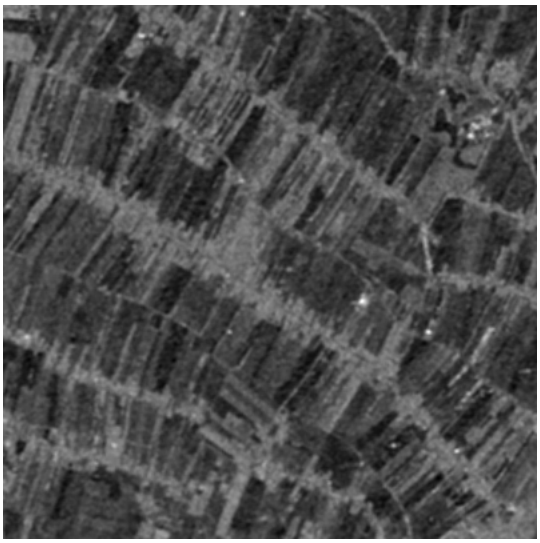
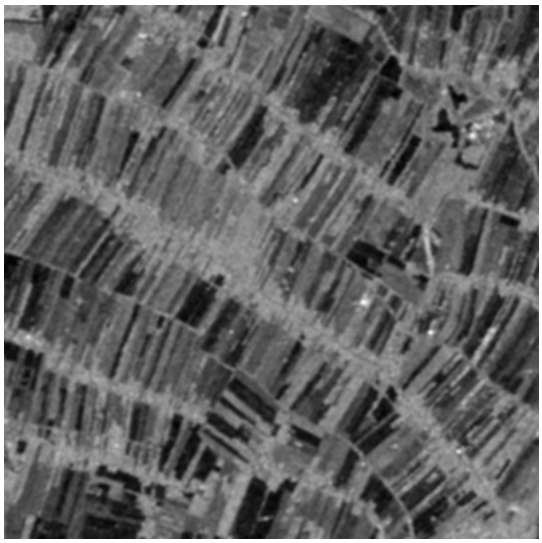
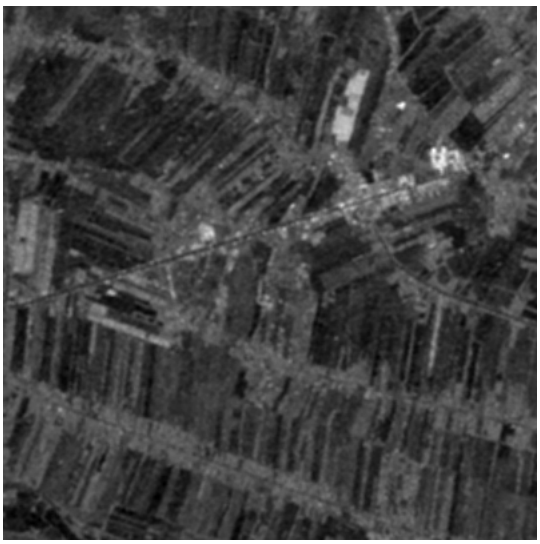
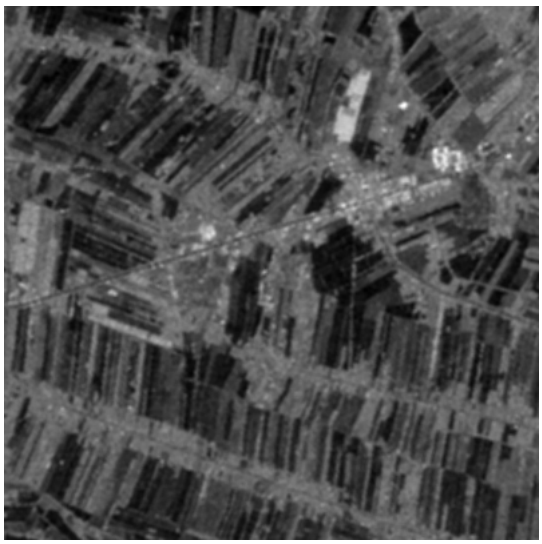







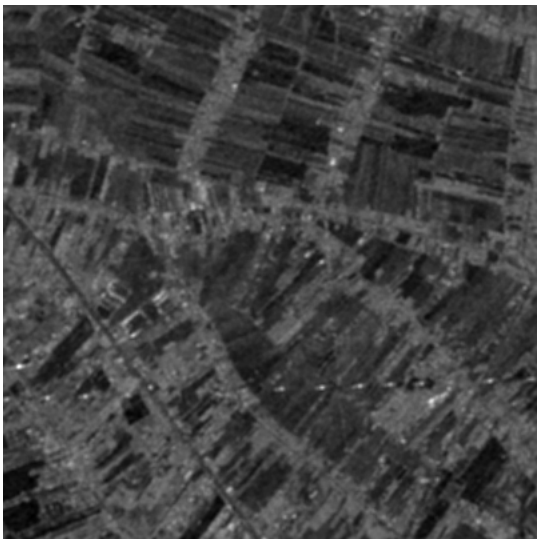
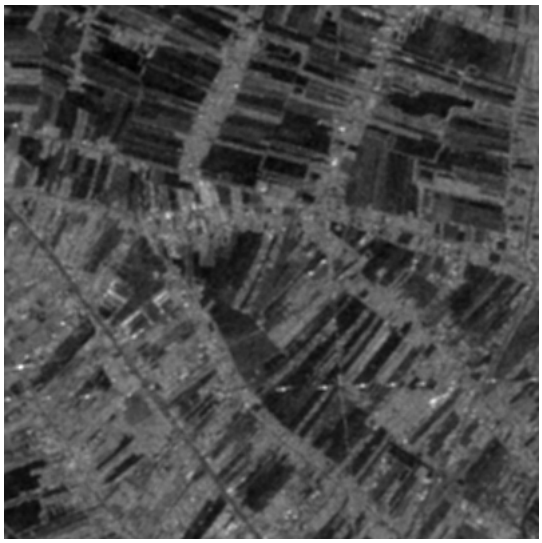
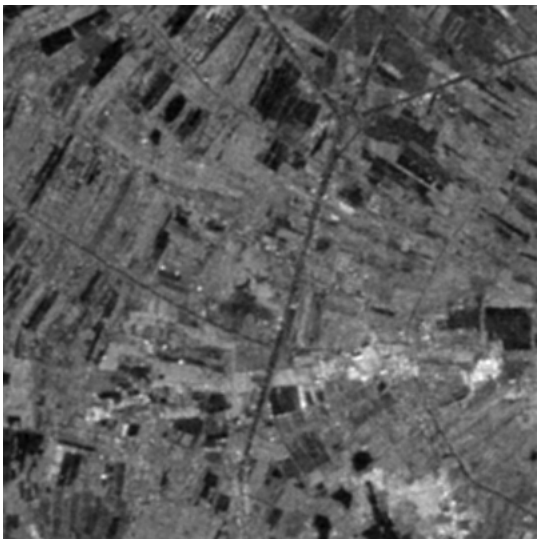
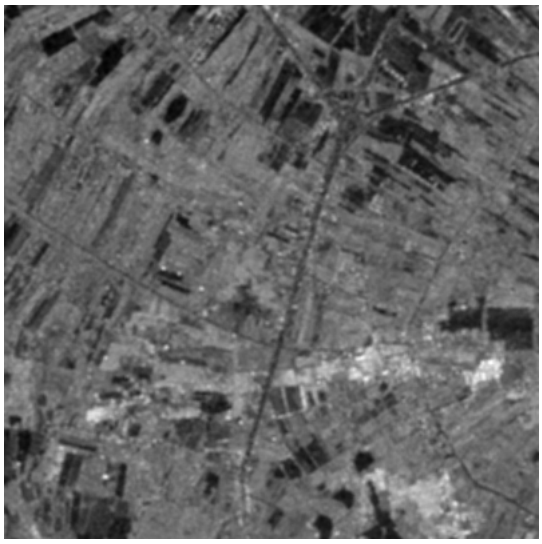
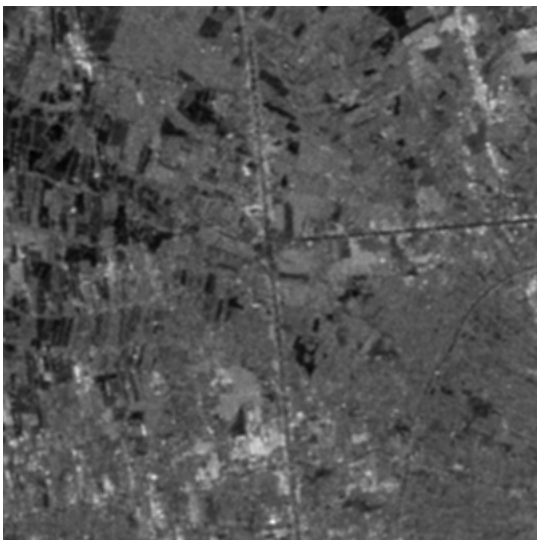
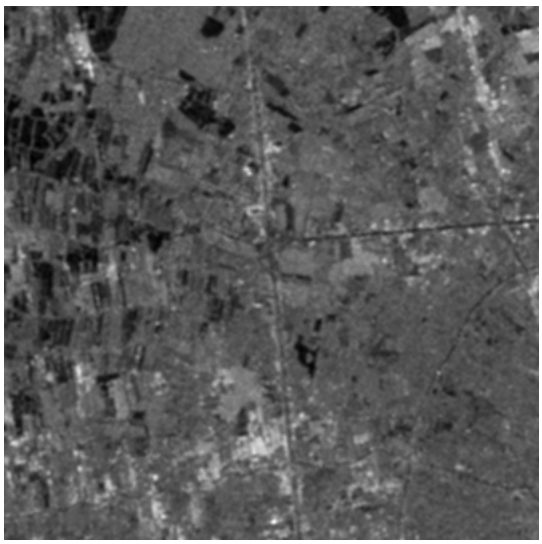







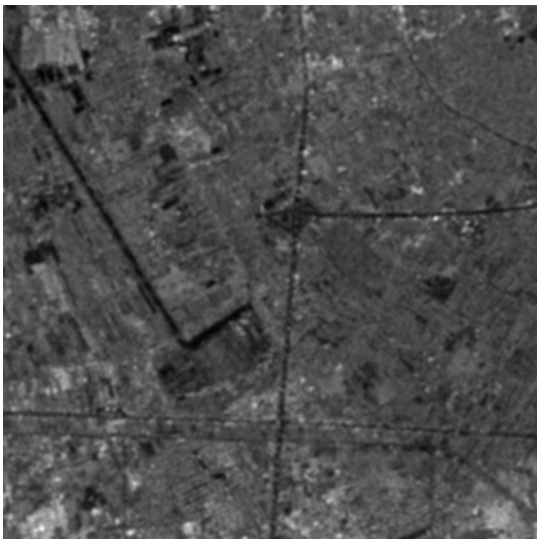
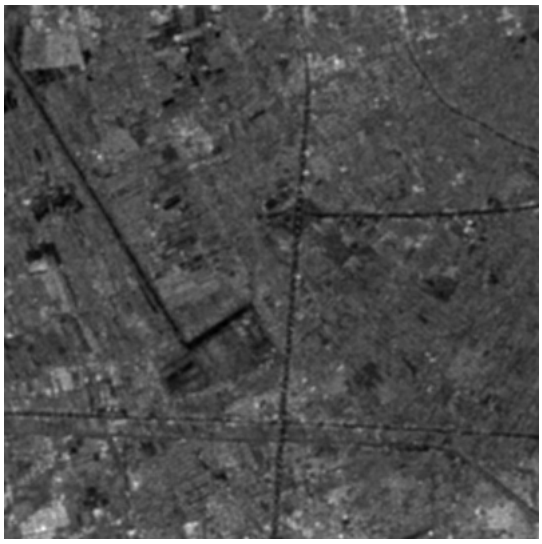
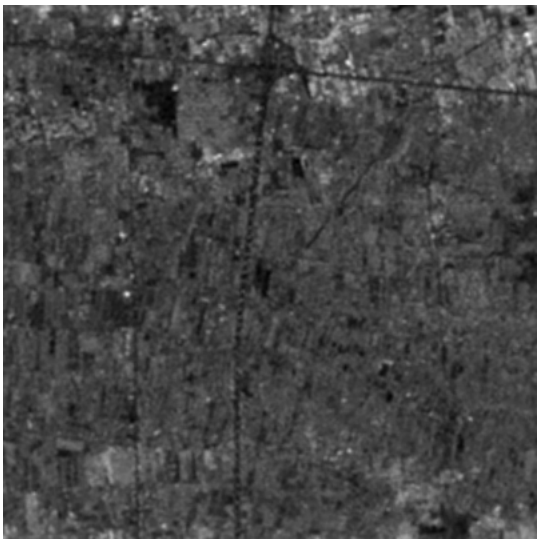
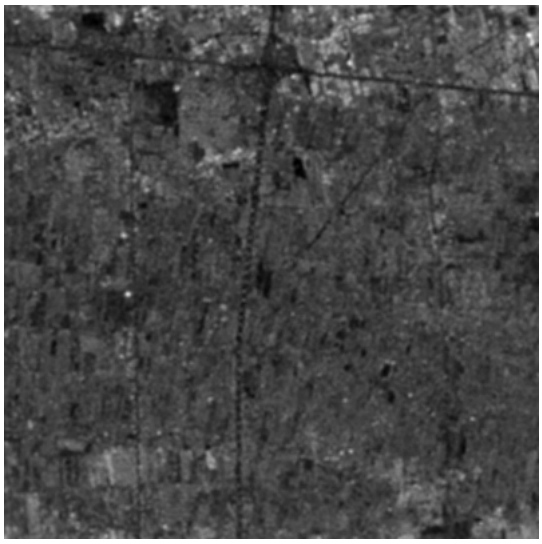
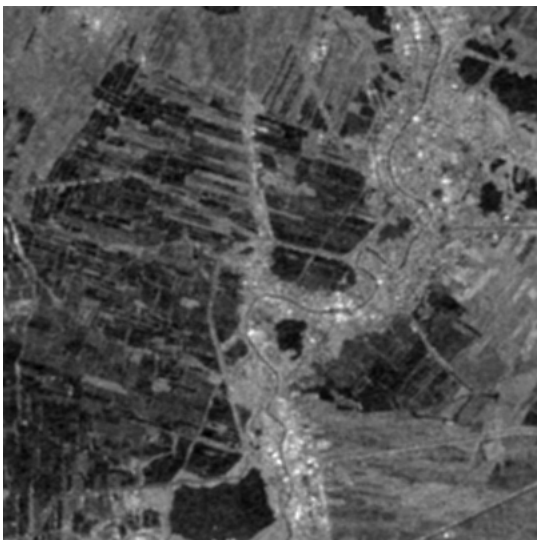
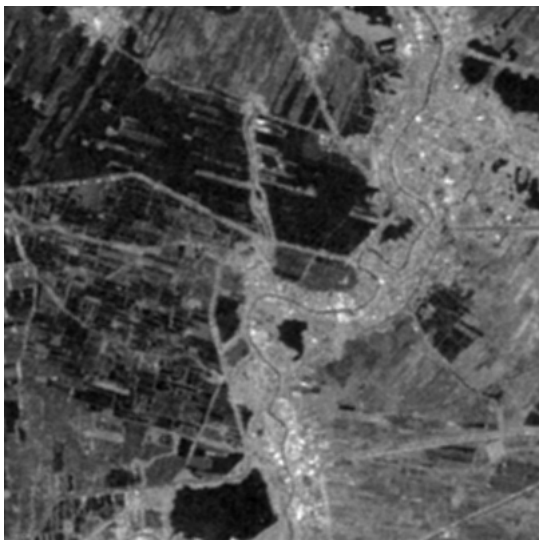


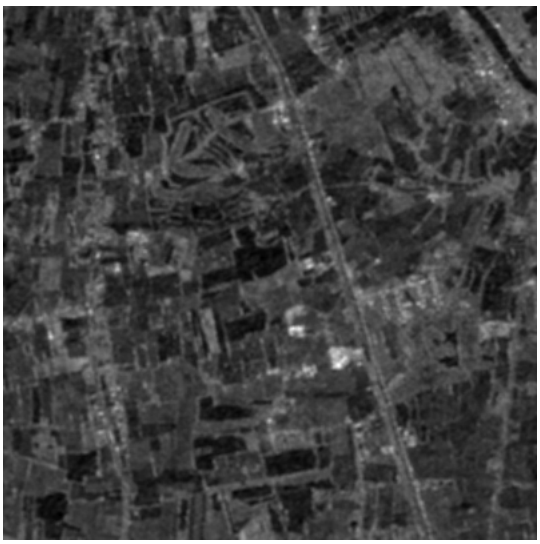
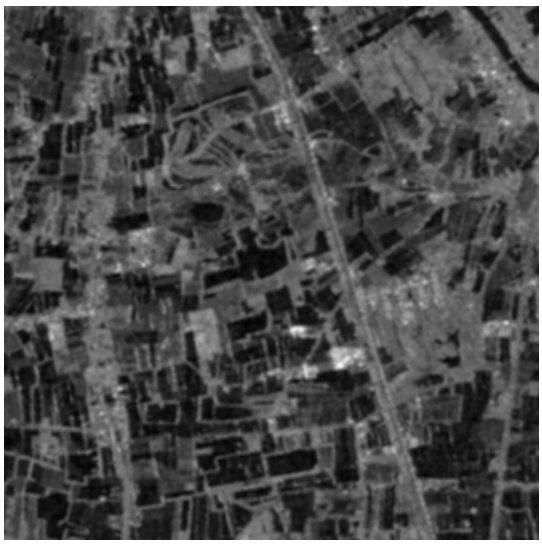







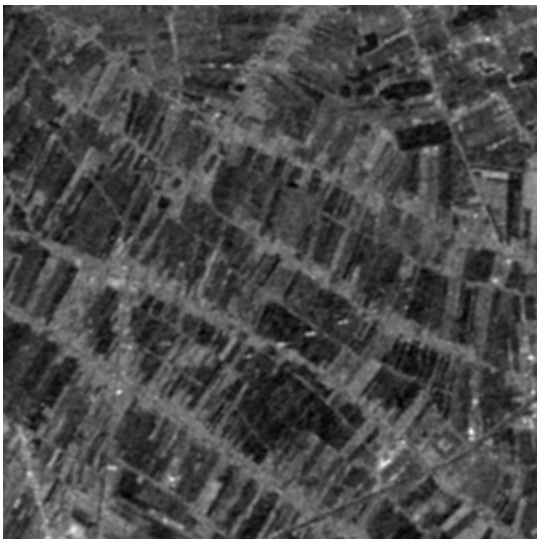
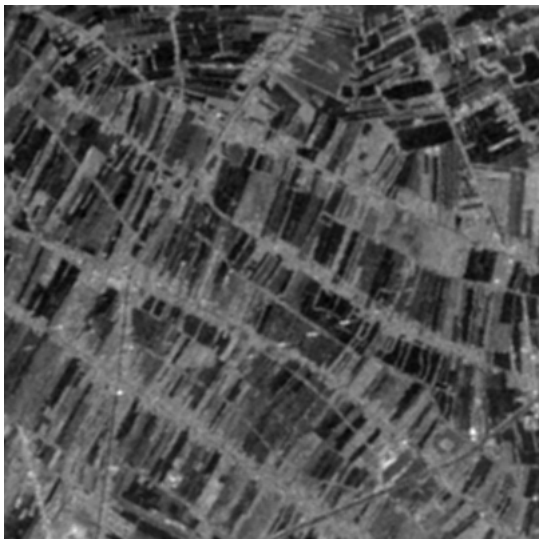
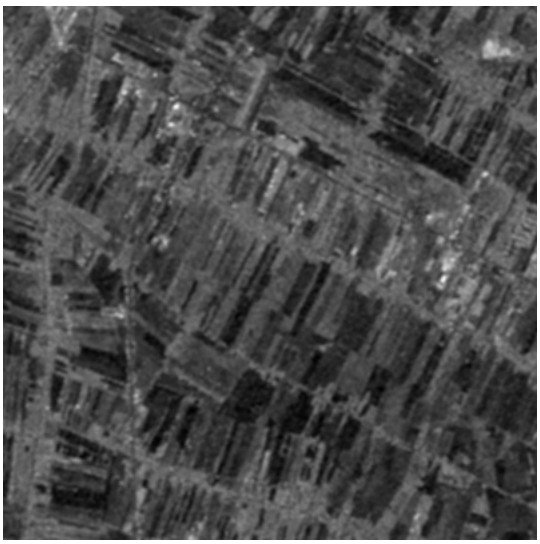
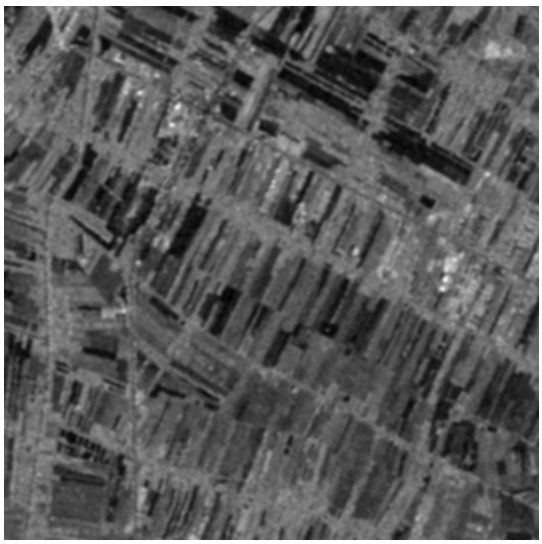







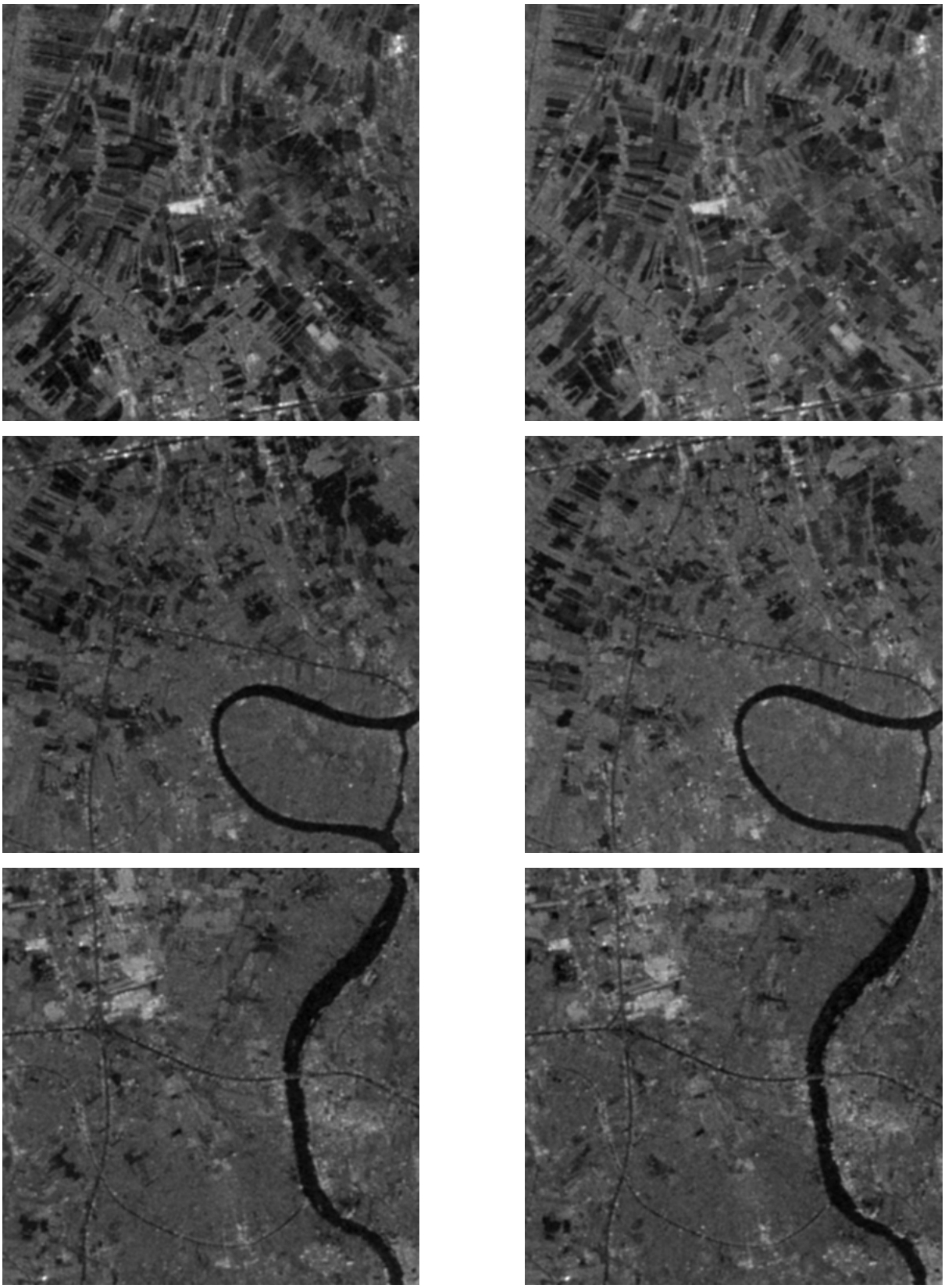


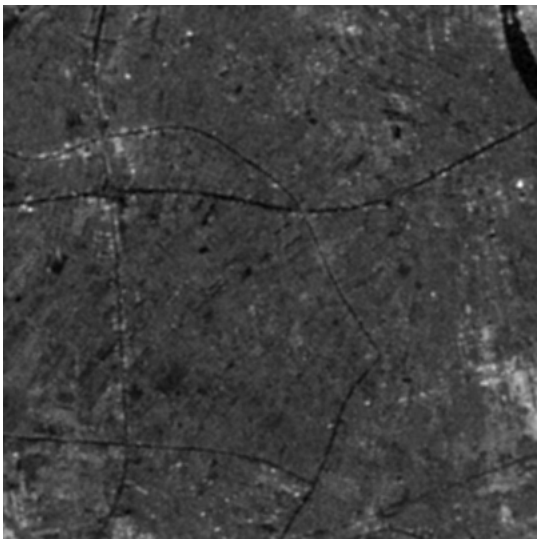
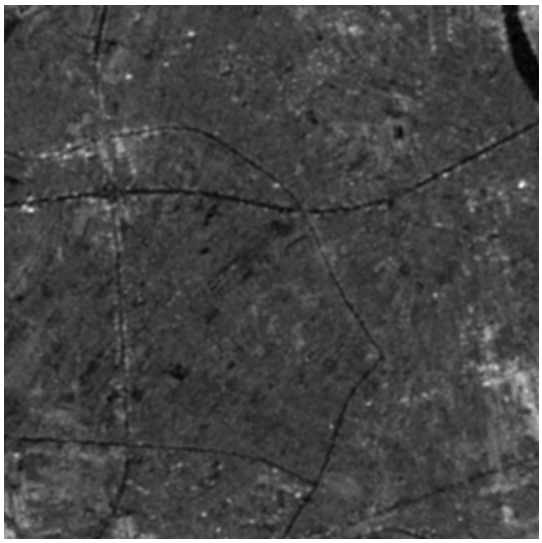
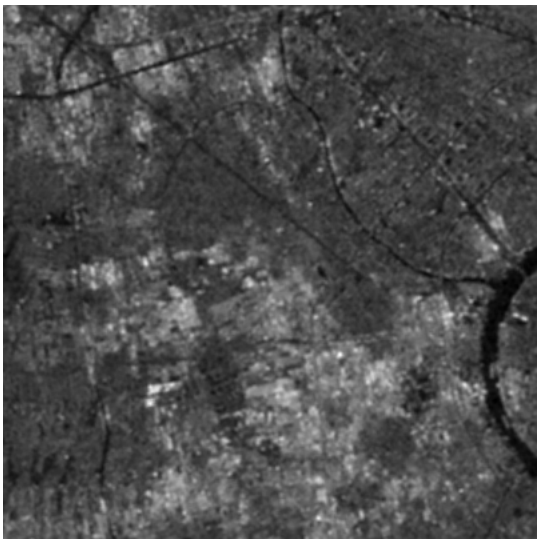
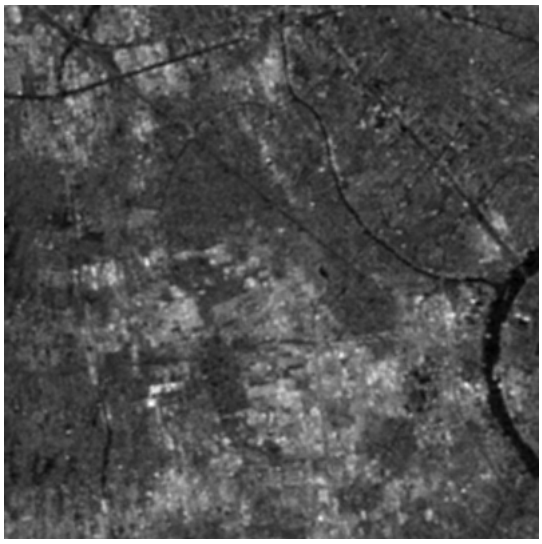


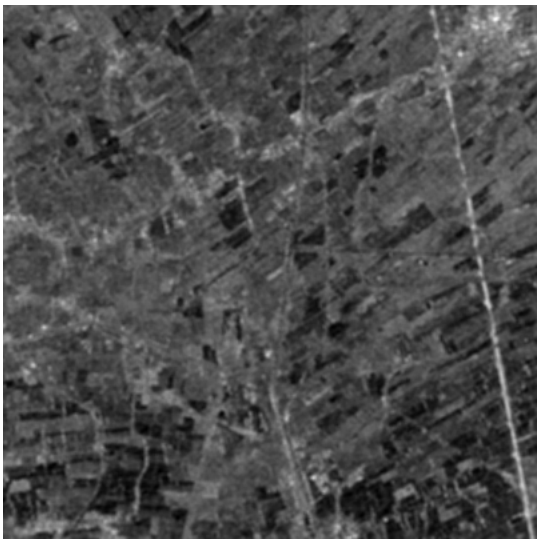
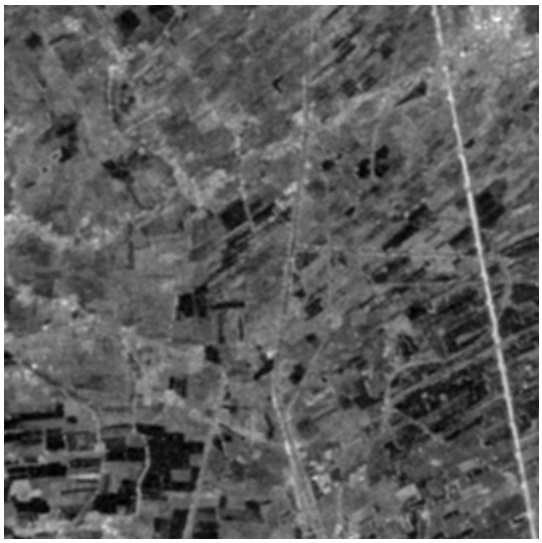



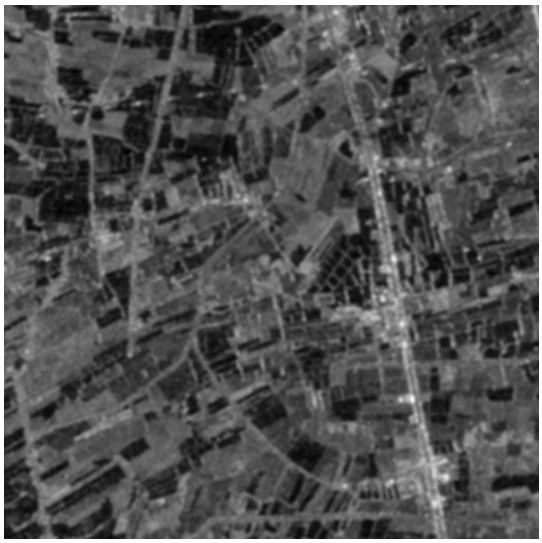




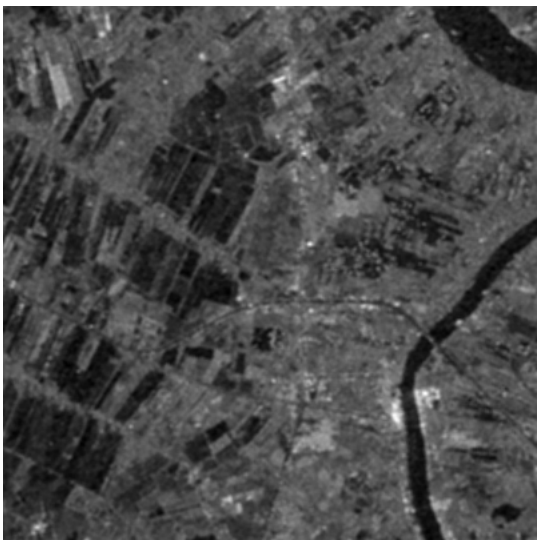
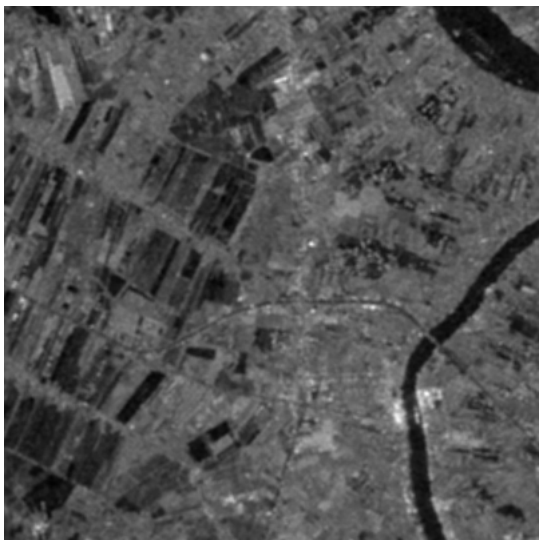







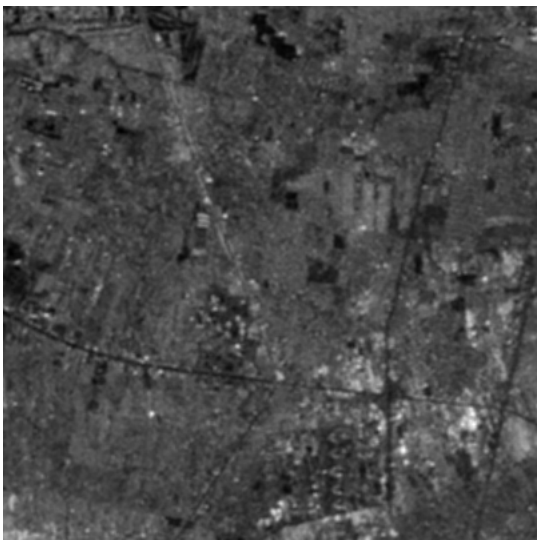
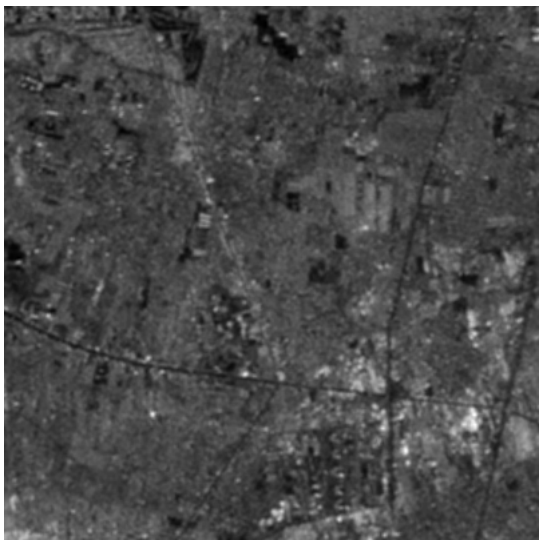


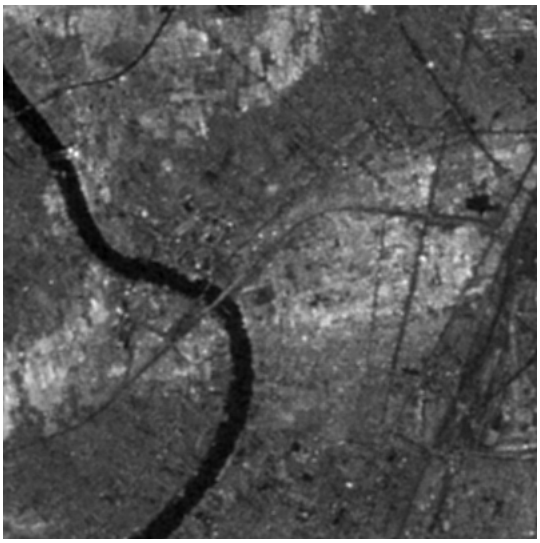
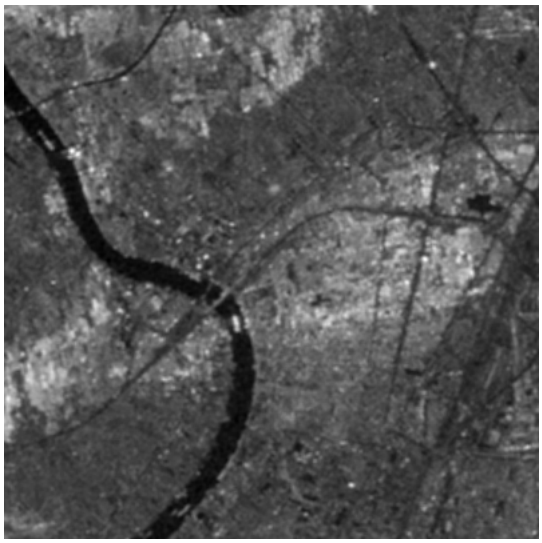
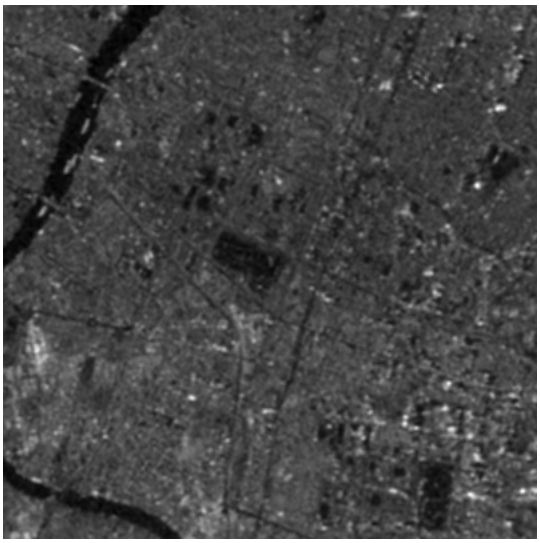
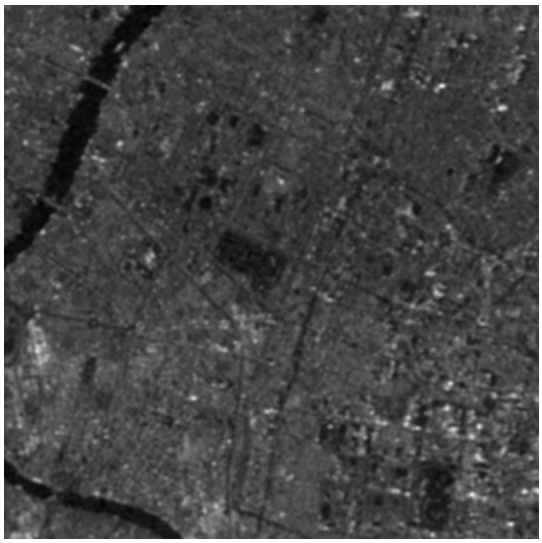




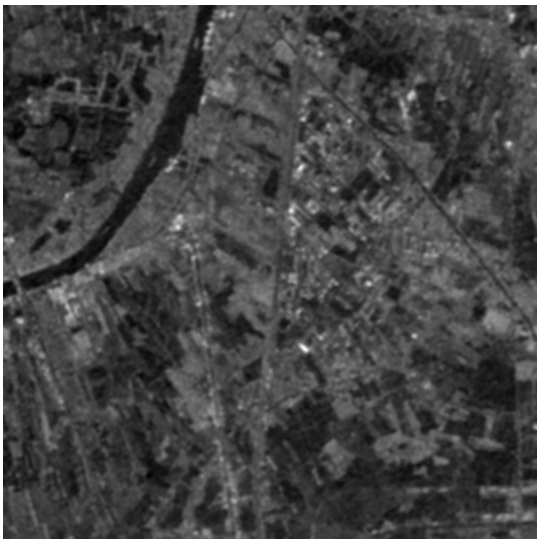
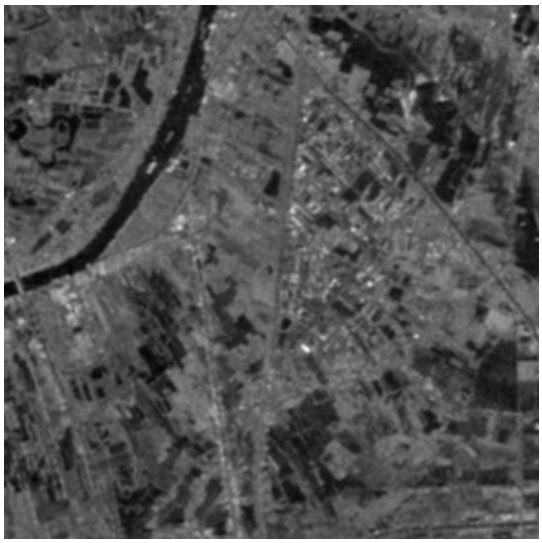


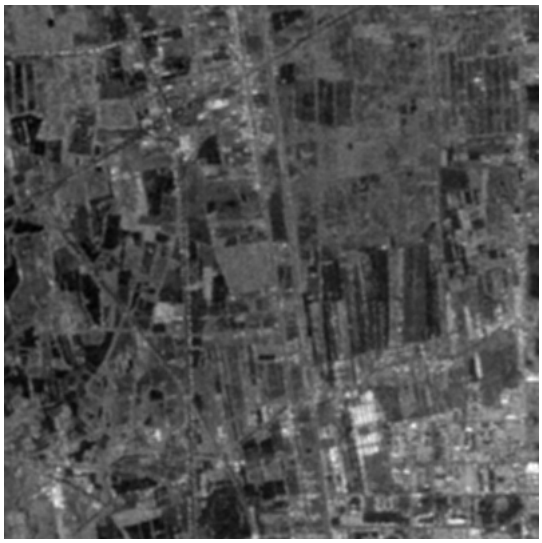
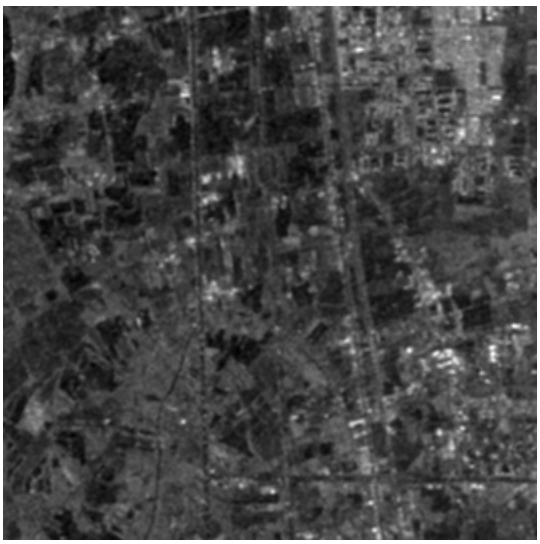
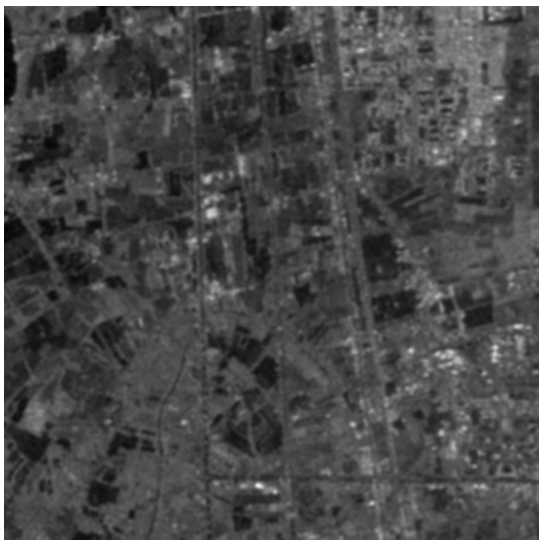





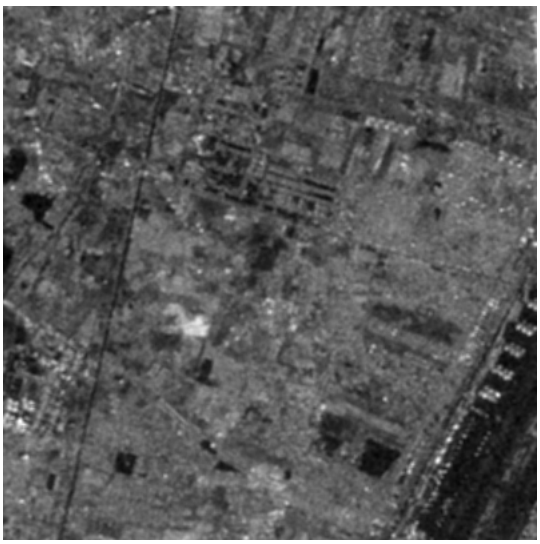
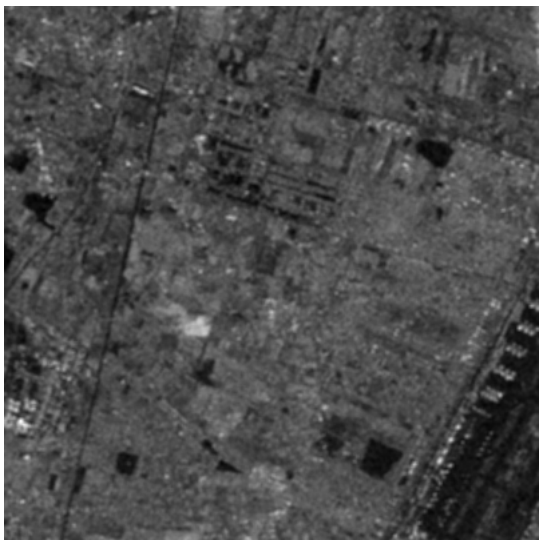


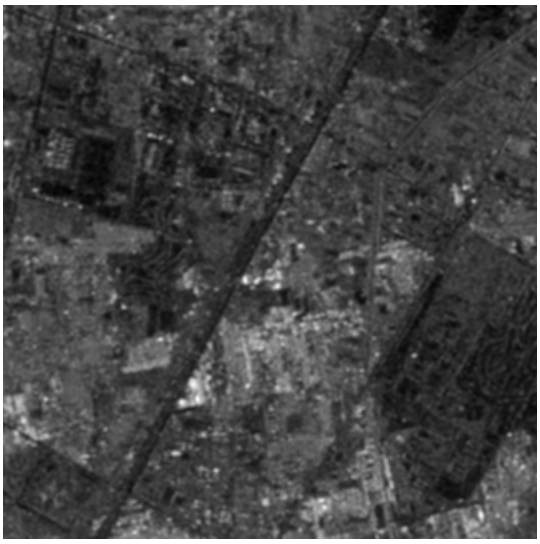




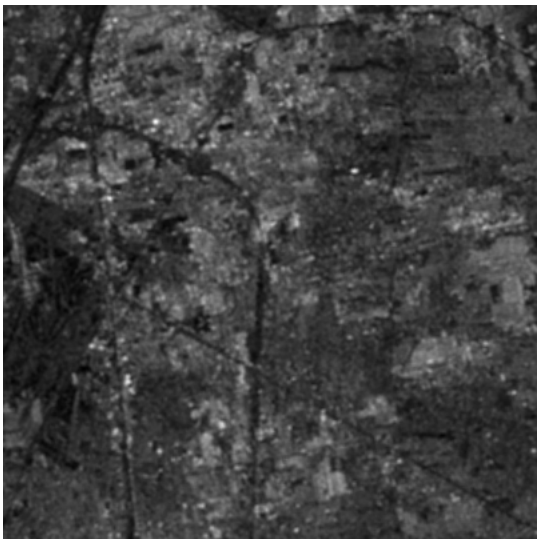




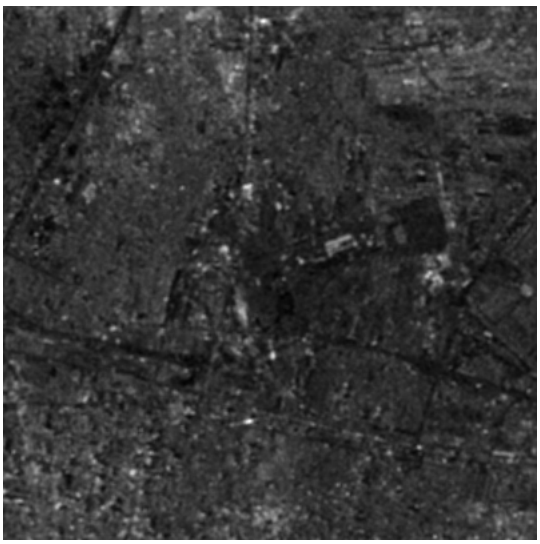
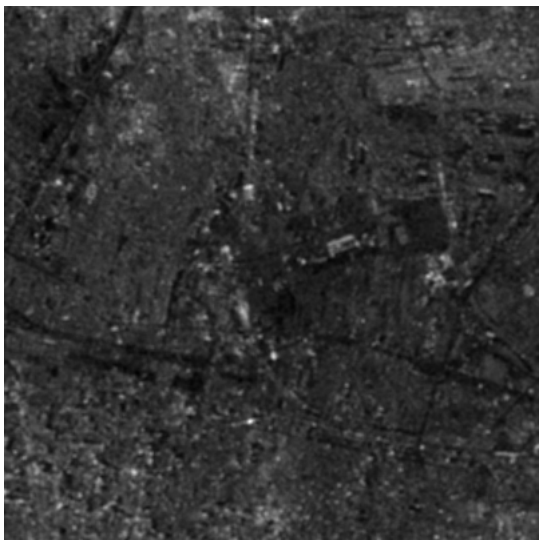


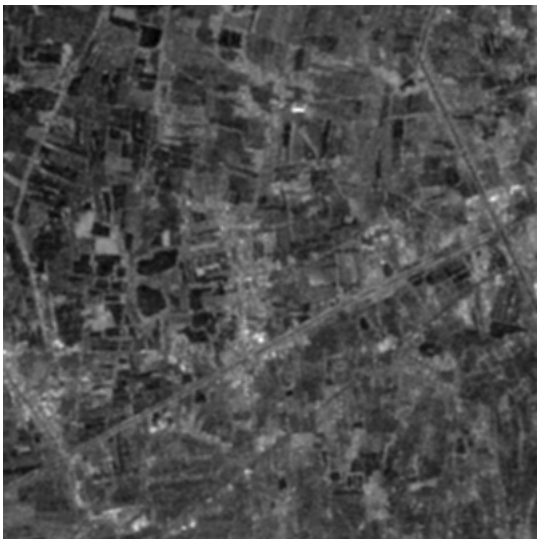
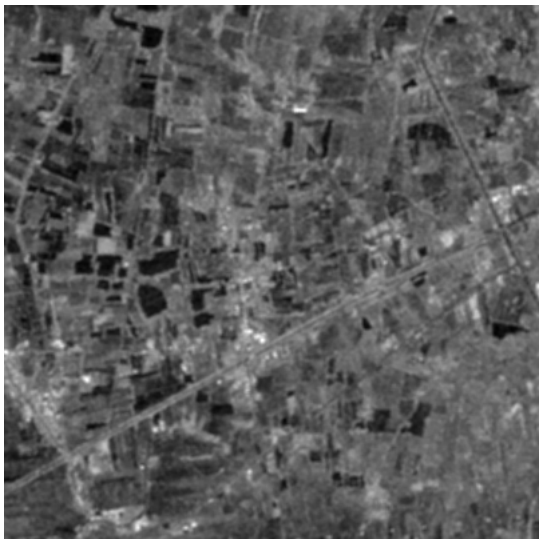
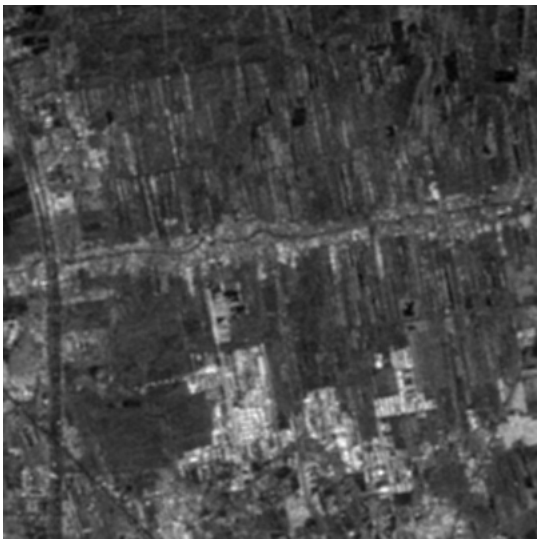


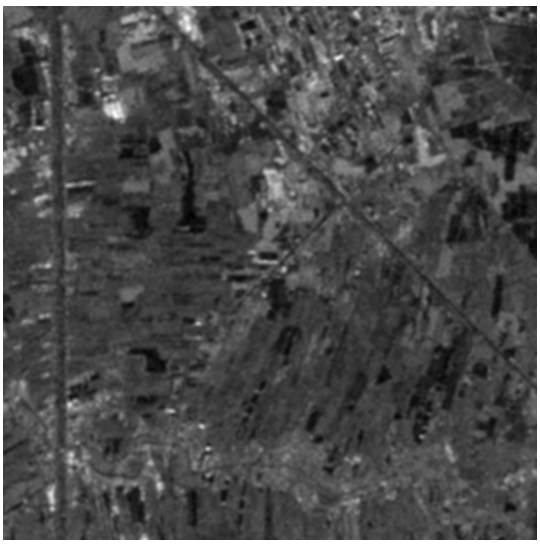
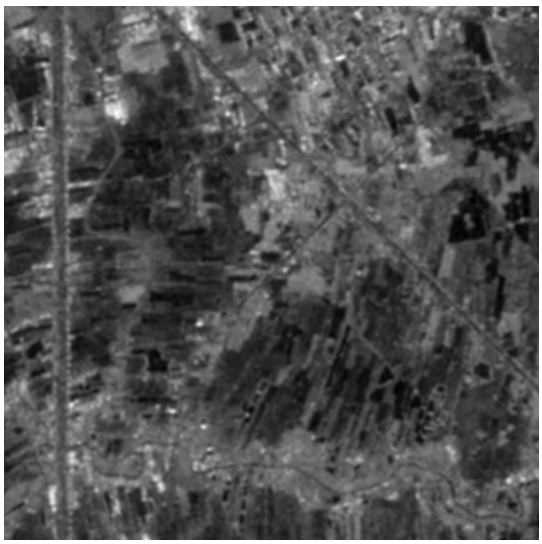



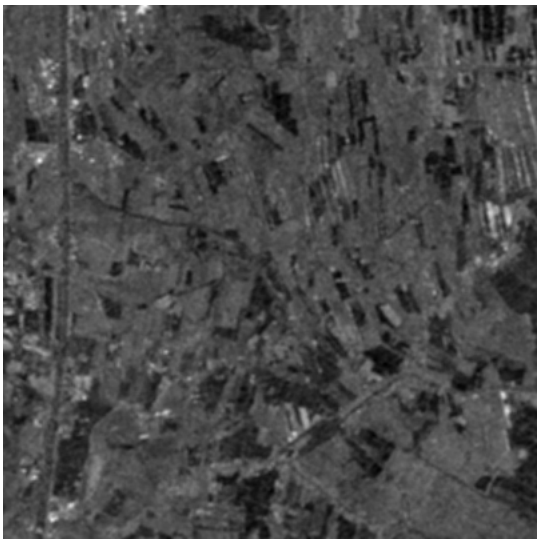
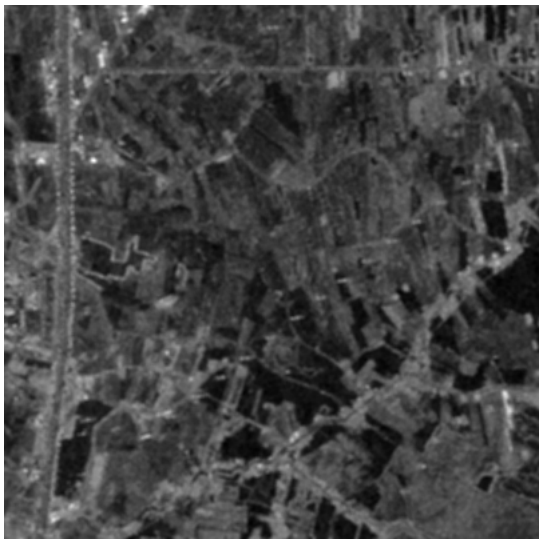




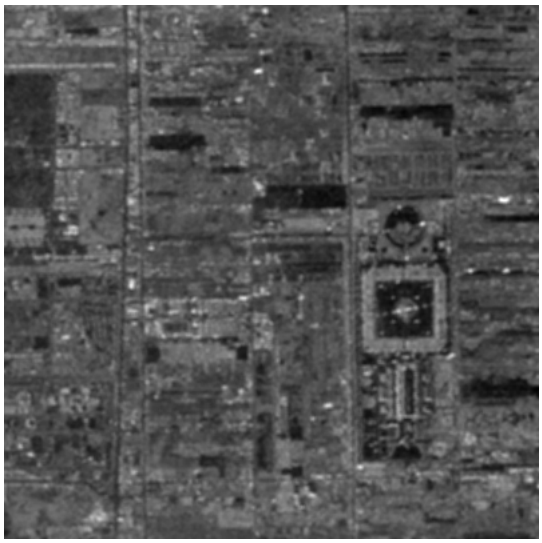
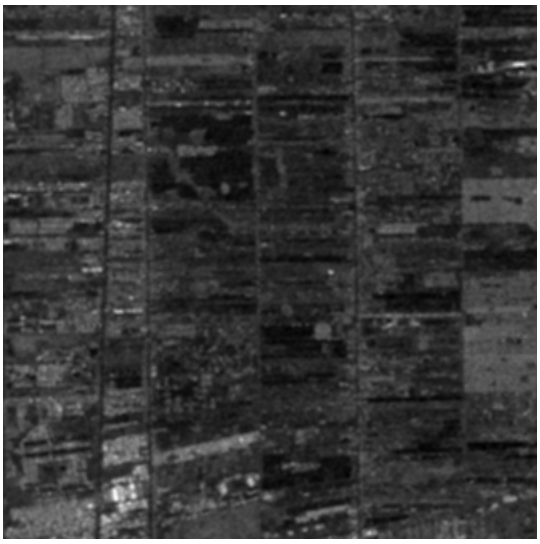
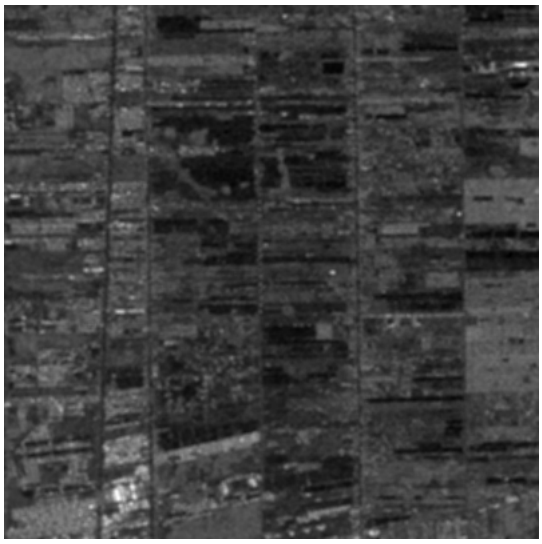



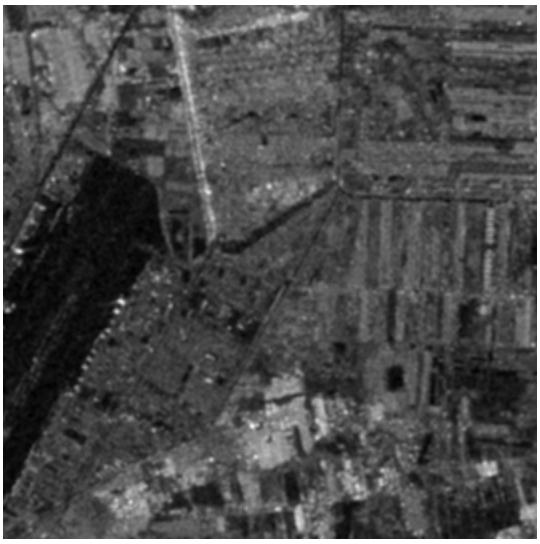
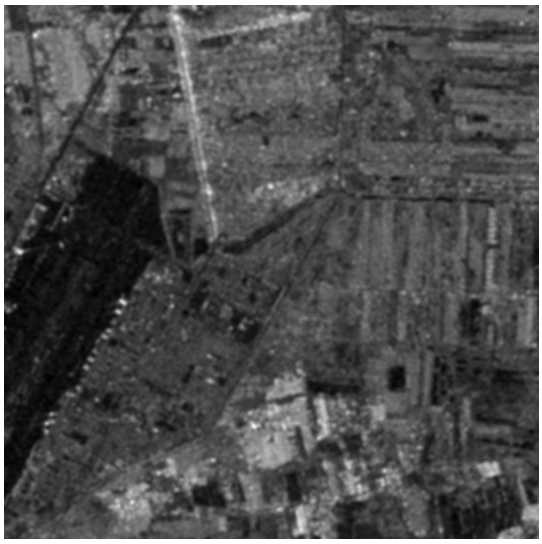
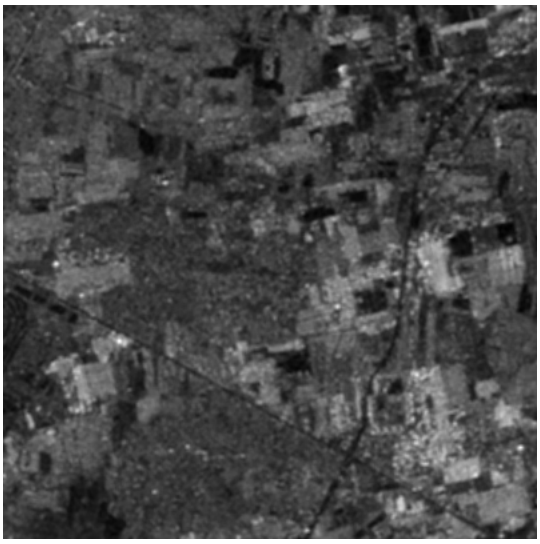
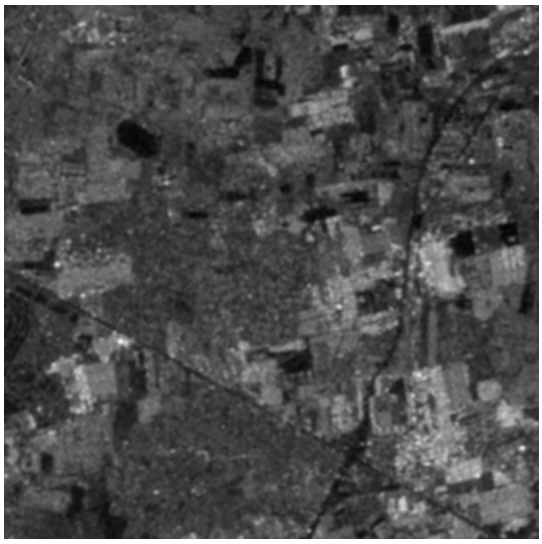
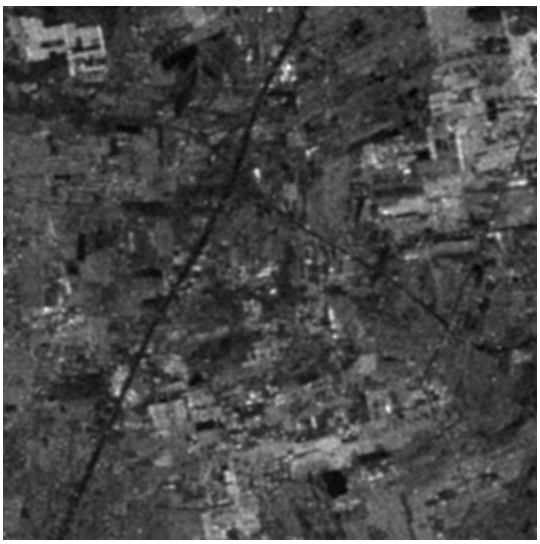
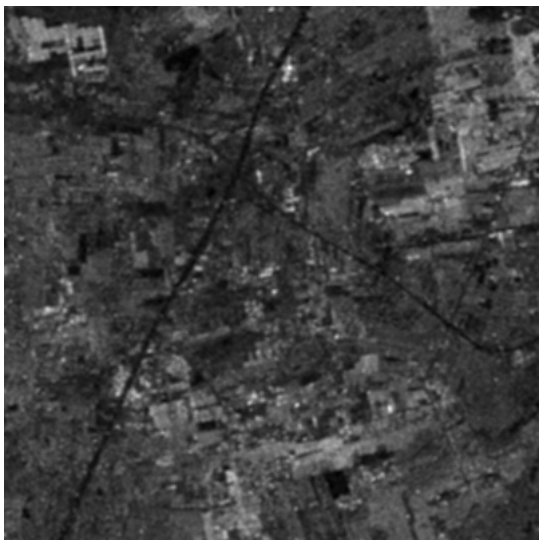




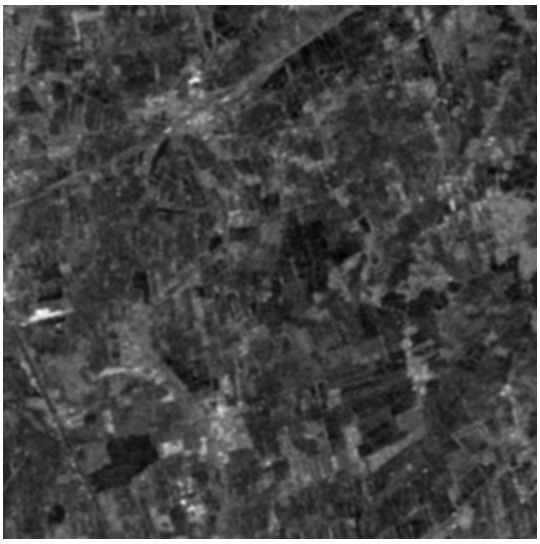
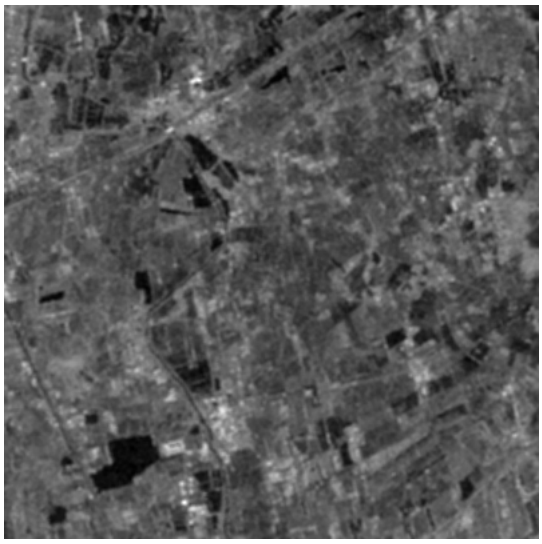
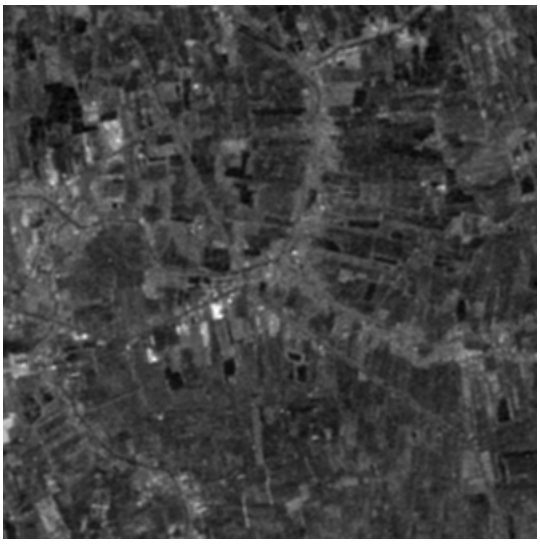
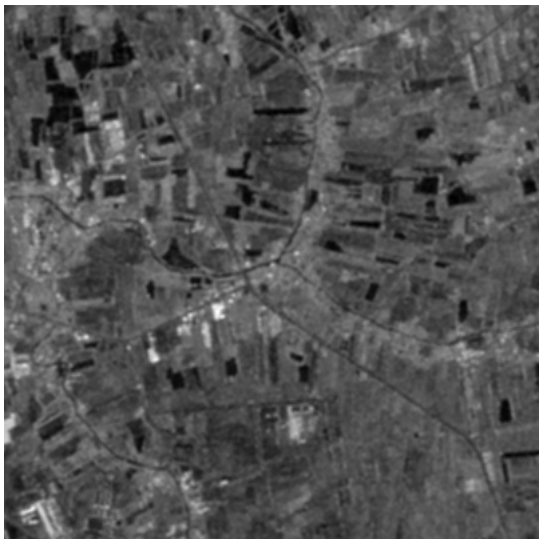



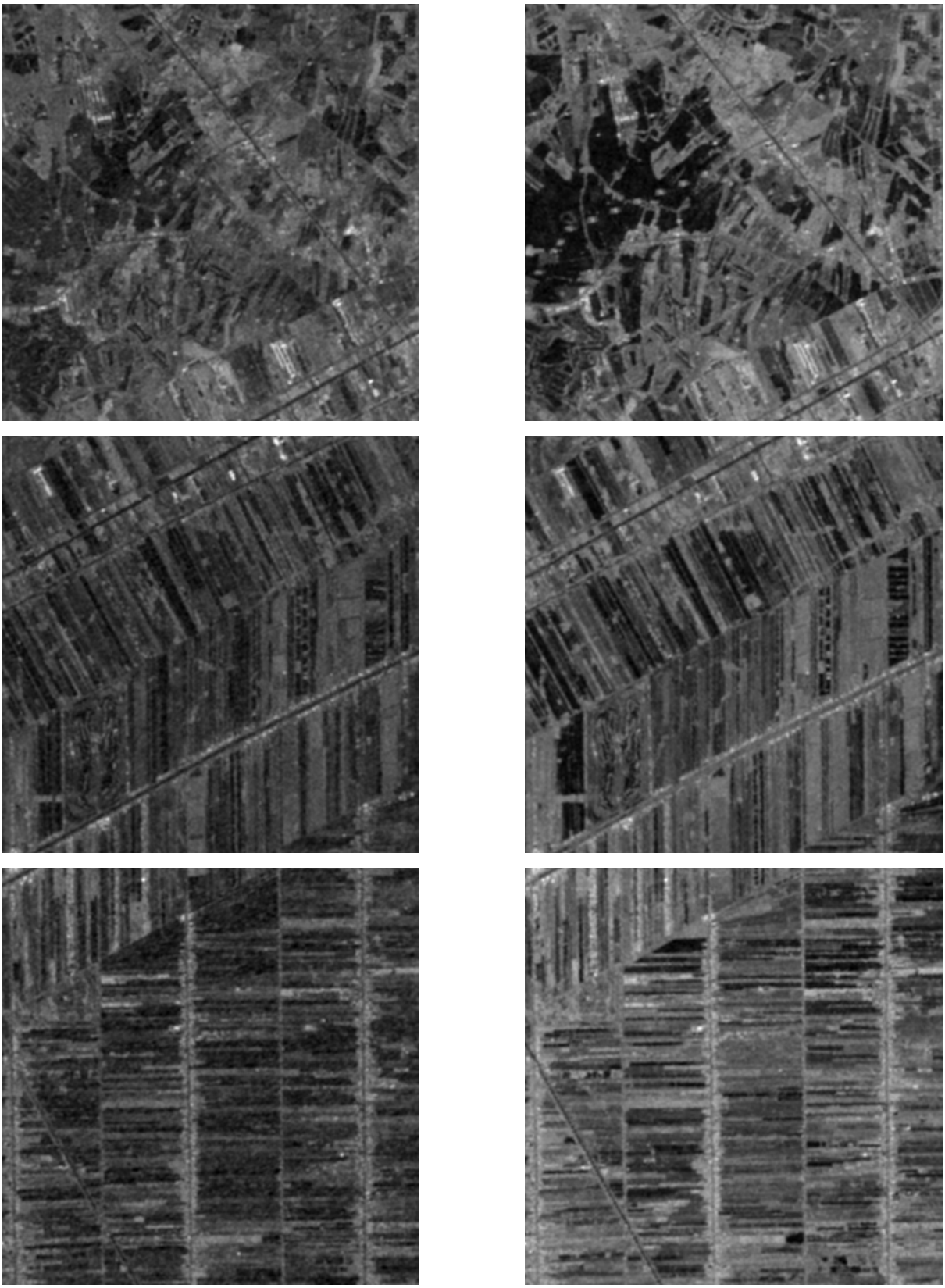



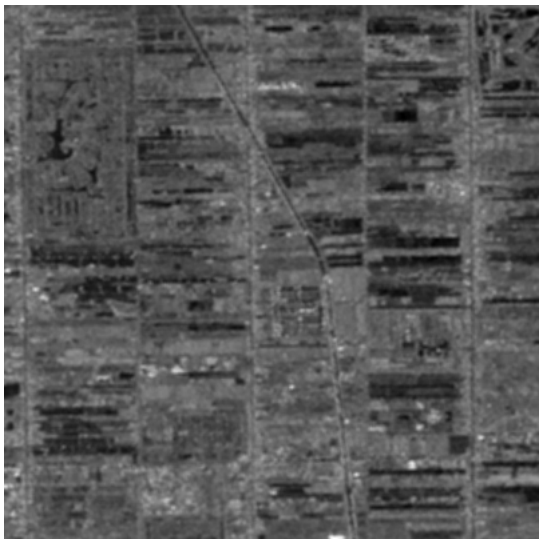


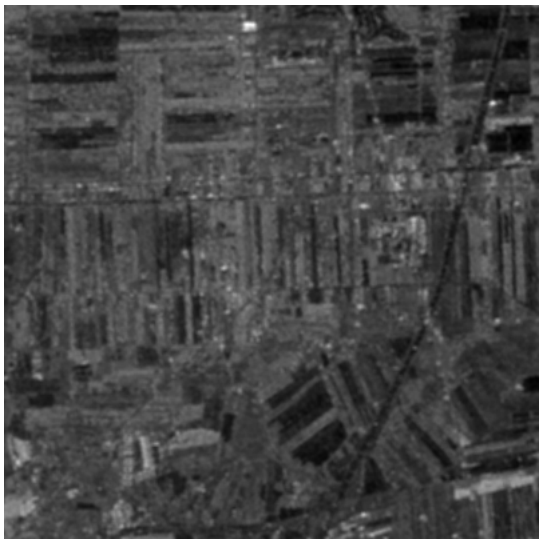







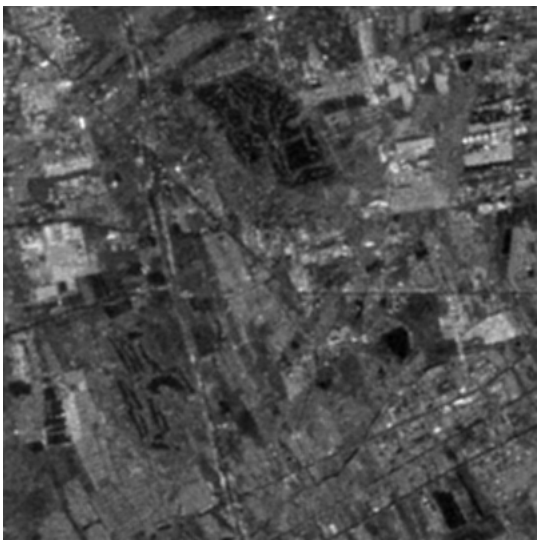
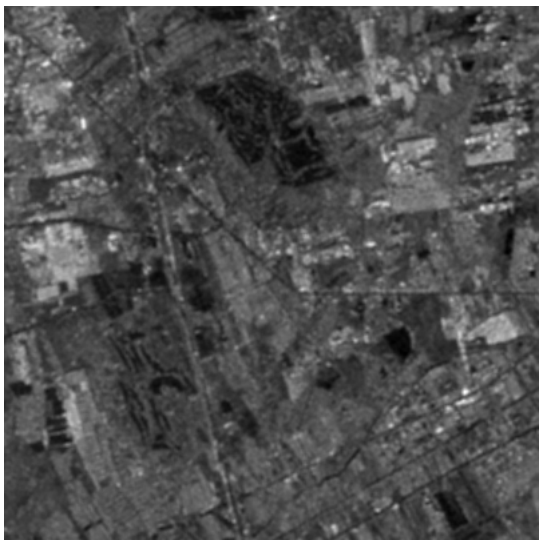





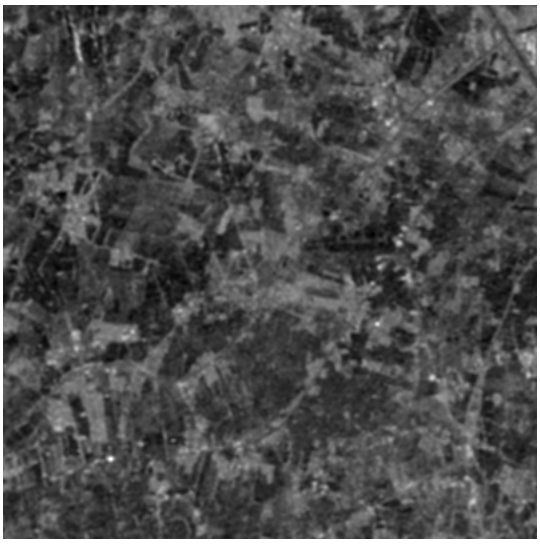
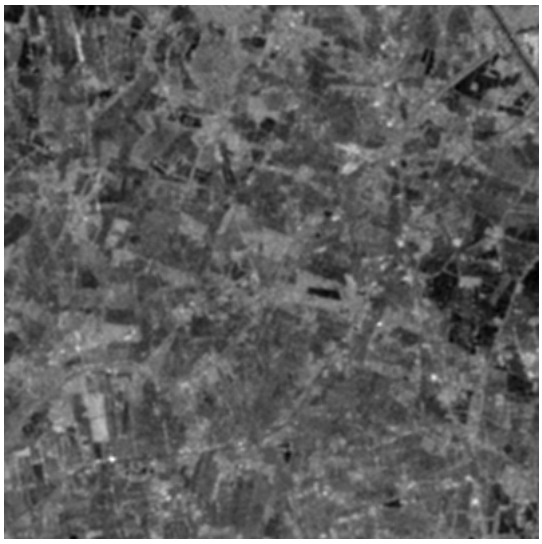
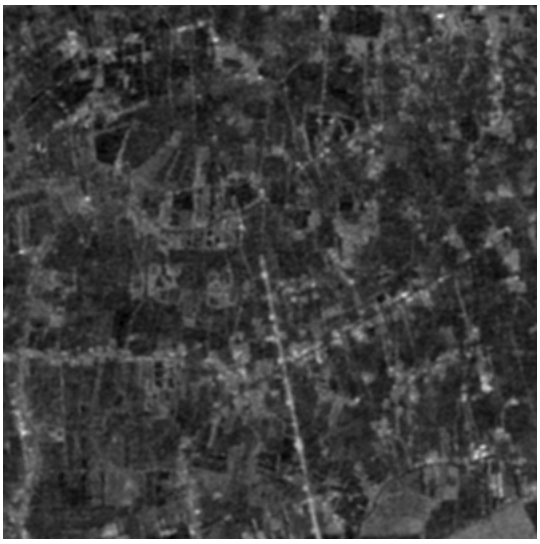
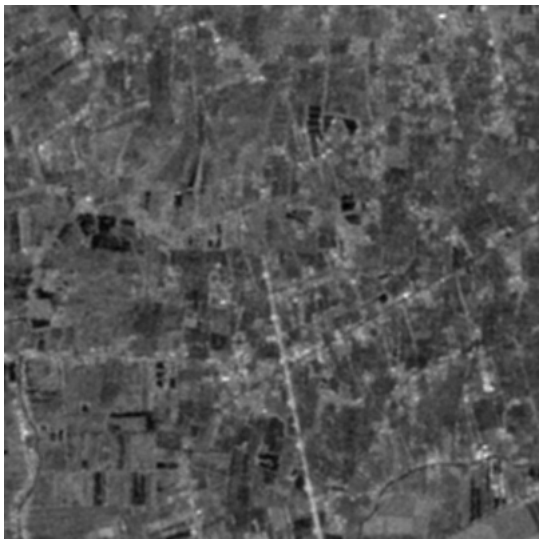
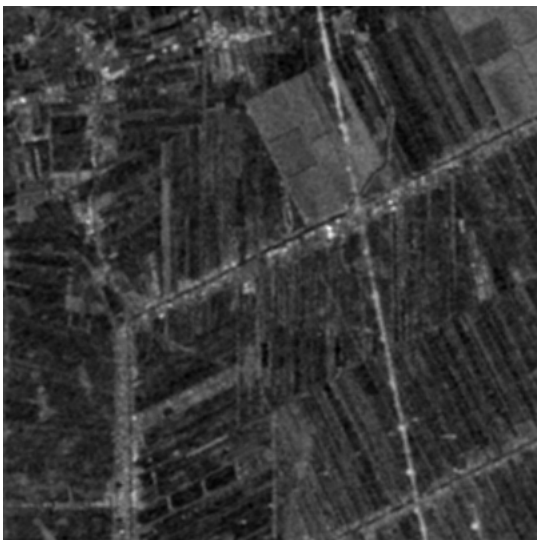
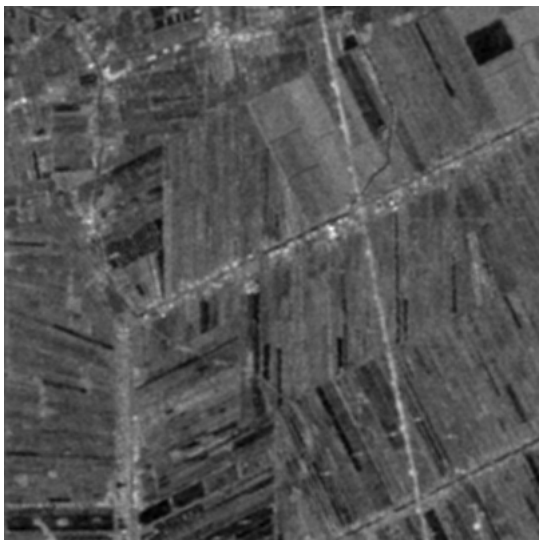


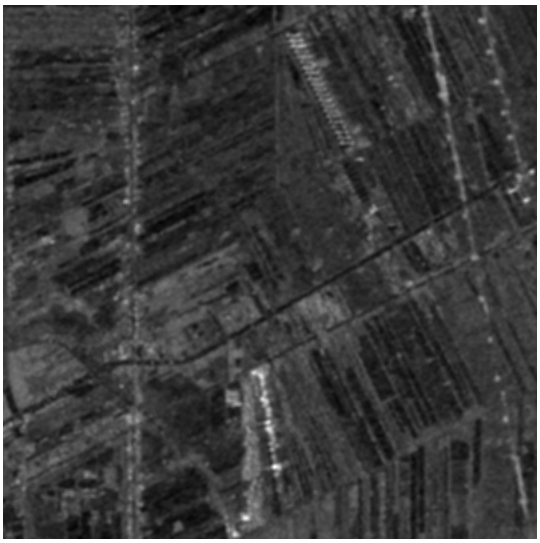
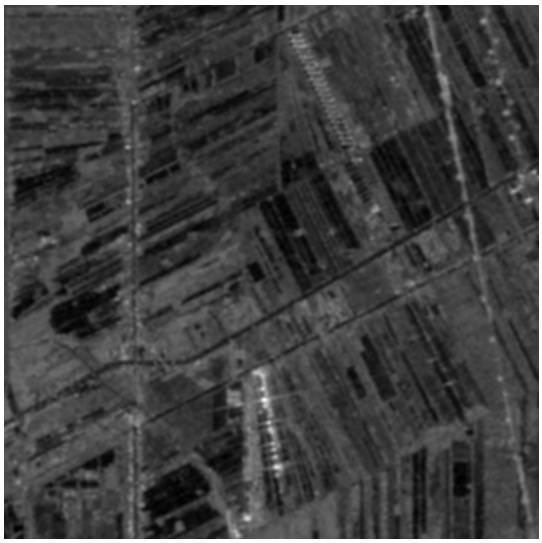
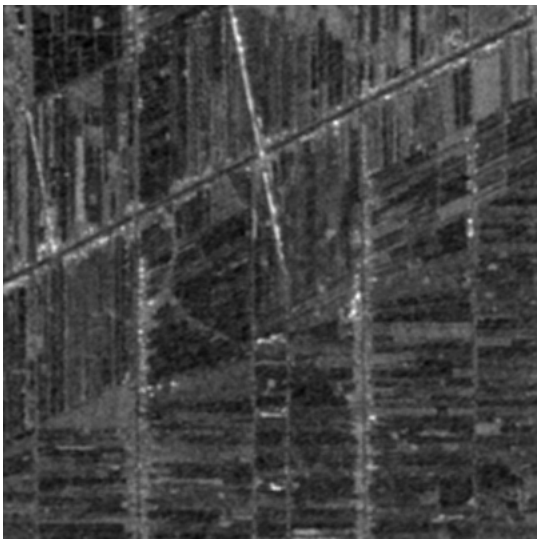



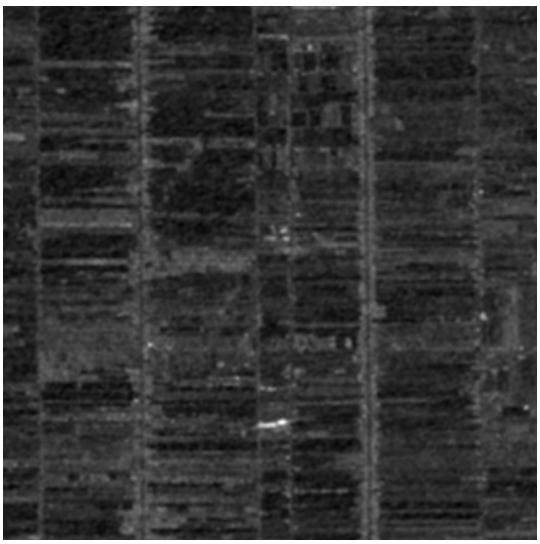
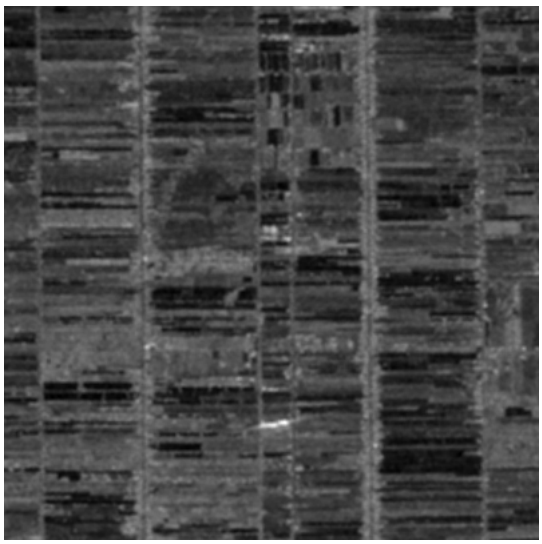


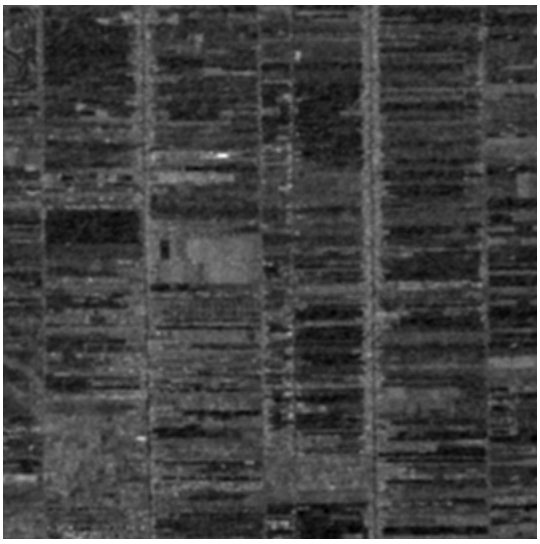



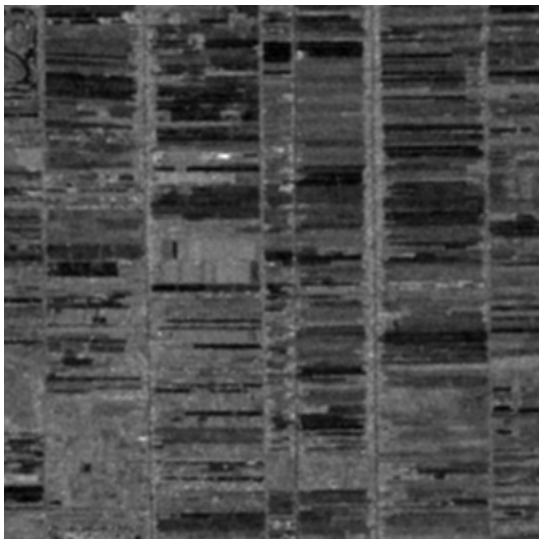




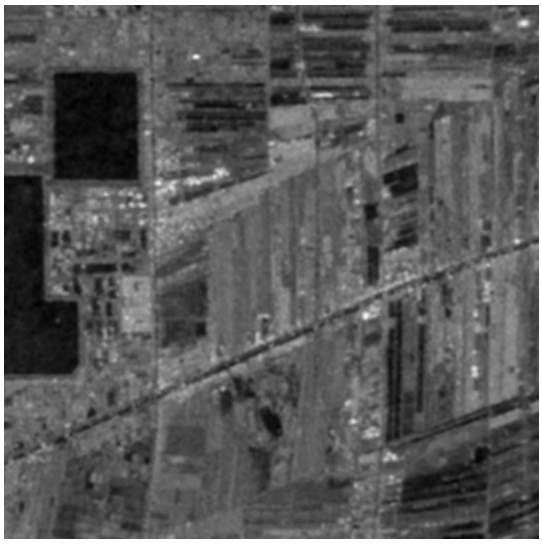







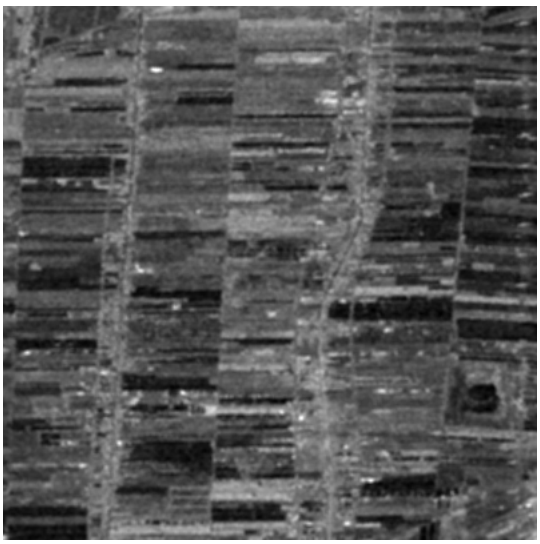




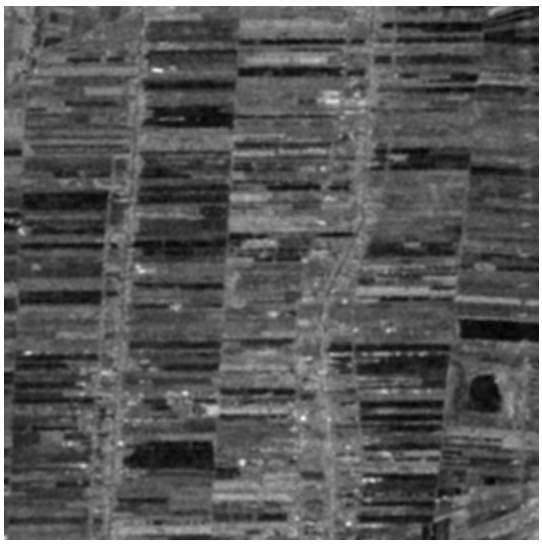



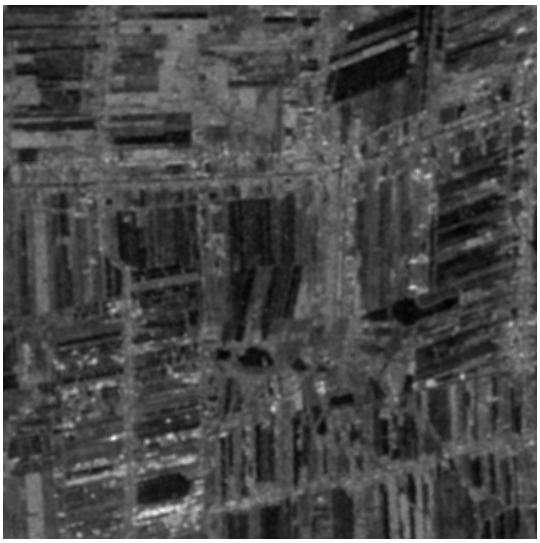
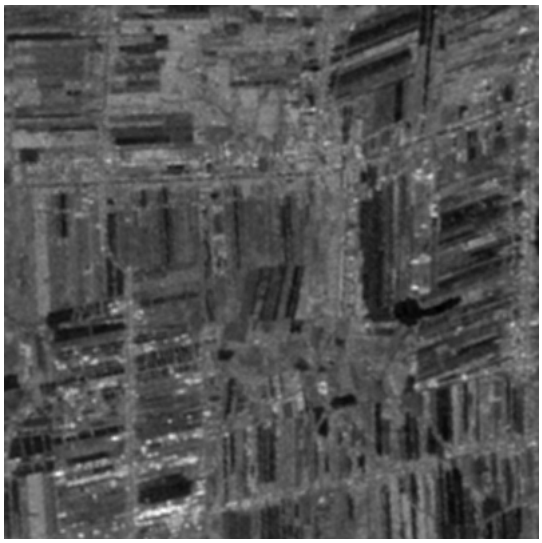


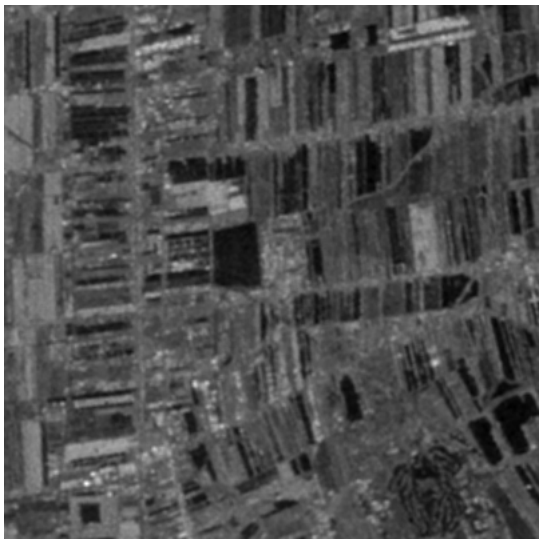
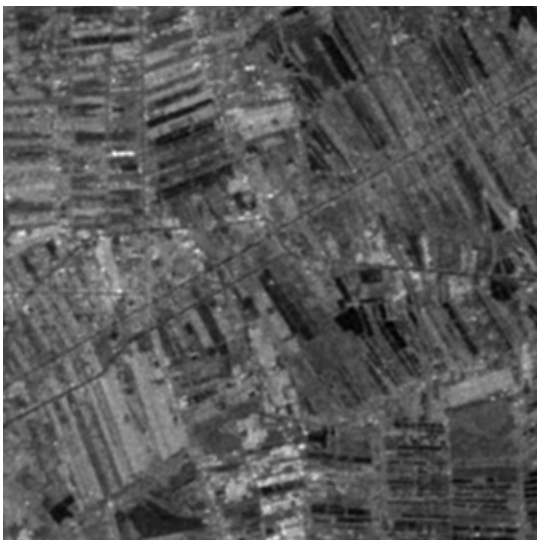
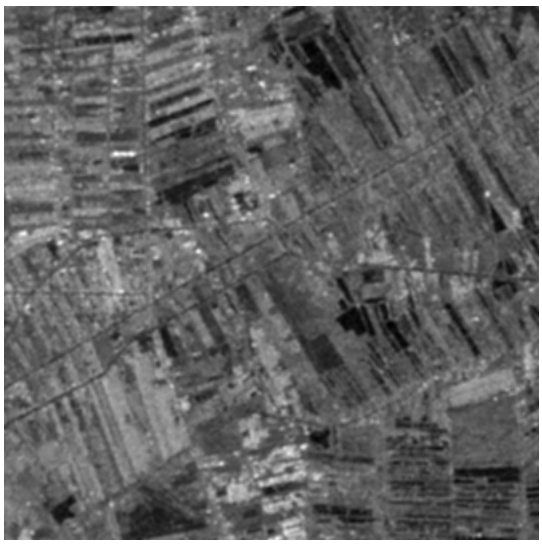


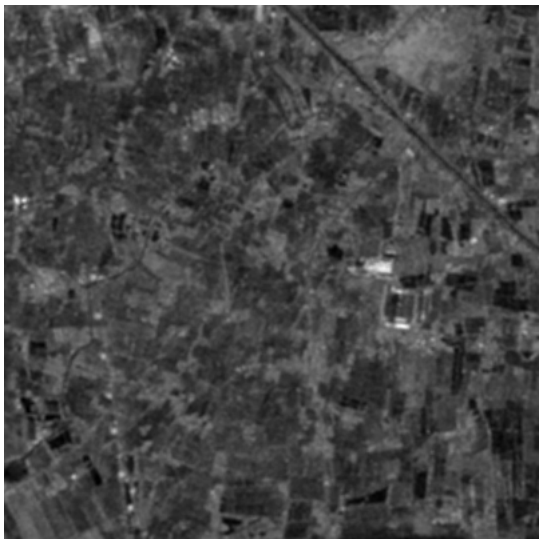


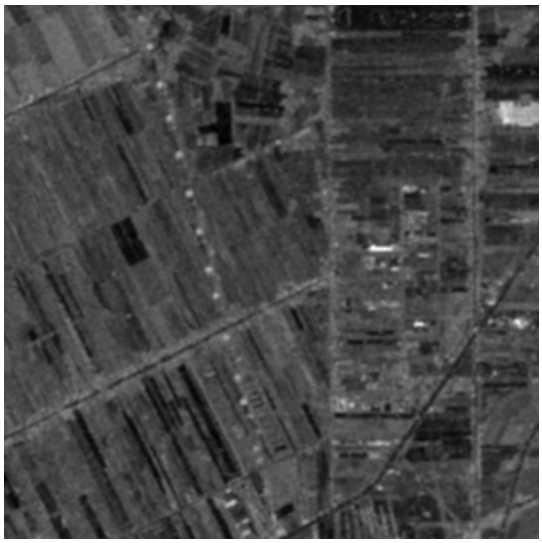
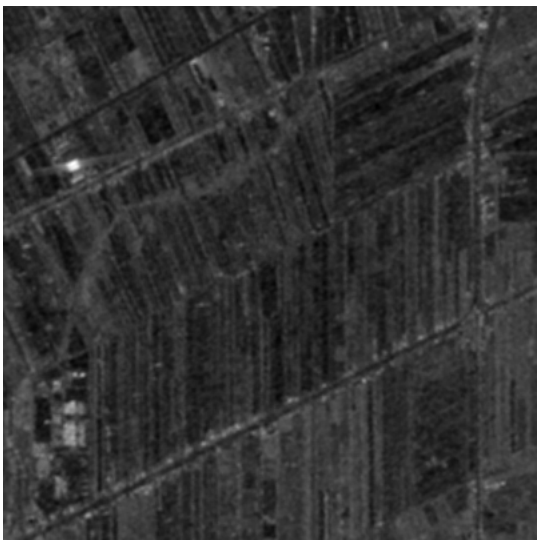
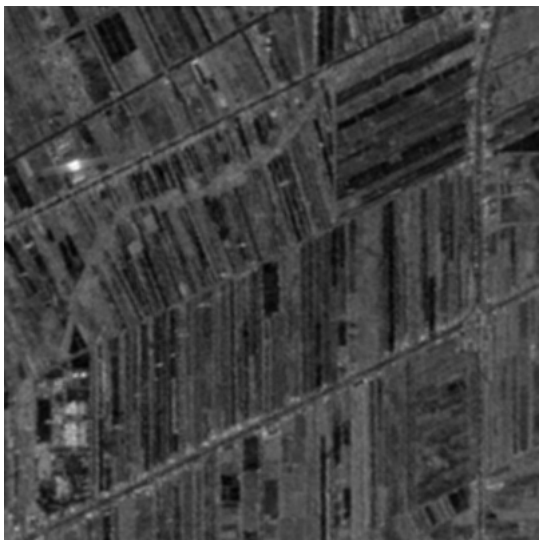


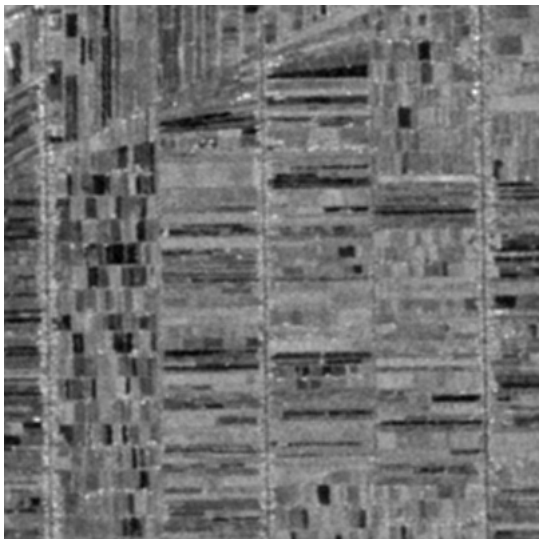
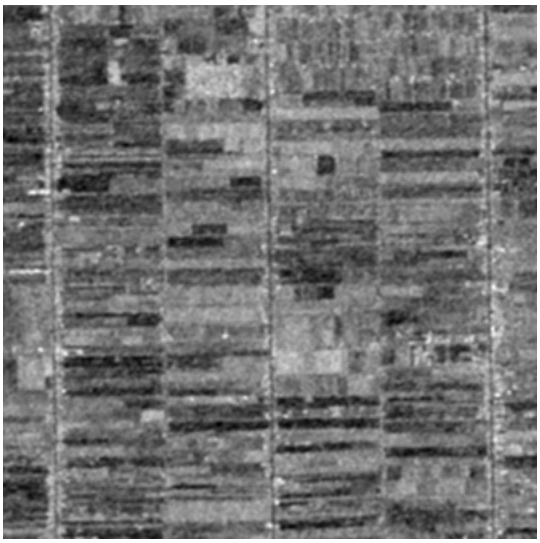
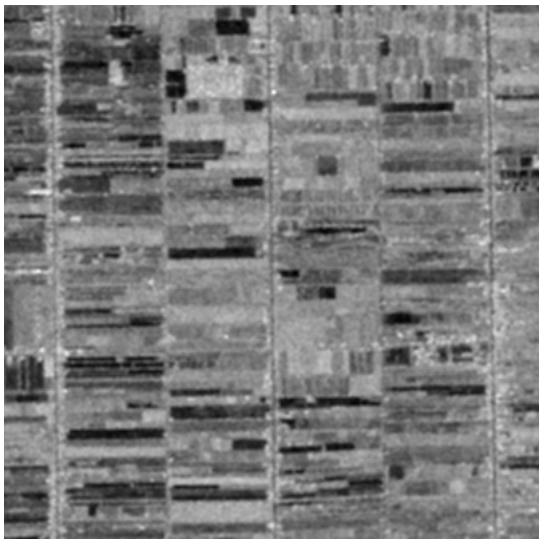
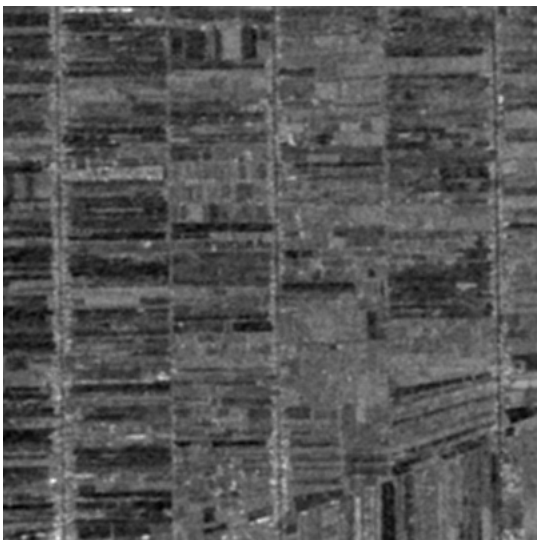
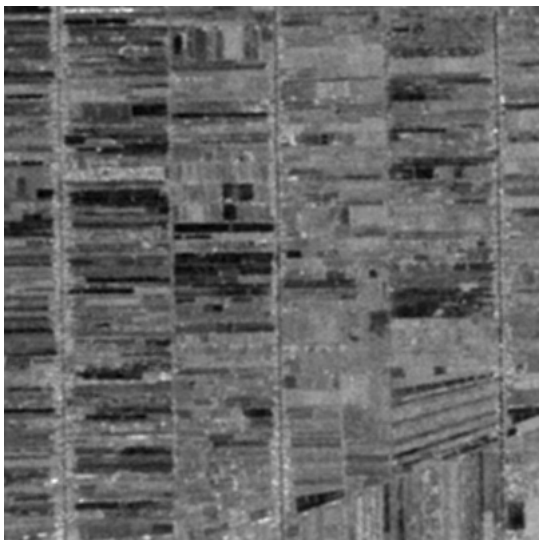


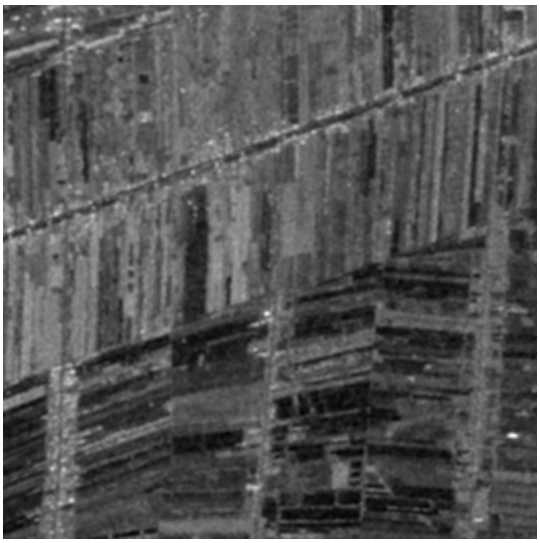
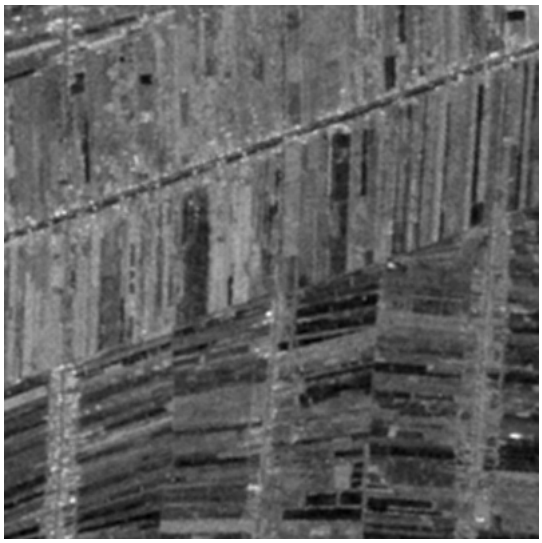



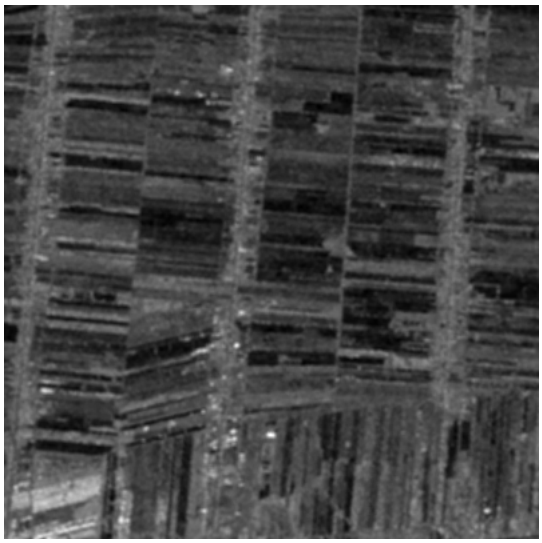
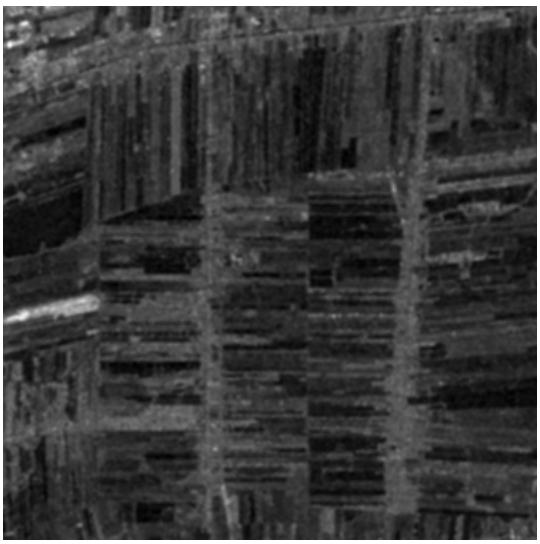
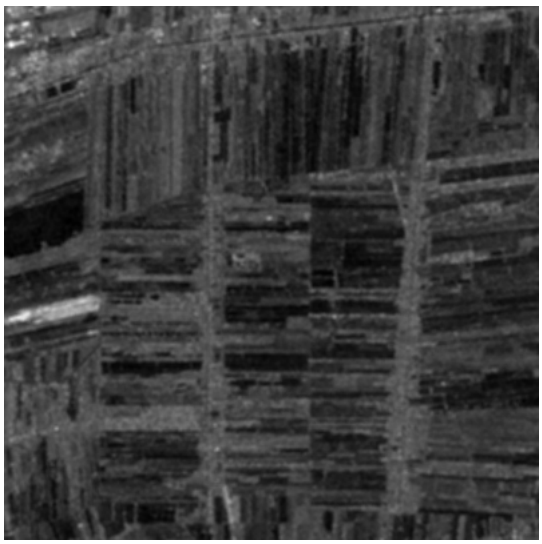




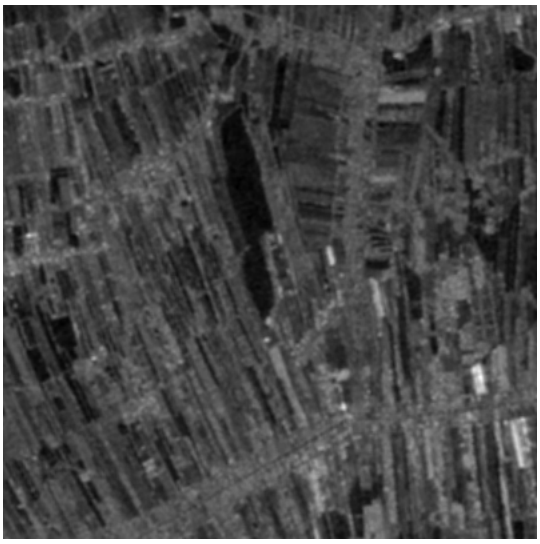
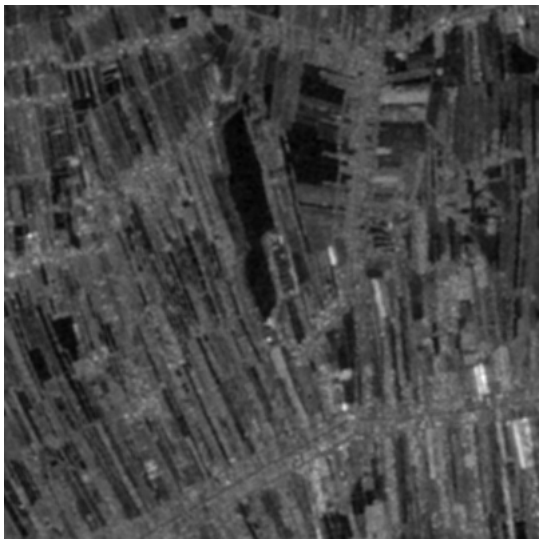
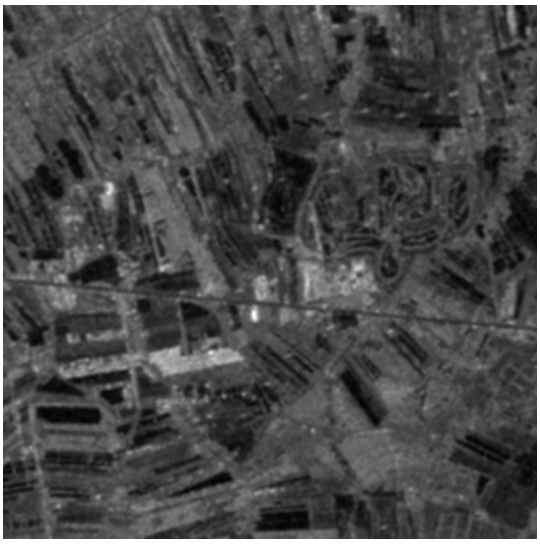
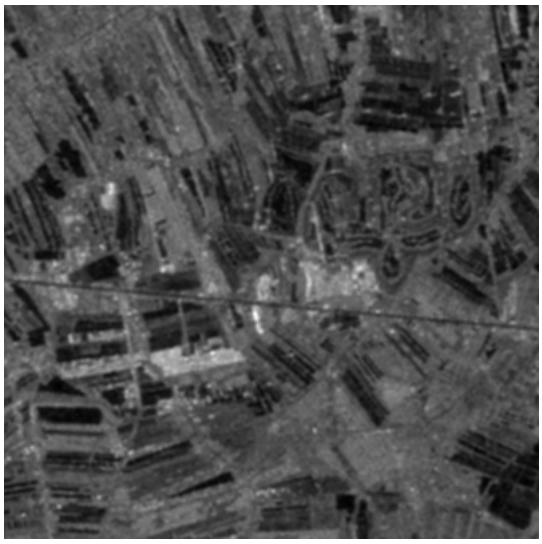



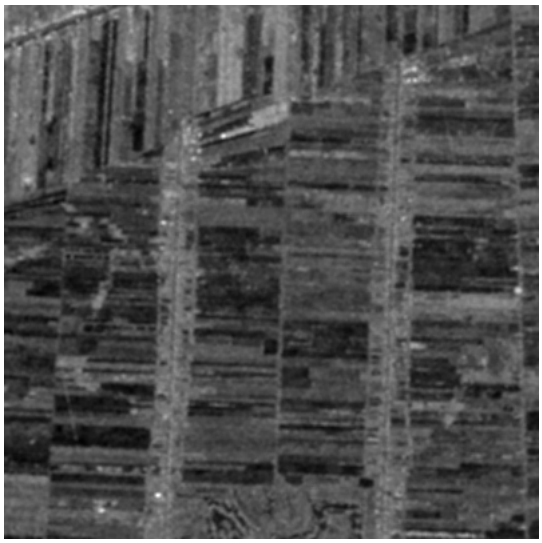
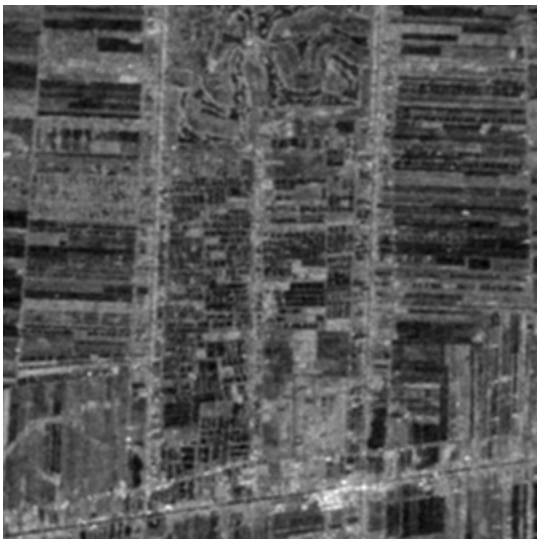
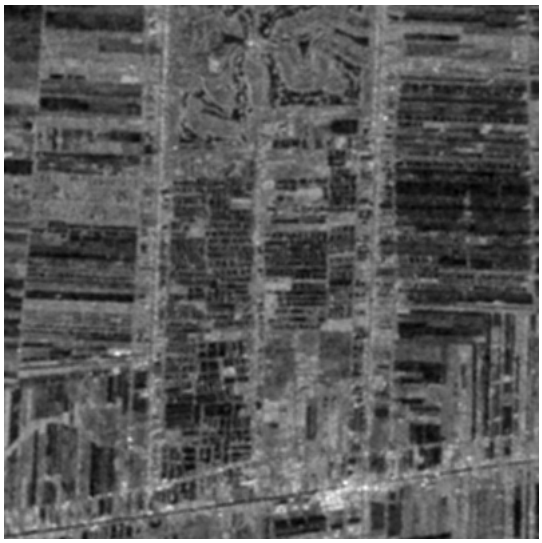
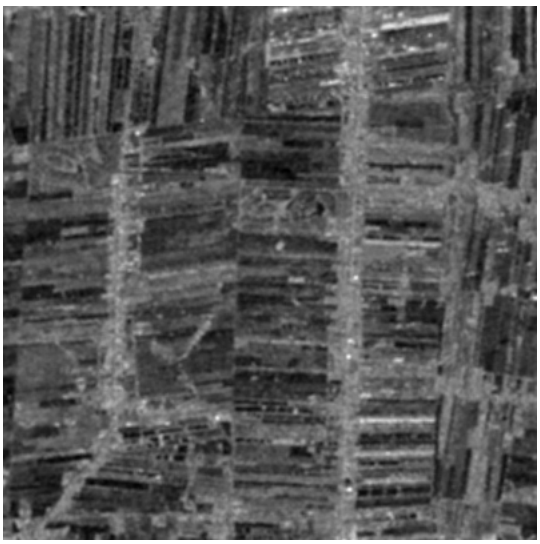
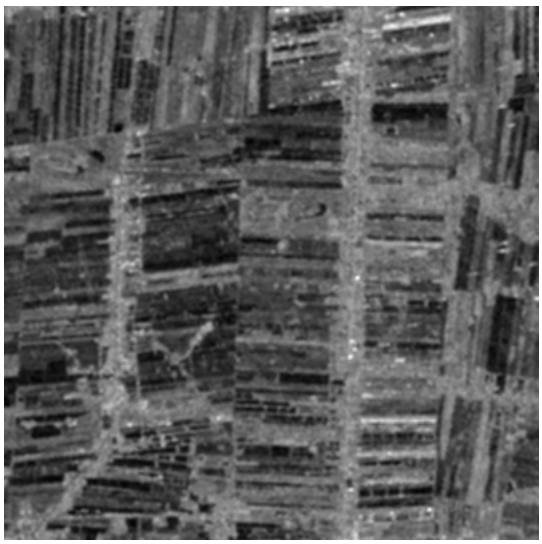




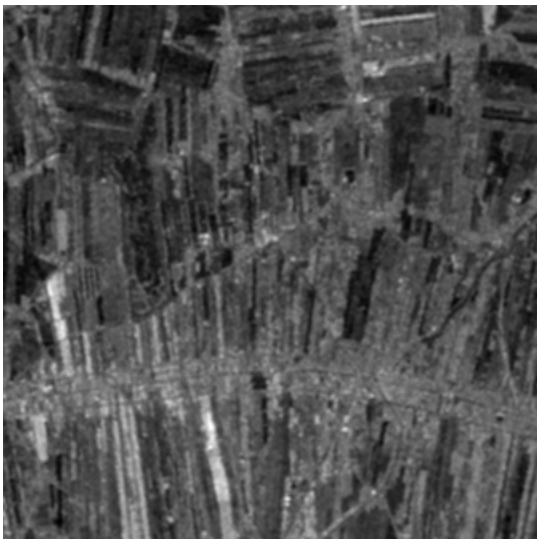







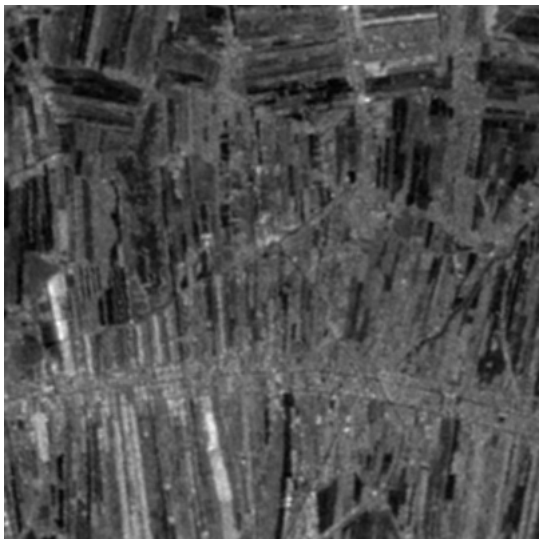



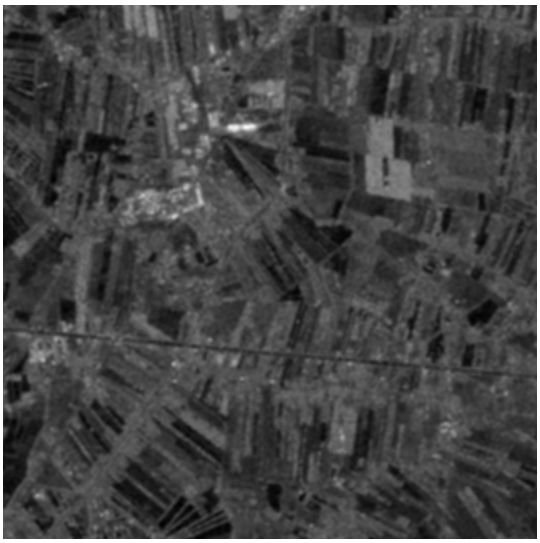




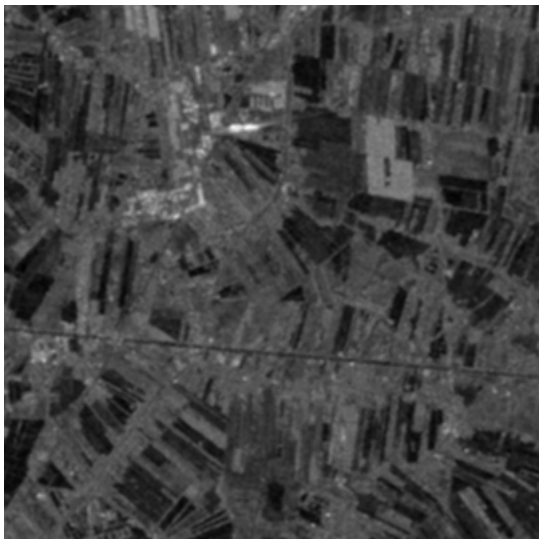







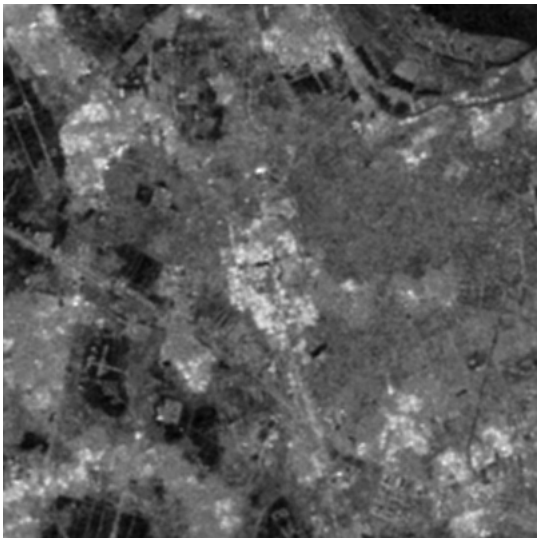


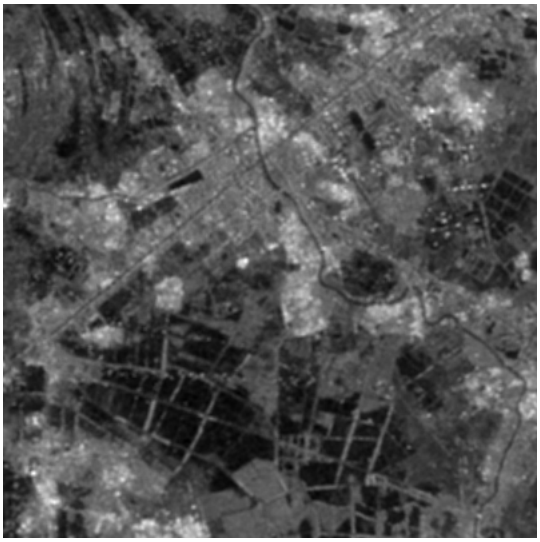




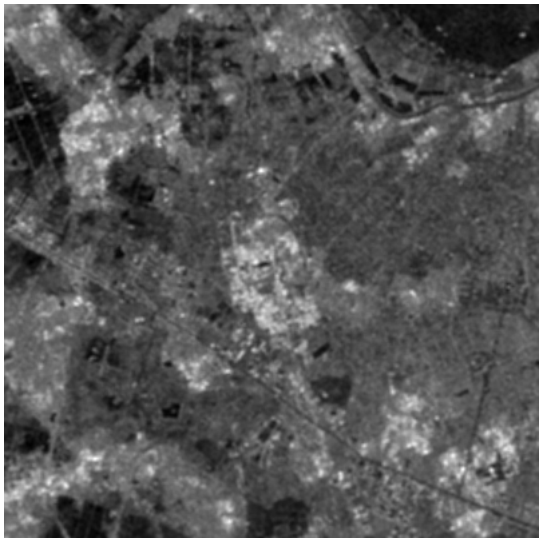
Hanoi area

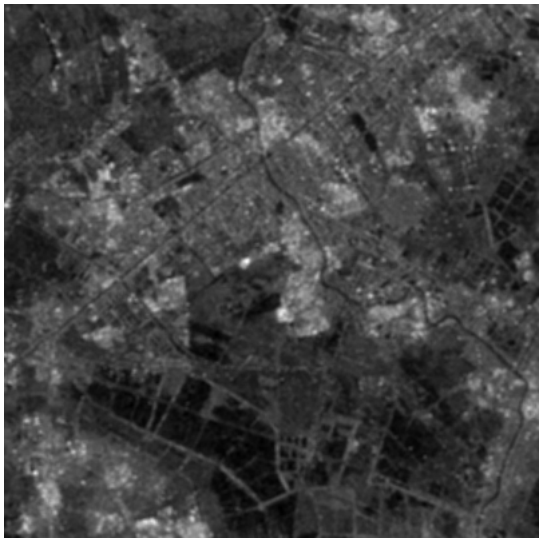
Time 1

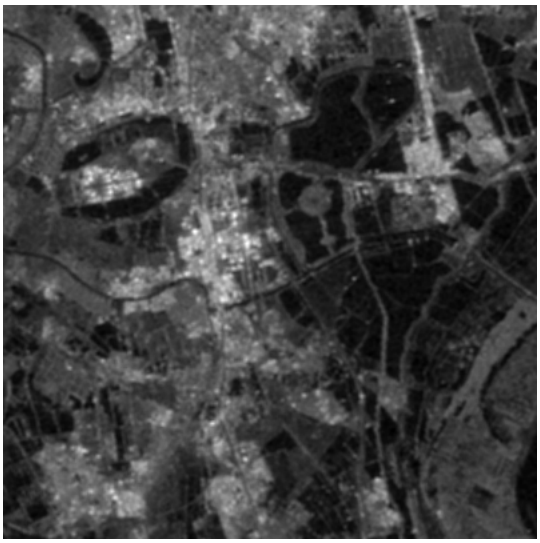


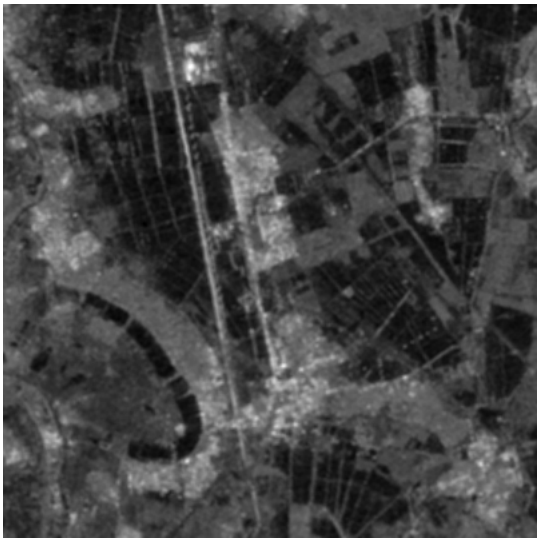
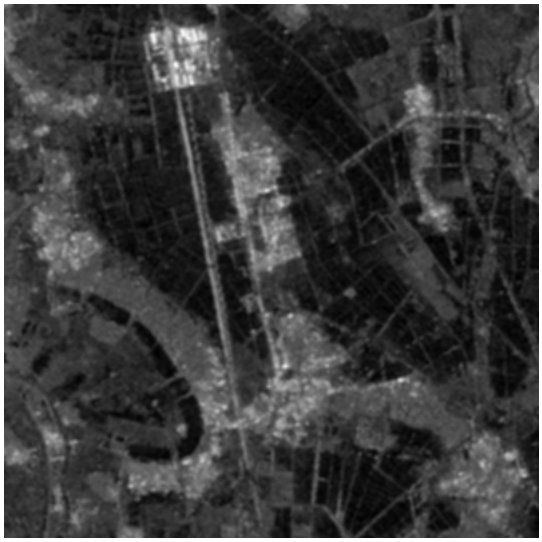
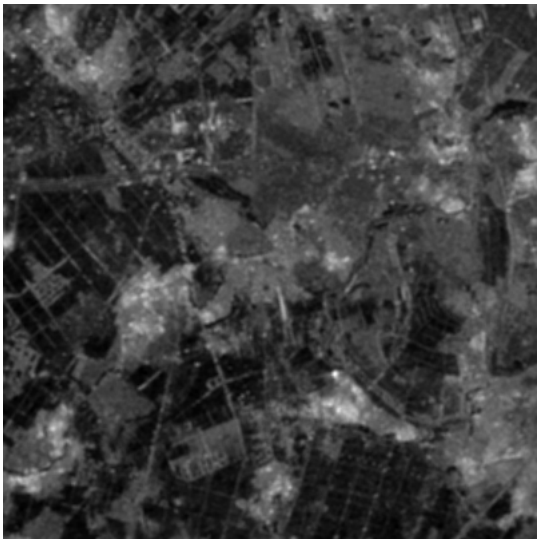
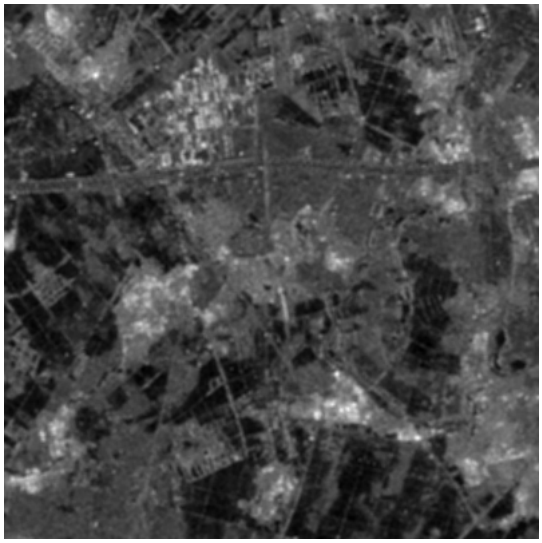


Time 2



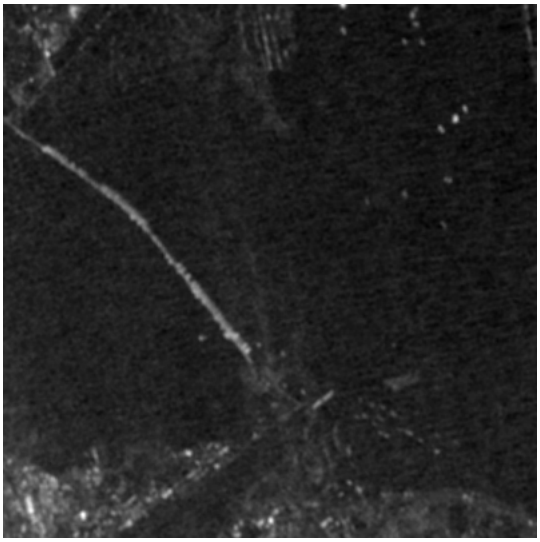


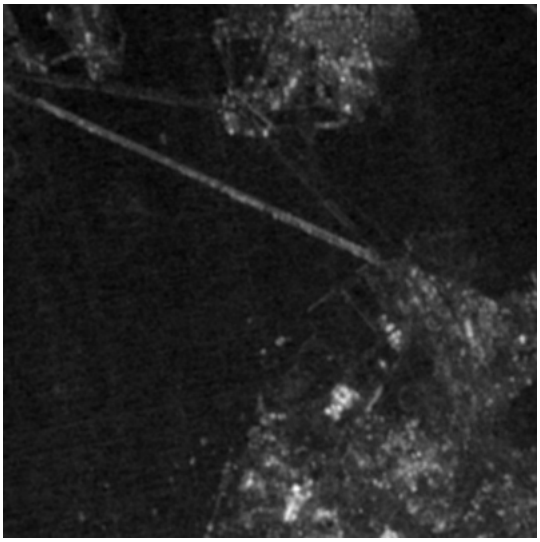


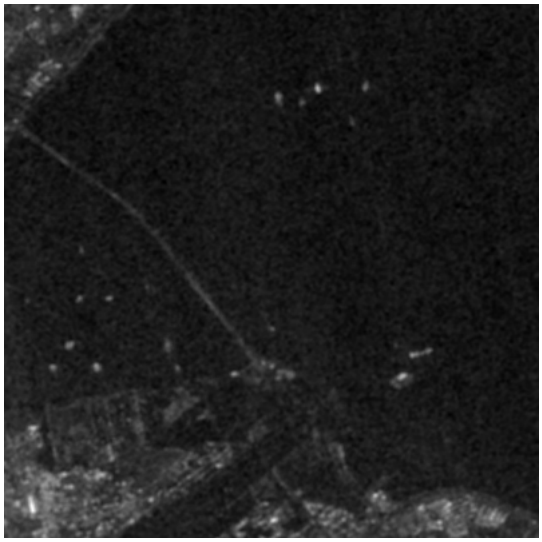
Xiamen area

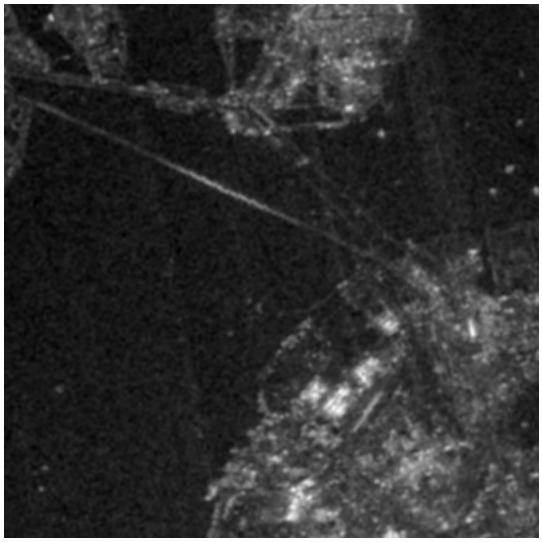
Time 1

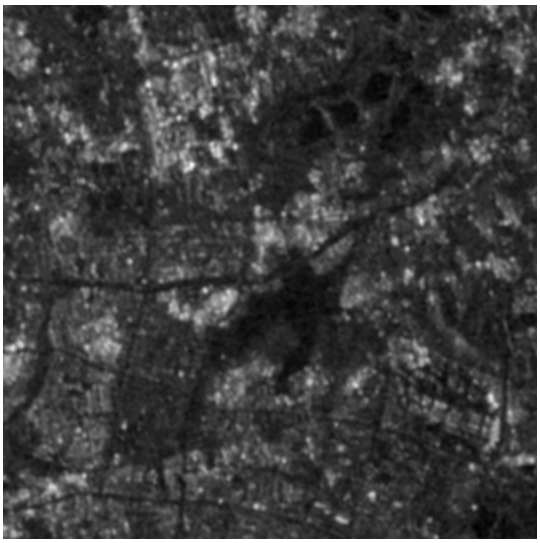
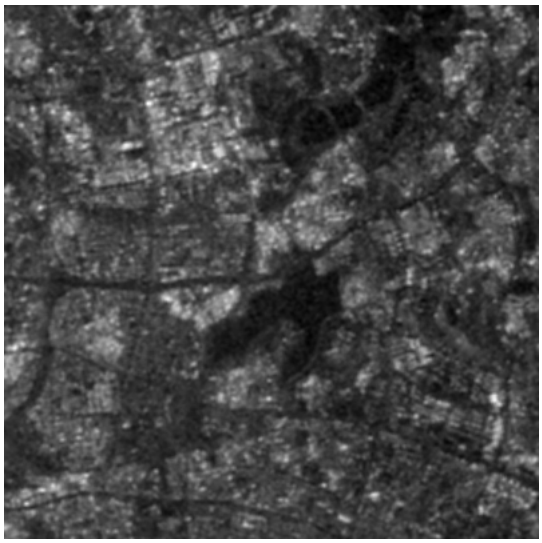
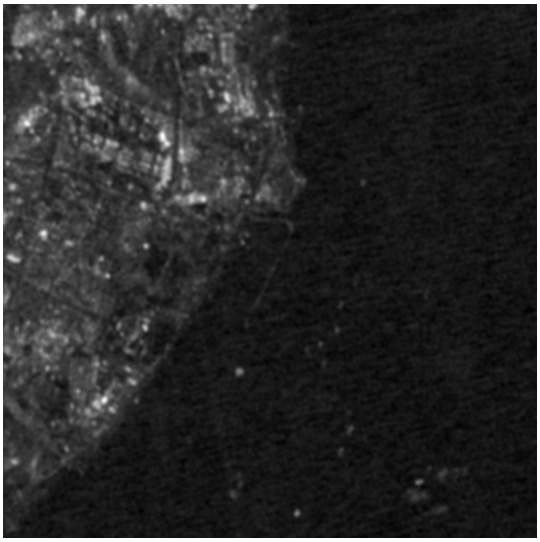
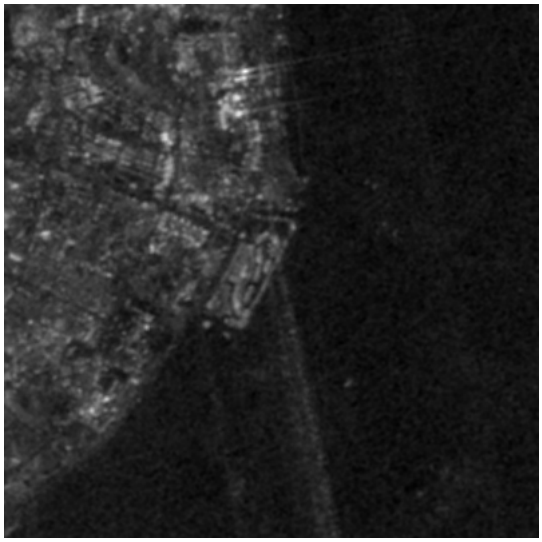
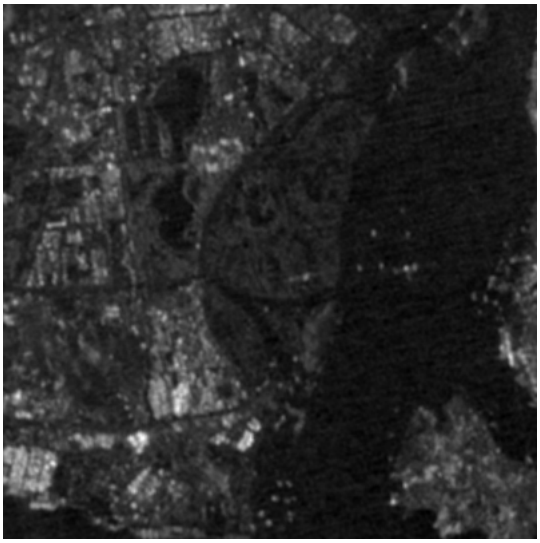
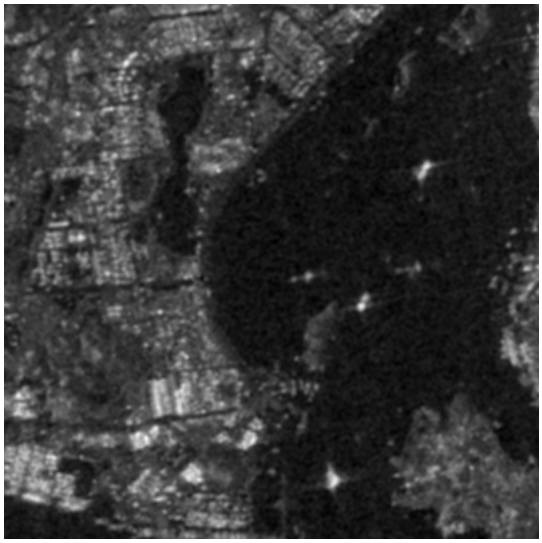




Time 2

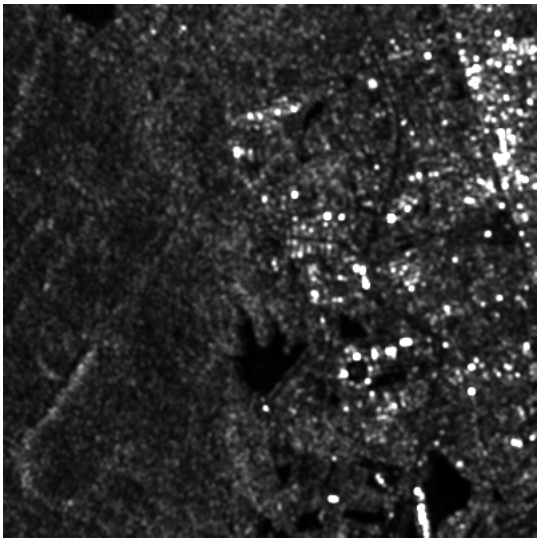




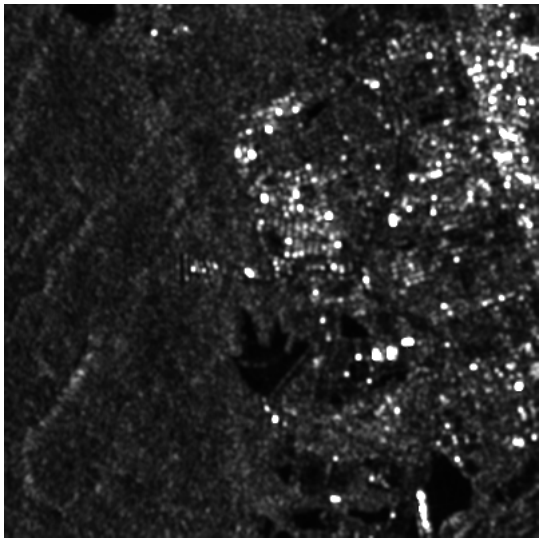







Chiang Mai area

Time 1



Time 2



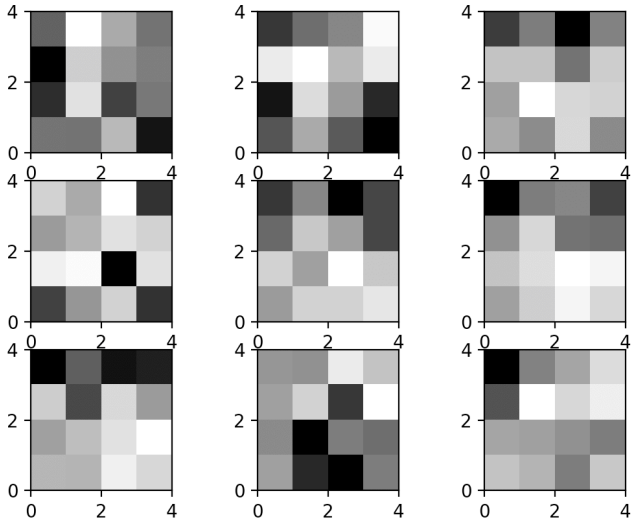
Appendix-C

Filter obtained from trained model

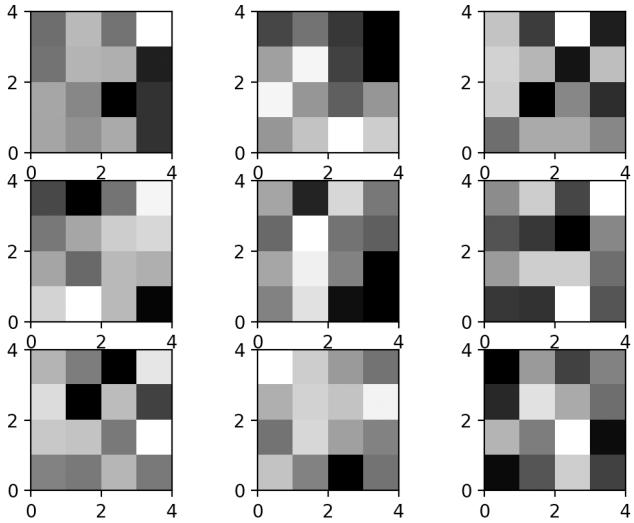
This section contains samples of the filters obtained from training of the U-net in Figure 4.2 and network proposed in this thesis (CORN) in Figure 5.6. In each filter, image axis represents the pixel number.

U-net

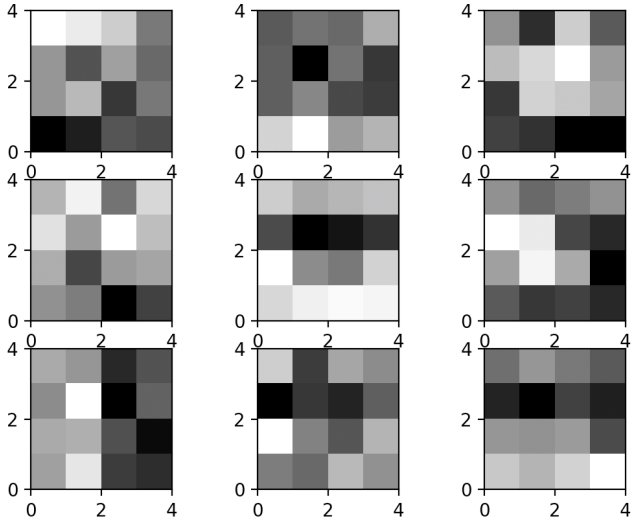
Encoder 1



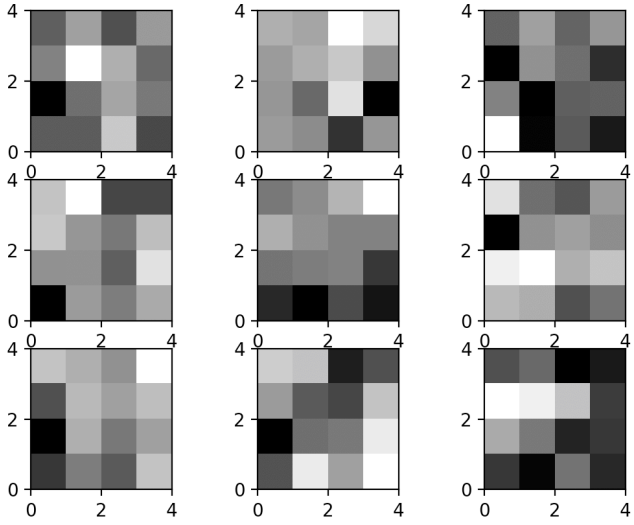
Encoder 2



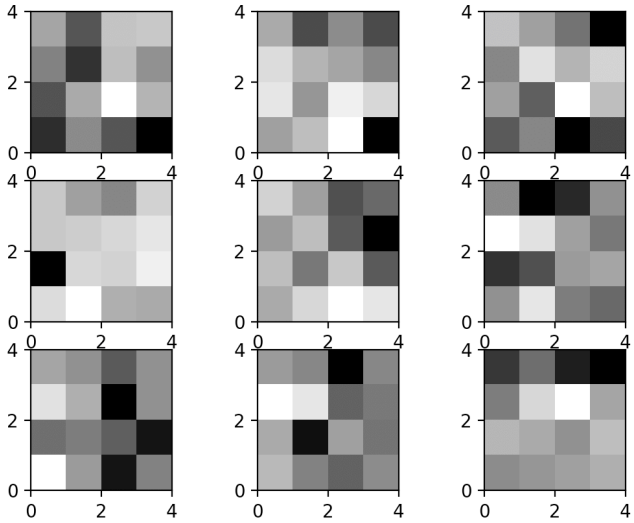
Encoder 3



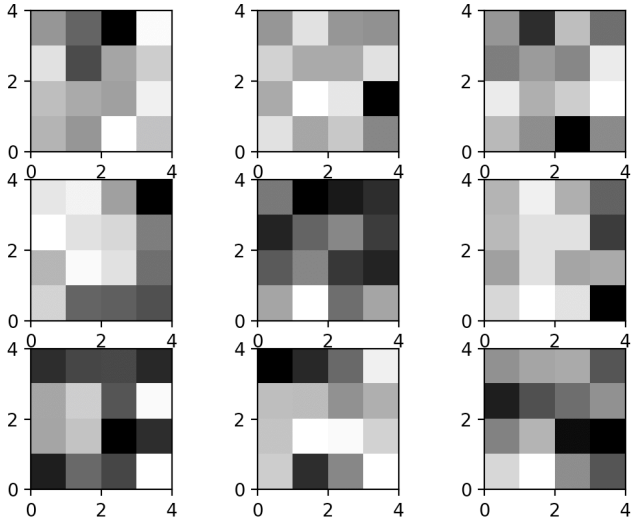
Encoder 4



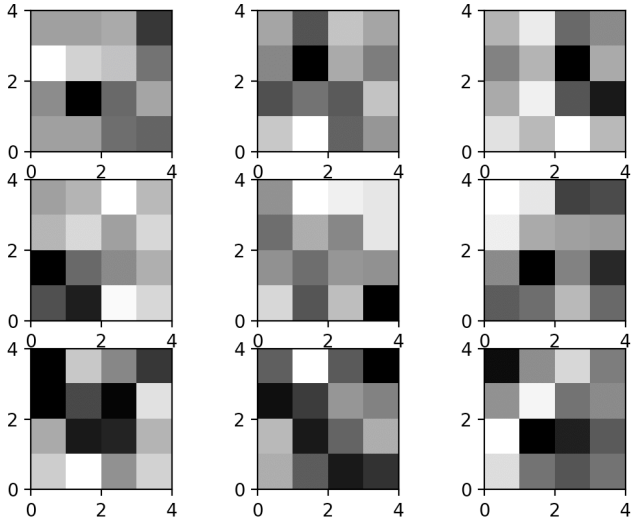
Encoder 5



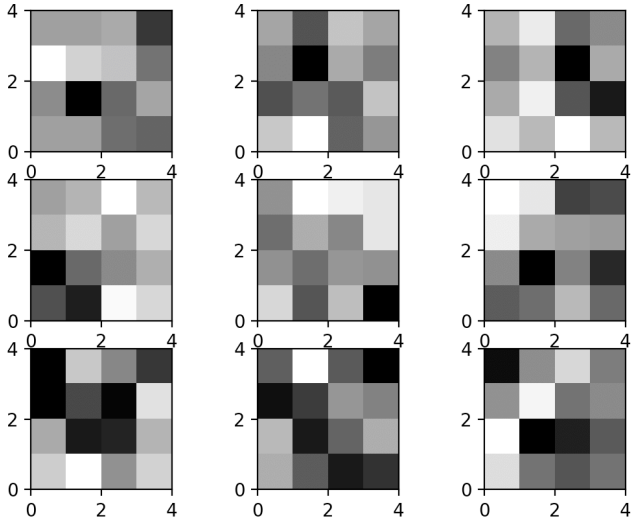
Encoder6



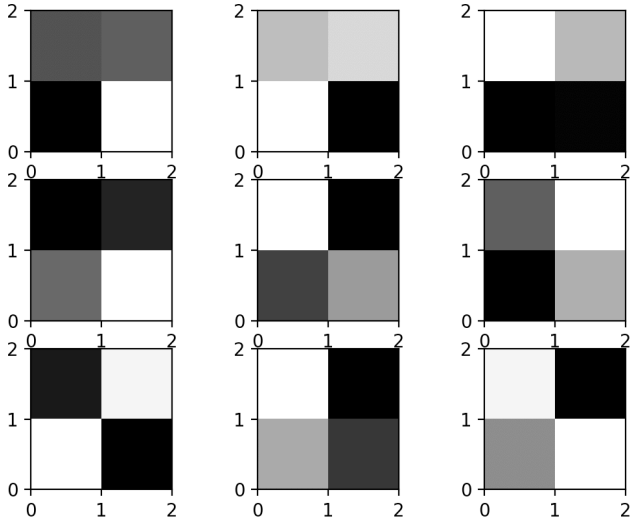
Encoder 7



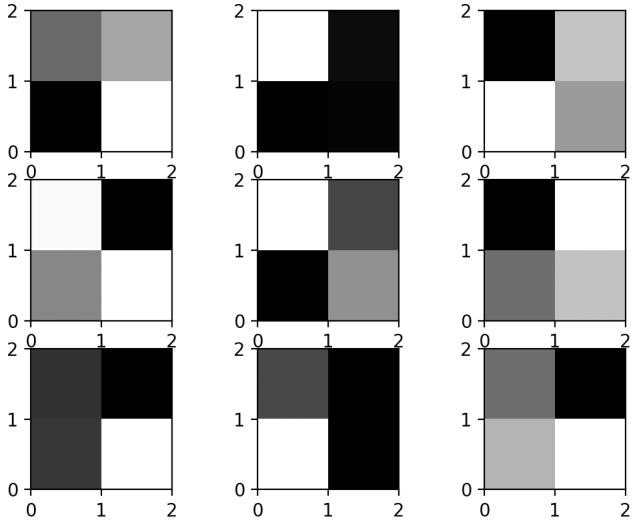
Encoder 8



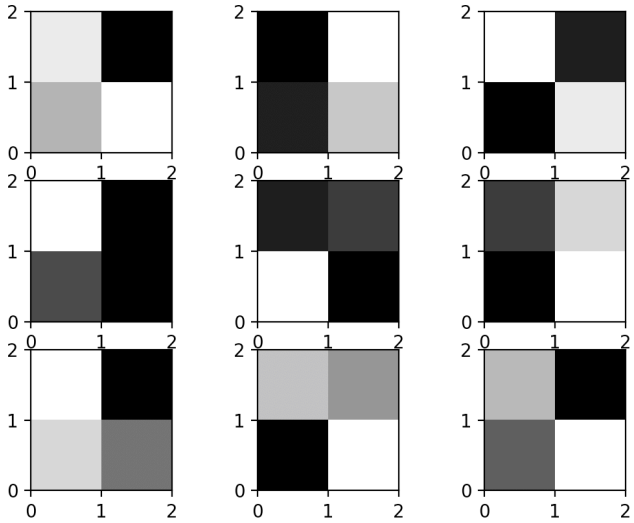
Decoder 1



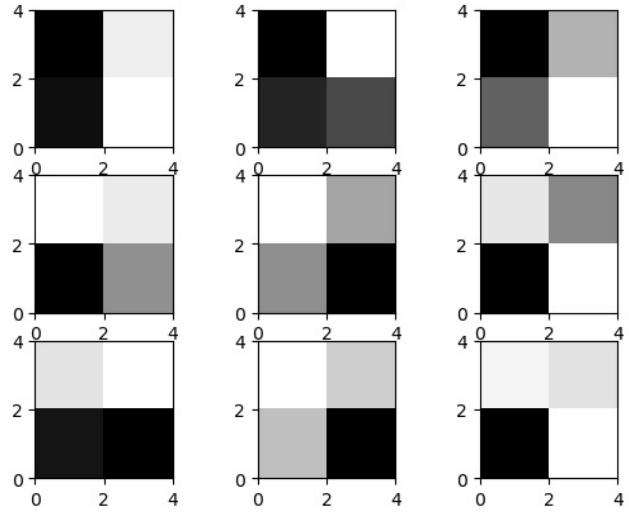
Decoder 2



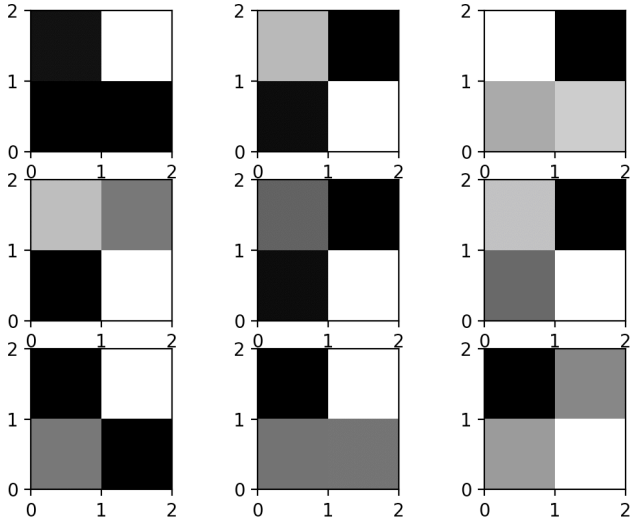
Decoder 3



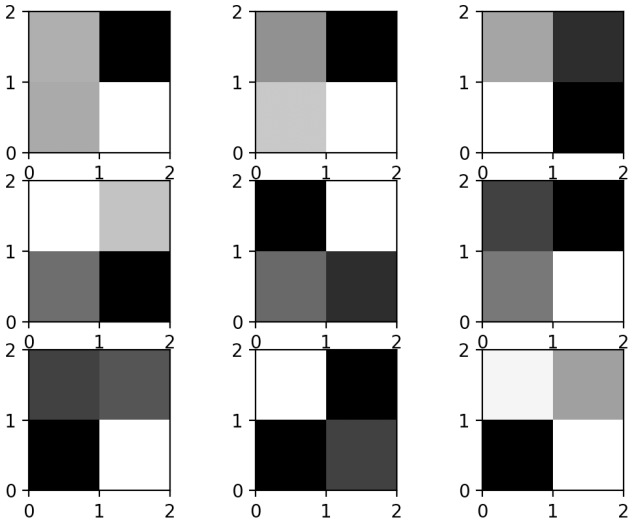
Decoder 4



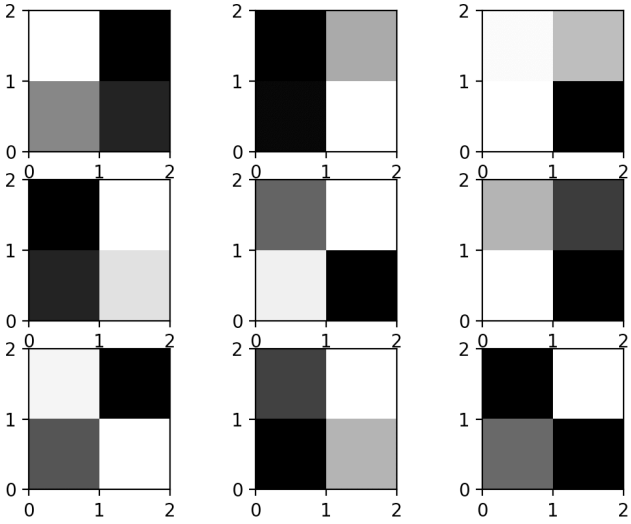
Decoder 5



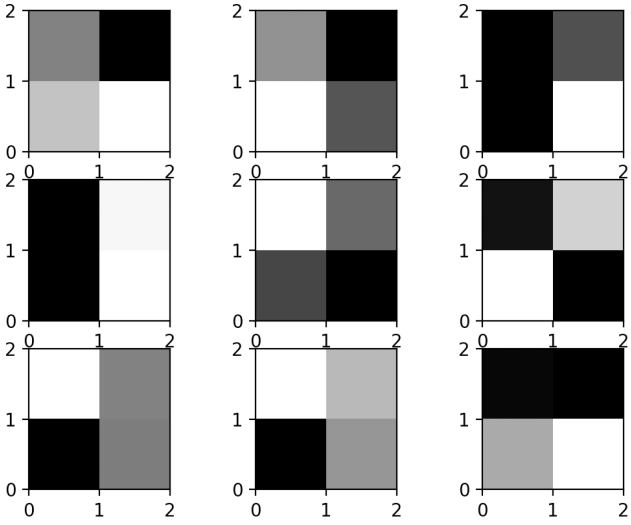
Decoder 6



Decoder 7

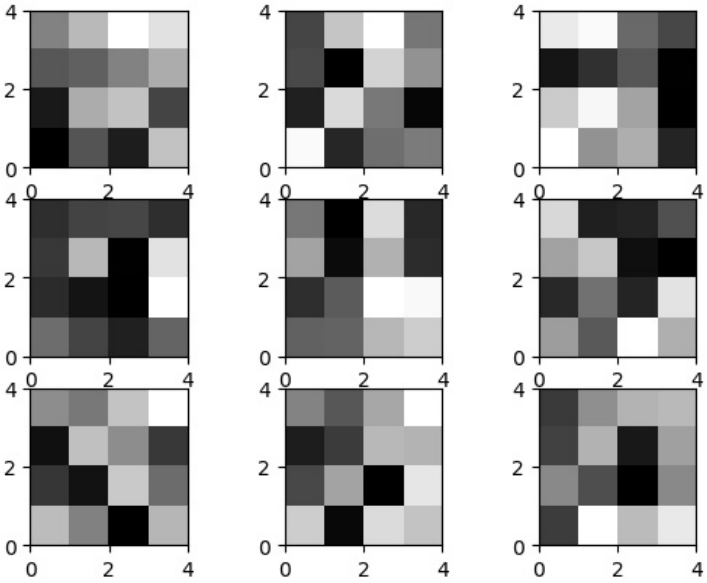


Decoder 8

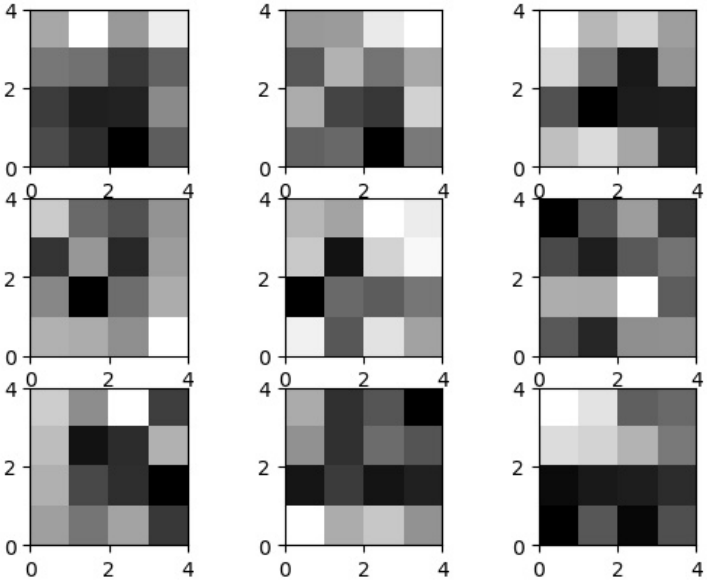


CORN

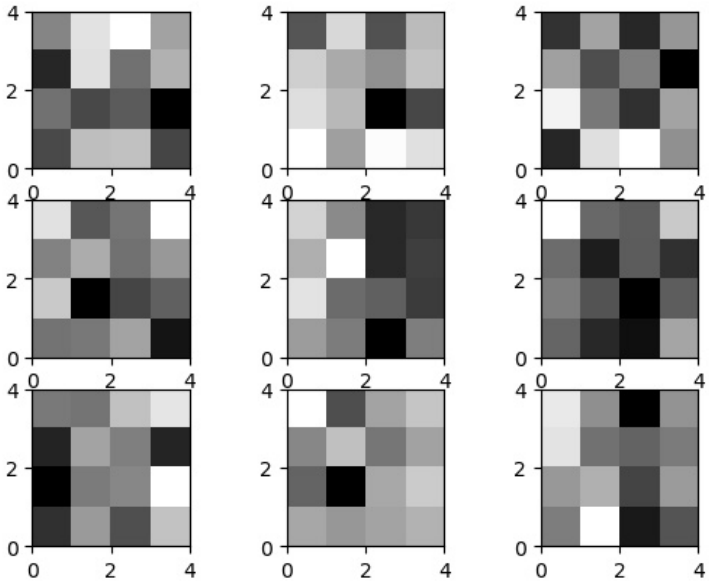
Encoder 1 (Time 1 – Time 2)



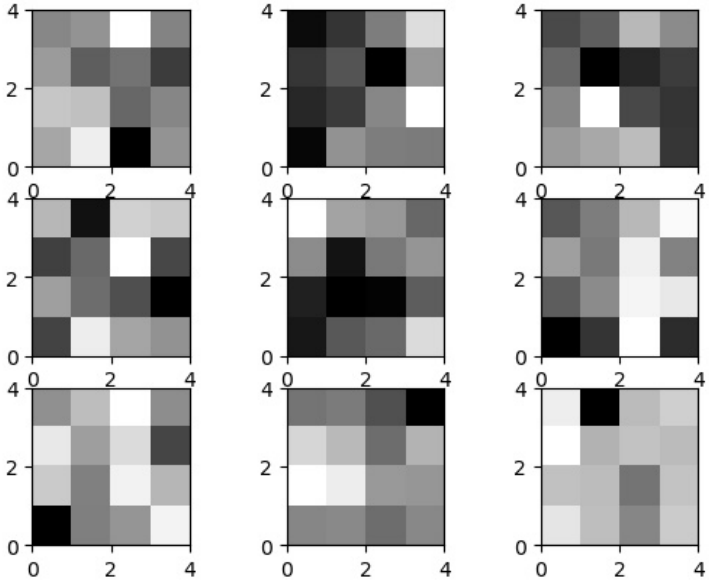
Encoder 2 (Time 1 – Time 2)



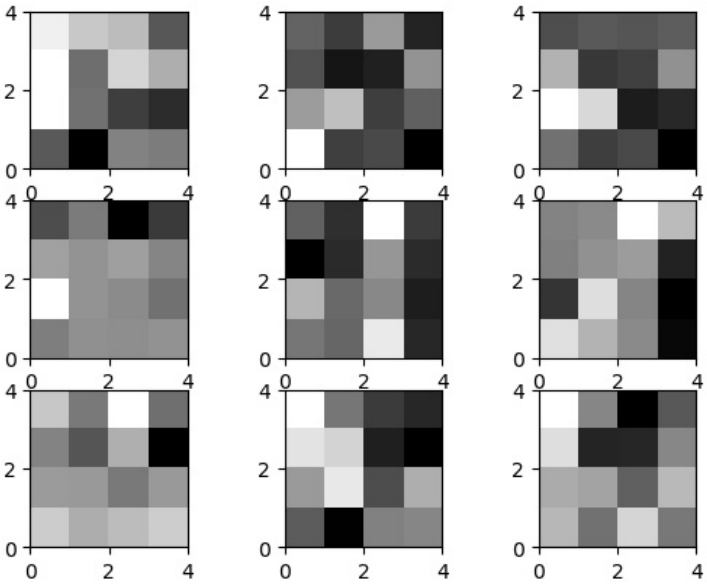
Encoder 3 (Time 1 – Time 2)



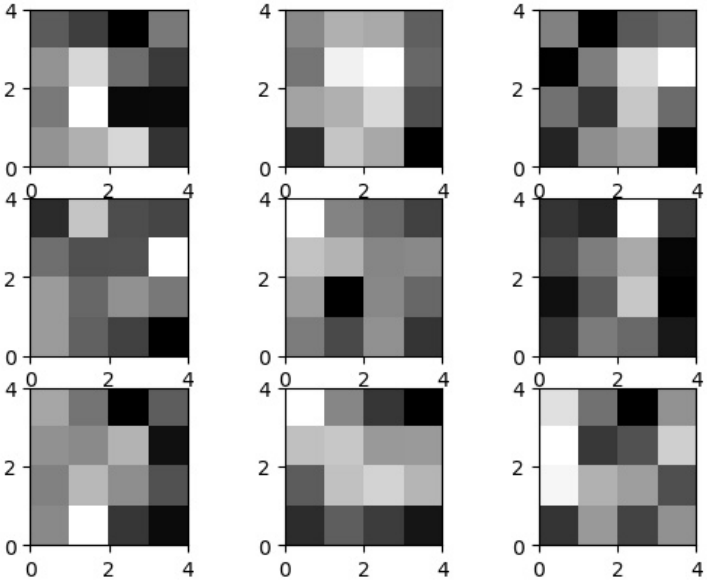
Encoder 4 (Time 1 – Time 2)



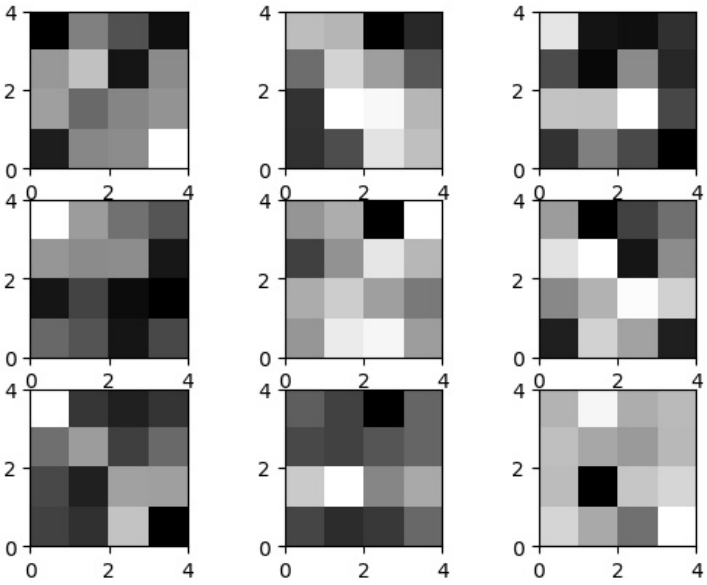
Encoder 5 (Time 1 – Time 2)



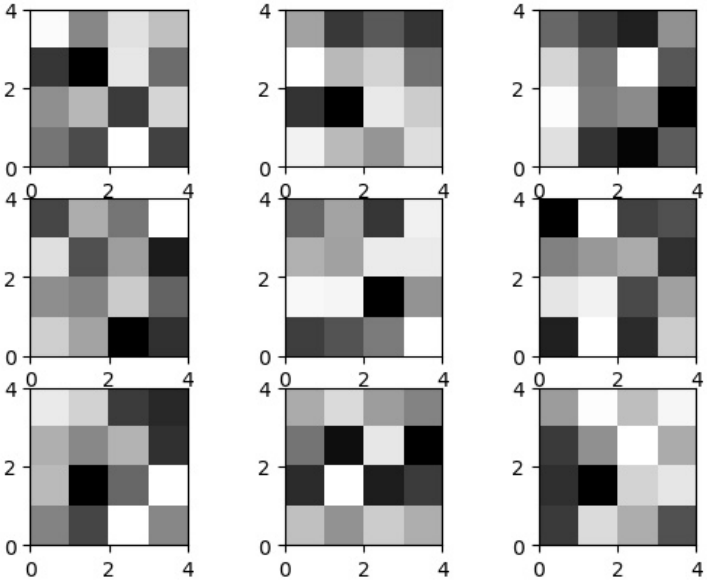
Encoder 6 (Time 1 – Time 2)



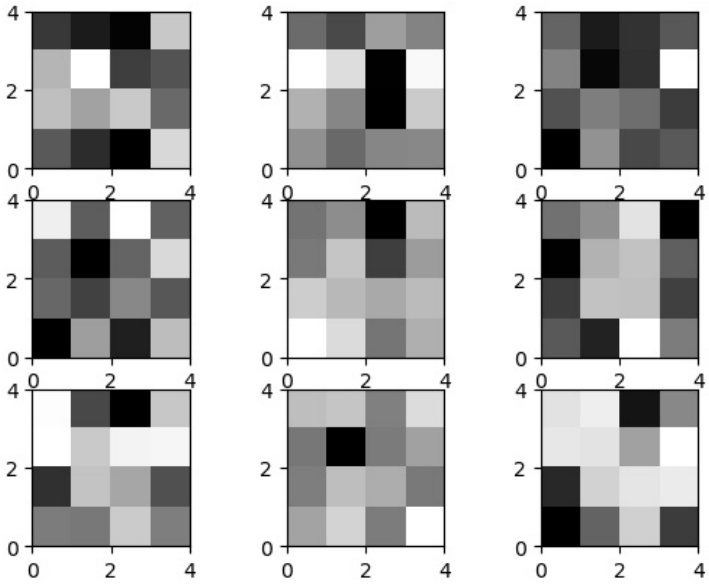
Encoder 7 (Time 1 – Time 2)



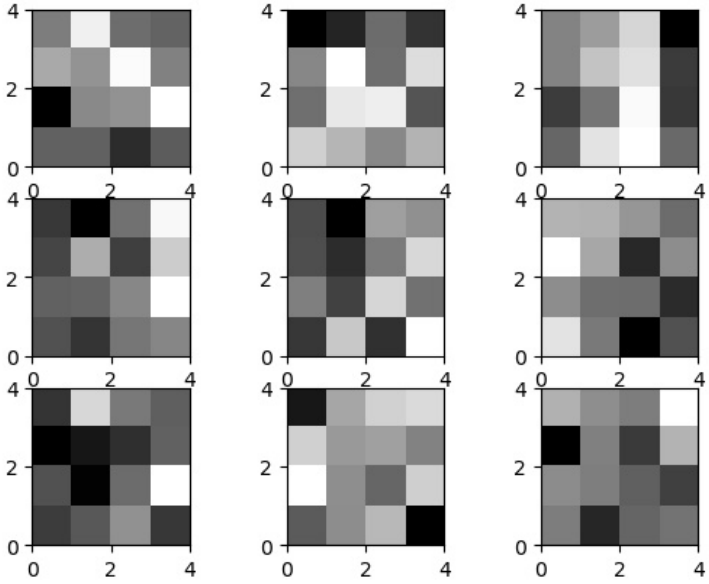
Encoder 8 (Time 1 – Time 2)



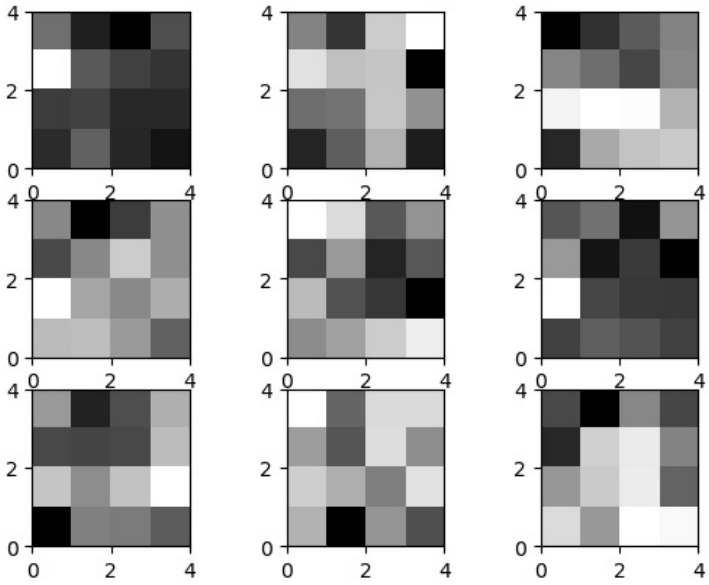
Encoder 1 (Time 2 – Time 1)



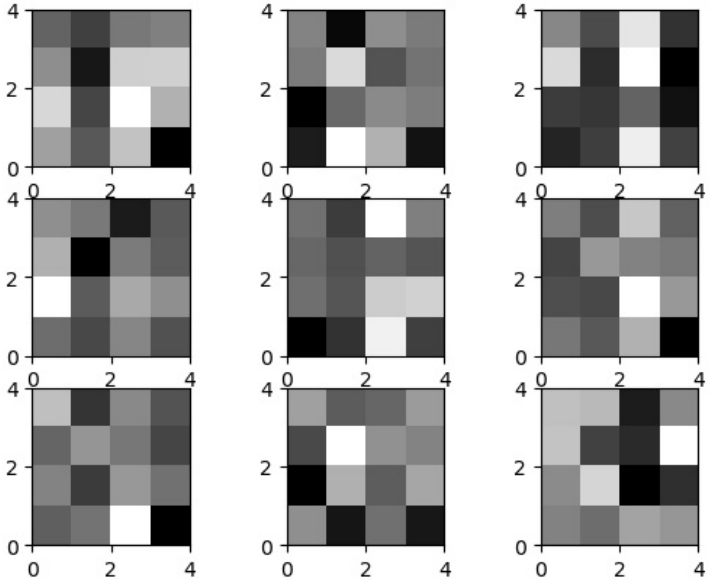
Encoder 2 (Time 2 – Time 1)



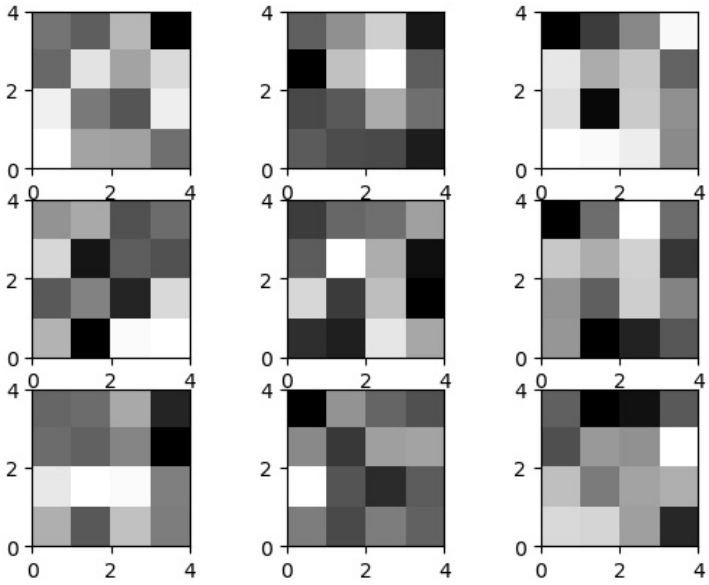
Encoder 3 (Time 2 – Time 1)



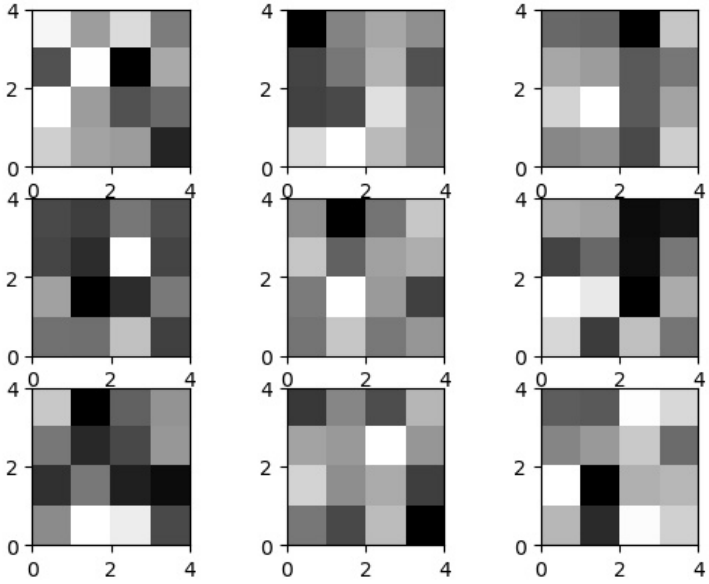
Encoder 4 (Time 2 – Time 1)



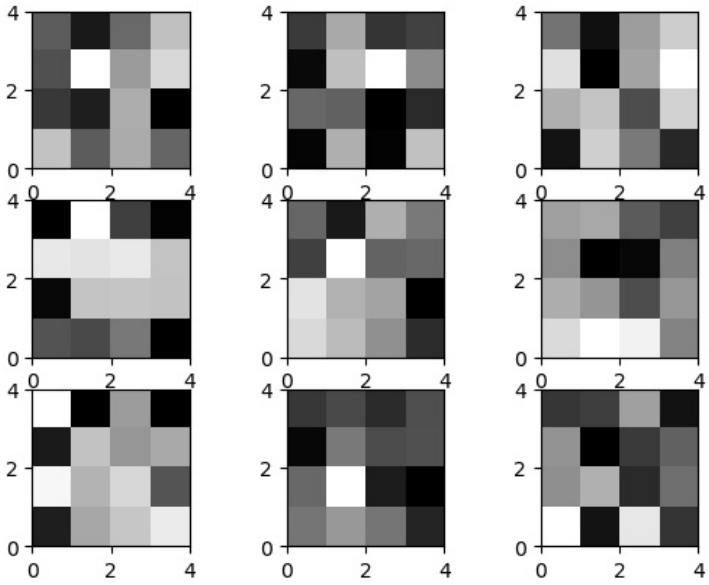
Encoder 5 (Time 2 – Time 1)



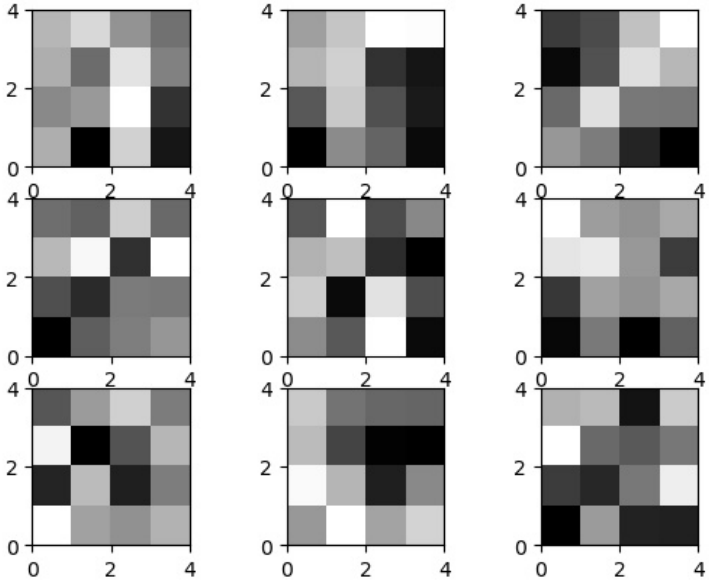
Encoder 6 (Time 2 – Time 1)



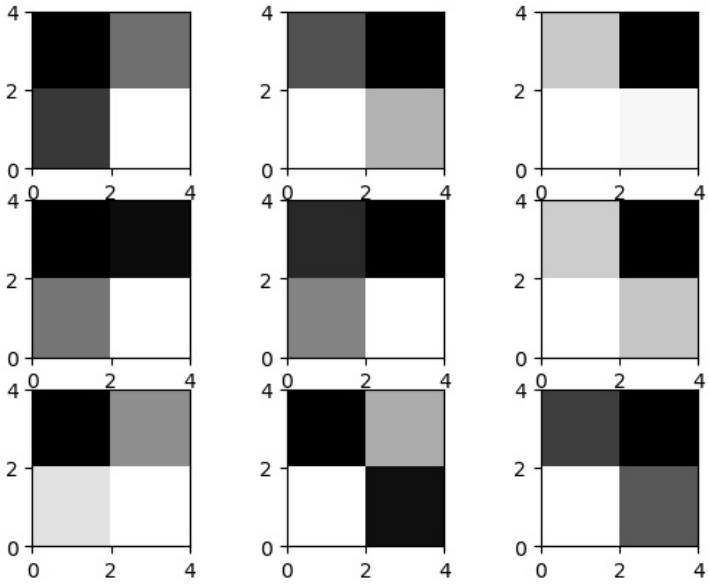
Encoder 7 (Time 2 – Time 1)



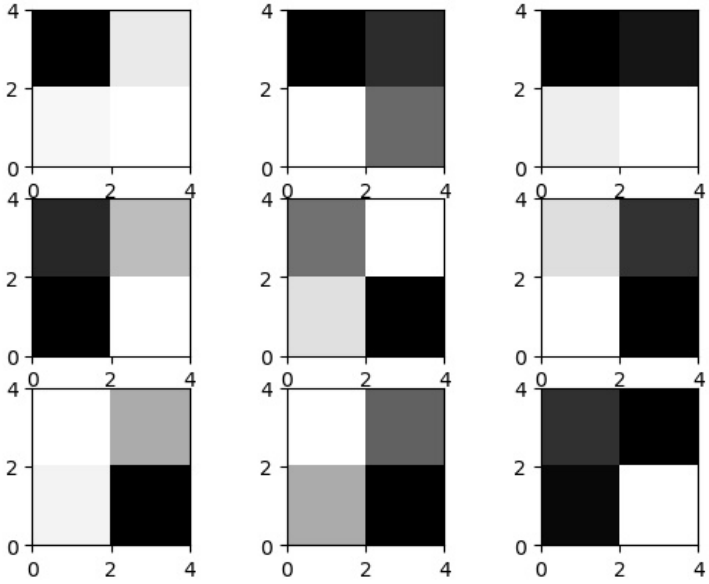
Encoder 8 (Time 2 – Time 1)



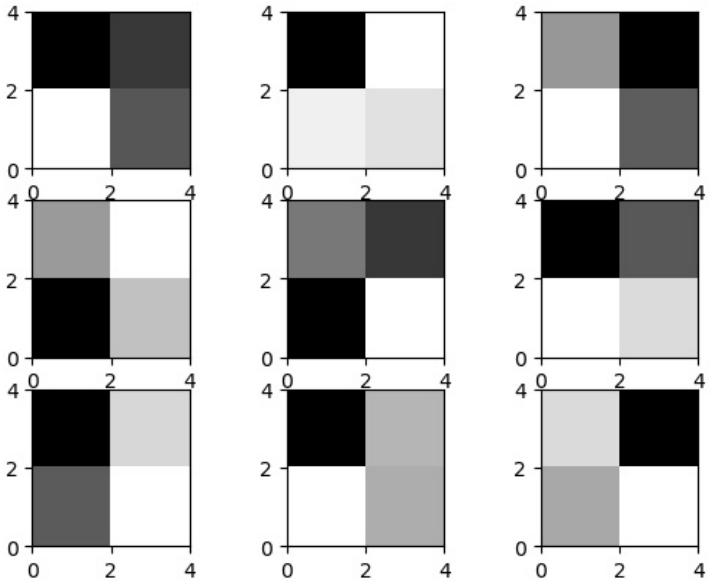
Decoder 1 (Time 1 – Time 2)



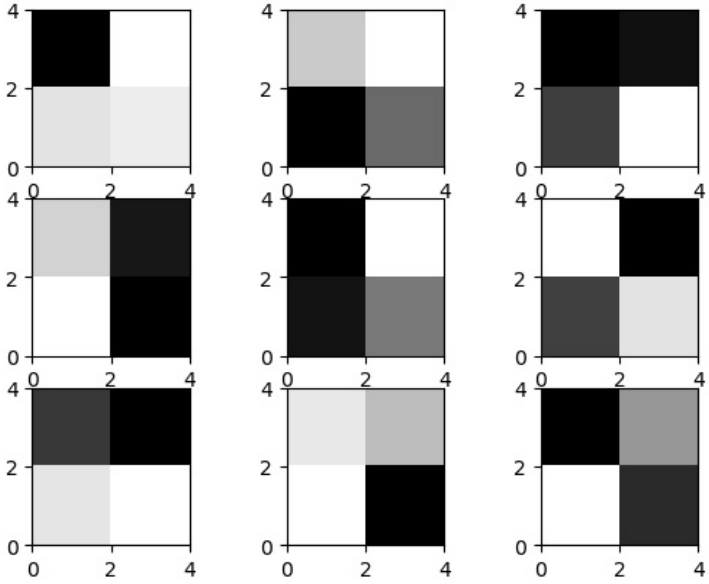
Decoder 2 (Time 1 – Time 2)



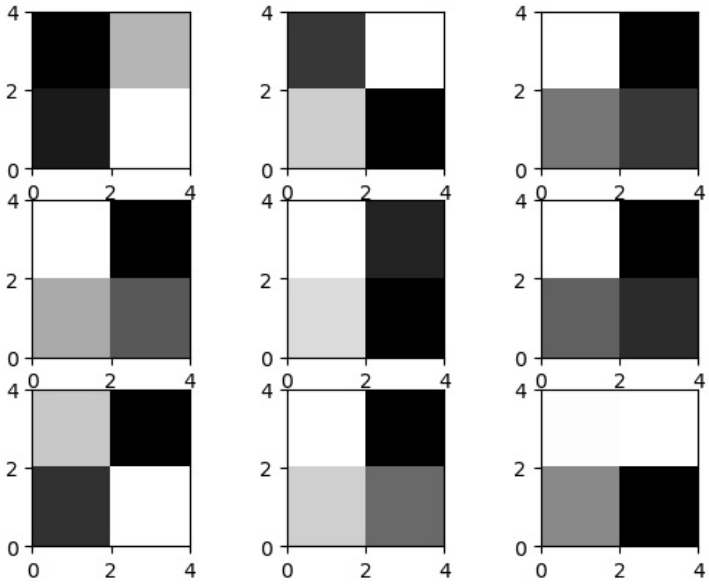
Decoder 3 (Time 1 – Time 2)



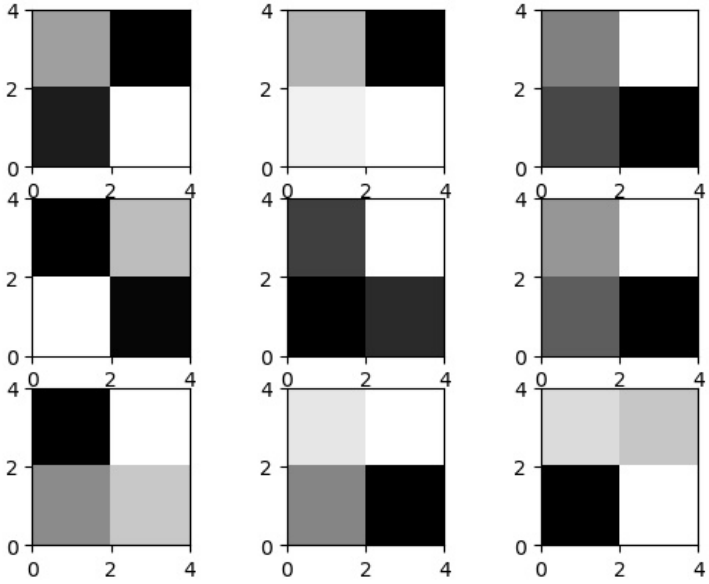
Decoder 4 (Time 1 – Time 2)



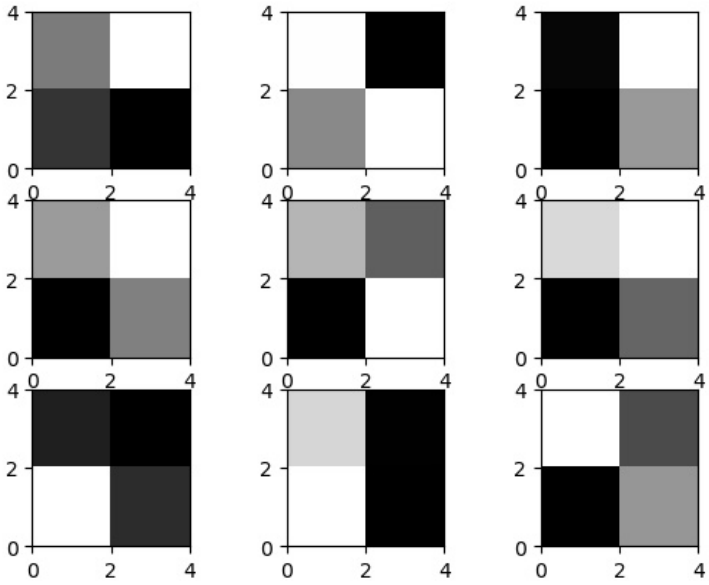
Decoder 5 (Time 1 – Time 2)



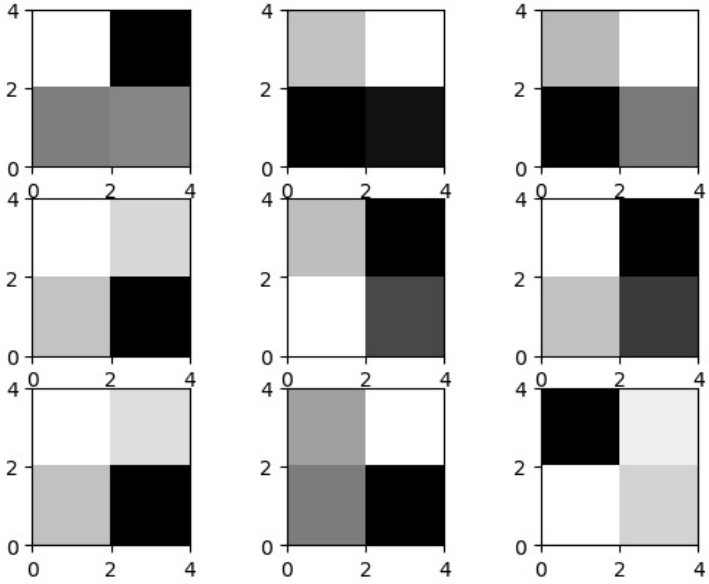
Decoder 6 (Time 1 – Time 2)



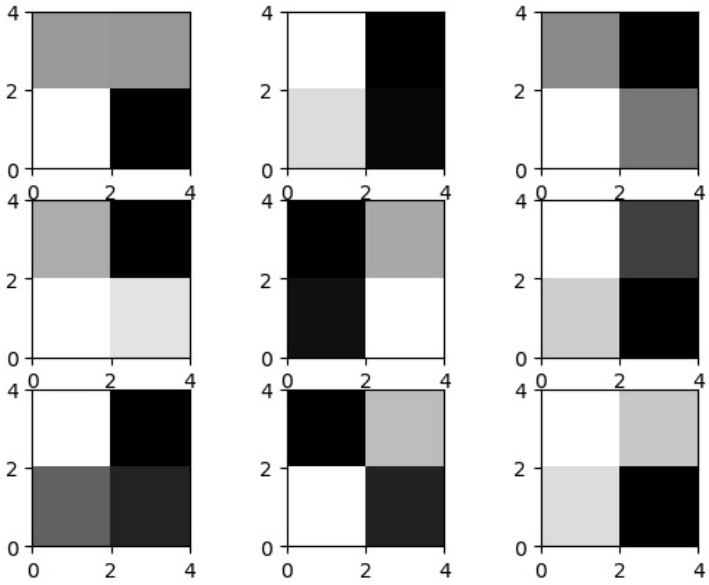
Decoder 7 (Time 1 – Time 2)



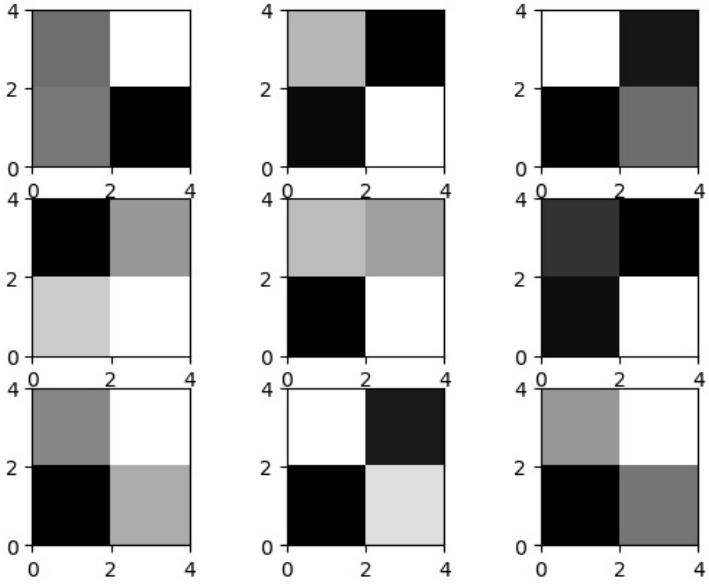
Decoder 8 (Time 1 – Time 2)



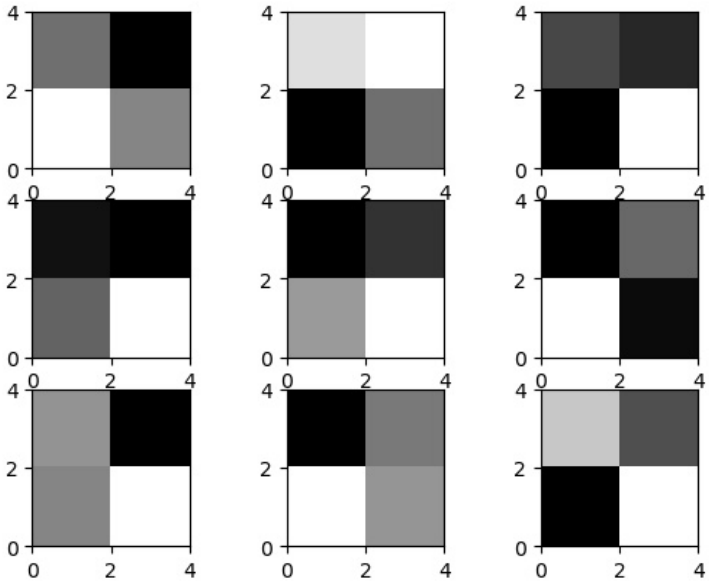
Decoder 1 (Time 2 – Time 1)



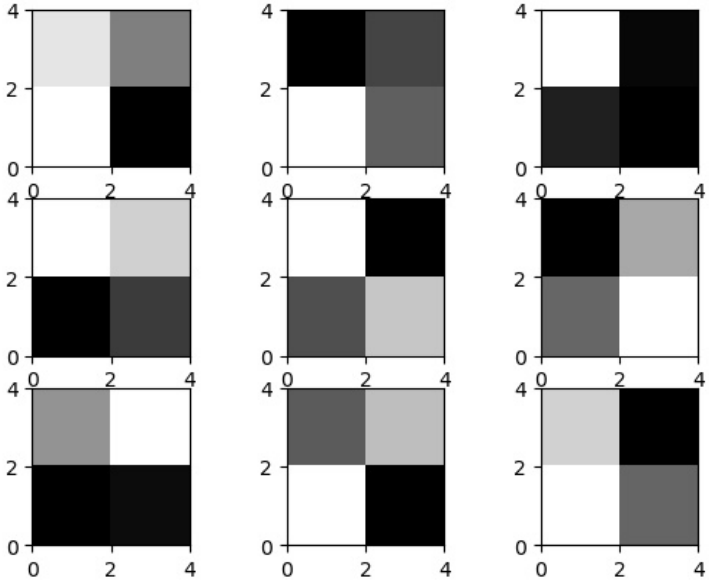
Decoder 2 (Time 2 – Time 1)



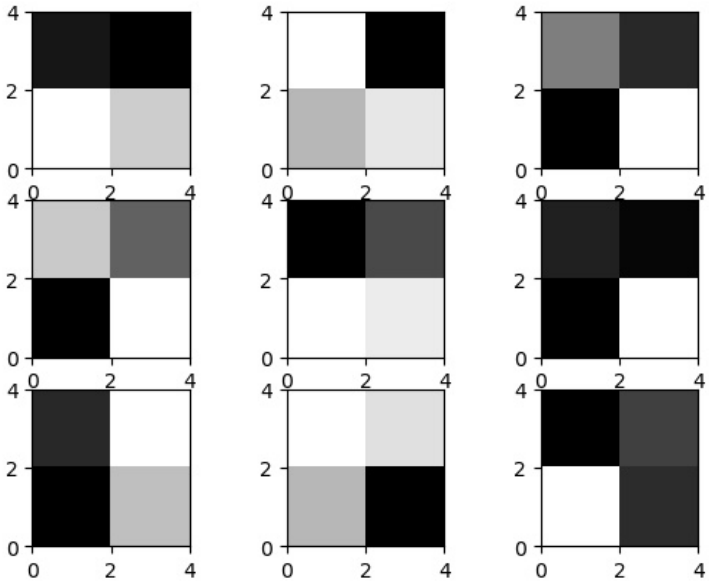
Decoder 3 (Time 2 – Time 1)



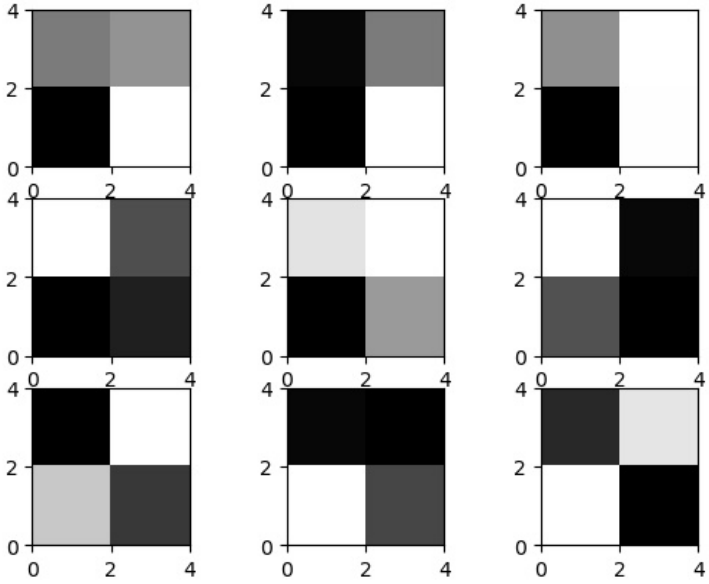
Decoder 4 (Time 2 – Time 1)



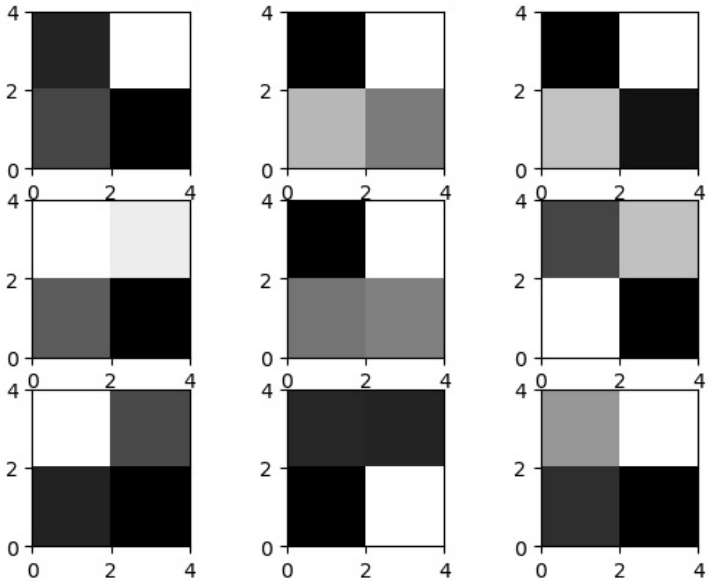
Decoder 5 (Time 2 – Time 1)



Decoder 6 (Time 2 – Time 1)



Decoder 7 (Time 2 – Time 1)



Decoder 8 (Time 2 – Time 1)