論文 / 著書情報
Article / Book Information

題目(和文)	深層学習を応用したレンズレス光学系による次世代イメージングとセンシング
Title(English)	Next-generation Imaging and Sensing: Lensless Optics Equipped with Deep Learning
著者(和文)	潘 秀曦
Author(English)	Xiuxi Pan
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12191号, 授与年月日:2022年9月22日, 学位の種別:課程博士, 審査員:山口 雅浩,SLAVAKIS KONSTANTINOS,熊澤 逸夫,渡辺 義浩,劉 載勲
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12191号, Conferred date:2022/9/22, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信　系 情報通信　コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of　（　工学　）
学生氏名： Student's Name	潘 秀曦（ハン　シュウギ） Pan Xiuxi	指導教員（主）： Academic Supervisor(main)	山口　雅浩
		指導教員（副）： Academic Supervisor(sub)	Konstantinos Slavakis

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Title: Next-generation Imaging and Sensing: Lensless Optics Equipped with Deep Learning

A mask-based lensless camera simply consists of a thin mask and an image sensor. The mask, placed closely in front of the sensor, optically encodes the scene and an unknown smudge is projected on the sensor. This smudge, known as the encoded pattern, can reveal sufficient visual information if a well-designed post-capture computational algorithm is applied. Free of lens, the lensless camera is naturally small, lightweight and low-cost in optical hardware. Nevertheless, the development of the lensless camera is still at its formative stage. This thesis advances the development progress of this technology by bringing in deep learning. By equipping the lensless optics with the proposed deep learning algorithms, this thesis not only achieves considerably improved imaging quality and but also unlocks the reconstruction-free recognition functionality for the lensless camera. Additionally, a novel learned mask pattern optimization method is also proposed.

The improvement of image reconstruction is one of the most important subjects in lensless imaging. Conventional reconstruction methods leverage knowledge of the physical system related to point spread function (PSF), known as PSF-based methods. PSF-based methods are susceptible to system modeling while the forward system of the lensless optics is non-linear, complicated and has not been well established by now. With imprecise system modeling, the performance of the PSF-based methods is limited. Reconstruction with pure data-driven deep neural network (DNN), which relies more on data than modeling, can avoid this limitation, thereby having potential to have better reconstruction quality. However, existing pure DNN reconstruction methods for lensless imaging do not turn out a better result than PSF-based methods. This thesis reveals that the multiplexing property in lensless optics makes global features essential in understanding the encoded pattern. Whereas, all existing DNN reconstruction methods apply fully convolutional networks (FCNs) which are not efficient in global feature reasoning. With this analysis, this thesis proposes the Transformer-based model, named as Lensless Imaging Transformer (LIT). With Transformer, LIT is better in global feature reasoning, hence enhancing the reconstruction. Besides, LIT has two unique designs that are dedicated to the lensless image reconstruction. Firstly, the Overlapped Patchify allows learning hierarchical feature presentations while preserving local continuity. Secondly, the traditional self-attention in Transformer is replaced by the axial-attentions to achieve computation reduction and easier training. The superiority imaging quality of the proposed LIT has been verified by comparing with PSF-based methods and FCNs in an optical experiment.

Considering that one significant application of the lensless camera could be computer vision task, such as object recognition. This thesis, in the first time, proposes to bypass reconstruction and perform object recognition directly on the encoded pattern. Avoiding image reconstruction not only saves computational resources but also averts errors and artifacts in reconstruction. To achieve reconstruction-free object recognition, two solutions are proposed. One solution is a novel data preprocessing approach for FCNs, named local binary patterns map generation (LBPMG). There is disturbance amplification issue caused by the multiplexing property, which makes encoded pattern recognition challenging. LBPMG is designed to alleviate disturbance amplification issue by transforming the encode pattern to an LBP map through comparing pivot pixel with its surrounding pixels. With alleviative disturbance, reconstruction-free recognition with FCNs become possible. Another solution is designing an end-to-end Transformer model, named as Lensless Recognition Transformer (LRT). LRT is inspired by LIT, aiming at enhancing global feature extraction by applying the Transformer to deal with the multiplexing property. Besides, separated convolution is utilized in the LRT to reduce training parameters. And the method of generating simulated training data for pretraining is also proposed. Both of them are important for LRT because the amount of real training data is insufficient and there is non-convex losses issue in Transformer training. The feasibility of LBPMG and LRT has been verified in optical experiments. LRT achieves a recognition accuracy (94.26%) closing to the lensed-camera-used method (97.00%) in the cats-vs-dogs classification task.

In this thesis, amplitude mask is utilized. For amplitude mask design, the aperture size can be optimized by considering the diffraction effect using the Fresnel number. Designing the mask pattern is a task that has attracted significant attention. This thesis proposes a novel mask pattern optimization method which optimizes the mask pattern through learning. Since the encoded pattern can be approximated as a convolution between the object and mask, the mask can be regarded as a convolutional filter. When a DNN is applied for the encoded pattern analysis, the mask can become a convolutional layer placed ahead of the following network. The mask pattern and weights in the following networks can be optimized simultaneously through training. The feasibility of the learned mask pattern optimization method has been verified in a simulation experiment.

In the end, this thesis evolves the lensless camera to be lite-yet-mighty by equipping it with deep learning. The reinforced lensless camera could find its irreplaceable application values as an extremely miniature and low-cost camera with qualified imaging ability and additional functionality of reconstruction-free recognition.