

## 无透镜成像系统中的无重建目标识别技术

陈凯余<sup>1,2,3,4,5</sup>, 李颖<sup>1,2,3,4</sup>, 李政岱<sup>1,2,3,4</sup>, 郭友明<sup>1,2,3,4\*</sup><sup>1</sup>中国科学院自适应光学重点实验室, 四川 成都 610209;<sup>2</sup>中国科学院光电技术研究所, 四川 成都 610209;<sup>3</sup>中国科学院大学, 北京 100049;<sup>4</sup>中国科学院大学电子电气与通信工程学院, 北京 100049;<sup>5</sup>光场调控科学技术全国重点实验室, 四川 成都 610209

**摘要** 无透镜成像系统使用掩模板替代镜头,在降低成本的同时使设备更加轻巧,然而在进行目标识别前需通过计算重建图像,涉及参数调优和计算耗时问题。基于此,提出一种无重建的目标识别方案,直接在无透镜相机拍摄的编码图像上训练网络识别目标,在节约计算资源的同时还提供隐私保护。使用具有相位掩模板和振幅掩模板的无透镜相机,仿真生成 MNIST 与 Fashion MNIST 数据集和实采 MNIST 数据集,然后在这些数据集上训练 ResNet-50 与 Swin\_T 网络进行目标识别。结果表明,在仿真 MNIST、Fashion MNIST 和真实 MNIST 数据集上,所提方案的最高识别准确率达 99.51%、92.31% 和 98.06%,与先重建目标后识别方案的准确率相当,证明所提方案是一种高效的、具有隐私保护的端到端方案,且在两种掩模板和两类常规骨干分类网络上得到了验证。

**关键词** 计算成像; 目标识别; 深度学习; 无透镜成像

中图分类号 O438; TP391

文献标志码 A

DOI: 10.3788/LOP230755

## Reconstruction-Free Object Recognition Scheme in Lensless Imaging Systems

Chen Kaiyu<sup>1,2,3,4,5</sup>, Li Ying<sup>1,2,3,4</sup>, Li Zhengdai<sup>1,2,3,4</sup>, Guo Youming<sup>1,2,3,4\*</sup><sup>1</sup>Key Laboratory on Adaptive Optics, Chinese Academy of Sciences, Chengdu 610209, Sichuan, China;<sup>2</sup>Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, Sichuan, China;<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China;<sup>4</sup>School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China;<sup>5</sup>National Key Laboratory of Optical Field Manipulation Science and Technology, Chengdu 610209, Sichuan, China

**Abstract** Lensless imaging systems use masks instead of lenses, reducing costs and making equipment lighter. However, before object recognition, reconstructing an image is necessary. This reconstruction involves parameter tuning and time-consuming calculations. Hence, a reconstruction-free object recognition scheme, which directly trains networks to recognize objects on encoded images captured via lensless cameras, that saves computing resources and protects privacy, is proposed herein. Using lensless cameras with a phase mask and an amplitude mask, the real MNIST dataset is collected and the simulated MNIST and Fashion MNIST datasets are generated. Subsequently, the ResNet-50 and Swin\_T networks are trained on these datasets for object recognition. The results show that with respect to the simulated MNIST, Fashion MNIST, and real MNIST datasets, the highest recognition accuracy achieved by the proposed scheme is 99.51%, 92.31%, and 98.06%, respectively. These accuracies are comparable to those achieved by the reconstructed object recognition scheme, proving that the proposed scheme is an efficient end-to-end scheme that provides privacy protection. Moreover, the proposed scheme is verified using two types of masks and two types of conventional backbone classification networks.

**Key words** computational imaging; object recognition; deep learning; lensless imaging

收稿日期: 2023-03-01; 修回日期: 2023-03-31; 录用日期: 2023-04-12; 网络首发日期: 2023-04-22

基金项目: 国家自然科学基金(12173041, 11733005)、中国科学院青年创新促进会(2020376)、中国科学院光电技术研究所前沿部署项目(C21K002)

通信作者: \*guoyouming@ioe.ac.cn

## 1 引言

人类感知世界 80% 依赖视觉,而成像技术客观上延伸了人眼视觉范围。成像大致经过光化学成像、胶片成像、数码成像再到计算成像 4 个阶段。其中计算成像被视为第四次成像革命,是一个联合优化光学系统和信号处理以实现特定成像功能与特性的新兴研究领域<sup>[1-2]</sup>。无透镜成像技术可以让设备体积更小巧、成本更低廉,是计算成像中的一种典型技术<sup>[3]</sup>,它抛弃了传统光学透镜成像概念,在成像传感器前通过掩模板对目标光进行编码或衍射,然后再对编码图像进行计算重建来生成目标图像,整个过程可以视为通过掩模板进行光学调制,然后再通过计算进行光学解调。

目前无透镜成像系统下进行目标识别的主流方案是先对编码图像进行计算重建得到清晰的目标图像,然后再使用深度学习训练网络进行分类。2022 年清华大学 Shi 等<sup>[4]</sup>利用一种无透镜光电混合神经网络架构对隐私保护人脸进行识别,其中电域卷积网络由重建网络和分类网络组成并进行交替训练,最终获得与非加密方法相当的识别准确率。2022 年浙江大学 Zhang 等<sup>[5]</sup>研究了在无透镜成像系统下文本的检测与识别工作,文中网络由 U-Net 重建网络和后续的文本检测与识别网络构成,实验结果表明,无透镜成像系统对简单背景下文本检测与识别的性能与带镜头相机的性能相似,随着文本字体背景等多样性的增加,性能会显著下降。

在无透镜成像系统下进行目标识别的另一种方案是不做目标图像的计算重建而直接在编码图像上利用深度学习训练网络进行分类。2021 年 Pan 等<sup>[6]</sup>在分析了基于掩模的无透镜光学的多路复用特性后提出了一种简化的 Transformer 架构<sup>[7]</sup>,跳过图像重建过程,直接对编码图像进行识别,该架构在 Fashion MNIST 数据集上的准确率达 91.47%,在猫狗数据集<sup>[8]</sup>上的准确率达 94.26%。2022 年浙江大学 Zhang 等<sup>[9]</sup>开展了无透镜相机拍摄的视频中的手势识别方面的研究工作,并提出了一种新颖的端到端深度神经网络模型,该模型对原始编码图像免重建而直接识别手势,在剑桥手势数据集上的准确率达 98.59%,与有透镜相机的结果相当。此外还有一些其他方案,针对编码图像的新型数据预处理方法,即局部二进制模式映射(LBP)<sup>[10]</sup>、改进的衍射深度神经网络(D<sup>2</sup>NNs)与电子神经网络集成混合分类器<sup>[11]</sup>。

考虑到以机器推理为最终目标的成像需求日益增多,成像更多地被看作为机器推理提供数据的前置过程。无透镜成像技术通过掩模板的光学调制作用在传感器上形成目标的编码图像,借鉴深度学习中编码器-解码器架构的思想,认为可以直接在编码图像上训练网络进行分类,重建目标不是必须的。为了验证这一观点,本文拟在 MNIST 和 Fashion MNIST 数据集<sup>[12]</sup>

上进行无透镜成像仿真和数据实采,然后利用主流的骨干网络在编码图像上进行分类。实验结果表明,所提方案可以获得与目标重建再识别方案相近的准确率,在节约目标重建的计算资源、加快识别速度的同时也天然地为目标提供光学加密、隐私保护。

## 2 相关理论

### 2.1 无透镜成像系统的简化模型

根据傅里叶光学的理论,透镜具有傅里叶变换的作用。基于两个假设,物体是不同光强度的点的集合,点扩散函数(PSF)是平移不变的,就可以运用该理论来创建基于掩模的无透镜成像系统的简化物理模型。该模型的物理意义就是每一个传感器单元接收的信号是不同目标点在不同编码权重调制下信息的线性叠加<sup>[13]</sup>。

图 1 是基于二值化振幅掩模板的无透镜成像的原理。

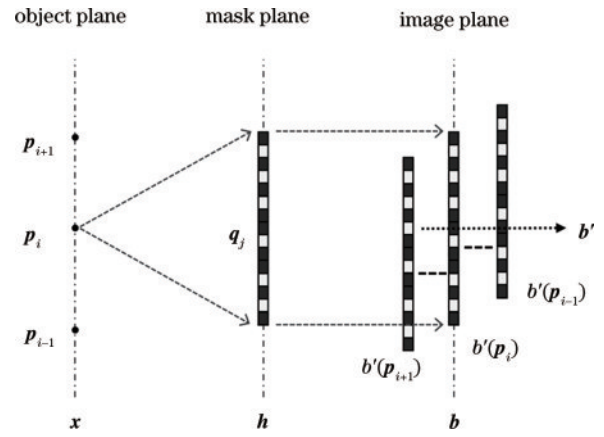


图 1 无透镜成像的原理

Fig. 1 Principle of lensless imaging

物面处的单个点  $p_i$  经编码后的成像公式为

$$b'(p_i) = \int x(p_i)h(q_j)dj, \quad (1)$$

式中:  $b'(p_i)$  是物面处单个点  $p_i$  在像面处的成像结果;  $x(p_i)$  是物面处的单个点的光强度;  $h(q_j)$  是编码面处的单个点。物体上单个点的成像过程则是点光强在编码图像上的积分。然后整个物面处的物体经编码面在像面处的成像公式可以表示为

$$\mathbf{b} = \int b'(p_i)di = \iint x(p_i)h(q_j)didi, \quad (2)$$

式中:  $\mathbf{b}$  代表整个物体经编码面在像面处所成的图像。整个物体的成像过程可视为物体图像与编码图像的卷积。在考虑噪声的情况下,得到最后的简化模型,为

$$\mathbf{b} = \mathbf{x} * \mathbf{h} + \mathbf{n} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (3)$$

式中:  $\mathbf{x}$  代表整个物体;  $\mathbf{h}$  代表掩模板的点扩散函数;  $\mathbf{n}$  代表高斯噪声;  $\mathbf{H}$  代表观测矩阵;  $*$  代表卷积运算符。

以上是以二值化的振幅掩模板为例对无透镜成像系统的简化模型进行原理阐述的内容,这套理论同样适用于相位掩模板情形。常用的掩模板有相位掩模板和振幅掩模板两种,它们都可以形成具有较多细节的

点扩散函数,实现对物像的光学调制。为了实验完整性,后续会使用两种掩模板进行实验。

## 2.2 目标识别

具体的图片分类方法有很多种,其中基于深度学习的一些骨干分类网络(比如 ResNet<sup>[14]</sup>、SwinTransformer<sup>[15]</sup>等)在图片分类上性能表现极好,为了尽量减少分类方法性能不佳导致整体方案目标识别准确率下降的可能性,本文在实验部分将采用常用的骨干分类网络。ResNet-50<sup>[14]</sup>作为卷积神经网络(CNN)中具有代表性的骨干分类网络,在常规图片分类上表现良好,选取 ResNet-50 作为后续实验的分类网络是具有代表性和参考价值的。此外 Pan 等<sup>[6]</sup>考虑

到无透镜成像系统的多路复用特性,在系统中引入 Transformer 架构,系统取得很好的识别性能,所以可以选取相同架构的 SwinTransformer 骨干网络来进行分类,让实验结果更加完备。其中 SwinTransformer 网络中的 tiny 模型 Swin\_T<sup>[15]</sup>在模型参数数量与运行时间上与 ResNet-50 相近,作为对比网络更加合适。

引言中介绍了当前在无透镜成像系统下进行目标识别的两种主流方案,为了对比和评价不同方案的优劣,参考 Pan 等<sup>[6]</sup>和 Zhang 等<sup>[9]</sup>的对比实验设计方案,本文设计的对比实验组如图 2 所示。对比实验组中的方案 1 就是在常规相机捕获的目标图像上训练分类网络的,然后将其识别准确率作为对比基准。

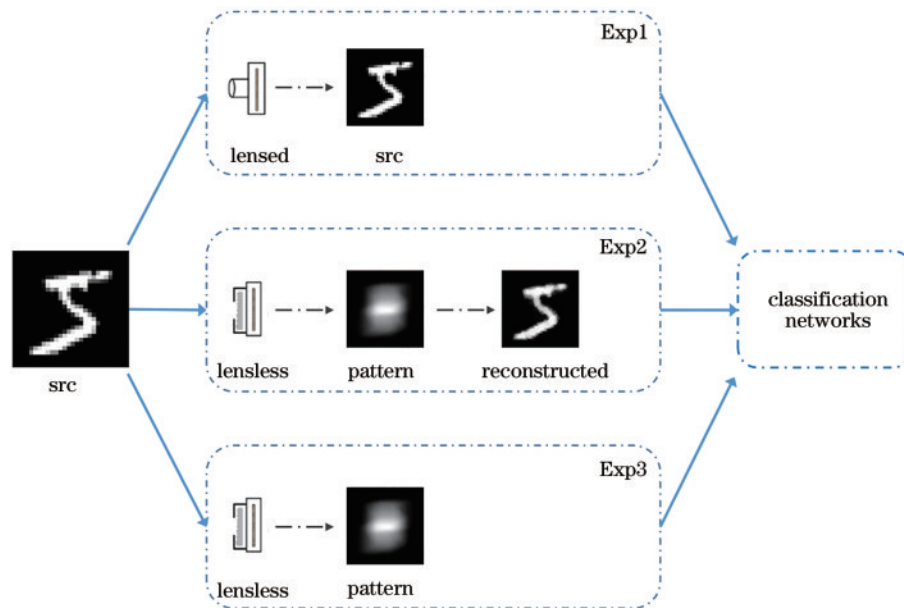


图 2 目标识别方案的对比实验组

Fig. 2 Comparative experimental group of object recognition schemes

对比实验组中的方案 2 是在无透镜相机下先对编码图像进行计算重建得到清晰的目标图像,然后再训练网络进行分类的。其中分类网络与基准实验组相同,不再进行过多阐述。对于无透镜成像系统下目标图像的重建算法,目前主要有 3 类:迭代优化法,比如交叉方向乘法(ADMM)<sup>[16]</sup>;纯深度神经网络,比如 U-Net<sup>[17]</sup>;将物理模型展开成深度网络,比如 ADMM-Net<sup>[18]</sup>。在不考虑时间复杂度的情况下,基于 ADMM 的物理模型迭代优化法能得到相对比较好的图像重建效果,且该方法模型收敛性很好,参数量相比神经网络较少,这样就更容易优化参数。鉴于以上优点,可以认为将 ADMM 作为重建算法能得到可靠的实验结果。ADMM 重建算法的模型与迭代求解过程参考 DiffuserCam<sup>[19]</sup>。

目前常用分类网络在常规相机所拍摄到的高清图片上的分类性能表现是值得肯定的,其中 CNN 类网络因为卷积操作,在局部特征提取方面具有优势;Transformer 类网络因为自注意力机制可以反映复杂

的空间变换和长距离特征依赖,能获得更好的全局特征。对比实验组中的方案 3 是在无透镜相机下直接对编码图像进行分类的,此时的编码图像可视为原图像经过掩模板光学调制后所得的结果,同样包含了原图像的所有信息。本文认为可以凭借深度网络强大的模型表示能力来直接提取特征,其中分类网络与基准实验相同,使用 ResNet-50 和 Swin\_T 作为分类网络。

## 3 实验与分析

### 3.1 仿真数据实验

#### 3.1.1 生成仿真数据集

使用的数据集是 MNIST 和 Fashion MNIST,它们中的样本属于单通道图片,构型相对简单,是广泛采用的公开基准数据集。根据无透镜成像系统的简化模型和 DiffuserCam<sup>[19]</sup>中的理论和处理过程,可以知道只需要常规数据集和 PSF 就可以生成仿真数据集,其中实验选用的常规数据集是公开且可以通过网络下载获取的。为了获取良好的 PSF 图像,可以使用位于中央

的点光源照射无透镜相机,此时图像传感器捕获到的就是 PSF 图像。本文使用的两种类型的掩模板对应的 PSF 图像如图 3 所示,其中左子图是随机相位掩模板对应的 PSF 图像,右子图是二值化的振幅掩模板<sup>[20]</sup>对应的 PSF 图像。

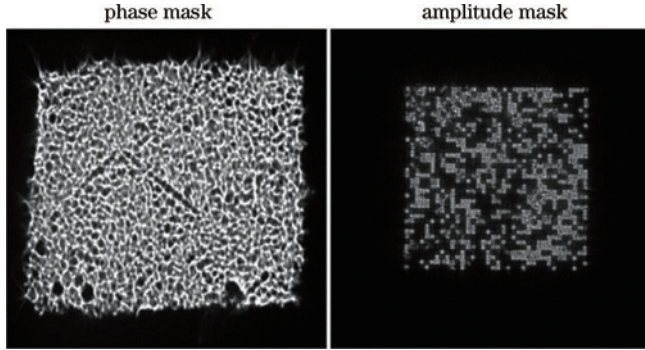


图 3 两种类型的掩模板对应的 PSF 图像

Fig. 3 PSF images corresponding to two types of masks

在仿真过程中,卷积操作可以通过傅里叶变换来加速实现,为了保证图片边缘的完整性,需要进行边缘零填充,而噪声则是通过均值为零的高斯分布来模拟,具体过程为

$$I_{\text{img}} = O_{\text{bj}} * P_{\text{sf}} + n = \mathcal{F}^{-1} \left\{ \mathcal{F} [P_{\text{ad}}(O_{\text{bj}})] \cdot \mathcal{F} [P_{\text{ad}}(P_{\text{sf}})] \right\} + \mathcal{N}(0, \sigma^2), \quad (4)$$

式中: $I_{\text{img}}$ 、 $O_{\text{bj}}$ 和 $P_{\text{sf}}$ 分别代表编码图像、原图像和 PSF 图像; $\mathcal{F}$ 和 $\mathcal{F}^{-1}$ 分别代表傅里叶变换与傅里叶逆变换; $P_{\text{ad}}$ 代表对图像进行零填充,至原尺寸的 2 倍; $\mathcal{N}(0, \sigma^2)$ 代表高斯分布。

使用的 ResNet-50 和 Swin\_T 分类网络均有 AdaptiveAvgPool2d 网络层,可以自适应地处理不同图

片尺寸的输入。一般而言图片尺寸越大,其分辨率就越高,纹理细节就越丰富,分类网络训练后模型准确率就越高,但代价是计算耗时就越长,这是一个权衡问题。考虑到 PyTorch 官网中 ResNet-50 和 Swin\_T 的预训练模型建议的图片输入尺寸为  $224 \times 224$ , 所以为了获取最高准确率的模型,仿真生成的编码图像的尺寸应该大于等于  $224 \times 224$ 。然后根据式(4)可知在仿真过程中,利用快速傅里叶变换(FFT)来加速卷积运算时为了能在原理上满足循环(圆)卷积的要求,会对图像进行零填充,至原尺寸的 2 倍,所以本文选用的图像尺寸为  $128 \times 128$ , 这样仿真生成的编码图像尺寸为  $256 \times 256$ , 既满足尺寸大于  $224 \times 224$  的要求, 同时也能减少 FFT 的运算时间和分类网络的运算时间。

根据设计的对比实验组方案,将原数据集作为常规相机采集的图像,无重建的目标识别方案可以使用仿真数据集,基于重建的目标识别方案需要在仿真数据集上利用 ADMM 进行重建生成重建数据集,为了增强实验结果的说服力就需要得到很好的重建效果。根据 Monakhova 等<sup>[21]</sup>的研究可知 ADMM 在经过 100 次迭代优化后可基本收敛,能得到效果很好的重建图片,本文考虑到运行时间上的消耗,利用惩罚系数的动态调节技巧,取 30 次迭代优化后的重建图像。

在 MNIST 和 Fashion MNIST 两个数据集上的部分仿真实验结果如图 4 所示,第 1 层 src 为数据集原图像,第 2 层 encoded pattern-1 和第 3 层 reconstructed-1 为使用相位掩模板所得的编码图像和对应的重建图像,第 4 层 encoded pattern-2 和第 5 层 reconstructed-2 为使用振幅掩模板所得的编码图像和对应的重建图像。根据这些结果明显可知:编码图像不易被人类直

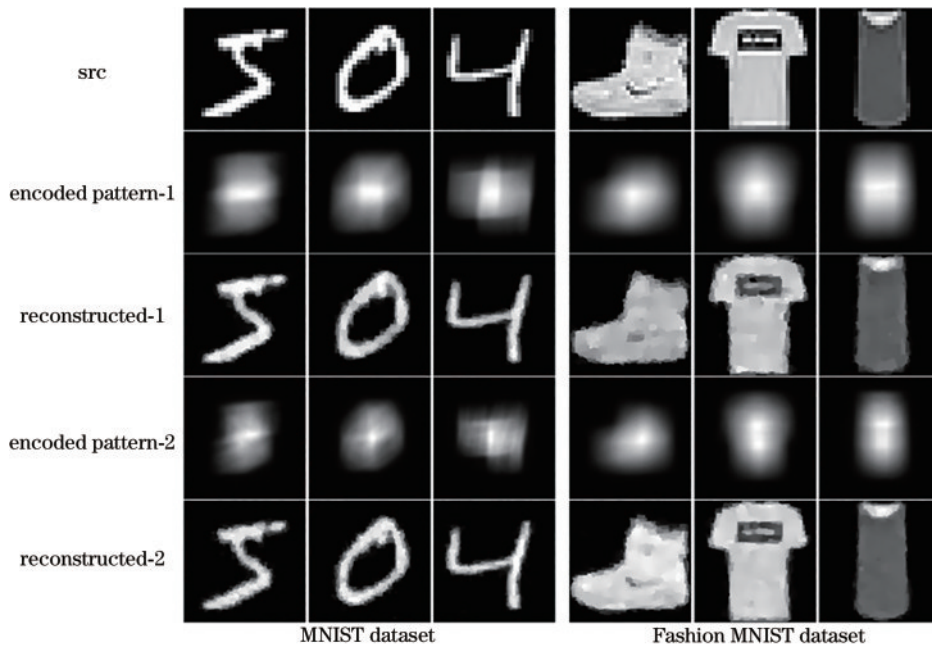


图 4 部分仿真实验结果

Fig. 4 Some simulation results

接理解识别,具有天然的光学加密效果;重建图像与原图像之间存在细微差异,可知计算重建过程会引入额外噪声和伪影,导致有效信息的可靠性下降。

### 3.1.2 实验结果

根据之前设计的对比实验组方案开展仿真实验,在 MNIST 数据集上的具体实验结果如表 1 所示。

表 1 在 MNIST 数据集上的实验结果

Table 1 Experimental results on the MNIST dataset

Mask type	Network	Experiment	Accuracy / %
Phase mask	ResNet-50	Exp1(lensed)	99.76
		Exp2(reconstructed)	99.68
		Exp3(lensless)	99.38
	Swin_T	Exp1(lensed)	99.77
		Exp2(reconstructed)	99.67
		Exp3(lensless)	99.45
Amplitude mask	ResNet-50	Exp1(lensed)	99.76
		Exp2(reconstructed)	99.68
		Exp3(lensless)	99.51
	Swin_T	Exp1(lensed)	99.76
		Exp2(reconstructed)	99.72
		Exp3(lensless)	99.46

根据表 1 中的实验结果可以看出,在使用相位掩模板的情况下,ResNet-50 和 Swin\_T 两种分类网络在三种方案下的识别准确率均在 99% 以上,这不仅说明使用 ADMM 进行重建的效果很好,也说明了在编码图像上直接使用常规骨干分类网络也能得到很好的分类结果。详细对比表中的不同实验组之间的准确率差异,可以知道 Transformer 在数据量规模不大时在常规图像上的分类性能与 ResNet-50 相当,在编码图像上的分类性能稍优于 ResNet-50,这与 Pan 等<sup>[6]</sup>的实验结果相呼应,但也不排除模型训练误差的影响。无论如何,表 1 数据至少可以说明,在编码图像上训练常用骨干分类网络进行无重建的目标识别与在重建图像上训练网络进行目标识别两种方案的准确率相当。使用振幅掩模板的实验结果同样印证了这一结论,说明掩模板的种类并不影响这一结论。

在 Fashion MNIST 数据集上的具体实验结果如表 2 所示。根据表 2 可以看出,在 Fashion MNIST 数据集上,使用相位掩模板的情况下,基于原数据的基准实验指标最高,准确率在 95% 以上,而另外两组对比实验虽然准确率稍低一些,但是这两种方案的准确率却是相近的。这说明针对略微复杂的数据集,无透镜相机的图像重建质量会略低,导致识别准确率下降,常规骨干分类网络在编码图像上的分类准确率也同样会略有下降。使用振幅掩模板的实验结果有同样的结论。总体来说,在 Fashion MNIST 数据集上,对比实验的结果同样证明了之前在 MNIST 数据集上得到的结论。

表 2 在 Fashion MNIST 数据集上的实验结果

Table 2 Experimental results on the Fashion MNIST dataset

Mask type	Network	Experiment	Accuracy / %
Phase mask	ResNet-50	Exp1(lensed)	95.77
		Exp2(reconstructed)	91.37
		Exp3(lensless)	91.28
	Swin_T	Exp1(lensed)	95.34
		Exp2(reconstructed)	92.26
		Exp3(lensless)	90.40
Amplitude mask	ResNet-50	Exp1(lensed)	95.77
		Exp2(reconstructed)	92.65
		Exp3(lensless)	92.31
	Swin_T	Exp1(lensed)	95.34
		Exp2(reconstructed)	92.98
		Exp3(lensless)	91.05

### 3.2 真实数据实验

为了验证在仿真数据集上得到的结果,在实验室搭建了无透镜成像实验系统,然后采集了 MNIST 数据集的真实数据。图 5 是搭建的无透镜成像系统的示意图,sensor 代表图像传感器,mask 代表随机相位掩模板,display 代表显示器屏幕,其上会显示 MNIST 数据集中的图片。

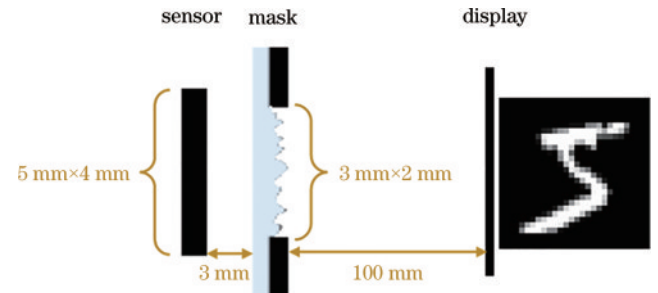


图 5 无透镜成像系统的示意图

Fig. 5 Schematic of the lensless imaging system

在光学实验台上搭建的无透镜成像系统的实验装置实物如图 6 所示。图 6(a)中 display 为高清显示器,lensless 代表无透镜相机,对常规相机用掩模板替代光学镜头;图 6(b)相机使用的是 CCD 工业相机,相机型号是 Basler acA780-75gm;掩模板采用的是思高双面胶带,因为胶带具有优良的透光性<sup>[22]</sup>,厚薄随机,所以可以作为一种随机相位掩模板,其对应生成的 PSF 图像如图 6(c)所示。

利用这套实验装置,图像传感器可以捕获到真实的无透镜成像系统的编码图像,采集到的部分数据如图 7 所示,第 1 层 src 代表 MNIST 数据集中的原图片,第 2 层 encoded pattern 代表实采的编码图像,第 3 层 reconstructed 代表重建后的目标图像。对比原图像与重建图像,可以发现所提 ADMM 在这类简单图像上的重建效果非常好,但仍可以发现少部分图像,比如图

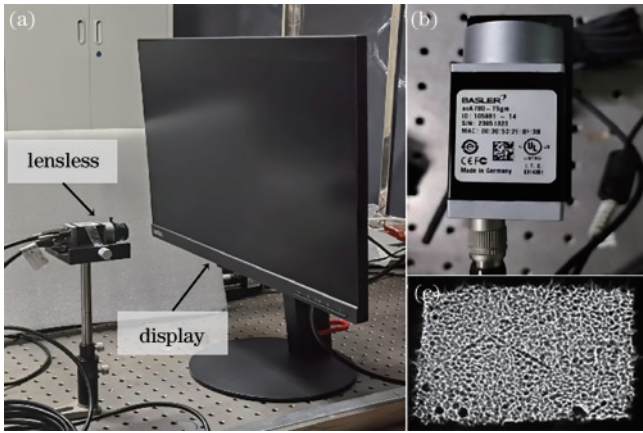


图6 无透镜成像系统的实验装置和 PSF 图像。(a)装置实物图;(b)相机;(c)随机相位掩模板的 PSF 图像  
Fig. 6 Experimental device and PSF image of the lensless imaging system. (a) Physical drawing of the device; (b) the camera; (c) PSF image of the random phase mask

中的最后两个数字 9 与数字 2,重建图像上有一些伪影,这可能是因为采集真实数据时相机曝光时间控制不当,在一次曝光时间内显示器依次显示了两幅图片,但这些质量稍差的实采图像在全数据集中占比很小,基本不影响后续实验。

参照仿真实验过程,在实采的 MNIST 数据集上的具体实验结果如表 3 所示。表 3 中的结果表明,使用 ResNet-50 的各对比实验组的目标识别准确率均在 98% 以上,考虑到实采的数据集本身存在一些瑕疵这一因素,Exp2(reconstructed)和 Exp3(lensless)对应的准确率可预见性地会略有下降,所以真实数据实验上的结果可以印证仿真数据实验的结果,证明了无透镜成像系统下无重建的目标识别方案在真实场景下也能取得与先重建目标后识别方案相近准确率的结论。Swin\_T 网络的实验结果呈现出相似的现象,强化证明了这一结论。

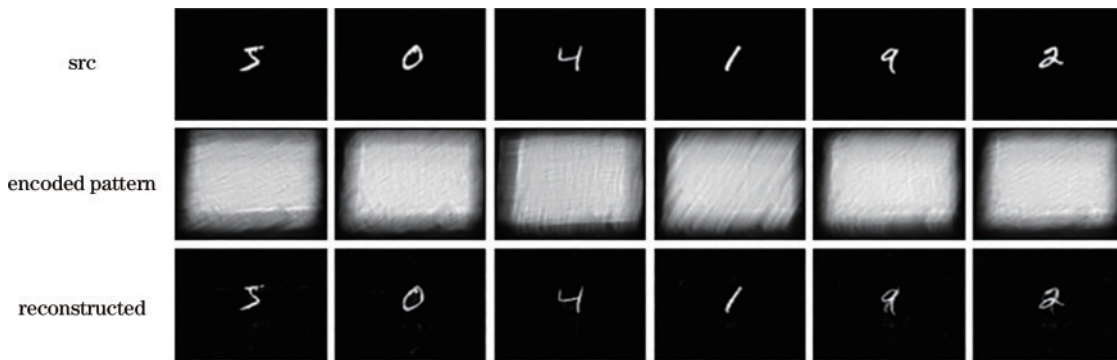


图7 部分无透镜成像的真实编码图像

Fig. 7 Some real encoded images of lensless imaging

表3 在真实数据集 MNIST 上的实验结果

Table 3 Experimental results on the real MNIST dataset

Network	Experiment	Accuracy /%
ResNet-50	Exp1(lensed)	99.43
	Exp2(reconstructed)	98.10
	Exp3(lensless)	98.05
Swin_T	Exp1(lensed)	99.32
	Exp2(reconstructed)	98.08
	Exp3(lensless)	98.06

### 3.3 实验设备和网络训练细节

所有的网络训练都在实验室服务器上运行。硬件配置:处理器是 Intel(R) Xeon(R) Platinum 8276L CPU @ 2.20 GHz,显卡是 Nvidia GeForce RTX 3090 24 GB,内存总量是 128 GB。软件环境配置:Ubuntu 22.04.1 LTS, Python 3.10.9, PyTorch 1.13.1, cuda 11.6。

本文所用仿真数据集和真实数据集都有 7 万张图片,采用相同的数据集划分方式,即后 1 万张图片为测试集,前 6 万张图片通过随机划分的方式分为含 5 万张

图片的训练集和 1 万张图片的验证集。网络训练使用的参数:批大小 batch size 是 256;训练轮次 epoch 是 300;优化器是 AdamW,附带 0.3 的权重衰减;初始学习率是 0.001;使用 StepLR 的学习率调度器,步长为 6,更新率为 0.75;同时使用指数移动平均(EMA)、标签平滑和数据增强(mixup)等技术增加图片的多样性。骨干分类网络 ResNet-50 和 Swin\_T 在开始训练前均加载了 PyTorch 提供的预训练模型,在训练过程中则是选择保存在验证集上准确率最高的模型作为最终模型。

## 4 结 论

在研究无透镜成像系统的目标识别方案时,认为基于掩模板的无透镜成像过程可以视为光学编码过程,得到的编码图像就是物体图像的一种编码形式,所以本文提出了可以直接利用深度网络强大的模型拟合能力来直接对编码图像进行目标识别的方案,同时认为作为解码过程的计算重建是不必要的。实验结果表明,常规的分类网络在编码图像上进行无重建的目标识别的准确率与先进行目标重建后识别的准确率相当,这就说明所提方案在保证了目标识别准确率的同时

时,不仅减少了计算重建过程带来的调参工作和计算资源,而且也利用了无透镜成像系统天然的光学加密效果。无透镜成像系统下无重建的目标识别是一种高效的、具有隐私保护的、端到端的优秀方案。

使用的数据集是背景与目标构型均相对简单的 MNIST 和 Fashion MNIST 数据集,为了拓展应用场景,后续工作可以选用背景与目标构型更复杂、更贴近真实生活场景的数据集进一步验证结论。同时为了得到所提方案的目标识别准确率,还可以针对特定的数据集有针对性地设计优化特殊的分类网络与掩模板。

无透镜成像系统中的无重建目标识别技术凭借低成本、小型化、低功耗和光学加密等特性,在隐私敏感型的监控或识别场景(比如家庭监控、家用体感设备)、物联网(万物互联可视)、植入式医学检测探头、小型无人机等领域都具有重要的应用前景,具有很高的研究价值。

### 参 考 文 献

- [1] 左超, 陈钱. 计算光学成像: 何来, 何处, 何去, 何从? [J]. 红外与激光工程, 2022, 51(2): 20220110.  
Zuo C, Chen Q. Computational optical imaging: an overview[J]. Infrared and Laser Engineering, 2022, 51(2): 20220110.
- [2] 纪铁男, 李海峰, 刘旭. 大视场单镜片计算成像系统图像分割学习方法[J]. 光电工程, 2022, 49(5): 210371.  
Ji Y N, Li H F, Liu X. Image segmentation learning method for large field single lens computational imaging system[J]. Opto-Electronic Engineering, 2022, 49(5): 210371.
- [3] 邵晓鹏, 刘飞, 李伟, 等. 计算成像技术及应用最新进展[J]. 激光与光电子学进展, 2020, 57(2): 020001.  
Shao X P, Liu F, Li W, et al. The latest development of computational imaging technology and its application[J]. Laser & Optoelectronics Progress, 2020, 57(2): 020001.
- [4] Shi W X, Huang Z, Huang H H, et al. LOEN: lensless opto-electronic neural network empowered machine vision [J]. Light: Science & Applications, 2022, 11(1): 1-12.
- [5] Zhang Y E, Wu Z Y, Lin P Y, et al. Text detection and recognition based on a lensless imaging system[J]. Applied Optics, 2022, 61(14): 4177-4186.
- [6] Pan X X, Chen X, Nakamura T, et al. Incoherent reconstruction-free object recognition with mask-based lensless optics and the Transformer[J]. Optics Express, 2021, 29(23): 37962-37978.
- [7] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12) [2022-11-08]. <https://arxiv.org/abs/1706.03762>.
- [8] Parkhi O M, Vedaldi A, Zisserman A, et al. Cats and dogs[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 3498-3505.
- [9] Zhang Y E, Wu Z Y, Lin P Y, et al. Hand gestures recognition in videos taken with a lensless camera[J]. Optics Express, 2022, 30(22): 39520-39533.
- [10] Pan X X, Nakamura T, Chen X, et al. Lensless inference camera: incoherent object recognition through a thin mask with LBP map generation[J]. Optics Express, 2021, 29(7): 9758-9771.
- [11] Mengu D, Luo Y, Rivenson Y, et al. Analysis of diffractive optical neural networks and their integration with electronic neural networks[J]. IEEE Journal of Selected Topics in Quantum Electronics, 2020, 26(1): 3700114.
- [12] Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms[EB/OL]. (2017-08-25) [2022-11-09]. <https://arxiv.org/abs/1708.07747>.
- [13] 张华, 曹良才, 金国藩, 等. 基于压缩感知算法的无透镜数字全息成像研究[J]. 激光与光电子学进展, 2020, 57(8): 080001.  
Zhang H, Cao L C, Jin G F, et al. Progress on lensless digital holography imaging based on compressive holographic algorithm[J]. Laser & Optoelectronics Progress, 2020, 57(8): 080001.
- [14] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [15] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 9992-10002.
- [16] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[M]. The Netherlands: Now Publishers Inc, 2010.
- [17] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [18] Yang Y, Sun J, Li H B, et al. ADMM-CSNet: a deep learning approach for image compressive sensing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(3): 521-538.
- [19] Antipa N, Kuo G, Heckel R, et al. DiffuserCam: lensless single-exposure 3D imaging[J]. Optica, 2018, 5(1): 1-9.
- [20] Pan X X, Chen X, Takeyama S, et al. Image reconstruction with transformer for mask-based lensless imaging[J]. Optics Letters, 2022, 47(7): 1843-1846.
- [21] Monakhova K, Yurtsever J, Kuo G, et al. Learned reconstructions for practical mask-based lensless imaging [J]. Optics Express, 2019, 27(20): 28075-28090.
- [22] 王江涛, 王艳红, 武京治. 基于散射元件的无透镜偏振成像系统设计[J]. 激光与光电子学进展, 2020, 57(6): 257-262.  
Wang J T, Wang Y H, Wu J Z. Design of lensless polarization imaging system based on scattering elements [J]. Laser & Optoelectronics Progress, 2020, 57(6): 257-262.