

硕士学位论文

(学术学位论文)

基于卷积神经网络的多模光纤图像重构研究

RESEARCH ON IMAGE RECONSTRUCTION OF
MULTIMODE FIBER BASED ON
CONVOLUTIONAL NEURAL NETWORK

龚志宝

哈尔滨工业大学

2022年6月

国内图书分类号：TP3

国际图书分类号：681

学校代码：10213

密级：公开

硕士学位论文

基于卷积神经网络的多模光纤图像重构研究

硕士研究生：龚志宝

导师：王正佳副教授

申请学位：工学硕士

学科：仪器科学与技术

所在单位：仪器科学与工程学院

答辩日期：2022年6月

授予学位单位：哈尔滨工业大学

ClassifiedIndex:TP3
U.D.C:681

Dissertation for the Master Degree

**RESEARCH ON IMAGE RECONSTRUCTION OF
MULTIMODE FIBER BASED ON
CONVOLUTIONAL NEURAL NETWORK**

Candidate:	Gong Zhibao
Supervisor:	Asso. Prof. Wang Zhengjia
Academic Degree Applied for:	Master of Engineering
Speciality:	Instrument Science and Technology
Affiliation:	School of Instrumentation Science and Engineering
Date of Defence:	June, 2022
Degree-Conferring-Institution:	Harbin Institute of Technology

摘要

内窥镜可以深入物体内部探测其组织结构，在医学诊断和工业检测领域应用广泛。传统内窥镜的光纤是由多根单模光纤组成的光纤束，光纤数量决定内窥镜成像分辨率，只有增加单模光纤数量才能提高其成像分辨率。多模光纤允许多个光学模式并行传输，相同的接入口径具有更高的成像分辨率。而多模光纤作为一种散射介质，内部每种光学模式都以不同的相速度传播，在输出端得到的是一个散斑图像。由散斑图像恢复原始图像成为研究热点问题。现阶段将卷积神经网络用于多模光纤的散斑图像重构成为一种主流的方式，但是使用卷积神经网络需要大量数据来拟合重构模型。为克服这一缺点，本文在卷积神经网络模型的基础上使用迁移学习，大幅度降低重构模型数据集的同时，实现了散斑图像的高质量重构。

散斑图像的采集与处理。通过对多模光纤基本传输理论的分析，搭建了自动采集散斑图像的实验平台，并对采集到的散斑图像进行批量裁剪、等比缩小、通道转换操作，制作成散斑图像与标签图像一一对应的数据集，为将其输入卷积神经网络奠定基础。

散斑重构网络模型的研究。通过对用于图像分类的经典卷积神经网络模型的分析，搭建了VGG、ResNet、U-net三种散斑重构网络模型。通过实验对比了三种网络模型在可训练参数和训练时间上的优劣。ResNet散斑重构网络模型在可训练参数上具有较大优势，U-net散斑重构网络模型在训练时间上较为突出。为验证模型的鲁棒性，给散斑图像增加了高斯噪声和椒盐噪声。模型能够在一定污染值下对散斑图像进行重构。

散斑重构网络模型的迁移学习研究。通过使用手写字母数据集对源域网络模型进行预训练，目标域数据集使用原来的九分之一即可达到未使用迁移学习的重构效果，同时缩短了训练时间。通过对预训练模型的迁移位置和迁移后网络模型输出尺寸的研究，确定了合适的迁移位置和输出尺寸。

关键词：多模光纤；散斑图像；卷积神经网络；迁移学习

Abstract

Endoscopes can probe the microstructure of objects deep inside and are widely used in medical diagnostics and industrial production. The traditional endoscopic fiber is a bundle of fibers composed of multiple single-mode fibers, and the number of fibers determines the imaging resolution of the endoscope, and only by increasing the number of single-mode fibers can improve its imaging resolution. Multimode fiber allows multiple optical modes to be transmitted in parallel with the same access aperture with higher imaging resolution. As a scattering medium, each optical mode inside propagates at a different phase speed, and a speckle image is obtained at the output. Recovering original images from speckled images has become a hot topic in research. At this stage, the use of convolutional neural networks for speckled image reconstruction of multimode fibers has become a mainstream approach, but the use of convolutional neural networks requires a lot of data to fit the reconstruction model. In order to overcome this shortcoming, this paper uses transfer learning on the basis of convolutional neural network model, which greatly reduces the reconstruction model data set and realizes the high-quality reconstruction of speckle images.

Acquire and process the speckled images. Through the analysis of the basic transmission theory of multimode optical fiber, an experimental platform for automatically collecting speckle images is built, and the collected speckle images are batch cropped, equal ratio reduction, and channel conversion, which are made into a dataset corresponding to the speckle image and the label image, laying the foundation for inputting it to the convolutional neural network.

The study of scattered reconfiguration network models. Through the analysis of the classic convolutional neural network model for image classification, three scattered plaque reconstruction network models of VGG, ResNet and U-net are constructed. Experimentally, the advantages and disadvantages of the three network models in terms of trainable parameters and training time are compared. The ResNet speckle reconstruction network model has great advantages in trainable parameters, and the U-net speckle reconstruction network model is more prominent in training time. In order to verify the robustness of the model, Gaussian noise and pretzel noise are added to the speckle image, and the experimental results show that the model can reconstruct the speckle image under certain pollution values.

Transfer learning of speckle reconstruction network model. By pre-training the source domain network model using a dataset of handwritten letters, the target domain dataset uses one-ninth of the original refactoring effect without transfer learning, while reducing training time. Through the study of the migration location of the pre-trained

model and the output size of the post-migration network model, the appropriate migration location and output size are determined.

Keywords: Multimode fiber; Speckled image; Convolutional neural network; Transfer learning

目 录

摘 要	I
ABSTRACT	II
目 录	IV
第1章 绪 论	1
1.1 课题背景及研究的目的和意义	1
1.2 国内外研究进展	2
1.2.1 国外研究进展	2
1.2.2 国内研究进展	7
1.3 本文主要研究内容	10
第2章 多模光纤传输理论与数据采集	11
2.1 引言	11
2.2 多模光纤传输理论	11
2.2.1 阶跃多模光纤与渐变多模光纤	11
2.2.2 阶跃多模光纤传输条件	13
2.3 实验平台搭建	14
2.3.1 硬件系统的设计	14
2.3.2 自动采集程序的实现	16
2.4 散斑图像处理	17
2.4.1 散斑图像的批量裁剪	17
2.4.2 散斑图像的等比缩小	18
2.4.3 散斑图像的通道转换	18
2.5 本章小结	19
第3章 卷积神经网络原理分析	20
3.1 引言	20
3.2 典型的卷积神经网络模型分析与研究	20
3.2.1 LeNet-5网络模型	20
3.2.2 AlexNet网络模型	20
3.2.3 VGG网络模型	22
3.2.4 ResNet网络模型	23
3.3 卷积神经网络基本概念的工作原理	24
3.3.1 卷积层的工作原理	24

3.3.2 激活层的工作原理	27
3.3.3 池化层的工作原理	28
3.4 本章小结	28
第4章 网络模型搭建与实验研究	30
4.1 引言	30
4.2 散斑重构图像评价指标	30
4.2.1 散斑重构图像正确率	30
4.2.2 散斑重构图像相似性	32
4.3 基于VGG-16散斑重构网络模型及实验结果	33
4.3.1 基于VGG-16散斑重构网络模型	33
4.3.2 VGG散斑重构实验结果	35
4.4 基于ResNet散斑重构网络模型及实验结果	38
4.4.1 基于ResNet散斑重构网络模型	38
4.4.2 ResNet散斑重构实验结果	39
4.4.3 改进散斑重构网络模型及实验结果	40
4.5 基于U-net散斑重构网络模型及实验结果	42
4.5.1 基于U-net散斑重构网络模型	42
4.5.2 U-net散斑重构实验结果	44
4.6 图像噪声	45
4.6.1 噪声类别	45
4.6.2 噪声实验结果	47
4.7 本章小结	50
第5章 迁移学习的应用及实验结果	51
5.1 引言	51
5.2 迁移学习概述	51
5.3 三种网络模型的迁移学习	52
5.3.1 VGG散斑重构网络模型的迁移学习	52
5.3.2 改进ResNet散斑重构网络模型的迁移学习	54
5.3.3 U-net散斑重构网络模型的迁移学习	56
5.4 本章小结	58
结论与展望	60
参考文献	62
哈尔滨工业大学学位论文原创性声明和使用权限	66
致谢	67

第1章 绪论

1.1 课题背景及研究的目的和意义

光纤作为一种光信号的传输介质，具有传输信息容量大、保密性好、不受电磁场干扰和使用简单等诸多优点^[1,2]。利用光纤和其他配套设备制成的内窥镜是集光学、精密机械、电子技术等技术于一体的现代精密检测仪器装置，目前在医疗领域^[3-6]和工业领域^[7-10]应用非常广泛。

阿拉伯内科医生Albukasim（公元936-1013）是最早发明身体孔洞检测方法人之一^[11]。根据人体不同的天然孔道制成的医用内窥镜有胃肠镜、鼻咽喉镜、支气管镜、胸腔镜等，它们具有使用灵活、结构柔软和体积小等特点，可以深入人体内部对病变组织进行检测，较大程度上缓解了患者在诊疗过程中的痛苦，也是医生在诊疗过程中的重要辅助工具。在一些高精密仪器设备上，内窥镜可以通过窄小的孔洞深入设备内部，对设备内部的结构和损伤情况进行检测，可以在不拆卸设备的情况下发现问题的原因，提高了检测故障的效率。

当前内窥镜中使用的光纤主要是由多根单模光纤组成的光纤束^[12,13]。每根单模光纤只能传输一个模式，将多根单模光纤按照一定规律组装在一起构成光纤束就可以对图像进行传输，一根单模光纤只对图像的一个像素点进行传输，光纤束成像原理如图1-1所示^[14]。同一孔径的光纤束，单模光纤数量越多，纤芯之间的间隙越小，成像的分辨率就越高，光纤束成像对外部环境要求不高，环境因素造成光纤束的弯曲和扰动一般不会影响成像质量。若想提高成像的范围，就需要增加单模光纤的数量，增加单模光纤的数量就会增加光纤束的体积，体积增大会降低内窥镜的灵活性，同时在诊疗过程中也会给患者增加痛苦。



图1-1 光纤束传输原理^[14]

随着数字空间调制器的使用，多模光纤具有给内窥镜带来革命性变化的潜力。多模光纤的芯径大于单模光纤，并且允许光线以多个的角度射入光纤端面，并在光纤内进行传播，多模光纤可独立实现大范围区域成像。原则上，利用一根只有微米级别的多模光纤可以取代目前内窥镜中使用的毫米级别的光纤束。多模光纤成像与光纤束成像相比，极大地减小了成像媒介的尺寸，同时也实现了高分辨率的要求，

这可能会形成一种新的侵入性较低的内窥镜并且实现对当前传统内窥镜无法触及的部分进行高分辨率成像^[15]。但是多模光纤是一个散射介质，在电磁理论中，在光纤内部每种模式都以不同的相速度传播，这样在多模光纤的输出端得到的是一个随机的散斑图像，工作原理图如图1-2所示^[14]。

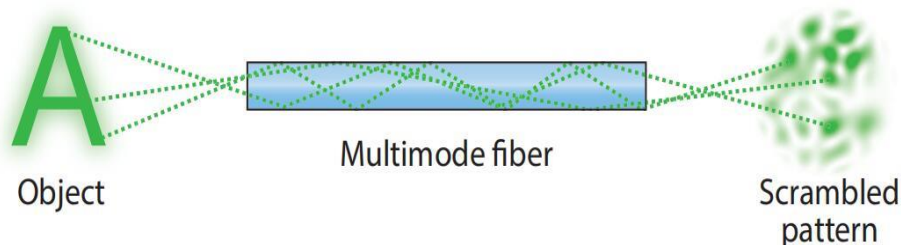


图1-2 多模光纤传输原理^[14]

如何由散斑图案恢复原始图像成为研究的热点问题。传统的方法主要有模拟相位共轭技术^[16,17]、数字相位共轭技术^[19,20]和传输矩阵技术^[21-23]等。传输矩阵技术是目前最为常用的技术，测量多个输入模式对应输出模式的振幅和相位，来构造一个从输入到输出的传输矩阵，这些方法都需要复杂的先验知识。而近年来卷积神经网络发展非常成熟，将卷积神经网络与不同的学科进行交叉融合，可以很好的解决一些传统的方法难以突破的瓶颈^[24-27]。卷积神经网络是一个典型的“端到端”的模型，即用输入和输出图像对网络模型进行训练，用训练好的网络模型就可对散斑图像进行重构，不需要复杂的先验知识，但是需要大量的数据来对模型进行训练。迁移学习^[28-32]可以利用训练数据、目标任务或者网络模型之间的相似性，将在原领域训练好的网络模型应用于目标领域，这将有望大大减少用于模型训练的数据量和训练时间。

1.2 国内外研究进展

1.2.1 国外研究进展

1990年，日本的Takao Matsumoto等人首先提出将神经网络用于光纤传输系统^[33]。他们提出两种方式：一种是由简单的光学元件和电神经网络组成的新型波分复用器；另一种是由多模光纤和神经网络形成新型光纤图像传输技术，在这两种技术中，多模光纤的传输模式在信号处理中都是非常重要的。两种方式应的结构图如图1-3所示。实验中使用的光纤芯径是150微米，数值孔径是0.4，长度为1米。该研究可以成功地将输入图像进行分类，证明了输入图像和输出图像存在一一对应的关系，由于硬件限制还无法实现对输入图像进行重构，但是该技术在当时已经可以用于一些传感器中。

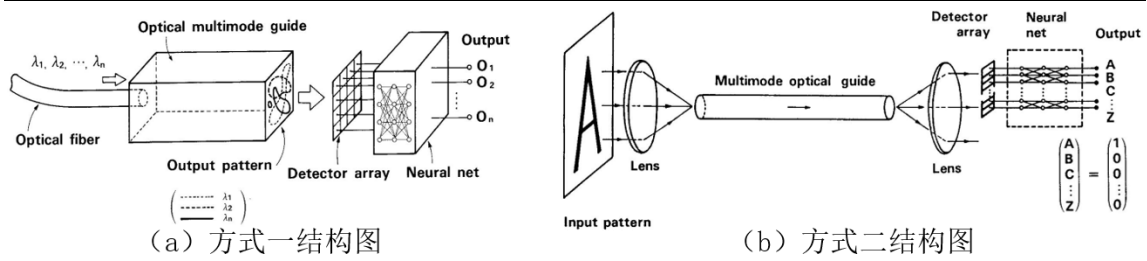


图1-3 两种方式结构图^[33]

将神经网络模型应用于多模光纤传输系统的思想已经存在了30多年，但是由于计算机的性能和网络模型深度的限制，神经网络模型在多模光纤传输系统中的应用一直停滞不前。近年来随着计算能力的不断进步，使得更加复杂的神经网络得以实现，涌现出非常多优秀的神经网络模型，他们可以处理更高维数据，在图像分类领域有着不俗的效果，将神经网络模型与具体的实际体相结合再次成为研究的热点问题。2018年，瑞士洛桑联邦理工学院的Babak Rahmani^[34]等人利用深度卷积神经网络ResNet和VGG网络模型对一段0.75米长的多模光纤的输入与输出关系进行学习，证明该网络可以有效的执行非线性反演任务，得到了98%的图像重构保真度和94%的图像预测准确率，并且创新的完成了网络可以重构不是训练集和验证集中的图片。图1-4 (a) 和图1-4 (b) 分别是通过CNN利用输出散斑振幅对输入振幅和相位的重构。

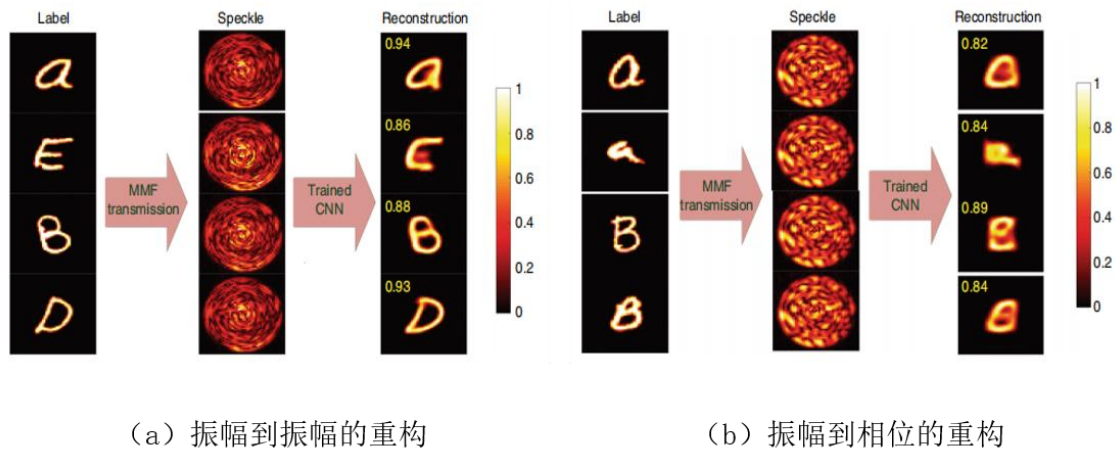


图1-4 ResNet网络重构结果^[34]

同年，同学院的Navid Borhani^[35]等人将VGG网络模型作为分类模型，U-net网络模型作为重构网络模型，网络模型结构如图1-5所示。利用两种模型主要研究了对于不同长度的多模光纤，散斑图像的分类准确率与重构效果。如图1-6 (a) 给出了不同光纤长度的分类准确率；图1-6 (b) 给出了0.1米,10米,1000米的GRIN光纤的输出散斑图的重建结果。从图中可以发现，随着光纤长度的增加，分类准确率和重

构效果均越来越差。

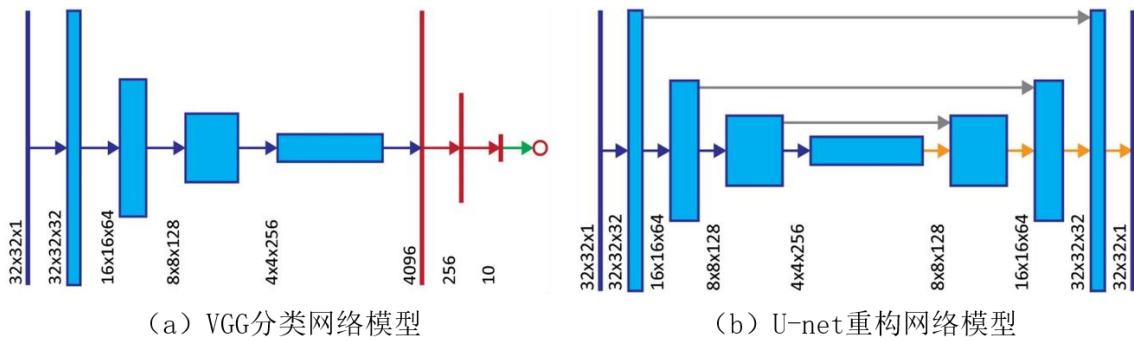


图1-5 两种网络模型^[35]

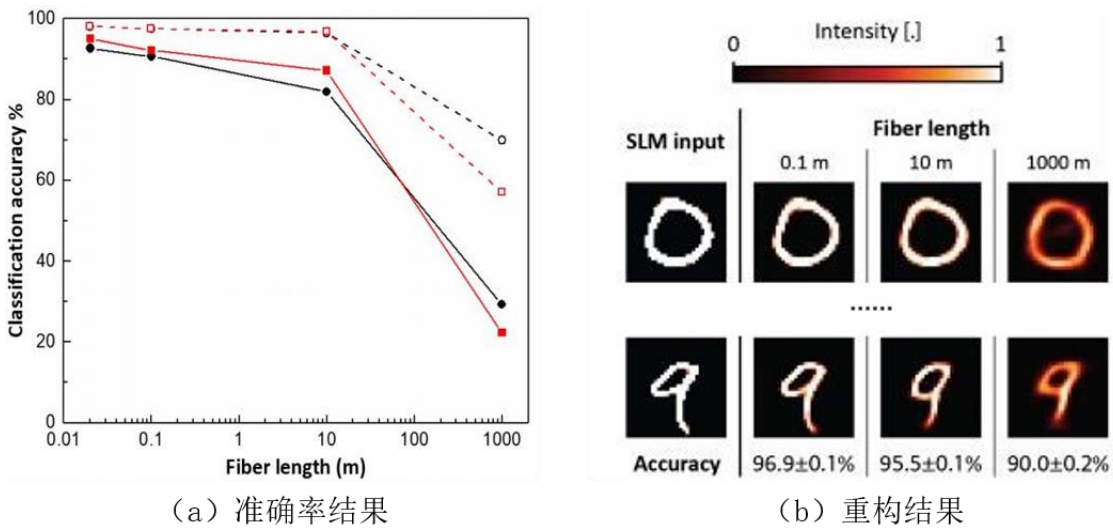


图1-6 准确率与重构结果^[35]

利用卷积神经网络模型对多模光纤输出的散斑图像进行重构的研究已经比较成熟，能够以较好的效果对散斑图案进行重构。多模光纤传输系统对环境和实验器件的要求都比较高，需要在环境和器件均相对稳定的状态下进行，如果存在一定的扰动还能否实现散斑图案的重构，相关学者对当光源存在波长漂移时和外在环境存在扰动时进行了相应的研究。2020年，瑞士洛桑联邦理工学院的Eirini Kakkava等人主要研究了在光源存在波长漂移的情况下，利用神经网络VGG网络模型对多模光纤输出的散斑图像进行分类^[36]，使用的VGG网络结构和图1-5 (a) 相同。他们使用波长范围700-1000纳米的激光器来产生一个带宽可控的波长漂移，来检验神经网络模型在不同状态下的性能以及存在的缺点。实验结果如图1-7所示，图中红线对应的是第一种情况，训练集的数据所用的光源波长为800纳米，测试集的数据所用的光源波长为803纳米、806纳米、812纳米和850纳米，从图中可以发现测试集数据的光源波长为803纳米时，分类的准确率下降到69%，当测试集数据的光源波长为

812纳米时，结果变得非常差，只有10%的分类准确率了。可见VGG网络模型的一般化只能是在波长漂移范围很窄的情况下有效，当波长漂移范围增加时，分类效果逐渐降低。图中的蓝线对应的第二种情况，训练集中的数据就包含光源波长存在漂移时的数据，随着波长漂移范围增加，分类准确率也有下降的趋势，但是即使漂移范围在100纳米的情况下，仍可以达到70%的分类准确率。这个结果证明了只要训练集中包含光源波长存在漂移时的数据，网络模型就可以表现出较好的鲁棒性。

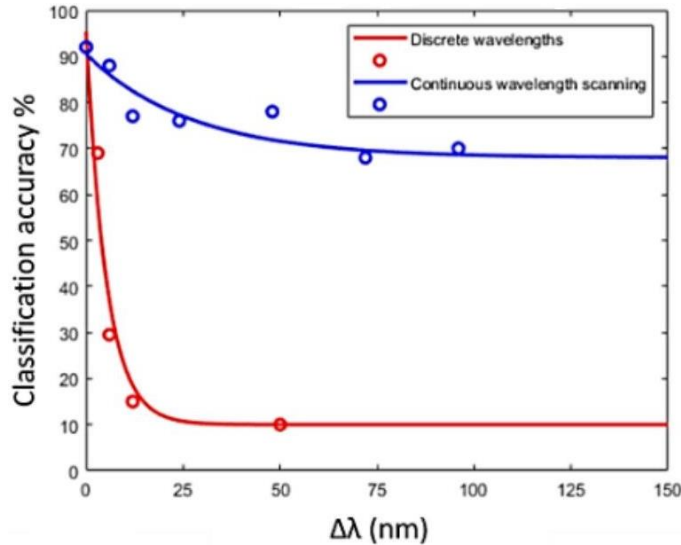


图1-7 训练集包含漂移与不包含漂移结果^[36]

多模光纤扰动的影响给电通信和内窥镜的使用带来了很大的困难。2021年，以色列希伯来大学拉卡物理研究所Shachar Resisi等人^[37]主要研究了在多模光纤存在强扰动的情况下的散斑图像重构。实验流程草图如图1-8所示，图像信息在数字显微镜上进行显示，用波长为632.8纳米的激光对其进行照射，将图像信息耦合进多模光纤，由CMOS相机对散斑图像进行采集。

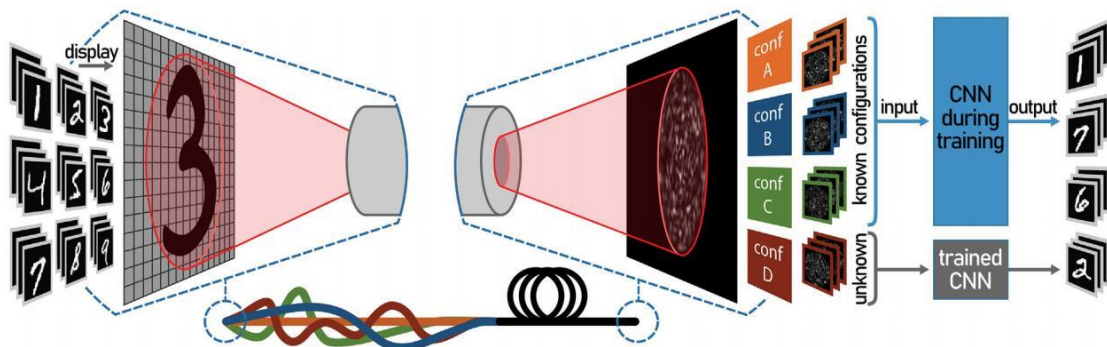


图1-8 重构过程示意图^[37]

对于相同的手写数字数据集，不同的扰动产生的散斑图案也是不相同的，用橙色、蓝色、绿色扰动模式下采集的散斑图像来对模型进行训练，用红色扰动模式下

采集的散斑图像对训练好的模型进行测试，实验中采用的网络模型是U-net网络模型，结构如图1-9所示。

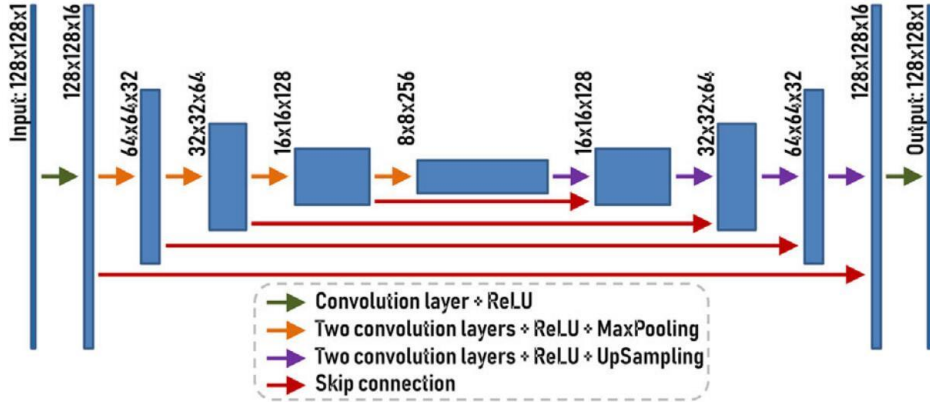


图1-9 网络结构图^[37]

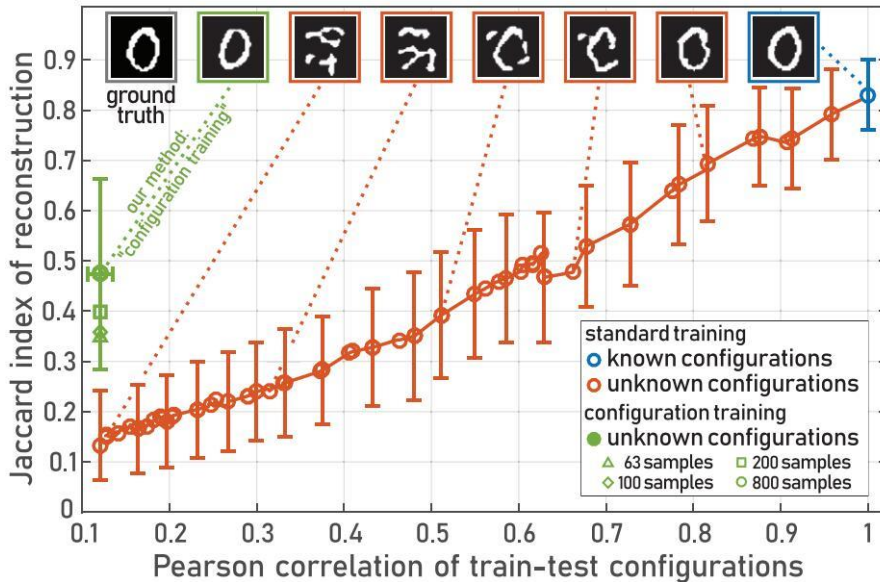


图1-10 实验结果^[37]

实验结果如图1-10所示，图中绿色表示该团队提出的方法，橙色的表示普通的方法，横轴表示训练集和测试集的相关性，纵轴表示重构结果，该值越接近于1表示重构效果越好。当训练集和测试集的相关系数越低的时候，普通方法的表现能力就越差，当橙色与蓝色重合，也就是训练集和测试集是同一个模式的时候，才能展现出较好的重构效果。而该团队提出的方法可以在训练集和测试集的相关系数很低的情况下给出非常好的重构效果。该实验证明了网络模型能够在多模光纤存在扰动的环境下对散斑图像进行重构，这一结果具有非常重要的实际意义，例如内窥镜的使用不可避免的会存在几何结构的变化。同年，澳大利亚阿德莱德大学Linhv Nguyen等人也利用深度学习对多模光纤存在扰动噪声时的表现进

行了研究^[38]。2021年，伦敦玛丽女王大学Pengfei Fan等人也利用神经网络对多模光纤传输空间分布的信息进行了研究，他们提出了一种基于深度学习的多模光纤的高密度信息传输框架，并且使用可扩展的半监督学习模型，使卷积神经网络模型实时适应时变的多模光纤的信息通道，来克服实验室环境的不稳定性^[39]。初步结果表明，深度学习在最大限度利用多模光纤的空间维度来传输数据信息方面具有很大的潜力。实验过程如图1-11所示，工作原理大致可描述为图1-11（a）为硬件原理图，依据硬件原理图采集到的神经网络模型所用的散斑数据，实验中激光器波长为532纳米。图1-11（b）表示利用神经网络模型对散斑图案进行重构。

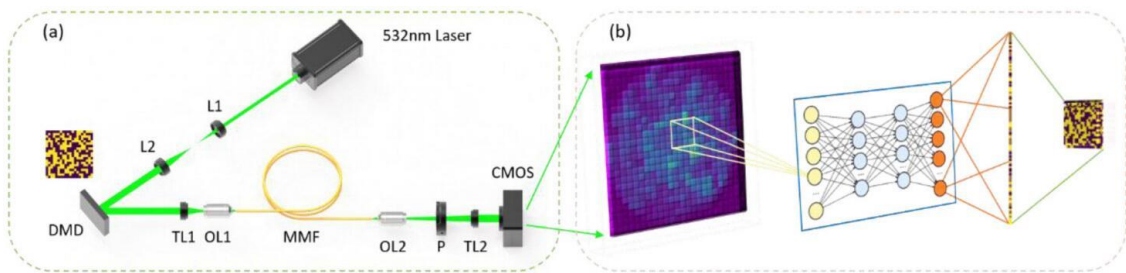


图1-11 实验流程图^[39]

1.2.2 国内研究进展

2018年，长安大学Ping Wang等人利用神经网络对多模光纤输出图像进行了分类研究^[40]。训练集中包含不同数量的图像在SVM和SpeckleNet两种方法上的分类准确率和训练时间的结果如图1-12所示。从结果图可以得出：（1）SpeckleNet在相同数据集上准确率高于SVM，但是训练时间也高于SVM；（2）随着训练集中数据量的增加，两种方法的准确率和训练时间均呈上升趋势。

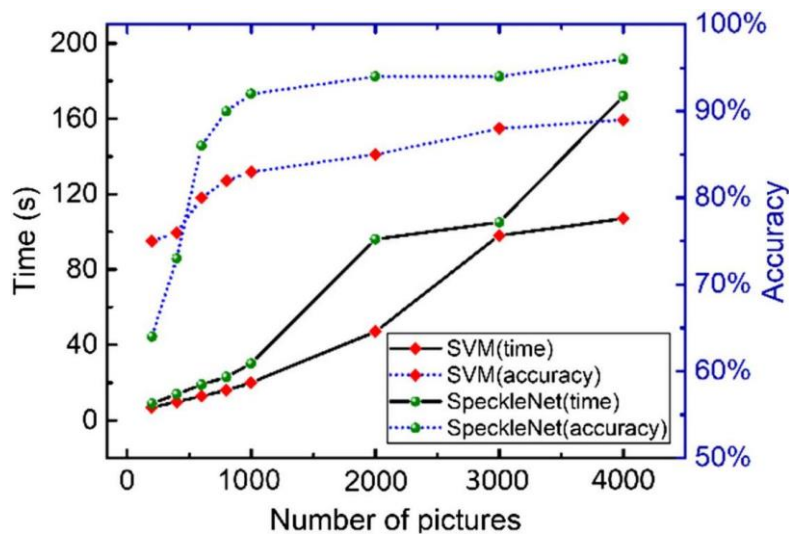


图1-12 准确率与训练时间结果^[40]

Ping Wang等人只研究了散斑图像的分类情况，未对散斑图像的重构情况做研究。2020年东北大学孟球等人采用加强U-net网络模型对多模光纤输出的散斑图案进行了分类和重构的研究^[41]。采用的网络结构总图如图1-13所示，a是卷积块，b是密集连接块，c是DenseUnet网络模型，d是依据VGG网络模型构造的深度卷积神经网络模型用于对重构图像进行分类。

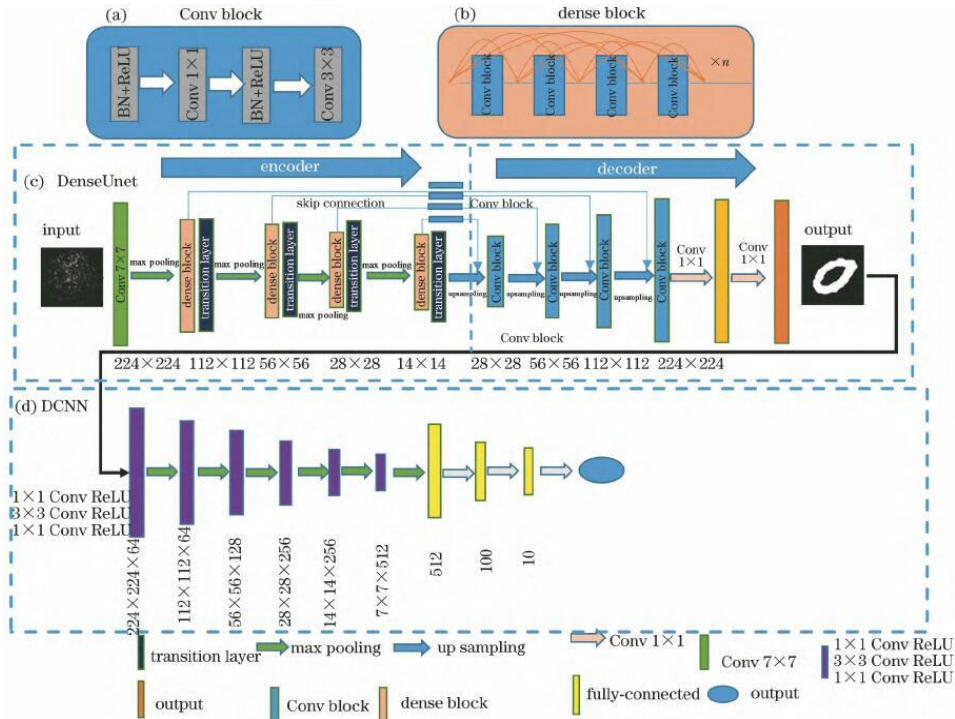


图1-13 网络结构图^[41]

其也对不同长度多模光纤输出散斑图像的重构结果做了相应的研究，结果如图1-14所示，从图中可以看出该模型在多模光纤长度为20米的时候仍然能够给出较好的重构结果。训练该模型需要10小时，端到端的微调需要5小时。

SLM input	fiber length			
	0.1 m	1 m	10 m	20 m

图1-14 不同长度多模光纤重构结果^[41]

2022年，清华大学Zhoutian Liu等人利用深度学习实现了全光纤高速图像检测，这对使用多模光纤的内窥镜实时观测生物组织的状态有关键影响^[42]。该研究与前面文献中提到的研究方法均不同，该方法主要是利用多模光纤中的模态色散将二

维空间信息转换为一维时间脉冲流，然后对神经网络进行训练，用训练好的神经网络对时间波形进行重构。能够以高质量对波形图像进行重构，重构结果如图1-15所示，每一个波形图像都对应一个二维图像，通过以非常高的频率记录波形信息就可以实现高帧率。

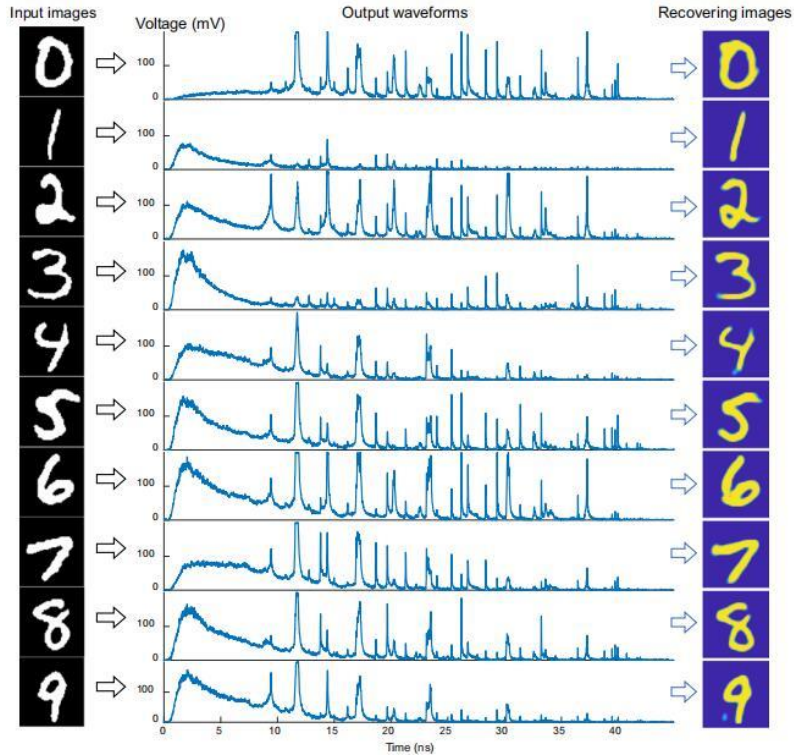


图1-15 重构结果^[42]

同时该方法与传统方法有明显区别，传统方法都是随着光纤长度的增加，利用散斑图像重构的效果下降，而该方法适合长距离多模光纤，不同长度的多模光纤的重构图像保真度如图1-16所示，

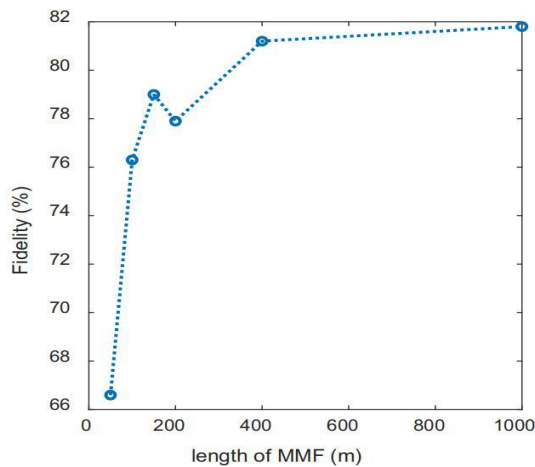


图1-16 不同长度多模光纤重构保真度^[42]

当多模光纤长度从1000米下降为400米时，重构图像保真度没有明显变化，但是当小于150米时，重构图像质量发生急剧下降，光纤的长度太短不能充分分割脉冲来将不同模式的信息分开。

总而言之，将神经网络模型用于多模光纤成像已经取得一定的进展。在系统存在一定噪声的情况下，仍能有较好的重构效果；能够实现高速图像检测，这对在内窥镜中使用该技术有重要影响。但是模型还存在训练时间长，需求数据量大的问题，所用图像较为简单等问题，如何降低模型的训练时间，减少需求的数据量，对复杂数据以较高保真度进行重构也应该是研究的关键问题。

1.3 本文主要研究内容

本课题主要的研究内容是研究不同卷积神经网络模型在对多模光纤输出散斑图案重构时的优缺点，并对散斑图案增加噪声来验证模型的鲁棒性，利用迁移学习来减少模型对数据量的需求，降低模型的训练时间。详细研究内容如下：

(1) 首先对多模光纤传输理论进行了详尽的分析和阐述，并从原理上分析了阶跃多模光纤和渐变多模光纤的特点和传输条件；搭建了相应的实验平台，编写了实验平台中需要的控制程序；对采集到的散斑图像进行了数据处理，以满足神经网络模型训练的要求。

(2) 对经典的卷积神经网络模型的工作原理做了详细的说明，为设计散斑重构神经网络模型提供了修改的理论依据。对网络模型中的卷积层、激活层、池化层进行了详细的说明，明确了设计卷积神经网络模型的基本思想。

(3) 给出了对散斑重构图像的评价指标；基于VGG-16网络模型，搭建了VGG散斑重构网络模型。对于VGG散斑重构网络模型分析了不同训练轮数对重构效果的影响，并且将网络模型进行了可视化。基于ResNet和U-net网络模型搭建了相应的散斑重构网络模型，对比了不同网络模型的优缺点。给散斑图像分别增加了高斯噪声和椒盐噪声，并给出了每种散斑重构网络模型在不同污染值的重构结果。

(4) 对迁移学习进行相关介绍，选择适合该课题的迁移学习方法；对搭建的三种网络模型进行迁移学习研究，从不同维度对迁移学习进行研究，观察实验结果。对比使用迁移学习和不使用迁移学习的重构结果，证明使用迁移学习的优势。

第2章 多模光纤传输理论与数据采集

2.1 引言

本课题主要思想是利用卷积神经网络对散斑图像进行重构，而利用卷积神经网络首先需要大量的数据，这也正是深度学习的特点。本章首先对多模光纤的传输理论进行了分析，为调整实验平台奠定理论基础；其次依据多模光纤传输理论搭建实验平台，并详细说明了实验平台中使用的各个器件的参数与功能，对使用计算机控制空间光调制器与CCD相机，实现散斑图像的自动采集做了详细介绍。最后对散斑图像进行处理，主要包括批量裁剪、等比缩小、通道转换，为下文研究卷积神经网络提供数据支撑。

2.2 多模光纤传输理论

光纤是一种纤芯折射率 n_1 大于包层折射率 n_2 的同轴圆柱形电介质波导。纤芯材料主要成分为 SiO_2 ，纯度高达99.999%，剩余成分为极少量的掺杂剂，常用的有 GeO_2 ，利用 GeO_2 来提高纤芯的折射率。包层材料一般也是 SiO_2 ，包层材料的作用就是把光限制在纤芯中。在包层材料外一般还会增加一层涂覆层，主要目的是为了增强光纤的柔软性、抗老化性和机械强度，光纤结构示意图如图2-1所示，光线主要在纤芯中传输，包层为光的传输提供了反射面，同时也起到了一定的机械保护作用。

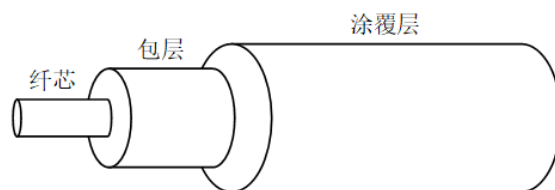


图2-1 光纤结构示意图

按照不同的分类方式，光纤可分为不同类别。按照芯径大小和内部模式数量不同，可将光纤分为单模光纤和多模光纤。按照折射率在纤芯和包层的径向分布情况，可将多模光纤分为阶跃多模光纤和渐变多模光纤。

2.2.1 阶跃多模光纤与渐变多模光纤

(1) 阶跃多模光纤

阶跃多模光纤在纤芯中折射率 n_1 为定值，到包层折射率突然变为 n_2 ，阶跃型多模光纤的折射率分布情况可用数学表达式表示为

$$n = \begin{cases} n_1 & r < a \\ n_2 & a \leq r \leq b \end{cases} \quad (n_1 > n_2) \quad (2-1)$$

在纤芯和包层界面处($r = a$), 折射率发生突变。光线在纤芯中以曲折形状传播, 如图2-2所示, 传光原理可以理解为, 光纤波导是一种透镜系统, 对于每种模式的光线都对应于一组焦距固定的透镜系统输入射光, 模式2的光线对应的透镜焦距短, 而模式3对应的透镜焦距长。光线在纤芯中的传输速度为

$$v = \frac{c}{n_1} \quad (2-2)$$

式中 c ——自由空间光速(m/s)。

纤芯中的折射率 n_1 保持不变, 所以路径2最曲折, 路径最长, 光线到大光纤末端所需的时间最长。路径1最短, 所需时间最短。由于光线沿不同路径到达光纤末端所需时间不同, 这就导致了在阶跃多模光纤的末端会形成散斑图案。

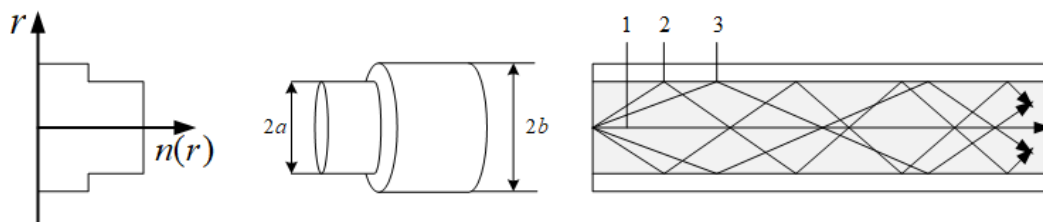


图2-2 阶跃多模光纤

(2) 渐变多模光纤

渐变多模光纤的纤芯折射率与阶跃多模光纤纤芯折射率不一样, 渐变多模光纤折射率 n_1 不是一个定值, 在纤芯中心处 n_1 最大, 沿径向外按抛物线形状逐渐减小, 直到包层减为 n_2 , 渐变多模光纤的折射率分布一般表示为

$$n = \begin{cases} n_1 \sqrt{1 - \Delta \cdot f\left(\frac{r}{a}\right)} & r \leq a \\ n_2 & a < r \leq b \end{cases} \quad (2-3)$$

式中 $\Delta = 1 - \left(\frac{n_2}{n_1}\right)^2$;

g ——折射率分布参数。

g 它决定了光线在纤芯中传播的形状。当 $g = \infty$ 时, 为阶跃光纤; 当 $g = 2$ 时, 为抛物线渐变多模光纤, 抛物线渐变多模光纤如图2-3所示。由公式 (2-2) 可得, 当 $n_{1,r \rightarrow 0}$ 时, n_1 最大, 速度最小, 所以沿光纤轴线传输的光线速度最慢。当 $n_{1,r \rightarrow a}$ 时, n_1 最小, 速度最大。所以虽然光线3到达末端传输距离最长, 但是它的传播速度最

快。光线在纤芯中沿不同的传输路径到达光纤末端的所需时间几乎相同，不会产生散斑图案。依据阶跃多模光纤和渐变多模光纤的特点，本次实验中采用阶跃多模光纤作为传输媒介，本文所提到的多模光纤均为阶跃多模光纤。

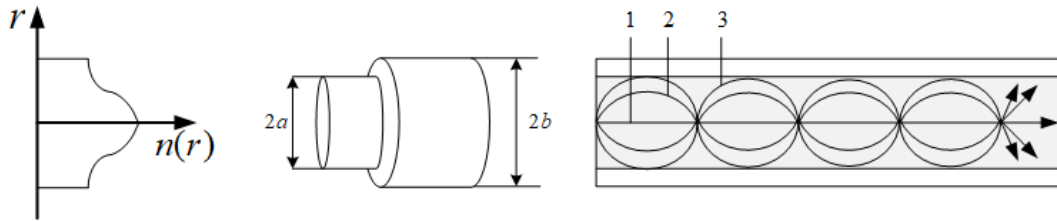


图2-3 渐变多模光纤

2.2.2 阶跃多模光纤传输条件

对于多模光纤， $\frac{2a}{\lambda}$ 远远大于 λ ，可用几何光学的光线理论近似分析光纤的传输原理与特性。由光学的基本知识可知，光波从折射率较大的介质射入到折射率较小的介质时，在边界会发生反射和折射，光波从折射率较大的介质以不同的入射角进入折射率较小的介质会出现以下三种不同的情况，如图2-4所示。当 $\theta_i < \theta_c$ 时，会同时发生反射和透射；当 $\theta_i = \theta_c$ 时，入射光沿交界面向前传播；当 $\theta_i > \theta_c$ 时，会发生全反射，没有折射光，只有反射光。

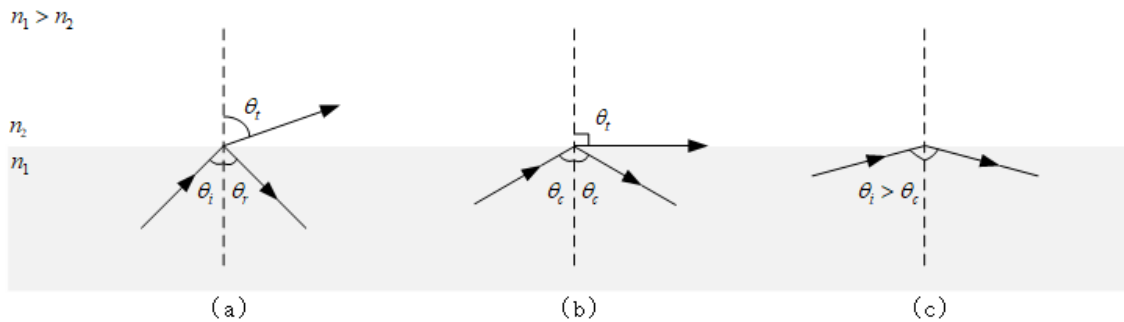


图2-4 光波从折射率较大介质射入折射率较小介质三种不同情况

光纤要想传输光波首先需要满足光线在纤芯和包层界面上发生全反射，其次还需满足传输过程中的相干加强条件，只有满足一定条件的光波才可以在光纤中进行有效的传输，这些指定的光波称为光纤模式，多模光纤支持多个传输模式。光线在多模光纤端面以不同角度 α 从空气入射到光纤中，只有光线在一定角度范围内射入光纤产生的折射光线才能够在光纤中传输。如图2-5 (a) 所示，光线在多模光纤端面入射角为 α ，在波导内光线与正交于光纤轴线的夹角是 θ_i ，当 $\theta_i > \theta_c$ 的光线会发生全反射，而 $\theta_i < \theta_c$ 的光线会进入包层。为使光线能够在多模光纤中传输，入射角 α 必须能够使进入多模光纤的光线在其内发生全反射，以折线形状向前传

播，如图2-5 (b) 所示，当 $\theta_i = \theta_c$ 时，对应的 α 角为最大的。再由空气入射纤芯界面，根据斯奈尔 (Snell) 定律可得

$$\frac{\sin \alpha_{\max}}{\sin \left(\frac{\pi}{2} - \theta_c \right)} = \frac{n_1}{n_0} \quad (2-4)$$

由全反射公式可知

$$\sin \theta_c = \frac{n_2}{n_1} \quad (2-5)$$

将公式 (2-4) 带入到 (2-3) 中可得

$$\sin \alpha_{\max} = \frac{(n_1^2 - n_2^2)^{\frac{1}{2}}}{n_0} \quad (2-6)$$

当光从空气进入光纤时， $n_0 = 1$ ，所以 $\sin \alpha_{\max} = (n_1^2 - n_2^2)^{\frac{1}{2}}$ 。定义数值孔径

$NA = (n_1^2 - n_2^2)^{\frac{1}{2}}$ ，因此可以使用数值孔径来表示光线最大入射角 $\sin \alpha_{\max} = NA$ 。角度 $2\alpha_{\max}$ 就是入射光线的最大射入范围。清楚光线在光纤中的传输原理，为调试硬件系统奠定基础。

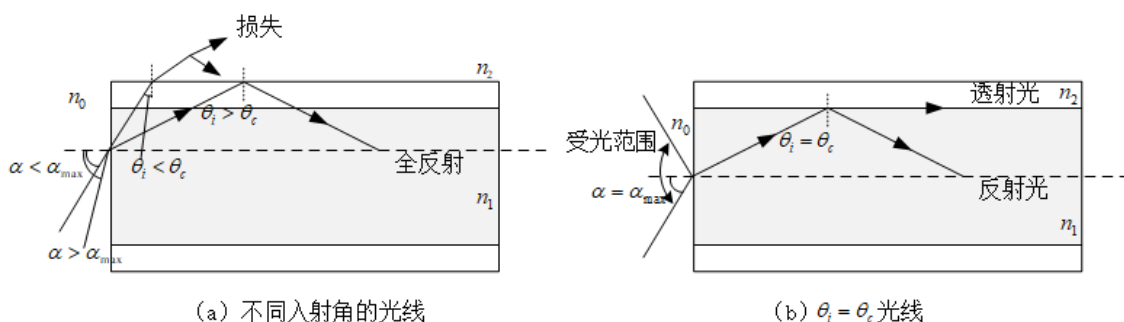


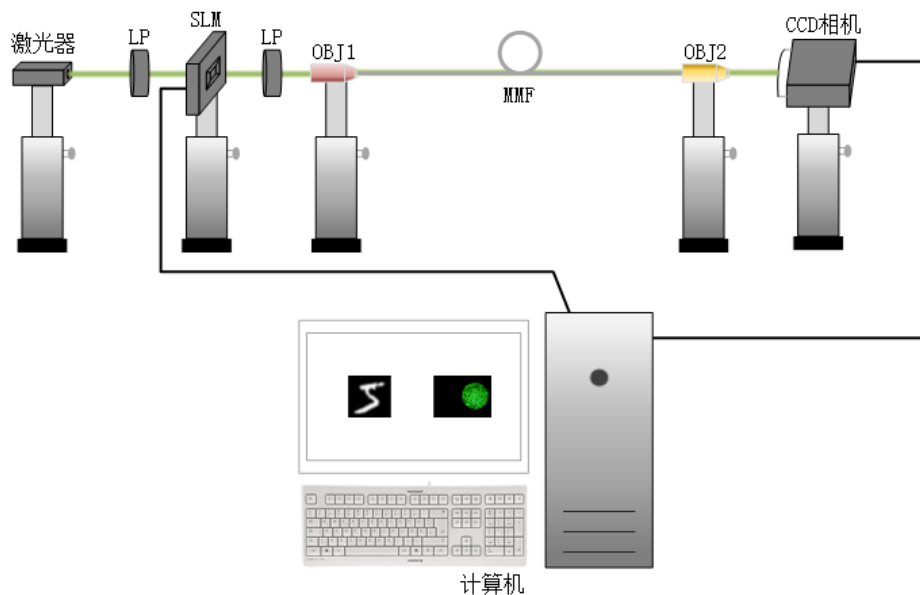
图2-5 光纤传输条件

2.3 实验平台搭建

2.3.1 硬件系统的设计

采集多模光纤输出端散斑图像的硬件系统设计的原理图如图2-6所示。多模光纤传输硬件系统主要包括脉冲激光器、偏振片、空间光调制器、多模光纤、CCD相机和计算机。脉冲激光器用来产生实验系统所需要的光源。计算机控制空间光调制器在光场中加载目标信息，偏振片用于对光照强度进行调节。平面显微物镜对光束进行耦合，平面显微物镜1将载有目标信息的光束耦合进多模光纤，平面显微物镜2将多模光纤输出端的散斑图像耦合进CCD相机。计算机控制CCD相机对散斑图像

进行大量采集。为避免自然光对采集散斑过程的干扰，实验需要在较黑暗的条件下进行。



LP:偏振片 SLM:空间光调制器 OBJ:平面显微镜 MMF:多模光纤

图2-6 硬件系统设计原理图

依据原理图搭建的实验系统如图2-7所示，在原理图的基础上在空间光调制器的输出端增加一片滤波片，降低光束的光照强度，保证CCD相机能够采集到带有目标信息的散斑图案，否则光束光照强度过大，CCD相机将采集到的是白色亮斑，无法确定光束载有的目标信息。



图2-7 实验系统实物图

其中重要的实验器件的简单介绍如下：

(1) 激光器：北京普华美光电公司生产，波长为532 nm，功率为80 mW，电源电压为12 V。

(2) 滤波片：美国Thorlabs公司生产，ND=3.0。

(3) 平面显微物镜：美国Thorlabs公司生产。多模光纤入射端的平面显微物镜参数与多模光纤出射端的平面显微物镜参数分别如表2-1。

表2-1

类别	目标类型	内嵌	放大倍数	数值孔径	工作距离	盖玻片厚度
入射端	平面消色差	空气	4 ×	0.10	18.5 mm	-
出射端	平面消色差	空气	10 ×	0.25	10.6 mm	-

(4) 多模光纤：上海昊量光电公司生产，型号为GZ200/260，NA=0.2，由公式(2-6)可知， $2\alpha_{\max}=23^\circ$ 。

(5) CCD相机：海康威视公司生产普通工业相机。

实验系统的调节与工作流程大致描述如下：脉冲激光器发出激光，光束透过空间光调制器，空间光调制器将图像中数字或字母信息耦合到光束上进行向后传输，偏振片和滤波片的主要作用是降低光束的光照强度，在光纤的输入端放置平面显微物镜1，将光束进行缩小。平面显微物镜1与光纤的距离不同，产生的光斑的大小就会不同，用白色纸片找一个合适的位置观察图像，当在纸片上可以看到一个带有图像信息的清晰偏暗的图案时，可以认为前半部分的光路调整是正确的，光束中包含了数据库中的图像信息。调整平面显微物镜与多模光纤输入端的距离，让带有数字或字母信息的光束恰好能入射到多模光纤的输入端。在多模光纤的输出端放置平面显微物镜2，对光束进行整形，调整平面显微物镜2与CCD相机的距离，在电脑屏幕上可以看到纹理清晰，明亮适中的散斑图像，并且在空间光调制器中的信息变化时，散斑图像也发生变化，可以基本确定整个光路的调试是正确的，采集的散斑图案数据可以用于后续的实验研究。

2.3.2 自动采集程序的实现

后续的实验研究需要大量的数据，所以设计自动采集程序成为关键问题，自动采集不仅可以保证图像的质量，还可以提高采集效率。设计自动采集程序的关键器件是空间光调制器和CCD相机，空间光调制器负责将图像的数字或字母信息耦合到光束上，CCD相机负责将多模光纤输出端的散斑图像采集到。设计的关键思想就是用计算机控制空间光调制器刷新数据库图像，在刷新的同时使用CCD相机进行采集，要求是刷新的图像与采集到的散斑图案必须是一一对应的，不能存在白色或

黑色的图像。

首先利用海康工业相机提供的接口，利用python语言编写相机自动采集程序，程序主要包含两大部分：一部分是python语言控制海康工业相机拍照程序，另一部分利用循环语句读取数据库中的图像信息，并在屏幕上显示。每张图像显示3秒后，激活相机采集程序，然后再使图像信息保留3秒，3秒之后显示下一张图像信息。目的就是使得采集到的散斑图像与显示的图像信息是一一对应的。采集过程示意图如图2-8所示。

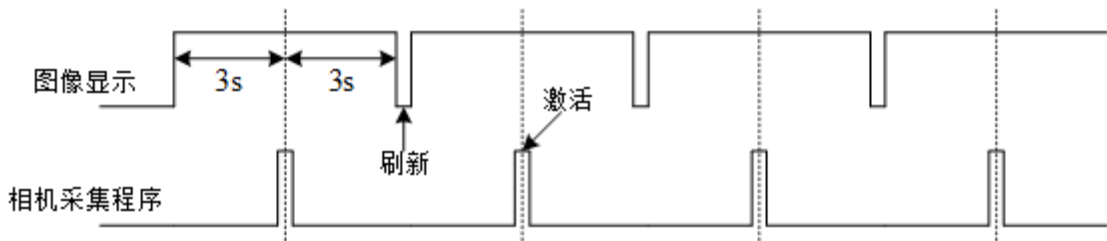


图2-8 采集过程原理示意图

2.4 散斑图像处理

2.4.1 散斑图像的批量裁剪

在完成实验平台的搭建以及自动采集程序的编写后，就可以对原始图像对应的散斑图像进行大量采集。原始图像采用手写数字数据集MNIST^[43]，该数据集由250人手写，共有70000张图像。原始图像与对应的散斑图像如图2-9所示。

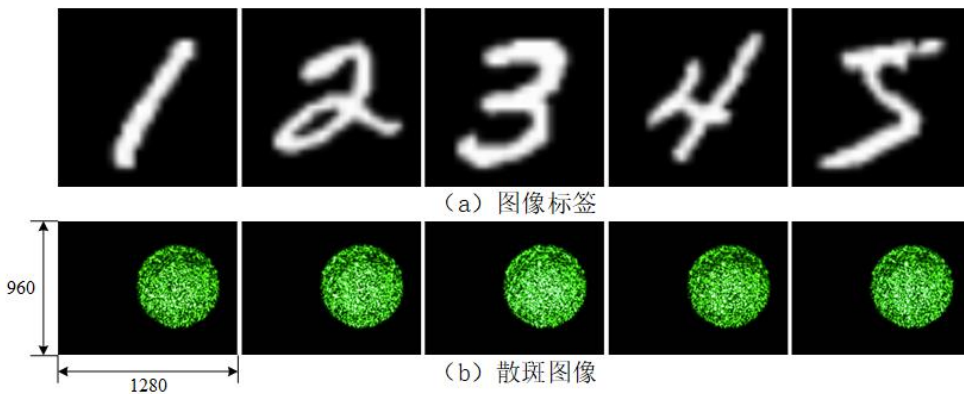


图2-9 图像标签与原始散斑图像

采集到的数据只有散斑图案可用于模型的训练，其余地方对实验并没有用。所以首先需要对采集图像进行批量裁剪，截取散斑图案，用于模型训练。批量裁剪程序使用python语言编写，主要包括读取图像、对图像进行裁剪、保存图像。原始图像的像素值为1280×960，批量裁剪之后的图像像素值为600×600。上图中的原始散斑图像批量裁剪之后如图2-10所示。

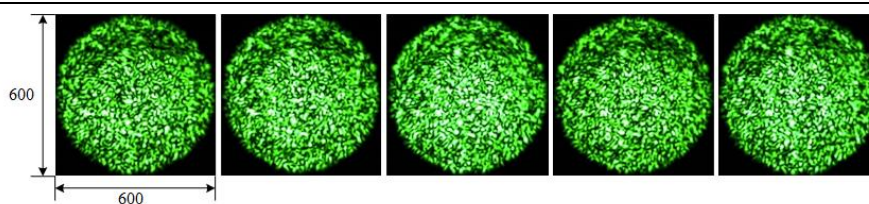


图2-10 裁剪之后散斑图像

2.4.2 散斑图像的等比缩小

图像在计算机中是以不同的像素存储的，在模型训练的时候，需要将图像的像素变小，来降低模型训练中的计算量。在降低像素的同时不能改变图像的纹理特征，但是当降低图像的像素的时候，图像将会变模糊。所以将图像降低到多大的像素需要结合实验效果等实际情况来综合考量。图2-11为将图像等比缩小为128×128。

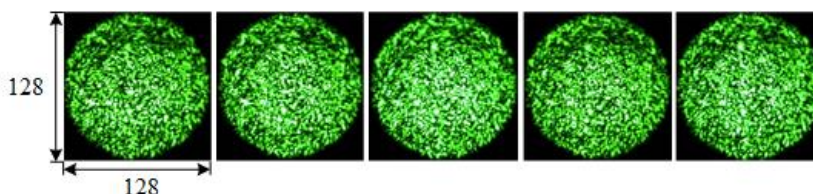


图2-11 等比缩小后散斑图像

2.4.3 散斑图像的通道转换

采集到的散斑图像是RGB三通道的彩色图像，每一个通道中的每一个像素值的大小都是0-255。本实验与识别人脸等图像识别有区别，无需保留图像的三个通道，可以将图像的三个通道转换为单通道，也就是将彩色图像转化为黑白图像，不改变图像的纹理特征即可。这样可以减少图像包含的数据量，降低模型训练的计算量，转换过程原理图如图2-12所示。

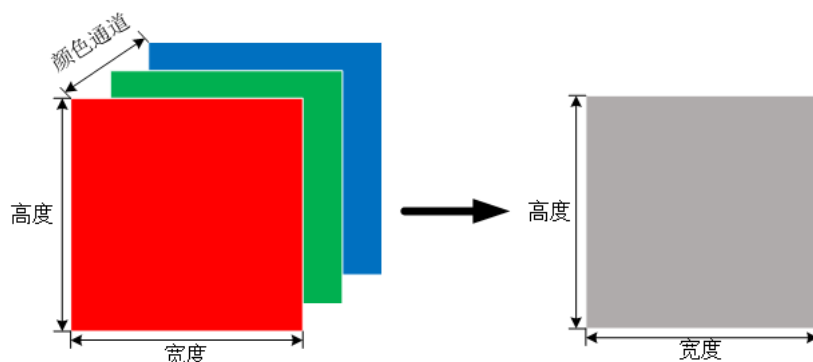
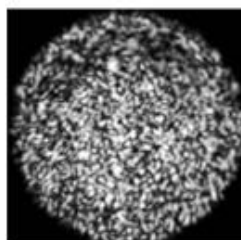


图2-12 通道转换原理图

图2-13 (a) 为数字5对应的散斑图像转换为灰度图，图2-13 (b) 为灰度图对应的矩阵，像素大小为16×16。灰度图像中0代表黑色，255代表白色。



(a) 散斑图像

```

[[ 1  1  1  1  1  1  1  15  3  9  2  1  1  1  1  1]
 [ 1  1  1  1  1 20 114 179 75 112 56 58  5  1  1  1]
 [ 1  1  1 23 171 142 185 44 56 40 66 30 155 66  1  1]
 [ 1  1 40 92 177 47 131 45 170 50 48 74 27 64 173  1]
 [ 1  5 83 101 110 63 22 246 60 146 69 75 111 115 59 37]
 [ 1 37 162 74 58 124 204 221 96 138 207 74 109 114 169 38]
 [ 1 123 89 22 16 140 42 115 112 40 161 177 173 173 166 20]
 [ 1 86 118 56 128 146 163 146 157 185 88 108 76 145 62 55]
 [ 3 50 55 92 203 166 38 200 53 51 52 55 185 167 107 131]
 [ 1 108 225 69 91 127 222 63 182 42 189 243 135 92 161 51]
 [ 1 126 123 182 61 232 184 155 156 118 214 204 122 87 29 184]
 [ 1 33 102 187 165 159 129 47 165 244 141 200 166 58 90 52]
 [ 1  1 34 192 84 167 188 212 211 119 186 156 124 146 107 15]
 [ 1  1 33 105 165 78 186 130 94 182 109 69 90 167 45  1]
 [ 1  1  1 28 31 156 69 137 75 182 85 79 132 102  1  1]
 [ 1  1  1  1  2 137 35 43 98 59 55 134  1  1  1  1]]
    
```

(b) 矩阵

图2-13 散斑图像与对应矩阵

2.5 本章小结

针对卷积神经网络需要大量数据的问题，本章分析了多模光纤的传输理论，依据传输理论搭建了实验系统，采集了实验数据。本章主要完成了以下工作：

(1) 首先从原理上分析了阶跃多模光纤和渐变多模光纤的特点，根据两种多模光纤的特点，阶跃多模光纤适合作为本实验的传输介质。对选择的阶跃多模光纤的传输条件进行了理论推导，为后续调整实验设备提供相应的理论支撑。

(2) 完成对实验器件的选型，搭建了采集散斑图像实验平台，编写了实验平台中需要的计算机控制CCD相机和空间光调制器的自动采集程序，最终完成了对散斑图像的采集。

(3) 对采集到的散斑图像进行了裁剪、缩小、通道转换处理，以满足神经网络模型训练的要求。制作了卷积神经网络模型需要的散斑与标签一一对应的数据集。详细说明了散斑图像数据在输入神经网络模型前的形式。

第 3 章 卷积神经网络原理分析

3.1 引言

目前深度学习的发展已经非常成熟，利用卷积神经网络解决相应的物理问题也已经是研究的热点。本章首先介绍了经典的卷积神经网络模型，经典的卷积神经网络有 LeNet-5^[43]、AlexNet^[44]、VGG^[45]、ResNet^[46]等。通过对经典卷积神经网络模型的认识，为搭建散斑重构网络模型做好方向指引。其次对神经网络模型中使用的具体概念做了详细的说明，从本质上理解卷积神经网络是如何工作的，为设计优化模型提供理论基础。

3.2 典型的卷积神经网络模型分析与研究

3.2.1 LeNet-5 网络模型

不包括输入层，LeNet-5 一共包含 7 层，每一层都包含不同数量的训练参数。LeNet 的结构图如图 3-1 所示。

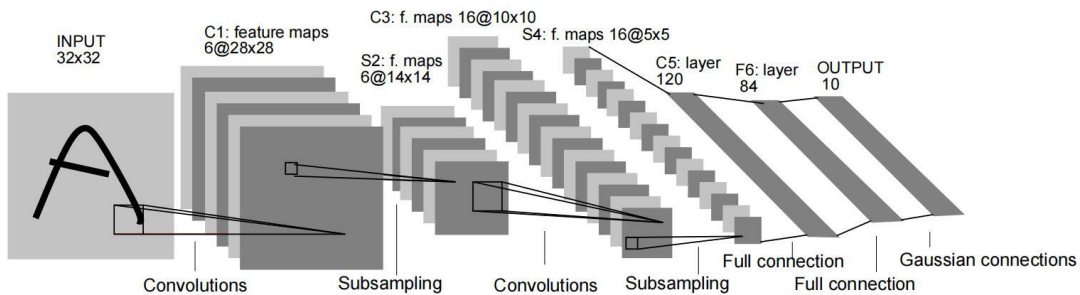


图 3-1 LeNet 网络结构图^[43]

在 LeNet-5 网络结构中，主要有卷积层、池化层、和全连接层三种操作方式。输入图像 A 的尺寸大小是 32×32 ，经过第一个卷积操作后，生成 6 个 28×28 的特征图。再经过一个池化操作后，将 6 个特征图的尺寸大小变为 14×14 。在池化层中，采用平均池化，将前层 4 个神经元通过平均池化，变成一个神经元，通过池化操作将图像的尺寸变成原来的一半。然后经过第二个卷积操作，输出 16 个 10×10 的特征图，经过池化操作将特征图的尺寸降低为 5×5 ，再经过两个分别具有 120 和 84 个连接单元的全连接层，最后对应 10 个数字的 10 个分类输出层。LeNet 是最早应用于数字识别的卷积神经网络，它对深度学习的发展起到了非常重要的作用。

3.2.2 AlexNet 网络模型

2012 年，加拿大多伦多大学的 Alex Krizhevsky 等人提出了 AlexNet，它使沉

寂已久的卷积神经网络又变得火热起来。他们使用更深、更大的卷积神经网络对包含 1000 个类别的 120 万张图像进行分类识别，在测试集上使用两种常用的评价指标 top-1 和 top-5 的错误率分别是 37.5%和 17.0%。AlexNet 的网络模型结构图如图 3-2 所示。

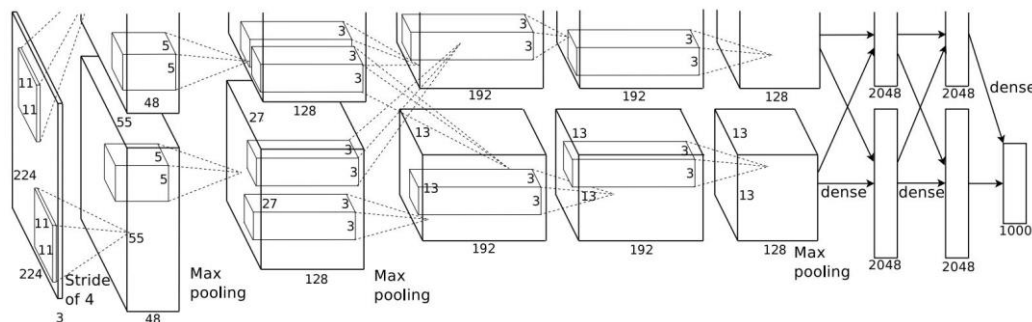


图 3-2 AlexNet 网络结构图^[44]

AlexNet 包含 5 个卷积层、3 个全连接层和输入层以及包含 1000 个类别单元的输出层。网络的输入层图像尺寸是 $224 \times 224 \times 3$ ，第一个卷积层包含有 96 个滤波器内核，尺寸是 $11 \times 11 \times 3$ ，步长是 4，所以对应生成 96 个特征图。第二个卷积层对应的滤波器内核尺寸是 $5 \times 5 \times 48$ ，生成 256 个特征图。前面两层的卷积层之后都有一个最大池化层，然后后面是 3 个卷积层相连，第五个卷积层之后又是一个最大池化层，然后是 2 个含有 4096 个单元的全连接层，最后是包含有 1000 个类别的输出层。AlexNet 在很多方面做出了创新性研究，主要包括以下几方面：

(1) 成功使用了 ReLU 非线性激活函数

相比于使用传统的饱和非线性激活函数，使用非饱和非线性激活函数在深度卷积神经网络的训练速度上有明显的优势。

(2) 使用多个 GPU 实现加速训练

将网络部署到两个 GPU 上，实现并行计算，利用计算机的硬件加速卷积神经网络模型的训练速度。

(3) 局部响应规范化 (Local Response Normalization)

规范化的本质就是将较大的输入抑制到指定的范围内。ReLU 函数是无边界函数，利用规范特性将无边界输出抑制到有边界的范围内，从而提升高频特征的提取效率。

(4) 重叠池化 (Overlapping Pooling)

步长 s 小于池化核 z ，这样就会得到重叠池化，使用重叠池化操作可以使输出彼此有重叠和覆盖，对特征提取的多样性进行丰富。

(5) 数据增强机制

深度学习的一个很大的特点就是使用大数据对模型进行训练，但是现实生活中，很难找到数量庞大的数据库来满足模型的训练需求。通过旋转、翻转、裁剪、缩放、平移、抖动等操作对图像进行几何变换，实现将一张图像变为多张图像，增加数据量，避免模型出现欠拟合现象。

(6) 使用随机失活机制

为了避免过拟合的发生，在前两个全连接层中使用了随机失活机制。

3.2.3 VGG 网络模型

VGG 网络模型主要研究了网络的深度对模型准确率的影响，网络模型的深度如图 3-3 所示，从左至右深度依次增加，每部分增加的深度用加粗字体表示。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 3-3 不同深度 VGG 网络模型构成图^[45]

VGG 成功地实现了 16~19 层的卷积神经网络，并且验证了增加网络模型的深度，能够在一定程度上影响网络的最终性能，使分类错误率大幅下降。上图中的 D 列和 E 列分别对应 VGG-16 和 VGG-19。VGG-19 比 VGG-16 更深，多出的三层为卷积层，VGG-19 的性能表现稍好于 VGG16，但是与此同时模型的参数增加了很多。研究团队将模型的网络层数定格在 VGG-19，不再继续增加更多层数。VGGNet 由 5 组卷积层和 3 个全连接层以及一层 Softmax 输出层组成，其中每组卷积中的卷积层个数不同，网络模型的深度不同也主要体现在卷积层的个数不同。每组卷积层之间使用最大池化操作，降低特征图的尺寸，所有隐含层的激活函数均采用 ReLU 函数。VGGNet 借鉴了 AlexNet 网络结构，但是采用小卷积核和多卷积层是 VGGNet 的一大亮点，采用小卷积核和多卷积层主要有两大优势，首先是可以减少网络参数，其次采用更多的卷积层就会进行更多次的非线性映射，从而提高网络模型的数据拟合能力。

3.2.4 ResNet 网络模型

随着 AlexNet 网络模型的提出，从 2012 年到现在，人工智能、计算机视觉领域发生了巨大的变化，ResNet 就是发生质的变化的典型代表，ResNet 太具有创新性、革命性，对人工智能深度学习产生了深远的影响。VGG 网络模型研究了网络模型的深度对分类准确率的影响，随着网络模型深度的加深，分类准确率会有一些的提高，但是并不是网络模型层数越深，分类准确率就越好，何恺明等人在深度残差学习在图像识别中的应用文章^[46]中指出，56 层的普通网络模型比 20 层的普通网络模型在 CIFAR-10 的测试集和训练集上的表现效果更差，也就是说随着网络模型的深度增加，会出现网络退化的问题。何恺明等人提出的 ResNet 网络模型成功解决了这一问题，解决这一问题的关键是采用了残差块，如图 3-4 所示，整个网络模型由多个这样的残差块构成。

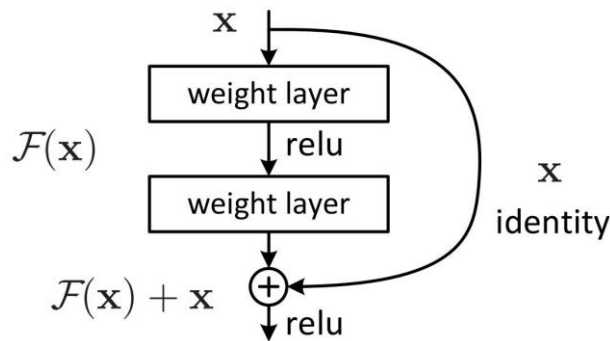


图 3-4 残差块^[46]

利用残差块可以实现不用拟合底层映射，只需拟合残差即可。通俗的理解就是

一部分输入通过卷积层、激活函数等操作，一部分输入什么都不改变，最后实现将两部分求和。如果恒等映射是最优的，将残差块置 0 比拟和非线性层堆叠的映射更加容易。残差网络减轻了网络的学习负担，网络模型需要学习输入与输出的差别信息，有效的解决了深层网络训练困难的问题，同时还保证了良好的性能。图 3-5 是 VGG-19、34 层一般卷积网络模型与 34 层残差网络模型结构图。

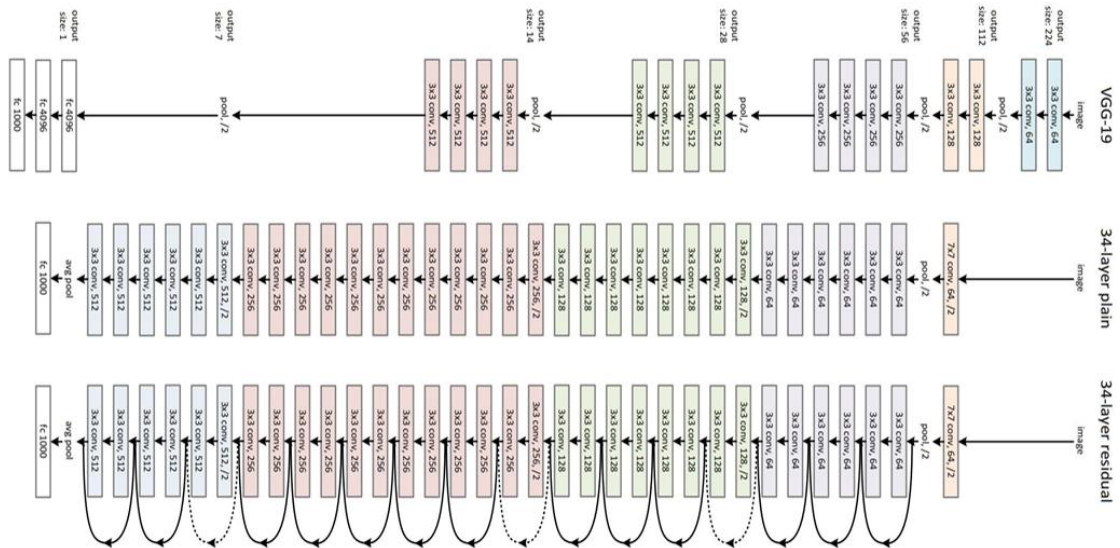


图 3-5 VGG-19、34 层一般卷积网络模型与 34 层残差网络模型结构图^[46]

3.3 卷积神经网络基本概念的工作原理

3.3.1 卷积层的工作原理

上一小结中介绍了典型的卷积神经网络模型，而卷积神经网络的核心就是卷积操作。在数学中卷积是这样定义的

$$h(x) = \int_{-\infty}^{+\infty} f(\tau)g(x-\tau)d\tau \quad (3-1)$$

$h(x)$ 称为函数 f 与 g 的卷积， f 为输入函数， g 称之为滤波器或者卷积核，输出结果 $h(x)$ 称之为特征映射或者是特征图谱，卷积操作简记为 $h(x) = (f * g)(x)$ 。公式

(3-1) 中的输入函数是连续的，但是在数字计算机处理场景下，卷积操作是不能处理连续信号的，所以需要将连续函数离散化，对于离散信号，卷积操作可用下式表示

$$h(x) = \sum_{\tau=-\infty}^{x+\infty} f(\tau)g(x-\tau) \quad (3-2)$$

在 2.3.3 节提到，散斑图像在计算机眼中是数字矩阵，对于灰度图像而言，每个元素都是 0~255 的像素值。卷积操作在图像处理中的功能主要是利用卷积核对

输入图像进行滤波操作，从而达到对特征进行提取的目的。通过卷积操作可以容易地从输入的数据矩阵中学到图像的特征，并且能够保留像素间的对应关系。下面详细说明在二维图像中使用卷积的过程。为了便于描述，将输入图像数字矩阵大小取为 6×6 ，卷积核大小取为 3×3 ，输入图像和卷积核中的各个图像取值均为随机取值，如图 3-6 所示。

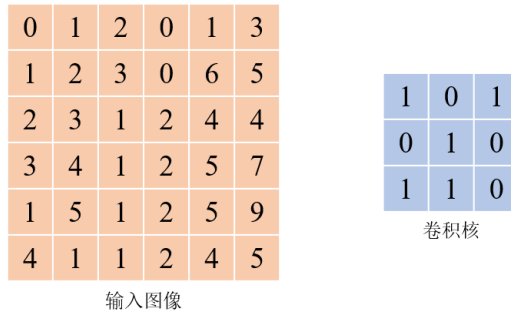


图 3-6 输入图像与卷积核

二维矩阵的卷积操作就是利用卷积核去对整个输入图像进行扫描，在卷积核的作用下，输入图像中的所有像素点会被线性变换组合，形成下一层的神经元节点，也称之为特征图谱。利用卷积核矩阵在原始图像上进行从左到右、从上到下的滑动，每次滑动 s 个像素，滑动的距离 s 称为步长，卷积操作过程如图 3-7 所示，步幅 s 取 1。

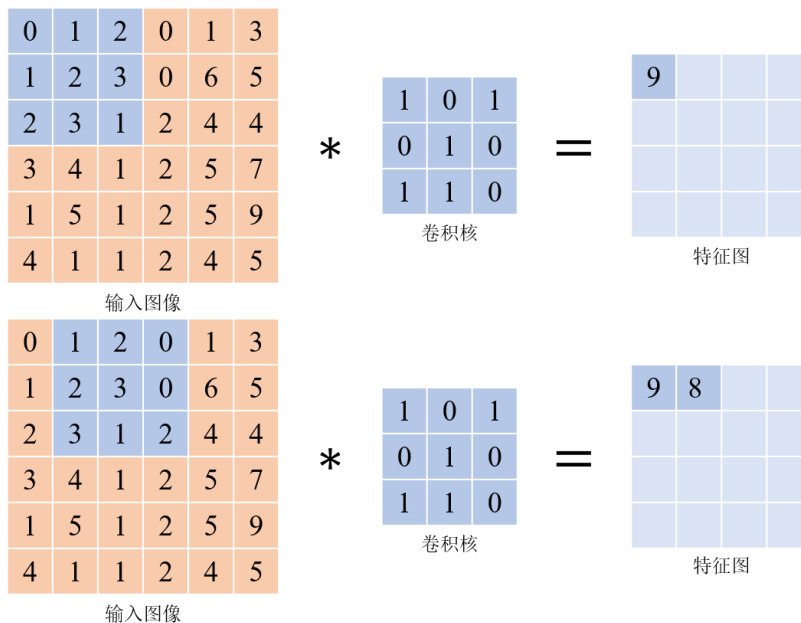


图 3-7 卷积操作过程

在上图的卷积操作过程中，输入图像的尺寸大小是 6×6 ，经过一个大小 3×3 的卷积核卷积之后，得到的特征图尺寸为 4×4 。在步长 s 为 1 时，输入图像和特征图

的尺寸关系为 (3-3)

$$\begin{aligned} H_{out} &= H_{in} - H_{kernel} + 1 \\ W_{out} &= W_{in} - W_{kernel} + 1 \end{aligned} \quad (3-3)$$

式中 H_{in} 、 H_{out} ——分别为输入图像、输出特征图像的高度；

W_{in} 、 W_{out} ——分别为输入图像宽度、输出图像宽度；

H_{kernel} 、 W_{kernel} ——分别为卷积核的高度、宽度。

对于卷积操作，如果我们所取的步幅 s 越大，那么得到的特征图的尺寸就会越小。当所取步幅与卷积核和输入图像的大小不匹配时，就无法完成卷积操作，例如当上述卷积过程取步幅 s 为 2 的时候，当向右第三次移动的时候，输入图像会缺少一列与卷积核进行卷积操作，这就会导致边界元素与卷积核卷积的次数少于非边界元素与卷积核卷积的次数，这时就需要对输入图像进行相应的填充。在卷积神经网络模型中常用的填充方式是等大填充 (Same Padding)，在等大填充这种模式下，会在输入矩阵的周围填充若干圈合适的值，在实际操作过程中一般取 0，这样就会使得输入矩阵边界处的矩阵大小刚好与卷积核的大小匹配，使得输入数据中的每个像素都可以参与卷积运算，从而保证输出的特征图与输入图像的尺寸保持一致。等大填充的卷积过程如图 3-8 所示。

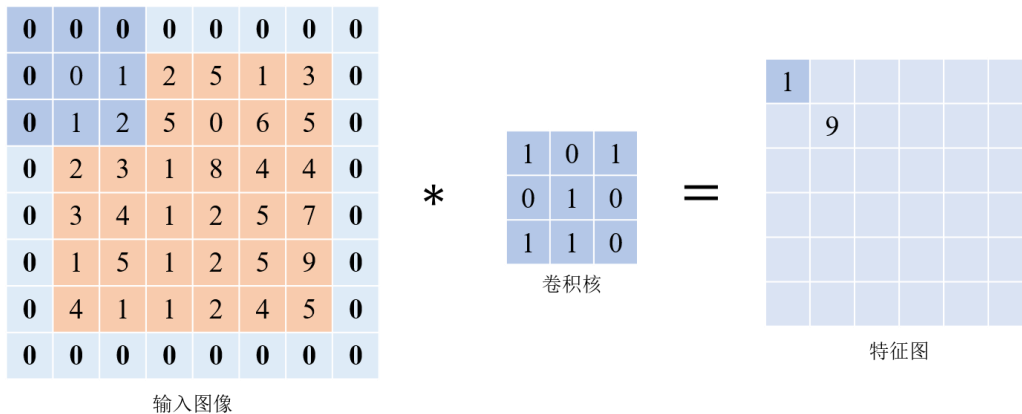


图 3-8 等大填充

输入图像补零的数目和卷积核移动的步长都会影响输出特征图的尺寸大小，其关系可以表示为

$$\begin{aligned} H_{out} &= \left\lfloor \frac{H_{in} + 2H_{padding} - H_{kernel}}{S_H} \right\rfloor + 1 \\ W_{out} &= \left\lfloor \frac{W_{in} + 2W_{padding} - W_{kernel}}{S_W} \right\rfloor + 1 \end{aligned} \quad (3-4)$$

式中 $H_{padding}$, $W_{padding}$ ——分别为在垂直维度上的补零高度、水平维度上的补零宽度;

S_H , S_W ——分别为在垂直维度上、水平维度上的步长大小;

$\lfloor * \rfloor$ ——向下取整。

在一般图像中,描述物体特征的像素之间的差距较大,像素值变化比较明显,通过卷积操作,可以过滤掉那些变化不明显的信息,这就是卷积核能够检测出特征的本质原因。

3.3.2 激活层的工作原理

从宏观角度来看,人工神经网络主要可以分为两个大的方面,一种是显层,另一种是隐层。显层就是输入层和输出层;隐层主要是卷积层、激活层、池化层等。如果隐层包含足够多的神经元,那么神经网络模型就能以任意精度逼近任意复杂的连续函数。在神经网络模型中,神经元与神经元的连接都是基于权值的线性组合,如果没有非线性的部分,在模型上叠加再多的网络层,最终都可能退化为一层神经网络,从而使神经网络的学习能力荡然无存。所以,在卷积神经网络中添加激活层与普通前馈神经网络一样,都是为了提高神经网络的表征能力而添加的一种非线性变换,有了这种非线性变换,卷积神经网络才具备分层的非线性映射学习能力。因此,激活函数是卷积神经网络中不可缺少的一部分。

常用的激活函数主要有 Sigmoid 函数、Tanh 函数、ReLU 函数。三种激活函数的表达式如下及其曲线图像如图 3-9 所示。

Sigmoid 函数:

$$S(x) = \frac{1}{1+e^{-x}} \quad (3-5)$$

Tanh 函数:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3-6)$$

ReLU 函数:

$$f(x) = \max(0, x) \quad (3-7)$$

由公式(3-5)和公式(3-6)可知, Sigmoid 函数和 Tanh 函数对输出值的值域进行了限定,前者的值域为(0,1),后者的值域为(-1,1),它们均为饱和型非线性函数。当输入数据 x 很大或者很小的时候,上述两种函数的导数将迅速趋近于 0,这就很容易产生所谓的梯度消失的现象,同时函数中存在幂运算,会增加计算压力,当网络层数较深时可能会引起梯度消失。而 ReLU 函数当 $x > 0$ 时,输出一直是 x ,

不会限制输入数据的值域，ReLU 函数是非饱和型线性函数。与 Sigmoid 函数和 Tanh 函数相比，ReLU 函数不需要进行指数计算，计算上更简单，计算成本也更低；它可以输出一个真正的零值，允许神经网络中的隐层激活包含一个或多个真零值，可以加速和简化模型；ReLU 函数使得利用硬件的提升和使用反向传播来训练具有非线性激活函数的深层网络成为可能。

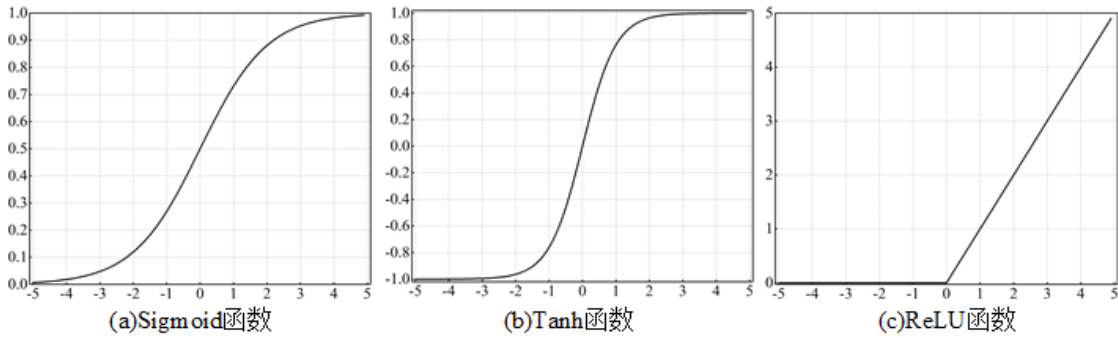


图 3-9 三种激活函数

3.3.3 池化层的工作原理

一般来讲，当卷积层完成对目标特征的提取之后，都会在卷积层之后添加一个池化层。添加池化层的主要作用有是减少下一层待处理的数据量，例如 3.1.1 节提到的 Lenet-5 网络模型中 C_1 层的特征图像尺寸为 28×28 ，经过一个步长为 2 的池化操作之后，特征图的尺寸大小变为 14×14 ，那么此时的数据量一下减少为池化前的 25%。减少了待处理的数据量，也就减少了参数数量，从而避免神经网络陷入过拟合状态。池化操作主要有两种方式：一种是最大池化操作，一种是平均池化操作。图 3-10 为步长为 2 的最大池化和平均池化操作，如果步长为 4，那输出的特征图的尺寸变为 1×1 。

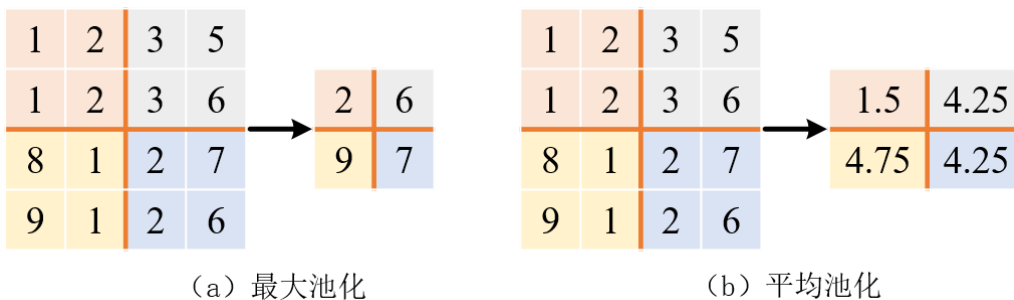


图 3-10 两种池化操作

3.4 本章小结

本章主要完成了以下工作：

(1) 对经典的 LeNet-5、AlexNet、VGG、ResNet 卷积神经网络模型的工作原理做了详细的说明，为设计散斑卷积神经网络模型提供了理论依据。

(2) 通过对卷积神经网络模型中的卷积层、激活层、池化层的详细说明，明确了设计卷积神经网络模型的基本思想。可以通过卷积层和池化层的搭配使用，可以生成需要的特征图，通过选择合适的激活函数，可以使网络模型拟合效果更好。

第4章 网络模型搭建与实验研究

4.1 引言

本课题主要研究利用卷积神经网络模型对散斑图像进行重构，所以卷积神经网络模型就是核心问题，也就是本课题的关键。模型的优劣直接导致重构效果的好坏，以及不同的卷积神经网络模型在各个方面都可能体现不同的优劣。因此本章主要研究基于不同经典卷积神经网络模型搭建的散斑重构网络模型在各个方面的指标，从而确定在不同情况下选择合适的重构网络模型。在众多指标中，最重要的就是评价重构图像的指标，因此本章首先对评价指标做了详细说明。

4.2 散斑重构图像评价指标

评价指标是衡量散斑重构网络模型重构效果的重要指标，也对网络模型进行优化、调参提供重要的依据。本次实验中，主要将重构图像的评价分为两大类：一类是评价重构图像的正确率，也就是评价重构图像与标签图像是否是同一个类别，例如散斑图像对应的图像标签是数字“2”，而重构出的图像是数字“3”，那么就认为是重构错误；另一类是评价重构图像的保真度，也就是通过数值来衡量重构图像和散斑图像对应的标签图像的相似度，相似度越高，说明重构效果越好。正确率应该是保真度的一个基础，如果正确率不高，保真度一定不高，同时也没有了意义，所以设计、优化模型的参考指标就是要高的正确率，在较高的正确率的基础上来提高保真度。除了重构图像的评价指标，对于模型的评价指标还有训练时间和可训练参数的个数。

4.2.1 散斑重构图像正确率

正确率的评价指标最直接，也是最准确的就是用人眼观察。例如散斑图像对应的图像标签是数字“2”，但是重构出来的效果既像数字“2”，也像数字“3”，这时，用人眼观察可以给出一个相对合适的结果。但是重构的图像并不是一张，所以用人眼观察的方法效率低下，无法适用。对于正确率的评价指标借鉴手写数字识别思想，设计卷积神经网络模型，卷积神经网络模型结构如图4-1所示。输入图像为 $28 \times 28 \times 1$ ，第一个卷积层的卷积核数为32，所以经过第一个卷积层变为 $28 \times 28 \times 32$ ，经过最大池化层后变为 $14 \times 14 \times 32$ ，在经过第二个卷积层变为 $14 \times 14 \times 64$ ，池化操作不会改变特征图的数量，只会改变特征图的尺寸大小，卷积操作不会改变特征图的尺寸大小，会改变特征图的数目。同理再经过一个最大池化操作和一个卷积操作，变为

7×7×128。

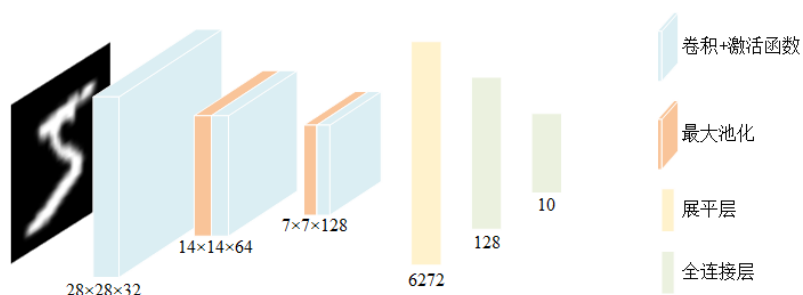


图 4-1 正确率网络模型

卷积神经网络模型前面的层在接入全连接层之前，必须先将高维张量展平成一维向量组，也就是图中的展平层，展平之后便于与后面的全连接层进行适配。展平操作的原理图如图4-2所示，共有128个特征图，按此展开将成为1×6272。网络模型共有896906个训练参数。

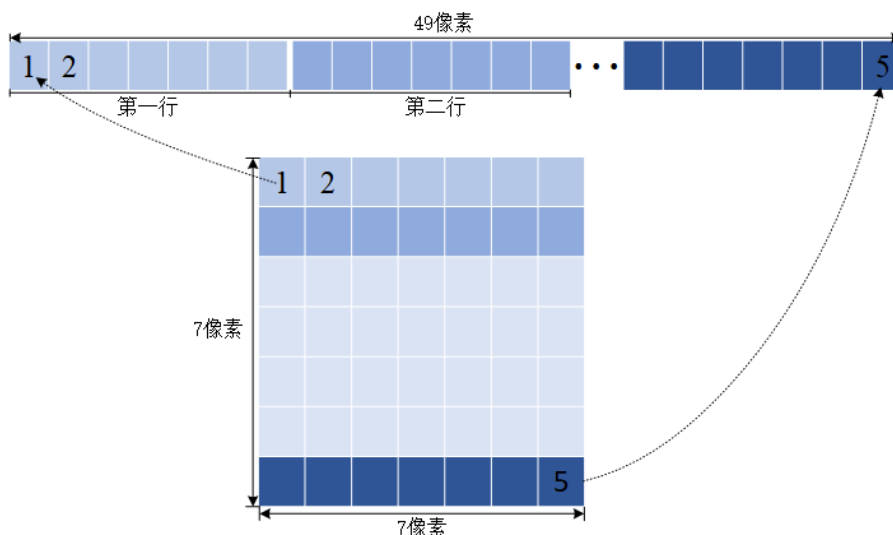


图 4-2 展平操作原理

卷积神经网络模型训练时计算机的环境配置如下：

- (1) 编程语言：python3.7
- (2) 框架：基于 TensorFlow-gpu-2.0.0
- (3) GPU：NVIDIA GTX1070
- (4) CPU：i7-7700k
- (5) 环境管理软件：Anoconda3
- (6) 电脑系统：Windows10

设计好卷积神经网络模型后，用MNIST数据集来对模型进行训练与测试，MNIST数据集包含60000张训练集图像和10000张测试集图像，共有10个类别。训练

50轮的测试集与训练集的损失值和准确率变化如图4-3所示。

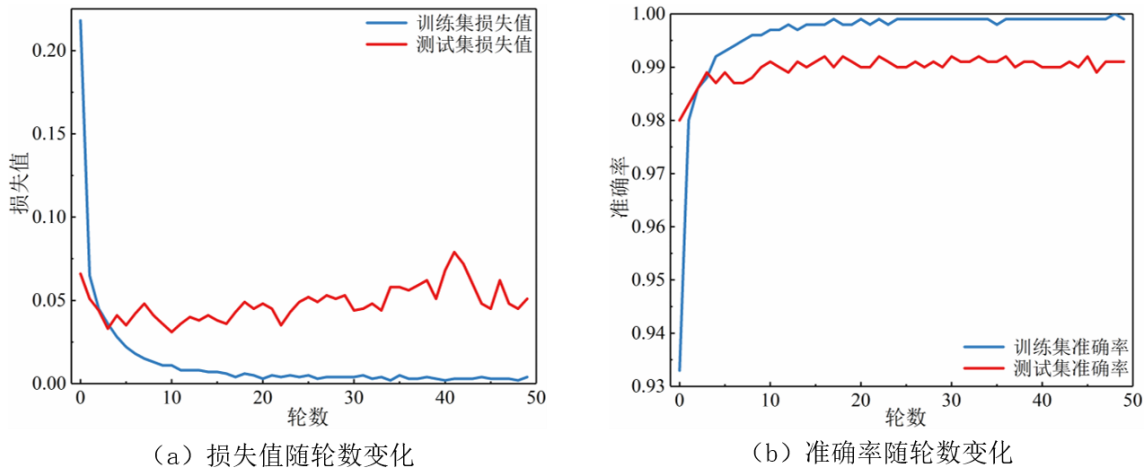


图 4-3 准确率模型的损失值和准确率变化

从图4-3 (a) 中可以看出训练集损失值随着训练轮数不断下降, 测试集损失值在0.05附近振荡, 说明模型对于手写数字识别可用。由图4-3 (b) 可以看出, 训练集与测试集的准确率都随着训练轮数都呈现一个先上升的趋势, 并分别收敛域99.99%和99.01%。通过训练结果表明, 该模型可以用于对散斑图像正确率的识别, 相当于训练了一双“人眼”, 在误差允许的范围内, 使用“人眼”模型可以快速得到重构图像的准确率, 提高效率。

4.2.2 散斑重构图像相似性

保证正确率的前提下, 还希望由散斑图像重构出的图像在亮度、对比度等方面不断的逼近其对应的图像标签, 若能达到人眼无法分辨是由模型重构出来的图像还是对应的图像标签, 那将是最好的效果。需要将这一客观感受的效果变为一种能够用数值量化出来的结果, 结构相似性 (Structural Similarity, SSIM) 指标很好的解决了这一问题。SSIM从亮度、对比度、结构三个方面综合对两张图像进行考量, SSIM的取值在0~1之间, 其越接近于1, 两张图像越相似, 当 $SSIM = 1$ 时, 表明两张图像完全相同。具体计算公式为

$$L(X, Y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4-1)$$

$$C(X, Y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (4-2)$$

$$S(X, Y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (4-3)$$

式中 $\sigma_{XY} = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X)(Y_i - \mu_Y)$;

μ_X 、 μ_Y ——图像 X ， Y 的均值;

σ_X ， σ_Y ——图像 X ， Y 的标准差;

σ_{XY} ——图像 X ， Y 的协方差;

C_1 、 C_2 、 C_3 ——常数。常数的作用是防止分母等于零。

由式 (4-1)、(4-2)、(4-3) 可得:

$$SSIM(X, Y) = L(X, Y)^\alpha \times C(X, Y)^\beta \times S(X, Y)^\chi \tag{4-4}$$

式中 $L(X, Y)$ 、 $C(X, Y)$ 、 $S(X, Y)$ ——图像 X ， Y 在亮度、对比度、结构上的差异;

当 $\alpha = \beta = \chi = 1$ ， $C_3 = \frac{C_2}{2}$ 时，得 $SSIM(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_X\sigma_Y + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}$ 。

在散斑重构网络模型中，均采用上述两种评价指标。在保证正确率的前提下，尽量提高 SSIM 值，充分发挥卷积神经网络在散斑图像重构方面的作用。

4.3 基于 VGG-16 散斑重构网络模型及实验结果

4.3.1 基于 VGG-16 散斑重构网络模型

在 3.1 节中详细介绍了经典的卷积神经网络模型和卷积神经网络模型的基本概念知识。VGG 网络模型是经典卷积神经网络模型在深度上能达到的极致，所以本小节主要介绍基于 VGG-16 网络模型进行搭建散斑重构网络模型，网络结构图如图 4-4 所示。

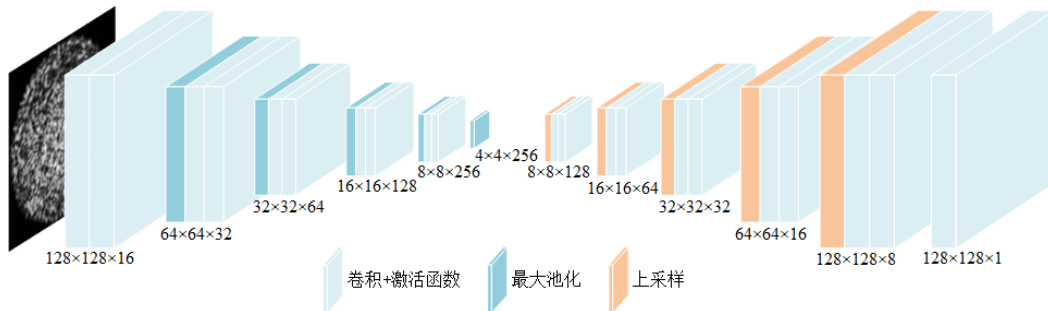


图 4-4 VGG 散斑重构网络模型

VGG-16 网络模型最后有三层全连接层，最后输出 1000 个类别。而对于散斑重构网络模型，输出的应该是与标签尺寸大小一致的图像。所以在设计基于 VGG-16 散斑重构网络模型时，先保留 VGG-16 前 5 组卷积层的设计，并且将后三组中的卷积个数由 3 个改为两个，同时将网络模型的输入由 $224 \times 224 \times 3$ 改为 $128 \times 128 \times 1$ ，

后期根据实验结果及需求在更改输入图像的尺寸大小。同时修改了每组卷积层中卷积核的数量，第一组卷积核数目由 64 修改为 16，第二组卷积核数目由 128 修改为 32，第三组卷积核数目由 256 修改为 64，第四组卷积核数目由 512 修改为 128，第五组卷积核数目由 512 修改为 256。最大池化层未做改变。经过五组卷积层和最后一个池化层之后，输出的特征图为 $4 \times 4 \times 256$ ，而图像标签为 $128 \times 128 \times 1$ 。此时需要增加五次上采样层和五组卷积层，将输出图像的尺寸改为 $128 \times 128 \times 1$ ，利用损失函数对输出图像与图像标签求损失值，通过增加训练轮数，不断对网络模型的参数进行更新，从而实现当输入一张散斑图像，利用网络模型可以对散斑图像进行重构，最大程度还原散斑图像所对应的图像标签。在图中，所标示的尺寸大小的卷积核数量为卷积层的，经过池化层和上采样层之后的特征图的数量是前一层卷积核的数量，例如在第二组卷积层中，经过池化层后特征图的尺寸大小与数量为 $64 \times 64 \times 16$ 。

这里采用的上采样为 UpSampling2D，上采样从本质上讲就是池化操作的逆过程，也正是因为有上采样和池化操作才使得我们在设计网络模型时，可以按照我们的需求随意更改输出特征图的尺寸大小，使网络模型更灵活。例如输入图像为 2×2 的矩阵，通过上采样操作将其宽和高变为原来的 2 倍，将得到一个 4×4 的矩阵；若将其宽扩大为原来 2 倍和高扩大为原来的 3 倍，将得到一个 4×6 的矩阵，操作过程如图 4-5 所示。

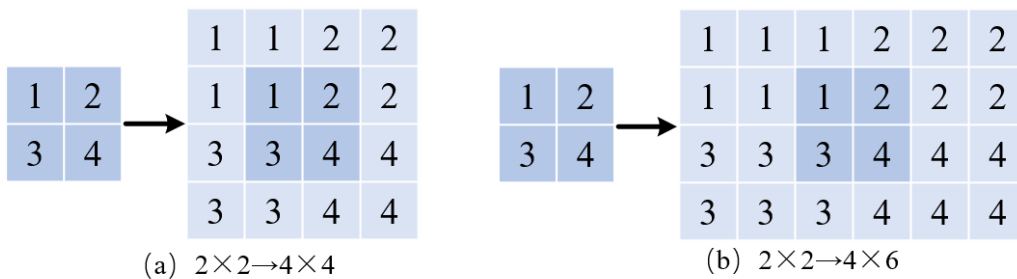


图 4-5 UpSampling2D 操作原理图

卷积神经网络不需要对图像进行复杂的特征提取，它可以直接从原始图像中找出视觉规律，所以人们也将卷积神经网络模型称为“黑箱”模型。为了更充分的了解“黑箱”模型，也为了更好地调试和优化程序，将基于 VGG-16 散斑重构网络模型卷积核和特征图进行可视化操作，模型的输入图像为 2.4.3 小结中的散斑图像。图 4-6 (a) 是对散斑重构网络模型中的第一组卷积层中的第一个卷积操作使用的卷积核进行可视化，卷积核大小为 3×3 ，共有 16 个卷积核；卷积核中第一个图对应的

矩阵为 $\begin{bmatrix} 0.163 & -0.184 & 0.023 \\ -0.098 & 0.004 & -0.129 \\ -0.012 & 0.150 & -0.105 \end{bmatrix}$ ，图中的亮暗对应矩阵中元素值的大小，元素值

较大，对应较亮的方块，元素值较小，对应较暗的方块。输入图像经过 16 个卷积核操作生成 16 个特征图，特征图尺寸大小为 128×128 ，如图 4-6 (b) 所示。

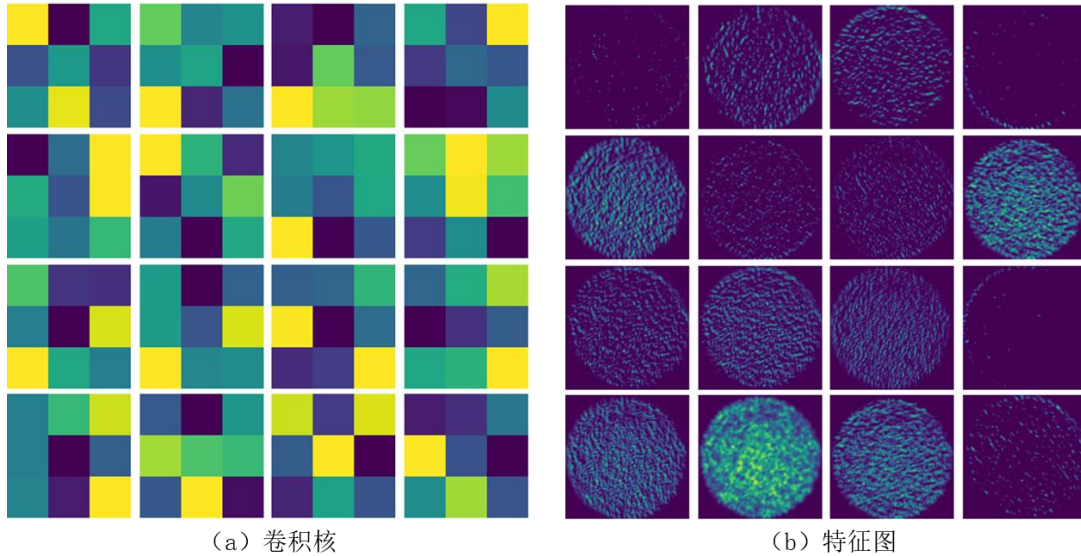


图 4-6 卷积核和特征图可视化

4.3.2 VGG 散斑重构实验结果

本次共采集 MNIST 数据集对应散斑图案 5000 张，其中 4500 张用作训练集，500 张作为测试集，500 张测试集中包含的图像为数字 0~9 是随机的。训练集与测试集比为 9:1。计算机的环境配置如下：

损失函数采用均方误差 (MSE)，均方误差在检测模型的预测值和真实值之间的偏差非常重用。训练集可表示为 $Train = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_N, y_N)\}$ ，其中 x_i 为训练集中散斑图像， y_i 为训练集中的散斑图案对应的图像标签。 N 为训练集的样本数，在这里为 4000。假设训练模型为 $f(x)$ ，模型的输出为 $f(x_i)$ ，则 MSE 可表示为

$$MSE = \frac{1}{M} \sum_{i=1+nM}^{(n+1)M} (y_i - f(x_i)) \quad (4-5)$$

式中 $n=0,1,2,\dots$;

M ——训练过程中的批量大小。

在基于 VGG-16 的散斑重构的网络模型中，批量大小为 64。优化器采用 Adam 优化器，Adam 优化器结合了 AdaGrad 和 RMSProp 两种优化算法的优点，其优点主要有 1) 实现简单，计算高效，对内存需求少；2) 参数的更新不受梯度的伸缩变换影响；3) 超参数具有很好的可解释性，且通常无需调整或仅需很少微调；4) 很适合应用于大规模的数据和参数的场景；5) 适用于不稳定的目标函数。

图 4-7 分别是训练 10 轮、30 轮、50 轮、100 轮损失值随训练轮数的变化趋势，训练所用时间分别是 1.416 分钟、3.893 分钟、6.307 分钟、12.733 分钟。从 10 轮和 30 轮图像可以看出，测试集和训练集损失值还有下降的趋势。50 轮训练集损失值收敛于 0.006，测试集损失值收敛于 0.016；100 轮训练集损失值收敛于 0.003，测试集损失值收敛于 0.014，从损失值角度看 100 轮效果稍微好于 50 轮，但是 100 轮训练时长约是 50 轮训练时长的 2 倍。

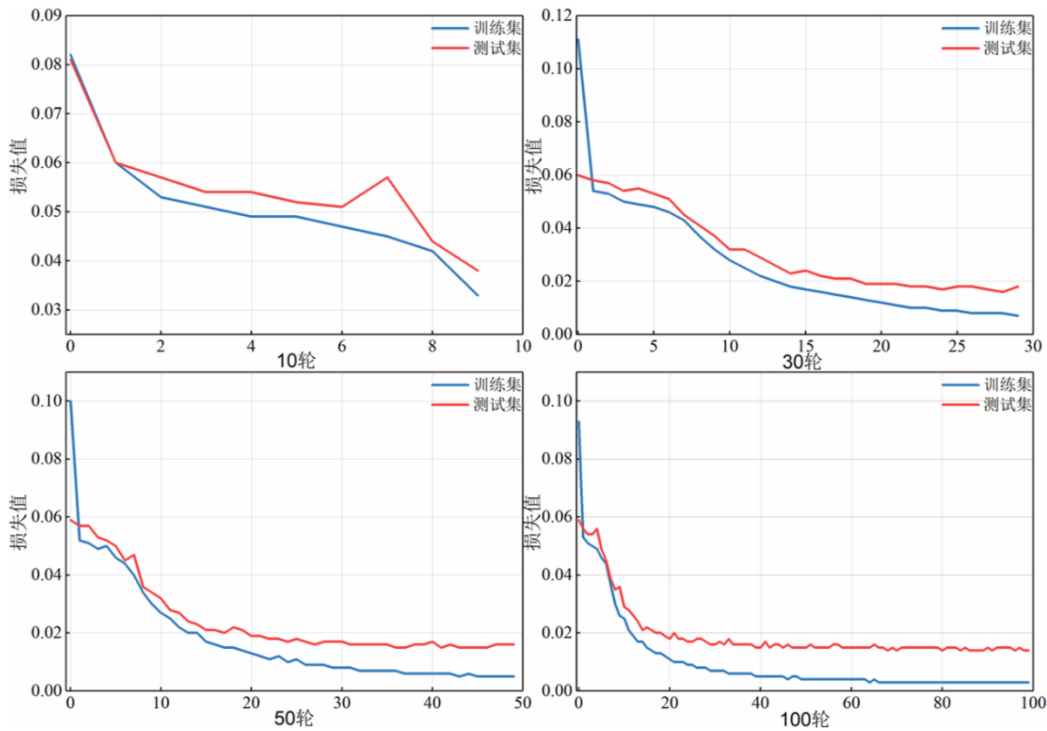


图 4-7 不同训练轮数的损失函数

用训练好的模型对测试集中的图像进行重构，用 3.2 节提到的两种评价指标来对重构图像进行评价。首先是利用训练好准确率模型对其进行评价，训练 10 轮、30 轮、50 轮、100 轮的准确率如图所示，从图 4-8 (a) 中可以发现准确率随着训练轮数逐渐上升并收敛于某一数值，当训练轮数为 100 轮的时候，准确率可达 93.4%；图 4-8 (b) 为训练 100 轮的混淆矩，混淆矩阵的纵轴为图像标签对应的数字真实值，横轴为“人眼”模型对由标签图像对应的散斑图像重构出的图像的预测值。当标签图像对应的标签是数字“0”，“人眼”模型对由标签图像对应的散斑图像重构出的图像的预测值为数字“0”个数为 50 个，预测为数字“1”的个数为 1 个。其中，标签图像对应的标签数字是“8”的时候，“人眼”模型对由标签图像对应的散斑图像重构出的图像的预测值错误个数最多，其中有 4 个预测为数字“3”。由于手写数字“8”和数字“3”书写非常相像，所以有散斑图像重构出的图像较难逼近与其对应的图像标签，图 4-9 (a) 为图像标签为“8”的散斑图像重构结果，使用“人眼”模型预测

结果为数字“3”，人眼可以发现，两张重构图像与标签图像还是非常接近的，这时就需要采用 SSIM 指标来对其进行评价，从多维度给出比较客观的结果。

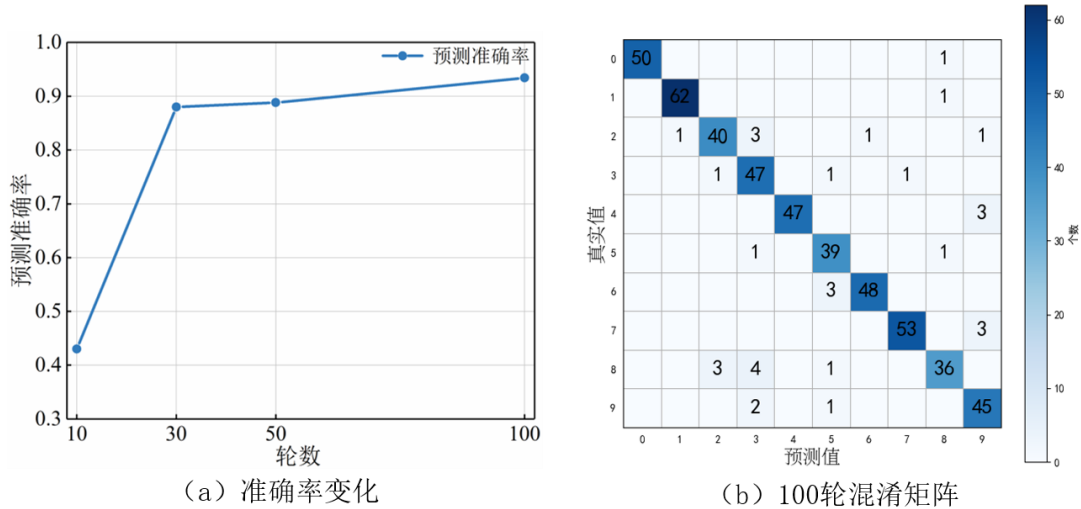


图 4-8 不同轮数准确率与 100 轮混淆矩阵

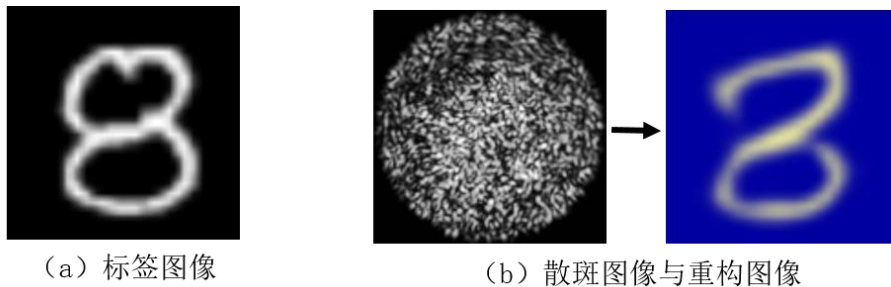


图 4-9 标签图像与重构结果

采用基于 VGG-16 散斑重构网络模型对数据集进行训练，训练不同轮数所需时间及 SSIM 均值如表 4-1 所示。

表 4-1 不同训练轮数所用时间与 SSIM 均值

轮数	10	30	50	100	150
时间 (分钟)	1.416	3.893	6.307	12.733	19.609
SSIM 均值	0.531	0.734	0.797	0.831	0.823

训练 10 轮、30 轮、50 轮、100 轮的测试集部分重构图像如图 4-10 所示。当 SSIM 均值为 0.531 时，重构图像的效果失真非常严重，只有数字“0”、“1”、“3”、“9”能够识别，其余数字均无法辨认。当 SSIM 均值可达 0.734 时，重构图像用肉眼基本可以分辨，并在数字“0”、“1”、“3”、“4”、“5”、“7”、“8”、“9”上有较好的重构效果。从 SSIM 均值这个维度上，训练 100 轮的效果要明显好于训练 50 轮，但是与此同时，也付出了更长训练时间的代价，因此根据需求选择合适的训练轮数。

图像标签	0	1	2	3	4	5	6	7	8	9
重构图像 10轮										
SSIM值	0.470	0.575	0.487	0.518	0.584	0.476	0.546	0.608	0.601	0.631
重构图像 30轮										
SSIM值	0.819	0.874	0.713	0.799	0.759	0.747	0.721	0.829	0.819	0.813
重构图像 50轮										
SSIM值	0.815	0.912	0.827	0.815	0.832	0.814	0.818	0.814	0.860	0.888
重构图像 100轮										
SSIM值	0.852	0.915	0.826	0.816	0.846	0.778	0.886	0.943	0.908	0.886

图 4-10 不同训练轮数重构结果

从上述结果看，基于 VGG 散斑重构网络模型在图像重构上的表现还不错，但是该网络模型中可训练参数有 1768297 个，略显臃肿。能否设计一个轻量化网络模型，在降低网络模型的可训练参数的前提下，其他指标争取达到与 VGG 散斑重构网络模型相当的水平成为下一小节主要研究内容。

4.4 基于 ResNet 散斑重构网络模型及实验结果

4.4.1 基于 ResNet 散斑重构网络模型

ResNet 网络模型可以使卷积神经网络模型网络层数加深的同时，并且降低可训练参数个数。本小节基于 ResNet 网络设计思想，首先设计网络残差块，并根据残差块设计散斑重构网络模型。残差块结构如图 4-11 所示。

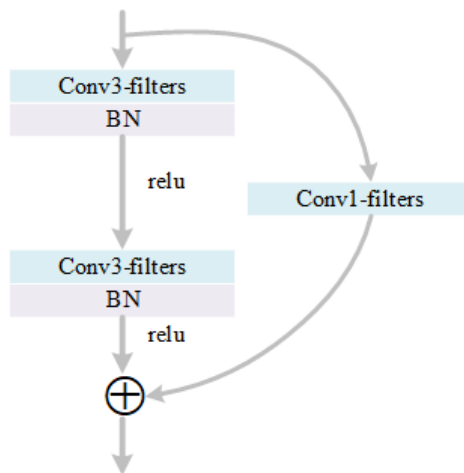


图 4-11 残差块

残差块主要包括两条路径：一条是学习路，学习路由两组卷积核大小为 3×3 的卷积层和批规范层组成；另一条是短接路，短接路由卷积核大小为 1×1 的卷积层，主要目的是让输入的 x 能够与经过学习路之后的输出进行相加。在学习路添加了批规范层，主要的目的是通过减少内部协变量偏移来加速深度网络训练速度。

利用残差块结合基于 VGG-16 搭建的散斑重构网络模型搭建基于 ResNet 的散斑重构网络模型，结构图如图 4-12 所示，其中损失函数和优化器与基于 VGG-16 搭建的散斑重构网络模型一致。

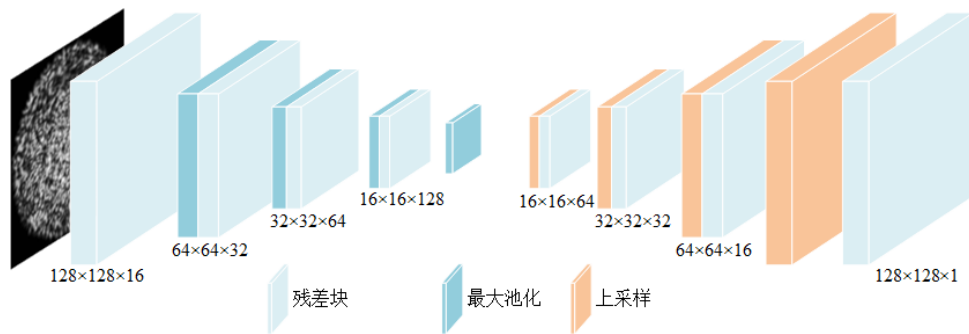


图 4-12 ResNet 散斑重构网络模型

4.4.2 ResNet 散斑重构实验结果

该散斑重构网络模型训练参数有 465299 个，与基于 VGG-16 的散斑重构网络模型的训练参数 1768297 个相比，训练参数减少了 1302998 个，充分体现了基于 ResNet 网络在参数减少上的优势，这对于加深网络模型非常有意义。采用该散斑重构网络模型对上述数据集进行训练与测试，训练轮数为 100 轮的训练集损失值和测试集损失值变化如图 4-13(a)所示，实验结果准确率的混淆矩阵如图 4-13(b)所示。

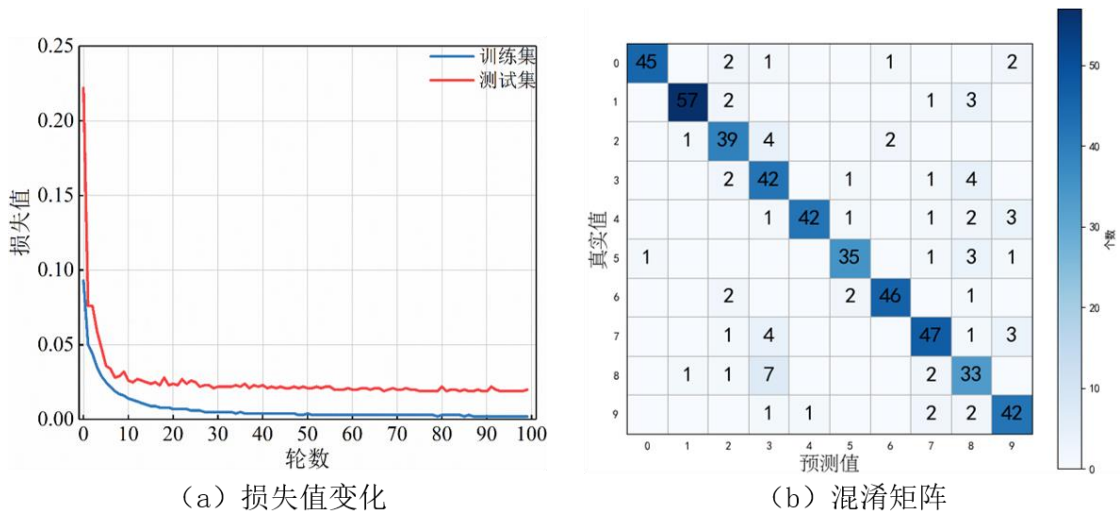


图 4-13 100 轮损失函数与混淆矩阵

训练集的损失值收敛于 0.002，测试集的损失值收敛于 0.020，利用“人眼”模型对由标签图像对应的散斑图像重构出的图像的预测值的准确率为 85.6%，从混淆矩阵可以看出，除了对角线以外，其他网格分布的数量增加了，相较于基于 VGG-16 的散斑图像重构模型来讲，准确率下降了 7.8%，所以只用损失函数不足以来判断模型的优劣，损失函数不断下降并收敛于某一数值只能证明网络模型对此问题能够进行拟合。

训练 100 轮所需时间为 15.567 分钟，测试集中部分图像重构结果如图 4-14 所示，500 张测试集重构图像的 SSIM 均值为 0.757，与基于 VGG-16 的散斑图像重构模型相比，SSIM 均值下降了 0.074。该 SSIM 均值比 VGG 散斑重构网络模型训练 30 轮的结果多 0.023，也就是相当的水平，但是从视觉效果上，该重构图像中出现的“白斑”较多，重构效果不够理想。

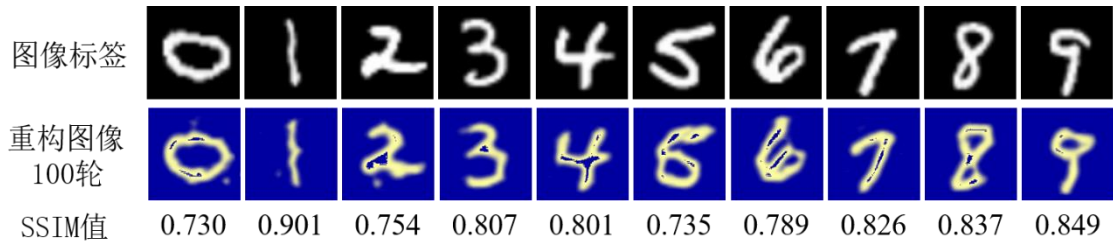


图 4-14 ResNet 散斑重构网络 100 轮重构结果

4.4.3 改进散斑重构网络模型及实验结果

基于 ResNet 散斑重构网络模型，只有在模型可训练参数上有明显的优势，但是在重构图像的准确率、SSIM 均值上都没有达到与 VGG 散斑重构网络模型相当的水平。因此对该模型进行相应的改进，来提高该模型对测试集中重构图像的 SSIM 均值和准确率。对上述模型的改进主要是又增加了一个残差快，其结构如图 4-15 所示。该残差块的短接路不包含任何卷积层，因此该残差快的输入和输出必须具有相同的特征图个数，这样才可以在输出出进行相加。结合两个残差块就需要对整体网络模型进行相应的调整。

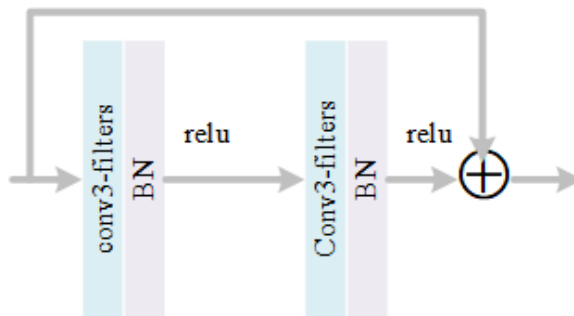


图 4-15 残差块 1

结合上述两个残差块的输出特点，将散斑重构网络模型修改为图 4-16，损失函数与优化器与未修改前一致。

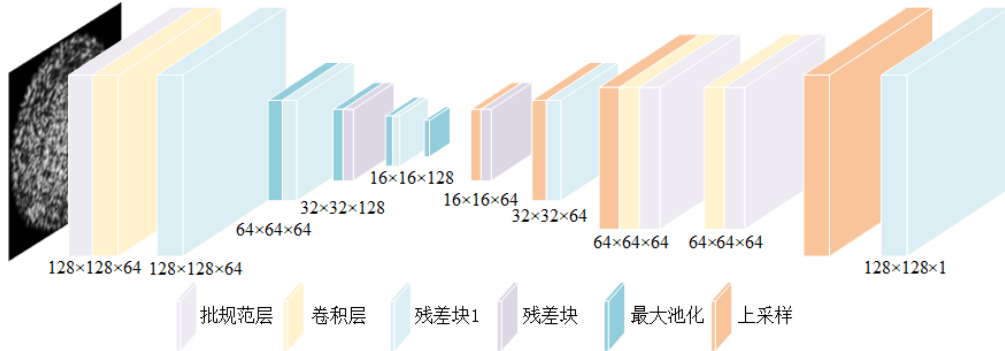


图 4-16 改进 ResNet 散斑重构网络

修改后的网络模型训练 100 轮的损失函数图像如图 4-17 (a) 所示,混淆矩阵如图 4-17 (b) 所示,准确率为 95%。

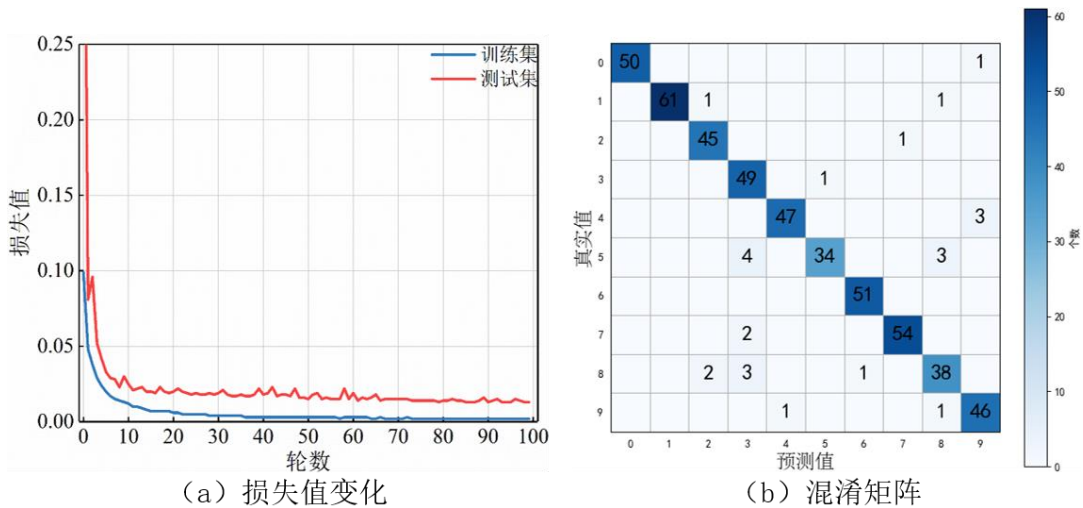


图 4-17 改进 ResNet100 轮损失函数与混淆矩阵

修改后的散斑重构网络模型可训练参数有 943235 个，比未修改的散斑重构网络模型的可训练参数增加了 477936 个。在可训练参数方面与基于 VGG-16 散斑重构网络模型相比仍具有较大优势。部分重构图像如图 4-18 所示，

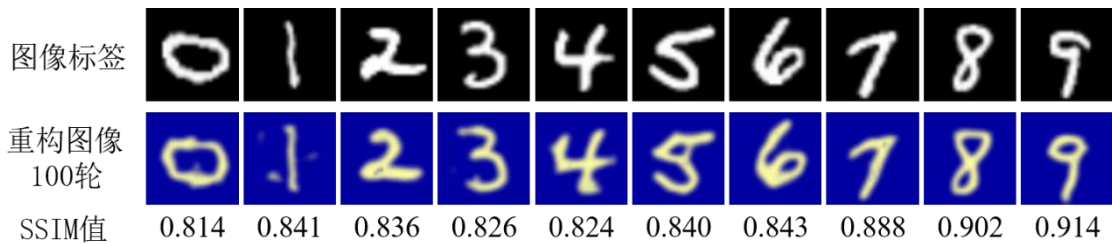


图 4-18 改进 ResNet 散斑重构网络 100 轮重构结果

修改后的散斑重构网络模型的训练集损失值收敛于 0.002，测试集的损失值收敛于 0.013，在测试集上的损失值低于未修改的散斑重构网络模型，从混淆矩阵和损失函数两个角度考虑，达到了修改网络模型的目的，提高了准确率，但是训练时间为 36.018 分钟，比未修改增加了 20.451 分钟。500 张测试集的 SSIM 均值为 0.815，从重构效果来看，不仅提高了 SSIM 均值，同时也修正了模型未修改前出现较多“白斑”现象。

4.5 基于 U-net 散斑重构网络模型及实验结果

4.5.1 基于 U-net 散斑重构网络模型

VGG 散斑重构网络模型可训练参数多，但是训练时间短，ResNet 与其正好相反。能否设计一个降低可训练参数或者降低训练时间，但是准确率和 SSIM 均值与 VGG 和改进 ResNet 散斑重构网络模型水平相当的散斑重构网络模型呢？本小节主要基于该问题进行散斑重构网络模型的设计。德国弗莱堡的 Olaf Ronneberger 等人提出的 U-net 网络模型^[47]在图像语义分割任务中应用非常成功，主要思想是在卷积神经网络的连续收缩层之后再添加连续扩张层，最终输出指定大小的特征图，反应了图像的像素级定位，这与将散斑图像重构为对应的原始图像的思想不谋而合，而 VGG 与 ResNet 网络模型都是图像分类上的经典网络模型。所以此小节介绍基于 U-net 网络模型的散斑重构网络模型，网络模型的整体结构图如图 4-19 所示。

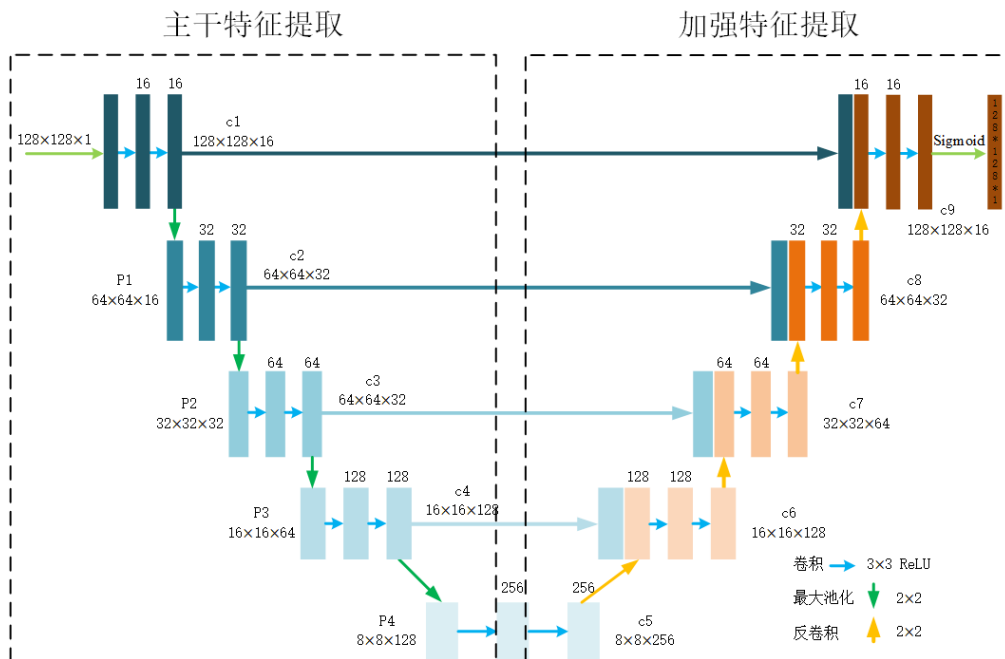


图 4-19 U-net 散斑重构网络

利用主干特征提取部分，可以获得 5 个初步有效的特征层，主干提取部分与 VGG 网络模型相似，主要包括卷积层和最大池化层，不同的是，在每一组的一个卷积层后面都增加了防止过拟合的 Dropout 机制^[48]，工作原理如图 4-20 所示。

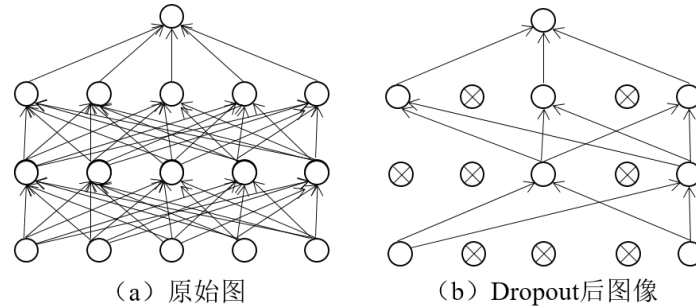


图 4-20 Dropout 工作原理

Dropout 机制就是以某种概率暂时丢弃一些单元，并丢弃与它相连的所有节点权值，若某节点被丢弃，那么输出为 0，从图中可以发现网络结构变简单了。卷积层的激活函数均采用 ReLU 激活函数，卷积层均采用等大填充，卷积操作结合池化操作可以得到需要的特征图的大小，并不一定是结构图中 2 的倍数，这里采用上述数值，一方面便于与输入输出图像匹配，另一方面便于设计网络结构。

利用加强特征提取部分对主干特征提取部分得到的特征层进行上采样，同时在相应的阶段进行特征融合，得到一个融合了所有特征的有效特征层。这里实现的上采样与上两节散斑重构网络中的上采样操作不同，这里上采样实现采用反卷积操作 (Conv2DTranspose)，反卷积是一种特殊的正向卷积操作，主要目的是实现特征图分辨率放大操作。采用等大填充的反卷积工作原理如图 4-21 所示，输入图像中尺寸为 2×2 的黄色矩阵为原始输入图像，周围的 0 为等大填充，经过反卷积之后特征图的大小变为 4×4，满足要求。最后一层卷积操作卷积核为 1×1，采用的激活函数为 Sigmoid 函数。

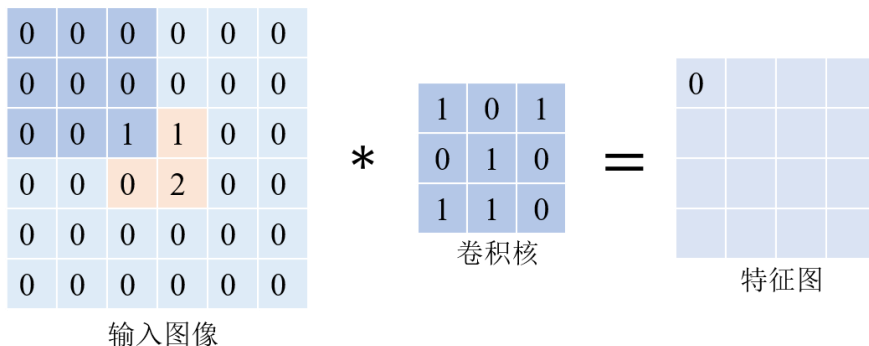


图 4-21 反卷积操作原理图

网络模型的损失函数采用交叉熵损失，其表达式为

$$L = \frac{1}{M} \sum_{i=1+nM}^{(n+1)M} -[y_i \log(f(x_i)) + (1 - y_i) \log(1 - f(x_i))] \quad (4-6)$$

式中 $n=0,1,2,\dots$;

M ——训练过程中的批量大小。

在基于 U-net 的散斑重构的网络模型中，批量大小为 64。 y_i 为散斑图像对应的标签， $f(x_i)$ 为模型的输出，当 y_i 与 $f(x_i)$ 值越接近时， L 值越小。

4.5.2 U-net 散斑重构实验结果

实验首先对 MNIST 数据集进行了测试，对散斑重构网络模型训练 50 轮的损失函数图像如图 4-22 (a) 所示，利用“人眼”模型对由标签图像对应的散斑图像重构出的图像的预测值的准确率为 92.2%，对应的混淆矩阵如图 4-22 (b) 所示。散斑重构刚落模型训练 50 轮时，训练集损失值收敛于 0.131，测试集损失值收敛于 0.149，采用不同的损失函数，损失值收敛于不同的数值，但是损失值变化都呈下降趋势，只有损失值有下降的趋势，才能说明该模型可以对此问题进行拟合。

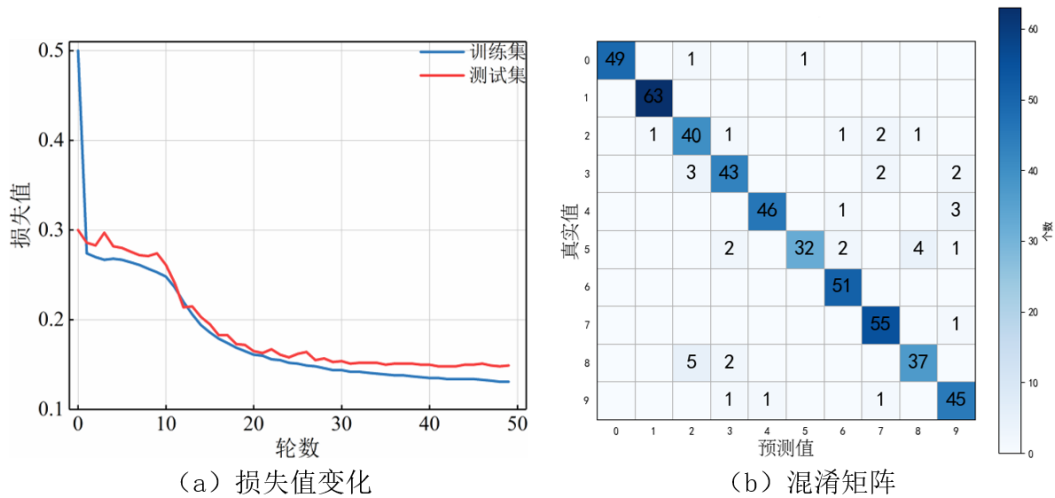


图 4-22 U-net50 轮损失函数与混淆矩阵

网络模型训练 50 轮时，测试集中 500 张测试集的 SSIM 均值为 0.813，训练所用时间为 8.743 分钟，部分重构图像如图 4-23 所示。

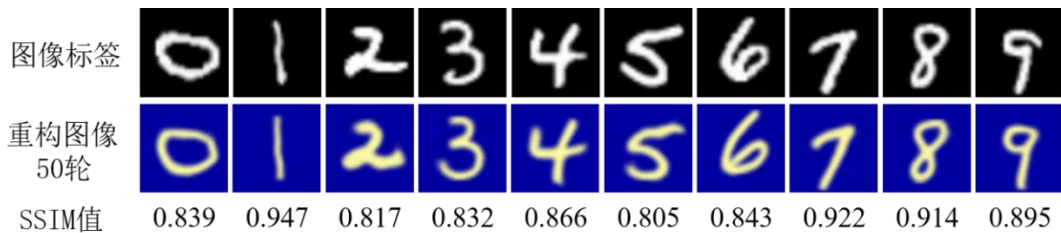


图 4-23 U-net 散斑重构网络 50 轮重构结果

将 VGG 散斑重构网络模型、改进 ResNet 散斑重构网络模型和 U-net 散斑重构网络模型的各项指标总结如下表 4-3。

表 4-3 三种网络模型指标对比

类别	可训练参数	训练时间（分钟）	SSIM 均值	准确率
VGG	1768297	12.733	0.831	93.4%
ResNet	943235	36.018	0.815	95.0%
U-net	1940817	8.743	0.813	92.0%

在准确率与 SSIM 均值达到与 VGG 散斑重构网络模型和 ResNet 散斑重构网络模型相当的情况下，该网络模型可训练参数有 1940817 个，比 VGG 散斑重构网络模型的可训练参数还多 172520 个，可见该网络模型在可训练参数上不占优势，但是该模型训练 50 轮所用的时间是 8.743 分钟，比 VGG 散斑重构网络模型训练 100 轮用时减少了 3.99 分钟。与 ResNet 网络模型训练 100 轮相比训练时间减少 27.257 分钟，但是网络模型可训练参数增加了 997582 个。训练轮数不重要，关注的是在测试集上取得相当的准确率和 SSIM 均值所用的时间。可见可训练参数的数量和训练的时间在该问题中是“鱼”和“熊掌”的关系。当数据集的数据比较复杂，类别较多时，采集到的数据必然会增加，数据集增加，训练时间定会增加，因此，从训练时间角度，对复杂数据集采用 U-net 散斑重构网络模型进行重构，验证在复杂数据集上的可行性。复杂数据集采用手写字母数据集^[49]，手写字母数据集包含大写字母和小写字母，难以区分大写和小写的字母归为一个类别，所以字母数据集共包含 37 个类别。因此对于手写字母数据集，共采集 55000 张散斑图像，其中训练集包含 50000 张散斑图像，测试集包含 5000 张测试图像，训练 100 轮的部分重构结果如图 4-24 所示，测试集 SSIM 均值为 0.815。从训练结果来看，模型对复杂数据集仍有较好的表现能力，这也为下一章节的迁移学习奠定基础。

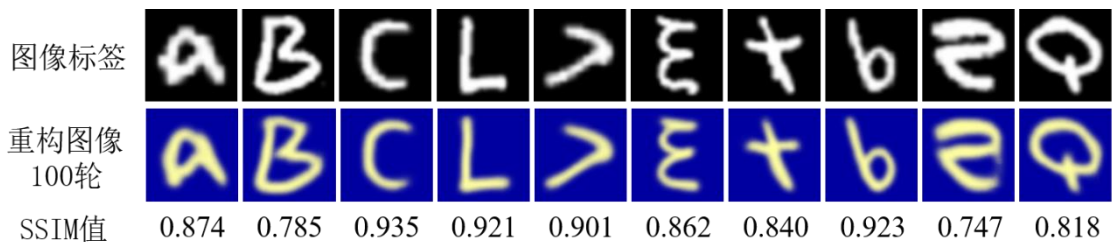


图 4-24 U-net 散斑重构网络字母重构结果

4.6 图像噪声

4.6.1 噪声类别

如果将散斑图像看作是原始信号，那么噪声就是干扰信号，那么增加噪声之

后的图像带有的原始信息就会被干扰，造成图像退化，这将会对散斑重构网络模型提出挑战。为验证散斑重构网络模型能否对加有噪声的散斑图像进行重构，选用两种典型的噪声类型，一种是高斯噪声，另一种是椒盐噪声^[50]。

(1) 高斯噪声

高斯噪声的概率密度函数是高斯分布，也就是正态分布，生活中极为常见，高斯随机变量的概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4-7)$$

式中 x —— 灰度值；
 μ —— x 的数学期望；
 σ —— x 的标准差；
 σ^2 —— x 的方差。

标准差和方差反应了灰度值的离散程度。取 $\mu=10$ ， σ 分别取值 50、100、150、200、250 所产生的噪声图像如图 4-25 所示， σ 值越大，产生的噪声就越大。

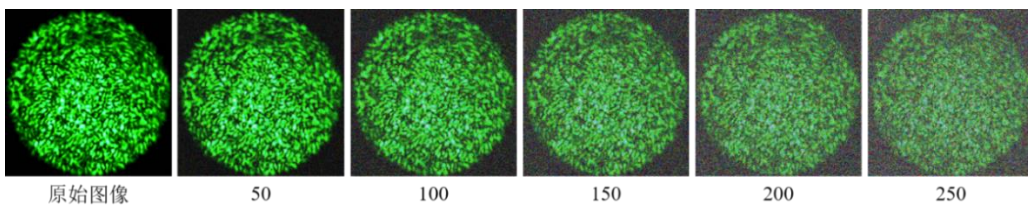


图 4-25 不同 σ 值污染后的散斑图像

(2) 椒盐噪声

经过椒盐噪声污染的图像，通常表现为图像上出现比较亮或者比较暗的点，在程序上椒盐噪声对应的像素值为 0，盐噪声对应的像素值为 255，黑白像素值的点在图像上随机生成，就会对图像产生椒盐噪声。椒盐噪声是视觉上感知最明显的一种噪声，噪声脉冲可正可负，每个像素点上的脉冲噪声通常在空间上是不相关的，并且与原图像信息无关。椒盐噪声的概率密度为

$$f(x) = \begin{cases} P_a & x = a \\ P_b & x = b \\ 0 & \text{其他} \end{cases} \quad (4-8)$$

图像被椒盐噪声污染时，图像中噪声的灰度值只会出现 a 和 b 两种类型，而这两种噪声灰度值出现的概率分别是 P_a 和 P_b 。如果将污染后的图像记为 $[f_{i,j}]_{m \times n}$ ，那么此时像素点 (i, j) 的灰度值可以表示为

$$f_{i,j} = \begin{cases} n_{i,j} & 1-p \\ 0_{i,j} & p \end{cases} \quad (4-9)$$

式中 $n_{i,j}$ ——被噪声污染后的灰度值；

$0_{i,j}$ ——原图像对应像素点的灰度值。

$0 \leq p \leq 1$ ，大小表示了图像被污染的程度。 p 分别取 0.1、0.3、0.5、0.7、0.9，污染后的图像如图 4-27 所示， p 值越小，污染程度越严重。

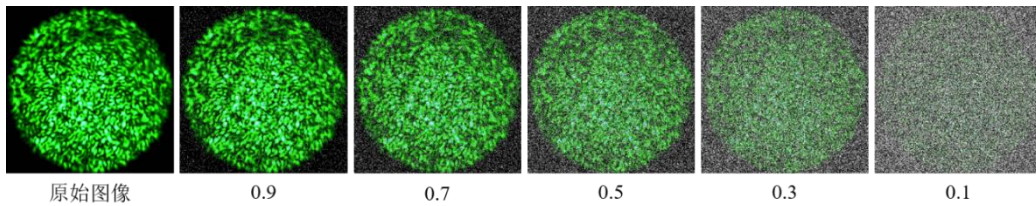


图 4-27 不同 p 值污染后散斑图像

4.6.2 噪声实验结果

(1) 高斯噪声

利用 VGG 散斑重构网络模型、改进 ResNet 散斑重构网络模型、U-net 散斑重构网络模型对 σ 值分别为 50、100、150、200、250 污染后的图像进行训练。以图像标签为数字“2”为例，不同网络模型在不同污染值下的重构图像如图 4-28 所示。

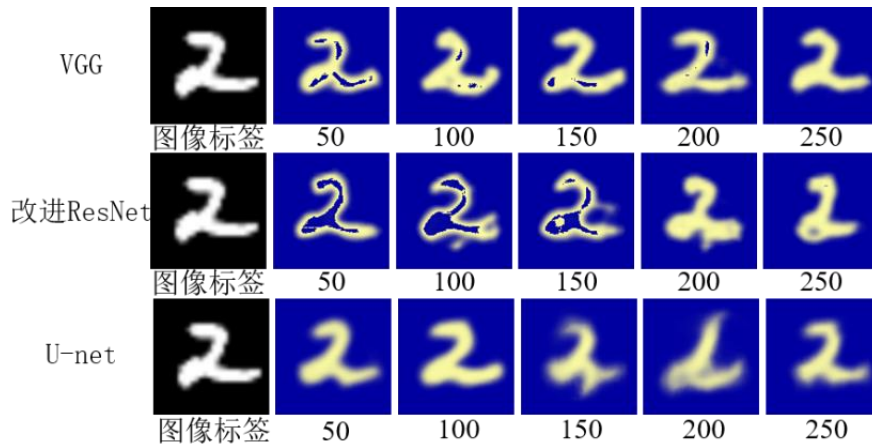


图 4-28 不同网络模型在不同高斯污染值下重构结果

利用准确率模型对测试集图像进行测试，得到其在不同污染值情况下的准确率，并求得测试集重构图像的 SSIM 均值，将两项评价图像重构质量的指标绘制如图 4-29 所示，主要以这两项指标来衡量模型的抗干扰能力大小，暂不讨论其可训练参数与训练时间。

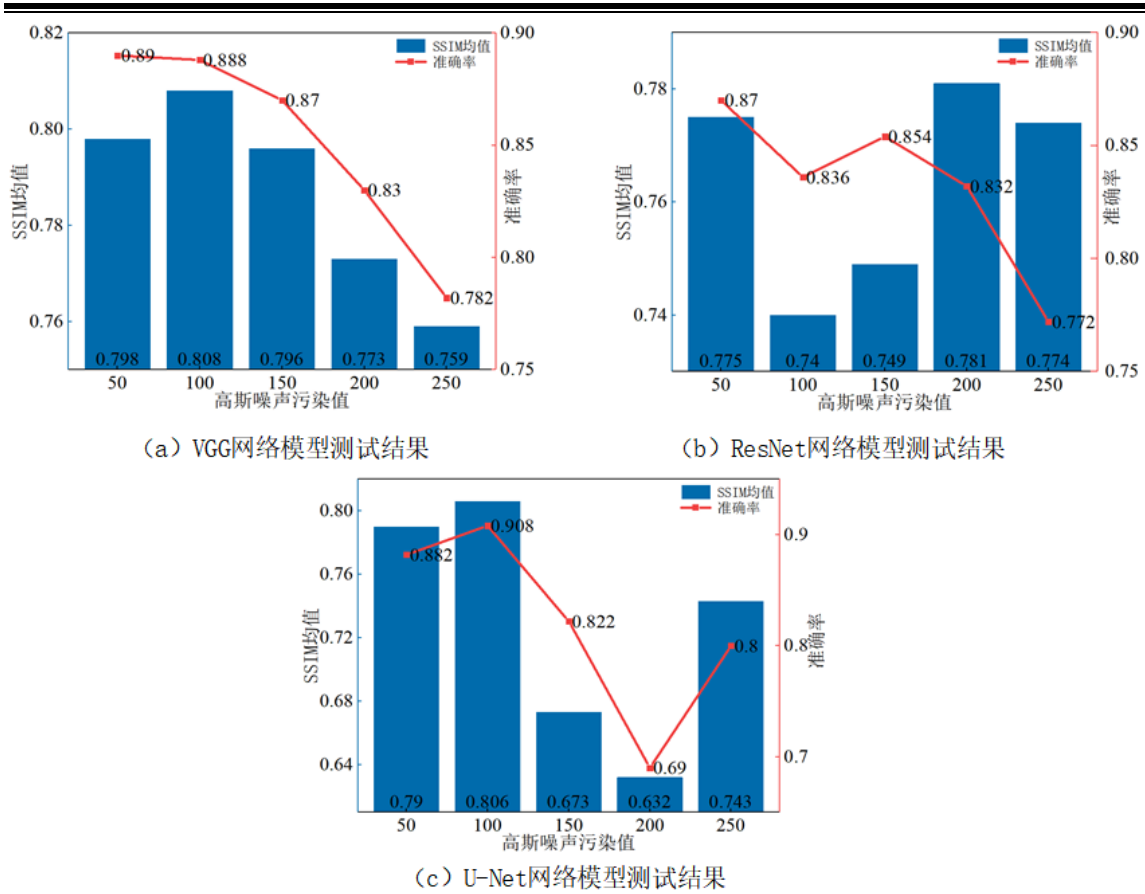


图 4-29 不同网络模型关于高斯噪声的准确率和 SSIM 均值变化

从上图的实验结果中可以得知：（1）在污染值较小的情况下，VGG 散斑重构网络模型与 U-net 散斑重构网络模型仍可以保持较好的重构效果，当污染值小于 100 的时候，准确率模型对重构图像识别的准确率均大于 88%。（2）ResNet 散斑重构网络模型受污染值影响较明显，波动较大，但是 SSIM 均值和准确率模型对重构图像识别的准确率总体呈现一个下降的趋势。（3）三种网络模型均体现了较好的鲁棒性，能够在一定程度上抵抗高斯噪声的影响，随着污染值的增加，重构效果也在波动状态下呈现一个下降的趋势。

（2）椒盐噪声

同样利用 VGG 散斑重构网络模型、改进 ResNet 散斑重构网络模型、U-net 散斑重构网络模型对 p 值分别为 0.9、0.7、0.5、0.3、0.1 污染后的图像进行训练。利用准确率模型对测试集图像进行测试，并求得测试集重构图像的 SSIM 均值，将两项评价图像重构质量的指标绘制如图 4-30 所示，同样主要以这两项指标来衡量模型的抗干扰能力大小，暂不讨论其可训练参数与训练时间。这里需要说明的是当网络模型无法对散斑图像进行重构的时候，将 SSIM 均值和准确率都看做是 0。

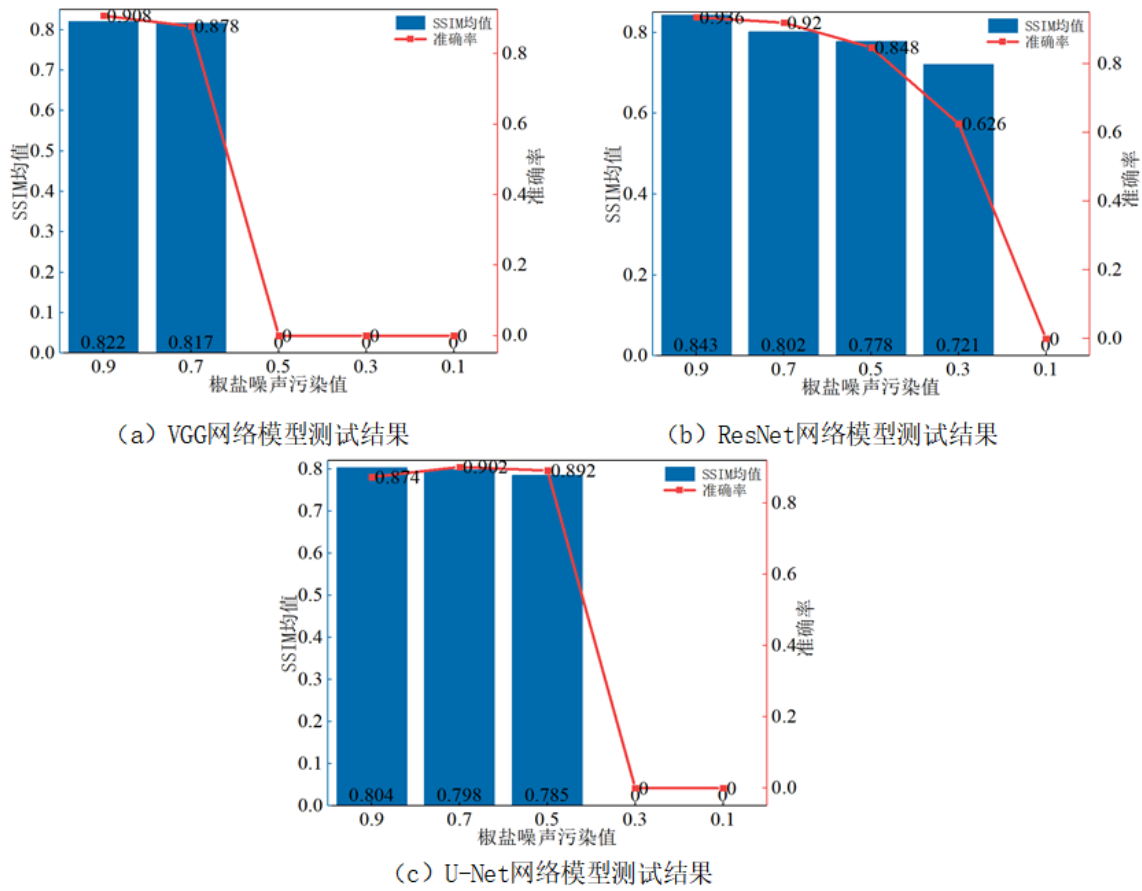


图 4-30 不同网络模型关于椒盐噪声的准确率和 SSIM 均值变化

同样以图像标签为数字“2”为例，不同污染值下的重构图像如图 4-31 所示。

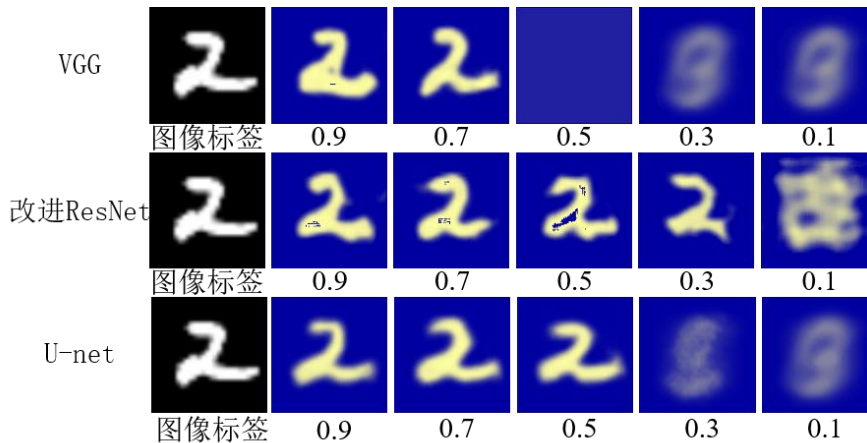


图 4-31 不同网络模型在不同椒盐污染值下重构结果

从 SSIM 均值与准确率图像和图像标签为数字“2”的散斑重构图像可知：（1）当 p 值为 0.9 和 0.7 时，三种网络模型均能对其有较好的重构效果，证明网络模型能够在一定程度上抵抗噪声带来的干扰。（2）当 p 值为 0.5 时，VGG 散斑重构网

网络模型已经不能对图像进行重构,改进 ResNet 散斑重构网络模型和 U-net 散斑重构网络模型仍具有较好的重构效果。(3) 当 p 值为 0.3 时,只有 ResNet 散斑重构网络模型能对其进行重构,VGG 散斑重构网络模型和 U-net 散斑重构网络模型均不能对其进行重构。(4) 当 p 值为 0.1 时,噪声影响太大,导致原始图像失真太严重,三种散斑重构网络模型均不能对其进行重构。

利用三种网络模型分别对高斯噪声和椒盐噪声处理后的图像进行训练与测试,证明在噪声值小于阈值时,散斑重构网络模型人能够对其进行重构,具有较好的鲁棒性,同时也随着噪声值的加大,重构效果会在波动中呈现一个下降的趋势,直至噪声太大将原始图像信息淹没,导致散斑重构网络模型无法对其进行重构。

4.7 本章小结

本章主要完成了以下工作:

(1) 本章首先给出了对散斑重构图像的评价指标,主要包括准确率和相似性两大指标,训练的准确率模型提高了判断重构图像与对应的标签图像是否为相同数字的效率。两项指标为评价模型的优劣和对模型进行修改提供了理论支撑。

(2) 不同的散斑重构网络模型在散斑重构图像的评价指标、训练时间、可训练参数等方面有较大差异。分别依据 VGG-16 网络模型、ResNet 网络模型、U-net 网络搭建了相应的散斑重构网络模型,实验使用 4500 张手写数字作为训练集,500 张作为测试集。基于 VGG-16 散斑重构模型实验结果表明,随着训练轮数增加,重构效果会变好,但是当轮数增加到一定数值时,重构效果无明显变化。基于 ResNet 的改进散斑重构网络模型在可训练参数方面要比 VGG 散斑重构网络模型少 825062 个,但是训练 100 轮所用的训练时间增加了 23.285 min。基于 U-net 散斑重构网络模型与 VGG 散斑重构网络模型相比在可训练参数方面没有较大优势,但是达到相同的重构效果训练时间减少了 3.99 min。

(3) 为验证散斑重构网络模型的鲁棒性,给散斑图像分别增加了高斯噪声和椒盐噪声。实验结果表明,随着污染值加大,重构效果呈现一个下降的趋势。当污染值超过一定阈值时,网络模型不能再对散斑图像进行重构。

第 5 章 迁移学习的应用及实验结果

5.1 引言

温故而知新，可以为师矣。能否在旧的训练集上进行训练，将训练得到的模型运用到新的训练集上，从而降低新的训练集的数量，缩短训练时间，达到事半功倍的效果呢？本章主要首先介绍了有关迁移学习的相关理论，而后利用迁移学习的相关理论将第四章中搭建的卷积神经网络模型进行迁移学习，对比了不同卷积神经网络模型迁移前和迁移后的相关指标。

5.2 迁移学习概述

迁移学习顾名思义就是将在一个方面学习到的成果运用到另一方面，从而节约在另一方面学习的时间，提高学习效率和学习质量。例如，如果我们学会了 java 语言，在学习 python 语言时，就可以将在 java 语言学习到的内容迁移到 python 语言上，我们再学习 python 语言时就可以学习 python 语言独有的内容，这样不仅可以节约学习时间，还可以取得非常好的学习效果。但是并非所有内容都可以进行迁移学习，源领域和目标领域两者之间具有某种相似性才可以进行，否则可能会出现“东施效颦”的结果。迁移学习主要有三个关键问题：1) 何时迁移，我们首先需要判断源域任务和目标域任务之间能否使用迁移学习，使用迁移学习之后的效果能否比未迁移效果要好；2) 何处迁移，在确定能够使用迁移学习方法后，就需要确定何处迁移，从网络模型的那个地方进行迁移可以有较好的效果；3) 如何迁移，当我们解决了前两个关键问题后，就需要确定如何迁移，如何迁移对应着众多迁移学习方法，选择适合的迁移学习方法成为关键问题。

迁移学习的方法根据不同的分类准则有不同的分类结果，在 Qiang Yang 的关于迁移学习的综述中^[32]，将迁移学习分为四大类：基于实例的迁移学习方法、基于特征的迁移学习方法、基于关系的迁移学习方法、基于模型的迁移学习方法。四种迁移学习的方法描述如表 5-1 所示。本次迁移学习实验的验证源域采用手写字母图像标签与对应的散斑图案制作成源域，共包含 50000 张图像。目标域采用第四章中使用的数据集，其中训练集只采用 4500 中的前 500 张，测试集不变。由于目标域数据集减少了，且目标域与源域的散斑数据相似度很高，所以这里采用的迁移学习的方法为基于模型的迁移学习。基于模型的迁移学习过程图如图 5-1 所示。利用源域字母图像数据集对源域网络模型进行预训练，得到一个具有泛化能力的模型，将得到的预训练网络模型迁移到目标域网络模型上，利用目标域网络模型对数字

图像数据集进行训练，实现降低目标域数据集和训练时间，并且达到与使用大数据集相同的效果，从而实现使用迁移学习的目的。

表 5-1 不同迁移学习方法与描述

方法	描述
基于实例的迁移学习	通过权重重用，对源域和目标域的样例进行迁移。不同的样本赋予不同的权重，相似的样本赋予较高的权重，完成迁移学习。
基于特征的迁移学习	通过对特征进行变换来完成迁移，例如源域和目标域的特征原来不在一个空间内，将他们通过一定手段变换到一个空间里，这样这些特征的相似性将会大大增加。
基于关系的迁移学习	主要挖掘和利用不同领域之间相似的关系进行迁移。例如老师上课、学生听课可以类比为公司开会场景。
基于模型的迁移学习	构建参数共享的模型，将模型的参数在源域和目标域中进行共享实现迁移。

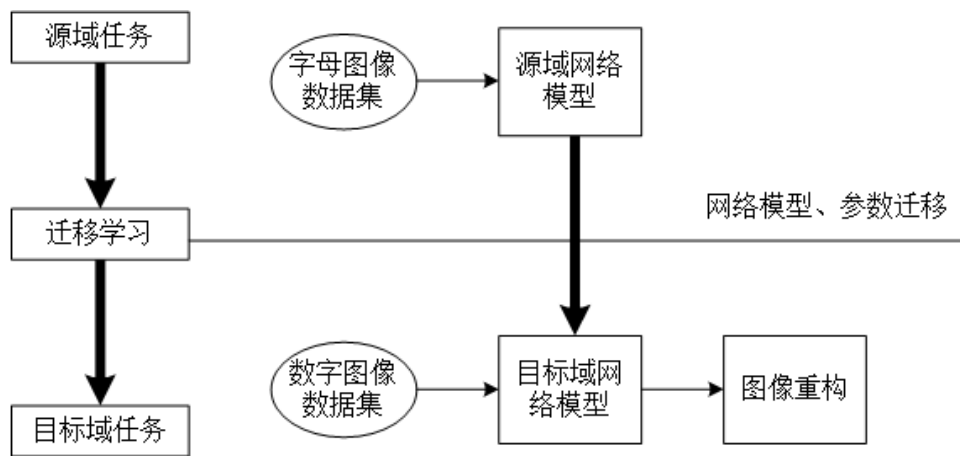


图 5-1 迁移学习过程图

5.3 三种网络模型的迁移学习

5.3.1 VGG 散斑重构网络模型的迁移学习

首先利用 4.3.1 节搭建的 VGG 散斑重构网络模型对源域数据进行训练，源域数据训练 50 轮损失函数如图 5-2 所示，训练集的损失函数收敛于 0.004，证明该网络在源域上有较好的学习效果，能够将在源域数据的低级特征运用在目标域上，从而实现目标域不再学习这些低级特征，提高在目标域模型上的收敛速度。

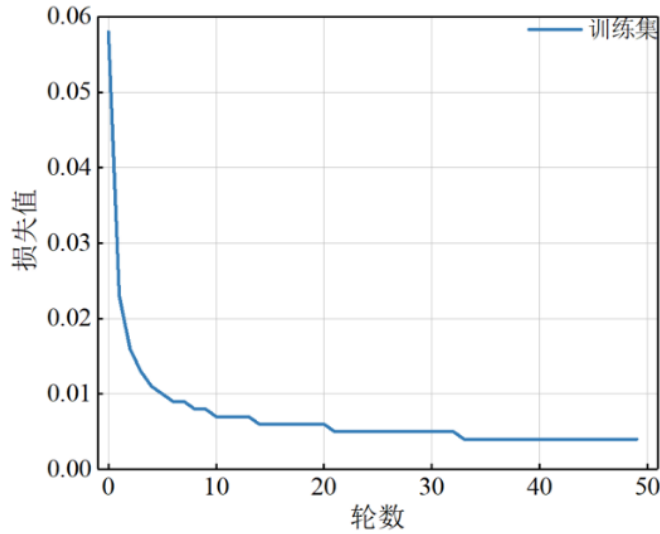


图 5-2 VGG 散斑重构网络源域模型损失函数

下面主要将在源域上训练好的网络模型运用在目标域数据集上，原理过程如图 5-3 所示。

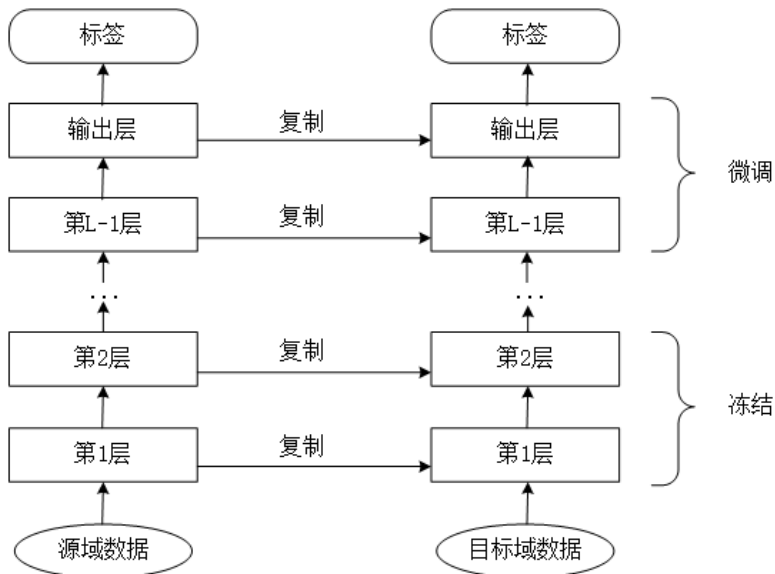


图 5-3 模型迁移原理图

在 VGG 散斑网络重构模型中不包括输入层共有 31 层，对哪些层的源域网络模型进行冻结，对哪些层进行微调就是下面研究的关键问题。当 L 取不同值的时候，会有不同的迁移效果，通过不同的迁移效果来探究合适的 L 。先取 L 为 7、16、31 来对网络模型进行实验。对 L 值前的源域网络模型进行冻结，保留在源域上得到的预训练权重；对 L 值后的源域网络模型进行微调，使其更加适应目标域。经过 50 轮训练，不同 L 值的重构效果如图 5-4 所示。

图像标签										
L=7										
SSIM值	0.856	0.834	0.800	0.846	0.859	0.810	0.852	0.883	0.841	0.864
L=16										
SSIM值	0.794	0.812	0.782	0.864	0.811	0.766	0.803	0.828	0.826	0.829
L=31										
SSIM值	0.767	0.661	0.725	0.771	0.750	0.730	0.740	0.774	0.783	0.826

图 5-4 不同 L 值迁移结果

当 L=7，对测试集中散斑图案重构的 SSIM 均值为 0.845，准确率为 0.974，训练时间为 2.012 分钟；当 L=16，对测试集中散斑图案重构的 SSIM 均值为 0.825，准确率为 0.966，训练时间为 2.011 分钟；当 L=31，对测试集中散斑图案重构的 SSIM 均值为 0.769，准确率为 0.958，训练时间为 1.431 分钟；当 L 值增大时，预训练网络模型被冻结的层数就会增加，可进行微调的层数就会减少，通过实验结果可以看出当 L 值增大时，重构图像的 SSIM 均值会有略微的降低，也就是重构效果会变差。神经网络的前几层基本都是通用特征，迁移的效果比较好，在深度迁移网络中加入微调，可以较好的克服数据之间的差异性，效果提升较大，可能会比原始网络模型的效果还好。现将未使用迁移学习和使用迁移学习（L=7）的各项指标对比如表 5-2。

表 5-2 未使用迁移学习和使用迁移学习（L=7）的各项指标对比

类别	训练集数量	训练时间（分）	SSIM 均值	准确率
未迁移	4500	12.733	0.831	93.4%
迁移	500	2.012	0.845	97.4%

通过对比迁移学习和未使用迁移学习的各项指标，充分表明使用迁移学习在散斑图像重构方面的优势，通过次实验也充分回答了迁移学习概述中的一个关键问题-何处迁移。

5.3.2 改进 ResNet 散斑重构网络模型的迁移学习

通过在 VGG 散斑重构网络模型上使用迁移学习，可以看到迁移学习巨大的优势，本小节继承上一小结讨论的结果，选择合适的 L 值（L=29），在字母数据集上进行预训练。后将与训练结果迁移到数字数据集上，网络结构图如图 5-5 所示，其中残差块 1 和残差块已经在 4.4.3 和 4.4.1 小节进行了介绍，损失函数与原来一致。

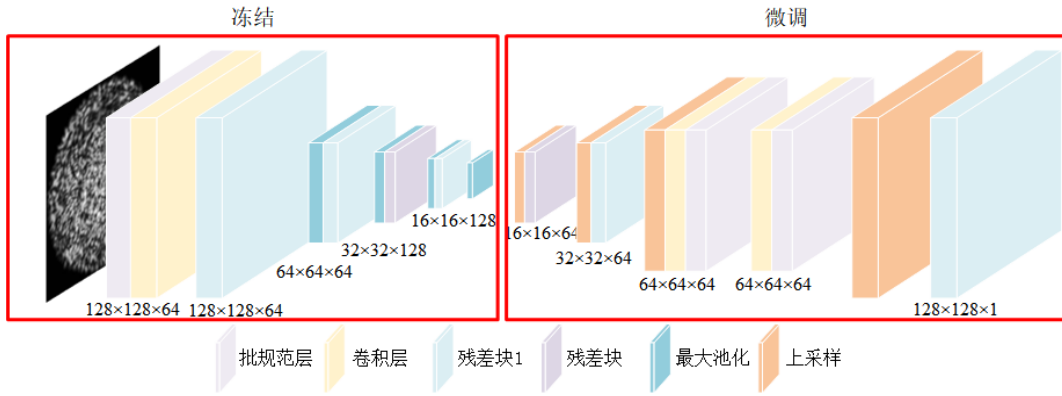


图 5-5 迁移网络结构图

将在字母数据集上预训练的网络模型中的下采样部分进行冻结，保留提取通用特征的权重。利用 500 张数字训练集对上采样部分进行微调，重构效果如图 5-6 所示。训练时间为 5.881 分钟，500 张测试集散斑图像重构的 SSIM 均值为 0.885，准确率为 98.8%。

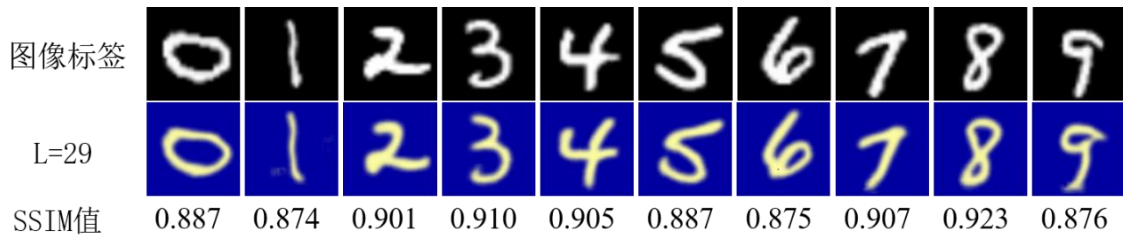


图 5-6 改进 ResNet 迁移后重构结果

使用迁移学习和未使用迁移学习的混淆矩阵如图 5-7 所示，从重构结果的 SSIM 均值、准确率和训练时间来看，对于改进 ResNet 散斑重构网络模型使用迁移学习在数字数据集上获得较好效果。

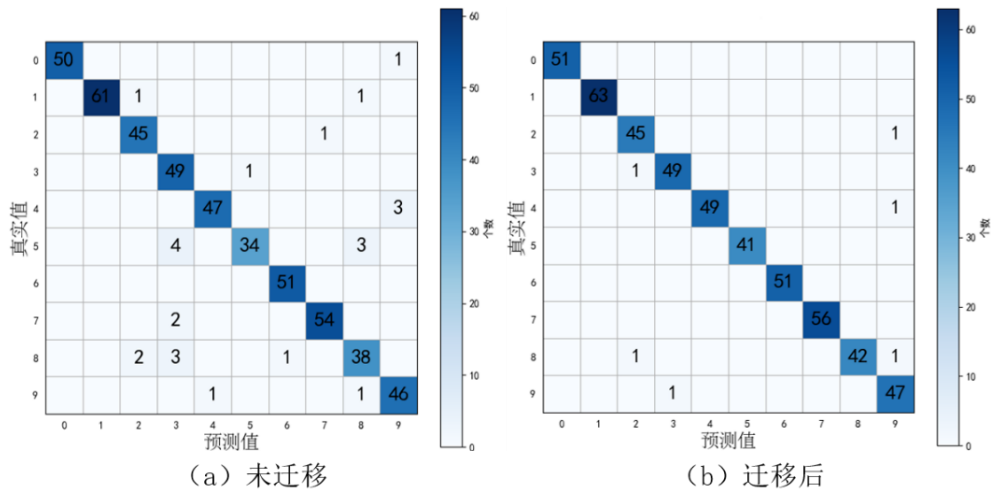


图 5-7 使用迁移学习和未使用迁移学习的混淆矩阵

迁移后的网络模型不仅可以进行微调，还可以对网络模型进行修改，将网络模型输出端的 $128 \times 128 \times 1$ 改为 $64 \times 64 \times 1$ ，需要在 $128 \times 128 \times 1$ 层后面增加一个池化层和卷积层，再经过最后的卷积层变为 $64 \times 64 \times 1$ ，网络结构图如图 5-8 所示，

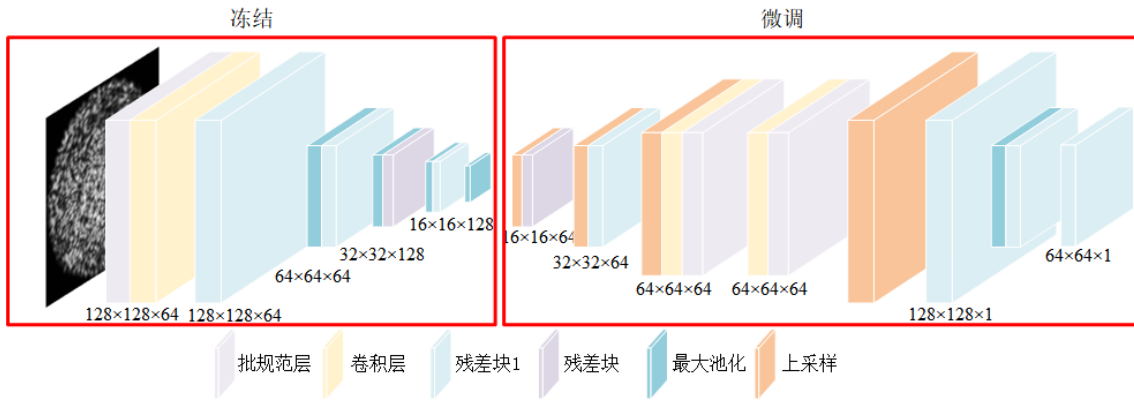


图 5-8 修改后 ResNet 迁移网络结构

其余部分与上一方法相同，500 张测试集重构结果的 SSIM 均值为 0.911，准确率为 97.6%。按照同样的方法，在上述基础上再增加一个池化层和卷积层将输出大小变为 $32 \times 32 \times 1$ ，500 张测试集重构结果的 SSIM 均值为 0.837，准确率为 97.2%。将输出分别为 $128 \times 128 \times 1$ 、 $64 \times 64 \times 1$ 、 $32 \times 32 \times 1$ 的网络模型对 500 张测试集重构结果表示为图 5-9 所示，根据上述实验结果，从 SSIM 均值角度来考虑，输出为 $64 \times 64 \times 1$ 的网络模型重构效果最好。对于基于改进 ResNet 散斑重构网络模型，使用迁移学习和未使用迁移学习在各项指标上均具有明显的优势。

图像标签										
128										
SSIM值	0.887	0.874	0.901	0.910	0.905	0.887	0.875	0.907	0.923	0.876
64										
SSIM值	0.934	0.861	0.901	0.922	0.907	0.910	0.910	0.916	0.942	0.910
32										
SSIM值	0.921	0.918	0.917	0.874	0.886	0.891	0.909	0.878	0.931	0.910

图 5-9 修改后 ResNet 迁移网络结构重构结果

5.3.3 U-net 散斑重构网络模型的迁移学习

对于使用 U-net 散斑重构网络模型进行迁移学习，选择合适的 L 值 ($L=21$)，并使目标域网络模型的输出为 $64 \times 64 \times 1$ ，网络结构图如图 5-10 所示，损失函数与

原网络模型一致。

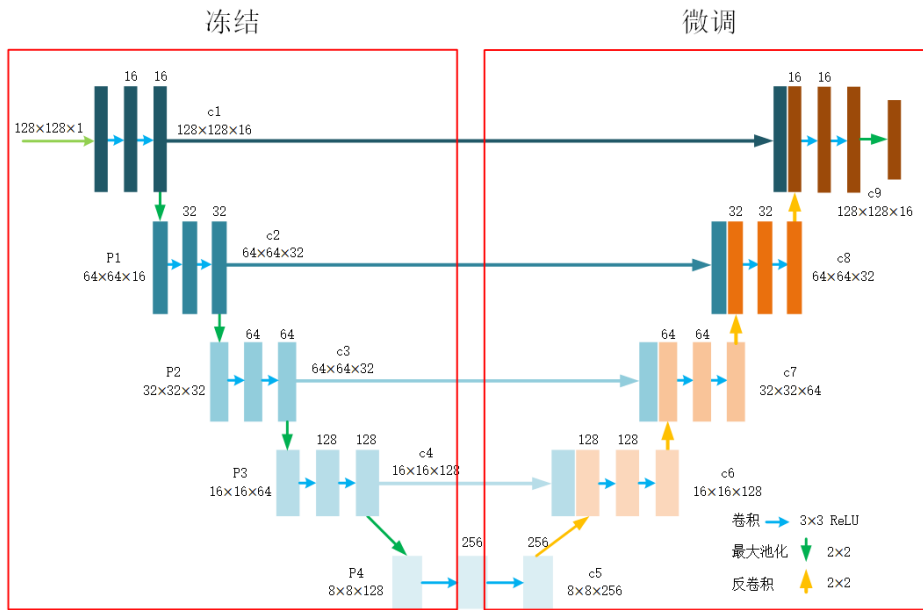


图 5-10 修改后 U-net 迁移网络结构

将使用未迁移的网络模型在 4500 张数字训练集上进行训练的损失函数和使用迁移学习后在 500 张数字训练集上进行训练的损失函数图像绘制如图 5-11 所示，未迁移网络模型的损失函数从 0.5 开始随着训练轮数逐渐下降，最终收敛于 0.131，使用迁移学习后损失函数从 0.274 开始下降，并最终收敛于 0.159。从图中可以看出使用迁移学习后，网络模型可以以更小的损失值快速收敛，使用迁移学习在 500 张数字训练集上所用训练时间为 1.502 分钟，训练时间是未使用迁移学习的 17.18%。达到了减少训练集数量，缩短训练时间的目标。

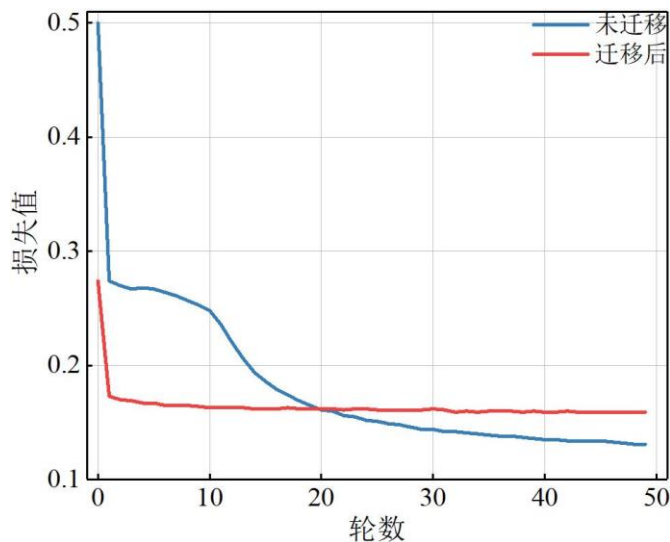


图 5-11 迁移前与迁移后损失函数

使用迁移学习后在 500 张测试集上的重构图像如图 5-12 所示，SSIM 均值为 0.856，准确率为 95.8%。未使用迁移学习的 SSIM 均值为 0.813，准确率为 92.0%，SSIM 均值提高了 0.043，准确率提高了 3.8%。





















图像标签										
64										
SSIM值	0.867	0.801	0.837	0.874	0.899	0.808	0.830	0.864	0.904	0.870

图 5-12 修改后 U-net 迁移网络结构重构结果

将 VGG、改进 ResNet、U-net 散斑重构网络模型使用迁移学习前和使用迁移学习后对 500 张测试集图像的 SSIM 均值整理如图 5-13 所示。从图中可以看出使用迁移学习后不仅达到了未使用迁移学习的重构效果（优于未使用迁移学习），还减少了使用的数据量和训练时间，采用迁移学习对将其应用在内窥镜方面有很大的优势。

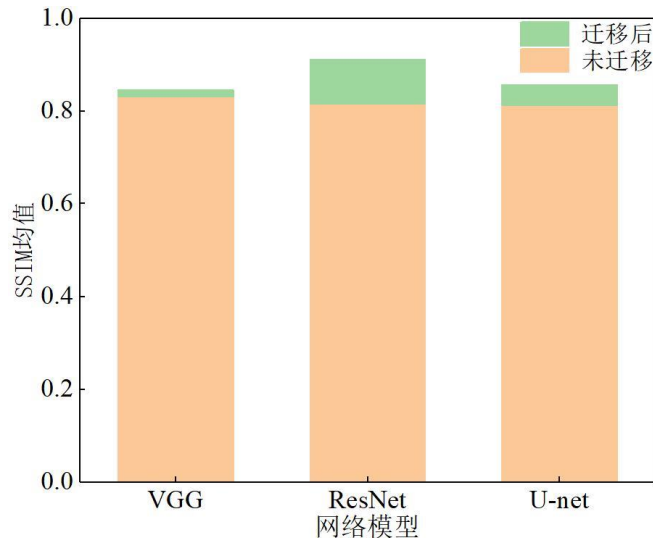


图 5-13 迁移前与迁移后三种网络模型的 SSIM 均值

5.4 本章小结

针对卷积神经网络模型需要大量数据的问题，本章使用了迁移学习来降低目标域的数据量和训练时间。本章主要完成了以下工作：

(1) 本章首先对迁移学习做了总体介绍，根据源域和目标域的相似性选择了基于模型的迁移学习方法。源域选择手写字母并对模型进行预训练，目标域选择手写数字并对模型进行微调。

(2) 基于 VGG-16 网络模型作为预训练网络模型，探究了“何处迁移”这个重

要问题,通过对比 L 值为 7、16、31 时的重构结果,确定了 L 值为 7 时, SSIM 均值为 0.845, 重构效果为最佳。

(3) 基于 ResNet 网络模型作为预训练网络模型,探究了对目标域网络模型进行修改后的迁移结果,通过对比输出尺寸为 $128 \times 128 \times 1$ 、 $64 \times 64 \times 1$ 、 $32 \times 32 \times 1$ 时的重构结果,当输出为 $64 \times 64 \times 1$ 时,重构结果的 SSIM 均值为 0.911, 重构效果最好。

(4) 基于 U-net 网络模型作为预训练网络模型,选择 L 值为 21, 确定输出尺寸为 $64 \times 64 \times 1$, 对比了未使用迁移学习和使用迁移学习的损失函数变化,使用迁移学习损失函数可以在 10 轮快速收敛,训练时间是未使用迁移学习的 17.18%。最后给出了三种网络模型未使用迁移学习和使用迁移学习后重构结果 SSIM 均值的对比, 三种网络模型使用迁移学习后 SSIM 均值均高于未使用迁移学习。

结论与展望

光纤作为光信号的传输介质具有传输大容量信息、不受电磁场干扰、使用简单等优点,利用光纤制成的内窥镜在工业探伤、体内诊疗等方面有着广泛应用。目前由多根单模光纤构成的光纤束在内窥镜中使用比较广泛,但是存在一定的缺点。多模光纤可以允许多个光学模式并行传输,独立实现大范围区域成像,将多模光纤运用在内窥镜上具有较好的发展前景。由于多模光纤是一个散射介质,当光信号透过多模光纤时,在光纤的输出端得到的是一个散斑图案,目前使用较多的方法为利用卷积神经网络对散斑图像进行重构,但是使用卷积神经网络模型需要大量的数据和较长的训练时间。本文针对该问题,在使用卷积神经网络模型对散斑图像重构的基础上使用了迁移学习,减少了目标域对数据量和训练时间的需求。

(1)完成了对卷积神经网络需要的散斑数据的采集。通过对阶跃多模光纤和渐变多模光纤的特点和传输条件分析,选择了阶跃多模光纤作为本实验的传输介质;实现了对散斑图像数据的自动采集,并对数据进行了裁剪、缩小、三通道转换为单通道处理,使数据集满足了卷积神经网络对输入数据的要求。

(2)基于VGG-16、ResNet、U-net搭建了散斑重构网络模型。在重构准确率和SSIM均值相当的情况下,对比了三种网络模型在可训练参数和训练时间两个指标。基于ResNet的改进散斑重构网络模型在可训练参数方面要比VGG散斑重构网络模型减少了825062个,但是训练100轮所用的训练时间增加了64.65%。基于U-net散斑重构网络模型与VGG散斑重构网络模型相比在可训练参数方面没有较大优势,但是达到相同的重构效果训练时间减少了31.34%。

(3)将迁移学习应用在三种网络模型上,成功实现了减少目标域对数据量和训练时间的需求。三种网络模型源域使用手写字母数据集,目标域使用手写数字数据集,基于VGG-16网络模型作为预训练网络模型,通过对比L值为7、16、31时的重构结果,当L值为7时,SSIM均值为0.845,重构效果最佳;基于ResNet网络模型作为预训练网络模型,通过对比输出尺寸为 $128 \times 128 \times 1$ 、 $64 \times 64 \times 1$ 、 $32 \times 32 \times 1$ 时的重构结果,当输出为 $64 \times 64 \times 1$ 时,重构结果的SSIM均值为0.911,重构效果最好;基于U-net网络模型作为预训练网络模型,选择L值为21,确定输出尺寸为 $64 \times 64 \times 1$,目标域损失函数在10轮快速收敛。三种网络模型使用迁移学习后,训练时间分别是未使用迁移学习的15.71%、16.32%、17.18%。

展望:关于此课题还有很多待研究的问题。例如,目前包括大量文献中的研究都是基于手写数字和手写字母等简单图像进行重构,对于一些复杂的真实的腔内

图片还没有进行重构研究。由复杂的真实的腔内图片难以大量获得，这时本课题研究的迁移学习就可以大展身手，利用易获得的散斑数据作为源域，真实复杂的腔内图片作为目标域，对复杂图像进行研究，为将该技术真正落地使用奠定基础。采集数据需要大量时间，能否将迁移学习和清华大学的研究成果结合起来也是该课题要深入研究的问题。

综上，使用卷积神经网络模型可以对多模光纤的输出进行重构，并利用迁移学习可以降低网络模型使用的数据量和训练时间，为将该技术应用于内窥镜奠定了基础。

参考文献

- [1] 原荣. 光纤通信技术[M]. 北京: 机械工业出版社, 2020: 36-39.
- [2] Sun B, Edgar M P, Bowman R, et al. 3D computational imaging with single-pixel detectors[J]. *Science*, 2013, 340(6134): 844-847.
- [3] Belykh E, Miller E J, Hu D, et al. Scanning fiber endoscope improves detection of 5-ala induced protoporphyrin IX fluorescence at the boundary of infiltrative glioma[J]. *World Neurosurgery*, 2018: e51.
- [4] Long Z, Nagamun K. Underwater surface reconstruction of narrow space by endoscope and optical fiber: application to minimally invasive surgery[C]//37th IEEE Annual International Conference Engineering in Medicine and Biology Society (EMBC). 2015.
- [5] Tanabe T, Yokota A, Sugie S. Cardiovascular fiber optic endoscopy- development and clinical-application[J]. *Surgery*, 1980, 87(4): 375-379
- [6] Bahcall J K, Barss J T. Fiber optic endoscope usage for intracanal visualization[J]. *Journal of Endodontics*, 2001, 27(2): 128-129.
- [7] Korczewski Z. Analysing the potential for application of the phase shift method in endoscopic examination of marine engines[J]. *Polish Maritime Research*, 2013, 20(1): 23-30.
- [8] Jiotode Y, Agarwal A K. In-cylinder combustion visualization of jatropha straight vegetable oil and mineral diesel using high temperature industrial endoscopy for spatial temperature and soot distribution[J]. *Fuel Processing Technology*, 2016, 153: 9-18.
- [9] Sanchez E J, Novotny L, Xie X S. Near-field fluorescence microscopy based on two-photon excitation with metal tips[J]. *Physical Review Letters*, 1999, 82(20): 4014-4017.
- [10] Chen Z, Jiang Z, Gui W, et al. A novel device for optical imaging of blast furnace burden surface: Parallel low-light-loss backlight high-temperature industrial endoscope[J]. *Ieee Sensors Journal*, 2016, 16(17): 6703-6717.
- [11] Spaner S J, Warnock G L. A brief history of endoscopy, laparoscopy, and laparoscopic surgery. *J Laparoendosc Adv Surg Tech A*. 1997;7(6):369-73.
- [12] Hughes M, Chang T P, Yang G Z. Fiber bundle endocytoscopy[J]. *Biomedical Optics Express*, 2013, 4(12): 2781-2794.
- [13] Gmitro A F, Aziz D. Confocal microscopy through a fiber-optic imaging bundle[J]. *Optics Letters*, 1993, 18(8): 565-567.
- [14] Loterie DC M. Microscopy and digital light shaping through optical fibers[R]. EPFL,

- 2017.
- [15] Rahmani, Babak, Oguz, Ilker, et al . Demetri and Moser, Christophe. Learning to image and compute with multimode optical fibers[J] *Nanophotonics*, 2022, 11(6) 1071-1082.
- [16] Yaqoob Z, Psaltis D, Feld MS. Optical phase conjugation for turbidity suppression in biological samples[J]. *Nature Photonics*, 2008; 2(2): 110-115.
- [17] Dunning G J, Lind R C. Demonstration of image transmission through fibers by optical phase conjugation[J]. *Optics Letters*, 1982, 7(11): 558-560.
- [18] Gover A, Lee C P, Yariv A. Direct transmission of pictorial information in multimode optical fibers[J]. *JOSA*, 1976, 66(4): 306-311.
- [19] Papadopoulos I N, Farahi S, Moser C, et al. Focusing and scanning light through a multimode optical fiber using digital phase conjugation[J]. *Optics Express*, 2012, 20(10): 10583-10590.
- [20] Papadopoulos I N, Farahi S, Moser C, et al. High-resolution, lensless endoscope based on digital scanning through a multimode optical fiber[J]. *Biomedical Optics Express*, 2013, 4(2): 260-270.
- [21] Choi Y, Yoon C, Kim M, et al. Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber[J]. *Physical Review Letters*, 2012, 109(20): 203901.
- [22] Gu R Y, Mahalati R N, Kahn J M. Design of flexible multi-mode fiber endoscope[J]. *Optics Express*, 2015, 23(21): 26905-26918.
- [23] Loterie D. et al. Digital confocal microscopy through a multimode fiber. *Opt. Express*, 2015, 23, 23845–23858.
- [24] Deng L, Yu D. Deep Learning: Methods and Applications[J]. *Foundations & Trends in Signal Processing*, 2014, 7(3):197-387.
- [25] Geert, Litjens, Thijs, et al. A survey on deep learning in medical image analysis[J]. *Medical Image Analysis*, 2017, 42: 60-88
- [26] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 1251-1258.
- [27] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 652-660.
- [28] Tang Y. Deep learning using linear support vector machines[J]. *arXiv preprint arXiv:1306.0239*, 2013.
- [29] Hussain M, Bird J J, Faria D R. A study on cnn transfer learning for image classification[C]//*UK Workshop on computational Intelligence*. Springer, Cham,

- 2018: 191-202.
- [30] Gao Y and Mosalam K M. Deep transfer learning for image-based structural damage recognition[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2018,33(9): 748-768
- [31] Larsen-Freeman D. Transfer of learning transformed[J]. *Language Learning*, 2013, 63: 107-129
- [32] Pan S J, Qiang Y. A Survey on Transfer Learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10):1345-1359.
- [33] Matsumoto T, Koga M, Noguchi K, et al. Proposal for neural-network applications to fiber-optic transmission[C]//1990 IJCNN International Joint Conference on Neural Networks. IEEE, 1990: 75-80.
- [34] Rahmani B, Loterie D, Konstantinou G, et al. Multimode optical fiber transmission with a deep learning network[J]. *Light: Science & Applications*, 2018, 7(1): 1-11.
- [35] Borhani N, Kakkava E, Moser C, et al. Learning to see through multimode fibers[J]. *Optica*, 2018, 5(8): 960-966.
- [36] Kakkava E, Borhani N, Rahmani B, et al. Deep learning-based image classification through a multimode fiber in the presence of wavelength drift[J]. *Applied Sciences*, 2020, 10(11): 3816.
- [37] Resisi S, Popoff S M, Bromberg Y. Image transmission through a dynamically perturbed multimode fiber by deep learning. *Laser & Photonics Reviews*, 2021, 15(10), 2000553.
- [38] Linh V, Nguyen, Cuong C, et al. Sensing in the presence of strong noise by deep learning of dynamic multimode fiber interference[J]. *Photonics Research*, 2021, 9(4): 0400B109
- [39] Fan P, Ruddlesden M, Wang Y, et al. Learning Enabled Continuous Transmission of Spatially Distributed Information through Multimode Fibers[J]. *Laser & Photonics Reviews*, 2021, 15(4): 2000348.
- [40] Wang P, Di J. Deep learning-based object classification through multimode fiber via a CNN-architecture SpeckleNet[J]. *Applied Optics*, 2018, 57(28): 8258-8263.
- [41] 孟球, 胡海峰, 胡金洲, 等. 基于深度学习的多模光纤散射介质成像重建[J]. *中国激光*, 2020, 47(12): 272-281.
- [42] Liu Z, Wang L, Meng Y, et al. All-fiber high-speed image detection enabled by deep learning[J]. *Nature Communications*, 2022, 13(1): 1-8.
- [43] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [44] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. *Advances in neural information processing*

- systems, 2012, 25.
- [45] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.
- [46] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [47] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [48] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv: 2012: 1207.0580.
- [49] Cohen, G, Afshar, S, Tapson J, van Schaik. A. EMNIST: An extension of MNIST to handwritten letters. Preprint at <https://arxiv.org/abs/1702.05373> (2017).
- [50] 倪臣敏. 数字图像滤波算法的研究与应用[M]. 北京: 电子工业出版社, 2020. 10: 3-7.

哈尔滨工业大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于卷积神经网络的多模光纤图像重构研究》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：

日期： 年 月 日

学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

致 谢

光阴似箭，转眼间硕士生活已接近尾声，回想这两年的生活，收获颇多，非常感谢遇到的每一位师长和同学。

衷心感谢导师王正佳老师对我的耐心指导。王老师于我来讲亦师亦友，在学术上，王老师不嫌我愚笨，总能从多角度考虑问题，给我启发和灵感，让我对课题有更深入的认识。在生活上，王老师对我们关怀备至，记得来学校第二天和老师的交流让我从一个很不自信的人变得积极自信起来，虽然现在在一些重要场合还是紧张，但是相比之前要好很多。老师严谨的治学态度和务实的工作作风给我带来了诸多的教益和启迪。

感谢张勇老师提供的实验设备和实验指导，在实验过程中遇到困难，张老师亲自帮我调试、检查实验设备。张老师严谨的工作态度对我产生了很大影响。

感谢孙晔老师和秦峰老师百忙之中对论文的指导。

感谢1006的罗健、黄飞强、张成刚、徐嵩浩、韩义、蔡子艺、周劼。感谢罗健在人文素养方面的“培养”；感谢张成刚在学习和工作上的指导；感谢徐嵩浩提供的各种零食，以一己之力提升了整个宿舍的消费水平；感谢黄飞强、韩义、蔡子艺、周劼在羽球上的指导。感谢声光所的所有同学们，和你们一起吃饭、玩耍、学习是最快乐的时光，也将成为我一生中最美好的回忆，愿各位前程似锦。

感谢父母亲人的支持与鼓励，感谢成长路上遇到的每一位帮助过我的人。