

中国科学技术大学

学士学位论文



基于深度学习的方法实现透过散射介质 成像

作者姓名： 韦璐

学科专业： 物理学

导师姓名： 王自强 老师

完成时间： 二〇二一年五月二十四日

University of Science and Technology of China
A dissertation for bachelor's degree



Deep learning based imaging for scattering medium

Author: Wei Lu

Speciality: Physics

Supervisor: Wang Ziqiang

Finished time: May 24, 2021

致 谢

光阴荏苒，白驹过隙，转瞬之间在科大已经经历了四年的寒暑更迭。曾经那些或曾欢乐或曾痛苦，或曾踌躇满志或曾受挫难过的日子都在这个夏天里逐渐飞驰远去，逐渐成为我的一段珍贵回忆，在我往后的人生里面投下或明或暗的影子，陪伴我踏上新的旅途。

回顾这四年的生活，我首先要感谢中国科学技术大学，它是我这几年生活的地方，是我的归属，我的第二故乡。在这里宽松自由、积极向上的氛围中我度过了非常充实的本科时光。学校领导和老师永远都把学生当成学校的主人，生活中每一个学生都能体会到学校领导、教职工给予的关怀和优先待遇，学习上每个学生都能体会到老师近乎苛刻的严厉，以及周边同学的刻苦、奋进！课堂上、实验室里老师的谆谆教诲、旁征博引、一丝不苟的治学态度，学术讲座上各种学术大咖们博学、精彩的分享，图书馆里昼夜不眠的灯火，学术讨论会上面红耳赤的争论都给我留下了深刻的印象。中科大特有的宽松自由的学术科研氛围、良好的学习和工作平台，让我在感受学习竞争压力巨大的同时，也深深的享受着自然科学的巨大魅力！虽然我自知是一只菜鸟，但是我既然勇敢的飞进了中科大校园，我就知道我必须要付出比别人更多的努力，四年来我从未敢放松过。其次我要感谢物理学院和少年班学院，学院的细致安排和培养给了我踏入物理学研究的准备，让我学习了基本知识并逐渐走进了前沿的科学研究。

四年紧张、充实的学习生活过的真快，转眼就来到了紧张的毕业论文研究撰写的毕业季。在本论文的研究工作中，我首先要感谢我的导师王自强老师，是他给我提供了一个宽松自由的学术环境，有问题时他都给予了我悉心的指导。其次我要感谢实验室的师兄师姐，他们给了我许多热情、无私的帮助。我还要感谢伍新明老师和徐云老师，他们给我提供了他们的计算资源，让我能够进行我的研究工作。

其次，我也要感谢曾蓓老师，她邀请我到深圳鹏程实验室交流的两个月期间，我学到了许许多多，也认识了许多朋友，度过了一段非常难忘的时光。我也要感谢学长贾治安，感谢他的讨论和交流。

我还要感谢我的室友宋宗燕和原思浩，感谢她们一直以来的陪伴和鼓励。“浮甘瓜于清泉，沉朱李于寒水，白日既匿，继以朗月，同乘并载以游后园”，那些曾经一起努力学习，一起在社团，一起游玩的快乐的时光一定会成为我往后的美

好回忆。

最后我要感谢我的父母，他们总是在我的背后默默支持鼓励我，给我力量，让我在人生道路上更加有勇气去前进。

“红专并进，理实交融”是中科大的校训！在中科大学习了四年，让我对她有了更深刻的理解，中科大建校的宗旨是科技救国！作为一名科大人科技报国永远是我的初心。当前的世界，正在面临着百年未有的大变局，科学技术的发展水平决定了一个国家的发展水平和综合竞争力。中国现在是一个富裕的大国，但还远不是一个强国，要走进世界强国的行列我们还有很远的路要走，这是中国青年的责任，更是中国科技青年的责任！中华民族复兴大业令人向往，但那不是等来的，需要青年人去奋斗！我是一名科大人，“红专并进，理实交融”已经融入我的血液，我必将不忘初心，牢记科技强国的科大人的使命，永远不断的努力，希望成为不辱使命的永远的科大人！

目 录

中文内容摘要	3
英文内容摘要	4
第一章 绪论	6
一、激光散斑图像恢复问题	6
二、论文结构安排	6
第二章 传统的散斑图像恢复方法	8
第一节 散斑图像的物理成因及其分类	8
一、散斑的物理成因	8
二、散斑的分类	10
第二节 散斑图像的透射矩阵描述	11
一、透射矩阵的定义	11
二、透射矩阵的实验测量	13
第三节 传统散斑图像恢复方法	16
一、散斑图像恢复的基本原理	16
二、直接求逆方法	17
三、光学相位共轭方法	17
四、Tikhnov 正则化算法	18
五、基于场强的方法	20
第四节 本章总结	21
第三章 神经网络图像处理	22
第一节 机器学习	22
第二节 卷积神经网络	25
一、前馈神经网络	26
二、卷积神经网络	27
第三节 卷积神经网络在图像处理方向的改进版本	31
一、VGG 神经网络	32
二、ResNet 神经网络	33

三、Unet 神经网络	35
四、DenseNet 神经网络	36
第四节 本章总结	38
第四章 基于神经网络的散斑图像恢复方法	39
第一节 神经网络架构	39
第二节 数据生成	41
一、仿真方法	41
二、实验方法	42
第三节 神经网络散斑图像恢复结果	43
一、数据预处理	43
二、训练过程	43
三、恢复结果讨论	44
四、相关工作的进展	45
五、本工作后续可拓展的内容	47
第四节 本章总结	47
第五章 总结与展望	49
参考文献	51
在读期间发表的学术论文与取得的研究成果	56

中文内容摘要

散斑图像是一种常见的光学现象，当激光被散射介质散射或者被粗糙光学表面所反射之后会形成一种颗粒状的斑纹图案，这种图案被称为散斑图像。本文中详细讨论了散斑图像恢复问题，包括它的传统恢复方法以及利用基于神经网络方法进行恢复的方法，并将两种方法做了简单的比较。我们首先讨论了散斑图像的物理成因，它的分类以及应用。其次我们讨论了散斑图像的透射矩阵方法，然后讨论了几种传统的散斑图像恢复方法，包括直接求逆法；光学相位共轭法；以及 Tikhnov 正则算子法。最后我们还介绍了一种基于场强的恢复方法。在此基础上，为了利用神经网络进行散斑图像恢复问题，我们首先对神经网络的基本内容做了简要的阐述。我们首先介绍了各种机器学习常用的模型，尤其是对神经网络进行了详细的阐述，我们介绍了前馈神经网络和卷积神经网络的基本内容，然后对它们的基本网络架构做了详细的分析。其次，我们对卷积神经网络在图像恢复等图像处理问题中的应用做了详细的讨论，我们介绍了几种常见的被用来做图像处理的神经网络架构，包括 VGG 神经网络，Res-Net 神经网络，Unet 神经网络和 DenseNet 神经网络。基于此，我们详细讨论了如何利用神经网络来进行散斑图像恢复问题。我们所使用的神经网络架构是一种结合了 Unet 和 DenseNet 的神经网络架构，即，我们在 Unet 里面引入了密集块。接着我们利用透射矩阵的方法，利用 MNIST 手写数字识别图片库进行了散斑图像的模拟生成，我们利用生成的图像作为训练数据集对神经网络进行了训练。然后我们选取了 1,9,5,8 四个数字对我们的神经网络进行了测试。我们发现，我们的具有密集块的 Unet 架构的神经网络可以很好地进行散斑图像恢复问题。我们也对传统的散斑图像恢复方法进行了模拟，并以直接求逆法和场强方法为例进行了数值实验，我们发现，传统方法虽然在整体上表现的不错，但在细节处理上不如神经网络方法。传统方法在恢复图像中仍然携带了部分散斑，而神经网络方法在细节上表现的十分好。而且它还可以用同一网络对不同介质的散斑图像进行恢复，这是传统方法很难做到的。综上所述，神经网络方法在散斑图像恢复问题中表现很好，有非常大的应用潜力。

关键词：散斑图像恢复；计算机视觉；深度学习；卷积神经网络

Abstract

Speckle image is a common optical phenomenon. When laser light is scattered by a scattering medium or reflected by a rough optical surface, a granular speckle pattern is formed. This pattern is called a speckle image. This article discusses in detail the problem of speckle image restoration, including its traditional restoration methods and methods based on neural networks for restoration, and a simple comparison between the two methods. We first discussed the physical causes of speckle images, its classification and applications. Secondly, we discussed the transmission matrix method of speckle images, and then discussed several traditional speckle image restoration methods, including direct inversion method; optical phase conjugation method; and Tikhnov regular operator method. Finally, we also introduce a recovery method based on field strength. On this basis, in order to utilize the neural network to recover the speckle image, we first briefly explained the basic content of the neural network. We first introduced various commonly used models of machine learning, especially neural networks. We introduced the basic contents of feedforward neural networks and convolutional neural networks, and then detailed their basic network architecture analysis. Secondly, we discussed in detail the application of convolutional neural networks in image processing problems such as image restoration. We introduced several common neural network architectures used for image processing, including VGG neural network, ResNet Neural network, Unet neural network and DenseNet neural network. Based on this, we discussed in detail how to use neural networks to restore speckle images. The neural network architecture we use is a neural network architecture that combines Unet and DenseNet, that is, we introduce dense blocks into Unet. Then we use the transmission matrix method to simulate the generation of speckle images using the MNIST handwritten digit recognition picture library, and we use the generated images as the training data sets to train the neural network. Then we selected four numbers 1, 9, 5, and 8 to test our neural network. We found that our neural network with dense block Unet architecture can perform speckle image restoration problems very well. We also simulated the traditional speckle image restoration method, and performed numerical experiments using the direct inversion method and the field strength method as examples. We found

that although the traditional methods perform well on the whole, it is not as good in detail processing as Neural network method. The traditional methods still carry some speckles in the restored image, while the neural network method performs very well in the details. And it can also use the same network to restore speckle images of different media, which is difficult to achieve with traditional methods. In summary, the neural network method performs very well in the problem of speckle image restoration, and has great application potential.

Key Words: Speckle image retrieve; computer vision; deep learning; convolutional neural network

第一章 绪论

一、激光散斑图像恢复问题

散斑图像处理问题是一个典型的光学图像处理问题，它在许多光学系统中都有重要的应用^[1-6]。在激光器实用化之后，激光器的使用者们发现，当激光照射光学粗糙的表面，比如纸面，墙面等，或者当激光透过折射率不均匀的介质，比如海水，生物组织等的时候，它会形成一种颗粒状的斑纹图案，这种图案被称为散斑图。散斑图像广泛出现在各种光学问题中。

散斑图像的物理成因是光学粗糙的表面或者散射介质将光路散射为许多随机的路径，这些随机的光束经过叠加可以形成强度分布图案，这种图案就是散斑图^[1-2]。散斑形成有两个要素：(i) 光学粗糙的散射表面或者光学粗糙的(折射率不均匀的)透射介质的性质；(ii) 照明光场的相干性质。激光因其具有很好的单色性、和相干性，且其亮度很高，方向性很好，所以它在散斑成像问题被广泛地进行了研究^[7]。除了对它的散斑图像的物理机理进行深入研究之外，散斑图像在光学系统中也有很多重要的应用，比如将散斑作为一种阴影掩膜可以用来测量金属的疲劳损伤，也可以将散斑图像用来测量物体表面的粗糙程度等等。

在这篇工作中，我们主要关心的问题是散斑图像恢复的问题。通常在光学系统中，为了使得我们的成像系统散斑减少，我们可以对波前进行调制，如果选用合理的波前空间调制，那我们成像系统中所获的图像的散斑就可以被很好的抑制。我们尝试利用神经网络来处理这种问题，因为卷积神经网络可以很好地用来处理各种图像恢复问题^[8-9]，所以我们很自然地可以去考虑将神经网络用在散斑图像恢复的问题里面^[10-11]。

我们详细分析了各种常被用于图像处理的卷积神经网络架构，最后我们采用的是带有密集块的 Unet 结构，我们发现这种结构在散斑图像恢复问题中的表现非常好。我们利用透射矩阵方法模拟生成了训练散斑图像对神经网络进行了训练，然后对训练好的神经网络进行了测试。我们还将神经网络方法和传统的恢复方法做了比较。

二、论文结构安排

在第二章，我们详细介绍了散斑图像形成的物理原理以及散斑图像恢复的一些传统方法。利用随机行走模型，我们给出了散斑图像的光强圆 Gauss 分布函

数。随后我们详细介绍了散斑图像的透射矩阵描述。通过将空间光场离散化，我们可以构造一个针对特定散射介质的入射光场和出射光场之间的矩阵描述，这个矩阵就是透射矩阵。我们详细介绍了实验中测量透射矩阵的方法，包括四步相移法，三步相移法和五步相移法。透射矩阵可以用来恢复散斑图像，我们详细介绍了集中基于透射矩阵的散斑图像恢复方法：直接求逆法；光学相位共轭法；以及 Tikhnov 正则算子法。最后我们还介绍了一种基于场强的恢复方法。

在第三章中我们将主要介绍基于神经网络的图像处理方法。我们首先对机器学习的常见工具进行了总结和讨论，机器学习依照其学习方式的不同主要分为三类：监督学习，非监督学习和强化学习。它常用的模型有神经网络，支持向量机等等，其中神经网络广泛地被用在图像处理问题里面。随后我们对神经网络做了详细的讨论。我们首先介绍了前馈神经网络以及卷积神经网络的基本概念，它里面所涉及的基本操作。然后我们针对图像处理问题介绍了经常用来在图像恢复问题的集中卷积神经网络，包括 VGG 网络，Res-net 网络，DenseNet 网络和 Unet 网络。这为我们后面利用神经网络做散斑图像恢复问题做了铺垫和准备工作。

在第四章中，我们讨论了利用神经网络方法数值模拟散斑图像恢复的过程，这也是本文的核心部分。我们首先利用透射矩阵随机地生成了散斑图像恢复的训练数据集。我们使用的 MNIST 数据集，利用随机透射矩阵我们可以生成每张手写数字图片的散斑图像然后对其进行标记。其次我们详细介绍了我们所使用的神经网络架构，我们的架构是 DenseNet 和 Unet 的结合形式。相当于在一个 Unet 里面引入 Dense block。接着我们对我们的数值模拟结果做了细致的分析。为了比较它和传统恢复方法的差别，我们也对传统的基于光强的散斑图像恢复方法做了数值模拟和讨论。

在最后一部分我们对我们的结果进行了总结，对我们方法的可能推广和应用简要地进行了讨论，对未来可能的研究方向也进行了一些展望。

第二章 传统的散斑图像恢复方法

散斑现象最早是在 1960 年代被观测到的, 当时连续波激光器刚刚商业化, 使用这些激光器的研究者们发现, 当激光被纸张、墙壁等表面平均起伏大于光波长量级的光学粗糙表面反射或者透过光学粗糙的介质时, 它会形成一个具有很高对比度的精细颗粒或者斑纹图案。这种图案被称为散斑图像 (speckle image)^[1-2]。这种图案的成因是无规则分布的光学粗糙界面或者无规则的介质的透射面元所散射的子光波进行叠加从而使得光场具有随机的空间光强分布, 从而显示出精细的颗粒状或者斑纹图案。在这一节我们将详细分析散斑图像的物理成因, 它的应用以及如何通过一些物理手段或者算法来复原散斑图像。

第一节 散斑图像的物理成因及其分类

一、散斑的物理成因

我们首先来介绍散斑图像的物理成因^[1-2]。散斑形成有两个要素: (i) 光学粗糙的散射表面或者光学粗糙的 (折射率不均匀的) 透射介质的性质; (ii) 照明光场的相干性质。激光因其具有很好的单色性、和相干性, 且其亮度很高, 方向性很好, 所以它在散斑成像问题被广泛地进行了研究^[7]。

从物理角度看, 经过一个不规则的光学粗糙表面散射或者折射率不均匀的介质散射之后, 经过出射面的光是一些随机光束, 散斑图像是这些无规则光束的相干叠加后所形成的强度分布斑纹。我们考虑单色光

$$u(x, y, z, t) = A(x, y, z)e^{i\theta(x, y, z)}e^{i\omega t} = \tilde{A}(x, y, z)e^{i\omega t}, \quad (2.1)$$

其中 ω 是光的圆频率, $A(x, y, z)$ 是振幅, $\theta(x, y, z)$ 是相位, 而 $\tilde{A}(x, y, z) = A(x, y, z)e^{i\theta(x, y, z)}$ 是复振幅。光场强度为

$$I(x, y, z) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} |u(x, y, z, t)|^2 dt = |\tilde{A}(x, y, z)|^2. \quad (2.2)$$

一个散斑图观测点的场是由许多度量的光场叠加给出的。我们来看一个关于不规则散射过程的最简单的模型, 随机相矢叠加 (random phasor sum) 模型^[1]。假设我们有 N 个独立的随机光场, $\tilde{a}_k = |\tilde{a}_k|e^{i\theta_k}$, 它们满足如下的几条假设:

- 不同波的振幅和相位之间统计独立，也就是说对于 $k \neq l$, $a_k = |\tilde{a}_k|$ 和 θ_k 与 $a_l = |\tilde{a}_l|$ 和 θ_l 统计独立；
- 同一个波的振幅和相位之间相互独立，即就是说 $a_k = |\tilde{a}_k|$ 和 θ_k 统计独立；
- 相位 θ_k 在 $(-\pi, \pi)$ 之间取均匀分布。

实际上这些场构成一个由空间位置 $k = 1, \dots, N$ 标记的随机过程。那么这些光场在经过叠加之后我们有

$$\tilde{A}(x, y, z) = \frac{1}{\sqrt{N}} \sum_{k=1}^N \tilde{a}_k = \frac{1}{\sqrt{N}} \sum_{k=1}^N |\tilde{a}_k| e^{i\theta_k}. \quad (2.3)$$

在复平面内看的话，这相当于是 N 个独立的向量 $\frac{1}{\sqrt{N}} \tilde{a}_k$ 进行相加，这相当于一个随机行走问题，参看图2.1 (c)。

我们分开来考虑叠加场 $\tilde{A}(x, y, z)$ 的实部 $\mathcal{R} = \text{Re}(\tilde{A})$ 和虚部 $\mathcal{I} = \text{Im}(\tilde{A})$ 。利用前面的关于各个叠加场的独立性假设以及相位分布的均匀分布假设，我们会有实部和虚部的期望

$$\mathbf{E}(\mathcal{R}) = 0, \quad \mathbf{E}(\mathcal{I}) = 0. \quad (2.4)$$

类似地，可以计算它们的方差

$$\sigma^2 = \sigma_{\mathcal{R}}^2 = \sigma_{\mathcal{I}}^2 = \frac{1}{2N} \sum_{k=1}^N \mathbf{E}(a_k). \quad (2.5)$$

然后利用中心极限定理，当 N 很大的时候，我们知道实部 $\mathcal{R} = \text{Re}(\tilde{A})$ 和虚部 $\mathcal{I} = \text{Im}(\tilde{A})$ 的分布趋向于正态分布，即

$$p(\mathcal{R}, \mathcal{I}) = \frac{1}{2\pi\sigma^2} e^{-\frac{\mathcal{R}^2 + \mathcal{I}^2}{2\sigma^2}}. \quad (2.6)$$

注意到 $I^2 = A^2 = \mathcal{R}^2 + \mathcal{I}^2$ 和 $\theta = \arctan \frac{\mathcal{I}}{\mathcal{R}}$ 。通过对概率分布进行变量替换，我们可以得到关于光强和相位的概率分布

$$p(I, \theta) = \begin{cases} \frac{1}{4\pi\sigma^2} e^{-\frac{I}{2\sigma^2}}, & I \geq 0, -\pi \leq \theta < \pi, \\ 0, & \text{其他情况.} \end{cases} \quad (2.7)$$

通过求边缘分布，我们可以分别求得光强和相位的分布，光强分布为

$$p(I) = \int_{-\pi}^{\pi} p(I, \theta) d\theta = \frac{1}{2\sigma^2} e^{-\frac{I}{2\sigma^2}}, \quad I \geq 0, \quad (2.8)$$

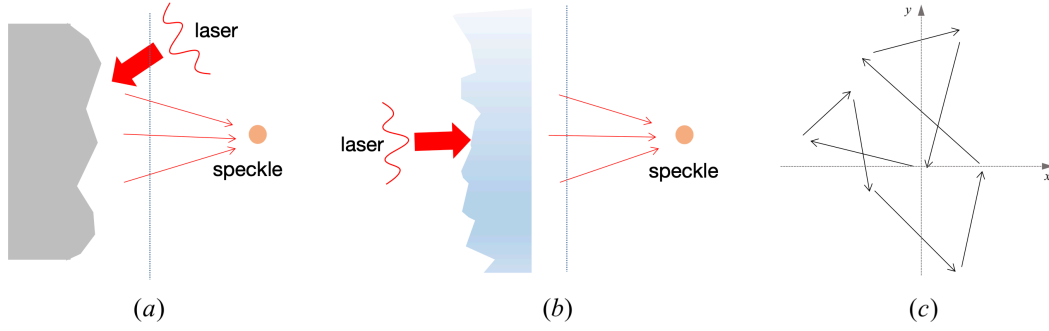


图 2.1 散斑图像原理示意图。(a) 光学粗糙的表面对激光进行颜色，每个独立的小面元相当于一个衍射单元，最后这些衍射的光场进行叠加导致散斑图像的形成，注意这里我们用虚线标出了出射平面；(b) 当激光经过折射率不均匀的介质时，大量无规则路径的光进行叠加产生散斑，这里我们也用虚线标出了出射平面；(c) 随机相矢叠加的示意图。

相位分布仍然为均匀分布。我们也可以去求关于振幅 $A = |\tilde{A}| = \sqrt{I}$ 的分布，利用变量替换方法，我们很容易求得

$$p(A) = \frac{A}{\sigma^2} e^{-\frac{A^2}{2\sigma^2}}, A \geq 0. \quad (2.9)$$

它满足 Rayleigh 分布。

上面的光强分布结果最早是由 Goodman 给出的^[12]，它所对应的散斑图像现在普遍被称为正态散斑或者高斯散斑，这个结果也被许多实验所验证^[13]。

对于一个光学粗糙的表面，我们可以将其分割成一个个小的面元，这些面元相当于是一个个的小衍射单元，而整个面就可以被看作是由许多衍射单元构成的一个类似光栅的结构。大量不规则的面元所衍射的光场经过叠加发生干涉，因为面元数量很多而且极度不规则，所以最后干涉产生的强度也是剧烈变化的，这种剧烈的强度变化就形成了无规则分布的颗粒状斑纹，也就是散斑图像。如图 2.1 (a) 所示。

对于一个光学散射介质，比如海水，生物组织，当光穿过的时候，由于它的折射度极度不均匀，最后的出射光非常随机，这些随机的光束经过相干叠加之后会产生剧烈变化的强度分布，也会形成散斑图。如图 2.1 (b) 所示。

二、散斑的分类

我们可以首先是依据散斑图案随着时间的变化关系分为将其分为静态散斑和动态散斑，在这篇文章中，我们将主要讨论静态散斑。但我们所发展的基于神经网络的散斑图像恢复方法实际上也可以应用于动态散斑图像问题。

此外我们也可以按照光场的传播方式来分类，因为粗糙表面的散射面元或者粗糙介质的投射面元都可以看成一个衍射单元，那么我们可以根据衍射的分类来对散斑进行分类。首先我们将其分为有成像系统的像面散斑和无成像系统的散斑，而无成像系统的散斑又可以进一步依据光源，衍射单元以及接收屏之间的远近分为：

- 对应于菲涅尔衍射的近场散斑；
- 对应于夫琅和费衍射的远场散斑。

这种分类对于散斑的研究和分析非常有用，因为这种分类本身携带了许多物理信息。

在研究中最常见最中的散斑分类是依据散斑的成因来进行分类的，它依赖于两个要素，一个是散射表面或者散射介质的性质，另一个是照明光场的性质。综合考量这两个要素可以对散斑进行分类，比如不同相干光对应不同散斑类型，不同的透射介质或者散射面对应不同的散斑类型。在本文中，我们也将主要采用这种分类方法对散斑图像进行研究。

第二节 散斑图像的透射矩阵描述

在折射率不均匀的介质的散斑图像问题的研究中，最重要的研究方法是透射矩阵 (transmission matrix) 方法，在这一节我们详细讨论这种方法。

一、透射矩阵的定义

通常对于给定的散射介质，因为其内部的光线传播自由程很短而且有多重散射效应的存在，出射光场相比于入射光场会显得极度无序，呈现出散斑结构。虽然经过散射介质之后，有序的稳定相干入射光变成了混乱的随机出射光，但这种散斑结构出射光并不是完全随机的，它内部也包含了大量信息。因为固定散射介质，相同的入射光总能给出相同的出射散斑图像，这意味着散射介质的入射光和出射光之间存在一定的关联，这种关联通常被称为散斑关联 (speckle correlation)^[3-6]。常用来研究散斑关联的方法是散斑关联散射矩阵 (speckle-correlation scattering matrix) 方法^[4-5]，而透射矩阵作为散斑关联矩阵的一种特殊情形被广泛用来刻画散斑关联。

我们知道，介质中的 Maxwell 方程的形式为

$$\nabla^2 A(\vec{r}, t) - \frac{n(\vec{r})^2}{c^2} \frac{\partial^2 A(\vec{r}, t)}{\partial t^2} = 0. \quad (2.10)$$

其中 $n(\vec{r})$ 是折射率, $A(\vec{r}, t)$ 表示 t 时刻空间位置 \vec{r} 处的场, c 是真空中光速。我们可以通过 Fourier 变换将其变为频域方程, 通常称为 Helmholtz 方程:

$$\nabla^2 A(\vec{r}, \omega) - k^2 \frac{\partial^2 A(\vec{r}, \omega)}{\partial t^2} = 0. \quad (2.11)$$

其中 $k = n(\vec{r})\omega/c$ 是波矢的模, ω 是光波的圆频率。可以用格林函数方法求解 Helmholtz 方程, 我们可以得到结果, 当空间点 \vec{r}' 处有一入射源时, 出射光场是这些源光场的带权重叠加

$$A(\vec{r}, \omega) = \int G(\vec{r}, \vec{r}'; \omega) A_s(\vec{r}'; \omega) d\vec{r}'^3. \quad (2.12)$$

上式中的 $G(\vec{r}, \vec{r}'; \omega)$ 就是格林函数, $A_s(\vec{r}', \omega)$ 是入射源场。虽然原理上来说, 如果我们知道散射介质的细节, 我们也就知道了对应的格林函数, 格林函数刻画了散斑关联。通过格林函数, 我们可以用上述的方法来求解出射光场。

但通常由于介质内部的折射率分布高度随机, 所以解析地去求解格林函数就变的极度困难。所以我们可以考虑近似去求格林函数的方法, 其中一种常用的技术就是将入射平面划分为一些小方格, 将出射平面也划分成一些小方格, 通过对各个方格内对应的场进行平均, 然后我们可以分析入射的有限个方格的场和出射的有限个方格的场之间的关联, 如果假设介质具有的关联具有线性性, 那么这种关联可以通过一个矩阵来描述, 这个矩阵就是透射矩阵。实际上, 从物理层面上来讲, 这种做法的合理性是基于散斑尺度的, 因为对波长为 λ 的光波, 可以证明其散斑尺度大约为 $(\lambda/4)^2$ 量级^[1], 这保证了上面的离散化方法的合理性, 我们可以将像元尺寸选得与散斑尺寸相对应进而满足我们的需求。

详言之, 对于入射平面, 我们将其划分为 $N = L \times L$ 个格子。我们将第 k 个格子对应的空间区域记为 S_k , 并将其中心区域对应的空间矢量记为 \vec{r}_k^{in} , 于是我们可以将这个格子内部的场先进行叠加从而得到

$$A_k^{in} = A_k^{in}(\vec{r}_k^{in}) = \int_{S_k^{in}} d\vec{r} A^{in}(\vec{r}). \quad (2.13)$$

对出射平面, 我们可以做类似的格子化处理, 将其划分为 $M = K \times K$ 个格子, 对于第 l 个格子 S_l^{out} , 我们将其中心所对应的空间矢量记为 \vec{r}_l^{out} , 我们可以将格子内部的场进行叠加

$$A_l^{out} = A_l^{out}(\vec{r}_l^{out}) = \int_{S_l^{out}} d\vec{r} A^{out}(\vec{r}). \quad (2.14)$$

据此，利用介质的线性假设，我们可以有一个 $M \times N$ 的矩阵 G_{lk} ，使得

$$A_l^{out} = \sum_{k=1}^N T_{lk} A_k^{in}. \quad (2.15)$$

这里的 T_{lk} 就称为是透射矩阵，它的矩阵元是一些复数。

实际上，依据场是在坐标空间表述还是在波矢空间表述，我们会有对应的坐标空间透射矩阵，类似地，在动量空间，我们会有动量空间对应的透射矩阵。在实际应用中，通常需要依据问题的需要选择合适的表述方法。

如果我们得知了透射矩阵，那么对于给定的输入光，我们就能得到输出光的信息，所以透射矩阵在研究散斑图像的时候是非常有用的工具。利用我们上一节关于不规则散射的随机相矢叠加模型的讨论我们知道，一个给定的入射面元 S_k^{in} 所对应的入射场在介质中经过随机散射之后得到的出射面元 S_l^{out} 处的散射所贡献的分布是一个高斯型分布。也就是说，透射矩阵的矩阵元实际上是服从高斯分布的。但要注意的是，高斯透射矩阵模型是一种白噪声模型，对于实际的散射介质，透射矩阵是比较复杂的，通常要通过实验来测量透射矩阵。

二、透射矩阵的实验测量

因为刻画散斑关联的透射矩阵有坐标空间和动量空间两种描述，所以它的测量也有坐标空间透射矩阵的测量^[14-15]和动量空间透射矩阵的测量^[16]两种。目前的技术水平能实现介观尺度下的传输矩阵测量。

我们这里来介绍一下实验测量坐标空间透射矩阵的测量，它常用的方法是四步相移干涉法 (four phases method)^[14-15]。测量过程的示意图如图2.3所示^[15]，一束激光经过透镜和偏振片之后，再经过一个空间光调制器 (spatial light modulator) 调制，其次再经过一个偏振片之后射入散射介质进行散射，出射光经过一个偏振片之后再由一个 CCD 相机成像。

考虑 N 个输入场 $A_1^{in}, \dots, A_N^{in}$ ，它们对应的第 m 个输出场的形式为

$$\begin{aligned} I_m^\alpha &= |A_m^{out}|^2 = \left| s_m + \sum_n e^{i\alpha} T_{mn} A_n^{in} \right|^2 \\ &= |s_m|^2 + \left| \sum_n e^{i\alpha} T_{mn} A_n^{in} \right|^2 + 2 \operatorname{Re} \left(e^{i\alpha} \bar{s}_m \sum_n T_{mn} A_n^{in} \right). \end{aligned} \quad (2.16)$$

其中 α 是相对相位偏移系数， s_m 第 m 个输出光场的参考光的复振幅。我们可以

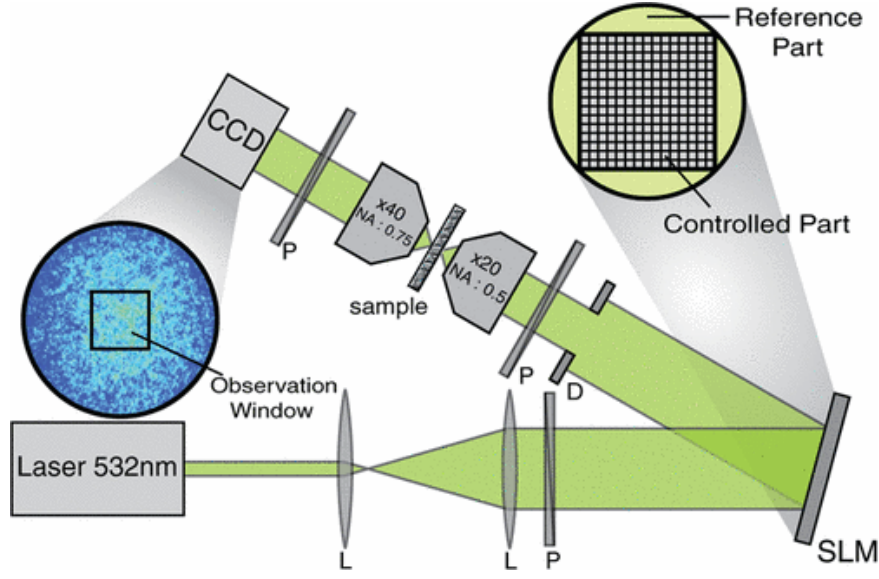


图 2.2 透射矩阵测量原理示意图^[15]。一束激光经过透镜和偏振片之后，再经过一个空间光调制器调制，其次再经过一个偏振片之后射入散射介质进行散射，出射光经过一个偏振片之后再由一个 CCD 相机成像。

取 $\alpha = 0, \pi/2, \pi, 3\pi/2$ 四个值，于是我们会得到

$$\begin{aligned}
 I_m^0 &= |s_m|^2 + \left| \sum_{n=1}^N T_{mn} A_n^{in} \right|^2 + 2 \operatorname{Re} \left(\bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right), \\
 I_m^{\pi/2} &= |s_m|^2 + \left| \sum_{n=1}^N T_{mn} A_n^{in} \right|^2 + 2 \operatorname{Re} \left(i \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right), \\
 I_m^{\pi} &= |s_m|^2 + \left| \sum_{n=1}^N T_{mn} A_n^{in} \right|^2 - 2 \operatorname{Re} \left(\bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right), \\
 I_m^{3\pi/2} &= |s_m|^2 + \left| \sum_{n=1}^N T_{mn} A_n^{in} \right|^2 - 2 \operatorname{Re} \left(i \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right).
 \end{aligned} \tag{2.17}$$

观察上面这些式子我们可以发现，它们与透射矩阵之间有非常密切的关联，具体来说，我们有

$$\begin{aligned}
 & \frac{I_m^0 - I_m^{\pi}}{4} + i \frac{I_m^{3\pi/2} - I_m^{\pi/2}}{4} \\
 &= \frac{2 \operatorname{Re} \left(2 \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right)}{4} + i \frac{2 \operatorname{Re} \left(-i 2 \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right)}{4} \\
 &= \operatorname{Re} \left(\bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right) + i \operatorname{Im} \left(\bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in} \right) \\
 &= \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in}.
 \end{aligned} \tag{2.18}$$

通过这种方法我们可以得到对应的透射矩阵，我们引入参考光矩阵 S_{ref} ，它是个对角矩阵，对角元为 s_m ，于是我们的观测透射矩阵形如

$$T_{obs} = S_{ref} T. \quad (2.19)$$

要得到透射矩阵，我们只需要左乘 S_{ref} 的逆矩阵。

上面讨论四步相移干涉法，实际上根据 α 取值的差异，也有三步相移干涉法和五步相移干涉法，他们分别对应的强度和透射矩阵的关系为：

$$\frac{2I_m^{4\pi/3} - (I_m^0 + I_m^{2\pi})}{3} + i \frac{\sqrt{3}(I_m^0 - I_m^{2\pi})}{3} = \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in}, \quad (2.20)$$

$$\frac{2I_m^\pi - I_m^{2\pi} - I_m^0}{5} + i \frac{2(I_m^{\pi/2} - I_m^{3\pi/2})}{5} = \bar{s}_m \sum_{n=1}^N T_{mn} A_n^{in}. \quad (2.21)$$

通常步数多精度高，但计算效率会下降，所以在实际应用中常采用精度和计算效率相折衷之后多四步相移干涉法。

透射矩阵测量过程中，核心的部件是空间光调制器。也就是说，我们需要对输入的光场进行空间调制，也就是说对输入光场的空间和振幅进行调制^[17]。传统的取 0, 1 的 Descartes 基的空间光调制器通常不能用来做透射矩阵测量。这是因为，一方面，逐个像素地去调制光场效率很低，而且会导致测量噪声比较大；另一方面，单像素地调控会使得光强变弱，经过散射介质之后光场会变得更弱，几乎无法测量。通常大家采用 Hadamard 基来进行调制，各个像素点取值为 ± 1 ，每个像素点都有光通过，且它们的振幅相等，而相位相差为 π 。1 对应的相位为 e^{0i} ，而 -1 对应的相位 $e^{i\pi}$ 。所以采用 Hadamard 基进行调制可以克服上面提到的一些问题。

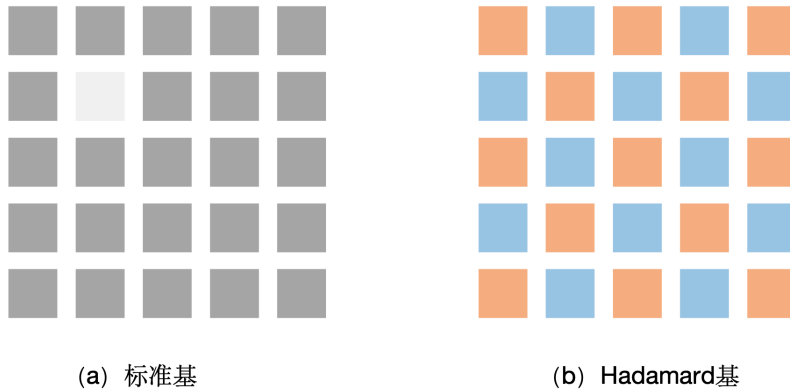


图 2.3 空间光调制的基选择。(a) 标准基，其中白色方块代表通过光，灰色方块代表不通过；(b) Hadamard 基，其中橙色方块代表 1 而蓝色方块代表 -1 ，它们都有光透过，并且相位差为 π 。

第三节 传统散斑图像恢复方法

散斑图像的恢复是计算成像领域的一个典型问题^[18,11]，许多传统的计算成像方法都可以用来处理散斑图像恢复问题。在本节我们详细讨论散斑图像的恢复问题以及传统的一些用于散斑图像恢复的方法。

一、散斑图像恢复的基本原理

散斑图像恢复问题考虑的是如下的问题，对于一个给定的散射介质，我们通过照明光对一个物体进行照明，然后让光透过一个散射介质，最后再通过一个相机记录得到的图像，这是一个散斑图像，我们的目标是对这个散斑图像进行恢复，得到原图像的信息。如图2.4所示。

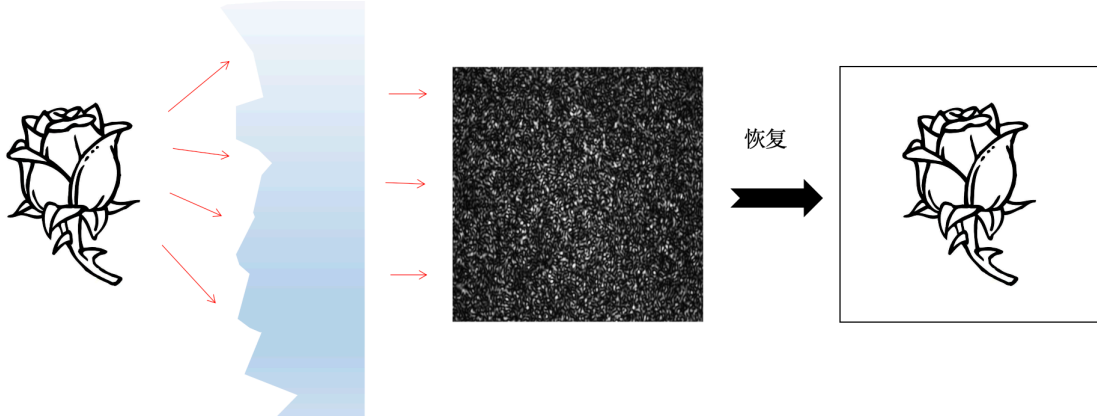


图 2.4 散斑图像恢复问题的示意图。

利用我们前面介绍的投射矩阵的框架，我们知道对于包含目标信息的输入光阵列 $A^{in} = (A_1^{in}, A_2^{in}, \dots, A_N^{in})$ ，经过散射介质的散射之后我们有出射光阵列 $A^{out} = (A_1^{out}, A_2^{out}, \dots, A_M^{out})$ ，利用理想的透射矩阵我们有

$$A^{out} = T A^{in}. \quad (2.22)$$

但在实际的实验中，会有两种误差，首先，透射介质的透射矩阵实验测量会有误差 T_e ，也就是说，观测值为 $T_{obs} = T + T_e$ ；其次在成像记录过程中也会引入误差 A_e ，于是最后实验获得的输出为

$$A_{obs}^{out} = T A^{in} + A_e. \quad (2.23)$$

我们是在知道 A_{obs}^{out} 以及实验测得的投射矩阵 T_{obs} 的情况下想去推断 A^{in} 。

实际上，这种问题是一种典型的病态逆问题。具体来说，我们的输入为 \vec{x} ，

经过透射矩阵 T 的作用之后在，在加上成像误差 \vec{e} ，最终的输出结果为 \vec{y} ，

$$\vec{y} = T\vec{x} + \vec{e}. \quad (2.24)$$

我们知道 \vec{y} ，目标是求出对应的 \vec{x} 。有许多经典的逆问题求解方法可以用来处理这类问题。

二、直接求逆方法

散斑图像恢复的最简单的思路是直接求逆。我们通过实验测得的投射矩阵 T_{obs} ，而对整个过程取逆。如果输入和输出阵列的尺寸一样，都为 N ，那么透射矩阵是一个方阵，我们直接取其矩阵逆 T_{obs}^{-1} 。如果输入和输出阵列的尺寸不一样，输入为 N ，而输出为 M ，那么透射矩阵是一个 $M \times N$ 矩阵，我们可以取其 Moore-Penrose 逆，

$$T_{obs}^{-1} = (T_{obs}^\dagger T_{obs})^{-1} T_{obs}^\dagger. \quad (2.25)$$

如果忽略成像误差，那么我们会有

$$\hat{A} = T_{obs}^{-1} A_{obs}^{out} = (T + T_e)^{-1} (T A^{in} + A_e). \quad (2.26)$$

这样的话，我们会得到散斑图像 A_{obs}^{out} 的一个复原图像 \hat{A} 。

直接求逆方法虽然思想上十分简单明了，它在散斑聚焦和成像问题中有许多应用，在噪声比较小的情况下，它的表现很好。但它对噪声十分敏感，随着噪声的增大，这种方法会快速失效^[19]。

从物理的角度来看，因为取逆过程是对投射矩阵的奇异值取倒数的过程，它会将 T_{obs} 中比较大的者奇异值变小，这样的话，原来图像过程中比较主要的信息通道在取逆以后就变成了比较次要的通道，而原来的次要通道反而变成了主要通道，这相当于放大的噪声而抑制了核心信息信号，所以在实际应用中这种方法常常有很大缺陷。

三、光学相位共轭方法

为了克服直接求逆方法的缺陷，我们可以尝试对透射矩阵的比较大的奇异值做特殊处理，这里我们介绍一种在这种思路下发展起来的方法，光学相位共轭 (optical phase conjugation) 方法^[20]。在这种方法里面我们为了恢复图像，直接对实验获得的散斑图像阵列 A_{obs}^{out} 作用实验测得的透射矩阵 T_{obs} 的 Hermitian 共轭矩阵 T_{obs}^\dagger ，即

$$\hat{A} = T_{obs}^\dagger A_{obs}^{out} = T_{obs}^\dagger (T A^{in} + A_e). \quad (2.27)$$

注意这里我们实际上也是直接忽略了成像误差的。

光学相位共轭方法是一种典型的时间反演方法。由透射矩阵的统计性质我们知道理想的透射矩阵满足 $T^\dagger T \simeq \chi I$ ，其中 χ 是一个归一化因子，这也是我们可以采用光学相位共轭的方法来恢复散斑图像的原因。

四、Tikhnov 正则化算法

前面介绍的两种方法对噪声项 A_e 都没有进行细致处理，也就是说，在噪声较小的情况下，它们才能取得比较好的结果。而要一般性地去处理这个问题，我们必须回过头去看原始的病态逆问题。一种常用的求解病态逆问题的方法是 Tikhnov 正则化算法 (Tikhnov Regularization Algorithm)^[21]，它也可以被用来处理散斑图像恢复的问题。

具体来说，我们要找一个矩阵 W ，恢复图像的过程可以表述为这个矩阵作用在观测到的图像阵列上的结果，即

$$\hat{A} = W A_{obs}^{out}. \quad (2.28)$$

为此，我们首先假设我们有一定的先验信息，它可以表述为一个正则矩阵 Q ，他所携带的信息表述为

$$\|Q(\hat{A} - A^{in})\|^2 = \|Q(W A_{obs}^{out} - A^{in})\|^2. \quad (2.29)$$

我们还要引入一个量化噪声的量，

$$\mu = \langle \|\hat{A} - A^{in}\|^2 \rangle = \langle \|W A_{obs}^{out} - A^{in}\|^2 \rangle. \quad (2.30)$$

那么我们的问题就转化为一个求条件极值的问题，也就是说，我们的恢复图像应该是如下的条件极值

$$\operatorname{argmin}_{\hat{A}} \{ \|T_{obs} \hat{A} - E_{obs}^{out}\|^2 + \mu \|Q(\hat{A} - A^{in})\|^2 \}. \quad (2.31)$$

所以 Tikhnov 正则化算法的核心思想是结合先验信息来求上面的极小值。但实际应用中， A^{in} 的信息往往无法预先得知，所以，可以采用平滑约束，将上面的条件极值问题变为

$$\operatorname{argmin}_{\hat{A}} \{ \|T_{obs} \hat{A} - E_{obs}^{out}\|^2 + \mu \|\hat{A}\|^2 \}. \quad (2.32)$$

这便是我们所关心的问题的核心表述。

我们现在来求解 $\|T_{obs}\hat{A} - E_{obs}^{out}\|^2 + \mu\|\hat{A}\|^2$ 的极小值, 首先将上式展开, 我们会有

$$\hat{A}^\dagger T_{obs}^\dagger T_{obs} \hat{A} - \hat{A}^\dagger T_{obs}^\dagger E_{obs}^{out} - (E_{obs}^{out})^\dagger T_{obs} \hat{A} + (E_{obs}^{out})^\dagger E_{obs}^{out} + \hat{A}^\dagger (\mu I) \hat{A}. \quad (2.33)$$

如果我们限制上式对与 \hat{A} 的偏导数为零, 那么我们会得到

$$\hat{A} = (T_{obs}^\dagger T_{obs} + \mu I)^{-1} T_{obs}^\dagger E_{obs}^{out}. \quad (2.34)$$

由此, 我们可以将图像恢复的矩阵写为

$$W = (T_{obs}^\dagger T_{obs} + \mu I)^{-1} T_{obs}^\dagger. \quad (2.35)$$

注意到 T_{obs} 的 Moore-Penrose 逆为 $(T_{obs}^\dagger T_{obs})^{-1} T_{obs}^\dagger$, 所以这里的 W 相当于直接求逆方法加入了噪声项。而噪声哪一项实际上可以看成是光学相位共轭方法所采用的复原矩阵项。那么这里的方法实际上就是将直接求逆方法和光学相位共轭方法结合在一起的一种方案。当噪声贡献 $\mu = 0$ 时, 它退化成直接求逆方法, 当 $\mu \gg 0$ 时, 它退化为光学相位共轭方法。

我们也可以利用奇异值分解对 W 做分析。我们首先对观测到的透射矩阵进行奇异值分解

$$T_{obs} = \sum_i s_i |u_i\rangle\langle v_i|, \quad (2.36)$$

其中 s_i 是 T 的奇异值, 而 u_i 和 v_i 分别是对应于 s_i 的左右奇异向量。这里为了方便我们采用了 Dirac 记号。在给定的基底下我们有

$$T_{obs} = \sum_i U^\dagger \Lambda V, \quad (2.37)$$

其中 Λ 是奇异值组成的对角阵并且我们约定按照奇异值从大到小的顺序排列。将其带入式 (2.35) 中我们会有

$$\begin{aligned} W &= \left[T_{obs}^\dagger \times T_{obs} + \mu \times I \right]^{-1} T_{obs}^\dagger \\ &= \left[(U^\dagger \Lambda V)^\dagger \times U^\dagger \Lambda V + \mu \times V^\dagger I V \right]^{-1} (U^\dagger \Lambda V)^\dagger \\ &= \left[V^\dagger (\Lambda^\dagger \Lambda + \mu I) V \right]^{-1} V^\dagger \Lambda^\dagger U \\ &= V^\dagger (\Lambda^\dagger \Lambda + \mu I)^{-1} V V^\dagger \Lambda^\dagger U \\ &= V^\dagger (\Lambda^\dagger \Lambda + \mu I)^{-1} \Lambda^\dagger U. \end{aligned} \quad (2.38)$$

注意到 $\Lambda^\dagger \Lambda + \mu I$ 是一个对角矩阵，它的对角元素为 $|s_i|^2 + \mu$ 。于是我们可以得到 W 矩阵的奇异值分解的对角阵， $\Sigma = (\Lambda^\dagger \Lambda + \mu I)^{-1} \Lambda^\dagger$ ，它的对角元素为

$$\frac{s_i}{s_i^2 + \mu}. \quad (2.39)$$

由此可见，噪声项 μ 的作用是将原来观测透射矩阵的比较大的奇异值放大，而对原来比较大的奇异值影响不大。这意味着 W 的作用是对原本 T_{obs} 中那些奇异值比较小的通道执行了近似于光学相位共轭方法的复原过程，而对原来 T_{obs} 中那些奇异值比较大的通道执行了直接求逆方法的复原过程。

五、基于场强的方法

前面几种方法都是基于光场的恢复方法，也就是说输入和输出都是相矢。这里我们来介绍一种基于场强的恢复方法^[4]。

假设输入光场的尺寸为 N ，而输出光场的尺寸为 M ，我们对输入光场取标准基 $a_i = (0, \dots, 0, 1, 0, \dots, 0)^T, i = 1, \dots, N$ ，那么一个输出光场可以表示为

$$A^{in} = \sum_{i=1}^N \alpha_i a_i. \quad (2.40)$$

类似地，对输出光场选择标准基 $b_j = (0, \dots, 0, 1, 0, \dots, 0)^T, j = 1, \dots, M$ 。如果我们将透射矩阵的每一列记为 t_p ，则 $T = (t_1, \dots, t_N)$ 那我们会有

$$T a_i = t_i. \quad (2.41)$$

如果入射光场是 $A^{in} = \sum_{i=1}^N \alpha_i a_i$ ，那么输出光场就是 $A^{out} = \sum_{i=1}^N \alpha_i t_i$ 。

这种方法的要点在于构造一个散斑散射关联矩阵，具体来说，这个矩阵如下

$$Z_{pq} = \frac{1}{\sum_p \sum_q} [\langle t_p^* t_q (E^{out})^* E^{out} \rangle_r - \langle t_p^* t_q \rangle_r \langle (E^{out})^* E^{out} \rangle_r], \quad (2.42)$$

其中 $\langle \cdot \rangle_r$ 是空间求平均，而 $\sum_p = \langle |t_p|^2 \rangle_r$ 。利用四阶矩和二阶矩之间的关系，我们可以将其写为

$$Z_{pq} = \alpha_p \alpha_q^* + \frac{1}{\sum_p \sum_q} \langle t_p^* (E^{out})^* \rangle_r \langle t_q E^{out} \rangle_r. \quad (2.43)$$

当 $\gamma = M/N$ 变得很大的时候，上面的结果可以进一步简化为

$$Z_{pq} = \alpha_p \alpha_q^*. \quad (2.44)$$

由此可以看出此时的 Z 矩阵的秩为 1，它的特征向量就是输入光场场强分布，所以我们只要求 Z 的本征值最大的本征向量即可。

第四节 本章总结

在本章中我们详细介绍了散斑图像相关的问题。我们首先对散斑图像的成像原理以及分类问题进行了详细的讨论，我们说明了散斑图像的分布是复 Gauss 分布。随后我们对散斑图像的透射矩阵描述进行了详细的分析和讨论，我们介绍了它的定义，以及它的实验测量。最后我们详细讨论了散斑图像恢复的问题，这个问题可以被归结为一个病态逆问题。我们介绍了传统的恢复散斑图像的几种方法，包括直接求逆法，光学相位共轭法，Tikhnov 正则化算法以及基于光强的方法。

第三章 神经网络图像处理

近些年,随着计算机算力的提高以及人工智能算法的飞速发展,机器学习已经开始在各个科学领域以及工业技术中有了很多的应用。在工业以及现实应用中的例子主要有,语音识别^[22-23],手写数字识别^[24-25],围棋等游戏的 AI 操作^[26],自动驾驶^[27],药物合成过程优化^[28],语言翻译^[29-30]等等。其在科学研究领域的应用更是无处不在,以物理学来说^[31],量子态的断层成像技术的研究和分析^[32],相变过程的研究和分析^[33-34],求解复杂物理系同的 Schrödinger 方程^[35],在量子纠错码领域的应用^[36],对开放系统的研究和分析^[37],在热力学中的应用^[38]等等。机器学习的应用和发展对科学研究和科技进步起到了很大的推动作用,反过来,这些应用也极大地刺激了对机器学习的进一步发展和更加深入的理解。

在这一章,我们将详细讨论机器学习的方法在图像处理问题中的应用,我们尤其会仔细讨论卷积神经网络以及它的各种改进版本。我们会仔细讨论他的各种技术细节以及理论基础,为我们在下一章讨论他们在光学图像处理领域中的应用进行准备。

第一节 机器学习

现在,整个机器学习领域已经变得非常丰富多彩,其内容很多,但其中有许多重要的算法和工具已经成为了教材中的标准内容。详言之,机器学习的主要算法按照其学习形式可以分为如下几类^[39-40]: 监督学习 (supervised learning), 非监督学习 (unsupervised learning) 以及强化学习 (reinforcement learning)^①。为了实现具体的机器学习过程,常常需要构建具体的学习模型,如神经网络 (neural network) 模型,支持向量机 (support vector machine) 模型,决策树 (decision trees) 模型等等。这些模型通常被用来做各种机器学习过程,按照其学习方式来分类,我们可以将这些模型大致进行如下的分类:

^①实际上还更多的类型,比如半监督学习 (semi-supervised learning), 自主学习 (self learning), 特征学习 (feature learning), 稀疏字典学习 (sparse dictionary learning), 关联规则学习 (association rule learning) 等等。但广为传播的主要是上面三种。

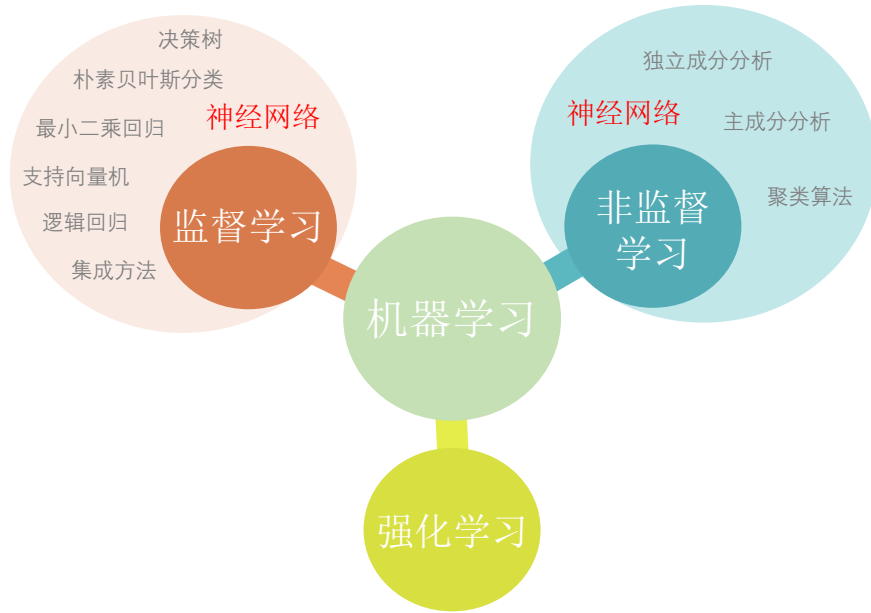


图 3.1 机器学习的一些常见模型的示意图。机器学习按照其学习形式主要分为监督学习，非监督学习和强化学习。这里我们给出了他们对应的一些常用的学习模型。

监督学习

监督学习这是目前理论上被研究得最清楚透彻的一种学习方式。在监督学习中，我们有一个训练数据集 $D_{train} = \{(x_i, y_i)\}$ ，其中 x_i 为数据点， y_i 是数据点的标记，也就是说，我们有一个待学习的目标函数 $f: X \rightarrow Y$ ， $f(x_i) = y_i$ 。监督学习是通过大量的训练数据来学习这个目标函数。最后我们可以用一组测试数据集 D_{test} 来测试我们所学得的函数的准确度。

常用来做监督学习的模型主要有分类 (classification) 模型和回归 (regression) 模型两种，具体来说，我们有下面的一些常用的例子：

- 神经网络 (nerual network) 模型。这也是本论文中会详细讨论的模型，它由大量神经元按照给定的拓扑结构相互连接而构成。它可以用来做分类分析和回归分析。它有很多类型，比如前馈神经网络 (feed-forward neural network)，Boltzmann 机 (Boltzmann machine)，卷积神经网络 (convolutional neural network) 等等。
- 决策树 (decision trees) 模型。连续取值的决策树也常被称为回归树 (regression tree)。它们常被用在数据挖掘里面，因为这种模型的结构比较简洁，所以它在机器学里面的应用非常广泛。它可以用来做分类和回归分析，常见的决策树算法有：分类树 (classification tree) 及回归树 (regression tree)，随机森林 (random forest)，ID3 算法 (Iterative Dichotomiser 3 algorithm)，C4.5 算

法 (C4.5 algorithm), Chi 方自动交互检测 (Chi-squared Automatic Interaction Detection, CHAID), 决策树桩 (decision stump) 等等。

- 朴素贝叶斯分类 (naive Bayesian classification), 它是概率性分类的一种, 主要是依据贝叶斯定理并且预先假定特征之间具有很强的独立性。
- 支持向量机 (support vector machine), 它是一种被广泛使用的监督学习模型, 常被用来做分类和回归分析。它通过对数据添加额外维度然后构造超平面的方法对数据集进行分类。
- 回归算法。比如最小二乘回归 (ordinary least squares regression)。这是很典型的一种线性回归的优化方法, 它通过最小化误差的平方和可以获得数据的最佳函数逼近, 在监督学习中用的很多。其他的回归方法还有逻辑回归 (logistic regression), 逐步式回归 (stepwise regression), 多元自适应回归样条 (multivariate adaptive regression splines) 等。
- 集成学习 (ensemble learning), 在这种方法里面, 将多个模型组成一个系统, 根据训练系统选择合适的模型去适应。

上面我们仅给出了部分的例子, 而且很多上面的模型都可以用到别的形式的学习中去, 比如神经网络也可以用来做非监督学习和加强学习。在本工作中, 我们也是在监督学习的框架下, 利用神经网络来处理图像问题。

非监督学习

与监督学习不同, 非监督学习的数据集没有标签, 它是通过自主学习从而构建数据内部的图像。常用来做非监督学习的有如下一些算法和模型:

- 聚类算法 (clustering algorithms), 它是对一系列给定对象进行训练分组, 使得那些特征相似的对象处于相同的集群中。最终对于一个给定的对象, 我们可以用训练好的结果的相似程度来判断它所属的类别。
- 主成分分析 (principal component analysis), 这是一种统计方法, 通过线性变换将原来的数据进行处理, 提取出主要成分。
- 独立成分分析 (independent component analysis), 它是将数据进行处理, 将数据变成一些独立分量。

神经网络也常常被用来做非监督学习, 通过自主学习, 神经网络可以学习数据内部的结构。

强化学习

另一种比较常用的学习方式是强化学习。在强化学习的框架下，一个智能体 (agent) 通过环境 (environment) 给他的反馈来学习。不同于前面两种学习方式，强化学习不需要预先提供数据，它有一个状态集合以及操作集合，通过环境对他的反馈去选择相应的状态以及相应的操作，据此获取学习信息并且调整模型参数。

常见的强化学习的模型是马尔可夫决策过程 (Markov decision processes)，它由下面的一些要素构成：

1. 一个有限状态的集合 S ，其中用 $s \in S$ 来标记智能体的一种状态；
2. 一个有限操作的集合 A ，其中用 $a \in A$ 来标记智能体采取的操作；
3. 转移概率 $p(s'|s, a)$ ，它表征了智能体从状态 s 和操作 a 中转移到状态 s' 的概率；
4. 奖励 $R(s, a)$ ，他表示环境从智能体的状态选择 s 和 a 中获得的奖励；
5. 从状态里面选取操作的规则 $\pi : S \times A \rightarrow [0, 1]$ ，具体来说，是用用条件概率 $\pi(a, s) = p(a|s)$ 刻画；

智能体每一步中利用上面这些要素进行学习，通过回报 (return) 函数 $V_\pi(s_0, s_1, \dots, s_t)$ 来衡量这个状态序列的最优性，进而完成学习，调整参数。

在本工作中，我们将主要讨论神经网络，尤其是卷积神经网络以及它的各种改进版本在图像处理问题中的应用，下面让我们详细讨论基于神经网络的图像处理。

第二节 卷积神经网络

为了介绍卷积神经网络及其各种改进版本，我们首先简要回顾一下神经网络的基础知识。神经网络是受生物神经网络的启发所提出的一种数学模型，它最早的形式是感知机 (perceptron)^[41]。随着感知机在许多问题上取得成功，人们进一步研究和发展了各种别的神经网络，比如典型的有下面两种：

- 前馈神经网络 (feed-forward neural network)，它包括各种感知机的变形，误差逆传播 (error backpropagation) 神经网络，径向基函数 (radial basis function) 神经网络，以及我们将着重讨论的卷积神经网络等。
- 统计回流神经网络 (stochastic recurrent neural network)，它包括 Hopfield 神经网络 (Hopfield neural network)，各种 Boltzmann 机 (Boltzmann machine) 的变形等等。

因为我们主要讨论的是卷积神经网络，所以，我们着重讨论一下前馈神经网络。

一、前馈神经网络

前馈神经网络的主要构成单元是神经元，它是生物神经元的数学抽象，我们知道，生物神经元之间相互连接，它们通过释放神经递质来向与之相连的神经元传递信息。神经元有高电位和低电位状态，分别代表它的激活和抑制状态。观察到这些信息，研究者们抽象出了神经元的数学结构。它是一个具有数个输入端以及一个输出端的单元，其自身有一个偏置来表征激活的阈值，进入神经元的输入经过带权重的叠加之后与偏置进行差值，然后用一个激活函数作用其上来表征其激活状态。

从数学角度来说，我们将神经元的 n 个输入记为 x_1, \dots, x_n ，将其相应的权重记为 w_1, \dots, w_n ，这些输入首先进行带权重的叠加 $\sum_i x_i w_i$ ，然后输入神经元，与神经元的偏置 b 进行比较，也就是说，做差值， $\sum_i x_i w_i - b$ 。最后再作用一个激活函数 f 上去得到输入结果 $y = f(\sum_i x_i w_i - b)$ 。

较为常用的单变量激活函数有以下一些：

- Heaviside 阶梯函数

$$H(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad (3.1)$$

它是最早被用在感知机里面的激活函数，分别用 0, 1 代表神经元的抑制和激活状态。

- 逻辑函数，也称为 Sigmoid 函数，

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (3.2)$$

它可以看成阶梯函数的连续性推广。

- 双曲正切函数

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (3.3)$$

- Relu (Rectified Linear Unit) 激活函数

$$\text{ReLU}(x) = \max\{0, x\}. \quad (3.4)$$

它在图像处理问题中用得很多。

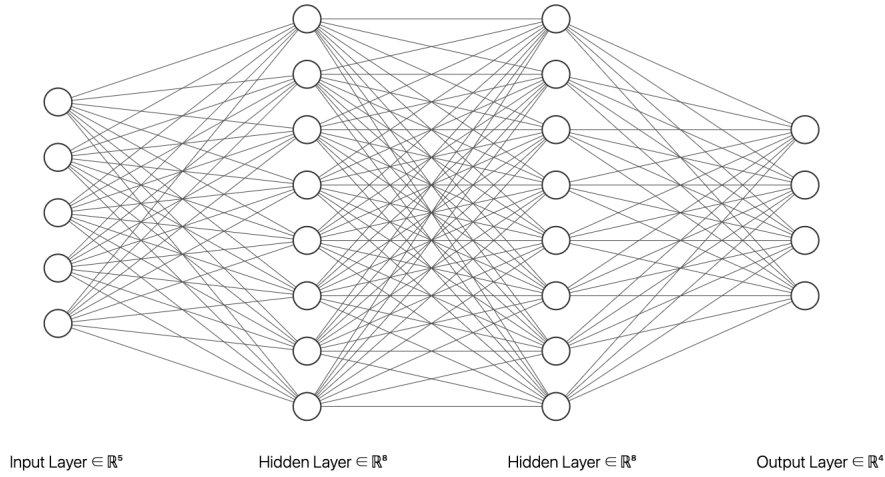


图 3.2 一个四层的全连接神经网络的示意图。

- 软加 (softplus) 函数

$$\text{SP}(x) = \ln(e^x + 1). \quad (3.5)$$

- ELU (Exponential Linear Unit) 函数

$$\text{ELU}(x) = \begin{cases} x, & x \geq 0, \\ \alpha(e^x - 1), & x < 0. \end{cases} \quad (3.6)$$

在神经网络的应用中，也常会用到一些多个自变量的激活函数，主要是在输出层进行处理的时候。常用的输入为 $\mathbf{x} = (x_1, \dots, x_n)$ 的多变量激活函数有

- 极大输出函数 $\max(\mathbf{x}) = \max\{x_1, \dots, x_n\}$.
- 软极大 (softmax) 函数, $\text{SM}(\mathbf{x})_j = \frac{e^{x_j}}{\sum_i e^{x_i}}$.

将大量神经元按照层状结构置于一一起，相邻两层之间的神经元相互连接，而层内神经元互相之间没有连接，据此我们可以构造出一个神经网络，这种类型的神经网络我们称为前馈神经网络。它的第一层为输入层，最后一层为输出层。中间的通常称为隐藏层。在图3.2中，我们给出了一个全连接神经网络的例子。

前馈神经网络在手写数字识别等分类问题中取得了巨大的成功，随后它的应用开始被拓展到各个领域，尤其是它在图像处理领域的变形和改进——卷积神经元，在现代图像处理算法中具有重要的地位。

下面我们详细来讨论卷积神经网络以及它在图像处理方面的应用。

二、卷积神经网络

不同于传统的全连接前馈神经原理，卷积神经网络利用了全新的思想和构造方法来构建神经网络，它主要用到了两个结构：卷积 (convolution) 和池化 (pool-

ing)。其基本结构是一个卷积层后面接一个池化层，然后将这种结构进行堆叠。

虽然卷积神经网络看上去和传统全连接神经网络差异很大，但我们可以证明，他实际上和全连接神经网络密切相关，我们可以证明卷积过程实际上是等价与一个神经元全连接结构的。

为了详细讨论卷积神经网络，我们先来看一下卷积和池化操作的一些细节。因为本论文主要讨论的是图像处理，我们假设我们的输入为一张图片，它有 $N \times M$ 个像素点。每个像素点上是一个具体的整数值。如果该图片是一张灰度图，那么图片只有一层，如果是一张 RGB 模式的彩色图片，那它有三层，分别代表 R 层，G 层以及 B 层。通常一个图片本身的层数也被称为通道 (channel) 数。下面的讨论为了简单，我们先假定输入的图片是一张灰度图，也就是说，只有一个通道。

1. 卷积

我们首先来看如何进行卷积操作，这是卷积神经网络的核心步骤。卷积过程是利用一个给定尺寸 $n \times m$ 的卷积核 (convolutional kernel) 来实现的，卷积核也常被称为过滤器 (filter)。卷积核在水平方向和竖直方向分别以一定的步长 l_1 和 l_2 扫完整张图片，获得图片在这个小区域的特征，进而获得整个图片的特征。

卷积核是一个 $n \times m$ 的矩阵，以 3×3 的卷积核为例，我们可以将其记为

$$K = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \quad (3.7)$$

我们假设它扫过的第 i 横步 j 纵步区域内的像素块的取值为

$$[X]_{ij} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}, \quad (3.8)$$

那么卷积之后我们所获的的相应的值为

$$Y_{ij} = \sum_{i,j=1}^3 a_{ij}x_{ij}. \quad (3.9)$$

我们将这个值取为输出图片第 i 行 j 列的像素值。值的指出的是，这里的卷积与数学中所使用的卷积略有差异，因为这里的卷积满足结合率。为了和让它满足结合率，可以先将扫描出来的像素块沿着次对角线翻转，然后在进行对应元素乘积

并且相加。但在实际应用中，因为前面所定义的方式已经能取得很好的效果了，所以一般没有翻转这个操作。

完成扫描之后我们将得到一个新的图片，新图片的尺寸是依赖于卷积核的尺寸以及扫描的步长的。我们看到，卷积操作会缩小图片尺寸，为了满足我们对图片尺寸大小的需求，大家提出了一些新的卷积方式。人们提出先对图片进行填充 (padding)，然后再进行卷积。填充过程通常是在图片边缘添加一些空白像素点，使得卷积之后的图片尺寸满足我们的要求。

在实际应用中，为了应对不同问题所给出的挑战，人们也提出了许多别的类型的卷积，比较常用的有以下几种：比较常见的有

- 扩张卷积 (dilated convolution)。扩张卷积是对原图像进行扩张然后再进行卷积，详言之，就是对原卷积核等间隔地添加空白行和列以扩大感受野的一种卷积方式。
- 形变卷积 (defomable convolution)。它是在原来标准卷积地基础上，加入了一些形变之后地卷积，这些卷积核的形状不再是标准的块状矩阵。
- 深度分离卷积 (depthwise separable convolution)。它是针对多通道图片的卷积，将图片的每个通道进行独立卷积。
- 空间分离卷积 (spatially separable convolution)。它是将一个 $n \times m$ 的卷积过程拆解成 $n \times 1$ 和 $1 \times n$ 两个中间步骤的卷积过程，它可以降低计算量。
- 平化卷积 (flattened convolution)。它是类似空间可分离卷积的，对于多通道图片，对于一个三维的卷积核在三个维度都进行拆分从而降低计算量。

除了上面提到的这些卷积之外，还有一种重要的卷积过程，我们将在本论文中使用，他就是反卷积 (deconvolution)，也常被称为转置卷积 (transpose convolution)。反卷积在图像复原，语义分割等问题里面是非常常用的一种卷积过程。

反卷积本质上不是卷积操作的逆操作，而是卷积操作的转置操作。具体来说，我们可以将一个 $N \times M$ 像素的图片 X 通过一个给定的卷积核 K 变成一个 $N' \times M'$ 像素的图片 Y ，通过将像素点拉直的方法，我们会发现，实际上卷积操作可以被写成一个矩阵变换

$$\vec{Y} = C_K \vec{X}. \quad (3.10)$$

其中 \vec{X} 和 \vec{Y} 是输入和输出图片拉直后的向量，而 C_K 是卷积操作所对应的矩阵。泛卷积操作时间上就是

$$\vec{X}' = C_K^T \vec{Y}'. \quad (3.11)$$

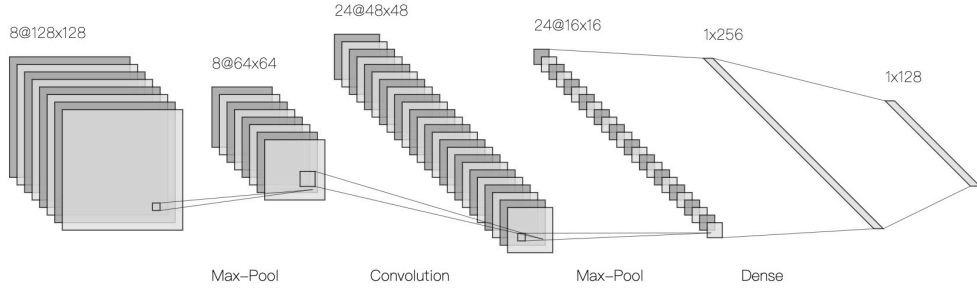


图 3.3 一个简单的卷积神经网络结构示意图，其中包含了一个池化层，紧接着是一个卷积层，然后再接一个池化层，最后以一个密集层，它的作用是将张量拉成向量。

这里我们用 C_K^T 表示 C_K 的转置操作。进行完这个变换之后，我们再将向量变为图片，于是我们就得到了一个人输入尺寸为 $N' \times M'$ ，输出尺寸为 $N \times M$ 的卷积操作，他称为前面所对应的卷积操作的反卷积操作。

2. 池化

卷积神经网络的另一个重要组成部分是池化 (pooling) 层，池化也常常被称为下采样 (subsampling)。池化层的主要作用是：对数据进行降维，降低信息冗余；提升模型的在大尺度上的不变性；另外还有一个很重要的作用就是防止过拟合。

直观上来讲，池化过程是提取局部的核心信息，而扔掉冗余的信息，这个过程可以保持图片的核心信息。常用的池化方式是最大池化 (max pooling) 和平均值池化 (average pooling)。具体来说，池化过程是按照一定的尺寸对图片进行分割，以一个 2×2 的小区域为例，它的像素值构成一个 2×2 的矩阵，

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (3.12)$$

最大池化就是对这几个像素值取极大值 $\max\{a_{11}, a_{12}, a_{21}, a_{22}\}$ ，而平均池化就是取平均值 $(a_{11} + a_{12} + a_{21} + a_{22})/4$ 。现代的神经网络理论中发展起来的池化方法有许多中，具体来说有^[42]：

- 最大池化，也即是说取最大值的池化方法；
- 平均值池化，也就是取平均值的池化方法；
- 混合池化，也即是混合了最大池化与平均值池化的池化方法；
- L_p -平均池化，取了类似 L_p 模的池化方法；
- 随机池化，按照像素值取值构造概率分布的池化方法。

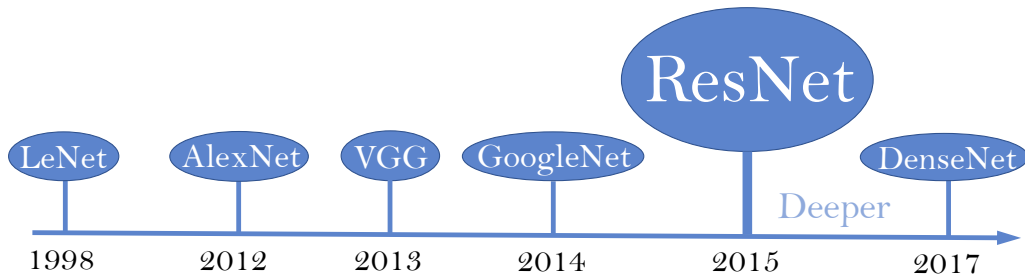
池化过程会缩小图片的尺寸，在图像处理过程中，也常常要用到反池化的过程，反池化也常常称为上采样 (upsampling)。与反卷积类似，反池化过程会将图片放大。以最大池化为例，反池化的过程是将池化过程中取极大值的那个像素点

的位置记住，然后将那个像素点对应的局部区域的像素点均设为与之相同的值。

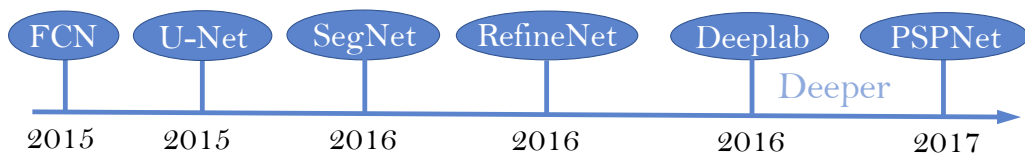
一个典型的卷积神经网络是由重复的卷积和池化层构成的，如图3.3所示。在经过多个卷积层和池化层结构之后，通常由一个平化层来将最后所得到的张量拉直成向量。也可以由一个全连接层最后的结果进行处理。

第三节 卷积神经网络在图像处理方向的改进版本

卷积神经网络属于深度神经网络模型的一种，其发展有几十年的历史^[43]：1959年，Hubel和Wiesel在对猫的视觉皮层细胞的研究中提出了感受野的概念^[44]，而这个概念在后来被用于在未能揭示卷积神经网络数学原理之前粗略解释神经网络的工作机制；1979年，日本科学家Fukushima在之前的感受野的基础上提出了神经认知机模型^[45]，这是第一个卷积神经网络模型；1989年，Lecun



(a) 分类任务



(b) 分割任务

图 3.4 各种神经网络的改进版本：(a) 完成分类任务的卷积神经网络的进化史。其中 ResNet 是一种重要的改进，它解决了深度卷积神经网络梯度回传消失的问题，使得神经网络的层数能够加深很多；而 DenseNet 在此基础上进行了改进，引入了短程连接和稠密模块 (Dense block) 的概念，这也是本文所使用的网络的核心结构；(b) 完成语义分割任务的卷积神经网络进化史。其中 U-Net 是最重要的一种，它首次提出了编码-解码 (Encode-Decode) 模式并就此成为经典，完全取代了 FCN。此后的分割网络都以 U-Net 为基线 (baseline) 进行改进，而本文中也是使用 U-Net 结构进行的图像修复。

等人首次提出了过滤器 filter 的权值共享技术^[46]，而在 1998 年，他们首次将卷积层和下采样层相结合，设计了现代卷积网络最主要的结构；进一步地，加拿大的学者 Xavier Glorot 和 Yoshua Bengio 在 2011 年提出了一种新的激励函数 $\text{ReLU}(x) = \max\{0, x\}$ ^[47]；在 2012 年的 ImageNet 竞赛中，冠军获得者 Hinton 和他的学生 Krizhevsky 设计的 Alexnet 就采用了 ReLu 作为激活函数^[48]。从此以后 ImageNet 竞赛上相继出现了许多的神经网络使得分类任务精度迅速上升，如图 3.4 (a) 所示。而同时完成分割任务的卷积神经网络也在飞速发展，如图 3.4 (b) 所示。

一、VGG 神经网络

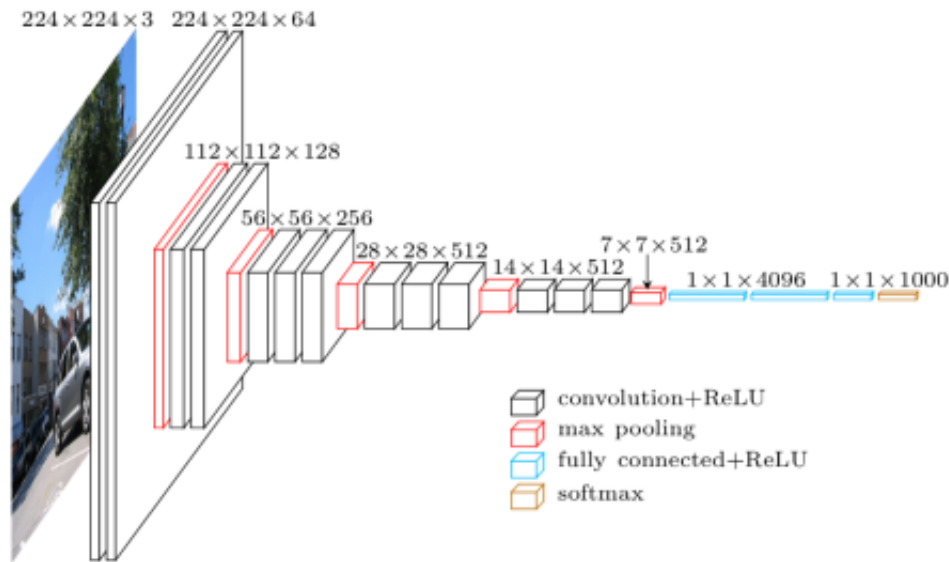


图 3.5 VGG^[49] 是在 AlexNet 上进行改进的，采用连续的几个 3×3 的卷积核代替原来 AlexNet 中的较大的 11×11 、 7×7 、 5×5 的卷积核，增加网络深度并减少参数个数来学习完成更复杂的任务。其中权重数目最多的地方就在第一个全连接层中，有许多公开的预训练权重包含的就是这些全连接层的权重。

VGG16^[50] 的名字来源于参加 ILSVRC 的团队名称，在他们的原作中，作者仅仅将该结构称为 ConvNets，即为 Convolutional networks 的简称。VGG 网络是在 AlexNet 网络上改进的用来进行图像分类的一种网络，其结构如图 3.5 所示。RGB 三通道的图像进来后，先经过几层步长为 1 的 3×3 的卷积。有时候也会使用 1×1 的卷积，因为可以看成一种对于输入通道的线性变换。然后经过五个步长为 2 的 2×2 的最大下采样层。卷积层之后是三个全连接层，前两个全连接层各自有 4096 通道，第三个全连接层有 1000 个通道。最后一层是使用 soft-max 做

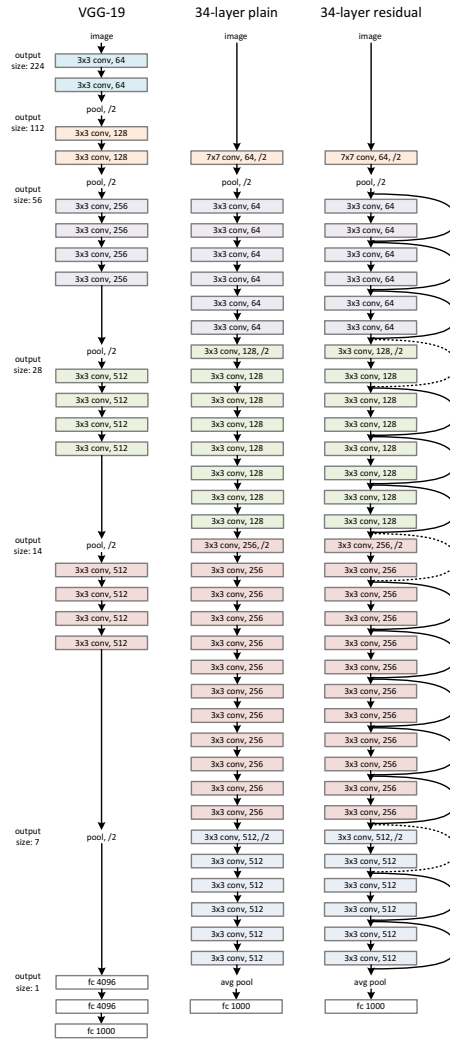


图 3.6 从图中可以看出 Kaiming He 发明 ResNet 的自然思路。从左到右分别是早先的 VGGNet，在其基础上改进的 plain baselines 以及在 plainnet 上改进的 Residual Network。

激活函数而非 ReLU 做激活函数的一层。当然，其具体的网络设置有许多，在卷积层层数会出现明显差别，而 VGG16 的名字来源就是卷积层为 16 的该类结构的网络。

VGGNet 的结构清晰明了，整张网络的卷积核和池化核尺寸统一，使用更多更小的卷积核提高了网络的性能，是早期的少见的深度网络。虽然在训练上 VGGNet 的权重很多，但是有公开的预训练权重可以使用^[51]，大大减轻了使用负担。

二、ResNet 神经网络

ResNet 的提出同样在 2015 年，是为了完成对 ImageNet 数据库的图像分类任务，作者设计 ResNet 的思路如图 3.6 所示。先是在 VGGNet 的启发下设计了平

常基线 (plain baselines) 网络。基线的结构主要由 3×3 的卷积层配合以下两个规则设计：(1) 如果输出特征图的分辨率大小相同的话，这样的神经层之间有同样数量的过滤器。(2) 如果输出特征图的分辨率大小被缩小了一半，那么过滤器的数量就要是原来的一倍，这样才能保持每一层的时间复杂度一致。下采样过程直接通过步长为 2 的卷积层完成。网络的最后是一个全局平均池化层和一个以 softmax 为激活函数的全连接层。平常基线网络总的层数为 34 层。按照上述规则设计的平常基线网络的权重文件的大小仅仅为 VGG19 的 18%。ResNet 的作者在平常基线网络的设计基础上，插入了短程连接，创造性地构造了残差网络 (Residual Network)。残差网络的核心结构单元有两种：(1) 当输入输出的维度相同时，使用的短程连接，其输入输出规则如式 $y = F(x, \{W_i\}) + x$ 所示，其位置如图 3.6 最右边的残差网络的实线连接模块所示；(2) 当输出的维度增加时，如图 3.6 最右边的残差网络中的虚线位置。对于这种情况 ResNet 的作者使用了两种处理方式：第一种是选择和实线位置相同的输入输出规则，而对于缺失的像素通过主动添加的方式来改变输出的维度，使其满足前述输入输出的规则所需要的维度数目，并且参数数目同前一种不会出现变化；第二种处理方式是通过 1×1 的卷积改变输入的维数，进而满足新的输入输出规则： $y = F(x, \{W_i\}) + W_s x$ 。

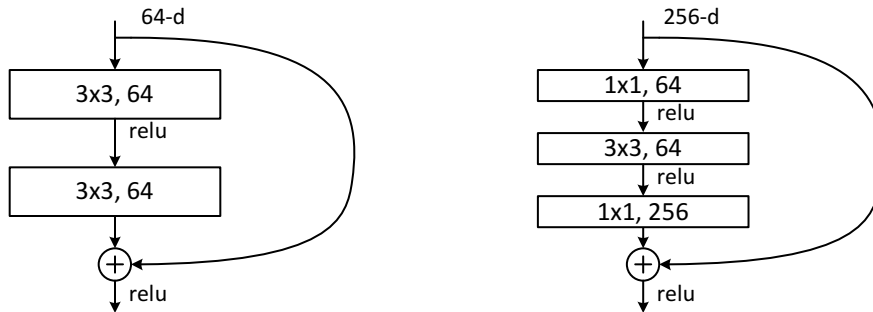


图 3.7 图左是在 ResNet-34 中的基本构成要素。图右是在 ResNet-50/101/152 中的“瓶颈”基本构成要素。

ResNet 的基本概念就是残差，而残差传播的基本结构单元就如图 3.7 左所示，在 ResNet-34 中用的就是这种结构，而在更深层的 ResNet 中，作者对其进行了如图 3.7 右所示的改进，使用三个分别为 1×1 、 3×3 、 1×1 的卷积层替代了原先的两个 3×3 的卷积层。上述的第二种结构被作者命名为更深的瓶颈结构 (Deeper Bottleneck Architecture)，该结构在 DenseNet 中也被作者所采用，同样被命名为瓶颈 (bottleneck) 层。瓶颈层的核心操作就是 1×1 的卷积操作，以改变输入数据的通道数目。

三、Unet 神经网络

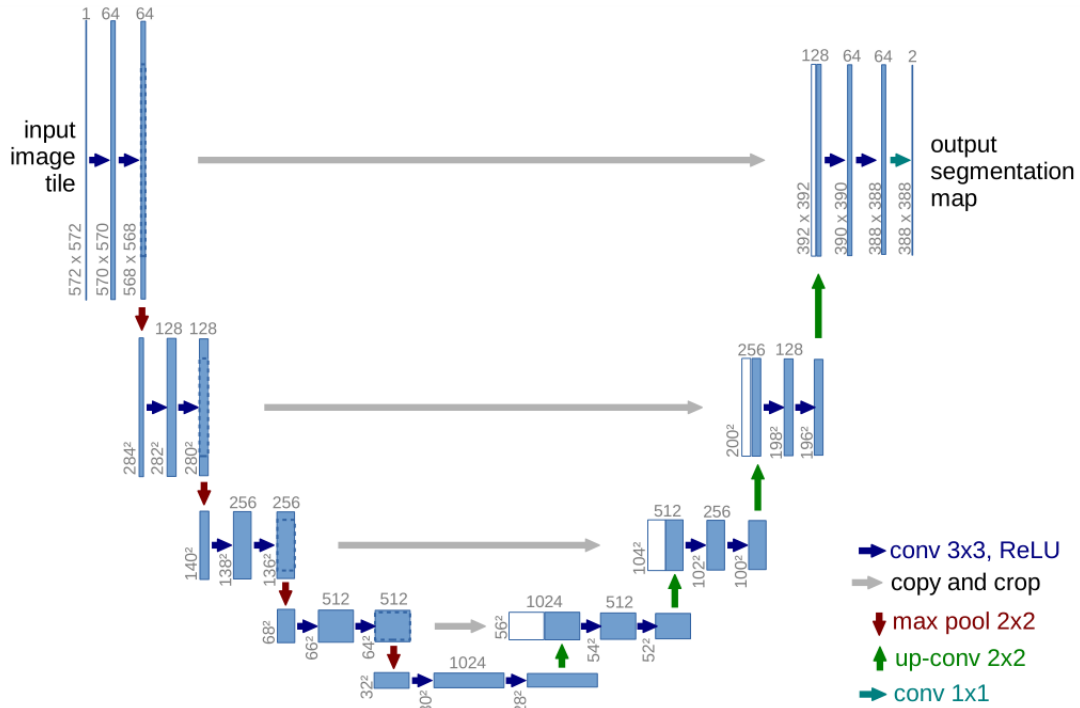


图 3.8 Unet^[8] 的网络架构，蓝色方框代表一个多通道特征图，通道数在盒子顶部，输入尺寸在盒子底部，白色方框代表复制拼接的特征图，蓝色的箭头代表做 3×3 卷积以后再进行线性整流，灰色箭头代表复制收缩路径每一步的特征层后直接拼接至膨胀路径每一步对应的输入中 (左边的虚线框就代表复制操作)，红色箭头代表了 2×2 的最大下采样，绿色箭头代表了 2×2 的反卷积操作，浅蓝色的箭头代表了 1×1 的卷积操作，该网络所能容许的最低分辨率为 32×32 。

Unet^[8] 最初的提出是为了解决生物医学图像分割问题。图像分割问题就是将图片中不同个体用不同颜色来标记区分。分割问题与卷积网络在此前用来处理的图像分类问题有本质的差别：图像分类问题一次只需要区分其中一个物体，也只有一个分类标签；而图像分割问题则需要一次区分其中多个物体，有多个分类标签，要对目标图片中的每个像素进行分类。

Unet 的结构如图 3.8 所示，包括收缩路径和与其逆过程的膨胀路径，其中左半部分为收缩路径，右半部分为膨胀路径，因其网络结构像英文字母 U，故称其为 Unet。收缩路径是一种典型的卷积网络架构，包括重复应用两个无填充的 3×3 卷积，每个卷积后面的 ReLU 线性整流单元以及最大池化操作，每一步的降采样都将特征通道的数量翻倍；在膨胀路径中每步都包含上采样和反卷积操作，并且会将每上一层所得到的特征图与之前对应层所训练得到的特征图层拼

接在一块，在最后一层完成和之前每一层一样的操作后，还使用 1×1 的卷积把每个 64 分量特征向量映射到所需的类数。前图3.8的网络中总共有 23 个卷积层。Unet 的使用中需要注意的是每一层的输出都会是偶数，所以需要输入可以写为 $2^n, n \in \mathbb{N}$ 。详细观察可以发现，Unet 网络的整体结构其实就是常见的深度神经网络加上一个拼接 (concatenate) 操作。普通的深度神经网络就是卷积层的堆叠，卷积层的具体操作常见为两次卷积和一次池化。Unet 的作者在此基础上提出了创新，将前面一部分卷积层提取出的特征图层拼接到后面的特征图层中作为后面卷积层的输入。这种拼接的操作使得神经网络在训练的时候，早先提取出的特征也能对后面的网络参数产生影响，使得提取出的特征能够达到最大的利用率。

Unet 作为 15 年发表的论文，网络结构是在早期用来解决同样问题的 FCN^[52] 网络结构上进行的改良，其收缩膨胀结构的卷积层之间提取出的特征层会进行拼接，使得网络能够融合不同层级之间的特征。Unet 网络能够很好地完成图像分割任务，直到现在还是完成该任务的主要基线。

四、DenseNet 神经网络

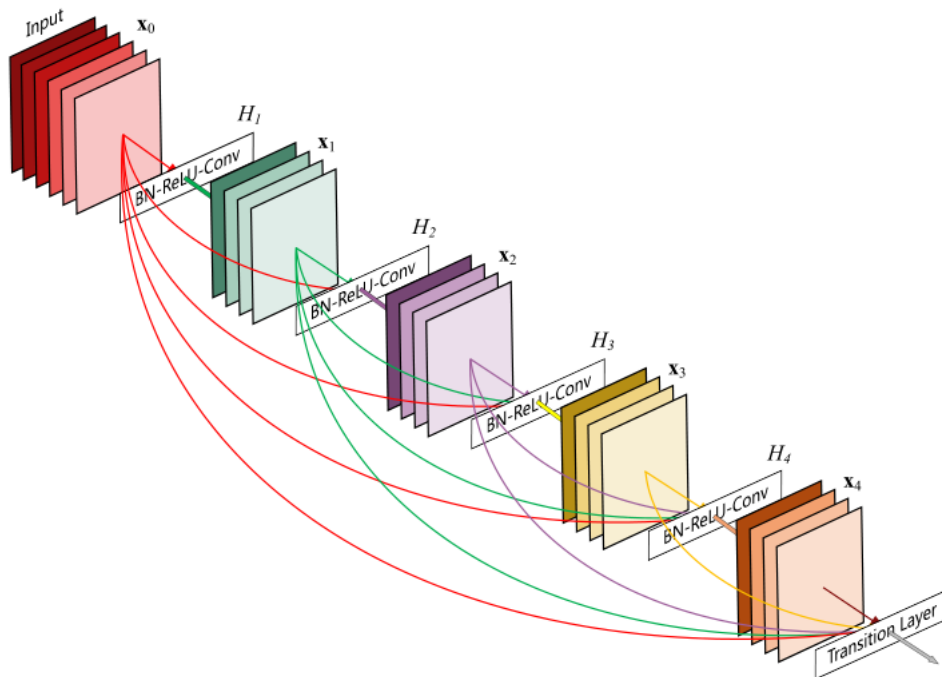


图 3.9 五层的密集块，其增长率为 4，每层的输入都是前面所有层的输出^[9]。

DenseNet^[9] 在 17 年提出，是完成分类任务的一个重要模型。其设计借鉴了 ResNet 中将之前的特征层直接传到后面的思想，但是在结构设计上进行了创新。

DenseNet 的网络结构由反复堆叠的密集块和传递层 (Transition Layer) 组成，

其详细的结构如表3.1所示, 其中密集块是 DenseNet 的设计核心。密集块的结构如图3.9所示, 是多个瓶颈的叠加。图3.9中每个黑色框都是一个瓶颈操作框, 其完整的操作顺序是 BN-ReLU-Conv(1×1)-BN-ReLU-Conv(3×3), 其中 Conv(1×1) 是为了对输入的特征层进行降维, 减少参数; 不同颜色的矩形框集是每一层瓶颈的输出, 可以看出其通道数是固定的。密集块结构的一个比较显眼的特点就是每一层的瓶颈的输入并非单纯的前一层瓶颈的输出, 而是前面所有瓶颈的输出的拼接; 另一个暗藏的特点就是其在瓶颈中采用了激活函数在前卷积层在后的顺序, 这个顺序被称为预激活 (pre-activation), 在 DenseNet 中拥有更好的表现, 而通常神经网络的顺序是后激活 (post-activation) 顺序, 也就是先卷积后进行激活函数操作。传递层处于相邻的密集块结构之间, 其具体步骤是 BN-ReLU-Conv(1×1)-Max Pooling(2×2), 其中 Conv(1×1) 的结构设计也是为了减少通道数。

在 DenseNet 的网络设计中有一个非常重要的概念就是增长率 k 。在图3.9中增长率为 4, 其实就是每个瓶颈所生成的特征层的层数, 所以可以看到图中每个瓶颈之后的矩形框的集合都是由四个矩形框组成的。如果输入的特征图的通道数为 k_0 , 那么第 ℓ 层的通道数就是 $k_0 + k \times (\ell - 1)$, 其中 ℓ 指的是在单个密集块中的次序。

我们可以从输入输出关系来看普通深度前馈卷积神经网络、ResNet 和 DenseNet 的区别, 用 \mathbf{x}_ℓ 来表示第 ℓ 层的输出, H_ℓ 来表示第 ℓ 层的操作, 具体包括连续的三个操作: batch normalization(BN), rectified linear unit, Conv(3×3)。普通神经网络的输入输出关系可以表示为:

$$\mathbf{x}_\ell = H_\ell(\mathbf{x}_{\ell-1}) \quad (3.13)$$

ResNet 的输入输出关系可以表示为:

$$\mathbf{x}_\ell = H_\ell(\mathbf{x}_{\ell-1}) + \mathbf{x}_{\ell-1} \quad (3.14)$$

DenseNet 的输入输出关系可以表示为:

$$\mathbf{x}_\ell = H_\ell([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1}]) \quad (3.15)$$

DenseNet 结构有许多优点, 由于密集块中每一层的输入都是前面所有层输出的特征层的叠加, 所以其误差传播丢失的问题比较小, 并且由于其瓶颈结构和传递层中特别设计的 Conv(1×1), 其参数总量相较相同层数的普通深度卷积神

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

表 3.1 上表是 DenseNet 的结构细节。可以观察到，不同层数的 DenseNet 之间的结构差别在于后两个密集块。每一种 DenseNet 都是先经过普通的卷积操作和最大池化操作之后，反复通过稠密块和传递层三次，最后经过一层稠密块以后，经过由平均池化层和全连接层组成的分类层进行输出。图中最左列是每一层的名字，第二列是每一层输出的特征图层的图片尺寸，后面的每一列都是不同层的 DenseNet。^[9]

经网络来说大大减少。作者在 CIFAR 和 ImageNet 的数据集上进行的测试显示参数个数和层数远远少于 ResNet 网络的 DenseNet 网络性能就能够超越前者。

第四节 本章总结

在本章中我们详细介绍了我们后面会用到的神经网络架构的。我们首先对机器学习的分类和模型做了介绍，它主要分为三种，监督学习，非监督学习和强化学习。然后我们介绍了一些常用的机器学习模型，其中最重要的也是本论文中要使用的模型就是神经网络模型。我们详细介绍了前馈神经网络的基本要素，神经元，激活函数等等。随后我们详细介绍了卷积神经网络，包括卷积操作，池化操作以及常见的卷积神经网络结构。最后一部分我们详细介绍了卷积神经网络在图像处理问题中的各种改进版本，包括 VGG 神经网络，Resnet 神经网络，Unet 神经网络以及 DenseNet 神经网络。其中 Unet 和 DenseNet 神经网络是我们后面会主要使用的架构，我们后面将会进行进一步的深入讨论。

第四章 基于神经网络的散斑图像恢复方法

前面几章中我们详细讨论了散斑图像的恢复问题，以及它的传统的基于透射矩阵的恢复方法和基于场强的恢复方法。接着我们讨论了卷积神经网络以及它在图像处理问题中的各种改进版本。在这一章我们来讨论使用神经网络进行散斑图像恢复的具体细节。我们将主要构造一种结合了 Unet 神经网络和 DenseNet 神经网络的神经网络架构，然后用这种神经网络来处理散斑图像恢复的问题。我们也会对基于神经网络的恢复方法和传统的恢复方法进行比较。

第一节 神经网络架构

本文的神经网络架构主要是在 tensorflow 上搭建的。Tensorflow 是谷歌开发的开源人工智能代码架构，为人工智能相关从业人员及学生提供服务，通过利用 tensorflow 可以提高神经网络代码编写的速度。目前 tensorflow 已经更新了许多版本，去年全面更新到了 2 系列，相比 1 系列有较大的简化，目前 1.x 的 tensorflow 的代码比较多，2.x 在使用的时候可以参考官网的 api 介绍。Tensorflow 可以被方便地部署到多种平台，包括 CPU、GPU、TPU 等设备进行加速计算。本文搭建的神经网络架构需要普通的包如 os, numpy, skimage, matplotlib 等等，神经网络相关的包有 keras.models、keras.layers、keras.optimizers 和 keras.callbacks 包中的 ModelCheckpoint、LearningRateScheduler、ReduceLROnPlateau、TensorBoard 模块等。其中检查点是相当重要的模块，在检查点中可以方便地设置神经网络的优化方法，批量 (batch size)，学习率，模型保存路径和保存方式等等。Tensorflow 平台为机器学习相关的任务提供了许多的便利。

我们前面已经阐述过如何通过测量透射矩阵然后通过逆转操作来得到原图了，现在我们想要通过神经网络来实现散斑图像恢复这一目标。我们使用的网络的原始结构可以参考 18 年的文章^[10]，作者称该网络为 IDiffNet，网络结构如图 4.1 所示，但并不是广为流传的结构。IDiffNet 使用了 DenseNet^[9] 中首次提出的密集块结构。密集块结构的特点是，单个块的每一层都和所有其他层连接在一起，加强了特征传递和特征再利用，并且实质上减少了神经网络的参数，具有更强的泛化能力。该网络也参考了 Unet，分为膨胀路径和收缩路径，中间有将特征层在两个路径之间跨层进行拼接。收缩路径中先对输入的图片进行卷积和下采样操作，然后在每一层进行反复的作者称为密集并下采样传递块 (dense

and downsampling transition block) 的操作，中间特殊标注进行了一层密集块，而后再在膨胀路径中反复进行作者称为密集并上采样传递块 (dense and upsampling transition block) 的操作，之后进行上采样转移块 (upsampling transition block) 和 convolution+ReLU 使得网络能够输出目标图片。其中密集并下采样传递块包含了两个部分：密集块和下采样转移块 (downsampling transition block)。其中密集并上采样传递块包含了两个部分：上采样转移块和密集块。下采样转移块包括 BN, ReLU, 1×1 的卷积，像素去除 (dropout) 和平均池化 (average pooling)；而上采样转移块包括 3×3 的卷积，ReLU, Upscaling, 3×3 的卷积，ReLU。IDiffNet 的下采样转移块的结构与 DenseNet 中设计的传递层完全一致。其结构的变化在于密集块的瓶颈结构被砍了一半，并且其卷积从普通卷积被调整为空洞卷积。除此以外，IDiffNet 的上采样转移块的精妙结构来自于 16 年推特的研究人员所提出的子像素提高操作 (subpixel upscaling operation)，能够提高输入图像的分辨率。

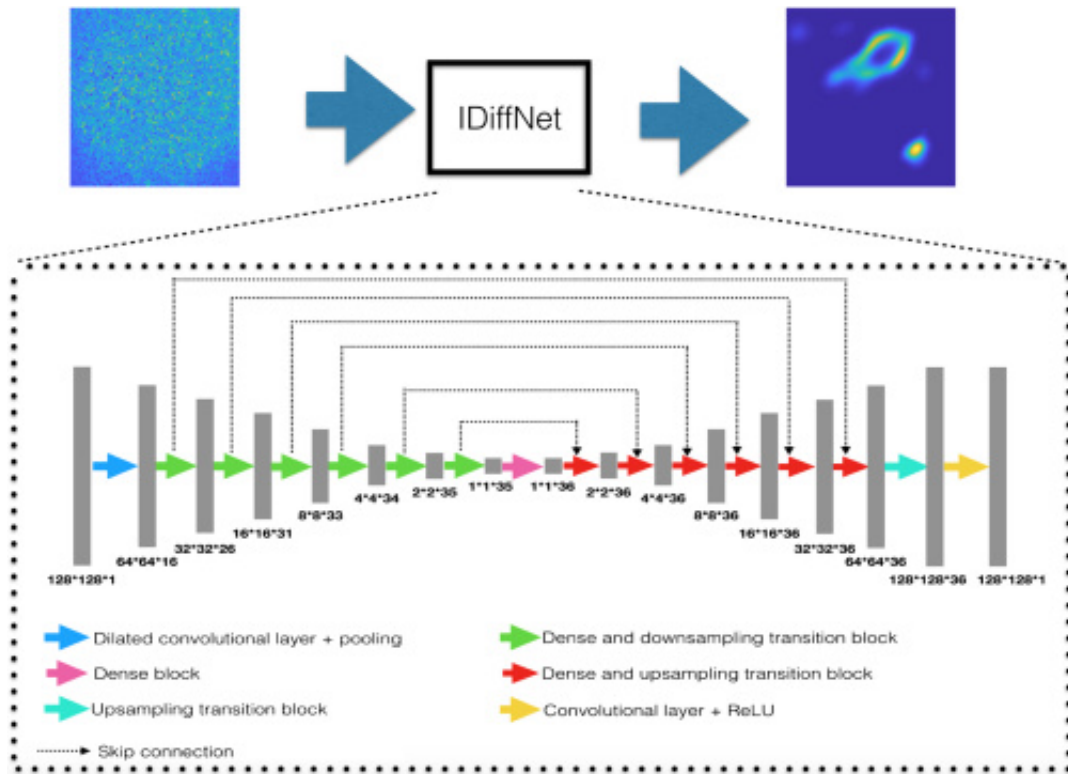


图 4.1 IDiffNet 的网络结构图^[10]。可以看到它具有和 Unet 一样的收缩路径和膨胀路径。

本文使用的神经网络的结构如图4.2所示，继承了上文所说的 IDiffNet 网络加入密集块的思想，大体结构类似于 Unet^[8](图3.8)，只是将 Unet 网络每一层第二列的单一的卷积操作都变成密集块操作。但是上述网络中的密集块中的瓶颈模块与传统的密集块的瓶颈操作的 BN-ReLU-Conv(1×1)-BN-ReLU-Conv(3×3) 模式有稍许不同，其修改为 BN-ReLU-DiConv(filter size: 5×5 dilation rate:2)。其

中，空洞卷积的引入是为了增加卷积层的滤波器的感受野。

神经网络的输入是我们仿真得到的散斑图片的 `numpy` 矩阵, 由于是灰度图, 并非由三原色组成, 所以只是单层的图片, 格式为 $32 \times 32 \times 1$ 。数据流向有两类分支。主线是膨胀和收缩路径: 数据在膨胀路径经过卷积、密集块和最大池化 (`maxpooling`) 下采样; 在收缩路径经过卷积、密集块和上采样。支线是数据拼接: 在膨胀路径和收缩路径的每一层, 都会有经过膨胀路径提取的特征层直接被拼接到收缩路径的每一层输入中。最后经过一层 1×1 的卷积把图片还原回去。

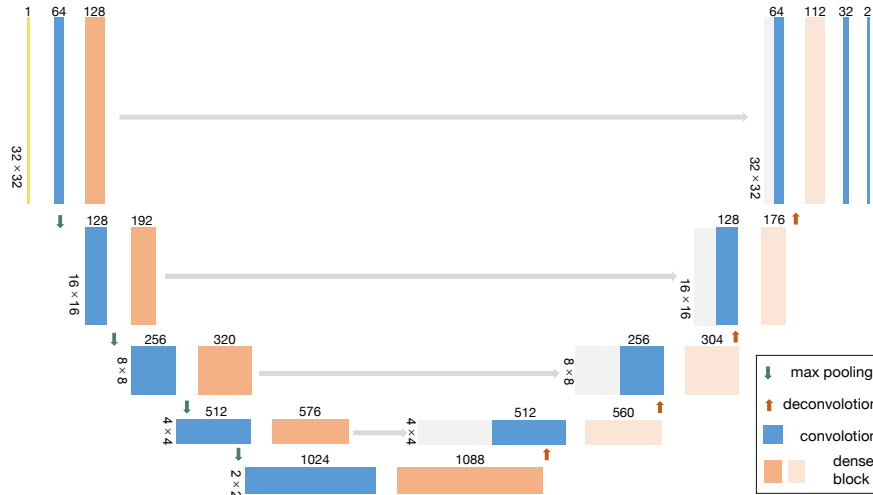


图 4.2 本文用来完成散斑图像恢复的神经网络的结构图。可以看到它具有和 `Unet` 一样的收缩路径和膨胀路径, 图中标注了不同处理之后特征层的层数, 而图像的分辨率可以很容易地通过原始图像大小推知。最终输出层的特征层个数与 `Unet` 保持相同为 2。经过笔者实际测试, 如果强行减少输出层的特征层数为 1 会导致学习效果极大降低。

第二节 数据生成

一、仿真方法

由于我们准备完成的任务是一种监督学习的任务, 需要的数据就是标签数据。神经网络的输入就是散斑图像的数据, 而输出就是还原的图像, 所以我们创建的训练数据集和测试数据集要包括原始的图片, 原始图片的黑白逆转图作为标签, 生成的散斑图作为输入。一般数据的来源有两种, 实验生成和仿真生成。实验步骤和细节我们会在后面稍加阐述。本论文中我们准备使用仿真的方法来生成数据, 而标准图片的来源是采用的 `MNIST` 数据集, 生成的方式是 `matlab` 生成。其中 `MNIST` 数据集中的图片需要从官网下载以后利用 `matlab` 编写代码来进

行解码，然后将图片文件以矩阵的形式存储在 `cell` 的第一列的单元格中，而标签文件在我们的实验中不需要使用，也就无需处理。由于 MNIST 文件中的图片格式是 28×28 ，所以我们需要对其进行重新处理成 32×32 的图片，然后再创建一个新的 `cell` 将原始图片和其黑白对调的方式存到对应的两列中。这里我们只使用的 MNIST 文件中测试数据集的前两千张图片来进行仿真，其中前 1600 张图片在我们的训练中被当作训练数据，而后 400 张图片被我们当作测试数据，训练完成之后再从原始的数据集中调图片出来做测试。我们使用最后生成的 `cell` 文件中的第一列的数据文件也就是 MNIST 数据库原始的图片进行散斑生成，我们通过构造 TM 矩阵进行了散斑的生成。将生成的散斑图片以图片和矩阵两种格式分别存在文件夹和 `mat` 下的 `cell` 中。前一种存储方式是为了提供另一种写成 `numpy` 文件的方式的，由于第二种方式成功了所以暂时不加使用。我们采用了第二种存储数据的方式，但是由于计算机内存和 `matlab` 处理数组的最大限制，我们最多只能存储 32×32 大小的图片。结合训练网络的时间，能使用的云存储内存综合考虑，我最终使用的训练数据也是这个格式。后续如果要拓展的话，我们可以使用 `python` 的 `numpy` 包重新改变数据的大小为其他更大的格式或者使用文件夹中我们生成的图片在使用 `numpy` 处理为矩阵的时候采用更大的图片格式。然而更大的图片对于我们的任务不是关键，只是需要更多的训练时间。我们的关键任务是为了展示神经网络在完成该任务上的优越性。在 `matlab` 上生成了需要的图片矩阵之后，我们将该 `mat` 文件上传了华为云，然后在 `python` 环境下使用 `python` 的 `scipy` 包进行了 `mat` 文件的读取，然后在云端直接创建了 `numpy` 的满足格式的空白数组将 `mat` 文件中的 `cell` 数据写入，使用 `z` 归一化方式对散斑图像数据进行处理，进行训练数据和测试数据的分批，就完成了我们的数据预处理。

二、实验方法

实验产生散斑图像的光路如图4.3所示^[10]。由氦氖激光器产生的光源经过空间滤波器 (spatial filter) 以后，经过瞄准镜和反射镜，然后穿过线性偏振器 (linear polarizer) 和光线分束器。同时 SLM 产生需要做散射的原始图像然后经过分光棱镜的反射后经过一个检偏镜，之后光线被玻璃散射器所散射。最后光线经过一个望远镜成像系统到达 CMOS 相机。由于本工作开展的时间较晚，所以我们采用的方法是仿真方法生成训练数据。

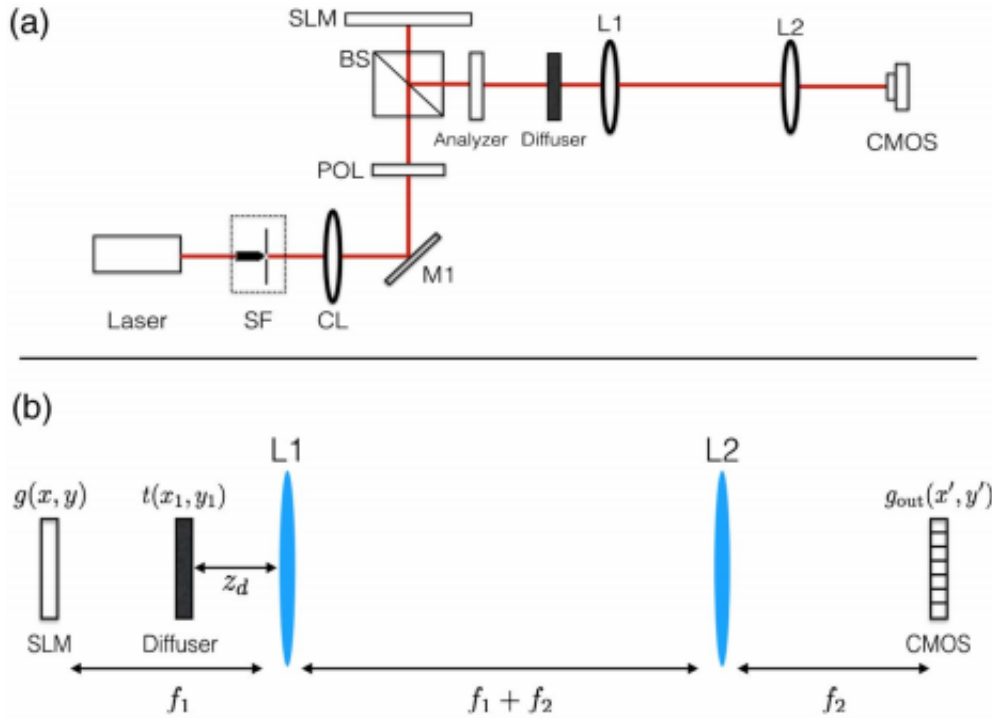


图 4.3 (a) 实验光路图。(b) 望远镜成像系统的细节。^[10]

第三节 神经网络散斑图像恢复结果

一、数据预处理

受限于我们的网络结构，输入图像矩阵的维度只能是 $2^n, n \in \mathbb{N}$ ，由于在前面一部分我们生成的 matlab 的图像的维度是 32×32 ，正好 n 等于 5，能够很好的满足这个要求。那么在经过上一节的数据生成之后，我们拥有原始数据的 numpy 数组了，数组里包含了原始图片的像素数据。出于训练速度等原因，我们需要对原先处于 0-255 之间的像素值进行 z-score 标准化，这一步标准化是在统计了原始数据的平均值和标准差之后代入公式 $x^* = \frac{x - \bar{x}}{\sigma}$ 完成的。

二、训练过程

要完成我们的训练任务需要 Tensorflow-GPU-Keras 的基本架构，并且受限于我们的数据量的两千张图片的大小，还对 GPU 显存的大小有所要求。作者在多个平台的普通主机和远程服务器上反复进行部署，有过多次失败的尝试。失败原因包括 GPU 显存容量不够；显存动态分配无法绘制图像；cuda 版本不吻合并且在 linux 系统下难以进行修改；没有 sudo 权限等等。最后采用的方案是使用华为云的计算资源，因为对方的云 jupyter notebook(GPU) 平台十分稳定高速，并且已经装好了所有可能用到的 python 包。其他尝试过的可用的平台有 colab，但是

由于时间不足等问题，没有继续学习该平台的使用。

对于写入 `mat` 文件的两千张图片，我们将其分为训练集和验证集，送入神经网络进行训练，训练各个参数设置如下：优化器选为 Adam，步长为 10^{-4} ，损失函数为 *binary_crossentropy*，batch size 设置为 20 进行训练。我们的神经网络的训练耗时十小时左右，最佳结果的 accuracy 能达到 0.6321。

三、恢复结果讨论

原始模型训练过程结束以后最终保存的模型文件(命名为 6321.hdf5)，重新载入数据和模型文件进行运行检查预测结果。我们以 MNIST 数据集中的 1,9,5,8 四张图片为例进行了测试。图4.4是 32×32 的手写数字 1,9,5,8 的测试结果: (a) 中是原始的 MNIST 数据集中的四张 32×32 的手写数字图片; (b) 中给出了他们的通过透射矩阵方法模拟生成的散斑图像; (c) 中是我们训练之后的神经网络所给出的恢复图片，我们将散斑图像作为输入，输出的图像非常接近原始图像。我们可以看到，边缘有些畸变之外，我们整体的恢复效果非常好。这主要得益于 Unet 和 DenseNet 相互结合的网络架构，Unet 可以从整体上很好地恢复图像，而密集块结构的引入使得神经网络的梯度丢失等问题得到了更好的处理，所以它在散斑图像恢复的问题上面表现出了很好的结果。

为了进一步的比较我们的神经网络方法对散斑图像恢复问题的处理结果和传统散斑图像恢复问题的处理结果，我们也分别利用了我们在第二章中所讨论的经典的直接求逆法和场强方法对散斑图像进行了恢复。模拟计算的结果如图 4.5所示: (a) 中是 MNIST 数据集中的原始图片; (b) 中给出了利用透射矩阵方法

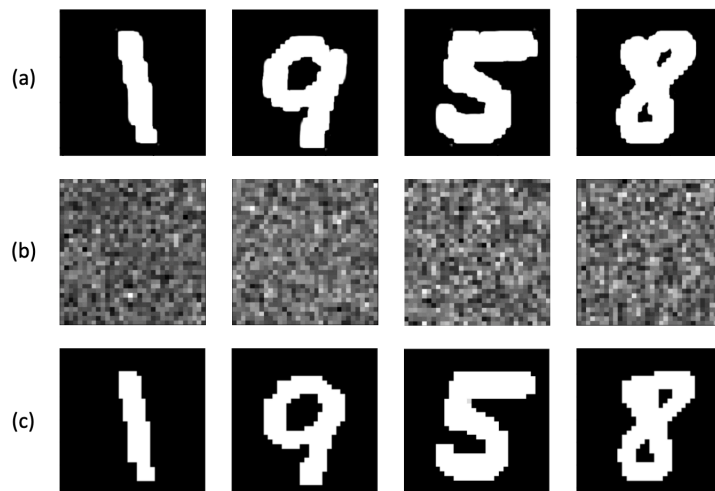


图 4.4 (a) 原始图片; (b) 散斑图片; (c) 恢复图片。

模拟生成的散斑图像；(c) 中给出了利用场强方法进行恢复的数值实验结果；(d) 中给出了利用透射矩阵直接求逆方法给出的数值实验结果。

我们可以看到，传统方法虽然对图像整体的恢复也不错，但其在细节的处理上显得不如神经网络的处理。传统恢复方法恢复的散斑图像仍然携带了一部分的散斑图案。这在数字 1 和 8 的恢复测试中显示得尤为明显。可以看到，具有密集块的 Unet 卷积神经网络在散斑图像恢复的问题中表现很出色。

进一步地，我们使用 numpy 将输入的图层分辨率由原先的 32×32 增加为 64×64 进行训练观察训练结果，发现虽然其准确率无法达到原先的水平，但是效果仍然很不错。考虑到时间因素，我们没有继续训练下去，但是可以调整训练优化器的步长进一步提高准确率。

四、相关工作的进展

最早提出某模块结合 Unet 架构网络来完成散斑图像恢复问题的文章是 17 年 Ayan Sinha 和 Justin Lee 等人发表在 Optica 上的 Lensless computational imaging through deep learning。该文章在 ResNet 的基础上提出了下残差块 (down-residual block)、上残差块 (up-residual block)、空洞卷积块 (dilated residual block) 等结构，并结合 Unet 的不同层的图层拼接的思想，首次构造出了用于散斑图像恢复的网络。该网络能够较好地恢复 ImageNet 数据库中图像经过散射介质后的散斑图像。

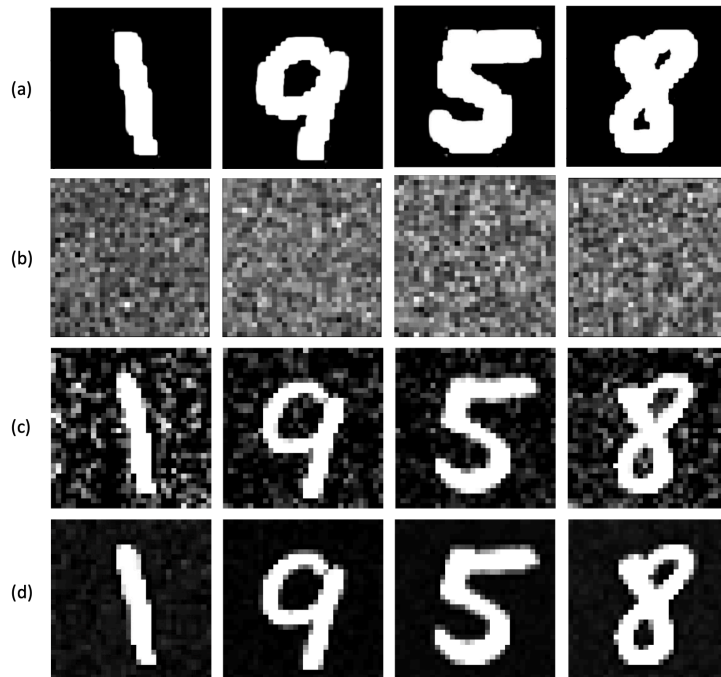


图 4.5 (a) 原始图片；(b) 散斑图片；(c) 场强方法恢复图片；(d) 逆传输矩阵方法恢复图片。

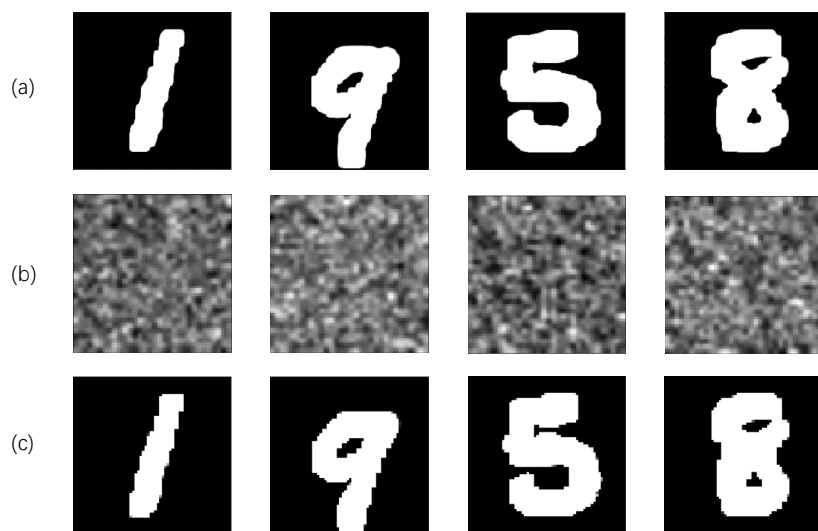


图 4.6 (a) 原始图片；(b) 散斑图片；(c) 恢复图片。

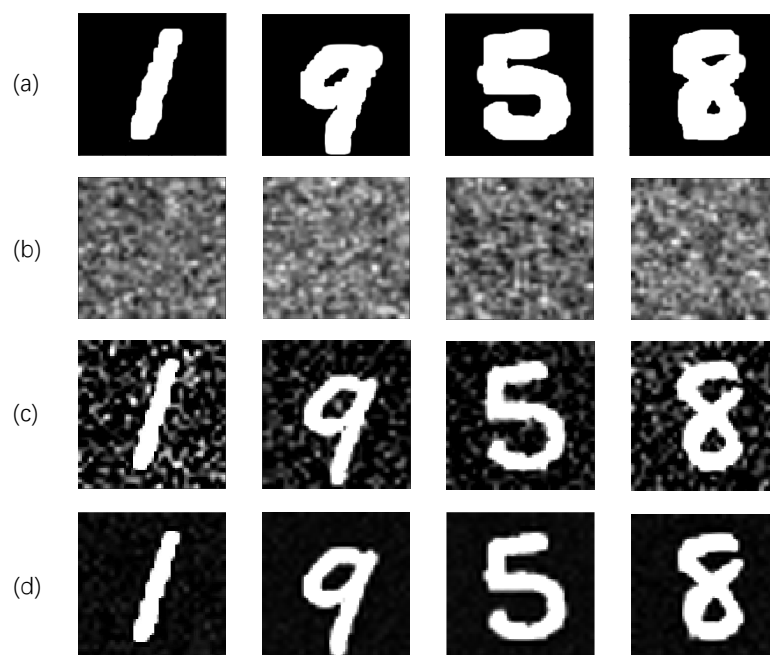


图 4.7 (a) 原始图片；(b) 散斑图片；(c) 场强方法恢复图片；(d) 逆传输矩阵方法恢复图片。

接下来是 18 年 Shuai Li 和 Mo Deng 等人在 18 年^[10]提出的 IDiffNet, 前文已经对该网络进行了介绍。该网络就是 DenseNet 结合了 Unet 的思想。作者在 MNIST 数据库和 ImageNet 数据库上进行了训练和测试, 并且使用 MAE 和 NPCC 作为损失函数以后发现效果都较好。

同样在 18 年, Yunzhe Li^[53] 等人使用 Shuai Li 等人的网络架构和真实的实验

数据进行了更详细的训练，得出了该网络架构的许多特性。该作者进行了四组实验：(1) 使用和训练集同样的原始图片经过不同的散射介质得到的测试图片；(2) 使用和训练集不同的原始图片经过不同的散射介质得到的测试图片；(3) 使用和训练集不同类型的原始图片经过不同的散射介质得到的测试图片；(4) 使用和训练集不同的原始图片经过相同的散射介质得到的测试图片；(5) 使用和训练集不同的原始图片经过不同的散射介质得到的测试图片 (该散射介质的图片并没有放入训练集)。经过多番对比发现，一个深度神经网络能够同时学习经过多种散射介质的图片，但是无法恢复通过未训练的散射介质的散斑图片。

19 年，Mu Yang^[54] 等人用训练好的神经网络恢复了经过散射介质上受训点以外的点的散斑图像。Yunqi Luo 等人进一步在 19 年^[55] 探究了不稳定的散斑图的成像问题。

五、本工作后续可拓展的内容

- 笔者在前期调研的时候，发现还没有人做神经网络预分类再还原的工作，这部分是可以继续做的。
- 由于该模型对于位置误差的容错性，所以可以用来做更贴近现实的实时图像还原。
- 本文使用的训练数据的来源是仿真，受限于 matlab 和电脑内存大小，我们无法产生分辨率大于 32×32 的图片进行训练，要使用更大分辨率更贴合应用的数据只能通过实验，产生散斑图像的实验过程并不复杂，就是需要花一定的时间，可以补上实验部分。
- 加深网络结构、修改网络结构或者改变数据的归一化方法进行训练比较训练效果的优劣，分析网络的性能。
- 系统地针对一批散射介质和某些特定的网络结构进行训练并保存下预训练权重方便后续继续训练。

第四节 本章总结

在这一章，我们详细讨论了基于卷积神经网络的散斑图像恢复方法。我们详细讨论了我们用来做散斑图像恢复的神经网络架构：具有密集块的 Unet 卷积神经网络架构。我们使用 MNIST 数据集，利用透射矩阵方法模拟生成了散斑图像，利用这些图像对神经网络进行了十个小时的训练。最后我们用 1,9,5,8 四个数字

进行了测试，我们发现，这种神经网络架构能够很好地复原散斑图像。我们还对传统的散斑图像恢复也进行了数值实验，并将结果与神经网络方法进行了比较。我们发现，神经网络方法在细节的处理上更加细腻。

第五章 总结与展望

在这篇工作中，我们详细讨论了如何利用卷积神经网络的方法来对散斑图像进行恢复。神经网络因其在计算机视觉领域所取得的巨大成功，现在已经逐渐成为了图像处理问题的一种标准方法。我们可以很自然地考虑如何将其运用在光学系统图像处理中去。散斑图像恢复问题是一个典型的光学图像处理问题，我们利用改造的卷积神经网络对散斑图像恢复问题进行了非常细致全面的探索研究。

我们首先详细介绍了散斑图像恢复问题的基础内容。我们详细介绍了散斑图像的物理成因及其分类。从物理上看，散斑图像是散射介质无规则散射的光束相互叠加所形成的强度图案。我们详细介绍了散斑图像的透射矩阵描述，并以此为基础讨论了传统的散斑图像恢复方法，包括直接求逆方法，光学相位共轭方法，Tikhnov 正则算子方法以及基于光强的方法。

随后我们详细介绍了卷积神经网络的基本内容以及它在图像处理问题中的应用。我们详细介绍了池化和卷积等基本操作，然后介绍了卷积神经网络的常见架构以及它作为分类器和图像处理工具的常见方法。我们尤其强调了它在图像处理问题中的应用，介绍了它常见的几种架构，包括 VGG 神经网络，Res-Net 神经网络，DenseNet 神经网络和 Unet 神经网络。

有了上面的准备，我们详细研究了利用卷积神经网络来处理散斑图像的问题。因为不同的神经网络架构有其各自的优势，我们这里所使用的是结合了 Unet 和 DenseNet 的一种架构，我们在 Unet 中引入了 Dense block。然后我们利用 MNIST 手写数字识别数据集，利用透射矩阵的方法生成了训练所用的散斑图像对神经网络进行了十个小时的训练。最后我们以 1, 9, 5, 8 为例对神经网络进行了测试。测试结果显示我们的神经网络可以很好地恢复散斑图像。我们也对传统的散斑图像恢复问题进行了数值实验，实验结果显示传统方法虽然整体结构恢复的不错，但在细节的处理上还有缺陷，恢复图片中仍然含有部分散斑，与之相比，我们的神经网络方法在细节的处理上表现更好。

虽然我们上面的架构在散斑图像恢复问题中取得了不错的效果，但限于我们使用的图像的尺寸为 32×32 ，对于更大的图像的处理的结果我们仍然不知道结果是否会更好。其次，我们的网络仍然有很多可以改进的地方，我们也可以尝试换用别的神经网络来对散斑图像进行处理，这部分内容在我未来的研究中将

继续探索。在使用方面，我们训练好的程序可以应用在实际实验中，通过实验生成实际的散斑图像，我们构造一个大的训练数据集库，我们甚至可以尝试变换不同的介质而将他们所对应的散斑数据集置于同一训练集里面，这样进行训练之后的神经网络很有可能能够同时处理许多不同介质的散斑图像恢复问题。这相比于传统的针对特殊介质必须特定地去测量透射矩阵然后据此来恢复散斑图像的方法来说，从效率和精度两方面都会有很大的提升。这些内容在我们将来的探索研究中将会继续研究。

参 考 文 献

- [1] GOODMAN J W. Speckle phenomena in optics: Theory and applications[M]. Roberts and Company, Greenwood Village, 2007.
- [2] 刘培森. 散斑统计光学基础[M]. 科学出版社, 1987.
- [3] SKIPETROV S E. Langevin description of speckle dynamics in nonlinear disordered media[J/OL]. Phys. Rev. E, 2003, 67: 016601. <https://link.aps.org/doi/10.1103/PhysRevE.67.016601>.
- [4] LEE K, PARK Y. Exploiting the speckle-correlation scattering matrix for a compact reference-free holographic image sensor[J]. Nature communications, 2016, 7(1): 1-7.
- [5] LEE K, PARK Y. Interpreting intensity speckle as the coherency matrix of classical light[J/OL]. Phys. Rev. Applied, 2019, 12: 024003. <https://link.aps.org/doi/10.1103/PhysRevApplied.12.024003>.
- [6] VAN PUTTEN E G, MOSK A P. The information age in optics: Measuring the transmission matrix[J]. Physics, 2010, 3: 22.
- [7] GOODMAN J W. Statistical properties of laser speckle patterns[M]//Laser speckle and related phenomena. Springer, 1975: 9-75.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, 2015: 234-241.
- [9] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [10] LI S, DENG M, LEE J, et al. Imaging through glass diffusers using densely connected convolutional networks[J]. Optica, 2018, 5(7): 803-813.
- [11] BARBASTATHIS G, OZCAN A, SITU G. On the use of deep learning for computational imaging[J]. Optica, 2019, 6(8): 921-943.
- [12] GOODMAN J W. Some effects of target-induced scintillation on optical radar performance[J]. Proceedings of the IEEE, 1965, 53(11): 1688-1700.
- [13] DAINTY J. Some statistical properties of random speckle patterns in coherent and

- partially coherent illumination[J]. *Optica Acta: International Journal of Optics*, 1970, 17(10): 761-772.
- [14] DUBOIS A, VABRE L, BOCCARA A C, et al. High-resolution full-field optical coherence tomography with a linnik microscope[J]. *Applied optics*, 2002, 41(4): 805-812.
- [15] POPOFF S M, LEROSEY G, CARMINATI R, et al. Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media[J/OL]. *Phys. Rev. Lett.*, 2010, 104: 100601. <https://link.aps.org/doi/10.1103/PhysRevLett.104.100601>.
- [16] CHOI Y, YANG T D, FANG-YEN C, et al. Overcoming the diffraction limit using multiple light scattering in a highly disordered medium[J/OL]. *Phys. Rev. Lett.*, 2011, 107: 023902. <https://link.aps.org/doi/10.1103/PhysRevLett.107.023902>.
- [17] LOVE G D. Wave-front correction and production of zernike modes with a liquid-crystal spatial light modulator[J]. *Applied optics*, 1997, 36(7): 1517-1524.
- [18] MAIT J N, EULISS G W, ATHALE R A. Computational imaging[J]. *Advances in Optics and Photonics*, 2018, 10(2): 409-483.
- [19] HANSEN P C, NAGY J G, O'LEARY D P. Deblurring images: matrices, spectra, and filtering[M]. SIAM, 2006.
- [20] VELLEKOOP I M, MOSK A. Focusing coherent light through opaque strongly scattering media[J]. *Optics letters*, 2007, 32(16): 2309-2311.
- [21] TIHONOV A N. Solution of incorrectly formulated problems and the regularization method[J]. *Soviet Math.*, 1963, 4: 1035-1038.
- [22] MIKOLOV T, DEORAS A, POVEY D, et al. Strategies for training large scale neural network language models[C]//2011 IEEE Workshop on Automatic Speech Recognition & Understanding. IEEE, 2011: 196-201.
- [23] HINTON G, DENG L, YU D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. *IEEE Signal processing magazine*, 2012, 29(6): 82-97.
- [24] NIELSEN M A. Neural networks and deep learning: volume 25[M]. Determination press San Francisco, CA, 2015.
- [25] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling[J]. *IEEE transactions on pattern analysis and machine intelligence*,

- 2012, 35(8): 1915-1929.
- [26] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge[J]. *nature*, 2017, 550(7676): 354-359.
- [27] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. *Journal of Field Robotics*, 2020, 37(3): 362-386.
- [28] MA J, SHERIDAN R P, LIAW A, et al. Deep neural nets as a method for quantitative structure–activity relationships[J]. *Journal of chemical information and modeling*, 2015, 55(2): 263-274.
- [29] JEAN S, CHO K, MEMISEVIC R, et al. On using very large target vocabulary for neural machine translation[J]. *arXiv preprint arXiv:1412.2007*, 2014.
- [30] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[J]. *arXiv preprint arXiv:1409.3215*, 2014.
- [31] CARLEO G, CIRAC I, CRANMER K, et al. Machine learning and the physical sciences[J/OL]. *Rev. Mod. Phys.*, 2019, 91: 045002. <https://link.aps.org/doi/10.1103/RevModPhys.91.045002>.
- [32] TORLAI G, MAZZOLA G, CARRASQUILLA J, et al. Neural-network quantum state tomography[J]. *Nature Physics*, 2018, 14(5): 447.
- [33] BROECKER P, ASSAAD F F, TREBST S. Quantum phase recognition via unsupervised machine learning[J]. *arXiv preprint arXiv:1707.00663*, 2017.
- [34] HUEMBELI P, DAUPHIN A, WITTEK P. Identifying quantum phase transitions with adversarial neural networks[J]. *Physical Review B*, 2018, 97(13): 134109.
- [35] CARLEO G, TROYER M. Solving the quantum many-body problem with artificial neural networks[J/OL]. *Science*, 2017, 355(6325): 602-606. <http://science.sciencemag.org/content/355/6325/602>. DOI: 10.1126/science.aag2302.
- [36] TORLAI G, MELKO R G. Neural decoder for topological codes[J/OL]. *Phys. Rev. Lett.*, 2017, 119: 030501. <https://link.aps.org/doi/10.1103/PhysRevLett.119.030501>.
- [37] TORLAI G, MELKO R G. Latent space purification via neural density operators[J/OL]. *Phys. Rev. Lett.*, 2018, 120: 240503. <https://link.aps.org/doi/10.1103/PhysRevLett.120.240503>.
- [38] TORLAI G, MELKO R G. Learning thermodynamics with boltzmann machines

- [J/OL]. Phys. Rev. B, 2016, 94: 165134. <https://link.aps.org/doi/10.1103/PhysRevB.94.165134>.
- [39] GOODFELLOW I, BENGIO Y, COURVILLE A, et al. Deep learning: volume 1 [M]. MIT press Cambridge, 2016.
- [40] SUTTON R S, BARTO A G. Reinforcement learning: An introduction[M]. MIT press, 2018.
- [41] MINSKY M, PAPERT S A. Perceptrons: An introduction to computational geometry[M]. MIT press, 2017.
- [42] GHOLAMALINEZHAD H, KHOSRAVI H. Pooling methods in deep neural networks, a review[J]. arXiv preprint arXiv:2009.07485, 2020.
- [43] RAWAT W, WANG Z. Deep convolutional neural networks for image classification: A comprehensive review[J]. Neural computation, 2017, 29(9): 2352-2449.
- [44] HUBEL D H, WIESEL T N. 8. receptive fields of single neurones in the cat's striate cortex[M]//Brain Physiology and Psychology. University of California Press, 2020: 129-150.
- [45] FUKUSHIMA K, MIYAKE S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]//Competition and cooperation in neural nets. Springer, 1982: 267-285.
- [46] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.
- [47] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010: 249-256.
- [48] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.
- [49] NASH W, DRUMMOND T, BIRBILIS N. A review of deep learning in the study of materials degradation[J]. npj Materials Degradation, 2018, 2(1): 1-12.
- [50] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [51] <https://github.com/fchollet/deep-learning-models/releases/>[Z].

- [52] 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.
- [53] LI Y, XUE Y, TIAN L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media[J]. Optica, 2018, 5(10): 1181-1190.
- [54] YANG M, LIU Z H, CHENG Z D, et al. Deep hybrid scattering image learning [J]. Journal of Physics D: Applied Physics, 2019, 52(11): 115105.
- [55] LUO Y, YAN S, LI H, et al. Towards smart optical focusing: Deep learning-empowered wavefront shaping in nonstationary scattering media[J]. arXiv preprint arXiv:1909.00210, 2019.

在读期间发表的学术论文与取得的研究成果

已发表论文

1. Zhih-Ahn Jia, **Lu Wei**, Yu-Chun Wu, Guang-Can Guo, Guo-Ping Guo, Entanglement Area Law for Shallow and Deep Quantum Neural Network States, New J. Phys. 22 053022 (2020)

待发表论文

1. Zhih-Ahn Jia, **Lu We**, Yu-Chun Wu, Guang-Can Guo, Quantum Advantages of Communication Complexity from Bell Nonlocality, arXiv:2004.05098