# 摘要

## 基于深度学习的图像去噪研究与应用

图像是一种常见的信息存储形式，其不仅能直观地表示大量相关的信息，还具有易于存储和传输等优势。由于各种环境和信道因素的影响，图像在采集、压缩和传输过程中难免受到噪声的干扰，导致图像信息失真或丢失。同时，因为噪声的存在，可能对后续的图像处理任务，如图像分类产生不利影响。真实噪声的无噪声标签也不易获取。因此，如何依靠少量训练样本并从含噪图像中恢复出有意义的信息是当今图像处理领域内的一个重要问题。

为了解决真实噪声去噪问题，本文提出了多尺度两阶段图像去噪网络模型（MTDNet）与基于元学习的多尺度两阶段去噪模型（MMTDNet）。本文的主要研究工作分为两个方面，分别为高性能图像去噪模型的构建、在少量训练样本中提高模型的去噪能力。最后将该模型投入到网站应用中。

具体相关研究工作如下：

（1）本文基于 UNet 提出了一种多尺度两阶段图像去噪网络模型，目标是在保留图像的高维信息条件下实现对图像中真实噪声的去除。在网络的两个阶段中，本研究利用注意力机制、空洞卷积与普通卷积提取不同尺度的特征，再将第一阶段空洞卷积与第二阶段普通卷积各自提取到的不同尺度的特征进行特征融合，使模型能够获取足够多的特征信息，进而完成去噪任务。

（2）为了在训练数据不足的真实噪声数据集上达到更加良好的去噪效果，将本文提出的去噪网络模型与元学习思想相结合，并划分出合成噪声子任务用以元训练，以达到可以使用少量的真实噪声训练样本（例如古文字去噪训练集）完成对去噪网络的训练，进而达到更加理想的去噪效果。

文中对本研究提出的 MTDNet 进行了有效性实验验证。实验结果表明，在 SIDD 测试集上，MTDNet 相较于基线模型 UNet，其评价指标 PSNR 提升 2.55，SSIM 提升 0.09，并且性能优于 DnCNN 等其他去噪模型；在甲骨文测试集上，去噪效果仍明显优于其他去噪模型，相较于基线模型 UNet，评价指标 PSNR 与

SSIM 分别提升 3.15 和 0.16。此外，本文对 MTDNet 内部添加或修改的各模块进行了消融实验，证明了各模块对去噪任务的有效性。

同样，文中对于本研究提出的 MMTDNet 也进行了有效性实验验证。实验结果表明，在 SIDD 测试集上，MMTDNet 相较于 MTDNet，其评价指标 PSNR 再次提升 2.35，SSIM 提升 0.05；在甲骨文测试集上，MMTDNet 相较于 MTDNet，评价指标 PSNR 与 SSIM 分别进一步提升 2.76 与 0.04。此外，本文对于元学习的泛化性也进行了有效性实验验证。实验表明，与迁移学习和普通的监督学习相比，元学习的泛化性也更占优。

**关键词：**

图像去噪，元学习，深度学习，计算机视觉

# Abstract

## Research and Application of Image Denoising Based on Deep Learning

Images are a prevalent means of storing information as they can visually convey a large amount of relevant data and are convenient for storage and transmission. However, when images are acquired, compressed, or transmitted, they are inevitably subjected to noise due to various environmental and channel factors, leading to distortion and loss of image information. The presence of noise can have negative consequences on subsequent image-processing tasks, such as image classification. Furthermore, obtaining noisy-clean labels for real noise can be challenging. Thus, it is crucial to address the issue of restoring meaningful information from noisy images, particularly with limited training samples.

This paper presents two novel models, namely, the Multi-Scale Two-Stage Image Denoising Network (MTDNet) and the Meta-Learning-based Multi-Scale Two-Stage Image Denoising Network (MMTDNet), to address the challenge of denoising real noise. The main focus of this study is divided into two aspects: firstly, the development of a high-performance image denoising model, and secondly, enhancing the denoising capability of the model when trained with a limited number of samples. Additionally, the proposed models are implemented as a web application.

This paper's contributions include:

(1) This paper propose a novel Multi-Scale Two-Stage Image Denoising Network model based on UNet architecture to effectively remove real noise and restore high-dimensional image information. The proposed model utilizes attention mechanisms, void convolution, and ordinary convolution in the two network stages to extract features at various scales. Subsequently, the features from the first stage void convolution and

the second stage ordinary convolution are merged to provide sufficient feature information for the model to accomplish the denoising task. This approach facilitates the restoration of image information while effectively removing noise.

(2) This paper introduce a denoising network model that integrates the concept of meta-learning and divides the synthetic noise sub-task for meta-training to improve the denoising performance on real noise datasets with limited training data. This approach aims to leverage a small set of real noise training samples, such as the ancient text denoising training set, to train the denoising network and achieve optimal denoising outcomes.

The effectiveness of MTDNet proposed in this paper is verified by experments. The experimental results show that in the SIDD test set, compared with the baseline model UNet, MTDNet's evaluation index PSNR increases by 2.55 and SSIM increases by 0.09, and its performance is better than other denoising models such as DnCNN; In the Oracle test set, the denoising effect is still significantly better than other denoising models. Compared with the baseline model UNet, the evaluation indicators PSNR and SSIM increase by 3.15 and 0.16. In addition, ablation experiments were carried out on each module added or modified in MTDNet, which proved the effectiveness of each module in the denoising task.

The effectiveness of MMTDNet proposed in this paper is also verified by experiments. The experimental results show that in the SIDD test set, compared with MTDNet, the evaluation index PSNR and SSIM of MMTDNet increase by 2.35 and 0.05 respectively; Compared with MTDNet, the evaluation indicators PSNR and SSIM of MMTDNet increased by 2.76 and 0.04 respectively in the Oracle test set. In addition, the generalization of meta-learning is also verified by experiments. The experiment shows that compared with transfer learning and general learning, the generalization of meta-learning is also better than transfer learning and general learning.

**Keywords:**

Image denoising, Meta-learning, Deep learning, Computer vision

# 关于学位论文使用授权的声明

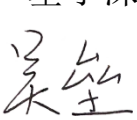
本人完全了解吉林大学有关保留、使用学位论文的规定，同意吉林大学保留或向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅；本人授权吉林大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或其他复制手段保存论文和汇编本学位论文。

（保密论文在解密后应遵守此规定）

论文级别：■硕士 □博士

学科专业： 软件工程

论文题目： 基于深度学习的图像去噪研究与应用

作者签名：  指导教师签名： 

2023 年 5 月 24 日

# 目　　录

# 第 1 章 绪论

## 1.1 研究背景与意义

图像是一种常见的信息存储形式，其不仅能直观地表示大量相关的信息，而且比较易于存储和传输。

噪声是指图像中不希望出现的干扰信息，它可能是由多种因素造成的，例如摄像机噪声和传输干扰等。由于各种环境和信道因素的影响，无论是在图像的采集、压缩还是在图像的传输过程中，都不可避免的受到噪声的干扰，从而导致图像信息失真和丢失。这些噪声可能会对后续的图像处理任务，如图像识别、目标检测等产生负面影响。因此，在现代的图像处理任务中，图像去噪常常作为其他任务的预处理手段之一，目标在于将图像所包含的噪声去除掉。然而，由于图像内不论是噪声信息还是图像本身的纹理信息与边缘信息都同属于高频信息，有时噪声信息与图像本身信息叠加，会导致这些信息之间难以区分，噪声可能会被误认为图像本身信息，而图像本身信息也可能会被误认为是噪声信息，因此去噪后的图像可能会面临一部分细节信息的损失，或是噪点未完全去除的问题。总体而言，如何从带有噪声的图像中恢复出有意义的高质量图像是当今一个重要的问题。

在古文字研究领域中，甲骨文，金文分别刻在甲骨和铜鼎上，由于这两种媒介长年受到氧化、风化等影响，遭到不同程度的腐蚀，在文字图片上呈白色噪声[1]，从而影响古文字学家对文字的考释以及古文字的数字化存储。如图 1.1 所示，不同古文字图像噪声各不相同，有些噪声存在于文字周围，有些噪声存在于文字笔划上，且在观感上异于高斯噪声等合成噪声或电路干扰的真实噪声。因此，对古文字图像去噪是一件值得研究的事情。此外，古文字相比现代文字数据量少且不可复制，数据十分珍贵，综上所述，将已有的古文字图像进行去噪处理对维护古代文化和助力现代的研究都有一定的帮助。

图 1.1 包含噪声的甲骨文图片

## 1.2 研究现状

去噪作为图像处理中一个常见的问题，其主要目标是尽可能地去除噪声，并还原原本图像该有的细节，提高图像质量。图像去噪按照技术手段通常可以分成两种，即传统图像去噪和基于深度学习图像去噪这两类算法。对于传统图像去噪算法，通常需要先对图像进行噪声等级预估，再使用指定强度的滤波器对图像进行去噪，无法对图像进行盲去噪（不需预估噪声等级，直接进行去噪）处理。基于深度学习的图像去噪算法通常可以用于盲去噪，同时，其对数据量的需求也较大，但实际上，例如在古文字图像去噪领域中，古文字噪声-干净图片对的数据量很少，且不易获取。那么如何在样本不足的情况下依然能让模型学习特定领域噪声依然是研究的关键。下文将对传统图像去噪算法、深度学习图像去噪算法、小样本学习算法的研究现状进行介绍。

### 1.2.1 传统图像去噪算法

传统图像去噪算法在发展上一般划分为两个阶段，即空间域方法和变换域方法。下文将对这两种方法进行介绍。

**空间域方法**

空间域去噪方法是指使用图像中某个像素周围的像素信息来对当前图像像素进行处理的方法。空间域去噪方法的基本流程是：先将图像划分为若干个小块，然后对每一个块使用相应的滤波器进行处理。常用的空间域去噪方法有均值滤波[2]、中值滤波[3]、高斯滤波[4]等。均值滤波法使用相邻像素的平均值来替换当前像素的值，这种方法在去除白噪声（例如扫描噪声或摄像头噪声）时效果较好。中值滤波法使用相邻像素的中位数来替换当前像素的值，能够有效地抑制大规模的噪声。高斯滤波器使用高斯函数来计算相邻像素的加权平均值，可以有效地模拟人眼对图像的感知。自适应滤波是一种动态调整滤波器参数的方法，以最大程度地保留图像中的细节，这种方法通常使用一个局部区域作为滤波器，并根据局部区域内的噪声水平调整滤波器参数。双边滤波法[5]同时考虑了像素的灰度值和距离，根据这两个信息来计算每个像素的值，这样可以有效地保留图像的边缘信息。导数滤波法通过计算图像的导数来检测和去除噪声。

空间域去噪方法的优点在于它们通常比频率域方法更容易实现，并且可以有效地保留图像的细节信息。但是同时存在一些缺点，如对于高频噪声的去除效果不太好，并且会使图像边缘变得模糊。另外，这些方法也可能会使图像处理变慢，因为它们通常需要对图像的每个像素进行计算。此外，它不能有效去除黑色噪声（例如椒噪声），并且使用空间域去噪方法可能会导致图像模糊，因为它会抹除图像中的细节。尽管如此，空间域去噪仍然是一种有效的图像处理技术，可以在许多应用中使用。例如，它可以用于去除扫描噪声、摄像头噪声以及其他类型的白噪声。

**变换域方法**

变换域去噪方法是另一种常用的传统图像去噪方法，这种方法利用图像信息和噪声在变换域中的特性不同进行去噪，使用图像域变换后的信息对图像进行处理。其基本思路是将图像转换到另一个域（如频域或小波域），利用图像信息和噪声在变换域中的特性不同，在变换域中提取信号的有用部分后再对图像进行处理，最后将图像转换回原域，从而实现去噪的目的。这种方法通常可以比空间域方法更有效地去除高频噪声，但是可能会丢失图像的细节信息。最初，变换域方法由傅里叶变换发展而来，此后逐渐出现了多种变换域方法，如余弦变换[6]、小波域方法[7]、块匹配三维滤波[8]（BM3D）等。

傅里叶变换是指通过使用傅里叶变换将图像从时域转换到频域，然后在频域对图像进行处理，最后将图像转换回原域，傅里叶变换能够有效地抑制周期性的噪声。小波域方法是指使用小波变换将图像从时域转换到小波域，然后利用小波域信息对图像进行处理，最后将图像转换回原域。小波变换能够有效地抑制小规模的噪声。离散余弦变换是一种线性变换，这种变换方式可以通过矩阵乘法实现，计算量较小，所以计算速度快，同时可以把信号从时域分析成频域，因此可以方便地进行频域分析。

### 1.2.2 深度学习图像去噪算法

随着神经网络的发展，卷积神经网络被广泛用于计算机视觉领域，可用于解决图像识别、目标检测、语义分割等任务。目前，卷积神经网络已成为图像去噪的主流方法[9, 10]，许多不同的网络结构被引入到了图像去噪领域中，其中包括最

初用于医学影像分割的 UNet[11]，以及起初用于图像分类的密集网络[12]（DenseNet）与残差网络[13]（ResNet），还有应用于图像分类和目标检测的注意力模块[14]（Attention）等。UNet 由于模型带有跳跃连接，可以帮助 Decoder 保留更多的特征信息，DenseNet 与 ResNe 各自通过密集连接与残差连接来减少梯度消失问题，帮助网络更好地学习噪声特征，Attention 模块可以对特征图地不同通道分别赋予不同权重，一定程度上解决模型在颜色空间不敏感问题。DnCNN[15]作为最著名的去噪模型，完成了对文字图像的去噪处理[16, 17]。这些基于神经网络的去噪模型对于文字图片去噪效果良好，甚至可以去除部分图片中的混合噪声。Neji 等人[18]设计了一个对抗自编码器来自动去除古文字图像中的椒盐噪声和高斯噪声。张等人引入了一种基于 GAN[19]（Generative Adversarial Network, GAN）的文字图像去噪器，通过学习混合的合成噪声来生成去噪后的文字图片[20]。一些基于其他的基于 GAN 的去噪器也都尝试从干净-噪声图片对中学习真实世界噪声，并对其进行建模[21]。由于干净-噪声图片对并不好获取，所以 Linh Duy Tran 等人在 2020 年提出了两步框架[22]，先使用 GAN 学习噪声并提取噪声分布，构建并得到训练数据对，再将这些生成好的数据输入到卷积神经网络中进行去噪训练，取得了较好的结果。相较于传统算法，无论是在数值还是视觉效果上，卷积神经网络的去噪效果都更加出色。

### 1.2.3 小样本学习

计算机视觉领域依靠深度神经网络模型在诸多任务上取得了不错的成就，但这背后却需要海量的高质量标注数据[23]，甚至有时训练数据的质量与数量相比于模型本身来说更加重要，也就是说数据对于训练结果来说，其重要性也非常高，当训练数据不充足，无法覆盖全部或大多数测试样本时，基于深度学习的模型会出现不同程度的性能退化。因此，对于某些数据不易获取或数据有限的任务来说，缺少了大量训练数据的支持，模型训练效果也会大打折扣。解决这种训练数据不充足的模型学习任务通常被称为小样本学习。小样本学习通常分为三个主要方向：（1）数据方面，由于训练数据不足，那么能在原始数据集的基础上扩充数据集[24]，将会是一个有效的方法。（2）模型方面，通过一些训练样本充足的相关任务的先验知识来缩小指定任务的假设空间的范围。（3）算法方面，通过一些训

练样本充足的相关任务的先验知识来修正在指定任务的假设空间中，任务模型寻找最优解的搜索策略[25]。

数据方面，通常使用已有的数据来创造新的样本，例如使用变形、缩放等方式进行数据增强，或使用 mixup[26]方法，通过线性插值将多个样本数据混合在一起，从而创造出新的样本。这种方式可以一定程度上降低对数据数量的要求，但在数据样本极少时仍无法满足任务需求的数据量。

基于模型的改进有多任务学习[27]（Multitask Learning），顾名思义，多任务学习就是让模型同时处理多个任务，同时，在处理不同任务时共享同一套参数，这可以让模型学习到不同任务之间的相关性且能同时解决多个任务，提高模型效率。

不论是数据方面还是模型方面的改进，在目标数据不足时的效果仍不够理想。对于算法方面的改进，例如元学习（Meta-learning）的思想则是让模型学会学习（Learning to learn），通过非目标数据学习解决某一大类任务的先验知识，在面对目标样本时，将先验知识应用于目标任务上，可以在少量的目标样本训练下取得较好的效果，因此也可以使用元学习来解决小样本问题。

## 1.3 研究内容

本文主要研究基于深度学习的图像去噪算法，尤其是古文字图像真实噪声去噪。

主要研究内容总结如下：

（1）为实现对真实噪声良好的去噪能力，且能够尽可能保留图像的高维信息，本文提出了一种多尺度两阶段图像去噪网络模型。在网络的两阶段中，利用注意力机制、空洞卷积与普通卷积提取不同尺度的特征，再将第一阶段空洞卷积与第二阶段普通卷积各自提取到的不同尺度的特征进行特征融合，进而完成去噪任务。

（2）为了在训练数据不足的古文字数据集上达到良好的去噪效果，将去噪网络与元学习思想相结合，并划分出噪声子任务用以元训练，目的是最终可以使用少量的古文字训练集对去噪网络进行训练即可达到较为理想的去噪效果。

## 1.4 论文组织结构

本文的章节安排如下：

第 1 章：绪论。本章首先简述了真实图像去噪与甲骨文图像去噪的研究背景与研究意义，然后详细介绍了传统图像去噪、基于深度学习的图像去噪与小样本学习的研究现状，最后介绍了本文的研究内容以及论文章节安排。

第 2 章：图像去噪和神经网络基本理论。本章首先简述图像去噪原理，包括噪声模型与图像去噪算法与基于深度学习的去噪方法，然后简述了卷积神经网络的相关知识，最后介绍了关于小样本学习的元学习理论。

第 3 章：多尺度两阶段图像去噪网络。首先详细介绍了本章提出的多尺度两阶段图像去噪网络模型结构以及内部不同模块之间的关系与功能，包括内部的基线模型 UNet、残差通道注意力模块等。然后介绍了两个阶段之间的特征融合方法，包括作为二阶段输入前的多尺度特征融合与一二阶段 UNet 的多尺度特征融合。最后在实验中，介绍了实验数据集与实验参数，通过对比实验证明了本模型在真实图像去噪方面的有效性，还通过消融实验证明了模型内部各模块的有效性。

第 4 章：基于元学习的多尺度两阶段图像去噪网络。首先详细介绍了 MAML 这种基于优化的元学习思想以及算法特点。随后讲述了对 MAML 这种训练方式的子任务划分问题。然后讲述了将去噪模型与 MAML 结合后的训练过程。最后在实验中，介绍了元学习实验数据集与实验参数，通过对比实验证明了本模型相较于前文提出的模型在小样本数据集上更加有效，还通过泛化性实验证明了元学习在小样本去噪数据集上的泛化性更佳。

第 5 章：去噪应用平台。首先对应用平台架构进行讲解，再详细介绍去噪功能模块的实现与逻辑过程，并展示了去噪功能的交互页面。

第 6 章：总结与展望。对本文全文的研究内容与研究结果进行总结，并对未来图像去噪方法的发展进行展望。

# 第 2 章 图像去噪和神经网络基本理论

## 2.1 图像去噪原理

### 2.1.1 噪声模型

图像的噪声是指遮挡了图像正常像素点的异常像素点，这些异常像素点被称为噪点。对于合成噪声，噪声图像$Y$可以使用$Y = X + N$进行表示，其中$N$表示不同的合成噪声分布。然而，这种表示方法在真实退化图像上的表现并不理想，因为真实字符图像的退化模型更为复杂，且不同于上述合成噪声[21]。尽管真实噪声对图像造成的退化效果各有千秋，但可以将包含真实噪声的图像抽象地认为是由真实图像与一个分布规律不确定的噪声组成的，可以用式 2.1 进行表示：

$$Y = X + \sum f(N_i) \cdots\cdots (2.1)$$

其中$X$表示真实图像，$N_i$表示特定噪声类别的分布情况，$f(\cdot)$表示不确定噪声水平的分布函数，$f(N_i)$的和表示混合噪声的分布。

### 2.1.2 图像去噪算法

图像去噪是图像处理的一个重要问题，随着计算机视觉和机器学习技术的发展，去噪方法也在不断演变，可以大致分为两个方向，一个是完整图像去噪，另一个是将完整图像分解，随后基于分解后的小图像块进行去噪。完整图像去噪是最初的图像去噪方法，它以整个图像为单位进行处理。传统的图像处理算法通常使用这种方式，如中值滤波、高斯滤波等。然而，随着图像复杂性的增加，完整图像去噪的效果变得不理想，因此又出现了基于图像块的图像去噪算法。基于图像块的图像去噪是指通过将图像分成若干个块，在块内对噪声进行去除处理，再对每个图像块的处理结果进行组合，从而得到去噪后的图像。较为经典的基于图像块的去噪算法有 BM3D 和 NLM[28]。BM3D 使用块匹配算法来在图像中寻找与当前图像块相似的图像块。然后利用这些相似的图像块来构建一个三维（包括时间）的数据集，并使用高斯滤波器对其进行处理。NLM 对每个像素都会在图像中搜索与其相似的图像块，随后计算图像块像素均值作为此像素的原始像素预测值，所以 NLM 方法对于每个像素都会生成一个去噪值，最终生成去噪后的图像。

近年来，基于图像块去噪的思想在基于深度学习的图像去噪算法内也有一定的体现，例如在提取特征时，采用感受野小的卷积核进行卷积计算，为了保证算法的普适性，一般不采用全连接结构，因为全连接结构由于参数固定，只能处理特定大小的图片块，同时具有过多的参数量，容易导致网络过拟合或浪费训练资源。

## 2.2 卷积神经网络

卷积神经网络（Convolutional Nerual Networks，CNN）是一种专门用来处理具有类似网格结构的数据的深度神经网络。卷积神经网络已经在计算机视觉领域内多次证明了其有效性，例如在图像识别[29]、目标检测[30]、语义分割[31]等任务上表现都尤为出色。相较于传统神经网络来说，卷积神经网络一方面加入了卷积运算；另一方面就是将全连接改为局部连接。所以，卷积神经网络的局部连接特性使得它能够更有效地学习图像中的局部特征，传统神经网络的全连接则更适合处理全局特征。因为全连接神经网络每个神经元都需要一个独立的权重和偏置参数，并只能处理固定分辨率大小的图片，当图片分辨率变大时，参数数量也会随之大幅增加，难以应对大规模图像处理。局部连接的优势在于它既可以学习到输入图像中的局部特征，也能够有效地减少参数数量和计算量。因为图像和视频数据通常具有局部性和平移不变性的性质，全连接的传统神经网络通常需要更多的参数和计算量来处理图像数据，而图像数据通常较大，难以计算，此外，全连接网络需要对整个图像进行学习，这也可能导致难以处理图像中的平移变换。因此 CNN 更适合处理图像和视频数据。

卷积神经网络内主要由三部分组成，分别为：卷积层、激活层、池化层。接下来将会对这三个基础组件进行详细介绍。

### 2.2.1 卷积层

卷积层是 CNN 的核心，每个卷积层都包含一些可以动态学习的卷积核，对数据的特征提取通常是由这些卷积核通过对输入的数据进行卷积运算来完成。卷积核大小通常为 3×3 或 5×5 的矩阵，当数据到达一个卷积层时，该层将使用卷积核与输入数据进行计算，从空间角度来看，可以将卷积核的运动轨迹看作是在输入数据平面上进行逐行滑动，卷积核每移动一次，就与当前所对应的局部输入

数据相乘，最终计算出此卷积核提取的特征图，一个卷积核可以计算出一个对应的特征，但一个特征所得到的任务信息远远不够，因此需要使用多个卷积核来提取多维度的特征，当有多个卷积核时，每个卷积核都可以提取出一个不同的维度特征，即卷积核数等于特征维度数，从而获得更丰富的特征信息。卷积层的参数包括卷积核矩阵内的权重和偏置项。权重可以学习到图像中的不同特征，偏置项可以对卷积计算后的结果进行偏移修正。其计算过程如图 2.1 所示，输入数据的值表示像素值，卷积核数据表示当前卷积核权重值。

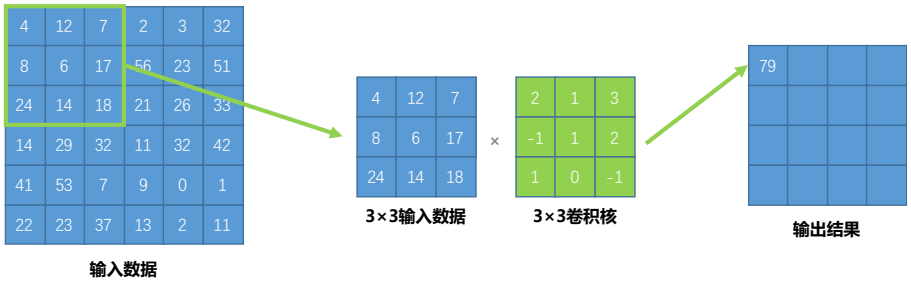



图 2.1 卷积运算

第$l$层卷积层的第$k$个特征图在坐标$(i,j)$处的特征值$z_{i,j,k}^{l}$可以用式 2.2 进行表示：

$$z_{i,j,k}^{l} = {W_k^l}^T x_{i,j}^l + b_k^l \cdots\cdots (2.2)$$

其中$W_k^l$和$b_k^l$是在$l$层中第$k$个卷积核的权重向量与偏移量，$x_{i,j}^l$是第$l$层中以坐标$(i,j)$为中心的输入块。从图 2.1 可以看出，卷积层中的每个神经元虽然要与所有数据进行计算，但每次计算时只会与一个局部区域连接，这种局部连接的机制使得卷积神经网络的参数量比全连接神经网络的参数量显著降低。卷积核在单次卷积的区域通常称为神经元的感受野，其大小与卷积核成正相关，但卷积计算会导致输出结果维度降低，为了解决这个问题，通常先对输入数据使用边缘填充法处理，这样得到的结果维度即可与输入维度保持一致。如图 2.2 所示，对于 3×3 的卷积核且步长为 1 的卷积层，如果想让输入数据与输出结果大小保持一致，需要在周围填充宽度为 1 的 0 像素值，对于填充宽度padding可以用式 2.3 进行表示：

$$\text{padding} = \left\lfloor \frac{\text{kernel}}{2} \right\rfloor \cdots\cdots (2.3)$$

其中，kernel表示卷积核大小。

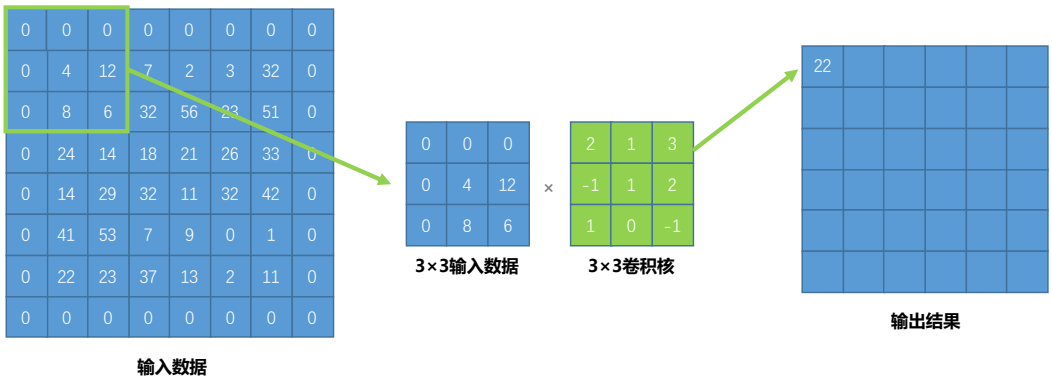



图 2.2 添加 padding 后的卷积运算

每个卷积核与输入数据进行卷积计算后都会对应一个特征图，将这些特征图进行堆叠，即可得到当前卷积层的整体多维特征输出结果。

### 2.2.2 激活层

激活层通常被放在卷积层的后面，其主要功能是对前面卷积层计算得到的结果进行非线性变换，使得网络能够学习更加复杂的非线性关系，增强网络的表示能力以及防止网络的过拟合问题。常见的激活函数有 Sigmoid 函数、ReLU 函数、Tanh 函数、Softmax 函数。

**Sigmoid 函数**

Sigmoid 函数属于非线性激活函数，常用于逻辑回归、神经网络分类等场景，可以用式 2.4 表示：

$$Sigmoid(x) = \frac{1}{1+e^{-x}} \cdots\cdots (2.4)$$

函数图像如图 2.3 所示，其优势有：输出范围为(0,1)，将接近 0 或 1 的结果看作是“否”或“是”，可以有效处理二分类问题；无论输入值的范围是多少，过大或过小，Sigmoid 的输出也只是接近于 1 或接近于 0，可以防止梯度爆炸问题，且能够有效地将非线性的输入转化为线性输出，把输入数据的不同范围缩放到同一范围，方便后续处理；此外，Sigmoid 函数在$x \in (-\infty, +\infty)$范围内均可导，易于梯度下降计算。但由于导数在$x$趋于无穷时很小，在深层网络中可能会导致梯度消失。

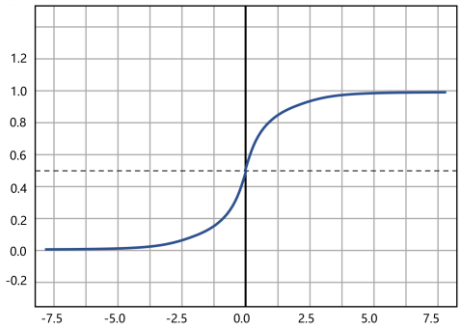



图 2.3 Sigmoid 函数图像

**ReLU 函数**

ReLU[32]函数属于非线性激活函数，常用于图像分类、语音识别、自然语言处理等场景，可以用式 2.5 进行表示：

$$f(x) = \max(0, x) \cdots\cdots (2.5)$$

ReLU 函数的优势如下：ReLU 函数为线性计算，相较于需要指数计算的激活函数计算量明显降低，例如 Sigmoid 函数，在进行反向传播计算梯度时，不但涉及指数计算，还涉及除法，计算量会更大，而使用 ReLU 激活函数可以大大减少计算量；此外，Sigmoid 函数在反向传播时的导数会接近于 0，容易出现梯度消失的问题，特征传递到深层网络时，浅层提取到的有效特征的权重相较于深层网络提取到的特征权重占比极小，导致信息丢失，难以完成深层网络的训练。ReLU 在计算时，使小于 0 的输入的输出结果为 0，造成了网络的稀疏性，由于部分结果为 0，也就减少参数的相互依存关系，同时可以减轻模型发生过拟合的问题。其函数图像如图 2.4 所示。

对于$x < 0$的情况，使用 ReLU 计算的输出结果如果永远为 0，会导致一些神经元永久性“坏死”。这种情况虽然满足了稀疏性，但也意味着它对应的权重参数不会再更新，也就是不再会对其他网络层产生贡献。为了既保持网络的稳定性与训练效率，同时又能避免神经元梯度消失，Leaky ReLU[33]对$x < 0$时的函数进行改进，使其斜率大于 0，可以用式 2.6 进行表示：

$$f(x) = \max(\alpha x, x) \cdots\cdots (2.6)$$

$\alpha$是一个小于 0 的系数，通常设置为 0.01，对小于 0 的输入可以得到一个非 0 输出，这样可以允许更多的非线性信息流动，避免神经元坏死，提高网络的表示能力。

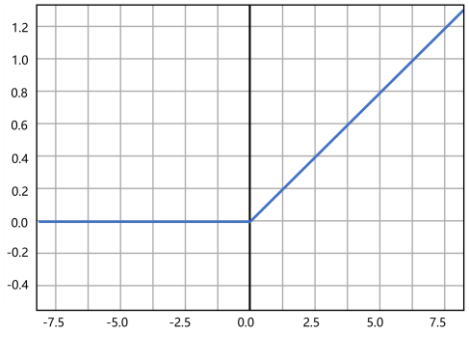



图 2.4 ReLU 函数图像

**Tanh 函数**

Tanh（双曲正切）函数常用于循环神经网络[34]（RNN）中，例如长短时记忆网络[35]（LSTM）和门控循环单元[36]（GRU），可以用式 2.7 进行表示：

$$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \cdots\cdots (2.7)$$

它可以把任意值限制在-1 和 1 之间，将其放在深度网络中的隐藏层能够解决梯度消失或梯度爆炸的问题，有助于模型收敛。

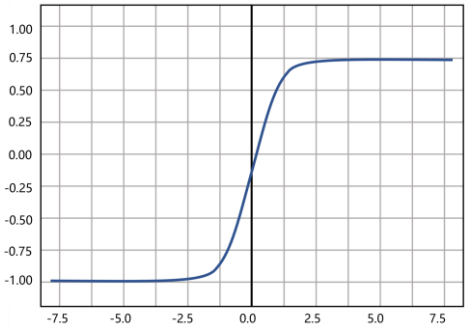



图 2.5 Tanh 函数图像

**Softmax 函数**

Softmax 函数输出各类别概率的和为 1，因此常用于多分类问题，可以用式 2.8 进行表示：

$$S(y_i)=\frac{e^{y_i}}{\sum e^{y_j}} \cdots\cdots (2.8)$$

其中，$y_i$为输入值，$S(y_i)$为输出值。Softmax 函数将输入值映射为概率值，并将所有输入值的和归一化为 1。

### 2.2.3 池化层

池化层通常放在两个卷积层之间，其作用是减小图像的空间尺寸并使网络更加健壮，同时增强了卷积层中检测到的特征的鲁棒性，并且能够进一步降低模型的参数量和计算复杂度。池化层的计算可以用式 2.9 进行表示：

$$y_{i,j,k}^{l} = pool\left(a_{m,n,k}^{l}\right), \forall (m,n) \in \mathcal{R}_{ij} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2.9)$$

$\mathcal{R}_{ij}$是特征图坐标$(i,j)$周围的局部邻域，$pool()$表示池化函数，池化层对输入中的每个局部区域进行操作，在此局部区域选择并保留一个特征值，从而降低输入数据的分辨率，例如最大池化是在区域内保留一个最大值，平均池化则计算并保留区域内所有值的平均值。在大多数卷积神经网络中，其核函数的维度为 2×2，并在空间维度以步长为 2 在输入数据上进行移动。这将特征图缩放到原始尺寸的 25%，但保持特征图的数量不变。

## 2.3 元学习理论

元学习与普通的监督学习的异处在于，其核心思想不仅仅是让模型针对指定任务达到最优解，而是让模型学会学习，即让模型像人一样拥有学习能力，能够在已有的知识基础上，快速学习指定的新任务，例如让一个分类模型在只能判断图片是否为猫的情况下，拥有迅速学习判断其他物体的能力，所以元学习也是解决小样本问题的常见方法之一。按照 Vinyals[37]的分类方法，可以将元学习划分为三种不同类别：基于度量的元学习技术；基于模型的元学习技术以及基于优化的元学习技术，以下将对这三种元学习技术进行详细介绍。

### 2.3.1 基于度量的元学习技术

基于度量的元学习技术的目标为获取元知识ω，ω通常由特征向量进行表示，由于其特征向量能够较好地表示所学的知识，因此在面对各种新任务时，都可以使用ω进行表示。在神经网络中，特征向量就是常说的模型权重$\theta$，基于度量的元学习技术将新的输入样例与元学习模型内已知标签的样例进行比较，进而学习新的任务。随后计算两者的相似度，新输入与已有样例的相似度越高，那么新的输入与样例输入具有相同标签的可能性越大。

基于度量的技术的“学习”不同于以往需要修改权重的训练，基于度量的技术在学习新任务时，网络不会做任何变化，其思想是通过简单地比较验证数据和训练数据，预测符合于训练数据的标签，在任务层面上进行非参数“学习”，学习一组任务之间的相对距离来学习任务的关系，根据相似性分数计算类预测。网络不针对新的任务信息进行训练和调整参数，取而代之的是学习一个相似度核或注意力机制$k_\theta$，其输入数据为$x_1$和$x_2$，输出是$x_1$与$x_2$的相似度得分，分数与相似度成正相关。然后，通过比较$x$和样例输入$x_i$，其中$x_i$的真实标签$y_i$是已知的，可以对新输入$x$进行类预测，即$x$和$x_i$的相似度越大，$x$有标签$y_i$的可能性越大。给定一个任务$\mathcal{T}_j = (D_{\mathcal{T}_j}^{train}, D_{\mathcal{T}_j}^{test})$和一个未知输入$x \in D_{\mathcal{T}_j}^{test}$，那么就可以通过相似度核$k_\theta$计算或预测类$Y$上的概率分布，并得到支撑集$D_{\mathcal{T}_j}^{train}$的标签的加权组合，可以用式 2.10 进行表示：

$$p_\theta(Y|x, D_{\mathcal{T}_j}^{train}) = \sum_{(x_i,y_i)\in D_{\mathcal{T}_j}^{train}} k_\theta(x, x_i)y_i \quad \cdots\cdots\cdots\cdots\cdots\cdots (2.10)$$

孪生网络[38]标志着基于度量的深度元学习技术在小样本学习环境中开始使用，他们首次使用比较支持集和查询集的输入来预测类别的思想，这一思想在图神经网络[39]中再次得到了应用。对于图神经网络，支持集是图的一部分，它代表图的邻接关系，例如，一个节点与其他节点之间的关系；查询集是图的另一部分，代表节点上的特征，例如，节点上的颜色、大小等。图神经网络通过计算每个节点的向量表示，以反映图结构的信息流参数。有了这些参数，图神经网络就可以通过梯度下降等方法进行训练。匹配网络[40]基于孪生网络出现，它们的核心思想相同，都对输入进行比较，然后生成预测结果。不同点在于匹配网络可以在小样本情况下进行训练，并且在相似度函数的使用上，不再使用距离度量函数，而是使用余弦相似度。因此，匹配网络删除了孪生网络的辅助二分类任务，作为替代添加了匹配模块，这个模块用于评估支持集和查询集之间的关系，从而计算出查询集与支持集之间的相似性分数。此外，匹配网络使用三元组损失进行训练，如果正例和负例之间的距离不够大，则会对网络进行惩罚；孪生网络使用对比损失进行训练，如果相似图像之间的距离不够小，则会对网络进行惩罚。一般来说，匹配网络相较于孪生网络更灵活，可以处理更复杂的比较任务，孪生网络更简单，训练速度更快。原型网络[41]提出了类原型（Class Prototype）的概念，类原型是

一个代表所有类别的向量，这个向量通常是每个类别样本的平均值。类原型表示了整个类别的中心，在测量样本与类别中心之间的相似度时，需要将每个输入的查询集与类原型进行比较，而不是与单个支持集实例进行比较，增加了比较鲁棒性的同时还降低了计算复杂度，将单个查询集所需的比较次数从$kN$减少到$N$。关系网络[42]使用神经网络替代匹配网络和原型网络中的相似性度量，并添加了关系信息，更加关注于图上的关系。最后，循环比较器[43]不再比较整个输入，而是采取了一种更符合生物学的方法，对输入的不同部分进行交错观察。

这些基于度量的技术的主要优点是：基于相似性预测的基本思想与概念比较简单；由于网络不需要对特定任务进行调整，因此当任务较小时，测试的速度会很快。然而，由于此方法无法将新的任务信息吸收到网络权重中，当元测试任务与元训练时的任务相似度低且距离更远时，网络性能也可能会下降。此外，当任务变大时，对输入成对比较可能会花费更大的代价。

### 2.3.2 基于模型的元学习技术

基于模型的元学习拥有一个可以面对不同情况的自适应内部状态，显然在面对不同任务时都需要依赖这个自适应内部状态，并不像基于度量的方法使用固定的神经网络。其内部维护一个任务的有状态内部表征，当输入一个任务时，模型将对任务支持集按照时间顺序进行处理。因此，对于每个任务来说，模型的内部状态都可以获取到特定的相关信息，并以此来对新的输入做出预测。由于预测结果是基于一个对外隐藏的内部动态表征，因此基于模型的元学习技术也叫黑箱模型。此外，为了让模型能够记住先前输入的信息，通常需要带有一个内部或外部的记忆模块。

基于模型的元学习技术会计算新输入$x$的类概率分布，可以用式 2.11 进行表示：

$$p_\theta(Y|x, D_{\mathcal{T}_j}^{train}) = f_\theta\left(x, D_{\mathcal{T}_j}^{train}\right) \cdots\cdots\cdots\cdots\cdots\cdots (2.11)$$

其中$\mathcal{T}_j$表示任务，$D_{\mathcal{T}_j}^{train}$是任务$\mathcal{T}_j$的支持集，$f$表示网络模型，$\theta$即为模型的参数。

最早出现的基于模型的元学习技术是记忆增强神经网络[44]（Memory-augmented neural networks，MANNs），其思想是按照顺序将整个支持集输入到

模型中，然后利用模型的内部状态对输入的查询集进行预测。基于模型的元学习的优势是系统内部的灵活性，与大多数基于度量的元学习相比泛用性更强。在监督学习任务中，基于模型的技术往往比基于度量的技术性能更高，但当数据集更大时，性能会稍逊一些，对于相关性稍低一些的任务来说，其泛化性也会比基于优化的技术低一些。

### 2.3.3 基于优化的元学习技术

基于优化的方法与前两种方法的学习角度不同，其通过显示的优化来实现快速学习。对于基于优化的元学习，可以将其抽象成一个双层优化问题。从内层来看，模型通过使用各种优化策略来对指定的任务进行更新，例如梯度更新策略。从外层来看，不同任务的性能都可以得到优化。可以用式 2.12 进行表示：

$$p(Y|x, D_{\mathcal{T}_j}^{train}) = f_{g_{\varphi\left(\theta, D_{\mathcal{T}_j}^{train}, \mathcal{L}_{\mathcal{T}_j}\right)}}(x) \quad \cdots\cdots (2.12)$$

其中$f$表示学习者，$g_{\varphi}$表示已经训练好的优化器，可以通过支持集$D_{\mathcal{T}_j}^{train}$与损失函数$\mathcal{L}_{\mathcal{T}_j}$在指定任务上对学习者的参数$\theta$进行更新。

## 2.4 本章小结

本章对图像去噪和卷积神经网络的相关知识进行了详细介绍，图像去噪的相关知识包括噪声模型、图像去噪算法，卷积神经网络的相关知识包括卷积计算与卷积神经网络模型的层次结构与功能。此外还对基于度量、模型、优化的三种元学习技术与理论思想进行了介绍。第 3 章与第 4 章将对于卷积神经网络结构以及与基于优化的元学习的结合训练进行详细介绍。

# 第 3 章 多尺度两阶段图像去噪网络

在深度学习图像去噪方面，DnCNN 与 UNet 都能够作为去噪网络来对图像进行去噪处理。然而受制于感受野、特征表征范围有限等原因，其去噪性能仍然有限。因此，为了提高去噪网络的去噪性能，本章将基于 UNet 进行模型改进，包括扩大网络感受野，对图像进行多尺度特征信息提取与多尺度特征信息融合，使得网络能够获取更多特征信息，从而提高去噪网络的性能。

## 3.1 去噪网络模型结构

图 3.1 中展示了本文去噪网络模型结构，其去噪过程分为两阶段，两个阶段各自在不同的特征尺度上进行去噪。在第一个阶段中，使用空洞卷积[45]来提取高尺度下的图像噪声的浅层特征，并经过第一阶段的 UNet[11]子网络得到第一阶段结果。在第二阶段中，使用普通卷积提取低尺度下的图像噪声的浅层特征，并对第一阶段结果进行多尺度特征融合，在融合过程中使用空洞卷积扩大感受野，将高低尺度下的图像噪声浅层特征充分融合后，再经过第二阶段的 UNet 子网络，得到最终去噪结果。

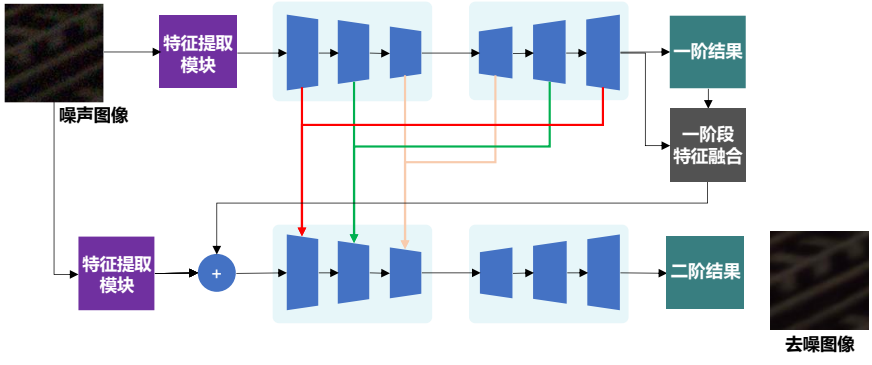



图 3.1 多尺度两阶段图像去噪网络模型整体架构

### 3.1.1 残差网络模块

去噪网络的性能与去噪网络能够获取到的特征信息数量之间的关系密不可分，而在深度学习中，在深层网络里想要尽可能保留更多信息，尤其是浅层网络提取到的信息，那么就需要使用到类似残差网络的思想。残差网络（Residual Network，ResNet）通常由数个残差块组成，残差网络与普通网络的结构与区别如图 3.2 所示。残差网络通过使用残差连接的思想，在提取特征的同时保留原始

输入特征，可以有效解决深度网络梯度消失或梯度爆炸的问题。残差网络在每一层输出时，都可以视作对前一层的输入加强，而非直接进行拟合。对于输入数据$x$的残差映射，可以用式 3.1 进行表示：

$$y = H(x) + x \cdots\cdots (3.1)$$

其中$H(x)$可以视作图 3.2 中网络层所代表的非线性计算层。残差网络的核心思想就是将输入$x$通过一个非线性层，例如卷积操作，得到一个非线性输出，随后再将$x$短连接到这个非线性输出上。这样可以在提取$x$的特征同时又有效保留上一层的输入特征，进而避免网络过深时导致的梯度消失。本文后续在 3.1.2 节使用到的残差通道注意力模块[46]与 3.2 节的特征融合中皆使用到这种残差连接思想来训练深层网络。

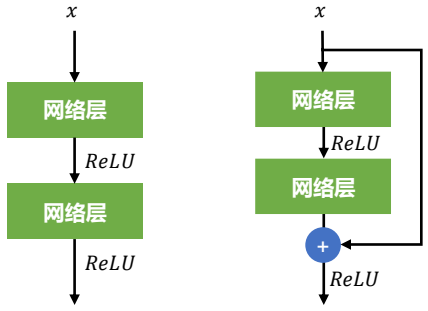



图 3. 2 普通网络与残差网络结构

### 3.1.2 残差通道注意力模块

一张带有噪声的图片往往有的区域是噪声，而其他区域不是噪声，通过卷积核提取出的多维特征图中，不同的通道也代表着图像所包含的不同的特征，残差通道注意力模块可以对特征图的不同通道分别赋值不同的权重，从而强调了关键信息，进一步提高了网络的注意力效果。通过学习针对特定通道的不同权重，可以一定程度上解决卷积神经网络在颜色空间上不敏感的问题。残差通道注意力模块（Residual Channel Attention Block，RCAB）可以看作是基于残差网络的扩展。利用残差通道注意力模块可以让网络更加充分利用不同通道之间的关系，更好地识别图像中的特征，并且能够更加适应不同的输入图像，让模型尽可能地去提取出那些含有噪声的特征，从而提高网络的性能。

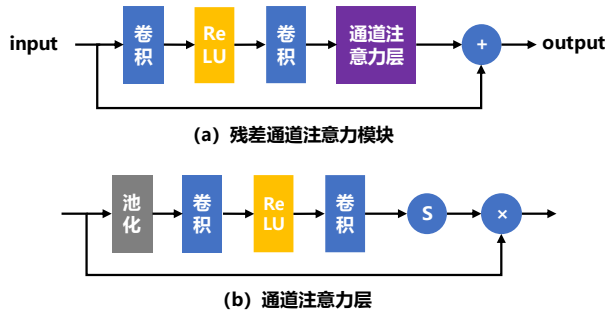



图 3.3 残差通道注意力模块内部结构

如图 3.3 所示，残差通道注意力模块是基于残差网络的扩展，其内部结构与残差网络类似，区别在于在非线性输出后，又添加了一个通道注意力层共同组成非线性输出层，随后再与原始输入进行短连接，也正是通过通道注意力层来完成前文所提到的对特征图不同通道赋值权重的功能。通道注意力层内部先对数据进行平均池化，卷积，随后使用 Sigmoid 激活，最终再将结果与原始输入数据相乘得到最终结果。因此，可以基于残差网络（式 3.1），用式 3.2 和式 3.3 表示出残差通道注意力模块：

$$H(x) = CA(conv(Relu(conv(x)))) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (3.2)$$

$$CA(x) = sigmoid(conv(Pklu(conv(pool(x))))) \otimes x \cdots\cdots\cdots\cdots (3.3)$$

其中式 3.2 的$H(x)$表示前文式 3.1 所提到的非线性层，$CA$表示通道注意力层。

### 3.1.3 基础网络模块

模型两阶段中都使用了基础网络模块 UNet，UNet 最早是在解决医学影像分割任务中提出的，可以应用于像素级图像分割任务。去噪任务亦属于精细的像素级图像重建任务，UNet 作为像素级重建网络自然很适合应用于图像去噪任务中。其结构如图 3.4 所示。

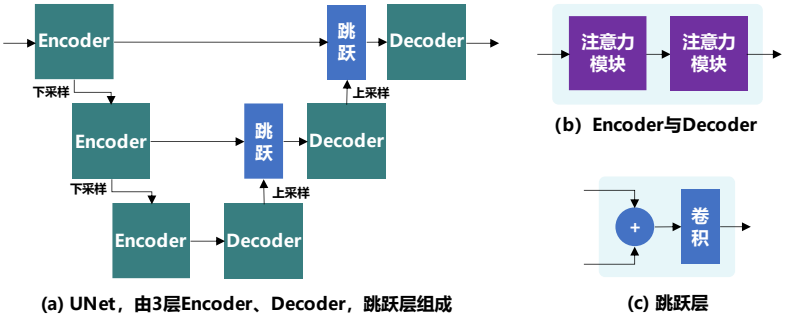



图 3.4 UNet 内部结构

UNet 宏观上来看由 Encoder 与 Decoder 组成，能够分别完成图像特征提取与重建任务。对于 Encoder 来说，数据从上层到下层需要不断的下采样，随着层数加深，提取到的特征分辨率逐渐降低。Decoder 从下层到上层则需要对 Encoder 提取到的低分辨率特征图增加分辨率，即进行上采样操作。但从低分辨率生成高分辨率图像在一定程度上属于无中生有行为，如果没有足够的辅助信息将无法保证上采样的正确性。因此，为了获取有效的辅助信息，UNet 在 Encoder 与 Decoder 之间加入了跳跃连接。当前层 Decoder 虽然没有高分辨率特征，但其上一层对应的 Encoder 拥有高分辨率特征，因此使用跳跃连接将当前层的 Decoder 上采样后的特征与上一层 Encoder 提取到的特征进行拼接，得到更充分的原图像高分辨率信息。跳跃连接能够在最大程度上保留网络在深层与浅层提取到的图像噪声特征，为 Decoder 重建高分辨率图像提供一定的辅助信息。

本文使用了 3 层 UNet 模型，同时对于 Encoder 与 Decoder 亦做了改进，将原本 Encoder 与 Decoder 中两个串联的普通卷积层替换成两个串联的残差通道注意力模块，用于提高提取和重建噪声特征的性能表现。

### 3.1.4 特征提取模块

本文提出的模型的两个阶段需要不同尺度卷积核提取到的特征信息作为 UNet 网络的输入，因此在输入前需要对图像进行不同尺度的特征提取。特征提取模块的目的是通过不同尺度的卷积核，从图像中提取出不同尺度的噪声特征信息，以便模型后续的噪声特征重建时使用。在将数据输入到 UNet 之前，对原始噪声图像进行不同尺度的空洞卷积，可以提取出不同尺度的多维特征信息，在不同尺度上完成对噪声的浅层信息特征提取。这样可以使模型更好地适应不同的输入，提高模型的鲁棒性。特征提取模块结构并不复杂，内部由卷积层与残差通道注意力模块组成，其详细结构如图 3.5 所示。

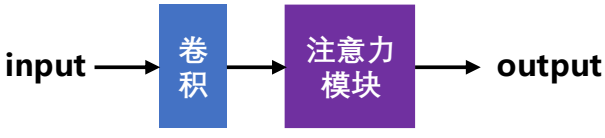



图 3. 5 特征提取模块内部结构

对于特征提取模块，可以用式 3.4 进行表示：

$$F_i = RCAB\left(conv_{d=i}(x)\right) \cdots\cdots (3.4)$$

其中$x$是噪声图片，$F_i$是特征提取后得到的浅层噪声特征，$conv_{d=i}$表示空洞值$d=i$的卷积运算。

## 3.2 特征融合

对于神经网络模型来说，通常情况下，网络的深度越深，提取到的特征也就越多。但随着网络加深的过程，会导致在浅层网络提取到的特征在向深层传递时不断减少，进而消失。同样，在不同阶段的生成网络中，如果仅仅将上一阶段的最终结果作为下一阶段的输入去传递给下一阶段，也会导致浅层网络提取到的特征无法有效地传递到深层网络，因此需要通过一些方式来尽可能地避免这种梯度消失问题。

所以在本文提出的模型中，使用了特征融合的方法，将不同阶段且不同尺度的噪声特征进行融合，这种做法可以在一定程度上增加模型的鲁棒性，因为不同的特征可能会在不同的情况下发挥作用。此外，特征融合也可以使深层网络得到信息更加丰富的特征图，尽可能地避免因为网络层数过多而导致深层网络特征减少的问题，也可以增加模型的泛化能力，提高其在新数据上的表现。

### 3.2.1 多尺度特征提取

合成噪声是人工构造的，符合先验噪声模型，其噪声按照指定的概率分布，通常是通过数学方法添加到图像上的，而真实噪声是在采集图像时产生或物体本身因磨损而产生的。因此，合成噪声的特征通常是确定的、可控的，而真实噪声的特征则更为复杂、不确定，其分布并不均匀，可能在图像中某些地方的分布较为密集，其他地方的分布又稀疏一些，所以真实噪声相较于合成噪声更难去除。

在图像处理中，图像的上下文信息对图像重建而言较为重要，那么，扩大感受野即可有效获取更多的上下文信息，对于真实噪声来说，扩大感受野也能够更好的捕捉到真实噪声的特征。在卷积神经网络中可以使用两种方式获取更大的感受野，第一种就是使用更大的卷积核来提取图像特征，因为卷积核变大，那么一次卷积运算所需要的数据在空间维度上也会变多，使感受野变大，但这样也会增加模型的参数量与计算量；另一种方式就是使用空洞卷积[45]。空洞卷积通过控制卷积核的内部扩展率来增大感受野范围，且不增加参数量，如图 3.6 所示，在保持$3\times3$大小的卷积核不变时，普通卷积的感受野为$3\times3$，空洞值为 2 的空洞卷

积感受野为$5 \times 5$。空洞卷积在与普通卷积参数量相同时，可以拥有更大的感受野。

同时为了让模型能更好地泛化到去除真实噪声任务上，本文通过使用不同大小的感受野来对图像的噪声进行特征提取。在不同尺度下提取噪声特征能够更好地捕捉图像中不同噪声在不同尺度下的特征，同时让模型适应不同尺度、不同类型的图像噪声，提高模型在面对不同分布规律的噪声去噪时的性能与鲁棒性，使模型对未知分布规律的去噪任务具有更好的泛化性。

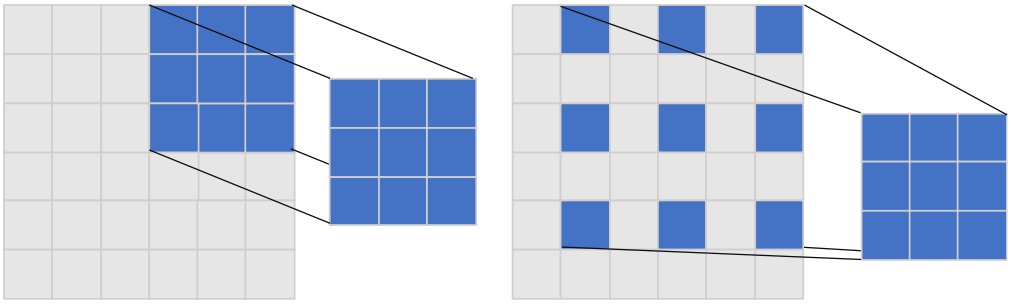



图 3.6 普通卷积与空洞卷积

### 3.2.2 一阶段特征融合

一阶段特征融合在第一阶段的尾部进行，为了避免上一阶段第一层 Decoder 重建噪声特征丢失，在此处需要将第一阶段的阶段性去噪结果与 Decoder 重建噪声后的特征先进行一阶段特征融合，将特征融合结果作为下一阶段的输入源之一。

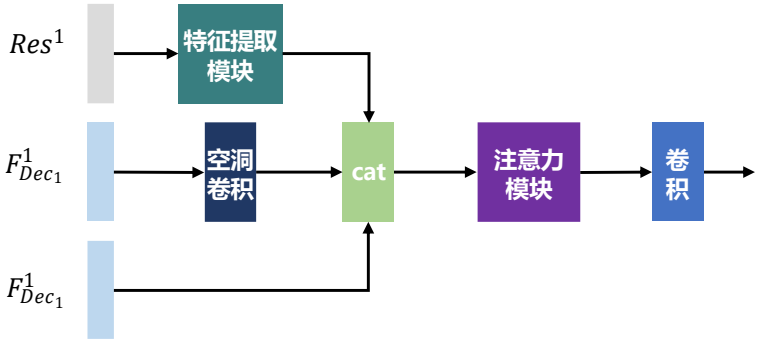



图 3.7 一阶段特征融合

在本文提出的模型中，第一阶段特征融合的详细过程如图 3.7 所示。可以用式 3.5 进行表示：

$$F_s^1 = RD([RCAB(Res^1), conv_{d=1}(F^1_{Dec_1}), F^1_{Dec_1}]) \cdots\cdots\cdots\cdots (3.5)$$

其中$F_S^1$表示第一阶段特征融合的结果，$F_{Dec_1}^1$表示第一层 Decoder 重建后的结果，$Res^1$表示第一阶段去噪后的结果图像，$RD$表示对拼接的特征进行进一步特征提取与降维。这里需要将第一阶段的去噪结果$Res^1$与第一层 Decoder 输出的噪声特征$F_{Dec_1}^1$进行特征融合。由于$Res^1$是输出图片，所以需要先对其按照噪声图片输入进行处理，即先对$Res^1$进行特征提取得到噪声特征。对于$F_{Dec_1}^1$，不仅需要其原本的特征，还需要使用更大的感受野在$F_{Dec_1}^1$上提取到高尺度噪声特征。最后，将$Res_1$的特征、$F_{Dec_1}^1$的特征与$F_{Dec_1}^1$使用拼接的方式合在一起，再使用注意力模块来进一步提取噪声特征。

### 3.2.3 两阶段特征融合

在神经网络使用不同感受野对噪声特征提取的情况下，网络会提取出不同尺度的噪声特征。第一阶段是在高尺度下提取到的噪声特征，第二阶段是在低尺度下提取到的噪声特征，两个阶段模型可以学习到不同尺度的不同特征。因此在第二阶段对噪声特征重建时，为了避免高尺度噪声特征被遗漏，采用类似于跳跃连接的方法，即将第一阶段 UNet 每层特征保存下来用以第二阶段做特征融合。通过两阶段特征融合将第一阶段 UNet 不同层级提取到的高尺度噪声特征与第二阶段 UNet 不同层级提取到的低尺度噪声特征进行融合，使得第二阶段编码器融合第一阶段编码器和解码器之间的多尺度信息流，得到兼具高低尺度噪声的特征。

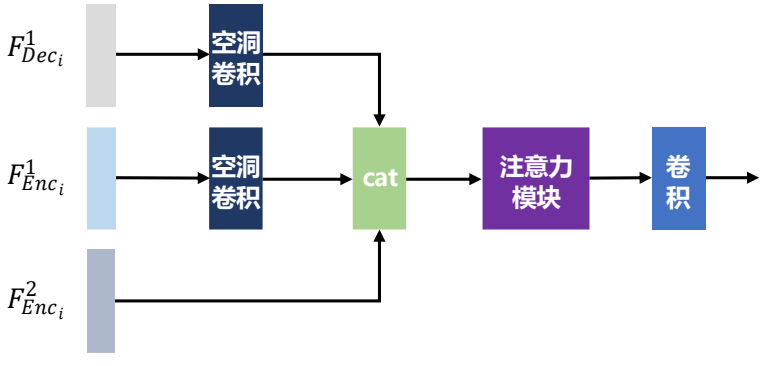



图 3.8 二阶段 UNet 特征融合

如图 3.8 所示，在第二阶段 Encoder 下采样之前，既需要第二阶段第$i$层的$F_{Enc_i}^2$特征，还需要融合第一阶段的第$i$层 Encoder 特征$F_{Enc_i}^1$与 Decoder 特征$F_{Dec_i}^1$。

为了保留特征且能进一步提取有用的噪声特征，这里先对$F_{Enc_i}^1$与$F_{Dec_i}^1$使用空洞卷积在更大感受野上进行初步的高尺度特征提取，随后与$F_{Enc_i}^2$进行拼接，得到不同尺度下的噪声特征。由于三个特征在拼接后，需要降低维度，保留那些真正有用的特征，但直接通过一层卷积降维又会在降维过程中很难提取出那些最有用的特征信息，导致效果不够理想，使特征融合的效果降低。基于此问题，本文在降维前使用注意力模块为拼接好的高维特征添加注意力机制，进行第一遍针对不同维度中不同尺度特征的提取，然后再使用一层卷积进行降维。这样可以有效地将不同尺度的噪声特征融合在一起而不增加维度。特征融合后的第二阶段第$i$层$Encoder$的特征结果可以用式 3.6 进行表示：

$$F_{Enc_i}^2 = conv_{d=1}(RD([F_{Enc_i}^2, conv_{d=2}(F_{Enc_i}^1), conv_{d=2}(F_{Dec_i}^1)])) \cdots\cdots (3.6)$$

其中$F_{Enc_i}^1$与$F_{Dec_i}^1$表示第一阶段中第$i$层 Encoder 与 Decoder 提取的特征，$F_{Enc_i}^2$表示第二阶段中第$i$层 Encoder 提取的特征，$conv_{d=i}$表示空洞值$d = i$的卷积运算，$RD$表示对拼接后的特征进行特征提取与降维。

## 3.3 实验结果分析

为便于实验说明，对于本文提出的多尺度两阶段图像去噪网络模型简称为 MTDNet。

### 3.3.1 数据集

为了方便实验进行和结果比较，并避免数据来源不确定导致的结果无法可靠验证的问题，本文选择 SIDD 公开数据集与甲骨文数据集作为真实噪声数据集。

**SIDD 数据集**

SIDD[47]（Smartphone Image Denoising Dataset）是一个用于图像去噪评估的数据集。本文使用的训练数据是 SIDD-Medium 数据集，共有 160 个不同场景的照片，其中每个场景有两张不同亮度、噪声的图片，共计 320 组图片对，并将每张图片裁剪至 256×256；测试数据是 SIDD 提供的 benchmark 测试集，共有 40 个不同场景的照片，其中每个场景有 32 张大小为 256*256 的图片，共计 1280 组图片对。

**甲骨文数据集**

甲骨文数据集的干净图片均由人工进行标注，共包括 900 组噪声-干净图片对。由于甲骨文文字图像大都是高度大于宽度，大致比例为 2:1，且分辨率较低，如果在训练数据时仍使用大小为 256×256 的切片可能会导致切片只保留了半个文字，从而导致切片内失去文字的完整语义，为了尽可能让训练数据保留完整的语义，本文在训练甲骨文数据集时选择在图像正中间裁剪出 256×512 大小的长方形，可以将大部分文字完整容纳进去，对于宽度不足 256 像素或高度不足 512 像素的图片，使用 0 像素值对空缺处进行填充。

### 3.3.2 损失函数

损失函数是一种评估算法对数据集建模好坏的方法，当需要判断模型的预测结果与真实值是否一致时，可以使用这种函数来计算模型的预测值与真实值之间的距离，即预测值与真实值的误差。在深度学习神经网络中，通常会希望模型的预测值能够最接近真实值，那么通过最小化损失函数计算到的误差值，即可让模型的预测结果逐渐接近真实值，这个误差值也被称为损失。通常情况下，损失越小，预测值与真实值的距离也就越近，从而说明预测结果更加准确。CNN 经过不断地训练，以损失函数为优化目标，最终收敛，进而达到最高性能。损失函数有很多种，显然，不同类型的模型一般来说也需要使用不同种类的损失函数。通常可以将损失函数分为回归与分类两大类。其中分类任务只关注模型能否正确分类的能力，只需计算待分类样本中共分对和分错多少样本。回归任务则需要预测具体的数字，更关注预测值与真实值之间的距离，而非像分类任务中只关注正确或错误。去噪任务显然更关心预测值与真实值之间的距离，属于回归任务，因此下文就将对回归任务常使用的损失函数进行讨论。

**L1 损失函数**

L1 Loss 用于最小化误差，误差是真实值和预测值之间所有差的绝对值的总和。可以用式 3.7 进行表示：

$$\mathcal{L}_{pixel\_L1}(\hat{I}-I)=\frac{1}{hwc}\sum_{i,j,k}\left|\hat{I}_{i,j,k}-I_{i,j,k}\right| \cdots\cdots\cdots\cdots\cdots\cdots (3.7)$$

$\hat{I}$与$I$表示预测值与真实值，$h,w,c$表示图片的高度和宽度以及通道数。$i,j,k$表示第$k$通道上坐标$(i,j)$对应的数值。