

学校代码: 10246
学 号: 17210240129

復旦大學

硕 士 学 位 论 文 (专业学位)

基于边缘感知的深度图像重建研究及其应用

**Research and Applications on
Edge-Aware Deep Image Restoration**

院 系: 计算机科学与技术

专 业: 计算机技术

姓 名: 孔昱

指 导 教 师: 薛向阳 教授

完 成 日 期: 2019 年 11 月 13 日

指导小组成员

张玥杰 教 授

金 城 教 授

李 斌 青年研究员

目 录

插图目录	iii
表格目录	v
摘要	vii
Abstract	ix
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 本文的主要贡献	3
1.3 组织结构	4
第 2 章 研究现状	5
2.1 图像边缘检测	5
2.2 生成式对抗网络	7
2.3 图像超分辨率	8
2.4 图像去模糊	9
2.4.1 图像的非盲去模糊	10
2.4.2 图像的盲去模糊	11
2.4.3 基于神经网络的图像去模糊算法	12
2.5 本章小结	12
第 3 章 边缘感知	15
3.1 边缘检测网络	15
3.2 边缘检测网络的训练	19
3.3 本章小结	20
第 4 章 基于边缘感知的深度图像超分辨率	21
4.1 问题的描述与定义	21
4.2 超分辨率网络结构设计	21
4.3 超分辨率网络的训练	22
4.3.1 平行训练	23
4.3.2 渗透性微调	23
4.3.3 损失函数	24
4.4 实验数据的准备	24

4.5 训练环境及超参数的设定	26
4.6 实验结果	26
4.6.1 评价标准	26
4.6.2 比较实验结果	27
4.7 边缘信息的必要性研究	29
4.8 本章小结	31
第 5 章 基于边缘感知的图像去模糊	33
5.1 问题的描述与定义	33
5.2 去模糊网络结构设计	34
5.3 去模糊网络的训练	36
5.4 实验数据的准备	37
5.5 训练环境及超参数的设定	39
5.6 实验结果	39
5.6.1 评价标准	39
5.6.2 比较实验结果	40
5.7 边缘提取网络的必要性研究	41
5.8 本章小结	42
第 6 章 总结与展望	47
6.1 本文总结	47
6.2 今后工作展望	48
参考文献	51

插图目录

1-1 散焦模糊 (左) 和运动模糊 (右)	2
2-1 不同算子的边缘检测结果。从左至右分别为: 输入图像, Sobel 算子检测结果, Canny 算法检测结果	6
2-2 SRCNN 和 FSRCNN 对比 ^[1]	9
2-3 SRGAN 结构示意 ^[2]	9
2-4 SRGAN ^[2] 的结果示例。SRResNet 为不采用对抗损失的训练结果 ^[2]	10
2-5 DCNN 完整网络结构 ^[3]	11
3-1 边缘检测网络结构示意	16
3-2 VGG-16 ^[4] 网络结构示意图	17
3-3 边缘检测网络判别器结构示意	18
3-4 边缘检测网络中第 1, 3, 5 层侧输出的结果示例。从左到右从上到下分别为: 模糊图像, 第 1 层侧输出结果, 第 3 层侧输出结果, 第 5 层侧输出结果	19
4-1 级联的超分辨率网络工作流程	22
4-2 DPNet 网络结构图	22
4-3 级联的细节保留网络训练策略。红色部分为参数冻结的 DPNet	23
4-4 ICDAR17-Textline 数据集	25
4-5 RVL-CDIP 数据集 ^[5]	25
4-6 RVL-DCIP Region 数据集上 4x 放大结果。从左至右: 低分辨率图像 (4x 退化), 高分辨率图像, 双三次插值结果, SRCNN ^[6] 结果, SRGAN ^[2] 结果, 我们的结果	28
4-7 ICDAR17-Textline 数据集上 4x 放大结果。从左至右: 双三次插值, SRCNN, SRGAN, 我们的结果。〈Null〉表示该图像从 ABBYY 中没有返回结果	29
4-8 边缘信息对文档图像超分辨率的影响 (a) 低分辨率图像 (4x 退化); (b) 不使用边缘损失的超分辨率结果; (c) 使用边缘损失的超分辨率结果; (d) 原始高分辨率图像	30
4-9 不同级联结构的实验结果。从左至右: 低分辨率图像 (4x 退化); 双三次插值和 DPNet 的 2x 超分辨率结果 (2x 退化); 双三次插值, 双三次插值 (2x) + DPNet (2x), DPNet (2x) + 双三次插值 (2x), 级联的 DPNet (4x) 的超分辨率结果; 原始高分辨率图像	31
5-1 基于边缘感知的图像去模糊网络工作流程	34
5-2 去模糊网络结构示意图	35

5-3 边缘检测网络 (a) 和去模糊网络 (b) 训练流程示意	36
5-4 GOPRO 数据集 ^[7] 示例。左边为清晰图像，右边为模糊图像	37
5-5 Kohler 数据集 ^[8] 。左边为清晰图像，右边为模糊图像	38
5-6 处理后的 COCO 数据集 ^[9] 。从左至右分别为：原始图像，生成的模糊图像，边缘图像	38
5-7 Kohler 数据集上的比较结果。从左到右分别为：输入图像, Tao ^[10] 结果和 EADNet 结果	42
5-8 GoPro 数据集上的比较结果。从左到右分别为：输入, Tao ^[10] 结果和 EADNet 结果	44
5-9 EADNet 模型处理的图像块实例。从上到下分别为：模糊图像，去模糊结果，清晰图像，模糊图像的 EdgeNet 边缘检测结果，去模糊后的边缘图像，清晰图像的边缘	45
5-10 是否采用边缘信息的结果毕竟。(a) 模糊图像; (b) 没有边缘信息的去模糊结果; (c) 有边缘信息的去模糊结果; (d) 真实清晰图像	46

表格目录

3-1 边缘检测网络的感受野和步长。加粗部分连接了侧输出层	15
4-1 基于边缘感知的深度图像超分辨率网络在两个数据集上的实验结果	28
4-2 RVL-DCIP Region 数据集上不同设置的实验结果	29
5-1 比较实验结果。加粗部分为最高得分	41
5-2 不同设置下的 EADNet 实验结果。加粗部分为最高得分	41

摘要

图像重建是数字图像处理的一个经典任务。受到成像条件的限制，或者人工操作不当，以及物理信息和数字信息保存传递过程中的信息丢失，还有各个环节中噪声的引入，图像退化现象大量出现在我们的日常生活中；如何从这些退化的图像中复原出原始图像的信息是一个非常基础而重要的研究课题。典型的图像退化场景包括模糊和低分辨率。退化的图像丢失了很多必要的细节信息，不仅为人类的识别和辨认以及使用带来困难，也会影响后续的计算机处理流程，特别是计算机视觉和模式识别等方向的后续应用难以顺利进行。传统的图像重建方法常利用大量的先验信息，针对于特定的场景，对图像的退化过程进行具体建模，通过对退化模型中各个因子的假设或估计，来求解原始图像，从而完成退化图像的重建工作。也有通过学习的方式，或者利用外部数据，从额外的数据信息中选取相似的部分，来补全退化图像中缺失的信息。由于图像重建任务的病态性，图像去模糊和超分辨率等任务仍然面临着很大的挑战。

我们针对于图像去模糊和超分辨率两个具体的应用场景，探索了图像的边缘感知信息在图像重建中的作用。本文首先介绍了图像重建任务的研究背景和研究意义，并对图像的边缘检测、图像去模糊和超分辨率等领域当前的研究进展进行了介绍。然后介绍了我们研究中所使用边缘感知信息的提取方式，包括网络模型的结构以及模型的训练和推理方式。

针对文档图像的超分辨率处理，我们提出了一个基于边缘感知的深度图像超分辨率网络。基于边缘感知的深度图像超分辨率网络使用级联的方式，用多个具有较小放大倍率的基础网络构件，组成能放大任意倍率的超分辨率网络。所有的基础网络构件拥有相同的网络架构和不同的参数。网络采用两阶段的训练方式，先单独训练每个构件，再组合为完整的超分辨率网络进行整体训练。训练过程中除了使用感知损失^[11] 和内容损失外，还加入了边缘损失指导网络保留字符结构细节。实验结果表明，利用边缘信息辅助文档图像的超分辨率，能有效的提升图像超分辨率效果，同时保留文档图像中的字符结构，提升后续字符识别流程的识别精度。

针对图像去模糊场景，我们构建了一个基于边缘感知的深度图像去模糊网络。基于边缘感知的深度图像去模糊网络由两个子网络构成，其中一个是用于提取边缘信息边缘检测网络，另一个是去模糊网络。边缘检测网络将首先被训练，用于从模糊的图像中提取图像边缘，再将提取出的边缘信息与模糊图像一起输入去模糊网络训练，辅助去模糊过程。同时，边缘检测网络也被用于损失函数的计算，比较去模糊图像和原始清晰图像的边缘差异。实验结果表明，加入边缘信息的去模糊网络，去模糊效果相较于当前最先进算法有了一定的提升，同时也保留并加强了图像清晰锐利的边缘，符合人类的视觉需求。

关键字：图像处理；边缘检测；图像超分辨率；图像去模糊；计算机视觉

中图分类号：TP391

Abstract

Image restoration is a classic task in digital image processing. Limited by imaging conditions, or improper manual operation, and loss of information in the process of storing and transmitting physical information and digital information, as well as the introduction of noise in various links, image degradation phenomena appear in our daily lives. How to recover the original image information from these degraded images is a very basic and important research topic. Typical image degradation scenarios include blurring and low resolution. Degraded images lose a lot of necessary details, which not only causes difficulties for human identification and identification, but also affects subsequent computer processing, especially for subsequent applications such as computer vision and pattern recognition. Traditional image restoration methods use a large amount of prior information to specifically model the degradation process of an image for specific scenes. The original image is derived by hypothesis or estimation of factors in the degraded model, thus completing the degraded image restoration. There are also some learning-based methods, or to use external data to extract similar parts from additional data information to complement missing information in the degraded image. Due to the ill-posed nature of image restoration, tasks such as image deblurring and super-resolution still great challenges.

We explored the role of image edge perception information in image restoration for two specific application scenarios: image deblurring and super resolution. This paper first introduces the research background and research significance of image restoration tasks, and introduces the current research progress in the field of image edge detection, image deblurring and super resolution. Then introduces the extraction method of edge perception information used in our research, including the structure of the network model and the training and reasoning methods of the model. After that, for the super-resolution processing of document images, we propose a edge-aware deep image deblurring network. The edge-aware deep image deblurring network uses a cascade method to form a super-resolution network capable of amplifying an arbitrary magnification by using a plurality of basic network components having a small magnification. All underlying network components have the same network architecture and different parameters. The network adopts a two-stage training method. Each component is trained separately and then combined into a complete super-resolution network for overall training. In addition to the use of perceptual loss and content loss, the edge loss guides the network to preserve the character structure details. The experimental results show that the edge information can be used to assist the super-resolution of the document image, which can effectively improve the image super-resolution effect, while retaining the character structure in the document image and improving the recognition accuracy of the subsequent character recognition process.

For image deblurring scenes, we construct a depth image deblurring network based on edge

Abstract

perception. The edge-aware depth image deblurring network consists of two sub-networks, one for extracting the edge information edge detection network and the other for the de-fuzzy network. The edge detection network will be trained first to extract the edge of the image from the blurred image, and then input the extracted edge information together with the blurred image to de-fuzzy network training to assist the deblurring process. At the same time, the edge detection network is also used for the calculation of the loss function, comparing the edge differences between the deblurred image and the original sharp image. The experimental results show that the deblurring network with edge information adds a certain improvement compared with the current state-of-the-art algorithm, and also preserves and enhances the sharp and sharp edges of the image, which is in line with human visual needs.

Keywords: Image Processing; Edge Detection; Image Super-resolution; Image Deblurring; Computer Vision

CLC number: TP391

第1章 绪论

1.1 研究背景及意义

随着互联网世界的飞速发展，特别是移动设备和移动互联网的极速扩张，数字图像作为一个重要的信息承载者，在数字世界中扮演着越来越重要的角色。带宽的增大让人们不再满足于仅仅通过文字来传递信息，图像，包括由大量图像帧组成的视频，也在渐渐成为信息交流的主要形式。数字图像来源广泛，包括且不限于由各类摄像设备直接采集而来的照片、扫描仪器产生的扫描结果、电子设备自身的截图，甚至超声、红外等其他领域信息的成像，最终都可以以数字图像的形式存储于计算机系统中。由于成像设备的精度和像素良莠不齐，成像条件受限，或者图像压缩和信息丢失，实际获取到的数字图像有可能出现退化现象，缺失部分欲取得的目标信息，为图像的下一步使用带来影响。模糊就是图像退化中的一种典型表现形式。为此，需要一些图像处理方法，从退化的图像中恢复出原始的清晰图像信号。常用的退化图像复原方法包括提高图像分辨率的超分辨率方法，以及提高图像质量的去模糊方法。

退化图像大量存在于日常的摄影、扫描等场景下，并具有不同的表现形式。模糊是一种常见的退化形式，包括散焦模糊和运动模糊等，如图 1-1。散焦模糊主要发生于拍摄时对焦不准的情况下，当相机焦点未对准在目标物体上时，会造成目标拍摄物体因失焦而模糊，产生虚化现象，不能清晰的表现出目标物体的细节信息。另外，在图像压缩存储的过程中，或者缩小放大图像，也会造成图像全局的高斯模糊。运动模糊则主要发生在拍摄成像的瞬间，设备与物体发生相对移动的场景，如拍摄过程中手的抖动，拍摄对象的高速运动等。运动模糊主要表现为图像中物体沿其相对运动方向有残影出现，例如在拍摄翅膀高速扇动的蜂鸟时，摄像机与背景、蜂鸟身体保持相对静止，能够清晰的捕捉到蜂鸟的身形，但蜂鸟的翅膀部分则在扇动所及范围内呈现出残影效果，鸟翼拉大变形，无法真实反映其本来面目。除此之外，还有因为图像曝光时间长短不同带来的模糊效果，延时摄影中车水马龙的光流就属于这类模糊。以及拍摄或扫描过程中，由像素数和 DPI 较低，无法完整还原出真实物体原本形态，存在信息丢失所带来的模糊，和在图像数据的流通过程中，因信息压缩、缺损导致的图像失真和模糊，比如在图像的多次流通过程中，每一次传输都使用了有损的压缩算法来压缩图像信息，此时最终用户接受到的图像数据就因流通过程中的大量信息丢失而显得模糊，图像上的目标物体或文字难以被辨认。另外，为了节约存储空间，有时会将图像缩小保存，或仅保存一份缩略图副本，当把这份缩小后的存储图像取出再放大查看时，需要通过一些算法或规则来补全放大后增加的像素信息。放大过程中常用的信息补全算法是插值法，将周围已知像素点的值之差均匀分配到中间的未知像素点，保持像素值均匀变化，形成过渡带。而这个像素值均匀变化的过程，是放大后图像模糊的来源。这些全局性的图像退化现象产生了模糊的低分辨率图像。尽管目前也涌现出很多算法来补全放大后图像中缺失的信息，但是仍然未能完全解决放大后图像存在模糊的现象。



图 1-1 散焦模糊 (左) 和运动模糊 (右)

在图像处理领域，图像的超分辨和去模糊都是低级视觉问题中基础而具有挑战性的一项任务。无论对人工辨认还是模式识别来说，从模糊或者低分辨率的图像中识别一个特定的物体，或者理解图像中蕴含的某些信息，都是一个巨大的挑战。人类识别一个物体，在一定程度上首先需要根据物体的外边缘轮廓确定物体的外形形态，将该物体与背景中的诸多其他物体区分开，再通过一些细节的纹理、颜色、景深等整体或局部特征来认知整个物体。而当图像中存在模糊现象，以及图像的分辨率较低时，人们赖以辨认物体的特征信息出现了丢失缺损，在识别物体的过程中就产生了不确定性，无法正常正确的识别模糊的物体或者图像中物体之间的关系。而在模式识别领域的研究中，图像模糊带来的影响更为严重。绝大多数的模式识别方法，都是针对图像中提取出的某些特征，通常表现为多维向量或矩阵的形式，结合统计学理论，对图像或特征进行分类，从而识别图像中所包含的特定物体或文字信息。低分辨率模糊图像中，由于信息的缺失，所能提取出的特征不完备，其数学表示与清晰图像中提取出的特征可能有较大的距离。将不完整的特征代入已设计优化好的统计模型进行计算时，就会出现较大的偏差，导致分类结果出错，不能准确的识别出模糊图像包含的物体。图像的超分辨率，就是将低分辨率的图像进行放大，在重建出高分辨率的图像同时，为图像增加合理的细节信息，并且保持图像原本的结构信息不发生改变。而图像的去模糊，就是要从模糊的退化图像出发，补全图像退化过程中丢失的信息，还原出一幅清晰原始图像。由于图像退化的原因多种多样，而对同一幅退化图像，也可以对应于多幅可能的清晰图像，因此图像的重建工作变得困难而病态。

针对实际应用中不同场景、不同成因的图像复原问题，学术界和工业界都已进行了一系列的研究。多年以来，人们都在不断尝试从不同的方向，利用不同的信息，来处理某些特定场景下的图像重建，根据重建侧重点的不同，也分为了不同的领域，如针对分辨率和细节信息的图像超分辨率，以及针对模糊的图像去模糊。有对通用的图像去模糊算法的相关研究，并产生了丰厚的研究成果。从学术研究的角度而言，人们通过对不同类型模糊成因的分析，

对模糊图像的产生过程分别进行了建模，如将模糊图像视为由清晰图像和点扩散函数(模糊核函数)进行卷积计算后，再加上独立的加性噪声所产生的结果^[12]。根据模糊图像产生过程的数学模型，分别对产生模糊图像的各个因素进行求解，从而以数学推导的形式解出模糊图像所对应的清晰图像。对于模型中的各个因子，可根据假设检验或统计分析的方式，用不同的概率分布来拟合特定情况下的参数，从而构建出针对特定模糊成因的去模糊算法。针对图像超分辨率的情况，常用的方法是从内部或外部寻找具有相似性的内容和结构，再利用这些额外的信息进行补全。也有根据样例通过学习从低分辨图像到高分辨图像的映射函数，来实现图像的超分辨率。在工业界中，相关技术人员也在不断的分析模糊图像的成因，从操作流程到硬件设计等多角度来干预模糊图像的产生。日常生活中常见的三脚架，就是消除因人身体抖动而导致运动模糊的最简易的装置，除此之外，还有自动对焦、光学防抖等技术，通过软硬件结合，来修正成像过程中一些可能产生图像模糊的操作。尽管如此，图像的去模糊问题远未达到已被解决的程度，无论去模糊后图像的清晰度，新产生细节的合理性，还是算法的适用场景，都还有很大的提升空间。因此，继续进行对退化的模糊图像进行去模糊操作的相关研究仍然是非常有意义的。本文主要着眼于实现一个端到端的去模糊算法，通过分析人类从图像中辨识物体的过程，引入图像边缘感知信息辅助，利用深度学习神经网络框架，来解决包括运动模糊在内的常见模糊图像复原问题。

1.2 本文的主要贡献

在本文中，我们致力于利用边缘感知信息，来进行退化图像的重建工作。我们构建了一个边缘检测网络，用于从图像中获取边缘信息，特别是从退化的图像中检测出清晰的边缘。在边缘感知信息的基础上，我们还探索了图像超分辨率和图像去模糊这两个具体的图像重建场景里，边缘信息的应用。我们构建了一个基于边缘感知的深度图像超分辨率网络，用于对低分辨率的文档图像进行超分辨率处理，并且在边缘感知信息的帮助下，尽可能的保留文档图像中的字符细节，为后续的字符识别过程带来帮助；另外，我们还构建了一个基于边缘感知的深度图像去模糊网络，将模糊图像的边缘信息与模糊图像一起输入网络，辅助整个图像去模糊过程。我们从这个两个具体的应用场景出发，分别用不同的实验验证了边缘感知信息在图像重建过程中的作用，相较于当前最先进算法，我们所设计的基于边缘感知的深度图像超分辨率网络和基于边缘感知的深度图像去模糊网络都取得了相当甚至更优的结果。

文章的主要研究工作有以下几点：

- 1) 探索了边缘信息在深度图像重建任务中的作用，从人类视觉认知的角度出发，充分挖掘利用图像自身所蕴含的信息来辅助图像的重建过程；
- 2) 提出了一种基于边缘感知的深度图像超分辨率算法，针对文档图像的场景，用级联的方式，将多个具有较小放大倍率的细节保留网络进行集成，构建出能够乘性倍率放大低分辨率图像的文档图像超分辨率网络。同时，在现有数据集和公开文档数据的基础上，构建了两个数据集，用于文档图像的超分辨率处理场景，并在这两个数据集上进行了实验；
- 3) 提出了一种基于边缘感知的深度图像去模糊算法，使用深度卷积神经网络，通过大样本学习的方式，端到端的恢复出清晰图像，避免了传统方法中需要预先假设模糊核函数以及噪声才能求解的问题；

- 4) 加入感知损失项 (Perceptual Loss)^[11] 作为损失函数的一部分，以提高图像重建结果的真实性。感知损失由李飞飞等人提出后，被大量应用于图像生成领域，能有效提升生成网络最终输出结果的真实性和实际效果，使人类肉眼无法简单地分辨出图像来源是真实场景拍摄得出还是由算法处理而来；
- 5) 采用生成式对抗网络的思想，生成更逼真的清晰图像。补全模糊图像缺失的信息是一个病态的问题，一个确定的输入状态对应于大量的解。由于人类对真实性的判断尚无清晰的概念能够表示并转化为数学语言，如何在庞大的解空间中寻找一个最优解是生成网络需要解决的问题。生成式对抗网络通过判别器学习的方式，将真实性的判断等效为一个函数，并通过不断的迭代学习优化这个函数，使其拟合于真实的数据和人类的认知，辅助生成网络更好的从解空间中寻找出满意解。

1.3 组织结构

本文的具体工作和内容安排如下：

第一章绪论部分简要介绍了图像重建领域研究的背景和意义，针对图像超分辨率和去模糊任务进行了详细描述，同时概述了本文的主要研究内容和工作贡献。

第二章研究现状部分详细分析了当前数字图像处理中边缘检测、图像超分辨率、去模糊算法方向的研究成果，对传统方法和深度学习算法方法分别进行了讨论，分析其各自的优缺点。同时也简要介绍了生成式对抗网络的研究进展。

第三章阐明了本文边缘感知的实现方式，同时详细介绍了所设计算法的具体网络结构设计和网络整体的训练流程；

第四章设计了一个基于边缘感知的深度图像超分辨率网络，用于文档分析场景中的文本图像超分辨率处理。这一章将介绍该网络的实现细节，包括损失函数的设定和网络的训练，同时构建了两个数据集，展示了我们的网络与同期最先进算法在这两个数据集上的效果；

第五章介绍一个基于边缘感知的深度图像去模糊算法所要解决的问题，同时详细介绍了所设计算法的具体网络结构设计和网络整体的训练流程。通过在公开数据集上的实验，将本文所介绍算法与同期其他最优算法进行了比较，并说明了实验所使用的数据集和实验环境；

第六章总结了我们基于边缘感知的深度图像重建研究和，同时对将来的研究工作与研究方向进行了展望。

第 2 章 研究现状

当前，在图像处理和计算机视觉领域，对退化图像的重建任务已经有了丰富的理论研究成果。同时，在图像的边缘检测这个基础视觉任务上，也有相当多的相关研究涌现。我们所研究的内容是基于边缘感知的深度图像重建研究及其应用，在本章中我们将从图像的边缘检测出发，并分别介绍图像重建任务中图像超分辨率和图像去模糊两个领域的研究现状。同时，本文所介绍的方法采用了生成式对抗网络的架构进行网络结构的设计，本章中也将涉及生成式对抗网络的一些基本介绍。在第一个部分中，我们首先会着重介绍当前主流的图像边缘检测算法，包括传统的边缘检测算子，以及传统方法和深度学习算法两个领域中各自的相关工作。第二个部分将介绍生成式对抗网络的基本概念，简述生成式对抗网络在图像生成方面的研究进展和近年来的代表性成果。随后，我们对当前图像超分辨率和图像去模糊两个领域的工作进行了梳理，从多个角度分介绍数字图像处理中这两个领域的研究成果。

2.1 图像边缘检测

图像的边缘检测是数字图像处理中的一类经典问题。边缘是图像中物体轮廓信息和结构信息的一种体现，边缘检测就是要查找图像内对象的边界。在图像中，边缘主要表现为亮度或深度的不连续，边缘点处图像的亮度或像素值发生了剧烈变化^[13]。边缘检测是数字图像处理的基本工具，图像边缘检测的结果，作为一种基础的图像特征，对计算机视觉的各个领域都是根本而重要的，包括视觉显著性分析，图像分割，物体的检测和识别，目标跟踪与运动分析，医学成像处理，三维重建等^[14]。

最早的边缘检测方法依赖于图像的梯度变化。边缘点是图像中信号剧烈变化的点，因此可以用一阶导数来表示图像的边缘强度，梯度方向则表示边缘方向，局部梯度的最大值就是边缘点。也可利用图像的二阶导数，图像在二阶导数的零穿越点取得了最大梯度值，以此来检索图像边缘。常用的边缘检测算子就是一类基于梯度变化的边缘检测算法。根据所选取梯度方向的、一阶导数还是二阶导数，以及阈值的选取不同，简化到数字图像处理的场景下，有很多不同的边缘检测算子，如 Sobel 算子^[15]、Prewitt 算子^[16]、Roberts Cross 算子。Sobel 算子是一个差分算子，包含横向和纵向两个方向的共计两组计算矩阵，通过计算像素点与周围八个点间的梯度，来近似的表示图像中单个点的边缘强度。利用 Sobel 算子计算图像横向和纵向的边缘强度过程如下：

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * A, \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A,$$

其中 A 为原始图像， G_x 和 G_y 分别为横向和纵向的边缘强度。由此可以得到图像中单个像素点的梯度大小：

$$G = \sqrt{G_x^2 + G_y^2},$$



图 2-1 不同算子的边缘检测结果。从左至右分别为：输入图像，Sobel 算子检测结果，Canny 算法检测结果

以及梯度方向：

$$\Theta = \arctan\left(\frac{G_y}{G_x}\right).$$

设定一定阈值后，图像中梯度值较大的点即可视为图像的边缘点。同样的，Prewitt 算子可以定义为如下：

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

不同于 Sobel 算子和 Prewitt 算子，Roberts Cross 算子是两个 2×2 大小的卷积核：

$$S_x = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, S_y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Canny 边缘检测算法^[17] 是当前应用最广的图像边缘检测算法。使用 Canny 算法进行图像的边缘检测可以分为以下几个步骤：首先将图像灰度化并进行高斯滤波去除噪声；再用一阶差分算子计算图像梯度值和方向；随后对计算好的梯度值进行非极大值抑制 (non-maximum suppression) 消除误检测边缘点；最后用滞后阈值来确定边缘点，得到最终的边缘检测结果。由于图像中的边缘可能存在多种方向，因此在 Canny 算法中使用了四个边缘检测算子来分别计算图像中横向、纵向，以及两个对角线方向的梯度值，所使用的边缘检测算子可以是任意一种算子，如 Sobel 算子^[15] 或 Prewitt 算子^[16]。以使用 Sobel 算子为例，计算出横向梯度 G_x 和纵向梯度 G_y ，以及梯度值 G 和梯度方向 Θ 后，根据求得的梯度方向 Θ ，将所有点的梯度方向近似到横向、纵向，以及两个对角线方向。由 Sobel 算子等方法计算出的边缘，可能会出现由多个边缘点组成的粗大边缘，形成模糊边缘图像。Canny 算法通过非极大值抑制来处理宽边缘，根据每个像素点的梯度方向，比较沿梯度方向前后的梯度值，仅保留梯度的局部极大值点而抑制其他非极大值点，将非极大值点的梯度置为 0，从而使每条边缘都仅有一个精确的边缘点。考虑到由噪声等因素产生的离散边缘点，Canny 算法在最后阶段采用了滞后阈值来进一步区分边缘。滞后阈值包括高阈值和低阈值两个阈值，根据像素点的梯度值将所有梯度值分为三类，低于低阈值的为非边缘点，高于高阈值的为强边缘点，在高低阈值之间的为弱边缘点。弱边缘点表示该点有可能真正的边缘点，而强边缘点则是已经确定的边缘点。当一条弱边缘有与其相连的强边缘点时，这条弱边缘被保留，否则抑制这条弱边缘。最终被保留的边缘点组成的边缘图像即 Canny 算法的边缘检测结果。不同方法的检测结果如图2-1。

除了早期根据图像梯度变化设计的边缘检测算子，还有通过人工设计特征来实现边缘检测的信息论方法，如统计边缘 (Statistical Edges)^[18]，Pb 方法^[19]，gPb 方法^[20]等。另一类方法是基于学习的方法，通过算法自动学习人工设计的一些特征来实现边缘检测，包括 Dollar 等人提出的 BEL 方法^[21]，Ren 提出的多尺度方法^[22]，Lim 等人提出的 Sketch Tokens 方法^[23]和结构化边缘 (Structured Edges) 算法^[24]。

近年来也有很多基于深度学习和卷积神经网络的边缘检测算法，通过自动的多级特征学习来检测图像边缘。包括 N^4 -Fields 算法^[25]，DeepContour 网络^[26] 和 DeepEdge 网络^[27]。HED(Holistically-nested Edge Detection)^[28] 通过在 VGG-16 网络卷积层后添加侧边输出，用多级网络输出多尺度的边缘检测结果，并将多尺度的边缘图像经过加权混合成为一个混合便往，网络最终将所有中间检测结果和混合边缘求取均值作为结果输出。对于每个尺度的边缘检测，都会计算相应尺度下的边缘检测损失并反向传播，增强了最终的边缘检测效果。

2.2 生成式对抗网络

生成式对抗网络 (Generative Adversarial Network, GAN)^[29] 是 Ian Goodfellow 于 2014 年提出的一种网络结构。对抗生成网络由两个子网络组成，其中一个是用于从输入人的随机噪声中生成目标分布样本的生成器 (generator)，另一个判别器 (discriminator) 子网络则用于预测输入数据的来源。其中，生成器的任务是用给定的输入数据，通常是一个 100 维的随机噪声，经过一定的计算后，得到一个目标分布的输出；判别器的任务是判定输入数据是来自目标的真实分布的概率。基于零和博弈的思想，在训练过程中，生成器将尽可能生成判别器无法判断的数据，以最大化判别器分类失败的概率为优化目标；而判别器则需要尽可能的将生成器产生的数据与真实数据区分开，提高分类成功率。当生成器成功的模拟出训练数据的分布，且判别器完全无法区分生成数据与真实数据，输出的概率值永远为 0.5 时，可认为生成器达到了最优的状态。生成式对抗网络的损失函数如下所示：

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log 1 - D(G(z))]$$

其中 G 为生成器， D 为判别器， p_{data} 为目标真实数据的分布， p_z 为生成器输入数据的分布。则对于生成器来说，优化目标是最小化：

$$\begin{aligned} C(G) &= \max_D V(G, D) \\ &= -\log(4) + 2 \cdot JSD(p_{data} || p_g) \end{aligned} \tag{2.1}$$

由于 JS 散度是一个非负值，则当 $p_g = p_{data}$ 时， $C(G)$ 取得最小值 $-\log(4)$ ，此时达到全局最优，JS 散度在 GAN 网络中被用来衡量生成器分布和真实数据分布间的距离，且整个训练过程是可收敛的。

在最初的生成式对抗网络中使用了 KL 散度和 JS 散度来衡量生成器数据分布与真实数据间的差异，并以此为基础来优化网络。Martin Arjovsky 等人证明^[30, 31]，当 p_g 与 p_{data} 距离相差过大而几乎没有重叠时，JS 散度始终为常数 $\log 2$ ，此时生成器的梯度消失为 0，无法再进行优化，同时还会出现模态坍缩的现象。随后提出了 WGAN^[31]，将衡量分布差异的距离由 JS 散度替换为推土机距离 (Wasserstein Distance)，用一阶利普希兹连续条件 (Lipschitz continuity) 来限制判别器的梯度，解决了当 p_g 与 p_{data} 分布差异过大时梯度消失的问题。WGAN 中，所有的网络参数都被截断到 $[-c, c]$ 范围内，而在实验中发现，大多数权重都落在 $-c$ 或 c 上，

影响了深度网络的学习能力；另外，由于权重被裁剪，梯度变得容易消失或爆炸。在 WGAN 的基础上，Gulrajani 等人提出 WGAN-GP^[32]，引入梯度惩罚 (Gradient Penalty)，在损失函数中额外增加一个正则项来约束梯度，引导判别器梯度的 L_2 范数在 1 附近。WGAN-GP 在训练上的稳定性在 WGAN 基础上进一步提升，同时能生成出质量更好的图像。

最初的生成式对抗网络用多层感知机实现，随后 Alec Radford 等人提出了 DCGAN^[33]，将该框架推广到卷积网络中，获得了不错的生成效果；Mirza 等人为生成式对抗网络加入条件约束，提出了 Conditional GAN^[34]，用条件概率分布代替 GAN 网络中的概率分布，即：

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log 1 - D(G(z|y))]$$

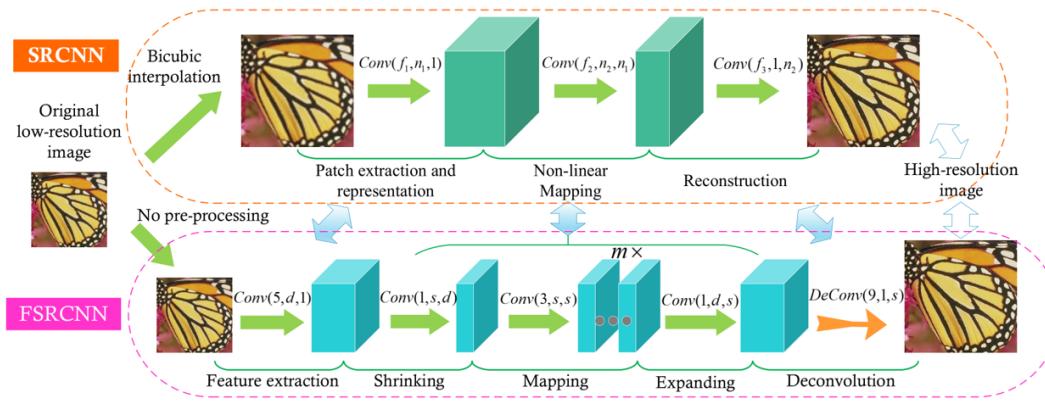
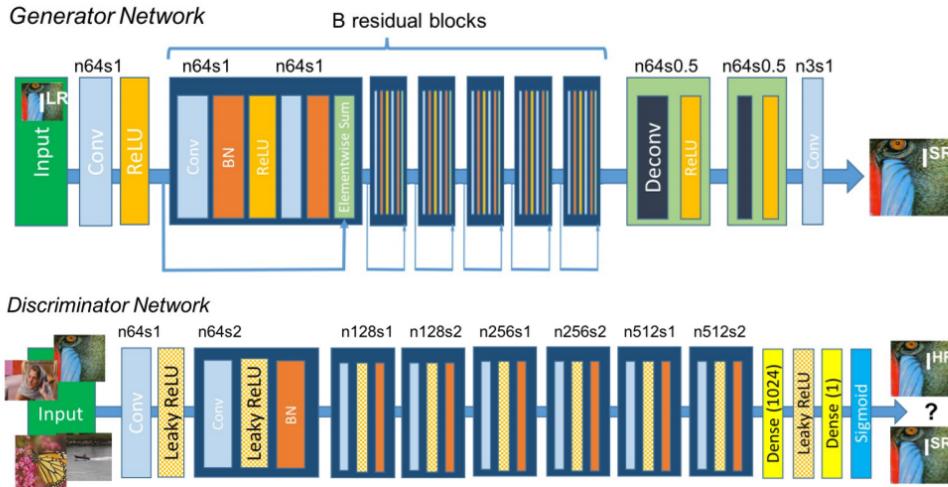
Conditional GAN 使得 GAN 能够在一定的条件约束下，根据约束类别生成对应的数据，拓展了生成式对抗网络的使用范围，在 Conditional GAN 的基础上，基于图像输入的生成式对抗网络得到了极大的发展。Phillip Isola 等人提出 Pix2Pix 网络^[35]，将生成器的输入数据由一个多维随机噪声，拓展到用一幅图像作为输入数据，用作图像翻译任务。CycleGAN^[36]、DiscoGAN^[37] 和 DualGAN^[38] 三个网络同时提出了一类成对的生成结构，实现两个不同类图像的相互翻译。以 CycleGAN 为例，网络包含两个生成器 G_X 和 G_Y ，以及两个判别器 D_X 和 D_Y 。 G_X 和 D_X 将属于 X 的图像转换到 Y，而 G_Y 和 D_Y 将属于 Y 的图像转换到 X，用无监督的学习方式实现图像域之间的转换。InfoGAN^[39] 也是一种无监督的生成式对抗网络结构。InfoGAN 利用互信息，试图让输入生成器的随机噪声具有可解释性，通过将隐变量 (latent code) 和随机噪声结合的方式，使用隐变量的不同维度来同时控制生成数据的多个属性，如用 MNIST 数据集进行训练时，用隐变量的第一维表示要生成的数字，第二维表示笔画的粗细等。

2.3 图像超分辨率

低分辨率图像是图像退化的一种常见表现形式，图像的超分辨率就是针对这种现象的一个图像复原任务。超分辨率任务的主要目的是将低分辨率的图像转换为高分辨率，同时还需要保持图像中原有的信息不被破坏，并且在恢复过程中得到加强。

图像的超分辨率是一个病态的问题。传统超分辨率方法包括线性插值、双线性插值、双三次插值等插值方法。还可以通过使用一些先验知识来解决。这些方法包括基于预测的方法^[40]，基于梯度轮廓的方法^[41]，基于图像统计的方法^[42, 43]，还有基于分块图像的模型^[44, 45]，或者利用图像的自相关性^[46] 进行超分辨率处理。另外，还有用外部学习或基于经验的图像超分辨率方法^[47, 48]。

近年来，深度学习的发展给视觉问题的解决也带来了帮助，目前也有相关研究转向用深度神经网络进行图像的超分辨重建。例如，Chao Dong 等人首先提出 SRCNN^[4] 网络结构，利用卷积神经网络进行图像的超分辨率重建。SRCNN 用三个卷积层来模拟稀疏编码 (sparse coding) 过程中的编码和重建过程，先将低分辨率图像用双三次插值放大至目标尺寸输入网络，再被网络处理成高分辨率图像输出。但由于 SRCNN 的感受野较小，网络层数少，对图像特征提取的不够，效果有限。在 SRCNN 基础上，Chao Dong 等人又提出了改进版本 FSRCNN^[1]，用反卷积代替了 SRCNN 中的双三次插值放大，并且在减小卷积核大小的同时增加了卷积层层数，加快了网络的运算速度，如图2-2。SRGAN^[2] 用生成式对抗网络进行图像的超分辨重建，如图2-3，同时用基于 VGG 网络的感知损失 (Perceptual Loss)^[11] 和均方误

图 2-2 SRCNN 和 FSRCNN 对比^[4]图 2-3 SRGAN 结构示意^[2]

差 (MSE) 构成损失函数，在生成器中使用了多个残差块结构和转置卷积，取得了很好的重建效果²⁻⁴。VDSR 网络^[49]相较于 SRCNN 增大了感受野，同时通过预测残差而非直接进行像素值的预测，来加快收敛速度。VDSR 网络将网络层数加深到 20 层，采用了自适应梯度裁剪 (Adjustable Gradient Clipping) 的方式来降低深层网络的训练难度，防止出现梯度消失或梯度爆炸。DRCN^[50] 使用深度递归卷积神经网络架构，用更少的网络参数达到了更好的效果。DRCN 用递归的推理网络 (Inference Network) 部分，将上一过程传来的特征图用同样的网络参数进行多次递归运算，每一次递归操作都将单独计算一个损失，并把每一轮递归输出的结果都输入下一个重建环节 (Reconstruction Network)。

2.4 图像去模糊

模糊图像的产生通常可建模为如下表示^[12, 51]:

$$\mathbf{B} = \mathbf{I} \otimes \mathbf{K} + \mathbf{N},$$

其中， \otimes 是卷积操作， \mathbf{B} 是生成的模糊图像， \mathbf{I} 是原始的清晰图像， \mathbf{K} 是模糊核函数， \mathbf{N} 是加性噪声。由这个建模可以得出，要从退化的模糊图像中还原出原始的清晰自然图像，必须首先知道导致模糊的模糊核函数 \mathbf{K} ，以及产生模糊图像过程中额外的加性噪声 \mathbf{N} 。然而，即

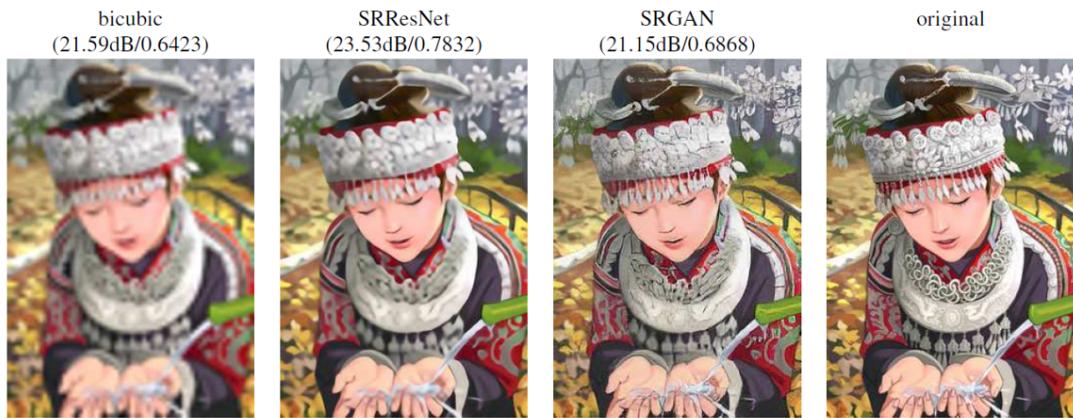


图 2-4 SRGAN^[2] 的结果示例。SRResNet 为不采用对抗损失的训练结果^[2]

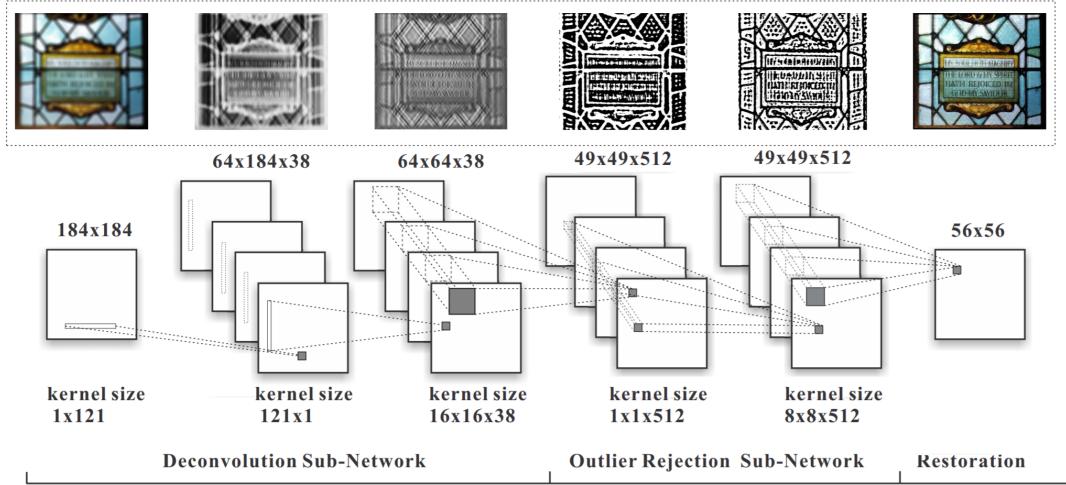
使已知模糊核函数，对模糊图像进行反卷积来求解清晰图像，仍是一个高度病态的问题。当已知模糊图像 B ，要求解清晰图像时 I ，其解不是唯一的，特别是在模糊核函数 K 未知的情况下，一幅模糊图像可能对应于多个清晰图像的解。

根据模糊核函数是否已知的情况，我们可以把去模糊算法分为两类。一类是已知模糊核函数的非盲去模糊 (Non-Blind Image Deblurring, NBID) 算法，另一类是核函数未知的盲去模糊 (Blind Image Deblurring, BID) 算法^[52]。

2.4.1 图像的非盲去模糊

当导致图像模糊的模糊核函数已知时，从模糊图像求解清晰图像的任务相对被简化。我们把这种已知模糊核函数情况下的去模糊操作称为图像的非盲去模糊。非盲去模糊假定有一个或一类已知的模糊核函数，再通过解反卷积的方式，求解出模糊图像对应的清晰图像。当模糊核已知时，理论上通过逆滤波计算，便能直接得到复原后的清晰图像。但由于加性噪声的存在以及不可估量，直接应用逆滤波方法得到的清晰图像复原效果并不理想。因此，非盲去模糊的重点在于如何更好的估计模糊核函数，以及消除噪声的影响。

早期的非盲去模糊方法包括使用维纳滤波 (Wiener Filtering)^[53] 和卡尔曼滤波 (Kalman Filtering)^[54]。维纳滤波器是一种线性滤波器，假定输入信号为目标信号和噪声的线性和，通过寻找一幅与输入图像均方差最小的输出图像，来求取最佳的滤波器参数，从而实现去模糊效果。求解过程可表示为 $\hat{I} = G \otimes B$ ，其中 \hat{I} 是计算出的对清晰图像的一个最小均方差估计， G 是我们目标的最佳滤波器， B 是输入的模糊图像。由此，我们的目标就转化为求取一个最优的 G ，使得生成的图像 \hat{I} 与期望的目标图像 I 间有最小的均方差。设计维纳滤波器的过程，需要用到期望输出与真实输出的信号，通过最小化期望输出与真实输出的均方差来优化滤波器参数。因此维纳滤波只适用于噪声频率特性已知的场景，通过过去所有观测值和当前观测值，来估计信号的当前值，消除特定的噪声。当输入噪声为频率随机的噪声时，维纳滤波效果较差，容易受到噪声波动的干扰。与维纳滤波相同，卡尔曼滤波也通过优化最小均方差来设计滤波器。不同的是，维纳滤波需要利用过去所有的观测值来进行参数优化，而卡尔曼滤波仅需要知道上一次的估计值和当前的观测值，就可计算出当前状态的估计值。卡尔曼滤波是一个递归优化的过程，首先用上一的估计来预测当前观测值的估计，再更新状态变量，获取更精确的当前观测估计。R-L(Richardson-Lucy) 算法是一个基于贝叶斯理论

图 2-5 DCNN 完整网络结构^[3]

的迭代优化算法，通过使期望最大来获得清晰图像最大似然估计。R-L 算法假设生成模糊图像的加性噪声服从泊松分布，无限次迭代求取期望输出的最大似然，进过多次迭代运算后可得到一个较好的结果，但随着迭代次数的继续增加，噪声会被放大，图像复原效果反而下降。同时，用 R-L 算法进行图像去模糊时，会产生振铃效应，在图像边缘处出现多层过渡带，影响去模糊结果。同样的，使用维纳滤波等方法进行去模糊时，也会出现振铃现象。

在此之后，有很多方法利用一些能够对自然图像进行统计建模的图像先验信息，如拉普拉斯先验，来缓解去模糊过程中的振铃效应^[55–57]。或者尝试处理噪声或饱和区域等额外信息^[58–60] 来提升去模糊效果。

2.4.2 图像的盲去模糊

在实际应用场景中，导致模糊的原因往往非常复杂，最终呈现的图像是由多种模糊因素和噪声共同作用的结果，而各种干扰因素也可能与输入信号本身的特性相关，不是一个简单固定的干扰。在这些情况下，人为的估计模糊核函数是一个非常困难的问题。我们把模糊核函数未知时的去模糊称为图像的盲去模糊。图像的盲去模糊方法更多的关注于如何通过有限的数据自动拟合出模糊核函数，再采用解卷积或其他非盲去模糊方法还原出清晰的图像；也有通过图像统计特性，端到端的直接还原出清晰图像的方法。Fergus 等人^[61] 针对单幅图像来取出相机抖动产生的运动模糊；Xu 等人^[62] 用两阶段的核函数估计来实现鲁棒性强的运动模糊去除算法；Cho 等人^[63] 提出了一个快速去运动模糊算法；Krishnan 等人^[64] 利用标准稀疏度 (normalized sparsity measure) 来进行解卷积过程；Whyte 等人^[65] 针对抖动和过饱和图像进行了去模糊；Pan 等人^[66] 用暗通道 (dark channel) 先验来进行图像去模糊。这些方法大都通过从模糊图像中预测一个运动方向，再利用已有的非盲图像去模糊算法来进行图像的盲去模糊^[67]。

也有采用神经网络来代替传统方法中某一过程的算法来进行图像去模糊。如 Jian Sun 等人^[68] 用两层卷积层和一层全连接层来预测图像中每一个 30*30 像素大小的图像块的运动模糊运动方向，从而构成造成图像运动模糊的模糊核。Xu 等人^[3] 通过核可分性定理 (kernel separability theorem)，将去模糊过程中用于解卷积的核函数转换为卷积神经网络，提出了一个用于图像去模糊的图像反卷积网络 DCNN 来对模糊核函数进行估计。随后，将 DCNN 与

一个用于去噪的卷积神经网络 ODCNN^[69] 结合，构成了完整的图像去模糊网络，如图2-5。

2.4.3 基于神经网络的图像去模糊算法

近年来随着深度学习的兴起，也涌现了很多直接利用卷积神经网络进行完整图像去模糊过程的算法。由于深度学习的强学习拟合能力，很多利用神经网络去模糊的方法都采用端到端的设计，将一幅模糊图像输入训练好的去模糊神经网络模型中，直接输出一幅处理后的清晰图像。Orest Kupyn 等人在 Pix2Pix 网络^[35] 基础上，提出了 DeblurGAN^[38]，将 GAN 框架应用于图像去模糊。DeblurGAN 是一个端到端的去模糊网络，主要用于运动模糊的盲去模糊。DeblurGAN 损失函数由对抗损失 (Adversarial Loss)、内容损失 (Content Loss) 两部分组成。对抗损失采用 WGAN-GP^[32] 中的损失函数以加强 GAN 的训练效果，内容损失采用感知损失^[11] 来增强生成图像的真实性。生成器采用了类似^[11] 中的结构，由两个步长为 $\frac{1}{2}$ 的卷积块，9 个残差块 (ResBlocks)^[70]，2 个转置卷积块组成。在残差块中，使用了全局的跳跃连接。判别器使用了 PatchGAN^[35] 中的结构。DeblurGAN 对于运动模糊的去除相较于同期工作有较好的表现，但实际输出结果不尽如人意，仍然存在一定程度的模糊。Seungjun Nah 等人^[7] 不进行模糊核函数的假设和估计，直接用端到端的方式恢复模糊图像。同时，采用多尺度训练的方式，逐层级放大图像尺寸，并在每一个层级都与对应尺寸的缩小清晰图像计算误差。生成器采用多个去除 BN 层的 ResBlocks^[70]，不同尺度采用相同的结构同时进行级联，损失函数为各尺度的均方误差与对抗损失。

2.5 本章小结

本章主要从图像超分辨率和图像去模糊两个方面，介绍了本文所研究的图像重建任务的研究现状。同时，针对于我们研究和应用过程中所用到的方法，我们还介绍了近年来在图像边缘检测、生成式对抗网络方面的研究进展。在图像边缘检测方面，传统的解决方法主要针对于图像的梯度分析，来检测和确定图像中物体的边缘轮廓。也有设计一些人工特征，通过统计或信息论的方法，来进行图像的边缘检测。近年来也有利用卷积神经网络来检测图像边缘。而生成式对抗网络是一种用于生成数据的网络架构，是数据生成领域的一类新方法，通过生成器和判别器的零和博弈对抗学习，来提高生成网络的生成效果，使得生成器能更好的拟合目标数据的分布，人为的制造出大量的目标分布样本。进过数年的发展，生成式对抗网络在图像数据的生成方面有了很多优秀的研究成果，同时也有各类方法来降低生成式对抗网络的训练难度。图像的超分辨率是用于还原低分辨率退化图像的一类方法。早期的超分辨率方法包括利用插值来补全缺失的信息，或者利用稀疏编码等方式将低分辨率图像映射为高分辨率图像，也有引入外部数据来进行补全的算法。而在深度神经网络出现后，也有相当多的研究集中与用卷积神经网络来生成高分辨率的重建图像。图像的去模糊是另一类常见的图像重建方法，主要针对于对模糊的退化图像进行重建。对于图像的去模糊问题，主要通过将模糊图像的生成建模为清晰图像与模糊核函数的卷积，再加上加性噪声的结果，通过对模糊核函数及噪声的假设或估计，以解卷积的方式，从模糊图像中还原出清晰图像。根据模糊核函数是否已知的情况，分为图像的非盲去模糊和图像的盲去模糊。早期去模糊方法有利用滤波的方式，在假设模糊核函数已知的情况下，用信号处理的方式抑制噪声的影响，从而求解出原始清晰图像。而盲去模糊方法很多则集中于对模糊核函数的估计，通过从模糊图像中估计出一个恰当的模糊核函数，再利用已有的非盲去模糊算法，来完成整个

图像去模糊过程。

第 3 章 边缘感知

图像边缘信息的提取一直是数字图像处理中的研究重点。作为区分物体的一个显著特征，边缘信息保留了图像中物体的结构性特征，减少了数据量，在图像处理和计算机视觉中，尤其是特征检测中扮演了重要的角色。图像的边缘通常被定义为亮度明显变化的点或线，在边缘的两侧，图像的像素值发生了剧烈的变化，反映出图像在边缘处的深度、亮度、物体属性或表面方向等属性的不连续。边缘信息主要用于描述图像的结构性信息。边缘图像中，相关性较弱的信息被省去，保留了更为重要的结构属性，大大减少了数量。同时，一个优秀的边缘图像，能够大大简化后续图像内容提取处理流程的复杂度。然而，想要获取一个精确的边缘图像是非常困难的。早期的边缘提取算法通过图像一阶导数的变化来查找边缘，将图像的边缘定义在梯度最大的点和方向上；或者通过图像的二阶导数，寻找二阶导数的零点来确定图像中的边缘。在此基础上，针对数字图像的离散矩阵形式表示，出现了如 Sobel 算子^[15]、Canny 算子^[17] 等边缘提取算子，以卷积的形式直接计算出图像的边缘点。但这些算法都存在一些人为的约束或阈值的限制，泛化性和适应性较差，不适用于任意场景的边缘检测工作。随后，我们采用基于神经网络的算法来进行边缘提取，并选择了一个基于 VGG 网络结构^[4] 的网络 HED(Holistically-nested Edge Detection)^[28]，作为我们边缘检测网络的基础结构，并对其进行剪枝操作。

3.1 边缘检测网络

本文所介绍的边缘检测网络被称为 EdgeNet，采用生成式对抗网络 GAN^[29] 的框架进行设计，整体网络分为生成器 (Generator) 和判别器 (Discriminator) 两个部分，生成器和判别器都是一个独立的网络，在优化其中一部分时，另一部分被看作是已经优化好的网络参与损失函数的计算，不断的交替循环迭代优化两个网络，最终达到一个全局最优解，此时的生成器能够完美拟合真实数据的分布，而判别器始终无法判别输入数据是来自于真实数据还是由生成器生成。

在边缘检测网络 EdgeNet 中，我们的生成器部分构示意图如图3-1。生成器整体依照 HED 网络^[28] 进行构建，采用 VGG-16 网络^[4] 作为生成器的基础架构，并添加了额外的侧输出层 (Side-output Layer) 和权重混合层 (Fuse Layer)，以更好的满足提取图像边缘信息的目标。EdgeNet 中生成器的隐藏层 (Hidden Layer) 由卷积层和池化层组成，基于在 ImageNet 挑战中具有出色的表现的 VGG-16 网络结构构建，并对其进行剪枝。VGG-16 网络是 VGG 网络

表 3-1 边缘检测网络的感受野和步长。加粗部分连接了侧输出层

网络层	c1_2	p1	c2_2	p2	c3_3	p3	c4_3	p4	c5_3	p5
感受野	5	6	14	16	40	44	92	100	196	212
步长	1	2	2	4	4	8	8	16	16	32

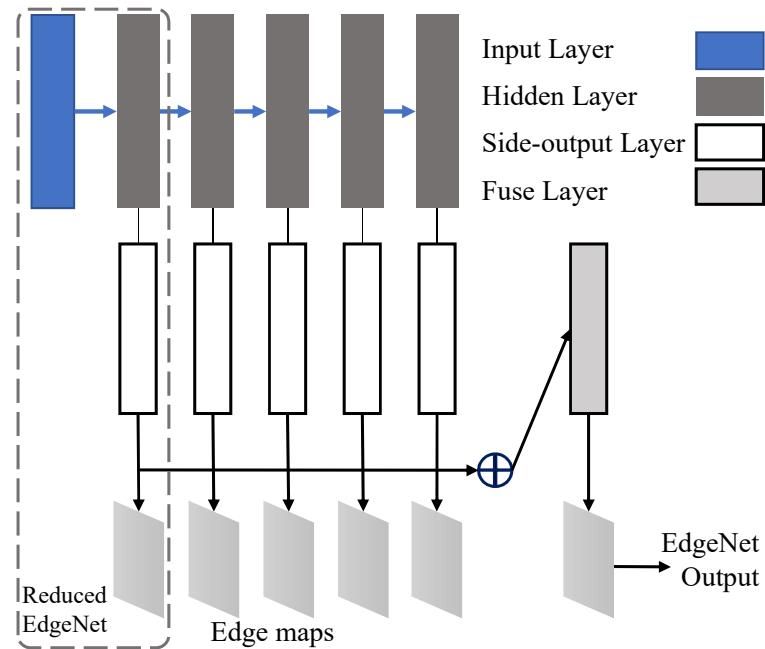
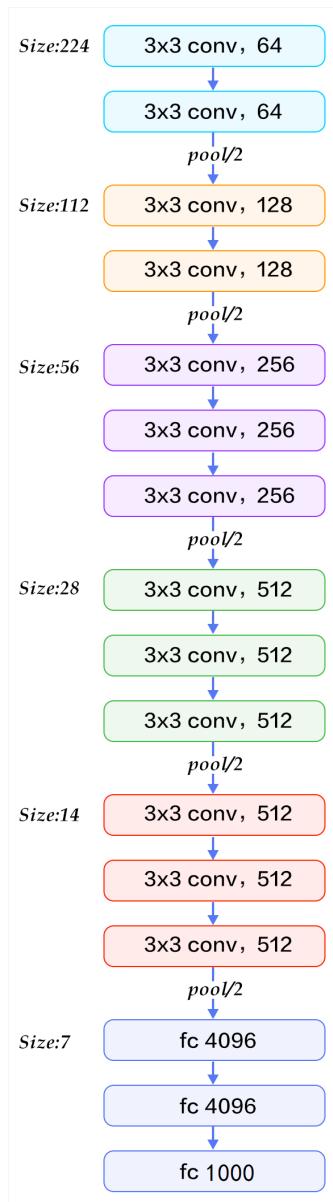


图 3-1 边缘检测网络结构示意

组中的一种配置，具有高达 16 层的足够的深度和步长为 1 的连接稠密度，并有 5 个步长为 2 的降采样层组成的块 (Block)。VGG-16 接受一个大小为 224*224 的 3 通道 RGB 图像作为输入，经过 5 个块的卷积和池化运算后，再接入 3 层全连接层，最终由 SoftMax 层输出一个表示分类结果的多维向量。VGG-16 中的每个块都由多个卷积层和一个最大池化层 (Max Pooling) 组成，前两个块中分别包含 2 个卷积核为 3 的卷积层，后三个块中分别包含 3 个卷积核为 3 的卷积层。在同一个块中，通道数对于每一个卷积层来说都是相同的，VGG-16 中 5 个块的通道数分别为 64、128、256、512、512。在每个块的多个卷积层之后，还分别接有一个最大池化层。最终 5 个块的输出将送入三层全连接层 (Full Connected Layer, FC) 中进行计算，三个全连接层的维度分别为 4096、4096 和 1000。最终的计算结果经 SoftMax 层处理后输出。VGG-16 的具体网络结构示意如图3-2。已经有很多的工作表明^[27]，利用在 ImageNet 等大数据集上预训练好的 VGG 分类器作为特征提取器或基础模型，并在这基础上进行微调 (Fine-tuning)，能够有效提升网络效果，缩短训练时间。通常在用预训练的 VGG 网络作为基础模型架构时，仅保留前面的 13 层卷积层，作为特征提取器，去掉后面的 3 层全连接层和最后的 SoftMax 层。

我们对构建隐藏层的 VGG 网络做出了一些修改，在每一个隐藏层阶段的最后一个卷积层之后，都插入有一个侧输出层 (Side-output Layer)，作为一个多尺度多级别的输出，即在 VGG-16 网络中每一个块 (Block) 最后一个卷积层 c1_2、c2_2、c3_3、c4_3 和 c5_3 之后，接入一个额外的卷积层作为对应的侧输出层，每个侧输出层的感受野与这些卷积层相对应。其中，c1_2 代表 VGG-16 中第 1 个块的第 2 个卷积层。另外，由于边缘检测网络的训练过程中需要用到不同尺度的侧输出层的边缘预测输出来增强训练效果，对于高层的输出结果，在进过多层卷积之后已经变得太小而失去了使用价值，放大插值后的结果变得模糊；而全连接层需要消耗大量的计算资源和存储资源。因此我们去除了 VGG 网络中第五个池化层，同时也去除了所有的全连接层。最终的边缘检测网络依照 VGG 网络形式构建，由 5 个块构成，每

图 3-2 VGG-16^[4] 网络结构示意图

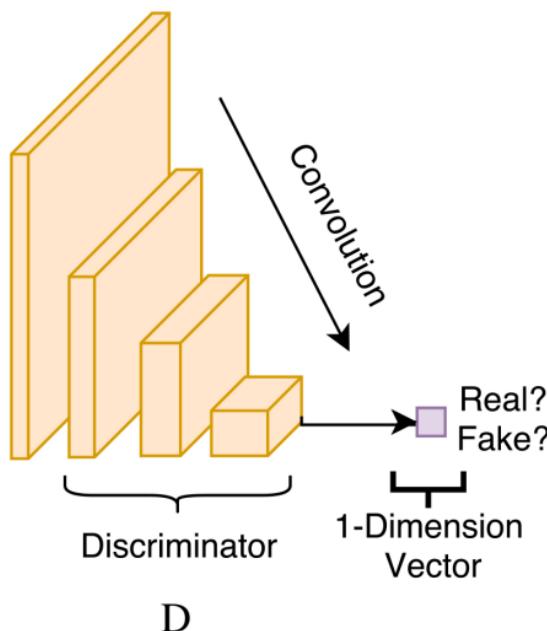


图 3-3 边缘检测网络判别器结构示意

个块的步长分别为 1、2、4、8、16，各个块相互间有不同大小的感受野，参数配置如表3-1。侧输出层是一层卷积层，用于将 VGG-16 网络中每一块的特征图输出，映射为该尺度下的一个边缘预测结果，每个侧输出层的输入维度和通道数与其对应的 VGG-16 卷积层相同，其输出为一个同尺度的单通道边缘预测结果。

而在网络的最后部分，还有一个额外的权重混合层 (Fuse Layer)，用于将所有不同尺度的输出进行结合，得到最终的边缘检测结果。由于不同的侧输出层预测的边缘图像尺度不一致，因此在最终组合所有侧输出层的预测结果时，将较小的边缘预测图像用线性插值进行放大，将所有的侧输出层预测结果放大到统一尺度后，再计算出最终的边缘预测结果。最终的边缘预测结果，为所有侧输出层预测结果在放大到同一尺度下的均值。在将侧输出层的预测结果进行上采样，使尺度归一化时，可以使用带可学习参数的转置卷积进行上采样^[71]，但在我们的实验中发现，采用转置卷积进行上采样时，其效果相较于直接使用线性插值并没有明显的提升，因此最终我们仍使用了线性插值作为放大侧输出层边缘预测图像的方法。

边缘检测网络 EdgeNet 的判别器与 PatchGAN^[35] 中相似。我们所采用的判别器是一个针对图像块 (Patch) 进行判断的判别器，判别器仅对特定 $N * N$ 大小图像块的真实性进行判断。对于一幅输入判别器的图像，首先需要将其分解为多个 $N * N$ 大小的图像块，分别输入判别器中，获取每一块的真实性判断结果。判别器最终的输出是所有图像块真实性评价的平均。本文所设计的判别器由 4 个卷积层组成，接收 $70 * 70$ 大小的图像块作为输入，所有卷积层的卷积核大小均为 4，步长均为 2，每个卷积层的通道数从上至下分别为 64、128、256、512。不同于 PatchGAN 中使用的批量归一化层 (Batch Normalization Layer)^[72]，我们在每个卷积层之后，还分别连接有一个谱归一化层 (Spectral Normalization Layer)^[73] 和一个 LeakyReLU^[74] 层。另外，判别器的最后，还有一个卷积层用于输出一个 1 维的最终判断结果。

尽管最初的 HED 网络在图像边缘检测上已经有了非常出色的表现，但对于模糊的图像

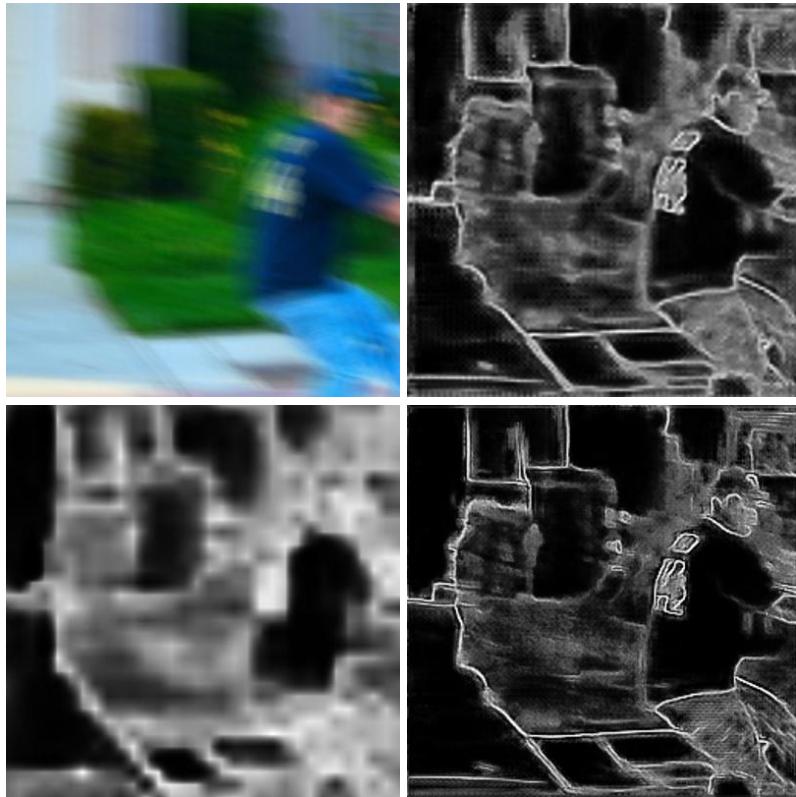


图 3-4 边缘检测网络中第 1, 3, 5 层侧输出的结果示例。从左到右从上到下分别为：模糊图像，第 1 层侧输出结果，第 3 层侧输出结果，第 5 层侧输出结果

仍然不能很好的检测出边缘，无法达到与清晰图像相匹配的效果。而在实验过程中我们发现，在多个不同尺度的侧输出层中，第一个侧输出已经保留了很多细节和局部边缘信息，其效果与最后一个侧输出相当，如图3-4。因此，我们在网络训练过程和网络推理过程采用了不同的策略。在训练过程中，整个 EdgeNet 都参与了网络训练，包括所有的侧输出层，以达到较好的多尺度边缘检测效果；而在具体的测试和推理应用阶段，我们采用了仅保留第一个侧输出层的 EdgeNet 子网络，在保留足够的边缘信息的同时，尽可能减少计算资源的占用。实验表明，采用这种策略进行推理，获得的边缘图像在后续的处理流程中能够达到与使用完整网络输出相媲美的结果。详细实验结果和分析见5.7。

3.2 边缘检测网络的训练

在 HED 网络的训练中，输入的训练图像都是清晰的图像，HED 网络仅需要学习如何从清晰图像中提取其中的边缘信息。而在去模糊的任务场景下，输入数据往往是模糊图像，并不存在清晰锐利的边缘，这为 HED 网络的训练带来了很大的阻碍。不同于 HED 网络中的训练策略，我们对边缘检测网络 EdgeNet 的训练进行了改进，引入了对抗损失，采用生成式对抗网络的方式进行训练，如图5-3(a)。在生成式对抗网络中，判别器通常具有很强的图像分布拟合能力，能很快的学习判断出输入图像是来自于真实数据还是来自于生成器的生成，这有助于帮助边缘检测网络 EdgeNet 中的生成器快速的学习如何从一幅模糊的图像中提取出边缘信息。

我们对边缘检测网络 EdgeNet 的损失函数定义如下：

$$L_{EdgeNet} = L_{Edge} + \lambda \cdot L_{Disc}$$

其中， L_{Edge} 为边缘损失项， L_{Disc} 为对抗损失项， λ 为对抗损失的权重。边缘损失 L_{Edge} 是一个基于类平衡的交叉熵损失 l_{CBCE} ^[28]：

$$L_{Edge} = \sum_i^n l_{CBCE} [\phi_{side_i}(I_B), E] + l_{CBCE} [\phi_{fuse}(I_B), E],$$

其中， E 表示真实的边缘图像， I_B 表示输入的模糊图像； $\phi_{side_i}(I_B)$ 是由第 i 个侧输出层输出的边缘预测图， $\phi_{fuse}(I_B)$ 是由权重混合层输出的边缘预测图。

对抗损失 L_{Disc} 的计算与最初的 GAN 网络^[29]是一致，这里的 D_{θ_D} 和 G_{θ_E} 分别表示边缘检测网络 EdgeNet 中判别器的参数和生成器的参数。

$$L_{Disc} = -\log [D_{\theta_D}(G_{\theta_E}(I_B))]$$

在实际的网络训练过程中，我们把边缘检测网络 EdgeNet 的判别器视为一个训练的加速器。在训练的最初始阶段，我们将 λ 设为 0.05 这样一个很小的权重，来防止判别器太快的对训练数据过拟合。经过 50 轮迭代训练后， λ 被设为 0。也就是说，我们的判别器仅用于作为一个预训练的加速器使用。

3.3 本章小结

本章详细介绍了在我们的研究中边缘感知的实现过程。实现了一个边缘检测网络 EdgeNet，用深度神经网络的方式，构建出一个端到端的，多尺度多层次的边缘特征提取方法。我们详细描述了边缘检测网络的构建策略，包括网络的架构设计，各个卷积层的大小及参数，以及网络的训练策略，损失函数的定义。本章介绍的边缘检测网络将分别应用在之后的图像超分辨率网络和图像去模糊网络中，辅助完成图像的重建任务。

第 4 章 基于边缘感知的深度图像超分辨率

本章将介绍一个基于边缘感知的深度图像超分辨率网络。图像的超分辨率是图像复原中的一个经典任务。低分辨率的图像由于在退化过程中丢失了很多细节信息，给后续的辨认和处理都带来了很大的挑战。图像的超分辨率在许多应用场景中都非常有效，特别在光学字符识别 (Optical Character Recognition, OCR) 场景下，能够有效的提升字符的检测和识别结果。文档图像的分辨率无论是对 OCR 系统还是人类视觉来说，都是识别图像中文字的一个很重要的因素。通常来说，文档图像的分辨率越低，视觉信息就丢失得越多，识别正确率也随之下降。文档图像的细节信息丢失会非常严重的影响字符检测和字符识别的效果，因此，超分辨率方法常被作为 OCR 系统中的一个预处理步骤，来提升系统的整体效果。我们针对于文档图像处理的场景，设计了一个基于边缘感知的深度图像超分辨率网络，用于将低分辨率的文档图像复原为高分辨率图像，并且保留文档中字符的结构信息，避免文档经过超分辨率处理后，字符结构被新添加的细节信息破坏而无法辨认，图像质量反而因此降低的情况发生。在本章中，我们首先将介绍该网络的设计目的，描述并定义其所要解决的问题。然后我们会给出这个基于边缘感知的深度图像超分辨率网络的网络结构设计，详细定义网络的训练策略和目标函数的设计。在本章的最后，我们给出了网络在基准数据集上的测试结果，并对结果进行了比较分析。

4.1 问题的描述与定义

本节我们将给出基于边缘感知的深度图像超分辨率网络所要解决问题的描述和定义。本章所介绍的基于边缘感知的深度图像超分辨率网络属于数字图像处理中的图像超分辨率领域。图像的超分辨率，就是将一幅退化的低分辨率图像 (Low-resolution Image, LR)，还原转变为一幅高分辨率的图像 (High-resolution Image, HR)，期间，在为低分辨率图像添加细节信息的同时，还需要保证原始低分辨率图像中的结构信息不被破坏。

4.2 超分辨率网络结构设计

本章要介绍的超分辨率网络主要针对于文档图像的应用场景。虽然低分辨率图像由于细节信息的丢失，导致文档图像中的字符难以被识别，降低了识别正确率，但是过高分辨率的文档图像也并不能更进一步的提升识别效果，反而带来了更重的计算负担。因此，考虑到现代 OCR 系统的实际应用背景，我们的超分辨率模型应当具有不同的放大倍率，以适应于不同程度的低分辨率情况。当文档图像的分辨率特别低时，我们的模型需要更大的放大倍率。作为一个预处理步骤，一个高效的超分辨率模型对整个 OCR 系统的处理都是有益的，这促使我们去构建一个轻量级的网络架构，并进一步以可组合的方式搭建我们的模型。

我们所构建网络的目标是以用可调整的放大倍率来超分辨率处理文档图像或者文字区

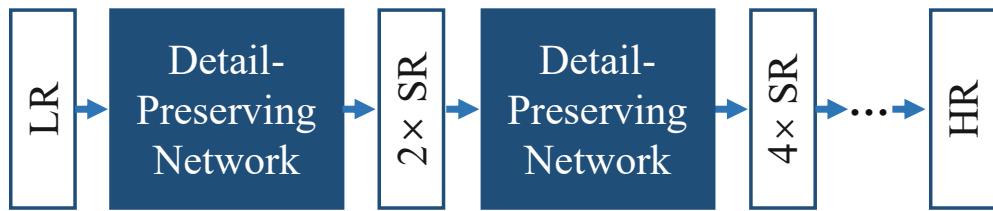


图 4-1 级联的超分辨率网络工作流程

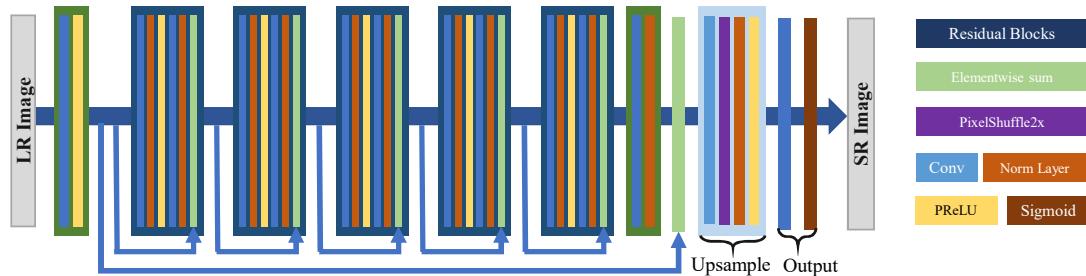


图 4-2 DPNNet 网络结构图

域。为此，我们用级联的方式来设计我们的超分辨率网络。如图4-1所示，模型整体由多个级联的细节保留网络 (Detail-Preserving Networks, DPNNet) 网络组成，每个 DPNNet 都有一个较小的超分辨率放大倍率 ($2\times$)。针对不同尺度文档图像训练的多个 DPNNet 网络具有相同的网络结构，但他们的参数是各自独立的。整个模型由多个细节保留网络 DPNNet 相连组成，每个 DPNNet 都是根据相邻输出的尺寸来训练的。模型整体输入的低分辨率图像将被连续的放大，产生了以乘数倍率放大的超分辨率图像。

本文所要介绍的细节保留网络 DPNNet 的结构如图4-2所示。DPNNet 使用了一个生成式的 CNN 卷积神经网络架构，遵循常见的单幅图像超分辨率模式进行设计，并由三个部分构成。第一部分是一系列的卷积神经网络，用于从输入图像中提取出具有固定大小的特征作为下一步的输入。在这部分中，我们首先使用了一层卷积核大小为 9 的卷积神经网络来从输入图像中提取出低级 (low-level) 图像特征。随后，再用 N 个残差块 (ResBlocks) 来从输入的低级特征中进一步提取图像的高级 (high-level) 特征。在本文所介绍的方法和实验中，我们选取了 $N = 5$ ，以保持模型效率和效果的平衡。所有残差块中的卷积核大小均为 3。同时，我们也将跳跃连接加入了这个部分，以更好的训练残差块，并混合图像的高级特征和低级特征。网络的第二部分是一个上采样模块。由于使用多个上采样块不能充分利用各个尺度间的特征，因此我们在这个部分仅采用了一个上采样块来进行放大。在上采样块中，我们使用了亚像素卷积层^[75] 来作为基本结构。在最后一部分中，我们使用了单层卷积神经网络和一个 Sigmoid 函数来生成最终的输出图像。

在我们的细节保留网络 DPNNet 中，所有的卷积块都使用了同一种归一化层和激活函数。我们选用了批归一化^[72] 方法来构建所有的归一化层，同时使用参数化的修正线性单元 (Parametric Rectified linear unit, Parametric ReLU)^[76] 作为激活函数。

4.3 超分辨率网络的训练

由于我们所设计的基于边缘感知的深度图像超分辨率网络采用了级联的结构，我们将训练流程分成了两个阶段进行。第一个阶段为平行训练阶段，第二个阶段为渗透性微调阶

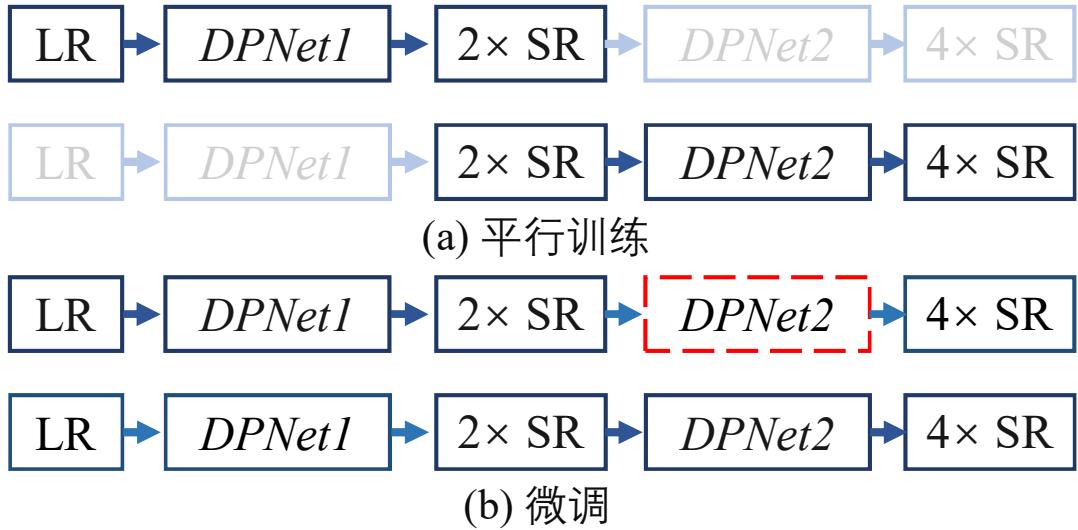


图 4-3 级联的细节保留网络训练策略。红色部分为参数冻结的 DPNet

段。模型训练策略的概述见图4-3。

4.3.1 平行训练

基于边缘感知的深度图像超分辨率网络模型是由多个细节保留网络 DPNet 级联而成，对于其中每个独立的 DPNet 网络组件，都是接收一个低分辨率图像作为输入，并输出一个高分辨率图像。在平行训练这个阶段，我们假设对应于不同尺度的 DPNet 网络是相互独立的，仅共享相同的结构设计，每个网络单独分别训练。图4-3(a) 中表示了这一过程，两个不同的 DPNet 网络被同时进行训练。以 $4\times$ 放大倍率的超分辨率网络为例，需要两个相互独立的 DPNet 来搭建整个模型。经过降采样处理，原始的高分辨率图像生成了 $2\times$ 和 $4\times$ 大小的低分辨率图像。 $4\times$ 的低分辨率图像被输入 DPNet1 中，经过 DPNet1 超分辨率处理后的输出结果将与 $2\times$ 的低分辨率图像相比较并计算损失。计算出的损失经反向传播后，被用来优化 DPNet1 中的网络参数。

相同的，DPNet2 将同时使用 $2\times$ 的低分辨率图像和原始高分辨率图像进行平行的训练。对于需要更大放大倍率的模型，每个单独的 DPNet 都能以同样的方式进行平行训练，这种训练方式在有多个 GPU 供使用时尤为有效。

4.3.2 渗透性微调

上一阶段的平行训练使每个独立的 DPNet 网络都具备了以较小放大倍数进行图像超分辨率的能力。然而，如前文所述，图像的复原问题是个病态的问题，任何模型都可能无法很快的找到一个与原始高分辨率图像相同的结果。因此我们设计了一个微调的阶段来使各个独立 DPNet 网络的参数相互适应，如图4-3(b)。

在每一步中，图上用箭头相连的所有 DPNet 网络都被用来进行渗透性微调。DPNet2 的网络参数在初始时被冻结，不进行更新。而此时整个网络将会接收低分辨率的图像作为输入，并且以 $4\times$ 的倍率来进行图像超分辨率处理，以此更新 DPNet1 的权重参数。整个基于边缘感知的深度图像超分辨率网络将在这个阶段以这种方式相继进行微调训练，微调顺序为从第 2 个 DPNet 到第 N 个 DPNet。

4.3.3 损失函数

对于训练过程中的每个阶段以及每个独立的 DPNet 网络，训练所使用的损失函数都是相同的。我们训练基于边缘感知的深度图像超分辨率网络所使用的损失函数由三个部分组成：

$$L = \lambda_1 \cdot L_{Pixel} + \lambda_2 \cdot L_{Perceptual} + \lambda_3 \cdot L_{Edge}$$

损失函数中的第一部分是一个像素内容损失 (pixel loss)，表示超分辨率结果和真实清晰图像间的点对点均方误差 (Mean-Square Error, MSE)。第二部分感知损失 $L_{Perceptual}$ ^[11] 是一个基于特征图 (Feature Map) 差异的损失项。感知损失的计算通过一个在 VGG-19 网络^[4]上，用 ImageNet^[7] 数据集预训练好的模型来完成，通过比较生成图像和目标图像在这个 VGG-19 网络中，不同卷积层所输出特征图的差异，来表现生成图像在人类感知层面上的真实性。两幅特征图间的差异越小，表示生成的图像在人类的感知中越接近于真实的图像。感知损失的定义如下：

$$L_{Perceptual} = \frac{1}{W_j H_j} \sum_x \sum_y (\phi_j(I_H R)_{x,y} - \phi_j(S(I_L R))_{x,y})^2$$

其中 ϕ_j 表示 VGG-19 网络中的第 j 个卷积层， (W_j, H_j) 是该层的特征图尺寸； $S(\cdot)$ 是超分辨率函数， $I_H R$ 为真实的清晰图像， $I_L R$ 为输入的低分辨率图像。我们将像素内容损失和感知损失统称为内容损失，并且通过 L_2 来计算均方误差 MSE。在早期实验中，我们发现仅采用感知损失，或者使用感知损失和 L_1 误差，会在生成结果上出现不真实的纹理，这一点在 Mechrez 等人^[78] 和之前的一些其他研究中也有说明。因此本文中采用 L_2 误差和感知损失共同构建内容损失。

损失函数的第三部分是一个边缘损失。边缘损失的计算用到了3.1中介绍的边缘检测网络 EdgeNet。我们使用类平衡的交叉熵损失 (class-balanced cross-entropy loss)^[28] 来进行边缘损失的计算。我们将经过超分辨率处理后的图像与原始的高分辨率图像分别输入已训练好的边缘检测网络 EdgeNet 中，得到各自对应的边缘图像，再用这两幅边缘图像来计算边缘损失。我们的边缘损失定义为：

$$L_{Edge} = l_{side}(\phi_{side_i}(I_H R), \phi_{side_i}(S(I_L R))),$$

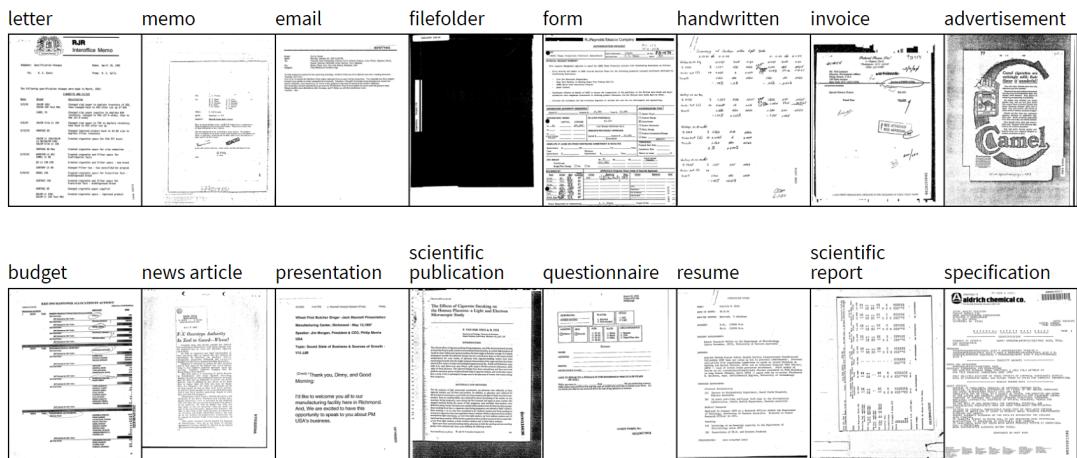
其中 ϕ_{side_i} 是 EdgeNet 中第 i 个侧输出层的检测结果， l_{side} 表示类平衡的交叉熵损失。在实际的使用过程中，我们在 EdgeNet 中取 $i = 1$ 以减少训练和推理时间。关于 i 值效果的相关实验结论见5.7。我们在基于边缘感知的深度图像超分辨率网络训练中使用边缘损失的目的有两个，一是通过增强的边缘信息来在较小的放大倍率中保留图像的细节信息；二是在我们之前的实验中发现，引入边缘损失能够加快训练过程中网络的收敛速度。

4.4 实验数据的准备

由于在文档应用场景中，并没有公开的数据集能直接应用于文档图像的超分辨率，为了验证我们所设计的基于边缘感知的深度图像超分辨率网络的有效性，我们搜集了两个文档图像数据集作为训练和测试的基础数据集。

and brings about 4% improvement over the state-of-
68.23 (68.55) Total assets Single-row Features (local)
 decoding path π , as long the output label $o[e]$,

图 4-4 ICDAR17-Textline 数据集

图 4-5 RVL-CDIP 数据集^[5]

RVL-CDIP Region 数据集

RVL-CDIP(Ryerson Vision Lab Complex Document Information Processing) 数据集^[5] 中包含有 16 类文档，每类有 25K 张文档图像，如图4-5。在这 400K 张灰度文档图像中，80% 被作为训练集，10% 为验证集，剩余的 10% 为测试集。我们从原始的 RVL-CDIP 数据集中，随机采样了 32K 个图像区域进行训练，同时分别采样了 4K 个图像区域作为验证图像和测试图像，以此来构建我们的 RVL-CDIP Region 数据集。每个图像区域大小均为 320×320 。

ICDAR17-Textline 数据集

除了从 RVL-CDIP^[5] 中构建我们的 RVL-CDIP Region 数据集外，我们还根据 ICDAR 2017 会议构建了一个 ICDAR17-Textline 数据集。ICDAR17-Textline 数据集是一个包含有识别结果标注的文本行区域图像数据集，如图4-4，用来同时验证超分辨率网络的效果，和经过超分辨率处理后的 OCR 系统识别精度。我们从 ICDAR 2017 会议中随机选取了 20 页数据，将他们进行打印，再分别以 300DPI 和 75DPI 的分辨率进行全页扫描，得到对应的高分辨率和低分辨率数字图像。在每一页中，我们随机裁剪出 30 个文本行区域，并对这些文本行图像进行了文本内容标注。对于所有的文本区域图像，其中的 400 张被分为训练集，剩余的 200 张作为测试集。

4.5 训练环境及超参数的设定

我们使用 PyTorch^[79] 深度学习库来搭建我们基于边缘感知的深度图像超分辨率网络。所有的实验均在 Intel Xeon-E5 CPU 和 NVIDIA Titan XP GPU 上完成。我们还在同样的环境和同样的数据集上，使用不同的微调网络参数评估了一些其他方法。

对于优化器的选择，我们使用了 Adam^[80] 来作为我们训练过程中的优化器。Adam 的参数配置我们使用了 $\beta_1 = 0.9$ 和 $\beta_2 = 0.999$ 。初始学习率被设为 0.001，并且在每 20 轮迭代后衰减为原来的十分之一。如4.3所述，网络的训练由两阶段组成，我们在平行训练中对每个 DPNet 训练了 50 轮迭代，而后的微调过程训练了 5 次迭代。

4.6 实验结果

我们在上一节中构建的 RVL-CDIP Region 数据集和 ICDAR17-Textline 数据集上分别进行了实验，来评估本文所介绍的基于边缘感知的深度图像超分辨率网络的效果，同时与其他超分辨率网络的结果进行了比较。我们从两个方面来评价网络对文档图像的超分辨率效果，一是传统的图像质量评价指标，包括峰值信噪比 PSNR 和结构相似性指标 SSIM，二是经过超分辨率处理后，文本行图像在 OCR 系统中的识别率。

4.6.1 评价标准

在图像重建领域，除了人工用肉眼比较重建效果外，常用的评价标准包括峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 和结构相似性 (Structural Similarity, SSIM)。

峰值信噪比 PSNR 是指信号本身最大可能功率与噪声功率的比值，常用于表示信号的质量以及经过处理后信号的品质，单位为分贝。在图像领域中，峰值信噪比通常作为信号重建质量的一个评判标准。峰值信噪比越高，表示图像的恢复效果越好。计算图像的峰值信噪比，常通过计算均方误差 MSE 来实现。两幅图像的均方误差可由下式得出：

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

其中 I 为原始图像， K 为处理后的图像， m 和 n 分别为图像的高和宽。而图像的峰值信噪比定义为：

$$PSNR = 10 \cdot \lg \left(\frac{MAX_I^2}{MSE} \right)$$

其中， MAX_I 表示原始图像 I 中的最大像素值。峰值信噪比在评价图像处理质量中应用广泛，但仍然有其缺陷。由于人眼感知会受周围信息的干扰，同时对于不同特性的信息敏感程度不同，峰值信噪比的得分高低无法完全代表人类视觉感知的图像质量好坏。在大量应用场景中都存在峰值信噪比得分高的图像，人类感知却认为其质量不如另一幅峰值信噪比得分更低的图像。

结构相似性指标 SSIM 用于评价两幅图像的相似程度。在图像的恢复重建领域中，我们希望图像重建过程既能补全丢失的信息，又能保持原始的结构不发生改变。结构相似性指标便是用于衡量恢复图像在结构上的失真程度，值越高代表两幅图像的相似程度越高，恢复效果越好。结构相似性指标从亮度、对比度和结构三个方面来衡量两幅图像的相似性，它的定义如下：

$$SIM(I, K) = l^\alpha(I, K) c^\beta(I, K) s^\gamma(I, K)$$

其中 $l(I, K)$ 比较图像 I 和 K 的亮度, $c(I, K)$ 比较对比度, $s(I, K)$ 比较结构; α, β, γ 为衡量三者重要性的权重参数。同时, 有:

$$l(I, K) = \frac{2\mu_I\mu_K + C_1}{\mu_I^2 + \mu_K^2 + C_1}$$

$$c(I, K) = \frac{2\sigma_I\sigma_K + C_2}{\sigma_I^2 + \sigma_K^2 + C_2}$$

$$s(I, K) = \frac{\sigma_{IK} + C_3}{\sigma_I + \mu_K + C_3}$$

其中, μ_I 和 μ_K 分别为 I 和 K 的均值, σ_I 和 σ_K 分别为标准差, σ_{IK} 为 I 和 K 的协方差, C_1, C_2, C_3 为常数。通常在实际使用中, 有 $\alpha = \beta = \gamma = 1$, 以及 $C_3 = \frac{C_2}{2}$ 。此时有:

$$SSIM(I, K) = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)}$$

无论峰值信噪比 PSNR 还是结构相似性指标 SSIM, 都是基于图像全局计算的评价标准。而人类感知评价图像恢复结果的好坏时, 也会考虑局部信息, 甚至由于一个局部的不够真实而对全图给出较低的评价。因此, 常会出现 PSNR 或 SSIM 得分高而人评价低的情况。仅用峰值信噪比和结构相似性指标来评判去模糊算法的有效性显得远远不够。

为了衡量文档图像在经过超分辨率处理后, 对后续字符识别过程性能的提升, 我们还使用了其他指标来评价超分辨率处理结果的好坏。在经过图像超分辨率处理后, 输出图像被输入商用 OCR 识别系统^①中进行字符识别。我们定义了两个识别精度指标, 分别是最长公共子序列 (Longest Common Subsequence, LCS) 得分和莱文斯坦距离 (Levenshtein Distance) 得分。最长公共子序列得分和莱文斯坦距离得分都在区间 $[0, 1]$ 内。最长公共子序列得分定义如下:

$$S_{LCS} = \frac{LCS_length(s, t)}{Maxlen(s, t)},$$

其中, s 和 t 分别表示预测结果和目标文本。最长公共子序列得分是一个最长公共子序列长度和 s 与 t 最大长度的比值, 即: $Maxlen(s, t) = \max(len(s), len(t))$ 。仅当 s 和 t 完全相同时, 最长公共子序列得分才取到最大值 1.0。莱文斯坦距离得分基于莱文斯坦距离计算。莱文斯坦距离又被称为编辑距离, 是一个衡量两个字符序列间差异的评价指标。我们使用莱文斯坦距离和最大长度 $Maxlen$ 来计算 s 和 t 的相似度。莱文斯坦距离得分定义如下:

$$S_{LD} = 1 - \frac{Levenshtein_distance(s, t)}{Maxlen(s, t)}.$$

4.6.2 比较实验结果

我们在自己构建的 RVL-DCIP Region 数据集和 ICDAR17-Textline 数据集上进行了实验, 并和当前图像超分辨率领域的最先进 (state-of-the-art) 算法进行了比较。实验结果见表4-1。我们用本文提出的基于边缘感知的深度图像超分辨率网络和传统的双三次插值 (Bicubic) 方法进行了比较, 同时还比较了最近的一些基于深度学习的图像超分辨率算法, 如 SRCNN^[6] 和 SRGAN^[2]。所有的基准方法 (baseline) 和我们所提出的网络框架都具有同样的 $4\times$ 放大倍

^① ABBYY Fine Reader 14: <https://www.abbyy.com/en-apac/finereader/>

表4-1 基于边缘感知的深度图像超分辨率网络在两个数据集上的实验结果

Method	RVL-DCIP Region		ICDAR17-Textline				
	PSNR	SSIM	PSNR	SSIM	S_{LCS}	S_{LD}	
Bicubic	20.74	0.7113	19.99	0.8025	0.6109	0.5771	
SRCCNN ^[6]	21.98	0.7225	22.77	0.8052	0.6395	0.6080	
SRGAN ^[2]	23.89	0.7403	25.85	0.8145	0.8617	0.8506	
Cascaded DPNets	25.27	0.7541	30.51	0.8779	0.9134	0.9089	

图4-6 RVL-DCIP Region 数据集上 4× 放大结果。从左至右：低分辨率图像（4×退化），高分辨率图像，双三次插值结果，SRCNN^[6]结果，SRGAN^[2]结果，我们的结果

率。实验结果表明，我们所设计的基于边缘感知的深度图像超分辨率网络，用级联 DPNet 网络的方式，在两个数据集上都有很好的表现，最终生成结果的峰值信噪比 PSNR 和结构相似性指标 SSIM 都高于其他方法。图4-6和图4-7展示了我们的网络分别在 RVL-DCIP Region 数据集和 ICDAR17-Textline 数据集上的定性评估和超分辨率输出结果。

我们成功的在不同文档类型和不同文字字体保留了文本区域图像中字符的细节信息，特别是在出现小字符的情况下，字符的细节信息得到了保留和增强。但是当有些字符特别小时，仍然会有失败的情况出现。另外，在多个太过邻近的连续字符出现时，也会有一些超分辨率处理失败，字符细节信息和结构信息被破坏的情况。使用 OCR 系统在 ICDAR17-Textline 数据集上的字符识别结果同样在4-7展示。从实验结果可以看出，将我们提出的基于边缘感知的深度图像超分辨率网络和 OCR 系统联合，能够有有效的提升后续步骤的识别精度。在类似于峰值信噪比和结构相似性指标这种全局图像质量指标上，我们的超分辨率结果相较于基线方法有了一定的提升。同时，也生成了高质量的文本字符和图像细节，能够为将来

following reasons: lol lowing rc4M>m:	following reasons: following rxvnx:	following reasons: following reasons:	following reasons:
66.41 (66.77) M4I j60 7T>	66.41 (66.77) M J I 4 W\77 >	66.41 (66.77) 66.41 (66.77)	66.41 (66.77) 66.41 (66.77)
no. 1, pp. 1-20, 2016. •u l.ro 1 · 3). »16	no. 1, pp. 1-20, 2016. no I · PV I-o. 2016	no. 1, pp. 1-20, 2016. no. I* pp · 1-Z0. 2016,	no. 1, pp. 1-20, 2016. no. I, pp. 1-20. 2016.
C. Aspect Ratio C. i Raiw	C. Aspect Ratio C. Axprc/ Ratio	C. Aspect Ratio C Arpect Ratio	C. Aspect Ratio C. Aspect Ratio
by using intra-cluster UMfig uiira-clusler	by using intra-cluster by uung in(ra<luMcr	by using intra-cluster by using intra-cluster	by using intra-cluster by using intra-cluster
in comparison with the in c<xnpuriMMi with the	in comparison with the in comparison, viih the	in comparison with the in comparison with the	in comparison with the in comparison with the
tion," in ICCV, 2011. r>IMt<rmlCCV0. XII	titice," in ICCV, 2011. ntxn." tn tCCV. 2011	nition," in ICCV, 2011. niiKYir in /CCV.2011.	nition? in /CCV, 2011.
feature triplets <Null>	feature triplets fcMure tnptk<	feature triplets fcafurc triplets	feature triplets feature triplets

图 4-7 ICDAR17-Textline 数据集上 4× 放大结果。从左至右：双三次插值，SRCNN，SRGAN，我们的结果。
(Null) 表示该图像从 ABBYY 中没有返回结果

表 4-2 RVL-DCIP Region 数据集上不同设置的实验结果

Method	PSNR	SSIM
Cascaded DPNet without Edge	24.96	0.7487
Cascaded DPNet with Edge	25.27	0.7541

(a) 边缘损失

Method	PSNR	SSIM
Bicubic (4×)	20.74	0.7113
Bicubic (2×) + DPNet (2×)	21.12	0.7218
DPNet (2×) + Bicubic (2×)	22.95	0.7361
Cascaded DPNet (4×)	25.27	0.7541

(b) 不同的级联结构

的后处理流程，如文档布局信息提取、文档字符识别等过程提供帮助。在推理过程中，以 128 × 128 大小的低分辨率图像作为输入时，基于边缘感知的深度图像超分辨率网络达到了 75 FPS 的推理速度，同时在单张 Nvidia GTX Titan XP GPU 上共占用了 2840M 显存。

4.7 边缘信息的必要性研究

本节中我们进行了一些附加实验，用于验证文档图像超分辨率一些可选实施方案的效果，特别是边缘信息在整个网络训练过程中的作用。考虑到 RVL-DCIP Region 数据集相较于 ICDAR17-Textline 数据集有更大的数据量，和更丰富的多样性，我们在本节中的实验均在 RVL-DCIP Region 数据集上进行。

在基于边缘感知的深度图像超分辨率网络的训练过程中，使用了边缘损失作为损失函数的一部分，来保留图像中的边缘信息。正如在 4.3 所提到的，边缘信息的引入对整个网络的训练非常重要。我们进行了一些比较实验来说明边缘信息在整个文档图像超分辨率过程中的作用。图 4-8 和表 4-2(a) 展示了是否使用边缘损失对文档图像超分辨率效果的影响。级联的 DPNet 在不使用边缘损失时，最终的效果超过了 SRGAN 网络^[2]，表明了我们的级联网



图4-8 边缘信息对文档图像超分辨率的影响 (a) 低分辨率图像 ($4\times$ 退化); (b) 不使用边缘损失的超分辨率结果; (c) 使用边缘损失的超分辨率结果; (d) 原始高分辨率图像

络架构在文档图像上的有效性。另外，从实验结果中也能看出，在添加边缘损失后，图像的超分辨率效果有一定的提升，并且经过超分辨率处理后的文本区域有着更好的轮廓细节和更清晰的字符，这对后续的字符识别过程是很有帮助的。

我们还评估了基于边缘感知的深度图像超分辨率网络中各个构件对超分辨率结果的影响。图4-9和表4-2(b)展示了将DPNet换为双三次插值模块的实验结果比较。从实验数据上看，包含有DPNet构件的模型在不同的级联配置中表现最好。多阶段的DPNet相较于级联双三次插值模块和DPNet，在峰值信噪比上带来了10.5%的提升，并且极大的提升了文档图像的超分辨率结果，如图4-9。



图 4-9 不同级联结构的实验结果。从左至右：低分辨率图像（ $4\times$ 退化）；双三次插值和 DPNet 的 $2\times$ 超分辨率结果（ $2\times$ 退化）；双三次插值，双三次插值（ $2\times$ ）+ DPNet（ $2\times$ ），DPNet（ $2\times$ ）+ 双三次插值（ $2\times$ ），级联的 DPNet（ $4\times$ ）的超分辨率结果；原始高分辨率图像

4.8 本章小结

在本章中我们介绍了一个基于边缘感知的深度图像超分辨率网络。基于边缘感知的深度图像超分辨率网络是一个针对文档图像的深度超分辨率网络，在这个网络中使用了细节保留网络 DPNet 作为网络的基础构件，并通过级联 DPNet 的方式组件一个任意大小的超分辨率网络。每个单独的细节保留网络构件都具有一个较小的放大倍率，能够保留图像内容信息，同时增强图像中字符的边缘信息。多个级联的 DPNet 网络被集成为一个拥有更大放大倍率的网络模型。我们详细介绍了细节保留网络的基本结构，并定义了整体基于边缘感知的深度图像超分辨率网络的损失函数和训练过程。为了验证所提出方法在文档图像上的有效性，我们还在公开数据集 RVL-CDIP 上构建了我们自己的 RVL-CDIP Region 数据集，并使用公开发表的 ICDAR 2017 会议文档构建了 ICDAR17-Textline 数据集。通过在文档图像上的一系列实验，我们展示了级联的 DPNet 网络相较于基线的深度学习方法更加有效，能够从低分辨率的文档图像中生成足够有竞争力的结果。除此之外，我们还进行了一些附加实验，以展示边缘信息对整个网络超分辨率效果的影响，以及细节保留网络构件对文档图像超分辨率效果的提升。

第 5 章 基于边缘感知的图像去模糊

本章将详细介绍我们基于边缘感知的深度图像去模糊网络。首先，我们会介绍本算法所要解决的问题的描述和定义，并给出其数学表示，所用到的数学符号的定义也将在此部分进行说明。随后，我们将介绍整体网络架构的设计，包括边缘检测网络 EdgeNet 的结构以及去模糊网络 DeblurNet 的结构，以及两个子网络是如何进行连接的；在此之后，我们定义了网络训练的目标函数，以及训练和推理过程中所使用的不同策略。最后，我们在基准数据集上与其他去模糊方法进行了比较实验，并对实验结果进行了分析。

5.1 问题的描述与定义

本节我们将给出基于边缘感知的深度图像去模糊网络所要解决问题的描述和定义。

本文所要介绍的基于边缘感知的深度图像去模糊网络，是一个数字图像处理中的退化图像复原问题。在数字图像处理中，图像被表示为一个矩阵进行存储和运算。我们可将一幅图像定义为 $I \in \mathbb{R}^{H \times W \times C}$ ，其中 H 为图像的像素高度， W 为图像的像素宽度， C 为图像的通道数。在这种定义下，图像的处理问题就被转化为一个多维矩阵的运算和求解问题。在下面的讨论中，我们仅讨论离散形式的数字图像的运算和求解过程，相应的连续信号表示不进行详细介绍。图像的退化过程，可以被建模为由退化函数和噪声共同作用在原始图像上的过程，噪声可以是与原始图像信号无关的加性噪声，也可以是随原始图像信号变化而变化的噪声。例如，我们可将加性噪声的图像退化过程建模如下：

$$B = I \otimes K + N,$$

其中 B 是退化后的图像，在本文所介绍的应用场景中是一幅模糊图像， I 原始的图像信号， K 是以卷积核形式表示的退化函数， \otimes 是卷积运算。

图像复原问题，就是要求解图像退化的逆过程，即根据退化图像，通过下式恢复出原始图像信号：

$$I = (B - N) \otimes K^{-1},$$

假设我们已知了退化函数 K 和噪声 N ，则可以通过反卷积的方式从退化图像中复原出原始图像。然而，在实际应用场景中，退化函数 K 和噪声 N 往往是无法准确估计的，另外，退化函数 K 的维度通常也远小于图像信号，在这种情况下，想要精确的求解原始图像变得极为困难，且需要在一个较大的解空间中寻找出一个满意的解，退化图像的复原变成了一个病态的问题。

更为一般的情况，我们可以将图像退化模型建模为如下过程：

$$B = F(I),$$

$F(x)$ 是一个图像退化函数。此时图像的复原过程便转换为寻找一个函数 $G(x)$ ，使得 $I = G(B)$ 。考虑到神经网络就是通过不断的学习优化过程，来拟合一个未知的函数，则可在已

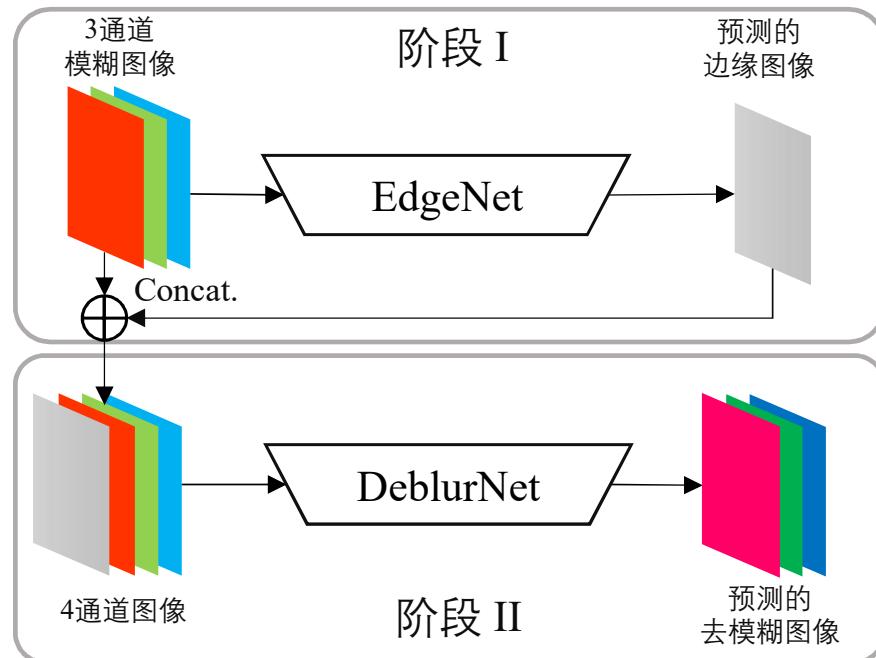


图 5-1 基于边缘感知的图像去模糊网络工作流程

知退化图像和原始图像样本对的情况下，通过深度学习算法，来拟合出一个函数 $G(x)$ ，从退化图像中还原出原始图像信息。

在本文中，我们设计了一个用于将模糊图像还原为清晰图像的卷积神经网络 EADNet，来作为图像去模糊过程中的还原函数，即

$$B = G_{EADNet}(I)$$

在我们的基于边缘感知的深度图像去模糊网络中，使用了图像边缘信息来辅助图像的去模糊过程，并包含有一个用于边缘检测的子网络 EdgeNet：

$$E = G_{EdgeNet}(I)$$

其中 E 是由边缘检测网络提取出的单通道边缘图像， $G_{EdgeNet}(x)$ 是边缘检测网络 EdgeNet。边缘检测网络从输入的模糊图像中提取出边缘图像信息后，边缘信息将与模糊图像一起输入用于去模糊的子网络 DeblurNet，得到最终经过去模糊处理后的清晰图像：

$$B = G_{DeblurNet}(I, E)$$

由此，我们基于边缘感知的深度图像去模糊网络可以表示为：

$$B = G_{DeblurNet} [I, G_{EdgeNet}(I)]$$

5.2 去模糊网络结构设计

人类识别区分一幅图像中的物体，是通过图像中的高频信息部分来完成的。图像的边缘轮廓信息是图像高频信息中非常具有代表性的一类特征。本文所描述的基于边缘感知的深

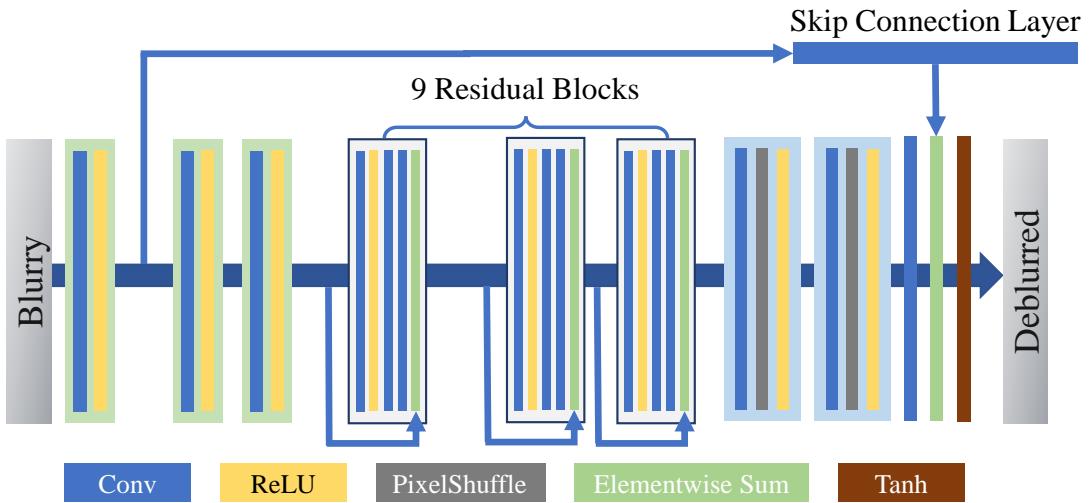


图 5-2 去模糊网络结构示意图

度图像去模糊网络 (Edge-Aware Deblur Network, EADNet)，主要任务是从模糊的退化图像中还原出原始的清晰图像，并且保持去模糊后的图像有清晰而锐利的边缘。本文所介绍方法分为两阶段进行处理，如图5-1，由两个子网络各自负责一个阶段。第一部分是用于提取图像边缘信息的边缘检测网络 EdgeNet，从模糊图像中预测物体的清晰轮廓和边缘信息，提取出的边缘信息将作为一个额外输入通道，与原始模糊图像一同参与下一阶段的去模糊过程；第二部分是利用提取出的边缘信息辅助，对模糊图像进行复原的去模糊网络 DeblurNet，接收一个由 3 通道模糊图像和单通道边缘图像组成的 4 通道输入，在对图像进行去模糊操作的同时，在边缘图像的辅助下增强图像的边缘细节信息，最终输出去模糊后的图像。边缘检测网络 EdgeNet 的具体结构设计和训练策略已在3.1 中详细介绍，在此不再进行赘述。本节将介绍去模糊网络 DeblurNet 的网络结构设计，并在下一节中介绍其训练策略和损失函数。

本文所介绍算法中的去模糊网络 DeblurNet 是一个多层的生成式卷积神经网络，其基本结构如图5-2。去模糊网络 DeblurNet 由 3 个卷积块，9 个残差块和 2 个上采样卷积块构成，同时还具有一个全局的跳跃连接层。另外，对于每个卷积层，我们都采用了权重归一化 (Weight Normalization)，以降低训练难度。

对于 DeblurNet 中的第一个卷积块，我们使用了一个较大的卷积核和较大的步长来提取与原始输入图像大小一致的低级 (low-level) 特征映射。此处所采用的卷积核大小为 9，步长为 4。随后的两个卷积块为降采样卷积块，每一个卷积块都会用一个较小的卷积核生成输入一半尺寸的大小的特征图，这两个卷积块所采用的卷积核大小为 3，步长为 2。对于所有的卷积块，都采用了修正线性单元 (Rectified linear unit, ReLU) 作为激活函数。

DeblurNet 中的残差块我们使用了 Jiahui Yu 等人在 CVPR NTIRE 2018 图像超分辨率比赛中提出的残差块架构设计 WDSR-B^[81]，并共计使用了 9 个残差块。在每个残差块中，首先由第一个卷积核大小为 1 的卷积层拓宽通道数，随后用修正线性单元进行激活。在此之后，一个卷积核大小为 1 的卷积层被用于减少通道数，并和一个用于空间特征提取的卷积核大小为 3 的卷积层组成一个高效的线性 low-rank 卷积组。

上采样卷积块由一个亚像素卷积层 pixel-shuffle 和一个卷积层组成，用于放大特征图 (feature map) 的尺寸，同时减少通道数。在整个 DeblurNet 的最后，我们还添加了一个全局

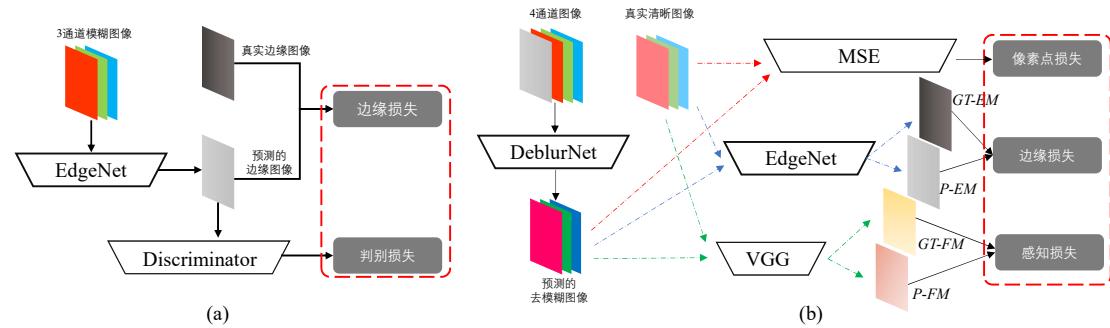


图 5-3 边缘检测网络 (a) 和去模糊网络 (b) 训练流程示意

的跳跃连接层来生成最终的输出结果。跳跃连接层是一个卷积核大小为 9 的卷积层，获取从第一个卷积块中得到的低级 (low-level) 特征，得到一个 3 通道的输出结果。同时，在经过 9 个残差块提取后得到的高级 (high-level) 特征，由两个上采样卷积块放大后，被输入另一个卷积核大小为 9 的卷积层处理，获取另一个 3 通道的输出。最终，我们将这两个卷积层输出的 3 通道结果进行像素点级别 (element-wise) 的求和运算，再由 Tanh 激活函数进行激活运算，得到最终的去模糊输出结果。

5.3 去模糊网络的训练

本文所介绍的基于边缘感知的深度图像去模糊网络分为边缘检测网络 EdgeNet 和去模糊网络 DeblurNet 两个子网络，两个子网络分成两个阶段独立进行训练。在第一个阶段中，我们将首先训练边缘检测网络 EdgeNet，用于提取模糊图像中的边缘信息。边缘检测网络 EdgeNet 采用生成式对抗网络的方式进行训练，其损失函数不仅使用了分类任务中常用的交叉熵，还引入了对抗学习损失，优化边缘检测效果。训练的第二个阶段，我们将利用已训练好的边缘检测网络 EdgeNet，首先预测出模糊图像的边缘信息，并将其作为额外的辅助信息，与模糊图像一起输入去模糊网络 DeblurNet 进行训练，同时边缘检测网络 EdgeNet 也将参与去模糊网络 DeblurNet 损失函数的计算，引导去模糊过程保留边缘信息。边缘检测网络 EdgeNet 和去模糊网络 DeblurNet 两部分网络的训练流程示意图如图5-3。图中 (a) 部分为边缘检测网络 EdgeNet 的训练流程，(b) 部分为去模糊网络 DeblurNet 的训练流程。其中，GT-EM 和 P-EM 分别表示真实的和预测的边缘图像；GT-FM 和 P-FM 分别表示真实的和预测的特征图 (feature map)。

去模糊网络 DeblurNet 的训练与边缘检测网络 EdgeNet 的训练直接相关。边缘检测网络 EdgeNet 不仅为去模糊网络 DeblurNet 提供模糊图像的边缘预测结果作为辅助输入，还将引导去模糊网络 DeblurNet 从模糊图像中生成具有清晰锐利边缘的去模糊结果。去模糊网络 DeblurNet 的训练过程如图5-3(b) 所示。

去模糊网络 DeblurNet 的损失函数由三部分构成，损失函数的定义如下：

$$L_{DeblurNet} = \lambda_1 \cdot L_{Pixel} + \lambda_2 \cdot L_{Perceptual} + \lambda_3 \cdot L_{Edge}$$

去模糊网络 DeblurNet 损失函数的第一项 L_{Pixel} 是一个像素内容损失，表示去模糊预测结果和真实清晰图像间的点对点均方误差 (Mean-Square Error, MSE)。第二项感知损失 $L_{Perceptual}$ ^[11] 是一个基于特征图 (Feature Map) 差异的损失项。感知损失的计算通过一个在 VGG-19 网络



图 5-4 GOPRO 数据集^[7]示例。左边为清晰图像，右边为模糊图像

上，用 ImageNet 数据集预训练好的模型来完成，通过比较生成图像和目标图像在这个 VGG-19 网络中，不同卷积层所输出特征图的差异，来表现生成图像在人类感知层面上的真实性。两幅特征图间的差异越小，表示生成的图像在人类的感知中越接近于真实的图像。感知损失的定义如下：

$$L_{Perceptual} = \frac{1}{W_j H_j} \sum_x \sum_y (\phi_j(I_C)_{x,y} - \phi_j(G_{\theta_D}(I_B))_{x,y})^2$$

其中 ϕ_j 表示 VGG-19 网络中的第 j 个卷积层， I_C 为真实的清晰图像。我们将像素内容损失和感知损失统称为内容损失，并且通过 L_2 来计算均方误差 MSE。

$L_{DeblurNet}$ 中的第三项是边缘损失，其计算过程与边缘检测网络 EdgeNet 中类似，都是一个基于类平衡的交叉熵损失。具体计算过程如下：

$$L_{Edge} = \sum_i^n \alpha_i \cdot l_{CBCE}(\phi_{side_i}(I_B), \phi_{side_i}(I_C)),$$

其中 α_i 是边缘检测网络 EdgeNet 第 i 个侧输出层的权重。我们将真实的清晰图像输入已训练好的边缘检测网络 EdgeNet 中，得到真实的边缘预测结果，再将去模糊网络 DeblurNet 输出的去模糊结果输入边缘测网络 EdgeNet 中，得到去模糊结果的边缘预测图像，最终计算这两幅边缘图像的类平衡交叉熵。实验中，我们取 $\alpha_1 = 0.7$, $\alpha_i = 0.1(i > 1)$ 。

5.4 实验数据的准备

我们在两个公开的去模糊算法基准数据集上进行了实验，并由此验证我们所设计的基于边缘感知的深度图像去模糊网络的有效性。两个基准数据集分别为 GOPRO^[7] 数据集和 Kohler^[8] 数据集。

GOPRO 数据集由 Nah 等人构建，通过使用高清摄像头快速而连续的拍摄一系列高清图像来获得清晰样本，并通过对同一批拍摄间隔非常短的图像求平均，模拟相机曝光时产生模糊的过程，来获取对应的模糊图像。数据集共包括 3214 对从不同场景中采集的模糊图像和清晰图像组成的样本对，如图 5-4。3214 对图像中，2103 对被划分为训练集，1111 对为测试集。清晰图像具有 720p 的清晰度，而模糊图像则是由清晰的视频图像所生成。Kohler 数据集是一个图像盲去模糊算法领域的一个标准基准测试集。Kohler 包括 4 张清晰的图像，对于每张清晰图像，都用 12 个不同的模糊核函数进行模糊化处理，如图 5-5。这些用于模糊处理的模糊核函数均来自于真实的摄像机运动记录数据分析，通过一个机器人运动平台上的数据采集过程，记录了摄像头的 6D 运动轨迹，从中生成了仿真的运动模糊核函数。

图 5-5 Kohler 数据集^[8]。左边为清晰图像，右边为模糊图像图 5-6 处理后的 COCO 数据集^[9]。从左至右分别为：原始图像，生成的模糊图像，边缘图像

对于一个边缘提取网络，通常而言，训练数据需要成对的输入图像与其对应的边缘图像。其中，输入图像数据一般是一幅清晰的图像，拥有较为清晰而锐利的边缘轮廓信息。但在去模糊的场景下，我们所设计的边缘检查网络 EdgeNet 的输入数据是一幅模糊的图像，并没有表现得很明显的边缘存在，物体的轮廓在梯度域上表现为梯度变换缓慢，相较于清晰图像梯度的变化更为平缓。当前可使用的公开数据集中，并没有数据集可以同时提供模糊图像和模糊图像对应的清晰边缘图像。这促使我们将现有的公开数据集进行处理，来获取我们训练边缘检测网络 EdgeNet 和去模糊网络 DeblurNet 所需要的训练数据。

为了保持我们所训练模型的去模糊能力，我们在模型的训练过程中使用了一个混合数据集。混合数据集由清晰图像，模糊图像，清晰图像产生的边缘图像三个部分构成，这些图像数据来自于 MS COCO^[9] 数据集和 GOPRO 数据集。GOPRO 数据集中已经包含有清晰图像数据和每一幅清晰图像所对应的模糊图像，我们仅使用了 Canny 算子^[17] 来获取这些清晰图像所对应的边缘图像。但由于 MS COCO 数据集中仅包含有清晰图像，因此我们从中选取了 2000 张具有锐利边缘的图像，并用 Canny 算子提取他们的边缘信息，得到对应的边缘图像，再对这些清晰的图像进行人工模糊化处理，如图 5-6。

我们随机选取了高斯模糊核函数和运动模糊核函数来从清晰的图像中生成模糊图像，以保证我们的算法在复杂图像场景中的适应性和鲁棒性。对于运动模糊图像的生成，我们采

用了 Boracchi 等人的方法^[82], 首先用马尔科夫过程 (Markov process) 生成一个随机向量来表示运动轨迹, 然后对生成的轨迹向量进行亚像素插值操作, 从中产生运动模糊所对应的模糊核函数。

需要特别说明的是, 使用 Canny 算子进行边缘提取, 提取结果的优劣非常依赖于人工设定的阈值, 同时很多被提取出的边缘并没有很好的连接在一起, 会出现很多零散的不连续断裂点。另外, Canny 检测出的边缘信息也会产生空间上的位移现象, 同时伴有不一致性。作为网络训练的真实数据 (ground truth), 由 Canny 提取出的边缘并不是一个很好的选择。但由于人工标注边缘的工作量太大, 耗时太长, 使用 Canny 提取边缘信息作为真实数据能够节省大量的时间, 同时, 实验结果也表明, Canny 算子提取出的边缘信息对于我们边缘检测网络 EdgeNet 的训练仍然是有帮助的。

5.5 训练环境及超参数的设定

对于本文所提出的基于边缘感知的深度图像去模糊网络中 EADNet 网络的训练, 我们在两个子网络中都使用了 Adam^[80] 作为训练的优化器, Adam 对应的参数选择为 $\beta_1 = 0.9$ 和 $\beta_2 = 0.999$ 。另外, 初始学习率 (Learning Rate) 我们设定为 0.0005, 并且在每 20 轮迭代后, 学习率衰减为原来的十分之一。训练数据的输入为从训练数据中随机裁剪出的, 大小为 256 * 256 的图像块 (patch), 并以 4 张模糊图像为一批 (batch) 进行训练。由于去模糊网络 DeblurNet 的训练依赖于边缘检测网络 EdgeNet, 因此我们首先对边缘检测网络 EdgeNet 进行了 50 轮迭代训练, 然后再用训练好的边缘检测网络 EdgeNet 辅助去模糊网络 DeblurNet 进行训练。

我们使用 PyTorch^[79] 深度学习库来构建我们的 EADNet 网络结构, 同时在 Intel Xeon E5 CPU 和 NVIDIA Titan X GPU 上进行模型的训练。

5.6 实验结果

我们在自己构建的混合数据集上对本文所介绍的基于边缘感知的深度图像去模糊网络进行了训练, 并在 GOPRO 测试集和 Kohler 数据集上分别验证了模型效果。同时, 也将我们所提出算法的性能与同期其他最先进算法 (state-of-the-art) 在峰值信噪比、结构相似性和运算时间上进行了比较。

5.6.1 评价标准

在图像去模糊领域, 除了人工用肉眼比较去模糊效果外, 常用的评价标准包括峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 和结构相似性 (Structural Similarity, SSIM)^[83]。

峰值信噪比 PSNR 是指信号本身最大可能功率与噪声功率的比值, 常用于表示信号的质量以及经过处理后信号的品质, 单位为分贝。在图像领域中, 峰值信噪比通常作为信号重建质量的一个评判标准。峰值信噪比越高, 表示图像的恢复效果越好。计算图像的峰值信噪比, 常通过计算均方误差 MSE 来实现。两幅图像的均方误差可由下式得出:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

其中 I 为原始图像, K 为处理后的图像, m 和 n 分别为图像的高和宽。而图像的峰值信噪比定义为:

$$PSNR = 10 \cdot \lg \left(\frac{MAX_I^2}{MSE} \right)$$

其中, MAX_I 表示原始图像 I 中的最大像素值。峰值信噪比在评价图像处理质量中应用广泛, 但仍然有其缺陷。由于人眼感知会受周围信息的干扰, 同时对于不同特性的信息敏感程度不同, 峰值信噪比的得分高低无法完全代表人类视觉感知的图像质量好坏。在大量应用场景中都存在峰值信噪比得分高的图像, 人类感知却认为其质量不如另一幅峰值信噪比得分更低的图像。

结构相似性指标 SSIM 用于评价两幅图像的相似程度。在图像的恢复重建领域中, 我们希望图像重建过程既能补全丢失的信息, 又能保持原始的结构不发生改变。结构相似性指标便是用于衡量恢复图像在结构上的失真程度, 值越高代表两幅图像的相似程度越高, 恢复效果越好。结构相似性指标从亮度、对比度和结构三个方面来衡量两幅图像的相似性, 它的定义如下:

$$SSIM(I, K) = l^\alpha(I, K) c^\beta(I, K) s^\gamma(I, K)$$

其中 $l(I, K)$ 比较图像 I 和 K 的亮度, $c(I, K)$ 比较对比度, $s(I, K)$ 比较结构; α, β, γ 为衡量三者重要性的权重参数。同时, 有:

$$\begin{aligned} l(I, K) &= \frac{2\mu_I\mu_K + C_1}{\mu_I^2 + \mu_K^2 + C_1} \\ c(I, K) &= \frac{2\sigma_I\sigma_K + C_2}{\sigma_I^2 + \sigma_K^2 + C_2} \\ s(I, K) &= \frac{\sigma_{IK} + C_3}{\sigma_I + \mu_K + C_3} \end{aligned}$$

其中, μ_I 和 μ_K 分别为 I 和 K 的均值, σ_I 和 σ_K 分别为标准差, σ_{IK} 为 I 和 K 的协方差, C_1, C_2, C_3 为常数。通常在实际使用中, 有 $\alpha = \beta = \gamma = 1$, 以及 $C_3 = \frac{C_2}{2}$ 。此时有:

$$SSIM(I, K) = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)}$$

无论峰值信噪比 PSNR 还是结构相似性指标 SSIM, 都是基于图像全局计算的评价标准。而人类感知评价图像恢复结果的好坏时, 也会考虑局部信息, 甚至由于一个局部的不够真实而对全图给出较低的评价。因此, 常会出现 PSNR 或 SSIM 得分高而人评价低的情况。仅用峰值信噪比和结构相似性指标来评判去模糊算法的有效性显得远远不够。

5.6.2 比较实验结果

我们将本文所介绍的基于边缘感知的深度图像去模糊网络与 Kim 等人^[84]、Sun 等人^[85]、Nah 等人^[7]以及 Tao 等人^[10]的算法进行了比较, 在 GOPRO 数据集上和 Kohler 数据集上分别比较了各自的峰值信噪比 PSNR 和结构相似性 SSIM, 所有的比较实验都在相同环境下进行。比较实验结果如表5-1所示。

从表5-1可以看出, 我们所提出的基于边缘感知的深度图像去模糊网络, 与同期的最先进算法相比, 无论在峰值信噪比 PSNR 上的得分, 还是结构相似性指标 SSIM, 都取得了比肩的效果。另外, 我们所提出的算法还具有更快的运算速度。

表 5-1 比较实验结果。加粗部分为最高得分

Method	GOPRO		Kohler		Time
	PSNR	SSIM	PSNR	MSSIM	
Kim ^[84]	23.64	0.8239	24.68	0.7937	1 hr
Sun ^[85]	24.64	0.8429	25.22	0.7735	20 min
Nah ^[86]	29.08	0.9135	26.48	0.8079	2.51 s
Tao ^[10]	30.10	0.9323	26.80	0.8375	0.67 s
EADNet	31.02	0.9123	26.91	0.8325	0.18 s

表 5-2 不同设置下的 EADNet 实验结果。加粗部分为最高得分

Method	GOPRO		Kohler	
	PSNR	SSIM	PSNR	MSSIM
EADNet w/o. EdgeNet	29.53	0.9014	25.97	0.8189
EADNet w. EdgeNet	30.78	0.9137	26.61	0.8297
EADNet w. Reduced EdgeNet	31.02	0.9123	26.91	0.8325

如上一节所提到的，用峰值信噪比和结构相似性指标都不能很好的代表去模糊算法的处理能力，因此我们还对去模糊图像的细节进行展示，以比较本文所述算法与其他算法的真实效果。我们选择了 Tao 等人^[10] 的算法进行比较。Tao 等人^[10] 用尺度循环网络 (scale-recurrent network) 从模糊图像中恢复出清晰图像。从图中可以看出，我们所提出的 EADNet 去除了大多数的模糊结构，和人造伪影，并且拥有更加鲜明而平滑的边缘。我们基于边缘感知的深度图像去模糊网络能够同时去除高斯模糊和运动模糊，同时保持图像中物体清晰的轮廓信息。与 Tao 等人^[10] 的方法相比，EADNet 中的去模糊网络 DeblurNet 学习到了生成更清晰图像的能力。

5.7 边缘提取网络的必要性研究

我们将在这个部分讨论基于边缘感知的深度图像去模糊网络中边缘提取网络 EdgeNet 的必要性，以及边缘信息的实际影响。首先，我们比较了不同阶段从图像中检测到的边缘结果，图5-2展示了去模糊后的边缘图像。在去模糊后的边缘图像中，像素亮度明显强于模糊图像中的边缘，并且具有更清晰的线条。同时，去模糊结果得到的边缘图像与清晰图像也更为相近。这种现象表明我们基于边缘感知的深度图像去模糊网络 EADNet 能够确实的生成具有清晰锐利边缘的去模糊图像。

我们还对没有边缘信息的情况进行了一系列的实验。在没有边缘信息的实验中，去模糊网络 DeblurNet 的边缘信息输入与边缘损失 EdgeLoss 均被移除。实验结果如表5-2中第一行所示。另外我们还试验了两种不同边缘信息的输入方式对去模糊结果的影响。第一种方式为完整的使用边缘检测网络 EdgeNet，采用 EdgeNet 的最终输出作为边缘检测结果；另一种方式采用一个更轻量的 Reduced EdgeNet 为去模糊网络 DeblurNet 提供边缘信息。轻量的边缘检测网络 Reduced EdgeNet 结构如图3-1中虚线范围所示，对于已训练好的 EdgeNet，在去模糊网络 DeblurNet 的训练过程中，我们仅取第 1 层侧输出层的边缘检测结果输入 DeblurNet。

图 5-7 Kohler 数据集上的比较结果。从左到右分别为: 输入图像, Tao^[10] 结果和 EADNet 结果

从表5-2中的峰值信噪比和结构相似性指标可以看出，采用 Reduced EdgeNet 的方案，能得到与采用完整边缘提取网络 EdgeNet 相当甚至更高的结果，同时大幅降低了计算资源消耗。另外，图5-10中是否使用边缘信息所产生的不同去模糊结果的比较也表明，在基于边缘感知的深度图像去模糊网络中，使用边缘检测网络 EdgeNet 能有效提高算法的去模糊效果。

5.8 本章小结

本章主要介绍了我们所提出的基于边缘感知的深度图像去模糊网络。首先我们给出了该算法所要解决问题的基本数学表示，随后详细介绍了我们所设计算法的基本网络结构，包括边缘检测网络 EdgeNet 和去模糊网络 DeblurNet 的设计和网络架构。另外，我们还给出了网络训练过程中目标损失函数的具体定义和计算方式，并总结了整个 EADNet 网络的算法流程。之后，还详述了我们基于边缘感知的深度图像去模糊网络在多个基准数据集上的测试结果。我们首先说明了本算法实验中所使用数据的来源，并对数据的处理过程进行了相应的描述。随后介绍了实验环境和一些超参数的选择。在具体的实验中，我们将本文所述算法与同期的其他最先进算法进行了横向比较，从峰值信噪比、结构相似度指标，以及视觉效

果三个层面展开对比，证明了所提出算法的有效性和效率。另外一些附加的实验还证明了边缘信息在我们的去模糊算法中起了重要的作用。

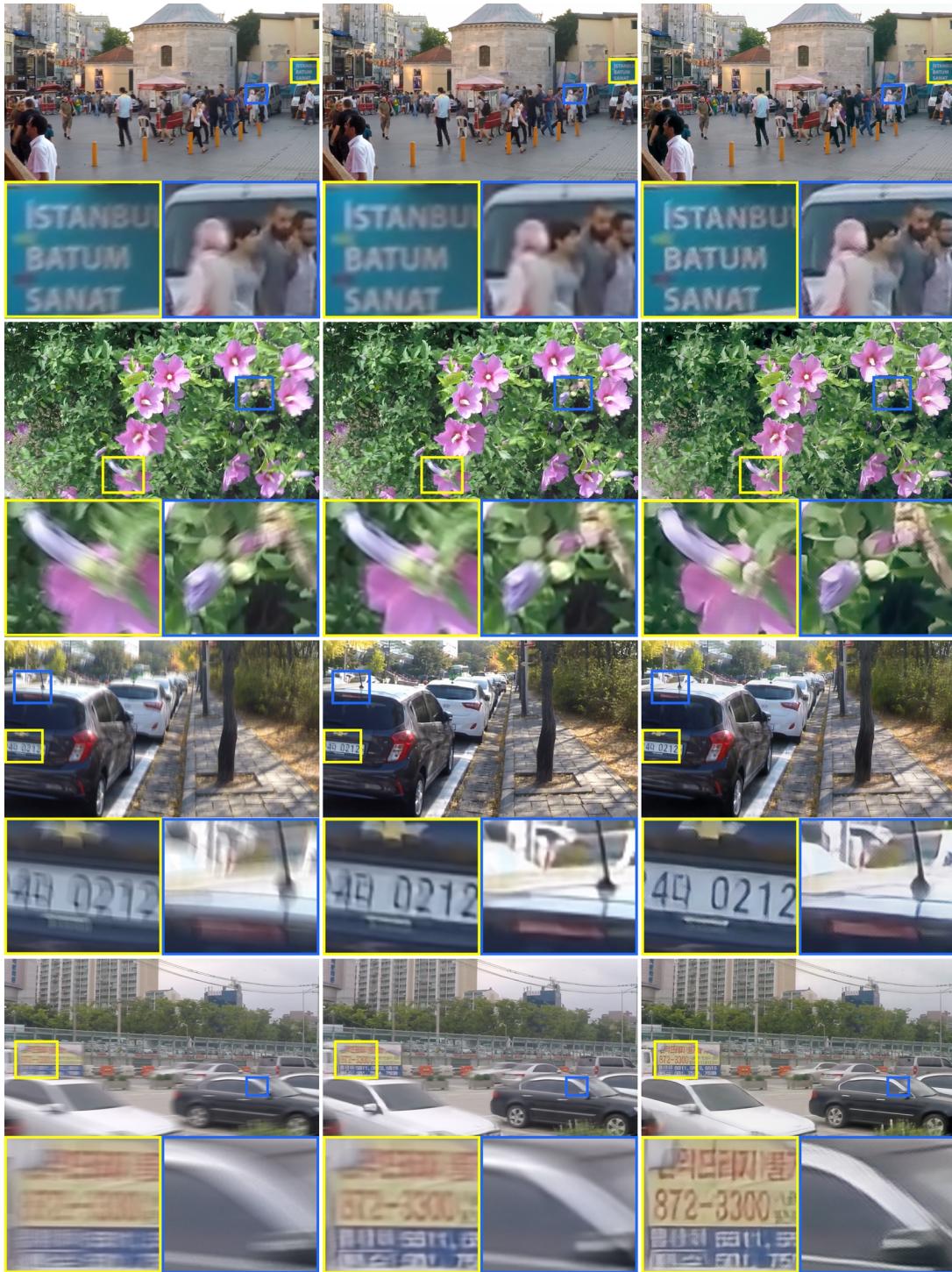


图 5-8 GoPro 数据集上的比较结果。从左到右分别为: 输入, Tao^[10] 结果和 EADNet 结果

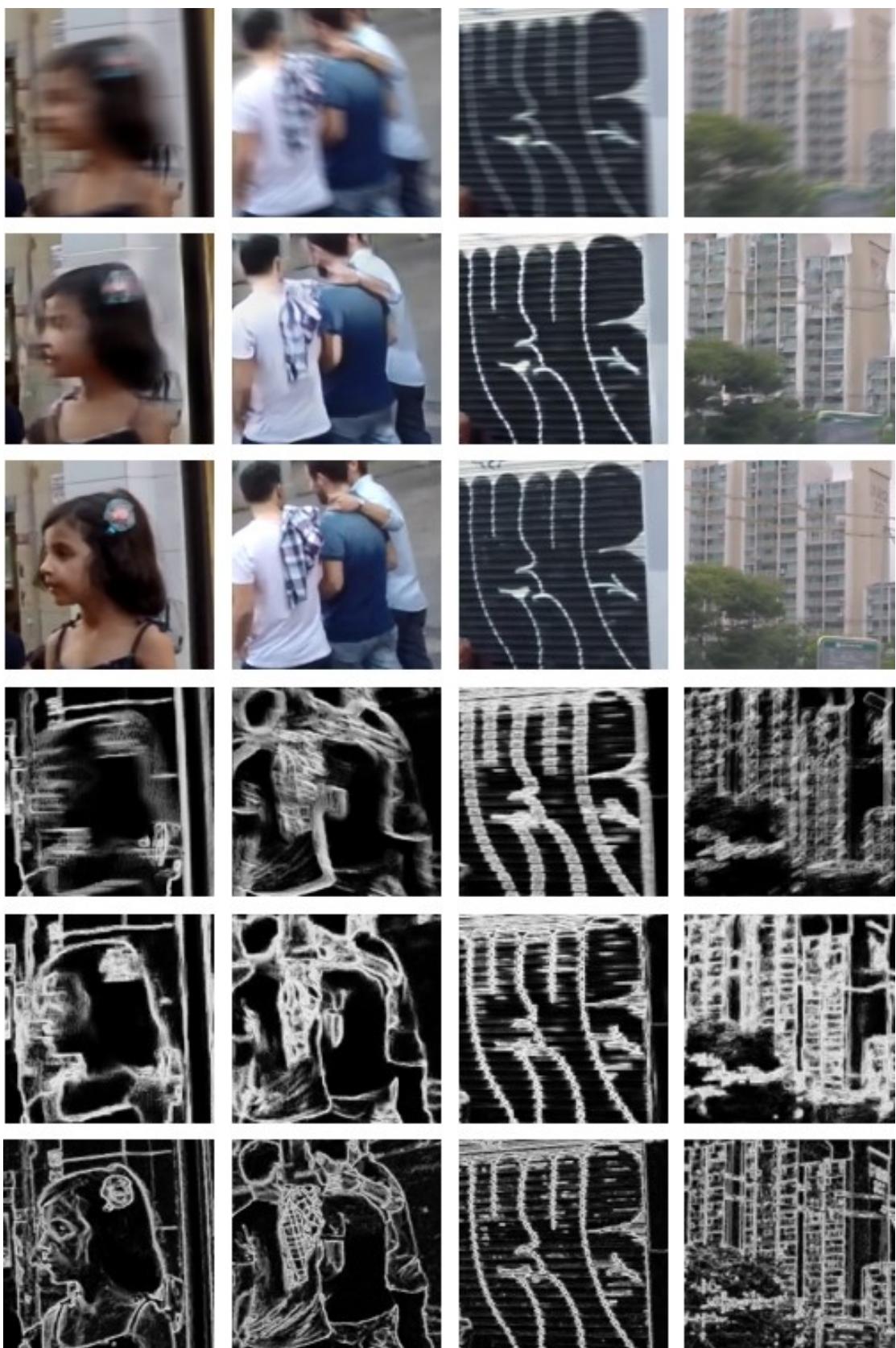


图 5-9 EADNet 模型处理的图像块实例。从上到下分别为：模糊图像，去模糊结果，清晰图像，模糊图像的 EdgeNet 边缘检测结果，去模糊后的边缘图像，清晰图像的边缘

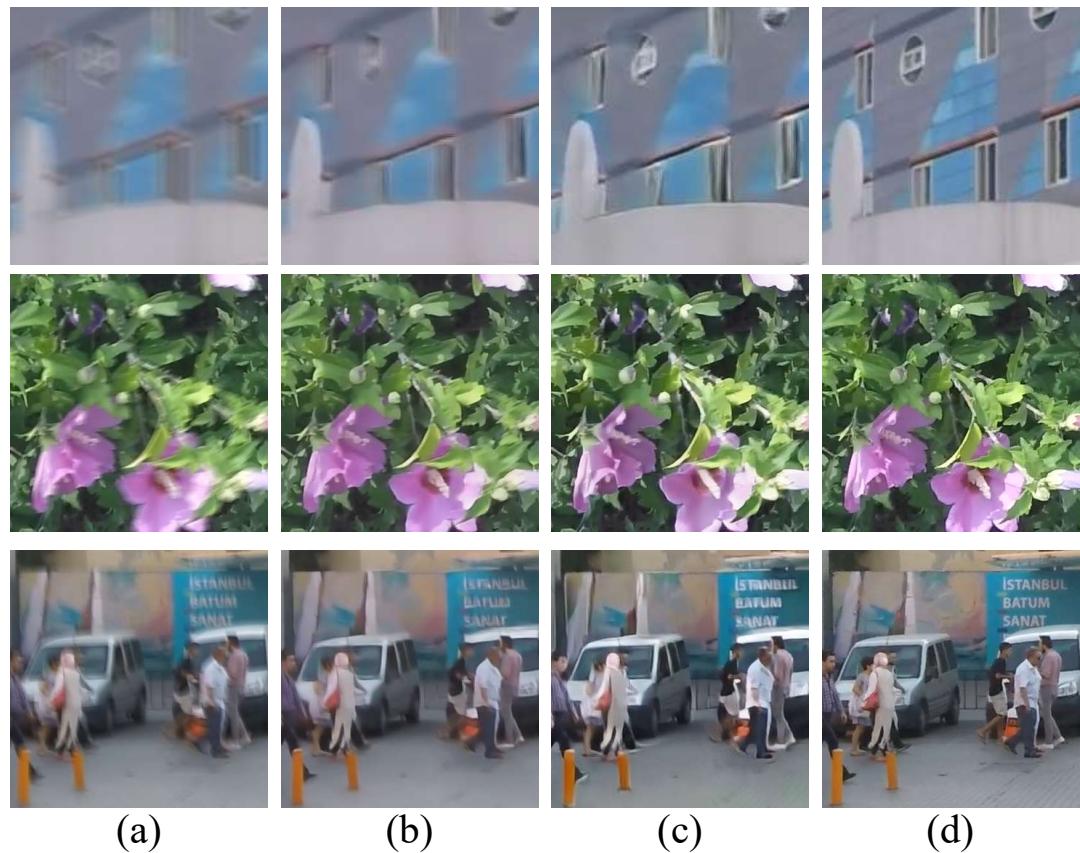


图 5-10 是否采用边缘信息的结果毕竟。(a) 模糊图像; (b) 没有边缘信息的去模糊结果; (c) 有边缘信息的去模糊结果; (d) 真实清晰图像

第 6 章 总结与展望

6.1 本文总结

本文的研究重点在于探索边缘感知信息在图像重建中的研究与应用。图像退化现象大量出现在我们的日常生活中如何从退化的图像中，很好的恢复出原始图像，是一个重要而富有挑战性的课题，退化的图像不仅难以辨认，还会对后去的图像应用带来影响，特别是针对于模式识别等计算机视觉处理，退化图像中的识别精度将会大幅降低。在图像退化过程中，我们往往很难确定导致图像退化的具体因素，即使针对特定的图像退化场景进行建模，模型中导致图像退化的核函数也是难以假设或估计的。在图像重建领域，已经有相当多的理论研究，致力于解决不同特性的退化图像重建问题，如提高图像分辨率的超分辨率处理，还原模糊图像的图像去模糊。但是，图像重建是一个病态的问题，一幅退化的图像可以对应于多个合理的清晰图像解，这为图像重建工作来了很大的挑战。我们探索了利用边缘感知信息来辅助图像重建过程。首先我们引入了一个边缘检测网络，用于从退化的图像中提取图像的边缘信息。我们详细介绍了边缘检测网络的构建方式，并对其训练过程进行了介绍。随后，针对低分辨率的退化文档图像，我们设计了一个基于边缘感知的深度图像超分辨率网络，用于对低分辨率退化图像进行超分辨率处理，采用级联的方式，逐步对文档图像进行放大。而对于模糊的退化图像，我们使用边缘信息进行辅助，将提取出的边缘信息与模糊图像一起进行去模糊过程，还原出清晰的图像。在多个数据集上的实验表明，在图像重建任务中引入边缘信息，特别是对文档图像超分辨率任务和图像去模糊任务，能够有效的提高图像超分辨率和图像去模糊的效果，并在保留图像的边缘细节信息的同时，增强边缘的清晰度和丰富度，生成具有清晰锐利边缘的重建图像。同时，我们所设计的两个图像重建网络在各自的实验中都取得了与当前最先进算法相当甚至更优的结果。

在第二章中，我们对当前图像边缘检测算法进行了回顾，包括早期基于梯度的边缘检测算子，以及后期基于统计、信息论基础，甚至深度神经网络的图像边缘检测方法。随后介绍了生成式对抗网络以及相关的研究进展。在此之后，我们针对图像超分辨率和图像去模糊这两个具体的图像重建应用场景，分别概述了他们研究现状。其中，对于图像的超分辨率领域，我们简要介绍了传统方法中常用的解决方案，同时比较了多个基于神经网络的图像超分辨率算法；对于图像去模糊领域，我们分析模糊图像的成因并对其进行建模，然后从图像的非盲去模糊和图像的盲去模糊两个方向分别介绍了各自的研究进展，包括用传统方法进行清晰图像求解的相关算法，和用深度学习代替部分求解步骤的方法，还有直接利用卷积神经网络端到端的还原出清晰图像的方案。

在第三章中，我们详细介绍了本文中所使用的边缘检测网络。本文后续应用中所涉及的边缘信息，均来自于本章介绍的边缘检测算法。本文中所使用的边缘检测网络，目的是从一幅图像中提取出清晰的边缘信息，来辅助其他研究和应用，特别是需要具备从退化图像中检测出清晰边缘的能力。我们首先对边缘检测网络的网络结构进行介绍，详细描述了神

经网络中每一层的大小以及作用，然后介绍了该网络的具体训练步骤，包括损失函数的构建和训练超参数的选取。

我们在第四章探索了边缘感知信息在图像超分辨率中的应用。本章中我们构建了一个基于边缘感知的深度图像超分辨率网络，用于将退化的低分辨率文档图像进行超分辨率处理，在提高图像清晰度的同时，还保留了文档中字符的结构信息和清晰的笔画边缘，提升了后续OCR系统的识别精度。基于边缘感知的深度图像去模糊网络由多个级联的子网络组成，每个子网络共享同样的网络结构，但具有不同的参数。对于单个子网络来说，能够以较小的倍率放大图像，提高图像的分辨率。而多个子网络进行级联后，能够以较大的放大倍率对输入图像进行超分辨率处理。我们详细介绍了子网络的结构，还有如何用多个子网络搭建完整的超分辨率网络。之后解释了网络的训练流程，包括两阶段训练方式的设计和损失函数的构建。最后我们在自己搭建的数据集上进行了相关实验，验证了我们所设计方法的有效性，并与同期其他超分辨率网络进行了比较。需要特别注意的是，我们在网络的损失函数中引入了边缘损失项，用第三章所介绍的边缘检测网络，分别对真实清晰图像和网络输出进行边缘检查，并将两者检测结果的差异作为损失反向传播，以提升超分辨率网络对图像边缘细节保留的能力，促使网络生成具有清晰边缘的超分辨率结果。

在第五章中，我们针对图像去模糊问题，提出了一个基于边缘感知度的深度图像去模糊网络，并对该网络进行了详细的介绍。首先我们简单的描述了去模糊问题及去模糊算法的数学表示。之后，对我们所提出的网络结构进行了详细的描述，包括使用第三章中的边缘检测网络作为去模糊网络中的一个子网络，网络整体所采用的设计方案，以及目标函数的定义。同时，还介绍了该算法的具体训练流程，对如何分别在两阶段中训练两个子网络进行了描述。随后，我们在两个公开的基准测试集上进行了验证，给出了相应的实验结果和输出示例。另外，我们还与当前几个最先进的算法进行了比较，从峰值信噪比、结构相似度指标以及视觉认知三个层面进行对比，展示了我们所提出算法的有效性。之后，我们还进行了一些附加实验，证明了边缘信息在我们去模糊算法中扮演了重要的角色。

6.2 今后工作展望

本文针对退化图像的重建问题进行了一些探索和研究，尝试了使用图像的边缘感知信息来辅助图像重建任务，并在两个具体应用中进行了相关实验。实验表明，边缘感知信息的引入，能够有效的提升图像超分辨率和图像去模糊的效果，在重建出清晰图像的同时，较好的保留并增强了退化图像中的边缘信息，生成出了具有清晰锐利边缘的重建结果。然而，我们的工作仍然存在很多不足之处，需要进一步完善。我们将目前发现的一些缺点总结如下：

1. 本文虽然验证了边缘信息对于图像的重建任务具有一定的提升效果，但并没有对边缘信息的质量和最终效果提升的程度进行进一步分析，仅简单的使用Canny算子检测出的边缘作为边缘检测网络的目标输出。由于Canny算子检测的边缘并不完整，边缘检测网络的训练结果并不完美，我们无法更有力的证明一个高质量的边缘信息是否能带来更大的提升，或者是带来相反的效果；
2. 边缘检测网络在纹理复杂的区域中，如密集的树丛，不能很好的检测出物体边缘，检测结果有太多的不连续细节，影响了后续试验的效果；

3. 对于所提出的文档图像超分辨率网络，在图像中字符较小，或者多个字符相距过近时，超分辨率结果并不理想，这部分字符区域即使经过了超分辨率处理，也仍然呈现出低分辨率的模糊状态，需要进一步优化算法以提升网络在这种情况下的性能；
4. 在图像去模糊网络中，虽然无论高斯模糊还是运动模糊都得到了一定的抑制，还原出了较为清晰的结果，但对于模糊区域还是有肉眼可见较为明显的模糊痕迹，图像的实际去模糊能力还需要进一步提高。另外，对于纹理复杂而密集的区域，由于边缘检测的结果不理想，这部分区域的实际去模糊效果不佳，输出结果仍然很模糊；
5. 尽管实验结果表明，我们所提出的网络在峰值信噪比和结构相似度指标上都有不错的成绩，但这两个评价标准并不能很好的说明说明算法的实际效果。还需要设计一个更为可靠的指标来评价我们算法得出的结果。

参考文献

- [1] C. DONG C C L, TANG X. Accelerating the super-resolution convolutional neural network [C]//European Conference on Computer Vision. [S.l.: s.n.], 2016.
- [2] C. LEDIG F H, L. Theis. Photo-realistic single image super-resolution using a generative adversarial network[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2017.
- [3] L. XU C L J J, J. S. Ren. Deep convolutional neural network for image deconvolution[C]// Advances in Neural Information Processing Systems. [S.l.: s.n.], 2014.
- [4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations. [S.l.: s.n.], 2015.
- [5] HARLEY A W, UFKES A, DERPANIS K G. Evaluation of deep convolutional nets for document image classification and retrieval[C]//International Conference on Document Analysis and Recognition. [S.l.: s.n.], 2015.
- [6] C. DONG K H, C. C. Loy, TANG X. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2):295-307.
- [7] NAH S, KIM T H, LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2017.
- [8] KöHLER R, HIRSCH M, MOHLER B, et al. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database[M]. [S.l.: s.n.], 2012: 27-40.
- [9] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// European Conference on Computer Vision. [S.l.: s.n.], 2014: 740-755.
- [10] TAO X, GAO H, WANG Y, et al. Scale-recurrent network for deep image deblurring[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [11] J. JUSTIN F F L, A. Alexandre. Perceptual losses for real-time style transfer and super-resolution[C]//European Conference on Computer Vision. [S.l.: s.n.], 2016: 694-711.
- [12] CHAN T F, WONG C K. Total variation blind deconvolution[M]. [S.l.]: IEEE Press, 1998: 370-375.

- [13] H.G. BARROW J T. Interpreting line drawings as three-dimensional surfaces[J]. *Artificial Intelligence*, 1981, 17(1-3):75–116.
- [14] UMBAUGH S E. Digital image processing and analysis : human and computer vision applications with cviptools (2nd ed.)[M]. [S.l.]: Boca Raton, FL: CRC Press, 2010.
- [15] R. DUDA P H. Pattern classification and scene analysis[M]. [S.l.]: John Wiley and Sons, 1973: 271-272.
- [16] PREWITT J M S. Object enhancement and extraction[J]. *Picture Processing and Psychopictorics*, 1970:75-149.
- [17] CANNY J. A computational approach to edge detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986(6):679-698.
- [18] SCOTT KONISHI J M C S C Z, A. L. Yuille. Statistical edge detection: learning and evaluating edge cues[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(1): 57-74.
- [19] D.R. MARTIN J M, C.C. Fowlkes. Learning to detect natural image boundaries using local brightness, color, and texture cues[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(1):530-549.
- [20] MALIK P A M M C F J. Contour detection and hierarchical image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 33(5):898-916.
- [21] P. DOLLAR S B, Z. Tu. Supervised learning of edges and object boundaries[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2006.
- [22] REN X. Multi-scale improves boundary detection in natural images[C]//European Conference on Computer Vision. [S.l.: s.n.], 2008.
- [23] J. J. LIM P D, C. L. Zitnick. Sketch tokens: A learned mid-level representation for contour and object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2013.
- [24] P. DOLLAR C L Z. Fast edge detection using structured forests[C]//Pattern Analysis and Machine Intelligence. [S.l.: s.n.], 2015.
- [25] YAROSLAV GANIN V L. N4-fields: Neural network nearest neighbor fields for image transforms[C]//Asian Conference on Computer Vision. [S.l.: s.n.], 2014.
- [26] WEI SHEN Y W, Xinggang Wang. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2015.

- [27] G. BERTASIUS L T, J. Shi. Deepedge: A multiscale bifurcated deep network for top-down contour detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2015.
- [28] XIE S, TU Z. Holistically-nested edge detection[J]. International Journal of Computer Vision, 2015, 125(1-3):1-16.
- [29] I. GOODFELLOW M M, J. Pouget-Abadie. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. [S.l.: s.n.], 2014.
- [30] MARTIN ARJOVSKY L B. Towards principled methods for training generative adversarial networks[C]//International Conference on Learning Representations. [S.l.: s.n.], 2017.
- [31] MARTIN ARJOVSKY L B, Soumith Chintala. Wasserstein gan[C]//International Conference on Machine Learning. [S.l.: s.n.], 2017: 214-223.
- [32] ISHAAN GULRAJANI M A, Faruk Ahmed. Improved training of wasserstein gans[J]. ArXiv e-prints, 2017.
- [33] A. RADFORD S C, L. Metz. Unsupervised representation learning with deep convolutional generative adversarial networks[C]//International Conference on Learning Representations. [S.l.: s.n.], 2016.
- [34] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. ArXiv, 2014, abs/1411.1784.
- [35] P. ISOLA T Z, JY. Zhu. Image-to-image translation with conditional adversarial networks[C]// International Conference on Learning Representations. [S.l.: s.n.], 2016.
- [36] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//International Conference on Computer Vision. [S.l.: s.n.], 2017.
- [37] KIM T, CHA M, KIM H, et al. Learning to discover cross-domain relations with generative adversarial networks[C/OL]//ICML'17: International Conference on Machine Learning. JMLR.org, 2017: 1857-1865. <http://dl.acm.org/citation.cfm?id=3305381.3305573>.
- [38] OREST KUPYN M M, Volodymyr Budzan. Deblurgan: Blind motion deblurring using conditional adversarial networks[J]. ArXiv e-prints, 2017.
- [39] CHEN X, DUAN Y, HOUTHOOFD R, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets[C]//NIPS. [S.l.: s.n.], 2016.
- [40] TOMER P, MICHAEL E. A statistical prediction model based on sparse representations for single image super-resolution[J]. IEEE Transactions on Image Processing, 2014, 23(6):2569-82.

- [41] JIAN S, XU Z, SHUM H Y. Image super-resolution using gradient profile prior[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2008.
- [42] EFRAT N, GLASNER D, APARTSIN A, et al. Accurate blur models vs. image priors in single image super-resolution[C]//International Conference on Computer Vision. [S.l.: s.n.], 2013.
- [43] FERNANDEZ-GRANDA C, CANDES E J. Super-resolution via transform-invariant group-sparse regularization[C]//International Conference on Computer Vision. [S.l.: s.n.], 2013.
- [44] QIANG W, TANG X, SHUM H. Patch based blind image super resolution[C]//International Conference on Computer Vision. [S.l.: s.n.], 2005.
- [45] AODHA O M, CAMPBELL N D F, NAIR A, et al. Patch based synthesis for single depth image super-resolution[C]//European Conference on Computer Vision. [S.l.: s.n.], 2012.
- [46] JIA-BIN HUANG A S, AHUJA N. Single image super-resolution from transformed self-exemplars[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2015.
- [47] FREEMAN W T, JONES T R, PASZTOR E C. Example-based super-resolution[J]. IEEE Computer Graphics and Applications, 2002, 22(2):56-65.
- [48] RADU TIMOFTE V D S, GOOL L V. A+: Adjusted anchored neighborhood regression for fast super-resolution[C]//Asian Conference on Computer Vision. [S.l.: s.n.], 2014.
- [49] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2016.
- [50] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:1637-1645.
- [51] GOLDSTEIN A, FATTAL R. Blur-kernel estimation from spectral irregularities[J]. European Conference on Computer Vision, 2012, 7576(1):622-635.
- [52] SZELISKI R. Computer vision: Algorithms and applications[M]. [S.l.]: Springer-Verlag New York, Inc., 2010: 2601-2605.
- [53] WIENER N. Extrapolation, interpolation, and smoothing of stationary time series[M]. [S.l.]: MIT press Cambridge, MA, 1949: volume 7.
- [54] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1):35-45.
- [55] A. LEVIN F D W T F, Y. Weiss. Understanding and evaluating blind deconvolution algorithms [C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2009.

- [56] D. KRISHNAN R F. Fast image deconvolution using hyper-laplacian priors[C]//Advances in Neural Information Processing Systems. [S.l.: s.n.], 2009.
- [57] D. ZORAN Y W. From learning models of natural image patches to whole image restoration [C]//International Conference on Computer Vision. [S.l.: s.n.], 2011.
- [58] S. CHO S L, J. Wang. Handling outliers in non-blind image deconvolution[C]//International Conference on Computer Vision. [S.l.: s.n.], 2011.
- [59] A. DANIELYAN K E, V. Katkovnik. Bm3d frames and variational image deblurring[J]. IEEE Transactions on Image Processing, 2012, 21(4):1715–1728.
- [60] A. KHERADMAND P M. A general framework for regularized, similarity-based image restoration[J]. IEEE Transactions on Image Processing, 2014, 23(12):5136–5151.
- [61] R. FERGUS A H S T R W T F, B. Singh. Removing camera shake from a single photograph [J]. ACM Transactions on Graphics, 2006, 25:787–794.
- [62] L. XU J J. Two-phase kernel estimation for robust motion deblurring[C]//European Conference on Computer Vision. [S.l.: s.n.], 2010.
- [63] S. CHO S L. Fast motion deblurring[J]. ACM Transactions on Graphics, 2009, 28:145.
- [64] D. KRISHNAN R F, T. Tay. Blind deconvolution using a normalized sparsity measure[C]// IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2011.
- [65] O. WHYTE A Z, J. Sivic. Deblurring shaken and partially saturated images[J]. International Journal of Computer Vision, 2014, 110(2):185–201.
- [66] J. PAN H P M H Y, D. Sun. Blind image deblurring using dark channel prior[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2016.
- [67] W.-S. LAI Z H N A M H Y, J.-B. Huang. A comparative study for single image blind deblurring [C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2016.
- [68] JIAN SUN Z X J P, Wenfei Cao. Learning a convolutional neural network for non-uniform motion blur removal[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2015.
- [69] D. EIGEN R F, D. Krishnan. Restoring an image taken through a window covered with dirt or rain[C]//International Conference on Computer Vision. [S.l.: s.n.], 2013.
- [70] KAIMING HE S R J S, Xiangyu Zhang. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2016.
- [71] J. LONG T D, E. Shelhamer. Fully convolutional networks for semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2015.

- [72] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C/OL]//ICML'15: Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37. JMLR.org, 2015: 448-456.
<http://dl.acm.org/citation.cfm?id=3045118.3045167>.
- [73] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[J]. International Conference on Learning Representations, 2018.
- [74] XU B, WANG N, CHEN T, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv:1505.00853, 2015.
- [75] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2016.
- [76] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//International Conference on Computer Vision. [S.l.: s.n.], 2015.
- [77] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3):211-252.
- [78] MECHREZ R, TALMI I, ZELNIK-MANOR L. The contextual loss for image transformation with non-aligned data[C]//European Conference on Computer Vision. [S.l.: s.n.], 2018: 768-783.
- [79] PASZKE A, GROSS S, CHINTALA S, et al. Automatic differentiation in pytorch[C]//NIPS Workshop. [S.l.: s.n.], 2017.
- [80] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]//International Conference on Learning Representations. [S.l.: s.n.], 2015.
- [81] YU J, FAN Y, YANG J, et al. Wide activation for efficient and accurate image super-resolution [Z]. [S.l.: s.n.], 2018.
- [82] BORACCHI G, FOI A. Modeling the performance of image restoration from motion blur[J]. IEEE Transactions on Image Processing, 2012, 21(8):3502-3517.
- [83] 宋刚王平 刘革新 韩慧健. 纹理细节保持的图像插值算法[J]. 计算机科学, 2019, 46(S1): 169-176.
- [84] KIM T H, AHN B, LEE K M. Dynamic scene deblurring[J]. International Conference on Computer Vision, 2013:3160-3167.
- [85] SUN J, CAO W, XU Z, et al. Learning a convolutional neural network for non-uniform motion blur removal[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2015:769-777.

- [86] NAH S, KIM T H, LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 257-265.

致谢

韶光易逝，年华易老，转眼间两年半的研究生生涯就将告一段落。两年多以来的求学时光仿如一瞬，初入复旦时校门的迷茫和喜悦都还历历在目。在各位老师的引领下，在各位同学的帮助下，两年来我接触到了很多当下最前沿的技术，对深度学习和智能媒体计算领域有了一定的了解。攻读硕士学位期间期间，复旦大学计算机科学技术学院与我们的导师薛向阳老师为我们提供了宽松而完备的科研环境与多样的前沿学术交流机会。同时，薛向阳老师和李斌老师等人也以自己丰富的学术实践经验，始终如一地支持我进行科研工作，为我探寻未知的旅程照亮了前路。我从最开始对图像处理和深度学习的一知半解，到后来通过阅读学术界的经典论文，系统的学习智能媒体计算方向的相关知识，再到自己开始动手实验，并选择图像生成方向作为今后的研究目标，最终在图像重建任务中实现了一些具体应用，这一路走来离不开各位老师的悉心教导。

感谢薛向阳老师用心栽培，带领我进入了计算机视觉和深度学习领域，以他站在学术前沿的视角用心指导我的成长。薛向阳老师严谨的治学态度深深影响了我，时刻提醒我不论在工作中还是生活上都要做到求真求实。除此之外，薛向阳老师还为我提供了充足的实际机会，让我能够学以致用，立足于社会实际需求来思考问题，使我获益良多。

感谢李斌老师在学术研究上的指导，从不同的角度指出我研究工作中存在的问题并给出建议，扩宽了我的眼界。感谢郑莹斌博士和叶浩博士的支持，带领我从入门开始不断实践自己的想法。

感谢我的同学周钊、马建奇、王丽和傅冰飞等人，他们对我学习和工作上遇到的问题知无不言，我从他们身上学到了很多进行科学的研究和项目实践的方法和精神，陪伴我在两年多的研究生生涯中不断成长。

感谢复旦大学两年来的培养，为我学习生活的各个方面提供了充足的条件，举办的各类学术交流也开拓着我的眼界，让我能不断的接触世界一流的技术，不断的成长。

复旦大学 学位论文独创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。论文中除特别标注的内容外，不包含任何其他个人或机构已经发表或撰写过的研究成果。对本研究做出重要贡献的个人和集体，均已在论文中作了明确的声明并表示了谢意。本声明的法律结果由本人承担。

作者签名：_____ 日期：_____

复旦大学 学位论文使用授权声明

本人完全了解复旦大学有关收藏和利用博士、硕士学位论文的规定，即：学校有权收藏、使用并向国家有关部门或机构送交论文的印刷本和电子版本；允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其它复制手段保存论文。涉密学位论文在解密后遵守此规定。

作者签名：_____ 导师签名：_____ 日期：_____