国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

“十四五”国家重点图书出版规划项目
图像图形智能处理理论与技术前沿

DEEP LEARNING-BASED IMAGE SUPER-RESOLUTION RECONSTRUCTION TECHNOLOGY AND APPLICATIONS

基于深度学习的图像超分辨率重建技术及应用

张浩鹏　编著

清華大學出版社
北京

图书在版编目（CIP）数据

基于深度学习的图像超分辨率重建技术及应用 / 张浩鹏编著. -- 北京 : 清华大学出版社, 2026. 6. --（图像图形智能处理理论与技术前沿）. -- ISBN 978-7-302-71972-4

Ⅰ. TP391.413

中国国家版本馆 CIP 数据核字第 20269TG723 号

责任编辑：刘　杨
封面设计：钟　达
责任校对：赵丽敏
责任印制：丛怀宇

出版发行：清华大学出版社
网　　址：https://www.tup.com.cn, https://www.wqxuetang.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-83470000　　邮　　购：010-62786544
投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn
印 装 者：涿州市般润文化传播有限公司
经　　销：全国新华书店
开　　本：170mm × 240mm　　**印　张：**13.75　　**字　数：**275 千字
版　　次：2026 年 6 月第 1 版　　**印　次：**2026 年 6 月第 1 次印刷
定　　价：65.00 元

产品编号：102995-01

丛书编委会名单

丛书序

“人工智能是我们人类正在从事的、最为深刻的研究方向之一，甚至要比火与电还更加深刻。”正如谷歌CEO 桑达尔·皮查伊所说，“智能”已经成为当今科技发展的关键词。而在智能技术的高速发展中，计算机图像图形处理技术与计算机图形学犹如一对默契的舞伴，相辅相成，为社会进步做出了巨大的贡献。

图像图形智能处理技术是人工智能研究与图像图形处理技术的深度融合，是一种数字化、网络化、智能化的技术。随着新一轮科技革命的到来，图像图形智能处理技术已经进入了一个高速发展的阶段。在计算机、人工智能、计算机图形学、计算机视觉等技术不断进步的同时，图像图形智能处理技术已经实现了从单一领域到多领域的拓展，从单一任务到多任务的转变，从传统算法到深度学习的升级。

图像图形智能处理技术被广泛应用于各个行业，改变了公众的生活方式，提高了工作效率。如今，图像图形智能处理技术已经成为医学、自动驾驶、智慧安防、生产制造、游戏娱乐、信息安全等领域的重要技术支撑，对推动产业技术变革和优化升级具有重要意义。

在《新一代人工智能发展规划》的引领下，人工智能技术不断推陈出新，人工智能与实体经济深度融合成为重要的战略目标。智慧城市、智能制造、智慧医疗等领域的快速发展为图像图形智能处理技术的研究与应用提供了广阔的发展和应用空间。在这个背景下，为国家人工智能的发展培养与图像图形智能处理技术相关的专业人才已成为时代的需求。

当前在新一轮科技革命和产业变革的历史性交汇中，图像图形智能处理技术正处于一个关键时期。虽然图像图形智能处理技术已经在很多领域得到了广泛应用，但仍存在一些问题，如算法复杂度、数据安全性、模型可解释性等，这也对图像图形智能处理技术的进一步研究和发展提出了新的要求与挑战。这些挑战既来自技术的不断更新和迭代，也来自人们对于图像图形智能处理技术的不断追求和探索。如何更好地提高图像的视觉感知质量，如何更准确地提取图像中的特征信息，如何更科学地对图像数据进行变换、编码和压缩，成为国内外科技工作者和创新企业竞相探索的新方向。

为此，中国图像图形学会和清华大学出版社共同策划了“图像图形智能处理理论与技术前沿”系列丛书。丛书包括21个分册，以图像图形智能处理技术为主

线，涵盖了多个领域和方向，从智能成像与感知、智能图像图形处理技术、智能视频分析技术、三维视觉与虚拟现实技术、视觉智能应用平台等多个维度，全面介绍该领域的最新研究成果、技术进展和应用实践。编写本丛书旨在为从事图像图形智能处理研究、开发与应用的人员提供技术参考，促进技术交流和创新，推动我国图像图形智能处理技术的发展与应用。本丛书将采用传统出版与数字出版相融合的形式，通过二维码融入文档、音频、视频、案例、课件等多种类型的资源，帮助读者进行立体化学习，加深理解。

图像图形智能处理技术作为人工智能的重要分支，不仅需要不断推陈出新的核心技术，更需要在各个领域中不断拓展应用场景，实现技术与产业的深度融合。因此，在急需人才的关键时刻，出版这样一套系列丛书具有重要意义。

在编写本丛书的过程中，我们得到了各位作者、审读专家和清华大学出版社的大力支持和帮助，在此表示由衷的感谢。希望本丛书的出版能为广大读者提供有益的帮助和指导，促进图像图形智能处理技术的发展与应用，推动我国图像图形智能处理技术走向更高的水平!

中国图像图形学会理事长

前言

图像超分辨率重建是利用单帧或多帧（视频）低分辨率的图像数据恢复高分辨图像数据的处理技术，而且是图像处理和计算机视觉领域的经典研究内容与重要研究方向。从医学影像的精细诊断到安防监控的细节捕捉，从文化遗产的数字化保护到遥感图像的高精度解析，图像超分辨率技术的应用场景无处不在。然而，传统的超分辨率重建方法往往受限于计算复杂度、细节恢复能力以及对噪声的敏感性，难以满足现代应用场景中对高质量图像的迫切需求。近年来，深度学习技术的崛起为图像超分辨率重建带来了前所未有的机遇。基于深度学习超分辨重建模型不仅能够学习图像的复杂结构和纹理特征，还能够通过大规模数据训练实现高效的特征提取和重建，从而显著提升图像的分辨率和视觉质量。

本书旨在系统性地介绍基于深度学习的图像超分辨率重建技术及其应用。首先阐述了超分辨率重建技术的背景和产生条件，介绍了图像超分辨率重建的基本概念，对技术发展情况和学术研究情况进行了简要的描述和概括；接着详细阐述了图像超分辨率重建内容的理论基础，针对单帧和多帧图像阐述了超分辨率的重建原理及流程，介绍了近些年来超分辨率领域内通用的图像质量评价方式；进一步详细介绍了一些应用于超分辨率重建领域的深度学习方法理论，包括卷积神经网络、生成式对抗网络、自编码器、循环神经网络等经典深度学习方法；聚焦基于深度学习的图像超分辨率重建技术，介绍了有监督、无监督、弱监督的图像超分辨率重建方法，同时从超分任务角度对多帧图像超分辨率重建方法进行了详述，从深度学习理论角度介绍了可解释的深度学习图像超分辨率重建方法；最后，介绍了超分辨率算法的实际应用。本书既包含对当前主流深度学习图像超分辨率重建理论和应用方法的整理和综述，也包含作者团队在该领域的研究成果。本书不仅是一本关于图像超分辨率技术的学术著作，也是一本面向实际应用和工程实践的指导手册，可以为高校学生、科研人员和工程技术人员提供理论与实践相结合的全面参考，期望帮助读者深入理解深度学习在图像超分辨率领域的应用潜力，并将其应用于实际问题，共同推动这一领域的技术进步。

北京航空航天大学宇航学院魏小源、要旭东、熊俊杰、韩喆鑫、李功哲、梅寒、张聪、王鹏睿等参与了本书的部分整理工作。本书的部分研究成果来源于国家自然科学基金面上项目“深度学习图像超分辨率模型可解释性原理分析及方法

研究”、国家重点研发计划课题“微纳卫星在轨遥感参数高精度一体化标定技术”和“临近空间遥感智能信息处理与应用技术”、北京航空航天大学中央高校基本科研业务费等项目。感谢中国图像图形学会、清华大学出版社和北京航空航天大学在本书出版过程中给予的支持和帮助！

由于作者水平有限，书中难免存在疏漏和不足之处，恳请读者批评指正并反馈宝贵意见。

编　者

2025年4月

目录

第1章

绪　论

1.1　图像超分辨率重建技术背景

随着计算机技术、信息处理技术和图像通信技术的迅速发展，人类已经步入崭新的信息化时代。在技术高速更迭的背景下，需要不断改进和发展信息处理技术以应对爆炸式增长的知识量，并为人们提供更加便捷、高效和多样化的服务。数字图像处理及其相关技术是众多信息处理技术中的重要组成部分，在许多领域得到了广泛应用。图像分辨率是描述数字图像质量的重要概念，高图像分辨率可以表示丰富和精确的图像细节。在特定场景下，高分辨率图像属于硬性需求。例如，高分辨率医学图像可以显示细微病灶，而这是肉眼无法直接观察到的；在遥感场景中，有时需要用航空图像识别人脸甚至证件信息特征；某些检测识别控制装置也需要足够高分辨率的图像以确保测量控制精度。因此，提高图像分辨率已经成为图像获取方面追求的重要目标之一。

从1970年至今，电荷耦合器件（charge-coupled device，CCD）和互补金属氧化物半导体（complementary metal oxide semiconductor，CMOS）图像传感器广泛应用于数字图像获取。在需要获取高分辨率图像的许多实际场景中，提高成像装置分辨率是最直接的解决方法。然而，受传感器阵列排列密度的限制，提高传感器空间分辨率变得越发困难。通常的方法是减小单个成像单元的尺寸以增加单位面积内像元的数量。数字摄像机常通过缩小图像传感器（例如CCD）单元尺寸来提高阵列密度和分辨率。但是这种硬件改进会导致成像设备价格的显著上涨。此外，越来越复杂的技术工艺也成为进一步提高分辨率的阻碍。另外，降低像元尺寸通常会同时降低每个像元接受的光照强度，当传感器单元变得十分微小时，接收的光电信号会相应变得非常微弱，容易被传感器自身和环境噪声污染甚至淹没，造成图像质量下降，因此像元尺寸不能无限制减小。具体表现为当CCD图像传感器阵列密度增加到一定程度时，图像分辨率不仅不会提高，还可能下降。此外，物体运动模糊、系统点扩散函数（point spread function，PSF）模糊、采样量化模糊

和电路噪声也会导致图像退化。这些或固有或随机的退化问题仅通过当前的硬件技术往往难以全面解决，因此通过提高硬件质量增强图像分辨率的方法是有限的。

与提升硬件的方法相对应，图像超分辨率重建技术可以有效克服图像传感器的限制。与昂贵的成像设备不同，图像超分辨率重建技术仅需通过计算机软件处理便可获得高分辨率图像，无论是在经济成本还是可操作性方面，都具备明显的优势。

1.2 图像超分辨率重建基本概念、原理与应用

图像超分辨率重建（image super resolution reconstruction，SR）[1] 是指用信号处理和图像处理的方法，通过软件算法从已有的低分辨率（low-resolution，LR）图像恢复原本高分辨率（high-resolution，HR）图像的技术。在图像超分辨率重建领域，研究将图像重建到原来的×2、×3、×4、×8这4种尺度的较多，其中×2代表将图像的边长放大2倍，即像素密度增加至4倍，×3、×4和×8与其同理。如何保证重建后的图像质量更接近真值（ground truth，GT）图像是超分辨率领域的研究重点之一，其目标如下：

$$\hat{y} = \arg\min_{y}[L(F_{\mathrm{sr}}(x), y)] + \lambda\Phi(y) \tag{1-1}$$

其中，x为LR图像，y为待求解的高分辨率图像，$F_{\mathrm{sr}}(x)$为运用图像超分辨率算法重建后的HR图像，λ为平衡参数，$\Phi(y)$为正则化项。

图像超分辨率重建[2] 在视频监控（video surveillance）、图像打印（image printing）、刑侦分析（criminal investigation analysis）、医学图像处理（medical image processing）和卫星成像（satellite imaging）等领域都有广泛的应用。

由于图像超分辨率重建技术可以在一定条件下克服成像系统的分辨率限制，提升输入图像的分辨率，其应用正在迅速增长，涵盖视频、遥感、医学和安全监控等多领域。

（1）在从数字电视（digital television，DTV）向高清晰度电视（high definition television，HDTV）的过渡时期，仅有部分节目以HDTV形式播出，大多数节目仍采用DTV格式。图像超分辨率重建技术可以将DTV信号转换为与HDTV接收机匹配的信号，提升电视节目的兼容性。

（2）在采集军事与气象遥感图像时，由于成像条件和系统分辨率的限制，高清晰度的图像经常难以获得。通过图像超分辨率重建技术，可以在不改变卫星图像探测系统的前提下，实现超出成像系统分辨率的图像获取。在公共安全领域，该技术也可以利用普通监控录像，重建更清晰的目标图像，方便相关人员辨识。

（3）在计算机断层扫描（computed tomography，CT）、磁共振成像（magnetic

resonance imaging，MRI）和超声波等医学成像系统中，图像超分辨率重建技术可以提高图像质量，便于对病变目标进行详细检测。在医学检测中，通常需要通过层析成像技术识别和确定病体的精确位置和详细情况，如阴影边缘、病体占位大小及位置等。由于硬件设备及现有成像技术的限制，往往难以获取高质量图像，而结合层析成像技术机理的图像超分辨率重建技术可以在该领域得到重要应用。

（4）在银行、证券等行业的安全监控系统中，当发生异常情况时，可以对监控录像进行图像超分辨率重建，提高关键部分的图像分辨率，为事件处理提供重要线索。

（5）图像超分辨率重建技术可用于图像压缩。存储或传输数据时采用低分辨率图像，需要时再利用重建技术获得不同分辨率的图像和视频。此外，该技术可以将图像由检出水平（detection level）转化为识别水平（recognition level），甚至进一步达到细辨水平（identification level），从而提高图像的识别能力和精度。

（6）地球资源卫星是一种可以搭载多光谱成像系统的卫星平台，可用于勘探鉴别地表资源。通过对多光谱图像的智能解译，可以获取植被分类及分布、区域地理结构、水资源分布面积等信息。然而，现有成像技术的分辨率限制了多光谱图像的判别和定位精度。通过图像超分辨率重建技术，可以提高资源与环境卫星遥感资料的获取精度。

总之，图像超分辨率重建拥有广阔的发展前景，随着技术的进一步迭代和完善，其应用领域将继续扩大。

1.3 图像超分辨率重建技术发展历程

图像超分辨率重建方法分为基于插值的图像超分辨率重建方法、基于重构的图像超分辨率重建方法和基于学习的图像超分辨率重建方法三类。近年来，随着深度学习技术的快速发展和硬件计算水平的提高，基于深度学习的图像超分辨率重建方法成为学界研究的主要方向，并取得了比以往其他技术更优越的效果和更广泛的应用。进而基于深度学习的图像超分辨率重建方法又分为基于判别式模型的方法和基于生成式模型的方法，下面分别介绍这两种方法的发展历程。

1. 基于判别式模型的方法

2014 年，Dong 等首次提出了超分辨率卷积神经网络（super-resolution convolutional neural network，SRCNN）[3]，这也是深度学习在图像超分辨率重建领域的开山之作。该方法采用预上采样的结构，首先采用插值方法将 LR 图像上采样为与 HR 图像具有相同大小的插值 LR 图像，随后用深度卷积网络直接学习从插值 LR 图像到 HR 图像的端到端映射，以重建具有更多纹理细节的 SR 图像。这一方法开创了基于深度学习的图像超分辨率重建技术研究的先河。2016 年，

Dong 等又进一步提出了一种快速超分辨率卷积神经网络（fast super-resolution convolutional neural network，FSRCNN）[4]，该方法采用后上采样的框架，保持 LR 图像的尺寸，在低维特征空间中，使用更小的卷积核，搭建更深的卷积层数实现变换，最后通过一个反卷积层放大到 HR 图像的尺寸，因为卷积过程主要在低维特征空间实现，因此具有更快的重建速度。

在残差网络提出以后，也有不少学者尝试在超分辨率重建网络中加入残差结构。2016 年，Kim 等提出 VDSR（image super-resolution using very deep convolutional network，极深卷积超分辨率网络）[5] 模型，在超分辨率重建模型中加入残差结构，从而训练出更深的网络结构，使重建效果获得显著提升。受到 VDSR、SRResNet（super-resolution residual neural network，超分辨率残差网络）[6] 等研究的启发，2017 年，Lim 等提出 EDSR（enhanced deep super-resolution network，增强型深度超分辨率网络）[7] 模型，将 ResNet（residual neural network）结构更好地应用于超分辨率重建任务，训练出更深的网络结构；作者还在 EDSR 的基础上进一步提出多尺度的超分辨率模型（multi-scale super-resolution network，MDSR），即用训练好的低倍上采样模型训练高倍上采样模型。

2017 年，Tong 等在 SRDenseNet（super-resolution dense network，超分辨率密集连接网络）[8] 中，将稠密连接加入网络结构，将所有特征层串联起来，充分利用浅层特征和深层特征的互补信息，进一步改善梯度消失现象，提高了超分辨率重建的性能。2018 年，Haris 等提出了深度反投影网络（deep back-projection network，DBPN）[9]，通过多次迭代上采样层和下采样层，并通过不同迭代阶段的跨层连接，计算每个阶段的投影误差，并通过误差反馈机制进行修正。DBPN 在高倍数的图像超分辨率重建任务方面取得了卓越的成果。

2．基于生成式模型的方法

超分辨率卷积神经网络重建方法取得了极为显著的成就，但这些模型生成的图像往往过于平滑，尽管在峰值信噪比（peak signal-to-noise ratio，PSNR）等像素级评价指标方面有很好的表现，图像的感知质量却往往不尽如人意。在生成式对抗网络（generative adversarial network，GAN）[10] 提出以后，一部分学者也将目光投向了 GAN。

2017 年，Ledig 等首次提出基于生成式对抗网络的超分辨率重建方法（super-resolution using a generative adversarial network，SRGAN）[6]，其中生成器采用残差连接结构，由多个残差块串联构成，损失函数采用基于 VGG（visual geometry group，视觉几何组网络）的内容损失和基于判别器的对抗损失的加权。通过对生成器和判别器的交叉迭代训练，得到一个可以生成高纹理细节 SR 图像的生成器。2018 年，Wang 等在 SRGAN 的基础上提出了 ESRGAN（enhanced super-resolution generative adversarial network，增强型超分辨率生成对抗网络）[11]，

ESRGAN 改进了 SRGAN 的网络结构与损失函数，并引入了 RRDB（residual-in-residual dense blcok，多级残差密集块）模块以替代 SRGAN 中的传统残差块结构，消除了 SRGAN 中存在的伪影，进一步提升了网络性能。

此外，从2017年起，计算机视觉领域学术会议，如CVPR（IEEE Conference on Computer Vision and Pattern Recognition，计算机视觉与模式识别会议）和ECCV（European Conference on Computer Vision，欧洲计算机视觉会议）等，专门举办超分辨率任务挑战赛，每年吸引大量的学者参加。比如，自2017年起，图像复原与增强新趋势挑战赛（The New Trends in Image Restoration and Enhancement，NTIRE）已经连续举办了六届，涵盖绝大多数的超分辨率相关任务，每届比赛都设置数十个赛题，上千支队伍参与，引领了超分辨率方向的前沿研究，涌现了EDSR、ESRGAN等代表性的深度学习超分辨率重建方法。又如，感知图像恢复与处理挑战赛（The Perceptual Image Restoration and Manipulation，PIRM）自2018年起每两年举办一届，更侧重于面向感知的图像复原方法，经历了三届激烈竞争，该赛事逐渐成为超分辨率领域的热点活动。这两项主要赛事每次举办都会出现众多有代表性的超分辨率重建方法和数据集，极大促进了超分辨率重建乃至整个图像复原领域的研究与发展，掀起了近年来超分辨率重建研究的新高潮。

1.4 图像超分辨率重建学术研究情况

图像超分辨率重建问题的解决涉及许多图像处理（image processing）、计算机视觉（computer vision）、最优化理论（optimization theony）等领域中的基本问题，例如图像配准（image registration）、图像分割（image segmentation）、图像压缩（image compression）、图像特征提取（image feature extraction）、图像质量评价（image quality estimation）、机器学习（machine learning）、最优化算法（optimization algorithm）等，图像超分辨率重建是这些基本问题的一个具体应用领域，同时对它们的研究进展起到了推动作用。因此图像超分辨率重建问题本身的研究具有重要的理论意义。目前图像超分辨率重建问题已经成为相关研究领域的热点之一。

在20世纪 80—90 年代，就有人开始研究图像超分辨率重建的方法。1984 年 Tsai 的论文[12]是最早提出这个问题的文献。在这之后很多相关的研究对图像超分辨率重建的问题进行了更深入的讨论。有关图像超分辨率重建问题的研究成果，在计算机视觉、图像处理与信号处理领域的顶级会议和期刊都有大量收录。1998年，Borman等发表了一篇图像超分辨率重建的综述文章[13]。2001 年，Kluwer 出版了一本详细介绍图像超分辨率重建相关领域前沿问题的书籍[14]。2003 年，*IEEE Signal Processing Magazine* 刊出了一期图像超分辨率重建的专刊[15]。这些早期

的综述文章主要介绍传统的基于重建的超分辨率问题的研究情况。

近年来，相关的图像超分辨率重建的综述文章[16]，包括介绍单帧图像超分辨率重建问题的文献与介绍基于重建的超分辨率问题的文献，总结了近年来提出的各类算法，并对研究的未来进行了展望。与这些综述文章不同，本书将图像超分辨率重建问题按不同的输入输出情况进行系统分类，综述近年来图像超分辨率重建算法与理论研究的进展，全面介绍图像超分辨率重建、视频超分辨率与单帧图像超分辨率等各类图像超分辨率重建问题的研究情况，对不同的图像超分辨率重建方法进行比较分析，以供相关领域的研究者参考。

1.5 本书内容及章节安排

本书研究了图像超分辨率重建算法的相关理论内容和应用场景，对已有的图像超分辨率重建算法进行了综述整理和系统化分类，期望读者通过此书获得对图像超分辨率重建领域的全面认识和了解。本书的具体章节安排如下。

第1章阐述了图像超分辨率重建技术的背景和产生条件，介绍图像超分辨率重建的基本概念、原理与应用，使读者清晰地意识到图像超分辨率重建的应用价值，并了解图像超分辨率重建内容的一些基本概念，方便对后续章节进行理解。随后对近年来的技术发展情况和学术研究情况进行了简要的描述和概括，最后对本书的章节安排进行了简要介绍。

第2章详细阐述了图像超分辨重建内容的理论基础。首先从成像的原理出发，介绍了光学成像模型及图像退化模型，阐述了退化图像的产生机理。随后针对单帧和多帧图像阐述了超分辨率重建的原理及流程。最后从重建图像的质量评价入手，介绍了近年来图像超分辨率重建领域通用的图像质量评价方法。

第3章概述了深度学习理论和典型方法。介绍了一些应用于超分辨率重建领域的深度学习方法理论，包括卷积神经网络、生成式对抗网络、自编码器、循环神经网络等经典神经网络学习方法，随后介绍了这些方法在图像超分辨率重建中的应用。

第4章介绍了有监督的图像超分辨率重建方法。首先介绍了这类问题的建模方式，随后将这类问题分为判别式超分辨率模型和生成式超分辨率模型，在判别式超分辨率模型中，分别介绍了残差学习、递归学习、课程学习、注意力机制和Transformer等方法中的典型算法；在生成式超分辨率模型中，介绍了生成式对抗超分辨率、先验生成式对抗超分辨率、循环一致性超分辨率算法和先进生成模型方法中典型的算法。

第5章介绍了无监督的图像超分辨率重建方法。先对这类问题进行了问题建模方式的介绍，随后阐述了零样本系列方法和深度图像先验系列方法的相关算法

与原理。

第6章介绍了弱监督的图像超分辨率重建方法，并选取三种典型方法：基于非成对图像训练、基于循环一致性损失和基于闭环卷积神经网络的方法，对这类图像超分辨率重建算法进行了详细的介绍。

第7章介绍了多帧图像超分辨率重建方法。先从视频角度出发，介绍了典型的视频超分辨率重建算法，随后介绍了参考图像的超分辨率重建方法，以及相关典型算法。

第8章介绍了可解释的深度学习图像超分辨率重建方法。先对这部分内容进行背景介绍，阐述了这部分问题产生的原因，随后解释了深度学习的不确定性，并介绍了典型可解释的超分辨率重建算法。

第9章介绍了超分辨率算法的实际应用。选取了几个典型的超分辨率重建应用场景和应用任务，给出了3种超分辨率重建方法的详细介绍。本章描述的应用场景包括遥感图像超分辨率、光谱图像超分辨率、人脸图像超分辨率、医学图像超分辨率和双目图像超分辨率，通过对多种应用场景进行详细介绍，阐述了超分辨率算法在实际生活中的重要作用。

第2章

图像超分辨率重建理论基础

2.1 引言

图像超分辨率重建作为图像处理和计算机视觉中的热门研究领域，具有独特的知识和理论体系，故在了解图像超分辨率重建技术的背景和产生条件后，本章将详细阐述图像超分辨率重建内容的理论基础。本章首先从成像的原理出发，介绍光学成像模型及图像退化模型，阐述退化图像的产生机理；随后介绍图像超分辨率重建的数学基础和信息学基础，展开超分辨率重建内容中包含的数学理论，并针对单帧和多帧图像阐述超分辨率重建的原理及流程；最后从重建图像的质量评价入手，介绍近年来超分辨率领域内通用的图像质量评价方法。

2.2 光学成像模型及图像退化模型

光学图像的成像是一个复杂的光电转化过程，完整地描述了从理想物理场景到电子图像的转化过程。图像降质模型是描述图像从理想的高分辨率到实际观测到的低分辨率图像的成像过程的数学模型。本节内容将从图像的物理成像模型和图像降质模型两个方面展开论述。为了准确分析成像系统中的降质因素并对其进行建模和数学分析，首先介绍图像的物理成像模型，从光学成像的角度分析光学镜头成像器件、下采样和系统噪声对生成图像的不同影响及其数学表达形式。

2.2.1 物理成像模型

影响图像成像质量的主要因素是镜头的光学模糊、成像器件模糊、下采样模糊和CCD噪声模糊。

1. 光学模糊

理想的光学系统假设传感器获得的辐射图像与景物在几何上完全一致。然而，

受光学衍射效应的影响，景物中的一点在成像系统中通常不是原物点成像，而是在此基础上受到一个小模糊核作用，如图2-1所示。现有文献将光学成像系统建模为线性时不变系统，假设景物用函数 $f(x,y)$ 表示，图像用 $g(x,y)$ 表示，模糊核是卷积形式的点扩散函数，其卷积形式如下：

$$g\left(x,\ y\right)=h\left(x,\ y\right)*f\left(x,\ y\right) \tag{2-1}$$

其中，$h(x,y)$ 表示点扩散函数，$*$ 表示卷积操作。对于衍射受限的光学系统，其卷积形式仍可用式(2-1)表示，其中，点扩散函数包含衍射项。

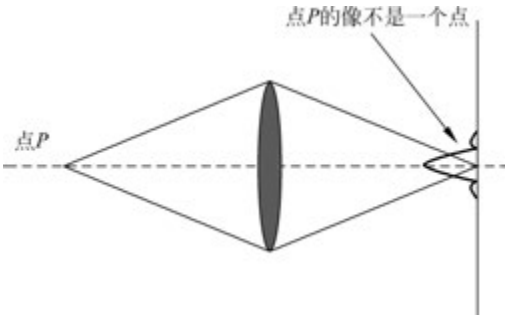

图 2-1　光学模糊

2．成像器件模糊

成像器件的模糊主要源于CCD模糊，可以分解为探测器的点扩散模糊和检测器的采样模糊两个过程。CCD采样导致的景物模糊可视为探测器孔径点扩散函数与景物图像的卷积。而如图2-2所示，检测器的采样模糊是由于光线到达检测器时，该检测器中心产生一个输出，但多个检测器中心之间不会产生输出，导致中心区域之外的信息丢失。基于这些模型，光学模糊、CCD模糊和采样混叠后的传递函数为

$$H_{\text{image}}\left(\xi,\ \eta\right)=H_{\text{optics}}\left(\xi,\ \eta\right)\times H_{\text{detector}}(\xi,\ \eta) \tag{2-2}$$

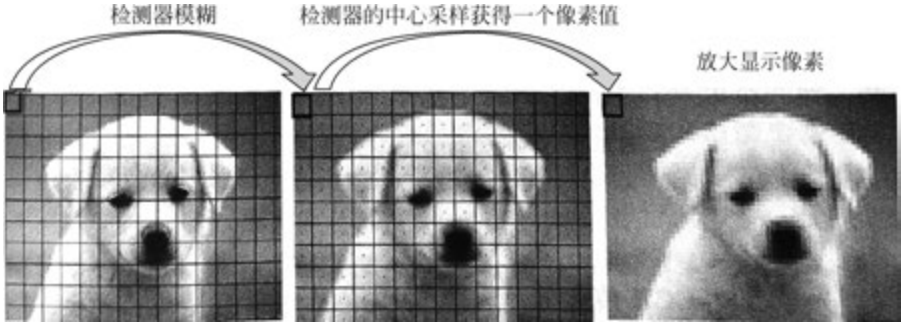

图 2-2　成像器件模糊

3．下采样模糊

受器件影响，CCD采样导致的图像成像具有与CCD一致的分辨率，达不到

与理想光学成像系统获得的高分辨率图像一样的分辨细节。将高分辨率图像转换为低分辨率图像的过程称为下采样，如图2-3所示。图2-3(a)展示了高分辨率图像的像素网格，图2-3(b)展示了相应低分辨率图像的像素网格，其中，图2-3(a)中的4个像素被下采样为图2-3(b)中的1个像素，公式表达如下：

$$A = \omega_1 A_1 + w_2 A_2 + w_3 A_3 + \omega_4 A_4 \tag{2-3}$$

其中，A表示像素值，w表示合成权重。常用的下采样在4个像素中取1个像素安排到低分辨率图像的像素网格中，或者取高分辨率4个像素的平均值作为低分辨率像素值，该过程实现的是2倍下采样。

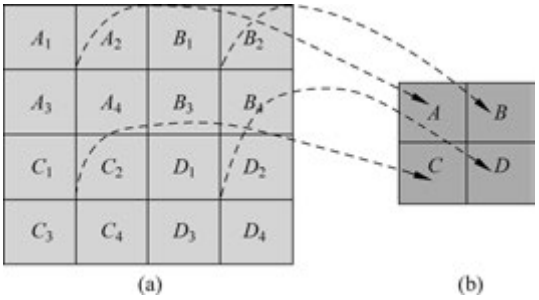

图 2-3　下采样模糊

4. CCD噪声模糊

当光子到达CCD传感器表面时，CCD检测器会生成与光子强度对应的电压值，这一过程会产生随机噪声。尤其是在光线较弱的情况下，到达CCD检测器表面的光子强度非常微弱，检测器生成的电压信号甚至与CCD自身的电荷相当，导致产生CCD暗电流噪声。随机噪声会增加信号的不确定性，通常用标准差衡量。如果各种噪声的分布是独立的，那么系统噪声方差就是各种噪声方差之和。对于各种不同的噪声，整体噪声的标准差可表示为

$$\sigma_{\text{noise}} = \sqrt{\sum_{n=1}^{N} \sigma_n^2} \tag{2-4}$$

其中，n为噪声的种类，当信号强度较大时，主要噪声是光子噪声，其主要是由到达检测器时的随机波动引起的。

综合以上各种物理退化模型，假设观测的低分辨率图像为y，从高分辨率图像到低分辨率图像的下采样过程定义为$\boldsymbol{D}$，大气扰动、光学系统、CCD成像模糊等因素根据统计特性都建模为高斯分布。根据高斯函数的特性，将各种模糊卷积叠加，可以定义成像系统的模糊为$\boldsymbol{B}$，图像的形变矩阵为$\boldsymbol{W}$，原始高分辨率图像为$\boldsymbol{H}$，n为系统噪声，降质过程的矩阵形式数学模型可表示为

$$y = \boldsymbol{DWBH} + n \tag{2-5}$$

2.2.2　图像降质模型

在超分辨率算法中，对图像降质过程估计的准确程度将直接影响超分辨率重建图像的质量。通常系统噪声可通过模拟高斯分布的白噪声近似。在图像处理中，下采样过程经常通过均值滤波技术实现。然而，图像质量下降的关键难题在于难以准确估计图像模糊核。

本书着重介绍图像降质过程中的模糊核估计。由于图像模糊核的估计往往是盲估计，即无法直接观察真实的模糊形式和原始图像，因此这一问题具有不确定性。相关综述指出，大多数模糊核估计问题可以转化为基于贝叶斯理论的最大后验概率优化问题。不同研究的区别在于它们所采用的优化函数和对输入图像降质的不同先验知识。例如，针对失焦模糊，研究者经常使用具有均匀强度的环形滤波器模型；而对于运动模糊，则倾向于采用具有均匀强度的直线段模型。

1999年，You和Kaveh对传统图像恢复方法中的振铃效应进行了深入分析，指出这一问题的根源在于未能有效利用像素间的依赖关系。他们提出利用各向异性正则项建模像素间的依赖关系，并据此约束模糊核的求解。这种方法通过选择与图像内容变化强度和方向相匹配的正则约束项，提高了模糊核参数的估计精度，从而实现了图像边缘的更好恢复。实验结果显示，该方法在处理某些线性运动模糊和失焦模糊时，能够获得较理想的图像复原效果。

2.3　单帧图像超分辨率重建原理及流程

单帧图像超分辨率重建（single image super resolution reconstruction，SISR）对于低分辨率图像进行像素点的补充，从而使图像的空间分辨率得到提升，但由于同一幅高分辨率图像经过不同的退化过程可能得到相似的低分辨率图像，这也造成图像超分辨率重建可能存在多个可能解，也就是图像超分辨率重建的解空间不唯一，所以单帧图像超分辨率重建也成为一个经典的“病态”问题。通过将图像超分辨率重建添加约束和辅助信息，可以将此问题正定化，从而使此问题得到的解尽可能与理想高分辨率图像一致。本节中，将深入探讨超分辨率重建领域的核心要素。首先，解析低分辨率图像是如何经历退化过程的，这是理解超分辨率重建问题的起点。其次，阐述图像超分辨率重建的理论依据，这是指导我们进行重建工作的基础。最后，介绍基于深度学习的超分辨率重建基础网络的网络框架，这是当前该领域的主流技术之一。

2.3.1　单帧图像退化模型

在成像过程中，由于成像器件带来的损失以及环境因素的影响，形成了低分辨率图像，这一过程就是低分辨率图像的退化过程。超分辨率重建问题拟合的就是这个退化过程的逆过程，在低分辨率图像的退化过程中，会受到成像仪器与成

像物体间的相对运动形成的运动干扰。在单帧图像积分时间内，成像仪器的微小运动也会造成图像模糊，在整个成像系统中，由整个图像信号先形成模拟图像最终形成数字图像，也会经历不断的采样和下采样，这也使图像的纹理细节受到损失，因此形成与理想的高分辨率图像不一致的低分辨率图像，整个低分辨率图像的退化过程如图2-4所示。

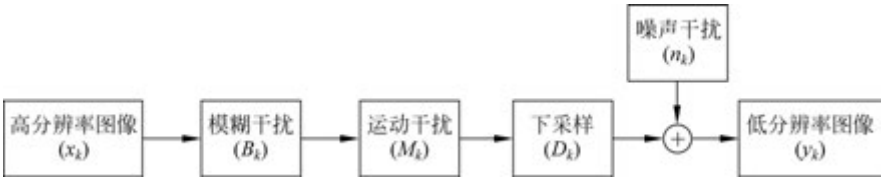

图 2-4　整个低分辨率图像的退化过程

可将整个退化过程用公式表示为

$$y_k = D_k M_k B_k x_k + n_k, \quad k = 1, 2, \cdots, N \tag{2-6}$$

其中，k表示为帧数序号。对于多帧序列图像，y_k为最终成像过程所得的低分辨率图像，D_k为下采样过程，M_k表示运动干扰，B_k则表示模糊干扰，x_k为原始的高分辨率图像，n_k是成像过程中受到的噪声干扰。对于同样的成像仪器成像，下采样和噪声干扰的影响可以认为是近似一致的，而对于单帧图像的退化过程，整个成像过程中也不会受到运动干扰的影响，单帧图像的退化过程可简化为

$$y = DBx + n \tag{2-7}$$

低分辨率图像的退化过程表现了高分辨率图像到低分辨率图像的完整退化，每个因素的影响都会使高分辨率图像降质。针对图像超分辨率问题，对每个影响因素进行处理，使模型更好地拟合退化过程的逆过程，从而实现理想的图像超分辨率重建的效果。

2.3.2　图像超分辨率重建理论基础

近年来，深度学习方法得到了迅速发展，很多研究人员将深度学习用于图像超分辨率重建，提出了很多基于监督深度学习的图像超分辨率重建方法，这些基于监督深度学习的图像超分辨率重建方法的主体思路为学习低分辨图像到高分辨率图像的映射关系，通过最小化迭代高分辨率真值图像和映射所得重建图像不断优化映射，训练好的网络可将输入的低分辨率图像映射为较理想的高分辨率图像，用下式表示映射优化的整个过程：

$$I^{\hat{\mathrm{HR}}} = \underset{I^{\mathrm{HR}}}{\operatorname{argmin}} \| I^{\mathrm{HR}} \downarrow_s - I^{\mathrm{LR}} \| + \lambda \Phi \left(I^{\mathrm{HR}} \right) \tag{2-8}$$

其中，单帧图像超分辨率重建网络的目标是学习映射函数$I^{\hat{\mathrm{HR}}}$表示与低分辨率图像配对的高分辨率图像，I^{LR}表示低分辨率图像，$\downarrow$是下采样操作，s表示下采样比例，$\Phi \left(I^{\mathrm{HR}} \right)$是正则化项，$\lambda$是对应的参数值。

2.4　多帧图像超分辨率重建原理及流程

基于单幅遥感图像的超分辨率重建方法虽然能够取得不错的重建效果，但由于输入数据不足，超分辨率重建的结果均为估计所得，重建得到的图像可信度较差，很难应用于后续遥感图像处理任务。基于序列遥感图像的超分辨率重建使用的数据更多，超分辨率重建的结果可以使用帧间的亚像素信息，因此，超分辨率重建的结果会更可靠，具有更大的应用价值。

序列遥感图像的超分辨率重建分为多帧图像的运动补偿与运动补偿后多帧图像的超分辨率重建两个步骤。其中，多帧图像的运动补偿采用光流法，包括光流估计与利用光流信息的Warp操作两部分。

2.4.1　多帧图像的运动补偿

1. 光流估计

由于遥感图像受大气或抖动的影响比较严重，并且遥感图像中存在汽车、飞机等运动目标，遥感图像帧间的运动较大。因此，序列遥感图像超分辨率重建首先要对多帧图像进行运动补偿，光流法是运动补偿的常用方法，光流是指空间中的运动物体在观测成像平面上呈现的像素运动的瞬时速度表现在图像处理领域，光流估计需求取图像A相对于图像B的像素运动数值$w(x)=(u(x),v(x))=(\Delta x,\Delta y)$。

2. 利用光流信息的Warp操作

得到光流信息$w(x)=(u(x),v(x))=(\Delta x,\Delta y)$后，需要根据光流信息对待运动补偿的图像进行运动补偿，运动补偿的方法采用光流法中的Warp操作。假定图像A为待运动补偿的图像，图像B为基准图像。光流信息$w(x)$为图像A到图像B的光流，现需根据图像A到图像B的光流和图像B求得图像A运动补偿后的图像。假定需求得图像A中坐标为(x,y)的点像素的灰度结果，首先根据图像A到图像B的光流张量w，获得坐标(x,y)处的光流值$(\Delta x,\Delta y)$，根据光流的实际意义，图像A坐标为(x,y)处的像素对应图像B中坐标$(x+\Delta x,y+\Delta y)$处的像素。由于光流的值为浮点数，因此图像B中坐标$(x+\Delta x,y+\Delta y)$处像素的灰度值需要根据其邻域像素的灰度值求得。Warp操作中，求坐标$(x+\Delta x,y+\Delta y)$处像素的灰度值一般利用周围4个像素，采用双线性（bilinear）的方法。

令$x_{\mathrm{B}}=(x',y')=(x+\Delta x,y+\Delta y)$为待求像素坐标，定义系数$\theta_x$，$\theta_y$：

$$\theta_x=x'-\lfloor x'\rfloor,\quad \theta_y=y'-\lfloor y'\rfloor \tag{2-9}$$

令$\widetilde{I}(x_{\mathrm{B}})$为求像素坐标为$x_{\mathrm{B}}$的像素值，则$\widetilde{I}(x_{\mathrm{B}})$的计算公式为

$$\begin{aligned}\widetilde{I}(x_{\mathrm{B}})=&(1-\theta_x)\ (1-\theta_y)\,I_{\mathrm{B}}(\lfloor x'\rfloor,\ \lfloor y'\rfloor)+(1-\theta_x)\ \theta_y I_{\mathrm{B}}(\lfloor x'\rfloor,\ \lfloor y'\rfloor+1)+\\&\theta_x\ (1-\theta_y)\,I_{\mathrm{B}}(\lfloor x'\rfloor+1,\ \lfloor y'\rfloor)+\theta_x\theta_y I_{\mathrm{B}}(\lfloor x'\rfloor+1,\ \lfloor y'\rfloor+1)\end{aligned} \tag{2-10}$$

在运动补偿时，由于图像中目标的运动，补偿后目标的坐标$x_{\mathrm{B}}=(x+\Delta x, y+\Delta y)$可能位于图像外，此时$\widetilde{I}(x_{\mathrm{B}})$为图像B中距离坐标$x_{\mathrm{B}}=(x+\Delta x, y+\Delta y)$最近的边界点。

2.4.2 多帧图像的重建

得到运动补偿后的多帧图像后，需采用多帧图像融合的方法对多帧图像进行超分辨率重建。当前使用多帧图像融合的方法主要基于深度学习进行，可以分为多帧图像按通道Concat的方法（VSRNet[17]等）、3D卷积的方法（DUF等）以及输入图像逐帧处理迭代更新的方法（VESPCN、FRVSR[18]等）。

多帧图像按通道Concat的方法，将输入的多帧图像在通道维度上连接在一起，假定帧图像为$B\times F\times H\times W\times C$，其中$B$、$F$、$H$、$W$、$C$分别为batch数量、图像帧数、图像高度、图像宽度、图像通道数。多帧图像按照通道维度连接在一起，图像张量变为$B\times H\times W\times CF$，之后采用2D卷积对其进行处理，与单帧图像超分辨率重建相同。

3D卷积从图像帧数、图像高度、图像宽度3个维度对多帧图像进行卷积，在卷积过程中逐渐减少图像的帧数，最终减少为1，完成对多帧图像的超分辨率重建。

输入图像逐帧处理迭代更新的方法一次仅针对两帧图像进行超分辨率重建。首先确定一帧图像为基准帧，依次输入其余帧图像，将每一次两帧图像超分辨率重建的结果图作为下一次超分辨率重建的基准帧，迭代更新基准帧图像。所有图像输入完毕后，最后得到的基准帧图像即为多帧图像的超分辨率重建结果。

2.5 重建图像的质量评价

图像质量通常涉及图像的视觉属性，这些属性主要影响观看者的感知。在评估图像质量时，存在两种主要的体系：主观定性评价体系（subjective qualitative assessment metric，SQAM）和客观定量评价体系（objective quantitative assessment metric，OQAM）。SQAM依赖人类视觉系统（human vision system，HVS）对图像质量进行主观评价，其结果受到观察者个人感知和经验的影响，因此具有较大的差异性。而OQAM通过建立模拟HVS的评价模型，计算量化参数得出具体的图像质量评价表达式。尽管SQAM更符合直观需求，但其耗时多且成本较高，OQAM目前已成为主流的评价方法。

在OQAM中，又可分为3种主要类型：全参考方法、减参考方法和无参考方法。全参考方法依赖完整的参考图像进行评估，而减参考方法基于提取的特征进行比较。相比之下，无参考方法不需要任何参考图像即可进行质量评估。尽管这些方法在技术上有所不同，但它们的目标都是更准确地评估图像质量。然而，由

于客观方法有时难以完全捕捉人类的视觉感知，因此评估结果可能存在一定的不一致性。

本节将介绍几种常用的图像质量评价（image quality assessment，IQA）方法（包括主观方法和客观方法）。

2.5.1　均方误差/均方根误差

均方误差（mean squared error，MSE）是一种常用的误差度量方法，它通过对估计值与真实值（或称参考值、基准值）之间差的平方进行平均计算，量化估计值与真实值之间的差异程度。MSE 的值越小，说明估计值与真实值之间的差异越小。给出具有 N 个像素的图像的真实值 I 和待评价的估计值 $\hat{I}$，I 与 $\hat{I}$ 之间的 MSE 定义为

$$\mathrm{MSE}=\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\hat{I}(i)\right)^2 \tag{2-11}$$

均方根误差（root mean square error，RMSE）通过计算均方误差的平方根获得，表达形式为

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\hat{I}(i)\right)^2} \tag{2-12}$$

MSE 与 RMSE 用于衡量估计值与真实值之间的偏差，是非常基本的全参考客观评价方法。

2.5.2　峰值信噪比

峰值信噪比是衡量图像质量，特别是评估图像在经过有损变换（如压缩或涂抹）后重建质量的重要指标之一。在图像超分辨率处理中，PSNR 的计算依赖最大像素值与图像之间的均方误差。PSNR 提供了一种量化评估图像重建质量的方式，特别是在处理图像超分辨率问题时，PSNR 被广泛用作评价指标。给定一幅具有 N 个像素的图像，设其参考值为 I，而待评价的估计值为 $\hat{I}$，二者之间的 PSNR 可进行以下定义：

$$\mathrm{PSNR}=10\log_{10}\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}\left(I(i)-\hat{I}(i)\right)^2} \tag{2-13}$$

在评估图像质量时，PSNR 是一个常见的全参考客观评价指标。它基于图像中每个像素 $I(i)$ 与参考图像之间的均方误差进行计算。其中，L 代表图像能表示的最大值，对于 8 位图像来说，这个值通常为 255。

然而，PSNR 有一个明显的局限性：它仅关注像素级的差异，而不考虑人类视觉感知。因此，在真实场景中评估图像重建质量时，PSNR 的表现可能并不理

想。特别是在图像超分辨率重建任务中，人们更关心图像在人类视觉感知下的质量，而不仅仅是像素级别的匹配。

尽管如此，由于缺乏一个完全准确的感知质量指标，PSNR 仍然是图像超分辨率重建模型中广泛应用的评价标准。它提供了一种简单而直接的度量方式，用于量化评估图像重建的精度。

2.5.3 结构相似性指数测度

鉴于人类视觉系统对图像结构的高度敏感性，研究者提出了结构相似性指数测度（structural similarity index measure，SSIM）以评估图像间的结构相似性。SSIM 基于亮度、对比度和结构三个方面进行独立比较，从而全面评价图像间的相似程度。

具体来说，对于一幅具有 N 个像素的图像I，亮度通常用图像的平均值衡量，而对比度则通过图像的标准差表示。这两个参数分别捕捉图像的整体亮度和对比度信息，是评估图像相似性的重要基础。

SSIM 通过综合考虑这些因素，能够更准确地反映图像之间的结构相似性，特别是在处理具有复杂结构和纹理的图像时，其表现尤为出色。因此，SSIM 在图像处理和质量评估等领域得到了广泛应用。SSIM 的计算公式为

$$\mu_I = \frac{1}{N}\sum_{i=1}^{N} I(i), \quad \sigma_I = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(I(i)-\mu_I\right)^2} \tag{2-14}$$

其中，$I(i)$ 代表图像中的第 i 个像素。对两幅图片亮度和对比度之间的对比分别用 $C_{\mathrm{l}}\left(I,\ \hat{I}\right)$ 和 $C_{\mathrm{c}}\left(I,\ \hat{I}\right)$ 表示，计算方法表示为

$$C_{\mathrm{l}}\left(I,\ \hat{I}\right) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \ \mu_{\hat{I}}^2 + C_1}, \quad C_{\mathrm{c}}\left(I,\ \hat{I}\right) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \ \sigma_{\hat{I}}^2 + C_2} \tag{2-15}$$

其中，C_1 和 C_2 分别为消除不稳定性的常数，计算方式为

$$C_1 = (k_1 L)^2, \quad C_2 = (k_2 L)^2, \quad k_1 \ll 1,\ k_2 \ll 1 \tag{2-16}$$

对于图像结构，SSIM 用归一化的像素值表示 $((I-\mu_I)/\sigma_I)$，利用 I 与 $\hat{I}$ 之间的内积，也就是二者之间的相关系数，衡量结构的相似性。结构之间的对比分别用 $C_{\mathrm{s}}\left(I,\ \hat{I}\right)$ 表示，计算方法定义如下：

$$\sigma_{I\hat{I}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(I(i)-\mu_I\right)\left(\hat{I}(i)-\mu_{\hat{I}}\right) \tag{2-17}$$

$$C_{\mathrm{s}}\left(I,\ \hat{I}\right) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I\sigma_{\hat{I}} + C_3} \tag{2-18}$$

其中，$\sigma_{I\hat{I}}$ 表示估计图像与真实图像之间的相关系数，C_3 是消除不稳定性的常数。

SSIM 定义如下：

$$\mathrm{SSIM}\left(I,\ \hat{I}\right)=\left[C_{\mathrm{l}}\left(I,\ \hat{I}\right)\right]^{\alpha}\left[C_{\mathrm{c}}\left(I,\ \hat{I}\right)\right]^{\beta}\left[C_{\mathrm{s}}\left(I,\ \hat{I}\right)\right]^{\gamma} \tag{2-19}$$

其中，α、β、γ分别是调整重要程度的控制参数。SSIM 作为一种基于 HVS 的图像质量评价指标，能够更贴近人类感知的评估重建质量，因而被广泛应用于图像超分辨率算法的评估。这种评估方式更能满足知觉评估的要求，使图像超分辨率算法的重建质量得到更准确的评估。

2.5.4　平均意见得分

平均意见得分（mean opinion score，MOS）是一种主观图像质量评估的常用方法，它通过收集人类评分者对被测图像的知觉质量评分获取平均评分。这种方法能够直接反映人类对图像质量的感知，因此在图像处理、计算机视觉等领域具有广泛的应用。通常情况下，分数从1（坏）到5（好），而最终的 MOS 为所有评分的算术平均值。

虽然 MOS 测试似乎是一种可信的 IQA 方法，但它有一些固有的缺陷，如非线性感知的尺度、不同评分者对图像的偏见和评级标准的差异等。在评估图像超分辨率重建模型的性能时，我们常发现一些模型在标准 IQA 指标（如 PSNR）下表现并不突出，但在实际的人类感知质量方面却远胜其他模型。此时，MOS 测试凭借其直接反映人类感知质量的特性，成为评估这些模型性能最可靠的主观 IQA 方法。

2.5.5　基于学习的质量感知方法

为了降低人工干预并提升图像感知质量的评估效率，研究人员致力于通过大型数据集上的学习自动评估图像质量。Ma 等提出了无参考方法，而 Talebi 等则开发了 NIMA 模型，两者均从视觉感知分数中学习，无须参考图像即可直接预测质量分数。Kim 等则进一步提出了 DeepQA，该模型通过结合失真图像、客观误差图和主观分数进行三合一训练，以预测图像的视觉相似度。

另外，Zhang 等构建了一个大规模的感知相似性数据集，并开发了 LPIPS（largescale perceptual image patch similarity）方法，该方法基于训练的深度网络的深度特征差异评估感知图像补丁的相似性。他们发现，卷积神经网络学习的深度特征在感知相似性建模方面显著优于无卷积神经网络的措施。

然而，尽管这些方法在模拟人类视觉感知方面取得了显著进展，但关于我们期望什么感知质量（如更真实的图像或与原始图像一致的图像）仍是一个悬而未决的问题。因此，目前主流的客观 IQA 方法，如 PSNR 和 SSIM，仍然占据重要的地位。

2.5.6 基于下游任务的质量感知方法

图像超分辨率重建的结果可以帮助其他下游任务获取更好的结果，因此可以通过其他任务指标评估图像超分辨率重建的性能。对此研究人员采用了一种高效的方法，他们首先将原始图像与重建的高分辨率图像一同输入预先训练好的模型，接着通过比较这些图像对模型预测性能的影响量化重建质量。这种评估方法能够涵盖多种视觉任务，包括但不限于物体识别、人脸识别、人脸对齐及图像解析等，从而全面地检验重建图像在不同应用场景下的表现。这种评估策略不仅提高了评估的效率和准确性，而且为图像重建技术的发展提供了有力的支持。

2.5.7 其他图像质量评估方法

在评估图像超分辨率重建方法时，除了常见的IQA方法外，还存在一些不太普及的度量评价指标。例如，多尺度结构相似性（multi-scale structural similarity index measure，MS-SSIM）相比单尺度的SSIM，在适应不同观看条件方面更具灵活性。而特征相似性（feature similarity index measure，FSIM）则通过结合相位一致性和图像梯度大小，聚焦人类感兴趣的特征点以评估图像质量。

此外，自然图像质量评估器（natural image quality evaluator，NIQE）是一种无须参考扭曲图像的方法，它基于自然图像中观察到的统计规律的可测量偏差评估图像质量。

近期的研究还揭示了一个有趣的发现，即传统的失真评价指标（如PSNR、SSIM）与感观质量评价（如MOS）之间存在矛盾。这意味着，随着失真度的降低，图像的感观质量并不一定提高。因此，如何精准地衡量图像超分辨率重建的质量，仍是一个值得进一步探索和研究的问题。

2.6 小结

本章全面介绍了图像超分辨率重建的理论基础，为读者提供了深入探索该领域的过程中所需的关键知识。首先从光学成像原理出发，详细解析了透镜衍射导致的光学模糊、传感器空间限制引起的器件模糊、像素转换中的下采样模糊以及电子感光产生的系统噪声等关键降质因素，并在此基础上构建了矩阵形式退化模型。随后，本章基于贝叶斯最大后验概率框架讨论了模糊核估计及降质过程的数学本质，阐述了超分辨率重建作为典型病态问题的根源，并指明了通过添加正则化约束和辅助信息实现问题正定化的基本理论路径。

本章的后半部分，深入研究了单帧图像超分辨率重建和多帧图像超分辨率重建的原理与流程。探讨了单帧图像超分辨率重建的基本思想，例如，如何通过学

习图像的内在特征和结构，实现从低分辨率到高分辨率的转换。同时，详细讨论了多帧图像超分辨率重建，利用多个观测图像的信息提高重建质量，更进一步地改善图像细节。

最后强调了重建图像质量评价的重要性，介绍了一些常用的评价指标，如均方根误差、结构相似性指数测度和峰值信噪比。这些指标有助于客观地量化重建结果的好坏，从而指导算法的选择和调优。还介绍了基于学习和基于下游任务的质量感知等重建图像的质量评估方法，以便更好地评估图像的感知质量。

通过本章的学习，读者将全面了解图像超分辨率重建的理论基础，为进一步深入研究和实践奠定坚实的基础。在接下来的章节中将进一步探索图像超分辨率重建的技术和应用，带领读者走进这个充满活力和创新的领域。

第3章

深度学习理论与典型方法概述

3.1 引言

随着机器学习领域的发展，深度学习算法迅速崛起并成为计算机视觉算法的热门。同样地，在图像超分辨率重建领域，基于深度神经网络的图像超分辨率重建方法迅速成为主流。采用深度学习的主要优势如下。首先，采用深度学习进行图像超分辨率重建是一个端到端的过程，无须对图像进行预处理和后续处理，重建过程简单方便。其次，深度神经网络具有很强的特征提取与描述能力，远远强于人工特征，能对低分辨率图像块进行很好的描述与重构。最后，深度学习具有不断学习和迁移学习的能力，学习后的网络能够根据实际问题需求进行迁移学习与调整。

本章将对深度学习理论及典型方法进行概述。首先介绍机器学习的定义及其思想，以及神经网络、深度学习的基本原理及其发展历程。然后介绍典型的深度神经网络，即卷积神经网络、生成式对抗网络、自编码器、循环神经网络等。最后介绍深度学习在图像超分辨率重建中的作用。

3.2 机器学习与神经网络

3.2.1 机器学习

机器学习领域的历史可追溯至20世纪50年代，Arthur Samuel提出了该领域的初步概念：在特定编程的情境下，赋予计算机自我学习能力的领域。他编写了西洋棋程序，并让程序自我对弈上万次，程序逐渐学会了识别哪些棋盘布局有利于胜利，哪些不利于胜利。

随着研究的深入，1997年Tom Mitchell提出了一个更精确和形式化的定义：一个程序被认为能从经验E中学习，解决任务T，达到性能度量值P，当且仅当有

了经验E后，经过P评判，程序在处理T时的性能有所提升。简言之，机器学习是探索如何利用计算技术，通过经验数据增强系统性能的学科。

在计算机系统中，经验通常是以数据形式存在的，因此机器学习的主要研究焦点在于开发能够从数据中生成“模型”的算法。从广义上讲，机器学习是一种使机器具备学习能力以完成直接编程难以达成的任务的方法。从实际应用的角度看，机器学习是使用数据训练模型，进而利用这些模型进行预测的过程。

3.2.2　神经网络

神经网络是一个由简单、适应性强的单元构成的网络结构，其互联的并行性高度模仿了生物神经系统的功能，特别是与真实世界的交互反应。在神经网络的核心，最基本的组成元素是神经元模型。

在模拟生物神经网络的过程中，每个神经元都被设计为与其他神经元相互连接。当某个神经元被激活时，它会通过发送化学信号（或称为神经递质）影响与其相连的神经元。这些信号会改变接收神经元的电位。一旦某个神经元的电位积累到一定程度，超过了设定的“阈值”，这个神经元也会被激活，进而继续传递信号给其他神经元。通过这种方式，神经网络能够模拟生物神经系统的基本工作原理，实现信息的处理和传递。

1943 年，McCulloch 等将上述生物神经元模型抽象为“M-P 神经元模型”。其中，神经元接收来自其他 n 个神经元的输入信号，这些信号通过加权连接传输。神经元将累积的输入值与自身的阈值进行比较，并通过激活函数进行处理，进而生成输出。在实践中，Sigmoid 函数常用作激活函数。将多个神经元模型中的神经元以不同方式连接，就能构成各种模式的神经网络。

3.3　深度学习基本原理与发展历程

3.3.1　深度学习及其基本原理

深度学习作为机器学习领域的新兴分支，专注于探索样本数据的内在规律与表示层次。在此过程中，获取的知识极大地助益对文字、图像、声音等数据的诠释。其终极愿景是使机器像人一样，具备分析学习的能力，从而辨识各类数据。深度学习算法在语音和图像识别方面的卓越表现已远超先前的技术。

深度学习属于模式分析范畴，具体涉及如下类别。

（1）卷积神经网络，即基于卷积运算的神经网络系统。

（2）包括自编码（autoencoding）与稀疏编码（sparse coding）在内的基于多层神经元的自编码神经网络。

（3）深度置信网络（deep belief network，DBN），通过多层自编码神经网络预训练，并结合鉴别信息优化神经网络权值。

过去机器学习应用于实际任务时，样本特征的描述多由专家设计，被称为“特征工程”。显然，特征的优劣直接影响模型的泛化能力，但优质特征的设计并不容易。而特征学习则通过机器学习技术自主产生良好特征，推动机器学习向自动化数据分析迈进。近年来，研究人员开始融合这些方法，如先结合有监督的卷积神经网络与无监督的自编码神经网络进行预训练，再利用鉴别信息微调网络参数，形成卷积深度置信网络。与传统方法相比，深度学习方法包含更多模型参数，因此训练难度更大，同时要求更大的数据量来支撑。

与传统浅层学习相比，深度学习的特色如下。

（1）强调模型结构的深度，常含有五六层甚至十余层的隐层节点。

（2）重视特征学习，即通过逐层特征变换，将样本特征由原空间转换至新空间，便于分类或预测。相较人工构造特征，大数据驱动的特征学习更能捕捉数据背后的丰富信息。

3.3.2 深度学习发展历程

1958 年，计算科学家 Rosenblatt 提出了由两层神经元组成的神经网络。他给这种模型起了一个名字——“感知器”（perceptron）（有的文献翻译为“感知机”，本书中统称“感知器”）。感知器是当时首个可以学习的人工神经网络，后来成为当今智能机器的核心和起源。Rosenblatt 现场演示了感知器学习识别简单图像的过程，当时引起了轰动，人们认为已经发现了智能的奥秘，许多学者和科研机构纷纷投入神经网络的研究。美国军方大力资助了神经网络的研究，并认为神经网络比“原子弹工程”更重要，这段时间直到 1969 年才结束，这个时期可以看作神经网络研究的第一次高潮。1960 年，Henry J. Kelley 开发了连续反向传播模型，为多层神经网络的训练打下了基础。最早的深度学习算法是 1965 年由 Alexey Grigoryevich Ivakhnenko（同时开发了数据处理的分组方法）和 Valentin Grigorevich Lapa（也是控制论与预测技术的作者）开发的。他们使用具有多项式（复杂方程式）激活函数的模型并进行统计分析，在每一层将统计上最佳的功能转发到下一层（缓慢的手动过程）。然而，深度学习和人工智能（artificial intelligence，AI）研究无法兑现诺言，从而影响了资金投入。20 世纪 70 年代，第一个 AI 寒冬开始了，深度学习的研究也停滞了。

由于得益于反向传播（back propagation，BP）算法的提出，这种停滞状态一直到 20 世纪 80 年代才有所改观。反向传播是指在网络训练中，通过计算损失函数对网络参数进行更新的过程，其概念在 20 世纪 60 年代初确实存在，但它笨拙且效率低下；1970 年，反向传播有了显著发展，Seppo Linnainmaa 的硕士论文中也包括用于反向传播的 FORTRAN 代码，但不幸的是，直到 1985 年，该概念才应用

于神经网络。Rumelhart、Williams 和 Hinton 证明了神经网络中的反向传播可以提供“有趣的”分布表示。从哲学上讲，这一发现使人们认识到人类理解是依赖符号逻辑（计算主义）还是分布式表示（联系主义）的认知心理学问题。1989 年，Yann LeCun 在贝尔实验室提供了反向传播的第一个实际演示。他将应用了反向传播算法的卷积神经网络应用到读取手写数字上。该系统最终成功应用于识别手写邮政编码数字和读取支票上。

福岛邦彦是首位设计出包含多个池化和卷积层的神经网络架构的先驱。1979 年，他构建了一种被命名为Neocognitron的人工神经网络，这一网络架构基于层次化的多层设计，赋予了计算机“学习”识别视觉图案的能力。这一网络与现代版本相似，但其独特的强化策略通过循环激活训练逐步增强，确保网络性能的优化。值得一提的是，福岛的设计还允许用户通过调整特定连接的“权重”手动强化重要功能，这种自上而下的连接策略以及新颖的学习机制，使各种神经网络得以实现。当多个模式同时展现时，其内置的选择性注意模型能够灵活地转移注意力，从而有效地区分和识别每种模式（这一点与我们在多任务处理时使用的机制颇为相似）。现代的Neocognitron不仅能够识别信息不完整的图案（例如部分缺失的数字5），还能通过补充缺失信息完善图像，这一过程被形象地称为“推断”。

第二个 AI 寒冬是1985—1990年，这也影响了神经网络和深度学习的研究。一方面，各种过于乐观的研究者夸大了人工智能的“即时”潜力并激怒了投资者，人工智能一度达到伪科学的地步；另一方面，随着计算机新技术的发展和算法的完善，研究人员发现反向传播会导致梯度消失问题出现：误差会随着网络层数的增加而逐渐消失，没有学习信号到达更深层次。该缺点的发现对于深度学习来说是一个致命打击，使深度学习的发展再度陷入停滞状态。但是在这段时间内，支持向量机（support vector machine，SVM）、决策树、随机森林等算法的发明，再加上它们对数据分类有着良好的效果，使基于统计思想的机器学习方法成为主流。幸运的是，一些研究者继续从事人工智能的研究，并取得了一些重大进展。梯度消失并不是所有神经网络的根本问题，只是那些采用基于梯度的学习方法的神经网络。问题的根源是某些激活功能，许多激活功能压缩了它们的输入，进而以某种混乱的方式缩小输出范围，这产生了在很小范围内映射的大面积输入，在这些输入区域中，大的变化将减小为输出的小变化，从而导致梯度消失。解决此问题的两个方案是逐层预训练和长短期记忆的开发。

在深度学习的演进历程中，1999年堪称关键的转折点。彼时计算机技术在处理数据方面实现了显著的加速，尤其归功于图形处理单元（graphics processing unit，GPU）的引入。GPU在图像处理方面的应用极大地提升了处理速度，短短十年间，其计算效能激增了千倍。这一时期，神经网络开始崭露头角，与支持向量机展开了激烈的竞争。尽管神经网络在速度上可能稍逊于支持向量机，但其在利用相同数据集时，展现出更优越的性能。更值得一提的是，神经网络具有一个

显著的优势，即它能够随着训练数据的不断增加持续进化和优化。

2001 年，META Group（现称 Gartner）的一份研究报告描述了随着数据源和类型范围的扩展，数据量的增加速度与之呈三次关系，这使学界开始为即将开始的大数据冲击做准备。2009 年，斯坦福大学 AI 教授李飞飞创建了 ImageNet，该数据库收集了超过 1400 万幅免费的带标签图像。李教授说：“我们的愿景是让大数据改变机器学习的工作方式，即数据驱动学习。”2011 年，GPU 的速度已显著提高，从而可以无须逐层进行预训练以训练卷积神经网络。随着计算速度的提高，深度学习在效率和速度方面具有了更明显的优势。一个例子是一种卷积神经网络 AlexNet，其体系结构在 2011 年和 2012 年间赢得了多项国际竞赛。

同样在 2012 年，Google Brain 发布了一个名为“猫实验”的项目的结果，这个自由奔放的项目探讨了“无监督学习”的方法和遇到的困难。那时深度学习通常使用“监督学习”，这意味着卷积神经网络是使用标记数据进行训练的。使用无监督学习，卷积神经网络将获得未标记的数据，然后在无标签数据中寻找重复模式。时至今日，获取大量数据已经不是难事，大量数据的标注反而耗时耗力耗材，故而无监督学习、弱监督学习或半监督学习仍然是深度学习领域的重要目标。

3.4 卷积神经网络

卷积神经网络（CNN）是一类包含卷积计算且具有深度结构的前馈神经网络（feedforward neural network），是深度学习（deep learning，DL）的代表算法之一。一个卷积神经网络一般包含 5 层，分别是数据输入层、卷积计算层、激活层、池化层和全连接层。CNN 主要用于图像分类、目标检测等领域。

1. 数据输入层

数据输入层要做的处理主要是对原始图像数据进行预处理，预处理手段包括去均值、归一化、主成分分析（principal component analysis，PCA）降维等操作。

去均值操作即各维度都减去对应维度的均值，使输入数据的各维度都中心化为 0。去均值是为了容易拟合。这是因为在神经网络中，特征值 x 比较大，会导致加权求和 $Wx+b$ 的结果也很大，进行激活函数输出时，会导致对应位置数值变化量太小，进行反向传播时因为要使用这里的梯度进行计算，所以会出现梯度消散问题，导致参数改变量很小，也就不易于拟合，效果不好。

归一化时可以使用最大最小值归一化和均值方差归一化等。最大最小值归一化可以决定一个范围，并把最大值归一化到范围上界，最小值归一化到范围下界。适用于本来就分布在有限范围内的数据。均值方差归一化一般是将均值归一化为 0，方差归一化为 1，适用于分布没有明显边界的情况。进行归一化后可以把各特征的尺度控制在相同的范围内，以便于找到最优解，提高收敛效率。

PCA是一种常见的数据分析方法，这一方法利用正交变换将由线性相关变量表示的数据转换为少数几个由线性无关变量表示的数据，线性无关变量称为主成分。主成分的数量通常小于原始变量数量，因此主成分分析常用于高维数据的降维，提取数据的主要特征分量。

2．卷积计算层

卷积计算层是卷积神经网络最重要的一个层次，也是“卷积神经网络”名字的来源。卷积层的参数设定涵盖卷积核大小、步长和填充，这三者共同确定输出特征图的尺寸，并在网络中被视为超参数。在设定卷积核大小时，可以选择小于输入图像尺寸的任意值，而较大的卷积核能够捕捉更复杂的输入特征。步长则定义卷积核在扫描特征图时两次相邻动作之间的距离，当步长为1时，卷积核会逐个扫描特征图的每个元素；而当步长设定为n时，则会在每次扫描后跳过$n-1$个像素。随着卷积层的不断叠加，特征图的尺寸会逐渐缩小，例如，一个16×16的输入图像在经过5×5的卷积核（无填充，步长为1）的处理后，会输出一个12×12的特征图。为了弥补这种尺寸上的缩减，可以采用填充的方法，即在特征图通过卷积核之前，人为地增加其尺寸，以抵消计算过程中的尺寸缩减。常见的填充方法包括0填充和边界值复制填充（replication padding）。

3．激活层

激活层对卷积层的输出结果进行非线性映射，常见的非线性激活函数有Sigmoid、Tanh、修正线性单元（the rectified linear unit，ReLU）、Maxout等。如3.2节中提到的Sigmoid在网络层数较深时会引起梯度消失问题，卷积神经网络采用的激活函数一般为 ReLU，它的特点是收敛快、求梯度简单。

4．池化层

池化层夹在连续的卷积层中间，用于压缩数据和参数的量，减小过拟合。在输入为图像的情况下，池化层的最主要作用是压缩图像。池化层常用的方法有最大池化和均值池化两种。以最大池化为例，对每个窗口滑过的部分，取其中的最大值作为对应输出相应位置的值。

5．全连接层

全连接层是神经网络中一种常见的层类型，全连接层的神经元连接形式与传统的神经网络中神经元的连接方式相同。在全连接层中，每个神经元都与上一层的所有神经元相连，每个输入特征都与每个神经元之间存在一定的连接权重。在训练过程中，神经网络通过反向传播算法优化每个神经元的权重和偏置，从而使输出结果更好地拟合训练数据。全连接层的作用是将输入特征映射到输出结果，通常在神经网络的最后一层使用，用于分类、回归等任务。全连接层的输出结果可以看作对输入特征的一种非线性变换，这种变换可以将输入特征空间映射到输出结果空间，从而实现模型的复杂性和非线性拟合能力。需要注意的是，全连接层

的参数量非常大，因此容易出现过拟合的情况。为避免过拟合，可以使用一些正则化方法，比如Dropout、L1/L2正则化等。

3.5 生成式对抗网络

生成式对抗网络（GAN）模型[10]通过框架中（至少）两个模块——生成模型（generative model）和判别模型（discriminative model）的互相博弈学习产生相当好的输出。以图片生成网络为例，可以设想有G和D两个网络在运作。其中，G网络负责图片的生成，它通过接受一个随机噪声z来创建图片，将其表示为$G(z)$。D网络则是一个鉴别器，它负责判断图片是否“真实”。具体来说，D输入图片x，输出$D(x)$表示x是真实图片的概率。若输出为1，则意味着x被判断为100%真实的图片；反之，若输出为0，则意味着x极可能是伪造的。在训练阶段，生成网络G的主要目标是创造足够逼真的图片，以迷惑判别网络D。而D的目标是准确区分G生成的图片与真实图片。因此，G和D之间形成了一个动态的“对抗过程”。当这种对抗达到理想状态时，G网络能够生成与真实图片几乎无法区分的图片$G(z)$，即达到“以假乱真”的效果。GAN在图像处理领域有着广泛的应用，例如提高图像分辨率、语义分割等任务。

3.6 自编码器

自编码器（autoencoder，AE）是一种在半监督和无监督学习场景中广泛应用的人工神经网络。其独特之处在于，它将输入信息直接设定为学习目标，进而实现输入信息的表征学习。自编码器由两大核心组件构成：编码器（encoder）与解码器（decoder）。从学习范式的角度看，自编码器可以细分为多种类型，如收缩自编码器（contractive autoencoder）、正则自编码器（regularized autoencoder）、变分自编码器（variational autoencoder，VAE）。其中，前两者属于判别模型范畴，而后者则是一种生成模型。在构筑类型方面，自编码器可以灵活构建为前馈结构或递归结构的神经网络。自编码器不仅拥有表征学习算法的基本功能，还在多个领域展现出应用价值，例如降维（维度缩减）和异常值检测。特别是在计算机视觉领域，自编码器结合了卷积层后，便能处理图像降噪、神经风格迁移等复杂问题。

3.7 循环神经网络

循环神经网络（recurrent neural network，RNN）是处理序列数据的一种神经网络架构，它通过链式连接的循环单元在序列的演进过程中进行递归操作。该领

域的研究始于20世纪八九十年代，并于21世纪初逐步演进为深度学习领域的重要算法。在众多RNN的变种中，双向循环神经网络（bidirectional RNN，Bi-RNN）和长短期记忆网络（long short-term memory network，LSTM）尤为常见。RNN在自然语言处理任务中发挥着关键作用，如语音识别、语言建模和机器翻译等，同时在时间序列预测等领域展现出其应用价值。为了处理包含序列输入的计算机视觉问题，研究者还引入了结合卷积神经网络结构的循环神经网络模型。

3.8　深度学习在图像超分辨率重建中的应用

3.8.1　基于卷积神经网络的图像超分辨率重建

由于卷积神经网络在图像分类、识别等领域有着出色的表现，因此研究者将卷积神经网络应用到图像超分辨率重建中，提出了SRCNN[3]、VDSR[5]、EDSR[7]等模型。其中，SRCNN是卷积神经网络应用于图像超分辨率重建的开山鼻祖，其对采集的高分辨率图像以一定的采样因子下采样，得到低分辨率图像，再利用双三次插值的方法将低分辨率图像重建为与原来高分辨率图像同等尺寸，而后将其作为卷积神经网络的输入，得到的超分辨率输出与相对应的高分辨率图像求损失函数，通过反向传播调整各卷积层的权值不断优化模型，直至损失函数收敛，这样模型就具有了图像超分辨率重建的功能。SRCNN 的提出具有里程碑式的意义，但它太过依赖小图像区域的上下文信息、训练时收敛较慢、网络仅适用于单一采样尺度，故研究者又在此基础之上对模型进行改进，提出新的模型[4, 8-9]，推动基于卷积神经网络的图像超分辨率重建向前发展。

3.8.2　基于生成式对抗网络的图像超分辨率重建

基于卷积神经网络的超分辨率重建放大将一系列低分辨率图像和与之对应的高分辨率图像作为训练数据，学习一个从低分辨率图像到高分辨率图像的映射函数，这个函数通过卷积神经网络表示。这种方法使重建结果有较高的信噪比，但是缺少高频信息，出现过度平滑的纹理，人眼观感并不是很好。故为了生成符合人眼感观的图像，研究者提出了 SRGAN 模型[6]，网络主体采用对抗神经网络，损失函数采用感知损失与对抗损失之和。虽然这种方法重建出的图像峰值信噪比不是最高，但是模型产生的图像更加自然清晰，更符合人眼的视觉效果。作为生成对抗网络在超分辨率重建方面的开山之作，SRGAN也有其局限性。尽管其取得了很好的视觉效果，但随着网络的加深，批归一化层可能会使图像出现伪影。随后研究者又提出了ESRGAN[11]、SFTGAN等模型，其重建效果在视觉上也越来越自然。

3.8.3 基于递归神经网络的图像超分辨率重建

为了生成高质量的图像，多数超分辨率重建的卷积神经网络不使用池化层，但随着网络的加深，网络参数会逐渐增多，进而出现过拟合问题。为了解决以上问题，同时考虑到模型存储的便利性，研究者采用递归神经网络进行超分辨率重建任务，提出了DRCN[19]、DRRN[20]等模型，将递归神经网络与残差学习相结合，在模型参数量和复杂度较低的情况下仍能取得较为优秀的重建效果。

3.8.4 基于通道注意力的图像超分辨率重建

多数基于神经网络的模型主要专注于设计更深或更宽的网络以学习更具判别力的高层特征，却忽略了层间特征的内在相关性，故有研究者将通道注意力机制和残差模块相结合进行图像超分辨率重建，如RCAN[21]模型。同时，大多数基于神经网络的超分辨率方法没有充分利用原始低分辨率图像的信息，这些信息可以通过长条约连接直接传到网络的最后一层，使网络重点学习高频信息，减轻网络学习负担。RCAN模型主要由4个部分组成，分别是浅层特征提取模块、RIR（residual in residual）深度特征提取模块、上采样模块和重构图像模块。其中除了RIR深度特征提取模块，其余模块与大多数超分辨率重建网络相同。RIR深度特征提取模块由残差组与一个长跳跃连接组成，每个残差组由残差通道注意力模块与短跳跃连接组成，RCAN模型的网络深度可以超过400层。

3.9 小结

本章对深度学习基本理论及其典型方法进行了简单描述。首先介绍了机器学习的定义及其基本思想，梳理了神经网络、深度学习的基本原理及其发展历程；其次介绍了典型的深度神经网络，即卷积神经网络、生成式对抗网络、自编码器、循环神经网络等；最后介绍了深度学习在图像超分辨率重建中的应用，简要列举了基于深度学习超分辨率重建的模型。总的来说，本章大致梳理了深度学习的发展脉络，覆盖了深度学习在图像超分辨率重建领域的大部分应用，有助于读者对深度学习技术形成初步了解。

第4章

有监督的图像超分辨率重建方法

4.1 引言

有监督的图像超分辨率重建方法，顾名思义即由高分辨率图像作为真值指导网络学习的超分辨率重建方法。早期的深度学习图像超分辨率重建方法大多基于有监督的范式，这些方法在训练网络时需要低分辨率图像及其相应的高分辨率图像组成训练对。训练时，网络将低分辨率图像重建后获得的超分辨率输出与相应的高分辨率图像计算误差，作为损失函数或损失函数的一项。根据对超分辨率过程建模思路的不同，可将有监督的图像超分辨率方法分为判别式超分辨率和生成式超分辨率两种。前者将超分辨率过程建模为求解条件概率分布的过程，即解决在某个低分辨率输入的条件下应该有什么超分辨率输出的问题；而后者将超分辨率过程建模为求解联合概率分布的问题，即解决低分辨率图像和高分辨率图像共同组成何种概率分布的问题。本章将从判别式超分辨率模型和生成式超分辨率模型入手，分别介绍经典的有监督的图像超分辨率重建方法，从而建立起研究超分辨率重建问题的基础。

4.2 方法介绍

4.2.1 判别式超分辨率模型

1. 残差学习

在SRCNN中，研究人员发现，通过增加更多卷积层增加感受野，可以获得更好的重建性能。然而，直接堆叠层将导致梯度消失/爆炸和退化问题[22]。同时，增加更多的层将导致更高的训练误差和更昂贵的计算成本。

在ResNet中，He等提出了一个残差学习框架[23]，其中需要残差图，而不是拟合整个基础图。在SISR中，由于LR图像和HR图像共享大部分相同的信息，因

此很容易在LR图像与HR图像之间显式建模残差图像。残差学习可以实现更深层次的网络，并缓解梯度消失和退化问题。借助残差学习，Kim[5]提出了一种非常深的超分辨率网络，也称为VDSR。在VDSR的论文阐述中，作者明确指出，输入的图像在低分辨率状态下与输出的高分辨率版本在整体结构上存在显著的相似性。具体而言，这种相似性体现在两者包含的低频信息方面，即低分辨率图像的低频信息与高分辨率图像中的相应部分极为接近。然而，在训练过程中，若过于关注这部分低频信息的匹配，无疑会增加不必要的计算和时间成本。实际上，仅需聚焦学习高分辨率图像与低分辨率图像之间的高频部分差异，即所谓的残差。这种残差网络结构的设计理念，在解决超分辨率问题方面展现出极大的优势，其影响力甚至波及后续的深度学习超分辨率技术。而VDSR模型作为残差学习的一个直观且显著的实例，其网络结构已在图4-1[22]中得以呈现。

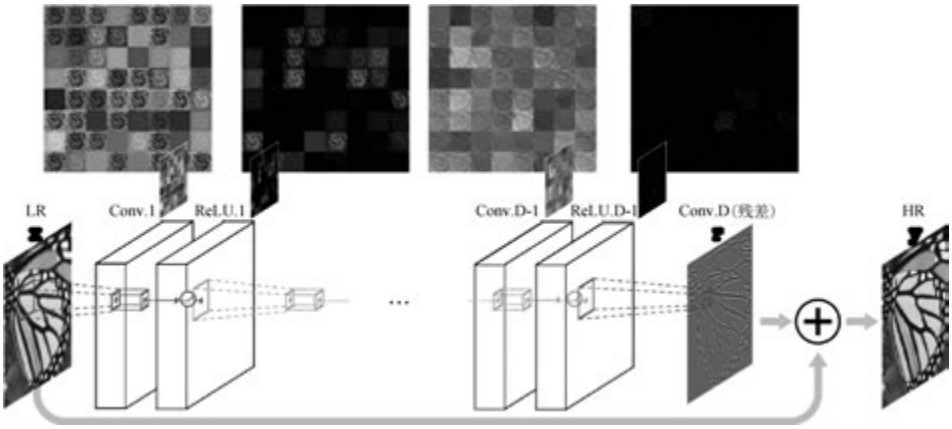

图 4-1　VDSR结构

在VDSR的设计中，网络结构得到了显著加深，达到20层，从而赋予了更深层网络更大的感受野。文章特别选取了3×3的卷积核，对于深度为D的网络，其感受野范围达到了$(2D+1)\times(2D+1)$。得益于残差学习的引入，由于残差图像的特性——较为稀疏，大部分值接近0或较小，训练过程得以快速收敛。此外，VDSR还巧妙地运用了自适应梯度裁剪策略，通过限制梯度的范围进一步加速了模型的收敛。为了确保特征图和最终输出图像在尺寸上的一致性，VDSR在每次卷积操作前都对图像进行了填充操作（补0），从而有效解决了图像在逐步卷积过程中尺寸逐渐减小的问题。值得一提的是，实验结果显示，这种填充操作还能改善对边界像素的预测效果。VDSR在训练过程中采用了独特的方法，将不同倍数的图像混合在一起进行训练，这样的训练策略使一个模型能够应对不同倍数的超分辨率问题，展示了其出色的泛化能力。

为了便于网络设计，残差块已逐渐成为网络结构中的基本单元。在卷积分支中，它通常有两个3×3卷积层、两个归一化层和一个ReLU激活函数。值得注意的是，由于EDSR[7]指出归一化层会消耗更多内存，但不会改善模型性能，因此

在SISR任务中通常会删除归一化层。

全局和局部残差学习：全局残差学习是从输入到最终重建层的跳跃连接，有助于改善信息从输入到输出的传输，并在一定程度上减少信息损失。然而，随着网络越来越深，大量图像细节经过这么多层后会不可避免地丢失。因此，提出了局部残差学习，它是在几个卷积层中执行的，而不是从输入到输出。在这种方法中，形成了多路径模式，承载了丰富的图像细节，并有助于梯度计算。此外，许多新的特征提取模块引入了局部残差学习，增强了学习能力[21, 24]。当然，将局部残差学习与全局残差学习相结合现在也非常流行[6, 9]。

在文献[6]中，作者使用GAN用于处理图像超精度SR，这是第一个对放大4倍自然图像做超分辨率的框架。为了实现这个框架，作者改进了目标函数，使用RestNet修复训练。对抗损失由判别器训练原始图像和SR图像的差异，使生成的图像更接近自然图像。内容损失由图像的视觉相似性生成，而不是像素空间的相似性。ResNet可以从下采样的图像恢复逼真的纹理。MOS测试作为图像效果的评判，最后的测试结果表明，采用SRGAN获得图像的MOS值比采用其他顶级方法获得图像的MOS值更接近原始的高分辨率图像。

有学者[19]设计了一个残差通道注意力网络RCAN，使网络变得更深并提高性能；提出了RIR结构，即用多个残差组和长跳跃连接构建粗粒度的残差学习，在残差组内部再堆叠多个简化的残差块并采用短跳跃连接（大的残差内部再套小残差）。这种结构可以使低频信息绕过网络，从而提高信息处理的效率。提出了通道注意力（channel attention，CA）机制，通过特征通道之间的相互依赖性重新调整特征权重。

从图4-2[19]中可以看出，RIR结构的最外层由G个残差组和一个长跳跃连接构成，从而形成一个粗粒度的残差学习。在每个残差组的内部，则是由B个残差通道注意力块（residual channel attention block，RCAB）和一个小的跳跃连接构成。简单来说，RIR就是大残差内部再嵌套小残差。长跳跃连接可以使网络在更加粗粒度的层次上学习残差信息。而短跳跃连接则是一种细粒度的跳跃连接，使大量网络不需要的低频信息得到过滤。为了进一步实现自适应的判别学习，作者提出了CA并在RCAB中进行了运用，其目的是给更有价值的通道赋予更高的权重。

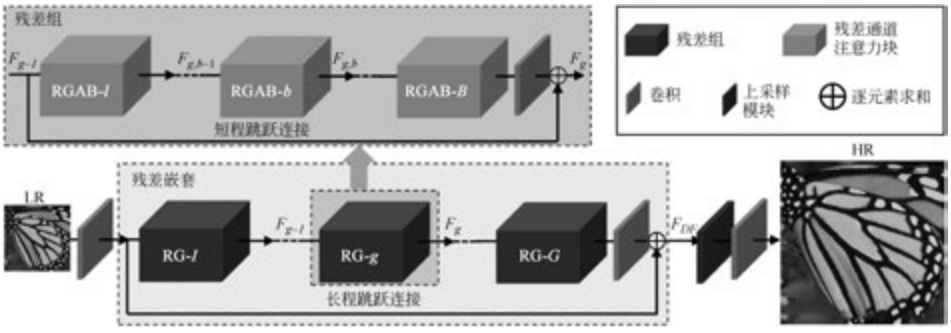

图 4-2　RCAN结构

在EDSR[7]中，Lim等发现，将特征图（通道维度）增加到一定水平会使训练过程数值不稳定。为了解决这些问题，他们采用了残差缩放，在将残差添加到主路径之前，通过乘以0和1之间的常数缩放残差。这种残差缩放的方法可以进一步提高模型性能。

在EDSR[7]中，作者对比了每个网络模型（原始ResNet、SRResNet和作者提出的网络）的基础模块。作者在网络中去除了批归一化（batch normalization，BN）层。由于批归一化层使特征标准化，同时去除了网络中的范围柔性，所以最好去除这些批归一化层。这一简单的修改可以大幅增加性能表现。另外，GPU 的内存使用率也会显著减少（因为批归一化层会消耗与之前卷积层等量的内存）。因此，可以创建一个更大型的模型，它在计算资源有限的情况下有着比传统ResNet更好的性能表现。在多尺度的架构设计中，作者创造性地纳入了尺度特定的处理单元，旨在实现对超分辨率的跨尺度精准控制。位于网络起始位置的预处理模块被用于减少输入图像在不同尺度上的变异性。每个这样的预处理模块均由两个配备5×5大小核的残差块组成。通过选择较大的核尺寸，网络在预处理阶段就能确保尺度特定的部分维持较浅的网络层次，进而在网络早期阶段就覆盖较大的感受野。作者并行布置了多个尺度特定的上采样模块，以达成多尺度的图像重构目标。

2. 递归学习

为了在不增加模型参数的情况下获得较大的感受野，针对SISR提出了递归学习，其中相同的子模块在网络中重复应用，并且共享相同的参数。在其他情况下，递归块是递归单元的集合，这些单元中的相应结构共享相同的参数。例如，同一卷积层在DRCN[19]中应用了16次，产生了41×41大小的感受野。DRCN结构如图4-3所示。

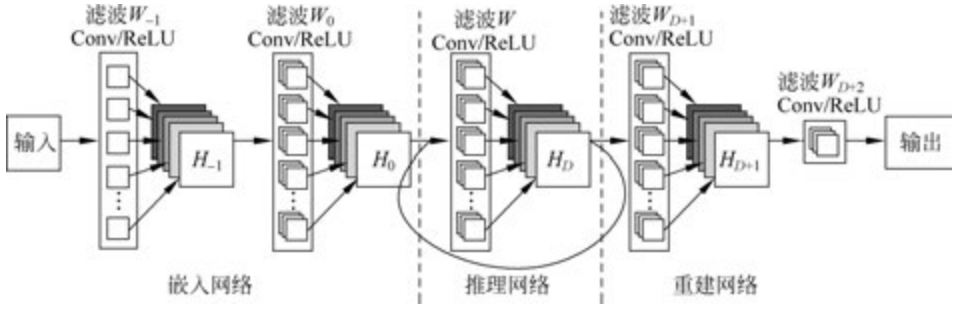

图 4-3 DRCN结构[19]

作者提出的基底模型由三个子网络组成：嵌入网络（embedding network）、推理网络（inference network）和重建网络（reconstruction network）。嵌入网络用于将给定图像表示为特征映射，推理网络加深网络深度，将嵌入网络的输出特征映射到更高维度。推理过程结束后，推理网络中获取的最终特征映射会作为输入，

被传递至重建网络，以产生期望的输出图像。图4-3中的输入图像是原始LR图像经过插值上采样后的图像。对于基准模型，作者并没能训练出成功的深度递归网络。为了解决梯度和最优递归问题，作者又提出了一种改进的模型。

监督递归：监督每层递归，以减轻梯度消失/爆炸的影响。假设在推理层中卷积过程反复使用相同的卷积核，使用相同的重建层预测每一次递归重建的SR图像。重建层输出D个预测图像，所有预测都在训练期间同时受到监督，这一步在公式中体现为增加了一部分损失。将所有D个预测图像通过加权求和计算最终输出(权重由网络学习得到)。通过递归减轻了训练递归网络的困难，通过反向传播对不同预测损失产生的反传梯度求和提供平滑效果，能够有效缓解梯度爆炸或消失。此外，鉴于监督机制能够有效利用来自所有中间层的预测数据，从而减弱对于确定最佳递归次数这一选择的依赖性。

跳跃连接：在图像重建任务中，鉴于输入和输出图像之间的紧密相关性，可以直接通过跳层连接将LR信息直接传输到SR重建层。该做法有两个优点：节约了远距离传输的复杂算力，极大程度地保留了完整的低频信息。

然而，在基于递归学习的模型中，过多的卷积层仍然会导致梯度消失/爆炸问题。因此，在DRRN[20]中，递归块基于残差学习进行。DRRN中引入了全局残差学习（global residual learning，GRL）。在视觉识别和图像恢复等应用领域，深度网络的使用有时可能导致性能不佳，这主要是由于经过多层网络处理后，图像中的众多细节信息可能会被逐渐削弱甚至丢失。为了解决这一问题，作者提出了一种增强的残差单元结构，即多路径模式局部残差学习（local residual learning，LRL），其特色在于识别分支，这一设计旨在将图像中丰富的细节信息传递至网络深处，以确保信息的有效流通和深层次的特征提取，GRL和LRL的主要区别在于LRL是在几个堆叠的层中执行，而GRL是在输入和输出图像之间执行，即DRRN有许多LRL，但只有一个GRL。为了保持模型的紧凑性，在DRRN中提出了残差单元的递归学习。与DRCN相比，DRRN有两个主要区别：①与DRCN不同，DRRN并未采用在卷积层间共享权重的做法，而是设计了一个递归块，这个块由几个共享同一权重集的残差单元组成。②为了应对极深模型中可能出现的梯度消失或爆炸问题，DRCN采取了监督每个递归的策略，确保早期递归的监督信息有助于反向传播。而DRRN则通过构建一个具有多路径结构的递归块降低这一挑战。最近，MemNet[25]、CARN[26]和SRRFN[27]等模型在其递归单元中引入了残差学习策略，以进一步优化模型性能。

3．课程学习

课程学习作为一种提升训练成效的策略，通过渐进地增加学习目标的难度优化训练过程。在初始阶段，课程学习的研究主要聚焦单一任务的训练，然而，随着研究的深入，Pentina等利用序列方法将课程学习的概念扩展至多任务场景，而

Gao等则将其引入图像恢复领域，旨在解决依赖问题（fixation problem）。鉴于这些网络模型需要做出一次性预测，研究者在网络训练轮次逐步累积的过程中，通过调整训练数据的复杂度实现课程学习，即按照任务的难易程度输入不同的训练样本。在图像超分辨率重建任务中，Wang等为金字塔结构设计了一种课程，通过将金字塔的新层级逐渐融合进先前训练好的网络，将一幅LR图像放大到更大的尺寸。相对于先前只注重单一降质过程的研究，SRFBN按多种降质方式得到退化LR图像以实现课程学习，解决这些越来越难课程的过程中可以逐渐重建退化LR图像。

SRFBN采用L1损失优化网络，采用T幅目标HR图像$(I_{\mathrm{HR}}^1, I_{\mathrm{HR}}^2, \cdots, I_{\mathrm{HR}}^T)$适配多种网络输出。每种降质模型共用相同的$(I_{\mathrm{HR}}^1, I_{\mathrm{HR}}^2, \cdots, I_{\mathrm{HR}}^T)$。对于复杂的降质模型，$(I_{\mathrm{HR}}^1, I_{\mathrm{HR}}^2, \cdots, I_{\mathrm{HR}}^T)$会根据任务的困难程度迭代$T$次顺序以实现课程。因此网络损失函数为

$$L(\Theta) = \frac{1}{T}\sum_{t=1}^{T} W^t ||I_{\mathrm{HR}}^t - I_{\mathrm{SR}}^t||_1 \tag{4-1}$$

其中，Θ表示网络参数，W^t表示第t轮迭代输出的权重的常系数，文中将该系数设置为1，即所有输出做出相同贡献，最终的超分辨率结果是最后一轮迭代输出I_{SR}^T。文中指出，随着T的增加，图像超分辨率重建质量也在提升，并在实验分析中将其设置为4。

4. 注意力机制

考虑一个人在复杂环境中寻找特定目标的场景，大脑往往对靠近视野中心的具有显著特征的物体更感兴趣，并会自动忽略周围不突出的物体。在这些感兴趣对象中分析并筛选目标，相比逐一对所有对象进行挑选，能有效降低思考压力、加快分析过程。总体上，这是一种将现有资源分配到输入中包含最有用信息成分的机制。这种机制在20世纪90年代被认知科学家发现并命名为注意力机制（attention mechanism），随后被相继应用于计算机视觉和自然语言处理领域，并在图像分类、目标检测、三维视觉、多模态任务和自监督学习等方面取得了优秀的成果。应用于计算机视觉领域的注意力方法可概括为对特征图的不同维度进行动态加权，为重要的特征分量赋予更大的权重，相对不重要的特征分量则分配更小的权重，赋权的过程可以看作注意力机制指导网络关注特征图中对任务来说更重要的部分。按照添加注意力机制的维度，可将计算机视觉领域主要的注意力方法分为通道注意力、空间注意力、时间注意力和分支注意力4类。

如图4-4所示，通道注意力的原理是在特征图的通道维度上生成一个掩模并选出重要的通道。深度神经网络生成的不同特征图的不同通道通常代表不同目标，通道注意力则对每个通道自适应地赋予权重，该过程可以看作目标选择过程。早期的通道注意力机制用于提取全局信息、捕获通道尺度的相关性和提升网络表示能

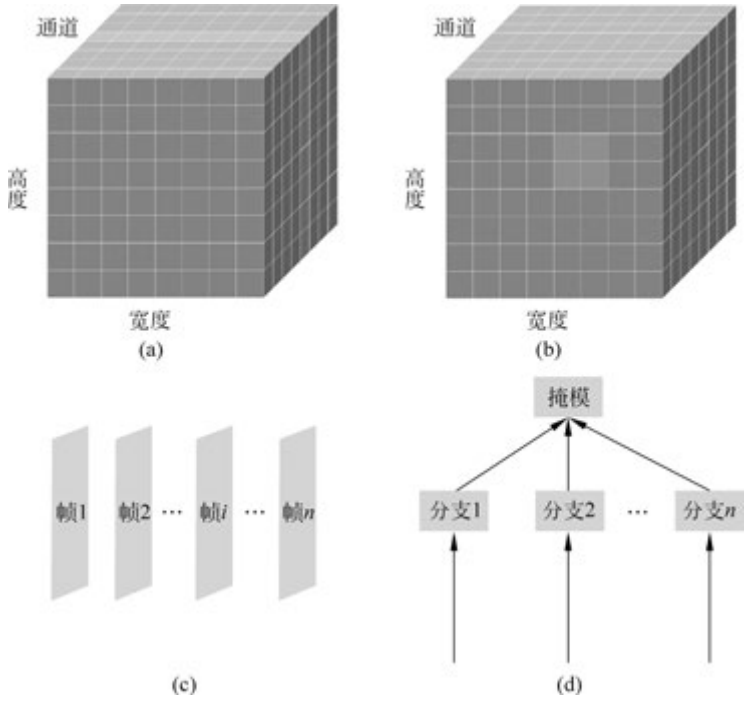

图 4-4　计算机视觉领域的注意力机制

(a) 通道注意力；(b) 空间注意力；(c) 时间注意力；(d) 分支注意力

力，但这些早期方法具有无法捕获复杂的全局信息和引入全连接层导致网络复杂度增加的缺点。随后有研究者提出用全局二阶池化（global second-order pooling）替代全局平均池化（global average pooling，GAP），在获取全局信息的同时对高阶统计信息建模。该方法提升了注意力模块收集全局信息的能力，但额外增加了计算量。因此基于减轻计算量和全连接层的复杂度的需求，研究者提出门控通道注意力（gated channel attention），更有效地获取信息和显式地对通道尺度的相关性进行建模。门控通道注意力的一个特点在于不仅将输入特征图和注意力向量相乘得到注意力特征，还引入跨层连接，将注意力特征和原始输入相加后作为最终输出。该方法参数量更少，并且它的轻量化使其可以添加到卷积神经网络的每层卷积层后。另一种高效通道注意力（efficient channel attention，ECA）方法使用一维卷积确定通道的相关关系，从而实现对输入和权重向量关系的直接建模，提升了结果质量。高效通道注意力模块可以用下式描述：

$$s = F_{\text{eca}}(X, \theta) = \sigma(\text{Conv1}D(\text{GAP}(X))) \tag{4-2}$$

$$Y = sX \tag{4-3}$$

其中，X 和 Y 分别是输入和输出特征图，s 为注意力模块根据参数 θ 计算得到的权重向量。在高效通道注意力模块中，输入 X 首先经由全局平均池化 $\text{GAP}()$ 提取全

局信息，然后通过一维卷积Conv1D()直接获得特征通道间的相关关系，最后由sigmoid函数σ()非线性化后得到s。s和X相乘可实现对特征图通道的加权，即完成一次对通道施加注意力的过程。高效通道注意力提供了速度和效果兼备的注意力模块，能方便地添加到多种卷积神经网络中。

5．Transformer

Transformer是一种基于自注意力（self-attention）机制的神经网络架构，最初于2017年由Google Brain团队的研究员提出并被用于自然语言处理领域中的机器翻译任务，该模型的表现远远超过了以往的序列模型，因此被认为是一种革命性的神经网络架构，引起了领域内的广泛关注。传统的用于自然语言处理任务的序列模型，如循环神经网络和长短期记忆网络，在处理长序列数据时存在梯度消失和梯度爆炸等问题，导致网络在学习过程中常会“忘记”之前学到的内容，也就是无法捕捉数据中的长期依赖（long-term dependencies）关系。但Transformer中的自注意力机制可以使模型关注输入序列中不同位置的信息，并根据这些信息进行加权计算，从而更好地捕捉序列中的长期依赖关系。

与上一节中的注意力机制不同，自注意力机制是一种在序列或集合中学习上下文相关性的机制，它允许模型在处理序列数据时，将注意力集中于序列中不同位置的元素。在自然语言处理任务中，序列通常被表示为单词或标记的序列，自注意力机制不仅考虑输入序列中不同位置或元素之间的关系，而且能够自适应地计算每个元素和其他元素之间的相似性，从而能够更好地捕捉序列或集合中元素之间的长期依赖关系，使模型更容易理解单词之间的依赖关系和语义信息。在自注意力机制中，每个单词或标记，即序列元素，都会经线性变换后生成3个向量，分别是查询向量（query vector）、键向量（key vector）和数值向量（value vector）。然后使用点积或其他函数计算查询向量和键向量之间的相似度，之后对于每个查询向量，根据其与键向量的相似度，计算出一个注意力分布（通常为了保证梯度能够计算，需要使用Softmax函数计算该分布），并将该分布与数值向量进行加权平均，得到该查询向量的自注意力表示。最后将所有查询向量的自注意力表示组合起来，形成整个序列的自注意力表示，整个自注意力表示的计算过程可由下式表示：

$$\text{Attention}(Q,K,V)=\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4-4}$$

式(4-4)中的Q、K和V分别是查询向量、键向量和数值向量组合成的矩阵，d_k指键向量的维度，在自注意力的计算过程中作为一项比例因子，用于缩放查询矩阵和键矩阵点乘的结果。

2020年，Google Brain团队首次提出用于计算机视觉领域的视觉Transformer（vision Transformer，ViT）。作为对自然语言处理中Transformer机制的发展和在计算机视觉领域的迁移应用，视觉Transformer同样用于提取图像中不同位置间

的依赖关系，因此可以视为能提取图像的全局语义信息的机制。与传统的卷积神经网络使用固定大小的卷积核不同，视觉 Transformer 使用自注意力机制进行特征提取，并且可以自适应地处理不同尺寸的输入图像。具体来说，视觉 Transformer 首先将输入图像划分为一系列小的图像块（patch），每个图像块都被视为一个“单词”，随后通过图像块嵌入（patch embedding）操作将这些图像块转换为向量表示。最后视觉 Transformer 将这些向量表示输入多个 Transformer 编码器层，进行特征提取和信息压缩。

这种 Transformer 编码器结构如图 4-5所示。在每个编码器层中，自注意力机制首先自动地学习图像块之间的关联性，从而捕捉图像中不同位置之间的语义关系。其次自注意力层之后的多层感知器对这些关系进行进一步学习和压缩，得到更抽象、更高级的特征表示。最后视觉 Transformer 将所有图像块的特征表示汇总为一个固定大小的特征向量，该特征向量随后用于图像分类或其他下游任务。

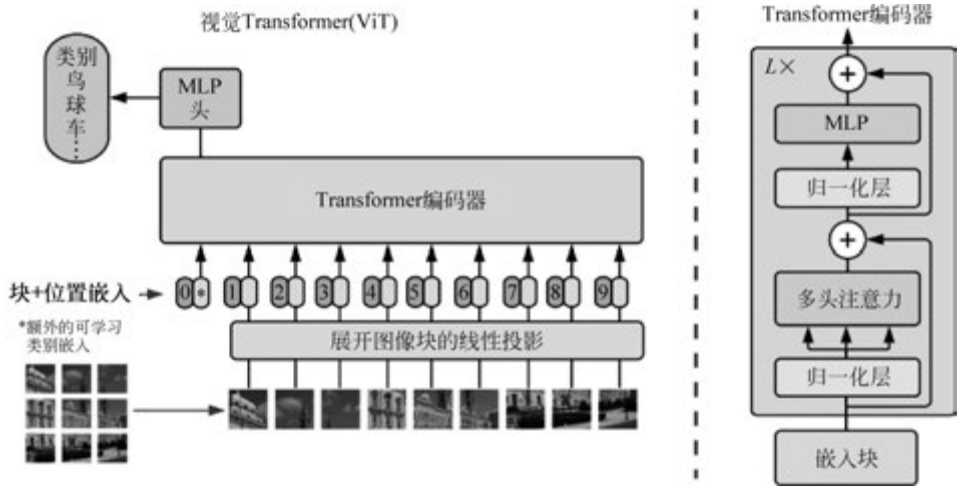

图 4-5　单层 Transformer 编码器结构

视觉 Transformer 的优点在于不受卷积核大小的限制，能够对任意大小的图像进行处理，并且能够通过自注意力机制学习到图像中不同位置之间的语义关系，捕获图像的全局语义信息。在图像分类、图像生成、目标检测和语义分割等计算机视觉任务方面，视觉 Transformer 已经取得了与同期最先进的卷积神经网络相当甚至更优秀的表现，目前 Transformer 在计算机视觉领域的应用还在不断发展中。

4.2.2　生成式超分辨率模型

1. 生成式对抗超分辨率

超分辨率重建的传统方法一般处理的是较小的放大倍数，当图像的放大倍数为 4 以上时，很容易使得到的结果过于平滑，而缺少细节上的真实感。这是因为传

统的方法使用的代价函数一般是最小均方差，即

$$l_{\mathrm{MSE}}^{\mathrm{SR}}=\frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I_{x,y}^{\mathrm{HR}}-G_{\theta_{\mathrm{G}}}(I^{\mathrm{LR}})_{x,y})^2 \tag{4-5}$$

该代价函数使重建结果有较高的峰值信噪比，但是缺少高频信息，出现过度平滑的纹理。有一种方法称为SRGAN，提出将深度残差网络作为生成器的生成对抗网络，与以往不同的是，ResNet的优化目标不只是MSE，还有VGG网络与判别器构成的感知损失函数。它主张在重建高分辨率图像时，确保这些图像不仅在像素级的低层次细节方面，而且在更高层次的抽象特征、整体概念乃至风格表现方面，都与真实的高分辨率图像保持相近的吻合度。整体概念和风格如何评估呢？可以使用一个判别器，判断一幅高分辨率图像是由算法生成的还是真实的。如果一个判别器无法区分，那么通过算法生成的图像就达到了以假乱真的效果。

因此，可以将代价函数改进为

$$l^{\mathrm{SR}}=l_X^{\mathrm{SR}}+10^{-3}l_{\mathrm{Gen}}^{\mathrm{SR}} \tag{4-6}$$

第一部分是基于内容的代价函数，第二部分是基于对抗学习的代价函数。基于内容的代价函数除了上述像素空间的最小均方差$L_{\mathrm{MSE}}^{\mathrm{SR}}$外，还包含一个基于特征空间的最小均方差，该特征是利用VGG网络提取的图像高层次特征：

$$l_{\mathrm{VGG}/i,j}^{\mathrm{SR}}=\frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\Phi_{i,j}(I^{\mathrm{HR}})_{x,y}-\Phi_{i,j}(G_{\theta_{\mathrm{G}}}(I^{\mathrm{HR}}))_{x,y})^2 \tag{4-7}$$

对抗学习的代价函数是基于判别器输出的概率：

$$l_{\mathrm{Gen}}^{\mathrm{SR}}=\sum_{n=1}^{N}-\log D_{\theta_{\mathrm{G}}}(G_{\theta_{\mathrm{G}}}(I^{\mathrm{HR}})) \tag{4-8}$$

其中，$D_{\theta_{\mathrm{G}}}$是一个图像属于真实高分辨率图像的概率，$G_{\theta_{\mathrm{G}}}(I^{\mathrm{LR}})$是重建的高分辨率图像。SRGAN使用的生成网络和判别网络如图4-6所示。

在训练SRGAN网络的过程中需要提供高分辨率图像，作者首先对高分辨率图像进行降采样得到低分辨率图像，然后将低分辨率图像输入，训练生成器，使之生成对应的高分辨率图像。训练生成器的过程与训练前馈神经网络一样，都是对网络参数进行优化，如下所示：

$$\hat{\theta_{\mathrm{G}}}=\mathrm{argmin}_{\theta_{\mathrm{G}}}\frac{1}{N}\sum_{n=1}^{N}l^{\mathrm{SR}}(G_{\theta_{\mathrm{G}}}(I_n^{\mathrm{LR}}),I_n^{\mathrm{HR}}) \tag{4-9}$$

进一步，定义判别器，如同Goodfellow提出的GAN网络一样，生成器和判别器交替优化下面这个式子：

$$\begin{aligned}&\min_{\theta_{\mathrm{G}}}\max_{\theta_{\mathrm{D}}}E_{I^{\mathrm{HR}}}\sim p_{\mathrm{train}}(I^{\mathrm{HR}})[\log D_{\theta_{\mathrm{D}}}(I^{\mathrm{HR}})]+\\&E_{I^{\mathrm{LR}}}\sim p_{\mathrm{G}}(I^{\mathrm{LR}})[\log(1-D_{\theta_{\mathrm{D}}}(G_{\theta_{\mathrm{G}}}(I^{\mathrm{HR}})))]\end{aligned} \tag{4-10}$$

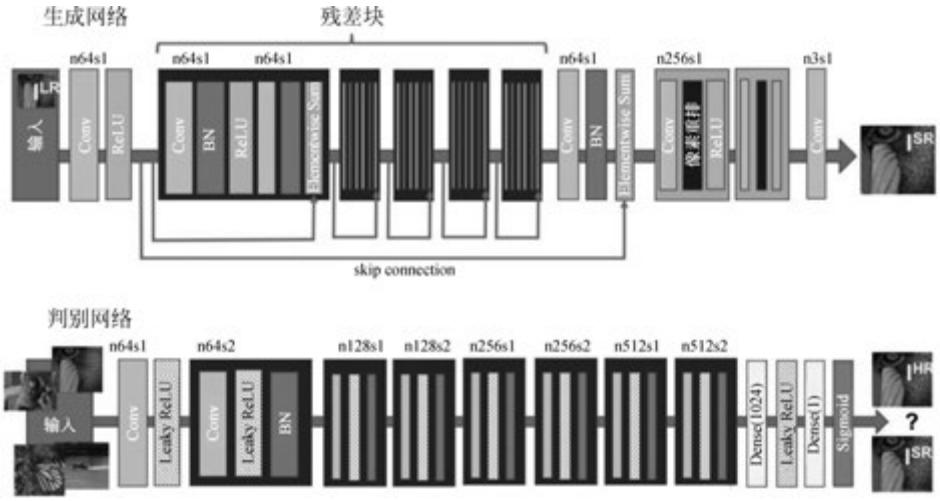

图 4-6　SRGAN 使用的生成网络和判别网络

在生成网络（具体为SRResNet）的构建中，多个残差块串联而成，每个残差块内均嵌入两个 3×3 的卷积层，其后紧跟的是BN层及PReLU作为激活函数。为了提升特征图的尺寸，设计中巧妙地采用了两个 2×2 的亚像素卷积层（sub-pixel convolution layers）。而在判别网络的设计中，则涵盖8个卷积层，这些卷积层随深度的增加，不仅特征数量递增，而且特征尺寸逐渐减小。选择LeakyReLU作为激活函数，并通过两个全连接层及最后的Sigmoid激活函数，输出图像被判别为自然图像的概率。SRGAN的损失函数构成如下：

$$l^{\mathrm{SR}} = l_X^{\mathrm{SR}} + 10^{-3} l_{\mathrm{Gen}}^{\mathrm{SR}} \tag{4-11}$$

在构建损失函数时，内容损失的一种可能形式是基于均方误差的度量方法，这种方法衡量了预测值与实际值之间的平均差异：

$$l_{\mathrm{MSE}}^{\mathrm{SR}} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I_{x,y}^{\mathrm{HR}} - G_{\theta_{\mathrm{G}}}(I^{\mathrm{LR}})_{x,y})^2 \tag{4-12}$$

内容损失也可以基于已训练好的VGG模型，并采用ReLU作为激活函数，其计算公式如式(4-7)所示。其中，i 和 j 标识了VGG19网络中特定位置的特征，即第 i 个最大池化层后的第 j 个卷积层提取的特征。至于对抗损失，其表达式见式(4-8)。

采用基于均方误差的损失函数训练SRResNet时，虽然能够获得较高的峰值信噪比，但这种方法往往会忽略一些高频细节，使生成的图像较为平滑。相对而言，SRGAN在视觉效果上更胜一筹。为了比较不同内容损失函数的效果，我们分别尝试了基于均方误差、基于VGG模型低层特征及基于VGG模型高层特征三种设置。实验结果显示，基于均方误差的内容损失表现最差，而基于VGG模型高层特征的内容损失能够生成更精细的纹理细节，优于基于VGG模型低层特征的情况。

近些年，还有一种增强型的SRGAN网络模型，称为ESRGAN。它引入残差

密集块，并且将残差密集块中的残差作为基本的网络构建单元而不进行批量归一化（有助于训练更深的网络），并且使用残差缩放（residual scaling）。此外，运用相对生成对抗网络的核心理念，使判别器专注于评估图像之间的相对真实性，而非追求绝对的真实度标准，即实现相对生成对抗网络的架构。利用VGG激活前的特征值改善感知损失，会使生成的图像有更清晰的边缘（为亮度的一致性和纹理恢复提供更强的监控）。

生成器部分生成网络的作用是输入一幅低分辨率图像，生成高分辨率图像。网络由以下几部分组成。

（1）浅层特征抽取网络，提取浅层特征。低分辨率图像进入后会经过一个卷积+ReLU函数，将输入通道数调整为64。

（2）RRDB网络结构，包含N个密集残差块（residual dense block，RDB）和1个残差连接，每个RDB都包含5个卷积+RELU。

（3）上采样网络，进入上采样部分并经过两次上采样后，原图的高宽变为原来的4倍，并实现了分辨率的提升。

其结构如图4-7所示。

图 4-7　ESRGAN使用的生成网络的结构

密集残差块的网络结构如图4-8所示：RRDB采用两层残差结构，外层由一个大的残差结构构成，主干部分由3个RDB密集残差块构成，将主干网络的输出与残差边叠加。每个RDB块都有5个卷积。

密集残差块和残差块、密集块的对比如图4-9所示。

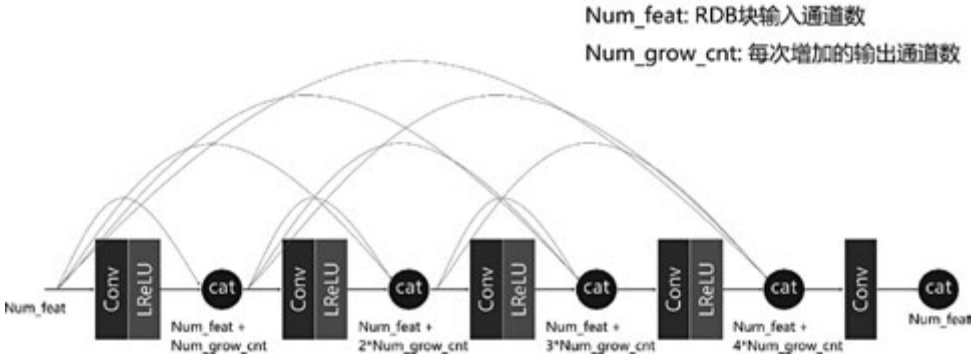

图 4-8　密集残差块的网络结构

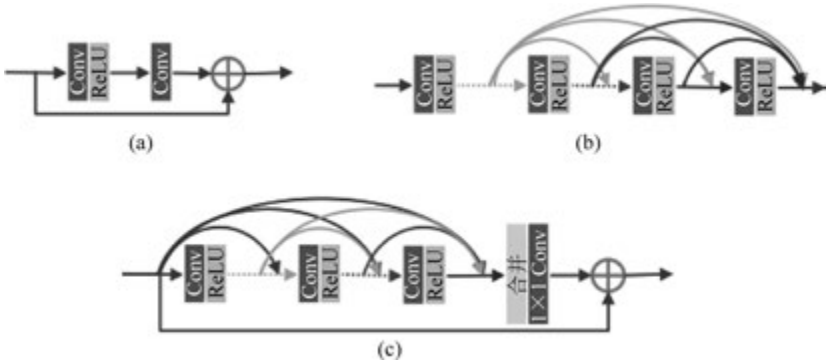

图 4-9　密集残差块和残差块、密集块的对比

(a) 残差块；(b) 密集块；(c) 密集残差块

RDB 的构造中融合了残差块与密集块的概念。通过密集的卷积层连接，这一结构可用于提取丰富的局部特征。其特色在于，将先前的RDB状态直接关联至当前RDB的每一层，形成一种直接的链接。随后，RDB利用这些局部特征进行融合，自适应地从先前和当前的局部特征中提炼出更有效的特征，从而确保训练的稳定性和高效性。

相较传统的密集块，这种网络结构选择了剔除BN层。在训练过程中，BN层会依赖当前批次的均值和方差归一化特征，而在测试时，它利用整个训练数据集的估算均值和方差。然而，当训练集与测试集在统计特性上存在显著差异时，BN层往往会引入不利的伪影，从而削弱模型的泛化能力。以经验观察到，BN层有可能在网络更深和在GAN下训练时带来伪影。在迭代过程中及不同的配置设定下，这些不稳定的伪影偶尔出现，明显与我们追求的稳定性能超越训练目标相悖。因此，决定去除BN层，以确保训练的稳定性和一致性。另外，由于批处理标准化层与前置卷积层的内存消耗相当，此举也显著降低了GPU的内存使用，进一步提升了训练效率。总之，去除BN层有助于提高泛化能力，降低计算复杂度和内存使用率。网络由多个基础块构成，每个基础块都由RRDB组成，其中RRDB即在级联3个RDB的基础上添加残差边。基于密集残差块的残差网络如图4-10所示。

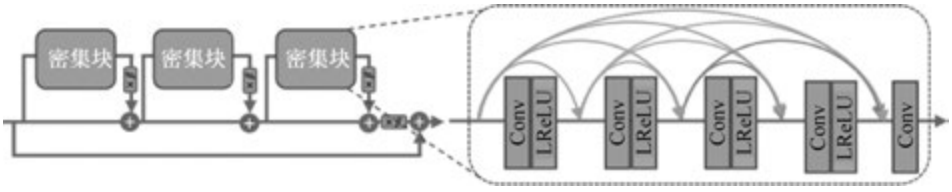

图 4-10　基于密集残差块的残差网络

鉴别器部分引入了一种相对鉴别器的概念。该鉴别器的工作方式是试图预测真实图像x_r相对于假图像x_f，具有更高真实性的概率，如图4-11所示。具体来说，用相对论（relativistic）平均鉴别器RaD代替标准鉴别器，如图4-11(b)所示。

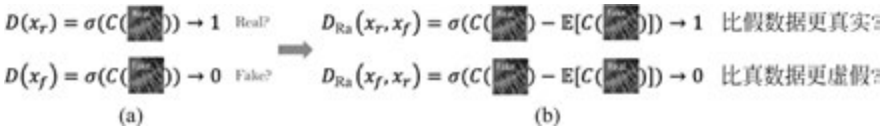

图 4-11 标准鉴别器和相对论平均鉴别器

(a) 标准鉴别器；(b) 相对论平均鉴别器

使用BCE函数计算生成损失和对抗损失，用于测量目标和输出之间的二进制交叉熵。损失函数有三部分，生成器的损失函数为

$$L_{\mathrm{G}} = L_{\mathrm{percep}} + \lambda * L_{\mathrm{G}}^{\mathrm{Ra}} + \eta L_1$$

$$\text{s.t.} \begin{cases} L_{\mathrm{percep}} = \|\phi(x_i^{\mathrm{SR}}) - \phi(G(x_i^{\mathrm{LR}}))\|_1 \\ L_{\mathrm{G}}^{\mathrm{Ra}} = -E_{x_i}[\log(1 - D_{\mathrm{Ra}}(x_r, x_f))] - E_{x_f}[\log(D_{\mathrm{Ra}}(x_f, x_r))] \\ L_1 = E_{x_i}\|G(x_i) - y\|_1 \end{cases}$$

其中，感知损失L_{percep}与SRGAN不同的是使用了激活前的特征图作为重构特征，并使用了一个在MINC数据集上微调后的VGG网络，其更关注纹理识别，而不是目标识别。

2．先验生成式对抗超分辨率

为了解决超分辨率任务中图像的纹理信息恢复不好的问题，提出了一种新的算法SFTGAN。一开始的超分辨率网络使用MSE损失函数，导致生成的图像比较模糊或者存在过于光滑的纹理。后来开始改进损失函数，使用感知损失函数和对抗损失函数，但是依然存在图像的纹理信息恢复不好的问题。SFTGAN在超分辨率的合成中使用语义图，语义图的生成使用了图像分割网络。文章探讨了不同分辨率下语义分割的误差，比较后发现，其实高低分辨率图像对于分割的精度影响不大。网络生成器整体结构如图4-12所示。

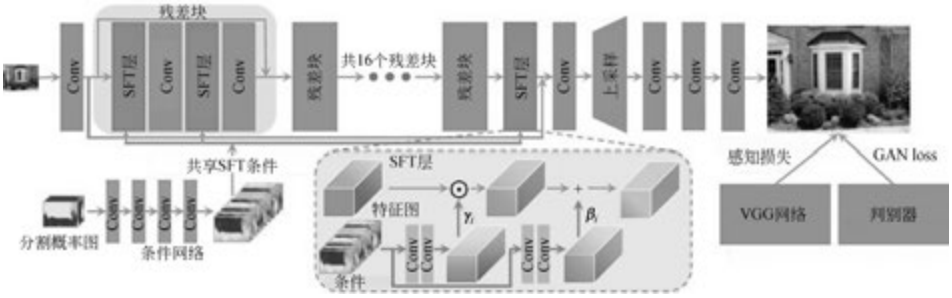

图 4-12 网络生成器整体结构

3．空间特征调制层

下一个问题是怎么把语义图和超分辨率模型结合起来，如果为每个类别训练一个超分辨率模型显然不可能。如果简单地将语义图和中间的特征图合并又无法充分发挥语义图的作用，作者的方案是采用空间特征调制（spatial feature trans-

form，SFT）层模块。它能将额外的图像先验（比如语义分割概率图）有效地结合到网络中，恢复出与所属语义类别特征一致的纹理。SFT层以语义分割概率图为条件，基于语义分割概率图，生成一对调制参数，以在空间上对网络的特征图应用仿射变换。SFT层有以下3点优势。

（1）节约参数。

（2）即插即用，容易与现有模型结合。

（3）可扩展，先验可以是语义图，也可以是深度图等。

其受到条件BN层的启发，但是条件BN层以及其他特征调制层（比如FiLM）往往忽略了网络提取特征的空间信息，即对于同一个特征图的不同位置，调制的参数保持一致。但是超分辨率等底层视觉任务往往需要考虑更多的图像空间信息，并在不同的位置进行不同的处理。基于这个观点提出了空间特征调制层，其结构如图4-13所示。

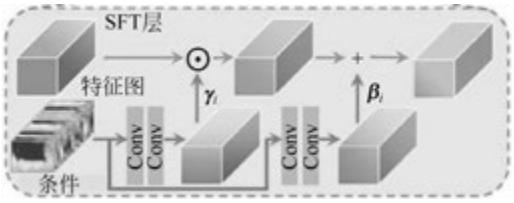

图 4-13　空间特征调制层结构

SFT层对网络的中间特征进行仿射变换，变换的参数由额外的先验条件经过若干层神经网络变换得到。若以F表示网络的特征，γ和β分别表示得到的仿射变换的尺度和平移参数，那么经过空间特征调制层得到的输出特征为

$$\mathrm{SFT}(F|\gamma,\beta)=\gamma\odot F+\beta \tag{4-13}$$

SFT层有两个输入，一个输入是条件网络的输出，另一个输入是上一层的输出F。条件网络计算出γ、β，继而计算出整个SFT层的输出，而整个SFT层又作为下一层的输入。空间特征调制层可以方便地被集成至现有的超分辨率网络，如SRResNet等。图4-12是本文使用的网络结构。为了提升算法效率，先将语义分割概率图经过一个条件网络得到共享的中间条件，然后将这些条件广播至所有的SFT层。算法模型在网络的训练中，同时使用了感知损失函数和对抗损失函数。

网路的输出为

$$\hat{y}=G_\theta(x|\gamma,\beta),\quad (\gamma,\beta)=M(\Phi) \tag{4-14}$$

其中，Φ是先验，M是一个映射函数，把先验映射为γ、β。映射函数可以是任意的函数，或者是一个神经网络，其参数随主网络一起训练。

4. 先验的语义分割图

低分辨率图像先经过双线性插值进行上采样，再经过一个分割网络作为输入。分割网络在COCO数据集上进行训练，在ADE数据集上进行微调。图4-14显示了

语义分割结果，当前基于深度学习的语义分割网络在低分辨率数据集上进行微调后，在大多数场景下都能够生成较满意的分割结果。

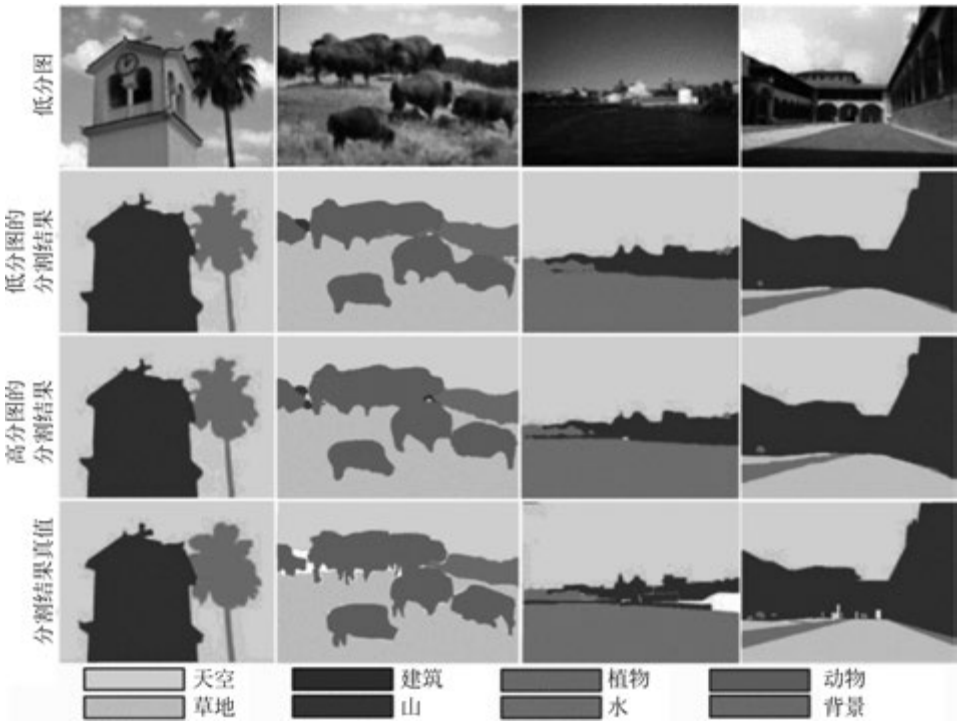

图 4-14　语义分割结果

研究了由低分辨率图像获得的分割图的准确性。在典型的超分辨率研究中，低分辨率图像从高分辨率图像以×4的比例因子进行下采样。在这种分辨率下，即使在基于现代CNN的分割模型的低分辨率图像上，仍然可以获得满意的分割结果。一些低分辨率图像和相应的语义分割结果如图4-14所示，低分辨率分割接近高分辨率分割。还没有尝试对小对象进行分割，因为这在图像分割领域仍然是一个具有挑战性的问题。在测试期间，不属于预定义的K分段类的类将被归类为“背景”类。在这种情况下，此方法仍然会生成一组默认值γ、β，将自身退化为SRGAN，即平等对待所有类。实际场景中，物体类别的分隔界限通常并不十分明显，比如植物和草的区域，它们之间的过渡是“无缝”且连续的，而作者使用的语义分割概率图以及调制层的参数也是连续变化的。因此，SFT-GAN可以更精细地调制纹理的生成。

5．总体网络结构

图4-12是生成器G_θ的结构。假设判别器为D_η，使用GAN的最小最大优化进行训练，目标函数为

$$\min_{\theta}\max_{\eta} E_{y\sim P_{\mathrm{HR}}}\log D_\eta(y)+E_{x\sim P_{\mathrm{LR}}}\log(1-D_\eta(G_\theta(x))) \tag{4-15}$$

网络分为两部分：条件网络和超分辨率网络。条件网络输入为语义图，经过4层卷积(卷积核1)，输出为中间条件。超分辨率网络使用16个残差块，每个残差块中都有SFT层，如图4-12所示，这些SFT层共享中间条件。不同的SFT层有不同的γ、β值。

6．损失函数

模型中应用感知损失和对抗损失。感知损失度量特征空间中的距离。为了获得特征图，使用预先训练的19层VGG网络，表示为ϕ。感知损失为

$$L_{\mathrm{P}}=\sum_{i}||\Phi((\hat{y})_i)-\Phi(y_i)||_2^2 \tag{4-16}$$

使用第5个最大池层之前的第4个卷积获得特征映射，并计算其特征激活的MSE。来自GAN的对抗性损失L_{D}也用于鼓励生成器支持多种自然图像中的解决方案。

$$L_{\mathrm{D}}=\sum_{i}\log(1-D_{\eta}(G_{\theta}(x_i))) \tag{4-17}$$

7．与其他算法比较结果

图4-15展示了SFT-GAN模型与其他模型比较的结果，可以看到基于GAN的算法模型SRGAN、EnhanceNet及SFT-GAN的视觉效果超过了以优化PSNR为目标的模型。SFT-GAN在纹理恢复方面能够生成比SRGAN和EnhanceNet更自然真实的结果（图4-15中的动物毛发、建筑物的砖块以及水的波纹）。

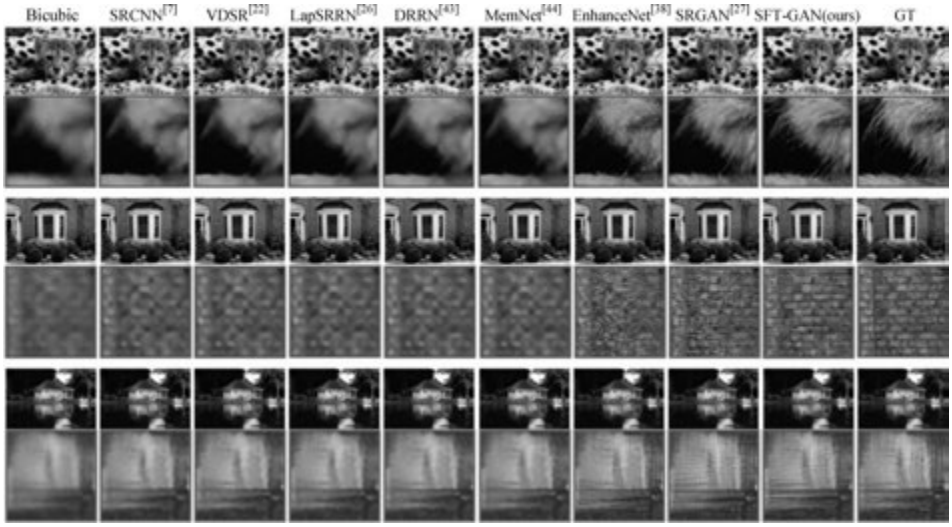

图 4-15　SFT-GAN模型与其他模型比较的结果

8．用户研究

进行一项用户研究以量化不同方法重建具有感知说服力图像的能力。为了更好地将方法与面向PSNR的基线和基于GAN的方法进行比较，将评估分为两个阶

段。第一阶段重点讨论面向PSNR的基线。要求用户对每幅图像的4个版本进行排名：SRCNN、MemNet、SFT-GAN和原始HR图像，根据他们的视觉质量进行排名。使用从户外场景测试中随机选取的30幅图片，所有图片都以随机方式呈现在图4-15中。第二阶段重点介绍了基于GAN的方法，以便用户专注于纹理质量。向受试者展示超分辨率图像对（描绘放大的纹理块以便比较）。每一对都由提议的SFT-GAN和SRGAN或EnhanceNet生成的对应物的图像组成。要求用户选择纹理更自然、更逼真的图像。共涉及96幅随机选择的图像。作者要求30名用户完成用户研究。第一阶段和第二阶段的结果分别显示在图4-16和图4-17中。

彩图 4-16

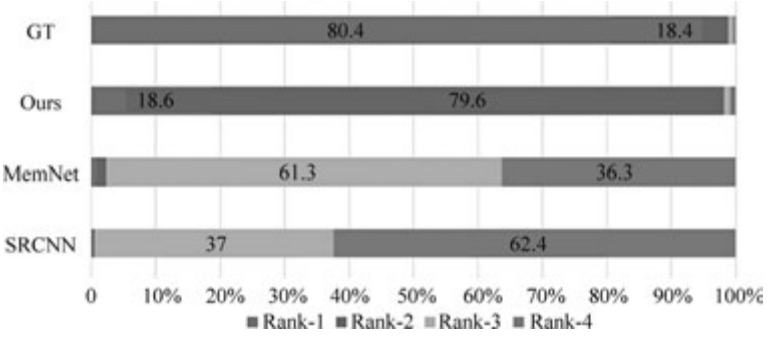

图 4-16 第一阶段结果

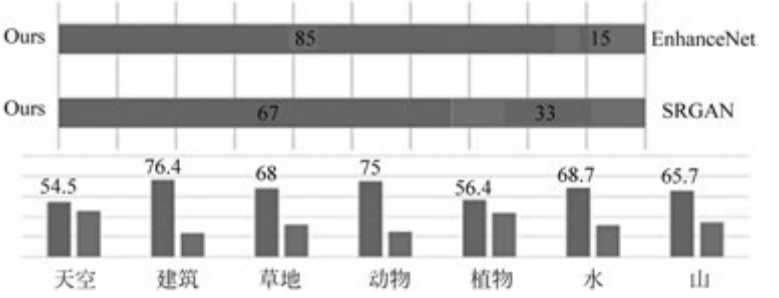

图 4-17 第二阶段结果

第一阶段的结果表明，SFT-GAN的性能大大优于面向PSNR的方法。这并不奇怪，因为面向PSNR的方法总是产生模糊的结果，尤其是在纹理区域。SFT-GAN有时会生成与HR相当的高质量图像，从而给用户造成混乱。在第二阶段中，SFT-GAN排名高于SRGAN和EnhanceNet，尤其是在建筑、动物和草地类别中。在天空和植物类别中可以找到类似的性能。

比较了结合其他先验条件的方式：①将图像和得到的语义分割概率图级联起来共同输入；②通过不同的分支处理不同的场景类别，然后利用语义分割概率图融合起来；③不考虑空间关系的特征调制方法 FiLM。（不同类别之间的纹理相互干扰）从图4-18中可以看到：方法①的结果没有SFT层有效(SFT-GAN模型中有多个SFT层能将先验条件更紧密地结合)；方法②的效率不够高(SFT-GAN只需要

进行一次前向运算)；方法③由于没有空间位置关系，不同类别之间的纹理会相互干扰。

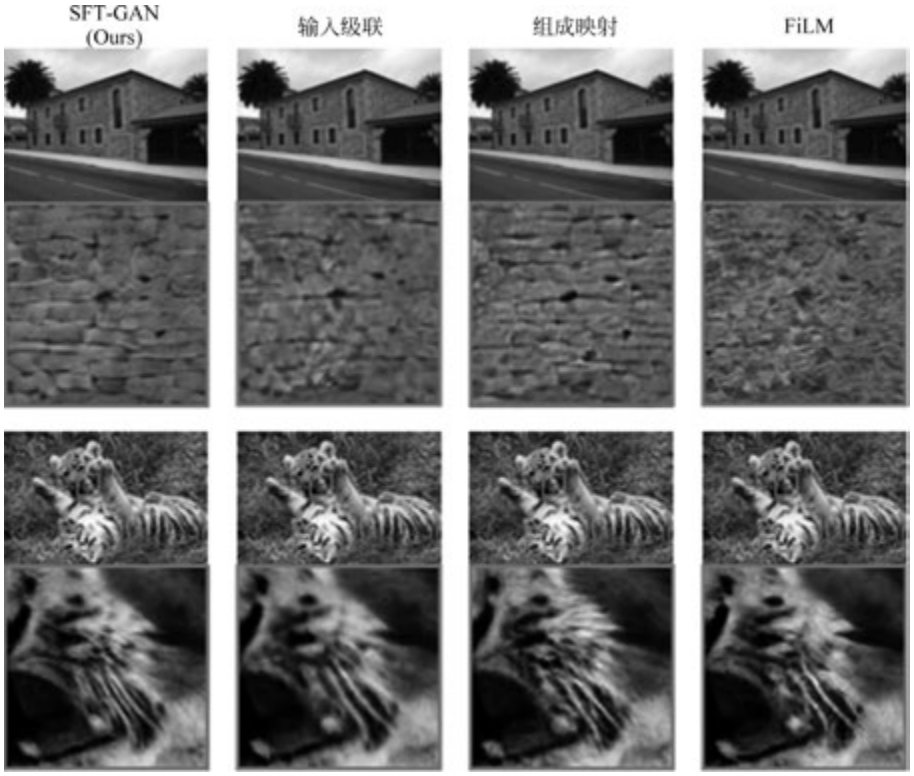

图 4-18　添加不同先验条件的结果

总的来说，SFT-GAN深入探讨了如何使用语义分割概率图作为语义先验约束超分辨率的解空间，使生成的图像纹理更符合真实且自然的纹理特性，还提出了一种新颖的空间特征调制层，有效地将先验条件结合到现有网络中。空间特征调制层可以与现有的超分辨率网络使用同样的损失函数，端到端地进行训练。测试时，整个网络可以接受任意大小尺寸的图像作为输入，只需要一次前向传播，就能输出结合语义类别先验的高分辨率图像。实验结果显示，相较现有超分辨率算法，SFT-GAN模型生成的图像具有更真实自然的纹理。

4.3　小结

本章主要介绍了有监督的图像超分辨率重建方法。首先对图像超分辨率重建问题进行了一般性建模，将问题抽象为最小化高分辨率估计图像和真值图像差异的过程，由于引入了真值高分辨率图像作为标签，因此该过程属于有监督的范式。接着介绍了在这种建模框架下的图像超分辨率重建方法，根据产生超分辨率图像

的出发点，可将这些方法分为判别式和生成式。

在判别式超分辨率重建方法中，首先介绍了残差学习。实际上不仅限于超分辨率重建领域，残差连接在许多深度学习模型中都有应用，其优点是在增加卷积层深度、提高网络表示能力的同时缓解梯度消失和爆炸问题。在卷积网络中，越深的层通常拥有越大的感受野，但是增加网络深度会引起模型参数量增加。因此有研究者提出基于循环学习的超分辨率重建方法，通过牺牲算力资源换取空间资源。其次介绍了逐渐提高学习任务难度以提升训练效果的课程学习策略。在超分辨率重建问题中，恢复按不同降质方式退化的高分辨率图像可以视为不同难度的任务，在训练过程中根据任务困难程度迭代学习恢复高分辨率图像可以有效提升模型性能。最后介绍了注意力和 Transformer 机制。受人类认知过程的启发，可以对卷积特征图的特定维度给予更高的权重，引导网络重点关注特征图的某一部分，从而获得更好的超分辨率重建效果。Transformer 是从自然语言处理迁移到计算机视觉的一种基于纯自注意力的方法，其优点是能提取图像的全局特征。相比传统的卷积网络，基于 Transformer 机制的超分辨率重建模型能取得明显提升。在生成式超分辨率模型中，主要介绍了生成对抗网络、带先验信息的生成对抗和扩散模型。生成对抗网络包含生成器和判别器，并在一般的超分辨率损失上添加了对抗损失项。生成对抗超分辨率结果通常具有更高的感知质量，但在像素级指标（峰值信噪比、结构相似度）上低于同规模的判别式模型结果。为了在超分辨率结果中重建更丰富的纹理细节，研究者提出将语义分割图作为先验信息输入生成对抗网络。这种先验生成对抗模型能生成具有真实自然纹理的超分辨率图像。

不难发现，本章介绍的有监督图像超分辨率重建方法有的是从网络结构出发提出的，有的是从数据特征出发提出的，还有的是受人类认知过程启发提出的，并都取得了一定提升。目前有监督超分辨率重建方法发展较为充分，未来的研究可针对具体应用场景进行适配性改进。

第5章

无监督的图像超分辨率重建方法

5.1 引言

图像超分辨率重建技术旨在通过从低分辨率图像中还原更高分辨率的细节，提高图像的视觉质量。在深度学习的发展历程中，无监督学习方法逐渐成为图像超分辨率领域的热点。本章将深入探讨无监督学习方法在图像超分辨率重建中的应用，并探讨其在实际场景中的优势。

与传统的监督学习方法不同，无监督学习方法在训练时不依赖大量的配对低分辨率和高分辨率图像。这种独特的特性使无监督学习方法更具灵活性和通用性，能够更好地适应各种应用场景。在实际项目中，获取大量配对数据可能面临各种挑战，如成本高昂、数据标注困难等问题，而无监督学习方法则能够在这些情况下发挥更大的作用。

本章将介绍几种基于深度学习的无监督图像超分辨率重建方法，涵盖零样本超分辨率（zero-shot super-resolution，ZSSR）、基于元迁移学习的零样本超分辨率（meta-transfer learning for zero-shot super-resolution，MZSR）、基于图像递归性的无监督超分辨率（image recursion super-resolution，IRSR）、基于约束重构的无监督超分辨率（unsupervised super-resolution generative adversarial network，UnSRGAN）等技术，并详细探讨这些方法的原理和优势。通过对比不同方法的性能，使读者更全面地了解无监督学习方法在图像超分辨率重建中的应用价值。

本章内容将围绕无监督学习的基本原理展开，深入探讨几种无监督图像超分辨率重建方法。通过研究旨在使读者建立对无监督学习在图像处理领域的深刻认识，为他们在研究中更有效地应用图像超分辨率重建技术提供指导。

5.2 方法介绍

5.2.1 问题建模

现有的超分辨率工作主要侧重监督学习，即使用匹配的LR-HR图像进行学

习。然而由于收集同一场景中分辨率各异的图像存在难度，SR数据集中的LR图像往往是通过对HR图像进行预设的降质处理而获取的。因此，经过训练的SR模型实际上是在学习如何逆转这种预设的降质过程。为了在不依赖人工降质先验知识的情况下学习真实的LR-HR映射关系，研究人员日益关注无监督SR领域。在这种情境下，训练过程仅依赖未配对的LR-HR图像，从而使模型更有可能应对实际场景中的SR问题。接下来将介绍几种现有的基于深度学习的无监督SR模型，除此之外仍有更多的方法等待我们发掘。

5.2.2 零样本超分辨率重建ZSSR

考虑到单个图像内的内部图像统计为SR提供了足够的信息，Shocher等提出了ZSSR[28]方法，通过在测试时训练特定于图像的SR网络应对无监督SR，而不是在大型外部数据集上训练通用模型。此外，他们利用了图像特定信息的跨尺度内部重现能力，而这使ZSSR不受基于面片的方法的限制。同时他们通过训练CNN从LR图像及其缩小版本（自我监督）推断复杂的图像特定HR-LR关系，并将这些学习到的关系应用于LR输入图像，以产生HR输出。这篇文章的突出贡献如下。

（1）这是第一个基于CNN的无监督SR方法。

（2）它可以处理非理想的成像条件，以及各种各样的图像和数据类型。

（3）它不需要预训练，并且可以使用少量的计算资源运行。

（4）它可以应用于任意大小的SR，并且理论上可以具有任意纵横比。

（5）测试时可以适应已知和未知的成像条件。

（6）2018年在“非理想”条件下图像上是最先进的SOTA（state-of-the-art）SR，在“理想”条件下训练的结果也可以与2018年基于监督的SOTA方法媲美。

针对无监督超分辨率任务，采用了一种特定的卷积神经网络，其融合了内部图像特定信息的预测能力、低熵特性和深度学习带来的泛化能力。在面临一个测试图像I时，由于外部训练样本的缺乏，ZSSR方法创造了一个专门针对该图像的CNN模型（见图5-1），以解决其特有的SR问题。为了训练这个CNN，利用测试图像I自身生成训练样本，即通过对I进行降采样创建其低分辨率（LR）版本$I\downarrow s$（其中s为期望的SR缩放因子）。接着采用了一个相对轻量级的CNN架构，并训练它使用$I\downarrow s$作为输入恢复原始测试图像I（见图5-1（b）上半部分）。训练完成后，将这个CNN应用于测试图像I，此时I作为网络的LR输入，通过该网络生成所需的高分辨率输出$I\uparrow s$（见图5-1（b）下半部分）。由于该CNN是完全卷积的，因此它能够适应并处理不同大小的图像。

对输入图像先进行各种比例的下采样（比例较小），标记为HR-fathers，再进行0°、90°、180°、270°旋转和垂直、水平翻转，再进行下采样，标记为LR-sons，构成LR-HR数据对。