

西北工业大学

学士学位论文



基于多尺度卷积神经网络的 单图光谱超分辨

作者姓名： 闫奕岐

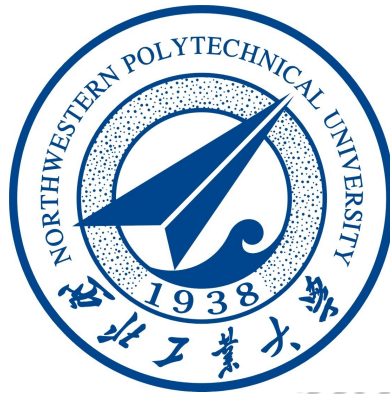
学科专业： 电子信息学院 通信工程

导师姓名： 魏巍 副教授（计算机学院）

完成时间： 二〇一八年五月三十一日

国奕岐-本科学位论文

Northwestern Polytechnical University
A dissertation for bachelor's degree



A multi-scale CNN for single image spectral super-resolution

Author: Yiqi Yan

Speciality: Communication Engineering

Supervisor: Prof. Wei Wei (School of Computer Science)

Finished time: May 31, 2018

国奕岐-本科学位论文

摘要

高光谱成像技术为很多计算机视觉问题提供了全新而高效的求解方案，但是由于硬件条件的限制，空间分辨率和光谱分辨率之间存在互相制约的关系，因此通常难以获取高（空间）分辨率的高光谱图像。另外，高光谱成像设备的成本较高，而且成像系统需要精细的配置。因此，超分辨算法在实践中有很重要的意义。

本论文主要关注的是光谱超分辨问题，主要任务是从单张 RGB 图像重建高光谱图像，也就是将三个离散的像素值 (r, g, b) 映射到一个高维连续的光谱。这本身是一个病态的逆问题，但是自然图像固有的内部相关性为此问题求解提供了很好的先验信息。在一幅图像中，对每一个像素点来说，都存在着很多局部和非局部的相似（相关）像素点，利用这些相关信息是求解该逆映射的关键。本文提出了一种多尺度卷积神经网络来进行从 RGB 到高光谱图像的映射。通过对称的上采样-下采样过程，我们的卷积神经网络能够同时编码局部和非局部的图像信息，并利用这些信息来进行很好的光谱重构。

我们在一个最新的高光谱数据集上进行了充分的实验，并用 5 种不同的评价指标和其他方法进行比较（绝对/相对均方根误差，光谱角等）。实验结果表明我们的模型在像素级和光谱的重构误差上都达到了当前最佳。另外，相对之前的模型而言，我们的模型更加稳定，对参数更加不敏感。

关键词：高光谱图像；光谱超分辨；多尺度卷积神经网络

Abstract

Hyperspectral imaging enhances the solution of many visual problems but suffers from low-resolution image data. Due to the trade-off between spectral and spatial resolution, it is hard to directly get high spectral-spatial resolution data. In addition, building a high-resolution hyperspectral imaging system can be really costly. Therefore, computational super-resolution methods mean a lot in practice.

This thesis focuses on one type of super-resolution method, spectral super-resolution. We aim to produce a high-resolution hyperspectral image from a signal RGB observation. Mapping three discrete intensity values to a continuous spectrum is highly under-constrained. Fortunately, the inherent correlation of natural images serves as a nice prior to help solve this problem. In fact, for each candidate pixel, there often exist locally and non-locally similar pixels. In this thesis, we propose a novel multi-scale convolutional neural network to explicitly map the input RGB image into a hyperspectral image. Through symmetrically downsampling and upsampling the intermediate feature maps in a cascading paradigm, the local and non-local image information can be jointly encoded for spectral representation, ultimately improving the spectral reconstruction accuracy.

We do experiments on a large hyperspectral database and prove that our method achieves state-of-the-art performance with regards to both pixel-level accuracy and spectral similarity. What's more, we experimentally show that our method is much more robust in that it is less sensitive to hyper-parameters compared to previous methods.

Key Words: Hyperspectral imaging, Spectral super-resolution, Multi-scale convolutional neural networks

致 谢

在研究学习期间，陕西省语音与图像信息处理重点实验室的魏巍副教授给予了细致的指导；两位师兄张磊和李勇在实验细节，论文撰写方面给予了很多帮助

感谢我的几位朋友给在我近一个月生活上遇到困境时给予的支持和鼓励：张玥（外国语学院），李明航（材料学院），刘沛宇（理学院），肖见睿（电子信息学院），也感谢在大学四年中给予我帮助的所有其他同学和朋友们。

另外，电子科技大学的张逸霄和居昊也在这几个月给予了我很多支持，向他们表示衷心的感谢。

最后，感谢家长和老师们的关怀。

特别致谢：偶像史蒂芬·库里（Stephen Curry）

在读期间研究成果与所获荣誉

已投稿论文

Accurate Spectral Super-resolution from Single RGB Image Using Multi-scale CNN, submitted to Chinese Conference on Pattern Recognition and Computer Vision (PRCV) 2018

所获荣誉

1. 2016-2017 学年 优秀学生标兵
2. 2016-2017 学年 国家奖学金
3. 2015-2016 学年 华为奖学金
4. 2016-2017 学年 校优秀学生

所获奖项

1. 2017 年 全国机器人锦标赛冠军（轮式机器人 1v1 足球）
2. 2016 年 国际水中机器人大赛一等奖（视觉挑战赛）
3. 2016 年 全国机器人锦标赛冠军（轮式机器人自主寻线导航）

目 录

中文摘要	I
英文摘要	II
在读期间研究成果与所获荣誉	IV
第一章 绪论	6
第一节 高光谱成像技术综述	6
第二节 现有超分辨率方法综述	9
一、空间超分辨率	9
二、光谱超分辨率	10
第三节 高光谱数据集	11
第四节 本文贡献	11
第二章 背景知识	13
第一节 插值算法	13
第二节 卷积神经网络	14
第三节 残差网 (ResNet) 和紧密连接网络 (DenseNet)	16
第四节 过拟合问题及应对方法	18
第五节 软件工具和硬件平台	18
第三章 模型结构	19
第一节 对比方法	19
一、基于稀疏表示的方法	19
二、基于深度学习的方法	21
第二节 多尺度卷积神经网络	22
一、组成模块	22
二、网络结构	24
三、讨论	24
第四章 实验	26
第一节 算法实现细节	26
第二节 评价指标	27

第三节 实验结果.....	28
一、收敛性能分析.....	28
二、数值评价指标.....	28
三、可视化分析.....	30
第四节 参数敏感度分析.....	31
第五章 结语.....	34
参考文献.....	35

国英歧-本科学位论文

插图清单

1.1 RGB 和高光谱成像对比图 (http://feilab.org/Research/Research_HSI.htm)	7
1.2 四种高光谱成像方式示意图	8
2.1 三种插值算法对比. 黑点: 插值点. 红/黄/绿/蓝点: 已知点 (观测点)	14
2.2 感知器和多层神经网络	15
2.3 卷积神经网络的两种基本模块	16
2.4 残差块和紧密连接块	17
3.1 两种稀疏表示方法的框架	19
3.2 全卷积紧密连接网络结构示意图	21
3.3 多尺度卷积神经网络结构图。“Conv m ”表示输出 m 个特征图的卷积层; 在绿色示意的模块中我们使用 3×3 卷积; 在红色示意的模块中我们使用 1×1 卷积; 灰色箭头代表特征图的堆叠	25
4.1 可见光光谱 (http://www.gamonline.com/catalog/colortheory/visible.php)	26
4.2 训练曲线和测试曲线	30
4.3 绝对误差的可视化。从左到右: RGB 图像, A+, Galliani <i>et al.</i> , 多尺度卷积网络	31
4.4 部分光谱重构效果可视化。第一排: RGB 图像; 第二排: 真实光谱 (实线) 和重构光谱 (虚线)。绘制光谱的四个位置在 RGB 图中用不同颜色的点标明。	32
4.5 Galliani <i>et al.</i> 和多尺度网络在有/没有 dropout 时的测试曲线 (只绘制了最后 50 轮)	33

表格清单

1.1 三个高光谱数据集的基本信息	11
3.1 全卷积紧密连接网络的组成模块	22
3.2 Galliani <i>et al</i> ^[1] 的完整结构信息	23
3.3 多尺度卷积神经网络的基本结构单元	24
4.1 两个深度卷积网络的超参数配置	27
4.2 测试集上各个评价指标的结果	29
4.3 Galliani <i>et al.</i> 和多尺度卷积网络有/没有 dropout 时的性能对比	32

西安交通大学 本科学位论文

符号说明

I_h	真实高光谱图像
I_{rgb}	RGB 图像
I_e	超分辨得到的近似高光谱图像
H	图像的高
W	图像的宽
C	高光谱图像的波段数
$p_h \in \mathbb{R}^C$	真实高光谱图像中的一个像素
$p_e \in \mathbb{R}^C$	近似高光谱图像中的一个像素
$p_l \in \mathbb{R}^3$	RGB 图像中的一个像素
$RMSE_1, RMSE_2$	均方根误差（有两种不同的计算方法）
$rRMSE_1, rRMSE_2$	相对均方根误差（有两种不同的计算方法）
SAM	光谱角
$D_H = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m\}$	含有 m 个元素的高光谱字典
$D_L = \{\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_m\}$	含有 m 个元素的 RGB 字典

第一章 绪论

高光谱图像将成像对象每一个像素点处的光谱响应量化编码为几十个甚至几百个波段，相邻波段之间的间隔十分紧密（例如只间隔 $10nm$ ）。和传统的图像不同，高光谱图像中的每个像素点包含连续的光谱响应，因此可以提供关于成像对象更加丰富的信息。例如，不同的物体或材料的反射光谱性质有差异，利用除了空间结构以外的附加（光谱）信息，很多计算机视觉问题可以更加高效和精确的求解，典型的应用包括：目标跟踪^[2]，图像语义分割^[3]，人脸识别^[4]，纸质文档分析^[5-6]，场景分类^[7-8]，异常检测^[9-10]，卫星遥感^[11-14]等。

然而，获取丰富的光谱信息是有一定代价的。当前有两个主要的问题限制高光谱图像的广泛应用。第一个问题是空间分辨率和光谱分辨率之间存在的制约关系。高光谱传感器成像的时候，由于每一个波段的范围较窄，每一个感光单元探测到的光子数目就会较少。在这种情况下，为了保持一个可以接受的信噪比（signal-to-noise ratio, SNR），传感器的瞬时视场角（Instantaneous Field Of View, IFOV）就需要增大^[15-16]。由于这种硬件上的限制，获取一个空间和光谱分辨率都很高的图像是很难的。第二个问题是高光谱成像设备成本比较高。由于高光谱传感器需要记录的是一个三维数据块，所以在长、宽、光谱三个维度上都需要进行信号采样操作，这大大增加了设备复杂度和配置难度。为了应对以上两方面的实际问题，研究者们提出了很多数值方法，也就是“超分辨”算法。

在绪论部分，我们首先介绍基本的高光谱成像原理，分析其优势与不足；然后综述当前已经被提出的超分辨方法，并且介绍三个常用的高光谱图像公开数据集。最后，我们将概叙本文的主要贡献。

第一节 高光谱成像技术综述

传统的成像传感器将光谱响应划分为几个相对较宽的波段，例如：普通 RGB 传感器在可见光范围内将光谱划分为 r, g, b 三段。与之相较，高光谱传感器能够同时在几十个甚至几百个连续的波段内成像。如图 1.1 所示。

高光谱传感器获取的图像是一个三维数据块，包括空间维度 (x, y) 和光谱维度 (λ)，这个数据块是从连续空间中进行信号采样得到的，这种采样操作则是通过“扫描”实现的。根据扫描的方式，可以将成像方法分成四种。他们的示意

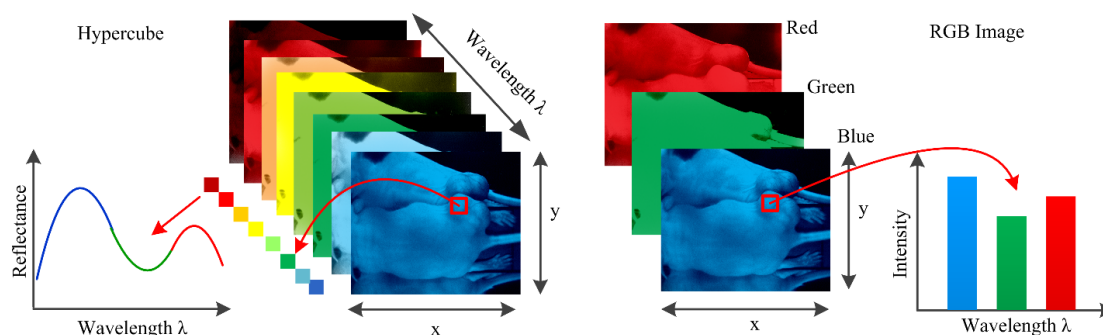


图 1.1 RGB 和高光谱成像对比图

(http://feilab.org/Research/Research_HSI.htm)

图如图 1.2 所示。

(1) 空间扫描 (Spatial scanning)

在空间扫描中，一个条形的“缝隙孔径” (slit aperture) 沿着 y 方向扫描，获取一些信号序列，然后组合成完整的高光谱图像。在每一个扫描位置，传感器获取一个二维的信号 (x, λ) 。从空间维度上看 (从 (x, y) 平面的角度观察)，每一个扫描位置是一个长条形区域，这个长条形区域扫过整个场景。因此，扫描的稳定性对于信号的“拼接”效果就至关重要 (例如对每一步扫描的步长的精确控制)。这种成像方式可以很好的获取较高的空间分辨率，但是逐步的扫描操作势必造成运动栅格效应 (motion artifacts)。

(2) 光谱扫描 (Spectral scanning)

光谱扫描也是逐步成像，每一步扫描获取一个二维信号，从这一点来说，它与空间扫描是一样的。它们的不同在于扫描操作作用于不同的维度。在光谱扫描中，每一步的二维输出是某一个波段的完整场景 (图像的某一个完整通道， (x, y))。这种扫描是通过可调的滤波器实现的 (每一步扫描通过滤波器选择相应的波段进行成像)。类似的，光谱扫描可以获得比较好的光谱分辨率，同时存在运动栅格效应 (motion artifacts)。

(3) 无扫描成像 (Non-scanning)

也叫做“快照成像” (snapshot imaging)。这种成像方式实际上是一次性把三个维度全部采样成像，不需要任何扫描操作。这种方式最大的好处就是成像十分迅速，因此系统吞吐量就可以很高。另外，因为没有扫描操作，运动栅格效应自然也不会存在。然而，由于需要一次性完整成像，因此传感器复杂度和成本都较高，成像分辨率也相对于扫描的方式低。

(4) 时空-光谱扫描 (Spatio-spectral scanning)

这种方式是空间扫描和光谱扫描的结合，每一次扫描都选取一部分的空间区域和一部分波段组合成像（可以用 $\lambda = \lambda(y)$ 来表示这种关系）。这种方式结合了空间扫描和光谱扫描的一些优势，同时也避免了它们的某些不足。

不管采用那种成像方式，空间分辨率和光谱分辨率往往无法兼顾。在实践中，我们可以很方便的获取高空间分辨率的 RGB 图像，但是它缺少光谱信息。反过来，获取丰富的光谱信息就要进行高光谱成像，需要牺牲空间分辨率。因此，使用数值方法“重构”光谱/空间维度（即“超分辨”）有很重要的实践意义。超分辨算法可大致分为两种：提升空间维度分辨率，提升光谱维度分辨率。接下来我们将介绍这两类方法。

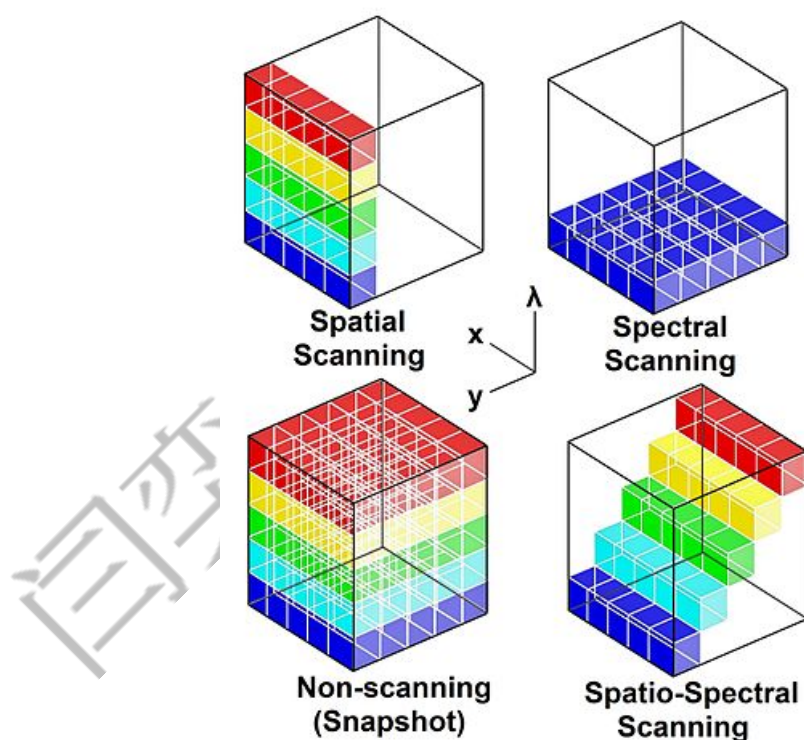


图 1.2 四种高光谱成像方式示意图

第二节 现有超分辨方法综述

一、空间超分辨

(1) 基于融合的超分辨方法

这类方法的主要思路是同时获取高空间分辨率的普通图像（例如：全色图像，RGB 图像）和高光谱分辨率的高光谱图像，通过数据融合算法获得空间和光谱分辨率都很高的图像^[17-18]。特别的，高光谱图像和全色图像（panchromatic image）的融合称为“pansharpening”，我们首先以此为代表介绍融合算法。现有的 pansharpening 方法可大致分为三类：构件替代（component substitution, CS），多分辨率分析（multiresolution analysis, MRA），贝叶斯方法（Bayesian）。构件替代 (CS) 的主要思路是用全色图去替代高光谱图像的一部分。典型的算法包括强度-色相-饱和度（intensity-hue-saturation^[19-21]），主成分分析（principal component analysis^[22-24]），以及施密特方法（Gram-Schmidt^[25]）。多分辨率分析 (MRA) 是一类将空间信息嵌入到高光谱图像中的方法，其中空间信息的提取依赖于对全色图的多分辨率分解。典型的多分辨率分解的方案有：抽样小波变换（decimated wavelet transform^[26]），非抽样小波变换（undecimated wavelet transform^[27]），拉普拉斯金字塔（Laplacian pyramid^[28]），以及不可分变换（nonseparable transforms^[29-30]）。贝叶斯方法（Bayesian）依赖已知的高光谱图像和全色图构建先验分布，并计算融合图像的后验概率分布，例如^[31-33]。以上所说的 pansharpening 算法可以比较好的拓展到高光谱图像和 RGB 图像的融合，最基本的思路可以是逐个波段使用 pansharpening 算法，然后将所有融合后的波段合成一个高分辨率的高光谱图像^[34-37]。

(2) 单图空间超分辨

基于融合的方法要求同时提供高光谱图像和普通图像，并要求它们严格配准，这在实践中是很难满足的。最新的一些方法致力于直接提高单张高光谱图像的分辨率，而不需要额外的输入数据。^[38] 使用卷积神经网络来编码图像的空间和光谱相关性，进而进行图像超分辨。^[39] 提出了一种三维全卷积网络结构，三维卷积可以更好的利用光谱结构信息，因为它考虑到了相邻波段的相关性。另外，^[39] 还针对传感器进行了建模，即使不使用目标场景的数据进行训练也可以获得很好的效果。在^[40-41] 中，一种光谱差异性卷积神经网络（spectral difference convolutional neural network）使用了一种空间约束的策略，可以在保持光谱结构信息的基础上矫正空间上的误差。^[42] 使用了残差网络的结构来进行超分辨，并

且将光谱角加入了损失函数（传统的损失函数只有均方误差）。

二、光谱超分辨

和空间超分辨相比，光谱超分辨的工作相对较少。它们大致分为以下三类。

(1) 早期成像方法

早期的很多方法直接使用 RGB 传感器来构建高光谱成像系统，它们大多工作在受控环境中。^[43] 通过使用光谱滤波器来利用主动照明（active lighting），这中条件只有在实验室环境中才可以满足。类似的，^[44-45] 也是要求在受控光照条件下捕捉原始 RGB 图像。在^[46] 中，作者提出对同一场景进行不同角度不同相机的 RGB 成像，然后利用这些 RGB 图像“组合”成高光谱图像，然而这套系统需要按照一定的规则很细致的配置和调试。

(2) 基于统计的方法

这类方法的核心将图像固有的某些统计特性作为图像先验，用以约束逆问题的求解。大部分基于统计的光谱超分辨方法都是基于字典学习和稀疏表示的，也就是首先从已有数据中构建一组字典，然后构造稀疏系数来作用于字典，字典元素在系数作用下的线性组合就是超分辨重构的结果。^[47] 是一个很典型的方法，它使用 K-SVD^[48] 来构建字典，在测试阶段使用正交匹配追踪（orthogonal matching pursuit, OMP^[49]）来计算得到稀疏系数。^[50] 基于 A+算法^[51-53] 进行光谱超分辨，在这个算法中，稀疏系数是通过求解一个带有稀疏约束的最小二乘问题来计算的。基于统计的方法往往是直接使用整幅图像抽取先验信息，而且所使用的先验往往是凭经验手动设计的，因此这类方法的泛化能力十分有限。

(3) 基于学习的方法

这一类研究致力于直接从数据中“学习”出从低分辨率到高分辨的映射关系。^[54] 利用径向基函数网络（radial basis function network）来学习这种映射，为了减弱光照差异性的影响，输入数据首先使用白平衡方程（white balancing function）进行预处理，这个预处理步骤的质量高低会影响到最终光谱超分辨的效果好坏。最近几年，由于深度学习方法在很多其他病态逆问题上的良好表现（例如图像去噪^[55]，普通图像超分辨^[56]），研究者很自然地考虑到使用深度神经网络（特别是卷积神经网络）来进行光谱超分辨。例如^[1] 就针对光谱超分辨任务改进了全卷积紧密连接网络（fully convolutional DenseNets, FC-DenseNets^[57]），不过这个网络对参数比较敏感，而且它的超分辨效果依然能够进一步改善。

第三节 高光谱数据集

足够数量和高质量的高光谱数据对于构建和测试超分辨算法十分重要。CAVE^[58]和 HARVARD^[59] 是两个常用的数据集，而 NTIRE2018^[60] 则是一个最新发布的数据集，它们的基本信息如表 1.1 所示。

(1) CAVE

CAVE 数据集包含 32 张空间尺寸为 512，具有 31 个波段的高光谱图像，覆盖的光谱范围是 400 到 700 nm 。图像内容主要是各种实物，例如人脸、水果等。

(2) HARVARD

HARVARD 数据集包含了 50 张图像，空间尺寸是 1024×1024 。图像的主要内容是各类自然场景，拍摄使用的传感器是 Nuance FX。

(3) NTIRE2018

这个数据集是从 ICVL 数据集^[47] 扩展而来的。ICVL 包含 203 张用 Specim PS Kappa DX4 相机获取的高光谱图像，每张图的空间尺寸是 1392×1300 ，覆盖了 400 ~ 1000 nm 范围内的 519 个波段。在实验中使用的是其中 31 个波段（在 400 ~ 700 nm 范围内以 10 nm 为步长提取 32 个波段）。在 NTIRE2018 中，53 张额外的高光谱被添加进来，构成的 256 张图像作为训练数据。除此之外，该数据集还提供了 5 张测试数据。NTIRE2018 针对所有的训练/测试数据都提供了相对应的 RGB 图像，由于其他任何数据集都不如 NTIRE2018 数据量大，数据的分辨率也不如 NTIRE2018 高，所以本论文中全部实验都在此数据集上面进行。

表 1.1 三个高光谱数据集的基本信息

	number of images	size	bands	spectral band
NTIRE2018	256 training + 5 test	1392×1300	31	400 ~ 700 nm
CAVE	32	512×512	31	400 ~ 700 nm
HARVARD	50	1024×1024	31	420 ~ 720 nm

第四节 本文贡献

在本文中，我们关注单图光谱超分辨问题，即从单张 RGB 图像中重构光谱信息。这个问题本身十分有挑战性，因为把三个离散的像素值 (r, g, b) 映射到一个高维连续的光谱是一个病态的逆问题（在下采样过程中很多光谱信息丢失

了)。为了解决这个问题，我们拟使用深度卷积神经网络来学习一个复杂的非线性映射。对于每一个待恢复的高光谱像素，都存在很多局部或非局部的相似像素点（也就是具有相似的光谱结构），这些相似像素点位置对应的颜色向量 (r, g, b) 可以看作是当前待恢复高光谱像素的一组多种相似类型的下采样。因此，进行光谱超分辨的时候，对 RGB 图像中的局部和非局部信息的合理利用就十分重要。处于这个考虑，我们设计了一种多尺度的卷积神经网络模型。通过几组对称的下采样-上采样操作，它可以同时编码局部和非局部的图像信息，用以提升光谱重建的准确度。通过大量实验，我们验证了我们的模型可以方便高效地进行端到端训练，并且在性能上超越了现有的方法。

我们的主要贡献可以总结如下：

- 我们设计了一种全新的多尺度卷积神经网络模型来进行光谱超分辨，多尺度结构使网络可以同时考虑到局部和非局部的信息。
- 我们在一个全新的数据集上进行了充分的实验，验证了我们的模型具有最佳的性能。

第二章 背景知识

这一章主要介绍与本论文相关的主要背景知识。第一部分是介绍插值算法的简单介绍，插值算法将作为本文的基础参考算法；其余部分主要是关于深度学习特别是卷积神经网络的介绍，包括：深度学习和卷积神经网络的基本概念，两种经典网络结构，深度学习中的过拟合问题和应对方法，以及深度学习的软硬件实现方案。

第一节 插值算法

插值算法是利用某函数一些已知的离散数据点来近似计算其他未知数据，在本文情境中，已知的数据点（r, g, b 像素值）是一个下采样的结果，因此插值算法就是一种“上采样”操作，插值的结果就可以作为重构的光谱信息。插值算法基于一种假设：观测值是整个函数的一个单纯的下采样（没有其他畸变），这一定程度上限制了插值算法的应用范围。由于本文的任务是进行光谱超分辨，因此这里我们只讨论一维信号的插值（图 2.1）。

(1) 最近邻插值 (Nearest neighbor interpolation)

这是思路最直接的一种插值方法：每一个插值点的直接赋予与其最近的一个观测点的值。这种插值信号是阶梯状的（参见图 2.1 左侧）。最近邻插值会造成较大的插值噪声和锯齿效应，不适合用于恢复丰富的光谱信息。

(2) 线性插值 (Linear interpolation)

已知两个观测点 (x_a, y_a) 和 (x_b, y_b) ，任意插值点处的值 (x, y) 可以用一下方式计算：

$$\frac{y - y_a}{x - x_a} = \frac{y_b - y_a}{x_b - x_a} \quad (2.1)$$

这个公式含义是： (x_a, y_a) 与 (x, y) 连线的斜率应当和 (x_a, y_a) 与 (x_b, y_b) 连线的斜率相同。换句话说，线性插值就是将插值点放置在相邻两个观测点的连线上（参见图 2.1 中间）。

(3) 样条插值 (Spline interpolation)

首先给出多项式插值的概念：给定一组观测点，多项式插值就是要找到一个多项式函数，使其能够通过全部观测点； n 次多项式插值指的是多项式最高次项

是 n 次的，例如 $n = 3$ 的时候就得到了三次插值（参见图 2.1 右侧）。样条插值是一种寻求这种多项式的一种特殊方法：相邻观测点之间用低次多项式来逼近，而不是使用高次多项式逼近整个函数，每个观测点处要保证是连续可导的。样条插值比传统的多项式插值具有很大的优势，因为样条插值使用较低次的多项式就可以获得较小的插值误差。

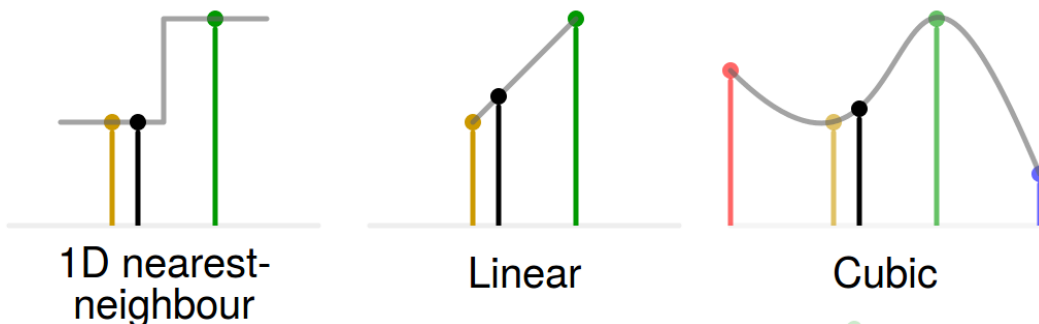


图 2.1 三种插值算法对比. 黑点：插值点. 红/黄/绿/蓝点：已知点（观测点）

第二节 卷积神经网络

在二十世纪五十年代，Frank Rosenblatt 受到生物神经元的启发，提出了一种数学模型：感知器（perceptron^[61]）。感知器算法后来被拓展为了多层神经网络（multi-layer neural networks）。一个人工神经元工作原理为：首先对输入进行一个线性变换，然后再施加一个非线性的激活函数（activation function）

$$y = \sigma(\mathbf{W}^T \mathbf{x} + b) \quad (2.2)$$

在最初的感知器模型中，只有一个神经元，激活函数是一个单位阶跃函数（神经元响应在达到一定阈值以后才会被激活），参见图 2.2（a）。在多层神经网络中（图 2.2（b）），每一层可以包含多个神经元，激活函数也有更加丰富的类型（例如 sigmoid 函数）。反向传播算法^[62]和梯度下降的出现使得神经网络的训练成为可能。

由于传统的神经网络是全连接的结构，它们不适用于图像数据。图像是多维数据并具有很丰富的局部相关性信息。一方面，使用传统神经网络会导致网络参数过多，难以优化；另一方面，传统神经网络无法利用图像的空间结构信息。^[63]提出了一种用来处理图像数据的全新网络结构：卷积神经网络（convolutional neural network, CNN），它与传统神经网络有以下两个主要的不同：

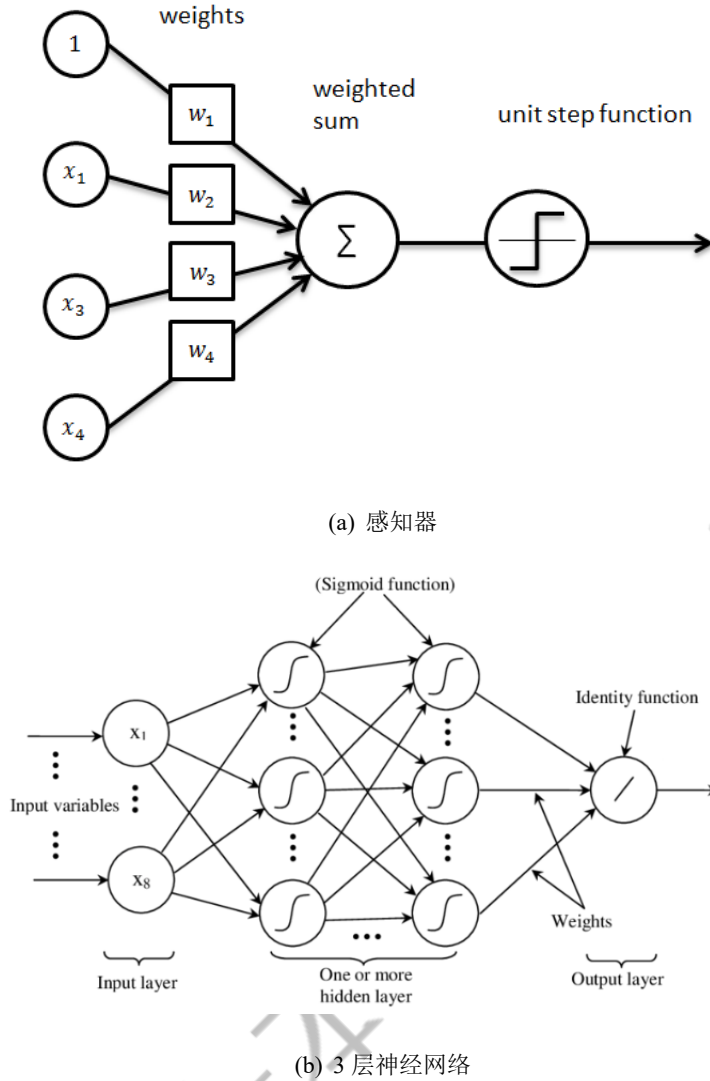
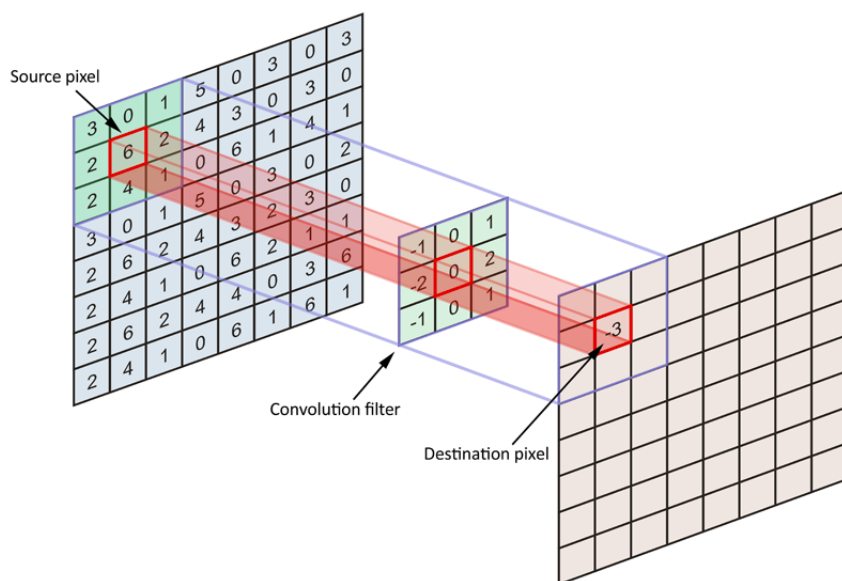
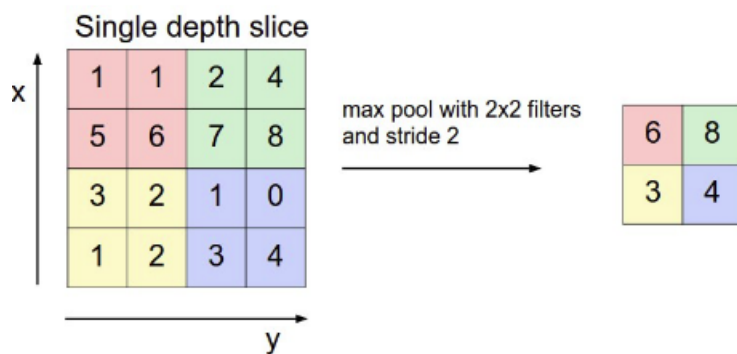


图 2.2 感知器和多层神经网络

- 在卷积神经网络中，网络权值是共享的。在网络的每一层，对输入施加的是卷积操作，卷积核（**kernels/filters**）与输入的局部区域相作用，也就是说输入的各个区域是贡献每一组参数的。参数共享的优势在于：（1）极大的缩减了参数量，使得训练深度网络成为可能；（2）有效利用了图像的空间结构信息。图 2.3（a）是一个标准卷积操作的示意图。
- 除了卷积操作，卷积神经网络的另一个典型结构就是池化（**pooling**）。池化的作用在于：（1）缩减隐层特征图的空间尺度，从而减小整体计算代价；（2）池化在一定程度上使网络具有旋转和平移不变性（**rotation invariance, shifting invariance**）。池化操作包括平均值池化（**average-pooling**）和最大值池化（**max-pooling**）两种，其中最大值池化是最常用的，它的示意图参见图 2.3（b）。



(a) 卷积



(b) 最大值池化

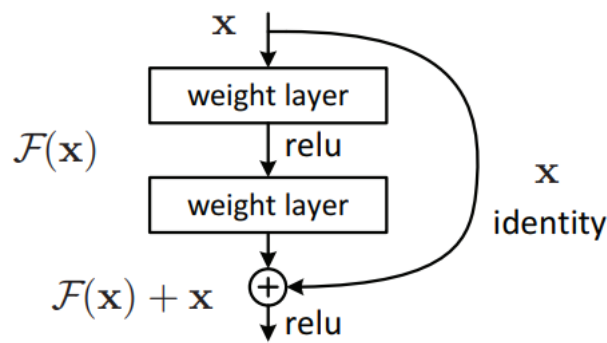
图 2.3 卷积神经网络的两种基本模块

过去几年间，有很多经典的卷积神经网络结构被提出，例如：LeNet^[63]，AlexNet^[64]，VGGNet^[65]，ResNet^[66-67]，DenseNet^[68]等。ResNet 和 DenseNet 与本文相关性较大，在下一小结我们将简要介绍这两种结构。

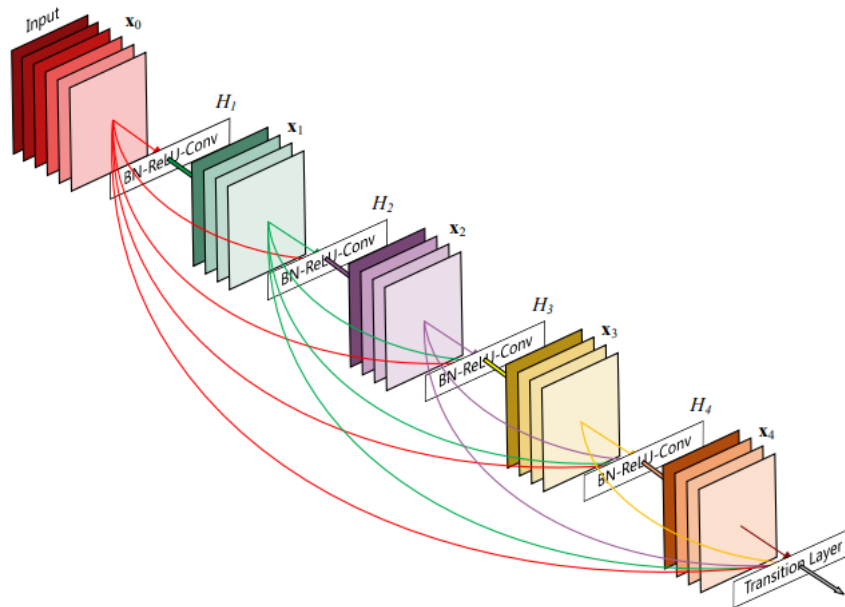
第三节 残差网（ResNet）和紧密连接网络（DenseNet）

深度残差网络（*deep residual network*, ResNet）是 2015 年 ImageNet 图像分类挑战赛的冠军。作者提出让网络去学习输入和输出之间的残差函数（*residual function*）而非直接去学习输入到输出的映射，学习残差函数是通过构造带有跨层连接（*skip connection*）残差模块（图 2.4（a））实现的。残差结构可以很好的抵抗梯度消失的现象，从而使训练上百层的深度网络成为可能。

紧密连接网络 (*densely connected convolutional network, DenseNet*) 使用了紧密的跨层连接 (残差网的跨层连接是“稀疏”的): 模块内部, 每个卷积的输入是之前全部卷积输出的组合 (图 2.4 (b))。另一个和残差网络的不同是: 紧密连接网络的跨层连接是将各个特征图堆叠起来 (残差网络是将它们相加)。紧密连接的优势在于: (1) 更好的对抗梯度消失现象; (2) 高效的特征重用机制; (3) 进一步缩减网络参数 (因为紧密连接模块靠着特征图的堆叠来扩展特征, 所以每一个卷积的自身的特征数目可以减少)。



(a) 残差模块 (ResNet 的基本单元)



(b) 紧密连接模块 (DenseNet 的基本单元)

图 2.4 残差块和紧密连接块

第四节 过拟合问题及应对方法

过拟合是广泛存在于各种深度学习任务中的一个问题，产生的主要原因是有限的训练数据不足以支撑大规模参数的训练，直接解决方案是收集更多数据，但是这往往不可行。除了扩展数据集之外，还有很多对抗过拟合的技术。

(1) 数据增广 (data augmentation)

对于低层计算机视觉问题(例如图像去噪, 图像超分辨), 用切割子图的方法进行数据增广是一种常用的数据增广手段。通过在原始图像中切割尺寸较小的子图, 在不增加图像数据的基础上就可以构造出几十倍甚至几百倍与图像数量的训练样本。这基本已经成为低层计算机视觉问题中一种默认的数据增广方式。

(2) 权值惩罚 (weight penalty)

又被称为正则化 (regularization)。通过在损失函数里加入网络参数的二次项, 可以抑制不合理的网络参数值。权值惩罚可以促使网络参数更加“均匀”, 是网络在输入数据上“均匀用力”而不是过于集中在某些部分。

(3) 随机抑制 (dropout)

过拟合主要发生在网络参数规模远大于训练数据规模的情况下。随机抑制^[69]就是在训练网络的过程中随机将部分神经元置零。这一定程度上等价于给网络引入了一定的训练噪声。

(4) 批归一化 (batch normalization)

批归一化^[70]的原理是将每一层的输出先进性归一化再传递到下一层去。深度网络训练的过程中, 各个层输出特征的均值和方差存在“漂移”现象, 批归一化就是抑制这种漂移, 使网络优化更加高效。

第五节 软件工具和硬件平台

在 GPU 加速和配套 CUDA 软件包的支持下, 大规模神经网络可以十分高效的进行训练。另外, 各大公司/研究机构开源其内部深度学习软件框架已经成为一个大趋势, 在这些框架的支持下, 构建和测试新的网络结构变得十分快捷。

当前最受欢迎的两大框架是 Tensorflow^[71] 和 PyTorch^[72], 分别由 Google 和 Facebook 开源与维护。本文的所有实验都是基于 PyTorch 框架, 因为它在构建动态计算图方面更有优势。我们的硬件平台是 8 块 GTX 1080 Ti。

第三章 模型结构

这一章里，我们介绍了本文涉及到的对比方法：^[47]中提出的稀疏表示方法 (Arad *et al.*)，^[50]中提出的另一种稀疏表示方法 (A+)，^[1]中提出的基于深度学习的方法 (Galliani *et al.*)。然后我们将介绍本文提出的多尺度卷积神经网络模型。

第一节 对比方法

一、基于稀疏表示的方法

Arad *et al.* 和 A+ 都是基于字典学习和稀疏表示的，算法框架如图 3.1 所示。

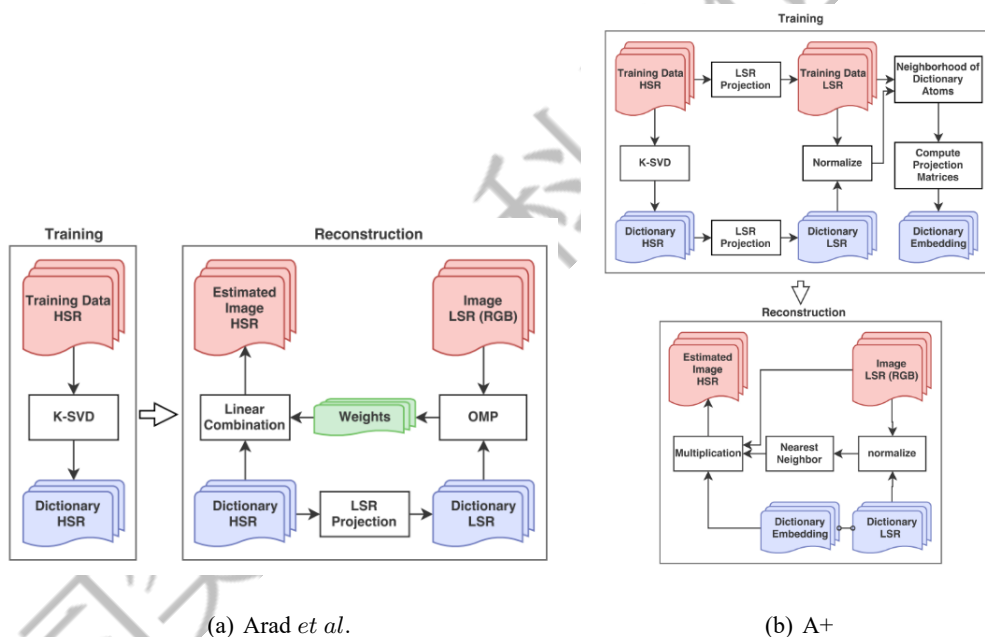


图 3.1 两种稀疏表示方法的框架

(1) Arad *et al.*

在训练阶段，Arad *et al.* 使用 K-SVD 算法^[48] 构建了含有 m 个元素的高光谱字典。

$$D_H = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m\} \quad (3.1)$$

由于从高光谱数据到 RGB 数据的隐射矩阵是已知的（这是本算法的一个假

设), 高光谱字典可以被投影到低维空间中, 得到降采样的字典。

$$D_L = \{l_1, l_2, \dots, l_m\} \quad (3.2)$$

在测试阶段, 对于一张输入 RGB 图像, 首先使用正交匹配追踪算法 (orthogonal matching pursuit, OMP^[49]) 将每一个 RGB 像素 $p_l = (r, g, b)$ 分解成低维字典的一个线性组合, 组合的系数 (w) 是稀疏的。

$$D_L \cdot w = p_l \quad (3.3)$$

得到了系数 w 之后, 相应的高光谱像素 p_h 就可以使用高光谱字典被重构出来 (使用和 RGB 图像相同的稀疏系数)。

$$p_h = D_H \cdot w \quad (3.4)$$

(2) A+算法

A+^[51-53] 最初是针对普通图像 (空间) 超分辨问题提出的, ^[50] 将其扩展到了光谱超分辨问题上, 但是保留了算法名称。在 A+ 中, 高光谱字典同样是通过 K-SVD 构建的, 也同样被投影到了低维空间。不过在 A+ 中, 高光谱数据本身也通过映射矩阵被投影到低维。接下来, 对于低维字典中的每一个元素 l_i , 稀疏系数通过如下方法估计:

$$\min_{\alpha} \|y_l - N_l \alpha\|_2^2 + \lambda \|\alpha\|_2^2 \quad (3.5)$$

其中, N_l 是 l_i 的最近邻元素集合, y_l 是投影到低维空间的图像数据。式 3.5 存在闭式解:

$$\alpha = (N_l^T N_l + \lambda I)^{-1} N_l^T \cdot y_l \quad (3.6)$$

由于高维和低维数据具有相关性 (是通过映射矩阵沟通的), 以下等式是近似成立的:

$$y_h = N_h \alpha \quad (3.7)$$

其中, N_h 是高维字典元素的最近邻集合, y_l 是高维数据 (高光谱图像)。这

时候，如果我们如下定义一个投影矩阵 P_i ：

$$P_i = N_h \cdot (N_l^T N_l + \lambda I)^{-1} N_l^T \quad (3.8)$$

那么，结合式 3.6~ 式 3.8，我们可以得到：

$$y_h = P_i \cdot y_l \quad (3.9)$$

也就是说，所构造的 P_i 就是从低维数据到高维数据的映射矩阵。在训练阶段，全部的投影矩阵 P_i 都基于训练数据被预先计算和存储，留作测试阶段使用。

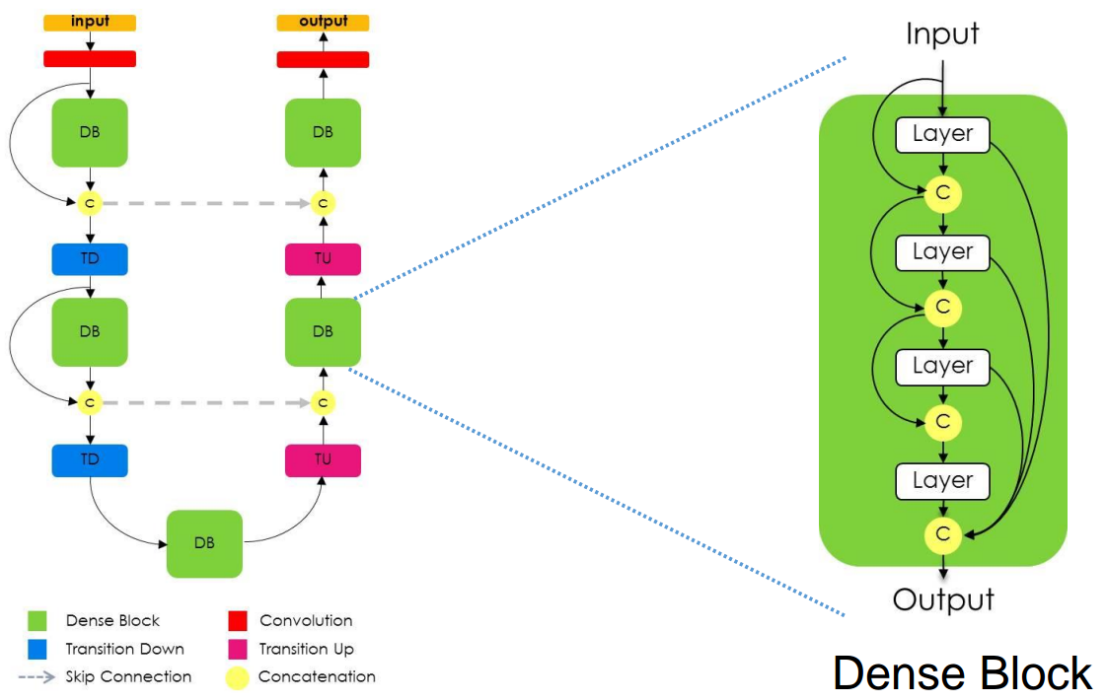


图 3.2 全卷积紧密连接网络结构示意图

二、基于深度学习的方法

Galliani *et al.*^[1] 使用了改进的全卷积紧密链接网络 (fully convolutional DenseNets, FC-DenseNets^[57])。后者最初是针对图像分割任务被提出的，它借鉴了紧密连接^[68]的思路。图 3.2 是全卷积紧密连接网络的示意图，表 3.2 是它的完整结构信息。这个网络中有三种基本组成模块 (表 3.1)。

(1) 紧密连接模块 (Dense block, DB)

在紧密连接模块内部，每一层输出 k 个特征图，并于本层的输入堆叠在一起作为下一层的输入。每一“层”依次包含：批归一化 (batch normalization)，漏

表 3.1 全卷积紧密连接网络的组成模块

(a)	(b)	(c)
紧密连接模块 (DB)	下采样模块 (TD)	上采样模块 (TU)
内部的一层结构	Batch normalization	Pixel shuffle
Batch normalization	Leaky ReLU	
Leaky ReLU	1×1 convolution	
3×3 convolution	Dropout	
Dropout	2×2 max-pooling	

隙 ReLU 激活函数 (leaky ReLU), 3×3 卷积, 随机抑制 (dropout) (参见表 3.1 (a))。紧密连接模块的输出是所有层输出的叠加。在^[1]中, 作者在每个基民连接块中设计了 4 层, 每层输出 16 个特征图, 因此每个模块输出 64 个特征图。

(2) 下采样模块 (Transition down block, TD)

下采样模块的作用是缩减特征图的空间尺寸, 在^[1]中, 作者不是只是用单一的池化操作, 而是在下采样模块中包含了批归一化 (batch normalization), 漏隙 ReLU 激活函数 (leaky ReLU), 随机抑制 (dropout) (参见表 3.1 (b))。

(3) 上采样模块 (Transition up block, TU)

^[1]使用了亚像素卷积 (sub-pixel convolution^[73]) 来进行上采样操作。原始的全卷积紧密连接网络^[57]使用的是转置卷积操作 (transposed convolution)。亚像素卷积的优势在于: (1) 不引入新的网络参数, 防止过拟合; (2) 防止出现棋盘格效应 (checkboard artifacts)。

第二节 多尺度卷积神经网络

在这一小节, 我们首先介绍网络的几个基本组成单元, 然后给出完整的网络结构。

一、组成模块

我们的网络有三种基本结构单元: 双重卷积 (double convolution, Double Conv, 表 3.3 (a)), 下采样 (downsample, 表 3.3 (b)), 上采样 (upsample, 表 3.3 (c))。

(1) 双重卷积 (Double Conv block)

包含两个 3×3 卷积操作，每一个层卷积都附加批归一化 (batch normalization)，漏隙 ReLU 激活函数 (leaky ReLU)，随机抑制 (dropout)。加入批归一化和随机抑制的目的是应对过拟合问题。

(2) 下采样模块 (Downsample block)

下采样模块使用了单个最大值池化 (max-pooling) 操作，它缩减特征图的空间尺寸，增大模型的感受视野。

(3) 上采样模块 (Upsample block)

和全卷积紧密连接网络类似，我们也采用了亚像素卷积来进行上采样，以达到缩减网络参数和防止棋盘格效应的目的。

表 3.2 Galliani *et al*^[1] 的完整结构信息

	Network components	Number of features
RGB	Input	3
Input	3×3 convolution	64
Downsampling Path	DB + TD	128
	DB + TD	192
	DB + TD	256
	DB + TD	320
	DB + TD	384
Bootleneck	DB	448
Uownsampling Path	TU+DB	400
	TU+DB	326
	TU+DB	272
	TU+DB	208
	TU+DB	144
Hyperspectral	3×3 convolution	31
Output	Output	31

表 3.3 多尺度卷积神经网络的基本结构单元

Double Conv	Downsample
3 × 3 convolution	2 × 2 max-pooling
Batch normalization	
Leaky ReLU	
Dropout	
Double Conv	Upsample
3 × 3 convolution	Pixel shuffle
Batch normalization	
Leaky ReLU	
Dropout	

二、网络结构

整体网络结构如图 3.3 所示，它由对称的两部分组成：编码部分（下采样通道）和解码部分（上采样通道）。在编码部分，每一个步骤都包含一个“双重卷积”模块（Double Conv）和一个下采样模块，特征图的空间尺度不断地缩小，特征图的数量每经过一步增加一倍。解码部分和编码部分是对称的，每一个步骤由一个上采样模块和一个“双重卷积”模块组成。特征图的空间尺度逐渐恢复，特征图数目没经过一个步骤就减半。最后，一个 1×1 卷积作用于输出特征图上面，将其映射到一个 31 波段的高光谱图像。在整个前馈结构之外，我们还引入了跨层连接来沟通编码部分和解码部分中相对应的特征图。

我们的模型可以很好的适应光谱超分辨的任务。首先，编码部分从 RGB 图像中逐步提取特征，在逐渐下采样的过程中，网络的感受视野逐渐增加，网络从获取局部信息逐渐扩展到非局部信息，这些信息统一被“编码”在逐渐增加的特征图中。解码部分利用这些深度特征逐步重构高光谱图像，而跨层连接则很好的引入了多尺度特征来辅助这个重构的过程，从而获得更好的重建精度。

三、讨论

U-Net^[74] 结构和本文提出的略有类似，我们的模型和 U-Net 的主要区别在于一下几个方面。

- 在我们的模型中，卷积操作使用了额外的零填充（zero padding）来保证卷

积操作不改变空间尺度。在 U-Net 中，由于卷积使得空间尺度缩减，在跨层连接部分需要将多余的边缘像素丢弃用以统一尺寸，这相当于直接丢弃了部分边缘特征。

- 我们的模型中增加了批归一化（batch normalization）和随机抑制（dropout）用以方式过拟合。
- 我们使用漏隙 ReLU（leaky ReLU）作为非线性激活函数，而 U-Net 采用的是 ReLU。
- 上采样部分，我们使用亚像素卷积而不是转置卷积。



图 3.3 多尺度卷积神经网络结构图。“Conv m ”表示输出 m 个特征图的卷积层；在绿色示意的模块中我们使用 3×3 卷积；在红色示意的模块中我们使用 1×1 卷积；灰色箭头代表特征图的堆叠

第四章 实验

第一节 算法实现细节

为了验证我们模型的有效性，我们和四种对比方法进行了对比：样条插值，Arad *et al.*^[47]，A+^[50]，Galliani *et al.*^[1]。^[47,50] 的作者开源了代码；^[1] 没有提供源码，所以我们自己实现了这个模型。下面我们简要介绍一下算法的实现细节。

(1) 样条插值

插值算法是本文中最基础的一个参照方法。对于每一个 RGB 像素 $p_l = (r, g, b)$ ，我们使用样条插值来进行上采样，获得 31 维的高光谱像素 (p_h)。根据可见光光谱的波长分布 (图 4.1)， r ， g ， b 三个值分别对应 $700nm$ ， $550nm$ ， $450nm$ 。

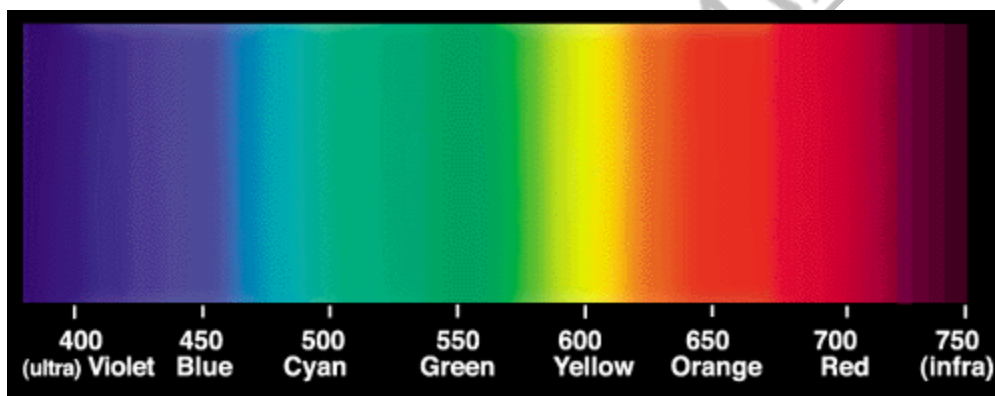


图 4.1 可见光光谱

(<http://www.gamonline.com/catalog/colortheory/visible.php>)

(2) Arad *et al.* 和 A+

低光谱分辨率图像被看作是高光谱图像的一个单纯的下采样，下采样是通过线性的映射矩阵完成的。在^[47,50]中，这个下采样矩阵要求是已知的。在我们的实验中，这个矩阵是使用全部训练数据用线性拟合的方式估计出来的。

(3) Galliani *et al.* 和多尺度卷积神经网络

我们通过实验确定每一个深度卷积网络的最佳超参数配置 (参见表 4.1)。对于 Galliani *et al.*，我们使用了 50% 的随机抑制概率 (dropout rate)，而多尺度网络使用了 20% 的随机抑制。全部的漏隙 ReLU 在负半轴的斜率都是 0.2。两个网络都使用 Adam 优化器训练 100 轮。两个网络的参数初始化方法和学习率也是不一样的：Galliani *et al.* 使用 HeUniform^[75] 来进行参数初始化，前 50 轮的学习

率是 2×10^{-3} ，后 40 轮是 2×10^{-3} 。对于我们提出的多尺度卷积模型，采用了 HeNormal^[75] 初始化方法，初始学习率是 5×10^{-5} ，然后每经过 10 轮学习率缩减为 0.93 倍。在数据增光部分，我们提取了 64×64 的子图，相邻子图有 40 个像素的重叠，最终我们获取了 267,000 个以上的训练数据。在测试阶段，我们直接将整幅图作为网络输入，通过一次前向传播重构出整幅图。

表 4.1 两个深度卷积网络的超参数配置

	Galliani <i>et al.</i>	Ours
Dropout rate	0.5	0.2
Slope for leaky ReLU	0.2	0.2
Initial learning rate	2×10^{-3}	5×10^{-5}
Weight penalty	1×10^{-6}	1×10^{-6}
Weight initialization	HeUniform	HeNormal

第二节 评价指标

我们使用了以下两个类别的评价指标来比较各种方法的优劣。

(1) 像素级重构误差

我们参考了^[50]中的评价指标，采用均方根误差（root mean square, RMSE）和相对均方根误差（relative root mean square, rRMSE）作为重构误差的评价指标。这两种指标各自有两种计算方式。

$$RMSE_1 = \frac{1}{n} \sum_{i=1}^n \sqrt{\left(I_h^{(i)} - I_e^{(i)}\right)^2} \quad (4.1)$$

$$RMSE_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(I_h^{(i)} - I_e^{(i)}\right)^2} \quad (4.2)$$

$$rRMSE_1 = \frac{1}{n} \sum_{i=1}^n \frac{\sqrt{\left(I_h^{(i)} - I_e^{(i)}\right)^2}}{I_h^{(i)}} \quad (4.3)$$

$$rRMSE_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{\left(I_h^{(i)} - I_e^{(i)}\right)^2}{\bar{I}_h^2}} \quad (4.4)$$

其中, $I_h^{(i)}$ 和 $I_e^{(i)}$ 分别表示真实的重构的高光谱图像中第 i 个元素 (标量值), \bar{I}_h 是 I_h 中所有元素的平均值, n 是一幅图中元素的总数目。

(2) 光谱相似度

由于光谱超分辨的目的是重构光谱信息, 我们还采用了光谱角 (spectral angle mapper, SAM) 作为一个评价指标。光谱角衡量的是真实的和重构的光谱夹角的平均值, 计算公式如下:

$$SAM = \frac{1}{m} \cos^{-1} \left(\sum_{j=1}^m \frac{(\mathbf{p}_h^{(j)})^T \cdot \mathbf{p}_e^{(j)}}{\|\mathbf{p}_h^{(j)}\|_2 \cdot \|\mathbf{p}_e^{(j)}\|_2} \right) \quad (4.5)$$

其中, $\mathbf{p}_h^{(j)}, \mathbf{p}_e^{(j)} \in \mathbb{R}^C$ 分别是真实的重构的高光谱图像中的第 j 个像素 (C 是总波段数), m 是一幅图像中的总像素数目。

第三节 实验结果

一、收敛性能分析

在图 4.2 中我们绘制了训练曲线 (训练集上的均方误差) 和测试曲线 (五个评价指标在测试集上的结果), 所有曲线都逐渐下降并最终收敛, 这说明我们的模型具有良好的收敛性能。

二、数值评价指标

表 4.2 显示了所有方法在五张测试图像上的数值评价结果。我们的方法在 $RMSE_1$ 和 $rRMSE_1$ 两个指标上超越了其他全部方法, 在 $RMSE_2$ 和 $rRMSE_2$ 上和 Galliani *et al.* 基本持平。更重要的是, 我们的方法在光谱角方面明显由于其他对比方法, 这说明我们的模型可以更好的重构光谱信息。需要指出的是, 像素级的重构误差和光谱相似性没有严格的正相关关系。例如, 如果我们简单的把像素打乱, $RMSE$ 和 $rRMSE$ 不会受到影响, 而光谱相似性会极大的变差。表 4.2 中的结果表示我们的方法在各个方面增强了光谱超分辨的性能。

表 4.2 测试集上各个评价指标的结果

$RMSE_1$						
	BGU_00257	BGU_00259	BGU_00261	BGU_00263	BGU_00265	Average
Interpolation	1.8622	1.7198	2.8419	1.3657	1.9376	1.9454
Arad <i>et al.</i>	1.7930	1.4700	1.6592	1.8987	1.2559	1.6154
A+	1.3054	1.3572	1.3659	1.4884	0.9769	1.2988
Galliani <i>et al.</i>	0.7330	0.7922	0.8606	0.5786	0.8276	0.7584
Our	0.6172	0.6865	0.9425	0.5049	0.8375	0.7177
$RMSE_2$						
	BGU_00257	BGU_00259	BGU_00261	BGU_00263	BGU_00265	Average
Interpolation	3.0774	2.9878	4.1453	2.0874	3.9522	3.2500
Arad <i>et al.</i>	3.4618	2.3534	2.6236	2.5750	2.0169	2.6061
A+	2.1911	1.9572	1.9364	2.0488	1.3344	1.8936
Galliani <i>et al.</i>	1.2381	1.2077	1.2577	0.8381	1.6810	1.2445
Ours	0.9768	1.3417	1.6035	0.7396	1.7879	1.2899
$rRMSE_1$						
	BGU_00257	BGU_00259	BGU_00261	BGU_00263	BGU_00265	Average
Interpolation	0.0658	0.0518	0.0732	0.0530	0.0612	0.0610
Arad <i>et al.</i>	0.0807	0.0627	0.0624	0.0662	0.0560	0.0656
A+	0.0580	0.0589	0.0612	0.0614	0.0457	0.0570
Galliani <i>et al.</i>	0.0261	0.0268	0.0254	0.0237	0.0289	0.0262
Ours	0.0235	0.0216	0.0230	0.0205	0.0278	0.0233
$rRMSE_2$						
	BGU_00257	BGU_00259	BGU_00261	BGU_00263	BGU_00265	Average
Interpolation	0.1058	0.0933	0.1103	0.0759	0.1338	0.1038
Arad <i>et al.</i>	0.1172	0.0809	0.0819	0.0685	0.0733	0.0844
A+	0.0580	0.0589	0.0612	0.0614	0.0457	0.0610
Galliani <i>et al.</i>	0.0453	0.0372	0.0331	0.0317	0.0562	0.0407
Ours	0.0357	0.0413	0.0422	0.0280	0.0598	0.0414
SAM (degree)						
	BGU_00257	BGU_00259	BGU_00261	BGU_00263	BGU_00265	Average
Interpolation	3.9620	3.0304	4.2962	3.1900	3.9281	3.6813
Arad <i>et al.</i>	4.2667	3.7279	3.4726	3.3912	3.3699	3.6457
A+	3.2952	3.5812	3.2952	3.0256	3.2952	3.2985
Galliani <i>et al.</i>	1.4725	1.5013	1.4802	1.4844	1.8229	1.5523
Ours	1.3305	1.2458	1.7197	1.1360	1.9046	1.4673

三、可视化分析

为了进一步说明光谱重构的准确性，我们可视化了真实的重构的高光谱图像之间的绝对误差，误差图是所有波段绝对误差的和（见图 4.3）。由于 A+ 在各个评价指标上面都绝对优于 *Arad et al.*，我们只是化了 A+ 来作为稀疏表示方法的代表。根据可视化结构可以看出我们的方法重构的图像更加平滑，误差更低。

另外，我们还随机选取了三张测试图像，并可视化了随机抽取的四个高光谱像素（图 4.4）。重构的光谱和真实光谱只有比较细微的差异。

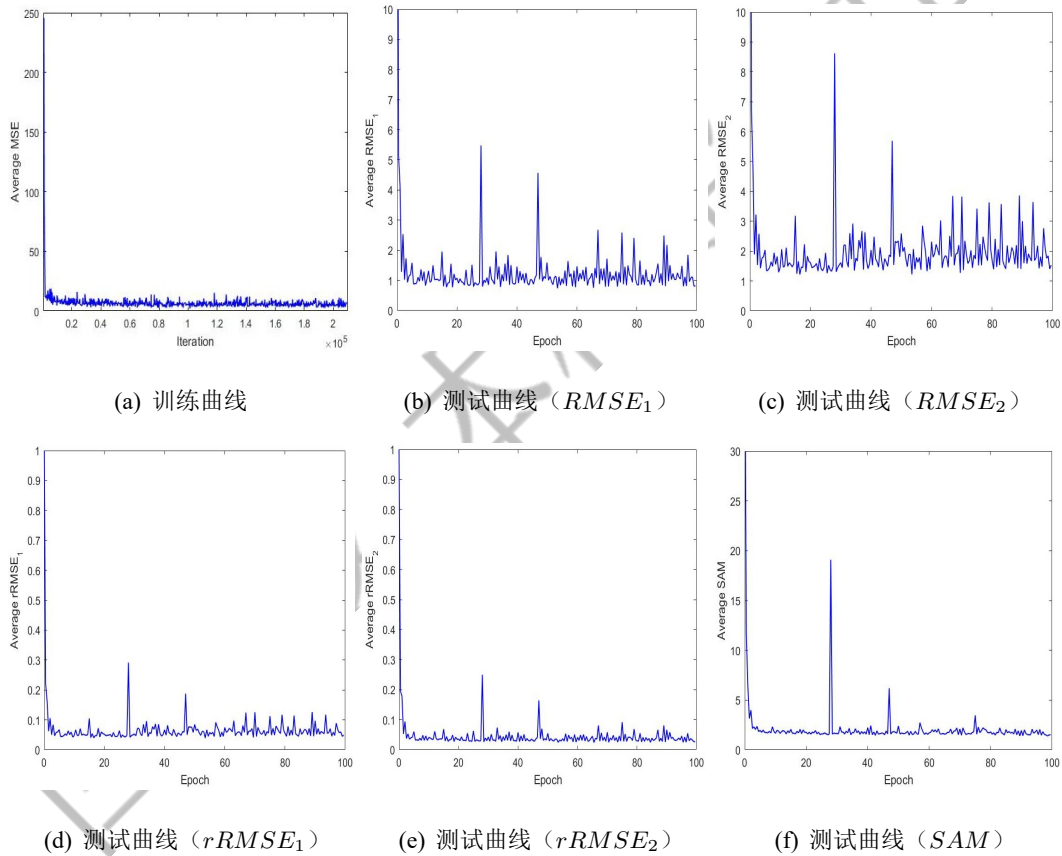


图 4.2 训练曲线和测试曲线

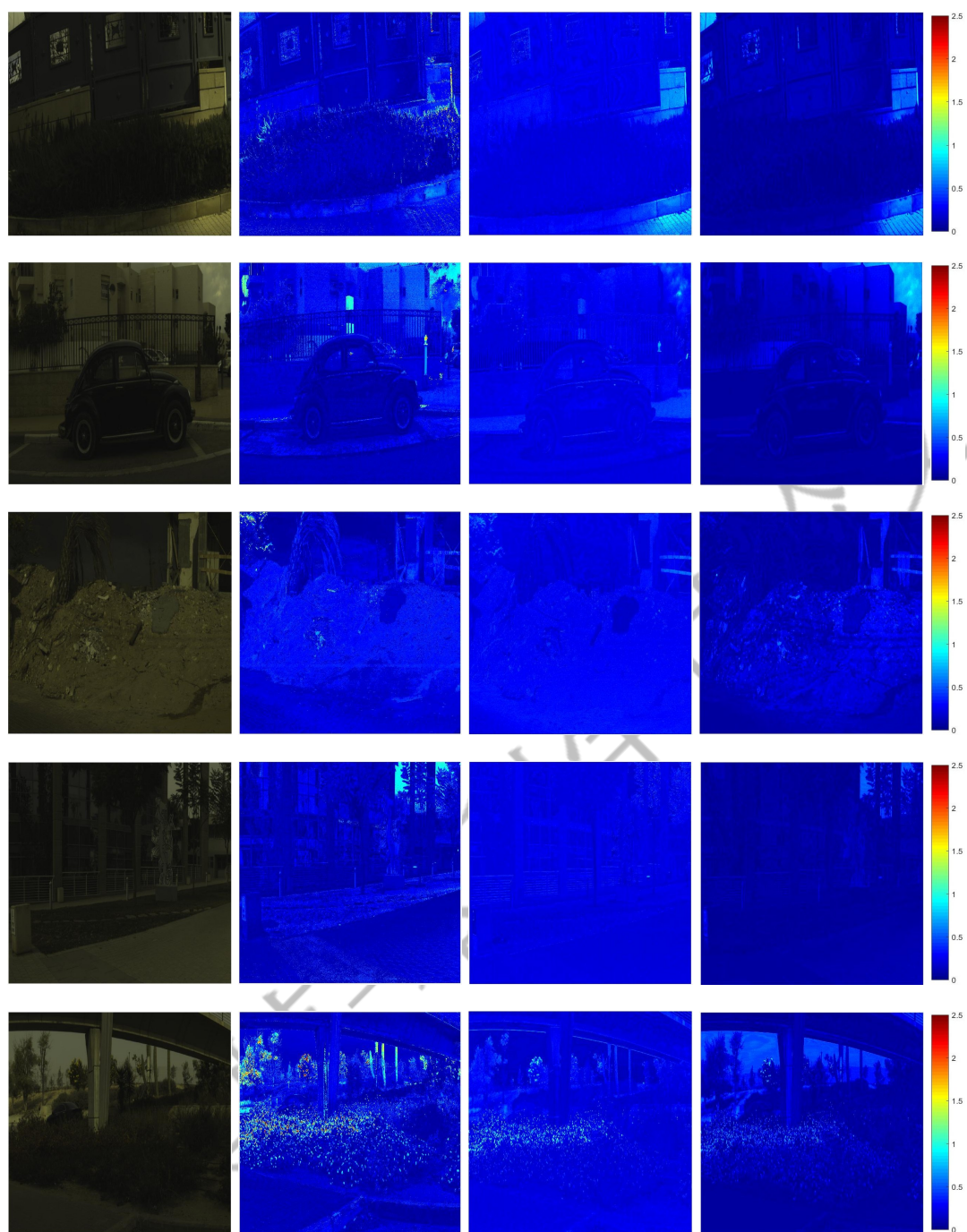


图 4.3 绝对误差的可视化。从左到右：RGB 图像，A+，Galliani *et al.*，多尺度卷积网络

第四节 参数敏感度分析

Galliani *et al.*^[1] 也采用了类似的对称网络结构，但是我们的模型更加稳定，对参数依赖性更低，为了说明这一点，我们丢弃了随机抑制（dropout）处理（也就是把抑制概率设置成 0），然后重新训练了两个网络。表 4.3 显示了重新训练的网络在测试集上的结果变化。虽然两个模型的性能都变差了，但是我们的模

型明显更加稳定。对于 Galliani *et al.*, 像素级的重构误差增加了至少 60%, 受影响最大的 $rRMSE_1$ 增大了 135.50% 之多。而我们的模型在各个指标上受到的影响都少于 50%。

表 4.3 Galliani *et al.* 和多尺度卷积网络有/没有 dropout 时的性能对比

	Galliani <i>et al.</i>	Galliani <i>et al.</i> (no dropout)	Increment (%)	Ours	Ours (no dropout)	Increment (%)
$RMSE_1$	0.7584	1.6092	112.18	0.7177	1.0662	48.56
$RMSE_2$	1.2445	2.0492	64.66	1.2899	1.8168	40.85
$rRMSE_1$	0.0262	0.0617	135.50	0.0233	0.0320	37.34
$rRMSE_2$	0.0407	0.0673	65.36	0.0414	0.0593	43.24
SAM	1.5523	2.1358	37.59	1.4673	1.6206	10.45

在图 4.5 中, 我们绘制了三个评价指标的测试曲线。当取消随机抑制之后, Galliani *et al.* 的各条测试曲线 (绿色虚线) 都有明显的“抬升”, 而我们的模型受到的影响就小得多 (各幅图中蓝色和品红色的实线都比较接近)。另外两个评价指标的测试曲线也呈现类似的趋势, 由于版面限制没有绘制。

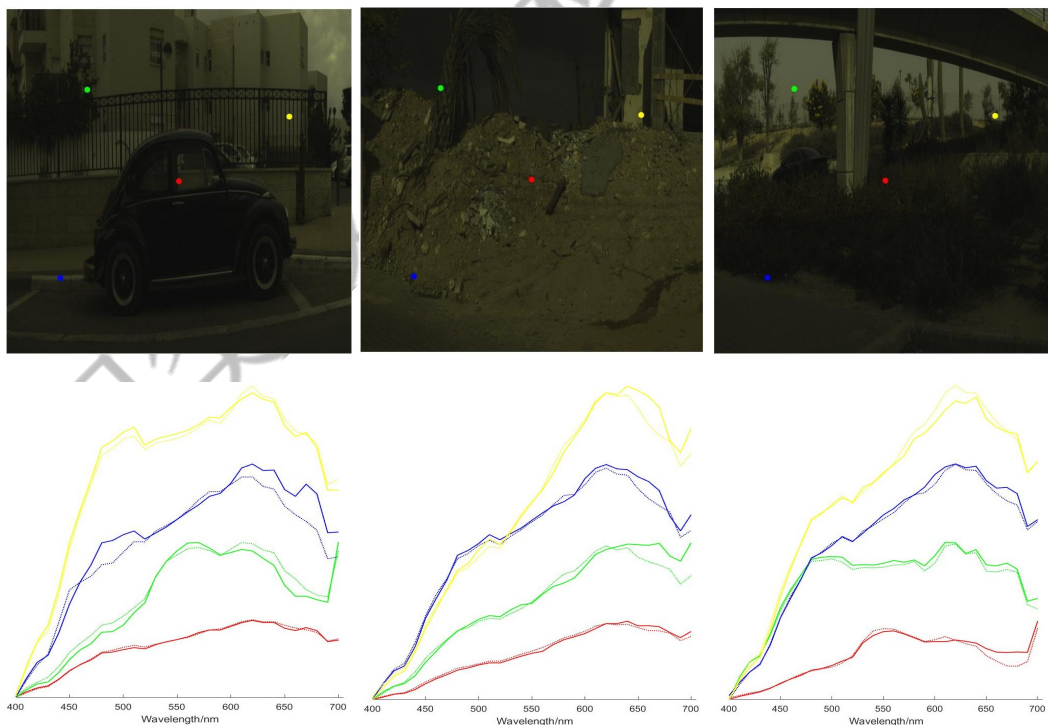
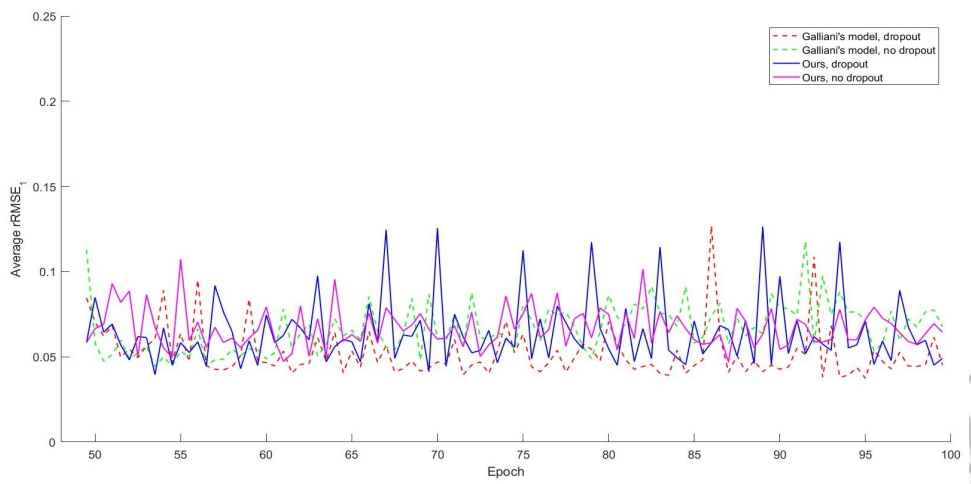
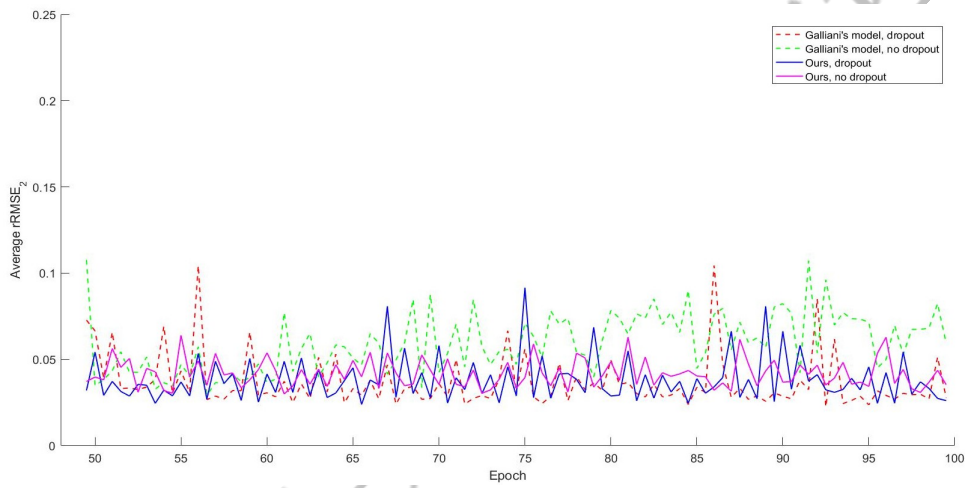
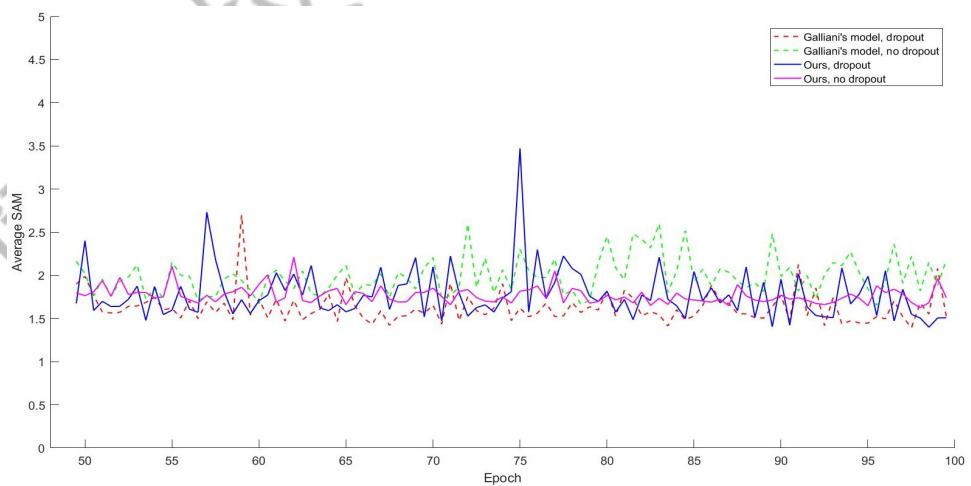


图 4.4 部分光谱重构效果可视化。第一排: RGB 图像; 第二排: 真实光谱 (实线) 和重构光谱 (虚线)。绘制光谱的四个位置在 RGB 图中用不同颜色的点标明。

 $rRMSE_1$  $rRMSE_2$  SAM 图 4.5 Galliani *et al.* 和多尺度网络在有/没有 dropout 时的测试曲线（只绘制了最后 50 轮）

第五章 结语

本文通过分析高光谱成像技术的优势和当前存在的挑战，提出了超分辨技术的现实意义，并着重于超分辨中的一类：光谱超分辨，也就是从 RGB 图像中恢复光谱信息。

我们说明了同时考虑局部和非局部信息对于光谱超分辨问题的重要性，基于这个，我们提出了一种多尺度的卷积神经网络模型。该模型包含了对称的下采样-上采样操作，能够同时编码局部和非局部的图像信息用以重构光谱结构。通过充分的实验，我们验证了该模型不仅具有更好的光谱重建精度，还具有更加稳定的性能。

参 考 文 献

- [1] Galliani S, Lanaras C, Marmanis D, et al. Learned spectral super-resolution. *CoRR*, 2017, abs/1703.09470. <http://arxiv.org/abs/1703.09470>.
- [2] Van Nguyen H, Banerjee A, Chellappa R. Tracking via object reflectance using a hyperspectral video camera//Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. IEEE, 2010: 44-51.
- [3] Tarabalka Y, Chanussot J, Benediktsson J A. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 2010, 43(7): 2367-2379.
- [4] Pan Z, Healey G, Prasad M, et al. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(12): 1552-1560.
- [5] Kim S J, Deng F, Brown M S. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognition*, 2011, 44(7): 1461-1469.
- [6] Padoan R, Steemers T A, Klein M, et al. Quantitative hyperspectral imaging of historical documents: technique and applications. *Art Proceedings*, 2008: 25-30.
- [7] Cheng G, Yang C, Yao X, et al. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns. *IEEE Transactions on Geoscience and Remote Sensing*, 2018.
- [8] Zou Q, Ni L, Zhang T, et al. Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(11): 2321-2325.
- [9] Kang X, Zhang X, Li S, et al. Hyperspectral anomaly detection with attribute and edge-preserving filters. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(10): 5600-5611.
- [10] Chang C I, Chiang S S. Anomaly detection and classification for hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 2002, 40(6): 1314-1325.
- [11] Haboudane D, Miller J R, Pattey E, et al. Hyperspectral vegetation indices and novel algorithms for predicting green lai of crop canopies: Modeling and valida-

- tion in the context of precision agriculture. *Remote sensing of environment*, 2004, 90(3): 337-352.
- [12] Hege E K, O'Connell D, Johnson W, et al. Hyperspectral imaging for astronomy and space surveillance//Imaging Spectrometry IX: volume 5159. International Society for Optics and Photonics, 2004: 380-392.
- [13] Belluco E, Camuffo M, Ferrari S, et al. Mapping salt-marsh vegetation by multi-spectral and hyperspectral remote sensing. *Remote sensing of environment*, 2006, 105(1): 54-67.
- [14] Borengasser M, Hungate W S, Watkins R. Hyperspectral remote sensing: principles and applications. CRC press, 2007.
- [15] Yilmaz O, Selimoglu O, Turk F, et al. Snr analysis of a spaceborne hyperspectral imager//Recent Advances in Space Technologies (RAST), 2013 6th International Conference on. IEEE, 2013: 601-606.
- [16] Laparrcr V, Santos-Rodriguez R. Spatial/spectral information trade-off in hyperspectral images//Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, 2015: 1124-1127.
- [17] Yokoya N, Grohnfeldt C, Chanussot J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 2017, 5(2): 29-56.
- [18] Loncan L, de Almeida L B, Bioucas-Dias J M, et al. Hyperspectral pansharpening: A review. *IEEE Geoscience and remote sensing magazine*, 2015, 3(3): 27-46.
- [19] CARPER W, LILLESAND T, KIEFER R. The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 1990, 56(4): 459-467.
- [20] Tu T M, Su S C, Shyu H C, et al. A new look at ihs-like image fusion methods. *Information fusion*, 2001, 2(3): 177-186.
- [21] Chavez P, Sides S C, Anderson J A, et al. Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic. *Photogrammetric Engineering and remote sensing*, 1991, 57(3): 295-303.
- [22] Kwarteng P, Chavez A. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens*, 1989, 55: 339-348.

- [23] Shettigara V K. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogram. Engineer. Remote Sen.*, 1992, 58: 561-567.
- [24] Shah V P, Younan N H, King R L. An efficient pan-sharpening method via a combined adaptive pca approach and contourlets. *IEEE transactions on geoscience and remote sensing*, 2008, 46(5): 1323-1335.
- [25] Laben C A, Brower B V. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. 2000.
- [26] Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 1989, 11(7): 674-693.
- [27] Nason G P, Silverman B W. The stationary wavelet transform and some statistical applications//Wavelets and statistics. Springer, 1995: 281-299.
- [28] Burt P J, Adelson E H. The laplacian pyramid as a compact image code//Readings in Computer Vision. Elsevier, 1987: 671-679.
- [29] Do M N, Vetterli M. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 2005, 14(12): 2091-2106.
- [30] Starck J L, Fadili J, Murtagh F. The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 2007, 16(2): 297-309.
- [31] Ballester C, Caselles V, Igual L, et al. A variational model for p+ xs image fusion. *International Journal of Computer Vision*, 2006, 69(1): 43-58.
- [32] Palsson F, Sveinsson J R, Ulfarsson M O. A new pansharpening algorithm based on total variation. *IEEE Geoscience and Remote Sensing Letters*, 2014, 11(1): 318-322.
- [33] He X, Condat L, Bioucas-Dias J M, et al. A new pansharpening method based on spatial and spectral sparsity priors. *IEEE Transactions on Image Processing*, 2014, 23(9): 4160-4174.
- [34] Grohnfeldt C, Zhu X X, Bamler R. Jointly sparse fusion of hyperspectral and multispectral imagery//Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International. IEEE, 2013: 4090-4093.
- [35] Grohnfeldt C, Zhu X X, Bamler R. The j-sparsefi-hm hyperspectral resolution en-

- hancement method—now fully automated//Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2014 6th Workshop on. IEEE, 2014: 1-4.
- [36] Grohnfeldt C, Zhu X X, Bamler R. Splitting the hyperspectral-multispectral image fusion problem into weighted pan-sharpening problems—the spectral grouping concept//Proceedings of 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing-WHISPERS 2015. IEEE Xplore, 2015: 1-4.
- [37] Selva M, Aiazzi B, Butera F, et al. Hyper-sharpening: A first approach on sim-ga data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 3008-3024.
- [38] Mei S, Yuan X, Ji J, et al. Hyperspectral image super-resolution via convolutional neural network//Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017: 4297-4301.
- [39] Mei S, Yuan X, Ji J, et al. Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, 2017, 9(11): 1139.
- [40] Li Y, Hu J, Zhao X, et al. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 2017, 266: 29-41.
- [41] Hu J, Li Y, Xie W. Hyperspectral image super-resolution by spectral difference learning and spatial error correction. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(10): 1825-1829.
- [42] Wang C, Liu Y, Bai X, et al. Deep residual convolutional neural network for hyperspectral image super-resolution//International Conference on Image and Graphics. Springer, 2017: 370-380.
- [43] Chi C, Yoo H, Ben-Ezra M. Multi-spectral imaging by optimized wide band illumination. *International Journal of Computer Vision*, 2010, 86(2-3): 140.
- [44] Han S, Sato I, Okabe T, et al. Fast spectral reflectance recovery using dlp projector//Asian Conference on Computer Vision. Springer, 2010: 323-335.
- [45] Park J I, Lee M H, Grossberg M D, et al. Multispectral imaging using multiplexed illumination//Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007: 1-8.
- [46] Oh S W, Brown M S, Pollefeys M, et al. Do it yourself hyperspectral imaging with

- everyday digital cameras.//CVPR. 2016: 2461-2469.
- [47] Arad B, Ben-Shahar O. Sparse recovery of hyperspectral signal from natural rgb images//European Conference on Computer Vision. Springer, 2016: 19-34.
- [48] Aharon M, Elad M, Bruckstein A. *rmk*-svd: An algorithm for designing over-complete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 2006, 54(11): 4311-4322.
- [49] Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition//Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on. IEEE, 1993: 40-44.
- [50] Aeschbacher J, Wu J, Timofte R, et al. In defense of shallow learned spectral reconstruction from rgb images//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 471-479.
- [51] Timofte R, De V, Van Gool L. Anchored neighborhood regression for fast example-based super-resolution//Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013: 1920-1927.
- [52] Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution//Asian Conference on Computer Vision. Springer, 2014: 111-126.
- [53] Timofte R, Rothe R, Van Gool L. Seven ways to improve example-based single image super resolution//Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. IEEE, 2016: 1865-1873.
- [54] Nguyen R M, Prasad D K, Brown M S. Training-based spectral reconstruction from a single rgb image//European Conference on Computer Vision. Springer, 2014: 186-201.
- [55] Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 2017, 26(7): 3142-3155.
- [56] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(2): 295-307.
- [57] Jégou S, Drozdal M, Vazquez D, et al. The one hundred layers tiramisu: Fully

- convolutional densenets for semantic segmentation//Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on. IEEE, 2017: 1175-1183.
- [58] Yasuma F, Mitsunaga T, Iso D, et al. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 2010, 19(9): 2241-2253.
- [59] Chakrabarti A, Zickler T. Statistics of real-world hyperspectral images//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 193-200.
- [60] NTIRE 2018 challenge on spectral reconstruction from rgb images// <http://www.vision.ee.ethz.ch/ntire18/>.
- [61] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 1958, 65(6): 386.
- [62] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *nature*, 1986, 323(6088): 533.
- [63] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [64] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks//Advances in neural information processing systems. 2012: 1097-1105.
- [65] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition//2014.
- [66] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [67] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks//European Conference on Computer Vision. Springer, 2016: 630-645.
- [68] Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks//Proceedings of the IEEE conference on computer vision and pattern recognition: volume 1. 2017: 3.
- [69] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, 2012, abs/1207.0580. <http://arxiv.org/abs/1207.0580>.

[//arxiv.org/abs/1207.0580](https://arxiv.org/abs/1207.0580).

- [70] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift//2015.
- [71] Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning.//OSDI: volume 16. 2016: 265-283.
- [72] Paszke A, Gross S, Chintala S, et al. Automatic differentiation in pytorch//2017.
- [73] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1874-1883.
- [74] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation//International Conference on Medical image computing and computer-assisted intervention. Springer, 2015: 234-241.
- [75] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.