信息科学技术专著丛书

基于深度学习的底层计算机视觉算法的研究与实现

董 谖 编著

RESEARCH AND IMPLEMENTATION OF
DEEP LEARNING BASED LOW-LEVEL COMPUTER VISION ALGORITHMS

北京邮电大学出版社
www.buptpress.com

信息科学技术专著丛书

基于深度学习的底层计算机视觉算法的研究与实现

董　譞　编著

北京邮电大学出版社
www. buptpress. com

内 容 简 介

本书是一本内容新颖、取材丰富的底层计算机视觉算法书籍，介绍了近年来基于深度学习的底层计算机视觉方向的热门问题，并给出了处理相关问题的具体算法和实验结果。

本书共分为11章，首先就目前底层计算机视觉算法的需求和发展对图像处理的重要性进行了介绍；然后分别对10个基于深度学习的底层计算机视觉算法问题进行了全面的讲解，包括相关问题的课题背景、研究现状、模型结构和算法结果等。全书以底层计算机视觉问题为主干、以深度学习算法为重点，附有大量实验数据和各算法效果图，清晰明了地展示了模型算法的处理流程和不同情况的处理效果。

图书在版编目(CIP)数据

基于深度学习的底层计算机视觉算法的研究与实现 / 董譞编著. -- 北京 : 北京邮电大学出版社，2022.5

ISBN 978-7-5635-6632-7

Ⅰ. ①基… Ⅱ. ①董… Ⅲ. ①计算机视觉—算法 Ⅳ. ①TP302.7

中国版本图书馆CIP数据核字(2022)第064110号

策划编辑：姚 顺 刘纳新 **责任编辑**：刘 颖 **封面设计**：七星博纳

出版发行：北京邮电大学出版社

社 址：北京市海淀区西土城路10号

邮政编码：100876

发 行 部：电话：010-62282185 传真：010-62283578

E-mail：publish@bupt.edu.cn

经 销：各地新华书店

印 刷：唐山玺诚印务有限公司

开 本：787 mm×1 092 mm 1/16

印 张：14.25

字 数：325千字

版 次：2022年5月第1版

印 次：2022年5月第1次印刷

ISBN 978-7-5635-6632-7 **定 价**：56.00元

·如有印装质量问题，请与北京邮电大学出版社发行部联系·

此为试读，需要完整PDF请访问：www.ertongbook.com

前　言

深度学习是近年来最为火热且发展最快的现代计算机技术之一，计算机视觉和自然语言处理是目前深度学习最为广泛研究和应用的两个方向。自深度学习方法提出以来，已经有许多有效的模型和算法被提出并应用于产品，但与此同时也有很多尚未解决或未达到理想效果的问题。

计算机视觉的目标是通过摄像机和计算机代替人眼对目标进行识别、跟踪和测量等。计算机视觉研究的相关理论和技术，意在建立能够从图像、视频或多维数据中获取或预测目标信息的人工智能系统。其经典问题包括图像识别、运动检测、场景重建和图像恢复。

本书旨在给那些想了解接触深度学习技术和计算机视觉算法的人提供研究思路和研究方法，特别是底层计算机视觉问题。因此，本书对 10 个课题提供了相应的基于深度学习的底层计算机视觉算法模型，均采用当下流行的 PyTorch 或 TensorFlow 框架实现。这些算法提供了相关图像处理问题的解决方法，同时展示了基本的深度学习技术和计算机视觉算法思想。另外，对于可以使用传统图像处理方法的问题，书中还将传统方法和基于深度学习的计算机视觉算法进行了对比，展示了两种方法各自的优缺点。

书中各课题所需的基本知识和研究背景均已在相关章节中做出了阐述，并在每个章节的最后列出了该章节的参考文献，因此不需要本书的读者具备过多的计算机视觉研究经验和图形学背景知识，但需要对机器学习的基本知识（如人工神经网络等）和 Python 等语言有一定的了解。

目　录

第 1 章　概述 …… 1

第 2 章　基于深度学习的图像滤波算法的研究与实现 …… 4

2.1　课题背景和任务 …… 4
2.2　相关技术介绍 …… 6
2.2.1　卷积神经网络 …… 6
2.2.2　引导滤波 …… 8
2.2.3　深度联合上采样 …… 9
2.2.4　结果评价指标介绍 …… 10
2.3　模型搭建 …… 11
2.3.1　模型结构分析 …… 11
2.3.2　利用 PyTorch 开源框架搭建网络模型 …… 14
2.4　数据收集与处理 …… 14
2.4.1　RGB/NIR 数据集 …… 14
2.4.2　RGB/Depth 数据集 …… 17
2.5　模型训练 …… 18
2.5.1　数据预处理 …… 18
2.5.2　深度联合滤波模型的训练 …… 19
2.6　实验结果 …… 20
2.6.1　RGB/NIR 降噪任务 …… 20
2.6.2　RGB/Depth 深度图上采样任务 …… 22
本章参考文献 …… 24

第 3 章　基于深度学习的图像修饰算法的研究与实现 …… 26

3.1　课题背景和国内外研究现状 …… 26
3.2　相关技术介绍 …… 29
3.2.1　基于深度学习的图像特征提取 …… 29
3.2.2　图像的颜色空间 …… 33
3.3　算法分析 …… 34

3.3.1 直方图分布 …… 34
3.3.2 回归问题 …… 37
3.4 算法与模型实现 …… 37
3.4.1 网络整体结构 …… 37
3.4.2 特征提取模块 …… 38
3.4.3 节点推断模块 …… 39
3.4.4 图像修饰模块 …… 39
3.4.5 数据集与预处理 …… 40
3.4.6 损失函数 …… 42
3.5 实验与分析 …… 42
3.5.1 评价指标 …… 42
3.5.2 模型验证 …… 44
3.5.3 对比实验 …… 46
3.5.4 消融实验 …… 48
3.5.5 拓展实验 …… 50
本章参考文献 …… 52

第 4 章 基于深度学习的图像超分辨算法的研究与实现 …… 55

4.1 课题背景和研究现状与发展趋势 …… 55
4.2 相关技术介绍 …… 57
4.2.1 图像处理技术 …… 57
4.2.2 超分辨 …… 58
4.2.3 深度学习 …… 59
4.3 基于 SRGAN 的单幅图像超分辨算法的设计与实现 …… 59
4.3.1 生成对抗网络 …… 59
4.3.2 感知损失 …… 63
4.3.3 算法流程描述 …… 64
4.3.4 实验分析与结论 …… 66
4.4 基于 SRCNN 的单幅图像超分辨算法的设计与实现 …… 69
4.4.1 卷积神经网络 …… 69
4.4.2 颜色空间 …… 73
4.4.3 算法流程描述 …… 74
4.4.4 实验分析与结论 …… 75
4.5 基于 VDSR 的单幅图像超分辨算法的设计与实现 …… 75
4.5.1 基于 SRCNN 的改进 …… 76
4.5.2 算法流程描述 …… 76
4.5.3 实验分析与结论 …… 77

本章参考文献 …… 79

第 5 章　基于深度学习的多幅图像超分辨算法的研究与实现 …… 81

5.1　课题背景与本章内容概述 …… 81
5.2　相关技术介绍 …… 85
5.2.1　分辨率 …… 85
5.2.2　图像降质退化模型 …… 86
5.2.3　插值 …… 86
5.2.4　重建图像的质量评估 …… 87
5.3　算法模块及网络结构设计 …… 88
5.3.1　动态上采样滤波器 …… 89
5.3.2　残差学习 …… 90
5.3.3　网络结构设计 …… 90
5.4　算法具体实现 …… 90
5.4.1　实验环境 …… 90
5.4.2　数据集 …… 91
5.4.3　训练 …… 91
5.5　实验结果分析及比较 …… 93
5.5.1　对比模型简介 …… 93
5.5.2　实验结果分析 …… 94
5.5.3　实验结果比较 …… 95
5.5.4　实验结果总结 …… 96
本章参考文献 …… 97

第 6 章　基于深度学习的图像立体匹配算法的研究与实现 …… 99

6.1　课题背景与任务 …… 99
6.2　立体匹配相关技术介绍 …… 100
6.3　模型的总体设计 …… 102
6.3.1　深度学习模型结构设计 …… 103
6.3.2　模型各模块搭建以及具体功能 …… 104
6.4　GC-NET 网络模型的详细构造和实现 …… 104
6.5　数据集获取和模型训练 …… 107
6.6　模型测试 …… 108
本章参考文献 …… 110

第 7 章　基于深度学习的光流算法的研究与实现 …… 113

7.1　课题背景与任务 …… 113

7.2 经典光流算法以及相关图像处理技术 …… 115
7.2.1 光流算法的主要发展历程 …… 115
7.2.2 早期经典方法及光流算法的基本假设 …… 116
7.2.3 用深度学习的方法求解光流 …… 118
7.2.4 其他主要技术 …… 121
7.2.5 图像处理的相关技术 …… 121
7.3 本章模型与端到端模型的区别 …… 122
7.3.1 端到端模型的一般结构 …… 122
7.3.2 使用特征金字塔的模型 …… 123
7.3.3 使用相似度金字塔的模型 …… 125
7.3.4 三种模型的对比 …… 127
7.4 实验与性能评测 …… 128
7.4.1 训练评测模型使用的数据集 …… 128
7.4.2 光流估计的误差定义 …… 130
7.4.3 本章模型在不同数据集上的实验结果 …… 131
7.4.4 实验结果详细情况 …… 131
本章参考文献 …… 132

第 8 章 基于深度学习的图像去噪算法的研究与实现 …… 134

8.1 课题背景与国内外研究现状 …… 134
8.2 相关知识介绍 …… 137
8.2.1 图像降噪概述 …… 137
8.2.2 噪声的来源 …… 137
8.2.3 图像噪声的种类 …… 137
8.2.4 图像噪声的模型 …… 138
8.2.5 图像降噪算法的分类 …… 138
8.2.6 质量评价指标 …… 139
8.3 传统图像降噪算法 …… 140
8.3.1 NLM 算法 …… 140
8.3.2 NLM 算法降噪结果 …… 141
8.3.3 BM3D 算法 …… 144
8.3.4 BM3D 算法的降噪结果 …… 146
8.4 基于深度学习的图像降噪算法 …… 149
8.4.1 DnCNN 简介 …… 149
8.4.2 残差学习和批量标准化 …… 149
8.4.3 DnCNN 网络模型结构 …… 151
8.4.4 DnCNN 模型训练 …… 152

8.4.5　DnCNN 算法的降噪结果 …… 152
8.5　图像降噪结果与分析 …… 153
8.5.1　数据预处理 …… 153
8.5.2　降噪结果 …… 153
8.5.3　对比与分析 …… 154
本章参考文献 …… 154

第 9 章　基于深度学习的遥感图像物体识别算法的研究与实现 …… 156

9.1　课题背景与国内外研究现状 …… 156
9.2　相关技术介绍 …… 159
9.2.1　R-CNN 家族算法 …… 159
9.2.2　YOLO 家族算法 …… 161
9.3　数据集选择与预处理 …… 162
9.3.1　数据集选择及数据集简介 …… 162
9.3.2　数据预处理 …… 163
9.4　Mask R-CNN 算法 …… 164
9.4.1　Mask R-CNN 网络结构 …… 164
9.4.2　模型实现与调优 …… 167
9.5　SOLO 算法 …… 168
9.5.1　SOLO 网络结构 …… 168
9.5.2　模型实现与调优 …… 171
9.6　算法实验结果与分析 …… 172
9.6.1　Mask R-CNN 实验结果与分析 …… 172
9.6.2　SOLO 模型实验结果与分析 …… 173
9.6.3　模型融合 …… 173
本章参考文献 …… 175

第 10 章　基于深度学习的多摄图像着色算法的研究与实现 …… 177

10.1　课题背景、国内外研究现状与本章主要内容概述 …… 177
10.2　相关理论基础 …… 180
10.2.1　图像着色理论 …… 180
10.2.2　色彩空间选择 …… 181
10.2.3　自监督学习 …… 181
10.2.4　循环卷积神经网络 …… 182
10.3　基于深度学习的双摄图像着色算法 …… 184
10.3.1　算法概述 …… 184
10.3.2　算法设计 …… 185

10.4 实验结果分析 …… 189
10.4.1 数据集与实验平台 …… 189
10.4.2 数据预处理 …… 191
10.4.3 实验流程 …… 192
10.4.4 算法测试流程 …… 192
10.4.5 实验结果分析 …… 193
本章参考文献 …… 194

第 11 章 基于深度学习的多摄图像高动态范围成像算法的研究与实现 …… 196

11.1 课题背景与任务 …… 196
11.2 相关技术介绍 …… 197
11.2.1 高动态范围成像 …… 197
11.2.2 色调映射 …… 197
11.2.3 多摄像头系统 …… 198
11.2.4 深度学习 …… 198
11.3 经典高动态范围图像生成算法的研究与实现 …… 199
11.3.1 经典算法调研 …… 199
11.3.2 HdrPlus 算法细节 …… 200
11.3.3 HdrPlus 算法实现 …… 204
11.4 基于深度学习的高动态范围图像生成算法的研究与实现 …… 205
11.4.1 深度学习方法在高动态范围成像领域的应用调研 …… 205
11.4.2 基于深度学习的高动态范围图像生成算法细节 …… 206
11.4.3 深度学习框架的选择 …… 209
11.4.4 基于深度学习的高动态范围图像生成算法的实现 …… 209
11.5 实验与结果分析 …… 210
11.5.1 评价指标 …… 210
11.5.2 HdrPlus 实验结果 …… 210
11.5.3 基于深度学习的高动态范围图像生成算法的实验 …… 212
11.5.4 结果对比及分析 …… 215
本章参考文献 …… 215

此为试读，需要完整PDF请访问：www.ertongbook.com

第1章

概　述

1. 背景

随着信息技术的发展，在图像与视频的拍摄、编辑、显示等方面，硬件设备以及软件算法不断迭代地提升，图像与视频的消费也与日俱增。在人们的日常生活中，YouTube、bilibili 等传统视频网站，以及 TikTok、快手等短视频应用，月活跃用户均已达到上亿人。除此之外，Twitch、斗鱼等直播网站，Facebook、微博等以图像、视频作为主要内容介质的社交网站也拥有大量的活跃用户。与此同时，智能手机、平板电脑、专业相机等拍摄设备也已经进入千家万户。因此，为了得到质量更高、视觉效果更好的图像和视频，广大的用户和众多厂商对于底层计算机视觉算法产生了巨大的需求，如图像去噪、图像超分辨、图像修饰、高动态范围成像等。

近年来，深度学习算法不断发展，涌现了大量高质量的图像、视频处理模型与方法。在数据训练方面，除最为传统的有监督学习和无监督学习外，还出现了半监督学习和自监督学习等深度学习算法。在网络模型方面，从 2012 年卷积神经网络模型 AlexNet 提出以来，陆续出现了一系列高质量卷积神经网络模型，如 VGG、UNet、ResNet。除卷积神经网络模型外，其他类型的深度学习模型也逐渐在底层计算机视觉领域取得了里程碑式的进展，如基于注意力机制的模型、Transformer 模型。可以说深度学习算法在底层计算机视觉领域已经有了长足的进步，并解决了许多具有很强实用性的问题，甚至在一些问题上深度学习方法处理的正确率已经超过了人工处理的正确率。

相比于传统图像、视频处理方法，基于深度学习的图像、视频处理方法具有不需要人工参与，在特定场景下适用性、可移植性强，处理后得到的视觉效果更符合人类视觉等优点。而且深度学习方法还可以解决很多传统方法无法解决的问题。但同时，深度学习也有计算量大、硬件成本高、模型设计复杂等缺点。

因此，围绕深度学习方法，本书重点研究了若干图像、视频处理问题，包括图像滤波、图像修饰、图像超分辨、图像立体匹配、光流、图像去噪、遥感图像物体识别、多摄图像着色、多摄图像高动态范围成像等问题。

2. 主要内容

接下来，此小节对本书将要讨论的基于深度学习的底层计算机视觉问题进行简要的介绍。

(1) 基于深度学习的图像滤波算法

含有引导图像的图像滤波称为联合滤波或引导滤波。此技术已成功应用于各种计算机视觉和计算机图形任务，如联合上升采样、图像增强等。引导滤波通过一张引导图对目标图像进行滤波处理，在引导图像的帮助下，使得最后的输出图像获得更好的质量。本书第 2 章构建了一个基于深度卷积神经网络的端到端联合滤波器。模型可选择性地传递同时存在于引导图像和目标图像中的空间结构，用于修正或增强目标图像中的结构纹理，同时避免原本不存在于目标图像中的纹理结构的负面影响。

(2) 基于深度学习的图像修饰算法

图像修饰可以提升图像的视觉质量，使得修饰后的图像恢复更复杂的细节与更丰富的色彩等。相比于传统的图像修饰集中于挖掘图像自身的信息，本书第 3 章提出了一种修饰效果多样、过程简单且高速的基于深度学习的图像修饰算法，可以从大量图像中挖掘共有信息。算法从图像的直方图出发，在某个颜色空间内，将图像修饰的过程近似为每个像素点发生若干不同的偏移得到修饰结果。算法的目标是每个像素节点的各自偏移量，从而将修饰问题转化为回归问题。

(3) 基于深度学习的图像超分辨算法

图像超分辨旨在对低图像分辨率图像进行处理，生成高分辨率、高精度的图像，此技术对于刑事侦查、军事领域、民用领域都具有重要意义。本书第 4 章研究并实现了三种基于深度学习的超分辨算法。一种是基于生成对抗神经网络的单幅图像超分辨算法，注重整体视觉效果的提升；另外两种是基于卷积神经网络的超分辨算法，注重局部细节的增强。

(4) 基于深度学习的多幅图像超分辨算法

本书第 5 章研究并复现了一种多幅图像的超分辨算法。多幅图像的超分辨算法大多依赖于精确的、显式的运动补偿，而该算法创新性地设计出动态学习上采样滤波器的网络结构，避免了显式的运动补偿，并利用了深度神经网络感知不同图像间的运动估计完成多幅图像的超分辨任务。

(5) 基于深度学习的图像立体匹配算法

立体匹配能够从双目图像中不同视图之间的像素点间的视差恢复图像中的场景深度，在自动驾驶、机器人避障、立体图像智能编辑等领域具有重要价值。本书第 6 章提出了一种可以适用于深度计算的端到端立体匹配的深度学习模型。此模型通过卷积神经网络提取图像特征，构建两张视图的代价匹配，并构建深度图。

(6) 基于深度学习的光流算法

光流算法的主要任务是通过视频中相邻几帧的数据去求解图像中物体的运动信息，因此经常被应用于自动驾驶、运动检测、行为识别等方面。光流是二维矢量场，包含各像

素点的瞬时运动速度矢量信息，通过给定图像序列，可以求其所含运动的运动场。本书第7章构建了两种基于深度学习与传统方法结合的光流算法模型。

(7) 基于深度学习的图像去噪算法

图像噪声主要来源于图像获取以及图像传输的过程，想获得完全不含噪声的图像难度是很高的，因此图像降噪具有巨大的需求。本书第8章研究了图像噪声的种类以及图像噪声的模型，在此基础上，实现了两种基于传统方法的降噪算法和一种基于深度学习的图像降噪算法，并将传统方法和深度学习算法的降噪结果进行对比，以验证深度学习算法的性能。

(8) 基于深度学习的遥感图像物体识别算法

遥感是与物体不直接进行物理接触的情况下获取有关物体的信息，通常指使用搭载在卫星或者飞机上的传感器接受的电磁波等信号。对遥感图像的处理，一般需要航空影像的专家进行。本书第9章设计了一种基于深度学习的在遥感图像上进行目标检测的算法，完成对遥感图像的实例分割，以便在一定程度上替代人工对遥感图像进行处理。

(9) 基于深度学习的多摄图像着色算法

黑白图像着色的传统方法主要有局部区域色彩扩展和基于参考图的颜色传递，这两种方法均需要人工参与和专业的图像处理。本书第10章以黑白-彩色双摄摄像系统为例，实现了基于深度学习的多摄图像着色算法。利用黑白图像信噪比高、成像质量高的特点，以彩色图像为参考，对黑图像进行着色，生成更高质量的彩色图像。

(10) 基于深度学习的多摄图像高动态范围成像算法

动态范围是指可变化信号的最大值和最小值的比值。日常生活中经常会出现拍摄对象最亮与最暗部分相差很大的情况，即高动态范围场景。而相机拍出的照片可能会出现曝光过度或曝光不足。高动态范围成像技术通过合成多张照片，以扩大动态范围获得质量更好的照片。本书第11章实现了一种基于深度学习的算法，通过捕获多张不同曝光的低动态范围图像，合成一张高动态范围图像，从而提升成像效果。

第2章

基于深度学习的图像滤波算法的研究与实现

2.1 课题背景和任务

图像滤波作为计算机视觉领域最为基础和广泛的任务之一，作为图像预处理中不可缺少的操作，其处理结果的优劣将对后续图像任务的处理和分析产生重要的影响。本节主要介绍基于深度学习的图像滤波算法的课题背景和课题任务。

1. 课题背景

传统图像滤波可大致分为线性滤波和非线性滤波。线性滤波包括著名的高斯滤波、均值滤波；非线性滤波包括引导滤波[2]、中值滤波、双边滤波[4]。

传统滤波方式，如高斯滤波、均值滤波、中值滤波、双边滤波，都是对图像本身进行处理，通过设定特定规模的滤波核，对滤波核内的像素点进行不同的处理达到不同的滤波效果。例如，高斯滤波处理去高斯噪声效果显著，而中值滤波则擅长处理脉冲滤波，并且不会模糊边缘。传统滤波常用于解决图像特定的噪声问题，适用范围比较单一，而引导滤波通过引入一张引导图对目标图像进行处理，可以处理一些超越平滑概念的图像任务，包括但不限于抠图、去雾等，且输入输出形式上跟本章的模型类似。2.3 节“相关技术介绍”部分将详细介绍引导滤波。

近几年随着硬件技术的发展、数据的爆炸式增长、并行计算框架的发布和成熟（如TensorFlow、PyTorch），深度学习逐渐应用于各种传统领域并在很多方向上（如图像增强[16,17]）取得显著的进步。图像滤波任务也逐渐出现基于深度学习和神经网络的方法。本章课题的主要任务是将传统的引导滤波（又称为联合滤波）与深度学习和神经网络的技术相结合，即深度联合滤波。通过搭建联合滤波器对应的神经网络模型，将一些开源图像数据集处理成联合滤波器所需的样本集（即每个样本包含的目标图像、引导图像、监督图像）作为输入引入模型用于模型训练。值得一提的是，该模型是一个端到端的通用模型，通过导入不同任务的数据集用于训练，可以得到可用于不同图像处理任务的模型。

目前应用该模型处理的图像任务包括两个，分别是 RGB/NIR 图像降噪任务和深度图上采样任务。

联合滤波器的基本思想是将引导图像通常包含的重要的结构细节传递给目标图像，用于增强由于噪声或低空间分辨率而退化的目标图像。

联合图像滤波器可以被分成两类。

一类是综合考虑引导图像给输出加上一些约束，并以此构建一个二次函数，需要做的就是优化该二次函数。然后通过求解一个用引导图像信息编码的大型稀疏矩阵来获得解。该方法已广泛应用于各种图像任务，如着色、图像去雾、多尺度分解等。这种基于优化的方法通常能得到优秀的结果，但它的代价就是需要很长的计算时间。

另一类是利用引导图像显示地构建滤波核[15]——可能会包括引导滤波器[3]和双边滤波器[4,19]。双边滤波器的输出图像在一个像素处的输出是附近像素的加权平均值，其中的权重取决于引导图像中的相似度，并且引导图像可以是原图像自身或者其他图像。联合双边滤波器可以平滑小的波动并且保留边缘。引导滤波跟双边滤波相似，但它的算法复杂度更低，引导滤波器对灰度图和 RGB 图像都有 $O(N)$（N 个像素点）的时间复杂度。

以上两类滤波器有一个共同的缺陷，就是这些滤波器只考虑引导图像中的结构信息，而当引导图像中存在一些和目标图像不一致的局部结构时，这些滤波器可能会将不一致的局部结构传递给目标图像，对滤波器输出造成负面影响。

本章所参考的网络模型是一种基于卷积神经网络的学习型联合滤波器，不同于现有的方法，该模型是完全数据驱动的，能同时学习所有网络参数，无须分段训练。最重要的一点是该网络同时考虑了目标图像和引导图像，自然地处理了引导图像和目标图像结构不一致的问题。

2. 课题任务

（1）课题内容

收集阅读相关文献，了解联合滤波概念、背景以及当前研究进展，总结前人工作，分析当前滤波算法的缺陷。参考本章文献[1]了解深度联合滤波的基本概念、算法原理，熟悉文献提出的模型框架，分析理解该模型如何解决目前相关算法的缺陷。参考提出的模型框架结构，自己动手搭建深度联合滤波器网络模型，收集实验所需的数据进行实验，训练模型，测试模型结果。总结分析结果并修改模型框架，尽量提高模型性能和处理效果。

（2）任务

① 文献收集与阅读，重点参考本章文献[1]，同时结合该文献的部分参考文献了解联合滤波的主要目的、期望解决的问题、前人的工作成果和研究方法，整体上形成对该任务较为清晰的认识。

② 参考上述文献提出的模型，应用 PyTorch 深度学习框架搭建深度联合滤波器模型。

③ 收集模型需要的数据集，并将其处理成适应模型训练和测试的数据样本集合。

④ 向模型输入数据，进行实验，得出并分析实验结果，依此调整模型超参数，重复实验，对比并分析实验结果。

⑤ 总结实验结果，并与原论文实验结果进行对比，得出结论。

2.2 相关技术介绍

本章参考的技术主要包括卷积神经网络、引导滤波、深度图联合上采样、各种图像结果评价指标。

2.2.1 卷积神经网络

卷积神经网络是计算机视觉领域一个经典的模型，广泛用于各种图像处理任务，以下分别从基本概念和特点两个方面进行介绍。

(1) 基本概念

卷积神经网络(CNN)[8]是一种前馈神经网络，尤其在图像处理领域表现十分出色。CNN 的基本结构由输入层、卷积层(Convolutional Layer)、池化层(Pooling Layer)、全连接层以及输出层组成。卷积层和池化层一般会有若干个，卷积层和池化层交替设置，通常是若干个卷积层连接一个池化层，依此类推。卷积神经网络示意图如图 2-2-1 所示。

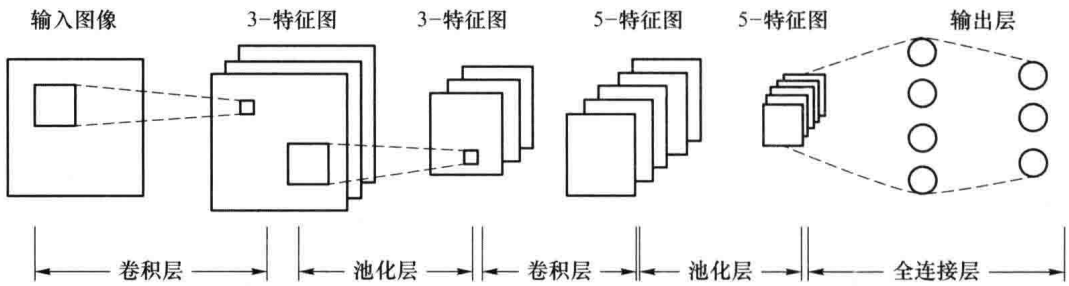

图 2-2-1 卷积神经网络示意图

常用架构模式可用如下公式表示：

$$\text{INPUT}\rightarrow[[\text{CONV}]*\text{N}\rightarrow\text{POOL}]*\text{M}\rightarrow[\text{FC}]*\text{K} \tag{2-2-1}$$

对于上述示意图展示的卷积神经网络可表示为：

$$\text{INPUT}\rightarrow[[\text{CONV}]*1\rightarrow\text{POOL}]*2\rightarrow[\text{FC}]*2 \tag{2-2-2}$$

(2) 特点

① 局部区域连接

在传统的神经网络结构中，神经元之间的连接是全连接的，即当前层的每个神经元都与上一层的所有神经元相连接。但是在卷积神经网络中，当前层的神经元只与上一层的部分神经元相连。这样能大大减少参数数量。

结合示例说明，假设处理一张 1 000×1 000 像素的图片，对于全连接神经网络而言，输入层就有 1 000×1 000=100 万个节点。设置隐藏层有 100 个节点，这在通常的全连接神经网络结构中并不算多，那么从输入层到隐藏层的参数就有(1 000×1 000+1)×100=1 亿个权重参数，这是很大的参数规模，在此基础上，图像增大一点点，参数数量就会多很多，所以，卷积神经网络的扩展性很差。与全连接神经网络对比，以该章的模型为例，每个子 CNN 包含三个卷积层，Channel 为[96,48,3]，Kernel Size 为[9,1,5]，则参数数量为(9×9×96+1)+(1×1×48+1)+(5×5×3+1)=7 902 个，且参数数量与上一层神经元个数无关，与全连接神经网络相比参数数量大大减少了，便于模型的训练。

② 权值共享

权值共享是卷积神经网络重要的特点，主要解决全连接神经网络不能利用图像像素之间的位置信息这个缺点，同时它也能够进一步减少网络参数。对于图像来说，每个像素都与其周围的一些像素之间存在较为紧密的联系，而和离得较远的像素之间的联系就会少很多。结合这点来分析全连接神经网络就会发现问题，全连接神经网络当前层每个节点都和上一层的每个节点相连，也就是说对每个像素而言，把图像的所有像素都同等看待，这显然跟上述分析相悖。因此，全连接神经网络在学习的过程中我们可能会发现权重矩阵中存在大量元素的值很小，即这些连接不重要。学习大量不重要的参数，这样的学习是非常低效的。

卷积神经网络卷积层中的卷积核类似于一个滑动窗口，在整个图像中按照预先设置好的步长滑动并扫描整张图片，经过卷积计算之后输出得到当前卷积层的 Feature Map 输出，权值共享具体指的就是一个卷积核的权重系数在扫描图片的过程中不变，当然 CNN 中每个卷积层可有不止一个卷积核。一方面，权值共享的卷积核大大减少了卷积核中参数的数量，降低了网络复杂度；另一方面，传统的神经网络和机器学习方法会对图像进行预处理提取特征，而卷积神经网络中卷积的操作具有利用图像空间局部性的优势，再结合多个卷积核，可提取图像多种特征。

③ Pooling 层

池化层的主要作用是下采样，通过去除 Feature Map 中不重要的样本，实际上可以理解为一个特征选择的过程，另一方面就是能进一步减少参数数量。Pooling 的方法很多，最常用的是 Max Pooling。Max Pooling 实际上就是在 $n\times n$ 的样本中取最大值，作为采样后的样本值。除 Max Pooling 外，常用的还有 Mean Pooling，即取各样本的平均值。对于深度为 D 的 Feature Map，各层独立做池化，因此池化后的深度仍为 D。另外，池化层可以在一定程度上提高空间不变性，如平移不变性、尺度不变性、形变不变性等。

④ 激活函数

该模型使用的激活函数是 ReLU 函数[9](Rectified Linear Unit)。ReLU 函数示意图如图 2-2-2 所示。

ReLU 函数的优势包括以下几个方面。

① ReLU 函数是一个分段线性函数，把负数输入变为 0，正数输入不变，这种操作称为单侧抑制。单侧抑制使得神经网络模型中的神经元具有了稀疏激活性。稀疏激活性

此为试读，需要完整PDF请访问：www.ertongbook.com