



同濟大學

TONGJI UNIVERSITY

硕士学位论文

(专业学位)

低分辨率人脸的多粒度识别研究

姓 名：饶梦

学 号：1732970

所在院系：电子与信息工程学院

专业学位名称：工程

专业领域：计算机技术

指导教师：苗夺谦 教授

二〇二〇年五月



同濟大學
TONGJI UNIVERSITY

A dissertation submitted to
Tongji University in conformity with the requirements for
the degree of Master of Engineering

Research on Multi-Granular Recognition of Low-resolution Face

Candidate: Rao Meng

Student Number: 1732970

School/Department: College of Electronics and
Information Engineering

Discipline: Engineering

Major: Computer Technology

Supervisor: Prof. Miao Duoqian

May, 2020

低分辨率人脸的多粒度识别研究

饶梦

同济大学

学位论文版权使用授权书

本人完全了解同济大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版本；学校有权保存学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或者部分的阅览服务；学校有权按有关规定向国家有关部门或者机构送交论文的复印件和电子版；在不以赢利为目的的前提下，学校可以适当复制论文的部分或全部内容用于学术活动。

学位论文作者签名：

年 月 日

同济大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

年 月

摘要

人脸识别是计算机视觉领域的研究热点,目前针对高分辨率人脸的识别研究成果显著,在一定程度上已经超过了人类专家的识别精度。但现实场景中,由于监控设备分辨率低或行人距摄像头位置远等因素,采集的人脸图像往往是低分辨率的,而低分辨率的人脸图像人脸尺寸小,图像模糊,缺乏有效信息,同时存在着噪声、光照、姿势变化的影响,使得提取差异化丰富的人脸特征较为困难,大大增加了识别的难度。为解决以上问题,本文开展了低分辨率人脸图像与已知高分辨率人脸图像之间的人脸验证工作的研究,主要研究内容包括:

1. 针对已知低分辨率人脸识别模型将低分辨率人脸图像和高分辨率人脸图像映射到同一尺度空间进行特征提取和匹配,忽略了不同尺度空间存在的一致性信息和互补性信息。本文提出了一种基于多尺度空间的多粒度三支识别模型。模型首先在高尺度空间通过深度卷积网络获取人脸图像面部区域的深度视觉特征,再基于三支决策理论进行粗分类,分为接受、拒绝和延迟;然后对粗分类结果中由于信息不足无法给予判断的延迟样本,通过提取低尺度空间与高尺度空间的互补信息,再进行细分类。并在 LFW、CelebA 数据集上开展实验,结果表明提出的模型取得了满意的结果,超越了传统的单粒度二支模型。

2. 为获取低分辨率人脸图像高尺度空间的深度视觉特征,需要通过低分辨率人脸图像重建高分辨率人脸图像。针对高分辨率图像重建算法 SRCNN 使用最小化均方差作为损失函数,缺乏高频信息,无法保证图像保真度,难以解决较大缩放比例下的图像重建问题,本文借鉴 SSIM 质量度量指标中结构相似性函数,将均方差与结构相似性函数融合,让模型自动约束重建图像与高分辨率图像结构的一致性,从而提升重建图像的效果。

3. 低分辨率数据作为高分辨率图像重建的输入,影响着高分辨率图像重建的质量。针对常用低分辨率人脸图像由高分辨率数据经过双三次、均值、最值等图像降采样算法获取,存在亮度不一致、边缘锯齿、梯度丢失、计算复杂等问题,本文综合已验证的能有效保留梯度信息的梅花均匀图像压缩算法和计算简单的均值图像降采样方法(属于图像退化范畴),提出了一种新的低分辨率人脸图像退化算法。并在不同高分辨率图像重建算法和数据集上进行了实验,结果表明提出的算法取得了良好的性能。

关键词: 低分辨率, 多粒度, 人脸识别, 三支决策

ABSTRACT

Face recognition is a research hotspot in the field of computer vision. At present, the results of research on high-resolution face recognition are remarkable, and it has exceeded the recognition accuracy of human experts to a certain extent. However, in real scenes, due to factors such as low resolution of the monitoring device or the distance of pedestrians away from the camera, the collected face images are often of low resolution, while the low-resolution face images have small face sizes, blurred images, and lack of effective Information. At the same time, there are the effects of noise, lighting, and posture changes, making it difficult to extract rich and diverse facial features, which greatly increases the difficulty of recognition. In order to solve the above problems, this paper carried out research on face verification between low-resolution face images and known high-resolution face images. The main research contents include:

1. According to the known low-resolution face recognition model, the low-resolution face image and the high-resolution face image are mapped to the same scale space for feature extraction and matching, and the consistent information and complementary information existing in different scale spaces are ignored. This paper proposes a multi-granular three-way recognition model based on multi-scale space. The model first obtains the deep visual features of the face area of the face image in a high-scale space through a deep convolutional network, and performs rough classification based on three-way decisions theory, divided into acceptance, rejection, and delay; Judgment delay samples, by extracting the complementary information of the low-scale space and the high-scale space, and then performing fine classification. Experiments were carried out on two datasets, LFW and CelebA. The results showed that the proposed model achieved excellent results, surpassing the traditional single-granular two-way model.

2. In order to obtain the depth visual characteristics of the high-resolution space of the low-resolution face image, it is necessary to reconstruct the high-resolution face image from the low-resolution face image. For the high-resolution image reconstruction algorithm SRCNN uses the minimized mean square error as the loss function, it lacks high-frequency information, cannot guarantee image fidelity, and it is difficult to solve

the problem of image reconstruction at large zoom ratios. The paper draws on the structure of SSIM quality metrics. The similarity function combines the mean square error and the structural similarity function to allow the model to automatically control the structural consistency of the reconstructed image and the high-resolution image, thereby improving the fidelity of the reconstructed image.

3. Low-resolution data is used as input for high-resolution image reconstruction, which affects the quality of high-resolution image reconstruction. For commonly used low-resolution face images obtained from high-resolution data through bicubic, mean, and maximum image downsampling algorithms, there are problems with inconsistent brightness, edge aliasing, gradient loss, and complex calculations. A plum blossom uniform image compression algorithm that effectively retains gradient information and a simple mean calculation method, a new low-resolution face image acquisition algorithm is proposed. Experiments have been performed on different high-resolution image reconstruction algorithms and data sets. The results show that the proposed algorithm has achieved excellent performance.

Key Words: Low-resolution, Multi-granularity, Face recognition, Three-way decisions

目录

摘要.....	I
ABSTRACT.....	II
目录.....	IV
第 1 章 绪论.....	1
1.1 引言.....	1
1.2 研究背景与意义.....	1
1.3 低分辨率人脸识别研究现状.....	2
1.4 本文工作.....	5
1.5 本文结构.....	6
第 2 章 低分辨率人脸识别的相关技术.....	8
2.1 引言.....	8
2.2 低分辨率人脸的多粒度识别模型.....	8
2.3 图像退化用于低分辨率生成.....	9
2.4 图像超分辨率重建用于低分辨率识别.....	11
2.4.1 基于插值的超分辨率算法.....	11
2.4.2 基于学习的超分辨率算法.....	13
2.4.3 针对人脸的超分辨率算法.....	14
2.5 图像质量评价标准.....	16
2.6 数据集介绍.....	17
2.6.1 超分辨率数据集.....	17
2.6.2 人脸识别数据集.....	18
2.7 本章小结.....	20
第 3 章 图像退化与超分辨率研究.....	21
3.1 引言.....	21
3.2 基于梅花均匀图像压缩的图像退化算法.....	21
3.3 基于结构约束的图像超分辨率算法.....	23
3.3.1 基于 CNN 的超分辨率模型.....	23
3.3.2 基于结构约束的损失函数设计.....	25
3.4 实验与分析.....	26
3.4.1 图像退化算法性能比较.....	26
3.4.2 图像超分辨率算法性能比较.....	30
3.5 本章小结.....	32
第 4 章 基于三支决策的低分辨率人脸的多粒度识别.....	33
4.1 引言.....	33

4.2 基于三支决策的多粒度识别模型设计.....	33
4.2.1 三支决策理论.....	33
4.2.2 基于三支决策的多粒度识别模型.....	34
4.3 人脸图像特征提取.....	36
4.3.1 高分辨率图像特征提取.....	36
4.3.2 低分辨率图像特征提取.....	37
4.3.3 特征融合.....	39
4.4 实验与分析.....	40
4.4.1 实验环境.....	40
4.4.2 评价指标.....	40
4.4.3 实验内容.....	41
4.4.4 实验分析.....	41
4.5 本章小结.....	43
第 5 章 总结与展望.....	44
5.1 总结.....	44
5.2 展望.....	45
致谢.....	46
参考文献.....	47
个人简历、在读期间发表的学术论文与研究成果.....	51

第 1 章 绪论

1.1 引言

低分辨率人脸的识别是人脸识别问题的一个子领域，旨在实现高分辨率基准人脸图像与现实监控系统场景下获取的低分辨率的人脸图像的验证。随着公共场所监控系统的广泛运用，远距离下非约束的低分辨率人脸识别系统的市场需求不断扩大，使得低分辨率人脸的研究成为必要。低分辨率人脸的识别研究最早始于 2000 左右，经过近 20 年的发展，针对低分辨率人脸图像尺寸小、质量差和缺乏特征信息等问题，众多国内外学者对其进行了研究，取得了一定的成果。

本章的主要结构如下：1.2 章节探索了低分辨率人脸识别的研究背景和研究意义；1.3 章节梳理了低分辨率人脸识别问题的研究现状；1.4 章节总结了本文的主要工作和相关贡献；1.5 章节概括了本文的组织架构。

1.2 研究背景与意义

近年来，随着深度学习技术在计算机视觉领域的运用，针对高分辨率清晰人脸的识别算法和模型，在广泛的大规模人脸数据集上已经取得了卓越成果，甚至超越了人类专家的识别精度。但是，为提取丰富的高维图像特征，这些算法对人脸图像的质量要求很高。而在社区、街道、游乐园等室外场所布置的监控摄像头像素质量一般，并且可能由于行人距离摄像头较远，无法获取高分辨率高质量大尺寸的人脸图像，这给低分辨率人脸的识别带来了极大挑战。

低分辨率人脸识别旨在实现监控系统下获取的低分辨率人脸数据与数据库中已知高分辨率人脸数据的识别验证，实现智能监控系统实时捕捉在逃人员。现有的高分辨率人脸识别系统提供了对高分辨率高质量人脸图像高尺度空间的特征描述，并且在已学习到的特征描述上建立了可区分的特征空间。受此影响，低分辨率人脸识别工作也尝试着获取高分辨率图像与低分辨率图像同一尺度空间的可区分特征的描述。

总结众多专家学者实现低分辨率与高分辨率图像数据同一尺度空间特征提取的策略主要分为三类：一是将高分辨率人脸图像转化为低分辨率图像，提取与低分辨率图像相同的低尺度鲁棒特征；二是将高分辨率人脸图像与低分辨率人脸图像映射到二者之间的一个尺度空间中，提取相同的特征；三是将低分辨率人脸图像重建到高分辨率人脸图像的尺度空间，称为图像超分辨率，从而提取相同的

高尺度特征信息。图 1.1 展示了实现低分辨率与高分辨率人脸图像同一尺度空间的三种策略。

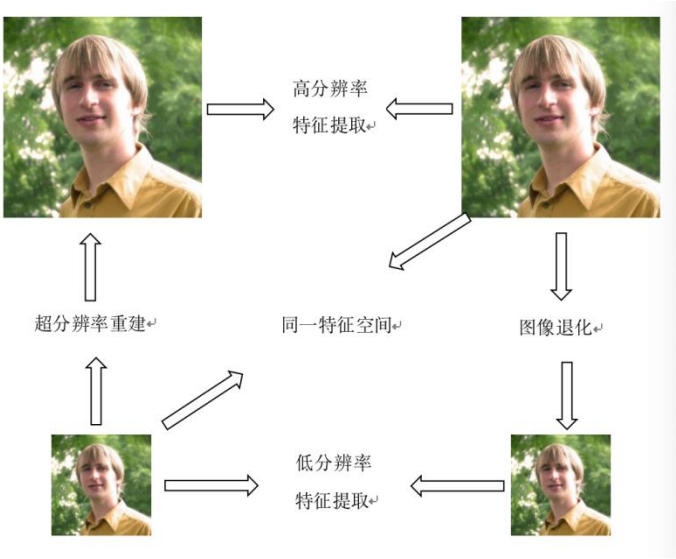


图 1.1 低分辨率人脸识别策略

人类视觉感知系统具有非常强的人脸辨识能力，其不仅可以通过一张正脸图片识别一个人，而且可以从一系列的行人侧脸、背面、形态等比对出行人身份，同时具有超强的鲁棒性。无论是图像模糊、存在遮挡、光照变化等，人类都可以准确的给出判断。而粒计算（Granular Computing, GrC）^[1]就是人工智能领域模拟人类思维 and 解决复杂问题的新方法，它涵盖了商空间理论、三支决策理论和粗糙集理论等主要内容，覆盖了有关粒度的主要理论、方法和技术，是研究复杂问题求解、大数据挖掘和不确定性信息处理等问题的有力工具。同样，三支决策（Three-way Decisions）也是一种基于人类认知的决策理论，其将贝叶斯风险决策方法与粗糙集理论^[2]中的正域、负域和边界域结合，形成了接受决策、拒绝决策和延迟决策的三支决策语义^[3]，对可以确定接受或拒绝的事物，立即决策；对不确定的事物，推迟判断。

以粒计算和三支决策的视角看待问题，给低分辨率人脸识别带来了新的思想。因此本文在现有的低分辨率人脸识别基础上，针对单一尺度空间判别的缺点，基于三支决策和粒计算的思想，根据图像的多个尺度特征，提出了低分辨率人脸的多粒度识别模型。同时，本文结合相关实验结果的分析证明了本文所提方法的有效性。

1.3 低分辨率人脸识别研究现状

低分辨率人脸识别技术的研究是在高分辨率人脸识别广泛在大规模公开数

据集上识别准确率达到较高水平,取得非凡成果之后,并且在此之前专家学者并未针对低分辨率人脸识别问题开展专门的研究,对人脸识别技术的研究可以追溯到计算机视觉和模式识别技术的开始。20世纪90年代,特征脸方法的出现使人脸识别研究逐渐流行;1990-2000期间,基于全局特征的方法占据了人脸识别领域的主导地位,出现了线性子空间、流形和稀疏表示等理论;2000年初,基于局部特征的方法产生, Gabor^[4]和 LBP^[5]以及它们的多级高维扩展利用局部滤波的部分不变性获得了鲁棒的性能,但这些手工制作的特征却缺乏独特性和紧凑性;2010年初,基于学习的局部描述符开始流行,局部滤波器学习了更佳的特性和紧凑性。但上述浅层的特征表示终究无法实现复杂的无约束环境下非线性人脸外形变化的鲁棒性表示。2012年, AlexNet^[6]深度学习网络以绝对的优势赢得了 ImageNet 的比赛。卷积神经网络 (Convolutional Neural Networks, CNN) 的多层处理单元学习了不同抽象级别的多个表示,多层抽象的组合表示提高了面部身份的稳定性,并且表现出对面部姿势、灯光和表情变化的鲁棒性。2014年, DeepFace^[7]在 Labeled Faces in the Wild (LFW)^[8]基准上达到了最先进的精确度,首次在无约束条件下接近人类的表现 (DeepFace: 97.35%, 人: 97.53%)。此时,低分辨率人脸的识别进入了专家学者的研究范围,成为了人脸识别领域又一研究热点。

与上述高分辨率场景下的人脸识别系统相比,低分辨率场景下的人脸识别系统具有噪声多、信息量少、缺乏有效方法等突出缺点。如果直接将高分辨率人脸系统移植到低分辨率场景,必然存在识别准确率低、算法失效等问题。因此,众多专家学者针对低分辨率人脸的识别,设计了新的有效的方法。主流的低分辨率人脸识别方法可以大概分为三类:一是基于低分辨率图像数据直接提取鲁棒的特征表示用于识别;二是将低分辨率人脸图像和高分辨率人脸图像映射到同一的中间的尺度特征空间获取人脸的可辨别特征;三是从低分辨率人脸图像重建对应的高分辨率图像,基于重建的高分辨率图像运用已有的高分辨率人脸识别算法提取特征从而提高识别的准确率。

1. 基于鲁棒特征的低分辨率人脸识别

基于鲁棒特征表示的方法旨在直接从低分辨率人脸图像上提取图像的可鉴别的、稳定的特征信息,用于低分辨率人脸的识别。而这些特征主要源于对高分辨率人脸特征的改进,例如改进的颜色空间^[9]和局部二进制模式描述符^[10];同时,基于特征表示的方法又分为全局和局部特征。基于全局特征的方法以含全局低频信息的单个高维矢量来表示整个低分辨率图像,可以隐式地保留全部细节纹理和形状信息,但与高分辨率人脸识别一样,易受姿态、光照、表情等变化的影响;

而基于局部特征的分类方法以一组包含局部高频信息的低维向量表示低分辨率图像,较全局方法相比,更加灵活,能够识别人脸的各个部分,并且对姿态、光照等变化具有更强的鲁棒性。

由于高频信息在低分辨率图像中的丢失,有必要在低频域中探索可靠的信息。Lei 等^[10]基于局部相位量化 (LPQ) 背后的思想,探索了低频域的模糊不变幅度和相位信息,提出了一种有效的局部频率描述符 (LFD),用于低分辨率人脸识别。识别低分辨率和模糊图像中的脸部是常见但具有挑战的任务,鉴于 LFD 是从频率平面每个像素的空间邻域中提取出来的,而与频率之间的相关性无关, Li 等^[11]探索了频率相关性,提出了增强型 LFD (ELFD),空间和频率信息的联合利用,使得 ELFD 比 LFD 更具描述性和判别力。Lu 等^[12]提出了一种半局部约束表示法,同时学习判别性表示法以及低分辨率和高分辨率特征之间的映射关系,低分辨率和高分辨率的局部流形几何都在表示过程中进行了编码,而且学习的映射关系可以将低分辨率特征转换为高分辨率特征,提高了流形一致,最后,将鲁棒性强的特征输入到基于稀疏表示的分类器中,以预测人脸标签。

2. 基于同一特征空间的低分辨率人脸识别

由于低分辨率人脸图像中原始高分辨率图像的面部纹理信息丢失,传统识别算法的性能急剧下降。Li 等^[13]基于耦合映射的方法将具有不同分辨率的面部图像投影到同一的、中间的特征空间中,并通过优化目标函数以最小化对应关系(即低分辨率图像与对应的高分辨率)之间的差异来学习这些耦合映射,还在目标函数中引入了惩罚加权矩阵,大大提高了识别性能。Jiang 等^[14]提出了一种称为耦合判别多流形分析的方法,首先探索了样本跨越的多流形空间的邻域信息以及局部几何结构;然后,通过监督的方式学习了将低分辨率和高分辨率人脸投影到同一的判别特征空间的两个映射,以最大化判别信息的距离。Zhou 等^[15]提出了一种有效的同时判别分析 (SDA) 方法,SDA 从低分辨率和高分辨率图像分别学习了两个映射到一个共同的子空间,通过映射到公共空间减小了低分辨率和高分辨率之间的数据间隙,同时保留大多数区分性信息。

3. 基于重建高分辨率图像的低分辨率人脸识别

针对低分辨率图像无法获取丰富的可区分的特征信息,一些学者将低分辨率人脸的识别分为两阶段,一阶段从低分辨率图像重建高分辨率图像,二阶段基于高分辨率人脸图像进行识别。Yang 等^[16-17]在基于稀疏表示的图像超分辨率方法上,实现了对人脸图像的重建;受该工作的启发, BilGazyev 等^[18]针对低分辨率图像丢失的高频信息,在面部数据中建立了一个高频分量字典,并将其添加到低分辨率输入图像中创建高分辨率图像,以提高人脸识别率。Shekhar 等^[19]对训练

集中每一类别的高分辨率样本生成不同光照下的图像,构成一个类别,以对抗光照变化,并在测试低分辨率图像时,先投影到每个学习词典中的原子跨度上,再使用残差矢量进行分类。Kim 等^[20]基于输入输出图像对,学习低分辨率图像与高分辨率图像的映射关系,并结合核匹配追踪和梯度下降的思想,降低了算法训练时间的复杂度。

1.4 本文工作

本文主要介绍了基于三支决策的低分辨率人脸的多粒度识别模型。为解决低分辨率人脸图像缺少原始高分辨率图像的纹理、结构等面部细节信息,无法获取有效的可辨别特征,并与高分辨率图像存在尺寸失配等问题,本文研究了低分辨率图像重建为高分辨率图像的超分辨率算法,针对在高分辨率图像重建过程加入对图像结构的约束,以提高重建人脸的保真度;此外,作为超分辨率算法输入的低分辨率数据影响着图像重建的质量,我们又将梅花均匀图像压缩算法引入到低分辨率图像的退化过程,以减少图像梯度信息的丢失。具体来说针对低分辨率人脸的识别和人脸图像重建中存在的一些问题,本文提出了“一种基于梅花均匀图像压缩的图像退化算法”、“一种基于结构约束的超分辨率算法”和“基于三支决策的低分辨率人脸的多粒度识别模型”。论文的主要工作包括:

1. 针对已知低分辨率人脸识别模型忽略了不同尺度空间存在的一致性信息和互补性信息,缺乏对高分辨率的高尺度和低分辨率的低尺度信息的共同建模和学习,我们提出了一种基于多尺度空间的多粒度三支识别模型。该模型将高分辨率人脸图像和低分辨率人脸图像的特征分别映射到高尺度和低尺度的两个空间,致力于提取不同尺度下人脸图像的不同尺度特征,以提高低分辨率人脸的识别准确率。具体来说,对于高尺度空间获取的深度视觉特征,基于欧式距离度量相似性,并通过双阈值实现人脸图像判别为接受、拒绝和延迟的三支粗分类;对于延迟分类的对象,融合低尺度空间获取的补充高尺度空间的多类别特征,再进行二次判断的细分类。此外,为保证高、低尺度空间特征融合的合理性,我们对不同特征设计了不同的度量函数。最后在两个尺度上实现了低分辨率人脸的识别,并在 LFW 和 CelebA 数据集上开展了实验,结果表明提出的模型取得了满意的结果,超越了传统的单粒度二支模型。

2. 针对常用低分辨率人脸图像获取的方法存在梯度信息表示缺乏,影响后续人脸图像高分辨率重建的问题,基于本文提出的一种基于梅花均匀图像压缩的图像退化算法,我们希望生成的低分辨率图像保留更多的结构信息,更好地用于高分辨率人脸图像的重建任务,以更佳地用于高尺度空间深度视觉特征的提取。

因此,我们研究了高分辨率图像映射为低分辨率图像的图像退化方法,在均值图像退化算法的基础上,引入了梅花均匀图像压缩的思想,提出了一种新颖的图像退化方法,保障了生成的低分辨率图像梯度信息的保留和边缘的平滑,为后续训练图像超分辨率模型提供了高质量的数据。

3. 针对基于学习的低分辨率图像重建为高分辨率图像的 SRCNN 模型,在学习过程中未对重建图像的结构进行约束,无法保障重建高分辨率图像与真实高分辨率图像结构的一致性问题,我们对原始 SRCNN 模型中使用的单纯均方差损失函数进行了改进,新的损失函数不仅包含了对图像亮度信息的约束,还基于图像结构相似函数对重建图像的结构进行了约束,提高了图像的保真度和视觉效果。并在 Set5、Set14、LFW 和 CelebA 四个公开数据集上开展了实验。实验结果表明,提出的新损失函数能够有效提高重建图像的视觉效果。

1.5 本文结构

本文的重点工作在于低分辨率人脸的识别,文章清晰介绍了低分辨率人脸识别的主流方法、图像退化算法、图像超分辨率技术以及我们提出的一种基于梅花均匀图像压缩的图像退化算法、一种基于结构约束的图像超分辨率算法,以及基于三支决策的低分辨率人脸识别的多粒度模型。

本文共分为五个章节,各章节内容安排如下:

第一章 绪论。本章首先介绍了低分辨率人脸识别问题的研究背景和意义;其次对低分辨率人脸识别的主流方法进行了详细介绍,主要包括基于鲁棒特征表示的低分辨率人脸识别、基于同一特征空间的低分辨率人脸识别和基于高分辨率图像重建的人脸识别;最后,总结了本文的主要工作。

第二章 低分辨率人脸识别的相关技术。本章首先介绍了本文提出的低分辨率人脸识别系统的流程;其次,梳理了基于高分辨率图像生成低分辨率图像的图像退化算法;接着,介绍了基于低分辨率图像重建高分辨率图像的超分辨率技术,并进一步将图像超分辨率方法分为:基于插值的、基于学习的和针对人脸图像的超分辨率方法;然后,对重建高分辨率图像的常用客观质量评价指标进行了介绍,并对这些指标的优缺点进行了比较分析;最后,介绍了目前常用于高分辨率图像重建和人脸识别的几个公开数据集。

第三章 图像退化与超分辨率研究。针对已有高分辨率图像生成低分辨率图像的图像退化算法忽视了对图像梯度信息的保留,以及基于低分辨率图像重建高分辨率图像的超分辨率模型忽略了对重建图像结构的约束这两方面的问题,本文提出了一种基于梅花均匀图像压缩算法有效保留图像梯度信息的图像退化算法,

和基于结构相似性指标改进的约束重建高分辨率图像与原始高分辨率图像结构一致的损失函数。同时还设计了一系列实验，展示了我们算法的有效性。

第四章 基于三支决策的低分辨率人脸的多粒度识别。针对已知基于重建的低分辨率人脸识别算法总是将低分辨率和高分辨率人脸图像映射到高尺度空间进行匹配识别，忽视了不同尺度空间的互补信息，本章提出了基于三支决策的低分辨率人脸识别的多粒度模型，保持了不同尺度空间中图像的可辨别特征，有效地为后续低分辨率人脸的识别提供帮助。并在 CelebA 和 LFW 数据集上开展实验，验证了所提模型的有效性。

第五章 总结与展望。本章凝练总结了本文的研究工作和主要贡献，并且明确了未来研究和改进的方向。

第 2 章 低分辨率人脸识别的相关技术

2.1 引言

近年来，随着深度学习技术的成熟和超分辨率算法的发展，众多专家学者基于图像超分辨率算法实现了从低分辨率图像重建高分辨率图像，并运用已有高分辨率模型解决了低分辨率人脸的识别问题，而超分辨率算法在学习低分辨率图像与高分辨率图像的映射关系时，又需要低分辨率数据作为输入。因此，低分辨率人脸识别问题涉及了低分辨率图像生成的图像退化算法、高分辨率图像重建的超分辨率算法等。

本章的主要结构如下：2.2 章节介绍了本文提出的基于三支决策的低分辨率人脸识别的多粒度模型；2.3 章节概括了图像退化算法的研究现状；2.4 章节梳理了图像超分辨率算法；2.5 章节介绍了高分辨率图像重建效果的评价指标；2.6 章节介绍了相关的超分辨率和人脸识别数据集；2.7 章节总结了本章的主要内容。

2.2 低分辨率人脸的多粒度识别模型

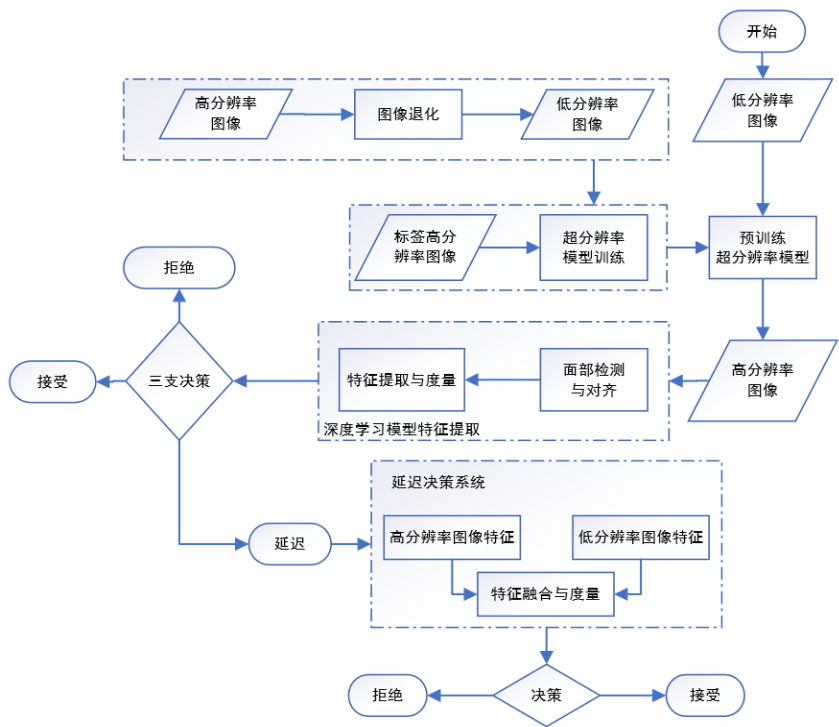


图 2.1 基于三支决策的低分辨率人脸识别的多粒度模型

为实现监控视频中低分辨率人脸的识别,本文研究了用于低分辨率图像生成的图像退化方法、用于高分辨率图像重建的超分辨率模型以及基于三支决策和两个尺度空间特征实现的低分辨率人脸识别的多粒度模型,如图2.1。

基于三支决策的低分辨率人脸识别的多粒度模型主要包括4个模块:一是选择合适的图像退化算法,获得低分辨率数据;二是使用获得的低分辨率数据和原始高分辨率标签数据训练超分辨率模型,输出预训练的模型;三是基于预训练的超分辨率模型重建清晰图像,利用深度学习模型提取面部特征并进行粗分类;四是构建细分类决策系统,对延迟决策的对象,融合高分辨率和低分辨率图像的特征进行二次分类,输出最终决策。

2.3 图像退化用于低分辨率生成

在基于三支决策的低分辨率人脸识别多粒度模型中,我们清晰图像基于预训练的图像超分辨率模型得到。而基于学习的超分辨率模型需要将低分辨率数据作为输入,高分辨率基准数据作为标签,从而学习到低分辨率数据和高分辨率数据的映射关系。然而,目前已有的大规模人脸数据集都是针对高分辨率人脸识别系统而设计,没有专门针对低分辨率人脸识别系统构建的数据集。因而,研究低分辨率数据生成的图像退化算法成为必要。

图像退化旨在尽量减小图像信息的基础上缩小图像的尺寸。若将二维图像看作矩阵,对一张分辨率为 $M \times N$ 的图像,进行 s 倍缩小,则是将 $s \times s$ 大小图像块的特征信息映射到一个像素点上。常用图像退化方法有基于最值、基于均值和基于双三次的方法。最值法选择 $s \times s$ 矩形图像块中的最大值或最小值来表示该矩形图像块;均值法则用 $s \times s$ 图像块内所有像素点的均值表示;而双三次方法获得的低分辨率图像中的一个坐标点的像素值是对应高分辨率图像坐标点距离最近的16个像素点的加权叠加,权重由BiCubic基函数求得。

除单纯缩小图像尺寸的图像退化方法外,图像退化还囊括其他众多复杂函数,许多科研学者对此进行了研究。在使用高斯回归的单图像超分辨率^[21]模型中,作者将图像退化过程看成是基于相机点扩散函数的连续空间域的退化,通过一个模糊内核 f 与高分辨率图像 I_u 进行卷积操作,之后使用比例因子 s ,生成低分辨率图像 I_l ,如公式(2.1)所示。

$$I_l = (f * I_u) \downarrow_s \quad (2.1)$$

在基于稀疏表示的单图像超分辨率统计预测模型^[22]中,作者运用模糊运算符 $Q \in R^{N_l \times N_l}$ 操作高分辨率图像 $I_u \in R^{N_h}$,再加上噪声 v 生成低分辨率图像 I_l ,如公式(2.2)所示。

$$I_l = QI_u + v \quad (2.2)$$

Yang 等^[23]定义低分辨率图像 I_l 生成公式如(2.3)所示。高分辨率图像坐标 (x_u, y_u) 由低分辨率图像坐标 (x_l, y_l) 计算得到, 如公式(2.4)和(2.5), 权重 w 由参数 σ 决定, 如公式(2.6)。

$$I_l(x_l, y_l) = \sum_{x,y} w(x - x_u, y - y_u) I_y(x, y) + v \quad (2.3)$$

$$x_u = s(x_l - 0.5) + 0.5 \quad (2.4)$$

$$y_u = s(y_l - 0.5) + 0.5 \quad (2.5)$$

$$w(\Delta x, \Delta y) = \frac{1}{Z} e^{-(\Delta x^2 + \Delta y^2)/2\sigma^2} \quad (2.6)$$

$x_l \in \{1, 2, \dots, m\}$, $y_l \in \{1, 2, \dots, n\}$, $x \in \{1, 2, \dots, s \times m\}$, $y \in \{1, 2, \dots, s \times n\}$, v 是噪声, s 是比例因子, $m \times n$ 是低分辨率图像的尺寸。

低分辨率图像作为超分辨率模型的输入, 针对高分辨率图像生成低分辨率图像的退化模型仍未统一。通常低分辨率的图像由高分辨率图像经过一个退化映射模型输出, 符号化表示如下。

$$I_x = D(I_u; \delta) \quad (2.7)$$

其中 I_u 代表原始的高分辨率图像, D 代表退化映射函数, δ 表示退化过程中的参数(例如缩放因子、噪声因子等)。而退化映射和参数都是未知的, 现有的许多工作认为函数 D 是图像降采样, 退化过程如公式(2.8)所示。

$$D(I_u; \delta) = (I_u) \downarrow_s, \{s\} \subset \delta \quad (2.8)$$

虽然大多数用于超分辨率的通用数据集都是基于公式(2.8)构建的, 采用具有抗锯齿的双三次方法, 但退化过程是未知的, 没有考虑散焦, 压缩伪影, 各向异性降级, 传感器噪声和斑点噪声等多种影响因素, Zhang 等^[24]用几种操作的组合构建了退化模型, 如公式(2.9)所示。

$$D(I_u; \delta) = (I_u \otimes k) \downarrow_s + n_\zeta, \{k, s, \zeta\} \subset \delta \quad (2.9)$$

$I_u \otimes k$ 表示模糊核 k 与高分辨率图像 I_u 之间的卷积, n_ζ 是增加的服从标准方差 ζ 的高斯白噪声。与公式(2.8)比, 公式(2.9)的组合退化模型与现实世界更接近, 已被证明对超分辨率有益。

在解决单图像超分辨率问题之前, 重要的是清楚了解退化模型, 多种组合操作的另一表达如下。

$$D(I_u; \delta) = (I_u) \downarrow_s \otimes k + n_\zeta, \{k, s, \zeta\} \subset \delta \quad (2.10)$$

2.4 图像超分辨率重建用于低分辨率识别

图像超分辨率是指基于低分辨率图像数据重建清晰高分辨率图像，包括传统基于插值的方法和基于学习的方法，以及针对人脸域的超分辨率技术，又称人脸幻想。

2.4.1 基于插值的超分辨率算法

基于插值的超分辨率算法假设图像的像素值变化是连续的，最为经典和常用的插值方法有最近邻、双线性 and 双三次。

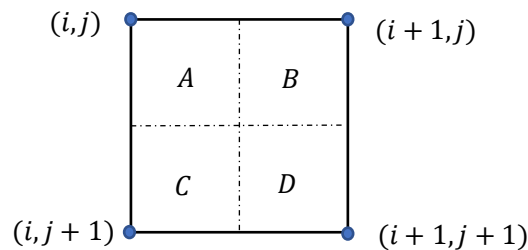


图 2.2 最近邻插值

最近邻插值是一种简单并且直观的算法，该方法为每个要插值的位置选择复制四邻域内最近的像素点，不考虑其他任何像素点，如图 2.2。点 (i, j) ， $(i + 1, j)$ ， $(i, j + 1)$ ， $(i + 1, j + 1)$ 是图像中相邻的四个点，在四点间基于最近邻方法插入新像素点，若新像素点坐标为 $(i + u, j + v)$ ，则新像素点由距该点最近的点决定。虽然该方法简单快速，但可能造成生成图像的灰度值不连续，出现锯齿状。

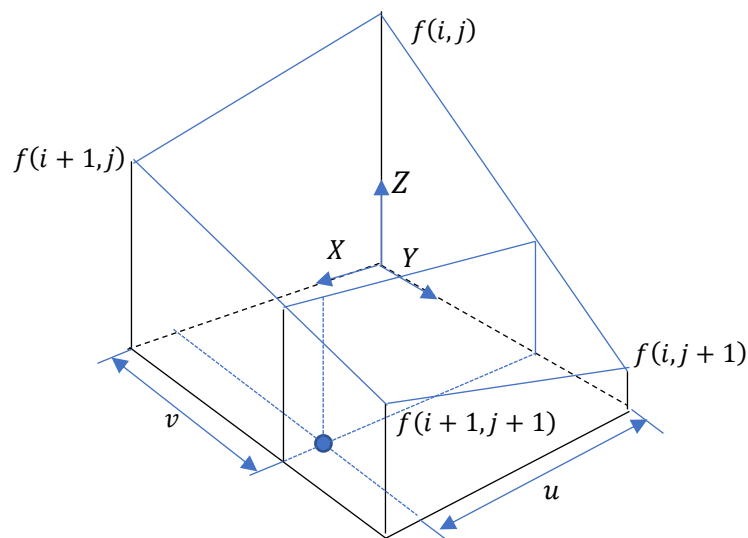


图 2.3 双线性插值

双线性插值需要进行两次线性运算完成插值。同样假设低分辨率图像存在四个相邻点 (i, j) , $(i + 1, j)$, $(i, j + 1)$, $(i + 1, j + 1)$, 待求像素点 $(i + u, j + v)$, 函数 f 表示求该像素点的值。图 2.3 介绍了双线性插值的详细原理。

该方法假设点 $(i, j + v)$ 的值 $f(i, j + v)$ 在三维空间点 $f(i, j)$ 和 $f(i, j + 1)$ 的直线上, 可以求得:

$$f(i, j + v) = (f(i, j + 1) - f(i, j)) * v + f(i, j) \quad (2.11)$$

同理, 得到点 $(i + 1, j + v)$ 的值 $f(i + 1, j + v)$ 在三维空间点 $f(i + 1, j)$ 和 $f(i + 1, j + 1)$ 的直线上, 求得:

$$f(i + 1, j + v) = (f(i + 1, j + 1) - f(i + 1, j)) * v + f(i + 1, j) \quad (2.12)$$

那么, $f(i + u, j + v)$ 则可以通过点 $(i + 1, j + v)$ 和 $(i, j + v)$ 求得, 最终 $f(i + u, j + v)$ 的值为:

$$f(i + u, j + v) = (f(i + 1, j + v) - f(i, j + v)) * u + f(i, j + v) \quad (2.13)$$

双线性插值先在一个轴上进行线性插值, 再在另一个轴上进行插值, 在插值位置和插值大小上都是线性的, 虽然计算复杂度上升了, 但是获得了比最近邻插值更好的效果, 图像更加平滑, 但是可能存在边界模糊的问题。

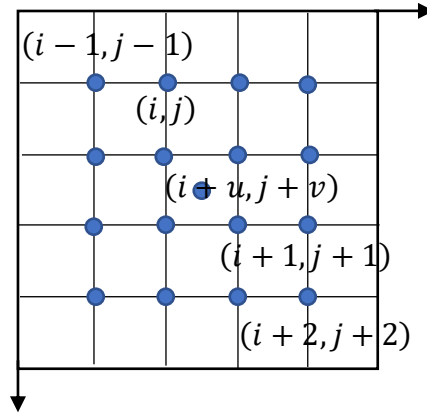


图 2.4 双三次插值

双三次插值^[25]利用三次多项式 $S(x)$ 去逼近最佳的插值函数 $\sin(x)/x$, $S(x)$ 是一个连续函数, 并且一阶偏导数连续, 对新插入的像素点 $(i + u, j + v)$ 的值由其周围 16 个相邻像素点加权得到, 如图 2.4 所示, 公式如(2.14)。

$$f(i + u, j + v) = \begin{pmatrix} S(1 + v) \\ S(v) \\ S(1 - v) \\ S(2 - v) \end{pmatrix}^T W \begin{pmatrix} S(1 + u) \\ S(u) \\ S(1 - u) \\ S(2 - u) \end{pmatrix} \quad (2.14)$$

$$W = \begin{pmatrix} f(i-1, j-1) & f(i-1, j) & f(i-1, j+1) & f(i-1, j+2) \\ f(i, j-1) & f(i, j) & f(i, j+1) & f(i, j+2) \\ f(i+1, j-1) & f(i+1, j) & f(i+1, j+1) & f(i+1, j+2) \\ f(i+2, j-1) & f(i+2, j) & f(i+2, j+1) & f(i+2, j+2) \end{pmatrix} \quad (2.15)$$

双三次插值在图像的每一个维度上执行三次插值，与双线性插值相比，生成的图像更加平滑，但计算复杂度高，是目前构建低分辨率数据集的主流方法。

上述经典基于插值的超分辨率算法都是基于图像相邻区域像素值来插入新的像素点，而通过插值得到的高分辨率图像得到的额外信息微乎其微。针对低分辨率图像丢失的高频信息，一些学者研究了从低频信息预测高频信息的插值算法，Chen 等^[26]基于可用的低频小波系数来估计原始图像的高频小波系数，同时运用多项式曲线拟合了实际高频小波系数与估计的高频小波系数之间的关系。另一方面基于图像本身的自相似性或普遍特征，Freedman 和 Fattal^[27]从输入图像极其局部的区域提取补丁，使用专用的新颖的非二阶滤波器组来实现小比例放大，实现了对低分辨率视频实时增强。

2.4.2 基于学习的超分辨率算法

基于学习的超分辨率算法一方面基于图像梯度先验知识和自然图片遵循的相似性特征，从而实现对高分辨率图像重建的有效约束；另一方面基于深度学习理论，通过大量的样本训练学习低分辨率图像和高分辨率图像的映射关系。

Sun 等^[28]从大量自然图像中预先学习了图像梯度轮廓的先验，获得了梯度的形状和清晰度的参数等信息，实现对重建高分辨率图像梯度的约束。Chang 等^[29]受多种组合方法的启发，发现低分辨率和高分辨率图像中的小图像块在两个不同的特征空间形成了具有局部相似的几何形状，在重建高分辨率图像时，实施局部兼容性和平滑度约束。Glasner 等^[30]结合多图像超分辨率和基于示例的超分辨率提出了一个统一的框架，尝试根据每个像素在尺度内和尺度上的补丁冗余，在每个像素点处恢复最佳分辨率。

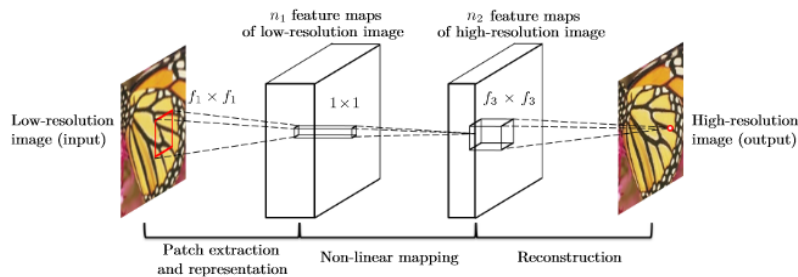


图 2.5 SRCNN 的网络结构^[31]

Dong 等^[31]最早将卷积神经网络引入到高分辨率图像重建问题中,学习了低/高分辨率图像之间端到端的映射,提出了 SRCNN^[31]。该模型不仅同时优化了所有层,还可对三个颜色通道同步处理,结构轻巧、速度快,如图 2.5 所示。Zhang 等^[23]基于数据影响模型的视角,设计了更加复杂的图像退化模型,并基于 CNN 的网络结构实现了难度更大的超分辨率模型。Kim 等^[32]发展了 Zhang 等^[23]构建的单纯 CNN 网络,提出了 DRCN (深度递归卷积网络);之后, Kim 等^[33]又将残差学习引入超分辨率模型,在收敛速度和视觉效果上都表现更佳。

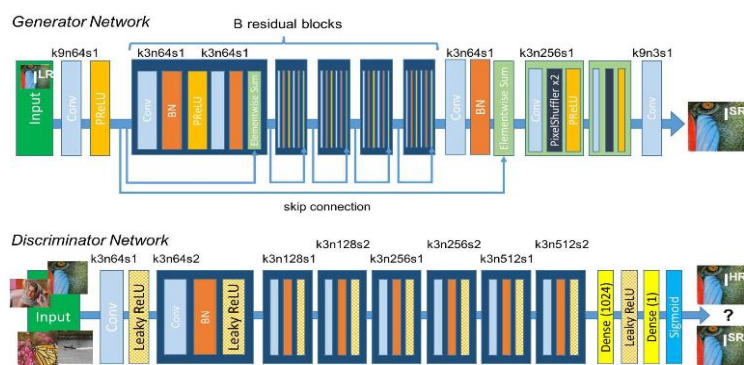


图 2.6 SRGAN 网络结构[34]

虽然基于 CNN 的超分辨率算法在准确性和速度上取得了突破,但仍然无法解决在较大的缩小比例下恢复图像细致的纹理特征。因此, Ledig^[34]提出了基于生成对抗网络 (GAN) 的超分辨率算法 (SRGAN), 可以对 4 倍放大因子推断出逼真自然图像; 同时, 构建了由对抗损失、内容损失和残差学习组成的感知损失函数, 使得 SRGAN 在公共基准上恢复了逼真的纹理, 网络结构如图 2.6 所示。

Wang 等^[35]深入研究了 SRGAN 的网络结构、对抗损失和感知损失三大关键部分, 对其进行改进, 提出了 ESRGAN。改进后的感知函数使用了激活函数之前的特征, 为亮度一致性和纹理恢复提供了更强的监督, 获得了最佳的视觉质量。针对现有基于 GAN 的单图像超分辨率模型无法直接优化不可区分的 PI, NIQE 和 Ma 等图像质量的指标, Zhang 等^[36]提出了带有 Ranker 的 RankSRGAN, 以实现在感知指标方向上优化生成器。

2.4.3 针对人脸的超分辨率算法

人脸幻想是人脸图像的超分辨率, 并始终使用具有强大内聚力的典型人脸先验知识。Zhou 等^[37]提出了一种基于双通道卷积神经网络 (Bi-channel CNN) 的新型网络结构, 从原始输入图像提取鲁棒的人脸表示, 并自适应地整合原始输入图

像和人脸表示两个通道的信息，以预测高分辨率图像。该方法在出现较大外观变化时，仍可持续提高面部图像的分辨率。

Zhang 等^[38]提出了超级身份卷积神经网络（SICNN）来恢复身份信息，以生成与真实身份接近的人脸，在超球空间定义了度量重建高分辨率与对应真实高分辨率人脸之间的超身份信息损失函数；此外，为解决直接使用该损失函数存在的发散问题，还构建了这两个域人脸健壮的身份特征，提出一种整合域的训练方法。

常规的人脸幻想仅限于 2 到 4 倍的缩放因子，Yu 等^[39]提出的判别式生成网络（UR-DGN），不仅可以将 16×16 像素的超低分辨率人脸图像放大 8 倍，还通过引入逐像素 L2 正则项与判别网络的反馈信息，使生成的人脸图像与真实人脸图像更加相似。

Song 等^[40]提出了一种两阶段的人脸幻想方法：第一阶段使用卷积神经网络产生脸部分量，代表基本的脸部结构；第二阶段通过训练高分辨率图像得到脸部更加细粒度的面部结构。两阶段的处理，粗细粒度的结合，使得生成的面部更加逼近真实的整体外观。

Zhu 等^[41]在一个框架内优化了人脸幻想和人脸对齐两个任务，提出了一种新的门控深层双向网络，以覆盖不同级别的纹理细节。Super-FAN^[42]是一个同时处理人脸幻想和对齐的端到端系统，通过回归热图和优化热图损失，将结构信息整合到算法中，以改进人脸幻想效果。MTUN^[43]使用多任务神经网络将人脸的结构信息显式地纳入人脸幻想过程，通过强度相似性获得低级信息，又通过面部热图，获取面部结构的中级信息，实现对非常小的未对齐人脸图像的超分辨。FSRNet^[44]充分利用几何学先验，先构建一个粗糙的超分辨率网络以恢复粗糙的高分辨率图像，再将重建的粗糙高分辨率图像输入精细的超分辨率编码器和先验信息估计网络中提取图像特征，估计面部关键点和解析图，最后融合图像特征和先验信息生成精细的高分辨率图像，结构如图 2.7 所示。

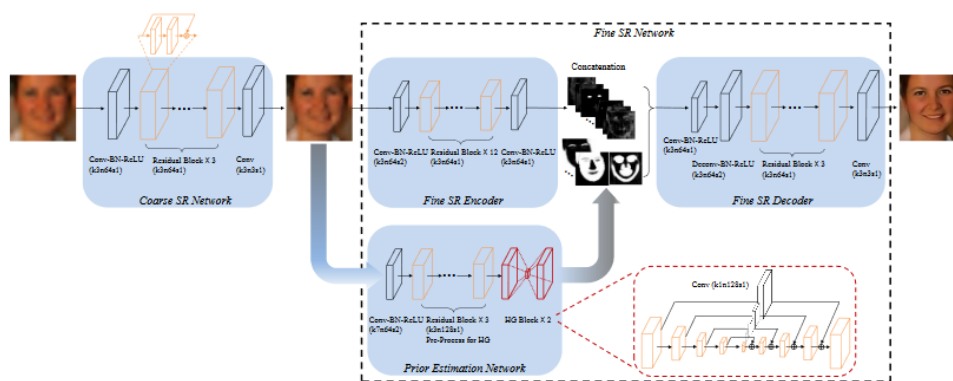


图 2.7 FSRNet 结构^[44]

2.5 图像质量评价标准

图像质量是图像重要的属性，与人类视觉感知重要相关。图像质量评价的方法可分为主观评价和客观度量两类。主观方法就是让人直接对图像进行评分，分数由 1 到 5，最后的平均主观得分（Mean Opinion Score, MOS），是所有人打分的均值。客观度量方法有峰值信噪比（Peak Signal to Noise Ratio, PSNR）、结构相似性（Structural Similarity Index, SSIM）等。

1. 峰值信噪比

峰值信噪比起初是描述信号特征的工程术语。它是信号最大可能功率与影响它表示精度的破坏性噪声的比值。在超分辨率技术领域，PSNR 通常用来度量重建图像的效果，描述重建清晰图像与原高分辨率图像之间的差距，通过最大可能像素值（表示为 L ）和图像之间的均方误差（MSE）定义，该值越大表示图像重建的效果越佳。给定真实图像为 I ，重建的图像为 I_s ，二者都包含 N 个像素点，则真实图像与重建图像之间的 MSE 和 PSNR 计算公式如下。

$$MSE = \frac{1}{N} \sum_{i=1}^N (I(i) - I_s(i))^2 \quad (2.16)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{L^2}{MSE} \right) \quad (2.17)$$

在使用 8bit 表示图像时， L 的值等于 255，PSNR 一般在 20 到 40 之间，值越大越好。固定 L ，PSNR 仅与图像之间像素层次的 MSE 相关，只关注相同位置像素点值的差异，不能包含人的视觉感知（即图像的外观如何）。这使得 PSNR 对真实场景中超分辨率模型重建的高分辨率图像的质量度量不佳。由于缺乏完全准确和统一的感知指标，PSNR 仍是目前最为广泛的评估的标准。

2. 结构相似性

结构相似性指标从图像的结构层次度量图像重建的效果，而峰值信噪比是从像素层级。当看到一幅图像时，人类视觉系统（HVS）往往倾向于记忆图像的结构信息。于是，Wang 等^[45]融合对比度、亮度、结构三个独立分量，提出了 SSIM 指标。假设真实图像为 I ，重建的目标图像为 I_s ，每张图片包括 N 个像素点，使用 $I(i)$ 代表第 i 个像素点，则关于图像亮度的均值和方差如下。

$$u_I = \frac{1}{N} \sum_{i=1}^N I(i) \quad (2.18)$$

$$\delta_I = \frac{1}{N-1} \sum_{i=1}^N (I(i) - u_I)^2 \quad (2.19)$$

重建图像 I_s 与真实图像 I 关于亮度和对比度的差异可以通过各自的均值和方差表示，如公式(2.20)和(2.21)。

$$D_l(I, I_s) = \frac{2u_I u_{I_s} + C_1}{u_I^2 + u_{I_s}^2 + C_1} \quad (2.20)$$

$$D_c(I, I_s) = \frac{2\delta_I \delta_{I_s} + C_2}{\delta_I^2 + \delta_{I_s}^2 + C_2} \quad (2.21)$$

在表示图像亮度和对比度的差异函数时，没有将图像像素点的值进行归一化处理。在考虑图像结构时，对像素点进行了归一化处理，用重建图像 I_s 和真实图像 I 归一化后像素点的协方差来描述结构相似性，协方差计算如公式(2.22)，关于结构的函数如公式(2.23)。

$$\delta_{II_s} = \frac{1}{N-1} \sum_{i=1}^N (I(i) - u_I)(I_s(i) - u_{I_s}) \quad (2.22)$$

$$D_s(I, I_s) = \frac{\delta_{II_s} + C_3}{\delta_I \delta_{I_s} + C_3} \quad (2.23)$$

最终，SSIM 融合对比度、亮度和结构，如公式(2.24)。

$$SSIM = D_l(I, I_s) \cdot D_c(I, I_s) \cdot D_s(I, I_s) \quad (2.24)$$

2.6 数据集介绍

2.6.1 超分辨率数据集

1. Set5 数据集

Set5^[46]数据集源于比利埃大学和法国贝尔实验室在 2012 年公开发布的 BSD 数据集，是基于非负邻域嵌入的低复杂度单图像超分辨率的数据集，包含女性、头部、蝴蝶、鸟和宝宝五张图片，如图 2.8 所示。



图 2.8 Set5 数据集

2. Set14 数据集

Set14^[47]数据集同样源于比利埃大学和法国贝尔实验室发布的 BSD 数据集，主要用于低分辨率图像重建高分辨率图像的单图像超分辨率算法的性能比较，包含 14 张尺寸不一的人、动物和自然照片，被广泛运用计算机视觉与图形学、医学成像和安全监控等领域。

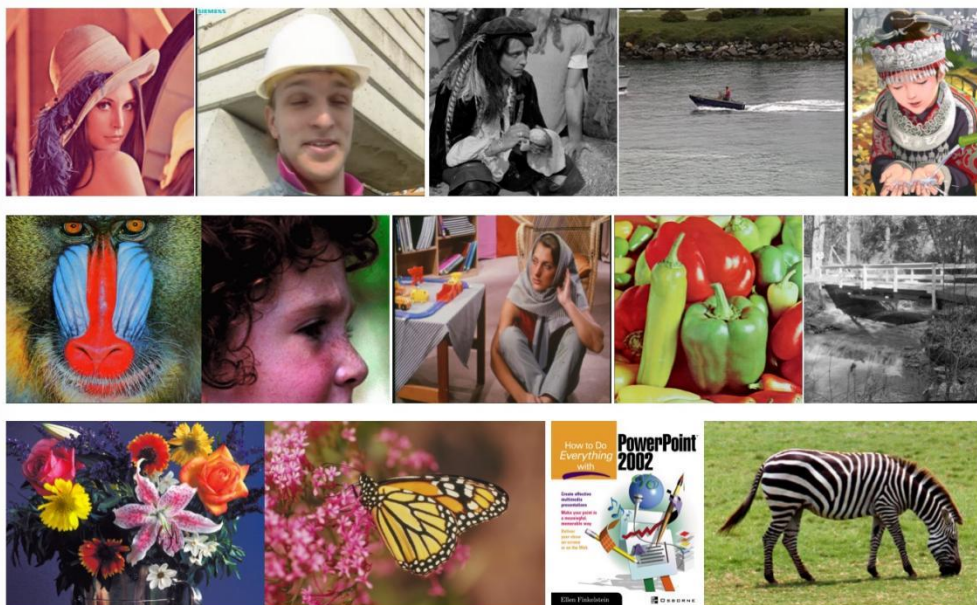


图 2.9 Set14 数据集

2.6.2 人脸识别数据集

1. CelebA 数据集

CelebA (CelebFaces Attributes Dataset) ^[48]人脸属性数据集是一个广泛被使用的大型数据集，由香港中文大学搜索、制作和公开。该数据集包含两大子数据集，一是直接从网络上爬虫，未经裁剪和处理的原始人脸图像；二是经过裁剪和人脸对齐的图像。该数据集包含 10177 个名人，共 202599 张照片，每张照片包括面部标注框、5 个面部关键点的坐标和四十个二进制表示的描述眉毛、秃头、头发、鼻子、下巴、皮肤、发际线、胡须、配饰等特征的标记。该数据集已经被广泛应用于人脸识别、人脸属性识别、人脸检测、人脸对齐和人脸合成等视觉任务上。图 2.10 给出了数据集的示例。



图 2.10 CelebA 数据集图像示例

2. LFW 数据集

LFW (Labeled Faces in the Wild) [8] 是一个专为研究无限制环境下人脸识别问题而设计的人脸照片数据库。该数据集包含从网络收集的 5749 个人的 13233 张面部图像，其中 1680 个人拥有两张或更多不同的照片，每张人脸都标有照片人物的名字。LFW 已被广泛应用于评价人脸验证算法的性能。



图 2.11 LFW 数据集图像示例

除上述数据集外，还有许多与人脸识别与人脸验证任务相关的公开数据集，相关数据集如下表 2.1 所示。

表 2.1 人脸识别数据集介绍

数据集	公开时间	图片数量	描述
Pubfig	2009 年	58797	200 个人
CASIAWebFace	2014 年	49414	10575 个人
MegaFace	2016 年	1000000	690,572 个人
BioID	-	1521	23 个人, 灰度图
CAS-PEAL-R1	2004 年	30900	1040 个人
FaceScrub	2014 年	106863	530 个人
Multi-PIE	2008 年	750000	337 个人
CMUVASC & PIE Face	200 年	40000	68 个人
IJB-A	2015 年	24327	49759 个人
Color feret	1997 年	10000+	1000+个人
CASIA-FaceV5		2500	500 个人
YouTube Faces	-	-	视频, 人脸验证
Person identification in TV	2013 年	-	视频

2.7 本章小结

本章介绍了低分辨率人脸识别问题涉及的相关技术。其中, 详细介绍了从高分辨率图像生成低分辨率图像的图像退化算法的研究现状; 从低分辨率图像重建高分辨率图像的超分辨率方法: 基于插值、基于学习和针对人脸的超分辨率。然后, 针对重建高分辨率图像的效果和质量, 介绍了客观的图像质量评价准则 PSNR 和 SSIM。最后, 详细介绍了用于图像超分辨率和人脸识别的公开数据集: Set5、Set14、CelebA 和 LFW。

第3章 图像退化与超分辨率研究

3.1 引言

图像退化算法的研究是为了重建质量更高的高分辨率图像。对于高分辨率图像重建问题，无论是基于插值还是基于学习的超分辨率算法，都直接受到低分辨率输入图像的影响。现有图像退化算法主要采用均值、最值和双三次实现对高分辨率图像尺寸的缩小，而这些方法往往直接缩小原始图像或仅依据周围像素点生成低分辨率图像，没有设计特定的特征构建手段，使得生成的低分辨率图像存在轮廓模糊、梯度信息消失、信息丢失突出等问题。因此，本章基于一种能够有效保留图像梯度信息的梅花均匀图像压缩算法，设计了一种新颖的保留图像梯度信息的图像退化算法。

图像超分辨率旨在从低分辨率图像重建高分辨率图像，不同的重建算法影响着重建高分辨率图像的效果。基于 CNN 的超分辨率模型通过最小化重建高分辨率图像和真实高分辨率图像的像素差值，学习了低分辨率和高分辨率图像之间的映射，有效提高了重建高分辨率图像的 PSNR 指标。然而，仅依据图像像素值设计的损失函数缺乏对重建高分辨率图像与真实高分辨率图像结构一致性的约束，本文基于图像结构相似性，在图像像素点约束的基础上，进一步融入了对重建高分辨率结构的约束。

本章的主要结构如下：3.2 章节提出一种基于梅花均匀图像压缩的图像退化算法；3.3 章节对基于 CNN 的超分辨率模型（SRCNN）的损失函数进行了改进；3.4 章节对本章提出的两个算法的有效性进行了验证；3.5 章节对本章的内容进行了总结。

3.2 基于梅花均匀图像压缩的图像退化算法

对于图像退化问题，Chrabaszcz 等^[49]认为从高分辨率图像序列间隔抽取构建的低分辨率图像应当倾向于表达与高分辨率图像相似的输入特征，并保留用于分类任务的大多数必要信息。Zissis 等^[50]依据用于分类的图像内容部分的像素邻域按比例小于获得全分辨率图像中相同内容的邻域，将卷积滤波器的接收场按每个维度减小相应比例，并对滤波器权重进行变换，保留原始滤波器对预期特征的提取能力。同时，文献[24,51]认为基于双三次退化生成的低分辨率图像学习的超分

分辨率模型并不是符合真实的低分辨率图像，在不同形式的低分辨率数据上，模型的效果将会减弱。

本章基于梅花均匀图像压缩^[52]和均值图像退化算法，提出了一种新的图像退化算法。梅花均匀图像压缩算法按照梅花的形状间隔抽取像素点，保留对角线相邻的点，丢弃每个像素点相邻的四个像素点，可以减小图像压缩过程中边缘信息丢失的可能性；均值图像退化算法在 s 倍缩小图像时，使用 $s * s$ 图像块内像素点的均值来表示新生成低分辨率图像的特征。图 3.1 给出了梅花均匀图像压缩算法的过程，其中黑色方格代表保留的像素点，而灰色则代表丢失的像素点。

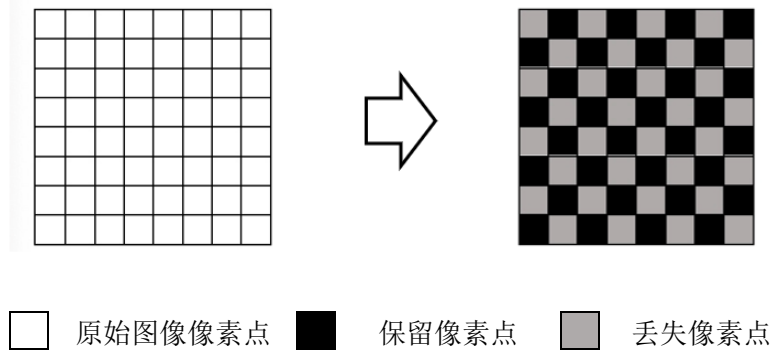


图 3.1 梅花均匀图像压缩算法

梅花均匀图像压缩算法有效保留了图像的梯度信息，但是由于抽取的图像序列不是长宽比例相同的，所以无法直接用于低分辨率图像的生成。因此，为实现图像长宽按相同比例缩小，我们引入了计算简单并且能表现图像局部信息的均值图像退化算法，在梅花均匀图像压缩算法生成的压缩图像上，进一步采用了均值图像退化算法。图 3.2 展示了图像缩小 2 倍时，本文提出算法的处理过程。

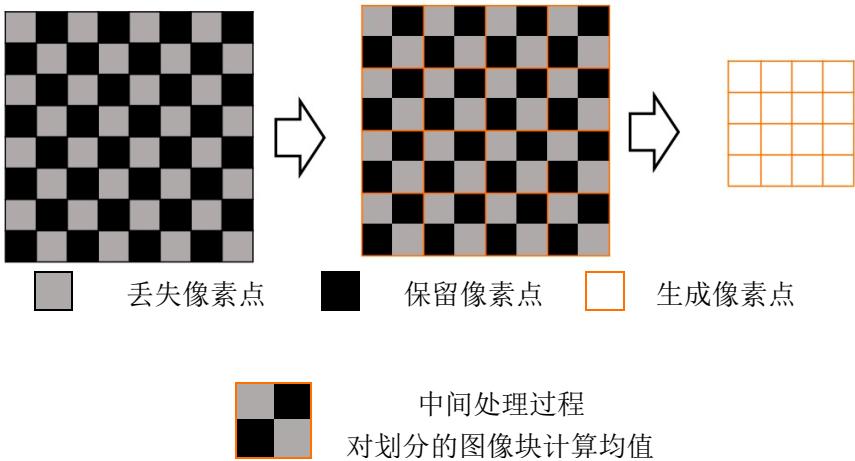


图 3.2 基于梅花均匀图像压缩的图像退化算法

为进一步具体描述本节提出的基于梅花均匀图像压缩的图像退化算法，公式(3.1)给出了数学表达。其中 I_u 是原始高分辨率图像，符号 H 代表梅花均匀压缩，符号 M 表示均值计算。

$$D(I_u; \delta) = MHI_u \quad (3.1)$$

真实的低分辨率图像不仅仅是高分辨率图像尺寸的缩小，还具有图像模糊、包含噪声等特点。因此，我们给出了更为复杂的基于梅花均匀图像压缩的图像退化关系，如公式(3.2)所示，其中 $\otimes k$ 代表模糊的卷积核，具体有均值模糊卷积核、高斯模糊卷积核等， n_ζ 是噪声，可以是高斯白噪声，也可以是椒盐噪声等。

$$D(I_u; \delta) = (MHI_u) \otimes k + n_\zeta, \{k, \zeta\} \subset \delta \quad (3.2)$$

3.3 基于结构约束的图像超分辨率算法

3.3.1 基于 CNN 的超分辨率模型

图像超分辨率与图像退化的过程相反，旨在从低分辨率图像重建尺寸放大、信息丰富、纹理清晰的高分辨率图像。随着深度学习技术的广泛运用，基于 CNN 的图像超分辨率模型（SRCNN^[31]）首先被提出。

3.3.1.1 模型设计

SRCNN 模型对低分辨率图像执行一次图像预处理，将低分辨率图像通过双三次插值法恢复到和高分辨率图像相同的尺寸，表示插值图像 Y ，为易于表达，仍然将 Y 称为“低分辨率”图像。SRCNN 通过补丁提取与表示、非线性映射以及图像重建学习了 Y 与真实高分辨率之间的映射 F ，如图 3.3 所示。

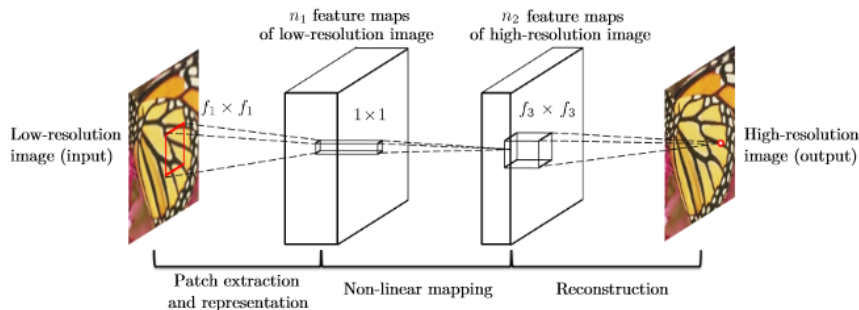


图 3.3 SRCNN 主干网络^[31]

输入图像经过三层卷积神经网络抽取特征, 其中 $f_1 = 9, n_1 = 64, f_2 = 1, n_2 = 32, f_3 = 5, n_3 = 1$, 第一层输出 64 张特征图; 第二层输出 32 张特征图; 第三层输出重建的高分辨率图像。

补丁提取和表示: 密集提取图像的补丁, 用一组预训练的基来表示它们, 例如 PCA, DCT, Haar 等, 是图像重建中普遍采取的策略。这相当于通过一系列的滤波器卷积图像, 每个滤波器就是一个基。此外, 网络还优化了这些基。第一层的操作 F_1 如公式(3.3)所示。

$$F_1(Y) = \max(0, W_1 * Y + B_1) \quad (3.3)$$

其中 W_1 和 B_1 分别代表滤波器和偏置。 W_1 的大小是 $c * f_1 * f_1 * n_1$, c 是输入图像中的通道数, f_1 是滤波器的大小, n_1 是滤波器的数量。激活函数采用整流线性单元 ($ReLU, \max(0, x)$) [53]。

非线性映射: 每个补丁上的 n_1 维特征在第一层网络被提取; 而在操作 F_2 中, 模型将这些 n_1 维向量通过 n_2 个滤波器映射到 n_2 维, 公式如(3.4)所示。

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \quad (3.4)$$

其中 W_2 的大小是 $n_1 * 1 * 1 * n_2$, B_2 是 n_2 维的向量。从概念上讲, 每个输出的 n_2 维向量都是将用于重构的高分辨率补丁的表示。

重建: 在最终完整的图像传统方法通常对预测的重叠高分辨率色块平均。平均可以看作是一组特征图上预定义的滤波器(其中每个位置都是高分辨率补丁的“矢量”形式)。为此, 模型定义了一个卷积层以产生最终的高分辨率图像, 如下式所示。

$$F(Y) = W_3 * F_2(Y) + B_3 \quad (3.5)$$

其中 W_3 的大小是 $n_2 * f_3 * f_3 * c$, B_3 是 c 维的向量。如果高分辨率色块的表示位于图像域中(即, 我们可以简单地对每个表示进行整形以形成色块), 即类似于平均滤波器; 如果高分辨率色块的表示形式在其他某些域中(例如, 以某些基表示的系数), 则我们期望 W_3 的行为类似于先将系数投影到图像域上, 然后取平均。无论哪种方式, W_3 都是一组线性滤波器。

3.3.1.2 损失函数

学习端到端映射函数 F 需要估计参数 $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$ 。这是通过最小化重建图像 $F(Y; \Theta)$ 与真实高分辨率图像 X 之间的损失来实现的。给定一组高分辨率图像 $\{X_i\}$ 及其对应的低分辨率图像 $\{Y_i\}$, 我们使用均方差 (MSE) 作为损失函数, 如下式所示。

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \Theta) - X_i\|^2 \quad (3.6)$$

参数 n 是代表训练样本的总数。损失函数的最小化使用了随机反向梯度下降和标准反向传播。使用 MSE 作为损失函数有利于实现高的 PSNR 指标。值得注意的是, 如果仅推导损失函数, 则卷积神经网络并不排除使用其他种类的损失函数, 在训练过程中给出了更好的感知动机指标, 则网络可以适应该指标。

3.3.2 基于结构约束的损失函数设计

尽管 SRCNN^[31]在 PSNR 指标上达到了良好的效果, 但是仍然存在许多不足。Gwang^[54]等针对 SRCNN 网络结构存在梯度消失和梯度爆炸引起的难以有效学习网络参数的问题, 提出了 MC-SRCNN^[54]模型, 该方法通过生成多种低分辨率图像来补充重建高分辨率图像的信息。并在 Set5 数据集上, 图像放大比例因子为 2 时, 平均 PSNR 增益为 0.21dB。Shi 等^[55]针对 SRCNN 使用单个滤波器将低分辨率输入图像放大到高分辨率空间增加了计算复杂度, 直接在低分辨率图像空间提取特征图, PSNR 指标较 SRCNN 平均提升了 0.15dB。刘等^[56]针对经典 SRCNN 存在边缘失真, 纹理细节恢复模糊的问题, 先在底层特征提取层以多种插值和锐化方法进行预处理, 再与一次插值和锐化的图像合成三维矩阵; 然后, 在非线性映射层将三维矩阵作为深层残差网络的多通道输入, 以获取更丰富的纹理信息, PSNR 和 SSIM 较 SRCNN 平均增加了 0.23dB 和 0.0066。

本章综合考虑了 SRCNN 在超分辨率图像问题上精确度和速度的优秀表现以及难以恢复图像细粒度的纹理、结构等特征的缺点, 针对 SRCNN 基于最小化均方差构建的目标函数缺乏高频细节的问题, 从图像结构感知的角度进行了改进, 研究分析了当前图像超分辨率领域的图像质量的常用度量方法, 提出了基于结构约束的超分辨率卷积网络 (Structural Constraint Super-Resolution Convolutional Neural Network, SC-SRCNN)。

2.5 小节详细介绍了常用客观图像质量评价指标 PSNR 和 SSIM, SRCNN 使用均方误差 (Mean Squared Error, MSE) 作为损失函数, 而 PSNR 由 MSE 构建得到, MSE 越小, PSNR 值越大, 故有利于获得较高的 PSNR, 而较高的 PSNR 并不一定能够保证重建图像与真实图像结构的一致性。因此, 本小节研究了图像的结构相似性特征, 构建了新的损失函数。

SSIM 结构相似性指标由图像亮度、对比度以及结构三个部分构成, 而亮度、对比度的函数与 PSNR 指标具有一定程度的关联, 因为都包含了 MSE。而对于图像结构, SSIM 对像素点进行了归一化处理, 用重建图像 I_s 和真实图像 I 归一化后的像素点的协方差和方差来描述结构相似性, 如公式(3.7), 当常数 c 取值为 0 时, 公式转化为度量图像相关程度的相关系数, 反应了图像的相关性, 如公式(3.8)。两变量相关系数的取值范围是 $[-1, 1]$, 而在图像超分辨率中, 我们期望重建的清晰

图像与原始图像是一致的，即相关系数趋于 1。故本小节基于 PSNR 和 SSIM 重新构建了公式(3.9)的损失函数。

$$D_s(I, I_s) = \frac{\delta_{II_s} + c}{\delta_I \delta_{I_s} + c} \quad (3.7)$$

$$\rho = \frac{\delta_{II_s}}{\delta_I \delta_{I_s}} \quad (3.8)$$

$$Loss = MSE \cdot (1 - \rho) \quad (3.9)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (I(i) - I_s(i))^2 \quad (3.10)$$

其中 I 为给定的真实图像， I_s 为重建的图像， N 为像素点个数， δ_{II_s} 是图像 I 与 I_s 的协方差， δ_I 是图像像素点的方差， c 为常数。在人脸图像数据上，由于各像素点一定存在差异，故 δ_I 的值一定是大于零的。

3.4 实验与分析

3.4.1 图像退化算法性能比较

3.4.1.1 实验环境

本小节的实验主要在 MacBook Pro 设备上开展，设备处理器：2.3GHz IntelCorei5；设备内存：8GB 2133 MHz LPDDR3；设备图形卡：Intel IrisPlusGraphics6551536MB；设备安装操作系统：macOS Mojave 10.14.3。本小节实验代码基于 Python 和 MATLAB 语言编写，Python 版本：3.6；MATLAB：R2018a。而 SRCNN 模型的训练在服务器上开展。

3.4.1.2 实验内容

为证明本文提出的梅花均值图像退化方法的有效性，本小节在标准公开图像超分辨率数据集 Set5 上，与 2 种经典图像退化算法均值(Mean)、双三次(Bicubic)在最近邻(Nearest)、线性(Linear)、双三次(Bicubic)和 SRCNN 这 4 种图像超分辨率算法上进行了比较。最后，分析了不同图像退化算法对高分辨率图像重建算法的增益。

3.4.1.3 实验分析

为直观表现本文提出的基于梅花均匀图像压缩的图像退化算法的有效性，图 3.4 展示了单张图片经过均值(Mean)、双三次(Bicubic)和我们的方法缩小图像 $s = 2$ 和 $s = 4$ 倍后的低分辨率图片。值得注意的是，基于我们的算法生成的低

分辨率图像，明显优于均值图像退化算法，不存在图像锯齿明显的现象；另外，在计算复杂度较双三次更低的情况下，达到了与其同样优秀的效果。

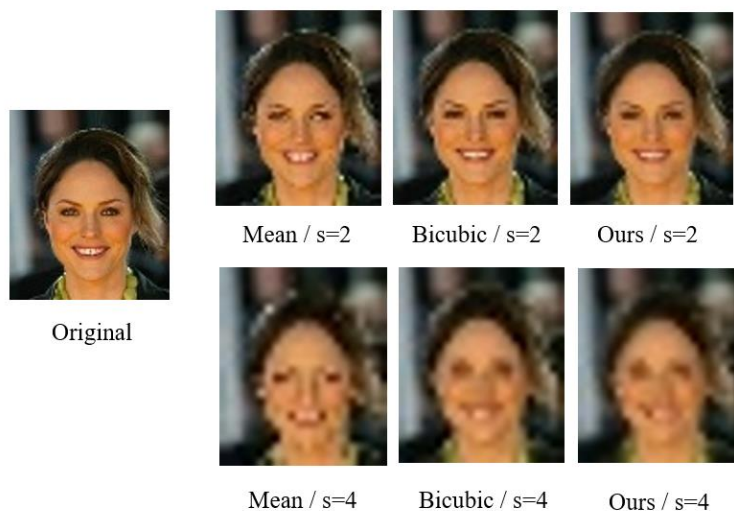


图 3.4 不同图像退化算法效果对比

我们在图像退化与超分辨率问题研究的公开数据集 Set5 包含的 5 张图片上验证了本文所提的图像退化算法的性能，图像缩小比例为 2，实验结果如表 3.1 和 3.2。显然，我们所提的基于梅花均值压缩的图像退化算法完全超越了均值图像退化方法，在基于学习的 SRCNN 超分辨率算法上更是超越了优异的双三次图像退化算法。尽管我们的算法在线性、双三次超分辨率算法上略逊于双三次图像退化算法，但我们所提的算法计算复杂度更低。值得注意的是，在计算复杂度更低的情况下，我们所提的算法在 PSNR 和 SSIM 指标上完全可以与双三次图像退化算法相媲美；而且，在不同的超分辨率算法上，我们所提的算法表现出了更佳稳定性，在传统插值超分辨率算法上结果基本与双三次图像退化算法一致，而在基于学习的超分辨率算法上的性能完全超越了双三次图像退化算法。因而，我们的算法对所有的超分辨率算法而言更具有普适性和稳定性。

表 3.1 基于不同超分辨率算法的图像退化算法性能比较（PSNR）

退化算法/超分辨算法	baby	bird	butterfly	head	women
Mean/Nearest	34.36	33.81	33.18	32.4	33.81
Bicubic/Nearest	34.90	34.26	32.50	32.50	33.69
Ours/Nearest	34.97	34.30	32.77	32.48	33.68
Mean/Linear	34.51	33.80	32.05	32.15	33.39
Bicubic/Linear	36.40	36.41	32.32	32.83	34.53
Ours/Linear	35.64	35.41	32.12	32.50	34.01
Mean/Bicubic	34.41	33.85	31.58	32.02	33.25
Bicubic/Bicubic	37.10	37.62	32.02	32.90	34.91
Ours/Bicubic	36.78	37.09	32.05	32.68	34.71
Mean/SRCNN	33.66	33.39	31.12	31.29	32.9
Bicubic/SRCNN	35.44	36.17	31.51	31.93	34.01
Ours/SRCNN	36.32	37.36	32.10	32.18	34.86

表 3.2 基于不同超分辨率算法的图像退化算法性能比较（SSIM）

退化算法/超分辨算法	baby	bird	butterfly	head	women
Mean/Nearest	0.883	0.892	0.819	0.712	0.88
Bicubic/Nearest	0.929	0.936	0.874	0.820	0.926
Ours/Nearest	0.927	0.935	0.869	0.786	0.923
Mean/Linear	0.911	0.925	0.855	0.752	0.908
Bicubic/Linear	0.947	0.965	0.901	0.816	0.947
Ours/Linear	0.934	0.952	0.883	0.787	0.932
Mean/Bicubic	0.915	0.929	0.862	0.743	0.913
Bicubic/Bicubic	0.958	0.975	0.915	0.831	0.958
Ours/Bicubic	0.952	0.970	0.908	0.803	0.952
Mean/SRCNN	0.884	0.906	0.814	0.673	0.882
Bicubic/SRCNN	0.938	0.961	0.894	0.794	0.938
Ours/SRCNN	0.947	0.969	0.910	0.780	0.952

进一步分析,均值图像退化算法在5张图片上的无论是PSNR指标还是SSIM指标都远低于本文所提的梅花均匀图像退化算法,原因也是我们指出的均值图像退化算法缺乏对原始高分辨率图像结构、梯度信息的保留能力;虽然双三次图像退化算法在传统的基于插值的最近邻、线性和双三次超分辨率算法表现良好,但在面对基于学习的SRCNN超分辨率算法时,在SSIM和SSIM效果都劣于本文所提算法。这些体现了我们算法的可靠性和有效性。

此外,我们还对表3.1和3.2的结果进行了进一步的定性分析,我们在不同的图像超分辨率算法上,展示了均值、双三次和本文算法在Set5数据集上的5张图片上的表现,直观了表现了本文所提的基于梅花均匀图像压缩的图像退化算法的优秀性能。可视化结果如图3.5和3.6所示。

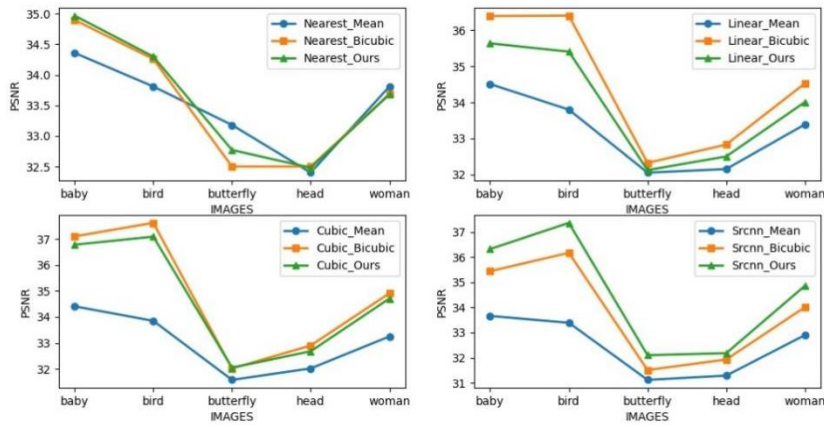


图 3.5 基于不同超分辨率算法的图像退化算法性能比较 (PSNR)

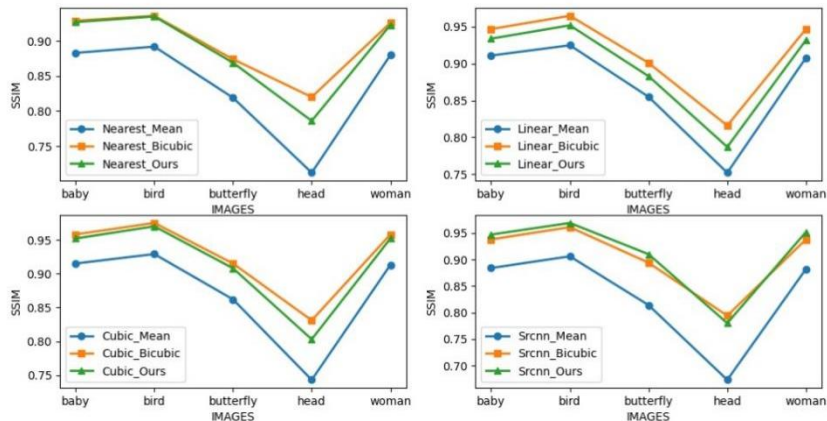


图 3.6 基于不同超分辨率算法的图像退化算法性能比较 (SSIM)

通过可视化的结果,我们可以直接得出,在最近邻超分辨率算法上,本文所提的算法与双三次退化算法平分秋色;在线性和双三次超分辨率算法上,性能处于均值和双三次之间,并且更接近双三次;在SRCNN算法上,优于双三次。综合来看,本文所提的算法广泛适用于各类超分辨率算法。

我们已经知道基于梅花均匀图像压缩的图像退化算法在基于学习的 SRCNN 超分辨算法上取得了比双三次图像退化算法更佳的性能。为进一步研究使用经过我们提出的图像退化算法得到的低分辨率图像训练 SRCNN 是否可以增益学习模型的性能。本小节在放大倍数为 4 的情况下，从 CelebA 数据集分别选取了 3000 张人脸图片，基于不同的图像退化方法生成的低分辨率图像训练模型，进一步验证本文算法的性能，结果如表 3.3 所示。

表 3.3 基于不同的低分辨率输入数据，观测 SRCNN 模型的表现，以比较不同图像退化算法对基于学习的 SRCNN 模型的增益效果。显然，我们提出的图像退化算法无论在 PSNR 指标还是 SSIM 指标上，都使 SRCNN 模型学习到了更好的参数表达，获得了更佳优秀的表现。

表 3.3 基于 SRCNN 的图像退化算法性能比较

图像退化算法	PSNR	SSIM	Loss(MSE)
Bicubic	23.431	0.901	314.816
Ours	25.174	0.924	213.135

此外，为进一步分析不同图像退化算法获取的低分辨率数据对 SRCNN 模型的具体影响，我们绘制了不同退化算法下 SRCNN 模型的损失函数变化曲线图，如图 3.7 所示。SRCNN 在双三次图像退化算法上经过近 50 次迭代，才达到较小的数值，而在梅花均值图像退化算法上收敛速度更快，达到更小的收敛结果。

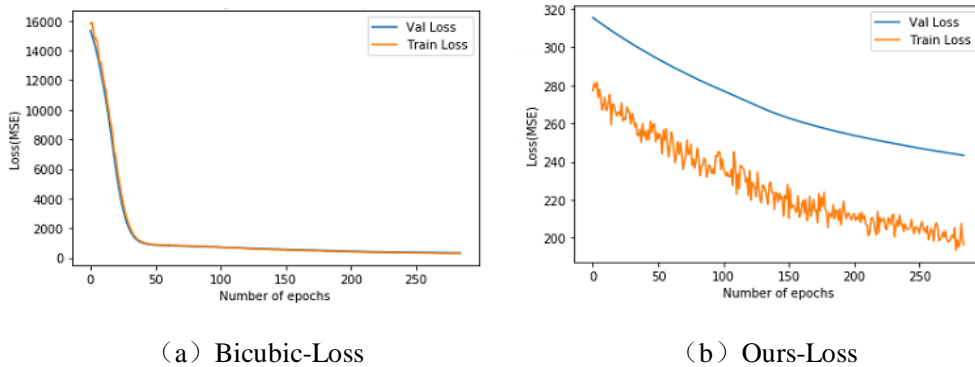


图 3.7 基于 SRCNN 的图像退化算法性能比较

3.4.2 图像超分辨率算法性能比较

3.4.2.1 实验环境

本小节的实验在服务器上开展，设备处理器：Intel®Core™i7-6800K CPU@ 3.40 GHz；设备内存：64GB；设备图形卡：NAVIDIA GeForce GTX 1080 10GB；

设备安装操作系统：CentOS 6。本小节实验代码基于 Python 语言编写，Python 版本：3.6，Pytorch 版本 1.1。

3.4.2.2 实验内容

为验证本文对原始 SRCNN 进行结构约束的 SC-SRCNN 图像超分辨率算法的有效性，我们将在更具挑战的图像 4 倍放大条件下，与图像超分辨率算法 Bicubic 和 SRCNN 在 Set5、Set14、CelebA、LFW 数据集上进行比较分析。由于深度学习模型的性能极易受训练数据集规模的影响，为减少训练时间，本文使用的 SRCNN 和 SC-SRCNN 模型统一由 CelebA 数据集上随机抽取的 3000 张人脸图像经过 300 次迭代训练而来，模型参数方面，学习率被设置为 $1e-5$ ，dropout 被设置为 0.2，批大小被设置为 512。

3.4.2.3 实验分析

图 3.8 中，我们直观展示了经过 Bicubic、SRCNN 以及本文改进的基于结构约束的超分辨率模型 2 倍重建的高分辨率图像与原始高分辨率图像的视觉对比效果。针对重建的图像，Bicubic 存在一定的锯齿现象，SRCNN 比 Bicubic 更清晰，而本文算法较 SRCNN 相比，色彩饱和度更佳，让视觉系统的感觉更真实。

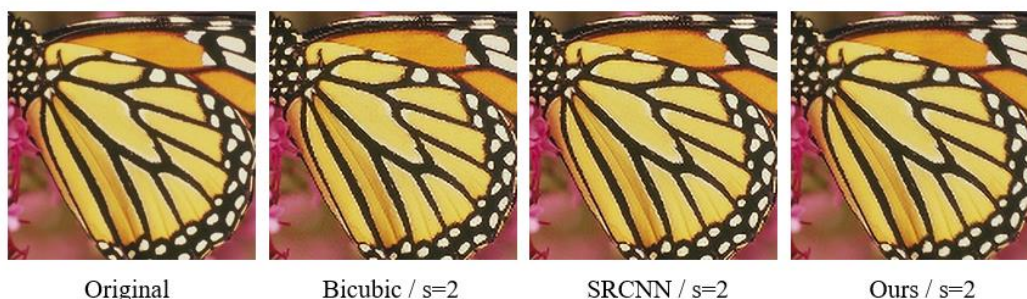


图 3.8 不同超分辨率算法效果展示

此外，基于图像的 4 倍放大，我们使用提出的基于结构约束的超分辨率算法在 Set5、Set14、CelebA、LFW 数据集上进行实验，验证算法在不同风格图像上的重建效果，实验结果如表 3.4 所示。提出的算法在 CelebA 和 LFW 数据集上取得比 SRCNN 更佳的结果，在 SSIM 指标上完全超越了 Bicubic 算法。尽管我们的算法在 Set5 和 Set14 中的 PSNR 的值比 SRCNN 较低，但是在 SSIM 指标上的值却和 SRCNN 一致，这说明我们提出的算法确实能够对生成图像的结构有更强的约束。值得注意的是，在四个数据集上，本文的算法更有稳定性，在自然、动物和人脸类型的图像数据上都达到了重建高分辨率图像与原始高分辨率图像的结构强一致性。

表 3.4 图像超分辨率算法性能比较(s=4)

数据集	Bicubic	SRCNN		Ours	
	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	0.710	20.33	0.809	20.30	0.809
Set14	0.573	18.76	0.702	18.74	0.702
CelebA	0.746	23.43	0.901	23.84	0.942
LFW	0.796	23.04	0.844	23.09	0.845

与此同时，为定性分析本文提出的带结构约束的超分辨率算法的有效性，我们绘制了如图 3.9 的折线图。

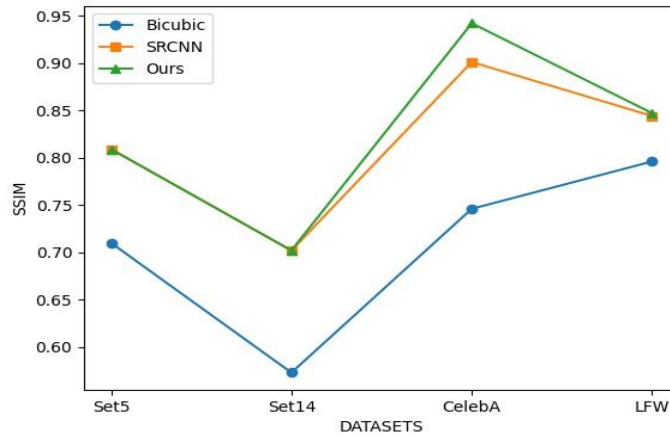


图 3.9 图像超分辨率算法性能比较（SSIM）

3.5 本章小结

本章针对均值图像退化算法在基于高分辨率生成低分辨率图像过程中缺乏对图像梯度信息的保留，导致出现图像细节信息的丢失以及双三次图像退化算法计算复杂度高的问题，结合梅花均匀图像压缩算法，提出了一种新颖的保留图像梯度信息并且计算简单的图像退化算法。针对超分辨率算法 SRCNN 的损失函数缺乏对重建高分辨率图像结构的约束，对其改进，引入了结构相似度量，提出了 SC-SRCNN 超分辨率模型，保证了重建高分辨率图像与原始高分辨率图像结构的一致性。本章首先研究了高分辨率图像生成低分辨率图像的图像退化算法，接着探索了基于低分辨率图像重建高分辨率图像的超分辨率算法，最后通过实验证明了利用图像的梯度特征能够在一定程度上保留低分辨率图像的细节信息，有利于高分辨率图像的重建，以及通过对重建图像结构的约束提高了重建图像的质量。

第4章 基于三支决策的低分辨率人脸的多粒度识别

4.1 引言

低分辨率人脸识别是计算机视觉技术在监控系统应用领域研究的热点,其主要研究路线是先基于超分辨率模型从低分辨率图像重建高分辨率图像,再运用现有高分辨率人脸识别技术,实现对低分辨率人脸的识别。虽然基于深度学习的众多人脸识别模型^[6,57-59]在基于特定数据集的高分辨率人脸识别问题上取得了优秀表现,但直接应用在超分辨率后重建的人脸图像上存在一定挑战,因为重建的图像与真实高分辨率图像相比,不仅缺乏有效信息,而且特征存在差异。除此以外,由低分辨率图像重建高分辨率图像的过程实质上是不断差异化原始图像,增加原始低分辨率图像的细节特征,即使是同一人脸的低分辨率图像可能因为不断差异化,尤其在缩放尺度过大时,造成与原始低分辨率图像相似特征的减弱。为弱化人脸识别模型对高尺度空间图像特征的过分依赖,本文提取了图像低尺度空间的多类别特征,并运用了粒计算和三支决策理论,辅助了低分辨率人脸识别系统的工作,提高系统的识别率。

本章的主要结构如下:4.2 章节设计了基于三支决策的低分辨率人脸识别模型;4.3 章节提取了人脸图像在高分辨率和低分辨率下的特征;4.4 章节对比了本文提出的多粒度三支决策与单粒度二支决策模型,验证了本文所提模型的有效性;4.5 章节对本章的内容进行了总结。

4.2 基于三支决策的多粒度识别模型设计

4.2.1 三支决策理论

三支决策理论是传统二支决策理论的重要推广,由加拿大华人学者姚一豫依据人类认知提出的一种处理不完备信息的决策分析方法。人类在对待某一事物掌握足够信息时,可以直接做出接受或拒绝的判断;而在掌握信息不足时,选择延迟判断,进一步探索补充的信息实现对事物的最终决策。三支决策理论对传统粗糙集理论^[71]引入了贝叶斯风险决策,依据最小化风险代价将对象划分到正域、负域和边界域,形成了接受决策、拒绝决策和延迟决策的三支决策语义^[1]。

决策粗糙集模型是贝叶斯决策理论的一个简单应用，对任意 $X \subset U$ 的子集，可以划分为属于或者不属于集合 X 的两个状态集合 $\Omega = \{X, \sim X\}$ ； $P(X|[x]_{AT})$ 表示在 $[x]_{AT}$ 条件下任意对象属于 X 的条件概率。对应粗糙集模型中的正域（ $POS(X)$ ）、负域（ $NEG(X)$ ）和边界域（ $BND(X)$ ），构造了一个决策动作集 $\tau = \{e_P, e_N, e_B\}$ 。不同的决策会导致不同的分类错误，从而产生不同的风险代价，可以由表 4.1 的代价矩阵定义。其中 λ_{PP} 、 λ_{BP} 、 λ_{NP} 分别代表某一对象属于集合 X 时，采取决策动作 e_P 、 e_B 、 e_N 的损失； λ_{PN} 、 λ_{BN} 、 λ_{NN} 分别代表某一对象不属于集合 X 时，采取决策动作 e_P 、 e_B 、 e_N 的损失^[1]。

表 4.1 决策粗糙集中的代价矩阵^[1]

动作	X	$\sim X$
e_P	λ_{PP}	λ_{PN}
e_B	λ_{BP}	λ_{BN}
e_N	λ_{NP}	λ_{NN}

Yao 对决策粗糙集模型中的代价矩阵做出进一步假设：

$$\lambda_{PP} \leq \lambda_{BP} \leq \lambda_{NP}, \lambda_{PN} \leq \lambda_{BN} \leq \lambda_{NN} \quad (4.1)$$

$$(\lambda_{NP} - \lambda_{BP}) \times (\lambda_{PN} - \lambda_{BN}) > (\lambda_{BP} - \lambda_{NP}) \times (\lambda_{BN} - \lambda_{NN}) \quad (4.2)$$

基于上述假设以及贝叶斯决策理论，形成了决策粗糙集中的三支决策，

(P)：如果 $P(X|[x]_{AT}) \geq \alpha$ ，则选择 $x \in POS(X)$ ；

(B)：如果 $\beta < P(X|[x]_{AT}) < \alpha$ ，则选择 $x \in BND(X)$ ；

(N)：如果 $P(X|[x]_{AT}) \leq \beta$ ，则选择 $x \in NEG(X)$ ；

其中参数 α 和 β 定义如下^[1]。

$$\alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \quad (4.3)$$

$$\beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \quad (4.4)$$

4.2.2 基于三支决策的多粒度识别模型

基于粒计算和三支决策的多粒度方法在文本情感分类领域已经得到应用。Zhang 等^[60]将文本数据的词、句、篇章三层粒结构融合，提出了具有强解释性的 SSS 特征，并结合三支决策将分类对象划分为三个区域，对延迟决策区域进一步使用 SVM 算法分类，提高了分类的准确性。

本文在上述研究的基础上,提出了图 4.1 的基于三支决策的低分辨率人脸的多粒度识别模型。该模型可将人脸图像信息划分为两个尺度空间,在高尺度空间提取了高分辨率图像深度视觉特征,并基于三支决策理论进行粗分类;在粗分类的基础上,基于低尺度空间提取了低分辨率图像的鲁棒、显著特征再进行细分类。针对直接运用基于超分辨率算法重建的高分辨率图像提取的深度视觉特征可能存在扩大人脸区域细节差异,造成错误,故引入了延迟决策域,进一步借助图像低尺度空间低分辨率图像的相似性特征降低差异,从而确定最终决策结果,以提高高分辨率人脸识别算法直接应用在低分辨率人脸识别问题上的准确率。

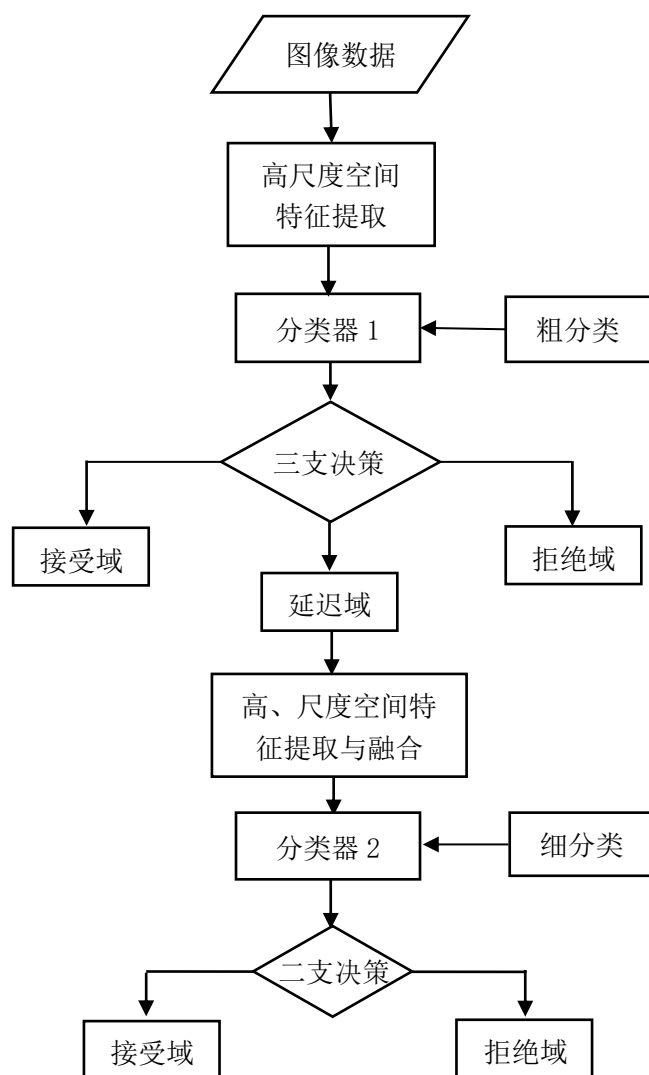


图 4.1 基于三支决策的多粒度识别模型

4.3 人脸图像特征提取

基于神经网络提取的深度视觉特征在针对高分辨率人脸识别的问题上已经取得了丰硕成果，但有效地针对低分辨率大规模的人脸验证和识别仍具有一定挑战。面对基于重建的低分辨率人脸的识别方法，由于重建过程差异化的加入，除在高尺度空间提取的高分辨率图像面部区域的特征外，仍需提取低尺度空间低分辨率图像相似的鲁棒特征，形成对输入低分辨率人脸图像更全面的表示。依据提出的基于三支决策的多粒度识别模型，三支粗分类中高尺度空间的深度视觉图像特征可以选取 FaceNet^[57]等深度学习模型提取的特征；在细分类中可以提取低尺度空间中低分辨率图像的 LBP、ORB 以及局部显著块等特征。

4.3.1 高分辨率图像特征提取

FaceNet^[57]高分辨率人脸识别模型与其他深度学习模型相比，并未采用 softmax 函数进行分类学习，而是直接学习了人脸图像到欧几里得空间的映射，用一个 128 维的 Embedding 特征向量表示人脸图像，特征向量间的距离代表着相似程度，结构下图所示。



图 4.2 FaceNet 网络结构^[58]

模型前部分是传统卷积神经网络，并在 L2 范数前实现了特征的归一化处理，构建了 Embedding 向量空间；L2 范式后，是一个三重损失函数，该损失函数并非传统的单个或者双个损失，解释如图 4.3。



图 4.3 Triplet loss^[58]

学习前，Anchor、Negative、Positive 三个样本距离都较近，认为都相似；而实际上 Anchor 应该离 Negative 更远，与 Positive 更近，故需要学习最小化 Anchor 与 Positive 的距离，最大化 Anchor 与 Negative 的距离。

FaceNet 模型提取的 Embedding 向量基于欧式距离度量人脸相似度，并给定了单一距离阈值 $d = 1.1$ 进行二支判别，识别流程如图 4.4 所示。

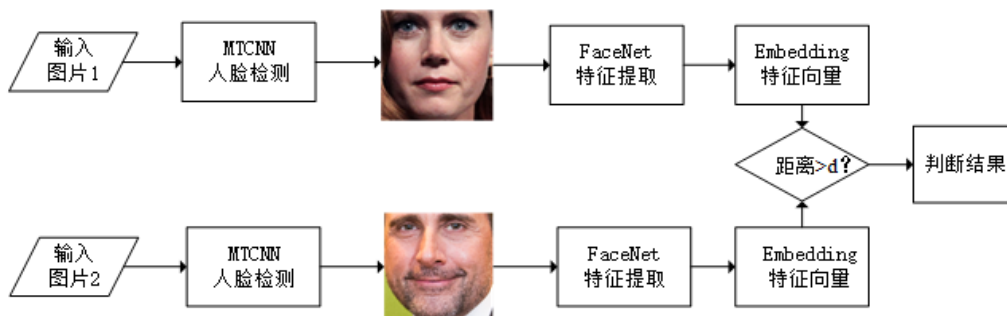


图 4.4 FaceNet 的人脸识别流程图

FaceNet 先用人脸检测与对齐算法（MTCNN^[61]）进行面部检测，再针对检测的面部区域提取特征。归一化的 Embedding 特征向量被映射在同一向量空间，并基于 N 维空间上的欧氏距离实现了人脸的相似性的计算，如公式(4.5)所示，其中 x, y 分别是两幅图像 Embedding 特征向量， $x = \{x_1, x_1, \dots, x_n\}$ ， $y = \{y_1, y, \dots, y_n\}$ ， $n=128$ 。

$$D_1(x, y) = \sqrt[n]{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (4.5)$$

4.3.2 低分辨率图像特征提取

4.3.2.1 基于粗糙不确定分割提取的局部显著块特征

现实生活中，人们经常通过一些局部显著特征来寻找或辨识一个人，而这些局部显著区域具有较强颜色、纹理等辨别特征。针对监控视频下低分辨率人脸识别问题，除面部特征外，行人上身的颜色特征具有强辨识度，若越多的图像块颜色直方图特征与某一图像块 A 越相似，则 A 越具有代表性和显著性。K-Means 算法是常用的数据聚类方法，但该算法孤立地看待每个样本数据，未考虑样本的邻域信息。本章基于邻域粗糙集模型^[62]在 K 均值算法初聚类基础上进行改进，提出了一种基于粗糙不确定分割的图像局部显著块的检测算法。

粗糙集理论在图像分割领域早已得到应用，Akash^[63]最早将模糊 C 均值的方法与经典粗糙集的上、下近似集结合，用于解决现实场景图片部分区域颜色渐变的分割问题。Dariusz^[64]等运用粗糙集模型中的粗糙熵优化了图像模糊边界分割。在上述研究基础上，我们提出了如下基于粗糙不确定分割的局部显著块检测算法，算法中涉及的相关邻域粗糙集模型的部分术语如定义 1-4 所示。

算法 1. 基于粗糙不确定分割的局部显著块检测

输入：数据集 U ，聚类中心 $C = \{c_1, c_2, \dots, c_k\}$ ，度量函数 D ，邻域半径 δ ，
局部显著块大小 $s * s$ ；

输出：每个样本的局部显著块。

1. 计算出 K 个聚类中心的上、下近似的差集 S_{out} ；
2. 根据度量函数和邻域半径，得出 S_{out} 每个样本邻域粒子 $\delta(x_i)$ ；
3. 类别 c_j 对样本 x_i 的包含度： $c(i, j) = \delta(x_i) \cap \delta(c_j) / \delta(x_i)$ ；
4. 根据包含程度重新划分 S_{out} 集合中的样本；
5. 依据最终分割结果，获取 K 个局部显著块。

数据集 $U = \{x_1, x_2, \dots, x_n\}$ 为图像中的每一像素点，可依据像素点灰度值、纹理及结构等特征分割图像，得到不同的分割 $U = \{X_1, X_2, \dots, X_k\}$ 。依据定义 1，对 U 中的样本粒化，得到对应的邻域信息粒子 $\{\delta(x_1), \delta(x_2), \dots, \delta(x_n)\}$ 。

定义 1. 邻域粒化^[62]：在 N 维实数空间上，给定空间上的非空有限集合 $U = \{x_1, x_2, \dots, x_n\}$ ，度量函数 D 以及邻域因子 δ ，对论域 U 上的任意对象 x_i 粒化，得到基本邻域信息粒子 $\delta(x_i)$ 如公式(4.6)所示，度量函数 D 满足交换定律。

$$\delta(x_i) = \{x | D(x, x_i) \leq \delta, x \in U\} \quad (4.6)$$

定义 2. 邻域关系矩阵^[62]：依据定义 1 给出邻域关系矩阵，对论域 U 上的 x_i 和 x_j ，若 $x_i \in \delta(x_j)$ ，则认为 x_i 在 x_j 的邻域内，定义 $d_{ij} = 1$ ，反之 $d_{ij} = 0$ 。由于度量函数 D 符合交换量，故 $d_{ij} = d_{ji}$ ，由此构成邻域关系矩阵 $R = (d_{ij})_{n \times n}$ 。

定义 3. 包含度^[62]：设论域上的子集 X_1 和 X_2 ， X_1 包含于 X_2 的程度如下：

$$C(X_1, X_2) = \frac{|X_1 \cap X_2|}{|X_1|}, X_1, X_2 \neq \emptyset \quad (4.7)$$

定义 4. 变精度邻域上、下近似空间^[62]：在邻域近似空间 $S = \langle U, R \rangle$ 上， U 为论域， R 为邻域关系矩阵， X 的下、上近似如公式(4.8)， $c \in [0.5, 1)$ 。

$$\begin{cases} RX = \{x_i | C(\delta(x_i), X) \geq c, x_i \in U\} \\ R\bar{X} = \{x_i | C(\delta(x_i), X) \geq 0, x_i \in U\} \end{cases} \quad (4.8)$$

我们可以对 RGB 三通道进行划分，将每一通道划分为 8 段，得到 $8*8*8=512$ 个颜色子空间，任一颜色必然落在 512 个子空间中。基于算法 1 提出的基于不确定分割检测的 K 个局部显著块，提取显著颜色特征，若两张图片总的颜色距离大于 K ，则认为两张图片颜色不一致，距离计算公式如下，其中 C_{1k} 和 C_{2k} 为图片 I_1 和 I_2 的第 k 个显著块的颜色。

$$D_2(I_1, I_2) = \sum_{i=1}^K C_{1k} - \sum_{i=1}^K C_{2k} \quad (4.9)$$

4.3.2.2 LBP 特征

LBP (Local Binary Patterns) 是一种提取图像纹理特征效果好且计算简单的算法, 因其具有光照不变性、旋转不变性等显著优点在计算机视觉领域被广泛应用, 并在人脸图像分析领域获得了成果。经典的 LBP 算法以 3×3 的图像块为单位, 将 8 邻域像素点灰度值与中心点比对, 若大于标记为 1, 反之标记为 0, 从而得到一串二进制的特征值, 换算为十进制, 取值范围 $[0, 255]$ 。本文通过将人脸图像划分为多个子区域, 在每个子区域利用 3×3 的滑动窗口, 提取 LBP 特征。而在人脸图像上我们通常采用统计子区域 LBP 特征直方图作为特征描述符, 度量 LBP 特征直方图相似度的方法有相关性比较、余弦距离和巴氏距离等。本文将多个子区域 LBP 特征拼接形成 LBP 特征向量, 采用余弦距离实现相似度量, 若 L_1 和 L_2 是图片 I_1 和 I_2 的 LBP 特征向量, 采用余弦距离计算的相似度如下式。

$$D_3(I_1, I_2) = \frac{L_1^T \cdot L_2}{\|L_1\| \times \|L_2\|} \quad (4.10)$$

4.3.2.3 ORB 特征

ORB (Oriented Fast and Rotated Brief) 是一种快速检测关键点并创建特征向量的算法, 在计算速度上优于 SIFT 和 SURF。该算法提取的特征对尺度、光照、旋转具有良好的不变性, 可以鲁棒的表示图像。ORB 特征提取由两个部分构成: 采用 FAST 算法检测关键点, 例如角点, 亮度梯度大的点等; 用 BRIEF 算法描述 FAST 检测到的特征点。本文通过 ORB 算法生成了人脸图像的特征描述子, 并通过 BF (Brute Force) 暴力匹配算法进行关键点匹配。

ORB 算法实现了两张图片的关键点检测和匹配, 假设图片 I_1 和 I_2 的检测的关键点个数分别为 N_1 和 N_2 , 通过 BF 暴力匹配算法得到匹配的关键点个数 M ($M \leq \min(N_1, N_2)$), 则关于 ORB 特征的度量函数可由下式定义。

$$D_4(I_1, I_2) = \frac{M}{\min(N_1, N_2)} \quad (4.11)$$

4.3.3 特征融合

对于低尺度空间的低分辨率人脸图像, 我们提取了图像的 LBP、ORB 和局部显著块颜色特征。为实现这些特征的融合, 我们选择了投票法, 如公式(4.12), 其中 $d_1(I_1, I_2)$ 、 $d_2(I_1, I_2)$ 、 $d_3(I_1, I_2)$ 、 $d_4(I_1, I_2)$ 分别代表 Embedding 特征、显著块特征、LBP 特征以及 ORB 特征的决策。若最终 $\text{dist}(I_1, I_2)$ 的值大于等于 3, 则识别 I_1 与 I_2 是同一人, 反之则不是, 决策如公式(4.17), 其中 μ_1 和 μ_2 是针对不同特征度量函数 $\text{dist}(I_1, I_2)$ 给定的阈值, 而 K 是显著块的数量。

$$dist(I_1, I_2) = d_1(I_1, I_2) + d_2(I_1, I_2) + d_3(I_1, I_2) + d_4(I_1, I_2) \quad (4.12)$$

$$d_1(I_1, I_2) = \begin{cases} 0, & D_1(I_1, I_2) \geq 1.1 \\ 1, & D_1(I_1, I_2) < 1.1 \end{cases} \quad (4.13)$$

$$d_2(I_1, I_2) = \begin{cases} 0, & D_2(I_1, I_2) \geq K \\ 1, & D_2(I_1, I_2) < K \end{cases} \quad (4.14)$$

$$d_3(I_1, I_2) = \begin{cases} 1, & D_3(I_1, I_2) \geq \mu_1 \\ 0, & D_3(I_1, I_2) < \mu_1 \end{cases} \quad (4.15)$$

$$d_4(I_1, I_2) = \begin{cases} 1, & D_4(I_1, I_2) \geq \mu_2 \\ 0, & D_4(I_1, I_2) < \mu_2 \end{cases} \quad (4.16)$$

$$\text{决策结果} = \begin{cases} \text{接受}, & dist(I_1, I_2) \geq 3 \\ \text{拒绝}, & dist(I_1, I_2) < 3 \end{cases} \quad (4.17)$$

4.4 实验与分析

4.4.1 实验环境

本小节的实验的硬件环境是 Intel®Core™i7-6800K CPU@ 3.40 GHz 处理器；内存 64GB；NAVIDIA GeForce GTX 1080 图形卡，显存 10GB；软件环境为 CentOS 6；编程环境是 Python3.6，opencv4，Pytorch1.1，Tensorflow1.7。

4.4.2 评价指标

在机器学习中用来评价模型性能的指标有许多，常用的几项有 TP（真正例）、TN（真负例）、FP（假负例）、FN（假正例），以及基于 TP、TN、FP、FN 计算的准确率（Accuracy）、召回率（Recall）、精确率（Precision）。准确率代表了模型计算准确的样本与总样本的比值；召回率是模型判断正例正确的样本与所有正例样本的比值；精确率是真正例占模型判断的所有正例的比例，各指标计算公式如下所示。

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.18)$$

$$Recall = \frac{TP}{TP+FN} \quad (4.19)$$

$$Precision = \frac{TP}{TP+FP} \quad (4.20)$$

4.4.3 实验内容

基于本章提出的基于三支决策的低分辨率人脸的多粒度识别模型，我们在 CelebA 和 LFW 数据集上开展了实验。并将三支模型和二支模型的性能进行了比较。本文所提的三支多粒度模型先在高分辨率图像上获取的深度视觉特征，进行一次粗分类，分为接受、拒绝和延迟；再对于延迟域中的样本，模型将高分辨率图像和低分辨率图像下获取的高低尺度信息进行融合，进行二次细分类，实现最终的决策。实验中，测试模型有效性的数据集介绍如下：

CelebA：该数据集是由香港中文大学通过网络收集的大规模人脸数据集，图像尺寸为 208*178 像素，人脸面部占整体 1/3 以上。为应用于低分辨率人脸识别问题，我们现将图片裁剪归一到 144*144 像素，先通过双三次图像退化生成 36*36 像素的低分辨率图像，再基于超分辨率算法得到 4 倍重建的高分辨率图像。实验中随机筛选了 100 个人的 1999 张图像。

LFW：该数据集为无限制环境下人脸识别问题设计，原始图片尺寸为 250*250 像素，为用于低分辨率人脸问题的识别，我们将图片归一化到 144*144 像素，作为高尺度空间下的高分辨率图像，而对应低尺度空间的低分辨率图像为 36*36 像素。我们选取了该数据集中 100 个人，共 180 张照片用来测试模型。

FaceNet 预训练模型版本为 20170512-110547，基于微软人脸识别数据集（MS-Celeb-1M）训练得到。

基于三支决策的低分辨率人脸的多粒度模型的阈值选定：依据三支决策的定义，延迟决策域的参数(α , β)应当由决策粗糙集中的代价矩阵计算而来，但本文所提的模型具有现实的背景，(α , β)的值可以依据对应的高分辨率人脸识别系统给出。实验中我们给定了不同的阈值，以便观测所提模型的有效性和稳定性。

4.4.4 实验分析

传统的二支决策模型假设当前已知的信息完全可以对未知对象实现划分，但该假设极大不切实际地简化了现实生活中极具复杂场景下的低分辨率人脸识别问题。为了将高分辨率人脸识别算法运用于现实场景，我们引入了模仿人类思维方式的三支决策理论，而三支决策的思想早已被成功运用在情感分析、图像分类等人工智能领域，在面对模型当前已知信息不足，无法直接判断的对象类别的情况下，我们选择延迟判断，尝试寻找其他补充信息实现最终的决策。基于三支决策的低分辨率人脸的多粒度识别模型将图像高、低尺度的一致性和互补性信息融入到了决策的过程，更加符合人类的推理判断的思维。

基于上述原因,对于低分辨率人脸的识别,我们在广泛使用的人脸数据集上开展了实验,并对本章提出模型的性能进行了评估,实验结果展示在表 4.2 和 4.3。

表 4.2 和 4.3 从多项指标比较了本文提出的模型与主流的 FaceNet 人脸识别模型,为区别我们的方法,我们把主流的 FaceNet 模型看作为单粒度二支决策类型。表 4.2 比较了基于高尺度空间高分辨率图像的深度视觉特征进行单粒度二支决策和基于高尺度空间和低尺度空间高、低分辨率图像特征融合度量的多粒度三支决策在 CelebA 和 LFW 数据集上的结果。表 4.3 基于表 4.2 的人脸验证结果,依据模型的准确率、召回率、精确度指标比较了本章所提的多粒度三支模型与单粒度二支模型的性能。该结果表明,我们提出的模型在 CelebA 和 LFW 数据集上获得了更高的识别准确率和精确率。我们认为,良好的识别准确率来源于我们提出的多粒度三支模型。

表 4.2 不同双阈值下单粒度二支模型与多粒度三支模型比较

数据集	指标	单粒度二支决策	多粒度三支决策(Ours)		
		d=1.1	(1.08, 1.12)	(1.07, 1.13)	(1.05, 1.15)
CelebA	TP	17641	16781	16357	15541
	TN	1952201	1957654	1959854	1963560
	FP	21316	15863	13663	9957
	FN	5843	6703	7127	7943
LFW	TP	194	193	191	190
	TN	14715	14759	14781	14816
	FP	1101	1057	1035	1000
	FN	100	101	103	104

表 4.3 不同双阈值下单粒度二支模型与多粒度三支模型性能比较

数据集	指标	单粒度二支决策	多粒度三支决策(Ours)		
		d=1.1	(1.08, 1.12)	(1.07, 1.13)	(1.05, 1.15)
CelebA	Acc	0.9864	0.9887	0.9895	0.9910
	Rec	0.7512	0.7145	0.6965	0.6618
	Pre	0.4528	0.5140	0.5449	0.6095
LFW	Acc	0.9254	0.9281	0.9294	0.9315
	Rec	0.6598	0.6565	0.6497	0.6462
	Pre	0.1498	0.1544	0.1558	0.1596

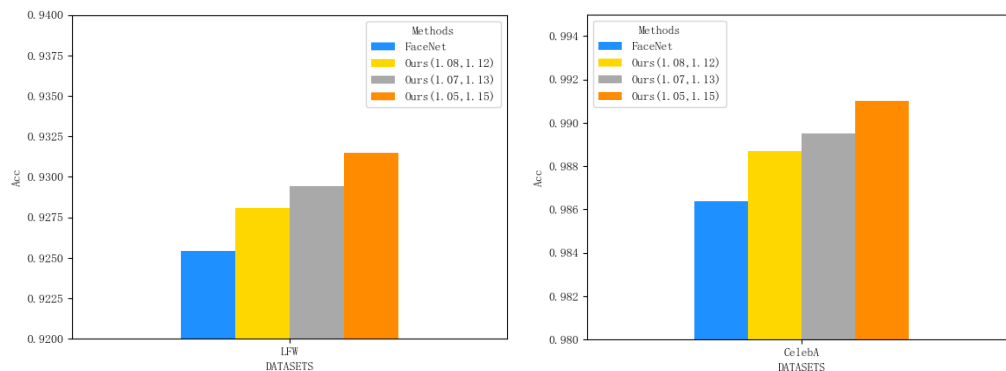


图 4.5 单粒度二支与不同阈值下的多粒度三支模型识别准确率比较

为更直观的展示本文所提的基于三支决策的低分辨率多粒度识别模型的优越性，我们绘制了图 4.5 以定性分析模型的识别准确率。值得注意的是，阈值区间的扩大，基于三支决策的多粒度识别模型的识别准确率和精确率在不断上升；尽管如此，但是我们不能无限的扩大阈值的区间，因为随着阈值区间的不断扩大，模型的召回率在降低，所以在不断提高准确率和精确率的同时，也要权衡召回率的变化，选择适中的阈值区间。

4.5 本章小结

针对直接基于重建的高分辨率人脸图像运用高分辨率人脸识别系统，由于重建图像带来的差异性，造成识别准确率降低的问题，本章基于三支决策和粒计算理论，综合图像高、低尺度空间信息的一致性和互补性，提出了一种基于三支决策的低分辨率人脸的多粒度识别模型，并在 CelebA 和 LFW 数据集上进行了实验，表明三支多粒度模型优于传统的二支单粒度模型。此外，本文还提出了一种基于粗糙不确定分割的图像显著区域检测算法。

第5章 总结与展望

5.1 总结

近年来,“智慧城市”建设的持续投入,使得视频监控系统和人脸、车辆采集设备能够全方位部署,产生了海量的视频图像数据。而现实场景的复杂性,例如遮挡、运动、光照变化、距离远近等因素,都给低分辨率人脸识别任务的实现带来了重重困难。

本文的研究工作回顾了人脸识别技术的发展历程、低分辨率人脸识别的研究现状,介绍了与低分辨率识别相关的技术和数据集,总结了图像退化与超分辨率方法,提出一种基于梅花均匀图像压缩的图像退化算法;改进了 SRCNN 模型的损失函数;提出了基于三支决策的低分辨率人脸的多粒度识别模型。主要工作如下:

1. 提出了一种保留图像梯度信息的基于梅花均匀图像压缩的图像退化算法。图像退化算法实现了高分辨率图像向低分辨率图像的转化,针对现有常用图像退化方法缺乏对图像梯度等信息的保留以及计算复杂的问题,我们借鉴了梅花均匀图像压缩算法对图像梯度信息的有效保留,在均值图像退化算法的基础上,提出了一种计算简单快速,并且能够有效保留图像梯度信息的图像退化方法。为了验证本文所提算法的有效性,我们在 Set5 数据集上基于 PSNR 和 SSIM 指标评估算法,通过与主流图像退化算法相比,我们的算法性能表现优秀。

2. 提出了一种结构约束的超分辨率模型。针对原始 SRCNN 模型的损失函数在学习过程中,单纯使用图像的像素值约束重建的图像,这使得重建的图像在 PSNR 指标上较高,但却缺乏真实性,我们受图像结构相似性指标启发,对原始 SRCNN 的损失函数进行改进,约束了重建高分辨率图像与原始高分辨率图像结构的一致,使得重建的图像更真实。并在 Set5、Set14、LFW 和 CelebA 四个公开数据集上进行了实验。

3. 提出了一种基于三支决策的低分辨率人脸的多粒度识别模型。我们基于图像高尺度空间和低尺度空间的信息的一致性和互补性,在高尺度图像特征无法判断当前样本的类别时,给予低尺度空间的信息补充,实现人脸更准确的判断。此外,在低尺度空间图像特征中,我们还提出了一种基于粗糙不确定分割的图像显著块检测算法,并基于此算法实现了图像显著块特征的提取。并在 LFW 和 CelebA 数据集上进行了实验,本文所提的模型提高了人脸验证的准确率。

5.2 展望

图像超分辨率技术推动了低分辨率人脸识别任务的发展,但现阶段仍无法广泛应用于现实场景。在上海、北京、广州、深圳等人口密集城市,监控视频采集的照片常被遮挡,并且存在光照、视角、运动、表情变化等问题,相比高分辨率人脸识别,该领域仍有许多问题等待解决。针对本文在低分辨率人脸识别任务上的工作,提出了如下改进方向:

1. 虽然本文考虑到对 SRCNN 的损失函数进行改进,对重建图像的结构进行约束,但是基于深度学习的超分辨率模型仍然存在训练时间长、参数繁多等实际问题,网络参数压缩是优化模型的另一个方向。
2. 虽然本文基于一种结构约束的超分辨率模型重建了良好的高分辨率图像,但是没有针对人脸特定的域,未来工作将研究如何融入人脸的先验知识,例如使用哪些先验知识、在哪一阶段融合先验以及各先验知识的重要程度等。

综上所述,低分辨率人脸识别任务仍具挑战,需要进一步探索,不断促进该领域的发展,使其最终应用于现实。

致谢

岁月匆匆，时光如虹。想当年，意气风发，怀揣理想，踏入秀丽多姿的校园；看今朝，基础扎实，满怀信心，挥别温情脉脉的母校。感谢同济大学，在这两年半的研究生学习生活中，让我收获了系统的专业知识，前沿的科学研究经验，导师的谆谆教诲，师兄师姐的研究建议，师弟师妹的欢声笑语，和蔼的食堂大叔，可爱的宿管阿姨，还有其他太多美好而珍贵的记忆。

首先要感谢我敬爱的导师——苗夺谦教授。苗老师治学严谨，学术成果丰厚，教学经验丰富，对待学生和藹可亲。他有着绅士一般的风度，谦逊温和的态度还有儒雅的气度。无论在学术研究上，还是在日常生活中，他都是我心中敬仰的榜样。感谢苗老师为我提供了良好的交流学习平台，多次邀请国内国外专家学者来学校和实验室交流，在这些交流中，有迸发的灵感，有开阔的视野；感谢苗老师让我有机会参加了 2019 年中国粒计算与知识发现学术会议，让我接触到了该领域的学术牛人和发展情况；感谢苗老师给我提供实践的机会，参与到工程项目中，不断提高自己的实践能力，将知识与工程联系在一起。

其次要感谢我可爱的同门们。与师兄师姐和师弟师妹的交流中，让我学习到了很多，由衷的感谢师兄师姐在学习、科研以及生活上对自己的帮助，让我更快地找到了自己的方向。同时，感谢已就业的师兄师姐在职业发展上给予的建议。感谢师弟师妹们，让我的科研生活增添了许多乐趣。同时，也感谢我的同学和朋友们，我们一起上课学习、准备春招、暑期实习、准备秋招、撰写毕业论文。很幸运，这一路我们能一同前行！

最后要感谢我亲爱的家人们，是你们一直在我背后默默地支持，在我成功时，不吝啬掌声与夸奖；在我失败，不缺少安慰和鼓励。让我可以一直带着笑容和勇气，面对挑战，不断前进！

2020 年 5 月

参考文献

- [1] 张燕平, 姚一豫, 苗夺谦, 等. 粒计算、商空间及三支决策的回顾与发展. 北京: 科学出版社, 2017.
- [2] Pawlak Z. Rough sets. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341-356.
- [3] 鞠恒荣, 杨习贝, 于化龙, 等. 决策粗糙集的属性约简准则研究. *南京师大学报(自然科学版)*, 2015, 38(1): 41-47.
- [4] Liu C, Wechsler H. Gabor feature-based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 2002, 11(4): 467-476.
- [5] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037-2041.
- [6] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012, 1:1097-1105.
- [7] Taigman Y, Yang M, Ranzato M A, et al. DeepFace: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 1701-1708.
- [8] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Proceedings of the European Conference on Computer Vision*, 2008.
- [9] Choi J Y, Ro Y M, Platanotis K N. Color face recognition for degraded face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B(Cybernetics)*, 2009, 39(5):1217-1230.
- [10] Lei Z, Ahonen T, Pietikäinen M, et al. Local frequency descriptor for low-resolution face recognition. *Proceedings of the International Conference on Face and Gesture*, 2011: 161-166.
- [11] Li G, Sun G, Zhang X. Robust face recognition in low resolution and blurred image using joint information in space and frequency. *IFIP International Conference on Network and Parallel Computing*. Springer Berlin Heidelberg, 2012.
- [12] Lu T, Yang W, Zhang Y D, et al. Very low-resolution face recognition via semi-coupled locality-constrained representation. *IEEE International Conference on Parallel & Distributed Systems*. IEEE, 2017.
- [13] Li B, Chang H, Shan S, et al. Low-resolution face recognition via coupled locality preserving mappings. *IEEE Signal Processing Letters*, 2010, 17(1):20-23.
- [14] Jiang J, Hu R, Han Z, et al. Coupled discriminant multi-manifold analysis with application to low-resolution face recognition. *International Conference on Multimedia Modeling*. Springer International Publishing, 2015.
- [15] Zhou C T, Zhang Z W, Yi D, et al. Low-resolution face recognition via Simultaneous

- Discriminant Analysis. 10.1109/IJCB.2011.6117595.
- [16] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010, 19(11): 2861-2873.
 - [17] Yang J, Wright J, Huang T S, et al. Image super-resolution as sparse representation of raw image patches. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.
 - [18] Bilgazyev E, Efraty B A, Shah S K, et al. Sparse representation-based super resolution for face recognition at a distance. *Proceedings of the British Machine Vision Conference*, 2011, 52: 1-11.
 - [19] Shekhar S, Patel V M, Chellappa R. Synthesis-based recognition of low resolution faces. *Proceedings of the International Joint Conference on Biometrics*, 2011: 1-6.
 - [20] Kim K I, Kwon Y. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(6): 1127-1133.
 - [21] He H, Siu W C. Single image super-resolution using Gaussian process regression. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011: 449-456.
 - [22] Peleg T, Elad M. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE Transactions on Image Processing*, 2014, 23(6): 2569-2582.
 - [23] Yang C Y, Ma C, Yang M H. Single-image super-resolution: A benchmark. *European Conference on Computer Vision*, 2014: 372-386.
 - [24] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 3262-3271.
 - [25] Keys R G. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Signal Processing*, 1981, 29(6).
 - [26] Chen C Y, Guo S M, Tsai C H, et al. Integer wavelet-based image interpolation in lifting structure for image resolution enhancement. *Applied Mathematics*, 2018, 9(10): 1156.
 - [27] Freedman G, Fattal R. Image and video upscaling from local self-examples. *ACM Transactions on Graphics*, 2011, 30(2): 12.
 - [28] Sun J, Xu Z, Shum H Y. Image super-resolution using gradient profile prior. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.
 - [29] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.
 - [30] Daniel Glasner, Shai Bagon, Michal Irani. Super-resolution from a single image. *Proceedings of the IEEE International Conference on Computer Vision*, 2009.
 - [31] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(2): 295-307.
 - [32] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1637-1645.
 - [33] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep

- convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1646-1654.
- [34] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681-4690.
- [35] Wang X. et al. ESRGAN: Enhanced super-resolution generative adversarial networks. Proceedings of the European Conference on Computer Vision, 2018.
- [36] Zhang W, Liu Y, Dong C, et al. RANKSRGAN: Generative adversarial networks with ranker for image super-resolution. Proceedings of the IEEE International Conference on Computer Vision, 2019: 3096-3105.
- [37] Zhou E, Fan H, Cao Z, et al. Learning face hallucination in the wild. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
- [38] Zhang K, Zhang Z, Cheng C W, et al. Super-identity convolutional neural network for face hallucination. Proceedings of the European Conference on Computer Vision, 2018.
- [39] Yu X, Porikli F. Ultra-resolving face images by discriminative generative networks. Proceedings of the European Conference on Computer Vision, 2016: 318-333.
- [40] Song Y, Zhang J, He S, et al. Learning to hallucinate face images via component generation and enhancement. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main Track, 2017: 4537-4543.
- [41] Zhu S, Liu S, Loy C C, et al. Deep cascaded bi-network for face hallucination. Proceedings of the European Conference on Computer Vision, 2016: 614-630.
- [42] Bulat A, Tzimiropoulos G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 109-117.
- [43] Yu X, Fernando B, Ghanem B, et al. Face super-resolution guided by facial component heatmaps. Proceedings of the European Conference on Computer Vision, 2018: 217-233.
- [44] Chen Y, Tai Y, Liu X, et al. FSRNet: End-to-end learning face super-resolution with facial priors. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2492-2501.
- [45] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [46] Bevilacqua, M., Roumy, A., Guillemot, C., Morel, M.L.A.: Low-complexity single image super-resolution based on nonnegative neighbor embedding. In: BMVC (2012).
- [47] Zeyde R, Elad M, Protter M. Curves and surfaces. Berlin Heidelberg: Springer, 2012: 711-730.
- [48] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild. Proceedings of the IEEE International Conference on Computer Vision, 2015: 3730-3738.
- [49] Chrabaszcz P, Loshchilov I, Hutter F. A downsampled variant of ImageNet as an alternative to the cifar datasets, arXiv preprint arXiv:1707.08819, 2017.
- [50] Poulos Z, Nouri A, Moshovos A. Training CNNs faster with dynamic input and kernel downsampling, arXiv preprint arXiv:1910.06548, 2019.
- [51] Zhang K, Zuo W, Zhang L. Deep plug-and-play super-resolution for arbitrary blur kernels. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019:

- 1671-1681.
- [52] 朱旭. 新型多描述图像编码算法研究. 四川: 电子科技大学, 2019.
- [53] Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: ICML. pp. 807-814, 2010.
- [54] G. Youm, S. Bae and M. Kim, Image super-resolution based on convolution neural networks using multi-channel input. 2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Bordeaux, 2016, pp. 1-5.
- [55] Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 1874—1883.
- [56] 刘月峰, 杨涵晰, 蔡爽, 等. 基于改进卷积神经网络的单幅图像超分辨率重建方法. 计算机应用, 2019, 39(05): 1440-1447.
- [57] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [58] Zhou E, Cao Z, Yin Q. Naive-deep face recognition: touching the limit of LFW benchmark or not?. Computer ence, 2015.
- [59] Sun Y, Liang D, Wang X, et al. Deepid3: Face recognition with very deep neural networks. arXiv preprint arXiv:1502.00873, 2015.
- [60] 张越兵, 苗夺谦, 张志飞. 基于三支决策的多粒度文本情感分类模型. 计算机科学, 2017, 44(12): 188-193.
- [61] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [62] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简. 软件学报, 2008, 19(3): 640-649.
- [63] Mohabey A, Ray A K. Fusion of rough set theoretic approximations and FCM for color image segmentation. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 2000, 2: 1529-1534.
- [64] Małyszko D, Stepaniuk J. Adaptive multilevel rough entropy evolutionary thresholding. Information Sciences, 2010, 180(7): 1138-1158.

个人简历、在读期间发表的学术论文与研究成果

个人简历

饶梦，女，1994 年 10 月生。

2017 年 7 月毕业于上海海事大学 网络工程专业 获工学学士学位。

2017 年 9 月入同济大学读硕士研究生。

已录用论文

[1] 饶梦, 苗夺谦, 罗晟. 一种粗糙不确定的图像分割方法. 计算机科学, 2020, 47(2): 72-75.

学术会议口头报告

[1] 中国粒计算与知识发现学术会议(CGCKD2019), 江苏镇江, 2019 年 8 月 9 日-11 日;

参与的项目

[1] 2017.12-2019.10 公安部重大专项 20170004。