南京大学

研究生毕业论文
(申请硕士学位)

论 文 题 目 基于超分辨率重建的低分辨率人脸识别

作 者 姓 名 王绪冬

学科、专业名称 计算机技术

研 究 方 向 计算机视觉

指 导 教 师 申富饶 教授 吴楠 副教授

2021 年 7 月 17 日

学　　号：MF1833077

论文答辩日期：2021 年 5 月 20 日

指 导 教 师：　　**(签字)**

Low-Resolution Face Recognition based on Face Hallucination Method

by

Xudong Wang

Supervised by

Professor Furao Shen, Assistant Professor Nan Wu

A dissertation submitted to
the graduate school of Nanjing University
in partial fulfilment of the requirements for the degree of
MASTER
in
Computer Technology

Computer Science and Technology
Nanjing University

May 29, 2021

南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目：基于超分辨率重建的低分辨率人脸识别

计算机技术 专业 2018 级硕士生姓名：王绪冬

指导教师(姓名、职称)：申富饶 教授 吴楠 副教授

摘　要

随着大数据与人工智能的发展，身份验证方式也逐渐发生变革。钥匙、指纹等传统的解锁方式正逐渐被人脸识别、声纹识别、步态识别等更为智能的方式所取代。在诸多新兴的识别方式中，人脸识别方法目前取得的效果最令人满意，已经在日常生活的许多场景中有所应用。在如今这样的数字化智能时代，将人脸识别技术应用于广泛部署的智能监控设备中，已经成为保障社会安全的需要。

然而，在许多场景中拍摄到的人脸并不理想，会因为拍摄距离、天气、拍摄目标移动等原因造成人脸过小或清晰度较低。但目前的人脸识别方法主要针对高分辨率人脸，对于公共场所监控所拍摄的通常较小的人脸，主流人脸识别模型的识别效果并不理想，解决这个问题最直观的方法便是对待识别的人脸进行超分辨率重建。和其他计算机视觉任务类似，基于深度学习的方法近几年在人脸超分辨率重建领域不断取得突破。但目前的研究普遍只针对于重建结果的视觉效果，而并没有关注重建人脸的识别准确率，使得人脸超分辨率重建领域很多研究成果的应用价值十分有限。

针对上述的实际需求以及目前存在的问题，本文研究通过超分辨率重建的方法，解决低分辨率人脸识别这一实际问题。结合近几年相关领域的一些模型及思想，本文提出了两个模型，并搭建了一套将模型应用于实际的身份验证系统，其主要内容如下：

1. 本文提出了一种**基于参考图片的人脸超分辨率重建模型**。该模型根据人脸数据集独有的特点，将基于参考的重建方法应用于人脸任务，并设计了一套参考图片选取流程。相比于以往的算法，我们的模型在视觉效果以及重建结果的识别准确率上都具有一定优势。

2. 本文提出了一种**基于身份信息的人脸超分辨率重建模型**。该模型结合了人脸识别、风格迁移、超分辨率重建等多个领域的思想，将身份信息融合到人脸图像重建的过程中，训练得到的模型能够在重建出更高分辨率人脸的同时保持原有的身份信息。相比于以往的方法，我们的模型在低分辨率人脸识别上具有显著优势。
3. 基于前面的工作，我们搭建了一**套身份验证系统**，以体现我们提出的模型具有实际应用价值。系统的人脸识别模块集成了前沿的人脸识别算法，同时将我们提出的重建模型加入到系统中作为识别流程中一个可选的模块。使得人脸识别模块能够对不同分辨率的人脸输入普遍保持较高的识别准确率，具有实际应用价值。

相关实验表明，本文的研究成果能够有效提高低分辨率人脸识别的准确率。按照本文的研究路线，还可以在该领域进行更多的研究工作。

关键词： 计算机视觉; 人脸识别；超分辨率重建

南京大学研究生毕业论文英文摘要首页用纸

THESIS: Low-Resolution Face Recognition based on Face Hallucination Method.

SPECIALIZATION: Computer Technology

POSTGRADUATE: Xudong Wang

MENTOR: Professor Furao Shen, Assistant Professor Nan Wu

ABSTRACT

With the development of big data and artificial intelligence, identity verification methods have gradually changed. Traditional unlocking methods such as keys and fingerprints are gradually being replaced by more intelligent ways such as face recognition, voiceprint recognition, and gait recognition. Among many emerging recognition methods, face recognition methods have achieved the most satisfactory results and have been applied in many scenes of daily life. In today's digital age, applying face recognition to widely deployed monitoring equipment has become necessary for social security.

However, faces images captured in many scenes are not ideal for the reasons such as the size is too small or the definition is low due to the shooting distance, weather, or object moving. However, almost of the current face recognition models only focus on high-resolution faces. The recognition effect of these models will not be ideal for the small faces taken by surveillance in public places. The most intuitive way to solve this problem is to reconstruct the low-resolution faces before inputting these images into the face recognition model. Similar to other fields in computer vision, deep learning models have continuously made breakthroughs in face hallucination in recent years. However, current researches generally focus on the visual effects of the reconstruction results. Those methods have not paid attention to the recognition accuracy of the reconstruction results, which makes them have limited application value.

In response to the actual needs and current problems which were mentioned above, we try to use the method of face hallucination to solve the actual problem of

low-resolution face recognition. Inspired by related work in recent years, we proposed two models and builds a set of models to be applied to the actual identity verification system in this paper. The main contents are as follows:

1. We proposed a reference-based super-resolution model for face reconstruction. According to the unique characteristics of the face data, the model applies the reference-based method to the face hallucination task. And we have designed a training procedure for choosing the reference image. Compared with previous methods, our model has advantages in visual effects as well as the recognition accuracy for reconstruction results.
2. We proposed a face hallucination model based on identity information. This model combines ideas in several fields such as face recognition, style transfer and super-resolution. Identity information has been used during the process of face reconstruction. The model can reconstruct a high resolution face as well as maintain the original identity information. Compared with previous methods, our model has significant advantages in low-resolution face recognition.
3. Based on the previous work, we have built an identity verification systems to show the practical value of our model. The face recognition module of the system integrates cutting-edge face recognition methods. At the same time, it can generally maintain a high recognition accuracy rate for face input of different definition, which has huge practical value.

Experiments show that the research results of this paper can effectively improve the accuracy of low-resolution face recognition. According to the research route of this paper, there is still much work that can be done.

KEYWORDS: Computer Vision, Face Recognition, Super-resolution

目　　录

插图清单

附表清单

第一章　绪论

1.1　研究背景及意义

随着大数据与人工智能时代的到来，身份验证方式也发生了巨大变化。ID卡、指纹、密码等传统的身份验证方式从安全性、便利性等多个角度都已经无法满足当前的实际需求。身份验证方式从用户行为的角度可以分为配合式与非配合式，传统的身份验证方式主要在配合式场景中使用，往往需要用户随身携带“通行证”(卡片、钥匙等)，或是在识别中需要用户配合机器(钥匙开门、刷卡等)，在当今快节奏的时代显得十分不便；非配合式识别是指用户无需做出某些动作配合机器的识别，只需要以正常的姿态经过识别设备，便可以自动完成识别，显然传统方法几乎无法胜任。非配合式识别由于其方便快捷的特点，在一些场景下显得十分必要的。例如，在火车站等人流量巨大的场景中，必须用非配合式的身份验证方式，更加自动化、智能化地对人群进行检测，才能够符合工作的高效要求[1][2]。在这样的时代背景之下，生物信息识别具有巨大的前景和研究意义。生物信息识别，包括人脸识别、步态识别等，与个体的“自然属性”相关联，不容易伪造，有利于非配合式的自动检测。在各种新兴的生物信息识别方法中，人脸识别历史最为悠久且效果相对稳定，目前在日常生活中已有广泛的应用。

但由于环境因素的影响，在很多实际场景下，清晰人脸的获取并不总是能够得到保证。例如，在火车站、机场、商业中心等人流量较大的场景中，虽然相关部门一直希望能够通过监控摄像头拍摄的图像识别过往行人身份，但在这些场景下，由于摄像头的架设位置普遍较高，拍摄结果中人脸所占区域通常很小。目前主流的人脸识别方法主要针对清晰度较高的人脸，且大多具有固定的输入尺寸。对于视频监控等场景下拍摄到的图像，人脸剪裁后得到的结果普遍不符合识别模型所要求的输入尺寸。通过实验可以看到，运用 SphereFace 人脸识别模型[3] 对于 LFW 数据集[4] 在 LFW-BLUFR 测试协议之下，能够达到 96.35% 的准确率。但如果将原本清晰的测试数据进行 4 倍降采样，然后再通过插值法恢复到需要的输入尺寸，同样的模型只能达到 50.84% 的识别准确率，主流人脸识

别模型对于低分辨率人脸的识别结果难以满足实际需求，这种情况为人脸识别投入实际应用造成了极大的阻碍。此外，在上述环境中，拍摄到的人脸图像经常会因分辨率较低、噪声污染、光照不足等环境因素使得清晰度较差，图片的质量会明显低于目前主流人脸识别数据集[5] [6] 中的图片。由于目前主流的人脸识别模型通常由这些高质量图片组成的数据集训练得到，上述实际场景中拍摄到的人脸经常会因为图片质量过低而无法识别。

这一类实际场景中十分常见的问题被统称为“低质量人脸识别” (Low Quality Face Recognition, 简称 LQFR)。图片的“低质量”包括分辨率低、噪声污染等多种情况，但在目前深度学习的背景下，各种用于提高图片质量的端到端模型大多能够相互借鉴。虽然本文的研究主要针对人脸图片分辨率过低的问题，但我们认为，对于解决噪声、光照等因素引起的图片质量问题也有一定借鉴意义。

在低分辨率人脸识别任务的多种解决方案中，超分辨率重建是最为直观的。该方案先将人脸进行重建，然后在对重建出的人脸进行识别，这种方法被称为“基于超分辨率重建的低分辨率人脸识别”。其他的解决方案，包括提取对低分辨率鲁棒的身份特征、寻找统一特征空间等，虽然得到的模型对低分辨率人脸的识别准确率有所提高，但在对于高分辨率人脸的识别准确率相比于主流识别模型也会有明显下滑。而在“重建 + 识别”的方案中，身份验证系统对低分辨率有较高识别率的同时，对于高分辨率人脸也保持前沿的识别准确率，下一节将具体介绍几类方法各自的特点。以上，就是基于超分辨率重建的低分辨率人脸识别研究的意义所在。

1.2 研究现状

早在 20 世纪七十年代，人脸识别就成为计算机视觉与图像处理领域一个备受关注的研究方向。不同于指纹识别、声纹识别等身份验证方式，人脸识别因为在使用时需要的用户配合较少，被认为是最为用户友好的一种身份验证方式。但人脸识别在应用中也面临许多环境带来的不确定因素，例如：昼夜引起的光照变化，用户或相机在拍摄时发生晃动，积尘、水渍等造成镜头污染。同时，随着年龄的增长，人的相貌会发生变化，但在很多档案系统中，可能长达数年才会

对用户的照片进行更新，这使得在实际场景中拍摄到的照片难以与系统中的存档进行身份验证。以上这些，都是人脸识别所要面临的挑战。

在上述人脸识别所能遇到的种种困难中，人脸分辨率较低的问题最为常见。随着科学技术的发展、交通的逐渐便利以及公共场所人流量逐年增大，为了应对社会安全的需要，“天眼” 数字监控系统的部署成为各种公共场所的必然需求。通过对人流密集公共场所的视频监控，在需要时可以通过调取、检索视频内容，来查找、追踪相关人员。但为了单个摄像头有更大的视野范围，公共场所的监控摄像头通常架设的位置较高，离单个行人的距离较远，同时，很多公共场所的监控为了节省成本，使用的录像设备成本较低，拍摄画面的分辨率也十分有限。而目前常用的人脸识别算法普遍面向质量较高的人脸输入，对于视频监控等场景拍到的人像，如果直接使用人脸识别算法进行识别，通常无法保证识别的准确率。因此，低分辨率人脸识别成为这些实际场景中关注的重点。在一些文献[7]中，低分辨率人脸识别任务的解决方案被分为四种：超分辨率重建、提取对低质量图片鲁棒的人脸特征、统一特征空间、模糊修复。本文的研究工作主要围绕超分辨率重建的方法，但为了使读者对这个领域的研究现状有一个大致的了解，本节接下来将分别介绍这四种解决方案。

1.2.1 超分辨率重建

超分辨率重建是最为直观同时也是目前最为有效的方案，本文的研究也主要围绕着基于超分辨率重建的低分辨率人脸识别展开。人脸超分辨重建 (Face Hallucination)，作为超分辨率重建的一个子领域，近年来也已经有诸多研究工作[8][9][10]。人脸超分辨率重建的工作主要可以分为两类，一类是更多的注重重建图片的视觉效果，另一类则更注重重建结果的识别准确率。虽然图片的质量 (人眼观察的效果) 与识别的准确率两者并不冲突，但对于图片质量的评价，目前并没有一个完善的指标，峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 与结构相似性 (Structural Similarity Index, SSIM) 等常用指标在一些研究[11] 中已经发现，其测试结果与人眼的观察并不完全一致。由于一般的超分辨率重建领域 (general image super-resolution) 模型大多只针对重建图片的视觉效果，这时直接将常用的模型，如 VDSR[12]、SRGAN[11]、EDSR[13] 等，在人脸数据集[4][6] 上进行训练，得到的人脸超分辨率重建模型也能够取得一定效果。在此基础上，专

门针对人脸的超分辨率重建模型通常在训练过程中参考了一些人脸相关的先验信息。例如，FSRNet[8] 与 SuperFAN[14] 等模型在重建过程中结合了人脸的特征点位置信息，使得重建结果更为清晰。但目前大多方法只关注重建结果的视觉效果，而没有将人脸超分辨率重建模型应用与识别相关的工作结合。关注识别效果的模型通常更加关注能否在重建前后维持图片中的身份信息，如 SICNN[15] 模型在训练过程中直接引入人脸识别相关的监督信息，保证模型在重建的过程中维持身份信息的一致性。人脸超分辨率重建是解决低分辨率人脸识别最为直观的方案，同时也是目前最为有效的方案，我们的研究工作也是围绕着人脸超分辨率重建展开。

1.2.2 低质量鲁棒的特征提取

低质量鲁棒的特征提取在过去的十几年间曾被广泛研究，这一类方法的目的是对人脸图片提取到更为鲁棒的特征，希望在图片质量下降时特征仍具有可区分性。这些特征提取方法通常是基于图片的全局或局部的纹理、色彩等特征而人工设计出特征提取方法，因此通常不需要基于学习。

LBP(Local Binary Pattern, 局部二值模式) 特征，因为具有一定可区分性，被广泛的应用于这一类方法中。Herrmann[16] 等人将改进后的 LBP 方法应用于视频中，通过融合不同的时间和尺寸比例下的头部姿态特征，避免了低分辨率人脸上出现稀疏 LBP 直方图的问题。Kim[17] 等人的工作则根据图片不同区域的清晰度来适应性调整 LBP 特征计算中邻域的范围。相比于直接使用 LBP 特征，他们的成果在低质量人脸识别上有了显著的提升。为了克服光线变化的问题，Zou 等人提出的 GLF 特征算法能够将人脸图片映射到光线不敏感的特征空间中。除了 LBP 以为，LPQ (Local Phase Quantization, 局部相位量化), HOG (Histogram of Oriented Gradient, 方向梯度直方图) 等经典的图像特征提取方法也都十分有助于对图片提取低质量鲁棒的特征[18] [19] [20]。

通过设计手工设计鲁棒特征的低质量人脸识别方法无需训练过程且计算速度较快，但因为大都依赖于图片的纹理，在图片质量过低时因为纹理过于模糊而难以取得有效特征。此外，这一类方法大多对光照、表情变化等因素十分敏感。Wang 等人总结了到 2014 年为止这一类方法的研究成果[21]，但随着后来深度学习的到来，这一类研究近几年进展较少。

1.2.3 统一特征空间法

不同于超分辨率重建以及低质量鲁棒特征的方案，统一特征空间法希望能够学习到一个高分辨率人脸与低分辨率人脸所共享的统一特征空间。在测试阶段，不同分辨率的图片都通过算法映射到一个特征空间中进行相似度匹配。

Biswas 等人的工作[22] 通过一种多维放缩方法将不同分辨率图片映射到同一个特征空间中，使得低分辨率图片 (LR) 在特征空间中与其对应的高分辨率图片 (HR) 距离相近。Wang 等人[23] 将 LR 与 HR 图片视为两组变量，并用 CCA(Canonical Correlation Analysis, 典型相关分析) 算法决定他们的映射关系，经过映射后将他们投影到同一个线性空间中。另外有许多方法[24][25] 通过字典学习找到 LR 与 HR 之间的映射关系，再将它们共同投影到一个特征空间中。可以看到，这一类方法与我们所研究的“超分辨率重建 + 识别”的方案有一定的相似之处。虽然在未使用深度学习之前，算法的特征提取、特征表示能力有限，但许多统一特征空间方法[26][27] 不仅仅关注属于同一个类的样本在特征空间中最小化类内距离，同样也对增大类间距离有关注。Haghighat 等人的工作[28] 提出了 DCA(Discriminant Correlation Analysis, 判别相关分析) 方法来改进 CCA，以更多的关注类间距离。

统一特征空间的低分辨率人脸识别也有很多关于深度学习的研究[29][30],Dan 等人[31] 将不同分辨率的人脸图片混合作为训练数据来训练一个深度模型，使得模型能够将图片映射到一个非线性的统一特征空间中。这一领域至今仍有很多值得研究的工作，并对一般的人脸识别研究具有借鉴意义。

1.2.4 基于模糊修复的低分辨率人脸识别

基于模糊修复 (Deblurring) 的低分辨率人脸识别，目的是通过修复图片中存在的模糊，从而提高图片质量，也同时能够提高识别准确率。图片的模糊主要是指在拍摄或存储时一些不利因素造成的质量下降问题，比如拍摄对象或相机在拍摄时发生抖动、镜头起雾、镜头磨损等等。图像去噪曾经是一个热门的研究领域，该领域的研究有三种分类方法：1. 将去噪方法分为盲去噪 (Blind image deblurring, BID) 与非盲去噪 (Non-blinding image deblurring, NBID), BID[32][33] 是指在进行去噪工作时，我们除了图片外没有其他先验信息，而 NBID 则是指事先

已经知道图片模糊的原因。NBID 方法[34] 通常只能应对一种模糊因素，在实际应用中具有较大局限性但通常效果较好，BID 方法有更为广泛的应用，但由于模糊的因素很多，在一些未知情况下难以取得好的效果；2. 局部模糊修复与全局模糊修复[35]。在一些情况下，例如相机发生抖动，相同类型的模糊会在图片全局普遍产生，而另一些情况下，图片不同的区域可能会有不同类型的局部模糊。显然，局部模糊修复通常更具有挑战性。3. 单一图片模糊修复 (Single image deblurring) 与多图片模糊修复 (multi-image deblurring)。由于不同的图片可能互相具有参考价值，在去模糊的过程中可以通过类似参考的方式，提高输出结果的质量，这种被称为多图片模糊修复。与此类似，在超分辨率重建领域也有单图片超分辨率重建与基于参考的超分辨率重建的区分。随着近些年来深度学习的兴起，图像模糊修复与图像超分辨率重建有着密切的关联，很多端到端的超分辨率重建模型只要适当的调整训练数据，也同样可以应用于模糊修复任务。

对于低分辨率人脸识别的几种研究并非各自独立，很多时候可以相互借鉴，甚至可以借鉴人脸识别以及一般的超分辨率重建等相关领域的研究。如前面介绍的那样，超分辨率重建与去噪两个领域有着紧密的联系。另外，也很容易看出，低分辨率鲁棒的特征提取与统一特征空间两类方法具有较大关联，同时也与人脸识别的研究工作很接近。近年来随着深度学习的发展，卷积神经网络在多种计算机视觉任务中发挥作用，相似的模型架构使得原本相关但不同的研究领域更容易相互借鉴。

1.3 本文工作

第1.2节介绍的几类方法中，低质量鲁棒的特征提取方法难以有效应对测试数据中人脸表情、角度等变化因素；统一特征空间方法虽然能够一定程度上提高低分辨率人脸的识别准确率，但会使高分辨率识别准确率有所下降；去噪的方法并不完全符合我们针对的问题，仅有一定的借鉴意义；而超分辨率重建的方法很大程度上可以克服前面几类方法存在的不足，能够有效解决低分辨率人脸识别问题。在此背景下，本文致力于通过超分辨率重建来解决低分辨率人脸识别问题，主要的研究成果包括以下几个方面：

1. 提出了一种基于参考的人脸超分辨率重建方法 (Ref-based Face Hallucination),

利用人脸数据集类内数据相似度较高的特性，在重建的过程中通过对比特征图相似性，将参考人脸的特征引入到重建结果中，增强重建结果的细节特征。

2. 提出了一种基于身份信息的人脸超分辨率重建模型，C-Face Network。设计了一个在重建过程中有助于保持身份信息的损失函数 (C-Face Loss)，并针对训练过程中不易收敛的问题，提出了一套新的训练方法。实验结果表明，通过我们的损失函数以及训练方法得到的模型能够有效的保持身份信息，在低分辨率人脸的识别上具有显著优势。
3. 搭建了一套多模态身份验证系统。系统包含多种身份验证方式，包括人脸识别、步态识别、声纹识别。当拍摄到的人脸质量较低时，系统中嵌入了我们的超分辨率人脸重建模型，有助于提高识别准确率。

1.4 本文组织结构

在本文的第一章中，我们从研究意义、研究背景、目前的研究现状等几个方面介绍了我们的研究课题，并再次基础上简要介绍了我们所做的工作。在第二章中，我们将用一定的篇幅介绍和我们的研究内容相关的一些工作。

根据我们的课题以及前面的介绍可以看到，我们的研究工作及成果主要与人脸识别以及超分辨率重建两个领域有较大关联。在下一章中，我们将分别介绍这两个相关领域在各个时期具有代表性的工作。首先介绍在深度学习到来之前的人脸识别以及超分辨率重建方法，然后我们将介绍深度学习的一些基本概念。在此之后，我们将分别介绍近年来人脸识别以及超分辨率重建两个领域在深度学习方面的工作。第三章将介绍我们提出的一种基于参考的人脸超分辨率重建方法，我们将从模型结构、训练过程、实验结果等几个方面进行介绍。第四章将介绍我们提出的另一个模型——基于身份信息的人脸超分辨率重建模型，也同样会从原理、实验等多个方面详细说明。为了体现本文工作的实际应用价值，我们将在第五章介绍一套我们自主研发的 RINC-ID 身份验证系统。通过该系统，我们将研究成果应用于实际场景中，我们将详细介绍系统的架构以及所用到的软硬件技术栈。最后，我们将在第五章总结所有的工作，同时分析目前人脸识别以及人脸超分辨率重建领域研究工作存在的不足之处，并指出未来的研究方向。

第二章　相关工作介绍

基于超分辨率重建的低分辨率人脸识别，是计算机视觉领域一项具有实际应用价值的研究工作。在许多场景下，针对低分辨率的人脸进行超分辨率重建成为了识别过程中必不可少的一环。人脸超分辨率重建是图像超分辨率重建的一个子领域，目前许多人脸超分研究的目标仅仅在于重建后的视觉效果。我们的研究则更希望模型在重建的过程中保持身份特征信息，重建得到结果的识别率相比于原本低分辨率人脸有明显上升。因此，我们的研究课题——基于超分辨率重建的低分辨率人脸识别研究，与人脸识别、超分辨率重建两个领域都有较大关联，几个相关领域之间的关系可以用图2-1大致概括。本章之后几节将会分别对几个相关领域进行介绍，包括相关的背景知识以及近年来各个相关领域具有代表性的研究工作。

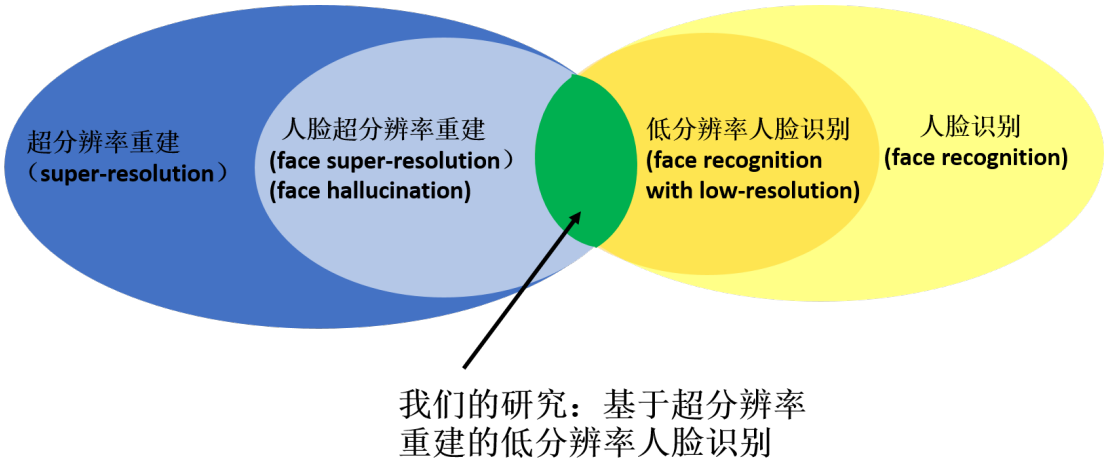

图 2-1: 几个相关研究领域之间的关系

2.1　传统人脸识别方法

目前大多数人脸识别系统都如图2-2所示，主要包括人脸检测、人脸对齐、人脸特征表示与人脸匹配四个模块。由于人脸识别系统的非配合性，摄像头拍到人脸的同时，难免也拍到一些无关的背景。这些背景内容与用户身份无关，会对后续的人脸特征表示以及人脸匹配造成干扰。因此，通常通过人脸检测与人

脸对齐两个步骤，从拍摄到的图片中剪裁出人脸部分，同时尽量少的包含背景信息。然后，通过仿射变换 (Affine Transformation) 适当调整拍摄到人脸的位置，能够消除姿势不同带来的误差。在此之后，将经过对齐的人脸图片输入到人脸特征表示模型 (也可以称为人脸识别模型) 中，得到一个特征向量。在系统的注册阶段，只需要将这个特征向量连同其他输入信息存储到数据库中。而在实际使用中，即识别阶段，将得到的特征向量与数据库中存储的向量进行相似度匹配，根据相似度是否超过阈值来确定是否为合法用户。

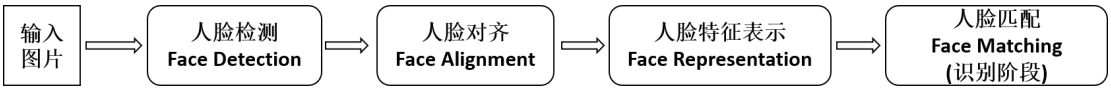

图 2-2: 人脸识别系统的各个模块

在图2-2的几个模块中，人脸特征表示通常被认为是对于人脸识别系统而言最为重要的模块，也是人脸识别领域主要的研究方向，本节将介绍人脸识别研究的早期几个具有代表性的研究工作。

早期有关人脸识别的研究主要是关注和运用人脸的几何特征[36][37]，通过对人脸图片运用一些轮廓检测或边缘检测方法，找到几个人脸关键点的坐标，通过计算关键点之间的相互位置，来区分不同人的身份，文献[38] 充分讨论了这一类运用人脸特征点与几何特征的人脸识别方法。这一类方法只能在一些很小的数据集上取得一定的效果，但这些研究在 70 年代为机器识别人脸提供了可能性。同时，相比于后来提出的方法，基于几何特征的方法 (geometry-based method) 运行速度更快且占用内存更小。另外，随着近年来深度摄像技术的发展，基于几何特征的方法在 3D 人脸识别上展现出了新的可能性[39][40]。

在此之后，统计子空间法 (statistical subspaces method) 成为人脸识别领域的主流方法，这一类工作将人脸图片通过一些线性或非线性方式映射投影到一个全新的向量空间中，再通过特征匹配来识别身份信息。相比于之前基于几何特征的方法，这一类算法在计算特征时考虑整个人脸的信息，因此也被称为人脸识别的“整体法”(holistic method)，其中最知名的便是基于 PCA 的人脸识别[41][42]。PCA(Principal Component Analysis，主成分分析) 是一种常用的数据降维算法，PCA 算法描述如下：

将 PCA 算法应用于人脸识别时，首先将训练数据中每一张图片通过列像素

Algorithm 2.1 PCA 算法

1: 输入：样本向量集合 $D = \{x_1, x_2, ..., x_n\}$, 降维后特征空间维度 d';
2: 对各个输入向量做去中心化处理:$x_i \rightarrow x_i - \frac{1}{n}\sum_{j=1}^{n} x_j$;
3: 去中心化的输入向量组成矩阵 X, 计算输入样本的协方差矩阵 XX^T;
4: 对上述协方差矩阵进行特征分解，得到特征值 $\lambda_1, ..., \lambda_d$ 与特征向量 $\omega_1, ..., \omega_d$
5: 取最大的 d' 个特征值对应的特征向量 $\omega_1, ..., \omega_{d'}$
6: 输出：投影矩阵 $W^* = (\omega_1, ..., \omega_{d'})$

的堆叠表示成一个列向量，然后将列向量组成的人脸样本矩阵作为上述 PCA 算法 2.1 的输入。得到的输出矩阵 W 用于将人脸识别中的输入图片映射到特征空间。在测试中，对于输入的人脸图片，用同样的方法表示为列向量，然后将列向量与矩阵 W 相乘得到人脸特征向量，根据不同人脸的特征向量相似度来判断是否属于同一个人。其中，组成矩阵 W 的特征向量 $\omega_1, ..., \omega_{d'}$ 可以通过拆分重组为二维矩阵而重新表示成图像，被称为特征脸，如图2-3所示。

图 2-3: PCA 特征脸

PCA 人脸识别能够有效处理较大的人脸数据 (相比于在它之前出现的人脸识别方法)，且无需过大的机器压力就能进行实时的人脸识别，但这种方法对于拍摄照片的光照、环境变化等因素十分敏感。同时，PCA 算法进行数据降维上也存在一定缺陷，例如在数据非高斯分布的情况下，PCA 方法得到的主元可能并非最优。

在此之后，LDA(Linear Discriminacant Analysis，线性判别分析) 人脸识别方法[43] 改进了 PCA 中存在的不足之处。PCA 与 LDA 方法关注人脸图片的全局结构信息，而 LPP[44] 方法则更加关注局部结构信息的相似性，使得算法对于表情等变化因素更为鲁棒。LPP、LDA 等 Holistic Method 方法在深度学习到来之前

一直是前沿的人脸识别算法，直到今日仍在部分场景中发挥着用途。

2.2 传统超分辨率重建方法

在传统数字图像处理领域，超分辨率重建的方式通常是通过插值法对图像进行放大。插值法的原始定义是指对于一个函数，已知有限个点的函数值，通过这些点估计出其他点的函数值。例如已知 (x_0, y_0) 与 (x_1, y_1)，计算区间 $[x_0, x_1]$ 内某一位置对应的 y 值。在图像的插值法中，这个被估计的函数值是图片中某些位置的像素值。

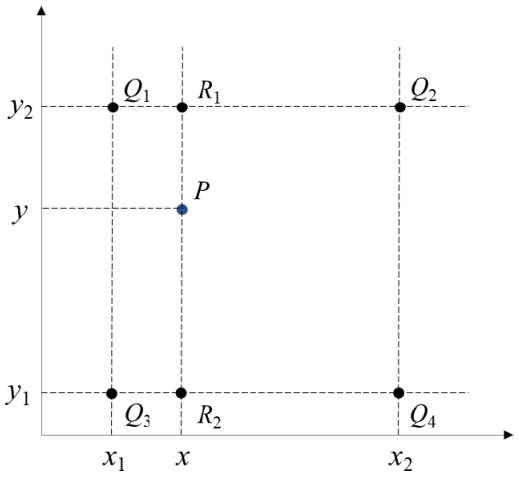

图 2-4: 插值法示意图 (最近邻插值，双线性插值)

图像插值法中最简单的是最近邻插值法，直接把待估计点的像素值设置为其周围四个点中，距离其最近的点的像素值。如图2-4所示，Q_1, Q_2, Q_3, Q_4 的坐标分别为 $(x_1, y_2), (x_2, y_2), (x_1, y_1), (x_2, y_1)$，且已知他们的像素值 $f(Q_1), f(Q_2), f(Q_3), f(Q_4)$，则 $P(x_p, y_p)$ 处的像素值 $f(P)$ 计算方法如下：

$$f(P) = \begin{cases} f(Q_1), x_p - x1 < x2 - x_p \text{ and } y2 - y_p < y_p - y_1 \\ f(Q_2), x_p - x1 > x2 - x_p \text{ and } y2 - y_p < y_p - y_1 \\ f(Q_3), x_p - x1 < x2 - x_p \text{ and } y2 - y_p > y_p - y_1 \\ f(Q_4), x_p - x1 > x2 - x_p \text{ and } y_2 - y_p > y_p - y_1 \end{cases} \tag{2-1}$$

从公式(2-1)可以看到，最近邻插值法原理十分简单，单从方法名便可明白其原理。但这种方法容易使生成的图像有许多锯齿。相比之下，双线性插值在计

算待求位置的像素值时，能够对周围四个点的像素做一个权衡，通常得到的结果更平滑。同样对于图2-4所示的情况，双线性插值在 x 与 y 两个方向上分别进行插值。首先在 x 方向上进行插值：

$$f(R_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_1) + \frac{x - x_1}{x_2 - x_1} f(Q_2), \text{ where } R_1 = (x, y_1) \tag{2-2}$$

$$f(R_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_3) + \frac{x - x_1}{x_2 - x_1} f(Q_4), \text{ where } R_2 = (x, y_2) \tag{2-3}$$

基于 $f(R_1)$ 与 $f(R_2)$ 的结果，在 y 方向进行插值，并按照公式(2-2)与公式(2-3)展开，便可得到二次插值法求 P 处像素值的结果:

$$\begin{aligned} f(P) =& \frac{y - y_1}{y_2 - y_1} f(R_1) + \frac{y_2 - y}{y_2 - y_1} f(R_2) \\ =& \frac{(y - y_1)(x_2 - x)}{(y_2 - y_1)(x_2 - x_1)} f(Q_1) + \frac{(y - y_1)(x - x_1)}{(y_2 - y_1)(x_2 - x_1)} f(Q_2) \\ &+ \frac{(y_2 - y)(x_2 - x)}{(y_2 - y_1)(x_2 - x_1)} f(Q_3) + \frac{(y_2 - y)(x - x_1)}{(y_2 - y_1)(x_2 - x_1)} f(Q_4) \end{aligned} \tag{2-4}$$

二次插值法相比于最近邻插值法，虽然增加了计算的复杂度，但由于待求点像素的最终值是周围四个点像素值加权的结果，得到的图像像素值更具有连续性。双三次插值与二次插值相比，在计算待求点像素时考虑到周围更大的范围。如图2-5所示，用双三次插值法计算 $(x + u, y + v)$ 处的像素点时，将对周围 16 个采样点的像素值进行加权平均。

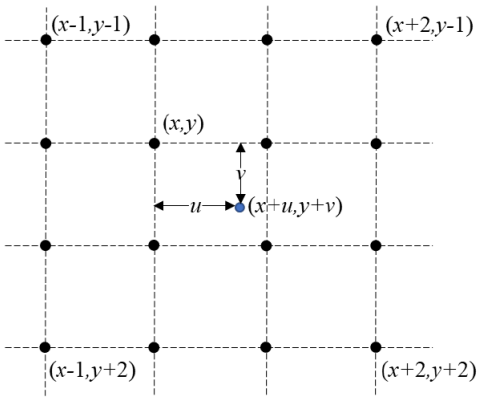

图 2-5: 插值法示意图 (双三次插值)

双三次插值法在计算的过程中首先需要计算插值核函数，插值核的表示方式如下：

$$w(x) = \begin{cases} |x|^3 - 2|x|^2 + 1, \text{for } 0 \le x < 1 \\ -|x|^3 + 5|x|^2 - 8|x| + 4, \text{for } 1 \le x < 2 \\ 0, \text{for } |x| \ge 2. \end{cases} \tag{2-5}$$

根据公式(2-5), 图2-5中 $P(x+u, y+v)$ 处的像素值 $f(P)$ 计算如下:

$$f(P) = ABC \tag{2-6}$$

$$A = [w(1+u) \;\; w(u) \;\; w(1-u) \;\; w(2-u))] \tag{2-7}$$

$$B = \begin{bmatrix} f(x-1,y-1) & f(x-1,y) & f(x-1,y+1) & f(x-1,y+2) \\ f(x,y-1) & f(x,y) & f(x,y+1) & f(x,y+2) \\ f(x+1,y-1) & f(x+1,y) & f(x+1,y+1) & f(x+1,y+2) \\ f(x+2,y-1) & f(x+2,y) & f(x+2,y+1) & f(x+2,y+2) \end{bmatrix} \tag{2-8}$$

$$C = [w(1+v) \;\; w(v) \;\; w(1-v) \;\; w(2-v)]^T \tag{2-9}$$

双三次插值法充分利用了周围的局部信息，相比于之前的方法，得到的结果有更高的清晰度。双三次插值法如今仍在数字图像处理领域发挥着重要作用，在 OpenCV 等软件包中，双三次插值是图像放缩的默认方法。

2.3 深度学习与卷积神经网络

随着近年来人工智能的兴起，深度神经网络在计算机视觉领域的许多任务上取得了突破进展，其中也包括人脸识别以及超分辨率重建领域。深度学习研究包括许多不同的神经网络模型，其中在人脸识别领域最常用的是卷积神经网络 (Convolutional neural network, CNN)。本节将简要介绍深度学习与卷积神经网络，为后面介绍基于深度学习的人脸识别以及超分辨率重建模型做好铺垫。

2.3.1 人工神经网络

人工神经网络 (Artificial Neural Network, 简称 ANN) 通过模仿生物神经元结构而提出的一种人工智能模型，相关的研究工作曾在上世纪 40 年代、80 年代以

及近十年来多次兴起。因为模型的结构与人类大脑的神经元结构相似，早期研究者希望 ANN 模型能够拥有与人脑相当的能力。类似于人脑中神经元通过突触前后连接，人工神经网络通常也有多个网络层。每层中包含多个神经元，在计算过程中，前一层神经元的输出结果作为后一层神经元计算的输入。因此，早期的 ANN 模型也被称为多层感知器。增加 ANN 模型的深度虽然会增加计算量，但也有助于增强模型的表现能力。有研究表明[45]，在神经元总数相同的情况下，增加网络模型的深度比增加广度更有利于使网络获得更强的表现能力。图2-6展示了一个 4 层全连接网络的结构，包括一个输入层、两个隐藏层、一个输出层。在前向传播的过程中，隐藏层与输出层的输出可用如下公式计算：

$$y_j^i = f(\sum_{k=1}^{n_{i-1}} y_k^{i-1}\omega_{k,j}^i + b^i) \tag{2-10}$$

其中 n_{i-1} 表示第 $i-1$ 层神经元的个数，y_j^i 表示第 i 层第 j 个神经元的值，$\omega_{k,j}^i$ 表示从第 $i-1$ 层的第 k 个神经元到第 i 层第 j 个神经元的连接权重。f 表示激活函数，有助于提高模型拟合非线性函数的能力。

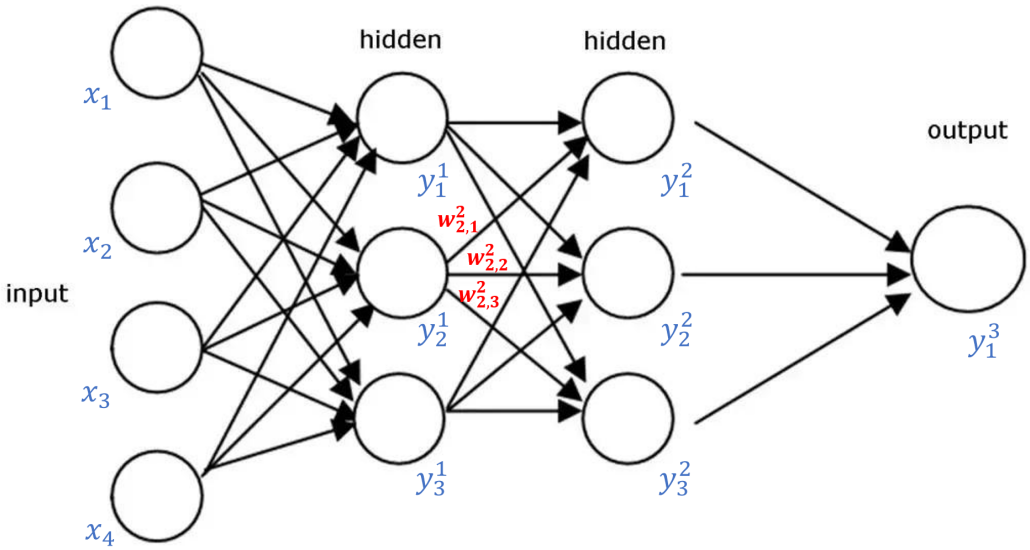

图 2-6: 全连接网络示意图

激活函数主要有 sigmoid, tanh, ReLU 几种，其中 ReLU 最为常用。ReLU 函数的定义可如下表示：

$$f(x) = max(0, x) \tag{2-11}$$

ReLU 函数的定义与数学表达都十分简洁，当输入小于 0 时输出为 0，输入 x 大于 0 时直接输出 x。相比与 sigmoid、tanh 等，ReLU 函数能够使模型的梯度更快收敛。同时，因为 ReLU 简单的数学定义，ReLU 作为梯度函数在计算梯度时远快于 sigmoid 或 tanh。但 ReLU 函数也同样具有缺点，如公式(2-11)所示，当输入小于 0 时，输出为 0。在训练的过程中，当梯度过大或学习率较大时，可能会使得某些神经元在后面的训练中不再对输入数据做出相应，导致部分神经元“坏死”。目前已经有一些对 ReLU 进行改进的激活函数，如 Leaky-ReLU、P-ReLU 等等。尽管有神经元“坏死”的问题，ReLU 仍然是目前应用最为广泛的激活函数。在本文接下来的内容中，如无特别说明，深度学习模型中所用的激活函数均为 ReLU 函数。

2.3.2 卷积神经网络

卷积神经网络 (Convolution Neural Network，CNN) 是指包含卷积计算层的深度神经网络，卷积层是卷积神经网络最重要的组成结构，同时也是大部分卷积网络中数量最多的层。卷积层通常由多个卷积核滤波器构成，卷积核可以看作滑动窗口。在网络前馈的过程中，卷积核在输入中滑动来捕捉图像中的特征。训练时，通过反向传播机制优化滤波器的参数。卷积神经网络是受生物视觉机制的启发而提出。Hubel 与 Wiesel 在 1962 年的研究中发现[46]，生物大脑中能够响应视觉环境中简单特征的初级视觉皮层中主要有两种不同类型的细胞——简单细胞与复杂细胞。其中简单细胞只在特定的空间位置对它偏好的方向产生响应机制，而能够对更大空间具有响应不变性的复杂细胞先对来自多个简单细胞的输入进行池化，再对得到的结果进行处理，进而产生响应。目前许多卷积神经网络仍是这种卷积层后跟随池化层的模式，如图2-7所示。

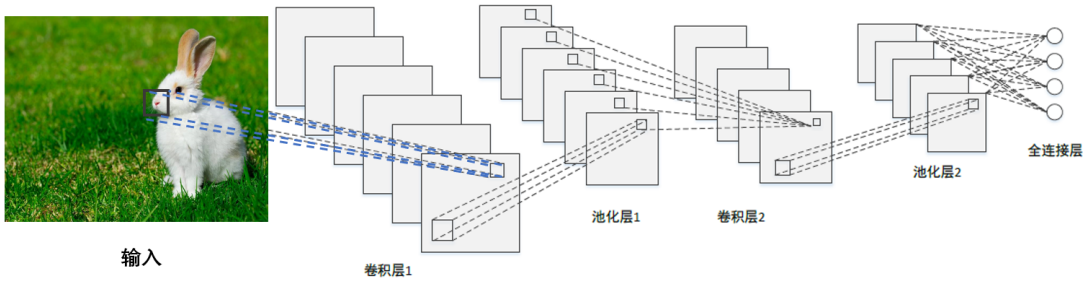

图 2-7: 卷积神经网络示例

卷积层用于提取网络的特征信息，一个 CNN 模型中较浅的卷积层通常会捕捉到图像中一些相对低级的信息，如线条或简单的轮廓，而处于较深层次的卷积层则会提取到一些比较复杂的特征。为了洞悉神经网络每一层究竟提取了什么信息，Matthew 等人的研究[47] 对卷积层的 feature map 进行了可视化，如图2-8所示。从图2-8中隐约可以看到卷积网络的各个层提取到了复杂程度不同的特征信息。

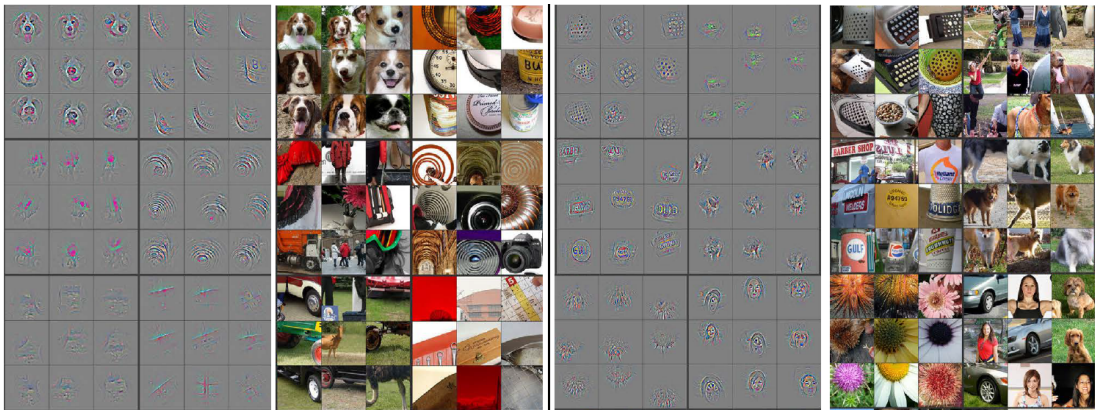

图 2-8: 卷积网络 feature map 可视化

对于如图2-6所示的传统全连接网络，在处理较为复杂的任务时，会因每一层参数过多而使层数受限。同时，全连接神经网络进行图像处理任务时，需要将输入图片“拉长”为一维向量作为网络输入，没有充分利用图像中相邻像素之间的关联性。卷积神经网络具有局部感知、权值共享等特性，能够克服全连接网络的许多缺陷，推动了深度学习在人工智能，尤其是视觉相关领域的发展。

2.4 基于深度学习的人脸识别算法

卷积神经网络是人脸识别领域最为常见的深度学习模型，通过大量人脸图片数据对网络进行训练，使得模型对于人脸具有身份识别与特征表示的能力。相比于传统方法 (将深度学习到来之前的方法统称为传统方法)，深度学习模型在数据集中学习到的特征通常更具有鲁棒性，更能够应对年龄、面部表情、光线等常见的变化因素。

卷积神经网络应用于人脸识别的一个常见方法是在训练过程将识别任务看作是一个图像分类问题，即训练数据集中每一个人对应一个类。目前主流的人脸识别模型[48] [49] 在训练阶段与图像分类任务相似，在若干卷积层后跟随几个全

连接层，其中最后一个全连接层用于分类。但人脸识别在应用中通常要求“开集识别”，即实际遇到的用户并没有在数据集中出现。因此，模型在应用时将抛弃分类层，使用原本模型的倒数第二层的输出作为输入人脸的特征向量。通过衡量不同人脸照片对应的特征向量相似与否，来判断他们是否属于同一个人。这样，在面对训练数据中没有出现的人的照片输入时，模型也能够有效进行人脸识别，有效应对开集识别任务。

在目前大多数图像分类任务中，卷积神经网络通常能够学习到具有鲁棒性的分类特征，在分类任务中取得较好成绩。但人脸识别模型相比于模型最终的分类结果，更关注对于输入图像的特征提取。且由于在训练过程中一个身份作为一个样本类，总的类别数目特别多。想要在面对数据集中未出现的人时能够通过提取的特征进行有效的人脸识别，要求提取的特征向量保证类间距离 (inter-class distance) 大于类内距离 (intra-class distance)。Softmax 函数作为损失函数，在训练时能够促进学习到的特征具有更大的类间距离，用 softmax 损失训练 CNN 网络在人脸识别的许多工作[50][51] 中取得了显著成效。Softmax 函数如下定义：

$$L = \frac{1}{N}\sum_{i} -\log(\frac{e^{f_{y_i}}}{\sum_{j\neq i} e^{f_j}}) \tag{2-12}$$

公式(2-12)所对应的二分类问题决策边界为 $(w_1 - w_2)x + b_1 - b_2 = 0$，在训练中没有要求类内距离大于类间距离。在 softmax 函数的基础上，如果限制权重 $||w_1|| = ||w_2|| = 1$, 并用 θ_1, θ_2 分别表示向量 x 与两个权重对应向量的夹角，则将 softmax 损失函数应用到二分类问题时，其决策边界可表示为:

$$||x||(\cos\theta_1 - \cos\theta_2) = 0 \tag{2-13}$$

公式(2-13)将人脸的特征向量映射到了一个高维的球面空间中，通过特征向量的角度可分性来保证模型对不同人的有效区分，不过公式(2-13)所表示的决策边界仍没有保证样本特征向量类间距离大于类内距离。为了使身份特征向量具有更大的角度可分性，同时保证类间距离大于类内距离，SphereFace[3] 模型在2-13的基础上加入了一个常数 m, 使得属于不同类的人脸，其特征向量的角度

表 2-1: 人脸识别中几种不同的决策边界

损失函数	对应决策边界
softmax loss	$\|\|x\|\|(\cos\theta_1 - \cos\theta_2) = 0$
A-softmax loss	$\|\|x\|\|(\cos(m\theta_1) - \cos\theta_2) = 0$
Cosine-Face loss	$s(\cos\theta_1 - m - \cos\theta_2) = 0$
ArcFace loss	$s(\cos(\theta_1 + m) - \cos\theta_2) = 0$

差异足够大。在二分类问题中，这种决策边界如下表示：

$$\begin{aligned} &||x||(\cos(m\theta_1) - \cos\theta_2) = 0, \text{ for class 1;} \\ &||x||(\cos(\theta_1) - \cos(m\theta_2) = 0, \text{ for class 2.} \end{aligned} \tag{2-14}$$

应用这种决策边界，sphereFace 中提出了 A-Softmax 损失函数，定义如下：

$$L_{A-Softmax} = \frac{1}{N}\sum_i -\log\Big(\frac{e^{\|x_i\|\psi(m\theta_{y_i,i})}}{e^{\|x_i\|\psi(m\theta_{y_i,i})} + \sum\limits_{j\neq y_i} e^{\|x_i\|\cos(\theta_{y_j,i})}}\Big). \tag{2-15}$$

其中 y_i 表示第 i 个训练样本所对应的分类，由于训练的卷积神经网络最后跟随全连接层用于分类，用 $\omega_1 ... \omega_n$ 表示最后一层全连接中对应于第 1 到 n 个类的权重向量。$\theta_{y_j,i}$ 表示第 i 个训练样本的特征向量与第 ω_{y_j} 与 x_i 的夹角。$\psi(m\theta_{y_i,i}) = (-1)^k\cos(m\theta_{y_i,i}) - 2k$, $\theta_{y_i,i} \in [\frac{k\phi}{m}, \frac{(k+1)\phi}{m}]$, $k \in [0, m-1]$。当 $m = 1$ 时，A-Softmax 退化为 softmax 损失。

SphereFace 人脸识别模型通过 A-Softmax 损失函数训练得到的模型能够提取到更为鲁棒的身份特征信息，这种在高维球面特征空间上改进特征向量角度可区分性的思想也成为了之后几年人脸识别研究的热点。在此之后，Cosface[52], Arcface[53] 等模型都以此为思路对公式(2-14)表示的决策边界进行了改进，表2-1对他们进行了对比。

尽管近年来基于深度学习的人脸识别模型不断取得突破，但大多只针对分辨率较高的人脸输入有效。在一些时候，输入的人脸图片已经具有足够识别的身份特征，但由于其分辨率低于模型训练数据的平均水平，而导致模型的识别效果不佳。对于这种问题，目前主流的人脸识别模型没有较好的解决方法。一个直观的解决方案就是对低分辨率人脸进行重建，然后在对重建得到的结果进行

人脸识别。

2.5 基于深度学习的超分辨率重建

近年来，深度学习在人工智能、尤其是计算机视觉领域的大多数任务中取得了突破进展。在 2015 年，Dong 等人提出 SRCNN 模型[54] 首次将 CNN 运用于超分辨率重建任务中。如图2-9所示，该模型包含三个模块。前两个模块由卷积层与激活函数构成，激活函数为 ReLU，第三个模块只有卷积操作，第三个模块的输出即为重建的结果。不同于分类、定位等任务，由于超分辨率重建任务最终输出的是尺寸不小于输入的重建图片，超分辨率重建任务中的深度学习模型通常不包含池化层。

SRCNN 的三个模块分别对应着三个阶段的操作：块提取与表示 (patch extraction and representation)，非线性映射 (non-linear mapping)，重建 (reconstruction)。目的是将基于稀疏表示的超分辨率重建技术应用于卷积神经网络中，通过训练过程，自动学习到从低分辨率图片到高分辨率图片的编码解码过程。

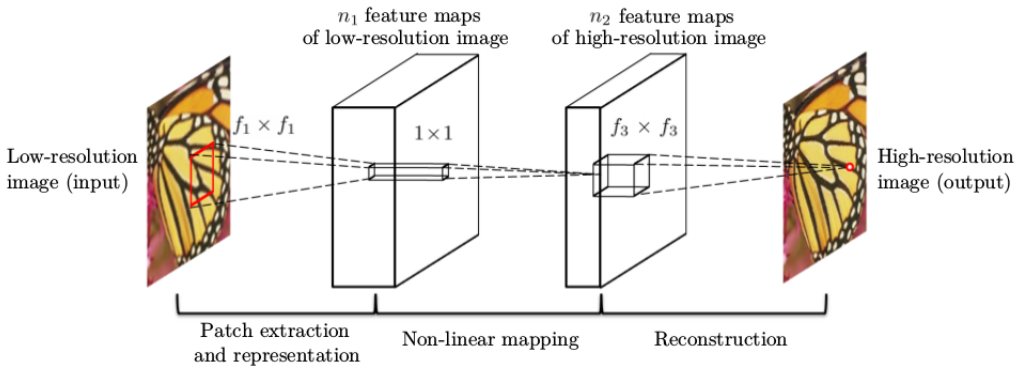

图 2-9: SRCNN 模型架构

SRCNN 模型相比于之前的方法，在 PSNR 与 SSIM 指标上有所提升，且是一个端到端的模型，在实际应用中无需过多的人工干预或多阶段计算。但在低分辨率图片输入到 SRCNN 之前，需要先进行预处理，通过插值法将图片的尺寸扩大，然后才能输入到 SRCNN 模型中进行重建。在之后几年的超分辨率重建模型中，通过在网络中加入子像素卷积 (sub-pixel convolutional) 来使得最后的输出尺寸变大，而无需在输入前进行预处理。

在 SRCNN 之后，基于深度学习的超分辨率重建模型层出不穷。同时另一方面，大规模人脸数据集也已经能够满足深度学习所需要的数据量。直接使用人脸数据重新训练超分辨率重建模型，在人脸重建任务上通常也能够取得一定的效果。但这种方式得到的模型没有在重建过程中充分利用人脸的特性。针对人脸的超分辨率重建，在英语文献中常常被称为“人脸幻视 (Face Hallucination)”，该任务的大多数方法在重建过程中会利用一些人脸相关的先验信息。如 FSRNet[8] 模型中，利用了人脸关键点热图 (landmark heatmap) 以及人脸解析图，来获得更好的面部重建图像，其大致模型架构如图2-10所示[8]。

FSRNet 模型可以分为两大部分，即粗糙重建网络 (Coarse SR Network) 与精细重建网络 (Fine SR Network)。前者 (Coarse SR Network) 是一个包含 3 个 Residual Block 结构的超分辨率重建模型，对人脸的重建与普通的超分辨率模型差别不大，目的是通过这个网络对低分辨率输入先进行一个大致的重建。精细重建网络还包含三个子模块，分别为精细超分编码器 (Fine SR Encoder)、先验估计网络 (Prior Estimation Network)、精细超分解码器 (Fine SR Decoder)。其中编码器网络受 ResNet 启发，通过多个 ResBlock 结构对前面粗略重建的结果进行特征提取。先验估计网络则是使用一个 HourGlass 结构来获取人脸的 landmark heatmap 与解析图。由于两者都是用于表示 2D 人脸的形状特征，所以对于提取两者的任务而言，网络结构可以共享，只要在 HourGlass 的末端通过两个分离的 1x1 卷积便可分别得到这两种人脸先验信息。最后，将先验估计网络提取到的 heatmap 与解析图，以及编码器网络输出的 feature map 一同输入到解码器网络中，输出更为清晰的人脸重建结果。

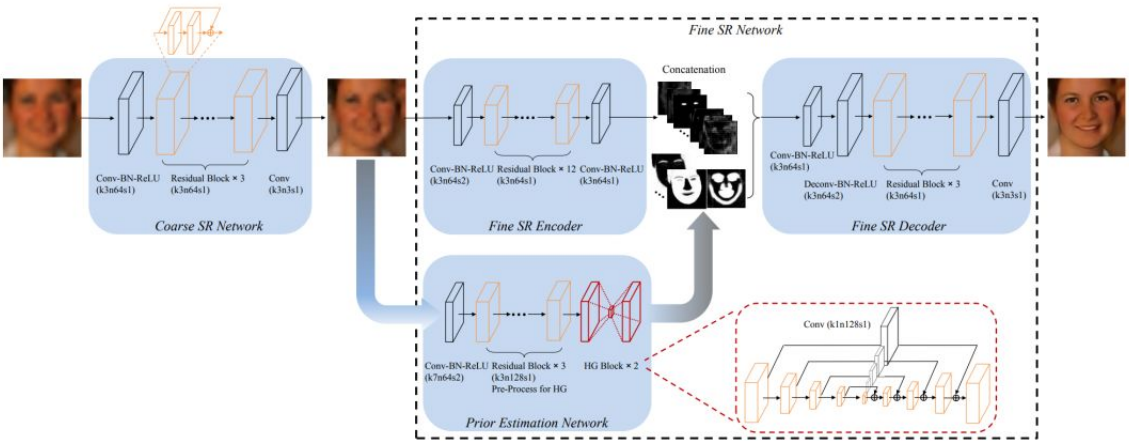

图 2-10: FSRNet 模型架构

FSRNet 所利用的先验信息主要涉及到提取人脸中五官的位置以及面部轮廓等信息，和人脸识别类似，提取人脸 landmark 的网络模型通过大量人脸数据训练得到，其训练图片大多较为清晰，模型在应用中也对清晰的图片有更好的效果。如果直接对低分辨率人脸进行提取先验信息，会由于输入图片过于粗糙而使结果不理想。因此，在 FSRNet 模型中，先构造一个重建模型进行粗略的重建，能够一定程度提高低分辨率输入的清晰度，然后再通过与特征提取相结合进行更为精准的人脸超分重建。

2.6 基于身份信息的人脸超分辨率重建

诸如 FSRNet[8], WaveletSR-Net[9] 等模型在进行人脸超分辨率重建过程中，利用了 2D 人脸特有的几何信息，在人脸重建任务中取得了比一般的超分辨率重建模型更好的效果。但人脸超分任务最大的实际用途在于服务人脸识别工作，将实际情况中遇到的低分辨率人脸进行重建，使其接下来能够被更好的识别。但前面提到的重建方法主要关注的是重建的视觉效果，并通过峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 与结构相似性 (Structural Similarity Index, SSIM) 两个指标作为测评标准，没有考虑到重建是否更有利于人脸的识别工作。

由于深度神经网络具有一定的“黑箱性”，对于目前主流的几种人脸识别网络，我们并不能完全透彻的了解哪些人眼可以观察到的信息能够对识别结果产生影响。况且，在 SRGAN[11] 等文章中已经指出，PSNR 与 SSIM 两个指标的高低于人眼观察的结果并不一致，那么在 PSNR 于 SSIM 指标上取得较好结果的模型也就难以保证他们相比于其他模型的重建结果更有利于人脸的识别。

希望在人脸超分辨率重建过程中保留身份信息以使得重建结果更有利于人脸识别，即基于低分辨率人脸识别的超分辨率重建，这一研究方向相比于一般的人脸超分研究一直不算是一个热门的研究方向。很大一个原因就是，许多人脸超分的研究者认为只要重建结果与真正的高分辨率原图足够的相似，那么重建结果一定能够有效的在人脸识别中使用。这种想法自然有其道理，但由于 SRGAN[11] 等论文在实验中发现的问题，目前用于评判与原图相似度的指标并不能支持这种想法。因此，有必要对重建结果进行人脸识别的测试，来评判在身份信息是否在重建结果中保留。在我们后面章节的实验中也可以看到，PSNR 与

SSIM 指标与人脸识别准确率指标并无强相关性。

想要重建的结果在取得良好视觉效果的同时能够被有效的识别，自然需要在训练过程中考虑保持模型输入输出图片之间的身份信息一致性。JunYu 等人的工作[55] 提出了一种 Joint Model，希望通过重建网络与识别网络相结合来达到在人脸重建过程中保持身份信息的效果。该模型的架构如图2-11所示[10]，模型包含了两个深度神经网络：重建网络 (SR Net)，输入低分辨率图片 I_i^l，输出重建后的图片 I_i^h；识别网络 (FR Net)，在模型的训练过程中将 SR Net 的输出 I_i^h 作为自己的输入，输出身份特征向量 x_i。

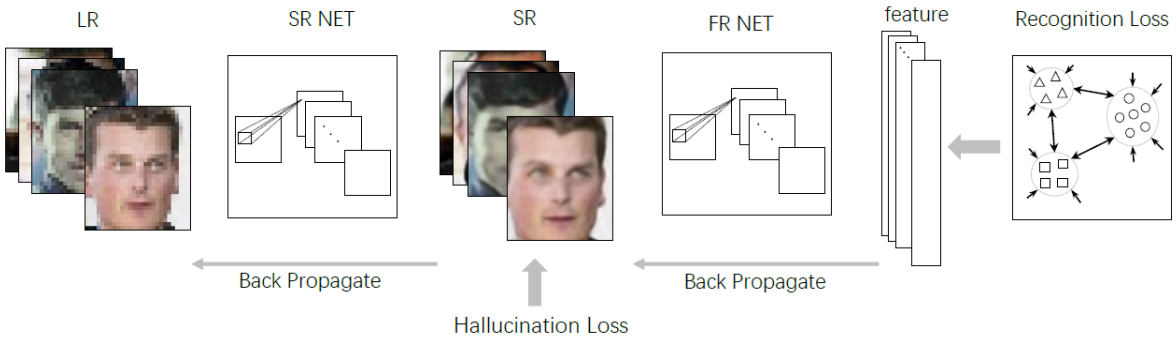

图 2-11: Joint Model 示意图

对于训练集中的某一张低分辨率输入图片 I_i^l，其对应的高分辨率原图为 $\tilde{I}_i^h$。图2-11中模型包括两部分损失函数，L_h 与 L_r。其中 L_h 保证重建结果的视觉效果，尽可能使 I_i^h 与 $\tilde{I}_i^h$ 相同，用公式表示如下:

$$L_h = \sum_{i=1}^{m} ||I_i^h - \tilde{I}_i^h||^2 \tag{2-16}$$

L_r 使用 cengter loss 使重建结果 I_i^h 保持与 $\tilde{I}_i^h$ 相同的身份信息。其数学表达式由两部分组成，表示如下:

$$L_r = -\sum_{i=1}^{n} \log \frac{e^{W_{c_i} x_i + b_{c_i}}}{\sum_{j=1}^{n} W_j x_j + b_j} + \sum_{i=1}^{n} ||x_i - m_{c_i}||^2 \tag{2-17}$$

其中 W_j 表示 softmax 对应权重矩阵 W 的第 j 列列向量，b_j 为对应的偏置系数。I_i^l 对应的高分辨率原图 I_i^h 属于人脸数据集中 c_i 类，m_{c_i} 表示 c_i 类对应特征空向的中心。center loss 相比于直接使用 softmax loss，更有利于特征空间中各个类的

类内距离减小。

2.7 本章小结

本章将与本文研究方向相关的几个领域都做了简要介绍，包括人脸识别以及超分辨率重建。首先介绍了传统的人脸识别方法，然后简要介绍了人工神经网络基础以及在本文的研究中最为常用的卷积神经网络。在介绍了卷积神经网络的基础知识后，对于近年来几种基于卷积网络的人脸识别模型原理做了介绍。之后，我们介绍了从传统方法到深度学习的超分辨率重建研究，以及如何将人脸特征引入到超分模型中以更好的对人脸进行重建。最后，简要介绍了基于超分辨率重建的低分辨率人脸工作相关的一个模型。通过本章的介绍可以看到，基于超分辨率重建的低分辨率人脸识别研究仍然存在很大的空白，也是一项具有实际应用价值的研究。在接下来的章节中，我们将就这一主题介绍我们自己提出的模型，并通过实验来验证其有效性。

第三章　基于参考图片的人脸超分辨率重建

第二章介绍的模型大都属于单张图片的超分辨率重建 (single image super-resolution，简称 SISR)。本章提出的模型与它们略有不同，属于基于参考图片的超分辨率重建方法 (reference-based super-resolution，简称 RefSR)，这一类方法在对低分辨率图片进行重建时，还会额外得到一张用于“参考”的高清图片，当参考图片与待重建图片内容相似时，模型能够从参考图片中学习部分特征信息，并用于重建结果中。我们给出了一套将这种基于参考的重建方法应用于人脸的方案，得到的新模型 (称为 Ref-Face Network) 对低分辨率人脸的重建结果在 PSNR、SSIM 指标上能够超越大多数模型，而在人脸识别的测试上也具有一定优势。

3.1　基于参考图片的超分辨率重建

单图片超分辨率重建 (简称 SISR) 在训练过程中输入低分辨率图片 I^{LR}，输出为重建的结果 I^{SR}，然后根据 I^{SR} 与原图 I^{HR} 计算损失函数来训练模型；在测试或实际应用中，除了 I^{LR} 以外再无其他信息输入模型。由于 I^{LR} 相比于 I^{HR} 通常存在着明显的信息损失，SISR 研究近几年一直难以取得突破性进展。

相比于 SISR 方法仅使用一张低分辨率图片作为模型的输入，基于参考的超分辨率重建方法使用另一张图片作为参考图片 (Reference image，后面简称 I^{Ref}) 来辅助重建的过程。通常来说，参考的图片需要与待重建图片有纹理或内容结构上的相似性。在目前的研究以及实际应用中，参考图片可以来自视频中的相邻帧[56]，也可以通过互联网检索得到相似图片[57]。RefSR 在重建过程中试图将 I^{Ref} 中的部分高频细节迁移到 I^{LR} 的重建结果 I^{SR} 中，由于来自 I^{Ref} 的高分辨率细节的辅助，RefSR 模型[58][59] 的表现通常具有一定竞争力。

我们本章的工作主要以 SRNTT 模型[58] 为基础，提出一种新的训练方法使其迁移到人脸超分辨率重建任务中。不同于以往一些基于参考的重建模型，SRNTT 并不会直接在图像像素层次比较 I^{Ref} 与 I^{LR} 的内容或纹理相似性，而是将他们

映射到特征空间后进行对比，希望找到更高阶的纹理特征相似性。模型对于每一组 I^{Ref} 与 I^{LR} 在特征空间中进行分块匹配，并将相似的特征块从参考图片交换到低分辨率图片 feature map 对应的区域中。在多个不同层次的特征空间进行 feature map 的相似特征块交换后，将得到的结果与重建模型对应层的输出以通道的方式进行拼接，使得最终的输出结果 I^{SR} 能够学习到 I^{Ref} 中某些高频特征。

由于 I^{Ref} 是高分辨率图片，与 I^{LR} 的分辨率不同，为了两者能够映射到同一个特征空间中进行特征相似性匹配，首先将 L^{LR} 通过插值法进行上采样得到与 I^{Ref} 相同的尺寸 (在我们的实验中需要 4 倍上采样)，记为 $I^{LR\uparrow}$。虽然此时 $I^{LR\uparrow}$ 与 I^{Ref} 具有相同的尺寸，但 I^{Ref} 相比于 $I^{LR\uparrow}$ 有更多纹理细节，直接进行特征匹配可能难以找到两者间相似的特征块。事实上，此时两者通常会在所有区域上普遍表现出较低的相似度。因此，在得到 $I^{LR\uparrow}$ 后我们还需要对 I^{Ref} 进行一些处理。先将 I^{Ref} 降采样到与 I^{LR} 相同的尺寸得到 $I^{Ref\downarrow}$，再通过插值上采样恢复到原尺寸，表示为 $I^{Ref\downarrow\uparrow}$。这样，我们通过在特征空间中对比 $I^{Ref\downarrow\uparrow}$ 与 $I^{LR\uparrow}$ 的 feature map，当 $I^{Ref\downarrow\uparrow}$ 的 feature map 在某个区域与 $I^{LR\uparrow}$ 相似时，便认为 I^{Ref} 与 I^{HR} 在对应区域上存在纹理特征相似。

将 $I^{LR\uparrow}$ 与 $I^{Ref\downarrow\uparrow}$ 分别映射到某个特征空间中得到 $\phi(I^{LR\uparrow})$ 与 $\phi(I^{Ref\downarrow\uparrow})$，并将它们分成若干图像块来计算相似度。对于 $\phi(I^{Ref\downarrow\uparrow})$ 第 j 个图像块 $P_j(\phi(I^{Ref\downarrow\uparrow}))$，$P_j(\phi(I^{Ref\downarrow\uparrow}))$ 与 $\phi(I^{LR\uparrow})$ 上各个图像块的相似度矩阵 S_j 计算如下：

$$S_j = \phi(I^{LR\uparrow}) * \frac{P_j(\phi(I^{Ref\downarrow\uparrow}))}{||P_j(\phi(I^{Ref\downarrow\uparrow}))||} \tag{3-1}$$

其中 $*$ 代表卷积运算，$\frac{P_j(\phi(I^{Ref\downarrow\uparrow}))}{||P_j(\phi(I^{Ref\downarrow\uparrow}))||}$ 相当于卷积核。对于上述计算得到的 S_j，$\phi(I^{LR\uparrow})$ 中以 (x, y) 为中心对应的图像块与 $\phi(I^{Ref\downarrow\uparrow})$ 第 j 个图像块之间的相似度用 $S_j(x, y)$ 来表示。根据公式(3-1)所描述的相似矩阵计算方法，我们构建一个经过交换的 feature map，表示为 M，并用 M 来表示 I^{LR} 经过 I^{Ref} 增强后的纹理特征。M 与 $I^{Ref\downarrow\uparrow}$、$I^{LR\uparrow}$ 在对应特征空间中的 feature map 尺寸相同，对于 M 中任意以 (x, y) 为中心的图像块 $P_{(x,y)}(M)$ 计算公式如下：

$$P_{(x,y)}(M) = P_{j^*}(\phi(I^{Ref})),\ j^* = \arg\max_j S_j(x, y) \tag{3-2}$$

注意，在公式(3-1)中，使用 $I^{Ref\downarrow\uparrow}$ 计算相似矩阵。但在上述公式(3-2)中，直接使用 I^{Ref} 计算特征交换后的 feature map。通过上述方式，与 I^{LR} 各个区域相似的特征被保存在 M 中，将 M 附加到超分辨率重建网络特定层的输出结果中，最终得到的重建结果 I^{SR} 将学习到来自 I^{Ref} 的某些特征信息。

由于 VGG19 网络[48]具有良好的纹理特征提取能力[60][61]，我们使用 VGG19 的 relu1_1、relu2_1、relu3_1 三层的输出分别作为 $\phi(.)$ 计算 M。输入 I^{LR} 的尺寸为 $A \times B$ 时，在 4 倍的超分辨率重建模型中存在三种尺寸的 feature map:$A \times B$、$2A \times 2B$、$4A \times 4B$。由于 VGG19 中存在最大池化层，对于输入的 $I^{LR\uparrow}$ 或 $I^{Ref\downarrow\uparrow}$ 尺寸为 $4A \times 4B$ 时，relu1_1、relu2_1、relu3_1 输出的 feature map 尺寸分别为 $4A \times 4B$、$2A \times 2B$、$A \times B$。这样，三种不同尺寸的 M 恰好可以以通道叠加的方式拼接到到重建网络不同层的输出中。图3-1[58]给出了 SRNTT 模型进行重建时的基本流程与原理，其中蓝色方框中是重建模型，篮框下方代表通过相似度进行特征交换的过程。通过 I^{Ref} 与 $I^{LR\uparrow}$ 计算得到的交换后的 feature map 是通过通道叠加的方式加到重建模型某些层的输出中，图中对层次结构进行了简化，只显示了需要与 M 进行通道叠加的网络层。

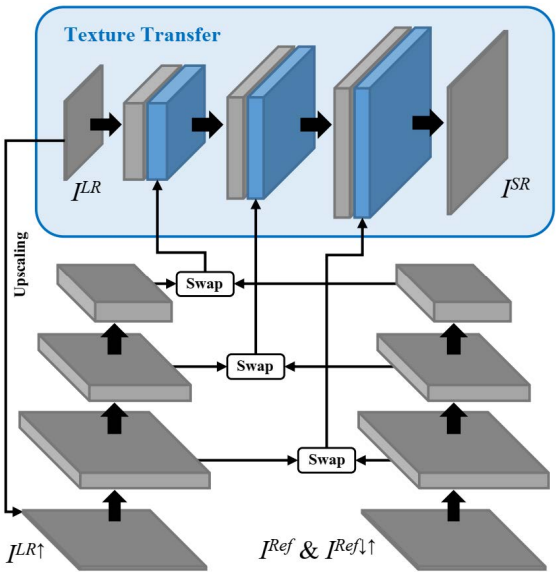

图 3-1: SRNTT 模型示意图

根据网络模型中各个层输出尺寸的不同，可以将模型分为若干模块，将第 l 个网络模块经过特征交换后的 feature map 表示为 M_l(可将 M_l 对应图3-1蓝色框内的灰色 feature map 部分)，对应网络层的输出表示为 ϕ_l，ϕ_l 的计算可用如下公

式表示：

$$\phi_l = [NetBlock(\phi_{l-1}||M_{l-1}) + \phi_{l-1}]\uparrow_2 \tag{3-3}$$

NetBlock 表示网络模块，‖ 表示两个张量进行通道叠加，$\uparrow_2$ 表示通过子像素卷积 (sub-pixel convolutional) 来将输入放大。因此，网络最终的输出，即重建结果 I^{SR} 的计算方式如下:

$$I^{SR} = NetBlock(\phi_{L-1}||M_{L-1}) + \phi_{L-1} \tag{3-4}$$

在我们的模型中 $L = 3$，ϕ_1 与 ϕ_2 的计算过程中已经经过了两次放大，使得 ϕ_2 的尺寸与最终需要的 I^{SR} 相同。因此，最终计算 I^{SR} 时无需再次进行放大。

在 SRNNT 模型的训练中，主要包括 4 部分损失函数：L1 损失，感知损失[62]，生成对抗损失[63]，纹理损失[58]。其中纹理损失 L_{tex} 通过重建结果 I^{SR} 在特征空间中的映射 $\phi(I^{SR})$，以及公式(3-2)计算的交换后的 feature map 计算得到，L_{tex} 促使重建结果从参考图片中学习到一些纹理特征信息。

3.2 基于身份信息的参考图片选取

超分辨率重建任务期望模型根据低分辨率图片重建出对应的高分辨率细节，某种意义上完成的是一种“无中生有”的任务。但很多时候由于低分辨率图片缺失了许多信息，而使得超分辨率重建任务面临“巧妇难为无米之炊”的困境。而基于参考的超分辨率重建方法通过参考图片提供部分高分辨率信息，一定程度上缓解了这一困境。

基于参考的重建模型其效果好坏很大程度上依赖于参考图片的选取，一般来说，如果参考图片 I^{Ref} 与希望重建出的原图 I^{HR} 在纹理特征上具有较多相似之处，则基于 RefSR 模型的结果通常具有优势；而如果 I^{Ref} 与 I^{HR} 的关联性不大，则 RefSR 模型难以保持优势，甚至某些 RefSR 模型在这种情况下的结果并不如大多 SISR 模型。

参考图片的选取在一般图片的 RefSR 任务上十分重要，且具有一定难度。一些时候可能需要人为的为输入图片选取参考，极大的妨碍了 RefSR 模型的实际应用。但在面对人脸数据时，参考图片的选取并不是一个困难的事情。由于人脸具有一些“普遍特性”，例如都有眼镜、鼻子、嘴巴、轮廓大多接近椭圆型，任意

两张属于同一个人的人脸图片就能够具有内容结构以及纹理特征的相似性，这使得基于参考的重建方法更适用于人脸超分辨率重建领域。我们在前一节所介绍的内容的基础上，用人脸数据集作为训练数据，并选取与 $\{I^{HR}, I^{LR}\}$ 属于同一个人且相似的人脸图片作为 I^{Ref}，使得模型在训练中学习到如何将相似的人脸特征转移到重建结果中，我们将得到的模型称为 RefFace。具体的，在训练阶段参考图片选取流程如算法3.1所述。

Algorithm 3.1 RefFace 模型训练时参考人脸的选取方法

Input: 训练集中任意一组数据 $\{I^{LR}, I^{HR}\}$，训练集中与 I^{HR} 属于同一个人的高分辨率图片：$\{I_1, I_2, ..., I_n\}$
Output: $\{I^{LR}, I^{HR}\}$ 对应的 I^{Ref}.

1: 对于所有的 $I_i (i = 1...n)$, 计算身份特征向量 $\phi(I_i)$;
2: 对于所有 $I_i (i = 1...n)$, 计算相似度 $m_i = \frac{\phi(I^{HR})\phi(I_i)}{||\phi(I^{HR})||\cdot||\phi(I_i)||}$;
3: 从上一步计算得到的 m_i 中找出其中最大的五个 $\{m_{k_1}, m_{k_2}, m_{k_3}, m_{k_4}, m_{k_5}\}$，从这五个中随机选取一个 m_{k_i}，将对应的 I_{k_i} 作为 I^{Ref}。

在算法3.1中，首先对于训练集中每一张图片 I 计算身份特征向量 $\phi(I)$。然后对于图片 I_i，将所有与其属于同一个人的图片的特征向量计算相似度。在得到的一组相似值中，选取值最大的五个。最后，从这五个相似值对应的图片中随机选取一个作为 I_i 的参考图片 I^{Ref}。这种结合了身份信息的参考图片选取方式主要有两方面优势：

首先，算法3.1使得 RefSR 方法能够十分方便的应用于人脸重建任务，避免了繁琐的人工。RefSR 方法在训练过程中本质上需要以 “三元组” $(I^{LR}, I^{Ref}, I^{HR})$ 为基本单位，类似的方法在计算机视觉领域里也曾出现，例如人脸识别领域的 Triplet Loss[64]，但三元组的匹配往往需要大量的人工数据处理工作。算法3.1则几乎不需要任何人工评判、分组的过程，整个流程都可以由机器自动完成。虽然需要对数据集中的每张图片计算身份特征，且第二步需要对数据集中属于每个人的 n 张图片计算 $C_n^2 = \frac{n(n-1)}{2}$ 次相似度计算，但对于支持 GPU 并行计算以及多线程处理功能的服务器而言，计算量在可承受范围内。

其次，算法3.1通过引入一定的随机性保证模型的充分收敛。算法在特征匹配度最高的 5 个中随意选取最终的参考图片，使得同一张图片 I 在训练的不同轮次中可能会匹配到不同的 I^{Ref}。这使得与图片 I 属于同一个人且最为相似的 5 张图片都有机会成为 I^{Ref}，以供模型学习它们中每一张图片与 I 之间的纹理相

似之处。在考虑一个机器学习问题的初期，一个很常见的想法就是先考虑人类在这个任务上的表现如何。由于人眼通常能够分辨出不同照片是否属于同一个人，这说明了属于同一个人的多张照片之间都具有很大的相似性。关联到我们的 RefSR 任务中，我们认为与 I 属于同一个人的图片都有一定必要让模型学习它们与 I 之间有哪些纹理特征相似处。但人脸数据集中属于同一个人的人脸可能出现差异较大的情况，例如不同年龄或不同拍摄角度，有些时候目前主流的人脸识别模型也不能保证将它们区分开。因此，为了我们的 RefFace 模型充分收敛，选取相似度最高的前 5 张图片作为参考图片的候选。

通过算法3.1为人脸数据集中每一张图片构建 $\{I^{LR}, I^{HR}, I^{Ref}\}$ 三元组，对前一节介绍的模型进行训练，得到的人脸超分辨率模型能够有效还原低分辨率人脸。特别地，尽管在训练过程中需要按照算法 3.1 所描述的流程对每一组数据选取参考图片，但在测试中只需要任意提供一张人脸图片模型便可以有效的进行重建，并不要求对测试中的每一张人脸输入提供属于同一个人的图片作为重建参考，这极大的提高了模型的实际应用价值。

在后面的实验章节可以看到，只要在训练中按算法 3.1 选取参考图片，在测试中是否提供属于同一个人的图片作为参考图片并不会对模型效果产生明显影响。

3.3 实验与分析

本章对使用前一节方法训练得到的人脸超分辨率重建模型进行实验分析。但在给出具体的实验结果之前，将先介绍一个对超分辨率重建模型进行测试时应该遵守的标准实验流程。我们将从两个方面对我们的 RefFace 模型进行测试：针对视觉效果的传统超分辨率重建指标，以及 (低分辨率) 人脸识别测试。

3.3.1 超分辨率重建领域数据预处理的标准流程

在介绍测试协议以及展示实验结果之前，有必要介绍一下我们认为在超分辨率重建研究的实验中应该遵守的一个处理流程，这种处理流程下训练得到的模型才能够有效应用于实际中。目前许多超分辨率重建研究都是人为降采样，得到高分辨率图片 I^{HR} 对应的低分辨率图片 I^{LR}，降采样的倍数通常为 4，我们的

实验数据也是同样的方法得到。但我们认为，有必要在进行模型的训练以前，完成对低分辨率图片的生成并保存为硬盘文件的形式。

具体来说，对于尺寸为 $A \times B$ 的高分辨率图片，从数据预处理到训练总共分为四个步骤：步骤一，从文件中读取高分辨率图片，用插值法将训练集中每一张高分辨率图片降采样为 $\frac{A}{4} \times \frac{B}{4}$ 分辨率的低分辨率图片 $\widehat{I^{LR}}$，将得到的图片存储为硬盘文件；步骤 2，从文件中读取高分辨率图片，作为 I^{HR}；步骤 3，从步骤 1 中存储的文件里读取与步骤 2 中 I^{HR} 对应的低分辨率图片 I^{LR}；步骤 4，将步骤 3 得到的 I^{LR} 输入到模型中，得到模型的输出 I^{SR}，用 I^{HR} 与 I^{SR} 计算损失函数以通过反向传播训练模型。

上述流程的关键是在于将低分辨率图片保存为硬盘文件，并在训练时直接从文件中读取。在测试阶段也同样将测试数据按照上述步骤 1 处理得到低分辨率图片的硬盘文件，然后直接从文件中读取数据并输入模型进行测试。在目前很多超分辨率模型的实验与测试中，将 I^{HR} 降采样并将得到的图片输入模型中的过程全部在计算机内存中进行，相当于用上述步骤 1 中 $\widehat{I^{LR}}$ 作为模型的输入。由于 CPU 进行插值法计算时精度较高，图片保存为文件后的像素值相比于插值计算得到的结果会有一定的舍入误差。也就是说，经过文件保存的 I^{LR} 相比于直接通过插值计算得到的 $\widehat{I^{LR}}$ 丢失了一定的精度信息，这部分“精度信息”来自于 I^{HR}，会有助于缩小 I^{SR} 与 I^{HR} 的差异。但在实际应用中，需要进行重建的是直接由摄像头拍摄得到的低分辨率图片，并不会有来自对应高分辨率图片的额外信息。图3-2给出了两种流程的示意图，(a) 为其他大多数论文的实验中采用的流程，(b) 为我们的实验中采用的流程。如果在实际应用中使用图3-2(a) 流程训练得到的模型，模型的输入与模型的训练数据具有一定的特征差异，重建结果并不理想。因此，在本文所有的实验中，均使用图3-2(a) 的训练流程。

3.3.2 超分辨率重建实验

超分辨率重建研究的初衷是让人眼看到高清的图像，但只通过人眼的主观视觉体验通常难以量化测评结果。因此，近年来超分辨率重建研究的成果通常由两种指标来客观评价重建图像的质量，分别为峰值信噪比[65](Peak Signal to Noise Ratio，简称 PSNR) 与结构相似性[66](Structural similarity Index，简称 SSIM)。

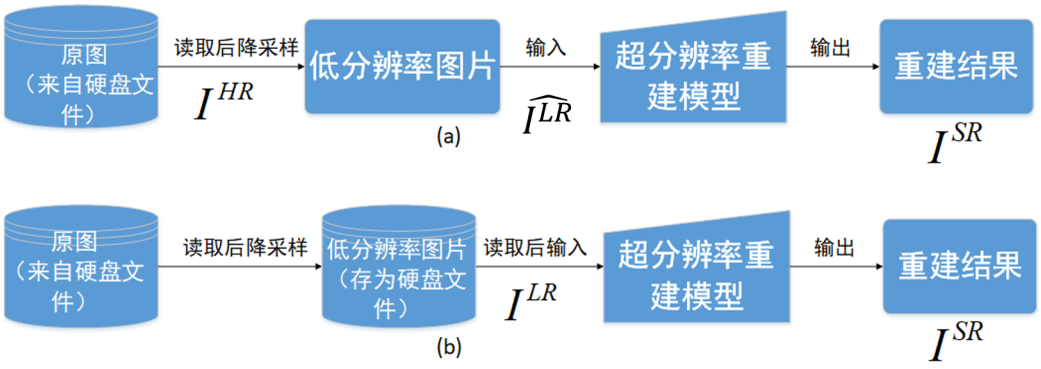

图 3-2: 处理流程对比

峰值信噪比 (PSNR) 通过两张图像之间像素的均方误差计算的相似度，对于两张同样尺寸的图片 P 与 Q，PSNR 计算如下:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|P(i,j)-Q(i,j)\right\|^2 \tag{3-5}$$

$$PSNR = 10\cdot\log_{10}\left(\frac{MAX_i^2}{MSE}\right) = 20\cdot\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{3-6}$$

其中 m,n 分别为图像的长与宽，MAX_I 表示图像颜色点的最大值，对于每个采样点用 8bit 表示的图像，$MAX_I = 225$。

结构相似性 (SSIM) 从图像亮度、对比度、结构三个方面来衡量两张图片的相似性。对于两张相同尺寸的图片 P 与 Q，亮度对比函数 $I(x,y)$, 对比度对比函数 $C(x,y)$，以及结构对比函数 $S(x,y)$ 分别定义如下:

$$L(X,Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \tag{3-7}$$

$$C(X,Y) = \frac{\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \tag{3-8}$$

$$S(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \tag{3-9}$$

其中 $\mu_X = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}X(i,j)$ 表示图像的均值，$\sigma_X^2 = \frac{1}{m*n-1}\sum_{i=1}^{m}\sum_{j=1}^{n}(X(i,j)-\mu_X)^2$ 为方差，标准差 $\sigma_X = \sqrt{\sigma_X^2}$。$\mu_Y,\sigma_Y^2,\sigma_Y$ 同理。C_1,C_2,C_3 为常数，为了避免分母为 0. 根据公式(3-7)，(3-8)，(3-9)，通常令常数 $C_3 = \frac{C_2}{2}$，两张图片的结构相似性

计算如下：

$$\begin{aligned} SSIM(X,Y) =& L(X,Y) * C(X,Y) * S(X,Y) \\ =& \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \end{aligned} \tag{3-10}$$

在 RefFace 的实验中，我们使用 CASIA-WebFace 数据集[6] 作为训练数据集，使用 LFW 数据集[4] 作为测试数据集合。CASIA-WebFace 数据集总共包含 494414 张图片，涉及 10575 个人。LFW 数据集是人脸识别领域一个经典的大规模数据集，最初是为了解决非限制人脸识别而建立，从互联网中抓取来自 5729 个人的 13233 张照片。其中超过 1680 个人的照片不止一张。LFW 所包含的图片具有更多变化因素，比如光线、表情等等，而在此之前的大多数人脸数据集大多是在实验室固定环境下招募志愿者采集得到。显然，非限制条件更附合实际生活中的人脸识别需求。由于 RefFace 模型在训练过程中较少引入身份信息，且我们用到的特征提取网络 VGG19 是针对一般图像的神经网络，因此本章的模型在提升视觉效果上更具有优势。本节首先将模型重建的结果用 PSNR 与 SSIM 两个指标进行测试，并与一些近年来主流的重建方法进行对比。我们将会在下一节给出重建结果在人脸识别上的测试结果。

本节实验我们选取的对比方法包括 WaveletNet[9]，SRGAN[11]，VDSR[12]，SICNN[15]，EDSR[13]，SRFBN[67]。在实验前，LFW 数据集与 CASIA 数据集中的数据均采用第3.3.1节所述的流程进行处理。经过相同训练集的训练之后，各个模型的人脸重建结果测试如表3-1所示。可以看到，我们的结果在 PSNR 指标上具有显著优势，在 SSIM 指标上的表现也尚可接受。

除了在两个超分辨率重建指标上能够取得效果，图3-3更为直观的展示了我们模型的实验效果。图3-3中分为三行，第一行为高分辨率原图，第二行为对应的低分辨率图片，第三行为将第二行图片作为 RefFace 模型输入后，模型输出的重建结果。可以看到，我们的模型能够重建出清晰的结果。

表 3-1: 超分辨率重建测试

方法名	PSNR	SSIM
Wavelet-srnet	29.53	0.831
SRGAN	28.28	0.748
VDSR	29.03	0.836
SICNN	28.27	0.763
EDSR	29.17	**0.844**
SRFBN	29.55	0.841
RefFace (ours)	**31.02**	0.80

图 3-3: Ref-Face 模型效果展示

3.3.3 低分辨率人脸识别测试

前一节通过 PSNR 与 SSIM 两个指标对 RefFace 模型的效果进行了测试，所用的两个指标主要针对重建图片的视觉效果，它们都属于误差敏感的评价指标。然而，在许多研究中的实验中也已发现，基于误差敏感评判方法与人眼的评判并非总是一致[68] [62]。由于人类的视觉系统是一个十分复杂的非线性系统，在计算机视觉领域至今仍无法找到一个能与人眼的评价保持一致的图像质量评价指标。然而，我们研究的初衷是希望通过超分辨率重建的方法解决低分辨率人脸识别任务，除了视觉效果相关的测评指标外，更有必要对重建结果的人脸识别准确率进行测试。

表 3-2: LFW 标准测试协议测试结果 (scale=4)

	LFW VERIFICATION
Original LFW	99.15%
Wavelet-srnet	95.83%
SRGAN	95.77%
VDSR	95.85%
EDSR	96.62%
SRFBN	96.07%
SICNN	**97.02**%
RefFace (ours)	96.37%

LFW 是一个用于人脸识别任务的数据集，有一个标准的测试协议。LFW 标准测试中，从整个数据集选取了 6000 对人脸，希望识别算法判断出每一个人脸对是否属于同一个人。在其中，有 3000 对是属于同一个人的人脸，另外 300 对人脸来自不同的人。在测试中，这些人脸对将被分为十组，每组分别进行测试，最后对十组的准确率取平均值作为最终的识别准确率。我们将 LFW 数据集经过第3.3.1节中的流程进行降采样以及重建后，对重建得到的结果进行 LFW 标准测试。测试结果如表3-2所示，我们的方法取得结果好于大多数对比模型。

在给定测试集的情况下，人脸识别的测试通常有两种方法：一对一验证 (人脸验证问题，face verification) 以及一对多验证 (人脸识别问题，face recognition，open-set identification)。在一对一验证中，对于待验证的输入图片，测试系统对应给出一张图片，要求人脸识别模型判断两张图片是否属于同一个人；在一对多验证中，假设数据库中一共有 K 个人的照片，需要人脸识别模型给出输入图片属于这 K 个 ID 中哪一个 (或是模型判定认为不属于其中任何一个)，一对多验证能够应用于更为广泛的场景。前面介绍的 LFW 标准测试协议显然属于一对一验证，但在由于判断属于同一个人的特征相似度没有样严格限定的阈值，使得测试只关注识别准确率，而没有保证足够低的误识率 (False Accept Rate，后面简称 FAR)。LFW 标准测试协议的评分高低并不足以证明方法能够应用于实际环境，目前也难以展现各个前沿识别模型的优势[3] [69] [52]。

表 3-3: RefFace 与各个对比模型在 LFW-BLUFR 上的测试结果

	FAR=0.1%, VR	FAR=1%, DIR
Original LFW	96.35%	78.94%
Interpolation	50.84%	14.46%
Wavelet-srnet	66.84%	26.36%
SRGAN	71.47%	36.49%
VDSR	73.07%	39.46%
EDSR	77.52%	42.27%
SRFBN	69.93%	35.42%
SICNN	76.03%	41.47%
RefFace (ours)	73.44%	31.97%

Liao 等人[70] 在 LFW 数据集的基础上提出了一个更为严格的测试协议，称为 LFW-BLUFR 协议。该测试分为一对一 (后面结果表格中 VR) 与一对多 (后面结果表格中 DIR) 两种测试，并对 FAR 进行严格的要求。在我们的低分辨率人脸识别测试中，我们将一对一测试的 FAR 限定为 0.1%，将一对多测试的 FAR 限定在 1%。这样，我们的模型与前面提到的各个对比模型在对低分辨率人脸进行重建后的识别准确率如表3-3所示。

表3-1所展示的结果中，我们的方法在各个对比方法中具有最高的 PSNR 测试值，SSIM 指标也高于大多数模型。但由表3-2与表3-3所展示的实验结果可以看到，我们的模型在低分辨率人脸识别中的表现还并不够突出。由于本章的研究属于基于参考的超分辨率重建模型，在模型的训练和测试中都需要对输入图片指定一张参考图片，且章节3.2介绍的训练过程中要求参考图片从 I^{HR} 同一个人的图片中选取。一个很自然的想法是，如果在测试中也同样选取一张与输入图片属于同一个人的图片作为 I^{Ref} 是否能够使重建结果具有更高的识别准确率? 我们同样对这种想法进行了实验测试，测试结果如表格3-4所示。我们看到，在测试中选取同一个人的图片作为参考，也并不能提升识别准确率。

以上测试结果说明了，我们提出的方法虽然能够重建出具有良好视觉效果的人脸，但在没有在重建过程中充分保持人脸图片中所蕴含的身份信息。但观察模型从训练到测试的整个过程，我们不难发现导致这种状况的原因。在模型的训练中，我们没有单独针对身份信息进行损失函数的约束。虽然在第3.2节中，我

表 3-4: 测试中对于 I^{Ref} 不同选取方式的实验结果对比

测试中 I^{Ref} 选取方式	FAR=0.1%，VR	FAR=1%，DIR
使用 $I^{LR\uparrow}$ (定义同章节3.1)	73.11%	29.74%
在测试集中属于同一个人的图片里随机选取一张	73.09%	30.15%
在测试集中随机选取一张	**73.44**%	**31.97**%

们通过人脸识别模型对人脸图片提取身份信息 $\phi(\cdot)$，并以此选取训练中的 I^{Ref}。但在对损失函数进行梯度下降的过程中，我们的方法更注重于让模型“意识”到 $\{I^{LR}, I^{HR}, I^{Ref}\}$ 三者间纹理特征的相似之处，而并没有强调身份特征的相似。基于参考的人脸重建模型 RefFace 能够重建出良好的视觉效果，具有一定的实际应用价值。尽管在低分辨率人脸识别上的效果不够突出，但本章所述的内容启发了我们后续的研究工作。我们将在下一章介绍一个新的模型，既蕴含 RefSR 方法的思想，同时能够将身份信息在超分辨率重建中的进行约束。

3.4 本章小结

本章提出了一种基于参考的人脸超分辨率重建模型。该模型以 SRNTT 为基础模型，将其应用于基于超分辨率重建的低分辨率人脸识别研究中。我们根据人脸数据集的特点，为每一组低分辨率输入选取合适的参考图片。得到的模型能够重建出更好的人脸图片，有助于提升视觉效果以及低分辨率人脸识别的准确率。虽然在训练阶段选取参考图片时要求与待重建图片属于同一个人，但在测试阶段无需这个要求，这使得模型能够应用于更广泛的场景。模型在视觉效果相关指标上具有突出表现，但在识别准确率方面不够突出。这种基于参考的思想，以及模型目前存在的不足，为后面的工作提供了启发。

第四章　基于身份信息的人脸超分辨率重建模型

本文前两章大致介绍了基于超分辨率重建的低分辨率人脸识别工作的目的，以及目前相关研究领域普遍存在的问题。在第三章中，我们根据基于参考图片的超分辨率重建方法 (Ref-based Super Resolution)，并将其与人脸重建任务结合，并通过在训练过程中对参考图片的选取设计特殊的规则，使得人脸重建取得了更好的效果。但由于模型的特性，其训练时需要对每一张图片单独进行特征提取与特征匹配。而其他大多数深度学习模型支持一批训练数据 (1 个 batch) 同时传入，在模型计算中通过相关软硬件支持可以并行优化。由于前一章的模型无法充分利用 GPU 并行优化的特性，在测试时运行时间较长，不利于对海量数据进行测试及实际应用。另外，RefFace 模型虽然能够对重建结果展现良好的视觉效果，但重建结果在识别准确率上并没有显示出充分的优势。本章提出了一种基于身份信息的人脸超分辨率重建模型，借鉴了前面介绍的 RefSR 模型的思想，但在训练中不只关注纹理特征的相似，而更多的关注是否在重建前后保持统一身份信息。本章提出的模型对于低分辨率人脸识别有更好的效果，也不存在前一章模型在训练与测试中计算繁琐的问题。

4.1　在人脸超分辨率重建中引入身份信息

在具体介绍我们提出的模型之前，本节先介绍一些相关的理论，同时大致介绍一下模型的设计思想。

想要在人脸超分辨率重建过程中确保人脸识别的准确率，一个最为直观的想法是：找一个具有身份标识的人脸数据集作为超分辨率重建模型的训练数据，然后在训练人脸超分模型时用模型输出的重建图片计算人脸识别损失函数，并将这个损失函数作为训练过程中总的损失函数的一部分，通过反向传播促使模型学习到识别身份信息的能力。前面第二章介绍的 JunYu 等人的工作[55] 就是直接将公式(2-17)所表示的人脸识别损失函数 center loss 在训练时作为重建模型损

失函数的一部分。但是，由于超分辨率重建与人脸识别这两个任务的相关性并不大，此时损失函数所包含的两部分损失 (人脸识别损失与超分辨率重建损失) 在反向传播时可能会将模型“拉”向两个不同的方向。在训练中，模型可能会希望“兼顾”两者而使得最终在任何一个任务上都没有突出表现。Zhang 等人的工作[15] 通过实验分析发现，这种直接用人脸识别损失训练超分辨率重建模型的方法会使模型无法充分收敛。

在 Zhang 等人的工作[15] 中，对于训练集中任意一组数据：低分辨率人脸与对应的高分辨率原图分别表示为 I^{LR} 与 I^{HR}，模型的输出表示为 I^{SR}，他们不要求人脸超分模型的参数能够最小化由 I^{SR} 与对应的身份类别计算得到的人脸识别损失，而只在训练过程中要求 I^{SR} 与 I^{HR} 在人脸识别中得到的结果尽可能相近。这种思想，得到的损失函数被称为 SI Loss(Super-Identity Loss)，用公式表示如下：

$$L_{SI}(I^{HR}, I^{SR}) = \left\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I^{HR})}{\| \phi(I^{HR}) \|_2} \right\|_2^2. \tag{4-1}$$

其中 $\phi(I)$ 表示将人脸图片 I 输入到人脸识别模型后网络输出的身份特征向量。相比于诸如公式(2-17)所表示的人脸识别损失，公式(4-1)表示的损失函数更容易被超分辨率重建模型所适应。在实验也同样表明，将 SI Loss 加入到人脸超分辨率重建模型的训练中将会提升低分辨率人脸识别的效果。虽然目前基于深度学习的人脸识别模型在训练中被看作一个分类网络，但在实际应用中往往会去掉最后的分类层，用网络的倒数第二层输出作为模型的输出，通过计算输出特征向量与原本保存的特征之间的相似度来判断识别结果。在这一背景下，对低分辨率人脸 I^{LR} 重建的过程中保持身份信息这一目标，某种程度上可以等同于希望 I^{SR} 与原图 I^{HR} 的身份特征向量具有足够的相似度。因此，将公式(4-1)加入损失函数来约束重建过程中身份信息的保持的做法，具有一定的合理性。

人脸识别任务相比于一般的识别任务 (比如对于 ImageNet[71] 的图像识别)，我们认为有两大特点：1. 包含的“种类”特别多，近几年的大规模人脸识别数据集中包含的不同 ID 的数量远超 ImageNet 等图像识别数据集中所包含的物体种类数; 2. 属于同一个类的图片，相互之间的相似度更大。如图4-1所示，图中前两列表示在普通的图像分类数据集中，同属于“瓶子类”的四张图片，而右侧两列则为人脸识别数据集中属于同一个人的四张不同照片。左侧四张图片，可能

图 4-1: 一般的图像数据集与人脸识别数据集

只是轮廓或基本造型比较相似，而右边的四张人脸图片，除了基本构造的相似性以为，五官等部位的细节也有极大的纹理相似性 (我们也是通过这种相似性来识别不同的人)。虽然左图中的图片也可以根据纹理细节区分为雪碧瓶、酒瓶等等，但在很多识别任务中并没不需要如此细致的分类。由此我们可以看除，人脸数据集的特点在于属于同一个类的样本，纹理特征具有更高的类内相似度。

由于上述的两个特点，想要取得更好的低分辨率人脸识别准确率，就必须尽量的对重建结果提取的身份特征向量增大类间距离，减小类内距离。此时，前一章所介绍的基于参考的超分辨率重建技术给我们提供了启发：在重建过程中，参考图片与低分辨率图片可以在某些图像块的纹理特征相似时进行参考，是否也可以在身份信息具有相似性时进行重建的参考？由于身份特征向量是对整个人脸进行特征提取得到，不再需要像前一章的模型那样先对图片分块再计算特征相似度，避免了许多繁琐的计算。在我们的模型中，将会以 SI-Loss 为基础模块，借鉴 RefSR 方法的思想，提出一个新的损失函数。

我们根据上面介绍的种种想法并参考近年来相关研究工作，提出了我们自己的基于身份信息的人脸超分辨率重建模型。由于我们的目的在于提升低分辨率人脸的识别准确率，因此不同于以往很多的人脸超分模型，我们并不过多的在意重建的视觉效果以及相关指标 (主要指 PSNR 与 SSIM)，而是更多的检验其在识别的准确率方面是否相比于原本的低分辨率图片有显著的提升。本章接下来几节将分别介绍模型的原理以及相关的实验分析。

4.2 模型架构

基于前一节所介绍的想法，我们提出了一个基于身份信息的人脸超分辨率重建模型。我们的模型涉及三个神经网络，其中两个网络构成一个 GAN 模型[63][11]，另有一个深度人脸识别网络，模型大致结构如图4-2所示。由于在训练过程中除了要对比重建结果 I^{SR} 与高分辨率原图 I^{HR}，还会用另一张属于同一个人的图片 (compare face) 与他们进行身份信息对比，我们将模型命名为 C-Face Newtork。

GAN[63](Generative adversarial network) 是最早由 Goodfellow 等人提出的一种通过生成对抗过程进行训练和估计的生成式模型。GAN 包含两个子网络: 生成网络 (G 网络) 与判别网络 (D 网络)。G 网络在训练过程中希望学习生成数据的数据分布，而 D 网络则希望判断出输入的数据是来自原本的数据集 (此时为高分辨率原图)，还是由 G 网络生成的伪数据 (此时输入 D 网络的是 G 网络的重建结果)。在对 GAN 进行训练的过程中，G 网络试图生成能够通过 D 网络判别的接近真实的数据，而 D 网络则希望判断出两种不同来源的数据。训练最终使得 G 网络生成的数据无法被 D 网络判别，即 D 网络判断准确的概率为 0.5，此时两个网络达到一种纳什均衡。在 SRGAN[11]、EDSR[13] 等模型中，GAN 结构被用于超分辨率重建任务。在超分辨率重建的 GAN 网络中，G 网络输入低分辨率图片 I^{LR} 而输出重建后的高分辨率图片 I^{SR}。D 网络在训练中试图判断其输入的图片是原始的高分辨率图片 I^{HR}，还是由低分辨率重建得到的 I^{SR}。

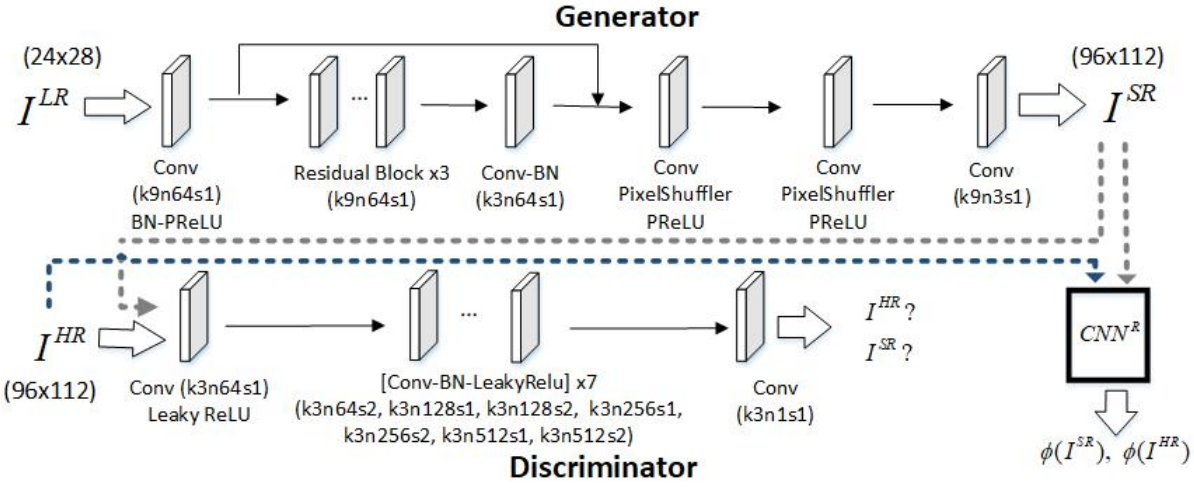

图 4-2: C-Face Network 模型框架

我们提出的模型 C-Face Network 中还包括一个人脸识别网络，在图4-2中表示为 CNN^R。理论上，CNN^R 可以是任何能够输出身份特征向量的深度人脸识别网络。我们在训练中引入一个人脸识别模型，来约束模型在对低分辨率人脸进行重建的过程中保持身份信息。类似于 GAN 的 D 网络，在训练过程中 CNN^R 模型的输入有两种情况: 原始的高分辨率图片 I^{HR}，以及经过重建得到的图片 I^{SR}。由于前一节所述的情况，我们并不会用 CNN^R 计算出损失函数来对 GAN 的生成网络进行反向传播训练，而只是用 CNN^R 计算身份特征向量，再用这个向量来进一步计算用于 G 网络优化的损失函数。

以上大致介绍了我们提出的模型的网络架构，以及各个模块之间的关系。我们的模型包括三个深度神经网络，对于人脸超分辨率重建任务来说，模型似乎显得有些“臃肿”。但在经过训练之后，在测试以及实际应用阶段，我们只需要调用 G 网络。需要注意的是，虽然我们在训练过程中使用了识别网络 CNN^R 来约束模型保持身份信息进而提高低分辨率人脸识别的准确率，同时在后面章节中将会看到，在训练过程中也会对该网络进行微调 (finetune)。但在实际测试过程中，我们并不会使用微调过的 CNN^R 作为测试模型。

在使用深度神经网络解决某个特定领域的问题时，主要有三方面研究: 网络结构、损失函数、训练数据采集。我们的研究主要针对损失函数，希望提出一个能够在人脸重建过程中更好的约束身份信息的损失函数。在训练数据方面，由于目前已存在较多的面向人脸识别以及超分辨率重建的数据集，且采集数据耗费成本巨大，我们没有亲自采集数据。但在实验过程中，我们将同样遵循第4.5.1中介绍的流程进行数据预处理。

在图4-2中所展示的模型架构中，GAN 网络的网络结构采用 SRGAN[11] 中使用的结构。其中 G 网络在接受输入 I^{LR} 后，首先是一个卷积 + 归一化 +pReLU 激活函数结构，对于通道数为 3 的输入图片 I^{LR}，卷积层的卷积核尺寸为 9，步长为 1，输出通道数为 64，在图4-2中用 “k9n64s1” 表示卷积的尺寸，其余几处标记含义类似。在此之后，G 网络将得到的通道数为 64 的 feature map 输入到 3 个连续的残差块结构中。残差结构在 ResNet[49] 网路中提出，其结构如图4-3所示。残差块结构通过在两个连续的卷积首尾处添加一个 “跳跃连接” (skip connection)，来解决神经网络深度不断增加时产生的梯度消失问题。

在三层残差块之后，有一个卷积 + 归一化结构。类似于残差结构的设计，

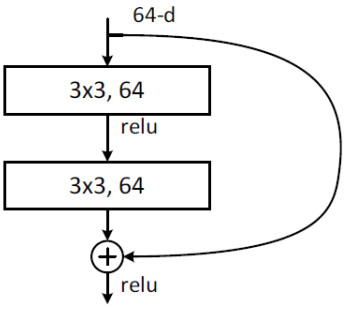

图 4-3: Residual Block 结构

在 G 网络中还会有一个更大的跳跃连接，将 G 网络中第一个卷积层输出的通道数为 64 的 feature map 与此时 (三个残差之后的卷积层的输出) 网络得到的 feature map 相加。这种结构一方面能够防止训练过程中出现梯度消失问题，同时对分辨率重建任务而言，还能够使得最终的结果充分利用网络浅层的信息，而不会因为网络过深而使得浅层信息对最终输出的直接影响较小。对于大多其他计算机视觉领域任务，例如分类、检测等等，超分辨率重建模型最终的输出与模型的输入有更为直接的关系，因此也有必要促使模型的输出中更多的考虑浅层信息。在跳跃连接之后，G 网络接下来有两个相同的模块：卷积层 + 归一化 +PixelShuffler+PReLU。其中卷积的输入通道数为 64，输出通道数为 256，步长为 1，卷积核尺寸为 3。PixelShuffle(像素重组) 是深度神经网络中用于 feature map 上采样的结构，其接受放大倍数作为其参数。两个模块参数设置完全相同，放大倍数都为 2。经过这两个模块，低分辨率输入的尺寸被放大 4 倍。最后，G 网络通过一个参数为 k9n3s1 的卷积网络，卷积网络后跟随 Tanh 激活函数，得到通道数为 3 的重建图片。D 网络的结构较为简单，由 9 层卷积层叠加得到。

需要在这里指出的是，我们后面提出的损失函数以及训练方法并不拘泥于某种特定的模型。在保持输入输出格式不变的情况下，将图4-2中 G 与 D 网络替换为不同的层次结构也同样可以取得保持身份信息的效果。下一节中，我们将介绍模型的损失函数以及训练过程的一些细节。

4.3 损失函数

C-Face Network 模型的损失函数分为两大部分，分别针对重建结果的视觉效果以及身份信息的保持。其中为了保证重建结果的损失函数包括两个部分，对

比损失 (adversarial loss) 与感知损失 (perceptual loss)，用于保持身份信息的损失是一个我们新提出的损失函数 (C-Face loss)。

4.3.1 对抗损失

生成对抗网络[63]，简称 GAN，是一种生成模型。前面章节已经介绍，GAN 模型包括两个子网络：生成网络与判别网络。两个子网络在训练过程中通过对抗的方式，逐渐使得生成网路 (G 网络) 生成能够 “以假乱真” 的数据。对于一般的 GAN，其初始输入为一个噪声，不妨表示为 z，希望生成网络的输出 $G(z)$ 为所需要的数据。GAN 的损失函数，即对抗损失函数 (adversarial loss) 如下表示：

$$min_{\omega_G} max_{\omega_D}\ L_{GAN} = \mathbb{E}_{x\sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (4\text{-}2)$$

其中 x 为真实的数据，$p(x)$ 表示其分布。在 GAN 的超分辨率重建任务中，G 网络的输入为低分辨率图片，判别网络 D 则需要进行判别其输入图片是原本的高清图片还是经过 G 网络重建得到。对于低分辨率图片 I^{LR}，G 网络对其重建的结果表示为 I^{SR}，对应的高分辨率原图表示为 I^{HR}，超分辨率重建任务的 GAN 网络的生成对抗损失用如下公式表示：

$$\begin{aligned} min_{\omega_G} max_{\omega_D}\ L_{GAN} &= \mathbb{E}[\log(N^D(I^{HR}))] + \mathbb{E}[\log(1 - N^D(N^G(I^{LR})))] \\ &= \mathbb{E}[\log(N^D(I^{HR}))] + \mathbb{E}[\log(1 - N^D(I^{SR}))]. \end{aligned} \quad (4\text{-}3)$$

其中 N^D 与 N^G 分别表示 GAN 结构的 D 网络与 G 网络，由于 G 网络的输出为重建结果 I^{SR}，公式(4-3)可以简写成第二行的形式。

公式(4-2)与4-3中的 ω_G 与 ω_D 分别表示 G 网络与 D 网络的参数。由于两个网络在训练中存在对抗性，他们对于公式(4-2)与4-3呈相反的优化方向。生成对抗损失的数学形式相比于其他损失函数显得更复杂一些，且有着严谨的数学理论支持，但其实现方式十分简单。值得一提的是，D 网络在实际应用中的输出通常并不是单一的值，而是一个矩阵，矩阵中包含若干个在 0 到 1 区间内的小数，每一个数值表示 D 网络对其对应感受野内图片真伪程度的判断。在软件实现中，

对于超分辨率重建任务，GAN 模型的优化方式可以如下表示：

$$\begin{aligned} &min_{\omega_G}\ ||N^D(I^{SR}) - valid||_2^2 \\ &min_{\omega_D}\ \frac{1}{2}\left(||N^D(I^{HR}) - valid||_2^2 + ||N^D(I^{SR}) - fake||_2^2\right) \end{aligned} \tag{4-4}$$

其中 $valid$ 表示一个与 D 网络输出矩阵的尺寸相同的全 1 矩阵，$fake$ 表示一个与 D 网络输出矩阵的尺寸相同的全 0 矩阵。由于 G 网络希望其输出的重建结果能够骗过 D 网络，在实际训练中表现为希望 $N^D(I^{SR})$ 为全 1 矩阵；D 网络则希望能够对真正的高分辨率图片 I^{HR} 输出为全 1 矩阵，而对重建的图片 I^{SR} 输出全 0 矩阵。公式(4-3)表示的对抗损失能够使 G 网络重建结果的特征分布更接近真实的数据，使得最终得到的结果更为逼真。通过对抗损失 L_{GAN} 与下一节将要介绍的感知损失，保证了我们的 C-Face 模型重建的结果具有良好的视觉效果。

4.3.2 感知损失

前面介绍的 GAN 网络以及相应的损失函数最初用于生成数据任务，如果只使用公式(4-2)到(4-4)描述的损失函数，并使用人脸图片作为训练数据，随后会得到用于生成人脸数据的生成模型。在必要的时候可以用这一类模型扩充人脸数据集，例如 DCGAN[72]，styleGAN[73]。但是，仅使用生成对抗损失并不能使得 G 网络学习到对低分辨率人脸重建的能力。超分辨率重建任务势必需要将重建结果 I^{SR} 与原图 I^{HR} 进行对比，并作为损失函数的一部分，最直观且简单的方法就是用 L_1 Loss 或 L_2 Loss，被许多基于深度学习的超分辨率重建模型所采用。

应用 L_1 Loss 或 L_2 Loss 训练得到的超分辨率重建模型，其输出结果能够获得较好的 PSNR 与 SSIM 指标，但这种这种结果通常过于平滑而未能重建出足够多的细节。在李飞飞等人的工作[62] 中提出了一种称为感知误差 (perceptual loss) 的损失函数，能够在超分辨率重建以及风格迁移任务中使得输出图片具有更多纹理细节。感知损失的原理如图4-4所示[62]，图中以风格迁移任务为例，但用于超分辨率重建任务时只会更加简单。

如图4-4所示，除了需要被训练的 “Image Transform Net” 模型，计算感知损失还需要一个已经训练好的 “Loss Network”，通常选取 VGG-16 网络，整个训

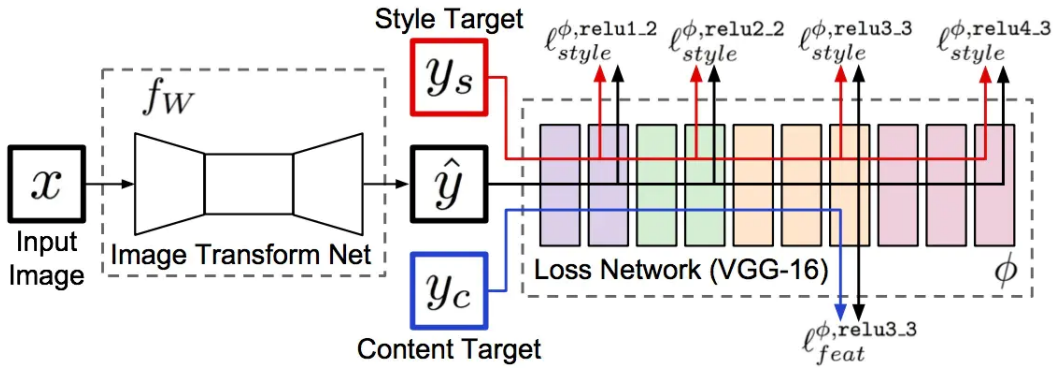

图 4-4: Perceptual Loss 原理示意图

练过程中这个网络并不参与训练，而只负责提供指定层的特征信息输出。风格迁移任务中，存在一张风格图片 y_s，一张内容图片 y_c，希望网络对于输入 x 输出的图像 $\widetilde{y}$ 能够用 y_s 的风格表现 y_c 的内容。但感知损失不用 $\widetilde{y}$ 直接与 y_s 或 y_c 进行比较，而是让 $\widetilde{y}$ 在 Loss Network 中某些层输出的 feature map 与 y_s、y_c 对应的 feature map 进行对比。在感知损失最初的论文[62] 中，$\widetilde{y}$ 与 y_c 的损失是通过对 feature map 计算 MSE 损失得到，而 $\widetilde{y}$ 与 y_s 的感知损失需要先通过对 feature map 做内积得到 Gram 矩阵，然后对两者的 Gram 矩阵计算 F 范数得到。在我们的模型中，只采用前一部分的感知损失。对于低分辨率人脸重建结果 I^{SR} 与原图 I^{HR}，模型的感知损失如下表示：

$$L_{perceptual-i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(I^{HR})_{x,y} - \varphi_{i,j}(I^{SR})_{x,y})^2. \tag{4-5}$$

其中，$\varphi_{i,j}(I)$ 表示 VGG 网络对于输入图像 I，其第 j 个卷积层之后且第 i 个最大池化层之前的 feature map。$W_{i,j}$ 与 $H_{i,j}$ 分别对应 $\varphi_{i,j}(I)$ 的长与宽。在我们的实验中，取 $i = 3$，$j = 5$。

感知损失的核心思想在于，希望 I^{SR} 与 I^{HR} 有更多高频特征的相似性，而不是得到过于平滑的结果。而在前面相关工作章节中，图2-8展示了深度神经网络在进行前向传播的过程中，不同的网络层会捕捉到不同级别的图像特征信息。通过使 I^{SR} 与 I^{HR} 在深度网络的 feature map 上有较高相似性，也就是保证了他们在一些细节特征上具有相似性，这有助于生成的图片包含更多真实的纹理细节。当然，感知损失函数也并不是没有缺点。在一些时候，由于输入图像某些地方过于模糊，使用感知损失可能会使生成得图片具有一些原本并不存在的细节，

如图4-5所示。图4-5[68] 中，左 1 为低分辨率图片，左 2 为 MSE 损失函数得到的结果，左 3 为感知损失得到得结果，最右侧为原图。可以看到，在头饰、围巾等地方，感知损失模型重建的结果相比于 MSE 模型有更多的纹理细节，具有更好的人眼观察效果。但许多纹理细节与原图并不一致，这使得感知损失得到的超分辨率重建模型在 PSNR、SSIM 指标上通常得分不高。在这种情况下，使用 MSE 训练得到的模型也未必能将真实的细节还原，很多时候只是通过将相应区域模糊化而获得更高的 PSNR 指标。

图 4-5: 感知损失的效果对比

感知损失能够使重建得到的人脸具有更多的纹理细节，I^{SR} 会因此有更好的视觉效果。同时，由于相比于 MSE 对每个像素点计算误差，感知损失函数在导数回传时，会更具有普适性，会使网络更快得收敛。

4.3.3 身份信息损失

本章前面几节内容介绍了模型的网络结构，以及损失函数的两个部分，即对抗损失函数以及感知损失函数。仅仅使用这两部分组成的损失函数来训练模型只能够保证生成的 I^{SR} 具有良好的视觉效果，而无法保证能够有更高的低分辨率人脸识别准确率。在前面的章节4.1已经提到，想要完成基于超分辨率重建的低分辨率人脸识别任务，势必需要在训练人脸超分辨率重建模型的过程中引入身份信息，以防止重建后的 I^{SR} 中丢失了原本 I^{LR} 中包含的身份信息。

诸如公式(2-15)所表示或表2-1列举的人脸识识别损失函数在训练网络时使得神经网络获得提取身份特征的能力，但直接用这些损失函数来训练超分变率重建模型则难以收敛。Zhang 等人的工作[15] 中通过公式(4-1)的形式促使重建结

果与原图具有统一的身份特征。但我们认为，应该进一步约束 I^{SR} 与属于同一个类的图片具有相似的身份特征，才能够促进重建结果在身份特征空间中具有更小的类内距离，进而提升低分辨率人脸识别的准确率。受前一章介绍的基于参考的超分辨率重建研究的启发，我们认为应该在对每一个样本训练时引入一张身份信息的参考图片，标记为 I_C^{HR}。这样，通过 $\{I^{SR}, I^{HR}, I_C^{HR}\}$ 的三元组来约束身份信息，构建一个有关身份信息的损失函数。

基于三元组的损失函数在深度学习中并不是新颖的事物，在 Schroff 等人的 FaceNet[64] 工作中提出的 triplet loss 就是通过对每一张训练数据 a 构建三元组 $\{a, p, n\}$ 来计算损失函数，进而使网络具有提取身份特征的能力，triplet loss 用公式表示如下：

$$L_{triplet}(a, pos, neg) = max(d(a, pos) - d(a, neg) + \text{margin}, 0) \tag{4-6}$$

其中 a (anchor，原点) 表示当前的训练图片，pos (positive) 表示一个与 a 属于同一个类别的样本，neg (negative) 表示与 a 不同类别的样本，$margin$ 为一个常数。在人脸识别模型的训练中最小化 $L_{triplet}$ 使得 $d(a, pos)$ 趋近于 0，而 $d(a, neg)$ 趋近于 $margin$，以此保证类内距离大于类间距离。

在人脸超分辨率重建中，数据集原本为成对图片 $\{I^{HR}, I^{LR}\}$。在计算损失函数时，我们使用低分辨率重建得到的 I^{SR} 与 I^{HR} 组成的二元组。这时对于数据集中每一个 I^{HR}，我们选取一张图片 I_C^{HR} 用于在接下来的损失函数中比较身份信息，组成了在我们的模型训练中特有的用于计算身份信息损失的三元组 $\{I^{HR}, I^{SR}, I_C^{HR}\}$。$I_C^{HR}$ 的下标 “C” 为 “compare” 的首字母，为了直观表示这张图片的用处；其上标 “HR” 为了表示它是一张高分辨率人脸图片。I_C^{HR} 与 I^{HR} 在数据集中属于同一个类别，即同一个人的不同人脸照片。分别对三元组的三张人脸图片用图4-2中的 CNN^R 提取身份特征向量，由于 I^{HR} 与 I_C^{HR} 属于同一个人，为了使重建后属于同一个人的图片在特征空间处于同一个流形中，我们希望 I^{SR} 的身份特征向量不仅与 I^{HR} 相近，同时也要与 I_C^{HR} 的特征向量具有较高

的相似度。因此，我们构建如下的损失函数：

$$\begin{aligned} L_{CF_1} =& \gamma_1 \left\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I^{HR})}{\| \phi(I^{HR}) \|_2} \right\|_2^2 + \gamma_2 \left\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I_C^{HR})}{\| \phi(I_C^{HR}) \|_2} \right\|_2^2 \\ =& \gamma_1 L_{SI}(I^{SR}, I^{HR}) + \gamma_2 L_{SI}(I^{SR}, I_c^{HR}) \end{aligned} \tag{4-7}$$

其中 $\phi(I^{SR})$，$\phi(I^{HR})$，$\phi(I_C^{HR})$ 分别表示三元组通过 CNN^R 网络提取得到的身份特征向量。我们使用特征向量归一化后计算 2 范数的方式来计算相似度，因此根据公式(4-1)中定义的 SI Loss，公式(4-7)的第一行可以简化成第二行的形式。这是我们提出的用于提高人脸超分辨率重建模型保持身份特征能力的损失函数，但也只是第一个版本。

在介绍如何对公式(4-7)所描述的 C-Face Loss 进行改进前，我们认为有必要先介绍一下我们打算如何为训练集中每一个 I^{HR} 选取 I_C^{HR}。Triplet loss 是深度学习领域通过三元组计算损失并以此提高模型提取身份特征能力的经典案例，但 Triple loss 用于训练模型时，如何构建三元组是一个困难但又至关重要的问题。在 FaceNet 论文的实验中，构建出的三元组又根据区分的难易程度分为 easy triplets，semi-hard triplets 与 hard triplets 三种。当数据量较大时，想要对训练数据中每一张图片都合理的构建三元组将需要巨大的工作量，否则很容易出现网络过拟合或欠拟合的情况，这也是在后来的研究中 triplet loss 没有被广泛使用的原因。

然而，在人脸识别任务中为庞大的训练数据集构建三元组的困难并不存在于人脸超分辨率重建任务中。为了让模型在重建中保持身份信息，需要选取一张与 I^{LR}、I^{HR} 身份信息相似度较高的图片，目前的人脸识别模型已经能够有效提取图片的身份特征。同时，通过前面4.1节图4-1可以看到，人脸数据集中属于同一个人的人脸图片相似度较高，不需要担心超分辨率重建模型无法收敛的问题。而对于一般的图片内容而言，属于同一个类的各个样本特征差距较大，难以利用本节讨论的三元组思想。我们首先对整个数据集中的人脸图片生成特征向量，与前面公式中相同，用 $\phi(I)$ 表示图片 I 对应的特征向量。对于 I^{HR}，在与其属于同一个类的图像中，找出与 I^{HR} 特征向量最为相似的 n 张图片，在模型训练时，从这 n 张图片中随机选取一张作为 I_C^{HR}。通过引入这个随机选区的过程，能够提高模型的鲁棒性。由于数据集中属于同一个类的图片 (即属于同一个人的不同照片) 时常有较大差异，因为年龄、光线、装扮等因素的变化，有时他们的

特征差异过大，会使得损失函数值难以在训练中下降。因此，n 的选取通常不要过大。如图4-6所示，这是我们从 CASIA-WebFace 数据集中选取了同一个人的几张照片，红框的是属于同一个人的照片中与绿框图片身份信息最为相似的三张。对于绿框标记的 I^{HR}，如果在 n 较大的范围内选取 I_C，用 C-Face Loss 在训练中会难以下降。正如我们不能直接使用形如公式(2-15)的人脸识别损失函数来训练人脸超分模型，我们在构建 $\{I^{SR}, I^{HR}, I_C^{HR}\}$ 计算用于保持身份信息的损失函数时也要注意超分模型能否对此收敛。在我们的实验中发现，$n = 3$ 取得的结果最好。

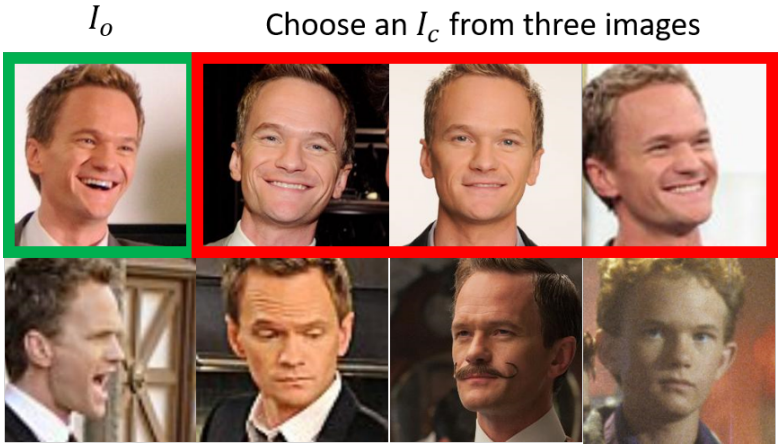

图 4-6: I_C^{HR} 的选取

公式(4-7)约束了两组身份信息的相似性：$\phi(I^{SR})$ 与 $\phi(I^{HR})$，以及 $\phi(I^{SR})$ 与 $\phi(I_C^{HR})$。其中，约束 $\phi(I^{SR})$ 与 $\phi(I_C^{HR})$ 的相似性本质上是因为 $\phi(I^{HR})$ 与 $\phi(I_C^{HR})$ 具有一定程度上的相似性。因此，我们认为，重建结果 I^{SR} 的身份特征向量与属于同一个人的图片 I_C^{HR} 身份特征的相似性不应该超过 $\phi(I^{HR})$ 与 $\phi(I_C^{HR})$ 的相似性。否则的话，会存在一定的过拟合行为，在训练中不利于 I^{SR} 恢复出 I^{HR} 所具有的一些细节。因此，我们对公式(4-7)中加号后的第二项做一个上限的约束，改进后的 C-Face Loss 损失函数公式表示如下:

$$
\begin{aligned}
L_{CF_2} =& \gamma_1 \left\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I^{HR})}{\| \phi(I^{HR}) \|_2} \right\|_2^2 + \\
& \gamma_2 \, max\left(\left\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I_C^{HR})}{\| \phi(I_C^{HR}) \|_2} \right\|_2^2 - \left\| \frac{\phi(I^{SR})}{\| \phi(I^{SR}) \|_2} - \frac{\phi(I^{HR})}{\| \phi(I^{HR}) \|_2} \right\|_2^2, \, 0 \right) \\
=& \gamma_1 L_{SI}(I^{SR}, I^{HR}) + \gamma_2 \, max\left(L_{SI}(I^{SR}, I_c^{HR}) - L_{SI}(I^{SR}, I^{HR}), \, 0 \right)
\end{aligned}
\tag{4-8}
$$

我们绘制了图4-7与图4-8来展现改进后 C-Face loss 的原理与效果。

图4-7通过与常规损失函数的原理进行对比，展现了 C-Face loss 的原理与特点。对于人脸超分辨率重建问题，以往的方法大多是约束 I^{SR} 与 I^{HR} 的一致性，如图4-7(a) 所示。这些方法要么试图为输出 I^{SR} 的深度模型设计更好的网络结构，要么试图对两者寻找某种更为合理的约束。即使部分模型在重建过程中考虑到身份信息的保持[15][10]，使用图4-7(a) 所示的模式训练得到的模型对于同属一个类的低分辨率人脸进行重建的结果在身份特征空间中的分布难进紧凑的处于同一个流形结构中。公式(4-8)使用图4-7(b) 的模式，能够使得人脸超分辨率重建模型对于低分辨率人脸识别具有更好的结果。

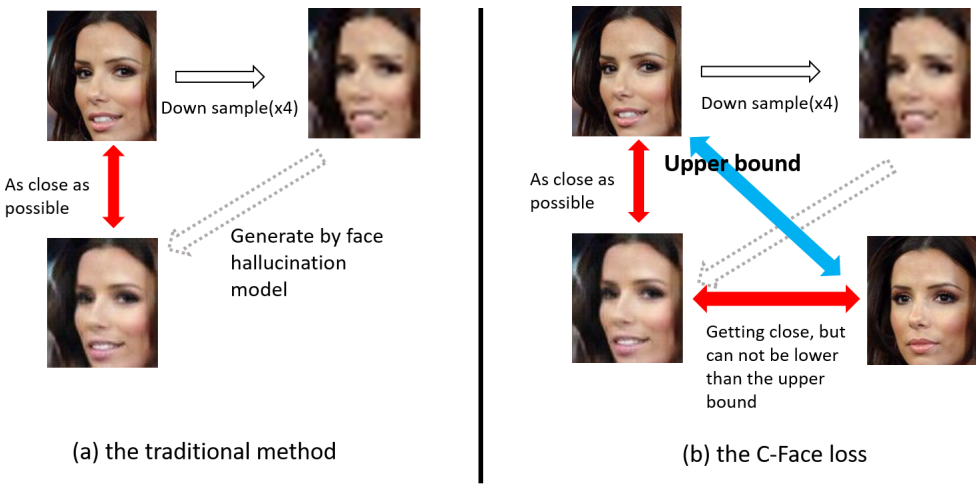

图 4-7: C-Face loss 原理示意图

我们希望通过图4-8更直观的展示我们期望通过 c-face loss 得到的效果。图4-8上下两部分分别对应我们的方法 (下) 与其他方法 (上) 训练前后的效果对比。其中，蓝色图标表示重建得到的图片的特征向量的位置，绿色图标代表高分辨率原图特征向量的位置，所有图片都属于同一个人。我们的方法得到的结果 (右下) 相比于以往的方法 (右上) 能够使得同一个类的重建结果在高维特征空间更紧密的处于同一个流形中。假设四张子图都代表着身份特征空间，图中每一个符号都代表着一张人脸图片的身份特征向量在特征空间中的位置。绿色表示高分辨率原图 I^{HR}，蓝色表示其对应的 I^{SR}，所有的图片都属于同一个人。假设上下两部分的左侧图片初始时刻 I^{HR} 与模型输出的 I^{SR} 的位置分布，图中箭头表示训练中的约束作用。对于普通方法 (上)，训练只能使每一个 I^{SR} 靠近 I^{HR}。但我们的方法希望 I^{SR} 在特征空间上靠近 I^{HR} 的同时也能够向属于同一个类的其他特征向量接近。我们希望这种方式，经过训练达到右上与右下的效果，

希望我们的损失函数训练得到的结果 (右下) 能够保证每一个类在特征空间中处于一个更为紧密的流形中。

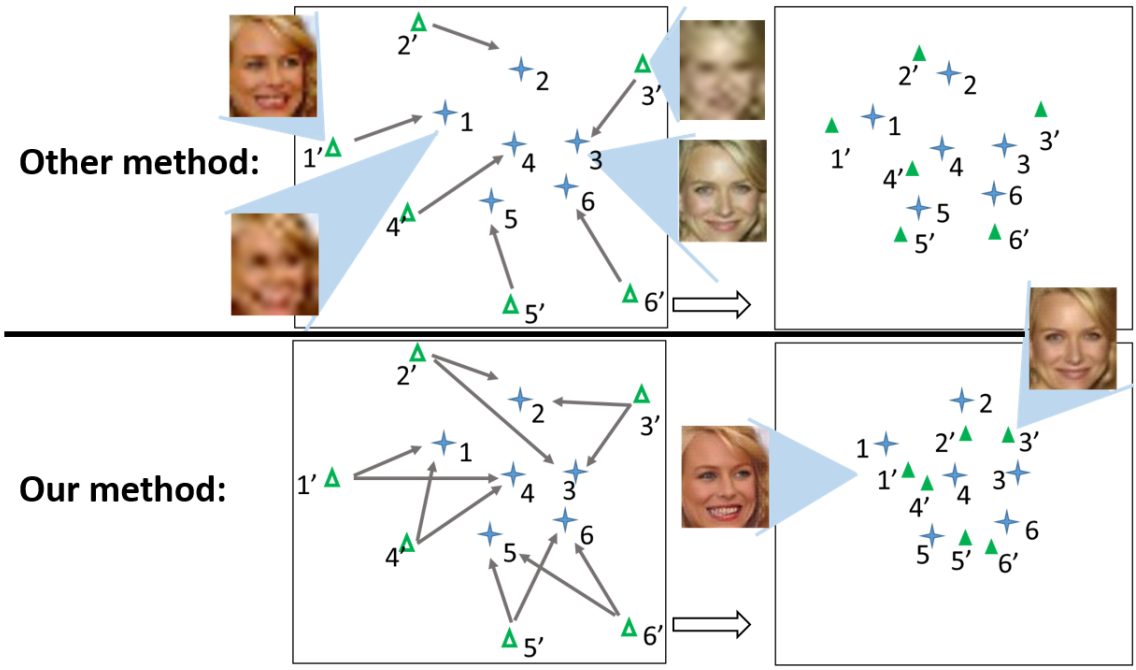

图 4-8: C-Face loss 效果示意图

从简单的对 I^{SR} 与 I^{HR} 计算 MSE 损失到 SI-Loss，再到我们提出的 C-Face Loss，我们认为这一系列研究是在人脸超分研究的一条“查找路径”上进行探索，目的是找到基于深度学习的人脸超分模型对于身份信息掌控能力的边界。这条路径目前已知由两个端点：其一是只约束 I^{SR} 与 I^{HR}，而另一个“端点”则是完全将人脸识别损失加入到人脸超分的训练中。因为实验已经证明这两个端点都不是合适的选择，才需要我们在不断尝试摸索出一个合理的中间位置。我们认为，本章的研究不仅仅提出了 C-Face-Network 模型，同时也提出了一种研究思路：将人脸识别、基于参考的超分辨率重建方法以及风格迁移的思想融入到人脸超分辨率重建中，并以此来解决低分辨率人脸识别问题。

当然，一个深度神经网络模型能否兼顾两种不太相关的任务，以及能够达到什么程度的平衡，一方面可能需要神经网络可解释性相关的研究成果，另一方面应该也与网络的规模以及结构设计有关。因此我们相信，基于本章的研究思路，还能得到更多有意义的成果。

4.4 训练流程

根据前一节的介绍，我们的模型完整的损失函数 L_{total} 如下表示：

$$L_{total} = \alpha L_{GAN} + \beta L_{perceptual} + L_{CF} \tag{4-9}$$

其中 β, α 都是常数。通过 L_{total} 对模型进行训练，能够使得 GAN 模型的 G 网络在训练结束后输出具有良好的视觉效果同时保持较高识别准确率的结果。但如果直接使用 L_{total}，在训练过程中观察损失函数各部分的结果可以发现，L_{CF} 会很快停止下降。原因在于 CNN^R 是使用高分辨率的人脸数据集训练得到，而重建得到的图片与原本就高分辨率的图片具有一定的差异。CNN^R 的训练使其具有将属于一个人的高分辨率人脸映射到特征空间的一个流形中，但它还不能很好的胜任将属于同一个人的重建图片 I^{SR} 也映射到与 I^{HR} 的同一个紧密流形中。因此，我们设计了一下训练流程，通过分阶段采用不同的训练策略，使得 CNN^R 逐渐生成更好的 $\phi(.)$，进而通过前面的 C-Face loss 使得模型在重建人脸过程中具有保持身份信息的能力。我们全部的训练流程如下：

Algorithm 4.1 C-Face Network 训练算法

Input: 经过高清人脸数据集 (D^{HR}) 训练好的 CNN^R，参数随机初始化的 GAN 网络 (结构如图4-2)，训练集中每一个数据都分为三元组 $\{I^{LR}, I^{HR}, I_c^{HR}\}$
Output: 训练好的 C-Face Network.

1: **stage 1**: 通过公式(4-9)训练 GAN 网络。CNN^R 只用于计算 C-Face Loss，其权重不参与训练；
2: **stage 2**: 用步骤一训练得到的 G 网络对训练数据中所有低分辨率图片进行重建，得到经过重建的数据集 D^{SR}；
3: **stage 3**: 将 D^{HR} 与 D^{SR} 混合，用混合后的数据集对 CNN^R 进行训练微调；
4: **stage 4**: 再次用 L_{total} 训练 GAN 模型，此时的 CNN^R 使用上一步微调后得到的模型。在本阶段，训练的每一步不仅通过三元组计算得到的 L_{total} 对 GAN 模型进行反向传播，也通过 I^{SR} 计算人脸识别损失，对 CNN^R 进行微调。

通过上述 4 个阶段的训练，得到我们最终的 C-Face Network 模型。算法流程4.1 所做的事情主要在不同阶段对用于提取特征的人脸识别模型 CNN^R 进行了微调训练，且分阶段进行。对 CNN^R 进行微调的理由前面已经说明，而将微调分阶段进行 (算法流程4.1 中的步骤 3 与步骤 4) 且每个阶段的微调方式并不相同，这样做的理由在于：我们认为随着模型的训练，生成的 I^{SR} 质量逐渐提

高，用于提取身份特征向量的 CNN^R 同步去适应当前模型重建人脸的特征分布特点。

通过算法 4.1 所描述的训练流程，使得图4-2中 GAN 网络的参数收敛效果更好。最终，G 网络输出的重建人脸能够有效解决低分辨率人脸识别问题。

4.5 实验与分析

为了证明我们前面提出的模型的有效性，将在本节展现我们的一些实验结果。首先将在下一节介绍如何在模型训练前准备好训练数据，并给出一些实验的设置。之后的几节中将分别对低分辨率人脸识别、超分辨率人脸重建等几个相关指标进行测试，将用我们的模型与人脸超分辨率重建以及一般的超分辨率重建方法进行对比。根据本节的内容，可以看到我们的模型的缺陷之处的同时也将看到我们的模型具有一定的实际应用价值。在介绍具体的实验结果前，为了读者对实验环境有充分了解，第4.5.1节与第4.5.2节将分别介绍数据的预处理以及一些模型训练细节以及超参数的设置。

4.5.1 数据预处理

根据我们的模型以及研究方向，我们选用 CASIA-WebFace 数据集[6] 作为训练数据，并选用 LFW 数据集[4] 作为测试数据，前面章节已经对这两个数据集做过简要介绍。虽然我们的 C-Face Network 模型属于一个人脸超分辨率重建模型，但我们真正的研究目的是解决低分辨率人脸识别问题，超分辨率重建只是我们使用的工具。因此，不同于以往的人脸超分辨率重建模型更多以 PSNR 与 SSIM 作为测试指标，我们更注重建结果在人脸识别中的准确率。

尽管近年来已经有诸如 MegaFace[5] 这样百万级人脸数据集，但人脸超分辨率重建模型更多的在于学习人脸的几何特征恢复，而并不能够学习到人脸识别的能力，CASIA-WebFace 的数据量已经足够作为我们任务的训练数据。原始的 CASIA-WebFace 数据集中每一张图片的大小为 250*250，数据集为每一张图片给出了类别信息以及人脸关键点信息。根据图4-6可以看到，原始的人脸图片包含很多一部分背景信息，目前通常的做法是：在人脸识别前，根据关键点信息，将图片剪裁成 96*112 大小的图片。我们将剪裁后 96*112 大小的图片作为训练过

程中的高分辨率原图 I^{HR}，我们将这些图片进行四倍降采样，得到 24*28 的图片作为训练过程中的低分辨率图片，作为 G 网络的输入。我们同样遵循第3.3.1节中所述的标准处理流程，将两种不同分辨率的数据预先准备好。

4.5.2 训练细节

在本文的实验中，相关模型的实现均在 PyTorch 框架下进行，实验的硬件环境为装有 NVIDIA 1080-Ti GPU 的服务器。在对 C-Face Network 模型进行训练时，我们统一使用 Adam 优化器，并使用衰减率 0.99，批量大小为 64。在算法 4.1 中，阶段 1 与阶段 4 的学习率分别为 2×10^{-4} 与 1×10^{-5}。阶段 1 总共训练 10 个 Epoch，阶段 4 训练 3 个 Epoch。在阶段 3 对 CNN^R 进行微调训练时，初始的学习率为 0.1，但每经过一个 Epoch 就会乘以 0.1，且一共进行 10 个 Epoch 的训练。对于公式(4-9)中的超参数，在算法 4.1 的所有阶段都是 $\alpha=0.1,\beta=1.0$。对于公式(4-8)与(4-8)所描述的改进前后的 C-Face Loss，在算法 4.1 的第 1 个阶段设置为 $\gamma_1=0.05,\gamma_2=0.1$，而在第 4 阶段中设置为 $\gamma_1=\gamma_2=0.05$。由于公式(4-8)是对于公式(4-7)的改进，使用公式(4-7)所描述的损失函数作为实验对比。在下面的实验结果中，如果没有特别说明，C-Face Network 默认使用公式(4-8)作为 L_{total} 中 L_{CF} 部分。

由于我们的研究是基于超分辨率重建的低分辨率人脸识别，自然涉及到重建模型的放大倍数。目前主流的人脸超分辨率研究中，通常在 2/4/8 中选取一个或多个，我们的研究主要针对放大倍率为 4 的情况。根据前一节对于数据预处理的描述可知，无论是对于我们模型的训练、测试过程，还是对于人脸识别的实际应用而言，96×112 已经属于“高分辨率图片”，对于这种分辨率而言，4 倍降采样后的 24×28 是一个合理的研究对象。如果将 96×112 缩小 8 倍得到 $12*14$ 作为低分辨率图片，此时模型的输入图片过于模糊。根据信息论的观点，无论是超分辨率重建亦或是低分辨率人脸识别，最终效果的好坏都取决于低分辨率输入中是否包含足够多的信息，而 $12*14$ 分辨率的图像显然难以包含足够多的信息。另一方面，当降采样倍数为 2 时，分辨率 48×56 的图片仍有一定的清晰度，在实验中可以看到，此时只需要用传统的插值法进行上采样，然后对采样的结果进行人脸识别测试，测试结果并没有比原图有太大的下降。此时使用基于深度学习的超分辨率重建方法，由于其需要巨大的内存占用同时运行时间明显长

于传统方法，显然不具有优势，自然也没有研究价值。基于以上讨论，我们有理由认为，在我们的课题中研究倍率为 4 的重建最有意义。在后面的实验中可以看到，我们的模型在 4 倍实验上相比于其他模型具有显著优势。

4.5.3 低分辨率人脸识别实验

我们选取的对比方法包括 WaveletNet[9]，SRGAN[11]，VDSR[12]，SICNN[15]，EDSR[13]，SRFBN[67]。图4-9展示了部分测试结果，从左到右依次为插值法，WaveletNet[9]，SRGAN[11]，VDSR[12]，SICNN[15]，C-Face-v1，C-Face-v2。仅通过直观的人眼观察，我们也可以大致对一些方法进行比较。我们可以看到，基于深度学习的方法得到的结果普遍优于传统的插值法 (插值法是许多软件中默认的图像缩放方法)；部分深度学习方法，诸如 VDSR[12]，得到的结果过于平滑，在 PSNR 与 SSIM 指标上这或许是有利的，但观察图4-9中第四列可以看到，这样会损失很多原本应有的细节。为了公平起见，我们的模型以及各个对比方法都是用相同的训练数据 (CASIA-WebFace 数据集) 进行训练得到，训练数据的预处理方式也均如章节4.5.1所述。

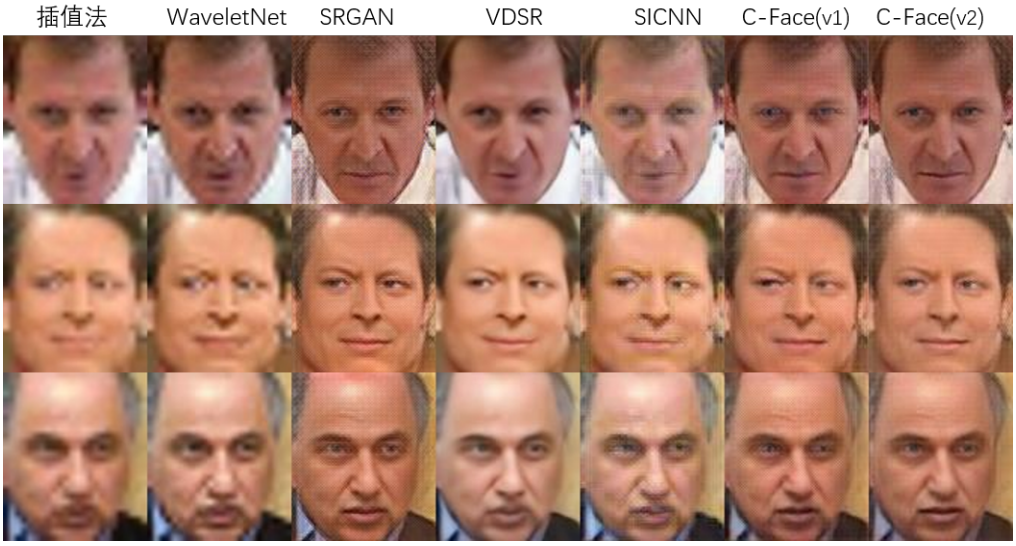

图 4-9: 几种模型的效果对比

但用人眼观察似乎难以辨别出各个重建结果对于识别准确率是否有提升，接下来我们对各个模型的结果进行人脸识别测试。在前面的章节中已经介绍了 LFW 标准测试协议，以及更有挑战性的 LW-BLUFR 测试协议，我们接下来将分别用这两种测试协议进行测试。表4-1给出了在 LFW 标准测试协议之下，各

表 4-1: LFW 标准测试协议测试结果 (scale=4)

	LFW VERIFICATION
Original LFW	99.15%
Interpolation	94.65%
Wavelet-srnet	95.83%
SRGAN	95.77%
VDSR	95.85%
EDSR	96.62%
SRFBN	96.07%
SICNN	**97.02**%
C-Face v1 (ours)	96.87%
C-Face v2 (ours)	96.72%

表 4-2: LFW-BLUFR 协议的测试结果 (scale=4)

	FAR=0.1%, VR	FAR=1%, DIR
Original LFW	96.35%	78.94%
Interpolation	50.84%	14.46%
Wavelet-srnet	66.84%	26.36%
SRGAN	71.47%	36.49%
VDSR	73.07%	39.46%
EDSR	77.52%	42.27%
SRFBN	69.93%	35.42%
SICNN	76.03%	41.47%
C-Face v1 (ours)	78.28%	39.87%
C-Face v2 (ours)	**78.66**%	**43.57**%

个模型的测试结果。其中 C-FACE 模型我们标注了“v1”与“v2”两种，其中“C-Face v1”表示仅通过公式(4-9)作为损失函数训练得到的模型，而“C-Face v2”则是按照算法4.1流程进行训练的得到的模型。从表4-1可以看到，我们的方法在 LFW 标准测试协议上，只是略低于 SICNN[15]，而优于其他各个方法。当然，后者也只比我们的模型高出 0.3%。

然而，如前面介绍的那样，LFW 标准测试协议并不是一个足够严格的测试，可以看到，即便时传统的插值法也能够取得 94.65% 的识别准确率，在 LFW 标准

测试协议上取得微弱的优劣势并不能真正说明问题。因此，接下来我们用 LFW-BLUFR 测试协议对各个模型进行测试。表4-2给出了各个模型在 LFW-BLUFR 测试协议下的测试结构，可以看到，在使用公式(4-8)所示的改进后的 C-Face 损失函数后，我们的模型在两个指标上都相比于其他模型占有优势；即使是使用公式(4-7)所示的改进前的 C-Face 损失函数，模型也在第一个指标上具有优势。

观察表4-1与表4-2的各项结果，容易产生两个疑问：其一，我们的模型在 LFW 标准测试协议上优势并不明显，而在更为苛刻的 BLUFR 测试下反而具有优势；其二，在表4-1的结果中，改进前的模型 (C-Face Network v1) 略高于改进后的模型 (C-Face Network v2)。我们认为这两个问题可以用相同的原因来解释：我们的模型的创新之处，包括 C-Face Loss 与算法 4.1，都更有利于人脸识别中的开集识别 (open-set identity test)。相比于 LFW 标准测试协议仅仅检测每一对人脸是否属于同一个人，我们的创新点更有利于重建结果的身份特征相比于其他类数据更有区分性。因此，在实际测试中对于 BLUFR 的第二个指标更有显著优势。同时，在 LFW-BLUFR 测试中取得更好的成绩相比于标准的 LFW 测试，更能说明模型实际应用价值。

4.5.4 视觉效果实验

前一节通过在 LFW 标准测试协议以及 LFW-BLUFR 测试协议的测试，证实了我们的模型能够有效应用于低分辨率人脸识别。但由于我们的模型归根到底仍然属于人脸超分辨率重建模型，因此还是有必要给出超分辨率重建领域的研究中常用指标的测试结果，即 PSNR 与 SSIM 测试。表4-3给出了各个测试模型在两个指标上的结果，可以看到，我们的模型在这两个指标上虽然不算突出，但与各个对比模型也并无明显差距，说明我们的模型输出结果的视觉效果仍然可以接受。

同时，将表4-3与表4-1、表4-2对比可以发现，模型识别效果的好坏与两个常用的视觉效果指标 (PSNR 与 SSIM) 并没有正相关性。例如，SICNN 模型[15]在表4-2显示的结果中，两个指标均高于 VDSR 模型[12]；但在表4-3中，VDSR 模型在两个视觉效果指标上均高于 SICNN。而另一方面，SRFBN 模型[67] 模型在 LFW-BLUFR 的两个指标上均低于 SRGAN[11]，但在 PSNR 与 SSIM 两个指标上均高于后者。PSNR 与 SSIM 是超分辨率重建领域最为常用的两个指标，但在

表 4-3: PSNR 与 SSIM 指标测试

	PSNR	SSIM
Wavelet-srnet	29.53	0.831
SRGAN	28.28	0.748
VDSR	29.03	0.836
SICNN	28.27	0.763
EDSR	29.17	**0.844**
SRFBN	**29.55**	0.841
C-Face v1 (ours)	28.66	0.740
C-Face v2 (ours)	28.62	0.802

面向低分辨率人脸识别的超分辨率重建任务上，这两个指标并不适用。因此，在这两个指标上不具有优势并不影响模型的有效性。

4.5.5 消融实验

为了充分证明本章模型的有效性，以及训练方法、损失函数设计的合理性，本节将分别对算法4.1以及 C-Face Loss 损失函数(4-8)中 I_C^{HR} 的选取方式进行消融实验。

在章节4.4中，我们为模型设计了一套训练流程。我们接下来通过拆解算法每一步的训练效果，来证明4.1算法的有效性。我们分别对算法4.1的四个子阶段的训练结果进行测试，测试的结果如表4-4所示。其中用“Version 1”表示只是用算法 4.1 中的 stage 1 来训练模型，相当于前一节实验结果中的“C-Face v1”模型；“Version 2”表示使用算法 4.1 来训练模型，即前一节对比实验中 C-Face v2，也是我们最终采用的模型；“Version 2-1”表示在算法 4.1 的 stage 3 中，只使用 D^{SR} 来对 CNN^R 进行微调；“Version 2-2”表示在算法 4.1 的 stage 3 中，使用 D^{SR} 重新训练出一个 CNN^R 以供后面的阶段使用；最后，“Version 2-3”则在 stage 2 以后，不进行算法 4.1 中的 stage 3，stage 4 照常进行而训练得到的模型。通过表4-4可以看到，采用整个4.1训练得到的模型在 LFW-BLUFR 两个指标上都取得了最好的结果。消融实验的结果说明了我们设计的训练流程能够得到表现更为优异的模型，充分的利用了 CNN^R 提取身份特征的能力，又能够使最终的模型结果不依赖于经过微调的 CNN^R。

表 4-4: 针对算法 4.1 的消融实验

	FAR=0.1%, VR	FAR=1%, DIR
version 1	78.28%	39.87%
version 2	**78.66**%	**43.57**%
version 2-1	78.44%	43.40%
version 2-2	77.60%	41.77%
version 2-3	74.25%	39.87%

接下来，我们将对 C-Face Loss 中参考人脸 I_C^{HR} 的选取方法进行实验测试。章节4.3.3介绍了如何对原本训练数据中每一对二元组 $\{I^{HR}, I^{LR}\}$ 选取 I_C^{HR} 来计算 C-Face Loss，并解释了采用这种选取方案的理由，这里我们同样给出实验分析来证明我们的想法。根据选取 I_C^{HR} 的基本思想是：从同属于同一类的数据中找一张与当前 I^{HR} 足够相似的图片，从最相似的 n 张图片中随意选取。在前面的对比实验中，我们采用 $n = 3$，接下来我们设置 n 分别为 1，2，3，4，5。除此之外，我们还尝试从同一个类的数据中完全随机选取一张图片作为 I_C^{HR}。通过以上几种不同的选取方案，用公式(4-7)作为 L_{total} 中的 L_{CF} 得到的结果如表4-5所示；用公式(4-8)作为 L_{CF} 得到的结果如表4-6所示。可以看到，不论使用改进前还是改进后的 C-Face Loss，$n = 3$ 都能取得做好的低分辨率人脸识别准确率。

表 4-5: v1 模型中对于 C-Face Loss 的消融实验结果

C-Face, v1	FAR=0.1%, VR	FAR=1%, DIR
$n = 1$	59.72%	24.43%
$n = 2$	76.25%	37.28%
$n = 3$	**78.28**%	**39.87**%
$n = 4$	72.89%	34.95%
$n = 5$	72.67%	34.76%
Randomly Choose	76.58%	36.08%

我们认为，在选取 I_C^{HR} 时，最佳的 n 取值可能也与不同的训练数据集有关。但总的来说，合适的 n 值应该保证对于给定的 $\{I^{HR}, I^{SR}\}$，I_C^{HR} 的备选不能与 I^{HR} 差异太大，否则用 C-Face Loss 在训练中将难以充分下降。我们认为这解释了为何在 $n \neq 3$ 时测试的结果并不理想。比较令人困惑的时当 n 值较小的情况，尤其

是表4-5与表4-6中当 $n = 1$ 时的结果，竟然相比于 n 为 2 或 3 时有明显的下降。我们初看到这一结果时也感到十分困惑，在进行 $n = 1$ 的实验之前，预计得到的结果至会比 $n = 2$ 时略有下降，结果居然在各个测试中均有 10 个左右百分点的下降。但通过观察训练数据的特点，我们给出以下解释：$n = 1$ 时就意味着对于 I^{HR} 选取同一个类中与之特征最为相似的图片记为 I_C^{HR}，图4-10给出了部分 $n = 1$ 时 I^{HR} 与 I_C^{HR} 的示例。对于图4-10中第一行的图片，数据集中与它们属于同一个人且相似度最高的图片如第二行所示。可以看到，在这些实例中，两张图片几乎完全相同。由于作为训练数据的大规模人脸数据集时在网上抓取得到的图片，来自于世界各国的演员、明星的新闻、社交平台、剧照等等。很可能会出现在同一地点很短的时间内拍摄多张相似度很高的图片这样的情况，图4-10的情况在 $n = 1$ 时并不少见。这会使得 $\phi(I^{HR})$，$\phi(I^{SR})$ 与 $\phi(I_C^{HR})$ 三者相似度过高而使得 L_{CF} 失去效果。

C-Face Loss 本质上是希望人脸超分辨率重建模型能够学习到两张高分辨率图片 (I^{HR} 与 I_C^{HR}) 的相似性，并对它们的身份特征差异值进行梯度下降。如果两张图片过于相似，那么它们身份特征向量的差异值接近于 0，C-Face Loss 自然也就失去了作用。况且由于超分辨重建任务存在一对多的病态问题，对于过于相似的两张图片，超分辨率重建模型无法学习到它们之间的差异。

因此，$n = 3$ 是一个最为恰当的选择，从两个方面防止 C-Face Loss 失去预期的作用。另外，由于数据集较大，也不能排除会出现对于某些类的某些图片作为 I^{HR} 时，$n = 1$ 时的差异刚好，而 $n = 3$ 时已经差异过大。通过从 $n3$ 张图片中随机选取一定程度上防止这种情况的出现，引入这种随机性一定程度上也从两个方面防止了 C-Face Loss 失去预期的作用。

表 4-6: v2 模型 (最终模型) 中对于 C-Face Loss 的消融实验结果

C-Face, v2	FAR=0.1%, VR	FAR=1%, DIR
$n = 1$	63.87%	29.75%
$n = 2$	72.99%	35.89%
$n = 3$	**78.66**%	**43.57**%
$n = 4$	77.52%	41.22%
$n = 5$	74.79%	38.85%
Randomly Choose	67.90%	33.51%

图 4-10: $n = 1$ 时选取的 I_C^{HR}

4.6 本章小结

本章介绍了一种基于身份信息的人脸超分辨率重建模型，称为 C-Face Network，通过在重建过程中充分保持身份特征信息的优势，在低分辨率人脸识别领域有巨大的实际应用价值。模型的创新之处主要有两个方面，包括一个基于身份信息的损失函数，以及一个新颖的模型训练流程。这两方面创新都是通过人脸识别与超分辨率重建模型结合，使得人脸超分辨率重建模型能在一定程度上 “分辨” 出身份信息，属于同一个类的重建结果在身份特征空间中处于一个流形结构。该模型借鉴了基于参考的超分辨率重建的思想，但在训练中对于参考图片的选取方式十分简单，且实际应用中无需选取参考。最后，我们通过实验证明了 C-Face Network 模型在低分辨率人脸识别上具有很大优势。相比于之前的重建模型，我们的方法不拘泥于视觉效果以及传统的指标 (PSNR 与 SSIM)，致力于解决基于超分辨率重建的低分辨率人脸识别问题，克服了前一章模型中存在的识别率较低以及计算速度较慢问题。

第五章　超分辨率重建与人脸识别系统

为了验证本文所提出方法的实用性，我们搭建了一套多模态身份验证系统，称为 RINC-ID。RINC-ID 系统包含多种前沿的识别方式，包括人脸识别、声纹识别、步态识别，能够面向更为广泛的场景。我们将前一章构建的模型 C-Face Network 嵌入到 RINC-ID 系统的人脸识别模块中，使得系统能够对拍摄到的较小人脸做重建处理，一定程度克服了人脸识别模型对低分辨率人脸识别准确率低的问题。通过该系统，证明了我们的研究成果具有很大的实际应用价值。本章将从相关背景、软件架构、硬件架构等多个角度介绍我们搭建的系统。

5.1　系统相关背景

正如第1章与第2章中所述，人脸识别已经融入到我们的日常生活中。无论是国家建设“天眼系统”，还是各种单位、场所的门禁系统，人脸识别这种身份验证方式都因其便捷性而时常出现。但根据我们的调查研究，目前日常中的人脸识别系统主要存在以下两方面问题：

第一，虽然近年来基于深度学习的人脸识别算法的研究不断取得突破，但由于新的模型参数量庞大、运行速度慢等原因，市面上许多“人脸识别机”仍然在使用传统的人脸识别算法。这种情况，在公共场所中普遍使用的人脸识别闸机最为常见。根据我们的市场调研，中低端价位且带有人脸识别的闸机，都是将传统算法集成到闸机内开发板中，作为“人脸识别一体机”出售。优点在于实现简单，且后台服务器压力较小甚至可以无需后端服务，商家可避免因为网络不稳定等因素而带来调试、售后方面的麻烦。但实际上，随着我国网络部署的不断升级，前后端分别部署相关模型并通过网络传输来获取识别结果在如今早已不是问题，且并不需要太多额外成本。在第2章中可以看到，近年来基于深度学习的人脸识别算法相比于之前的算法有显著的优势，使用前沿的模型已成为升级身份验证系统的必然需求。

第二，在人脸识别应用系统中，缺乏对于低质量人脸的特殊处理过程。由于人脸识别算法一般都有固定的人脸输入尺寸，识别机在摄像头捕获的画面中利

用人脸检测模型捕捉到人脸后，用图像处理方法 (通常为三次样条插值法) 将捕获到的人脸缩放为模型输入所需要的大小。当摄像头捕捉到的画面中人脸区域较小时，系统可能会直接放弃识别并在前端提示用户人脸过小。根据不同系统的设定，也可能对检测到的人脸进行尺寸放缩后同样进行识别。相对而言，拒绝识别低分辨率人脸的方案看似更安全，但在例如视频监控等应用场景下，画面中出现的人脸普遍比较小，此时将使识别系统无法工作；直接对低分辨率人脸进行缩放后识别也是不可取的方法，在第4章表4–2中的结果可以看到，直接使用插值法将低分辨率人脸进行缩放，其识别结果远不够理想。

基于以上问题，我们搭建了一套面向低分辨率人脸的身份验证系统。通过将前一章提出的模型融入到人脸识别系统中，我们的身份验证系统有更大的实用价值。

5.2 系统介绍

本节将对我们搭建的 RINC-ID 身份验证系统做一个全面的介绍，包括系统需求、工作流程、软硬件架构几个方面。

5.2.1 系统需求

作为一个可在实际中使用的身份验证系统，我们认为必须具备以下需求：第一，准确的人脸识别。目前人脸识别模型在大规模数据集上的识别能力已经超过人眼，有必要将前沿的人脸识别模型集成到系统中以保证实际应用中识别的准确率；第二，低分辨率人脸的处理。在某些实际场景下，拍摄到较小的人脸是普遍出现的情况，一般的人脸识别模型应对的都是高分辨率人脸。此时为了保证识别的准确率，在输入到识别模型前需要先对低分辨率人脸进行重建处理。第三，实时运行。在诸如门禁等场景中，只有在保证准确率的同时保证实时性，才能充分体现人脸识别相比于传统方法的优势；第四，可扩展性。生物信息识别在不断发展，包括声纹识别、步态识别等多种方式逐渐趋于完善。兼容多种识别方式既能提高安全性，同时也能够促使系统应用场景更加广泛。

我们提出的 RINC-ID 身份验证系统能够满足以上四方面需求。相比于仍在使用传统人脸识别算法的部分产品，我们的 RINC-ID 系统集成基于深度学习的

人脸识别算法[3][53][69]，能够随着学术界的发展不断更新算法。RINC-ID 系统使用本文工作提出的 C-Face Network 模型对拍摄到的低分辨率人脸进行重建，保证系统对于低分辨率人脸识别的准确率，同时也证明了人脸重建模型的实用性。为了在使用深度学习模型的同时保证系统实时识别的能力，我们将系统分为前后端部署。前端负责采集数据以及一些简单的预处理工作，后端服务器搭载 GPU 显卡运行相关模型进行重建以及特征提取工作。前后端通过网络通信传输识别的结果，在百兆乃至千兆路由得到普及的今天，网络传输带来的延迟并不会影响身份验证系统的实时性要求。在可扩展性上，在5.3节所展示的界面中可以看到，我们的 RINC-ID 系统包括人脸识别、步态识别等多种识别方式，能够灵活应用于不同的场景。但由于我们的研究课题是“基于超分辨率重建的低分辨率人脸识别”，且本章主要目的是体现我们提出的人脸超分辨率重建模型的实用性，本章接下来内容仅围绕 RINC-ID 系统的人脸识别模块展开。

5.2.2 使用流程

本小节介绍 RINC-ID 系统运行的流程，分为注册与识别两个阶段。其中系统的人脸识别模块在注册阶段运行流程如下：

前端部分：用户输入姓名、ID，并正脸朝向镜头拍摄照片。若未在拍摄到的画面中检测到人脸，则提示重新拍摄，否则根据检测到的人脸位置，将人脸所在的长方形区域剪裁。剪裁的结果随用户填写的信息一同发送给后端，格式为 {name:***，id:***，img:***.jpg}。

后端部分：首先对前端传来的人脸图片大小进行判断，如果判定人脸过小，则在人脸对齐之后将人脸缩放为 24×28 尺寸 (图5-1中“对齐与缩放 1”)，则进行在人脸对齐后直接缩放为 96×112(图5-1中“对齐与缩放 2”)。经过上述步骤后，将分辨率为 96×112 的人脸输入到人脸识别模型中，得到长度为 512 的特征向量。将特征向量与姓名、id 信息一同存入到数据库中 (id 为主键)。

类似的，人脸识别模块在识别阶段运行流程如下：

前端部分：用户站在镜头前拍照，前端系统对拍摄到的照片进行人脸检测，如为检测到人脸则提示重拍，否则将图片编码后发送给后端;

后端部分：对于从前端接受到的人脸，识别阶段对图片生成长度为 512 的特征向量的流程与注册阶段一致，此处不再重复。在得到特征向量后，将特征向

量与后端数据库中已经存储的各个特征向量进行相似度对比，如果相似度大于预先设定的阈值，则识别成功，向前端返回数据 {True, Name:***, Id:***}，否则认为识别失败，向前端返回 {False, Name:None, Id:None}。

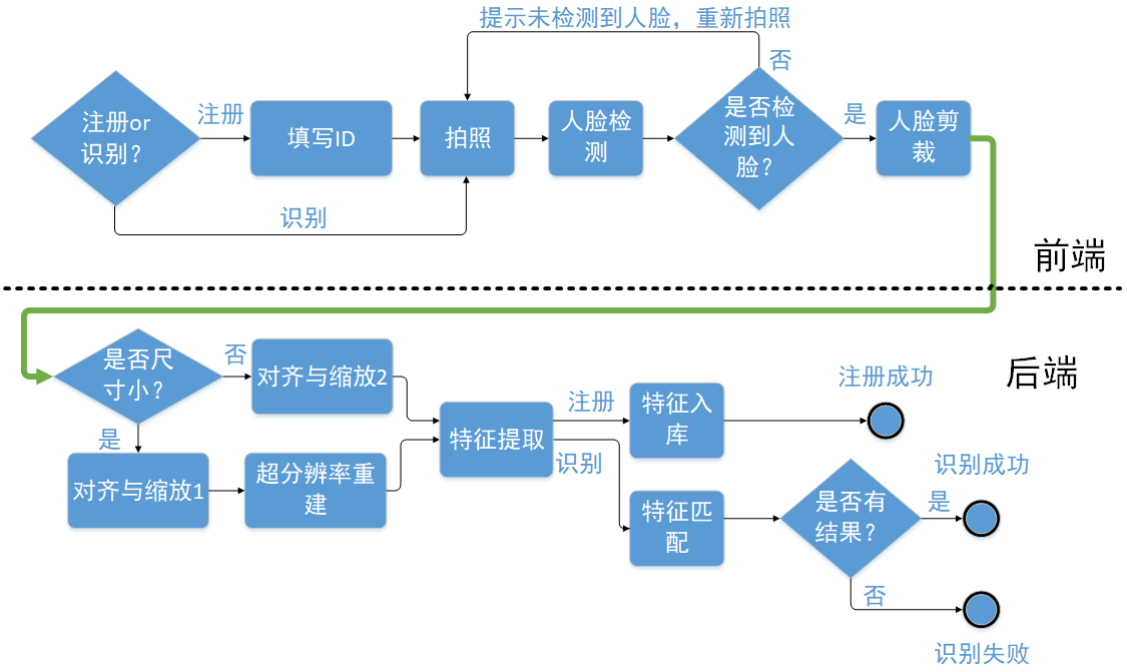

图 5-1: 系统使用流程

图5-1展示了上述的流程，其中绿色箭头代表该处需要进行网络传输。在前面的流程说明以及图5-1中，我们对于注册与识别阶段所遇到的尺寸较小的人脸进行重建，将重建后的结果作为特征提取的输入。但为了保证识别时的效率与准确率，我们建议在注册阶段尽量拍摄清晰的人脸。

在识别阶段需要对前端传来的人脸与库中特征进行相似度匹配，我们使用余弦相似度来计算。当前接受到图片提取得到的特征向量表示为 α，对于数据库中任意一个已有的特征向量 β_i，两者的余弦相似度 s_i 的计算公式如下：

$$s_i = \frac{\alpha \times \beta_i}{|\alpha|_2 \cdot |\beta_i|_2}. \tag{5-1}$$

在上述两个阶段的流程介绍以及图5-1中可以发现，其中有两个步骤分别被称为“对齐与缩放 1”与“对齐与缩放 2”。这两个步骤的原理相同，仅在最后缩放得到的尺寸有所区别。对齐的目的在于对人脸图片中不同的倾斜程度、五官位置做适当的矫正，使得后续模型的输入中人脸大小、倾斜程度等因素基本保持一致。首先通过 Dlib 人脸检测得到关键点的位置，dlib 库进行检测时能够

获得 68 个特征点，涉及眼、鼻、口以及轮廓。我们选取鼻子与上嘴唇两个特征点 (x_1, y_1) 与 (x_2, y_2)，以及左嘴角 (x_3, y_3)、右嘴角 (x_4, y_4)，首先用如下公式计算旋转角度 θ:

$$\theta = arctan\frac{x_1 - x_2}{y_1 - y_2} \tag{5-2}$$

相似变换能够对图像进行旋转、平移、缩放，对于原图中的任意一点 (x, y)，进行相似变化后的位置 (x', y') 计算如下：

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s\ cos(\theta) & -s\ sin(\theta) & t_x \\ s\ sin(\theta) & -s\ cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{5-3}$$

其中 (t_x, t_y) 为平衡矢量，s 为缩放因子，$s = \frac{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}}{47}$。近年来几个主流人脸识别模型[3][53][69] 的输入尺寸均为 96×112，在经过上述相似变换后，对于原本分辨率较高的图像，用 OpenCV 内置的 resize() 函数将图片尺寸缩放为 96×112；而对于低分辨率人脸，在相似变换后缩放为 24×28，这样经过 C-Face Network 模型重建后的分辨率也恰好为 96×112。

5.2.3 软件架构

整个系统使用 python 语言开发，前后端都需要 python 语言环境以及 Numpy 库、python-opencv 库用于一些数据处理。前端主要功能为用户界面的展示，以及一些简单的数字图像处理。如图5-2所示的前端界面使用 py-Qt 库结合 Qt-Creator 工具设计完成。

前后端通信通过 ProtoBuf 序列化需要传输的数据进行通讯。ProtoBuf 全称为 “protocol buffers”，是一种高效的数据序列化机制，且具有语言无关、平台无关的特点。ProtoBuf 与 XML 类似，但比 XML 占用空间更小、且速度更快，用于数据通信时更为高效。使用 ProtoBuf 进行通信具有高度的灵活性，可以自定义系统需要传输的数据结构。用类似于 json 的格式将需要传输的数据结构创建为.proto 文件，通过 protoc 编译即可获得文件读写接口。

后端主要设计三大模块：人脸重建，身份特征提取以及数据库。人脸重建模型采用上一章我们提出的 C-Face Network 模型，身份特征提取使用人脸识别

模型 SphereFace[3]。经过近几年的发展，已经有很多成熟的深度学习框架，例如 Tensorflow[74]，MaxNet[75]。对于本系统所涉及到的两个深度学习模型，我们采用 Pytorch[76] 框架进行开发。Pytorch 是基于 python 语言的深度学习框架，相比于其他框架，所办含的函数功能丰富，构建模型的步骤更为简单易懂，有利于快速实现模型。

为了在面对大规模用户量的同时保持系统在进行身份特征检索时的及时性，我们的系统采用非关系型数据库 levelDB 存储数据，并通过 Plyvel(levelDB 的 python 接口) 进行管理。LevelDB 是谷歌公司开源的一款非关系型数据库引擎，近年来许多都使用 levelDB 作为底层存储引擎。LevelDB 能够完成键值数据的高性能存储，相比于 MySql 等传统的关系型数据库，在快速检索方面有着巨大优势。同时相比于近年来火热的非关系型数据库 Redis，levelDB 很大程度上解决了 Redis 在持久化存储方面存在的问题。采用 levelDB 作为后端数据库，既有利于系统在未来应对大规模的用户，同时也有利于未来将存储从单机扩展为分布式部署。

以上介绍了系统在软件实现中主要需要用到的工具，部分软件包在安装前还需要其他其他软件库的支持，推荐使用 Anaconda 工具配置前后端所需的软件环境。前端可以在 windows 或 linux 系统上搭建，后端由于驱动等问题，需要在 linux 环境上搭建。

5.2.4 硬件架构

我们的系统采用前后端分离部署，通过网络进行通讯。硬件设备可以多种选择，但总的来说前端设备注重轻便、部署方便，后端设备注重系统稳定与算力充足。

前端设备我们有两种方案，其一是树莓派 + 外接摄像头 + 屏幕。经测试，在树莓派 3b+ 与树莓派 4 都可以部署前端系统；摄像头使用罗技 C920e，拍照像素较高且对亮度等因素具有一定预处理能力；屏幕可以根据部署场景选择合适的型号与尺寸。这种方案的优势在于有更高的灵活性，几乎每个部件都有选择的余地，可以根据价格以及部署环境等因素进行选择。另外，人脸识别的实际应用场景广泛，前端设备的部署需要适应实地环境，这种嵌入式 DIY 的方式更有利于灵活部署。

另一种方案中，我们直接使用微软公司生产的的 surface 平板电脑作为前端设备一体机。该方案在成本上自然远高于前一种方案，但在部署过程中无需考虑前端各个模块如何相互适配。这种方案同时体现出我们的系统具有跨平台特性，在各种常见的 PC 设备上均可以部署前端，为系统的广泛应用提供了更多可能性。

后端部署的服务器装有英伟达公司生产的 1080-TI 显卡，为运行人脸识别以及人脸重建两个模块各自的深度学习模型提供充足的算力。

5.3 效果展示

图5-2至5-5展示了系统各个流程的效果。其中图5-2为初始界面，对于已注册用户，点击左侧进入相应的识别模块。图5-2左侧区域包含三个按钮，分别为人脸识别 (图中“Recognize Face”按键)、声纹识别 (“Speaker Verification”)、步态识别 (“Recognize Gait”)、侧脸识别 (“Face Rotate”) 功能对应的进入按钮。对于新用户而言，首先需要进行注册：在右侧填写姓名 (Name) 与唯一的 id 后，点击“login”进入图5-3所示界面。

与图5-3左侧的识别区相似，图5-3中对应多个模块的注册的按钮，点击右侧“Register Face”后进入人脸注册界面。

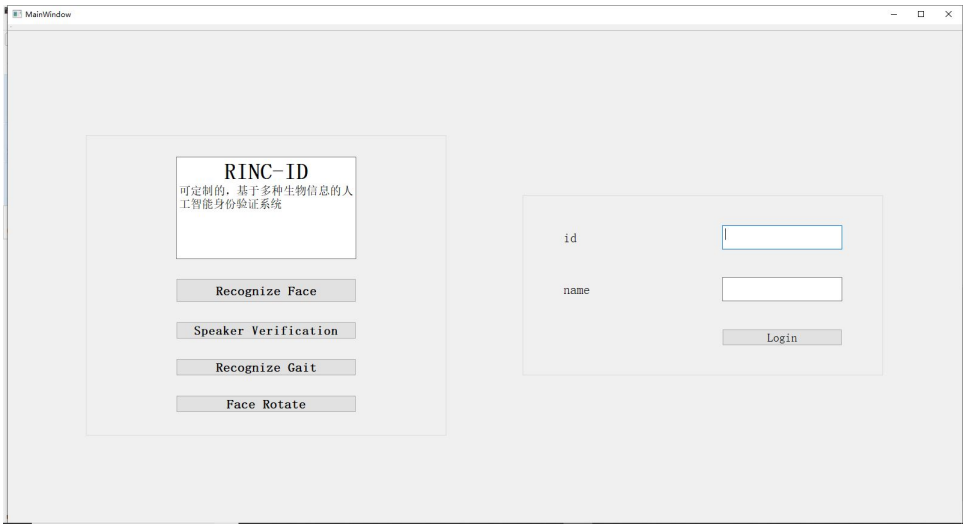

图 5-2: 前端界面 1

在图5-4所示的人脸识别注册界面包含四个按键：打开相机、取照片、人脸识别、确认并退出。为了保护用户的隐私安全，我们的身份验证系统在任何阶段

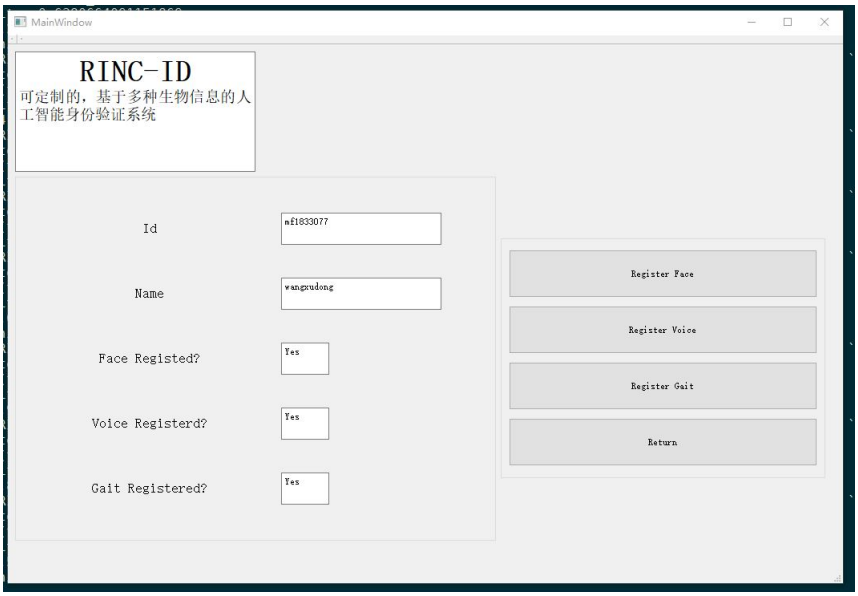

图 5-3: 前端界面 2：注册功能初始界面

都不会自动打开用户摄像头，只有用户手动点击“打开相机”按钮的情况下摄像头才会被打开。打开摄像头后点用户在调整好拍摄位置、角度后点击“取照片”进行人脸拍摄，拍摄到一张图片后相机会自动关闭。在如果用户认为拍摄的照片并不理想，可以再次点击“打开相机”重新拍摄。当用户认为拍摄到合适的相片用于注册时，点击图5-4界面中“人脸注册”按钮进行注册，当前端系统检测到用户实施这一步操作后，会首先对拍摄到的图片进行人脸检测，如果认为画面中不存在人脸，则系统要求用户重新拍摄。在某些情况下 (比如图5-4中拍摄到的画面)，照片中可能存在多张人脸，此时系统的处理方法是，选取图片中面积最大的人脸用于注册，并在前端界面中框出候选人脸，在识别阶段如果出现画面中有多张人脸的情况也会以同样方式进行处理。在检测到照片中存在人脸后，系统将图片连同前面填写的姓名、id 信息用 protobuff 协议发送到后端服务器上。后端服务器将提取到的人脸特征向量连同其他信息存入数据库中并向前端返回注册结果，后者将同时在界面中通知用户注册是否成功。

对于已注册用户，识别时需要点击图5-2所示界面左侧的“Recognize Face”按键，点击后用户将如图5-5所示，进入识别界面。识别界面相比于注册界面更为简洁，只有两个按键，这是考虑到身份验证的实际应用场景以及使用时的效率而做出的设计。在第一章中我们提到，非配合式识别要比配合式更为高效，我们的搭建的系统在开启识别后便能够进行非配合式工作。点击图5-5中“打开相机”按钮后，系统自动对拍摄到的画面提取人脸，并传输到后端服务器上进行

图 5-4: 前端界面 3：人脸注册界面

特征相似度匹配，如果相似度超过阈值则向前端发出识别成功的信息完成识别，整个过程中无需用户过多操作。此外，在一次成功识别后，经过某个特定的时间后，系统会自动的重新开始抓取照片进行识别。这样，在门禁之类的场景中，人群在排队前进的过程中无需对识别机器进行任何操作便可以不断被识别，实现了非配合式的工作方式。

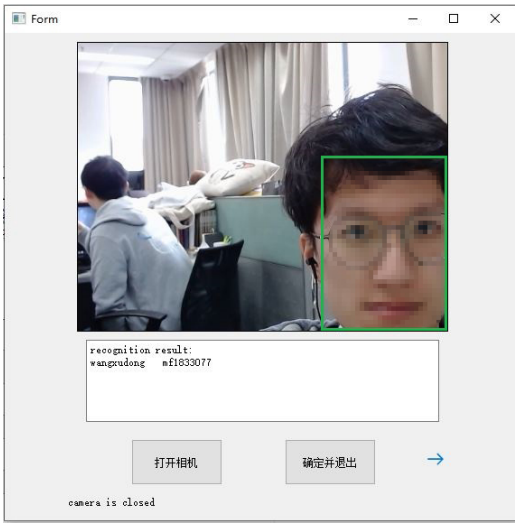

图 5-5: 前端界面 4：人脸识别界面

我们的身份验证系统使用灵活，能够应用于实际场景。根据前一章的实验效果也可以看出，集成了我们的 C-Face Network 重建模型后的身份验证系统能够有效识别低分辨率人脸，使得系统能够应用于更多的场景中。目前 RINC-ID 系统已经在学校、公司的多个场景进行试用，希望未来能够将其进一步推广。当然，目前系统尚处于雏形阶段，仍有许多地方可以改进。虽然已经同时支持 Windows

与 Linux 系统，但只有 PC 端应用，注册功能可以在更多的平台 (安卓/IOS 等) 以更多样的形式 (网页端、手机 App 等) 面向用户开放。同时，近年来数据存储以及信息检索方面的研究有许多新进展，RINC-ID 系统的服务端也应该在数据存储以及数据检索上做更多的探索与改进。

5.4 本章小结

本章主要介绍了将低分辨率人脸识别模型应用于实际的案例：RINC-ID 身份验证系统。将基于身份信息的人脸重建模型融入到人脸识别系统中，使得人脸识别系统应用场景更为广泛。充分验证了人脸超分辨率重建模型的实用性与有效性，证明了该领域的研究成果能够从理论走向实际应用。基于超分辨率重建的低分辨率人脸识别是低分辨率人脸识别领域最为直观的解决方案，同时也是最有实际应用价值的方案。

第六章　总结与展望

低分辨率人脸识别是目前的人脸识别研究走向实际应用的过程中一个不可忽视的问题，符合许多实际应用的需求，仅能有效识别高分辨率输入的人脸识别系统只能在十分有限的场景内发挥作用。本文的研究通过使用超分辨率重建方法试图解决低分辨率人脸识别问题，希望在经过重建后能够有效提高低分辨率人脸的识别准确率。在过去的工作中，人脸超分辨率重建方法大多只注重建图像的视觉效果，而没有关注是否在重建后有足够高的识别准确率。本文从提高重建结果的准确率出发提出了两个超分辨率重建模型，希望图片在重建前后保持一致的身份信息。

虽然超分辨率重建与人脸识别两个任务有一定相关性，但从任务目标的角度并没有较强关联，直接使用人脸识别中约束身份信息的损失函数对超分辨率重建模型进行训练可能难以收敛；因为同样的原因，迁移学习之类的方法也无法产生满意的效果。为了使模型在训练中学习到如何提取身份特征，本文提出的第一个模型 RefFace 使用了基于参考的超分辨率重建方法，在训练中选取与输入图片属于同一个人的照片作为参考图片，希望在训练过程中通过特征交换的方式重建出更好的结果。最终得到的结果虽然在视觉效果上具有突出的表现，但由于在训练过程中只是比较纹理特征的相似性，并没有直接涉及到身份信息的提取，RefFace 模型对低分辨率人脸识别准确率的提升十分有限。

受基于参考的超分辨率重建启发，并分析 RefFace 模型存在不足之处的原因，我们的第二个模型 C-Face Network 采用了在训练中对身份信息进行参考的方法。我们为模型设计了一个用于保持身份信息的损失函数，以及一套模型训练流程。通过我们提出的训练流程，模型参数能够充分收敛，在测试中表现出了优异的低分辨率人脸识别准确率。C-Face 模型的损失函数结合了人脸识别以及基于参考的重建思想，克服了传统人脸识别函数在重建模型中难以收敛的问题。此外，在基于参考的模型中，无论训练还是测试阶段，都要选取合适的参考图片才能有好的效果。但我们的模型虽然在训练中需要相对复杂的流程为训练数据配对，在测试中却并不需要选取任何的参考，只需要输入低分辨率人脸便可得

到重建结果。

在以上研究的基础上，我们搭建了一套身份验证系统，并将我们提出的 C-Face 模型集成到系统中，以此体现我们的研究具有实际应用价值。相比于一般的身份验证系统，我们系统中的人脸识别模块能够对低分辨率人脸做特殊的处理，从而对各种不同分辨率的人脸都能有效的识别，使得我们的系统能够应用于更加广泛的场景。

按照本文的研究思路，可以继续开展进一步的工作。首先，可以借鉴风格迁移的思想，考虑如何进一步的在重建过程中保持身份信息。其次，我们认为本文的许多思想也可以用于生成式模型中，用于高分辨率人脸的生成工作。最后，如何将人脸重建模型结合适当的软硬件环境部署到实际场景中，为实际生产生活做出贡献，也是十分值得研究的问题。

参考文献

[1] 刘晖龚知资. 基于人脸识别的人物信息在线检索平台[J]. 现代信息科技, 2020, v.4(13):82-84+89.

[2] 李国和, 乔英汉, 吴卫江, 等. 深度学习及其在计算机视觉领域中的应用[J]. 计算机应用研究, 2019, 036(012):3521-3529,3564.

[3] LIU W, WEN Y, YU Z, et al. Sphereface: Deep hypersphere embedding for face recognition[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2017.

[4] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments[J]. Technical Report 07-49, 2007.

[5] KEMELMACHER-SHLIZERMAN I, SEITZ S M, MILLER D, et al. The megaface benchmark: 1 million faces for recognition at scale[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016: 4873-4882.

[6] YI D, LEI Z, LIAO S, et al. Learning face representation from scratch[J]. arXiv preprint arXiv:1411.7923, 2014.

[7] LI P, PRIETO L, MERY D, et al. Face recognition in low quality images: a survey [J]. arXiv preprint arXiv:1805.11519, 2018.

[8] CHEN Y, TAI Y, LIU X, et al. Fsrnet: End-to-end learning face super-resolution with facial priors[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2018.

[9] HUANG H, HE R, SUN Z, et al. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.: s.n.], 2017: 1689-1697.

[10] HSU C C, LIN C W, SU W T, et al. Sigan: Siamese generative adversarial network for identity-preserving face hallucination[J]. IEEE Transactions on Image Processing, 2019, PP(99):1-1.

[11] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2017: 4681-4690.

[12] KIM J, KWON LEE J, MU LEE K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2016: 1646-1654.

[13] LIM BEE K H N S, Son Sanghyun, MU L K. Enhanced deep residual networks for single image super-resolution[J]. CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017, 2017:1132-1140.

[14] BULAT A, TZIMIROPOULOS G. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2018: 109-117.

[15] ZHANG K, ZHANG Z, CHENG C W, et al. Super-identity convolutional neural network for face hallucination[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.: s.n.], 2018: 183-198.

[16] HERRMANN C. Extending a local matching face recognition approach to low-resolution video[C]//2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. [S.l.]: IEEE, 2013: 460-465.

[17] KIM H I, LEE S H, RO Y M. Adaptive feature extraction for blurred face images in facial expression recognition[C]//2014 IEEE International Conference on Image Processing (ICIP). [S.l.]: IEEE, 2014: 5971-5975.

[18] XIAO Y, CAO Z, WANG L, et al. Local phase quantization plus: A principled method for embedding local phase quantization into fisher vector for blurred image recognition[J]. Information Sciences, 2017, 420:77-95.

[19] EL MESLOUHI O, ELGARRAI Z, KARDOUCHI M, et al. Unimodal multi-feature fusion and one-dimensional hidden markov models for low-resolution face recognition[J]. International Journal of Electrical and Computer Engineering, 2017, 7(4):1915.

[20] PENG Y, SPREEUWERS L, VELDHUIS R. Designing a low-resolution face recognition system for long-range surveillance[C]//2016 International Conference of the Biometrics Special Interest Group (BIOSIG). [S.l.]: IEEE, 2016: 1-5.

[21] WANG Z, MIAO Z, WU Q J, et al. Low-resolution face recognition: a review[J]. The Visual Computer, 2014, 30(4):359-386.

[22] BISWAS S, BOWYER K W, FLYNN P J. Multidimensional scaling for matching low-resolution face images[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 34(10):2019-2030.

[23] WANG Z, YANG W, BEN X. Low-resolution degradation face recognition over long distance based on cca[J]. Neural Computing and Applications, 2015, 26(7): 1645-1652.

[24] WEI X, LI Y, SHEN H, et al. Joint learning sparsifying linear transformation for low-resolution image synthesis and recognition[J]. Pattern Recognition, 2017, 66: 412-424.

[25] HEINSOHN D, VILLALOBOS E, PRIETO L, et al. Face recognition in low-quality images using adaptive sparse representations[J]. Image and Vision Computing, 2019, 85:46-58.

[26] JIANG J, HU R, HAN Z, et al. Coupled discriminant multi-manifold analysis with application to low-resolution face recognition[C]//International Conference on Multimedia Modeling. [S.l.]: Springer, 2015: 37-48.

[27] SHI J, QI C. From local geometry to global structure: Learning latent subspace for low-resolution face image recognition[J]. IEEE Signal Processing Letters, 2014, 22(5):554-558.

[28] HAGHIGHAT M, ABDEL-MOTTALEB M. Low resolution face recognition in surveillance systems using discriminant correlation analysis[C]//2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). [S.l.]: IEEE, 2017: 912-917.

[29] LI P, PRIETO L, MERY D, et al. On low-resolution face recognition in the wild: Comparisons and new techniques[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(8):2000-2012.

[30] LU Z, JIANG X, KOT A. Deep coupled resnet for low-resolution face recognition [J]. IEEE Signal Processing Letters, 2018, 25(4):526-530.

[31] ZENG D, CHEN H, ZHAO Q. Towards resolution invariant face recognition in uncontrolled scenarios[C]//2016 International Conference on Biometrics (ICB). [S.l.]: IEEE, 2016: 1-8.

[32] ZHANG H, YANG J, ZHANG Y, et al. Close the loop: Joint blind image restoration and recognition with sparse representation prior[C]//2011 International Conference on Computer Vision. [S.l.]: IEEE, 2011: 770-777.

[33] JIN M, HIRSCH M, FAVARO P. Learning face deblurring fast and wide[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. [S.l.: s.n.], 2018: 745-753.

[34] DODGE S, KARAM L. Understanding how image quality affects deep neural networks[C]//2016 eighth international conference on quality of multimedia experience (QoMEX). [S.l.]: IEEE, 2016: 1-6.

[35] SHEN Z, LAI W S, XU T, et al. Deep semantic face deblurring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2018: 8260-8269.

[36] KANADE T. Picture processing system by computer complex and recognition of human faces[J]. 1974.

[37] KELLY M D. Visual identification of people by computer: number 130[M]. [S.l.]: Department of Computer Science, Stanford University., 1970.

[38] SHI J, SAMAL A, MARX D. How effective are landmarks and their geometry for face recognition?[J]. Computer vision and image understanding, 2006, 102 (2):117-133.

[39] DANIYAL F, NAIR P, CAVALLARO A. Compact signatures for 3d face recognition under varying expressions[C]//2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. [S.l.]: IEEE, 2009: 302-307.

[40] GUPTA S, MARKEY M K, BOVIK A C. Anthropometric 3d face recognition [J]. International journal of computer vision, 2010, 90(3):331-349.

[41] SIROVICH L, KIRBY M. Low-dimensional procedure for the characterization of human faces[J]. Josa a, 1987, 4(3):519-524.

[42] KIRBY M, SIROVICH L. Application of the karhunen-loeve procedure for the characterization of human faces[J]. IEEE Transactions on Pattern analysis and Machine intelligence, 1990, 12(1):103-108.

[43] ETEMAD K, CHELLAPPA R. Discriminant analysis for recognition of human face images[J]. Josa a, 1997, 14(8):1724-1733.

[44] HE X, YAN S, HU Y, et al. Face recognition using laplacianfaces[J]. IEEE transactions on pattern analysis and machine intelligence, 2005, 27(3):328-340.

[45] DELALLEAU O, BENGIO Y. Shallow vs. deep sum-product networks[J]. Advances in neural information processing systems, 2011, 24:666-674.

[46] HUBEL D H, WIESEL T N. Early exploration of the visual cortex[J]. Neuron, 1998, 20(3):401-412.

[47] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]//European conference on computer vision. [S.l.]: Springer, 2014: 818-833.

[48] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[49] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. 2016.

[50] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10,000 classes[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2014: 1891-1898.

[51] WANG F, XIANG X, CHENG J, et al. Normface: L2 hypersphere embedding for face verification[C]//Proceedings of the 25th ACM international conference on Multimedia. [S.l.: s.n.], 2017: 1041-1049.

[52] WANG H, WANG Y, ZHOU Z, et al. Cosface: Large margin cosine loss for deep face recognition[J]. 2018.

[53] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019: 4690-4699.

[54] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//European Conference on Computer Vision. [S.l.: s.n.], 2014.

[55] WU J, DING S, XU W, et al. Deep joint face hallucination and recognition[J]. arXiv preprint arXiv:1611.08091, 2016.

[56] CABALLERO J, LEDIG C, AITKEN A, et al. Real-time video super-resolution with spatio-temporal networks and motion compensation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2017: 4778-4787.

[57] YUE H, SUN X, YANG J, et al. Landmark image super-resolution by retrieving web images[J]. IEEE Transactions on Image Processing, 2013, 22(12):4865-4878.

[58] ZHANG Z, WANG Z, LIN Z, et al. Image super-resolution by neural texture transfer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019: 7982-7991.

[59] ZHENG H, JI M, WANG H, et al. Crossnet: An end-to-end reference-based super resolution network using cross-scale warping[C]//Proceedings of the European conference on computer vision (ECCV). [S.l.: s.n.], 2018: 88-104.

[60] GATYS L A, ECKER A S, BETHGE M. Texture synthesis using convolutional neural networks[J]. arXiv preprint arXiv:1505.07376, 2015.

[61] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016: 2414-2423.

[62] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution[C]//European Conference on Computer Vision. [S.l.]: Springer, 2016: 694-711.

[63] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems. [S.l.: s.n.], 2014: 2672-2680.

[64] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015: 815-823.

[65] HUYNH-THU Q, GHANBARI M. Scope of validity of psnr in image/video quality assessment[J]. Electronics letters, 2008, 44(13):800-801.

[66] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4):600-612.

[67] ZHEN LI Z L X Y G J, Jinglei Yang, WU W. Feedback network for image super-resolution[C]//IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. [S.l.: s.n.], 2019: 3867-3876.

[68] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016: 1874-1883.

[69] FENG, WANG, JIAN, et al. Additive margin softmax for face verification[J]. IEEE Signal Processing Letters, 2018.

[70] LIAO S, LEI Z, YI D, et al. A benchmark study of large-scale unconstrained face recognition[C]//IEEE international joint conference on biometrics. [S.l.]: IEEE, 2014: 1-8.

[71] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. [S.l.]: Ieee, 2009: 248-255.

[72] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[73] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019: 4401-4410.

[74] ABADI M, AGARWAL A, BARHAM P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems[J]. arXiv preprint arXiv:1603.04467, 2016.

[75] WYDROWSKI B, ANDREW L L, ZUKERMAN M. Maxnet: A congestion control architecture for scalable networks[J]. IEEE Communications Letters, 2003, 7(10):511-513.

[76] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. arXiv preprint arXiv:1912.01703, 2019.

[77] ZHANG S, LIN Y, SHENG H. Residual networks for light field image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019: 11046-11055.

[78] ZHANG K, ZUO W, ZHANG L. Deep plug-and-play super-resolution for arbitrary blur kernels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019: 1671-1681.

[79] GU J, LU H, ZUO W, et al. Blind super-resolution with iterative kernel correction [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019: 1604-1613.

[80] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[J]. 2018.

简历与科研成果

基本信息

王绪冬，男，汉族，1996 年 1 月出生，辽宁省大连人。

教育背景

2018 年 9 月 — 2021 年 6 月 南京大学计算机科学与技术系 硕士

2014 年 9 月 — 2018 年 6 月 江南大学信息与计算科学系 本科

攻读硕士学位期间完成的学术成果

1. Xudong Wang, Furao Shen, Jian Zhao, "C-Face Network: Face Hallucination for Maintaining Identity Information," in *Proc. Journal of Artificial Intelligence Research*, 2021, Mar. 2021.

攻读硕士学位期间的发明专利

1. 申富饶,**王绪冬**,李俊,赵健."一种基于身份信息的人脸图像重建方法"(ZL201911024313.X)

攻读硕士学位期间参与的科研课题

1. 国家自然科学基金面上项目 ''基于深度感知增量式联想记忆神经网络的信息融合系统研究''（课题年限 2019.01 ~ 2022.12），负责神经网络模型相关研究。

致谢

毕业论文工作接近尾声，也意味着研究生三年的生涯即将告一段落。在南京大学三年的读研时光里，我在南京大学计算机系 RINC 实验室学习到了很多宝贵的知识与经验。

首先最应该感谢的便是我的导师申富饶教授以及吴楠副教授。老师们治学严谨，深受我们大家的爱戴。我仍然记得当初我刚刚跨专业考研来到南大时，对许多专业知识、专业工具都不能熟练运用，申老师总是鼓励我不能急于求成，才使我的研究工作逐渐步入正轨。在科研中申老师总是教导我们不要迷信学术权威，多注入自己的思考；在生活中，他锻炼身体、劳逸结合，关心大家的身心健康。在学术与生活上都是我们的良师益友。

同时还要感谢赵健老师。赵老师讲授矩阵论课程，同时定期为我们做组会报告，为我们介绍科研工作或者论文书写的经验。在我们因为没有经验而只能用蹩脚的英语写出粗糙的论文时，赵老师对我们的每篇文章都进行了多次修改，为我们的科研工作提供了极大的帮助。

还要感谢 RINC 实验室的同学们，他们在学习和生活中都为我提供了许多的帮助，在 426 教室三年的朝夕相伴必然是人生一段宝贵的回忆。

最后特别感谢我的父母，他们十多年的含辛茹苦的付出才能使我有今天的小小成绩。如今即将走向社会，终于能够用自己的劳动回报他们昔日的付出。

在南京大学的三年里遇到了许多小伙伴，他们中有的人专注于科学技术事业，有的人立志于从政而为社会奉献，也有的人身怀才艺而成为舞台上的焦点。每个人兴趣、特长不同，但都心怀自己的理想。正是认识了这些有趣的小伙伴，我才能有这三年丰富多彩的生活。

学位论文出版授权书

本人完全同意《中国优秀博硕士学位论文全文数据库出版章程》（以下简称“章程”），愿意将本人的学位论文提交“中国学术期刊（光盘版）电子杂志社”在《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》中全文发表。《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》可以以电子、网络及其他数字媒体形式公开出版，并同意编入《中国知识资源总库》，在《中国博硕士学位论文评价数据库》中使用和在互联网上传播，同意按“章程”规定享受相关权益。

作者签名：_____

_____年_____月_____日

论文题名	基于超分辨率重建的低分辨率人脸识别				
研究生学号	MF1833077	所在院系	计算机科学与技术系	学位年度	2018
论文级别	☑ 硕士　☐ 硕士专业学位 ☐ 博士　☐ 博士专业学位　（请在方框内画勾）				
作者 Email	wangxd@smail.nju.edu.cn				
导师姓名	申富饶 教授　吴楠 副教授				

论文涉密情况：

☐ 不保密

☑ 保密，保密期：_____年_____月_____日　至　_____年_____月_____日

注：请将该授权书填写后装订在学位论文最后一页（南大封面）。