

硕士学位论文

**临床皮肤疾病图像的辅助诊断分类算法及
应用研究**

**Research on Application and Classification Algorithm
of Computer Aided Diagnosis for Skin Disease Based
on Clinical Image**

学 科 专 业： 控制工程

专 业 领 域： 模式识别与智能系统

作 者 姓 名： 赵新宇

指 导 教 师： 谢斌

中 南 大 学
2021 年 5 月

中图分类号 TP391.4

学校代码 10533

UDC 004.8

学位类别 专业学位

硕士学位论文

临床皮肤疾病图像的辅助诊断分类算法及应用研究 Research on Application and Classification Algorithm of Computer Aided Diagnosis for Skin Disease Based on Clinical Image

作者姓名：赵新宇

学科专业：控制工程

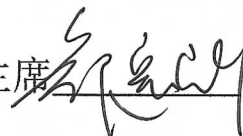
专业领域：模式识别与智能系统

研究方向：医疗辅助诊断系统

二级培养单位：自动化学院

指导教师：谢斌

论文答辩日期 2021年5月27日 答辩委员会主席



中南大学
2021年5月

学位论文原创性声明

本人郑重声明，所呈交的学位论文是本人在指导教师指导下进行的研究工作及取得的研究成果。尽我所知，除了论文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中南大学或其他教育机构的学位或证书而使用过的材料。与我共同工作的同志对本研究所作的贡献均已在论文中作了明确的说明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

作者签名： 赵新宇 日期： 2021 年 6 月 3 日

学位论文授权使用授权书

本学位论文作者和指导教师完全了解中南大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交学位论文的复印件和电子版；本人允许本学位论文被查阅和借阅；学校可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用复印、缩印或其它手段保存和汇编本学位论文。

保密论文待解密后适应本声明。

作者签名： 赵新宇

日期： 2021 年 6 月 3 日

导师签名： 谭毅

日期： 2021 年 6 月 3 日

临床皮肤疾病图像的辅助诊断分类算法及应用研究

摘要: 皮肤疾病的患病率极高,且种类繁多,对于某些高致死率的疾病,如果可以在早期被诊断出来,将具有很大概率被治愈。然而我国的注册皮肤科医生总量少,且存在较为严重的医疗资源不平衡问题,皮肤疾病的辅助诊断研究在具备非侵入性、无主观性等优势下,已经逐渐成为研究热点。本文针对这一课题展开了深入的研究,具体的工作内容如下所示:

1.针对于现有的公开数据集多为欧美人种的皮肤镜图像,而皮肤镜图像涵盖的疾病种类少,获取条件苛刻且不利于实际应用的问题,本文建立了以国内皮肤病人群为主的大规模临床图像数据集,包含了 **541 类** 有病理检验上的结果作为依的皮肤疾病,总数量达 **47075 张**。为了评估数据集的性能,本文在四种常用卷积神经网络上进行了基准实验,并通过与国外数据集对比,突出了不同人种皮肤病特征的差异性。

2.针对于粗粒度皮肤疾病类别的类内差异较大的问题,提出一种基于精细粒度的两阶段分类方法。本文以银屑病这样一种数据量充足且研究价值较大的疾病为例,建立了包含四种银屑病近亲以及四种银屑病远亲的数据集,研究银屑病的两分类任务。与直接将其他 8 类疾病归为“非银屑病”的单阶段方法相对,本文通过两阶段的方式,先进行细粒度的 9 分类预测,并通过输出决策,将结果改出为两元输出。**实验结果显示,以 InceptionV3 网络为基础的两阶段 AUC (Area Under the Curve) 可达 0.981, 而单阶段的 AUC 只有 0.966。**

3.针对于多病种分类任务中,不同皮肤疾病的类间差异不一致,训练难度大的问题,本文提出了一种**基于粒度迁移的分层诊断算法**。在构建了包含 19 类子类皮肤病以及 6 类大类皮肤病的两层分类结构基础下,本文先采用了细粒度特征迁移的方式将子类疾病训练的模型参数部分迁移至大类疾病的训练网络中,之后在预测过程中,结合粗粒度的大类疾病预测结果,完成对子类细粒度的分层诊断。通过对比实验,这样的分层诊断方式在 12612 张数据组成的严格按照病例拆分的 5 折分层交叉验证中达到了 **F1: 0.591, ACC (Accuracy): 0.600**, 高于未分层的结果。

4. 针对于现有的分类算法敏感性及特异性难以达到临床应用标准的问题,本文以皮肤肿瘤恶性程度分级作为切入点,提出了面向病人的健康管理应用设计。考虑到皮肤病学上目前没有具体的恶性程度量化指标,本

文使用了代表三种恶性程度的 10 类皮肤肿瘤，利用分类任务的损失函数完成了模型的训练，并通过建立输出决策的方式，给出了每张图像的参考分值及具体的恶性程度级别。通过与 20 位专业医生进行对比，本应用可以更准确更客观地对皮肤肿瘤的恶性程度进行分级。

图 29 幅，表 7 个，参考文献 78 篇

关键词：辅助诊断；临床皮肤疾病图像；迁移学习；恶性程度分级；卷积神经网络

分类号：TP391.4

Research on Application and Classification Algorithm of Computer Aided Diagnosis for Skin Disease Based on Clinical Image

Abstract: Skin diseases are various with a high prevalence. For some diseases with high mortality, if they can be diagnosed in the early stage, they will be easily cured. However, the number of registered dermatologists in China is small, and there is a serious imbalance about medical resources. The research on the computer aided diagnosis of skin diseases has become a hot topic because of the advantage of non-invasive and objectivity. This paper has carried out a deep study on this subject, and the specific work contents are as follows:

The existing public datasets are mostly composed of European and American dermoscopic image, and cover less diseases, the acquisition conditions are harsh and not conducive to practical application. This paper builds a large-scale clinical image dataset based on the domestic skin patients, including 541 types of skin diseases based on pathological examination, with a total number of 47075. In order to evaluate the performance of datasets, this paper conducts benchmark experiments on four kinds of convolutional neural networks, and highlights the differences of dermatological characteristics of different races by comparing this dataset with foreign datasets.

A two-stage classification method with more fine-grained features is proposed to solve the problem of large intra-class difference in the classification of coarse-grained skin diseases. In this paper, psoriasis, a disease with sufficient data and high research value, is taken as an example, and a dataset of four close relatives and four distant relatives of psoriasis is built to study the binary classification tasks of psoriasis. Compared with the single-stage method that other 8 diseases are defined as "non psoriasis", this paper conducts fine-grained 9 classification prediction with two-stage method firstly, and then changes the result to binary output through output strategy. The experimental results show that the AUC (Area Under the Curve) of two stages based on the concept V3 network can reach 0.981, while the AUC in single stage is only 0.966.

In view of the problems of classification for multi-category skin diseases with inconsistent inter-class difference, which are difficult to train, a hierarchical diagnosis algorithm based on granularity transfer is proposed. Based on the construction of two-layer taxonomy including 19 sub-class dermatoses and 6 meta-class dermatoses, this paper first to transfer the model parameters of sub-class disease to the training network of meta-class diseases. Then, in the process of prediction, combined with the prediction results of meta-class diseases, the fine-grained hierarchical diagnosis of sub-class diseases is completed. Through the extensive experiment, the results of this method were F1: 0.591 and ACC (Accuracy): 0.600 with using stratified 5 folds of 12612 images, which was strictly in accordance with the case splitting, which was higher than that of the non-hierarchical model.

In view of the problems that the sensitivity and specificity of the existing classification algorithms are difficult to meet the clinical application standards, this paper presents the application of health management for patients based on the grading of skin tumor malignancy. Considering that there is no specific quantitative index of the degree of malignancy in dermatology, this paper uses 10 kinds of skin tumors representing three kinds of malignant degree. The model is trained by using the loss function of classification task. The reference score and the specific level of malignant degree of each image are given by establishing the output decision-making method. Compared with 20 professional dermatologists, the application can be more accurate and objective to classify the malignant degree of skin tumors.

Keywords: Computer aided diagnosis; Clinical skin disease image; Transfer learning; Grade of malignancy; Convolution neural networks

Classification: TP391.4

目 录

第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 皮肤疾病数据集	2
1.2.2 皮肤疾病图像分类算法	6
1.3 研究目的与主要研究内容	9
1.3.1 研究目的	9
1.3.2 研究内容及本文结构安排	9
第 2 章 临床皮肤疾病数据集——XiangyaDerm	12
2.1 数据集的建立	12
2.1.1 数据的获取与清洗	12
2.1.2 数据集的标注与后处理	13
2.2 数据集的评估	15
2.2.1 评价指标	15
2.2.2 数据集的基准表现	15
2.2.3 国内外数据集差异性研究	16
2.3 本章小结	18
第 3 章 基于精细粒度的两阶段分类方法	19
3.1 分类网络的细粒度	19
3.1.1 理论基础	19
3.1.2 两阶段分类方法的网络结构	20
3.2 两阶段分类实验方案设计	20
3.2.1 实验数据集	20
3.2.2 算法评估准则	21
3.2.3 两阶段皮肤疾病分类的实验评估	22
3.2.4 泛化性能验证	24
3.3 本章小结	26
第 4 章 基于粒度迁移的分层诊断算法	28
4.1 数据集与数据增强	28

4.2 参数选择	30
4.2.1 超参数设置	30
4.2.2 基于进化算法的超参数选择	31
4.3 皮肤疾病的细粒度分类特征迁移	34
4.3.1 特征迁移原理	34
4.3.2 迁移参数的选择与对比	35
4.4 结合粗粒度特征的皮肤疾病分层诊断	38
4.4.1 分层诊断算法	38
4.4.1 算法评估	40
4.5 本章小结	41
第 5 章 皮肤肿瘤恶性程度分级的应用研究	42
5.1 面向病人的应用设计	42
5.1.1 应用的功能	42
5.1.2 应用的设计与实现	43
5.2 应用性评价	46
5.3 本章小结	47
第 6 章 总结与展望	48
6.1 论文总结	48
6.2 研究展望	49
参考文献	50
攻读学位期间的主要研究成果	57
致 谢	59

第 1 章 绪论

1.1 研究背景及意义

皮肤疾病普遍存在于全球不同地区、不同人种以及不同年龄段的人群中，占到 30-70%人口^[1]。在不良环境因素（如辐射及生活方式）的诱因下，通常表现出包含癌症、炎症以及传染性疾病在内的病症，皮肤疾病种类多达 2000 多种，其中以具有高致死率的恶性黑素瘤危害最大^[2]。早期诊断是皮肤疾病患者康复的决定性因素，尽管某些疾病的致死率很高，但如果能在早期及时诊断，这类疾病同样也具有很高的治愈率。然而在我国，医疗资源分布存在着明显的不均，注册的皮肤科专业医生仅为 2 万余人，而现有的患者数量却远大于此。一些相对偏远的基层医疗机构，更是没有医疗机构，这给我国皮肤疾病的早期诊断和筛查带来了极大的困难。

随着近几年人工智能技术的快速发展，通过在高性能的计算机部署机器学习算法已经渐渐成为解决供需不平衡、辅助医生进行早期疾病筛查的重要研究方向。该技术相对于传统的医学手段存在以下优势：一、具有非侵入的特性。只需要借助疾病的病患处图像以及一些疾病信息，就可以通过算法推断出可能患有疾病及症状，避免了直接接触患者。二、改善了诊断过程中的主观性。由于部分医师医学经验的差异，以及外界因素干扰（如长期疲劳）导致的个人主观产生偏差，可能会导致一些误诊情况。三、降低医生的工作量，优化工作效率。计算机辅助诊断系统可以在一定程度提供诊断的参考结果，帮助医生进行病情初筛，患者也可以通过初筛的结果有参照地就医。

研发辅助诊断系统关键因素之一是医疗数据，就皮肤病的图像数据而言，目前还存在缺少大规模数据集的问题。在皮肤疾病的诊断流程中，不同病种的视觉差异不明显，导致诊断过程中需要借助不同的医疗工具，比如皮肤镜、电子显微镜等。皮肤镜图像可以更加清晰地辨认出皮肤肿瘤相关的疾病，可以大大提高诊断过程中准确率，然而医疗资源不足的部分地区却难以获取。利用显微镜或电子显微镜对细胞组织进行切片观察，是目前诊断皮肤病的“金标准”，但这对医疗条件的要求又进一步提升。在临床上，利用智能手机或者专业摄像机拍摄的病患部位的临床图像，对很多疾病是可以有效诊断的，这种皮肤病图像的获取要相较于其他类型图像要便利的多。但对于研发有效的辅助诊断系统，无论从病种还是数量而言，目前的数据集都是不够的。更值得关注的问题是，由于皮肤病患者群体人种的差异，在全球范围内的皮肤病临床图像特征是不完全统一的。据有效统计，目前可公开获取的大规模数据集以欧美人为主群体，对于可靠的亚洲群体甚至中国人群的数据集的需求，是目前国内辅助诊断系统开发急需解决的重要问题。

此外，就病种而言，目前皮肤疾病领域的计算机辅助诊断系统研究大多是基于部分病种（如相关研究最多的皮肤肿瘤类）。原因一方面体现在，这类疾病有着较高的

致死率,且有相关的数据集,但另一方面,则是由于缺乏行之有效的多病种诊断体系,使得很多常用的方法在多病种这一复杂相似的类间特征差异下,变得表现不佳,难以胜任。然而对于辅助诊断系统而言,能够覆盖更多的病种是一重要的研究方向,也是现在需要克服的难点。此外对于系统的应用而言,高的敏感性以及特征性是诊断必须要达到的要求,迫于数据、技术以及现阶段设备的压力,要短期实现这种医疗应用的落地是十分困难的。实际上,在日常生活中,病人的健康管理及预防是更加重要的,但是有效的面向病人的皮肤病健康管理系统的出现。

基于以上问题,本文介绍了以中国人为主的皮肤疾病临床图像数据集的建立过程以及在常见算法上的表现,并通过研究分类粒度对皮肤病图像分类网络的影响,构建了基于粒度迁移的分层诊断算法。该算法通过模拟医生临床诊断的思路,形成了一套针对于多病种的辅助诊断体系,相较于现有的分类方法,具有更好的准确率以及一定程度的可解释性。考虑到实际应用方面,本文还给出了面向患者的健康管理思路,通过对病人所患皮肤肿瘤的分级,给出合理的建议,帮助病人有参考地就诊。

1.2 国内外研究现状

大规模、高质量的皮肤疾病图像在辅助诊断算法中一直充当着驱动力的角色。本文在介绍现阶段用于皮肤病分类的常用算法之前,会先对现有的不同种类的公开数据集进行统计,并详细介绍针对这些数据集的相关研究。

1.2.1 皮肤疾病数据集

现有的公开可获取的皮肤病数据集主要包含三种:皮肤镜图像、组织病理学图像以及临床图像。在实际的皮肤诊断中,往往还涉及到其他类型:比如反射式共聚焦显微镜(Reflectance Confocal Microscopy, RCM)图像^[3]、光学相干断层扫描(Optical coherence tomography, OCT)图像^[5]等,但由于没有公开的数据集,涉及到的研究很少,在此不作进一步介绍。

一、皮肤镜图像

在医学诊断领域中,大部分的皮肤疾病可以由多个皮肤科专家直接观察诊断鉴别(如白癜风、银屑病等皮损),但这需要多年的问诊经验以及相应的医学知识素养。而有些皮肤病(如细胞癌)的皮损特点很难被直接观察到,这就需要借助物理手段放大患病皮肤部位。皮肤镜便是利用这样的特性来优化皮肤疾病诊断流程的技术,它是目前皮肤病学中最广泛使用的图像采集方法之一,是一种非侵入性的 PSL 成像技术。皮肤镜技术允许通过手持入射光放大设备(由高分辨率数字单镜头反射和光学显微镜组成)和浸液(具有使皮肤角质层对光更透明并消除反射的折射率)对其表面下结构进行可视化,使皮下结构比传统的临床图像更容易看到^[6]。

皮肤镜图像的获取过程是先用酒精或者水等液体覆盖皮损部位,接着用皮肤镜照

相机来获取得到皮肤镜图像。通过这种方式得到的图像一般光线均匀，分化较多，不仅可以扩大了皮肤的病变面积，也在一定程度上消除了一些光的干扰。这种技术能够通过增强皮损特征来辅助诊断分析，帮助专业医生对皮损形态结构的识别和检测。有文章已经证明皮肤镜检查会降低没有经验的皮肤科医生的诊断准确性^[9]。

由于皮肤镜图像在图像质量上的优势，这种类型的公开数据集出现在绝大多数的研究中。最常使用的是由国际皮肤成像协作组织（International Skin Imaging Collaboration, ISIC）提供的公开的数据集。该组织是由国际皮肤数字成像学会赞助的，旨在提高黑色素瘤诊断水平的国际组织。ISIC 档案库则是为了帮助临床知识培训，以及举办图像挑战赛而专门设立的大规模皮肤镜图像国际存储库，包括了来自国际临床中心收集的上万张图像。本文对目前的公开数据集进行了图像数量及疾病类别上的统计，如图 1-1 所示，图中缩写的全称为痣（Nev）、黑色素瘤（Mel）、脂溢性角化病（SK）、基底细胞癌（BCC）、光化性角化病（AK）、皮肤纤维瘤（DF）、血管瘤（VASC）、良性角化病（BK）、鳞状细胞癌（SCC）、上皮内癌（IEC）、化脓性肉芽肿（PYO）。值得注意的是，ISIC2020 是根据两分类任务比赛设置的标签，但实际有 18%的数据给出了 9 类标签。

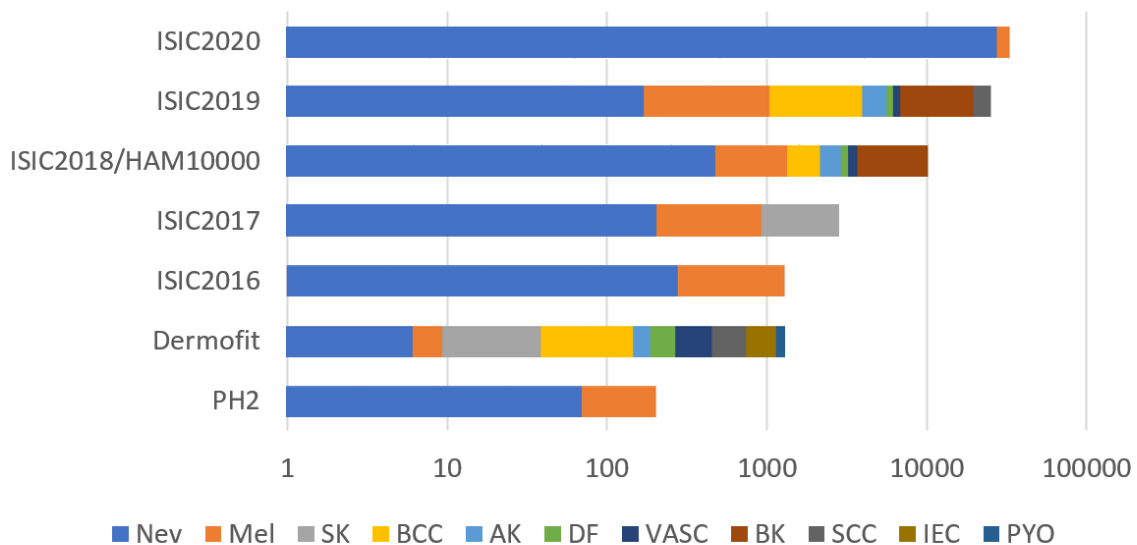


图 1-1 皮肤镜图像的公开数据集统计

观察上述的几个皮肤镜图像数据集，可以发现以下几点：

- 1) 由于皮肤镜图像的获取条件较为苛刻，公开的数据集的数据总量不是很大。在机器学习算法的大背景下，如果直接应用于皮肤疾病的分割或分类，模型参数的训练学习容易过拟合，诊断效果会受很大影响。
- 2) 皮肤病种类偏少。基于公开数据集的现有研究任务，研究对象大多是以皮肤肿瘤类皮肤疾病为主，在实际临床中细分种类极多的情况下，需要把握到具

体病症才可以提供特异的诊治，研究的模型效果泛化性可能不足够强

- 3) 公开数据集中的类别间数量不平衡。使用不平衡数据，模型会容易产生偏差，倾向于把对象分到大样本的类别，同时少样本数的类别包含的特征信息有很大的局限性，造成该类别的识别率很低。

二、组织病理学图像

皮肤活检组织病理学图像被广泛应用于皮肤科，可以通过病理学的定量分析来理解皮肤疾病发病的潜在原因，对皮肤科医生至关重要。因此这种方式诊断出来的结果被普遍认为皮肤病诊断领域的金标准^[10]。组织病理学的制备活检过程需要借助载玻片提供疾病组织更全面的视图，保留了潜在的组织结构，能够观察到细胞层面的细微特征。所以在图像的自动分析任务中，适当利用这一微观空间信息往往可以从病理角度获取皮肤病更具体、更细化的表征。但由于组织病理学图像的采集以及处理过程都有着最严格的要求，且图像本身带有超大数量级的像素表达（据研究：采集包含 12 到 20 个样本的患者对象，将产生 25-40 亿像素），使得这些图像的分析变得更加困难，对算法处理高分辨率的性能要求更高^[11]。

考虑到组织病理学图像应用的困难性以及获取途径的复杂性，在针对组织病理学图像的皮肤病辅助诊断研究中，公开的大型数据集极其稀有。2020 年，国际癌症机构公布了包含 2860 张图像的数据集：癌症基因组图集，如图 1-2 所示。

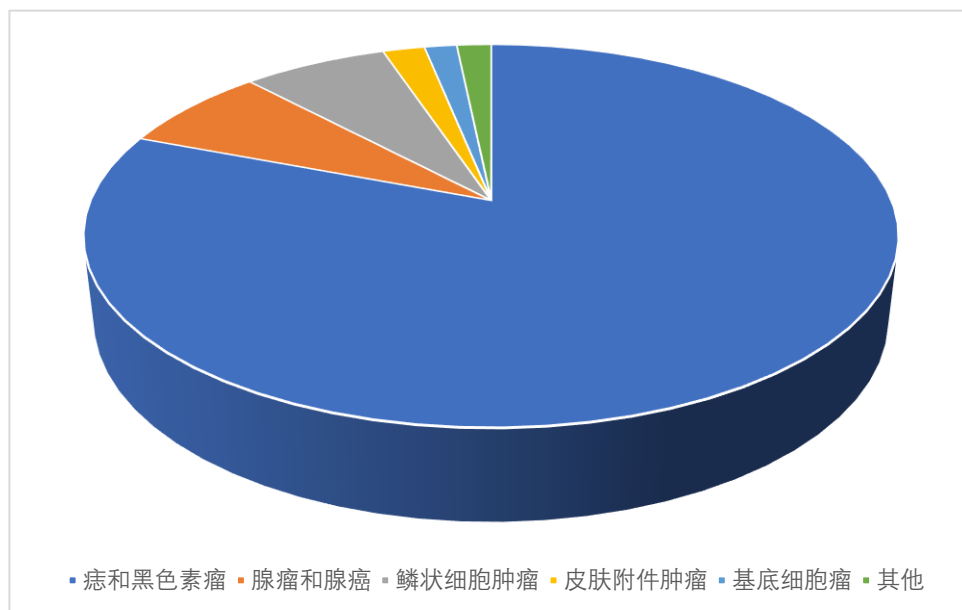


图 1-2 公开组织病理学图像数据集的病种分布

该数据集的属性特点可总结为了以下几点：

- 1) 数据类别间的不平衡问题仍然存在，由疾病的发病率以及就诊率实际情况不同引起的。
- 2) 相比其他类型图像，组织病理学图像的文件很大，具有很高的分辨率并附带

标注信息。

三、 临床图像

与皮肤镜图像以及组织病理学图像相比，临床图像的获取方式要便利得多，可以通过专业摄像机或者智能手机对准病患部位，即可得到供临床诊断用的图像。皮肤镜技术考虑的是增强病灶面积，但是病灶区域的大小、分布位置等视觉特征会被忽略。所以临床图像可以弥补皮肤镜的不足，其获取的便利性将为未来辅助诊断的应用提供绝对的驱动力，如可用于远程检查、患者的医疗记录等。

临床图像也叫宏观图像，这些图像再现了临床医生用肉眼看到的在活生物体中的位置^[12]。但由于拍摄角度，光线强度的影响，临床图像几乎都是在不同的光照条件，不均匀的病灶焦距下获取的。这会导致原始临床图像受到的外界干扰会比较大，也是目前的研究中需要解决的问题

皮肤镜图像目前有许多大规模公开的数据集，其诊断标签质量都能得到国际研究团队的认可。相比较而言，临床图像数据集的质量欠缺，导致相关的研究相对较少，表 1-1 列出几个在研究中相对使用频繁的数据集：

表 1-1 公开临床皮肤病图像数据集的分布统计

数据集	病种数	图像总数
MED-NODE	2	170
Danderm	91	3000
SD-198	198	6584
DermIS	735	7172
AtlasDerm	534	10129
Asan dataset	12	17125
Dermnet	626	18974
Derm101	525	22979
MoleMap	25	102451

现有的临床数据集可以总结为以下几点：

- 1) 数据量比起皮肤镜图像数据集要大不少，图像基本都存在着水印，其中 MoleMap 数据集数量最多，由 102451 张图像构成，但均为皮肤肿瘤类图像。
- 2) 存在涵盖较多病种的数据集，DermIS 数据集有最多的 735 类皮肤疾病，但是每类疾病的图像最多不超多 200 张，还存在很多个位数的情况。

综上所述可以看出，由于皮肤镜图像质量较高，现有的公开数据集以皮肤镜图像为主，其他图像类型的数据集相对较少。但是临床图像获取门槛低，具有更高的应用价值，且临床图像数据集中涉及的疾病种类丰富。然而目前还没有以国内人群为主的公开数据集。

1.2.2 皮肤疾病图像分类算法

一、传统的机器学习分类算法

传统的皮肤病图像分类研究主要是从较低水平的视觉特征出发，如颜色，纹理，形状等，来构建模型所需的描述子（如 SIFT^[13]、HOG^[14]、CH^[15]等），以监督学习的方式（如支持向量机^[16]、K-近邻算法^[17]、决策树^[18]等）来帮助自动快速诊断。这种方法得到的模型可解释性较高，可以较为清晰地理解模型判断的依据，但由于参数量比深度学习模型少，且手动设置的特征表征能力有限，在大规模数据下往往比不上深度学习模型，所以近年来相关研究数量变得稀少。

传统机器学习分类算法多应用在皮肤镜图像上，因为基于皮肤镜图像的描述子会受到更少环境噪声干扰，相关的模型也就表现得更出色。Afza 等人^[19]利用对比度线性拉伸处理来改善病灶边缘低对比度的情况，并提出熵值来最优化颜色特征选择，他们的未来工作是采用深度学习手段来自动化特征选择过程。Mporas 等人^[20]计算颜色的统计值作为特征，并验证了几种分类算法的性能，效果最佳的是带有随机森林分类器的 AdaBoost 算法。

单一的特征有时候不能提供皮损的全部信息，包含多个特征评价指标的 ABCD 准则^[21]在皮肤科专家诊断中被频繁使用，是一套比较可靠的评定规则。Monisha 等人^[22]开发一个自动的模型，依照 ABCD 规则、LBP 与 GLCM 来实现病灶分割，并输入到 BPN 中分类。Chatterjee 等人^[23]基于改善传统的 ABCD 的评价指标，重构了高鲁棒性的专家系统良恶性识别。尽管临床图像和皮肤镜图像有不同的含义组合，但它们包含相同的皮肤损伤方面特征，如结构、颜色和形状等，同样可以用来进行模型研究。Yang 等人^[24]则基于临床图像，研究遵循皮肤科医生的病变标准来设计几款符合 ABCD 准则的视觉特征，并且在小型数据集上得出了不弱于深度学习模型的临床图像分类效果。考虑到视觉特征的偏重各异，Dhivyaa 等人^[25]基于决策树的随机森林分类算法，以轻量的计算代价实现了较高的性能。

二、基于卷积神经网络的分类算法

随着 GPU 的快速发展，具有深层特征学习能力的卷积神经网络（Convolutional Neural Network, CNN）渐渐成为了研究皮肤病图像分类算法的首选。CNN 模型具备端对端的学习能力，即模型能够直接学习得到非具象的特征，以此有时能得到远超以人工特征为主的视觉特征的效果。作为具有里程碑意义的文章，Esteva 等人^[26]最先掀起了使用 CNN 来实现皮肤病图像分类的热潮。他们使用 129450 张图像来训练二元（良性/恶性）分类器，其性能可以达到与专业认证的皮肤科医生相同的水平。此后，一些国内外的学者在此文章的鼓舞下也开始相关研究，何小宇等人^[27]通过五种主流网络结构验证了 CNN 在基底细胞癌和色素痣上的分类性能，王诗琪等人^[28]也利用 CNN 建立了色素痣和脂溢性角化病的二分类模型。Milton 等人^[29]对比了 5 个经典卷积神经网络

模型的病种分类诊断系统,在 ISIC2018 数据集上获得了最好的效果,并提出更大的数据集以及更优的超参数将有助于模型效果进一步提升的结论。此外,还有一些针对于具体大类病种的多分类任务,如 Singhal 等人^[30]在皮肤疾病的七分类任务上对比了不同网络的性能。Polevaya 等人^[31]利用深度学习网络对图像中主要形态学部分进行四分类任务。

深度学习的方法会对数据量的需求带来爆炸式的增长,但数据标注会占用大量的时间成本和计算成本,因此领域内相关的文献研究也致力于解决这个矛盾。除了一些数据增强方法,从数据本身角度提升分类性能,如 Qin 等人^[32]利用生成对抗式网络生成高质量的病变图像,切实地提升了分类模型的效果。此外还有以迁移学习为主的非数据增强方法,这也能够让深度学习模型在局限的数据量中获得令人满意的效果。如 Jaworek 等人^[33]采用预训练的 VGG-19 网络进行术前黑色素瘤厚度评估。实验结果表明,该算法能在皮肤镜图像上实现了最先进的黑色素瘤厚度预测结果。同样的, Hekler 等人^[34]利用预训练的 ResNet-50 网络在组织病理学图像上诊断痣和黑色素瘤两分类。Gu 等人^[35]提出渐进式迁移学习的概念,即在 ImageNet 和目标数据集之间,借由另一个皮肤数据集作为中间数据集执行先一步的微调,效果要远优于一步式迁移学习。

现有模型中的某些超参数或者模块并不能适配所有的分类任务,有些研究通过调整超参数的方式来优化网络性能。为了减缓人工的试错调参过程, Kwasigroch 等人^[36]提出了神经体系结构搜索方法来自动化这一过程,在减少参数量的路上作了很多努力。在已有 CNN 模型的基础上, Muckatira 等人^[37]得出通过修剪网络的部分参数能相比原网络提升超过 10%,修剪是通过在整个模型的范围内将低于设定阈值的权重置零来实现的。随着 CNN 模型被进一步的研究论证,构造出了一系列改进的 CNN 网络模型。深度学习的最新进展包括已知与传统 CNN 相比在相同的计算复杂度下具有更高准确性的空洞卷积。Ratul 等人^[38]基于空洞卷积对比了 4 个经典网络架构的转移学习效果,比任何已知方法都具有更好的总体和每个类的准确性。然而,由于 CNN 直接输出病灶类别,这样的诊断普遍存在缺乏可解释性的现象。碰巧的是,基于可解释性内容的图像检索有望能作为临床决策的补充^[39]。Allegretti 等人^[40]提出一种皮肤图像检索方法,嵌入层串接在特征层后,通过其中的距离损失来评估与目标图像之间的相似程度,能达到人机大战的预测平均水平。

此外,在分类网络中加入注意力机制也是一种提升网络性能的方向,其本质上是给输入特征向量学习一个权重,以便突出显著特征并抑制不相关的特征,最近已广泛应用于皮肤病分类^[41]。Barata 等人^[42]提出了一个深度注意力模型,用注意力模块结合神经网络与 LSTM,有助于能够识别皮肤损伤中的相关区域并指导分类决策。Aggarwal 等人^[43]针对类间相似性以及皮损关键特征的关注度偏重,提出了一种注意力引导的深度卷积神经网络来实现黑素瘤的良恶性二分类,结果显示,一个有注意力的模型可

以将一个正常的 CNN 模型结构的精确度提高超过 10 个百分点。同样的, Zhang 等人^[44]也提出一个 ARL 注意模块来利用上层学习到的特征来生成低层的注意力特征图。

虽然研究文献在权威数据集上的准确率不断刷出新高,但是实际的任务要求会导致无论网络提出的新旧,在具体皮肤疾病的诊断上都会各有千秋,所以集成学习^[45]的思路被用来减少预测的误差。有的研究会在特征层面集成,比如 Tang 等人^[46]提出结合图像的全局以及区域局部特征,并集成 4 种不同尺度的图像输入来进一步提高分类的能力。Sabbaghi 等人^[47]利用特征包模型来聚类 SIFT 特征和颜色特征的组合,并输入到堆叠稀疏自动编码器来通过无监督方案完成皮肤病变分类任务。而有的研究会侧重于最终的决策阶段,比如 Wang 等人^[48]尝试利用拼接五种经典的卷积神经网络模型的输出向量来实现皮肤病变的七分类,实验中还结合贝叶斯搜索算法得到的带有最佳参数的分类器,达到了高效准确率。Mahbod 等人^[49]所提出的方法由侧重不同特征的不同体系结构的多组 CNN 构成,在后期融合了多个神经网络的高级特征,实验得出融合后的特征具有更好以及更充分的特征描述能力。Perez 等人^[50]在 ISIC 2017 数据集上对黑色素瘤分类进行反复评估得出 135 个模型,发现无论怎么组合,多个模型的集成都比单一模型要有优势。

此外,以优化策略和目标函数为研究点的方法也为提高模型性能方面提供了很大的帮助。现有的皮肤病研究应用的 CNN 模型大多采用经典损失函数,如交叉熵损失函数等,而在具体的皮肤疾病任务上,这样就可能限制了模型从皮肤图像中进一步学习识别特征的能力。为了解决上述问题,Ahmad 等人^[51]提出了一个新的框架,尝试基于三重态损失函数和微调模型来改进皮肤病的分类。除此外,针对训练过程中暴露出数据集问题,Lin 等人^[52]描述了一种新的损失函数用来改善不平衡数据集以及难分样本的干扰。而 CNN 模型在测试和训练阶段,不同的激活函数解决数据集非线性因素的能力不同。因此,考虑到模型在数据集的特异性,Goceri 等人^[53]分析和对比了两种活化函数在四个网络模型上的临床图像自动诊断效果。

除了改进损失函数这种训练层次的策略,在预处理层次也能做出贡献。为减少具有成本效益的皮肤数据标注代价,Shi 等人^[54]仅采用样本选择策略的主动学习以及进一步添加样本扩充策略来实现高水平的皮肤病变分析。实验用一半的数据集在两种不同的任务上达到最先进的性能。同时,Bdair 等人^[55]提出一种新的半监督学习模型,这是一个通过共享训练样本的知识来生成未标注样本伪标签的过程,相比于基线提高了超过 15%。现有文献中的“出色性能”并没有在现实中得到广泛应用,可能是因为真实数据存在的不确定性,这容易导致对自动诊断缺乏信心或结果解释中的错误。与之前方法不同,Bagchi 等人^[56]充分利用有限的数据来提高精度。在诊断模型之后,串接了一个重识别网络用以确认诊断模型是否预测正确,其他未知类别的数据用来训练这个网络。Combalia 等人^[57]使用蒙特卡罗遗漏和几种能增加训练数据不确定性的估计

技术来提高分类器的真实诊断性能。

从皮肤疾病图像分类算法的应用来看, 尽管目前大量的研究都基于皮肤镜图像获得了和医师诊断水平相当的结果, 但由于皮肤镜图像获取条件的苛刻性, 最适合实际应用的数据类型还是临床图像。而对于此类的图像的实际应用研究还是不足的, Khamparia 等人^[58]为皮肤图像中的黑素瘤分类提供一种新颖的, 由物联网驱动的深度学习的框架, 可以远程使用以协助医学专家诊断和治疗皮肤癌。Hameed 等人^[59]提出一个可移动的皮肤损伤分类专家系统应用。该系统能快速实现临床图像皮肤类别的四分类, 既便于患者自查, 又能较大程度地简化专业医师的初步诊断步骤, 从而使得专家可以更专注于疑难疾病的精细诊断。但是, 当该模型在不同照明条件和距离下由不同相机拍摄的多个智能手机图像上进行测试时, 很难实现相同的诊断精度。

通过以上的文献内容, 皮肤疾病分类算法的研究现状可以总结为以下几点:

- 1) 目前的研究方法主要以深度学习方法为主。由于深度学习方法具有很强的特征提取能力, 在大数据的加持下, 已经在很多疾病分类任务上远远超过了人工提取特征的结果。
- 2) 对于疾病的种类而言, 目前的研究主要针对于少数病种比如皮肤肿瘤, 但缺少对大范围多病种的分类。主要原因在于, 一方面目前的数据集存在数量和种类的局限性, 另一方面, 则是不同病种间的特征差异过于复杂, 目前皮肤疾病的分类学不够完善。
- 3) 缺少基于临床图像的应用研究。临床图像具有获取便利的优势, 但同样也具备了质量相对较低的劣势。目前的分类方法很难可以达到实际应用的水平, 且往往单纯的分类任务在敏感性和特异性不高的前提下很难具备应用价值。

1.3 研究目的与主要研究内容

1.3.1 研究目的

在国内医疗资源不平衡的背景下, 皮肤疾病辅助诊断的研究愈发重要。目前以国内人群为主的临床皮肤疾病数据存在空缺, 且目前的辅助诊断算法针对的病种单一, 缺少应对多病种的算法结构, 此外临床图像的应用研究方面存在明显不足。本文旨在建立起一个以中国人群为主的大规模多病种临床皮肤疾病图像数据集, 并在此数据集的基础上研究出以多病种为主要目标的辅助诊断分类算法, 致力于将临床皮肤病图像的分类应用于实际的医疗环境中。

1.3.2 研究内容及本文结构安排

本文的主要研究内容为如何建立高质量的适应于辅助诊断算法的临床图像数据集, 并以此为基础, 通过优化算法分类过程的方式提升算法的性能并实际应用, 具体

分为以下几点：

(1) 临床皮肤病图像数据集的建立。本文联合湘雅医院收集了以中国人群为主的临床图像数据集。由于临床图像相较于其他类型的数据，存在质量偏低的问题，如何对数据进行整理与清洗是需要研究的内容，此外还需要对建立的数据集进行最终的评估。

(2) 分类粒度对辅助诊断分类算法的影响研究。由于皮肤疾病的类别在实际的临床诊断中也存在不同的粒度问题，越细的粒度往往具有越强的监督信息。为了之后的多病种甚至全病种研究做铺垫，需要研究精细粒度的标签对分类算法的影响。

(3) 针对于多病种的分层诊断算法研究。皮肤疾病的种类过于繁多，不同类别间的类间差异相差较大，如何通过建立具有层次性的皮肤疾病分类体系是值得研究的问题，并需要研究在此基础上如何进行有效的训练及预测。

(4) 基于恶性程度分级的健康管理应用研究。现有的分类算法难以实际的应用于临床环境。本文以皮肤肿瘤为切入点，研究如何对皮肤肿瘤的临床图像进行恶性程度分级，从而做到对病人进行病情初筛，帮助病人进行健康管理，缓解临床诊断压力。

本文共分六个章节，结构安排如下：

第一章：绪论。首先本章会对皮肤疾病辅助诊断领域的背景以及算法背景作介绍，并解释本文的研究意义。之后，本章会对国内外的数据集以及分类算法方面的研究现状作详细介绍，总结出现阶段该领域研究存在的问题，提出本文的研究目的，以及研究的主要内容。

第二章：介绍临床皮肤病数据集的建立过程及评估。本章会详细介绍数据集的获取途径及方式，并侧重于之后的数据清洗、标注以及后处理过程。最后，根据现有的评价指标，在常用的卷积神经网络上进行基准实验，并与现有的国外数据集进行比较，突出建立国内数据集的必要性。

第三章：研究皮肤病分类粒度对分类算法的影响，并提出两阶段分类方法。本章在第二章的数据基础上，以银屑病为例，研究不同“非银屑病”的分类粒度在不同网络下的分类效果，对比出两阶段网络结构的优势，并通过一系列实验及“人机大战”验证其泛化性能。

第四章：研究针对于多病种的分层诊断算法。本章选取了 19 类具有高发病率且具有代表性的疾病作为研究对象。在同时考虑皮肤病学分类以及视觉相似性的条件下，先将 19 类疾病分成了不同粒度的两层分类结构，并以病例为单位建立了分层交叉验证数据集。之后，本章会介绍对于该数据集的数据增强方式，以及参数选择方法。最后，本章参考第三章的研究成果，提出了基于细粒度的特征迁移方式，并结合粗细粒度介绍了分层诊断算法的详细步骤及算法评估。

第五章：介绍皮肤肿瘤恶性程度分级的应用研究。本章将首先介绍目前分类算法

难以应用于临床环境的背景以及研究皮肤病恶性程度意义。之后，本章会介绍应用的基本的功能以及算法实现，并通过实验对其应用性进行评价。

第六章：总结与展望。总结出本文的优势以及不足，并提出未来的改进思路

第 2 章 临床皮肤疾病数据集——XiangyaDerm

目前已有的公开可获取数据集大多为欧美人口为主，图像上存有水印，皮肤病种的覆盖面不够，因此本文联合湘雅医院构建了以中国皮肤病人群为基础，具有病人病理检验结果作为金标准的高质量大规模数据集。本章将从数据集的获取途径、数据清洗、标注及后处理方面阐述数据集的建立过程，同时通过几种深度学习网络的基准表现，验证数据集的性能，并通过与国外数据集在算法适用性上的对比，突出构建面向中国人群的皮肤病数据集的必要性。

2.1 数据集的建立

2.1.1 数据的获取与清洗

由于数据来源于湘雅附一医院，所以数据集命名为 XiangyaDerm。数据集的收集获得了湘雅医院伦理委员会批准，并取得所有参与者的知情同意。所有临床图像均由湘雅医院皮肤科医生在标准光照下，使用索尼 DSC-HX50（350dpi）、佳能 IXUS 50（180dpi）、尼康 D40（300dpi）、尼康 COOLPIX L340（300dpi）四种不同摄像机捕获，分别采用以下几种分辨率：3888×5184、1944×2592、2000×3008、3864×5152。全部数据集共获得了 47075 张图像，覆盖 541 种皮肤病，涵盖 99% 的中高发病率疾病。每幅图像的诊断标签都是以病理学检验结果作为金标准进行验证的，并附有患者完整的病史为依托，图 2-1 中展示了一些示例图像。从图中可以看出，这些图像质量高，皮损区域主要集中在图片中央。但是不同疾病之间同样存在很高的相似性，比如基底细胞癌与脂溢性角化（图中第一行和最后一行），这两种疾病的皮损特征主要呈现出黑色的原点状病变；而患有类天疱疮与天疱疮（图中六行和第七行）的患者，皮肤上则会出现透明状的大疱，即使医生单纯借助图像也很难分辨出差异；同样的情况还存在于湿疹、扁平苔藓、红斑狼疮、银屑病等均存在伴有大面积红色斑块的皮肤疾病，这些极其相似的特征，无不给皮肤图像的识别任务带来了很大难度。

本文协同湘雅医院皮肤科医师一起完成了数据清理过程。在这个过程中，为保证每张图像都存有疾病的关键性视觉特征，存在以下五种情况的图像会被清除：

- 1) 由于拍摄不当，图像质量低，存在高度模糊的情况，或者没有采集到关键的皮损区域。
- 2) 皮损区域被肉眼可见的治疗药物或者有颜色的药物残留覆盖，可能对算法训练过程造成一定的不利影响。
- 3) 图像大部分被身体毛发或其他器官遮挡，导致很难观察到清晰完整的病变区域。

- 4) 一些特殊部位比如手指尖、耳尖等，皮肤面积很小，不利于观察与提取特征。
- 5) 渗出液过多，导致正常皮肤或者病变区域丢失原来的纹理和外观。



图 2-1 XiangyaDerm 数据集中的样例

2.1.2 数据集的标注与后处理

数据的标注过程由专业皮肤科医生完成，这些医生均具有 5 年以上的临床工作经验，且所有的标签都经过了病理组织学的检验，并由相应的诊断报告作为依托，如图 2-2 所示。在标注过程中，团队成员会被分为两组，每组在负责一半图像的标注的同时，会检查来自另一组图像标注是否正确。标注的工具是可获取开放源码的图像标注软件 labelImg，医生会对图像上的明显病变区域使用带标签的边界框框出，既定位出了病变区域的具体位置，同时也标出了疾病的类别标签。

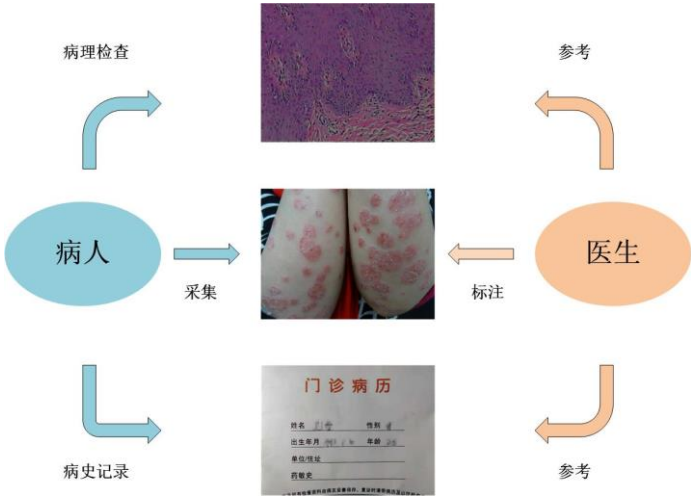


图 2-2 数据集的采集与标注

考虑到实际的临床图像相较于皮肤镜图像而言，一般存在着较大面积的背景干扰区域或者正常皮肤区域。而在常用的图像识别算法中，图像在输入端都会进行一定程度的压缩，这样会导致本身占比不大的皮损特征，会被进一步的缩小，从而丢失一定的细微特征。相较于 ImageNet 的物体识别，本身就属于细粒度分类的皮肤病分类则更加关注于这些细微的特征，所以对原始图像进行一定的后处理是很有必要的。

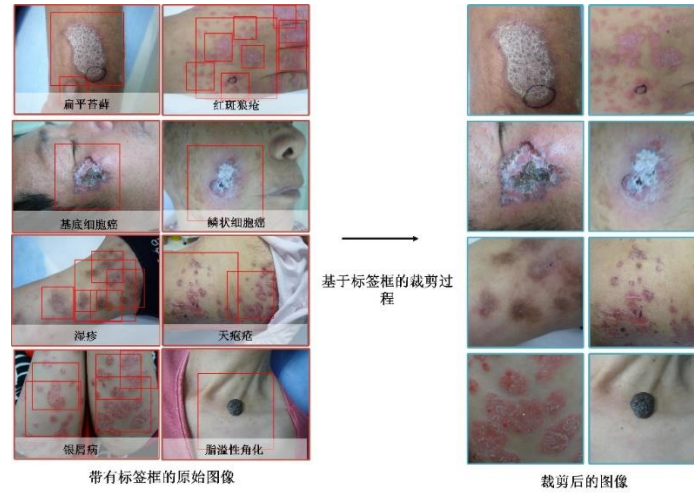


图 2-3 数据集的裁剪过程

如图 2-3 所示，文本利用已有的边界框裁剪出主要的皮损区域，通过裁剪操作，不仅去除了临床皮肤病图像的复杂背景，而且增加了数据集的图像数量，因为一张图像可能有几个孤立分布的皮损块。最终 XiangyaDerm 的数据量从 44108 增加到 107565，共有 541 类皮肤病，其中银屑病的数据量最大，为 67066 张，占总整个数据集的 62%。除了本身的发病率高之外，湘雅附一医院具有很多全国知名的银屑病专家也是其中一个原因，每年都有很多银屑病患者前去就诊。除银屑病外，其余 540 种皮肤病的数据分布如图 2-4 所示，水平轴和垂直轴代表疾病及其相应的数据量。

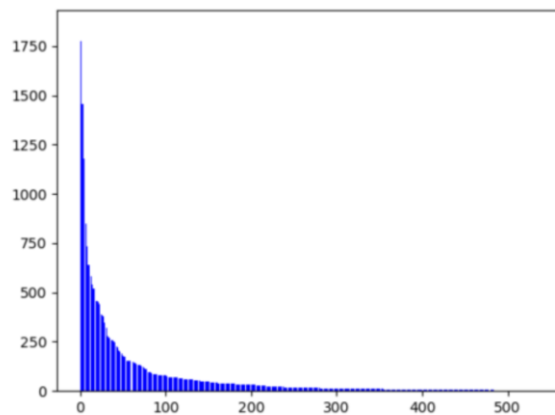


图 2-4 XiangyaDerm 的数据分布（横坐标表示病种的索引号，纵坐标表示数量）

2.2 数据集的评估

2.2.1 评价指标

为了能评估出数据集在常见 CNN 下的性能，获取其基准表现。本章使用到的评价指标为以下几种形式。

(1) 准确率 (Accuracy, ACC)。准确率是所有模式识别问题中，最基础、直接的评价指标，可以表示一个模式识别任务中，被正确识别数量的占比，如公式 (2-1) 所示，其中 N_{true} 表示正确分类或预测的数量， N_{all} 表示需要被分类或预测的总数量。

$$ACC = \frac{N_{true}}{N_{all}} \quad (2-1)$$

(2) 混淆矩阵。混淆矩阵也称为可能性表格，是一种使用特征矩阵来表示分类算法性能的可视化方法。混淆矩阵可以明显地看出类别之间的分类情况，相比于准确率 ACC，可以详细地分析结果的好坏。一般情况下，每行代表一个真实的类别 i ，而每一列代表每一个预测的类别 j 。所以混淆矩阵中的每个元素 (i, j) 就代表了在真实标签为 i 类的情况下，预测为 j 类的数量或者百分比。由于皮肤病数据天然存在不平衡的现象，为了使每类间可以进行比较，本文的混淆矩阵均表示百分比。值得一提的是，混淆矩阵的对角线上的值越高，表示模型的分类性能越好，每个矩阵元素的百分比值就等于这个类别的准确率 ACC。

2.2.2 数据集的基准表现

考虑到皮肤病数据集天然具有很高的类间相似性，特征复杂，越先进的网络不一定具有越好的表现，本章选择了 4 种常用的具有不同基础结构的 CNN 网络，包括 InceptionV3^[60]、InceptionResNetV2^[61]、DenseNet121^[62]和 Xception^[63]，对 80 种较为常见皮肤病进行分类。这 80 种常见的皮肤病的数据量均大于 100，为了使整体的数据集相对平衡，本章刻意去除了一些疾病的数据量，使得每类的数据量最大不超过 1000，比如银屑病的数据过多，如果使用全部的银屑病数据，则会导致银屑病的准确率比较高，而其他疾病则很大概率被预测为银屑病。

为了保证每类测试集的数量足够，本章采用四折交叉验证，以 3:1 的比例分成训练集和测试集，考虑到某些疾病的病例数不多，本次实验没有按照病例区分训练集与测试集。整个训练过程是在三块 NVIDIA TITAN Xp 的显卡上完成的。InceptionV3、InceptionResNetV2 和 Xception 的图像输入大小均为 $299 \times 299 \times 3$ ，Densenet121 的图像输入大小为 $224 \times 224 \times 3$ 。其他参数及实验条件均保持了一致性，比如：均在 ImageNet 数据集上加载了预训练权重，最大训练次数 5000，基本学习率 0.001，批次大小为 25，优化器 Adam，采用了交叉熵损失函数，最终的实验结果如表 2-1 和图 2-5

所示。

表 2-1 各类算法实验结果对比

方法	InceptionV3	DenseNet121	Xception	InceptionResNetV2
Top1 ACC	0.470	0.494	0.523	0.588
Top3 ACC	0.671	0.696	0.707	0.764

从图 2-5 所示的混淆矩阵中可以看出，这些 CNN 都具有相对较好的性能，混淆矩阵的深绿色区域主要分布在对角线上，而其他区域的颜色相对较浅。结果表明，80 种皮肤病中，除少数不易鉴别外，大部分可由所采用的 CNN 进行有效鉴别疾病。分析图中的浅色区域发现，这些疾病大多具有较少的数据量，同时很多疾病对于医生来说，也很难从中找到有效的区分特征。从表 3-1 的 80 分类实验结果中还可以得出初步的结论：在 XiangyaDerm 数据集中，InceptionResNetV2 比其他 3 种网络具有更好的性能。值得一提的是，本节的目标不是找到最佳的识别网络，而是验证所提出的数据集的有效性并给出它的基准表现。

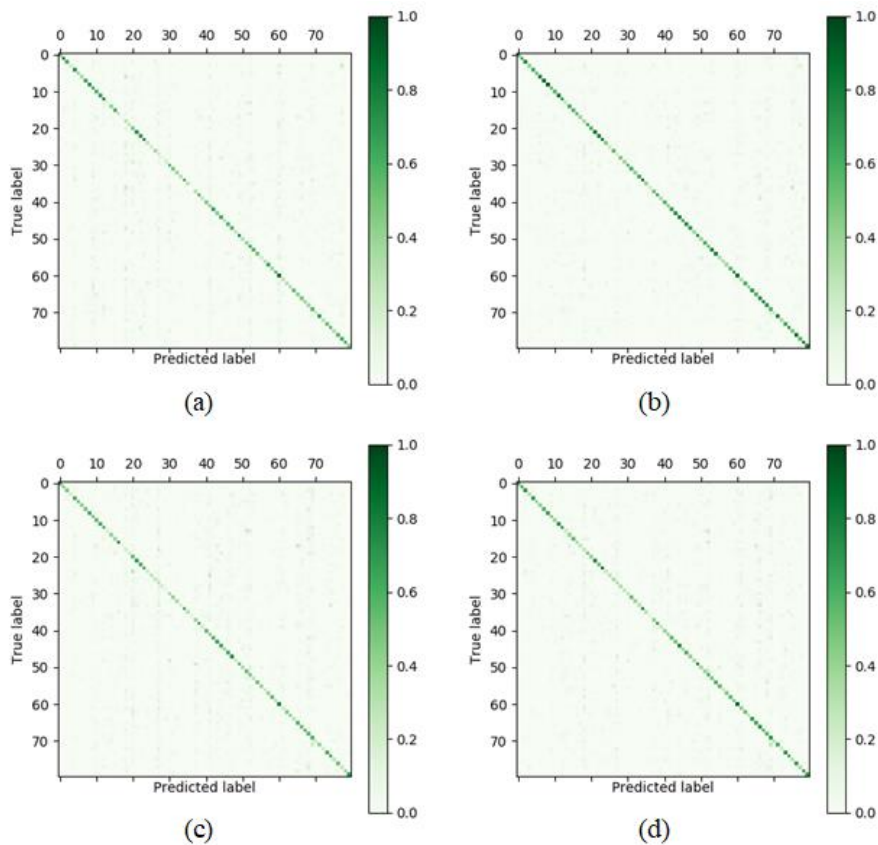


图 2-5 80 类分类实验的混淆矩阵(a:InceptionV3, b:DenseNet121, c:Xception, d:InceptionResNetV2)

2.2.3 国内外数据集差异性研究

由于不同人种之间的皮肤病患病区域的视觉特征是存在差异的，为了在算法层面

上,体现出这种差异性,本节在不同的数据集间进行了交叉测试(在其中一个数据上训练,其他的数据集上测试),从而验证建立国内皮肤疾病数据集的重要意义。本节用到的国外数据为 Derm101,且已经获得了数据集官方的研究许可。在病种上,本节选择了 6 种高发病率的常见疾病,包括基底细胞癌,表皮样囊肿,牛皮癣,酒渣鼻,脂溢性角化病和淤积性皮炎,它们分别在 Derm101 以及 XiangyaDerm 中有 100 张以上的数据量,形成了两个子数据库 Derm101-6 和 Xiangya-6。而数据集 Mix-6,则是由 Derm101-6 和 Xiangya-6 混合组成的数据集,用来验证拥有两种混合特征训练集的训练效果。在实验设置上,本节使用 InceptionResNetV2(这是 80 分类实验中效果最好的 CNN)对 Derm101-6, Xiangya-6 和 Mix-6 进行交叉测试,损失函数,批次大小和其他参数设置均保持与以前的实验相同。该实验获得的测试结果如表 2-2 和图 2-6 所示。图中,(a)表示在 Derm101-6 上训练和测试,(b)表示在 Derm101-6 上训练, Xiangya-6 上测试,(c)表示在 Xiangya-6 上训练, Derm101-6 上测试,(d)表示在 Xiangya-6 上训练和测试,(e)表示在 Mix-6 上训练, Derm101-6 上测试,(f)表示在 Mix-6 上训练, Xiangya-6 上测试。

表 2-2 国内数据集与国外数据集交叉测试下的 ACC 值

	Derm101-6(测试)	Xiangya-6(测试)
Derm101-6(训练)	0.800	0.193
Xiangya-6(训练)	0.213	0.720
Mix-6(训练)	0.671	0.621

从图 2-6 的 (a) 和 (d) 中可以看出,深色正方形集中在对角线上,这表明每种疾病的分类效果在各自数据训练和测试时效果比较不错,而当交换测试集时,从 (b) 和 (c) 可以看得出,准确性却差得多。而 (b) 和 (c) 均属于在其中一个人种的皮肤病数据集上训练,在另外一个人种的皮肤病数据集上测试,这样的差异本质上是由于不用人种的肤色以及基因不同导致所患疾病在临床图像的视觉特征上出现差异性。为了解决这种因特征不一致导致的泛化性能降低,本文尝试在训练时,加入来自两种人种的皮肤疾病特征(如 (e) 和 (f) 所示),从结果可以看出,要远好于之前单一特征的结果,但由于同一类别的类内差异增加,导致整体性能上会比在在同一数据集上进行训练的结果要差,这就需要在泛化性能上和准确率做出一个权衡。

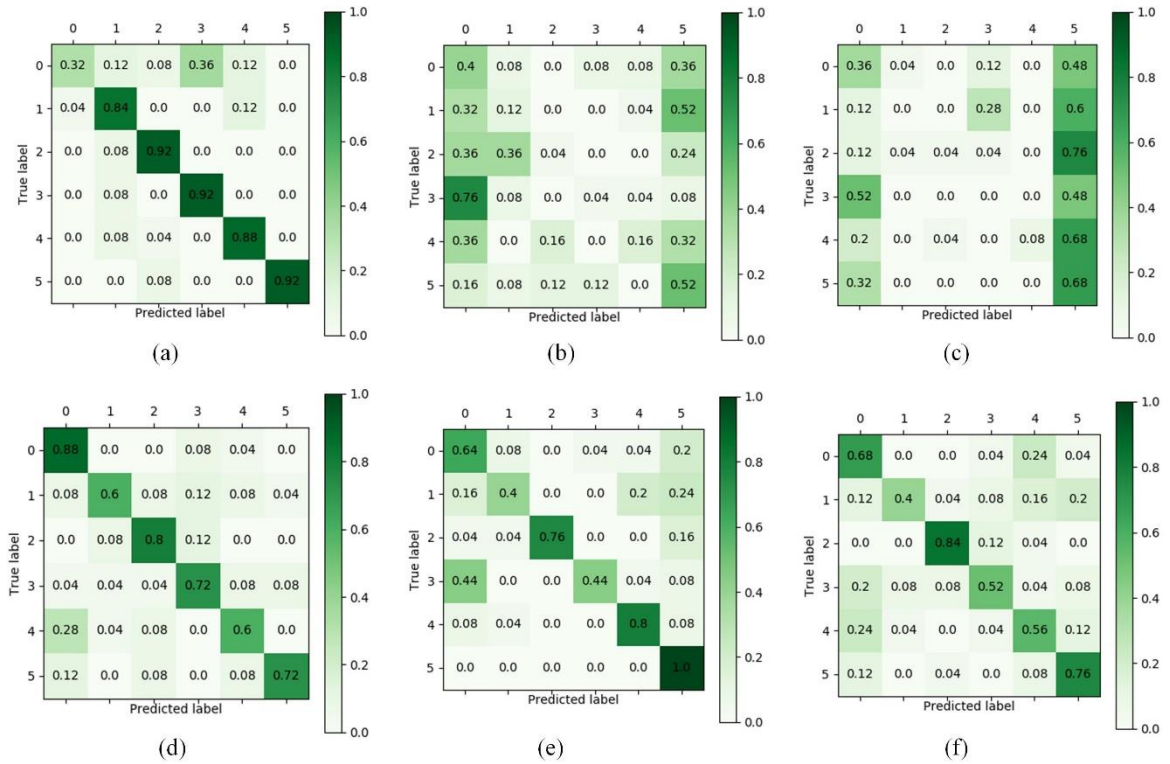


图 2-6 6 分类交叉测试的混淆矩阵

2.3 本章小结

本章主要介绍了一个全新的皮肤疾病临床图像数据集——XiangyaDerm，数据来源于中国皮肤病人群，覆盖了 541 类皮肤疾病，每张图像都有病理检验上的结果作为依托，后处理后的总数量达到了 10 万量级。本章具体介绍了数据获取、清洗、标注以及后处理的一些细节，为了评估数据集在深度学习分类算法上的整体表现，选取了四种常见的深度学习网络进行了训练与测试，并利用混淆矩阵进行了结果上的对比。本章的最后，利用 XiangyaDerm 与国外最权威的临床数据集之一的 Derm101 进行了对比，从算法层面上，说明了国内外皮肤病图像的特征差异，提出了拥有国内数据集的必要性。

第3章 基于精细粒度的两阶段分类方法

皮肤疾病图像的分类本身属于细粒度分类，理论上，越是能获取到更加精细的分类粒度特征，越是有利于分类网络的性能提升。本章将探索在皮肤疾病分类任务中采用更加精细的分类任务是否有利于最终的识别结果，并提出了一种基于精细粒度的两阶段分类方法。由于在 XiangyaDerm 数据集中，银屑病的数量占绝大多数，且银屑病识别在临床上本身就具有重要意义，所以，本次实验采用的疾病类别为银屑病及其四种“近亲”皮肤疾病和“远亲”皮肤疾病（根据视觉上的相似度定义），采用的基础结构为四种常见的卷积神经网络，通过两阶段方法的实验效果以及和一阶段的对比结果，验证两阶段是否有效，并通过“人机大战”进一步确实其泛化性能及实际意义。

3.1 分类网络的细粒度

3.1.1 理论基础

卷积神经网络算法本质上是一种优化算法，学习表达该系统的最优参数，而寻求最优参数的核心算法即为梯度下降算法——以损失函数为目标，计算该目标下各个参数的梯度，顺着梯度减小的方向更新参数。使用卷积神经网络的分类网络中，常用的损失函数为交叉熵损失。在解释交叉熵之前，需要了解信息熵概念。信息熵表示的是对所有信息量的期望，而信息量则是与信息发生的概率成反比，所以信息量 $I(x)$ 和信息熵 $H(X)$ 可以用以下公式表示：

$$\begin{cases} I(x) = -\log(p(x)) \\ H(X) = -\sum_i^n p(x_i) \log(p(x_i)), X = x_1, x_2, \dots, x_n \end{cases} \quad (3-1)$$

交叉熵表示的是两种不同概率分布的差异程度，在分类算法中，就表示为预测概率分布与真实概率分布的差异值，预测和真实概率分别用符号 $q(x_{ij})$ 和 $p(x_{ij})$ 表示，所以多分类的交叉熵可以用如下公式表示：

$$H(p, q) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n p(x_{ij}) \log(q(x_{ij})) \quad (3-2)$$

其中 m 表示批次的数据量， n 表示类别数，如果 n 的数量为 2，由于真实标签为 0-1 分布，所以公式可以直观地表现为以下形式：

$$H(y, p) = -\frac{1}{m} \sum_{i=1}^m [y \log(p(x_{ij})) + (1 - y) \log(1 - p(x_{ij}))] \quad (3-3)$$

粒度在统计学上表示的是同一维度下，数据统计的粗细程度^[64]，在分类任务中，就可以表示为类别划分的粗细程度。就同一份数据而言，粒度越细，所划分出来的类别数量就越多，类别内的差异就相对越小。从以上的公式也可以看出，在数据一致的前提下，类别越多，交叉熵损失函数所带来的约束就越强，理论上多类别有着更强的监督信息，更有利于卷积神经网络进行特征的提取。所以本文尝试进行细化分类粒度，

构造两阶段来提升分类的性能。

3.1.2 两阶段分类方法的网络结构

基于通过上节内容的构想，本节提出了两阶段网络的基本结构，通过获取到更多监督信息，构建相对细粒度网络，作为中间结果，最后通过分类决策，得到最终的相对粗分类的分类器。网络结构如图 3-1 所示。

在第一阶段，可以采用 ImageNet 预训练的方式，在更多分类标签的情况下，训练一个相对于最终分类器具有更加精细粒度的模型。这个模型可以利用卷积操作进行特征提取，并利用 softmax 激活函数得到所有类别的预测值。这里的预测值可以预测更多的类别，可以相对减少训练时的类内差异，有效地提高拟合效果效率。第二阶段则是融合策略，利用先验的类别知识，将属于同一粗粒度类别的结果融合到一起，最后进行不同粗粒度的分值的对比，得到最终的分值。综上，该网络结构具有两点优势：一是利用强监督信息，可以优化特征提取过程，二是降低分类的类内差异，获取的特征更具有针对性。

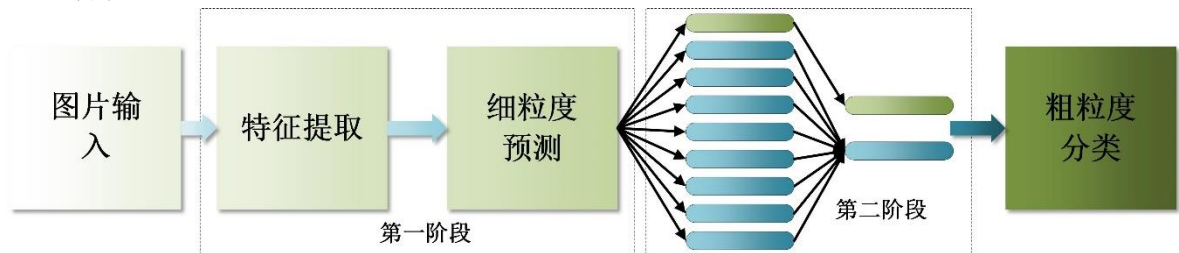


图 3-1 基于精细粒度的两阶段分类网络

3.2 两阶段分类实验方案设计

本节将对提出两阶段网络进行实验评估与验证，通过几种不同的网络的两阶段算法的实验结果，说明算法的有效性，同时通过与一阶段方法的对比实验，验证基于精细粒度的两阶段分类方法的优越性。在本节的最后，会通过与实际医生的临床图像识别能力对比，验证其泛化性能。

3.2.1 实验数据集

在 XiangyaDerm 数据集中，银屑病（又称为牛皮癣）的图像数量占总数据集的一大半，这种慢性自身免疫病十分容易复发，目前没有药物可以根治，所以对银屑病的早期诊断和及时发现极为重要，可以有效避免严重并发症（如心血管，肝和肾并发症）。据统计，全世界已有超过 1.25 亿的牛皮癣患者被记录^[65]，患病率为 1-3%。但是，由于医疗资源的限制以及皮肤病学的医学知识有限，很多病人会被漏诊或牛皮癣被误诊为其他疾病。因此，本节内容以银屑病作为实验的识别对象，是十分合理地选择。首

先在直觉上，牛皮癣的识别任务可以看作是“银屑病”类和“非银屑病”类的二元分类任务。然而，“非银屑病”类在实际的皮肤病学上，并不是由一种皮肤病疾病组成，排除银屑病在内的所有疾病，都可以组成“非银屑病”，但是这样“非银屑病”类别的类内特征差异就会十分大，模型训练过程收敛变得困难。为了通过精细粒度的分类任务，来更有效获取到银屑病的关键特征，在本节实验中，“非银屑病”类被定义为八类皮肤病，其中四类为银屑病视觉上的近亲病种：扁平苔藓、副银屑病、红斑狼疮、湿疹，另外四类为银屑病视觉上的远亲病种：基底细胞癌、鳞状细胞癌、天疱疮、脂溢性角化。通过以区别远亲疾病为学习目标，可以得到一些更宏观的特征，比如颜色和形状，通过以区别近亲疾病为学习目标，则可以得到一些更细微的特征，比如纹理等。为了降低数据集不平衡对网络带来的影响，在数据量的设定上，这九类病种保持着高度一致性：除了基底细胞癌 832 张，脂溢性角化 889 张，其余均为 900 张。

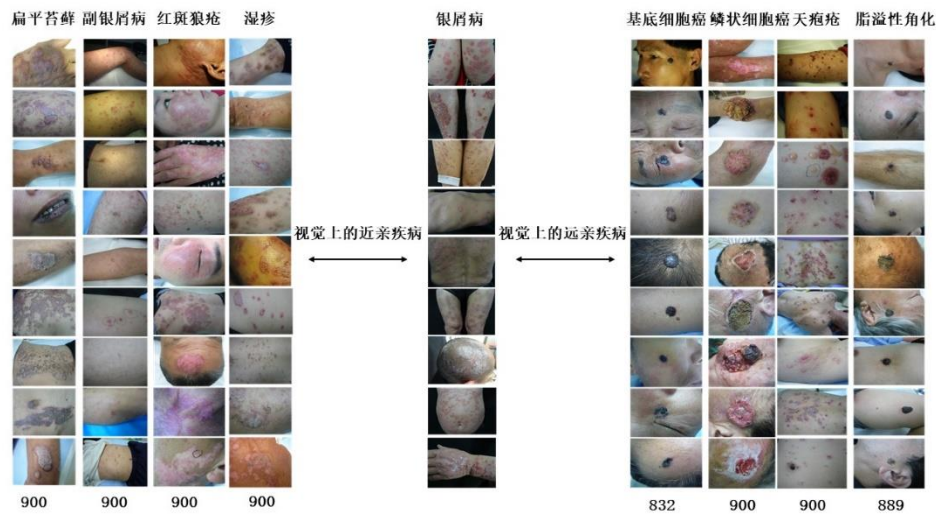


图 3-2 两阶段分类方法数据集

3.2.2 算法评估准则

由于本章的任务为银屑病的二分类任务，所以评估准则会采用常用的 ROC 曲线（受试者工作特征曲线）以及 AUC 值，它与 P-R 曲线（精确率-召回率曲线）相比，具有更好的稳定性，能够降低测试集数据分布不均带来的影响，从而更加客观合理地评估模型。ROC 曲线的横坐标为假阳率（False Positive Rate, FPR），纵坐标为真阳率（True Positive Rate, TPR），其概念可以从二分类的混淆矩阵来解释，如图 3-3 所示，

预测数量		预测项	
		阳性 (Positive)	阴性 (Negative)
观测项	阳性 (Positive)	真阳 (True Positive, TP)	假阴 (False Negative, FN)
	阴性 (Negative)	假阳 (False Positive, FP)	真阴 (True Negative, TN)

图 3-3 二分类的混淆矩阵及指标

所以假阳率 FPR，真阳率 TPR 则可通过以下公式得到

$$FPR = \frac{FP}{FP+TN} \quad (3-4)$$

$$TPR = \frac{TP}{TP+FN} \quad (3-5)$$

在二分类的任务中，预测值的阳性和阴性往往是根据模型预测时的置信度来设定的，当模型的输出置信度大于某一阈值时，将其认定为阳性，而低于阈值的则为阴性，因此不同的阈值设定，就会产生不同组假阳率与真阳率值。在 ROC 曲线的绘制过程中，通过将假阳率设为横坐标，真阳率设为纵坐标，每组值就可以成为坐标轴中的一个点，通过调节阈值，就可以得到一整组坐标点，再通过连接坐标点，便得到了该模型的 ROC 曲线。不难看出，分类器的真阳率越高，越能将所有的阳性样本预测出来，其敏感性（sensitivity）就越高，假阳率越低，越不容易将阴性样本错误地预测为阳性样本，其特异性（specificity）就越高，换言之，曲线中最靠近左上角的位置为该模型的最佳阈值，可以通过约登指数（Youden index, Y_i ）来确定该位置的坐标：

$$Y_i = TPR - FPR = sensitivity + specificity - 1 \quad (3-6)$$

对于模型整体的性能而言，ROC 曲线下的面积越大，则表示性能越好，这个面积则为 AUC（Area Under the Curve）值。

3.2.3 两阶段皮肤疾病分类的实验评估

本节将数据集按照 6:2:1 的比例分成三个子集：训练、验证和测试，并确保这些数据集之间没有重叠。选取了四种经典的 CNN：DenseNet、InceptionV3、InceptionResNetV2 和 Xception 作为基础网络结构。网络计算所用的设备显卡为 TITAN Xp，为了更好的训练效果，采用 ImageNet 的预训练模型，对数据进行一定程度的增强效果：旋转，缩放和水平翻转，以增加数据多样性，并对数据进行乱序处理以保持相同批次的相似图像不被训练，然后根据经验以及硬件条件将批次大小设置为 25。在

优化器的选择上,通过试验与其他优化器进行比较之后,选择 RMSprop 作为优化器。

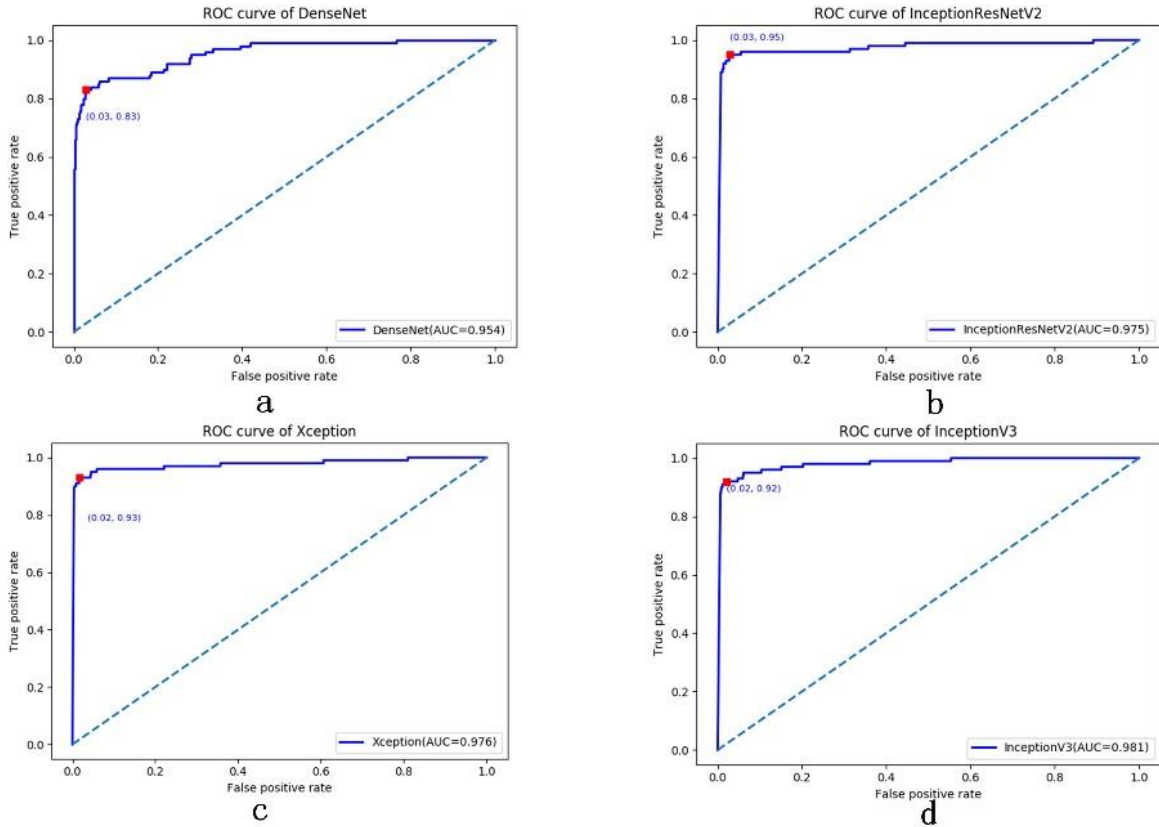


图 3-4 不同网络的 ROC 曲线(a:DenseNet, b:InceptionResNetV2, c:Xception, d:InceptionV3)

得到的 ROC 曲线如图 3-4 所示,图中的红色点表示最大约登指数的坐标点,该坐标下的具体特异性和敏感性值见表 3-1。为了评估模型的整体表现,计算各个模型的 AUC 值,用 DeLong 方法计算了 95%置信区间内的区间值。从结果上可以看出,在两阶段方法中 InceptionV3 的 AUC 值最好,为 0.981 ± 0.015 ,其次是 Xception、InceptionResNetV2 和 DenseNet121。从敏感性和特异性的最高数值上也可以看出,银屑病被漏诊的概率不超过 2%,被误诊的概率不超过 5%。

表 3-1 模型分类的 AUC 值以及最优点敏感性和特异性 (1 表示单阶段, 2 表示两阶段)

方法	AUC 值 (95%CI)	敏感性	特异性
InceptionV3-1	0.966 ± 0.014	0.96	0.91
DenseNet121-2	0.954 ± 0.024	0.97	0.83
InceptionResNetV2-2	0.975 ± 0.022	0.98	0.95
Xception-2	0.976 ± 0.022	0.98	0.93
InceptionV3-2	0.981 ± 0.015	0.98	0.92

为了进一步说明两阶段模型的优势,本节加入单阶段的模型进行了对比。由于在

上一步的实验中，InceptionV3 在四种网络中 AUC 值最优，便以 InceptionV3 为例，测试其在单阶段模型下的性能。在单阶段模型中，精细粒度的 9 分类层会被从两阶段模型中移开，即直接训练银屑病临床图像的多元分类器。数据集的划分与两阶段模型中的划分相同。为避免偶然性，实验应用了 bootstrap 策略，进行了 8 次实验。如图 3-5 以及表 3-2 所示，提出的两阶段模型（AUC: 0.981 ± 0.015 ）确实优于单阶段模型（AUC: 0.966 ± 0.014 ）。从视觉特征的角度来看，单阶段模型将 8 种疾病的所有病例分配到同一类别“非银屑病”中，而这 8 种疾病的视觉特征模式各不相同（例如基底细胞癌呈黑色凸起圆点状皮损，而副银屑病则是更接近于银屑病的红色斑块状皮损），类内特征差异较大，确实会导致“非银屑病”类别的分类边界更加模糊。

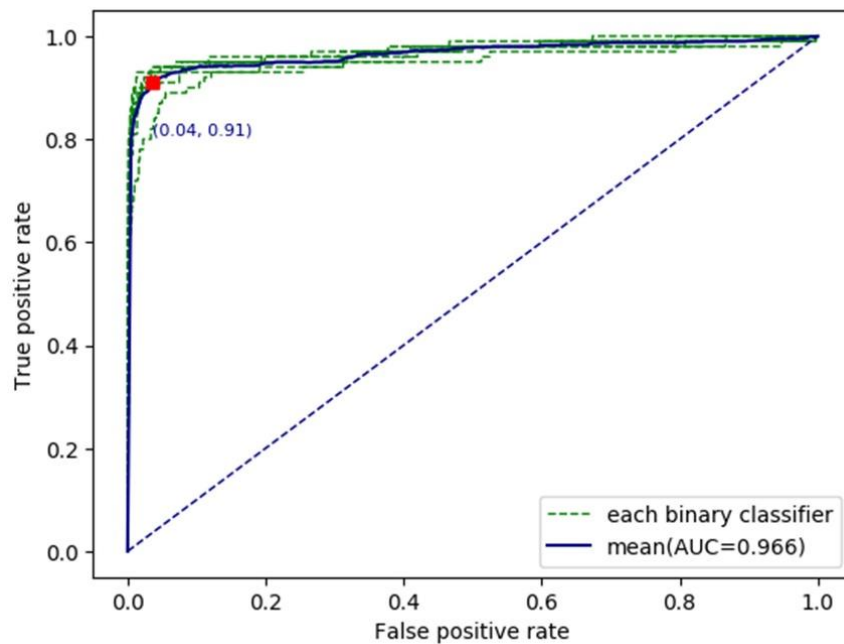


图 3-5 两阶段模型与单阶段模型对比

3.2.4 泛化性能验证

为了验证分类器在识别银屑病临床图像上的实际效果以及它的泛化性能，本文联合湘雅医院的 25 位皮肤科医生进行了一次人机大战实验。数据上，由医生选择了 100 个未包含在训练集中且较难识别的图像，作为新的测试集，并选择了 InceptionV3 作为比赛的分类器。结果如图 3-6 所示，橙色圆点表示医生的结果（绿色三角形表示均值），蓝色线表示两阶段方法的结果。从图中的最优约登指数可以看出（红色正方形点），本方法可以达到的最优假阳率为 0.04，即误诊率为 4%，可以达到的最优真阳率为 0.97，即漏诊率为 3%；而 25 位皮肤科医生的假阳率平均值为 0.10，即误诊率为 10%，可以达到的真阳率为 0.81，即漏诊率为 19%，明显低于分类器的水准。值得一提的是，

这 25 位医生是包含一部分实习医生，从图 3-6 中也可以看出，明显有一部分医生的表现要低于平均值，然而就皮肤科专家医生而言，只是略低于分类器的识别能力。在实际的诊断中，医生通常还会考虑其他信息，例如患者的病史或者斑块的厚度等等，所以就临床图片的识别而言，专业医生比不上深度学习训练出来的模型也是合理的。但是，从结果上至少可以说明一点，通过算法得到分类器可以在一定程度上辅助实习医生进行临床诊断。

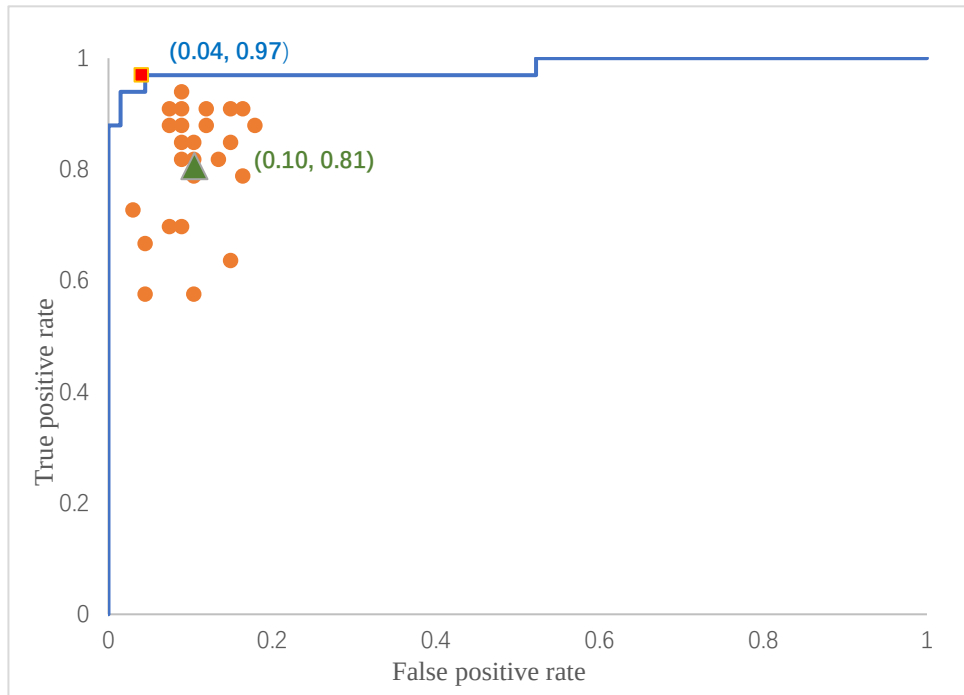


图 3-6 两阶段分类方法与医生的对比

在临床诊断中，“非银屑病”不仅限于以上网络训练中选择的几种疾病，为了进一步探索模型的泛化性能，本节从湘雅医院获取了两个新测试集进行实验，均为以上实验中“非银屑病”未包含到的病种。第一个测试集包含 50 张皮肤疾病临床图像，涉及到疤痕组织、鲍恩病、传染性软疣、苔藓样淀粉样变、汗孔角化病、黑色素瘤、尖锐湿疣和寻常疣等，这些疾病与银屑病有着差异较大的外观，属于未包含在训练集中的视觉上的“远亲”病种。第二组测试集同样由 50 个图像组成（如玫瑰糠疹），但是这些图像与银屑病有着相似的外观，属于未包含在训练集中的视觉上的“近亲”病种。实验结果显示，ACC 可以达到 0.88，这表明通过测试未出现在训练数据集中皮肤病种，模拟真实临床诊断中场景，本文提出的方法仍然具有良好的性能，适用于实用的计算机辅助诊断系统。

在未出现的病种上也有较高的准确率是值得思考的问题，可以猜测这个模型很好得找出了银屑病的关键性特征，因为在训练数据集的设定上，本章有意识地加入区别于银屑病的远亲病种以学习直观上的特征，加入区别于近亲病种以学习更难以观察的

抽象特征，这样在之后遇到的新病种，也可以对应到训练集中的“非银屑病”中。从图 3-7 四种两阶段模型的对于九类疾病（四种银屑病远亲疾病：LP-扁平苔藓、Par-副银屑病、LE-红斑狼疮、Ecz-湿疹，四种银屑病近亲疾病：BCC-基底细胞癌、SCC-鳞状细胞癌、Pem-天疱疮、SK-脂溢性角化）的混淆矩阵可以看出，这九种疾病之间没有明显的诊断偏差，且银屑病预测的准确性明显高于其他疾病，除了 DenseNet 的 0.8 以外，均高于 0.9 的值。

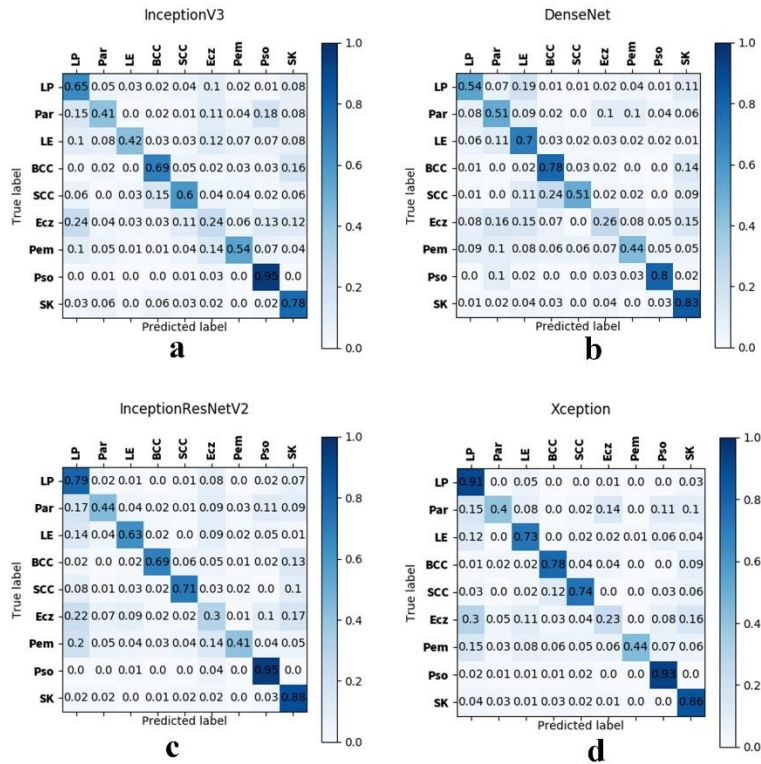


图 3-7 银屑病、四种近亲疾病以及四种远亲疾病的混淆矩阵

除此之外，每个网络中银屑病的误诊率都非常低，最低的仅为 2.5% (DenseNet)。可以看出，CNN 可以很好地提取皮肤图像的各个层次的特征，以区分银屑病与其他皮肤疾病。但是进一步分析所选的八种皮肤疾病在四种网络上平均准确度会发现，四种皮损与银屑病相似的临床疾病平均准确性较差，CNN 容易混淆它们，但这也反映了临床上的现实情况，因为即使是专业的皮肤科医生，也无法在不考虑患者病史的信息的情况下立即分辨出这些疾病。

3.3 本章小结

本章针对于皮肤疾病的分类粒度展开了研究，基于理论分析，初步认定在数据相同的情况下，精细粒度的分类具有更强的监督信息以及更小的类内差异，从而更有利于

于模型训练。基于此，本章提出了两阶段分类网络，先获取细粒度下的预测结果再通过决策融合生成最终的分类结果，并以银屑病为例，通过实验对比证实了理论猜想。同时发现合理地训练类别设置，得到的银屑病分类器具有很好的泛化性能，在“人机大战”中超过了人类医生的平均水平，并在面临一些未知病种，也能较好的判断是否为银屑病，所以本章的实验成果可以在一定程度上，辅助一些实习医生进行银屑病的诊断。

第 4 章 基于粒度迁移的分层诊断算法

通过上一章的内容，可以发现分类粒度对皮肤疾病分类网络的影响：用细粒度分类作为训练目标可以优化粗粒度分类的性能。然而在多病种的诊断问题中，分类的粒度往往不统一，某些疾病可以被细分，而某些疾病则很难再被细分，考虑到实际的临床诊断中，医生往往也不是直接得出病人所患的具体病种，所以其在诊断逻辑上，则是由粗粒度再到细粒度的过程。得益于医生的诊断思路，本文提出了一种针对于多病种的辅助诊断体系，并结合上一个章的研究成果，构建了粒度更具层次性的数据集，并提出了基于粒度迁移的分层诊断算法。其算法流程主要分为以下四个步骤：1.训练基于细粒度标签的子类皮肤疾病的基础分类模型；2.基于细粒度基础模型的特征迁移；3.训练基于粗粒度标签的大类皮肤疾病分类模型；4.策略融合。本章将在第 4 节做详细介绍。

4.1 数据集与数据增强

皮肤病多达 2000 余种，要想对全部病种进行分类显然对现在的技术来说是不现实的，一方面目前还不存在涵盖面如此广的数据集，另一方面，从皮肤性病学角度来说皮肤疾病的分类学仍比较混乱，存在各种分类标准共存的现象^[66]。综合考虑数据集 XiangyaDerm 的数据分布，参考《皮肤性病学第 8 版》的分类标准，以及结合本身疾病的视觉相似性，本章对 XiangyaDerm 优化了后处理过程，并从中选出了 19 类疾病组成 6 大类疾病的分类学来研究分层诊断问题，如图 4-1 所示。为了保证实验结果的稳定性，本章使用了 5 折分层交叉验证的方式。首先将数据按照病例分组，因为一个病人可能会存在几张照片，为了使一个病人的照片不会同时出现在训练集和测试集，按照随机的病例顺序将 19 类皮肤疾病数据分成 5 份。从整体数据集的角度而言，不管使用什么样的标签，训练集的数据分布总是等比例的，这样使得不同折之间的数据具有较高的稳定性，极大地排除了实验的偶然性。

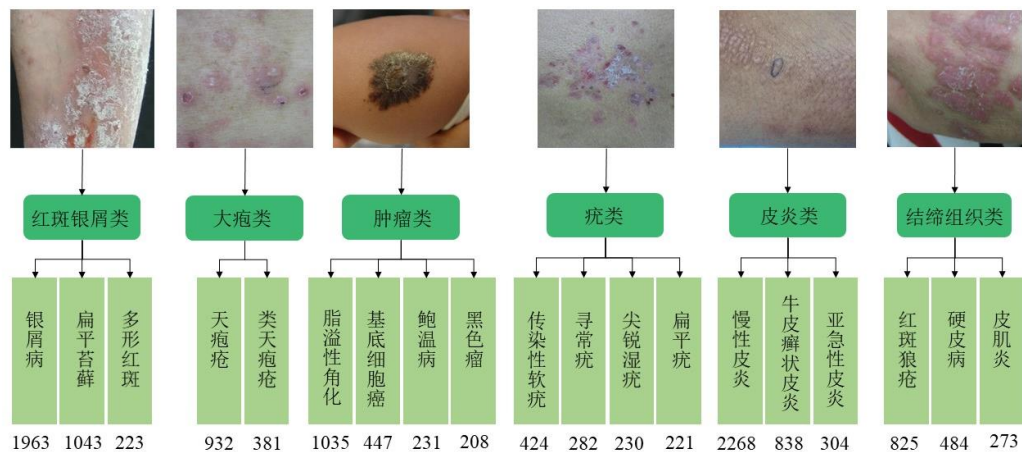


图 4-1 分层诊断算法的数据集分布

由于不同皮肤病的发病率是存在天然差异的,从数据集的分布上也可以看出,数据集总量较小,所以本章考虑采用有监督的数据增强方式弥补数据的不足。有监督的数据增强手段主要包含有基于单个数据和多个数据的方法,单个数据的增强主要有几何变换操作(随机剪切、缩放、旋转、拉伸、水平和垂直翻转等)、色彩变换操作(噪声模糊、颜色抖动和色彩空间等)^[67],以及 Cutout^[68]等方式。许多研究证明,这些数据增强操作,提高了皮肤癌的诊断^[69]。此外基于多个数据的增强方法主要有 SMOTE^[72]和 Mixup^[73]等,但由于皮肤疾病图像普遍相似度较高,这些数据增强方式的效果不佳。

本章使用的数据增强效果图如图 4-2 所示,包含了以下几种数据增强方式:

- 1) HSV (Hue, Saturation, Value) 颜色空间变换。其中 H 表示的为色调,也叫做色相,可以用度数来表示; S 表示为饱和度,即颜色靠近光谱色的程度,程度越高,其饱和度就越高; V 表示颜色的明度,即颜色的明亮程度。通过此种变换,提高了图片在色彩空间的多样性。
- 2) 几何变换。从图中可以看出,图像经过了一定的随机旋转、平移、缩放以及剪切操作。图像变换之后,空出的区域会用固定值进行填补。通过此种变换,在保留关键特征不变的条件,增强无关信息的多样性。
- 3) Cutout 操作。在原始的操作中,会通过固定大小的矩形块对图像进行一定程度的遮挡,然后将矩形区域的值填充为某一固定值。在这里,为了进一步加强多样性,本次实验将固定大小改为一定范围内的随机大小,同时颜色的填充值也变为一定范围的纯色值,此外为了防止皮损被完全遮挡,会减少图像中间位置的矩形数,以及大小。考虑到人眼是可以过滤掉这些简单干扰,通过此种变换,增强了网络表征全局特征的能力。

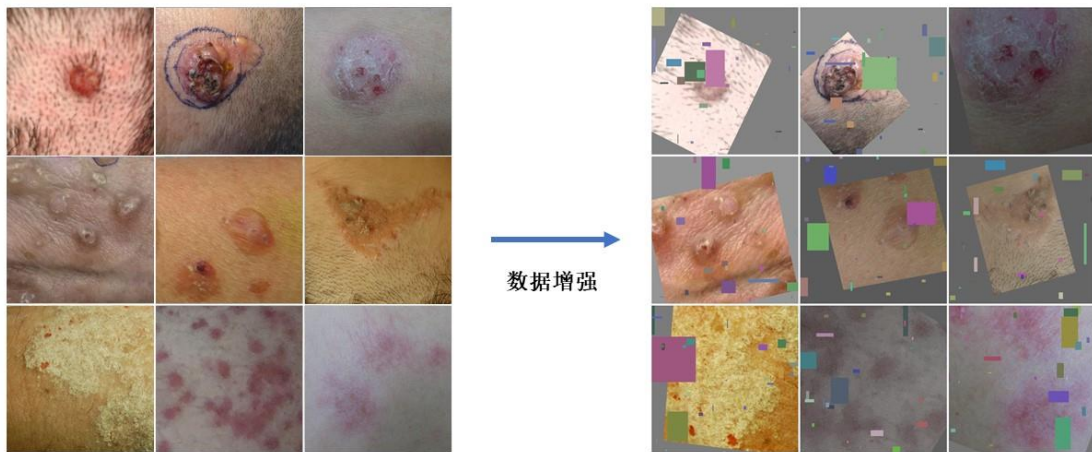


图 4-2 数据增强效果图

4.2 参数选择

4.2.1 超参数设置

在深度学习模型中超参数的选择十分重要，为了尽可能地获得更好的网络性能，除了使用之前提到的数据增强方法，还会添加一些学习率的调优策略。由于超参数的选择具有一定的不可解释性，所以在参数的选择上需要使用者有着丰富的经验，为了降低人工选择的成本以及更方便地获取到合理的参数，本节之后会基于进化算法，对超参数进行一定程度的选择。表 4-1 为实验中需要调优的超参数以及超参数的初始值。

表 4-1 超参数及经验初始值

参数符号及名称	初始值
lr0(初始学习率)	0.00465
lrf(最终学习率)	1e-5
momentum(动量)	0.951
weight_decay(权重衰减)	0.000355
hsv_h(色调变化量)	0.0138
hsv_s(饱和度变化量)	0.678
hsv_v(明度变化量)	0.36
rotation(旋转量)	10
translation(平移量)	0.15
scale(缩放量)	0.358
shear(裁剪量)	0.9
burn_in(初始学习策略阈值)	50

本次实验的初始值是基于经验获得的超参数，是通过以往的实验验证得到的具有一定优越性的超参数。在这里，使用的优化器为带动量的随机梯度下降算法（Stochastic Gradient Descent, SGD）^[74]，它与更早的 SGD 相比，可以保留一部分之前状态的梯度，具有更好的稳定性。公式可以用如下符号表示：

$$\begin{cases} v_{t+1} = \mu \times v_t + g_{t+1} \\ p_{t+1} = p_t - lr \times v_{t+1} \end{cases} \quad (4-1)$$

其中， p_t 表示 t 时刻的模型参数， g_t 表示 t 时刻的模型梯度， v_t 表示 t 时刻的速度， μ 表示动量。在模型训练的开始阶段，一般设置较大的学习率，便于模型更快地收敛，在之后的阶段，为了找到模型参数的最优解，往往会降低学习率。本次实验使用的方法为余弦退火策略，如图可以看到优化器的学习率变化曲线如图 4-3 所示：

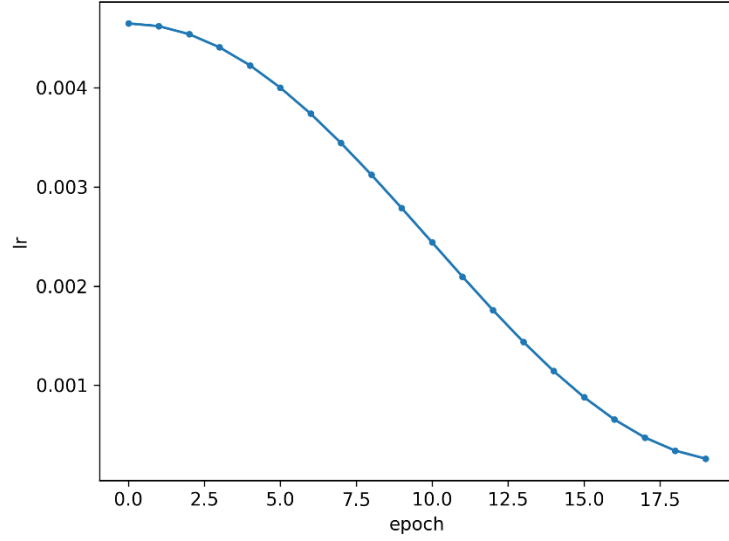


图 4-3 基于余弦退火的学习率更新策略

其中起始的学习率为 lr_0 ，最终的学习率即为参数 lrf 。为了防止模型过拟合，在模型的损失中加入了 L2 正则项。

$$Obj = loss + \frac{\lambda}{2} \times \sum_i^N w^2 \quad (4-2)$$

原来的参数更新公式为：

$$w_{t+1} = w_t - \frac{\partial Obj}{\partial w_t} = w_t - \frac{\partial loss}{\partial w_t} \quad (4-3)$$

加入正则项的公式为：

$$w_{t+1} = w_t - \frac{\partial Obj}{\partial w_t} = w_t - \left(\frac{\partial loss}{\partial w_t} + \lambda \times w_t \right) = w_t(1 - \lambda) - \frac{\partial loss}{\partial w_t} \quad (4-4)$$

通过对比公式 (4-3) 和公式 (4-4) 可以看出，L2 正则项会使参数在更新的过程中造成一定的衰减，所以在实际的操作中，只需要在模型参数的设置上加入了 `weight_decay` 参数。此外，为了防止出现网络刚开始训练时跳过全局最优解的情况，在模型训练中会加入 `burn_in` 参数，表示在 `burn_in` 次更新参数前，优化器会遵循学习率从小到大的变化过程，如公式 (4-5) 所示。

$$lr = lr_b \times \left(\frac{N_{batch}}{N_{burn_in}} \right)^n \quad (4-5)$$

通过实验经验，这种方式的学习率变化策略可以使模型达到更优秀的性能，但是具体的超参数值设置仍然需要选择。

4.2.2 基于进化算法的超参数选择

为了得到比随机设置甚至经验设置更好的超参数，本节使用进化算法来搜索超参数。值得一提的是，由于深度学习算法对设备的要求过高，在基于单张 Titan Xp 的显

卡条件下,使用 EfficientNet-B6^[75]模型得到一次完整训练的时间往往需要数小时,所以为了降低训练成本,从而在有限时间内得到较为满意的超参数,本次超参数选择使用基本结构一致的 EfficientNet-B3 作为深度学习模型。此外,数据上也做了相应的处理,从 5 折分层交叉验证中随机选取 2 折进行实验,取 2 折实验的平均值作为一次迭代的结果。本次的超参数搜索的具体步骤如下:

- 1) 初始化种群。种群数设置为 5,本次实验需要优化 12 个超参数,所以一个个体是由 12 维向量组成,超参数的初始值由经验给出,如上节中的表 4-1 所示。
- 2) 计算适应度。适应度即目标函数,是由分类算法的评价指标构成。在超参数搜索中使用的分类任务为本次皮肤病分类学的最后一层,即 19 类皮肤的分类任务。使用的评价指标包括:召回率 (Recall)、精确率(Precision)、F1 分数以及 AUC 值。通过对以上几个指标进行加权求和,可得到最终的综合分数 S ,即适应度,公式如 (4-6) 所示。

$$\begin{cases} Recall = \frac{TP}{TP+FN} \\ Precision = \frac{TP}{TP+FP} \\ F_1 = \frac{2(Precision \times Recall)}{Precision+Recall} \\ S = w_1 \times Recall + w_2 \times Precision + w_3 \times F_1 + w_4 \times AUC \end{cases} \quad (4-6)$$

- 3) 种群选择。为了确保较小的迭代次数下可以找到优化的参数,这里仅选择适应度最高的 5 个体作为下一步参与变异的种群。
- 4) 变异方式。为了得到新的种群个体,对种群中的 5 个体就均值后进行变异操作,具体的公式如 (4-7) 所示。其中变异的概率设置为 0.9,根据超参数的重要性,设置学习率、旋转量以及裁剪量的 g 设为 10, δ 设为 0.2, r 和 r_g 分别表示 0 到 1 的随机数。其中每个参数根据参数特性会将其限定在固定范围内。

$$\begin{cases} m_t = \min(3, \max(g \times r_g \times r \times \delta + 1, 0.3)) \\ sp_{t+1} = sp_t \times m_t \end{cases} \quad (4-7)$$

- 5) 停止条件。由于本次的超参数搜索涉及到的参数数量较多,且每次迭代的实验代价较高。在仅仅需要次优解的条件下(比手动设置参数高即可),算法迭代的停止条件为达到指定的迭代次数: 50。

超参数的搜索结果如图 4-4 所示(横坐标表示参数值,纵坐标表示适应度),图中每个橙色圆点表示每次变异的各参数分数,蓝色圆点则表示起始经验参数的分数,红色三角形则表示最终得分最高的超参数。由图可以看出,图中的一半橙色点都在起始点之上,且最优个体往往处于较为集中的区域,说明变异的效果是有的,但可以明显看出,最终的超参数得到的分数相较于一开始的经验参数增长的并不多。一方面,经

验参数是本质上就是经过一系列实验及知识验证过的较优的参数,提升的空间并不大;另一方面,由于所要搜索的参数目标过多,设置的迭代次数不够,导致算法很难找到全局最优解。然而,本次实验的目的是为了搜索相较于经验参数更好的参数,从结果来看,已经达到了实验目的。

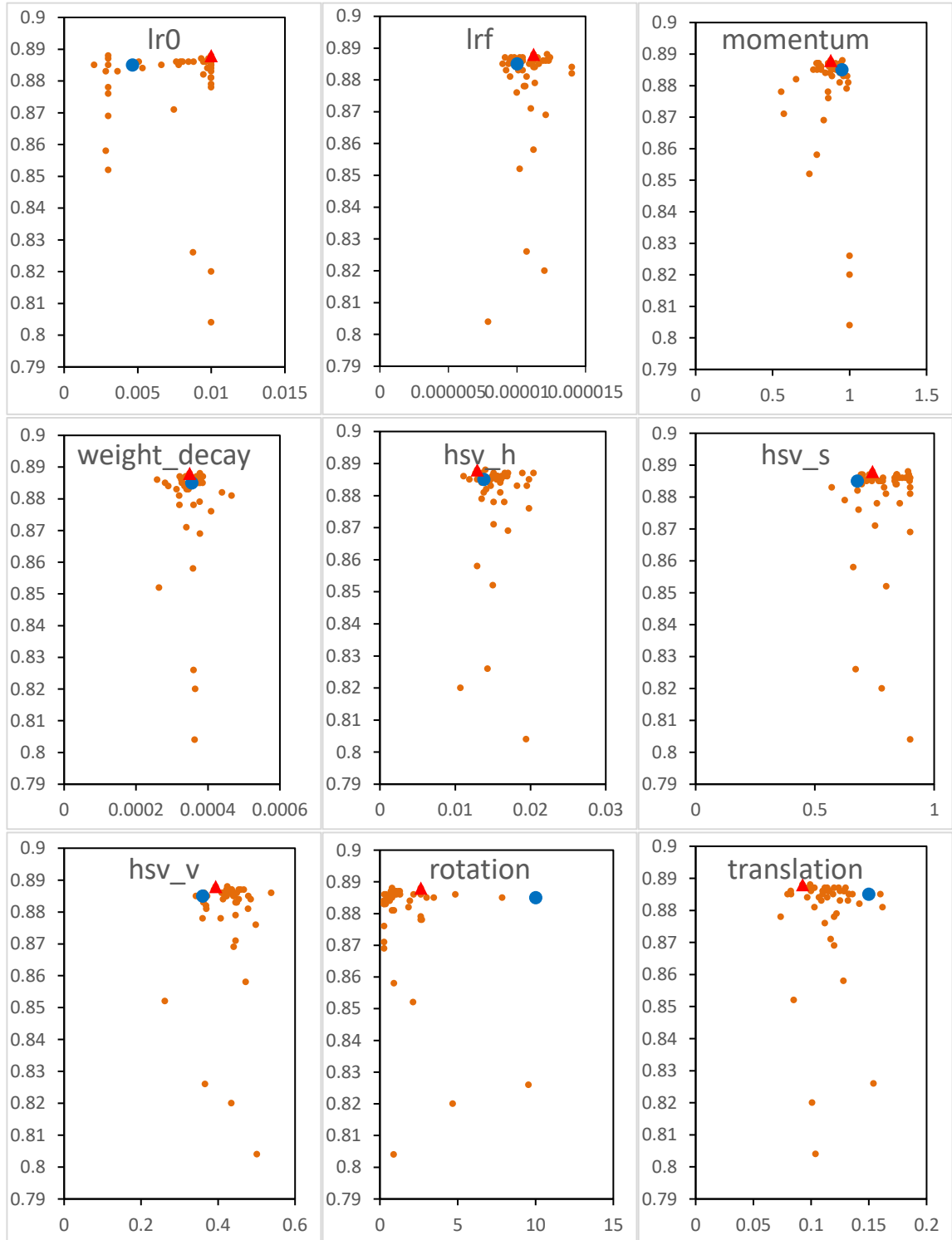


图 4-4 超参数的选择过程(横坐标表示参数值, 纵坐标表示适应度)

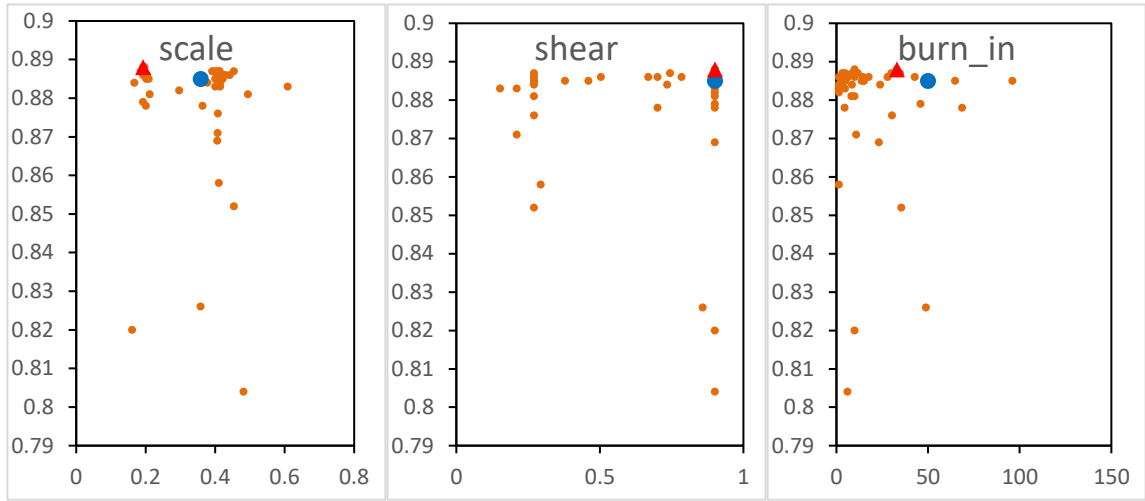


图 4-4 (续) 超参数的选择过程(横坐标表示参数值, 纵坐标表示适应度)

4.3 皮肤疾病的细粒度分类特征迁移

4.3.1 特征迁移原理

在面对数据集较小的问题时,除了利用一些数据增强方法,迁移学习的方法也是一种有效的手段。相较于图 4-5 中的方法 1: 随机初始化参数,目前常用的迁移学习方法,是方法 2: 基于 ImageNet 图像数据集的 1000 类物体的粗粒度分类预训练模型。通过去除最后的全连接层及 softmax 函数,将剩下的模型参数直接加载到相同的网络结构中,即可完成迁移过程,之后根据任务情况,可以选择全部参数继续训练,也可以选择冻结一部分参数微调训练。相关研究表明,迁移学习确实可以更快地达到更好的效果^[76]。根据前面章节的研究结果,相同数据内容下,更细粒度的分类可以获得更好的性能效果。所以为了优化整体的皮肤疾病分层诊断算法的性能,在训练第一层的相对粗粒度的分类任务时,本文利用了第二层分类任务的细粒度特征参数作迁移学习。考虑到 19 类皮肤疾病的分类任务使用的数据量相较于 ImageNet 还是属于小规模数据集,实际进行参数迁移时,如使用方法 4: 使用全部参数迁移,可能会起到反作用,所以下节将进行实验对比,对比在不同位置进行迁移的效果。

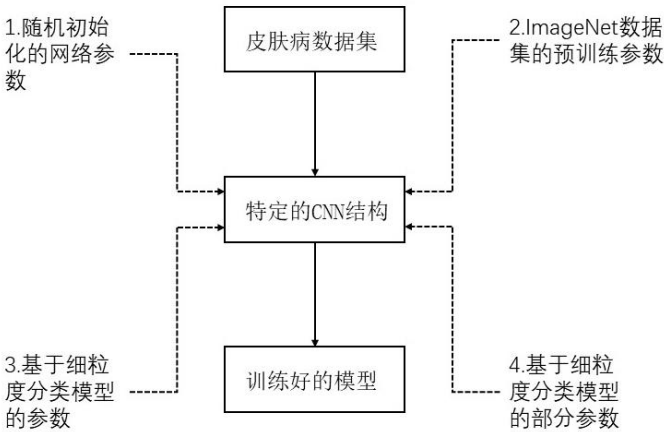


图 4-5 可以使用到的迁移学习方法

4.3.2 迁移参数的选择与对比

为了在有限的运算资源下，尽可能地提高本次整体分类性能的结果，本节实验使用的网络结构为 EfficientNet-B6。理论上在资源允许的情况下，卷积神经网络的规模（即网络深度、网络宽度以及网络输入图像的分辨率）越大，其性能表现就越强。但是通过人工的调整这些规模参数，往往具有很高的不确定性。所以，EfficientNet 正是通过建立相应的复合系数，来约束规模参数之间的关系，并通过神经结构搜索技术以最高准确率为目标来获取最优的参数。最终的基础网络结构 EfficientNet-B0 可如表 4-2 所示，表中的 MBConv 表示移动翻转瓶颈卷积（mobile inverted bottleneck convolution，MBConv）模块，是由深度可分离卷积以及 SENet^[78]构成。EfficientNet-B6 则是由更多相应的 MBConv 卷积块构成。

表 4-2 EfficientNet-B0 的网络结构

步骤	操作	分辨率	通道数	层数
1	Conv 3*3	224*224	32	1
2	MBConv1, k3*3	112*112	16	1
3	MBConv6, k3*3	112*112	24	2
4	MBConv6, k5*5	56*56	40	2
5	MBConv6, k3*3	28*28	80	3
6	MBConv6, k5*5	14*14	112	3
7	MBConv6, k5*5	14*14	192	4
8	MBConv6, k3*3	7*7	320	1
9	Conv1*1&Pooling&FC	7*7	1280	1

在整个网络中,如何选择参数迁移的位置是值得探讨的问题。为了选择合理的位置,本节将进行实验来选择合理的迁移位置。考虑到 EfficientNet-B6 总共有 45 个 MBConv 模块(block),本次实验选择了三个中间位置:从 10block 前进行参数迁移,(即 10block 前采用的是 19 类细粒度皮肤疾病分类模型的参数,10block 之后采用 ImageNet 的参数),从 23block 前进行参数迁移以及从 30block 前进行参数迁移。此外,还使用了完全使用 ImageNet 预训练参数以及完全使用 19 类细粒度皮肤疾病分类模型的参数作为对比。实验结果可以从图 4-6 可以看出,其中每个折线图表示五折分层交叉验证的验证数据集,每个颜色表示一种参数迁移方式,圆点表示准确率 ACC 的值,而三角形表示 AUC 值。为了更清晰看出最优的结果,不同指标下的最高分坐标由迁移方式对应的颜色给出,横坐标表示训练的轮数。

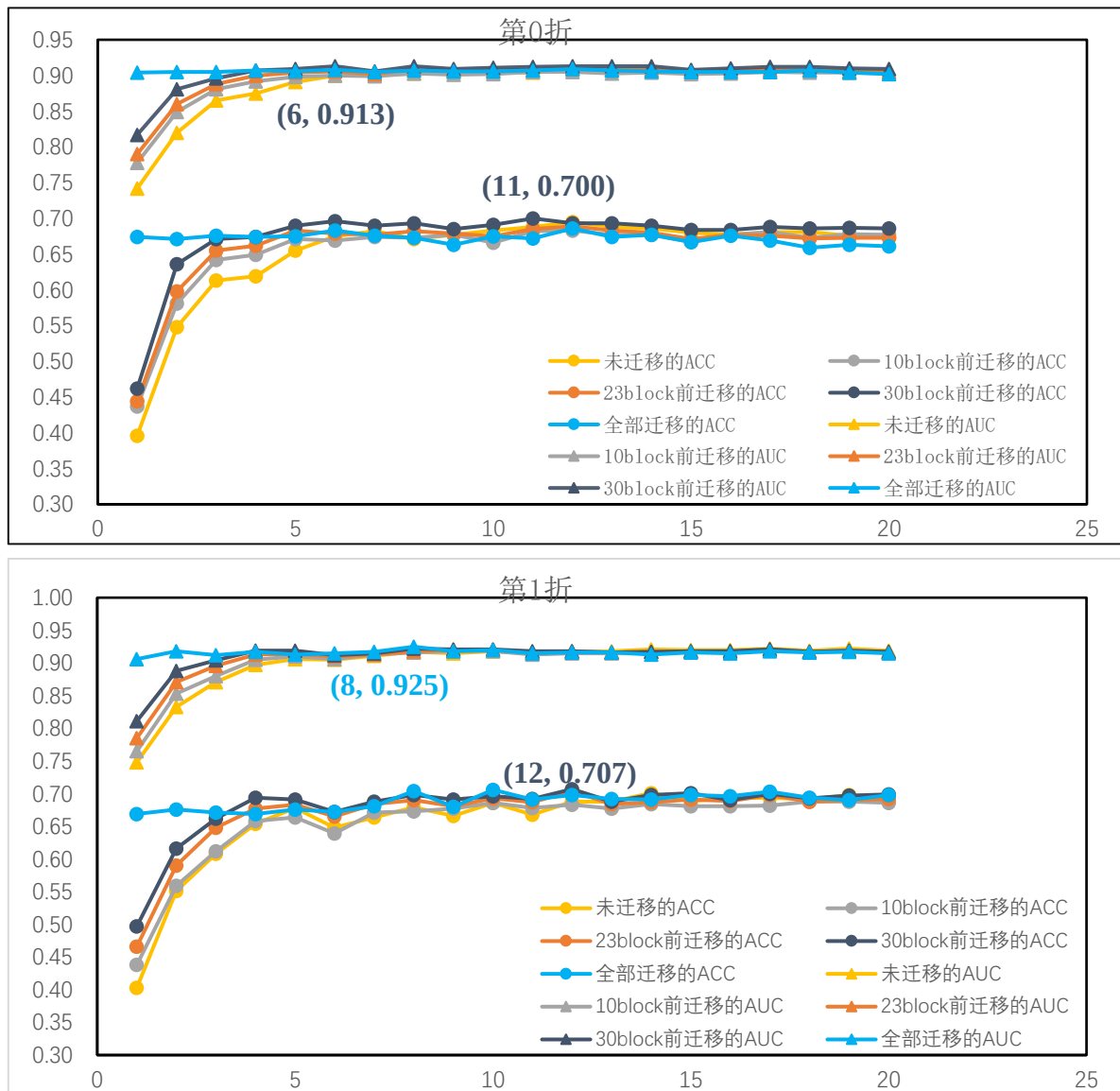


图 4-6 特征迁移在 6 分类任务上实验结果(横坐标表示训练轮数, 纵坐标表示 AUC 及 ACC 值)

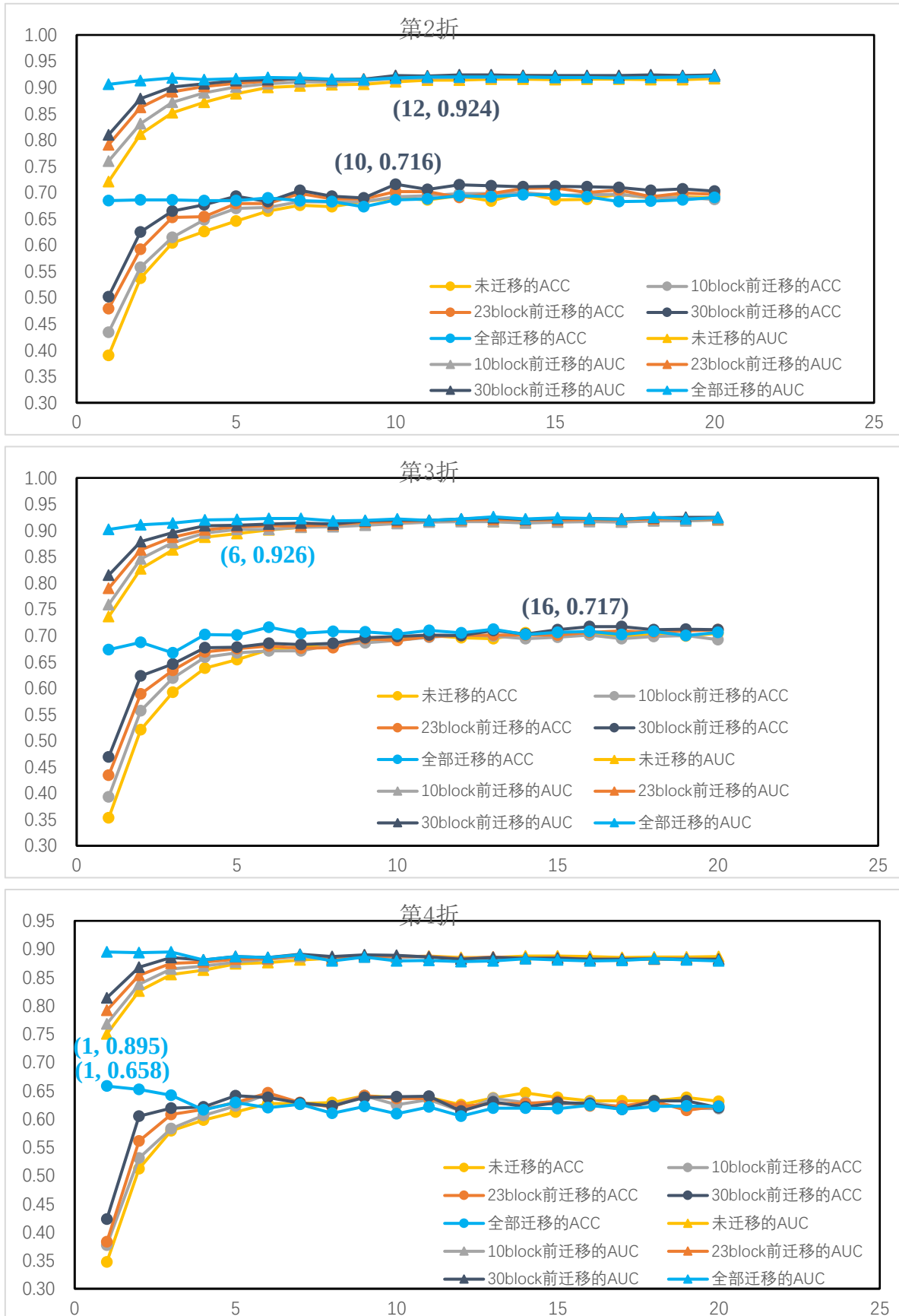


图 4-6 (续) 特征迁移在 6 分类任务上实验结果(横坐标表示训练轮数, 纵坐标表示 AUC 及 ACC 值)

从实验结果可以很明显地看出，使用细粒度分类模型的参数进行迁移，实验可以更快地达到更高的结果，而且随着迁移参数数量的增加，这种现象越明显。但是值得注意的是，使用全部参数进行迁移时，虽然可以在第一轮就能达到很高的结果，但是随着训练轮数的增加，其验证集的评价指标存在下降的趋势，且从总体最高结果而言，使用全部参数迁移大多不如选择部分参数进行迁移的结果好（除第 4 折可能存在一定的偶然性以外）。

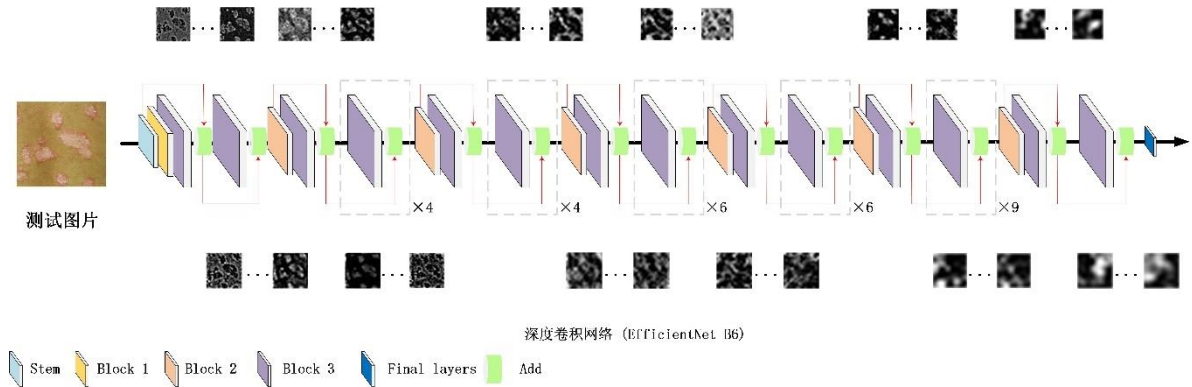


图 4-7 不同位置特征图的展示

考虑到在卷积神经网络中（如图 4-7 所示），随着网络层数加深，网络提取的特征往往更加抽象，更加地细微。而在训练本文所提到的皮肤病分类结构中的第一层时，这些类内的同一特征更加地宏观，所以更适合只迁移前面一部分的参数，如果采用全部参数可能会产生一些反作用。值得一提的是，具体的最优位置还需要大量的实验来进一步确定，本次实验仅仅是为了选择较好的迁移方式，相对于不迁移而言，有着较大的优势。

4.4 结合粗粒度特征的皮肤疾病分层诊断

4.4.1 分层诊断算法

医生在进行临床诊断疾病时，会先对疾病有个粗略的诊断结构，即可以理解为粗粒度的分类结果，然后会根据粗粒度的结果进一步细化，最终确定为某一更细精确的结果。受到此思路的启发，本章设计出了基于粒度迁移的分层诊断算法，其步骤如图 4-8 所示。

1. 训练基于细粒度标签的子类皮肤疾病的基础分类模型。由于分类的最终目标是对更细粒度的第二层疾病进行分类，所以第一步必须要先获得基准模型，之后会对基准模型的预测结果进行决策修改。
2. 基于细粒度基础模型的特征迁移。根据上面的研究，可以说明细粒度标签的较强的监督信息，有利于模型的训练。所以这里使用步骤 1 的模型参数，用

作之后粗粒度模型的初始化参数，根据上一节的实验，这里选择 30block 的位置作为迁移的临界点，30block 之前为步骤 1 的模型参数，30block 以及之后为 ImageNet 的模型参数。

3. 训练基于粗粒度标签的大类皮肤疾病分类模型。在加载完初始参数后，使用同步骤 1 相同的数据，但是标签为步骤 1 中 19 类皮肤疾病合并后的 6 大类皮肤疾病。
4. 策略融合。在最终的输出阶段，采用决策融合方式，将粗粒度模型的结果融入到细粒度模型中，从而实现最终的分层诊断。

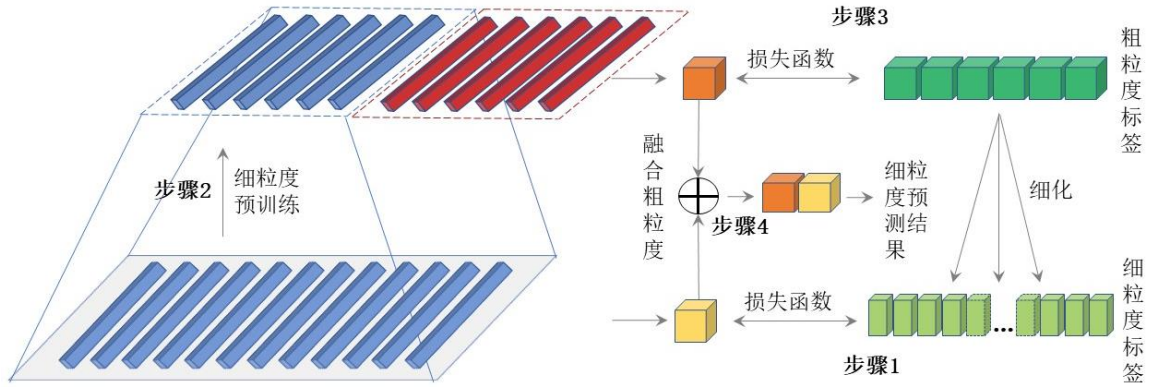


图 4-8 基于粒度迁移的分层诊断算法研究

最终的决策过程可以建模为两种方式。方法一：假设第一层的 6 类疾病的预测结果可以表示为 $p^I = (x_1^I, x_2^I, x_3^I, x_4^I, x_5^I, x_6^I)$ ，第二层的 19 类疾病的预测结果可以表示为 $p^II = (x_1^{II}, x_2^{II}, x_3^{II}, \dots, x_{19}^{II})$ ，并且用符号 $[x_i]$ 表示对 x_i 进行公式 (4-8) 的操作。

$$x_i = \begin{cases} 1, & \text{if } x_i = \max(x) \\ 0, & \text{if } x_i \neq \max(x) \end{cases} \quad (4-8)$$

则前面所述的预测过程得到的结果 p' 可以表示为公式 (4-9) 以下形式：

$$p' = ([x_1^I] \times x_1^{II}, [x_1^I] \times x_2^{II}, [x_1^I] \times x_3^{II}, [x_2^I] \times x_1^{II}, \dots, [x_6^I] \times x_{19}^{II}) \quad (4-9)$$

这种分层诊断的思路本质上是对第二层皮肤疾病的预测进行修正，即最终的结果参考了第一层预测的结果，且第一层的诊断结果起到了绝对引导性的作用：非最大分数值的子类疾病预测值全部置为 0。这种方式直观上比较极端，如果选择的大类疾病被错误判断，会直接导致最终的正确子类疾病的分数为 0。所以本章在此基础上，为了提高容错率，提出了方法二：直接将第一层的预测结果与第二层对应的子类预测结果相乘，得到 p'' 。其过程可用公式 (4-10) 表示。

$$p'' = (x_1^I \times x_1^{II}, x_1^I \times x_2^{II}, x_1^I \times x_3^{II}, x_2^I \times x_1^{II}, \dots, x_6^I \times x_{19}^{II}) \quad (4-10)$$

从逻辑上而言，第二种更为简单，且避免了极端的情况。

4.4.1 算法评估

为了验证分层诊断算法的真实性能，本章对未采用分层诊断算法，采用第一种方法以及第二种方法进行了实验对比。实验采用的是 5 折分层交叉验证的方式，每一折的测试集数量约为 2523 张，通过统计总计 12612 张测试数据的 F1 值以及 ACC 值，将其整理在了表 4-3 中。

表 4-3 分层诊断算法对 19 类疾病的整体实验结果对比

方法	F1	ACC
未分层	0.579	0.588
分层方法一	0.577	0.583
分层方法二	0.591	0.600

根据表格中的结果可以看出，分层方法二的结果最高，明显起到了效果，但分层方法一相当于在第一层的预测结果上加一个强限定条件，反而起到了反作用，略微降低了性能。由于本次的结果是在万张数据以上得出的结果，具有较高的稳定性，所以也验证了本次分层诊断算法的有效性，但由于数据集本身的难度较高，且完全按照病例分训练集与测试集，所以效果提升不大。

本文对分层方法二的结果进行了混淆矩阵分析，见图 4-9（图中标签的顺序为图 4-1 中的疾病顺序，横轴为预测标签，纵轴为真实标签），可以看出绝大多数的疾病都能较好的区分。其中银屑病（标签 0）本身由于具备明显特征且数据量较高，所以分类的效果最好。而值得一提的是慢性皮炎（标签 13）的数据量与银屑病处于同一量级，却具有很低的精确度，即大部分疾病都被预测成了慢性皮炎。这是由于慢性皮炎这一类疾病本身还可以继续划分，它是较多慢性皮肤炎症的统称，还包括了比如：特应性皮炎，脂溢性皮炎，日光性皮炎，神经性皮炎等多种疾病，其类内差异依旧很大。在数据量够多的情况下，“慢性皮炎”这一类保证了一定的召回率，但由于其本身的具备较大类内差异，导致网络只能拟合出它的粗粒度特征，而这一粗粒度特征与很多其他疾病都具有一定的相似性，比如与某些红斑银屑病、疣类以及其他皮炎类疾病都存在红肿、丘疹等特征，而肿瘤类（标签 5-8）的大多为黑色，所以不会或极少被预测为“慢性皮炎”类。这样的结果也侧面反映了直接对所有疾病进行直接分类的弊端，因为很多疾病在分类粒度不统一的前提下，极易出现结果偏向于粗粒度类别的情况。所以更加突出了采用分层诊断，再将疾病分类结构细化到多层，细化到同一粒度条件下的必要性。

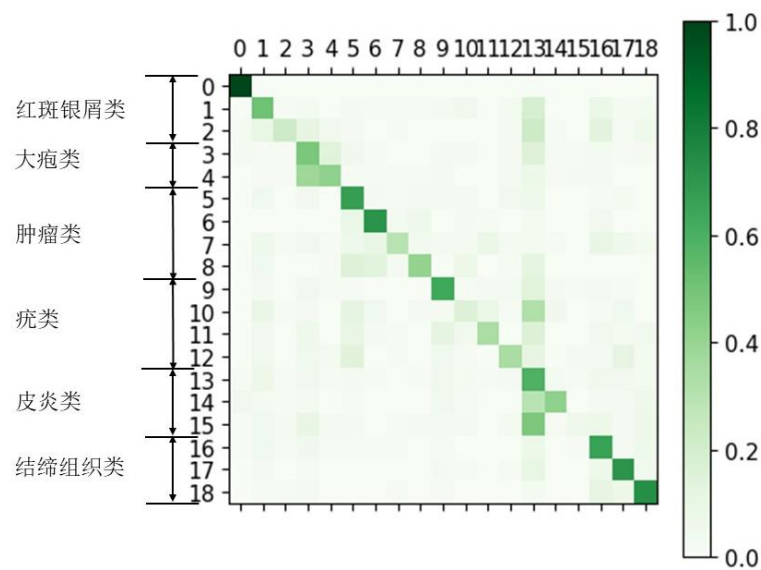


图 4-9 分层方法二的混淆矩阵

4.5 本章小结

本章针对于多病种的分类问题，提出了基于粒度迁移的分层诊断算法。首先，本章基于 XiangyaDerm 数据集，选出了 19 类具有代表性的皮肤疾病，并依旧《皮肤性病学》以及视觉相似性，将 19 类疾病分成 6 大类疾病从而构建成了两层的结构性数据集。在参数选择的过程中，本章使用了基于进化算法的超参数选择方式，并基于此参数进行了后面的实验。本章提出的分层诊断算法，首先根据第 3 章的研究结果，利用了细粒度的特征进行了部分参数迁移从而学习粗粒度特征，并将基于粗粒度特征的预测结果融合到了最后的细粒度分类预测中完成了整体性能的提升。

第5章 皮肤肿瘤恶性程度分级的应用研究

根据前面的研究可以发现,针对于较多类别的皮肤疾病临床图像的分类任务,很难达到足以实际应用的准确率及泛化性能。从应用角度考虑,可以把实际的辅助诊断问题,转化为病人的健康管理问题,通过算法提供给病人高准确性的范围性结果,从而给予病人一些参考信息以及指导性的意见。比如,黑色素瘤是致死率最高的一种皮肤肿瘤,然而和它具有高相似性的色素痣却只属于良性肿瘤,对于普通人而言,要想从临床图像区分出皮肤肿瘤的恶性程度是很难的,但这从健康管理学来说是十分重要的。本章节以皮肤肿瘤为例,通过深度学习方法提供给病人所患皮肤肿瘤的恶性程度信息,从而帮助患者有选择性地就诊。

5.1 面向病人的应用设计

5.1.1 应用的功能

对于面向病人的健康管理系统而言,皮肤肿瘤是最为值得关注的一类疾病,其本身具有很高的发病率,多数为良性肿瘤,但小部分皮肤肿瘤可能转化为恶性皮肤癌(如黑色素瘤),这对患者来说是致命的。由于皮肤癌的严重影响和高发病率,评估皮肤肿瘤的恶性程度便成为了皮肤病健康管理中的一项重要任务。在临床上,皮肤肿瘤的恶性程度的“金标准”需要活检和病理检查,这对患者来说既费时又费钱。而且国内的皮肤科医生严重短缺,皮肤科医生和病人的比例严重失衡。在少数的农村地区,更是几乎没有皮肤科医生,所以病人在去医院很不方便的环境下,可能造成忽略自身的皮肤健康状况,导致皮肤肿瘤进一步恶化。

在这种情况下,如果有一种的更加智能方便的初步筛查方法,可以识别出有风险的患者,并提醒他们去医院做进一步检查。例如,如果皮损图像被准确识别为良性,患者无需担心,只要不断监测病情即可;如果皮损图像被准确识别为低恶性,则建议患者到附近的医院进行进一步诊断;如果为高度恶性,则建议患者紧急前往三甲医院做精确诊断。在这种情况下,可以避免恶性黑色素瘤患者延误治疗,也可以让普通色素痣患者不会整天诚惶诚恐。

系统的功能结构可由图 5-1 表示,主要包含以下步骤:

- 1) 病人通过智能手机上传自己的患病处图像。由于其他种类的皮肤病图像获取条件苛刻,临床图像的最大优势即为获取的便利性。且目前智能手机的发展迅速,其拍照性能甚至已经达到专业相机的级别,为后面的图像分析提供有利前提条件。
- 2) 上传图像至服务器端。服务器会部署 WEB 服务以及基于深度学习的算法模

- 块，用户上传图片前，服务器会处于任务等待的状态。当有新的请求出现时，服务器会将收到的图像输入到深度学习算法中。
- 3) 基于深度学习的皮肤肿瘤恶性程度分析。上传的临床图像经过训练好的模型，会给出相应的恶性程度及分数。
 - 4) 生成反馈信息。根据得到的恶性程度预测结果，生成对应的医学建议并以文档形式返回给病人。当低风险时建议病人注重自我防预并推荐一些皮肤管理的资料，可以暂时不考虑去医院就诊，当高风险时建议就近的医院就诊，当结果处于危险级别时，则建议病人尽量前往三甲级别的医院就诊。

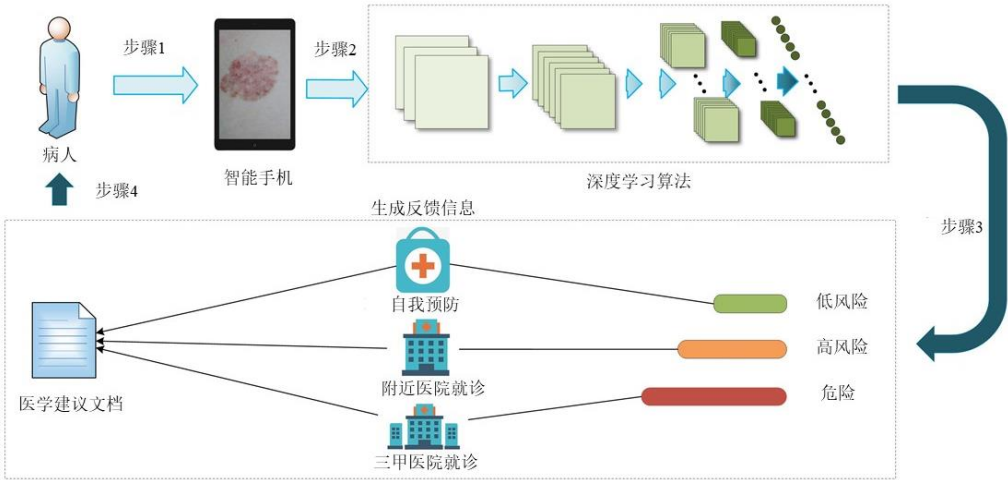


图 5-1 应用的基本功能

5.1.2 应用的设计与实现

本应用主要由三部分内容组合，客户端主要用于病人拍照上传自己的皮损区域图像，服务器端主要用于处理客户端的请求，并调用部署在上面的算法，并根据算法端返回结果，生成相应的报告，而算法端主要用于分析服务器接收的图像，并根据深度学习模型，分析出该图像的恶性程度等级。其中，客户端与服务器端采用 HTTP 协议进行通信，下面将分以下几点介绍。

(1) 客户端

客户端主要基于 Html5 技术开发的，其与原生 APP 具有相当的功能，涉及到的语言为 HTML、CSS 以及 JavaScript。客户端的基本工作流程图如图 5-2 所示。病人可以选择两种方式上传疑似皮肤肿瘤的图像，一种为直接打开摄像头即时拍照的方式，一种采用本地选取图像的方式。通过这样的设计，不仅可以上传自己的图像，也可以帮助他人进行上传。由于临床图像难免具有一定的额外背景或者其他光照原因，所以

会在上传之前可以进行图像的后处理，包括图像裁剪，以及颜色空间调整等。最后客户端将会一直等待服务器端的结果，并预设固定的超时时间，当超过一定的时长未得到本次结果，将会中断本次的皮肤肿瘤恶性程度检测过程。

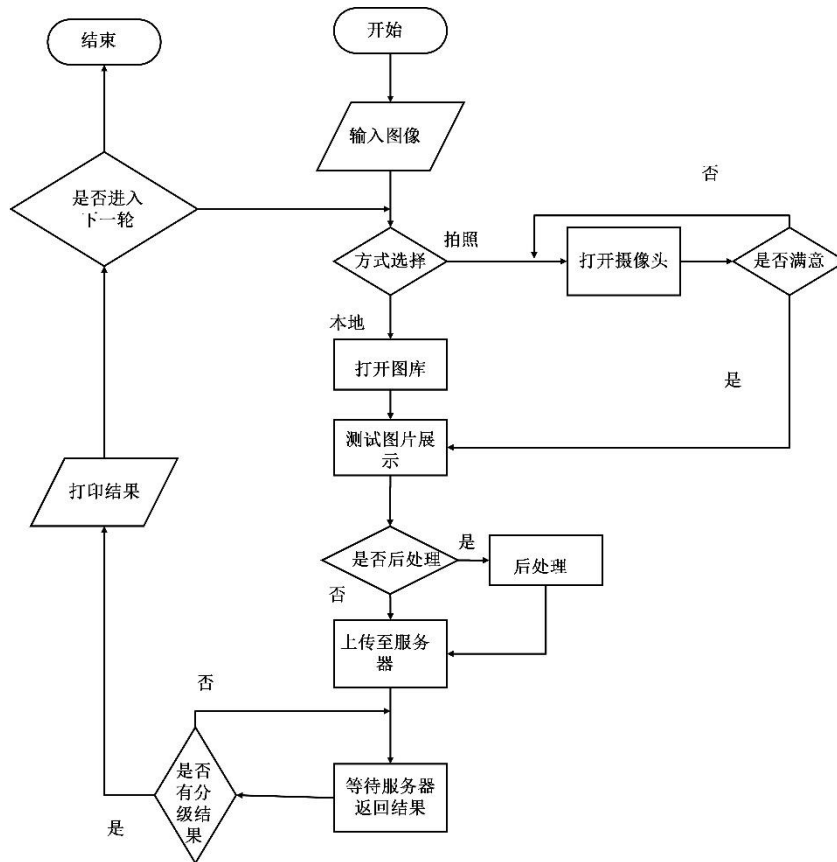


图 5-2 客户端工作流程

(2) 服务器端

服务器端包括 WEB 服务平台、文件存储器以及数据库，分别基于 Django、NAS 以及 MySQL 实现。服务器的主要工作流程如图 5-3 所示。在接收到客户端的辅助诊断请求之前，服务器会一直处于等待请求中，当有新的图像请求传来，服务器端会将图像和其地址信息分别存储进 NAS 网络存储以及 MySQL 中。之后，服务器便会调用部署在上面的基于深度学习方法的皮肤肿瘤分级代码，在得到分级结果或者超时之前，服务器都会处于等待结果中。如果因为超时原因，未得到有效结果，会将超时的请求信息的记录，并返回给客户端超时提示。若正常得到分级的结果，服务器会生成对应的医学建议以及相关的医院信息。

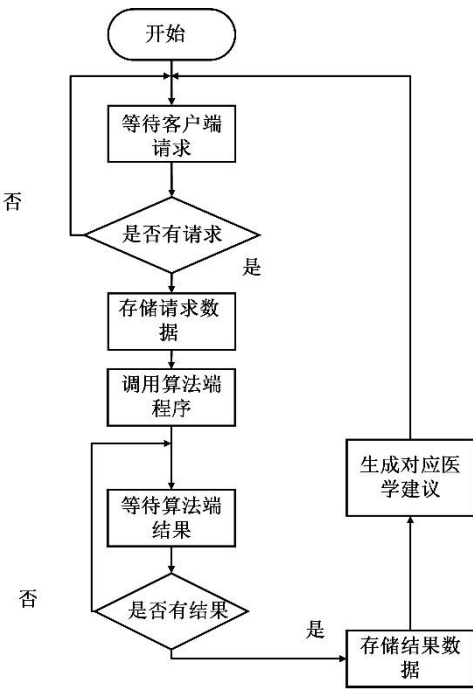


图 5-3 服务器端工作流程

(3) 算法端

算法端部署了预先训练好的深度学习模型，其中用于训练的数据是关键。为了能够尽可能地从图像中提取到表征恶性程度的特征，本次实验的数据是从 XiangDerm 中选取的属于皮肤肿瘤大类的 10 种皮肤病，如图 5-4 所示。图中的绿色框内的皮肤病（交界痣、皮内痣、皮肤纤维瘤、脂肪瘤、脂溢性角化）表示为良性的疾病，黄色框内的皮肤病（鲍温病、基底细胞癌、光化性角化病）表示的为低恶性程度的疾病，红色框内的皮肤病（鳞状细胞癌、恶性黑色素瘤）表示的为危险程度的疾病，总的数据量为 4500 张图像。

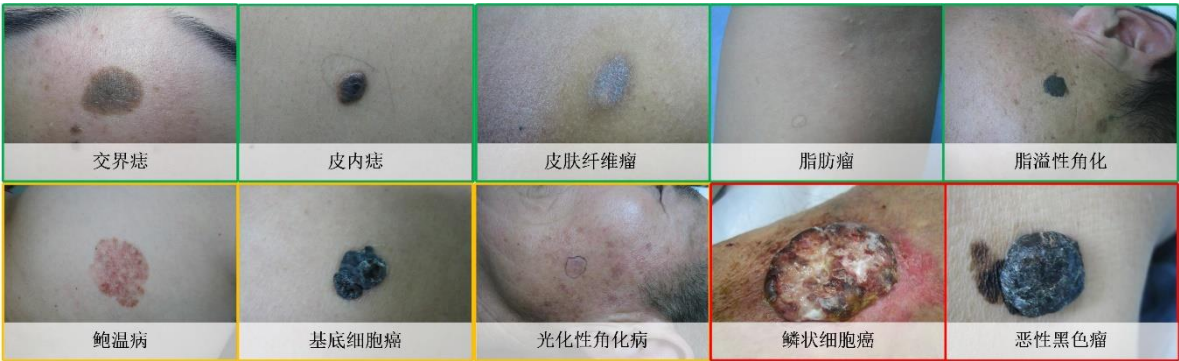


图 5-4 不同恶性程度的皮肤肿瘤

由于在皮肤病学上，没有具体的恶性程度打分标准，即从数据标签的角度，无法给出每张图像的真实分数。目前的不同种皮肤病的恶性程度是根据皮肤病书籍以及医

生们的建议得到的具有一定专业性的等级划分。所以恶性程度分级任务本质上可以看成是一个分类任务，其损失函数依旧可以是交叉熵损失函数。为了给出参考性分值，这里给每个疾病类别一个预设分值 s ，低恶性程度的疾病类别的分值为 1，高恶性程度的疾病类别的分值为 2，低恶性程度的分值为 3。网络的输出会经过 softmax 函数，如公式（5-1）所示，得到每个类别的预测分值 P_i ，以 P_i 为权重得到 s_i 的累计值，即为最终的恶性程度分数 S 。

$$\begin{cases} s = [1,1,1,1,1,2,2,2,3,3] \\ P_i = \frac{e^{V_i}}{\sum_i^C e^{V_i}} \\ S = \sum_i^C P_i s_i \end{cases} \quad (5-1)$$

5.2 应用性评价

本次实验以 Xception 网络为例，搭建本应用的核心分级算法用于评价其应用价值。本节将通过两个测试集来评估系统的性能。首先，对 1125 张临床图像进行了分类测试，计算阳性样本中正确预测比例即敏感性，阴性样本中正确预测比例即特异性。并通过调整阈值，得到一系列由灵敏性和特异性组成的集合，即可得到受试者工作特性曲线。实验结果如图 5-3 中 test1 所示，良性、低恶性程度和高恶性程度的下面积 AUC 分别为 0.959、0.919 和 0.947。各恶性程度最优约登指数下的指标如下：良性病变的敏感性（与漏诊率相加为 1）和特异性（与误诊率相加为 1）分别为 0.93、0.88，低恶性程度分别为 0.85、0.85，高恶性程度分别为 0.86、0.91，从结果上来看，系统误诊和漏诊的概率都很小，分级的整体准确率也可以达到 0.827，具有良好的性能。

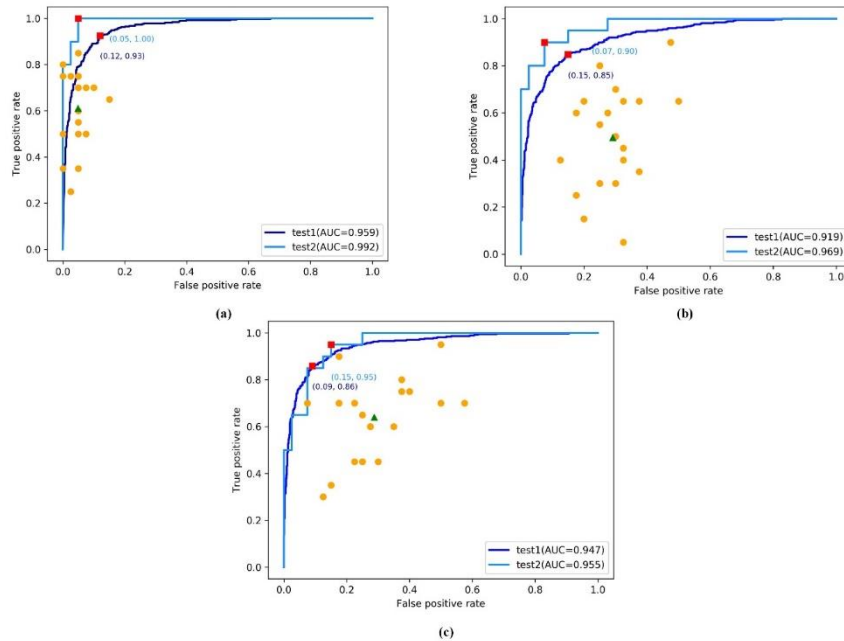


图 5-5 验证集的 ROC 曲线（test1）以及“人机大战”结果（test2）

为了进一步体现其应用性价值, 本节从湘雅医院中采集包含三种恶性等级的 60 张新的测试图像, 每个等级包含 20 张图片, 与 20 位皮肤科医生进行恶性程度分级的“人机大战”。实验结果展示在图 5-5 中 (test2 表示与医生同一数据下的算法结果), 橙色圆点表示 20 名皮肤科医生的假阳性率和真阳性率分布, 平均值用绿色的三角形表示, 三个等级的均值分别为 (0.049,0.61), (0.29,0.495) 和 (0.29,0.64), 低于本应用的评分结果。并且从结果上可以看出, 医生的分级结果大多低于曲线, 这说明本应用在恶性程度分级任务上要优于绝大多数皮肤科医生。此类的健康管理系统相较于医生团队, 有着明显的优势。因为目前在皮肤病学上, 没有具体的恶性程度分级细则, 医生在判断此类任务上具有较高的主观性 (从图中橙色点的分散程度就可以看出), 但对病人所患皮肤病的恶性程度进行预测, 是临床初筛中十分有价值的一环, 利用深度学习算法可以完全避免主观性。只要制定的准则具有相应的权威性, 算法就能很好学到这方面的准则。此外, 利用该算法开发的应用, 可以有效地突破地域的限制, 在医疗资源不平衡的背景下, 引导病人就诊, 并起到皮肤病健康管理的作用。

5.3 本章小结

本章从应用角度, 介绍了以皮肤肿瘤为例的面向病人健康管理的应用设计思路及算法实现。针对于目前的分类算法难以达到实际的应用水准, 而皮肤肿瘤的恶性程度分级的应用没有尚未被研究的问题。本文采用了卷积神经网络, 提取皮肤肿瘤临床图像的抽象特征, 利用分类任务作为算法基础进而进行分数统计的思路, 得到了每张图像的分级结果及分数。为了更好地获取恶性程度的表征, 本文遵循国内三甲医院医生团队的标准, 建立了恶性程度分级的数据集, 为以后相关的研究打下了基础。最后的实验结果显示, 本应用可以很好地对单张皮肤肿瘤图像进行恶性程度的分级, 并且在“人机大战”中赢得了医生。

第 6 章 总结与展望

6.1 论文总结

皮肤疾病发病率高，种类多且特征复杂，在医疗资源分布不平衡的条件下，研究皮肤疾病的辅助诊断分类算法以及应用便显得尤为重要。在数据集的限制下，目前相关研究多为皮肤镜图像，而具有更高应用价值的临床图像却研究较少。此外，现有的研究多为单一病种，缺少相关的应用研究。在此背景下，本文做了以下工作：

(1) 建立了以国内人群为主的临床图像数据集。由于现有的公开数据集多为欧美人种的皮肤镜图像数据集。皮肤镜图像由于获取条件苛刻不利于之后的应用开发。而欧美人种的皮损特征与国内存在一定的差异，难以应用于亚洲人种。本文构建了覆盖了 541 类皮肤疾病，每张图像都有病理检验上的结果作为依托，后处理后的总数量达到了 10 万量级的临床皮肤疾病图像数据集。为了评估数据集的质量，本文还做了不同网络的基准实验，并通过与国外数据集对比，反映了人种差异的问题。据了解，本文的数据集是目前最大规模、种类最多的国内人群数据集。

(2) 提出了一种基于精细粒度的两阶段分类方法。本文首先在理论层面分析了精细的粒度信息对网络训练的影响。考虑到本文提出的数据集大部分数据为银屑病数据且银屑病本身具有很高的研究价值，所以便以银屑病为例，建立了包含四种银屑病近亲以及四种银屑病远亲的 9 类皮肤病子数据集。针对于银屑病与非银屑病的两分类任务，本文利用 9 类疾病的细粒度标签完成训练，在预测过程中，会先预测 9 类疾病，再通过输出决策，从而实现银屑病的高性能的两分类模型。经实验及“人机大战”验证，本文提出的分类方法要明显优于单阶段的结果，且高于医生的平均水平，具有较好的泛化性能。

(3) 提出了基于粒度迁移的分层诊断算法。在面对多病种问题时，不同皮肤疾病的类间差异不一致，训练难度大，考虑到医生的诊断过程也是从粗粒度到细粒度的过程，本文建立了包含 19 类子类疾病组成的 6 大类疾病的分层皮肤病分类结构。在前面的研究基础上，本采用了细粒度特征迁移的方式将 19 类子类疾病训练的模型参数部分迁移至大类疾病的训练网络中，提升了大类疾病的训练速度以及训练效果。在最终预测过程中，本文会结合粗粒度的大类疾病预测结果，对子类细粒度预测进行集成。通过实验验证，这样的分层诊断方式在 12612 张数据的严格按照病例拆分的 5 折分层交叉验证中，有了一定的效果提升，并说明了在数据分类结构更合理的情况下，结果会更好。

(4) 提出了一种皮肤肿瘤恶性程度的评估方式及应用。目前的分类算法由于敏感性及特征性不够，涉及到的皮肤病种类也较少，很难实际应用于临床环境。考虑到皮

肤肿瘤的高关注度，以及潜在风险，本文从皮肤肿瘤的恶性程度分级角度，提出了一种面向病人的健康管理应用设计。由于皮肤病学上目前没有具体的恶性程度量化指标，本文使用了代表三种恶性程度的 10 类皮肤病，利用分类任务的方式，完成了模型的训练，并通过建立输出决策的方式，给出了每张图像的参考分值及具体的恶性程度级别。经最后的实验验证，本文提出的应用设计，在算法性能可以达到实际应用水平，并可以避免主观性同时突破地域的限制。

6.2 研究展望

本文对临床皮肤疾病图像的辅助诊断算法及应用展开的了深入的研究，并取得了一定的进展，但是仍然存在一定的不足需要改进：

(1) 目前建立的数据集存在较为严重的数据量不平衡问题。数据集的总体病种量虽然多，但大部分为银屑病数据，很多病种由于天然的发病率低，数据量较少，需要进一步研究皮肤疾病的小样本算法才能有效使用，或者还需要进一步收集更多的数据。

(2) 本文提出了皮肤疾病的分层诊断结构还可以进一步完善。从第 4 章的实验结果可以看出，“慢性皮炎”类造成了较大的干扰，如果可以对“慢性皮炎”进行进一步细化标签，则可以使最终的分层诊断效果更好，但这需要专业医生进一步标注。

(3) 从整体的临床图像辅助诊断而言。单纯依靠图像对疾病进行区分存在一定的局限性，在实际的诊断流程中，专业的医生往往还会借助一些额外的非视觉信息，询问病人比如是否存在瘙痒等。这就需要建立一套多源异构的皮肤病数据集，并建立相应的多模态模型。

参考文献

- [1] Hay R J , Johns N E , Williams H C , et al. The Global Burden of Skin Disease in 2010: An Analysis of the Prevalence and Impact of Skin Conditions[J]. Journal of Investigative Dermatology, 2014, 134(6):1527-1534.
- [2] Wernli, Karen, J, et al. Screening for Skin Cancer in Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force.[J]. JAMA: Journal of the American Medical Association, 2016, 316(4):436-447.
- [3] Bozkurt A, Kose K, Alessi-Fox C, et al. A multiresolution convolutional neural network with partial label training for annotating reflectance confocal microscopy images of skin[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018: 292-299.
- [4] Bozkurt A, Gale T, Kose K, et al. Delineation of skin strata in reflectance confocal microscopy images with recurrent convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 25-33.
- [5] Fujimoto J G. Optical coherence tomography for ultrahigh resolution in vivo imaging[J]. Nature biotechnology, 2003, 21(11): 1361-1367.
- [6] Pellacani G, Seidenari S. Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized-light videomicroscopy[J]. Clinics in dermatology, 2002, 20(3): 222-227.
- [7] Alfred, W, Kopf, et al. The use of dermoscopy and digital imaging in the diagnosis of cutaneous malignant melanoma[J]. Skin Research & Technology, 1997.1-7.
- [8] Menzies S W. Automated epiluminescence microscopy: human vs machine in the diagnosis of melanoma[J]. Archives of dermatology, 1999, 135(12): 1538-1540.
- [9] Binder M, Schwarz M, Winkler A, et al. Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists[J]. Archives of dermatology, 1995, 131(3): 286-291.
- [10] Rubin's pathology: clinicopathologic foundations of medicine[M]. Lippincott Williams & Wilkins, 2008.
- [11] Gurcan M N, Boucheron L E, Can A, et al. Histopathological image analysis: A review[J]. IEEE reviews in biomedical engineering, 2009, 2: 147-171.
- [12] Day G R, Barbour R H. Automated melanoma diagnosis: where are we at?[J]. Skin Research and Technology, 2000, 6(1): 1-5.
- [13] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International

- journal of computer vision, 2004, 60(2): 91-110.
- [14]Surhone L M, Tennoe M T, Henssonow S F, et al. Histogram of Oriented Gradients[J]. Betascript Publishing, 2016, 12(4):1368 -1371.
- [15]Swain M J, Ballard D H. Color indexing[J]. International journal of computer vision, 1991, 7(1): 11-32.
- [16]Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [17]Altman N S. An introduction to kernel and nearest-neighbor nonparametric regression[J]. The American Statistician, 1992, 46(3): 175-185.
- [18]Kamiński B, Jakubczyk M, Szufel P. A framework for sensitivity analysis of decision trees[J]. Central European journal of operations research, 2018, 26(1): 135-159.
- [19]Afza F, Khan M A, Sharif M, et al. Skin Lesion Classification: An Optimized Framework of Optimal Color Features Selection[C]//2020 2nd International Conference on Computer and Information Sciences (ICCIS). IEEE, 2020: 1-6.
- [20]Mporas I, Perikos I, Paraskevas M. Color models for skin lesion classification from dermoscopic images[M]//Advances in Integrations of Intelligent Methods. Springer, Singapore, 2020: 85-98.
- [21]Abbasi N R, Shaw H M, Rigel D S, et al. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria[J]. JAMA, 2004, 292(22): 2771-2776.
- [22]Monisha M, Suresh A, Bapu B R T, et al. Classification of malignant melanoma and benign skin lesion by using back propagation neural network and ABCD rule[J] Cluster Computing, 2019, 22(5): 12897-12907.
- [23]Chatterjee S, Dey D, Munshi S, et al. Dermatological expert system implementing the ABCD rule of dermoscopy for skin disease identification[J] Expert Systems with Applications, 2021, 167: 114204.
- [24]Yang J, Sun X, Liang J, et al. Clinical Skin Lesion Diagnosis Using Representations Inspired by Dermatologist Criteria[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2018.
- [25]Dhivyaa C R, Sangeetha K, Balamurugan M, et al. Skin lesion classification using decision trees and random forest algorithms[J]. Journal of Ambient Intelligence and Humanized Computing, 2020: 1-13.
- [26]Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. nature, 2017, 542(7639): 115-118.

- [27]谢斌, 何小宇, 黄伟红,等. 基于卷积神经网络的基底细胞癌和色素痣的临床图像鉴别[J]. 中南大学学报(医学版), 2019, v.44(09):113-120.
- [28]王诗琪, 刘洁, 朱晨雨,等. 皮肤科医师与深度卷积神经网络诊断色素痣和脂溢性角化病皮肤镜图像比较[J]. 中华皮肤科杂志, 2018, 51(7):486-489.
- [29]Milton M A A Automated skin lesion classification using ensemble of deep neural networks in ISIC 2018: Skin lesion analysis towards melanoma detection challenge[J] arXiv preprint arXiv:190110802, 2019.
- [30]Singhal A, Shukla R, Kankar P K, et al. Comparing the capabilities of transfer learning models to detect skin lesion in humans[J]. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2020, 234(10): 1083-1093.
- [31]T. Polevaya, R. Ravodin, A. Filchenkov, Skin lesion primary morphology classification with end-to-end deep learning network, in: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), IEEE, 2019, pp. 247–250.
- [32]Qin Z, Liu Z, Zhu P, et al. A GAN-based image synthesis method for skin lesion classification[J]. Computer Methods and Programs in Biomedicine, 2020, 195: 105568.
- [33]Jaworek-Korjakowska J, Kleczek P, Gorgon M. Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.
- [34]Hekler A , Enk A H , Berking C , et al. Pathologist-level classification of histopathological melanoma images with deep neural networks[J]. European Journal of Cancer, 2019, 115:79-83.
- [35]Gu Y, Ge Z, Bonnington C P, et al. Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification[J] IEEE Journal of Biomedical and Health Informatics, 2019, 24(5): 1379-1393.
- [36]Kwasigroch A, Grochowski M, Mikołajczyk A. Neural architecture search for skin lesion classification[J]. IEEE Access, 2020, 8: 9061-9071.
- [37]Muckatira S. Properties Of Winning Tickets On Skin Lesion Classification[J]. arXiv preprint arXiv:2008.12141, 2020.
- [38]Ratul M A R, Mozaffari M H, Lee W S, et al. Skin lesions classification using deep learning based on dilated convolution[J]. BioRxiv, 2020: 860700.
- [39]Tschandl P, Argenziano G, Razmara M, et al. Diagnostic accuracy of content - based dermoscopic image retrieval with deep classification features[J]. British Journal of

- Dermatology, 2019, 181(1): 155-165.
- [40] Allegretti S, Bolelli F, Pollastri F, et al. Supporting Skin Lesion Diagnosis with Content-Based Image Retrieval[C]//2020 25th International Conference on Pattern Recognition (ICPR). 2020.
- [41] Barata C, Celebi M E, Marques J S. Explainable skin lesion diagnosis using taxonomies[J]. Pattern Recognition, 2021, 110: 107413.
- [42] Barata C, Marques J S, Emre Celebi M. Deep attention model for the hierarchical diagnosis of skin lesions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.
- [43] Aggarwal A, Das N, Sreedevi I Attention-guided deep convolutional neural networks for skin cancer classification[C]//2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA) IEEE, 2019: 1-6
- [44] Zhang J, Xie Y, Xia Y, et al. Attention residual learning for skin lesion classification[J] IEEE transactions on medical imaging, 2019, 38(9): 2092-2103.
- [45] Lee I, Kim D, Kang S, et al. Ensemble Deep Learning for Skeleton-Based Action Recognition Using Temporal Sliding LSTM Networks[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.
- [46] Tang P, Liang Q, Yan X, et al. GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification[J]. IEEE journal of biomedical and health informatics, 2020, 24(10): 2870-2882.
- [47] Sabbaghi S, Aldeen M, Garnavi R A deep bag-of-features model for the classification of melanomas in dermoscopy images[C]//2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) IEEE, 2016: 1369-1372.
- [48] Wang W, Sun G Classification and Research of Skin Lesions Based on Machine Learning[J] Computers, Materials & Continua, 2020, 62(3): 1187-1200.
- [49] Mahbod A, Schaefer G, Ellinger I, et al. Fusing fine-tuned deep features for skin lesion classification[J] Computerized Medical Imaging and Graphics, 2019, 71: 19-29.
- [50] Perez F, Avila S, Valle E. Solo or ensemble? choosing a cnn architecture for melanoma classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.
- [51] Ahmad B, Usama M, Huang C M, et al. Discriminative Feature Learning for Skin Disease Classification using Deep Convolutional Neural Network[J] IEEE Access,

- 2020, PP (99):1-1.
- [52]Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J] IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.
- [53]Goceri E Analysis of Deep Networks with Residual Blocks and Different Activation Functions: Classification of Skin Diseases[C]//2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA) IEEE, 2019: 1-6.
- [54]Shi X, Dou Q, Xue C, et al. An active learning approach for reducing annotation cost in skin lesion analysis[C]//International Workshop on Machine Learning in Medical Imaging. Springer, Cham, 2019: 628-636.
- [55]Bdair T, Navab N, Albarqouni S. Peer Learning for Skin Lesion Classification[J]. arXiv preprint arXiv:2103.03703, 2021.
- [56]Bagchi S, Banerjee A, Bathula D R. Learning a meta-ensemble technique for skin lesion classification and novel class detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 746-747.
- [57]Combalia M, Hueto F, Puig S, et al. Uncertainty estimation in deep neural networks for dermoscopic image classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 744-745.
- [58]Khamparia A, Singh P K, Rani P, et al. An internet of health things - driven deep learning framework for detection and classification of skin cancer using transfer learning[J]. Transactions on Emerging Telecommunications Technologies, 2020: e3963.
- [59]Hameed, Nazia, et al. "Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques" Expert Systems with Applications 141 (2020): 112961.
- [60]Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [61]Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2017, 31(1).
- [62]Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

- [63]Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
- [64]周建. 数据挖掘之数据分析专业名词阐释[J]. 科技信息, 2011, 1.
- [65]Langan S M, Seminara N M, Shin D B, et al. Prevalence of metabolic syndrome in patients with psoriasis: a population-based study in the United Kingdom[J]. Journal of Investigative Dermatology, 2012, 132(3): 556-562.
- [66]张学军. 皮肤性病学.第 8 版[M]. 人民卫生出版社, 2013.
- [67]Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.
- [68]DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv preprint arXiv:1708.04552, 2017.
- [69]Shah V, Autee P, Sonawane P. Detection of Melanoma from Skin Lesion Images using Deep Learning Techniques[C]//2020 International Conference on Data Science and Engineering (ICDSE). IEEE, 2020: 1-8.
- [70]Pham T C, Luong C M, Visani M, et al. Deep CNN and data augmentation for skin lesion classification[C]//Asian Conference on Intelligent Information and Database Systems. Springer, Cham, 2018: 573-582.
- [71]Al-Masni M A, Al-Antari M A, Choi M T, et al. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks[J]. Computer methods and programs in biomedicine, 2018, 162: 221-231.
- [72]Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J] Journal of Artificial Intelligence Research, 2002, 16(1):321-357.
- [73]Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
- [74]Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning[C]//International conference on machine learning. PMLR, 2013: 1139-1147.
- [75]Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [76]Jaworek-Korjakowska J, Kleczek P, Gorgon M. Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer

- learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 0-0.
- [77]A. Hekler, J. S. Utikal, A. H. Enk, C. Berking, J. Klode, D. Schadendorf, P. Jansen, C. Franklin, T. Holland-Letz, D. Krah, et al., Pathologist-level classification of histopathological melanoma images with deep neural networks, *European Journal of Cancer* 115 (2019) 79–83.
- [78]Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

攻读学位期间的主要研究成果

一、发表的学术论文

- [1] 本人第一作者. The application of deep learning in the risk grading of skin tumors for patients using clinical images[J]. Journal of medical systems, 2019, 43(8): 1-7. (**Q1, IF: 3.058**) 对应于本文的第 5 章内容.
- [2] 导师第一作者, 本人第二作者. 584 Scoring of the clinical images for malignant degree of skin cancer using deep learning[J]. Journal of Investigative Dermatology, 2019, 139(5): S100. (**Q1, IF: 7.143**)
- [3] 老师并列第一, 本人其次. Smart identification of psoriasis by images using convolutional neural networks: a case study in China[J]. Journal of the European Academy of Dermatology and Venereology, 2020, 34(3): 518-524. (**Q1, IF:5.248**) 对应于本文的第 3 章内容.
- [4] 本人第五作者. XiangyaDerm: a clinical image dataset of asian race for skin disease aided diagnosis[M]//Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention. Springer, Cham, 2019: 22-31. (**MICCAI Workshop Best Paper**)
- [5] 本人第五作者. Studies on different CNN algorithms for face skin disease classification based on clinical images [J]. IEEE Access, 2019, 7: 66505-66511. (**Q1, IF:3.745**)
- [6] 本人第五作者. Assistant Diagnosis of Basal Cell Carcinoma and Seborrheic Keratosis in Chinese Population Using Convolutional Neural Network[J]. Journal of healthcare engineering, 2020, 2020. (**Q3, IF:1.803**)

二、参加科研项目

- [1] 基于临床图像的皮肤病辅助诊断及算法研究, 2018.9-现在, 项目组主要成员.
- [2] 多模脑部图像的肿瘤特征分析, 2018.9-2019.5, 项目组主要成员.

三、参加的竞赛及获得的奖励

- [1] 第一届中国高校计算机大赛人工智能创意赛全国一等奖.团队主要成员.
- [2] 2018 年第四届中国“互联网+”大学生创新创业大赛全国总决赛银奖.团队主要成员.
- [3] 2019 年研究生国家奖学金.

四、硕士期间参加学术交流活动

- [1] 2019.5.8, SID2019 (Society for Investigative Dermatology, SID). Chicago, USA.
- [2] 2019.10.13, MICCAI2019 (International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI). Shenzhen, China.

致 谢

致谢的第一句话往往是感慨时光的荏苒，深吸一口气，以此时为起点的反方向时间轴才更加清晰深刻。我开始在这不长的时间轴上，找寻关键的时间节点，某个可能影响我一生的节点，某个可能挖掘到宝藏的节点，这些节点亦或是正相关，亦或是负相关。不得不说的是，从成长的角度而言，所有鼓励与批判都值得被敬重，所有的孤独与欢笑都有其实际的意义，这些都拼凑在了我与中南的回忆里。

我想起了，在升华楼与导师谢斌老师的第一句对话，“你未来的打算是什么？”。在一间不大的，窗外透着郁郁葱葱的办公室里，我愣了几秒钟。的确，在此之前，我连之后是读博还是工作都未曾想过，脑海里只有“读个研就对了”，目标对一个人而言是多么重要。漫无目的地活着，漫无目的地死去也好，就怕某个时刻开始后悔自己的漫无目的。此后，在同样郁郁葱葱的办公室里，经历了无数次讨论与沟通，自己的目标也越发清晰。常常有人将老师比作指路灯，相较于路途中遇到的远光灯而言，我们的导师更显得警醒而不刺眼。在批评时，他总能照顾到学生的自尊心，在商讨时，总能尊重学生的想法，有进步时，更能不吝啬自己的表扬。

我想起了，从师兄师姐身上感受到的第一股力量。我清晰地记得，在忙完一个通宵后的何小宇师兄，依旧聚精会神地参会时的双眼；在某个清晨，伏案在实验室桌前的张文婷师姐，起身开门时的脸颊。这些场景偶尔会浮现在我代码改不出的时刻，论文写不来的时刻，没有耐心看英文的时刻。师兄师姐的引导真的挺重要的，我在与何小宇师兄一起搞项目搞论文的那段时间，我从他身上看到了很多（关于状态、效率等等），此外，我还从李沁师姐身上看到了活力，从梁照栋师兄身上看到了温柔，从宋迪师兄身上看到了领导力，从秦觅师兄身上看到了亲和力，从林学斌师兄身上看到了坦然感。我偶尔会反思一下自己有没有做到一位合格的师兄，说实话，大多数时间没有。

我想起了，在这个团队中体会到的无数次凝聚力。在参与“互联网+”大赛的期间，汗水留在了培训室和实验室，泪水留在最终的颁奖台，口水则流在了厦门的餐桌上。在组织团队活动期间，我们曾将欢笑声绽放在农家乐、桌游室、真人CS靶场……我在KTV学会了你们的歌，在互相调侃时记住了你们的口头禅，我的回忆里除了提到的师兄师姐外，还有你们的身影：徐勇、丘文杰、彭哲、程清、孙铭蔚、吴迪、万思远、李旭辉、后国炜、邝逸灵、解振南、聂海涛。

还有太多的人值得被感谢，所有帮助过我的老师以及我的父母，当我下次深吸一口气时，依旧可以在时间轴上找到你们的身影。

2021.4.17
于新校信息楼