



汕 頭 大 學

SHANTOU UNIVERSITY

硕 士 学 位 论 文

题 目 基于深度学习的毛发镜图像中
毛囊单位计数的研究

英文题目 Study on Deep Learning-Based Follicular
Unit Counting in Trichoscopy Images

姓 名 彭杰华 学 号 151909093

所在学院 工学院 导师姓名 范衡

专 业 电子与通信工程

入学日期 2019 年 9 月 答辩日期 2022 年 5 月 28 日

学位论文原创性声明

本论文是我个人在导师指导下进行的工作研究及取得的研究成果。论文中除了特别加以标注和致谢的地方外，不包含其他人或其他机构已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在论文中以明确方式标明。本人完全意识到本声明的法律责任由本人承担。

作者签名： 胡杰华 日期： 2022 年 5 月 30 日

学位论文使用授权声明

本人授权汕头大学保存本学位论文的电子和纸质文档，允许论文被查阅和借阅；学校可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，可以采用影印、缩印或其他复制手段保存和汇编论文；学校可以向国家有关部门或机构送交论文并授权其保存、借阅或上网公布本学位论文的全部或部分内容。对于保密的论文，按照保密的有关规定和程序处理。

作者签名： 胡杰华 导师签名： 范衡

日期： 2022 年 5 月 30 日 日期： 2022 年 5 月 30 日



硕士学位论文

论文中文题目：基于深度学习的毛发镜图像中毛囊单位计数的研究

论文英文题名：Study on Deep Learning-Based Follicular Unit
Counting in Trichoscopy Images

指导教师：范衡 教授

申请人：彭杰华

论文答辩委员会成员

主席：肖 刚 教授 （韩山师范学院）

委员：陈耀文 教授 （汕头大学工学院电子系）

邹安民 教授 （汕头大学工学院电子系）

赖李洋 副教授 （汕头大学工学院电子系）

李 兵 副教授 （汕头大学工学院电子系）

摘要

在临床实际中,对毛发镜图像中的毛囊单位或毛发数量进行计数是一件费时耗力且枯燥乏味的工作。如果可以通过机器快速、自动地进行可靠的计数,不仅能够降低计数成本、提高计数效率,还可以减轻临床医生的工作负担。

然而,当前对毛囊单位数量进行自动化计数的方法绝大多数仅针对短发图像,且不能有效处理毛发交叉、遮挡等复杂情况。同时,标注毛囊单位的具体类别信息和空间位置信息意味着巨大的人工标注工作量。基于此,本文提出了一个基于深度卷积神经网络、仅利用毛囊单位总计数这一弱标签信息进行学习的端到端回归计数模型,以自动评估含长发的毛发镜图像中的毛囊单位数目。

为了充分利用弱标签信息,本文提出了一种网络结构正则化机制,即在卷积神经网络的深层特征图上添加辅助预测分支,以实现标签信息在不同特征尺度空间上的复用。针对毛发镜图像中普遍存在的毛发遮挡难点,本文提出了一种图像数据正则化机制,即通过随机丢弃输入图像中部分像素的方式来模拟遮挡模式,迫使回归模型去学习图像中的空间上下文信息以弥补局部细节的丢失。针对计数标签数据分布不平衡的特点,本文提出了一种标签数据正则化的机制,即重建一个相对平滑的计数标签数据分布,以提升回归模型的学习能力和泛化能力。

在仅利用毛囊总计数这一弱标签信息、同时减少人工标记工作量的情况下,本文所提出的网络模型实现了对毛发镜图像中毛囊单位数量的快速、鲁棒估计,并取得了与专家计数标签一致性高的预测结果。在含有 80 张毛发镜图像的验证集上,模型回归预测的平均绝对值误差为 5.49,平均绝对百分比误差为 11.32%。对比实验结果表明本文所提出的网络模型以及若干正则化机制具备有效性,它们在皮肤科临床医学中评估患者毛囊密度任务方面具有广泛应用前景。

关键词: 毛发镜图像, 毛囊单位计数, 弱标签信号, 深度监督分支, 卷积神经网络

Abstract

In clinical practice, counting the number of follicular units or hairs in trichoscopy images is a time-consuming and tedious task. If reliable counting can be performed quickly and automatically by machines, it can not only reduce counting costs and improve counting efficiency, but also reduce the workload of clinicians.

However, most of the current methods for automatically counting the number of follicular units are restricted to short hair images, and cannot effectively deal with complex situations such as hair intersection and occlusion. Meanwhile, labeling the specific category information and spatial location information of follicular units means a huge workload of manual annotation. To this end, an end-to-end regression model based on deep convolutional neural networks, which uses solely the weak label information of the total number of follicular units, is proposed to learn automatically estimate the number of follicular units in hair trichoscopy images.

In order to make full use of weak label information, a network structure regularization mechanism, that is, to add an auxiliary prediction branch onto the deep feature map of convolutional neural network, is proposed in the thesis, so that the weak label information is reused in different feature scale spaces. The thesis also highlights the ubiquitous hair occlusion difficulty in trichoscopy images, and proposes an image data regularization mechanism, that is, by randomly discarding some pixels in the input image to simulate the occlusion pattern, the regression model is forced to learn the spatial context information in the image so as to compensate for the loss of local details. The unbalanced distribution of label data is underlined, and correspondingly, a regularization mechanism on reconstructing a relatively smooth distribution of data label of the counting numbers is proposed to improve the learning ability and generalization ability of the regression model.

Under the condition of only utilizing the weak label information of the total follicle number in each trichoscopy images, while reducing the workload of manual annotation, the proposed model achieves a fast and robust prediction on the validation set containing 80 trichoscopy images, where the mean absolute error of regressing the number of follicular units is 5.49, and the mean absolute percentage error is 11.32%. The comparative experimental results show that the network model proposed in this thesis and the regularization mechanisms mentioned above are effective, and they have a wide application prospect in the evaluation of patients' hair follicle density in clinical dermatology.

Keywords : Trichoscopy Image, Follicular Units Counting, Weak Label Information, Deeply-supervised Branch, Convolutional Neural Network

目 录

摘要	I
Abstract.....	II
目 录	V
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	1
1.2.1 人工计数方法	2
1.2.2 基于经典数字图像处理技术的毛发计数方法	4
1.2.3 基于深度学习的毛发计数方法	7
1.3 研究难点与挑战	9
1.3.1 图像数据集难点	9
1.3.2 计数标签分布稀疏及不平衡的特点	9
1.3.3 弱标签学习的特点	11
1.4 主要研究内容	11
1.5 论文章节安排	11
第 2 章 相关工作及基础技术研究	13
2.1 卷积神经网络基础	13
2.1.1 发展历史	13
2.1.2 经典深度卷积神经网络的基本组件	14
2.1.3 现代卷积神经网络	18
2.2 基于深度学习的复杂场景中目标回归计数	21
2.2.1 基于强标签的回归计数	21
2.2.2 基于弱标签的回归计数	22
2.3 基于弱标签的深度学习图像回归预测	22
2.4 本章小结	23
第 3 章 数据集的建立	24
3.1 毛发镜图像数据集信息	24
3.1.1 毛发镜图像的收集与标记	24
3.1.2 毛发镜图像数据集的统计信息	26
3.2 训练集和验证集的划分	27

第 4 章 基于深度学习的毛发镜图像毛囊单位计数模型	29
4.1 融合深度监督信息的卷积神经网络架构	29
4.2 损失函数设计	32
4.3 数据增强技术	33
4.4 标签平滑技术	34
4.5 本章小结	35
第 5 章 实验结果与分析	37
5.1 实验环境及参数设置	37
5.1.1 实验环境	37
5.1.2 参数设置	37
5.2 相关评价指标	37
5.3 实验设计	39
5.4 对比实验结果与分析	40
5.4.1 骨干网络的选取	40
5.4.2 颜色空间信息的重要性分析	41
5.4.3 迁移预训练模型的有效性分析	42
5.4.4 标签平滑有效性分析	43
5.4.5 基于 Cutout 的数据增强有效性分析	43
5.4.6 深度监督信息分支有效性分析	45
5.4.7 实验总结	45
5.5 回归预测模型性能分析	46
5.5.1 整体模型性能分析	46
5.5.2 困难样本分析	48
第 6 章 总结与展望	49
6.1 总结	49
6.1.1 本文工作总结	49
6.1.2 不足	50
6.2 展望	50
参考文献	52
附录 A 骨干网络选取实验的详细结果	61
附录 B 基于 Cutout 数据增强的网格搜索实验详细结果	62
附录 C 引入注意力机制的实验训练过程及结果	63
附录 D 引入多任务学习的实验训练过程及结果	64
附录 E 训练集和验证集分布一致性的 K-S 检验	65

附录 F 皮尔逊相关性检验	66
附录 G 模型回归预测毛囊单位计数的具体结果	67
攻读学位期间主要研究成果	68
致谢	69

第1章 绪论

1.1 研究背景及意义

近年来,“脱发”与“植发”成为了当前年轻人茶余饭后的新生热点话题之一,其背后反映的是随着生活节奏的加快、社会压力的增大,越来越多的人遭遇脱发问题。据国家卫生健康委员会 2019 年调查数据显示,我国脱发人数超 2.5 亿,平均每六人中就有一人脱发,其中男性约 1.63 亿,女性约 0.88 亿。与此同时,国内毛发种植行业市场迅猛发展,《2021 中国植发行业报告》显示,中国毛发移植行业市场规模从 2016 年的 58 亿元增长到了 2020 年的 134 亿元。

毛囊密度或毛发密度检测作为临床门诊中评估患者脱发水平、衡量药物治疗脱发效果、展示毛发移植手术成效中最直观的量化手段之一,是临床实际中不可避免的重要环节。但当前绝大多数医疗机构和科研单位中的专业人员进行毛发密度检测工作时,一般的操作流程是首先使用毛发镜(皮肤镜)或毛囊镜捕获某块头皮区域一定面积大小的图像,然后手动标记并计数该图像区域中的发根数量或毛囊(follicle)^①个数,最后近似计算这块头皮区域的毛发密度或毛囊密度。显然,在临床实际中,通过人工手动的方式对毛发镜中的发根数目进行计数是一项极其耗时且枯燥的工作。

为了将临床医生和研究专家从繁琐的计数工作中解放出来,研究一套能够由计算机自动化地对毛发镜图像中的毛囊进行可靠计数的智能算法具有广泛的应用场景和重要研究意义。

1.2 国内外研究现状

毛发密度,衡量的是毛发所覆盖区域单位面积内的毛发数量。毛发密度随毛发部位、个体年龄、遗传因子、种族差异等因素不同而存在差异。在临床实际中,常常对不同头皮部位的毛发进行取样,并以取样样本的毛发密度近似为相应毛发部位的毛发密度。准确评估毛发密度是包括人类种族学、皮肤病学、医美整型、烧伤护理、动物学研究等众多研究领域在内都十分关注和频繁使用的技术环节。

在人类学和种族学的毛发分布信息研究内容当中,毛发密度特征是种群信息库的重要组成部分。王冠等人^[1]通过计算中国健康汉族人头顶、后枕部及颞部三个区域的平均毛囊单位密度,填补了汉族人口毛发分布信息的空白,为国内毛发移植手术所

^① 毛囊,一种向下生长的指状表皮细胞。它是发干产生的地方,一个毛囊单位通常可生长出一根毛发,部分毛囊可以长出多根毛发,但一般不会超过四根。

需移植密度提供定量依据。张春韶^[2]通过评估广东地区健康成年人毛发密度,探讨了其分布特征与中医体质之间的关系。在毛囊移植领域,毛发密度或毛囊密度检测是评估术前患者秃发水平、计算术中待移植毛囊数量,衡量术后毛囊移植成活率的关键指标^[3]。韩兴斌等人^[4]提出在毛发移植手术过程中遵循五单位排列计数法有助于准确计算移植的毛囊总数,保证手术过程中毛囊分离质量的稳定性。在评估研发的新药物^[5-7]、新技术手段(如激光治疗^[8],局部电刺激治疗^[9]等)对患者脱发病况的治疗效果时,患者的毛发密度变化情况是最直观和最客观的量化指标。通常来说,研究人员会基于临床随机对照试验,持续地跟踪患者在接受治疗后每隔两个星期或每隔四个星期时的毛发密度、毛干直径等指标,并记录患者主观的治疗满意程度。理想的情况下,医生和患者都希望在接受若干疗程的治疗之后,从毛发密度、毛发厚度等硬性指标上观察到可观提升,且其数值与治疗前的相应数据具有统计学意义上的显著性差异。在烧伤护理研究领域,观测并计算瘢痕性脱发^①患者头皮区域的毛发存活率同样有赖于毛囊单位的计数与密度估计^[10]。另外,毛发密度同时还是影响哺乳动物毛皮绝缘性能的重要结构参数,观测物种在其进化历史进程中毛发密度的变化情况对研究进化生物学和生态生理学具有一定指导意义^[11]。

毛发计数技术,作为毛发密度估计的先行环节,是本文的主要研究内容。下文将主要从技术发展的侵入式到非侵入式、人工参与的多寡两方面简要回顾国内外研究学者在进行毛发根数或毛囊单位数量计数时采用的相关技术和方法,并介绍这些技术的应用现状。

1.2.1 人工计数方法

人工手动计数尽管是一项极其琐碎且耗时的工作,但它至今仍是皮肤科领域内计数毛发或毛囊单位数量时采用的最为普遍的方式,其计数模式和规则鲜有变化,但获取毛发观测数据的过程经历了从侵入式到半侵入式,再到非侵入式的发展历程。

最典型的侵入式研究方法是头皮活体组织检查(scalp biopsy),简称头皮活检。头皮活检几乎适用于所有瘢痕性脱发(cicatricial alopecia)病例和未确诊的非瘢痕性脱发病例。其标准操作程序是在对患者施加局部麻醉的情况下,利用直径至少4毫米的圆柱形穿刺沿着毛发生长方向获取4毫米穿刺活检标本的水平切片,并对其进行人工检查后确定毛囊单位密度,毳毛(vellus hair)^②、终毛(terminal hair)^③个数以及毛囊生长期^④(anagen)和静止期^⑤(telogen)的比例^[12,13]。由于在切片中的毛囊单位数

^① 瘢痕性脱发是指突发的局部脱发,通常是由于精神压力大、营养不足等原因引起。

^② 毳毛,又称毫毛,细小的短毛纤维,厚度一般不超过3个红细胞。

^③ 与毳毛相比,终毛既粗又长。健康人的头发绝大多数是终毛。

^④ 处于生长期阶段的毛囊最大,发干生长。

^⑤ 静止期,毛发生长周期的间歇期。

目通常较少，且它们在空间位置上相互独立、分明，因此通过头皮活检可以得到最为准确的毛囊密度。然而这种侵入式的研究手段，不可避免地会给患者带来术后创伤。同时，对临床医生来说，获取活检切片并最终对毛囊数量进行人工计数是一项繁琐的工作。因此，若想在一个治疗周期内频繁地监测脱发患者毛发密度的变化情况，头皮活检不是理想的选择。

半侵入式的毛发计数技术与侵入式计数技术相比，其对人的伤害较小，理想的情况下有可能达到微创。基于脱发量的计数方法，其首先使用均匀的纤维尖笔，在头皮上标记出一块固定大小的区域，通过一些脱发技术来获得该区域的毛发数，并借助光学显微镜或毛发成像镜（trichogram）来辅助人工计数毛发根数^[14]。与此同时，实验者还可以观测到毛干直径，毛发生长期/休止期比率等参数。这种方法考验研究人员是否掌握相当成熟的脱发水平，否则很可能会损伤患者的头发。另外，在收集毛发过程中容易遗漏毳毛或处于生长早期的毛发，从而导致计数错误。这种基于脱发量的毛发计数技术已鲜有在临床实验中被使用。

随着数字成像设备的发展，各式各样的专用皮肤镜乃至全局摄像系统逐渐被应用到毛发计数工作当中。较之前面所提的头皮活检以及脱发量计数，基于图像的分析与计数具有可重复（reproducible）的特点。例如，对一套全局摄像系统而言，通过搭配立体视觉和定位标定设备，限制患者的移动，并控制摄像头的角度、放大倍率、光照等条件与上一次摄像时一致，便可以进行精确的连续检测^[15]。与医生和患者的主观评价相比，这种将前后配对的全局头发图像进行比对的方法可以更精准地了解毛发密度等参数的变化情况。但这也存在一个问题，即复现条件苛刻，它要求患者头发的长度、颜色、形状以及发型在每一次摄像阶段都保持一致^[16]。又如，基于毛发显微成像（phototrichogram, PTG）^[17]的技术要求研究人员首先将患者顶部脱发区域中某个 1cm^2 的子区域内（通常会在观测区域的中心位置上，点下一个红色的纹身标记以供后续配准摄像视角）的头发尽量剃短至靠近头皮表皮（大约是 $2\sim 4\text{mm}$ ），剃发后即刻通过摄像设备拍照并记录毛发密度、毛发厚度、毛发长度等参数，并在两天后对同一标记区域进行二次拍照。研究人员通过手动比对前后两次不同时刻捕获的毛发图像，进行相应人工计数便可获知患者的毛发生长变化信息。Hoffman 改进了 PTG，研发了一款商业上可用的自动化系统 Trichoscan^[18]。Trichoscan（毛发纤维图像分析系统）基于标准化照相设备和校准处理程序，依托数字图像分析，能够自动地生成毛发生长分析结果。TrichoScan 的有效性虽已通过头皮活检和连续切片的组织学检查的检验，但业界^{[19][11]} 仍存在质疑其准确性的声音。学者 Van Neste 等人^[20]通过实验分析得出 Trichoscan 不仅不能处理边缘效应，而且当图像中出现染发或毛发密集、交叉等情况时，Trichoscan 极易漏记。尽管如此，在临床实际中，当研究人员能够严格控制毛发剪短

至 2~4mm（无交叉和密集情况）、确保每次测量区域完全一致^[21]时，Trichoscan 及类似软件仍然是他们在分析脱发患者毛发密度参数变化时乐意使用的辅助工具。

上述介绍的获取毛发密度信息的技术或方法都不可避免地会对患者的头部造成一定负面影响。如侵入式的头皮活检会在患者头部留下伤疤；半侵入式的毛发显微成像技术则要求将患者的毛发剃短，为了在配套的图像分析软件中取得更好的分析效果时甚至会将患者的头发进行染色。在临床实际中，很多患者并不愿意接受头皮活检手术或剪短头发。基于此，手持式皮肤镜（毛发镜，dermoscopy）或数字视频皮肤镜（video dermoscopy）作为非侵入式获取患者毛发信息的工具受到了研究人员和患者的青睐，它们也是当前医疗机构和研究单位中最普遍使用的摄像工具。

借助手持式皮肤镜，临床医生可以观察到放大 10 倍后的毛发镜图像，并对患者的头皮及毛发情况作初步诊断。皮肤镜检查也是头皮活检前一个步骤，它帮助临床医生找到正确的待活检头皮区域的位置，从而避免不必要的活组织检查^{[19]5}。视频皮肤镜能够支持更大的图像放大倍率，且一般拥有更高清的图像分辨率。通过与计算机相连，视频皮肤镜还可以将实时视频中捕捉的毛发镜图像存储起来，以备进一步地查看和分析。Folliscope[®] (LeadM Corporation, Seoul, Korea) 是皮肤科领域内比较出名的一款商用视频皮肤镜系统，并配备 Folliscope PT[®] 软件（LeadM Corporation，版本号 2.8）辅助研究人员对收集的毛发镜图像进行后续的分析。在 Folliscope PT 软件中，一般选择在 50 倍的放大倍数下通过点击鼠标的方式对毛发图像中的终毛毛囊和毳毛毛囊进行人工分类和标记，最终的计数结果会由软件根据人工标记结果自动地统计得出。使用类似 Folliscope[®] 的辅助计数系统成为了皮肤科领域研究人员十余年来获取毛发密度信息的主流范式^[13,22-24]，但当毛发镜图像中存在毛发交叉、遮挡严重等不利于计数的情形时，人工计数工作就要求计数人员具备一定的专业知识和计数经验。不幸的是，在临床实际中，不同的计数人员针对同一张毛发图像进行计数时，通常会存在数值偏差；甚至一个专业的计数人员在不同时刻对同一张毛发图像进行计数时，两次计数结果之间也很可能存在差异。

基于光学相干断层扫描（optical coherence tomography, OCT）的成像技术可以帮助研究人员捕获同一头皮区域的正面视角和垂直视角所组合成的正交视角，研究人员可以更准确地判别重叠区域的毛发和毛囊单位，从而改善毛发中存在交叉、遮挡的不利情况^[25]。

1.2.2 基于经典数字图像处理技术的毛发计数方法

在摄像设备还不发达的年代，Robert 等人^[26]较早使用数字图像处理技术来探索如何在数字图像中评估脱发患者毛发密度变化情况。他们认为可以通过量化脱发患者头皮顶部的灰度图像中头发区域与头皮区域的像素比例变化来近似分析毛发密度的变

化。通过高斯混合模型将头皮顶部的灰度图像的灰度频率直方图分解为关于头发和头皮的两个不同正态分布的灰度频率直方图,通过统计两个分布直方图中的各自的灰度频率,并以头发灰度频率的占比来量化毛发密度。随着成像技术与数字图像处理技术的发展,如今研究人员已经能够获取和保存毛发镜下的高清显微图像,许多学者尝试利用计算机算法来自动地提取出毛发镜图像中的毛发,并对毛发根数进行计数,从而将临床医生从手工计数环节中解放出来。

使用经典数字图像处理技术在处理毛发计数问题时,对毛发镜图像的处理流程大致如下图 1-1 所示。



图 1-1 经典数字图像处理技术对毛发镜图像中毛发进行计数的一般流程

对毛发镜图像进行预处理的目的是尽可能去除硬件设备、光照等因素带来的噪声,增强图像中的毛发像素,为下一步的毛发检测任务做准备。最初获取的毛发镜图像中可能存在光晕现象^①,通常需要手动截取出图像的中心区域。由于头发颜色一般深于头皮颜色,通过拉伸图像的对比度可以进一步强化前景毛发与背景头皮的像素差异。经典数字图像处理技术一般认为图像中的颜色容易受光照等因素影响,且彩色信息仿佛并不为物体检测提供关键信息。而灰度图像保留了原始图像中的梯度^②信息,在减少图像计算量的同时,有利于后续的特征提取任务。因此,在去除噪声后,通常会使用 Karhunen-Loeve transform (KLT)^[27]将彩色图像转换为灰度图像。接着,使用 Otsu 阈值分割技术^[28]将灰度图像二值化,进一步地区分毛发像素与头皮像素。在检测毛发阶段,通过采用形态学操作 (morphological operation), Canny 边缘检测以及基于霍夫变换 (Hough transform, HT) 的线检测^[29]等技术,研究人员手工设计出能捕捉图像中的物体特征的模式。在计数毛发阶段, k 均值聚类 (k-means clustering) 算法及其变体是常被使用的技术,通过算法收敛时的聚类簇数来表示图像中的毛发根数。

毛发镜图像中主要存在三大制约经典数字图像处理技术性能的因素:

(1) 图像中存在的随机亮点 (光斑)

光斑的产生是由于湿发或油性头皮表面将来自补充光的光线反射到成像硬件设备中的感光器件上。这会导致传感器在这些位置处的传感数据饱和度变高,从而将头发像素替换为亮点。若不能处理这些亮点,在对灰度图像进行二值化操作后,毛发像素中就可能出现影响毛发检测的白点。然而白点出现的位置是随机的,无法通过提前

^①相机镜头边缘接收到的光线较弱,导致成像的图像边缘较图像中心的亮度偏暗。

^②一般认为,图像中的梯度信息指示着图像中物体的边缘,是识别物体时的关键因素。

抹除像素的办法消除它们，且我们希望尽可能地保留原有的头皮、毛发纹理。许多空间滤波器算子可以抑制这些白点。如非线性中值滤波器可以将这些突兀的白点当作椒盐噪声滤除，但由于亮点大小和位置的随机性，算子大小难以确定合适的值。空间平滑滤波器可以减弱亮点的强度，但同时也带来了其他像素区域的模糊。

(2) 弯曲的长发或卷发

在对直发进行检测时，常用的方法是基于霍夫变换的线检测。霍夫等学者观察发现，笛卡尔坐标系空间中的某一点很容易转换为相应的参数空间的某条直线，若在参数空间中存在另外一条直线与之相交，则在像素空间中必然存在另一点与该点共线^[30]。基于此，在给定待检形状（如直线、圆等）时，基于霍夫变换的线检测能将图像像素空间的点映射到相应参数空间，并在参数空间中对候选点提取局部最大值，再将选出的点转换回像素空间，最终找到同一直线上或符合某种形状的点。而毛发镜图像中曲发乃至卷发的存在，则直接挑战了此方法的适用性。尽管在给定曲线方程的情况下，基于霍夫变换的检测还是有可能找到曲线上的所有点，但对于毛发镜图像中的曲发来说，其曲率通常难以被捕捉到一致的模式^[31]。

(3) 毛发交叉、遮挡

当毛发镜图像中的毛发是长发，以及拍摄设备的放大倍率有限制的情况下，毛发镜图像中便极有可能出现毛发交叉、重叠等不利于毛发检测的情况。它们很可能会导致在对图像进行形态学膨胀操作时，将交叠毛发中原本不相连的像素连通成起来，进而导致毛发数目的漏检。

针对上述难点，Shih 通过定义空间结构中像素相关性的强弱，提出基于颜色的数学形态学方法来去除图像中的与其他像素相关度低的亮点；通过联合在多个尺度图像空间上的基于霍夫变换的线检测结果（若干连续的小线段可以近似地组合成一条曲线），提升毛发检测的准确率；通过迭代机器学习算法聚类若干个小的线元标记，使得它们组合成一条条完整的毛发标记，这些的毛发标记应与上一步毛发检测任务中检测到的毛发线的匹配度尽可能高。最终这些线元组成的簇数（即毛发标记的生成数目）便是算法给出的毛发根数的预测值^[32]。然而，使用多尺度的霍夫变换是极其耗时的，在时间和精度权衡之下，作者发现联合两个尺度空间的线检测结果，并没有取得良好的毛发检测效果。

针对毛发镜图像中存在的曲发或卷发，Vallotton 等人^[33]通过反复合并小线段的方法近似表示曲线，但他们的算法对光照条件敏感，且主要面向短发计数。

针对毛发图像中存在的遮挡模式，Kim 等人^[34]提出了一种新的毛发计数方法。他们发现实际拍摄中毛发镜图像取皮面积恒定为 $5\text{mm} \times 5\text{mm}$ 的矩形区域，而当一根毛发的长度超过 5mm 时，它必然超出这个矩形区域。基于此观察，他们假设每一根头

发可以表示为一个起始点和一个终端点，起点在矩形区域内部，终端点在矩形区域的外边框上。通过遍历毛发骨架图像^①中每条骨架线（skeleton line）上是否同时存在矩形内部的起点以及矩形边框上的终端点来计算毛发数目。此方法的有效性首先有赖于图像处理前期进行的形态学处理没有破坏完整的毛发信息，既不因腐蚀操作将细微的毛发去除，又不因膨胀操作将密集相连的毛发合二为一；且当图像中存在终端点在矩形内部的短发时，算法同样会出现漏检的情况。

Seo 等人^[35]基于平均毛发长度是图像整体网格长度的一半的假设，提出网格线选取和直方图统计的方法来计算二值化图像中的毛发。然而，当两个相邻毛发像素（它们很可能分属于不同的毛发）与图像某条图像网格线毗邻时，算法会将这两个像素归为同一根毛发的像素，故此时算法会失效。

综上，基于经典图像处理技术的毛发计数方法依赖于人工设计的特征提取器，且它们对采集图像的光照条件、视角变化、尺度变换等干扰因素十分敏感。一旦某些条件发生改变，算法往往会“牵一发而动全身”，许多参数就需要重新设计。

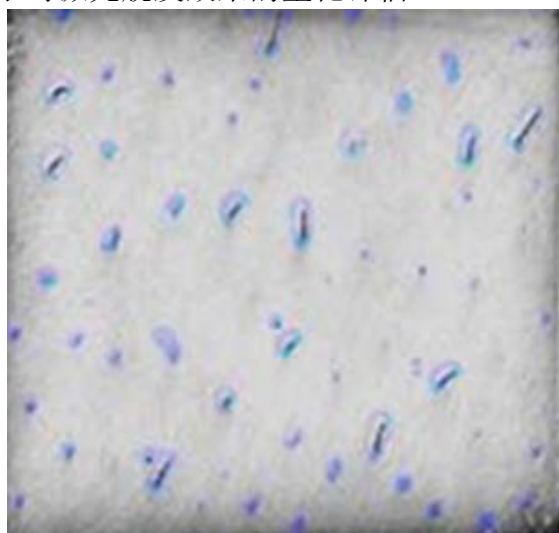
1.2.3 基于深度学习的毛发计数方法

随着深度学习近十年来的爆发式发展，计算机视觉领域的研究已被深度卷积神经网络（convolutional neural network, CNN）所主导，CNNs 在图像分割、图像分类、物体检测与识别等基础图像任务上展示了其卓越性能^[36]。较之经典图像处理技术，CNNs 对各种干扰因素更不敏感，鲁棒性更好。一些学者率先将深度学习技术运用到与毛发、头皮相关的图像任务当中。Benhabiles 等人^[37]检验了 Kaggle, DEX-IMDB-WIKI 以及 DEX-ChaLearn 三个经典网络在基于人脸图像的脱发水平分类任务上的性能。Chang 等人^[38]基于 Faster R-CNN 和 Inception ResNet_v2_Atrous 两个基础模型开发了一款能够诊断头皮屑、毛囊炎、脱发、油性头皮四类头皮病症的头皮健康检测系统。高树辉等人^[39]分析了 CNN 应用于物证检验领域中毛发的无损检验的可行性，并开发了对毛发物证显微图像进行自动分类的 Hair-Net^[40]。

Gallucci 等人^[41]是较早使用深度学习技术来直接对毛发进行计数的学者。他检验了经典神经网络模型（包括 VGG16, ResNet50, U-Net 等）在计数皮肤图像中毫毛数量任务上的性能，验证了这些常被用于分类或分割任务的模型稍作改动后可用作构建一个计数毫毛数量的回归器。基于 U-Net（最早被提出用于医学图像分割）^[42]回归计数的性能最好这一实验结果，作者认为对毫毛进行回归计数的过程与毫毛分割任务存在很高程度的耦合，因此多任务学习（multi-task learning）或可被引入。然而，该文处理的图像数据是被剃短后的皮肤毫毛图像（如图 1-2（a）所示），应用场景仅限

^① 通过形态学细化（thinning）操作得到，仅保留像素连通区域中的中心像素，方便下一步的线检测。

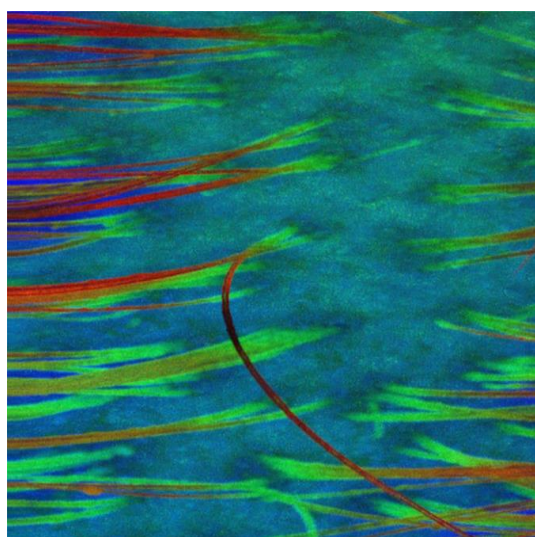
于对激光脱发效果的量化评估。



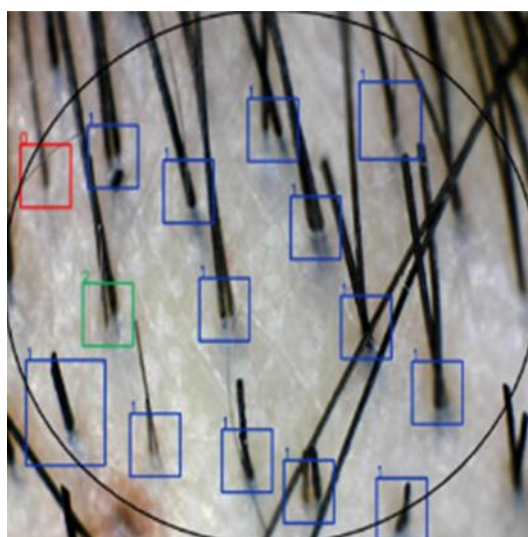
(a) 皮肤毫毛图像^[41]



(b) 剃短的头发表^[43]



(c) 基于 OCT 的头皮毛发彩色成像^[44]



(d) 一般的 RGB 彩色长发图像^[45]

图 1-2 基于深度学习的毛发计数方法处理的图像数据样式

Jarek 等人^[43]基于短毛发的线段监督信息（线段的一头是短发的发尖，另一头是发根位置），利用 UNet 实现对发尖位置的检测，并以检测点的数目代表毛发计数。但该方法要求患者将毛发剃短至 0.5mm（如图 1-2（b）所示），系半侵入式检测方法，无法满足那些有毛发密度或毛囊密度检测需求而不愿意剃发的门诊患者。

Gregor 等人^[44]利用 OCT 成像技术获得的头皮毛发的三维灰度信息，在尽可能保存深度信息的情况下，将其转换为彩色的二维图像，如图 1-2（c）所示。他们构建了一个类似 YOLOv3 的多任务全卷积神经网络以同时进行毛囊单位的分类和计数任务。尽管在毛发总数值上的计数误差小于人类专家之间计数误差，但模型对不同类别毛囊

单位的分类与计数上出现了比较大的误差。

Kim 等人^[45]检验了基于深度学习的目标检测网络（包括 EfficientDet、YOLOv4 等）在识别和分类不同毛囊类别（单毛毛囊和多毛毛囊）时的检测性能。尽管在实验结果中，作者发现检测网络在毛囊分类检测环节中可能出现误检（如将三毛毛囊识别为双毛毛囊；无法检测白发的毛囊等），但还是在一定程度上支撑了当拥有足够多训练数据集的情况下，将基于深度学习的算法模型用于自动的头发密度测量具有可行性。

1.3 研究难点与挑战

不同于上文提及学者的工作，本文所处理的毛发镜图像数据集更具挑战性，难点主要包括以下三个方面：图像数据集样本难点；计数标签分布稀疏、不平衡的特点；仅有毛发镜中毛囊单位总计数这一软标签作为监督信息的特点。

1.3.1 图像数据集难点

毛发镜图像数据集中的难样本例图如图 1-3 所示，它们主要存在以下四类不利于毛囊检测与计数的模式：

（一）存在部分图像质量一般的图像数据

如子图(a)，图像中整体像素偏暗，从视觉效果上较难辨别毛囊信息；子图(b)中存在相机镜头运动模糊。

（二）毛发颜色多样

除了大部分毛发是黑色之外，还存在部分白发和染发，如图(c)、(d)。

（三）难以辨别的纹理信息

如图(e)中头皮上出现粗细与毛发线段相近的纹理；图(f)中部分毫毛纹理过细，不仅人工标定存在困难，机器也容易出现漏检。

（四）复杂的遮挡模式

由于本文图像数据集中的毛发绝大多数是长发，因此广泛地存在复杂的交叉、遮挡模式，如图(g)所示，部分毛发的发干会遮挡其他毛发的毛囊；图(h)中某些异物（如头皮屑）会遮挡毛囊。

可见，毛发镜图像样本难点众多，这为构建一个鲁棒的毛囊计数模型设置了障碍。

1.3.2 计数标签分布稀疏及不平衡的特点

如图 3-3 所示，毛囊单位计数标签在极端值处的样本量少，而相对中间的数值区间中的样本量多。同时，部分计数数值对应的毛发镜图像样本缺失。这种标签分布稀疏与不平衡的特性为构建一个准确的计数模型提出了难点。



(a) 对比度低



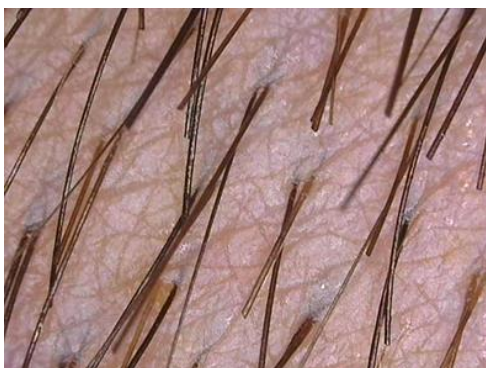
(b) 图像模糊



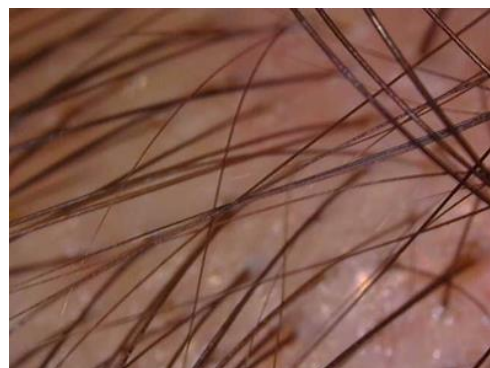
(c) 存在白发



(d) 存在染发



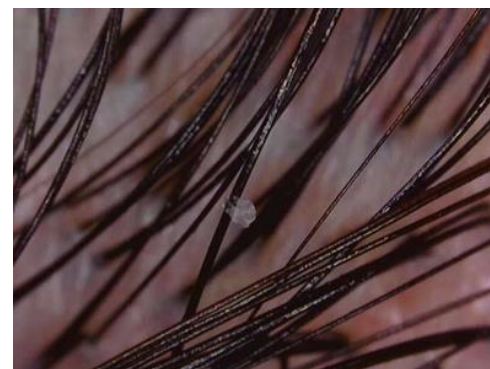
(e) 头皮纹理



(f) 细发难辨



(g) 交叉遮挡



(h) 异物遮挡

图 1-3 毛发镜图像数据集中存在的难点与挑战

1.3.3 弱标签学习的特点

本文数据集的标签信息仅有毛发镜图像中毛囊单位的总计数这一弱标签,而关于毛囊单位的具体空间信息、类别信息缺失,这是典型的弱监督学习中的不确切监督(inexact supervision)学习^[46]。只标注毛囊单位的总计数信息,一方面减轻了标注人员的工作量,但同时也为算法模型如何在仅有弱标签信息的情况下实现可靠预测带来挑战。

1.4 主要研究内容

本文围绕毛发镜中毛囊单位计数这一任务,基于深度学习技术,意图构建一个端到端的回归计数模型,以实现毛发镜图像中的毛囊单位进行快速、自动、可靠的预测。主要研究内容有:

(一) 研究基于深度卷积神经网络的回归计数模型,该网络模型架构应能在仅利用毛囊计数标签这一弱监督信号的情况下,对毛发镜图像数据集中毛囊单位的分布进行良好的学习,并在模型验证阶段取得可靠的计数结果。

(二) 针对毛囊计数标签不平衡的特点,研究在网络模型在训练阶段适宜的损失函数,尽可能缓解对数据集中低频样本学习欠佳的状况。

(三) 针对毛发镜图像数据集样本存在的多重难点,研究适宜的数据增强方式,以增强图像数据样本的多样性,模拟毛发间普遍存在的遮挡模式,从而提高网络模型的鲁棒性。

(四) 针对毛囊计数标签分布稀疏的特点,研究重建一个更平滑的标签数值分布空间,以使得回归计数模型对连续标签数值区间中的极端值、缺失值所对应样本具有更好的感知能力。

1.5 论文章节安排

本文共包含六个章节,各章具体内容如下:

第1章:绪论。本章介绍了毛囊单位计数的研究背景与意义,概述国内外的研究历史、进展与现状,并确立本文的研究难点、主要研究内容及章节安排。

第2章:相关工作及基础技术研究。本章介绍了深度学习的基础知识,并调研了与毛囊单位计数任务相近、相关的工作。

第3章:数据集的建立。本章描述了本文所使用毛发镜图像数据集的整体信息,包括毛发镜图像的获取与标定,训练集与验证集的划分。

第4章:基于深度学习的毛发镜图像毛囊单位计数。本章阐述了本文针对毛囊单位回归计数任务所提出的算法模型,包括融合深度监督信息分支的卷积神经网络架构,

加权计数标签样本密度的损失函数。同时介绍了针对任务中存在的不同难点而分别提出的正则化技术方法，包括标签数据的正则化，图像数据的正则化。

第 5 章：实验结果与分析。本章介绍了本文开展数据实验的环境及相关参数设置；描述了回归任务的相关评价指标；阐述了本文进行的各组实验并分析了实验结果。

第 6 章：总结与展望。本章对全文的研究内容进行总结，阐述了本研究的主要贡献、创新点及不足，并对未来工作的展望。

第2章 相关工作及基础技术研究

2.1 卷积神经网络基础

深度学习(deep learning, DL)是机器学习(machine learning, ML)领域的一个分支。人类在不断的实践中进行学习、总结经验,并利用经验做出有效决策。那么,机器也能利用这些“经验”吗?计算机科学家认为人类的这些“经验”可以被各式各样的传感器采集并转换为计算机二进制世界中的“数据”形式。因此机器学习的研究内容就是通过计算机在数据中学习并产生决策模型的算法,即“学习算法”^[47]。

深度学习的核心要义在于“学习”,它研究的是如何通过某些特定的计算单元如卷积神经网络(convolutional neural network, CNN)和非线性处理机制如激活函数(activation function)来学习样本数据中的内在规律和表示层次,最终让机器获得类人的分析数据和学习数据的能力^[48]。深度学习的典型特征在于“深度”,通过增加卷积神经网络中隐藏层的层数来构造深层的神经网络。隐藏层多了,意味着神经网络中参数容量的增多。一般认为,参数量越多,模型复杂度越高,深度学习模型就能完成更复杂的学习任务。

深度学习的深层学习理论可以被动物视觉机理的研究发现所支撑^[49,50],研究证明了动物大脑的视觉皮层具有分层结构,且不同神经元对各种空间位置和方向偏好不同(表现为神经元释放出的电信号强弱程度)。基于此,深度学习的分层学习理论认为,通过设置深层神经网络构建特征的多重抽象表征,第一层特征通常是比较简单的边缘特征,第二层在第一层的特征基础上可以学到角或曲度等稍复杂的特征;随着层的加深,更复杂的特征、更具语义信息的特征将被习得^[51]。

本节将主要介绍深度学习中基于卷积运算的神经网络 CNN。

2.1.1 发展历史

深度学习的历史最早可追溯到上世纪 40 年代,彼时随着生物学习理论^[52,53]的发展,研究人员陆续构造出 McCulloch-Pitts (M-P) 神经元和感知机^[54] (perceptron),但由于彼时的感知机模型仍是线性模型,它甚至无法解决简单逻辑运算异或(XOR)问题^[55],这导致深度学习研究经历了第一次低谷。

深度学习的第二次浪潮伴随着上世纪 80 年代联结主义(connectionism)的潮流而出现。联结主义认为,当网络将若干个简单的计算单元连连接至一起时可以实现智

能行为，现在看来这与物理世界中群体智能^①的涌现不谋而合。联结主义时期形成的分布式表示^[56]以及反向传播（back propagation, BP）算法^[57]至今仍在深度学习中占据重要地位，此时借助 BP 算法已经可以训练具有两个隐藏层的神经网络。然而，含多隐藏层的神经网络仅使用经典 BP 算法依然难以训练且在实际部署中成效惨淡；与此同时，核方法发展出完整的泛函理论，通过核函数计算能将机器学习问题变成凸优化问题^[58,59]，且其模型较神经网络更显式，后者逐渐淹没了神经网络。

2006 年，深度信念机（deep belief machine）^[60]被提出，通过逐层预训练深度神经网络的方式使得训练一个深层神经网络成为可能，深度学习由此开始立足脚跟。至 2012 年，深度卷积神经网络 AlexNet^[61]在 ImageNet 分类竞赛中拔得头筹，并以绝对精度优势超过第二名，由此拉开了深度学习延续至今的爆发式发展的序幕。如今，深度学习的影响力在不断蔓延，并愈发地在生物、物理等其他基础学科研究中发挥着重要作用^[62,63]，催生着科研新范式，书写 AI for Science 的新时代。

2.1.2 经典深度卷积神经网络的基本组件

2.1.2.1 M-P 神经元模型

图 2-1 展示了卷积神经网络中最基本的运算单元 M-P 神经元模型及其神经科学基础的启发由来。如子图(a)所示，在生物神经网络中，每个神经元的树突中接收来自前一神经元传来的化学物质，这些化学物质可以改变神经元胞体内的电位；若神经元中的电位超过一个阈值，则该神经元就会被“激活”，并通过轴突将上一神经元传来的神经冲动传至下一个神经元或效应细胞。子图(b)中的 M-P 神经元模型尝试抽象出(a)中的“激活”模式。每个神经元会接收到多个输入信号，通过对这些信号加权连接得到总输入刺激，当刺激值高出阈值 θ 时，神经元会被激活，并通过激活函数获得非线性输出。而所谓神经网络就是很多这样的计算单元通过嵌套、迭代得到的一个数学系统^[47]。

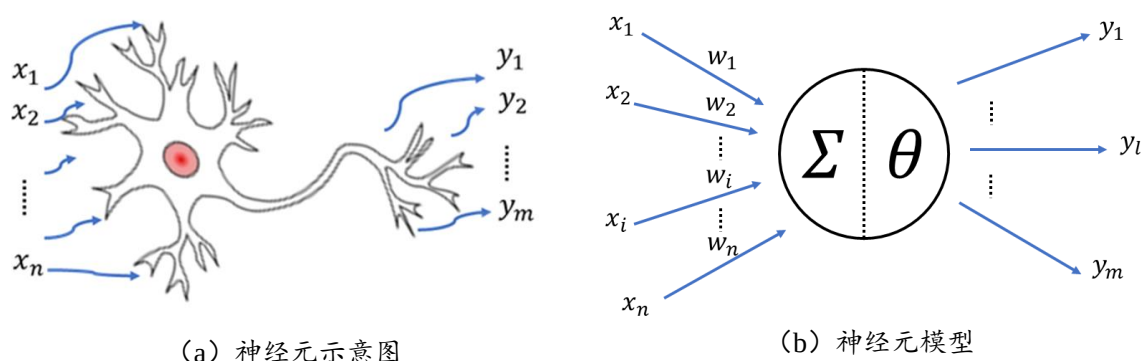


图 2-1 M-P 神经元模型

^① 群体智能源于对社会性昆虫的群体行为的研究。群体行为具有并行分布式的特点。

2.1.2.2 激活函数

为了使神经元模型获得类似阈值控制门的性质，最早使用的激活函数是 sigmoid 函数，其计算式如公式(2-1)所示，导数形式为公式(2-2)。如图 2-2 中子图(a)所示，sigmoid 能将任意范围的输入挤压到(0,1)的数值范围，且其连续、光滑的性质便于基于梯度的学习，一时受到研究人员的追捧。但通过观察图 2-2 中子图(b)可知，在很大一部分数值区域里内，sigmoid 的导数都近于 0，这很容易使得模型在训练阶段出现梯度消失。

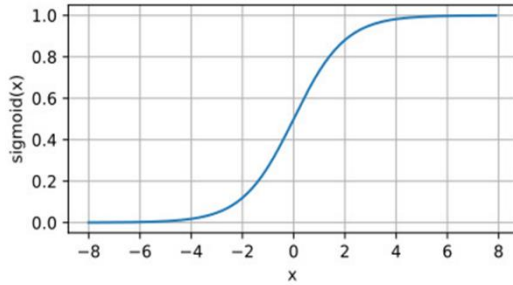
$$S(x) = \frac{1}{1+e^{-x}} \quad (1-1)$$

$$\frac{dS(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = S(x)(1-S(x)) \quad (1-2)$$

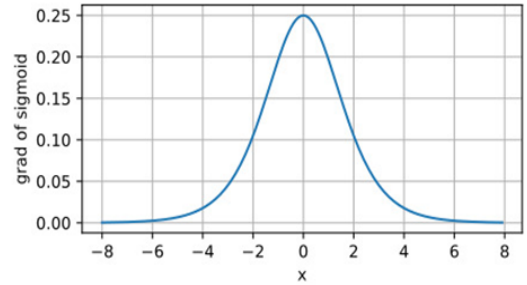
陆续有学者发现修正线性单元（rectified linear unit, ReLU）的价值与优势^[64,65]，如今人们在构建 CNN 时，通常会选取 ReLU 作激活函数，其计算公式及其导数如下：

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (1-3)$$

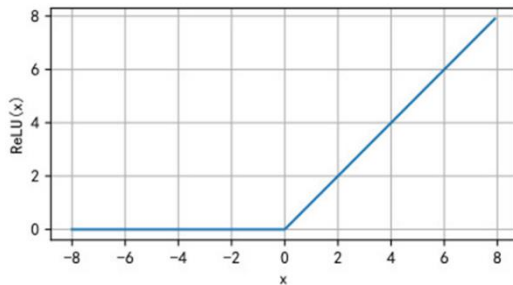
$$\frac{df(x)}{dx} = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (1-4)$$



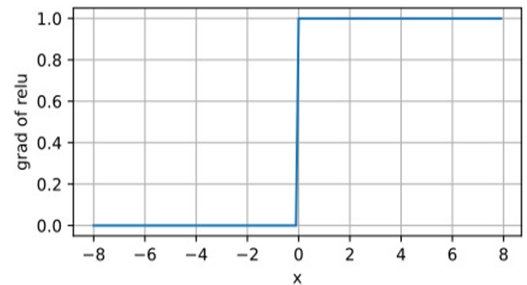
(a) sigmoid



(b) sigmoid 的一阶导



(c) ReLU



(d) ReLU 的一阶导

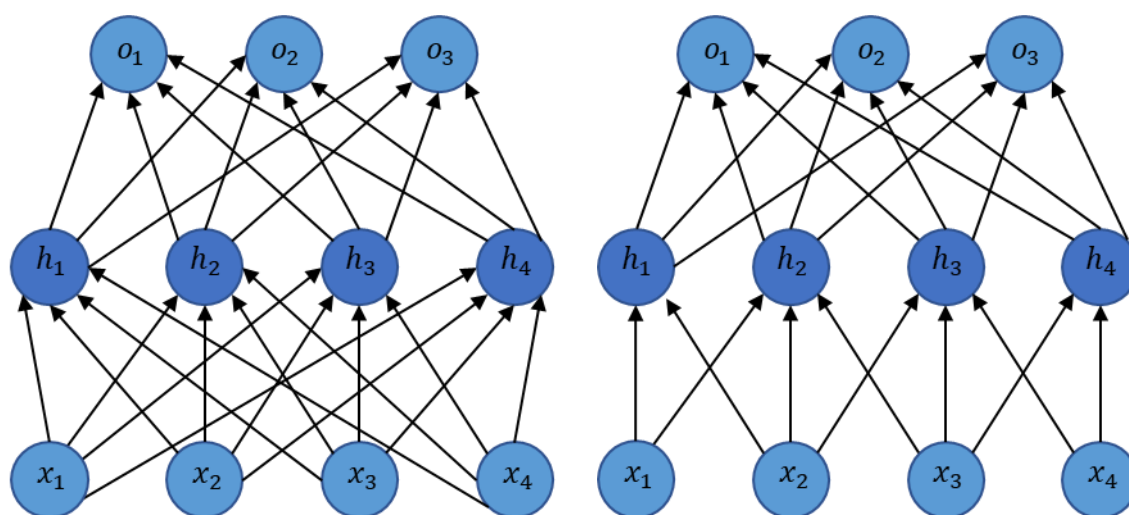
图 2-2 常用激活函数

2.1.2.3 全连接层与多层感知机

图 2-3 中子图(a)展示了多层感知机 (multilayer perceptron, MLP), 这是最简单一种深层神经网络。MLP 具有层级结构, 每层神经元与下一层所有神经元均有连接, 故称这样的层为全连接层。由于同层间神经元不连接, 也不存在跨层连接, 这样的神经网络结构也被称为多层前馈神经网络 (feedforward neural networks)。

图中 MLP 具有 4 个输入和 3 个输出, 隐藏层包含 4 个隐藏单元, 由于输入层不涉及计算, 故该 MLP 层数为 2。其中, 隐藏层与输出层的每一个基本运算单元都是 M-P 神经元模型。每一条连线都代表上一层神经元向下一层神经元传递信息的权重, 而所谓神经网络学习到的神经网络的学习过程, 就是根据训练数据来调整这些权重。

通常会在 MLP 的线性层后紧接激活函数以神经网络获得非线性特性, 否则线性层的叠加依然是线性层。



(a) 多层感知机中的全连接

(b) 卷积神经网络中的稀疏连接

图 2-3 多层感知机的全连接与卷积神经网络的稀疏连接

2.1.2.3 从全连接层到卷积层

MLP 中的全连接层无法对数据中的结构信息进行利用, 尤其是对于图像这种高维感知数据而言。我们容易从图像结构信息中提炼两种先验知识, 平移不变性和局部性。平移不变性是指当检测对象出现在图像中不同位置时, 其特征并不会发生改变。局部性是指可以首先通过获取局部图像信息, 最后聚合局部特征的方式近似地获得整个图像级别的特征信息。

我们希望神经网络能够利用这些先验结构信息做一些适应性改变, 于是卷积神经网络 (CNN) 应运而生。CNN 中的神经元对同一对象在不同图像区域的计算结果应

具有相似输出；神经网络的前几层只需检测图像局部区域信息，而在后面的层可以聚合这些局部特征，进而获得图像的全局特征。

公式(2-5)描述了二维卷积运算的计算公式，可见，CNN 中的卷积层执行的计算实际上是互相关运算。对比数学意义上的二维卷积运算，如公式(2-6)所示，卷积层中的卷积核没有进行权重参数的翻转。图 2-4 中子图(a)展示了使用一个 2×2 的卷积核（无填充）依步幅大小 1 对图像进行二维卷积运算的一个例子。通过使用不同卷积核，卷积神经网络可以获取不同的局部特征。

$$y_{i,j} = \sum_{a=1}^h \sum_{b=1}^w w_{a,b} x_{i+a,j+b} \quad (1-5)$$

$$y_{i,j} = \sum_{a=1}^h \sum_{b=1}^w w_{-a,-b} x_{i+a,j+b} \quad (1-6)$$

如下图中的子图(b)所示，对图像进行二维卷积时，一个固定大小（为满足局部性）的卷积核会在全局图像数据上进行扫描（以相同卷积核提取图像中每一处的同一特征模式）并作互相关运算，从而满足了平移不变性和局部性。

CNN 可以看作是一种特殊的 MLP，只不过对每一层的权重参数作了先验结构信息的假设，实现了稀疏连接（如图 2-3 中(b)图所示）与权重共享。正因为如此，CNN 中的卷积层较全连接层需要更少的参数。同时，通过额外设置填充和步幅，控制输入通道、输出通道个数，可使 CNN 在每个空间位置上获取图像的多方面特征。

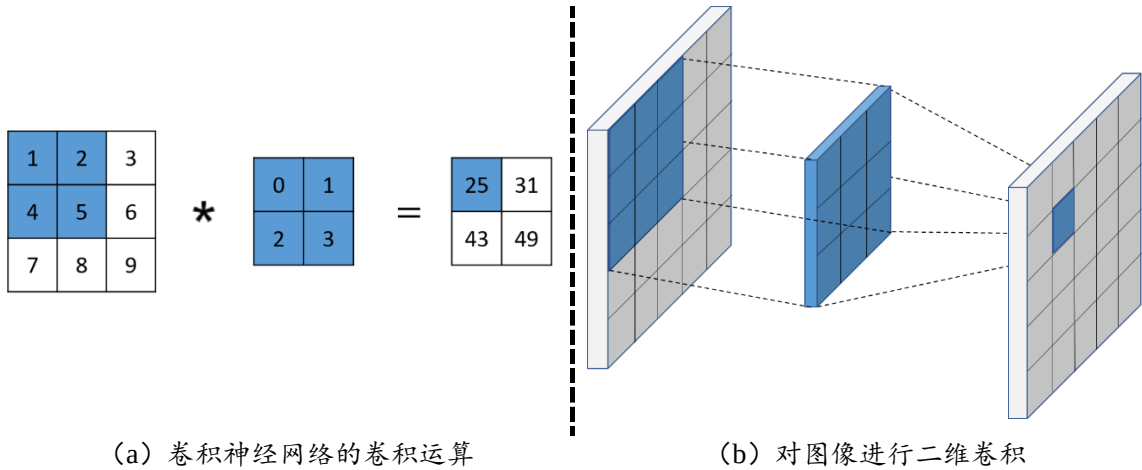


图 2-4 图像二维卷积运算示意图

2.1.2.4 汇聚层

卷积层能够提取到图像某一个尺度空间（某一分辨率的特征图）上的局部特征信息，需要进一步通过汇聚层^①[47,66]来聚合特征以获得更大的感受野（receptive field）。图 2-5 展示了两种汇聚层的运算过程，汇聚层的计算同样由一个固定大小的窗口依步

^① 汇聚层（pooling），中文翻译多称之“池化”。然“池化”参考“pool”的本义“水池”一词，不能体现其聚合前一层卷积层提取到的特征信息的作用。故本文参考领域前沿学者对“pooling”的解释。

幅大小 1（无填充）在输入的特征图上滑动，最大汇聚层输出窗口中的最大值，平均汇聚层输出窗口中的平均值。通过引入汇聚层，在汇聚浅层局部特征信息、增大深层卷积层感受野的同时降低了卷积层对位置的敏感性。

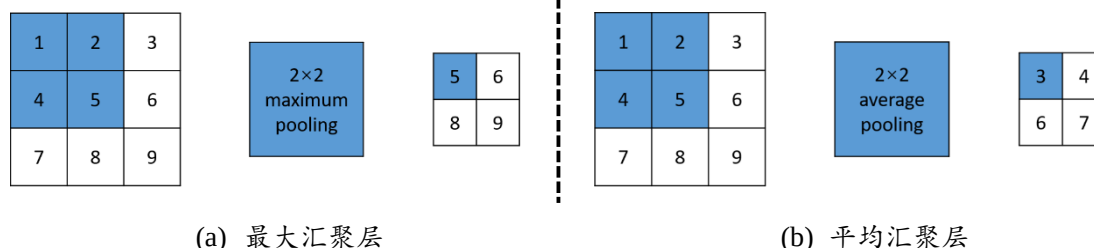


图 2-5 最大汇聚层与平均汇聚层计算示意图

2.1.2.5 卷积神经网络 LeNet

本节以最早发布的卷积神经网络 LeNet^[67]为例介绍构建一个完整 CNN 所需的基本组件。如图 2-6 所示，LeNet 由卷积编码器（由两个卷积层、两个汇聚层组成）和三个全连接层组成。

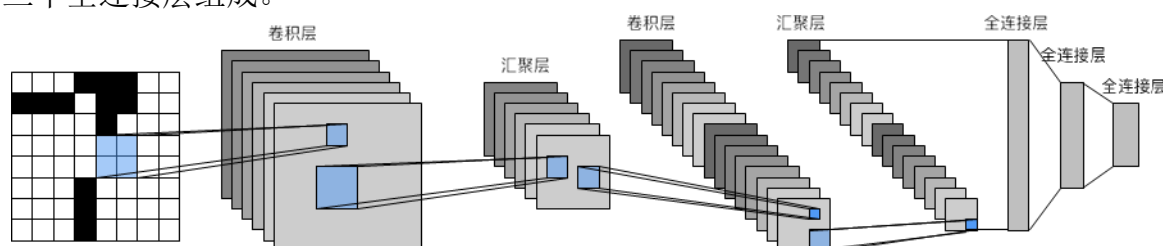


图 2-6 LeNet 网络示意图

LeNet 对输入图像进行两层特征提取、两层特征汇聚，并最终通过全连接层预测出手写字符识别的结果。每个卷积块的基本单元是一个卷积层、一个激活函数和平均汇聚层。卷积层中不同卷积核提取图像中不同模式的局部特征信息；激活函数（sigmoid）为特征信息变换引入非线性；平均汇聚层在下采样的同时聚合了前一层卷积层提取的局部特征信息。

值得注意的是，随着卷积层的加深，特征图的空间分辨率会降低，同时特征通道数增加，而越到深层的特征更容易被全连接层解码输出好的结果。这种平滑的信息压缩、特征提取的处理方式成为了往后卷积神经网络特征提取的范式。

2.1.3 现代卷积神经网络

本节将介绍领域研究人员在经过大量探索与试错后的研究成果结晶，它们被奉为经典，时至今日仍然启发着研究人员对现代卷积神经网络的设计。

2.1.3.1 模块化设计

VGG^[68]是首个尝试对卷积神经网络进行模块化构造的网络，它提出的 VGG 块由一系列卷积层及用于下采样的最大汇聚层构成，如图 2-7 子图(a)所示。模块式设计

有助于更加复杂的神经网络。

另外，VGG 作者 Simonyan 等人认为使用深层卷积网络且选择较小的卷积核比在浅层使用大的卷积核更有效。其中，小卷积核的设定几乎变得约定俗成了，直至最近才遭到质疑^[69]。

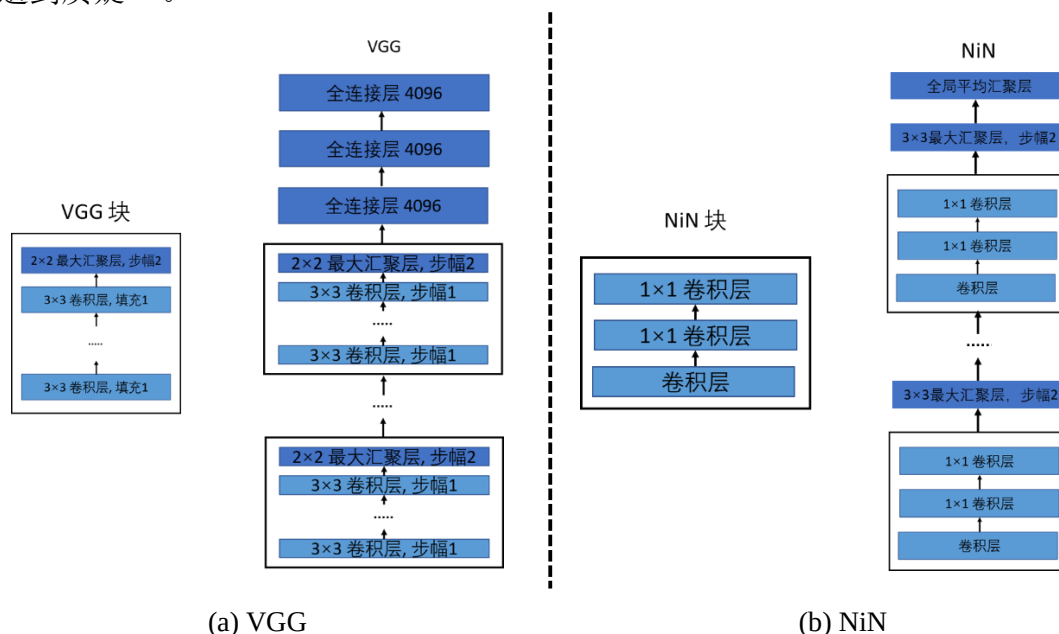


图 2-7 VGG 块与 NiN 块

2.1.3.2 全局平均汇聚层与 1×1 卷积层

Lin 等人针对浅层网络中使用全连接层可能会对表征空间造成破坏^{[66]257} 的问题，提出将全连接层替换为 1×1 的卷积层，如图 2-7 子图(b)所示，通过在卷积层后接两个 1×1 的卷积层（支持更多像素非线性，转换特征通道数）构造了一个 NiN 块^[70]。进一步地，深层卷积层后接的全连接层也被替换为全局平均汇聚层（对特征图上所有位置进行求和）。通过以上设计，显著降低模型参数数量的同时减少过拟合。NiN 中提出的全局平均汇聚层被后续网络广泛沿用。

2.1.3.3 含并行连结的网络

以往的 CNN 中卷积核大小的选择相对单一且多是串联结构，GoogLeNet^[71]吸收了 NiN 中的串联卷积层的思想，并探索了各卷积核在拓扑结构上更有效的组合，提出了如图 2-8 所示的 Inception 块。Inception 块有四条并行支路，其中三条路径分别使用 1×1 、 3×3 、 5×5 的卷积层以抽取不同大小的空间局部信息，另外一条路径对上一层特征信息作聚合。为了使得这四条路径的特征信息组合不至于造成模型复杂度的巨幅提高，在每条支路中普遍地使用了 1×1 卷积层以降低特征通道数。通过这种并行卷积层、连结不同大小的图像特征的方式，特征图信息的多样性增多。

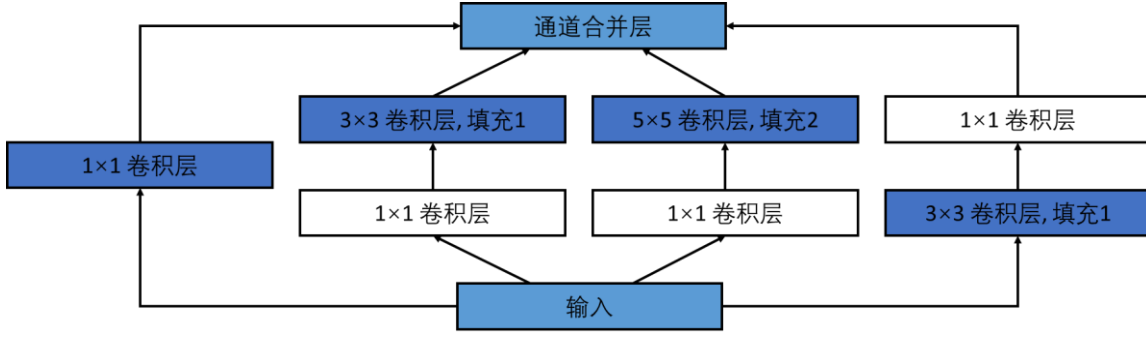


图 2-8 Inception 块

2.1.3.4 批量归一化层

深度卷积神经网络在模型训练的后期，越是靠近输入数据的底层，得到梯度值越小，很容易出现梯度消失的情况。Ioffe 等人^[72]认为，输入变量的内部协变量偏移（internal covariate shift）是导致梯度消失的原因之一，即中间特征层的变量在网络的训练过程中可能会产生比较大的数值偏差。为此，他们提出了一种简单而有效的方案，批量归一化层（batch normalization, BN，也称批量规范化层）。

与在输入数据阶段将输入变量进行标准化（使其均值为 0，方差为 1）类似，BN 层对每一次训练批次中传入的输入变量也作归一化，并对变量额外添加了随机偏移和缩放，如下式：

$$x_{i+1} = \gamma \frac{x_i - \hat{\mu}_B}{\hat{\sigma}_B} + \beta \quad (1-7)$$

其中， x_i 为当前批次 B 中第 i 个样本的输入变量， x_{i+1} 为归一化后的输入变量， $\hat{\mu}_B$ 、 $\hat{\sigma}_B$ 分别为当前批次 B 中所有输入变量的均值、方差， γ 、 β 分别为 BN 层中的两个用于控制输入变量的缩放与偏移的可学参数。

BN 层的引入可以加速网络模型的收敛速度，但一般并不会带来模型精度提升。有学者认为，BN 层并没有减小协变量的偏移，而是通过缩放与偏移两个参数向批量数据中加入了噪声以控制模型复杂度^[73]。

2.1.3.5 残差网络

构造一个深度卷积神经网络的模式通常是在已有深度的基础上额外添加更深的卷积层，从最早只有 2 层卷积层的 LeNet 到夺得 2014 年 ImageNet 图像识别挑战赛冠军、具有 11 层卷积层的 GoogLeNet，CNN 的深度呈现上升的趋势。但由于训练深度网络的难以，CNN 深度的提升并不总能带来模型性能的提升。

残差网络（ResNet）^[74]中引入了残差块，试图使上一块的输出辅助新加入块的学习，如图 2-9 子图(b)所示，通过将上一块的输出 $f(x)$ 跳连(shortcut)到新加入块的输出部分。这种残差、跳连结构构造了一个恒等映射（identity mapping）^[75]，使得前向传

播的特征信息可以从底层传递到高层，同时后向传播的梯度信息又可以传递到底层，从而极大地改善了梯度消失的问题。这种残差、跳连的结构模式，几乎成为了现代卷积神经网络设计的标配。

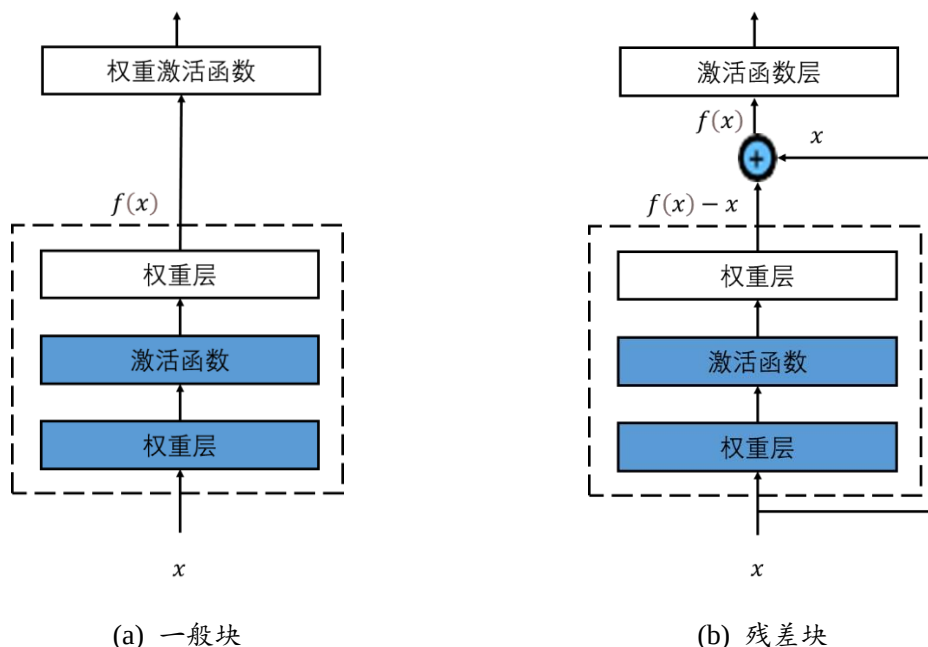


图 2-9 残差块中的跳连结构

2.2 基于深度学习的复杂场景中目标回归计数

2.2.1 基于强标签的回归计数

以线、框、像素块等大像素级别的标注方式来标定物体在图像中的具体空间位置能获得一种强力的监督信息，它们常用于有监督地训练 CNN 精准检测与定位图像中的各个物体。使用一阶段、二阶段的深度学习目标检测检测任务可以解决绝大多数简单场景中物体的检测，如无人机下的计数^[76]，检测与计数花朵数目^[77]等。乃至实例分割也可用于辅助计数物体数量，如显微镜下细胞计数^[78,79]。然而，当物体对象相对密集，且存在交叉、遮挡等复杂情况时，对物体对象进行大像素级别标注十分困难，且标注信息过于靠近也不利于检测。

Lempitsky 等人^[80]回避了在密集场景中检测和定位物体对象实例的难度，通过获得图像中物体的点标注信息，将物体计数问题转换为估计物体密度的问题。受此启发，现在主流的密集人群计数方法^[81,82]大多基于人群头部的点标注信息构建人群密度图的金标准（ground truth），并尝试构建一个基于 CNN 的密集人群密度图回归模型，对密度图进行累加计算便得全图的人群计数数目。

基于点标注信息的目标检测与计数的方法在农业采果^[83]、细胞计数^[84]等任务也

有应用。

2.2.2 基于弱标签的回归计数

上一小节中描述的人群计数方法多采用像素级的人员点标注信息,这种标注方式代价昂贵,而基于人群总计数的弱标签计数方法是一种更经济的标记方法。

Wang 等人^[85]是最早借助 CNN 直接回归计数密集人群人员数量的学者之一,为了削弱图像背景中的建筑物、树木等物体的影响,他们在训练数据中增加了人员数目为零的负样本图像。传统 CNN 一般只具有图像的局部感受野,Liang 等人^[86]同样基于弱标签信息,借助具有全局感受野的 Transformer^[87]构建了关于密集人群的回归计数模型,其预测性能超过现有的所有基于 CNN 的弱标签计数方法,并与基于点标注强监督信息的计数方法具备可比性。

无独有偶,在拥堵场景下的车辆计数^[88]、农业场景中的水果计数^[89]、植物叶片计数^[90]等任务中,都有学者利用 CNN 构造一个回归器直接对图像中相应对象的数目进行回归预测。

2.3 基于弱标签的深度学习图像回归预测

深度学习还在另外一些需要对复杂、隐晦的图像信息具有解释与预测能力的任务中大放异彩。

如基于人脸图像的年龄预测任务,图 2-10 展示了人脸图像例图。容易想到,人脸图像中的毛发、五官、皱纹等特征信息共同构成了对于一个人年龄判断的依据,但人们仿佛很难言明其中的微妙关系,且常常只能近似估计人的年龄。

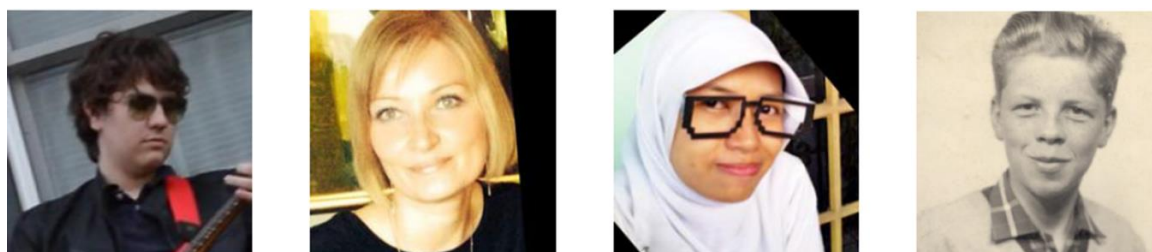


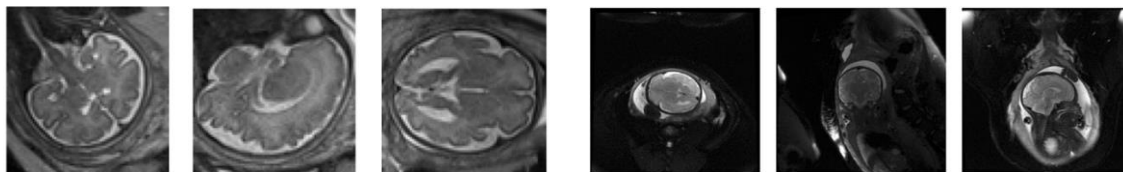
图 2-10 用于年龄预测的人脸图像例图^{[91]4}

Levi 等人^[92]仅使用了一个含三层卷积层、两层全连接层的深度卷积神经网络就取得了超越基于支持向量机的人脸年龄预测的性能。Rasmus 等人^[91]将人脸年龄回归问题转化为一个深度分类问题,通过在 VGG-16 的最后一层全连接层末尾接入 softmax 层,得到输入人脸图像分类为不同年龄(0~100 岁)的概率,网络模型最终的年龄预测值是此概率向量与离散年龄向量的内积结果。此基于 CNN 的人脸年龄回归预测模型取得了 2015 年 ChaLearn LAP^[93]挑战中年龄预测竞赛的第一名,并超出人类的预测水平。

又如基于核磁共振成像（magnetic resonance imaging, MRI）的年龄预测任务。Jonsson 等人^[94]使用具有残差、跳连结构的三维卷积神经网络对成人大脑 MRI 的四种不同图像样式进行并行的特征提取，并结合四条分支的预测结果给出更准确的大脑年龄预测，且其预测精度高于其他机器学习方法。

如图 2-11 中子图(a)所示，胎儿大脑图像中几乎没有与胎儿年龄直接显式相关的信息，人们只能通过某些组织生长的大小或形状信息来大致评估其生长阶段。Liao 等人^[95]构建了一个基于多分支的可变形卷积神经网络的回归网络模型，通过融合多个视角 MRI 成像的特征信息，对胎儿脑龄进行预测的平均绝对值误差低至 0.75 周。

如图 2-11 中子图(b)所示，孕妇子宫 MRI 并无显式信息与孕龄挂钩，然而深度学习有潜力挖掘其中与孕龄有关的特征信息。Shen 等人^[96]设计了一个注意力机制引导的双分支残差神经网络以解析子宫 MRI 图像中的复杂特征信息，实现了高精度的孕龄检测。



(a) 胎儿 MRI^{[95]3}

(b) 孕妇子宫 MRI^{[96]7}

图 2-11 基于 MRI 的年龄预测

综上，不难发现，即使只拥有关于图像全局语义信息的弱标签情况下，深度卷积神经网络也能尝试学习图像数据中复杂信息与弱标签信息之间抽象的、深层的表示关系。

2.4 本章小结

本章围绕本文所使用的核心技术（深度学习）及所处理的核心任务（基于图像的回归/计数任务），简要介绍了相关工作及基础技术研究，展示了即使在使用弱标签作为监督信号的情况下，卷积神经网络依然具有强大的特征提取能力和对复杂表示关系的学习能力。

第3章 数据集的建立

3.1 毛发镜图像数据集信息

3.1.1 毛发镜图像的收集与标记

本文实验中所使用的毛发镜图像数据由广东省揭阳市人民医院健康管理中心提供。头皮毛发显微图像由一款可移动的智能毛发和头皮分析设备（Hair and Scalp Analysis Machine, Salana Beauty Technology Co., Ltd）拍摄获得。该设备的检测仪部分装载有最高可支持一千两百万像素的摄像头，通过 USB 串口将设备与计算机连接，操作人员可以将捕获的毛发镜图像（480×640 像素大小）保存到本地硬盘中，以备后续的标定与分析工作。临床实际中，医生会对患者四个不同头皮部分（顶部，枕部，左颞部，右颞部）中感兴趣的区域（area of interest, AOI）分别拍摄若干图像。

同时，该系统配备分析软件辅助临床医生分析患者的毛发、头皮健康情况。软件整体的功能和操作页面如图 3-1 所示。

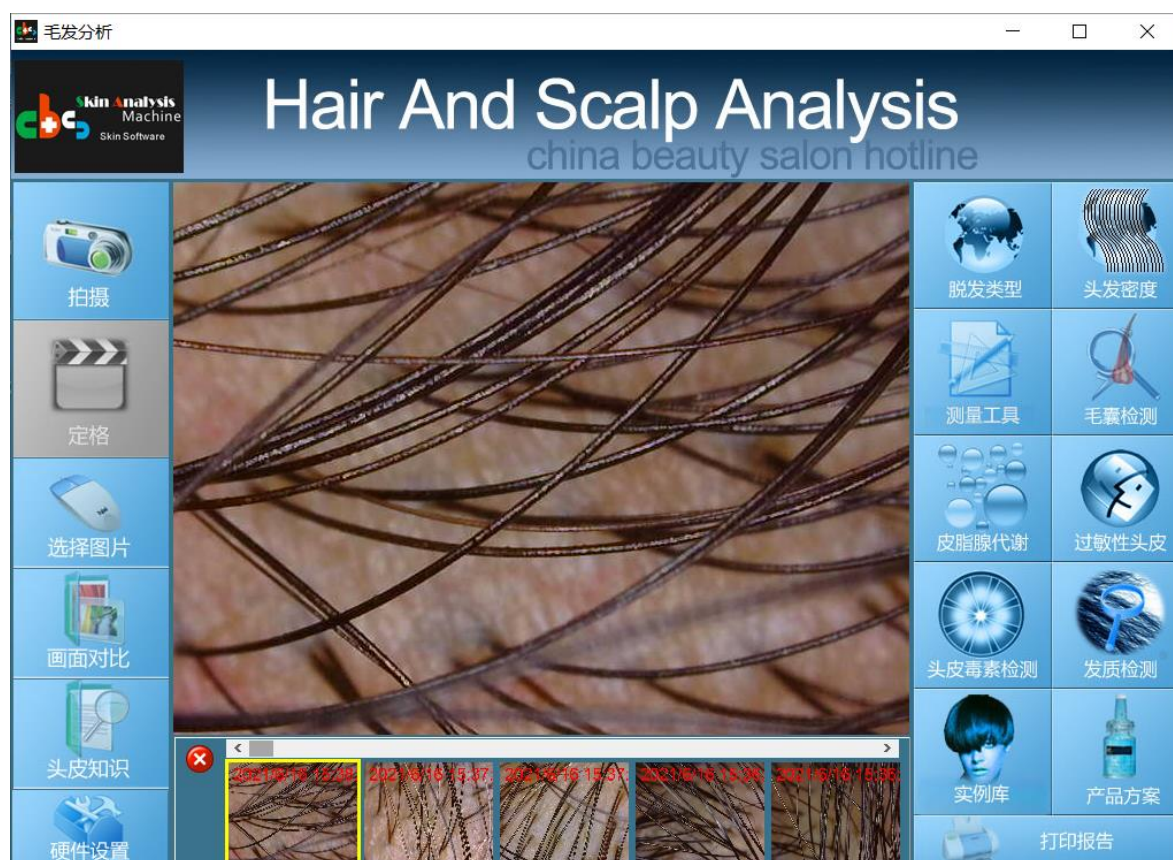


图 3-1 Hair and Scalp Analysis 软件界面

皮肤科临床医生使用上述设备采集了门诊患者的头皮毛发镜图像，采集时间跨度从 2021 年 6 月 23 日到 2021 年 11 月 20 日，收集了共计 781 张（存在重复图像）毛发镜图像。通过使用软件界面右上方的“毛发密度”功能子块，皮肤科医生可以对毛发镜图像中的毛发根数进行手动计数。

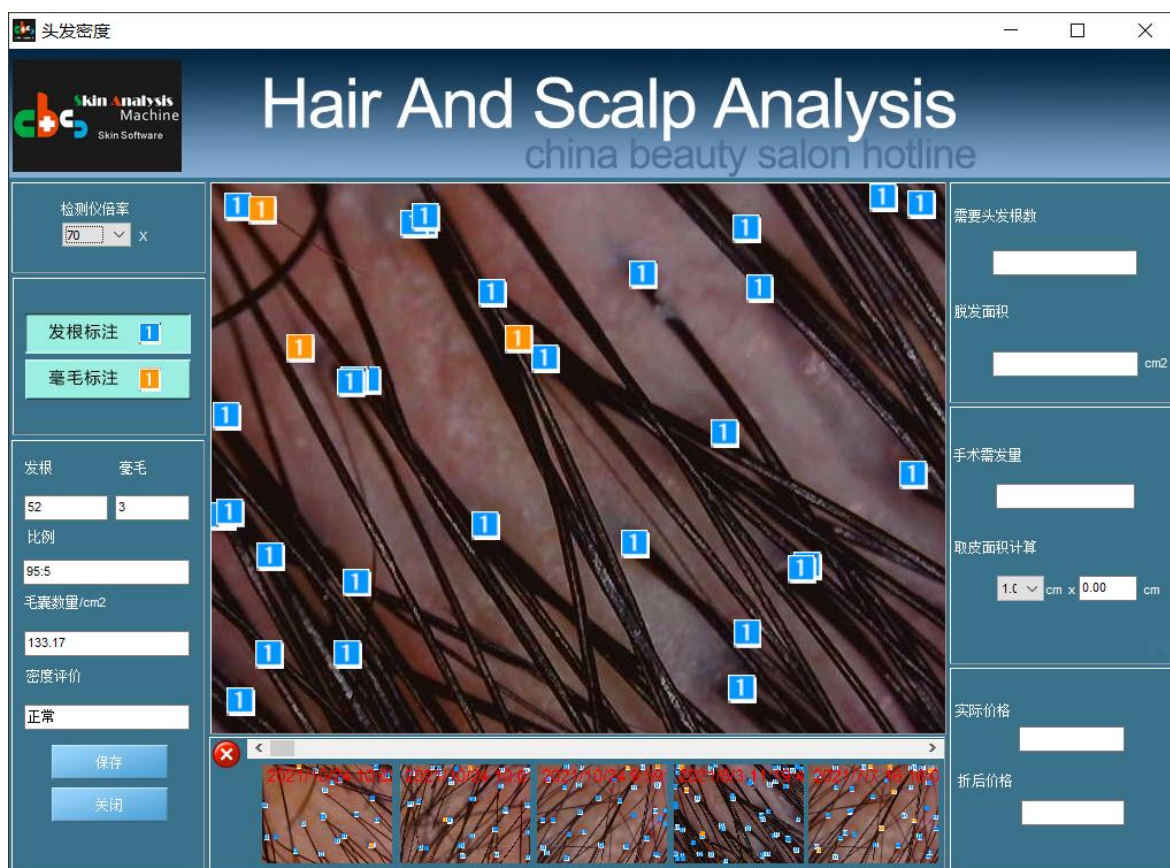


图 3-2 医生通过鼠标点击的方式区分和计数毛发

如图 3-2 所示，医生首先匹配好检测仪拍摄毛发图像时对应的放大倍率（一般 70 倍），再选择选择不同的标注样式对发根（对应图中蓝色填充标记的“1”）和毫毛（对应图中橙色填充标记的“1”）进行标注。在医生手动进行标注的过程中，软件会自动记录不同类型毛囊的数目，并将计数结果显示在软件界面左边的“发根”、“毫毛”数据栏处。同时，软件会根据实际取皮面积自动计算出毛囊密度，并给出该头皮区域内的毛发密度评价^①。如上图，根据医生的标记，该毛发镜图像中有 52 个终毛毛囊单位和 3 根毳毛毛囊单位，共计 55 个毛囊单位，在 41.30 mm²^②取皮面积上，毛囊密度为 133.17 根/cm²，属于正常密度范围。

值得注意的是，由于绝大多数毛发镜图像中普遍存在毛发长度较长、毛发重叠和

^① 软件中给出的参考毛发密度评价范围有：非常低（低于 100 根/cm²），低（100~125 根/cm²），正常（125~150 根/cm²），以及高（高于 150 根/cm²）。

^② 本数据集中的毛发镜图像，其取皮面积均为 41.30 mm²。

遮挡等不利于人工计数的复杂情况，因此此类毛发镜图像的标定任务要求标注人员具备一定的专业知识和临床经验。本文所使用的毛发镜图像数据集的计数标签均由两名专业的皮肤科门诊医生借助上述软件标定给出，笔者整理并挑选了其中 400 张（总计 20243 个毛囊单位）带有计数标签的的毛发镜图像作为数据集。

3.1.2 毛发镜图像数据集的统计信息

毛发镜图像数据集中包含了门诊患者四个不同头皮部位、不同毛囊密度等级的 400 张图像，其对应的图像数据量如表 3-1 所示。

表 3-1 不同头皮部位、不同密度等级的毛发镜图像数量

密度 部位	非常低	低	正常	高	总计
顶部	4	21	53	25	103
枕部	6	16	60	16	98
左颞部	35	35	28	2	100
右颞部	29	37	33	0	99
总计	74	109	174	43	400

统计数据集图像数量的分布情况后，不难发现毛发镜图像中毛囊单位数量相对集中在“低”和“正常”两个密度等级区间。更进一步地，本文统计了数据集中毛囊单位计数标签数据的频次分布图如图 3-3 所示。

从图 3-3 可以观测到，毛囊单位的计数标签分布呈现出“中间高”、“两头矮”的特点，这是典型的数据不平衡。本文数据集中存在的数据不平衡问题一方面是由于训练样本有限，无法完全模拟真实世界中的毛囊数据分布；最根本的原因或许在于真实世界中毛囊数量的分布本身就是不平衡的。现有的处理不平衡数据或长尾分布数据的方法大多针对分类问题，而关于解决不平衡回归计数问题的研究偏少。

在机器学习中，当数据分布不平衡时，训练模型便更容易产生过拟合现象。这是因为模型在训练阶段对那些在数据集中占比高的数据样本进行了更充分的学习，而“轻视”了少样本的学习。具体而言，模型在训练批次中，针对高频率样本的训练误差可以越来越低，而在低频率样本上的训练误差很可能会居高不下。随着训练的迭代，尽管总体或平均的训练误差会降到很低，但模型在预测新样本阶段很可能对低频率样本完全失效。因此，数据分布的不平衡分布是本研究的难点之一。后文实验结果分析部分将对此议题作进一步的讨论。

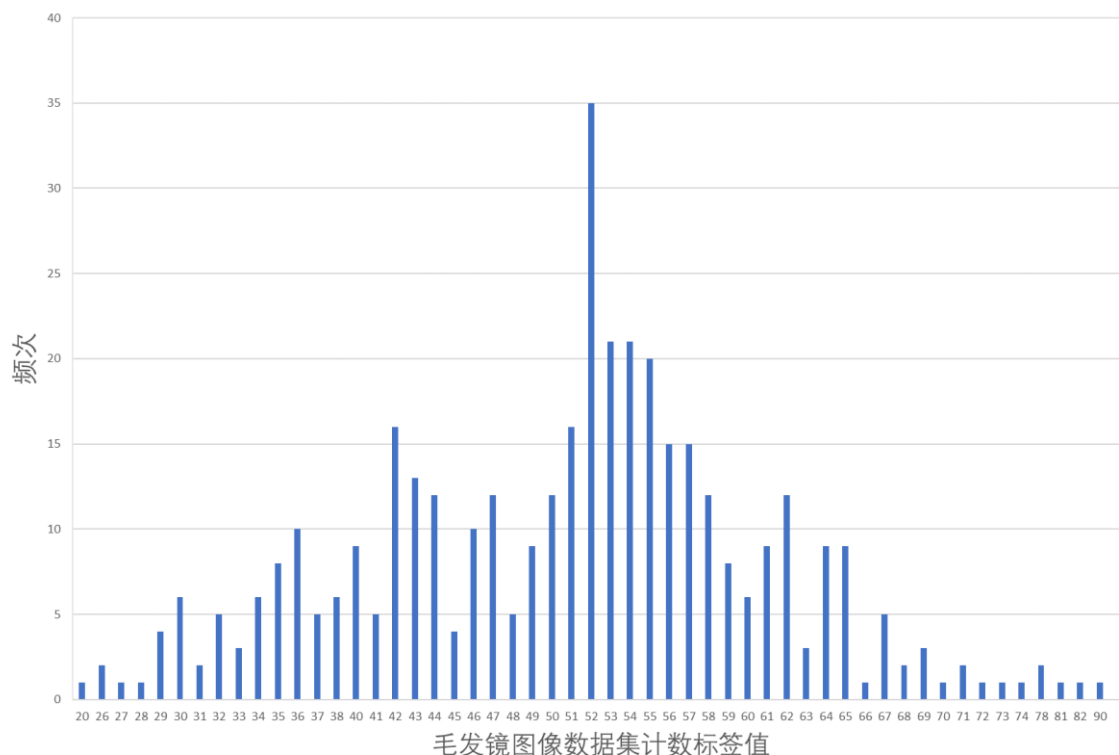


图 3-3 毛发镜图像数据集中不同毛囊单位计数数值标签的数据量分布

3.2 训练集和验证集的划分

近些年来，深度学习之所以发展迅速，一个重要原因在于如今许多任务通常可以获取海量数据来对神经网络进行训练。在现实世界中采样越多数据，采样数据的分布就越接近真实的数据分布本身。理想情况下，测试集与训练集应有相同或尽可能相近的数据分布。而在数据量不充足的情况下，训练集数据与测试集数据之间极大可能拥有不同的分布，其间的分布偏移（也称协变量偏移）是影响模型无法获得良好泛化性能的关键所在。本文毛发镜图像数据集仅有 400 张图像，因此在划分训练集和验证集^①时，使得它们的数据分布尽可能一致便更为重要了。

本文随机采样了毛发镜图像数据集中的图像，并按 4:1 的比例划分出训练集（320 张，平均一张毛发镜图像含有 50.61 个毛囊单位）和验证集（80 张，平均一张毛发镜图像含有 51.31 个毛囊单位），它们各自的统计分布信息如下页图 3-4、图 3-5 所示。

使用 IBM SPSS Statistics 25 对训练集和验证集的毛发计数标签数据进行两独立样本柯尔莫戈洛夫-斯米诺夫（Kolmogorov-Smirnov, K-S）^[97]检验。两个分布之间的最大距离 Z 值为 0.725，渐进显著性（双尾）p 值为 0.669（具体检验结果见附录 E），可以认为训练集和验证集来自同一分布。

^① 本文中不设置测试集，通过检验训练模型在验证集上的泛化能力来模拟实际测试时的情况。

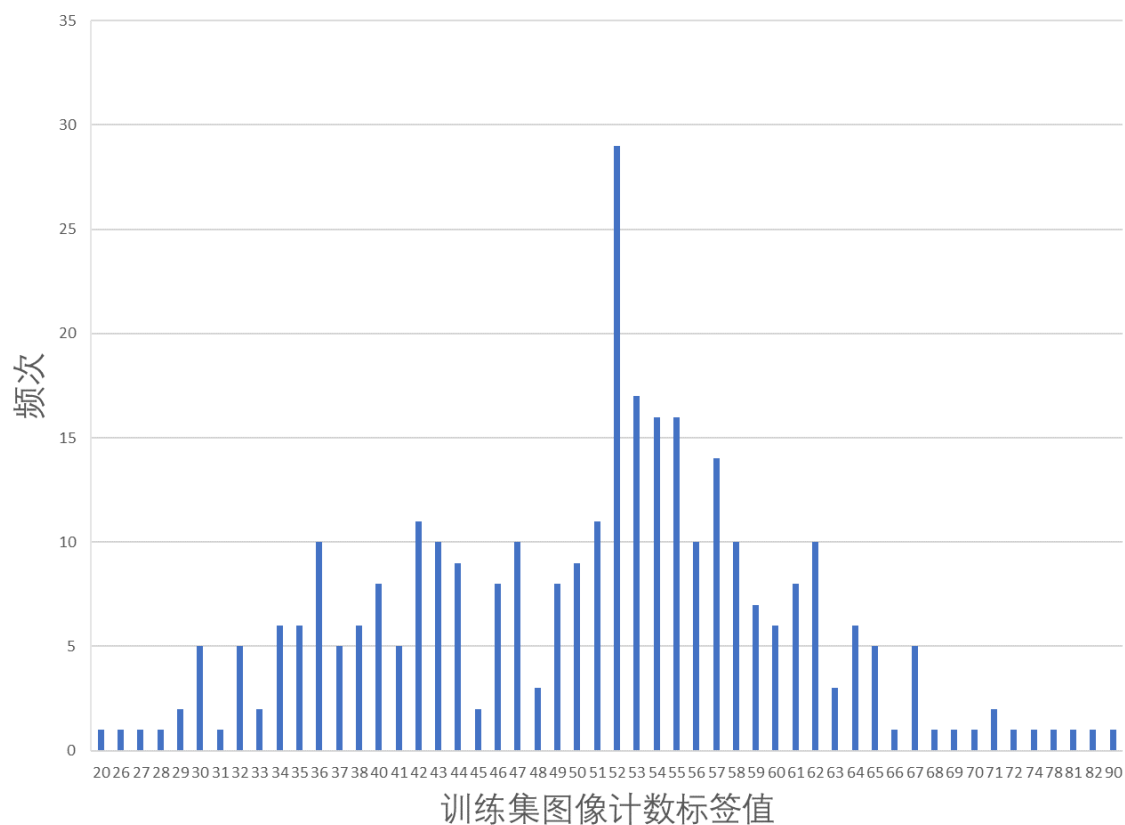


图 3-4 训练集中不同毛囊计数标签值的数据量分布

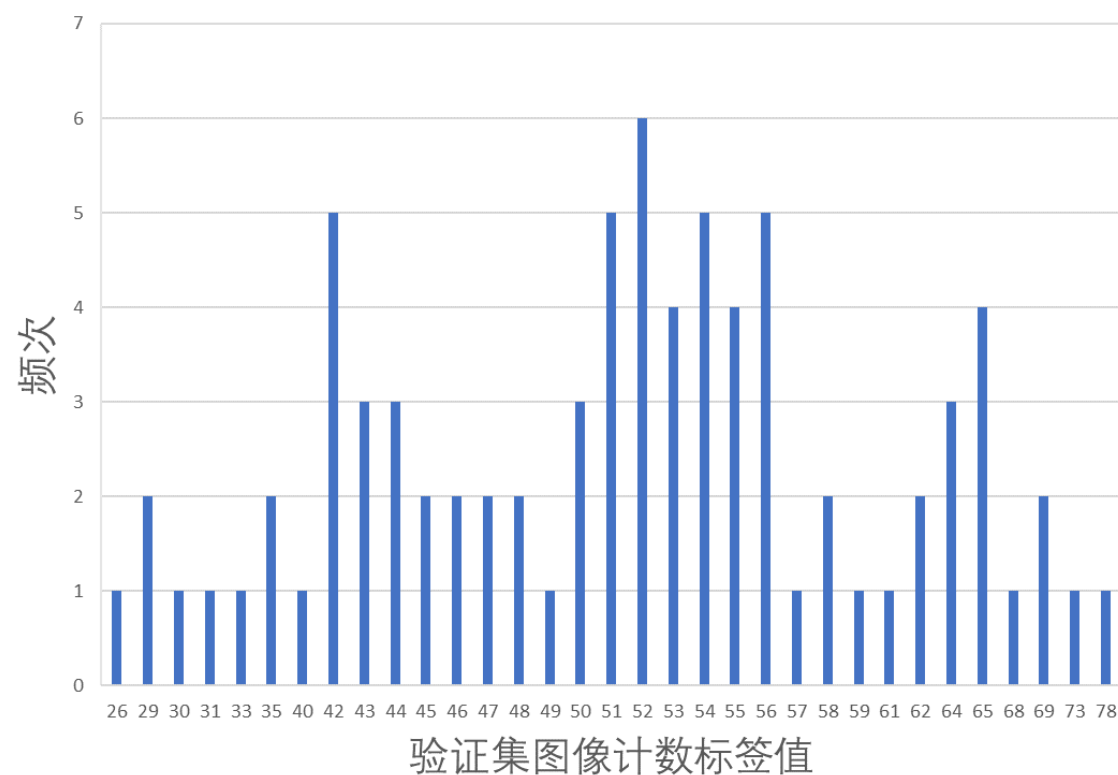


图 3-5 验证集中不同毛囊计数标签值的数据量分布

第4章 基于深度学习的毛发镜图像毛囊单位计数模型

4.1 融合深度监督信息的卷积神经网络架构

由于本文毛发镜图像数据集仅有图像中毛囊单位的总计数这一偏弱的标签信息，同时缺少关于毛囊单位具体空间位置信息以及类别信息等更细粒度的标签信息，因此基于毛囊位置检测的方法无法适用于本研究。与此同时，我们在第2章相关工作调研部分发现即使是仅有弱标签信息的情况下，深度卷积神经网络依然展示了其强大的特征提取能力。鉴于此，本文致力于借助深度卷积神经网络来构造一个端到端的回归计数模型，通过网络结构上的正则化以平衡回归模型对部分预测结果的“过度自信”，从而增强回归模型对毛囊单位分布的学习能力。

受集成学习^[47]及深度监督网络^[98]启发，本文构建的卷积神经网络架构在扩展 ResNet18 的基础上添加了深度监督信息辅助分支，网络架构如图 4-1 所示。

其中，骨干网络 ResNet18 包含 4 个由残差块组成的模块，每个模块如图中蓝色虚线框内所示具有 4 个 3×3 卷积层；因其加上第一个 7×7 卷积层和最后一个全连接层，共 18 层，故称之为 ResNet18。ResNet18 具有 F1~F5 五个不同尺度空间的特征图信息。为引入结构正则化，本文首先在原始 ResNet18 的最后一个残差模块尾部(图中特征图 F5)与全局平均汇聚层之间额外添加了一个包含残差模块的基本卷积模块 (basic block)^[74] (图中蓝色虚线框框出的部分)，从而获得关于毛发镜图像在六种不同尺度空间^①上的特征图。然后在第四个特征层输出尾部 ((图中特征图 F4)) 添加一个辅助预测分支 (图中红色虚线框框出的部分，包含一个 basic block 以及用于输出回归预测结果的解码器部分)，从而在深度特征层水平上实现对弱标签信号的重复利用。回归模型的最终输出是对两个分支预测结果的均衡加权。

记标量 $f_{aux}(x_i)$ 为辅助任务分支对每张毛发镜图像 $x_i \in \mathcal{X}$ 的预测输出，标量 $f_{main}(x_i)$ 为为主干任务分支的预测输出，则回归网络模型总的预测输出：

$$y_i = f(x_i) = \frac{1}{2}(f_{main}(x_i) + f_{aux}(x_i)) \quad (3-1)$$

本文选用 ResNet18 作为主干网络，并在其基础上增加上述额外的卷积神经网络部件及分支，主要基于以下考量：

(一) 选择 ResNet18 作为骨干网络的原因

我们首先检验了 VGG 系列^[68]、AlexNet^[61]、Inception 系列^[61]、ResNet 系列^[74]

^① 每经过一次下采样操作，就会得到一个新的尺度空间。

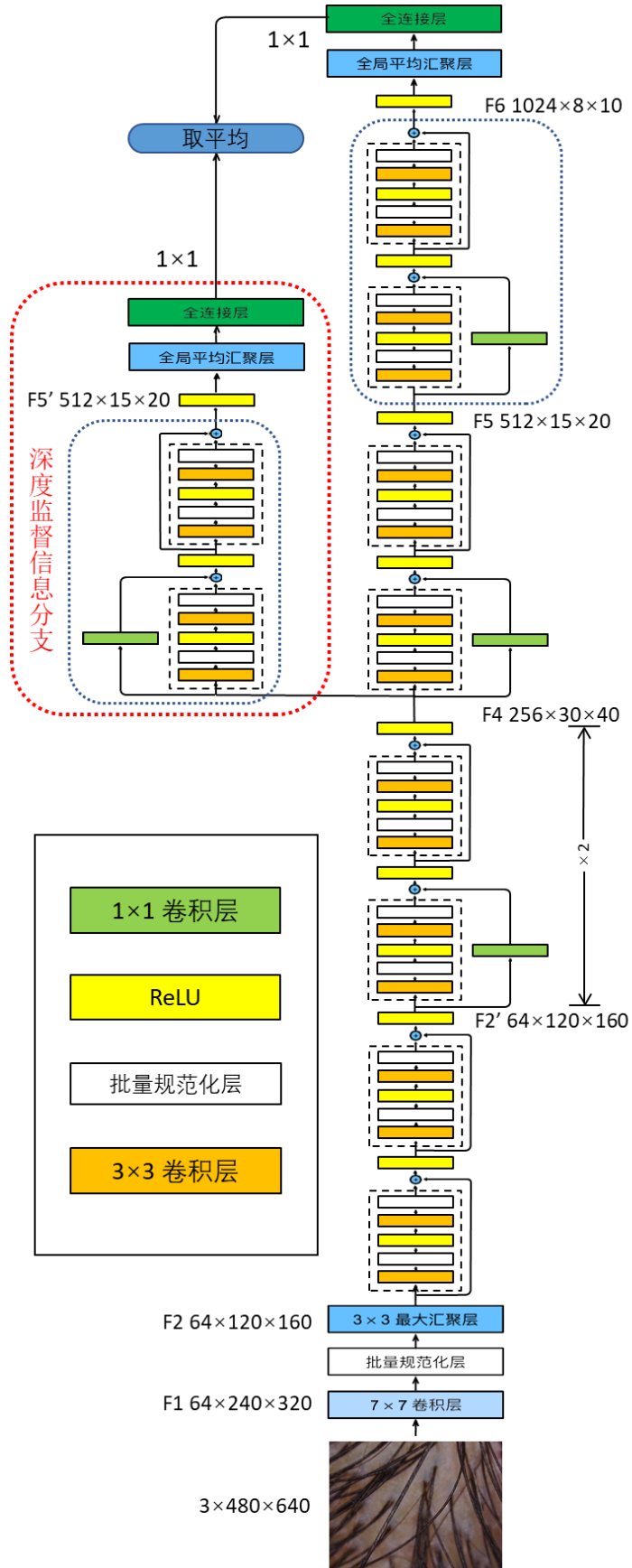


图 4-1 融合深度监督信息分支的毛囊单位回归计数模型

经典神经网络架构模型在毛发镜图像毛囊单位回归计数任务中的潜力，如实验 5.4.1 节所示。基于实验结果分析，本文发现具有残差、跳连结构的 ResNet 系列网络模型具有较优的学习能力，而同时在 ResNet 系列网络中具有更少模型参数量的 ResNet18 取得了更好的泛化性能。故将具有 5 个不同尺度空间特征信息的原始 ResNet18 作为本文回归网络模型的骨干网络。

（二）设置辅助预测分支的动机

①最大化利用弱标签信息。一个比较符合直觉的认识是，通过将弱标签信息在不同尺度空间上进行复用，并融合该深度监督信息到主干网络，相当于在回归模型中加入了“自我纠错”机制。一方面，不同毛发镜图像中毛囊单位分布在不同尺度特征空间上的学习难度可能是不同的，通过基于多尺度特征空间预测的方式很有可能使预测分支产生互补的机制，从而提高对整体样本的回归预测性能；另一方面，基于某个特征尺度空间的回归预测结果若出现较大误差，则它会被另一个预测分支所平滑。在这种相互“钳制”的作用下，模型的过拟合会被缓解，从而提高模型的泛化性能。集成学习 (ensemble learning) 通过构建并结合多个学习器来完成学习任务。在本文所提的融合深度监督信息的卷积神经网络架构中，两个基学习器（主干分支与辅助分支均系卷积神经网络）分别基于不同尺度空间的特征图信息来完成各自回归计数，从而实现差异化或多样性的预测结果^{[47]171~173}。尽管这两个基学习器不满足基学习器的误差相互独立的假设，但根据上文的分析可知，它们是有可能“好而不同”的。

②实现回归网络模型的结构正则化。

可以将辅助分支的回归预测与主干分支的回归预测视作一种特殊的多任务学习范式，只不过这里主任务与辅助任务的目标是一致的，即毛囊单位回归计数。通过共享骨干网络中前面四层尺度空间的特征信息，两个预测分支分别基于不同尺度空间的特征信息进行回归预测，每个预测分支都可以充当彼此的正则化项^[99]。从模型参数更新的角度来说，每个预测分支都有其归纳偏置 (inductive bias)，从而在更新前面四层较低层的特征层参数时能共享互补信息；从损失函数的角度来说，如公式(4-1)所示，总的预测输出总是这两个分支的加权叠加，因此经验误差也总包含这两者，这显式地表明了它们互为正则的关系。

（三）选择在深层特征层上添加辅助预测分支的原因

实验证明了具有五层不同尺度特征图信息的 ResNet18 能够在一定程度上对毛囊计数任务进行学习。考虑到残差、跳连结构具有的恒等映射特点，它使得前向传播中的特征信息和后向传播的梯度信息可以直接从一个 block 传递到任意一个 block^[75]，我们在 ResNet18 骨干网络的基础上额外增加了一个带有残差、跳连结构的 basic block，这在保证预测模型不出现模型退化的同时，获得了基于更深层特征信息的回归预测。

一方面，底层卷积特征信息中含有大量噪声，直接在低层特征层后直接连接分类器或回归器等解码器，模型无法取得良好的性能^[100]。同时，一般认为深层特征层具有更大的感受野，包含了更高级的语义特征，更适合解码特征图信息到输出的映射。因此，我们在第四个尺度空间的特征层（深层特征）输出端接上一个辅助预测分支以重建（一）中具有五层尺度空间特征信息的原始 ResNet18，并用作回归预测模型的辅助分支。由于辅助预测分支是基于深层特征信息进行回归预测，且其预测结果最终会联合主干分支的预测结果作均衡加权。因此，本文将这种引入辅助预测分支以实现弱标签信息进行复用的方式称为融合深度监督信息的学习方式。

4.2 损失函数设计

毛发镜图像毛囊单位计数是一项数值回归任务，我们沿用在数值回归任务中普遍使用的均方误差损失（mean square error, MSE）函数，其计算公式如下：

$$L_{mse}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3-2)$$

其中， n 为当前训练批次的样本总数， y_i 为第 i 个毛发镜图像样本对应的计数标签， $\hat{y}_i$ 为回归模型对第 i 个样本的预测结果。

受 Focal Loss^[101]启发，针对毛囊单位计数标签对应毛发镜图像样本存在数据不平衡的特点，本文在训练阶段对数据集中的低频样本（计数标签值对应的图像数据量少）分配更大的权重，同时减少高频样本的权重，从而调整后的整体损失函数计算公式如下：

$$L_{mod}(y, \hat{y}) = \sum_i^n w_i \cdot (y_i - \bar{y})^2 \cdot \frac{1}{n} (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_i^n w_i (y_i - \bar{y})^2 (y_i - \hat{y}_i)^2 \quad (3-3)$$

其中， $\bar{y}$ 为毛囊计数标签的平均值， w_i 为第 i 个计数标签样本在标签数据集中出现频率的倒数^[63]，有：

$$w_i = \frac{1}{freq_i} = \frac{N}{num_i} \quad (3-4)$$

其中 $freq_i$ 为第 i 个计数标签在标签数据中出现的频率， num_i 为第 i 个计数标签在标签数据中出现的频次， N 为数据集容量大小。

如图3-3所示，计数标签所对应的样本在标签值的极大值和极小值附近分布稀疏，而较多的计数标签样本相对集中在标签值的均值（约为 50.61）附近，因此本文通过在公式(4-3)的 $(y_i - \bar{y})^2$ 项迫使模型在训练阶段更多地关注那些计数标签与标签均值偏移较大的毛发镜图像样本。

4.3 数据增强技术

为了提高模型的鲁棒性,并显著减少过拟合,在训练过程中使用了数据增强技术,这是一种广泛应用的技术,在这种技术中,训练数据以小的、随机的方式进行扰动变化,从而提高数据样本的多样性,以迫使模型在一定噪声扰动的条件下依然做出正确的预测。我们对训练集图像进行包括水平、垂直翻转,颜色抖动(包括亮度,对比度,饱和度,色调四个参数)等常规数据增强。

特别地,为了模拟毛发镜图像数据集中普遍存在的毛发遮挡模式(包括毛发间的交叉遮挡、头皮屑等异物遮挡),本文引入了 Cutout^[102]的数据增强技术。结合常规数据增强,图 4-2 展示了一次训练批次中训练集数据进行数据增强的模式。



图 4-2 训练集中进行数据增强的模式

Cutout 是指在训练阶段,在将输入图像传入神经网络之前随机地裁剪图像中某一方块大小的像素区域。具体而言,通过对图像中随机位置区域附加一个固定大小的零

掩码,从而遮挡了其背后原有的像素,如上图被白色方块遮挡的图像所示。这种数据增强方式可以看作是 Dropout^[103]的一种变种。经典的 Dropout 会随机丢弃上一层神经网络中部分神经元的输出,而 Cutout 在图像数据层面随机丢弃一些图像像素,从而迫使模型更多地考虑全局的图像特征,通过捕捉空间中的上下文信息来弥补局部细节信息的丢失^{[102]3}。因此这可以看作是在图像数据层面对模型进行正则化。

注意到,在 Cutout 原文中,针对 CIFAR-10、CIFAR-100 数据集,作者分别寻找到了 16×16 、 8×8 不同的遮挡像素区域大小的数据增强方案^{[102]5}。同时,他们认为对于更细粒度的检测任务(CIFAR-100)来说,图像的上下文信息对判别类别的用处会变小,而那些更微妙的细节信息用处则更大,因此此时更合适选择较小的遮挡区域面积^{[102]5}。

为了找到适合本文毛发镜图像数据集的 Cutout 策略,本文在实验部分对裁剪的像素区域大小进行网格搜索。由于毛发镜图像数据集中常常存在多处遮挡,则在数据增强阶段裁剪的像素区域个数也可以是多变的。因此,与原文工作不同的是,我们在训练阶段执行裁剪像素区域时,对每张毛发镜图像中随机地生成 1~10 个裁剪区域,进一步丰富了数据增强的多样性。

4.4 标签平滑技术

注意到,毛发镜图像数据集的标签数据存在不平衡分布,这会极易造成深度卷积神经网络对低频率样本数据学习能力欠佳;同时,由于计数标签连续区间中存在部分缺失值,重建一个有效的、平滑的标签数据分布对缓解模型过拟合有重要作用。

使用标签平滑的好处在于:

(1) 通过使用标签平滑(一般来说会对标签的数值区间进行缩放,使得标签数据变得“扁平”,让稀疏的标签数据趋于集中),标签数据的数值分布变得更加平滑,特征空间也随之平滑化。数值变动的平滑化使得网络模型在训练阶段不至于对太大或太小的极端值过于敏感,因此能够习得更平滑的数据特征分布。同时相对较窄的数值区间更利于回归模型对连续区间中缺失的部分标签数据进行感知和学习^[104]。同时,更小数值区间(更聚拢)的标签数值区间鼓励模型参数也在小范围内收缩(shrinkage),因此标签平滑相当于一个隐式的正则技术,有助于缓解过拟合现象。

(2) 加速网络模型的收敛,使训练过程稳定。另外,有学者^[105]认为标签平滑技术甚至帮助降低训练标签中的噪声。

在分类任务中,标签平滑正则化(label smoothing regularization, LSR)在 Inceptionv_3^[106]中被首先提出。如公式(4-5)、(4-6)所示,通过加权平均硬标签^①与一个

^① Hard label, 也称 hard target, 即原始数值标签。

均匀分布的噪声，使得原始的、稀疏的 one-hot 编码变为带有噪声项的形式，从而避免了模型对预测的过度自信。从数值上看，软标签的小数分布较硬标签的 0、1 分布更紧凑、更平滑。

$$q'(k|x) = (1-\epsilon)\delta_{k,y} + \frac{\epsilon}{K} \quad (3-5)$$

其中， $q'(k|x)$ 为训练样本 x 的标签 $k \in \{1 \dots K\}$ 被重建后的标签分布， ϵ 为标签平滑因子（原文中经验性地将其设置为 0.1）， $\delta_{k,y}$ 为硬标签，有

$$\delta_{k,y} = q(k|x) = \begin{cases} 1 & \text{if } k = y \\ 0 & \text{otherwise} \end{cases} \quad (3-6)$$

Mixup^[107]数据增强从某种程度上来说也支持了标签平滑的理论。在模型的训练阶段，两幅图像被混合到同一幅图像中，它们相应的 one-hot 编码也被重加权，从而得到新图的标签编码表示^[108]。

在回归任务中，常用的标签平滑策略有归一化和取对数。如深度学习中广泛存在的图像输入数据归一化、层归一化等，数值标签也可以进行归一化。对标签数据取对数，常用在房价预测^[109,110] 等具有较大的硬标签数值差距的回归预测任务上。

值得注意的是，标签平滑仿佛并不总是有效的。本文在实验部分检验了三种常用的标签平滑方式对回归网络模型性能的影响，包括标签数据归一化，取对数，以及平滑因子 $\epsilon = 0.1$ （即将原始标签数据的数值乘以 0.1）。

4.5 本章小结

本章提出了一个端到端的、基于深度卷积神经网络的回归网络模型以实现对毛发镜图像中毛囊单位数量的回归预测。

针对仅有毛囊单位整体计数信息而无更细粒度的监督信息的特点，本文在网络架构中引入一种结构正则化的机制。通过在深层特征层中引出一条辅助预测分支，在对弱标签信息进行复用的同时，融合预测分支在多个特征尺度空间上对毛囊单位进行回归计数的结果，进而提高回归模型的稳定性。

针对毛发镜图像数据集中存在的不利于识别毛囊单位的特点以及存在复杂的交叉、遮挡模式的难点，本文提出了一系列图像数据正则化的机制。通过翻转图像、引入颜色扰动增强训练数据集的多样性；通过裁剪输入图像部分像素的方式迫使回归模型利用空间上下文信息。

针对计数标签分布不平衡的特点，本文提出了重加权的损失函数，以提高毛发图像数据集中低频计数标签样本在训练阶段的经验误差。

针对计数标签分布稀疏的特点，本文提出了关于标签正则化的一个新机制。通过

对标签数据数值进行平滑,重建一个更适合回归学习的计数标签分布,以降低回归模型对原始标签分布空间中的极端值的敏感度,同时使回归模型对于连续标签数值区间中的缺失样本具有一定的感知、预测能力。

第5章 实验结果与分析

5.1 实验环境及参数设置

5.1.1 实验环境

本文实验使用深度学习框架 Pytorch 1.8.1^[111]在 GPU 集群 Linux 服务器上进行实验。服务器系统版本为 Ubuntu16.04.5 LTS, CPU 型号为 Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz, 配备四张显存 12GB 的 NVIDIA TIAN Xp 型号的 GPU。使用 IBM SPSS Statistics 25 对实验数据进行统计分析。

5.1.2 参数设置

在实验参数方面,模型的最大训练迭代代数设为 500 (在前期骨干网络搜索阶段为 200 代)。初始学习率为 0.0001 (加载预训练模型参数时,预训练模型的卷积层、BN 层参数拥有较低的初始学习率 0.0001;用于构建深度监督信息分支的各层,其参数拥有较大的初始学习率为 0.001)。学习率更新方式选择余弦退火衰减方式^[112],衰退周期 T_max 设置最大训练迭代代数,未设置重启动。模型优化器选择 Adam^[113],权重衰退系数 weight_decay 设置为 0.001。

图像预处理阶段,本文保持原始输出图像 480×640 分辨率(若缩放会导致模型精度的下降),并对图像各通道像素进行归一化,归一化参数 mean=[0.367, 0.281, 0.269], std=[0.145, 0.138, 0.143]。

数据增强阶段,我们除了执行常见的一些增强方式,如水平方向图像翻转、垂直方向图像翻转、颜色抖动^①,还添加了适应本文毛发镜图像数据集的 Cutout 增强^②。每种数据增强方式的执行概率均为 0.5。

5.2 相关评价指标

毛发镜图像中毛囊单位计数是一项回归任务,回归任务中常用的以下评价指标来衡量回归模型在预测阶段的预测精度或拟合程度。

(1) 平均绝对值误差 (mean absolute error, MAE)

MAE 是指预测值与真实值偏差的绝对值的平均数,其计算公式如下:

^① 亮度,对比度,饱和度,色调四个参数均设为 0.5,即上下抖动 50%的水平。

^② 毛发镜图像中随机插入 1~10 个遮挡毛发镜图像的 32*32 白色方块。

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4-1)$$

其中, n 为一个训练批次中的毛发镜图像样本数, y_i 为训练批次中的第 i 个毛发镜图像样本的毛囊单位计数标签, $\hat{y}_i$ 为模型对第 i 个毛发镜图像样本预测的毛囊单位数目。在本文任务中, MAE 反映了毛囊单位预测值与计数标签之间的平均绝对误差。

(2) 平均绝对百分比误差 (mean absolute percentage error, MAPE)

MAPE 是指预测值与真实值偏差相比于真实值的绝对值的平均数, 它在平均绝对值误差 MAE 的基础上结合了真实值的量纲, 反映了偏差的相对大小 (即相对误差), 其计算公式如下:

$$MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (4-2)$$

注意到, 当 MAE 相同的情况下, 对于不同标签值, 会得到不同的 MAPE。例如, 假设计数标签分别为 50、100, 而模型预测值分别为 40、90 的情况下, 尽管它们的 MAE 是相等的 (均为 10), 但对于 MAPE 来说, 显然前者超出了后者 ($20\% > 10\%$)。

因此, 在本文任务中, MAPE 反映了模型预测毛囊单位数目出现偏差的程度, 相比于 MAE 更能描述模型预测的可靠程度, 后文实验结果部分关于模型泛化性能的分析也主要围绕该指标进行。

(3) 决定系数 (coefficient of determination)

决定系数, 记作 R^2 , 反映了在回归模型中因变量的方差变化可由自变量的方差变化解释的比例, 用于判定回归模型的拟合程度 (goodness-of-fit), 其计算公式如下:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = \frac{SS_{reg}}{SS_{tot}} \quad (4-3)$$

其中, SS_{res} 为残差平方和, SS_{reg} 为回归平方和, SS_{tot} 为总离差平方和。 $\bar{y}$ 为因变量的样本均值, 在本文中指数计数标签样本的均值。

$$SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4-4)$$

$$SS_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (4-5)$$

$$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4-6)$$

R^2 介于 0~1 之间, 越接近 1, 模型回归拟合的效果越好。

(4) 皮尔逊相关系数 (Person correlation coefficient)

皮尔逊相关系数, 用于衡量两个变量之间的线性相关水平, 在本文回归分析中, 记作 $\rho_{y_i, \hat{y}} = r(y, \hat{y})$, 其计算公式如下:

$$r(y, \hat{y}) = \frac{\sum_i^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_i^n (y_i - \bar{y})^2 \sum_i^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (4-7)$$

在本文中， $\bar{y}$ 是计数标签样本的均值， $\bar{\hat{y}}$ 是模型对样本进行预测的均值。

5.3 实验设计

为了论证本文提出的网络模型中所采用的各个机制的有效性(包括骨干网络的选取、标签平滑、基于 Cutout 的数据增强、深度监督信息分支等)，特设计多组对比实验如下：

(一) 骨干网络的选取

本实验检验了 VGG 系列、AlexNet、Inception 系列以及 ResNet 系列深度卷积神经网络模型用于本文毛发镜图像毛囊单位回归计数任务上的潜力。我们对上述每个模型分别训练了 5 次(不加载预训练模型参数，在训练阶段仅使用常规数据增强)，观测模型参数量与回归精度的关系，从中选取适宜作为骨干网络的基础网络模型。

(二) 颜色空间信息的重要性分析

经典数字图像处理环节中，通常会将彩色图像转换到灰度空间，在减少计算量的同时便于后续的特征提取。基于实验(一)中的骨干网络，在只对训练集进行翻转增强的条件下，本文检验当输入图像为单通道灰度图像、三通道灰度图像、RGB 三通道彩色图像时，网络模型的回归性能的差异。接着，我们在训练阶段添加颜色抖动这一数据增强环节，进一步说明颜色空间信息对图像处理任务的重要性。

(三) 加载 ResNet18 的预训练模型参数

在实验(二)的基础上，我们迁移了 ResNet18 在 ImageNet^[114]数据集上的预训练模型。由于 ImageNet 数据集中的图像样式与本文毛发镜图像数据集中的图像样式存在较大差异，我们怀疑预训练模型是否依然可以有良好的特征提取能力。

(四) 检验标签数据平滑的有效性

在实验(三)的基础上，我们对标签数据作平滑处理。为了寻找相对合适的标签平滑方式，我们检验了三种常用的平滑方式在本文回归任务上的效果，分别为标签平滑因子 $\epsilon = 0.01$ ，数据标签归一化，数据标签取对数。

(五) 检验基于 Cutout 的数据增强的有效性

在实验(四)的基础上，我们通过网格搜索找到最适合本文任务的 Cutout

数据增强模式。

（六）检验深度监督信息分支的有效性

在实验（五）的基础上，我们检验了通过引入深度监督信息分支对回归模型性能的影响。

下面将对本文进行的各组实验结果进行报告及分析，并论证所提机制的有效性。

5.4 对比实验结果与分析

5.4.1 骨干网络的选取

首先对各个经典卷积神经网络在毛囊单位计数任务上的学习能力与泛化能力进行检验，每个网络均被训练五次，每次均迭代模型 200 代，具体实验结果见附录 A。根据实验数据，我们绘制了网络模型参数量与其在本文回归任务上平均的学习性能、泛化性能的关系如图 5-1、图 5-2 所示。

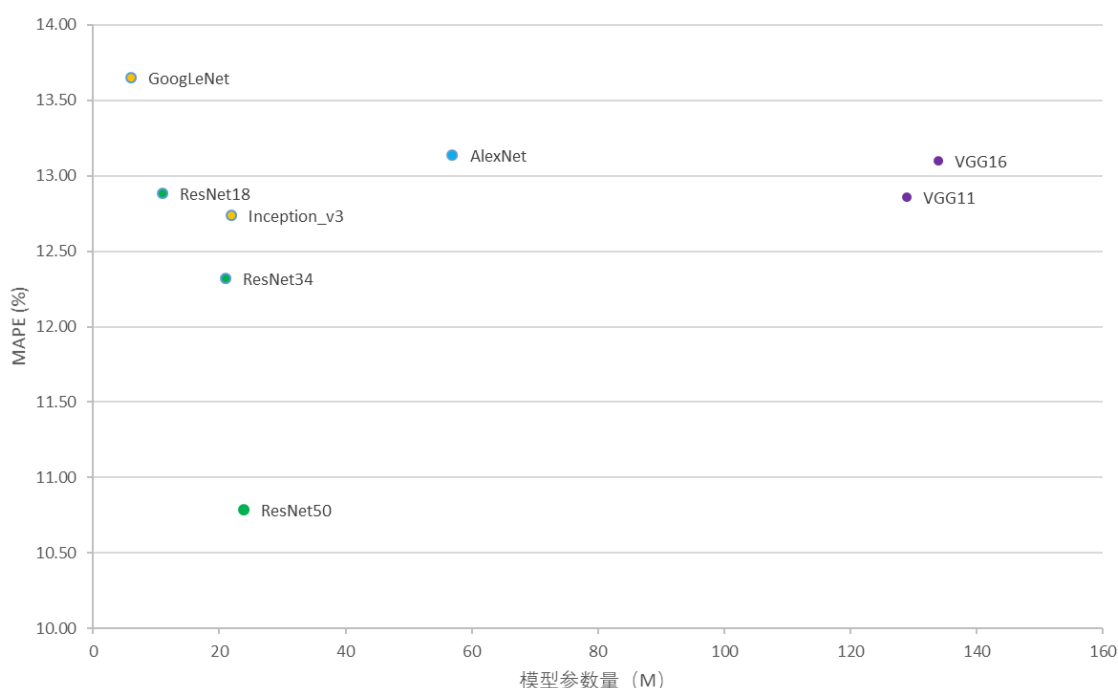


图 5-1 经典卷积神经网络在毛囊单位计数任务训练集上的学习能力

从图 5-1 可知，ResNet 系列（在图中以绿色标记）经典网络在模型参数量更少的情况下在本文毛囊单位计数任务上具有优于其他经典网络的学习能力。这符合奥卡姆剃刀原理^[115]，即对于一个数据集容量小、标签信息相对简单的任务而言，一个更轻量化的网络模型较参数量巨大的网络模型反而可能更有效。同时，对于同系列（在图中以同色标记）卷积神经网络而言，随着模型参数量的小幅提升，网络的学习能力也随之提升。

从图 5-2 可知，ResNet18 与 Inception_v3 这两个经典网络模型的泛化性能最好，且模型的参数量均偏低。综合考察了模型参数量、模型学习能力以及泛化能力指标之后，本文选取 ResNet18 作为骨干网络，并在此基础上进行扩展。

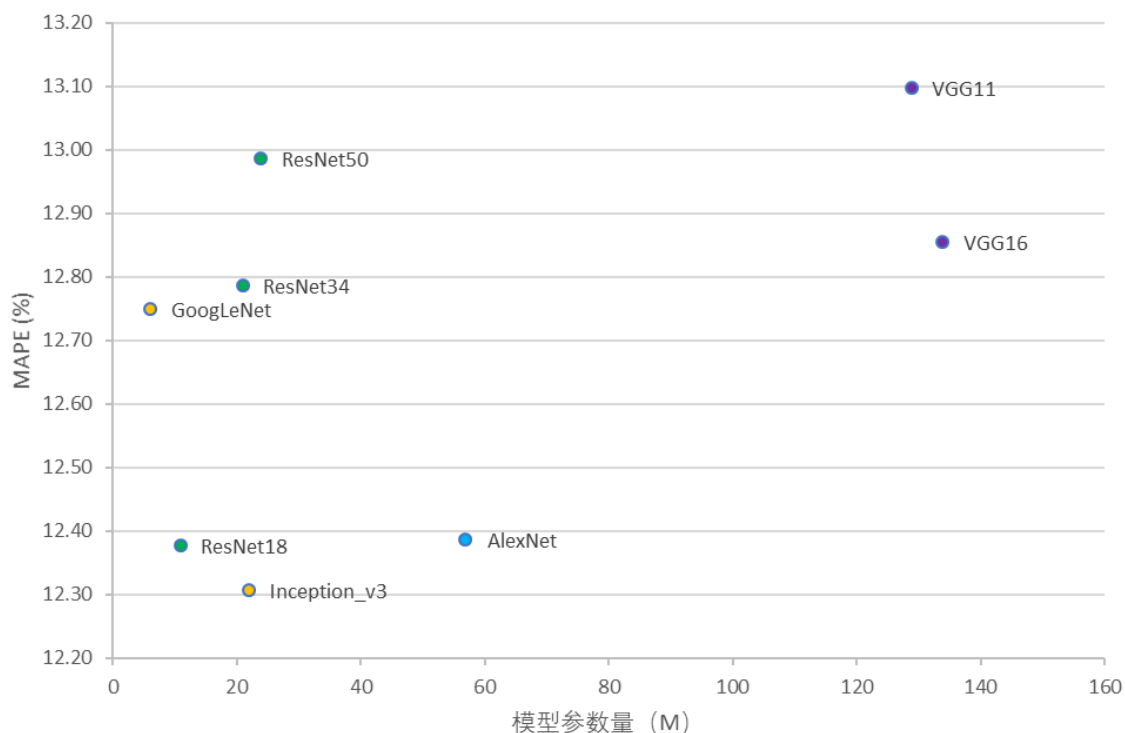


图 5-2 经典卷积神经网络在毛囊单位计数任务验证集上的泛化能力

5.4.2 颜色空间信息的重要性分析

表 5-1 改变输入图像颜色空间信息对毛发镜图像毛囊单位计数的影响

基于 ResNet18	训练集 MAPE	验证集 MAPE	最好	最差
	(五次平均)	(五次平均)		
单通道灰度图像	5.24 ± 0.33	15.37 ± 0.51	14.52	15.82
三通道灰度图像	4.99 ± 0.33	15.07 ± 0.39	14.45	15.47
RGB 彩色图像	5.99 ± 0.21	13.94 ± 0.41	13.39	14.45
RGB + 颜色抖动	12.18 ± 0.24	12.38 ± 0.18	12.16	12.60

基于毛发和头皮的颜色信息为区分它们提供重要信息的假设，在经典数字图像处理工作中，一般会在图像预处理阶段通过对比度拉伸的方式来提高前景头发像素与背景头皮像素的区分度。但在后续处理环节中将彩色图像灰度化则导致图像丢失大部分颜色空间信息，这意味着经典数字图像处理中对灰度图像进行特征提取时，其特征本身是有损的。除此之外，数字图像处理技术对光照条件敏感，常常需要手动消除图像中可以存在的亮点能颜色空间上的噪声。而对于基于深度学习的计算机视觉技术来说，

在训练阶段甚至可以对输入的 RGB 图像作颜色抖动数据增强。

为说明颜色空间信息对毛囊单位计数任务的重要性，我们做了一系列实验，结果如上表 5-1 所示。可以看到，当输入图像为单通道灰度图像或三通道灰度图像^①时，网络模型在验证集中泛化能力相当（MAPE 约为 15%），而使用 RGB 彩色图像作为输入，验证集上的 MAPE 提升了超过 1 个百分点。在对引入颜色抖动的数据增强之后，MAPE 又额外上升了 1.56 个百分点。

一个有趣的现象是，随着验证集上 MAPE 的提升，RGB 彩色图像在训练集上拟合输入图像数据的能力却在“降低”，尤其上在引入颜色抖动之后，回归网络模型在训练集上的 MAPE 从 6%的水平上升到 12.18%。除此之外，此时回归网络模型在训练集上和验证集上的回归预测精度相当，甚至在验证集上可以看到回归性能指标 MAPE 稍好于训练集！这是因为在训练阶段，我们使用了数据增强策略提高了数据样本的多样性，这使得训练集上的数据样本学习难度升高，从而在难度偏易的验证集数据样本上获得高出学习能力的泛化性能。

综上，本文认为，在毛发镜图像的毛囊单位计数任务中，颜色空间信息对回归网络模型的性能至关重要。

5.4.3 迁移预训练模型的有效性分析

表 5-2 迁移预训练模型对回归网络模型影响的实验结果

是否迁移预训练模型	训练集 MAPE (五次平均)	验证集 MAPE (五次平均)	最好	最差
否	12.18±0.24	12.38±0.18	12.16	12.60
是	3.12±0.40	12.17±0.44	11.63	12.73

ImageNet 数据集中收集了自然场景中超过一千五百万张图像数据，有超过两万类图像。一般认为，基于大容量数据集的预训练模型，其在浅层特征层中学习到了各类物体在检测、识别、分类等任务中通用的底层特征，而这些底层特征通常是图像的边缘信息。随着特征层的加深，越高层的特征将进一步地抽象图像中的语义特征。通过加载 ResNet18 在 ImageNet 上预训练模型参数，并使得迁移的这些参数在训练模型阶段具有较低的学习率（我们希望尽可能地使得底层的特征提取器在微调环节变动小），而对本文提出的具有深度监督信息分支的回归网络模型中新加入的一些组件（如卷积层、批量规范化层、全连接层）进行随机初始化，并施加更高的学习率。

本节实验结果如表 5-2 所示。可以看到，在训练同样代数（500 代）的情况下，回归模型在训练集上 MAPE 从 12.18%提升至 3.12%，这说明通过迁移预训练模型，

^① 将灰度化后的单通道灰度图像复制为三通道灰度图像。

回归网络模型的特征提取能力有所提升,拟合数据的能力增强,从而加速了模型的收敛进程。尽管这同时也使得网络模型开始出现过拟合现象,但它仍然提升了网络模型的回归预测性能(验证集上的平均 MAPE 提升了 0.21 个百分点)。

5.4.4 标签平滑有效性分析

表 5-3 不同标签平滑方式对回归网络模型性能的影响

标签平滑方式	训练集 MAPE (五次平均)	验证集 MAPE		
		(五次平均)	最好	最差
无标签平滑	3.12±0.40	12.17±0.44	11.63	12.73
平滑因子 $\epsilon = 0.1$	3.16±0.35	11.95±0.16	11.71	12.13
标签数值归一化	5.99±0.21	13.05±0.84	12.12	14.31
标签数值取对数	16.01±0.94	256.19±14.79	231.38	277.81

我们通过引入标签平滑,目的是希望对存数据分布不平衡、数据缺失的标签数据重建一个数值分布更平滑的标签空间(一般来说对数值区间进行缩放以使得稀疏的标签数据趋于集中),从而让模型对异常值不那么敏感,对那些夹在在已有数值标签连续区间之中、但缺失样本的数据有一定感知能力(从邻近标签空间中的样本去近似地学习其间缺失的样本)。

然而,从表 5-3 中可知,并不是所有标签平滑方式对回归网络模型有效。在本节实验中,我们检验了三种标签平滑的方式,只有对标签数据引入平滑因子 $\epsilon = 0.1$ 才对本文回归网络模型的预测性能有提升效果。

注意到,在本文任务中,毛发镜图像毛囊单位计数标签数据的原始数值分布为 20~90(若视作分类问题,则至少有 71 个类别),对标签分别进行平滑因子 $\epsilon = 0.1$ 、标签数值归一化、标签数值取对数后,标签数据的数值分布变为 2.0~9.0, 0.2222~1.0000, 2.9957~4.4998。笔者认为,当标签的数值区间过窄时,回归网络模型无法在数值区间范围内对特征边界进行良好的学习。即使在训练阶段,网络模型可以记住所有样本,但在预测阶段,其学到的特征分类边界很可能是无效的。如表中对标签数值取对数的实验结果,验证集上 MAPE 达到 256%,说明此时回归网络模型完全失效,模型无法对回归任务取得学习效果。

基于本节实验,本文认为,深度学习是数据驱动的科学,尤其在基于深度学习的回归任务上,对于数值标签数据,通过选取适宜的标签平滑预处理工作来改善数值分布情况是有可能对模型回归预测的性能产生积极影响的。

5.4.5 基于 Cutout 的数据增强有效性分析

为适应本文任务所使用的数据集,我们对最佳的像素裁剪区域大小进行网络搜索。

实验结果如表 5-4 所示。

表 5-4 不同的 Cutout 像素区域大小对回归网络模型性能的影响

Cutout 像素大小	训练集 MAPE (五次平均)	验证集 MAPE (五次平均)	最好	最差
无遮挡	3.20 ± 0.28	11.95 ± 0.16	11.71	12.13
4×4	1.09 ± 0.10	11.78 ± 0.42	11.25	12.38
8×8	1.15 ± 0.25	11.92 ± 0.27	11.68	12.33
16×16	1.06 ± 0.04	11.91 ± 0.42	11.30	12.62
24×24	1.16 ± 0.18	11.97 ± 0.41	11.29	12.41
32×32	1.20 ± 0.11	11.64 ± 0.08	11.52	11.73
40×40	1.07 ± 0.11	11.94 ± 0.21	11.57	12.18
48×48	1.46 ± 0.19	12.04 ± 0.47	11.45	12.78

根据上表数据,可得裁去毛发镜图像中不同像素大小图像片块对回归网络模型泛化性能的影响如图 5-3 所示。

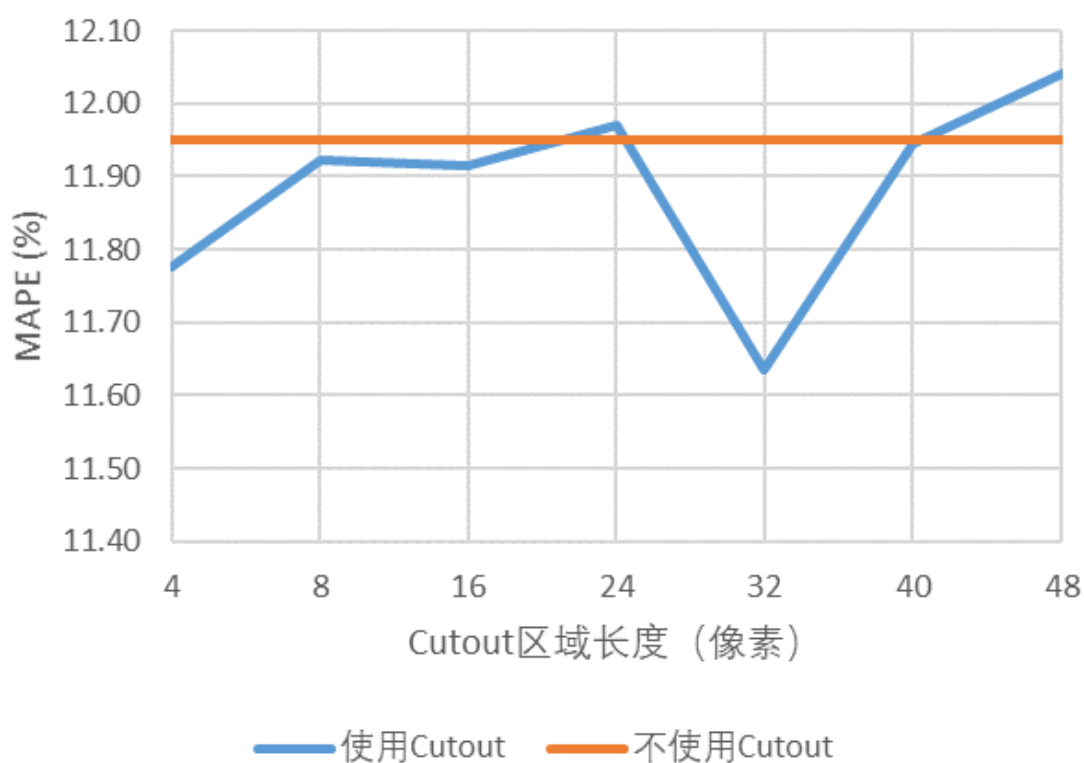


图 5-3 不同 Cutout 像素大小对回归网络模型泛化性能的影响

由上面的图表可知,并不是所有的 Cutout 策略都是有效的。采用较小的裁剪尺寸似乎总是有效的,而当裁剪掉过大的图像片块时,很可能会将重要的毛囊单位信息

丢失，同时使图像中的上下文信息遭到比较严重破坏，从而不利于毛囊单位的回归预测。本文通过网格搜索寻找到对于毛发镜图像较适宜的裁剪像素大小为 32×32 。在此设置下，网络模型拟合数据的能力得到提升（训练集 MAPE 从 3% 水平降低至 1% 水平），泛化性能也同时得到小幅提升（验证集上的 MAPE 降低了 0.31 个百分点）。

5.4.6 深度监督信息分支有效性分析

表 5-5 加入深度监督分支对回归网络模型性能的影响

基于 ResNet18					训练集 MAPE (五次平均)	验证集 MAPE (五次平均)		验证集 MAE
常规数据增强	迁移预训练模型	标签分布平滑	模拟遮挡模式	深度监督信息分支		最好	最差	
√	√	√	√		1.20 ± 0.11	11.64 ± 0.08	11.52 11.73	5.65 ± 0.07
√	√	√	√	√	3.22 ± 0.21	11.32 ± 0.29	10.99 11.68	5.49 ± 0.16

如表 5-5 所示，在前面所提机制的基础上，通过额外地在骨干网络中添加深度监督信息分支的方式，可以缓和模型的过拟合。同时，注意到，此时回归模型在训练集上的 MAPE 从 1% 的水平上升到 3%，但这并不意味着此时模型拟合数据能力的降低了。更可能的情况是，由于毛囊计数标签本身具有一定噪声，且训练集与验证集的数据分布并不可能完全一致，因此当模型越是对训练数据拟合的能力越强，便越可能在泛化阶段出现更高的误差。而在临床实际中，我们更关注的是回归模型的泛化性能，也就是对未见过的数据的预测能力，因此，在训练精度上有轻微“退化”是可以接受的。同时，正则化正是以偏差的增加为代价来换取方差的下降。

通过在深层卷积层上引入辅助回归预测分支的方式，使得最终整个回归网络的预测输出由主干网络与辅助分支的两者输出共同组成，则主干网络上在对样本产生的“过度自信”的有偏预测就会被辅助分支“拉回”，从而使得模型在训练阶段不至于过分拟合数据，增强在预测阶段的泛化性能。因此，通过在深层特征层上构建一条额外的辅助分支，一方面上缓和了主干网络的过拟合趋势。另一方面，辅助分支使得弱标签信息在深层及更深层的特征层间共享，实现了对弱标签信息的复用。

5.4.7 实验总结

在进行一系列探索实验后，本文选取了 ResNet18 作为本文所提网络模型的骨干网络，并以此为基准模型开展进一步的实验。针对图像数据标签分布不平衡的特点，我们提出了标签平滑的技术以使标签数值分布更平滑；针对毛发镜图像中普遍存在的

毛发交叉、遮挡的复杂情况，我们提出了基于 Cutout 的数据增强以模拟毛发中的遮挡模式；针对本数据集仅有毛发镜图像中毛囊单位计数这一弱监督信息的特点，我们提出通过在主干网络中添加深度监督信息分支的方式以实现弱标签信号在不同特征尺度空间上的复用。

整理 5.4 节的实验结果如下表 5-6 所示，本文提出的正则化技术方法都有效地提升了深度网络模型回归预测毛囊单位数目的性能。

表 5-6 本文所提方法在毛发镜图像毛囊回归计数任务上的实验结果

基于 ResNet18					训练集 MAPE	验证集 MAPE	最好	最差	验证集 MAE
常规数据增强	迁移预训练模型	标签分布平滑	模拟遮挡模式	深度监督信息分支					
✓					12.18 ± 0.24	12.38 ± 0.18	12.16	12.60	-
✓	✓				3.12 ± 0.40	12.17 ± 0.44	11.36	12.73	-
✓	✓	✓			3.20 ± 0.28	11.95 ± 0.16	11.71	12.13	-
✓	✓	✓	✓		1.20 ± 0.11	11.64 ± 0.08	11.52	11.73	5.65 ± 0.07
✓	✓	✓	✓	✓	3.22 ± 0.21	11.32 ± 0.29	10.99	11.68	5.49 ± 0.16

5.5 回归预测模型性能分析

5.5.1 整体模型性能分析

如表 5-6 所示，本文提出的关于毛发镜图像中毛囊单位计数任务的回归网络模型，在 80 张验证集上，预测值与计数标签的平均绝对值误差为 5.49，平均相对误差为 11.32%。

进一步地，我们选取了本文所提出的回归网络模型的一次预测结果（具体结果见附录 G），并绘制模型预测与计数标签之间的散点图如图 5-4 所示。回归分析的 R^2 分数为 0.619，皮尔逊相关系数 0.787，组内一致性系数 $ICC^{\text{①}}$ 为 0.734（具体检验结果见附录 F），说明模型预测结果与专家计数标签一致性高。

^① 用于衡量和评价观察者间信度和复测信度的信度系数，其值介于 0~1 之间。一般认为，信度系数大于 0.75 表示信度良好。

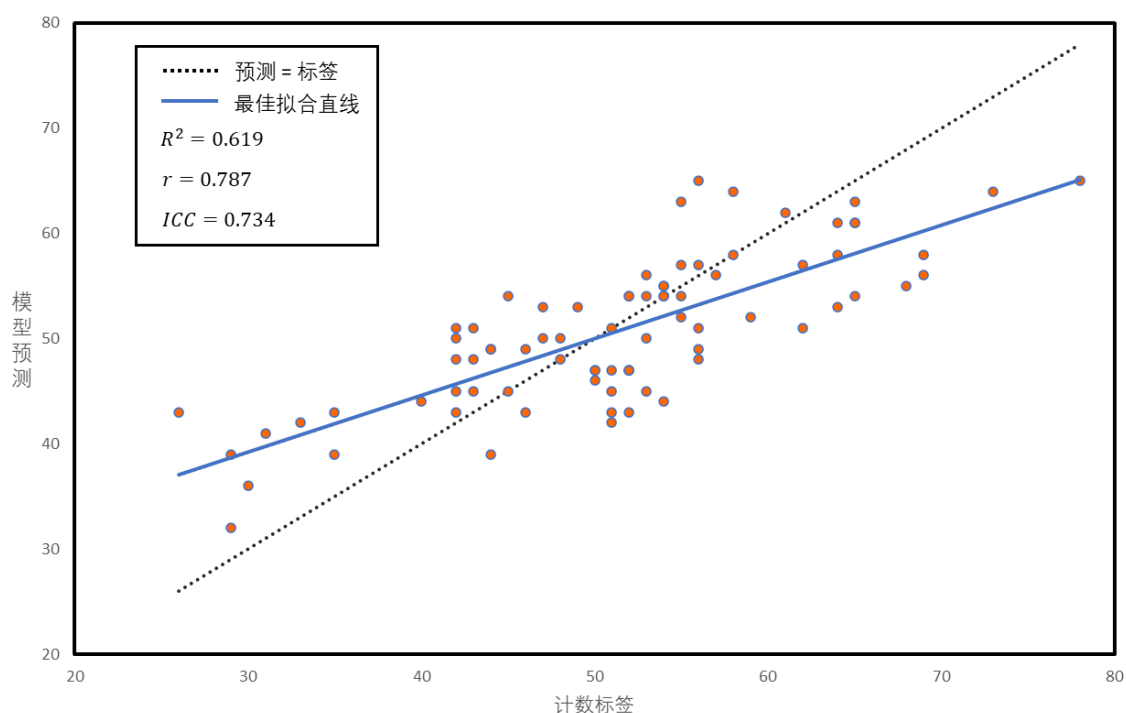


图 5-4 计数标签与模型预测的散点图

表 5-7 与已有深度模型的性能对比

模型	数据集 图像对象	毛发 长短	标注信息	数据集 容量 (张)	数据集 统计信息 (根)	MAPE	MAE	ICC
Gallucci ^[41]	体表皮肤 毫毛图像	短	毫毛位置 像素信息	共 4358	平均 54.46 数值范围 0~389	24	6.99	-
Jarek ^[43]	头部毛发 显微图像	短	毛发位置 像素信息	训练集 227 验证集 72	-	-	-	0.97
Gregor ^[44]	OCT 毛发图像	长	毛囊位置 像素信息	训练集 50 验证集 20	平均 44.9	11.8	-	-
本文所提	头部毛发 镜图像	长	毛囊数目 图像信息	训练集 320 验证集 80	平均 50.61 数值范围 20~90	11.32	5.49	0.734

与已有的一些关于毛发计数或毛囊单位计数的深度模型相比,本文所提回归预测模型仅需要图像级的弱标签信息,即毛发镜图像中的毛囊单位数目的标量值。相对那些要求精确到像素级的点标注或线、框标注的强标签信息的深度模型,本文实验方法显著减少了人工标注工作量。与已有深度模型的性能分析对比如上表 5-7 所示。尽管

各深度模型所处理的毛发图像数据集不同，但本文的毛发镜图像数据集难度更大，本文所提模型关于毛囊数目的回归预测性能指标均与已有的深度模型的性能具有可比性。

5.5.2 困难样本分析

同时，为进行困难样本分析，本文统计了验证集中的不同计数标签样本的平均预测误差 MAPE 如图 5-5 所示。可以看到，验证集上的样本预测误差（图中以橙色柱状图显示）呈现出“两头高”、“中间矮”的形势。值得一提的是，回归模型对计数标签为 26、29、30、31、33、35 对应样本的预测误差偏大，这与平均相对误差 MAPE 这个评价指标的非对称性质具有一定关系，因为 MAPE 会放大那些低标签数值对应样本的相对误差。

进一步地，联合计数标签样本在数据集中出现的频率（图 5-5 中以蓝色柱状图显示）信息进行分析可知，对于那些预测误差值相对较高的样本，它们的图像数据量在整体数据集中占比偏少，属于低频样本。本文检验了验证集中样本的预测误差与其在数据集中出现频率之间的存在负相关关系（皮尔逊相关系数 $r = -0.453$ ，双尾显著性 $p = 0.008$ ，具体结果见附录 F），说明样本数据出现频率的高低的确影响了回归模型对其预测结果的好坏。这反映了高频样本在网络模型训练的过程中获得了更好的学习，而低频样本则欠佳。

网络模型对低频样本的学习不如高频样本好，这是批次训练策略无法避免的，即网络模型在每一次的训练批次中，旨在降低所有样本的总体损失值，尽管我们尝试以重加权损失函数的方式提高低频样本的训练损失，但依然无法根除低频样本对模型学习能力的消极影响。

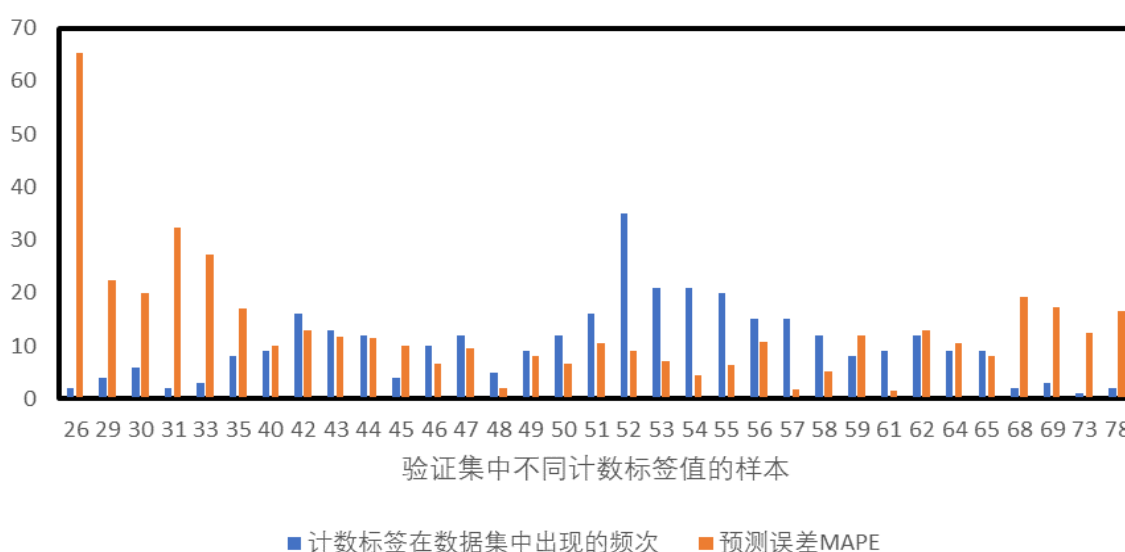


图 5-5 验证集中不同计数标签样本的预测误差与其计数标签在数据集中出现频次的关系

第6章 总结与展望

6.1 总结

6.1.1 本文工作总结

毛囊密度检测或毛发密度检测是人类种族学、皮肤病学、医美整型、烧伤护理、动物学研究等众多学科领域中普遍存在的工作环节。然而在实际研究工作中,这些繁琐的计数环节通常由研究人员手动计数来完成。即使有关于毛囊单位计数的自动化工作,也多针对短发图像。本文围绕皮肤科临床实际中,对患者头皮区域中的毛囊单位进行计数这一任务,提出了基于深度学习的毛发镜图像毛囊单位计数的回归预测模型。本文的主要贡献有:

- (1) 首个提出基于深度卷积神经网络的回归预测模型用于回归计数毛发镜图像中长发的毛囊单位数目,实现了端到端的、快速、可靠计数。回归模型在 GPU 上对一张毛发镜图像进行推理的时间仅需 0.01 秒;在 80 张验证集上取得的预测结果与毛囊单位计数标签的平均绝对值差异为 5.49,平均相对误差为 11.32%;回归模型的预测结果与计数标签一致性高, $R^2=0.619$, 组内相关系数为 0.734。相较前人工作,本文毛发镜图像数据集的应用场景更复杂,无需将患者头发剃短。同时,由于不需要毛囊单位的具体类别信息和空间信息,极大地减少了临床医生的数据标定工作量。
- (2) 为了尽可能地利用仅有的毛囊单位总计数这一偏弱的标签信息,本文提出了融合深度监督信息的辅助预测分支,实现对弱标签信息的复用,通过联合预测分支结果的方式显式地校准回归模型的预测输出。
- (3) 为了在训练阶段让低频标签对应的图像样本受到更多的关注,本文对回归任务中常用的损失函数(均方误差,常被直接使用而未作适应性修改)做了重加权标签样本密度的修改。
- (4) 为了增强回归模型的特征提取器对部分毛发镜图像中普遍存在的毛发交叉、异物遮挡等模式的识别能力,本文提出适应于毛发镜图像数据集的 Cutout 数据增强,通过在原始输入图像中随机添加若干个固定大小的零掩码,迫使特征提取器去更多地考虑图像空间中的上下文信息,以弥补局部细节信息的丢失。
- (5) 为了使回归模型不对极端值敏感,同时对标签连续数值区间中缺失数值标签的样本具有一定感知和预测,本文提出对计数标签应用平滑因子 $\epsilon = 0.1$

以重建标签数据的数值分布。

6.1.2 不足

本文的不足主要有以下三点：

尽管本文提出的模型在毛囊单位计数任务上取得了整体上与专家计数标签一致性高的预测结果，但对低计数标签的毛发镜图像样本的预测误差较大。一方面，这是由于这类极端值所对应的数据样本量偏少，回归模型对其学习不充分；另一方面，整体计数标签分布稀疏。回归模型很难对低数值区间上的样本学习到相对平滑且准确的特征边界。非均衡样本学习是领域内难点问题，尤其是对于回归任务而言^[63]。基于 5.5 节的实验分析，本文提出的重加权损失函数设计并不能彻底解决数据集中低频样本难以学习的问题。

对于毛发密集、遮挡严重的毛发镜图像而言，绝对精确的计数本身就有难点。加之不同数据标定人员之间标定经验与资质的差异化，可能导致原本相同或相近的标签数据出现较大的分布偏移，因此目前数据集标签中存在一定噪声。另一方面，鉴于数据集容量较小、标签数据分布不平衡、标签连续区间中部分数据缺失的情况，本文认为深度学习模型很容易过拟合，因此在开展实验阶段将工作重点放在研究正则化技术方面，而缺少对标签噪声的考量。Li 等人^[116]尝试用半监督学习技术来学习带有噪声标签的数据，Northcutt^[117]等人基于置信学习（confident learning）来估计标签噪声的联合分布，这些工作或可在后续研究中对本文工作产生启发。

注意到，由于人工计数毛发镜图像中毛囊单位的难点，不同的专家计数人员在对同一张毛发镜图像进行计数时，很大几率会存在计数偏差；甚至一个专业的计数人员在不同时刻对同一张毛发镜图像进行计数时，两次计数结果之间也很可能存在差异。

本文的研究目的是希望机器能自动对毛发镜图像中的毛囊单位进行可靠计数。尽管我们的最佳模型能在 80 张毛发镜图像验证集上取得与专家计数标签相比，具有 11.32% 的平均相对误差，但由于在标注毛发镜图像数据时，两位专家标注人员并没有对验证集中的图像样本进行冗余计数，这就导致了本文所提模型无法与专家计数之间的差异进行对比。理想的情况是，机器自动计数与专家计数之间的差异应接近（甚至低于）不同专家计数人员之间的计数差异，从而更好地说明本文所提模型的可靠性。

6.2 展望

随着毛发镜图像数据集容量的与日俱增，数据量不足的情况对神经网络模型性能的制约将有所缓和。由于当前毛发镜图像数据集中，毛囊单位具体的位置标签信息和类别标签信息并没有作为一种强的监督信号加入到标签空间当中，后续可以将基于弱标签信号的学习与基于强标签信号的学习进行对比，进一步地评估在减少人工标记量

的情况下，基于弱标签信息的深度学习网络模型的性能与价值。

在实验阶段，本文探索过引入近年来夺目的注意力机制（如 Convolutional Block Attention Module, CBAM）^[118]以及多任务学习^[119]范式（如加入判别毛发镜图像中毛发密度等级的辅助任务分支），但基于实验结果分析（见附录 C、D），笔者认为它们暂时还不能对本文工作产生有效启发。

注意到，在网络模型的预测阶段，即使本文已使用了多种正则化技术（包括结构正则化，数据正则化，标签正则化），同时在训练阶段使数据集中的低频样本获得更高的损失值，模型依然出现了较为明显的过拟合现象，我们期待在后续的研究中能找到一些更强、更有效的正则化技术来缓和它们。

参考文献

- [1] 王冠, 苗勇, 冯传波, 等. 中国汉族人头皮毛发分布研究[J]. 实用医学杂志, 2012, 28(09): 1441-1444.
- [2] 张春韶. 广东健康成年人头发密度及其与中医体质的关系[D]. 广州中医药大学, 2014.
- [3] 程潮江, 吴信峰. 自体毛发移植技术临床应用现状及进展[J]. 中国美容医学, 2019, 28(12): 166-169.
- [4] 韩兴斌, 李会民, 候佳琪. 五单位排列计数法在毛发移植手术中的应用[C]. 2012 全国中西医结合医学美容学术交流大会, 2012: 34-35.
- [5] 王璇, 王瑾. 雄激素源性脱发的机制及药物治疗进展[J]. 临床医药文献电子杂志, 2018, 5(33): 186-188.
- [6] 张璐璐, 王志朋, 张栋益. 雄激素性脱发的非手术治疗进展[J]. 中国临床医生杂志, 2020, 48(09): 1028-1031.
- [7] Hegde P, Relhan V, Sahoo B, et al. A randomized, placebo and active controlled, split scalp study to evaluate the efficacy of platelet - rich plasma in patchy alopecia areata of the scalp[J]. Dermatologic Therapy, 2020, 33(6): e14388.
- [8] Alhattab M K, Al Abdullah M J, Al - Janabi M H, et al. The effect of 1540 - nm fractional erbium - glass laser in the treatment of androgenic alopecia[J]. Journal of Cosmetic Dermatology, 2020, 19(4): 878-883.
- [9] Yao G, Jiang D, Li J, et al. Self-activated electrical stimulation for effective hair regeneration via a wearable omnidirectional pulse generator[J]. ACS Nano, 2019, 13(11): 12345-12356.
- [10] Blome-Eberwein S A, Roarabaugh C, Gogal C. Assessment of hair density and sub-epidermal tissue thickness in burn scars using high-definition ultrasound imaging[J]. Journal of Burn Care & Research, 2020, 41(2): 421-426.
- [11] Cui L Y, Liu W, Xu Y C, et al. A new method to estimate hair density of small mammal pelage[J]. Journal of Mammalogy, 2020, 101(4): 1205-1212.
- [12] Lee H-J, Ha S-J, Lee J-H, et al. Hair counts from scalp biopsy specimens in Asians[J]. Journal of the American Academy of Dermatology, 2002, 46(2): 218-221.
- [13] Leerunyakul K, Suchonwanit P. Evaluation of hair density and hair diameter in the adult thai population using quantitative trichoscopic analysis[J]. BioMed Research

- International, 2020, 2020: 2476890.
- [14] Rushton H, James K, Mortimer C. The unit area trichogram in the assessment of androgen - dependent alopecia[J]. *British Journal of Dermatology*, 1983, 109(4): 429-437.
- [15] Canfield D. Photographic documentation of hair growth in androgenetic alopecia[J]. *Dermatologic Clinics*, 1999, 17(2): 261-269.
- [16] Kurtti A, Jagdeo J, Eisinger A, et al. New diagnostic tools to evaluate hair loss[J]. *Dermatologic Clinics*, 2021, 39(3): 375-381.
- [17] Saitoh M, Uzuka M, Sakamoto M. Human hair cycle[J]. *Journal of Investigative Dermatology*, 1970, 54(1).
- [18] Hoffmann R. TrichoScan: Combining epiluminescence microscopy with digital image analysis for the measurement of hair growth in vivo[J]. *European Journal of Dermatology*, 2001, 11(4): 362-8.
- [19] Dhurat R, Saraogi P. Hair evaluation methods: merits and demerits[J]. *International Journal of Trichology*, 2009, 1(2): 108.
- [20] Van Neste D, Trüeb R. Critical study of hair growth analysis with computer - assisted methods[J]. *Journal of the European Academy of Dermatology and Venereology*, 2006, 20(5): 578-583.
- [21] 吴园琴, 范卫新, 董青, 等. 毛发显微图像分析系统在雄激素性秃发疗效分析中的应用[J]. *临床皮肤科杂志*, 2018, 47(01): 3-7.
- [22] Birnbaum M R, Mclellan B N, Shapiro J, et al. Evaluation of hair density in different ethnicities in a healthy American population using quantitative trichoscopic analysis[J]. *Skin Appendage Disorders*, 2018, 4(4): 304-307.
- [23] Kim J E, Lee J H, Choi K H, et al. Phototrichogram analysis of normal scalp hair characteristics with aging[J]. *European Journal of Dermatology*, 2013, 23(6): 849-856.
- [24] Lee B S L, Chan J-Y L, Monselise A, et al. Assessment of hair density and caliber in Caucasian and Asian female subjects with female pattern hair loss by using the Folliscope[J]. *Journal of the American Academy of Dermatology*, 2012, 66(1): 166-167.
- [25] Ekelem C, Feil N, Csuka E, et al. Optical coherence tomography in the evaluation of the scalp and hair: common features and clinical utility[J]. *Lasers in Surgery and Medicine*, 2021, 53(1): 129-140.
- [26] Gibbons R D, Fiedler-Weiss V C, West D P, et al. Quantification of scalp hair-a computer-aided methodology[J]. *Journal of Investigative Dermatology*, 1986, 86(1): 78-82.

- [27] Gastpar M, Dragotti P L, Vetterli M. The distributed karhunen–loeve transform[J]. IEEE Transactions on Information Theory, 2006, 52(12): 5177-5196.
- [28] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [29] Ballard D H. Generalizing the Hough transform to detect arbitrary shapes[J]. Pattern Recognition, 1981, 13(2): 111-122.
- [30] Hart P E. How the Hough transform was invented [DSP History][J]. IEEE Signal Processing Magazine, 2009, 26(6): 18-22.
- [31] Illingworth J, Kittler J. A survey of the Hough transform[J]. Computer Vision, Graphics, and Image Processing, 1988, 44(1): 87-116.
- [32] Shih H-C. An unsupervised hair segmentation and counting system in microscopy images[J]. IEEE Sensors Journal, 2014, 15(6): 3565-3572.
- [33] Vallotton P, Thomas N. Automated body hair counting and length measurement[J]. Skin Research and Technology, 2008, 14(4): 493-497.
- [34] Kim H, Kim W, Rew J, et al. Evaluation of hair and scalp condition based on microscopy image analysis[C]. 2017 International Conference on Platform Technology and Service (PlatCon), 2017: 1-4.
- [35] Seo S, Park J. Trichoscopy of alopecia areata: hair loss feature extraction and computation using grid line selection and eigenvalue[J]. Computational and Mathematical Methods in Medicine, 2020, 2020: 6908018.
- [36] Liu W, Wang Z, Liu X, et al. A survey of deep neural network architectures and their applications[J]. Neurocomputing, 2017, 234: 11-26.
- [37] Benhabiles H, Hammoudi K, Yang Z, et al. Deep learning based detection of hair loss levels from facial images[C]. 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2019: 1-6.
- [38] Chang W-J, Chen L-B, Chen M-C, et al. ScalpEye: A deep learning-based scalp hair inspection and diagnosis system for scalp health[J]. IEEE Access, 2020, 8: 134826-134837.
- [39] 高树辉, 姜晓佳. 卷积神经网络在物证检验中的应用与毛发自动识别的展望[J]. 科学技术与工程, 2019, 19(23): 1-9.
- [40] 姜晓佳, 高树辉. 基于改进卷积神经网络的毛发显微图像自动分类[J]. 激光与光电子学进展, 2019, 56(23): 83-90.
- [41] Gallucci A, Znamenskiy D, Pezzotti N, et al. Hair counting with deep learning[C]. 2020 International Conference on Biomedical Innovations and Applications (BIA), 2020: 5-

9.

- [42] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [43] Sacha J P, Caterino T L, Fisher B K, et al. Development and qualification of a machine learning algorithm for automated hair counting[J]. International Journal of Cosmetic Science, 2021, 43: S34-S41.
- [44] Urban G, Feil N, Csuka E, et al. Combining deep learning with optical coherence tomography imaging to determine scalp hair and follicle counts[J]. Lasers in Surgery and Medicine, 2021, 53(1): 171-178.
- [45] Kim M, Kang S, Lee B-D. Evaluation of Automated Measurement of Hair Density Using Deep Neural Networks[J]. Sensors, 2022, 22(2): 650.
- [46] Zhou Z-H. A brief introduction to weakly supervised learning[J]. National Science Review, 2018, 5(1): 44-53.
- [47] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [48] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [49] Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex[J]. The Journal of Physiology, 1959, 148(3): 574.
- [50] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of Physiology, 1962, 160(1): 106.
- [51] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Cambridge: MIT Press, 2016.
- [52] Hebb D O. The organization of behavior: A neuropsychological theory[M]. New York: Psychology Press, 2005.
- [53] Mcculloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.
- [54] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386-408.
- [55] Minsky M, Papert S. Perceptrons[M]. Cambridge: MIT Press, 1969.
- [56] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines[J]. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1986, 1(282-317): 2.
- [57] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [58] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin

- classifiers[C]. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 1992: 144-152.
- [59] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [60] Hinton G E, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [61] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25.
- [62] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold[J]. Nature, 2021, 596(7873): 583-589.
- [63] Yang Y, Zha K, Chen Y, et al. Delving into deep imbalanced regression[C]. International Conference on Machine Learning, 2021: 11842-11851.
- [64] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]. Proceedings of the fourteenth international conference on artificial intelligence and statistics, 2011: 315-323.
- [65] Jarrett K, Kavukcuoglu K, Ranzato M A, et al. What is the best multi-stage architecture for object recognition?[C]. 2009 IEEE 12th international conference on computer vision, 2009: 2146-2153.
- [66] 阿斯顿, 李沐, 李普顿, 等. 动手学深度学习[M]. 北京: 人民邮电出版社, 2019.
- [67] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [68] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [69] Ding X, Zhang X, Zhou Y, et al. Scaling up your kernels to 31x31: Revisiting large kernel design in CNNs[J]. arXiv preprint arXiv:2203.06717, 2022.
- [70] Lin M, Chen Q, Yan S. Network in network[J]. arXiv preprint arXiv:1312.4400, 2013.
- [71] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 1-9.
- [72] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International Conference on Machine Learning, 2015: 448-456.
- [73] Santurkar S, Tsipras D, Ilyas A, et al. How does batch normalization help optimization?[J]. Advances in Neural Information Processing Systems, 2018, 31.

- [74] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [75] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]. European Conference on Computer Vision, 2016: 630-645.
- [76] Cai Y, Du D, Zhang L, et al. Guided attention network for object detection and counting on drones[J]. arXiv preprint arXiv:1909.11307, 2019.
- [77] Farjon G, Krikeb O, Hillel A B, et al. Detection and counting of flowers on apple trees for better chemical thinning decisions[J]. Precision Agriculture, 2020, 21(3): 503-521.
- [78] Falk T, Mai D, Bensch R, et al. U-Net: deep learning for cell counting, detection, and morphometry[J]. Nature Methods, 2019, 16(1): 67-70.
- [79] Morelli R, Clissa L, Amici R, et al. Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet[J]. Scientific Reports, 2021, 11(1): 1-11.
- [80] Lempitsky V, Zisserman A. Learning to count objects in images[J]. Advances in Neural Information Processing Systems, 2010, 23.
- [81] Li Y, Zhang X, Chen D. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1091-1100.
- [82] Zhang C, Li H, Wang X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 833-841.
- [83] Akiva P, Dana K, Oudemans P, et al. Finding berries: Segmentation and counting of cranberries using point supervision and shape priors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 50-51.
- [84] Paul Cohen J, Boucher G, Glastonbury C A, et al. Count-ception: Counting by fully convolutional redundant counting[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 18-26.
- [85] Wang C, Zhang H, Yang L, et al. Deep people counting in extremely dense crowds[C]. Proceedings of the 23rd ACM International Conference on Multimedia, 2015: 1299-1302.
- [86] Liang D, Chen X, Xu W, et al. Transcrowd: Weakly-supervised crowd counting with transformer[J]. arXiv preprint arXiv:2104.09116, 2021.
- [87] Parmar N, Vaswani A, Uszkoreit J, et al. Image transformer[C]. International Conference on Machine Learning, 2018: 4055-4064.

- [88] Onoro-Rubio D, López-Sastre R J. Towards perspective-free object counting with deep learning[C]. European Conference on Computer Vision, 2016: 615-629.
- [89] Rahnemoonfar M, Sheppard C. Deep count: fruit counting based on deep simulated learning[J]. Sensors, 2017, 17(4): 905.
- [90] Dobrescu A, Valerio Giuffrida M, Tsiftaris S A. Leveraging multiple datasets for deep leaf counting[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 2072-2079.
- [91] Rothe R, Timofte R, Van Gool L. Dex: Deep expectation of apparent age from a single image[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015: 10-15.
- [92] Levi G, Hassner T. Age and gender classification using convolutional neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015: 34-42.
- [93] Escalera S, Fabian J, Pardo P, et al. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015: 1-9.
- [94] Jónsson B A, Bjornsdottir G, Thorgeirsson T, et al. Brain age prediction using deep learning uncovers associated sequence variants[J]. Nature Communications, 2019, 10(1): 1-10.
- [95] Liao L, Zhang X, Zhao F, et al. Multi-branch deformable convolutional neural network with label distribution learning for fetal brain age prediction[C]. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020: 424-427.
- [96] Shen L, Zheng J, Lee E H, et al. Attention-guided deep learning for gestational age prediction using fetal brain MRI[J]. Scientific Reports, 2022, 12(1): 1-10.
- [97] Massey Jr F J. The Kolmogorov-Smirnov test for goodness of fit[J]. Journal of the American statistical Association, 1951, 46(253): 68-78.
- [98] Lee C-Y, Xie S, Gallagher P, et al. Deeply-supervised nets[C]. Artificial Intelligence and Statistics, 2015: 562-570.
- [99] Vandenhende S, Georgoulis S, Proesmans M, et al. Revisiting multi-task learning in the deep learning era[J]. arXiv preprint arXiv:2004.13379, 2020, 2: 3.
- [100] Wang L, Lee C-Y, Tu Z, et al. Training deeper convolutional networks with deep supervision[J]. arXiv preprint arXiv:1505.02496, 2015.
- [101] Lin T-Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-

2988.

- [102] Devries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv preprint arXiv:1708.04552, 2017.
- [103] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [104] Müller R, Kornblith S, Hinton G E. When does label smoothing help?[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [105] Lukasik M, Bhojanapalli S, Menon A, et al. Does label smoothing mitigate label noise?[C]. International Conference on Machine Learning, 2020: 6448-6458.
- [106] 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.
- [107] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
- [108] Santos C F G D, Papa J P. Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks[J]. ACM Computing Surveys (CSUR), 2022.
- [109] Afonso B, Melo L, Oliveira W, et al. Housing prices prediction with a deep learning and random forest ensemble[C]. Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional, 2019: 389-400.
- [110] Wang F, Zou Y, Zhang H, et al. House price prediction approach based on deep learning and ARIMA model[C]. 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), 2019: 303-307.
- [111] Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [112] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.
- [113] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [114] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [115] Blumer A, Ehrenfeucht A, Haussler D, et al. Occam's razor[J]. Information Processing Letters, 1987, 24(6): 377-380.
- [116] Li J, Socher R, Hoi S C. DivideMix: Learning with noisy labels as semi-supervised

- learning[C]. International Conference on Learning Representations, 2019.
- [117] Northcutt C, Jiang L, Chuang I. Confident learning: Estimating uncertainty in dataset labels[J]. Journal of Artificial Intelligence Research, 2021, 70: 1373-1411.
- [118] Woo S, Park J, Lee J-Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [119] Ruder S. An overview of multi-task learning in deep neural networks[J]. arXiv preprint arXiv:1706.05098, 2017.

附录 A 骨干网络选取实验的详细结果

模型	Params (M)	训练集 MAPE						
		第一次	第二次	第三次	第四次	第五次	平均	标准差
VGG11	129	12.78	12.84	13.38	12.46	12.82	12.86	0.30
VGG16	134	13.34	12.04	13.53	13.79	12.77	13.09	0.62
AlexNet	57	13.48	12.81	13.30	12.89	13.20	13.14	0.25
GoogLeNet	6	13.08	12.95	14.21	12.88	15.12	13.65	0.88
Inception_v3	22	12.30	12.80	12.22	13.02	13.33	12.73	0.42
ResNet18	11	12.51	13.04	13.03	12.78	13.06	12.88	0.21
ResNet34	21	11.89	12.21	11.92	12.21	13.35	12.32	0.53
ResNet50	24	11.44	9.12	11.84	10.80	10.72	10.78	0.93

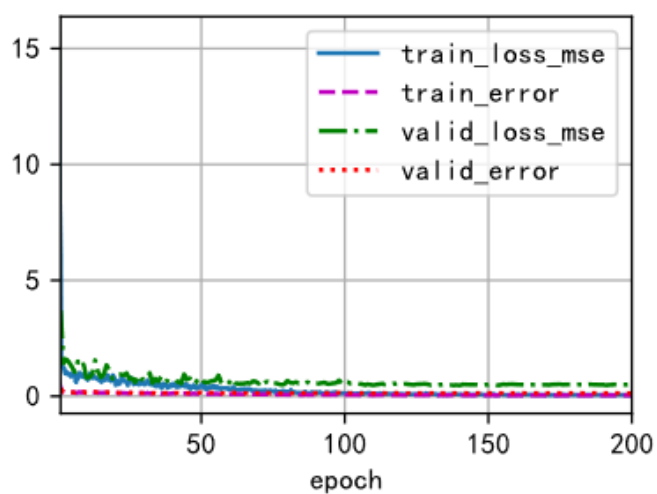
模型	Params (M)	验证集 MAPE						
		第一次	第二次	第三次	第四次	第五次	平均	标准差
VGG11	129	12.55	13.32	13.20	13.40	13.34	13.10	0.39
VGG16	134	12.70	12.94	12.76	12.83	13.04	12.85	0.12
AlexNet	57	12.33	12.66	12.33	12.72	11.89	12.39	0.30
GoogLeNet	6	12.03	13.30	12.62	12.82	12.97	12.75	0.42
Inception_v3	22	13.02	11.70	12.61	12.47	11.73	12.31	0.52
ResNet18	11	12.60	12.16	12.35	12.20	12.57	12.38	0.18
ResNet34	21	12.88	13.10	12.89	12.92	12.14	12.79	0.33
ResNet50	24	12.74	13.65	12.49	13.00	13.05	12.99	0.39

附录 B 基于 Cutout 数据增强的网格搜索实验详细结果

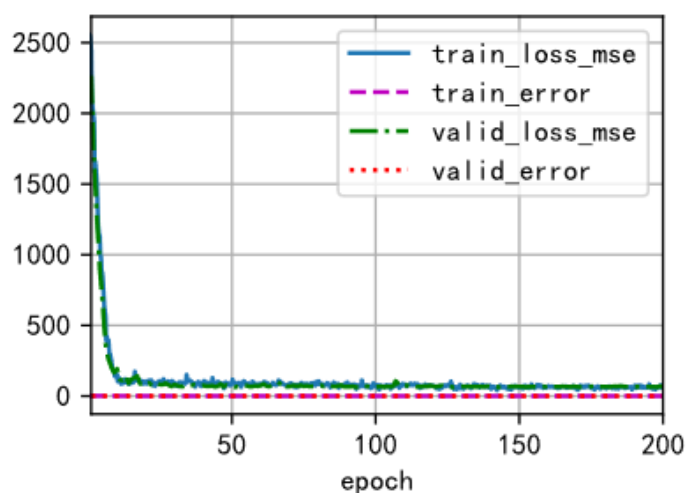
设置	训练集 MAPE						
	第一次	第二次	第三次	第四次	四五次	平均	标准差
4×4	1.12	1.14	1.03	0.94	1.24	1.09	0.10
8×8	0.90	1.06	1.62	1.14	1.04	1.15	0.25
16×16	1.02	1.12	1.05	1.04	1.09	1.06	0.04
24×24	1.18	1.41	1.00	0.93	1.30	1.16	0.18
32×32	1.11	1.31	1.27	1.27	1.04	1.20	0.11
40×40	0.99	0.94	1.08	1.08	1.25	1.07	0.11
48×48	1.55	1.43	1.36	1.76	1.21	1.46	0.19

设置	验证集 MAPE						
	第一次	第二次	第三次	第四次	四五次	平均	标准差
4×4	11.25	11.68	11.44	12.38	12.13	11.78	0.42
8×8	11.75	11.7	12.15	12.33	11.68	11.92	0.27
16×16	11.9	11.96	11.79	11.3	12.62	11.91	0.42
24×24	11.8	11.45	11.8	12.78	12.37	12.04	0.47
32×32	11.56	11.71	11.66	11.52	11.73	11.64	0.08
40×40	11.57	11.97	12.18	11.89	12.11	11.94	0.21
48×48	11.29	12.41	11.91	11.87	12.37	11.97	0.41

附录 C 引入注意力机制的实验训练过程及结果



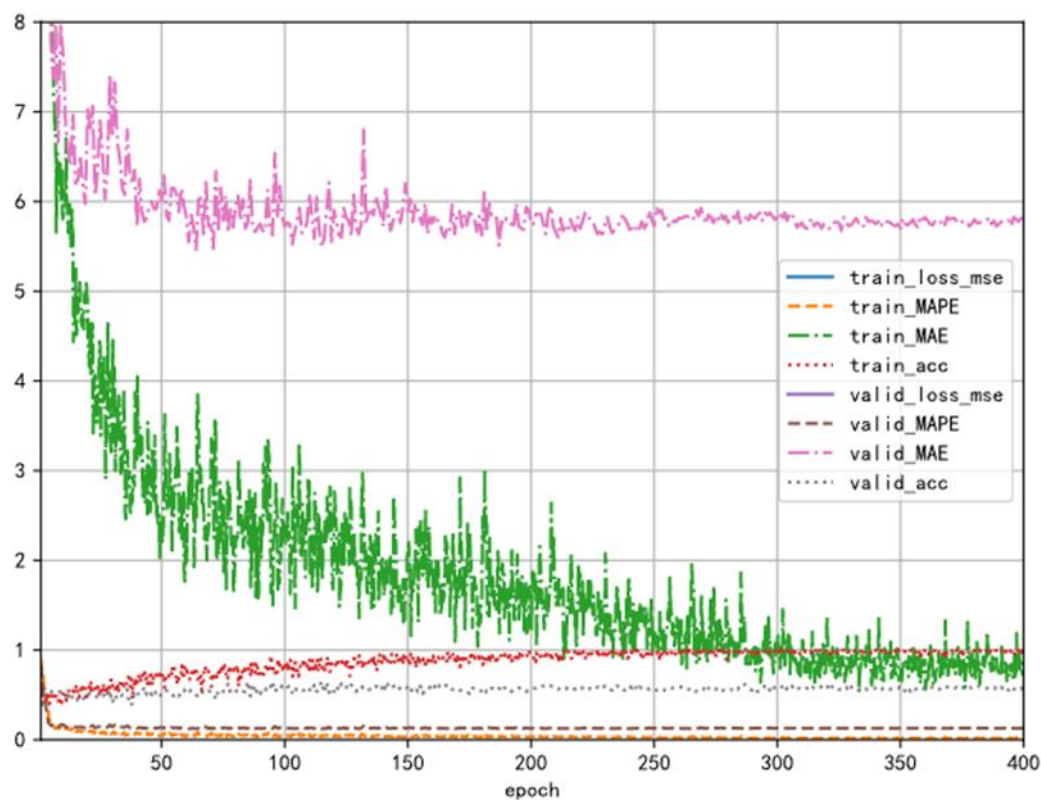
(a) ResNet18+CBAM



(b) ResNet50+CBAM

模型	训练 MAPE	验证 MAPE
ResNet18 (baseline)	3.12 (五次平均)	12.12 (五次平均)
ResNet18 + CBAM	3.88	12.58 (最好结果)
ResNet50 (baseline)	1.47 (五次平均)	12.19 (五次平均)
ResNet50 + CBAM	12.98	12.89 (最好结果)

附录 D 引入多任务学习的实验训练过程及结果



模型	训练 MAPE	验证 MAPE
ResNet18 (baseline)	3.12 (五次平均)	12.12 (五次平均)
+ 毛发密度等级分类任务分支	1.44	12.81(最好结果)

附录 E 训练集和验证集分布一致性的 K-S 检验

描述统计

	个案数	平均值	标准 偏差	最小值	最大值	第 25 个	百分位数 第 50 个 (中 位数)	第 75 个
2	400	50.6075	10.51689	20.00	90.00	43.0000	52.0000	57.0000
分组	400	1.20	.401	1	2	1.00	1.00	1.00

双样本柯尔莫戈洛夫-斯米诺夫检验

频率

	分组	个案数
2	1	320
	2	80
	总计	400

检验统计^a

		2
最极端差值	绝对	.091
	正	.091
	负	-.031
柯尔莫戈洛夫-斯米诺夫 Z		.725
渐近显著性 (双尾)		.669

a. 分组变量: ^1

附录 F 皮尔逊相关性检验

相关性

描述统计			
	平均值	标准 偏差	个案数
计数标签	51.31	10.399	80
模型预测	50.68	7.128	80

相关性			
		计数标签	模型预测
计数标签	皮尔逊相关性	1	.787**
	Sig. (双尾)		.000
	个案数	80	80
模型预测	皮尔逊相关性	.787**	1
	Sig. (双尾)	.000	
	个案数	80	80

** 在 0.01 级别 (双尾), 相关性显著。

相关性

描述统计			
	平均值	标准 偏差	个案数
平均MAPE	.13297649	.116307124	33
总体频次	10.2121	7.24033	33

相关性			
		平均MAPE	总体频次
平均MAPE	皮尔逊相关性	1	-.453**
	Sig. (双尾)		.008
	个案数	33	33
总体频次	皮尔逊相关性	-.453**	1
	Sig. (双尾)	.008	
	个案数	33	33

** 在 0.01 级别 (双尾), 相关性显著。

附录 G 模型回归预测毛囊单位计数的具体结果

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
y	62	53	51	49	55	45	46	52	69	42	46	56	52	29	51	78	42	43	50	54
$\hat{y}$	51	50	43	53	52	45	49	43	58	43	43	57	54	39	51	65	50	45	47	44
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
y	54	30	55	42	47	50	52	51	31	59	44	58	56	55	42	35	64	42	51	47
$\hat{y}$	54	36	57	48	50	47	47	42	41	52	49	58	49	63	51	39	58	45	45	53
	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
y	65	73	35	61	64	54	54	55	51	56	45	26	53	52	62	43	64	43	58	40
$\hat{y}$	54	64	43	62	53	55	55	54	47	48	54	43	54	47	57	51	61	48	64	44
	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
y	65	48	48	56	29	56	52	44	57	54	44	33	52	53	50	53	65	68	69	65
$\hat{y}$	61	48	50	51	32	65	54	49	56	54	39	42	47	45	46	56	61	55	56	63

攻读学位期间主要研究成果

一、专利

[1] 冯靖安, 郑锐强, 范衡, 彭杰华. 一种基于卷积神经网络的中华白海豚背鳍识别方法. 公开号: CN110837818A

二、软件著作权

[1] 黄文宁, 韦家弘, 彭杰华, 朱智健. 场馆人群数量估计与检测系统 V1.0. 登记号: 2020SR1718224

三、拟发表论文

[1] 彭杰华, 范衡. 基于深度学习的毛囊单位回归计数方法.

致谢

相聚有时，总是别离。匆匆这年，行将结束这一瞬三年的研究生生活。值此交付论文之际，回首过往，感慨良多，无以言表，唯有感恩。向在我整个研究生学习、生活期间所有关心和帮助我的人们致以我最真挚的谢意。

感谢我的导师范衡教授。他永葆科研热情，身体力行、严谨治学的态度，垂范后辈，一直鼓舞着我沉下身、静下心来研究和学习。在学习中，他帮助我们梳理、凝练科研项目中的科学问题，向我们传授宝贵的科研经验、科学严谨的逻辑思辨能力；他总是不吝表扬与肯定我取得的即使是很小的成绩，而当我遇到暂时难以突破的难题时，又总是激励我再坚持下去。在生活中，他平易近人、宽容大度、谦虚务实，这些品质同样感染着我。三年师恩，点滴难以尽述，而我从中收获到的学识却很可能使我受益终生。

感谢林常敏教授，是她组织和牵头了本文的工作。同时，还要感谢林教授的研究生魏奕娜、倪娜同学对我的工作的理解、支持与配合。

感谢帮助过我、给我提供建议的所有师兄师姐、同门以及师弟。

感谢阿木木人工智能实验室的卓文浩院长和首席执行官吴德政在我研究生生活中给予我的帮助与陪伴。

感谢李良慧、邱德卓、付子杰、王萍、练瑜琦、肖昱晨、陈家兴的关心与陪伴，虽则他们不在我身边，但从他们身上我获得过短暂的治愈。

感谢邹秋云师兄。感谢李宗杰、冯靖安师兄。他们无私的帮助让我动容。金麟只是池中物，误落尘网三四年。等闲却见青天过，喘喘人作花下魂。

最后，感谢我的家人一直以来对我含辛茹苦的培养。有志有识，且尽人事，何惧无为。