



兰州大学

大学生创新创业行动计划 项目申报书

项目名称：基于卷积神经网络的皮肤癌检测方法研究

所在学院：信息科学与工程学院

起止时间：2020 年 4 月 至 2020 年 11 月

Hollow Man

共青团兰州大学委员会 制

填 报 说 明

一、项目团队所有成员和指导老师应严格按照有关规定执行。鼓励学生跨年级、跨专业组队申报项目，促进不同学科之间的交叉融合。

二、同一年度个人最多参与两项，最多担任一个项目负责人。

三、团队成员根据实际情况填写（包括负责人在内最多不超过 5 人）；
“所在学院”一栏填写项目负责人所在的学院。

四、申报书请如实填写，表达明确严谨。所需签字之处，必须由相应人员亲笔签名。如有弄虚作假现象，一经核实，将按照撤项处理。

五、项目实施时间的起止年月一般按自然年份填写，每个项目原则上在当年内完成，以半年为一阶段。申报书的各项内容，要实事求是，表达要明确、严谨。第一次出现的缩写词，需注出全称。

六、项目及子项目经费预算中的科目名称可根据预算情况填写实验耗材、药品、试剂、图书资料、实地调研，参加校内外会议，生产采购、仓储运输、广告营销，场地租赁等。

七、所有项目均需负责人通过团委综合业务管理系统进行项目申报，同时上传电子版《申报书》。电子版、纸质版申报书报送学院，纸质版要求用 A4 纸填写或打印（签字之处不得打印），于左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。电子版填表字体用小四号宋体，单倍行距。

八、所列各项内容不能简单标注“见附件”，否则视为不合格。

项目名称	基于卷积神经网络的皮肤癌检测方法研究		
项目类型	<input checked="" type="checkbox"/> 自然科学类 <input type="checkbox"/> 社会科学类 <input type="checkbox"/> 医药科学类 <input type="checkbox"/> 创业计划类 <input type="checkbox"/> 创业实践类		
申请经费	200 （元）	起止时间	2020 年 4 月 至 2020 年 11 月
<p>一、项目简介（严格控制在 200 字左右）</p> <p>本项目利用卷积神经网络（CNN）实现皮肤病图像的分类与识别。根据皮肤病图像的特征，利用卷积神经网络对良性和恶性黑色素瘤进行分类和识别。针对医学图像数据集（病例数据）有限问题研究适合的数据增强方法以提高卷积神经网络的分类精度，增强训练模型的学习性能。此项目成果主要面向疑似皮肤病患者和皮肤病医生，实现对皮肤癌的初筛。通过实现低成本的皮肤癌初筛诊断，从一定程度上减轻患者与医生的医疗负担。</p>			

二、项目创新点（严格控制在 100 字左右）

1. 本项目使用卷积神经网络（CNN）深度学习技术来分类识别黑色素瘤。卷积神经网络模型在有监督学习方式下与传统的分类方法相比往往可以获得更准确的高级语义信息以便于提高皮损图像分割的精度。
2. 针对目前病理图像训练数据集较匮乏的情况，本项目将通过迁移学习等数据增强方法进行数据增广的 CNN 分类模型学习。
3. 尝试使用多数表决融合或加权融合等模型融合方法提高判断的准确度。

三、申请理由（可另附页。1500 字左右，包括项目实施目的、背景、内容、自身/团队具备的知识、条件、特长、兴趣、前期准备，以及参加哪些科技事件创新活动和取得的成绩等。需对项目可行性分析作专门论述，背景分析、成本分析、社会效益或经济效益分析等。）

目的：

此项目意在寻找适合于分类的皮肤癌病理图片特征信息，通过对待检测者疑似患有皮肤癌部位的皮肤纹理数据集预处理后，将其输入基于卷积神经网络的皮肤分类网络模型进行训练调参，最后尝试使用多模型融合方法以尽可能提高准确度。

背景：

根据皮肤科经典教科书的记载，不同皮肤病诊断名称共有超过 2000 种。除了种类繁多，很多皮肤恶性肿瘤在早期很容易与其他皮肤疾病相混淆，等到确诊时已是晚期。比如脂溢性角化病和色素痣、恶性黑素瘤之间的鉴别诊断，以及一些特殊部位皮肤肿瘤的诊断在临床中就有着较高的误诊率。由于大多数皮肤肿瘤的表现都较为复杂，这就给了误诊、漏诊可趁之机，这样的情况在基层医疗机构更为突出。

有数据显示，我国各类医疗机构每年的皮肤科门诊量超过 2 亿，而皮肤科医生数量仅略超过 2 万，这就意味着，皮肤科医生的年均接诊量数以万计。除了皮肤科

医生数量存在严重不足，不同地域、不同医疗机构医生的诊断水平也存在着显著差异。这已经成为皮肤科学面临的重大临床问题。

我们希望该项目的成功实现可以让皮肤科医生的诊断拓展到临床之外，帮助患者及早筛查识别出皮肤癌，提高重视程度，同时帮助医生进行辅助排查诊断，减轻了医疗压力，而且一定程度上能帮助医生减少误诊率。因而，我们基于卷积神经网络的皮肤癌识别能够实现低成本皮肤癌排查诊断，降低诊断成本，具有不错的经济价值。

项目内容：

基于卷积神经网络进行的智能皮肤癌检测可以简单分为 6 个阶段，即：获取数据集、数据集分类及划分测试训练集、搭建神经网络、训练模型、测试模型、模型优化。

我们使用当下较为流行的卷积神经网络进行网络结构的搭建，然后使用 ISIC(International Skin Imaging Collaboration)的皮肤病开源数据集，整理出其中我们需要的皮肤癌图片并将其分类（这里我们先将黑色素瘤及痣的图片分别整理，更多分类功能后续可逐步添加）。训练模型阶段，我们考虑采用 Faster R-CNN、ResNet 这两种深度学习网络模型进行训练，考虑到训练模型需要的时间以及计算能力，我们使用 ImageNet 中已经训练好的模型进行迁移学习，并且进行模型的参数调整以及数据增强，直至测试结果准确率达到标准。最后，根据两个模型的测试结果进行模型融合（具体融合方法要根据结果判断），或对两个模型结果取加权平均，以期得到的结果更加准确。

项目前期准备：

①关于研究选择：

背景分析中，我们已经了解到，皮肤癌是世界上增长最快的癌症，我国皮肤科医生数量严重不足，黑色素瘤等皮肤癌早期发现的治愈率将大大提高。而目前可以辅助医生确诊的智能系统少之又少，可以普及给大众进行皮肤健康自测的有关模型目前存在缺口。我们希望就卷积神经网络入手，对黑色素瘤这种将为常见的皮肤癌进行智能识别和分类。

②关于信息收集：

我们通过采用关键词：卷积神经网络，深度学习，皮肤癌，病变，黑色素瘤等相结合搜索 Google Scholar, NCBI, CNKI, ScienceDirect 和 Web of Science 数据库，以保证所用信息具有足够科学依据。

③关于 CNN（卷积神经网络）：

CNN 是具有特定架构的神经网络，已被证明在诸如图像识别和分类等领域具有传统机器学习方法不能比拟的效果。CNN 已被证明比人类更能识别面部，物体和交通标志，因此在机器人和自动驾驶汽车领域已经有了成功的应用。

CNN 是一种有监督的学习方法，使用带有相应类别标记的数据进行训练。从本质上讲，CNN 会学习输入对象和类标签之间的关系，并包括两个组件：在其中提取要素的隐藏层以及在处理结束时用于实际分类任务的完全连接的层。与常规神经网络不同，CNN 的隐藏层具有特定的体系结构。在常规神经网络中，每一层由一组神经元组成，并且一层的一个神经元连接到前一层的每个神经元。CNN 中隐藏层的架构略有不同。一层中的神经元未连接到前一层的所有神经元；相反，它们仅连接到少数神经元。这种对本地连接的限制以及将本地神经元输出汇总为一个值的附加池化层，可以简化训练过程并降低模型复杂度。

团队成员介绍：

我们团队由五人组成，他们分别来自信息科学与工程学院 2018 级和 2019 级，由计算机和电信学科的跨年级跨专业同学组成，都熟练掌握 MATLAB 语言，对人工智能方面的知识都有一定程度的掌握和了解。总体上，本团队的每个成员都有着认真负责的工作态度，对科研项目有着高度的热情，同时跨年级跨专业的团队组成确保了团队的创新性与活力。

项目可行性分析：

我们所采取的方法是对最新公开发表的研究成果进行优化，集成和创新。此外，我们在卷积神经网络模型的基础上，进行数据增广并尝试使用多数表决融合或加权融合等模型融合方法以提高判断的准确度。皮肤是人体最大的器官，皮肤病的种类不但繁多，且在形态学上大多相似，前期难以分辨。我们通过调查、阅读论文了解到，如果只通过图片进行诊断，机器的诊断正确率以及所用时间明显优于医生。无论作为医生的辅助诊断工具还是大众的健康自测，我们的研究项目都不会缺少市场。此次项目中我们所采取的工具及数据集大多开源，研究成本不会太高。

四、项目方案（可另附页。1500 字左右，包括目标任务，技术路线或运作模式，行动方案，以及人员分工和资源整合等。）

目标任务：

近年来，由于机器学习与医学诊断相结合的方式极大地提高了医生对于疾病诊断的精确性，且随着计算机科学技术的发展，机器学习能够提取出图像的高级特征，对皮肤病变图像进行分类的研究已成为机器学习领域的一个热点。研究表明，将卷积神经网络（CNN）用于端对端学习（End-to-End learning）分类模式的皮肤样本的研究，相对于其他网络模型具有无可比拟的优势。虽然基于深度学习进行分类的方法，其精度已获得显著提高，但由于数据集获取困难、人工标注成本较高等诸多因素的限制，深度学习在黑色素瘤等皮肤病的分类问题方面依然存在着分类精度低、数据集缺乏等问题。我们将研究基于卷积神经网络的对皮肤病变进行分类的模型，着重于就黑色素瘤的颜色和纹理等特征进行识别、分类，最后希望融入多种模

型算法，以帮助医师提高判断的准确度。

技术路线：

首先是前期准备，小组成员一起学习探究现在有关皮肤癌的医学影像特征和检测流程、皮肤癌图像特征分类方法、基于卷积神经网络（CNN）原理和算法的图像识别和处理技术，了解其研究现状及存在的主要问题。同时根据自身情况，自主学习相关理论知识，熟悉 MATLAB 和 Python 的图像和数据处理及相关的工具包，并广泛阅读国内外有关论文成果。接下来进入正式的研究阶段，在寻找到合适数据库并对数据集进行扩充、预处理后，将可用数据集分为测试、训练和验证三部分。在对图像分类任务分析后，使用基于 CNN 深度学习技术分类识别疑似病变的皮肤镜图像，使输入图像中像素的强度与对应类别标签的概率形成对应关系，最后尝试使用多数表决融合或加权融合等模型融合方法提高判断的准确度。

行动方案：

我们先深入研究当前卷积神经网络的发展现状并搜集阅读国内外有关论文成果，从中了解当前 CNN 网络模型有关应用（未必一定是皮肤识别、分类研究）、清楚 CNN 处理图像的有关机理，并基于 CNN 网络模型的相关研究汇集大家对现有模型的改进想法，形成共同的努力方向。与此同时，小组成员根据自身情况学习有关理论知识，如线性代数、MATLAB 仿真基础与应用、Python 实践与应用等，为顺利学习、理解 CNN 网络模型夯实基础。之后根据每个团队成员自身的兴趣、特点，对大家在各方面的核心任务进行分工。此外，在本学期内（特殊情况除外）保证每两周进行小组会议以分享交流讨论项目进展。我们希望在暑假（7、8 月）前找出合适的数据集并完成使用卷积神经网络常用的图像处理方法割出皮肤图像中的受损皮肤图像。暑假期间将基于卷积神经网络的皮肤分类网络模型进行训练调参，并重复进行“训练（调参和数据增强等）、讨论、修正”的过程，同时对比结果进行选型，尽可能提高其准确度，不断重复这个过程直至最终得到令人较为满意的结果。然后是模型的修改完善阶段，此时模型已经过一定的训练，我们将尝试使用多数表决融合或加权融合等有助于提高模型准确度的模型融合方法，进而达到优化模型、提升算法精度的目的。

人员分工：

由于 A 同学曾经进行过对 CNN 模型的研究对此模型较为熟悉，B 和 C 同学也分别具有较好的编程和建模能力，因此他们三人负责根据其他同学的意见和想法对模型进行改进。由 B 同学所学大数据科学专业特点、及其专业基础扎实可知，她擅长进行数据处理，C 和 D 同学擅长 MATLAB 编程，因此他们三人负责探究如何进行数据增强、增广的任务。此外在模型形成后，由 A、C、D 同学进行模型的训练工作，并尝试使用多数表决融合或加权融合等模型融合方法提高识别的准确度。由于 D 同学长期担任班干部职务，具有较多的工作经验，因此他主要负责项目进程记录和提

出模型改善想法的任务。（具体安排根据实际情况而定）

资源整合：

目前，我们准备先学习图像处理的基础理论，并已有较多入门书籍，如《数字图像处理》、《MATLAB 图像处理实例详解》等，导师也会不定期给我们发来前沿论文，指导我们的研究。此外，学校图书馆各种实体、电子资源众多十分方便我们进行资源查阅。

五、项目进度安排（300-500 字，包括进度计划安排，实施阶段和步骤等）

进度计划	时间安排	目标任务
准备阶段	2020 年 4 月下旬---5 月	<p>①理论方面：学习有关黑色素瘤的医学影像特征和检测流程、黑色素瘤图像特征分类方法、基于卷积神经网络（CNN）原理和算法的图像识别和处理技术以及图像增广、迁移学习等技术。</p> <p>②处理工具方面：其一，熟悉 Matlab 和 python 的与图像和数据处理及可视化相关的工具包（如 Tensorflow、Matplotlib 等）；其二，搜集格式和内容较为规范的、合适的数据集，使用适合 CNN 分类模型的数据处理方法对图像数据进行增广。</p>
实施阶段	6-8 月	<p>比对不同 Faster R-CNN、ResNet 等学习模型，使用较为合适的模型搭建基础的 Tensorflow 架构，建立简单的深度学习模型，采用 CNN 半监督学习的方式，对部分增广后</p>

		数据对模型不断进行初步训练，与测试集进行比较。在图像可被较为准确识别的基础上建立有分类机制的系统。
	9-10 月	在模型得到一定训练的基础上尝试使用多数表决融合或加权融合等模型融合方法，提升模型的准确度，并对在训练中发现的问题进行突破。
测试和改进	11 月	对模型进行更细致的研究，进一步优化模型、提升算法。在研究的过程中做好各阶段研究记录的整理。
项目结题	2020 年 12 月初	对研究内容及各阶段的文字资料整理归档，进行成果检查和整合、书写结项书等收尾工作，准备结项。

（各项更为具体的研究工作计划根据实际情况做出安排）

六、项目完成预期成果与形式（可另附页。500-1000 字，请简述：专著或编著、调研或实验报告、论文篇数及论文级别，或专利、设计、产品、服务、公司性质及规模、社会融资或风险投资、营业额和利润额等。）

预期成果：

①探索较为合理的适用于深度学习网络训练的目标检测算法模型。

现有对于图像识别方面的各类方法、工具的研究都已取得了一定的进展，但目前还没有较为完善的、能够被大众认可的深度学习图像分类方法，许多团队和个人都在进行这方面的探索。图像分类前要对图像进行预处理，预处理的结果需要有所偏重且足够细致，会对分类更加有益。由于训练数据集较为匮乏，我们希望通过深入研究、比较和归纳已有的研究，结合一般黑色素瘤所具有的特点，通过数据增强和迁移学习等方式，尝试一些能够实现图像标准化裁剪、图像层次化分解、图像精准识别等功能且占据主流的网络模型预处理方案，选择合适的算法模型（如 Faster R-CNN、ResNet 等）处理照片数据，用于卷积神经网络的训练。作为方法研究的具体实例和载体，我们也希望使用具有特定体系结构特征的 CNN，找到一种较为高效且能够被普遍接受的参数模型，结合现有技术对模型进行

图像分类识别的应用和实践。同时，在研究和实践中，我们将理解模型如何完成对于黑色素瘤图像的辨别和图像处理，并对其的优点与不足进行讨论。因此，我们将在此基础上探究如何改进，细化这一过程，并将后期图像分类所需的指标作量化处理，更便于图像分类，方便医生诊断。

②使用基于 CNN 的神经网络模型对黑色素瘤图像识别、处理和分类。

采用不同的算法、不同的参数将模型的训练效果进行数据和图像等多角度对比，以得出对于黑色素瘤图像识别、处理和分类的可行度较高、较为合适的方案。在卷积神经网络经过一定训练的基础上，尝试使用多数表决融合或加权融合等模型融合方法，提高判断的准确度、进一步优化算法。

预期形式：

我们将把取得的成果以结项论文的形式进一步整理、汇报。

七、项目成员承诺

在获得立项后，本人将与项目组成员团结一致，努力做好该项目的研究及实施工作，实现制定的目标。如果因主观原因导致项目执行不力，未达到预期目标，本人与项目组成员愿意承担相应损失，并接受学校相应处理。

项目全体成员签名：

年 月 日

八、经费预算

科目名称	预算经费 (单位：元)	备注（预算依据与具体说明）
论文翻译及印刷	200	翻译软件（英译汉）付费

九、项目审批

指导教师意见（需附具体的指导计划）：

本项目针对目前皮肤癌皮损图像训练数据集匮乏且分类精度不高问题进行研究。由于公开的皮肤癌训练数据集较少，并且皮损图像病变部位干扰因素诸多，使得利用机器学习方法来对皮肤癌图片进行分类成为目前研究的难点和热点。该项目利用卷积神经网络来进行病理图片的分类，并对存在的问题进行研究，旨在提高分类精度，得到较为适合的分类模型。

指导计划如下：

2020.4.1-2020.6.30 指导学生学习有关黑色素瘤的医学影像特征和检测流程、黑色素瘤图像特征分类方法、基于卷积神经网络（CNN）原理和算法的图像识别和处理技术。

2020.7.1-2020.7.30 指导学生对当前最新发表的基于卷积神经网络的皮肤癌图像分类方法进行研究，提出改进的方法。

2020.8.1-2020.10.30 指导学生对改进的方法进行具体的实现，训练模型、对比结果。

2020.11.1- 2020.11.31 指导学生对项目进行归纳，总结，撰写结项报告书。

签名： 吕鑫

2020 年 3 月 27 日

学院团委意见：

签章：

年 月 日

备
注