

华中科技大学

2023

计算机视觉实验

· 实验报告 ·

专    业：          计算机科学与技术

班    级：          CS2106

学    号：          U202115514

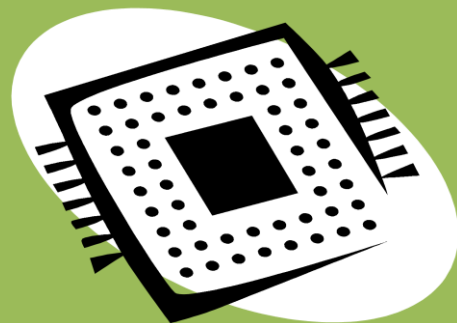
姓    名：          杨明欣

电    话：          13390396012

邮    件：          ymx@hust.edu.cn

指导老师：          刘康

完成日期：          2023-12-27



计算机科学与技术学院

## 目录

1	实验内容 .....	2
1.1	实验要求 .....	2
1.2	实验内容 .....	2
1.2.1	卷积神经网络的可解释性 .....	2
1.2.2	Grad-CAM 算法 .....	2
1.2.3	LayerCAM 算法 .....	3
1.3	实验方法 .....	4
1.3.1	网络架构分析 .....	4
1.3.2	可解释性分析 .....	4
1.4	实验设计 .....	6
1.4.1	数据集 .....	6
1.4.2	实验环境 .....	6
1.5	实验分析与结果 .....	7
1.5.1	Grad-CAM 实验结果 .....	7
1.5.2	LayerCAM 实验结果 .....	9
1.6	实验总结 .....	12
2	实验心得 .....	14
3	附录 .....	15
3.1	数据集 .....	15
3.2	Grad-CAM 对 Maxpooling 层的可解释性分析 .....	16

## 1 实验内容

### 1.1 实验要求

针对已训练好的卷积神经网络，给定一张输入图片，生成该图片对于特定类别的可解释性分析结果。

具体要求：实验将提供基于 PyTorch 和 TensorFlow 的两个不同版本的二分类模型，该模型可用于猫和狗的分类（class 0 为猫，class 1 为狗）。注：PyTorch 使用的网络架构是 AlexNet，TensorFlow 使用的是 VGG16，两者略有不同，请任选一个模型进行实验。

实验将同时提供三张输入图片，对于每张图片，分别针对猫和狗的类别，进行 Grad-CAM 和 LayerCAM 的可解释性分析。

### 1.2 实验内容

#### 1.2.1 卷积神经网络的可解释性

卷积神经网络的可解释性是指对网络中各层的输出进行可视化，从而理解网络如何做出决策的过程。对于一个深层的卷积神经网络，通过多次卷积和池化后，其最后一层卷积层包含了最丰富的空间和语义信息。然而，再往下的全连接层和 softmax 层所包含的信息是难以被人类理解的，也很难以可视化的方式展示出来。

为了提高卷积神经网络的可解释性，一种方法是将全连接层替换为全局平均池化（GAP）。这种方法被称为类别置信度映射（CAM）。通过这种方式，可以更好地理解网络的分类结果。

#### 1.2.2 Grad-CAM 算法

Grad-CAM 是一种独特的可视化技术，它能帮助我们深入理解深度神经网络在预测过程中的关键决策点。通过定位到图像的特定区域，Grad-CAM 使我们能够直观地揭示神经网络如何进行分类，从而使决策过程更加透明。

# 华中科技大学课程实验报告

其核心思想在于，神经网络的最后一个卷积层所输出的特征图，对分类结果起到了决定性的作用。因此，我们可以通过对这一层的梯度进行全局平均池化，从而计算出每个通道的权重。这些权重可以进一步用来加权特征图，生成一个 Class Activation Map (CAM)。在这张 CAM 中，每个像素都反映了该特定区域对于分类结果的重要性。其模型架构如图 1-1 所示。

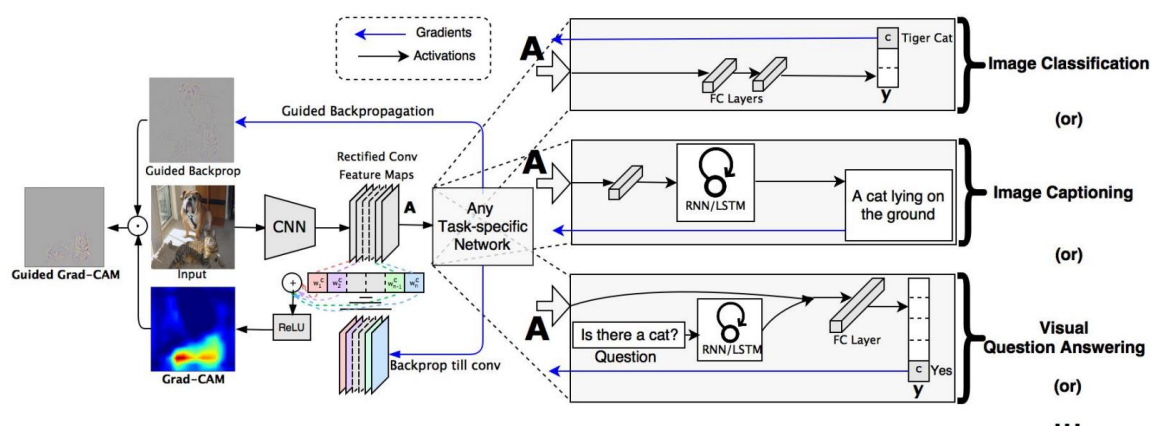


图 1-1 Grad-CAM 整体模型框架示意图

与传统的 CAM 方法相比, Grad-CAM 具有更广泛的适用性。它无需修改网络结构, 也不依赖于特定的层结构, 因此可以应用于任何类型的神经网络。此外, Grad-CAM 不仅限于对特征的可视化, 还能深入分析网络中的特定层或单元。

总的来说, Grad-CAM 提供了一种有力的工具, 帮助我们理解深度神经网络的决策过程, 使这一复杂的过程变得可视化、可解释。

## 1.2.3 LayerCAM 算法

LayerCAM 算法是一种针对卷积神经网络的可解释性算法, 旨在解决深层特征图大小不够细粒度的问题。该算法通过计算输入图像在经过卷积层操作后, 在最后一层卷积层输出的特征图的梯度, 得到每个通道的权重。这些权重被用来加权特征图, 生成一个 Class Activation Map (CAM), 其中每个像素都代表了该像素区域对于分类结果的重要性。与传统的 CAM 方法相比, LayerCAM 算法提供了像素级的权重, 而不是特征图级的权重, 使得神经网络的决策过程更加透明和可解释。

## 1.3 实验方法

为了实现针对已训练好的卷积神经网络，给定一张输入图片，生成该图片对于特定类别的可解释性分析结果，本实验主要分成两个部分，第一个部分是分析已经训练好的卷积神经网络 Alexnet 的网络架构，同时获取三张图片的预测结果，第二部分是实现 Grad-CAM 算法和 LayerCAM 算法对卷积神经网络框架进行可解释性分析。

### 1.3.1 网络架构分析

本实验中已训练好的卷积神经网络框架为 AlexNet，原始的 AlexNet 网络结构相对简单，使用了 8 层卷积神经网络，前 5 层是卷积层，剩下的 3 层是全连接层。具体结构示意图如图 1-2 所示。

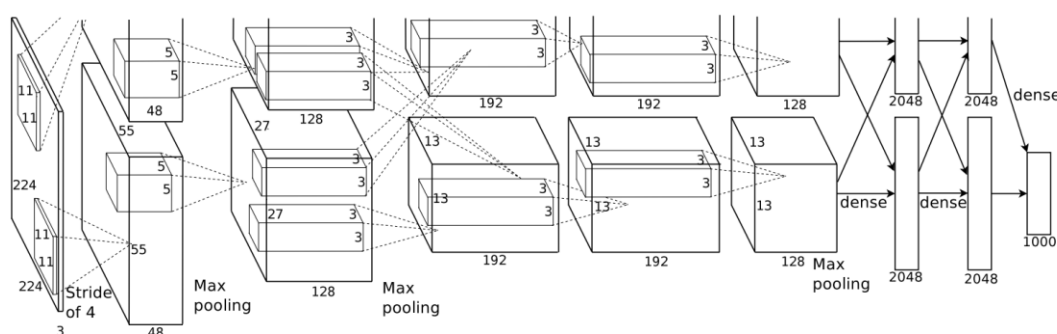


图 1-2 AlexNet 网络架构示意图

同时我们利用预训练好的模型进行三张图片的分类，最终三张图片的分类结果为 both.jpg 分类的结果为 1 标签，对应为狗，cat.jpg 分类的结果为 0 标签，对应为猫，dog.jpg 分类的结果为 1 标签，对应为狗，通过上面的结果可以得知模型在三张图片的分类任务上达到了很高的准确率。

在接下来的实验中我们要完成的就是对于三张图片在 AlexNet 网络中的可解释性分析。

### 1.3.2 可解释性分析

在部署的模型上完成分类任务后，我们可以获得 AlexNet 网络中最后一层卷积层（包括激活函数激活）对于三张图片的 256 个通道的输出。

both.jpg 图片对应最后一层卷积层的输出结果可视化如图 1-3 所示。

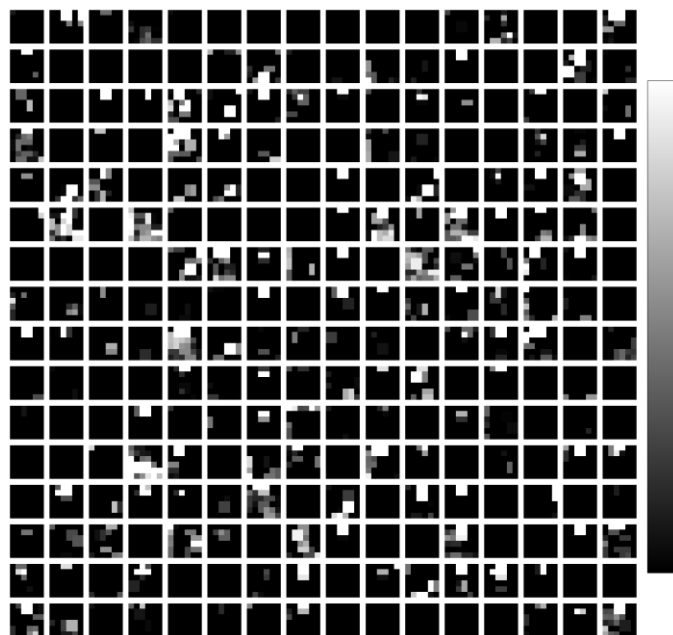


图 1-3 both.jpg 图片对应最后一层卷积层的输出结果

cat.jpg 图片对应最后一层卷积层的输出结果可视化如图 1-4 所示。

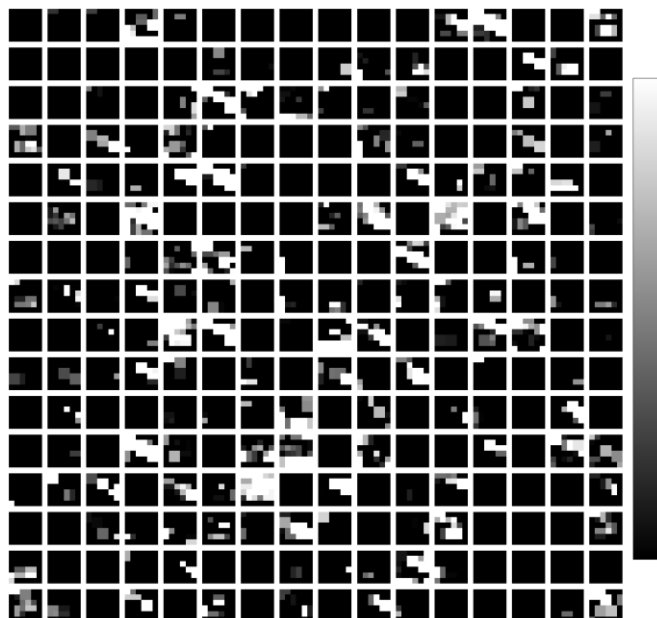


图 1-4 cat.jpg 图片对应最后一层卷积层的输出结果

dog.jpg 图片对应最后一层卷积层的输出结果可视化如图 1-5 所示。

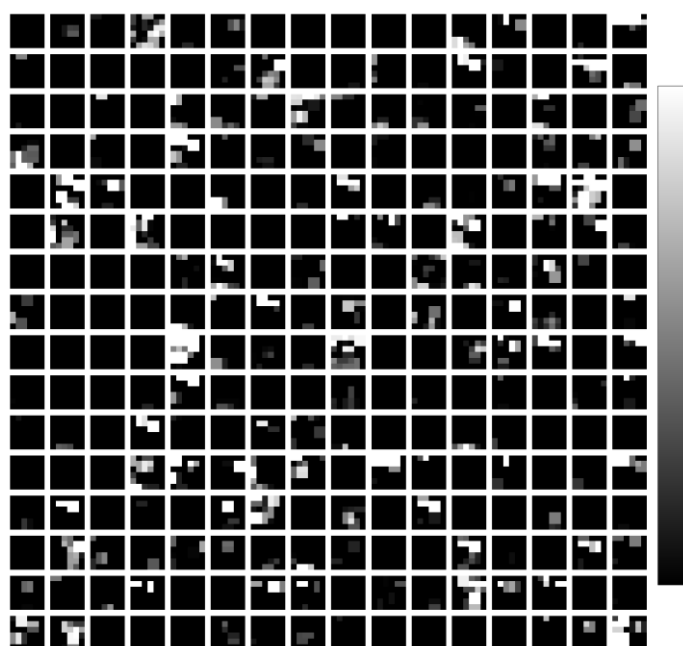


图 1-5 dog.jpg 图片对应最后一层卷积层的输出结果

进一步我们将要根据 Grad-CAM 算法和 LayerCAM 算法原理对卷积神经网络框架进行可解释性分析，同时将结果可视化展现在实验结果部分。

## 1.4 实验设计

### 1.4.1 数据集

本实验针对三张图片进行可视化分析，三张图片请见附录部分。

### 1.4.2 实验环境

系统版本：Windows 11 专业版

处理器：13th Gen Intel(R) Core(TM) i9-13900HX 2.20 GHz

机带 RAM：16.0 GB (15.7 GB 可用)

系统类型：64 位操作系统, 基于 x64 的处理器

显卡版本：NVIDIA GeForce RTX 4060 Laptop GPU

显存：8188MiB

CUDA 版本：12.0

Python 版本：3.9.18

torch 版本: 2.1.1+cu118

## 1.5 实验分析与结果

### 1.5.1 Grad-CAM 实验结果

此部分主要展示 Grad-CAM 算法对于 AlexNet 最后一层卷积层在三张图片上的可解释性分析（附录展示了最后一层平均池化层的结果，没有最后一层卷积层的效果好，轮廓感不清晰，因此接下来仅使用最后一层卷积层）。

对于 both.jpg 图片模型的预测标签为 1，即狗标签，观察这张图片可知，上面同时有一只猫和一只狗。首先对于这张图片在 1 标签上的可视化分析结果如图 1-6 所示。

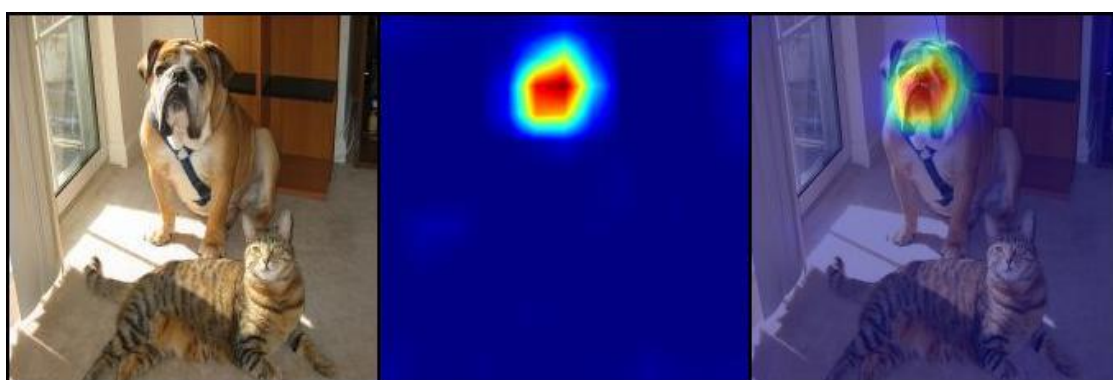


图 1-6 both.jpg 图片在 1 标签上的可视化分析

通过上面的图片可以看出，卷积层正确关注到了狗的面部特征，模型的可解释性较好。进一步对于这张图片在 0 标签上的可视化分析结果如图 1-7 所示。

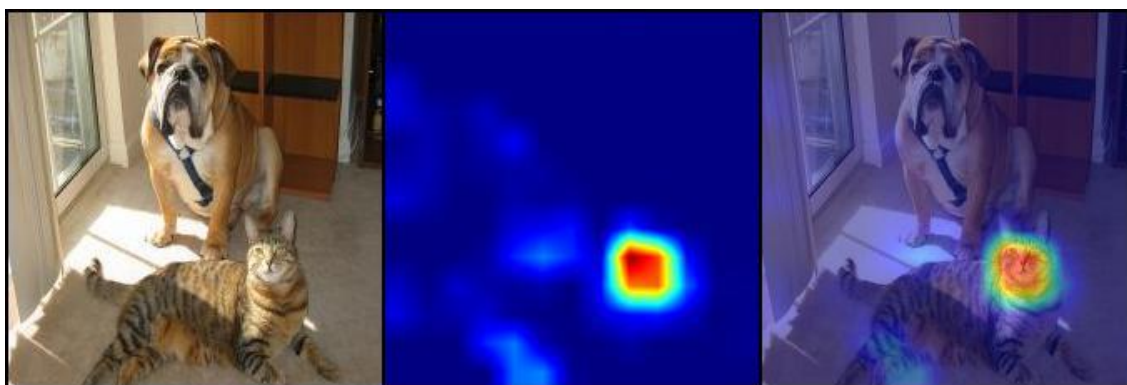


图 1-7 both.jpg 图片在 0 标签上的可视化分析

通过上面的图片可以看出，虽然模型没有将这张图片分类为标签 0（即猫），但是



卷积层正确关注到了猫的面部特征，进一步证明了模型的可解释性较好。

对于 cat.jpg 图片模型的预测标签为 0，即猫标签，观察这张图片可知，上面有 3 只猫。首先对于这张图片在 0 标签上的可视化分析结果如图 1-8 所示。

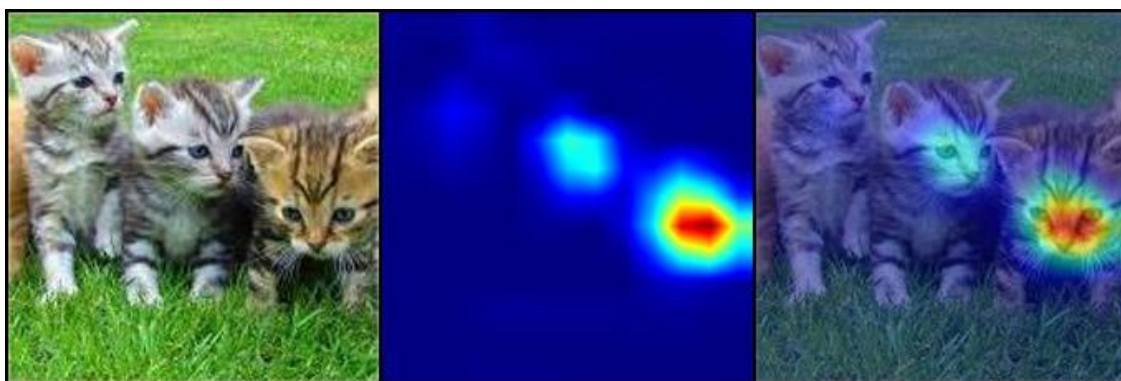


图 1-8 cat.jpg 图片在 0 标签上的可视化分析

通过上面的图片可以看出，卷积层正确关注到了猫的面部特征，模型的可解释性较好。进一步对于这张图片在 1 标签上的可视化分析结果如图 1-9 所示。

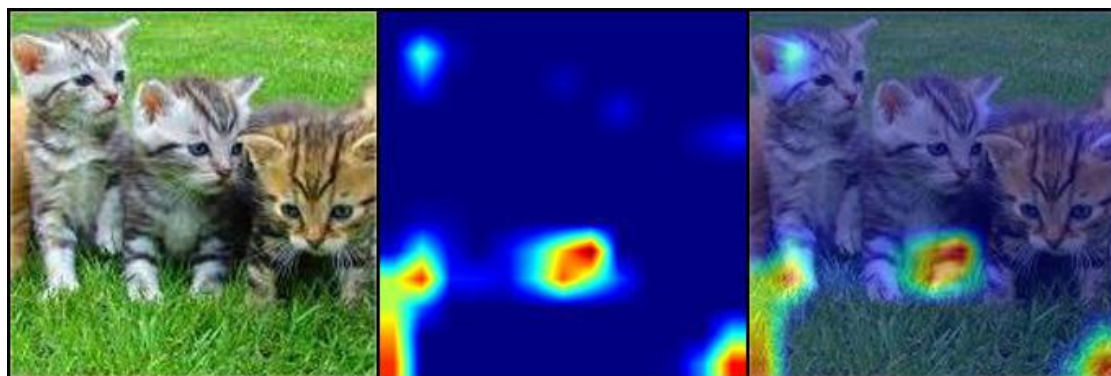


图 1-9 cat.jpg 图片在 1 标签上的可视化分析

通过上面的图片可以看出，卷积层的关注点发散，同时聚焦的位置没有太多的有效信息，无法对于预测错误标签进行合理的解释，反过来证明了模型的可解释性较好。

对于 dog.jpg 图片模型的预测标签为 1，即狗标签，观察这张图片可知，上面有 2 只狗。首先对于这张图片在 1 标签上的可视化分析结果如图 1-10 所示。

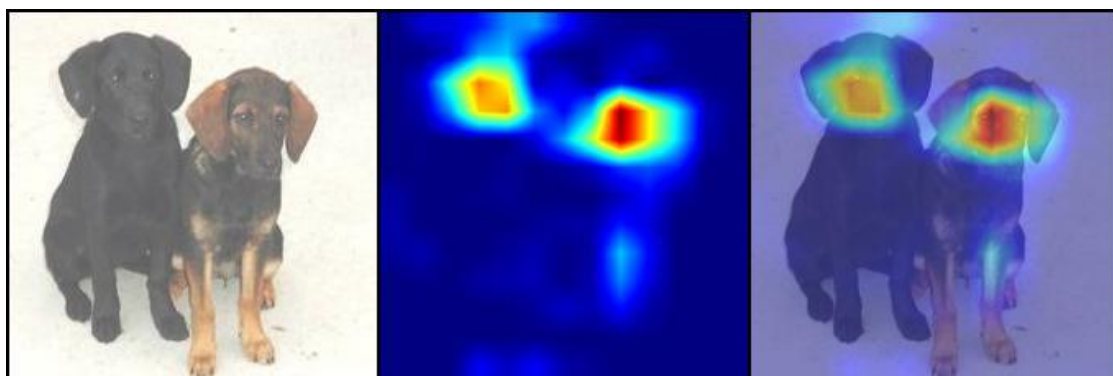


图 1-10 dog.jpg 图片在 1 标签上的可视化分析结果

通过上面的图片可以看出，卷积层正确关注到了狗的面部特征，模型的可解释性较好。进一步对于这张图片在 0 标签上的可视化分析结果如图 1-11 所示。

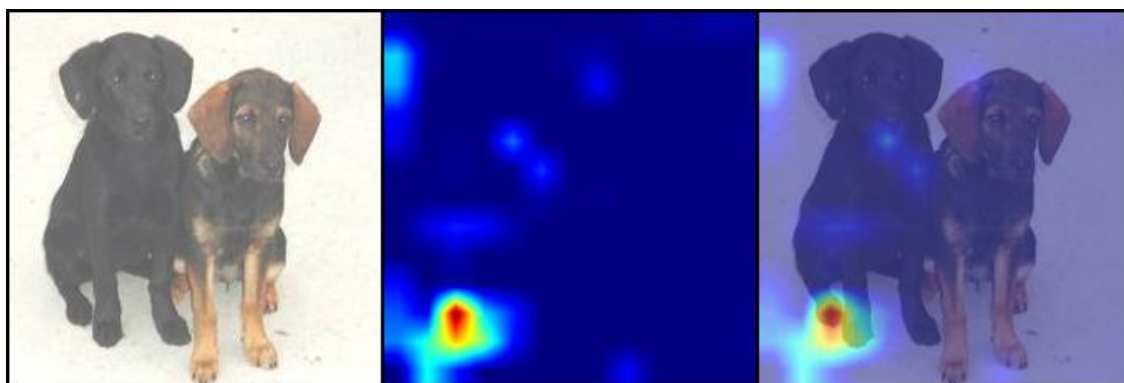


图 1-11 dog.jpg 图片在 0 标签上的可视化分析结果

通过上面的图片可以看出，卷积层的关注点发散，同时聚焦的位置没有太多的有效信息，无法对于预测错误标签进行合理的解释，反过来证明了模型的可解释性较好。

总的来说，Grad-CAM 很好地完成了在 AlexNet 网络上对于三张图片的可解释性分析并且获得了正确的分析结果。

## 1.5.2 LayerCAM 实验结果

此部分主要展示 LayerCAM 算法对于 AlexNet 最后一层卷积层在三张图片上的可解释性分析。

对于 both.jpg 图片模型的预测标签为 1，即狗标签，观察这张图片可知，上面同时有一只猫和一只狗。首先对于这张图片在 1 标签上的可视化分析结果如图 1-12 所示。

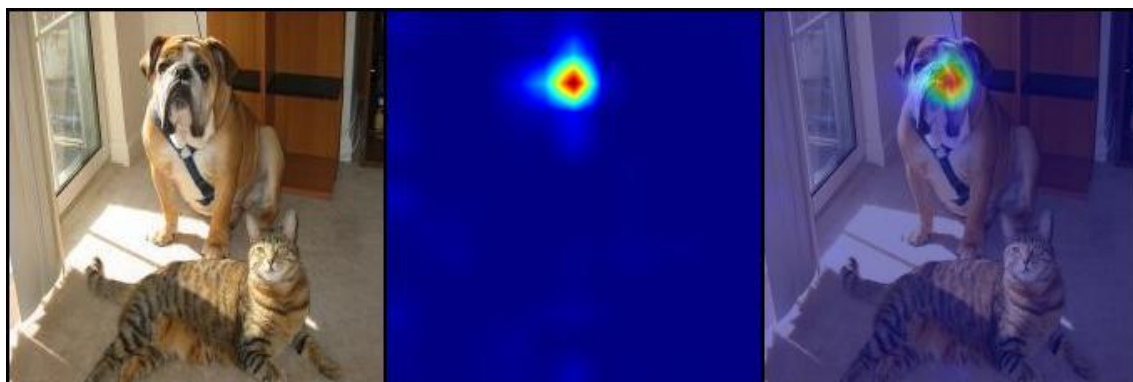


图 1-12 both.jpg 图片在 1 标签上的可视化分析

通过上面的图片可以看出，卷积层正确关注到了狗的面部特征，模型的可解释性较好。进一步对于这张图片在 0 标签上的可视化分析结果如图 1-13 所示。

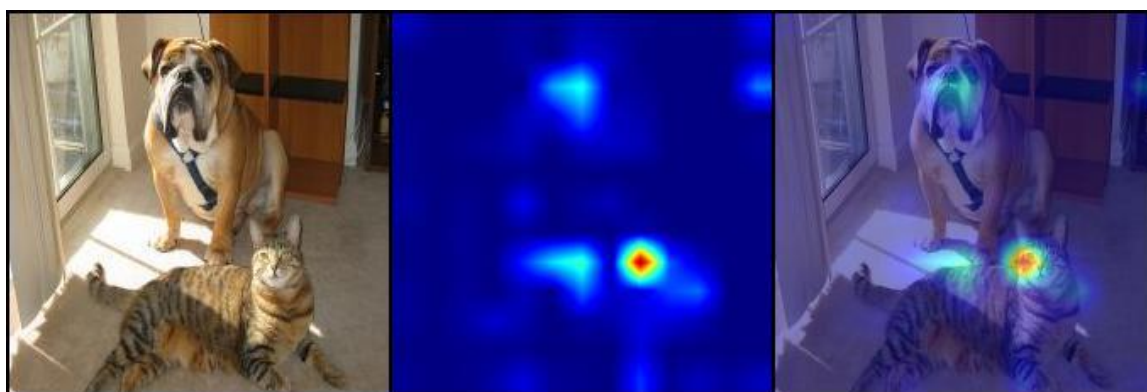


图 1-13 both.jpg 图片在 0 标签上的可视化分析

通过上面的图片可以看出，虽然模型没有将这张图片分类为标签 0（即猫），但是卷积层正确关注到了猫的面部特征，进一步证明了模型的可解释性较好。

对于 cat.jpg 图片模型的预测标签为 0，即猫标签，观察这张图片可知，上面有 3 只猫。首先对于这张图片在 0 标签上的可视化分析结果如图 1-14 所示。



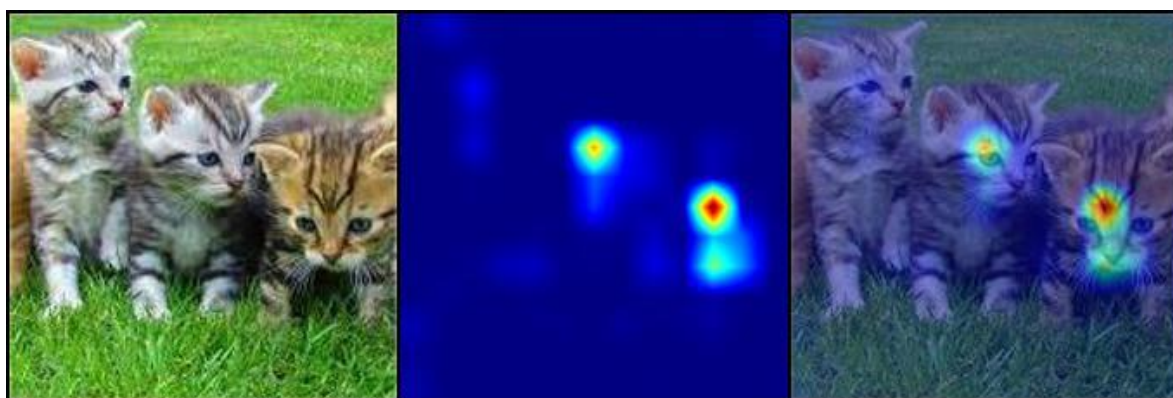


图 1-14 cat.jpg 图片在 0 标签上的可视化分析

通过上面的图片可以看出，卷积层正确关注到了猫的面部特征，模型的可解释性较好。进一步对于这张图片在 1 标签上的可视化分析结果如图 1-15 所示。

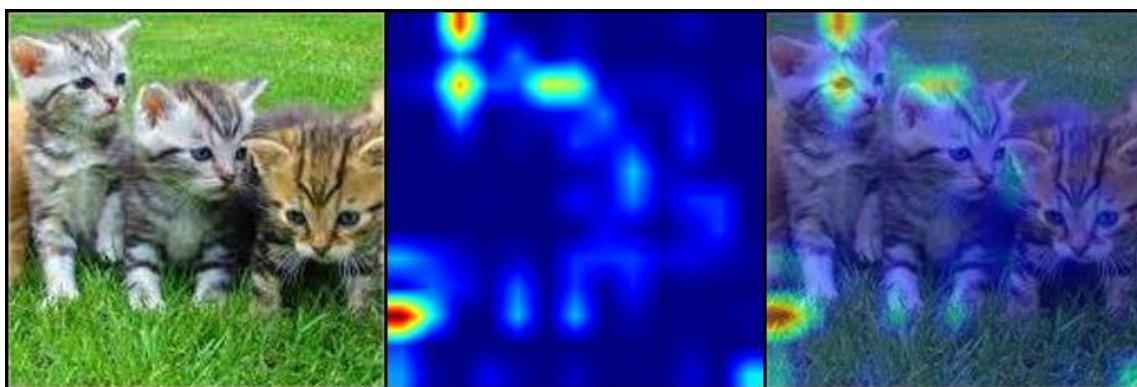


图 1-15 cat.jpg 图片在 1 标签上的可视化分析

通过上面的图片可以看出，卷积层的关注点发散，同时聚焦的位置没有太多的有效信息，无法对于预测错误标签进行合理的解释，反过来证明了模型的可解释性较好。

对于 dog.jpg 图片模型的预测标签为 1，即狗标签，观察这张图片可知，上面有 2 只狗。首先对于这张图片在 1 标签上的可视化分析结果如图 1-16 所示。

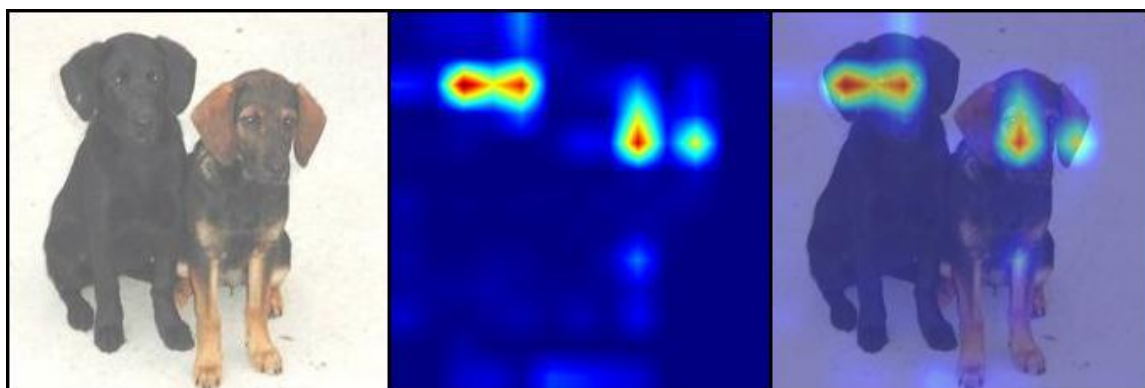


图 1-16 dog.jpg 图片在 1 标签上的可视化分析结果

通过上面的图片可以看出，卷积层正确关注到了狗的面部特征，模型的可解释性较好。进一步对于这张图片在 0 标签上的可视化分析结果如图 1-17 所示。

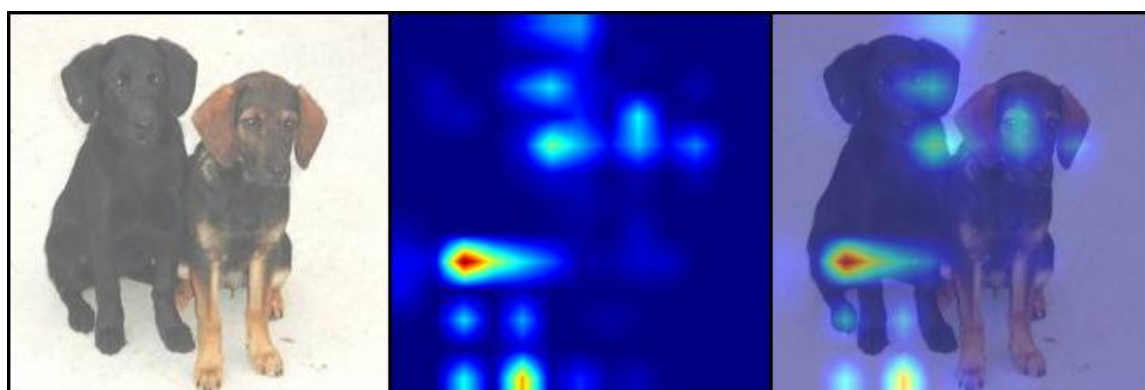


图 1-17 dog.jpg 图片在 0 标签上的可视化分析结果

通过上面的图片可以看出，卷积层的关注点发散，同时聚焦的位置没有太多的有效信息，无法对于预测错误标签进行合理的解释，反过来证明了模型的可解释性较好。

总的来说，LayerCAM 很好地完成了在 AlexNet 网络上对于三张图片的可解释性分析并且获得了正确的分析结果。

## 1.6 实验总结

本次实验主要完成了如下几点工作：

- 1) 针对已训练好的卷积神经网络，给定一张输入图片，生成该图片对于特定类别的可解释性分析结果。
- 2) 设计并实现了 Grad-CAM 算法对于 AlexNet 最后一层卷积层在三张图片上的

# 华中科技大学课程实验报告

---

可解释性分析。

- 3) 设计并实现了 LayerCAM 算法对于 AlexNet 最后一层卷积层在三张图片上的可解释性分析。

## 2 实验心得

在完成卷积神经网络的可解释性分析的实验过程中，我获得了许多宝贵的经验。

- 1) 对于卷积神经网络的可解释性分析有了更加深刻的认识。目前，缺乏可解释性仍然是在许多应用中采用深层模型的一个关键障碍，通过提高卷积神经网络的可解释性，人们可以更好地理解网络的内部工作机制和决策过程。这有助于增强人们对网络决策的信任，并使人们能够更好地识别和纠正任何潜在的偏见或歧视。此外，可解释性还有助于发现网络中的漏洞或弱点，从而提高网络的安全性和可靠性。
- 2) 学会了通过 Grad-CAM 和 LayerCAM 算法对卷积神经网络模型进行可解释性分析。

总的来说，这次实验让我对卷积神经网络的可解释性分析有了更深入的了解。我不仅学会了对于训练好的模型进行调用，更加认识到了卷积神经网络可解释性分析的重要意义，收获很大。

## 3 附录

### 3.1 数据集



图 3-1 both.jpg



图 3-2 cat.jpg



图 3-3 dog.jpg



## 3.2 Grad-CAM 对 Maxpooling 层的可解释性分析

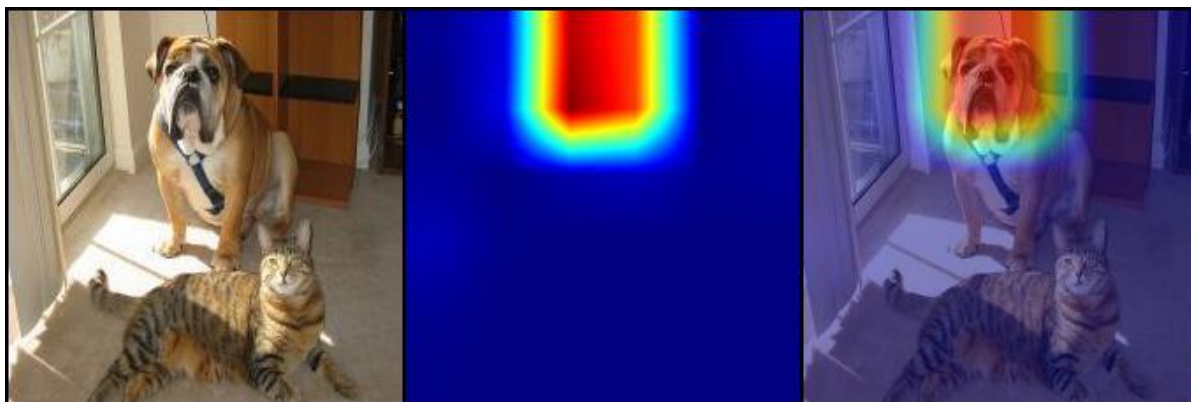


图 3-4 both.jpg 图片在 1 标签上的可视化分析

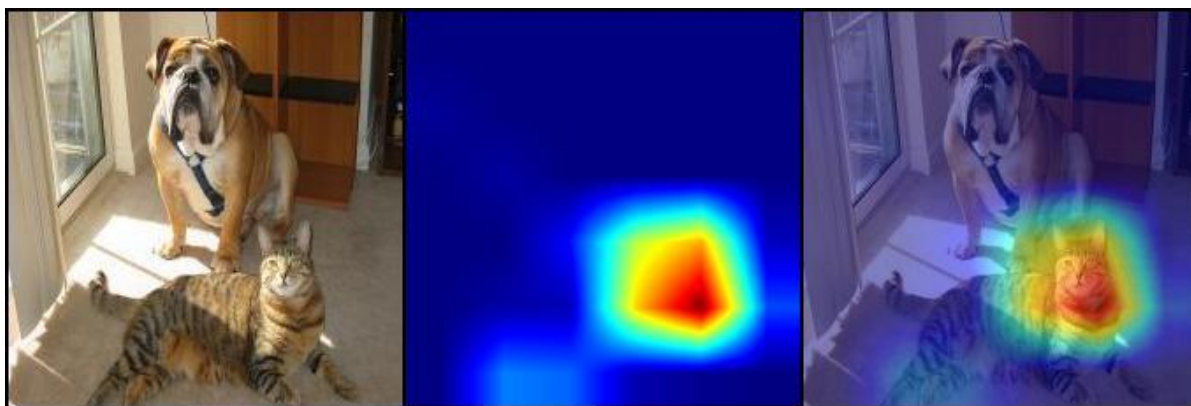


图 3-5 both.jpg 图片在 0 标签上的可视化分析

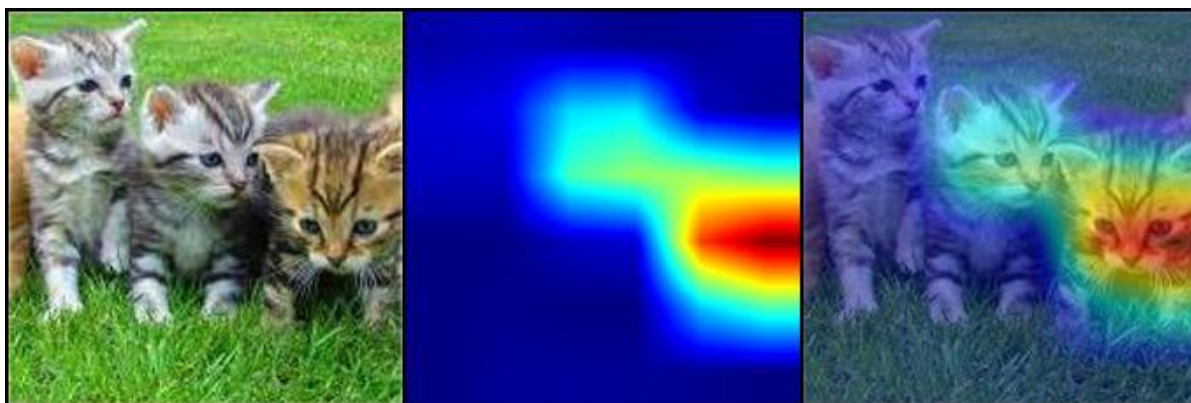


图 3-6 cat.jpg 图片在 0 标签上的可视化分析

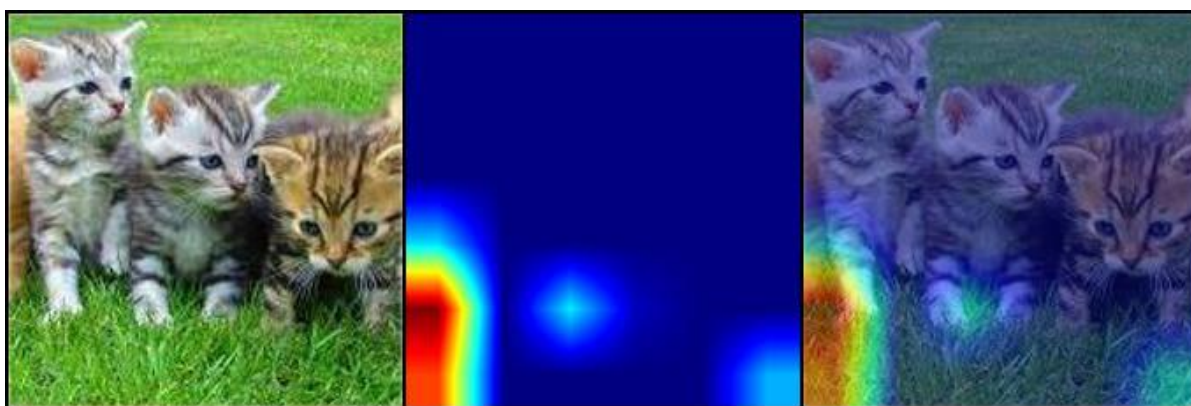


图 3-7 cat.jpg 图片在 1 标签上的可视化分析

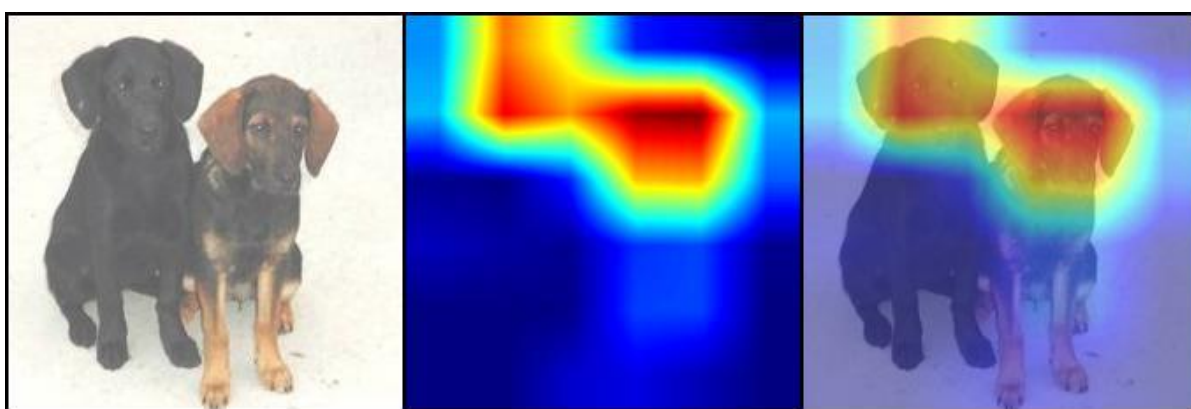


图 3-8 dog.jpg 图片在 1 标签上的可视化分析

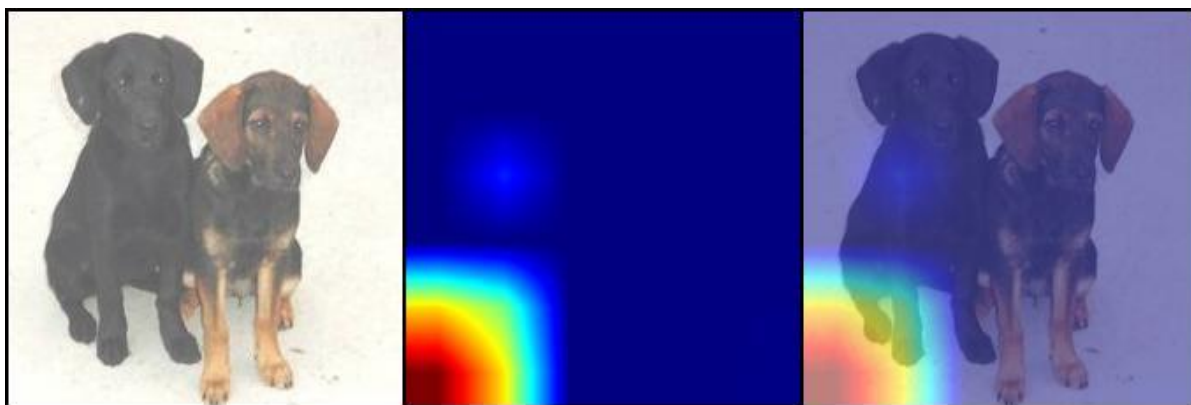


图 3-9 dog.jpg 图片在 0 标签上的可视化分析

## • 指导教师评定意见 •

### 一、原创性声明

本人郑重声明本报告内容，是由作者本人独立完成的。有关观点、方法、数据和文献等的引用已在文中指出。除文中已注明引用的内容外，本报告不包含任何其他个人或集体已经公开发表的作品成果，不存在剽窃、抄袭行为。

特此声明！

作者签字：杨明欣



### 二、对课程实验的学术评语（教师填写）

### 三、对课程实验的评分（教师填写）

评分项目 (分值)	课程目标 1 工具应用 (10 分)	课程目标 2 设计实现 (70 分)	课程目标 3 验收与报告 (20 分)	最终评定 (100 分)
得分				

指导教师签字：\_\_\_\_\_