

2025—2026 秋季学期

机器学习课程设计

实验报告

评 分	
--------	--

小组成员: 周炜杰 李天牧

学 号: 20231610 20233098

整体工作占比: 周炜杰 60% 李天牧 40%

专业班级: 人工智能 2301

学 院: 计算机科学与工程学院

任课教师: 曹鹏 吴宏林 曹春红

课设代码仓库

https://github.com/Double-Z-wj/Machine_Learning_Course

工作分配

	具体工作内容	整体工作占比
周炜杰	<p>任务一：建机器学习流水线，完成基于 SVM 的初版分类器，并主导特征工程优化。</p> <p>任务二：搭建 ResNet/CNN 基础训练框架，确立性能 Baseline。</p> <p>任务三：完成面部表情分类的 CNN 基线构建，并进行初步超参数调整和部分优化。</p> <p>任务四：调研并尝试部署 vascx models 等 SOTA 方案进行可行性验证。并辅助优化模型。</p> <p>任务五：主导 FR-UNet 全分辨率模型的复现与工程适配，解决分辨率对齐问题，完成多种模型的对比实验。</p>	60%
李天牧	<p>任务一：完成数据增强策略的改进，协助进行特征选择。</p> <p>任务二：在基线基础上引入 DenseNet、Mixup 等策略，解决过拟合问题，产出最终高分方案。</p> <p>任务三：引入强数据增强方法，显著提升模型在验证集上的鲁棒性。</p> <p>任务四：构建中心凹定位的回归基线 (UNet)，并引入重心法等后处理策略提升定位精度。</p> <p>任务五：尝试构建血管分割模型，辅助复现模型</p>	40%

任务一：植物分类（特征工程）

一、任务重述

使用纯机器学习的方法（不包括深度学习）完成植物分类任务。

数据集包含 5 种植物，每种植物 100 张图片。类别如下

‘Black-grass’ ‘Common wheat’ ‘Loose Silky-bent’ ‘Scentless Mayweed’
‘Sugar beet’。

二、实验与分析过程

首先分析数据集，站在 2025 年看，这份数据集有如下特点：

数据量极小：总共仅约 500 张图片，不过用于特征工程足够了。

背景噪声大：图片背景包含大量碎石、土壤，部分图片还包含白色的标尺或条形码，信噪比低。

类间相似度高：Black-grass 和 Loose Silky-bent 同属禾本科，形态极度相似，仅在茎部颜色和微观纹理上有细微差别。人眼几乎无法分辨区别。

这份数据集源于 Kaggle 的 “Plant Seedlings Classification” 竞赛（子集）(2017 年)。但在 2017 年，高分方案几乎都是使用深度学习的，很难找到一份对我们有借鉴意义的方案，因此我们需要从零开始构建方案。

本实验采用了图像预处理→数据增强→多维特征工程→模型训练的流水线策略。

做机器学习，最重要的就是先 train 一发试试能不能行，建立一个基线模型然后改进。首先是建立一个基准模型，我使用了 SVM 作为基础分类器，使用基本的 HSV 颜色直方图统计颜色分布和简单的 Flatten 像素提取特征。针对背景中的碎石颜色各异，且存在苔藓等绿色干扰，导致特征提取不纯净的问题，我使用了 HSV 阈值掩码将植物主体从背景提取出来。至此，我完成了初版代码。

在初版代码的基础上，我加入了几何形态特征和数据增强，在特征加入了计算叶片的实心度、长宽比、延展度，这对于区分阔叶和细草非常有效。在读取训练集时，对每张图片进行翻转和旋转，将训练样本量扩大了 6 倍，提升模型的泛化能力。获取了第二版代码。

之后，我通过分析数据集图像信息，继续改进。原始图像背景由复杂的碎石组成，部分石块表面附着苔藓或因光照产生偏绿色的噪点。仅依靠简单的 HSV 颜色阈值分割，极易将这些背景噪点误判为植物像素，导致后续计算的纹理和形状特征混入大量噪声。因此，我引入了形态学开运算。通过该操作，可以有效滤除面积较小、孤立的绿色背景噪点，同时保留面积较大的植物主体结构，显著提高了感兴趣区域的纯净度。通过对训练集样本的色彩空间分布进行统计分析，我们调整了 HSV 阈值策略（提高饱和度 S 和亮度 V 的下限）。确保了算法仅提取色泽鲜艳、特征显著的植物叶片部分，剔除了灰暗的阴影和发白的过曝区域，使提取的颜色直方图更能代表植物本身的色素特征。

于此同时，我求助 AI 工具，引入植物学先验知识，Scentless Mayweed（无味母菊）与 Sugar beet（甜菜）在颜色和纹理上差异不大，是分类的难点。观察发现，Sugar beet 的叶片呈完整的卵形或勺状，边缘光滑，其轮廓面积与凸包面积极为接近；而 Scentless Mayweed 的叶片呈羽状分裂，边缘破碎且分叉多，其轮廓面积远小于凸包面积。引入实心度作为量化指标。Sugar beet 的实

心度通常接近 1.0，而 Scentless Mayweed 则显著较低。该特征将复杂的形态差异转化为单一的数值差异，极大地提升了 SVM 对这两类植物的可分性。禾本科植物（如 Black-grass, Common wheat）形态均为细长条状，纹理相似，极易混淆。通过仔细观察高分辨率样本，发现 Black-grass（黑草）的茎基部在幼苗期呈现出特征性的紫色或红褐色，而 Common wheat（小麦）通常通体呈现纯绿色。紫红斑占比针对性地计算 HSV 空间中紫色和红色色相区间的像素在植物总像素中的占比。这一特征捕捉了黑草独有的生物学标记，使其能从形态高度相似的草类植物中被准确识别。总而言之，引入了实心度与紫红斑占比等植物学特征，提升模型能力。通过以上改进，我使用第三版代码获取了 0.85 的准确率。

之后，我又进行了一些不成功的改进。首先，通过划分验证集，我发现当前模型在识别 Black-grass 和 Loose Silky-bent 时极易混淆，而其他植物的分类几乎没错。换言之，所有的错误几乎都是 Black-grass 和 Loose Silky-bent 的错分。我引入了分层分类、以及 ExG 指数等特征，尝试了多种方法，效果都不好，最终也没能超过之前对方法。

三、 最终使用的模型

使用 SVM（支持向量机）作为核心分类器，因为它在小样本、高维特征数据上表现最佳。

3.1 模型架构

特征向量：由颜色直方图、特殊颜色占比、GLCM 纹理、Hu 矩、几何形态特征拼接而成，经过 StandardScaler 进行标准化处理，消除量纲影响。

分类器： sklearn.svm.SVC

核函数： RBF（径向基函数）。由于特征在高维空间非线性可分，RBF 核能有效将数据映射到高维空间进行划分。

3.2 图像预处理

尺寸归一化： 将所有图像统一缩放至 256×256 。

HSV 阈值掩膜：

利用 cv2.inRange 设定阈值：lower=[30, 50, 50], upper=[90, 255, 255]。阈值设定较为激进，过滤掉了低饱和度（石头、标尺）和低亮度（阴影）的像素。虽然牺牲了部分植物边缘，但最大程度地保证了提取特征的纯净度（即只提取最鲜艳的绿色和部分紫色区域）。

形态学操作： 使用开运算去除噪点，闭运算填补叶片空洞。

3.3 特征工程

我们将多种特征拼接成一个高维向量：

颜色直方图 (Color Histograms)：

在 HSV 空间提取 Hue (60 bins), Saturation (16 bins), Value (16 bins) 直方图。Hue 直方图是分类的核心，不同植物的绿色有细微差异。

特异性颜色特征：

统计 Mask 区域内紫色/深红色像素的占比。专门用于区分 Black-grass（茎部带紫）和 Sugar beet（茎部带红）。

纹理特征：

基于灰度图计算灰度共生矩阵，提取 Contrast, Dissimilarity, Homogeneity, Energy, Correlation。捕捉叶片表面的纹理细节。

形状与几何特征：

Hu Moments: 7 个不变矩，具有旋转、缩放不变性。

Solidity (实心度), Aspect Ratio (长宽比), Extent。有效区分阔叶植物（圆润、实心度高）和禾本植物（细长）。

3.4 参数优化

使用了 GridSearchCV (网格搜索) 进行超参数调优：

C (惩罚系数): 最终设定为较大值 (如 100 或 500)，以在训练集准确率和泛化能力之间寻找平衡。

Gamma: 设为 scale，自动适应特征方差。

3.5 训练流程

1. 输入扩充后的 3000 张图片特征。
2. 使用 5 折分层交叉验证 (Stratified K-Fold) 评估模型稳定性，优化 F1-Macro 指标。
3. 使用全量数据重新训练最佳参数模型，并对测试集进行预测。

四、总结

4.1 实验结论

1. **最终性能:** 基于 HSV 颜色直方图与 SVM 的模型在测试集上达到了 0.85 的 F1-Score，为所有尝试方案中的最高分。
2. **特征有效性:** 在碎石背景干扰严重的数据集中，颜色特征 (尤其是 Hue 分布) 是最稳健的分类依据。复杂的形状 (HOG) 和纹理 (LBP) 特征容易受到背景噪声的污染，反而降低模型性能。

4.2 思考与感悟

在简单的小规模任务下，传统机器学习、特征工程的方法仍能达到一个不错的效果。

特征工程依赖大量的先验知识，需要大量专业知识，而深度学习通过“力大砖飞”规避了这一点。

综上所述，解决机器学习问题的关键不在于堆砌复杂的算法，而在于对数据特征的深刻理解以及选择最适合当前数据分布 (噪声、样本量) 的处理策略。

任务二：植物分类（特征学习）

一、任务重述

任务二与任务一的任务和数据集都是一样的，唯一不同的是任务二允许使用深度学习的方法。

二、实验与分析过程

本次实验旨在探索深度学习在小样本、细粒度植物图像分类任务中的应用。考虑到数据集源于 Kaggle 的 “Plant Seedlings Classification” 竞赛（子集），且原始竞赛历史较久，许多早期的 CNN 方案（如 AlexNet、VGG）在特征提取效率上已不如当前的主流模型。因此，我们并未直接沿用历史方案，而是从数据特性出发，逐步构建并优化了一套现代化的深度学习解决方案。

针对数据集背景复杂（泥土、石块干扰大）且各类样本稀缺（每类仅约 100 张）的难点，我们采取了“由浅入深、逐步优化”的实验策略，通过不断的迭代，将模型性能从最初的 Baseline 逐步提升至最终的 0.93。

2.1 基线模型与初步优化

我们为了确立基准性能，首先构建了一个基于 ResNet18 的 Baseline 模型，使用 ImageNet 预训练权重进行迁移学习。仅对图像进行简单的缩放（Resize）和随机裁剪（RandomCrop），使用交叉熵损失函数进行训练。

由于背景像素占比过大，模型如果过多关注了土壤的颜色和纹理，而非植物本身的话可能导致问题。如前文特征工程时做的那样，我们考虑背景去除这一预处理，利用植物叶片在 HSV（色相、饱和度、明度）色彩空间中独特的绿色范围，构建掩膜（Mask）分割前景。这一步极大地提升了图像的“信噪比”，强制让我们的模型专注于叶片的形状和纹理。此时验证集准确率提升至 80%–85%。但由于特征过于简单，模型出现了严重的过拟合现象。

为了解决过拟合问题，以及提升性能，我采用了多种策略。我将模型由 ResNet18 切换为 EfficientNet-B3。该模型在参数量和特征提取能力之间取得了极佳平衡，特别适合细粒度纹理识别。加入 Mixup 数据增强，通过将两张图片按比例（Beta 分布， $\alpha=0.2$ ）线性混合，强制模型学习“部分特征”而非死记硬背。使用余弦退火帮助模型跳出局部最优解。

在这一阶段，模型在 Kaggle 上的得分为 0.88。

2.2 架构升级与高阶正则化

考虑到数据量较小，我们将 ResNet18 换为 DenseNet121。DenseNet 通过密集连接的机制实现了特征的极致复用，在参数量显著少于 ResNet 的同时，能够提取更丰富的细粒度特征，更加适合小样本任务。

为了消除单次数据划分的偶然性，我们采用了分层 5 折交叉验证。将训练集分为 5 份，轮流使用其中 4 份训练，1 份验证，训练出 5 个独立的 DenseNet 模型。在测试阶段，我们将这 5 个模型的预测概率相加取平均。

引入测试时增强（TTA）。在最终预测时，不仅输入原图，还输入水平翻转后的图片，取两者的平均预测结果。

实验结论

经过上述改进，通过结合 HSV 去背景、DenseNet121 轻量化架构、Mixup 混合增强以及 5-Fold 模型融合，我们在 Kaggle 测试集上最终取得了 0.93 的准确率。这一过程证明了在小样本任务中，数据清洗（去背景）与集成学习策略的重要性远大于单纯的模型深度堆叠。

三、 最终使用的模型

3.1 模型架构

使用 torchvision 提供的 DenseNet121 模型，加载 ImageNet 预训练权重。

DenseNet 的密集连接结构使得每一层都能直接访问前面所有层的特征图，缓解了梯度消失问题，并强化了特征传播。

修改全连接层（Classifier），增加了 Dropout ($p=0.3$) 层以进一步防止过拟合，最终输出 5 个类别的概率。

3.2 图像特征预处理：HSV 颜色分割

将 RGB 图像转换为 HSV 空间。

设定绿色的阈值范围：Lower [25, 30, 30] 至 Upper [95, 255, 255]。

进行形态学操作（闭运算）填充叶片内部孔洞，生成掩膜并移除背景。

3.3 高级数据增强

在训练过程中，以 $\text{alpha}=0.4$ 的参数进行 Mixup 操作。

输入模型的数据不再是单一图片，而是

$$x = \lambda x_i + (1 - \lambda) x_j$$

损失函数相应调整为：

$$\text{Loss} = \lambda \text{Loss}(y_i) + (1 - \lambda) \text{Loss}(y_j)$$

3.4 训练策略

优化器： AdamW，初始学习率为 $1e-4$ ，提供比 SGD 更快的收敛速度。

学习率调度： 使用 CosineAnnealingLR（余弦退火），使学习率随 Epoch 周期性下降，帮助模型收敛到更平坦的极小值区域。

集成学习（Ensemble）： 利用 5-Fold CV 训练出的 5 个模型权重，对测试集进行预测并取平均概率，消除了单模型的偏差。

3.5 预测时 TTA

在预测时，不仅预测原图，还将图片进行水平翻转（Horizontal Flip）后再次预测，取两次预测结果的平均值作为最终结果，进一步提升了预测的稳健性。

四、总结

本次实验针对植物幼苗的细粒度分类问题，从零构建了一套基于深度学习的完整解决方案。通过从基准模型到最终高分方案的迭代优化，我们在测试集上最终达到了 93% 的准确率。

实验初期，模型难以区分背景杂草与目标植物，导致准确率低。通过引入 HSV 颜色空间分割，我们成功剔除了图像中的无关背景（土壤、石块），显著提高了图像的“信噪比”。这一过程再次印证了在深度学习中“Garbage in, Garbage out”的原理——对于背景噪声极大的数据集，单纯增加模型深度不如优化数据输入来得有效。

本实验选用的 DenseNet121 展现了优异的性能。相比于 ResNet，DenseNet 的密集连接机制（Dense Connectivity）实现了特征的复用。在仅有每类约 100 张样本的情况下，DenseNet 能更有效地利用浅层特征，且参数效率更高，这在很大程度上避免了深层网络常见的梯度消失和过拟合问题，证明了其在小规模数据集上的适用性。

为了解决过拟合问题，我们采用了 Mixup 数据增强。它通过线性插值的方式构建虚拟样本，使得分类边界更加平滑，防止了模型在训练数据上“死记硬背”，迫使模型学习局部特征，以及真正有用的特征。采用余弦退火的学习率策略，模型能够在训练后期跳出局部极小值，寻找更平坦、泛化性更好的解空间。

实验中发现，单次训练的模型在不同验证集划分下表现波动较大。为了进一步提高性能，冲击高分，采用 5 折交叉验证结合测试时增强（TTA）的策略，有效地平滑了模型的预测方差，消除了单次训练的随机性误差。

不足与改进

模型局限：DenseNet 虽然有效，但在处理极度相似的杂草（如 Black-grass 与 Loose Silky-bent）时，对纹理细节的捕捉仍不足。更加新颖的 EfficientNet 系列模型在尝试中表现不佳。ViT 在这个任务中过于“大材小用”了，寻找一个合适的模型仍是问题。

分割缺陷：HSV 阈值分割法较为生硬，可尝试使用 U-Net 等 Pixel2Pixel 的模型进行更精细的背景剔除，但是算力有限，而且精细的背景剔除本身也是一个复杂困难的问题，因此并未进行过多的尝试。

增强策略：可引入 CutMix 或 Mosaic 增强，进一步强迫模型关注植物的局部关键特征。

综上所述，本实验验证了一个充满噪声的小样本数据集上，通过合理的数据预处理、模型选择以及正则化策略，可以构建出高鲁棒性的分类系统。

任务三：图像情感分类

一、任务重述

本次实验的任务是基于面部表情图像进行情感分类。本次竞赛的核心任务是构建一个高效的卷积神经网络模型，对面部表情图像进行自动分类。输入数据为 48x48 像素的单通道灰度图像，输出为 6 种基本情感类别之一：愤怒 (Angry)、恐惧 (Fear)、快乐 (Happy)、中性 (Neutral)、悲伤 (Sad)、惊讶 (Surprise)。

本次竞赛数据集包含一系列 48x48 像素的单通道灰度图像。每张图像对应一个特定的情感标签。

输入：48x48 灰度人脸图像。

输出：6 个情感类别 (Angry, Fear, Happy, Neutral, Sad, Surprise)。

数据集存在明显的类别不平衡现象。其中 Happy 类别样本最多 (7192 张)，而 Surprise (3202 张) 和 Angry (3963 张) 相对较少。

二、实验与分析过程

2.1 数据集分析

首先分析数据集，本次实验使用的数据集包含一系列 48x48 像素的单通道灰度图像。不难发现，数据集存在明显的类别不平衡现象。其中 Happy 类别样本最多 (7192 张)，而 Surprise (3202 张) 和 Angry (3963 张) 相对较少。图像分辨率较低，且存在光照不均、遮挡等干扰因素。

通过分析数据集，我们发现问题有以下难点：

1. **低分辨率**：48x48 的像素难以保留细腻的面部纹理特征。
2. **类别不平衡**：模型容易偏向于预测数量较多的类。
3. **类间相似性**：部分情感（如 Fear 和 Surprise）在视觉特征上极其相似，容易混淆。
4. **过拟合风险**：相对于深度学习模型的参数量，训练样本数量较少，极易发生过拟合。

2.2 基线构建与初步尝试

这次实验我们按照数据预处理 → 模型构建 → 训练调优的流程进行。

首先要构建一个简单的 baseline，这里我们组使用 ResNet18 微调，为了适配尺寸，将 48x48 的灰度图 Resize 到 224x224，并将单通道灰度图复制为 3 通道，以伪装成 RGB 图像输入。

为了扩充数据集的多样性，我们的训练集采用了 RandomHorizontalFlip (水平翻转)、RandomRotation (±15 度旋转)、RandomAffine (平移与缩放) 以及 ColorJitter (亮度与对比度扰动)。

而验证集和测试集仅进行 Resize 和 Normalize，不做几何变换。

针对 Happy 类别过多而 Surprise 过少的问题，我们计算了**类别权重**。将该权重传入 CrossEntropyLoss，使得模型在分类错误的少样本类别时受到更大的惩罚，让模型关注“难分类”的样本。

使用简单的 ResNet18 微调，发现训练集准确率迅速达到 90% 以上，但验证集准确率停留在 65% 左右，出现了严重的过拟合。

我们采取了以下的策略：

加深模型：将 ResNet18 换为 ResNet34。

引入 Dropout：在全连接层前加入了 $p=0.5$ 的 Dropout 层，随机丢弃一半的神经元连接，显著降低了过拟合。

学习率调度：使用了 Cosine Annealing (余弦退火) 策略。学习率在训练初期较大，加速收敛；在后期平滑下降，帮助模型收敛到更优的局部最小值。

2.3 强数据增强

尽管采取了上述措施，模型泛化能力虽有提升，但我们发现模型在训练后期依然存在对训练数据“死记硬背”的迹象，且对于面部遮挡（如手捂嘴、眼镜反光）的样本识别能力较弱。为了进一步提升模型的鲁棒性并突破性能瓶颈，我们引入了更强的数据增强策略：

引入 Random Erasing (随机擦除)：考虑到真实场景中常存在手部或物体遮挡面部的情况，我们在训练预处理中加入了随机擦除操作。通过随机选取图像中的矩形区域并填充噪声，强迫模型不再单纯依赖某个显著特征（如嘴型），而是学习利用眉毛、眼睛等局部特征进行综合判断。

引入 Mixup 数据增强：为了解决过拟合和决策边界生硬的问题，我们实施了 Mixup 策略。在训练过程中，不再直接使用单张图像，而是将两张随机抽取的图像按比例（如 λ 和 $1 - \lambda$ ）进行线性叠加，同时标签也进行对应的线性混合。这使得模型在处理“似笑非笑”或“悲喜交加”等模糊情感时，具有更平滑的预测能力，显著增强了模型的泛化性。

2.4 进一步训练

通过分析之前训练的结果，通过分析损失下降曲线，正确率曲线，我们发现采用强数据增强后，模型完全避免了过拟合问题，甚至在训练 30 个 epoch 后训练集准确率低于验证集准确率。我们判断，当前模型继续训练就能取得更好的结果，无需修改。因此我们加载了之前的最佳权重，继续微调训练了 30 个 Epoch。这一过程帮助模型在损失函数的曲面上找到了更优的收敛点。

2.5 测试时增强

在最终生成提交文件时，我们摒弃了单次预测模式，转而采用 TTA 策略。对于测试集中的每一张图像，分别将“原图”和“水平翻转后的图像”送入模型，得到两组预测概率分布 (Logits)，然后对这两组结果取平均值。这种方法利用了模型对镜像图像预测的一致性，有效修正了部分边缘样本的预测偏差。

实验结果

经过上述一系列改进 (ResNet34 架构、Mixup/Random Erasing 增强、TTA)，我们的模型最终在 Kaggle 测试集上取得了 **0.712** 的准确率，有效克服了低分辨率和类别不平衡带来的问题。

三、 最终使用的模型

最终选定的模型架构及配置如下：

1. 骨干网络：ResNet-34

相比于 VGG，ResNet 的残差连接（Residual Connections）解决了深层网络的梯度消失问题。相比于 ResNet-18，ResNet-34 拥有更深的层数，能提取更丰富的面部纹理特征；而相比于 ResNet-50/152，ResNet-34 参数量适中，不易在小规模数据集上过拟合。

加载了 ImageNet 的预训练权重，极大地加速了特征提取的学习过程。

2. 网络结构修改

全连接层（FC Head）：移除了原始的 1000 类输出层，替换为：

Dropout ($p=0.5$)：随机丢弃神经元，防过拟合。

Linear Layer：输出节点数为 6，对应 6 种情感类别。

3. 优化器与损失函数

优化器：AdamW。相比标准 Adam，AdamW 将权重衰减（Weight Decay $1e-2$ ）与梯度更新解耦，提供了更好的泛化性能。

损失函数：加权交叉熵损失（Weighted CrossEntropy）+ 标签平滑（Label Smoothing）。

4. 训练策略

模型的最终性能很大程度上归功于以下几种“训练技巧”的组合使用，它们显著增强了模型的泛化能力和鲁棒性：

Mixup 数据混合增强：

我们在训练过程中引入 Mixup 策略。该策略不直接使用原始图像和标签进行训练，而是随机抽取两张图像及其对应的 One-hot 标签，按 Beta 分布生成的比例系数进行线性加权叠加。Mixup 使得模型在训练时面对的是“混合特征”和“软标签”，这不仅扩充了训练数据分布，更重要的是平滑了类别间的决策边界。在 FER 这种表情过渡模糊（如悲伤与中性难以区分）的任务中，Mixup 有效防止了模型对特定样本的死记硬背，预防了过拟合。

标签平滑：

我们将传统的 One-hot 标签（如 $[0, 1, 0, 0, 0, 0]$ ）替换为软化标签（如 $[0.02, 0.9, 0.02\dots]$ ）。考虑到面部表情标注具有主观噪声，Label Smoothing 惩罚了模型对预测结果的“过度自信”，迫使模型学习更具包容性的特征，从而提升了在测试集上的泛化表现。

类别加权损失：

针对数据集中 Happy 类样本过多而 Surprise 类样本过少的问题，我们在 CrossEntropyLoss 中引入了类别权重。通过赋予少样本类别更高的损失惩罚权重，确保模型不会为了追求高整体准确率而忽略少数类别的识别。

四、总结

本次实验我最大的收获是要学会看数据，在 ResNet-34 上训练时，最初的 30 个 epoch 并未收敛，如果此时放弃并寻找其他方案的话，一个简单的提升 1.5 个百分点的方法便失之交臂了。

之前我读过一篇论文《DepthAnything》，这是一篇半监督学习的工作，它使用的强数据增强（CutMix）来促进学习，在本次实验中我使用了 Mixup，这是个很好的方法，显著的抑制了过拟合。

实验结论：

单纯增加网络深度并不一定能带来性能提升，数据增强和正则化才是关键。

Mixup 和 RandomErasing 的引入显著抑制了过拟合。在没有这些策略的初期实验中，模型极易出现过拟合的情况，而最终模型在训练集 73% 的情况下，测试集依然能达到 71.2%，证明了模型学到了真正的特征而非噪声。

ResNet-34 结合迁移学习证明了其在小图特征提取上的有效性，通过 Resize 放大图片虽然增加了计算量，但保留了预训练权重的空间敏感性。

通过合理的特征工程、稳健的模型架构选择以及先进的训练策略，实验成功达到了 71.2% 的高准确率，这在带有噪声的 FER 数据集上是一个具有竞争力的结果。

任务四：中心凹检测定位

一、任务重述

1.1 实验目标

本次实验的核心目标是自动预测眼底图像中黄斑中心凹（Fovea Centralis）的像素坐标(X, Y)。中心凹是视网膜上视觉最敏锐的区域，其精确位置的确定对于眼科疾病（如黄斑变性、糖尿病视网膜病变）的诊断具有重要临床意义。

1.2 数据集概况

训练集：仅包含 80 张已标注坐标的眼底图像。此外，任务还提供了 train_location 标注框，暗示了“先检测黄斑区域、再定位中心凹”的两阶段逻辑。

测试集：包含 40 张图像，用于最终的性能评估。

数据特点：样本量极小（仅 80 张），图像背景复杂（存在血管、视盘干扰），光照不均，且中心凹与周围黄斑区域对比度较低。

1.3 评价指标

采用 MSE（均方误差）。计算公式为：

$$MSE = \frac{1}{n} \sum_{i=1}^n [(X_{pred} - X_{true})^2 + (Y_{pred} - Y_{true})^2]$$

由于平方项的存在，哪怕是 10 像素的偏差也会产生 100 的误差值。因此，该指标对定位的精确度要求达到了亚像素级。

二、实验与分析过程

2.1 建立基准与初步尝试（256px + 灰度 UNet）

在实验的初期，通过对 80 张小样本数据的初步观察，我确立了“模型结构简单化、数据预处理标准化”的原则。

策略选择：放弃了参数量巨大的深层网络，选择了经典的 UNet 架构。这里我们考虑到彩色图像在不同设备下色差显著，通过 cv2.IMREAD_GRAYSCALE 将输入统一为灰度图，强制模型学习中心凹的几何特征和明暗对比。

基准结果：在 256x256 分辨率下，利用整数级的 argmax 提取坐标，获得了 14.14 的 MSE 分数。这验证了基础架构的有效性，但也暴露了离散坐标带来的瓶颈。

2.2 探索高分辨率与复杂模型（512px 的挫败）

为了追求更高的定位精度，我曾尝试打破 256px 的限制，将分辨率提升至 512x512，并试图引入 ResNet-34 预训练权重和 Attention Gate 机制。

偏差分析：实验结果出乎我们的意料，MSE 瞬间飙升至 250 以上。通过分析发现，在仅有 80 张图的极端环境下，高分辨率显著放大了背景噪声（如血管阴影和反光点），并且复杂的模型产生了严重的过拟合，导致模型在预测测试

集时彻底“迷路”。

这次挫败让我深刻认识到，在极小样本任务中，**稳定性远比复杂度重要**。一味提升分辨率而忽略了噪声干扰，会导致模型无法在巨大的像素空间内稳定收敛。

3. 回归稳定性并寻求亚像素突破（重心法的引入）

在经历了 512px 的失败后，我果断回溯并锁定了最初获得 14.14 分的 256px 灰度 UNet 基因。

坐标精细化：为了突破 14 分的瓶颈，我意识到不能再死记硬背整数像素索引。我引入了 **3x3 极窄窗口重心法**。

物理意义：将定位从离散的“点寻找”转变为连续的“质心估计”。由于 256 缩放回原图有近 10 倍的放大效应，这种亚像素级的微调能够微量补偿模型预测时的偏移，将原图上的物理误差降至最低。

4. 最终集成策略（检测+定位的闭环）

在实验的最后阶段，我将任务建议的“先检测、再定位”逻辑，通过热力图回归巧妙地整合进了端到端模型中：

隐式检测阶段：模型生成的 $\text{Sigma} = 5$ 的高斯圆斑覆盖了中心凹及其周围的黄斑区域。在训练过程中，模型必须先学会“检测”出这块特定区域，才能产生有效响应。

显式定位阶段：在检测到的响应区域内，配合 TTA（测试时增强）对原图与镜像图进行结果融合，利用双路对冲消除方位偏见，最终配合重心法实现精准定位。

通过这一时序性的探索与调优，实验成功将 MSE 压低至 13.xx 的水平。这一过程证明了在底层视觉任务中，“看清数据 \rightarrow 尊重基准 \rightarrow 物理微调”才是最有效的优化路径。

三、最终使用的模型

3.1 模型架构：标准 UNet

Encoder（编码器）：通过 4 层双卷积和最大池化，逐步下采样提取图像的高级语义信息（识别哪里是黄斑区域）。

Decoder（解码器）：通过转置卷积逐步恢复空间分辨率。

Skip Connection（跳跃连接）：将浅层的纹理细节与深层的语义信息融合，确保在定位时能够利用原始图像的边缘细节。

3.2 坐标提取：3x3 极窄窗口重心法（核心提分点）

这是将分数从 15 分压进 13 分的关键技术。

传统方法的缺陷： argmax 只能返回整数索引。在 256 分辨率下，缩放回原图时，哪怕是 1 像素的舍入误差，在原图上也会放大 8 倍以上。

重心法原理：在预测热力图的最大值点周围取 3×3 邻域。利用该邻域内 9 个像素的概率值作为权重进行加权平均。

数学公式：

$$X_{subpixel} = \frac{\sum_{i=1}^9 (x_i \cdot w_i)}{\sum_{i=1}^9 w_i}$$

优势: 该方法实现了亚像素级定位, 能预测出类似 125.48 这种精确的小数, 彻底消除了量化误差。

3.3 推理策略: 双路 TTA (Test Time Augmentation)

实施方法: 模型同时预测原图和水平镜像图。将镜像图的预测结果再次水平翻转后, 与原图结果取平均。

效果: 通过“自集成”抵消了模型由于训练样本单一而产生的方向性偏见。实验证明, TTA 稳定降低了约 1 分左右的 MSE。

3.4 损失函数与优化器

MSE Loss: 计算预测热力图与真值高斯图之间的差异。

ReduceLROnPlateau: 动态学习率调整。当连续 10 个 Epoch 验证集 Loss 不下降时, 将学习率乘以 0.5。这保证了模型在训练后期能以更小的步长进行精细搜索。

四、总结

本次实验我最大的收获是学会通过数据回溯本质。在尝试将分辨率提升至 512px 以及引入 Attention 机制时, MSE 分数曾一度飙升至 250 以上, 这让我深刻意识到在极小数据集(80 张图)下, 分辨率与模型复杂度的盲目提升反而会放大背景噪声。如果当时因为高分挫败就彻底放弃 UNet 架构, 那么后续通过“回归 256px 稳定版”并结合“亚像素微调”实现 13 分突破的机会便失之交臂了。

之前我研究过关于坐标回归精度限制的相关工作, 意识到 argmax 提取出的整数索引在经过 8 倍缩放还原后, 会产生不可消除的阶梯量化误差。为此, 本实验引入了**亚像素重心拟合法**。这与情感分类中使用 Mixup 平滑决策边界的逻辑有异曲同工之妙: 它不再死记硬背某一个像素点的索引, 而是通过 3x3 区域的能量加权, 在连续空间内寻找“质心”, 显著抑制了缩放造成的精度损失。

任务五：血管分割

一、任务重述

将 2D 眼底图像中的血管区域分割出来。训练集，包含眼底图像及其分割标签各 20 张。测试集，只含眼底图像 20 张。在提交前，需要使用提供的预测分割图像转 csv 的代码 (segmentation_to_csv.py)，生成最后的提交 csv 文件。

二、实验与分析过程

本课设使用的数据集是 DRIVE 数据集，图片统一裁剪为 565*584 像素。考虑在 Paper with Code 上找一个不错的方法复现。

阅读 segmentation_to_csv.py 代码发现，在获取 csv 文件前，代码先将原图 resize 为 512*512 像素，脚本内部包含特定的像素索引映射逻辑。这意味着，模型的输出尺寸必须与提交脚本对齐（原图尺寸或者直接在 512*512 分辨率上做掩码），否则将产生严重的坐标偏移。

在初期调研中，我首先尝试复现了 Paper with Code 上排名较高的 SOTA 方法，包括 vascx models 和 FSG-Net。这些模型在理论上拥有更强的特征提取能力。

然而，在对接最终的 CSV 生成流程时，遇到了严重的分辨率失配问题：原始 DRIVE 数据集图像尺寸为 565×584 ，而这些模通常在经过 Padding 或 Crop 后的更大的尺寸上训练，必须对这些模型的输出掩码进行 Resize 操作。然而眼底血管（尤其是末端微血管）只有 1-2 像素宽。使用双线性插值或最近邻插值进行 Resize 时，会导致微细血管断裂、模糊或消失。在血管分割这种及其精细的分割任务中，非原图（或者直接 resize 到 512*512）的推理在后处理时会造成灾难性的破坏。

同时我尝试了调整超参数，FSG-Net 的预处理模块支持 resize 的方式，我尝试在 512*512 的分辨率下进行推理，由于 FSG-Net 并不是在这个尺寸下训练的，因此模型效果不佳。

基于上述分析，我重新确立了模型选型标准：在原图尺寸训练，输出与原图尺寸完全一致的掩码，避免任何缩放操作。最终，我选中并复现了 FR-UNet，取得了最佳结果。

在复现过程中，由于本课设提供的 DRIVE 数据集文件结构与官方不同，且缺少官方测试集标签。需要对数据处理部分代码进行修改，并为测试集制作伪标签。值得一提的是，由于 FR-UNet 代码处理数据时的文件读取顺序（字符串排序）与 CSV 要求的顺序不一致，最终生成的掩码需要重排序后，在由脚本处理生成 csv 文件，确保提交结果的对应关系正确。

三、最终使用的模型

FR-UNet 是针对医学图像分割（特别是视网膜血管和冠状动脉造影）提出的一种深度学习模型，其核心是为了解决细小血管在下采样过程中丢失空间信息的问题。

FR-UNet 的主要特点如下：

全分辨率架构：针对传统 U-Net 因多次下采样导致微小结构空间信息丢失的局限性，FR-UNet 创新性地移除了池化操作，始终保持特征图的全分辨率状态，

从而最大程度保留了精细的边缘与纹理信息。

特征聚合模块: 为了兼顾全局语义理解, 模型引入了特征聚合模块 (FAM) 来融合多尺度上下文特征。

双阈值迭代后处理: FR-UNet 独创了双阈值迭代 (DTI) 后处理算法, 通过模拟区域生长策略, 有效修复了低对比度下的血管断裂问题。

四、总结

本次实验通过对 DRIVE 数据集进行视网膜血管分割, 最终生成了符合提交规范的 CSV 文件。在实验过程中, 我们对比了多种主流分割模型, 最终确认了全分辨率模型 FR-UNet 在当前问题环境下处理细微血管结构上的优势。

实验最大的收获在于认识到了数据处理流水线对模型选型的决定性影响。尽管 FSG-Net 等模型架构更新颖, 理论上性能更强, 但在必须对接特定转换脚本的约束下, 避免重采样带来的插值损失更加重要, 因而理论性能更弱, 但适配转换脚本的方法取得了更好的成绩。

最终, 通过成功复现并适配全分辨率模型 FR-UNet, 不仅解决了微细血管在缩放中丢失的问题, 也顺利生成了符合规范的高质量提交文件。这证明了在实际工程中, 最合适的模型不一定是理论性能最强的, 但一定是与系统约束最匹配的。