医学图像二分类技术报告

基于度量学习与双模型集成的方法

汪谨遵

ZY2557445

摘要

本报告详细介绍了医学图像二分类（正常/疾病）比赛的技术方案。我们采用基于度量学习（Metric Learning）的方法，使用 Triplet Loss 训练双骨干网络（ConvNeXt Large + ViT Large）提取特征，并通过 K 近邻（KNN）分类器进行最终预测。通过双模型软投票集成策略，最终在测试集上取得了 **0.93** 的准确率。实验表明，度量学习方法显著优于直接分类方法，高分辨率输入和差分学习率是成功的关键因素。

1 引言

医学图像分类是计算机辅助诊断的重要任务。本次比赛要求对医学图像进行二分类，判断图像属于正常还是疾病类别。由于医学图像数据集通常规模较小且类别不平衡，传统的基于交叉熵损失的直接分类方法往往效果不佳。

本方案的核心思想是采用度量学习方法，通过 Triplet Loss 将图像映射到特征空间，使同类样本距离更近、异类样本距离更远，然后使用 KNN 分类器进行预测。这种方法在小样本学习任务中表现优异。

2 数据集分析

2.1 数据集概况

训练集包含 1639 张医学图像，分为两个类别：

- **正常（Normal）**：993 张图像，占比 60.6%
- **疾病（Disease）**：646 张图像，占比 39.4%

测试集包含 250 张未标注图像。

2.2 数据预处理

由于医学图像包含丰富的细节信息，我们采用较高的输入分辨率：

表 1: 数据预处理配置

参数	值
输入分辨率	448×448
归一化均值	[0.485, 0.456, 0.406]
归一化标准差	[0.229, 0.224, 0.225]
验证集比例	15%

2.3 数据增强策略

训练阶段采用以下数据增强方法：

1. **随机裁剪**：先将图像缩放至 480×480，再随机裁剪为 448×448
2. **随机水平翻转**：概率 50%
3. **随机垂直翻转**：概率 50%
4. **随机旋转**：$\pm 15°$ 范围
5. **颜色抖动**：亮度和对比度各 $\pm 15\%$

3 方法论

3.1 整体架构

整体方案的数据流如图1所示：

图 1: 整体技术方案流程

3.2 骨干网络

我们选择两个互补的预训练模型作为特征提取器:

3.2.1 ConvNeXt Large

ConvNeXt 是一种现代化的卷积神经网络架构，融合了 Transformer 的设计理念。其主要特点包括:

- 使用 7×7 深度可分离卷积
- Layer Normalization 替代 Batch Normalization
- GELU 激活函数
- 输出特征维度: 1536

我们使用在 ImageNet-22K 上预训练并在 ImageNet-1K 上微调的版本:`convnext_large.fb_in22k`

3.2.2 Vision Transformer (ViT) Large

ViT 将图像分割为 16×16 的 patch，通过 Transformer 编码器处理。其主要特点包括:

- Patch 大小：16×16
- 24 层 Transformer 块
- 16 个注意力头
- 输出特征维度：1024

我们使用在 ImageNet-21K 上预训练并在 ImageNet-1K 上微调的版本：`vit_large_patch16_224.a`

3.2.3 双骨干网络的互补性

选择 ConvNeXt 和 ViT 组合的原因：

- **CNN + Transformer**：ConvNeXt 擅长捕捉局部纹理特征，ViT 擅长建模全局上下文关系
- **不同归纳偏置**：两种架构的特征表示具有互补性
- **经验验证**：实验中该组合性能优于其他组合（如 Swin + EfficientNet）

3.3 Adapter 网络

骨干网络输出的 2560 维特征通过 Adapter 网络映射到 128 维嵌入空间：

$$\text{Adapter} : \mathbb{R}^{2560} \to \mathbb{R}^{512} \to \mathbb{R}^{128} \tag{1}$$

具体结构：

```
Adapter = Sequential(
    Linear(2560, 512),
    BatchNorm1d(512),
    ReLU(),
    Dropout(0.5),
    Linear(512, 128)
)
```

输出经过 L2 归一化，使所有嵌入向量位于单位超球面上。

3.4 Triplet Loss

Triplet Loss 是度量学习中常用的损失函数，其目标是使锚点（anchor）与正样本（positive）的距离小于与负样本（negative）的距离：

$$\mathcal{L}_{\text{triplet}} = \max(0, d(a, p) - d(a, n) + m) \tag{2}$$

其中：

- a：锚点样本
- p：与锚点同类的正样本
- n：与锚点不同类的负样本
- $d(\cdot,\cdot)$：欧氏距离
- m：间隔参数（margin），本方案设为 0.5

我们采用 Batch Hard 策略选择三元组：

- 正样本：同类中距离最远的样本（最难正样本）
- 负样本：异类中距离最近的样本（最难负样本）

3.5 KNN 分类器

训练完成后，使用 K 近邻分类器进行预测：

$$\hat{y} = \arg\max_{c} \sum_{i \in N_k(x)} w_i \cdot \mathbb{I}(y_i = c) \tag{3}$$

其中：

- $N_k(x)$：样本 x 的 k 个最近邻
- $w_i = 1/d(x, x_i)$：距离加权
- 距离度量：欧氏距离

最佳 K 值通过验证集搜索确定，候选值为 $\{5, 9, 15, 20, 25, 30, 40, 50\}$。

3.6 测试时增强（TTA）

测试阶段采用 4 视角 TTA 增强预测稳定性：

1. 原始图像
2. 水平翻转
3. 垂直翻转
4. 旋转 90°

四个视角的特征向量取平均后再进行 L2 归一化：

$$f_{\text{TTA}} = \text{Normalize}\left(\frac{1}{4}\sum_{i=1}^{4} f_i\right) \tag{4}$$

3.7 模型集成策略

我们训练了两个模型并采用软投票集成：

表 2: 集成模型配置

模型	验证集准确率	集成权重
Model 1 (baseline)	0.90	0.90
Model 2 (metric ensemble)	0.92	0.92

软投票公式：

$$P_{\text{ensemble}} = \frac{w_1 \cdot P_1 + w_2 \cdot P_2}{w_1 + w_2} \tag{5}$$

最终预测：

$$\hat{y} = \arg\max P_{\text{ensemble}} \tag{6}$$

4 训练细节

4.1 差分学习率

差分学习率是本方案成功的关键因素之一：

表 3: 学习率配置

参数组	学习率	说明
Adapter 层	1×10^{-3}	快速学习任务特定映射
骨干网络（解冻部分）	5×10^{-6}	保持预训练知识，微调适应

4.2 骨干网络微调策略

为平衡迁移学习效果与计算效率，我们仅解冻骨干网络的最后几层：

- **ConvNeXt**：解冻 Stage 3 和 Norm 层
- **ViT**：解冻最后 2 个 Transformer 块和 Norm 层

4.3 训练超参数

表 4: 训练超参数

参数	值
批次大小	16-24
训练轮数	20-25
优化器	AdamW
权重衰减	1×10^{-3}
学习率调度	Cosine Annealing
Triplet Loss margin	0.5
Dropout 率	0.5

5 实验结果

5.1 消融实验

表 5: 不同方法对比

方法	测试集准确率
CrossEntropy 直接分类	0.57
Swin + EfficientNetV2	0.74-0.79
ConvNeXtV2 + BEiT	0.71
ConvNeXt + ViT (Triplet Loss)	0.90
ConvNeXt + ViT (Metric Ensemble)	0.92
双模型软投票集成	**0.93**

5.2 关键发现

1. **度量学习显著优于直接分类**: Triplet Loss + KNN 方法 (0.90+) 远超 CrossEntropy 直接分类 (0.57)
2. **骨干网络选择至关重要**: ConvNeXt + ViT 组合效果最佳
3. **高分辨率输入有益**: 448×448 优于 384×384
4. **简单集成策略有效**: 软投票将 0.90 和 0.92 模型提升至 0.93
5. **低准确率模型不应加入集成**: 加入准确率低于 0.85 的模型会引入噪声

5.3 模型一致性分析

两个高准确率模型（0.90 和 0.92）在 250 个测试样本上：

- 预测一致：241/250（96.4%）
- 预测不一致：9/250（3.6%）

软投票在这 9 个争议样本上做出更好的决策，从而将整体准确率提升至 0.93。

6 最终成绩

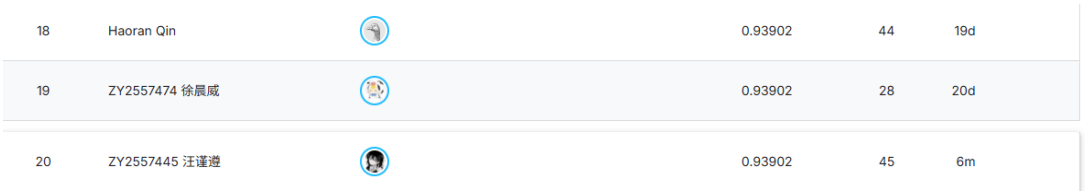

图 2: 比赛排名截图

最终在测试集上取得 **0.93** 的准确率。

7 结论

本报告介绍了一种基于度量学习的医学图像二分类方法。主要贡献包括：

1. 采用 Triplet Loss + KNN 的度量学习框架，显著优于直接分类方法
2. 使用 ConvNeXt + ViT 双骨干网络，融合 CNN 和 Transformer 的互补优势
3. 设计差分学习率策略，平衡预训练知识保持与任务适应
4. 通过双模型软投票集成，进一步提升预测准确率

未来可以探索的方向包括：更大规模的预训练模型、自监督预训练、以及更复杂的集成策略。