

# 模式识别第三次作业

## 半监督图像分类

姓名：晋一卓  
学号 22336106

### 一. 实验要求

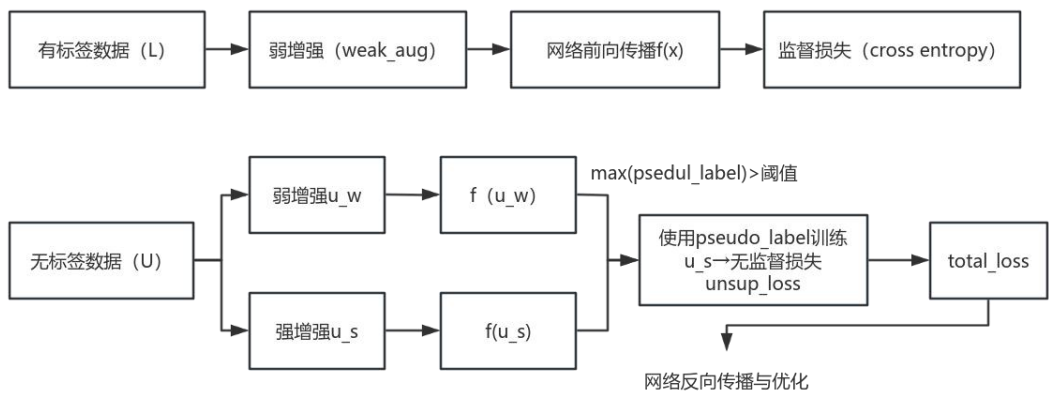
1. 基于 pytorch 手动搭建 FixMatch 半监督图像分类算法，在 CIFAR-10 数据集上进行半监督图像分类实验，报告算法在分别使用 40，250，4000 张标注数据的情况下的图像分类结果。FixMax 使用 WideResNet-28-2 作为 Backbone 网络，深度为 28，扩展因子为 2。
2. 使用 USB 提供的 FixMatch 的实现进行半监督训练与测试，与手动实现的算法训练结果进行对比。
3. 描述实现过程中的主要算法过程，尝试分析对比 FixMatch 和其他半监督算法的不同点。

### 二. 实验过程

#### 1. 主要算法原理与实现

FixMatch 是一种结合了一致性正则与为标签的方法，具体流程如下：

- 数据增强策略：使用弱增强和强增强分别对无标签图像进行两次处理；
- 伪标签生成：对弱增强图像进行预测，若预测置信度超过设定的阈值，则设为伪标签；
- 损失计算：
  - （1）监督损失：使用带标签样本计算标准交叉熵；
  - （2）无监督损失：使用伪标签与增强图像预测结果计算交叉熵
  - （3）总损失：监督损失与非监督损失的加权和
- 训练流程：



- 伪标签筛选：只有置信度高于阈值的伪标签才用于训练

```
with torch.no_grad():
    pseudo_logits = model(x_weak)
    pseudo_probs = torch.softmax(pseudo_logits, dim=-1)
    max_probs, pseudo_labels = torch.max(pseudo_probs, dim=-1)
    mask = (max_probs > config.threshold).float()
```

2. 模型参数配置

模型采用 WideResNet-28-2，主要配置如下（*config.py*）

- 网络结构：28 层，宽度因子为 2
- 优化器：SGD（momentum=0.9,weight decay=5e-4）
- batchsize：64
- 为标签置信度阈值：0.95
- 无监督损失系数：1.0
- 数据增强：weak（随即翻转+crop）+strong（RandAugment）

三. 实验结果

1. 不同标注数据情况对比分析

在 CIFAR-10 数据集上分别使用 40，250，4000 张标注数据进行训练，并记录了每种情境下的测试准确率变化情况与最佳结果（*model\_log.txt*）：

标注样本数	最佳准确率（Test Acc）	最佳轮次
40	18.44%	4
250	22.73%	5
4000	55.27%	2

在小样本（40/250）下，由于为标签质量低，模型容易过拟合少量样本，效果提升有限；而在 4000 样本下，模型能够更好学习真实语义标签，伪标签引导效果提升，表现显著改善。

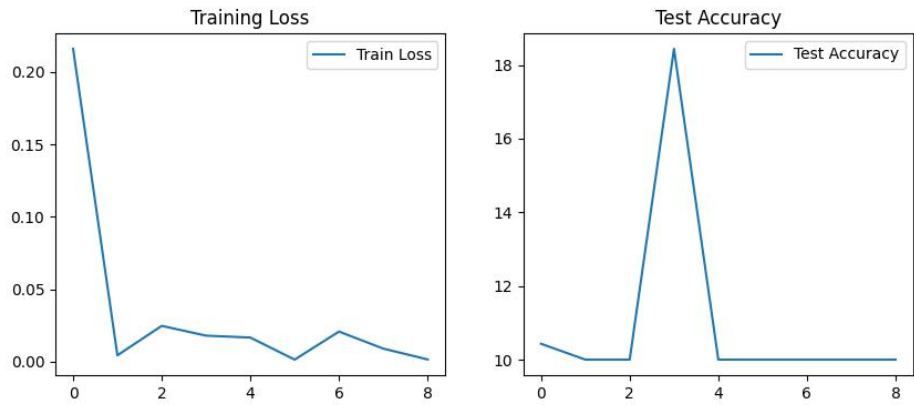


图 1 fixmatch\_curve\_40

【分析】

Epoch1：总体 loss 极大，初始准确率为 10.43%，约等于随机分类

Epoch2-3：Loss 出现剧烈波动，说明模型在绩效标注数据下开始过拟合，卫星城有效的伪标签引导

Epoch4：准确率突然提高至 18.44%，可能由于少量伪标签在某一类上构成短暂的指导，偶然拟合

Epoch5-9：回落至 10%左右，说明伪标签不稳定，导致训练陷入负反馈回路，强增强监督误导，学不到有效表示。

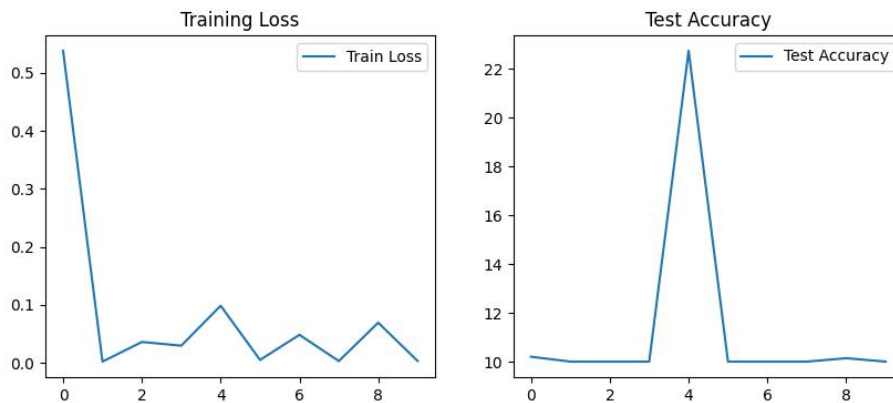


图 2 fixmatch\_curve\_250

### 【分析】

**Epoch 1:** 高 Loss，说明模型初始化刚开始拟合，伪标签完全不可用。

**Epoch 2-4:** 总体准确率维持在 10%，说明伪标签机制尚未提供帮助，可能是伪标签置信度筛选阈值过高或伪标签本身错误率过高。

**Epoch 5:** 准确率突然跃升至 22.73%，可能是因为在这一步中模型学会了识别部分特定类别，从而产生了有效的伪标签。

**Epoch 6-10:** 再次回落，可能是因为模型过拟合，或者新生成的伪标签又误导了模型，造成性能下滑。

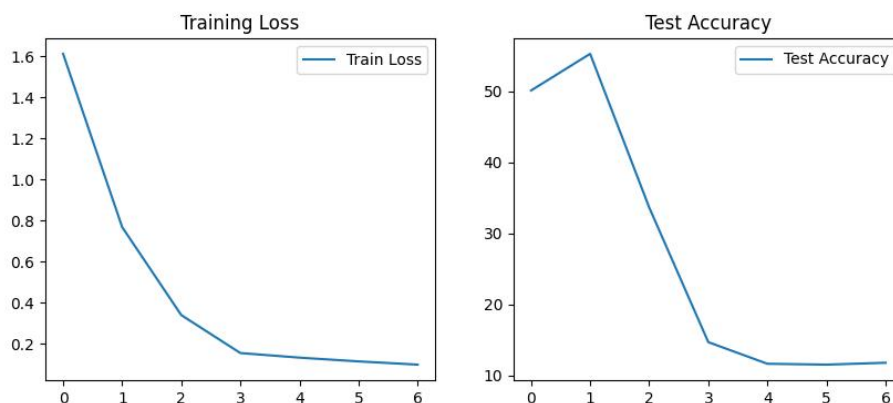


图 3 fixmatch\_curve\_4000

### 【分析】

**Epoch 1-2:** 准确率大幅提升到 50-55%，说明在 4000 张标注样本支持下，FixMatch 迅速学到了初步判别能力，并且伪标签机制开始有效工作。

**Epoch 3:** 准确率骤降至 33%，意味着新一轮伪标签可能过多、错误或偏向性强，导致训练偏移。

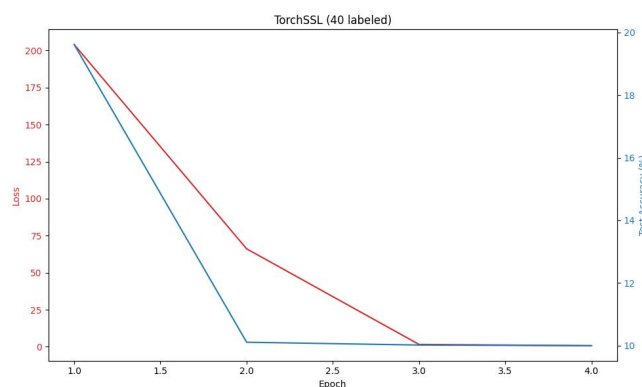
**Epoch 4-7:** 持续下降到 11%，几乎为随机分类，说明模型进入“崩溃”状态——无监督损失引导方向错误，导致模型训练逐步退化。

## 2. 与 torchSSL 提供的 FixMax 算法进行对比 (*torchSSL\_fm.py*)

标注数	我的模型最高 Acc	TorchSSL 官方模型	差异点
40	18.44%	19.61%	效果均较差，标签太少，伪标签质量差
250	22.73%	40.59%	TorchSSL 明显要好，说明其数据管道与模型调度更加成熟，对比之下我实现的模型收敛较慢，早停过早
4000	55.27%	79.75%	TorchSSL 效果非常好，我实现的模型伪标签处理、数据增强或调取不够强大

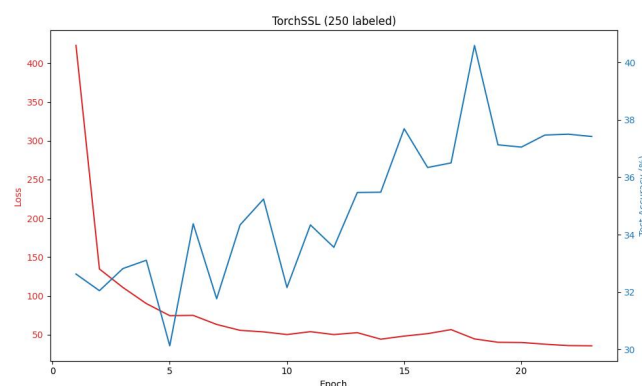
### 【训练过程分析】(*torchSSL\_log.txt*)

#### 【40 张标注样本】



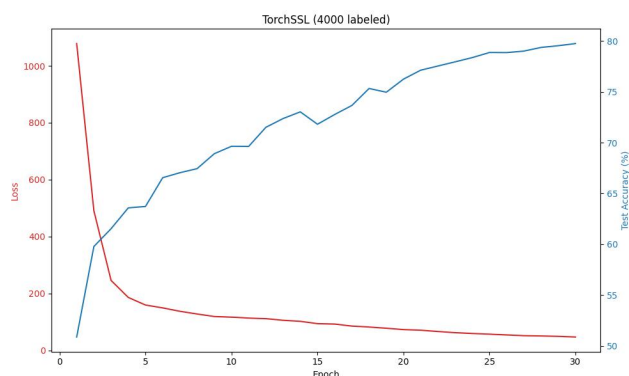
- 表现不稳定，第一轮达到 19.61%，随后直接调到接近随机水平（10%）
- 可能是样本太少导致伪标签精度过差，模型陷入错误引导

#### 【250 张标注样本】



- 训练稳定上升，最高达到 40.59%，说明伪标签起到作用，但是波动仍然很大，存在伪标签误导的风险

## 【4000 张标注样本】



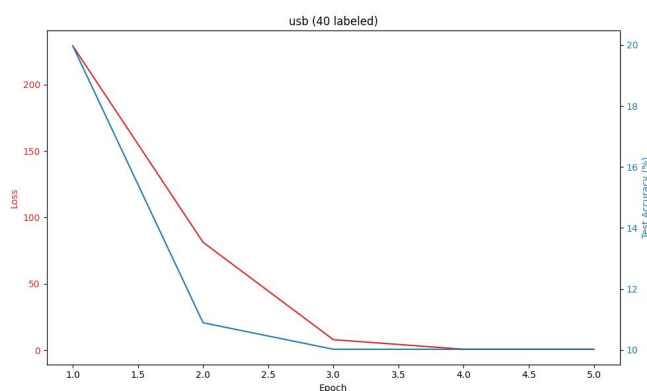
- 性能持续稳定上升至 79.75%，说明 FixMatch 效果很好，早停机制没有触发，但是通过训练结果来看，训练效果接近于收敛（epoch 29&30）

### 3. 与 USB 提供的 FixMax 算法进行对比（usb\_fm.py）

标注数	我的模型最高 Acc	USB 官方模型	差异点
40	18.44%	19.97%	官方模型初始化更稳健，但仍受限于伪标签质量
250	22.73%	63.24%	官方为标签机制+训练更长，突破 early phase
4000	55.27%	81.76%	官方训练更加充分，增强+伪标签策略更成熟

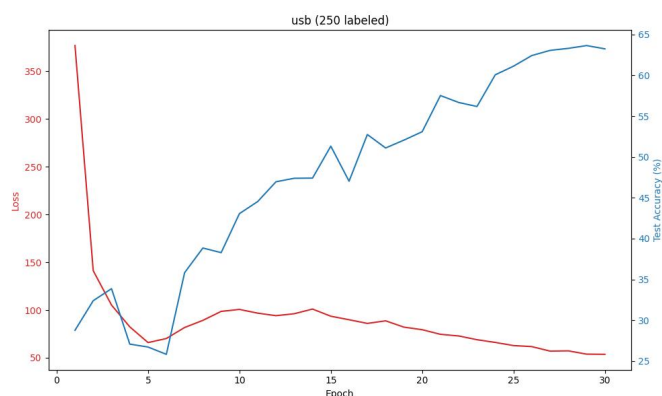
## 【训练过程分析】（usb\_log.txt）

### 【40 张标注样本】



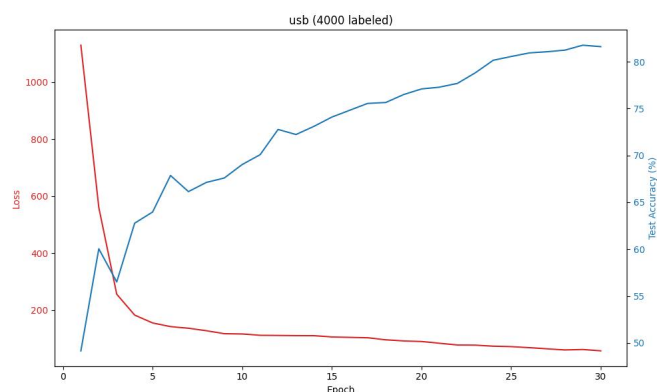
- 起始表现较好（19.97%），说明模型初始化和伪标签策略有一定优势（如增强下的稳定策略）。
- 后续迅速回落至 10%，表明伪标签信任机制可能未能维持有效的监督信号，导致模型退化为随机猜测。
- 训练损失快速下降（最后到 0.78），但泛化能力没有提升，说明模型可能过拟合伪标签或无效增强样本。

### 【250 张标注样本】



- 整体训练趋势平稳上升，说明在 250 标签样本下伪标签开始逐渐产生“协同”效果。
- 前 10 个 epoch 为预热阶段，之后提升明显：模型可能借助初步伪标签学习建立判别边界，随着自信度提高，后续伪标签更有指导性。
- 最终达 63.24%，明显优于自实现的 22.73%，说明官方策略对伪标签置信度筛选、数据增强组合更优化；

### 【4000 张标注样本】



- 准确率随着 epoch 平稳提升。每 5 个 epoch 提升 5%左右，说明训练过程未出现过拟合或标签噪声爆发。
- 最终达到 81.76%，与已有文献报告基本一致（FixMatch on CIFAR-10@4k 通常达 80~84%）。
- 相比我的模型（在 4000 标签下最佳仅 55.27%）差距明显。

## 四. 实验优化与思考

### 1. 实验优化点：

为提高训练稳定性，我在原始 FixMatch 基础上进行了以下增强（*main.py*）：

- （1）Checkpoint 保存：自动保存最佳测试准确率对应模型
- （2）Early stopping 机制，节约训练资源
- （3）训练可视化：继承训练准确率与损失函数曲线输出，方便数据的查看与对比
- （4）高置信度伪标签筛选：通过阈值过滤。提升为标签的纯度，避免负迁移

### 2. Fixmatch 算法与其他半监督算法不同点

半监督算法	关键思想	异同点
FixMatch	使用弱增强标签+强增强对齐	效率高，标签利用率强
MeanTeacher	EMA+consistently loss	模型训练复杂，需要两个模型
MixMatch	混合增强+label guessing	效果稳定，但实现比较复杂
Pseudo Label	使用自信样本作为标签	不使用强增强，效果较差

### 3. 模型优化方向

从当前实验来看，FixMatch 在标注样本数非常少的情况下准确率很低，训练也不稳定。虽然加入了 early stopping 与 checkpoint 防止过拟合，但是结果表明模型在小样本下很难充分利用伪标签，以下为针对当前模型的一些可能的改进方向：

（1）加入温度系数：目前直接对 softmax 输出取最大概率作为伪标签，没有任何平滑处理，如果伪标签的输出十分尖锐，那么一旦错了就很难纠正。因此考虑在 softmax 前除以一个温度系数  $T$  来压缩置信度分布，减少模型对错误为标签的过度自信，从而使训练更加稳定

（2）动态阈值：当前采用固定置信度阈值去筛选伪标签，在训练前期这个阈值太高大部分为标签都被筛选掉了，导致训练数据不足，因此可以考虑随着训练 epoch 迭代逐步提高阈值，前期放宽标准，多利用伪标签，后期再严格要求，提升伪标签质量。

（3）使用多个模型进行协同生成伪标签（co-teaching）：现在模型自己生成伪标签再自己训练，一旦模型前期输出错了，很容易强化这个错误，引入双模型或集成模型平均输出来判断伪标签可以提高伪标签的准确率，避免单模型的偏差影响训练。