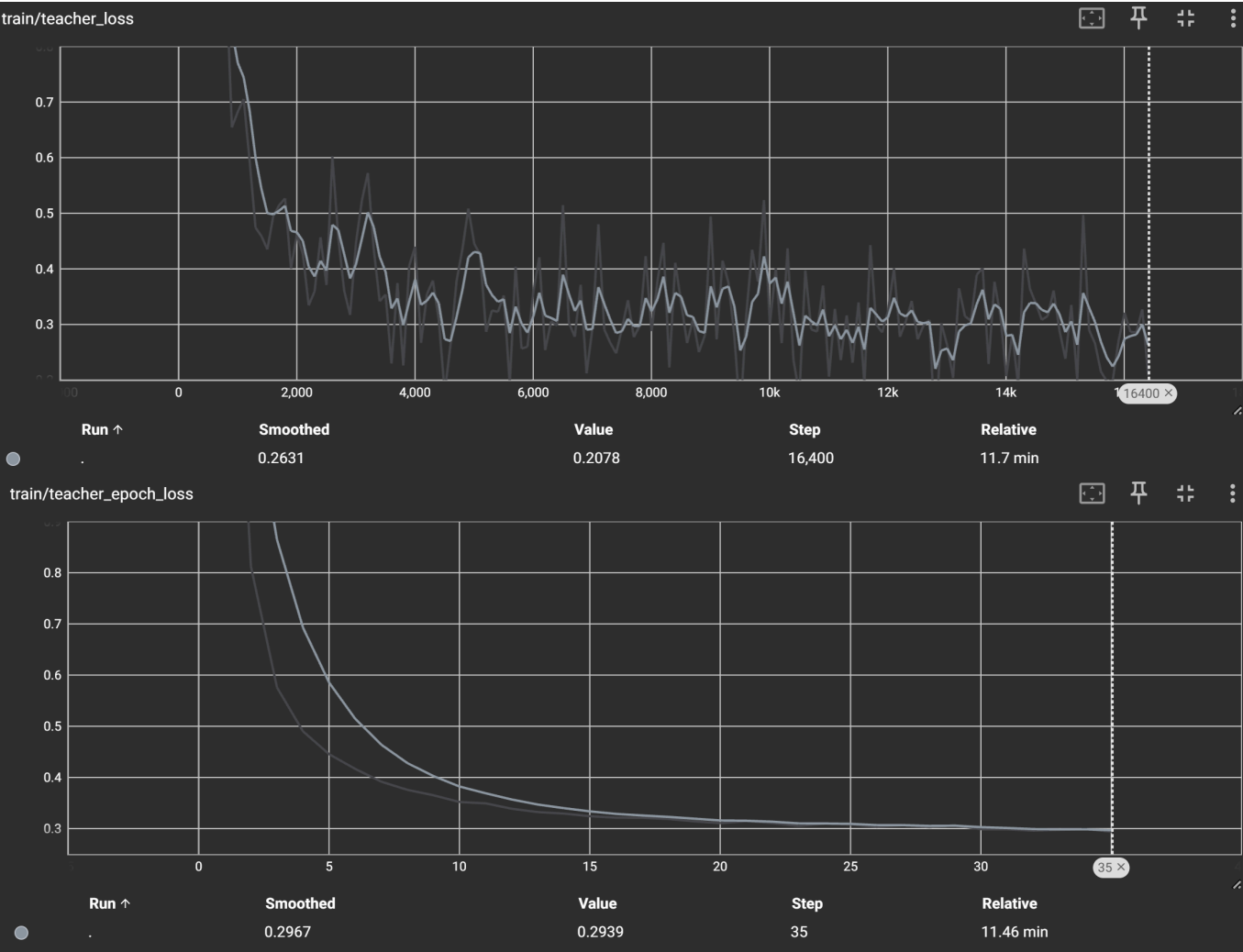


由于一些历史因素，本页笔记见PDF版本

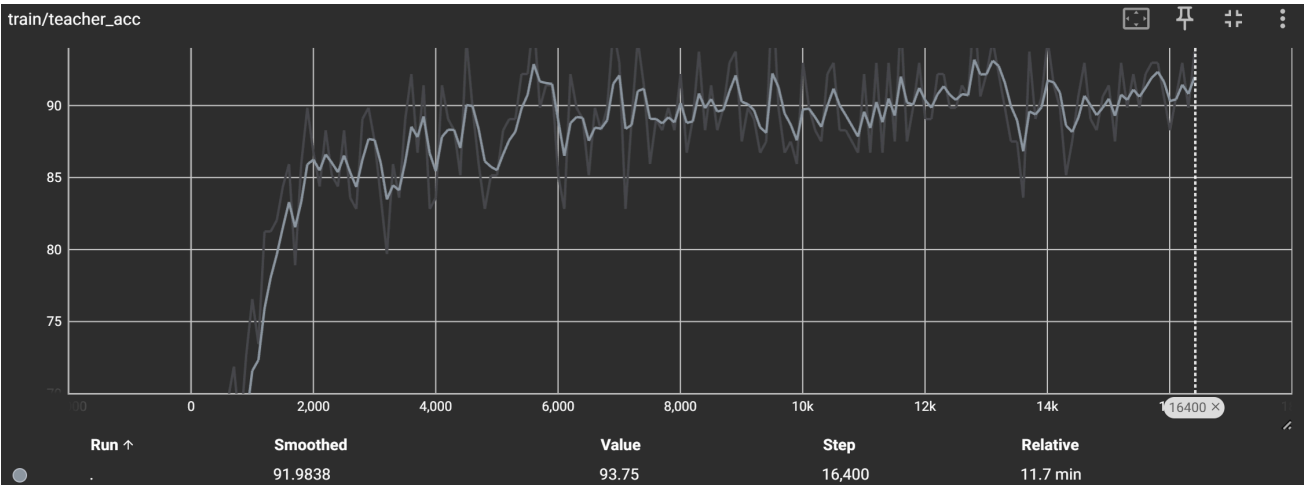
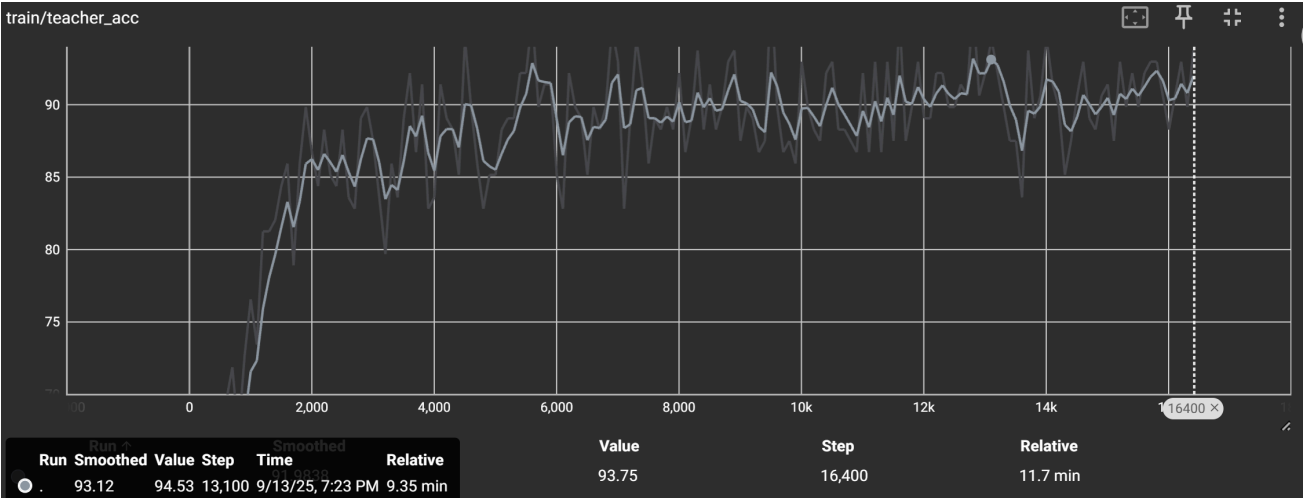
# 实验

## 试验结果1——自定义模型

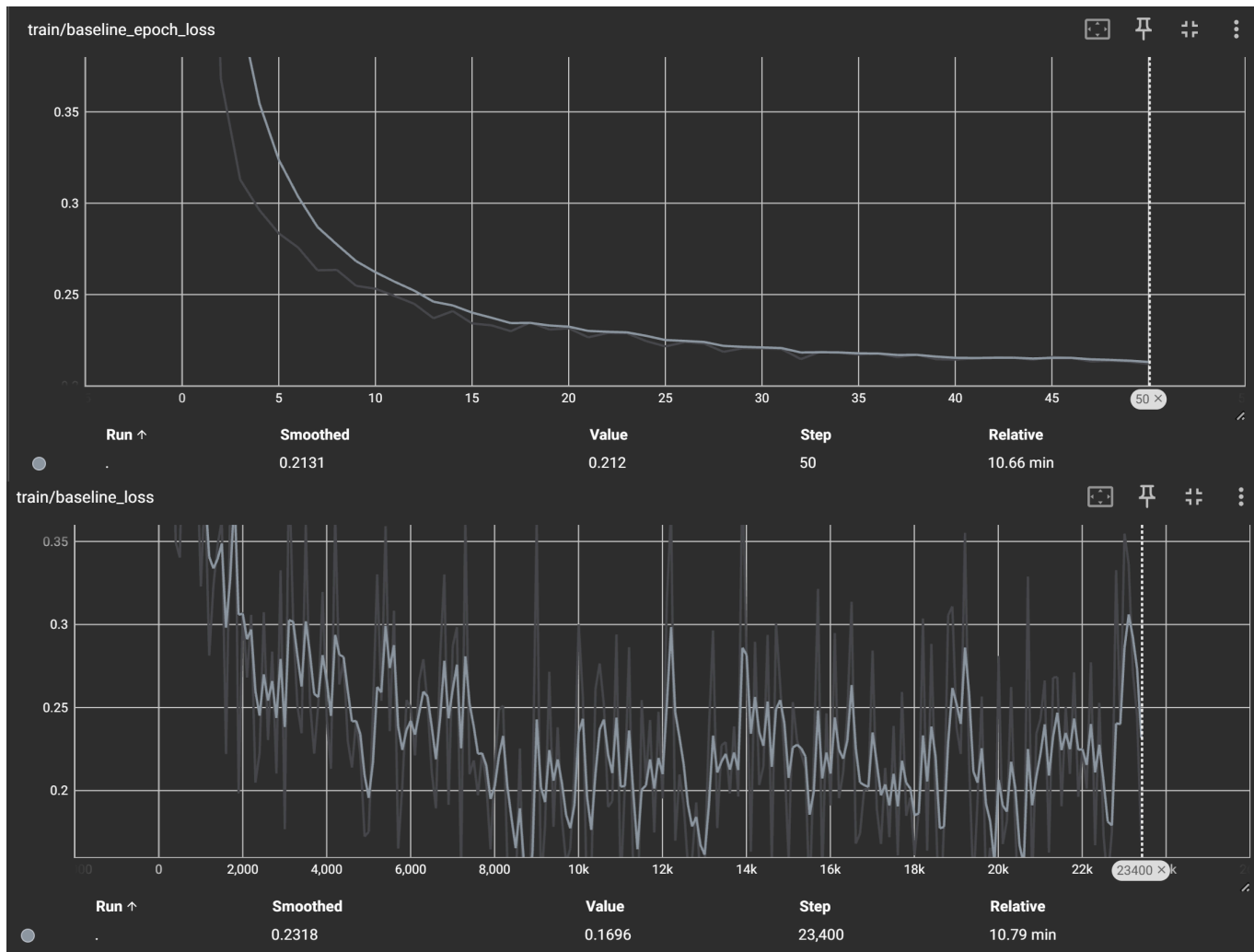
### 教师

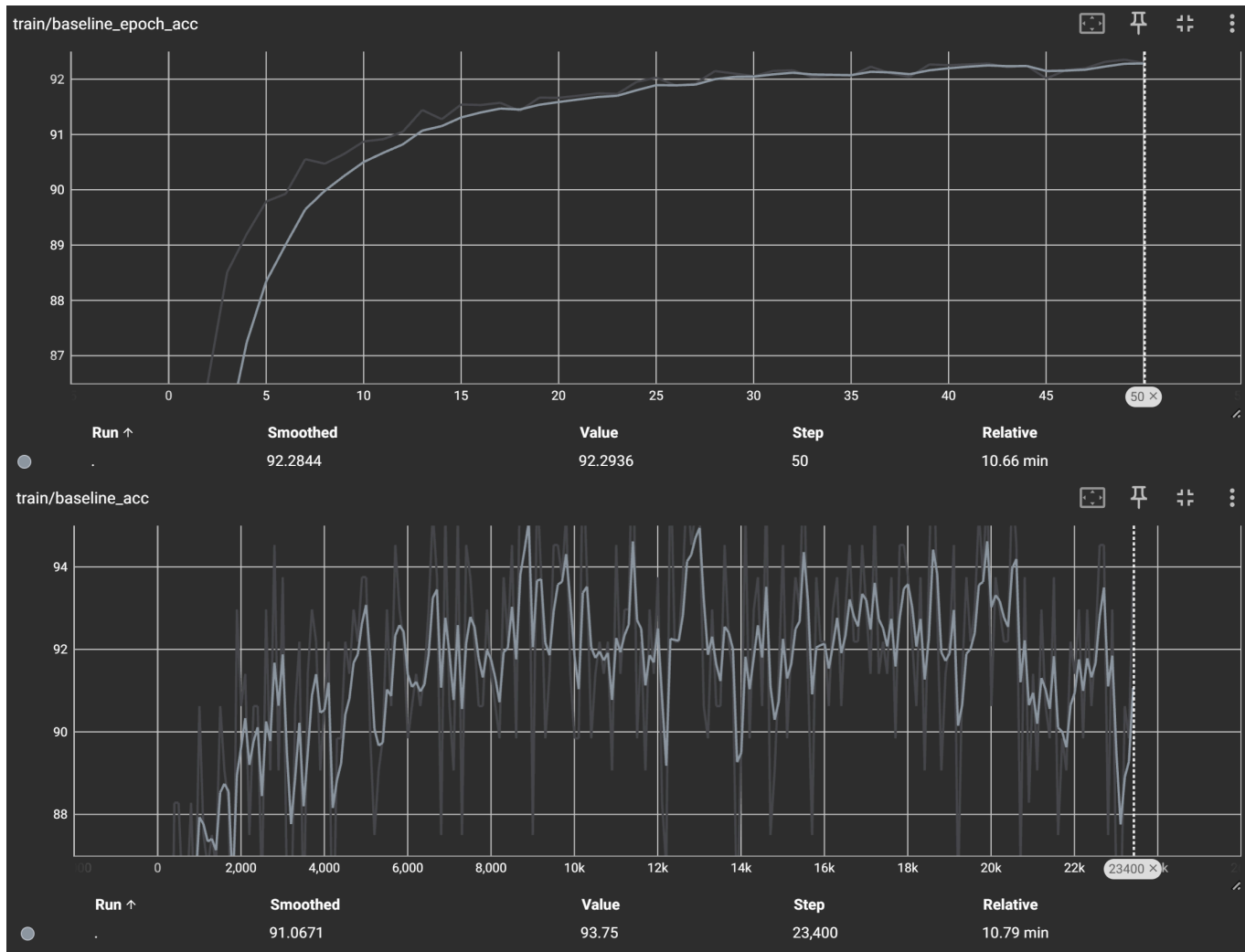


• 最高准确率达

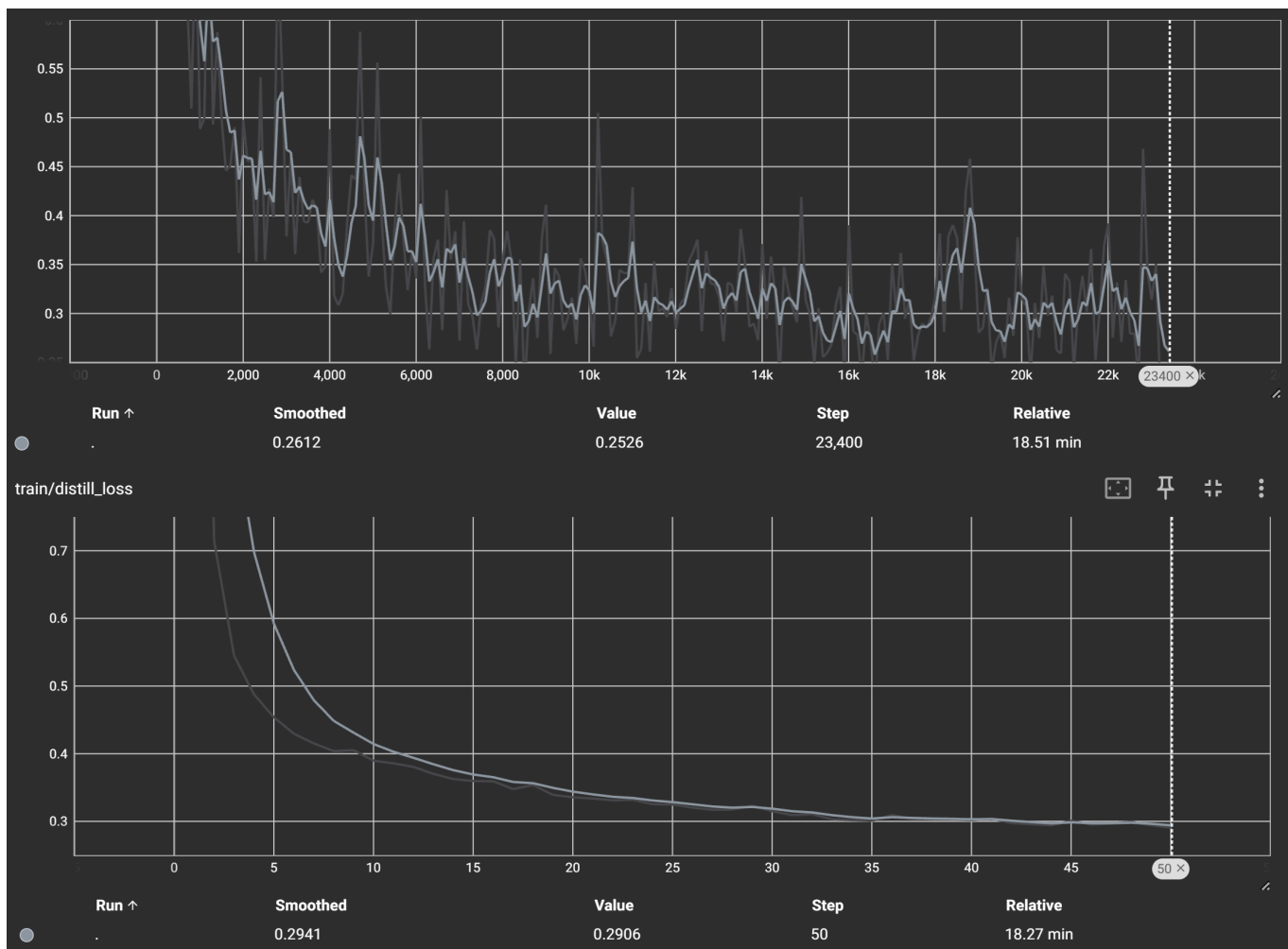


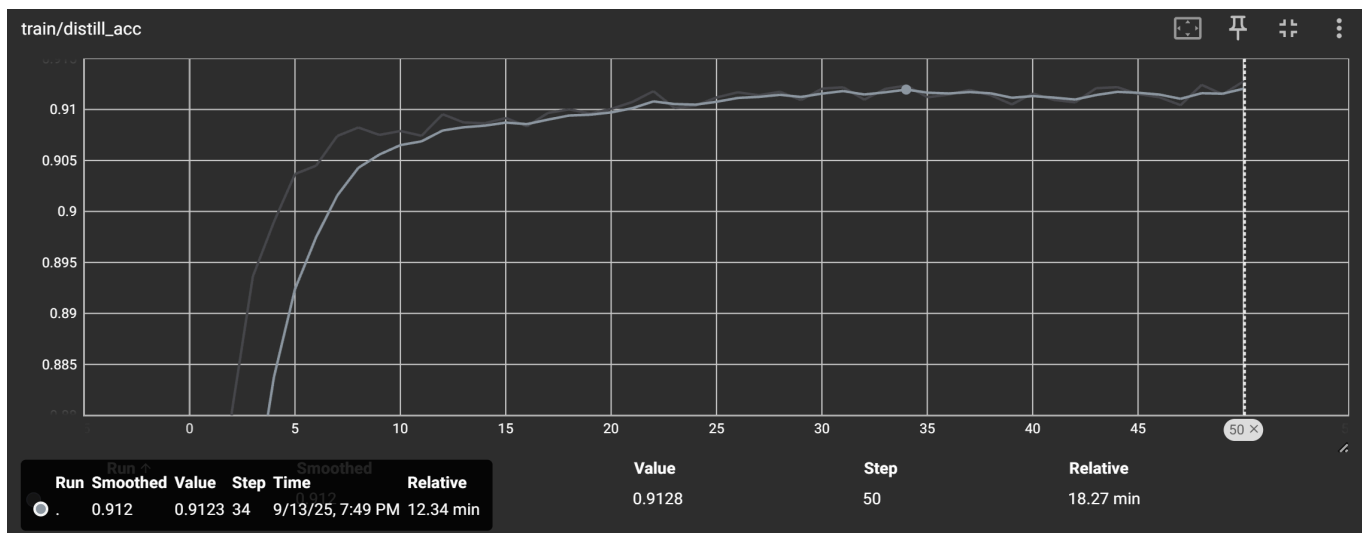
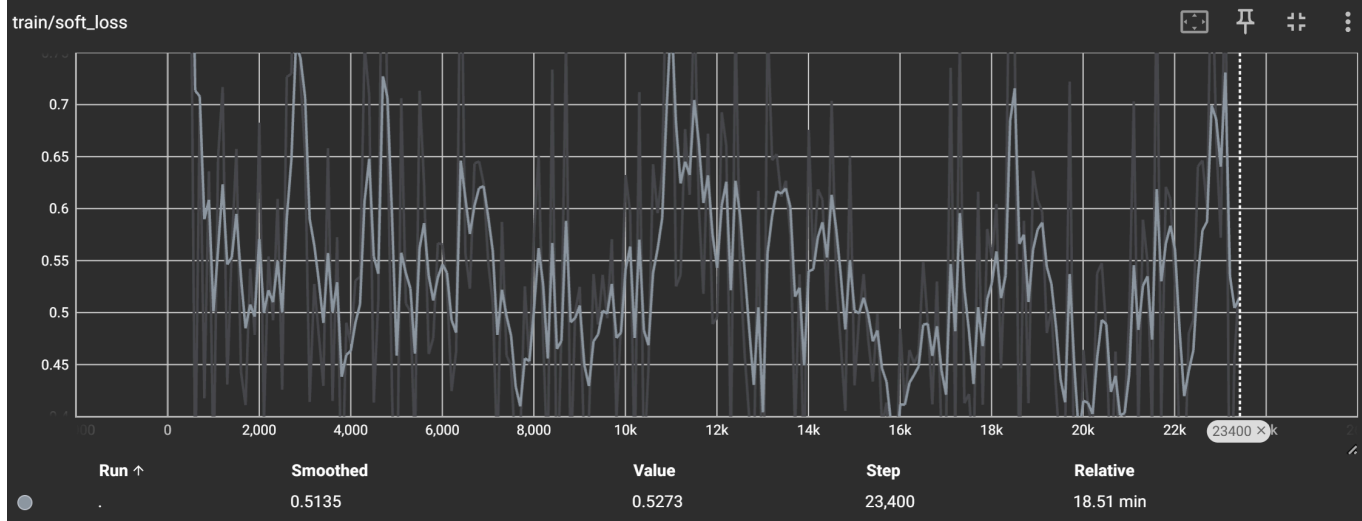
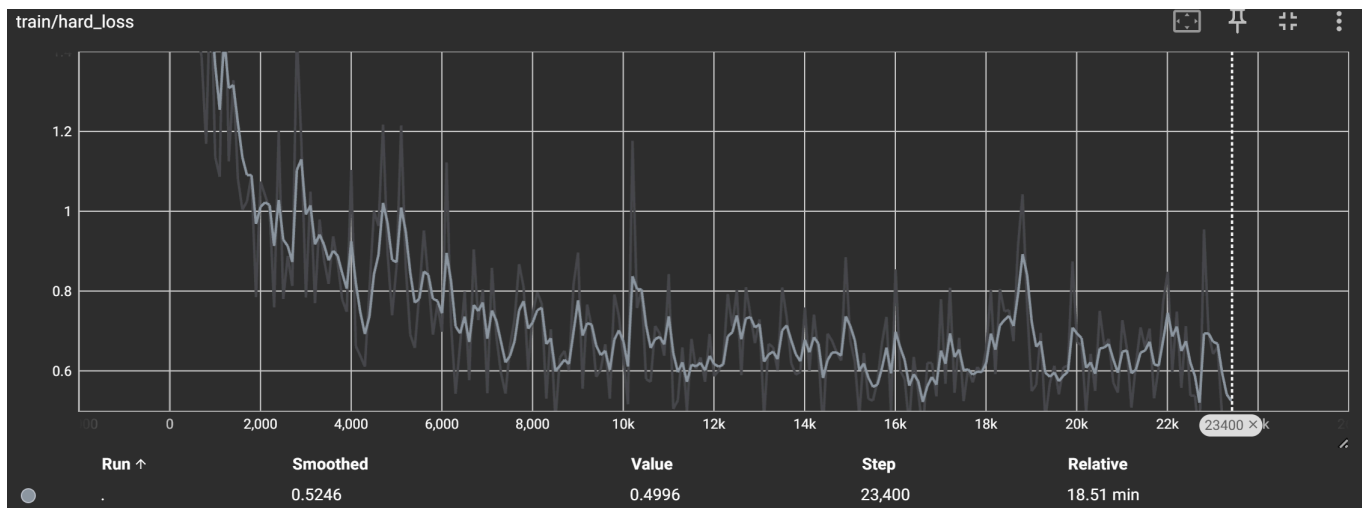
基线



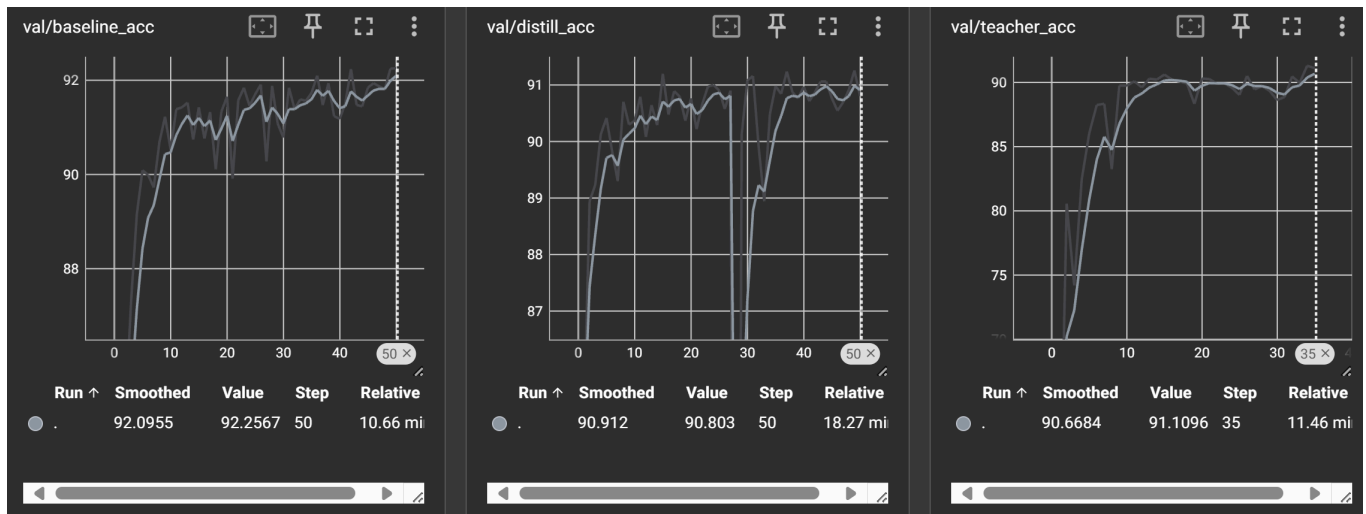


学生





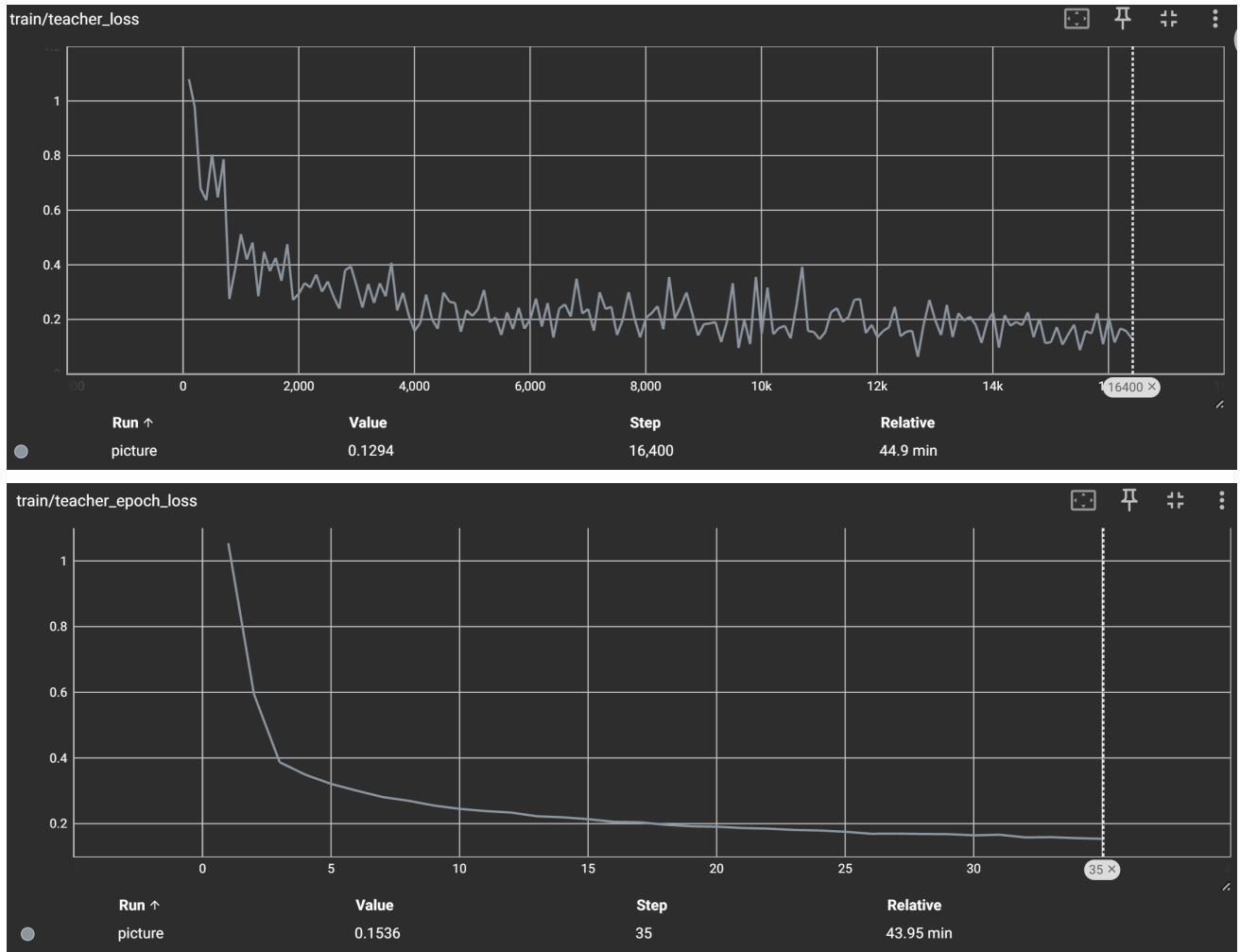
验证集



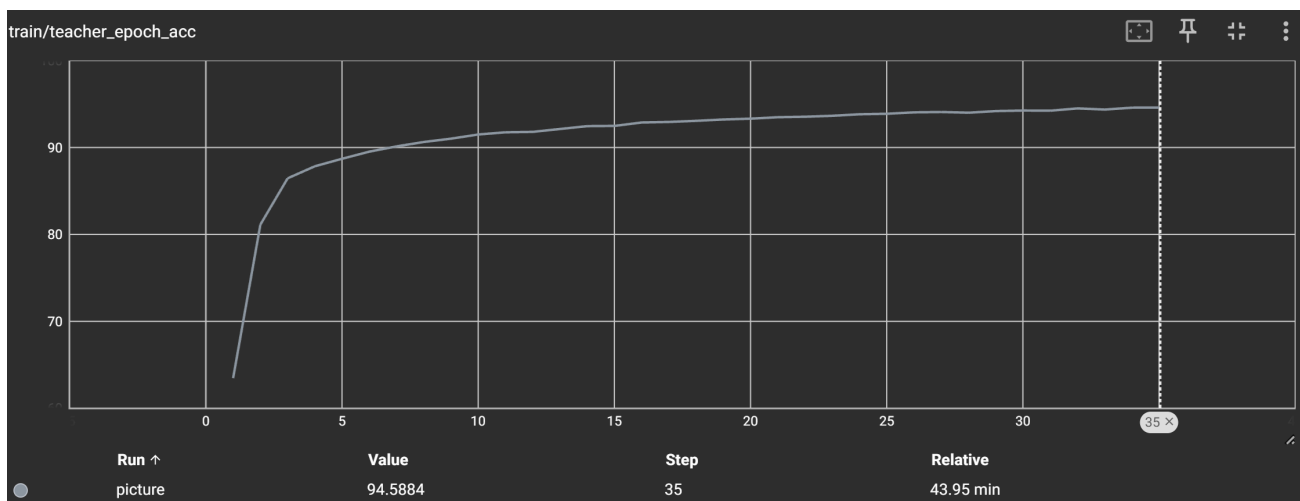
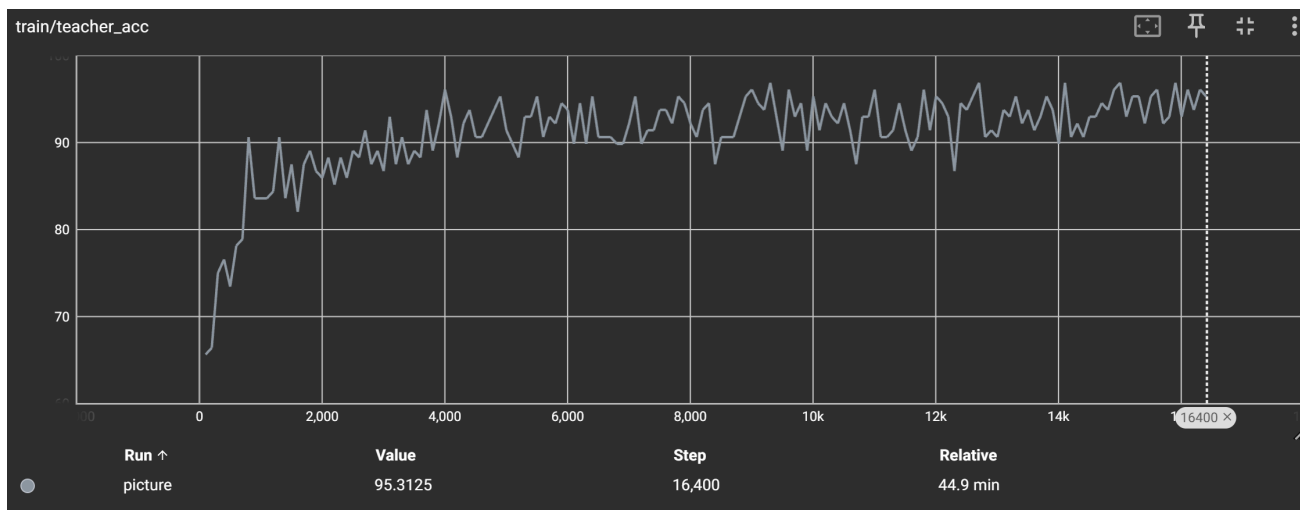
## 试验结果2——ResNet模型

### 教师

- 损失

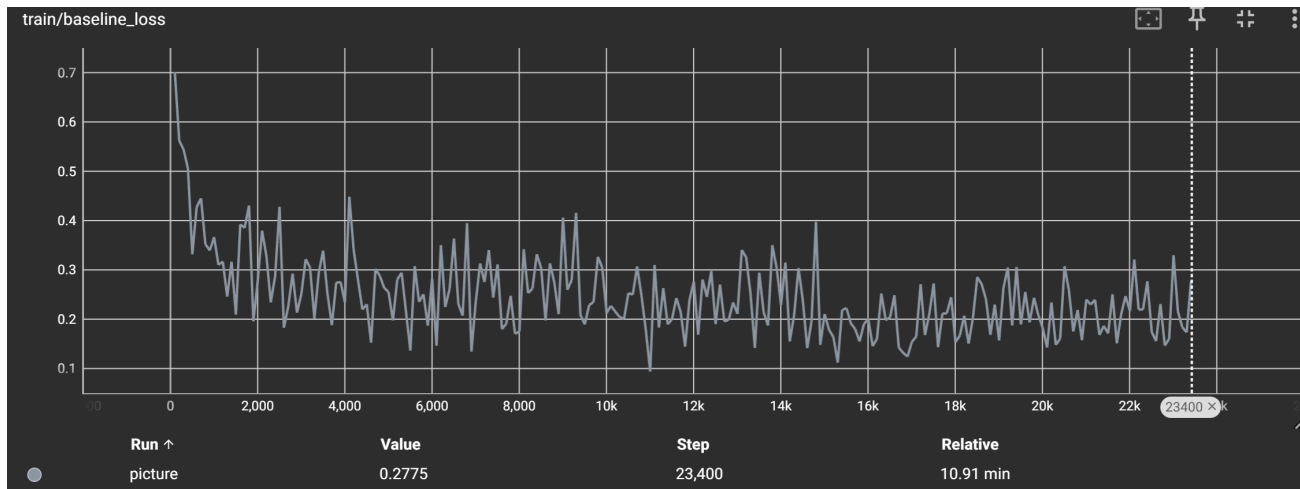


- 准确率 —— 在训练集上达到了94.5以上，非常亮眼

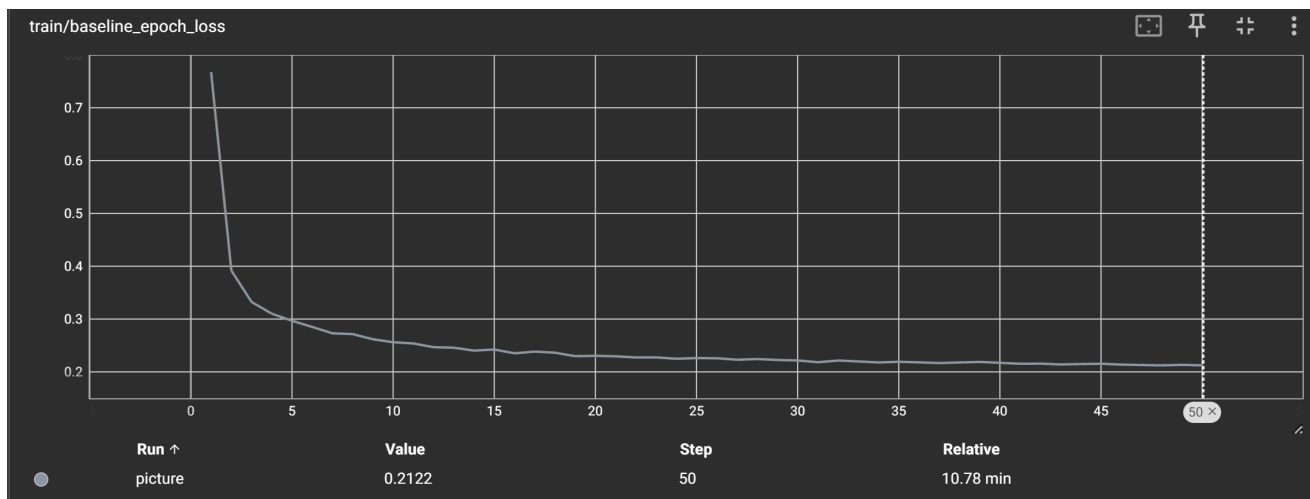


## 基线

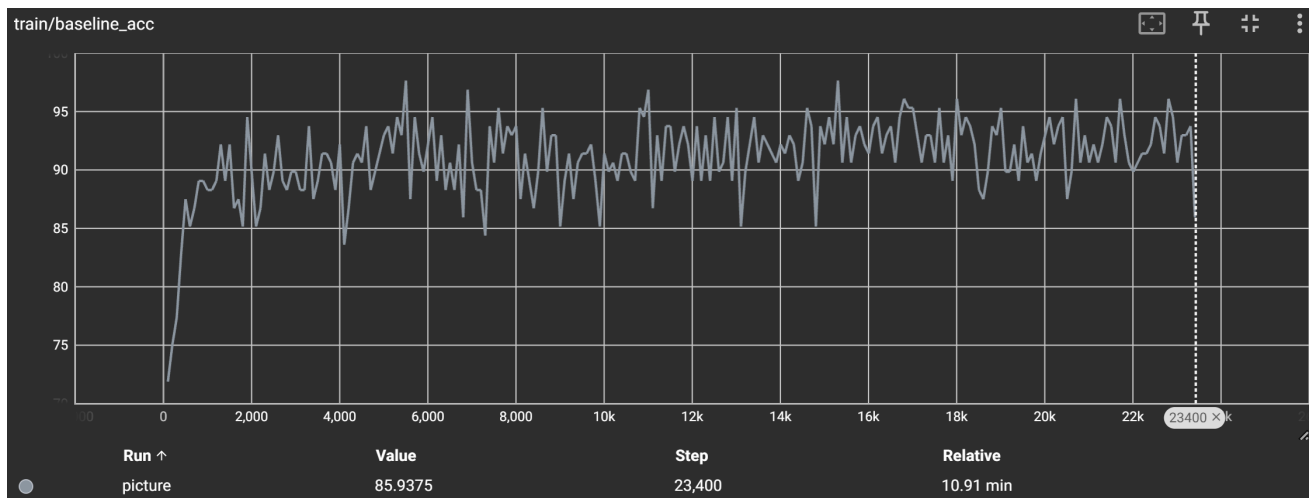
- 损失，在2以上，相较于teacher逊色许多，不过训练时不稳定，震荡程度大



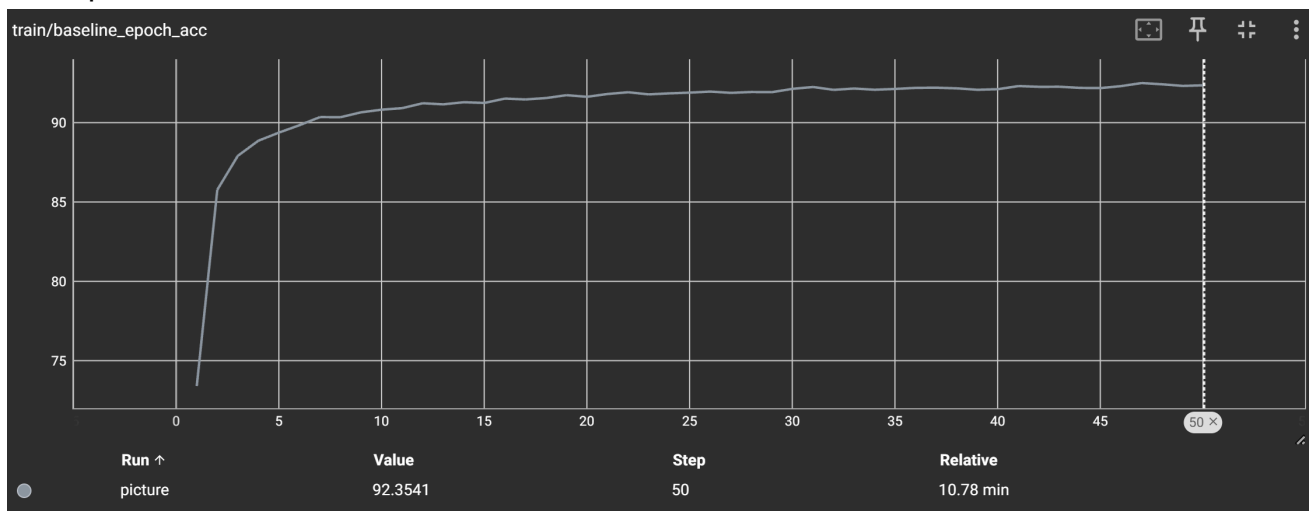




- 准确率——震荡十分明显，可能是由于lr高

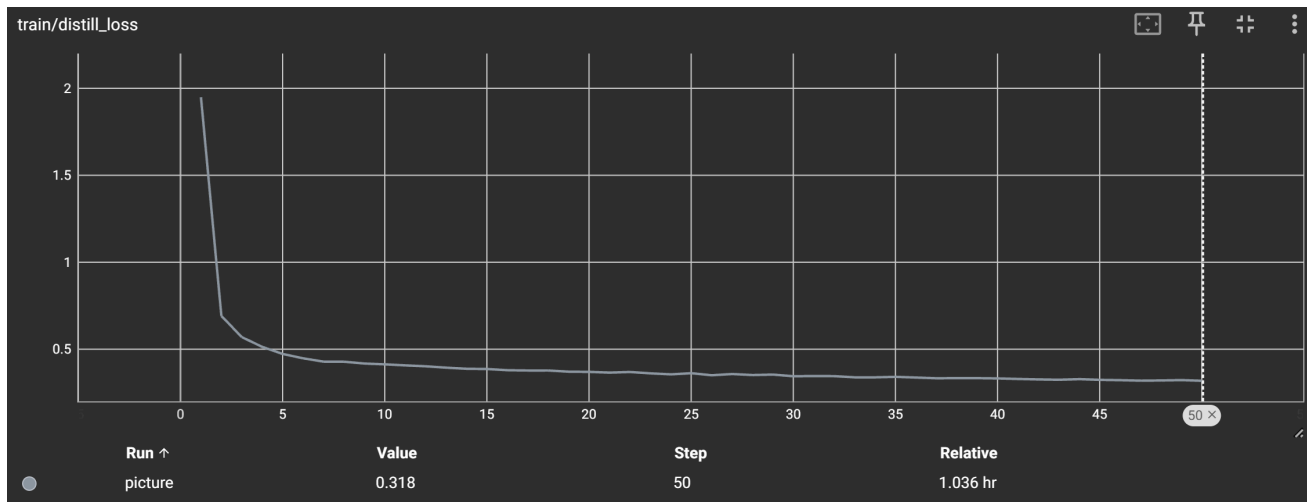
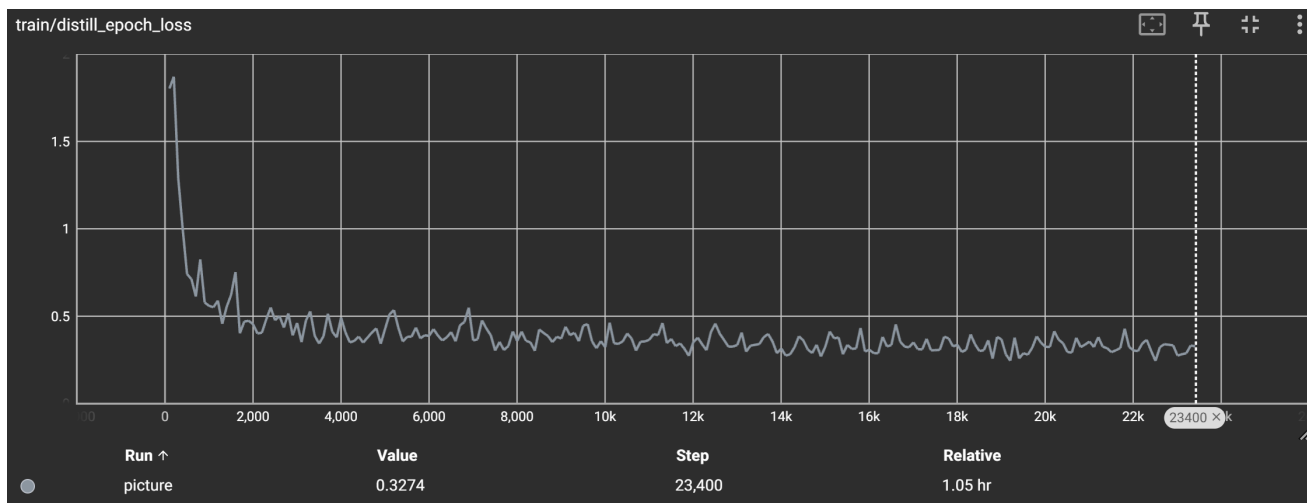


对于epoch的训练集准确率到达了92.5的水平，也是稍微逊色



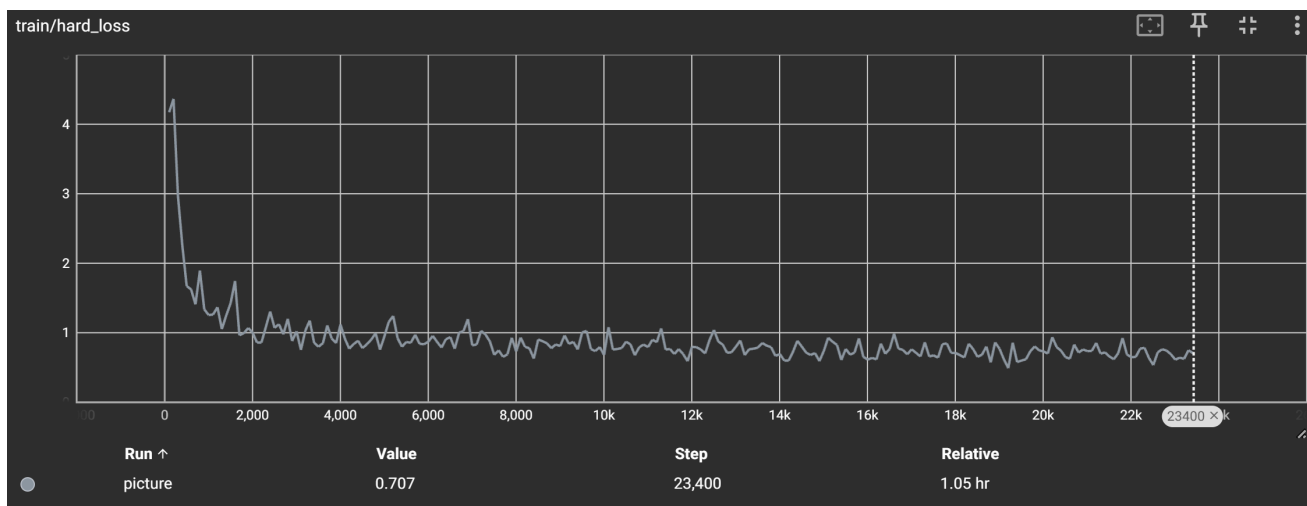
## 学生

- 损失（两张图标签颠倒了）  
震荡程度很小，但是损失值比较大

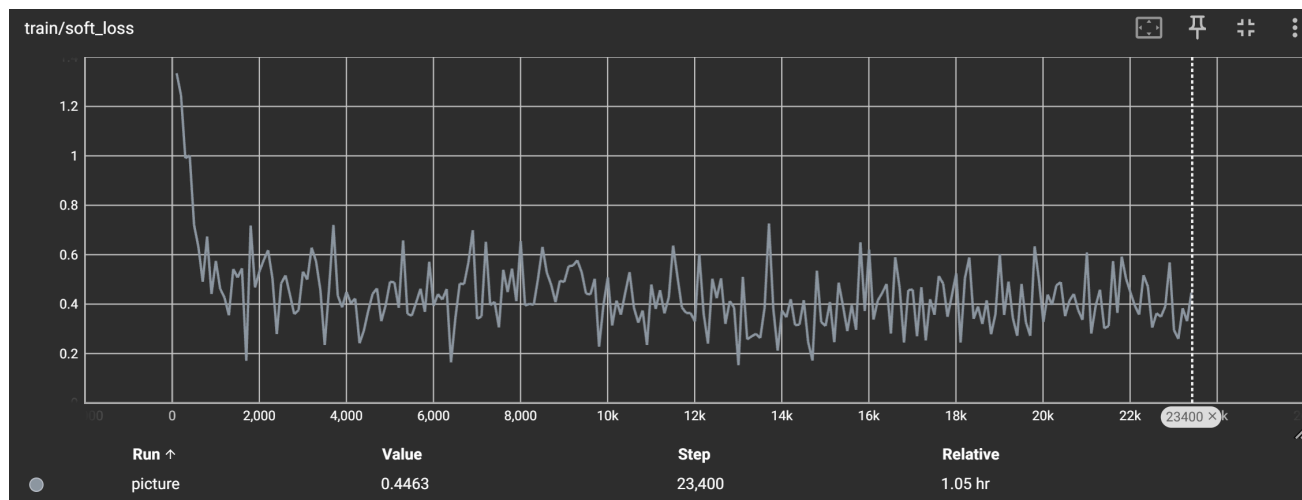


- 分类损失

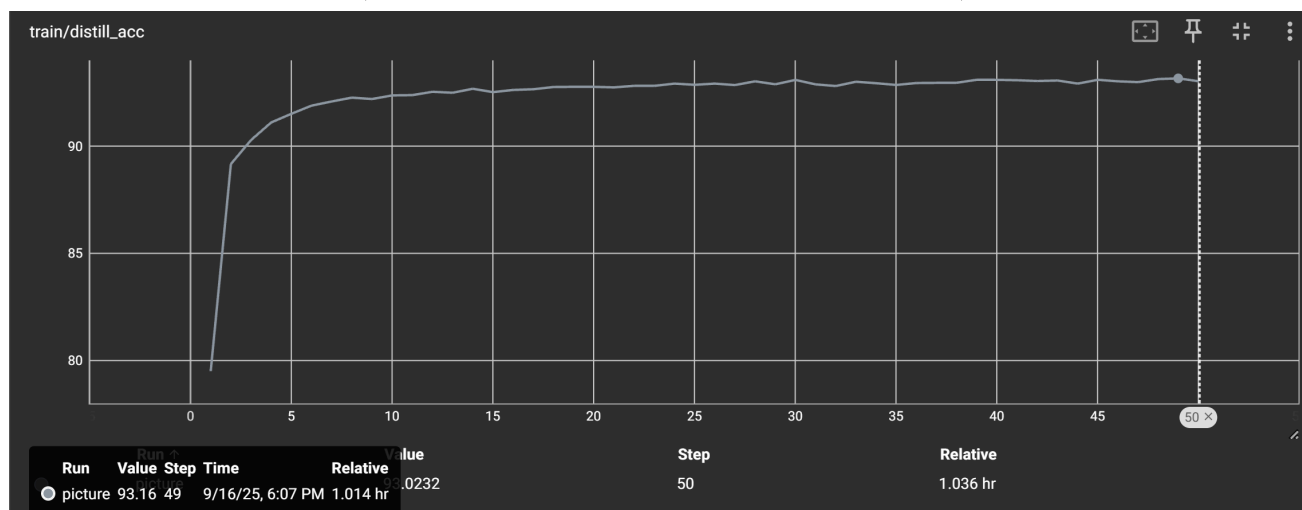
硬标签的损失十分的高，他和正确标签之间存在一些距离（一部分来源于软标签的T值）



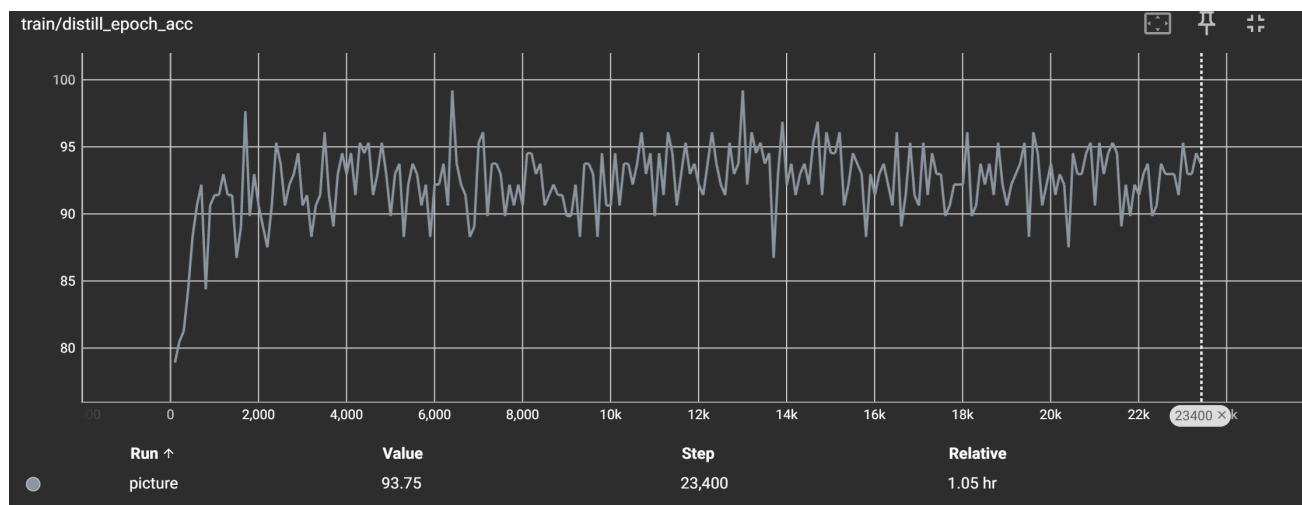
软标签的震荡十分大，也是保持在0.3之上，值也很大



- 准确率——准确率十分高，在baseline的92.5和教师模型的94.5之间，达到了93



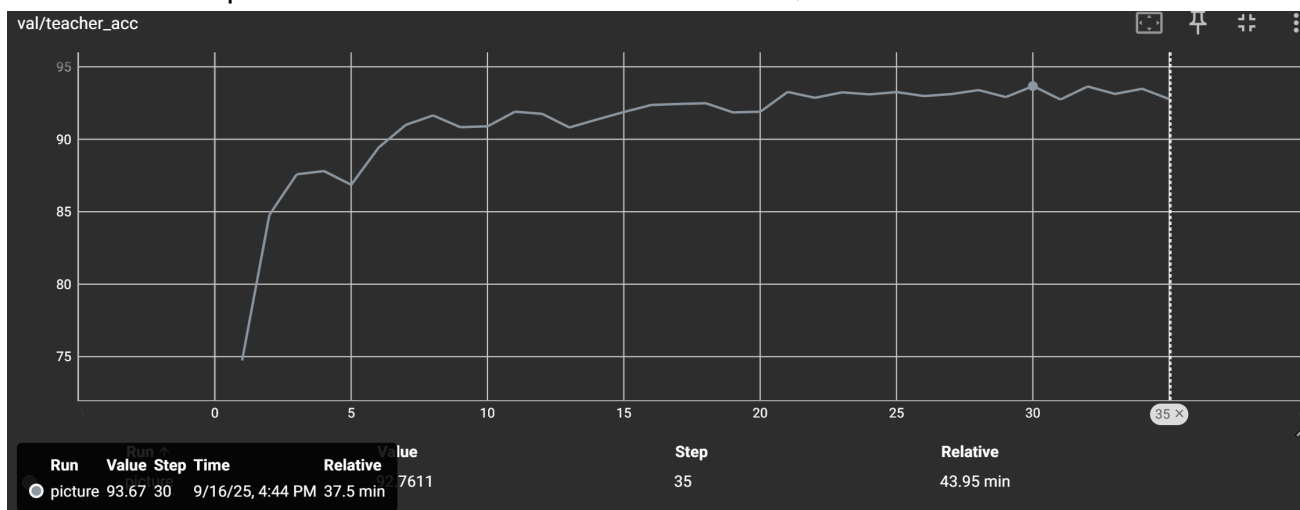
小模型震荡的比较严重



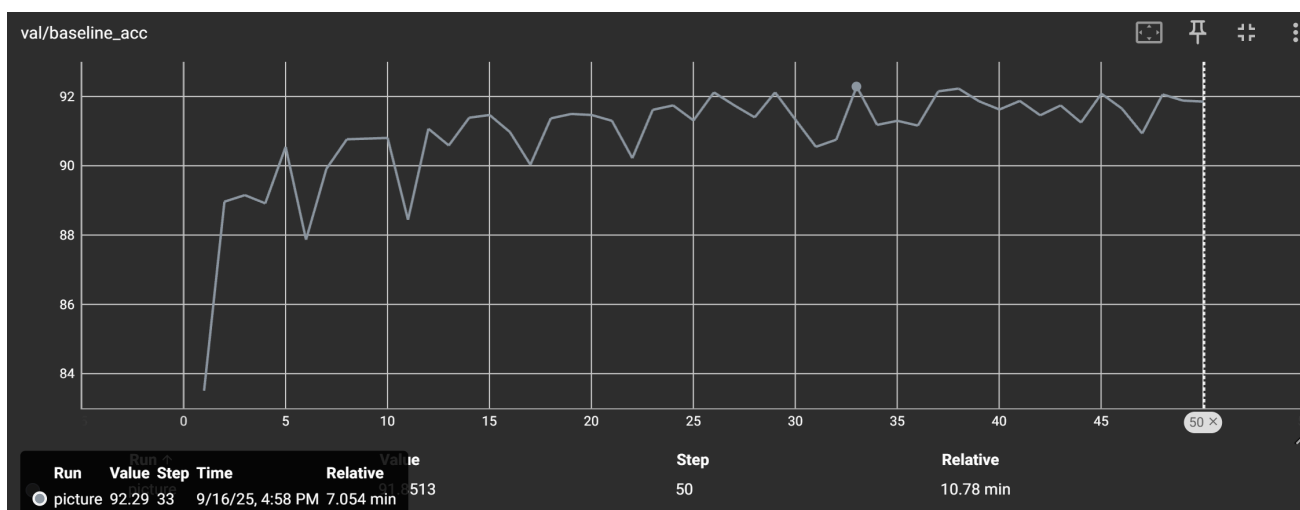
## 验证集

来看看泛化的表现

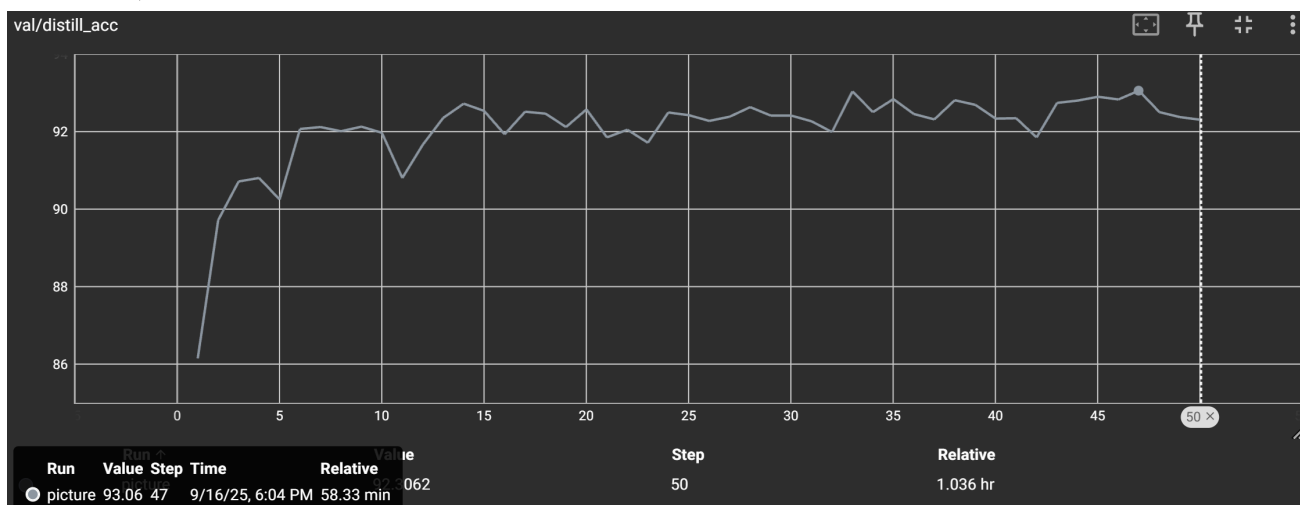
- teacher——在epoch30最佳模型达到了93.67%的准确率，在跳出了局部最小值后十分平稳



- baseline——在epoch33最佳模型达到了92.29%的准确率，训练不平稳，比教师模型差了1.5%



- student——在epoch33和epoch47都达到了93.05%左右的正确率，收敛后基本平稳，相较于基线模型，有接近0.8%的提升。



## 结果分析

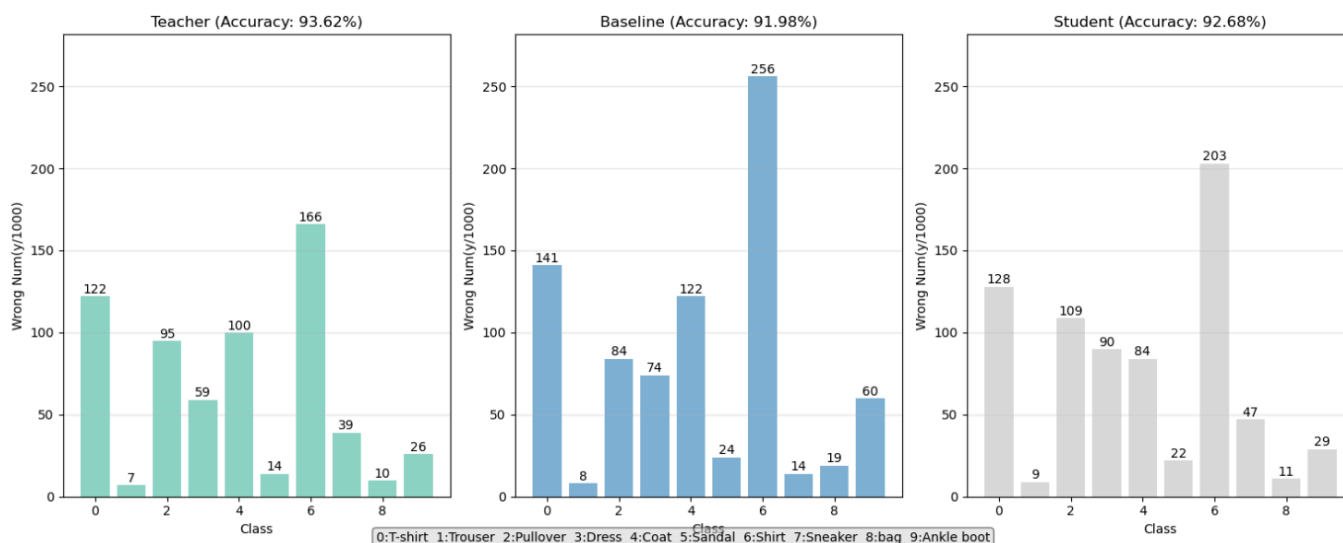
## 1. 模型的表现

模型	参数量	训练集Accuracy	测试集Accuracy
teacher (ResNet)	1351.1K ~ 1.35M	<b>95.59%</b>	<b>93.67%</b>
baseline (简单卷积)	175K ~ 1.75M	92.50%	92.29%
student (同baseline)	175K ~ 1.75M	93.09%	93.05%

在模型架构和参数量相等的条件下

- **训练速度**：蒸馏模型比空白模型收敛更快
- **训练稳定性**：在损失函数方面，蒸馏模型不稳定；然而在体现模型真实表现的验证集上，却蒸馏模型更稳定
- **准确性/泛化能力**：在拟合训练集时，蒸馏模型获得更好效果，搞了绝对0.5%，并且在泛化能力上，蒸馏模型也更加强大，接近教师模型的表现

## 2. 模型出错分析



观察teacher和baseline模型，两者出奇的一致，分布很相同。

学生模型在蒸馏之后，对于类别6明显出错有所减少。

相较于baseline，student模型出错的概率情况比较平均。