

# 实验小结

Wang Yunyi

2025.10.01

## 1 实验

实验对象共有 4 个，分别为：基础 MLP 模型（MLP），小型 CNN 模型（SmallCNN），参考 ResNet 思路并加入了残差块的 CNN 模型（ResCNN），简化版 ResNet18 模型（SmallResNet）。

设置 EarlyStopping 的 patience 为 7，最大 epoch 数为 32，所有生成随机数的种子均为 608，使用 AdamW 优化器和 CosineAnnealingLR，其他超参数均保持相同，位于 config.yaml 中。经过训练，4 个模型在训练集、验证集上的的准确率和损失如图 1。

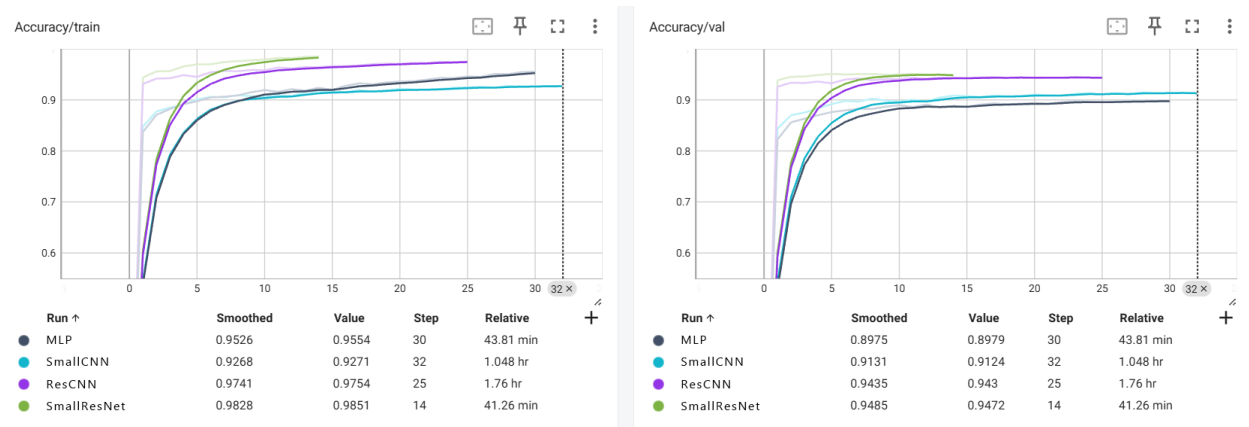


图 1: Accuracy

输出模型在验证集、测试集的准确率、loss 如表 1。

Model	Validation		Test	
	Accuracy	Loss	Accuracy	Loss
MLP	0.89763	0.045164	0.90038	0.04607
SmallCNN	0.91478	0.039486	0.91553	0.03975
ResCNN	0.94423	0.034856	0.94409	0.03511
SmallResNet	0.95100	0.031972	0.94779	0.03377

表 1: Accuracy and Loss

由实验结果可见：

- 当在 CNN 中添加了具有跳跃连接的残差块时，准确率的提升最大，达到了约 0.03；其次是在基础 MLP 模型上添加卷积、BN 等层，可以使准确率提升约 0.015。
- MLP 在达到约 10 个 epoch 时，验证集上的准确率不再提高，而在测试集上的准确率仍然缓慢提高，推测出现了过拟合。
- ResCNN 的参数量明显小于 SmallResNet，但是其在测试集上的准确率只相差 0.004 左右，这可能是由于 EMNIST 数据集仅为灰阶图片，不需要太多参数即可达到较高准确率。

四次实验的混淆矩阵如图 2。

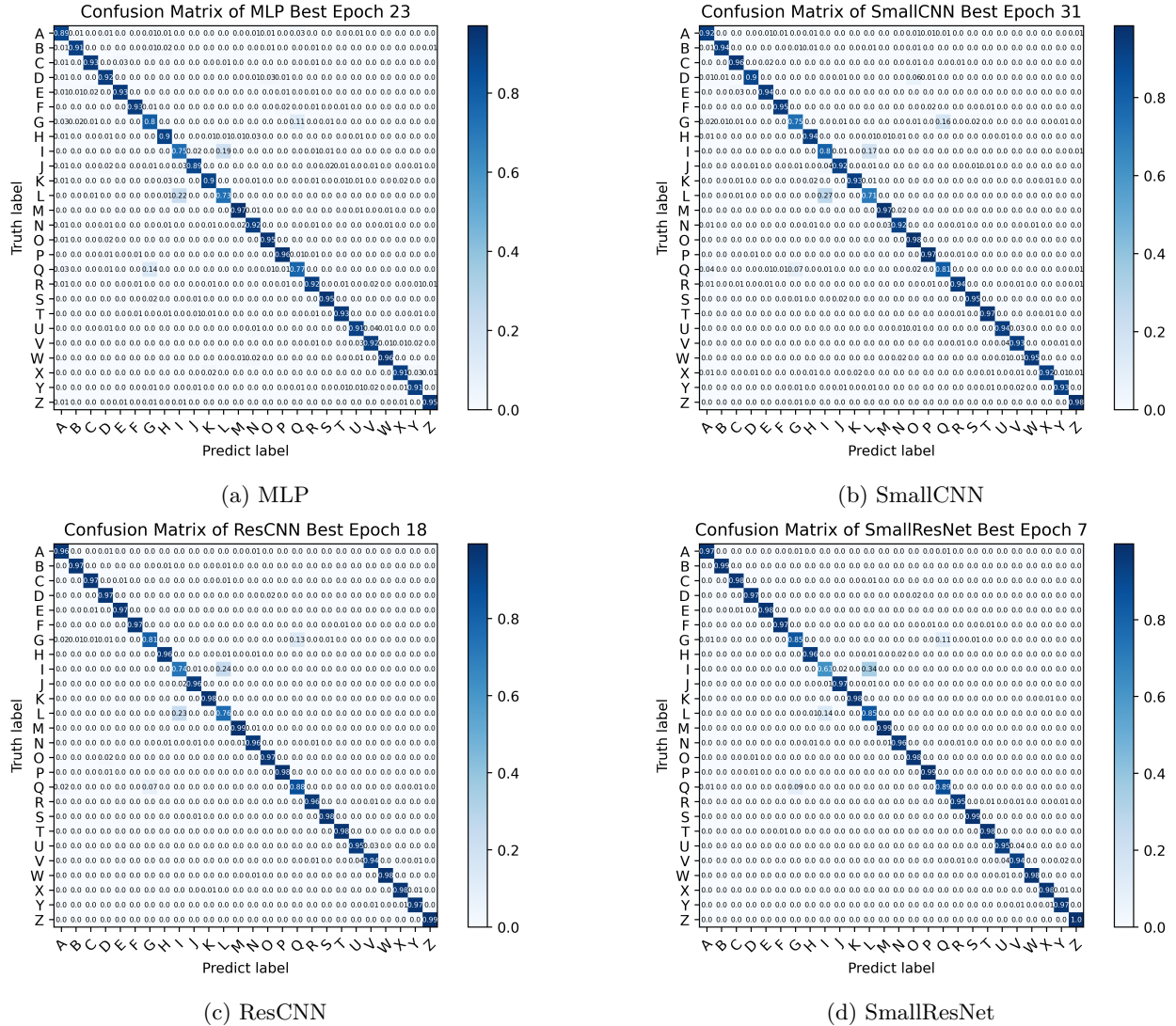


图 2: Confusion Matrix

整理所有错误率在 0.1 以上的结果为表 2。

可见，现在仍存在的最容易混淆的是 I 和 L，其次是 Q 和 G。在字形上，这两对字母比较相似，错误概率高属于正常现象。但是，注意到 SmallResNet 上 I 误判成 L 的概率达到了 0.34，明显高于其他模型，这有待进一步实验来确认原因。

Model	MLP				SmallCNN			ResCNN			SmallResNet		
Truth	L	I	Q	G	L	I	G	I	L	G	I	L	G
Predict	I	L	G	Q	I	L	Q	L	I	Q	L	I	Q
Possibility	0.22	0.19	0.14	0.11	0.27	0.17	0.16	0.24	0.23	0.13	0.34	0.14	0.11

表 2: Results with Error Rate over 0.1

## 2 计划

基于当前的混淆矩阵，我下周的计划是对当前的模型进行优化并调整相应超参数，尝试引入 attention 机制，使模型的准确率达到 0.95 以上。