

讨论模型的不同配置设计对 MNIST 分类任务的影响

白锦琪

2025 年 3 月 13 日

本实验基于 MNIST 手写数字分类任务，探究全连接神经网络中三个关键设计要素（网络深度、激活函数、学习率）对模型性能的影响。通过对比不同配置下的训练曲线和最终指标，来说明不同设定下模型性能变化。

一. 模型设计

1. 网络架构

表 1: 网络架构参数设置

组件	参数设置
输入层	784 节点 (28×28 图像展平)
隐藏层	动态配置深度和宽度
输出层	10 节点 (Softmax 输出)
优化器	Adam
损失函数	交叉熵损失
训练轮数	15 epochs
Batch Size	64

2. 实验变量配置

实验变量配置如下：

- 深度对比组：3 层、5 层、8 层
- 激活函数对比组：ReLU、Sigmoid、Tanh
- 学习率对比组：0.1、0.01、0.001
- 综合对比组：不同深度、宽度、激活函数组合

具体变量配置代码片段如下所示：

```
experiments = [  
# 深度对比组 (1)  
{ 'depth':3, 'width':256, 'activation':'relu', 'lr':0.001},  
{ 'depth':5, 'width':256, 'activation':'relu', 'lr':0.001},  
{ 'depth':8, 'width':256, 'activation':'relu', 'lr':0.001},  
  
# 激活函数对比组 (2)  
{ 'depth':4, 'width':256, 'activation':'relu', 'lr':0.001},  
{ 'depth':4, 'width':256, 'activation':'sigmoid', 'lr':0.001},  
{ 'depth':4, 'width':256, 'activation':'tanh', 'lr':0.001},  
  
# 学习率对比组 (3)  
{ 'depth':4, 'width':256, 'activation':'relu', 'lr':0.1},  
{ 'depth':4, 'width':256, 'activation':'relu', 'lr':0.01},  
{ 'depth':4, 'width':256, 'activation':'relu', 'lr':0.001},  
  
# 综合对比组 (4)  
{ 'depth':3, 'width':128, 'activation':'relu', 'lr':0.001},  
{ 'depth':5, 'width':512, 'activation':'tanh', 'lr':0.0005},  
{ 'depth':4, 'width':256, 'activation':'sigmoid', 'lr':0.005}  
]
```

二. 实验结果分析

1. 网络深度影响

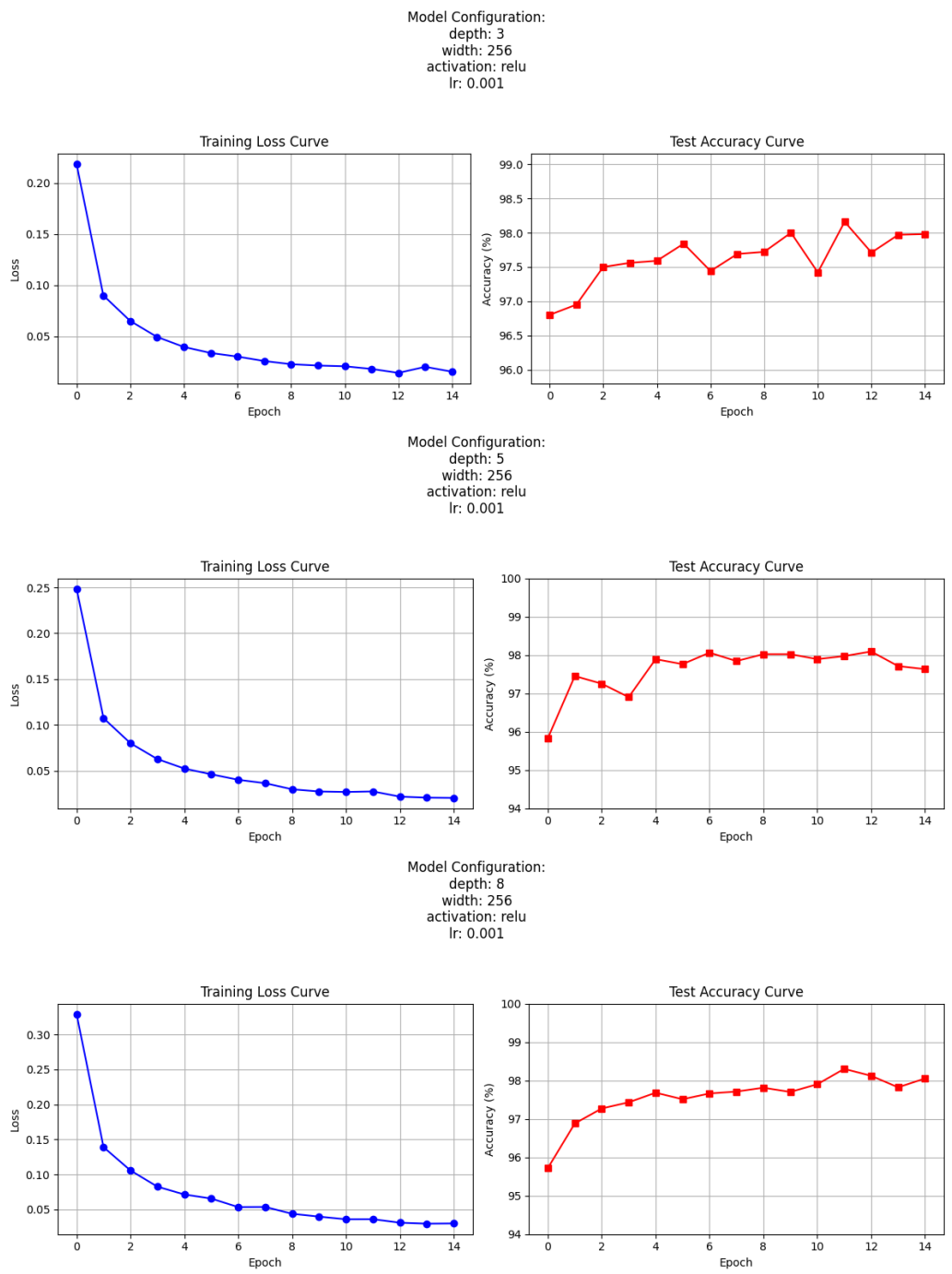


图 1: 不同深度网络的训练曲线对比

表 2: 网络深度对模型性能的影响			
深度	最终准确率	训练损失	收敛稳定性
3 层	98.10%	0.0163	持续提升
5 层	97.63%	0.0203	后期波动
8 层	98.05%	0.0299	收敛缓慢

2. 激活函数影响

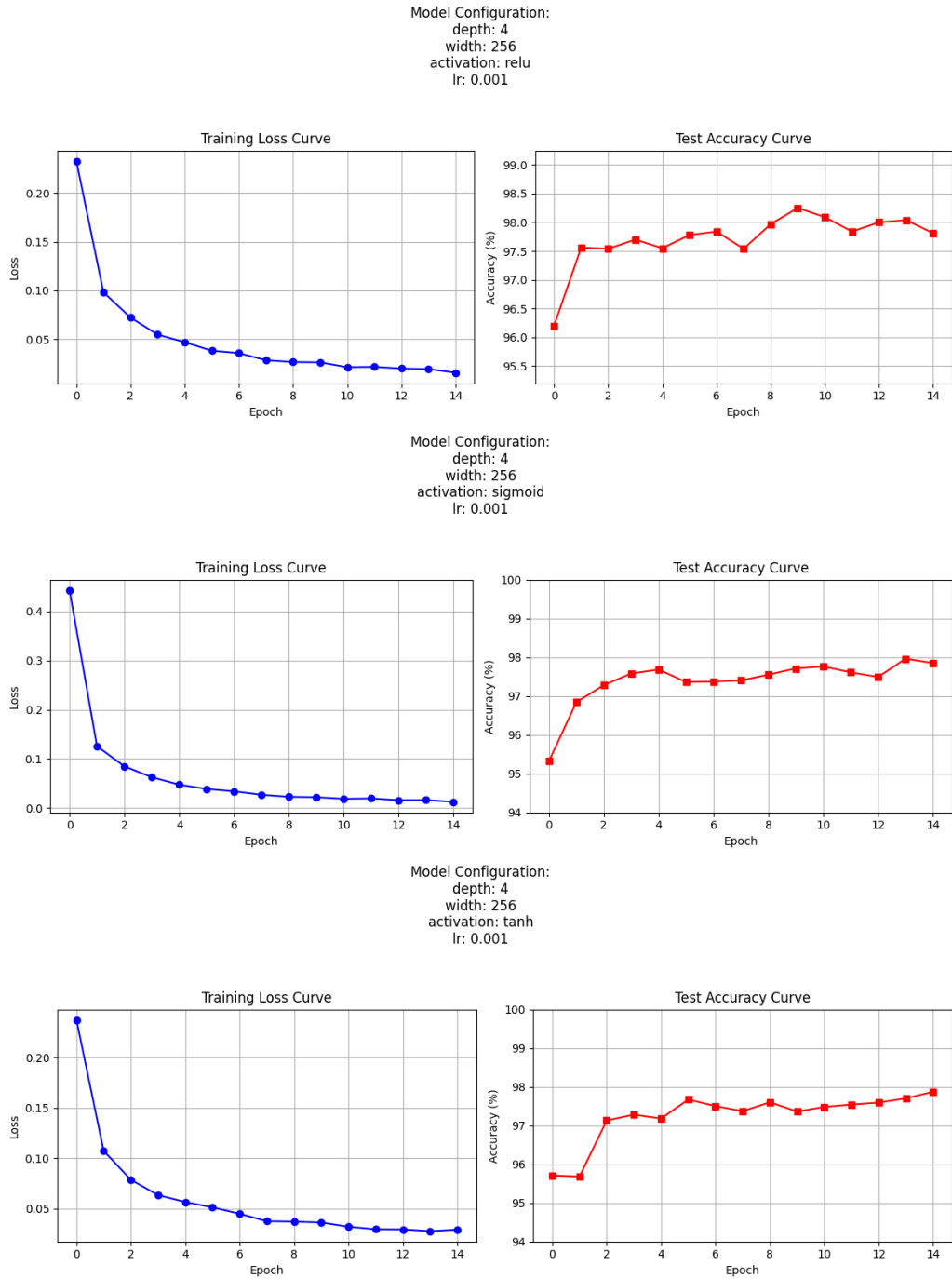


图 2: 不同激活函数的训练曲线对比

表 3: 激活函数对模型性能的影响

激活函数	最终准确率	训练损失	收敛速度
ReLU	97.83%	0.0157	最快
Sigmoid	97.85%	0.0126	中等
Tanh	97.87%	0.0292	最慢

3. 学习率影响



图 3: 不同学习率的训练曲线对比

表 4: 学习率对模型性能的影响

学习率	最终准确率	训练损失	收敛状态
0.1	10.09%	2.3100	完全未收敛
0.01	96.46%	0.1357	震荡收敛
0.001	97.81%	0.0157	平稳收敛

三. 结论

1. 性能对比表格

表 5: 不同影响因素的最优配置			
影响因素	最优配置	次优配置	需避免配置
网络深度	3 层 (98.10%)	5 层 (97.63%)	8 层 (收敛慢)
激活函数	ReLU (97.83%)	Tanh (97.87%)	Sigmoid (损失高)
学习率	0.001 (97.81%)	0.01 (96.46%)	>0.01 (无法收敛)
综合配置	3 层 ReLU+0.001	5 层 Tanh+0.0005	Sigmoid+ 高学习率

2. 结论

在本次实验中，结合上述结果和分析，我们可以得出以下结论：

- 深度选择：3-5 层为最佳区间，更深网络需配合残差连接
- 激活函数：优先使用 ReLU，复杂场景可尝试 Leaky ReLU
- 学习率设置：推荐 0.001 作为起点，配合学习率调度器
- 宽度影响：512 宽度比 128 宽度准确率提升 0.12

当然，上述结论只针对使用全连接神经网络在 MNIST 分类任务上得到的结果，不代表最终结果，不过也可以提供一些模型配置建议。想要得到更好更准确的结果，可以采用不同神经网络模型和不同复杂度的数据集及任务进行对比测试。