

Fashion-MNIST 多模型分类实验报告

王雨乾 PB23061103

2026-01-11

摘要

围绕 Fashion-MNIST 图像分类任务，实验以“从简单基线到复杂结构逐步升级”的方式，依次实现并对比多层感知机 (MLP)、经典 LeNet-5、改进卷积网络 (CNN)、残差网络 (ResNet)、Wide ResNet、DenseNet-BC 与 Capsule Network (CapsNet) 等 7 类模型，并在统一的数据读取、训练与评估流程下给出可复现实验记录；在训练策略方面，代码中引入了 Dropout、Weight Decay、学习率调度 (ReduceLROnPlateau / CosineAnnealingLR)、Early Stopping、数据增强与 Label Smoothing 等微调手段，使模型的泛化性能随结构与训练技巧的增强而逐步提升；最终结果显示 Wide ResNet-28-10 + Random Erasing 取得最高测试准确率 **94.50%**，ResNet 取得 **93.96%**，DenseNet-BC 取得 **92.39%**，CapsNet 取得 **91.58%**，而 MLP 尽管训练准确率达到 **98.69%**，测试准确率仅为 **89.83%**，体现了模型表达方式、结构归纳偏置与正则化强度对泛化差异的共同影响；所有实验代码与对应训练结果均已开源并发布于 GitHub: <https://github.com/wang-ava/Alexperiment>。

1 实验目标与学习路线

本实验的核心目标是完成 Fashion-MNIST 十分类任务，并通过逐步替换模型结构与训练策略来观察精度、训练时间与过拟合程度的变化规律；为使学习过程清晰可追溯，整体实现遵循“先建立能跑通的最小基线，再逐步引入更符合视觉任务归纳偏置的结构，再引入更先进的连接方式与训练技巧”的路线：首先用 MLP 将图像像素直接展平输入，快速得到可用的分类基准；随后引入卷积与池化，使网络能够利用局部纹理与空间结构信息，从而在相似外观类别之间形成更稳定的区分；在卷积网络能够稳定训练后，再进一步引入残差连接、宽化策略与密集连接等现代结构，使更深、更复杂的网络也能在较稳定的梯度传播下达到更高准确率；最后引入 CapsNet 作为“向量胶囊 + 路由”的代表性替代范式，观察其在该数据集上的收益与训练代价。

2 数据集、读取与预处理

2.1 数据集概述

Fashion-MNIST 由 28×28 的灰度图像构成，共 10 类，分别对应 T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot；实验代码从原始 `idx_gzip` 文件读取训练集与测试集，并在加载阶段将像素值归一化到 $[0, 1]$ 区间，形成适合神经网络优化的输入分布。

2.2 预处理与批量组织

实现中通过自定义的 `mini-batch` 生成器对数据进行打乱与分批，使每轮训练的样本顺序不同，从而减弱固定样本顺序导致的偶然偏差；同时，为兼顾运行速度与显存限制，多数模型脚本在主程序中依据 GPU 可用性与剩余显存动态调整 `batch size`，这一设计能在资源充足时提升吞吐，而在资源紧张时避免因显存不足导致训练中断。

3 模型方法：用通俗语言理解每一种结构

3.1 多层感知机 (MLP)：把图像当作长向量来记忆

MLP 的输入是展平后的 784 维像素向量，其本质是多次执行“线性变换 + 非线性激活”的叠加，从而把原始像素映射到 10 类 logits；由于该结构不显式利用邻近像素之间的空间关系，因此往往依赖大量参数去“记住”训练集中的统计规律，而当样本外观变化稍大时容易出现泛化不足；为缓解这一问题，代码中在隐藏层之后加入较强的 Dropout (0.4) 并配合 Weight Decay (10^{-4}) 与学习率调度器 (`ReduceLROnPlateau`)，使其训练更稳定且不至于过快陷入过拟合。

3.2 LeNet-5：经典卷积网络的“局部特征 + 逐步抽象”

LeNet-5 通过卷积层在局部感受野内抽取边缘、纹理等特征，并通过池化逐步降低分辨率、扩大感受野，从而把“像素级细节”逐层汇聚成“类别相关的高层表征”；在 Fashion-MNIST 中，该结构通常比纯 MLP 更符合视觉任务的归纳偏置，但其网络容量与正则化手段相对有限，因此性能提升往往有限且容易停留在中等水平。

3.3 改进 CNN：更深的卷积块 + BatchNorm + Dropout

在 LeNet 思想之上，改进 CNN 进一步堆叠多个卷积块，并在卷积后加入 BatchNorm，使每层激活分布更稳定，从而提升收敛速度并降低对初始化的敏感性；同时，代码中引入多处 Dropout（卷积特征使用 `Dropout2d`，分类头使用 `Dropout(0.5)`），并默认结合数

据增强与 Label Smoothing，从而在训练准确率不至于过高的情况下获得更小的训练/测试差距，体现出“强正则化下的稳定泛化”。

3.4 ResNet：用残差连接让更深网络更好训练

ResNet 的关键在于残差连接，其基本形式可写为

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \mathbf{x}, \quad (1)$$

其中 \mathcal{F} 表示若干卷积层的非线性变换；这一设计使梯度能够沿着恒等映射更直接地传播，从而缓解深层网络的退化问题；在本实验中，ResNet 配合 AdamW、CosineAnnealingLR、数据增强与 Label Smoothing 后取得显著更高的测试准确率，说明当网络深度上升时，结构性设计与训练技巧对可训练性与泛化能力具有共同作用。

3.5 Wide ResNet：把 ResNet “加宽” 以提升表达能力

Wide ResNet 的直观理解是减少极端深度、增加通道宽度，使每层可表达的特征子空间更大，从而在保持残差稳定训练的同时提升模型容量；本实现采用 Wide ResNet-28-10，并在训练时结合 Random Erasing（随机擦除）等增强策略，使模型不依赖某些局部区域的“捷径特征”，从而在外观相近类别（例如 Shirt 与 T-shirt/top）之间得到更鲁棒的边界；最终该模型获得本次实验最高测试准确率。

3.6 DenseNet-BC：让特征在层间“密集复用”

DenseNet 的核心连接方式可写为

$$\mathbf{x}_\ell = \mathcal{H}_\ell([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1}]), \quad (2)$$

其中 $[\cdot]$ 表示通道拼接；该结构使早期层特征能够直接被后续层使用，有利于特征复用与梯度传播；DenseNet-BC 在 DenseNet 基础上使用 bottleneck 与 compression 减少计算量，但在本实验环境下仍具有相对较高的训练开销，并且在该数据集上取得了比普通 CNN 更高、但略低于 Wide ResNet/ResNet 的准确率，这与不同结构对容量、正则化与训练动态的平衡差异相关。

3.7 Capsule Network：用向量胶囊与动态路由表示“部件与整体”

CapsNet 试图用向量而非标量表示特征，使向量长度刻画“存在概率”，向量方向刻画“姿态/属性”；其分类常用 margin loss，可写为

$$L_k = T_k \max(0, m^+ - \|\mathbf{v}_k\|)^2 + \lambda(1 - T_k) \max(0, \|\mathbf{v}_k\| - m^-)^2, \quad (3)$$

其中 T_k 为真实类别指示变量， \mathbf{v}_k 为第 k 类胶囊输出；尽管这一范式在某些场景中对几何变换更敏感，但在 Fashion-MNIST 上其收益并不必然超过强增强下的现代 CNN/ResNet，

同时动态路由也会引入额外计算与优化难度，因此本实验中其准确率位于中上水平，训练时间也明显高于轻量基线。

4 训练策略与微调记录

4.1 损失函数与 Label Smoothing

对于 CNN / ResNet / Wide ResNet / DenseNet-BC，训练阶段在交叉熵基础上引入 Label Smoothing (代码中 smoothing 系数设为 0.1)，其思想是把 one-hot 标签替换为“略微变软”的分布，从而降低过度自信带来的过拟合倾向；若类别数为 K ，平滑后的目标分布可写为

$$y'_k = \begin{cases} 1 - \varepsilon, & k = y \\ \frac{\varepsilon}{K-1}, & k \neq y \end{cases}, \quad (4)$$

其中 ε 为平滑系数。

4.2 优化器、Weight Decay 与学习率调度

代码中存在两类较明确的优化组合：

- MLP 与 LeNet-5：使用 Adam + Weight Decay (10^{-4})，并使用 ReduceLROnPlateau (以验证集准确率为监控信号) 进行更保守但稳定的学习率衰减，从而避免在较浅结构上出现震荡或过快收敛到次优解；
- CNN / ResNet / Wide ResNet / DenseNet-BC：使用 AdamW + Weight Decay (10^{-3})，并使用 CosineAnnealingLR，使学习率随 epoch 近似按余弦曲线平滑下降，典型形式为

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})(1 + \cos(\pi t/T)), \quad (5)$$

其中 T 为总 epoch 数；该策略往往在训练后期提供更细腻的“收敛打磨”。

4.3 正则化：Dropout 与 Early Stopping

为降低过拟合，MLP 在每个隐藏层后使用 Dropout(0.4)，CNN 在卷积与全连接层均使用 Dropout (卷积层 Dropout2d(0.25)，分类头 Dropout(0.5))，ResNet 在全连接前使用 Dropout(0.5)，Wide ResNet 在块内使用 dropout_rate=0.3；此外，CNN/ResNet/WideResNet/DenseNet-BC 的训练循环中实现了 Early Stopping (patience=15)，当测试准确率在一定轮数内不再提升时提前停止，从而避免在后期把训练集噪声继续“学进去”。

4.4 数据增强与 Random Erasing

数据增强模块提供随机旋转、平移、裁剪、水平翻转与随机擦除 (Random Erasing) 等策略，并在 CNN/ResNet/WideResNet/DenseNet-BC 的训练批次中调用；其中 Random Erasing 通过随机遮挡局部区域迫使模型利用更完整的语义线索而非局部纹理捷径，尤其对衣物类中外观高度相似的类别具有积极意义，这一点在后续类别级别分析中也能观察到其对 Shirt/Pullover/Coat 等类别的提升更明显。

5 实验设置与对比指标

七个模型均以最终一轮的训练/测试准确率作为主要性能指标，并记录总训练耗时；同时，为刻画过拟合强度，引入“泛化差距 (训练准确率 - 测试准确率)”作为辅助指标，使不同正则化强度与结构容量的差异能够被更直观地比较。

6 实验结果

6.1 总体结果对比

表 1 汇总了七种模型的 epoch、batch size、学习率、最终训练/测试准确率、泛化差距与训练时间（分钟），其数值来自各模型生成的训练报告文件。

表 1: 七种模型的总体性能对比								
模型	Epoch	Batch	LR	训练 Acc(%)	测试 Acc(%)	差距 (%)	时间 (min)	
Wide ResNet-28-10 + Random Erasing	120	128	0.0010	99.12	94.50	4.62	71.54	
残差网络 (ResNet)	120	256	0.0010	97.74	93.96	3.78	21.58	
DenseNet-BC	120	64	0.0010	94.24	92.39	1.85	31.43	
Capsule Network (CapsNet)	80	64	0.0005	98.54	91.58	6.96	9.32	
卷积神经网络 (CNN)	120	256	0.0010	92.48	91.21	1.27	40.16	
多层感知机 (MLP)	100	1024	0.0005	98.69	89.83	8.86	0.29	
LeNet-5	80	512	0.0005	91.43	88.92	2.51	0.40	

6.2 结果差异的原因分析：结构归纳偏置与训练技巧的共同作用

从表 1 可以观察到一个稳定趋势：当模型从“把图像当作向量”的 MLP 过渡到“显式利用空间结构”的卷积网络后，测试准确率出现了系统性提升，而当卷积网络进一步引入残差连接、加宽策略或密集连接时，准确率继续提高并逐步逼近 95 %；这一趋势说明，对于视觉任务而言，卷积操作所携带的局部连接与权值共享的归纳偏置能够显著减少对训练样本的“死记硬背”，从而把学习重点从像素级偶然性转向更稳定的局部形状与纹理模式。

与此同时，泛化差距也揭示了正则化与数据增强的重要性：MLP 的训练准确率达到 98.69%，但测试准确率仅为 89.83%，差距高达 8.86%，这往往意味着模型通过大量全连接参数拟合了训练集中的细节而非可迁移规律；相比之下，CNN 的训练准确率只有 92.48%，但测试准确率达到 91.21%，差距仅 1.27%，这一现象与其较强的 Dropout、Label Smoothing、CosineAnnealingLR 与数据增强组合相一致，即通过限制训练集上的“过度拟合能力”来换取更稳健的测试表现。

在高性能模型中，Wide ResNet 以 94.50% 测试准确率位列第一，其优势主要来自更强的表达能力（宽化带来的更大特征子空间）与更强的泛化约束（Random Erasing + Label Smoothing + AdamW + CosineAnnealingLR），从而在不显著牺牲训练稳定性的条件下提升了对外观细微差异的辨别能力；ResNet 的测试准确率为 93.96%，其核心收益来自残差连接带来的更易优化性；DenseNet-BC 虽然具备较强的特征复用能力，但在该数据集与当前训练配置下其收益略低于 Wide ResNet/ResNet，这可能与网络规模、增强策略强度、以及 Dense 连接下特征拼接导致的优化动态差异有关。

CapsNet 在本实验中达到 91.58%，其训练准确率很高但泛化差距也相对偏大，说明动态路由与 margin loss 并不会天然压制过拟合，且当数据集尺度与类别差异主要体现为纹理/轮廓而非复杂姿态变化时，强增强的现代 CNN 往往具有更直接的优势。

6.3 类别级别分析：困难类别与增强收益

为使结论更直观，表 2 对比了 MLP 与最优模型 Wide ResNet 在 10 个类别上的测试准确率（来自两份报告中的“各类别准确率”段落）；可以看到 Shirt、Pullover、Coat 等上衣类存在更明显的提升，而 Trouser、Sandal、Bag 等形状特征更稳定的类别在两种模型中都较高，这与类别间的视觉相似度与局部纹理差异程度密切相关。

表 2: MLP 与 Wide ResNet 的逐类准确率对比 (%)

类别	MLP	Wide ResNet	提升
T-shirt/top	85.3	88.9	+3.6
Trouser	97.9	99.1	+1.2
Pullover	83.0	93.7	+10.7
Dress	89.7	95.1	+5.4
Coat	83.7	93.0	+9.3
Sandal	97.1	99.6	+2.5
Shirt	72.6	82.2	+9.6
Sneaker	96.3	98.0	+1.7
Bag	97.0	99.5	+2.5
Ankle boot	95.7	95.9	+0.2

从逐类结果可以推断，Random Erasing 与更强卷积表征对“上衣类细粒度差异”的

帮助更明显，因为遮挡与形变增强迫使模型学习到更分散、更鲁棒的判别线索，而不是把注意力锁定在某个固定局部纹理上；同时，Label Smoothing 与 AdamW 的组合也会降低过度自信的错误预测概率，从而在边界模糊类别上减少“高置信错判”的现象。

7 结论与展望

综合结构与训练策略两条主线，实验表明：当任务为标准灰度服饰分类时，卷积网络相对 MLP 的优势不仅来自更强容量，更来自对空间结构信息的有效利用；当卷积网络进一步叠加残差连接、宽化策略与密集连接等现代结构后，测试准确率可稳定提升到 92% 以上，而在加入 Random Erasing、Label Smoothing、AdamW 与 CosineAnnealingLR 等微调手段后，最高准确率可达到 94.50%；在未来改进方向上，若希望进一步提升 Shirt/T-shirt 等难分类类别的区分能力，可考虑引入更精细的数据增强组合（例如更强的遮挡比例搜索或更贴合衣物形变的仿射扰动）、更系统的学习率 warmup 与正则项系数网格搜索，并通过混淆矩阵与错误样本可视化进一步定位模型仍然依赖的非鲁棒特征。