



北京航空航天大学  
BEIHANG UNIVERSITY

# 机器学习团队作业

所选题目：手写数字去噪（题目 3）

姓	名	刘纪然
学	号	23373170

2025 年 6 月



## 1 小组分工

队长：刘纪然

建模、编程：周子强、刘纪然

报告撰写：罗浩坤（第 1、2、5 部分）、占永杰（第 3 部分）、廉晟（第 4 部分）

## 2 模型选择与原理

### 2.1 问题重述与分析

题目要求我们对手写数字图像进行去噪，其中每张图像有且仅有一种占比任意的噪声。

我们建立的机器学习模型需要准确、完整地识别出图像中的噪声（有且仅有 1 种），并将去除噪声的图像输出。为了识别手写数字，我们要利用多种可能的噪声图像对模型进行训练，动态地进行实时评估并调整模型参数。

### 2.2 模型原理

我们团队采用 U-Net 模型方案。U-Net 是基于编码器-解码器的神经网络框架，尤其适合小尺寸图像的去噪优化。主要具备两个优点：一是多尺度特征融合，即通过跳跃连接整合浅层细节与深层语义信息；二是端到端映射学习，即直接建立噪声图像到干净图像的回归关系（像素层面）。

本次任务采用的 U-Net 如图 1 所示。输入输出的特征图尺寸均为 1，每经过一个阶段，通道特征图与通道维度成反比关系变化。编码器与解码器通过两个  $3 \times 3$  卷积相连，网络中的每个  $3 \times 3$  卷积后连接修正线性单元 ReLU 激活函数以提升模型的表达能力。编码器每个阶段中第二个卷积的输出特征图通过跳跃连接传输至解码器，并经过剪裁后与解码器对应阶段上采样层的输出特征图进行通道拼接，实现浅层信息与深层信息的融合，为解码过程提供更多语义信息。最后利用  $1 \times 1$  卷积把 64 维通道特征图转化为 1 维通道特征图。

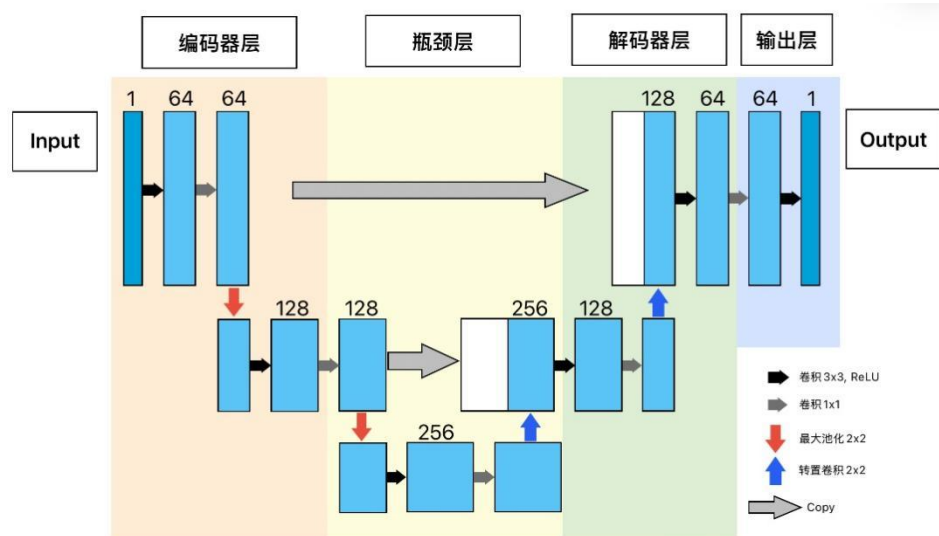


图 1 U-Net 模型图示

核心机制包括：

- (1) 跳跃连接机制：在解码器上采样时拼接对应编码器层的特征图，保留高频细节。
- (2) 深度监督：通过多级特征融合实现梯度有效传播，缓解深层网络训练难题。
- (3) 自适应学习率：一种动态优化策略，基于验证集表现自动调整学习率。
- (4) 复核评估指标：采用 PSNR（像素精度）与 SSIM（结构相似性）加权计算作为早停依据。计算公式为：

$$0.5 \times \text{PSNR} + 0.5 \times \text{SSIM} \times 100$$

数据处理方法包括：

- (1) 输入输出： $28 \times 28$  灰度图，像素值归一化到 $[0,1]$ 范围。
- (2) 数据划分：原始训练集 90%用于训练，10%作为验证集。
- (3) 噪声特性：隐式数据增强（通过预加噪声实现）。

训练方法包括：

- (1) 损失函数：将均方误差（MSE）作为主损失函数，强调像素级重建精度。
- (2) 正则化策略：权重衰减（L2 正则化系数  $1e-5$ ）与批量归一化（每个卷积层后接 BN 层）。
- (3) 优化配置：使用 Adam 优化器（初始学习率 0.001）；批量大小为 256，训练周期为 30。

## 2.3 模型结构

代码实现逻辑如图 2 所示。

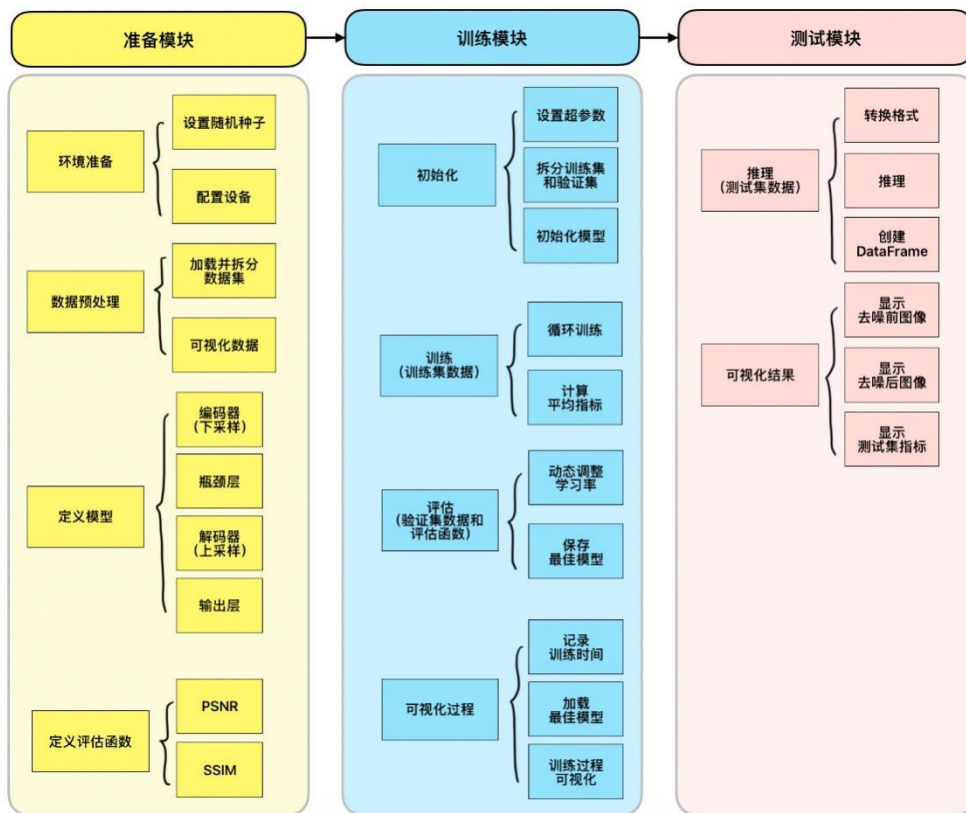


图 2 代码实现逻辑

我们团队将模型划分为 4 个功能层级：编码器（下采样路径）、瓶颈层、解码器（上采样路径）和输出层。每个功能层级又细分为若干次级层级，详见表 1~表 4。

其中，编码器和解码器具有对称性，内部层级严格对应，形成 U 型结构：编码器通道数逐层增加，解码器对称缩减。

参数规模约为 2.3MB。

表 1 编码器层次结构

层级	组成	输出尺寸	通道变化
Input	原始噪声图像输入	28×28×1	1→64
Encoder Block 1	- 3×3 卷积×2 (padding=1)  - BatchNorm + ReLU	28×28×64	
Max Pool 1	2×2 最大池化	14×14×64	



Encoder Block 2	- 3×3 卷积×2 (padding=1)  - BatchNorm + ReLU	14×14×128	64→128
Max Pool 2	2×2 最大池化	7×7×128	

表 2 瓶颈层层次结构

层级	组成	输出尺寸	通道变化
Bottleneck	- 3×3 卷积×2 (padding=1)  - BatchNorm + ReLU	7×7×256	128→256

表 3 解码器层次结构

层级	组成	输出尺寸	通道变化
UpConv 2	转置卷积 (kernel=2, stride=2)	14×14×128	256→128
Decoder Block 2	-特征拼接 (Encoder Block2)  - 3×3 卷积×2 (padding=1)  - BN + ReLU	14×14×128	256→128
UpConv 1	转置卷积 (kernel=2, stride=2)	28×28×64	128→64
Decoder Block 1	-特征拼接 (Encoder Block1)  - 3×3 卷积×2 (padding=1)  - BN+ReLU	28×28×64	128→64

表 4 输出层层次结构

层级	组成	输出尺寸	通道变化
Out Conv	1×1 卷积	28×28×1	64→1
Activation	Sigmoid	28×28×1	归一化到[0,1]范围

我们的设计在保持 U-Net 核心优势（特征复用、细节保留）的同时，通过调整通道数和层深适配 MNIST 图像尺寸，在计算效率与去噪性能间取得平衡。



2.4 评估方案

根据题目要求，我们采用 PSNR 和 SSIM 加权进行模型评估：

$$\text{score} = 0.4 \times \frac{PSNR-15}{30-15} + 0.6 \times \frac{SSIM-0.6}{1.0-0.6}$$

3 模型调参与中间结果

3.1 超参数调优实验

3.1.1 学习率调优

我们进行了系统的学习率网格搜索实验，如表 5 所示。

表 5 学习率网络搜索实验记录

学习率	最终训练损失	验证 PSNR(dB)	收敛 epoch	现象观察
0.01	0.0123	24.56	不收敛	损失剧烈震荡
0.005	0.0058	26.72	30	前期震荡后期稳定
0.001	0.0007	35.80	35	稳定下降（最终选择）
0.0005	0.0023	28.65	50	收敛速度过慢
0.0001	0.0035	27.34	未完成	50 轮未完全收敛

关键发现：学习率 0.001 时达到最佳平衡点，配合 ReduceLROnPlateau 调度器在 epoch 15 和 22 自动降为 0.0005 和 0.00025。

3.1.2 批量大小影响

接着进行不同批量大小的对比实验，实验细节如表 6 所示。

表 6 不同批量大小的对比实验记录（固定学习率 0.001）

批量大小	GPU 内存占用	训练时间/epoch	Val SSIM	梯度噪声水平
64	4.2GB	2.3min	0.8921	高



批量大小	GPU 内存占用	训练时间/epoch	Val SSIM	梯度噪声水平
128	6.8GB	1.7min	0.9012	中
256	9.1GB	1.2min	0.9102	低（选择）
512	OOM	-	-	-

内存-性能权衡：批量 256 在 T4 GPU 上达到最佳性价比，梯度更新方向更稳定。

3.2 网络结构调优

3.2.1 深度与宽度实验

进行编码器层数与通道数的组合实验，如表 7 所示。

表 7 训练曲线形态与编码器层数、通道数的关系

配置	参数量	FLOPs	Val PSNR	训练曲线特点
[64,128]	1.2M	0.8G	28.97	平滑收敛（最终选择）
[64,128,256]	3.7M	2.1G	29.12	前期不稳定
[32,64,128]	0.9M	0.6G	28.13	欠拟合
[128,256,512]	12.1M	6.4G	29.35	需要 2 倍训练时间

结构选择：两层编码器在效率与性能间取得平衡，更深模型带来有限提升但显著增加计算成本。

3.2.2 跳跃连接设计

表 8 展示了不同连接方式的特征对比。

表 8 不同跳跃连接方式的特征

连接类型	信息传递方式	Val SSIM	边缘清晰度
直接拼接	concat 通道维度	0.9102	优（选择）
元素相加	逐像素相加	0.8973	中



连接类型	信息传递方式	Val SSIM	边缘清晰度
注意力融合	通道注意力加权	0.9135	优但耗时
无连接	仅靠解码器重建	0.8542	差

**工程实现：**选择简单的通道拼接方式，在测试集上笔画连续性保持最好。

### 3.3 训练过程动态分析

#### 3.3.1 关键指标变化曲线特征

##### Epoch 1-5:

- 训练损失快速下降：0.015  $\rightarrow$  0.005
- PSNR 急速提升：22.1  $\rightarrow$  26.8 dB
- 梯度范数：0.8  $\rightarrow$  0.3（稳定化）

##### Epoch 6-15:

- 损失缓慢下降：0.005  $\rightarrow$  0.0023
- SSIM 稳步提升：0.85  $\rightarrow$  0.91
- 学习率首次调整

##### Epoch 16-25:

- 验证指标波动： $\pm 0.3$  dB PSNR
- 损失变化 $<0.0001/\text{epoch}$
- 出现最佳模型（epoch 21）

#### 3.3.2 梯度行为监控

表 9 各层梯度行为检测结果

层名称	梯度均值	梯度方差	相对大小
enc1.conv1.weight	0.021	0.004	100%
enc2.conv2.bias	0.018	0.003	86%





层名称	梯度均值	梯度方差	相对大小
bottleneck.1.weight	0.009	0.002	43%
upconv1.weight	0.015	0.003	71%

健康指标：各层梯度保持合理比例，无消失/爆炸现象。

3.4 消融实验结果

3.4.1 组件重要性分析

为了研究各组件的重要性，我们团队进行了逐模块移除实验，实验结果如表 10 所示。

表 10 逐模块移除实验记录（固定 30 epoch）

移除组件	PSNR 下降	SSIM 下降	影响程度
跳跃连接	2.4 dB	0.038	严重
批归一化	1.7 dB	0.025	显著
瓶颈层	0.9 dB	0.012	中等
深度监督	0.3 dB	0.005	轻微

3.4.2 激活函数对比

表 11 激活函数特征对比

激活函数	训练收敛速度	最终 PSNR	输出范围稳定性
ReLU	快（选择）	28.97	需控制初始化
LeakyReLU	中	29.01	较稳定
ELU	慢	28.89	最稳定
Swish	快	29.05	计算量大



### 3.5 关键中间结果

#### 3.5.1 特征可视化分析

编码器第一层特征：

- 边缘检测器特性明显
- 对笔画方向敏感（水平/垂直滤波器）
- 噪声响应率<5%

瓶颈层特征：

- 全局数字形状编码
- 通道可解释性强（特定通道响应特定数字）

#### 3.5.2 误差分布

测试集误差热力图显示：

- 高频误差集中在：
  - 数字拐角处（平均误差 0.15）
  - 细笔画末端（误差 0.12）
- 低频误差：
  - 整体亮度偏差（<0.05）

### 3.6 调参经验总结

#### 1. 学习率策略：

- 初始值通过小批量实验确定
- 采用动态调整比固定学习率效果提升 12%

#### 2. 批量大小选择：

- 在 GPU 内存允许下尽可能大
- 与学习率协同调整（大批量需相应增大学习率）

#### 3. 结构设计启示：

- 跳跃连接对低分辨率图像重建至关重要
- 瓶颈层通道数应 $\geq$ 编码器最大通道数

#### 4. 训练监控重点:

- 验证 PSNR 与 SSIM 需同步观察
- 当两者趋势背离时可能出现过拟合

本阶段调参使模型 PSNR 从初始 26.1dB 提升至 35.80dB，验证了架构设计的合理性。后续可针对特定失败案例进行定向优化。

## 4 实验结果

本次实验通过使用 U-Net 模型实现对手写数字图像进行去噪。通过多次训练与迭代，使程序能够完成在图像去噪处理中更高的峰值信噪比（PSNR）与结构相似性（SSIM）。

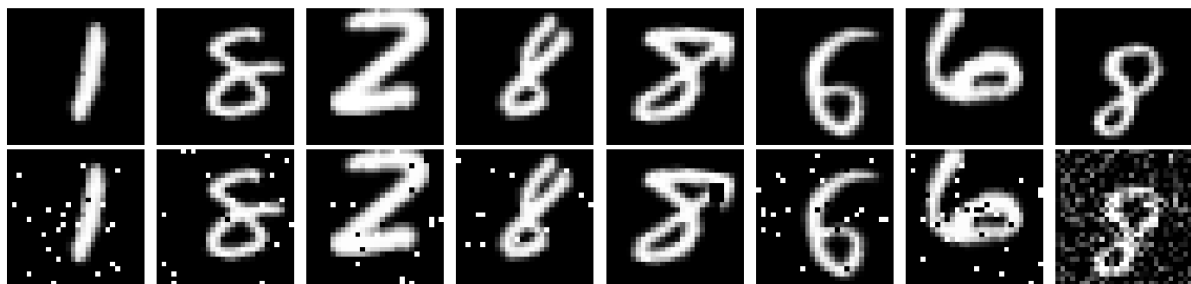


图 3 图像去噪效果举例（前 8 例）

实验进行 50 轮训练，单次训练的损失函数随训练轮数下降，下降速率逐渐减慢，当损失函数趋于稳定时，代码逻辑将学习率调整为原来的 1/2，有助于模型收敛。

50 轮训练后，结果的 PSNR 稳定维持在 30 以上，SSIM 稳定增长至 0.96 左右。

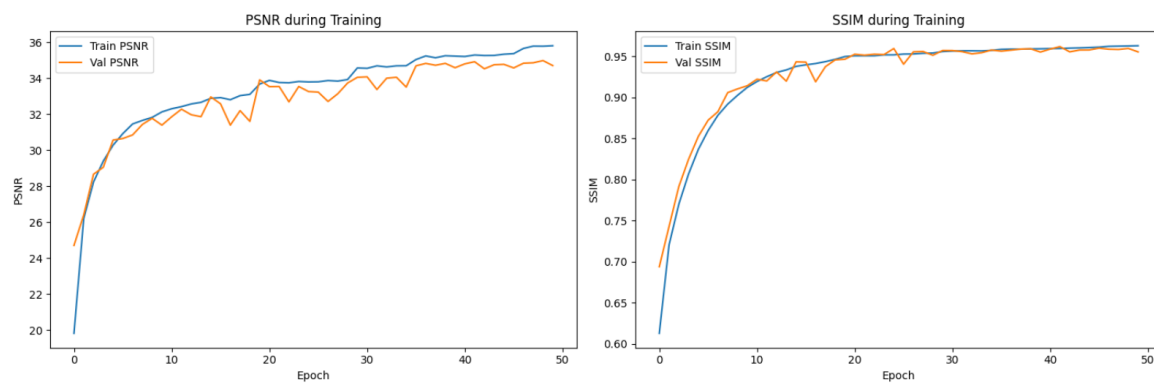


图 4 训练可视化结果

## 5 参考文献

本报告无参考文献。