

Task-3

一、计算机图像表示方式/图像清晰度影响因素：

1、计算机如何表示图像：

在计算机中，图像通常以**数字矩阵**形式存储。对于灰度图像，可表示为二维矩阵，矩阵中每个元素对应像素的灰度值，(0-255, 0 为黑, 255 为白)；彩色图像常采用 RGB 模型，以三维矩阵存储（高度×宽度×3），3 个通道分别对应红，绿，蓝分量的强度值。从底层存储来看，本质是将连续的视觉信息离散化为像素点的数值集合，这一过程涉及采样和量化。

2、图像清晰相关因素：

从**成像角度**，与光学系统的分辨率（区分或分辨物体细节的能力，即能清晰呈现两个相邻点，线或微小结构的最短距离）高度相关；原理：

从物理本质来说，受光的衍射现象影响，光学系统对点光源的成像并非理想点，而是形成具有一定分布的衍射光斑。根据瑞利判据，当一个光斑的中心与另一个光斑的边缘重合时，这两个点刚好能被分辨，此时两个点间的距离为系统的分辨率极限。

而影响分辨率主要因素有：

(1) 光学元件的性能，镜头焦距，孔径大小等，孔径越大，衍射影响越小，分辨率通常较高；

(2) 光源波长，波长越短，衍射效应越弱，系统越易分辨细节，分辨率越高；

从**数字图像本身**，像素间的对比度，边缘锐利度，噪声水平是影响图像视觉清晰度的核心因素；

像素间对比度：指相邻像素或区域间的亮度/色彩差异，对比度高时，明暗，色彩界限分明，人眼能快速区分不同区域，图像显得层次清晰，反之对比度低会导致区域边界模糊；

边缘锐利度：边缘是图像中物体轮廓，纹理变化的关键区域，梯度大小量化了像素值在边缘处的变化速率；

噪声水平：噪声是图像中随机的无用信号。噪声水平高时会干扰像素的真实信息，掩盖细节和边缘；

二、是否存在 PSNR 高但人眼看起来很差的图像？如何解决评估误差？

PSNR（峰值信噪比）是基于像素误差的评估指标，PSNR 计算公式：

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

其中 MAX_I 是像素的最大值，MSE 是均方误差。当图像存在结构性失真，但像素误差整体不大时，PSNR 可能不大，可人眼对结构，语义等信息更敏感，会觉得质量差。例如对添加细微，规律的块状噪声，若块内像素误差小，PSNR 下降不明显，但人眼易察觉瑕疵。

解决评估误差方法：

- （1）引入**感知驱动**的评估方法，如 SSIM（结构相似性），通过对比图像局部结构信息（亮度，对比度，结构）评估质量；
- （2）采用基于深度学习的**感知指标**，如 LPIPS，利用预训练的深度卷积网络，提取图像的高层语义和感知特征，通过计算两张图像在感知特征空间上的距离来衡量它们在感知视觉上的差异，LPIPS 更贴近人类主观感受；

三、任务阐述，技术原理，现实应用：

这是一个图像复原任务，目标是从退化图像中恢复出高质量，接近原图的结果。

技术原理：早期以**模型驱动方法**为主，如基于变分法的去模糊：

基于变分法的去模糊是将图像去模糊问题转化为一个能量函数的优化问题，假设退化图像由清晰图像经过模糊和噪声干扰生成，通过构建包含“数据保真项”和“正则化项”的能量函数，找到使得该函数最小化的清晰图像估计值；

其中 **TV 正则化**是典型代表，其**数据保真项**通常衡量退化图像与清晰图像经模糊操作后结果的差异（如基于均方误差）确保估计的清晰图像与退化图像在数据层面的一致性，而 **TV 正则化项**则通过约束图像的总方差，鼓励图像保持空间连续性和边缘的锐利性，从而在去模糊的过程中抑制噪声放大，通过求解该能量函数的最小值，即可得到

去模糊化后的图像。参考《Image restoration by sparse 3D total variation》

随着深度学习发展，基于 CNN 的方法成为主流，如 DnCNN 用于去噪：

核心思路是通过“残差学习”简化去噪目标—模型不直接学习从退化图像到清晰图像的映射，而是学习退化图像中噪声的残差，再用退化的图像减去预测的残差，得到去噪结果。参考《Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising》

EDSR 用于超分辨率

其核心特点包括：

- **强化残差学习**：通过增加残差块数量和调整残差连接方式，增强网络对图像细节特征的学习能力，缓解深层网络训练中的梯度消失问题。
- **移除批量归一化（BN）层**：在超分辨率任务中，BN 层可能引入不必要的噪声或破坏图像细节，移除后能更精准地恢复高频信息（如边缘、纹理）。
- **高效上采样模块**：采用亚像素卷积（PixelShuffle）作为上采样操作，在提升图像分辨率的同时，保持特征提取的连续性，避免传统插值方法导致的模糊。

EDSR 在多个公开数据集（如 Set5、Set14）上表现优异，能有效将低分辨率图像恢复为高分辨率版本，为后续的图像分析、识别等任务提供更清晰的输入，相关原理可参考论文《Enhanced Deep Residual Networks for Single Image Super-Resolution》。



后续又发展出 Transformer 结合的模型，如《Restormer: Efficient Transformer for High-Resolution Image Restoration》

现实应用：

安防领域图像去模糊，去噪以还原清晰监控画面；文物数字化中修复老旧，破损图像；医学影像处理中提升 CT，MRI 图像质量辅助判断；还可用于数字取证，像本题判定图像真伪；

四、图像退化的建模，常见的退化类型：

退化建模：一般将退化过程建模为： $y = H(x) + n$ ， x 是原图， y 是退化图， $H(\cdot)$ 是退化算子， n 是噪声。例如运动过程可建模为线性卷积 $y = x * k + n$ ， k 是模糊核；

下采样退化是对原图进行采样操作后结合噪声引入；

常见的退化类型有：

模糊（运动模糊，高斯模糊等），噪声（高斯噪声，椒盐噪声（部分噪声值被强制设置为最大/最小值），压缩失真，几何畸变等

五、训练数据集的生成：

核心目标为合成训练所需的“退化图像-原图对”，任务可拆解为：

- (1) 利用 GitHub 上的 original 清晰图像集，模拟真实退化，构建训练数据集；
- (2) 生成的数据集需与任务情景匹配，为后续模型训练提供“输入-标签”的配对；

实现思路：

- (1)：使用 GitHub 仓库中的 original 目录的清晰图像作为数据基础；
- (2)：通过随机选择模糊类型（高斯模糊和运动模糊）和噪声类型（高斯噪声，椒盐噪声）模拟真实场景的图像退化，增加数据多样性；
- (3)：为每个清晰原图生成三个退化样本

运行代码后结果：生成结果在 train_dataset 里面

```
训练集生成完成！  
原图保存路径： /Users/duqiu/Desktop/焦点计划/train_dataset/original  
退化图保存路径： /Users/duqiu/Desktop/焦点计划/train_dataset/degraded  
(ML) duqiu@duqiudembp Task-3 %
```

合成方式与实际退化的差距：

1、退化机制简化性差异：

代码中采用“单一模糊类型（高斯/运动）+单一噪声类型（高斯/椒盐）的线性叠加模式”与实际退化中的多因素耦合的非线性退化机制存在显著差异；

真实图像退化是光学系统，成像过程，传输存储等多环节的累积效应，还包含：

光学像差（球差，色差，导致边缘细节失真而非均匀模糊）；

压缩失真（JPEG 等有损压缩引入的块效应）

光照畸变（明暗不均，眩光，导致局部像素值非线性偏移）

传感器非线性噪声（低光环境下的泊松噪声，与光照成正相关）

2、退化过程的线性化假设偏差：

代码是先模糊后叠加噪声的做法，而真实退化过程具有非线性耦合特性：

例如，相机抖动（运动模糊）发生时，传感器同时接收光信号并产生噪声，模糊与噪声存在“同步时空耦合”

3、模糊核的统计分布偏差：

代码中运动模糊仅支持水平/垂直两个方向，且核内像素值呈“均匀分布”，而真实相机抖动的运动轨迹服从高斯分布或者 Levy 飞行分布，导致模糊核呈现“中间强，两边弱”的渐变特性；

且真实失焦模糊的高斯核标准差与镜头光圈大小呈负相关，且服从正态分布，但在代码中随机生成（均匀分布）

六、本题中图像噪声类型的评估/验证的方式：

1、高斯噪声判定：

对退化图与原图的残差（ $N = I_{degraded} - I_{original}$ ）进行统计，若残差符合零均值高斯分布（通过 Shapiro-Wilk 正态性检验，且 p 值大于 0.05），且噪声方差在图像平滑区域保持稳定，则判定为高斯噪声；

2、椒盐噪声判定：

计算残差中像素值为 255/0 的比例，若此类像素占比大于 0.1%，且呈随机分布（通过了空间自相关分析，Moran's I 接近于 0），则判定为椒盐噪声；

3、混合噪声判定：

若残差直方图“高斯分布基底加少量极值点”，且正态性检验 $p < 0.05$ ，则 1 判定为高斯与椒盐混合噪声；

验证方法：

定性：

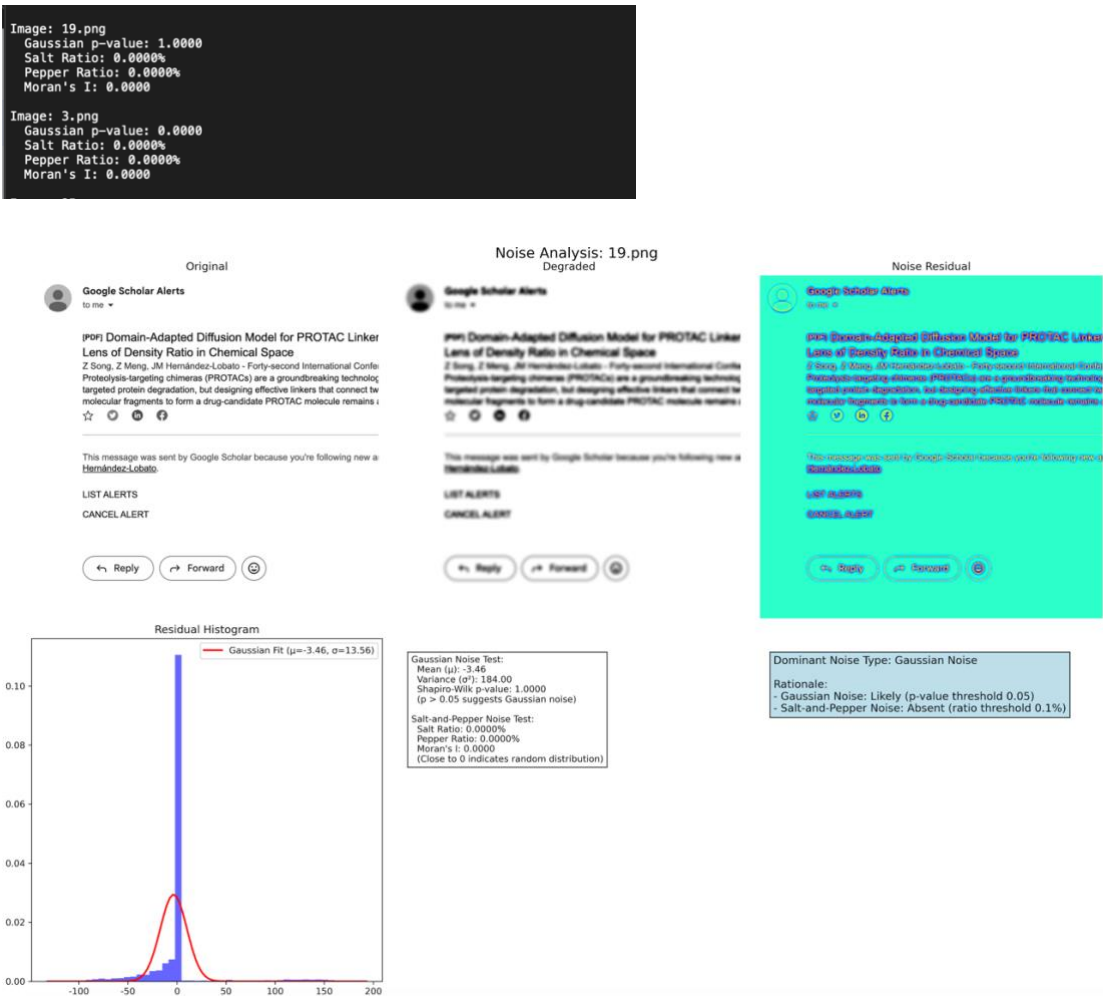
选取典型样本，通过视觉对比观察噪声形态，高斯噪声表现为“颗粒状均匀分布”，椒盐噪声表现为随机黑白点；

定量：

绘制残差图像的概率密度函数，与理论高斯分布图像对比；

计算噪声的峰值信噪比与噪声方差，分析不同退化样本中噪声强度的分布规律；

以某个 png 图像得到结果（执行 evaluate_noise.py 后）为例：（截取部分）



七、适合的模型及其依据：

1、适合的模型：

MIRNetV2（CVPR 2022）：

适配原因：设计**多尺度特征交互与动态选择**机制，可同时处理噪声（高斯/椒盐/混合），模糊，压缩等复合退化。对文本截图，多尺度分支能够精确捕捉字符“局部笔画”和“全局布局”依赖，在真实复杂退化中经过大量实验验证有效；

学术支撑：“MIRNetV2: Learning Enriched Features for Real Image Restoration and Enhancement”

HINet：

适配原因：采用**层次化信息蒸馏**压缩冗余特征，提升模型的效率与性能，尤其适用于处理含多元噪声，模糊等复杂退化的图像。

层次化信息蒸馏：模型通过多个子网络（蒸馏者与学习者）进行特征传递，逐步压缩冗余信息，保留关键复原特征，这种机制在降低计算量的同时，增强了对细节的捕捉能力。

学术支撑：*"Hierarchical Information Distillation for Image Restoration"*

改造型的 DnCNN（多专家+噪声判别）

在经典 DnCNN 基础上，新增噪声判别分支（识别高斯/椒盐/混合）与多专家模块（不同专家处理特定噪声），保留 CNN 高效性的同时，动态适配多元噪声。针对文本情景，可精确区分噪声类型并针对处理，小数据集训练也不容易过拟合；

学术支撑：

"Adaptive Multi-Noise Removal with CNNs"

2、不适合的模型：

纯传统方法（BM3D，中值滤波等）：依赖显示噪声先验，例如 BM3D 对高斯假设强，中值滤波仅适用于椒盐，多元噪声下无法自适应切换策略。处理文本截图时，易破坏字符边缘，丢失笔画细节

单一噪声专用 CNN：仅优化特定噪声，多元噪声场景下泛化能力骤降，文本截图含有复合退化时，无法同时处理模糊，压缩失真，导致字符复原不完整。

无结构先验的大模型（纯 ViT，未经优化的 Restormer）：缺乏文本结构先验，注意力机制易忽略字符“小尺度笔画细节”，复原后字符模糊，可读性差，且模型参数多，训练成本高。

八、数据增强：

数据增强是在不改变数据核心语义的（如图像的文本内容）的前提下，通过对原始数据施加合理的变换（如噪声，几何，色彩调整等），生成更多样化的训练样本的技术。核心目的是：

- 1、扩大训练集规模，缓解模型过拟合；
- 2、模拟真实场景中的数据变异，提升模型的鲁棒性；
- 3、让模型学到更通用的特征

由于是基于原有数据做模糊加噪声生成的训练集，在这次做数据增强时候仅仅考虑几何变换和光照调整方面。

见 data_augmentation.py 的代码；

九、跨语言训练数据可行性：

图像去噪复原的本质是学习“退化模式->清晰图像的映射关系”，其中

1、核心理论基础：

底层特征（通用）： 噪声分布（如高斯，椒盐），模糊核，压缩失真等退化模式与语言无关；

而**高层特征（语言相关）：** 文本的字符结构（如中文笔画复杂度，英文字母形态），排版规律**存在语言特异性；**

2、中文图像的适配性分析：

（1）适合作为训练数据的场景：模糊/噪声主导的退化：

若任务中退化以通用噪声（高斯），模糊为主，中文图像的底层退化特征可迁移，例如高斯噪声的统计分布与语言无关，模型学习的“高斯噪声抑制”能力可复用。且中文图像引入的字符结构多样性，可迫使模型学习更泛化的“结构保留”能力，间接提升英文场景的字符边缘复原效果；

（2）不适合的场景：语言特异性退化修复：

若任务需要精确复原英文特有的字符细节，中文图像的高层结构差异会干扰模型学习。[“Language-Specific Image Restoration: A Case for Latin vs. Non-Latin Scripts”](#), ACM MM 2023, 验证语言特异性结构对复原的影响)。

3、德语、法语的适配性

（1）高适配性依据：

字符结构相似性： 德语，法语与中文同属拉丁语体系，字符基于拉丁字母演化，高层文本结构高度兼容。

噪声分布共性： 若训练数据与任务数据的退化模式一致，德语/法语图像的底层退化特征可完全复用，模型无需重新学习噪声抑制

（2）局限性依据：

语言特有的字符/排版： 若德语文本含有大量的哥特手写字符，法语含有特殊排版，与

英文场景差异较大，可能引入干扰；

4、总结：

中文图像数据或者**英文/德文数据可适量引入**，利用跨语言退化特征的共享性提升模型的鲁棒性，但需要控制非英语数据的比例，避免语言特异性结构过度干扰；