

生成模型实现多张 LDR 重构 HDR 及 DRDB 改进

田晶怡¹ 张玉梅¹

1. 中国海洋大学, 青岛 266400

Email: tianjingyi55@gmail.com

2826586780@qq.com

摘要 本文聚焦于 NTIRE 2022 高动态范围挑战赛, 探讨了从低动态范围 (LDR) 图像恢复高动态范围 (HDR) 图像的问题, 这些 LDR 图像可能受噪声、量化误差和曝光问题影响。参考论文提出了一种新方法, 利用深度神经网络结合注意力机制从不同曝光的 LDR 图像中提取最可见信息, 并通过空间对齐模块生成 HDR 图像。该研究强调了特征提取的重要性, 但存在噪点和图像错位的问题。我们的工作受此启发, 分为两部分: 一是创新特征提取, 二是改进和完善模型。何恺明教授的团队提出的扩散模型作为特征提取器, 能补充噪点问题, 并可能生成更好的 HDR 图像。针对噪点和错位, 我们借鉴了 AHDRnet 网络中的 DRDB 模块, 通过更深的网络和更大的感受野来抑制这些问题。

关键词 多张 LDR 恢复 HDR, 深度学, GAN 网络, 扩散模型, 扩张残差密集块

1 介绍

图像处理是一项关键的计算机视觉任务, 旨在恢复降级的图像内容、填充缺失的信息或实现所需目标所需的转换或其他操作 (关于处理此类图像的应用程序的感知质量、内容或性能)。近年来, 视觉和图形界对这些基础研究主题的兴趣日益浓厚。不仅相关论文不断增加, 而且取得了实质性进展。本文的研究问题是基于 NTIRE 2022 高动态范围挑战赛的研讨问题, 即从一个或多个输入低动态范围 (LDR)

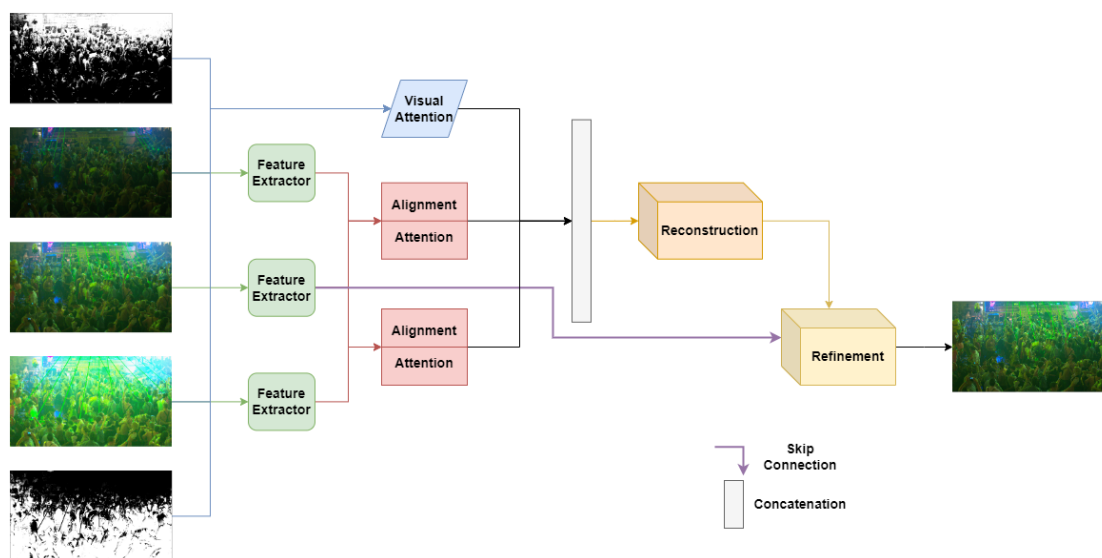
图像中恢复 HDR 图像的任务, 这些图像受到噪声、量化误差的影响, 并且由于传感器的限制, 可能会受到曝光过度和曝光不足的影响。

论文《High Dynamic Range Imaging via Visual Attention Modules》提出了一种新颖的想法, 即用深度神经网络框架提取短曝光、中曝光和长曝光三组 LDR 图像的最可见信息, 然后再利用注意力机制对遮罩图和神经网络提取到的特征进行逐元

素乘法，即去除过亮或过暗的区域；然后使用空间对齐模块将不同组别的特征图进行组合重构，最后得到效果较好的 HDR 图，实现了从多个 LDR 图像恢复 HDR 图像的任务。该文章相比起其他关注将不同曝光度的图像相结合的方法，重点和创新点在于图像有用信息特征提取。但同时文章模型存在两个缺点：一是过程中产生的噪点会在输出时依然存在；二是多组图片的组合可能会产生图像错位，即输出图像存在“鬼影”。

该论文给予了我们启发。我们的工作主要分为两部分：一是特征提取部分的创新工作，二是对该论文模型的改进和缺点完善。首先特征提取部分，何恺明教授及其团队的新作《Deconstructing Denoising

Diffusion Models for Self-Supervised Learning》给了我们灵感：扩散模型也可以作为优秀的特征提取器使用，并且由于其特殊的性质（通过噪点学习图像信息）可以弥补该文章的噪点问题；同时作为生成模型，能否从一系列不同曝光度的 LDR 学习特征来生成较好的 HDR 图也是我们一直比较好奇的问题。第二是针对参考文献噪点和错位问题，我们通过另一篇论文《Attention-guided Network for Ghost-free High Dynamic Range Imaging》了解到了 DRDB 模块，即通过更深的网络和更大的感受野来对特征信息进行更深层的挖掘，该方法已被证明能够较好地抑制噪点和“鬼影”的产生，我们基于此对参考文献的模型进行修改。



图一 论文模型结构图

2 相关工作

高动态范围成像技术分为单曝光 LDR 成像和多曝光 LDR 成像。

如果使用单曝光 LDR 进行成像，可以直接避免图像的对齐和鬼影问题。但是只使用单曝光的图像难以获取过暗或过亮区域的细节信息，近年来很

多研究者都在努力解决这个问题，Eilertsen 等人[8]提出的深度自动编码器网络 HDRCNN 方法解决了预测饱和图像区域中丢失的信息的问题。Endo 等人[9]使用基于深度学习的方法从一张 LDR 图像合成了多个 LDR 图像，然后通过合并它们来重建 HDR 图像。但这些方法都不具有我们所实现的模型的灵活性。

我们所研究的课题为目前较为热门的多曝光 LDR 进行成像并结合深度学习，自动学习图像的特征和融合规则，从而提高融合效果。Wang 等人[4]提出了一种利用快速局部拉普拉斯滤波（FLLF）来仅增强最亮和最暗区域细节的局部细节增强机制，但无法解决鬼影问题。进一步，Ali 等人提出的模型能够整合由视觉注意力模块提取的每张图像中最可见区域的信息，却不能处理好错位和去噪问题。他们都注重信息的提取，当图像发生错位时，结果往往不太理想。[10]提出借助光流法对齐图像，然后使用深度神经网络将它们组合在一起。但光流估计误差也会产生鬼影。[3]所提出的非基于流量的 AHDRNet 方法能够有效的解决鬼影问题，其运用了扩张残余密集块（DRDB）来充分利用层次特征并增加感受野，以产生幻觉缺失的细节。目前这些方法在达到好的效果时，会需要大量的计算资源。

何恺明教授及其团队在最新的论文中提到扩散模型也具有较好的特征提取效果，因此我们希望能够尝试结合扩散模型作为特征提取器，使其自

动根据提取到的特征生成 HDR 图像，同时我们也构建了 GAN 网络模型进行另一种方式地生成图像尝试并取得了初步的成果。最后我们对参考文章的模型进行改进，利用 DRDB 获取更多的特征信息来重建 HDR 图像，抑制鬼影和噪声问题。

3 方法的具体细节

3.1 生成模型构建

3.1.1 GAN 模型的构建

GAN(Generative Adversarial Network)即生成对抗网络，其实质可理解为两个互补网络的组合。具体而言，生成器（Generator）负责模拟数据的创建，而鉴别器（Discriminator）则负责鉴别输入数据的真伪。生成器的目标是使其生成的数据足够逼真，以至于鉴别器难以辨别；而鉴别器则致力于提升自身的鉴别能力，以求更为精准地判断数据的真实性。以上二者均是基于神经网络的架构。

该模型的核心在于通过对网络内部神经元间权重参数的调整来实现优化。在这个过程中，生成器和鉴别器利用各自的损失函数，结合误差反向传播（Backpropagation）算法以及随机梯度下降法、牛顿形法等优化方法，不断地进行参数的调整和优化，进而提升整个 GAN 的性能。

生成网络的损失函数：

$$L_G = H(1, D(G(z))) \quad (1)$$

上式中，G 代表生成网络，D 代表判别网络，H 代表交叉熵，z 是输入随

机数据。D(G(z)) 是对生成数据的判断概率，1 代表数据绝对真实，0 代表数据绝对虚假。H(1, D(G(z))) 代表判断结果与 1 的距离。

判别网络的损失函数：

$$L_D = H(1, D(x)) + H(0, D(G(z))) \quad (2)$$

上式中，x 是真实数据，这里要注意的是，H(1, D(x)) 代表真实数据与 1 的距离，H(0, D(G(z))) 代表生成数据与 0 的距离。根据公式推断，识别网络如果要想取得良好的效果，就要满足真实数据与 1 的距离小，生成数据与 0 的距离小。

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k , is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{\text{data}}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[\log D(x^{(i)}) + \log (1 - D(G(z^{(i)}))) \right].$$

end for

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z^{(i)}))).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

图二 GAN 网络训练过程伪代码

3.1.2 扩散模型的构建

Diffusion Models (扩散模型) 是通过逐步向训练数据中引入高斯噪声，进而学习数据恢复的过程，训练完成后，这些模型可将随机噪声样本输入，通过学习去噪过程来生成新数据，本质上，扩散模型属于隐变量模型范畴，它借助马尔可夫链 (Markov Chain, MC) 将数据映射至隐空间。在每一个时间步 t 中逐渐将噪声添加到数据 x_i 中以获得后验概率

$q(x_{1:T}|x_0)$ ，其中 x_1, \dots, x_T 代表输入的数据同时也是隐空间。也就是说 Diffusion Models 的隐空间与输入数据具有相同维度。Diffusion Models 分为正向的扩散过程和反向的逆扩散过程，以下分别介绍：

扩散过程：

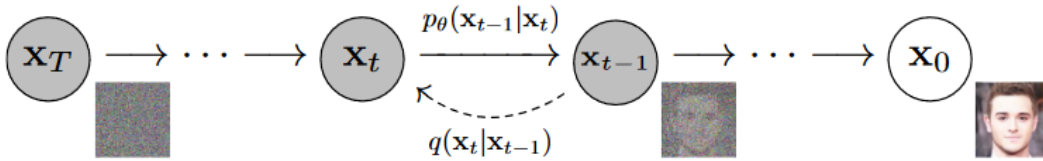
$$\begin{aligned} q(x_{1:T}|x_0) &:= \prod_{t=1}^T q(x_t|x_{t-1}) \\ &:= \prod_{t=1}^T N(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \end{aligned} \quad (4)$$

上式中, β_1, \dots, β_T 是高斯分布方差的超参数, 随着 t 的增大, x_t 越来越接近纯噪声。当 T 足够大的时候, x_T 可以收敛为标准高斯噪声 $N(0, I)$ 。

逆扩散过程:

$$\begin{aligned} p_{\theta}(x_{T:0}) &:= p(x_T) \prod_{t=1}^T p_{\theta}(x_{t-1}|x_t) \\ &:= p(x_T) \prod_{t=1}^T N(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)) \end{aligned} \quad (5)$$

根据马尔可夫规则表示, 逆扩散过程当前时间步 t 只取决于上一个时间步 $t-1$, 所以有:



图三 DM 的扩散和逆扩散过程

3.2 扩大感受野构建更深层网络

首先分析一下噪点产生的可能原因: 我们在进行扩散模型构建训练的过程中发现部分图像在初始时间步 ($t=0$) 的时候已经产生了噪点, 该类产生噪点的图像共同特征就是, 基本都为高曝光图像。所以我们首先想到将输入图像进行预处理, 对图像的曝光度进行调整, 但这样的方法是行不通的, 这样不仅对“鬼影”问题毫无用处, 对于原问题的研究也毫无意义 (参考论文模型本来就是从一组不同曝光的 LDR 图像恢复成 HDR 的, 再预处理降低曝光度是无意义的)。于是在查找资料阅读文献的过程中, 我们发现了一种端到端的深度神经网络 AHDRNet, 它所使用的 DRDB 模块能够在过程中较好地抑制噪点和“鬼影”的产生。实验证明, 该方法在多个数

$$p_{\theta}(x_{t-1}|x_t) := N(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t)) \quad (6)$$

通过以上公式, 我们已对 Diffusion Models 的扩散与逆扩散过程有了初步理解。然而, 在马尔科夫具体实现的过程中, 求解步骤往往复杂且关键。研究员们通常采用蒙特卡洛方法来进行采样, 随后对得到的结果的有效性进行评估。这一步骤对于确保模型的准确性和效率至关重要。

据集上均能取得最佳的数量和质量结果, 且无需基于光流的方法, 避免了由光流估计误差产生的图像伪影。

DRDB 是指 dilated residual dense block, 扩张残差密集块, 它是一种卷积神经网络结构, 用于图像处理任务中。DRDB 是基于 ResNet 中的 residual block 思想, 并在其基础上加入了 dilated convolution 来实现更深的网络和更大的感受野。

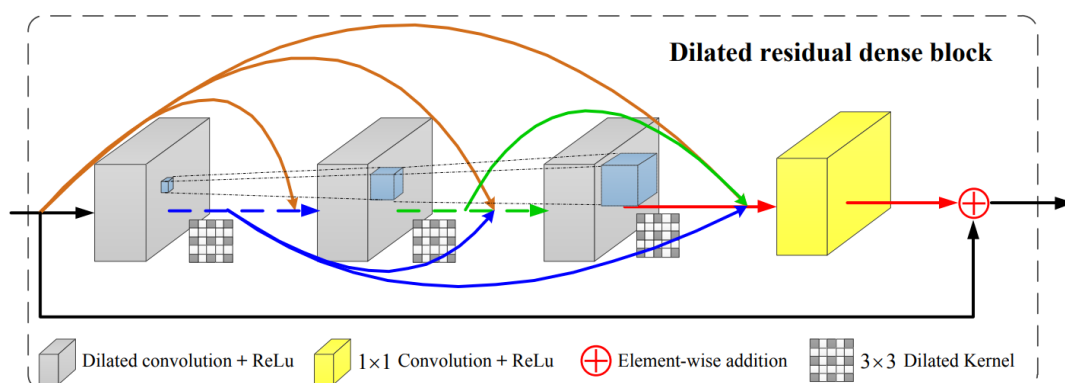
具体来说, DRDB 的主要特点包括:

- 它由多个卷积层组成, 每个卷积层后接 ReLU 激活函数。
- 在卷积层之间使用稠密连接 (dense connection), 即将前一层的所有特征图与当前层输入特征图进行拼接。

- 在卷积层中使用 dilated convolution, 通过在卷积核之间插入空洞来实现更大的感受野, 避免过拟合。
- 每个 DRDB 模块内部使用局部残差连接, 输入特征图与输出特征图进行拼接, 有助于训练。
- DRDB 模块可以堆叠, 形成 DRDB 网络, 实现更深的网络结构。

DRDB 在图像处理任务中表现优异, 可以提取更丰富的特征, 并有助于缓解过拟合问题。它在图像超分辨率、HDR 图像合成等任务中都有广泛应用。

本项目实现中, DRDB 用来充分利用层次化特征并扩大感受野, 以生成缺失的细节。



图四 DRDB 模块结构

4 结果

4.1 使用扩散模型的 HDR 成像及 GAN 网络模型成像

4.1.1 数据集

在这个模型的训练中, 我们使用的是一组一共有十三张不同曝光下的照片的数据集, 为了减少计算并且提高效率, 我们对输入图像进行了预处理, 使其形状为 128*128。

4.1.2 设备

我们在 jupyter 上搭载了所需环境, 但由于需要更大的内存, 我们只能选择使用 cpu 来运行我们的代码。

4.1.3 训练结果

我们构建及训练了一个扩散模型以及 GAN 网络模型, 扩散模型能够通过学习去

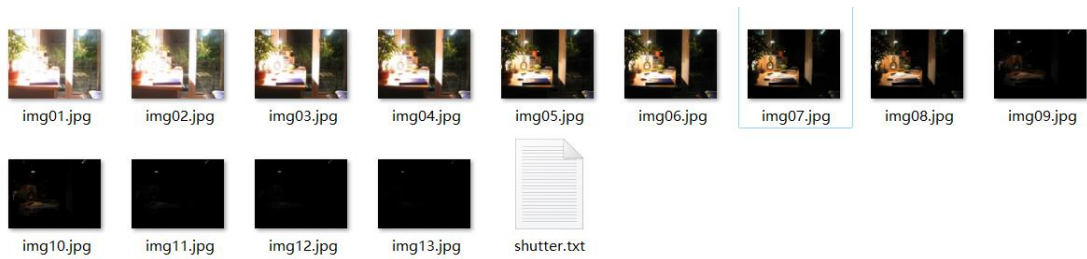
噪过程来生成新图像, GAN 网络模型则是通过对抗地学习生成新图像。

对于扩散模型, 我们采用了随机噪声输入, 进行 500 步的增加高斯噪声过程, 如图六。然后让模型逆向学习去噪过程, 以生成新图像。我们一共训练了 10 个 Epoch。用训练好的模型我们得到了生成的图像, 如图七 (左)。在图七 (右) 是我们模型在从噪点图中学习生成新图像的过程图。可以看到效果不是很好, 因为生成的图像是模糊不清的, 我们据此推测是时间步的学习生成不够, 但由于我们设备的算力有限, 当我们想要提高模型学习去噪的工作量, 将加噪过程增加到 1000 步甚至更多时, 发现这需要大量的时间及大量的内存, 我们的设备上 cuda 内存太小, 而 cpu 训练时间太长, 做不到大量的训练, 后续我们尝试了转移到 colab 上使用 cuda

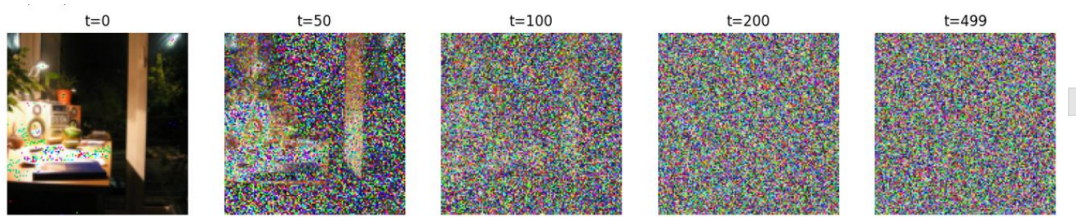
进行训练，发现仍需要很长时间，所以对于参数的适当性调整以及精确化的工作未能顺利完成。但由目前结果可知，提高调整参数等一系列方法，我们的训练结果已有一个大概形状，如图七，后续可能需要进行更多的训练可能能够生成更多的图像细节，达到更满意的效果。

对于 GAN 网络模型，其模型复杂度远小于扩散模型，算力要求较低，因此我们

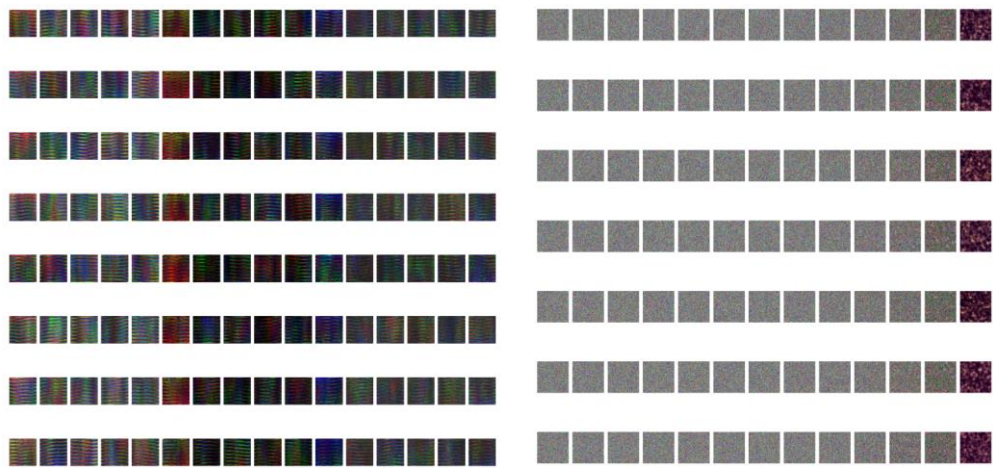
可以看到其生成了一个类似的结果图（图八）。以此验证或许通过生成模型来实现多张 LDR 重构 HDR 或许是可行的，至于效果不是那么清晰的原因，我们讨论归结于是数据集过小。事实上在神经网络的训练中，较大的数据集往往能取得更好的效果，后续我们会更换数据集继续研究。



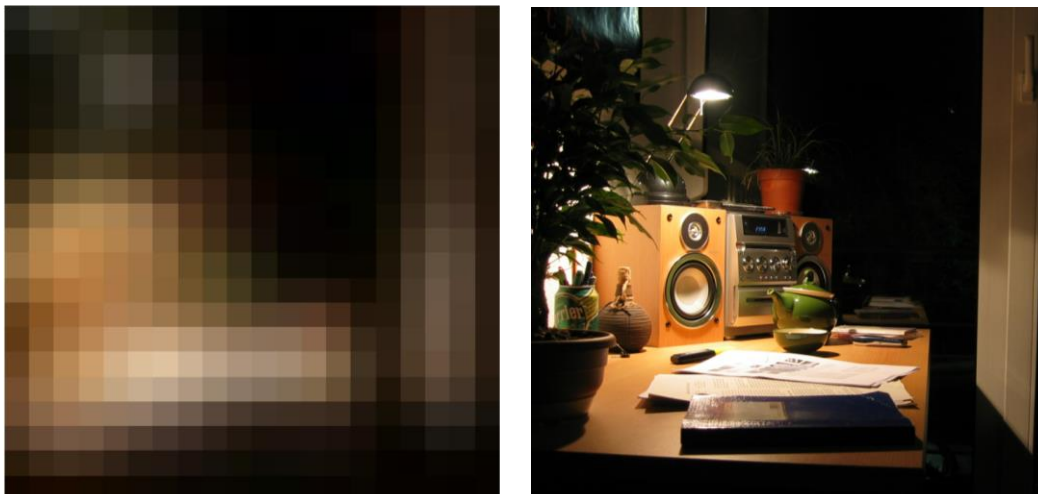
图五 扩散模型数据集



图六 扩散模型加噪过程



图七 扩散模型生成图像（左）和去噪过程（右）



图八 GAN 生成图像(左)和参考图像(右)

4.2 使用 DRDB 模块的 HDR 成像

4.2.1 数据集

我们向 NTIRE 2022 高动态范围挑战赛申请使用其提供的数据集进行训练, 在这个数据集中, 包含了约 1500 对用于训练的 LDR 图像, 40 对验证集以及 200 对分辨率为 1900×1060 的测试集图像。其中每对图像中都包含了长, 中, 短三种曝光时间的图像, 且图像都在动态场景中拍摄。我们选择了测试集中的 60 对图像用以进行我们模型的训练。

4.2.2 设备

为了能够训练这个模型, 我们在本地搭建了虚拟环境, 并使用 PyCharm 运行代码。

4.2.3 训练结果

由于参考论文提到无法解决在过程中生成噪声和鬼影问题, 我们在此基础上引入的 DRDB 模块, 扩展了特征提取的网络层数以期提取到更多的特征信息。为了验证我们的模型是否能够达到良好的去噪效果, 我们人为地给输入图像增加了噪声, 如图九(左)。将图像传入改进后的网络进行训练, 最后成功去除了噪声如图九(右)。

经过验证, 通过 DRDB 模块提供的更深的网络层数扩大感受野来对特征信息进行更深层的挖掘的操作, 能够有效的解决图像的鬼影问题。



图九 加噪输入图像(左)和生成图像(右)

5 总结和讨论

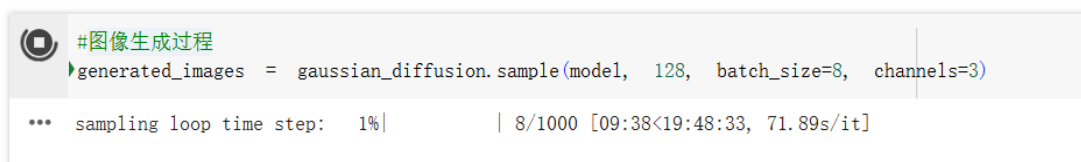
在本次实验过程中，我们聚焦于多张低动态范围图像重构高动态范围图像的课题研究，首先是进行了利用生成模型进行相关工作的尝试，其次改进了参考论文网络的噪声和“鬼影”问题。此过程中，我们在环境配置上遇到了很大的挑战，并且由于要对生成模型进行构建训练，算力也成了让我们头疼的问题。由于所选论文对环境有所要求，我们首先想到了在本地配置虚拟环境。但由于我们的设备内存不够充足，电脑配置的 cuda 使用不便，给我们的代码运行带来了极大的阻碍。然后我们转战到 jupyter，发现这也要求我们内存足够。所以后来我们尝试使用 colab，却发现相同参数下，colab 甚至需要更长的训练时间。这也许对我们今后实验中环境的选择起到一个启示性作用。

同时，我们大胆尝试了利用生成模型自主学习特征来重构特征的想法，实现一个端到端过程，尽管效果没有那么尽如人意，但确有初步的成果显现，我们认为若是能够有更好的算力对模型参数进行精调的话，或许能够达到我们理想当中的效果。在这个过程中，我们对扩散模型以及 GAN 网络模型有了一个深入的了解。但是模型训练的过程需要投入大量的时间与精力，往往一次生成图像就需要二十个小时，如

图十，导致我们模型的调整十分困难。寻找好的算力资源也是研究过程中的一个关键条件。

为了解决鬼影问题，在初期我们尝试通过提取特征值，获取多对关键点进行匹配，让图像能够对齐。但这样的思路是基于神经网络的端到端过程，需要网络自选点，这样选点的准确性难以保证，导致图像无法实现对齐。图十一是我们尝试使用两张图片进行关键点匹配以进行图像对齐的结果，可以看到，由于不同曝光下场景内显示的信息不完全相同，导致获取的关键点不能正确匹配上，所以我们放弃了这个方法。

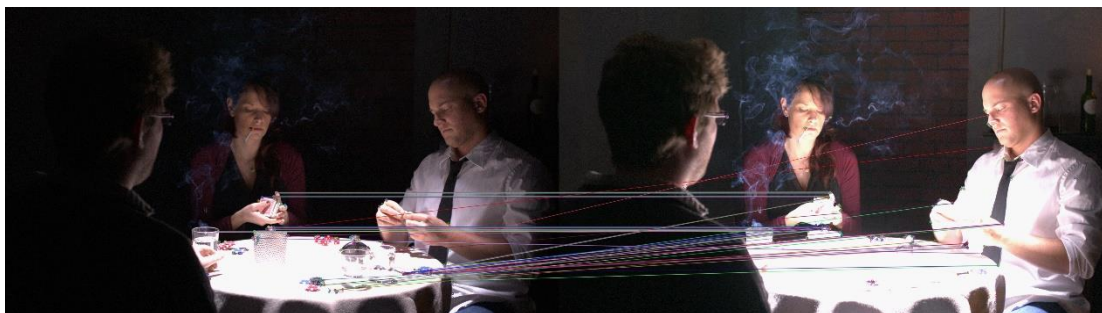
前面的尝试并没有达到我们预期的效果，所以我们继续查阅各种论文，学习了各种相关的方法，终于发现了 AHDRnet 网络，其中的扩张残差密集块（DRDB），此模块已被证实能有效抑制图像重构过程中噪声和鬼影的生成问题，因此我们将其添加进参考论文的网络中进行一个改进。由于原网络也是基于端到端的深度学习框架，因此进行中间步骤的修改替换往往会牵扯到其他部分的实现，但因为 DRDB 实现的是对特征提取层的扩大所以将原网络增加到 6 通道的位置进行修改即可。在此过程中我们学习到了如何对神经网络进行修改，以及对它的工作机制都有了更深刻的理解。

A terminal window with a dark background. The title bar is "#图像生成过程". The first line of code is "generated_images = gaussian_diffusion.sample(model, 128, batch_size=8, channels=3)". The second line shows the progress of a sampling loop: "... sampling loop time step: 1% | 8/1000 [09:38<19:48:33, 71.89s/it]".

```
#图像生成过程
>generated_images = gaussian_diffusion.sample(model, 128, batch_size=8, channels=3)

... sampling loop time step: 1% | 8/1000 [09:38<19:48:33, 71.89s/it]
```

图十 生成图像耗时



图十一 获取关键点进行匹配

6 个人贡献声明

田晶怡（60%）：生成模型的创意提出，模型构建，代码编写，论文绪论及方法部分的撰写，结果及总结部分的补充。

张玉梅（40%）：改进 DRDB 模块的想法提出，模型调整，训练运行，论文相关工作及结果部分的撰写，总结部分的补充。

参考文献

- 1 Xinlei Chen, Zhuang Liu, Saining Xie, Kaiming He et al. Deconstructing Denoising Diffusion Models for Self-Supervised Learning. arxiv. 2024. 14404.
- 2 Ali Reza Omrani and Davide Moroni. High Dynamic Range Imaging via Visual Attention Modules. Arxiv. 2023. 14705.
- 3 Qingsen Yan, Dong Gong, Qinfeng Shi. Attention-guided Network for Ghost-free High Dynamic Range Imaging. arxiv. 2019.
- 4 Wang C M, He C, Xu M F. Fast exposure fusion of detail enhancement for brightest and darkest regions[J]. The Visual Computer, 2021, 37(5): 1233-1243.
- 5 Tang L, Lu R S, Shi Y Q, et al. High dynamic range imaging method based on YCbCr spatial fusion[J]. Laser & Optoelectronics Progress, 2022, 59(14): 1415029

- 6 Xu H, Ma J Y, Zhang X P. MEF-GAN: multi-exposure image fusion via generative adversarial networks[J]. IEEE Transactions on Image Processing, 2020, 29: 7203-7216.
- 7 Hu Y T, Zhen R W, Sheikh H. CNN-based dehazing in high dynamic range imaging[C]//2019 IEEE International Conference on Image Processing (ICIP), September 22-25, 2019, Taipei, China. New York: IEEE Press, 2019: 4360-4364.
- 8 Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K. Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. ACM TOG, 2017. 1, 2, 3, 5, 6, 7, 8
- 9 Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deepreversed tone mapping. ACM Transactions on Graphics, 36(6):1-10, 2017.
- 10 N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Trans. Graph., vol. 36, jul 2017.