

基于 GAN 和 U-Net 的低光照图像增强算法^①



李晨曦¹, 李 健²

¹(中国科学院 计算机网络信息中心, 北京 100190)

²(中国科学院大学, 北京 100049)

通信作者: 李晨曦, E-mail: lichenxi@cnic.cn

摘 要: 夜间、低光照等条件下产生的图像数据, 存在画面过暗、细节丢失的问题, 对理解图像内容、提取图像特征造成阻碍. 研究针对此类图像的增强方法, 恢复图像的亮度、对比度和细节, 在数字摄影、上游计算机视觉任务中有着重要的应用价值. 本文提出一种基于 U-Net 的生成对抗网络, 生成器采用带有混合注意力机制的 U-Net 模型, 其中混合注意力模块将非对称的 non-local 的全局信息和通道注意力的通道权重信息相结合, 提高网络的特征表示能力. 判别器采用基于 PatchGAN 的全卷积网络模型, 对图像不同区域进行局部处理. 本文引入多损失加权融合的方法, 从多个角度引导网络学习低光照图像到正常光照图像的映射. 通过实验证明, 该方法在峰值信噪比、结构相似性等客观指标上取得较好的成绩, 同时合理的恢复了图像的亮度、对比度和细节, 直观上改善了图像的感知质量.

关键词: 低光照图像增强; 生成对抗网络; U-Net; 混合注意力; non-local; 深度学习

引用格式: 李晨曦, 李健. 基于 GAN 和 U-Net 的低光照图像增强算法. 计算机系统应用, 2022, 31(5): 174-183. <http://www.c-s-a.org.cn/1003-3254/8431.html>

Low-light Image Enhancement Algorithm Based on GAN and U-Net

LI Chen-Xi¹, LI Jian²

¹(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The image data generated at night, under low light conditions, etc., has the problems of too dark images and loss of details, which hinders the understanding of image content and the extraction of image features. The research of enhancing this type of images to restore the brightness, contrast and details is meaningful in the applications of digital photography and upstream computer vision tasks. This study proposes a U-Net-based generative adversarial network. The generator is a U-Net model with a hybrid attention mechanism. The hybrid attention module can combine the asymmetric non-local global information and the channel weight information of channel attention to improve the feature representation ability of the network. A fully convolutional network model based on PatchGAN is taken as the discriminator to perform local processing on different regions of the images. We introduce a multi-loss weighted fusion method to guide the network to learn the mapping from low-light images to normal-light images from multiple angles. Experiments show that this method achieves better results regarding objective indicators such as peak signal-to-noise ratio and structural similarity and reasonably restores the brightness, contrast and details of the images to intuitively improve their perceived quality.

Key words: low-light image enhancement; generative adversarial network (GAN); U-Net; mixed attention; non-local; deep learning

① 基金项目: 国家重点研发计划 (2019YFC0507405)

收稿时间: 2021-07-13; 修改时间: 2021-08-04; 采用时间: 2021-08-12; csa 在线出版时间: 2022-04-11

目前,基于深度学习的计算机视觉算法通过强大的图像特征学习能力,在图像识别、目标检测等领域取得巨大成功.深度卷积神经网络能够很好的提取图像中的关键信息,理解图像语义,但这依赖于高质量的图像数据.光线不足的环境下产生的图像数据,往往存在亮度和对比度低、细节丢失、噪声多等现象^[1],这类低光照图像不但影响人类的主观感受,而且对上游的视觉算法任务造成阻碍.所以,不依赖于昂贵的图像采集设备的图像增强算法^[2]具有非常重要的研究意义和应用价值.本文着手于提高低光照图像的亮度、对比度,改善图像感知质量,展开相关算法研究,共分为6个章节:第1节对相关工作展开调研和分析.第2节介绍生成对抗网络算法基本原理.第3节提出了一种基于生成对抗网络的增强算法,设计了带有混合注意力模块的U-Net作为生成器,基于PatchGAN的全卷积网络作为判别器.第4节通过大量实验来证明所提出模型的有效性.第5节通过消融实验证明本文所提出的模型组件和加权融合损失函数对提高图像质量有积极的影响.第6节总结本文的工作,展望未来的工作方向.

1 相关工作

低光照图像增强方法分为传统图像处理方法和基于深度学习的图像增强方法.传统方法中直方图均衡化^[3]方法通过均匀拉伸灰度直方图,扩展图像灰度级的动态范围,来增加对比度.限制对比度自适应直方图均衡化^[4]算法把图像划分为多个不重叠区域,并对区域内灰度直方图最大值进行限定,超出最大值的像素均匀分配到其余灰度级,之后再直方图均衡化,能够有效抑制区域内噪声放大和局部区域过度增强,但仍不能有效处理噪声,恢复图像细节.Retinex理论^[5]模拟人类视网膜成像原理,将图像分解为入射分量和反射分量,入射分量决定了像素的动态范围,反射分量则反映了图像中物体的本质内容.通过估计两种分量,去除入射分量,只保留图像中物体的反射属性,得到不受光照影响的图像内容,从而实现图像增强.根据这一理论出现了单尺度Retinex(SSR)^[5],多尺度Retinex(MSR)^[6]等方法.SSR用高斯函数与图像进行卷积,近似估计入射分量,从而求得反射分量.MSR使用不同尺度的高斯核函数,可以看作多个不同尺度的单尺度Retinex线性加权求和.带彩色恢复的多尺度Retinex

算法(MSRCR)^[6]引入颜色恢复因子,弥补图像局部区域对比度增强而导致的图像颜色失真的缺陷.此外,NPE^[7]利用亮度滤波分解图像将反射分量限制在 $[0, 1]$,并结合log双边转换平衡光照分量对细节和自然度的增强,避免过度增强.LIME^[8]首先估计亮度图,再根据Retinex公式反推出反射分量,同时使用BM3D算法进行去噪.BIMEF^[9]引入相机响应模型加强图像曝光度,并通过光照估计加权融合得到增强图像.

近年来,深度学习凭借强大的图像理解能力在一些低级的图像视觉任务,如超分辨率,去噪去雾等中获得成功,LLNet^[10]将深度学习引入低光照图像增强任务,构建了堆叠稀疏去噪自编码器,并人工合成低光照数据,模拟低光照环境,对低光照有噪声图像进行增强和去噪.相比传统算法增强图像质量明显提升,但模型结构简单,没有完全利用深度学习的优势,仍有巨大的进步空间.MBLEN^[11]提出多分支网络,分别对应图像增强中亮度增强、对比度增强、去噪去伪影等多种功能需求,并对不同层次的特征进行融合,达到多方面提高图像质量的效果.ALIE^[12]提出了一种注意力机制引导的增强方法和多分支网络结合的体系结构,通过生成注意力和噪声图来引导区域自适应性的弱光增强和去噪.Wang等^[13]将Retinex理论和神经网络相结合,通过卷积神经网络估计光照分量,调整曝光程度,得到期望的正常曝光图像,并加入平滑损失提高对比度、三通道颜色损失提高鲜艳程度.EnlightenGAN^[14]首创地在低光领域使用GAN技术,即使不成对的数据也能进行训练学习,并利用局部和全局判别器处理局部和全局的光照条件.基于深度学习的方法通过改变网络结构、学习不同类型特征以及优化损失函数等多元化的手段可以显著提高增强效果,但对于图像质量和细节的恢复还有很大提升空间.

2 生成对抗网络原理

本文针对传统增强方法的不足和现有基于深度学习算法的特点,提出使用生成对抗网络作为模型框架进行低光照图像增强的方法.本节介绍生成对抗网络的基本原理,及目前主流的PatchGAN思想.

2.1 生成器和判别器

生成对抗网络由生成器 G 和判别器 D ^[2]组成,生成器学习真实数据的潜在分布,产生生成数据.判别器本质是一个二分类器,判断输入数据是生成数据还是

真实数据. 两个模型训练过程中相互博弈, 使得生成器的生成数据不断接近真实数据, 判别器无法判断其真假, 最终达到生成器和判别器的动态平衡. GAN 的优化属于极大极小博弈问题^[2], 其目标函数公式如下.

$$\min_G \max_D (D, G) = E_{x \sim P_x(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

其中, $E(\cdot)$ 表示分布函数的期望值, $P_x(x)$ 表示真实样本的分布, $P_z(z)$ 表示随机噪声的分布, $G(z)$ 表示随机噪声 z 输入生成器 G 产生的生成数据, $D(x)$ 和 $D(G(z))$ 分别表示判别器对真实数据和生成数据的判别结果. GAN 的训练流程如算法 1 所示.

算法 1. GAN 算法的训练流程

```

for number of training epochs do
  for k steps do
    抽样 m 个来自噪声分布的样本 {z1, z2, ..., zm};
    抽样 m 个来自真实分布的样本 {x1, x2, ..., xm};
    噪声样本经过生成器和判别器得到输出 D(G(zi));
    真实样本经过判别器得到输出 D(yi);
    计算判别器损失函数 Ld:
    Ld = 1/m * sum_{i=1}^m [-log D(yi) - log (1 - D(G(zi)))]
    通过 Adam 梯度下降算法优化判别器的参数:
    theta_d = Adam(∇theta_d(Ld), theta_d)
    计算生成器损失函数 Lg:
    Lg = 1/m * sum_{i=1}^m [log (1 - D(G(zi)))]
    通过 Adam 梯度下降算法优化判别器的参数:
    theta_g = Adam(∇theta_g(Lg), theta_g)
  end for
end for
    
```

2.2 PatchGAN 方法

普通二分类判别器采用基于 CNN 的分类模型, 将输入数据映射为二维向量, 表示该数据为真实数据和生成数据的概率, 研究表明该方法在低光照图像增强领域并不适用, 因为图像增强不仅是一个整体的二分类问题, 而且还需要对图像中不同区域进行调整. PatchGAN 是一种关注局部图像特征的判别器框架^[15], 思想是使用全卷积网络提取高级图像特征, 输出为一个 $N \times N$ 的矩阵, 其中每一个元素表示一个感受野, 能够表示出原图像中对应区域的图像特征.

3 主要方法

本文提出的模型基于 GAN 架构, 生成器采用带有混合注意力机制的 U-Net, 输入为低光照图像, 输出为同尺寸的增强图像. 判别器借鉴 PatchGAN 的思

想, 采用全卷积网络, 输出为矩阵张量. 具体结构如图 1 所示.

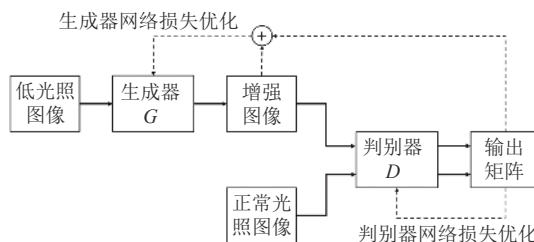


图 1 GAN 框架模型图

3.1 生成器网络结构

生成器是 GAN 的核心, 主要作用是进行图像增强、去噪和细节恢复. U-Net^[16] 在图像分割领域应用广泛, 是由编码器、解码器和跳跃连接^[17] 组成的卷积神经网络. 编码器使用 3 次最大池化层缩小特征图尺寸, 获得感受野更大的特征图. 解码器使用 3 次反卷积从高级语义特征恢复到高分辨率图像, 跳跃连接将编码器和解码器对应层次的特征图进行叠加, 避免了网络深度加深造成的浅层特征丢失, 同时聚合多层特征合成高质量图像. 本文提出的生成器网络在 U-Net 的基础上进行改进, 在网络头部提取注意力图, 在编码器部分加入混合注意力模块 MixAttBlock, 在解码器部分加入注意力卷积模块 ConvAttBlock, 提高网络的特征表示能力, 有助于恢复图像细节. 生成器网络具体结构如图 2 所示.

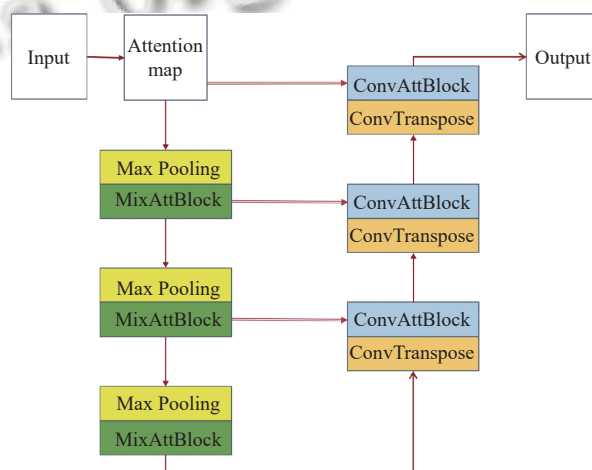


图 2 生成器网络结构图

注意力卷积模块: SENet^[18] 提出的通道注意力机制, 通过全局平均池化获得每个通道的代表值, 再使用

全连接层及激活函数学习通道间的关系, 获得通道的权值. 注意力机制能够引导网络关注更重要的特征, U-Net 的解码器阶段特征图通道数逐层减少, 直至恢复至三通道, 我们在此引入通道注意力, 形成注意力卷积模块 (ConvAttBlock), 在通道恢复时有助于保留重要特征, 其网络结构如图 3 所示.

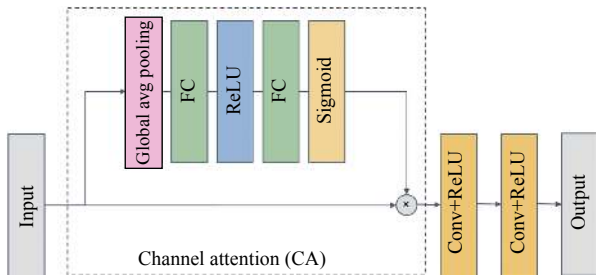
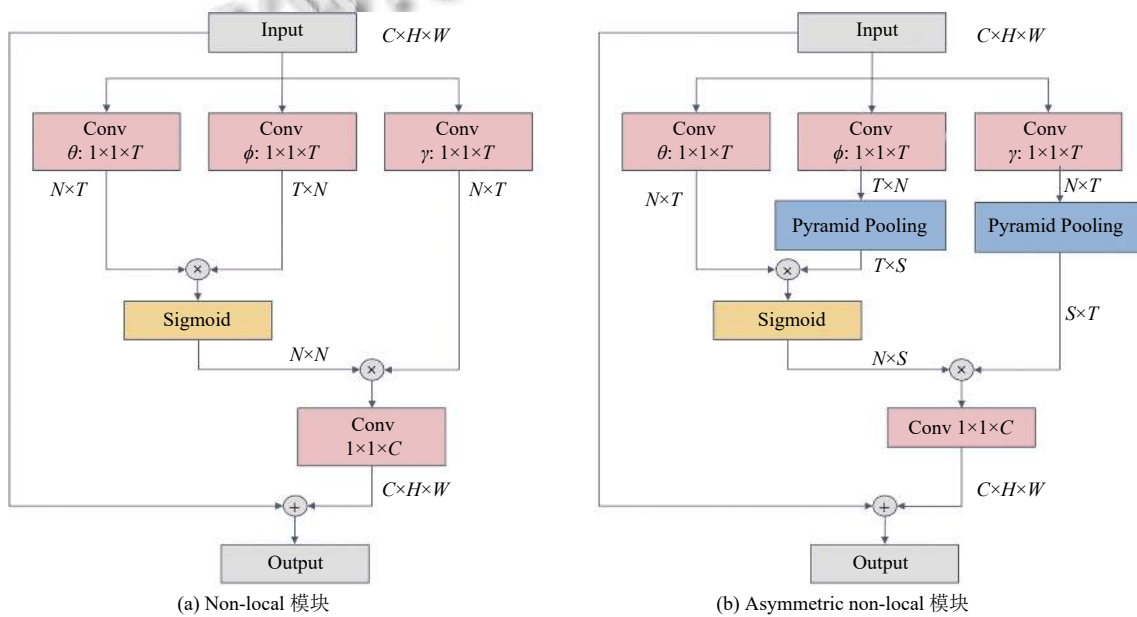


图 3 ConvAttBlock 结构图

非对称的 non-local 模块: 卷积层只在局部像素范围进行卷积运算, 因此需要堆叠大量卷积层才能获得较大感受野, 同时也增加了网络参数和网络优化难度. Non-local^[19] 旨在通过计算特征图任意像素位置之间的远程依赖的方式代替堆叠大量卷积层, 来提高感受野.

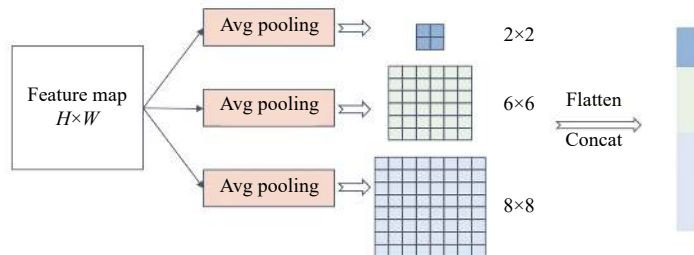
其通用公式为式 (2), 其中, x_i 表示特征图上任意位置, f 是相似度函数, g 是映射函数, 将点映射成向量, 即求任意位置的特征表示, $C(x)$ 表示归一化. Non-local 的具体网络结构如图 4(a) 所示, 其中, θ 和 ϕ 两个卷积用于压缩通道数, 矩阵乘法用于计算相似度, Softmax 进行归一化, γ 卷积即为映射函数.

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j) \quad (2)$$



(a) Non-local 模块

(b) Asymmetric non-local 模块



(c) 金字塔池化层

图 4 Non-local 和 ANN 模块结构图

Non-local 引导网络利用更大范围的信息这一特性, 使得低光照图像增强时能够保留更丰富的信息. 同

时, 该模块计算过程中会产生大型的矩阵张量, 导致内存占用量和计算量激增. 为进一步优化网络性能, 我

们采用了非对称的 non-local 模块 (asymmetric non-local, ANN)^[20], 在 φ 和 γ 卷积后面引入金字塔池化层, 结构如图 4(c) 所示. 多尺度的池化可以在减小特征图尺寸的同时, 不会丢失过多的特征信息. ANN 的具体结构如图 4(b) 所示, 其中, $N=H \times W$, 且 $S \ll N$.

表 1 展示了 non-local 与 ANN 模块的内存占用和乘加计算量的对比结果, 输入不同尺寸特征图时, ANN 均能大幅减小内存占用和计算量, 有效提高网络性能.

混合注意力模块: 我们将 ANN 模块和通道注意力结合, 提出混合注意力模块 (MixAttBlock). 具体结构如图 5 所示, 该模块可以利用全局位置信息和通道注意力信息^[21], 获得更好的特征表示.

表 1 Non-local 与 ANN 内存和乘加计算量对比

输入	模块	内存 (MB)	乘加计算量 (M)
(128, 128, 32)	Non-local	261.00	267.12
	ANN	11.51	70.91
(64, 64, 64)	Non-local	18.50	79.04
	ANN	4.15	67.73

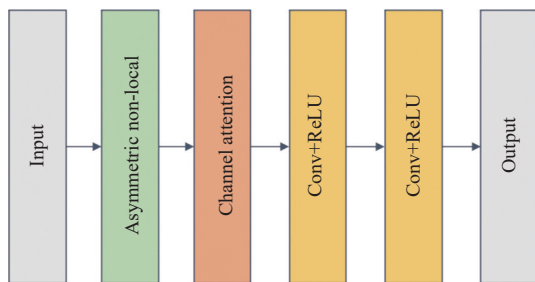


图 5 MixAttBlock 结构图

3.2 判别器网络结构

本文的判别器采用 13 层全卷积网络, 均使用 3×3 的卷积核, 除最后的卷积层外, 其余卷积层后均使用 LeakReLU 激活函数. 全卷积网络提取高层图像特征, 并输出一个 $1 \times 4 \times 4$ 的矩阵张量, 充分考虑图像不同区域的影响, 注重图像全局质量和局部细节的提升. 网络结构如图 6 所示.

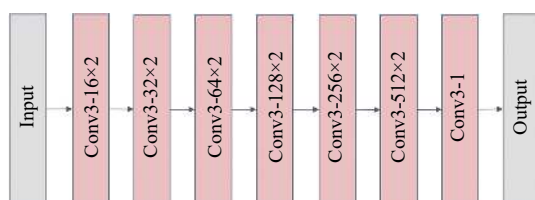


图 6 判别器网络结构图

3.3 损失函数

3.3.1 生成器损失函数

图像增强中常用损失函数为平均绝对误差 (mean absolute error, MAE) 损失^[11], 即增强图像和正常标签图像的像素绝对值损失, 记作 L_{pix} , 计算公式如下:

$$L_{pix} = \frac{1}{M} \sum_{i=1}^M MAE(y_i, G(x_i)) \quad (3)$$

其中,

$$MAE(A, B) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |A(i, j) - B(i, j)| \quad (4)$$

这类损失只能使生成图像与对应的标签图像像素值绝对误差减小, 没有考虑图像的内容相关性. 为了能够更好的恢复图像细节, 我们采用了多损失加权融合的方法, 加入对抗损失、结构损失和感知损失, 构造新的生成器损失函数, 公式如下:

$$loss = \lambda_1 L_{ad} + \lambda_2 L_{pix} + \lambda_3 L_{str} + \lambda_4 L_p \quad (5)$$

对抗损失: 该损失基于生成器和判别器的对抗机制^[17], 引导生成器网络学习正常光照图像的亮度、对比度、纹理等特征. 对抗损失定义为生成图像输入判别器得到结果矩阵, 与相同尺寸的全 1 矩阵张量的均方误差值 (mean square error, MSE). 当结果矩阵每个元素都接近 1, 则判别器不能判断其是否为生成图像, 从而达到生成器和判别器的动态平衡. 计算公式如式 (7), 其中, G 表示生成器网络输出, D 表示判别器网络输出, I_{one} 表示全 1 矩阵张量.

$$MSE(A, B) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (A(i, j) - B(i, j))^2 \quad (6)$$

$$L_{ad} = \frac{1}{M} \sum_{i=1}^M MSE(I_{one}, D(G(x_i))) \quad (7)$$

结构损失^[22]: 该损失旨在通过衡量生成图像和标签图像的结构差异性, 提高生成图像的整体视觉质量. 该损失根据常用的图像质量评估指标结构相似性 (structural similarity, SSIM)^[23] 和多尺度结构相似性 (multi-scale structural similarity, MS-SSIM)^[24] 构成. SSIM 从亮度、对比度、结构 3 方面度量图像相似性, 计算公式如式 (8) 所示. 亮度、对比度、结构度量函数表达式分别为式 (9)、式 (10)、式 (11). 其中, μ_x 和 μ_y 分

别表示图像 x 和 y 的均值, σ_x 和 σ_y 分别表示图像 x 和 y 的方差, σ_{xy} 表示图像 x 和 y 的协方差, C_1 、 C_2 和 C_3 是常数. 实际计算时, 使用高斯加权滑动窗口将图像划分为多个局部区域, 整张图像的 $SSIM$ 取所有区域的均值.

$$SSIM(x, y) = l(x, y) \times c(x, y) \times s(x, y) \quad (8)$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_2} \quad (9)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (10)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (11)$$

$MS-SSIM$ 考虑不同图像分辨率的影响, 对图像进行多尺度的缩小, 每个尺度具有不同权值, 仅在原尺度上计算亮度度量函数^[24], 即式 (9). 公式表达如式 (12). 其中, M 表示尺度个数, α_j 、 β_j 和 γ_j 表示第 j 尺度上的 3 个分量的权值.

$$MS-SSIM(x, y) = [l_M(x, y)]^{\alpha_j} \times \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (12)$$

结构损失定义为 L_{SSIM} 和 $L_{MS-SSIM}$ 的和, 计算公式如下:

$$L_{str} = L_{SSIM} + L_{MS-SSIM} \quad (13)$$

其中,

$$L_{SSIM} = \frac{1}{M} \sum_{i=1}^M 1 - SSIM(x_i, y_i) \quad (14)$$

$$L_{MS-SSIM} = \frac{1}{M} \sum_{i=1}^M 1 - MS-SSIM(x_i, y_i) \quad (15)$$

感知损失: L_{pix} 、 L_{str} 等指标只关注了图像中的底层信息, 而没有考虑高层特征信息. 我们认为越相似的图像, 通过通用特征提取器获得的特征图也越相似, 而高层特征图也是提高图像视觉质量的重要标准, 这在 SRGAN^[25] 中定义为感知损失. 由于 VGG19 网络在图像特征提取上有着良好的表现, 我们采用 ImageNet 上预训练的 VGG19 模型作为特征提取器^[1], 提取其第 2 和第 5 个池化层的输出构成特征图, 计算特征图的均方差作为感知损失, 通用公式如下,

$$L_p = \frac{1}{W_{i,j}H_{i,j}C_{i,j}} \times \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \sum_{z=1}^{C_{i,j}} (\varphi_{i,j}(E)_{x,y,z} - \varphi_{i,j}(T)_{x,y,z})^2 \quad (16)$$

其中, E 、 T 分别表示增强图像和正常光照图像, $W_{i,j}H_{i,j}C_{i,j}$ 表示 VGG19 网络中第 i 块第 j 层的特征图的尺寸, $\varphi_{i,j}(\cdot)$ 表示第 i 块第 j 层输出的特征图.

3.3.2 判别器损失函数

生成图像和正常光照图像输入判别器时, 生成图像应该判别为假, 而正常光照图像应该判别为真, 又由于判别器的输出为 $1 \times 4 \times 4$ 的矩阵张量, 则使用全 0 和全 1 矩阵张量分别与生成图像和正常光照图像的输出矩阵计算 MSE , 损失函数定义为:

$$L_{ad} = \frac{1}{M} \sum_{i=1}^M MSE(I_{one}, D(y_i)) + MSE(Z_{zero}, D(G(x_i))) \quad (17)$$

其中, Z_{zero} 表示全 0 矩阵, I_{one} 表示全 1 矩阵.

4 实验分析

4.1 实验环境

本文实验环境为 Ubuntu 18.04, Intel Xeon E5-2630 @ 2.20 GHz, 32 GB RAM, TITAN RTX 24 GB, PyTorch 深度学习框架.

4.2 数据集

真实环境下很难捕捉到成对的低光照和正常光照图像, 根据之前的研究, 本文采用 MBLLEN 提供的基于 PASCAL VOC 图像数据集的合成数据集^[11], 该数据集通过对每个通道随机伽马非线性调整产生低光照图像, 公式表达为:

$$I_{out} = A \times I_{in}^{\gamma}, \gamma > 1 \quad (18)$$

其中, A 表示图像最大像素值, γ 服从均匀分布 $U(2, 3.5)$. VocDark 包含训练集 16 925 张和测试集 144 张.

4.3 实验指标

图像增强通常采用峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似性 (SSIM) 作为定量指标^[26]. PSNR 通过评估生成增强图像与正常光照图像间像素差异来衡量整体增强效果, 单位是分贝 (dB), 公式如下:

$$PSNR = 10 \times \log_{10} \left(\frac{\max_I^2}{MSE} \right) \quad (19)$$

其中, \max_I 表示像素的最大值, MSE 即图像像素值的均

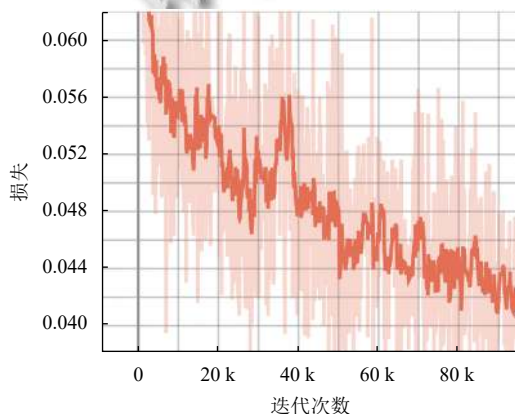
方误差,公式见式(6). *SSIM* 权衡图像间结构差异,用于表现图像的细节相似程度,公式见式(8). 实验取 VocDark 测试集的平均 *PSNR*、*SSIM* 进行比较,数值越高,表明模型增强效果越好.

4.4 参数细节

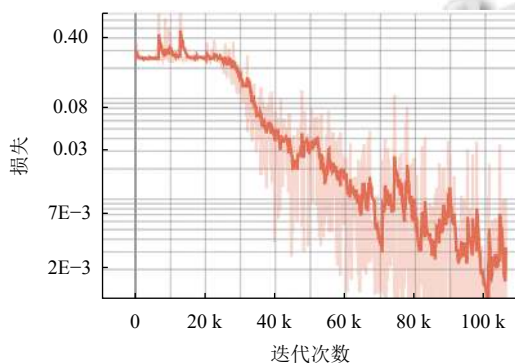
本文采用 Adam 优化器,初始学习率为 $1E-3$,并使用学习率衰减,每个 epoch 衰减 1%. 网络训练的 epoch 设置为 100, batch-size 设置为 16. 采用随机裁剪、旋转和翻转扩充数据,并指定输入尺寸为 256×256 . 金字塔池化层输出尺寸设置为 [2, 6, 8], 损失函数权值 $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ 设置为 (0.006, 0.85, 0.14, 0.004).

4.5 实验评估

实验网络模型训练损失随迭代的收敛曲线如图 7 所示,其中,图 7(a) 是生成器训练损失曲线,图 7(b) 是判别器训练损失曲线. 可以看出,模型在 100 epoch 后接近收敛.



(a) 生成器训练损失曲线



(b) 判别器训练损失曲线

图 7 训练损失曲线

本文提出的模型与一些经典算法在 VocDark 测试集上的对比结果如表 2 所示. 其中,加粗字体表示

最高值,下划线表示次高值. 实验通过客观指标对比证明本文算法要优于其他算法,说明基于 U-Net 的生成对抗网络在低光照图像增强任务中具有比较明显的优势.

测试集的直观增强效果如图 8 所示,MSRCR 算法的结果存在颜色失真,曝光过度的问题. NPE 算法一定程度上缓解了颜色失真的问题,但亮度提升不足,且局部区域模糊. LIME 算法的结果色彩不自然,亮度、对比度过高,如 Image 1 中沙发偏蓝,Image 2 中狗的毛色过亮. BIMEF 算法结果偏暗,如 Image 4 和 Image 5 整体亮度增强效果不明显. MBLLEN 算法的效果较好,但局部区域较暗,且与自然光照图像存在偏差,如 Image 4 中羊腿下的区域和 Image 5 中车底的区域较暗. 本文算法的结果与自然光照图像非常接近,并且本文算法在增强整体的亮度、对比度的同时,细节恢复上也更加出色,如 Image 1 中地面光线轮廓更真实,Image 2 中地面、Image 4 中羊嘴和 Image 5 中车牌的细节更加明显. 从人眼视觉感知效果上可以看出,本文算法的增强图像质量高于其他算法.

表 2 与经典算法的对比结果

算法	<i>PSNR</i>	<i>SSIM</i>
Input	12.29	0.44
HE ^[3]	17.01	0.68
MSRCR ^[6]	14.99	0.71
NPE ^[7]	16.84	0.70
LIME ^[8]	16.28	0.66
BIMEF ^[9]	19.80	0.87
MBLLEN ^[11]	<u>26.56</u>	<u>0.89</u>
Ours	29.21	0.90

5 消融实验

本文通过消融实验验证提出的网络模型中每个组件的有效性. 实验通过逐步添加各类组件来比较指标的变化. 在这些实验中,均保持训练过程中的超参数设置不变,将所有网络训练 100 epoch 达到收敛状态,选择 *PSNR*、*SSIM* 作为衡量不同模块对网络性能影响的指标.

首先选择 U-Net 作为骨干网络,分别加入 ConvAtt-Block、MixAttBlock 以及其组合,结果如表 3 所示,说明本文所提模块组合后能够取得最佳效果.

图 9 展示了添加不同模块的效果,通过视觉对比

可以发现, U-Net 作为骨干网络在低光照图像增强中能取得较好的效果, 但仍存在亮度、对比度提升不足, 边缘细节模糊的问题. 加入 ConvAttBlock 后增强效果不明显, 加入 MixAttBlock 后局部区域提升明显, 如 Image 1 中羊毛颜色和细节得到恢复, Image 5 中人脸细节更加明显. 本文所提方法产生的增强图像视觉效果最好, 说明这些模块有助于恢复低光图像各方面属性, 使其成

为清晰真实的自然光照图像.

其次, 本文对多种损失的组合进行对比, 测试集结果如表 4 所示, 其中, *号表示选中该损失参与加权融合, 组成新的生成器损失函数. 可以看出同时使用 L_{pix} 和 L_{str} 能够显著增强图像的 PSNR, 说明该损失能够正确引导网络学习低光照图像到自然光照图像的映射关系. 并且 L_p 的加入有助于提高增强图像质量.

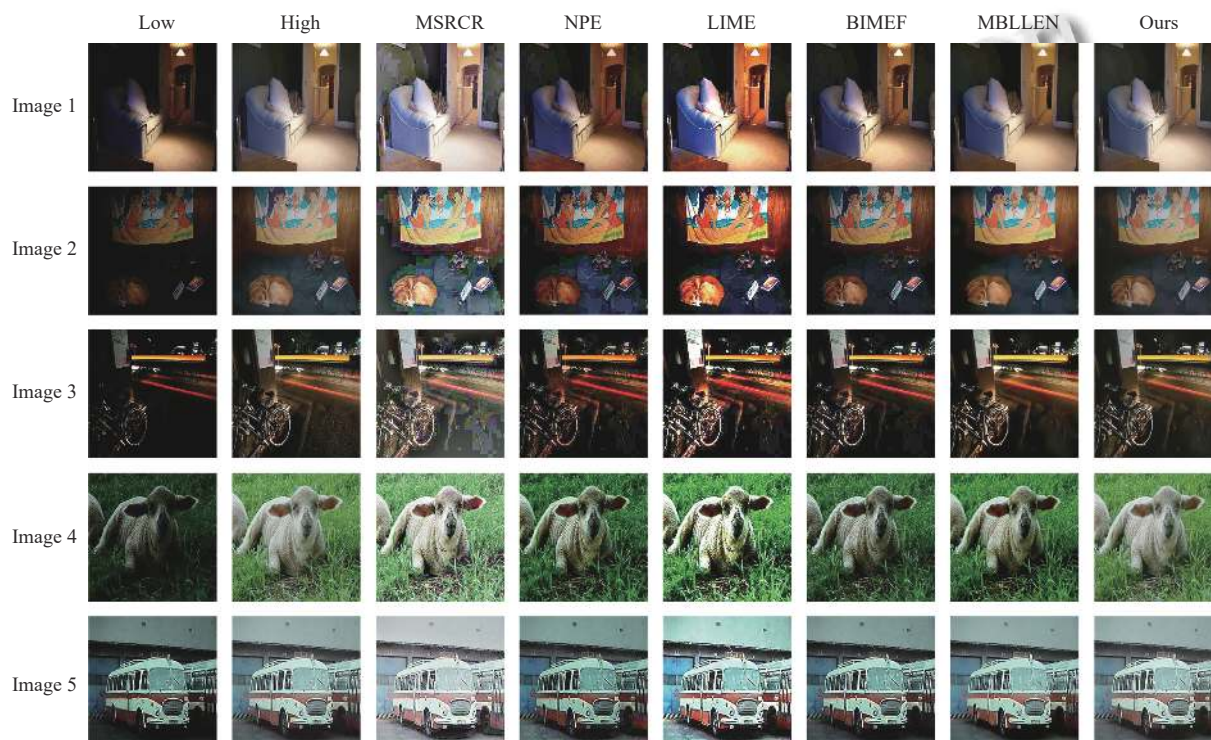


图 8 测试集效果对比

表 3 不同模块组合结果对比

模块组成	PSNR	SSIM
U-Net	26.73	0.88
ConvAttBlock	25.96	0.87
MixAttBlock	<u>28.13</u>	<u>0.89</u>
MixAttBlock+ConvAttBlock	29.22	0.90

6 总结

本文针对低光照图像增强的问题, 提出了一种基于 U-Net 生成对抗网络的低光照图像增强算法. 生成器采用带有混合注意力的 U-Net 网络, 该方法利用非对称的 non-local 模块减少网络复杂度的同时提高感受野, 与通道注意力结合, 获得更丰富的特征表示. 判别

器借鉴 PatchGAN 的思想, 从普通的二分类网络改为输出为矩阵的全卷积网络, 以考虑图像局部区域差异, 提高生成器的增强效果. 实验证明本文的方法能够获得高对比度、高亮度、噪声和颜色失真较小、细节更显著的增强图像. 并且, 通过在公开数据集上对 PSNR、SSIM 等评价指标的客观比较, 证明本文所提方法具有更好的效果. 最后, 通过消融实验证明本文提出的算法及参数配置能够取得最佳的效果. 在今后的工作中, 我们将从两个方向继续探索和应用低光照图像增强技术: 一是结合图像分类、目标检测等上游任务, 提高其在低光照领域的可行性和准确性; 二是进一步优化网络结构和训练方法, 解决 GAN 难训练, 容易出现梯度消失和梯度爆炸的问题.



图9 模块消融实验效果对比

表4 不同损失组合结果对比

L_{ad}	L_{pix}	L_{str}	L_p	PSNR	SSIM
*	*			26.77	0.88
*		*		26.49	<u>0.89</u>
*	*	*		<u>29.04</u>	0.90
*	*	*	*	29.22	0.90

参考文献

- 1 陈清江, 李金阳, 胡倩楠. 基于并联残差网络的低照度图像增强算法. 激光与光电子学进展, 2021, 58(14): 1410015.
- 2 蔡文成. 基于生成对抗网络的低照度图像增强方法研究 [硕士学位论文]. 武汉: 湖北工业大学, 2020.
- 3 He KM, Sun J, Tang XO. Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341–2353. [doi: 10.1109/TPAMI.2010.168]
- 4 Zuiderveld K. Contrast limited adaptive histogram equalization. Graphics Gems IV, 1994: 474–485.

- 5 Jobson DJ, Rahman Z, Woodell GA. Properties and performance of a center/surround Retinex. IEEE Transactions on Image Processing, 1997, 6(3): 451–462. [doi: 10.1109/83.557356]
- 6 王小明, 黄昶, 李全彬, 等. 改进的多尺度 Retinex 图像增强算法. 计算机应用, 2010, 30(8): 2091–2093.
- 7 Wang SH, Zheng J, Hu HM, et al. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 2013, 22(9): 3538–3548. [doi: 10.1109/TIP.2013.2261309]
- 8 Guo XJ, Li Y, Ling HB. LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 2017, 26(2): 982–993. [doi: 10.1109/TIP.2016.2639450]
- 9 Ying ZQ, Li G, Gao W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv: 1711.00591, 2017.
- 10 Lore KG, Akintayo A, Sarkar S. LLNet: A deep autoencoder

- approach to natural low-light image enhancement. *Pattern Recognition*, 2017, 61: 650–662. [doi: [10.1016/j.patcog.2016.06.008](https://doi.org/10.1016/j.patcog.2016.06.008)]
- 11 Lv FF, Lu F, Wu JH, *et al.* MBLLEN: Low-light image/video enhancement using CNNs. *British Machine Vision Conference 2018*. Newcastle: BMVA Press, 2018. 1–13.
 - 12 Lv FF, Li Y, Lu F. Attention-guided low-light image enhancement. *arXiv: 1908.00682*, 2019.
 - 13 Wang RX, Zhang Q, Fu CW, *et al.* Underexposed photo enhancement using deep illumination estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 6842–6850.
 - 14 Jiang YF, Gong XY, Liu D, *et al.* EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 2021, 30: 2340–2349. [doi: [10.1109/TIP.2021.3051462](https://doi.org/10.1109/TIP.2021.3051462)]
 - 15 Sajjadi MSM, Schölkopf B, Hirsch M. EnhanceNet: Single image super-resolution through automated texture synthesis. *IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 4501–4510.
 - 16 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
 - 17 江泽涛, 覃露露. 一种基于 U-Net 生成对抗网络的低照度图像增强方法. *电子学报*, 2020, 48(2): 258–264.
 - 18 Hu J, Shen L, Albanie S, *et al.* Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011–2023. [doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372)]
 - 19 Wang XL, Girshick R, Gupta A, *et al.* Non-local neural networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7794–7803.
 - 20 Zhu Z, Xu MD, Bai S, *et al.* Asymmetric non-local neural networks for semantic segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019. 593–602.
 - 21 Zhang C, Yan QS, Zhu Y, *et al.* Attention-based network for low-light image enhancement. *2020 IEEE International Conference on Multimedia and Expo (ICME)*. London: IEEE, 2020. 1–6.
 - 22 Tao L, Zhu C, Xiang GQ, *et al.* LLCNN: A convolutional neural network for low-light image enhancement. *IEEE Visual Communications and Image Processing*. Saint Petersburg: IEEE, 2018. 1–4.
 - 23 Wang Z, Bovik AC, Sheikh HR, *et al.* Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, 13(4): 600–612. [doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861)]
 - 24 Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. *Proceedings of the IEEE 37th Asilomar Conference on Signals, Systems & Computers*. Pacific Grove: IEEE, 2003. 1398–1402.
 - 25 Ledig C, Theis L, Huszár F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2016. 105–114.
 - 26 卫依雪, 周冬明, 王长城, 等. 结合多分支结构和 U-net 的低照度图像增强. *计算机工程与应用*, 2021: 1–12. [doi: [10.3778/j.issn.1002-8331.2104-0432](https://doi.org/10.3778/j.issn.1002-8331.2104-0432)]