

一种基于 MR-VAE 的低照度图像增强方法

江泽涛¹⁾ 伍旭¹⁾ 张少钦²⁾

¹⁾(桂林电子科技大学广西图像图形与智能处理重点实验室 广西 桂林 541004)

²⁾(南昌航空大学土木建筑学院 南昌 330063)

摘要 针对低照度图像多重失真特点(低亮度、多噪声和模糊等),本文基于变分自编码器提出了一种多重构变分自编码器(Multiple Reconstruction-Variational AutoEncoder, MR-VAE),逐步增强、从粗到细地生成高质量低照度增强图像. MR-VAE由特征概率分布捕获、全局重构和细节重构三个模块构成,核心思想是将全局特征与局部特征分阶段重建、将多重失真问题逐步解决,全局重构模块构建图像全局特征,提高全局亮度,得到较粗糙的图像;细节重构模块权衡去噪与去模糊,生成细节更逼真、噪声更少与局部亮度更合适的图像;此外,本文定义了一个多项损失函数替代 l_2 损失,以引导网络生成高质量图像.实验结果表明,多重构与多项损失函数的设计提高了网络生成复杂图像、处理多重失真的低照度图像性能,且提高了生成图像的质量、信噪比和视觉特性.

关键词 低照度图像增强;多重构;多项损失;多重失真;变分自编码;残差网络

中图法分类号 TP391 DOI号 10.11897/SP.J.1016.2020.01328

Low-Illumination Image Enhancement Based on MR-VAE

JIANG Ze-Tao¹⁾ WU Xu¹⁾ ZHANG Shao-Qin²⁾

¹⁾(Guangxi Key Laboratory of Image and Graphics Intelligent Processing, Guilin University of Electronic Technology, Guilin, Guangxi 541004)

²⁾(School of Civil Architecture, Nanchang Hangkong University, Nanchang 330063)

Abstract According to our investigation, low-illumination images have multiple distortion characteristics (including low light, high noise, blur, etc.). In order to better enhance the low-illumination image, we propose a Multi-Reconstruction Variational AutoEncoder (MR-VAE) based on the Variational AutoEncoder (VAE) to gradually enhance the image and generate high quality low-illumination enhanced image from coarse to fine. MR-VAE is mainly composed of three modules: Feature Probability Distribution Capture (FPDC), Global Reconstruction (GR), and Detail Reconstruction (DR). Its key idea is to generate the global and local features of the image in stages, and solve the problem of multiple distortion step by step. Finally, MR-VAE can capture low-illumination to normal illumination, noisy to noiseless, fuzzy to clear composite nonlinear mapping. The FPDC module is mainly used to capture hidden variables that cover the entire image feature, which is equivalent to encoding the entire image. The GR module is mainly used to capture the nonlinear mapping from low-illumination to normal illumination, and uses the hidden variables to generate global features of the image step by step (global features include: scene, color distribution, illumination characteristics), and finally obtains appropriate brightness enhancement, rough quality image. The DR module is mainly used to capture composite nonlinear maps from noisy to noiseless and clear to fuzzy, finding a good balance between removing noise and retaining detailed information, helping the network to generate high-quality images which

details more realistic, less noisy, and more suitable for local brightness. More importantly, we redefine a multi-loss function to replace the l_2 loss function, which improves the image quality criteria and guides the network to generate high quality images. This multi-loss function consists of “hidden variable content loss”, “global reconstruction loss” and “detail reconstruction loss”. Hidden variable content loss helps the FPDC module capture hidden variables that better reflect the probability distribution of the essential content of the image. Global reconstruction loss is used to help the GR module generate large image features and overall brightness. Detail reconstruction loss helps the DR to generate detailed features between rough and high-resolution images, and helps DR to be more robust to noise. In addition, based on the MS COCO dataset, we have created pairs of low-light images for training our network. Gamma correction is used to adjust the brightness, including four different brightness levels. Gaussian noise is used to simulate actual noise, the noise level and brightness level meet a certain relationship. Gaussian blur is used to simulate the blur of the image to produce a composite image that is very close to the actual low-illumination image. In the experimental stage, We did a complete experiment, including the ablation experiment of the network structure and the comparison experiment with other methods. The ablation experiment is mainly to verify the depth of the network structure and the role of the multi-loss function. The comparison experiment includes the traditional method and the deep learning methods. The experimental results show that the multi-reconstruction and multi-loss function can help the network to generate complex images and improve the network’s enhanced performance for low-intensity images with multiple distortions. Indicating that our method can better Enhances low-illumination images and has better ability to remove noise.

Keywords low-illumination image enhancement; multiple reconstruction; multi-loss function; multiple distortion; variational AutoEncoder; residual network

1 引言

低照度环境导致图像成像质量下降,图像噪声较多、对比度较低,用于图像分类、目标识别、图像理解分析、超分辨率重建等图像处理时,效果不理想.因此需要对该类图像进行照度增强,即提高图像整体和局部的对比度、去噪,适当调整图像背景和边缘.

目前处理低照度图像主要通过物理途径和低照度增强方法提高图像质量.物理途径可在一定程度上提高成像质量,但实用性不强;如提高设备感光度、增大光圈和延长曝光时间等.经典的低照度图像增强方法主要有基于直方图均衡化(Histogram Equalization, HE)^[1]和基于 Retinex 理论^[2]的方法. HE 是以累计分布函数为基础的直方图修改方法,将图像直方图调整为均衡分布以拉伸图像的动态范围,从而提高图像对比度.基于 HE 的改进方法有 Kim^[3]、Celik^[4]等,该类方法操作简单、效率高,但生成图像易受伪影影响、真实感不强.

基于 Retinex 理论的方法由 Land 等人^[2]提出.该类方法模拟人类视觉系统处理图像的过程,遵循将图像分为反射、照明两个分量的假设.早期基于 Retinex 理论的算法有 Single-Scale Retinex(SSR)^[5]和 Multi-Scale Retinex(MSR)^[6].后续算法有 Dong 等人^[7]提出的 DeHZ 增强算法;Fu 等人^[8]提出的加权可变模型;Guo 等人^[9]通过查找 R、G、B 通道中的最大值重构低照度图像的光照图,以此估计环境光照;文献[10-12]也是基于 Retinex 理论进行改进的.由于图像通道间的非线性、数据复杂性,难以准确估计光照分量,导致上述方法易出现过度增强、局部失真等问题.

近年来,基于深度学习的低照度图像增强方法陆续提出. Lore 等人^[13]提出的 LLNet 通过构造栈式稀疏去噪自编码器实现低照度增强;Fu 等人^[14]提出了基于融合的方法;Shen 等人^[15]基于 MSR 算法,搭建了一个 MSRNet 网络用于对比度增强和去噪;Yu 等人^[16]提出了基于对抗性学习的方法;Chen 等人^[17]构建了深度卷积神经网络实现低照度增强;Chen 等人^[18]基于 Retinex 理论,提出了 Retinex-Net

用于低照度增强;Gharbi 等人^[19]、Wang 等人^[20]提出了学习图像预测二次变化,将其与输入结合的增强方法.上述深度学习方法较经典方法取得了更好的增强效果,但通常侧重于增强亮度,而忽略了实际低照度图像会出现多重失真,如噪声、模糊、不同照度等.

为提高对多重失真的低照度图像的处理能力,解决增强图像不真实、模糊等问题,本文提出了 MR-VAE 用于低照度图像增强. MR-VAE 由特征概率分布捕获、全局重构和细节重构三个模块构成.“特征概率分布捕获”用于获取涵盖了图像全局特征的隐变量;“全局重构”主要用于增强图像全局亮度,生成亮度显著提升、具有全局特征(场景、颜色分布和照度特征等)的图像;“细节重构”主要用于权衡去噪、去模糊,去噪的同时保留较多细节特征(细节锐度、色差和视觉特性等),同时微调局部亮度,得到细节锐度更高、视觉特性更好的图像.此外,本文定义了一个涵盖图像质量评估、内容生成和细节构建的多项损失函数替代 l_2 损失.实验结果表明,多重重构与多项损失函数提高了网络生成复杂图像、增强多重失真的低照度图像能力,使其可生成高质量、高信噪比且更符合视觉特性的图像.

2 相关工作

2.1 变分自编码器

变分自编码器(Variational AutoEncoder, VAE)^[21]是一种生成模型,其关键思想是对可能生成数据 X 、符合高斯分布的隐变量 Z 进行采样,并从中计算概率分布模型 $P(X)$.对于图像生成任务,给定数据集 $X = \{x^{(i)}\}_{i=1}^N$,VAE 能够学习图像特征的概率分布.为此,VAE 训练一个 θ 参数化的概率分布模型 $p_\theta(X)$ 以近似数据分布,并用最大化对数似然来优化模型^[22].如式(1)所示.

$$\begin{aligned} \log p_\theta(X) &= \log p_\theta(x^{(1)}, x^{(2)}, \dots, x^{(N)}) \\ &= \sum_{i=1}^N \log p_\theta(x^{(i)}) \end{aligned} \quad (1)$$

但 X 是高维数据,其结构化图模型的潜变量之间相互作用导致 $P_\theta(X)$ 求解困难,因此无法估计边缘似然、无法使用 EM 等算法.故 VAE 引入低维隐变量 Z ,构造更易求解的近似推断模型 $q_\varphi(Z|X)$,采用最大化似然对 $q_\varphi(Z|X)$ 进行有效优化.单个样本的最大似然如式(2)所示.

$$\begin{aligned} \log p_\theta(x^{(i)}) &= L(\varphi, \theta; x^{(i)}) + \\ &D_{\text{KL}}(q_\varphi(z|x^{(i)}) \| p_\theta(z|x^{(i)})) \end{aligned} \quad (2)$$

因 $D_{\text{KL}}(q_\varphi(z|x^{(i)}) \| p_\theta(z|x^{(i)})) \geq 0$,则有

$$\log p_\theta(x^{(i)}) \geq L(\varphi, \theta; x^{(i)}) \quad (3)$$

$$\begin{aligned} L(\varphi, \theta; x^{(i)}) &= E_{q_\varphi(z|x^{(i)})} [\log p_\theta(x^{(i)}|z)] - \\ &D_{\text{KL}}(q_\varphi(z|x^{(i)}) \| p_\theta(z)) \end{aligned} \quad (4)$$

其中, $D_{\text{KL}}(q_\varphi(z|x^{(i)}) \| p_\theta(z))$ 为 $q_\varphi(z|x^{(i)})$ 与 $p_\theta(z)$ 间的 KL 散度; $L(\varphi, \theta; x^{(i)})$ 为 $x^{(i)}$ 的变分下界.由式(2)~(4)可知,VAE 将求解 $p_\theta(X)$ 转换成优化 $L(\varphi, \theta; x^{(i)})$ 最大值的,相当于最小化 KL 散度.如图 1 所示.

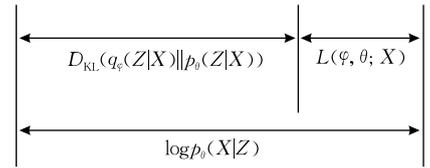


图 1 VAE 变分下界与 KL 散度的关系

2.2 残差网络

残差网络是 He 等人^[23]于 2015 年提出的深度卷积网络,主要由残差块以串联形式组成,残差块由卷积操作和跳跃结构构成.跳跃结构在未增加参数量和计算复杂度的前提下,将浅层网络提取的特征信息与卷积输出以逐通道、逐元素方式相加后作为深层网络输入的一部分,可定义成式(5)所示.目前残差网络的设计思想已应用于图像分类、目标识别等领域.

$$y = F(x, \{W_l, b_l\}) + W_s x \quad (5)$$

其中, x 与 y 分别为残差块的输入和输出; W_l 和 b_l 分别为第 l 层卷积操作的权重和偏移量, $F(x, \{W_l, b_l\})$ 为残差块的卷积操作; W_s 为调整残差块输入通道数的矩阵.

3 MR-VAE 结构设计

MR-VAE 是由卷积、池化和上采样操作构成的全卷积网络结构^[24].使用成对的“低照度-正常照度 (I_l-I_n)”图像作为训练集,其中 I_l 由 I_n 加噪、伽马校正和模糊处理后得到,以模拟多重失真的真实低照度图像;采用“端对端”训练策略,利用 MR-VAE 隐含层学习低照度到正常照度、含噪图像到去噪图像和模糊到清晰的复合非线性映射,即捕获真实低照度到正常照度图像的近似非线性映射.则低照度图像增强问题可表示如式(6)所示.

$$I_l = \mathbf{M} \times I_n \quad (6)$$

其中, \mathbf{M} 表示导致 I_n 退化成 I_l 的高维、非线性矩阵,因此低照度图像增强问题还可表示如式(7)所示.

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) dz$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right) d\left(\frac{z-\mu}{\sigma}\right) \quad (9)$$

3.3 全局重构模块

GR 主要用于捕获从低照度到正常照度的非线性映射关系, 利用卷积与上采样操作, 逐步将编码得到的低维特征映射到高维空间, 构建图像全局特征, 得到亮度适当增强, 但细节信息较少、噪声未完全去除的粗糙图像。

GR 初始输入为 FPDC 获得的低维隐变量 Z , 其涵盖了所有特征分布信息; GR 每层输入由 FPDC 跳跃连接送来的 $f_{en}^{(l)}$ 与上采样信息 $f_{de}^{(l-1)}$ 以拼接的方式组合得到, 以弥补编码下采样丢失的特征信息; GR 每层对输入信息进行卷积和上采样操作, 其中卷积模块的设计与编码部分保持一致; 上采样操作采用“Resize conv”的方式, 避免了上采样后出现“棋盘伪影”, 且不限输入图像的大小。

3.4 细节重构模块

细节重构模块 (Detail Reconstruction, DR) 主要用于捕获含噪到去噪图像、模糊到清晰的复合非线性映射关系, 在去噪与保留细节特征间找到合适的关系, 调整局部亮度。其输入由 GR 输出、FPDC 第一个编码块输出组成。前者具有丰富的场景、颜色分布和照度特征, 但下采样过程丢失了部分信息; 后者为浅层特征图, 包含了大量的基础特征信息, 但颜色分布和照度特征不明显 (可添加特征图示意图)。两个结合后具有丰富的特征信息, 有利于 DR 捕获复合非线性映射关系, 从而较好地重构图像细节特征、去除噪声和调整局部图像亮度, 以忠实地将低照度图像还原或近似还原到正常照度图像。

DR 的提出相当于增加了 MR-VAE 的网络深度, 以学习更复杂的复合非线性映射关系, 但更深的网络会带来梯度消失或爆炸、网络学习能力退化等问题。为此, Ioffe 等人^[25] 提出了批归一化处理 (Batch Normalization, BN), 提高训练时模型对不同超参 (学习率、初始化) 的鲁棒性^[26]; He 等人^[23] 提出了残差网络, 通过跳跃连接将浅层网络提取的特征信息作为深层网络输入的一部分, 这两种方法都有效保持了深层网络的学习能力, 有助于训练高效的深层网络模型。

鉴于此, DR 引入残差思想以确保特征信息的有效传递, 添加 BN 层以提高网络性能, 故 DR 主要

由跳跃连接、残差模块和普通卷积模块组成。跳跃结构将 FPDC 第一层的卷积输出和 GR 的输出以拼接的方式组合后送入 DR。残差模块如图 3 所示, Conv k3n64s1 表示卷积核大小为 3×3 、卷积核数量为 64、卷积步长为 1 的卷积操作, ‘ \oplus ’ 表示逐像素的相加。普通卷积模块由 3 个卷积操作组成。

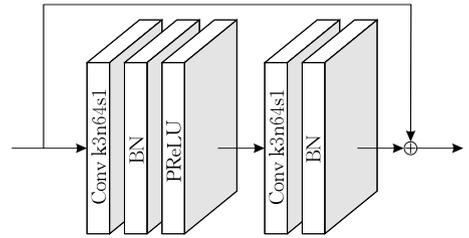


图 3 DR 残差块结构

4 多项损失函数设计

本文定义了一个涵盖图像质量评估、内容生成和细节构建的多项损失函数替代 l_2 损失, 以弥补生成图像质量评判标准不足, 引导网络生成高质量图像。如式 (10) 所示, $L_{Z-Content}$ 为隐变量内容损失; L_{GR} 为传统解码模块损失; L_{DR} 为细节重构模块损失。 ω_1 、 ω_2 和 ω_3 为各项损失的权重, 采用网格法调参并结合实验验证, 最终设置 $\omega_1 = 1$ 、 $\omega_2 = 10$ 和 $\omega_3 = 1$ 。

$$Loss_{MR-VAE_\theta} = \omega_1 L_{Z-Content} + \omega_2 L_{GR} + \omega_3 L_{DR} \quad (10)$$

4.1 隐变量内容损失

I_l 较于 I_n 噪声更多、亮度和对对比度更低, 但两者本质内容是一致的 (场景、主要物体和轮廓等), 为捕获更能反映图像本质内容概率分布的隐变量, 本文提出了隐变量内容损失。此外, 该损失函数还可激励 MR-VAE _{θ} 网络生成具有感知相似性的图像、提高图像信噪比, 且允许生成图像与目标图像有细微的变化, 进而减少过拟合并提高网络的泛化能力。将 I_l 、 I_n 同时送入编码器, 分别得到隐变量 Z_l 、 Z_n , 其内容损失如式 (11) 所示, N 表示 I_l 数量。

$$L_{Z-Content} = \frac{1}{N} \sum_{i=1}^N (Z_l^{(i)} - Z_n^{(i)})^2 \quad (11)$$

4.2 全局重构损失

GR 的损失函数由网络输入 x 与 GR 输出间误差、 $q_\varphi(z|x^{(i)})$ 与 $p_\theta(z)$ 间的 KL 散度组成, 其中 KL 散度相当于一个正则项, 确保 MR-VAE 不会退化成普通自编码器。如式 (12) 所示。

$$L_{GR} = L_{MSE} + L_{KL} \quad (12)$$

令 J 为隐变量 Z 的维度数量, μ_i 、 σ_i 分别为第 i 张图像的均值和方差, 由式 (3)、(4) 知:

$$L_{\text{MSE}} = -E_{q_{\varphi}(z|x^{(i)})} [\log p_{\theta}(x^{(i)}|z)]$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^N (x^{(i)} - g_{\theta}(z^{(i)}))^2 \quad (13)$$

$$L_{\text{KL}} = D_{\text{KL}}(q_{\varphi}(z|x^{(i)}) \| p_{\theta}(z))$$

$$= \frac{1}{2} \sum_{j=1}^J (1 + \log(\sigma_j)^2 - \mu_j^2 - \sigma_j^2) \quad (14)$$

4.3 细节重构损失

DR 主要目的是重构粗糙图像和高分辨率图像间的细节特征,且对噪声具有更强的鲁棒性. DR 损失函数受偏离正常范围的离群样本影响小、鲁棒性更好,有利于提高图像轮廓对比度、构建图像细节信息,如式(15)所示, N 为低照度图像数量; $I_n^{(i)}$ 为第 i 张正常照度图像; $DR^{(i)}$ 为第 i 张 DR 的增强图像.

$$L_{\text{DR}} = \frac{1}{N} \sum_{i=1}^N |I_n^{(i)} - DR^{(i)}| \quad (15)$$

5 实验结果与分析

5.1 数据集制作

本文从 MS COCO^[27] 数据集中随机选择 2200 张图像用于制作 MR-VAE 的训练和部分测试数据集; 2000 张用于制作训练集 Synthetic_train_set,

200 张用于制作测试集 Synthetic_test_set. 此外, 本文还从 ExDark^①、DICM^[28]、TID2013^[29] 数据集分别随机选取 120、70、25 张图片构成真实低照度图像测试集 True_test_set.

本文对 2200 张原始图像添加噪声、降低亮度和模糊处理,以模拟多重失真的低照度图像. 对于 Synthetic_train_set, 每张图像在 4 个光照等级上分别处理 3、7、7、3 次, 得到 20 张合成图, 共 40000 张; 对于 Synthetic_test_set, 1 张图像处理 1 次, 共 200 张. 其中, 采用 4 个不同参数范围的伽马校正模拟四个不同的光照等级, 如式(16)所示, I_l 为低照度图像; I_n 为原始图像; 根据亮度越低, 噪声越多的特点, 设定高斯噪声参数与伽马校正参数成指数关系, 如式(17)所示; 使用 $\sigma \in [0, 1.1]$ 的高滤波器对图像做模糊化处理. 合成低照度图像示例如图 4 所示, 可见上述合成策略可较好地模拟不同照度条件. 顶行图像参数为相机曝光系数, 底行图像参数分别为伽马校正系数、高斯噪声系数和高斯模糊系数. 合成图像的系数设置如表 1 所示.

$$I_l = (I_n)^{\gamma}, \gamma \in [1.5, 4.5] \quad (16)$$

$$\sigma = \begin{cases} e^{\gamma}, & \gamma \in [1.5, 4.5] \\ 10, & \gamma > 3 \end{cases} \quad (17)$$



图 4 合成低照度图像示例(顶行为相机拍摄,底行为合成的低照度图像)

表 1 每张原图制作合成图像的系数设置

光照等级	Canon5D markIII 曝光系数	伽马校正系数 γ	高斯噪声系数 σ	合成图像占比
1	$[-0.5, -1]$	$[1.5, 2]$	e^{γ}	3
2	$[-1, -2]$	$[2, 3]$	e^{γ}	7
3	$[-2, -3]$	$[3, 4]$	10	7
4	$[-3, -3.5]$	$[4, 4.5]$	10	3

5.2 训练细节与参数

训练前对低照度图像进行像素值归一化处理, 有助于提高网络的收敛速度与精度、减缓过拟合问

题. 输入图像大小为 512×512 , 4 张/次. FPDC、GR 和 DR 模块的卷积核为 3×3 , 卷积步长为 1, 分别使用 $ReLU$ 、 $ReLU$ 和 $PReLU$ 作为激活函数; FPDC 的最大化池化操作步长为 2; 初始学习率为 0.001, 每 5000 次以 0.96 的衰减率减小; 共 6 批次, 10000 次训练/批次; 使用 Adam 优化器. MR-VAE 详细设计如表 2 所示.

① <https://github.com/cs-chan/Exclusively-Dark-Image-Dataset/blob/master/Dataset/README.md>

表 2 MR-VAE 详细参数

Output size	FPDC	Output size	GR	Output size	DR	Params
256×256	[conv 3×3,64,s1]×2 max pool,2×2,s2	64×64	[conv 3×3,64,s1]×2 resize 512×512 conv 3×3,64,s1	512×512	$(\text{conv } 3\times 3, 64, \text{BN, prelu}) \times 5$	67 980 745
128×128	[conv 3×3,128,s1]×2 max pool,2×2,s2	128×128	[conv 3×3,64,s1]×2 resize 512×512 conv 3×3,64,s1	512×512		
64×64	[conv 3×3,256,s1]×2 max pool,2×2,s2	256×256	[conv 3×3,64,s1]×2 resize 512×512 conv 3×3,64,s1			
32×32	[conv 3×3,512,s1]×2 max pool,2×2,s2	512×512	[conv 3×3,64,s1]×2 resize 512×512 conv 3×3,64,s1	512×512	conv 3×3,64 conv 3×3,3	
32×32	[conv 3×3,1024,s1]×2	512×512	conv 3×3,64,s1 conv 3×3,3,s1			

5.3 网络结构与损失函数作用评估

5.3.1 细节重构模块消融实验

3.4 节提出的 DR 相当于增加了网络的深度,一方面有助于提取更多的细节特征,一方面有利于捕获复合非线性映射关系。

DR 模块消融实验的结果如图 5 所示。可见,图 5 顶行船的左上角区域,有 DR 结构去噪效果更好,能够更好地处理光渐变、图像更自然;图 5 底行街景图的招牌字体,有 DR 结构的边缘更好,细节信息保留的较多,且字体背景相对更平滑,噪声更少。



图 5 细节重构模块消融实验结果比较

5.3.2 细节重构模块特征图分析

DR 各残差块生成的特征如图 6 顶行所示,可知网络逐步生成更多的纹理信息,如图 6 中的救生圈、天空的颜色渐变;DR 普通卷积模块的第一个卷积组合 DR 输入和残差模块的输出,故特征信息较乱,第二个卷积整合特征信息,进一步提高图像的对

比度,如图 6 中底行的 conv-1,conv-2. 最终将通道数调整为 3,得到输出。

5.3.3 多项损失函数消融实验

本文提出了多项损失函数替代原始的单一损失函数,其有利于在抑制噪声的同时,保留较多细节信息、提高增强图像的视觉特性。消融实验的结果如图 7 所示。

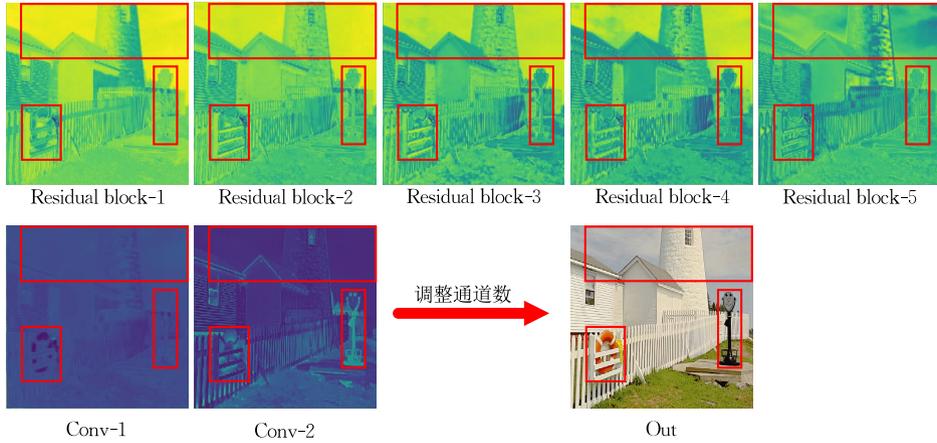


图 6 细节重构各阶段生成的特征图

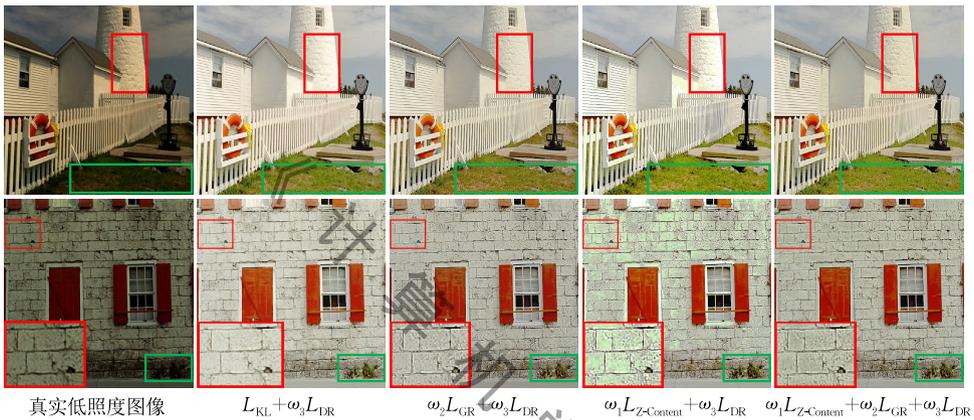


图 7 损失函数消融实验结果比较

当仅有 $(L_{KL} + \omega_3 L_{DR})$, 低照度图像较亮的区域易被增强过度, 如图 7 顶行图像中红框内的白墙, 且图像细节信息较少, 如底行的红框内的墙砖线;

当仅有 $(\omega_2 L_{GR} + \omega_3 L_{DR})$, 增强后的颜色容易失真, 如图 7 顶行图像中绿框内的草地;

当仅有 $(\omega_1 L_{Z-Content} + \omega_3 L_{DR})$ 两项, 低照度图像稍亮的区域极易被过度增强、失真, 如图 7 中红框内的墙。

当为多项损失 $(\omega_1 L_{Z-Content} + \omega_2 L_{GR} + \omega_3 L_{DR})$, 图像细节信息保留的较好、亮度适当增强、视觉特征更好, 整体的增强效果最好。

损失函数消融实验在 Synthetic_test_set 上的 PSNR/SSIM 均值如表 3 所示, 可知多项损失函数的 PSNR/SSIM 均值最高, 客观说明了其在该函数的引导下, 网络生成的图像质量最好。

表 3 损失函数消融实验的 PSNR/SSIM 均值比较

测试集	$L_{KL} + \omega_3 L_{DR}$	$\omega_3 L_{GR} + \omega_3 L_{DR}$	$\omega_1 L_{Z-Content} + \omega_3 L_{DR}$	$\omega_1 L_{Z-Content} + \omega_2 L_{GR} + \omega_3 L_{DR}$
Synthetic_test_set	22.847/0.776	22.276/0.772	22.765/0.789	23.761/0.806

5.4 训练模型评估

分别在 Synthetic_test_set、True_test_set 上比较 MR-VAE 与 MSRRCR^[30]、DeHZ 和 LIME 传统增强方法, 及 MSRNet 的效果。在 True_test_set 上比较 MR-VAE 与 MSRNet、RetinexNet 两种基于深度学习的方法的效果。对比实验中, 本文优先使用作者开源的代码、参数或依据相关论文的描述复现算法; 对于基于深度学习的方法, MSRNet 使用 Synthetic_train_set 进行训练, RetinexNet^① 直接使用作者提

供的模型文件。

5.4.1 模型复杂度分析

为捕获低照度到正常照度、含噪图像到去噪图像和模糊到清晰的复合非线性映射, 本文设计了更深的 MR-VAE, 其由 GR、DR 两个模块组成。下面将简要分析 MR-VAE 与对比实验中的基于深度学习方法的模型参数量、模型大小, 如表 4 所示。

① <https://github.com/weichen582/RetinexNet>

表 4 模型复杂度分析

Network	Params	Model size/KB
MSRNet	9974	132
RetinexNet	444 613	8263
MR-VAE	67 980 745	796 655

5.4.2 主观视觉评估

由图 8 知,MR-VAE 在亮度、内容和细节锐度等方面的增强效果是相对较好的.图 8 为 MR-VAE 与传统、基于深度学习方法在 True_test_set 的增强

效果.分析可知,MSRCR 的图像内容基本生成,但整体偏暗;DeHZ 的整体效果不错,但局部亮度过度增强、边缘易出现伪影,如第二个场景中的招牌、第三个场景中的窗户边缘;LIME 处理后的颜色存在明显过度增强、不自然,如第四个场景的草地,成深绿色;MSRNet 出现了局部颜色失真,如第三、四个场景中的草;RetinexNet 的细节信息保留较好,但色彩增强过度、图像整体失真.



图 8 不同方法在 True_test_set 上的低照度增强效果

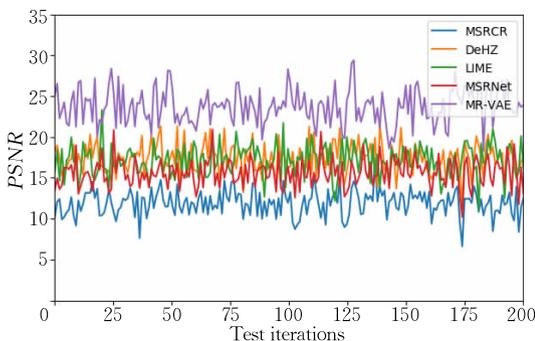
5.4.3 客观定量评估

采用 PSNR、SSIM 指标衡量增强效果的结构相似度、对比度与信噪比.由表 5、图 9 可知,MR-

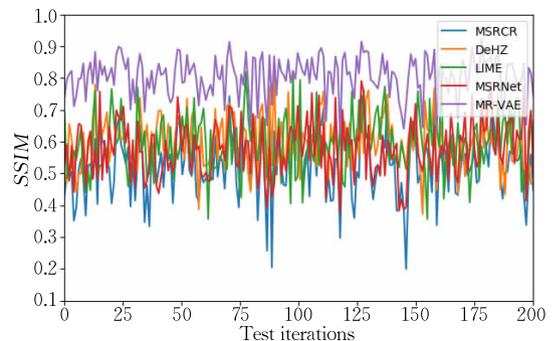
VAE 方法较于其他几种方法的 PSNR/SSIM 值更高,客观验证了 MR-VAE 增强结果噪声更少,对比度、结构更接近原始图像.

表 5 几种方法在 Synthetic_test_set 上的 PSNR/SSIM 均值

测试集	MSRCR	DeHZ	LIME	MSRNet	MR-VAE
Synthetic_test_set	11.925/0.526	17.406/0.597	17.138/0.513	15.625/0.578	23.761/0.806



(a) 几种方法在 Synthetic_test_set 上的 PSNR



(b) 几种方法在 Synthetic_test_set 上的 SSIM

图 9 几种方法在 Synthetic_test_set 上的 PSNR 和 SSIM 比较

图 10 展示了不同方法在 Synthetic_test_set 的低照度增强结果,右下角表示该图像的 PSNR/SSIM. 分析可知,对于亮度低、噪声多或模糊的图像,MR-VAE 仍然得到整体质量更接近于原始图像的结果. 如图 10 所示,MSRCR 方法亮度没有明显增强;DeHZ 和 LIME 还原了亮度,但边缘存在伪影,色彩还原过度,如第一个场景中的窗户边缘比较模糊,第二个场景的绿色被过度增强、人脸比较模糊;MSRNet 亮度还原效果较好,但细节信息较少,整体偏模糊,如第一个场景中的窗户线条;此外,上述对比方法降噪能力较弱,如图 10 中的噪声明显.

此外,采用谷歌公司提供的 GCVA^①(Google Cloud Vision API)检测和识别 MR-VAE 与几个基于深度学习方法增强结果中的物体信息,评估各方法对提高其他计算机视觉任务性能的作用. 其中 GCVA 具备物体识别、OCR 文字识别等功能,且识别准确率高.

由图 11 可知,GCVA 在原图上检测出 5 个物体,GCVA 在 MSRNet 的增强结果检测出的物体数量最少,RetinexNet 次之,MR-VAE 最多,有 7 个物体,尤其是检测到“car”,是其他方法都未检测到的. 从上述分析可知,MR-VAE 对提升 GCVA 识别性能的作用较好.

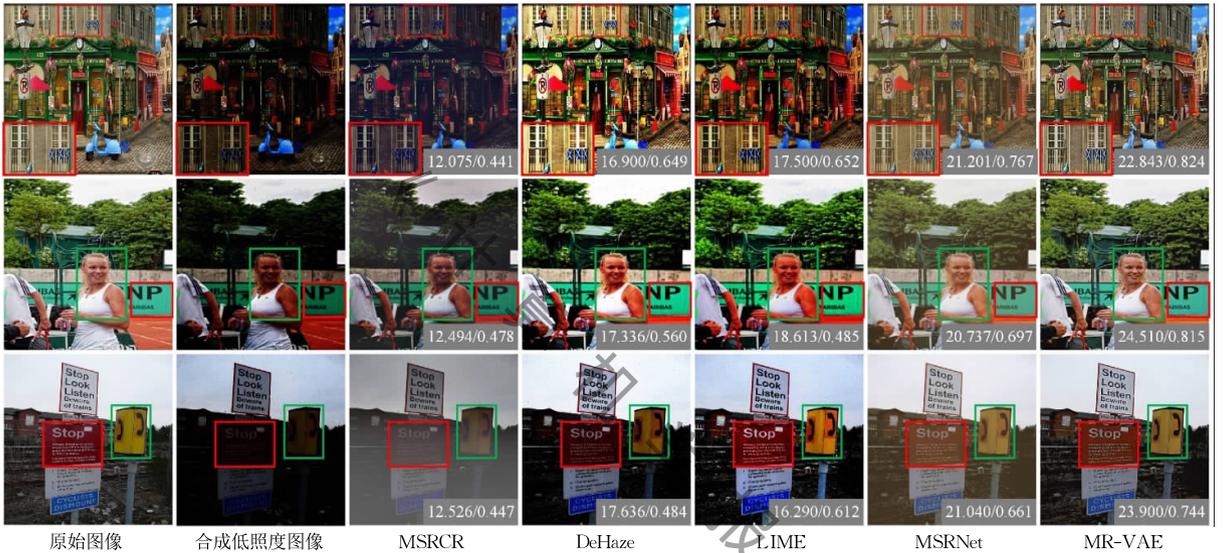


图 10 不同方法在 Synthetic_test_set 上的低照度增强效果

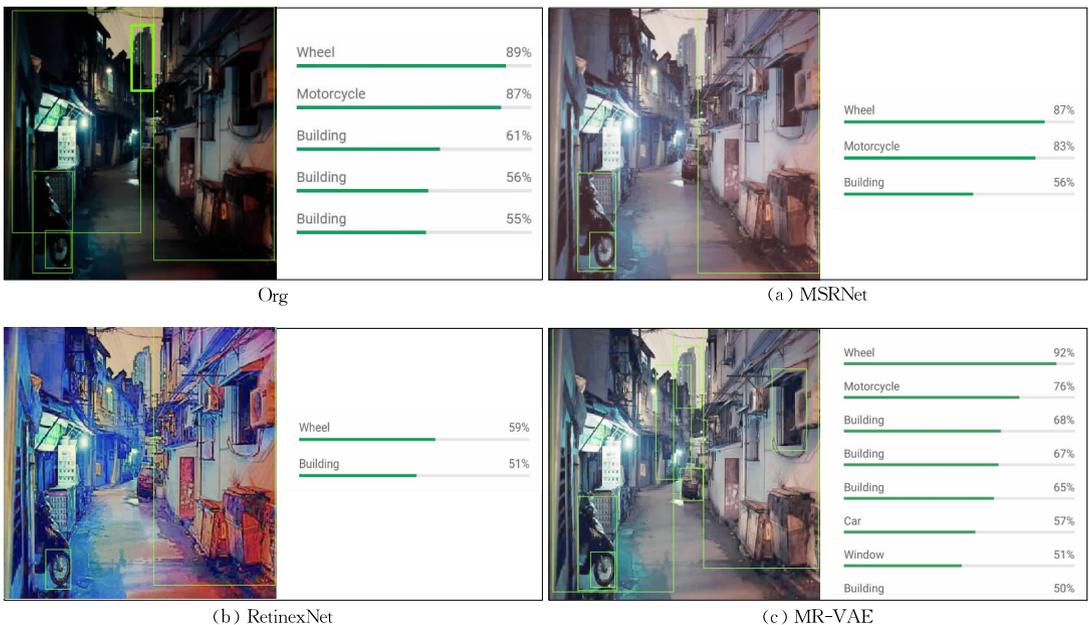


图 11 不同方法增强结果在 GCVA 的处理效果

① <https://cloud.google.com/vision/>

综上所述,相较于对比实验中的传统方法、基于深度学习的方法,MR-VAE 增强效果的色彩和整体亮度更接近原图、细节信息更多。可调整全局亮度,且在局部过曝区域不会过度增强,这得益于 GR 模块基于整张图像调整亮度、重构图像;且图像噪声较少、细节锐度较好,这得益于 DR 模块恰当地权衡了去噪、保留细节间的关系。

6 结 论

本文通过借鉴变分自编码、残差网络设计思想,提出了一种基于 MR-VAE 的低照度图像增强方法,逐阶段、由粗到细生成高质量图像。MR-VAE 主要由特征分布概率捕获、全局重构和细节重构模块组成,特征分布概率捕获模块用于捕获图像特征信息,全局重构模块生成涵盖图像基本内容、粗糙的图像,细节重构模块在此基础上生成细节锐度更好、噪声更少、色差更小的图像。此外,本文提出了多项损失函数替代 l_2 损失函数,该函数涵盖了图像质量评估、内容生成和细节构建,可引导网络生成高质量图像。实验验证了本文提出的方法对增强多重失真的低照度图像有较好的性能。

参 考 文 献

- [1] Cheng H D, Shi X J. A simple and effective histogram equalization approach to image enhancement. *Digital Signal Processing*, 2004, 14(2): 158-170
- [2] Land E H, McCann J J. Lightness and Retinex theory. *Journal of the Optical Society of America*, 1971, 61(1): 1-11
- [3] Kim W, Jeong J, You J. Contrast enhancement using histogram equalization based on logarithmic mapping. *Optical Engineering*, 2012, 51(6): 067002
- [4] Celik T. Spatial entropy-based global and local image contrast enhancement. *IEEE Transactions on Image Processing*, 2014, 23(12): 5298-5308
- [5] Jobson D J, Rahman Z, Woodell G A. Properties and performance of a center/surround Retinex. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 1997, 6(3): 451-462
- [6] Jobson D J, Rahman Z, Woodell G A. A multiscale Retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 1997, 6(7): 965-976
- [7] Dong X, Wang G, Pang Y, et al. Fast efficient algorithm for enhancement of low lighting video//*Proceedings of the 2011 IEEE International Conference on Multimedia and Expo. Barcelona, Spain*, 2011: 1-6
- [8] Fu X, Zeng D, Huang Y, et al. A weighted variational model for simultaneous reflectance and illumination estimation//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA*, 2016: 2782-2790
- [9] Guo X, Li Y, Ling H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 2016, 26(2): 982-993
- [10] Park S, Yu S, Moon B, et al. Low-light image enhancement using variational optimization-based Retinex model. *IEEE Transactions on Consumer Electronics*, 2017, 63(2): 178-184
- [11] Shi H, Kwok N, Wu H, et al. Logarithmic profile mapping multi-scale Retinex for restoration of low illumination images//*Proceedings of the 9th International Conference on Graphic and Image Processing (ICGIP 2017). International Society for Optics and Photonics. Qingdao, China*, 2018, 10615: 106152H
- [12] Zhang J, Zhou P C, Zhang Q. Low-light image enhancement based on iterative multi-scale guided filter Retinex. *Journal of Graphics*, 2018, 39(1): 1-11
- [13] Lore K G, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 2017, 61: 650-662
- [14] Fu X, Zeng D, Huang Y, et al. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 2016, 129: 82-96
- [15] Shen L, Yue Z, Feng F, et al. MSR-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017
- [16] Yu K, Liu W, Zhang Y, et al. DeepExposure: Learning to expose photos with asynchronously reinforced adversarial learning//*Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada*, 2018: 2149-2159
- [17] Chen C, Chen Q, Xu J, et al. Learning to see in the dark//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA*, 2018: 3291-3300
- [18] Wei C, Wang W, Yang W, et al. Deep Retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018
- [19] Gharbi M, Chen J, Barron J T, et al. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics*, 2017, 36(4): 118
- [20] Wang R, Zhang Q, Fu C W, et al. Underexposed photo enhancement using deep illumination estimation//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA*, 2019: 6849-6857
- [21] Kingma D P, Welling M. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013
- [22] Cai L, Gao H, Ji S. Multi-stage variational auto-encoders for coarse-to-fine image generation//*Proceedings of the 2019 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics. Alberta, Canada*, 2019: 630-638

- [23] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778
- [24] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3431-3440
- [25] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015
- [26] Santurkar S, Tsipras D, Ilyas A, et al. How does batch normalization help optimization?//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2018: 2483-2493
- [27] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 740-755
- [28] Lee C, Lee C, Kim C S. Contrast enhancement based on layered difference representation of 2D histograms. IEEE Transactions on Image Processing, 2013, 22(12): 5372-5384
- [29] Ponomarenko N, Ieremeiev O, Lukin V, et al. Color image database TID2013: Peculiarities and preliminary results//Proceedings of the European Workshop on Visual Information Processing (EUVIP). Paris, France, 2013: 106-111
- [30] Rahman Z, Jobson D J, Woodell G A. Multi-scale Retinex for color image enhancement//Proceedings of the 3rd IEEE International Conference on Image Processing. Lausanne, Switzerland, 1996: 1003-1006



JIANG Ze-Tao, Ph. D. , professor.

His research interests focus on computer vision.

WU Xu, M. S. candidate. His current research interest is computer vision.

ZHANG Shao-Qin, M. S. , professor. Her research interest is optimization calculation.

Background

To improve the enhancement ability of low-illumination images with multiple distortions (low light, multi-noise and blur, etc.) and solve the problems of unrealistic and fuzzy enhancement of images. We are inspired by the concept of Variational AutoEncoder and Residual Network, design and proposes MR-VAE for low-illumination image enhancement, select “end-to-end” training strategy and using the MR-VAE hidden layer to learn low illuminance-to-normal illumination, noisy-to-noiseless, and fuzzy-to-clear composite nonlinear mapping, equivalent to capture an approximate nonlinear mapping from true low-illumination to normal illumination images.

First of all, MR-VAE consists of three modules: feature probability distribution capture, global reconstruction and detail reconstruction. ‘Characteristic probability distribution capture’ is used to capture hidden variables that cover the global features of the image. ‘Global Reconstruction’ mainly captures non-linear mapping from low to normal illumination, enhances global brightness of the image, and produces images with significantly improved brightness and global features (scene, color distribution, and illuminance characteristics, etc.). ‘Detail Reconstruction’ is mainly used to weigh the denoising, deblurring, and capture the appropriate noisy-

to-noiseless, fuzzy-to-clear composite nonlinear mapping. Denoising retains more detail features (detail sharpness, chromatic aberration, and visual characteristics), while fine-tuning local brightness to obtain images with higher detail sharpness and better visual characteristics.

In addition, this paper defines a Multi-loss function replacement l_2 loss that cover image quality assessment, content generation, and detail construction to help guide the network to generate high quality images.

Finally, the multi-reconstruction and multi-loss function improve the network’s ability to generate complex images and enhance multiple distortions of low-illumination images, enabling them to generate images with high quality, high signal-to-noise ratio and more visual characteristics.

This work was supported by the National Natural Science Foundation of China (Nos. 61572147, 61876049, 61762066), the Guangxi Science and Technology Project (No. AC16380108), the Guangxi Key Laboratory of Image Graphics and Intelligent Processing (Nos. GIIP201701, GIIP201801, GIIP201802, GIIP201803), the Guangxi Graduate Education Innovation Program Funding Project; YCBZ2018052 Funding.