

VQCNIR: 使用矢量量化码本恢复更清晰的夜间图像

Wenbin Zou¹, Hongxia Gao^{1,2*}, Tian Ye³, Liang Chen⁴,
Weipeng Yang¹, Shasha Huang¹, Hongshen Chen¹, Sixiang Chen³

1 华南理工大学自动化科学与工程学院, 广州2 脑机接口研究中心, 琶洲实验室, 广州3
香港科技大学, 广州4 福建师范大学光子与电子工程学院, 福州hxgao@scut.edu.cn

Abstract

夜间摄影经常面临因黑暗环境和长时间曝光而导致的低光和模糊等挑战。目前的方法要么忽略先验, 直接拟合端到端网络, 导致光照不一致, 要么依靠不可靠的手工先验来约束网络, 从而给最终结果带来更大的误差。我们相信数据驱动的高质量先验的力量, 并努力提供可靠且一致的先验, 规避手动先验的限制。在本文中, 我们提出使用矢量量化码本 (VQCNIR) 的清晰夜间图像恢复来实现在现实世界和综合基准上获得显著且一致的恢复结果。为了确保细节和照明的忠实恢复, 我们建议结合两个基本模块: 自适应照明增强模块 (AIEM) 和可变形双向交叉注意 (DBCA) 模块。AIEM 利用特征的通道间相关性来动态维护降级特征和高质量码本特征之间的照明一致性。同时, DBCA 模块通过双向交叉注意力和可变形卷积有效地集成了纹理和结构信息, 从而增强了并行解码器的细粒度细节和结构保真度。大量实验验证了 VQCNIR 在增强图像质量方面的显著优势弱光条件下, 展示了其在合成和真实数据集上的最先进的性能。代码可在 <https://github.com/AlexZou14/VQCNIR> 获取。

Introduction

为了在夜景中获得可靠的图像, 通常使用长时间曝光以允许更多可用光线照亮图像。然而, 以这种方式捕获的图像仍然存在可见度低和颜色失真的问题。此外, 长时间曝光容易受到外部场景干扰, 例如相机抖动和动态场景, 这可能会导致图像中的运动模糊和噪声 (2022)。因此, 夜间图像经常表现出复杂的退化问题 (2022a; 2022; 2023), 例如低照度和模糊, 使得恢复

* 通讯作者。版权所有 © 2024, 人工智能促进协会 (www.aaai.org)。版权所有。

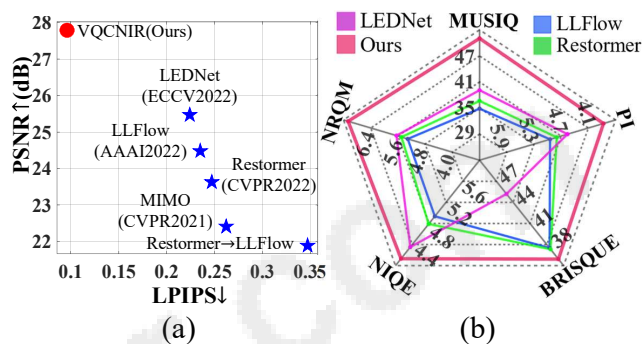


图 1: 与最先进方法的定量比较。(a) LOL-Blur 数据集上的 PSNR 和 LPIPS 结果。(b) Real-LOL-Blur 数据集上五个感知指标的结果。对于 PSNR、MUSIQ (2021) 和 NRQM(2017), 越高越好, 而对于 LPIPS(2018a)、NIQE (2012)、BRISQUE (2012) 和 PI (2018), 越低越好。

具有真实纹理和正常照明条件的高质量图像极具挑战性。随着深度学习方法 (2022; 2023d; 2023b; 2023c; 2023; 2023a) 在图像恢复方面的巨大成功, 许多基于深度学习的算法已经得到了应用。准备应对这一具有挑战性的任务。目前, 大多数研究人员只考虑夜间图像中的低光问题, 并提出了许多低光图像增强 (LLIE) 方法 (2017; 2018; 2019; 2020; 2021; 2019; 2021)。尽管这些 LLIE 方法可以产生视觉上令人愉悦的结果, 但它们的泛化能力在真实夜间场景中受到限制。这主要是由于 LLIE 方法主要关注增强图像亮度和减少噪声, 而忽略了模糊引起的空间退化, 从而导致清晰图像的无效恢复。一个直观的想法是将图像去模糊方法与 LLIE 方法相结合来解决这个问题。然而, 大多数现有的去模糊方法 (2021; 2021; 2019; 2022; 2022b) 都是在正常照明条件下捕获的数据集上进行训练的, 这使得它们不适合夜间图像去模糊。特别是, 由于夜间图像黑暗区域的可见度较差, 这些方法可能无法有效捕获运动模糊线索, 导致去模糊性能不理想。因此, 只需级联

LLIE 和去模糊方法不能产生令人满意的恢复结果。为了更好地处理低照度和模糊的联合退化过程, Zhou 等人。(2022) 首先提出了 LOL-Blur 数据集和称为 LEDNet 的端到端编码器-解码器网络。LEDNet 可以在合成的 LOL-Blur 数据集上实现高性能。然而, 其在真实场景中的泛化能力仍然有限。上述夜间恢复方法难以从低质量的夜间图像中恢复正确的纹理和可靠的照明。这是由于缺乏稳定可靠的先验, 因为大多数现有先验都是从低质量图像生成的。例如, 基于 Retinex 的技术 (2018; 2019; 2021) 通过分解低质量图像来进行照明估计, 而模糊内核则使用相同的降级输入进行估计。然而, 先验的偏差估计会导致最终结果出现累积误差。因此, 我们引入矢量量化 (VQ) 码本作为可信且可靠的外部特征库, 为纯粹数据驱动的图像恢复提供高质量的先验, 而不是依赖于 VQ 码本是由 aVQGAN (2021) 生成的隐式先验, 并在大量高保真干净图像的语料库上进行训练。因此, 训练有素的 VQ 码本可以为复杂退化图像提供全面、高质量的先验, 有效解决复杂退化问题。此外, 当直接使用码本进行重建时, 夜间图像的退化特征与 VQ 码本中的原始特征之间的不一致照明和不正确匹配会导致不令人满意的视觉效果。它甚至可能会放大模糊并在恢复的图像中产生伪影。因此, 利用码本先验来恢复夜间模糊图像的关键步骤在于精确对齐高质量的码本特征。在本文中, 我们提出了一种称为矢量量化码本的 ClearerNight 图像恢复

(VQCNIR) 的新方法, 用于夜间图像修复。为了解决上述关键考虑因素, 我们提出的 VQCNIR 包含两个专用模块。具体来说, 我们设计了自适应照明增强模块 (AIEM), 它利用特征的通道间相关性来估计曲线参数并自适应地增强特征中的照明。这有效地解决了退化特征和高质量 VQ 码本特征之间的照明不一致问题。为了改善退化特征和高质量特征之间的特征不匹配, 我们提出了一种集成可变形双向交叉注意 (DBCA) 的并行解码器。这种并行设计有效地结合了高质量的码本功能, 同时有效地融合了来自并行编码器的纹理和结构信息。我们提出的 DBCA 在高质量和低质量特征之间进行上下文建模, 自适应地融合它们以逐渐恢复精细细节, 从而提高整体质量。如图 1 所示, 我们的方法不仅在合成数据上实现了卓越的性能, 而且还可以很好地推广到现实世界的场景。对公开数据集的广泛实验表明

我们的方法在失真和感知指标上都超越了现有的最先进的方法。我们的主要贡献总结如下:

- 我们提出了 VQCNIR, 一种新框架, 通过引入高质量码本先验, 将夜间图像恢复制定为退化特征和高质量特征之间的匹配和融合问题。这解决了以前仅依赖于低质量输入的方法的局限性, 并实现了卓越的性能。
- 我们提出了一种自适应照明增强模块, 它利用通道间依赖性来估计曲线参数。这有效地解决了退化特征与高质量 VQ 码本特征之间的光照不一致问题。来自并行解码器的特征并恢复更准确的纹理细节。

相关工作

图像去模糊

深度学习技术的最新进展极大地影响了计算机视觉领域。针对单图像和视频去模糊任务, 人们提出了大量的深度学习方法 (2014; 2017; 2018; 2019; 2019; 2020; 2021), 并且表现出了优越的性能。随着去模糊任务的大型训练数据集的引入 (2009; 2017; 2019), 许多研究人员

(2009; 2019) 采用端到端网络来直接恢复清晰图像。尽管端到端方法优于传统方法, 但它们在严重模糊的情况下可能无效。为了提高网络性能, 一些方法 (2017; 2018; 2021) 使用多尺度架构来增强不同尺度的去模糊。然而, 这些方法在低光条件下 (特别是在黑暗区域) 捕获正确模糊线索的能力有限, 阻碍了它们在处理低光模糊图像方面的有效性。为了解决这个问题, 周等人。(2022) 引入夜间图像模糊数据集并开发端到端 UNet 架构, 该架构包含可学习的非线性层, 可有效增强黑暗区域, 而不会过度曝光其他区域。

低光图像增强

自第一项开创性工作 (2022 年) 以来, 近年来, 基于深度学习的微光图像增强 (LLIE) 取得了令人瞩目的成功。已经提出了许多端到端方法

(2017; 2019) 来使用编码器-解码器框架增强图像照明。为了进一步提高 LLIE 的性能, 研究人员受 Retinex 理论的启发, 开发了基于深度 Retinex 的方法 (2018; 2019; 2021), 该方法采用专用子网络来增强照度和反射率分量并实现更好的恢复性能。然而, 此类方法存在局限性, 因为增强结果强烈依赖于训练数据的特征。

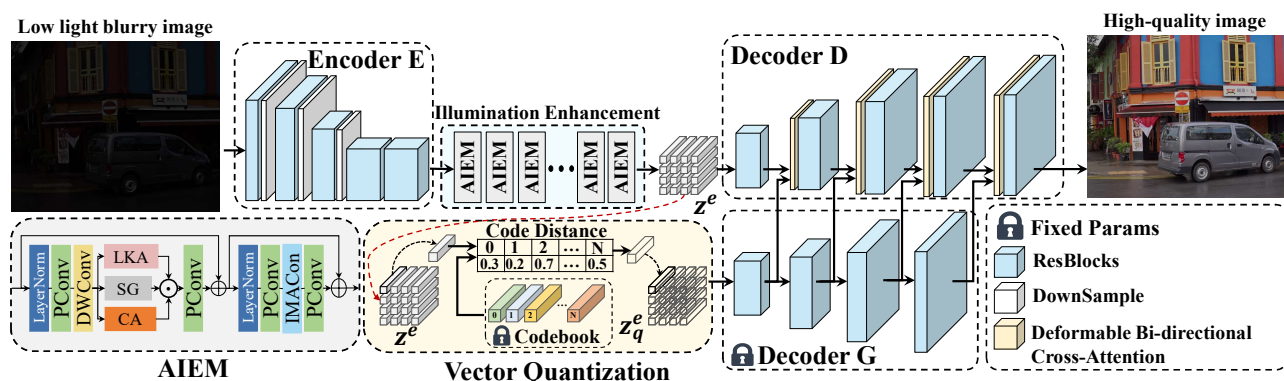


图 2: 拟议的 VQCNIR 框架。它由编码器、一些自适应照明增强模块 (AIEM) 和具有可变形双向交叉注意 (DBCA) 的并行解码器组成, 允许网络有效地利用高质量的码本先验信息。

为了提高网络的泛化能力, 研究人员 (2020; 2021; 2021) 提出了许多无监督方法。例如, Jiang 等人 (2021) 将自正则化和不配对训练引入 LLIE with EnlightenGAN。此外, Guo 等人 (2020) 提出了一种快速灵活的方法来估计图像增强深度曲线, 在训练过程中不需要任何法线照明参考图像。

Verctor-Quantized Codebook

VQVAE (2017) 首次将矢量量化 (VQ) 技术引入基于自动编码器的生成模型中, 以实现卓越的图像生成结果。具体来说, 编码的潜在变量在可学习的码本中被量化为它们最近的邻居, 并且所得的量化的潜在变量用于重建数据样本。在 VQVAE 的基础上, 后续工作对码本学习提出了各种改进。例如, VQGAN (2021) 利用生成对抗学习和精炼码本学习来进一步增强重建图像的感知质量。训练有素的码书可以作为高质量的先验, 可用于各种图像恢复任务, 例如图像超分辨率和人脸恢复。为此, 陈等人。 (2022a) 引入了用于盲图像超分辨率的 VQ 码本先验, 它将扭曲的 LR 图像特征与来自预训练的 HRprior 的无失真 HR 特征进行匹配。此外, Gu 等人。 (2022) 探索内部码本属性对重建性能的影响以及扩展离散码本技术对人脸图像恢复的影响。从这些作品中汲取灵感, 我们在夜间图像修复之前应用高质量的码本。

Methodology

框架概述

为了提高从包含复杂退化的夜间图像中恢复具有真实纹理和正常照明的高质量图像的能力, 我们引入了矢量量化码本作为高质量先验信息

设计夜间图像恢复网络 (VQCNIR)。VQCNIR 框架的概述如图 2 所示。VQCNIR 包括编码器 E、自适应照明增强模块、高质量码本 Z 和两个解码器 G 和 D。解码器 G 是来自 VQGAN 的具有固定参数的预训练解码器。解码器 D 代表主解码器, 它通过融合解码器 G 中的高质量特征来逐步恢复精细细节。

先验 VQ 码本

VQ Codebook: 我们首先简要描述 VQGAN (2021) 模型及其 codebook, 更多详细信息可以参考 (2021)。给定正常光下的高质量图像 $x_h \in \mathbb{R}^{H \times W \times 3}$, 编码器 E 将图像 x_h 映射到其空间潜在表示 $\hat{z} = E(x) \in \mathbb{R}^{h \times w \times n_z}$, 其中 n_z 是潜在的维度向量。然后, 通过逐元素量化过程 $q(\cdot)$, 将码本中的每个元素 $z_i \in \mathbb{R}^{n_z}$ 欧氏距离最近向量 z_k 找到为 VQ 表示 z_q 。如下所示:

$$z_q = \mathbf{q}(\hat{z}) := \left(\arg \min_{z_k \in \mathcal{Z}} \|\hat{z}_i - z_k\|_2^2 \right) \in \mathbb{R}^{h \times w \times n_z}, \quad (1)$$

其中码本是 $\mathcal{Z} = \{z_k\}_{k=1}^K \in \mathbb{R}^{K \times n_z}$, 具有 K 个离散码。然后, 解码器 G 将量化表示 z_q 映射回 sRGB 空间。总体重建过程可以表述如下:

$$\hat{x}_h = G(z_q) = G(\mathbf{q}(E(x))) \approx x_h, \quad (2)$$

用于夜间图像恢复的 VQ 码本: 为了充分探索 VQ 码本先验对夜间图像恢复的效果, 进行了一些初步实验来分析 VQGAN 的优缺点。首先, 我们使用训练有素的 VQGAN 来重建真实图像。实验结果如图 3 上方所示。从图中我们可以看到, VQGAN 可以在重建图像中生成生动的纹理细节。然而, 一些结构信息在矢量量化过程中丢失, 导致重建图像失真和伪影。因此, 重建仅取决于图像中的量化特征。

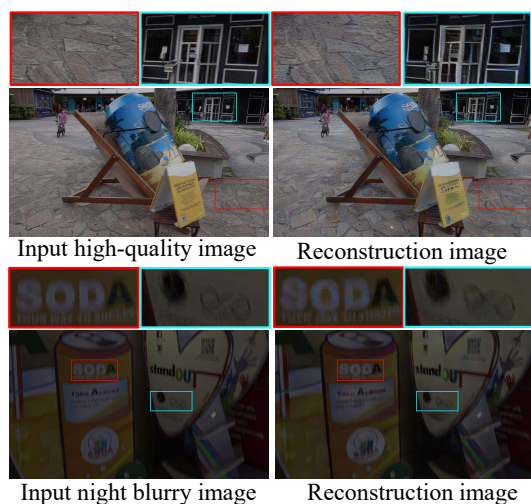


图 3: VQGAN 重建结果。左边是输入图像, 右边是重建图像。VQ-GAN 可以为高质量图像提供丰富的细节, 但会导致一些结构失真。在退化的图像中, 图像失真更加严重, 因为退化的特征与正确的高质量码本特征不匹配。

码本并没有产生令人满意的恢复结果。最直观的想法是将码本中的量化特征生成的纹理信息与潜在表示的结构信息相结合, 以避免图像的结构失真。随后, 我们探讨了 VQGAN 在高质量图像上进行训练, 用于重建夜间降级图像。如图3底部所示, 由于输入图像和 VQGAN 训练集的光照不一致, 恢复图像无法恢复到正常光照。此外, 我们发现 VQ-GAN 进一步恶化了模糊纹理并产生伪影。这是由于网络难以匹配正确的 VQ 码本特征, 导致 VQGAN 在高质量图像训练下无法从低照度和模糊中恢复。因此, 我们分别针对上述低光和模糊问题设计了自适应照明增强模块和可变形双向交叉注意力。

自适应照明增强

基于之前的观察和分析, 我们设计了自适应照明增强模块 (AIEM) 来解决量化特征与编码器获得的潜在特征之间的照明不一致问题, 如图2所示。该模块由两部分组成: 分层信息提取 (HIE) 和照明相互注意增强 (IMAE)。分层信息提取: 夜间经常观察到局部照明, 例如光源环境。然而, 全球运营往往会过度或不足地增强这些局部区域。因此, 我们采用通道注意力和大内核卷积注意力来提取 spa-

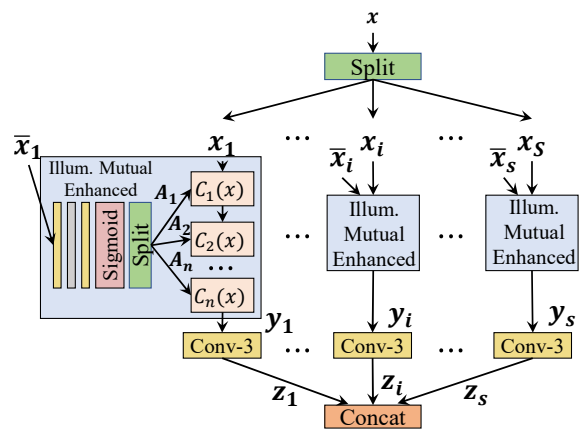


图 4: 照明相互注意卷积 (IMACConv) 的架构。

不同层次结构的原始信息。具体来说, HIE 首先采用层归一化来稳定训练, 然后进行不同感受野的空间信息融合。剩余捷径用于促进训练收敛。在归一化层之后, 使用逐点卷积和 3×3 深度卷积来捕获空间不变特征。然后, 使用三个并行算子来聚合通道和空间信息。第一个算子使用 SimpleGate (2022b) 对空间不变特征应用非线性激活。第二个算子是通道注意力 (2018b) 来调节特征通道。第三个是处理空间特征的大核卷积注意力 (2022)。三个分支输出相同大小的特征图。逐点乘法用于直接融合来自三个分支的不同特征。最后, 通过逐点卷积调整输出特征。光照相互注意增强: 根据从 HIE 获得的不同感受野的层次信息, IMAE 首先利用层归一化来稳定训练, 然后进行光照增强被应用于特征。具体来说, 我们设计了新颖的照明相互注意卷积 (IMACConv), 它利用特征通道之间的依赖关系来估计曲线参数, 从而调整特征的照明。使用两个逐点卷积来调整 IMACConv 的输入和输出特征。残差连接用于促进训练收敛。照明相互注意卷积: 考虑到特征通道之间的照明变化相似, 我们受 ZeroDCE (2020) 的启发, 引入曲线估计和通道相互映射, 提出照明相互注意卷积: 调整特征的像素范围来增强照明, 如图4所示。具体来说, 给定 $\text{IMACConv } x \cdot f \in \mathbb{R}^{C_{in} \times H \times W}$ 的输入特征 f , 我们首先沿着通道将 x 一次分成 S 部分, 如下所示:

$$x_f^1, x_f^2, \dots, x_f^S = \text{split}(x_f), \quad (3)$$

其中 $\text{split}(\cdot)$ 表示分割操作。对于每个部分 x 如果 ε

\in

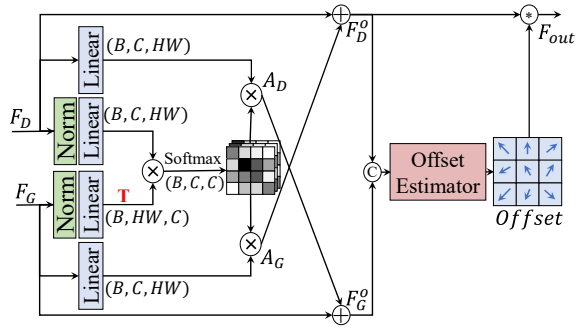


图 5: 可变形双向交叉注意力 (DBCA) 的架构。偏移估计器模块由许多大型内核卷积组成, 这些内核卷积使用来自大型感受野的信息来帮助融合两个解码器的特征。

$S \times H_f \times W_f$, 我们将排除 x_{if} 的通道特征连接在一起作为 x_i 的补充, 表示为 asx_{if} 。 x_{if} 和 x_i 都传入光照互增强, 通过曲线估计网络 F 估计多个曲线参数 $A = \{A_i\}_{i=1}^N$ 。 A_i 用于调整特征中像素值的范围。整个过程表述为:

$$A_1, A_2, \dots, A_n = \text{split}(\mathcal{F}(\bar{x}_i)), \quad (4)$$

$$y_f^i = C_n(x_f^i, A_1, A_2, \dots, A_n), \quad (5)$$

其中 $F(\cdot)$ 和 $C_n(\cdot)$ 表示曲线估计网络和高阶曲线映射函数。曲线估计网络 F 由三个内核大小分别为 5、3 和 1 的卷积层、两个激活函数和一个 sigmoid 函数组成。对于高阶曲线映射函数 C_n , 可以表示为:

$$C_n(x_f^i) = \begin{cases} A_1 x_f^i (1 - x_f^i) + x_f^i, n = 1 \\ A_n C_{n-1}(x_f^i) (1 - C_{n-1}(x_f^i)) + C_{n-1}(x_f^i), n > 1 \end{cases}$$

(6) 光照增强后, 对于所有 $y_{1f}, y_{2f}, \dots, y_{Sf}$, 我们使用 3×3 卷积层生成特征 $z_{if} = \text{Conv}_{we}(y_{if})$ 。最后, 不同的特征 $z_{1f}, z_{2f}, \dots, z_{Sf}$ 连接起来形成 IMACConv 的输出。

$$z_f = \text{Concat}(z_{f1}^1, z_{f1}^2, \dots, z_{f1}^S), \quad (7)$$

可变形双向交叉注意力

如前所述和分析, 从码本获得的高质量量化特征并非完美无缺。结构扭曲和纹理失真导致高质量 VQ 码本特征和原始降级特征之间更严重的失准。因此, 我们提出了可变形双向交叉注意 (DBCA) 来融合高质量的 VQ 码本特征和降级特征。与传统的交叉注意方法 (2021) 不同, 我们的 DBCA 旨在使用

双向交叉注意机制, 并采用可变形卷积来有效纠正退化特征中的模糊退化。如图 5 所示, 给定输入解码器 D 和 G 特征 F_D 和 F_G , 它们首先通过归一化和线性层映射到相应的 $Q_D = W_p D \text{LN}(F_D)$ 和 $Q_G = W_p G \text{LN}(F_G)$ 。我们进一步利用线性层将这些特征映射到相应的值 V_D 和 V_G 。我们将上述 Q_D 、 Q_G 、 V_D 和 V_G 重塑为 $(B, C, H \times W)$ 的形状, 并使用以下双向交叉注意力公式融合这两个特征: $A_D = \text{Softmax}(Q_D Q_G^T /$

$\sqrt{C}) V_G$, (9) $F_{oD} = \gamma_D A_D G + F_D$, (10) $F_{oG} = \gamma_G A_G D + F_G$, (11) 其中 $\text{Softmax}(\cdot)$ 表示 softmax 函数。 A_D 和 A_G 分别表示特征 D 和特征 G 的注意力图。 γ_D 和 γ_G 是可训练的通道尺度, 并用零初始化以稳定训练。为了更好地将高质量码本先验特征融合到退化特征中, 我们首先生成一个通过连接两个输出特征来抵消。然后, 我们使用可变形卷积中生成的偏移来扭曲纹理特征并有效地去除模糊退化, 其可以形式化如下: $\text{offset} = \text{LKConv}(\text{Concat}(F_{oD}, F_{oG}))$, (12) $F_{out} = \text{DeformConv}(F_{oD}, \text{offset})$, (13) 其中 $\text{LKConv}(\cdot)$ 和 $\text{DeformConv}(\cdot)$ 表示 7×7 卷积和可变形分别进行卷积。

VQCNIR 的培训目标

VQCNIR 的训练目标包括四个部分: (1) 像素重建损失 L_{pix} , 最小化输出与地面实况之间的距离; (2) 代码对齐损失 L_{ca} 强制夜间图像的代码与相应的地面实况对齐; (3) 感知损失 L_{per} , 在特征空间中操作, 旨在增强恢复图像的感知质量; (4) 用于恢复真实纹理的对抗性损失 L_{adv} 。具体地, 我们采用像素域中常用的 L1 损失作为重建损失, 表示为: $L_{pix} = \|x_h - VQCNIR(x_n)\|_1$, (14) 其中 x_h 和 x_n 分别表示高质量地面实况图像和夜间图像。提高夜间图像代码与高质量图像代码的匹配性能。我们采用 L2 损失来衡量距离, 可以表示为: $L_{ca} = \|z_e - z_{eq}\|_2$, (15) 其中 z_e 和 z_{eq} 分别是夜间图像代码和地面实况代码。总的训练目标是上述损失的组合: $L_{VQCNIR} = \lambda_{pix} L_{pix} + \lambda_{ca} L_{ca} + \lambda_{per} L_{per} + \lambda_{adv} L_{adv}$, (16) 其中 λ_{pix} 、 λ_{ca} 、 λ_{per} 和 λ_{adv} 分别表示每个损失函数的比例因子。

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Zero-DCE \rightarrow MIMO	17.68	0.542	0.510
LLFlow \rightarrow Restormer	21.50	0.746	0.357
LLFlow \rightarrow Uformer	21.51	0.750	0.350
MIMO \rightarrow Zero-DCE	17.52	0.570	0.498
Restormer \rightarrow LLFlow	21.89	0.772	0.347
Uformer \rightarrow LLFlow	21.63	0.758	0.342
KinD++*	21.26	0.753	0.359
DeblurGAN-v2*	22.30	0.745	0.356
DMPHN*	22.20	0.817	0.301
MIMO*	22.41	0.835	0.262
Restormer*	23.63	0.841	0.247
LLFlow*	24.48	0.846	0.235
LEDNet*	25.74	0.850	0.224
Ours	27.79	0.875	0.096

表1: LOL-Blur数据集上的定量评估。符号*表示网络在LOL-Blur数据集上重新训练。最佳值和次佳值分别用粗体文本和下划线文本表示。

Experiments

数据集和训练细节

我们在 LOL-Blur 数据集 (2022) 上训练 VQCNIR 网络, 该数据集由 170 个序列 (10,200 对) 训练数据和 30 个序列 (1,800 对) 测试数据组成。我们对增强训练数据使用 90、180、270 的随机旋转、随机翻转和随机裁剪到 256×256 大小。我们使用 Adam (2014) 优化器训练我们的网络, $\beta_1=0.9$, $\beta_2=0.99$, 总共 500k 次迭代。小批量大小设置为 8。初始学习率设置为 1×10^{-4} 并采用 MultiStepLR 逐步调整学习率。我们凭经验将 λ_{pix} 、 λ_{ca} 、 λ_{per} 和 λ_{adv} 设置为 $\{1, 1, 1, 0.1\}$ 。所有实验均在配备 Intel Core i7-13700K CPU、32G RAM 和带有 CUDA 11.2 的 Nvidia RTX 3090 GPU 的 PC 上进行。

LOL-Blur 数据集的结果

在本节中, 我们将我们提出的 VQCNIR 与 LOL-Blur 测试集 (2022) 上的所有上述方法进行定量和定性比较。我们使用两个最广泛评估的指标: PSNR 和 SSIM 对所有方法进行公平评估。此外, 我们采用 LPIPS 指标来评估恢复图像的感知质量。定量评估。表 1 显示了我们的方法和其他方法在 LOL-Blur 数据集上的定量结果。如图所示, 我们的方法在 PSNR 和 SSIM 方面分别优于最先进的 LEDNet 2.05 dB 和 0.025。此外, 当通过感知质量指标进行评估时, 我们的方法比现有方法表现出明显的优势。我们的方法的有效性已得到充分证明。定性评估。图6显示了所有比较方法的视觉效果。如图所示, 大多数方法无法有效消除严重模糊区域的模糊效果, 不可避免地会在恢复的图像中引入伪影。相比之下, 我们的方法可以有效地恢复

Method	MUSIQ \uparrow	NRQM \uparrow	NIQE \downarrow
RUAS \rightarrow MIMO	34.39	3.322	6.812
LLFlow \rightarrow Restormer	34.45	5.341	4.803
LLFlow \rightarrow Uformer	34.32	5.403	4.941
MIMO \rightarrow Zero-DCE	28.36	3.697	6.892
Restormer \rightarrow LLFlow	35.42	5.011	4.982
Uformer \rightarrow LLFlow	34.89	4.933	5.238
KinD++*	31.74	3.854	7.299
DMPHN*	35.08	4.470	5.910
MIMO*	35.37	5.140	5.910
Restormer*	36.65	5.497	5.093
LLFlow*	34.87	5.312	5.202
LEDNet	39.11	5.643	4.764
Ours	51.04	7.064	4.599

表 2: 对 Real-LOL-Blur 数据集的定量评估。符号*表示网络在 LOL-Blur 数据集上重新训练。最佳值和次佳值分别用粗体文本和下划线文本表示。

通过使用高质量的先验信息来获得正确的纹理特征。因此, 这些结果提供了充分的证据表明我们的方法提出的先验码本特别适合夜间图像恢复的任务。

真实数据集的结果

为了更好地说明我们的方法在真实场景中的有效性, 我们在真实的 Real-LOL-Blur 数据集 (2022) 下将我们提出的 VQCNIR 与上述方法进行定量和定性比较。由于真实场景缺乏相应的参考图像进行评估, 因此使用了三种非参考评估指标进行评估: MUSIQ (2021)、NRQM (2017) 和 NIQE (2012)。MUSIQ 指标主要评估色彩对比度和清晰度, 更适合此任务。定量评估。表 2 显示了我们的方法和其他方法在 Real-LOL-Blur 测试集上的定量结果。如表2所示, 我们的方法获得了最高的 NIQE和NRQM分数, 表明我们的方法的恢复结果具有更好的图像质量并且与人类感知一致。而且, 我们拥有最高的MUSIQ, 这意味着我们的结果在色彩对比度和清晰度方面是最好的。定性评估。图 7 显示了所有评估方法的视觉比较结果。从图中可以明显看出, 简单的级联去模糊和低光增强技术可能会导致图像中的过度曝光和饱和区域模糊等问题。即使是在 LOL-Blur 数据集上进行端到端再训练的方法也会出现不良的严重伪影和模糊问题。相比之下, 我们提出的 VQCNIR 在视觉质量方面优于这些方法, 显示出更少的伪影和模糊。这一改进可归因于将高质量码本预先成功集成到网络中, 这有助于生成高质量纹理。真实世界图像的比较结果进一步证明了我们提出的方法的优越性。

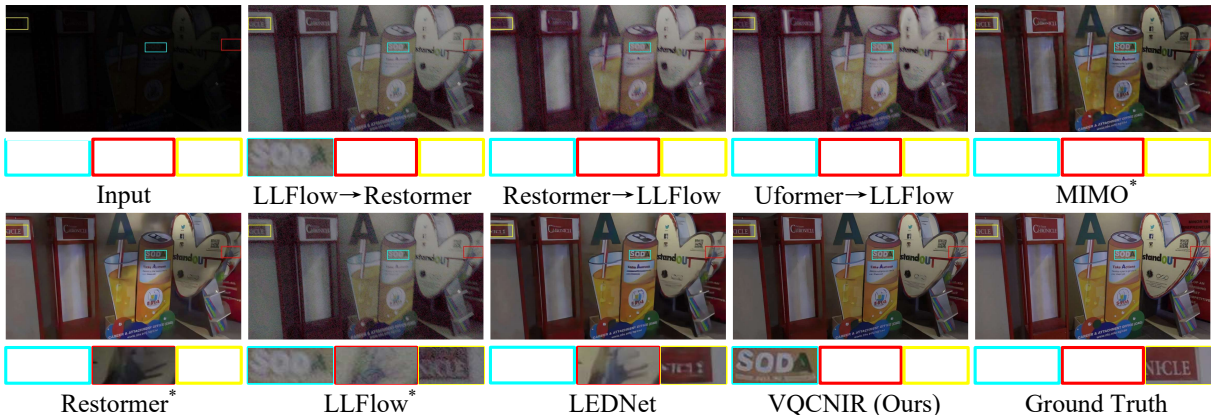


图 6: LOL-Blur 数据集 (2022) 的视觉比较结果。符号 * 表示网络在 LOL-Blur 数据集上重新训练。所提出的方法产生视觉上更令人愉悦的结果。（放大以获得最佳视野）

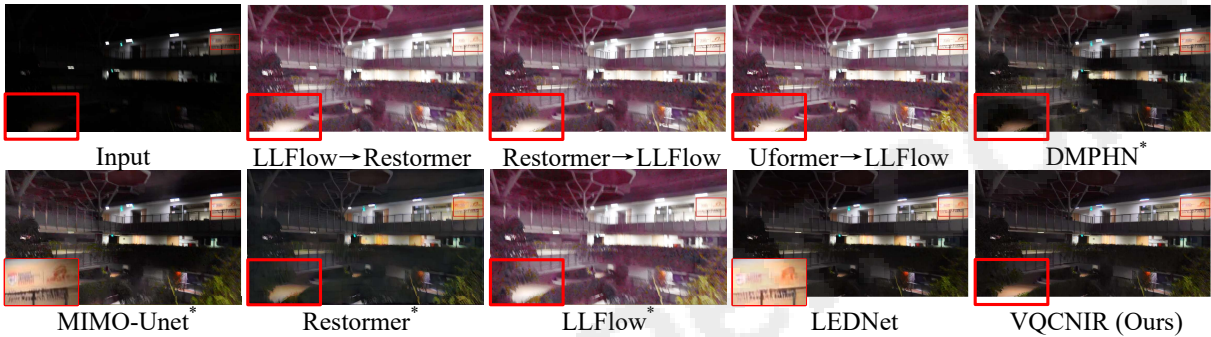


图 7: Real-LOL-Blur 数据集 (2022) 的视觉比较。符号 * 表示网络在 LOL-Blur 数据集上重新训练。所提出的方法产生视觉上更令人愉悦的结果。（放大以获得最佳视野）

Models	Configuration			LOL-Blur	
	Decoder D	AIEM	DBCA	PSNR	SSIM
VQGAN				10.79	0.3028
Setting 1	✓			26.58	0.8486
Setting 2	✓		✓	26.89	0.8599
Setting 3	✓	✓		27.48	0.8692
VQCNIR	✓	✓	✓	27.79	0.8750

表 3: 不同成分的消融研究。我们报告了 LOL-Blur 数据集上的 PSNR 和 SSIM 值。

消融研究

在本节中，我们实施了一系列消融实验，以更好地验证我们提出的每个模块的有效性。为了验证我们提出的操作的有效性，提出了一系列消融实验，结果如表 3 所示。最初，我们使用 VQGAN 作为基线模型。表 3 显示，VQ-GAN 不能有效解决低光和模糊退化问题，因为 VQGAN 是先从高质量自然图像中学习的码本，无法正确匹配退化特征。通过设计相应的并行解码器，网络可以有效地利用高质量的优先级

或协助重建退化的特征。然而，退化特征和码本先验之间的照明不一致会妨碍高质量先验特征的准确匹配，从而导致伪影的出现。此外，降级的特征与高质量的特征还有一定距离。因此，利用AIEM和DBCA可以有效提高网络性能和图像质量。

Conclusion

在这项工作中，我们引入了高质量的码本先验，并提出了一种称为 VQCNIR 的夜间图像恢复新范例。通过分析，我们发现直接应用码本先验会导致退化特征与高质量码本特征之间的不正确匹配。为了解决这个问题，我们提出了自适应照明增强模块（AIEM）和可变形双向交叉注意（DBCA）模块，利用估计的照明曲线和双向交叉注意。通过融合码本先验和降级特征，VQCNIR 有效地从夜间图像中恢复正常照明和纹理细节。大量的实验证明了我们方法的最先进的性能。

Acknowledgments

该工作得到广州市科技计划项目202103010003、佛山市重点领域科技计划项目2020001006285的资助。

References

- 布劳, Y.; 梅赫雷斯, R.; 蒂莫特, R.; 米凯利, T.; 和 Zelnik-Manor, L. 2018. 2018 年 PIRM 感知图像超分辨率挑战。欧洲计算机视觉会议 (ECCV) 研讨会论文集, 0-0. Chen, C.; 施X.; 秦, Y. 李, X. 韩X.; 杨, T. 和 Guo, S. 2022a. 通过与隐式高分辨率先验进行特征匹配来实现现实世界的盲超分辨率。第 30 届 ACM 国际多媒体会议论文集, 1329-1338. Chen, C.-F. R.; 范, Q. Panda, R. 2021. Crossvit: 用于图像分类的交叉注意多尺度视觉变换器。IEEE/CVF 国际计算机视觉会议论文集, 357-366. Chen, L.; 楚, X. 张X.; 和 Sun, J. 2022b. 图像恢复的简单基线。计算机视觉 - ECCV2022: 第 17 届欧洲会议, 以色列特拉维夫, 2022 年 10 月 23-27 日, 会议记录, 第七部分, 17-33. 施普林格, Chen, L.; 张, J. 林, S. 方, F. 和 Ren, J. S. 2021. 饱和图像的盲去模糊。IEEE/CVF 计算机视觉和模式识别会议论文集, 6308-6316. Chen, S.; 然而.; 白, J. 陈, E. 石, J.; 和朱 L. 2023a. 具有不确定性驱动排序的稀疏采样变压器, 用于统一去除雨滴和雨条纹。IEEE/CVF 国际计算机视觉会议论文集, 13106-13117. Chen, S.; 然而.; 刘, Y. 白, J. 陈, H. 林, Y. Shi, J.; 和 Chen, E. 2023b. CPLFormer: 用于图像除雪的跨尺度原型学习变压器。第 31 届 ACM 国际多媒体会议论文集, 4228-4239. 然而.; 刘, Y. 廖, T.; 蒋, J. 陈, E. 和 Chen, P. 2023c. MSP-former: 用于单图像除雪的多尺度投影变换器。ICASSP 2023-2023 IEEE 国际声学、语音和信号处理会议 (ICASSP), 1-5. IEEE. Chen, S.; 然而.; 薛, C.; 陈, H. 刘, Y. 陈, E. 和 Zhu, L. 2023d. 用于高效单图像除雪的不确定性驱动的动态退化感知和背景建模。第 31 届 ACM 国际多媒体会议论文集, 4269-4280. Cho, S.-J.; 吉, S.-W.; 洪, J.-P.; 荣格, S.-W.; 和 Ko, S.-J. 2021. 重新思考单图像去模糊中从粗到细的方法。IEEE/CVF 国际计算机视觉会议论文集, 4641-4650. Esser, P.; 罗姆巴赫, R.; 和 Ommer, B. 2021. 驯服变压器以进行高分辨率图像合成。IEEE/CVF 计算机视觉和模式识别会议论文集, 12873-12883. 顾, Y. 王X.; 谢L.; 董, C.; 李, G. Shan, Y.; 和 Cheng, M.-M. 2022. Vqfr: 使用矢量量化字典和并行解码器进行盲脸恢复。计算机视觉 - ECCV 2022: 第 17 届欧洲会议, 以色列特拉维夫, 2022 年 10 月 23-27 日, 会议记录, 第 XVIII 部分, 126-143. 施普林格, 郭, C.; 李, C. 郭, J. 洛伊, C.C.; 侯, J. Kwong, S.; and Cong, R. 2020. 低光图像增强的零参考深度曲线估计。IEEE/CVF 计算机视觉和模式识别会议论文集, 1780-1789. Guo, M.-H.; 卢, C.-Z.; 刘, Z.-N.; 郑, M.-M.; 和胡, S.-M. 2022. 视觉注意网络。arXiv preprint arXiv:2202.09741. Hu, Z.; 曹, S. 王, J. 和 杨, M.-H. 2014. 用光条纹去模糊低光图像。IEEE 计算机视觉和模式识别会议论文集, 3382-3389. 龚X.; 刘, D.; 程, Y. 方, C.; 沉X.; 杨J.; 周, P. 和 Wang, Z. 2021. Enlightengan: 无需配对监督的 Deeplight 增强。IEEE 图像处理交易, 30: 2340-2349. Ke, J.; 王Q. 王, Y. 米兰法尔, P.; 和 Yang, F. 2021. Musiq: 多尺度图像质量转换器。IEEE/CVF 国际计算机视觉会议论文集, 5148-5157. Kingma, D.; Ba, J. 2014. Adam: 随机优化方法。计算机科学。Kupyn, O.; 马蒂纽克, T.; 吴, J. 和 Wang, Z. 2019. Deblurgan-v2: 更快更好地去模糊 (数量级)。IEEE/CVF 国际计算机视觉会议记录, 8878-8887. Levin, A.; 韦斯, Y.; 杜兰德, F.; 和 Freeman, W. T. 2009. 理解和评估盲解卷积算法。2009 年 IEEE 计算机视觉和模式识别会议, 1964-1971. IEEE. Li, C.; 郭, C.; 汉, L.; 蒋, J. 郑, M.-M.; 顾, J. and Loy, C. C. 2022. 使用深度学习增强低光图像和视频: 一项调查。IEEE 模式分析和机器智能汇刊, 44(12): 9396-9416. Li, C.; 郭, C.; 和 Loy, C. C. 2021. 通过零参考深度曲线估计学习增强低光图像。IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(8): 4225-4238. Liu, Y.; 严, Z. 陈, S. 然而.; 任, W.; 和陈, E. 2023. Nighthazeformer: 使用先前的查询变压器去除单个夜间雾霾。第 31 届 ACM 国际多媒体会议论文集, 4119-4128. 严, Z. 吴, A. 然而.; Li, Y. 2022. 基于变分分解模型的夜间图像去雾。IEEE/CVF 计算机视觉和模式识别会议论文集, 640-649. Lore, K. G.; 阿金塔约, A.; Sarkar, S. 2017. LLNet: 用于自然低光图像增强的深度自动编码器方法。模式识别, 61: 650-662.

苹果。;杨, C.-Y.; 杨X.; 和杨, M.-H. 2017.学习单图像超分辨率的无参考质量度量。计算机视觉和图像理解, 158: 1-16。米塔尔, A.; 穆尔西, A.K.; 和 Bovik, A. C. 2012. 空间域中的无参考图像质量评估。IEEE 图像处理汇刊, 21(12): 4695-4708。Mittal, A.; 桑德拉拉詹, R.; 和 Bovik, A. C. 2012. 制作“完全盲”图像质量分析器。IEEE 信号处理快报, 20(3): 209-212。Nah, S.; 炫金, T.; 和 Mu Lee, K. 2017. 用于动态场景去模糊的深度多尺度卷积神经网络。IEEE 计算机视觉和模式识别会议论文集, 3883-3891。Shen, Z.; 王, W. 卢X. 沉, J. 凌, H.; Xu, T.; 和 Shao, L. 2019. 人类感知运动去模糊。IEEE/CVF 国际计算机视觉会议论文集, 5572-5581。高, H. 沉X. 王, J. 和 Jia, J. 2018. 用于深度图像去模糊的尺度循环网络。IEEE 计算机视觉和模式识别会议论文集, 8174-8182。Van Den Oord, A.; 维尼亚尔斯, O.; 等人。2017. 神经离散表示学习。神经信息处理系统的进展, 30。Wang, R.; 张, Q. 傅, C.-W.; 沉X. 郑W.-S.; and Jia, J. 2019. 使用深度照明估计的曝光不足照片增强。IEEE/CVF 计算机视觉和模式识别会议论文集, 6849-6857。万, R. 杨, W. 李, H. 周, L.-P.; 和科特, A. 2022a. 通过归一化流进行低光图像增强。AAAI 人工智能会议记录, 第 36 卷, 2604-2612。存X. 鲍, J.; 周, W. 刘, J. 和 Li, H. 2022b. Uformer: 用于图像恢复的通用U形变压器。IEEE/CVF 计算机视觉和模式识别会议记录, 17683-17693。Wei, C.; 王, W. 杨, W. 和 Liu, J. 2018. 用于低光增强的深度视网膜分解。arXiv preprint arXiv: 1808.04560。Ye, T.; 陈, S. 刘, Y. 柴, W.; 白, J. 邹, W.; 张, Y. 江, M. 陈, E. 和薛, C. 2023. 视频恢复的顺序亲和学习。第 31 届 ACM 国际多媒体会议论文集, 4147-4156。张, Y. 江, M. 陈L.; 刘, Y. Chen, S.; 和 Chen, E. 2022. 图像去雾的密度感知和建模。欧洲计算机视觉会议, 130-145。施普林格·扎米尔, S.W.; 阿罗拉, A.; 汗, S.; 哈亚特, M.; Khan, F.S.; 和 Yang, M.-H. 2022. Restormer: 用于高分辨率图像恢复的高效变压器。IEEE/CVF 计算机视觉和模式识别会议论文集, 5728-5739。

张, H. 戴, Y. 李, H. 和 Koniusz, P. 2019. 用于图像去模糊环的 Deepstacked 分层多补丁网络。IEEE/CVF 计算机视觉和模式识别会议论文集, 5978-5986。罗, W.; 钟, Y. 马, L. 斯坦格, B.; Liu, W.; 和 Li, H. 2020. 通过真实模糊进行去模糊。IEEE/CVF 计算机视觉和模式识别会议论文集, 2737-2746。伊索拉, P.; 埃夫罗斯, A.A.; 谢赫特曼, E.; 和王, O. 2018a. 深度特征作为感知度量的不合理有效性。IEEE 计算机视觉和模式识别会议论文集, 586-595。郭X. 马, J. 刘, W.; 和 Zhang, J. 2021. 超越增亮低光图像。国际计算机视觉杂志, 129: 1013-1037。李, K. 李, K. 王L.; 钟B.; 和傅, Y. 2018b. 使用非常深的残差通道注意力网络进行图像超分辨率。欧洲计算机视觉会议记录 (ECCV), 286-301。张, J. 和 Guo, X. 2019. 点燃黑暗: 一种实用的低光图像增强器。在第 27 届 ACM 国际多媒体会议记录中, 1632-1640 年。Zheng, C.; 石, D. 和 Shi, W. 2021. 用于低光图像增强的自适应展开全变分网络。IEEE/CVF 国际计算机视觉会议论文集, 4439-4448。Zhou, S.; 李, C. 和 Change Loy, C. 2022. Lednet: 联合低光增强和黑暗中去模糊。计算机视觉 - ECCV 2022: 第 17 届欧洲会议, 以色列特拉维夫, 2022 年 10 月 23-27 日, 会议记录, 第六部分, 573-589。施普林格。