

RIDCP: 通过高质量码本先验重振真实图像去雾

Anonymous CVPR submission

Paper ID 03324

Abstract

现有的去雾方法由于缺乏配对的真实数据和稳健的先验知识，难以处理现实世界中的雾霾图像。在这项工作中，我们从合成更真实的雾霾数据和在网络中引入更稳健的先验知识两个角度，提出了一种新的真实图像去雾范式。具体而言，(1) 我们重新思考了真实雾霾图像的退化过程，提出了一种考虑多种退化类型的现象学管道，而不是采用事实上的物理散射模型。(2) 我们提出了一种基于高质量码本先验的真实图像去雾网络 (RIDCP)。首先，VQGAN 在大规模高质量数据集上进行预训练，获得离散的码本，封装高质量的先验 (HQP)。通过使用 HQP 替代雾霾带来的负面影响后，包含新型归一化特征对齐模块的解码器可以有效利用高质量的特征，生成干净的结果。尽管我们的退化管道大大减少了合成数据和真实数据之间的域间差距，但要完全避免这种差距仍然是困难的，这也对在自然环境下 HQP 的匹配带来了挑战。因此，我们通过一种可控匹配操作重新计算特征在与 HQP 进行匹配时的距离，促使其找到更好的对应项。我们提供了一种基于可解释方案的匹配控制建议，用户还可以根据自己的喜好灵活调整增强程度。大量实验验证了我们数据合成管道的有效性以及 RIDCP 在真实图像去雾中的卓越性能。代码和数据将会公开发布。

1. 引言

图像去雾旨在从雾霾图像中恢复干净的图像，这对计算摄影和高层次任务来说非常重要 [20, 32]。雾霾图像的公式通常通过物理散射模型描述为：

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad (1)$$

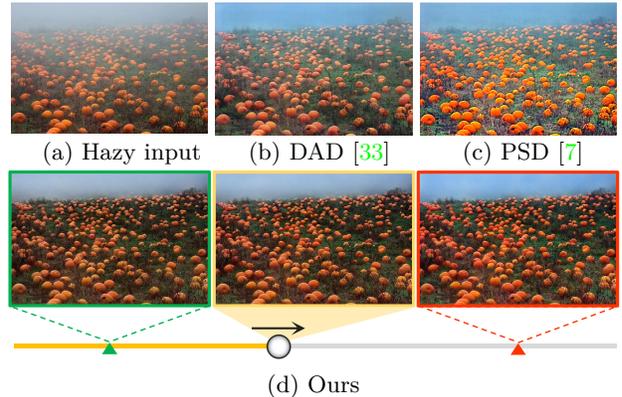


图 1. 对典型雾霾图像的视觉比较。所提出的方法生成的结果比另外两种最先进的真实图像去雾方法更干净。我们的结果的增强程度可以通过在真实域自适应阶段采用不同的参数灵活调整。带有金色边框的图像是根据我们推荐的参数获得的结果。

其中， $I(x)$ 表示雾霾图像， $J(x)$ 是其对应的清晰图像。变量 A 和 $t(x)$ 分别是全球大气光和透射图。透射图 $t(x) = e^{\beta d(x)}$ 取决于场景深度 $d(x)$ 和雾霾密度系数 β 。

对于给定的雾霾图像，恢复其清晰版本是一个高度不适定问题。为缓解这一问题的不适定性，现有的传统方法提出了各种先验知识，例如暗通道先验 [16]、颜色衰减先验 [45] 和颜色线条 [12] 等。然而，这些统计先验无法涵盖现实场景中的多样情况，导致去雾效果不理想。

随着深度学习的兴起，图像去雾取得了显著进展。现有方法要么采用深度网络来估计物理参数 [5, 21, 31]，要么直接恢复无雾图像 [10, 15, 27, 30, 40]。然而，由于难以收集大规模且完美对齐的成对训练数据，以及在没有稳健先验的情况下解决不适定问题的不确定性，图像去雾神经网络在真实场景中的泛化能力有限。具体来说，1) 收集大规模且完美对齐的雾霾图像及其对应

的清晰图像是极其困难的，甚至可以说是不可能的。因此，大多数现有的深度模型使用合成数据进行训练，其中雾霾图像通过公式 (1) 生成，这导致了对多个退化因素的忽视。确实存在一些具有配对数据的真实雾霾图像数据集 [2,3]，但其规模和多样性不足。此外，这些数据集与在真实场景中捕捉到的雾霾图像存在偏差。这些缺陷不可避免地降低了深度模型在真实场景中的能力。2) 真实图像去雾是一个高度不适定的问题。通常，解决一个不确定的映射问题往往需要先验知识的支持。然而，获得能够涵盖真实雾霾图像多样场景的稳健先验是困难的，这也限制了去雾算法的性能。最近，许多针对真实图像去雾的研究试图通过数据生成的领域自适应 [33,39] 或先验引导 [7,23] 的角度来解决这两个问题，但仍然无法获得理想的结果。

在这项工作中，我们提出了一种新的真实图像去雾范式，旨在解决上述两个问题。为了获得大规模且完美对齐的成对训练数据，我们通过观察大量真实的雾霾图像，重新思考雾霾图像的退化过程，并提出了一种考虑多种退化因素的新数据生成流程。为了解决不定问题的不确定性，我们尝试在高质量图像上训练 VQGAN [11]，以提取更加稳健的高质量先验 (HQPs)。VQGAN 仅学习高质量图像的重建，因此它自然包含稳健的 HQPs，可以帮助雾霾特征跳跃到清晰领域。第 4.1 节中的观察进一步验证了我们的动机。因此，我们提出了基于高质量码本先验的真实图像去雾网络 (RIDCP)。其中码本和 VQGAN 的解码器用来提供 HQPs。RIDCP 还配备了一个编码器，用于帮助找到正确的 HQPs，和一个新的解码器用于利用来自固定解码器的特征，从而生成最终结果。此外，我们提出了一种新型的归一化特征对齐 (NFA) 方法，可以减轻失真并平衡特征，从而实现更好的融合。

与先前引入码本进行图像修复的方法相比 [6,14,43]，我们进一步基于 VQGAN 的特性和统计结果设计了一种独特的真实域自适应策略。直观上，我们提出了一种可控高质量先验匹配 (CHM) 操作，该操作通过在推理阶段对特征和 HQPs 之间的距离施加精心设计的权重，来取代最近邻匹配。权重由可控参数 α 和第 4.3 节中的 HQPs 激活的统计分布差异决定。通过调整 α ，可以改变 HQPs 激活的分布。此外，我们提出了一种理论上可行的解决方案，通过最小化两个概率分布的 Kullback-Leibler 散度来获得最佳的 α 值。更重

要的是， α 的值可以通过图像的增强程度在视觉上反映出来，如图 1(d) 所示，用户可以根据自己的偏好调整去雾结果。我们的 CHM 方法有效、灵活且具有可解释性。

与最先进的真实图像去雾方法相比，例如 DAD [33] 和 PSD [7]，只有我们提出的 RIDCP 能够有效处理在自然环境中拍摄的雾霾图像，并生成可调节的结果，如图 1 所示。我们工作的贡献可以总结如下：

- 我们提出了一种新范式，以推动基于深度学习的图像去雾技术向真实场景发展。
- 我们是首个在真实图像去雾任务中利用高质量码本先验的研究。提出的可控 HQPs 匹配操作旨在克服合成域和真实域之间的差距，并生成可调节的结果。
- 我们重新制定了真实雾霾图像的退化模型，并提出了一种现象学退化管道，以模拟在真实环境中捕获的雾霾图像。

2. 相关工作

2.1. 单幅图像去雾

图像去雾。早期的单幅图像去雾尝试通过对无雾图像的先验知识 [4,12,16,35,45]，来估计大气散射模型 (如公式 (1) 中所示) 的参数。这些方法取得了显著的效果。然而，基于经验观察的手工设计先验在多样化场景中表现不佳。例如，暗通道先验 (DCP) [16] 在天空区域的假设并不适用。我们提出的方法通过在大规模数据集上预训练离散码本，获得了高质量图像的先验，这更为可靠且全面。

随着深度学习技术的发展，如何利用数据驱动的理念去除雾霾引起了广泛关注。在早期阶段，许多研究 [5,21,31] 尝试采用卷积神经网络 (CNNs) 来估计公式 (1) 中退化模型的参数。此外，为了避免参数估计中的累积误差，一些端到端的网络 [10,15,27,30,40] 被提出用于直接估计无雾图像。上述基于学习的方法在合成数据集上取得了优异的性能。然而，它们在真实世界数据上表现显著下降的问题亟待解决。

真实图像去雾。最近，一些工作开始关注真实图像去雾问题。一类研究方向是利用 GANs [13] 来生成符合真实雾霾域的雾霾数据。Shao 等人 [33] 基于

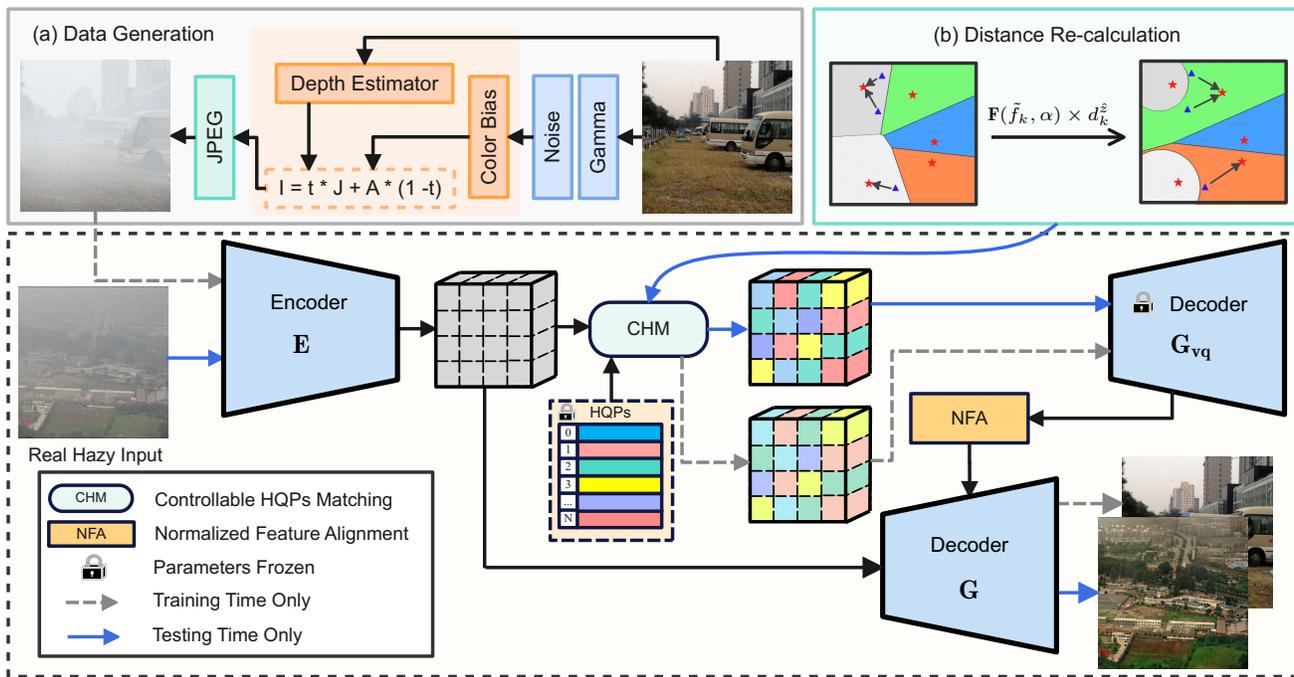


图 2. RIDCP 的概述。在训练阶段，我们在由数据生成流程合成的数据上训练去雾网络，如图 (a) 所示。该网络基于预训练的 HQPs 码本和 VQGAN 的相应解码器 G_{vq} 。我们还设计了可控 HQPs 匹配 (CHM) 操作，通过重新计算特征与 HQPs 之间的距离 $d_k^z = \|\hat{z} - z_k\|$ ，来实现真实域适应。图 (b) 展示了通过两个 Voronoi 图进行距离重新计算的过程，彩色区域表示属于更优的 HQPs，而灰色区域则相反。三角形代表特征点，星形点代表 HQPs。可以看到，在距离重新计算后，原本属于灰色区域的点被我们的 CHM 强制分配到了彩色区域。

CycleGAN [44] 框架设计了一种领域自适应策略。Yang 等人 [39] 提出了一个名为 D4 的无配对去雾框架。该框架可以估计雾霾图像的场景深度，并生成具有不同厚度的雾霾数据，以有利于去雾模型的训练。然而，直接使用手工设计的先验方法无法避免基于先验方法的固有缺陷。另一类研究方向旨在通过损失函数或网络架构引入先验知识。Li 等人 [23] 提出了一个半监督流程，采用基于先验的损失函数在真实数据集上训练网络。PSD [7] 在预训练的去雾模型上增加了一个基于物理的子网络，并进一步提出了一个先验损失，以无监督的方式在真实世界数据上微调网络。然而，GANs 容易在生成结果中产生伪影，这对模型的训练是有害的。在我们的研究中，我们通过提出一种新的数据生成流程并利用潜在的高质量先验，来探索克服这两类真实图像去雾方法的缺点。

2.2. 离散码本学习

最近，VQ-VAE [36] 提出了一个向量量化的自编码器框架，该框架在潜在空间中学习离散码本。离散

表示有效地解决了自编码器 [19] 架构中的“后验崩溃”问题。VQGAN [11] 通过引入对抗性监督来进行码本学习，进一步提高了重构结果的感知质量。学习到的离散码本在许多低级视觉任务中有助于提升性能，包括人脸修复 [14, 43] 和图像超分辨率 [6]。Gu 等人 [14] 将向量量化技术引入人脸修复，并设计了一个并行解码器，以在视觉质量和保真度之间取得平衡。Zhou 等人 [14] 将盲人脸修复任务视为一个编码预测问题，并提出了基于 Transformer 的预测网络，取代最近邻匹配操作，以更好地匹配相应的编码。FeMaSR [6] 将离散码本学习扩展到盲超分辨率任务中。受这些方法出色表现的启发，我们首次利用高质量的码本先验 (HQPs) 进行真实图像去雾。我们提出了一种新颖且可控的 HQPs 匹配操作，以进一步缩小合成数据与真实数据之间的差距，这是处理真实场景时不可避免的问题。

3. 真实图像去雾的数据准备

重新设计数据生成流程已被证明是解决真实世界低级视觉任务的有效方法 [37, 38, 41]。基于这些工作，我

我们在合成用于训练去雾网络的配对数据时考虑了多种退化因素，以减轻与真实数据之间的域差异。为了简洁起见，我们将公式 (1) 表示为 $I(x) = \mathcal{P}(J(x), t(x), A)$ 。雾霾图像的生成过程可以写成：

$$I(x) = \text{JPEG}(\mathcal{P}(J(x)^\gamma + \mathcal{N}, e^{\beta d(x)}, A + \Delta A)). \quad (2)$$

公式 (2) 的具体细节如下：

光照条件差异。 $\gamma \in [1.5, 3.0]$ 是一个亮度调整因子， \mathcal{N} 是高斯噪声分布。这两个组件可以模拟雾霾天气中经常出现的光照条件差。

透射图。作为退化模型中的关键参数，我们采用深度估计算法 [18] 来估计深度图 $d(x)$ ，并使用 $\beta \in [0.3, 1.5]$ 来控制雾霾的密度。

具有颜色的雾霾。为了获得更多样化的雾霾图像，我们考虑了全局大气光的色彩偏差，通过一个三通道向量 $\Delta A \in [-0.025, 0.025]$ 实现。 A 的取值范围在 $[0.25, 1.0]$ 之间。

JPEG 压缩。我们观察到去雾算法会放大 JPEG 压缩伪影。因此，在去雾的同时消除这些伪影是理想的。 $\text{JPEG}(\cdot)$ 表示最终结果中的 JPEG 压缩处理。

我们选择了 500 张干净图像来构建配对数据，雾霾数据则在训练阶段动态生成。此外，在所提出的数据生成流程中，低光照和 JPEG 压缩以 50% 的概率出现。

4. 方法论

我们工作的关键思路是采用离散码本，将高质量的先验 (HQPs) 引入去雾网络。所提出方法的总体框架如图 2 所示。训练过程可以分为两个阶段。在第一个训练阶段，我们在高质量数据上预训练 VQGAN [11]，从而获得具有 HQPs 的潜在离散码本 \mathcal{Z} 和相应的解码器 \mathbf{G}_{vq} (见 Sec.4.1)。在第二阶段，基于预训练 VQGAN 的 RIDCP 在由我们所提出的合成流程生成的雾霾图像上进行训练 (见 Sec.4.2)。此外，为了帮助网络找到更准确的码，我们提出了一种基于高质量图像码激活分布的可控调整特征匹配策略 (见 Sec. 4.3)。训练目标的细节可以在补充材料中找到。

4.1. 用于高质量先验的潜在码本

我们首先简要介绍 VQGAN 的工作原理。给定一个高质量图像块 x ，它作为 VQGAN 编码器 \mathbf{E}_{vq} 的输



(a) Hazy input (b) Reconstruction

图 3. 由预训练的 VQGAN 重构的结果中，虽然雾霾被去除，但引入了失真的纹理。

入，编码器输出潜在特征 \hat{z} 。然后， \hat{z} 的每一个“像素” \hat{z}_{ij} 都会与码本 $\mathcal{Z} \in \mathbb{R}^{K \times n}$ 中最近的 HQPs 匹配，从而获得离散表示 z^q_{ij} ，其表达式为：

$$z^q_{ij} = \mathcal{M}(\hat{z}_{ij}) = \arg \min_{z_k \in \mathcal{Z}} (\|\hat{z}_{ij} - z_k\|_2), \quad (3)$$

其中， K 表示码本的大小， n 是 \hat{z} 的通道数， $\mathcal{M}(\cdot)$ 代表匹配操作。最后，输入 x 通过 \mathbf{G}_{vq} 解码重构：

$$x' = \mathbf{G}_{vq}(z^q) = \mathbf{G}_{vq}(\mathcal{M}(\mathbf{E}_{vq}(x))), \quad (4)$$

其中， x' 为重构结果。

观察 1。为了理解码本中高质量先验 (HQPs) 的潜力并更好地加以利用，我们对由预训练的 VQGAN 重构的结果进行了观察。如图 3 所示，我们的 VQGAN 在不进行微调的情况下可以去除薄雾并恢复雾霾输入图像的鲜艳色彩。我们分析认为，采用 HQPs 的匹配方式可以替换退化的特征，从而帮助其跃迁到高质量域。然而，由于匹配正确码的难度，VQGAN 的去雾能力有限。此外，由于向量量化阶段的信息丢失，产生了一些失真的纹理。因此，直接采用 \mathbf{G}_{vq} 的特征是次优的选择。直观来看，我们的下一步是训练一个能够帮助先验匹配的编码器 \mathbf{E} ，以及一个能够利用从 HQPs 重构特征的解码器 \mathbf{G} 。

4.2. 通过特征匹配的图像去雾

基于 Sec. 4.1 中的观察，图像去雾被分解为两个子任务：匹配正确的码和消除纹理失真。

用于 HQPs 匹配的编码器。我们采用了 SwinIR [25]，该方法在图像恢复中展示了强大的特征提取能力，并以此设计我们的编码器 \mathbf{E} 。具体来说，浅层特征提取模块由一组残差层 [17] 组成，并对特征进行 $4 \times$ 的下采样。接下来是 4 个残差 Swin Transformer 模块 [28]，它们作为深层特征提取模块。

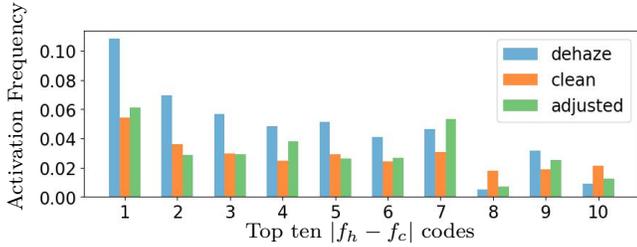


图 4. 不同情况下码激活的频率。'dehaze' 表示使用真实雾霾图像作为去雾网络的输入，'clean' 表示将干净图像输入预训练的 VQGAN 网络。'adjusted' 表示将真实雾霾图像输入配备可控 HQPs 匹配操作的 RIDCP，并使用推荐参数进行调整。

具有归一化特征对齐的解码器。我们提出了归一化特征对齐 (NFA) 方法，以帮助解码器利用从 HQPs 重构的特征。首先，VQGAN 由于向量量化操作 [14, 43] 带来的信息丢失，往往会降低结果的保真度。我们的解决方案是通过在 HQPs 匹配之前的特征引导，消除失真。具体而言，在第 i 层中，我们采用可变形卷积 [9] 将来自 \mathbf{G}_{vq} 的特征 F^{i}_{vq} 与来自 \mathbf{G} 的特征 F^i 对齐，可以表示为：

$$F_w^i = DCONV(F_{vq}^i, CONV(Concat(F_{vq}^i, F^i))), \quad (5)$$

其中 F_w^i 表示变形后的特征， $DCONV$ 是可变形卷积层， $CONV$ 是用于偏移生成的卷积层。此外，我们注意到 F_w^i 和 F^i 的比值不稳定，导致组合不充分。因此，我们通过强制它们在同一数量级上平衡各自的贡献，公式为：

$$F^i = F^i + \sum \frac{F^i}{F_w^i} F_w^i. \quad (6)$$

关于 NFA 方法的影响将在 Sec. 5.4 中详细介绍。

4.3. 可控 HQPs 匹配操作

观察 2. 借助编码器 \mathbf{E} 和解码器 \mathbf{G} ，我们的 RIDCP 取得了相对令人满意的结果。然而，仍然存在一些局限性，例如在一些具有挑战性的真实数据中，色彩饱和度较低。这并不是因为 HQPs 本身，正如在观察 1 中所展示的，它们已经表现出了强大的重构鲜艳图像的能力。主要原因是难以找到正确的 HQPs，这归因于合成数据和真实数据之间的域差异。尽管我们提出的合成流程在缩小域差异方面比之前的工作 [22, 42] 有了显著的改进，但仍然不可能通过我们的流程覆盖所有真实世界中的雾霾条件。

为了验证这一说法，我们进行了如下观察。我们随机收集了 200 张高质量的干净图像作为预训练 VQGAN 的输入，并计算每个码的激活频率 $f_c \in \mathbb{R}^K$ 。类似地，将 200 张真实雾霾图像输入去雾网络，以计算激活频率 $f_h \in \mathbb{R}^K$ 。图 4 展示了 f_h 和 f_c 之间差异最大的前十个码的激活频率。我们可以看到显著分布偏移。这一观察结果证明了不可避免的域差异导致了 HQPs 匹配的分散性。因此，HQPs 仍然具有未被充分挖掘的潜力，这也是我们需要进一步探索的下一个步骤。

通过距离重新计算的可控匹配。基于上述观察，在遇到真实雾霾图像时，匹配更优的 HQPs 是必不可少的，即匹配在干净图像中高频出现的先验。影响 HQPs 匹配的两个主要因素是编码器 \mathbf{E} 和匹配操作 $\mathcal{M}(\cdot)$ 。由于在没有参考图像的情况下，很难在真实雾霾图像上重新训练编码器 \mathbf{E} ，因此定义一个新的匹配操作 $\mathcal{M}'(\cdot)$ 是一个简单但必要的解决方案。我们提出了可控 HQPs 匹配 (CHM)，其尝试通过在匹配阶段分配不同权重来重新计算距离。CHM 可以写成：

$$\mathcal{M}'(\hat{z}) = \arg \min_{z_k \in Z} (\mathbf{F}(\tilde{f}_k, \alpha) \times \|\hat{z} - z_k\|), \quad (7)$$

其中， $\mathbf{F}(\tilde{f}_k, \alpha)$ 是根据频率差异 $\tilde{f}_k = f_h^k - f_c^k$ 并通过参数 α 调整后生成的权重函数。在设计 \mathbf{F} 时，有三个目标：1) 由于较高的 \tilde{f}_k 意味着较少的激活需求， \mathbf{F} 应与 \tilde{f}_k 单调变化，以确保趋势调整的一致性。2) 当 $\mathbf{F}(0, \alpha) \equiv 1$ 时，表示在干净和雾霾数据上具有相同频率的 HQPs 不进行调整。3) 调整的幅度可以通过 α 单icontrol，例如： $\forall \tilde{f}_1 > \tilde{f}_2, \forall \alpha_1 > \alpha_2 \rightarrow \frac{\mathbf{F}(\tilde{f}_1, \alpha_1)}{\mathbf{F}(\tilde{f}_2, \alpha_1)} > \frac{\mathbf{F}(\tilde{f}_1, \alpha_2)}{\mathbf{F}(\tilde{f}_2, \alpha_2)}$ 。巧合的是，指数函数具有这些属性，因此 \mathbf{F} 可以被公式化为：

$$\mathbf{F}(\tilde{f}_k, \alpha) = e^{\alpha \times \tilde{f}_k} = e^{\alpha \times (f_h^k - f_c^k)} \quad (8)$$

图 2(b) 使用两个 Voronoi 图来模拟特征匹配过程中在高维空间中发生的变化。正如我们所见，原本属于灰色区域的点在距离重新计算后匹配到了彩色区域，即找到了更优的 HQPs。

推荐 α 的可能解决方案。我们的方法能够基于上述策略控制 HQPs 的匹配。最终目标是找到一个合适的 α ，以使网络适应真实域。根据大数定律，频率 f_c^k, f_h^k 可以替换为对应的概率 $P_c(x = z_k), P_h(x = z_k | \alpha)$ 。去雾结果与干净域之间的差距可以通过这两个概率分布



(a) Hazy image (b) MSBDN [10] (c) Dehamer [15] (d) PSD [7] (e) D4 [39] (f) DAD [33] (g) RIDCP

图 5. 在 RTTS 数据集上的视觉对比 [22].

的差异来表示。因此，真实域适应问题转化为计算一个最优参数 $\hat{\alpha}$ ，以最小化 $P_c(x = z_k)$ 和 $P_h(x = z_k|\alpha)$ 的前向 Kullback-Leibler 散度，这也相当于 α 的最大似然估计：

$$\begin{aligned}
 \hat{\alpha} &= \arg \min_{\alpha} KL(P_c || P_h) \\
 &= \arg \min_{\alpha} \sum_{i=1}^K P_c(x = z_i) \log \frac{P_c(x = z_i)}{P_h(x = z_i|\alpha)} \\
 &= \arg \max_{\alpha} \sum_{i=1}^K P_c(x = z_i) \log P_h(x = z_i|\alpha) \\
 &= \arg \max_{\alpha} \prod_{i=1}^K P_c(x = z_i) P_h(x = z_i|\alpha).
 \end{aligned} \tag{9}$$

我们使用二分搜索算法迭代地找到 $\hat{\alpha}$ 的近似最优解。最终确定的 $\hat{\alpha} = 21.25$ ，更高精度的计算对结果影响不大。需要注意的是， $\hat{\alpha}$ 可能并不是所有情况的确定选择。用户可以根据自己的需求灵活调整 α 值。

5. 实验

5.1. 数据集

高质量数据集。为了从预训练的 HQPs 中获得高质量的结果，VQGAN 需要在包含高分辨率和纹理清晰图像的大规模数据集上进行训练。在我们的工作中，我们使用了 DIV2K [1] 和 Flickr2K [26] (包含 4,250 张图像) 来训练第一阶段。这两个数据集广泛应用于高质量重建任务 [6, 24, 25]。

真实雾霾数据集。我们在 RTTS 数据集 [22] 上对我们的去雾网络进行了定性和定量评估，该数据集包含超过 4,000 张场景多样、分辨率多样、退化问题复杂的真实雾霾图像。此外，我们还使用了 Fattal 的数据集 [12]，其中包含 31 个经典的真实雾霾案例，用于进一步的视觉对比。

5.2. 实现细节

对于 VQGAN 和 RIDCP 的训练，我们使用 Adam 优化器，默认参数为 $(\beta_1 = 0.9, \beta_2 = 0.99)$ 。在训练阶段，学习率固定为 0.0001，批量大小设置为 16。为了进

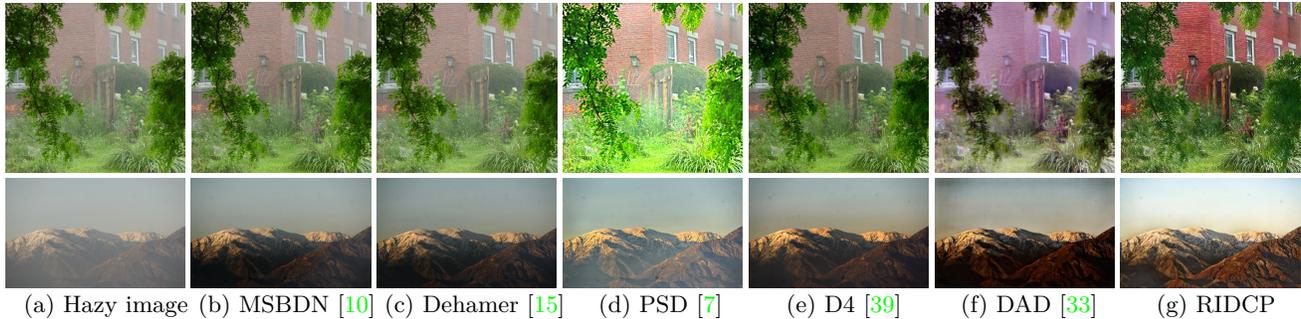


图 6. 在 Fattal 数据集上的视觉对比 [12].

表 1. 在 RTTS 数据集上的定量比较和用户研究。红色表示最佳结果，蓝色表示第二好的结果。‘US’ 表示用户研究中的投票百分比。

Method	FADE↓	BRISQUE↓	NIMA↑	US↑
Hazy image	2.484	37.011	4.3250	0.030
MSBDN	1.363	28.743	4.1401	0.046
Dehamer	1.895	33.866	3.8663	0.041
DAD	1.130	32.727	4.0055	0.143
PSD	0.920	25.239	4.3459	0.105
D4	1.358	33.206	3.7239	0.079
RIDCP	0.944	18.782	4.4267	0.556

行数据增强,我们随机调整输入大小并裁剪为 256×256 的尺寸,并以 50% 的概率进行翻转。在第一个训练阶段中,我们的 VQGAN 在 DIV2K 和 Flickr2K 数据集上预训练了 350K 次迭代。随后,所提出的 RIDCP 在通过我们提出的合成流程生成的数据上训练了 10K 次迭代。所有实验都是在 PyTorch 框架下使用 4 个 NVIDIA V100 GPU 进行的。

5.3. 与最新方法的比较

我们将所提出的方法与几种最新的去雾方法进行性能比较。实验从定量和定性两个角度进行设计。此外,我们还进行了用户研究,以验证我们方法的主观性能。通过这些实验,我们能够评估 RIDCP 方法在图像去雾任务中的表现,并展示其在高频细节恢复、色彩还原以及视觉效果方面的优势。

定量比较。由于真实雾霾数据集中没有对应的真实图像,我们使用一些无参考指标进行定量比较。首先,我们采用雾感知密度评估器 (FADE) [8] 来估计雾霾密度。此外,还包括两个广泛使用的图像质量评估指标 BRISQUE [29] 和 NIMA [34] 进行性能评估。我们

在 RTTS 数据集上进行了定量比较,比较对象包括两种在合成雾霾图像数据集 [22] 上表现优异的去雾方法 (MSBDN [10] 和 Dehamer [15]) 以及三种用于真实去雾的最新方法 (DAD [33]、PSD [7]、D4 [39])。表 1 展示了结果。可以看到,RIDCP 在 BRISQUE 和 NIMA 指标上取得了最佳表现,分别提升了 25.57% 和 1.86%。对于 FADE,我们的方法排名第二,仅略低于 PSD。然而,正如图 5 所示,PSD 倾向于产生过度增强的结果,这导致评估的不准确。总体而言,我们的方法在定量指标上取得了最好的结果,后续实验将进一步证明该方法的优越性。

定性比较。我们在 RTTS 和 Fattal 的数据集上进行了定性比较,结果如图 5 和图 6 所示。可以观察到,Dehamer、MSBDN 和 D4 在处理真实雾霾图像时表现不佳。PSD 虽然能够生成较亮的结果,但去雾能力有限。DAD 在去雾方面有效,但其结果存在颜色偏差和暗色调的问题。相比其他方法,我们的方法在亮度、色彩丰富度和雾霾残留方面生成了最佳的感知结果。更多结果可以在补充材料中找到。

用户研究。我们进行了一项用户研究,主观评估所提出的方法与其他方法的比较。我们从 RTTS 数据集中随机选择了 100 张图像进行比较,并邀请了 5 位具有图像处理背景的专家和 5 位普通观察者作为志愿者。在用户研究之前,我们向观察者提供了一些提示: 1) 主要关注点是雾霾是否被去除,尤其是远处的浓雾。2) 注意自然色彩是否被恢复,过度增强并不意味着高质量。3) 一个好的方法应生成无伪影的结果。随后,图像分组显示给观察者。每组图像包含输入图像、用于定量比较的五种方法的结果以及我们的方法结果。观察者在至少观察 10 秒后需要选择其中表现最好的一个。我们统计了每种方法被选为最佳的百分比,最终得分列于

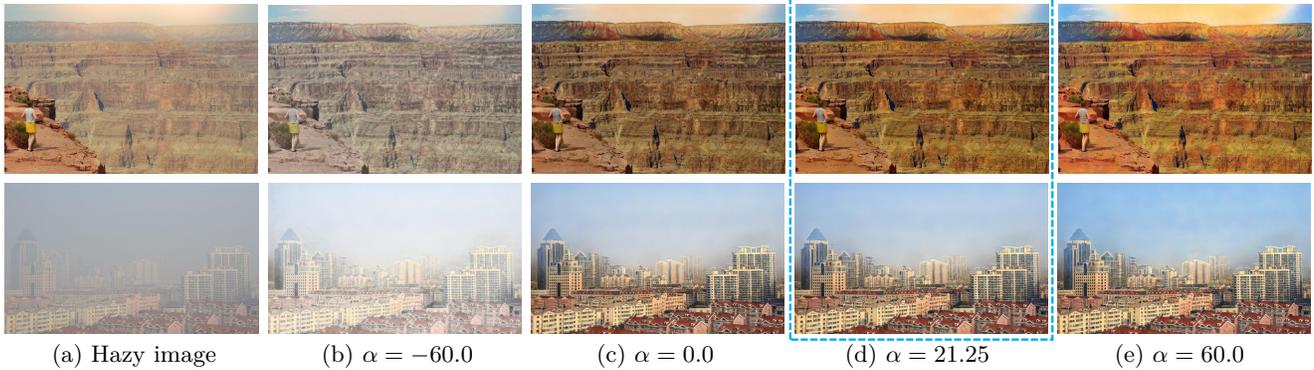


图 7. 不同调整程度下的结果。所提出的 CHM 允许用户根据需求调整增强的程度，范围从低 ($\alpha = -60.0$) 到高 ($\alpha = 60.0$)。推荐的值为 $\alpha = 21.25$ ，此值有助于网络生成最自然的结果（由蓝框标出）。请放大查看最佳效果。

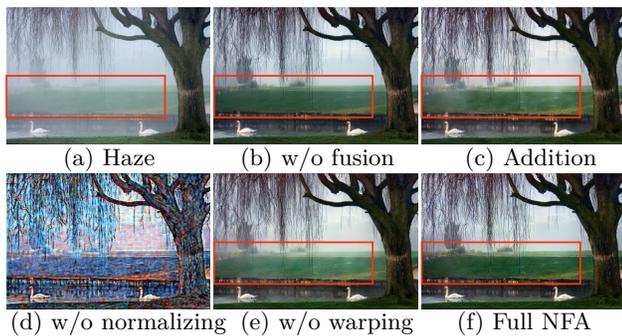


图 8. NFA 消的融实验结果。

表 1. 所提出的方法获得了最高分，且远远领先于第二名，进一步证明了我们方法的卓越去雾能力。

5.4. 消融实验

为了验证每个关键组件的有效性，我们进行了系列消融实验。在本节中我们讨论了 CHM、NFA 和现象学退化流程的有效性。

调整参数的影响。 真实域适应的程度由公式 (8) 中的参数 α 控制，通过调整 α 可以灵活调节最终结果。因此，我们好奇不同的 α 值会带来什么影响。正如图 7 所示， α 的值与图像增强效果呈线性关系。当 $\alpha > 0$ 时，可以生成更具视觉吸引力，甚至过度增强的结果。有趣的是，当 α 朝相反方向调整时，我们可以得到效果较弱的结果。

NFA 的有效性。 我们的 NFA 可以在保留从 HQPs 重构的有用信息的同时，帮助消除由于特征匹配引起的失真纹理。NFA 可以分为两个关键部分：基于可变形卷积的变形操作和基于归一化的加法操作。为了分析每个部分的作用，我们提出了 4 个替代 NFA 的变体，分别为：1) 没有任何融合操作。2) 直接相加。3)

归一化相加但不进行变形。4) 变形后直接相加。图 8 展示了一组比较。观察红框中的草地区域，变体 1 的结果较暗且残留有薄雾。变体 2 和 4 在某些区域也有不均匀的雾残留。变体 3 产生了明显的伪影，这是因为将未对齐的特征强行归一化到同一数量级并相加，使得网络难以训练。只有完整的 NFA 在亮度和去雾方面取得了最佳效果，能够生成最好的去雾结果，并消除失真纹理。

现象学退化流程的有效性。 为了证明我们提出的用于生成配对数据的退化流程能够增强去雾能力，我们在两个广泛使用的合成数据集上重新训练了我们的 RIDCP，分别是 OTS [22] 和 Haze4K [42]。值得注意的是，Haze4K 经过了 DAD [33] 的后处理。此外，我们还将 OTS 训练集替换为我们的合成数据，分别用于基于 Transformer 的 Dehazer 和基于 CNN 的 MSBDN，从而证明它可以带来普遍的提升。比较结果如图 9 所示。我们可以观察到，在 OTS 和 Haze4K 训练下，我们的去雾网络无法有效去除雾霾。此外，借助我们的训练数据，Dehazer 和 MSBDN 可以生成雾霾更少、亮度更高的结果。然而，与我们的方法相比，它们在色彩恢复方面仍然存在困难，这也进一步证明了 HQPs 和我们的域适应策略的有效性。

6. 结论与未来工作

在本文中，我们提出了一种新颖的范式，旨在使深度去雾网络更加贴近真实世界。我们提出的现象学退化流程能够合成更真实的雾霾数据，从而在去雾能力上取得了显著提升。基于我们的观察和分析，我们将 VQGAN 中的高质量先验引入去雾网络，并逐步发挥其优势，最终通过高质量码本先验 (HQPs) 构建了我



图 9. 提出的数据生成流程的消融实验结果展示如下。我们在 OTS (b) 和 Haze4K (c) 数据集上重新训练了我们的去雾网络，以验证我们生成流程的有效性。Dehamer 和 MSBDN 也在我们的合成数据上进行了重新训练，结果标记为 †。

们的真实图像去雾网络 (RIDCP)。大量实验表明，所提出的范式具有明显的优势。然而，RIDCP 在处理极度浓密且不均匀分布的雾霾图像时，仍会生成不理想的结果。这类雾霾难以模拟，现有方法也难以应对。我们将把解决这一具有挑战性的案例作为未来的研究方向。

参考文献

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 126–135, 2017. 6

[2] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1014–1018. IEEE, 2019. 2

[3] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 444–445, 2020. 2

[4] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, 2016. 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for

single image haze removal. *IEEE Transactions on Image Processing (TIP)*, 25(11):5187–5198, 2016. 1, 2

[6] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM)*, pages 1329–1338, 2022. 2, 3, 6

[7] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7180–7189, 2021. 1, 2, 3, 6, 7

[8] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing (TIP)*, 24(11):3888–3901, 2015. 7

[9] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 764–773, 2017. 5

[10] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2157–2167, 2020. 1, 2, 6, 7

[11] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12873–12883, 2021. 2, 3, 4

[12] Raanan Fattal. Dehazing using color-lines. *ACM Transactions on Graphics (TOG)*, 34(1):1–14, 2014. 1, 2, 6, 7

[13] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio. Generative adversarial nets. In *Neural Information Processing Systems (NeruIPS)*, 2014. 2

[14] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *Proceedings of the European*

918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971

- 972 *Conference on Computer Vision (ECCV)*, 2022. 2, 3, 1026
973 5 1027
974 1028
975 [15] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin 1029
976 Cong, Wenqi Ren, and Chongyi Li. Image dehazing 1030
977 transformer with transmission-aware 3d position em- 1031
978 bedding. In *Proceedings of the IEEE/CVF Conference 1032*
979 *on Computer Vision and Pattern Recognition (CVPR)*, 1033
980 pages 5812–5820, 2022. 1, 2, 6, 7 1034
981 1035
982 [16] Kaiming He, Jian Sun, and Xiaoou Tang. Single image 1036
983 haze removal using dark channel prior. *IEEE Trans- 1037*
984 *actions on Pattern Analysis and Machine Intelligence 1038*
985 *(TPAMI)*, 33(12):2341–2353, 2010. 1, 2 1039
986 1040
987 [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian 1041
988 Sun. Deep residual learning for image recognition. In 1042
989 *Proceedings of the IEEE Conference on Computer Vi- 1043*
990 *sion and Pattern Recognition (CVPR)*, pages 770–778, 1044
991 2016. 4 1045
992 1046
993 [18] Mu He, Le Hui, Yikai Bian, Jian Ren, Jin Xie, 1047
994 and Jian Yang. Ra-depth: Resolution adaptive 1048
995 self-supervised monocular depth estimation. *arXiv 1049*
996 *preprint arXiv:2207.11984*, 2022. 4 1050
997 1051
998 [19] Geoffrey E Hinton and Richard Zemel. Autoencoders, 1052
999 minimum description length and helmholtz free energy. 1053
1000 *Advances in Neural Information Processing Systems*, 6, 1054
1001 1993. 3 1055
1002 1056
1003 [20] Shih-Chia Huang, Trung-Hieu Le, and Da-Wei Jaw. 1057
1004 Dsnet: Joint semantic learning for object detection in 1058
1005 inclement weather conditions. *IEEE Transactions on 1059*
1006 *Pattern Analysis and Machine Intelligence (TPAMI)*, 43(8):2623–2633, 2020. 1 1060
1007 1061
1008 [21] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, 1062
1009 and Dan Feng. Aod-net: All-in-one dehazing network. 1063
1010 In *Proceedings of the IEEE International Conference 1064*
1011 *on Computer Vision (ICCV)*, pages 4770–4778, 2017. 1065
1012 1, 2 1066
1013 1067
1014 [22] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan 1068
1015 Feng, Wenjun Zeng, and Zhangyang Wang. Bench- 1069
1016 marking single-image dehazing and beyond. *IEEE 1070*
1017 *Transactions on Image Processing (TIP)*, 28(1):492– 1071
1018 505, 2019. 5, 6, 7, 8 1072
1019 1073
1020 [23] Lerenhan Li, Yunlong Dong, Wenqi Ren, Jinshan Pan, 1074
1021 Changxin Gao, Nong Sang, and Ming-Hsuan Yang. 1075
1022 Semi-supervised image dehazing. *IEEE Transactions 1076*
1023 *on Image Processing (TIP)*, 29:2766–2779, 2019. 2, 3 1077
1024 1078
1025 [24] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, 1079
Gwanggil Jeon, and Wei Wu. Feedback network for im-
age super-resolution. In *Proceedings of the IEEE/CVF
Conference on Computer Vision and Pattern Recogni-
tion (CVPR)*, pages 3867–3876, 2019. 6
[25] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang,
Luc Van Gool, and Radu Timofte. Swinir: Image
restoration using swin transformer. In *Proceedings of
the IEEE/CVF International Conference on Computer
Vision (ICCV)*, pages 1833–1844, 2021. 4, 6
[26] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah,
and Kyoung Mu Lee. Enhanced deep residual net-
works for single image super-resolution. In *Proceed-
ings of the IEEE Conference on Computer Vision and
Pattern Recognition Workshops (CVPRW)*, pages 136–
144, 2017. 6
[27] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen.
Griddehazenet: Attention-based multi-scale network
for image dehazing. In *Proceedings of the IEEE In-
ternational Conference on Computer Vision (ICCV)*,
pages 7314–7323, 2019. 1, 2
[28] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei,
Zheng Zhang, Stephen Lin, and Baining Guo. Swin
transformer: Hierarchical vision transformer using
shifted windows. In *Proceedings of the IEEE Interna-
tional Conference on Computer Vision (ICCV)*, pages
10012–10022, 2021. 4
[29] Anish Mittal, Anush K Moorthy, and Alan C Bovik.
Blind/referenceless image spatial quality evaluator.
In *2011 Conference Record of the Forty Fifth Asilo-
mar Conference on Oignals, Systems and Computers
(ASILOMAR)*, pages 723–727. IEEE, 2011. 7
[30] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie,
and Huizhu Jia. Ffa-net: Feature fusion attention net-
work for single image dehazing. *Proceedings of the
AAAI Conference on Artificial Intelligence*, 34(07),
2020. 1, 2
[31] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun
Cao, and Ming-Hsuan Yang. Single image dehazing via
multi-scale convolutional neural networks. In *Proceed-
ings of the European Conference on Computer Vision
(ECCV)*, pages 154–169. Springer, 2016. 1, 2
[32] Christos Sakaridis, Dengxin Dai, Simon Hecker, and
Luc Van Gool. Model adaptation with synthetic and

- real data for semantic dense foggy scene understanding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 687–704, 2018. 1
- [33] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2808–2817, 2020. 1, 2, 6, 7, 8
- [34] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing (TIP)*, 27(8):3998–4011, 2018. 7
- [35] Robby T Tan. Visibility in bad weather from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 2
- [36] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in Neural Information Processing Systems*, 30, 2017. 3
- [37] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14791–14801, 2021. 3
- [38] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1905–1914, 2021. 3
- [39] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2037–2046, 2022. 2, 3, 6, 7
- [40] Tian Ye, Mingchao Jiang, Yunchen Zhang, Liang Chen, Erkang Chen, Pen Chen, and Zhiyong Lu. Perceiving and modeling density is all you need for image dehazing. *arXiv preprint arXiv:2111.09733*, 2021. 1, 2
- [41] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021. 3
- [42] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16180–16189. IEEE, 2021. 5, 8
- [43] Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *Neural Information Processing Systems (NeurIPS)*, 2022. 2, 3, 5
- [44] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. 3
- [45] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single image dehazing using color attenuation prior. In *British Machine Vision Conference (BMVC)*. Citeseer, 2014. 1, 2