

# 条件式变分图像去雨网络

Yingjun Du<sup>#</sup>, Jun Xu<sup>#</sup>, Xiantong Zhen, Ming-Ming Cheng, Ling Shao

**摘要**—图像去雨是一项重要但具有挑战性的图像处理任务。尽管确定性的图像去雨方法表现出不错的性能，但它们难以学习用于概率推断和各种预测的灵活表示形式。此外，去雨强度在空间位置和整个色彩通道中都不同，这使这项任务更加困难。在本文中，我们利用条件式变分自动编码器（CVAE）的独有生成能力，为雨天图像提供多种预测，提出了一种条件式变分图像去雨（CVID）网络，以实现更好的去雨效果。为了执行空间自适应去雨，提出了一个空间密度估计（SDE）模块来估计每个图像的雨密度图。由于雨水密度在不同的颜色通道之间变化，因此还提出了一种通道方式（CW）去雨方案。对合成数据集和真实数据集进行的实验表明，所提出的 CVID 网络比以前的确定性方法在图像去雨方面具有更好的性能。广泛的消融研究验证了 CVID 网络中提出的 SDE 模块和 CW 方案的有效性。代码可在 <https://github.com/Yingjun-Du/VID> 中找到。

**Index Terms**—条件式变分自动编码器, 单图像去雨, 空间注意力图, 通道维去雨.

## I. 引言

雨水的存在会降低人类感知图像的视觉真实性，并严重阻碍视觉系统的性能 [19]。图像去雨旨在从低质量图像中去除雨痕，并恢复其干净的背景。由于其在视频监控、目标检测和目标分割等许多实际应用中的先决作用，已受到越来越多的关注。

以往的图像去雨方法大致可以分为三类：基于优化的方法 [3], [24], [52]、基于判别学习的方法 [8], [29], [46] 和基于生成学习的方法 [23], [28], [51]。其中，基于优化的方法 [3], [4] 采用适当的正则化器在线性加法复合模型 [2], [19], [40] 或非线性屏幕复合模型 [24] 下恢复雨天图像。基于判别学习的方法 [9], [22], [46] 直接学习从雨天图像到其干净背景的确定性（非线性）映射函数。生成方法 [23], [28], [51] 利用高斯混合模型 [41],

这项研究得到了国家自然科学基金和天津市自然科学基金的部分支持 (17JCJQJC43700, 18ZXZNGX00110)。

YJ Du 在荷兰阿姆斯特丹的阿姆斯特丹大学工作。J Xu 和 MM Cheng 在中国天津 300071 南开大学计算机学院 TKLNDST 工作。XTzhen 就职于中国广东石油化工大学计算机学院和阿联酋阿布扎比 Inception 人工智能研究所。L Shao 就职于阿联酋阿布扎比的 Inception 人工智能研究所。Xiantong Zhen (zhenxt@gmail.com) 是通讯作者。<sup>#</sup> 前两位作者的贡献相同。这项工作是我们会议论文 [6] 的延伸。

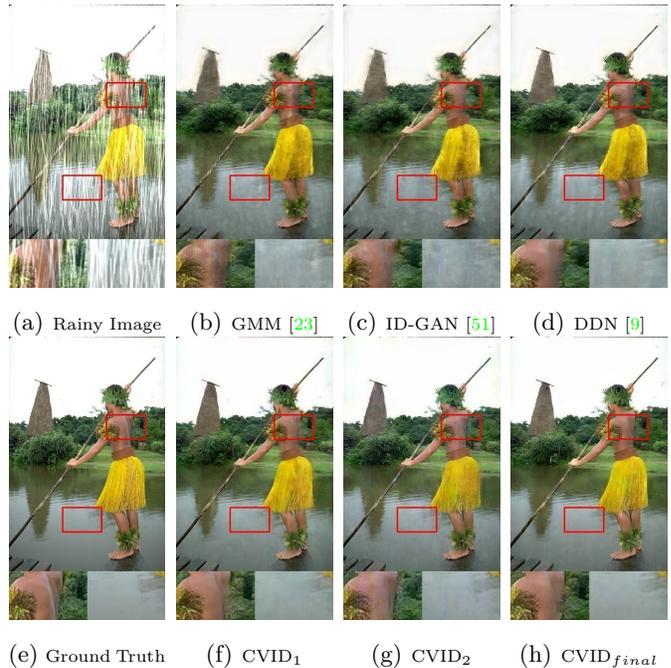


图 1: 不同图像去雨方法的去雨图像。CVID<sub>1</sub> (f) 和 CVID<sub>2</sub> (g) 是本文提出的 CVID 网络的两个候选预测，CVID<sub>final</sub> (h) 是使用它们的平均值得到的最终去雨图像。

[44] 或条件生成对抗网络 (GAN) [11], [26] 的生成建模功能，其灵感来自于它们在合成具有视觉吸引力的图像方面的成功。

尽管它们在图像去雨方面取得了成功，但这些方法存在两个主要局限性。首先，当前基于优化的方法中使用的复合模型不能完全反映雨图像的真实组成，因此正则化器在表征背景图像和雨层方面仍然不足，限制了这些方法在不同场景下的稳健去雨性能。其次，基于判别或生成学习的方法 [18] 主要学习映射函数并生成确定性去雨图像 (图 1(b)-(d))。然而，由于图像去雨固有的性质，很难为现实世界的雨天图像定义最佳去雨图像。

在本文中，为了解决上述挑战，我们通过利用最近开发的条件变分自动编码器 (CVAE) [34] 框架的强大生成能力，提出了一个条件变分图像去雨 (CVID) 网络。CVAE 提供了强大的能力来对图像先验的潜在分

布进行建模，从中可以生成干净的图像。我们提出了一个基于 CVAE 的 CVID 网络，与以前的方法中学习确定性映射函数不同，CVID 同时学习干净图像先验的潜在表示并预测多个可能的去雨图像（图 1(f)-(g)）。在 CVID 网络的学习阶段，给定一对干净和下雨的图像，编码器学习将这些干净的图像映射到一个潜在分布，该分布共享干净背景图像的公共信息，而解码器根据潜在分布空间中的采样变量恢复去雨的图像（如图 2 所示）。在推理阶段，我们使用先验网络从干净图像的潜在分布中采样多个潜在变量，并采用蒙特卡罗方法 [33] 对每个采样变量进行确定性预测（如图 3 所示）。这些预测被平均以产生最终的去雨图像（图 1 (h)）。

雨天图像的一个重要观察结果是，雨条纹通常在整个图像中分布不均，无论是空间位置还是颜色通道（如图 4 所示）。因此，如何在空间上和通道上明智地自适应地去除雨纹，以实现图像的去雨，是需要认真处理的问题。为了解决这个问题，张等人 [50] 提出了一种利用全局降雨密度信息的密度感知去雨方法。但是这种方法会在局部区域产生不准确的去雨结果，并且忽略了不同通道降雨分布不同的事实。为此，我们提出了一种空间密度估计 (SDE) 模块和一种通道式 (CW) 方案，用于对我们的 CVID 网络进行更自适应的图像去雨。在基准数据集上的实验证明了我们的 CVID 网络在图像去雨方面优于以前的竞争者，以及我们的 SDE 模块和 CW 方案的有效性。

总之，我们的贡献是三方面的：

- **一种新颖的生成网络，可输出多个预测以获得更好的图像去雨性能。** 我们利用强大的条件变分自动编码器 (CVAE) 框架 [34] 进行图像去雨。提出的 CVID 网络有效地执行概率去雨并产生多个互补的去雨预测以获得更好的性能。据我们所知，我们的 CVID 网络是第一个在 CVAE 框架下解决图像去雨问题的方法。
- **新的空间注意力模块和通道式去雨方案。** 我们提出了空间密度估计 (SDE) 模块和通道式 (CW) 方案，以赋予我们的 CVID 网络执行空间和通道方式自适应去雨的能力。§IV-D 中的大量消融研究验证了所提出的 SDE 模块和 CW 方案的有效性。
- **在不同的数据集上有更好的去雨性能。** 在三个合成和一个真实世界的雨天图像数据集上进行的实验表明，所提出的 CVID 网络始终能够比以前最先进的图像去雨方法取得更优异的结果。

本文的其余部分安排如下。在 §II 中，我们调查了相关工作。在 §III 中，我们提出了用于图像去雨的 CVID 网络。在 §IV 中进行了广泛的实验，将所提出的 CVID 网络与合成和真实世界图像数据集上的最先进图像去雨方法进行比较。最后在 §V 中给出了结论。

## II. 相关工作

在过去的十年中，已经提出了许多方法 [8], [38], [51] 来解决图像去雨问题。在此，我们简要回顾一下相关工作。

**基于优化的方法** [4], [19], [30] 基于雨图像由干净的背景图像层和雨层组成的事实提出。图像去雨可以通过在两层上使用有效的正则化器来制定，并通过适当的优化算法解决。Kang 等人 [19] 将下雨图像的高频部分分解为下雨和非下雨分量，仅对下雨分量进行处理以去除雨痕。Luo 等人 [24] 提出了一种基于图像块的判别稀疏编码框架。之后，Chen 等人 [4] 提出了一种用于去除雨水条纹的低阶外观模型。相似的，Chang 等人 [3] 利用雨条纹的低秩属性，这些属性通过基于低秩的层分解被去除。然而，由于 [4], [19], [24] 工作中使用的复合模型无法完全探索下雨图像的真实组成，这些方法使用的正则化器不足以表征背景和雨层，限制了这些方法对不同图像去雨性能的鲁棒性。

**判别式学习方法。** 最近，基于深度学习的方法在图像去雨方面取得了可喜的成绩 [8], [22], [38]。据我们所知，DerainNet 可能是第一个为图像去雨而开发的深度学习网络。之后，深度细节网络 (DNN) [9] 直接缩小从输入到输出的映射范围。深度联合雨检测和去除 (JORDER) 网络 [47], [48] 的工作是通过使用循环扩张网络为图像去雨而开发的。它可以检测下雨图像每个像素上的雨区，但不反映像素的密度信息。[51] 中的工作还提出了使用多流密集网络 (DID-MDN) 进行密度感知图像去雨的工作。DID-MDN 考虑了雨条纹的全局密度，但忽略了雨图像中局部空间位置的密度差异。[22] 中的工作引入了循环挤压和激发上下文聚合网络 (RESCAN) 来解决图像去雨中雨条纹层重叠的问题。[29] 提出了一个简单的基线网络，通过保留有价值的去雨模块来进行单幅图像去雨。在 [38] 中还开发了一个半监督图像去雨网络。然而，这些方法没有考虑不同渠道降雨密度的差异，从而可能产生不准确的结果。

**生成方法** [23], [28], [51] 也被开发用于图像去雨。Li 等人 [23] 提出对背景和雨层使用简单的基于补丁的先验。

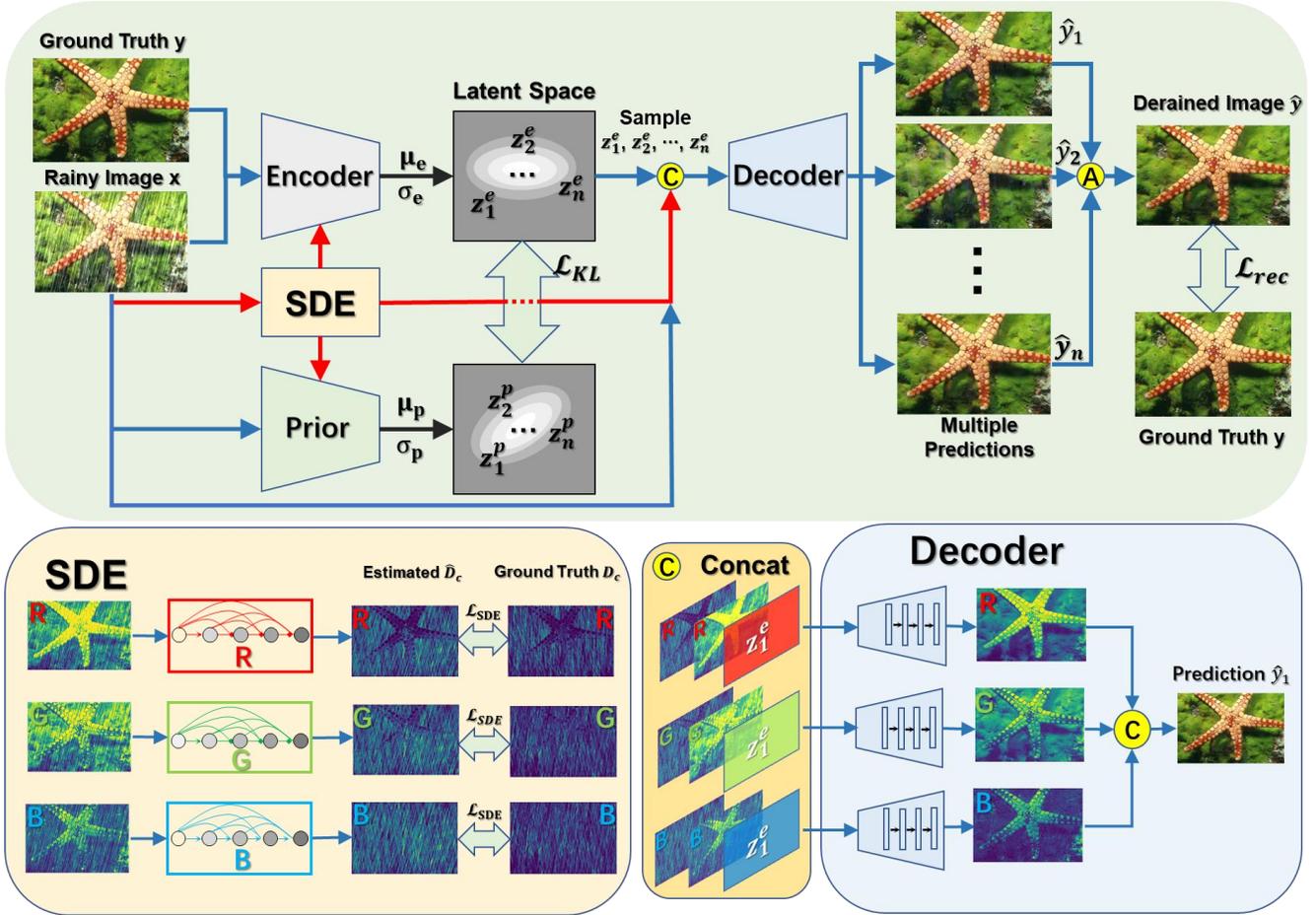


图 2: CVID 网络的学习阶段. 该编码器通过提出的 SDE 模块将降雨图像  $x$ 、干净图像  $y$  和估计的降雨图  $\hat{D}_c$  串联起来作为输入。解码器输出基于 SDE 估计的去雨图像  $\hat{y}$ 、降雨图像  $x$  和从  $\{z_i^e\}_{i=1}^n$  潜在分布  $\mathcal{N}(\mu_e, \sigma_e)$  中采样的去雨图像。黑色箭头表示潜在空间的产生，蓝色箭头是标准 CVAE 模型的数据流，而红色箭头是我们 SDE 模块的附加数据流。请注意，编码器、SDE 模块和解码器是按通道执行的关于降雨图像  $x$ ，干净图像  $y$  和估计降雨图。“C”和“A”表示连接和平均操作。

在 [51] 中，Zhang 等人利用条件生成对抗网络 (GAN) 来防止背景图像在从雨天图像中提取时退化，其中使用学习得到的鉴别器网络作为合成无雨图像的指导。[28] 的工作将视觉注意力引入 GAN 的生成网络和判别网络，并了解雨滴区域及其周围环境以去除雨滴。

**条件变分自编码器 (CVAE)** [34] 是一种基于变分自编码器 (VAE) 的条件生成模型，最初被提出用于结构化预测任务，例如图像分割和标记。Kohl 等人 [21] 将 U-Net [32] 与 CVAE 相结合，该 CVAE 能够有效地产生无限制数量的合理假设，以处理医学图像分割的固有歧义。最近，Bao 等人 [1] 提出了将 VAE 与生成对抗网络相结合的 CVAE-GAN 网络，用于细粒度图像泛化。CVAE-GAN 本质上是一个条件生成模型，以细粒度的类别标签作为输入，生成特定类别的图像。Ham

等人 [12] 将感知损失纳入 VAE 模型，并证明其对图像修复的有效性。在我们的 CVID 中，我们使用 CVAE 进行监督学习 [35]。我们对干净图像的潜在分布进行建模，并预测输入雨天图像的特定干净图像，该图像被视为条件。在学习阶段，我们学习了一个先验网络，该网络将雨天图像作为输入并估计其相应干净图像的潜在分布。在推理阶段，使用来自先验的采样潜在变量，解码器网络可以从雨天恢复干净的图像。此外，我们结合了雨天图像的密度图估计，作为充分探索 CVAE 对图像去雨的能力的额外条件。

### III. 学习条件变分图像去雨网络

在本节中，我们首先介绍在条件变分自动编码器 (CVAE) 框架下开发的条件变分图像去雨 (CVID) 网络。在 §III-A 中，我们在 §III-B 中提供了 CVAE 的预备

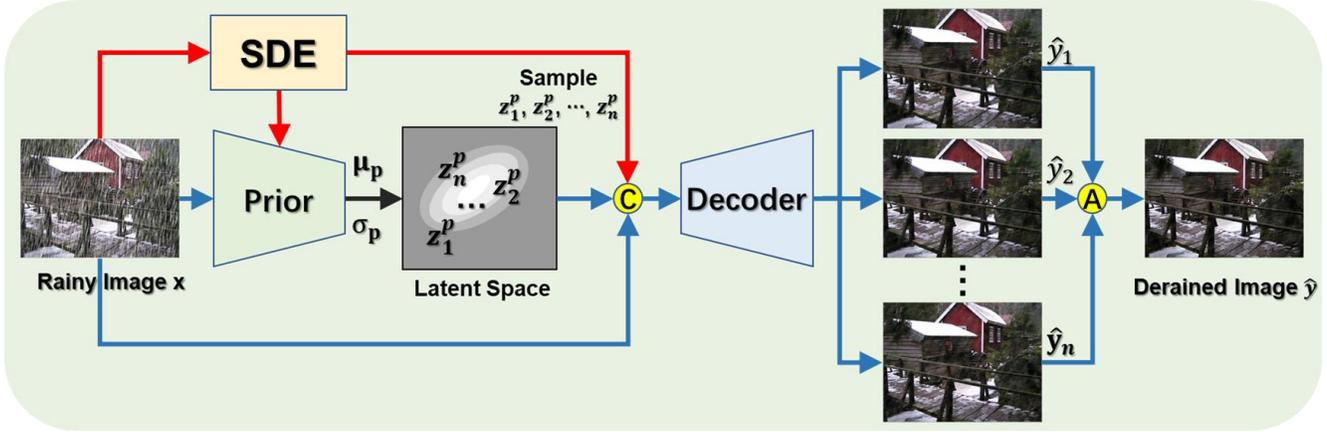


图 3: CVID 网络的推理阶段。我们从先验分布  $\mathcal{N}(\mu_p, \sigma_p)$  中采样多个潜在变量  $z_i^p$  ( $i = 1, \dots, n$ ), 并采用蒙特卡罗方法 [33] 执行确定性推理, 然后进行加权平均。黑色箭头表示潜在空间的生成过程。“A”和“C”表示平均和连接操作。SDE 是空间密度估计模块, 将在 III-C 中介绍。

知识。然后我们描述了所提出的用于图像去雨的 CVID 网络的学习阶段。我们在 §III-C 和 §III-D 中分别提出了空间密度估计 (SDE) 模块和通道 (CW) 去雨方案。在 §III-E 中提供了 CVID 的优化。最后, 我们在 §III-F 中介绍了 CVID 的推理阶段。

#### A. 初步 CVAE

变分自动编码器 (VAE) 是一个强大的生成框架, 用于学习复杂数据的潜在分布 [15], [20], [31]。VAE 的生成过程如下: 编码器以数据  $x$  为输入, 输出一个潜在向量  $z$  的数据条件分布  $q_\phi(z|x)$ 。编码条件重构分布  $p_\theta$  中抽取样本  $z \sim p_\theta$ , 然后由解码器用于确定输入数据  $x$  上的分布  $p_\theta(x|z)$ 。VAEs 的目标是最大化  $p_\theta(x)$  的变分下界:

$$\log p_\theta(x) \geq -D_{\text{KL}}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{q_\phi(z|x)} \log p_\theta(x|z), \quad (1)$$

其中  $D_{\text{KL}}$  是 Kullback-Leibler (KL) 散度函数并且  $z = g_\phi(x, \epsilon)$ ,  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ 。

尽管 VAE 具有对潜在分布进行建模和保留干净图像的共同特征的先天能力, 但它只能接收和输出相同的雨天图像, 而不能从雨天输入中输出去雨图像。因此, VAE 不能直接应用于图像去雨。

最近, Sohn 等人已将 VAE 扩展为更强大的条件 VAE (CVAE), 它以边信息为条件对潜在变量和数据进行建模, 例如雨天图像  $x$  的干净图像  $y$ 。通过考虑干净图像  $y$  的条件信息, 我们可以重写下界方程 (1) 为:

$$\begin{aligned} \tilde{\mathcal{L}}_{\text{CVAE}} = & -D_{\text{KL}}(q_\phi(z|x, y)||p_\theta(z|x)) \\ & + \mathbb{E}_{q_\phi(z|x, y)} \log p_\theta(y|z, x), \end{aligned} \quad (2)$$

其中  $z = g_\phi(x, y, \epsilon)$ ,  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ 。在这里,  $p_\theta(z|x)$  假设为各向同性高斯分布并且  $p_\theta(x|y, z)$ , 同时  $q_\phi(z|x, y)$  是高斯分布。

自从被引入以来, CVAE 已经在各种计算机视觉任务中展示了其强大的能力, 例如轨迹预测 [36]、图像着色 [5]、图像生成 [7] 和多模态人体动态生成 [45] 等。

图像去雨是一个高度不适定的问题, 因为为真实世界的雨天图像定义最佳干净背景并非易事。CVAE 可以从输入的雨天图像中生成多个去雨图像的预测。因此, 通过整合这些预测可以获得更准确的去雨结果。在这项工作中, 据我们所知, 我们是最早探索 CVAE 模型对单幅图像去雨的生成能力的人之一。

#### B. 学习用于图像去雨的 CVID 网络

在这项工作中, 我们利用 CVAE 强大的生成能力作为我们的条件变分图像去雨 (CVID) 网络进行图像去雨的支柱。CAVE 主干基本上由编码器、先验网络和解码器组成, 如图 2 所示。我们将编码器和先验网络中的滤波器大小设置为 3, 卷积滤波器的数量设置为 16。在编码器和先验网络的最后一层, 前半部分是  $\mu$ , 后半部分是  $\sigma$ 。我们将卷积滤波器的数量设置为 1。对于我们的 CVID 网络, 我们将编码器、先验网络和解码器的深度设置为 7。我们采用 Leaky ReLU [39] 作为激活函数。每层后使用批归一化 [17]。

编码器以雨天图像  $x$  为条件, 学习对相应干净图像  $y$  的信息进行加密的潜在分布  $\mathcal{N}(\mu_e, \sigma_e)$ 。为了保证来自潜在分布的采样潜在变量  $z$  和输入  $x$  在推理过程中密切相关, 我们引入了一个先验网络 (图 3)。先验网络学习将雨天图像  $x$  映射到“先验”潜在分布  $\mathcal{N}(\mu_p, \sigma_p)$ ,

该潜在分布对雨天图像的分布信息进行编码。解码器的目标是基于来自“编码器”潜在分布  $\mathcal{N}(\mu_e, \sigma_e)$  的采样潜在变量  $z_e$  重建去雨图像  $\hat{y}$ ，也以雨图像  $x$  为条件。具体来说，解码器的输入是雨天图像、雨量密度估计图和采样后的  $z_e$ 。对于解码器，我们将滤波器大小设置为 3，将反卷积滤波器的数量设置为 16。为了方便地计算梯度，我们使用重新参数化技术 [20] 通过  $z = \mu(x) + \epsilon * \sigma(x)$  对潜在变量  $z$  进行采样，其中  $\epsilon$  是来自高斯分布  $\mathcal{N}(0, I)$  的采样噪声。

为了学习 CVAE 骨干网络，我们需要最大化方程 (2) 中定义的条件变分下界。方程 (2) 中的第一项充当正则化项以最小化数据条件分布  $q_\phi(z_e|x, y)$  和先验分布  $p_\theta(z_p|x)$  之间的差异。在这里，我们将 Kullback-Leibler (KL) 散度作为惩罚函数来最小化两个高斯分布  $q_\phi(z_e|x, y)$  和  $p_\theta(z_p|x)$  之间的差距。方程 (2) 中的第二项是测量采样的潜在编码  $z_e$  和干净图像  $y$  之间的信息损失的重建误差。我们最大化条件对数似然  $\mathbb{E}_{q_\phi(z_e|x, y)}[\log p_\theta(\hat{y}|x, z_e)]$  以进行准确重建。在实践中，误差可以计算为干净图像  $y$  和重建图像  $\hat{y}$  之间的  $\ell_2$  损失。

**CVAE 的损失函数.** 训练 CVAE 网络以最大化方程 (2) 中第二项的条件对数似然。由于这个目标函数是难以处理的，我们在方程 (2) 中最大化其变分下界。我们最小化数据条件分布  $q_\phi(z|x, y)$  和先验分布  $p_\theta(z|x)$  之间的 KL 散度，以减轻学习和推理阶段潜在变量编码之间的差异：

$$\mathcal{L}_{\text{KL}} = \sum_{i=1}^N q_\phi(z_i|x_i, y_i) \log\left(\frac{q_\phi(z_i|x_i, y_i)}{p_\theta(z_i|x_i)}\right), \quad (3)$$

其中  $q_\phi(z_i|x_i, y_i) = \mathcal{N}(\mu_e, \sigma_e)$ ， $p_\theta(z_i|x_i) = \mathcal{N}(\mu_p, \sigma_p)$ ， $N$  是训练图像的数量。

为了最大化用于重建  $x$  的  $\mathbb{E}_{q_\phi(z|x, y)}[\log p_\theta(y|x, z)]$ ，我们将损失  $\mathcal{L}_{\text{rec}}$  定义如下：

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{i=1}^N \sum_{c \in \{r, g, b\}} \|y_{i,c} - \hat{y}_{i,c}\|_F^2, \quad (4)$$

其中  $\hat{y}_{i,c} = f_c^{\text{rec}}(x_{i,c}, y_{i,c}, D_{i,c})$  是与第  $c$  个通道相关的 CVAE。CVAE 将雨图像  $x$ 、干净图像  $y$  的每个单独颜色通道和通道  $c$  (Fig. 2，这部分将在 §III-D 中解释) 中估计的雨密度图  $D_c$  作为输入，并输出  $c$  通道下的去雨的图像  $\hat{y}_c$ 。总的来说， $\mathcal{L}_{\text{CVAE}}$  损失函数是  $\mathcal{L}_{\text{KL}}$  和  $\mathcal{L}_{\text{rec}}$  之和：

$$\mathcal{L}_{\text{CVAE}} = \mathcal{L}_{\text{rec}} + \beta \mathcal{L}_{\text{KL}}, \quad (5)$$

其中  $\beta > 0$  是正则参数。

---

### Algorithm 1 学习用于图像去雨的 CVID 网络

---

**学习:** 输入成对的雨天和干净的图像  $\{x_i, y_i\}_{i=1}^N$   
 $\theta, \phi \leftarrow$  初始化参数

repeat

SDE:  $\hat{D}_{i,c} \leftarrow \text{SDE}_\theta(x_i)$

Encoder:  $\begin{cases} \mu_e, \sigma_e \leftarrow E_\phi(x_i, y_i, \hat{D}_{i,c}) \\ z_e \leftarrow \mu_e(x_i) + \epsilon * \sigma_e(x_i), \epsilon \sim \mathcal{N}(0, I) \end{cases}$

Prior:  $\begin{cases} \mu_p, \sigma_p \leftarrow P_\theta(x_i, \hat{D}_{i,c}) \\ z_p \leftarrow \mu_p(x_i) + \epsilon * \sigma_p(x_i), \epsilon \sim \mathcal{N}(0, I) \end{cases}$

Decoder:  $\hat{y}_i \leftarrow D_\theta(x_i, z, \hat{D}_{i,c})$

$g \leftarrow \nabla_{\theta, \phi} \mathcal{L}(\theta, \phi; x, y, \epsilon)$

$\theta, \phi \leftarrow$  Update parameters using gradients  $g$

until convergence

return  $\theta, \phi$

---

**推理:** 输入雨天图像  $x$

SDE:  $\hat{D}_c \leftarrow \text{SDE}_\theta(x)$

先验:  $\begin{cases} \mu_p, \sigma_p \leftarrow P_\theta(x, \hat{D}_c) \\ z_p \leftarrow \mu_p(x) + \epsilon * \sigma_p(x), \epsilon \sim \mathcal{N}(0, I) \end{cases}$

Decoder:  $\hat{y} \leftarrow \frac{1}{n} \sum_{j=1}^n D_\theta(x, z, \hat{D}_c)$

return 区域图像  $\hat{y}$

---

### C. 提出的空间密度估计模块

雨条纹通常在雨图像中分布不均。忽略降雨分布空间方差的方法不可避免地会产生不准确的去雨结果。尽管 [50] 通过将降雨强度分级为不同级别来考虑全球密度估计，但在局部地区仍然不可避免地会出现不准确的去雨结果。此外，由于雨条纹在雨图像中随机分布，很难在不同的图像中一致地定位雨区。

在这项工作中，我们为 CVAE 骨干网络提出了一个空间密度估计 (SDE) 模块，以学习输入雨天图像的密度估计图，并使其在空间上适应图像去雨。所提出的 SDE 模块被实现为一个具有五层的紧凑密集连接的卷积块 [16]。每层的输入是通过连接所有先前层的输出获得的。滤波器大小设置为 3 并且滤波器的数量为 16。每一个卷积层后都跟了一个批归一化层 [17]，使用 ReLU [25] 进行激活操作。对于最后一层，我们使用 Sigmoid 激活函数来制作  $[0, 1]$  内的密度估计图

密度估计图的学习是以完全监督的方式进行的。它将整个雨天图像作为输入，并为每个颜色通道输出一个密度估计图。具体来说，我们从对应的干净图像  $y$  (“真值图”) 中减去雨天图像  $x$ ，并产生一个残差图像，表示为  $\hat{R}e$ 。 $\hat{R}e_c$  代表  $\hat{R}e$  中的第  $c$  个色彩通道，其中  $c \in \{R, G, B\}$ ，并且  $\hat{R}e_c(x)$  表示每个通道上位置  $x$  的

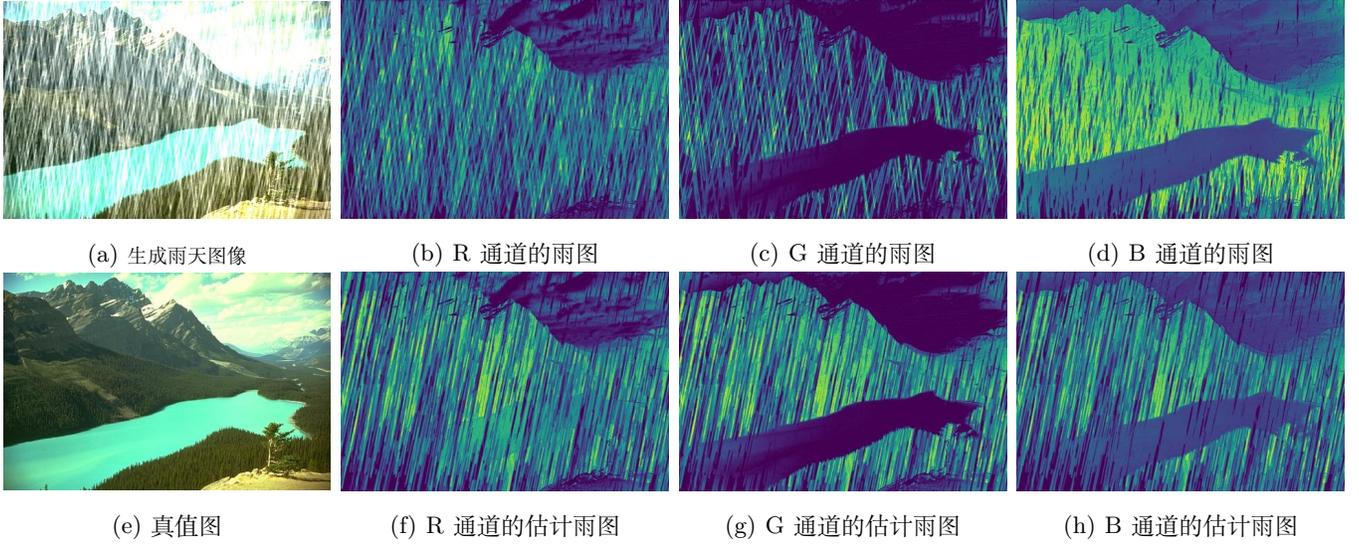


图 4: 分别在  $R$ 、 $G$ 、 $B$  通道上的雨图和估计雨密度图. (a) 和 (e) 分别是干净的和对应的雨天图像. (b), (c) 和 (d) 分别展示了  $R$ ,  $G$ ,  $B$  通道下的雨图. (f), (g) 和 (h) 分别是所提出的 SDE 模块对三个颜色通道的估计雨密度图.

像素值.  $\hat{R}e_c(x) = 0$  在  $x$  位置没有雨,  $\hat{R}e_c(x) \neq 0$  表示该位置的雨的强度. 基于残差图  $\hat{R}e_c$ , 我们使用  $D_c(x)$  生成用于密度估计图监督学习的地面实况图像,

$$D_c(x) = \begin{cases} 0 & \hat{R}e_c(x) = 0 \\ \sigma(\hat{R}e_c(x)) & \hat{R}e_c(x) \neq 0 \end{cases}, \quad (6)$$

其中  $D_c$  是  $c$  通道的真值,  $\sigma$  是 Sigmoid 激活函数.

我们在图 4(b), (c) 和 (d) 中分别为雨天图像的  $R$ ,  $G$ ,  $B$  通道绘制了  $D_c$  的示例. 可以看出, 雨条纹随机分布在空间位置和三个通道中. 主要原因是不同光源发出的光强度不同, 如阳光、白色地板、绿色水面. 我们还分别为雨天图像的  $R$ ,  $G$ ,  $B$  通道绘制了由图 4(f), (g) 和 (h) 中提出的 SDE 模块估计的密度图. 可以看出, 密度图与图 4(b), (c) 和 (d) 中所示的地面真值图非常接近. 这说明 SDE 模块可以准确定位三个颜色通道的降雨区域. 因此, 所提出的 SDE 模块嵌入 CVAE 网络可以通过密度图准确估计降雨强度来获得自适应去雨性能.

**SDE 的损失函数.** SDE 模块也以完全监督的方式进行培训. 它将雨天图像作为输入并估计每个颜色通道的密度图  $D_c$  ( $c \in \{R, G, B\}$ ). SDE 模块的损失函数如下:

$$\mathcal{L}_{\text{SDE}} = \frac{1}{N} \sum_{i=1}^N \sum_{c \in \{R, G, B\}} \|D_{i,c} - \hat{D}_{i,c}\|_F^2, \quad (7)$$

其中  $\hat{D}_{i,c} = f_c^D(x_{i,c})$  和  $f_c^D(\cdot)$  是与第  $c$  通道相关的 SDE 模块. 惩罚损失方程 (7) 旨在最小化估计密度图

和地面实况图之间的差异. 然后将获得的密度估计图输入到 CVAE 骨干网络, 以在不同的局部区域进行自适应去雨性能.

#### D. 提出的通道去雨方案

除了空间分布不均匀外, 不同颜色通道的降雨密度也有明显的分布. 以前的图像去雨方法在很大程度上忽略了这一点. 受明亮通道先验 (BCP) [10] 的启发, 我们提出了一种通道 (CW) 去雨方案, 以进一步提升 CVAE 骨干网络以进行图像去雨. 到目前为止, 我们提出的条件变分图像去雨 (CVID) 网络是通过使用 CVAE 主干和建议的 SDE 模块和 CW 去雨方案来实现的. BCP 先验 [10] 描述了在自然场景中, 对于每个像素, 至少存在一个高强度的颜色通道. 具体来说, BCP 先验定义为:

$$J^{\text{bright}}(x) = \max_{y \in \Omega(x)} (\max_{c \in \{R, G, B\}} J^c(y)), \quad (8)$$

其中  $J^c$  是图像  $J$  的第  $c$  色彩通道,  $\Omega(x)$  是一个以位置  $x$  为中心的局部补丁. 除非在光线不足或阴影占主导地位 [10] 的情况下,  $J^{\text{bright}}$  的强度应该接近 1 (强度取值为  $[0, 1]$ ). 依据 BCP 先验 (8), 我们提出了如下提议来验证我们的 CW 去雨策略 (证明在 §VI 中提供):

**Proposition 1.** 记  $\bar{B}$  和  $B$  分别为不区分和区分不同颜色通道的情况下的去雨图像.  $\|\cdot\|_0$  是  $\ell_0$  归一, 计算非零值的数量. 那么,  $\bar{B}$  中像素的强度远低于  $B$ . 即  $\bar{B}$

中最亮的像素数往往比  $B$  中的少。所以，我们有

$$\|1 - B\|_0 \leq \|1 - \bar{B}\|_0. \quad (9)$$

请注意，最亮像素的强度为 1，并且当且仅当每个通道上的雨条纹分布相同时，才能满足小于或等于符号。为了提供更直观的说明，在图 5 中，我们将我们方法的结果与以前的代表性方法进行比较，例如，DNN [9] 和 JORDER [47], [48] 这些不分离三个颜色通道的方法。从中间一行可以看出，DDN 和 JORDER 得到的去污图像的最亮像素的强度大多低于我们的 CVAE 网络，如红色圆圈所示。图 5（底行）显示了去雨图像的明亮通道的强度分布。可以看到，使用我们的通道方案的去雨图像包含比其他方案更多的最亮像素，证明了所提出的通道方案的有效性。单独处理每个颜色通道的优势在其他低级视觉任务中也得到了验证 [13], [42], [43]。

#### E. 优化

通过联合最小化负条件变分下界 (2) 和方程 (7) 中定义的 SDE 模块的损失来优化所提出的 CVID 网络。具体来说，我们将 CVID 网络的目标函数公式化为方程 (5) 中的 CVAE 损失和方程 (7) 中的 SDE 损失：

$$\mathcal{L}_{\text{CVID}} = \mathcal{L}_{\text{CVAE}} + \lambda \mathcal{L}_{\text{SDE}}, \quad (10)$$

其中  $\lambda > 0$  是平衡  $\mathcal{L}_{\text{SDE}}$  和  $\mathcal{L}_{\text{CVAE}}$  重要性的正则化参数。我们观察到，当我们平等对待  $\mathcal{L}_{\text{SDE}}$  和  $\mathcal{L}_{\text{CVAE}}$  时，即  $\lambda = 1$  时，CVID 不断获得令人满意的性能。在我们的 CVID 网络中，CVAE 损失  $\mathcal{L}_{\text{CVAE}}$  (5) 和 SDE 损失  $\mathcal{L}_{\text{SDE}}$  (7) 通过梯度下降以端到端的方式通过反向误差传播共同最小化。

#### F. 推理阶段

为了在推理过程中获得确定性输出，我们从先验网络学到的先验分布  $p_\theta(z_p|x)$  中提取  $n$  个潜在代码  $\{z_j^p\}_{j=1}^n$ ，并简单地取  $n$  个后验的平均值作为最终预测。具体来说，我们使用蒙特卡罗方法 [33] 计算潜在干净图像  $y$  的边际似然：

$$p_\theta(y|x) \approx \frac{1}{n} \sum_{j=1}^n p_\theta(y|x, z_j^p), z_j^p \sim p_\theta(z_p|x). \quad (11)$$

对于 (2) 的前两项，我们使用蒙特卡罗抽样来估计其条件对数似然 (CLL)。初始实验表明 100 个样本足以获得 CLL 的准确估计。在算法 1 中，我们总结了所提出的用于图像去雨的 CVID 网络的学习和推理过程。

## IV. 实验

在本节中，我们进行了大量实验来证明所提出的条件变分图像去雨 (CVID) 网络对图像去雨的有效性。还进行了全面的消融研究以验证不同组件的有效性。补充文件中提供了更多结果。

#### A. 实现细节

在我们的 CVID 网络中，我们设置  $\beta = 0.1$ ,  $\lambda = 1$ ,  $n = 100$ 。对于网络训练，我们从每个训练集中随机生成 2,000 对大小为  $64 \times 64$  的图像块。我们使用带有默认参数的 Adam 优化器，权重衰减为  $10^{-10}$ ，小批量大小为 32。学习率初始化为 0.01，并在每个 epoch 除以 10，轮数为 4。

#### B. 实验方案

**数据集。**我们在 3 个合成数据集和 1 个真实世界数据集上进行实验。第一个合成数据集提供于 [9] 并包含 14,000 个合成的干净/雨天图像对。按照 [50] 中的设置，13000 张图片用于学习，其余 1000 张图片用于测试（记作 D1）。第二个合成数据集提供于 [48]，由 1800 对大雨图像和 200 对小雨图像组成，供学习使用。两组 (Rain100L 和 Rain100H) 用于测试（记作 D2）。第三个合成数据集包含 12,000 个合成的晴雨图像对，其中包括 4,000 张大雨图像、4,000 张中雨图像、4,000 张小雨图像。用于测试的 1,200 对干净/雨天图像表示为 D3。据我们所知，这是对这三个数据集进行实验评估的第一项工作。我们测试的真实下雨图像来自真实世界的数据集 [49]，该数据集是作者从网上下载的。

**评估指标。**我们采用三个常用指标，即峰值信噪比 (PSNR)、结构相似性指数 (SSIM) [37] 和基于感知的指标 NIQE [27] 来评估合成数据集上的去雨性能。由于现实世界的雨天图像没有“地面真值”图像，我们仅通过竞争方法比较去雨图像的视觉质量。

#### C. 与最先进技术的比较

**对比方法。**我们将提出的 CVID 网络与 6 种最先进的图像去雨方法进行比较，包括高斯混合模型 (GMM) [23]、深度细节网络 (DDN) [9]、联合雨检测和去除 (JORDER) [48]、使用条件生成对抗网络的图像去雨 (ID-GAN) [51]、密度感知去雨 (DID-MDN) [50] 和循环挤压和激发上下文聚合网络 (RESCAN) [22]。

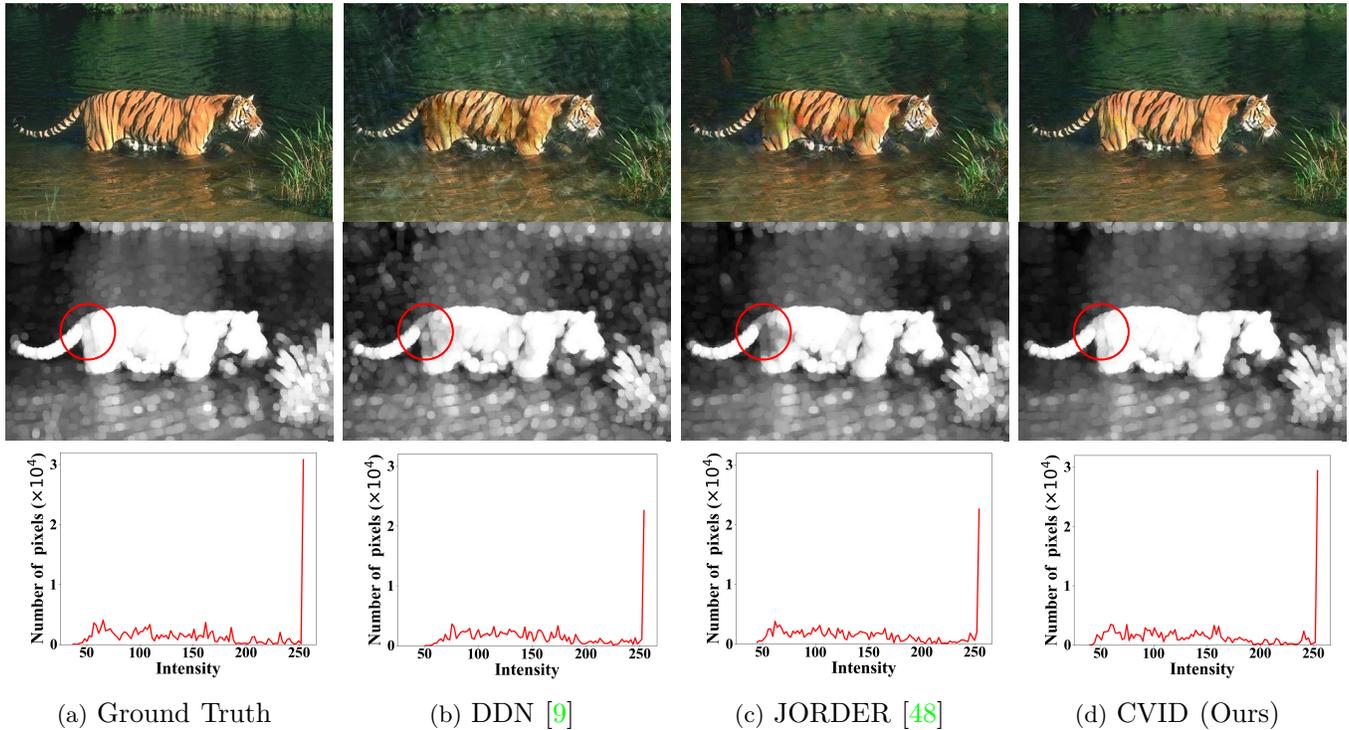


图 5: 顶部: 真值图 (a), 不同方法的区域图像 (b)-(d)。中部: 对应亮度通道 底部: 去雨图像明亮通道的强度分布。

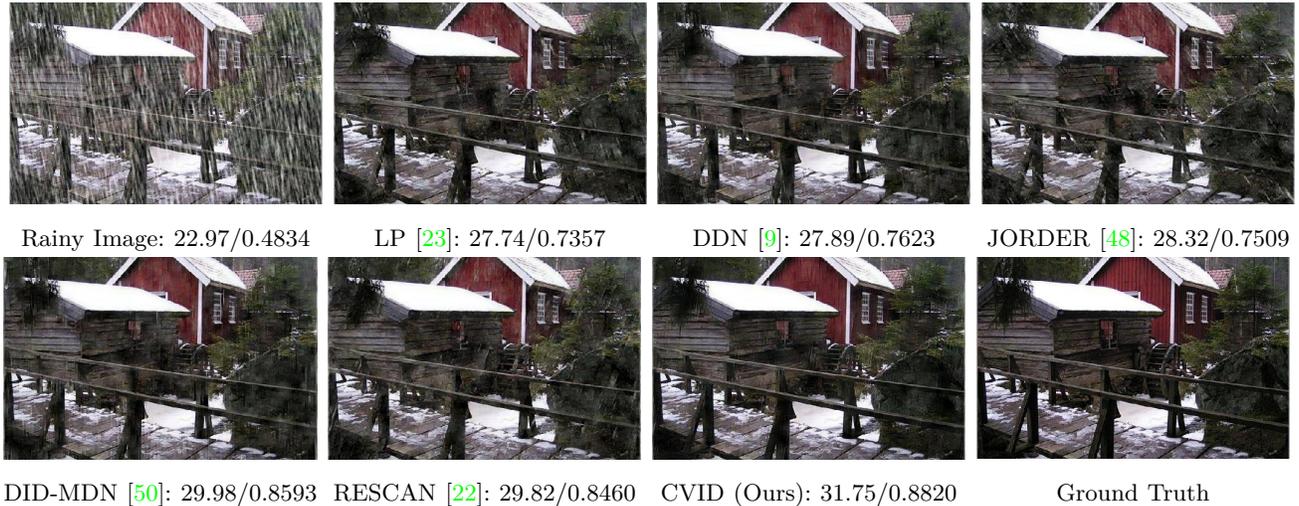


图 6: 在来自 D1 [9] 的合成雨天图像上通过不同方法得到的去雨图像和 PSNR (dB)/SSIM.

**合成除雨的结果.** 定量比较见表 I。我们的 CVID 网络在所有三个数据集上都大大超过了以前的方法。特别是, 在 D1 上, 我们的方法在 PSNR、SSIM 和 NIQE 方面分别优于第二好的方法, 即 ID-GAN [51] 2.42 dB、0.0233 和 0.44。卓越的性能证明了我们的方法对单幅图像去雨的巨大有效性。在图 6 和 7, 我们通过不同的方法比较了去雨图像以及 PSNR/SSIM 结果。我们观察到, 我们的 CVID 更清晰地去除了雨痕, 同时比以前的方法更好地保留了图像细节。

**实际除雨的结果.** 在处理现实世界的雨天图像时, 在合成雨天图像上学习的去雨方法将遭受巨大的性能下降, 这是一个常见的挑战。然而, 这个问题在很大程度上被引入的 CVAE 框架的专有生成特性所缓解, 它可以为一个下雨输入输出多个候选解决方案。为此, 我们将提出的 CVID 网络应用于去除真实雨天照片中的雨痕。提出的 CVAE 网络是在 DID-MDN [50] 中使用的训练集上学习的。我们使用 [49] 中的下雨图像, 包括 4 个不同的代表性场景 (在图 8 (a) 中展示): 小雨、中雨、大



图 7: 在来自 D2 [48] 的合成雨天图像上通过不同方法得到的去雨图像和 PSNR (dB)/SSIM

Dataset	D1 [9]			D2 Rain100L [48]			D2 Rain100H [48]			D3 [50]		
	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓
Rainy Input	19.31	0.7695	9.17	23.52	0.8332	7.93	12.13	0.3702	14.27	21.15	0.7781	10.15
GMM [23]	24.35	0.8312	4.83	32.02	0.9137	3.14	14.26	0.5444	7.76	25.23	0.8514	5.37
JORDER [48]	22.36	0.8405	4.97	36.02	0.9712	2.74	23.45	0.7382	5.39	24.32	0.8622	5.64
DDN [9]	25.63	0.8851	3.92	33.75	0.9213	3.09	22.26	0.6928	6.89	27.33	0.8978	4.98
ID-GAN [51]	26.31	0.8932	3.43	35.89	0.9631	2.96	23.15	0.7120	6.33	28.18	0.9102	4.01
DID-MDN [50]	26.07	0.9092	3.47	35.73	0.9602	3.02	23.25	0.7315	6.18	27.95	0.9087	3.71
RESCAN [22]	25.45	0.8812	3.90	37.27	0.9813	2.39	26.45	0.8458	4.96	26.19	0.8712	4.01
CVID (Ours)	28.96	0.9375	2.99	37.83	0.9882	2.13	27.93	0.8765	4.71	30.97	0.9374	3.33

表 I: 在三个合成数据集上通过不同方法比较 PSNR (dB)、SSIM [37] 和 NIQE [37] 结果。

Image Size	DDN [9]	JORDER [47]	RESCAN [22]	CVID <sub>n=1</sub>	CVID <sub>n=10</sub>	CVID <sub>n=100</sub>
500 × 500	0.41	0.18	0.45	0.12	0.35	0.82
1024 × 1024	0.76	0.82	1.81	0.42	0.68	1.38

表 II: 在不同大小的雨天图像上不同方法的运行时间 (以秒为单位)。

雨、雪 (从上到下)。如图 8 (b)-(f) 所示, 我们的 CVID 在各种真实场景中的表现都优于之前的竞争对手。补充文件中提供了更多结果

**速度。** 速度对比结果见表 II。我们观察到, 我们使用  $n = 1$  生成样本的 CVID 比其他方法快, 而使用  $n = 100$  的 CVID 在大小为  $500 \times 500$  的图像上比其他方法慢, 但在大小为  $1024 \times 1024$  的图像上仍然比 RESCAN [22] 快。

#### D. 提出的 CVID 网络的验证

为了进一步验证我们的 CVID 网络的工作机制, 我们对雨天图像数据集 D2 [48] 进行了更深入的分析。具体来说, 我们评估的内容包括 1) 采用的 CVAE 框架的重要性; 2) 提出的通道去雨方案的效果; 3) 提出的 SDE 模块对 CVID 网络的影响; 4) 样本数  $n$  如何影

响我们的 CVID 网络的性能; 5) 超参数  $\beta$  and  $\lambda$  的影响; 6) 我们的 CVID 网络中的这些组件对累积误差分布 (CED) 的影响; 以及 7) 在我们的 CVID 中选择 ReLU 还是 PReLU。在所有实验中, 由于其简单的网络架构, DDN [9] 都被用作基线。

1) 所采用的 CVAE 框架对图像去雨的重要性如何? 为了评估我们采用的 CVAE 主干网络对图像去雨的重要性, 我们比较了基线 DDN 网络 [9] 和我们的 CVAE 主干网络。如表 III 所示 (第二列和第三列), CVAE 在 PSNR/SSIM 上的结果为 35.38 dB/0.9574, 远高于基线 DDN [9] 的结果 (33.75 dB/0.9213)。更重要的是, 当比较 DDN 和 CVAE 的通道变体时, 我们观察到通道 CVAE 实现了 36.03 dB/0.9627 的 PSNR/SSIM 结果, 仍然比通道 DDN (34.83 dB/0.9493)。在视觉质量上, 从图 9 (b) 和 (c) 中, 我们观察到我们的 CVAE

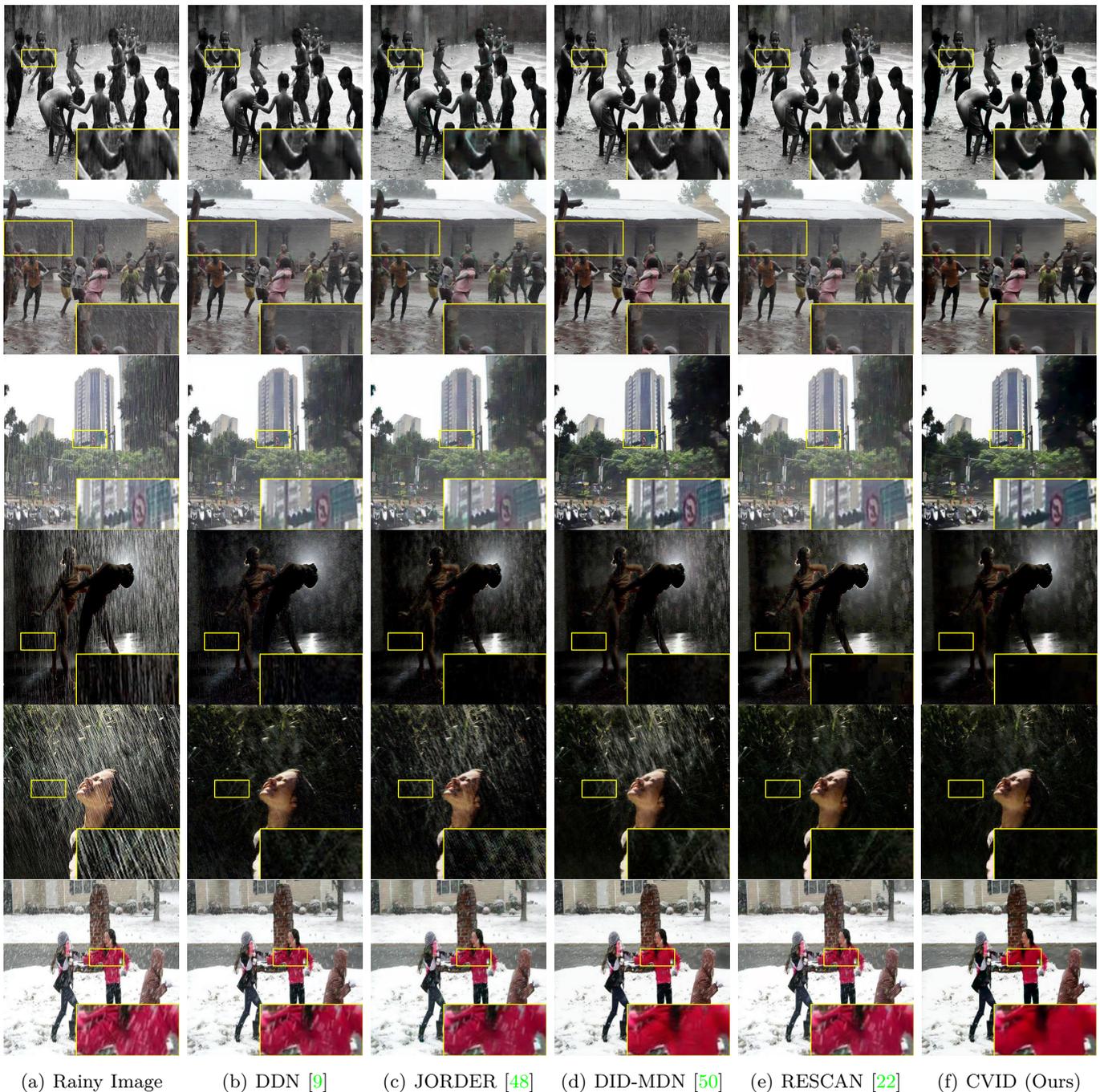


图 8: 通过不同方法对来自数据集的五种具有代表性的真实雨景场景的去雨图像 [49]: 分别是小雨、中雨、大雨、夜间、深色背景和雪 (从上到下)。

主干比 DDN [9] 获得了更清晰的结果。所有这些结果都证明了采用 CVAE 主干对 ResNet 主干 [14] 上的图像去雨的重要性。

2) 提议的通道方式 (CW) 去雨方案的效果如何? 为了研究所提出的 CW 方案的效果, 我们将其嵌入到 DDN 网络 [9] 和我们的 CVAE 骨干网中, 我们分别称它们为 +cw 和 CVAE+cw。从表 III (第 4 到第 6 列), 我们观察到, 通过我们提出的 CW 去雨方案, PSNR/SSIM 上的

基线 DDN 结果从 33.75 dB/0.9213 显著提高到 34.83 dB/0.9493(DDN+cw)。同样, 使用 CW, 我们的 CVAE 主干在 PSNR/SSIM 上的结果也从 35.38 dB/0.9574 提高到 36.03 dB/0.9627(CVAE+cw)。对比图 9(b) 和 (d), (c) 和 (f), 我们观察到 DNN+cw 和 CVAE+cw 分别比基线 DDN [9] 和我们的 CVAE 主干获得更好的视觉质量。所有这些结果清楚地证明了我们的 CW 去雨方案的有效性。

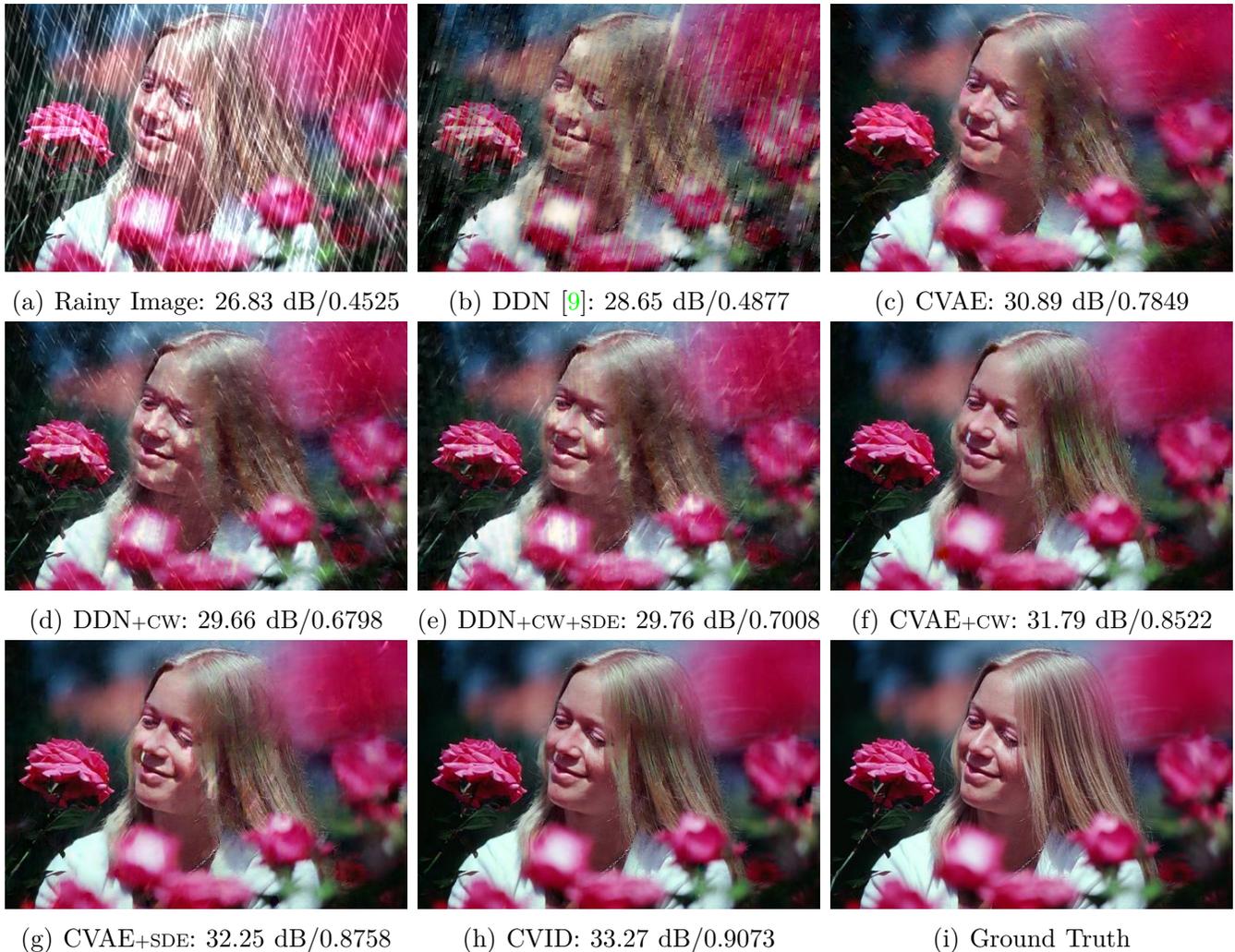


图 9: 在来自 D2Rain100H [48] 的合成下雨图像上 DDN [9] 的不同变体和提出的 CVAE 主干的去雨图像和 PSNR/SSIM 结果。我们的 CVID 网络 (CVAE+SDE+cw) 是 CVAE 主干、所提出的空间密度估计 (SDE) 模块和通道 (CW) 去雨方案的集成。

3) 提出的 SDE 模块如何影响我们的 CVID 网络? 为了回答这个问题, 我们将提议的 SDE 模块嵌入到变体 DDN+cw、提议的 CVAE 主干和变体 CVAE+cw 中, 产生的变体称为 DDN+cw+SDE、CVAE+SDE 和 CVAE+cw+SDE (CVID)。这些变体在数据集 D2 (Rain100L 和 Rain100H) 上的对应结果列于表 III (第 5、7、8 列)。可以看出, 配备我们的 SDE 模块的所有这些变体的性能明显优于相应的基线。例如, DDN+cw+SDE 在 PSNR/SSIM 上达到 35.78 dB/0.9596, 比 DDN+cw 高 0.95 dB/0.0103。在视觉质量方面, 从图 9(c) 和 (g)、(d) 和 (e)、(f) 和 (h) 中, 我们观察到我们的 SDE 模块持续提高了几个变体的性能。这些改进验证了所提出的 SDE 模块确实在很大程度上提高了我们的基线。

4) 我们的 CVID 网络如何在更大的样本数  $n$  下执行? 为了研究这一点, 我们在 D2 [48] 中的 Rain100L 和 Rain100H 数据集上运行具有不同样本数  $n$  的 CVID 网络。表 IV 中列出了 PSNR (dB) 的结果。我们观察到我们的 CVID: 1) 在  $n = 1$  的情况下, RESCAN [22] 的结果较差, 在 Rain100L 和 Rain100H 数据集上分别达到 37.27 dB/0.9813 和 26.45 dB/0.8458; 2) 在  $n = 5, 10, 100, 200$  生成的图像时始终表现更好, 但在  $n = 200$  处收敛, 当  $n = 300, 500$  时变得更糟; 3) 在  $n = 1$  时比 RESCAN 快, 但需要更多的运行时间和更多的样本数, 如表 II 所示。这再次表明, 与比较方法采用的确定性框架相比, 我们的 CVAE 框架在精度-速度权衡方面更加灵活。

5) 参数  $\beta$  和  $\lambda$  如何影响我们的 CVID 网络的性能? 我

	DDN [9]	CVAE	DDN+cw	DDN+cw+SDE	CVAE+cw	CVAE+SDE	CVID
CVAE	✗	✓	✗	✗	✓	✓	✓
SDE	✗	✗	✗	✓	✗	✓	✓
CW	✗	✗	✓	✓	✓	✗	✓
D1	25.63/0.8851	26.57/0.8994	25.96/0.8901	26.08/0.8932	27.38/0.9138	27.69/0.9193	28.96/0.9375
Rain100L	33.75/0.9213	35.38/0.9574	34.83/0.9493	35.78/0.9596	36.03/0.9627	36.79/0.9783	37.83/0.9882
Rain100H	22.26/0.6928	25.25/0.7738	24.39/0.7637	26.11/0.8157	26.75/0.8332	27.04/0.8485	27.89/0.8721
D3	27.33/0.8978	27.91/0.9073	27.55/0.9006	27.87/0.9065	29.19/0.9198	29.38/0.9207	30.97/0.9374

表 III: 在数据集 D1 [9], D2 [48] 和 D2 [48] 上对 PSNR (dB) 和 SSIM 进行不同变体的定量比较。CW 表示通道方案, SDE 表示空间密度估计模块。我们的 CVID 也是“CVAE+cw+SDE”。

	CVID <sub>n=1</sub>	CVID <sub>n=5</sub>	CVID <sub>n=10</sub>	CVID <sub>n=100</sub>	CVID <sub>n=200</sub>	CVID <sub>n=300</sub>	CVID <sub>n=500</sub>
Rain100L	36.91	37.19	37.75	37.83	37.87	37.84	37.81
Rain100H	25.73	26.18	27.38	27.89	27.95	27.91	27.88

表 IV: 在 D2 上使用样本数为  $n = 1, 5, 10, 100, 200, 300, 500$  的 CVID 网络的 PSNR (dB) 结果 [48]

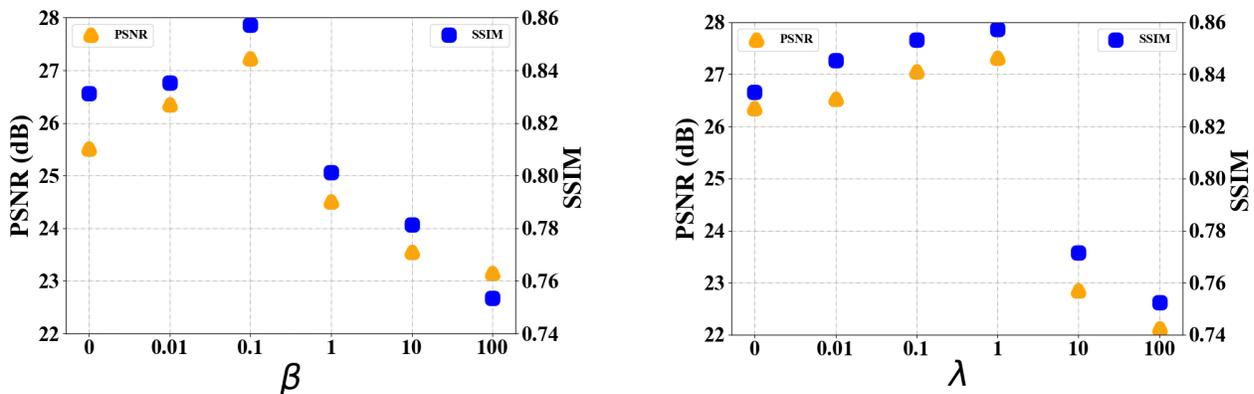


图 10: 我们的 CVID 网络在 D2 Rain100H [48] 上具有不同的  $\beta$ (左) 和  $\lambda$  (右) 的 PSNR (dB) 和 SSIM 结果

	CVID with ReLU	CVID with Leaky ReLU
Rain100L	37.13/0.9817	37.83/0.9882
Rain100H	26.97/0.8597	27.89/0.8721

表 V: 我们的 CVID 在 D2 数据集 [48] 上具有不同激活函数的 PSNR (dB) 和 SSIM [37] 结果

我们的 CVID 有两个超参数  $\beta$  and  $\lambda$ , 分别平衡了方程 (5) 中 KL 散度损失和方程 (7) 中 SDE 损失的重要性。在我们的 CVID 网络中, 我们设置  $\beta = 0.1, \lambda = 1$ 。为了研究它们对我们的 CVID 的影响, 我们在 D2 Rain100H 数据集 [48] 上进行了实验。我们一次改变一个参数, 同时维持另一个不变。结果如图 10 所示。我们观察到当  $\beta$  从 0 增加到 0.1 时, 我们的 CVID 表现更好, 但当  $\beta = 1, 10, 100$  时表现更差。在参数  $\lambda$  对我们的 CVID 的影响上可以发现类似的趋势

6) 我们的 CVID 网络在累积误差分布 (CED) 上的性能。为了进一步验证我们的 CVID 网络的有效性, 我们采用 CED 作为 PSNR 和 SSIM 的补充方式进行性能测量。在这里, CED 计算两个图像之间绝对强度误差的累积直方图曲线。较早的饱和曲线表示两个图像的更准确和近似。在图 11 中, 我们根据基线 DDN [9] 和 CVAE 主干的不同变体, 根据去污图像和地面实况之间的绝对像素强度误差绘制 CED 曲线。我们观察到我们的 CVID 网络的曲线比 DDN [9] 和我们的 CVAE 主干的那些变体更早饱和。此外, 我们的 CVID 网络的曲线在所有竞争变体中实现了最低的误差均值和方差, 再次证明了我们的 CVID 网络对单个图像去雨的有效性。

7) 在我们的 CVID 中选择 ReLU 还是 Leaky ReLU。我们在 CVID 网络中使用 Leaky ReLU [39], 因为 Leaky

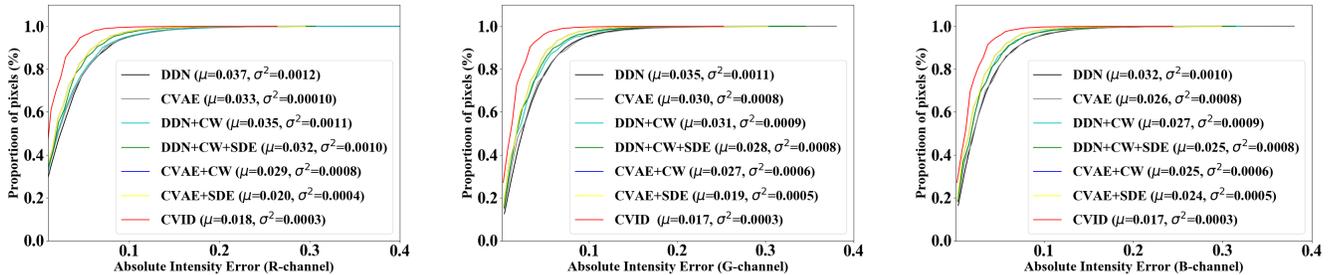


图 11: 以绝对强度误差表示的累积误差分布 (CED) 曲线图示, 以及它们分别在 R (左)、G (中) 和 B (右) 通道上的均值 ( $\mu$ ) 和方差 ( $\sigma^2$ )。

ReLU 在非线形激活方面通常比普通 ReLU [25] 更好。表 V 中列出了我们的 CVID 与 ReLU 和 Leaky ReLU 操作的比较。我们观察到结果也支持在我们的 CVID 网络中选择 Leaky ReLU 来进行图像去雨任务。

## V. 总结

在本文中, 我们提出了一个条件变分图像去雨 (CVID) 网络来解决图像去雨问题。CVID 利用条件变分自动编码器 (CVAE) 框架的强大生成能力对干净图像先验的潜在分布进行建模, 从中生成多个去雨图像用于图像去雨。此外, 我们提出了空间密度估计模块和通道方式去雨方案, 以在不同颜色通道中实现更自适应的图像去雨。开发了空间密度估计模块以提高对不均匀雨天图像的空间自适应去雨性能。在合成数据集和真实数据集上的实验表明, 我们的 CVID 网络始终比以前最先进的图像去雨方法获得更好的性能。

## VI. 附录

**命题 1 的证明:** 记  $\bar{B}$  ( $B$ ) 和  $\bar{R}$  ( $R$ ) 为去雨图像和去除雨条纹层, 分别没有 (有) 区分不同的颜色通道。我们有

$$O_c = B_c + R_c, \quad O_c = \bar{B}_c + \bar{R}_c, \quad (12)$$

其中  $O$ ,  $B$ , 和  $R$  表示下雨图像、干净的背景图像和雨条纹层,  $c \in \{R, G, B\}$  是颜色通道索引。从 (12), 我们有:

$$\bar{B}_c - B_c = R_c - \bar{R}_c \quad (13)$$

在之前不区分颜色通道的去雨模型中, 灰色雨条纹  $R$  被简单地添加到干净的 RGB 图像上, 这导致三个颜色通道的雨条纹密度分布相同。然而, 在我们的模型中, 我们通过为每个颜色通道指定雨条纹  $R_c$  来分别处理三个颜色通道, 因为雨条纹明显地分布在三个颜色通

道中。因此, 它认为  $R_c(y) \leq \bar{R}_c(y)$ 。之后我们可以得到  $B_c \geq \bar{B}_c$ , 并且自然地有  $\max_{y \in \Omega(x)} (B_c(y)) \geq \max_{y \in \Omega(x)} (\bar{B}_c(y))$ , 这也就导致:

$$\max_{y \in \Omega(x)} (\max_{c \in \{R, G, B\}} B_c(y)) \geq \max_{y \in \Omega(x)} (\max_{c \in \{R, G, B\}} \bar{B}_c(y)) \quad (14)$$

$$J^{bright}(B_c)(x) \geq J^{bright}(\bar{B}_c)(x), \quad (15)$$

这表明, 在没有通道区分的情况下被去除的图像的明亮通道的强度将低于按通道去除的图像的强度。

结论是, 区分颜色通道去雨图像最亮像素的数量, 大于没有区分颜色通道的图像的数量, 即,

$$\|1 - B_c\| \geq \|1 - \bar{B}_c\|_0 \quad (16)$$

和

$$\|1 - B\|_0 \leq \|1 - \bar{B}\|_0, \quad (17)$$

证明结束。

## 参考文献

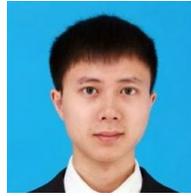
- [1] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. CVAE-GAN: fine-grained image generation through asymmetric training. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 2745–2754, 2017. 3
- [2] P. C. Barnum, S. Narasimhan, and T. Kanade. Analysis of rain and snow in frequency space. International Journal of Computer Vision, 86(2):256, Jan 2010. 1
- [3] Y. Chang, L. Yan, and S. Zhong. Transformed low-rank model for line pattern noise removal. In IEEE International Conference on Computer Vision (ICCV), pages 1726–1734, 2017. 1, 2
- [4] Y.-L. Chen and C.-T. Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In IEEE International Conference on Computer Vision (ICCV), pages 1968–1975, 2013. 1, 2
- [5] A. Deshpande, J. Lu, M.-C. Yeh, J. M. Chong, and D. Forsyth. Learning diverse image colorization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 6837–6845, 2017. 4

- [6] Y. Du, J. Xu, Q. Qiu, X. Zhen, and L. Zhang. Variational image deraining. In *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2020. 1
- [7] P. Esser, E. Sutter, and B. Ommer. A variational u-net for conditional appearance and shape generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8857–8866, 2018. 4
- [8] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. 1, 2
- [9] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3855–3863, 2017. 1, 2, 7, 8, 9, 10, 11, 12
- [10] X. Fu, D. Zeng, Y. Huang, X. Ding, and X.-P. Zhang. A variational framework for single low light image enhancement using bright channel prior. In *IEEE Global Conference on Signal and Information Processing*, pages 1085–1088, 2013. 6
- [11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *In Advances in Neural Information Processing Systems (NeurIPS)*, pages 2672–2680, 2014. 1
- [12] C. Ham, A. Raj, V. Cartillier, and I. Essa. Variational image inpainting. In *NeurIPS workshop on Bayesian Deep Learning*, 2018. 3
- [13] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010. 7
- [14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 10
- [15] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013. 4
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017. 5
- [17] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning (ICML)*, pages 448–456, 2015. 4, 5
- [18] X. Jin, Z. Chen, and W. Li. Ai-gan: Asynchronous interactive generative adversarial network for single image rain removal. *Pattern Recognition*, 100:107143, 2020. 1
- [19] L.-W. Kang, C.-W. Lin, and Y.-H. Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4):1742–1755, 2012. 1, 2
- [20] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 4, 5
- [21] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. A. Eslami, D. J. Rezende, and O. Ronneberger. A probabilistic u-net for segmentation of ambiguous images. In *Advances in Neural Information Processing Systems*, pages 6965–6975, 2018. 3
- [22] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *European Conference on Computer Vision (ECCV)*, pages 254–269, 2018. 1, 2, 7, 8, 9, 10, 11
- [23] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2736–2744, 2016. 1, 2, 7, 8, 9
- [24] Y. Luo, Y. Xu, and H. Ji. Removing rain from a single image via discriminative sparse coding. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3397–3405, 2015. 1, 2
- [25] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *International Conference on Machine Learning (ICML)*, 2013. 5, 13
- [26] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 1
- [27] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 22(3):209–212, 2013. 7
- [28] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2482–2491, 2018. 1, 2, 3
- [29] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2
- [30] D. Ren, W. Zuo, D. Zhang, L. Zhang, and M.-H. Yang. Simultaneous fidelity and regularization learning for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2
- [31] D. J. Rezende, S. Mohamed, and D. Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014. 4
- [32] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3
- [33] R. Y. Rubinstein. *Simulation and the Monte Carlo Method*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1981. 2, 4, 7
- [34] K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. In *In Advances in Neural Information Processing Systems (NeurIPS)*, pages 3483–3491, 2015. 1, 2, 3
- [35] H. Sun, Y. Du, J. Xu, Y. Yin, X. Zhen, and L. Shao. Learning to learn kernels with variational random features, 2020. 3
- [36] J. Walker, C. Doersch, A. Gupta, and M. Hebert. An uncertain future: Forecasting from static images using variational autoencoders. In *European Conference on Computer Vision (ECCV)*, pages 835–851. Springer, 2016. 4
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7, 9, 12
- [38] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3877–3886, 2019. 2

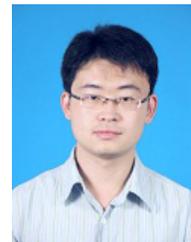
- [39] B. Xu, N. Wang, T. Chen, and M. Li. Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853, 2015. 4, 12
- [40] J. Xu, Y. Hou, D. Ren, L. Liu, F. Zhu, M. Yu, H. Wang, and L. Shao. Star: A structure and texture aware retinex model. *IEEE Transactions on Image Processing*, 29:5022–5037, 2020. 1
- [41] J. Xu, L. Zhang, and D. Zhang. External prior guided internal prior learning for real-world noisy image denoising. *IEEE Transactions on Image Processing*, 27(6):2996–3010, June 2018. 1
- [42] J. Xu, L. Zhang, and D. Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In *European Conference on Computer Vision (ECCV)*, September 2018. 7
- [43] J. Xu, L. Zhang, D. Zhang, and X. Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 7
- [44] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In *IEEE International Conference on Computer Vision (ICCV)*, pages 244–252, 2015. 1
- [45] X. Yan, A. Rastogi, R. Villegas, K. Sunkavalli, E. Shechtman, S. Hadap, E. Yumer, and H. Lee. Mt-vae: Learning motion transformations to generate multimodal human dynamics. In *European Conference on Computer Vision (ECCV)*, pages 265–281, 2018. 4
- [46] W. Yang, J. Liu, S. Yang, and Z. Guo. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE Transactions on Image Processing*, 28(6):2948–2961, 2019. 1
- [47] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1357–1366, 2017. 2, 7, 9
- [48] W. Yang, R. T. Tan, J. Feng, J. Liu, S. Yan, and Z. Guo. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2, 7, 8, 9, 10, 11, 12
- [49] H. Zhang and V. M. Patel. Convolutional sparse and low-rank coding-based rain streak removal. In *WACV*, pages 1259–1267. *IEEE*, 2017. 7, 8, 10
- [50] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 695–704, 2018. 2, 5, 7, 8, 9, 10
- [51] H. Zhang, V. Sindagi, and V. M. Patel. Image deraining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019. 1, 2, 3, 7, 8, 9
- [52] L. Zhu, C. Fu, D. Lischinski, and P. Heng. Joint bi-layer optimization for single-image rain streak removal. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2545–2553, 2017. 1



Yingjun Du received the B.E. degree from Hainan University, Hainan, China in 2016 and his Master degree from the College of Software in Beihang University, Beijing, China in 2020. He is currently pursuing his PhD degree in Informatics Institute, University of Amsterdam, The Netherlands. His research interests include Machine Learning and Computer Vision.



Jun Xu is an Assistant Professor in College of Computer Science, Nankai University, Tianjin, China. He received the B.Sc. degree in Pure Mathematics and the M. Sc. Degree in Information and Probability both from the School of Mathematics Science, Nankai University, China, in 2011 and 2014, respectively. He received the Ph.D. degree in the Department of Computing, The Hong Kong Polytechnic University in 2018. He worked as a Research Scientist in Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, UAE. More information can be found on his homepage <https://csjunxu.github.io/>.



Xiantong Zhen received the B.S. and M.E. degrees from Lanzhou University, Lanzhou, China in 2007 and 2010, respectively and the Ph.D. degree from the Department of Electronic and Electrical Engineering, the University of Sheffield, UK in 2013. He worked as postdoctoral fellows with the University of Western Ontario, London, Canada and the University of Texas at Arlington, Texas, U.S.A. from 2013 to 2017. He is with Inception Institute of Artificial Intelligence (IIAI) in the United Arab Emirates and College of Computer Science, Guangdong University of Petrochemical Technology, China. His research interests include Machine Learning and Computer Vision.



Ming-Ming Cheng received his PhD degree from Tsinghua University in 2012. Then he did 2 years research fellow, with Prof. Philip Torr in Oxford. He is now a professor at Nankai University, leading the Media Computing Lab. His research interests includes computer graphics, computer vision, and image processing. He received research awards including ACM China Rising Star Award, IBM Global SUR Award, CCF-Intel Young Faculty Researcher Program, etc. He is on the editor board of *IEEE Transaction on Image Processing*.



Ling Shao is the Executive Vice President and Provost of the Mohamed bin Zayed University of Artificial Intelligence. He is also the CEO and Chief Scientist of the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, United Arab Emirates. His research interests include computer vision, machine learning, and medical imaging. He is a fellow of IAPR, IET, and BCS.