单幅图像的暗通道辅助景深

Moushumi Medhi and Rajiv Ranjan Sahay

***Abstract*—在本论文中，我们利用暗通道作为补充线索来估计其由于有效隐式捕捉模糊图像和场景结构的局部统计，从单个空间变化的散焦模糊图像中提取深度信息。现有的基于散焦深度（DFD）技术通常依赖于具有不同光圈或焦点设置的多张图像来恢复深度信息。由于问题的不确定性，关于从单个散焦图像中提取深度信息的研究很少。我们的方法利用局部散焦模糊与对比度变化之间的关系作为关键深度线索，以增强场景结构估计的整体性能。整个流程是以端到端的方式对抗训练的。在真实数据上进行的实验中，通过真实的深度诱发散焦模糊，证明将暗通道先验引入到单图像 DFD 中可以产生有意义的深度估计结果，从而验证了我们方法的有效性。**

***Index Terms*—Depth-from-defocus, dark channel, local variation map.**

单输入的

I. 引言

D由散焦计算深度 的目标是从单张失焦图像中估计场景深度。单张失焦图像由系统或机器人即时捕获而不依赖自动对焦，可以快速提供深度线索。可以利用由于光学限制自然产生的模糊作为优势，从而能够在传统的全焦方法失效的情况下提取深度。本文提出了一种新颖的方法，从单个使用固定光圈设置拍摄的失焦模糊图像中估计深度。现有的失焦深度（DFD）方法往往依赖于收集多个具有不同光圈或焦点配置的图像来推导深度信息。这些方法利用了在不同焦点设置的图像之间观察到的失焦关系。例如，[1] 共同训练两个网络，DefocusNet 和 FocusNet，其中 DefocusNet 处理失焦图像以预测深度，之后与输入的全焦（AIF）图像结合生成合成焦点堆栈。然后，FocusNet 从这个焦点堆栈中估计深度，其输出与 AIF 图像结合以重建失焦图像。在训练期间，网络利用深度和失焦图像一致性损失进行自我监督，但在推理时，可以仅从一个失焦图像或从焦点堆栈中进行深度估计。与这些依赖视频序列、训练或推理期间多个帧或多个线索融合及传统优化技术的工作不同，我们的研究探索了基于深度学习和一种新颖的暗通道方法，来解决单图像 DFD 的不适定问题，其中在训练和测试期间都仅使用单个失焦图像。这一区别至关重要，因为我们的方法是为仅有单一图像可用（例如单目成像系统）设计的，与先前基于视频或多线索的方法相比，具有根本的不同和挑战性。

尽管多图像 DFD 技术常常优于单图像方法，但单图像 DFD 仍然是一个限制更多且更具挑战性的任务。相比之下，由于问题的复杂性，目前关于使用单个失焦图像进行 DFD 的研究相对较少。这些方法利用端到端神经网络在有监督学习环境中，通过使用真实深度数据来估计深度图。为了提高 DFD 结果，[2] 还计算了去模糊的模糊核，而 [3] 则从预测的深度图中推导出镜头参数（模糊因子和焦距差距），用于失焦模糊估计。[4] 从单个全焦图（AIF）图像中估计深度，并利用失焦图像仅在训练期间进行监督。在另一种从全焦图像进行深度估计的实例中，[5] 使用从暗通道计算的透射图作为网络的第四通道输入。在本研究中，我们提出了一种利用暗通道推导局部失焦模糊与对比度变化之间关系的新方法，以推断失焦模糊的存在及程度，并为深度估计提供线索。暗通道先验（DCP）广泛用于从雾霾/水下图像估计深度，[6]–[8] 其中 DCP 被利用来计算场景透射图，透射图是深度的函数。然而，DCP 最近也被适应用于基于去模糊图像中暗通道的稀疏性进行空间变化模糊分析和去模糊 [9]–[11] 。虽然失焦模糊降解是由于相机的光学效应，其与物理介质中的雾霾光散射不同，暗通道在这两种类型的降解图像中都起着类似的作用。在失焦模糊图像中，靠近焦平面的区域通常表现出较少的模糊。暗通道通过增加的强度变化突出显示这些区域。相反，暗通道在远离焦平面且显著模糊的区域中显示出降低的强度方差，并且由于模糊的平滑效应缺乏清晰细节。我们利用了失焦图像及其暗通道的综合局部强度偏差，即局部失焦和暗通道变化（LDDCV）图，以改善 DFD 性能。NYU-Depth V2（NYU-v2）数据集 [12] 在图 **??** 中的核密度估计（KDE）图有助于可视化暗通道强度差异和 LDDCV 图差异如何随归一化空间变化模糊程度改变，这是场景深度的函数。此外，我们使用对抗网络在训练期间使用失焦模糊图作为对抗监督信号来监督我们的 DFD 模型。我们基于单帧图像的 DFD 方法也为传统的多图像或硬件密集型方法提供了一个有希望的替代方案，能够从有限的数据中快速推断深度以提高系统效率。可以设计一个系统，故意使用固定焦点、光圈大的相机（这自然会产生失焦模糊）来被动推断单帧图像的深度。与主动深度感知技术相比，这种方法降低了系统复杂性和成本，使其成为现实中自动化应用的实用和可扩展的解决方案。我们的实验证明，对失焦图像应用 DCP 可产生有意义的深度估计结果。

我们通过在以像素 i 为中心且大小为 $\Omega(i) \times \Omega(i)$ 的局部窗口中，从三个颜色通道（红 r 、绿 g 、蓝 b ）的最小强度值计算失焦图像 I_{df} 的最暗场景辐射 J_{df} 。

$$J_{df}(I_{df})(i) = \min_{p \in \Omega(i)} \left(\min_{c \in \{R,G,B\}} I^c_{df}(p) \right) \tag{1}$$

暗通道强调阴影的长度、边缘及较暗的结构元素，这可以通过一种非常规的方式提供附加的背景去理解场景的三维布局，比如场景中物体的空间排列和相对距离。尽管在细腻的纹理和细节方面存在信息丢失，暗通道仍然保留了主要的场景结构和对应于显著深度变化的边缘。这有时会因为减少噪音和平滑小的变化而导致更大结构元素的更清晰表现。我们已将从单一散焦图像中提取的特征与暗通道的特征结合起来，以获得用于深度估计模型的增强结构信息。局部散焦和暗通道变化 (LDDCV) 图是通过连接局部散焦变化 (LDV) 和局部暗通道变化 (LDCV) 图得到的双通道强度变化图。它们描绘了某个局部区域内相邻像素之间的最大强度偏差，并通过捕捉与深度相关的散焦模糊的细微之处形成适当的表示。LDV 和 LDCV 图分别突出显示了

M. Medhi is in the Advanced Technology Development Center, Indian Institute of Technology, Kharagpur, India, 721302. e-mail: medhi.moushumi@iitkgp.ac.in

R. R. Sahay is with department of Electrical Engineering, Indian Institute of Technology, Kharagpur, India, 721302.

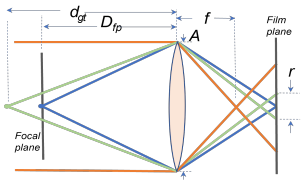

Fig. 1. 用于散焦模糊形成的薄透镜近似模型。

I_{df} 和 J_{df} 中的局部变化。数学上，

$$LDDCV(J,I)(i,j) = \{\max|J(i,j) - J(p,q)|, \max|I(i,j) - I(p,q)| \,|\, p = i-1, i, i+1, q = j-1, j, j+1\} \quad (2)$$

散焦模糊通过减少尖锐变化和降低局部区域内的最大值平滑了图像的整体外观。由于散焦模糊在局部区域的同质化效应，具有高散焦模糊的区域在 LDDCV 图中显示较低的局部变化。相反，低散焦模糊的区域显示出略高的 LDDCV 值。这一观察可以用来确定图像中散焦模糊的存在和程度，并提供评估单一失焦图像深度的见解。我们的网络的示意性架构如图 **??** 所示。对于给定的失焦图像 $I_{df} \in \mathbb{R}^{H\times W\times 3}$，使用一个预训练的 ResNeXt101-32x8d-wsl [13] 作为编码器主干（在图 **??** 中标记为 (a)），利用来自不同编码器层 i （$i = 1, 2, 3, 4$ ）的多尺度特征 $F_i^{df} \in \mathbb{R}^{H_i\times W_i\times C_i}$。这里，$H_i$、$W_i$ 和 C_i 分别表示高度、宽度和通道维度。类似地，多尺度特征 $F_i^l \in \mathbb{R}^{H_i\times W_i\times C_i'}$ 从 LDDCV 嵌入网络（LDDCV-Net）中提取，标记为 (b)。此外，采用一个平行的遮罩介导的稀疏池化网络（MMSP-Net）(标记为 (c)) 从输入的 LDDCV 映射及其有效性遮罩 (1 if |LDDCV| > T， where $T = 0.05$ 为阈值）中提取多尺度池化特征 $F_i^{lv} \in \mathbb{R}^{H_i\times W_i\times C_i''}$，然后与 F_i^l 串联。暗通道 J_{df} 突出显示的结构信息在通过全局平均池化（GAP）并被平展以获得特征 $z \in \mathbb{R}^{1\times Q}$ 之前，由暗通道嵌入网络 (D-Net) 嵌入到潜在空间（标记为 (d)）中。嵌套特征调制和融合模块（$Nest(FM)^2$ ），标记为 (e)，被构造成嵌套的、多层次的组，以分层方式提取细微的暗通道嵌入特征 z，以调制主要特征 $F^b \in \mathbb{R}^{\eta_{h_i}\times\eta_{w_j}\times Q}$。ARU、核心特征变换块（CFTB）、多层级特征增强块（MFEB）和分层残差细化（HR^2B ）的嵌套重复在多个层级上促进了广泛的特征提取和细化。暗通道注入的特征提升（DIFB）单元如图 **??** 所示。我们发现将 N 的重复设置为 2 可以在内存效率和 DFD 性能之间达到平衡。一个后续残差模块包含一个深度可分离非对称多尺度金字塔融合（DSA-MSPF）块（标记为 (f)），通过在传递给解码器（标记为 (g)）进行深度 d 重建之前充当多尺度上下文聚合先验来巩固学习到的表示。我们在整个深度生成模型中采用了蓝图可分离卷积（BSConv）[14]，以减少我们的模型参数 $\approx 49\%$。一个鉴别器 D （标记为 (h)）在对抗训练过程中将真实或估计的深度图 d_{gt}/d、真实或估计的散焦模糊图 $r(d_{gt})/r(d)$ (在章节 II-A 中解释）和散焦图像 I_{df} 作为输入，分别用来区分真实数据和生成数据。

A. 目标函数

回归像素级深度值的目标函数包括空间保真度损失 $\mathcal{L}_{\text{spafid}} = |d - d_{gt}|_1$、频域损失 $\mathcal{L}_{\text{freq}} = |\text{DCT}(d) - \text{DCT}(d_{gt})|_1$ 和对抗损失 $\mathcal{L}_{\text{adv}} = 0.5 \cdot \mathbb{E}_{d\sim p_d}\left[(D(d, r(d), I_{df}) - 1)^2\right]$ 项。离散余弦变换（DCT）定义为：$\text{DCT}(x_i) = x_k = \sum_{i=0}^{L-1} x_i \cos\left[\frac{\pi}{L}\left(i + \frac{1}{2}\right)k\right]$，其中 L 是信号中的总数据点数，k 是正在计算的 DCT 系数的索引。$\mathbb{E}_{d\sim p_d}$ 表示预测深度图 d 的分布 p_d 的期望值。联合损失函数被表述为：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{spafid}} + 0.1 \cdot \mathcal{L}_{\text{freq}} + 0.1 \cdot \mathcal{L}_{\text{adv}} \quad (3)$$

II. 实验

A. 数据集

NYU-Depth V2 (NYU-v2) 数据集 [12]：NYU-v2 数据集 [12] 包含 1,449 对空间匹配的 RGB 和深度图像，这些图像是使用 Microsoft Kinect 采集的。根据先前的工作 [16], [17]，我们采用了标准的训练/测试拆分，即 795/654 张图像。为了在 AIF NYU-v2 RGB 图像 I 中生成光学上真实的深度依赖散焦效果，我们选择了参数，对应于一个焦距（f ）为 9mm，焦平面（D_{fp} ）在 0.7m，光圈数（F_n ）为 2，以实现浅景深（DoF），传感器尺寸 p_x 为 7.5µm，以及光圈 $A = f/F_n$ 的合成相机。我们通过将 AIF 图像 I 与一个具有核半径 r 和位置索引 x, y 的点扩散函数（PSF）$G(x, y, r)$ 卷积来生成散焦模糊图像 I_{df}：按照图 1 所示的薄透镜模型，r 作为场景距离 d_{gt} 与相机的距离的函数计算为：

$$r(d_{gt}) = \frac{1}{\sqrt{2}\cdot p_x}\frac{Af}{(D_{fp} - f)}\frac{|d_{gt} - D_{fp}|}{d_{gt}} \quad (4)$$

EBD 数据集 [15]：EBD 数据集 [15] 包含 1,305 张高分辨率（1600 × 1024）真实散焦图像，没有真实深度图注释。这些图像具有一个光圈 F_n 为 1.8 的浅景深。请注意，我们仅将 EBD 数据集 [15] 用于测试。我们在表格 **??** 中报告了 NYU-v2 测试数据的消融结果。未使用黑暗通道作为补充线索（DDC）的模型得到的结果最不理想（① ）。将 DDC（② ）引入模型，通过将拼接的黑暗通道和散焦 RGB 图像作为四通道输入传递给图像编码器，同时保留 LDDCV-Net 和 MMSP-Net，标志着性能的提高。在这种配置中，D-Net 和 $Nest(FM)^2$ 被排除。当我们在没有对抗监督（③ ），即没有判别器的情况下训练我们的模型时，与完整模型（★ ）相比，我们观察到性能略有下降。

我们尝试了一种从单个空间变异散焦图像推断深度的新方法。我们研究了黑暗通道及其局部强度变化对基于其模糊特征进行深度估计的指导作用。在经过现实建模的合成数据集和实际散焦数据上的实验结果证明了我们方法的潜力。

References

[1] Y. Lu, G. Milliron, J. Slagter, and G. Lu, "Self-supervised single-image depth estimation from focus and defocus clues," *IEEE Robot. Autom. Lett. (RAL)*, vol. 6, no. 4, pp. 6281–6288, 2021.

[2] S. Anwar, Z. Hayder, and F. Porikli, "Depth estimation and blur removal from a single out-of-focus image." in *Brit. Mach. Vis. Conf. (BMVC)*, vol. 1, 2017, p. 2.

[3] D. Piché-Meunier, Y. Hold-Geoffroy, J. Zhang, and J.-F. Lalonde, "Lens parameter estimation for realistic depth of field modeling," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 499–508.

[4] S. Gur and L. Wolf, "Single image depth estimation trained via depth from defocus cues," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 7683–7692.

[5] Y. Li, C. Jung, and J. Kim, "Single image depth estimation using edge extraction network and dark channel prior," *IEEE Access*, vol. 9, pp. 112 454–112 465, 2021.

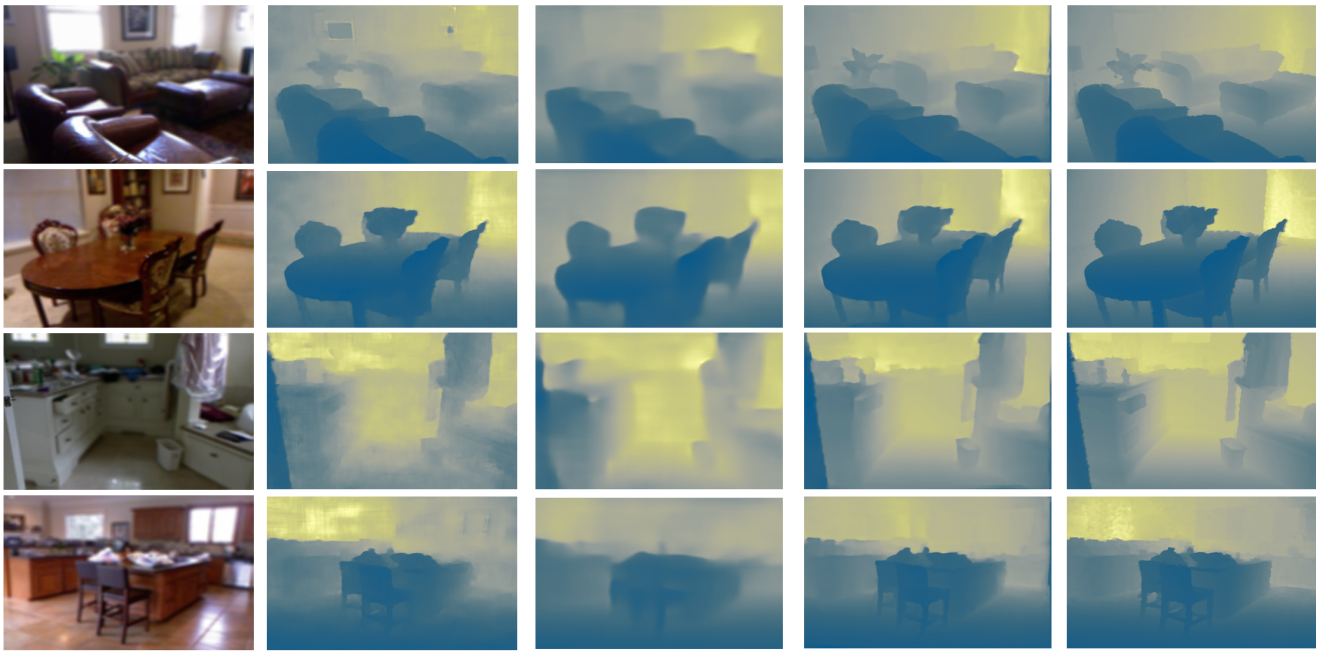

Fig. 2. 纽约大学-v2 数据集 [12] 中四张合成散焦模糊图像（a）的深度估计结果。估计的深度图（b）-（d）以及真实值（GT）（e）。

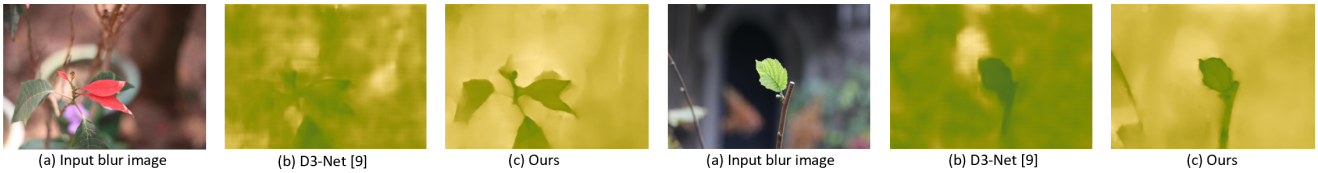

Fig. 3. 未经过微调的 EBD 数据集 [15] 中的两张高分辨率真实散焦模糊图像（a）的深度估计结果。

[6] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2010.

[7] J. Chen and L.-P. Chau, "An enhanced window-variant dark channel prior for depth estimation using single foggy image," in *IEEE Int. Conf. Image Process. (ICIP)*, 2013, pp. 3508–3512.

[8] J. Zhou, Q. Liu, Q. Jiang, W. Ren, K.-M. Lam, and W. Zhang, "Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction," *Int. J. Comput. Vis.*, pp. 1–19, 2023.

[9] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4003–4011.

[10] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Deblurring images via dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2315–2328, 2017.

[11] J. Cai, W. Zuo, and L. Zhang, "Dark and bright channel prior embedded network for dynamic scene deblurring," *IEEE Trans. Image Process.*, vol. 29, pp. 6885–6897, 2020.

[12] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 746–760.

[13] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1623–1637, 2020.

[14] D. Haase and M. Amthor, "Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 14 600–14 609.

[15] Y. Jin, M. Qian, J. Xiong, N. Xue, and G.-S. Xia, "Depth and dof cues make a better defocus blur detector," in *IEEE Int. Conf. Multimedia Expo (ICME)*, 2023, pp. 882–887.

[16] M. Carvalho, B. Le Saux, P. Trouvé-Peloux, A. Almansa, and F. Champagnat, "Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks?" in *Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 0–0.

[17] G. Song, Y. Kim, K. Chun, and K. M. Lee, "Multi image depth from defocus network with boundary cue for dual aperture camera," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 2293–2297.