陈永生, 李幸. 改进 CycleGAN 网络的危险驾驶场景生成[J]. 智能计算机与应用,2026,16(7):229-234. DOI:10.20169/j.issn.2095-2163.24111302

改进 CycleGAN 网络的危险驾驶场景生成

陈永生, 李 幸
(同济大学, 上海 200093)

摘 要: 在自动驾驶领域的研究与应用中,模拟和重现危险驾驶场景是一个成本高昂且劳动密集的过程。为了降低实验成本, 本文探讨了利用虚拟场景替代部分真实场景以进行相应实验的可能性,具体采用了模型生成危险驾驶场景的方式。为此,本文选择生成对抗网络(GAN)作为主要的图像生成模型。然而,该模型在训练过程中经常面临判别器训练中出现的梯度消失问题,这导致训练不稳定并最终影响图像的生成质量。针对这一问题,本文提出了一个改进的模型-uvCycleGAN。该模型结合了改进的视觉变换器(Vision Transformer)和 U-Net 架构,以期在生成危险驾驶场景图像时,显著提升生成图像的质量与真实感。为验证生成数据集的有效性,使用 fast-mot 算法在检测工具 mot trackeval 下对模型的表现进行了评估。实验结果表明,在开放数据集 KITTI 及特定的危险场景数据集 RADIATE 中,uvCycleGAN 模型的性能均明显超过了传统的 CycleGAN,从而显示出其在自动驾驶场景生成中的潜力与应用价值。

关键词: 生成对抗网络; Transformer; 场景生成; 驾驶; CycleGAN
中图分类号: TP391.1 **文献标志码**: A **文章编号**: 2095-2163(2026)07-0229-07

Enhanced CycleGAN network for generating dangerous driving scenarios

CHEN Yongsheng, LI Xing
(Tongji University, Shanghai 200093, China)

Abstract: Simulating hazardous driving scenarios in autonomous driving research is expensive and time-consuming. To mitigate these experimental costs, this paper explores the possibility of using virtual scenarios to replace certain real-world settings for corresponding experiments, specifically through the generation of hazardous driving scenarios using models. For this purpose, this study employs Generative Adversarial Networks (GANs) as the primary image generation model. GANs often face the challenge of gradient vanishing when training the discriminator, leading to unstable training processes, consequently, impacting the quality of the generated images. To address this challenge, this paper proposes an improved model, termed uvCycleGAN. This model combines an advanced Vision Transformer with a U-Net architecture, with the objective of substantially enhancing the quality and realism of images generated for hazardous driving situations. To assess the effectiveness of the generated dataset, the model's performance is analyzed using the fast-mot algorithm within the mot trackeval detection framework. Experimental findings indicate that the uvCycleGAN model substantially surpasses the traditional CycleGAN in performance on both the open KITTI dataset and the specialized hazardous scenario dataset, RADIATE. This emphasizes its potential and relevance in generating scenarios for autonomous driving.

Key words: Generating Adversarial Network; Transformer; scene generation; driving; CycleGAN

0 引 言

在自动驾驶领域,模型能够使用大量数据进行训练是至关重要的,尤其是在危险场景下,厂家往往缺乏足够的数据进行训练,导致问题频出。为了提升自动驾驶系统的安全性和可靠性,充足且多样化的训练数据显得尤为关键[1]。然而,由于危险场景的稀缺性和获取成本高昂,从现实世界中直接收集这些数据既不现实、也不经济。

因此,如何有效地生成高质量、覆盖全面的危险场景数据成为当前自动驾驶研究中的一个重要课题。通过使用生成对抗网络(GAN)等先进的深度

作者简介: 陈永生(1966—),男,博士,教授。主要研究方向:城市轨道交通或国铁列车控制系统,模式识别等。Email:lixing991206@163.com; 李 幸(2000—),男,硕士研究生。主要研究方向:图像生成。

收稿日期: 2024-11-13

学习技术,可以在人为控制的条件下合成各种危险场景数据[2],这不仅能够大幅降低数据获取的成本,还可以确保生成的数据具有足够的多样性和代表性,从而更好地满足自动驾驶模型的训练需求。

1 相关工作

1.1 生成对抗网络

生成对抗网络(GAN)在天气生成领域的研究主要集中于图像风格迁移和天气条件处理,其核心目标是通过跨域转换实现天气条件的变化。经典 GAN 通过生成器和判别器的博弈训练,成功学习了数据分布。然而,经典 GAN 由于输入为随机噪声,生成结果通常是不可控的,难以应用于具体任务。为了解决这一问题,条件 GAN(cGAN)在生成过程中引入了条件信息,如类别标签或其他辅助信息,从而实现了特定类别的图像生成。这一改进使得 GAN 能够更好地适应天气生成任务的需求。

基于 cGAN 的 Pix2Pix 方法进一步将生成对抗网络应用于成对数据的图像转换任务,通过结合 U-Net 结构和 $L1$ 损失,提升了生成图像的质量和细节表现。然而 Pix2Pix 依赖于高度一致的成对训练数据,而这样的数据在现实中往往难以获取,比如同一场景在不同天气条件下的图像对。因此,CycleGAN 通过引入循环一致性损失,解决了成对 数据不足的问题,实现了无监督条件下的跨域图像转换。CycleGAN 成为一种通用的图像风格迁移框架,广泛应用于风格转换、亮度调整等研究中。

1.2 循环一致性生成对抗网络 CycleGAN

循环一致性生成对抗网络 CycleGAN[3],在 GAN 的基础上增加了一个生成器和一个判别器,来实现样本 X 和 Y 之间的双向转换,如图 1 所示。样本 X 通过生成器 G 得到与样本 Y 相似的样本 Y^*,Y^* 通过生成器 F 得到与样本 X 相似的样本 X^*,判别器 DY 和 DX 分别判断 Y 与 Y^*、X 与 X^* 的相似程度。样本 Y 也会以同样的原理反向转换。

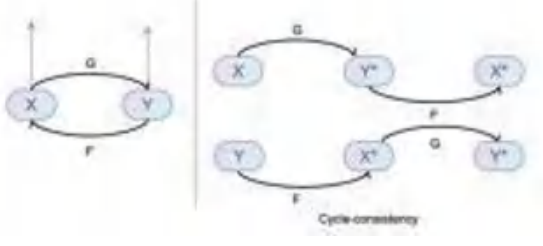


图 1 CycleGAN 结构图

Fig. 1 CycleGAN structure diagram

然而,CycleGAN 的生成策略存在全局转换的问题,即在对图像进行风格转换时,将倾向于对整幅图像进行全局处理,改变整体风格,而忽略了特定任务中需要专注的局部区域。在天气生成任务中,这一问题尤为明显。天气条件的变化通常只涉及图像中的部分区域(例如天空、云、地面反射等天气线索区域),其他区域(如建筑物、树木、道路等)应保持不变。然而,CycleGAN 的全局转换策略可能会对这些无关区域进行不必要的修改,导致生成图像出现伪影、不自然或细节丢失。例如,在晴天到多云的转换中,CycleGAN 可能对整幅图像进行风格化处理,使地面或建筑物的颜色或纹理也发生不必要的变化,削弱了生成结果的真实性和细节保留能力。

2 融合注意力机制和 U-Net 的 UV-CycleGAN 算法

2.1 生成器结构

CycleGAN 的生成器由 3 个部分组成,分别为编码器、转换器以及解码器(见图 2)。CycleGAN 的编码和编码器类似 U-Net 的 3 层卷积层和 3 层反卷积层组成。

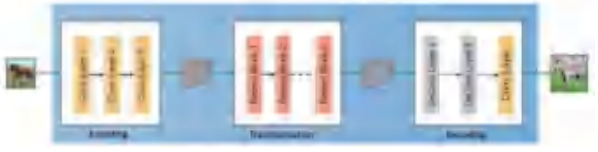


图 2 CycleGAN 编码器结构

Fig. 2 CycleGAN encoder structure

随着深层神经网络的发展,梯度消失问题成为训练的主要挑战之一。ResNet 通过引入残差连接,有效缓解了这一问题,并成为计算机视觉领域的主流模型。表 1 对比了 ResNet 与 Vision Transformer(ViT)的设计差异。

在 ViT 前后,使用原本的 U-Net 结构。uvCycleGAN 生成器结构如图 3 所示。

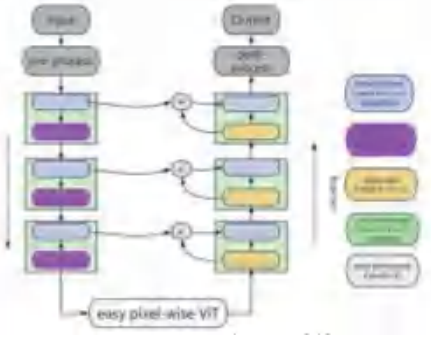


图 3 uvCycleGAN 生成器结构图

Fig. 3 uvCycleGAN generator structure diagram

表 1　ResNet 和 ViT 对比总结表

Table 1　Comparison summary table of ResNet and ViT

特性	ResNet[4]	Vision Transformer(ViT)[5]
基础单位	卷积层和残差块	Multi-Head Attention
输入处理	整个图像卷积	图像块嵌入(Patch Embedding)
计算复杂度	计算密集型,依赖卷积操作	自注意力,复杂度约为 $O(N^2)$
参数数量	随深度线性增加	通常较多,集中在 Transformer 编码器
训练和数据要求	数据需求较低,可在小数据集上训练	需要大规模数据集,依赖预训练

因为特征在输入 ViT 之前已经过了 UNet,达到了减少空间维度和提取特征的作用,在 easy pixel-wise ViT 中(见图 4),要去除 Vit Patch Embedding 中的 Conv2d(16×16),直接输入 Flatten 层,运算后求得特征矩阵,因为经过 U-Net 处理后,输入特征的空间维度已经相对较小,且已经具备了较强的语义特征,因此不再需要通过 Conv2d(16×16)进一步提取空间特征;经过 positional encoding,原本的 vit encoder block 中存在 MLP block[6],为了降低模型复杂度和避免过拟合,将多层感知机改为由 linear 和 leakRelu 组成的简单 LLr 层。

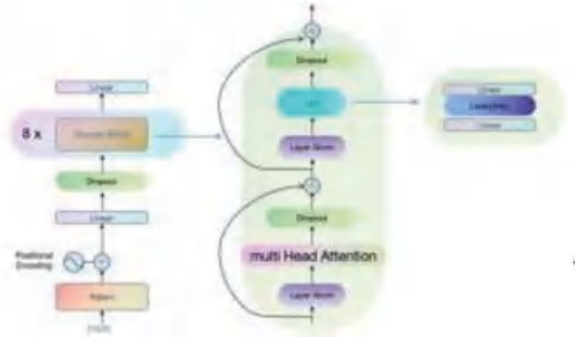


图 4　easy pixel-wise ViT 结构图

Fig. 4　easy pixel-wise ViT structure diagram

2.2　判别器结构

本文的判别器采用 PatchGAN[7] 结构不变。其核心思想是:PatchGAN 判别器不对整张图像进行判别,而是将图像划分为多个小块(patches),并对每个小块进行独立的判别。PatchGAN 与传统的 GAN 判别器的主要区别在于,普通 GAN 判别器输出一个实数,表示输入图像是否为真样本的概率,而 PatchGAN 输出一个 $N \times N$ 矩阵,每个元素表示图像中对应小块区域为真样本的概率,最终结果通过计算这些值的平均值得出。

PatchGAN 的优势在于 PatchGAN 能够关注图像的局部细节,特别适用于高分辨率图像生成。通过对每个小块的独立判断,更细致地捕捉图像的局部特征,从而生成更精细的图像。相比之下,普通 GAN 判别器可能忽略局部信息,导致生成图像细节不足。

2.3　损失函数

2.3.1　原始损失函数

CycleGAN 的损失函数主要包括 2 部分:对抗损失和循环一致性损失。

(1)对抗损失。CycleGAN 包含 2 个生成器和 2 个判别器。判别器的对抗损失用于训练判别器,使其能够区分真实图像和生成图像。判别器的损失公式如下,以 D_Y 举例:

$$L_{D_Y} = E_{y \sim P_{\text{data}}}(y)[\log D_Y(y)] + E_{X \sim P_{\text{data}}(x)}[\log(1 - D_Y(G(x)))] \tag{1}$$

其中,E 表示期望,"+"左边第 1 部分期望判别器 D 能够将真实数据 y 的概率 D_Y 最大化,判别器需要更好地识别真实数据。第 2 部分期望判别器能够将生成器生成的假样本 $G(x)$ 最大化,判别器需要更准确地识别生成的数据为假数据。

生成器的对抗损失用于训练生成器,使其生成的图像尽可能逼真,以欺骗判别器。数学公式如下:

$$L_G = E_{x \sim P_{\text{data}}(x)}[\log D_Y(G(x))] \tag{2}$$

可以用以下 min-max 博弈问题,$\max\limits_{D_Y}$ 的目标是提升判别器的能力以提高区分真实数据的能力。$\min\limits_{G}$ 的目标是生成更逼真的数据,从而欺骗判别器。用于描述生成器 G 和判别器 D 的对抗性训练过程,定义公式如下:

$$\min_G \max_{D_Y} L_{\text{GAN}}(G, D_Y, X, Y) \tag{3}$$

(2)循环一致性损失。CycleGAN 的核心创新点是循环一致性损失,确保图像在双向转换后能够恢复到原始状态。具体来说,正向和反向的循环一致性损失分别保证了图像从源域到目标域的转换,以及从目标域回到源域的转换,最终生成的图像尽可能与原始图像保持一致,正向循环一致性损失如下:

$$L_{\text{cyc}}(G, F, X) = E_{x \sim P_{\text{data}}}(x)[\| F(G(x)) - x \|_1] \tag{4}$$

综上,CycleGAN 的总损失函数表示如下:

$$L = L_{\mathrm{GAN}}(G, D_Y, X, Y) + L_{\mathrm{GAN}}(F, D_X, X, Y) + \lambda_1 L_{\mathrm{cyc}}(G, F) \quad (5)$$

其中,λ_1 表示循环一致性损失的权重;$L_{\mathrm{GAN}}(G, D_Y, X, Y)$ 表示从 $X \rightarrow Y$ 的对抗损失;$L_{\mathrm{GAN}}(F, D_X, X, Y)$ 表示从 $Y \rightarrow X$ 的对抗损失;$L_{\mathrm{cyc}}(G, F)$ 表示循环一致性损失。

2.3.2 损失函数的改进

(1)对抗损失。在原始的 CycleGAN 网络中,一般选用 $L2$ 作为对抗损失[8],虽然这能够加快损失值的收敛速度,但是对于损失的放大缩小是比较直接的:真实样本和生成样本之间的损失很小或很大的时候,都会降低模型的整体性能。为了降低此类误差带来的影响,本文模型使用 Wasserstein 损失[9],可以缓解训练过程中的不稳定性。Wasserstein 距离也称为 Earth-Mover(推土机)距离,其公式如下:

$$W(P_r, P_g) = \inf_{r \sim \Pi(p_r, p_g)} E_{(x,y)\sim r}[\| x - y \|] \quad (6)$$

其中,$\Pi(P_r, P_g)$ 表示所有可能的联合分布的集合;$\inf_{r \sim \Pi(p_r, p_g)}$ 表示寻找最优移动计划,找到使传输代价最小的计划;E 表示期望代价。

使用 Wasserstein 距离度量,判别器(称为判别器或评估器)的输出不再是概率,而是一个实数值。

对于判别器 D,可推得:

$$L_D = -E_{x \sim P_{\mathrm{data}}(x)}[D(x)] + E_{z \sim P_z(z)}[D(G(z))] \quad (7)$$

对于生成器 G,可推得:

$$L_G = -E_{z \sim P_z(z)}[D(G(z))] \quad (8)$$

式(7)、式(8)中,真实数据 x 来自分布 P_{data},噪声 z 来自分布 P_z,生成器 G 将噪声 z 转换为生成样本 $G(z)$。

判别器 D 输入数据给出一个实数评分,用于衡量样本的真实性。判别器损失 L_D 通过对真实样本评分 $D(x)$ 的最大化、及对生成样本评分 $D(G(z))$ 的最小化,以更好地区分真实数据和生成数据;而生成器损失 L_G 则最小化生成样本的评分 $D(G(z))$,以生成更加真实的样本,从而迷惑判别器。

(2)循环一致性损失。循环一致性损失是 CycleGAN 中的核心组成部分,有助于保持图像转换的一致性。在循环一致性损失的训练过程中,由于生成器的目标是生成逼真的图像,而不是完全保留原始图像的所有信息,生成器可能会过度平滑化图像,导致一些细微的纹理和细节被模糊掉。

LPIPS 提供了一种更贴近人类视觉感知的图像相似度度量,通过神经网络提取高级特征,超越传统的 $L2$ 距离和 PSNR 等方法。即通过学习生成图像和真实图像之间的反向映射,专注于感知相似度。因此,本文采用 LPIPS[10] 作为循环一致损失函数,以提升生成图像质量。推得的公式如下:

$$d(x, x_0) = \sum_L \frac{1}{H_L W_L} \sum_{hw} \| W \odot \Theta(\hat{y}^L_{hw} - \hat{y}^L_{0hw}) \|_2^2 \quad (9)$$

其中,$d(x, x_0)$ 表示 LPIPS 感知相似度度量,通过预训练的神经网络提取图像 x 和 x_0 在特定特征层 L 的特征 $\hat{y}^L_{hw}$ 和 $\hat{y}^L_{0hw}$,对每个特征位置 (h, w) 的特征差值计算 $L2$ 范数,并用所学习到的权重 w 加求和。最终对特征图的空间分辨率(HLWL)进行归一化。

3 实验

3.1 数据集

本文选用 KITTI[11] 数据集作为目标数据集,CARLA[12] 仿真软件采集的虚拟数据作为源数据集,以实现仿真图像转化为真实世界图像的目标。在 CARLA 中,采集了以下 3 种场景数据图:障碍物、糟糕天气、高速紧急避障[13] 这 3 个场景。



(a) CARLA 数据集 (b) KITTI 数据集

图 5 CARLA 和 KITTI 数据集图

Fig. 5 Images of CARLA dataset and KITTI dataset

3.2 实验参数与评价指标

3.2.1 实验环境与参数设置

本文的实验在 Linux 系统上完成,显卡为 4080 16 GB。实验的输入和输出图像为 256×256。训练轮次不等,100~400 轮不等,学习率从 0.000 2 逐渐衰减到 0[14]。

3.2.2 评价指标

为了量化实验结果,比较生成的图像质量与人工标记的图像质量的差异性,本文选取 FID[15] 和 KITTI vision benchmark suite[16-17] 两个指标评估生成图像的质量。

(1)FID 通过计算生成图像与原图像的特征向量之间的距离衡量 2 组图像的相似度,FID 的分数越低,代表 2 组图像越相似,公式如下:

$$\mathrm{FID} = \| u_1 - u_2 \|^2 + T_Y(C_1 + C_2 - 2(C_1C_2)^{\frac{1}{2}}) \tag{10}$$

其中，u 表示均值；C 表示常数；T_Y 表示矩阵的迹。

（2）KITTI vision benchmark suite mot trackEval 适用于多目标跟踪（MOT）领域中用于评估图像算法性能的指标，参数包括 HOTA、DetA、AssA、DetRe、DetPr、AssRe、AssPr 和 LocA。这些参数介于 0~1 之间，数值越高、算法性能越好。使用加权平均的方法，将各个指标按照其重要性赋予权重，统一到一个总体评分公式中，

$$\begin{aligned}\mathrm{Score} = {} & w_1 \cdot \mathrm{HOTA} + w_2 \cdot \mathrm{DetA} + w_3 \cdot \mathrm{AssA} + \\ & w_4 \cdot \mathrm{DetRe} + w_5 \cdot \mathrm{DetPr} + w_6 \cdot \mathrm{AssRe} + \\ & w_7 \cdot \mathrm{AssPr} + w_8 \cdot \mathrm{LocA}\end{aligned} \tag{11}$$

其中，权重 $w_1 \sim w_8$，根据任务需求进行调整，满足和为 1。

3.3　对比实验

本文以 CycleGAN 模型为基础，对其生成器、判别器和损失函数进行改进。为了验证本文模型改进的有效性，使用 3 800 张 256×256 的 KITTI 数据集、RADIATE 数据集（世界首个恶劣天气开源数据集）和 CARLA 虚拟场景数据集，其中 KITTI 和 RADIATE 分别包含正常天气和暴雨天数据集，在多种训练轮次下进行对比实验。

方法 A 使用 CycleGAN 原网络，训练 70~300 次；方法 B 在方法 A 的基础上，加入上述的改进，训练 70~300 次。在正常天气下，虚拟场景和真实场景数据图数分别为 4 862 和 6 500 张图，数据量比率为 0.75。

在阴雨天气下分别为 4 553 和 5 700 张图，数据比例约为 0.8。

（1）实验结果 1。CycleGAN 对比 uvCycleGAN 在不同天气不同轮次 FID 数值结果见表 2。从实验结果可以看出：正常天气条件下，原始 CycleGAN 在训练轮次 220 次时达到了最低的 FID 值 74，但在 280 次时 FID 值有所回升。改进后的 uvCycleGAN 在 100 次训练轮次时 FID 值为 64，显著低于 CycleGAN 的最佳值，且在 300 次训练轮次时 FID 值进一步降低至 61。暴雨天气条件下，uvCycleGAN 在 100 次训练轮次时 FID 值为 106，但在进一步训练至 300 次时、FID 值显著降低至 82。

实验结果表明，改进后的 uvCycleGAN 模型在生成图像质量上明显优于原始 CycleGAN 模型，尤其是在正常天气条件下，uvCycleGAN 的 FID 值显著低于 CycleGAN。在暴雨天气条件下，虽然 uvCycleGAN 在初始训练阶段的 FID 值较高，但随着训练轮次的增加，生成图像质量逐步提升，表现出比原始 CycleGAN 更好的稳定性和生成效果。

表 2　CycleGAN 对比 uvCycleGAN 在不同天气不同轮次 FID 数值

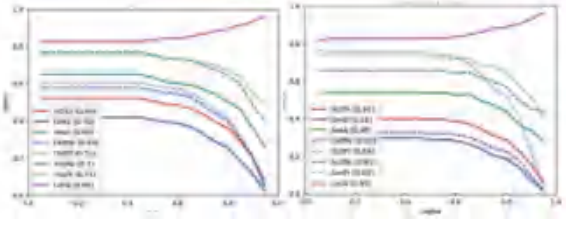
Table 2　Table of FID values for CycleGAN vs uvCycleGAN across different weather conditions and epochs

模型（正常天气）	FID	模型（阴雨天气）	FID
CycleGAN-70	88.99	CycleGAN-100	117.13
CycleGAN-170	78.85	CycleGAN-200	106.67
CycleGAN-220	74.13	CycleGAN-300	107.32
CycleGAN-280	82.49	uvCycleGAN-100	106.31
uvCycleGAN-100	64.14	uvCycleGAN-200	87.91
uvCycleGAN-300	61.24	uvCycleGAN-300	82.10

（2）实验结果 2。本文选定最简单、最快速的方式：在图像数据上使用 YOLOv3 进行图像识别，并将识别结果通过矩阵运算映射到 3D 点云图像中。使用该算法，对生成的图像和点云数据集（在正常天气下）进行对比实验。

使用算法直接检测原始的 CARLA 数据集，将 A 作为 ground true 真实数据集，检测 CycleGAN 和 uvCycleGAN 生成的数据集。在 DFR-FastMot[18] 算法下检测结果的分数如图 6 所示，指标越接近 1 越好：

alpha 参数是指标计算中的一个重要参数，可控制检测结果和跟踪结果之间的权衡。α 参数的取值范围通常在[0, 1] 之间。当 $0 < \alpha < 1$ 时，MOTA 会同时考虑检测结果和跟踪结果，权重由 α 控制。

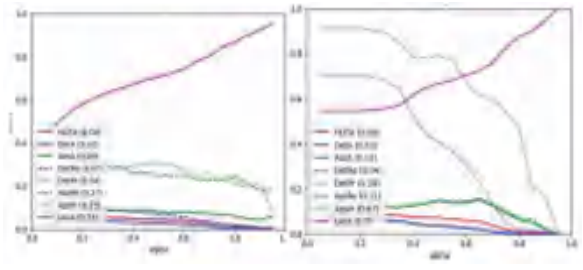


（a）CycleGAN　　（b）uvCycleGAN

图 6　CycleGAN 和 uvCycleGAN KITTI 评价指标图

Fig. 6　KITTI evaluation metrics chart of CycleGAN and uvCycleGAN

由图 6 结果对比可得出结论：使用 uvCycleGAN 生成的数据集在 DFR-FastMOT 算法上取得了更好的效果。除了对比正常场景，在相同训练参数下，又对基于 RADIATE 数据集这种极端场景数据集下的表现进行了对比，结果如图 7 所示。



(a) CycleGAN (b) uvCycleGAN

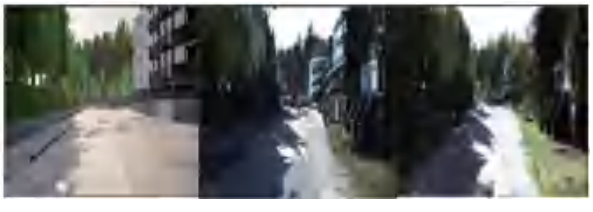
图 7 CycleGAN 和 uvCycleGAN 极端天气下 KITTI 评价指标图

Fig. 7 KITTI evaluation metrics chart in extreme weather conditions of CycleGAN and uvCycleGAN

从图 7 数据中可以看出,2 种 GAN 在极端浓雾场景下的表现都比较差,大部分的车辆都没有被 DFR-FastMot 算法识别出来。可能的原因有 2 种:

① GAN 在处理恶劣天气条件下的图像时,生成的合成数据往往无法完全恢复被遮挡的细节[19],特别是在遮蔽严重的情况下,如大雾、雨雪等环境。雾霾情况下的图像数据往往缺乏清晰的边缘和高对比度特征,GAN 的生成网络难以从这些模糊的输入中有效学习到关键特征,从而导致生成的数据质量下降,影响最终的检测效果。

② 传统的检测和跟踪算法在极端条件下,特别是在能见度极低的环境中表现较差,检测器难以有效分割目标或提取清晰的边缘信息,导致跟踪误差增加。DFR-FastMot 算法由于缺乏针对低对比度、低分辨率图像的训练,在浓雾等恶劣天气下,其检测率和跟踪精度显著下降,难以区分目标物体与背景噪声,从而影响整体的多目标跟踪效果[20]。晴朗天气下数据生成的结果对比如图 8 所示。



(a) 原图 (b) 结果 1 (c) 结果 2

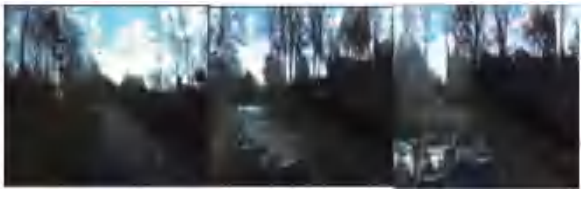
图 8 CARLA 原图-CycleGAN-uvCycleGAN 生成结果图

Fig. 8 Diagram of CARLA original image-CycleGAN-uvCycleGAN generated results

CycleGAN 生成的图像,整体颜色较暗,阴影部分区域的亮度降低;建筑物和地面细节较原图模糊,边缘有轻微的伪影。CycleGAN 在训练时对颜色和纹理的映射学习较弱,特别是在复杂的阴影和细节区域,导致风格迁移不稳定。

uvCycleGAN 生成的图像与 CycleGAN 相比有所改进,亮度和阴影区域的平衡更好,整体颜色与真实场景接近;建筑物的边缘和地面的纹理较为清晰,较少出现模糊和伪影。

uvCycleGAN 相较于 CycleGAN,能够更好地处理复杂的光照和纹理,提升了风格迁移的质量和一致性。生成雨天地面有积水的结果对比如图 9 所示。



(a) 原图 (b) 结果 1 (c) 结果 2

图 9 原图-CycleGAN-uvCycleGAN 生成结果图

Fig. 9 Diagram of original image - CycleGAN - uvCycleGAN generated results

图 9(b)是 CycleGAN 生成的雨天积水场景。虽然图像加入了雨天积水的效果,但积水区域的分布显得不够真实,存在一定的模糊感,且积水的反射纹理不自然。图 9(c)是 uvCycleGAN 生成的雨天积水场景。相比 CycleGAN, uvCycleGAN 在积水区域的纹理表现更加自然,积水的反射效果更真实,符合实际雨天积水的光学特性。

4 结束语

目前的危险场景下的数据获取方法主要是模拟场景和少量的实地读取。本文的方法主要应用于模拟场景转换为真实场景的方法,使用本文方法可以减少运用模拟场景带来的数据差别,增强鲁棒性。

本文主要是在图像场景下将危险虚拟场景转化为真实场景,下一步的研究方向为:

(1)在多模态数据中, 点云数据和图像数据是一起出现的,在 CARLA 生成的暴雨天点云数据不包含噪声,这时需要转换为现实暴雨场景下的点云数据。

(2)地图数据现已应用在自动驾驶中,危险变道场景[21]的研究工作也在开展。经过点云、图像和地图数据[22]的多模态数据生成, 可以有效提高自动驾驶在危险场景下的表现能力。

参考文献

[1] 王淳浩, 闭家铭, 阮利, 等. 自动驾驶安全关键场景生成技术综述[J]. 信息与控制, 2024, 53 (1): 17-32.

[2] Xu S, Mi L, Gilpin L H. A framework for generating dangerous scenes for testing robustness [R/OL]. (2022-01-01). https://openreview.net/pdf? id=ZjN2AuXgu1.

[3] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 2223-2232.
[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778.
[5] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv, 2010.11929, 2020.
[6] Qiu G. MLP for adaptive postprocessing block-coded images[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(8): 1450-1454.
[7] Demir U, Unal G. Patch-based image inpainting with generative adversarial networks[J]. arXiv preprint arXiv, 1803.07422, 2018.
[8] 邹承明, 胡佑璞. 引入生成对抗网络的室外场景单目深度估计[J]. 计算机工程与应用, 2021, 57(6): 176-183.
[9] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks[J]. arXiv preprint arXiv, 1701.07975, 2017.
[10] 杜佳俊, 兰红. 基于多模态特征融合的图像编辑模型[J]. 计算机科学与应用, 2024, 14(6): 164-176.
[11] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The kitti dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
[12] Dosovitskiy A, Ros G, Codevilla F, et al. CARLA: An open urban driving simulator[J]. arXiv preprint arXiv, 1711.03938, 2017.
[13] 周兵, 郑康强, 王茹, 等. 考虑道路可用宽度的自适应避撞控制[J]. 汽车工程, 2024, 46(5): 893-905.
[14] 冯宇旭, 李裕梅. 基于 LSTM 神经网络的沪深 300 指数预测模型研究[J]. 数学的实践与认识, 2019, 49(7): 308-315.
[15] Jayasumana S, Ramalingam S, Veit A, et al. Rethinking FID: Towards a better evaluation metric for image generation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 9307-9315.
[16] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 3354-3361.
[17] Athar A, Luiten J, Voigtlaender P, et al. Burst: A benchmark for unifying object recognition, segmentation, and tracking in video[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE, 2023: 1674-1683.
[18] Nagy M, Khonji M, Dias J, et al. DFR-FASTMOT: Detection failure-resistant tracker for fast multi-object tracking based on sensor fusion[C]//Proceedings of 2023 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ: IEEE, 2023: 827-833.
[19] Mefta H L, Cherif A, Braham R. Improving autonomous vehicles maneuverability and collision avoidance in adverse weather conditions using generative adversarial networks[J]. IEEE Access, 2024, 12: 89679-89690.
[20] Tahir N U A, Zhang Z, Asim M, et al. Object detection in autonomous vehicles under adverse weather: A review of traditional and deep learning approaches[J]. Algorithms, 2024, 17(3): 103.
[21] 赵祥模, 赵玉钰, 景首才, 等. 面向自动驾驶测试的危险变道场景泛化生成[J]. 自动化学报, 2023, 49(10): 2211-2223.
[22] 杨振凯, 华一新, 訾璐, 等. 浅析高精度地图发展现状及关键技术[J]. 测绘通报, 2021(6): 54-60.