

# 基于改进 StyleGAN 乡村道路图像数据增强方法<sup>①</sup>



希仁娜<sup>1,2,3</sup>, 张太红<sup>1,2,3</sup>, 姚芷馨<sup>1,2,3</sup>

<sup>1</sup>(新疆农业大学 计算机与信息工程学院, 乌鲁木齐 830052)

<sup>2</sup>(智能农业教育部工程研究中心, 乌鲁木齐 830052)

<sup>3</sup>(新疆农业信息化工程技术研究中心, 乌鲁木齐 830052)

通信作者: 张太红, E-mail: [zth@xjau.edu.cn](mailto:zth@xjau.edu.cn)

**摘要:** 针对现有乡村道路图像数据集数据样本少且类别不均衡的问题, 提出了一种基于改进 StyleGAN 的数据增强方法. 在原有 StyleGAN 框架基础上引入了解耦映射网络, 以降低  $W$  空间隐编码的耦合度; 综合卷积与 Transformer 的优点, 设计了一个卷积耦合迁移模块 (CCTB), 该模块中的核心十字形窗口自注意力机制增强了网络对复杂上下文和空间布局的捕捉能力. 这两项改进使网络性能大幅提升. 通过消融实验对比改进前后的 StyleGAN 网络,  $IS$  指标从 42.38 提高到了 77.31,  $FID$  值从 25.09 降至 12.42, 表明了数据生成质量和真实性得到了显著提升. 为了验证数据增强对模型性能的影响, 采用了 2 种经典且主流的目标检测算法进行测试, 比较了原数据集与增强后数据集的性能差异, 进一步证实了改进措施的有效性.

**关键词:** 数据增强; StyleGAN; 解耦映射; 卷积耦合迁移

引用格式: 希仁娜, 张太红, 姚芷馨. 基于改进 StyleGAN 乡村道路图像数据增强方法. 计算机系统应用, 2025, 34(4): 45-54. <http://www.c-s-a.org.cn/1003-3254/9838.html>

## Data Augmentation Method of Rural Road Images Based on Improved StyleGAN

XI Ren-Na<sup>1,2,3</sup>, ZHANG Tai-Hong<sup>1,2,3</sup>, YAO Zhi-Xin<sup>1,2,3</sup>

<sup>1</sup>(School of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China)

<sup>2</sup>(Intelligent Agriculture Engineering Research Center of the Ministry of Education, Urumqi 830052, China)

<sup>3</sup>(Xinjiang Agricultural Informatization Engineering and Technology Research Center, Urumqi 830052, China)

**Abstract:** To address the issues of limited sample size and imbalanced categories in existing rural road image datasets, a data augmentation method based on an improved StyleGAN is proposed. This approach introduces a decoupled mapping network into the original StyleGAN framework to reduce the coupling degree of the  $W$ -space latent code. By integrating the advantages of convolution and Transformer, this study designs a convolution-coupled transfer block (CCTB). The core cross-window self-attention mechanism within this module enhances the network's ability to capture complex context and spatial layouts. These two improvements significantly boost network performance. Ablation experiments comparing the original and improved StyleGAN networks show that the  $IS$  index increases from 42.38 to 77.31, and the  $FID$  value decreases from 25.09 to 12.42, demonstrating a substantial improvement in data generation quality and authenticity. To verify the impact of data augmentation on model performance, two classic and mainstream object detection algorithms are used for testing. Performance differences between the original and augmented datasets are compared, further confirming the effectiveness of the improved methods.

**Key words:** data augmentation; StyleGAN; decoupled mapping; convolutional coupling transfer

① 收稿时间: 2024-10-09; 修改时间: 2024-10-21, 2024-11-19; 采用时间: 2024-11-29; csa 在线出版时间: 2025-03-04  
CNKI 网络首发时间: 2025-03-06

随着无人驾驶技术的迅猛发展,针对乡村道路环境的机器视觉与自主导航技术正成为研究的新焦点.深度学习<sup>[1]</sup>技术以其高精度性、强鲁棒性以及低成本在自动驾驶领域取得了巨大成功.深度学习模型在多种任务中表现优异的关键,在于其训练数据集是否兼具广泛性与多样性.然而,现有的乡村道路数据集样本量有限,且乡村道路因其独特性,在结构、环境和使用模式上与城市道路存在显著差异,这些差异使得乡村道路的数据集在样本量和多样性上更具挑战性.例如,乡村道路可能更狭窄、曲折,且周围环境更加多变,这些特点在数据集中可能难以全面覆盖.若要快速解决样本量有限问题,数据增强是一种有效方式.

数据增强的意义不仅在于扩充样本数量,更在于提升模型的泛化能力.通过数据增强,模型能够接触到更多道路的变体,包括不同天气条件、光照变化以及道路两旁的植被变化等,从而提高模型在不同乡村道路环境下的泛化能力.此外,数据增强有助于解决乡村道路数据集中可能存在的样本不平衡问题.某些类型的乡村道路可能在数据集中占比较小,通过数据增强可以增加这些类别的样本数量,从而提高模型对这些少数类别的识别能力.

基于图像的数据增强方法<sup>[2]</sup>分为基于传统图像处理技术的图像数据增强和基于机器学习<sup>[3]</sup>的图像数据增强技术.传统的数据增强方法包括对原始图像进行几何变换、随机调整亮度和对比度、添加各类噪声等操作.王建明等<sup>[4]</sup>设计了一种基于目标前景提取,使用纯色替换原始背景以实现数据增强的方法.Perez等<sup>[5]</sup>对传统的数据增强方法和基于深度学习的图像增强方法进行了对比,相较于传统数据增强方式基于深度学习的数据增强方法可以更加有效地提高分类算法的精度.深度学习,尤其是利用生成对抗网络GAN<sup>[6]</sup>进行数据增强,是当前研究的热点.这些方法通过学习数据的底层分布来生成新的、逼真的数据样本,从而增加数据集的多样性和数量.田晨璐等<sup>[7]</sup>提出了一种基于CycleGAN<sup>[8]</sup>的道路多场景图像数据增强与目标检测方法,利用CycleGAN生成黑夜、黄昏、阴雨天等多种场景下的图片,并将生成的图像与原始图像混合作为新的数据集训练YOLOv5.李艳玲等<sup>[9]</sup>提出一种基于改进DCGAN<sup>[10]</sup>的图像扩充方法,将DCGAN原模型中的损失函数进行替换,并将自注意力机制模块融入原有的DCGAN模型中,提高模型对病害斑块特征的学

习能力.StyleGAN是Karras等<sup>[11]</sup>提出的一种生成对抗网络,其核心特点是独特的架构设计,它能够生成高质量、高分辨率的图像,并且在控制图像的特定风格和内容方面具有前所未有的灵活性.

综上所述,鉴于现有乡村道路图像具有分辨率较高、图像中各类别之间耦合度较高等特点,为了生成高质量的乡村道路图像,并进一步提高乡村道路环境下图像识别算法的精度和鲁棒性,本文对StyleGAN网络结构进行了优化:1)为精细调控各类别的视觉样式生成,减少特征间的相互依赖关系,本文将原有的映射网络架构8层全连接层划分为4个子层组,每2层输出中间隐编码并在仿射变换后快速引入AdaIN<sup>[12]</sup>模块,降低了W空间隐编码的耦合度.2)设计了十字形窗口自注意力机制(CSWin-Attention),增强了网络对复杂上下文和空间布局的捕捉能力.本研究基于乡村道路数据集进行数据增强,以生成更多样本.为了评估改进后的数据增强方法对模型识别精度的影响,选取了Mask R-CNN<sup>[13]</sup>和YOLOv8n<sup>[14]</sup>这两个神经网络模型,并对比分析了数据增强前后在这两个模型上的识别精度.

## 1 StyleGAN网络改进

生成对抗网络在图像生成任务中的优异表现,在很大程度上归因于其神经网络架构的强大拟合能力.因此,生成图像的质量与多样性直接受到神经网络架构设计的影响<sup>[15,16]</sup>.本文在深入分析StyleGAN网络架构的基础上,提出了改进方案,在原生生成器上引入解耦映射网络和卷积耦合迁移块(CCTB).

### 1.1 解耦映射网络

在StyleGAN网络中,W空间的特征均源自同一个映射网络F,这意味着各维度特征之间可能存在某种程度的相关性,进而造成不同风格特征之间的耦合效应.鉴于乡村道路数据集图像中包含多种多样的类别样式,为了更好地实现对各类别独立风格的精细化生成,有必要降低这种耦合度.因此构建了解耦映射网络,即将原有的8个FC层分为4组,每两层FC网络的输出即作为一个中间隐编码.这些隐编码在经过专门设计的仿射变换处理后,直接供给到适应性实例规范化(AdaIN)模块中,这一操作旨在减少W空间隐编码间的强耦合关系,从而促进生成网络生成更多样化且风格之间区分度更高的图像.改进后的网络结构如图1所示.

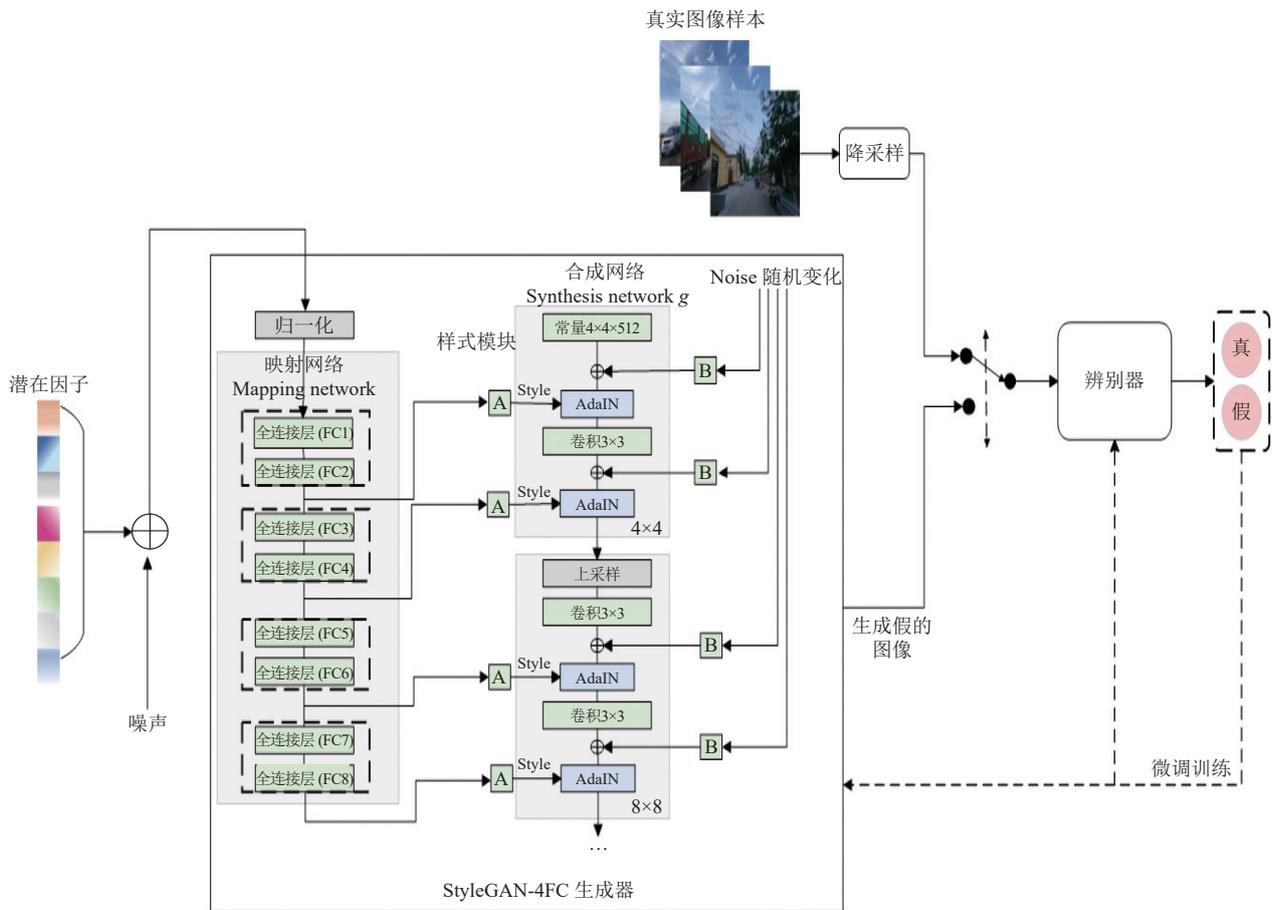


图1 改进后的网络结构图

### 1.2 卷积耦合迁移模块 (CCTB)

在 StyleGAN 中, 每一层的学习模块仍然采用基础的卷积模块进行, 卷积模块的感受野有限, 仅能从生成器多尺度逐步生成的特征中生成图像, 图像语义信息较强时, 生成的图像不管从连续性, 还是从生成质量上看都较差. 在乡村道路图像数据集中道路、树木、天空等目标的生成依赖于长距离感知. 相较于卷积, Transformer<sup>[17]</sup>具备感知长距离依赖的能力. 利用 Transformer 进行生成任务的设计, 能够更好地生成这些目标. 然而, Transformer 相较于普通卷积也存在一定的局限性, 例如其收敛速度较慢, 需要大量数据来进行训练. 因此, 综合卷积与 Transformer 的优点, 本文设计了一个卷积耦合迁移模块 (CCTB).

该模块是由局部编码模块 (LEM)、十字形窗口自注意力 (CSWin-Attention)、归一化 (BN, LN) 和多层感知器 (MLP) 层构成, 其结构如图 2 所示. 公式定义为式 (1) 和式 (2).

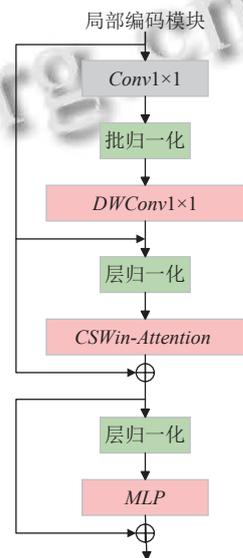


图2 CCTB 模块结构图

$$Z'_l = \text{CSWin-Attention}(\text{LN}(\text{LEM}(Z_{l-1}) + Z_{l-1}) + Z_{l-1}) \quad (1)$$

$$Z_l = \text{MLP}(\text{LN}(Z'_l)) + Z'_l \quad (2)$$

其中,  $Z_l$  表示第 1 个 CCTB 的输出。

### (1) 局部编码模块 (LEM)

物体外观和位置的变化经常给图像处理任务带来挑战。在乡村道路数据集的图像中, 各对象之间的边界往往是模糊的。CNN 凭借其强大的局部特征提取能力, 能够有效地识别各对象的边缘。此外, 为了表示标记之间的相对位置, Transformer 为每个对象集成了特定的位置编码, 这在一定程度上增加了平移、缩放和失真的不变性。然而, CNN 通过局部感受野和共享权重熟练地捕捉了这种局部结构, 从而减轻了上述限制。本文使用 LEM 提取局部特征, 并将其融合为强全局特征编码由 Transformer 通过残差连接。这部分由式 (3) 定义。

$$LEM(Z_l) = DWConv(BN(Conv(Z_l))) + Z_l \quad (3)$$

其中,  $DWConv$  表示  $3 \times 3$  步长为 1 的深度可分离卷积,  $Conv$  表示  $1 \times 1$  步长为 1 卷积。

### (2) 十字形窗口自注意力 (CSWin-Attention)

尽管自注意力机制<sup>[18]</sup>提供了强大的长距离上下文建模能力, 但其计算复杂度与特征图尺寸呈二次关系, 这意味着当输入为高分辨率图像时, 计算所需时间较长。因此, 将特征图划分为多个窗口并在窗口内部及窗口之间计算自注意力是一种可行且可靠的选择方案, 有助于解决计算耗时的问题。

然而, 考虑到乡村道路数据集中场景具有复杂的上下文信息和空间关联性, 如树木半遮挡交通标志、监控系统等问题, 传统的注意力机制可能无法充分捕捉到这种二维空间内的长距离依赖关系。针对此类问题, 提出了一种创新的解决方法 CSWin-Attention。该模型设计了特殊的注意力计算方式, 通过将输入特征图  $Z \in \mathbb{R}^{C \times H \times W}$  按照头维度 (通道维度) 进行水平和垂直的拆分, 这样可以分别从横向 (水平方向) 和纵向 (垂直方向) 独立地计算注意力权重。这意味着模型会生成两个不同的注意力分支, 即垂直注意力 (Ver-Attention) 和水平注意力 (Hor-Attention):

**垂直注意力 (Ver-Attention):** 该分支关注输入特征图在垂直方向上的相关性, 对应于特征向量集合  $h_{\text{ver}} = \{h_1, h_2, \dots, h_{k/2}\}$ , 这些向量沿着垂直轴进行自注意力计算, 从而能够捕获上下或左右相邻位置之间的语义关联。

**水平注意力 (Hor-Attention):** 该分支则聚焦于输

入特征图在水平方向上的依赖关系, 对应的特征向量集合  $h_{\text{hor}} = \{h_{k/2+1}, h_{k/2+2}, \dots, h_k\}$ , 同样在相应方向上进行自注意力计算, 以理解同一行或列内元素间的语义交互。其中  $k$  表示输入  $Z$  的数量, 通常为偶数。CSWin-Attention 的垂直分支  $h_{\text{ver}}$  的计算方法如下所示。首先, 给定  $h_i \in h_{\text{ver}}$ , 将  $h_i$  均匀地分成宽度为  $sw$  的非重叠垂直带  $\{Z^1, Z^2, \dots, Z^M\}$ , 其中  $i = 1, 2, \dots, k/2$ ,  $M = H/sw$ ,  $sw$  根据不同阶段特征图的大小进行经验性设置。然后, 通过对每个垂直分支  $Z_j$  进行线性投影获得相应的查询 ( $Q_j$ )、键 ( $K_j$ ) 和值 ( $V_j$ ), 其中组索引  $j = 1, 2, \dots, M$ 。自注意力的计算方法可以用式 (4)、式 (5) 表示:

$$F^j = \text{Softmax} \left( \frac{Q_j \cdot K_j^T}{\sqrt{d}} \right) V_j + LEPE(V_j) \quad (4)$$

$$\tilde{h}_i = [F_i^1, F_i^2, \dots, F_i^M] \quad (5)$$

其中,  $d$  表示收缩因子,  $LEPE$  是由深度可分离卷积实现的具有局部信息的位置编码,  $\tilde{h}_i$  表示每个垂直分支的自注意力的串联矩阵。因此, 垂直分支 Ver-Attention 的自注意力矩阵是  $[\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_{k/2}]$ 。类似地, 水平带的自注意力矩阵 Hor-Attention 可以通过相同的方式得到。

最后, 通过融合水平和垂直方向上的注意力结果, CSWin-Attention 能够更有效地整合图像在二维空间中的全局上下文信息, 提升模型对于复杂场景的理解能力, 尤其适合处理类别交织、目标之间存在强语义关联的视觉任务。其结构如图 3 所示。

## 2 实验及结果分析

### 2.1 实验数据集

实验采用乡村道路数据集“Rural roads dataset”, 该数据集由 1285 张  $3840 \times 2160$  高清图像构成, 以新疆北部地区的乡村道路作为图像采集对象, 囊括了多种路面类型, 包括柏油路、水泥路、砂石路及土路, 并覆盖了日间和黄昏两个不同的时间段和晴、多云、阴、下雨 4 种天气情况。该数据集具有 40 种类别, 其中 20 种为实例类别, 主要类别有: 车辆、行人、牲畜、植物、交通标志、路灯、栅栏、广告牌等。数据集样本如图 4 所示。此外该数据集的所有类别对象进行像素级标注, 标注文件格式为 JSON, 可用于语义分割或实例分割计算机视觉算法的训练。

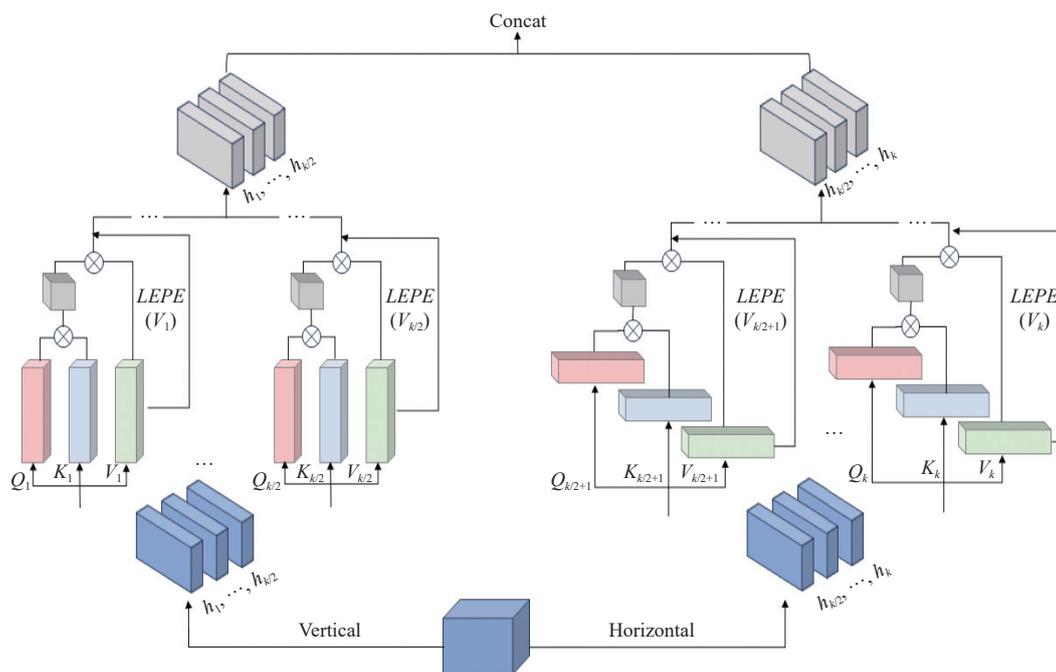


图3 十字形窗口自注意力结构图



图4 数据集样本集图

## 2.2 实验环境及参数设置

实验训练参数设置如表 1. 运行环境是基于 Python

3.7 的 PyTorch 1.7.0 框架, 使用了 4 块型号为 NVIDIA-RTX3090-24G 的 GPU, 运算平台为 CUDA 11.3, 为了

加快训练速度提高训练效果,将模型包装成了 DDP 模型进行并发训练. 实验中采用 Adam 作为优化器,将生成器的学习率设置为 0.001,判别器的学习率设置为 0.0001, batch\_size 设为 8, 惩罚项参数  $\lambda$  设为 10, 在迭代 2000 个 kimg 后停止训练. 因 StyleGAN 网络目前生成图像最大大小为  $1024 \times 1024$ , 因此在实验前对原数据集图像进行裁剪得到图像大小为  $1024 \times 1024$  的乡村道路样本集, 通过 StyleGAN 的工程将该样本集批量转换成 StyleGAN 生成格式的数据集共 1285 张图像.

表 1 模型训练参数设置

参数名	含义	数值
batch_size	一个批次的大小	8
learning rate-G	生成器学习率	0.001
learning rate-D	判别器学习率	0.0001
kimg	迭代次数	2000
batal	优化指数衰减率	0.5

### 2.3 实验设计

为了验证两种改进措施在网络性能上的增益作用, 设计了实验来评估这些改进对模型的影响. 首先, 选择了 StyleGAN 作为基准模型. 在此基础上, 分别实施了 2 种改进: 一是加入解耦映射网络, 二是引入 CCTB 模块; 最后组合这 2 种改进创建了 StyleGAN-ALL 模型. 通过定性和定量 2 个维度的评估对比, 验证了所提出改进措施的有效性.

### 2.4 评价指标

在目前的研究中, 对生成图像质量的评估主要从定性评估和定量评估两个方面进行<sup>[15]</sup>. 定性评估依赖于观察者的主观判断, 通过视觉观察和相似度评价来给出评分, 最终统计得出一个总体指标, 但这种方式主观性较强<sup>[15,16]</sup>. 相比之下, 定量评估提供了一种更客观的方法, 常用的定量评估方法包括 inception score (IS)<sup>[19]</sup>、Fréchet inception distance (FID)<sup>[20]</sup>等. 本实验为了客观地评估生成图像的质量, 选择了 IS 和 FID 作为主要的评估指标.

IS 是从图像生成质量和多样性方面客观地进行评估, 利用预先训练好的 inception 模型来计算边缘分布和条件分布之间的 KL (Kullback-Leibler) 散度<sup>[15,16]</sup>. 计算公式如式 (6) 所示:

$$IS(G) = \exp\left\{E_{x \sim P_g} D_{KL}[p(y|x) \| p(y)]\right\} \quad (6)$$

其中,  $x \sim P_g$  为生成图像;  $D_{KL}$  为 KL 散度. IS 得分越高, 即 KL 值越高, 说明生成图像的质量越高, 多样性越

丰富.

FID 是计算真实图像与生成图像特征向量之间的距离, FID 值越小, 说明生成图像越接近于真实图像, 生成的图片质量越好<sup>[15,16]</sup>. FID 的基本思想是用 inception 网络的卷积特征层作为一个特征函数  $\phi$ , 并且用特征函数将真实数据分布  $P_r$  和生成数据分布  $P_g$  建模为两个多元高斯随机变量. 这样就可以计算多元高斯分布的均值  $\mu_x$ ,  $\mu_g$  和方差  $\Sigma_x$ ,  $\Sigma_g$ <sup>[15]</sup>. 计算公式如式 (7) 所示:

$$FID(X, G) = \|\mu_x - \mu_g\|_2^2 + \text{tr}\left(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}\right) \quad (7)$$

其中,  $\mu_x$  为真实图像特征的均值;  $\mu_g$  为生成图像特征的均值;  $\Sigma_x$  为真实图像特征的方差;  $\Sigma_g$  为生成图像特征的方差.

## 2.5 实验结果

### 2.5.1 改进模型图像生成结果

为了全方位展示乡村道路场景在不同迭代阶段的生成效果, 如图 5 所示, 其中“kimg”代表模型训练的迭代周期数. 当 kimg 为 500 时, 生成的图像呈现出明显的失真问题: 道路形态扭曲, 树木与天空的界限不清, 车辆外形模糊, 显示了算法尚处于优化的早期阶段. 随着训练深入至 kimg 为 1000, 可以观察到显著的改善: 道路的不自然扭曲大幅减少, 路面的车道线变得明确, 尽管路边的植被细节仍略显模糊, 且树木与天空的分界线有待进一步区分, 这标志着生成质量正处于稳步提升的过程中. 进入 kimg 为 1500 的迭代时期, 图像的真实感获得飞跃: 天空与树木之间的边界变得清晰可辨, 细节如电线杆被精细地再现, 环境元素的多样性也有了显著增加, 反映了生成模型对复杂场景把握能力的增强. 最终, 当迭代至 kimg 为 2000 时, 所生成的乡村道路图像几近逼真, 不仅道路及周边环境的细节高度还原, 连同路边的建筑物也得以构建. 这一过程不仅展现了改进的 StyleGAN 模型通过迭代训练逐步克服失真、提升生成质量的轨迹, 也凸显了其在复杂场景图像合成方面的卓越潜力.

### 2.5.2 消融实验结果分析

在本次消融实验中, 旨在探究解耦映射网络和 CCTB 模块对 StyleGAN 网络在生成乡村道路场景图像性能的影响. 本文以 StyleGAN 作为基准模型, 构建了 4 个不同配置的变体模型: 模型 1 是原始的 StyleGAN; 模型 2 (StyleGAN-4FC) 在模型 1 的基础上增加了解耦

映射网络; 模型 3 (StyleGAN-CCTB) 在模型 1 的基础上增加了 CCTB 模块; 模型 4 (StyleGAN-ALL) 则同时集成了这 2 个模块. 通过定性和定量评估, 比较了这 4 种模型的性能, 发现模型 2 在减少图像扭曲方面表现较好, 但场景构成简单; 模型 3 在类别多样性上有所提

升, 但对复杂特征的表现不足; 模型 4 则在图像质量、边缘清晰度和场景复杂度上都有显著提升, 证明了两个模块联合使用时的增益效果. 最终, 实验结果揭示了每个模块对模型性能的具体贡献, 为进一步优化模型提供了指导. 定性评估结果如图 6 所示.



图 5 迭代训练效果图

为了更客观地评估各网络模型生成图像的效果, 对 4 种模型的 *IS* 和 *FID* 进行了对比, 结果如表 2 所示.

根据表 2 的数据, 可以观察到 StyleGAN-ALL 在 *IS* 和 *FID* 两个关键评估指标上均展现出显著的优势. 具体来说, StyleGAN-ALL 的 *IS* 值达到了 77.31, 较 StyleGAN 的 42.38 和 StyleGAN-4FC 的 48.27 有了显著提升, 显示出其生成图像质量的大幅改善. 此外, StyleGAN-ALL 的 *FID* 值降至 12.42, 相较于 StyleGAN-CCTB 的 16 和 StyleGAN 的 25.09 也有了显著降低, 这进一步证

实了 StyleGAN-ALL 在生成图像的真实性和多样性方面的卓越表现. 这些结果表明, 通过结合解耦映射网络和 CCTB 模块, StyleGAN-ALL 在图像生成任务中实现了性能的显著提升, 证明了这些改进措施的有效性和对模型性能的积极影响.

为了评估本研究设计的 StyleGAN-ALL 网络的性能, 对网络训练过程中的生成器损失、生成器正则化、路径长度正则化进行了监控, 监控结果如图 7 所示.



图6 消融实验生成图

表2 消融实验 IS、FID 值

模型	IS值	FID值
StyleGAN	42.38	25.09
StyleGAN-4FC	48.27	23.78
StyleGAN-CCTB	74.89	16
StyleGAN-ALL	<b>77.31</b>	<b>12.42</b>

在图7训练损失图的图7(a)中展示了生成器损失随训练迭代的变化.从图7(a)中可以看出,损失值从较高的数值开始,随着训练的进行逐渐下降,到训练到4000次的时候图像开始趋于平稳.在图7(b)中展示

了生成器正则化损失的变化.图中数据表明,在训练到将近4000次的时候损失值逐渐下降并趋于稳定,这表明正则化策略正在有效地帮助模型学习.图7(c)展示了路径长度正则化损失的监控结果.路径长度正则化是StyleGAN系列的一个创新点,它通过控制生成图像的路径长度来提高图像质量.从图7(c)中数据可知在训练到将近4000次的时候损失值逐渐下降并趋于稳定.这表明路径长度正则化策略正在有效地帮助模型学习.

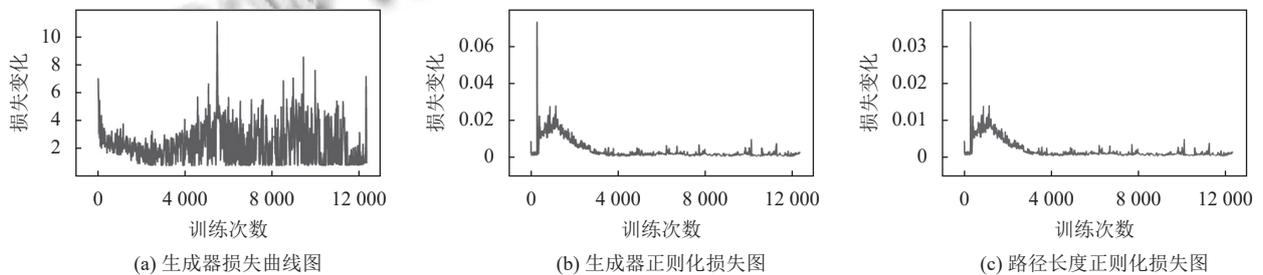


图7 训练损失图

### 3 目标检测应用评价及分析

#### 3.1 实验设计

在本实验中,采用了一种融合混合注意力<sup>[21]</sup>机制

和低频-高频分解技术的图像去噪算法.为了全面评估该算法的性能,准备了2个数据集:一个包含1285张真实噪声图像的原始数据集,以及一个合成数据集.合

成数据集由本文改进的 StyleGAN-ALL 生成对抗网络创造的 100 张合成图像组成, 这些图像经过标注后与真实图像合并, 扩充了数据集至 1385 张. 这种结合真实图像与合成图像的方法不仅增强了模型对不同噪声条件的适应性, 还提高了其在多样化环境下的鲁棒性, 从而为算法的有效性提供了更严格的验证.

为了评估数据增强对模型性能的具体影响, 对 Mask R-CNN 和 YOLOv8n 两个模型分别使用了原始数据集和经过 StyleGAN-ALL 增强后的数据集进行了实验训练. 同时构建了一个全面的模型性能评价体系, 采用如表 3 所示关键指标.

表 3 模型性能评估指标

指标名称	描述
平均交并比 (mIoU)	衡量预测边界与真实边界重叠程度的平均值
精确率 (Precision)	预测为正的样本中, 实际为正确的比例
召回率 (Recall)	实际为正的样本中, 预测为正确的比例
平均精度 (AP)	对于不同阈值下的精确率和召回率曲线下的面积
边界位移误差 (BDE)	预测边界与真实边界之间的平均距离

这些指标共同确保了实验用数据集不仅能够满足模型训练的需求, 还能为模型提供高质量的数据支持, 从而验证数据增强策略在提升模型性能方面的有效性. 通过这一综合评价框架, 可以准确地衡量和比较不同数据集对模型性能的具体影响, 确保实验结果的可靠性和模型的实用性.

### 3.2 实验结果及分析

上述 2 种实例分割模型在原始和合成数据集上的检测框与实例分割 mAP 指标值对比情况如表 4 所示, 其中 (B) 为目标检测任务, (M) 为实例分割任务.

表 4 检测框与实例分割 mAP 指标表 (%)

数据	算法	mAP50 (B)	mAP50 (M)	mAP50-95 (B)	mAP50-95 (M)
原始	Mask R-CNN	54.92	49.26	38.79	28.60
合成	Mask R-CNN	<b>55.84</b>	<b>50.07</b>	<b>39.23</b>	<b>29.08</b>
原始	YOLOv8n	84.13	77.12	68.64	49.40
合成	YOLOv8n	<b>84.84</b>	<b>77.24</b>	<b>68.96</b>	<b>49.52</b>

通过对比在这两种不同数据条件下的性能表现, 对于 Mask R-CNN 算法, 使用合成数据相比原始数据在  $f$  上提升了 0.92, 在 mAP50 (M) 上提升了 0.81, 同时在 mAP50-95 (B) 和 mAP50-95 (M) 这两个跨越多 IoU 阈值的指标上也有显著提升. 这些提升说明合成数据不仅提高了 Mask R-CNN 在较高置信度下的检测性能, 也改善了其在不同 IoU 阈值下的稳定性. 对于 YOLOv8n

算法, 合成数据在 mAP50 (B) 上带来了 0.71 的提升, 同时在 mAP50-95 (B) 上也有一定增长. 尽管在 mAP50 (M) 和 mAP50-95 (M) 上的提升幅度较小, 但整体上仍证明了合成数据对于 YOLOv8n 算法性能的正向影响. 基于实验数据的分析, 观察得到合成数据对于不同算法的性能指标产生了积极的影响.

此外, 将上述 2 种实例分割算法分别应用于原始数据集和合成数据集上, 对其在实例分割任务中的精度和召回率得到的结果如表 5 所示.

表 5 精度和召回率对比情况表 (%)

数据	算法	Precision (B)	Precision (M)	Recall (B)	Recall (M)
原始	Mask R-CNN	69.15	65.04	49.81	44.89
合成	Mask R-CNN	<b>69.25</b>	<b>67.28</b>	<b>50.96</b>	<b>46.82</b>
原始	YOLOv8n	92.86	87.76	77.81	71.61
合成	YOLOv8n	<b>93.44</b>	<b>88.16</b>	<b>78.77</b>	<b>72.70</b>

在对 Mask R-CNN 算法的测试中, 使用经 StyleGAN-ALL 增强的合成数据集在精确度 (Precision) 上取得了进步: 在目标检测任务中精确度从 69.15% 微升至 69.25%; 而在实例分割任务中精确度的提升更为显著, 从 65.04% 增至 67.28%. 此外, 合成数据集在召回率 (Recall) 方面也带来了显著提升, 无论是在目标检测任务提升了 1.15 个百分点, 还是在实例分割任务提升了 1.93 个百分点. 这些结果表明, 合成数据集对于提高 Mask R-CNN 算法的精确度和召回率均有积极作用, 尤其是在物体分割方面效果更为显著.

对于 YOLOv8n 算法, 合成数据集虽然在精确度上的提升不如 Mask R-CNN 显著, 但 Precision (B) 从 92.86% 增至 93.44%, Precision (M) 从 87.76% 增至 88.16%, 依然显示出正面影响. 在召回率方面, 合成数据集的影响更为显著, Recall (B) 提升了 0.96 个百分点, 而 Recall (M) 提升了 1.09 个百分点. 这说明合成数据集在提高 YOLOv8n 算法的召回率方面尤为有效.

综合实验结果, 引入 StyleGAN-ALL 数据增强技术后, 合成数据集在多种目标检测算法上均显示出性能指标的提升. 这一发现进一步证实了合成数据集在增强算法的精确度和召回率方面的有效性, 从而验证了数据增强技术在目标检测领域的应用价值.

## 4 结论与展望

当前, 乡村道路环境下的无人驾驶技术正成为研究的热点. 然而, 由于乡村道路图像数据集样本量有限, 这在一定程度上限制了深度学习模型的训练效果. 为

了扩充数据集并提高样本图像的数量,本研究基于 StyleGAN 网络,提出了改进算法 StyleGAN-ALL 网络模型.该模型引入了解耦映射层来降低图像中不同类别目标之间的耦合度,并设计了卷积耦合迁移块(CCTB)以增强模型对复杂场景的理解能力.

通过消融实验,对比了改进前后模型的 *IS* 和 *FID* 指标.结果显示,改进后的网络在 *IS* 值上提升了 82%, *FID* 值降低了 50%,表明生成的图像质量更为真实.此外,通过比较 Mask R-CNN 和 YOLOv8n 在原始数据集与经 StyleGAN-ALL 增强后的合成数据集上的性能,进一步验证了生成网络在提升实例分割模型的分割效能方面的显著作用.这不仅提高了模型的分割精度,还增强了其泛化能力.

展望未来,本研究将继续优化模型,以进一步提高生成图像的清晰度,特别是针对小目标的生成质量.这将有助于进一步提升无人驾驶系统在乡村道路环境下的性能和可靠性.

### 参考文献

- 1 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436–444. [doi: 10.1038/nature14539]
- 2 冯晓硕, 沈樾, 王冬琦. 基于图像的数据增强方法发展现状综述. *计算机科学与应用*, 2021, 11(2): 370–382. [doi: 10.12677/csa.2021.112037]
- 3 Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*, 2015, 349(6245): 255–260. [doi: 10.1126/science.aaa8415]
- 4 王建明, 陈响育, 杨自忠, 等. 不同数据增强方法对模型识别精度的影响. *计算机科学*, 2022, 49(6A): 418–423. [doi: 10.11896/jsjcx.210700210]
- 5 Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621, 2017.
- 6 Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 2018, 35(1): 53–65. [doi: 10.1109/MSP.2017.2765202]
- 7 田晨璐, 万耿宏, 翟远, 等. 基于 CycleGAN 的道路多场景图像数据增强与目标检测方法研究. *软件*, 2023, 44(12): 1–7. [doi: 10.3969/j.issn.1003-6970.2023.12.001]
- 8 Almahairi A, Rajeswar S, Sordani A, et al. Augmented CycleGAN: Learning many-to-many mappings from unpaired data. *Proceedings of the 35th International Conference on Machine Learning*. Stockholm: PMLR, 2018. 195–204.
- 9 李艳玲, 张博翔, 董萍, 等. 基于改进 DCGAN 的小样本玉米病害图像扩充方法. *河南农业大学学报*. <https://link.cnki.net/urlid/41.1112.S.20240925.1413.002>. (2024-09-25).
- 10 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *Proceedings of the 4th International Conference on Learning Representations*. San Juan: OpenReview.net, 2016.
- 11 Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 4396–4405. [doi: 10.1109/CVPR.2019.00453]
- 12 Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017. 1510–1519. [doi: 10.1109/ICCV.2017.167]
- 13 He KM, Gkioxari G, Dollár P, et al. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 386–397. [doi: 10.1109/TPAMI.2018.2844175]
- 14 Elaoua A, Nadour M, Cherroun L, et al. Real-time people counting system using YOLOv8 object detection. *Proceedings of the 2nd International Conference on Electronics, Energy and Measurement (IC2EM)*. Medea: IEEE, 2023. 1–5. [doi: 10.1109/IC2EM59347.2023.10419684]
- 15 陈佛计, 朱枫, 吴清潇, 等. 生成对抗网络及其在图像生成中的应用研究综述. *计算机学报*, 2021, 44(2): 347–369.
- 16 王威, 李玉洁, 郭富林, 等. 生成对抗网络及其文本图像合成综述. *计算机工程与应用*, 2022, 58(19): 14–36. [doi: 10.3778/j.issn.1002-8331.2205-0119]
- 17 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017. 6000–6010.
- 18 李真如. 基于注意力机制改善的图像描述生成 [硕士学位论文]. 上海: 上海交通大学, 2020. [doi: 10.27307/d.cnki.gsjtu.2020.001375]
- 19 Barratt S, Sharma R. A note on the inception score. arXiv: 1801.01973, 2018.
- 20 Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local NASH equilibrium. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: ACM, 2017. 6629–6640.
- 21 吴炳剑, 高琳, 李衍志, 等. 结合生成对抗网络与混合注意力机制的街景图像语义分割. *软件导刊*, 2024, 23(11): 187–192. [doi: 10.11907/rjdk.232155]

(校对责编: 张重毅)