DOI:10.19651/j.cnki.emt.2209413

基于生成对抗网络的自动装卸目标物标注数据集生成方法*

卢国杰[1] 王桂棠[1,2] 陈泳铮[1] 甘仕文[1] 林宗杰[1]
(1.广东工业大学机电工程学院 广州 510006; 2.佛山沧科智能科技有限公司 佛山 528225)

摘 要:针对建立无人起重装卸目标检测深度学习标注数据耗时问题,设计了货物图像检测生成对抗网络,构成准确的含语义标注和关键点标注的数据集,该数据集可用于有监督深度学习语义分割模型的训练。通过融合 StyleGAN 与 DatasetGAN 的生成对抗网络,对实际应用中存在的语义特征变形问题进行改进,将生成器的样本归一化层进行修改,去除均值操作,修改噪声模块和样式控制因子的输入方式;对纹理特征单一的物体的空间位置编码能力弱的问题,将生成网络的常数输入替换为傅里叶特征,并提出一个融合非线性上下采样的模块;最后引入 WGAN-GP 对目标函数进行改进。应用实验生成标签数据集,使用 Deeplab-V3 作为评价网络,以 DatasetGAN 方法作为基线,在语义标签生成任务上,Deeplab-V3 输出 mIOU 值提高 14.83%,在关键点标签生成任务上,L2 损失平均降低 0.4×10^{-4},PCK 值平均提高 5.06%,验证了改进的生成对抗网络生成语义及关键点标注数据的可行性和先进性。

关键词:生成对抗网络;标注数据生成;DatasetGAN;起重装卸
中图分类号:TP391.4 **文献标识码**:A **国家标准学科分类代码**:520.6040

Generation method of annotation data set of automatic loading and unloading objects based on generative adversarial network

Lu Guojie[1] Wang Guitang[1,2] Chen Yongzheng[1] Gan Shiwen[1] Lin Zongjie[1]
(1. School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China;
2. Foshan Cangke Intelligent Technology Co., Ltd., Foshan 528225, China)

Abstract: Aiming at the time-consuming problem of establishing deep learning labeling data for unmanned lifting target detection, a cargo image detection generation admision network was designed to construct an accurate data set containing semantic labeling and key point labeling, which could be used for the training of supervised deep learning semantic segmentation model. The generative adversation network of StyleGAN and DatasetGAN was fused to improve the semantic feature deformation in practical applications. The sample normalization layer of generator was modified to remove the mean operation and modify the input mode of noise module and style control factor. To solve the problem of weak coding ability of spatial position of objects with single texture feature, the constant input of generating network is replaced by Fourier feature, and a module integrating nonlinear up-down sampling is proposed. Finally, WGAN-GP is introduced to improve the objective function. Using deeplab-V3 as evaluation network and DatasetGAN as baseline, the output mIOU value of Deeplab-V3 increases by 14.83% on average in semantic label generation task, and L2 loss decreases by 0.4×10^{-4} on average in key point label generation task. PCK value is increased by 5.06% on average, which verifies the feasibility and advance of the improved generative adversarial network generation semantics and key point annotation data.

Keywords: adversarial network generation; annotation data generation; DatasetGAN; lift loading and unloading

0 引 言

在未知场景下实现目标货物自动识别检测是无人起重装卸的关键技术。目前最有效的是采用基于有监督的深度学习方法,但由于其建立带真值标签数据十分困难且耗费时间,导致推广应用时受限。

收稿日期:2022-03-26
* 基金项目:“佛山广工大研究院创新创业人才团队计划项目”(20191108)资助

深度学习在很多计算机视觉任务应用广泛,拥有相当优秀的模型预测精度。通常来说,用于计算机视觉任务的深度学习模型以有监督的方式进行训练,需要使用大量含有真值标签的数据,但标注数据获取困难,因此半监督学习、无监督学习受到了越来越多的关注,在无监督学习任务中,生成对抗网络(generative adversarial networks,GAN)[1]是最有前途的技术之一。迄今为止,GAN 在图像生成、语音合成、风格迁移等已经有不错的研究与应用[2]。

最近,GAN 网络如 proGAN[3]、StyleGAN[4]等,专注于使用对抗目标在大型数据集上进行训练后合成高质量图像,但后续仍需人工进行打标签工作。在域适应方法中,有几项研究[5-8]旨在通过利用图像到图像的翻译技术将标记的大型图像数据集翻译到另一个域,但由于存在不同数据集之间的风格和统计数据分布的差异性,使得其在特定数据集上训练好的深度学习模型也很难较好地迁移到数据分布不同的新场景中[9],从而效果较差甚至失效。

一些半监督学习方法[10-13]旨在对给定大量未标记图像和少量注释图像,学习比单独使用监督数据更好的分割网络。这些方法中的大多数将分割网络视为生成器,并使用少量真实注释对其进行对抗性训练。伪标签[14-15]和一致性正则化[16]也被探索用于语义分割,其中关键思想涉及在小型标记数据集上进行训练,并使用混合真实的标记数据和对未标记图像的置信度的预测。

分析现有方法可知,无论生成对抗网络或域适应方法要么耗费大量时间用于数据标记工作,要么存在域偏移问题,均无法有效解决带标记数据匮乏问题。

在本文工作中引入基于样式的生成对抗网络(StyleGAN)生成可控高质量图像,引入 DatasetGAN 网络[17]融合 StyleGAN 强大的渲染丰富语义信息能力,并结合具体应用场景对网络进行改进,通过使用少量人工注释货物图像进行数据扩增,生成准确的像素级标签数据集,最后设计实验进行评价。实验部分首先进行生成数据定性分析,然后使用评价网络 Deeplab-V3 作定量分析,以迁移学习、半监督学习和 DatasetGAN 为基线,对网络输出的 mIOU 值、L2 损失和 PCK 值进行对比。应用试验证明,该生成的标签数据集能够作为有监督深度学习语义分割模型的训练数据且训练后能达到较好的分割效果,有效解决了标注数据耗时问题。

1 基本原理和方法

1.1 生成对抗网络概述

GAN 是 Goodfellow 等[1]在 2014 年提出的一种深度学习网络模型,以学习训练数据内在分布的目标,通过生成器(generator,G)和判别器(discriminator,D)两个基本的模块构成的博弈最终达到纳什平衡状态。其中生成模型负责生成接近真实数据的伪数据,而判别模型 D 试图区分真实数据与生成模型 G 创造出的伪数据。生成器 G 的训练目标是为了生成让判别器 D 无法区分或者分辨样本来自真实数据还是生成器[18]。

GAN 的结构如图 1 所示。假设一个输入随机数据,进入生成器得到最初生成图像,并与真实图像进入判别器进行判别,得到判别结果后计算判别损失,再对判别器和生成器进行参数更新,反复迭代,直至判别器收敛。换句话说,D 和 G 玩极大极小值游戏,其价值函数为:

$$\min_G \max_D V(D,G) = \mathrm{E}_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1-D(G(z)))] \tag{1}$$

式中:$D(x)$表示数据 x 来自数据的概率;$p_z(z)$为输入噪声;将 $p_z(z)$到数据空间的映射表示为 $G(z)$。

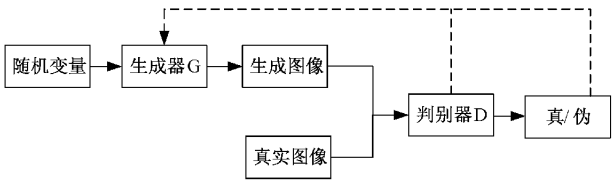

图 1 GAN 模型结构

1.2 StyleGAN 概述

基于样式的生成对抗网络(StyleGAN)是英伟达公司于 2019 年在 proGAN 基础上提出的一种生成对抗网络的变体模型。与其他对抗生成网络不同,StyleGAN 借鉴了风格迁移并引入可调噪音,实现了无监督地分离高级属性和随机变化的自动学习,生成高度可控图像。

在网络具体结构上,StyleGAN 延续了 GAN 的思想,其主要集中在生成器方面进行创新。在 StyleGAN 中将生成器 G 拆分为如图 2 所示的映射网络(mapping netowrk)f 与合成网络(synthesis network)g 两个子网络。在映射网络 f 中,通过改进对潜码 z 作的映射,将会使隐藏空间得到有效的解耦。合成网络 g 的作用与传统的 GAN 模型中的 G 相同,起到生成图像的功能,而且也采用了与传统 G 类似的 Block 结构[19],但在每个 Block 中都额外地输入由潜在空间 w 产生的仿射变换(A),用于控制生成图像的风格;输入转换后的随机噪声(B),用于丰富生成图像的细节。StyleGAN 中合成网络 g 的另一个重要改进是在每个 Block 中都添加了两个自适应样本归一化(adaptive instance normalization,AdaIN)层。AdaIN 层的工作过程如式(2)所示。

$$AdaIn(x_i,y) = y_{s,i}\frac{x_i-\mu(x_i)}{\sigma(x_i)} + y_{b,i} \tag{2}$$

由式(2)可以看出 AdaIN 需要首先对每个特征 x_i 作标准化处理,样式 y(放缩因子 $y_{s,i}$ 与偏差因子 $y_{b,i}$)会与标准化之后的卷积输出做一个加权求和,就完成了一次 W 影响原始输出 x_i 的过程。

1.3 DatasetGAN 概述

StyleGAN 虽然能利用无监督的数据集生成高清的、

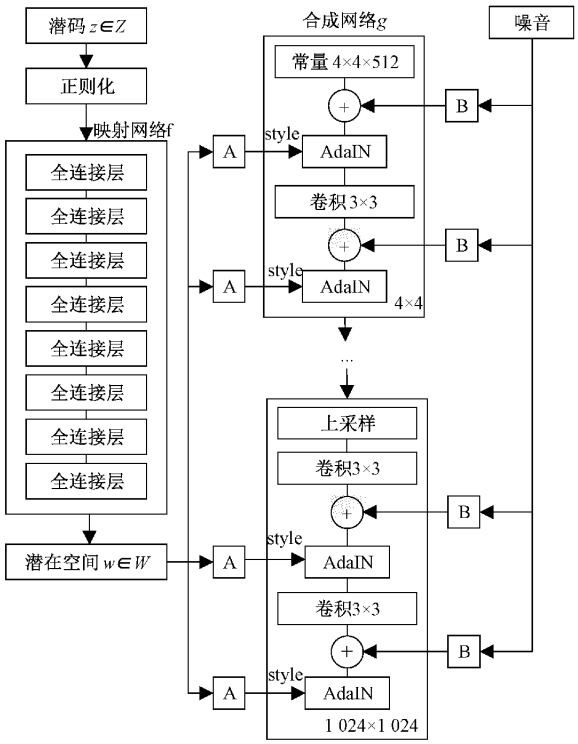

图 2　StyleGAN 的生成器结构

不同风格、不同属性、逼真的图像，但生成的数据无法直接用于目标检测，仍需耗费大量时间进行数据标注工作。而 DatasetGAN 训练解码器只需要几个带标签的示例即可泛化到潜在空间的其余部分，从而产生无限的带注释数据集生成器。然后，这些生成的数据集可以像真实数据集一样用于训练任何计算机视觉架构。

DatasetGAN 主要关注像素级标注任务，例如语义分割和关键点预测，因为它们是最耗时的手动标注任务的典型示例。DatasetGAN 的关键见解是，生成模型必须在其高维潜在空间中获取语义知识，然后经过训练以合成高度逼真的图像。例如，StyleGAN 等架构中的潜在代码包含控制 3D 属性（如视点和对象身份）的解开维度[4,20]。已证明在两个潜在代码之间进行插值可以产生逼真的生成，这表明 GAN 还学会了在语义和几何上对齐对象及其部分。DatasetGAN 旨在利用图像 GAN 的这些强大特性。具体来说，就是将来自 StyleGAN 的特征图上采样到最高分辨率，以便为合成图像上的所有像素构建逐像素特征向量。然后训练一组 MLP 分类器，将像素特征向量中的语义知识解释为其部分标签。

2　基于无人装卸场景数据的生成对抗网络模型构建

2.1　现有模型存在问题及改进

许多学者在 GAN 提出后便不断地对其进行优化并且衍生出了其他 GAN 模型。具体实践中，利用 GAN 生成装卸场景标注数据时还存在其它一些问题，针对这些问题，本文结合了其它模型的优点和特性进行改进。

1）为生成器引入 StyleGAN 网络

为了解决 GAN 模型黑盒运行，缺乏对图像合成过程各个方面的理解且生成图像模糊的问题，引入了基于样式的生成对抗网络（StyleGAN），从而有效控制生成图像质量和特征。

2）为生成器引入 DatasetGAN 网络生成语义标签信息

StyleGAN 虽然能生成不同属性和风格的图像，却没有利用其渲染能力捕获丰富语义信息进行处理，因此引入 DatasetGAN 生成带像素级标签的图像，以少量详细标注的图像生成大量带标签的数据集。

3）语义特征变形问题改进

生成图像存在语义特征变形的问题，据分析是主要因为均值操作破坏了在特征量级中相对于彼此发现的任何特征信息。受 StyleGAN2[21] 启发，本文对生成器的自适应样本归一化层进行修改，去除均值操作，修改噪声模块、放缩因子 w_i 和偏差因子 b_i 的输入位置。具体在 3.5 节设置消融实验验证其有效性。

4）提高空间位置提取能力

实际应用中，装卸场景下待识别物体如圆桶、木箱等，大多纹理特征单一，造成生成器无法有效地提取物体的空间位置信息，生成的语义标签像素粘连，没有空间层次。据分析，一是因为训练数据量较少，二是虽然 StyleGAN 等架构中的潜在代码包含控制 3D 属性（如视点和对象身份）的编码，但对纹理特征单一的物体的空间提取能力仍然较弱。具体地，本文将生成网络的常数输入替换为傅里叶特征，由于特征映射有一个无限的空间范围，通过引入一个固定大小的边界来近似，每一层操作之后再对权重参数进行裁剪（crop）。在 crop 操作前的将非线性部分融合进上采样过程中，即进行上采样 4×-LeakyReLU 非线性激活-下采样 2×。为了增加空间旋转等变形，在所有层上用 1×1 卷积代替 3×3 卷积，来增加特征映射数量。

5）目标函数的改进

生成器的目标是生成与真实数据分布接近的数据，那么目标函数就是影响到 GAN 模型的直接因素。DatasetGAN 中使用 JS（Jensen-Shannon）散度[22]作为像素的不确定性度量，通过最小化 JS 散度实现最小化生成器的损失函数。然而 JS 散度由于其自身函数域的问题，在训练中容易发生梯度消失的现象，从而导致生成数据与真实数据分布差异较大。本文引入 WGAN-GP（wasserstein GAN-gradient penalty）[23]代替 JS 散度。WGAN-GP 通过梯度惩罚机制代替了权重裁剪以强制执行 Lipschitz 约束，解决了 JS 散度因函数域导致的梯度消失问题和 WGAN（wasserstein GAN）[24] 中权重剪裁导致的梯度消失、参数不集中等问题，还有更好的稳定性和更多的多样性。

WGAN-GP 如式(3)所示。

$$W(P_{data},P_G)=\max_D\{E_{x\sim P_{data}}[D(x)]-E_{x\sim P_G}[D(x)]-\lambda E_{x\sim P_{penalty}}[(\|\nabla xD(x)\|-1)^2]\} \tag{3}$$

式中:P_{data} 是真实分布,P_G 是生成器分布,隐含定义了从 P_{data} 和 P_G 采样的点对之间的直线均匀采样的 $P_{penalty}$。式中第 3 项为惩罚项,目的是让梯度尽可能趋向 1。一个“好”的判别器应该在 P_{data} 附近是尽可能大,要在 P_G 附近尽可能小。

2.2　模型构建

改进后的完整网络结构如图 3 所示。网络首先省略输入层,并从一个学习的常数开始。给定输入潜在空间 Z 中的潜码 z,其是一个隐藏表达向量(512×1),进行归一化处理后输入到由 8 个全连接层(fully connected layers,FC)组成的映射网络,将输入向量学习的仿射变换得到中间向量 w(512×1),从而将特征解缠后的中间向量 w 变换为样式控制向量,即非线性映射网络 $f:Z\rightarrow W$ 首先产生 $w\in W$。

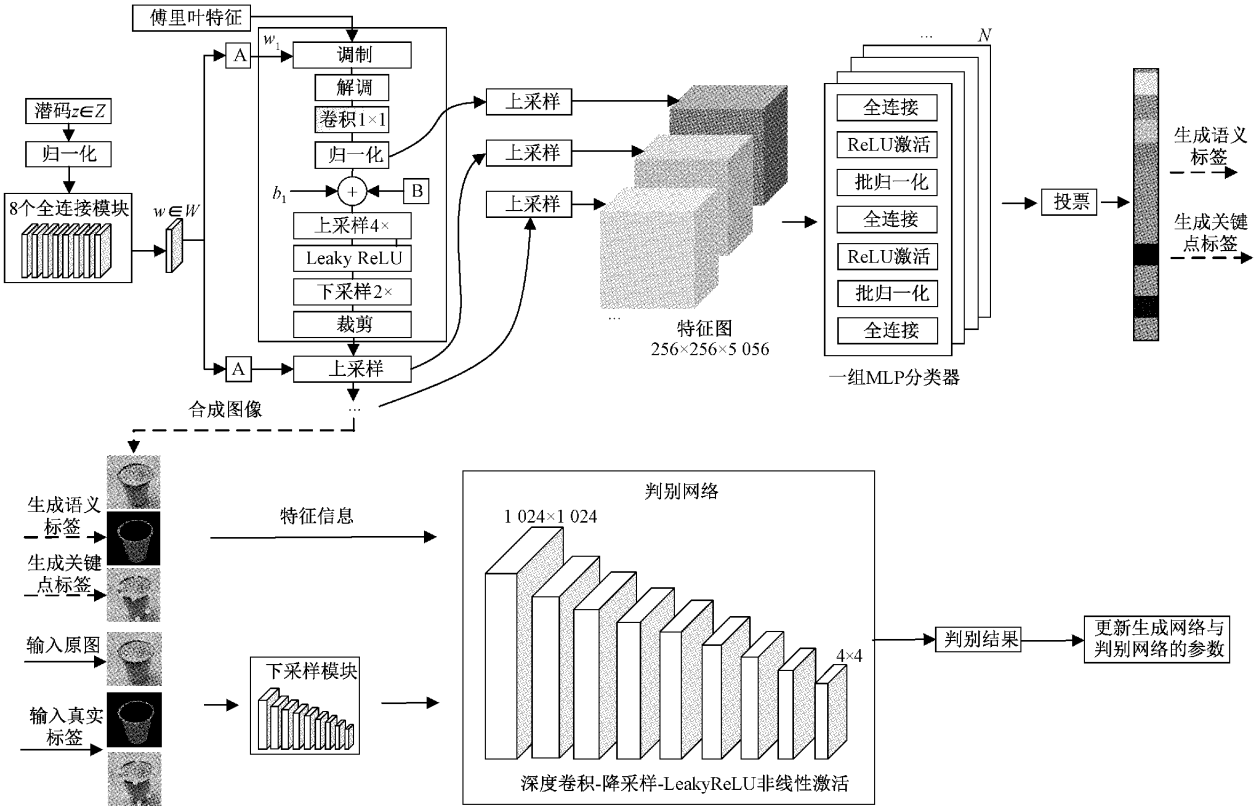

图 3　基于装卸场景数据的生成对抗网络模型

生成器由于从 4×4,上采样到 8×8,并最终上采样到 1 024×1 024 的特征向量,所以它由 9 个合成块组成。首先输入傅里叶(Fourier)特征,每个合成块由上采样块、调制、解调、1×1 卷积块和归一化组成,受样式控制变量 A 和噪音 B 施加影响。其中一个样式控制向量在上采样之后对其影响一次,另外一个样式控制向量在归一化之后对其影响一次。与 StyleGAN 不同的是,噪音模块的输入位置移到样式的有效区域之外,即在归一化后。调制根据传入的样式 $w_{i,j}$ 缩放卷积的每个输入特征图,具体的通过缩放卷积权重来实现,如式(4)所示。

$$w'_{i,j}=v_i\cdot w_{i,j} \tag{4}$$

式中:w 和 w' 分别是原始权重和调制权重,v_i 是与第 i 个输入特征图相对应的比例,而 j 是枚举输出特征图。

实例归一化的目的是从卷积输出特征图的统计数据中去除 v 的影响。如式(5)所示,经过调制和卷积后,输出激活的标准差,随后进行解调,缩放每个输出特征图 j 为:

$$w''_{i,j}=\frac{1}{\sqrt{\sum_{i-1}^{n}w'^{2}_{i,j}}} \tag{5}$$

因此,修改后的 AdaIN 层的工作过程如式(6)所示。

$$AdaIn(x_i,y)=w''_{i,j}\frac{x_i-\mu(x_i)}{\sigma(x_i)} \tag{6}$$

借助 AdaIN 层对特征 x_i 的处理与加入外部信息,合成网络 g 的每一层就可以通过接收外部的单通道显式噪声 B 和样式控制偏差因子 b_i 向生成的图像中添加局部随机的信息,从而直接生成随机细节特征图$\{S^0,S^1\cdots,S^k\}$。

上采样模块采用上采样 4×-LeakyReLU 非线性激活-下采样 2× 的结构,上采样后进行裁剪(crop)操作。LeakyReLU 激活函数如式(7)所示。

$$LeakyReLU(x)=\max(0,x)+negative\_slope*\min(0,x) \tag{7}$$

图像的上采样操作完成后，进行语义标签特征学习。将 AdaIN 输出的特征图$\{S^0, S^1 \cdots, S^k\}$上采样到最高输出分辨率(1 024×1 024)，并将它们连接起来得到一个 3D 特征张量 $\boldsymbol{S}^* = (S^{0,*}, S^{1,*} \cdots, S^{k,*})$，其大小为 256×256×5 056。输出图像中的每个像素 i 都有自己的像素特征向量 $\boldsymbol{S}^* = (S_i^{0,*}, S_i^{1,*} \cdots, S_i^{k,*})$，其参数为 1×1×5 056。为简单起见，我们共享所有像素的权重。

由于特征向量 $\boldsymbol{S}_i^*$ 具有高维数(5 056)，并且特征图具有高空间分辨率(最多 1 024)，则网络不能轻易地批量消耗所有图像特征向量。因此，对每个图像的特征向量进行随机采样，从而确保从每个标记区域中至少采样一次。对于语义分割标签生成任务，使用交叉熵损失训练分类器，如式(8)所示。

$$C = -\frac{1}{n}\sum_x [y\ln a + (1-y)\ln(1-a)] \tag{8}$$

而对于关键点预测标签生成任务，为训练集中的每个关键点构建一个高斯热图，并使用多层感知机(multilayer perceptron，MLP)函数去拟合每个像素的热值。MLP 包括 3 层：输入层、隐层和输出层。层与层之间是全连接的，这里使用 ReLU 激活函数代替常用的 *sigmod* 激活函数进行非线性映射。

为了摊销随机抽样的影响，训练了 N 个 MLP 分类器的集合，取 $N=10$。在测试时对每个像素使用多数投票进行语义分割。对于关键点预测，平均每个像素的 N 个分类器预测的 N 个热值。最后将 MLP 输出的像素特征向量中的语义知识解释为其部分标签和生成关键点标签。

将生成器合成的高清图像、语义标签、关键点标签与下采样后的真实图像、真实语义标签、真实关键点标签送入判别器 D，来判断合成数据的真实性。判别器 D 采用常用的深度卷积、降采样和 LeakyReLU 非线性激活的网络结构。通过提取各层的特征信息来判断其类别，输出二值标签，指示结果的真实与否。最后使用 WGAN-GP 目标函数计算真实数据的概率分布与生成数据的相似度，更新生成网络和判别网络的参数。

3　应用试验与分析

在本文作者开发的无人起重装卸智能测控系统中，使用装卸场景中的圆桶和方箱两种类型货物对设计的生成对抗网络模型进行测试验证和评估。

3.1　装卸场景数据集

测试所采用的数据集包含 2 000 张圆桶真实图像，2 000 张方箱真实图像，分辨率为 1 024×1 024，使用 LabelMe 分别对其中 40 张图像进行语义标注和关键点标注，平均每张圆桶图像注释时间为 92 s，木箱为 76 s。GAN 圆桶图像有两个注释对象，分别为上表面(tsurface)和侧面(side)；GAN 方箱图像有 3 个注释对象，分别为 tsurface、side 和木桩(timpile)；GAN 圆桶关键点有 3 个注释对象，分别为上表面(ktsurface)、侧面(kside)和底部(kdsize)；GAN 木箱关键点有两个注释对象，分别为角点(angular)和木桩点(buttom)。注释完成后按 5∶2∶3 的比例划分训练集、验证集与测试集。通过本文模型，生成 12 000 张带标签信息的合成图像数据集。

3.2　试验环境

为验证本文模型的网络性能，需要搭建相应的实验平台，该平台由硬件平台和软件平台组成。其中硬件平台主要是深度学习工作站，具备较强的并行计算能力，主要的配置信息为：32 G DDR4 2 400 MHz(内存条)、GeForce GTX 1080Ti(显卡)等。而软件平台包括操作系统 Ubuntu18.04(64 位)，深度学习框架 Pytorch 1.4.0。

3.3　评价指标

为了简单起见，对于语义注释生成任务，本文与 DatasetGAN 保持一致，使用带有 ImageNet 预训练的 ResNet151[25]主干的 Deeplab-V3[26]作为要在生成的数据集上训练的语义分割网络。让 Deeplab-V3 为每个像素在所有部分标签上输出一个概率分布，即平均交并比(mean intersection over union，mIOU)，其是语义分割任务标准评价指标，指计算真实值(ground truth)和预测值(prediction)两个集合的交集和并集之比。mIOU 的取值范围为[0，1]，值越大，预测的分割图越准确。

对于关键点检测任务的完全监督基线，采用相同的策略和设置，只是该任务的模型输出的是热图而不是概率图，并且使用 L2 损失而不是交叉熵损失。L2 值越小，关键点预测越接近。关键点检测的另一项评价指标 PCK(Percentage of Correct Keypoints)，是计算检测的关键点与其对应的 Ground truth 间的归一化距离小于设定阈值的比例。

3.4　训练流程

在 DatasetGAN 中，在训练开始前，需要每个类别的 StyleGAN 预训练模型[17]。本文为了便于训练，将训练过程分了两个步骤。第 1 阶段将网络拆分，屏蔽生成网络中的标签生成模块，先进行图像生成网络的预训练，同时基于随机采样激活标签生成模块，进行参数学习。第 2 阶段，取消屏蔽标签生成模块，进行完整的网络训练。

在训练过程中，模型的超参数设置如表 1 所示。

表 1　模型的超参数

类目	类型/数值
优化器	Adam 优化器
学习率	0.000 1
训练批量	60
训练迭代次数	150

3.5　试验结果及分析

1)定性结果

在进行模型训练时尝试使用不同分辨率图像进行训

练,注意到使用更高分辨率的图像,会产生更准确的合成注释。因此本文图像采用 1 024×1 024 分辨率。如图 4 所示,展示了本文生成图像注释的定性结果,可见生成语义标签很好地标记了物体,但在类别边缘区域略粗糙。

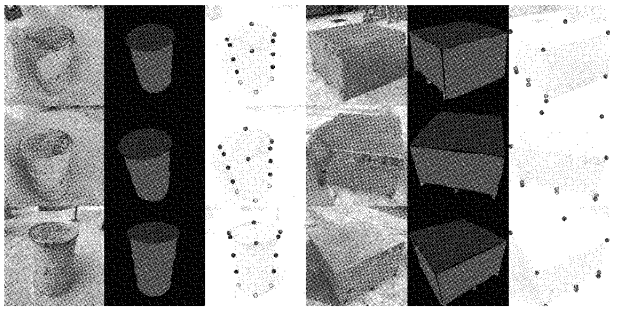

图 4　生成注释图像定性结果

2)不同种类数据集的试验结果

由于本文提出的网络是基于装卸场景货物数据集的,因此无法像 DatasetGAN 使用开源数据集训练,将结果与迁移学习和半监督模型基线进行比较,但依然列出 DatasetGAN 使用多种数据集(ADE-Car-5、Car-20、CelebA-Mask-8(Face)、Face-34、Bird-11、Cat-16、Bedroom-19)的实验结果以及本文方法在本文数据集的实验结果(如表 2 所示),以便与本文方法进行一定的参照比较。

表 2　不同种类数据集的实验结果

方法	数据集	mIOU/%
DatasetGAN	ADE-Car-5	57.77
	Car-20	62.33
	CelebA-Mask-8 (Face)	70.01
	Face-34	53.46
	Bird-11	36.76
	Cat-16	31.26
	Bedroom-19	36.83
本文方法	圆桶和方箱	76.26

由表 2 可知,本文方法在圆桶和方箱语义标签生成任务上比 DatasetGAN 在 Car、Face、Bird、Cat 等数据集上的 mIOU 值高出不少。

3)圆桶和方箱数据集在不同任务上的表现

进一步地,将本文方法与原 DatasetGAN 方法、迁移学习和半监督基线进行比较(如表 3 所示),使用本文的图像训练这些方法在生成的未标记真实图像上。对于迁移学习基线,使用预先训练的 MSCOCO[27] 语义分割权重初始化网络,并以监督方式微调本文人工注释数据集的最后一层,需对源域进行迁移至本文目标域。对于半监督基线,采用文献[28]并使用本文相同的预训练主干,即 Deeplab-V3。如图 5 所示,展示了在本文生成数据集上训练的 Deeplab-V3 的预测与真值标签进行比较的定性结果,虽然不完美,但结果显示本文方法有不错的质量标签输出。

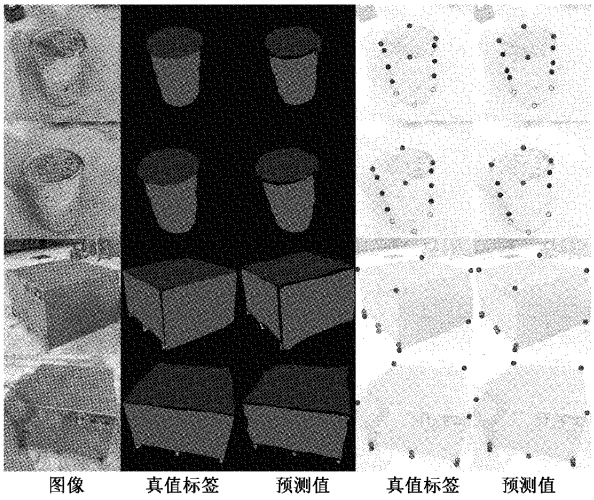

图 5　预测与真实标注的定性结果

表 3　圆桶和木箱数据集在不同任务上的表现

类目	圆桶 IOU/%	木箱 IOU/%
迁移学习	48.36	49.16
半监督	51.22	50.01
DatasetGAN	62.53	60.33
本文方法	77.62	74.89

由表 3 可知,针对本文数据集,本文方法在两类物品上的 IOU 值都大大优于迁移学习和半监督基线,且在圆桶上比 DatasetGAN 高出 15.09%,在木箱上高出 14.56%,平均提高 14.83%。同时,如图 6 所示,展示了本文标记数据集中训练图像的数量与 mIOU 的关系,当提供的标记图像数量达到 25 张时,本文方法即生成可用于监督网络训练的标签数据集,并有相当优秀的结果。

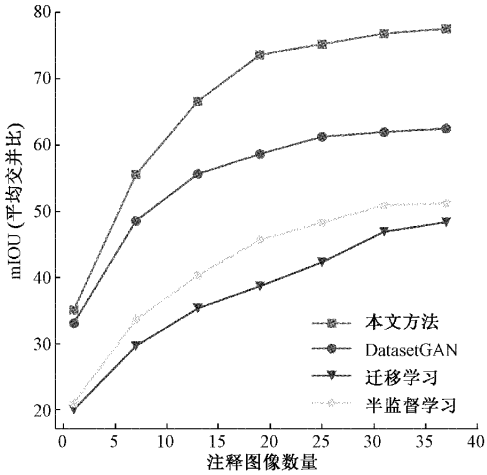

图 6　标记数据集中训练图像的数量与 mIOU 的关系

4)消融实验

(1)生成数据集大小的消融研究

使用 25 个训练示例消融了圆桶和方箱数据集生成数

量大小的选择，如表 4 所示，生成示例从 3 000 张增加到 15 000 张时可以提高性能，但进一步添加更多数据时，改进将微不足道，这与 DatasetGAN 在 16 个汽车类别示例的生成实验结果相近。在本文以上实验中，我们将生成的数据集的大小设置为 15 000 张，并过滤掉前 20%的不确定实例，即实际生成数据集大小为 12 000 张。

表 4　生成数据集大小的消融研究结果

生成数据集大小/张	圆桶 IOU/%	方箱 IOU/%
3 000	65.36	66.51
5 000	70.49	70.24
10 000	74.24	72.01
15 000	77.62	74.89
20 000	77.81	75.12

(2)改进内容消融研究

针对语义特征变形的改进、空间提取能力的改进和目标函数的改进，设计消融实验。如表 5 所示，展示了消融实验的结果，对于语义特征变形的改进，本文方法比 DatasetGAN 输出的 IOU 值分别提高 2.74%和 3.48%；对于空间提取能力的改进，分别提高 4.99%和 5.20%；对于目标函数的改进，分别提高 1.41%和 1.69%。通过消融实验可见各改进点对模型均产生积极影响，其中对模型空间提取能力的改进提高比例最大，改进效果最好。

表 5　改进内容消融研究结果

方法	圆桶 IOU/%	木箱 IOU/%
DatasetGAN	62.53	60.33
语义特征变形的改进	65.27	63.81
空间提取能力的改进	67.52	65.53
目标函数的改进	63.94	62.02
本文方法	77.62	74.89

5)关键点标签生成任务评价实验

本文遵循关键点检测的常见做法，即预测高斯热图而不是关键点位置，使用与部分分割实验相同的策略和设置，并且使用 L2 损失代替交叉熵损失。与语义标签评价一样，使用圆桶和木箱数据集，将本文的方法与 DatasetGAN、迁移学习基线进行比较。关键点标签生成任务性能评价实验结果如表 6 所示，定性结果如图 5 所示。

表 6　关键点标签生成任务性能评价实验结果

数据集	指标	迁移学习	DatasetGAN	本文方法
圆桶	L2 损失	4.2×10^{-4}	2.3×10^{-4}	2.0×10^{-4}
	PCK^{k}_{meam}	42.51	78.50	83.68
木箱	L2 损失	4.6×10^{-4}	2.4×10^{-4}	1.9×10^{-4}
	PCK^{k}_{meam}	39.95	76.31	81.25

结果表明，L2 损失平均降低 0.4×10^{-4}，PCK 值平均提高 5.06%，本文的方法均优于使用相同注释数量的其它方法。

4　结　　论

本文针对装卸场景货物标签数据生成进行研究，提出了一种功能强大的半监督学习方法，利用改进的 StyleGAN 和 DatasetGAN 网络学习潜在空间和训练分类器，证明了可以仅从极少数人工注释的图像即可生成庞大的语义及关键点标记数据集，该数据集能有效用于有监督目标检测任务网络的训练。并且本文方法被证明在使用生成的装卸场景数据集训练的视觉任务中显著优于迁移学习、半监督学习以及 DatasetGAN 基线。本文研究的不足在于仅针对特定数据集进行网络改进，下一步将结合其他数据集的共性，研究提高生成数据集网络模型的通用性。

参考文献

[1] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.

[2] 汪美琴，袁伟伟，张继业. 生成对抗网络 GAN 的研究综述[J]. 计算机工程与设计，2021，42(12)：3389-3395.

[3] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[C]. International Conference on Learning Representations, 2018, DOI: 10.48550/arXiv.1710.10196.

[4] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4401-4410.

[5] ZOU Y, YU Z, KUMAR B, et al. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 289-305.

[6] MUREZ Z, KOLOURI S, KRIEGMAN D, et al. Image to image translation for domain adaptation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4500-4509.

[7] VU T H, JAIN H, BUCHER M, et al. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 2517.

[8] CHOI J, KIM T, KIM C. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation[C]. Proceedings of the IEEE/

CVF International Conference on Computer Vision, 2019: 6830-6840.

[9] 孙琦钰, 赵超强, 唐漾, 等. 基于无监督域自适应的计算机视觉任务研究进展[J]. 中国科学:技术科学, 2022, 52(1): 26-54.

[10] HUNG W C, TSAI Y H, LIOU Y T, et al. Adversarial learning for semi-supervised semantic segmentation [C]. 29th British Machine Vision Conference, BMVC 2018.

[11] LUC P, COUPRIE C, CHINTALA S, et al. Semantic segmentation using adversarial networks [C]. NIPS Workshop on Adversarial Training, 2016, DOI: 10.48550/arXiv.1611.08408.

[12] SOULY N, SPAMPINATO C, SHAH M. Semi supervised semantic segmentation using generative adversarial network [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 5688-5696.

[13] KE Z, QIU D, LI K, et al. Guided collaborative training for pixel-wise semi-supervised learning [C]. European Conference on Computer Vision, 2020: 429-445.

[14] BERTHELOT D, CARLINI N, GOODFELLOW I, et al. Mixmatch: A holistic approach to semi-supervised learning [J]. Advances in Neural Information Processing Systems, 2019, DOI: 10.48550/arXiv.1905.02249.

[15] SOHN K, BERTHELOT D, CARLINI N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence [J]. Advances in Neural Information Processing Systems, 2020, 33: 596-608.

[16] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results [J]. Advances in Neural Information Processing Systems, 2017, DOI: 10.48550/arXiv.1703.01780.

[17] ZHANG Y, LING H, GAO J, et al. Datasetgan: Efficient labeled data factory with minimal human effort[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10145-10155.

[18] 王桂棠, 林桢哲, 符秦沈, 等. 联合生成对抗网络的肺结节良恶性分类模型[J]. 仪器仪表学报, 2020, 41(11): 188-97.

[19] 董虎胜, 刘诚志, 朱晶, 等. 基于 StyleGAN 的动漫图像生成[J]. 甘肃科技纵横, 2021, 50(7): 14-16.

[20] ZHANG Y, CHEN W, LING H, et al. Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering [C]. International Conference on Learning Representations, 2020, DOI: 10.48550/arXiv.2010.09125.

[21] KARRAS T, LAINE S, AITTALA M, et al. Analyzing and improving the image quality of stylegan [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 8110-8119.

[22] BELUCH W H, GENEWEIN T, NüRNBERGER A, et al. The power of ensembles for active learning in image classification [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9368-9377.

[23] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein GANs[J]. Advances in Neural Information Processing Systems, 2017, DOI: 10.48550/arXiv.1704.00028.

[24] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[J]. Advances in neural information processing systems, 2017, DOI: 10.48550/arXiv.1701.07875.

[25] 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, 2016: 770-778.

[26] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

[27] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]. European Conference on Computer Vision, 2014: 740-755.

[28] MITTAL S, TATARCHENKO M, BROX T. Semi-supervised semantic segmentation with high-and low-level consistency [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(4): 1369-1379.

作者简介

卢国杰,硕士研究生,主要研究方向为深度学习、半监督学习、计算机视觉。

E-mail: 635496946@qq.com

王桂棠(通信作者),教授、硕士生导师,主要研究方向为仪器科学与技术、机器视觉。

E-mail: wanggt@gdut.edu.cn

陈泳铮,硕士研究生,主要研究方向为计算机视觉、点云处理。

E-mail: 1154955157@qq.com

甘仕文,硕士研究生,主要研究方向为深度学习、密集型运算加速处理。

E-mail: 1249835519@qq.com

林宗杰,硕士研究生,主要研究方向为深度学习,计算机视觉。

E-mail: 113026704984@163.com