航拍场景下的车辆生成

陶晓力,刘宁钟
(南京航空航天大学 计算机科学与技术学院,江苏 南京 211106)

摘　要:随着智能交通的提出,结合无人机的航拍车辆检测有着越来越多的应用。目前在车辆检测方面,基于 CNN 的目标检测方法如 faster-rcnn、yolo 等都达到了很高的水准,但也存在着需要收集大量标注数据进行训练的问题。而通过图像生成方法解决训练样本的获取是一个可行的解决方案。但一般的生成模型要么只能生成车辆,没有背景信息,要么只能拟合背景,生成车辆严重失真。对此,文中在 pix2pixGAN 的基础上提出多条件约束的生成对抗网络,用以在真实航拍场景图像中生成带位置标注信息的车辆。通过在生成对抗网络中设立多判别器的方法分别约束背景的拟合以及图像中车辆的生成,将图像中预先设置的噪声区域完美转化成车辆图像。对比实验结果显示,该车辆生成模型能够很好地在航拍图像中生成较为逼真的车辆。

关键词:GAN;车辆生成;pix2pix;多条件约束
中图分类号:TP31　　**文献标识码**:A　　**文章编号**:1673-629X(2019)12-0162-05
doi:10.3969/j.issn.1673-629X.2019.12.029

Vehicle Generation in Aerial Scenes

TAO Xiao-li,LIU Ning-zhong
(School of Computer Science and Technology,Nanjing University of Aeronautics and Astronautics,
Nanjing 211106,China)

Abstract:With the introduction of intelligent transportation,aerial vehicle detection combined with UAV has more and more applications. At present,in terms of vehicle detection,CNN-based target detection methods,such as faster-rcnn,yolo,have reached a high level,but there is still a problem that a large amount of labeled data needs to be collected for training. It is a feasible solution to obtain training samples by image generation. However,the general generated model can only generate vehicles without background information,or it can only fit the background and generate vehicle with severe distortion. Based on pix2pixGAN,we propose a multi-condition constrained generation adversarial network to generate vehicles with positional annotation information in real aerial scene images. The noise region preset in the image is perfectly converted into a vehicle image by constraining the fitting of the background and the generation of the vehicle in the image by respectively setting up a multi-discriminator in the generation confrontation network. The comparison experiment shows that the proposed vehicle generation model can generate a more realistic vehicle in the aerial image.

Key words:GAN;vehicle generation;pix2pix;multi-condition constraint

0　引　言

随着无人机飞控技术的快速发展以及摄像云台的成熟,民用无人机航拍的应用领域逐渐扩大。尤其是政府对于智能交通的逐渐重视,无人机在交通事故检测、道路勘探等方面起着越来越重要的作用[1]。这其中,航拍车辆检测作为智能交通领域的核心,更是研究的重点。

目前,基于卷积神经网络(CNN)的目标检测虽然已经在多个数据集上取得了不错的结果[2-5],但在航拍小目标检测中,由于真实场景的多变性,需要大量的人力物力和精力来收集多场景下的图片并进行标注。同时,有限的训练数据很难让检测器达到一个较好的性能。因此,文中基于生成对抗网络提出了一种真实场景下的航拍车辆生成方法。

收稿日期:2018-12-24　　**修回日期**:2019-04-25　　**网络出版时间**:2019-09-24
基金项目:国家自然科学基金(61375021);南京航空航天大学研究生创新基地(实验室)开放基金资助(kfjj20171608);中央高校基本科研业务费专项资金
作者简介:陶晓力(1993-),男,硕士,CCF 会员(89015G),研究方向为计算机视觉和模式识别;刘宁钟,教授,博导,研究方向为计算机视觉和模式识别。
网络出版地址:http://kns.cnki.net/kcms/detail/61.1450.TP.20190924.1534.008.html

1　相关工作

不同于判别模型学习样本的条件概率分布,生成模型通过从训练集和标签中学习联合概率分布,并产生与训练集同样分布的图像。传统机器学习的生成模型很难拟合真实的样本分布,而随着深度学习的发展,Goodfellow 等提出的生成对抗网络[6](generative adversarial nets,GAN)已经在图像合成领域获得了很多的成功应用[7]。

文献[8]提出了一种 DCGAN 模型,将 CNN 与 GAN 结合,使用卷积网络代替了 GAN 中的全连接网络,利用卷积网络强大的特征提取能力来提高生成网络的学习效果。文献[9]提出了一种图像风格迁移模型(pix2pixGAN),通过图像对的方式进行一对一的监督训练,能够将准备好的原始图像转换成所希望的风格。

但是要想使得 GAN 生成的目标图像能够辅助检测器的训练,还面临两大挑战:一是 GAN 能生成足够真实的目标图像;二是如何使得生成的目标图像融合到周围的场景中[10-12]。

2　模型和方法

生成对抗网络是由一个生成网络和一个判别网络进行竞争博弈的过程。不同于传统的 GAN,文中的 GAN 模型采用了一个生成网络 G 和两个判别网络 D_c 和 D_b。

其中,生成器 G 将带有噪声块的图像变成航拍车辆图像,D_c 用来判别车辆的真假,D_c 用来判断背景的真假。通过生成器以及两个判别器的对抗学习,最终生成足够真实的航拍车辆图像。

文中提出的基于生成对抗网络的航拍车辆生成方法,与 PS-GAN 一样,在 pix2pix 的基础上,通过增加判别器以及对于不同目标任务采用不同的损失函数,以达到对生成器 G 的约束。网络结构如图 1 所示。

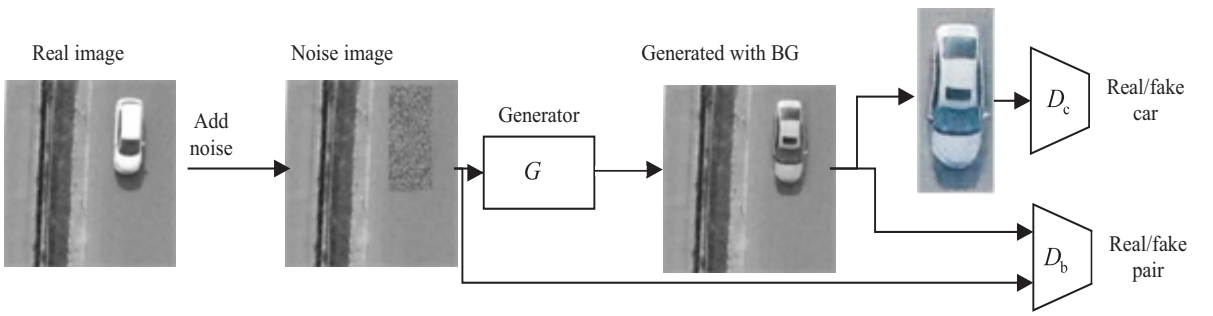


图 1　多判别器约束的 pix2pixGAN

2.1　生成模型

生成器 G 的目标是学习一种 x 到 y 的映射,使得 $G(x)=y$。其中 x 是输入的噪声图像,y 是目标图像。文中的生成网络 G 采用 U-Net[13] 作为网络结构,遵循编码解码结构。编码解码结构将输入图像 x 通过卷积层(如下采样层)来减少信息量,然后将编码得到的信息输入到解码器中通过上采样的方式进行解码,完成图像到图像之间的转换,如图 2 所示。但是传统的编码解码器会在数据编码的过程中出现信息丢失的情况,使得映射不稳定。U-Net 生成器结构在编码解码过程中使用跳步连接将下采样和上采样的镜像层连接,复制编码层的特征图给对应的解码层,保留了更多的原图信息,解决了信息传递丢失的问题。U-Net 结构如图 2 所示。

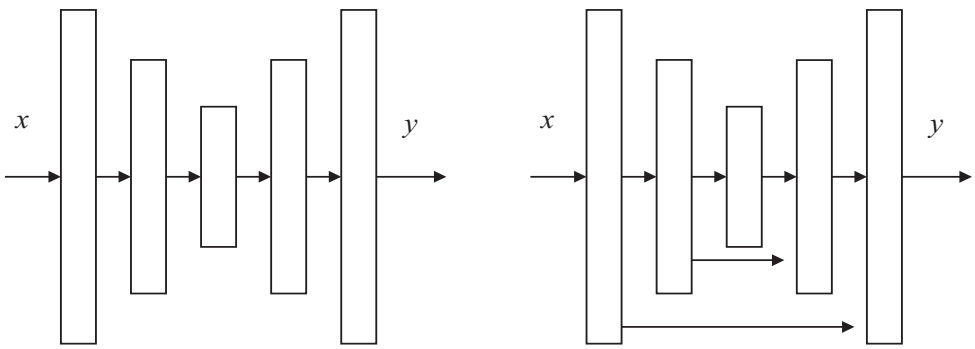


图 2　编码解码器和 U-Net

2.2　判别模型

对于车辆判别模型 D_c 来说,关注点在于噪声区域。根据车辆的位置标注信息,将生成图像 $G(x)$ 中相应位置的生成车辆 $G(z)$ 裁剪出来作为负样本。同样的,从原始图像 y 中将对应的真实车辆 y_c 裁剪出来作为正样本。判别器 D_c 的作用是判别输入的车辆

$G(z)$ 或 y_c 是真还是假。这将迫使生成器 G 学习到一种将噪声 z 转换到真实车辆 y_c 的映射，$G(z) \rightarrow y_c$，其中 z 是位于噪声图像 x 中的噪声区域。

判别器 D_c 的结构如图 3 所示。使用 5 层卷积网络进行特征提取。由于 D_c 的输入是从生成图像或者是原始图像中裁剪出来的不同大小的车辆，因此通过卷积网络得到的特征图大小也不一样。为解决这个问题，使用空间金字塔池化层[14]（spatial pyramid pooling，SPP）来得到固定长度的池化特征。这里，使用 3 层空间金字塔（1 * 1，2 * 2，4 * 4），最终会得到 21 个特征点，然后进行损失函数的计算。

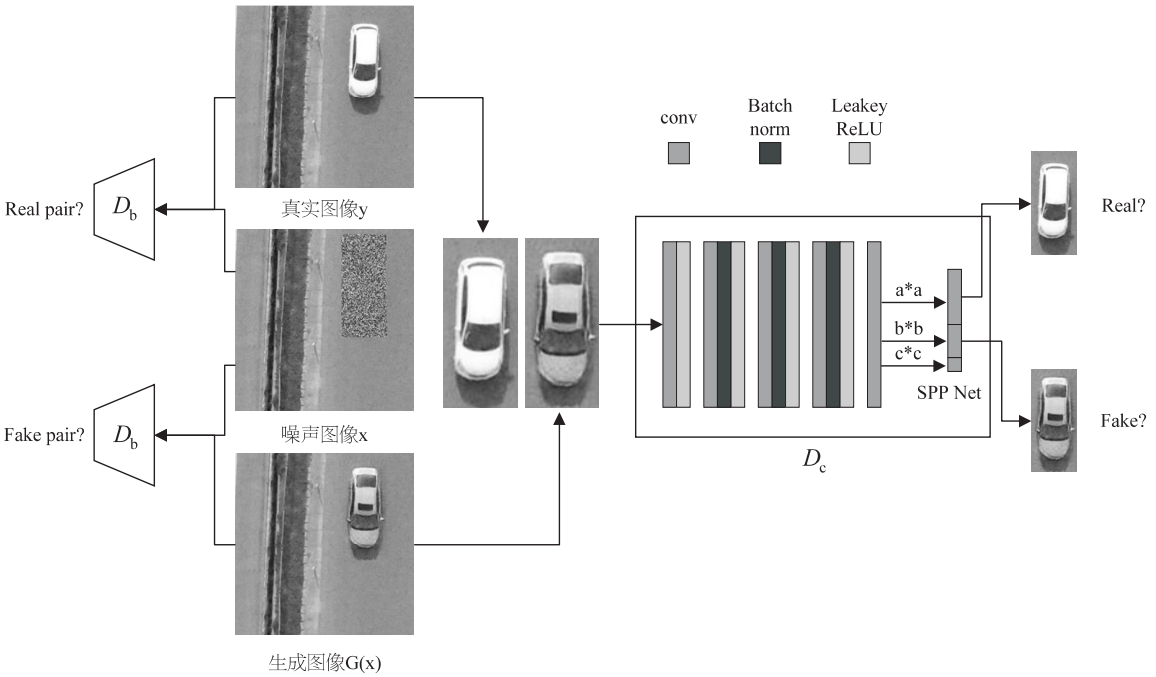


图 3 双判别器网络

对于背景判别模型 D_b 来说，关注点在于全局信息。目标不仅是生成逼真的车辆，更重要的是确保生产车辆能融入到背景中去。像 pix2pix 中的图像对一样，D_b 用来判别真实的图像对以及生成的图像对。真实图像对由原始图像 y 和噪声图像 x 组成，生产图像对由生成图像 $G(x)$ 和噪声图像 x 组成。

2.3 目标函数

如图 1 所示，该网络主要由一个生成网络和两个对抗网络组成。

对于 D_b 来说，使用最小二乘损失函数[15]来进行训练：

$$L_{\text{lsgan}}(D_b, G) = E[(D_b(y) - 1)^2] + E[(D_b(G(x)))^2] \quad (1)$$

对于 D_c 来说，使用交叉熵损失函数进行训练：

$$L_{\text{gan}}(D_c, G) = E[\log D_c(y_c)] + E[\log(1 - D_c(G(z)))] \quad (2)$$

对于 G 来说，为了使得生成的图片更逼近于真实图像，在训练过程中增加一个 L_1 正则化约束，具体如下：

$$L_{l1}(G) = E(\| y - G(x) \|_1) \quad (3)$$

所以，生成器 G 的最终的损失函数形式为：

$$L(G, D_b, D_c) = L_{\text{lsgan}}(D_b, G) + L_{\text{gan}}(D_c, G) + \lambda L_{l1}(G) \quad (4)$$

其中，y 表示原始图像，x 表示噪声图像，y_c 表示原始图像中的真实车辆，z 表示噪声图像中的噪声区域，$G(x)$ 表示生成图像，$G(z)$ 表示生成图像中的生成车辆，λ 控制着生成器的 L_1 损失的权重

3 实验结果与分析

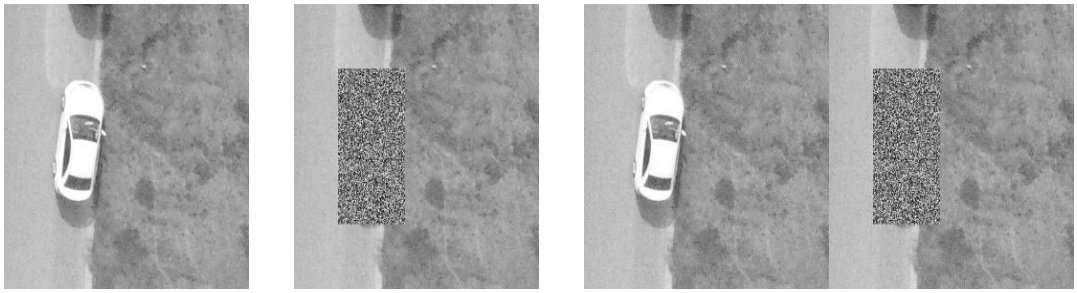
训练数据如图 4 所示，其中（a）是真实图片，（b）是噪声图片，（c）是将（a）、（b）在第二维度上拼接得到的图像对。将标注数据保存为 json 文件，并和训练图像一一对应，标注数据格式如下：{"y_1": 58, "x_1": 46, "y_2": 116, "x_2": 198}。其中 y_1，x_1 是标注车辆包围框的左上角，y_2，x_2 是其包围框的右下角。

在装备的数据集上使用多个模型分别训练了 200 代，在测试集上的部分结果如图 5 所示。其中（a）是原图，（b）是噪声图像，（c）是 pix2pixGAN 得到的图像，（d）是两个判别器 D_b 和 D_c 都使用最小二乘损失函数得到的模型，（e）是文中方法得到的图像，D_b 使用最小二乘损失函数，D_c 使用交叉熵损失。

通过比较可以发现，pix2pixGAN 只能勉强生成车辆的部分形状，内容十分模糊，而（d）中的车辆已经有了较好的外形，效果最好的是文中方法。对于（e）和

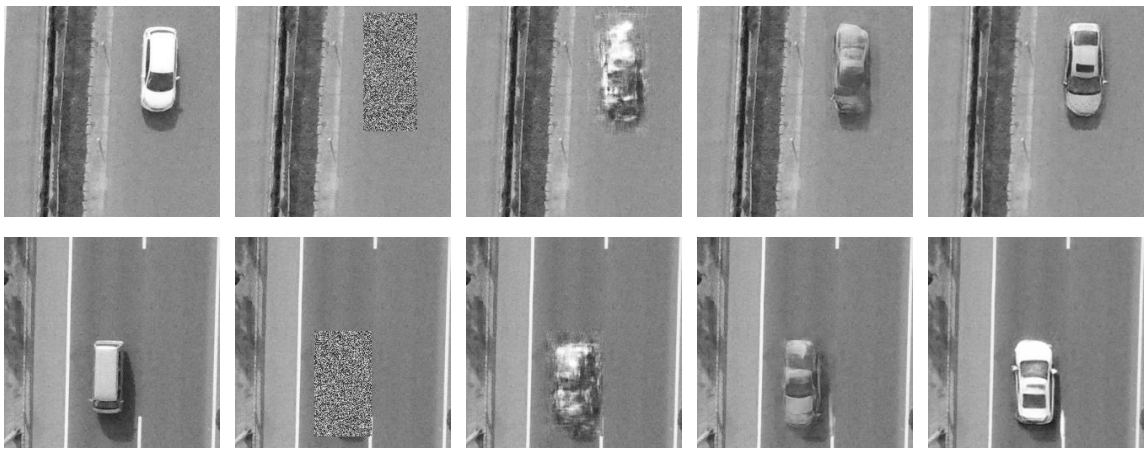
(d)两种模型得到的结果,可以推测是因为最小二乘损失函数相对于交叉熵损失能够获得更大的损失值,使得 D_c 在训练集上产生了过拟合,因此生成器 G 在测试集上面对新的背景以及噪声时,没能产生较高质量的航拍车辆。

图 6 是在真实航拍图像中生存车辆的结果,左图是原始的航拍图像,右图是生成图像。通过观察对比可以发现,原始图像中的车辆比较少,通过车辆生成方法,在其中的多处空白背景处生成了车辆,相比于真实车辆,生成车辆也比较真实。



(a)真实车辆图片　(b)噪声图片　(c)训练图像对

图 4　车辆生成数据准备



(a)原图　(b)噪声图像　(c)pix2pix　(d)ls loss　(e)文中方法

图 5　不同方法生成车辆对比

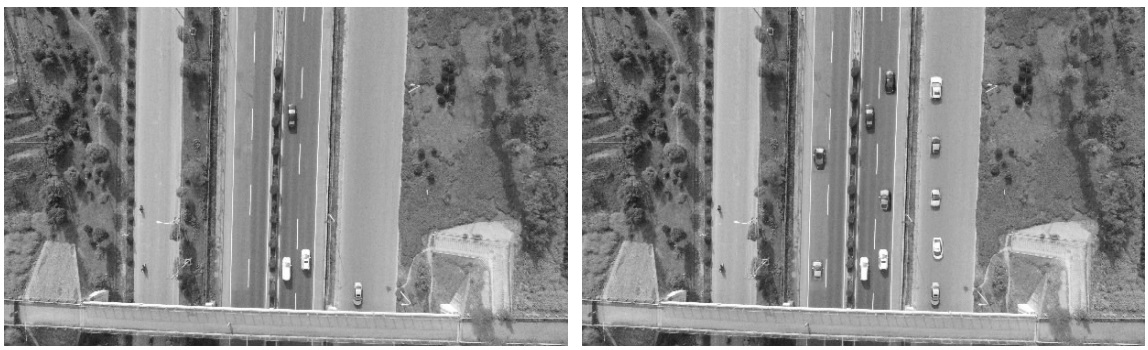


图 6　完整航拍图像中的车辆生成

好的生成结果。

4　结束语

为了解决航拍车辆检测模型训练需要大量标注样本的问题,提出了一种基于生成对抗网络在真实航拍图像中生成车辆的方法。在基于 pix2pixGAN 的方法上,使用两个判别器分别约束背景的拟合和车辆生成,通过多条件约束将噪声区域转化为航拍车辆。通过实验分析对比不同的方法,结果表明该生成方法具有较

参考文献:

[1] 周　敏. “无人机+交通”大势所趋[J]. 中国公路,2017(23):28-31.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on

computer vision and pattern recognition. San Francisco: IEEE,2014:580-587.
[3] REN Shaoqing,HE Kaiming,GIRSHICK R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(6):1137-1149.
[4] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:unified,real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. San Francisco:IEEE,2016:779-788.
[5] LIU Wei,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//European conference on computer vision. Amsterdam,The Netherlands:Springer,2016:21-37.
[6] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al. Generative adversarial nets[C]//Proceedings of the 27th international conference on neural information processing systems. Montreal,Canada:MIT Press,2014:2672-2680.
[7] 曹仰杰,贾丽丽,陈永霞,等. 生成式对抗网络及其计算机视觉应用研究综述[J]. 中国图象图形学报,2018,23(10):1433-1449.
[8] RADFORD A,METZ L,CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[C]//International conference on learning representations. [s. l.]:[s. n.],2015:67-68.
[9] ISOLA P,ZHU Junyan,ZHOU Tinghui,et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. San Francisco:IEEE,2017:5967-5976.
[10] 陈文兵,管正雄,陈允杰. 基于条件生成式对抗网络的数据增强方法[J]. 计算机应用,2018,38(11): 3305-3311.
[11] 曹志义,牛少彰,张继威. 基于半监督学习生成对抗网络的人脸还原算法研究[J]. 电子与信息学报,2018,40(2): 323-330.
[12] PATHAK D,KRAHENBUHL P,DONAHUE J,et al. Context encoders:feature learning by inpainting[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas,NV,USA:IEEE,2016:2536-2544.
[13] RONNEBERGER O,FISCHER P,BROX T. U-net:convolutional networks for biomedical image segmentation[C]//International conference on medical image computing and computer-assisted intervention. Munich,Germany:Springer,2015:234-241.
[14] HE Kaiming,ZHANG Xiangyu,REN Shaoqing,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//European conference on computer vision. Zurich,Switzerland:Springer,2014:346-361.
[15] MAO Xudong,LI Qing,XIE Haoran,et al. Least squares generative adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. San Francisco:IEEE,2017:2794-2802.

(上接第 161 页)

adaptation:a survey of recent advances[J]. IEEE Signal Processing Magazine,2015,32(3):53-69.
[7] LONG Mingsheng,CAO Yue,WANG Jianmin,et al. Learning transferable features with deep adaptation networks[C]//Proceedings of the 32nd international conference on machine learning. Lille:JMLR,2015:97-105.
[8] LONG Mingsheng,WANG Jianmin,JORDAN M. Deep transfer learning with joint adaptation networks[C]//International conference on machine learning. [s. l.]:[s. n.],2017:2208-2217.
[9] GANIN Y,USTINOVA E,AJAKAN H,et al. Domain-adversarial training of neural networks[M]//Domain adaptation in computer vision applications. [s. l.]:Springer,2016:189-209.
[10] SIMONYAN K,ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International conference on learning representations. [s. l.]:[s. n.],2015:1150-1210.
[11] ZEILER M D,FERGUS R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Zurich,Switzerland:Springer,2014:818-833.
[12] CHEN Yihua,LI Wen,SAKARIDIS C,et al. Domain adaptive faster R-CNN for object detection in the wild[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]:IEEE,2018:3339-3348.
[13] GANIN Y,LEMPITSKY V. Unsupervised domain adaptation by back propagation[C]//International conference on machine learning. [s. l.]:[s. n.],2015:1180-1189.
[14] CORDTS M,OMRAN M,RAMOS S,et al. The cityscapes dataset for semantic urban scene understanding[C]//2016 IEEE conference on computer vision and pattern recognition. [s. l.]:IEEE,2016:3213-3223.
[15] GEIGER A,LENZ P,STILLER C,et al. Vision meets robotics:the KITTI dataset[J]. International Journal of Robotics Research,2013,32(11):1231-1237.