计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(1):194−199 [doi: 10.15888/j.cnki.csa.007756]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于图像风格迁移的端到端跨域目标检测①

吴泽远, 朱　明

(中国科学技术大学 信息科学与技术学院, 合肥 230026)
通讯作者: 吴泽远, E-mail: wzyt@mail.ustc.edu.cn

摘　要: 跨域目标检测是最近兴起的研究方向, 旨在解决训练集到测试集的泛化问题. 在已有的方法中利用图像风格转换并在转换后的数据集上训练模型是一个有效的方法, 然而这一方法存在不能端到端训练的问题, 效率低, 流程繁琐. 为此, 我们提出一种新的基于图像风格迁移的跨域目标检测算法, 可以把图像风格迁移和目标检测结合在一起, 进行端到端训练, 大大简化训练流程, 在几个常见数据集上的结果证明了该模型的有效性.

关键词: 跨域; 目标检测; 风格迁移; 端到端

引用格式: 吴泽远,朱明.基于图像风格迁移的端到端跨域目标检测.计算机系统应用,2021,30(1):194–199. http://www.c-s-a.org.cn/1003-3254/7756.html

End-to-End Cross-Domain Object Detection Based on Image Style Transfer

WU Ze-Yuan, ZHU Ming

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: Cross-domain object detection is a new research direction, which aims to solve the problem of generalization from training set to test set. In the existing methods, using image style transfer and train the model on the converted data set is an effective method. However, this method has the problems of not end-to-end training, low efficiency, and tedious process. Therefore, we propose a new cross domain target detection algorithm based on image style migration, which can combine image style migration and target detection to carry out end-to-end training, and greatly simplify the training process. The results on several common datasets show the validity of the model.

Key words: cross-domain; object detection; style transfer; end-to-end

目标检测是计算机视觉领域最基础也最重要的问题, 在现实生活中具有广泛的应用, 如自动驾驶[1], 视频监控[2], 人脸识别[3], 所谓目标检测, 就是在图像中检测到物体的位置和类别. 随着深度学习的蓬勃发展, 基于卷积神经网络 (CNN) 的方法[4–9] 在目标检测标准数据集[10–12] 上取得了明显进步. 尽管效果显著, 却只是在标注好的干净的数据集上进行实验, 而真实世界中目标检测要面临的情况要更加复杂, 数据的视角, 外观, 背景, 光照等方面的差异导致训练好的模型难以落地使用, 即训练数据称为 (源域) 和测试数据 (称为目标域) 分布不同, 导致训练好的模型泛化性较差, 这一问题称为域偏移问题.

为了处理这个问题, 一个具有吸引力的办法是无监督域适应[13], 即将在源域上训练的模型用于目标域, 却不需要对目标域进行数据标注. 最近, 基于图像风格迁移的域适应方法[14,15] 取得了不错的结果, 这一类方法基于图像风格迁移[16], 基本思想是利用图像风格迁移技术将有标签的源域数据转换为目标域风格, 然后在转换后的源域图像上进行训练, 这样便相当于在有标签的目标域数据集上训练模型, 最后, 将训练好的模

① 基金项目: 安徽省重点研发计划 (201904a05020035)
Foundation item: Key Research and Development Program of Anhui Province (201904a05020035)
收稿时间: 2020-06-06; 修改时间: 2020-07-07; 采用时间: 2020-07-10; csa 在线出版时间: 2020-12-31

型用于目标域即可.

然而, 上述基于图像风格迁移的方法有几点缺陷: 1) 这些方法通常将先进行图像风格迁移, 然后再对迁移后的图像进行训练, 流程繁琐. 2) 图像风格迁移网络和检测网络分开训练, 训练速度慢, 同时两个网络无法共享数据, 不能充分利用数据.

为了处理上述问题, 本文设计了基于图像风格迁移的端到端跨域检测网络. 在该网络中我们设计了两个模块, 包括图像风格迁移模块, 和目标检测模块. 其中图像风格迁移模块采用比较流行的风格迁移方法, 对源域图像进行风格转换. 目标检测模块采用通用的 Faster R-CNN[4] 网络对转换后的图像进行检测.

本文设计的网络在 4 个标准数据集上取得了相当甚至超过最佳方法的结果.

1 相关工作

1.1 目标检测

目标检测网络通常分为两阶段方法[4–6] 和单阶段方法[7–9]. 两阶段方法在第 1 阶段通常采用区域候选网络[4](RPN) 和区域池化模块[4](ROI Pooling) 提取目标特征, 然后在第 2 阶段采用全连接层对提取的目标特征进行分类和位置回归. 单阶段网络省去了第 1 阶段, 直接回归输出目标的类别和位置, 网络流程的简化加快了检测速度, 当然, 在追求速度的同时也牺牲了精度. 然而, 这些方法通常用于常规检测, 即训练集和测试集来自于同一数据域, 无法处理域偏移问题. 在本文中, 我们选取 Faster R-CNN 作为基础检测器, 并且结合图像风格迁移技术改善其泛化性能.

1.2 跨域目标检测

DA Faster[17] 是跨域目标检测的开山工作, 该方法在图像级别和目标级别分别设计判别器达到域适应的目的. DTPL[14] 采用图像风格迁移技术, 通过 CycleGAN[18] 将源域图像转换为目标域风格的图像, 并在转换后的带标签图像上训练检测器. DM[15] 对 CycleGAN 损失函数进行修改, 生成多个中间域图像, 并训练多域判别器. 在 MAF[19] 中, 通过在图像级特征层面设置多层判别器对检测器达到域适应. FAFR-CNN[20] 将少样本与跨域目标检测相结合, 专注于少样本情况下的域迁移. 与上述方法相比, 本文方法的不同之处在于将图像风格迁移和目标检测放到一个端到端的框架中.

1.3 生成对抗网络

生成对抗网络 (GAN)[21] 由 GoodFellow 等人于 2014 年发明, 并引发了无数的后续相关研究. 主要的变体有条件生成对抗网络[22](cGAN), DCGAN[23], WGAN[24] 等. DCGAN 用卷积神经网络代替原始 GAN 中的全连接层, 能够生成更清晰的图像. cGAN 在网络的输入中加入条件, 从而可以控制输出的类别. WGAN 修改了原始 GAN 的目标函数, 使得训练更加稳定.

1.4 图像风格迁移

基于深度学习的图像风格迁移算法可按照数据输入分为两类, 一类是需要成对数据的算法, 以 Pix2Pix[25] 为代表, 这一系列还有 Pix2PixHD[26], Vid2Vid[27]. 其中, Pix2Pix 利用成对数据作为输入, 即一一对应的源域风格数据和目标域风格数据, 利用生成器将源域数据进行转换, 然后利用判别器对真实目标域数据和虚假数据进行判别. Pix2PixHD 是 Pix2Pix 的高清版本, 采用多层金字塔以生成高清图像. 而 Vid2Vid 是 Pix2Pix 的视频版本, 用以生成视频. 另一类是无需成对数据的算法, 以 CycleGAN 为代表, 这一系列还有 StarGAN[28]. CycleGAN 无需成对数据, 输入源域图像和目标域图像, 利用生成器 GA 将源域图像转换为目标域风格, 并利用判别器 DA 进行判别, 同时将生成的虚假图像利用生成器 GB 转换为原来风格, 并约束图像尽可能还原. 对于目标域图像, 同样存在上述循环. StarGAN 相比于 CycleGAN, 可以解决多个域之间的风格转换, 主要改进在于判别器不光进行真假判断, 同时也对数据所属的域进行分类.

2 基于图像风格迁移的端到端网络

如图 1 所示, 我们的网络主要由两个模块组成, 一个是图像风格迁移模块, 另一个是目标检测模块. 关于检测模块我们采用常见的 Faster R-CNN 网络, 下面重点阐述图像风格迁移模块的设计.

2.1 图像风格迁移模块

图像风格迁移模块部分共有 4 个组成部分, 分别是源域到目标域的生成器 G-A2B, 目标域判别器 DB, 以及对称的目标域到源域的生成器 G-B2A, 源域图像判别器 DA. 我们用 A 指代源域图像, B 指代目标域图像, fake_B 代表生成的虚假目标域图像, fake_A 代表生成的虚假源域图像, 其中, G-A2B 负责将 A 转换为 fake_B, 并由判别器 DB 对真实的图像 B 与 fake_B 进

行判别, 同时为了保证 fake_B 在图像内容上仍保留原图像 A 的内容, 对 fake_B 再次转换, 并对转换后的 fake_A 与 A 进行内容一致性约束.

上面是 A 到 B 的一次循环, 与之对称, B 到 A 也有相同的生成器转换与判别器判别流程, 这样的循环生成网络可以保证生成器在保证内容不变的条件下对图像风格进行改变, 值得注意的是, 由于空间有限, 在图 1 中并没有画出 B 到 A 的对称流程.

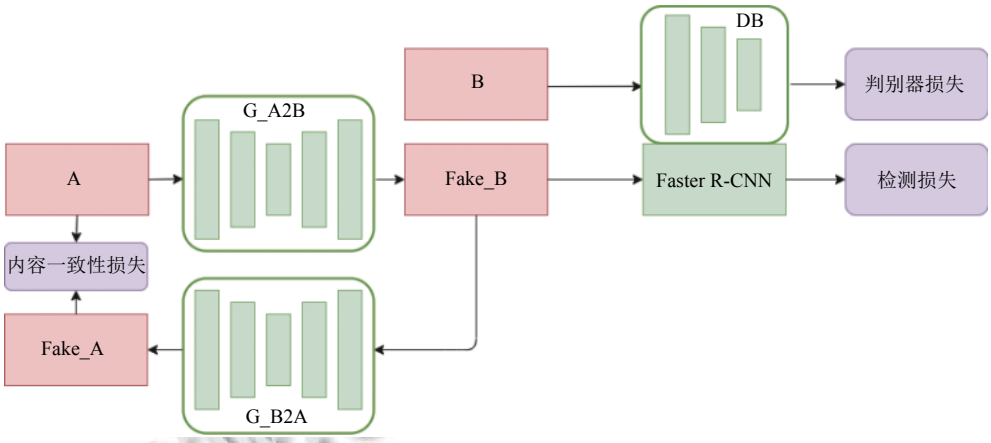

图 1 基于图像风格迁移的跨域目标检测框架

两个生成器具有相同的网络结构, 采用编码器-解码器网络, 即对输入图像先压缩缩小分辨率, 再上采样还原分辨率. 具体地, 编码器网络采用 3 层卷积, 每层卷积后有相应的池化层, 图像分辨率缩小为 1/8, 解码器部分, 采用转置卷积进行上采样, 同样有 3 层转置卷积, 将图像还原为输入分辨率. 所有卷积层采用 3x3 卷积核, 池化层步长为 2, 采用 max pooling 池化. 判别器部分, 采用 3 层全连接网络, 输出为 2 分类输出, 损失函数采用二值交叉熵损失. 另外, 为了保证转换后的图像仍保留原图像的内容, 对转换后的图像输入生成器进行还原, 还原后的图像与原图像间采用 L1 损失进行内容一致性约束. 所以, 在图像风格迁移模块, 存在 4 个损失, 分别是两个判别器的分类损失, 以及两个内容一致性损失, 损失函数如下:

$$L_{DA} = \log(1 - DA(fake\_A)) + \log(DA(A)) \quad (1)$$

$$L_{DB} = \log(1 - DB(fake\_B)) + \log(DB(B)) \quad (2)$$

$$L_{CA} = |A - fake\_A| \quad (3)$$

$$L_{CB} = |B - fake\_B| \quad (4)$$

2.2 目标检测模块

对于转换后的源域图像, 利用已有的检测器进行训练, 由于源域图像有标签, 这样便相当于在有标签的目标域图像上进行训练. 目标检测模块采用常见的 Faster R-CNN 检测器, 检测器采用两阶段设计, 第一阶段采用 RPN 网络提取候选区域, 第二阶段提取候选区域特征并进行分类和位置回归. 网络基础特征提取 (backbone) 部分采用 ResNet[29]. 网络的损失函数定义如下:

$$L_{\text{det}} = J(x, y) \quad (5)$$

其中, J 代表 RPN 网络的分类和回归损失, 以及检测器最后分类器和回归器的损失, x 和 y 代表输入和标签.

2.3 总目标函数

网络的总目标函数即包括图像风格迁移部分的损失和检测器损失:

$$L_{\text{total}} = \alpha(L_{DA} + L_{DB} + L_{CA} + L_{CB}) + L_{\text{det}} \quad (6)$$

其中, α 是判别器损失和内容一致性损失的权重因子. 在实验部分我们会详细说明网络的训练细节.

3 实验分析

在这一部分, 我们阐述使用的数据集, 评估场景, 基线方法和实验细节, 然后给出实验结果分析.

3.1 数据集

我们利用下列几个数据集构建域适应场景并执行了实验. (1) PASCAL VOC[11]. 这个数据集包含 20 类常见物体, 我们将 PASCAL VOC 2007 和 2012 的训练集和验证集用于训练, 作为源域, 总共 15 000 张图片.

(2) WaterColor[30]. WaterColor 包含 2K 张水彩风格图像, 6 类物体, 属于 PASCAL VOC 的 20 类物体的子集. 1K 张用于训练, 1K 张用于测试. (3) Sim10K[31]. 这个数据集由合成的驾驶场景图像组成, 共 10K 张图像, 我们将其作为目标域, 并只对汽车目标进行检测. (4) CityScape & FoggyCityScape[32,33]. CityScape 中的图像由车载相机捕捉而来, FoggyCityScape 是利用 CityScape 添加雾噪声得到. 两个数据集规模相同, 包含 2975 张训练集, 500 张测试集.

3.2 评估场景

本文共建立 3 个域适应场景, 包括: (1) 场景 1. PASCAL VOC 到 WaterColor, 用于捕捉真实数据到艺术风格数据的偏移. (2) 场景 2. CityScape 到 FoggyCityScape, 用于捕捉正常天气到雾天的偏移. (3) 场景 3. Sim10K 到 CityScape, 用于捕捉合成图像到真实数据的偏移.

3.3 基线方法

我们的方法以两阶段检测器 Faster R-CNN 为基础, 同时也与一些目前的跨域检测方法包括 DA Faster[17], DTPL[14], DM[15], MAF[19], FAFR-CNN[20], ST[34], Strong-Weak[35], SCDA[36] 进行比较, 这些方法的结果引用于文献 [35,36].

3.4 实现细节

在实验中, 我们设置 batch size 为 1, 初始学习率为 0.001, 每 5 个周期乘以 0.1, 共训练 20 个周期. 优化器采用随机梯度下降 (SGD), 动量设置为 0.9, 权重衰减设为 0.0001. 对于所有实验, 我们采用 PASCAL VOC 的阈值 0.5 作为评估标准. 关于超参数, 设置α=0.2. 所有场景基础网络部分都采用 ResNet.

3.5 结果

在这一部分, 我们展示在 3 个场景的实验结果并作出详细分析.

场景 1. 考察真实图像与艺术风格图像的域适应. 采用 PASCAL VOC 作为源域, WaterColor 作为目标域实验过程中, 将 ResNet 在 ImageNet[10] 上预训练, 然后作为基础特征网络. 候选区目标的数目是 128, 每个维度是 2304. 表 1 表明我们的方法超出所有方法至少 1.0 MAP, 表明该方法在真实图像到艺术风格图像这种域偏移较大的场景上表现良好, 尤其相比于 DTPL 和 DM 这两个同样采用风格迁移技术的方法, 我们的方法流程更简单, 效果更好. 同时可以看到, DA Faster 相比于 Faster R-CNN 提升有限, 而 DTPL 和 DM 等基于图像转换的方法效果突出, 表明对于域风格差异较大的场景, 图像转换可以起到良好作用.

表 1 PASCAL VOC 到 WaterColor 的域适应结果

Method	Bike	Bird	Car	Cat	Dog	Person	MAP
Faster RCNN	68.8	46.8	37.2	32.7	21.3	60.7	44.6
DA Faster	75.2	40.6	48.0	31.5	20.6	60.0	46.0
ST	75.6	45.8	49.3	34.1	30.3	64.1	49.9
DTPL	76.5	54.9	46.0	37.4	38.5	72.3	54.3
DM	–	–	–	–	–	–	52.0
Strong-Weak	82.3	55.9	46.5	32.7	35.5	66.7	53.3
Ours	95.3	51.9	49.1	39.0	48.9	62.3	55.3

场景 2. 这个场景中考察正常天气到雾天下的表现. 我们用 CityScape 作为源域, FoggyCityScape 作为目标域. 如表 2 所示, 我们提出的方法与表现最佳的 DM 相差无几. 这一结果表明我们的方法在两个域不相似和相似时都能表现良好. 在这一设置中, DA Faster 和 FAFR-CNN 相比于 Faster R-CNN 分别带来 7.3 和 11.0MAP. 相比于 Strong-Weak, 我们的方法在这一场景下表现不如场景 1, 这表明在域偏移较小时关注全局对齐就能有效缓解域偏移.

表 2 CityScape 到 FoggyCityScape 的域适应结果

Method	Bus	Bicycle	Car	Bike	person	Rider	Train	Truck	MAP
Faster RCNN	22.3	26.5	34.3	15.3	24.1	33.1	3.0	4.1	20.3
DA Faster	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
FAFR-CNN	29.1	39.7	42.9	20.8	37.4	24.1	26.5	29.9	31.3
SCDA	33.5	38.0	48.5	26.5	39.0	23.3	28.0	33.6	33.8
MAF	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
Strong-Weak	36.2	35.3	43.5	30.0	29.9	42.3	32.6	24.5	34.3
DM	30.8	40.5	44.3	27.2	38.4	34.5	28.4	32.2	34.6
Ours	44.5	34.6	43.9	24.5	30.6	40.2	34.2	26.4	34.2

场景 3. 这里, 我们评估我们的方法在合成图像到真实图像上的域适应效果. 我们采用 Sim10K 作为源域. 至于目标域, 我们采用 Cityscape. 两个域都是驾驶场景, 但是在光照, 视角上有明显不同. 相比于场景 2, 域偏移更大. 结果展示在表 3. 在这一场景中, 我们的方法相比于 Strong-Weak DA 有 1.4MAP 的提升, 与 FAFR-CNN 效果相当. 仅低于最好的方法 SCDA.

表 3 CityScape 到 SIM10K 的域适应结果

	Faster	DA Faster	FAFR-CNN	SCDA	Strong-Weak	Ours
Car	34.6	38.9	41.2	43.0	40.1	41.5

3.6 消融试验

这一部分, 我们执行一些消融试验以分析一些超参数和模型中不同模块的影响, 所有实验采用 PASCAL

VOC 到 WaterColor 这一场景.

α的影响结果展示在表 4 中, 我们采用场景 2 进行实验. 我们尝试了不同的参数设置, 发现最好的结果是 0.2.

表 4 权重因子α的影响

α	Bike	Bird	Car	Cat	Dog	Person	MAP
0.2	95.3	51.9	49.1	39.0	48.9	62.3	55.3
0.5	93.0	51.8	43.6	36.5	48.6	58.3	52.9
0.7	91.6	49.4	41.8	37.3	48.6	55.7	51.7
1.0	89.7	48.9	41.4	36.0	44.7	51.5	49.6

4 结论与展望

在本文中, 我们提出来一个新颖的方法, 可以将图像风格迁移与目标检测放在一个统一的框架中, 并进行端到端的训练, 简化了训练流程. 在标准数据集上的实验验证了所提出方法的有效性. 后面的工作中我们会把重点放在图像风格迁移模块的改进以及寻求风格迁移模块与检测模块如何更好地融合, 更好地提高在目标域上的泛化性能.

参考文献

1 Li PL, Chen XZ, Shen SJ. Stereo R-CNN based 3D object detection for autonomous driving. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 7636–7644.

2 Hattori H, Lee N, Boddeti VN, *et al*. Synthesizing a scene-specific pedestrian detector and pose estimator for static video surveillance: Can we learn pedestrian detectors and pose estimators without real data? International Journal of Computer Vision, 2018, 126(9): 1027–1044. [doi: 10.1007/s11263-018-1077-3]

3 Turk MA, Pentland AP. Face recognition using eigenfaces. Proceedings of 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Maui, HI, USA. 1991. 586–591.

4 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2015. 91–99.

5 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.

6 Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 6154–6162.

7 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 6517–6525.

8 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.

9 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 21–37.

10 Deng J, Dong W, Socher R, *et al*. ImageNet: A large-scale hierarchical image database. Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. 2009. 248–255.

11 Everingham M, Eslami SMA, Van Gool L, *et al*. The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 2015, 111(1): 98–136. [doi: 10.1007/s11263-014-0733-5]

12 Lin TY, Maire M, Belongie S, *et al*. Microsoft coco: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 740–755.

13 Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France. 2015. 1180–1189.

14 Inoue N, Furuta R, Yamasaki T, *et al*. Cross-domain weakly-supervised object detection through progressive domain adaptation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 5001–5009.

15 Kim T, Jeong M, Kim S, *et al*. Diversify and match: A domain adaptive representation learning paradigm for object detection. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach. CA, USA. 2019. 12448–12457.

16 Chen HY, Fang IS, Cheng CM, *et al*. Self-contained stylization via steganography for reverse and serial style transfer. Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass Village, FL, USA. 2020. 2152–2160.

17 Chen YH, Li W, Sakaridis C, *et al*. Domain adaptive faster R-CNN for object detection in the wild. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern

Recognition. Salt Lake City, UT, USA. 2018. 3339–3348.
18 Zhu JY, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 2242–2251.
19 He ZW, Zhang L. Multi-adversarial faster-RCNN for unrestricted object detection. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 6667–6676.
20 Wang T, Zhang XP, Yuan L, *et al*. Few-shot adaptive faster R-CNN. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 7166–7175.
21 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2014. 2672–2680.
22 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv: 1411.1784, 2014.
23 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: 1511.06434, 2015.
24 Gulrajani I, Ahmed F, Arjovsky M, *et al*. Improved training of wasserstein GANs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA. 2017. 5769–5779.
25 Isola P, Zhu JY, Zhou TH, *et al*. Image-to-image translation with conditional adversarial networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 5967–5976.
26 Wang TC, Liu MY, Zhu JY, *et al*. High-resolution image synthesis and semantic manipulation with conditional GANs. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 8798–8807.
27 Wang TC, Liu MY, Zhu JY, *et al*. Video-to-video synthesis. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2018. 1152–1164.
28 Choi Y, Choi M, Kim M, *et al*. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 8789–8797.
29 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
30 Torralba A, Murphy KP, Freeman WT, *et al* Context-based vision system for place and object recognition. Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France. 2003. 273–280.
31 Johnson-Roberson M, Barto C, Mehta R, *et al*. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? Proceedings of 2017 IEEE International Conference on Robotics and Automation. Singapore. 2017. 746–753.
32 Cordts M, Omran M, Ramos S, *et al*. The cityscapes dataset for semantic urban scene understanding. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 3213–3223.
33 Sakaridis C, Dai DX, Van Gool L. Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision, 2018, 126(9): 973–992. [doi: 10.1007/s11263-018-1072-8]
34 Kim S, Choi J, Kim T, *et al*. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 6091–6100.
35 Saito K, Ushiku Y, Harada T, *et al*. Strong-weak distribution alignment for adaptive object detection. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 6956–6965.
36 Zhu XG, Pang JM, Yang CY, *et al*. Adapting object detectors via selective cross-domain alignment. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. 2019. 687–696.