

# 控制与决策

Control and Decision

## 基于生成对抗网络学习被遮挡特征的目标检测方法

安珊, 林树宽, 乔建忠, 李川皓

引用本文:

安珊, 林树宽, 乔建忠, 等. 基于生成对抗网络学习被遮挡特征的目标检测方法[J]. *控制与决策*, 2021, 36(5): 1199–1205.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1319>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### [基于条件对抗生成孪生网络的目标跟踪](#)

Conditional generative adversarial siamese networks for object tracking

*控制与决策*. 2021, 36(5): 1110–1118 <https://doi.org/10.13195/j.kzyjc.2019.1215>

### [基于卷积神经网络的云雾遮挡舰船目标识别](#)

Obscured ship target recognition based on convolutional neural network

*控制与决策*. 2021, 36(3): 661–668 <https://doi.org/10.13195/j.kzyjc.2019.0781>

### [抗遮挡与尺度自适应的改进KCF跟踪算法](#)

Improved KCF tracking algorithm based on anti-occlusion and scale transformation

*控制与决策*. 2021, 36(2): 457–462 <https://doi.org/10.13195/j.kzyjc.2019.0394>

### [尺度自适应的多特征融合相关滤波目标跟踪算法](#)

Scale adaptation and multi-feature fusion correlation filtering object tracking algorithm

*控制与决策*. 2021, 36(2): 429–435 <https://doi.org/10.13195/j.kzyjc.2019.0445>

### [复杂背景下全景视频运动小目标检测算法](#)

Panoramic video motion small target detection algorithm in complex background

*控制与决策*. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

# 基于生成对抗网络学习被遮挡特征的目标检测方法

安 珊, 林树宽<sup>†</sup>, 乔建忠, 李川皓

(东北大学 计算机科学与工程学院, 沈阳 110169)

**摘 要:** 实际生活中目标间存在的遮挡会造成待检测目标的特征缺失, 进而使得检测准确度降低. 鉴于此, 提出一种用于被遮挡特征学习的生成对抗网络 (generative adversarial networks for learning occluded features, GANLOF). 被遮挡特征学习网络分为被遮挡特征生成器、鉴别器两个部分. 首先对数据集生成随机遮挡, 作为模型的输入; 然后利用生成器恢复被遮挡图片的池化特征, 通过鉴别器区分恢复后的被遮挡池化特征与无遮挡图片池化特征, 同时使用检测损失监督生成器, 使恢复的被遮挡特征更准确. 所提出被遮挡特征学习网络可以作为组件插入到任意的两阶段检测网络中. 与 Faster RCNN 等已有模型相比, 所提出模型在 PASCAL VOC2007 和 KITTI 数据集上的 mAP (mean average precision) 指标均有不同程度的提升.

**关键词:** 目标检测; 目标遮挡; 特征缺失; 恢复被遮挡特征; 生成对抗网络; 被遮挡特征生成器

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1319

开放科学(资源服务)标识码(OSID):



**引用格式:** 安珊, 林树宽, 乔建忠, 等. 基于生成对抗网络学习被遮挡特征的目标检测方法[J]. 控制与决策, 2021, 36(5): 1199-1205.

## Object detection via learning occluded features based on generative adversarial networks

AN Shan, LIN Shu-kuan<sup>†</sup>, QIAO Jian-zhong, LI Chuan-hao

(College of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

**Abstract:** Object detection is a fundamental task in computer vision. There often exist occlusions between objects in real life, which result in that some features of an object are missing, and detection accuracy is reduced. Therefore, we propose a generative adversarial network for learning occluded features (GANLOF). It is divided into two parts: the generator of occluded features and the discriminator. Firstly, we generate random occlusions for pictures in datasets, and the occluded pictures are the inputs of the network. Then we use the generator to restore pooling features in occluded regions, and the occluded pooling features and the un-occluded image pooling features are distinguished by the discriminator. Meanwhile, we use the detection loss to supervise the generator, so that the recovered occluded features are more accurate. The proposed GANLOF can be used as a component added into two-phase object detection networks. Compared with the Faster RCNN and other models, the mean average precision (mAP) of model is improved on the PASCAL VOC2007 dataset and the KITTI dataset.

**Keywords:** object detection; object occlusion; feature missing; occluded feature recovery; generative adversarial net; occluded feature generator

## 0 引言

目标检测是计算机视觉中的一个重要研究内容, 其主要任务是定位图像中目标的位置并判断目标类别, 广泛应用于无人驾驶、障碍物检测等领域. 目前的主要做法是利用卷积神经网络 (convolutional neural networks, CNN)<sup>[1]</sup> 提取图像特征, 并通过建立检测模型对目标进行分类和定位. 在实际应用中, 图像中的目标经常存在被遮挡、形变、模糊等现象, 使得提取到

的目标特征不明显或者缺少部分特征信息, 进而导致分类错误. 目前, 对于目标检测任务而言, 充分提取被遮挡目标的特征并利用目标无遮挡部分的特征对目标进行分类回归仍然是一个挑战.

本文基于生成对抗思想提出一种用于被遮挡特征学习的生成对抗网络 GANLOF, 通过被遮挡特征生成器 (occluded feature generator) 恢复被遮挡目标的特征, 通过鉴别器 (discriminator) 监督被遮挡

收稿日期: 2019-09-19; 修回日期: 2019-12-18.

基金项目: 国家自然科学基金项目 (61272177).

<sup>†</sup>通讯作者. E-mail: linshukuan@ise.neu.edu.cn.

特征学习的准确性. 本文的目标检测模型主要分为RoI (region of interest)池化特征提取、被遮挡特征生成器、鉴别器、检测器4个部分. 被遮挡特征生成器用于学习被遮挡目标缺失特征, 鉴别器用于判断生成器学到的特征与无遮挡目标特征的相似程度, 从而使得被遮挡特征学习更准确. 通过随机遮挡图片中目标的部分区域, 生成遮挡图片, 通过VGG16<sup>[2]</sup>提取图片特征, 对感兴趣区域特征进行池化, RoI池化特征作为训练生成器的样本, 而数据集中原始图片的RoI池化特征作为训练生成器的真实标签.

本文检测器采用Faster RCNN<sup>[3]</sup>模型的检测网络. 为了验证所提出模型的有效性, 将其与不进行被遮挡特征学习的Faster RCNN模型以及其他先进方法进行对比, 主要贡献为: 1) 提出一个恢复被遮挡特征的生成对抗网络GANLOF, 可以作为一个组件加入到任何两阶段目标检测模型中; 2) 训练过程中, 融合了特征学习损失 $L_{mse}$ 、对抗损失 $L_{adv}$ 、检测损失 $L_d$ 三部分损失函数, 对模型进行端到端训练, 生成的随机遮挡图片相当于困难样本, 增加了训练数据的多样性, 从而增强了模型的鲁棒性; 3) 与Faster RCNN模型相比, 所提出模型在PASCAL VOC2007(下文简称VOC2007)数据集<sup>[4]</sup>和KITTI数据集<sup>[5]</sup>上的mAP分别取得了0.2%和0.46%的提升, 同时也优于其他先进的模型.

## 1 相关工作

### 1.1 目标检测

近期目标检测方法主要分为基于候选区域的两阶段方法和基于回归的一阶段方法. 两阶段方法将目标检测分为提取候选区域和对候选区域进行分类与回归两个阶段. 一阶段方法将目标检测看作回归问题, 直接在图像的多个位置进行目标位置的回归. 典型的两阶段方法包括R-CNN<sup>[6]</sup>、SPP-Net<sup>[7]</sup>、Fast-RCNN<sup>[8]</sup>和Faster-RCNN<sup>[3]</sup>等. 文献[6]提出一种基于区域的卷积神经网络结构R-CNN, 通过选择性搜索算法(selective search)<sup>[9]</sup>生成候选区域, 利用CNN网络对归一化后的候选区域进行特征提取, 最后使用SVM进行分类并利用线性回归对位置进行微调. 文献[7]利用CNN在整张图像上提取特征, 提升了计算速度. 文献[8]提出快速区域卷积神经网络Fast RCNN, 并提出共享特征提取网络和RoI池化, 提升了检测速度和精度. 文献[3]是对文献[8]所提方法的进一步改进, 提出RPN (region proposal network)网络结构提取候选区域, 实现端到端的训练. 文献[10]提出目标关系图卷积网络ORGCN (object relation

graph convolution networks), 基于自适应注意力机制, 动态构建关系图聚合目标及其相邻目标间的信息. 典型的一阶段方法包括YOLO (you only look once)<sup>[11]</sup>、SSD (single shot detector)<sup>[12]</sup>等. 文献[11]将目标检测任务看作回归问题, 直接在图像的多个位置对目标位置进行回归. 文献[12]提出SSD模型, 借鉴两阶段方法Faster RCNN中的锚框(Anchor)机制实现多尺度目标检测, 相较于YOLO模型检测精度有了很大提升.

### 1.2 遮挡处理

文献[13]针对分类和弱监督目标定位提出在训练过程中对图像进行随机遮挡的策略, 该策略分别在AlexNet<sup>[14]</sup>和GoogleNet<sup>[15]</sup>网络上对图像分类进行实验, 分类精度均有所提升. 文献[16]提出一种基于区域分解与合成的目标检测方法R-DAD (region decomposition and assembly), 该方法通过分割候选区域建议, 整合RoI中响应值高的区域, 增加被遮挡目标无遮挡部分特征的权重, 以得到更可靠的特征, 同时利用多尺度的候选区域增强目标的上下文信息.

### 1.3 基于生成对抗网络的目标检测

文献[17]提出生成对抗式网络结构GAN (generative adversarial net), 通过生成器与判别器两个网络的对抗使生成器从给定的噪声中生成更接近真实数据分布的合成数据. 此后, GAN被广泛应用于计算机视觉各类基本任务中. 在小目标检测任务中, 一些研究将小目标模糊的特征作为生成器的输入, 通过生成器生成接近清晰图像的特征, 并进行目标检测<sup>[18-19]</sup>. 文献[20]基于生成对抗的思想, 提出对抗空间丢弃网络ASDN (adversarial spatial dropout network), 生成使目标分类损失最大的遮挡掩码作为遮挡模板来训练ASDN网络, 生成被遮挡的困难样本以增加检测模型的鲁棒性. 文献[10,16]从上下文信息角度出发, 利用目标间的语义关系增强目标的特征表示, 能够缓解由遮挡带来的特征缺失而导致的检测精度降低的问题. 以上方式仅利用未被遮挡的部位特征进行分类, 当特征缺失严重时, 会降低检测模型的准确度. 本文从恢复被遮挡特征的角度出发, 提出了学习被遮挡特征的生成对抗网络, 通过生成对抗网络对被遮挡目标缺失的特征进行恢复, 得到用于分类的更可靠的特征表示.

## 2 基于生成对抗网络学习被遮挡特征的目标检测模型

本节将详细介绍目标检测中用于被遮挡特征学习的生成对抗网络GANLOF. 训练过程通过被遮挡

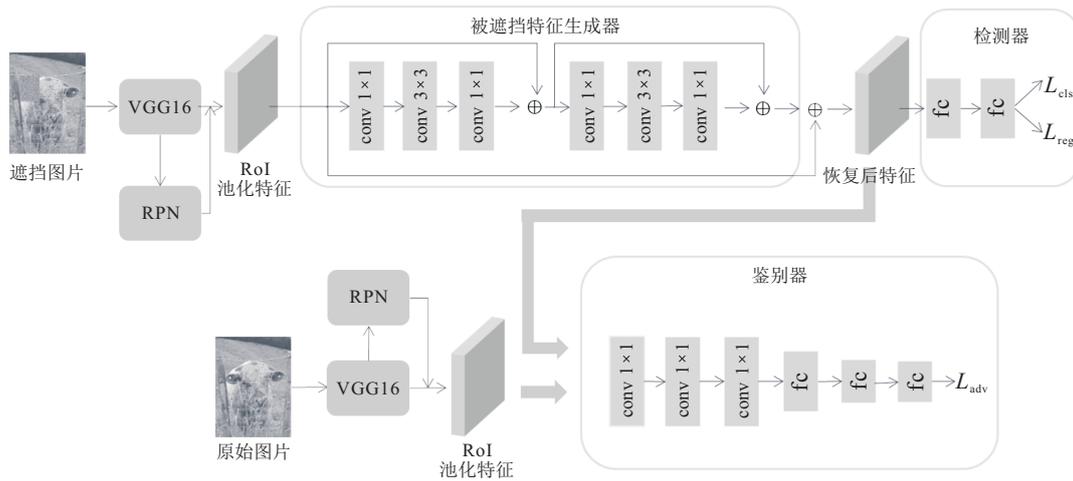


图 1 带有GANLOF模块的目标检测模型

特征生成器学习被遮挡目标缺失的特征,通过鉴别器使恢复的特征接近无遮挡特征,测试过程只使用生成器对被遮挡特征进行恢复.

### 2.1 网络结构

目标检测网络结构如图 1 所示,主要分为RoI池化特征提取、被遮挡特征生成器、鉴别器、检测器 4 个部分.

#### 2.1.1 模型的输入和输出

在训练带有 GANLOF 模块的目标检测网络时,将遮挡图片与原始图片构成的图片对作为网络的输入. 虽然数据集图片中存在遮挡情况,但是并未提供被遮挡目标的无遮挡特征标签,所以将图片中每个目标进行随机遮挡,形成由遮挡图片和原始图片构成的图片对,用于训练由被遮挡特征生成器和鉴别器构成的生成对抗网络如图 1 所示. 关于遮挡大小,若遮挡过小则会使得遮挡无效;若遮挡过大则在实际应用中不具有代表性. 为生成大小适当的遮挡图片,将遮挡大小设置为  $(w/3, h/3)$ , 其中  $w, h$  分别为目标真实标签的宽、高. 模型的输出为 RoI 对应目标的类别和回归后位置.

#### 2.1.2 特征提取

目标检测中的遮挡一般分为两种情况:一是目标被其他待检测目标遮挡,二是目标被无关物体遮挡. 对于前者,如果直接在图像上进行像素的图像恢复,则使得遮挡物的特征丢失. 因此,本文提出的模型基于 RoI 池化特征进行被遮挡目标特征的学习. 通过 VGG16<sup>[2]</sup> 提取特征,利用 RPN 网络<sup>[3]</sup> 在特征图的每个位置生成不同大小和尺度的候选区域. 由于每个 RoI 池化特征间是相互独立互不影响的,被遮挡特征生成器在恢复被遮挡目标的 RoI 池化特征时,不会改变遮挡物对应的 RoI 池化特征,在 RoI 池化特征上进

行恢复能够尽可能地减少在学习被遮挡目标特征过程中对遮挡物特征的影响.

遮挡图片通过特征提取、RoI 池化生成含有遮挡信息的池化特征  $F_{occ}^i$ , 原始图片在特征提取、RoI 池化后得到池化特征  $F_{un}^i$ ,  $(F_{occ}^i, F_{un}^i)$  构成 GANLOF 的训练样本,其中  $i$  表示第  $i$  个 RoI.

#### 2.1.3 被遮挡特征生成器

被遮挡特征生成器的作用是学习被遮挡目标与无遮挡目标之间的特征差异,进而将被遮挡目标特征恢复成无遮挡目标特征. 被遮挡特征生成器的输入是 RoI 池化特征  $F_{occ}^i$ , 输入特征图的大小为  $d \times d \times c$ , 其中  $d$  为特征图宽、高,  $c$  为特征图通道数(本文使用 VGG16<sup>[2]</sup> 骨干网络,输入图片大小为  $1000 \times 600$ , RoI 池化后  $d = 1000, c = 600$ ), 输出为被遮挡特征与无遮挡特征之间的差值,输出的特征图大小与输入特征图大小相同. 最后被遮挡特征  $F_{occ}^i$  与被遮挡生成器的输出  $G(F_{occ}^i)$  进行对应元素相加,得到恢复后的被遮挡特征  $F_{res}^i$  为

$$F_{res}^i = F_{occ}^i \oplus G(F_{occ}^i). \quad (1)$$

被遮挡特征生成器的网络设置如表 1 所示,由 6 个卷积层构成,为了避免深层网络带来的梯度消失等

表 1 被遮挡特征生成器网络设置

层	操作	尺寸	输入	输出
0	conv	1×1×128	7×7×512	7×7×128
1	conv	3×3×128	7×7×128	7×7×128
2	conv	1×1×512	7×7×128	7×7×512
3	res	—	7×7×512	7×7×512
4	conv	1×1×128	7×7×512	7×7×128
5	conv	3×3×128	7×7×128	7×7×128
6	conv	1×1×512	7×7×128	7×7×512
7	res	—	7×7×512	7×7×512

问题,设置两个残差块<sup>[21]</sup>,每个残差块包含两个 $1 \times 1$ 卷积、一个 $3 \times 3$ 卷积,使用ReLU<sup>[22]</sup>作为激活函数。

### 2.1.4 鉴别器

鉴别器用来判断输入数据是来源于被遮挡特征生成器的生成数据还是真实数据,其输入为恢复后的被遮挡特征与原始池化特征,输出为对输入特征来源的判断值.输出值介于0、1之间,接近0代表输入数据来源于生成数据,接近1代表输入数据来源于真实数据.鉴别器由3层卷积与3层全连接构成,激活函数全部采用ReLU,为了避免过拟合,在全连接过程中进行dropout<sup>[23]</sup>.鉴别器的网络设置如表2所示,卷积大小为 $1 \times 1$ ,前两层全连接的输出单元数分别为2048、1024.

表2 鉴别器网络设置

层	操作	尺寸	输入	输出
0	conv	$1 \times 1 \times 256$	$7 \times 7 \times 512$	$7 \times 7 \times 256$
1	conv	$3 \times 3 \times 128$	$7 \times 7 \times 256$	$7 \times 7 \times 128$
2	conv	$3 \times 3 \times 128$	$7 \times 7 \times 128$	$7 \times 7 \times 128$
3	fc	2048	6272	2048
4	fc	1024	2048	1024
5	fc	1	1024	1

### 2.1.5 检测器

GANLOF网络检测器采用Faster RCNN<sup>[3]</sup>检测结构,包含两层共享的全连接、一层全连接构成的分类分支以及一层全连接构成的回归分支.本文将被遮挡特征生成器输出的恢复后的被遮挡池化特征作为检测器的输入,检测器的输出则为目标类别概率以及参考坐标值.

## 2.2 损失函数

按照图1所示的目标检测结构,提出由特征学习损失 $L_{\text{mse}}$ 、对抗损失 $L_{\text{adv}}$ 、检测损失 $L_d$ 构成的联合损失函数,以实现端到端的训练.其中特征学习损失 $L_{\text{mse}}$ 用于指导生成器学习被遮挡池化特征缺失的信息;对抗损失 $L_{\text{adv}}$ 用于判别特征的来源,区分其是生成特征还是原始特征;检测损失 $L_d$ 用于指导检测过程,使分类和回归的结果更加准确.在端到端训练过程中,通过最小化联合损失,使得生成器学习到更接近原始特征的被遮挡特征,使鉴别器能够更准确地区分输入来源,同时使目标检测的结果更加准确.

特征学习损失 $L_{\text{mse}}$ 采用MSE(mean square error)损失,有

$$L_{\text{mse}} = \frac{1}{N} \sum_{i=1}^N \|F_{\text{res}}^i - F_{\text{un}}^i\|^2. \quad (2)$$

其中: $F_{\text{res}}^i$ 为恢复后的被遮挡特征, $F_{\text{un}}^i$ 为原始图片

RoI池化特征, $N$ 为RoI的总数量.

对抗损失 $L_{\text{adv}}$ 采用交叉熵损失,有

$$L_{\text{adv}} = \min_G \max_D \frac{1}{N} \sum_{i=1}^N p_i^* [\log D(F_{\text{un}}^i) - \log(1 - D(F_{\text{res}}^i))]. \quad (3)$$

其中 $D(\cdot)$ 表示对输入特征来源的预测值,在训练鉴别器过程中,当输入为原始池化特征时, $D(F_{\text{un}}^i)$ 的标签为1;当输入为生成的恢复特征时, $D(F_{\text{res}}^i)$ 的标签为0.在训练生成网络的过程中, $D(F_{\text{res}}^i)$ 的标签为1.通过最小化损失函数,使得生成器生成的恢复特征越来越接近原始池化特征.本文对RoI池化特征进行恢复,RoI中存在大量负样本,即背景候选区域,而本文仅对目标进行随机遮挡,所以遮挡图片的负样本RoI池化特征与原始图片的负样本RoI池化特征相同,不需要区分负样本RoI来源于遮挡图片还是原始图片.在进行梯度回传时只计算正样本RoI的损失,其中当RoI为正样本时, $p_i^*$ 值为1,否则为0.

本文采用的检测损失为Fast RCNN分类和回归损失<sup>[8]</sup>,有

$$L_d(p, u, t^u, v) = L_{\text{cls}}(p, u) + \lambda[u \geq 1]L_{\text{loc}}(t^u, v). \quad (4)$$

其中: $p$ 为预测框类别概率, $u$ 为真实类别标签, $t^u$ 为类别 $u$ 预测边框的坐标矩阵, $v$ 为类别 $u$ 对应的真实边框的坐标矩阵.

最终的联合损失函数如下所示:

$$L = L_{\text{mse}} + \alpha L_{\text{adv}} + \beta L_d. \quad (5)$$

其中 $\alpha$ 、 $\beta$ 为对抗损失和检测损失的权值,根据实验结果本文 $\alpha$ 设置为0.001, $\beta$ 设置为0.1.

## 3 实验分析

### 3.1 数据集

为了验证GANLOF的有效性,本文在两个公开数据集VOC2007<sup>[4]</sup>与KITTI<sup>[5]</sup>上进行实验.VOC2007数据集共包含20个类别,训练数据包含5011张图片,测试数据包含4952张图片.在VOC2007数据集上,与Fast RCNN<sup>[8]</sup>、Pruning Reduced RPN Multi-Feature<sup>[24]</sup>、Faster RCNN<sup>[3]</sup>、ASDN<sup>[20]</sup>、ORGCN<sup>[10]</sup>模型进行对比.

同时,将本文方法与Faster RCNN模型在KITTI数据集上进行比较.KITTI数据集中训练数据包含7481张图片,测试数据包含7518张图片.由于KITTI测试数据只包含图片信息,不含目标类别、位置等标签信息,无法计算模型在KITTI测试数据上的mAP值,本文对KITTI训练数据进行划分,分为训练验证

集、测试集两个部分,便于对模型进行评测.划分后的训练验证集和测试集分别含有3740张、3741张图片. KITTI数据中包含目标遮挡分级信息,为验证所提出方法对被遮挡目标的检测效果,将测试集划分为3类,分别为完全可见测试集(fully visible)、部分遮挡测试集(partly occluded)和大面积遮挡测试集(largely occluded).

### 3.2 实现细节

为得到被遮挡目标与无遮挡目标训练样本,分别对VOC2007和KITTI数据集中的图片进行随机遮挡,形成训练GANLOF网络的被遮挡RoI池化特征、无遮挡RoI池化特征训练样本对.将图片中每个目标分为3×3的网格,随机选择一个网格进行遮挡,并使用随机数对遮挡位置的像素进行填充.遮挡后的数据集作为训练样本,无遮挡的原始数据集作为对应遮挡图片的标签样本.

本文使用Faster RCNN模型作为检测网络,模型代码使用Tensorflow版本<sup>[25]</sup>,并使用基于ImageNet数据训练的VGG16作为提取特征的骨干网络.首先对Faster RCNN网络进行训练,使模型具有对特征分类的能力,训练得到的Faster RCNN在VOC2007上的mAP值为70.7%.最后对生成的对抗网络(包括生成器和鉴别器)进行端到端的训练,学习和恢复被遮挡特征.训练生成器时利用Adam技术<sup>[26]</sup>进行优化,学习率设置为0.0002;训练鉴别器使用SGD(stochastic gradient descent)优化方法,学习率设置为0.001.

### 3.3 mAP评价指标

mAP是用于评价目标检测模型的评价指标,主要通过精度(precision,  $P$ )和召回率(recall,  $R$ )计算得到. mAP的计算过程可以分为判断预测结果、计算精度和召回率、计算mAP三个步骤.

1) 判断预测结果:首先根据预测结果与真实边框的IoU判断其属于真正例(true positives, TP)还是假

正例(false positives, FP),将IoU大于阈值,且在对应同一目标的预测结果中置信度最大的预测结果记为TP,否则记为FP.

2) 计算精度和召回率  $P$ :精度是指每个类别预测结果中的TP数量与该类别预测结果总数的比值,即

$$P = \frac{TP}{TP + FP} \quad (6)$$

召回率为类别预测结果中的TP数量与该类别目标数量的比值,即

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

3) 计算mAP:根据PR(准确率-召回率)曲线计算每个类别的平均精度AP,即PR曲线与坐标轴的面积,再计算所有类别AP的平均值,得到最终的mAP.

## 3.4 实验结果

### 3.4.1 与Faster RCNN模型对比结果

1) VOC2007数据集实验结果对比.

本文方法在VOC2007测试集上的mAP值为70.9%,比基准方法Faster RCNN提升0.2%.表3给出了GANLOF与Faster RCNN在VOC2007数据集上20个类别的AP值结果,本文模型对大部分类别的AP值均有提升.同时比较了不同网络mAP值随迭代次数的变化情况,如图2所示.由图2可见,本文模型在60000次迭代时mAP达到70.9%,Faster RCNN在70000次迭代时达到70.7%.

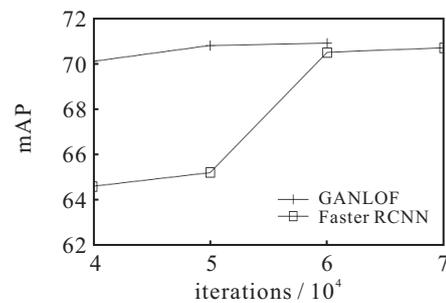


图2 不同模型训练收敛情况

表3 在VOC2007数据集上的mAP值对比

model	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Faster RCNN	70.7	69.9	<b>78.3</b>	65.6	56.8	<b>55.7</b>	78.2	<b>82.4</b>	84.6	52.6	<b>76.9</b>	67.8	81.5	83.7	74.4	<b>77.8</b>	<b>43.6</b>	<b>70.7</b>	<b>65.6</b>	75.0	<b>73.0</b>
GANLOF(ours)	<b>70.9</b>	<b>70.8</b>	77.8	<b>68.2</b>	<b>58.2</b>	55.1	<b>78.4</b>	81.4	<b>86.1</b>	<b>52.6</b>	76.0	<b>67.9</b>	<b>82.0</b>	<b>83.9</b>	<b>75.8</b>	77.6	41.8	70.0	65.5	<b>75.9</b>	72.9

2) KITTI数据集实验结果对比.本文模型与Faster RCNN使用相同的训练集,共包含图片3740张.按照第3.1节KITTI测试集划分方式,将测试集划分为完全可见数据集、部分被遮挡数据集以及大面积被遮挡数据集.如表4所示,本文模型在全部测试

集上的mAP高出Faster RCNN模型0.46%,在3个子测试集中,与Faster RCNN相比分别高出0.03%、1.61%和0.29%.可以看出,GANLOF在无遮挡数据(fully visible)上的检测精度提高有限,这是因为本文方法主要是恢复被遮挡目标的RoI池化特征,对无遮

挡目标特征没有进一步加强.而在部分被遮挡子测试集 (partly occluded) 中, GANLOF 提高效果较为明显, 提升了 1.61%; 在大面积被遮挡子测试集 (largely occluded) 上提升了 0.29%, 这是由于本文在对图像进行遮挡时仅随机遮挡了目标的 1/9, 而大面积遮挡数据中目标至少被遮挡 50%, 对大面积被遮挡目标的特征恢复没有部分被遮挡目标的特征恢复效果好. 所以, 当待检测目标被遮挡时, 通过 GANLOF 网络对被遮挡特征进行恢复, 能使检测器更准确地检测出目标.

表 4 在 KITTI 数据集上的 mAP 值对比

model	KITTI fully visible	partly occluded	largely occluded	
Faster RCNN	78.41	88.45	85.22	72.64
<b>GANLOF</b>	<b>78.87</b>	<b>88.48</b>	<b>86.83</b>	<b>72.93</b>

同时验证了 IoU 阈值  $T_{IoU}$  分别为 0.5、0.6、0.7 三种情况下模型 mAP 值的提升效果, 如表 5 所示, 相对 Faster RCNN 分别提升了 0.46%、0.62% 和 0.2%. 可以看出, 当检测要求更严格时, 本文方法仍有一定的提升效果.

表 5 不同 IoU 阈值下 GANLOF 与 Faster RCNN 结果对比

model	$T_{IoU} = 0.5$	$T_{IoU} = 0.6$	$T_{IoU} = 0.7$
Faster RCNN	78.41	74.75	65.66
<b>GANLOF</b>	<b>78.87</b>	<b>75.37</b>	<b>65.86</b>

### 3.4.2 与其他模型对比结果

表 6 给出了在 VOC2007 数据集上本文方法与其他目标检测模型 mAP 值的对比结果. 由表 6 可见, 本文方法相比于 Fast RCNN<sup>[8]</sup>、Pruning Reduced RPN Multi-Feature<sup>[26]</sup>、ASDN<sup>[20]</sup> 和 ORGCN<sup>[10]</sup> 方法, mAP 值也有一定的提升, 原因在于 GANLOF 通过生成器生成被遮挡目标恢复后的池化特征, 为检测器对目标进行分类提供了更有力的特征图.

表 6 实验结果对比

model	backbone	mAP
Fast RCNN	VGG16	66.9
pruning reduced RPN multi-feature	VGG16	69.1
Faster RCNN	VGG16	70.7
ASDN	VGG16	70.8
ORGCN	VGG16	70.83
<b>GANLOF</b>	<b>VGG16</b>	<b>70.9</b>

图 3 展示了模型在 VOC2007 数据集上对被遮挡特征的学习效果. 特征图中矩形框为遮挡位置的特征, 不同亮度代表不同的特征值. 由图 3 可见, 被遮挡

特征生成器生成的恢复后的池化特征接近无遮挡 RoI 池化特征.

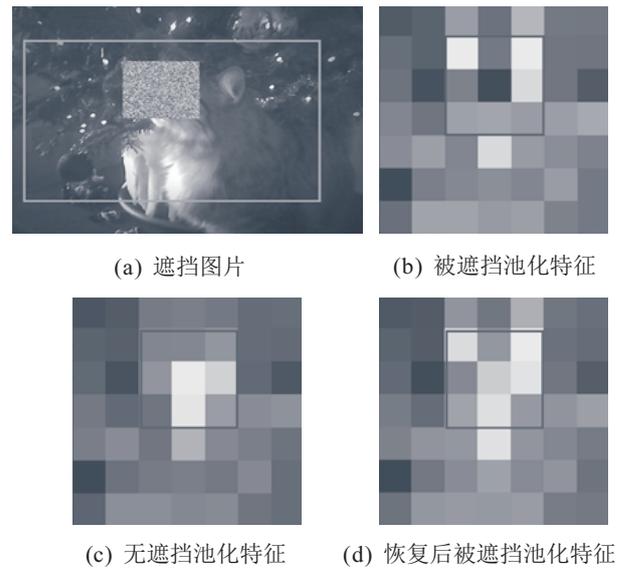


图 3 被遮挡特征学习效果

## 4 结论

本文提出了一种用于被遮挡特征学习的生成对抗网络, 能够有效对被遮挡特征进行恢复, 提高对被遮挡目标的检测效果. GANLOF 模型可以作为通用组件应用到任何两阶段目标检测模型中, 对 RoI 池化特征进行特征恢复. 目前, 本文将目标真实边框划分为  $3 \times 3$  网格, 通过随机选择网格进行遮挡生成遮挡图片, 这种方式生成的遮挡可能位于目标边框中的背景区域, 并未对检测目标形成遮挡, 在一定程度上影响了 GANLOF 对于被遮挡目标特征的学习效果. 所以在未来工作中将考虑研究更好的图片遮挡方法, 进一步提升被遮挡特征学习的效果, 以及继续研究有效的生成器和训练方法, 提高特征恢复的质量.

## 参考文献 (References)

- [1] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. 2014, arXiv: 1409.1556.
- [3] Ren S, He K, 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, 2015, 39(6): 1137-1149.
- [4] Everingham M, Winn J. The PASCAL visual object classes challenge 2007 (VOC2007) development kit[J]. International Journal of Computer Vision, 2006, 111(1): 98-136.
- [5] 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.
- [6] 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. Columbus: IEEE, 2014: 580-587.
- [7] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904-1916.
- [8] Girshick R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [9] Uijlings J R R, Sande K E A. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [10] He Chen-hang, Lai Shun-cheung Lai, Lam Kin-Man. Improving object detection with relation graph inference[C]. International Conference on Acoustics, Speech and Signal Processing. Brighton: IEEE, 2019: 2537-2541.
- [11] 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. New York: IEEE, 2016: 779-788.
- [12] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.
- [13] Singh K K, Lee Y J. Hide-and-Seek: Forcing a network to be meticulous for weakly-supervised object and action localization[C]. Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 3544-3553.
- [14] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems. Lake Tahoe: MIT Press, 2012: 1097-1105.
- [15] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818-2826.
- [16] Bae S H. Object detection based on region decomposition and assembly[C]. Association for the Advance of Artificial Intelligence. Hawaii: AAAI, 2019: 8094-8101.
- [17] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. Advances in Neural Information Processing Systems. Montreal: MIT Press, 2014: 2672-2680.
- [18] Li J, Liang X, Wei Y, et al. Perceptual generative adversarial networks for small object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1951-1959.
- [19] Bai Y C, Zhang Y Q, Ding M L, et al. SOD-MTGAN: Small object detection via multi-task generative adversarial network[C]. Proceedings of the European Conference on Computer Vision. Munich: Springer, 2018: 206-221.
- [20] Wang X L, Shrivastava A, Gupta A. A-Fast-RCNN: Hard positive generation via adversary for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 3039-3048.
- [21] 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. Las Vegas: IEEE, 2016: 770-778.
- [22] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: JMLR, 2011: 315-323.
- [23] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.
- [24] Shih Kuan-Hung, Chiu Ching-Te, Pu Yen-Yu. Real-time object detection via pruning and a concatenated multi-feature assisted region proposal network[C]. International Conference on Acoustics, Speech and Signal Processing. Brighton: IEEE, 2019: 2537-2541.
- [25] Xinlei Chen, Abhinav Gupta. An implementation of faster RCNN with study for region sampling[J]. 2017, arXiv: 1702.02138.
- [26] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. 2014, arXiv: 1412.6980.

### 作者简介

安珊(1995—),女,硕士生,从事人工智能、计算机视觉等研究, E-mail: An943314718@163.com;

林树宽(1966—),女,教授,从事人工智能、模式识别、数据挖掘等研究, E-mail: linshukuan@ise.neu.edu.cn;

乔建忠(1964—),男,教授,博士生导师,从事模式识别、并行计算等研究, E-mail: qiaojianzhong@ise.neu.edu.cn;

李川皓(1996—),男,硕士生,从事人工智能、计算机视觉等研究, E-mail: 13359853236@163.com.

(责任编辑: 郑晓蕾)