

改进YOLOv2卷积神经网络的多类型合作目标检测

王建林, 付雪松, 黄展超, 郭永奇, 王汝童, 赵利强

引用本文:

王建林, 付雪松, 黄展超, 等. 改进YOLOv2卷积神经网络的多类型合作目标检测[J]. *光学精密工程*, 2020, 28(1): 251–260.

WANG Jian-lin, FU Xue-song, HUANG Zhan-chao, et al. Multi-type cooperative targets detection using improved YOLOv2 convolutional neural network[J]. *Optics and Precision Engineering*, 2020, 28(1): 251–260.

在线阅读 View online: <https://doi.org/10.3788/OPE.20202801.0251>

您可能感兴趣的其他文章

Articles you may be interested in

基于改进YOLOv2的快速安全帽佩戴情况检测

Fast helmet-wearing-condition detection based on improved YOLOv2

光学精密工程. 2019, 27(5): 1196–1205 <https://doi.org/10.3788/OPE.20192705.1196>

利用卷积神经网络的自动驾驶场景语义分割

Autonomous driving semantic segmentation with convolution neural networks

光学精密工程. 2019, 27(11): 2429–2438 <https://doi.org/10.3788/OPE.20192711.2429>

多模深度卷积神经网络应用于视频表情识别

Video-based facial expression recognition using multimodal deep convolutional neural networks

光学精密工程. 2019, 27(4): 963–970 <https://doi.org/10.3788/OPE.20192704.0963>

基于三维卷积神经网络的低剂量CT肺结节检测

Detection of low dose CT pulmonary nodules based on 3D convolution neural network

光学精密工程. 2018, 26(5): 1211–1218 <https://doi.org/10.3788/OPE.20182605.1211>

基于卷积神经网络的光学遥感图像检索

Optical remote sensing image retrieval based on convolutional neural networks

光学精密工程. 2018, 26(1): 200–207 <https://doi.org/10.3788/OPE.20182601.0200>

文章编号 1004-924X(2020)01-0251-10

改进 YOLOv2 卷积神经网络的多类型合作目标检测

王建林*, 付雪松, 黄展超, 郭永奇, 王汝童, 赵利强
(北京化工大学 信息科学与技术学院, 北京 100029)

摘要:针对大型构件三维精密测量中构件结构复杂、测量环境变化等导致的合作目标检测精度低的问题,提出一种改进 YOLOv2 卷积神经网络的多类型合作目标检测方法。首先,利用 WGAN-GP 生成对抗网络扩增合作目标图像样本数量;其次,采用卷积层密集连接代替 YOLOv2 基础网络的逐层连接增强图像特征信息流,引入空间金字塔池化汇聚图像局部区域特征,构建改进 YOLOv2 卷积神经网络的多类型合作目标检测方法;最后,采用增强的目标图像样本数据集训练改进 YOLOv2 卷积神经网络的多类型合作目标检测模型,实现多类型合作目标检测。实验结果表明:采用多类型合作目标图像数据集测试,多类型合作目标检测精度达到 90.48%,目标检测速度为 58.7 frame/s。该方法具有较高的检测精度和速度,鲁棒性好,满足大型构件三维精密测量中多类型合作目标检测的要求。

关键词:合作目标;目标检测;数据增强;改进 YOLOv2;卷积神经网络

中图分类号:TP394.1;TH691.9 **文献标识码:**A **doi:**10.3788/OPE.20202801.0251

Multi-type cooperative targets detection using improved YOLOv2 convolutional neural network

WANG Jian-lin*, FU Xue-song, HUANG Zhan-chao, GUO Yong-qi,
WANG Ru-tong, ZHAO Li-qiang

(College of Information Science and Technology, Beijing University of
Chemical Technology, Beijing 100029, China)

* Corresponding author, E-mail: wangjl@mail.buct.edu.cn

Abstract: In the three-dimensional (3D) precision measurement of large component, the detection accuracy of cooperative targets is low due to complex structure of large components and various measurement environment. To solve this problem, a multi-type cooperative target detection method using improved YOLOv2 convolutional neural network was proposed. Firstly, the data augmentation method combined with WGAN-GP was employed to amplify the number of cooperative target images. Secondly, the convolutional layer dense connection was used instead of the YOLOv2 basic network layer-by-layer connection to enhance image feature information flow, and the spatial pyramid pooled was introduced to convergence image local area feature. Base on those two parts, the multi-type cooperative targets detection method with improved YOLOv2 convolutional neural network was constructed. Fi-

收稿日期:2019-07-08;修订日期:2019-09-12.

基金项目:国家重点研发计划资助项目(No. 2017YFF0107303)

nally, the multi-type cooperative targets detection model with improved YOLOv2 convolutional neural network was trained by the augmentation dataset for detecting the multi-type cooperative targets. The experimental results of multi-type cooperative target detection indicate that, detection precision of the proposed method is up to 90.48%, and detection speed is 58.7 frame per second by using image dataset of multi-type cooperative targets to test. This method has higher precision, rapid speed and strong robustness, which can satisfy the multi-type cooperation targets' detection requirements for 3D precision measurement of the large component.

Key words: cooperation target; object detection; data augmentation; improved YOLOv2; convolutional neural network

1 引言

高端制造业及大型装备研制的快速发展,迫切需求对大型构件三维几何参量实现高精度测量。在大型构件三维几何参量精密测量中,由于被测目标结构复杂、测量环境变化等因素,现有的测量仪对合作目标的检测精度较低。目标视觉检测具有较高的检测精度,可同时检测多种类型目标,充分利用视觉检测的优点,准确检测多类型合作目标,能有效地提高大型构件三维几何参量精密测量的精度。

现有目标视觉检测方法主要有:基于人工设计特征的视觉检测方法和基于卷积神经网络(Convolutional Neural Networks, CNN)的视觉检测方法^[1]。

基于人工设计特征的目标视觉检测方法利用被测目标图像的人工设计特征训练分类器,通过分类器对滑动窗口图像区域的人工设计特征进行目标分类和目标定位。Felzenszwalb 等^[2]提出了一种将目标检测分解为多个部件检测的思想,设计了一种可变形部件模型(Deformable Parts Models, DPM)目标检测框架,将方向梯度直方图(Histogram of Oriented Gradient, HOG)特征与支持向量机(Support Vector Machine, SVM)分类方法紧密结合并应用于目标检测,显著地提高了目标检测精度与效率。在合作目标检测应用中,李新德等^[3]通过融合 Hu 矩、归一化转动惯量、仿射不变矩等多种特征检测目标,提高了合作目标检测精度;初广丽等^[4]通过直线和圆等几何特征检测航天器合作目标,具有对噪声、光照和旋转变化的优点;罗志伟等^[5]采用高斯混

合模型并结合圆度特征对 BGA 焊球进行缺陷检测,获得了较高的准确度和可靠性;王慧利等^[6]利用边缘-方向梯度直方图特征检测复杂背景下的舰船目标,具有边缘特征鲁棒性强、检测精度高的优点。然而,基于人工设计特征的目标视觉检测方法采用滑动窗搜索目标效率较低,人工设计特征描述目标能力有限,不同人工设计特征适用的目标类型也不相同,难以满足复杂环境下大型构件三维几何参量精密测量中多类型合作目标的检测需求。

随着数据资源的不断丰富和图形处理器性能的提升,深度卷积神经网络方法在目标检测领域得到了应用。Girshick 等^[7]首次提出了 R-CNN (Regions with CNN features)目标检测方法,采用深度卷积神经网络提取目标特征,取得了较高的目标检测精度。在此基础上,Fast RCNN^[8]、Faster R-CNN^[9]方法相继出现,目标检测精度和速度进一步提升;但是上述方法在检测目标时仍然需要分阶段进行候选区域生成和目标分类,检测速度难以满足实时性要求。Redmon 等^[10]基于回归预测的思想提出了 YOLO(You Only Look Once)目标检测方法,实现了目标实时检测,较大地提升了检测速度,但是该方法对于图像中的小尺寸目标检测精度较低,定位误差较大。Liu 等^[11]提出了一种单次多目标框检测(Single Shot MultiBox Detector, SSD)方法,引入先验候选框,在多尺度特征图上进行目标检测,目标检测精度得到提高。Redmon 和 Farhadi^[12]借鉴 SSD 方法的先验候选框策略提出了 YOLOv2 方法,相较于 YOLO 方法,目标检测精度和速度有了明显提高,然而该方法的基础网络特征提取能力较弱,未充分利用图像多尺度局部区域特征,制约了目标

检测精度的提高。随后, DSSD^[13]和 YOLOv3^[14]将基础网络改为特征提取能力更强的深度残差网络^[15],提高了目标检测精度,但是网络层数过多导致检测速度严重下降。Zhou 等^[16]提出的可变尺度检测网络(Scale-Transferrable Detection Network, STDN)方法采用密集连接网络^[17]对 SSD 进行改进,在有效提升检测速度的同时保持了良好的检测精度。Jeong 等^[18]提出了改进 SSD 目标检测方法,但是现有研究中对于 YOLO 方法的改进仍然较少。

本文针对大型构件三维精密测量中构件结构复杂、测量环境变化等导致的合作目标检测精度低的问题,提出一种改进 YOLOv2 卷积神经网络的多类型合作目标检测。利用 WGAN-GP 生成对抗网络扩增合作目标图像样本数量,采用卷积层密集连接代替 YOLOv2 基础网络的逐层连接,引入空间金字塔池化汇聚图像局部区域特征,构建改进 YOLOv2 卷积神经网络的多类型合作目标检测方法,实现多类型合作目标检测。

2 改进 YOLOv2 卷积神经网络的多类型合作目标检测方法

2.1 WGAN-GP 训练数据增强

在基于深度卷积神经网络方法的合作目标检测应用中,难以获得大量带有标注的合作目标图像样本用于模型训练,样本数量少会导致模型的泛化能力弱,出现过拟合,降低目标检测精度。为了解决合作目标图像数量不足的问题,采用基于生成对抗网络(Generative Adversarial Network, GAN)的 WGAN-GP^[19]方法扩增合作目标图像样本数量,使得目标检测模型的泛化能力和鲁棒性得到提高。

在 GAN 中,生成模型 G 和判别模型 D 之间的二元极小极大博弈表示为^[20]:

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

式中: $E(\cdot)$ 表示期望; x 为服从分布 $p_{\text{data}}(x)$ 的真实训练样本; z 为服从分布 $p_z(z)$ 的输入生成器噪声; $p_g(x)$ 表示生成样本的数据分布。

采用 Wasserstein 距离衡量两个分布的相似程度,引入梯度惩罚加快模型收敛^[19],目标函数为:

$$W(p_{\text{data}}, p_g) = \max_{D \in 1\text{-Lipschitz}} E_{x \sim p_{\text{data}}} [D(x)] - E_{x \sim p_g(x)} [D(x)] \approx \max_D \{ E_{x \sim p_{\text{data}}} [D(x)] - E_{x \sim p_g(x)} [D(x)] - \lambda_{\text{penalty}} E_{x \sim p_{\text{penalty}}} [(\| \nabla_x D(x) \| - 1)^2] \} \quad (2)$$

式中: $D \in 1\text{-Lipschitz}$, 对于任意的 x 都有 $\| \nabla_x D(x) \| \leq 1$, $\nabla_x D(x)$ 表示判别模型输入对于输出的梯度; 设生成样本分布中的点为 x_g , 真实样本分布中的点为 x_{data} , x_g 和 x_{data} 连线上的点满足分布 $p_{\text{penalty}}(x)$, λ_{penalty} 表示惩罚参数。

2.2 YOLOv2 目标检测

YOLOv2 目标检测方法划分输入图像为 $S \times S$ 个网格,每个网格采用卷积神经网络预测 K 个边界框的位置、边界框包含目标属于 C 种类型的条件概率 $\Pr(\text{Class}_i | \text{Object})$ 、目标的置信度 $\Pr(\text{Object}) * \text{IoU}$ (Intersection over Union, IoU, 表示预测框与真实框的交并比)。每个预测框的类别置信度为:

$$\Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IoU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IoU}_{\text{pred}}^{\text{truth}} \quad (3)$$

表征了目标真实边框与目标预测边框的重合程度和目标属于每个类别的概率。YOLOv2 卷积神经网络预测输出层的维度为 $S \times S \times (K \times (5 + C))$ 。

YOLOv2 卷积神经网络基础网络提取特征的中间层采用卷积-下采样的方式,梯度逐层反向传播将越来越小,最终出现“梯度消失”,且 YOLOv2 偏重于融合不同层的全局特征,未充分利用多尺度局部区域特征,限制了 YOLOv2 目标检测的精度。

2.3 卷积层密集连接

针对 YOLOv2 基础网络特征提取能力弱以及反向传播的“梯度消失”问题,采用改进的卷积层密集连接代替 YOLOv2 基础网络的逐层连接,增强 YOLOv2 基础网络特征提取能力和图像特征信息流,提高目标检测精度。

设卷积神经网络第 l 层的输入为 x^l , 卷积核权重为 w^l , 偏置为 b^l , 中间变量 $y^l = x^l * w^l + b^l$, 激活函数为 $f(\cdot)$, 损失函数为 $L(\cdot)$; 卷积层密集连接将前 $l-1$ 层输出的特征图堆叠为 $[x^0, x^1, \dots, x^{l-1}]$ 作为第 l 层的输入,当卷积神经网络正向传播时:

$$\mathbf{x}^l = f(\mathbf{w}^l * [\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^{l-1}] + \mathbf{b}^l). \quad (4)$$

当卷积神经网络反向传播时,损失函数的梯度:

$$\delta^{l-1} = \delta^l * \text{rot180}(\mathbf{w}^l) \odot f'(\mathbf{w}^{l-1} * [\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^{l-2}] + \mathbf{b}^{l-1}), \quad (5)$$

其中: * 表示卷积;rot180(·)表示对权重参数矩阵逆时针旋转 180°,⊙表示 Hadamard 积。

与 YOLOv2 的激活函数导数项 $f'(\mathbf{w}^{l-1} * \mathbf{x}^{l-2} + \mathbf{b}^{l-1})$ 相比, $f'(\mathbf{w}^{l-1} * [\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^{l-2}] + \mathbf{b}^{l-1})$ 包含输入特征图 \mathbf{x}^0 以及之前各层的输出特征;此连接方式将输入特征传入网络的每一层,能够充分利用输入特征,并且网络可以直接从损失函数计算梯度,减轻“梯度消失”现象,进而提高目标检测精度。

由于网络中过多添加密集连接将会影响目标检测速度,因此将基础网络中包含丰富特征的最后一个卷积模块的逐层连接改为密集连接;同时,由于密集连接模块位于网络的较深层,所提取的特征为深层抽象特征,所含信息较丰富,因此采用 3×3 卷积层提取特征,并采用 1×1 卷积减少特征图数量,最终卷积层密集连接单元的非线性映射表示为 BN-ReLU-Conv(3×3)-BN-ReLU-Conv(1×1)。

2.4 空间金字塔池化

为了充分利用图像局部区域特征,在 YOLOv2 中引入了改进的空间金字塔池化模块,实现局部不同尺度区域的特征池化融合,并将不同卷积层的全局特征和同一卷积层不同尺度的局部特征有效融合,进而提高目标检测精度。

改进的空间金字塔池化模块如图 1 所示。首先采用 1×1 卷积减少输入特征图数量,其次,对特征图进行多尺度池化,用 $S_{\text{pool}} \times S_{\text{pool}}$ 表示池化

窗口尺寸, $S_{\text{map}} \times S_{\text{map}}$ 表示输入特征图尺寸,则:

$$S_{\text{pool}} = \lceil S_{\text{map}} / n_i \rceil. \quad (6)$$

取 $n_i = 1, 2, 3$, 分别进行最大池化,每个尺度池化步长均为 1,利用边界填充(padding)以保持池化后的特征图尺寸不变;最终得到 3 个尺寸为 $S_{\text{map}} \times S_{\text{map}} \times 512$ 的特征图。

与 He K 等^[21]提出的空间金字塔池化相比较,改进的空间金字塔池化将 3 个维度为 $S_{\text{map}} \times S_{\text{map}} \times 512$ 的特征图池化后与输入特征图进行叠加,得到 $S_{\text{map}} \times S_{\text{map}} \times 2\,048$ 维的特征图作为输出,此方式避免了原空间金字塔池化中各层特征图变换为固定尺寸特征向量的局限性,汇聚了图像局部区域特征,提高了目标检测精度。

2.5 改进 YOLOv2 卷积神经网络的多类型合作目标检测方法

采用上述卷积层密集连接和空间金字塔池化改进 YOLOv2,构建改进 YOLOv2 卷积神经网络的多类型合作目标检测方法。图 2 为改进 YOLOv2 卷积神经网络多类型合作目标检测模型。

首先利用 5 个下采样的层叠卷积-池化模块提取目标的特征;然后采用包含由 4 个密集连接单元的密集连接模块提取目标的抽象特征,密集连接模块的每个卷积层输出 m 幅堆叠的特征图,密集连接模块的第 l 层输入 $m_0 + m \times (l-1)$ 幅堆叠的特征图,其中 m_0 为每个密集连接模块输入的特征图通道数;其后是 3 层的空间金字塔池化模块,该模块先采用 1×1 卷积减少输入特征图数量,然后汇聚多尺度局部区域特征,将输入特征图与多尺度池化特征图叠加,得到 $S_{\text{map}} \times S_{\text{map}} \times 2\,048$ 维的特征图;最后为多尺度目标检测模块,根据空间金字塔池化模块输出的特征图尺寸重构密集连接模块过渡层下采样前的特征图,使二者尺寸相同并叠加,采用 $1 \times 1 \times (K \times (5+C))$ 的卷积核对叠加后的特征图进行卷积,得到 $S \times S \times (K \times (5+C))$ 特征图并进行目标检测。

改进 YOLOv2 卷积神经网络多类型合作目标检测方法最终输出的预测框可表示为:

$$\mathbf{b} = [b_x, b_y, b_w, b_h, b_c]^T, \quad (7)$$

其中: (b_x, b_y) 为预测框中心点坐标; b_w 和 b_h 分别为预测框的宽和高; b_c 表示置信度; b_x, b_y 中相对网格左上角的偏移量 t_x, t_y 和 b_c 采用 Sigmoid 函数约束至 $[0, 1]$ 。

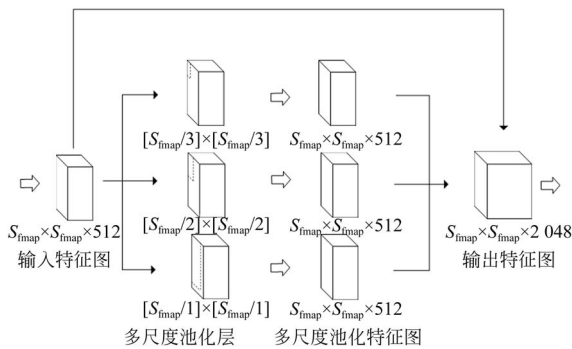


图 1 改进的空间金字塔池化

Fig. 1 Improved spatial pyramid pooling

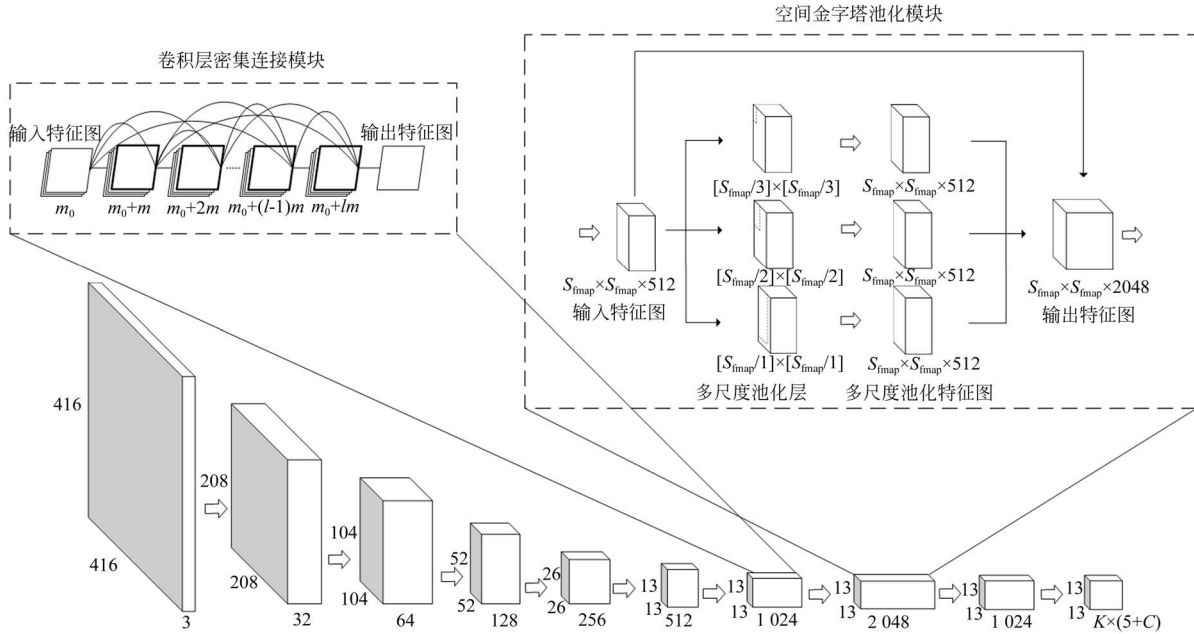


图 2 改进 YOLOv2 卷积神经网络多类型合作目标检测模型

Fig. 2 Multi-type cooperative target detection model based on improved YOLOv2 convolutional neural network

训练集中目标边界框的真实信息可表示为 $\mathbf{g} = [g_x, g_y, g_w, g_h, g_c]^T$, 目标的分类情况表示为 $\mathbf{Class} = [Class_1, Class_2, \dots, Class_C]^T$, 目标真实类别概率为 $Pr(Class_i)_{i \in C}$, 目标预测类别概率为

$\hat{Pr}(Class_i)_{i \in C}$. 目标检测损失采用目标真实类别与目标分类预测概率之间的交叉熵以及边界框真实值与边界框预测结果之间的误差平方和表示。构建损失函数:

$$L = \lambda_{noobj} \sum_{i=0}^S \sum_{j=0}^S \sum_{k=0}^K 1_{ijk}^{noobj} \cdot ((g_{c,ij} - b_{c,ijk}) \cdot ?_{\sigma}(b_{c,ijk}))^2 +$$

$$\lambda_{obj} \sum_{i=0}^S \sum_{j=0}^S \sum_{k=0}^K 1_{ijk}^{obj} \cdot ((g_{c,ij} - b_{c,ijk}) \cdot ?_{\sigma}(b_{c,ijk}))^2 +$$

$$\lambda_{coord} \sum_{i=0}^S \sum_{j=0}^S \sum_{k=0}^K 1_{ijk}^{obj} \cdot (((g_{x,ij} - b_{x,ijk}) \cdot ?_{\sigma}(b_{x,ijk}))^2 + ((g_{y,ij} - b_{y,ijk}) \cdot ?_{\sigma}(b_{y,ijk}))^2 +$$

$$(g_{w,ij} - b_{w,ijk})^2 + (g_{h,ij} - b_{h,ijk})^2) + \lambda_{class} \sum_{i=0}^S \sum_{j=0}^S \sum_{k=0}^K 1_{ijk}^{obj} \sum_{l=1}^C (-Pr_{ij}(Class_l) \log(\hat{Pr}_{ijk}(Class_l))), \quad (8)$$

其中:当预测框与真实框的交并比 IoU_{pred}^{truth} 的最大值大于阈值时, $1_{ijk}^{obj} = 1, 1_{ijk}^{noobj} = 0$, 否则 $1_{ijk}^{obj} = 0, 1_{ijk}^{noobj} = 1$; $?_{\sigma}(\cdot)$ 表示对 Sigmoid 函数的导数;超参数 $\lambda_{noobj}, \lambda_{obj}, \lambda_{coord}, \lambda_{class}$ 和 λ_{prior} 分别为各部分损失的权重系数。

WGAN-GP 的方法进行数据增强(Data Augmentation, DA),增加训练样本数量。根据训练样本的标注采用 k-means 聚类生成先验候选框^[12], 聚类距离指标为训练样本各目标边界框与聚类中心目标边界框的交并比 IoU, 即:

$$dist_{centroid}^{box} = 1 - IoU_{centroid}^{box}. \quad (9)$$

3 多类型合作目标检测算法

如图 3 为改进 YOLOv2 卷积神经网络的多类型合作目标检测方法的算法流程,包括训练数据集构建、目标检测模型训练以及目标检测三个步骤。

首先,为了防止模型训练过拟合,采用基于

其次,设置训练参数,加载卷积神经网络模型,构建损失函数,迭代更新模型权重参数使得损失函数收敛,获得改进 YOLOv2 卷积神经网络多类型合作目标检测模型。

最后,输入测试样本,加载已训练的改进 YOLOv2 卷积神经网络多类型合作目标检测模型进行多类型合作目标检测。

多类型合作目标检测算法步骤如下:

步骤 1: 将输入图像划分为 $S \times S$ 个网格, 每个网格根据先验边界框生成 K 个边界框;

步骤 2: 采用改进 YOLOv2 卷积神经网络提取特征, 通过回归预测边界框 $\mathbf{b} = [b_x, b_y, b_w, b_h, b_c]^T$ 以及分类概率 $\mathbf{Class} = [Class_1, Class_2, \dots, Class_c]^T$;

步骤 3: 选取具有最大置信度的边界框, 判断其置信度与阈值的关系; 若置信度大于阈值, 则认为此边界框中包含目标, 得到边界框的中心点坐标和尺寸参数, 否则认为该边界框不包含目标;

步骤 4: 选取包含目标的边界框预测概率最大的类型作为预测目标类型;

步骤 5: 采用非极大值抑制方法进行局部最大搜索, 抑制冗余边框, 输出并显示目标检测结果。

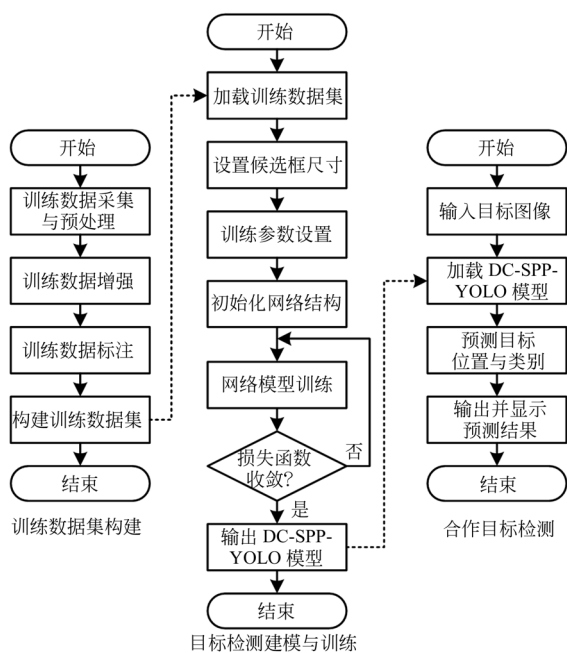


图 3 改进 YOLOv2 卷积神经网络的多类型合作目标检测算法流程

Fig. 3 Flow of multi-type cooperative target detection algorithm based on improved YOLOv2 convolutional neural network

4 实验及分析

4.1 多类型合作目标检测实验系统

实验系统计算机配置为 Intel(R) Xeon(R)

E5-2643 3.3 GHz CPU, 32.00 GB 内存, Nvidia GTX 1080Ti GPU, 11.00 GB GPU 显存, Windows10 系统 Visual Studio 2017 平台, 所使用的深度学习框架为 Darknet, 采用 C/C++ 语言编程实现。实验系统如图 4 所示。

实验结果采用平均精度 (Average Precision, AP)、平均精度均值 (mean Average Precision, mAP)、平均 IoU 和目标检测速度作为性能评价指标, 评价改进 YOLOv2 方法的目标检测性能。AP 表示网络对每类目标检测的平均精确度, 计算方式为每类目标查准率-查全率曲线 (Precision-Recall 曲线) 下方的面积, mAP 为各类目标检测的 AP 均值; 平均 IoU 为全部预测框 IoU 的均值; 目标检测速度采用每秒传输帧数表示; Precision、Recall、AP、mAP 和 IoU 可分别表示为:

$$precision = TP / (TP + FP), \quad (10)$$

$$recall = TP / (TP + FN), \quad (11)$$

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, 0.2, \dots, 1\}} precision_{max}(r), \quad (12)$$

$$mAP = \frac{1}{m} AP, \quad (13)$$

$$IoU = TP / (TP + FP + FN), \quad (14)$$

其中: FN 表示预测不是该类目标但实际是该类目标的目标数量; FP 表示预测是该类目标但实际不是该类目标的目标数量; TP 表示预测与实际均为该类目标的目标数量; $precision_{max}(r)$ 表示最大的 precision; m 表示目标类别数量。



图 4 多类型合作目标检测实验系统

Fig. 4 Experimental system of multi-type cooperative target detection

多类型合作目标数据集包含平面反射片合作目标、球形角锥棱镜合作目标, 训练集图像 3 536 幅, 验证集图像 3 178 幅, 测试集图像 2 750 幅, 测

试图像包括实验室环境和自然环境下拍摄的被测合作目标。多类型合作目标检测实验中阈值 IoU_{pred}^{truth} 设置为 0.5。

4.2 训练数据增强

采用 WGAN-GP 方法扩增训练集和验证集各 3 000 幅图像。原始训练样本图像包含不同类型合作目标且图像整体尺寸较大,导致生成高分辨率复杂环境多类型合作目标图像较困难;因此,截取训练图像中已标注的目标边界框子区域并提取区域中的目标类型,利用子区域和其中的目标类型训练 WGAN-GP 模型生成对应类别的合作目标图像。

如图 5(a)所示,用于训练的平面反射片目标子区域图像为 26 850 幅、球形角锥棱镜目标子区域图像为 21 350 幅,图像尺寸均为 128×128 pixel,其他设置与 Gulrajani 等^[19]相同。图 5(b)所示为 WGAN-GP 生成的多类型合作目标图像。将生成的合作目标图像与原始图像目标子区域叠加获得合成图像样本,如图 5(c)所示。根据添加的合作目标类型以及随机添加的目标子区域位置,生成合成图像样本的先验标记信息,与原始训练样本一起构建被测合作目标检测训练数据集。

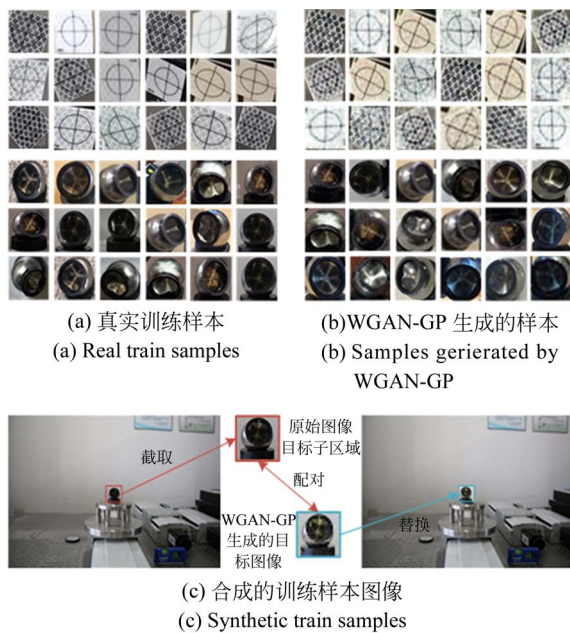


图 5 基于 WGAN-GP 的数据增强结果

Fig. 5 Experimental results of data augmentation based on WGAN-GP

采用训练数据集训练 YOLOv2 目标检测模型,表 1 为分别采用原始测试集中的真实目标和 WGAN-GP 增强数据集进行目标检测实验的结果。

表 1 YOLOv2 检测真实目标和 WGAN-GP 生成目标的结果

Tab.1 Detection results of real targets and WGAN-GP generated targets using YOLOv2 (%)

测试数据集	AP- Reflector	AP- Spherical target	mAP	平均 IoU
	原始测试集	78.20		
WGAN-GP 增强数据集	77.05	90.56	83.80	62.40

注:AP-Reflector 表示平面反射片目标平均检测精度,AP-Spherical target 表示球形角锥棱镜目标平均检测精度。

由实验结果可知,增强数据集的 mAP 为 83.80%、平均 IoU 为 62.40%,原始测试集的 mAP 为 84.45%、平均 IoU 为 64.28%,增强数据集的目标检测精度和定位准确度与原始测试集相近;表明通过 WGAN-GP 数据增强方法可生成高质量的多类型合作目标合成图像,能够较为真实地模拟多类型合作目标图像;证明了基于 WGAN-GP 方法扩增训练样本的有效性和可行性。

4.3 多类型合作目标检测

分别采用原始合作目标训练数据集和数据增强的合作目标训练数据集训练改进 YOLOv2 卷积神经网络多类型合作目标检测模型,并同时训练 YOLOv2 目标检测模型作为对照组;利用原始合作目标训练数据集训练的改进 YOLOv2 目标检测模型表示为改进 YOLOv2,利用数据增强的合作目标训练数据集训练的改进 YOLOv2 目标检测模型表示为改进 YOLOv2 + DA, YOLOv2 模型的表示方法与改进 YOLOv2 相同。

图 6、图 7 为改进 YOLOv2+DA 方法的合作目标检测结果;表 2 为在合作目标测试数据集上的目标检测实验的统计结果。

采用原始合作目标训练集训练,采用合作目标测试数据集测试,改进 YOLOv2 目标检测模型的合作目标检测 mAP 达到 89.78%,相较于

YOLOv2 模型提高了 5.33%，表明所提改进 YOLOv2 卷积神经网络的多类型合作目标检测方法能够显著提高合作目标检测精度；改进 YOLOv2 模型检测的合作目标平均 IoU 为 67.46%，相较于 YOLOv2 模型提高了 3.18%，表明了改进 YOLOv2 模型对于合作目标的定位和尺寸预测比 YOLOv2 模型更加准确；合作目标检测速度达到 58.9 frame/s，仍然保持了较快的检测速度。

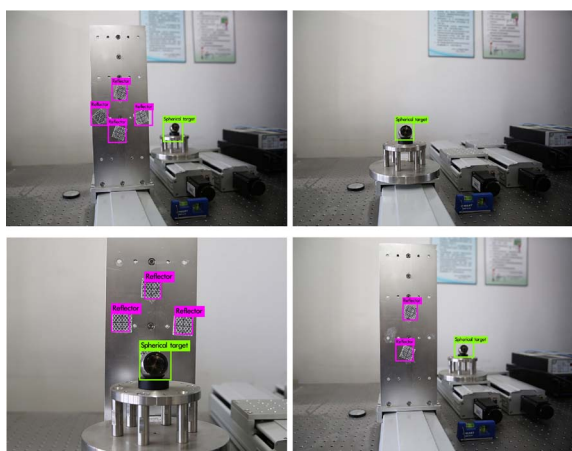


图 6 实验室环境下的合作目标检测结果

Fig. 6 Results of cooperative target detection in the cooperative laboratory environment

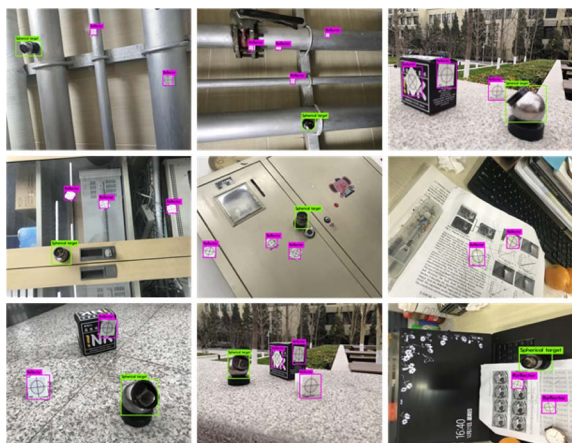


图 7 自然环境下的合作目标检测结果

Fig. 7 Results of cooperative target detection in the natural environment

采用数据增强的合作目标训练集训练，采用合作目标测试数据集测试，改进 YOLOv2+DA 目标检测模型的合作目标检测 mAP 达到

90.48%，相较于改进 YOLOv2 目标检测模型提高了 0.7%；改进 YOLOv2+DA 目标检测模型检测的合作目标平均 IoU 达到 70.89%，相较于改进 YOLOv2 目标检测模型提高了 3.43%；表明采用 WGAN-GP 方法扩增样本训练合作目标检测模型，能够有效提高合作目标检测精度。

综合比较分析 YOLOv2、YOLOv2+DA、改进 YOLOv2 和改进 YOLOv2+DA 目标检测模型的合作目标检测结果；改进 YOLOv2+DA 目标检测模型的目标检测 mAP 相较于以上三种模型分别提高了 5.89%、4.24% 和 0.7%，表明了采用卷积层密集连接和空间金字塔池化改进 YOLOv2 方法的有效性，以及基于 WGAN-GP 的数据增强对于提高合作目标检测精度的可行性。同时，基于改进 YOLOv2 和数据增强的合作目标检测仍然保持 58.7 frame/s 的较快目标检测速度，能够实现多类型合作目标的实时检测。

由图 6 和图 7 可知，所提改进 YOLOv2 卷积神经网络的多类型合作目标检测方法能够实现多类型合作目标的高精度检测，对于环境变化具有较好的鲁棒性，适用场景较广。

表 2 改进 YOLOv2 和 YOLOv2 的合作目标检测结果比较

Tab. 2 Comparison of cooperation target detection results between Improved YOLOv2 and YOLOv2

目标检测模型	AP-Reflector /%	AP-Spherical target /%	mAP /%	平均 IoU /%	速度 / (frame/s)
YOLOv2	78.20	90.70	84.45	64.28	61.3
YOLOv2+DA	81.95	90.53	86.24	64.32	61.2
改进 YOLOv2	89.21	90.34	89.78	67.46	58.9
改进 YOLOv2+DA	90.45	90.50	90.48	70.89	58.7

5 结 论

本文采用卷积层密集连接代替 YOLOv2 基础网络的逐层连接，引入空间金字塔池化，建立了改进的 YOLOv2，在此基础上，提出了一种改进 YOLOv2 卷积神经网络的多类型合作目标检测方法，有效地增强了目标图像特征信息流，提高了

目标图像局部区域特征利用率,具有较高的多类型目标检测精度和速度。实验结果表明:采用 WGAN-GP 扩增图像数据样本具有可行性和有效性,采用 YOLOv2 检测扩增图像的 mAP 达到 83.80%;采用数据增强样本训练改进 YOLOv2 合作目标检测模型,能够使合作目标

检测的 mAP 提高了 0.7%;增强数据集训练的改进 YOLOv2 合作目标检测模型检测多类型合作目标的 mAP 达到 90.48%,检测速度达到 58.7 frame/s;所提方法能够准确地检测多类型合作目标,满足大型构件三维测量中多类型合作目标的检测要求。

参考文献:

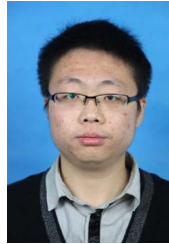
- [1] HAN J, ZHANG D, CHENG G, *et al.*. Advanced Deep-Learning techniques for salient and category-specific object detection: a survey[J]. *IEEE Signal Processing Magazine*, 2018, 35(1):84-100.
- [2] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, *et al.*. Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627-1645.
- [3] 李新德, 杨伟东, DEZERT JEAN. 一种飞机图像目标多特征信息融合识别方法[J]. *自动化学报*, 2012, 38(8):1298-1307.
Li X D, YANG W D, DEZERT J. An airplane image target's multi-feature fusion recognition method [J]. *Acta Automatica Sinica*, 2012, 38(8):1298-1307. (in Chinese)
- [4] 初广丽, 王延杰, 邸男, 等. 复杂场景中航天器靶标的快速识别[J]. *光学精密工程*, 2016, 24(4): 865-872.
CHU G L, WANG Y J, DI N, *et al.*. Fast recognition of aircraft target in complex scenes[J]. *Opt. Precision Eng.*, 2016, 24(4): 865-872. (in Chinese)
- [5] 罗志伟, 杨玉龙, 李志红. BGA 焊球视觉检测算法及系统设计[J]. *光学精密工程*, 2018, 26(9):63-70.
LUO ZH W, YANG Y L, LI ZH H. Design of vision detection algorithm and system for BGA welding balls[J]. *Opt. Precision Eng.*, 2018, 26(9): 63-70. (in Chinese)
- [6] 王慧利, 朱明, 蔺春波, 等. 光学遥感图像中复杂海背景下的舰船检测[J]. *光学精密工程*, 2018 (3):723-732.
WANG H L, ZHU M, LIN CH B, *et al.*. Ship detection of complex sea background in optical remote sensing image[J]. *Opt. Precision Eng.*, 2018, 26(3): 723-732. (in Chinese)
- [7] GIRSHICK R, DONAHUE J, DARRELL T, *et al.*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Columbus, OH, USA: IEEE, 2014: 580-587.
- [8] GIRSHICK R. Fast R-CNN[C]. 2015 *International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, 2015: 1440-1448.
- [9] RE 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*, 2017, 39(6): 1137-1149.
- [10] REDMON J, DIVVALA S, GIRSHICK R, *et al.*. You Only Look Once: Unified, Real-time Object Detection[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [11] LIU W, ANGUELOV D, ERHAN D, *et al.*. SSD: Single Shot Multi-box Detector[C]. 2016 *European Conference on Computer Vision (ECCV)*. Amsterdam, The Netherlands: Springer, 2016, 9905: 21-37.
- [12] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017: 6517-6525.
- [13] FU C-Y, LIU W, RANGA A, *et al.*. DSSD: Deconvolutional single shot detector[J]. arXiv:1701.06659, 2017.
- [14] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv: 1804. 02767, 2018.
- [15] HE K, ZHANG X, REN S, *et al.*. Deep Residual Learning for Image Recognition[C]. 2016 *IEEE Conference on Computer Vision and Pattern Rec-*

- ognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [16] ZHOU P, NI B, GENG C, *et al.*. Scale-Transferable Object Detection[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, Utah, USA: IEEE, 2018: 528-537.
- [17] HUANG G, LIU Z, LAURENS V D M, *et al.*. Densely Connected Convolutional Networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [18] JEONG J, PARK H, KWAK N. Enhancement of Ssd by Concatenating Feature Maps for Object Detection[C]. British Machine Vision Conference, 2017
- [19] GULRAJANI I, AHMED F, ARJOVSKY M, *et al.*. Improved training of wasserstein GANs [C]. Advances in Neural Information Processing Systems, 2017: 5767-5777.
- [20] GOODFELLOW I J, POUGET A J, MIRZA M, *et al.*. Generative Adversarial Nets[C]. Neural Information Processing Systems, 2014: 2672-2680.
- [21] 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, 2015, 37(9): 1904-1916.

作者简介:



王建林(1965—),男,陕西西安人,教授,1993年和1997年于天津大学获得硕士和博士学位,主要从事视觉检测技术、智能检测与传感技术等方面的研究。E-mail: wangjl@mail.buct.edu.cn



付雪松(1990—),男,内蒙古赤峰人,博士研究生,2013年于北京化工大学获得学士学位,主要从事智能检测、视觉检测等方面的研究。E-mail: 2015400133@mail.buct.edu.cn

(本栏目编辑:秦 思)