

# 改进的 YOLO V3 算法及其在小目标检测中的应用

鞠默然<sup>1,2,3,4,5\*</sup>, 罗海波<sup>1,2,4,5</sup>, 王仲博<sup>1,2,3,4,5</sup>, 何淼<sup>1,2,3,4,5</sup>, 常铮<sup>1,2,4,5</sup>, 惠斌<sup>1,2,4,5</sup>

<sup>1</sup>中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;

<sup>2</sup>中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016;

<sup>3</sup>中国科学院大学, 北京 100049;

<sup>4</sup>中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;

<sup>5</sup>辽宁省图像理解与视觉计算重点实验室, 辽宁 沈阳 110016

**摘要** 针对图像中小目标检测率低、虚警率高等问题,提出了一种 YOLO V3 的改进方法,并将其应用于小目标的检测。由于小目标所占的像素少、特征不明显,提出对原网络输出的 8 倍降采样特征图进行 2 倍上采样,将 2 倍上采样特征图与第 2 个残差块输出的特征图进行拼接,建立输出为 4 倍降采样的特征融合目标检测层。为了获取更多的小目标特征信息,在 YOLO V3 网络结构 Darknet53 的第 2 个残差块中增加 2 个残差单元。利用 K-means 聚类算法对目标候选框的个数和宽高比维度进行聚类分析。用改进的 YOLO V3 算法和原 YOLO V3 算法在 VEDAI 数据集上进行对比实验,结果表明改进后的 YOLO V3 算法能有效检测小目标,对小目标的召回率和检测的平均准确率均值都有明显的提升。

**关键词** 机器视觉; 小目标检测; YOLO V3; VEDAI 数据集; K-means 聚类算法

**中图分类号** TP391.4

**文献标识码** A

**doi:** 10.3788/AOS201939.0715004

## Improved YOLO V3 Algorithm and Its Application in Small Target Detection

Ju Moran<sup>1,2,3,4,5\*</sup>, Luo Haibo<sup>1,2,4,5</sup>, Wang Zhongbo<sup>1,2,3,4,5</sup>, He Miao<sup>1,2,3,4,5</sup>,  
Chang Zheng<sup>1,2,4,5</sup>, Hui Bin<sup>1,2,4,5</sup>

<sup>1</sup>Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China;

<sup>2</sup>Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences,  
Shenyang, Liaoning 110016, China;

<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China;

<sup>4</sup>Key Laboratory of Opt-Electronic Information Processing, Chinese Academy of Sciences,  
Shenyang, Liaoning 110016, China;

<sup>5</sup>The Key Laboratory of Image Understanding and Computer Vision, Shenyang, Liaoning 110016, China

**Abstract** This study proposes an improved detection algorithm of YOLO V3 specially applied in small target detection to solve the problems of low detection and high false alarm rates of small targets in an image. The resolution of small targets is low, and their features are not obvious; thus, this study proposes  $2\times$  upsampling for the feature map down-sampled by  $8\times$  of the previous network, and the feature map upsampled by  $2\times$  is concatenated with the output of the second ResNet block unit. A feature fusion target detection layer, whose feature map is down-sampled by  $4\times$ , is established. Two ResNet units in the second ResNet block unit of Darknet53 in the YOLO V3 network structure are added to obtain more features of the small target. The K-means clustering algorithm is used to select the number of candidate anchor boxes and aspect ratio dimensions. A comparative experiment is performed based on the improved YOLO V3 algorithm on the VEDAI dataset and YOLO V3 algorithm. The results show that the improved YOLO V3 algorithm can efficiently detect small targets and improve the mean average precision and recall rate of small targets.

**Key words** machine vision; small target detection; YOLO V3; VEDAI dataset; K-means clustering algorithm

**OCIS codes** 150.0155; 150.1135; 100.4996

收稿日期: 2019-01-11; 修回日期: 2019-02-27; 录用日期: 2019-03-22

\* E-mail: jumoran@sia.cn

0715004-1

## 1 引言

目前,目标检测已被广泛应用于军事和民用等领域中。小目标检测作为目标检测的一项关键技术,近年来已成为研究热点。远距离目标通常会呈现小目标的特性。由于小目标像素少、特征不明显,与大目标相比,其检测率低、虚警率高,因此小目标检测仍然是目标检测中的一个具有挑战性的研究课题<sup>[1]</sup>。

随着深度学习技术的不断发展<sup>[2]</sup>,多种多样的基于深度学习的目标检测<sup>[3-4]</sup>和目标跟踪算法<sup>[5]</sup>被提出。深度卷积神经网络可以利用目标数据集对要检测的目标进行自主学习,并完善自己的模型。目前基于深度学习且应用比较广泛的目标检测算法可以分为两类:1) 双步目标检测算法,如 Fast R-CNN (region-convolutional neural network)<sup>[6]</sup>、Faster R-CNN<sup>[7]</sup>、Mask R-CNN<sup>[8]</sup>等,这些算法都是将目标检测分为两个阶段,即先使用区域候选网络(RPN)来提取候选目标信息,再利用检测网络完成对候选目标的位置和类别的预测和识别;2) 单步目标检测算法,如 SSD (single shot Multibox detector)<sup>[9]</sup>、YOLO (you only look once)<sup>[10]</sup>、YOLO 9000<sup>[11]</sup>、YOLO V3<sup>[12]</sup>等,此类算法不需要使用 RPN,直接

通过网络来产生目标的位置和类别信息,是一种端到端的目标检测算法。因此,单步目标检测算法具有更快的检测速度。

为了提高网络对小目标的召回率和检测的准确率,本文以单步目标检测算法的深度卷积神经网络 YOLO V3 为基础,将航拍图像 VEDAI (vehicle detection in aerial imagery)数据集<sup>[13]</sup>中的车辆作为训练和检测的目标,对 YOLO V3 的结构进行改进。首先,对原网络输出的 8 倍降采样特征图进行 2 倍的上采样,将 2 倍上采样特征图与第 2 个残差块的输出进行拼接,以融合高层中的特征语义信息,并在 YOLO V3 网络结构 Darknet53 的第 2 个残差块中增加 2 个残差单元,以获取网络低层中准确的位置信息。最后,为避免梯度消失,将 YOLO V3 网络的目标检测输出层前的 6 个卷积变成 2 个卷积和 2 个残差单元。将改进后的网络结构与 YOLO V3 进行对比实验,结果表明改进后的网络结构对小目标的召回率和检测的准确率都有明显的提升。

## 2 YOLO V3 网络介绍

YOLO V3 采用了 Darknet53 的网络结构,共包含 53 个卷积层,如图 1 所示。

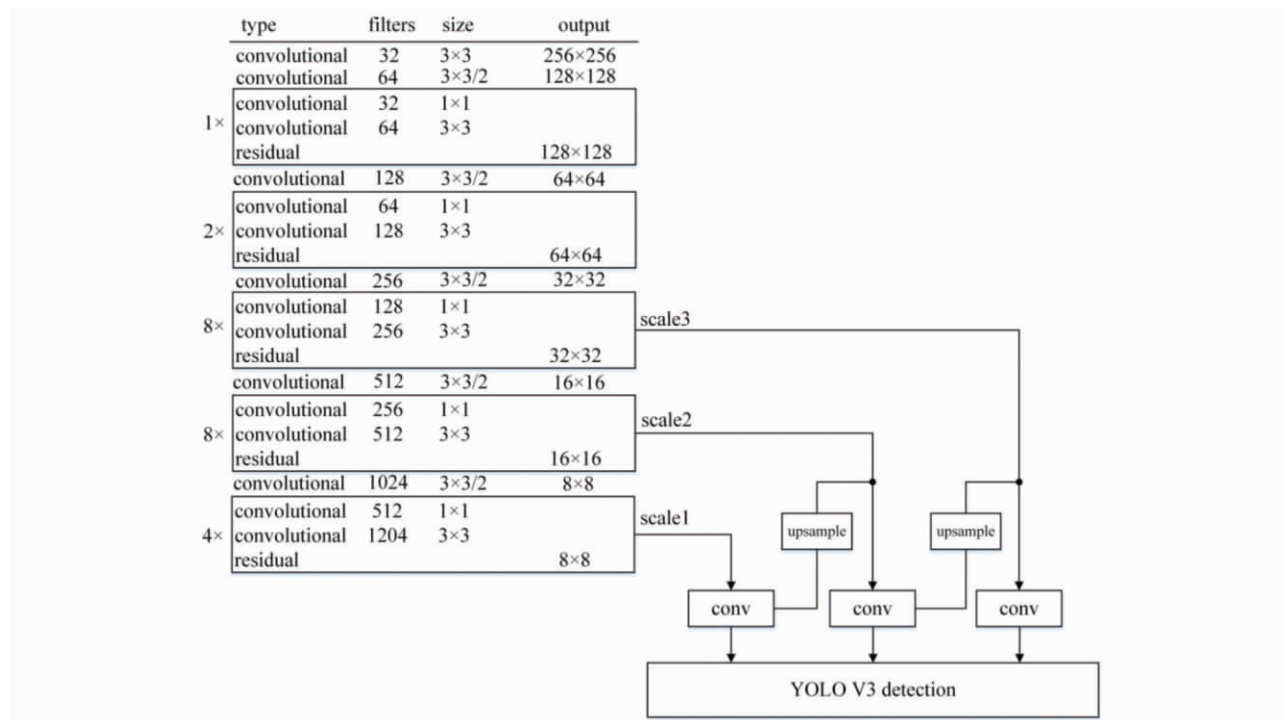


图 1 YOLO V3 网络结构

Fig. 1 YOLO V3 network structure

Darknet53 由 5 个残差块构成,借鉴了残差神经网络的思想<sup>[14]</sup>。每个残差块由多个残差单元组成,通过输入与两个数码累积造型(DBL)单元进行残差操作,构建残差单元,如图 2(a)所示。其中,

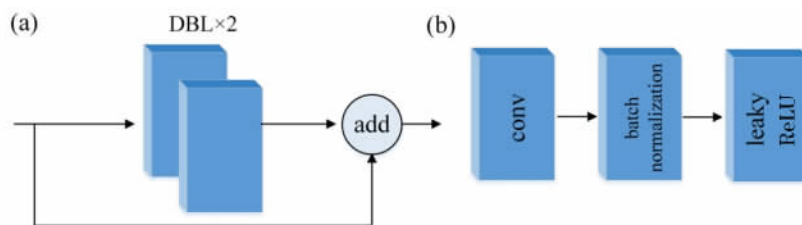


图 2 Darknet53 结构单元。(a)残差单元;(b) DBL 单元

Fig. 2 Structural unit of Darknet53. (a) ResNet unit; (b) DBL unit

YOLO V3 对输入图片进行了 5 次降采样,并分别在最后 3 次降采样中对目标进行预测。最后 3 次降采样中包含了 3 个尺度目标检测的特征图。小特征图提供深层次的语义信息,大特征图则提供目标的位置信息,小特征图经过上采样后与大特征图融合,因此该模型既可以检测大目标,也可以检测小目标。

### 3 改进的 YOLO V3 检测模型

本文主要研究小目标检测问题,原网络定义的 anchor boxes 与网络的层级结构不适用于本文的研究对象。因此,首先利用 K-means 聚类算法对数据集中的小目标进行聚类分析,然后针对小目标检测,修改网络的层级结构。

#### 3.1 数据集目标框的聚类分析

YOLO V3 引入了 Faster R-CNN 中所使用的 anchor boxes 的思想,anchor boxes 是一组宽高固定的初始候选框,对初始 anchor boxes 的选择会直接影响网络对目标的检测精度和速度。YOLO V3 利用 K-means 聚类算法对 COCO 数据集中目标框的宽高进行聚类。采用平均重叠度(Avg IOU)作为目标聚类分析的量度,对 VEDAI 数据集进行聚类分析。聚类的 Avg IOU 目标函数  $f$  可表示为

$$f = \operatorname{argmax} \frac{\sum_{i=1}^k \sum_{j=1}^{n_k} I_{\text{IOU}}(B, C)}{n}, \quad (1)$$

式中: $B$  表示样本,即 ground truth 中的目标; $C$  表示簇的中心; $n_k$  表示第  $k$  个聚类中心中样本的个数; $n$  表示样本的总个数; $k$  表示簇的个数; $I_{\text{IOU}}(B, C)$  表示簇的中心框和聚类框的交并比; $i$  表示样本序号; $j$  表示聚类中心中样本的序号。

选取  $k=1 \sim 9$ , 分别对数据集中样本进行聚类分析,得到  $k$  与 Avg IOU 之间的关系如图 3 所示。

DBL 单元包含卷积、批归一化和 leaky ReLU 激活函数,如图 2(b)所示。通过引入残差单元来增大网络深度,避免梯度消失。

随着  $k$  值的增大,目标函数变化越来越平稳,变化的拐点可以认为是最佳的 anchor boxes 的个数。在  $k$  值大于 3 时,曲线开始变得平稳,所以选取 anchor boxes 的数量为 3,既可以加快损失函数的收敛,又可以消除候选框带来的误差。对应预测框的大小设置为 3 个聚类中心,其在 VEDAI 数据集上对应聚类中心的宽和高分别为(22,9),(10,21),(20,19)。

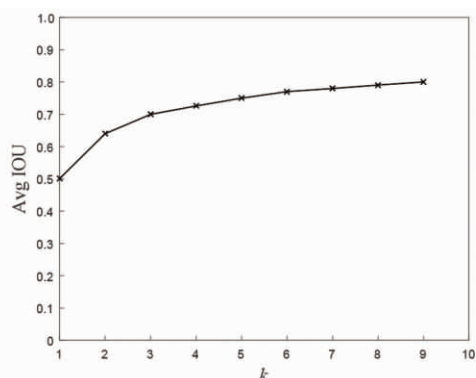


图 3 K-means 聚类分析结果

Fig. 3 K-means clustering analysis result

#### 3.2 改进的 YOLO V3 检测模型

YOLO V3 网络利用 8 倍降采样输出的特征图对小目标进行检测,这意味着当目标小于  $8 \text{ pixel} \times 8 \text{ pixel}$  时,网络对目标的预测就会出现困难,并且特征图为 8 倍降采样的目标检测层对小目标位置信息的检测能力是有限的。为了使网络获取更多小目标的特征信息,提高对小目标的检测率,利用原网络中输出的 4 倍降采样特征图对目标进行检测,因为它含有更多小目标的位置信息。对 YOLO V3 输出的 8 倍降采样特征图进行 2 倍上采样,将 2 倍上采样特征图与 Darknet53 中第 2 个残差块输出的 4 倍降采样特征图进行拼接,建立输出为 4 倍降采样的特征融合目标检测层,以检测小目标。同时,为了获

取更多低层的小目标位置信息,在原网络的第 2 个残差块中增加 2 个残差单元。

YOLO V3 采用了 3 个尺度对 VOC 和 COCO 数据集进行预测。在 COCO 数据集中尺度 1 对应的预测框是(116,90),(156,198),(373,326),尺度 2 对应的预测框是(30,61),(62,45),(59,119),尺度 3 对应的预测框是(10,13),(16,30),(33,23)。为了提高网络对小目标的召回率和检测的准确率,根据 VEDAI 数据集的聚类结果,取消 YOLO V3

在原 3 个尺度上的输出检测,直接利用输出的 4 倍降采样特征融合目标检测层对小目标进行检测。

YOLO V3 网络的目标检测输出层前包含 6 个 DBL 单元和一个  $1 \times 1$  的卷积。受 DSSD (deconvolutional single shot detector) 网络的启发<sup>[15]</sup>,为了避免梯度消失并增强特征的复用,将 6 个 DBL 单元变成 2 个 DBL 单元和 2 个 ResNet 单元,如图 4 所示。

改进的 YOLO V3 网络结构如图 5 所示,将第

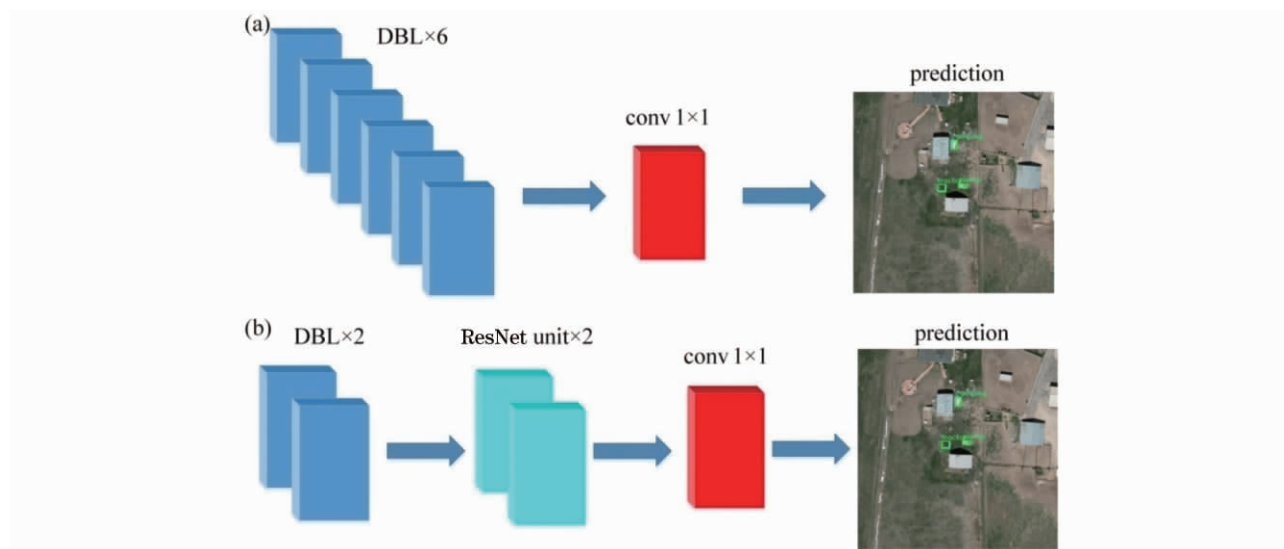


图 4 各网络的输出结构。(a) YOLO V3 网络;(b)改进的 YOLO V3 网络

Fig. 4 Output structure of each network. (a) YOLO V3 network; (b) improved YOLO V3 network

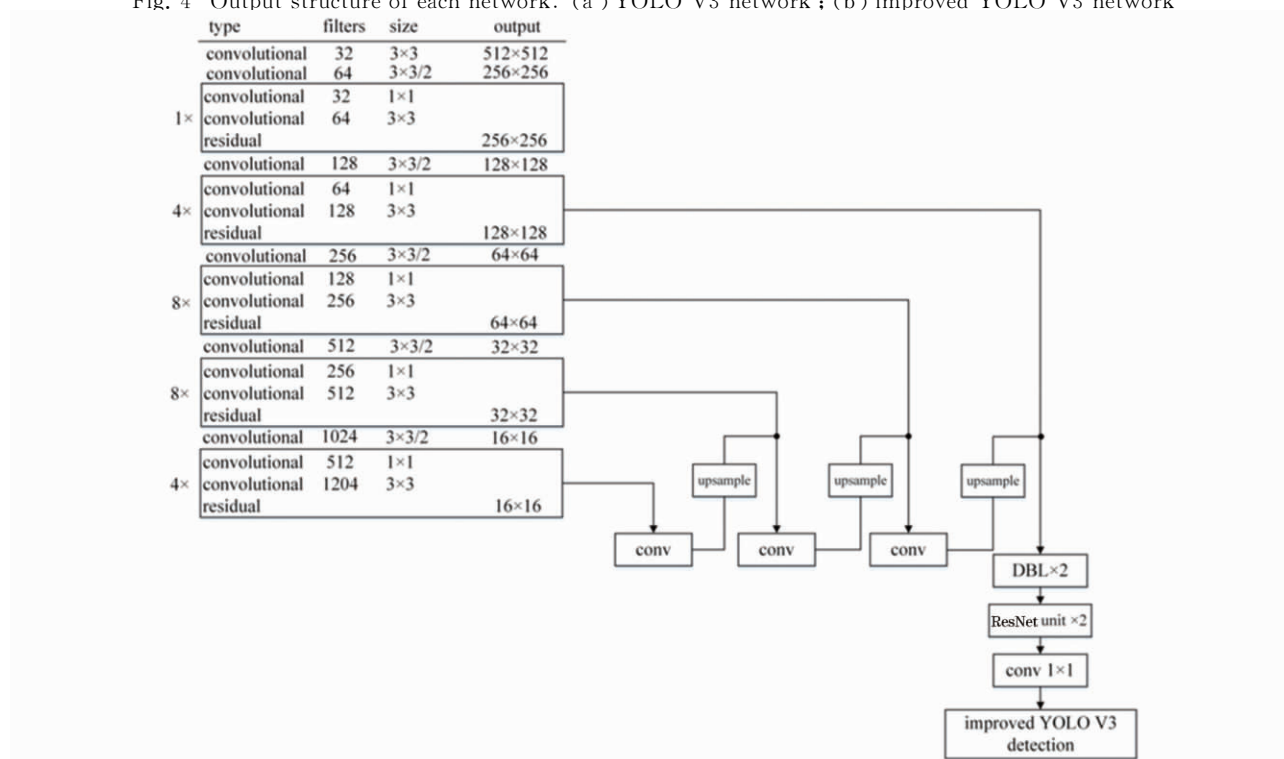


图 5 改进的 YOLO V3 网络结构

Fig. 5 Improved YOLO V3 network structure

2个残差块输出的4倍降采样特征图与经过2倍上采样的8倍降采样特征图进行拼接,建立输出为4倍降采样的特征融合目标检测层,对小目标进行检测。

YOLO V3网络会对检测的图像进行缩放或裁剪,无论输入的图像多大,最后都会被缩放或裁剪成 $416\text{ pixel} \times 416\text{ pixel}$ 的图像。采用 $512\text{ pixel} \times 512\text{ pixel}$ 的VEDAI航拍数据集,如果将图像进行缩放或裁剪,原图像中小目标的分辨率会变得更低或视场变小,直接影响网络对小目标的检测。因此,改进的YOLO V3网络不对图像进行压缩或裁剪处理,维持 $512\text{ pixel} \times 512\text{ pixel}$ 的分辨率不变,以使输入网络的图像分辨率保持不变,从而提高网络对小目标的检测性能。

#### 4 实验结果与分析

YOLO V3在多个尺度下进行目标检测,是目前能较好兼顾大、小目标的代表性算法,对小目标具

有较好的检测性能。因此,将所提的改进YOLO V3算法与YOLO V3目标检测算法进行对比实验。

利用VEDAI航拍图像作为数据集,对YOLO V3算法和改进的YOLO V3算法进行对比实验。VEDAI数据集是将原始大视场卫星图像分割成 $1024\text{ pixel} \times 1024\text{ pixel}$ 的图像,包含各种各样的车辆,背景和混淆对象。数据集中包含可见光和红外两种图像。所有的图像都采用了与地面相同的距离,是典型应用于监视侦察的场景。此外,数据集中还包含了对 $1024\text{ pixel} \times 1024\text{ pixel}$ 图像进行下采样得到的分辨率为 $512\text{ pixel} \times 512\text{ pixel}$ 的图像,为评估提供更小、更具挑战性的目标图像。VEDAI数据集中平均每张图像包含约5.5个车辆目标,并且它们只占图像总像素的0.7%,是典型的小目标检测数据集。数据集中共有9类对象,包括:airplane, boat, car, truck, tractor, camping, pickup, vans, others。9类对象中每种类别的目标数量如表1所示。

表1 VEDAI数据集中各类目标的数量

Table 1 Number of each target in VEDAI dataset

Class name	Boat	Camping	Car	Others	Pickup	Tractors	Truck	Vans	Airplane
Total number	170	390	1340	200	950	190	300	100	47

采用两种方法来验证改进后的YOLO V3网络对小目标的检测性能:1)采用分辨率为 $512\text{ pixel} \times 512\text{ pixel}$ 的图像数据集,将数据集中最小的3类目标car,vans,pickup归为一类car,并随机选择80%的图像进行训练,其余的20%图像用于测试;2)采用分辨率为 $512\text{ pixel} \times 512\text{ pixel}$ 的图像数据集,对数据集中的9类目标进行检测。

实验条件为:操作系统为Ubuntu 14.04,深度学习框架为Darknet;CPU为i7-5930K,内存为64 GB,GPU为NVIDIA GeForce GTX TITAN X。

##### 4.1 网络的训练

对YOLO V3和改进的YOLO V3分别进行训练,训练阶段初始学习率为0.001,衰减系数为0.0005,当训练迭代次数为20000次和25000次时,分别将学习率降低为0.0001和0.00001,使损失函数进一步收敛。利用旋转图像、增加对比度等方法对数据集中的图像进行增强和扩充。

改进的YOLO V3网络训练过程中损失值的收敛曲线如图6(a)所示,目标框与ground truth的Avg IOU曲线如图6(b)所示。

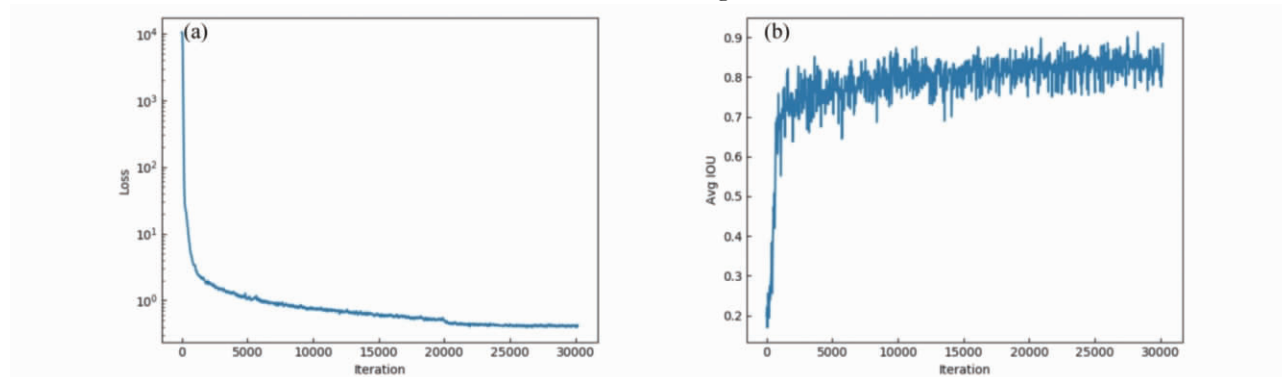


图6 改进的YOLO V3损失值函数曲线和Avg IOU曲线。(a)损失值函数曲线;(b) Avg IOU曲线

Fig. 6 Improved YOLO V3 loss curve and Avg IOU curve. (a) Loss function curve; (b) Avg IOU curve



大约经过 30000 次迭代之后,各参数基本趋于稳定,最后的损失值下降到 0.2 左右,Avg IOU 逐渐接近 1,最终稳定在 0.85 左右。从此参数的收敛情况分析可知,改进的 YOLO V3 网络的训练结果比较理想。

## 4.2 网络的测试

### 4.2.1 实验 1

将数据集中最小的 3 类目标 car、vans、pickup 归为一类(car),分别对 2 个网络进行测试,计算其对目标的召回率和检测的准确率。

目标召回率  $R$  和检测的准确率  $P$  可分别表示为

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}}, \quad (2)$$

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}}, \quad (3)$$

式中: $X_{TP}$ 表示正确检测出来的目标数; $X_{FN}$ 表示没有被检测出来的目标数; $X_{FP}$ 表示被错误检出的目标数。

测试的 166 幅图像中共有 430 个目标,使用两种目标检测算法在 VEDAI 数据集上进行测试,分

别计算  $R$  和  $P$ ,结果如表 2 所示。

表 2 不同算法目标检测结果的对比

Table 2 Comparison of target recognition results with different algorithms

Detection algorithm	$X_{TP}$	$X_{FP}$	$X_{FN}$	$P / \%$	$R / \%$
YOLO V3	373	69	57	84.3	86.7
Improved YOLO V3	392	56	37	87.5	91.2

与 YOLO V3 相比,改进的 YOLO V3 算法对小目标的检测准确率由 84.3% 提高到 87.5%,召回率由 86.7% 提高到 91.2%。

在 VEDAI 数据集上,分别利用两种网络计算小目标检测的平均精准度(AP),平均精准度是从召回率和准确率两个角度来衡量检测算法的准确性,是评价检测模型准确性的直观评价标准,可以用来分析单个类别的检测效果<sup>[16]</sup>。检测结果分别如图 7 所示,改进的 YOLO V3 网络将目标检测的平均精准度由 85.37% 提高到 90.38%。

改进的 YOLO V3 网络 and 原 YOLO V3 网络在 VEDAI 数据集下的检测结果如图 8 所示。

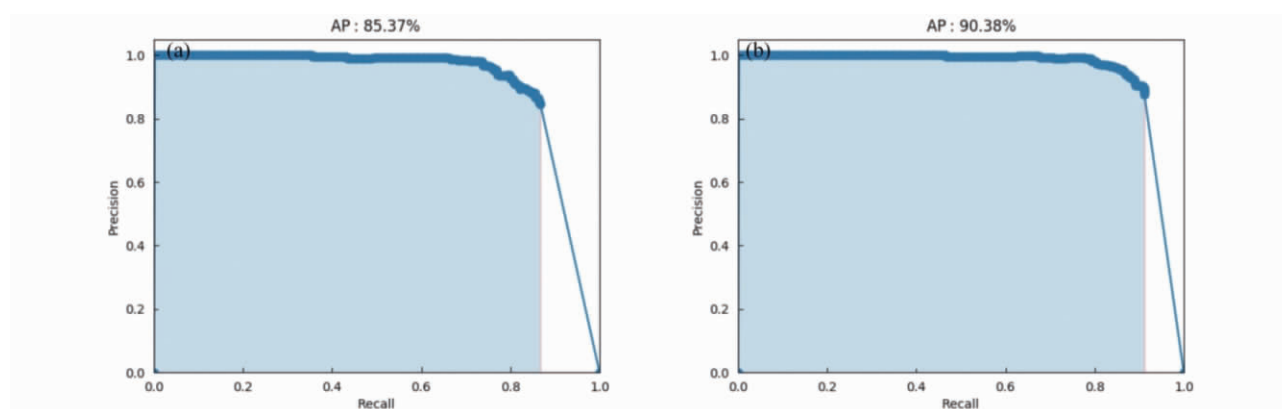


图 7 各网络的 AP 曲线。(a) YOLO V3 网络的 AP 曲线;(b)改进 YOLO V3 网络的 AP 曲线

Fig. 7 AP curves of each network. (a) AP curve of YOLO V3 ; (b) AP curve of improved YOLO V3

图 8(a)、(c)是 YOLO V3 目标检测算法对小目标的检测结果,图 8(b)、(d)是改进 YOLO V3 目标检测算法对小目标的检测结果。对比图 8(a)、(b),可以发现 YOLO V3 对小目标的检测存在漏检的问题,而改进的 YOLO V3 网络可以检测出原 YOLO V3 网络漏检的目标。对比图 8(c)、(d),可以发现图 8(c)中的深色框代表错检的目标,YOLO V3 对小目标检测存在错检的问题,而改进的 YOLO V3 网络可以避免对小目标的错检。

### 4.2.2 实验 2

分别利用 YOLO V3 和改进的 YOLO V3 对数据集中的 9 类目标进行检测,分别计算各类目标的

平均准确率,并计算 9 类目标的平均准确率均值(mAP)。mAP 值越高表示模型在全部类别中检测的综合性能越好。测试结果如表 3 所示,改进的 YOLO V3 网络将目标检测的平均精准度由 55.81% 提高到 62.36%。

改进的 YOLO V3 网络 and 原 YOLO V3 网络在 VEDAI 数据集的 9 类目标的检测结果如图 9 所示。

图 9(a)、(c)是 YOLO V3 目标检测算法对小目标的检测结果,图 9(b)、(d)是改进 YOLO V3 目标检测算法对小目标的检测结果。对比图 9(a)、(b),可以发现图 9 中的深色框代表错检的目标,YOLO V3 对小目标检测存在错检的问题,而改进

的 YOLO V3 网络可以避免对小目标的错检。对比图 9(c)、(d), 可以发现 YOLO V3 对小目标检测存在漏检的问题, 而改进的 YOLO V3 网络可以检测出原 YOLO V3 网络的漏检目标。

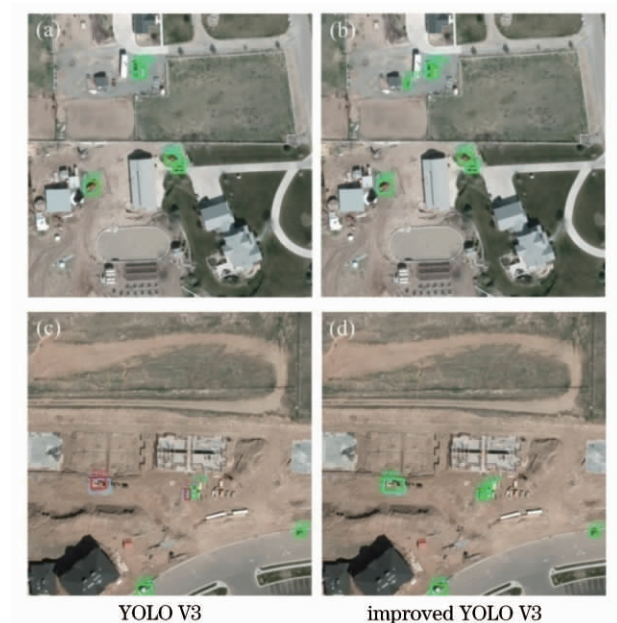


图 8 改进的 YOLO V3 算法与原 YOLO V3 算法对小目标的检测结果对比

Fig. 8 Comparison of detection results obtained by improved YOLO V3 algorithm and original YOLO V3 algorithm for small object

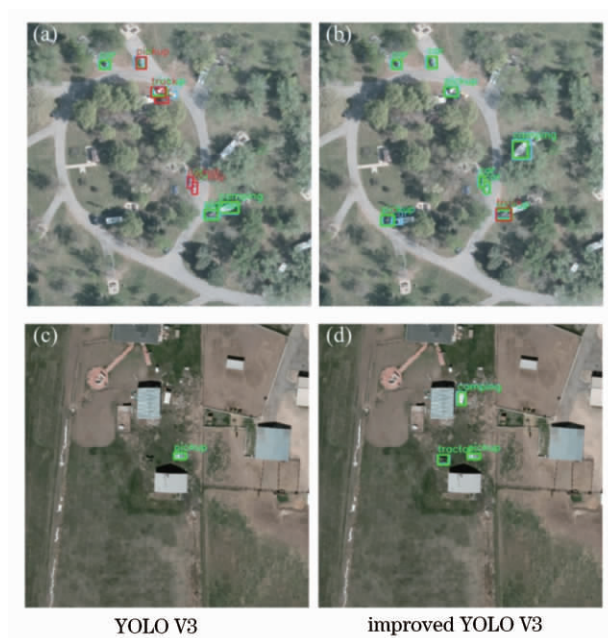


图 9 改进 YOLO V3 算法与原 YOLO V3 算法对小目标的检测结果对比

Fig. 9 Comparison of detection results obtained by improved YOLO V3 algorithm and original YOLO V3 algorithm for small object

#### 4.2.3 结果分析

根据实验 1 和实验 2 的结果, 分别从定量和定性两个方面对改进的 YOLO V3 网络 and 原 YOLO V3 网络进行性能对比。

表 3 不同算法目标检测结果的对比

Table 3 Comparison of target detection results with different algorithms

Detection algorithm	AP / %									mAP / %
	Boat	Camping	Car	Others	Pickup	Tractors	Truck	Vans	Plane	
YOLO V3	21	61	78	22	70	43	55	52	100	55.81
Improved YOLO V3	42	61	81	34	76	66	31	70	100	62.36

1) 定量对比。在实验 1 中, 改进的 YOLO V3 网络对小目标检测的准确率较原 YOLO V3 网络提高了 3.2%, 召回率提高了 4.5%, 对小目标检测的 mAP 值提高了 5.01%。在实验 2 中, 改进的 YOLO V3 网络对 9 类小目标检测的 mAP 值较原 YOLO V3 网络提高了 6.55%。

2) 定性对比。从图 8 和图 9 可以看出, 原 YOLO V3 对场景中的小目标存在漏检和错检的情况, 而改进的 YOLO V3 网络可以有效地检测出场景中的小目标, 并且可以避免对小目标的错检。

## 5 结 论

提出了一种 YOLO V3 改进方法, 并将其用于

小目标检测中。首先, 对数据集的样本进行聚类分析, 得到数据集上对应的聚类中心。然后对 YOLO V3 输出的 8 倍下采样特征图进行 2 倍上采样, 将 2 倍上采样特征图与 Darknet53 中的第 2 个残差块输出的 4 倍降采样特征图进行拼接, 建立输出为 4 倍降采样的特征融合目标检测层。最后, 为了获取更多的小目标特征信息, 在 YOLO V3 网络结构 Darknet53 的第 2 个残差块中增加 2 个残差单元, 并将 YOLO V3 网络的目标检测输出层前的 6 个 DBL 单元变成 2 个 DBL 单元和两个 ResNet 单元。实验结果表明, 改进后的 YOLO V3 算法对小目标的召回率、检测的平均准确率均有明显的提高。但是, 改进后的 YOLO V3 算法离实

时性距离工程应用仍有差距。如何在降低检测性能的前提下精简网络结构、降低计算量,将是未来的主要研究方向。

## 参 考 文 献

- [1] Zhang X Y, Ding Q H, Luo H B, *et al.* Infrared dim target detection algorithm based on improved LCM [J]. *Infrared and Laser Engineering*, 2017, 46(7): 0726002.  
张祥越, 丁庆海, 罗海波, 等. 基于改进 LCM 的红外小目标检测算法[J]. *红外与激光工程*, 2017, 46(7): 0726002.
- [2] Zhang R, Li W P, Mo T. Review of deep learning [J]. *Information and Control*, 2018, 47(4): 385-397, 410.  
张荣, 李伟平, 莫同. 深度学习研究综述[J]. *信息与控制*, 2018, 47(4): 385-397, 410.
- [3] Hua X, Wang X Q, Wang D, *et al.* Multi-objective detection of traffic based on improved SSD[J]. *Acta Optica Sinica*, 2018, 38(12): 1215003.  
华夏, 王新晴, 王东, 等. 基于改进 SSD 的交通大场景多目标检测 [J]. *光学学报*, 2018, 38(12): 1215003.
- [4] Wang W X, Fu Y T, Dong F, *et al.* Infrared ship target detection method based on deep convolution neural network [J]. *Acta Optica Sinica*, 2018, 38(7): 0712006.  
王文秀, 傅雨田, 董峰, 等. 基于深度卷积神经网络的红外船只目标检测方法 [J]. *光学学报*, 2018, 38(7): 0712006.
- [5] Luo H B, Xu L Y, Hui B, *et al.* Status and prospect of target tracking based on deep learning[J]. *Infrared and Laser Engineering*, 2017, 46(5): 0502002.  
罗海波, 许凌云, 惠斌, 等. 基于深度学习的目标跟踪方法研究现状与展望[J]. *红外与激光工程*, 2017, 46(5): 0502002.
- [6] Girshick R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [7] Ren S Q, He K M, 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.
- [8] He K M, Gkioxari G, Dollár P, *et al.* Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2980-2988.
- [9] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector [M]//Leibe B, Matas J, SbeN, *et al.* *Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [11] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [12] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08)[2018-12-25]. <https://arxiv.org/abs/1804.02767>.
- [13] Razakarivony S, Jurie F. Vehicle detection in aerial imagery: a small target detection benchmark [J]. *Journal of Visual Communication and Image Representation*, 2016, 34: 187-203.
- [14] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [15] Fu C Y, Liu W, Ranga A, *et al.* DSSD: deconvolutional single shot detector [EB/OL]. (2017-01-23)[2018-12-25]. <https://arxiv.org/abs/1701.06659>.
- [16] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved faster R-CNN [J]. *Acta Optica Sinica*, 2018, 38(6): 0615004.  
冯小雨, 梅卫, 胡大帅. 基于改进 Faster R-CNN 的空中目标检测 [J]. *光学学报*, 2018, 38(6): 0615004.