

基于 YOLOV3 的改进目标检测识别算法

王战涛¹, 张 策², 王晓田²

(1. 中国人民解放军 95889 部队, 甘肃 酒泉 735018; 2. 西北工业大学 航天学院, 陕西 西安 710072)

摘 要: 经过近几十年不断的研究和发展, 红外目标检测识别在侦察、导弹制导等领域取得了卓越的成就和广泛的应用, 亦成为当今的热门话题。为进一步提高模型的检测识别性能, 提出一种基于 YOLOV3 改进的目标检测识别算法。首先, 通过分析红外目标的检测特性, 改进了原始算法的特征提取网络, 融合 KL-LOSS, 在原网络预测目标位置的基础上, 进一步预测了位置的准确度标准差, 并结合 Soft-NMS 算法用于改善网络的检测准确度; 其次, 针对红外目标相对三通道彩色图像的特征量少的问题, 在检测层前融合了 SKNET 模块, 使网络更加关注目标的有用特征; 最后, 给出改进网络训练的新的损失函数及前向传播算法流程。实验结果表明: 改进的 KS-YOLO 网络在目标域(实拍空中红外目标数据集)上的平均 AP 性能值要优于原来的 YOLOV3 网络 2.4 个百分点, 预测时间比 YOLOV3 实用性更好、更快。

关键词: 红外目标检测; YOLOV3; SKNET 网络; Soft-NMS 算法; KS-YOLO

中图分类号: V 19

文献标志码: A

DOI: 10.19328/j.cnki.2096-8655.2021.06.009

Improved Target Detection and Recognition Algorithm Based on YOLOV3

WANG Zhantao¹, ZHANG Ce², WANG Xiaotian²

(1. Unit 95889 of the Chinese People's Liberation Army, Jiuquan 735018, Gansu, China;

2. School of Astronautics, Northwestern Polytechnic University, Xi'an 710072, Shaanxi, China)

Abstract: After decades of continuous research and development, the technology of infrared target detection and recognition has been widely applied in the fields such as military reconnaissance and missile guidance, and has made remarkable achievements. Therefore, the technology has become a hot topic today. In order to improve the detection and recognition performance of the technology, an improved target detection and recognition algorithm based on YOLOV3 is proposed in this paper. First, by analyzing the detection characteristics of infrared targets, the feature extraction network of the original algorithm is improved, and the KL-LOSS function is integrated. The target location is predicted based on the original network, based on which the accuracy standard deviation of the location is further predicted, and the detection accuracy of the network is improved in combination with the Soft-NMS algorithm. Then, in order to solve the problem that the number of feature quantities of an infrared target is less than that of a three-channel color image, the SKNET module is integrated before the detection layer so that the network could pay more attention to the useful target features. Finally, a new loss function and the forward propagation algorithm process of the improved network training are presented. The experimental results show that the average precision (AP) performance of the improved KS-YOLO network in the target domain (real-shot aerial infrared target dataset) is 2.4% better than the original YOLOV3 network, and the prediction time is better and faster.

Key words: infrared target detection; YOLOV3; SKNET network; Soft-NMS algorithm; KS-YOLO

收稿日期: 2021-03-26; 修回日期: 2021-05-20

基金项目: 航天科技创新基金项目(CASC201105)

作者简介: 王战涛(1981—), 男, 工程师, 主要研究方向为试验鉴定总体技术。

通信作者: 张 策(1997—), 男, 本科生, 主要研究方向为智能对抗。

0 引言

红外目标检测识别一直是国内外计算机视觉技术研究的热点问题,经过近几十年不断的研究和发展,其在侦察、导弹制导、安防检测及医学影像等领域上取得了卓越成就和广泛应用。由于红外成像技术能在全天候进行清晰成像显示,在一些恶劣的天气状况下也能有很好的抗干扰能力,因而得到快速发展,已经成为精确制导武器的主要研究方向之一。

随着人工智能理论与技术的发展,深度学习技术及相应的一些神经网络模型有着很强的特征学习能力,已经在图片分类、目标识别、场景理解等任务中取得了重大成果。其中,卷积神经网络在图像分类识别任务上表现出非常好的效果,成为研究的热点。基于卷积神经网络的目标检测识别算法从大体思路分为两阶段(Two-stage)目标检测识别算法与一阶段(One-stage)目标检测识别算法。

两阶段目标检测识别的经典算法有R-CNN^[1]、SPP-Net^[2]、FastR-CNN^[3]、FasterR-CNN^[4]、CascadeR-CNN算法^[5]等。而现在很多一阶段的目标检测和两阶段的目标检测相比,速度更快、精度更高。一阶段目标检测识别的神经网络经典算法有YOLO^[6]、SSD^[7]、YOLOV2^[8]、YOLOV3^[9]算法等。YOLO^[10]算法直接使用一个卷积网络就可以同时预测目标的位置和类别。YOLO算法的输入为整幅图像,统揽全局信息以降低目标的误检率,同时也可以实现实时的检测。YOLOV3算法是在原有的YOLOV2网络上进行了改进。首先,YOLOV3算法采用了多尺度预测融合的方式进行预测,其次,YOLOV3算法的Darknet-53特征提取网络对于目标的特征信息具有较好的提取效果。

YOLOV3主要有以下改进点:1)类别预测方面主要是将原来的Softmax分类损失修改为交叉熵损失。2)借鉴了特征金字塔^[11](FPN)融合不同尺度特征图的思想,采用多尺度(scale)融合的方式做预测,进一步提升了YOLOV3算法在小目标识别上的精度。3)研究表明先验框(Anchor BOX)的设定将影响算法最终的识别效果,合理的先验框可以降低算法的收敛难度提升算法性能。YOLOV3通过遍历训练数据标签的真实边界框,采用K-means聚类算法生成初始的Anchor BOX。相比之前采用人为经验设定的算法(Faster-RCNN,SSD),得到的Anchor BOX质量更高更合理。4)一方面,网络结构(Darknet-53)采用步长为2的卷积实现特征图的缩放,相比

于之前的最大池化方法(Max Pooling),卷积实现可以保留更多的特征信息;另一方面,引入了残差residual结构(YOLOV2中还是类似VGG那样直筒型的网络结构,层数太多在训练过程中容易出现梯度消失和梯度爆炸等问题,得益于ResNet的residual结构,精度提升较明显。

而本文论述的是基于改进YOLOV3的目标检测识别算法。首先,通过分析红外目标的检测特性,改进了原始算法的特征提取网络,并且融合KL-LOSS^[12];然后,再结合Soft-NMS^[13]算法用于改善网络的检测准确度,同时针对红外目标相对三通道彩色图像的特征量少的问题,在检测层前融合了SKNET模块^[14]使网络更加关注目标的有用特征;最后,给出改进网络训练的新的损失函数以及前向传播算法流程。实验结果表明,改进的KS-YOLO网络在目标域(实拍空中红外目标数据集)上的平均精度(Average Precision, AP)性能值要优于原来的YOLOV3网络2.4个百分点,预测时间要比YOLOV3更快。

1 基于YOLOV3的改进目标检测识别算法

1.1 改进特征提取网络 Backbone

如图1所示,由于对红外目标图缺乏纹理、颜色等特征,为避免模型过于复杂导致过拟合,本文降低了特征提取网络的参数量,减少原网络的3个检测层前的多残差层,其余部分仍然采用1×1卷积和3×3卷积串联的残差结构。

神经网络	卷积核个数	卷积核尺寸/步长	输入尺寸	输出尺寸
卷积	32	3*3/1	416*416	416*416
卷积	32	3*3/2	416*416	208*208
1 X	卷积	32	1*1/1	208*208
	卷积	64	3*3/1	208*208
	残差		208*208	208*208
2 X	卷积	128	3*3/2	208*208
	卷积	64	1*1/1	104*104
	卷积	128	3*3/1	104*104
8-> 2 X	卷积	256	3*3/2	104*104
	卷积	128	1*1/1	52*52
	卷积	256	3*3/1	52*52
8-> 2 X	卷积	512	3*3/2	52*52
	卷积	256	1*1/1	26*26
	卷积	512	3*3/1	26*26
4-> 2 X	卷积	1024	3*3/2	26*26
	卷积	512	1*1/1	13*13
	卷积	1024	3*3/1	13*13

图1 改进的特征提取网络

Fig.1 Improved feature extraction network

1.2 融合 KL-LOSS 损失函数

通过分析大量红外目标的图像发现,在标记真实目标边界框时,实质上不可能完全对所有目标精确地标注出来,所以在标注目标边界框时,边界框的模糊性会发生。边界框的模糊性会导致网络学习困难,并降低网络性能。特别对于部分目标对象被遮挡时,边界框的边界将变得更加模糊。YOLOV3 算法依靠回归边界框来定位目标,算法的边界框回归损失即方差损失并未考虑边界框的模糊性。KL-LOSS 损失是由 HE 等于 2019 年提出的一种新的边界框回归损失,KL-LOSS 让网络可以同时学习边界框回归和定位的不准确性,以提升网络的检测定位精度。KL-LOSS 针对边界框的移动与位置间的方差进行学习,所学习到的方差可用于后处理阶段。研究者提出了方差投票(Variance Voting)方法^[15],可在非极大值抑制(Non-maximum Suppression, NMS)期间使用由预测的方差加权的临近位置来投票得到边界框的位置,在同等计算量下提升不同结构定位的准确率。KL 边界框回归损失可以获取到数据集的模糊,从而降低预测边界框的回归损失。且所学习到的概率分布是可解释的,因为其反映了边界框预测的不确定性。

1.2.1 检测框建模

具体来说,为了得到边界框预测的不确定性,将边界框的预测建模为高斯分布,设四维向量边界框为 $b_{ax} = (x_1, y_1, x_2, y_2)$, 其中, x_1, y_1, x_2, y_2 为预测边界框的坐标位置,即左上角、右下角坐标; $t_{x_1}, t_{x_2}, t_{y_1}, t_{y_2}$ 为网络预测边界框的偏移量; $t_{x_1}^*, t_{x_2}^*, t_{y_1}^*, t_{y_2}^*$ 为真实边界框的偏移量; $x_{1a}, x_{2a}, y_{1a}, y_{2a}, w_a, h_a$ 来自先验框的坐标与尺寸。

$$t_{x_1} = \frac{x_1 - x_{1a}}{w_a}, t_{x_2} = \frac{x_2 - x_{2a}}{w_a} \quad (1)$$

$$t_{y_1} = \frac{y_1 - y_{1a}}{h_a}, t_{y_2} = \frac{y_2 - y_{2a}}{h_a} \quad (2)$$

$$t_{x_1}^* = \frac{x_1^* - x_{1a}}{w_a}, t_{x_2}^* = \frac{x_2^* - x_{2a}}{w_a} \quad (3)$$

$$t_{y_1}^* = \frac{y_1^* - y_{1a}}{h_a}, t_{y_2}^* = \frac{y_2^* - y_{2a}}{h_a} \quad (4)$$

因可单独优化每个坐标,简单起见,设边界框坐标为 x , 则可得边界框预测模型:

$$P_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_e)^2}{2\sigma^2}} \quad (5)$$

式中: θ 为可学习神经网络的参数; x_e 为网络预测估计的边界框位置。

σ 标准差用于测量预测估计值的不准确度。当标准差 σ 趋于 0 时,表示神经网络对边界框位置的预测非常准确:

$$P_D(x) = \delta(x - x_g) \quad (6)$$

式中: x_g 为真实边界框位置。

得到 2 个模型后,我们的目标是通过估计 θ 以最小化 N 个样本的 $P_{\theta}(x)$ 与 $P_D(x)$ 之间的 KL 散度。KL-LOSS 的提出,不光考虑单个预测框的正确性,而是对于 N 个样本的 KL 散度最小化,即

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{N} \sum D_{KL}(P_D(x) \| P_{\theta}(x)) \quad (7)$$

1.2.2 KL 散度

KL 散度又称 KL 距离,用于衡量 2 个概率分布之间的差异性:

$$D(P \| Q) = \sum P(x) \log \frac{P(x)}{Q(x)} \quad (8)$$

KL 散度具有 2 个重要性质:

1) 不对称性

$$D(P \| Q) \neq D(Q \| P) \quad (9)$$

2) 非负性

$$D(P \| Q) \geq 0 \quad (10)$$

1.2.3 KL-LOSS 函数的计算

KL 散度可以作为衡量 2 个概率分布之间的距离,于是可以使用 KL 散度作为边界框回归的损失函数 L_{reg} 。对于单个样本图像:

$$\begin{aligned} L_{reg} &= D_{KL}(P_D(x) \| P_{\theta}(x)) = \\ &= \int P_D(x) \log P_D(x) dx - \int P_D(x) \log P_{\theta}(x) dx = \\ &= \frac{(x_g - x_e)^2}{2\sigma^2} + \frac{\log(\sigma^2)}{2} + \frac{\log(2\pi)}{2} - H(P_D(x)) \end{aligned} \quad (11)$$

对于式中后 2 项 $\frac{\log(2\pi)}{2}$ 和 $H(P_D(x))$ 并不依赖估计参数 x_e 与 σ , 因此到得矢量 (m_w, θ_w) , 则

$$L_{reg} \propto \frac{(x_g - x_e)^2}{2\sigma^2} + \frac{\log(\sigma^2)}{2} \quad (12)$$

当 σ 趋于 1 时, KL-LOSS 退化为标准欧式损失:

$$L_{reg} \propto \frac{(x_g - x_e)^2}{2} \quad (13)$$

KL-LOSS_{reg} 对于估计参数 x_e 与 σ 是可微的:

$$\frac{d}{dx_e} L_{\text{reg}} = \frac{x_g - x_e}{\sigma^2} \quad (14)$$

$$\frac{d}{d\sigma} L_{\text{reg}} = \frac{(x_e - x_g)^2}{\sigma^3} + \frac{1}{\sigma} \quad (15)$$

由于 σ 位于分母中,可能会在训练模型时发生梯度爆炸。为避免梯度爆炸,使网络预测 $a = \log(\sigma^2)$,而不是实际中的 σ ,在实际使用 σ 值时再由公式转换成标准差,于是 KL-LOSS 变为

$$L_{\text{reg}} \propto \frac{e^{-a}}{2} (x_g - x_e)^2 + \frac{1}{2} a \quad (16)$$

当 $|x_g - x_e| > 1$ 时,定义 KL-LOSS 为

$$L_{\text{reg}} = e^{-a} \left(|x_g - x_e| - \frac{1}{2} \right) + \frac{1}{2} a \quad (17)$$

于是,可得到新的边界框损失(KL-LOSS)为

$$L_{\text{reg}} = \begin{cases} \frac{e^{-a}}{2} (x_g - x_e)^2 + \frac{1}{2} a, & 0 < |x_g - x_e| < 1 \\ e^{-a} \left(|x_g - x_e| - \frac{1}{2} \right) + \frac{1}{2} a, & |x_g - x_e| > 1 \end{cases} \quad (18)$$

KL-LOSS 不仅计算预测框的坐标位置,同时也计算位置的标准差,其中计算损失的示意图如图 2 所示。

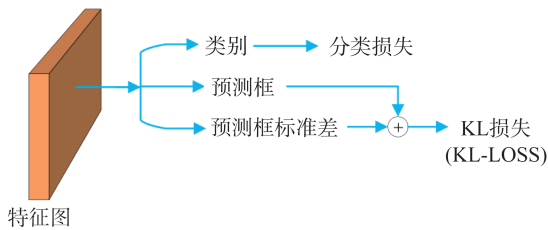


图 2 KL-LOSS

Fig.2 KL-LOSS

1.3 融入 Soft-NMS 算法

目标检测中,首先产生一些候选区域,其次通过分类网络得到类别得分,与此同时通过回归网络得到更加精确的位置参数在目标检测算法中,一个点可能对应多个预测结果,容易出现目标被多个预测框包围的情况,因此,需要采用非极大抑制的方法,保留质量最高的预测框。最后通过非极大值抑制得出最后的检测结果。目标检测算法会输出多个检测边框,尤其是在真实目标周围会有很多置信度高的检测边框,这是为除去重复的检测边框,达到每个物体有且只有一个检测结果的目的。非极大值抑制是一种获取局部最大值,抑制非极大值的

算法,在计算机视觉中有着广泛的应用。其核心思想是一个迭代一遍历一消除的过程,重叠率大于固定阈值的低分框会被高分框抑制。

非极大值抑制的具体步骤对每个类别而言,首先是设置交并比(IOU)阈值,对候选框依据类别得分进行排序,选中类别得分最高的候选框,遍历其余的候选框,如果和当前类别得分最高的候选框的 IOU 大于 IOU 阈值,将其剔除,从未处理的边框中继续选一个类别得分最高的。重复上述过程,直到处理完所有的边框为止。保留下来的候选框即为检测结果。

非极大值抑制是物体检测流程中重要的组成部分。算法首先按照得分从高到低对检测框进行排序,然后分数最高的检测框被选中,其他框与被选中建议框有明显重叠的框被抑制。该过程被不断递归地应用于其余检测框。

传统的非极大值抑制算法,首先在被检测图片中产生一系列的检测框 B 及对应的分数 S 。当选中最大分数的检测框 M 时,该框从集合 B 中移出并放入最终检测结果集 D 。与此同时,集合 B 中任何与检测框 M 的重叠部分大于重叠阈值的检测框也将随之移除。如果一个物体在另一个物体重叠区域出现,即当 2 个目标框接近时,非极大值抑制算法中会将相邻检测框的分数均强制归零,即分数更低的框就会因为与之重叠面积过大而被删掉,这就是最大问题所在。

此时,如果重叠区域出现目标,则存在目标漏检情况,甚至导致检测失败,从而降低了算法平均检测率(Average Precision, AP)。检测算法本来应该输出 2 个检测框,但是由于传统的非极大值抑制可能准确的检测框的得分较低且两框的 IOU 大于设定的阈值,因此会被过滤掉,导致只检测出一个目标,显然这样的算法设计是不合理的。NMS 直接粗暴地将和得分最大的 box 的 IOU 大于阈值的 box 的得分置零,而 Soft-NMS 用一个稍微小一点的分数替代原有的分数,而非直接粗暴地置零。该算法是在非极大值抑制算法上作出改进,设置了衰减函数,使得重叠部分相邻检测框的分数不会置零。如果检测框与 M 重叠部分较大,则得分低;如果检测框与 M 重叠部分较小,则此部分原始得分不会发生改变。在标准数据集 PASCAL VOC 2016 和 MS-COCO 2018 等标准数据集上,加入 Soft-NMS

的目标检测算法在目标有重叠部分时检测平均准确率显著提升。

相对于传统 NMS 算法, Soft-NMS 算法能够提高模型检测与分类效果, 且不增加额外计算量。对重叠和小目标有较好检测效果, 能很好地融入本文算法, 降低漏检率, 加快运算速度。Soft-NMS 算法流程如下。

输入: 检测框集合 $B = \{b_1, b_2, \dots, b_N\}$, 相应的检测分数 $S = \{s_1, s_2, \dots, s_N\}$, NMS 阈值 N_t

```

begin
 $D \leftarrow \{ \}$ 
while  $B \neq \text{empty}$  do
 $m \leftarrow \arg \max S$ 
 $M \leftarrow b_m$ 
 $D \leftarrow D \cup M$ ;  $B \leftarrow B - M$ 
for  $b_i$  in  $B$  do
 $s_i \leftarrow s_i * f(\text{iou}(M, b_i))$ 
end
end
return  $D, S$ 
end

```

对于传统的非极大值抑制算法, 当前检测框和最高得分检测框的 IOU 大于阈值时, 直接将该检测框的得分置零, 这将导致重叠区域较大的目标框被漏检:

$$s_i = \begin{cases} s_i, & \text{IOU}(M, b_i) < N_t \\ 0, & \text{IOU}(M, b_i) \geq N_t \end{cases} \quad (19)$$

式中: s_i 为当前检测框的得分; N_t 为 IOU 的阈值; M 为得分最高的检测框。

而 Soft-NMS 算法将当前检测框得分乘以一个权重函数, 该函数会衰减与最高得分检测框 M 有重叠的相邻检测框分数, 越是与 M 框高度重叠的检测框, 其得分衰减越严重, Soft-NMS 的改进有两种形式, 一种是线性加权:

$$s_i = \begin{cases} s_i, & \text{IOU}(M, b_i) < N_t \\ s_i(1 - \text{IOU}(M, b_i)), & \text{IOU}(M, b_i) \geq N_t \end{cases} \quad (20)$$

另一种是高斯函数作为权重函数, 从而修改其删除检测框的规则, 其高斯权重函数为

$$s_i = s_i e^{-\frac{\text{IOU}(M, b_i)^2}{\sigma}} \quad (21)$$

1.4 融合 KL-LOSS 与 Soft-NMS 算法

使用 KL-LOSS 训练后, 可以得到预测位置的方差, 根据相邻目标框的学习方差对候选目标框的位置进行方差决策, 通过融合 KL-LOSS 与 Soft-NMS 算法得到 var voting 算法。

在 Soft-NMS 循环内, 投票选择目标框的位置, 在选择最高的目标得分检测框 b 后即 $b = \{x_1, y_1, x_2, y_2, s, s_{x_1}, s_{y_1}, s_{x_2}, s_{y_2}\}$, 根据自身及其相邻边检测框计算出新的位置, 为较近且不确定度较低的检测框分配更高的权重。设 x 为预测框的坐标, x_i 为第 i 个预测框的坐标, 则新的坐标计算如下:

$$p_i = e^{-\left(1 - \text{IOU}(b_i, b)\right)^2 / \sigma_i} \quad (22)$$

式中: σ_i 为可调参数。

$$x = \frac{\sum_i p_i x_i / \sigma_{x,i}^2}{\sum_i p_i / \sigma_{x,i}^2}, \quad \text{s.t. } \text{IOU}(b_i, b) > 0 \quad (23)$$

在算法投票选择目标框的过程中, 两种类型的相邻边界框会得到更低的权重: 1) 有大方差的框; 2) 候选框的 IOU 小的框。分类得分并不参与投票选择, 因为分类得分较低的框可能具有更高的坐标置信度。

var voting 算法流程如下。

输入: 检测框集合 $B = \{b_1, b_2, \dots, b_N\}$, 相应的检测分数 $S = \{s_1, s_2, \dots, s_N\}$, NMS 阈值 N_t 相应的方差集合 $C = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2\}$, 方差决策可调参数 σ_i 。

```

begin
 $D \leftarrow \{ \}$ 
 $T \leftarrow B$ 
while  $T \neq \text{empty}$  do
 $m \leftarrow \arg \max S$ 
 $T \leftarrow T - b_m$ 
 $S \leftarrow S * f(\text{IOU}(b_m, T))$ 
 $idx \leftarrow \text{IOU}(b_m, B) > 0$ 
 $p \leftarrow \exp\left(-\left(1 - \text{IOU}(b_m, B[idx])\right)^2 / s_i\right)$ 
 $D \leftarrow D \cup b_m$ 
end while
return  $D, S$ 
end

```

1.5 检测层前融合 SKNET 算法

1.5.1 SKNET 概述

对动物的初级视觉皮层神经元的局部感受野的研究激发了卷积神经网络的构建,并继续激励着现代卷积神经网络(Convolutional Neural Networks, CNN)结构的构建。例如,在视觉皮层中,同一区域神经元的局部感受野大小是不同的,这使得神经元能够在同一处理阶段收集多尺度的空间信息。这种机制在最近的 CNN 中得到了广泛的应用。

然而,在神经网络的设计中,神经元的一些其他局部感受野特性并没有得到重视,其中之一就是局部感受野大小的自适应变化。大量实验表明,视觉皮层神经元的局部感受野大小不是固定的,而是受刺激的调节。但是,这种性质在构建深度学习模型时并没有受到太多的关注。在同一层中包含多尺度信息的模型,如 InceptionNet^{page90},由于下一层卷积层线性地聚集了来自不同分支的多尺度信息,因此具有根据输入内容调整下一卷积层神经元感受野大小的内在机制,但这种线性聚集方法可能不足以提供神经元强大的适应能力。

SKNET 算法是一种非线性的多核信息聚合方法来实现神经元的自适应感受野大小的算法,可以很方便地嵌入现在的网络结构(如 ResNet、Inception、ShuffleNet)中,实现精度的提升。SKNet 使用了 2 个思路提高精度^[16]:1) 很多网络使用了各种方法来降低计算量,比如 ResNeXt,计算量大大减少,精度却略有提升。那么如果不牺牲那么多计算量,能否精度提高一些呢? 比如使用大一点的 Kernel,如 5×5 的卷积提升精度。2) 采用注意力机制。加了上面 2 个操作后,显然计算量会增大,于是作者再增加了一个分组卷积来做换取。

算法引入了一种“选择性核”(sk)卷积,它由 3 个算子组成:分裂算子(Split)、融合算子(Fuse)和选择算子(Select)。分裂算子产生多条不同核大小的路径,对应于不同的神经元感受野大小,融合算子将来自多条路径的信息进行组合和聚合,以获得选择权重全局综合表示,选择算子根据选择权重聚合不同大小内核的特征映射。Sk 卷积采用分组卷积,其在计算上是轻量的,只会稍微增加参数与计算成本。

同时算法采用注意力机制的好处已经在一系列

神经网络任务中得到了体现,从自然语言处理中的神经机器翻译到图像理解中的图像字幕。注意力机制加强了对网络任务用处更大的特征表达,同时抑制不太有用的特征表达,通过引入注意机制来明确关注神经元的自适应感受野大小。

1.5.2 SKNET 基本原理与融合方法

SKNET 中,神经元的感受野大小可以在前向传播过程中自适应地改变。为了使神经元能够自适应地调整它们的感受野大小,网络能够自适应选择感受野,即在不同核大小的多个核之间进行“选择性”卷积即 sk 卷积。具体地说,通过拆分、融合和选择 3 个步骤来实现 sk 卷积,如图 3 展示了具有不同的内核大小的 2 个分支的情况。

分裂算子是指对于任意给定的特征图 $X \in \mathbf{R}^{H' \times W' \times C'}$,首先进行 $\tilde{F}: X \rightarrow \tilde{U} \in \mathbf{R}^{H \times W \times C}$ 和 $\hat{F}: X \rightarrow \hat{U} \in \mathbf{R}^{H \times W \times C}$ 两个变换,卷积核大小分别为 3×3 和 5×5 。其中, $\tilde{F}$ 和 $\hat{F}$ 都是由有效的分组深度可分离卷积、批处理规范化和 relu 函数按顺序组成的特征提取器。为了进一步提高效率,可将传统的 5×5 卷积核替换为 3×3 卷积核和 2 倍的扩展卷积。

融合算子:如前面所述,目标是使神经元能够根据输入数据自适应地调整其感受野大小。其基本思想是利用阀门来控制来自多个分支的信息流,这些分支携带不同尺度的信息进入下一层的神经元。

为实现这一目标,阀门需要整合来自所有通道的信息,首先通过每个通道的像素级求和来融合自多个通道的特征:

$$U = \tilde{U} + \hat{U} \quad (24)$$

然后,通过简单的全局平均池化来嵌入全局特征,生成通道统计量。具体地说, s 的第 c 个元素是通过在空间尺 $H \times W$ 上压缩 U 来计算:

$$s_c = F_{sp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (25)$$

最后,建立一个紧凑的特征映射,使之能够进行精确的自适应选择,这里通过一个简单的全连接层(FC)实现,使用全连接层降低维度以获得更好的效率:

$$z = F_{fc}(s) = \delta(B(Ws)) \quad (26)$$

式中: δ 为 relu 函数; B 为批标准化; $W \in \mathbf{R}^{d \times c}$ 。

d 是对模型效率的影响,并使用一个比率 r 来控制 d 的值:

$$d = \max(C/r, L) \quad (27)$$

式中: L 为 d 的最小值。

选择算子: softmax 回归用于自适应地选择不同通道间的特征重要性, 该特征重要性通过压缩特征 z 计算得到。具体地说, 使用 softmax 回归得到如下注意力权重:

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}} \quad (28)$$

$$b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \quad (29)$$

式中: $A \in \mathbb{R}^{c \times d}$ 和 a 为 U_1 的 softmax 注意力向量; $B \in \mathbb{R}^{c \times d}$ 和 b 为 U_2 的 softmax 注意力向量; $A \in \mathbb{R}^{c \times d}$ 式中为 A 的第 c 行; a_c 表示 a 的第 c 个元素, 并且有

$$a_c + b_c = 1 \quad (30)$$

最后的特征映射 V 是通过不同核上的注意力权重得到

$$V_c = a_c \cdot \tilde{U}_c + b_c \cdot \tilde{U}_c \quad \text{s.t. } a_c + b_c = 1 \quad (31)$$

式中:

$$V = [V_1, V_2, \dots, V_c], V_c \in \mathbb{R}^{H \times W} \quad (32)$$

如图 3 所示, SKNET 模块可以对特征通道进行重要性加权, 使最终得到的特征图更加关注对检测识别等任务有用的特征通道, 并且该模块会自动选择调整最优的卷积算子, 进而在理论上改善检测识别等任务的性能。因此, 在改进的 YOLOV3 特征提取层后加入 SKNET 模块, 原因是在对图像特征提取后, 紧接着对通道进行重要性加权, 然后送入目标检测层检测。以此训练的网络在特征提取后的特征将是更加关注训练目标的加权特征, 使用加权特征进行目标检测理论上将提高目标的检测效率, 从而改善检测识别性能。

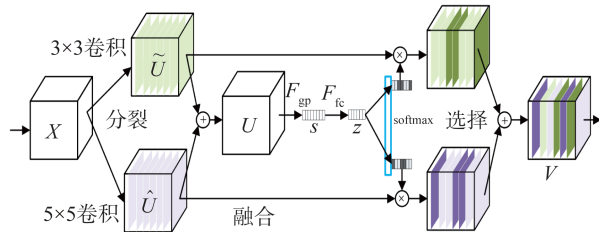


图 3 SKNET 结构

Fig.3 Structure of the SKNET

1.6 融合网络层

通过融合 KL-LOSS 得到新网络的新损失函

数, 并且使用 Soft-NMS 算法与通过 KL-LOSS 训练得到的方差结合可以得到更加准确的目标位置, 改进特征提取层与融合 SK-NET 和 KL-LOSS 的新网络称为 KS-YOLO, 如图 4 所示。

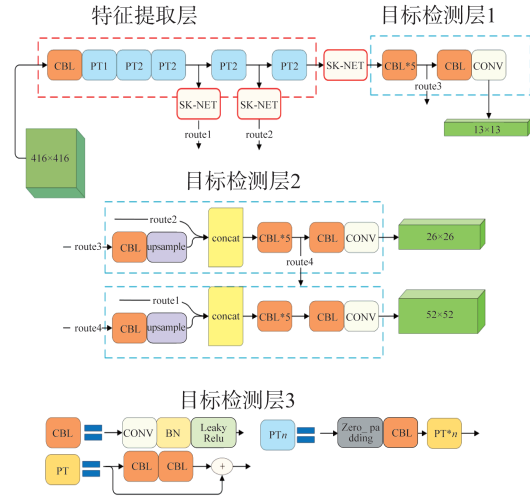


图 4 KS-YOLO 结构

Fig.4 Structure diagram of the KS-YOLO

图中, Conv 为卷积, BN 为批标准化, LeakRelu 为激活函数, Concat 为张量合并操作。对于新网络的预测框预测结果变为 $b = \{b_x, b_y, b_w, b_h, s, \sigma_{b_x}, \sigma_{b_y}, \sigma_{b_w}, \sigma_{b_h}\}$, 令 $a = \log(\sigma^2)$, 则算法实际预测 a 值。最终目标的预测由 var-voting 算法给出, 如图 5 所示。

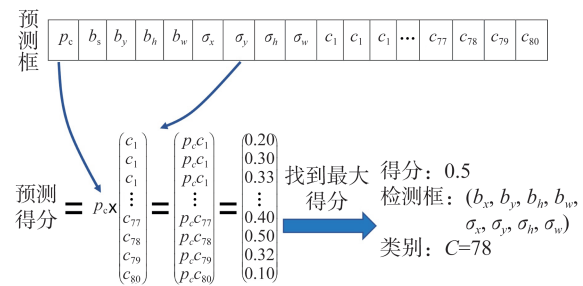


图 5 KS-YOLO 位置及其标准差与类别预测

Fig.5 Location, standard deviation, and category prediction of KS-YOLO

图中, P_c 为当前预测框有无目标的概率值, C_i , $i \in (1, 2, \dots, 80)$ 为当前预测框属于某一类别的概率, $\sigma_x, \sigma_y, \sigma_h, \sigma_w$ 为边界框的中心以及宽高值, 最后给出融合改进后的新网络的训练用反向传播目标损失函数为

$$\begin{aligned}
\text{loss}(\text{object}) = & \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} \left[\begin{cases} \frac{e^{-a_x} (x_i - \hat{x}_i)^2}{2} + \frac{1}{2} a_x & 0 < |x_i - \hat{x}_i| < 1 \\ e^{-a_x} \left(|x_i - \hat{x}_i| - \frac{1}{2} \right) + \frac{1}{2} a_x & |x_i - \hat{x}_i| > 1 \end{cases} \right] + \\
& \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} \left[\begin{cases} \frac{e^{-a_y} (y_i - \hat{y}_i)^2}{2} + \frac{1}{2} a_y & 0 < |y_i - \hat{y}_i| < 1 \\ e^{-a_y} \left(|y_i - \hat{y}_i| - \frac{1}{2} \right) + \frac{1}{2} a_y & |y_i - \hat{y}_i| > 1 \end{cases} \right] + \\
& \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_i \times h_i) \left[\begin{cases} \frac{e^{-a_w} (w_i - \hat{w}_i)^2}{2} + \frac{1}{2} a_w & 0 < |w_i - \hat{w}_i| < 1 \\ e^{-a_w} \left(|w_i - \hat{w}_i| - \frac{1}{2} \right) + \frac{1}{2} a_w & |w_i - \hat{w}_i| > 1 \end{cases} \right] + \\
& \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_i \times h_i) \left[\begin{cases} \frac{e^{-a_h} (h_i - \hat{h}_i)^2}{2} + \frac{1}{2} a_h & 0 < |h_i - \hat{h}_i| < 1 \\ e^{-a_h} \left(|h_i - \hat{h}_i| - \frac{1}{2} \right) + \frac{1}{2} a_h & |h_i - \hat{h}_i| > 1 \end{cases} \right] - \\
& \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - \hat{C}_i) \right] - \\
& \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{noobj}} \left[\hat{C}_i \log(C_i) + (1 - \hat{C}_i) \log(1 - \hat{C}_i) \right] - \\
& \sum_{i=0}^{K \times K} \sum_{c \in \text{classes}} I_{ij}^{\text{obj}} \left[\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c)) \right] \quad (33)
\end{aligned}$$

新网络 KS-YOLO 前向传播算法步骤如下。

输入: 图像 S

输出: 目标位置边界框 D , 目标类别得分 C

KS-YOLO 的 Backbone 网络特征提取层提取图像特征得到 3 个特征图 map1、map2 与 map3

利用特征融合层上采样(upsample)map1 并与 map2 融合(concat)计算出新特征图 map_2

利用特征融合层上采样(upsample)map2 并与 map3 融合(concat)计算出新特征图 map_3

利用融合 SK-NET 的新检测层检测 map1、map_2 与 map_3 计算出边界框集合 B 与目标置信度集合 S 及条件类别概率 $P(\text{class}|S)$ 集合

利用 var-voting 算法从 B 、 S 中筛选出最佳边界框 D 及其得分 S

计算目标类别得分 $C = P(\text{class}|S) * S$

返回 D , C

2 结果

定量上, 评价模型检测识别性能有以下指标即精准度(Precision, P)、召回率(Recall, R)、准确率

(Accuracy, Acc)、平均精准度(Average Precision, AP)、均值平均精准度(mean Average Precision, mAP)、交并比(Intersection over Union, IOU)、检测速度 FPS 或检测预测时间(Inference time)。

如果设测试集图像中目标为正样本(Positive), 背景为负样本(Negative)。网络将预测正样本与负样本, 设正确预测正样本为目标的个数为 T_{positive} , 正确预测负样本为背景的个数为 T_{negative} , 错误预测负样本为目标的个数为 F_{positive} , 错误预测正样本为背景的个数为 F_{negative} 。

则 P 又叫查准率, 定义为识别成目标的个数中正确的比例:

$$P = \frac{T_{\text{positive}}}{T_{\text{positive}} + F_{\text{positive}}} \quad (34)$$

R 又叫查全率, 定义为正确预测为目标的个数占测试集总的该类目标个数的比例:

$$R = \frac{T_{\text{positive}}}{T_{\text{positive}} + F_{\text{negative}}} \quad (35)$$

Acc 定义为正确预测出正样本与负样本的个数占测试集总正负样本个数的比例:

$$A_{acc} = \frac{T_{positive} + T_{negative}}{T_{positive} + T_{negative} + F_{positive} + F_{negative}} \quad (36)$$

AP 定义为每一类的精准度的均值:

$$A_{AP} = \int_0^1 P(R) dR \quad (37)$$

mAP 定义为所有类的平均精准度的均值:

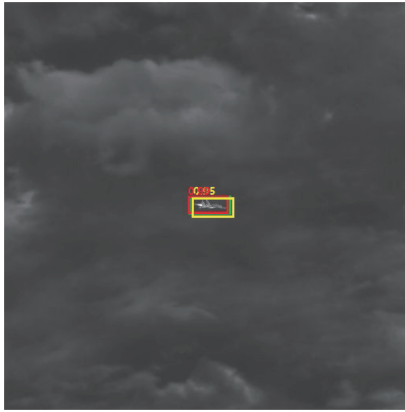
$$m_{AP} = \frac{\sum_{i=1}^C A_{AP}(i)}{C} \quad (38)$$

IOU 表征了目标检测的准确率, 定义为目标检测的区域 R_p 与实际目标区域 R_g 的交集与并集之比:

$$I_{IOU} = \frac{R_p \cap R_g}{R_p \cup R_g} \quad (39)$$

由于本文针对的是典型的空战红外目标, 如第三代战斗机、教练机等红外目标, 所以使用 AP50 及算法预测的时间 2 个指标, 其中指标 AP50 因其蕴含了 P 、 R 及交并比指标, 准确率指标又可通过 P 、 R 指标得到, 所以 AP50 指标已经成为深度学习目标检测识别领域的一个综合性指标。

以绿色框表示目标本身真实位置, 如图 6~图 7 所示, 黄色框表示原始 YOLOV3 算法的检测识别结果, 红色框表示改进的 KS-YOLO 算法的检测识别结果。图 6 是表示实验室仿真图像初始 YOLOV3 和改进的 KS-YOLO 算法对比结果; 图 7 是实拍空中红外目标图像初始 YOLOV3 和改进的 KS-YOLO 算法对比结果。左图与右图是从不同方位视角下对红外目标图像进行检测。



(a) 近距离红外目标图像检测



(b) 远距离红外目标图像检测

图 6 实验室仿真图像初始 YOLOV3 和改进的 KS-YOLO 算法对比结果

Fig.6 The comparison results of the initial YOLOV3 and the improved KS-YOLO algorithm in laboratory simulation images



(a) 下视红外目标图像检测



(b) 侧向红外目标图像检测

图 7 实拍空中红外目标图像初始 YOLOV3 和改进的 KS-YOLO 算法对比结果

Fig.7 Comparison results of initial YOLOV3 and improved KS-YOLO algorithm for aerial infrared target images

从实验结果看到改进的算法相对于原始算法具有更高的位置准确率,这是因为算法融合了 KL-LOSS,网络预测了位置的不准确度自动校正坐标位置;同时,还可以看到改进的算法比原始算法有更高的目标分类识别得分,这是因为融合的 SKNET 模块对特征通道进行了加权,并自动选择了对目标任务更有用的卷积算子,使网络更加关注对目标有用的特征,从而提升了目标识别的准确度。

统计两个算法的召回率与准确性,可以计算出评价算法综合性能的 AP50 指标。实验统计了图像在 800×800 、 736×736 、 608×608 、 416×416 、 320×320 的尺度上的性能,得到性能对比图如图 8 所示,具体性能对比值见表 1。

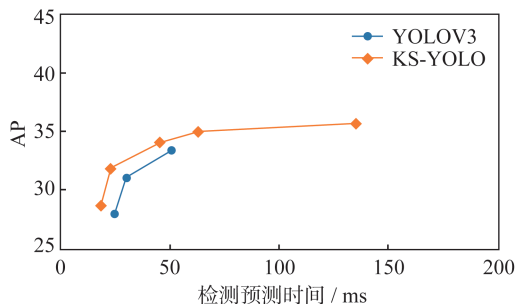


图 8 目标域数据集上速度与综合性能对比图

Fig.8 Speed versus comprehensive performance on the target domain dataset

表 1 目标域数据集上速度与综合性能对比

Tab.1 Comparison of speed and comprehensive performance on the target domain dataset

方法	AP50	时间/ms
YOLOV3-320	28.0	24
YOLOV3-416	31.1	30
YOLOV3-608	33.5	51
KS-YOLO-320	28.8	19
KS-YOLO-416	31.9	23
KS-YOLO-608	34.1	45
KS-YOLO-736	35.0	63
KS-YOLO-800	35.7	135

3 结束语

红外目标的检测识别在侦查、红外预警及许多领域都有着重要的意义。本文针对空战典型红外目标,使用卷积神经网络实现目标的检测与识别,并针对红外图像目标的特性进行了基础检测识别

网络的改进。针对红外图像目标特征量不足检测困难、训练网络时模型容易过拟合及检测定位不准确的问题,在原 YOLOV3 的基础上改进了特征提取网络,训练时融合 KL-LOSS 损失并结合 Soft-NMS 算法提升了目标的检测定位精度,融合 SKNET 模块提升了网络的目标检测准确性。总体来说改进的算法提升了网络对红外目标的检测精度与识别效率。

本文使用深度学习神经网络模型为基础网络,其中还有很多方面可以进行深入研究:1) 对于原始 YOLOV3 网络提取层的改进只是简单地减少残差层的数量,目的是为了避免模型复杂导致过拟合,对于改进特征提取层的卷积核与卷积方式也是非常好的思路。2) 在算法实时性方面,YOLOV3 目标检测识别算法,在实时性上优于许多同期的检测算法,后续研究工作可进一步在实时性表现更好的基础网络上应用迁移学习技术,以减少人工标注时间,提升模型的检测识别性能。

参考文献

- [1] 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. Washington D. C., USA: IEEE Press, 2014: 580-587.
- [2] 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.
- [3] GIRSHICK R. Fast R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems. Washington D. C., USA: IEEE Press, 2015: 91-99.
- [5] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 6154-6162.
- [6] 王瑶,胥辉旗,姜义,等.基于深度学习的舰船目标检测技术发展综述[J].飞航导弹,2021(2):76-81.

- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// European Conference on Computer Vision. Berlin, Germany: Springer, 2016:21-37.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 7263-7271.
- [9] REDMON J, FARHADI A. YOLOV3: an incremental improvement[J]. ArXiv Preprint, ArXiv: 1804.02767, 2018.
- [10] 李向荣, 焦学军, 刘冬冬. 基于 YOLOv3 算法的车牌识别优化算法[J]. 电子设计工程, 2021, 29(22): 166-170.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 2117-2125.
- [12] HE Y, ZHU C, WANG J, et al. Bounding box regression with uncertainty for accurate object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 2888-2897.
- [13] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS: improving object detection with one line of code [C]// Proceedings of the IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 5561-5569.
- [14] LI X, WANG W, HU X, et al. Selective kernel networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 510-519.
- [15] HE Y, ZHU C, WANG J, et al. Bounding box regression with uncertainty for accurate object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 2888-2897.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7132-7141.

(上接第 46 页)

- [5] 党庆庆, 金磊, 徐世杰. 基于极点配置的空间站角动量管理[J]. 北京航空航天大学学报, 2016, 42(8): 1715-1723.
- [6] 薛锐, 崔雁, 王晓磊. 嫦娥五号飞行试验器服务舱环月期间的角动量管理[J]. 空间控制技术与应用, 2016, 42(4): 53-56.
- [7] 王淑一, 雷拥军, 王新民, 等. 轨控大干扰下的系统角动量管理[J]. 空间控制技术与应用, 2018, 44(2): 23-28.
- [8] 马传令, 刘勇, 陈明, 等. CE-4 中继卫星使命轨道维持与动量轮卸载联合控制方法[J]. 宇航学报, 2020, 41(4): 389-397.
- [9] 马雪, 韩冬, 汤亮. 电推进卫星角动量卸载研究[J]. 中国空间科学技术, 2016, 36(1): 70-76.
- [10] 张学钢, 朱振才, 陈宏宇. 基于气动力矩和模型预测的 LEO 卫星角动量管理算法[J]. 控制与决策, 2019, 34(11): 183-189.
- [11] 李连军, 戴金海. 双翼箱式航天器太阳光压干扰力矩模型[J]. 航天控制, 2005, 23(4): 13-17.
- [12] 刘墩, 赵志萍. 卫星太阳光压力矩计算中有效作用面积的计算[J]. 南京理工大学学报, 2007, 31(6): 685-688.
- [13] 刘善伍, 万松, 容建刚. 航天器空间环境干扰力矩分析与仿真研究[J]. 航天控制, 2015, 33(2): 78-81.
- [14] 洪振强, 宋效正, 吕旺等. 大型三轴气浮台转动惯量和干扰力矩高精度联合辨识技术[J]. 航天器环境工程, 2017, 34(1): 28-34.
- [15] 周杨森. 导航卫星太阳光压建模方法与模型特性分析[D]. 长沙: 湖南大学, 2016.