

# 一种基于注意力模型的无锚框交通标志识别算法

褚晶辉, 黄浩, 吕卫\*

天津大学电气自动化与信息工程学院, 天津 300072

**摘要** 针对交通标志只在图像中占极小的区域且难以准确识别的问题, 提出一种基于注意力模型的无锚框交通标志识别算法, 利用密集连接网络 DenseNet-121 作为骨干网络并对特征进行提取。为了解决小型交通标志准确率低的问题, 在骨干网络中加入注意力模型, 可以对特征图进行空间和通道上的自适应调整, 通过加强或抑制特征图中元素的权重可以提升对小型交通标志的识别性能。为了减小编码路径与解码路径间的语义鸿沟, 引入残差网络的连接方式并提出一种语义连接路径。为了解决锚框中正负样本不均衡的问题, 采用无锚框的检测方式可以定位交通标志的中心点、回归边界框的位置与尺寸信息。对所提算法在 TT100K 数据集上进行验证, 实验结果证明所提算法具有优越性。

**关键词** 图像处理; 深度学习; 交通标志识别; 注意力模型; 无锚框

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.1610020

## Anchor-Free Traffic Sign Recognition Algorithm Based on Attention Model

Chu Jinghui, Huang Hao, Lü Wei\*

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

**Abstract** Aiming at the problem that traffic signs only occupy a very small area in the image and are difficult to accurately identify, an anchorless frame traffic sign recognition algorithm based on the attention model is proposed. The densely connected network DenseNet-121 is used as the backbone network and features are extracted. In order to solve the problem of low accuracy of small traffic signs, an attention model is added to the backbone network to make adaptive adjustments to the space and channel of the feature map. The recognition performance of small traffic signs can be improved by strengthening or suppressing the weight of elements in the feature map. In order to reduce the semantic gap between the encoding path and the decoding path, the residual network connection method is introduced and a semantic connection path is proposed. In order to solve the problem of the imbalance of positive and negative samples in the anchor frame, the detection method without anchor frame can locate the center point of the traffic sign to regression the position and size information of the boundary box. The proposed algorithm is verified on the TT100K dataset, and the experimental results prove the superiority of the proposed algorithm.

**Key words** image processing; deep learning; traffic sign recognition; attention model; anchor-free

**OCIS codes** 100.2000; 200.4260; 150.1135

## 1 引言

根据世界卫生组织统计, 全球每年有超过一百万人因交通事故死亡, 而且该事故已经成为青少年

以及儿童的首要死亡原因之一。为了减少交通事故的发生, 各国的科研院校和高科技企业都投入了大量的资金来研发先进驾驶辅助系统, 而交通标志的识别是驾驶辅助系统中非常重要的一项

收稿日期: 2020-08-18; 修回日期: 2020-09-07; 录用日期: 2020-09-30

通信作者: \*luwei@tju.edu.cn

研究内容。

传统的交通标志识别算法利用交通标志特殊的形状和颜色对其进行分类<sup>[1]</sup>。Madani 等<sup>[2]</sup>提出了一种基于监督距离分类器的颜色分割算法,使用 6 种随机值代表 HSV(Hue, Saturation, Value)颜色空间<sup>[3]</sup>中的主要颜色并进行训练,该算法能够在不同的天气条件下对 6 个颜色区域中的标志进行分割。随着深度学习的发展,Faster R-CNN(Faster Region-Convolutional Neural Network)<sup>[4]</sup>和 YOLO(You Only Look Once)<sup>[5]</sup>等各种卷积神经网络模型均被应用在交通标志识别的领域,而且取得了出色的效果。此外,许多学者都对经典的卷积神经网络进行了改进,并应用在交通标志识别的领域。张毅等<sup>[6]</sup>提出了一种基于改进 Faster R-CNN 的交通标志识别算法,通过拟合交通标志特征来设计锚框以获取更好的推荐区域,从而进一步降低误检率与漏检率。童英等<sup>[7]</sup>对 YOLOv2 网络模型进行了改进,并使用标准的边界框宽度和高度来计算损失函数。常用的目标检测算法都将锚框作为先验框,一些学者对锚框的生成方式进行了改进并提出了无锚框的方法。CenterNet 模型<sup>[8]</sup>将目标建模为其边界框的中心点,检测目标的过程中利用关键点估计来找到中心点,再回归出其余目标的属性,例如

边界框的大小、3D 位置、方向甚至姿势等。虽然这些方法在复杂的现实场景中均能准确识别近距离的交通标志,但对于远距离和小尺寸的交通标志,识别率仍旧较低,如何准确识别小型目标仍然是交通标志识别领域的一大难点。

本文提出一种基于注意力模型的无锚框交通标志识别网络,利用密集连接网络(DenseNet-121)<sup>[9]</sup>作为骨干网络并用来对特征进行提取,特征的重复利用可缓解梯度消失的现象,有效加强了特征传播。网络整体采用编码器-解码器的结构,在编码路径中加入注意力模型可对提取到的特征图进行空间和通道上的自适应调整,用来获取有益于交通标志分类的信息。本文提出一种语义连接路径,可以将编码路径的浅层语义信息与解码路径的深层语义信息进行融合。网络采用无锚框的检测方式,通过定位交通标志的中心点来回归边界框的位置与尺寸信息,有效解决锚框所带来的正负样本不均衡问题。

## 2 算法原理

### 2.1 卷积神经网络结构

基于注意力模型的无锚框卷积神经网络(AAFCNN)模型的结构如图 1 所示,该模型采用编码器-解码器的结构,整个网络大致分为编码路径与

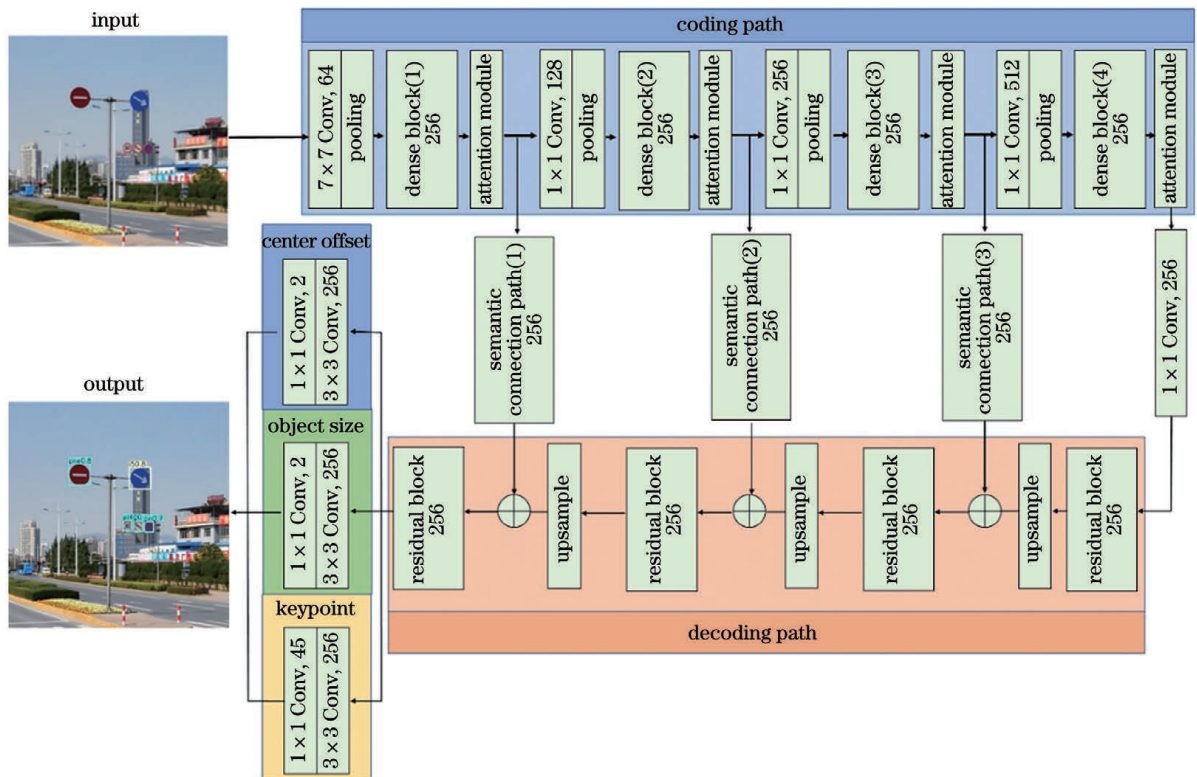


图 1 AAFCNN 模型的结构

Fig. 1 Structure of AAFCNN model

解码路径,其中 $\oplus$ 为特征图对应元素相加。编码路径采用密集连接网络(DenseNet-121)作为骨干网络,可以使其充分提取图像中的浅层语义信息;解码路径由残差连接模块与上采样层组成,将编码器网络输出的低分辨率特征图映射回输入图像的尺寸,从而逐步修复图像的细节信息。在编码路径的密集连接模块之间加入注意力模型,能够有效提取对交通标志分类有益的信息。编码路径与解码路径之间存在三条语义连接路径,编码路径提取到的浅层语义特征会经过语义连接路径并与解码路径的深层语义特征进行融合,能够帮助解码器修复交通标志的轮廓信息。

受到 CenterNet 模型的启发,AAFCNN 模型采用中心点预测的方式来识别交通标志。输入的街景图像经过编码器-解码器结构中的全卷积网络后,得到包含交通标志语义信息的热力图。使用网络模型对热力图进行  $3 \times 3$  卷积与  $1 \times 1$  卷积操作后,得到关键点热力图、中心点偏移量与目标尺寸。其中关键点热力图的峰值点就是交通标志的中心点,检测与识别交通标志的过程就是提取峰值点的过程。首先针对每个类别分别提取关键点热力图中的峰值,然后检测出所有大于或等于其相邻 8 个响应点的峰值点并保留前 100 个点。令  $\hat{S}_k$  表示  $k$  个类别交通标志中所检测到的  $n$  个中心点  $\hat{S} = \{(\hat{x}_j, \hat{y}_j)\}_{j=1}^n$  的集合,其中  $(\hat{x}_j, \hat{y}_j)$  为检测到第  $j$  个中心点的坐标,得到的交通标志边界框坐标为

$$\begin{aligned} &(\hat{x}_j + \partial\hat{x}_j - \hat{w}_j/2, \hat{y}_j + \partial\hat{y}_j - \hat{h}_j/2, \\ &\hat{x}_j + \partial\hat{x}_j + \hat{w}_j/2, \hat{y}_j + \partial\hat{y}_j + \hat{h}_j/2), \end{aligned} \quad (1)$$

式中:  $(\partial\hat{x}_j, \partial\hat{y}_j) = \hat{O}_{(\hat{x}_j, \hat{y}_j)}$  为中心点偏移量;  $(\hat{w}_j, \hat{h}_j) = \hat{S}_{(\hat{x}_j, \hat{y}_j)}$  为交通标志尺寸的预测值。

### 2.2 语义连接路径

对于一些经典的编码器-解码器结构,如 U-Net<sup>[10]</sup>和 V-Net<sup>[11]</sup>等均会在编码路径的池化层和解码路径的反卷积层之间加入跳跃连接路径。因此,编码路径中提取到的多尺度特征便可以传播到解码路径,编码路径中经过池化操作后丢失的空间信息得到了恢复。虽然保留了编码路径中丢失的空间特征,但跳跃连接路径仍然存在缺陷。编码路径的特征是在浅层网络中所提取到的低层次语义特征,而解码路径中的特征是深层神经网络所提取到的高层次语义特征,二者之间存在语义鸿沟<sup>[12]</sup>。将这两组不兼容的特征直接进行融合,这会影响神经网络的学习,而且会对交通标志的识别产生不利影响。

为了减小编码路径与解码路径之间的语义鸿沟,可以在跳跃连接路径中加入一些卷积层,用来对编码路径传播过来的特征进行更多处理。此外,随着卷积神经网络的加深,编码路径中的特征图与解码路径中的特征图之间的语义鸿沟逐渐减小。由于语义鸿沟的大小与卷积层的数量呈正相关,所以每个跳跃连接路径中加入的卷积层数量也应该不同。综上,提出语义连接支路,结构如图 2 所示。利用卷积核尺寸为  $1 \times 1$  与  $3 \times 3$  的卷积层加深编码路径传入特征的语义层次,同时引入残差连接可以使神经网络更易于学习<sup>[13]</sup>。

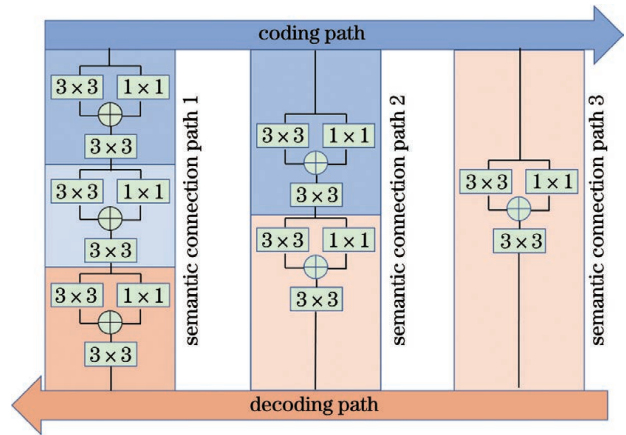


图 2 语义连接路径的结构

Fig. 2 Structure of semantic connection path

### 2.3 注意力模型

注意力机制是一种模拟人类大脑处理视觉任务的机制,其可以引导神经网络关注交通标志区域,排除交通场景中其他背景特征的干扰。受到文献<sup>[14]</sup>的启发,在解码路径中加入注意力模型,该注意力模型由通道注意力模块与空间注意力模块构成。

通道注意力模块利用不同特征通道间的关系来生成通道注意力图,使其更关注输入特征图中不同通道信息的重要性。通道注意力模块的结构如图 3(a)所示,其中 FC 表示全连接。首先对输入的特征图  $Y \in \mathbb{R}^{K \times W \times H}$  进行全局最大池化和全局平均池化等操作,然后对特征图中的元素进行相加,可以扩张得到包含输入特征图各个通道权重信息的通道注意力图  $Y_{CA} \in \mathbb{R}^{K \times W \times H}$ ,其中  $K$  为注意力图的通道数, $W$  和  $H$  分别为注意力图的宽和高。特征图的不同通道对交通标志识别任务所做的贡献不同,将输入特征图与通道注意力图的对应元素相乘,可以根据通道的重要程度来对特征图进行自适应调整。

空间注意力模块利用特征之间的空间关系来

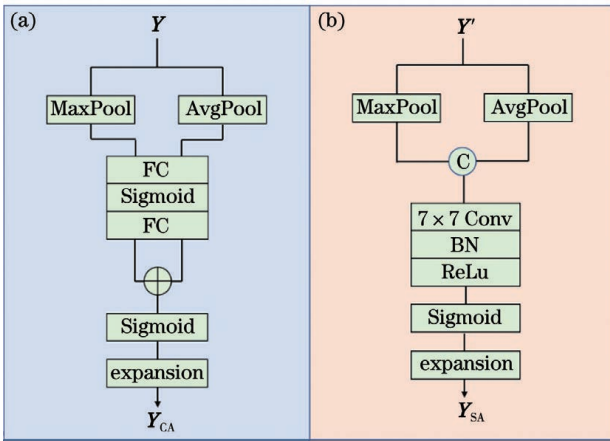


图 3 不同的注意力模块。(a)通道注意力模块；  
(b)空间注意力模块

Fig. 3 Different attention modules. (a) Channel attention module; (b) spatial attention module

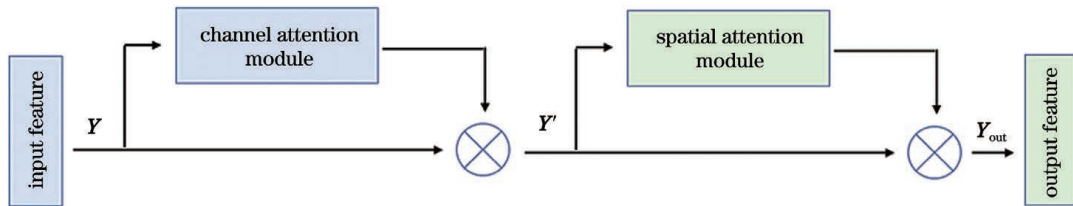


图 4 注意力模型的结构

Fig. 4 Structure of attention model

$$Y_{out} = Y \otimes Y_{CA} \otimes Y_{SA}, \quad (2)$$

式中： $\otimes$ 为对应元素相乘符号。

### 2.4 无锚框预测方式

目标检测算法中通常使用滑动窗口的方式来产生大量的锚框,然后以锚框作为先验框生成候选区域,但锚框中负样本的数量远大于正样本,这为模型的训练带来困难。为了避免此类问题,实验采用无锚框的方法对标志进行识别,首先确定交通标志的中心点与类别信息,再回归出中心点的偏移量与边界框的尺寸。

采用基于编码器-解码器结构的全卷积网络处理输入的街景图像,目的是获得  $C$  通道的关键点热力图,其中  $C$  表示关键点热力图的种类,即数据集中的 45 类交通标志。对于真值图中的每个关键点,

生成空间注意力图,使其更关注输入特征图中不同位置信息的重要性。空间注意力模块的结构如图 3(b)所示,其中 $\odot$ 为特征图拼接操作,BN 表示批量归一化。首先对输入特征图 $Y' \in \mathbb{R}^{K \times W \times H}$  基于通道维度进行全局最大池化与全局平均池化操作后进行通道拼接,然后经过  $7 \times 7$  的卷积处理,可以扩张得到包含特征图不同坐标权重信息的空间注意力图 $Y_{SA} \in \mathbb{R}^{K \times W \times H}$ 。将输入特征图与空间注意力图的对应元素相乘可以增强图像中交通标志区域的权重,降低背景区域的权重,从而提高交通标志的识别率。

集成的注意力模型如图 4 所示。输入的特征图  $Y$  经过通道注意力模块后生成通道注意力图,将其与原始特征图进行元素相乘以得到  $Y'$ ,接着经过空间注意力模块后生成空间注意力图,将其与  $Y'$  进行元素相乘以得到最终改善的特征  $Y_{out}$ ,具体可表示为

一个低分辨率的等效点可表示为

$$\tilde{p} = \left[ \frac{p}{R} \right], \quad (3)$$

式中: $R$  为输出的跨度,实验中  $R$  值为 4; $p$  为原始真值图中的关键点; $\tilde{p}$  为低分辨率的等效关键点; $[\cdot]$ 为映射符号。

使用  $\sigma_p$  表示目标自适应尺寸的标准差,卷积神经网络在训练过程中利用高斯核公式将真值图中的所有关键点映射到热力图中,表达式为

$$Y_{xy,C} = \exp \left[ -\frac{(x - \tilde{p}_x)^2 + (y - \tilde{p}_y)^2}{2\sigma_p^2} \right], \quad (4)$$

式中: $(x, y)$  表示热力图中对应点的平面坐标。交通标志中心点预测的损失函数  $L_k$  可表示为

$$L_k = \frac{1}{N} \sum_{xy,C} \begin{cases} (1 - \hat{Y}_{xy,C})^\alpha \ln(\hat{Y}_{xy,C}), & \text{if } Y_{xy,C} = 1 \\ (1 - Y_{xy,C})^\beta (\hat{Y}_{xy,C})^\alpha \ln(1 - \hat{Y}_{xy,C}), & \text{otherwise} \end{cases}, \quad (5)$$

式中: $\alpha$  和  $\beta$  为损失函数的超参数; $N$  为街景图像中关键点的数量; $\hat{Y}_{xy,C}$  值为 1,表示检测的点为交通标志关键点, $\hat{Y}_{xy,C}$  值为 0,表示检测的点为街景图像的背景点。

网络中的编码器-解码器会对输入的图像进行多次池化操作,但在重新映射回原图像后会产生位置偏差,因此使用位置偏差损失函数  $L_{off}$  来补偿交通标志中心点的位置偏差,表达式为

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \hat{O}_p - \left( \frac{p}{R} - \tilde{p} \right) \right|, \quad (6)$$

式中:  $\frac{p}{R} - \tilde{p}$  为交通标志中心点的位置偏差;  $\hat{O}_p$  为交通标志中心点位置偏差的预测值。

使用  $(x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)})$  表示待测交通标志的边界框, 则交通标志中心点  $p_k$  的坐标为  $\left( \frac{x_2^{(k)} - x_1^{(k)}}{2}, \frac{y_2^{(k)} - y_1^{(k)}}{2} \right)$ , 交通标志的尺寸  $s_k$  为  $(x_2^{(k)} - x_1^{(k)}, y_2^{(k)} - y_1^{(k)})$ , 其中  $k$  为第  $k$  个关键点。使用  $\hat{S}$  表示交通标志尺寸的预测值, 使用尺寸损失函数  $L_{\text{size}}$  来预测交通标志的尺寸, 表达式为

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^N |\hat{S}_{p_k} - s_k|. \quad (7)$$

综上, 网络的整体损失函数由中心点预测损失  $L_k$ 、位置偏差损失  $L_{\text{off}}$  和尺寸损失  $L_{\text{size}}$  构成, 表达式为

$$L = L_k + \lambda_s L_{\text{size}} + \lambda_o L_{\text{off}}, \quad (8)$$

式中:  $\lambda_s$  和  $\lambda_o$  分别为控制位置偏差损失  $L_{\text{off}}$  和尺寸损失  $L_{\text{size}}$  的超参数, 实验中  $\lambda_s$  和  $\lambda_o$  分别设为 0.1 和 1.0。

### 3 实验结果与分析

#### 3.1 交通标志识别的数据集

TT100K 数据集<sup>[15]</sup>是非常具有挑战性的交通标志数据集之一, 共包含  $1 \times 10^5$  张从腾讯街景全景图中收集的图像, 图像涵盖各种天气条件和光照变化。数据集中街景图像的分辨率为  $2048 \text{ pixel} \times 2048 \text{ pixel}$ , 分辨率比其他数据集更大, 并且涵盖的交通标志种类更多, 尺寸更小。在  $[0, 32]$  像素之间的交通标志约占总数的 42%, 这些小型交通标志所占面积不到街景图像面积的  $1/4000$ 。在  $[32, 96]$  像素之间的交通标志约占总数的 49%, 而在  $[96, 400]$  像素之间的大型交通标志仅占总数的 9%, 图 5 为

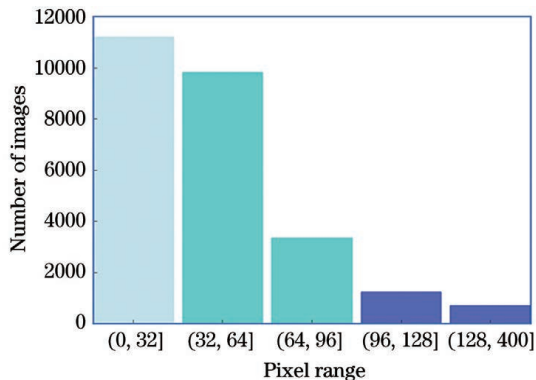


图 5 TT100K 数据集中交通标志的尺寸分布

Fig. 5 Size distribution of traffic signs in TT100K dataset

TT100K 数据集中交通标志的尺寸分布。

由于 TT100K 数据集中不同类别的交通标志数量不均匀, 一些罕见的交通标志数量稀少, 为此实验忽略了在数据集中出现频次小于 100 的类别, 对剩余的 45 类交通标志进行训练。为了平衡样本并确保每个类别至少包含 1000 张样本, 实验对数据集中数量少于 1000 的类别进行重采样操作。为了证明所提算法的鲁棒性, 将数据集中的交通标志按照像素范围分为小型 (像素区间为  $(0, 32]$ , S)、中型 (像素区间为  $(32, 96]$ , M) 与大型 (像素区间为  $(96, 400]$ , L) 三种尺度并进行指标评测。

#### 3.2 实验设置

神经网络是使用深度学习框架 PyTorch 搭建的并在 Ubuntu 16.04 的环境中进行训练, 实验中使用的显卡型号为英伟达 GTX1080Ti。网络的输入是经过裁剪与数据增强后的街景图像, 分辨率为  $512 \text{ pixel} \times 512 \text{ pixel}$ 。实验选择的优化算法为随机梯度下降 (SGD) 算法, 网络的初始学习率设为 0.00125, 批量处理的数据规模为 8, 共进行 120 轮迭代训练, 分别在第 50 轮、80 轮和 110 轮学习率降低 90%。

#### 3.3 实验结果分析

##### 3.3.1 交通标志识别性能评价

在交通标志数据集 TT100K 中进行对比实验, 将所提模型的实验结果与 TT100K 数据集的提出者 Zhu 等<sup>[15]</sup>使用的网络模型、三个经典的目标检测网络 Faster R-CNN<sup>[4]</sup>、FPN (Feature Pyramid Networks)<sup>[16]</sup> 和 RetinaNet<sup>[17]</sup> 以及两个无锚框的目标检测网络 FCOS (Fully Convolutional One-Stage Object Detection)<sup>[18]</sup> 和 CenterNet<sup>[8]</sup> 进行比较, 各对比实验中均对 TT100K 数据集进行相同处理, 实验结果如表 1 所示。从表 1 可以看到, Faster R-CNN 对小型交通标志的识别性能最差, 只有 72.0% 的召回率与 76.1% 的精确率; FPN 与 RetinaNet 均在网络中加入特征金字塔结构, 利用图像的多尺度信息识别小型交通标志可以获得一定的提升, 但仍存在较大的改进空间; 两种无锚框的目标检测网络 FCOS 与 CenterNet 能够解决锚框所带来的正负样本不均衡问题, 而且均取得了较好的效果, 说明研究无锚框的交通标志识别算法是有意义的; AAF CNN 模型采用编码器-解码器结构可以充分利用街景图像的多种尺度信息, 对各种尺寸的交通标志均具有出色的识别效果; 与 CenterNet 模型相比, 所提的 AAF CNN 模型识别小型交通标志的精确率提高了

1.1 个百分点,对中型交通标志与大型交通标志的精确率分别提高了 0.6 个百分点与 1.9 个百分点,并且网络的参数量仅为 CenterNet 的 25.1%。此外,引入 F1 分数来全面评估网络模型的识别性能,

其中 AAFCNN 模型在三种尺度下均取得了最高的得分,并在小型交通标志的情况下获得了 90.9% 的得分,取得了比其他目标检测网络更出色的效果,这充分说明了所提模型的优越性。

表 1 不同交通标志识别方法的性能对比

Table 1 Performance comparison of different traffic sign recognition methods

Method	Backbone	Params /10 <sup>6</sup>	Index	S / %	M / %	L / %
Faster R-CNN	ResNet-101	52.2	Recall	72.0	91.3	91.5
			Precision	76.1	87.5	86.1
			F1-score	74.0	89.4	88.7
Faster R-CNN + FPN	ResNet-101	60.1	Recall	86.6	95.5	<b>95.1</b>
			Precision	85.0	92.9	92.3
			F1-score	85.8	94.2	93.7
Ref. [15]		81.2	Recall	87.4	93.6	87.7
			Precision	81.7	90.8	90.6
			F1-score	84.5	92.0	89.1
RetinaNet	ResNeXt-101	94.7	Recall	87.4	95.1	93.1
			Precision	84.3	95.9	94.2
			F1-score	85.8	95.5	93.6
FCOS	ResNeXt-101	89.7	Recall	88.7	95.6	92.4
			Precision	85.6	96.4	93.5
			F1-score	86.8	96.0	93.0
CenterNet	HourglassNet	191.3	Recall	89.7	<b>96.0</b>	92.4
			Precision	90.1	96.7	94.9
			F1-score	89.9	96.3	93.6
AAFCNN	DenseNet-121	<b>48.1</b>	Recall	<b>90.6</b>	95.6	93.1
			Precision	<b>91.2</b>	<b>97.3</b>	<b>96.8</b>
			F1-score	<b>90.9</b>	<b>96.4</b>	<b>94.9</b>

图 6 为 7 种识别方法的准确率-召回率曲线,根据交通标志的尺度信息分为三幅图。从图 6 可以看

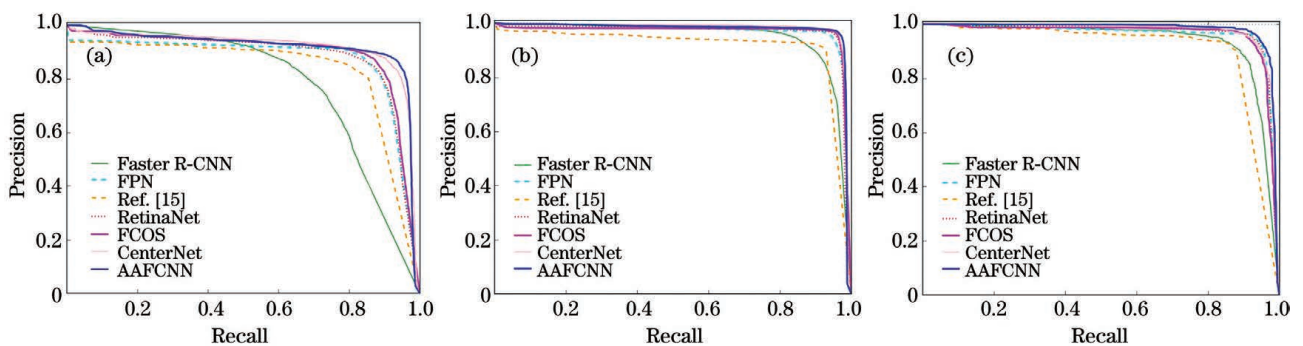


图 6 三种尺度交通标志的准确率-召回率曲线。(a) (0,32] 的像素区间;(b) (32,96] 的像素区间

Fig. 6 Accuracy-recall curves of traffic signs at three scales. (a) Pixel interval of (0,32]; (b) pixel interval of (32,96];

(c) pixel interval of (96,400]

到,对于(32,96]的像素区间,所提模型的性能与 FCOS 模型和 CenterNet 模型相当,显著优于 Faster R-CNN 与文献[15]模型;对于(0,32]与(96,400]的像素区间,所提模型的性能均优于其余方法。图 7 为 TT100K 测试集中部分图像的可视化识别结果,为了便于观察,将图中的交通标志使用

矩形框标出并在图像的右下角放大。第一行图像为交通标志离散分布下的识别结果,第二行图像为交通标志聚集下的识别结果。从图 7 可以看到,在复杂的交通场景下,无论是近距离的交通标志还是远距离的交通标志,所提的 AAFCNN 模型均能准确地定位与识别交通标志。

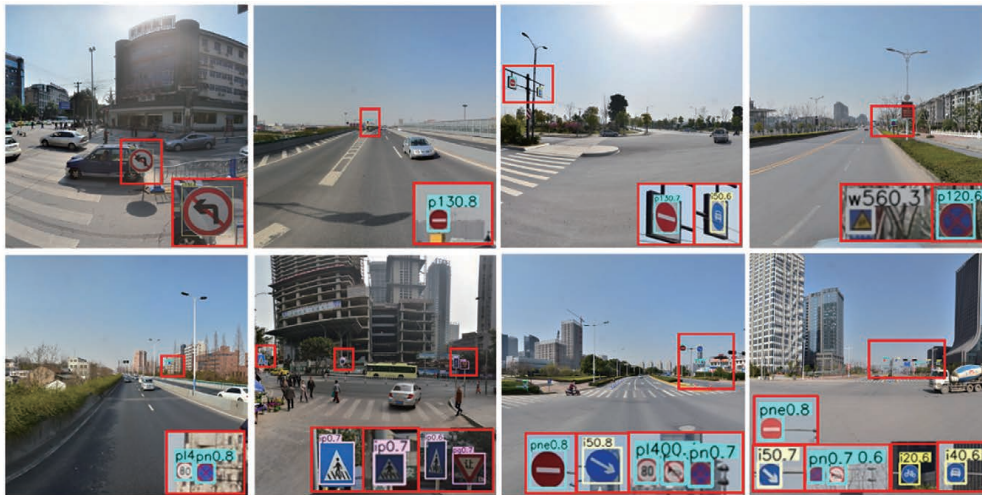


图 7 AAFCNN 模型的部分可视化识别结果

Fig. 7 Part of visual recognition results of AAFCNN model

### 3.3.2 模块有效化验证

为了探究解码路径中密集连接网络的深度对网络整体性能的影响,在 TT100K 数据集中分别利用 DenseNet-121、DenseNet-161、DenseNet-169 和 DenseNet-201 4 种骨干网络对特征进行提取,实验结果如表 2 所示。为了更直观地给出识别效果,实验使用 AP(Average Precision)来衡量性能<sup>[19]</sup>。从表 2 可以看到,编码路径中利用 DenseNet-121 对特征进行提取,网络的实验性能最佳,但该网络会过多地加深编码路径,而且不会提升网络性能。加深编码路径中的特征提取网络会降低识别小型交通标志的精确率,并且引入大量的无用参数。

表 2 密集连接网络的深度对识别性能的影响

Table 2 Effect of depth of densely connected network on recognition performance

Backbone	Params /10 <sup>6</sup>	AP /%		
		S	M	L
DenseNet-121	<b>48.1</b>	<b>63.4</b>	<b>80.1</b>	<b>86.1</b>
DenseNet-169	65.4	62.5	79.9	<b>86.1</b>
DenseNet-201	101.4	61.7	79.7	85.7
DenseNet-264	154.8	61.9	80.0	85.0

在网络的编码路径与解码路径中均加入注意力

模型,用来研究注意力模型的位置对网络识别性能的影响,结果如表 3 所示。从表 3 可以看到,在编码路径中加入注意力模型可以使网络的识别性能最优,对于小型交通标志的识别,比解码路径中加入注意力模型的识别率高出 1.3 个百分点;同时在解码路径与编码路径中加入注意力模型后,网络的识别性能不会得到更多的增益,解码路径中的注意力模型会抑制编码路径注意力模型的性能。在网络提取特征的过程中加入注意力引导,可以提取出更多有益于交通标志定位与分类的语义特征,因此注意力模型加在网络的编码路径中会使识别结果更理想。

表 3 注意力模型的位置对识别性能的影响

Table 3 Effect of location of attention model on recognition performance

Location	Params /10 <sup>6</sup>	AP /%		
		S	M	L
In coding path	48.1	<b>63.4</b>	<b>80.1</b>	<b>86.1</b>
In decoding path	<b>47.8</b>	62.1	80.0	85.8
Both coding path and decoding path	48.2	62.6	80.0	85.1

为了验证注意力模型与语义连接路径的有效性,对模型进行了消融实验,实验结果如表 4 所示,其中 base 代表采用跳跃连接方式并且移除注意力

模块的基准模型, AM 表示注意力模型, SCP 表示语义连接路径。从表 4 可以看到, 加入语义连接路径后, 由于需要对密集连接网络输入的特征图进行多次卷积运算, 所以网络的参数量显著增加, 但网络的识别性能得到明显提升, AP 值分别提高 1.1、0.3 与 1.3 个百分点; 在网络中加入注意力模型后, 虽然在一定程度上降低网络对大型交通标志的识别性能, 但显著提升网络对小型交通标志的识别率; 对于小型交通标志, 加入注意力模型与语义连接路径后, 所提模型的 AP 值比基准模型提升了 2.6 个百分点, 充分说明了增加的模块对小型交通标志识别的有效性。

表 4 各模块的性能对比

Table 4 Performance comparison of each module

Model	Params / $10^6$	AP / %		
		S	M	L
Base	<b>14.1</b>	60.8	79.7	85.9
Base+AM	14.2	61.7	79.8	85.1
Base+SCP	47.8	61.9	80.0	<b>87.2</b>
Base+AM+SCP	48.1	<b>63.4</b>	<b>80.1</b>	86.1

## 4 结 论

本文提出一种基于注意力模型的无锚框交通标志识别网络, 其由编码路径与解码路径两部分构成。编码路径采用密集连接网络 DenseNet-121 作为骨干网络并进行特征提取, 可以获取交通标志的轮廓信息; 解码路径通过上采样层逐步修复交通标志的细节信息。为了减小编码路径与解码路径之间的语义鸿沟, 本文设计一种语义连接路径。实验结果表明, 加入语义连接路径后, 神经网络对小型、中型与大型交通标志的精确率分别提高 1.1、0.3 与 1.3 个百分点。为了提升对小型交通标志的识别效果, 本文在编码路径中加入注意力模型, 可以抑制特征图中背景元素的权重, 能够使网络关注交通标志区域。通过采用交通标志中心点的检测方式, 本文成功解决正负样本的不均衡问题。与目前领先的几种目标检测网络相比, 所提的 AAFCNN 模型在 TT100K 数据集中获得了更出色的识别效果。

## 参 考 文 献

- [1] Qu Z H, Shao Y M, Deng T M, et al. Traffic sign detection and recognition under complicated lighting conditions[J]. *Laser & Optoelectronics Progress*, 2019, 56(23): 231009.
- [2] 屈治华, 邵毅明, 邓天民, 等. 复杂光照条件下的交通标志检测与识别[J]. *激光与光电子学进展*, 2019, 56(23): 231009.
- [3] Madani A, Yusof R, Maliha A. Traffic sign segmentation using supervised distance based classifiers[C]//2015 10th Asian Control Conference (ASCC), May 31-June 3, 2015, Kota Kinabalu, Malaysia. New York: IEEE Press, 2015: 15436782.
- [4] Wang Y Q, Liu L M, Zhao Y. Traffic sign detection based on fixed color combination and intensity restraint[C]//2009 International Symposium on Computer Network and Multimedia Technology, January 18-20, 2009, Wuhan, China. New York: IEEE Press, 2009: 11069873.
- [5] 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.
- [6] 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 Press, 2016: 779-788.
- [7] Zhang Y, Gong Z Y, Wei W W. Traffic sign detection based on improved Faster R-CNN model[J]. *Laser & Optoelectronics Progress*, 2020, 57(18): 181015.
- [8] 张毅, 龚致远, 韦文闻. 基于改进 Faster R-CNN 模型的交通标志检测[J]. *激光与光电子学进展*, 2020, 57(18): 181015.
- [9] Tong Y, Yang H C. Traffic sign recognition based on improved neural networks[J]. *Laser & Optoelectronics Progress*, 2019, 56(19): 191002.
- [10] 童英, 杨会成. 基于改进神经网络的交通标志识别[J]. *激光与光电子学进展*, 2019, 56(19): 191002.
- [11] Zhou X, Wang D, Krahenbuhl P, et al. Objects as points [EB/OL]. (2019-04-16) [2020-08-17]. <https://arxiv.org/abs/1904.07850v1>.
- [12] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [13] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W, et al. *Medical image computing and computer-*



- assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [11] Milletari F, Navab N, Ahmadi S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C] // 2016 Fourth International Conference on 3D Vision (3DV), October 25-28, 2016, Stanford, CA, USA. New York: IEEE Press, 2016: 565-571.
- [12] Ibtehaz N, Rahman M S. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation [J]. Neural Networks, 2020, 121: 74-87.
- [13] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning [EB/OL]. [2020-08-17]. [https://www.researchgate.net/publication/301874967\\_Inception-v4\\_Inception-ResNet\\_and\\_the\\_Impact\\_of\\_Residual\\_Connections\\_on\\_Learning](https://www.researchgate.net/publication/301874967_Inception-v4_Inception-ResNet_and_the_Impact_of_Residual_Connections_on_Learning).
- [14] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [15] Zhu Z, Liang D, Zhang S H, et al. Traffic-sign detection and classification in the wild [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2110-2118.
- [16] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [17] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [18] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea. New York: IEEE, 2019: 9626-9635.
- [19] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [M] // Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.