计算机科学
COMPUTER SCIENCE

ETF-YOLO11n:交通图像的多尺度特征融合目标检测方法

夏淑芳, 尹昊楠, 瞿中

引用本文

夏淑芳, 尹昊楠, 瞿中. ETF-YOLO11n:交通图像的多尺度特征融合目标检测方法[J]. 计算机科学, 2025, 52(12): 150-157.

XIA Shufang, YIN Haonan, QU Zhong. ETF-YOLO11n:Object Detection Method Based on Multi-scale Feature Fusion for TrafficImages [J]. Computer Science, 2025, 52(12): 150-157.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于改进YOLOv8的城市交通视域下的目标识别算法

Target Recognition Algorithm in Urban Traffic Field of View Based on Improved YOLOv8

计算机科学, 2025, 52(11A): 241200131-8. https://doi.org/10.11896/jsjkx.241200131

融合注意力机制的道路场景三维目标检测算法

Three-dimensional Object Detection Algorithm of Road Scene Based on Attention Mechanism

计算机科学, 2025, 52(11A): 241100112-7. https://doi.org/10.11896/jsjkx.241100112

基于特征提取增强和金字塔结构的实时Transformer小目标检测模型

Real-time Transformer Small Target Detection Model Based on Feature Extraction Enhancement and Pyramid Structure

计算机科学, 2025, 52(11A): 250100139-11. https://doi.org/10.11896/jsjkx.250100139

面向全天候多场景的多模态融合目标检测方法

Multi-modal Fusion Based Object Detection for All-day and Multi-scenario Environments

计算机科学, 2025, 52(11A): 241100137-10. https://doi.org/10.11896/jsjkx.241100137

C2P-YOLO:一种轻量级的风电塔筒裂缝检测算法

C2P-YOLO:A Lightweight Crack Detection Algorithm for Wind Turbine Towers

计算机科学, 2025, 52(11A): 250100126-6. https://doi.org/10.11896/jsjkx.250100126

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.241200021

ETF-YOLO11n:交通图像的多尺度特征融合目标检测方法

夏淑芳　尹昊楠　瞿　中
重庆邮电大学软件工程学院　重庆 400065
(xiasf@cqupt.edu.cn)

摘　要　近年来深度学习算法在计算机视觉领域取得了显著的进展,但是由于复杂交通图像中存在目标尺寸小、特征信息不明显、易受干扰等问题,目标检测精度依旧不高。针对这一问题,对最先进的检测模型 YOLO11 进行改进,设计了多尺度特征融合模型 ETF-YOLO11n(Effective Traffic Feature YOLO)。首先,设计了三重特征融合模块 TFF,将主干网络提取到的不同尺寸特征信息进行有效融合;其次,设计了基于混合空洞卷积的特征加强模块 HDCFE,并添加至模型的颈部网络中整合不同感受野提取到的特征,降低因为遮挡等情况对模型的干扰;最后,用提出的 GeoCIoU 替代 CIoU,通过两个不同的惩罚项,模型能更精准地反馈检测框与真实框的匹配情况。所提出的 ETF-YOLO11n 在交通数据 KITTI 上 AP 达到 65.6%,mAP@0.5 达到 90.7%,与基线模型 YOLO11n 相比分别提升了 2.4 个百分点和 1.2 个百分点,体现了良好的检测效果。此外,ETF-YOLO11n 在 COCO-Traffic 数据集上 AP 和 mAP@0.5 分别达到了 42.5%和 59.8%;所提出的方法迁移至 YOLOv8 模型,在 KITTI 数据集上 AP 和 mAP@0.5 分别达到 66.9% 和 91.5%。实验结果表明,所提出的方法能显著提升模型的检测能力,且对不同模型和数据集都有较好的泛化性,在精度与参数量上达到了很好的平衡[1)]。

关键词:目标检测;多特征融合;交并比;特征加强;复杂交通场景
中图分类号　TP391.41

ETF-YOLO11n:Object Detection Method Based on Multi-scale Feature Fusion for Traffic Images

XIA Shufang,YIN Haonan and QU Zhong
School of Software Engineering,Chongqing University of Posts and Telecommunications,Chongqing 400065,China

Abstract　Deep learning algorithms have made significant progress in the field of computer vision in recent years,but the accuracy of object detection in complex traffic scenes is still unsatisfactory due to the small size of traffic objects,inconspicuous feature,and susceptibility to interference. To address this problem,this paper improves the state-of-the-art YOLO11 and designs the ETF-YOLO11n based on multi-scale feature fusion. Firstly,it designs TFF,which effectively fuses the feature information of different sizes extracted from the backbone. Secondly,it designs HDCFE,effectively integrates the features extracted from different receptive fields and reduces the interference on the detection effect of the model due to occlusion and overlapping. Finally,the proposed GeoCIoU is used to replace CIoU,and the model can provide more accurate feedback on the matching of the predicted box and the ground-truth box through the two different penalization terms. The ETF-YOLO11n achieves an AP of 65.6% and mAP@0.5 of 90.7% on KITTI dataset,which is improved by 2.4 percentage points and 1.2 percentage points. In addition,ETF-YOLO11n achieves 42.5% AP and 59.8% mAP@0.5 on COCO-Traffic,and EFT-YOLOv8n achieves 66.9% AP and 91.5% mAP@0.5 on KITTI dataset. The results show that the proposed methods significantly improve the performance and have good generalization ability to different models and datasets,achieve a good balance between the accuracy and parameters. The source code has been opened.

Keywords　Object detection,Multi-feature fusion,Inter over Union,Feature enhancement,Complex traffic scenarios

---

1) https://gitee.com/yin_haonan_handsome/etf-yolo11

到稿日期:2024-12-02　返修日期:2025-03-30
基金项目:国家自然科学基金(62576058,62571077);重庆市教育委员会科学技术研究项目(KJZD-M202300604);重庆市自然科学基金(2023NSCQ-MSX1781)
This work was supported by the National Natural Science Foundation of China(62576058,62571077),Scientific and Technological Research Program of Chongqing Municipal Education Commission(KJZD-M202300604) and Natural Science Foundation of Chongqing,China(2023NSCQ-MSX1781).
通信作者:瞿中(quzhong@cqupt.edu.cn)

1 引言

目标检测是计算机视觉领域中一个非常具有挑战性的任务，旨在对图像中的目标进行定位和分类[1]。在过去的几十年中，目标检测经历了从基于人工设计的方法到基于深度学习的方法的不同发展阶段。相比于传统的手工特征设计方法，基于深度学习的目标检测方法避免了繁琐的手工设计过程，学习快速，泛化能力强，可以分为单阶段算法、两阶段算法和基于 Transformer 的检测算法。单阶段算法直接在输入图像上进行目标检测，具有较高的实时性，包括 SSD（Single Shot Multi-box Detector）[2]、YOLO（You Only Look Once）系列算法[3-9]；两阶段算法先生成一组候选区域，再在候选区域内进行检测，以 Faster R-CNN（Faster Region Convolutional Neural Network）[10] 为代表；Transformer 系列算法的代表 DETR（DEtection TRansormer）[11]，采用编码器-解码器构架，实现端到端目标检测。

据公安部相关数据显示，截至 2024 年 6 月底，全国汽车保有量为 4.4 亿辆[12]。对交通图像实施精准检测能准确识别出车辆、道路标志等物体，从而有效降低交通事故发生率，保障驾驶人员与行人的生命安全，对人类社会、经济具有重大意义，但是目前自动驾驶系统的可靠性与智能性还不能完全满足实际应用需求。

YOLO 系列算法采用的路径聚合网络中，自下而上路径将低层特征转移到高层特征，自上而下路径将特征从高层转移到低层，多采样过程会导致特征信息丢失，影响特征提取的效果，且不同层之间存在语义鸿沟，若将其直接融合，效果通常不好；交通图像目标分布不均匀、视觉特征不明显，缺乏边缘、颜色等细节信息，网络很难获得准确的细节特征。

为解决上述问题，本文提出了 ETF-YOLO11n 算法，具体贡献如下：(1)设计三重特征融合模块 TFF（Triple Feature Fusion），将提取到的多特征层进行融合；(2)设计基于混合卷积的特征加强模块 HDCFE（Hybrid Dilated Convolution Feature Enhancement），增强模型的全局信息处理能力；(3)提出 GeoCIoU（Geometrically Impr-oved Complete IoU），加入点对点距离度量，提升准确性。

本文在 KITTI[13] 上验证模型的有效性，同时在 COCO-Traffic 数据集和 YOLOv8[7] 上验证泛化性。结果表明，ETF-YOLO11n 的 AP 和 mAP@0.5 分别达到 65.6% 和 90.7%，在提升 0.3×10^6 参数量的情况下，平均检测精度提升 2.4 个百分点，能精准检测交通图像中的目标。

2 相关工作

2.1 YOLO11 网络

YOLO 是 Joseph Redmon 在 2015 年提出的实时目标检测方法，YOLO11[9] 是 Ultralytic 公司在 2024 年 9 月发布的最新一代目标检测算法。如图 1 所示，YOLO11 由输入（Input）、主干（Backbone）、颈部（Neck）、输出（Output）4 部分组成。YOLO11 精度与速度全面提升，广泛应用于检测、分割等领域，其主要创新如下：(1)设计了 C3k2，C3k2 是 CSP（Cross Stage Partial Network）的更快实现；(2)改进 C2PSA 模块，增强特征提取能力。YOLO11 有 n，s，m，l，x 共 5 种模型，但交通目标检测要在提升精度的同时兼顾模型大小，因此本文在 nano 模型上进行实验。

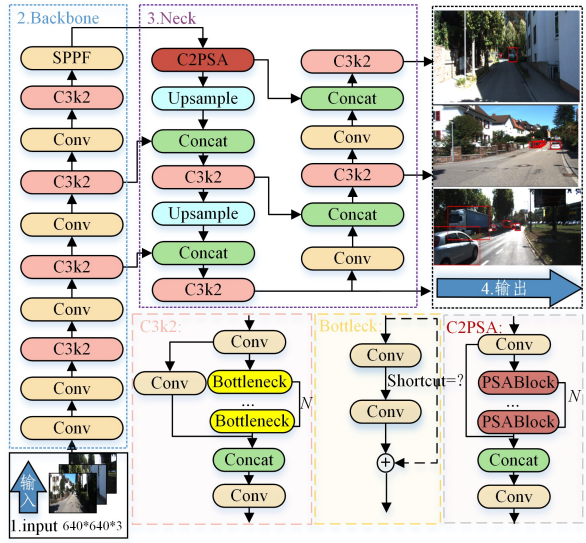


图 1 YOLO11 模型结构

Fig. 1 Architecture of YOLO11 model

2.2 多尺度特征融合

大量的研究表明，多尺度特征融合应用到检测任务中效果显著，是因为网络深层特征具有较强语义特征，低层级特征具有更高分辨率和更多细节信息，通过各层的组合达到互补的效果，模型能得到更高的检测精度。

Lin 等[14] 提出了 FPN（Feature Pyramid Network），对多级预测特征层进行检测。Liu 等[15] 在 FPN 的基础上加入了自顶向下和自底向上的双向融合骨干网，提出 PANet（Path Aggregation Network）。Lan 等[16] 提出 FFAN（Feature Fusion and Attention Network）。Lu 等[17] 提出 CA-FPN（Channel-space Adaptive Enhancement FPN），通过提取不同特征的内在关系，实现多特征自适应融合。

本研究基于多尺度特征融合，通过有效融合 Backbone 和 Neck 提取到的不同特征信息，有效弥补了特征在采样中丢失的不足，解决了非相邻层级特征融合效果较差的问题。

2.3 边界框回归损失函数

IoU（Inter over Union）[18] 是目标检测的常用指标，用于衡量预测框与真实框之间的重叠情况。基于 IoU 的损失函数能引导模型朝提高 IoU 的方向更新，通过直接优化两框的重叠程度来准确地定位目标，提升检测性能。

Rezatofighi 等[19] 提出了 GIoU（Generalized-IoU），引入最小外接矩形，能关注不重叠区域。Zheng 等[20] 提出了 DIoU（Distance IoU）和 CIoU（Complete IoU），DIoU 引入两框中心点的距离，CIoU 引入高宽比作为惩罚项。Gevorgyan 等[21] 提出了 SIoU（SCYLLA-IoU），重新定义角度惩罚项，同时包括角度损失、距离损失、形状损失、IoU 损失。Luo 等[22] 提出 UIoU（United-IoU），动态地将模型的注意力从低质量预测框转移到高质量预测框。将本文策略与现有方法进行对比，结果如表 1 所列。

表 1 不同 IoU 回归损失方法的对比

Table 1 Comparison of different IoU regression loss methods

方法	贡献	存在的问题
IoU[18]	能够准确描述预测框和真实框之间的匹配程度	当真实框与预测框无重叠时，IoU 为 0，无法回归
GIoU[19]	引入最小外接矩形，能更精准反映预测框与真实框的重合度	两框完全重叠时，GIoU 退化为 IoU，同样存在损失为 0 的情况，无法精准定位
DIoU[20]	将真实框和预测框中心点之间的欧氏距离作为回归对象	当 IoU 值和两个框的中心点距离一样时，DIoU 的值一致无法优化模型
CIoU[20]	在 IoU 的基础上，引入真实框与预测框的高宽比作为回归对象	预测框与真实框的长宽比一致时，惩罚项为 0，此时模型无法优化
SIoU[21]	考虑了两框之间的角度差异，重新定义距离损失	需要设置多个 IoU 阈值来进行计算，需要根据不同数据集而调整，泛化性差
UIoU[22]	兼顾 IoU 权重和置信度信息，对不同质量预测框的权重，关注高质量预测框	需要提前为不同数据集设置不同超参数“ration”，泛化性较差
GeoCIoU(Ours)	最小化预测框和真实框之间的左上和右下点距离，通过两种不同的惩罚项，使得训练反馈更准确	

本研究也基于 IoU 回归损失，对 YOLO11 算法使用的 CIoU 进行改进，提出 GeoCIoU，引入点对点距离作为额外惩罚项，有效解决了预测框与真实框具有相同高宽比时无法回归的问题，提升了回归的准确性。

2.4 注意力机制

注意力机制是一种常用的提升模型性能的方法。它通过对输入数据进行加权处理，使模型更专注于图像中的重要信息，能有效聚焦于图像中与目标相关的区域。

Hu 等[23]提出了通道注意力 SE(Squeeze and Excitation)。Woo 等[24]提出了结合信道和空间信息的注意力机制 CBAM(Convolutional Block Attention Module)。Ren 等[25]提出了用于自动驾驶的即插即用注意力机制 CuDAM(Curio-sity-driven Attention Mechanism)。Vaswania 等[26]提出的 Transformer 模型采用自注意力机制 SA 对输入序列中的每个位置进行加权求和，捕捉向量之间的依赖关系。Carion 等[11]结合自注意力与编码器和解码器，设计了端到端的目标检测模型 DETR(Deter Transformer)。Anthony 等[27]提出了定向注意力 DA(Directed Attention)，它将自注意力头限定在固定视野内，提高了模型在目标检测中的泛化能力。

注意力机制被广泛应用于提升模型的检测精度，在之后的工作中也将尝试使用 Transformer 结构和混合注意力机制对交通图像的精准目标检测进行研究。

3 基于多尺度特征加强的交通目标检测

本章首先介绍了改进后的网络整体结构，然后依次介绍了 TFF 模块、HDCFE 模块以及 GEOCIoU 函数在 YOLO11 模型上的应用。

3.1 ETF-YOLO11n 模型结构

本文提出了 ETF-YOLO11n 模型，模型整体结构如图 2 所示。本文设计了 TFF 模块和 HDC 模块，TFF 模块将 Backbone 提取到的不同层级特征进行融合；HDCFE 模块获取不同感受野的特征信息，增强模型的表达能力；此外，还设计了 GeoCIoU 来替换 CIoU 并训练模型。

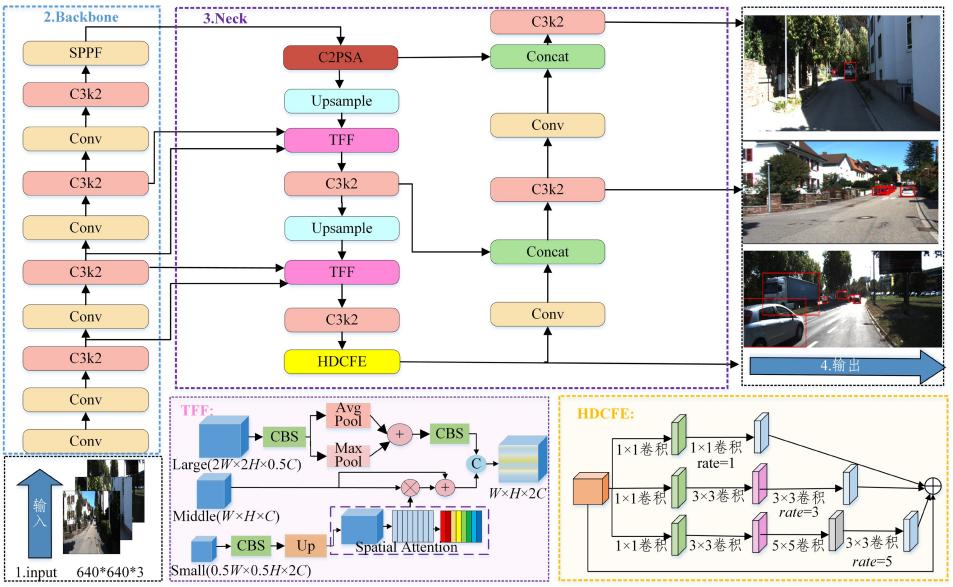


图 2 ETF-YOLO11n 网络结构

Fig. 2 Architecture of ETF-YOLO11n network

3.2 三重特征融合模块

在特征提取网络中，浅层特征细节信息丰富而语义性较差；而深层特征经过更多层卷积等操作后，语义信息更强，但是分别率较低，在卷积大量卷积和上采样过程中，特征信息会被削弱。为增强主干网络和颈部网络在特征提取和融合方面的能力，设计了 TFF 模块(Triple Feature Fusion)。

如图 3 所示，TFF 是一种高效的多特征融合模块，有 3 个不同尺寸的输入，分别为大(Large)、中(Middle)、小(Small)特征图。该模块具体流程如算法 1 所示，通过自适应最大池化和自适应平均池化，将 l 层特征图调整为与中间层特征图 m

相同的大小；通过最近邻插值将 s 的尺寸调整为与中间特征图 m 相同的大小。由于较深层特征包含更多语义信息，因此本文获取 s 层的空间注意力，并将生成的注意力作用于 m 层特征图，以提高模型对不同目标的理解能力。将经过处理的特征图 l、增强后的特征图 m 沿通道维度拼接。与普通场景相比，交通图像中目标遮挡问题严重，*TFF* 综合利用不同尺度的信息，提高了模型对交通场景中细节特征的提取。

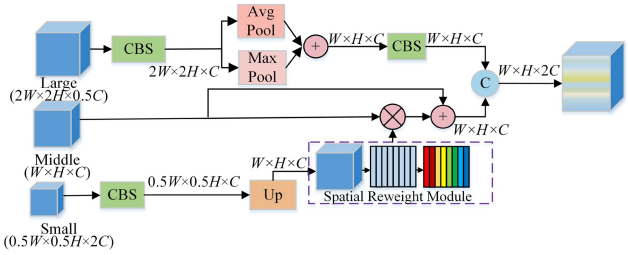


图 3　三重特征融合模块结构

Fig. 3　Architecture of the triple feature fusion module

TFF 模块的伪代码如算法 1 所示。

算法 1　三重特征融合模块(TFF)

输入：不同尺寸特征图 l，m 和 s

输出：融合后的特征图 output

1. 调整大尺度特征图：使用自适应最大池化和平均池化，调整 l 为目标尺寸。

   l_resized＝ad_maxpool(l，目标尺寸)＋ad_avg pool(l，目标尺寸)

2. 调整小尺度特征 s：通过最近邻插值将 s 调整为目标尺寸。

   s_resized＝interpolate(s，目标尺寸，mode＝'nearest')。

3. 对 m 层特征进行加强。

   3.1. 获取小尺度特征图 s_resized 空间注意力。

   attention_map＝SpatialAttention(s_resized)

   3.2. 利用该注意力值增强中尺寸特征图。

   m_enhanced＝m ＊ attention_map＋m

4. 特征拼接：将 l_resized 和 m_enhancd 沿通道维度拼接。

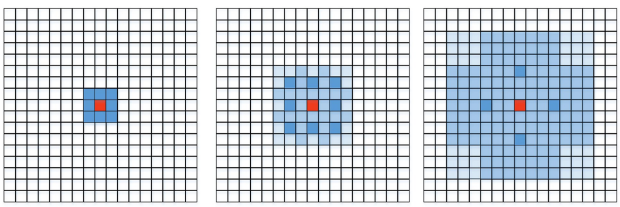
   output＝concatenate(l_resized，m_enhanced)

3.3　基于混合空洞卷积的特征加强

3.3.1　混合空洞卷积

空洞卷积在扩大特征图感受野的同时降低了传统卷积带来的资源消耗，网络保持高分辨率的同时，能捕捉到更广泛的上下文信息。然而空洞卷积采取的是一系列非连续性采样，也正因如此，图像的特征表示会变稀疏，模型在提取信息时会忽略某些局部信息，有很多的特征无法被采集，从而影响模型的检测效果。

混合空洞卷积 HDC(Hybrid Dilated Convolution)是整合一组拥有不同组膨胀率的空洞卷积操作。如图 4 所示，采用一组膨胀空洞率为 1，2，3 的混合空洞卷积，就能在有效扩展感受野的同时，缓解非连续采样导致的特征丢失。



(a)空洞率为 1　(b)空洞率为 2　(c)空洞率为 3

图 4　混合空洞卷积

Fig. 4　Hybrid dilated convolution

3.3.2　基于混合空洞卷积的特征加强模块

为了增强模型对交通目标特征的表达能力，本文通过引入不同大小空洞率的空洞卷积，对目标信息进行整合提取，以便更精准地识别和定位图像中的目标信息，提升网络对全局信息的理解能力。

如图 5 所示，本文首先通过一组 1×1 的卷积对特征图进行分流，精心设计各路径的空洞率分别为 1，3，5(没有公约数，避免采样点过于集中)，同时引入残差连接，得到增强后的特征图。该模块充分利用空洞卷积能改变特征图感受野的优势，提取到不同感受野下的特征信息，最大程度确保了特征提取过程中的连续性与高效性。交通图像的背景元素与目标具有相似的局部特征，HDCFE 通过整合不同感受野分支特征信息，使模型对交通图像全局信息进行特征感知的能力得到加强。

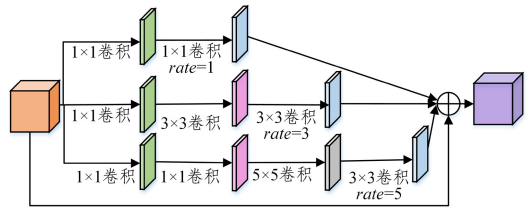


图 5　基于混合空洞卷积的特征加强模块结构

Fig. 5　Architecture of hybrid dilated convolution feature enhancement module

3.4　GeoCIOU

YOLO11 使用 CIoU 作为边界框损失函数，考虑了目标框的长宽比、中心点距离等因素。计算式如式(1)－式(3)所示：

$$CIoU=IoU-\left(\frac{\rho^2(box_1,box_2)}{c^2}+\alpha v\right) \tag{1}$$

$$\alpha=\frac{v}{(1-IoU)+v} \tag{2}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{\omega_{\text{box1}}}{h_{\text{box1}}}-\arctan\frac{\omega_{\text{box2}}}{h_{\text{box2}}}\right)^2 \tag{3}$$

其中，ρ 代表两框中心点之间的欧氏距离，c 代表最小外接矩形对角线长度，α 是权重参数，v 用于衡量两框的宽高一致性。

此时，若预测框与真实框两者具有相同高宽比，那么仅将高宽比作为惩罚项的 CIoU 将失去有效性，这将极大程度地限制模型的精度，导致样本不平衡。因此，本文设计了一种新的基于 IoU 的度量方法 GeoCIoU。在 CIoU 的基础上引入另一惩罚因子，通过最小化预测框和真实框之间的左上和右下点距离，通过两种不同的惩罚项，使得训练反馈更准确。其计算过程如式(4)－式(6)所示：

$$GeoCIoU=IoU-\omega_1\times\left(\frac{\rho^2(box_1,box_2)}{c^2}+\alpha v\right)-\omega_2\times\left(\frac{d_1^2}{w^2+h^2}+\frac{d_2^2}{w^2+h^2}\right) \tag{4}$$

$$d_1=\sqrt{(x_1^{\text{pred}}-x_1^{\text{gt}})^2+(y_1^{\text{pred}}-y_1^{\text{gt}})^2} \tag{5}$$

$$d_2=\sqrt{(x_2^{\text{pred}}-x_2^{\text{gt}})^2+(y_2^{\text{pred}}-y_2^{\text{gt}})^2} \tag{6}$$

其中，d_1 和 d_2 如图 6 所示，分别表示预测框与真实框间左上角的距离和右下角的距离；ω_1 和 ω_2 是两个惩罚项的权重系数，关于此系数，将在 4.4 节中详细介绍。

复杂交通图像中的小目标较多，长宽比例差别较小，加入点对点的距离度量，能够精细地反映框的每个边界、每个角落

之间的差异，从而增强框的位置和形状的稳定性，从更多维度来评估框的匹配度，包括目标框边界的逐点位置、形状等。基于此，训练过程中的损失反馈能更加精确地引导模型去学习准确的边界框位置，更适配交通场景中各类目标的检测需求。

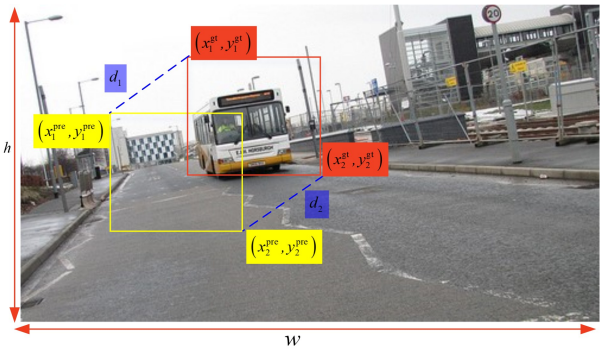


图 6 GeoCIoU 回归损失

Fig. 6 GeoCIoU regression loss

3.5 小结

本章介绍基于多尺度特征加强的交通目标检测方法，提出 ETF-YOLO11n 模型，设计 TFF 模块融合多尺度特征、HDCFE 模块利用混合空洞卷积增强表达，并改进边界框损失函数为 GeoCIoU，以提升模型检测精度。

4 实验结果分析

本章介绍了实验中使用的数据集和评价指标，描述了实验的细节，对 GEOCIoU 策略系数进行了讨论，将本文方法与目前国际前沿方法进行对比，通过消融实验对各个策略的有效性进行了探究，并分析说明了模型复杂度等情况。

4.1 数据集

本文主要在交通类数据集 KITTI 上进行实验，此外在 COCO-Traffic 数据集上验证了模型的泛化能力。

(1)KITTI[13]。该数据集是自动驾驶常用数据集之一，有汽车、行人、有轨电车、自行车、货车、卡车 6 个类别，随机分配给训练集 6981 幅图像，测试集 500 幅图像。

(2)COCO-Traffic。COCO 数据集是目标检测领域最具挑战性的数据集，共 80 类别，164 000 幅图像，包含大量小目标和密集目标。COCO-Traffic 从中提取 78 694 幅交通图像，包含行人、自行车、汽车等 12 类，其中 62 016 幅图用于训练，16 678 幅图用于验证。

4.2 实验设置

为了确保实验的公平性，本文所涉及的全部实验均与 YOLO11n 基线模型一致且满足以下条件：(1)使用深度学习框架 PyTorch，其版本为 1.7.1；(2)CPU 为 4-core Intel(R) Xeno(R) Sliver，GPU 为 40GB 的 Tesla A100；(3)训练轮次 epochs 设为 100，批量大小 batchsize 为 16；(4)使用 SGD 优化器进行训练，初始学习率设置为 0.01，动量设置为 0.937。

4.3 评价指标

本文使用准确率(Precision，P)、召回率(Recall，R)、平均检测精度(Average Precision，AP)、参数量(Parameters，Params)、计算量(Floating Point Operations，FLOPs)、每秒传输帧数(Frames Per Second，FPS)衡量模型的效果。定义如式(7)—式(12)所示：

$$P=\frac{TP}{TP+FP} \tag{7}$$

$$R=\frac{TP}{TP+FN} \tag{8}$$

$$AP=\int_0^1 P(R)\mathrm{d}r \tag{9}$$

$$Params=(C_{in}k^2+1)C_{out} \tag{10}$$

$$FLOPs=2HW(C_{in}k^2+1)C_{out} \tag{11}$$

$$FPS=\frac{1}{Average\ inference\ time\times 10^{-3}\times B} \tag{12}$$

其中，真阳性(True Positive，TP)、假阳性(False Positive，FP)、假阴性(False Negative，FN)分别表示正确预测的正样本的数量、预测错误的负样本的数量、预测错误的正样本的数量；H 和 W 分别表示输入图像的高、宽；C_{in} 和 C_{out} 是输入和输出的通道数；*Average inference time* 是模型平均推理时间；B 是训练过程中的批量大小。AP 表示所有类别的的平均检测值，与非极大抑制中交并比的阈值有关；$mAP@0.5$ 是阈值为 0.5 时 AP 的值。

4.4 GeoCIoU 的对比实验

如式(4)所示，GeoCIoU 损失函数通过两个不同的惩罚项探究预测框与真实框的匹配情况，因此需要为这两个不同的惩罚项分配合理的系数。为了更好地平衡框的形状和位置差异，本文进行进一步实验，探究不同的权重系数 ω_1 和 ω_2 对模型检测效果的影响。

如表 2 第 2—4 行所列，与 CIoU 相比，GeoCIoU 增加惩罚项后都能帮助模型提升检测精度。当 $\omega_1=0.6$，$\omega_2=0.4$ 时，平均检测精度达到 64.8%，与基线模型相比提升了 1.6 个百分点，此时 GeoCIoU 的优化效果达到最佳。因此本文后续实验将 ω_1 和 ω_2 分别设为 0.6 和 0.4。

表 2 GeoCIoU 不同权重系数对比实验结果

Table 2 Experimental results with different weights of GeoCIoU

Methods	ω_1	ω_2	AP/%	$mAP@0.5$/%
baseline(CIoU)	—	—	63.2	89.5
+GeoCIoU	0.5	0.5	63.9	89.5
+GeoCIoU	**0.6**	**0.4**	**64.8**	**89.9**
+GeoCIoU	0.7	0.3	64.5	89.5

4.5 对比实验

本文在 KITTI 上进行实验，不仅与 YOLO 系列算法、Faster R-CNN[10]、DETR[11]进行比较，还与交通领域专门方法 S-PANet[28]和 YOLOM[29]进行比较。S-PANet 对 PANet 进行优化并加入改进的 CBAM，YOLOM 综合利用多空间金字塔特征和焦损函数，它们被证明在交通场景中有良好的检测能力。如表 3 所列，本方法的 AP 和 mAP@0.5 分别达到 65.6%和 90.7%，参数量却仅有 2.9×10^6，检测性能优于绝大多数先进的检测方法，参数量是 DETR 模型的 1/15，却有着更好的检测精度。本文方法参数量和检测精度都优于 S-PANet-tiny；ETF-YOLO11 的 nano 模型检测效果与 YOLOMs 模型相近，但参数量和计算量远少于 YOLOMs。

表 3　在 KITTI 数据集上的对比实验

Table 3　Comparative experiments on the KITTI dataset

Methods	Backbone	*Epoch*	*AP*/%	*mAP*@0.5/%	*Params*	*FLOPs*
Faster R-CNN[10]	Resnet-50	100	48.5	60.4	—	—
DETR[11]	Resnet-50	100	54.9	88.9	4.13×10^6	101.3×10^9
DETR[11]	Resnet-101	100	57.4	89.5	6.02×10^6	186.0×10^9
YOLOv5n[3]	CSPDarkNet53	100	63.3	89.1	2.50×10^6	7.1×10^9
YOLOv6n[4]	CSPDarkNet53	100	61.4	87.3	4.30×10^6	11.8×10^9
YOLOX-tiny[5]	CSPDarkNet53	100	55.1	84.4	6.50×10^6	6.5×10^9
YOLOv7-tiny[6]	CSPDarkNet53	100	41.7	64.2	6.00×10^6	13.9×10^9
YOLOv8n[7]	CSPDarkNet53	100	65.4	90.3	3.00×10^6	9.0×10^9
S-PANet-tiny[28]	CSPDarkNet53	100	54.2	84.3	3.80×10^6	5.7×10^9
YOLOMs[29]	CSPDarkNet53	100	65.2	91.5	9.30×10^6	26.9×10^9
YOLO11n(baseline)[9]	CSPDarkNet53	100	63.2	89.5	2.60×10^6	6.3×10^9
ETF-YOLO11n(Ours)	CSPDarkNet53	**100**	**65.6(+2.4)**	**90.7(+1.2)**	**2.90×10^6**	**10.6×10^9**

4.6　消融实验与结果分析

为验证本文所提出 3 种策略的有效性，如表 4 所列，进行了消融实验，训练轮次为 100，实验参数按 4.2 节设置。

本文的基线模型为 YOLO11n，根据表中前 4 行的数据，TFF 模块、HDCFE 模块、GeoCIoU 回归损失都适用于复杂交通场景中的目标检测，在增加参数量不多的情况下，平均检测精度分别提升了 1.3 个百分点、1.2 个百分点、1.6 个百分点。表 4 中第 5—7 行将 3 种策略两两分组，结果显示，平均检测精度都有所提升，这验证了本文提出策略的有效性。

3 种策略组合使用结果如第 8 行所示，AP 和 mAP@0.5 分别达到 65.6%和 90.7%，与基线模型相比分别提升了 2.4 个百分点和 1.2 个百分点，这说明本文方法能有效提升模型对交通图像目标的检测精度。

表 4　在 KITTI 数据集上的消融实验结果

Table 4　Ablation experiment results on the KITTI dataset

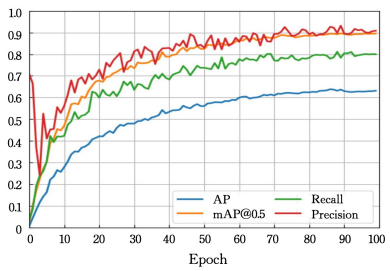
TFF	HDCFE	GeoCIoU	*AP*/%	*mAP*/@0.5%	*Params*	*FLOPs*	*FPS*
×	×	×	63.2	89.5	2.6×10^6	6.3×10^9	672
√	×	×	64.5	90.0	2.7×10^6	6.8×10^9	625
×	√	×	64.4	90.1	2.9×10^6	10.2×10^9	587
×	×	√	64.8	89.9	2.6×10^6	6.4×10^9	652
√	√	×	65.0	90.4	2.9×10^6	10.6×10^9	557
×	√	√	65.1	90.3	2.9×10^6	10.2×10^9	561
√	×	√	65.2	90.1	2.7×10^6	6.8×10^9	593
√	√	√	**65.6(+2.4)**	**90.7(+1.2)**	**2.9×10^6**	**10.6×10^9**	**529**

图 7 为 ETF-YOLO11n 方法在 KITTI 数据集上的可视化结果，第 1 行为原图，第 2 行为热力图，第 3 行为最终结果图。结果显示，本文提出的 ETF-YOLO11n 方法在目标特征提取以及检测方面都展示出良好性能。

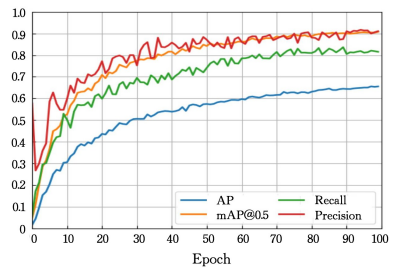
图 7　ETF-YOLO11n 在 KITTI 数据集上的可视化结果

Fig. 7　Visualization results of ETF-YOLO11n on KITTI

如图 8 所示，本文将 ETF-YOLO11n 的各项精度与基线模型 YOLO11n 进行了可视化比较，其中图 8(a)是 YOLO11，图 8(b)是 ETF-YOLO11n，横坐标代表训练的轮次，而纵坐标代表检测精度，不同颜色曲线代表不同的精度评价指标。可以直观看到，本文方法的召回率、精确率、AP、mAP@0.5 等指标都有所上升。



(a) YOLO11n



(b) ETF-YOLO11n

图 8　KITTI 数据集上的精度曲线

Fig. 8　Accuracy curves on KITTI dataset

4.7　泛化性验证

(1)在不同数据集上的泛化性验证

本文使用 COCO-Traffic 数据集对 ETF-YOLO11n 模型

在不同数据集上的泛化性进行了验证。如表 5 所列，在该数据集上 AP 和 mAP@0.5 的值分别达到了 42.5%和 59.8%，与基线模型相比，提升了 1.3 个百分点和 1.2 个百分点。与 KITTI 数据集上的结果相似，无论精度还是参数量，本文方法在 COCO-Traffc 上的结果都优于大多数先进的目标检测算法。在 COCO-Traffic 数据集上的可视化结果如图 9 所示。

表 5 COCO-Traffic 数据集上的泛化性实验结果

Table 5 Generalization experimental results on COCO-Traffic

Methods	Backbone	*AP*/%	*mAP*@0.5/%	*Params*
SSD[2]	VGG16	48.4	28.7	—
RetinaNet[30]	VGG16	36.1	56.4	—
YOLOv5s[3]	CSPDarkNet53	36.9	55.5	7.1×10^6
YOLOXs[5]	CSPDarkNet53	39.0	59.7	9.0×10^6
YOLOv7-tiny[6]	CSPDarkNet53	41.7	64.2	6.0×10^6
YOLOv8n[7]	CSPDarkNet53	38.9	54.8	3.0×10^6
YOLO11n (baseline)[9]	CSPDarkNet53	41.2	58.6	2.6×10^6
ETF-YOLO11n (Ours)	CSPDarkNet53	**42.5 (+1.3)**	**59.8 (+1.2)**	**2.9×10^6**

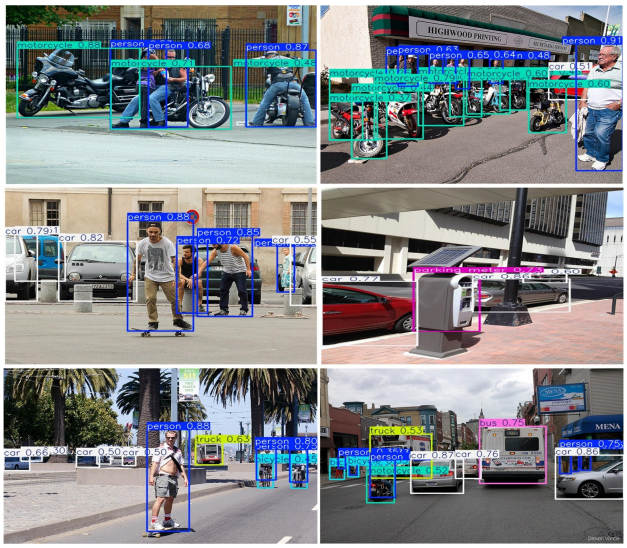


图 9 ETF-YOLO11n 在 COCO-Traffic 数据集上的可视化结果

Fig. 9 Visualization results of ETF-YOLO11n on COCO-Traffic

(2)在模型上的泛化性验证

如表 6 所列，将所提出的 ETF-YOLO 方法应用于 YOLOv8n，并在 KITTI 上进行验证。结果显示，ETF-YOLOv8n 的 AP 和 mAP@0.5 分别达到 66.9%和 91.5%，提升了 1.5 个百分点和 1.2 个百分点。由此可见，本文方法在不同模型上都表现出很好的泛化性。

表 6 在 YOLOv8 模型上的泛化性实验结果

Table 6 Generalization experimental results on YOLOv8 model

Methods	*AP*/%	*mAP*@0.5/%	*Params*
YOLOv8n[7]	65.4	90.3	3.0×10^6
ETF-YOLOv8n(Ours)	**66.9(+1.5)**	**91.5(+1.2)**	**3.3×10^6**

4.8 模型复杂度分析

交通图像目标检测需兼顾准确率和精度，如表 7 所列，本文算法的参数量和计算量约为 DETR 的 1/15，却有更好的检测效果；检测精度虽不如 YOLOv9s，但它引入辅助分支增加检测头，参数量和计算量是本文算法的 4 倍左右；与 YOLOv5n，YOLOX-tiny，YOLOv7-tiny，YOLOv8n 等算法相比，本文算法有最高的检测精度。

表 7 模型复杂度对比

Table 7 Comparison of model complexity

Methods	*AP*/%	*Params*	*FLOPs*
DETR[11]	54.9	41.3×10^6	101.3×10^9
YOLOv5n[3]	63.3	2.5×10^6	7.1×10^9
YOLOX-tiny[5]	55.1	6.5×10^6	6.5×10^9
YOLOv7-tiny[6]	41.7	6.0×10^6	13.9×10^9
YOLOv8n[7]	65.4	3.0×10^6	9.0×10^9
YOLOv9s[8]	69.1	9.6×10^6	38.7×10^9
ETF-YOLO11n(Ours)	**65.6**	**2.9×10^6**	**10.6×10^9**

本文算法 ETF-YOLO11n 的参数量相比基线模增加 0.3×10^6，计算量增加4.0×10^9 FLOPs。然而，这一改变是基于交通目标检测场景复杂性与精准性需求权衡后的结果。综上，本文算法在复杂度和检测精度中做到了较好平衡，有较高的应用价值。

4.9 小结

本章通过设置公平实验条件，对比 GeoCIoU 不同权重系数确定最佳值，将所提方法与前沿算法进行对比，开展消融实验验证本文策略的有效性，且验证了模型泛化性、分析复杂度，证明了 ETF-YOLO11n 性能优，实用性强。

结束语 为研究复杂交通场景中的目标检测中特征信息不明显、易丢失等问题，本文提出了基于特征加强的单阶段目标检测算法 ETF-YOLO11n，实现交通场景图像目标的高性能检测。设计 TTF 模块，融合不同尺度特征信息，进一步提升模型对交通目标的捕捉能力；在模型的 Neck 中，设计并引入了 HDCFE 模块，以增强模型上下文全局特征提取能力；设计 GeoCIoU 函数，使模型训练过程更专注于高质量锚框。实验结果显示，本文方法能有效提升模型对交通图像中目标的检测能力，优于多数先进的目标检测算法，在不同数据集和不同模型上都表现出较强的泛化能力。

对未来工作的展望：(1)无人驾驶等实际交通监控场景需要实施传输数据，对模型复杂度提出更高要求，本文方法计算量有一定增长，之后将精简冗余结构，在不损失检测精度的前提下削减参数量与计算量，并尝试在无人驾驶汽车上进行部署；(2)鉴于交通图像里不同类别样本数量存在差异，后续将深入探究如何在目标检测流程里实现困难样本的自动甄别与选取，减少样本不均衡导致的检测偏差。

参 考 文 献

[1] CHEN H, WAN W W, MATSUSHIT A, et al. Automatically Prepare Training Data for YOLO Using Robotic In-Hand Observation and Synthesis [J]. IEEE Transactions on Automation Science and Engineering, 2024, 21(3): 4876-4982.

[2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector [C]// European Conference on Computer Vision. Springer, 2016: 21-27.

[3] JOCHER G. YOLOv5[EB/OL]. (2020-06-28) [2024-11-17]. https://github.com/ultralytics/yolov5.

[4] LI C Y，LI LL，JIANG H L，et al. YOLOv6：A Single-Stage Object Detection Framework for Industrial Applications[J]. arXiv：2209.02976，2022.

[5] ZHENG G，LIU S T，WANG F，et al. YOLOX：Exceeding YOLO Series in 2021[J]. arXiv：2107.08430，2021.

[6] WANG C Y，BOCHKOVSKIY A，LIAO H Y M. YOLOv7：Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2023：7464-7475.

[7] JOCHER G. YOLOv8[EB/OL]. (2023-01-10) [2024-11-17]. https://github.com/ultralytics/yolov8.

[8] WANG C Y，YEH I H，LIAO H Y，et al. Yolov9：Learning What You Want to Learn Using Programmable Gradient Information[C]// European Conference on Computer Vision. Springer，2024：1-21.

[9] JOCHER G. YOLO11[EB/OL]. (2024-09-30) [2024-11-17]. https://GitHub-ultralytics/ultralytics：Ultralytics YOLO11.

[10] GIRSHICK R，REN S，HE K，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.

[11] CARION N，MASSA F，SYNNAEVE G，et al. End-to-end Object Detection with Transformers[C]// European Conference on Computer Vision. Springer，2020：213-219.

[12] Government of the People's Republic of China. MotorVehicles Nationwide Reached 440 Million in the First Half of 2024 [EB/OL]. (2024-07-09) [2024-11-17]. https://www.gov.cn/lianbo/bumen/202407/content_6961935.htm.

[13] GEIGER A，LENZ P，STILLER C，et al. Vision Meets Robotics：The Kitti Dataset[J]. The International Journal of Robotics Research，2013，32(11)：1231-1237.

[14] LIN T Y，DOLLAR P，PIOTR G，et al. FeaturePyramid Networks for Object Detection [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2017：936-944.

[15] LIU S，QI L，QIN H F，et al. Path Aggregation Network for Instance Segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2018：8759-8768.

[16] LAN L X，CHI M Y. Remote Sensing Change Detection Based on Feature Fusion and Attention Network [J]. Computer Science，2022，49(6)：193-198.

[17] LU H T，FANG M Y，QIU Y X，et al. An Anchor-Free Defect Detector for Complex Background Based on Pixelwise Adaptive Multiscale Feature Fusion[J]. IEEE Transactions on Instrumentation and Measurement，2023，72：1-12.

[18] YU J H，JIANG Y N，WANG Z Y. et al. Unitbox：An advanced object detection network [C]// International Conference on Multimedia. ACM，2016：516-520.

[19] REZATOFIGHI H，HAMID T，NATHAN G，et al. Generalized Intersection over Union：A Metric and a Loss for Bounding Box Regression[C]// IEEE International Conference on Computer Vision. IEEE，2019：658-666.

[20] ZHENG Z H，WANG P，LIU W，et al. Distance-IoU loss：Faster and Better Learning for Bounding Box Regression [C]// AAAI Conference on Artificial Intelligence. AAAI，2020：12993-13000.

[21] GEVORGYAN Z. SIoU Loss：More Powerful Learning for Bounding Box Regression [J]. arXiv：2205.12740，2022.

[22] LUO X，CAI Z，SHAO B，et al. Unified-IoU：For High-Quality Object Detection [J]. arXiv：2408.06636，2024.

[23] HU J，SHEN L，SUN G. Squeeze-and-excitation Networks [C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE，2018：7132-7141.

[24] WOO S P，LEE J C，KWEON J Y. CBAM：Convolutional Block Attention Module [C]// European Conference on Computer Vision. Springer，2018：3-19.

[25] REN X X，LI M，LI Z H，et al. Curiosity-driven Attention for Ano-maly Road Obstacles Segmentation in Autonomous Driving [J]. IEEE Transactions on Intelligent Vehicles，2022，8(3)：2233-2243.

[26] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is All You Need[J]. arXiv：1706.03762，2017.

[27] ANTHONY F，DANIEL G，KYROLLOS Y Y，et al. LookHere：Vision Transformers with Directed Attention Generalize and Extrapolate[J]. arXiv：2405.13958，2024.

[28] GAO L Y，QU Z，WANG S Y，et al. A Lightweight Neural Network Model of Feature Pyramid and Attention Mechanism for Traffic Object Detection[J]. IEEE Transactions on Intelligent Vehicles，2024，9(2)：3422-3435.

[29] WANG S Y，QU Z，GAO L Y，et al. Multi-spatial Pyramid Feature and Optimizing Focal Loss Function for Object Detection [J]. IEEE Transactions on Intelligent Vehicles，2023，9(1)：1054-1065.

[30] LIN T S，GOYAL P，GIRSHICK R，et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，99：2999-3007.

XIA Shufang，born in 1980，Ph.D. Her main research interests include computer vision，machine learning and artificial intelligence.

QU Zhong，born in 1972，Ph.D，professor. His main research interests include computer vision，machine learning and artificial intelligence.

（责任编辑：喻藜）