人机交互下的知识嵌入与小样本目标识别技术研究

人工智能专业　　章星宇
指导教师　　吴金建

[摘　要] 本文针对小样本目标识别开展两方面的研究工作：一是提出了一个融合道路信息的目标检测网络，针对无人机采集的数据，利用 D-LinkNet 算法对图像进行道路分割，并将道路信息作为先验知识嵌入目标检测网络中，引导模型识别道路中的汽车目标，有效提升了模型的检测识别性能。二是搭建了一套基于人机交互的目标检测系统。通过人工干预的方式，对检测模型的输出进行检查和纠正，纠正完的数据会重新输送到模型中，指导模型进行二次训练。该系统确保了检测结果输出的可靠性，增强了模型的可迭代能力。
[关键词] 目标检测 道路分割 知识嵌入 人机交互

[**Abstract**] This paper focuses on two research aspects for small-sample object recognition: First, a detection network that incorporates road information is proposed. Specifically, for data collected by drones, the D-LinkNet algorithm is used for road segmentation in images. The road information is embedded as prior knowledge into the object detection network, guiding the model to recognize car objects on the road. This approach effectively improves the detection and recognition performance of the model. Second, a human-computer interaction-based object detection system is developed. Through manual intervention, the system checks and corrects the outputs of the detection model. The corrected data is fed back into the model for retraining, guiding the model in a second training phase. This system ensures the reliability of the detection results and enhances the model's iterability.
[**Key Words**] Object Detection　Road Segmentation　Knowledge Embedding　Human-Computer Interaction

一、引言

深度学习在目标检测任务中取得了显著的成果，但是相关模型通常依赖于大规模高质量训练数据。在很多场景下，训练数据的标注成本很高，难以获得充足的训练数据，导致现有模型在样本量不足时，往往出现难以收敛，效果不佳的情况。

在实际生活中，人类对新物体有很好的学习能力。例如，给出一个物体的少量照片，人类通过快速学习之后，就能在现实中识别出这个物体。受此启发，研究人员希望使用少量标记样本训练模型来得到具备一定泛化能力的检测模型，该任务被称为小样本目标识别[1]。

目前，小样本目标识别主要是在普通目标检测的基础上使用一些优化策略对其改进，主要包括基于迁移学习的改进、基于多尺度学习的改进和基于上下文学习的改进。

迁移学习是指先通过包含大量标注数据集对检测模型进行预训练，然后再通过一定参数微调，使其应用于小样本目标数据集进行训练和检测。Wang 等人[2]提出了 TFA 模型，模型在第一个阶段使用 Faster R-CNN 在大量标注样本上进行训练，第二个阶段在不改变模型参数的前提下使用余弦相似度对分类器进行微调。

多尺度学习指的是在深度神经网络提取图片信息时，将深层的语义信息和浅层的表征信息

进行结合。Adelson 等人[3]提出的特征金字塔结构，能够在不同尺度下，对输入图片进行检测。Lin 等人[4]提出了 FPN 结构，将深层语义特征融合进浅层特征图，丰富了目标的空间特征。Liu 等人[5]进一步提出了 PAN 结构，将深层语义特征和浅层表征特征进行融合，有效提升了网络的检测性能。

上下文学习指的是深度神经网络在学习中，不仅仅考虑目标本身的特征，进一步将“目标与场景”，“目标与目标”之间的关系信息纳入检测过程。Liu 等人[6]提出一种结构推理网络 SIN，考虑了场景与目标之间的关系，有效提升了检测的性能。Xu 等人[7]提出了 Reasoning-RCNN 网络，通过构建知识图谱来编码目标之间的关系，并利用先验关系来影响检测效果。

本文使用的是无人机拍摄的汽车目标数据集，基于数据集的特点，针对所需检测识别的汽车目标，将“汽车目标基本只出现在道路等可行区域内”这一特点作为先验信息。基于此先验信息，利用 D-LinkNet 算法[8]对无人机遥感图像进行道路分割。之后，基于 YOLOv5 目标检测算法[9]，将提取的道路信息和检测网络提取的特征进行融合，输出检测结果。实验表明，该方法有效提升了模型对小样本目标的检测性能。同时，本文搭建了基于人机交互的目标检测系统，利用人工干预的方式对模型的输出结果进行检查和纠正，并指导模型进行二次训练。提升了模型输出结果的可靠性，增强了模型的可迭代能力。

二、基于知识嵌入的小样本目标检测模型

本文基于 YOLOv5 目标检测算法，提出了一个嵌入道路知识的小样本目标检测模型。模型的整体框架如图 1 所示。

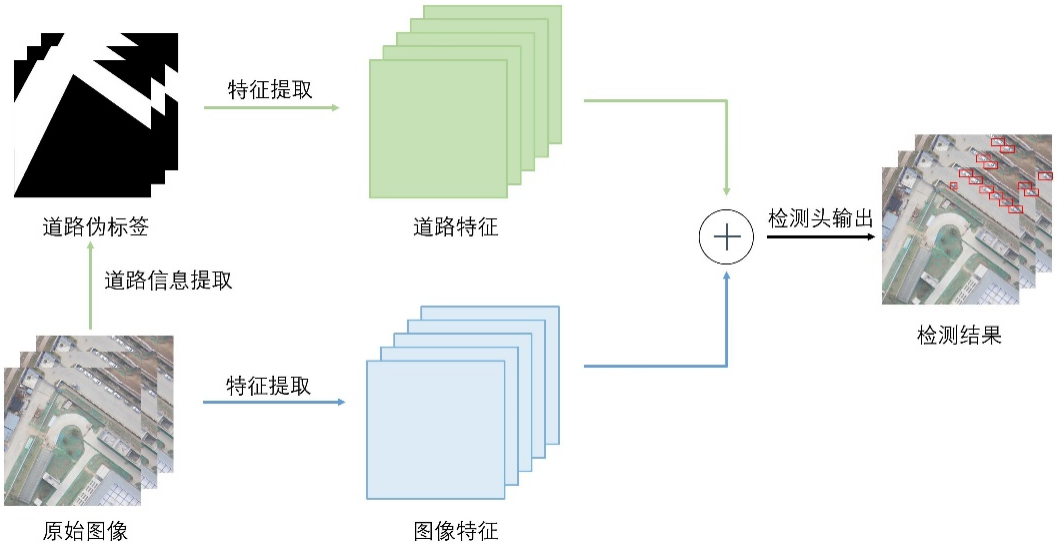

图 1 嵌入道路知识的小样本目标检测模型框架

无人机图像在输入网络之后，一方面通过 D-LinkNet 进行道路信息的提取，另一方面直接通过 YOLOv5 的骨干网络提取特征。提取的道路信息再通过特征提取后，将该特征和原图提取的特征进行相加融合，最后送入检测头输出结果。整套框架主要包含道路信息提取和道路信息嵌入两个部分。

1. 道路信息提取

本文所使用的数据集是在陕西省西安市王曲村附近用无人机采集的可见光影像，图像大小为 7360×4912 像素，分辨率约为 7 厘米/像素。所需检测的是汽车单类目标，目标平均尺寸约为 60×30 像素，部分数据如图 2 所示。

图 2 汽车目标数据集预览

为了提升道路分割的效果，本文使用分割算法 D-LinkNet 在卫星遥感数据集 DeepGlobe[10]上进行预训练，该数据集图像的分辨率为 0.5 米/像素。由于无人机遥感图像和卫星遥感图像存在高度差异，无人机遥感图像的分辨率通常在 3 厘米/像素-8 厘米/像素范围之间。因此，为了使 D-LinkNet 能够对无人机遥感图像进行道路分割，通过下采样的方式，将无人机遥感图像缩放到分辨率为 0.5 米/像素左右，从而实现道路信息的分割提取。道路信息提取的流程如图 3 所示。

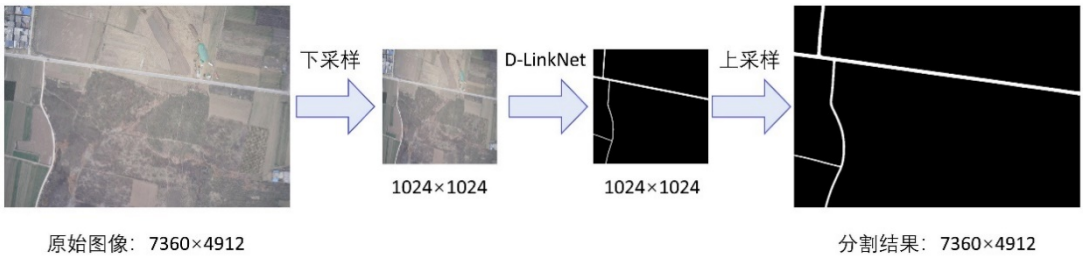

图 3 道路信息提取流程图

2. 道路信息嵌入

道路信息嵌入是指将该提取到的道路信息进一步提取特征，然后将原始图像提取到的特征和道路特征进行融合。特征提取网络包含三个卷积层结构和两个 C3 结构，如图 4 所示。

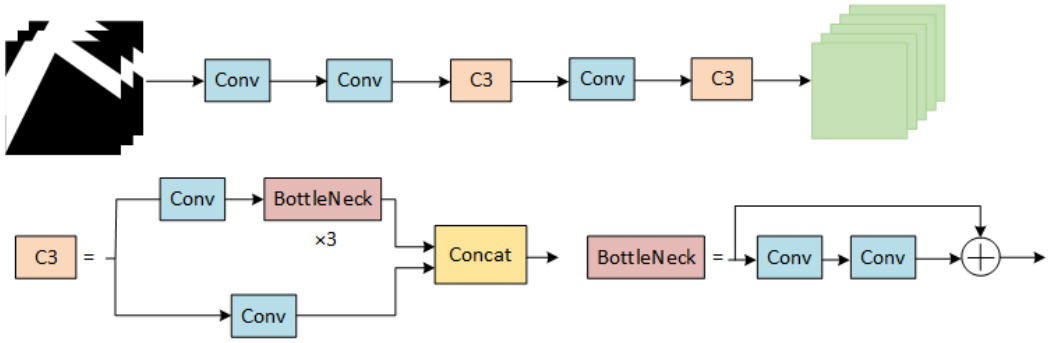

图 4 特征提取网络结构图

输入网络的特征向量每经过一层卷积层后，向量的尺度会缩小为原始的一半，通道数会增加至原始的两倍。同时，网络中添加了两个 C3 结构，通过堆叠残差结构来增加网络的深度，从而是网络能够提取到模型更深层的语义信息。一个 C3 结构的计算公式如公式 1 所示。

$$y = Concat(W_1 x + b_1, B((W_2 x + b_2)) \tag{1}$$

式中，x表示输入的特征向量，W表示卷积层权重矩阵，b表示卷积层偏置，y表示输出的特征向量。

在提取到道路特征和原始图像特征之后，模型会将道路特征和原始图像特征两部分进行加权

融合，融合的计算公式如公式 2 所示。

$$y = \lambda_1 y_1 + \lambda_2 y_2 \tag{2}$$

式中，y_1表示道路特征向量，y_2表示道路特征向量，λ_1和λ_2表示权重，用来衡量两者的重要程度，y表示融合之后的特征向量。

3. 实验评价指标

在目标检测任务中，通常根据以下这些指标来衡量算法的性能。

1）精确率

精确率为检测结果为正样本的正样本数占全部检测为正样本总数的比例，精确率计算方式如公式 3 所示。

$$P = \frac{T_P}{T_P + F_P} \tag{3}$$

式中，T_P为被正确识别的正样本数，F_P为被错误识别的正样本数。

2）召回率

召回率为检测结果为正样本的正样本数占实际全部正样本数的比例，召回率计算方式如公式 4 所示。

$$R = \frac{T_P}{T_P + F_N} \tag{4}$$

式中，T_P为被正确识别的正样本数，F_N为实际的正样本检测为负样本的数量。

3）平均准确率均值

平均准确率均值为各类别平均准确率的均值，平均准确率为 PR 曲线下的面积。平均准确率均值计算方式如公式 5 所示。

$$mAP = \frac{1}{N}\sum_{k=1}^{N} P(K)\Delta R(k) \tag{5}$$

式中，N 表示类别总数，对于本项目仅包含一类汽车类别，N 取值为 1。$P(K)$为不同置信度下的准确率，$\Delta R(k)$表示不同置信度下召回率插值。

根据 IOU 阈值的选取，mAP 通常细分为$mAP_{0.5}$和$mAP_{0.5:0.95}$。$mAP_{0.5}$是指 IOU 阈值设定为 0.5，大于 0.5 的预测框作为正样本。$mAP_{0.5:0.95}$是指 IOU 阈值从 0.5 一直取到 0.95，每隔 0.05 计算一次 mAP 值，最后计算所有 mAP 的均值。

4）单帧检测耗时

单帧检测耗时为一张图片检测完成需要的时间，用以衡量模型的检测速度。

4. 实验结果

数据集图像共包含 5016 张 1280×1280 像素大小的图像。本实验随机将 4500 张图像划分成训练集，其余图像用作测试集。本实验使用 YOLOv5m 作为基准模型，主要衡量原始模型和融合道路信息之后的模型性能。

1）收敛性分析

如图 5 所示，图中展示了 Loss 和$mAP_{0.5:0.95}$随 epoch 的变化情况，蓝色的曲线为 YOLOv5 原始模型，橙色的曲线为融合道路信息之后的检测模型。从图中可以发现，在经历 100 轮迭代后，Loss 和 mAP 曲线均变得平缓，可见两种模型都已经收敛。同时可以看出，融合道路信息的检测

模型 Loss 整体更小，$mAP_{0.5:0.95}$整体更大，说明收敛性更好。

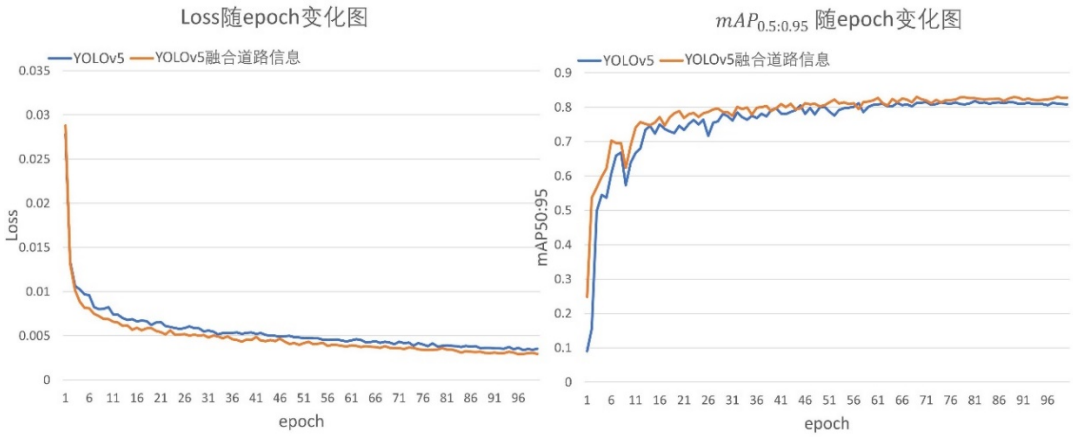

图 5 Loss 和$mAP_{0.5:0.95}$随 epoch 变化图

2）模型性能评估

改进前后的两种模型相关性能指标如表 1 所示。

表 1 不同算法性能对比

模型名称	P	R	$mAP_{0.5}$	$mAP_{0.5:0.95}$	单帧检测耗时
YOLOv5	0.940	0.863	0.923	0.823	0.43s
YOLOv5 融合道路信息	0.958	0.878	0.930	0.830	0.56s

从实验结果可以看出，将道路信息融合进 YOLOv5 之后，虽然检测速度有略微下降，但准确率、召回率等各项指标都有所提升。

为了进一步探究道路信息对模型的提升效果，本文选取了一处局部场景，来对比改进前后的模型检测效果，如图 6 所示。

YOLOv5　　YOLOv5融合道路信息

图 6 不同模型检测结果对比图

左图为 YOLOv5 原始模型的检测结果，可以看到，模型误将屋顶上形状类似车的杂物识别成了车辆。而在右图融合道路信息之后，模型更聚焦于道路上的目标信息，从而正确地检测出了目标，排除了干扰。由此可见，道路信息的融合对模型的检测性能有所提升。

三、基于人机交互的目标检测系统

由于目标检测模型的输出结果存在不可靠性，因此本文提出了一种人机交互的检测方式，通过人来对模型的输出进行检查和纠正，并对模型的学习进行指导。本文搭建了一套可视化系统，可使用户在无需修改源代码的前提下，完成模型训练到检测的整个流程。

1. 人机交互机制设计

图 7 展示了人机交互机制的设计流程，图中红色目标框表示模型的检测结果，此模型用于检测图中的车辆目标，可以看到模型出现了漏检情况。此时通过人工干预，将模型漏检的车辆目标标注出来，图中用绿色框表示，从而得到正确的输出结果。同时，此结果可以再次作为数据集输送到模型中进行二次训练，从而使模型再次遇到类似场景时，提升识别的准确性。

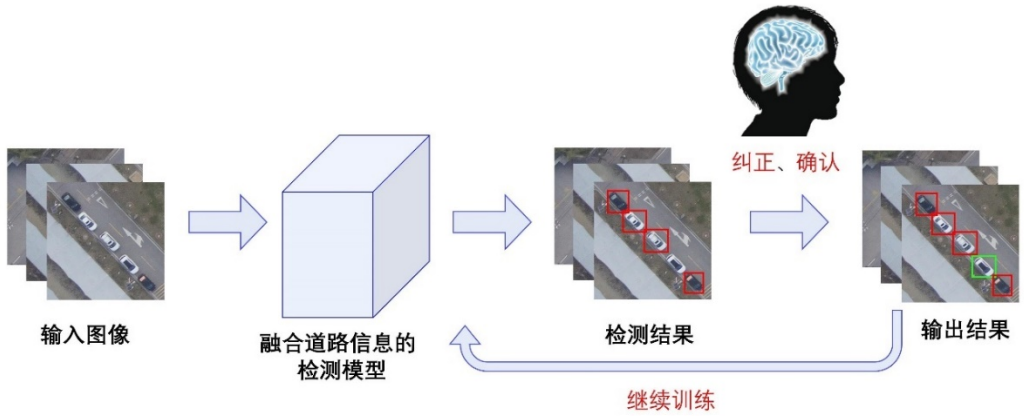

图 7 人机交互机制设计图

2. 系统功能模块设计

系统的基础界面如图 8 所示。界面左侧为一系列可点击按钮，实现不同功能，界面中心用来显示主要图片，右侧用于切换图片。

图 8 人机交互机制设计图

系统主要包含以下五个功能模块：

•加载数据。点击左侧按钮可以批量导入图片数据和模型数据。

•目标检测。点击“目标检测”按钮，可以直接加载导入的模型数据对图片数据进行检测。

•图片裁剪。若图片的尺寸较大，可以在导入图片之后点击“图片裁剪”按钮，图像会被裁剪为 1280×1280 像素大小的图像块。

•人机交互标注。点击“人机交互标注”按钮，可以在弹出的子界面中，根据检测结果，调整目标框的位置和类型，并输出最终结果和模型训练标签。

•模型训练。点击“模型训练”按钮，可以在弹出的子界面中，调整相关的模型训练超参数，对模型进行训练。

3. 系统操作流程

人机交互检测系统的操作流程如下。

（1）首先导入数据和模型，直接进行目标检测，下方进度条会显示检测进度，检测完成后，可以直接在主界面看到检测结果，如图 9 所示。

图 9 目标检测图层展示

点击左侧的图层栏，可以切换不同的图层，图 10 显示的是道路分割图层。

图 10 道路分割图层展示

（2）如果图像尺寸较大，点击“图片裁剪”按钮，下方进度条会显示检测进度。裁剪完成后，点击“人机交互标注”按钮，进入到子界面。

（3）如图 11 所示，在子界面中，点击“打开目录”，加载裁剪好的图像，之后点击“自动标注”，模型会自动进行检测，输出检测标签。此时，可以通过左侧其它功能键来实现对标注结果的修改。修改过程中，系统会自动生成标签，以供后续训练。

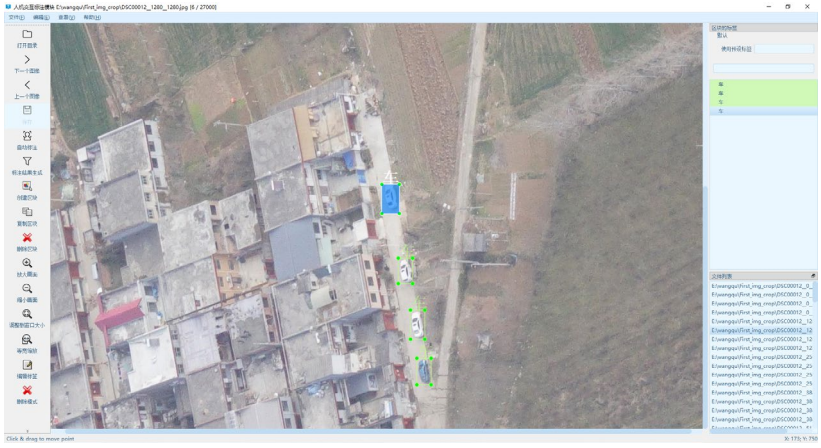

图 11 人机交互功能展示

（4）最后点击主界面上的“模型训练”按钮，在弹出的子界面上，可以修改常用的一些超参数，修改完成后点击下方按钮，模型会开始训练，可以在最下方的窗口中查看训练进度和日志信息。

图 12 模型训练模块界面

四、总结

本文针对小样本目标检测精度低，识别难度大等问题，提出了融合道路信息的目标检测网络。对于所需检测识别的汽车目标，构建了“汽车目标基本只出现在道路等可行区域内”的先验信息。基于此先验信息，使用 D-LinkNet 算法对无人机遥感图像进行道路分割，并设计相关网络提取道路特征，将道路信息嵌入到检测网络中。同时，本文提出了一套基于人机交互的目标检测系统。通过人工干预的方式，对检测模型的输出进行检查和纠正，纠正完的数据会重新输送到模型中，指导模型进行二次训练。该系统确保了检测结果输出的可靠性，增强了模型的可迭代能力。

参考文献

[1] 姜钧舰,刘达维,刘逸凡,任酉贵,赵志滨.基于孪生网络的小样本目标检测算法[J/OL].计算机应用:1-7[2023-05-19].http://kns.cnki.net/kcms/detail/51.1307.TP.20230419.1150.002.html

[2] Wang X, Huang T E, Darrell T, et al. Frustratingly simple few-shot object detection [EB/OL]. [2022-08-11].

[3] Adelson E H,Anderson C H,Bergen J R,et al.Pyramid methods in image processing[J].RCA Engineer,1984,29(6): 33 41.

[4] 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. 2017: 2117-2125.

[5] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[6] Liu Y, Wang R, Shan S, et al. Structure inference net: Object detection using scene-level context and instance-level relationships[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6985-699

[7] Xu H, Jiang C H, Liang X, et al. Reasoning-RCNN: Unifying adaptive global reasoning into large-scale object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 6419-6428.

[8] Zhou L, Zhang C, Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition Workshops. 2018: 182-186.

[9] Jocher,G.(2020). YOLOv5 by Ultralytics [Computer software]. https://doi.org/10.5281/zenodo.3908559

[10] Demir I, Koperski K, Lindenbaum D, et al. Deepglobe 2018: A challenge to parse the earth through satellite images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 172-181.