班 级 1920031
学 号 19200300029

西安电子科技大学

本科毕业设计论文

题 目 人机交互下的知识嵌入与小样本目标识别技术研究

学 院 人工智能学院

专 业 人工智能

学生姓名 章星宇

导师姓名 吴金建

摘要

深度学习在目标检测识别领域取得了显著的成果，然而，在数据样本量不足的情况下，相关模型无法训练出有效的特征提取器，效果较差。为了提高小样本目标检测结果的准确性和可靠性，本文开展了基于知识嵌入和人机交互的小样本目标识别算法研究，具体内容包括以下三个方面：

（1）为了增加目标检测模型所获取的信息量，对于数据集中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。基于此信息，使用 D-LinkNet 对无人机遥感图像进行道路分割：先使用卫星遥感影像数据集 DeepGlobe 对 D-LinkNet 进行预训练，然后对无人机影像进行下采样，使其接近训练数据的分辨率，再输入网络中实现道路分割，得到原始图像的道路信息。

（2）为了提升目标检测模型对小样本目标的检测识别能力，提出了一个融合道路信息的检测网络。具体方式是先对 D-LinkNet 分割得到的图像道路信息进行特征提取，然后将道路特征和 YOLOv5 本身提取到的图像特征进行融合，最后输出检测结果。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。

（3）为了提升目标检测模型对小样本目标的检测结果的可靠性，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

Deep learning achieves significant advancements in the field of object detection and recognition. However, the related models fail to train effective feature extractors when the data sample size is limited. In order to improve the accuracy and reliability of few sample detection, this paper conducts research on few sample recognition algorithms based on knowledge embedding and human-computer interaction. The specific contents include the following three aspects:

(1) To increase the information acquired by the object detection model, the characteristic that "car objects mostly appear in feasible areas such as roads" is taken as prior knowledge for the targets in the dataset. Based on the knowledge, D-LinkNet is used for road segmentation in drone remote sensing images. First, D-LinkNet is pretrained using the DeepGlobe dataset of satellite remote sensing imagery. Then, the drone images are downsampled to approximate the resolution of the training data and input into the network to obtain road information from the original images.

(2) To enhance the detection and recognition capability of the object detection model for few samples, this paper proposes a detection network that integrates road information. The specific approach involves extracting features from the road information obtained by D-LinkNet and fusing them with the image features extracted by YOLOv5. The detection results are generated based on this fusion. Experimental results demonstrate that this method effectively improves the detection performance of the model.

(3) To enhance the reliability of the object detection results for few samples, this paper presents a human-computer interaction-based detection system. Through manual intervention, the outputs of the detection model are checked and corrected. The corrected data is fed back into the model for retraining, guiding the model for iterative improvement. This system ensures the reliability of the detection result outputs and enhances iterative ability of the model.

Keywords: Object Detection, Road Segmentation, Knowledge Embedding, Human-Computer Interaction

目录

第一章 绪论

1.1 研究背景及意义

目标检测是指从图像或视频中定位和识别目标的技术，是计算机视觉领域中的一个重要研究方向。目前，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]。

小样本目标识别在现实场景中非常常见，比如在医学上的病灶检测[2]和稀有动物检测[3]的场景中，数据集的获取成本高昂，且包含有效目标的样本量十分稀少，对相关目标的检测识别有助于提升医疗水平和生态环境保护。在无人机遥感场景中，受限于无人机的续航能力，能够获取到目标样本同样不足。在无人机拍摄的图像中，准确检测出车辆，行人等位置和信息，有助于政府机构遏制毒品和人口贩运。

综上所述，小样本的检测识别具有重大的研究意义和广泛的应用价值。对小样本的检测识别有助于推动目标检测领域的发展，扩充目标检测在现实生活中的应用场景，提升中国的科技创新水平，助力中国全面步入智能化时代。

1.2 国内外研究现状

小样本目标识别的主要难点是样本较少，所携带的信息量不足。因此，如何利用有限的目标样本，更大程度地提取样本中所包含的信息，成为小样本识别的关键问题。针对此问题，国内外研究学者进行了一系列探索。从算法演进的历史出发，大致可分为传统的目标检测和基于深度学习的目标检测。

1.2.1 传统的目标检测

传统的目标检测思路是先通过滑动窗口的方式找到图片中的感兴趣区域（ROI），然后利用 HOG[4]、SIFT[5][5]等特征提取算法对每一个区域进行特征提取，最后根据这些区域提取的特征，使用 SVM[6]、Adaboost[7]等分类算法进行分类，最终得到目标的位置和类别信息。

不过传统的目标检测算法往往依赖人工设计特征[8]，效率低且精度不高，存在较大的局限性。

1.2.2 基于深度学习的目标检测

随着深度学习技术的不断发展，逐渐涌现出一大批基于深度学习的目标检测算法。针对小样本目标的特点，这些算法的优化思路大致可归纳为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[9]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[10]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[11]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[12]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[13]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[14]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。

1.3 本文的主要工作

本文主要研究内容包括：

（1）为了增加目标检测模型所获取的信息量，本文基于数据集的特点，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，本文使用 D-LinkNet 对无人机遥感图像进行道路分割。本文先使用卫星遥感影像数据集 DeepGlobe 对 D-LinkNet 进行预训练，然后对无人机影像进行下采样，使其接近训练数据的分辨率，再输入网络中实现道路分割，得到原始图像的道路信息。

（2）为了提升目标检测模型对小样本的检测识别能力，本文提出了一个融合道路信息的检测网络。具体方式是先对 D-LinkNet 提取到的图像道路信息进行特征提取，然后将道路特征和 YOLOv5 本身提取到的图像特征进行融合，最后输出检测结果。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。

（3）为了提升目标检测模型对小样本的检测结果的可靠性，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。

1.4 本文的内容组织

本文的主要内容安排如下：

第一章：概述了小样本识别的研究背景、研究意义，从传统的目标检测和基于深度学习的目标检测两个方面介绍了国内外研究现状，并阐述了本文的主要工作与结构安排。

第二章：首先介绍深度学习和卷积神经网络的基础知识，然后分析目标检测和图像分割的常用算法内容，为后续章节奠定理论基础。

第三章：首先详细分析实验中所涉及的 D-LinkNet 道路分割算法和 YOLOv5 目标检测算法，然后详细阐述将道路信息嵌入到检测网络中的方式，最后提出人机交互实现小样本准确检测机制的相关设计。

第四章：对嵌入道路信息的检测网络进行实验，比较嵌入信息前后的效果，并

对实验结果进行具体分析。

第五章：详细阐述人机交互的目标检测系统的功能模块，并介绍系统使用的相关流程。

第六章：对本文进行全面的总结，阐述本文所取得的成果，同时指出工作的不足之处及后续研究工作的努力方向。

第二章 目标检测和图像分割相关知识

2.1 深度学习概述

2012 年，AlexNet[15]的提出，将深度神经网络带入了爆发期。得益于其强大的学习能力，很快就在图像检索、机器翻译、语音识别、自动驾驶等各领域得到广泛应用，掀起一波热潮。

追溯神经网络的发展历程，起初它是受到生物神经元的启发提出来的。在生物神经系统中，神经元细胞之间的联络依赖于突触结构。信息通过突触，不断地在一层层神经元细胞中传递，最终形成了一个庞大的神经网络。人工神经网络模拟了生物神经网络的架构，通过堆叠神经元节点，形成输入层、隐藏层、输出层等各种神经网络层级。每层之间通过带有不同权重的边进行相连，这些权重会随着网络的训练不断得到更新，最终使整个网络的性能达到最优。图 2.1 展示了一个人工神经网络的基础架构。

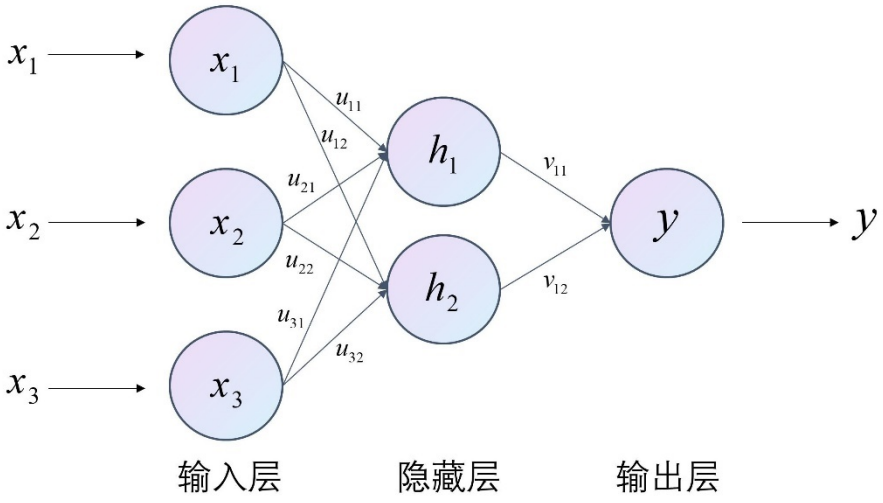

图2.1 人工神经网络架构图[15]

神经网络的训练包括正向传播和反向传播两个阶段。在正向传播中，数据通过输入层，加权处理后进入隐藏层，经过层层处理后，最后映射到输出层，输出结果。该结果和已知的正确结果进行对比，通过相应损失函数计算得到两者的误差，然后进入到误差的反向传播阶段。

在反向传播中，误差经过网络的各神经元时，通常使用优化器来更新神经元连接边的权重参数，从而使误差无限最小化。常见的优化器主要包括随机梯度下降法

（SGD）[16]、动量梯度下降法（Momentum）[17]、自适应梯度下降法 AdaGrad[18]、Adam[19]等。

随着神经网络的不断发展，网络层数也不断加深，逐渐衍生出深度学习。相比于传统的特征提取方法，深度学习能够学习到数据中高级的语义特征，从而具有强大的特征表达能力，大大提升了模型的鲁棒性，从而在各领域都逐渐形成主流。

2.2 卷积神经网络

在计算机视觉领域，通常使用卷积神经网络（CNN）来进行特征提取。卷积神经网络最早来源于 LeNet-5[20]，其主要由输入层、卷积层、池化层、全连接层、输出层组成，网络架构如图 2.2 所示。

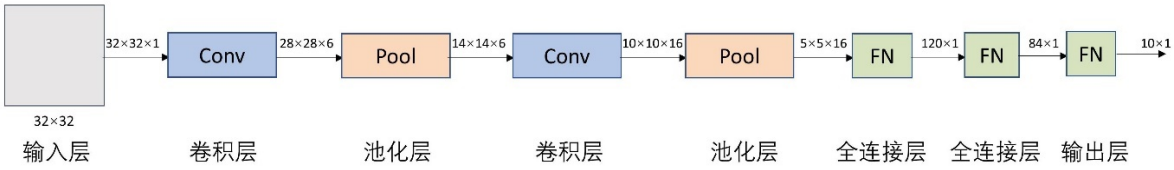

图2.2 LeNet-5 网络架构图[20]

2.2.1 卷积层

每一个卷积层都包含多个卷积核，各个卷积核通过卷积运算来提取图像特征。图 2.3 表示了一个卷积计算的计算过程。

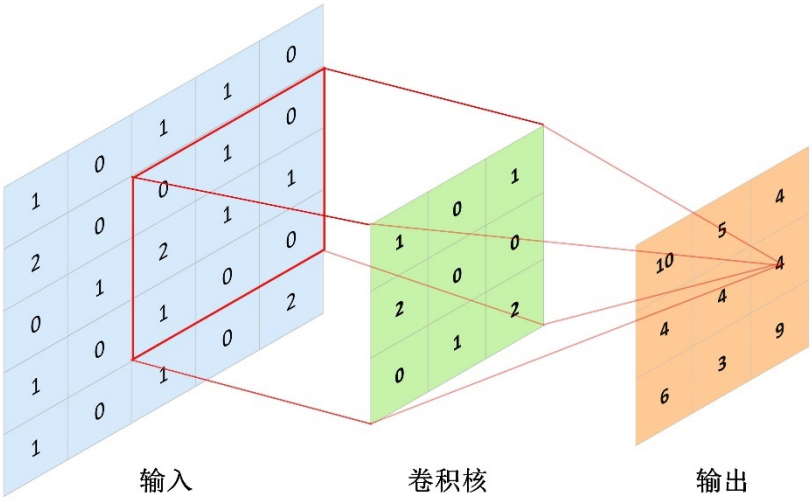

图2.3 卷积计算示意图[20]

卷积核和输入图像相同大小的区域进行对应点相乘求和，计算结果映射到输

出特征图上的一个点。之后，卷积核在输入图上进行一定步长的滑动，就可以得到整幅输出特征图。

在进行卷积操作时，距离输入端较近的卷积层得到的特征图分辨率较大，感受野较小，能够提取到输入图像的浅层特征，比如纹理、颜色、边缘等细节信息。距离输入端较远的卷积层得到的特征图分辨率较小，感受野较大，能够提取到输入图像的深层特征，包含更多高级语义信息。由于每一个卷积核的初始值并不相同，因此可以提取不同的特征信息，作为后续处理的基础。

2.2.2 池化层

池化层主要用来进一步提取特征图上最有代表性的特征，从而降低网络参数量，减小模型的训练开销。

如图 2.4 所示，常见的池化方式包括最大池化和平均池化两种。最大池化是将一定区域内的特征值用区域最大值代替，这样可以有效抑制一些噪声的干扰。平均池化是将一定区域内的特征值用该区域的平均值代替，这样可以有效提升模型的鲁棒性。

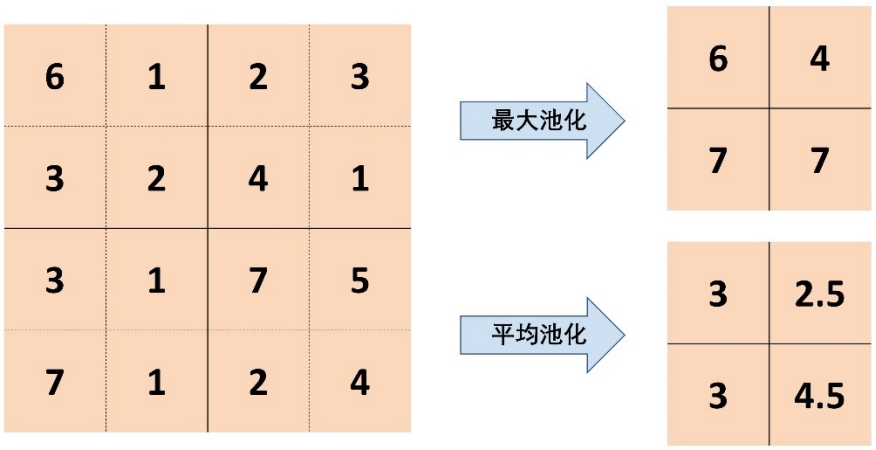

图2.4 两种常见的池化方式[20]

2.2.3 全连接层

全连接层处于卷积神经网络中的末端，主要用于将高维特征图映射成一维向量，整合网络前几层提取的特征，最终连接输出层，输出网络结果。图 2.5 展示了一个全连接层的基本结构。

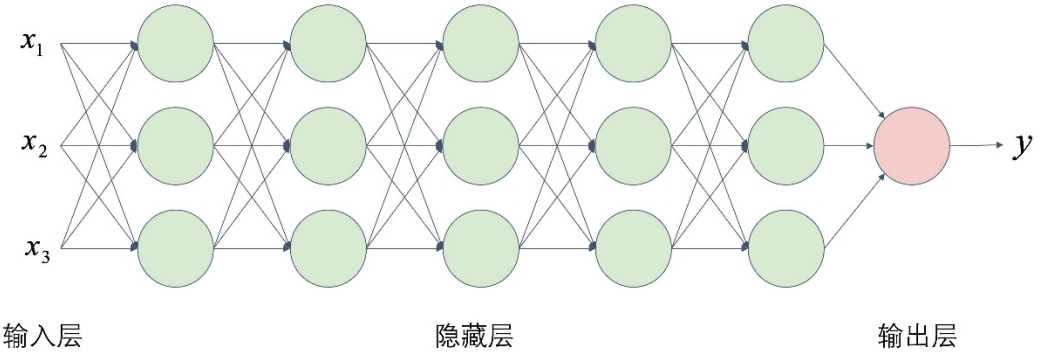

图2.5 全连接层基本结构[20]

2.2.4 激活函数

由于卷积和池化操作都属于线性运算，这导致神经网络无论深度如何，始终是线性模型。为了提高神经网络的非线性表达能力，在每一个神经元的信息传播过程中，加入了激活函数，对神经元的输入信号进行非线性映射。常用的激活函数主要有 Sigmoid 函数、Tanh 函数、Softmax 函数、ReLU [21]函数等。

对于一个二分类问题，通常会在输出层添加一个 Sigmoid 函数，将其映射成（0,1）之间的数值。Sigmoid 函数表达式如公式 2-1 所示。

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2-1}$$

对于多分类问题，通常会在输出层添加一个 Softmax 函数，该函数可以将输入值映射到（0,1）之间，得到各类别的概率分布。Softmax 函数表达式如公式 2-2 所示。

$$f(z_i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}} \tag{2-2}$$

式中，z_i表示第i个节点的输出值，C 表示节点总数。

为了解决网络在训练过程中，梯度消失的问题，通常会使用 ReLU 函数，ReLU 函数的收敛较快，目前已成为应用最广泛的激活函数之一。ReLU 函数表达式如公式 2-3 所示。

$$f(x) = \max(0, x) \tag{2-3}$$

2.2.5 损失函数

损失函数用来计算模型输出值和目标值之间的误差，误差越小说明模型越可能收敛。因此合适的损失函数能够在训练过程中，更好地帮助模型更新参数，达成收敛效果。

损失函数大致可分为回归损失和分类损失。回归损失主要针对连续型变量，通常使用均方误差来进行计算。均方误差指的是模型预测值与目标值之差的平方和，计算方式如公式 2-4 所示。

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \tag{2-4}$$

分类损失主要针对离散型变量，通常使用交叉熵损失来进行计算。二分类交叉熵（BCE）损失函数如公式 2-5 所示。

$$Loss_{BCE} = -[ylog(p) + (1-y)\log(1-p)] \tag{2-5}$$

式中，y 表示类别，使用 0 或 1 表示两个分类类别；p 表示模型预测为正样本的概率，通常通过 Sigmoid 激活函数得到（0,1）范围内的值。

多分类交叉熵损失函数如公式 2-6 所示。

$$Loss_{CE} = -\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{2-6}$$

式中，i表示样本数；M表示类别数；y_{ic}表示样本i的真实类别是否为c，若是则为 1，不是则为 0；p_{ic}表示样本i预测为类别c的概率，该值通常使用激活函数 Softmax 得到。

2.3 目标检测算法

在深度学习进入目标检测领域之后，基于深度学习的目标检测算法层出不穷。从算法的处理方式来分，主要可以分为单阶段目标检测算法和双阶段目标检测算法。单阶段目标检测算法实现了从图像到检测结果端到端的方式，即输入图像之后，可以一次性得到图像中目标的检测结果。双阶段目标检测算法主要分成两个阶段，第一阶段是通过算法来确定目标的预选范围，第二阶段再将预选范围的目标进行分类。总体而言，单阶段目标检测算法检测速度较快，但准确度往往不如双阶段目标检测算法。

2.3.1 单阶段目标检测算法

单阶段目标检测算法中的经典算法主要代表是 YOLO[22]和 SSD[23]。

YOLO 算法的核心思想是先将不同尺寸的图片缩放到同一尺寸，然后对每一张图片划分成等大的网格，每一个网格负责预测目标的位置和其所属的类别。

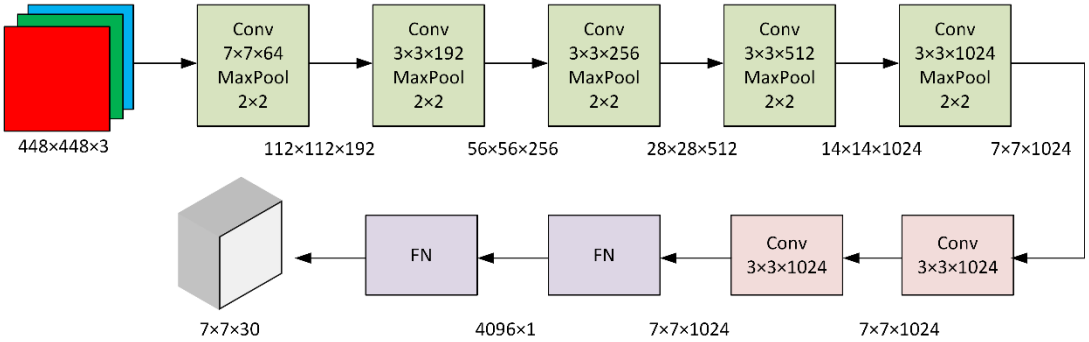

图2.6 YOLOv1 网络结构图[22]

如图 2.6 所示，448×448×3 的图像经过一系列卷积和池化操作后，最终会输出一个 7×7×30 的特征向量。这里的 7×7 表示原始图片被划分为 7×7 个网格，每一个网格输出长度为 30 的内容，包含中心点坐标、目标长宽和目标类别。然而，当同一个网格区域存在多个目标时，YOLO 只能检测出一个。

针对 YOLO 存在的漏检问题，SSD 在不同特征层上用不同尺寸的框来检测目标，有效提升了检测的精度，同时保持了较快的检测速度。

2.3.2 双阶段目标检测算法

双阶段目标检测算法最早起源于 RCNN[24]。RCNN 的核心思路是先通过启发式区域提取算法 Selective Search 对图像进行划分，得到 1000-2000 个候选区域，然后将这些候选区域送入卷积神经网络中提取特征，最后使用非极大抑制算法（NMS）去除重复的预测框，并通过支持向量机（SVM）对预测框中的目标进行分类。

RCNN 相比于传统的目标检测算法性能得到了较大的提升，然而，由于 RCNN 需要提取大量候选区域，且对每一块候选区域都需要单独进行回归和分类计算，会产生较大的计算量，这使 RCNN 并未得到广泛应用。针对 RCNN 速度慢的缺点，Fast RCNN[25]对候选区域进行池化操作来取代 RCNN 中的归一化操作，大大减少了模型的参数量，并且，其将模型最后阶段的回归和分类进行统一，进一步减少了

算法的计算量，使检测速度更快。然而，Fast RCNN 在获取候选区域时，仍花费了较长时间。因此 Faster RCNN[26]对其进行进一步优化。

Faster RCNN 使用区域提出网络（RPN）来代替 Selective Search 算法来提取候选区域，这使所有后续模块无需单独对每一块候选区域进行单独操作，而只需共用一幅特征图，大大减小了计算量，使模型的计算速度得到进一步提升。

2.4 图像分割算法

基于深度学习的图像分割最早起源于全卷积神经网络[27]（FCN）。在一般的用于分类的卷积神经网络中，模型最后会连接三个全连接层，这导致模型输入必须为固定大小。而在 FCN 中，移除了全连接结构，整个网络仅使用卷积层和池化层，这可以使模型能够处理不同输入大小的图像。图 2.7 展示了 FCN 的具体结构。

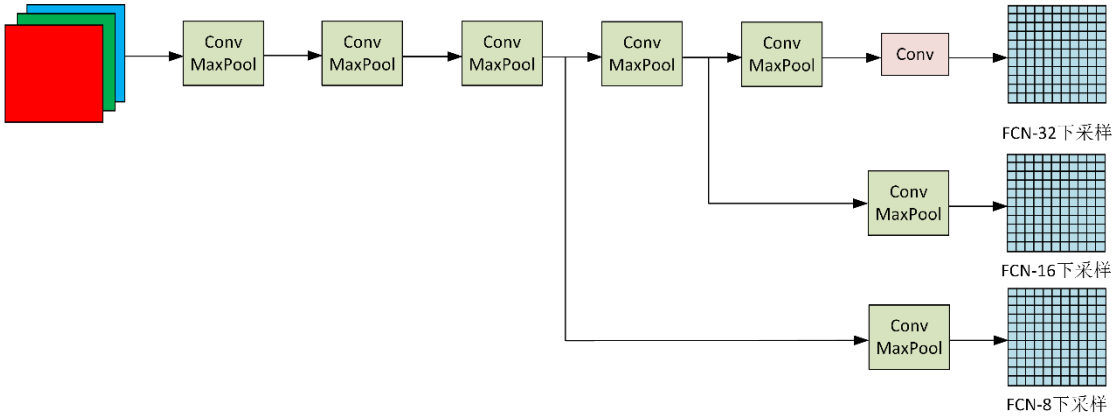

图2.7 FCN 模型结构[27]

除了全卷积的特点之外，FCN 还引入了反卷积和跳跃连接两种处理方式。反卷积是指在图片经过数层卷积和池化操作后，特征图大小远小于原始图大小，此时通过反卷积将特征图大小恢复成原始图大小，实现将特征图中每个像素和原始图中一一对应。跳跃连接是指 FCN 将模型浅层的细节信息和高层的语义信息结合起来，从而使模型的性能和鲁棒性更强。

续 FCN 之后，针对医学图像分割任务，新的图像分割模型 U-Net[28]被提出。U-Net 使用了编码器-解码器结构，并同样使用了跳跃连接。

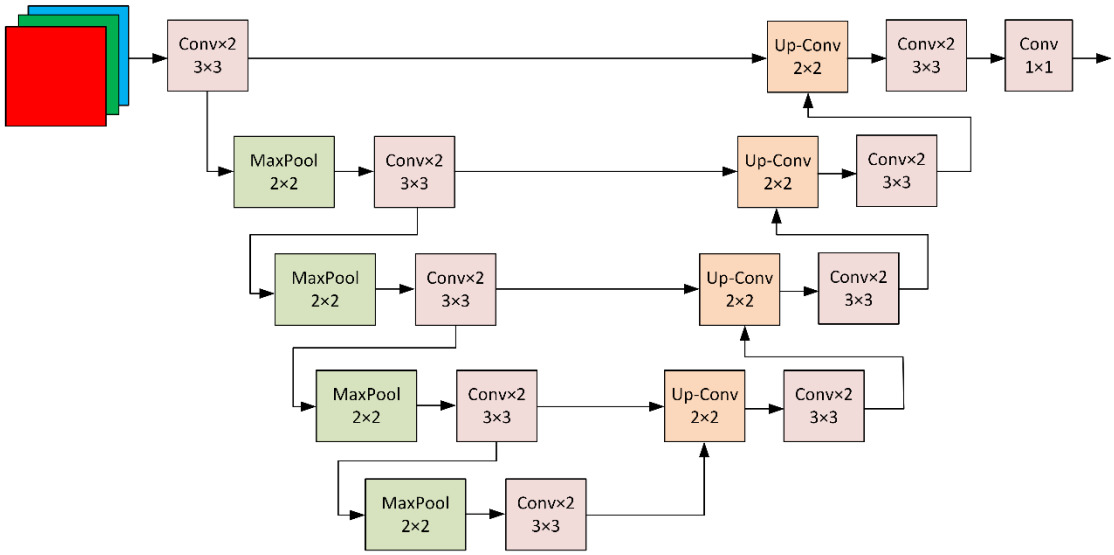

图2.8 U-Net 模型结构[28]

如图 2.8 所示，U-Net 对于每一张图片首先在编码器部分进行 4 次下采样，每次下采样之后，通道数变为原来的 2 倍。之后送入解码器进行上采样，每次上采样都与相同次数的下采样特征图进行特征融合，最后使用 1×1 的卷积核获得需要的分类数。

尽管 U-Net 的分割性能得到了一定提升，但它的参数量较大，训练较为缓慢，且容易发生过拟合。针对 U-Net 的这些缺陷，LinkNet[29]被提出来，主要贡献是在获得较高分割准确率的基础上，尽可能减少模型的推理时间。

LinkNet 仍采用了编码器-解码器的模型架构设计，整体结构和 U-Net 相类似。为了减少模型的参数量，LinkNet 在编码器部分使用 ResNet18 作为主干网络，解码器部分使用通道约减来压缩体积，这使整个网络在保证一定分割准确率的同时推理速度大大加快。

2.5 本章小结

本章节从深度学习的基础出发，首先介绍了深度学习的起源和前向传播、反向传播等基础概念。然后简要介绍了卷积神经网络中卷积层、池化层、全连接层、激活函数、损失函数等模块的原理。最后介绍了几个目标检测和图像分割中的经典算法，为后续章节提供理论支持。

第三章 知识嵌入与人机交互的小样本目标检测

3.1 引言

小样本数据量较少的问题给检测识别带来严峻的挑战。常规的目标检测往往只关注于目标本身的信息，根据目标自身的大小、轮廓、颜色等信息作为特征来进行识别。而在现实场景中，光照剧烈变化、目标遮挡、目标尺度变化等一系列实际存在的问题，都会对模型的检测构成强烈的干扰。

本文对此的改进思路是引入更多环境信息，作为先验知识，嵌入检测模型中，增加模型学习的信息量，更好地利用有限的样本信息，同时利用环境信息对模型的检测结果进行约束和引导。在全新的场景中，模型检测的结果很难做到百分之百正确，因此，本文又引入了人机交互的目标检测机制，通过“人在回路”的设计理念，让人对模型输出的结果进行检查和修正，同时，人工修正后的结果会再次输入到模型中进行二次训练。本章节将详细阐述知识嵌入和人机交互目标检测的具体设计。

3.2 D-LinkNet 算法原理

本文主要针对的是在野外场景中，无人机视角下汽车目标的检测和识别。野外场景中包含农田、建筑、公路等区域，而所需要检测识别的汽车目标大概率只会出现在道路上。因此，本文尝试在检测模型中引入道路信息，以帮助模型更好地将检测范围引导在道路上。因此，首先需要通过图像分割，将道路信息提取出来。

D-LinkNet[30]是 CVPR2018 遥感图像道路分割的冠军算法。它在 LinkNet 的基础上，引入空洞卷积模块，它可以在保证特征图卷积之后分辨率不变的同时，扩大感受野，使其能够捕获到更多信息。同时，D-LinkNet 采用了测试时增强策略（test time augmentation, TTA），即在图像测试时，对每一张图像进行水平翻转、垂直翻转、对角翻转等数据增强方式，最后合并每一种增强之后的检测结果。D-LinkNet 的模型结构如图 3.1 所示。

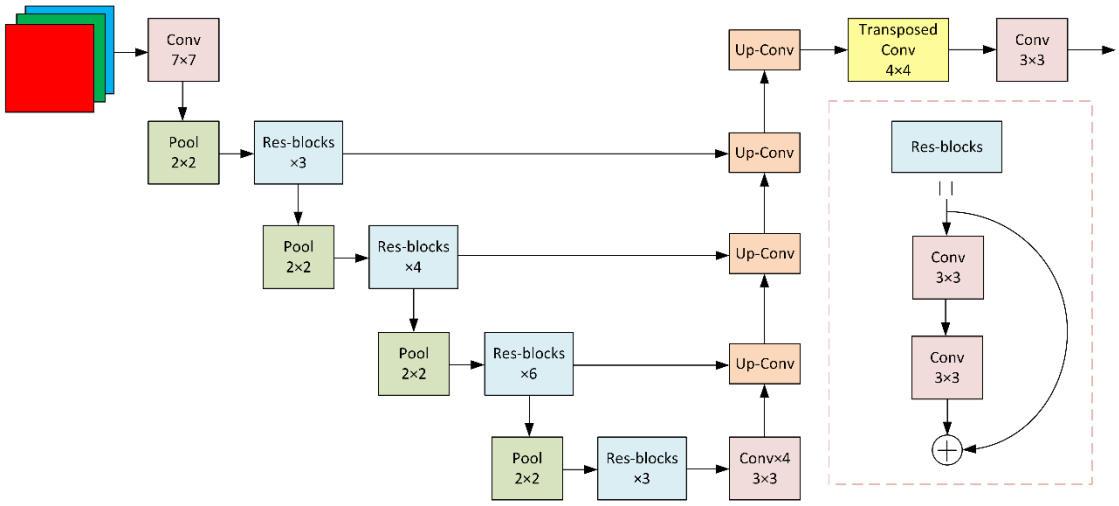

图3.1 D-LinkNet 模型结构图[30]

D-LinkNet 提出者采用了 DeepGlobe 数据集[31]对其进行训练，该数据集包括 6226 张大小为 1024×1024 的图片，图片分辨率为 0.5 米/像素。D-LinkNet 的分割效果如图 3.2 所示。

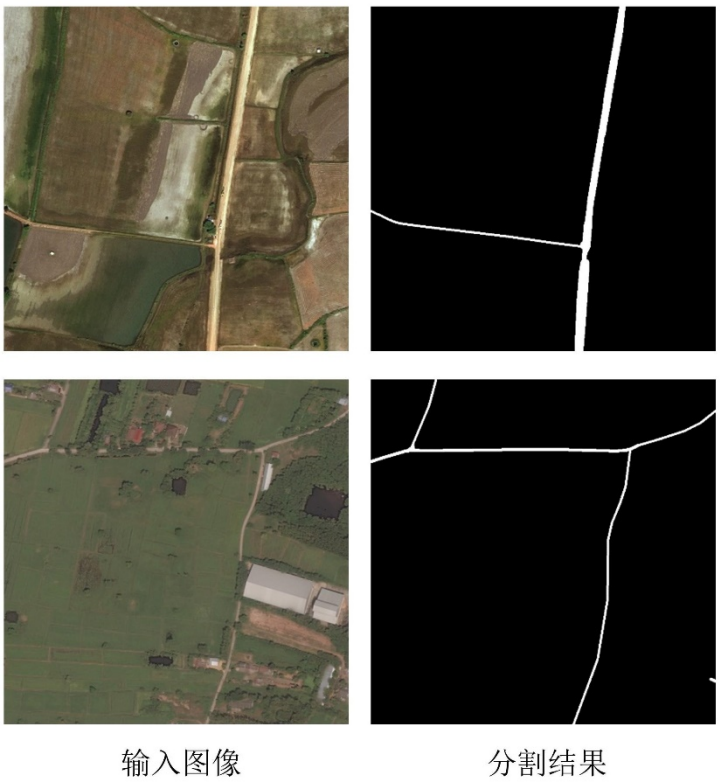

图3.2 D-LinkNet 对 DeepGlobe 部分数据的道路分割结果

3.3 YOLOv5 算法原理

作为单阶段目标检测算法，YOLOv5 主要特点是速度快。为了兼顾模型检测的速度和精度，YOLOv5 通过调整网络的深度和宽度，依据模型的参数量，从小到大依次包括 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 五种模型。相比于 YOLO 系列的早期版本，YOLOv5 增加了许多创新，有效解决了 YOLOv1 版本中，单个区域只能检测一类目标的缺陷。图 3.3 展示了 YOLOv5 模型的整体结构，根据结构不同的所处位置，YOLOv5 可将模型分为输入层（Input）、骨干结构（Backbone）、颈部结构（Neck）、检测头（Detect Head）这几个部分。

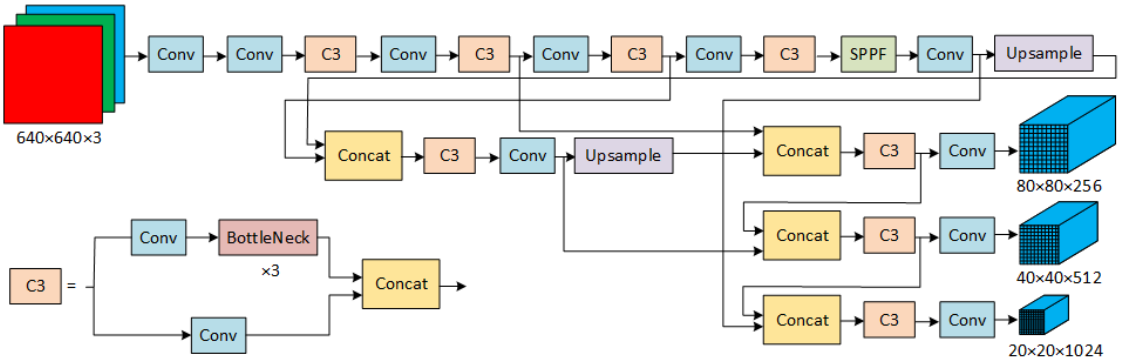

图3.3 YOLOv5 模型结构图[32]

对于输入层部分，YOLOv5 主要做了以下三点改进：

（1）Mosaic 数据增强。YOLOv5 除了使用图像平移、图像旋转、颜色变换等常规数据增强手段外，还沿用了 YOLOv4[32]中的 Mosaic 数据增强方式。Mosaic 是指将经过随机缩放、裁剪的四张图像合并成一幅图像，输入到模型中进行训练。该方法一定程度地将所有的目标都进行缩小，因此，对于小目标检测的效果提升比较明显。

（2）自适应锚框计算。YOLOv2[33]引入了锚框机制，使网络输出位置时，只需要输出对应锚框的偏移量，提升了目标位置检测的精准度。然而，锚框的尺寸大小依赖于人工经验设置，不具备良好的通用性。因此，YOLOv5 使用 K-means 聚类的方式，根据目标本身的大小，自适应计算锚框的大小，提升了模型的泛化性能。

（3）自适应图片缩放。由于 YOLOv5 在进行特征提取过程中，需要进行五次下采样，每次下采样，图像的宽高都会缩小为原始的二分之一。因此，输入模型的图像尺寸宽高必须为 32 的倍数。对于不满足此条件的图像，YOLOv5 会进行自适应图片缩放，为图片填充少量像素，使其满足输入条件。

YOLOv5 的骨干结构主要用来提取特征，在此过程中，YOLOv5 加入了一些创新性模块，比如图 3.3 中的 C3 模块，该模块将一个卷积层提取的特征和另一个附带多个残差结构提取的特征进行拼接，这样能在减少计算量的同时学习到更多信息，同时降低了内存消耗。

YOLOv5 的颈部结构借鉴了 FPN[11]和 PAN[12]结构的思想，将网络的浅层特征下采样后和深层特征进行融合，既保留了浅层网络提取到的细节信息，又加入了深层网络提取到的语义信息，有效提升了算法性能。

YOLOv5 设计了三个检测头结构，每个检测头的输出表示原始图像被划分的网格尺寸，根据网格尺寸的大小，分别用来检测大目标、中目标和小目标。在检测头输出结果后，YOLOv5 会进行非极大抑制处理（NMS），用来过滤掉多余的检测框。NMS 的计算思路是对所有检测框的置信度进行排序，选择置信度最高的检测框，其它检测框和它求交并比（IOU），若大于设定的阈值，则将该检测框删除。之后，在未删除的检测框中，继续进行此操作，循环往复，基本可以剔除所有重复的检测框，使每一个目标仅保留一个检测结果。

训练过程中，在检测结果输出之后，YOLOv5 输出的检测框需要和真实目标框进行比较，通过损失函数来计算损失。YOLOv5 的损失函数由目标损失、分类损失、定位损失三部分构成，损失的计算方法如公式 3-1 所示。

$$Loss = \lambda_1 L_{obj} + \lambda_2 L_{cls} + \lambda_3 L_{loc} \tag{3-1}$$

式中，λ_1、λ_2和λ_3为调节各损失重要性的权重，通过人工设置。

其中，目标损失和分类损失采用的是交叉熵损失函数，定位损失采用的是 CIOU，该损失函数充分考虑了预测框和真实框的重叠面积、中心点距离和长宽比。

3.4 融合道路信息的检测网络

本文将 D-LinkNet 和 YOLOv5 相结合，提出了一种融合道路信息的检测网络。网络具体分为两个步骤，首先通过 D-LinkNet 分割出图像中的道路信息，获得道路伪标签，然后将其嵌入到 YOLOv5 检测网络中。

3.4.1 道路信息提取

D-LinkNet 预先使用了 DeepGlobe 数据集进行训练，该数据集为卫星遥感图像，

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。

3.4.2 道路信息嵌入

获得图像中的道路信息之后，下面就需要将该信息嵌入到检测网络之中，具体的网络架构如图 3.4 所示。

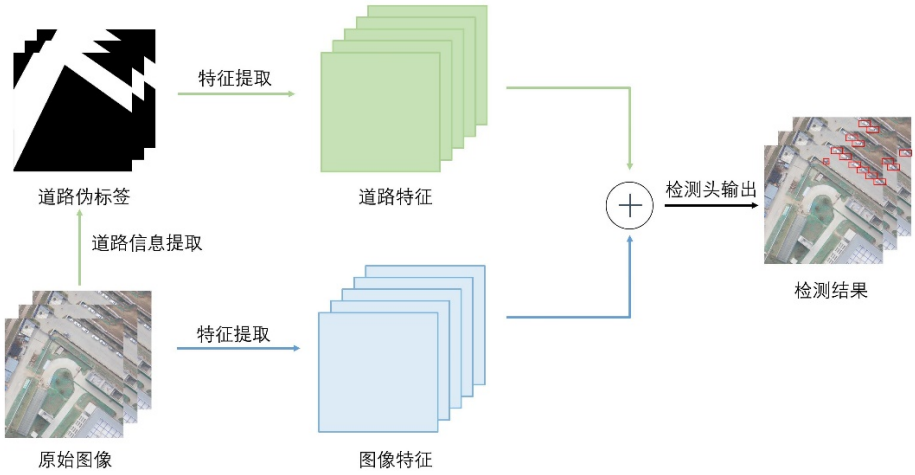

图3.4 融合道路信息的检测网络架构图

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。

如图 3.5 所示，道路信息的特征提取部分沿用了 YOLOv5 的部分结构。

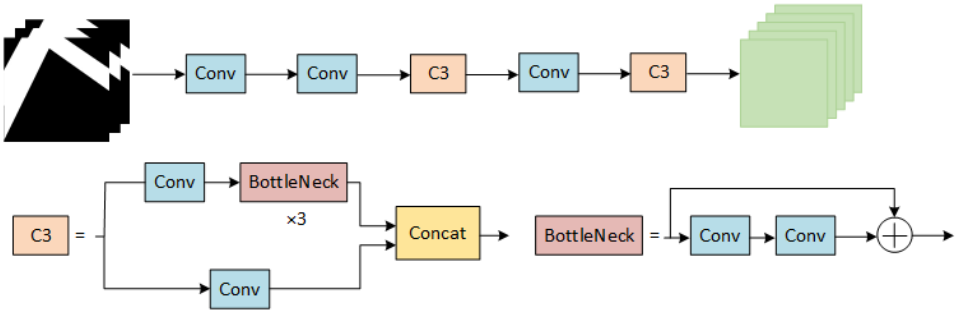

图3.5 道路信息特征提取结构图

3.5 人机交互机制设计

深度学习输出的往往都是概率值，目标检测会给予每一个目标框一个置信度，置信度范围为（0,1）。置信度越接近 1，说明模型认为其预测得越准确。然而，即使是准确率很高的模型，其检测结果也很难做到百分之百确定。特别是对于检测难度较高的小目标，模型往往会存在漏检或者误检的情况。

因此，本文引入了一套人机交互机制，通过人工干预的方式，对模型的输出结果进行检查和纠正。同时，纠正完成的信息会形成数据标签，可以再次送入模型中，对模型的学习情况进行指导。

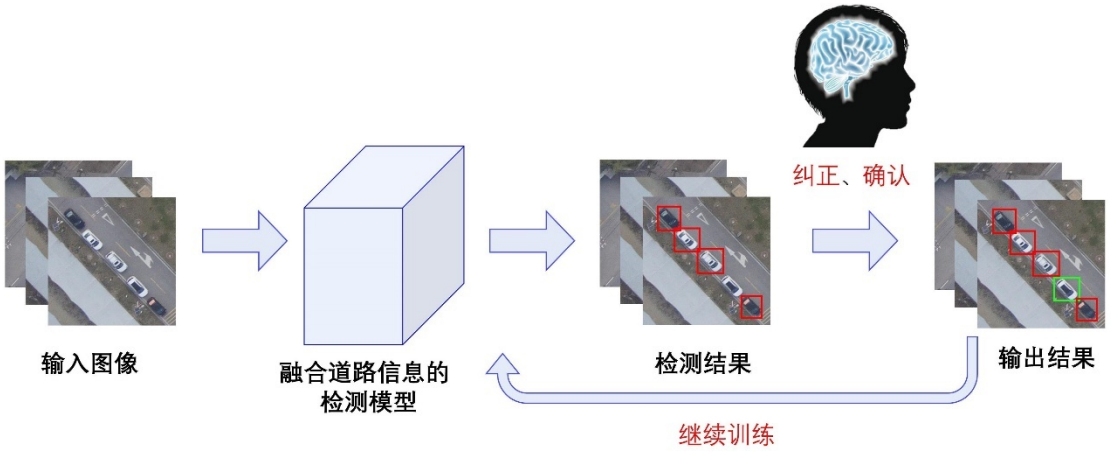

图3.6 人机交互机制设计图

图 3.6 展示了人机交互机制的流程，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。

3.6 本章小结

本章节首先介绍了 D-LinkNet 图像分割的原理和 YOLOv5 目标检测的原理和网络细节。之后从道路信息提取和道路信息嵌入两个角度详细介绍了本文提出的融合道路信息的检测网络设计思路和模型结构。最后阐述了人机交互的目标检测机制设计流程，通过该方式提升了目标检测结果的可靠性和可迁移性。

第四章 实验结果与分析

4.1 实验环境与超参数设置

本文的实验平台为 Windows10，使用 Pycharm 作为编辑器，具体的软件和硬件系统配置如表 4.1 所示。

表4.1 实验环境

环境配置	名称	信息
硬件配置	CPU	英特尔酷睿 i9-13900K 主频 3.0GHz
	GPU	七彩虹 RTX4090
	内存	宏基炫光 D5 6000 64GB
	显存	24GB
	硬盘	西部数据 Sn850X 固态硬盘 2TB
软件环境	操作系统	Windows 10 专业版 22H2
	Pycharm	Community Edition 2023.1
	Python	Python 3.8.15
	Cuda	11.7
	cuDNN	8.1.0
	Pytorch	1.13.0

本文实验所设置的超参数如表 4.2 所示。

表4.2 超参数设置

名称	数值
训练图片分辨率(image_size)	1280×1280×3
迭代运行次数(epochs)	100
批大小(batch size)	16
优化器(optimizer)	SGD
初始学习率(lr0)	0.01
周期学习率(lrf)	0.01
学习率动量(momentum)	0.937

4.2 数据集与数据预处理

4.2.1 数据集简介

本文所使用的数据集是课题组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。所需要检测的是汽车单类目标，目标平均尺寸约为 60×30 像素。

本文在其中筛选了 209 张带有包含汽车目标的图像作为数据集，部分数据如图 4.1 所示。

图4.1 汽车目标数据集预览

4.2.2 数据预处理

由于数据本身的尺寸较大，直接送入模型中进行检测会带来较大的计算开销，且无法直接利用 YOLOv5 原作者在 ImageNet[34]大型数据集上预训练得到的模型参数进行迁移学习。因此，本文将数据集图像切分成 1280×1280 的图像块。图 4.2 展示了一个图像切分的示例，红框表示 1280×1280 的图像块。为了防止在切分过程中，目标被切分开导致检测失败，每两个相邻图像块之间保留了 10%的重叠度。

图4.2 图像切分示例

其次，为了得到每一个切分后图像块的道路信息，本文先将每一张原始图像进

行下采样，缩放到 1024×1024，送入 D-LinkNet 提取道路信息，然后上采样得到原始图像的分割结果，整个流程如图 4.3 所示。

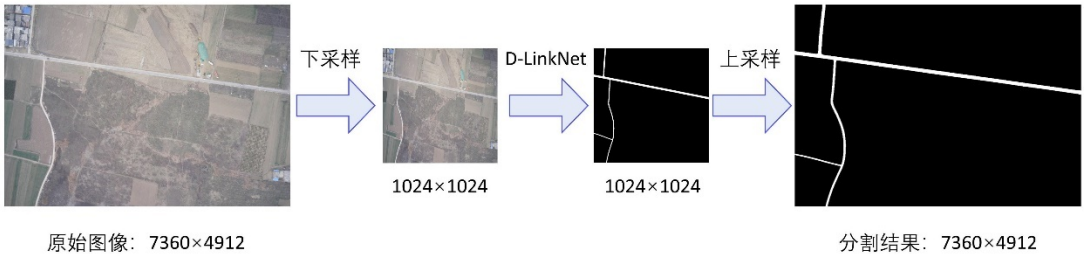

图4.3 道路分割流程图

得到每一张原始图像的道路分割结果后，再根据上一步的切分方式，即可得到每一个图像块对应的道路信息。

4.3 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 P：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 4-1 所示。

$$P = \frac{T_P}{T_P + F_P} \tag{4-1}$$

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 R：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 4-2 所示。

$$R = \frac{T_P}{T_P + F_N} \tag{4-2}$$

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 mAP：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 4-3 所示。

$$mAP = \frac{1}{N}\sum_{k=1}^{N} P(K)\Delta R(k) \tag{4-3}$$

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。

4.4 实验性能评估

在经过数据预处理之后，数据集图像共包含 5016 张 1280×1280 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。

4.4.1 收敛性分析

如图 4.4 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。

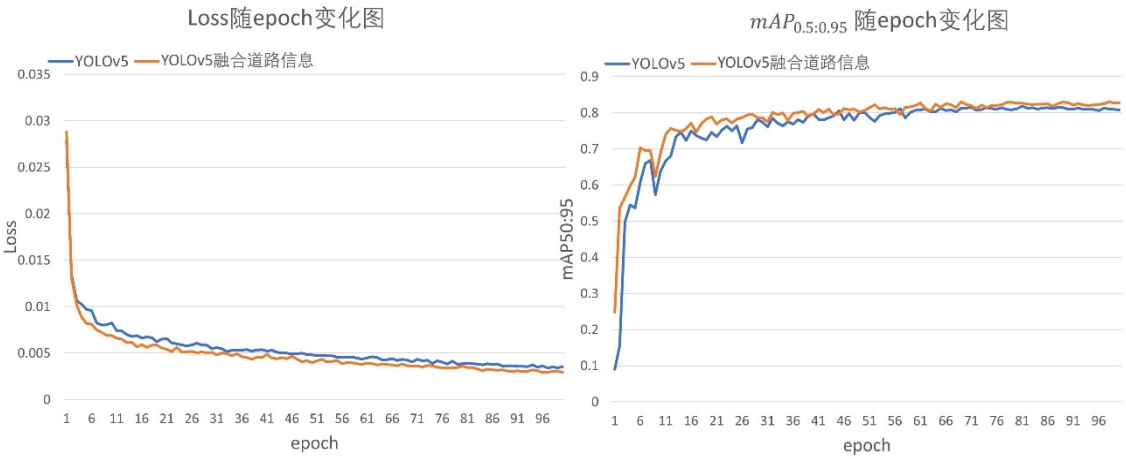

图4.4 Loss 和$mAP_{0.5:0.95}$随 epoch 变化图

4.4.2 模型性能评估

改进前后的两种模型相关性能指标如表 4.3 所示。

表4.3 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 4.5 所示。

图4.5 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，道路信息的融合确实对模型性能有所提升。

4.5 本章小结

本章节对嵌入道路信息的检测网络和原始检测网络进行了对比实验，详细阐述了数据来源、数据预处理方式、实验性能的评估指标。之后从模型收敛性和模型性能两个角度对道路信息嵌入模型的有效性进行评估。实验表明，嵌入道路信息之

后，虽然模型的推理速度会有所下降，但模型的性能会得到提升。最后用一个场景的检测结果对比来说明了道路信息的有效性。

第五章 基于知识嵌入的人机交互检测系统

5.1 系统概述

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。

5.2 功能模块设计

系统的基础界面如图 5.1 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5.1 人机交互目标检测系统基本界面

系统主要包含以下几个功能模块。

（1）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。

（2）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。

（3）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 的图像块。

（4）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。

（5）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。

5.3 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 5.2 所示。

图5.2 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 5.2 显示的是目标检测图层，图 5.3 显示的是道路分割图层。

图5.3 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 5.4 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。

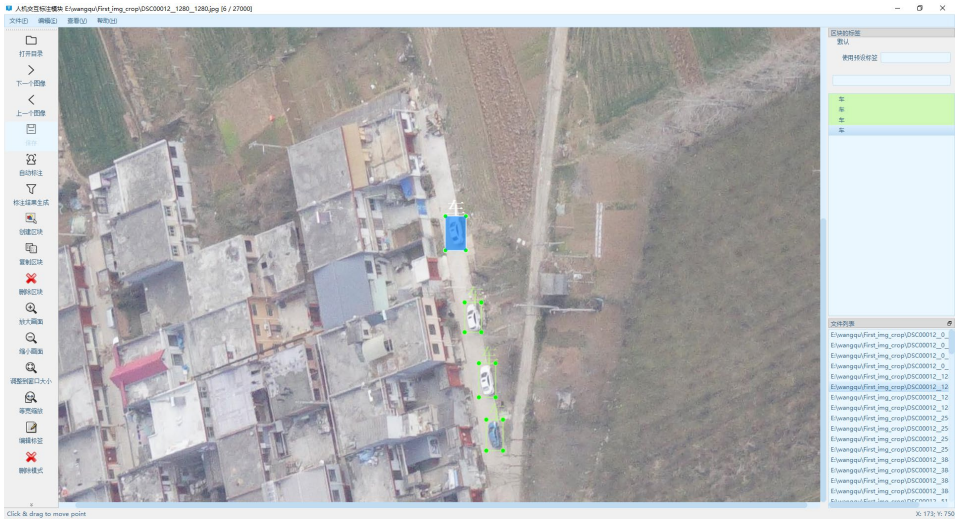

图5.4 目标检测图层展示

（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5.5 模型训练模块界面

5.4 本章小结

本章节详细阐述了一套基于知识嵌入的人机交互检测系统的设计思路和操作流程。通过本系统，用户可以在不接触源代码的情况下，可视化地完成数据加载、数据预处理、检测结果查看、人机交互标注、模型训练等功能。本系统大大降低了目标检测的使用门槛，并提升了检测结果的可靠性，具有广泛的应用价值。

第六章 总结与展望

6.1 本文总结

深度学习在目标检测识别领域取得了显著的成果，但是相关模型在数据样本量不足的情况下，无法训练出有效的特征提取器，效果较差。为了提高小样本检测结果的准确性和可靠性，本文开展了基于知识嵌入和人机交互的小样本识别算法研究，具体内容包括以下三个方面：

（1）为了增加目标检测模型所获取的信息量，本文基于数据集的特点，针对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，本文使用 D-LinkNet 对无人机遥感图像进行道路分割。本文先使用卫星遥感影像数据集 DeepGlobe 对 D-LinkNet 进行预训练，然后对无人机影像进行下采样，使其接近训练数据的分辨率，再输入网络中实现道路分割，得到原始图像的道路信息。

（2）为了提升目标检测模型对小样本的检测识别能力，本文提出了一个融合道路信息的检测网络。具体方式是先对 D-LinkNet 提取到的图像道路信息进行特征提取，然后将道路特征和 YOLOv5 本身提取到的图像特征进行融合，最后输出检测结果。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。

（3）为了提升目标检测模型对小样本的检测结果的可靠性，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。

6.2 工作展望

在更多实际的小样本目标识别应用中，算法所面临的场景更为复杂，且场景中所蕴涵的信息更为丰富。本文仅仅关注场景中所包含的道路信息，所做的工作仍有许多不足之处，个人认为还可以从以下几个角度进行进一步研究：

（1）引入更多的先验知识。实际场景中，不仅仅只包含道路知识，环境的亮暗、色调、灰度均可作为先验知识辅助检测。同时，多于多目标的识别任务，目标

与目标之间的先验关系也可以作为一种知识来帮助模型进行检测。

（2）样本不均衡问题。小样本目标由于通常样本量较少，通常存在比较明显的样本不均衡问题。本文并没有对此展开研究。后续如果采用一定方法缓解小样本不均衡的问题，可能会对检测效果有所帮助。

（3）检测速度问题。现实场景下，除了对目标检测的精度有一定要求外，对于检测的速度也往往存在要求。如何在不影响速度的情况下，提升模型的精度，也是未来的一项重要课题。

致谢

时光荏苒，本科生阶段的学习走入尾声。回首过去四年的学习时光，十分感谢陪我一起经历的老师、同学、朋友和家人。正是他们的支持和帮助，我才能顺利地走到今天。

首先，我要衷心感谢我的指导老师吴金建教授。吴老师思维敏锐、治学严谨，成果丰硕。我正是在吴老师的谆谆教导下，提升了自己的学习能力、工作态度和思维方式。吴老师严格律己的工作态度深深打动了我，促使我不断在知识学习、科研探索中取得进步。在我完成此次毕设期间，吴老师为我提供了良好的科研环境，并对我的毕设方向进行指导。在此，我向吴老师表示衷心的感谢。

其次，我要感谢我的母校西安电子科技大学。在母校的庇荫下，我深深沉醉在母校优秀的科研氛围中。母校不仅提供了舒适的住宿环境，还给予了每个学生充沛的时间和充足的图书资源，以便我能够在课余时间中不断提升自己。在此，我深深表达对母校的感谢，祝母校越办越好。

此外，我还要感谢大学四年期间传授我知识的老师们和陪伴我学习的同学们。每当我遇到困难和挫折时，总能获得及时的帮助。这得以让我在日积月累中，不断构建自己的知识体系，夯实自己的专业基础，顺利完成所有的毕业要求。

最后，我要感谢我的家人。感谢父母含辛茹苦的养育和栽培，感谢长辈们一直以来的理解和支持，使我能够全身心投入到学业中，我会带着你们的期望继续向更高的殿堂进发。

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] Shen D G, Wu G R, Suk H I. Deep learning in medical image analysis[J]. Annual Review of Biomedical Engineering, 2017, 19: 221-248.

[3] 黄元涛. 基于深度学习的藏羚羊检测与跟踪[D].西安：西安电子科技大学，2020：3-69.

[4] 宋一言,唐东林,吴续龙,周立,秦北轩.改进穿线法与 HOG+SVM 结合的数码管图像读数研究[J].计算机科学,2021,48(S2):396-399+440.

[5] 黄海波,李晓玲,聂祥飞,张月,冯丽源.基于 SIFT 算法的遥感图像配准研究[J]. 激光杂志,2021,42(06):97-102.DOI:10.14016/j.cnki.jgzz.2021.06.097.

[6] 范博文,段敏.采用模糊支持向量机算法的前车识别系统[J].重庆理工大学学报 (自然科学),2022,36(09):172-178.

[7] 何松华,章阳.基于快速检测和 AdaBoost 的车辆检测[J].计算机工程与设计,2020,41(01):203-207.

[8] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C] //Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE 2005, 1: 886-893.

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

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

[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. 2017: 2117-2125.

[12] 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.

[13] 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

[14] 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.

[15] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[16] Bottou L. Stochastic gradient descent tricks[M]. Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012: 421-436.

[17] Qian N. On the momentum term in gradient descent learning algorithms[J]. Neural networks, 1999, 12(1): 145-151.

[18] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of machine learning research, 2011, 12(7)

[19] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[20] Lecun Y , Bottou L . Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[21] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]. Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2011: 315-323.

[22] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, 2016:

[23] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision, Springer, 2016: 21-37.

[24] 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, 2014: 580-587.

[25] Girshick R. Fast r-cnn[C]. Proceedings of The IEEE International Conference on Computer Vision, 2015: 1440-1448.

[26] Ren S, He K, 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, 2016, 39(6): 1137-1149.

[27] Long J,Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2015 [C].

[28] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation; proceedings ofthe International Conference on Medical image computing and computer-assisted intervention, F, 2015 [C]. Springer.

[29] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation[C]//2017 IEEE visual communications and image processing (VCIP). IEEE, 2017: 1-4.

[30] 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.

[31] 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.

[32] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[33] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[34] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009: 248-255