中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于 FPN 融合因子的弱小目标检测研究

作者姓名：宫宇琦

指导教师：韩振军 副教授

中国科学院大学

学位类别：工程硕士

学科专业：电子与通信工程

培养单位：中国科学院大学电子电气与通信工程学院

2021 年 6 月

Research on Tiny Object Detection Based on FPN Fusion Factor

A thesis submitted to

University of Chinese Academy of Sciences

in partial fulfillment of the requirement

for the degree of

Master of Engineering

in Electronic and Communication Engineering

By

Gong Yuqi

Supervisor Han Zhenjun

School of Electronic, Electrical and Communication Engineering

University of Chinese Academy of Sciences

June 2021

中国科学院大学

研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：宫宇琦

日　　期：2021 年 6 月

中国科学院大学

学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：宫宇琦　　导师签名：韩振宇

日　　期：2021 年 6 月　　日　　期：2021 年 6 月

摘　要

弱小目标检测（Tiny Object Detection，TOD）有广泛的应用前景和科研价值。弱小目标检测不仅对通用目标检测在弱小尺度目标方面研究的补充，而且已经应用到多个领域，比如安全监控，快速海上救援，自动驾驶等。对于弱小目标检测，通常沿用通用目标检测的框架，然而这不是最优解。基于特征金字塔（Feature Pyramid Network，FPN）的检测器在通用目标检测上取得了很好的效果，如在MS COCO和PASCAL VOC等数据集上已经取得了优异的性能。然而，这些检测器在弱小目标检测场景下，检测精度并不高，原因在于通用目标检测框架没有对小目标进行适应性的改进。

本研究中发现FPN中相邻层之间自顶向下的连接是影响弱小目标检测性能的关键。为了验证此发现，分别在五个不同数据集上进行了实验分析，实验结果表明，FPN特征层之间的链接不仅决定了不同特征层融合的比例，更决定了深层任务参与浅层进行小目标相关检测任务的程度。本文主要工作如下：

1、提出了一个新的概念——融合因子，用来描述深层传递给浅层的信息量，分析融合因子在特征融合时带来的双面影响。进一步本文对如何设置有效的融合因子进行了深入研究和分析，比对了五种不同的融合因子设置方法，从实用性以及对性能提升两个方面进行分析，得到了最佳的基于统计的融合因子设置方法，在不增加网络的参数的情况可以嵌入到网络中，从而使基于统计方法设置的FPN适应弱小目标检测。

2、对融合因子的作用机理进行了探究，在融合因子隐式学习的方面，从FPN相关卷积参数初始化方式，数据集体量分析了影响融合因子隐式学习的因素。并且探究了基于融合因子的方法对多尺度数据集性能的影响，同时分析了融合因子在网络梯度回传中造成的影响并进行了解释。

3、在实验方面，基于统计的融合因子设置方法在不同的数据集，不同的检测器，不同的骨干网络下都取得了性能上的一致性提升，验证了方法的有效性。并且与不同通用检测器性能的对比实验结果表明对FPN的适应性改变对小目标检测是有益的，其中基于尺度匹配和基于统计的融合因子设置方法合用在

Tinyperson上取得了性能上的进一步提升，证明融合因子是和其他方法兼容的，说明方法的可拓展性。

关键词：弱小目标检测，特征金字塔，融合因子

Abstract

Tiny Object Detection(TOD) has a wide range of application prospects and scientific research value. Tiny Object Detection is not only a supplement to the research of general object detection in tiny-scale object, but also has been applied to many fields, such as security monitoring, fast sea rescue, automatic driving, and so on. For Tiny Object Detection, the framework of general object detection is usually used, but this is not the optimal solution. Detectors based on Feature Pyramid Network (FPN) have achieved good results in general object detection. For example, they have achieved excellent performance on datasets such as MS COCO and PASCAL VOC. However, these detectors do not perform well in scenarios of Tiny Object Detection, because the general object detection framework does not make adaptive improvements to tiny targets.

In this study, we found that the top-down connection between adjacent layers in FPN is the key to the detection performance of tiny targets. In order to verify this finding, experiments were carried out on five different datasets. The experimental results show that the cross-layer link in FPN not only determines the ratio of the fusion of different feature layers, but also determines the degree to which deep-level tasks participate in shallow-level tasks related to tiny object. The main work of this paper is as follows:

1.Propose a new concept, fusion factor, which is used to describe the amount of information transmitted from the deep layers to the shallow layers and analyze the double-sided influence of the fusion factor during feature fusion. And the paper also studied how to set effective fusion factor, compared five different fusion factor setting methods, analyzed from two aspects of practicability and performance improvement, and got the best statistical-based fusion factor setting. The method can be embedded into the network without increasing the parameters of the network, so that the FPN set based on the statistical method can be adapted to the detection of tiny targets.

2.Explored the mechanism of the fusion factor. In terms of the implicit learning of the fusion factor, from the FPN-related convolution parameter initialization method and the amount of data, the factors affecting the implicit learning of the fusion factor were analyzed. And explored the influence of the method based on the fusion factor on the performance of the multi-scale dataset, and also analyzed the influence of the fusion factor in the network gradient backpropagation and explained it.

3.In terms of experiments, the statistical-based fusion factor setting method has achieved a consistent improvement in performance under different datasets, different detectors, and different backbone networks, which verifies the effectiveness of the method. And we conducted a comparison experiment with the performance of different general detectors, the experimental results show that the adaptive change to FPN is beneficial for tiny object detection. Among them, the combination of scale matching and statistics-based fusion factor setting methods has achieved further performance improvement on TinyPerson, which proves that the fusion factor is compatible with other methods and shows the scalability of the method.

Key Words: Tiny Object Detection, Feature Pyramid Network, Fusion factor

目 录

图目录

表目录

第 1 章 引言

1.1 研究的背景及意义

计算机视觉（Computational Vision）是运用图像采集设备和计算机来获取人类所需的信息，用计算机“理解”图像的任务。具体过程为由相机或摄像头等图像采集设备获得图片，通过计算机对图像中的目标进行识别以及语义上的理解，输出人类可以理解的信息。计算机视觉是机器学习在图像处理领域的应用，是人工智能领域的一个重要部分。它的研究内容可以概括为：采集图片或视频，对图片或视频进行处理分析，从中获取信息。

计算机视觉这一学科有着漫长的发展历史，1959 年 David Hubel 和 Torsten Wiesel 通过猫的视觉观察实验，首次发现了大脑中初级皮层神经元对于移动边缘刺激敏感，发现了视功能柱结构，为视觉神经研究奠定了基础。并且同年 Russell 研制了第一台可以把图片转化为被二进制机器所理解的灰度值的数字图像扫描仪第一台，这让计算机处理数字图像开始成为可能。

从上世纪六十年代到八十年代，随着电子计算机的发展，计算机视觉技术开始出现雏形。逐渐出现了用计算机进行图像处理，这时期的工作主要分为两个方向：一是三维视觉理解，人看到的东西是二维的，但是人对图像的理解是三维的，所以将二维图像处理成三维信息是机器理解的前提，Lawrence Roberts 描述了三维视觉理解的最初模式，这个方向的相关工作主要聚焦于物体的三维重构；二是对图像的知识理解进行初步的探索，比如给机器建立一个预先的知识库，探索在知识的条件下对物体的识别效果，这一阶段，研究的主要对象如光学字符识别、工件表面、显微图片和航空图片的分析和解释等[1]。在 1982 年，David Marr 描述了视觉的基本框架，标志着计算机视觉已经成为了一门独立学科。

九十年代后，图像的特征工程成为研究领域的重点。1999 年，David Lowe 发表《基于局部尺度不变特征（SIFT 特征）的物体识别》[2]，标志着研究人员开始停止通过创建三维模型重建对象，而转向基于特征的对象识别。2001 年，第一台具有实时性的相机问世标志着计算机视觉技术在日常任务的正式应用。Dalal

和 Triggs 在 2005 年提出了用行人检测的 Histogram of Gradients(HoG)[3]取得巨大的成功，Felzenszwalb，McAllester等在2008年提出DPM（Deformable Part Model）[4]进一步提高在行人检测上的性能。同时随着 CPU 和 GPU 的发展，基于深度深度学习的计算机视觉任务取得了前所未有的发展，在全球权威的计算机视觉竞赛 ILSVR(ImageNet Large Scale Visual Recognition Competition)上[5]，千类物体识别 Top-5 错误率在 2010 年和 2011 年时分别为 28.2%和 25.8%，从 2012 年引入深度神经网络之后，后续 4 年分别为 16.4%、11.7%、6.7%、3.7%，出现了显著突破。在工业界，诸如人脸识别、字体识别、车牌识别等任务也已经逐渐成熟已应用于日常生活。

随着互联网时代的到来，产生信息与消费信息的方式也越来越多种多样，每个人的都是产生信息的基本单位，图片和视频作为内容中最重要的信息载体，各种各样内容不同的图片和视频每天都在源源不断的产生，人工无法处理亿量级的数据只能依赖于机器，但这也对机器的处理能力提出了巨大的挑战，计算机视觉的应用场景也快速扩展，包括自动驾驶中的视觉系统、短视频处理中的内容理解、医疗领域的医学影像处理与智能诊断。

目标检测是计算机视觉领域的基础任务，这个任务在图像分类的任务基础上，增加了图中物体位置回归这一任务，将图片级别的分类任务升级为物体级别分类与回归任务，产生的位置信息对实际应用有重大意义。因此目标检测被作为一个基础任务被广泛研究，且衍生了不同场景下的具体目标检测任务，如通用检测、行人检测、人脸检测、弱小目标检测等等。

弱小目标检测是计算机视觉领域的一个研究分支，具有广泛的落地应用前景，包括视频监控、辅助驾驶和海上快速救援等。这些应用场景中的目标具有弱小目标的特征和复杂的多样性。另外对无人驾驶这一任务，越清晰的摄像头使得机器获得更远距离的图像，从而能更早的做出一些操作和处理，但是越远的目标就会变得越小，弱小目标检测在这种场景中是个亟待解决的问题。此外，对于很多国防军事任务，例如精确防卫或精确打击，或边境安全监测等任务中，弱小目标检测都是场景需求，比如在遥感图像处理中，汽车和船只往往只是很小部分的像素点，同时这类任务也对弱小目标检测的精度也提出了比较高的要求。通用目标检

测一般研究的多是大尺度的物体，很难适用于这些任务，因此弱小目标的研究可以帮助目标检测在尺度上形成更全面的体系，同时对于实际应用场景中更是意义重大，其重要性不亚于对现有通用目标检测的研究。图 1.1 为弱小目标检测典型场景，图片来源于 TinyPerson 数据集[6]。

图 1.1　弱小目标检测场景示例[6]

Figure 1.1　Sample of tiny object detection scene

从研究的角度，弱小目标检测作为目标检测的子任务，是通用目标检测在尺度研究上的一个细化拓展，通用目标检测中往往对这一尺度的物体很少涉及，因为其数据集都是近距离拍摄的图片，图中的目标尺度都比较大，不会涉及弱小目标检测，但弱小目标检测在实际的视觉任务中是一个常见的现象，所以对弱小目标检测的研究是对通用目标检测的一个很好补充。

另外，无人机无论在军用或者民用的领域中应用也越来越广泛[7]，随着无人机技术的成熟与商业化，基于无人机的相关技术也越来越受到重视，基于航拍目标检测完成自动化监测等就是其中之一。航拍通常会配备极度高分辨率的摄像头以获得高质量的图像，但在进行高空拍摄时地面目标，目标分辨率依旧会很低，因为图像是远距离广视角下产生，目标所占像素点数就非常少，导致目标所占像

素的点不够，目标就会存在边界模糊和容易受到噪声干扰的问题，因此航拍目标检测就是天然的弱小目标检测。人作为经常被检测的对象，对于人体的检测是弱小目标检测的一个重要应用分支，其应用场景也很多，比如用于海难或其他灾难中的搜救任务[6]，比如用于监控分析中或者用于日常娱乐中的定位任务，还可以用于海岸情况监控，控制海滩人员密度，保持安全的社交距离。另外，弱小人体目标检测虽然研究的对象是人体，但其中人体姿态、角度等具有很大的多样性，可以为其他的弱小目标检测提供参考。在 VisDrone[8]数据集中详细描述了无人机航拍的应用场景，比如用于太阳能发电厂对太阳能电池板的缺陷检测，机器检测替代人工排查可以大幅度提高工作效率；或者用于植物的早期病害的检测；还可以用来公共安全领域的鲨鱼侦测。对于这些应用场景中的目标也同样具有弱小目标的特征和复杂的多样性，因此很多在弱小人体目标检测中的方法应当也能对这个问题起到很大的参考作用。

对于很多国防军事任务，例如精确防卫或精确打击或者是舰船检测，或边境安全监测等任务中，图片卫星采集，或者为不同信息源的遥感目标，相较于无人机拍摄的图片，距离更远导致目标也更小，这也是弱小目标检测任务的一种。但是在这种任务中，由于任务的特殊性，对目标的识别也提出了更高的精度要求，也说明了弱小目标检测的重要性。

因此对于弱小目标的研究对目标检测在尺度上形成更全面的体系有很大意义，无论在军用、商用、民用或者日常生活领域中，弱小目标都是一直存在的情况，对各种应用场景都有着重大意义。

1.2 本文研究内容

本文利用基于特征金字塔的融合因子方法对弱小目标检测任务进行研究。弱小目标检测任务有四个特点：1.目标的绝对尺度小，整个数据集的目标尺度平均大小为 18 个像素，小于其他目标检测数据集，这增加了检测任务的难度；2.目标的相对大小比较小，这说明了图片的结构特点，小目标一般是在远距离场景下拍摄的，所以图片分辨率比较大但目标尺度小所以相对比例小；3.目标信息量弱，因为目标尺度小，目标占据的像素点有限，所以目标会呈现出边界模糊，细节特

征少和上下文信息弱的情况；4.易受到噪声干扰，因为目标信息量弱，所以缺少语义判别信息，图片背景复杂时容易与目标本身产生混淆。以上四个特点也是弱小目标检测任务的四个难点。

在众多的卷积神经网络中，基于特征金字塔[9]（Feature Pyramid Network，FPN）的检测器由于其有效的特征融合机制和目标分层机制而成为应用最广泛的目标检测器之一。一方面，融合机制将高层语义信息与底层细节特征融合在一起，形成鲁棒的特征表达。另一方面，深层特征层也通过融合机制参与浅层特征层的分类和回归任务，不同尺度大小的目标也分别分配 FPN 不同的特征层上，让大目标在深层做分类和回归，小目标在浅层做分类和回归，也让神经网络有了更好的特征表达能力。基于 FPN 的检测器是处理弱小目标检测问题的常用方法。

基于特征金字塔的检测器通过自顶向下和侧向连接融合形成多尺度特征并结合目标分层策略，在常用的目标检测数据集上取得了巨大的成功，如 PASCAL VOC[10]、MS COCO[11]和 CityPersons[12]。然而，这些检测器在弱小目标检测检测精度并不高，例如在 TinyPerson[6]数据集上，Faster RCNN-FPN[13]的 AP_{50}^{tiny} 性能为 44%远低于同等条件下在 MS COCO 数据集上训练的结果。

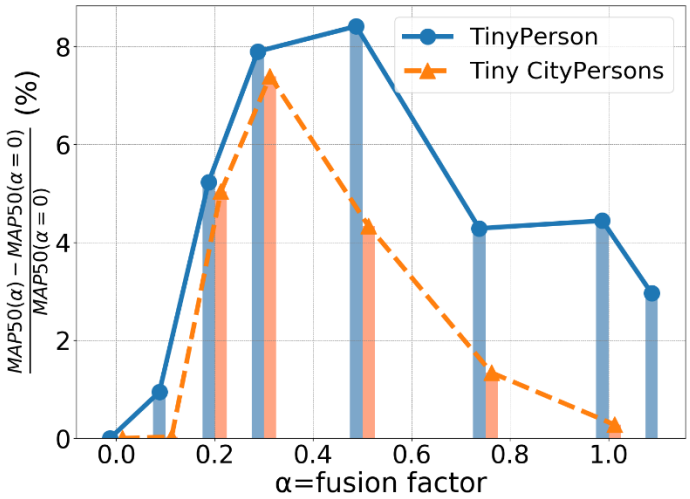

图 1.2　融合因子对弱小目标检测数据性能影响

Figure 1.2　The influence of fusion factor on the performance of tiny target detection

实验结果表明融合因子是影响弱小目标检测的精度的因素。融合因子（fusion factor，简称为α）为在 FPN 中融合相邻两层特征时，对较深层的加权系数。基

于特征金字塔的检测器在弱小目标检测的实验结果如图 1.2 所示，图的横坐标为融合因子的大小，从 0 到 1.1 递增，纵坐标为融合因子为某一值相对于融合因子为 0 时的性能，即融合因子对性能的影响程度。曲线显示随着深层传送到浅层的信息量增加，性能先增后降。

FPN 的结构图如图 1.3，首先将经过卷积处理后的深层特征上采样成与浅层特征层分辨率相同然后与同样经过卷积处理的浅层特征线性相加（深层与浅层是相对概念，P_3对于P_2来说是深层但对于P_4就是浅层），经过一次卷积计算就得到对应 FPN 的输出层特征。通常基于 FPN 的检测器将融合因子设置为 1。下图展示了此过程。果 FPN 融合了P_2，P_3，P_4，P_5，P_6级的特征，存在三个不同的融合因子α，包括α_2^3、α_3^4、α_4^5，分别代表两个相邻层之间的融合因子。由于P_6是通过直接对P_5进行下采样而生成的，因此P_5和P_6之间没有融合因子。

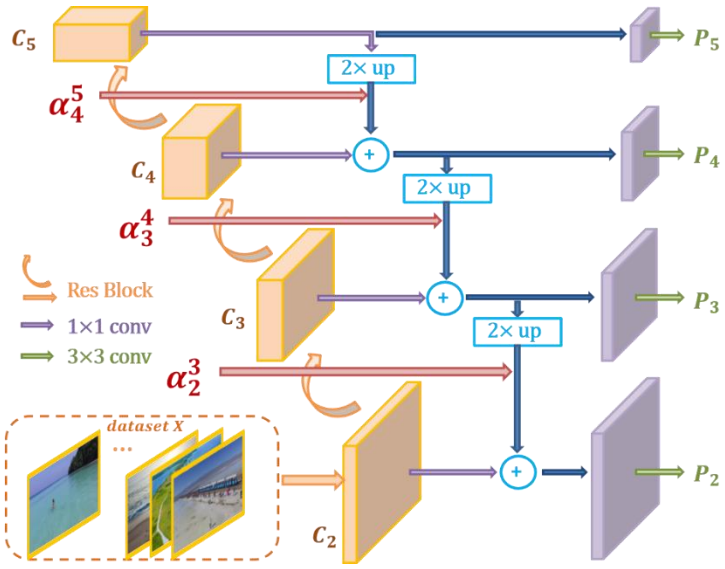

图 1.3　特征金字塔结构示意图

Figure 1.3　Schematic diagram of feature pyramid structure

FPN 的工作原理：1.通过自上而下和侧向连接的特征融合机制帮助检测器得到更好的特征表达，融合机制将富含细节信息的浅层特征和富含纹理和语义信息的深层特征融合；2.分层匹配机制将不同大小的目标对应到不同分辨率的特征层上学习，大尺度目标分配到深层特征层上，小尺度目标分配到浅层特征层，不同分辨率的特征层可以更专注于适合当前特征层的分辨率大小目标的学习。

高层特征往往是高度抽象并且具有高语义性，同时低层特征抽象程度不够但却含有丰富的细节特征，这有助于较小目标的识别，但是对于 TinyPerson 数据集来说不是这样，由于 TinyPerson 尺度分布的长尾效应，目标尺度都集中在[0,40)像素。FPN 目标分层时，属于深层的目标数量非常少，并且在特征融合的时候，深层特征不再含有高度抽象的语义信息，默认设置的特征层一比一的融合方式不合理，改变这种融合方式的是必要的。

本文从多个方面探讨了如何显式地学习 FPN 中的有效融合因子，以提高 FPN 在弱小目标检测中的性能，并提出合适的融合因子设置方法，同时从各个角度分析了融合因子的作用机理并给出相关解释，完成逻辑闭环。大量的实验结果表明，基于 FPN 融合因子显著提高了常用基于 FPN 的检测器在弱小目标检测中的性能。

1.3 本文主要贡献

本文对弱小目标检测这一任务提出了基于特征金字塔的融合因子方法，主要贡献总结如下：

1、提出了一个新的概念——融合因子，用来描述特征金字塔中相邻层的耦合程度，分析融合因子在特征融合时带来的双面影响；

2、从实用性以及性能提升两个方面进行分析，得到了最佳的基于统计的融合因子设置方法，在不增加网络的参数的情况提升检测精度；

3、对融合因子的原理进行了探究，从融合因子的显示学习和隐式学习两个方面，研究融合因子的作用前提和作用机理并给出相关解释与证明。

相关结果均得到实验验证，说明了方法的有效性，结论的可靠性，论文的研究成果具有理论意义和实用价值。

1.4 本文组织结构

第一章引言，分为研究的背景及意义与本文组织结构两部分。研究的背景及意义首先介绍了计算机视觉这一学科的发展历程，阐明了深度学习给计算机视觉这一学科带来的巨大变化，介绍了弱小目标检测的场景、应用前景、科研价值等，描述了目前小目标检测中常见的问题。然后详细介绍了常用于小目标检测的特征

金字塔网络的结构和作用机理，指出其不适用与小目标检测之处，同时提出一个影响弱小目标检测的性能的新概念——融合因子并加以介绍。本文组织结构部分按照章节简略介绍了论文各章的重点内容。

第二章国内外本学科领域的发展现状与趋势，分为了四个部分。第一个部分介绍了目标检测算法的发展以及相关算法和重要影响的工作；第二部分介绍了用于目标检测领域的几种典型数据集，划分为通用目标检测数据集，行人检测数据集，人脸检测数据集和弱小目标检测数据集并介绍代表工作；第三部分介绍小目标检测相关算法，分为用于多尺度目标检测中的小目标检测和弱小目标检测两个类别；第四部分从修改特征金字塔结构和改变特征融合策略两个角度分别介绍特征融合相关的研究进展。

第三章研究内容与方法，研究内容方面首先介绍了弱小目标检测的难点与特点，接着介绍了融合因子的研究背景，通过对融合因子在五个不同种类的数据集上实验的对比分析，验证了融合因子对弱小目标检测性能有关键影响，同时说明了特征金字塔的作用原理以及融合因子在其中起到的效果。在研究方法方面，从固定参数方法、可学习参数方法、注意力机制参数方法、统计参数方法等各个角度阐述了最佳的融合因子设置方法，并且给出了适用于弱小目标检测的特征补充模块帮助检测器提高性能。

第四章实验结果与分析，实验结果方面首先介绍了目标检测领域及论文采纳平均准确率（Average Precision，AP）和丢失率（Miss Rate，MR）两种评价指标的原理和计算方式，然后介绍了两种基准实验的实验设置，接下来分别介绍实验方法结果对比，基于统计的融合因子横向与当前主流算法对比体现出性能优势，纵向在不同实验条件下的性能都取得了提升。另外在不同数据集上也验证了方法的可推广性，最后介绍了特征补充模块的实验结果。实验分析方面为网络自主学习能力对融合因子有效性的影响并且分析了融合因子在多尺度数据集上的表现，以及从网络梯度回传的角度解释了融合因子起作用的原因。

第五章结论与展望，总结了全文的行文逻辑，并且从论文章节的角度概括了每章的重点。并且从不同角度展望了未来的研究工作。

第 2 章　国内外本学科领域的发展现状与趋势

在深度学习出现之前，计算机视觉的任务主要是以图像数据计算、特征工程为主的图像压缩、增强、去噪、识别、配准等等。深度出现之后，基于卷积神经网络的图像处理成为主流，在各个应用方向上均取得了长足的进步，比如图片分类、目标检测、场景分割、目标跟踪等等。

本章将从相关的检测算法发展开始阐述，并对有关数据集进行介绍与特点比对，然后介绍小目标检测算法的前沿研究并对特征金字塔机制进行探讨。

2.1　目标检测算法的发展

目标检测是计算机视觉领域中长期存在的一个基本任务，近些年也是研究的重点领域。目标检测的任务一般定义为给定一张图像，判断图中是否存在预先定义的类别实例，如果存在则返回类别和空间的最小外接矩形的坐标，是图像级别的分类任务的扩展，对图片的处理结果也从图像级别转化为实例级别。在 2012 年前，目标检测领域的方法还是以提取人工设计的特征为主流，包括 SIFT[2]、HOG[3]、DPM[4]、HOG-LBP[14]等，这时候的处理问题流程一般为：区域选择，特征提取、分类回归三步，虽然在一些问题上取得了初步的成果，但是却有两个难以处理的问题：一是区域选择算法效果差，计算量大，基于滑动窗口策略需要进行像素级的计算，这就导致窗口数量太大，而且随着像素的数量呈指数级增长，如果为了涵盖多尺度或者不同长宽比的物体，计算量又会成倍增长，无法做到实时性的检测；二是基于人工设计特征提取泛化能力不好，同一个方法在不同的数据上表现差异可能很大。Krizhevsky[15]等在 2012 提出了用于图片分类的深度卷积神经网络，标志着目标检测也正式进入深度学习时代。

深度学习之后，检测算法按照处理阶段分类主要可以分为两种：一是不使用区域提议阶段的单阶段（one stage）算法，最后的结果选自于预先在图中密集设定的锚点框；二是基于区域提议的两阶段（two stage）算法，这类算法第一步得到无类别差异的前景提议区域，然后对提议区域进行分类和回归。

在两阶段算法中，Ross Girshick 在 2013 首次提出了影响重大区域卷积神经网络（RCNN）[16]，这也是卷积神经网络第一次应用于目标检测任务，后续的工作基本都遵循了其两阶段算法的处理思路。RCNN 对传统的基于滑动窗口的算法改进在于：1.传统的检测算法一个窗口就会完成一次检测过程，但相邻窗口像素重叠大，因此带来了计算冗余，RCNN 使用了一个启发式方法（Selective search），先生成候选区域再检测，降低信息冗余程度，从而提高检测效率；2.使用了卷积网络提取了目标特征避免了传统算法提取特征鲁棒性不够。但是这个算法仍有许多不足，首先，其训练是多阶段的，较为繁琐和耗时；其次，由于在高密度的候选区域上反复进行特征提取，其检测速度很慢（GPU 下每张图 40 秒，640×480 像素）[17]。

何恺明在 2014 年提出了 SPP Net[18]进一步改进了 RCNN：1.提出了先卷积计算再生成候选区域的策略，之前的先生成候选区域再卷积计算的方法仍然在临近区域有大量的重复计算，特征被重复提取，如果先卷积计算就可以通过一次计算满足特征提取的需求；2.提出了空间金字塔池化层（Spatial Pyramid Pooling）的特征池化方法，由于全连接层的存在，经过深度卷积处理后的特征必须处理成相同大小才能输入到全连接层，这就要求图片在输入网络的时候要统一到相同的尺寸，为了将不同大小归一化，一般会采取裁剪或者缩放的手段，但无论是哪种方法都会强制图片变形，破坏图片的结构，而空间金字塔池化层允许输入不同大小的图片输入网络，但在全连接计算前可以将不同大小的特征归一到同一大小，打破图片输入的尺寸必须统一这一束缚。

在 2015 年 Ross Girshick 又提出了 Fast RCNN[19]在 RCNN 的基础上又充分吸收了 SPP Net 的优点，摒弃了单独训练 SVM 的分类器的方式，并且将分类器和回归器并行设计，极大提高了计算速度。同年任少卿提出了第一个端到端的两阶段检测器 Faster RCNN[13]，两阶段算法中第一阶段生成的预选区域的质量直接决定了第二阶段的效果，在以往选择候选区域都采用的是启发式的算法，如果生成的候选区域太多会造成计算冗余，如果太少又会产生误检，而且卷积计算都是由 GPU 实现的的，而启发式算法是在 CPU 实现，这也降低了计算效率，在 Faster RCNN 首次提出了 RPN（Region Proposal Networks），将提取候选区域用卷积网

络实现，大大减少了计算量和耗费时间。并且首次引入了 Anchor（锚点框）的概念，通过设定密集分布的 Anchor 判断对应区域是否可以作为目标候选区域，为后来的基于锚点框的方法打下来基础。何恺明在 2017 年又提出了 Mask RCNN[20]，就 ROI Pooling 中造成的特征不对齐问题提出了 ROI Align 解决，并且将目标分割和特征金字塔引入同一框架中也显著提升了性能。之后有很多文章对这些进行了一处或者多处的改变，比如 RefineDet[21]，Cascade RCNN[22]，Grid RCNN[23]，Libra RCNN[24]都取得了不错的性能提升。

一阶段算法没有采用了二阶段算法的“粗检测+细检测”的模式，直接在由卷积提取的特征上进行分类回归，经过后处理后产生检测结果，和二阶段检测算法相比，虽然损失了性能有所下降但是却极大提升了速度，可以满足实时性需求，代表算法有 YOLO[25]系列、SSD[26]、RetinaNet[27]等。

YOLO（You Only Look Once）[25]是第一个一阶段检测算法，由 Joseph 和 Girshick 等人在 2015 年提出。算法直接将整张图像统一到固定大小作为网络的输入，经过一系列卷积计算后再接全连接层直接得到预测结果。并且 YOLO 中没有锚点框（Anchor）的概念，引入了 grid 来做区域性划分达到分而治之的目的，但是子区域（grid）的划分过大这就导致 YOLO 很容易丢失小目标，导致最后的检测精度不足，好处是大幅度提高了检测速度，该算法的增强版本在 GPU 上速度为 45 帧/秒，快速版本速度为 155 帧/秒（640×480 像素）[17]。

Wei Liu 等人于 2015 年提出 SSD[26]。SSD 算法在 YOLO 速度快和 Faster RCNN 的基础上做了进一步改进。主要贡献有三点：1.在一阶段检测算法中首次引入了锚点框（Anchor）的概念；2.实现了在不同尺度和深度的特征图上做检测，这让不同大小的物体在适合的特征层上做检测，提高了检测效率；3.采用了难样本挖掘的策略一定程度上解决了一阶段检测算法中的正负例样本数量不平衡问题。在 VOC2007 上取得了接近 Faster RCNN 的准确率（mAP=72%），同时保持了极快的检测速度（58 帧/秒，640×480 像素）。

在 Tsung-Yi Lin 提出 RetinaNet[27]之前，单阶段检测器虽然在速度上优于两阶段检测器但是性能一直低于同期的双阶段检测器。Tsung-Yi Lin 等人分析了原因认为由于单阶段算法没有候选框提取这个过程，没有经过初步筛选的样本全部

参与到分类和回归任务中导致正负例样本比例不平衡，单个简单负样本虽然产生的梯度较小，但是由于数量巨大在梯度中起了主导地位给网络学习带来负面影响。为了解决这个问题，RetinaNet 中提出了 Focal Loss，通过两个参数的限制降低网络训练过程中简单负样本的学习权重，使网络更专注于其他样本的学习，取得了和两阶段检测器类似的性能。

2.2 用于检测的数据集

目标检测是计算机视觉中的基础任务，主要解决的是图片中的目标分类与坐标回归。目标检测在实际应用中有着丰富的场景，针对不同的场景问题需求有不同的数据集已经被发表出来。按照任务的类型主要有以下：1.通用目标检测如 PASCAL VOC[10]、MS COCO[11]、LVIS[28]等；2.人脸检测如 WiderFace[29]；3.行人检测如 Caltech USA[30]，CityPerson[12]；4.弱小目标检测如 TinyPerson[6]。各种任务的代表场景如图。图 2.1 为不同种类的数据集的典型场景示例（a）弱小人体检测——Tinyperson; (b)行人检测——Citypersons; (c)通用目标检测——MS COCO;（d）人脸检测——WiderFace。

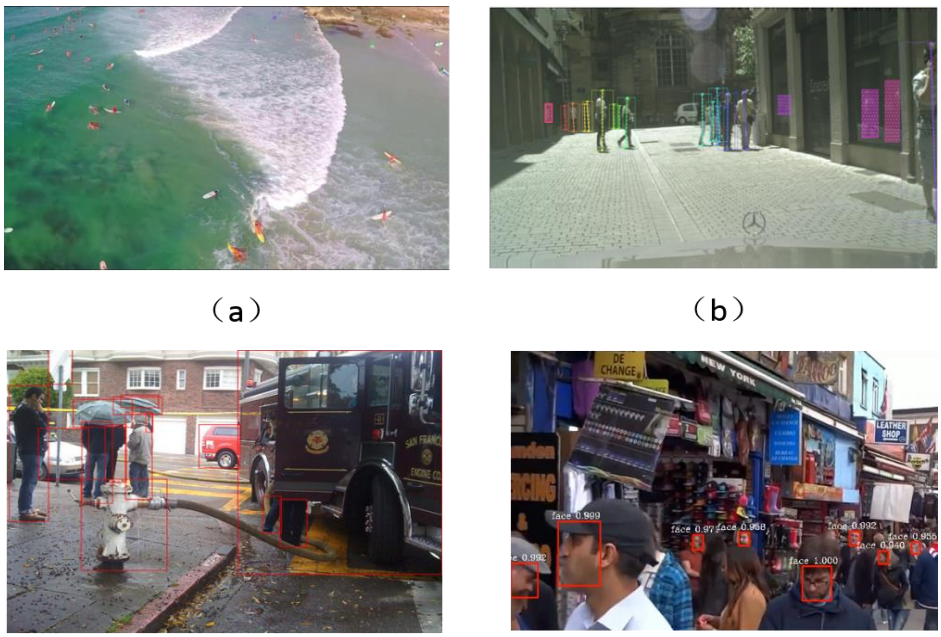

图 2.1 不同数据集的典型图例

Figure 2.1 Typical illustrations of different datasets

通用目标检测最早的代表数据集是 PASCAL VOC[10]数据集（简称 VOC）有 VOC 2007 和 VOC 2012 两种版本，在其上举办的比赛对目标检测领域的发展起到了极大的推动作用。数据集总共分为 4 个大类 vehicle，household，animal，person，总共 20 个小类，两个数据集训练和验证部分总计有 16551 张图片 40058 个标注框，测试部分有 16492 张图片和 39482 个标注框，总计有 33043 张图片和 79540 个标注。

微软团队在 2014 年发布了 MS COCO[11]其全称是 Microsoft Common Objects in Context，这是一个跨尺度数据集。其标注信息十分丰富，可以用来进行目标检测、关键点检测、语义分割、字幕生成等任务。MS COCO 数据集中的图像分为训练、验证和测试集。由于巨大的数据体量、丰富的类别和完整的标注信息，是通用目标检测中最具有影响力的数据集。该数据集包含 33 万张图像、150 万目标实例、80 个目标类、91 个物品类以及 25 万目标关键点。相较于 VOC，MS COCO 有着更细致的类别划分，更多的标注数量和更大的尺度覆盖，是目标检测领域目前最有影响力的数据集。

FAIR 开放了 LVIS[28]，一个大规模细粒度词汇集标记数据集，包含了 164k 图像，并针对超过 1000 类物体进行了约 200 万个高质量的实例分割标注。由于细致的分类标注，所以这个数据集的物体类别数量具有天然的长尾分布特性（大部分类别物体数量丰富，但一些类别的物体数量非常少）也满足目标检测的实际任务场景，即如何让网络有效地从小样本中学习，这也对检测任务提出更大的挑战。2019 年，旷视研究院发布了新的大型目标检测数据集 Objects365[31]，它拥有超过 600,000 个图像，365 个类别和超过 1000 万个高质量的边界框，进一步提升了数据集的体量，对推动了目标检测领域发展。

行人检测在实际场景中应用广泛，一直是计算机视觉领域的热点应用方向。随着研究的深入，具有更大的容量、更丰富的场景和更好注释的行人检测数据集相继被发表出来，如 INRIA[3]、ETH[32]、Daimler[33]、Caltech USA[30]，KITTI[34]和 CityPersons[12]，同时这也代表了对泛化能力更好的算法和更高的性能的追求。Caltech USA 是加州理工学院发布的行人检测数据集，来自在城市环境中正常行驶的车辆拍摄的视频的切帧图片，视频大约 250,000 帧（137 分钟左右），共计

350,000 个标注框和 2300 个被标注的独立行人。张珊珊在 Cityscapes[36]数据集上建立了 CityPersons 数据集，在 5000 张图像上标注了 35000 个行人，13000 个忽略区域，同时遮挡场景进行了很好的标注，而且图像在不同的季节采集于德国 27 个不同的城市，保证了图像场景的丰富程度。

WIDER FACE[29]数据集是人脸检测的一个 benchmark 数据集，由香港中文大学选择了 61 个事件类别，共包含 32203 图像，以及 393,703 个标注人脸，其中，158,989 个标注人脸位于训练集，39,496 个位于验证集，检测难度划分为 Easy，Medium，Hard。WIDER FACE 场景丰富，涉及了人脸在不同的尺度，姿态，光照，表情，妆容，遮挡条件下的情况。

作者作为参与者构建了弱小目标检测代表数据集 TinyPerson[6]，TinyPerson 发表于 2019 年，共计 1610 张图片，72651 个标注框，图片由无人机在远距离广视角下拍摄，因此图片中的目标天然具有绝对尺度小这个特点，整个数据集的目标平均绝对大小为 18 个像素，这也是弱小目标检测的最大特点。数据集还对标注类别进行了细致的划分，包括位于陆地上的人，位于海里的人和不确定为人的区域，可满足不同任务的需要。

TinyNet[35]涉及远程遥感目标检测。VisDrone[8]数据集由无人机采集于中国 14 个不同城市的不同城市/郊区，保证了场景的多样性。整个数据集共包括 288 个视频片段，261908 帧和 10209 张图像，共计 260 多万个常用类别的标注框。

从数据的目标平均大小来看，遥感目标检测和人脸检测都与弱小目标检测相近，但是在其他方面仍存在差异。以 Wider Face 为例，Wider Face 的数据集目标平均大小为 32.8 像素，接近于弱小目标的标准的 20 像素，但是人脸检测的目标长宽比固定而且人脸检测问题可以利用上下文信息来提高识别精度，但是对于视角多样，信息量弱，长宽比不一的弱小目标检测数据集，这些信息便无法利用。遥感目标检测在尺度上更接近于弱小目标检测，但是遥感图像中的目标多为舰船、飞机等目标，具有特定的长宽比和密集排列等特点，这些也是不同于弱小目标检测的地方。表 2.1[6]为各种典型数据集的统计特征对比，其中正负号前边为数据的平均数，后边为数据的标准差。

表 2.1 典型数据集统计特性[6]

Table 2.1 Statistical characteristics of typical datasets

数据集	绝对大小	相对大小	长宽比
TinyPerson	18.0±17.4	0.012±0.010	0.676±0.416
COCO	99.5±107.5	0.190±0.203	1.214±1.339
Wider face	32.8±52.7	0.036±0.052	0.801±0.168
Citypersons	79.8±67.5	0.055±0.046	0.410±0.008

2.3 小目标检测

目前国内外对于小目标检测的研究主要可以分为两大类：一、针对跨尺度检测中的小目标，在 COCO 的评测标准中按照尺度（物体面积开根号后得到的值）对物体的大小区间进行划分，[0,32)像素为小尺度目标区间，[32,96)像素为中尺度目标区间，[96,+∞)像素为大尺度目标区间，COCO 就是一个典型的涵盖了各种尺度目标的跨尺度数据集。当研究目标为跨尺度数据集时，检测大尺度目标的难度要低于小尺度，提升整体性能关键在于提升小尺度目标检测的性能，设计出针对小目标策略但是同时要保证其他尺度目标性能不能下降；二、针对弱小目标检测，TinyBenchmark[6]提出对于[0,32)尺度区间可以进行更细粒度的划分，[2,20)为弱小尺度区间，[20,32)为小尺度区间，物体尺度小于 2 个像素的时候一般不会参与网络训练。划分原因是 COCO 中缺少尺度小于 20 像素的目标，这一部分目标检测难度更大更需要设计特定的策略。TinyBenchmark 同时提出由弱小目标占主导的数据集 TinyPerosn，提升性能只需要设计针对弱小目标的策略不需要考虑其他尺度目标的性能变化。

2.3.1 基于多尺度的小目标检测

针对跨尺度检测中的小目标的大多数研究都是为了获取尺度不变性时，发现小目标性能很差，再对小目标进行处理，因此绝大多数关于小目标检测的研究都是将小目标检测作为通用目标检测中一个附带的子问题来进行研究。已有许多学

者对小目标的检测也进行了广泛的研究。SNIP[37]和 SNIPER[38]利用图像金字塔同时使用尺度正则化策略来保证目标的大小在一个固定范围内，将小目标放大提升检测精度。SNIPER 采用区域抽样的方法进一步提高训练效率。

小目标性能精度比较低的重要原因是小目标自身尺度小，信息量少，细节特征不够多。超分辨率（Super Resolution）常用于恢复低分辨率目标的信息，用网络的学习能力自动补充小目标缺乏的细节特征，提高其分辨率，因此一些工作已经将其引入到小目标检测中。Noh J[40]认为小目标检测在 ROI 特征提取过程中容易失真，因此提出了一种利用高分辨率目标特征作为监督信号的超分辨率方法，通过用相对容易被网络学习的大尺度目标引导小尺度目标的学习，并且在小尺度目标上取得了性能提升。Chen[41]等人分析了 COCO 中不同尺度目标所占损失的比重，发现小目标贡献的损失并不主导网络的学习，因此提出了一种反馈驱动的数据提供策略，将包含大目标的图片拼接产生小目标，利用损失统计信息来平衡小目标检测的损失。

三叉戟网络（TridentNet）[42]中分析了影响检测器性能的三个因素：网络深度、网络下采样倍率、感受野大小，所提网络结构如图 2.2[42]，构建了不同感受野的平行多分支，并生成了更具辨别力的小目标特征，同时采用不同分支共享参数的办法降低参数量，并且让不同大小的目标适应于不同的感受野分支学习，提高了小目标检测的性能。IPG-Net[43]提出了浅层特征富含空间和细节特征但是缺少高级的语义信息，深层有语义信息但缺少空间特征，这种信息不平衡是阻碍性能提升的原因。因此作者提出了图像金字塔引导网络，通过图像金字塔信息转化模块和图像金字塔融合引导模块解决信息不平衡问题从而减轻了小目标特征的信息丢失。这些上述方法在一定程度上提高了小目标检测的性能，但都权衡了其他尺度的性能变化，保证了检测器在各尺度目标的检测精度。

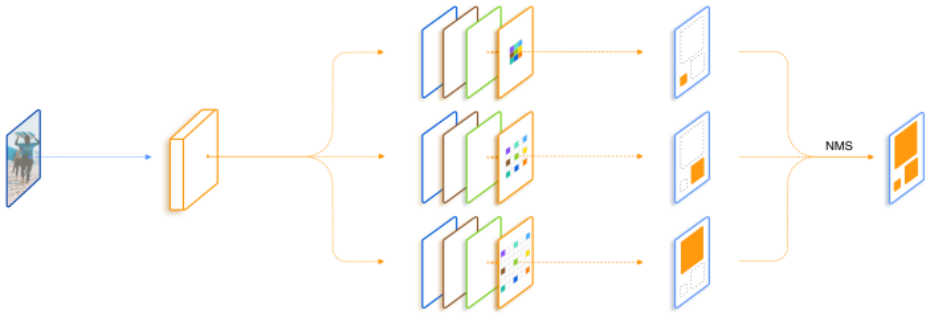

图 2.2 三叉戟网络结构示意图[42]

Figure 2.2 Schematic diagram of TridentNet structure

2.3.2 基于单尺度的小目标检测

基于单尺度的小目标检测即弱小目标检测，因为弱小目标数据集的物体尺度分布的平均大小集中于 20 像素左右，整体数据集为单一小尺度的数据集，因此设计算法的时候只考虑提升小目标的检测精度即可。

扩展特征金字塔（EFPN）[39]的作者认为小尺度目标和中尺度目标都集中于 FPN 中的浅层特征层，对小目标检测不利，因此利用超分辨率的思想构造了新一个具有更多几何细节的特征层，如图 2.3[39]，通过特征纹理转移模块将适合做小目标检测的特征充分融合，并进一步接受来自更浅层骨干网络的特征信息组成适合小目标检测的特征层。

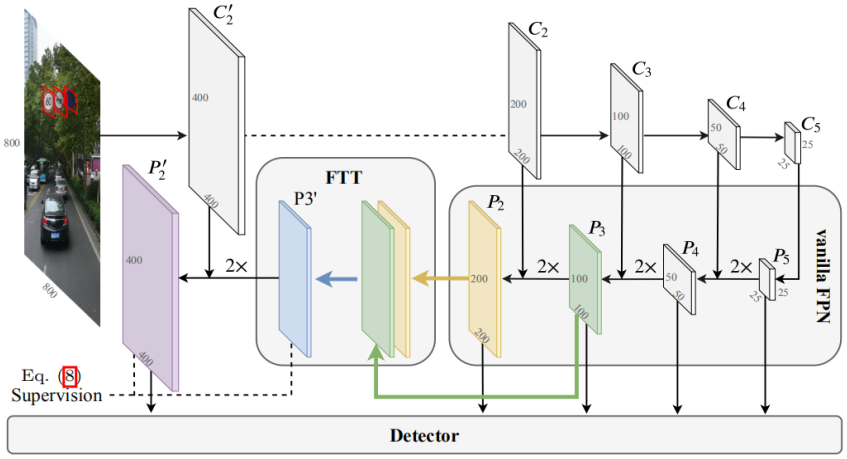

图 2.3 扩展特征金字塔结构示意图[39]

Figure 2.3 Schematic diagram of extended feature pyramid structure

在 TinyBenchmark[6]中，提出了一个迁移预训练数据集的尺度分布向目标数据集的尺度分布接近的方法来提升性能，从数据集预训练的角度为弱小目标检测提供了一种新的研究思路。作者在实验中发现，用于网络的权重初始化的预训练数据集和用于检测器训练的目标数据集之间的尺度分布的不匹配会弱化网络的特征表示能力和降低检测器检测精度。因此，作者提出了一种简单而有效的尺度匹配方法，将两个数据集之间的尺度分布对齐，在预训练的过程中让网络学习到目标数据集的尺度分布特性的相关知识，以获得对有利于弱小目标特征表示能力。实验结果证明基于尺度匹配的算法在不同的检测器有显著的性能提高。具体过程如下图[6]所示：

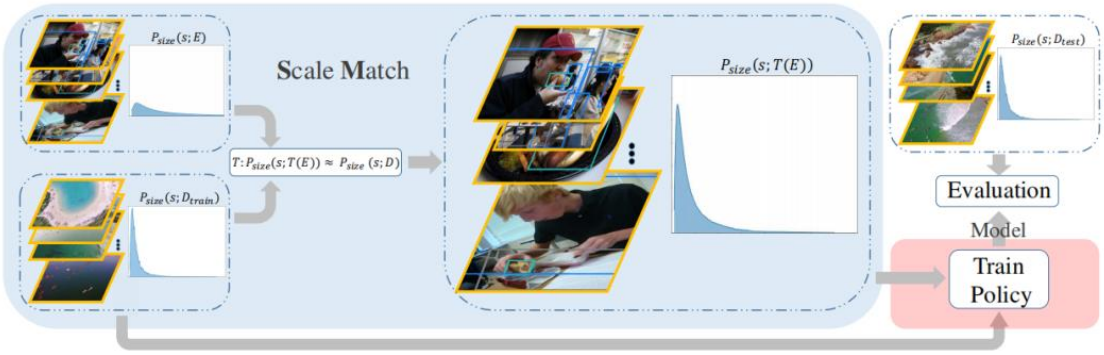

图 2.4　尺度匹配算法流程示意图[6]

Figure 2.4　Schematic diagram of scale matching algorithm

2.4　特征金字塔

为了解决目标检测中的多尺度问题，一般有建立图像金字塔和建立特征金字塔两种方法。图像金字塔方法是将图片进行不同尺度上的缩放，构成一个依据图片尺寸建立，图片由小到大排布的“金字塔”状结构，其目的是让目标大小处于一个合适的范围，让原本的大目标在小尺度图片检测，让原本的小目标在大尺度图片上检测，代表工作有 SNIP[37]，SNIPER[38]。图像金字塔的方法显著提升了小目标检测的性能，但是由于不同尺度的图片的相同区域都需要网络计算，带来了计算冗余，同时放大后的图片也会成倍增加计算量，带来了沉重的计算负担。因此特征金字塔方法[9]（FPN，Feature Pyramid Network）被提了出来，作者将固定尺寸的图片在网络计算不同阶段不同大小的特征图提取出来，构成了“金字塔”

状结构，同时引入了特征融合机制，在小幅增加计算量的情况下，带来了在多尺度目标检测精度提升。

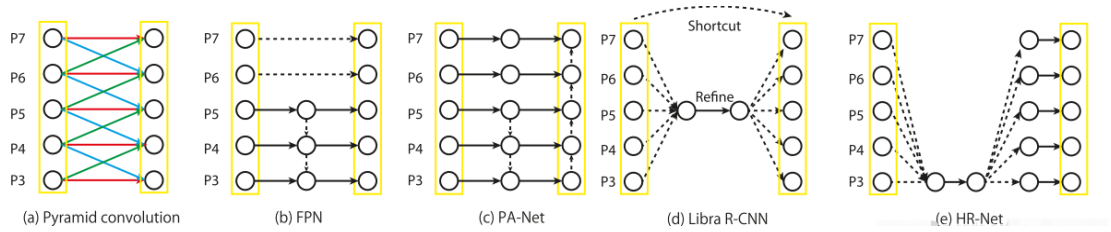

图 2.5　不同的特征融合方式[48]

Figure 2.5　Different feature fusion methods

特征融合的方法已经在不同的检测算法都取得了性能上的提升[9][24][44][48]。图片经过深度卷积神经网络得到的所有特征层中，浅层特征层经过较少的卷积计算，尺度较大且保留了丰富的纹理和几何细节而缺少抽象语义信息，适合小目标检测任务。而深层特征层与浅层相反，深层特征层经过大量的卷积，富含抽象的高级语义信息适合大目标检测任务。上图列举了几种不同的特征融合方式。FPN[9]（图 2.5 的 b[48]）采用了自顶向下和侧向连接的特征融合方式，将经过上采样的深层特征线性相加于浅层特征，以构建特征金字塔。在此之后出现了一些改进 FPN 的工作，总结下来可以分为两种：1.基于结构改进特征金字塔，比如扩展网络结构新加信息通路，让然后不同尺度的特征更充分融合；2.基于策略改进特征金字塔，比如增加多尺度融合次数得到更好的特征表达。

2.4.1　基于结构改进特征金字塔

PANet[44]（图 2.5 的 c[48]）在 FPN 的基础上增添了一条自底向上的信息聚合通路，可利用浅层丰富的细节特征增强所有的特征层的定位信息，同时也缩短了深层特征到浅层特征的路径长度，提高了网络学习效率。Nie[45]在 SSD 的基础上引入了 MSCF 模块通过提取多尺度的上下文特征来丰富从骨干网络提取出来的特征，并使用级联优化方案处理自下而上的特征层次的分类和回归任务。HRNet[46]（图 2.5 的 e[48]）增加了高分辨率到低分辨率的特征子网，并且以并行链接方式连接多分辨率子网，通过重复交叉并行卷积反复进行多尺度特征融合使得特征充分融合增强特征表达。Nas-FPN[49]改变了 FPN 人为设计的层间结合方

式，探索了使用 AutoML 在给定搜索空间内寻找每一层特征融合的最佳组合方法。

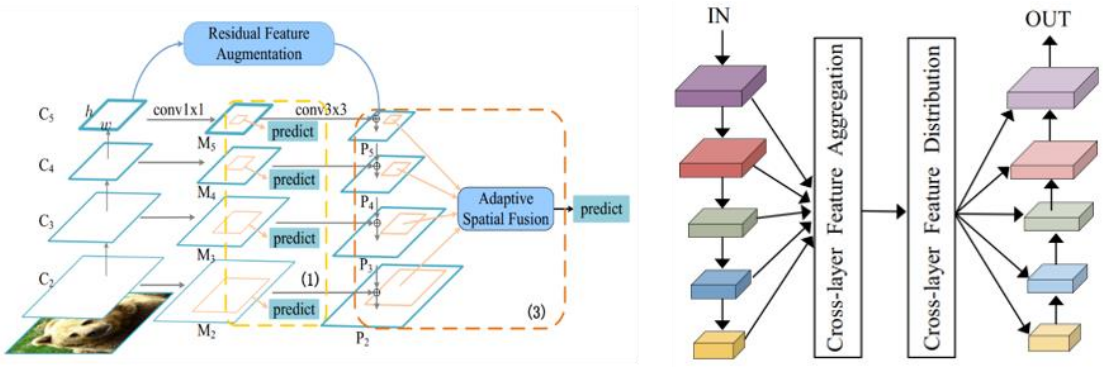

图 2.6　AugFPN 结构示意图[51]（左）和 CL-FPN 结构示意图（右）[52]

Figure 2.6　AugFPNstructure diagram (left) and CL-FPN structure diagram (right)

2.4.2　基于策略改进特征金字塔

Libra RCNN[24]（图 2.5 的 d[48]）提出无论是 FPN 还是 PANet 都是考虑了相邻分辨率层的特征融合，跨越特征层的特征融合会将非相邻层的语义信息稀释，因此提出了平衡特征金字塔（BFPN），生成一个由所有特征层参与的特征平衡层去补充各层的多尺度信息得到更平衡的特征表达。ASFF[47]改变了传统的线性相加的特征融合方式，引入了注意力机制自适应地进行特征融合，达到了更好的多尺度融合效果。

SEPC[48]（图 2.5 的 a[48]）提出了金字塔卷积算法来挖掘内相邻特征层在尺度的关联性，以提高相邻特征层的特征融合效率。Tan[50]提出了 BiFPN,相较于 FPN 增加了跨层链接通道，增强了多尺度特征融合，虽然在 NAS-FPN 已经存在了跨层链接通路，但是搜索算法的效率带来了巨大的计算量。BiFPN 增加了特征融合时的可学习权重，改变了传统方法直接堆叠的方式，网络自行学习不同输入特征层的融合的权重。

AugFPN[51]（图 2.6 左侧）通过增加一致性监督，减少了不同分辨率的特征层之间的语义差异，在特征融合过程中采用残差特征增强的方法按按比例融合特征，减少深层特征在融合过程的语义丢失，并且还通过在不同的特征层上软化提取 ROI 特征选择方式为后续检测任务提供更好的 ROI 特征。CL-FPN[52]（图 2.6 右侧）认为多尺度特征融合是影响显著性目标检测的关键，提出了将 FPN 不同

层的特征先聚合后分散机制，让不同的分辨率的特征通过聚合机制充分融合方便形成具有多尺度融合结果的各层特征。这些方法从不同方面提高了特征融合的效果。但是，他们都没有考虑特征融合受数据集尺度分布的影响。

2.5 本章小结

本章从四个方面详细的阐述了与本研究相关的国内外本学科领域的发展现状与趋势，分别是检测算法的发展历程、检测相关的数据集、小目标检测、特征金字塔。检测算法方面主要介绍了主流检测算法的原理和发展历程；数据集方面介绍了用于不同目标检测任务的数据集并进行尺度特点的比对；小目标检测介绍了多尺度和单尺度下的两种小目标检测；特征金字塔介绍了与 FPN 相关的算法发展历程。下一章将介绍研究内容与方法。

第 3 章　基于融合因子的弱小目标检测方法

3.1　融合因子研究背景及意义

弱小目标检测的难点主要在于两方面，一是目标信号弱、信息量少，二是目标尺度小，而尺度小又有相对尺度小和绝对尺度小两层含义。目标信号弱在目标所处环境单一且与目标区别明显的时候并不构成关键难点，但在环境复杂且与存在大量与目标相似度很高的内容时将成为一个核心难点，这一情况下处理方式一般为引入上下文信息利用目标周围的环境信息帮助目标的分类和回归或者放大图片，但是由于目标本身信息量少过度放大图片也不能显著提升性能，可以采用一些超分辨的方法提升目标区域的信息量。

FPN 作为处理多尺度检测任务的手段，是适合处理弱小目标检测的算法，影响小目标检测的 FPN 性能有两个主要因素，包括下采样因子和相邻特征层之间的融合比例。先前的研究[39]已经探索了前一因素，并得出结论，下采样比率越低，特征图越大，适合小目标检测的细节特征越丰富，越适合做小目标检测，性能将越好，但计算复杂度会成倍增加；然而，相邻特征层之间的融合比例这一因素常被忽略。

FPN 的特征融合的方式可以表示为图 1.3，表示为公式形式为：

$$P_i = f_{layer_i}(f_{inner_i}(C_i) + \alpha_i^{i+1} * f_{upsample}(P'_{i+1})) \tag{3.1}$$

其中f_{inner}是用于通道匹配的 1×1 卷积运算，$f_{upsample}$表示用于分辨率匹配的 2x 上采样运算，f_{layer}通常是用于特征处理的 3×3 卷积运算，而α_i^{i+1}表示用于第 i 层和第 $i+1$ 层的融合因子。常规检测算法中将 α 默认设置为 1。图 1.3 展示了此过程。 实际上，如果 FPN 融合了P_2，P_3，P_4，P_5，P_6级的特征，则存在三个不同的 α，包括α_2^3，α_3^4和α_4^5，分别代表两个相邻层之间的融合因子。（由于P_6是通过直接对P_5进行下采样而生成的，因此P_5和P_6之间没有融合因子）。融合时，可通过设置不同的 α 来调整来自不同层的特征的融合比例。FPN 中具体不同层的融合公式如下：

$$P_5 = f_{layer_5}(f_{inner_5}(C_5)) \tag{3.2}$$

$$P_4 = f_{layer_4}(f_{inner_4}(C_4) + \alpha_4^5 * f_{upsample}(P_5')) \tag{3.3}$$

$$P_3 = f_{layer_3}(f_{inner_3}(C_3) + \alpha_3^4 * f_{upsample}(P_4')) \tag{3.4}$$

$$P_2 = f_{layer_2}(f_{inner_2}(C_2) + \alpha_2^3 * f_{upsample}(P_3')) \tag{3.5}$$

在实验中发现，通过调整融合因子可以对弱小目标检测的性能造成明显影响，于是进一步设计实验探究最佳的融合因子设置方式和最佳的数值区间，和融合因子能对弱小目标检测性能起作用的深层并且给出相关分析。

为了进一步探索融合因子有效的原因及作用范围，首先研究是数据集哪方面的因素会影响融合因子的有效性，探究融合因子作用的前提是什么。首先假设数据集的四个属性会影响融合因子的有效性：1.目标的绝对尺度大小；2.目标的相对尺度大小；3.数据集的数据量；4. FPN 中每层的目标分布。

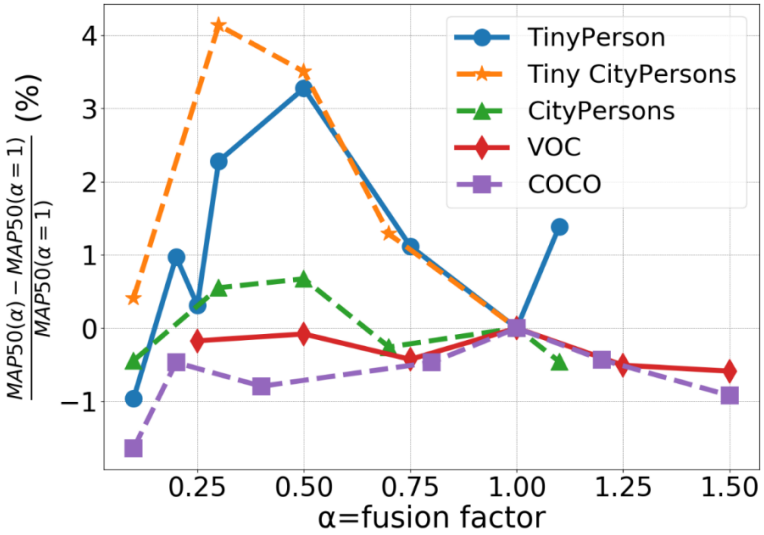

图 3.1　融合因子的变化对不同数据集的性能影响

Figure 3.1　The influence of fusion factor on the performance of different datasets

首先，实验以评估融合因子对不同数据集的影响。实验结果在图 3.1 中给出，图的横坐标为融合因子的值，取值范围是[0,1.5]，纵坐标为融合因子取特定值时相对融合因子为 0 时性能提升的程度。在不同的融合因子下，不同的数据集的性能变化表现出不同的趋势，例如曲线峰值大小和出现的位置。跨尺度数据集 CityPersons，VOC 和 COCO 对 α（融合因子）的变化不敏感，除非当 α= 0 时（这

意味着没有特征融合）。但是，在 TinyPerson 和 Tiny CityPersons 数据集上，性能随 α 的增加先升后降，这意味着融合因子是影响其性能的关键因素，并且存在一个最佳取值范围。由于当融合因子大于 1.1 难以在 TinyPerson，Tiny CityPersons 和 CityPersons 上进行收敛，因此图中未显示相关实验结果。

TinyPerson 和 Tiny CityPersons 数据集的共同特征是标注框的平均绝对大小小于 20 个像素，由此带来的正负例不平衡和分类回归难度变大等给网络的学习带来了巨大的挑战。为了验证是否是数据集标注框绝对大小导致的不同数据集对融合因子的变动敏感程度不同，在实验中调整了 CityPersons 和 COCO 数据集输入网络训练时图像的大小以获得不同的绝对尺度大小数据集，CityPersons 中的图像分别缩小 2 倍和 4 倍，分别为 Tiny CityPersons×2，Tiny CityPersons，COCO 中的图像分别缩小 4 倍和 8 倍，分别为 COCO200，COCO100。

如图 3.2，横坐标为融合因子的值，取值范围在[0,1.1]，纵坐标为对应融合因子取值下的性能相对于融合因子为 0 时变化的相对比例，图中的曲线体现出了明显的趋势，在融合因子递增的情况下，随着 CityPersons 输入图片尺度的变小，性能上的峰值现象越来越明显。Tiny CityPersons 的峰值性能变化程度是 CityPersons 的峰值性能变化程度的 4 倍。

图 3.3 与图 3.2 不同之处在于图 3.3 目标数据集是 COCO，横坐标融合因子的值区间更大，为 0 到 2，COCO800 的训练网络用的是 RetinaNet（用P_3，P_4，P_5，P_6，P_7构建 FPN），COCO200 和 COCO100 用的都是 adaptive RetinaNet（用P_2，P_3，P_4，P_5，P_6构建 FPN），但变化趋势与图 3.3 类似，随着 α 的变化，性能都是先增后降，而且输入图片尺寸越小峰值现象越明显。另外，对于 Tiny CityPersons 和 CityPersons，数据量和目标的相对大小完全相同。但是，当融合因子增加时，性能变化趋势不同，Tiny CityPersons 有超过 4%的性能波动，CityPersons 只有不到 1%的性能波动。

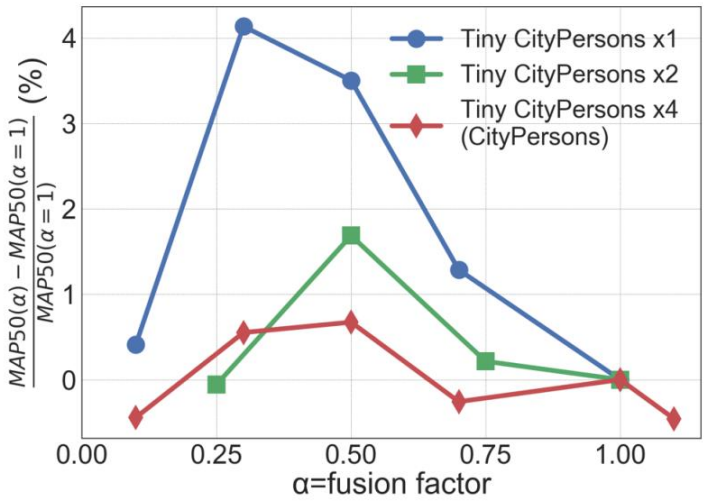

图 3.2　融合因子的变化对不同尺度的 CityPersons 的性能影响

Figure 3.2　The influence of α on the performance of Citypersons with different scales

FPN 每层中目标的分布将决定每个特征层上的正例的训练样本是否足够以及正例样本和负例样本的比例，这直接影响每层卷积的特征表达能力。CityPersons 与 TinyPerson 和 Tiny CityPersons 具有类似的 FPN 分层。尽管通过 CityPersons 的 4 倍下采样获得了 Tiny CityPersons（标注框和图片的大小缩小 4 倍），但由于 Tiny CityPersons 的锚点框（anchor）的大小也减少了 4 倍，因此在 anchor 与标注框的匹配机制中，标注框的变小与 anchor 大小的变小产生影响相互抵消，FPN 中 CityPersons 的分层仍然与 Tiny CityPersons 相似。由于小目标数据集中的弱小目标的数量占主导地位，大量的弱小目标集中在 FPN 的 P_2 和 P_3 层中，而导致 FPN 深层中的学习样本不足。但是，融合因子在 CityPersons 上的性能趋势不同于 TinyPerson 和 Tiny CityPersons。

目标（标注框）的绝对大小是影响融合因子的有效性的关键因素，其他三个因素不是融合因子作用的前提。以下给出融合因子为何起作用以及如何起作用的分析：融合因子通过在梯度反向传播中重新对加权来自不同层梯度来确定 FPN 中的深层参与浅层的学习的程度。当数据集中的目标是弱小目标级别的，FPN 中每层的检测任务的学习变得困难。因此，每层的学习能力都是不够的，深层没有额外的能力来参与到浅层的检测任务学习。当每层的学习难度增加和降低 α 的同时，FPN 中深层和浅层之间学习能力的供求关系发生了变化，这表明每层都会更

加专注于本层检测任务的学习。

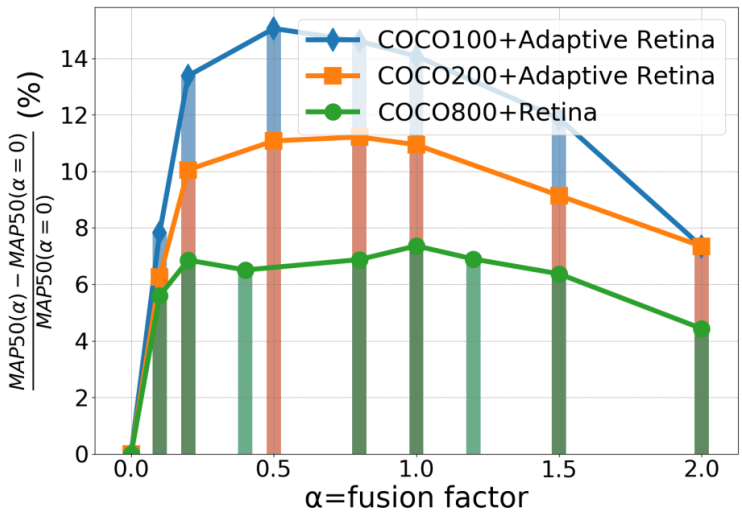

图 3.3 融合因子的变化对不同尺度的 COCO 的性能影响

Figure 3.3 The influence of α on the performance of COCO with different scales

3.2 融合因子实现方式

在上一节得到了融合因子对弱小目标检测有效的前提，数据集为弱小尺度的数据集，即数据集中目标的平均大小应小于 20 像素，但是在以上实验中，融合因子的设定方式为人为固定设置，即在网络训练开始前已经将三个融合因子固定为同一数值。

这种设置方式有两个问题：1.三个融合因子本身所处位置不同，控制的参与融合的特征层也不同，这样统一设置成固定值的方式可能不是最优，三个融合因子可能分别存在最佳取值区间；2.所有的融合因子在网络训练前已经依据经验值固定，并不会随着选择的数据集或网络的参数变化而变化，缺乏动态性。

下文将给出不同的融合因子设置方式对以上两个问题进行分析讨论，采用不同的方法进行融合因子设置的实验。通过检测精度和实用性两方面的对比给出最有效的融合因子设置方法。

3.2.1 基于可学习参数的融合因子

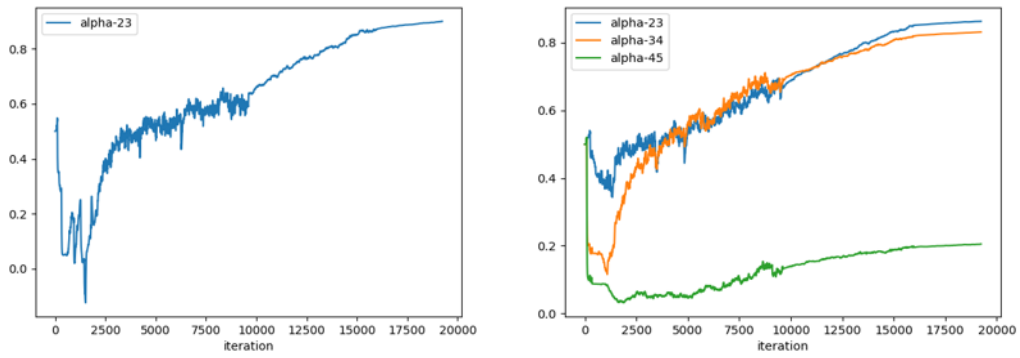

图 3.4 单可学习融合因子（左）和三个可学习融合因子（右）收敛情况

Figure 3.4 Convergence of single learnable α (left) and three learnable α (right)

因为融合因子对网络学习能力有影响，首先进行了添加可学习参数的融合因子的实验，在不外加融合因子监督信息的情况下，探究可学习参数能否在自主学习到融合因子的最合适的值区间。

添加可学习融合因子实验可以分为两类：1.三个融合因子共用一个可学习参数；2.三个融合因子分别采用三个不同的可学习参数。实验的具体设置为 Adaptive RetinaNet 在 TinyPerson 数据集上，训练 12 个周期（19236 的迭代）。两种设置下的融合因子都随着训练迭代的进行，数值逐渐收敛，在单参数情况下，融合因子的收敛值为 0.899，如图 3.4 左图，AP_{50}^{tiny} 最高性能为 46.86%；在三个参数情况下，分别收敛到α_2^3-0.863、α_3^4-0.832、α_4^5-0.205，如图 3.4 右图，AP_{50}^{tiny} 最高性能为 47.66%。研究发现，虽然可学习的参数能带来一定程度的性能改善，但并没有得到与最佳固定融合因子相同的结果，说明网络自主学习无法获得融合因子的最佳取值范围，并且三个不同的可学习参数性能好于单个可学习参数说明三个融合因子有各自最合适的值区间。

3.2.2 基于注意力机制的融合因子

基于 query，value，gallery 机制实现的注意力（Attention）方法在计算机视觉和自然语言处理等领域都取得了不错的性能提升，借鉴以往的 Attention 方法，本节设计了一种基于 Attention 机制产生融合因子的方法。

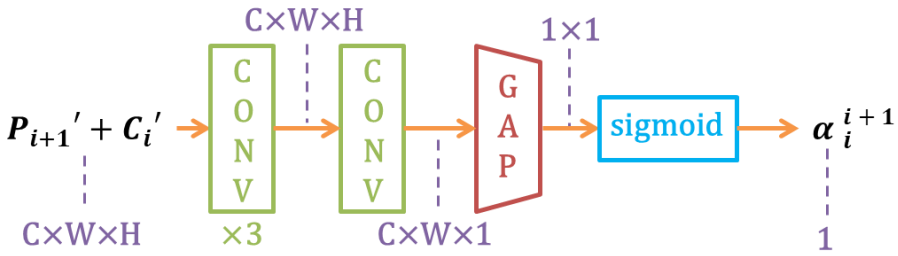

图 3.5　基于注意力机制的融合因子示意图

Figure 3.5　Schematic diagram of fusion factors based on attention mechanism

如图 3.5 所示，将P'_{i+1}（P'_{i+1}为高层特征经过上采样后形状与C'_i相同的特征）与C'_i（C'_i为将从 backbone 中直接取得的特征C_i经过 FPN 中卷积核大小为 1×1 的f_{inner}卷积处理后通道数为 256 的特征层）进行特征图的相加操作，此时得到的特征图的形状可以表示为 C×W×H，C 为通道数为 256，W 和 H 为特征图的长和宽，为了简化暂时不考虑批的（batch）维度。之后会经过两个 3×3 的f_{layer}卷积计算，得到形状为 1×W×H 的特征，后经过 GAP 计算（全局平均池化，Global Average Pooling）得到一组 1×1 的参数，经过 sigmoid 计算后的值作为对应位置上的融合因子。

3.2.3　基于监督信息的融合因子

在可学习融合因子设置方式之外还进行了设定有监督信息的可学习融合因子设置方式的实验。如图 3.6 所示，在融合的时候加入了 FRM（Fuse Ration Modular）模块，预测深层和浅层融合因子 α。形状为（b，c，w，h）的特征F_{deep}和形状为（b，c，w，h）的特征$F_{shallow}$作为 FRM 模块的输入，b 代表批大小，c 代表通道数，w 和 h 分别代表特征图的宽和高，经过 4 个 3×3 卷积模块后，F_{high}的形状将变为（b，1，w，h），再经过一个 Global Average Pooling 后F_{high}的形状将变为（b，1，1，1），过一个 sigmoid 激活函数后得到融合因子 α，这样就可以根据当前的特征图来预测对应的融合因子，来自适应的改变融合过程，得到更适合做弱小目标检测的特征图。预测值的监督信息来源于固定值的设置。

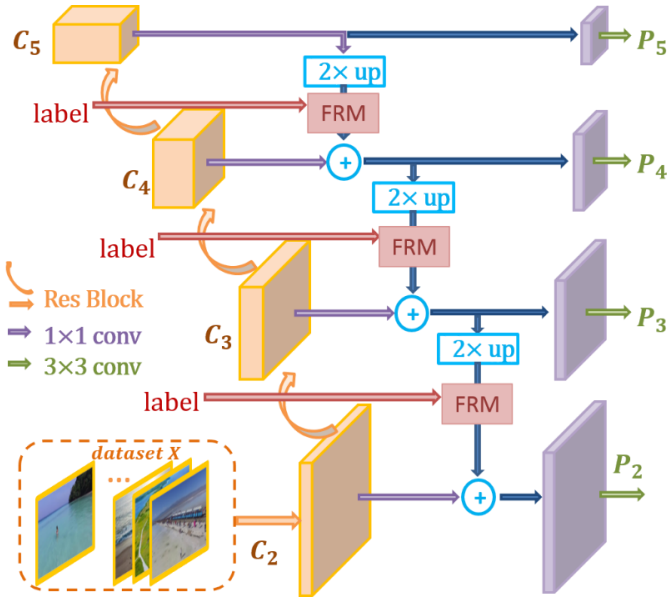

图 3.6　有监督的可学习融合因子示意图

Figure 3.6　Schematic diagram of supervised learnable fusion factors

3.2.4　基于统计的融合因子

这节介绍基于统计的融合因子设置方法，基于统计的方法称为 S-α，根据 FPN 中相邻层之间的目标数比例设置 α，如下图中的标红公式所示，计算结果是整个数据集中统计的结果。本研究基于以下事实设计数学公式：对于弱小目标检测，由于整个数据集都是弱小目标，导致 FPN 中的各层都在学习小目标或者弱小目标的分类和回归，而且由于弱小目标检测任务本身的难度，导致 FPN 的每一层都难以学习具有代表性的检测任务特征，从而加剧了层之间的竞争。FPN 中的每一层都希望它们的参数可以学习到提起其相应检测任务的合适特征的能力。然而某些层可能比其他层具有更少的训练样本，从而导致在更新共享参数时，这些层的梯度与其他特征层产生的梯度相比不处于主导地位。因此，当$N_{p_{i+1}}$比较小或者N_{p_i}比较大时（N_{p_i}为基于统计的融合因子的被统计层划分为正例的数量，统计规则细节见下文），该方法设置一个小的 α 以减小由P_i层中的检测任务产生的梯度对P_{i+1}产生的影响，反之亦然，这会促使网络每一层中的检测任务都能平衡地学习。因此，弱小目标检测任务学习效率得以提升。方法如下图：

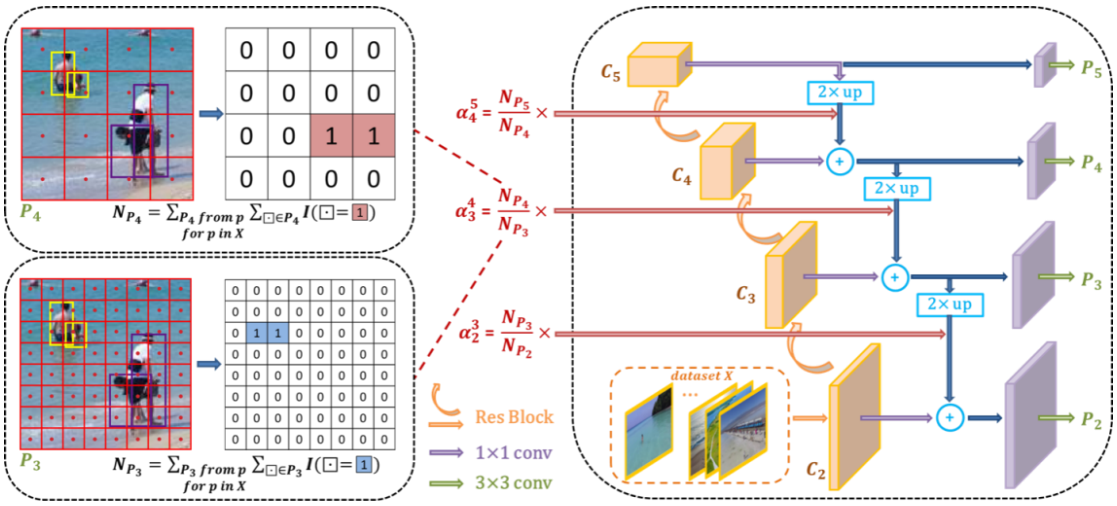

图 3.7　基于统计的融合因子的计算过程图

Figure 3.7　Calculation process chart of fusion factor based on statistics

上图是方法框架示意图，图中的示例图片取自 TinyPerson。左侧的虚线框显示N_{p_i}的统计过程，中间计算公式如下，表示融合因子的计算过程，右侧虚线框内为 FPN 的结构示意图以及融合因子作用位置。

$$\alpha_i^{i+1} = \frac{N_{p_{i+1}}}{N_{p_i}} \tag{3.6}$$

在左侧虚线框内矩阵中的 1 和 0 分别代表匹配到 ground truth（标注框）被划分为正例的 anchor（锚点框）和没有匹配到标注框被划分为负例的 anchor。红色框和红色点表示预先设定好的 anchor 和锚框点（anchor point）。为了简化表示，一个 anchor point 仅显示一个 anchor。黄色框和蓝色框均代表 ground truth，不同的是黄色框代表匹配到P_3层的 anchor 的 ground truth 和匹配到P_4层的 anchor 的 ground truth。通过基于统计的方法获得有效的融合因子 α。

每层目标数N_{p_i}的统计过程中的匹配原则概括如下：1）以交并比（IoU）匹配为原则，在图像中选择与 ground truth 最大的 IOU 的 anchor。2）基于划分为正例 anchor 以及预先设定的 anchor 结构的情况，FPN 每层匹配到的 ground truth 数可以被计算出来。之后对数据集中的每个图像重复步骤 1 和步骤 2，以获得整

个数据集上累加的统计结果，然后根据如图 3.6 左虚线框所示的等式计算 α。由于 anchor 是预先定义的，并且数据集提供了 ground truth，因此计算过程不涉及网络的前向计算。整个过程表示为详细算法如下：

表 3.1　统计算法步骤

Table 3.1　Statistical algorithm steps

输入：M（M为所有图片的 ground truth 与 anchor 的 IOU 匹配结果集合，M_i表示第 i 张图片的匹配结果），A（预先定义好的 anchor 集合，A_i 表示设定在 FPN 的第 i 层上面的 anchor），$List_{tn}$（$List_{tn}$表示一个记录 FPN 每层按照匹配原则的匹配结果的数组，$[N_{p_2}, N_{p_3}, N_{p_4}, N_{p_5}, N_{p_6}]$）。
输出：$List_{\alpha}$（融合因子的计算结果，α_2^3、α_3^4、α_4^5）。
步骤（1）：将$List_{tn}$初始化为空集[0，0，0，0，0]；
步骤（2）：把M中的M_i依据 ground truth 与 anchor 的匹配原则（一个 ground truth 最后只与一个 anchor 匹配），得到匹配结果 R，R_i表示第 i 张图的匹配结果；
步骤（3）：对于 R 中每一个R_i，按照 A 中每一个A_i可以对R_i进行分层统计匹配数目，累加到$List_{tn}$。$List_{tn}$在 Tinyperson 和 Tiny Cityperson 的计算结果如表 3.1 所指示；
步骤（4）：依照$List_{tn}$计算 $List_{\alpha}$，其中$\alpha_2^3 = N_{p_3}/N_{p_2}$，$\alpha_3^4 = N_{p_4}/N_{p_3}$，$\alpha_4^5 = (N_{p_5} + N_{p_6})/N_{p_4}$。

表 3.2　算法步骤 3 统计结果

Table 3.2　Statistical results of algorithm step 3

数据集	N_{p_2}	N_{p_3}	N_{p_4}	N_{p_5}	N_{p_6}
TinyPerson	20733	7638	2958	919	166
Tiny Citypersons	6156	2577	1460	335	9

3.3 特征补充融合

在 FPN 用来做弱小目标检测任务中，用来做分类预测和回归预测的五层分别是P_2，P_3，P_4，P_5，P_6，如图 1.3，分别对应骨干网络（backbone）的残差网络（ResNet）的第 2 个阶段（stage）特征图C_2，第 3 个 stage 特征图C_3，第 4 个 stage 特征图C_4，第 5 个 stage 特征图C_5，其中P_6为P_5下采样得到，通过实验发现绝大数的弱小尺度的目标在P_2得到，尺度在 20 像素以下的弱小目标在 4 倍降采样的P_2上会小于 5 个像素，等价于小于 25 个像素点区域，实验表明上下文信息在弱小目标检测中并不一定能帮助提升检测器检测精度，所以目标本身的细节特征对弱小目标检测更为关键，但是由于经过特征融合后高层的抽象特征会弱化原本P_2的细节特征，检测难度还是比较大的，所以需要检测层包含更多的细节特征，但是由于 GPU 和内存和计算量的限制，已经不能将 ResNet 的第一个 Stage 特征图C_1构建 FPN。因此，在原本的 FPN 中，骨干网络中最浅层的特征C_1没有直接参与到 FPN 的构建导致特征的细节信息在一定程度上的缺失。这时候可以采取特征补充的方式来提升对应检测层的细节特征，如图 3.7，使其更适合做弱小尺度的目标检测。

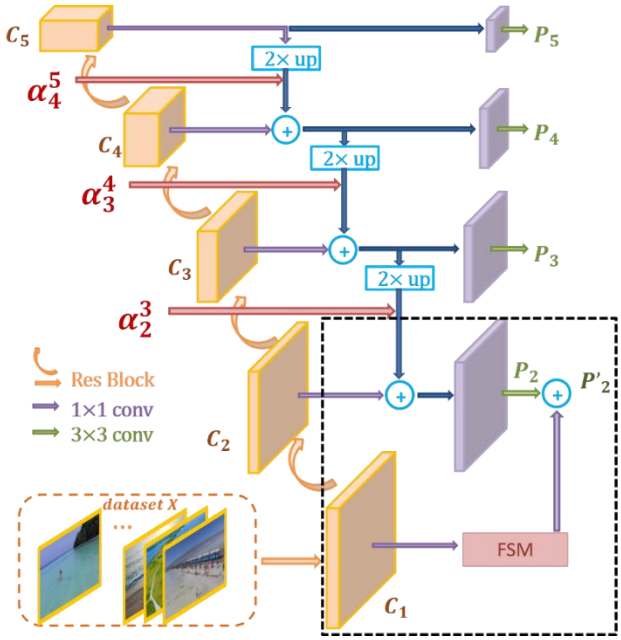

图 3.8 特征补充模块作用位置图

Figure 3.8 Function location map of feature supplement module

如图 3.7 所示，FSM（Feature Supply Modular）作用对象为C_1与P_2，作用区域为图中黑色虚线框部分，其包含 2 层不同的卷积，为（卷积核大小 3×3 的卷积，空洞率为 1），（卷积核大小 3×3 的卷积，步长为 2），得到的特征来自不同的感受野，在一定区域内的特征来源于不同感受野会让特征的上下文信息更加丰富。将C_1经过 FSM 得到与P_2相同形状的特征C_1'，然后将C_1'与P_2按照公式（3.1）进行融合，得到新的P_2'，P_2'与P_2的形状相同，但是相较于原本的P_2，P_2'在融合过程中加进了来这C_1的低层特征的信息，这些特征包含着大量细节特征和纹理信息，与P_2融合后，P_2'就含有更多的细节特征，这些特征将帮助弱小尺度的目标检测，同时可以利用融合因子的作用来控制融合比例。

3.4 本章小结

本章主要介绍了研究内容与方法，研究内容方面介绍了特征金字塔的作用机理以及弱小目标检测对其的影响，还包括从数据集的角度介绍融合因子作用的前提，并结合实验进行分析，说明对融合因子的研究是弱小目标检测的关键点。研究方法方面介绍了与人为固定融合因子不同的另外四种融合因子设置方式，并对思想原理和算法流程进行了详细的介绍。下一章将介绍不同方法设置的融合因子的性能以及选定的基于统计的融合因子方法在各种条件下的性能，证明方法的泛化能力。

第 4 章　实验结果及融合因子原理分析

4.1　评测指标选择

本研究主要使用类别平均精度（Mean Average Precision - MAP）和丢失率（Miss Rate - MR）进行评测。MAP 是在各种检测任务中广泛使用的指标，反映了 Recall（查全率）和 Precision（准确率）在类别上平均后的检测结果的指标。本文主要实验是在是在弱小目标检测的代表数据集 TinyPerson，同时因为 TinyPerson 是行人数据集，因此 MR 也用作评测标准。 并且将评测选取的交并比（IoU）的阈值设置为 0.25、0.5 和 0.75。在 Tinybenmark 进一步将 tiny [2，20)（像素）分为 3 个子区间：tiny1 [2，8)，tiny2 [8，12)，tiny3 [12，20)。本研究更多地关注是否可以找到对象，而不是位置精度。因此，我们选择 IoU = 0.5 作为评测的主要阈值。

首先介绍 Recall 和 Precision 的计算过程。在训练集上学习到检测模型之后，测试集上的每一张图片经过网络的前向计算后都会得到本张图的检测结果集合 D，D 中每个检测样本外接矩形框 $d_i(x, y, w, h, c, score)$，$i$ 代表 d 在 D 中的位序，$x, y, w, h, c, score$ 具体含义分别是代表矩形框的中心坐标的 x，y，矩形框宽和高 w，h，还有预测的类别 c，以及得分置信度 $score$。根据 x, y, w, h 可以计算出当前样本与图中标注框（ground truth）的 IoU。IoU 的计算如下：

$$\mathrm{IoU} = \frac{\text{预测框区域} \cap \text{标定框区域}}{\text{预测框区域} \cup \text{标定框区域}} \tag{4.1}$$

对于 Citypersons 来说，可以采用 IoU 标准进行绩效评估，因为 Citypersons 大多数忽略区域的大小与行人的大小相同。然而，在 TinyPerson 数据集中大多数忽略区域比人的大得多。因此，对于忽略区域，我们将 IoU 标准更改为 IoD（IoD 标准仅适用于忽略区域，非忽略区域仍然使用 IoU 标准）。

$$\mathrm{IoD} = \frac{\text{忽略区域} \cap \text{标定框区域}}{\text{标定框区域}} \tag{4.2}$$

根据样本的得分 $score$ 依据预先设定的得分阈值可以判断样本是正例还是反

例，对所有样本的正反例评判有如下四种情况：1. True Positive（TP）：正确的正例，正例样本被检测模型正确的判定为正例的样本；2. False Positive（FP）：错误的正例，反例样本被检测模型错误的判定为正例的样本；3.True Negative（TN）：正确的反例，反例样本被检测模型正确的判定为反例的样本；4.False Negative（FN）：错误的反例，正例样本被检测模型错误的判定为反例的样本。基于上述四个定义，Recall 和 Precision 可以用如下公式计算：

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.4}$$

在给定 IoU 阈值以及确定类别情况下，对所有同一类测试样本的得分降序排序，得分越高的排序越靠前。接下来划定一个得分与之，依照降序排序遍历所有的检测结果，每个检测结果和标注框（ground truth）都会计算出一个 IoU,如果大于给定 IoU 阈值则计为匹配成功（TP），并且对应的 ground truth 不会参与到后续匹配过程，如果没有超过阈值则为 FP，重复匹配过程，最后小于给定得分阈值但是大于给定 IoU 阈值的检测结果则为 FN。给定一个得分阈值就可以得到一个 Recall 和一个 Precision，通过不断的调整正反例样本的得分阈值形成不同的 Recall 和 Precision，最后会得到一条 PR 曲线。对 PR 曲线有两种计算方法：1. 将横坐标的 Recall 采样 11 个等分点，然后将 11 个等分点对应的 Precision 计算平均值得到 MAP；2.对 PR 曲线下的面积求积分得到 MAP。第二种方法求出来的 MAP 更为精确，但是考虑到不同数据集的数量差异对计算结果的影响，一般采用第一种方法，即通过采样 11 个等间距点的方法可以对 PR 曲线下求面积得到 AP，对于每个类别都会计算得到一个 AP，对于所有的类别的 AP 计算平均值就可以得到我们采用的评测指标 MAP。

作为另一种性能指标，我们还可以采用与 Recall 相对应的 MR（Miss Rate，丢失率）进行评测，MR 指的是图中没有检测出正例的标注框占总体正例目标标注框的比例。

$$\text{MR} = \frac{FN}{TP + FN} \tag{4.5}$$

MR 在具体计算时依赖于 MR- FPPI（Miss rate against false positives per

image），FPPI 为单张中 FP 的数量，对于每一个阈值情况下的 FP 都会有对应的 TP 和 FN,因此就有对应的 MR，一个 MR 就会对应一个 FPPI 值，一般来说，分数阈值越高，FPPI 会越低 MR 就会越高，设置不同的得分数阈值也可以得到一组 MR- FPPI 值，从而得到 MR- FPPI 曲线。在比较不同的算法或评测算法性能时一般将 FPPI 设定为一个较有意义的值如 1，10，100 来得到对应得 MR。

4.2 实验设置

代码基于 Tinybenmark。如果没有特殊声明，网络初始权重选择 ImageNet 预训练的骨干网络权重。检测实验主要用的数据集为 TinyPerson，TinyPerson 中目标的平均绝对大小是 18 个像素，但是 TinyPerson 中目标的长宽比变化很大。而且考虑到光照和角度的因素，样本的多样性也更加复杂，使得检测任务更加困难。TinyPerson 的训练集和测试集分别包含 794 张和 816 张图像。TinyPerson 中的大多数图像的尺寸都很大。因为 GPU 显存的因素，图片不适合直接作为网络的输入，数据统计如表 4.1，表中对比了切割图片处理后的 Tinyperson（TinyPerson-Cut）和 CityPersons，在数据量上 TinyPerson-Cut 大于 Cityperson，。因此，在训练和测试过程中，原始图像被分割成重叠的子图像作为网络的输入。由于 TinyPerson 中的一些图像中有 200 多个密集的目标，只选择了 200 个以下的目标图像进行训练和测试。在数据增强方面，实验中只采用了水平翻转。

表 4.1 TinyPerson-Cut 与 CityPersons 对比

Table 4.1 Comparison of TinyPerson-Cut and CityPersons

	TinyPerson-Cut	CityPersons
图片数量	25949	3457
标注数量	72651	19683
图片总面积	14853692	5032337

实验部分选择了不同的检测器做对比实验，在一阶段算法中选择 RetinaNet 作为代表，二阶段算法中选择了 Faster RCNN-FPN 作为代表。表 4.2 和表 4.3 罗列了两种算法部分设置的细节。

表 4.2　RetinaNet 实验设置

Table 4.2　RetinaNet experimental setting

网络参数	值
RETINANET_ON	True
BACKBONE	R-50-FPN
RETINA.ANCHOR_SIZES	(8, 16, 32, 64, 128)
RETINA.ASPECT_RATIOS	(0.5, 1., 2)
BASE_LR	0.005
MAX_ITER	19236
STEPS	(9618, 16030)
IMS_PER_BATCH	2
NUM_GPU	1
TEST_ITER	1603

表 4.3　Faster RCNN-FPN 实验设置

Table 4.3　Faster RCNN-FPN experimental setting

网络参数	值
RETINANET_ON	False
BACKBONE	R-50-FPN
RPN.ANCHOR_SIZES	(8.31, 12.5, 18.55, 30.23, 60.41)
RPN.ASPECT_RATIOS	(0.5, 1.3, 2)
BASE_LR	0.01
MAX_ITER	19236
STEPS	(9618, 16030)
IMS_PER_BATCH	2
NUM_GPU	1
TEST_ITER	1603

4.3 实验结果

本节介绍了不同方法在不同数据集上的实验结果，基准实验的训练共有 12 周期，共 19236 次迭代，每 1603 次迭代测试一次网络性能，取在一次训练中单次测试中性能最高的结果作为整个训练的结果。

4.3.1 不同实现方式结果及分析

表 4.4 不同 α 的实现方式性能比较

Table 4.4 Performance comparison of different α implementations

方法	AP_{50}^{tiny}	MR_{50}^{tiny}
baseline	46.56	88.31
one-α	46.86	88.31
three-α	47.66	87.98
atten-α	47.88	87.80
sup-α	47.89	87.66
bf-α	48.33	87.94
S-α	48.34	87.73

上表是，融合因子的不同实现方法在 TinyPerson 上的性能比较。 基准实验中的 α 默认设置为 1。one-α 和 three-α 分别表示使用一个和三个可学习的参数（one-α 代表将三个融合因子α_2^3、α_3^4、α_4^5设置为同一个可学习参数, three-α 将三个融合因子设置成三个不同的可学习参数）。atten-α 代表基于 attention 的实现方式。α-bf 表示通过暴力解的最优值，sup-α 代表有监督的融合因子，S-α 代表基于统计的融合因子，上表中的性能都是以 RetinaNet 作为算法框架获得的。较低的 MR（丢失率）意味着更好的性能。

通过对上表的分析得到以下结论：第一，暴力搜索找到了最佳 α。但是，它包含冗余计算，而且三个融合因子的最优值并不相同，如果采用暴力搜索的方法，这是指数级的参数搜索空间，而且用于一个新数据的时候这个方法应用成本太大，这限制了该方法的大规模应用。第二，所有非固定的 α 设置都优于基线性能（baseline），其中 α 设置为 1，可学习参数无监督的方法性能低于最佳性能。在

另外的实验中发现，可学习参数有监督的方法性能近似于基于最佳性能，但和注注意力的方法一样都增加了不可忽略的计算量。第三，只有基于统计的方法才能获得与暴力搜索可比的性能。

4.3.2 S-α 实验结果

4.3.2.1 S-α 在 TinyPersons 上的结果

表 4.5 S-α 在 TinyPerson 上结果

Table 4.5 The results of S-α on TinyPerson

检测算法	Backbone	AP_{50}^{tiny}	MR_{50}^{tiny}
RetinaNet	ResNet-50	46.56	88.31
Faster RCNN	ResNet-50	47.34	87.57
RetinaNet with S-α	ResNet-50	48.34	87.73
Faster RCNN with S-α	ResNet-50	48.39	87.29

综合考虑下选择了基于统计的融合因子设置方法，简称 S-α，上表是在用 Adaptive-RetinaNet(后续实验结果如果没有特别说明，表中的 RetinaNet 均代表 Adaptive-RetinaNet)和 Faster RCNN（如果没有特殊说明，就为 Faster RCNN-FPN 的简化表示）的基准实验和使用了 S-α 方法的对比，数据集为 TinyPerson，RetinaNet 原本使用从骨干网络比如 ResNet-50 中提取出来的C_3，C_4，C_5作为 FPN 的输入，然后由下采样计算得到P_6，P_7，由线性及卷积计算得到P_5，P_4，P_3。而 Adaptive-Retina 则使用 ResNet-50 中第 2 个 stage 到第 5 个 stage 即C_2-C_5作为骨干网络的输出，然后只有P_6是由下采样得到的，其他 FPN 层均是由卷积和线性计算组合得到，相当于将原本 RetinaNet 的骨干网络整体前移，这样做的好处是在于可以更利于小目标检测，因为小目标检测更多是在 FPN 浅层，浅层富含更多的细节信息，目标匹配过程大多也是在浅层完成的。可以发现在 RetinaNet 上 AP_{50}^{tiny} 取得了 1.78%的提升，MR_{50}^{tiny} 取得了 0.58%的提升，在 Faster RCNN 上 AP_{50}^{tiny} 取得了 1.05%的提升，MR_{50}^{tiny} 取得了 0.44%的提升。说明了方法在不同检测器上的有效性。

表 4.6 S-α 在不同骨干网络的结果

Table 4.6 The results of S-α in different backbones

检测算法	Backbone	AP_{50}^{tiny}	MR_{50}^{tiny}
RetinaNet	ResNet-50	46.56	88.31
RetinaNet	ResNet-101	46.99	88.16
RetinaNet with S-α	ResNet-50	48.34	87.73
RetinaNet with S-α	ResNet-101	47.99	87.81

上表是基于统计的融合因子在不同骨干网络条件下得到的试验结果，试验数据集为 TinyPerson，检测器为 Adaptive-Retina。

表中性能表明：基于统计的融合因子设置方法在不同骨干网络的条件下下均取得了性能上的提升，在 ResNet-50，AP_{50}^{tiny} 取得了 1.78%的提升，MR_{50}^{tiny} 取得了 0.58%的提升（MR 为丢失率,越低代表取得了越好的性能），在 ResNet-101，AP_{50}^{tiny} 取得了 1.00%的提升，MR_{50}^{tiny} 取得了 0.35%的提升。基于统计的融合因子在 ResNet-50 上的提升大于在 ResNet-101 上提升原因可能在于，虽然一般认为更深的网络应该获得更大的性能提升，但是对于弱小目标检测这个任务而言，目标的分类和回归都是在 FPN 的浅层特征层，ResNet-101 相较于 ResNet-50 多出的 51 层卷积实际上位于骨干网络 ResNet 的第 4 个阶段，这些计算量对 FPN 的浅层细节特征影响很小，在网络深层增加的计算并不能显著来帮助弱小目标的性能提升。

4.3.2.2 S-α 与其他检测算法对比

表 4.7 不同方法在 TinyPerson 上的 AP 性能

Table 4.7 AP performance of different methods on TinyPerson

检测算法	AP_{50}^{tiny}	AP_{50}^{small}	AP_{25}^{tiny}	AP_{75}^{tiny}
FCOS[53]	17.90	40.54	41.95	1.50
RetinaNet*[27]	33.53	48.26	61.51	2.28

续表：

检测算法	AP_{50}^{tiny}	AP_{50}^{small}	AP_{25}^{tiny}	AP_{75}^{tiny}
FreeAnchor[54]	41.41	59.61	63.38	4.58
Libra RCNN[24]	44.68	62.65	64.77	6.26
RetinaNet[27]	46.56	59.97	69.6	4.49
Grid RCNN[23]	47.14	62.48	68.89	6.38
Faster RCNN-FPN[13]	47.35	63.18	68.43	5.83
RetinaNet-SM	48.48	63.01	69.41	5.83
RetinaNet-MSM	49.56	63.38	71.24	6.16
Faster RCNN-FPN-SM	51.33	66.96	71.55	6.46
Faster RCNN-FPN-MSM	50.89	65.76	71.28	6.66
RetinaNet with S-α	48.34	61.73	71.18	5.34
Faster RCNN-FPN with S-α	48.39	65.15	69.32	5.78
RetinaNet-SM with S-α	52.56	65.69	73.09	6.64
RetinaNet-MSM with S-α	51.60	64.39	72.60	6.43
Faster RCNN-FPN-SM with S-α	51.76	66.81	72.19	6.81
Faster RCNN-FPN-MSM with S-α	51.41	65.97	72.25	6.69

表 4.8　不同方法在 TinyPerson 上的弱小尺度的 AP 性能

Table 4.8　Tiny scale AP performance of different methods on TinyPerson

检测算法	AP_{50}^{tiny1}	AP_{50}^{tiny2}	AP_{50}^{tiny3}
RetinaNet[27]	27.08	52.63	57.88
Faster RCNN-FPN[13]	30.25	51.58	58.95
RetinaNet-SM	29.01	54.28	59.95
RetinaNet-MSM	31.63	56.01	60.78
Faster RCNN-FPN-SM	33.91	55.16	62.58
Faster RCNN-FPN-MSM	33.79	55.55	61.29

续表：

检测算法	AP_{50}^{tiny1}	AP_{50}^{tiny2}	AP_{50}^{tiny3}
RetinaNet with S-α	28.61	54.59	59.38
Faster RCNN-FPN with S-α	31.68	52.20	60.01
RetinaNet-SM with S-α	33.90	58.00	63.72
RetinaNet-MSM with S-α	33.21	56.88	62.86
Faster RCNN-FPN-SM with S-α	34.58	55.93	62.31
Faster RCNN-FPN-MSM with S-α	34.64	55.73	61.95

表 4.9　不同方法在 TinyPerson 上的 MR 性能

Table 4.9　MR performance of different methods on TinyPerson

检测算法	MR_{50}^{tiny}	MR_{50}^{small}	MR_{25}^{tiny}	MR_{75}^{tiny}
FCOS[53]	96.28	84.16	90.34	99.56
RetinaNet*[27]	92.66	82.84	81.95	99.13
FreeAnchor[54]	89.63	74.38	78.21	98.77
Libra RCNN[24]	89.22	74.86	82.44	98.39
RetinaNet[27]	88.31	74.05	76.33	98.76
Grid RCNN[23]	87.96	73.16	78.27	98.21
Faster RCNN-FPN[13]	87.57	72.56	76.59	98.39
RetinaNet-SM	88.87	71.82	77.88	98.57
RetinaNet-MSM	88.39	72.18	76.25	98.57
Faster RCNN-FPN-SM	86.22	68.59	74.16	98.28
Faster RCNN-FPN-MSM	85.86	68.76	74.33	98.23
RetinaNet with S-α	87.73	72.82	74.85	98.57
Faster RCNN-FPN with S-α	87.29	70.75	76.58	98.42
RetinaNet-SM with S-α	87.00	69.25	74.72	98.41
RetinaNet-MSM with S-α	87.07	70.35	75.38	98.41

续表：

检测算法	MR_{50}^{tiny}	MR_{50}^{small}	MR_{25}^{tiny}	MR_{75}^{tiny}
Faster RCNN-FPN-SM with S-α	85.96	69.35	73.92	98.30
Faster RCNN-FPN-MSM with S-α	86.18	69.28	73.90	98.24

表 4.10　不同方法在 TinyPerson 上弱小尺度的 MR 性能

Table 4.10　Tiny scale MR performance of different methods on TinyPerson

检测算法	MR_{50}^{tiny1}	MR_{50}^{tiny2}	MR_{50}^{tiny3}
RetinaNet[27]	89.65	81.03	81.08
Faster RCNN-FPN[13]	87.86	82.02	78.78
RetinaNet-SM	89.83	81.19	80.89
RetinaNet-MSM	87.8	79.23	79.77
Faster RCNN-FPN-SM	87.14	79.60	76.14
Faster RCNN-FPN-MSM	86.54	79.2	76.86
RetinaNet with S-α	89.51	81.11	79.49
Faster RCNN-FPN with S-α	87.69	81.76	78.57
RetinaNet-SM with S-α	87.62	79.47	77.39
RetinaNet-MSM with S-α	88.34	79.76	77.76
Faster RCNN-FPN-SM with S-α	86.57	79.14	77.22
Faster RCNN-FPN-MSM with S-α	86.51	79.05	77.08

在 TinyPerson，分别对比了其他最先进的检测算法和添加 S-α 的性能。由于目标的尺度极小（目标平均大小为 18 像素），在通用目标检测上表现得很好的检测算法的性能显着下降，如表 4.7 至表 4.10 展示了部分算法在弱小尺度上性能（FreeAnchor 使用P_2，P_3，P_4，P_5，P_6构建 FPN 并将 anchor 大小调整为[8、16、32、64、128]， RetinaNet*是初始本版本。）。由于在 TinyPerson 上，正例样本数量和负例样本数量的比例不平衡问题变得更加严重，对这个问题处理比较好的两

级检测器的性能优于一级检测器。使 Faster RCNN-FPN with S-α 在无需增加网络参数的情况下，分别将 AP_{50}^{tiny} 和 MR_{50}^{tiny} 的性能提高了 1.04％和 0.28％。实验结果表明：对 FPN 的修改对两级检测器和一阶段检测器都是有益的。另外的实验证明 FreeAnchor 在融合因子为固定设置 0.5 条件下，在 AP_{50}^{tiny} 提高了 0.92％。

RetinaNet with S-α 的性能优于带有 SM / MSM 以外的其他检测器。 SM / MSM 需要通过尺度匹配策略（Scale Matching）算法/单调尺度匹配策略（Monotonous Scale Matching）算法将用于网络预训练 COCO 和 TinyPerson 之间进行尺度匹配，然后在 TinyPerson 上进行微调，预训练需要比较大的计算和时间资源。而 RetinaNet with S-α 仅使用已经公开的 ImageNet 上的预训练模型，只进行一次微调，训练代价大大降低。如表 4.7 至表 4.10 所示，RetinaNet with S-α 无需添加新的网络参数即可达到近似的性能。SM / MSM（SOTA 方法）与 S-α 同时使用可以使性能获得更大增益，说明 SM / MSM 和 S-α 两个方法是互补的，SM / MSM 方法从预训练数据的角度为网络提供了更好的初始化权重，S-α 针对目标数据集的特点对网络结构做出了适应性改变，两个方法从网络训练的不同阶段改进提升了小目标检测的性能。RetinaNet with S-α+ SM 实现了新的最高性能，并且与 RetinaNet + SM 相比，在 AP_{50}^{tiny} 和 MR_{50}^{tiny} 分别提高了 4.08％和 1.87％。

4.3.2.3 S-α 在 Tiny Citypersons 和 Tiny COCO 上的结果

表 4.11 S-α 在 Tiny CityPerson 上结果

Table 4.11 The results of S-α on Tiny CityPerson

检测算法	AP_{50}^{tiny}	MR_{50}^{tiny}
RetinaNet	36.36	78.03
RetinaNet with bf-α	38.94	75.91
RetinaNet with S-α	38.60	76.45

上表是 RetinaNet 在另一个弱小目标数据集 Tiny Cityperson 上的性能结果，bf-α 为暴力设置融合因子取得的最佳性能，可以发现的是在 Tiny Cityperson 上性能依旧取得了提升，AP_{50}^{tiny} 取得了 2.24%的提升，MR_{50}^{tiny} 取得了 1.58%的提升。

表 4.12　S-α 在 Tiny COCO 上结果

Table 4.12　The results of S-α on Tiny COCO

检测算法	AP	AP_{50}^{all}
RetinaNet	14.60	27.96
RetinaNet with bf-α	14.68	28.09
RetinaNet with S-α	14.86	28.27

上表是 RetinaNet 在另一个弱小目标数据集 Tiny COCO（将 COCO 的图片输入到网络的时候，短边统一缩小到 100 像素）上的性能结果，S-α 为暴力设置融合因子取得的最佳性能，可以发现的是在 Tiny COCO 上性能依旧取得了提升，AP_{50}^{all} 取得了 0.31%的提升，AP 取得了 0.26%的提升。基于统计的融合因子在不同的弱小目标检测数据集上都取得了性能提升，说明了算法的泛化能力，也适用了其他弱小目标检测数据集。

4.3.3　特征补充实验结果

表 4.13　特征补充模块的实验结果

Table 4.13　Experimental results of feature supplement module

实验条件	AP_{50}^{tiny}
RetinaNet	46.56
RetinaNet+FSM with No GN,α=1	46.63
RetinaNet+FSM with GN,α=1	47.09
RetinaNet+FSM with GN,α=0.5	47.39
RetinaNet+FSM with GN,α=0.25	46.33

上表是特征补充模块在 RetinaNet 为基本框架在 TinyPerson 上的实验结果，FSM 为特征补充模块（Feature Supplement Module）全称的缩写，with GN 代表卷积带有组归一化（Group Normalization），α 代表在特征补充模块中的融合因子，后边的数值为实验中具体融合因子的取值。

通过比对结果可以发现，如果单纯添加无 GN 的特征补充模块，性能相较于

基准实验的性能几乎没有变化，但是再附带 GN 处理后 AP_{50}^{tiny} 提升了 0.53%，如果再选择合适的融合因子则性能提升可以进一步扩大到 0.83%，说明富含更多细节特征的浅层特征层经过合适的卷积计算后是可以帮助到弱小目标检测，用来创造出一个更适合做弱小目标检测的特征层。

4.4 融合因子的分析与解释

4.4.1 对融合因子隐式学习的研究

在前文中分别对融合因子的作用前提以及如何设置一个有效的融合因子分别进行了研究，在这些实验中，无论是采用可学习参数的方式还是固定融合因子的方式，都是将融合因子作为网络中的一个参数显式的表达了出来，由此产生的问题是：为什么没有融合因子的网络自身没有学习到具有有融合因子网络的性能？网络中的参数都具有学习能力，但是在学习过中原本的网络参数并没有将融合因子的表达能力兼容，融合因子没有被隐式的被网络学习到，下边将从 FPN 相关卷积的初始化和目标数据集体量两个方面分别进行研究和分析。

4.4.1.1 对 FPN 相关卷积初始化研究

从网络结构的原理上分析：FPN 的输入为从骨干网络中不同的阶段提取的特征（C_2~C_5或C_2~C_5，FPN 的输出一般是 5 层，输入不足的层数会由C_5下采样产生），输入 FPN 后，不同的特征层都会各自经过一组 1×1 的卷积，这些卷积的目的是将来自不同阶段的通道数目不同的特征改变为通道数相同的特征（默认为 256 个通道），方便后续的线性计算。在深层的特征经过上采样后和浅层特征进行像素级别的点对点相加融合，之后不同大小的特征又会过一组 3×3 的卷积，经过这层卷积后这个特征才会作为 FPN 的输出。卷积本身是具有学习能力的，但在没设定融合因子的条件下，这些具有学习能力的卷积没能学到融合因子的能力，这是值得进一步的探究。

在分析了 FPN 的结构后，初步推测可能是相关卷积的初始化导致的，因为在网络开始训练的时候，如果添加融合因子，就变相缩小了相关卷积的初始化区间。原版 FPN 中的f_{inner_i}与f_{layer_i}卷积的初始化方式均采用 kaiming_uniform 初始

化方式。通过数学上的推导，找到了另一种初始化方式，称为隐融合因子初始化，具体设置如下图。这种方式是在不添加融合因子的情况下只改变相关卷积的初始化区间，等效于在不改变网络初始化的情况下添加融合因子。在不同初始化条件下的检测结果比较如表 4.14 所示。

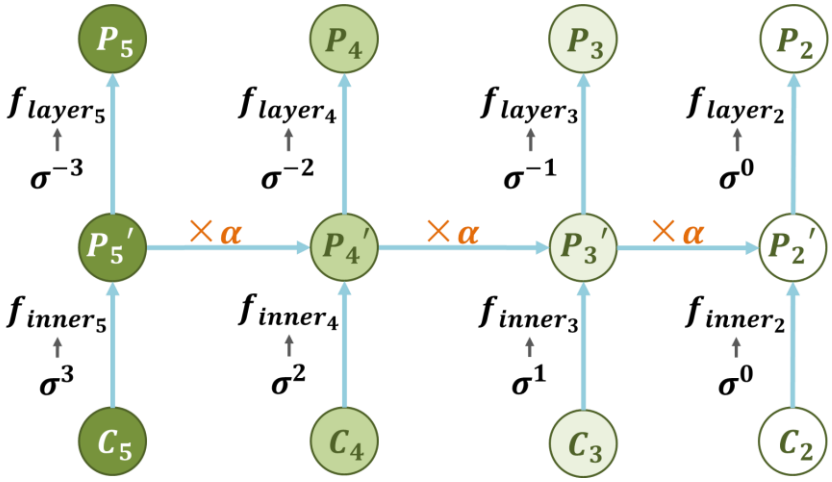

图 4.1　特征金字塔网络的初始化修改方式

Figure 4.1　Initialization and modification of feature pyramid network

上图中，橙色 α 代表融合因子作用的位置，绿色图标代表特征层，箭头指代特征金字塔网络的计算过程，颜色越深代表经过的网络计算层数越多，f_{inner_i}与f_{layer_i}代表网络中的相关卷积，σ^i代表对相关卷积修改的程度。

在原版的 FPN（α= 1）中，将f_{inner_i}的参数乘以 σ^{i-2} 并将f_{layer_i}的参数除以 σ^{i-2} 等效于保持f_{inner_i}，f_{layer_i}固定并设置融合因子 α=σ。以P_4为例分析，在原版初始化和隐融合因子初始化条件下可以分别表示为公式（4.6）和公式（4.7），在 α=σ 时，公式（4.6）和公式（4.7）得到的结果是相同的。

$$P_4 = f_{layer_4}(f_{inner_4}(C_4) + \alpha_4^5 * f_{upsample}(f_{inner_5}(C_5))) \tag{4.6}$$

$$P_4 = \sigma^{-2} * f_{layer_4}(\sigma^2 * f_{inner_4}(C_4) + \sigma^{-3} * f_{upsample}(f_{inner_5}(C_5))) \tag{4.7}$$

因此，传统的 FPN 具有隐式学习有效 α 的潜在能力。通过实验进一步研究如何通过调整f_{layer_i}参数的初始化来激活该能力。使用不同的f_{inner_i}和f_{layer_i}初始值通过将它们的对应系数相乘进行试验，如图 4.1 所示。

如表 4.14 所示，该设置 0.5-power + α=1 无法提升基准实验性能，0.5-power 代表 σ=0.5。我们进一步进行实验：将 α 设置为 σ，并保持f_{inner_i}和f_{layer_i}的上述初始配置(0.5-power + α=0.5)，其性能类似于保持基准实验的初始化条件但直接加融合因子固定值为 0.5 的性能，表 4.14 说明了改变相关卷积初始化区间的方式无法隐式学习融合因子的在特征金字塔中起到的作用。

表 4.14　TinyPerson 上的 σ 幂次方初始化的结果

Table 4.14　The result of initialization of σ power on TinyPerson

方式	AP_{50}^{tiny}	MR_{50}^{tiny}
baseline	46.56	88.31
0.5-power + α=1	46.94	87.98
0.5-power + α=0.5	48.17	87.17

4.4.1.2　对数据集体量研究

神经网络的学习过程是数据驱动的，越多的数据量可以使网络模型具有越好的泛化性能，因为越多的数据量就可以让网络学习过越多种类的数据，网络在测试集上泛化性能才会变得越好。

在检测领域最具有代表性的数据集是 COCO，一是因为 COCO 种类丰富（80 类），二是 COCO 的数据量大（十万数据量级的图片数量），在 COCO 上训练得到的网络一般被认为是具有比较好的泛化性能。Tiny Cityperson 和 TinyPerson 对不同的融合因子是敏感的，它们的数据量相似（千数据量级图片数量，万数据量级标注数量），和 COCO 并不是一个数据体量。

因此在 COCO 上设计实验来验证大体量数据集能否能促使 FPN 隐式学习融合因子具有得表达能力。在 COCO100（即 Tiny COCO，图片在网络输入时统一将短边设置为 100 个像素）上设置不同的融合因子，得到在不同融合因子下的性能，观察在大体量数据集条件下融合因子的变动是否能影响性能。

在图 3.3 中，由融合因子的变动引起的性能峰值现象是明显的。说明在 COCO

这个数据集体量下训练，网络没能完全隐式地学习到融合因子对网络表达能力得提升效果。从数据集的角度出发，在其他弱小目标检测数据集上都可以通过设置融合因子得方法来提升性能。

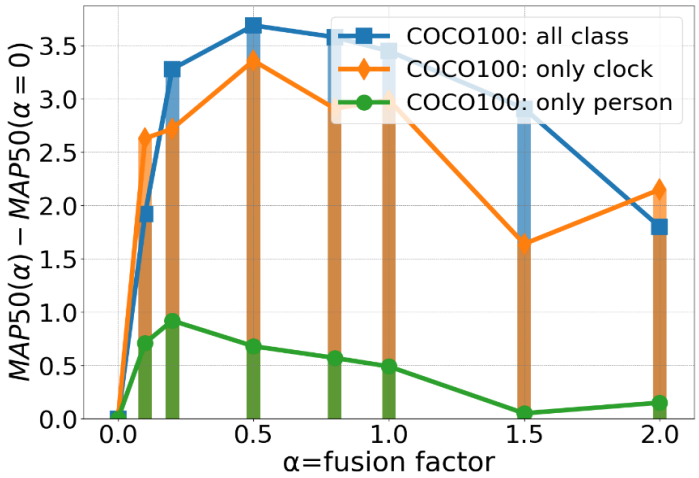

图 4.2　COCO100 中融合因子对不同类性能的影响

Figure 4.2　The influence of fusion factors on different types of performance in COCO100

同时本研究对弱小目标检测中的多类别问题也进行了进一步的探究：COCO 是一个长尾的多类别数据集。例如，人数接近 COCO 中总标注框总量的二分之一，而其他类别则相对较少。因此，进行了融合因子对不同数据量的不同类别的影响的实验与分析。

如图 4.2 所示，在同次实验对不同类别分类统计，由融合因子引起的峰值现象在不同类别上表现并不相同，对少样本数量的 clock 类影响较大，对多样本数量的人影响较小。结果表明，当训练数据集足够大时，会出现不同类别受到融合因子影响不一致的现象。但在 COCO 中，绝大多数类别数据量不够大其受融合因子的影响比较大，导致最终整体的性能对融合因子设置敏感。融合因子对于不同类别物体的影响可以在后续研究中继续探索。

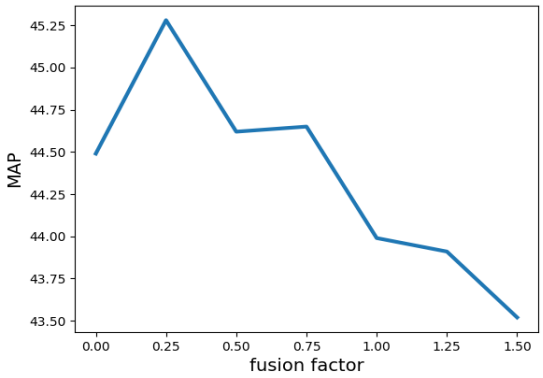

图 4.3　融合因子对 COCO100-人类数据集的性能影响

Figure 4.3　Influence of fusion factor on the performance of coco100-human dataset

为了进一步讨论融合因子对类别影响以及更严谨的实验结果，进行 COCO100-人类数据集的实验，即选取 COCO100 中所有属于人这一类的数据构成一个新的数据集 COCO100-人类，因为 COCO 中人类的标注占据总体标注数量的大约 50%，这就保证了新的数据集体量仍然是 COCO 这一级别的同时也排除了类别间的相互作用可能对结果产生的影响。实验结果如图 4.3，图的横坐标为融合因子的取值区间，纵坐标为相应融合因子条件下检测器的性能，实验框架为 RetinaNet，从图中可以看出存在明显的峰值现象，说明排除类别影响后，基于融合因子的方法仍然对大体量的弱小目标检测数据集有性能提升的效果。

4.4.2　对融合因子在多尺度数据集的研究

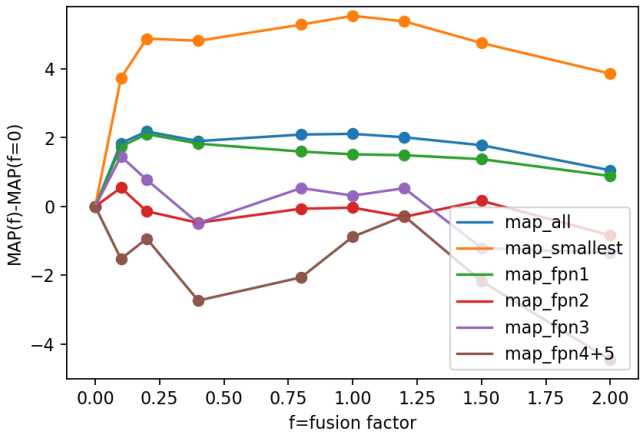

图 4.4　融合因子对 COCO 数据集的影响（未修正）)

Figure 4.4　Influence of fusion factor on coco dataset (uncorrected)

本小节探究了融合因子对跨尺度数据集的影响，实验框架选择为 RetinaNet，实验数据集为 COCO，保持图片短边 800 像素输入网络。实验结果如图 4.4，其中图例中 map_all 为[0,+∞]，map_smallest 为[0,32)，map_fpn1 为[32,64)，map_ fpn2 为[64,128)，map_ fpn3 为[128,256)，map_ fpn4+5 为[256,1024)，对 COCO 涉及到的尺度区间进行了详细的划分，横坐标为融合因子的值，纵坐标为性能变化的相对值。如上图所示，在融合因子为 0 点之外，除了 map_ fpn4+5 曲线显示出明显的波动情况，其他尺度的曲线并没有显示出明显的变化趋势。我们仔细分析了原因并研究了 COCO 评测代码，发现在 COCO 的评测代码中，对物体尺度的划分是依据标注中的 area（面积）字段，这个字段在 COCO 的标注中代表的是物体掩膜（mask）的面积，这个面积是小于标注框的面积，这与 TinyPerson 的标注规则不同，因此 COCO 中许多物体的区域面积是大于 area 字段面积，相当于把所有物体在一定程度上按缩小处理。为了公平比较得出结论，我们替换 COCO 中的 area 字段为标注框的面积，重新得到性能曲线，如下图。

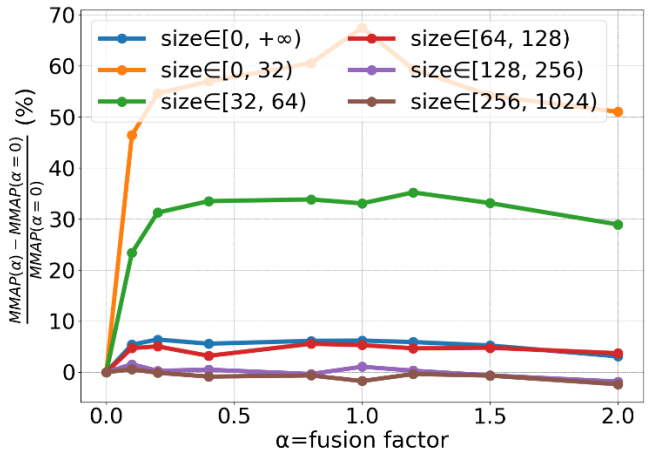

图 4.5　融合因子对 COCO 数据集的影响（已修正）

Figure 4.5　Influence of fusion factor on coco dataset (Revised)

图 4.5 为修正后融合因子对 COCO 数据集的性能影响趋势图，图的横坐标为融合因子的值，纵坐标为性能变化的相对值，不同尺度的物体曲线用不同颜色来表示。如上图所示，小尺度区间物体的性能明显受到了融合因子变化的影响，在融合因子取 1 处出现了明显的峰值，其他尺度的物体受到的影响不大，最优值区

间与弱小目标检测数据集的最优值区间不相同，可以得到的结论是融合因子影响了不同类型数据集中的小尺度物体的性能，但在 COCO 这类跨尺度数据集中，由于不同尺度目标的相互影响最优值区间不同于弱小目标数据集，这在未来的研究中可以更深入的探究其原因。

4.4.3 对融合因子梯度反传的研究

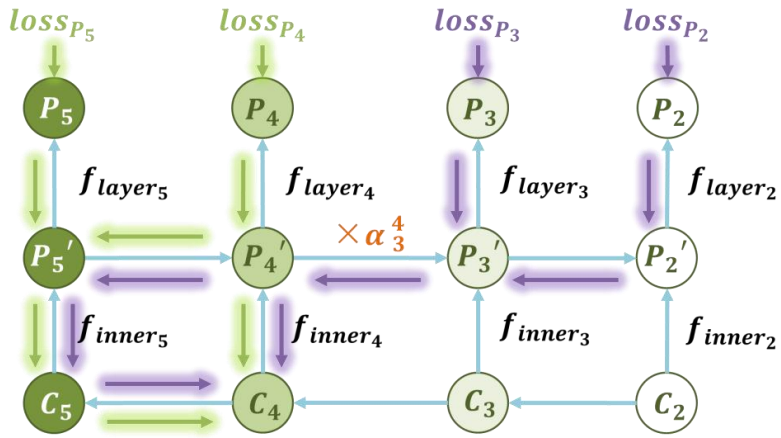

图 4.6 特征金字塔网络中以 C4 为例的梯度流向图

Figure 4.6 Gradient flow graph with C_4 as an example in FPN

这节将从梯度传播的角度讨论融合因子（α）的影响和作用。在不损失一般性的情况下，以α_3^4对C_4的梯度更新为例分析：FPN 中的融合因子如何影响骨干网络（backbone）的参数优化。上图为梯度流向示意图，其中，紫色箭头代表来自于浅层特征层的梯度（$loss_{p_2}$，$loss_{p_3}$），绿色箭头表示来源于深层特征层的梯度（$loss_{p_4}$，$loss_{p_5}$），越深的特征层被越深的绿色表示，融合因子α_3^4用橙色表示为了简化示意图，只有与C_4相关的梯度流动被显示表达出来。另外，深浅是相对的，P_4对P_3是深层，对P_5是浅层。

$$\Delta C_4 = -\eta * [\sum_{j=1}^{N_{p_4}} \frac{\partial(loss_{P_4}^j)}{\partial C_4} + \sum_{j=1}^{N_{p_5}} \frac{\partial(loss_{P_5}^j)}{\partial C_4} + \alpha_3^4 * (\sum_{j=1}^{N_{p_2}} \frac{\partial(loss_{P_2}^j)}{\partial P_3'} + \sum_{j=1}^{N_{p_3}} \frac{\partial(loss_{P_3}^j)}{\partial P_3'}) * \frac{\partial P_3'}{\partial C_4}] \tag{4.8}$$

C_4层的梯度更新过程可以表示为上式（4.8）。其中$loss_{p_i}$表示与第 i 层相对应的分类和回归损失，N_{p_i}表示 FPN 的第 i 层上样本，j 表示对应的样本序列的位序。ΔC_4获得的梯度可以划分为两个部分，来自浅层特征层的梯度$\Delta C_{4_{shallow}}$由$loss_{p_2}$和$loss_{p_3}$构成，来自深层特征层的梯度$\Delta C_{4_{deep}}$由$loss_{p_4}$和$loss_{p_5}$构成。

上式表示C_4需要学习两种任务：较深层的检测任务（P_4，P_5）和较浅层的检测任务（P_2，P_3）。当使用更大的α时，C_4将接受更多用于浅层检测任务的信息，而丢失了更多用于深层检测任务的信息，当使用更小的α时，C_4将接受更多用于深层检测任务的信息，而丢失了更多用于浅层检测任务的信息。

FPN 是多任务学习，将不同尺度的目标划分为不同的人任务。具体而言，如果省略 FPN 中自上而下的连接，那么每一层只需专注于检测与尺度高度相关的目标，即浅层学习小目标，深层学习大目标。然而，在 FPN 中，由于自上而下和侧向连接机制，每一层也收到来自其他层的间接梯度，每一层需要学习几乎所有大小的目标，深层特征也需要学习小目标的特征表示。对于弱小目标检测而言，按照样本的尺度划分，小目标都基本都集中于浅层学习，深层特征对弱小目标帮助不大。当数据集为弱小目标数据集，大部分目标为弱小目标，导致适合学习大目标的深层样本数量不够。连接机制又要求每一层不仅需要关注其对应的尺度的目标，还需要从其他特征层获得更多的训练梯度，每一层由于学习任务难度的变大，导致网络学习能力不足，从而导致不同层之间竞争的加剧。融合因子控制着这两个需求的优先级，并在两者之间取得平衡。因此，传统的 FPN 对应的融合因子为 1，不适用于弱小目标的检测。

当目标数据集为物体尺度较大的数据集中（例如 COCO800），关于物体的信息非常丰富，任务的难度相对较低。 这时如果为了深层检测放弃浅层部分信息（应用较小的融合因子），最终性能几乎不会降低，而如果保留它们（应用较大的融合因子），则性能也不会得到很大改善。所以在这种大尺度数据集上，α的设置不太敏感。 并且数据集物体尺度越大，α设置的灵敏度越低。 换句话说，在较大范围内设置α的性能几乎是相同的，分析结果与图 4.5 一致。

对于弱小目标检测数据集而言，任务难度变大，信息量较少，这决定了在每

一层可以学习的信息量较少。 因此，放弃任何信息都是危险的。 因此，无论是深层还是浅层中的检测任务都倾向于C_4可以保留更多有益于他们的信息，他们都倾向于占据更大的C_4梯度比例。P_2和P_3中的检测任务希望α_3^4越大，P_4和P_5倾向于α_3^4越小。最后，最佳性能取决于折衷值，与该值的偏差越大，性能将越差，因为它太倾向于深层任务或浅层任务，更容易丢失另一个任务的重要信息。

4.5 本章小结

第四章介绍了研究结果及分析，4.1 小节阐述了目前目标检测领域主要的评价指标以及其计算过程，并且介绍了小目标检测领域更加适用的指标。4.2 小节介绍了本研究中两种基准实验的条件和具体实验设置。4.3 小节罗列了不同融合因子实现方式的性能比较和选定的基于统计的设置融合因子为主方法的原因，还罗列基于统计设置的方法在不同的数据集，不同检测器，不同的骨干网络上带来的性能提升，并进行了和其他先进检测算法的横向比较，当基于统计的融合因子和尺度匹配方法合用时也取得了性能的提升，证明了所提方法的兼容性和推广性。4.4 小节探讨了融合因子的作用机理主要分为三方面：1.融合因子是否可以隐式学习，并给出了在相关卷积初始化和数据集体量两方面的解释；2.讨论了融合因子在多尺度数据集上的表现并分析其原因；3.深入分析了融合因子在网络梯度回传时产生的影响并结合特征金字塔的原理做出分析与解释。下一章将是本论文的总结和展望。

第 5 章　结论与展望

弱小目标检测（Tiny Object Detection）作为计算机视觉领域的子课题，有着广泛的应用前景和科研价值。本研究提出了一个新的概念——融合因子，并且探究了融合因子在弱小目标检测中的作用，从融合因子的方法与原理都进行了详细的阐述，同时详尽的实验结果也证明了研究的可靠性。本节将对研究的内容进行整体归纳并规划未来的研究内容。

5.1　全文总结

本文首先介绍了计算机视觉这一领域的发展历程，然后着重介绍了弱小目标检测这一子任务，并且说明了其存在的重要科研价值和应用前景。同时还总结了弱小目标检测任务的难点，指出了基于特征金字塔网络的算法框架难以在弱小目标检测任务上高效发挥的原因是缺乏适应性的改进，并且提出了一个影响弱小目标检测关键性能的新概念——融合因子。

第二章对本研究涉及到几个方面：检测算法、检测数据集、小目标检测、特征融合这四个方面的前沿与经典工作进行了总结与陈述。

接下来论文从融合因子影响弱小目标检测性能的现象出发，研究了特征金字塔网络的作用机理，通过比较融合因子在不同数据集对算法性能影响程度不同，分析了融合因子影响弱小目标检测性能的原因，并探讨了如何设计一个有效的融合因子，并给出基于可学习参数、注意力机制、统计方法等不同的融合因子设计方式，最后从方法的实用性和性能指标综合下选择了基于统计融合因子设置方法作为本研究的方法。

第四章基于选定方法进行了一系列科学严谨的实验，包括所提方法与不同算法的比较，在不同的实验条件下都能取得性能上的一致提升，也证明了方法的有效性与泛化能力。

论文进一步从网络隐式学习、多尺度训练接、特征金字塔网络中梯度传播等角度分析融合因子的原理，并给出解释。结果表明，通过调整特征金字塔网络相

邻层的融合因子，可以自适应地促使浅层聚焦于弱小目标的学习，从而提高了对弱小目标的检测能力。

5.2 未来展望

科研工作的目的是可以运用到军用、民用、商用等各个领域发挥更大的价值，但是从目前的深度学习的“黑盒”算法的特点来看，科研成果距离实际落地应用还有不小的距离，未来的工作也可以从缩短这一距离、拓展任务场景以及对原理进一步探究追寻本质三个方面开展。

对融合因子的原理可以进行进一步的探究，融合因子对单尺度的弱小目标检测数据集和多尺度的通用目标检测数据集中的小尺度目标性能均有影响，但表现的现象却不相同，比如性能峰值对应的融合因子的值不在同一区间，为什么最优值不同影响程度也不同？多尺度的数据集在除小目标以外的其他尺度对小目标的学习影响具体是什么？这些都是可以继续探究得问题。对融合因子的原理继续深入有利于形成一个全面而又完备的知识体系。

对于融合因子的应用场景也可以进一步拓展，基于融合因子的方法在其他的比较难的检测任务上的表现如何，除了在多尺度数据集的表现之外，在遮挡问题、密集检测问题、人群计数问题等课题上的研究可以拓展融合因子应用的场景。另外，也可以将融合因子拓展到更多的基于特征金字塔的网络中，设计出更适合做小目标检测的检测算法。

对于小目标任务本身而言，现有的数据集主要分为两种：一是通过下采样原始大图得到，这会破坏原本的图片结构；二是类似于 TinyPerson 的弱小行人检测，场景和目标类别相对单一，这也限制了基于融合因子的研究在实际中的应用。在未来，如果可以在场景更加完善，类别更加丰富的弱小目标检测数据集上进行研究，可以让基于融合因子的研究在实际应用中发挥更大的价值。

参考文献

[1] 迟健男. 视觉测量技术[M]. 机械工业出版社, 2011.

[2] Lowe D G . Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[3] Dalal N , Triggs B . Histograms of Oriented Gradients for Human Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2005: 886-893.

[4] Felzenszwalb P F , Mcallester D A , Ramanan D . A discriminatively trained, multiscale, deformable part model[C] In Proceedings of IEEE Computer Vision and Pattern Recognition, 2008:

[5] 陈方芳. 基于目标对筛选和联合谓语识别的视觉关系检测[D]. 浙江大学, 2019.

[6] Yu X, Gong Y, Jiang N, et al. Scale Match for Tiny Person Detection[C]. In Proceedings of IEEE Winter Conference on Applications of Computer Vision, 2020: 1246-1254.

[7] 薛政钢. 基于多群体蚁群算法的多无人机协同搜索方法研究[D]. 河南大学, 2018.

[8] Zhu P , Wen L , Xiao B , et al. Vision Meets Drones: A Challenge[C] . In Proceedings of IEEE European Conference on Computer Vision. Springer, Cham, 2018.

[9] Lin T, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2017: 936-944.

[10] Everingham M , Gool L V , Williams C , et al. The Pascal Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision, 2010, 88(2):303-338.

[11] Lin T, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[C]. In Proceedings of European Conference on Computer Vision, 2014: 740-755.

[12] Zhang S, Benenson R and Schiele B. CityPersons: A Diverse Dataset for Pedestrian Detection[C].In Proceedings of IEEE Computer Vision and Pattern Recognition, 2017: 4457-4465.

[13] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C]. In Proceedings of Neural Information Processing Systems,

2015: 91-99.

[14] Wang X , Han T X , Yan S . An HOG-LBP human detector with partial occlusion handling[C]. In Proceedings of IEEE International Conference on Computer Vision, 2009: 32-39.

[15] Krizhevsky A, Sutskever I and Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[C]. In Proceedings of Neural Information Processing Systems, 2012: 1106-1114.

[16] Girshick R , Donahue J , Darrell T , et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2014: 580-587.

[17] 宋姚烨. 复杂交通环境下的车辆检测算法研究[D]. 江苏大学, 2019.

[18] He K , Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.

[19] Girshick R. Fast R-CNN[C]. In Proceedings of IEEE International Conference on Computer Vision,2015: 1440-1448.

[20] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]. In Proceedings of IEEE International Conference on Computer Vision, 2017: 2980-2988.

[21] Shifeng Zhang, Longyin Wen, Xiao Bian, Zhen Lei, Stan Z. Li. Single-Shot Refinement Neural Network for Object Detection[C] In Proceedings of IEEE Computer Vision and Pattern Recognition, 2018: 4203-4212.

[22] Cai Z and Vasconcelos N. Cascade R-CNN: Delving Into High Quality Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2018: 6154-6162.

[23] Lu X , Li B , Y Yue, et al. Grid R-CNN[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2019: 7363-7372.

[24] Pang J, Chen K, Shi J, et al. Libra R-CNN: Towards Balanced Learning for Object Detection[C].In Proceedings of IEEE Computer Vision and Pattern Recognition, 2019: 821-830.

[25] Redmon J, Divvala S, Girshick R, Et al. You Only Look Once: Unified, Real-Time Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2016: 779-788.

[26] Liu W，Anguelov D，Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. In Proceedings of European Conference on Computer Vision, 2016: 21-37.

[27] Lin T, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]. In Proceedings of IEEE International Conference on Computer Vision, 2017: 2999-3007.

[28] Gupta A , P Dollár, Girshick R . LVIS: A Dataset for Large Vocabulary Instance Segmentation[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2019: 5356-5364.

[29] Yang S, Luo P, Loy C, et al. WIDER FACE: A Face Detection Benchmark[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2016: 5525-5533.

[30] Dollar P , Wojek C , Schiele B , et al. Pedestrian Detection: An Evaluation of the State of the Art[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(4): 743-761.

[31] Shao S , Li Z , Zhang T , et al. Objects365: A Large-Scale, High-Quality Dataset for Object Detection[C]. In Proceedings of IEEE International Conference on Computer Vision,2019: 8429-8438.

[32] Andreas Ess, Bastian Leibe, Konrad Schindler, Luc Van Gool:.A Mobile Vision System for Robust Multi-Person Tracking[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2008.

[33] Enzweiler M , Gavrila D M . Monocular Pedestrian Detection: Survey and Experiments[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31:2179-2195.

[34] Geiger A , Lenz P , Urtasun R . Are we ready for autonomous driving? The KITTI vision benchmark suite[C. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2012: 3354-3361.

[35] Pang J , Li C , Shi J , et al. R²-CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images[J]. IEEE Transactions on Geoscience & Remote Sensing, 2019:1-13.

[36] Cordts M , Omran M , Ramos S , et al. The Cityscapes Dataset for Semantic Urban Scene

Understanding[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2016: 3213-3223.

[37] Singh B and Davis L. An Analysis of Scale Invariance in Object Detection - SNIP[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2018: 3578-3587.

[38] Singh B, Najibi W, and Davis L. SNIPER: Efficient Multi-Scale Training[C]. In Proceedings of Neural Information Processing Systems, 2018: 9333-9343.

[39] Deng C , Wang M , Liu L , et al. Extended Feature Pyramid Network for Small Object Detection[J]. IEEE Transactions on Multimedia, 2021.

[40] Noh J , Bae W , Lee W , et al. Better to Follow, Follow to Be Better: Towards Precise Supervision of Feature Super-Resolution for Small Object Detection[C]. In Proceedings of IEEE International Conference on Computer Vision,2019: 9724-9733.

[41] Chen Y, Zhang P, Li Z, et al. Stitcher: Feedback-driven Data Provider for Object Detection[J].ArXiv, abs/2004.12432, 2020.

[42] Li Y, Chen Y, Wang N, et al. Scale-Aware Trident Networks for Object Detection[C]. In Proceedings of IEEE International Conference on Computer Vision, 2019: 6053-6062.

[43] Liu Z , Gao G , Sun L , et al. IPG-Net: Image Pyramid Guidance Network for Small Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition Workshop, 2020: 4422-4430.

[44] Liu S , Qi L , Qin H , et al. Path Aggregation Network for Instance Segmentation[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2018: 8759-8768.

[45] J Nie, Rao M A , Cholakkal H , et al. Enriched Feature Guided Refinement Network for Object Detection[C]. In Proceedings of IEEE International Conference on Computer Vision,2019: 9536-9545.

[46] Sun K , Xiao B , Liu D , et al. Deep High-Resolution Representation Learning for Human Pose Estimation[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2019: 5693-5703.

[47] Liu S , Huang D , Wang Y . Learning Spatial Fusion for Single-Shot Object Detection[J]. ArXiv, abs/ 2019. 1911.09516.

[48] Wang X , Zhang S , Yu Z , et al. Scale-Equalizing Pyramid Convolution for Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2020: 13356-13365.

[49] Ghiasi G , Lin T Y , Le Q V . NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2019: 7036-7045.

[50] Tan M , Pang R , Le Q V . EfficientDet: Scalable and Efficient Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2020: 10778-10787.

[51] Guo C , Fan B , Zhang Q , et al. AugFPN: Improving Multi-scale Feature Learning for Object Detection[C]. In Proceedings of IEEE Computer Vision and Pattern Recognition, 2020: 12592-12601.

[52] Li Z , Lang C , Liew J , et al. Cross-layer Feature Pyramid Network for Salient Object Detection[J]. IEEE Transactions on Image Processing, 2021.

[53] Tian Z, Shen C, Chen H, et al. FCOS: Fully Convolutional One-Stage Object Detection[C]. In Proceedings of IEEE International Conference on Computer Vision, 2019: 9626-9635.

[54] Zhang X, Wan F, Liu C, et al. FreeAnchor: Learning to Match Anchors for Visual Object Detection[C]. In Proceedings of Neural Information Processing Systems, 2019: 147-155.

致　谢

感谢韩振军老师在我读研期间的谆谆教诲，不仅在科研方面基于我非常大的帮助，同时在生活中也对我十分照顾，在我研究生一年级的时候，由于自身转专业读研，在基础知识方面需要很多补充，韩老师从编程以及基础的机器学习知识方面给我很多可行的建议和具体指导。研究生二年级时，在我科研上遇到问题举步维艰的时候，韩老师从研究方向和思考细节上教会我形成了良好的科研思维也带我开拓了科研视野。研究生三年级时，韩老师除了在学习上的指导，还经常与我交流职业选择方面的问题，也让我受益良多。感谢韩老师三年来的帮助与教诲！

感谢师兄师姐一直以来的帮助，余学辉师兄在我读研期间给予我的巨大帮助，在我编程基础薄弱的时候一点一滴教我如何写好程序，同时他也扮演了深度学习求知路上的引路人角色，从深度学习的损失函数给我讲起帮我搭建了整个知识框架。同时每当我做实验遇到困难时，找他讨论，他总是不厌其烦地帮我解决问题，在这个过程中我也学会了设计实验中的基本方法与思考方式。另外，感谢丁瑶师姐在投稿论文时的巨大帮助，在她一字一句和对文章的梳理中我学到了如何组织行文逻辑以及文意表达，十分感谢在投稿前夜支撑不住之时给我了莫大的鼓励。

感谢同届蒋楠同学一直以来的帮助，我从蒋楠快速而又严谨的思维以及过人的理解能力中学习到了很多。感谢同实验室的其他师弟师妹们，韩许盟，彭潇珂，王岿然，吴狄，陈鹏飞，黄志勋，一直以来的帮助。感谢三年邻居魏君轩，五年博士刘冰昊和其他同届同学总能给我带来生活中的幽默与快乐。另外，感谢张聚良对论文支持。

最后最要感谢我的父母和家人，感谢父母对我从小到大的辛勤付出，把我培育成人并对我的学业一直鼎力支持。在研究生就读期间，家庭永远是最温暖的岗位，每每有不顺利的事情，父母一直在背后从各个方面给予我全力支持，让我坚定信心鼓起勇气一次次挑战未知。

在接下来的人生之路中，我会带着所有人的希望与期限继续坚定出发。

作者简历及攻读学位期间发表的学术论文与研究成果

作者简历：

姓名：宫宇琦　性别：男　出生日期：1994 年 11 月 30 日　籍贯：山西

2014 年 09 月——2018 年 06 月，在北京邮电大学经济管理学院获得学士学位。

2018 年 09 月——2021 年 06 月，在中国科学院大学电子电气与通信工程学院系攻读硕士学位。

获奖情况：

2020 年 11 月 国家奖学金

2020 年 6 月 三好学生

已发表（或正式接受）的学术论文：

[1] Gong Y , Yu X , Ding Y , Peng X, Ding Y, Han Z. Effective Fusion Factor in FPN for Tiny Object Detection[J]. 2021. IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2021.

[2] X Yu, Gong Y , Jiang N , Ye Q, Han Z. Scale Match for Tiny Person Detection[C]// 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2020.

[3] Yu X , Han Z , Gong Y , et al. The 1st Tiny Object Detection Challenge:Methods and Results[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020.

申请或已获得的专利：

[1] 韩振军，宫宇琦等. 基于 FPN 的融合因子的弱小目标检测方法：中国，202010752490.6[P]

[2] 韩振军，余学辉，宫宇琦等. 一种基于尺度匹配的弱小人体目标检测方法：中国，201910918836.2[P] （已授权）

[3] 韩振军，韩许盟，余学辉，宫宇琦等. 基于多源信息融合的弱小目标检测方法：中国，202010215165.6[P]

[4] 韩振军，张如飞，王攀，余学辉，宫宇琦，蒋楠等. 一种基于无监督深度孪生网络的视频去重方法：中国，202010214485.X[P] （已授权）

[5] 韩振军，蒋楠，余学辉，陈鹏飞，宫宇琦等. 基于精确尺度匹配的弱小人体目标检测方法：中国，202010746942.X[P]

另有软件著作权四篇：

[1] 韩振军，宫宇琦等.《基于融合因子的弱小目标检测软件》证书号:2020SR1052810

[2] 韩振军，余学辉，蒋楠，张如飞，宫宇琦等《基于尺度匹配的弱小航拍人体目标检测软件》 证书号:2019SR1047577

[3] 韩振军，余学辉，蒋楠，张如飞，宫宇琦等《基于多源信息融合的弱小目标检测软件》证书号:2020SR0129500

[4] 韩振军，蒋楠，余学辉，宫宇琦等《基于精确尺度匹配的弱小人体目标检测软件》证书号:2020SR1052818