



激光杂志

Laser Journal

ISSN 0253-2743,CN 50-1085/TN

《激光杂志》网络首发论文

题目: YOLO-Plane: 一种基于改进 YOLOv5 的飞机检测算法
作者: 梅礼坤, 陈智利
网络首发日期: 2023-08-07
引用格式: 梅礼坤, 陈智利. YOLO-Plane: 一种基于改进 YOLOv5 的飞机检测算法
[J/OL]. 激光杂志. <https://link.cnki.net/urlid/50.1085.TN.20230807.1449.004>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

YOLO-Plane：一种基于改进 YOLOv5 的飞机检测算法

梅礼坤，陈智利*

(西安工业大学光电工程学院 西安 710021)

摘 要：精准识别和定位飞机目标是航空安全和信息化战争胜利的关键，针对传统飞机目标识别抗干扰性差，对遮挡、光照、尺度敏感难应对复杂场景需求的问题，提出了一种基于改进 YOLOv5 的飞机目标检测算法。通过 IOU-NWD Similarity Metric for Bounding Boxes 策略解决了 IOU 机制对飞机小目标的标签分配歧义问题；使用 GFPN based on NLnet 模块完成了“跨层”与“跨尺度”的自适应融合，更加丰富和具有代表性的特征信息；使用 soft-NMS 方法解决了在目标密集区域中飞机小目标存在的漏检问题。在飞机数据集上进行实验，结果表明：与原始 YOLOv5 相比，改进后的模型在 Precision、Recall、mAP0.5、mAP0.5: 0.95 分别提高了 1.9%、10.4%、3.6%和 5.8%。该算法通过针对性的网络调整和模块迁移来提高模型对小型和遮挡的飞机目标的检测效果，并通过实验验证了该算法的优越性，实验结果表明，AIR-YOLO 在检测精度和鲁棒性方面优于 YOLOv5，解决了原始 YOLOv5 算法的小飞机目标误检的问题。

关键词：IOU-NWD；GFPN；NLnet；soft-NMS

中图分类号：TP391 文献标志码：A

YOLO-Plane: An aircraft detection algorithm based on improved YOLOv5

MEI Likun, CHEN Zhili*

(School of Optoelectronic Engineering, Xi'an Technological University, Xi'an 710021, China)

Abstract: Accurate identification and positioning of aircraft targets is the key to the victory of aviation safety and information war. In view of the problems that traditional aircraft target identification has poor anti-interference performance and is sensitive to occlusion, illumination and scale and difficult to cope with complex scene requirements, an aircraft target detection algorithm based on improved YOLOv5 is proposed. IOU-NWD Similarity Metric for Bounding Boxes solves the ambiguity of label assignment for aircraft small targets by IOU mechanism. Using GFPN based on NLnet module, the "cross-layer" and "cross-scale" adaptive fusion is completed, and more abundant and representative characteristic information is obtained. soft-NMS method is used to solve the problem of missing detection of small aircraft targets in crowded target areas. The experimental results show that compared with the original YOLOv5, the Precision, Recall, mAP0.5 and MAP0.5:0.95 of the improved model are increased by 1.9%, 10.4%, 3.6% and 5.8%, respectively. Through targeted network adjustment and module migration, the algorithm can improve the detection effect of the model on small and blocked aircraft targets. The superiority of the algorithm is verified by experiments. The experimental results show that AIR-YOLO is superior to YOLOv5 in terms of detection accuracy and robustness, which solves the problem of false detection of small aircraft targets in the original YOLOv5 algorithm.

Key words: IOU-NWD；GFPN；NLnet；soft-NMS

1 引言

飞机在军事作战中扮演着重要的角色，

能够提供迅速的机动能力和强大的火力支援，执行战略轰炸、对地和对海攻击、空中拦截、电子干扰等任务，为战争的胜利发挥

基金项目：国家级科研项目（G20210101）、陕西省科技厅（2023-YBGY-369）

作者简介：梅礼坤（1996.5-），男，硕士研究生，主要研究方向：数字图像处理、深度学习。E-mail:2112868362@qq.com

通讯作者：陈智利（1973.7-），男，博士，教授，硕士生导师，主要研究方向：光学加工、光学设计。E-mail:medichen@163.com

了重要作用^[1-3]。随着技术的进步，对飞机进行目标检测是一项具有实际意义和应用价值的任务^[4-5]。

传统飞机目标识别方法通常需要手工设计特征和分类器，需要大量的人工工作，而且很难处理复杂的变化和背景干扰等问题^[6-7]。此外，传统方法通常对遮挡、光照、尺度和角度等变化较为敏感，难以应对复杂的场景需求^[8-9]。深度学习方法具有出色的自适应性和泛化能力，自动地学习特征，减少了对人工特征的依赖，并且可以通过增加网络深度和复杂度来适应更加复杂的场景需求。此外，深度学习方法可以对图像中的背景干扰和目标变化进行自适应处理，并且可以通过大量的数据进行训练和优化，从而提高检测精度和鲁棒性。因此，深度学习技术在飞机目标识别任务中具有很大的优势。YOLO（You Only Look Once）^[10]是一种流行的目标检测算法，它通过使用深度卷积神

经网络（CNN）将整张图像作为输入，同时预测图像中的所有目标的类别和位置。相对于其他基于区域提取的目标检测算法，YOLO 具有更快的检测速度和更高的精度，因此被广泛应用于计算机视觉领域的各种应用中，如自动驾驶^[11-12]、安防监控^[13]等。

针对传统的飞机识别算法存在实时性不佳、容易受到环境影响、对遮挡和小尺寸目标的识别能力较弱、难以在大视野复杂环境下实现飞机目标检测等问题。本文提出了一种基于深度学习技术的飞机目标检测方法。该方法利用了卷积神经网络(CNN)和基于单阶段的目标检测的最新进展，能够在复杂场景中准确检测飞机。本文的主要贡献在于提出了一种鲁棒性较强的新型飞机目标检测方法，可以有效地处理遮挡和小尺寸目标，并且能够识别不同方向和尺度的飞机。该方法解决了在大视野复杂环境下飞机目标检测的难题，具有广泛的应用前景。

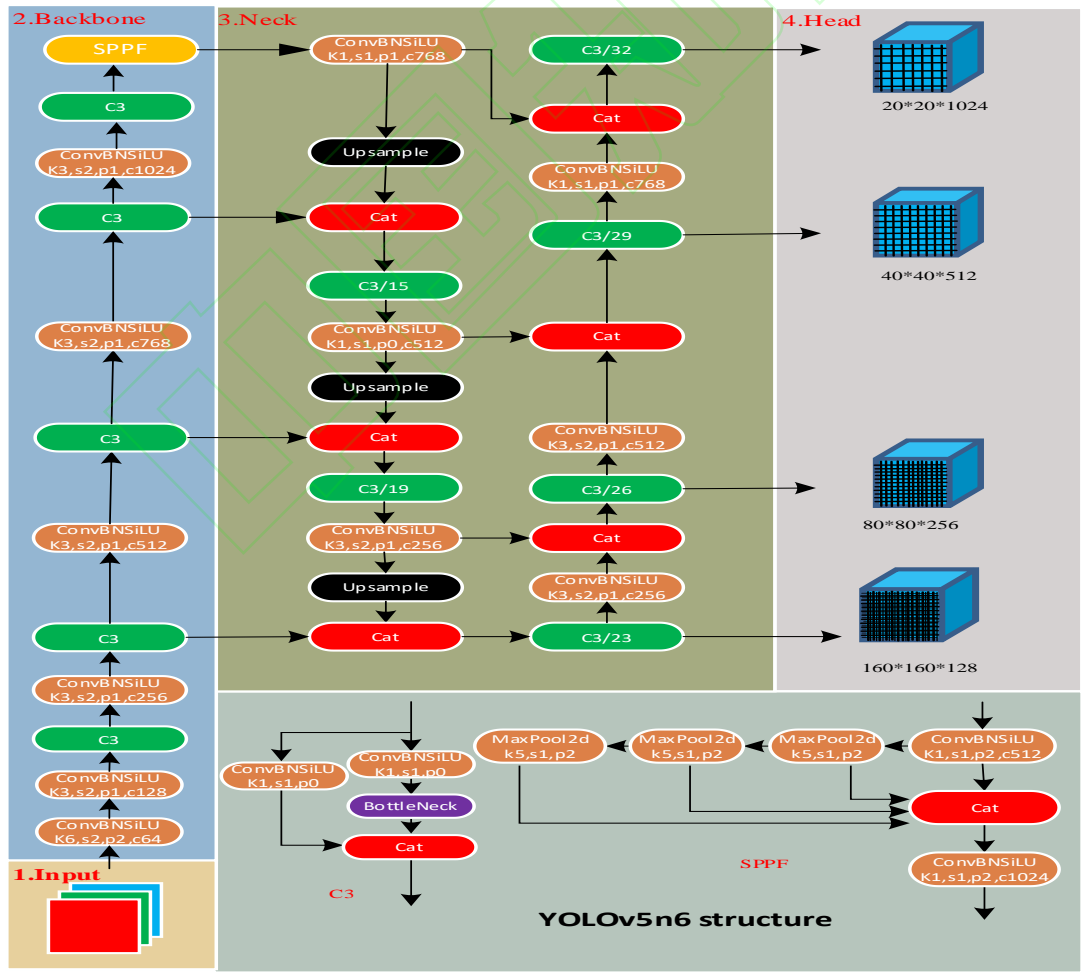


图 1 YOLOv5n6 整体架构图

2 YOLOV5 算法原理

YOLOv5 是一种轻量级的目标检测算法，它采用了一种基于中心点的检测方式，具有检测速度快、模型轻量、准确度高等特点。与 YOLOv3、YOLOv4 相比，YOLOv5^[14] 在算法结构、网络架构等方面都有了改进，因此具有更好的性能和可扩展性。YOLOv5 的输入是一张图片，输出是图片中所有目标的类别、置信度和位置，如图 1 所示。在 YOLOv5 中，首先采用一种图像增强的方式对输入图片进行处理，然后通过一个主干网络进行特征提取，再通过一个 neck 网络进行特征融合，最后通过一个 head 网络进行目标检测和分类。针对更大分辨率的输入图像（1024*1024），增加检测层、在更大的下采样倍率（64）上进行特征提取，可以进一步增强检测器对于细节信息的感知能力和对小目标的识别能力，由于在更大的下采样倍率下，特征图的分辨率会降低，因此需要通过增加检测层来提高感受野，同时保留更多的空间信息。此外，特征融合也是提高检测器性能的关键技术之一。在 64 倍下采样的特征图上进行特征融合，可以将来自不同层次的特征信息结合起来，从而进一步提高检测器的准确性和鲁棒性，通过特征融合，可以综合利用不同层次的特征信息，提取出更具有代表性的特征，从而进一步提高检测器的性能。

YOLOv5 采用的主干网络是 CSPNet（Cross Stage Partial Network）^[15]，它由一系列 CSP blocks 组成，CSPNet 的核心思想是将卷积操作分为两个阶段：中心卷积和周边卷积。中心卷积对输入数据进行卷积操作，而周边卷积则对卷积输出进行操作。这种分离卷积的方法可以减少网络参数和计算量，同时保持网络的表现力。CSP Block 是 CSPNet 的关键组件，每个 CSP block 由一个主干路径和一个残差路径组成，其中一条路径是一个小型的卷积神经网络，另一条路径是直接将输入数据进行降维处理，再通过一个大型的卷积神经网络进行卷积操作。两个路径的卷积输出进行拼接，形成 CSP Block 的输出。主干路径是一个深度卷积网络，可

以提取图像特征。残差路径则是一个浅层的卷积网络，可以对图像特征进行微调。通过这种方式，CSPNet 可以在保持高准确率的同时，大幅度减少计算量。

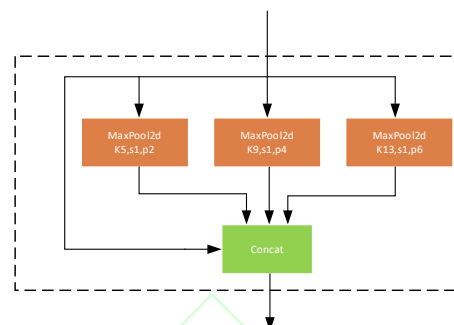


图2 SPP 模块

SPP (Spatial Pyramid Pooling)^[16] 是一种用于卷积神经网络的池化操作，用于处理不同大小的输入图像。传统的池化层是固定大小的，这意味着当输入图像尺寸发生变化时，必须重新训练模型。SPP 将多个池化层合并，池化层大小不再固定，而是可以处理不同大小的输入图像。具体来说，SPP 将输入图像分割成多个不同大小的区域，对每个区域应用池化操作，并将不同大小的池化结果合并在一起，形成一个固定大小的特征向量，以供后续的全连接层使用，如图 2 所示。

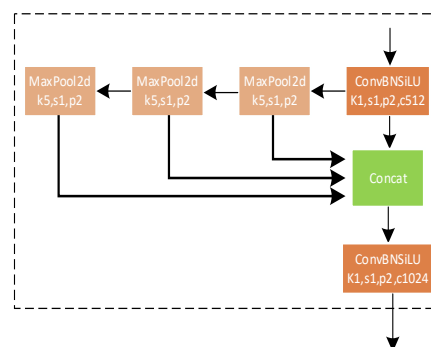


图3 SPPF 模块

SPPF (Serial Spatial Pyramid Fusion) 是对 SPP 的改进，主要在计算效率上进行了串行池化的改进操作，在保持二者相同的池化能力的同时提高了池化速率。在 SPP 中，不同层级的金字塔是并行地进行池化的，这就导致了计算效率较低的问题。为了解决这个问题，SPPF 提出了串行池化的思想，即将 SPP 中并行的池化操作改为串行。SPPF 在空间金字塔的每一层进行串行池化，将每层

的输出串联起来，最终得到一个固定大小的特征向量，如图 3 所示。

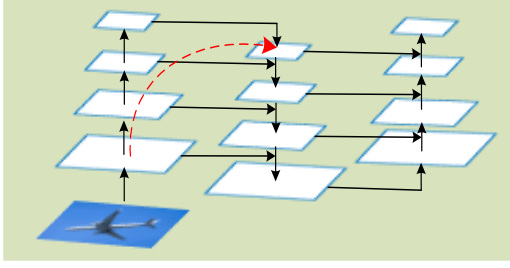


图 4 Neck 结构图

在目标检测中，neck 一般紧跟在主干网络后面，可以认为是特征提取的第二阶段。它的主要作用是将主干网络提取的特征进一步加工和优化，以便更好地进行目标检测。neck 网络是 YOLOv5 的重要组成部分，用于融合不同尺度的特征图。YOLOv5 采用了一种新的 neck 网络结构，即 FPN (Feature Pyramid Network) [17] + PANnet (Path Aggregation Network) [18]，如图 4 所示。FPN 可以从网络中不同层级提取出不同分辨率的特征图，然后通过上采样和下采样等操作将它们融合成一个金字塔形结构。PANnet 则是一种新的特征融合方式，可以在不同尺度的特征图之间进行信息交互和共享，从而进一步提高检测精度。

head 网络是 YOLOv5 的输出层，它是一个多任务的分类和回归网络。head 网络包括多个输出层，每个输出层都用于预测不同尺度和不同类别的目标。具体来说，每个输出层都包括一个卷积层和一个全连接层，用于预测目标的置信度、类别和位置。通过这种多层输出的方式，YOLOv5 可以有效地识别多个尺度、多个类别的目标。

3 改进 YOLOv5

3.1 IOU-NWD Similarity Metric for Bounding Boxes

小目标大小通常定义为目标在图像中占据的像素数量小于或等于 16x16 个像素，某些应用场景中，目标的尺寸可能更小。由于缺乏外观信息，最先进的检测器也不能在小目标上得到令人满意的结果。在目标检测、语义分割等计算机视觉任务中，交并比 (Intersection over Union, IoU) 通常被用作衡量两个有限样本集之间相似性的指标，它计算的是这两个集合的交集与并集之间的比率，即 Jaccard 相似系数，如式 1 所示。在目标检测任务中，通常需要将检测框与真实标注框进行比较，以判断检测算法的准确性，计算重叠程度的方法通常是使用交并比指标。IOU 指标是将检测框与真实标注框的重叠面积除以它们的并集面积，得到一个 0 到 1 之间的数值，如果检测框与真实标注框之间的 IOU 值高于一定的阈值，通常设置为 0.5 或 0.7，那么就认为该检测框正确地检测到了目标。在 YOLO 中，IOU 被用来判断检测框与真实标注框是否匹配，并根据 IOU 值来确定正负样本的分类。此外，IOU 还用于非极大值抑制 (Non-Maximum Suppression, NMS)，即在一个图像中，多个检测框可能会重叠在同一个目标上。在这种情况下，NMS 通过计算检测框之间的 IOU 值，选择具有最高得分且不重叠的检测框，从而确保每个目标只被检测一次。通过这种方式，NMS 可以有效地减少重复检测，提高算法的效率和准确性。但是，基于 IOU 的指标，如 IOU 本身及其扩展 (GIOU、CIoU)，对小目标的位置偏差非常敏感，在基于 Anchor 的检测器中使用，严重降低了检测性能。

$$IOU = \frac{Intersection\ Area}{BoundingBox1Area + BoundingBox2Area - IntersectionArea} \quad (1)$$

IOU 通常被用于衡量两个样本集之间的相似度，但是对于像素级别的小目标，轻微的位置偏差会导致明显的 IOU 下降，导致标签分配不准确。然而，轻微且相同的位置偏差对于不同像素的前景目标产生的

IOU 影响相差很大。这种现象意味着 IOU 度量对离散位置偏差的目标尺度是变化的，对 anchor_based 检测器中 Pos/Neg 训练样本的分配产生了歧义。

Wasserstein Distance，也称为 Earth

Mover's Distance^[19], 是一种用于衡量两个概率分布之间差异的距离度量。该距离度量可以应用于多种应用领域, 包括图像生成、自然语言处理、经济学等。Wasserstein 距离的本质是考虑将一种概率分布变换为另一种概率分布所需的最小代价。这种代价通常是两个分布之间对应元素之间的距离或权重乘积。通过求解这个问题, 我们可以获得一个距离度量, 它可以比 KL 散度等其他度量更准确地反映两个分布之间的相似性和差异性。KL 散度是一种非对称的度量方式, 用于测量两个概率分布之间的信息差异。它衡量的是两个分布之间的相对熵 (即相对于第一个分布的第二个分布的期望负对数似然), 并不是真正意义上的距离度量。

相比之下, Wasserstein 距离是一种基于运输问题的距离度量, 它衡量的是从一个分布将“质量”转移成另一个分布所需的最小代价。Wasserstein 距离具有以下特点:

1. Wasserstein 距离是对称的。即对于两个概率分布 P 和 Q , 它们之间的 Wasserstein 距离等于 Q 和 P 之间的 Wasserstein 距离。
2. Wasserstein 距离能够较好地处理非凸的分布形状, 对于多峰分布等情况, Wasserstein 距离可以更准确地描述两个分布之间的距离。
3. Wasserstein 距离在某些场景下具有更好的数学性质, 例如它可以更容易地进行优化求解, 而且它的估计不会受到零概率贡献的影响。

因此, Wasserstein 距离相对于 KL 散度在某些场景下有更好的优点和应用。例如在生成对抗网络 (GAN) 的训练中, 使用 Wasserstein 距离可以帮助生成器生成更真实的图像, 避免 GAN 训练过程中的梯度消失和模式崩溃等问题。两个二维高斯分布 P_1 和 P_2 的 Wasserstein 距离可以表示为式 2 所示:

$$W_2(P_1, P_2) = \sqrt{\frac{\min}{\pi \Gamma(P_1, P_2)} \int_{\mathbb{R}^2 \times \mathbb{R}^2} \|x - y\| d\gamma(x, y)} \quad (2)$$

式 1 中, $P_i(P_1, P_2)$ 表示所有联合概率分布 $\gamma(x, y)$ 的集合, 使得 $\gamma(x, y)$ 在 $P_1(x)$

和 $P_2(y)$ 上的边缘分布为 $P_1(x)$ 和 $P_2(y)$ 。对于高斯分布闭式解计算可得如式 3 所示:

$$W_2^2(P_1, P_2) = \|\mu_1 - \mu_2\|^2 + \text{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2})^{1/2}) \quad (3)$$

式 3 中, $\|\mu_1 - \mu_2\|^2$ 表示欧几里得范数, Tr 表示矩阵的迹。基于 Wasserstein 距离的小目标检测评估方法。具体而言, 该方法首先将边界框 (Bounding Box) 建模为二维高斯分布。然后提出了一种新的相似性度量标准, 称为归一化 Wasserstein 距离 (Normalized Wasserstein Distance, NWD)^[20], 通过计算两个高斯分布之间的距离来量化它们之间的相似性。这种 NWD 度量可以轻松集成到任何基于 Anchor 的检测器的分配、非最大抑制和损失函数中, 以替代常用的 IOU 度量。通过使用 NWD 度量, 可以更全面地评估两个边界框之间的相似度, 因为它考虑了更多的特征。与 IOU 相比, NWD 可以捕捉到更多的细节和空间信息, 因此在处理具有高度重叠的边界框时表现更好。其具有以下优点: 无论小目标之间有没有重叠都可以度量分布相似性; NWD 对不同尺度的目标不敏感, 更适合测量小目标之间的相似性。

然而, IOU 直接替换为 NWD 的策略会显著降低网络的收敛速度, 带来极大的时间成本。将 IOU 和 NWD 结合, 可以更全面地评估两个边界框之间的相似度。定义一个权重参数来控制两个度量值的贡献比例, 可以将两个度量值进行加权平均, 以得到最终的相似度分数。最终相似度分数如式 4 所示:

$$\text{similarity} = w1 * \text{IoU} + w2 * (1 - \text{nwd}) \quad (4)$$

$w1$ 和 $w2$ 分别为 IOU 和 NWD 的权重参数, 它们的和为 1。通过调整这两个权重参数, 可以得到不同的相似度计算方式, 以适应不同的应用场景和任务需求。

3.2 GFPN based on NLnet

在目标检测任务中, 神经网络需要能够同时检测不同大小的物体, 因为物体的尺寸和比例是多种多样的。然而, 神经网络的卷积层由于固定的卷积核大小, 只能对特定尺寸的物体具有较好的检测效果, 而对于不同尺寸的物体可能会出现检测不准或漏检的情况。为了解决这个问题, 通常使用图像金字塔 (Image Pyramid) 和特征金字塔 (Feature Pyramid) 方法, 将不同尺度的特征图融合在

一起，图像金字塔常见与主干网络，伴随着卷积的深入和采样的倍率逐步增大，可以在一定程度解决尺度变化问题，但是它需要消耗大量的计算资源和时间。特征金字塔方法中的 NECK，通常指的是神经网络中用于融合不同尺度特征的层或模块。常见的特征金字塔方法包括 FPN、PANet、BiFPN 等，相较于图像金字塔而言更加高效。FPN 是一种基于图像金字塔和特征金字塔的方法，通过在特征图上逐层上采样和下采样，将多尺度的特征图融合在一起。FPN 的优点在于简单易实现，同时在各种目标检测任务上都有着不错的性能表现，但它的缺点是没有充分利用高层特征对低层特征的修正作用。PANet 是一种类似于 FPN 的方法，但在特征融合时引入了一条上采样的路径，使得高层特征能够更充分地修正低层特征。因此，PANet 在处理尺度变化问题时具有更好的性能表现。但 PANet 的计算量比较大，训练和测试时间较长。BiFPN 是一种新兴的特征金字塔方法，采用了双向流的思路，在特征图上进行自下而上和自上而下的信息流动，同时使用多尺度的特征融合方式，以进一步提升模型的性能。BiFPN 在解决尺度变化问题时表现出了较好的效果，同时在计算量和时间上也有一定的优势。但在相关研究中表明，一定数量的堆叠可以获得更高的精度，但由于 BIFPN block 模块相互独立，堆叠过深可能存在梯度消失的问题。

Densenet 是一种深度残差网络 (ResNet) 的改进，它采用了密集连接 (Dense Connection) 的方式，通过将每一层的输入都连接到后面的所有层，使得特征可以更好地传递和共享，从而提高模型的性能和减少参数数量。Densenet 的主要思想是：每个层的输入都包含了前面所有层的输出，这种全连接的结构可以更好地利用每一层的特征，提高特征的复用和表征能力。在 Densenet 中，每个密集块 (Dense Block) 由多个卷积层和一个连接层 (Transition Layer) 组成，每个卷积层的输入是前面所有卷积层的输出，包括它自己的输出，连接层则用来进行下采样和特征图的压缩，如式 5 所示。

$$y_l = HI([y_0, y_1, \dots, y_{l-1}]) \quad (5)$$

式 5 中， y_l 表示第 l 的输出， HI 表示第 l 层的非线性转换函数， $[y_0, y_1, \dots, y_{l-1}]$ 表示前 $l-1$ 层所有特征的拼接，即密集连接的输入。Densenet 也有一些缺点，例如内存消耗较大、计算复杂度高等。在此基础上，log2n-link 复杂度由 $o(l^2)$ 下降至 $o(l \log 2l)$ ，在反向传播的过程中，Log2N-Link 使用了更长的反向传播路径，从而将最短距离从 1 提升到了 $1 + \log 2N$ 。

$$y_l = HI([y^{l-2^n}, y^{l-2^{n-1}}, \dots, y^{l-2^0}]) \quad (6)$$

因此，在同等的计算量限制下，Log2N-Link 可以更深度地扩展 GFPN^[21]，并保留有效的特征复用。

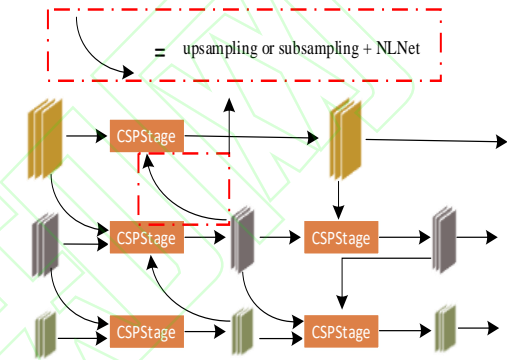


图 5 GFPN 结构示意图

此外，为了增加更多的特征融合，每一层不仅仅接收上一个节点的输入，Feature Fusion Block 还接收来自斜上方和斜下方节点的输入，如图 5 所示。同时，为了避免使用 sum 操作带来的信息丢失，Feature Fusion Block 采用了 concat 方式进行特征融合，如图 6 所示。

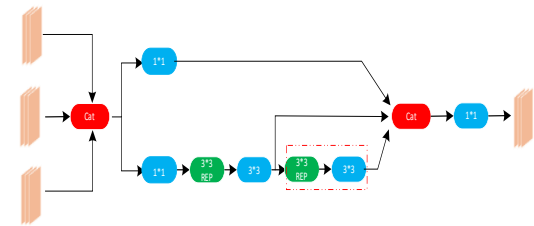


图 6 CSP Stage

在 GFPN 中，特征融合的过程是通过上采样和下采样来实现的，采样操作是一种粗略的下采样过程，会导致特征信息的损失和空间信息的模糊。引入 NLnet^[22] 可以更好地保留重要的特征信息并且减少不必要的信息损失。在每个层级中，将上一层级的特征进行上采样，与下一层级的特征进行融合，

得到一个新的特征。将 NLnet 添加到 GFPN 中的采样层中以实现 NLnet 的特征增强。在上采样层中，将上采样后的特征和下一层的特征输入到 NLnet 中，通过 NLnet 得到特征的权重，然后将这些权重应用到上采样的特征上，得到一个新的特征，再与下一层的特征进行融合。在下采样层中，将下采样的特征和上一层的特征输入到 NLnet 中，通过 NLnet 得到特征的权重，然后将这些权重应用到下采样的特征上，得到一个新的特征，再与上一层的特征进行融合。GFPN based on NLnet 可以对特征进行更加全面和深入的分析，得到更加丰富和具有代表性的特征信息。这样可以提高目标检测算法的检测精度和性能。

3.3 soft-NMS

非极大值抑制 (NMS) 是目标检测中常用的一种技术，它用于剔除重复的边界框，以保留最可能包含目标的边界框。IOU 阈值作为硬阈值处理会造成边界框的信息丢失。在传统的 NMS 方法中，如果一个边界框与另一个边界框的 IOU 大于阈值，则会被丢弃，而其他的边界框则会保留。

1. 对所有边界框按照置信度 (confidence) 从大到小排序。
2. 选择置信度最高的边界框，并将其保留。
3. 遍历剩余的边界框，如果与已选择的边界框的交并比 (IOU) 大于等于一定阈值 (通常取 0.5)，则将其舍弃。
4. 重复步骤 2 和 3，直到所有的边界框都被处理完毕。

传统的 NMS 算法通常会选择分数最高的框作为基准框，然后根据重叠面积来抑制其他与之重叠度高的框，保留重叠度较低的框。这种方法虽然简单有效，但是存在以下几个缺点：

1. 可能会漏掉重要目标：如果存在多个物体重叠在一起，传统 NMS 只会保留分数最高的物体，可能会漏掉其他物体，尤其是小目标。
2. 非极大值抑制过程中的重复计算：传统

NMS 算法中，对于每个检测框都要计算其与其他检测框之间的重叠程度，因此计算量较大，速度较慢。

3. 对于密集目标的处理不理想：如果场景中存在密集的目标，传统 NMS 算法可能会误判一些框，将其判定为重复框并抑制掉。

使用权重函数可以避免这种情况发生，因为它可以降低重叠的边界框的得分，而不是完全剔除。Soft-NMS 的算法，它与传统 NMS 的不同之处在于，它在计算重叠程度时不是简单的二值化阈值，而是使用一种类似高斯函数的方式，将重叠程度转化为一个在 0-1 之间的实数，然后根据这个实数对所有框进行排序。在抑制过程中，不再直接抑制与基准框重叠度高的框，而是根据重叠程度计算一个权重，对所有框进行加权求和，最终得到加权和最高的框。这种方法不仅能够更好地处理密集目标，还能够减少重复计算，提高速度。如式 7 所示。

$$S_i^{soft} = (1 - \max_{j \in S \setminus i} IoU_{i,j}) S_i \quad (7)$$

式 7 中， s_i 是第 i 个边界框的置信度， S 是已选择的边界框集合， $IoU_{i,j}$ 是第 i 和第 j 个边界框的交并比， $\max_{j \in S \setminus i} IoU_{i,j}$ 是 i 与

所有已选择的边界框的最大交并比。当检测到多个重叠的边界框时，soft-NMS 会为每个边界框计算一个置信度，根据置信度对边界框进行排序，并在排名较低的边界框上应用一个加权函数，降低它们的置信度。这样，即使一个框与其他框有重叠，它也可能被保留。因此，soft-NMS 相对于传统 NMS 更具有鲁棒性和灵活性。具体来说，soft-NMS 可以分为以下三个步骤：

1. 检测边界框：使用目标检测算法检测图像中的边界框，并为每个边界框分配一个得分。
2. 带权重排序：根据每个边界框的得分进行排序，得分高的边界框排在前面。然后，对于排名较低的边界框，应用一个权重函数来降低它们的得分。这个权重函数可以是高斯函数或线性函数等。
3. 重新排名：对调整后的得分进行重新排名，然后使用 NMS 方法选择得分最高的边界框。

soft-NMS 的优点不仅在于其鲁棒性和灵活性,而且在于它可以提高检测的准确性。使用 soft-NMS 可以防止过多的边界框被剔除,从而提高了检测结果的召回率。此外,soft-NMS 还可以减少检测结果中的重复框数目,使得最终输出的结果更加简洁。

4 实验结果与分析

4.1 数据集与预处理

实验所用飞机数据集共计 2000 张,尺寸



图 7 数据增强示例

4.2 训练参数设置

YOLOv5 模型从小到大分为 s、m、l、x、nano,其在初始阶段均对输入数据进行 resize (640*640) 操作,为适应飞机数据集的更大尺寸,在 YOLOv5 的基础上增加了更大的下采样倍率 (64 倍),分别在 23、26、29、32 层使用四组不同尺度的 anchors 进行预测任务。初始学习率为 0.033,使用余弦退火动态降低学习率,余弦退火超参数为 0.15184,学习率动量为 0.74747。训练的 epochs 为 70, batch-size 为 32,损失函数包含了三个部分:矩形框回归预测部分(Reg)、置信度预测部分 (Obj) 和分类预测部分 (Cls)。其中,Reg 部分采用 CIOU 损失,Obj 部分和 Cls 部分则采用 BCE Loss (二元交叉熵损失) 来计算。CIOU 损失是一种特殊的距离度量损失函数,能够减小预测框与真实框之间的重合度误差。BCE Loss 则是在分类问题中常用的损失函数,可以用于计算模型对每个类别的预测值与真实类别之间的差距。通过在这三个部分的损失函数进行加权求和,YOLOv5 模型可以对目标框的

尺寸大小为 1024*1024*3,使用 Labelimg 工具为您的图片创建了 pascal VOC 格式的标签,标签以 XML 格式存储,训练集和验证集划分比例为 8: 2。为了增强模型的鲁棒性和泛化能力,采用了 Mosaic 数据增强技术来处理原始数据,如图 7 所示。这项技术可以对图像进行随机缩放、翻转、裁剪等多种处理方式,从而为模型提供更多的样本数据,增加其对于不同场景的适应性。

位置、置信度和类别进行全面的预测和优化。在同一环境下训练,改进前后 YOLOv5 网络的损失曲线如图 8 所示。

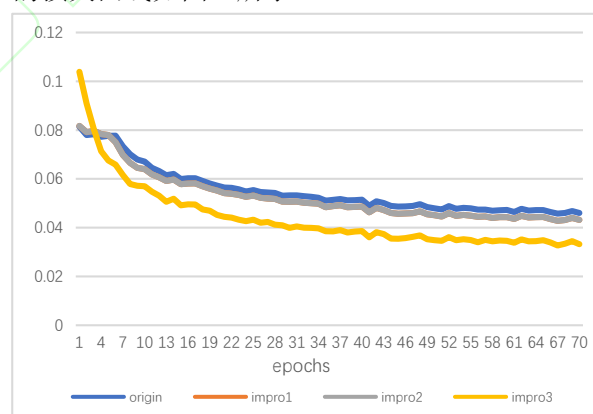


图 8 改进前后 YOLOv5 的训练 Loss 曲线

4.3 实验结果

为了验证算法模型的改进效果,我们使用了前文介绍的数据集以及相同的训练参数,对原始算法模型和改进后的模型进行了训练和测试,并对比它们的性能表现。根据 Precision (精确率)、Recall (召回率)、mAP_0.5 (平均精度, IOU 阈值为 0.5)、mAP_0.5:0.95 (平均精度, IOU 阈值在 0.5 到 0.95 之间) 等四项指标,我们可以发现,改

进算法模型相比于原始模型都有不同程度的提升。这表明，我们所采用的改进方法在

提高模型性能方面是有效的，可以更好地适应实际场景的需求，如图 9 所示。

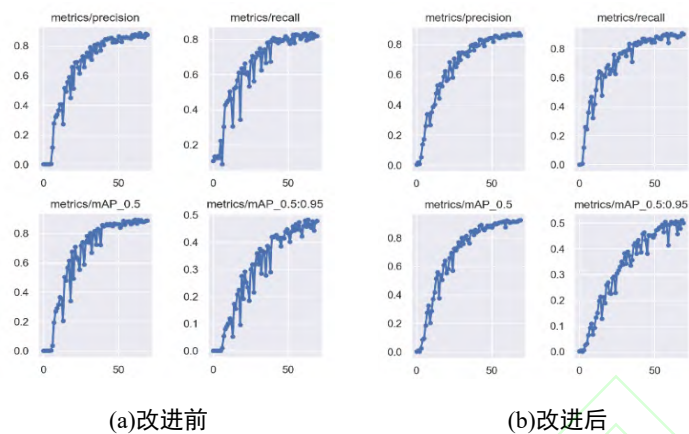


图 9 YOLOv5 改进前后指标对比

为了评估 YOLO 算法在飞机目标检测任务中的性能，我们在相同数据集以及参数设置的基础上进行了相关实验，并使用常见的性能指标进行了评估。实验结果表明，YOLO 算法在飞机目标检测任务中表现出了较高的准确性和鲁棒性。具体来说，采样相同的数据划分比例以及超参数设置，分别对 origin、impro1、impro2、impro3 进行的 epochs 为 70 的模型训练，在训练结束后使用同一验证数据集评估模型的性能，选取精度（Precision，P）、召回率（Recall，R）、平均精度（mAP0.5）以及 mAP0.5:0.95 对模型进行全方位的评价，如表 1 所示。

表 1 消融实验

Model	P	R	mAP_0.5	mAP_0.5:0.95
origin	0.863	0.821	0.891	0.484
impro1	0.844	0.837	0.893	0.473
impro2	0.845	0.837	0.895	0.503
impro3	0.875	0.906	0.923	0.512

为了展示 YOLO 算法在飞机目标检测任务上的表现，我们提供了一些效果图，如图 10 所示。这些图像显示了 YOLO 算法在不同场景下准确检测飞机的能力，例如在军事基地等环境中。同时，这些图像还突出了 YOLO 算法的实时性和鲁棒性，即使在飞机遮挡、遥远视距和复杂背景等情况下，算法仍能够准确地检测目标。

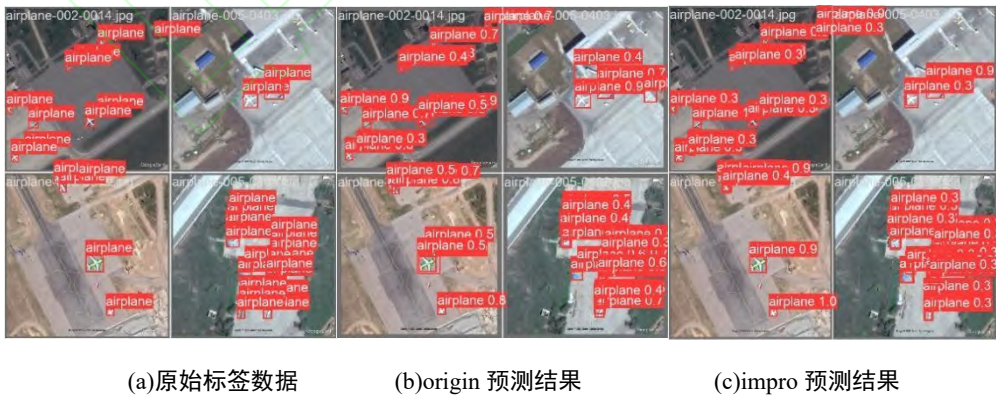


图 10 YOLOv5 改进前后检测结果对比

图 10（a）为数据原始标签展示，对数据集样本进行了准确定位与分类。通过 10（a）（b）两幅图像观察可以看到在左下角和右上角 origin 模型发生了不同程度的“误检”的

错误识别，在 impro 模型中这种错误被纠正，而且精度显著优于前者，通过 impro 预测结果也成功排除了原始数据标签的错误标注的原因。

5 结论

在本研究中,我们针对传统飞机目标识别在抗干扰性、对复杂场景的需求等方面存在的问题,提出了一种基于改进 YOLOv5 的飞机目标检测算法,包括 IOU-NWD Similarity Metric for Bounding Boxes 策略、GFPN based on NLnet 模块和 soft-NMS 方法等三个改进点。

实验结果表明,我们改进后的模型在 Precision、Recall、mAP0.5、mAP0.5:0.95 等指标上都有所提升,与原始 YOLOv5 相比,改进后的模型分别提高了 1.9%和 10.4%的准确率和召回率。特别地,改进后的模型在 MAP0.5 和 MAP0.5:0.95 评分方面分别为 92.3%和 51.2%,相对原始模型提高了 3.6%和 5.8%。这表明我们提出的算法对于不同 IOU 阈值下的检测性能都有着不错的表现,并且通过针对性的网络调整和模块迁移,能够提高模型对小型和遮挡的飞机目标的检测效果,解决了原始 YOLOv5 算法的小飞机目标误检的问题。

因此,我们的改进算法在飞机目标识别任务中表现出更好的性能,并具有更高的鲁棒性和准确性,有望在实际应用中发挥重要作用。

参考文献 (References)

- [1] A. Moussaoui, S. El Hajjaji, M. Boutaoune, et al. "A Survey of AI Applications for Military Airplane Maintenance and Repair," [J]. IEEE Access, 2021, vol. 9, 61407-61421.
- [2] J. Yang, X. Liu, S. Fan, "A Study of the Aerodynamic Characteristics of a Military Aircraft with Varying Angles of Attack," [J]. Journal of Aerospace Engineering, 2021, vol. 34, no.3, 040 21026.
- [3] X. Liu, H. Wu, Q. Zhu, Y. Zeng, "Research on Stealth Technology of Military Aircraft Based on Machine Learning," [C]. Journal of Physics: Conference Series, 2021, 012032.
- [4] J. Cheng, L. Liu, Y. Wang, "Aircraft Detection in Remote Sensing Images Using Progressive Anchor Boxes," [J]. Remote Sensing, 2021, vol. 13, no. 8, pp. 1447.
- [5] S. Zhang, X. Ma, W. Zhu, C. Zhao, "An Object Detector Based on Feature Pyramid Network for Aircraft Detection in Optical Remote Sensing Images," [J]. Remote Sensing, 2021, vol. 13, no. 8, pp. 1597,
- [6] J. Wang, J. Wang, Y. Yao, L. Liu, "Aircraft recognition in optical remote sensing images based on HOG-SVM," [J]. Journal of Applied Remote Sensing, 2021. vol. 15, no. 2, pp. 026514
- [7] S. Zhang, W. Zhu, C. Zhao, "Aircraft detection in high resolution remote sensing images based on visual attention and SVM," [J]. Journal of the Indian Society of Remote Sensing, 2021. vol. 49, no. 5, pp. 821-831.
- [8] X. Liao, Q. Huang, "An Automatic Target Recognition Method for SAR Images Based on Scattering Characteristic and Morphological Features," [J]. IEEE Access, 2021. vol. 9, pp. 43987-43998.
- [9] Z. Wang, C. Jin, C. Guan, Y. Jiang, "Aircraft target recognition method based on spectral and morphological features," [J]. Journal of Applied Remote Sensing, 2021. vol. 15, no. 1, pp. 016524.
- [10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," [C]. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
- [11] 牛洪超, 胡晓兵, 罗耀俊. 基于 M-YOLO 的自动驾驶下目标识别算法[J]. 计算机工程与设计, 2022, 43(08): 2213-2220.
- [12] 茅智慧, 朱佳利, 吴鑫等. 基于 YOLO 的自动驾驶目标检测研究综述[J]. 计算机工程与应用, 2022, 58(15): 68-77.

- [13] 潘惠苹, 王敏琴, 张福泉. 基于优化 YOLO-V4 的交通标志检测识别方法[J]. 计算机科学, 2022, 49(11): 179-184.
- [14] 梅礼坤, 陈智利, 牛恒. 基于改进 YOLOv5 的坦克目标检测算法[J/OL]. 激光杂志: 1-7[2023-05-14].
- [15] J. Wang, K. Liu, Z. Wu, J. Yuan, and X. Gu, "CSPNet: A new backbone that can enhance learning capability of CNN,"[C]. in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, pp. 3501-3510.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition,"[C]. in European Conference on Computer Vision (ECCV). 2014, pp. 346-361.
- [17] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature Pyramid Networks for Object Detection. [C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 2117-2125.
- [18] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," [C]. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, pp. 8759-8768.
- [19] H. S. Bhatia, R. K. Katiyar, and P. K. Biswas, "Wasserstein Generative Adversarial Networks (WGAN) with Improved Training," [C]. in Proceedings of the International Conference on Pattern Recognition and Machine Intelligence (PReMI). 2020, pp. 307-315.
- [20] Jinwang Wang、Chang Xu、Chang Xu. "A Normalized Gaussian Wasserstein Distance for Tiny Object Detection" Computer Vision and Pattern Recognition (cs.CV).[C]. 2021, arXiv:2110.13389.
- [21] Yiqi Jiang, Zhiyu Tan, Junyan Wang. " GiraffeDet: A Heavy-Neck Paradigm for Object Detection" Computer Vision and Pattern Recognition (cs.CV).[C]. 2022, arXiv:2202.04256.
- [22] Jianfeng Wang, Zhiqiang Shen, Qi Qian. "Selective Kernel Networks (SKNet) European". [C]. Conference on Computer Vision (ECCV) . 2018, 11205, 222-238.