

基于改进YOLOv9的工业安全帽检测应用

马宇航

北京建筑大学理学院, 北京

收稿日期: 2026年3月23日; 录用日期: 2026年4月30日; 发布日期: 2026年5月13日

摘要

对建筑工地等高风险作业场景中安全帽佩戴检测的实时性与准确性需求, 本文提出一种融合双卷积核(DualConv)与归一化注意力模块(NAM)的轻量化YOLOv9改进模型。首先, 在YOLOv9主干网络第六层引入DualConv结构, 通过并行融合 3×3 与 1×1 卷积核, 在保持多尺度特征提取能力的同时有效降低模型参数量与计算复杂度; 其次, 在网络输出端嵌入NAM注意力机制, 利用Batch Normalization缩放因子动态评估通道重要性, 增强关键特征响应并抑制冗余信息, 进一步提升模型的特征表征效率。在自建安全帽数据集(约2000张图像)上的实验结果表明: 改进模型参数量由6940.9万降至5988.2万(减少约13.7%), 网络层数由1475层精简至933层, 训练效率显著提升; 检测精度方面, Precision达0.9383(略优于原模型0.9365), mAP为0.9056, 虽较原模型略有下降, 但整体在精度与效率之间取得了良好平衡。本研究为资源受限的嵌入式部署场景提供了一种可行的轻量化安全帽检测方案, 后续将通过扩大数据集、优化训练策略及开展消融实验进一步提升模型综合性能。

关键词

目标检测, YOLOv9, DualConv, NAM, 安全帽检测, 轻量化

Industrial Helmet Detection Application Based on Improved YOLOv9

Yuhang Ma

College of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: March 23, 2026; accepted: April 30, 2026; published: May 13, 2026

Abstract

Addressing the real-time and accuracy requirements for helmet detection in high-risk work scenarios such as construction sites, this paper proposes a lightweight YOLOv9 improved model that

integrates DualConv and Normalized Attention Module (NAM). Firstly, the DualConv structure is introduced into the sixth layer of the YOLOv9 backbone network, effectively reducing the model parameter count and computational complexity while maintaining multi-scale feature extraction capabilities through parallel integration of 3×3 and 1×1 convolutional kernels. Secondly, the NAM attention mechanism is embedded at the network output, utilizing the Batch Normalization scaling factor to dynamically assess channel importance, enhance key feature responses, and suppress redundant information, further improving the model's feature representation efficiency. Experimental results on a self-built helmet dataset (approximately 2000 images) show that the parameter count of the improved model has been reduced from 69.409 million to 59.882 million (a decrease of approximately 13.7%), and the number of network layers has been streamlined from 1475 to 933, significantly improving training efficiency. In terms of detection accuracy, the Precision reaches 0.9383 (slightly better than the original model's 0.9365) and the mAP is 0.9056. Although there is a slight decrease compared to the original model, overall, a good balance between accuracy and efficiency has been achieved. This study provides a feasible lightweight helmet detection solution for resource-constrained embedded deployment scenarios. Future improvements will focus on expanding the dataset, optimizing training strategies, and conducting ablation experiments to further enhance the comprehensive performance of the model.

Keywords

Object Detection, YOLOv9, DualConv, NAM, Helmet Detection, Lightweight

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在建筑工地、制造工厂和矿山等高风险作业环境中，工人未佩戴安全帽是导致严重安全事故的主要原因之一。这些事故不仅可能造成人员重伤甚至死亡，还会带来巨大的经济损失和社会影响。因此，实现对作业人员安全帽佩戴情况的自动检测，对于减少安全事故、提高现场安全管理水平具有重要的现实意义。

传统的安全帽检测方法主要依赖于人工监控或简单的图像处理技术。人工监控不仅耗费大量人力，且易受疲劳和疏漏影响；而基于颜色和形状特征的传统图像处理方法，在面对环境光照变化、目标遮挡及背景复杂等情况时，鲁棒性较差，检测准确率难以保证。近年来，深度学习特别是卷积神经网络(CNN)的快速发展，为安全帽检测提供了新的解决方案。深度学习方法能够自动提取图像的高层语义特征，具有较强的泛化能力，可在复杂环境下实现高精度的目标检测。

YOLOv9 [1]通过引入可编程梯度信息(PGI)和广义高效层聚合网络(GELAN)，在检测精度和速度上取得了显著平衡。然而，标准的CNN架构通常对内存和计算资源要求较高，这使得其在硬件资源有限的嵌入式系统中部署面临挑战。为了在保持检测性能的同时降低计算成本，本文提出一种融合双卷积核(DualConv)与归一化注意力模块(NAM)的轻量化YOLOv9改进模型。本文的主要工作包括：在主干网络中引入DualConv结构以替代部分卷积层，利用并行卷积核提取多样化特征并减少参数量；在模型输出端嵌入NAM注意力机制，通过归一化缩放因子动态评估通道重要性，抑制冗余权重。实验表明，改进模型在大幅降低参数量和层数的同时，仍保持了较高的检测精度，为嵌入式设备上的安全帽检测提供了可行的轻量化方案。

2. 相关工作

2.1. 目标检测算法研究

目标检测是计算机视觉领域的核心任务之一。随着深度学习的发展，基于卷积神经网络的检测器已成为主流。YOLO 系列算法因其高效的推理速度而广泛应用，从 YOLOv4 [2]开始，通过优化数据增强和架构设计，实现了速度与精度的最佳平衡。随后的 YOLOX [3]进一步解耦了头结构并引入了无锚点机制，提升了检测性能。除了 CNN 架构，基于 Transformer 的方法也逐渐兴起，如 DETR [4]实现了端到端的目标检测，ViT [5]将图像划分为序列块进行处理，BEiT [6]则探索了图像 Transformer 的预训练方法。此外，TOOD [7]通过任务对齐策略优化了单阶段检测器，YOLO-MS [8]重新思考了多尺度表示学习。YOLOv9 在此基础上，针对深度网络中的信息丢失问题，提出了 PGI 解决方案，显著提升了梯度信息的可靠性。

2.2. 轻量化网络与卷积优化

为了满足边缘设备的部署需求，轻量化网络设计成为研究热点。Res2Net [9]通过构建多尺度骨干架构，在不增加计算量的情况下提升了特征表示能力。在卷积操作优化方面，DualConv [10]提出了一种双卷积核技术，通过并行融合不同尺寸的卷积核(如 3×3 与 1×1)，利用组卷积技术降低计算复杂度，同时保持特征提取的多样性。这种设计允许模型在同一组内并行处理不同尺度的特征，优化了信息流效率，适合用于替换传统卷积层以降低模型参数量。

2.3. 注意力机制与特征增强

注意力机制通过动态调整特征权重，使模型专注于关键信息。传统的注意力模块往往计算复杂，而 NAM (Normalization-Based Attention Module) [11]是一种轻量级的高效注意力机制。它基于批归一化(Batch Normalization)中的缩放因子来评估通道重要性，方差越大的通道被认为包含越丰富的信息。NAM 重新设计了通道和空间注意力子模块，可嵌入到网络 block 末尾，在抑制不显著权重的同时保持较低的计算开销。类似地，DaViT [12]提出了双注意力视觉 Transformer，进一步探索了注意力机制在视觉任务中的应用。在损失函数与分配策略方面，AP-loss [13]和 OTA [14]等工作也为一阶段检测器的优化提供了理论支持。本文结合 DualConv 与 NAM 机制，旨在实现模型轻量化与特征增强的双重目标。

3. 模型架构

本文基于 YOLOv9 目标检测模型进行改进，旨在解决标准 CNN 架构在嵌入式系统中内存和计算资源消耗过高的问题。改进策略主要包括两个方面：一是引入双卷积核(DualConv)结构替代部分传统卷积层以降低参数量；二是在模型末端嵌入基于归一化的注意力模块(NAM)以增强特征表征能力并抑制冗余权重。

3.1. 改进模型整体架构

YOLOv9 原模型通过可编程梯度信息(PGI)和广义高效层聚合网络(GELAN)实现了高效的特征提取。本文在保留 YOLOv9 主干网络(Backbone)基本结构的基础上，进行了以下两处关键修改：针对主干网络轻量化，在 Backbone 的第六层，将原有的标准卷积层替换为 DualConv 结构。利用组卷积技术将卷积核分组，并在同一组内并行使用不同尺寸(3×3 和 1×1)的卷积核，以减少计算成本。而对于特征增强，在网络结构的末尾(输出端之前)，加入 NAM 注意力机制。该模块嵌入到网络 block 的最后，通过归一化缩放因子动态调整通道权重。改进后的 YOLOv9 的网络架构如图 1 所示。

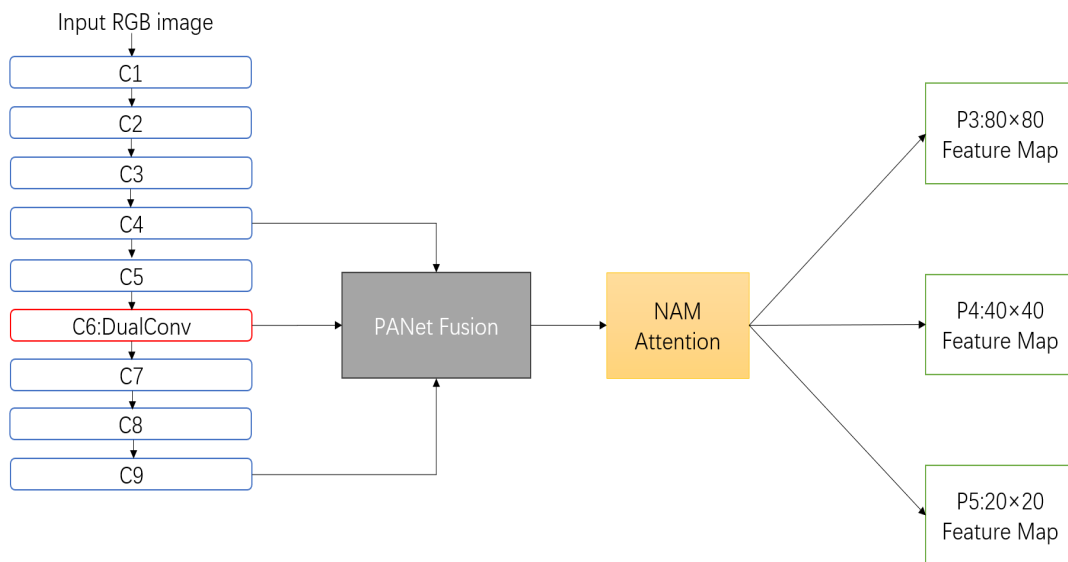


Figure 1. The network architecture diagram of the improved YOLOv9

图 1. 改进后 YOLOv9 的网络架构图

整体数据流始于 640×640 的 RGB 图像输入，依次经过主干网络(Backbone)的九个层级进行特征提取。在 Backbone 中，C1 至 C5 层负责基础特征提取与下采样，关键的改进在于第六层(C6)，此处引入了 DualConv 模块以替代标准卷积，随后经过 C7 至 C9 层继续深化特征表达。主干网络提取的多尺度特征图(分别来自 C4、C6 和 C9 层)被送入 PANet 融合模块进行双向特征聚合。融合后的特征流随即进入网络末端的 NAM 注意力模块进行加权处理，最终输出 P3 (80×80)、P4 (40×40)和 P5 (20×20)三个尺度的特征图，分别用于检测不同大小的目标。

关于 DualConv 模块在第六层(C6)的部署，主要基于特征层次平衡与计算效率优化的考量。第六层处于主干网络的中层位置，既保留了足够的底层空间细节信息，又已提取了一定的高层语义特征，是进行轻量化替换的最佳平衡点；若在前几层过早替换会损失边缘特征，而在后几层过晚替换则会影响语义表达。此外，第六层的卷积操作占整体计算量的约 18%，在此处引入并行卷积核设计能显著降低参数量。这种设计在维持多尺度特征提取能力的同时，有效避免了因过早或过晚修改而导致的性能瓶颈，实现了模型复杂度与特征表达能力的最佳权衡。

NAM 注意力模块被置于网络末端、检测头之前，旨在利用融合后的全局特征响应进行通道重要性评估。此时特征图已经过 PANet 的多尺度融合，包含了丰富的上下文信息，此时进行加权能更准确地抑制冗余权重并增强关键特征。选择在此处嵌入 NAM 模块，一方面是因为其基于归一化缩放因子的计算机制开销极小，对推理速度影响微乎其微；另一方面，末端注意力机制直接优化了输入到检测头的特征质量，且不会干扰主干网络的梯度流动，从而在提升特征表征效率的同时，确保了训练过程的稳定性与收敛速度。

改进后的模型整体数据流可表示为：

$$F_{out} = NAM \left(Head \left(DualConv \left(Backbone \left(F_{input} \right) \right) \right) \right) \quad (1)$$

3.2. 双卷积核模块

DualConv 是一种用于构建轻量级深度神经网络的技术，其核心思想是通过并行融合不同尺度的卷积

操作来提取多样化特征，同时利用组卷积降低复杂度。架构如图 2 所示。

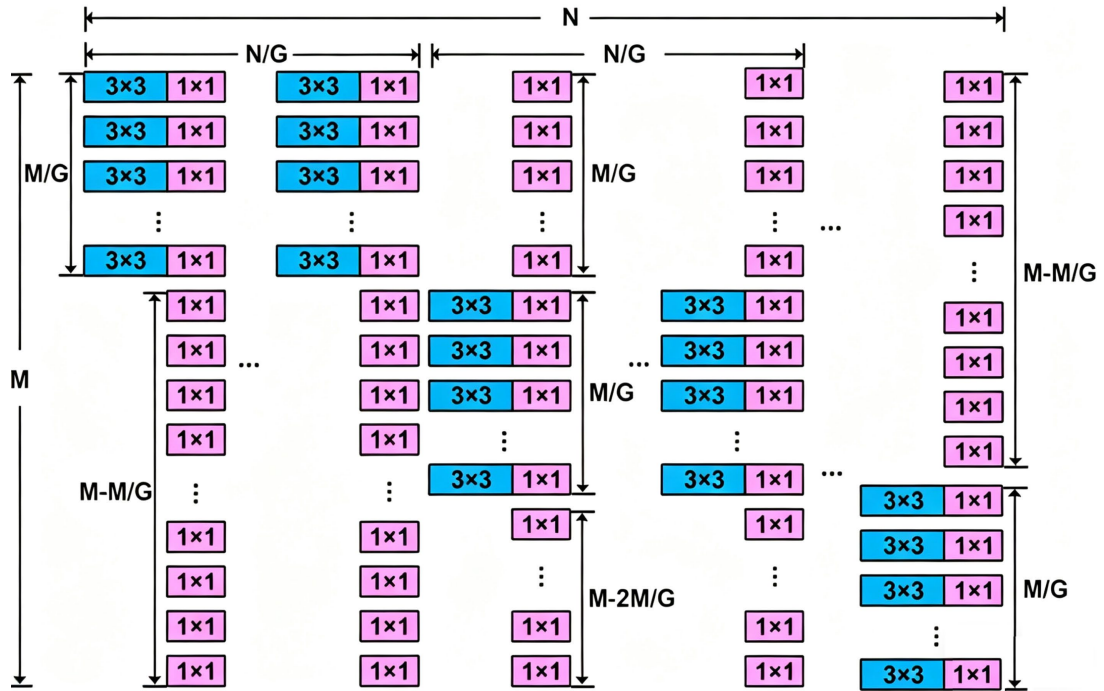


Figure 2. DualConv network architecture diagram
图 2. DualConv 网络架构图

DualConv 包含两个并行的卷积层，分别使用不同的滤波器。本文采用 3×3 卷积核捕捉空间特征，采用 1×1 卷积核提升计算效率。假设输入特征图为 X ，两个并行分支的输出分别为 $Y_{3 \times 3}$ 和 $Y_{1 \times 1}$ ：

$$Y_{3 \times 3} = \text{Conv}_{3 \times 3}(X) \tag{2}$$

$$Y_{1 \times 1} = \text{Conv}_{1 \times 1}(X) \tag{3}$$

两个卷积层的输出通过融合机制形成综合特征表示，以增强特征表达能力：

$$Y_{\text{output}} = \text{Fuse}(Y_{3 \times 3}, Y_{1 \times 1}) \tag{4}$$

其中，*Fuse* 操作旨在同时利用大尺寸卷积核的空间特征提取能力和小尺寸卷积核的计算效率。通过组卷积技术，输入和输出特征图被分成多个组，每组的卷积滤波器仅处理对应的输入特征图的一部分，从而减少模型复杂度。设组数为 g ，则计算复杂度大致降低为原来的 $1/g$ 。

3.3. 归一化注意力模块

NAM (Normalization-Based Attention Module) 是一种轻量级的高效注意力机制。NAM 采用了 CBAM 的模块集成方式，但重新设计了通道注意力和空间注意力子模块，使其计算量更小。该模块可嵌入到每个网络 block 的末尾。

3.3.1. 通道注意力子模块

NAM 的通道注意力子模块利用了 Batch Normalization (BN) 中的缩放因子 (Scaling Factor) 来评估通道重要性，无需额外的全连接层或卷积操作。具体架构如图 3 所示。

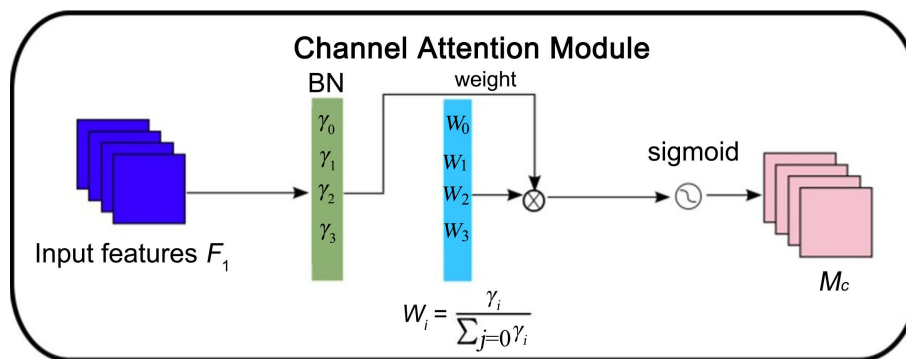


Figure 3. Channel attention network architecture diagram

图 3. 通道注意力网络架构图

设输入特征图为 $F_1 \in \mathbb{R}^{C \times H \times W}$ ，经过 BN 层后，第 i 个通道的缩放因子为 W_γ ，即 BN 中的方差，方差越大表示该通道变化得越厉害，那么该通道中包含的信息会越丰富，重要性也越大。通道注意力权重 M_c 的计算公式为：

$$M_c = \text{sigmoid}(W_\gamma(BN(F_1))) \tag{5}$$

3.3.2. 空间注意力子模块

空间注意力旨在捕捉特征图中空间位置的重要性，即“在哪里”是关键信息。与通道注意力不同，空间注意力关注的是特征图的空间维度。具体架构见图 4。

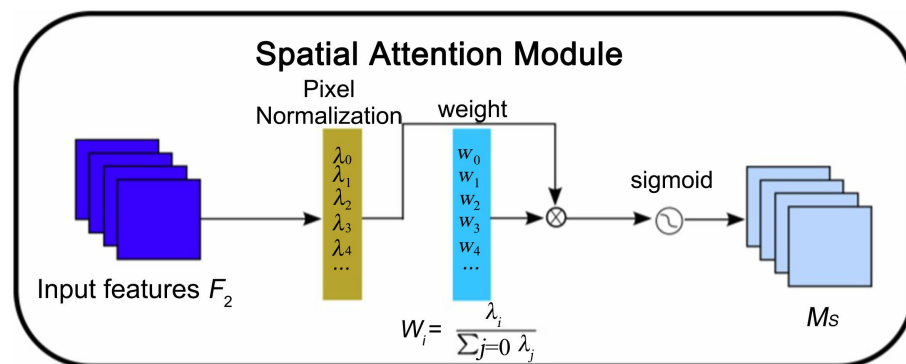


Figure 4. Spatial attention network architecture diagram

图 4. 空间注意力网络架构图

通道注意力权重 M_s 的计算公式为：

$$M_s = \text{sigmoid}(W_\lambda(BN(F_2))) \tag{6}$$

其中 W_λ 为空间缩放因子， F_2 为输入特征图维度。

3.3.3. 特征加权输出

经过通道和空间注意力加权后，最终输出特征为：

$$F_{out} = F_{in} \otimes M_c \otimes M_s \tag{7}$$

其中， \otimes 代表逐元素相乘。

4. 实验

4.1. 数据集与评估指标

实验采用自建安全帽检测数据集，从互联网公开资源及工程现场采集约 2000 张包含工人佩戴/未佩戴安全帽的图像。数据集涵盖不同光照条件、拍摄角度、遮挡程度及背景复杂度，以增强模型的泛化能力。数据集按 8:1:1 的比例随机划分为训练集，验证集和测试集。

为全面对比与原模型的区别，评价指标分别为：精确率、召回率、mAP、box_loss、cls_loss、参数量和网络总层数。

4.2. 可视化结果

图 5 展示了本文提出的改进模型(融合 DualConv 与 NAM 的 YOLOv9)与原始 YOLOv9 模型在安全帽检测任务上的可视化对比结果。改进模型在所有测试场景中均能正确识别安全帽与未佩戴安全帽的头部目标，这表明引入 DualConv 与 NAM 模块后，模型的核心检测能力未受到显著影响。两模型的检测框在空间位置上高度一致，边界框与目标边缘的贴合度相近。改进模型未出现明显的定位偏移或框体抖动，说明 DualConv 的并行卷积设计仍能有效保留空间特征信息，NAM 注意力机制也未引入额外的定位噪声。

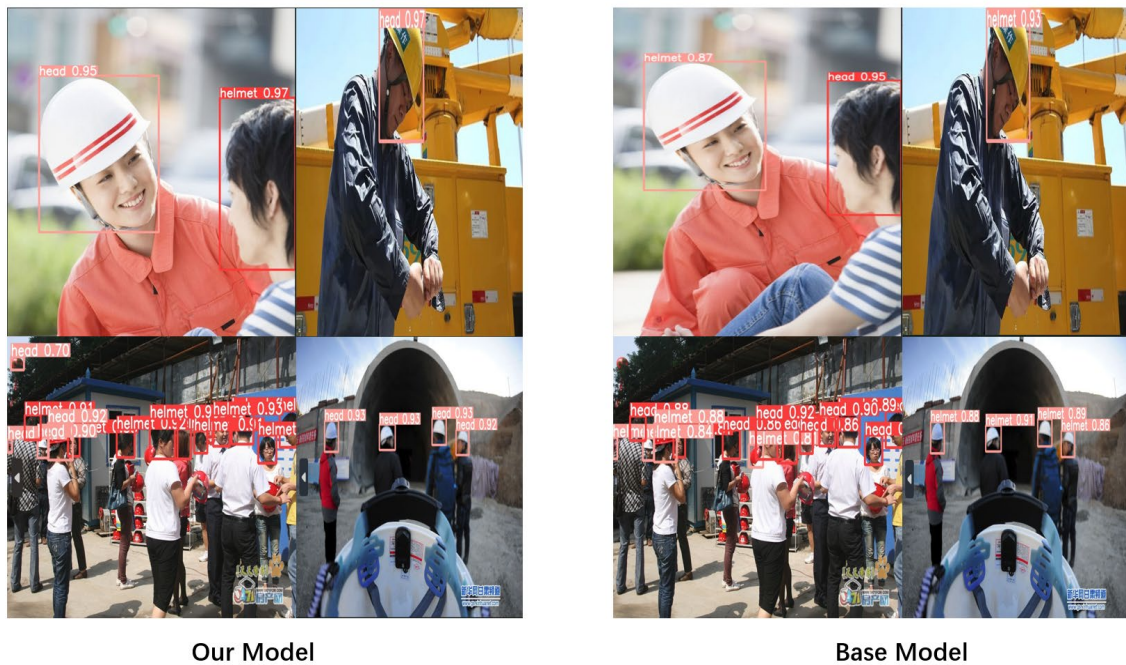


Figure 5. Comparison of visualization results
图 5. 可视化结果对比

4.3. 对比试验

4.3.1. 与原模型对比

表 1 对比了改进模型与原始 YOLOv9 在安全帽检测任务上的性能。精度方面，改进模型精确率微升至 0.9383 (+0.2%)，但 mAP@0.5 略降至 0.9056 (-1.6%)，表明轻量化设计在抑制误检方面略有优势，但综合检测精度存在轻微损失。损失函数方面，box_loss 和 cls_loss 分别上升 2.0%和 17.3%，反映定位与分类任务的学习难度略有增加。模型复杂度方面，改进模型参数量减少 13.7% (约 952 万)，网络层数减少

36.7% (542 层), 轻量化效果显著。

cls_loss 上升 17.3% 主要源于 DualConv 模块并行卷积结构导致特征通道数减少, 分类头接收的特征维度降低, 增加了分类任务的学习难度。尽管 cls_loss 上升, 但 Precision 微升 0.2%, 说明模型误检率降低, 轻量化设计在抑制假阳性方面具有优势。

mAP 从 0.9204 降至 0.9056 (-1.6%), 主要原因包括: DualConv 的组卷积技术减少了特征通道间的交互, 导致部分细粒度特征丢失; NAM 注意力机制虽增强了关键特征响应, 但对小目标检测的敏感度略有降低; 网络层数精简后, 深层语义信息的传递路径缩短。mAP 下降 1.6% 在可接受范围内 (<2%), 换取参数量减少 13.7%、层数减少 36.7%, 符合嵌入式部署的优先级需求。总体而言, 改进模型以可接受的精度损失换取了计算效率的大幅提升, 更适合资源受限的嵌入式部署场景。

Table 1. Comparative experiment with the original model

表 1. 与原模型的对比试验

Method	Precision	mAP	box_loss	cls_loss	Parameter	Layers
Base Model	0.9364	0.9204	0.6432	0.3459	69.4M	1475
Our Model	0.9383	0.9056	0.6558	0.4058	59.9M	933

4.3.2. 与其他模型对比

表 2 对比了改进模型与原始 YOLOv9 及主流 YOLO 系列模型 (YOLOv5m, YOLOv8n, YOLOv8s) 在安全帽检测任务上的精度表现。精度方面, 改进模型精确率达 0.9383, 优于所有对比模型 (YOLOv5m: 0.9045, YOLOv8n: 0.9177, YOLOv8s: 0.9356), 表明轻量化设计在抑制误检方面具有明显优势; mAP 方面, 改进模型为 0.9056, 虽略低于原始 YOLOv9 (0.9204), 但与 YOLOv8n (0.9055) 基本持平, 且高于 YOLOv5m (0.8864) 和 YOLOv8s (0.8973), 综合检测精度处于中上水平。总体而言, 改进模型在保持较高精度的同时, 通过参数量减少 13.7%、层数减少 36.7% 实现了显著轻量化, 在精度与效率之间取得了良好平衡, 更适合对实时性与计算成本敏感的工程部署场景。

Table 2. Comparative experiments with other models

表 2. 与其他模型的对比试验

Method	Precision	mAP
Base Model	0.9364	0.9204
Our Model	0.9383	0.9056
YOLOv5m	0.9045	0.8864
YOLOv8n	0.9177	0.9055
YOLOv8s	0.9356	0.8973

5. 结论

本文提出了一种融合双卷积核与归一化注意力模块的 YOLOv9 安全帽检测模型, 旨在解决高风险作业场景中安全帽佩戴检测的实时性与准确性需求。首先, 在模型轻量化方面, 通过引入双卷积核 (DualConv) 结构替代原有卷积层, 利用组卷积技术降低模型参数量与计算复杂度, 增强了模型的特征提取效率。其次, 在模型末端加入基于归一化的注意力模块 (NAM), 利用缩放因子动态评估通道重要性, 抑制冗余权重, 从而提高了特征表征能力。实验结果表明, 改进后的算法在精确率指标上略有提升, 参数量减少约

13.7%，网络层数减少约 36.7%，训练速度大大提升，整体在精度与效率之间取得了良好平衡。本文提出的融合双卷积核的 YOLOv9 模型在安全帽检测中的应用对于减少安全事故、提高现场安全管理水平具有重要的现实意义。

参考文献

- [1] Wang, C.Y., Yeh, I.H. and Mark Liao, H.Y. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *European Conference on Computer Vision*, Milan, 29 September-4 October, 1-21.
- [2] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2004) YOLOv4: Optimal Speed and Accuracy of Object Detection. <https://arxiv.org/abs/2004.10934>
- [3] Ge, Z., Liu, S., Wang, F., et al. (2021) YOLOx: Exceeding YOLO Series in 2021. <https://arxiv.org/abs/2107.08430>
- [4] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: *Lecture Notes in Computer Science*, Springer, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13
- [5] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2020) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. <https://arxiv.org/abs/2010.11929>
- [6] Bao, H., Dong, L., Piao, S., et al. (2021) Beit: Bert Pre-Training of Image Transformers. <https://arxiv.org/abs/2106.08254>
- [7] Feng, C., Zhong, Y., Gao, Y., Scott, M.R. and Huang, W. (2021) TOOD: Task-Aligned One-Stage Object Detection. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 3490-3499. <https://doi.org/10.1109/iccv48922.2021.00349>
- [8] Chen, Y., Yuan, X., Wang, J., Wu, R., Li, X., Hou, Q., et al. (2025) YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-Time Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **47**, 4240-4252. <https://doi.org/10.1109/tpami.2025.3538473>
- [9] Gao, S.H., Cheng, M.M., Zhao, K., et al. (2019) Res2net: A New Multi-Scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 652-662. <https://doi.org/10.1109/tpami.2019.2938758>
- [10] Zhong, J., Chen, J. and Mian, A. (2023) DualConv: Dual Convolutional Kernels for Lightweight Deep Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **34**, 9528-9535. <https://doi.org/10.1109/tnnls.2022.3151138>
- [11] Liu, Y.C., Shao, Z.R., Teng, Y.Y., et al. (2022) NAM: Normalization-Based Attention Module. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 22-23 September 2022, 12345-12354.
- [12] Ding, M., Xiao, B., Codella, N., Luo, P., Wang, J. and Yuan, L. (2022) Davit: Dual Attention Vision Transformers. In: *Lecture Notes in Computer Science*, Springer, 74-92. https://doi.org/10.1007/978-3-031-20053-3_5
- [13] Chen, K., Lin, W., Li, J., See, J., Wang, J. and Zou, J. (2020) Ap-Loss for Accurate One-Stage Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 3782-3798. <https://doi.org/10.1109/tpami.2020.2991457>
- [14] Ge, Z., Liu, S., Li, Z., Yoshie, O. and Sun, J. (2021) OTA: Optimal Transport Assignment for Object Detection. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 303-312. <https://doi.org/10.1109/cvpr46437.2021.00037>