E-Commerce Letters 电子商务评论, 2025, 14(10), 2045-2053
Published Online October 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.14103364

Hans汉斯

基于混合注意力机制的YOLOv5模型及其在电商透明商品检测中的应用

张　伟

贵州大学大数据与信息工程学院，贵州　贵阳

收稿日期：2025年9月13日；录用日期：2025年9月26日；发布日期：2025年10月28日

摘　要

随着电子商务的发展，面向透明商品(比如玻璃瓶、塑料包装化妆品)的检测识别变得尤为重要。由于其特殊光学特性，传统目标检测模型YOLOv5在此类物体上常出现特征提取困难、定位不准和漏检率高等问题，严重制约了电商运营效率与用户体验。为解决这一难题，本研究提出了一种改进的YOLOv5目标检测模型。本文在YOLOv5s的主干与颈部网络结合一种混合的轻量级注意力机制：CLEAR-Attn (Channel-Linear External-SimAM Attention with Residuals)，该注意力机制针对透明的电商物品进行改进，提升了针对电商物品检测的稳健性。该模型融合了外部注意力机制(External Attention, EA)与SimAM注意力机制，EA注意力机制能够以两层线性映射与双重归一化实现外部记忆建模，捕获长程依赖。在其输出后串接SimAM以强化边界与高光区域的响应，最后通过残差连接输出最后的特征，从而能够实现对透明电商商品的有效检测。在Trans10K透明商品数据集上进行实证，采用与基线一致的训练策略。结果显示，在增加少量参数情况下，本文模型在整体精度上取得稳定提升5个百分点，检测精度达到了96%。在进一步的应用分析表明，该技术可广泛应用于电商平台的图像搜索、智能货柜的商品识别、自动化仓库的包裹分拣及库存盘点等具体场景，有效提升识别准确率与作业自动化水平。

关键词

目标检测，YOLOv5，透明商品，电子商务，计算机视觉，电子商务部署

A YOLOv5 Model with Hybrid Attention Mechanism for Transparent Product Detection in E-Commerce

Wei Zhang

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: September 13, 2025; accepted: September 26, 2025; published: October 28, 2025

文章引用: 张伟. 基于混合注意力机制的 YOLOv5 模型及其在电商透明商品检测中的应用[J]. 电子商务评论, 2025, 14(10): 2045-2053. DOI: 10.12677/ecl.2025.14103364

Abstract

With the growth of e-commerce, detecting transparent or highly reflective products (e.g., glass bottles and plastic-wrapped cosmetics) has become increasingly important. Due to their optical properties, conventional detectors such as YOLOv5 often struggle with feature extraction, precise localization, and miss rates on such objects, which constrains operational efficiency and user experience. To address this, we propose an improved YOLOv5 model. Built on YOLOv5s, we integrate a lightweight hybrid attention mechanism, CLEAR-Attn (Channel-Linear External-SimAM Attention with Residuals), into both the backbone and neck to enhance robustness for transparent product detection. CLEAR-Attn fuses External Attention (EA) and SimAM: EA models long-range dependencies via two linear projections with double normalization to form an external memory, while SimAM—placed after EA—amplifies responses along object boundaries and specular highlights. A residual connection is then applied to stabilize training and preserve information. We validate the approach on a Trans10K transparent-product subset under training protocols aligned with the baseline. Results show that, with only a small increase in parameters, our method yields a ~5 percentage-point improvement in mAP50 to reach 96% over the baseline YOLOv5s and is particularly stable in scenarios with strong reflections, low contrast, and small objects. The proposed technique can be readily applied to e-commerce image search, smart-cabinet product recognition, automated warehouse parcel sorting, and inventory auditing, thereby improving recognition accuracy and the level of operational automation.

Keywords

Object Detection, YOLOv5, Transparent Products, E-Commerce, Computer Vision, E-Commerce Deployment

1. 引言

随着人工智能的飞快发展，计算机视觉在电子商务中发挥着很大的作用。从基于内容的商品图像搜索、智能货柜的自动结算，到大型自动化仓库中的机器人分拣与库存盘点，高效、精准的目标检测算法在这些场景中扮演着至关重要的角色。与常规的商品不同，透明的商品(例如玻璃杯，化学试管)由于其独特的光学特性(如折射、反射和缺乏纹理)，对于传统的计算机视觉任务造成了严重的挑战[1]。在计算机视觉任务中，此类物体的外观高度依赖于环境背景，其边缘和内部特征往往与背景融为一体，导致基于常规卷积神经网络(CNN)的检测模型(如 YOLO、Faster R-CNN 等)难以提取有效特征，普遍存在定位模糊、漏检率高等问题。

近年来随着深度学习的发展，出现了一系列的目标检测模型，其中 YOLOv5 模型有着精度与速度的优势[2]。然而 YOLOv5 的主要核心网络是卷积神经网络(CNN)，在建模长距离的依赖关系以及全局上下文信息等方面存在局限性，而这正是准确感知和分割透明物体模糊边界所必需的能力，在处理透明物体模糊边界与高光伪影等问题时，传统的基于 CNN 的模型就难以发挥出作用。近年来，Transformer 由于其全局建模能力受到广泛的应用，但是在电子商务的透明商品检测中，由于其计算复杂度以及显存开销等难以部署在实时的目标检测模型中。所以在电子商务平台中，一个实时、轻量的透明电商物品检测模

型就十分重要。

近年来，许多人提出各种注意力机制，其中外部注意力(EA)通过两个共享的外部记忆矩阵以两层线性映射与双重归一化实现近似线性复杂度的全局建模。SimAM 则基于能量函数的零参数空间注意力，能在不引入可学习参数的前提下增强像素级显著性。这些工作为在实时检测器中引入全局—局部互补提供了新的可能。

基于上述的方法，本文在 YOLOv5 框架上提出一种混合注意力检测网络，核心是新设计的 CLEAR-Attn (Channel-Linear External-SimAM Attention with Residuals)模块。该模块以增强型外部注意力为骨干，通过两层线性映射与双重归一化引入外部可学习记忆，在近似 1 × 1 卷积的复杂度下聚合长程依赖[3]。其输出串接 SimAM 这一零参数空间注意力，显式放大边界与高光线索，提高模型对透明商品的局部感知能力[4]，最后由通道注意力(SENet)进一步融合两种注意力机制，增强模型的表达能力。

本研究的核心贡献在于提出了一个名为 CLEAR-Attn 的混合注意力模块，旨在解决上述研究空白。在 Trans10K 透明商品数据集上[5]，本文在保持训练策略与检测头、损失函数不变的前提下进行验证。实验结果显示，相比基线 YOLOv5s，所提方法在电商透明商品目标场景下的平均精度 mAP50 均有稳定提升，满足电商业务对低时延、易部署的工程要求。

2. 文献综述

目标检测模型是计算机视觉领域的重要领域，近年来的目标检测模型主要分为单阶段与两阶段检测器。单阶段的目标检测器以 YOLO 为主要代表，其中 YOLOv5 以其精度与速度的优势在目标检测中成为一个经典的框架。两阶段检测器常常以 Faster R-CNN 为主要代表，这一类的检测器虽然精度高但是检测速度稍慢[6]。其中 YOLOv5[7]在自动驾驶，安防监控、电子商务等方面都有着很多的作用，但是由于其主要的网络是 CNN，能够获得商品局部的特征细节，但是在面对透明商品这样需要全局建模能力的场景时，往往会遇到检测不精准的问题，会有一定的局限性。

传统的 CNN 模型由于其只能提取局部特征，但建模全局依赖关系需要很深的网络结构，网络的结构过深时可能会面对梯度爆炸等问题。Transformer 开始广泛应用于自然语言处理问题，但是其核心的自注意力(Self-Attention)机制能够直接计算序列中任意两个元素之间的关系，从而高效地捕获全局上下文信息[8]。现在 Transformer 也常常地应用于视觉领域，但是 Transformer 会带来更大的计算量，而且 Transformer 在一些小数据集上还会出现过拟合的问题，使得模型易在边界与伪影处产生误检与漏检[9]。近两年来虽然也出现了 RT-DETR 等实时目标检测模型[10]，但是由于其依赖于 Transfomer 所以也存在着很大的计算开销。

学界围绕该问题开展了数据集与方法方面的探索：ClearGrasp 针对透明体三维形状与深度估计构建了 RGB-D 基准[11]，同时 ClearGrasp 方法提出来一种大型的数据集，为后来的许多工作奠定了基础。由于透明物体的深度在检测时会存在失真等问题，所以目前的方法更多地聚焦于深度重建来恢复透明物体的深度[12]，例如 Yingjie Tang 提出一种自注意力多抗网络从原始 RGB 图中恢复深度图。同时针对透明物体的分割也有很多的研究，Trans10Kv2 [13]提出来一种基于 Transformer 的方法来实现对透明物体的分割，该方法利用了 Transformer 的全局建模能力来提高分割的精度。LingTong [14]提出一种针对透明物体的语义分割方法，该方法通过特征选择与特征增强来提高对透明物体的分割能力。这些方法充分说明了 Transformer 模型针对透明物体的优势，但是 Transformer 由于其二次计算复杂度，会增加很多的计算推理开销。方法层面，针对玻璃的分割、边缘与高光建模、折射恢复等议题均有进展。但就实时检测而言，如何在不明显增加推理开销的前提下，同时抑制反射诱发的背景误激活并强化弱边界，仍缺乏工程友好的通用方案。

Transformer 能够在透明物体的图像分割中发挥出很大的作用核心就是它的自注意力机制，能够对物体进行全局建模，但是自注意力机制带来的计算复杂度难以实现轻量实时的透明物体检测。为了平衡性能和效率，一系列轻量级注意力变体被提出。CBAM (Convolutional Block Attention Module)是一种轻量级的注意力机制[15]，能够在空间和通道上进行 Attention 操作，但是针对透明物体时还是难以实现全局建模的能力。SE (Squeeze-and-Excitatio)注意力机制[16]，通过显式地建模出卷积特征通道之间的相互依赖性来提高网络的表示能力，但是也面临着 CBAM 同样的问题。外部注意力(External Attention)自 2021 年提出， 利用两个共享的外部记忆单元替代了自注意力中的 Key 和 Value 矩阵[17]，将计算复杂度从二次降为线性，并隐式地学习了数据集级别的全局先验，既保留了全局依赖，又显著降低资源消耗。另一方面无参注意力机制，如 SimAM 基于神经科学中的显著性理论，无需任何可学习参数即可推导出 3D 注意力权重，实现了轻量化。这些工作为本研究设计高效、轻量科部署的混合注意力机制提供了基础。

综上所述，当前透明商品检测领域存在一个明显的研究空白：一方面，性能优异的 Transformer 模型因计算负担过重而难以落地实时应用；另一方面，现有的轻量级 CNN 检测器[18] (如 YOLOv5)又缺乏对透明物体特有的全局光学属性和微弱边缘的有效感知能力。虽然已有多种注意力机制被提出，但如何将它们高效地融合，以兼顾全局依赖建模、局部显著性增强与轻量化部署，仍未得到充分探索。本文的研究工作正是为了填补这一空白。本研究的主要贡献在于：提出了一种名为 CLEAR-Attn 的混合注意力模块，创造性地将线性复杂度的 EA 机制与零参数的 SimAM 机制深度融合，使 YOLOv5 基线模型同时获得了强大的长程上下文建模能力和空间显著性感知能力。该方案并非简单的模块堆叠，而是通过精心的设计(包括通道注意力筛选与残差连接)实现了两种注意力的优势互补，最终显著提升了模型在复杂电商场景下对透明商品的检测精度与鲁棒性，为实时高效的透明商品检测提供了一个实用、可靠的解决方案。

3. 方法

3.1. 整体网络架构

本文提出一种基于混合注意力机制改进的 yolov5s 模型，整体框架以 YOLOv5s 为基础，其核心创新在于将 External Attention (EA)模块与 SimAM 模块深度融合，构建了一个能够同时捕获全局上下文与局部显著性的高效检测器。我们的模型以 Yolov5s 为基线，其主干网络是 CSPDarknet，颈部网络是特征金字塔网络[19]。本文改进的策略是：使用一个高效的 CLEAR-Attn 注意力模块放在主干网络和颈部网络 C3 层之后，本研究分别在主干网络的第 3 和第 4 个 C3 模块后加入了 CLEAR-Attn 注意力模块，在颈部网络中分别在第 3、4 个 C3 模块后加入了 CLEAR-Attn 注意力模块。这种放置策略确保了多尺度特征在生成和融合后都能得到增强，使模型能够从低层到高层、从局部到全局地关注透明物体的关键特征，这对于识别具有复杂光学特性的透明商品包装至关重要。

3.2. CLEAR-Attn 注意力网络

本文提出 CLEAR-Attn 注意力模块结构如图 1 所示，首先将输入的 RGB 图像经过 YOLOv5 的 C3 模块，提取图像的特征(Feature) F_{in}，Project1x1 代表着 1 × 1 卷积投影，将前面得到的特征投影到低维便于后续的注意力计算，将特征保持为 B × C × H × W。B 代表着 batch，C 代表着通道数，H，W 代表着图像特征的尺寸。随后得到的特征首先经过 EA 模块，输入的特征经过卷积后得到了 Query，Query 通过查询外部记忆后，与外部 key 记忆 M_k 相乘生成了注意力图，M_k 把通道特征“投影”为若干原型，这些原型对整个数据集共享，把每个空间位置的通道向量映射到 K 个记忆槽。随后 K 个记忆槽经过了归一化(Norm)后与外部 value 记忆 M_v 相乘，根据注意力图把 K 个槽的聚合结果回写为 C 维通道特征，得到了经过 EA 的特征 F_{ea}。EA 外部注意力通过获得的数据集先验，能有效地增强模型对透明物体的全局建模能力，而

且其线性时间复杂度远小于自注意力机制。

输入的特征 F_{ea} 经过了 SimAM 模块。首先 SimAM 先基于能量函数在每个位置每个通道上生成空间权重，随后经过扩展(Expansion)将通道注意权重广播到空间维度，将输入的特征与扩展后的 3-D 权重进行了融合之后就得到了具有空间显著性的权重，SimAM 注意力机制基于能量函数为每个像素生成显著性权重，强调边界与高光区域，抑制背景伪影。随后经过了一个通道注意力 SENet 筛选了与“透明”相关的通道，下一步经过了一个融合模块 Fusion 后，融合了空间显著性与通道注意力，得到了增强后的特征 F_{ca}，进一步增强了特征的表达。

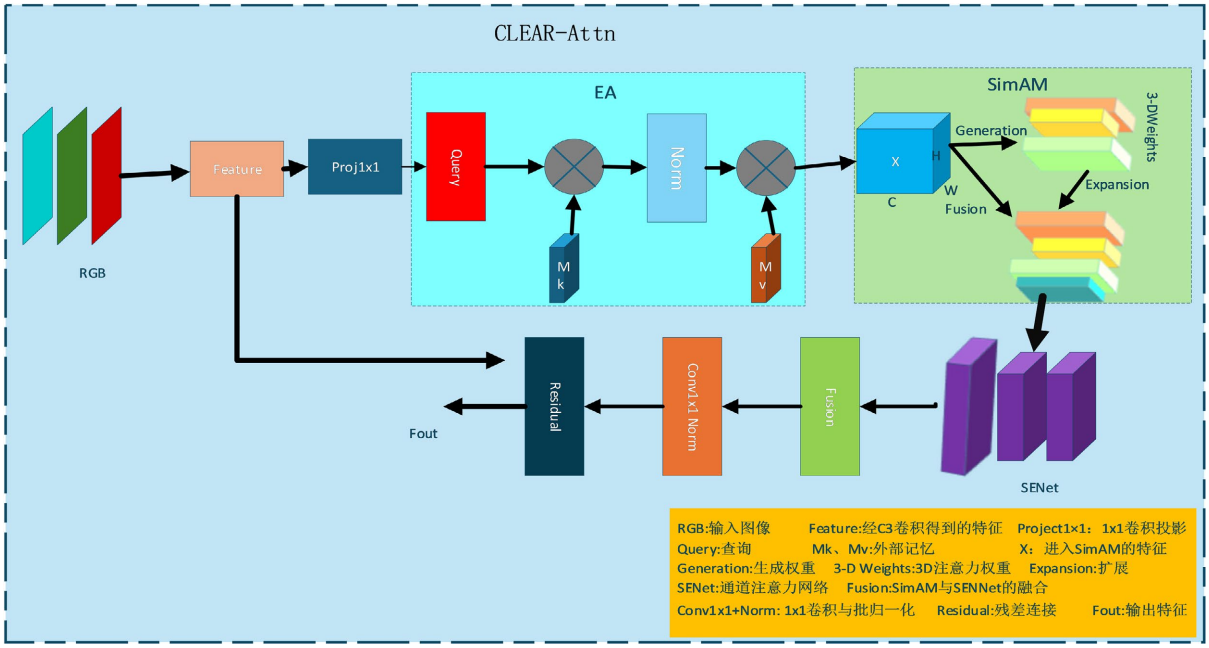

Figure1. The CLEAR-Attn attention network architecture
图 1. CLEAR-Attn 注意力网络结构

最后经过了一个 1 × 1 的卷积($\text{Conv}_{1\times1}$)与 Batch Norm 归一化确保了与输入特征的一致性，随后经过了 Residual 残差块连接了初始的一个特征，最后的输出见公式 1：

$$F_{out} = \text{ReLU}\left(\text{BU}\left(\text{Conv}_{1\times1}\left(F_{ca}\right)\right) + F_{in}\right) \tag{1}$$

其中 F_{out} 代表着最终的输出特征，ReLU 是激活函数，BN 代表着归一化， $\text{Conv}_{1\times1}$ 是一个 1 × 1 的卷积，F_{ca} 是经过几个注意力机制增强后的特征，F_{in} 代表着最开始的输入特征。最后的输出 F_{out} 就是 CLEAR-Attn 模块的输出特征，用于后续的检测头。

经过上面的一个 CLEAR-Attn 注意力模块，一方面能够有效地增强对透明物体的全局建模能力，另一方面也没有带来很大的计算开销，实现了一个轻量、实时的透明物体检测的模块能够部署在电商平台的实时检测模型当中。

4. 实验

4.1. 实验数据集及指标

本研究采用 Trans10K 数据集，这是一个大型的透明物体数据集，由 10,428 张真实场景的图像组成，并带有仔细的手动注释，数据集中透明物体类别包括玻璃器皿、塑料瓶、窗户等，具有丰富的尺度、形

状和背景变化。Trans10K 数据集所有的图像均采集于日常生活中的常见透明物体，采用了 5000、1000、4428 张图像分别用于训练、验证、测试。

本研究基于 Windows10，GPU 选用 RTX4070，12G 显存，使用 PyTorch2.3.0，CUDA12.1，Python3.8.2，输入图像尺寸为 640 × 640，训练迭代次数为 100 轮，batchsize 为 16，采用 SGD 优化器，学习率 0.01。

为了评估模型的效果，本研究采用目标检测领域通用指标：平均精度 mAP50，为了评估模型的参数，本文还对每种模型的参数进行对比。

4.2. 消融实验

为了验证模型的有效性，我们在 Trans10K 数据集上进行了一系列的消融对比实验。首先模型 1 就是在 YOLOv5s 基线模型下进行实验，YOLOv5s + SimAM 是在 YOLOv5s 在基线模型上加入 SimAM 空间注意力机制，YOLOv5s + EA 是在基线模型上加入了 EA 注意力机制，YOLO5s + EA + SimAM 是在基线模型上简单地串联加入了 SimAM 与 EA 注意力机制的模型，Ours 模型就是本文所提出来的 YOLOv5-CLEAR-Attn 模型，Parameter 代表着模型的参数大小。消融实验结果如表 1 所示，可以看出我们的模型比单独的各个模块精度更高，虽然参数量有所增加但是带来的精度提升是很有必要的。

Table 1. The results of the ablation study on the Trans10K dataset
表 1. Trans10k 数据集上消融实验的结果

Model	EA	SimAM	mAP50	Parameter
YOLOv5s	×	×	0.91	7.2 M
YOLOv5s + SimAM	×	√	0.92	7.2 M
YOLOv5s + EA	√	×	0.94	7.6 M
YOLOv5s + EA + SimAM	√	√	0.95	7.7 M
Ours	√	√	0.96	7.8 M

4.3. 对比实验

为了验证本研究的模型的效果，我们在 Trans10K 数据集上进行了一系列的对比实验。其中 Yolov5-CBAM 代表着加入 CBAM 注意力机制的 YOLOv5s 模型，RT-DETR 是近年来效果比较好的实时目标检测模型，Ours 代表着本文提出来的模型。分别在平均精度(mAP50)，计算量(GFLOPs)，推理速度(FPS)，参数等指标上进行了对比。为了保证实验结果的客观性，几个模型都采用了 4.1 节的训练策略。对比实验结果如表 2 所示，可以看出本文提出的模型在精度和效率上取得了平衡。

Table 2. Comparative experimental results on the Trans10K dataset
表 2. Trans10K 数据集上对比实验结果

Model	mAP50	GFLOPS	FPS	Parameter
YOLOv5-CBAM	0.93	17.0	203	7.4 M
RT-DETR	0.952	103.4	52.4	31.9 M
Ours	0.96	21.9	146	7.8 M

4.4. 实验结果分析

从表 1 中可以看出我们提出的方法比原本的基线模型 YOLOv5s 提高了 mAP50 提高了 5 个百分点，相比于单独加入 SimAM 模块或者 EA 模块，我们所提出的融合了 CLEAR-Attn 注意力网络的模型也有着更优的性能。原始的 YOLOv5s 的模型参数为 7M，本研究的模型的参数是 7.8M，在尽可能减少了计算量的同时也提高了模型的性能。在 4.3 的对比实验可知，本研究提出的模型在增加少量参数的情况下模型的检测精度高于 YOLOv5-CBAM，推理速度 146 低于 YOLOv5-CBAM 的 203，这主要是因为参数量的增加带来了额外的计算量，但是带来的精度提升远超微小的速度损失，这样的结果是满足电商平台的实时检测需求的。对比于实时目标检测模型 RT-DETR，本研究提出来 YOLOv5-CLEAR-Attn 模型在参数量和计算量均远小于 RT-DETR 的情况下，精度也更高。实验最终可视化的结果如图 2 所示。

Figure2. Visualization results
图 2. 可视化结果

5. 结论

本研究面向电商场景中透明/高反射商品难检测的问题，本文在 YOLOv5 框架上提出了基于混合注意力的改进模型 CLEAR-Attn，将 External Attention 的长程依赖建模与 SimAM 的零参数空间显著性增强进行深度融合，并在网络关键尺度以即插即用方式部署，实现对多尺度全局–局部信息的协同强化。在 Trans10K 透明商品数据集的实证结果显示，在增加少量参数情况下，模型相较基线 YOLOv5s 的 mAP50 提升约 5 个百分点，在强反射、弱对比及小目标场景下表现更为稳定。方案由标准算子构成、易于导出与部署，能够直接服务于电商平台的上架审核、图像搜索、智能货柜识别与仓配分拣/盘点等实时业务，具有较好的工程落地价值。

同时，本研究仍存在局限：其一，电商环境中强背光、镜面高光、复杂环境反射以及低照度噪声会导致目标/背景对比度过低或出现高亮饱和区，影响 EA 的全局汇聚与 SimAM 的显著性评估；其二，电商环境中遮挡物体与不透明物体重叠或遮挡的情况下对透明目标的检测精度有一定下降；其三对折射率较高、曲面复杂的玻璃器皿，或者纹理极弱的透明薄塑料包装情况下，易变形的透明商品会对定位造成干扰。后续工作将从两个方面推进：(1) 在面对复杂的光照条件下，引入折射、反射先验的数据增强与仿真，在仿真环境中模拟更复杂光照、材质，提高在真实环境下的适配。(2) 针对复杂的遮挡情况，选择偏振光、红外、深度等信息来辅助实现感知，通过多模态感知方法来解决当 RGB 失效时的感知问题。

综上，本文所提出的混合注意力机制 CLEAR-Attn 为电商透明商品检测提供了一种低改动、可部署、可扩展的结构化解法，可作为电商视觉系统的通用增强组件，并具备面向更广泛零售与供应链场景的推广潜力。一方面本研究所提出的方法可以在电商运营当中降低其成本，基于低成本 RGB 摄像头的计算机

视觉技术，能胜任此前必须依赖人工目检或特殊传感器的透明商品识别任务。另一方面由于本研究的方法的高精度以及高推理速度优势，在电商运营环境中例如仓储物流的自动化分拣环节可以降低误检率提升电商仓库运营效率。最后精准的检测能力意味着更低的货损率和差错率。更重要的是，可靠的自动化流程减少了人工干预，降低了用工成本与管理成本。

参考文献

[1] 谭鑫, 齐福霖, 王楠, 等. 基于视觉失真的玻璃表面检测方法[J]. 计算机辅助设计与图形学学报, 2025, 37(5): 832-843.

[2] 罗文沛, 李军. 基于 YOLOv5s 的无人机视角下的小目标检测算法[J]. 计算机应用, 2025, 45(S1): 235-238.

[3] Guo, M.H., Liu, Z.N., Mu, T.J. and Hu, S.M. (2023) Beyond Self-Attention: External Attention Using Two Linear Layers for Visual Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 5436-5447.

[4] Yang, L.X., Zhang, R.Y., Li, L., *et al.* (2021) SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Proceedings of the* 38*th International Conference on Machine Learning*, **139**, 11863-11874.

[5] Xie, E., Wang, W., Wang, W., Ding, M., Shen, C. and Luo, P. (2020) Segmenting Transparent Objects in the Wild. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV* 2020., Springer, 696-711. https://doi.org/10.1007/978-3-030-58601-0_41

[6] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[7] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* (*NIPS*'17), Long Beach, 4-9 December 2017, 6000-6010.

[9] Dosovitskiy, A., *et al.* (2021) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.

[10] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., *et al.* (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 16-22 June 2024, 16965-16974. https://doi.org/10.1109/cvpr52733.2024.01605

[11] Sajjan, S., Moore, M., Pan, M., Nagaraja, G., Lee, J., Zeng, A., *et al.* (2020) Clear Grasp: 3D Shape Estimation of Transparent Objects for Manipulation. 2020 *IEEE International Conference on Robotics and Automation* (*ICRA*), Paris, 31 May-31 August 2020, 3634-3642. https://doi.org/10.1109/icra40945.2020.9197518

[12] Tang, Y., Chen, J., Yang, Z., Lin, Z., Li, Q. and Liu, W. (2021) DepthGrasp: Depth Completion of Transparent Objects Using Self-Attentive Adversarial Network with Spectral Residual for Grasping. 2021 *IEEE*/*RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Prague, 27 September-1 October 2021, 5710-5716. https://doi.org/10.1109/iros51168.2021.9636382

[13] Xie, E., Wang, W., Wang, W., Sun, P., Xu, H., Liang, D., *et al.* (2021) Segmenting Transparent Objects in the Wild with Transformer. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, Montreal, 19-27 August 2021, 1194-1200. https://doi.org/10.24963/ijcai.2021/165

[14] Tong, L., Song, K., Tian, H., Man, Y., Yan, Y. and Meng, Q. (2023) SG-Grasp: Semantic Segmentation Guided Robotic Grasp Oriented to Weakly Textured Objects Based on Visual Perception Sensors. *IEEE Sensors Journal*, **23**, 28430-28441. https://doi.org/10.1109/jsen.2023.3321742

[15] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[16] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/cvpr.2018.00745

[17] Huang, Z., Liang, M., Qin, J., Zhong, S. and Lin, L. (2023) Understanding Self-Attention Mechanism via Dynamical System Perspective. 2023 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Paris, 1-6 October 2023, 1412-1422. https://doi.org/10.1109/iccv51070.2023.00136

[18] Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J. (2022) A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 6999-7019. https://doi.org/10.1109/tnnls.2021.3084827

[19] 王慧云, 赵俊生, 王禹, 等. 面向无人边防的复杂环境遮挡小目标检测算法[J]. 电子测量技术, 2024, 47(21): 168-177.