Modeling and Simulation 建模与仿真, 2025, 14(4), 1186-1199
Published Online April 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.144365

Hans汉斯

基于改进YOLOv8-seg的车辆智能定损

王飞扬

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2025年3月30日；录用日期：2025年4月23日；发布日期：2025年4月30日

摘　要

在中国科技快速发展的大背景下，智能化、自动化逐渐渗透到各个领域。车辆的智能定损研究正在有序推进，传统的定损过程需要进行现场勘察和测量，这一过程繁琐且耗时。而智能定损技术通过分析大量的汽车相关数据和损伤图像，能够在几秒钟内完成对损伤的识别和评估。但目前相关应用泛化性不足，受图片分辨率影响，在小目标及区域分割精确度上仍有较大进步空间。针对这些问题，本文以YOLOv8-seg图像分割算法为基准模型提出了一种精确度高的车辆智能定损模型。通过引入NWD (Normalized Wasserstein Distance)损失函数，强化了模型的小目标识别及分割能力，以及对低分辨率图像的特征提取能力，提升了模型的鲁棒性。提出DLKA (Deformable Large Kernel Attention)可变形大核注意力机制与YOLOv8-seg主干网络C2f模块相融合，实现了在计算复杂度略微增加的前提下，模型分割能力的大幅度提升，强化模型的泛化能力。通过实验验证，本文的改进算法相比原始算法，平均精度上提升了8.1%，在车辆智能定损研究中表现出色。

关键词

YOLOv8-seg，车辆智能定损，实例分割，D-LKA

Vehicle Intelligent Damage Assessment Based on Improved YOLOv8-seg

Feiyang Hang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 30th, 2025; accepted: Apr. 23rd, 2025; published: Apr. 30th, 2025

Abstract

In the context of China's rapidly developing technology, intelligence and automation are gradually permeating various fields. Research on intelligent vehicle damage assessment is being carried out

文章引用: 王飞扬. 基于改进 YOLOv8-seg 的车辆智能定损[J]. 建模与仿真, 2025, 14(4): 1186-1199.
DOI: 10.12677/mos.2025.144365

in an orderly manner. Traditional damage assessment processes require on-site inspections and measurements, which can be cumbersome and time-consuming. Intelligent damage assessment technology analyzes large amounts of car-related data and damage images to identify and evaluate damages within seconds. However, current applications lack generalization capabilities and are affected by image resolution, leaving significant room for improvement in accuracy for small targets and regional segmentation. In response to these issues, this paper proposes a high-precision intelligent vehicle damage assessment model based on the YOLOv8-seg image segmentation algorithm. By introducing the NWD (Normalized Wasserstein Distance) loss function, the model enhances its ability to recognize and segment small targets as well as extract features from low-resolution images, thereby improving its robustness. The proposed DLKA (Deformable Large Kernel Attention) mechanism is integrated with the C2f module of the YOLOv8-seg backbone network, achieving a substantial increase in model segmentation capability with only a slight increase in computational complexity, thus enhancing the model's generalization ability. Experimental validation shows that the improved algorithm presented in this paper achieves an average precision improvement of 5.9% compared to the original algorithm, demonstrating excellent performance in research on intelligent vehicle damage assessment.

Keywords

YOLOv8-seg, Vehicle Intelligent Damage Assessment, Instance Segmentation, D-LKA

1. 引言

随着汽车保有量的持续增长和交通事故频发，车辆定损作为保险理赔和维修评估的核心环节，面临着效率低、成本高、主观性强等挑战。传统的车辆定损依赖人工经验，存在耗时长、易受评估者主观因素影响等问题，尤其在复杂损伤(如细微划痕、凹痕、裂纹等)的识别中，误检和漏检率较高。近年来，人工智能与计算机视觉技术的快速发展为车辆定损提供了新的解决方案。智能定损是一种基于人工智能技术对车辆外观和损伤部位进行自动化判定的系统。能在车险理赔、汽车维修和二手车市场等各方面提供智能定损的服务。基于深度学习的自动检测系统能够通过图像分析快速定位损伤区域并分类损伤类型，相比传统的人工评估，显著提升了定损效率与客观性。

在早期阶段，研究者通常基于物理特征或者传统机器视觉的方式进行车辆外观损伤检测。例如利用超声波传感器[1]，高频声波反射检测车身裂纹，适用于金属结构的无损检测，但对微小损伤敏感度较低。Jayawardena S 提出一种名为 3D CAD [2]的模型，通过对比完好的汽车推断待检测汽车受损坏的程度，但存在由于汽车的外表不时发生变化导致算法的应用性不强。射线检测[3]主要采用 X 射线照射待检测物体，通过判断待检测车辆中是否存在缺陷，进而确定缺陷的位置以及损伤程度。针对划痕损伤，边缘检测方法采用不同的边缘检测算子[4]检测表面划痕，但是当被测物纹理复杂或者划痕与背景对比度差时，很难获取划痕的边缘特征，很容易误检。也可以利用阈值分割和模板匹配在规则纹理下提取损伤区域，例如通过 Canny [5]边缘检测定位车漆划痕，但同样是在复杂背景下泛化能力差。

近年来，随着深度学习的高速发展，相比用传统机器视觉等方式进行智能定损，目标检测与实例分割技术的突破为车辆损伤自动化检测提供了新思路。2021 年文献[6]使用了 Mask R-CNN 图像分割算法来检测汽车的损坏区域，但其中类别只有单一的“划痕”，且模型推理速度较慢，难以满足实时定损需求。

2023 年文献[7]制作了新的汽车损坏图像数据集，开源了 9000 多张图片以及 6 种损伤类别，此外进行了大量实验分析，为后续研究者提供了有效的数据集。文献[8]使用了传统 CNN 检测模型用于定位车辆损伤并将这些损伤分为十二个类别，且能够在各种复杂场景下准确地检测小目标损伤。为了实现在嵌入式设备中实时的智能定损，需要保证精确度同时，尽可能对模型进行轻量化压缩。Lee [9]等人使用了 EfficientNet-B0、MobileNet-V2 等五种轻量化网络模型，介绍了使用四分之三视角汽车图片尽可能看全整体以及一套二元分类模型来评估性能。在过去车辆损伤检测领域，两阶段模型(如 Faster R-CNN)曾因高精度占据主导地位，但其计算复杂度高限制了其实时性。目前单阶段模型如 YOLO [10]系列通过锚框与分类并行优化，在速度与精度间取得了平衡。谢东升[11]使用了 YOLOv3 目标检测方法对车辆损伤部位进行标注，再利用 Mask R-CNN 对车辆各个部位进行识别分割，两者相结合即可准确判断出损伤部位和类型。金浩然[12]基于 YOLOX 模型使用了坐标注意力和焦点损失函数减少了模型参数和计算量同时提高了精度，并且进行了 Web 端部署，满足方便且快速的检测需求，但其精度及损伤种类仍需提高。

然而，现有系统在实际应用中仍面临诸多挑战：微小损伤(如划痕)因像素占比低，易被常规卷积操作忽略，检测精度不足。传统损失函数对不规则损伤边界敏感度不足。复杂光照条件下的图像噪声较大。车身损伤常分布于不同尺度区域(如车灯局部与大范围凹陷)。本文使用 YOLOv8-seg 图像分割模型作为基准模型。引入了 NWD (Normalized Wasserstein Distance) [13]损失函数作为分割损失函数，使用 Wasserstein 距离进行小目标分割，大大提升对车辆划痕、小范围凹陷等类型的分割能力。D-LKA (Deformable Large Kernel Attention) [14]的注意力机制，利用其大卷积核的广阔感受野和可变形卷积的灵活性可有效地处理复杂的视觉信息这一特性，在较小增加计算量同时提升精度。

2. YOLOv8-seg 相关理论

2.1. YOLOv8-seg 模型概述

YOLOv8-seg 是 Ultralytics 团队在 YOLOv8 目标检测模型基础上扩展的实例分割版本，其核心思想是将目标检测与像素级分割任务结合，实现端到端的实例分割。如图 1 所示，模型分为 Backbone (主干网络)、Neck (特征融合层)和 Head (检测与分割头)三部分。整体模型架构延续了 YOLO 系列的经典设计，在 Backbone 部分 YOLOv8 将 YOLOv5 中的 C3 模块替换为 C2f 模块，通过多分支残差连接和梯度分流设计，进一步减少参数量并提升特征融合效率，实现轻量化。沿用 YOLOv5 的 SPPF 模块，通过串行最大池化层高效捕获多尺度上下文信息。在 Neck 部分简化了 PAN-FPN 结构以及引入新的 Path Aggregation Network (Rep-PAN)，增强了多尺度特征的融合能力。且在 Head 进行了针对性改进以支持分割任务。本课题会以 YOLOv8-seg 作为主要的研究模型。

2.2. YOLOv8 Backbone (主干网络)

YOLOv8-seg 是 Ultralytics 团队在 YOLOv8 目标检测模型基础上扩展的实例分割版本，其核心思想是将目标检测与像素级分割任务结合，实现端到端的实例分割。如图 1 所示，模型分为 Backbone (主干网络)、Neck (特征融合层)和 Head (检测与分割头)三部分。整体模型架构延续了 YOLO 系列的经典设计，在 Backbone 部分 YOLOv8 将 YOLOv5 中的 C3 模块替换为 C2f 模块，通过多分支残差连接和梯度分流设计，进一步减少参数量并提升特征融合效率，实现轻量化。沿用 YOLOv5 的 SPPF 模块，通过串行最大池化层高效捕获多尺度上下文信息。在 Neck 部分简化了 PAN-FPN 结构以及引入新的 Path Aggregation Network (Rep-PAN)，增强了多尺度特征的融合能力。且在 Head 进行了针对性改进以支持分割任务。本课题会以 YOLOv8-seg 作为主要的研究模型。Backbone 主要是对图片进行特征提取的，输出三个不同尺寸的特征图为后续 Neck 部分的多尺度融合提供基础。主干网络的模块主要由 CBS 模块、C2f 模块、SPPF

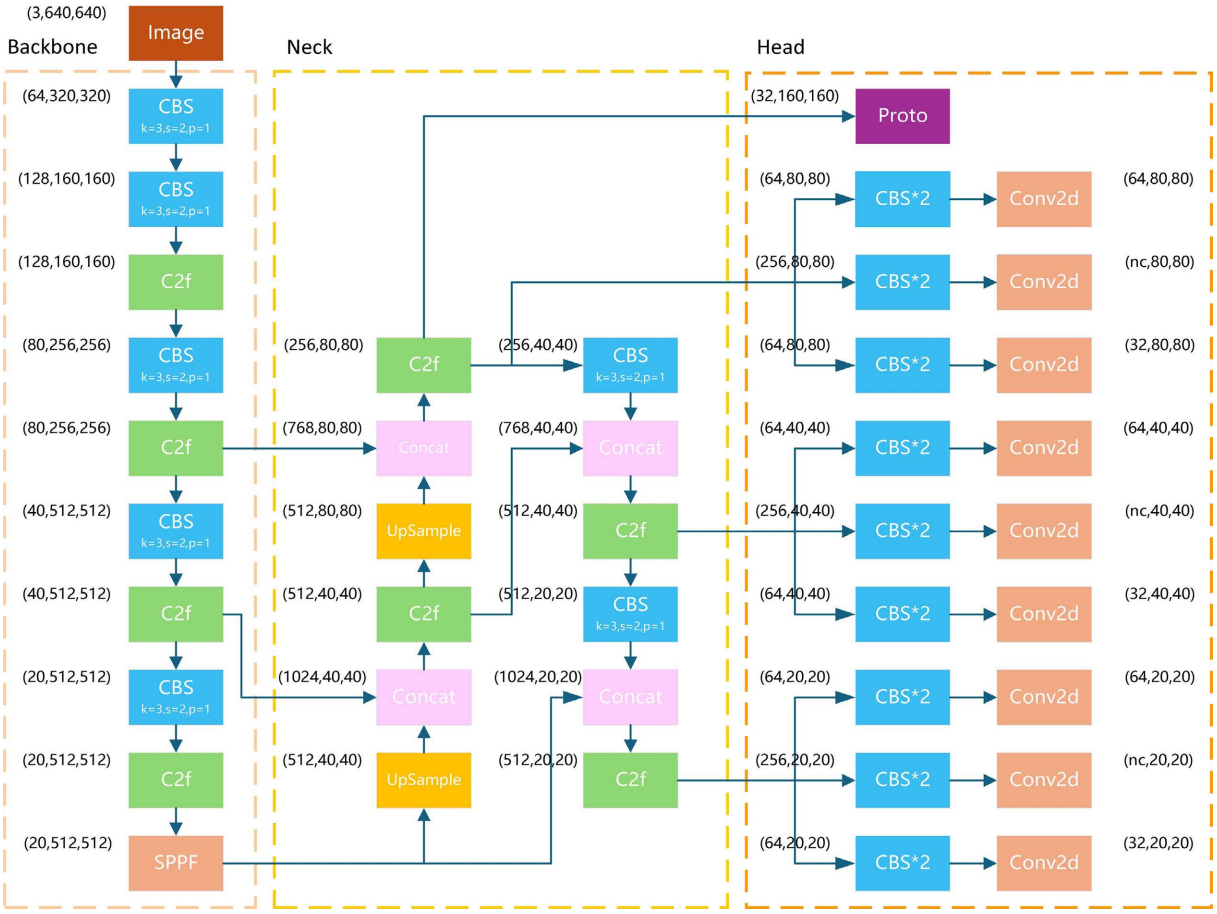

Figure 1. YOLOv8-seg network model structure diagram
图 1. YOLOv8-seg 网络模型结构图

模块组成。对于每个模块都会输出对应尺寸和通道数的特征图，比如(64, 320, 320)表示输出 64 通道数和 320 × 320 大小的特征图。其中 CBS 模块由 Conv 卷积层、BN 批量归一化层和 SiLU 激活函数组成，实现对输入特征图的特征提取和通道数调整，k 代表卷积核大小、s 代表步长、p 代表填充值。C2f 模块可以对特征进行深度提取，其输入输出的特征图大小不变，但其内部会将输入特征图被分割为两部分，一部分直接传递，另一部分经过 n 个 Bottleneck 处理，最终拼接后输出。SPPF 模块可实现快速池化操作，采用多个小池化层代替一个大核的池化操作，减少了计算量，提升了推理速度。

2.3. YOLOv8 Neck (颈部网络)

Neck 部分负责融合 Backbone 输出的多尺度特征，Neck 部分采用 PANet 的网络 PAN-FPN (Path Aggregation Network + Feature Pyramid Network)结构，负责特征融合与增强，提升模型对不同尺度目标的适应能力。把 Backbone 三个不同尺度的特征图通过上采样的方式进行 Concat，经过两次上采样后输出第一个特征图，然后通过两个 CBS 模块对特征图进行下采样输出第二个和第三个特征图到 Head 部分。Concat 模块可将上采样后的上采样后的深层特征与 Back 输出的渐层特征融合，连接不同层次的语义信息。UpSample 模块通过插值算法将特征图尺度放大与渐层特征图实现尺寸一致，可以恢复部分空间细节信息。

2.4. YOLOv8 Head (头部网络)

YOLOv8-seg 继承了 YOLACT 的检测–分割并行框架，在 Head 部分采用解耦头设计，包含检测分支和分割分支，分别处理目标定位与掩码预测。YOLOv8 的检测模型会生成用来预测边界框和分类的特征图。相对于 YOLOv8 的网络结构，YOLOv8-seg 只有 Head 部分不一样。在三个尺度下，除了 YOLOv8 检测分支的特征图，还会生成 32 通道的 Mask Coefficients 特征图当成 Mask 系数。此外，还会通过上采样和卷积生成 32 通道数 160 × 160 分辨率的 Prototype Mask 特征图，该分辨率下像素损失更少，会有更好的分割效果。

2.5. YOLOv8-seg 损失函数

YOLOv8-seg 的损失函数包含分类损失、定位损失和分割损失。相比于 YOLO 系列以往的损失函数，一个重要的创新是针对不同大小目标的检测加入了动态权重调整机制，优化了小目标检测性能。此外，通过自适应锚框和多尺度训练策略调整损失函数的权重，使模型在不同尺度下均能实现高质量检测。分类损失采用二元交叉熵(BCE Loss)通过计算预测类别与真实标签的差异，逐类别判断目标是否存在。BCE Loss 可表示为：

$$L_{BCE} = \frac{1}{N}\sum_{i} -\left[y_i \cdot \log\left(p_i\right) + \left(1-y_i\right)\cdot \log\left(1-p_i\right)\right] \tag{1}$$

其中 $y_i \in \{0,1\}$ 代表真实标签(否属于某类别)， p_i 代表模型预测的概率值。

定位损失函数由完全交并比 + 分布聚焦损失(CIoU Loss + DFL Loss)构成，这两个损失函数共同工作。对于 CIoU Loss，相当于在 IoU 基础上增加中心点距离惩罚项和长宽比差异项，优化边界框重合度。CIoU Loss 可表示为：

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2}{c^2} + \frac{v^2}{1-\mathrm{IoU}+v} \tag{2}$$

其中 ρ 代表预测框与真实框中心点的欧氏距离，c 表示最小包围框对角线长度，v 是长宽比的计算。DFL Loss 主要用于提高边界框回归的精度。DFL Loss 可表示为：

$$L_{\mathrm{DFL}} = -\left(\left(y_{i+1}-y\right)\log\left(S_i\right) + \left(y-y_i\right)\log\left(S_{i+1}\right)\right) \tag{3}$$

y_{i+1} 和 y 为标签的实际值， S_i 和 S_{i+1} 是预测值(预测的边界框分布概率)。

分割损失采用的是和分类损失一样的二元交叉熵(BCE Loss)。计算预测 Mask 与真实 Mask 的逐像素差异，结合 Prototype Mask 生成实力分割结果。其表达式可表示为：

$$L_{seg} = \frac{1}{N}\sum_{i} -\left[y_i \cdot \log\left(m_i\right) + \left(1-y_i\right)\cdot \log\left(1-m_i\right)\right] \tag{4}$$

其中 $y_i \in \{0,1\}$ 代表真实 Mask 掩膜像素标签， m_i 代表预测 Mask 的概率值。

3. 模型改进

3.1. NWD (Normalized Wasserstein Distance)损失函数

在车辆外观损伤中，存在许多小范围形变、裂纹、划痕等损伤。由于这种类型因小尺寸和像素问题，导致特征提取十分困难。由图 2 可知，训练当中微小的检测偏差会对小目标的 IoU 影响较大，使得损失函数变化较大，模型训练难以收敛。针对这一问题，本文引用了 Normalized Wasserstein Distance (NWD)损失函数，一种基于 Wasserstein 距离的度量方法，可用于提高微小物体的检测能力。

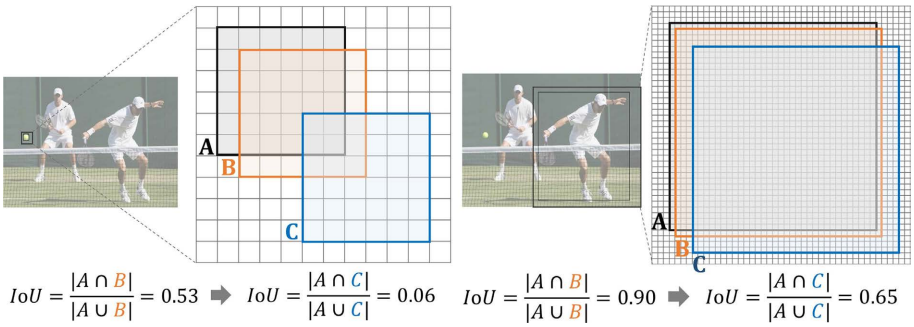

Figure 2. The sensitivity contrast of IoU on tiny and normal scale objects
图 2. IoU 对微小和正常尺寸检测敏感性对比

文[13]中提到在微小物体检测场景中，IoU 对位置偏差过于敏感，这在基于锚框的检测中使用时会极大地降低标签分配的质量，影响模型训练。NWD 损失函数通过将每个边界框建模为二维高斯分布。这种建模方式可以更全面地描述边界框的位置信息和尺寸信息，而不是仅仅依赖于重叠面积。利用 Wasserstein 距离来度量两个高斯分布之间的差异，可以在两个边界框没有重叠或重叠极少时依然提供有效的距离衡量。与 IoU 不同，Wasserstein 距离是连续且平滑的，这使得在梯度传递过程中更稳定，特别适合小目标这种细微变化较大的情况。为了消除不同尺度目标之间的影响，NWD 对 Wasserstein 距离进行了归一化处理。归一化后的指标在数值上更稳定，且具有尺度不变性，这对于小目标检测尤为重要，因为它们的尺寸变化不会引起过大的损失波动。

由于微小目标的边界框内往往会包含较大部分背景像素。目标像素多集中于中心区域，而背景像素则主要分布在边缘。为了更精确地反映边界框内不同位置像素的重要性，可以将边界框建模为二维高斯分布，在此模型下，中心像素拥有最大的权重，而权重会沿着中心向边界逐渐降低。具体来说，可以将边界框建模为二维高斯分布图 3。

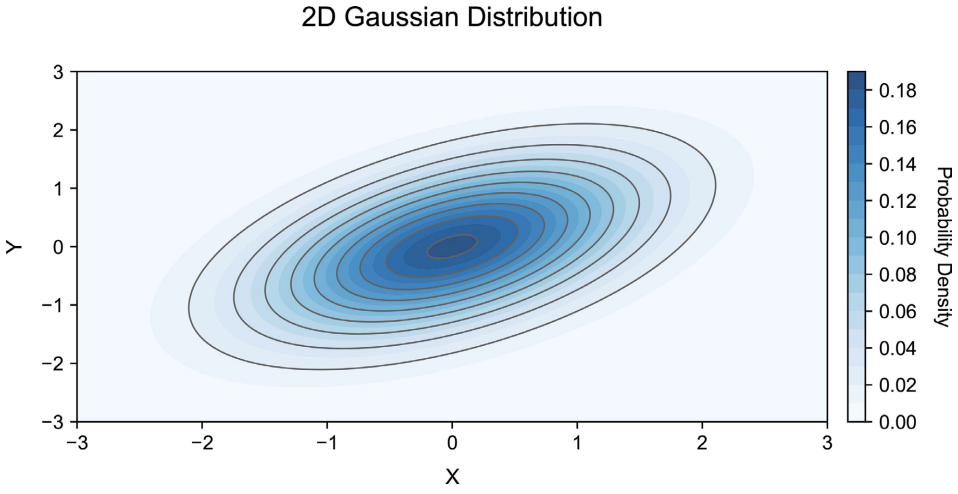

Figure 3. 2D Gaussian distribution diagram
图 3. 二维高斯分布示意图

可设边界框为 $R=(cx,cy,w,h)$，其中 (cx,cy) 代表均值，w 和 h 分别表示中心坐标的宽度和高度。那么其内切椭圆的方程可以表示为：

$$\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}=1 \tag{5}$$

式中(μ_x,μ_y)为椭圆的中心坐标，(σ_x,σ_y)为沿 x、y 轴的半轴长度。因此，$\mu_x=cx$，$\mu_y=cy$，$\sigma_x=\frac{w}{2}$，$\sigma_y=\frac{h}{2}$。二维高斯分布的概率密度函数可以表示为：

$$f(x|\mu,\Sigma)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right)}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{6}$$

式中 μ 和 Σ 可表示为：

$$\mu=\begin{bmatrix}c_x\\c_y\end{bmatrix},\Sigma=\begin{bmatrix}\frac{w^2}{4}&0\\0&\frac{h^2}{4}\end{bmatrix} \tag{7}$$

对于预测框和真实框的两个二维高斯分布 $\mu_1=\mathcal{N}(m_1,\Sigma_1)$ 和 $\mu_2=\mathcal{N}(m_2,\Sigma_2)$，它们之间的二阶 Wasserstein 距离公式为：

$$W_2^2(\mu_1,\mu_2)=m_1-m_{22}^2+\Sigma_1^{\frac{1}{2}}-\Sigma_{2F}^{\frac{1}{2}2} \tag{8}$$

然后将 Wasserstein 距离转换为 0~1 的相似性度量(归一化)：

$$\mathrm{NWD}(\mu_1,\mu_2)=\exp\left(-\frac{\sqrt{W_2^2(\mu_1,\mu_2)}}{C}\right) \tag{9}$$

本文将分割损失二元交叉熵(BCE Loss)替换为 NWD 损失函数。其中 C 需要根据本身的数据集自行调整，对于本文实验设置为 12.8。回归损失函数可以定义为：

$$L_{seg}=1-NWD(\mu_1,\mu_2) \tag{10}$$

该方法核心优势在于对微小目标的尺度变化不敏感，适合不同大小的 Mask。相比 IoU 的突变性，NWD 对位置偏差的响应更平缓，减少误判。即使预测 Mask 与真实 Mask 无重叠，仍可计算相似度，提供有效的梯度优化。以及增加正样本数量，缓解小区域因 IoU 阈值过高导致的监督信号不足。

3.2. D-LKA (Deformable Large Kernel Attention)可变形大核注意力

注意力机制类似于人脑和人眼的感知机制，它可以让神经网络选择性地关注某个重要信息，比如对图像加强某个区域，使得模型更加高效。车辆外观损伤图像当中，存在因损伤带来的不规则形变。传统的注意力机制使用固定的感受野，可能在局部区域内表现较好，但当面对形变较大或不规则的损伤区域时，难以捕捉到足够的全局上下文信息。本文引用了 D-LKA (Deformable Large Kernel Attention)可变形大核注意力，提高对不规则形变的特征提取能力。

Azad R [14]等人在一篇医学图像分割领域论文中提出了可变形大核注意力(Deformable Large Kernel Attention, D-LKA)，这是一种采用大卷积核来充分理解上下文信息的简化注意力机制。大核注意力(Large Kernel Attention, LKA)通过分解大核卷积为深度可分离卷积与膨胀卷积，平衡感受野与计算效率。可变形卷积(Deformable Convolution)通过动态调整采样位置，增强对不规则形状(如器官边界)的适应性。D-LKA

结合了大卷积核的广阔感受野和可变形卷积的灵活性，形成两阶段注意力机制，通过 LKA 捕获全局上下文信息，模拟自注意力的感受野。引入可变形卷积对 LKA 的输出进行动态调整，增强对局部形变的建模。这一机制通过动态调整卷积核的形状和大小来适应不同的图像特征，提高了模型对目标形状和尺寸的适应性。可变形大卷积核通过不断地训练学习，改变它的感受野，可以更有效率且减少去计算无关的局部信息，而从去关注全局上面更需要关注的关键信息。

图 4 为可变形大核注意力模块架构示意图，其中包括标准二维卷积(Conv2D)用于提取输入特征的基本空间信息。带有偏移量的变形卷积(Deformable Convolution, Deform-DW Conv2D)通过引入可学习的偏移量，使卷积核能够根据输入特征自适应地调整其感受野，捕捉车辆外观不规则形变。偏移场(Offsets Field)的计算是由标准卷积层生成，用于指导变形卷积层如何调整其采样位置，确保卷积操作能够适应输入特征的几何形变。GELU 激活函数用于增加网络的非线性表达能力，帮助模型更好地拟合复杂的特征映射。

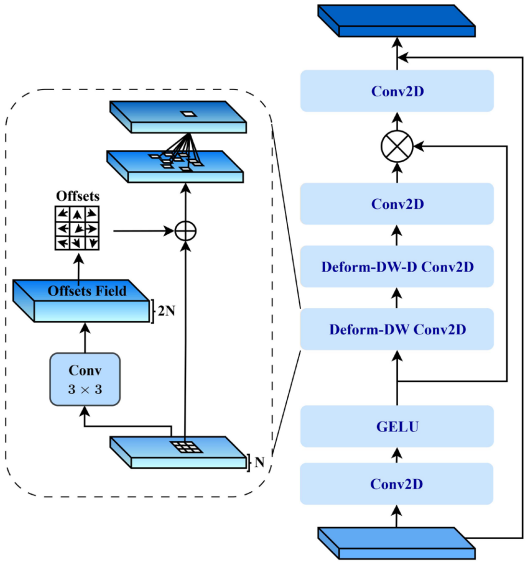

Figure 4. Architecture of the deformable LKA module
图 4. 可变形大核注意力架构

为求得可变形大核卷积中采样点位置的偏移量。对于无变形卷积操作而言，得到的输出特征值可以是每一个卷积窗口的权重乘以输入特征图对应区域的特征值的和。图 5 中可看到，$\mathcal{R}$ 为卷积核，p_n 是相对 p_0 采样点的位置，由于可变形卷积核的滑动窗口每个采样点都会发生偏移，设偏移量 offsets 为 Δp_n

$$R = \begin{matrix} (-1,-1) & (-1,0) & (-1,1) \\ (0,-1) & (0,0) & (0,1) \\ (1,-1) & (1,0) & (1,1) \end{matrix}$$

Figure 5. Relative positions of sampling points in a convolution kernel
图 5. 卷积核内采样点相对位置

那么可变形卷积核计算公式可以为：

$$y(p_0)=\sum_{p_n\in\mathcal{R}} w(p_n)\cdot x(p_0+p_n+\Delta p_n) \tag{11}$$

其中 $y(p_0)$ 是对 p_0 区域卷积操作的特征值， $w(p_n)$ 是对应区域的权重。偏移量 Δp_n 可以通过一个额外的卷积层从输入特征中学习偏移量。该层的输入是原始特征图，输出是每个采样点在 X 方向和 Y 方向的二维偏移量。如使用一个 3 × 3 的卷积核，那么就会输出通道数 channel 为 2 乘以卷积核的采样点数，3 × 3 的卷积核对应输出通道数就为 18。将 18 个通道数对应中心采样点的特征值进行 reshape 可得到偏移量 Δp_n 的矩阵。偏移量可以通过网络的反向传播算法更新，其卷积层的梯度通过链式法则传递，调整其权重以最小化损失函数，使得偏移量的参数更加合理。即可得到 $p_0+p_n+\Delta p_n$ 对应的位置。那么该位置对应的亚像素点的值需要用双线性插值法得到。整个图像的像素值是线性均匀变化的，利用离该亚像素点最近的四个像素值，根据距离加权平均来求得该亚像素点的值。

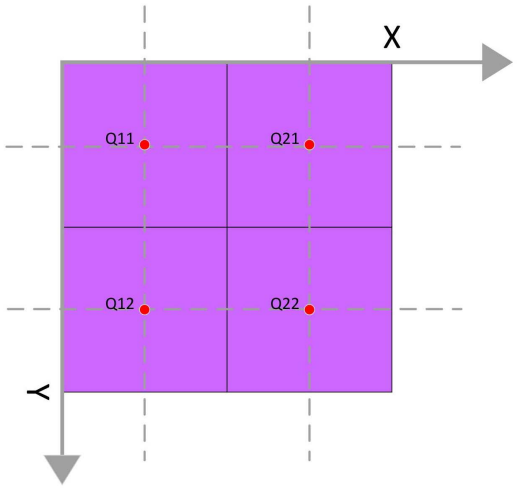

Figure 6. Bilinear interpolation diagram
图 6. 双线性插值示意图

由图 6 可得，$(Q_{11},Q_{21},Q_{12},Q_{22})$ 为相邻四个区域像素值，求 P 点的亚像素值。可以计算 x 和 y 方向上的加权平均。那么 x 向上的两次单线性插值：

$$f(R_1)=\frac{x_2-x}{x_2-x_1}\times f(Q_{11})+\frac{x-x_1}{x_2-x_1}\times f(Q_{21}) \tag{12}$$

$$f(R_2)=\frac{x_2-x}{x_2-x_1}\times f(Q_{12})+\frac{x-x_1}{x_2-x_1}\times f(Q_{22}) \tag{13}$$

y 方向上的单线性插值：

$$f(P)=\frac{y_2-y}{y_2-y_1}\times f(Q_{11})+\frac{y-y_1}{y_2-y_1}\times f(Q_{21}) \tag{14}$$

双线性插值法得到的亚像素值得出的即为该点的值。偏移量可以使卷积核能不是聚焦在颜色上或者纹理上，而是在不规则形状或变形目标，比如本文数据集的边界信息。对于这一特点，再结合有损车辆通常不是发生颜色或者纹理的变化而是形状的变化，就这一相似性能够使本文模型的分割能力得到有效提升。而且仅增加一个轻量级卷积层，参数量远小于全局注意力机制。本文为了不过多地增加模型的参

数量，选择把这个注意力机制融合到主干网络的最后一层 C2f 模块当中作为 C2f_DLKA 模块。

4. 实验结果与分析

4.1. 数据集及预处理

本文所使用的有损车辆数据集为网络上收集的公开数据集，一共有 7239 张图片。该数据集为 COCO 数据集，使用 JSON 文件形式存储，包含图像数据，分割 Mask 坐标信息与类别 ID 等。为了能让本文模型使用该数据集，需转化成 YOLO 格式，每个图像对应一个 TXT 文件，每行表示一个分割 Mask 存储归一化的坐标信息和类别 ID。图像分辨率有 800 × 800、1000 × 667 两种，在训练时设置超参数 imgsz = 640，会被统一缩放到 640 × 640 分辨率，且会进行 Mosaic 数据增强，即四张图片进行随机裁剪，再拼接到一张图上作为训练数据。该数据集中将图片划分为训练集 5213 张，验证集 1194 张，测试集 832 张。损伤类型分为 7 类，分别是汽车部分裂纹(Car-Part-Crack)、变形(Deformation)、瘪胎(Flat-Tire)、玻璃破裂(Glass-Crack)、灯泡损坏(Lamp-Broken)、划痕(Scratch)、侧视镜裂纹(Side-Mirror-Crack)。数据集图片示例如图 7 所示：

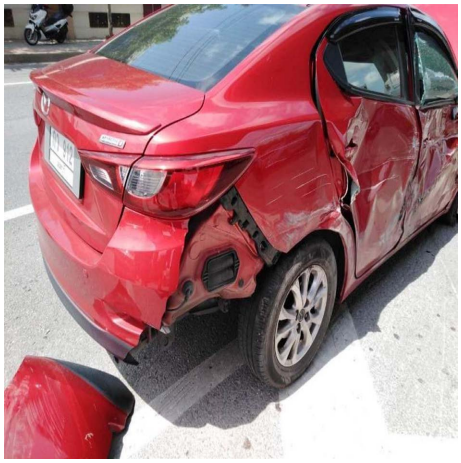

Figure 7. Damaged vehicle diagram
图 7. 有损车辆示意图

4.2. 实验环境配置

为保证实验结果的有效性，控制无关变量的影响，本文的所有的实验都基于同一硬件软件平台。操作系统是基于 Ubuntu20.04，CPU 为 Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz，GPU 选用 RTX2080Ti(11GB)。软件配置版本是 Cuda11.8，Pytorch 为 2.0.0，以及 Python3.8。默认训练验证参数和超参数设置也须一致，训练轮次 epochs 为 400，批大小 batch 为 32，图像尺寸 imgsz 为 640，马赛克增强关闭 close_mosaic 为 10，Mask 采样比例 mask_ratio 为 4，非最大值抑制交并比 iou 为 0.7。

4.3. 评价指标介绍

对于以往的实例分割模型而言，评价指标主要是看精确率(Precision)、召回率(Recall)和 mIoU，mIoU 是每个类别中真实 Mask 区域 A 和预测 Mask 区域 B 覆盖像素的交集比上并集的平均值。先给 IoU 设定一个 0.5 的阈值，即每个预测的 IoU 大于这个值则认为预测正确，小于这个值则表示预测错误。

$$\mathrm{IoU} = \frac{A \cap B}{A \cup B} \tag{15}$$

机器学习中有四个指标 TP (True Positives)表示预测为正样本且实际上也是正样本的数量、FP (False Positives)表示预测为正样本然而实际上却是负样本的数量、FN (False Negatives)表示预测为负样本然而实际上却是正样本的数量、TN (True Negatives)表示预测为负样本而且实际上也是负样本的数量。精确率和召回率可以表示为：

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{16}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{17}$$

提高 IoU 的阈值，那么预测为正确的样本里面为正样本的概率越高，但是正样本的数量为减小，所以精确率(Precision)会上升，召回率 recall 下降。IoU 减少阈值，召回率会上升，精确率下降。可以根据 IoU 阈值从 0 到 1 的增加绘制一条 xy 轴分别是召回率和精确率的曲线，即 P-R curve (精确率﹣召回率曲线)。P-R 曲线包围的面积越大说明模型的性能越好，该面积称为 AP (平均精确度)，对于所有类别都有平均精确度，mAP (mean Average Precision)表示所有类别的平均精确度。本文将 mAP50 作为性能评估指标，即在 IoU 为 0.5 的阈值下的 mAP 值。

$$\mathrm{mAP} = \frac{\sum_{i=1}^{K} \mathrm{AP}_i}{K} \tag{18}$$

4.4. 消融实验

为了对比不同改进方案的效果，通过控制变量法验证不同改进模块对模型性能的贡献，且引入了 GFLOPs 计算复杂度作为参考，在不大幅度的增加 GFLOPs 的情况下，模型的性能提升是值得且有效的。从基准模型 YOLOv8n-seg 开始实验，分别添加 NWD 损失函数改进及 C2f_DLKA 模块，以及本文的最终模型 NWD + DLKA 改进方式。从表 1 中可以看到，基准模型的平均精度(mAP@0.5)为 52.7%，对于单独一项的改进，NWD 和 DLKA 分别有 2.7 个百分点和 6 个百分点不同幅度的性能提升。同时加入这两项改进之后，模型性能显著提升。mAP@0.5 由基准模型的 52.7%提升至 60.8%，且 GFLOPs 仅增加 1.3。这一实验数据验证了将两者改进结合应用于 YOLOv8-seg 模型中对于平均精度的提升是极为有效的。

Table 1. Ablation experiment data
表 1. 消融实验数据

模型	mAP@0.5	GFLOPs
YOLOv8n-seg	52.7%	12.1
+NWD	55.4%	12.1
+DLKA	58.7%	13.4
+NWD+DLKA	60.8%	13.4

4.5. 数据结果分析

如图 8 所示为实验测试集平均精度 PR 曲线图，横坐标表示召回率，纵坐标表示精确率。一共有 7 条曲线表示每个类别精确率和召回率因 IoU 变化的对应关系，每条曲线下方和 XY 轴围成的面积即为 AP 平均精度。其中汽车部分裂纹(Car-Part-Crack)类别平均精度相对最低为 0.133，瘪胎(Flat-Tire)类别平均精度最高为 0.929，所有类别平均精度为 0.608。

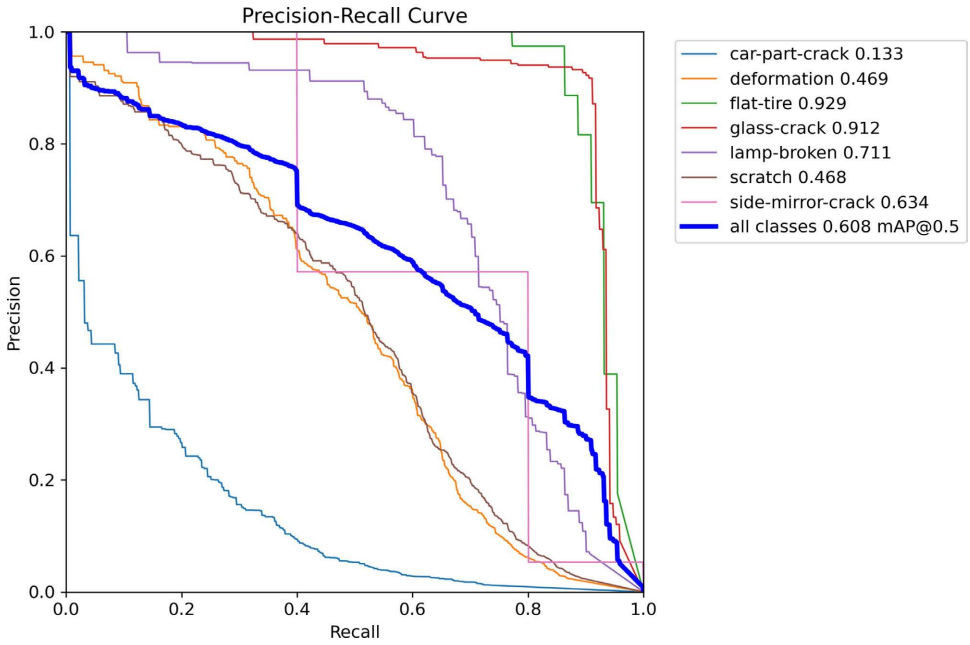

Figure 8. P-R curve
图 8. P-R 曲线图

如图 9 所示为损失计算值随训练轮次变化的曲线，已知训练轮次为 500 轮。从左到右依次是边框损失、分割损失、分类损失以及分布焦点损失。由图可知，训练集损失下降放缓，验证集损失实现稳定，算法趋于拟合。

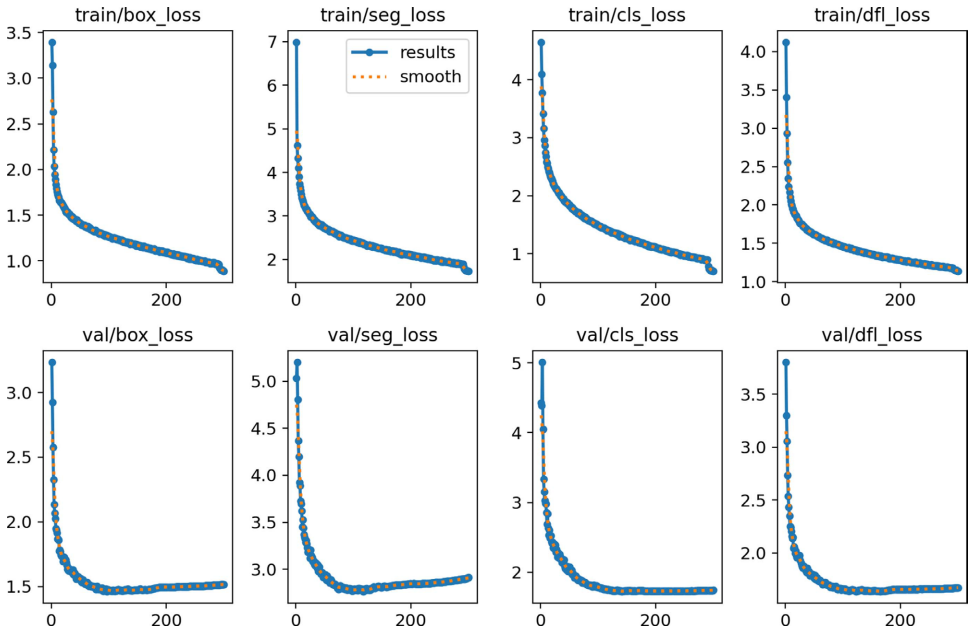

Figure 9. Loss trend
图 9. 损失函数曲线图

如图 10 所示 mAP 平均精度在 200 轮后逐渐趋于稳定，模型实现收敛。

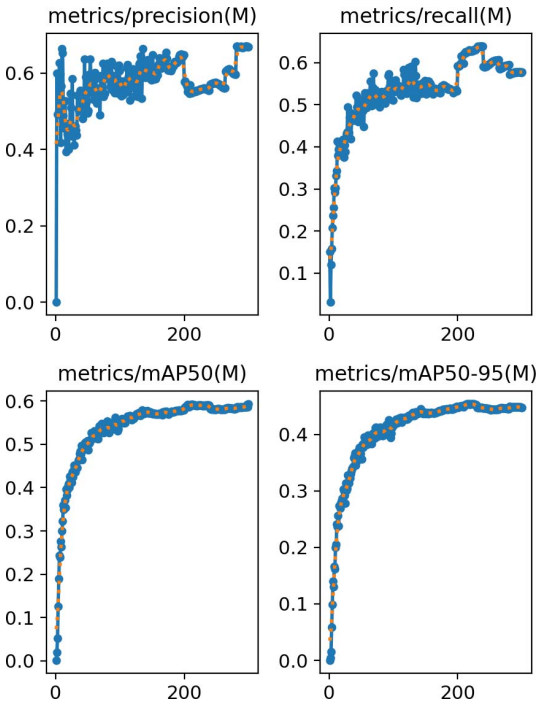

Figure 10. Index result
图 10. 指标结果图

图 11 为模型效果图，可以看到改进后的模型有效地实现了车辆损伤的分类以及损伤部位的 Mask 分割，展现了模型的实用性。

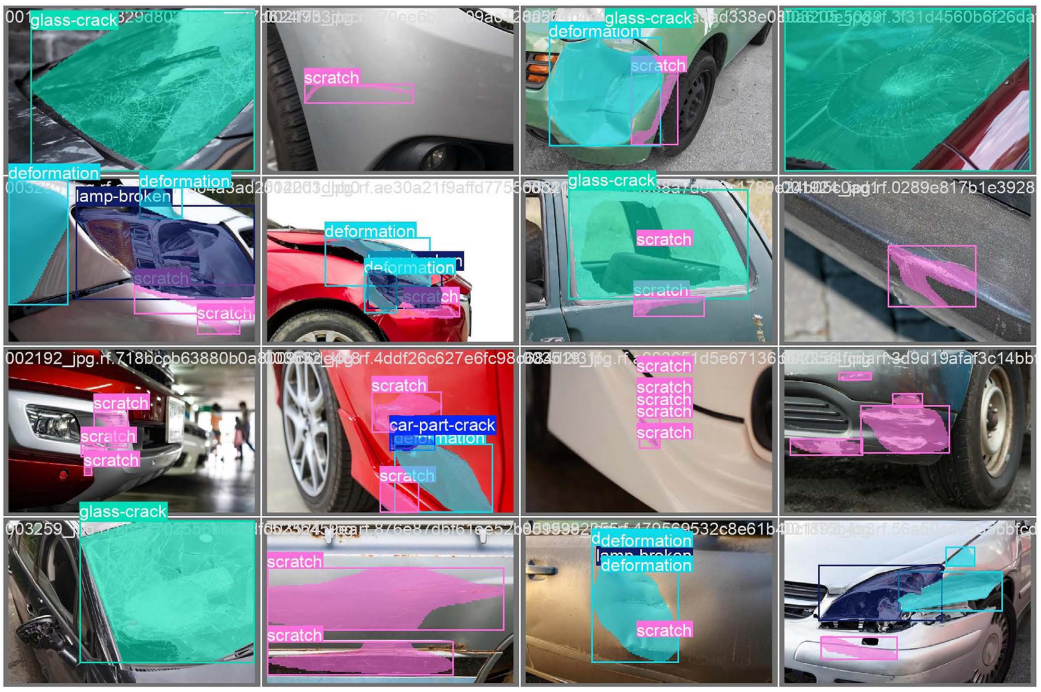

Figure 11. Model renderings
图 11. 模型效果图

5. 总结与展望

本文提出了一种基于改进 YOLOv8-seg 算法的车辆智能定损模型，实现了对有损车辆数据集中车辆外观损伤部位的精确区域分割以及损伤类型的识别。通过融合可变形大核卷积注意力模块以及 NWD 分割损失函数的替换。模型的计算复杂度略微增加，GFLOPs 从 12.1 增至 13.4。但实现了相比于 YOLOv8 自带的原始分割算法 13.3%的性能提升，平均精度 60.8%，提升了约 8.1%。有效地减少了损伤类型的识别错误和有损部位的错误划分，为车辆智能定损领域提供了有效的方案。相比基模型有更强的泛化性和鲁棒性。未来的研究可以集中于模型的轻量化、剪枝蒸馏以及嵌入式部署并进一步提高精度，实现快速且高精度的现场实时智能定损。

参考文献

[1] Gontscharov, S., Baumgärtel, H., Kneifel, A. and Krieger, K. (2014) Algorithm Development for Minor Damage Identification in Vehicle Bodies Using Adaptive Sensor Data Processing. *Procedia Technology*, **15**, 586-594. https://doi.org/10.1016/j.protcy.2014.09.019

[2] Jayawardena, S. (2013) Image Based Automatic Vehicle Damage Detection. Ph.D. Thesis, The Australian National University.

[3] 黄爱民. 射线检测技术在无损检测中的应用[J]. 山东工业技术, 2015, 14(1): 218-219.

[4] 陈春谋. 基于直方图均衡化与拉普拉斯的铅条图像增强算法[J]. 国外电子测量技术, 2019, 38(7): 131-135.

[5] Rong, W., Li, Z., Zhang, W. and Sun, L. (2014) An Improved Canny Edge Detection Algorithm. 2014 *IEEE International Conference on Mechatronics and Automation*, Tianjin, 3-6 August 2014, 577-582. https://doi.org/10.1109/icma.2014.6885761

[6] Dorathi Jayaseeli, J.D., Jayaraj, G.K., Kanakarajan, M. and Malathi, D. (2021) Car Damage Detection and Cost Evaluation Using MASK R-CNN. In: Peng, S.L., Hsieh, S.Y., Gopalakrishnan, S., Duraisamy, B., Eds., *Lecture Notes in Networks and Systems*, Springer, 279-288. https://doi.org/10.1007/978-981-16-3153-5_31

[7] Wang, X., Li, W. and Wu, Z. (2023) CarDD: A New Dataset for Vision-Based Car Damage Detection. *IEEE Transactions on Intelligent Transportation Systems*, **24**, 7202-7214. https://doi.org/10.1109/tits.2023.3258480

[8] van Ruitenbeek, R.E. and Bhulai, S. (2022) Convolutional Neural Networks for Vehicle Damage Detection. *Machine Learning with Applications*, **9**, Article 100332. https://doi.org/10.1016/j.mlwa.2022.100332

[9] Lee, D., Lee, J. and Park, E. (2024) Automated Vehicle Damage Classification Using the Three-Quarter View Car Damage Dataset and Deep Learning Approaches. *Heliyon*, **10**, e34016. https://doi.org/10.1016/j.heliyon.2024.e34016

[10] Jocher, G., Chaurasia, A. and Qiu, J. (2023) Ultralytics YOLO (Version 8.0.0) [Computer Software]. https://github.com/ultralytics/ultralytics

[11] 谢东升. 基于深度学习的车辆智能定损算法研究[D]: [硕士学位论文]. 天津: 天津大学, 2019.

[12] 金浩然. 面向智能辅助定损的车辆外观损伤识别方法研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2023.

[13] Xu, C., Wang, J., Yang, W., Yu, H., Yu, L. and Xia, G. (2022) Detecting Tiny Objects in Aerial Images: A Normalized Wasserstein Distance and a New Benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, **190**, 79-93. https://doi.org/10.1016/j.isprsjprs.2022.06.002

[14] Azad, R., Niggemeier, L., Hüttemann, M., Kazerouni, A., Aghdam, E.K., Velichko, Y., *et al*. (2024) Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image Segmentation. 2024 *IEEE/CVF Winter Conference on Applications of Computer Vision* (*WACV*), Waikoloa, 3-8 January 2024, 1276-1286. https://doi.org/10.1109/wacv57701.2024.00132