arXiv:2303.04275v1 [cs.CV] 7 Mar 2023

一种基于Transformer预测头改进YOLOv5的计算机视觉损坏检测模型

Arunabha M. Roy[1] 和 Jayabrata Bhaduri[2]

[1]航空航天工程系，密歇根大学，安娜堡，*MI 48109*，美国

[2]*Capacloud AI*，深度学习 & 数据科学部，加尔各答，*WB 711103*，印度。

摘要

目标。 基于计算机视觉的最新精准损伤分类与定位，对于基础设施监测、土木基础设施的安全与适用性至关重要。然而，目前最先进的基于深度学习（DL）的损伤检测模型，往往缺乏在复杂且嘈杂环境中的卓越特征提取能力，限制了精准可靠目标区分的发展。方法。 为此，我们提出了 DenseSPH-YOLOv5，这是一种基于深度学习的实时高性能损伤检测模型。该模型将 DenseNet 块集成到骨干网络中，以改进对关键信息的保留与复用。此外，我们采用了卷积块注意力模块（CBAM），以提升注意力机制的性能，增强深层空间特征提取能力与判别能力，从而在各种复杂环境下实现卓越检测。此外，我们还添加了额外的特征融合层和 Swin-Transformer预测头（SPH），利用先进的自注意力机制，更高效地检测多尺度目标尺寸，同时降低计算复杂度。结果。 在大规模道路损伤数据集上评估模型性能。

（RDD-2018）中，DenseSPH-YOLOv5 的检测速度为 62.4 帧/秒，平均精度均值（mAP）为 85.25%，F1分数为 81.18%，精确率（P）为 89.51%，优于当前最先进模型。意义。本研究提出了一种有效且高效的损伤定位模型，通过提供高精度的边界框预测，弥补了现有基于深度学习的损伤检测模型的不足。本研究是迈向能够应用于实时现场场景的准确、稳健自动化损伤检测系统的一步。

关键词：自动化损伤检测；You Only Look Once（YOLOv5）算法；Swin Transformer目标检测（OD）；计算机视觉；深度学习（DL）

1. 引言：

近年来，自动化损伤检测在各种工业应用中发挥着重要作用，包括产品质量评估（Agarwal 和 Singh，2015；Hanzaei 等人，2017）、基础设施监测（Eisenbach 等人，2017；Gopalakrishnan，2018），以及土木基础设施的安全性和适用性（Koch 等人，2015；Hartmann 和 Trappey，2020）。早期阶段的准确裂缝检测对于路面损伤评级以及后续的封缝或修复活动至关重要（Chen 等人，2021；Chen 和 Cho，2022）。因此，道路基础设施工程师准确检测路面裂缝十分重要，从而能够制定并实施最具成本效益的养护和修复方案（Ni 等人，2022；Dong 等人，2021）。传统的损伤检测技术主要包括目视检查，但这类劳动密集型方法存在效率低、成本高以及个人偏差等缺点（Xu 等人，2022；Shang 等人，2023）。此外，由于需要具备特定领域经验、知识和技能的合格人员，这些方法在可重复性、可靠性和客观性方面也受到限制（Fang 等人，2020）。为解决这些问题，近年来，各种自动化和半自动化裂缝检测算法受到了广泛关注（Gopalakrishnan，2018）。

在过去三十年中，基于图像的裂缝检测（Mohan and Poobal，2018；Koch 等人，2015）一直是研究重点，其中涵盖了多种图像处理方法，例如边缘检测（Zhao

等人，2010年；Hanzaei 等人，2017；Li 等人，2022年）、动态阈值处理（Oliveira 和 Correia，2009；Wang 等人，2021b）以及各种形态学操作（Anitha 等人，2021；Li 和 Zhao，2021），这些方法一直是复杂现实场景中损伤检测的核心研究重点。然而，上述方法对噪声和不断变化的光照强度相当敏感，因此不适用于复杂的现实环境（Koch 等人，2015）。为解决这些问题，后来引入了传统机器学习（ML）方法进行损伤检测（Mohan 和 Poobal，2018；Koch 等人，2015）。通常，这类方法会使用经过训练的分类器（如支持向量机（SVM）），并基于局部特征描述符进行分类；这些描述符可以是局部二值模式（LBP）（Varadharajan 等人，2014；Quintana 等人，2015），也可以是方向梯度直方图（HOG）（Kapela 等人，2015）。尽管与传统图像处理方法相比，基于机器学习的模型显著提高了损伤检测的准确率和效率，但分类性能存在较大误差，仍然是此类模型在现实应用中部署的一大瓶颈（Fang 等人，2020）。

近年来，以多层神经网络（NN）为特征的深度学习（DL）（LeCun 等，2015）在模式识别领域取得了显著突破，应用于图像分类（Rawat and Wang, 2017; Jamil et al., 2022; Khan et al., 2022b,a）、计算机视觉（Voulodimos et al., 2018; Roy and Bhaduri, 2021; Roy et al., 2022c; Roy and Bhaduri, 2022; Roy et al., 2022a）、目标检测（Zhao et al., 2019a; Chandio et al., 2022; Roy et al., 2022b; Singh et al., 2023a）、脑机接口（Roy, 2022b,a,c; Singh et al., 2023b）、信号分类（Jamil and Roy, 2023, 2022）以及各种科学学科（Bose and Roy, 2022; Roy and Bose, 2023b; Roy and Guha, 2022; Roy and Bose, 2023a; Roy and Guha, 2023）。随着这一领域的发展，越来越多的研究开始运用深度学习技术开展损伤分类任务，其中大多采用卷积神经网络（CNN），例如残差网络（Bang et al., 2018）、AlexNet（Dorafshan et al., 2018; Li et al., 2018）、VGG网络（Gopalakrishnan et al., 2017; Silva and Lucena, 2018）以及其他多种网络（Chow et al., 2020; Nath et al., 2022; Li et al., 2021）。尤其是在目标定位方面，深度学习方法展现出更高的准确率（Han et al., 2018），可分为两类：两阶段检测器和单阶段检测器（Lin et al., 2017a）。两阶段检测器包括区域卷积神经网络（R-CNN）（Girshick, 2015）、Faster R-CNN（Ren

等，2016）、Mask R-CNN（He et al., 2017）等，它们显著提高了目标定位的准确率。近年来，You Only Look Once（YOLO）的各种变体（Redmon et al., 2016; Redmon and Farhadi, 2017, 2018; Bochkovskiy et al., 2020）相继提出，将目标分类与定位统一起来。在（Roy et al., 2022c; Roy and Bhaduri, 2022, 2021）中，研究者将DenseNet（Huang et al., 2017）模块与空间金字塔池化（SPP）（He 等人，2015）结合，并集成了改进的路径聚合网络（PANet）（Liu et al., 2018），以增强原始YOLOv4中感受野的表征能力和重要上下文特征的提取能力，从而显著提升检测速度和准确率。为提升梯度性能并降低计算成本，YOLOv4（Bochkovskiy et al., 2020）设计了跨阶段部分（CSP）网络。为进一步提高检测准确率，YOLOv4采用Mish激活函数（Misra, 2020）和CIoU损失（Zheng et al., 2020）。近期，Scaled-YOLOv4（Wang et al., 2021a）在检测各种应用中大量线性缩放目标方面展现出卓越性能。作为YOLO系列的最新一代，YOLOv5（Jocher et al., 2021）继承了上述所有优势，被评为最先进目标检测模型中的佼佼者。因此，本研究将YOLOv5作为多类别损伤检测的基准模型。近期，有研究提出基于Transformer预测头的改进型YOLOv5（TPH-YOLOv5）（Zhu et al., 2021），将卷积块注意力模块（CBAM）（Woo et al., 2018）集成其中，用于密集目标检测；此外，还设计了融入基于归一化的注意力模块（NAM）的Swin Transformer增强型YOLOv5（SPH-YOLOv5）（Gong et al., 2022）。这些方法在显著提高准确率的同时降低了模型的计算复杂度，这也是当前工作开展网络架构设计的主要动机。

1.1 相关研究：

本节重点介绍了道路损伤检测领域近期一些相关的深度学习研究。近年来，多项研究采用了各种基于机器学习和深度学习的方法，实现路面损伤分类与检测自动化（Zhang 等人，2017a；

Stricker 等人，2019；Bi.cici 和 Zeybek，2021）例如，有研究提出了一种基于智能手机的监督式深度卷积神经网络（D-CNN），用于路面损伤分类（Zhang 等人，2016）。类似地，深度神经网络（DNN）架构已被用于检测裂缝和坑洞（Anand 等人，2018；Silva 和 Lucena，2018），以及评估路面状况（Fan 等人，2018）。Nhat-Duc 等人（2018）证明，在路面裂缝检测方面，基于深度卷积神经网络的方法优于基于边缘检测的方法。Maeda 等（2018）提出了一种基于单次多框检测器（SDD）的实时道路损伤检测模型，该模型使用公开的大规模道路损伤数据集（RDD-2018）进行训练，涵盖八种不同类别的道路损伤。由于该数据集广受欢迎，研究者尝试采用多种方法进一步提升检测性能，其中包括 YOLO（Alfarrarjeh 等人，2018）、Faster R-CNN（Kluger 等人，2018）、采用残差网络-152的 Faster R-CNN（Wang 等人，2018a）、结合 Faster R-CNN 和 SSD 的集成模型（Wang 等人，2018b），以及 RetinaNet（Angulo 等人，2019）。此外，RetinaNet（Angulo 等人，2019）还被用于经过修改的 RDD-2018 数据集，并显著提升了性能。继 Maeda 等（2018）的研究之后，研究者采用结合泊松融合的渐进式生长生成对抗网络（PG-GANs）（Maeda 等人，2021）生成新的训练数据（RDD-2019），以提高道路损伤检测的准确率。近来，研究者提出了基于迁移学习（TL）的道路损伤检测模型（Arya 等人，2021a），并引入了考虑多个国家情况的大规模开源数据集 RDD2020（Arya 等人，2020，2021b）。Naddaf-Sh 等人（2020）采用高效检测器-D7检测沥青路面病害。另一方面，Mandal 等人（2020）使用 YOLO CSPDarknet53 进行道路损伤检测。同样，YOLO网络也被用于从高分辨率图像中检测路面病害（Du 等人，2021）。Majidifard 等人（2020）使用 YOLOv2 模型，根据谷歌街景图像对路面病害进行分类。类似地，Patra 等人（2021）采用了一个基于 Google API 图像训练的卷积神经网络预测模型，用于检测坑洞。在另一项研究中，Guan 等人（2021）部署了一种集成立体视觉的基于分割的深度学习模型，并采用改进的深度可分离卷积U-Net进行裂缝和坑洞检测；该模型使用多特征图像数据集进行训练。近期，一种基于半监督深度学习的像素级分割

模型（Karaaslan 等人，2021年）提出了一种利用注意力引导来定位裂缝和剥落的方法，可显著降低计算成本。另一项独立研究提出了一种改进的 YOLOv5 道路损伤检测算法（Guo 和 Zhang，2022年），该算法采用 MobileNetV3 作为骨干特征提取器。在 Hacıefendioˇglu 和 Basaˇga （2022年）的研究中，Faster R-CNN 被用于在各种光照和天气条件下进行混凝土路面裂缝检测。尽管目前已有多项损伤检测领域的先进研究，包括多类别损伤定位模型，但这些方法往往存在准确率低、漏检以及计算开销相对较大的问题（Cao 等人，2020年；Azimi 等人，2020年；Naddaf-Sh 等人，2020年）。

1.2 研究动机：

尽管当前先进的深度学习算法在损伤检测方面表现出色，但仍需进一步改进，因为其细粒度上下文特征提取能力不足，导致对具有多种纹理、形状、尺寸和颜色的各类损坏/裂缝出现漏检和错误目标预测（Cao 等人，2020年；Azimi 等人，2020年；Naddaf-Sh 等人，2020年）。不同损伤类别之间，由于光照条件变化显著、能见度低、具有不同宽高比的多种物体类别共存，以及其他形态特征的影响，准确检测和定位颇具挑战（Azimi 等人，2020年；Naddaf-Sh 等人，2020年）。此外，视觉相似性、复杂背景以及其他各种关键因素，也给先进的损伤检测模型带来了额外困难（Naddaf-Sh 等人，2020年）。因此，目前的研究旨在同时开发高效、稳健的损伤分类模型和准确的损伤定位模型，并兼顾训练时间和计算成本；而近期的先进研究尚未解决这一问题。

1.3 研究贡献：

为解决上述不足，本研究提出了 DenseSPH-YOLOv5，该模型基于改进后的先进 YOLOv5 检测模型，可实现精准的实时

损伤检测。本研究的主要贡献可概括如下：

- 在 Dense-SPH-YOLOv5 中，我们将 DenseNet 块与 CSPDarknet53 中的 CSP模块相结合，以保留关键特征图，并高效复用具有判别性的特征信息。

- 其次，我们在所提出的 DenseSPH-YOLOv5 网络的检测头部分加入了一个附加检测头，专门用于检测微小目标。

- 此外，我们采用了卷积块注意力模块（CBAM），在通道和空间维度上构建覆盖范围广的渐进式特征注意力，以便在目标检测过程中实现强有力且具有判别性的深层空间特征提取。

- 此外，我们将 YOLOv5 中常规的 CNN预测头（CPH）升级为 Swin Transformer预测头（SPH）。这些预测头采用了 Swin Transformer（STR）编码器块，并利用先进的自注意力机制，高效检测多尺度目标尺寸，同时降低计算复杂度。

- 我们将空间金字塔池化（SPP）紧密连接到骨干网络，以增强感受野的表征能力并提取重要的上下文特征。

- 最后，利用改进的路径聚合网络（PANet），通过在多尺度特征金字塔图中进行特征融合，高效保留细粒度定位信息。

通过上述改进，模型在多尺度目标检测方面的检测准确率显著提高。为同时优化检测准确率和检测速度，我们针对不同的骨干网络-颈部架构组合进行了广泛的消融研究。所提出的 DenseSPH-YOLOv5 用于检测八种不同的损伤类别，在各种复杂且具有挑战性的环境下均能实现出色而准确的检测。本研究有效弥补了现有基于深度学习的裂纹检测模型的不足，并展现了其

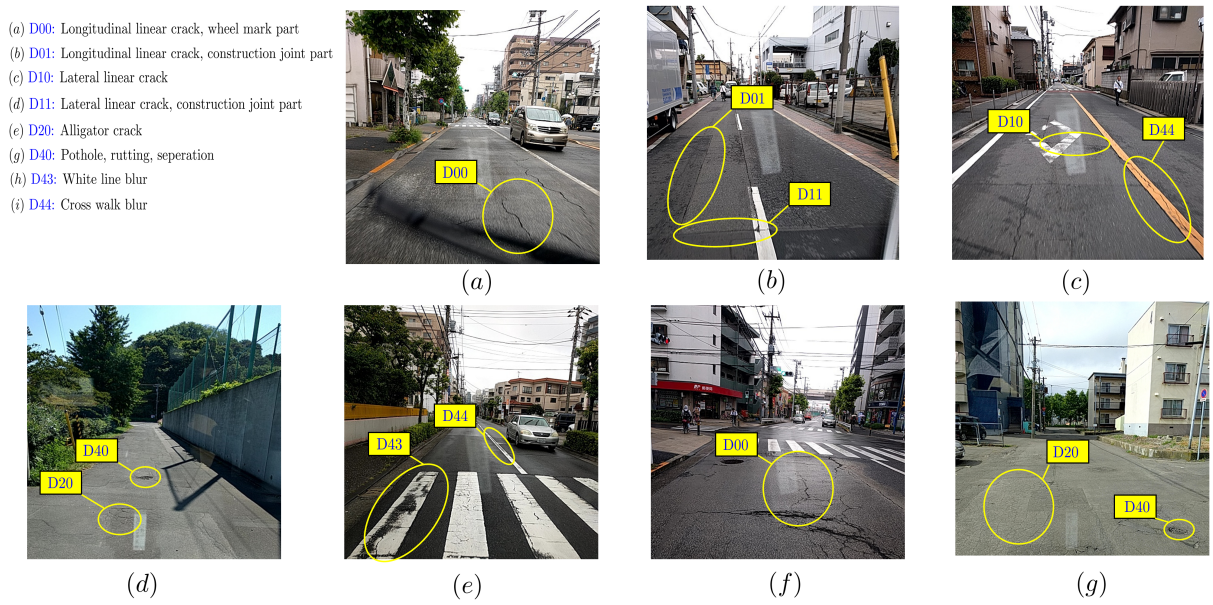

图 1：RDD-2018 数据集中的样例图像（Maeda 等人，2018）：(a) 至 (g) 分别对应图例中的八个类别。

在实时现场应用中的巨大潜力。本文其余部分安排如下：第 2 节介绍数据集；第 3 节阐述所提出的损伤检测方法；第 4 节和第 5 节分别讨论所提出模型的相关研究发现及讨论。最后，第 6 节总结本研究的结论。

2. 道路损伤数据集：

本研究使用了大规模道路损伤数据集（RDD-2018）（Maeda 等人，2018），该数据集包含 9,053 张已标注的道路损伤图像，分辨率为 600×600 像素，共包含 15,435 个标注边界框，涵盖八种不同类型的道路损伤。为避免偏差，这些图像是在日本不同地区（包括千叶、室兰、市原、隅田、长久手和沼津）于各种天气及光照条件下拍摄的。标注过程中，聘请了道路损伤方面的专业人员来核验各种损伤类别，以确保数据集的可靠性。表 1列出了各种损伤类型及相应的类别标识符。每种

表1：RDD-2018 数据集中各种道路损坏类型及其对应的类别标识符（Maeda 等人，2018）。

类别标识符	损坏 type	走向	详情
D00	线性裂缝	纵向	车辙部分
D01	线性裂缝	纵向	施工缝部位
D10	线性裂缝	横向	等间距
D11	线性裂缝	横向	施工缝部位
D20	龟裂	-	局部路面，整体路面
D40	其他裂缝	-	坑槽、车辙、脱落
D43	其他裂缝	-	白色标线模糊
D44	其他裂缝	-	人行横道模糊

损坏类型如图 1所示。首先，损坏被分为裂缝或其他破坏类型。随后，裂缝又分为线性裂缝和鳄鱼裂缝。而其他破坏类型则包括坑洞和车辙，以及白色标线模糊等其他道路损坏类别。

3. 损伤检测的拟议方法：

在目标检测中，会同时执行目标对象分类和定位，将目标类别分类并与背景区分开来。目标定位旨在通过在包含完整目标的输入图像上绘制边界框（BBs）来确定目标的位置。这对于准确调查濒危物种数量尤其有用。为此，本研究的主要目标是开发一种高效、稳健的损伤分类与精准损伤定位模型。在这方面，YOLO 的不同变体（Redmon 等人，2016年；Redmon 和 Farhadi，2017，2018；Bochkovskiy 等人，2020年）是精度较高的单阶段目标检测模型中的佼佼者。近来，

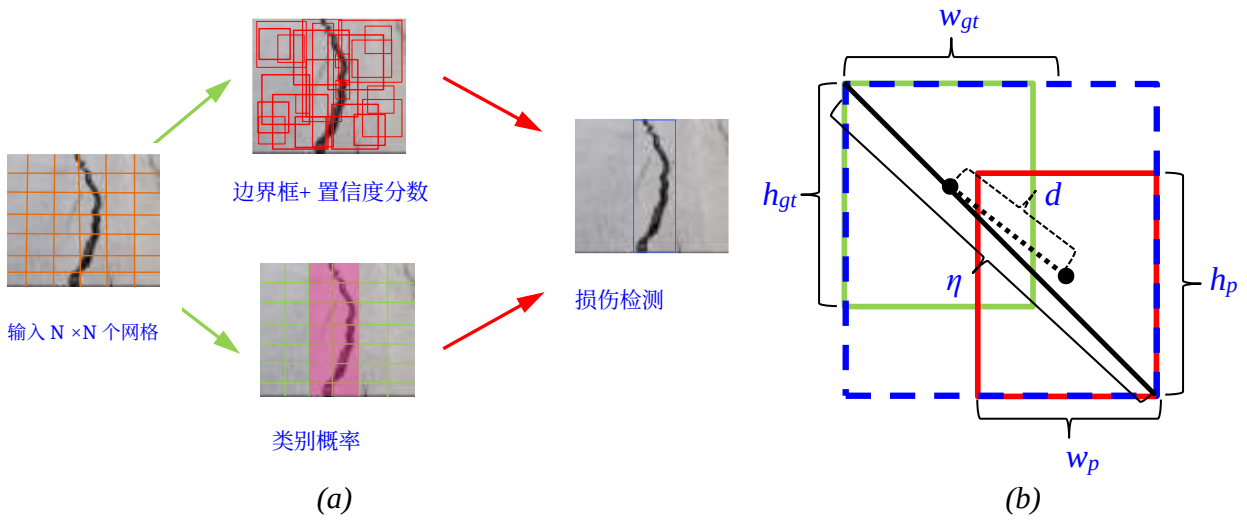

图 2: （a）YOLO 损伤定位目标定位过程示意图；（b）目标边界框预测的 CIoU 偏移回归示意图。

YOLOv5（Jocher 等人，2021年）已被提出，目前具有最佳检测性能，并包含四种不同的模型变体：YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x，这些变体的模型宽度和深度各不相同。通常，YOLOv5 的整体架构由以下部分组成：用于深层特征提取的骨干网络、用于汇集语义特征融合的颈部网络，以及用于目标分类和定位的头部网络。YOLOv5 的原始版本采用 CSPDarknet53（Wang 等人，2020年；Bochkovskiy 等人，2020年）与 SPP 和 PANet（Liu 等人，2018）分别作为骨干网络和颈部网络。而 YOLO 检测头（Redmon 等人，2016年）则被用于检测头。YOLO 模型通过生成边界框坐标和各类别的概率，将目标检测任务转化为回归问题，如图 2所示。在此过程中，输入图像尺寸被均匀划分为 $N \times N$ 个网格，并在其中生成B 预测边界框。随后，如果目标对象落入特定网格，就会为其分配一个置信度分数。当真实标注的中心位于指定网格内时，模型便会检测出相应类别的目标对象。在检测过程中，每个网格会预测 N_B 个边界框，其置信度值 Φ_B 为：

$$\Phi_B = \mathcal{P}_r(obj) \times \mathrm{IoU}_p^t \qquad \vee\ \mathcal{P}_r(obj) \in 0, 1 \tag{1}$$

其中， $\mathcal{P}_r(obj)$推断边界框预测的准确率，即 $\mathcal{P}_r(obj) = 1$ 表示目标类别位于网格内，否则， $\mathcal{P}_r(obj) = 0$。真实标注与预测边界框之间的重叠程度由尺度不变评估指标交并比（IoU）描述，可表示为

$$\mathrm{IoU} = \frac{\mathbf{B}_p \cap \mathbf{B}_{gt}}{\mathbf{B}_p \cup \mathbf{B}_{gt}} \tag{2}$$

其中，$\mathbf{B}^{gt}$ 和$\mathbf{B}^{p}$ 分别为真实标注和预测边界框。

3.1 边界框回归中的损失：

为进一步改善边界框回归并缓解梯度消失问题，研究者引入了广义交并比（GIoU）（Rezatofighi 等人，2019年）和距离交并比（DIoU）（Zheng 等人，2020年），并考虑了重叠边界框的宽高比和方向。近期，为提高边界框预测的准确率并加快收敛速度，研究者提出了完全交并比（CIoU）（Zheng 等人，2020年），其可表示为

$$\mathcal{L}_{\mathrm{CIoU}} = 1 + \beta\xi + \frac{\alpha^2(\mathbf{b_p}, \mathbf{b_{gt}})}{\eta^2} - \mathrm{IoU} \tag{3}$$

$$\xi = \frac{4}{\pi^2}\left(\tan^{-1}\frac{w_{gt}}{h_{gt}} - \tan^{-1}\frac{w_p}{h_p}\right)^2; \quad \beta = \frac{\xi}{(1-\mathrm{IoU}) + \xi'} \tag{4}$$

其中，$\mathbf{b}^{gt}$ 和 $\mathbf{b}^{p}$ 分别表示 $\mathbf{B}^{gt}$ 和 $\mathbf{B}^{p}$的质心； ξ 和 β 分别为一致性参数和权衡参数。如图 3 - (b)所示， η 是 $\mathbf{B}_p \cup \mathbf{B}^{gt}$的最短对角线长度； w_{gt}、 w_p 分别为 $\mathbf{B}^{gt}$ 和 $\mathbf{B}^{p}$的宽度， h_{gt}、 h_p 分别为其高度。随着 w_p/h_p的增加，根据式 4 可得 $\xi \to 0$ 。因此，为了优化 ξ 对 CIoU 的影响，可为检测模型适当选择 w_p/h_p 。最后，可通过对不同尺度应用非极大值抑制（NMS）算法（Ren 等，2016）获得最佳边界框预测。在 YOLO 框架的目标检测中，总损失函数 Δ_l 由边界框坐标预测误差 Δ_{cor}、交并比误差 Δ_{IoU}和分类误差项 Δ_{cl}组成。损失

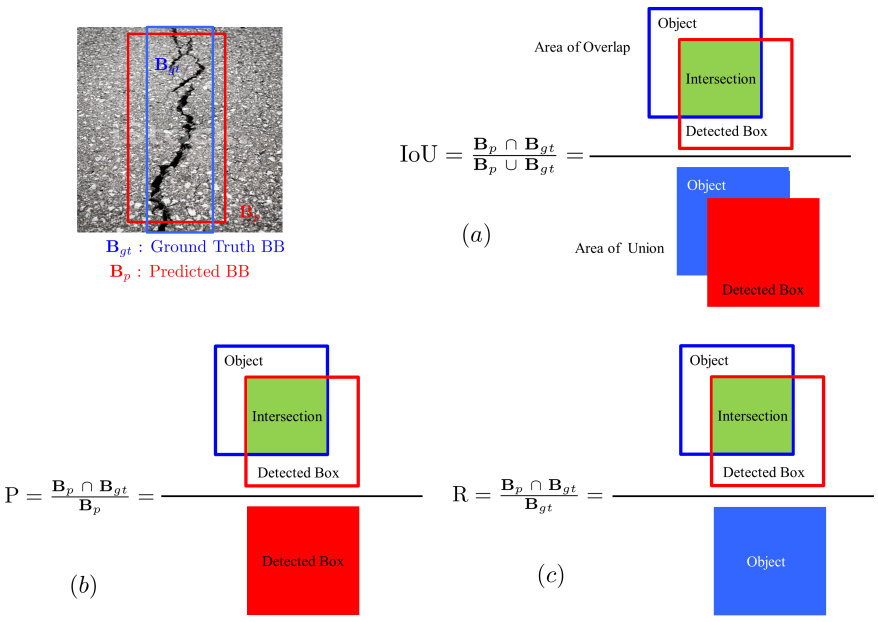

图 3：测量各种性能指标的示意图：(a) 交并比（IoU）；(b) 精确率（P）；(c) 损伤检测过程中的召回率（R）。

用于边界框回归的函数 Δ_l 可表示为：

$$\Delta_l = \Delta_{cor} + \Delta_{cl} + \Delta_{IoU}. \tag{5}$$

3.2 性能指标：

本研究采用常见的标准指标评估目标检测模型的性能（Ferri 等人，2009），包括平均精确率（平均精度）、精确率（精确率）、召回率（召回率）、交并比、F1 分数、平均精度均值等。评估过程中生成的混淆矩阵对测试结果作出如下分类：真正例、假正例、假负例和真负例。在目标分类中，当交并比 ≥ 0.5时，可将分类出的目标定义为真正例；而当交并比 < 0.5时，则可将其分类为假正例。根据上述定义，分类器的精确率

可根据其正确区分目标类别的能力来定义：

$$\mathrm{P} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})}; \tag{6}$$

目标类别被正确预测的比例称为分类器的召回率，可按以下公式计算：

$$\mathrm{R} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})} \tag{7}$$

精确率和召回率的值越高，检测能力越强。而 F1 分数是精确率和召回率的算术平均值，计算公式如下：

$$\mathrm{F1} = 2 \times \left(\frac{\mathrm{P}\ \mathrm{R}}{\mathrm{P}\ +\ \mathrm{R}} \right) \tag{8}$$

较高的 F1 分数表明检测模型具有较强的鲁棒性。性能指标平均精度可定义为精确率-召回率曲线下的面积（Davis和Goadrich，2006年），如下所示：

$$\mathrm{AP} = \int_0^1 \mathrm{P}\ (\mathrm{R})\, \mathrm{dR} \tag{9}$$

平均精度值越高，表明对各种物体类别的预测准确率越高。此外，$\mathrm{AP}_{50:95}$ 表示交并比 $=0.50$ 从 0.05 到 0.95 时的平均精度；AP_{50} 和 AP_{75} 分别表示交并比阈值为 50% 和 75% 时的平均精度。小物体、中等物体和大物体的检测平均精度分别可通过 AP_S、AP_M和 AP_L衡量。最后，平均精度均值可通过所有平均精度的平均值计算得出：

$$\mathrm{mAP} = \frac{1}{N_c} \sum_{i=1}^{N} \mathrm{AP}_i. \tag{10}$$

在YOLOv5中，可以采用包括Sigmoid函数、leaky-ReLU和SiLU在内的激活函数组合（Hendrycks和Gimpel，2016年），以提升模型针对特定检测任务的性能。此外，还可以采用免费赠品与特别技巧包（Bochkovskiy 等人，2020年）进一步优化YOLOv5的检测架构。作为YOLO系列的最新一代，YOLOv5已被证明能够提供业界领先的检测性能。因此，本研究选择其作为基线模型。

3.3 DenseSPH-YOLOv5架构：

近年来，人们已将多种基于深度学习的计算机视觉模型应用于损伤检测，例如Faster R-CNN（Kluger 等人，2018；Wang 等人，2018a）、SSD（Maeda 等人，2018；Wang 等人，2018b）、RetinaNet（Angulo 等人，2019）、YOLO（Alfarrarjeh 等人，2018；Mandal 等人，2020）、YOLOv2（Majidifard 等人，2020）、YOLOv5（Guo和Zhang，2022）等。尽管上述技术已展现出卓越的性能，但损伤检测任务仍面临诸多挑战，尤其是复杂且嘈杂的背景、光照条件的显著变化、低能见度、密集类别与重叠，以及具有不同宽高比和其他形态特征的多目标类别共存等问题（Azimi 等人，2020；Naddaf-Sh 等人，2020）。在这方面，与YOLOv4相比，YOLOv5能够同时提供更高的准确率和更快的检测速度，因此可作为一种合适的模型。业界领先的YOLOv5相对轻量但十分有效，通过多次下采样将浅层语义信息转换为高层语义信息，从而有效减少参数数量。然而，这不可避免地会损失语义信息，导致网络无法有效提取和融合具有判别性的语义特征。因此，原始YOLOv5网络的细粒度上下文特征提取能力不足，仍需进一步改进，否则会对具有广泛纹理、形状、尺寸和颜色范围的各种损坏/裂缝造成漏检和错误目标预测（Cao 等人，2020；Azimi 等人，2020；Naddaf-Sh 等人，2020）。为了在损伤检测中实现高准确率和高效率，我们基于业界领先的YOLOv5网络提出了一种新型目标定位算法DenseSPH-YOLOv5，以增强特征提取、保留细粒度定位信息并改进特征融合，从而在各种复杂环境下提供优异的损伤检测性能。DenseSPH-YOLOv5模型的整体网络如图4所示。为提升分类准确率和目标定位性能，我们开展了大量实验，并提出了多项改进，具体内容将在后续章节中详细介绍。

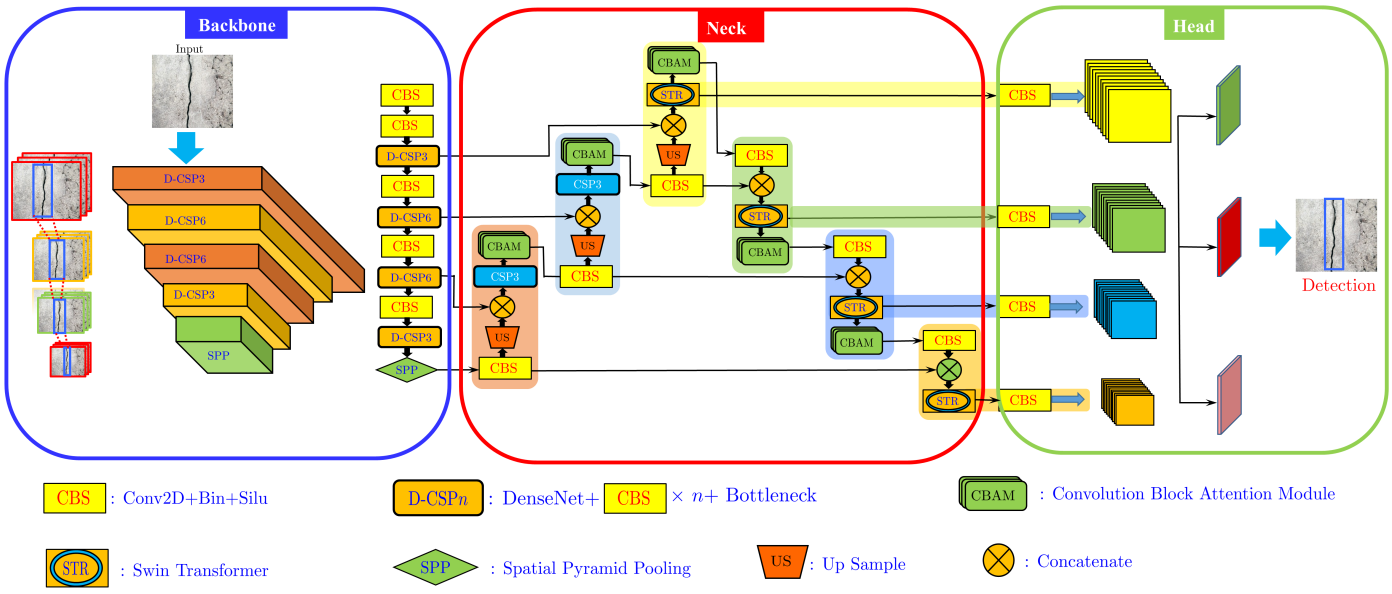

图4：所提出的 DenseSPH YOLOv5 示意图，其中 CSPDarknet53 骨干网络融合了 DenseNet，改进后的颈部网络中加入了 CBAM 模块；四个 SPH 利用改进后的颈部网络中 STR 编码器块提供的分层特征图。

3.3.1 DenseNet 块：由于 YOLOv5 通过卷积和下采样处理来缩减输入图像的特征图，这会导致特征传递过程中显著的语义特征损失。为此，我们在 YOLOv5 原有的 CSP Darknet53 中引入 DenseNet（Huang等人，2017），并将其连接到 CSP 模块，以保留关键特征图并高效复用判别性特征信息，如图 4所示。在 DenseNet 中，每一层都以 n前馈模式与其他层相连，第 ξ_n 层可以接收来自所有先前层的关键特征信息 ξ_0，ξ_1，...，ξ_{n-1}如下：

$$\xi_n = \Theta_n[\xi_0, \xi_1, ..., \xi_{n-1}] \tag{11}$$

其中 Θ_n 是第 n层的特征图函数。DenseNet块网络结构示意图如图 5-(d) 所示。通过大量实验，我们发现 DenseNet 能够改善特征传递，并缓解检测网络中的过拟合。因此，在所提出的 Dense-SPH YOLOv5 网络中，如图 3所示，我们引入了四个 DenseNet 块：第一个块（Dense B-1）连接在跨阶段部分连接块 CSP3 之前；第二个块（Dense B-2）置于 CSP6 之前；

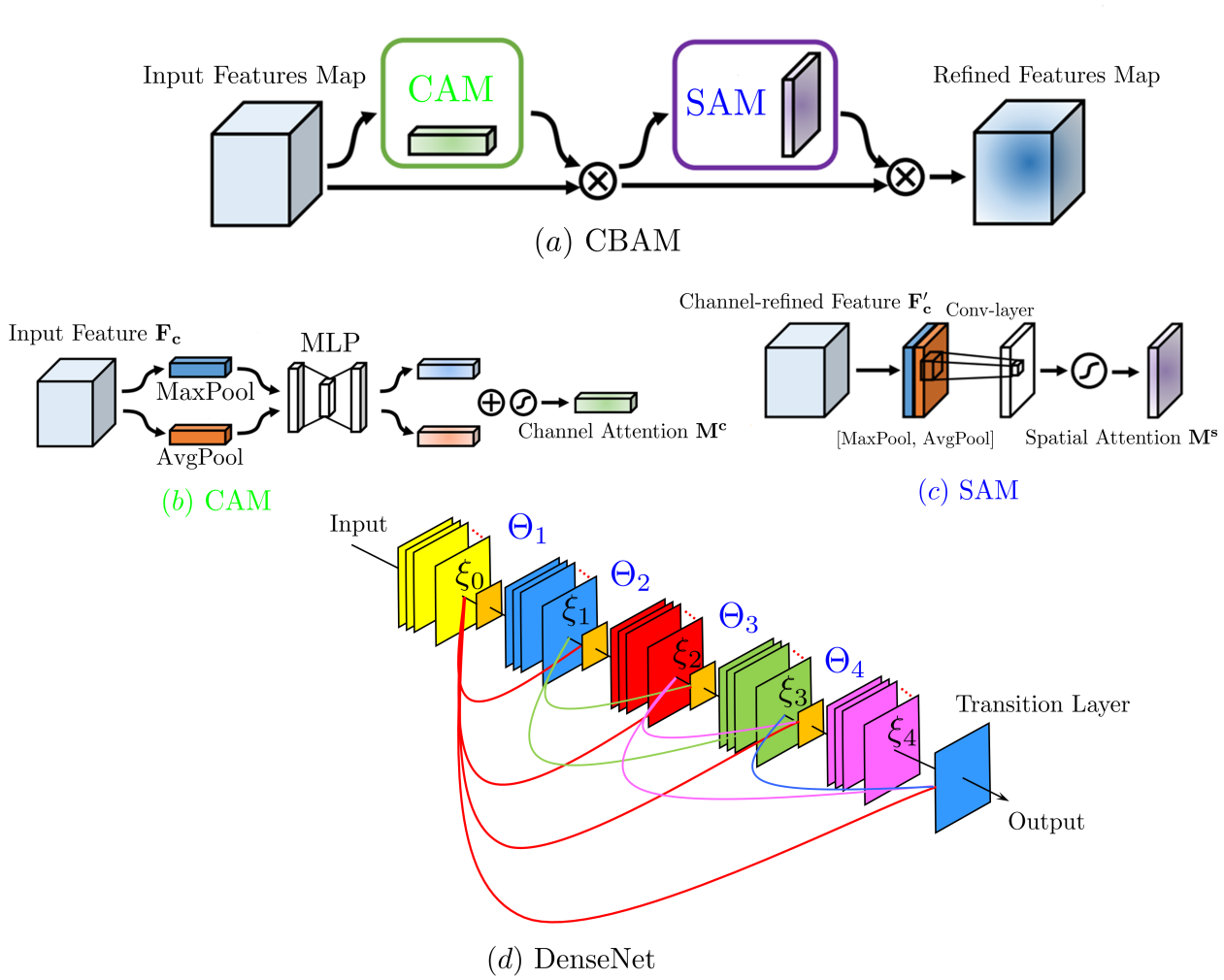

图 5：（a）CBAM 示意图，其包含两个顺序排列的子模块：（b）CAM 和（c）SAM，用于对每个中间卷积块中的特征图进行自适应细化；（d）Dense-SPH YOLOv5 检测模型中使用的 DenseNet 块网络架构。

而第三个（Dense B-3）和第四个（Dense B-4）块分别添加在 CSP6 和 CSP3 之前，从而增强了特征传播。此外，通过减少冗余的特征操作，Dense-CSPDarknet53 网络提高了计算速度。

3.3.2 卷积块注意力模块（CBAM）： 卷积块注意力模块（CBAM）（Woo 等人，2018）是一种轻量且有效的注意力块，可端到端地集成到目标检测模型中，并展现出卓越的检测准确率（Zhu 等人，2021年）。对于输入的特征图，CBAM 依次沿通道维度和空间维度应用注意力图。如图 5-(a) 所示，CBAM 模块的网络结构由两个依次连接的子模块组成：通道注意力模块（CAM）和空间注意力模块（SAM）。它们通过将输入特征图与注意力图相乘，在每个中间卷积块对特征图进行自适应细化。对于输入的特征图 $\boldsymbol{F_c} \in \mathbb{R}^{C\times H\times W}$，CAM 生成一维通道注意力图 $\boldsymbol{M^c} \in \mathbb{R}^{C\times 1\times 1}$，随后 SAM 推断出二维空间注意力图 $\boldsymbol{M^s} \in \mathbb{R}^{1\times H\times W}$，如图 4-(a-c) 所示。CBAM 的整体注意力处理过程可表示为：

$$\boldsymbol{F_c'} = \boldsymbol{M^c}(\boldsymbol{F_c}) \otimes \boldsymbol{F_c}; \quad \boldsymbol{F_c''} = \boldsymbol{M^s}(\boldsymbol{F_c'}) \otimes \boldsymbol{F_c'}; \tag{12}$$

其中；$\boldsymbol{F_c''}$ 表示经过 CBAM 优化的输出；$\otimes$ 表示逐元素乘法。具体而言，CAM 会生成两种不同的空间描述符，包括平均池化特征$\boldsymbol{F^c_{avg}}$和最大池化特征 $\boldsymbol{F^c_{max}}$，并通过多层感知机（MLP）生成通道注意力图 $\boldsymbol{M^c} \in \mathbb{R}^{C\times 1\times 1}$。CAM 注意力可表示为：

$$\boldsymbol{M^c}(\boldsymbol{F_c}) = \sigma(\boldsymbol{W_1}(\boldsymbol{W_0}(\boldsymbol{F^c_{avg}}))) + \boldsymbol{W_1}(\boldsymbol{W_0}(\boldsymbol{F^c_{max}})) \tag{13}$$

其中，$\boldsymbol{W_0} \in \mathbb{R}^{C/r\times C}$和 $\boldsymbol{W_1} \in \mathbb{R}^{C\times C/r}$ 是 MLP 权重；σ 表示 Sigmoid 激活函数。在 SAM 中，空间间注意力图 $\boldsymbol{M^s(F_c)} \in \mathbb{R}^{H\times W}$经过聚合，生成二维图：$\boldsymbol{F^s_{avg}} \in \mathbb{R}^{1\times H\times W}$和 $\boldsymbol{F^s_{max}} \in \mathbb{R}^{1\times H\times W}$。根据这两个特征图，

空间注意力可计算为:

$$\boldsymbol{M_s}(\boldsymbol{F}) = \sigma(f^{n\times n}([\boldsymbol{F^s_{avg}}; \boldsymbol{F^s_{max}}]) \tag{14}$$

其中， $f^{n\times n}$ 表示滤波器尺寸为 $n \times n$、具有 $n = 7$的卷积操作。在所提出的 DenseSPH-YOLOv5 中引入 CBAM模块，通过针对密集目标给予特定关注，并区分易混淆的噪声区域，显著提高了损伤检测的检测准确率，这证明了该模块的有效性。总体而言，它有助于学习更具表现力的特征，从而显著提高本文所考虑的道路损坏数据集上的检测准确率。

3.3.3 Swin Transformer编码器块： 受视觉Transformer（Dosovitskiy等人，2020年）在密集目标检测和遮挡目标检测中的卓越性能启发，在所提出的模型中，如图3所示，Swin Transformer（STR）编码器块（Liu 等人，2021年）已融合到 DenseSPH-YOLOv5架构的全部四个检测头中。该实现利用自注意力机制（Vaswani等人，2017）改进了全局语义特征提取和上下文信息融合，并在密集目标检测中展现出卓越性能（Gong等人，2022年）。

对于输入的特征图 $\boldsymbol{F_s} \in \mathbb{R}^{H\times W\times C}$，经过线性投影和重塑操作后，将其转换为 $\boldsymbol{Q^s}\ \boldsymbol{K^s}\ \boldsymbol{V^s}$,,,,,, $V^s \in \mathbb{R}^{N\times C'}$，其中$N = H \times W$，以输入多头自注意力（MSA）模块。MSA输出的特征图 $\boldsymbol{Z^s}$ 聚合了全局信息，可表示为

$$\boldsymbol{Z^s} = \boldsymbol{A^s V^s}; \quad \boldsymbol{A^s} = \text{softmax}\ \boldsymbol{Q^s K^{sT}}; \tag{15}$$

其中， $\boldsymbol{A^s} \in \mathbb{R}^{N\times N}$ 是自注意力矩阵，表示特征图元素与其余元素之间的关系。尽管 Transformer 中的 MSA 能够通过整合多个独立子空间有效生成自注意力矩阵，但对于高分辨率图像而言，其全局计算开销显著过高。因此，对于密集预测或处理高分辨率图像，Transformer 中 MSA 的计算复杂度并不适用，因为其复杂度会随词元数量呈二次增长。

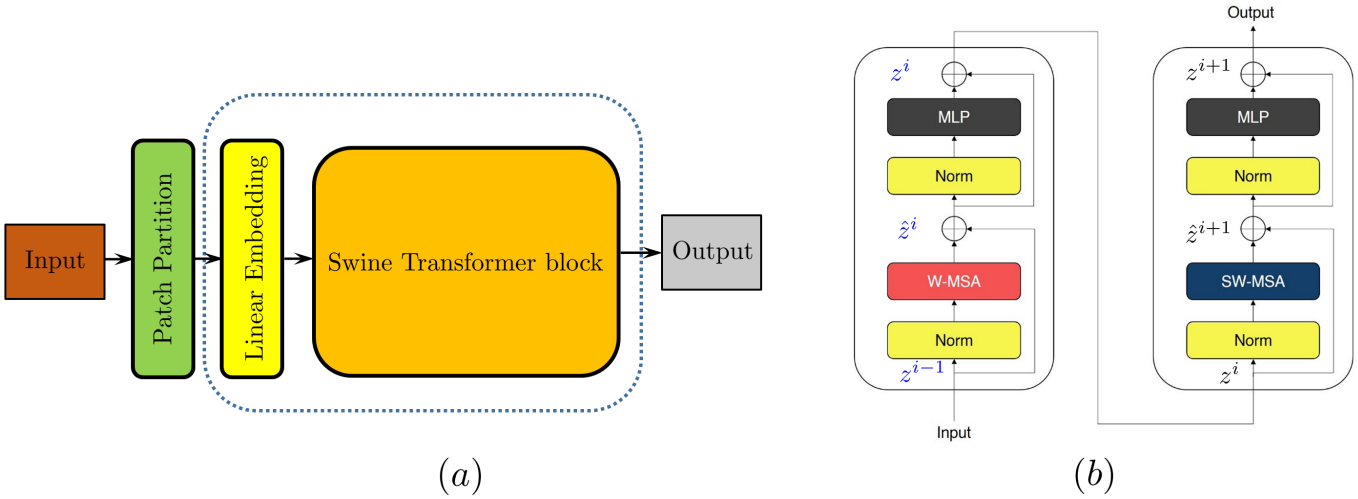

图6：示意图：(a) STR 编码器架构，其中包含 (b) 用于 DenseSPH-YOLOv5架构检测头的常规窗口多头自注意力和移位窗口多头自注意力。

(Liu 等人，2021年）。为此，STR 能够显著提高 MSA 的计算效率，其计算复杂度随图像尺寸呈线性增长，从而提升模型在检测速度和准确率方面的性能。因此，在所提出的 DenseSPH-YOLOv5 检测模型中，STR 编码器模块被融合到四个不同的预测头中。如图 6 - (b) 所示，每个 STR 编码器包含两个子层，其中包括基于移位窗口的多头自注意力（MSA）模块，后接带有 GeLU 非线性的全连接多层感知器。每个 MSA 模块之后均采用残差连接。随后，在 MSA 和 MLP 之前加入层归一化（LN）。在 STR 中，W-MSA 模块将得到的特征图划分为互不重叠的独立窗口。随后，在每个局部窗口内计算自注意力。采用 SW-MSA 分区后，连续的 STR 块可计算如下：

$$\hat{\boldsymbol{z}}^{i} = \mathbf{W\text{-}MSA}(\mathbf{LN}(\boldsymbol{z}^{i-1})) + \boldsymbol{z}^{i-1}; \quad \boldsymbol{z}^{i} = \mathbf{MLP}(\mathbf{LN}(\hat{\boldsymbol{z}}^{i})) + \hat{\boldsymbol{z}}^{i} \tag{16}$$

$$\hat{\boldsymbol{z}}^{i+1} = \mathbf{SW\text{-}MSA}(\mathbf{LN}(\boldsymbol{z}^{i})) + \boldsymbol{z}^{i}; \quad \boldsymbol{z}^{i+1} = \mathbf{MLP}(\mathbf{LN}(\hat{\boldsymbol{z}}^{i+1})) + \hat{\boldsymbol{z}}^{i+1} \tag{17}$$

其中 $\boldsymbol{z}^{i}$ 和$\hat{\boldsymbol{z}}^{i}$ ^^分别表示来自MLP和移位窗口多头自注意力模块的输出特征。对于局部窗口大小为 $m \times m$ 的特征图 $\boldsymbol{F}_{\boldsymbol{s}} \in \mathbb{R}^{H \times W \times C}$，其计算复杂度

ϒ可通过以下方式获得

$$\Upsilon(\mathbf{MSA}) = 4HWC^2 + 2(HW)^2C; \quad \Upsilon(\mathbf{W\text{-}MSA}) = 4HWC^2 + 2(HW)M^2C \tag{18}$$

其中 HW 为图块数量。显然，当 HW较大时，全局多头自注意力的计算开销极高，而窗口多头自注意力具有良好的可扩展性。因此，采用STR可显著降低模型的计算复杂度。

3.3.4 感受野增强：卷积神经网络的一项要求是输入图像必须具有固定尺寸。然而，由于图像的宽高比各不相同，在卷积过程中需要通过裁剪和变形将其调整为固定尺寸，这会导致重要特征丢失。针对这一问题，SPP（He 等人，2015）采用了一种高效策略，可在多个尺度上检测目标对象。为此，我们在 Dense-CSPDarknet53 骨干网络中添加了一个与 DCSP-3 集成的 SPP 模块，以增强感受野表示能力并提取重要的上下文特征，如图 4所示。

4. 结果：

本节讨论所提出的 DenseSPH-YOLOv5 框架的性能和检测准确率。该框架在 RDD-2018 数据集上进行了评估，该数据集包含 8 种不同的损伤类别。为使边界框表示更加清晰，每种损坏类型都标注了相应的类别标识符，如表 1所示。通过广泛的消融研究，对 DenseSPH-YOLOv5 网络的性能进行了优化。最后，我们详细研究了所提出模型的性能，并将其与若干先进的目标检测模型进行了比较。

4.1 训练流程：

在本研究中，我们开展了广泛而深入的研究，以探究所提出的 DenseSPH-YOLOv5 模型在道路损伤定位任务中的比较性能分析。根据 RDD-2018 数据集，共有 80% 和 20% 的图像分别被

随机选取用于训练和验证。对于所有实验，我们使用了一台基于 Windows 10 Pro（64 位）的计算系统，该系统配备 Intel Core i5-10210U 中央处理器，主频为 2.8 GHz $\times 6$、32 GB DDR4 内存，并采用 NVIDIA GeForce RTX 2080 进行 GPU 并行化。作为数据增强的一部分，除光度畸变和几何畸变等传统方法外，还结合使用了 MixUp（Zhang 等人，2017b 年）、CutMix（Yun 等人，2019年）和 Mosaic（Bochkovskiy 等人，2020年）等额外的数据增强策略，有助于提升 DenseSPH-YOLOv5 的性能。此外，Mosaic 增强可以显著丰富背景信息。由于该方法会增大批量大小，因此采用了批归一化。除非另有说明，批量大小设为 16，总训练步数保持为 80。初始学习率设为 0.01，并采用 SGD 优化器。训练数据集使用可用的预训练权重文件（Lin 等人，2014）。

4.2 网络性能优化：

首先，我们开展了大量实验，以选择合适的主干网络-颈部组合模块，从检测准确率和检测速度两方面优化所提出的 DenseSPH-YOLOv5 模型的性能。对于不同的主干网络-颈部配置组合，其在 AP、AP_{50}, AP_{75}, AP_S、AP_M以及 AP_L 等参数方面的检测准确率，以及检测速度（以 FPS 表示）已列于表 2中。从表 2可以看出，引入附加检测头（ADH）后，通过采用额外的特征融合机制，模型性能得到了提升。值得注意的是，与基线原始 YOLOv5 相比，AP_{50}提高了 2.3%；AP_L 提高了 6.1%。此外，在采用 ADH 的基础上引入 DenseNet 可进一步提升性能，尤其是在相对较小目标检测方面，AP_L 提高了 3.6%，如图 7-(a)所示。然而，这也导致检测速度显著下降（从 69.1 FPS 降至 55.2 FPS）。在检测模型的颈部部分引入 CBAM 后，与 DensNet+ADH 配置相比，检测性能得到进一步提升，尤其是在中大型目标尺寸检测方面，AP_M 和 AP_L分别提高了 3.7% 和 3.6%，但检测速度略有下降。因此，事实证明，CBAM 的实现是一种有效的

表2：锚框尺寸为 416×416 时，DenseSPH-YOLOv5架构中各种残差块与密集块组合的性能。

骨干网络 附加模块	Neck 附加模块	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	FPS
（YOLOv5） -	-	69.2	71.7	83.2	61.2	78.3	81.4	70.2
-	ADH	70.4	74.1	85.9	67.3	79.7	84.7	69.1
DenseNet	ADH	74.5	79.7	88.5	70.9	81.2	85.1	55.2
DenseNet	ADH+CBAM	79.1	83.6	90.2	75.1	84.9	88.7	56.4
DenseNet	ADH+CBAM+STR	**81.4**	**85.2**	**91.4**	**78.2**	**86.5**	**89.5**	**62.4**

策略，可通过空间和通道注意力机制保留语义信息及精细的空间信息，从而提升模型的整体性能。最终，当 CBAM 和 STR 分别集成到颈部网络和检测头中时，模型取得了最佳性能，准确率指标得到了显著提升，尤其是 AP、AP_{50}, AP_{75}, AP_M 和 AP_L 分别提高了 12.2%、13.5%、8.2%、8.3% 和 8.1%，相比原始 YOLOv5 配置均有所提升。此外，我们还观察到，在检测头中引入 STR 后，检测速度也得到了提升；与仅采用 ADH+CBAM 颈部配置相比，其计算复杂度有所降低，如图 7-(a)所示。因此，对于本文所考虑的道路损坏数据集，DenseSPH-YOLOv5 采用这种配置可在检测准确率和检测速度方面实现最佳性能。总之，结合合适的激活函数和改进的主干网络-颈部组合，DenseSPH-YOLOv5 能够在复杂场景下提供一种高效且高性能的损伤检测模型。

4.3 最先进方法比较：

本节将所提出的 DenseSPH-YOLOv5 的检测性能与一些现有的最先进检测模型进行了比较（Zhao 等人，2019b年）。在性能比较中，我们考虑了 Faster R-CNN（Ren 等人，2016年）、RetinaNet（Lin 等人，2017b年）、SSD（Liu 等人，2016年）、YOLOv4（Bochkovskiy 等人，2020年）、Dense-YOLOv4

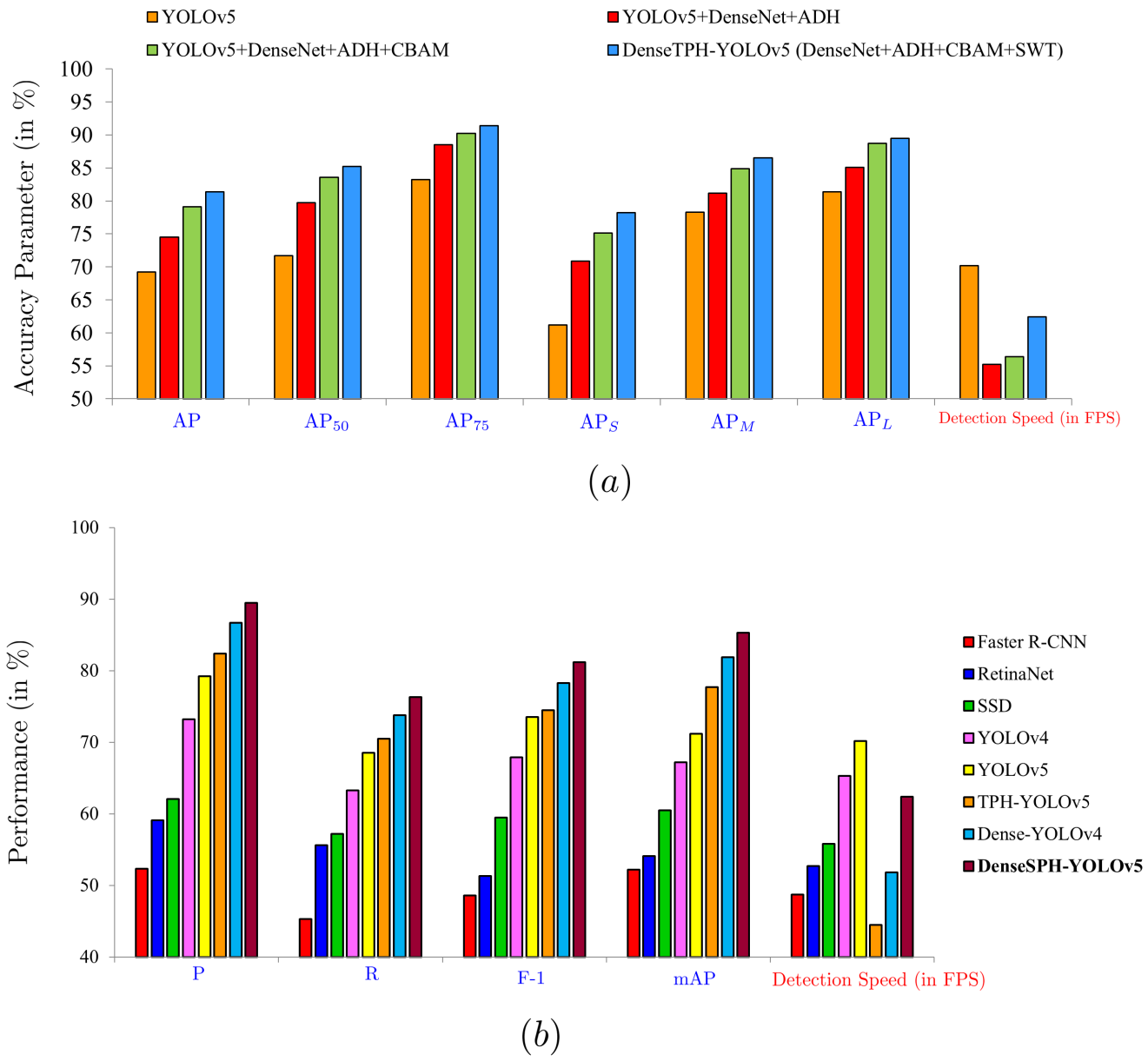

图 7：精度参数与检测速度对比柱状图：（a）检测模型不同骨干网络和颈部网络组合的对比；（b）不同先进模型与 DenseSPH-YOLOv5 的对比。

表3：DenseSPH-YOLOv5 与其他最先进模型在精确率、召回率、F1值、平均精度均值和检测速度（帧/秒）等性能参数方面的比较，其中粗体表示最佳性能值。

模型	精确率 (%)	召回率 (%)	F1分数（%）	平均精度均值（%）	检测时间（毫秒）	FPS
Faster R-CNN	52.32	45.39	48.60	52.17	20.50	48.7
RetinaNet	59.11	55.67	51.35	54.11	18.87	52.7
SSD	62.13	57.19	59.56	60.52	17.92	55.8
YOLOv4	73.22	63.35	67.93	67.13	15.31	65.3
YOLOv5	79.17	68.47	73.43	71.13	**14.24**	**70.2**
TPH-YOLOv5	82.37	70.51	74.52	77.62	22.47	44.5
Dense-YOLOv4	86.75	73.75	78.28	81.87	19.30	51.8
DenseSPH-YOLOv5	**89.51**	**76.25**	**81.18**	**85.25**	16.02	62.4

（Roy 和 Bhaduri，2022）、在 RDD-2018 数据集上训练的 YOLOv5（Jocher 等人，2021）以及 TPH-YOLOv5（Zhu 等人，2021）。表3展示了这些模型在精确率、召回率、F1分数、平均精度均值和检测速度等不同性能参数上的比较结果。比较结果表明，与图7-(b)柱状图所示的 YOLO变体相比，Faster R-CNN、RetinaNet 和 SSD 的准确率明显较低。在 YOLOv4 和 YOLOv5 之间，YOLOv5 的性能更好，精确率和平均精度均值分别提高了 5.95% 和 4.01%。我们发现，Dense-YOLOv4 的性能优于 TPH-YOLOv5，其精确率、召回率、F1值和平均精度均值分别提高了 4.38%、3.24%、3.76% 和 4.25%。然而，如图7-(b)所示，本文提出的 DenseSPH-YOLOv5 性能最佳，其精确率、召回率、F1值和平均精度均值分别达到 89.51%、76.25%、81.18% 和 85.25%。此外，DenseSPH-YOLOv5 的实时检测速度高达 62.4 帧/秒，分别比 TPH-YOLOv5 和 Dense-YOLOv4 模型高出 28.68% 和 16.98%。总之，DenseSPH-YOLOv5 在检测准确率和速度方面均优于一些优秀的检测模型，展现出其用于自动化高性能损伤检测模型的适用性。

4.4 与现有最先进损伤检测模型的比较：

此外，我们还将 DenseSPH-YOLOv5 的性能与若干现有的最先进道路损伤检测模型进行了比较，这些模型均在 RDD-2018 数据集上进行评估。如表4所示，我们将本文的 Dense-YOLOv4 和 DenseSPH-YOLOv5 模型与多种深度学习模型进行了性能比较，包括 SSD Inception v2（Maeda 等人，2018）、SSD MobileNet（Maeda 等人，2018）、YOLO（Alfarrarjeh 等人，2018）、Faster R-CNN（Kluger 等人，2018）、采用 ResNet-152 的 Faster R-CNN（Wang 等人，2018a）、采用 Faster R-CNN 和 SSD 的集成模型（Wang 等人，2018b），以及 RetinaNet（Angulo 等人，2019）。请注意，在 Angulo 等人（2019）的研究中，研究人员向 RDD-2018 数据集添加了额外图像，以提升检测性能。直接比较结果显示，目前两种最先进的模型——采用 ResNet-152 的 Faster R-CNN（Wang 等人，2018a）和集成模型（Wang 等人，2018b）——F1值均达到 62.55%，较 SSD MobileNet（Maeda 等人，2018）提高了 2.63%。尽管 RetinaNet（Angulo 等人，2019）表现出显著的性能提升，Dense-YOLOv4 在原始 RDD-2018 数据集上的表现更好，F1值达到 78.28%。与上述模型相比，本文提出的模型在当前最先进的损伤检测模型中取得了最高的 F1值，即 81.18%。在其他精度指标方面，与 RetinaNet（Angulo 等人，2019）相比，精确率和召回率分别提高了 3.64% 和 10.51%。在检测速度方面，DenseSPH-YOLOv5 的表现可与 SSD Inception v2（Maeda 等人，2018）相媲美，进一步凸显了其在实时损伤检测方面的优势。

4.5 DenseSPH-YOLOv5 的整体性能：

从第 4.3 节可以看出，与其他最先进模型相比，TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 的表现更佳。因此，如表 5所示，我们从平均精度均值、F1值、交并比、最终损失和平均检测时间等方面对这三种模型进行了比较。所提出的 DenseSPH-YOLOv5 达到了 0.803 的最高平均交并比，表明其在目标检测过程中的边界框精度优于另外两种模型。此外，它还通过达到

表4：在 RDD-2018 数据集上评估的 DenseSPH-YOLOv5 与其他 SOAT 道路损伤检测模型的各项性能参数对比，其中粗体表示最佳性能值。

参考文献	方法	性能	检测速度（帧/秒）
Maeda 等人(2018)	SSD Inception v2	F1：52.61；精确率：81.10；召回率：38.97	**63.1**
Maeda 等人(2018)	SSD MobileNet	F1：59.92；精确率：81.11；召回率：47.52	30.6
Alfarrarjeh 等人 (2018)	YOLO	F1：62.00；精确率：-；召回率：-	-
Kluger 等人 (2018)	Faster R-CNN	F1：61.00；精确率：-；召回率：-	-
Wang 等人 (2018a)	Faster R-CNN (ResNet-152)	F1：62.55；精确率：-；召回率：-	-
Wang 等人(2018b)	集成 (Faster R-CNN+ SSD)	F1：62.55；精确率：-；召回率：-	-
Angulo 等人（2019年）*	RetinaNet	F1：74.97；精确率：85.87；召回率：65.75	-
（Roy 和 Bhaduri，2022年）	Dense-YOLOv4	F1：78.28；精确率：86.75；召回率：73.75	51.8
我们的方法（本研究）	**DenseSPH-YOLOv5**	**F1：81.18；精确率：89.51；召回率：76.25**	62.4

表5：TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 的整体性能比较。

检测模型	IoU	F1	mAP	验证损失	检测时间（毫秒）	检测速度（帧/秒）
TPH-YOLOv5	0.740	0.745	0.776	18.07	22.47	44.5
Dense-YOLOv4	0.781	0.782	0.819	10.13	19.30	51.8
DenseSPH-YOLOv5	0.803	0.811	0.852	7.18	16.02	62.4

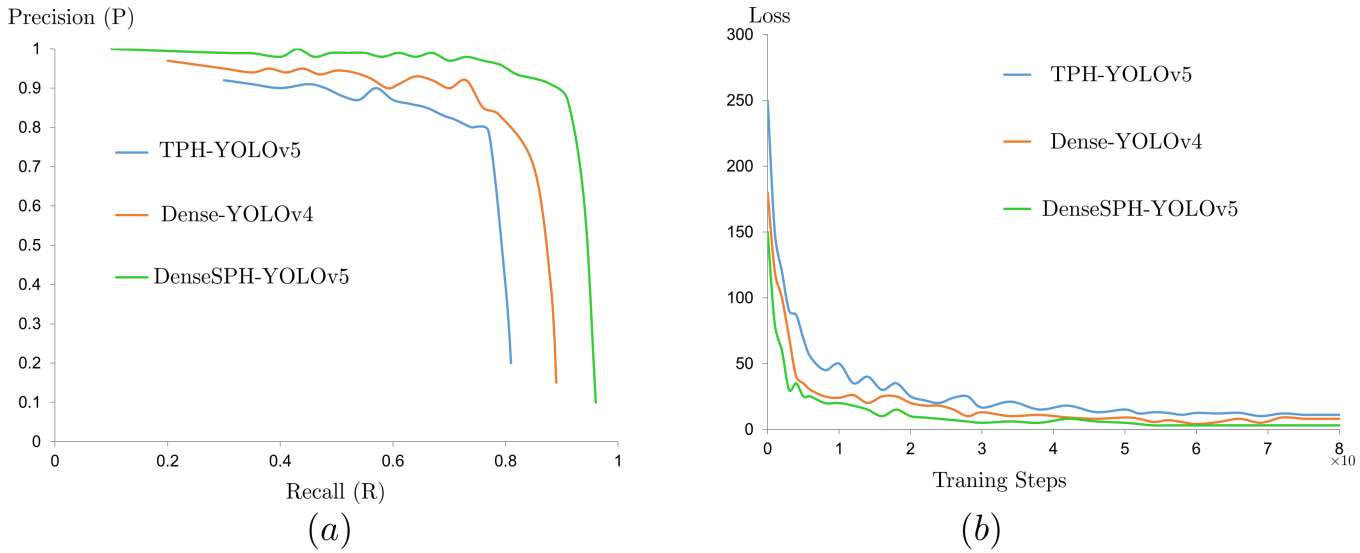

图8：TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 的（a）精确率-召回率曲线；（b）损失演化曲线对比。

最高 F1值和平均精度均值分别为 0.811 和 0.852，较 TPH-YOLOv5 提高了 6.61% 和 7.63 %。此外，DenseSPH-YOLOv5 的检测速度达到 62.4 帧/秒，高于 TPH-YOLOv5 和 Dense-YOLOv4。因此，与另外两种最先进模型相比，该模型能够以更高的准确率实现实时道路损伤检测。图 8-(a) 展示了这三种模型的精确率-召回率曲线对比。从曲线对比中可以看出，对于特定的召回率，DenseSPH-YOLOv5 能够取得更高的精确率。它的精确率-召回率曲线下面积最大，表明其检测性能优于 TPH-YOLOv5 和 Dense-YOLOv4。接下来，我们比较图 8-(b) 所示的损失演化曲线。在初始阶段经历数次波动后，DenseSPH-YOLOv5 的损失

表6：TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 各类别检测结果的比较

模型	类别 →	**D00**	**D01**	**D10**	**D11**	**D20**	**D40**	**D43**	**D44**	**平均值**
TPH-YOLOv5	P	0.87	0.62	0.84	0.87	0.84	0.87	0.87	0.81	0.82
	R	0.61	0.89	0.49	0.51	0.71	0.74	0.81	0.88	0.71
	F1	0.72	0.73	0.62	0.64	0.77	0.79	0.83	0.84	0.75
	mAP	0.81	0.79	0.78	0.82	0.73	0.78	0.76	0.74	0.77
Dense-YOLOv4	P	0.89	0.69	0.91	0.92	0.89	0.91	0.89	0.84	0.87
	R	0.65	0.93	0.52	0.51	0.78	0.78	0.86	0.87	0.74
	F1	0.75	0.79	0.66	0.65	0.83	0.84	0.87	0.85	0.78
	mAP	0.82	0.81	0.84	0.89	0.77	0.82	0.81	0.79	0.82
DenseSPH-YOLOv5	P	0.92	0.74	0.93	0.94	0.93	0.92	0.91	0.87	**0.89**
	R	0.68	0.95	0.58	0.54	0.81	0.79	0.88	0.87	**0.76**
	F1	0.78	0.83	0.71	0.68	0.86	0.85	0.89	0.87	**0.81**
	mAP	0.84	0.89	0.81	0.91	0.82	0.85	0.85	0.85	**0.85**

模型在约 50 个训练步数后趋于稳定，最终损失值为 7.18。而另外两种模型的损失演化波动更大，最终损失值也更高。显然，所提出的 TPH-YOLOv5 更易于训练，且具有更快的收敛特性，这从计算角度证明了其有效性。

为了进一步深入了解这些模型的性能，表6展示了 TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 的检测结果对比，其中包含各类别的精确率（P）、召回率（R）、平均精度均值（mAP）和 F1 值。DenseSPH-YOLOv5 在多个类别的精确率和召回率方面均有显著提升，尤其是在检测纵向和横向线性裂缝、鳄鱼裂缝及坑洞方面。因此，DenseSPH-YOLOv5 能够有效地最大化所有类别的真正例（TP）值，同时降低假正例（FP）和假负例（FN）值。与 Dense-YOLOv4 相比，所提出的模型在精确率方面提升了 2。01%，召回率方面提升了 2。11%。综合对比结果表明，在检测各种损伤类别方面，DenseSPH-YOLOv5 的表现最佳，在精确率和准确率方面优于另外两种最先进模型。

4.6 各种损伤类别的检测：

在本节中，我们展示了 DenseSPH-YOLOv5 对八种不同损伤类别的检测结果，并将其与 TPH-YOLOv5 和 Dense-YOLOv4 进行了比较。

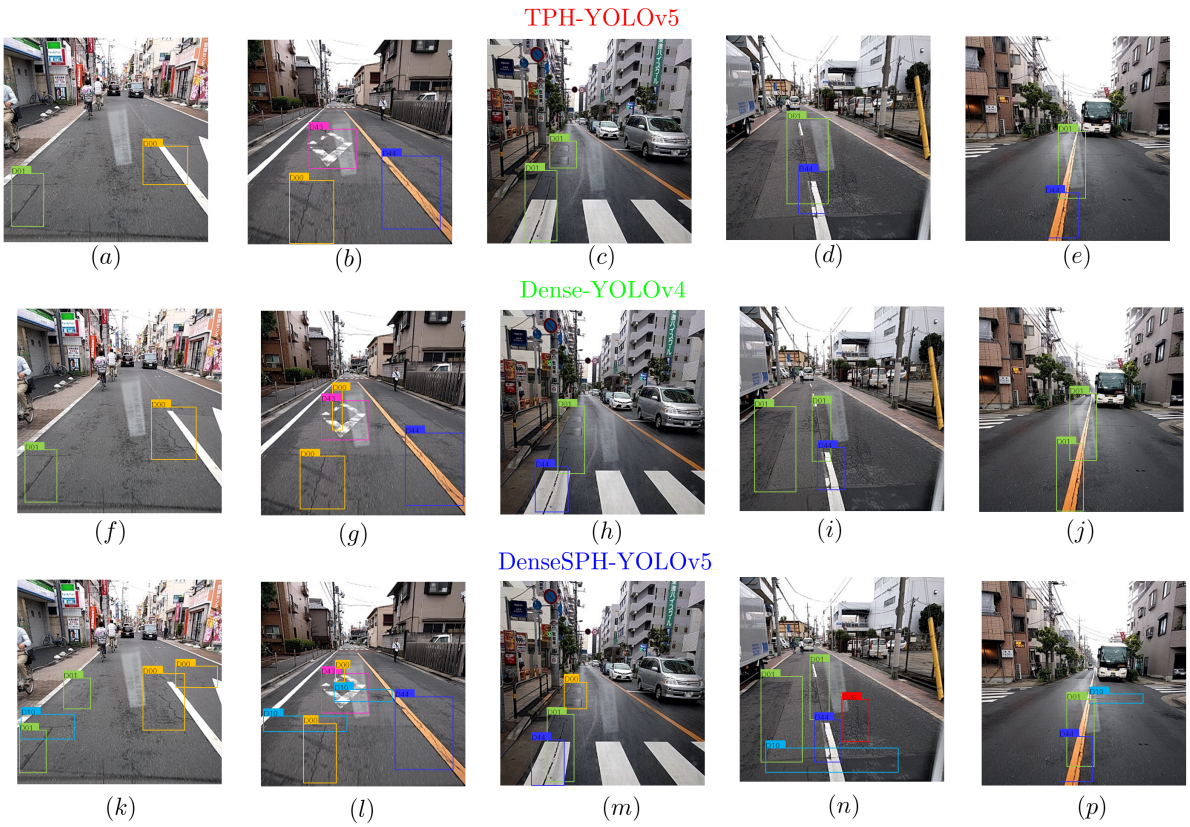

图9：不同损伤检测结果对比：(a-e) TPH-YOLOv5；(f-j) Dense-YOLOv4；(k-p) DenseSPH-YOLOv5。表中列出了详细检测结果及平均置信度指标：7。

表7：图中所示不同损伤类别的 TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 的详细检测结果：9- 10。

图号	模型	检测	误检/漏检	平均置信度分数
9 (a-e)	TPH-YOLOv5	13	8	0.71
9 (f-j)	Dense-YOLOv4	16	5	0.75
9 (k-p)	**DenseSPH-YOLOv5**	20	1	0.83
10 (a-e)	TPH-YOLOv5	13	7	0.67
10 (f-j)	Dense-YOLOv4	14	6	0.72
10 (k-p)	**DenseSPH-YOLOv5**	18	2	0.79

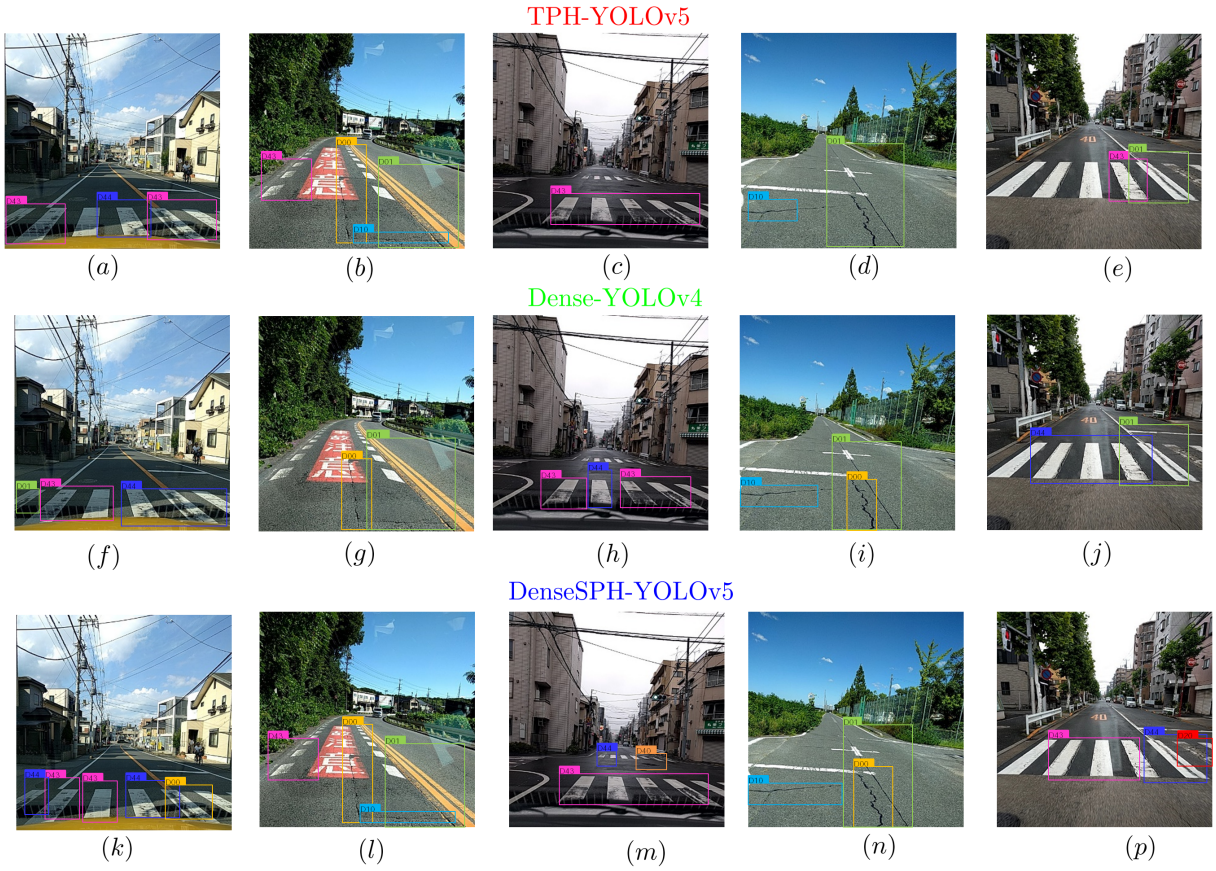

图10：不同损伤检测结果的对比：(a-e) TPH-YOLOv5；(f-j) Dense-YOLOv4；(k-p) DenseSPH-YOLOv5。包含平均置信度分数的详细检测结果见表7。

检测结果的可视化呈现采用带类别标识符的边界框标出（见表1），如图9-10所示。表中报告了相应的详细检测结果，包括检出和漏检的目标类别数量以及平均置信度分数。7。综合比较可得，所提出的模型能够以较高的平均置信度指数值精确检测目标对象，展现出良好的性能。同时，与 TPH-YOLOv5 和 Dense-YOLOv4 模型相比，该模型减少了误检和漏检结果。尤其是在多个目标对象之间存在显著重叠时，如图9-(l, n) 和图10-(k, n) 所示，所提出的 DenseSPH-YOLOv5 能够相当准确地预测边界框并检测出各个目标对象。

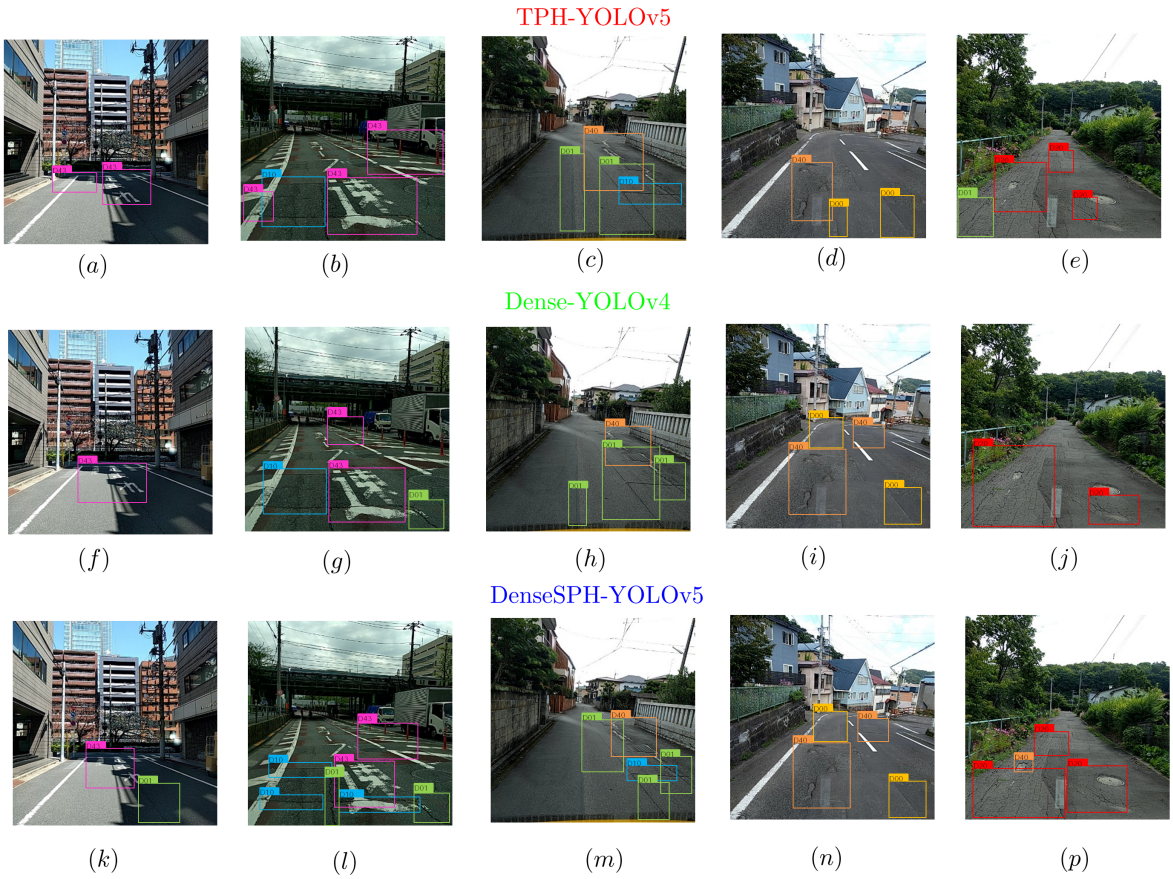

图11：复杂场景下损伤检测结果的比较：(a-e) TPH-YOLOv5；(f-j) Dense-YOLOv4；(k-p) DenseSPH-YOLOv5。具有平均置信度指标的详细检测结果已在表6中展示。

4.7 复杂场景下的检测：

本节进一步探讨了图11 所示的一些复杂场景下的检测情况，以展示所提出的 DenseSPH-YOLOv5 的性能。表8报告了详细检测结果，包括检出和漏检的目标类别总数以及平均置信度分数。这里我们考虑了各种复杂背景和复杂环境；由于存在阴影，目标类别通常难以检测。此外，不同目标类别的边界框之间还存在一定程度的遮挡和重叠。对于目标检测模型而言，逐一检测各个目标对象是一项具有挑战性的任务。因此，出色的特征提取对于高效、准确地检测各种目标类别至关重要。在此类情况下，所提出的 DenseSPH-YOLOv5 在边界框精度方面展现出优越性，

表8：图11中不同损伤类别的 TPH-YOLOv5、Dense-YOLOv4 和 DenseSPH-YOLOv5 的详细检测结果。

图号	模型	检测	误检/漏检	平均值置信度分数
11 (a)-(e)	TPH-YOLOv5	13	10	0.68
11 (f)-(j)	Dense-YOLOv4	15	8	0.69
11 (k)-(p)	**DenseSPH-YOLOv5**	19	4	0.77

表9：图中所示原始 RGB、灰度、低分辨率以及不同亮度强度（75% 和 50%）图像的 DenseSPH-YOLOv5 检测结果对比12。

图编号	原文		灰度		低分辨率		亮度-75%		亮度-50%	
	Det.	未检测	Det.	未检测	Det.	未检测	Det.	未检测	Det.	未检测
图 12-(a-e)	6	0	5	1	4	2	6	0	5	1
图 12-(f-j)	3	1	4	0	2	2	3	1	3	1
图 12-(k-p)	5	0	5	0	2	3	5	0	3	2

与另外两个模型相比，平均置信度分数更高。值得注意的是，该模型显著减少了误检或漏检（与 TPH-YOLOv5 相比，从 10 次降至 4 次），因此置信度分数也有所提高，这一点在表8中清晰可见。总体而言，根据对各类损伤类别的检测结果，所提出的 DenseSPH-YOLOv5 展现出更强的检测能力，尤其是在检测阴影方面，优于 TPH-YOLOv5 和 Dense-YOLOv4 模型。

4.8 灰度、低分辨率及不同光照强度下的检测：

本节进一步测试了 DenseSPH-YOLOv5 在灰度和像素化低分辨率（100×100）图像上的检测准确率，如图 12所示。此外，还研究了不同光照强度（即亮度75%和亮度50%）下的预测结果，并将其与原始RGB图像的检测结果进行了比较。表 9报告了详细的检测结果及置信度分数。对于灰度图像，所提出的模型表现出色，能够准确预测边界框，且边界框置信度分数更高，如图 12-(b, g,

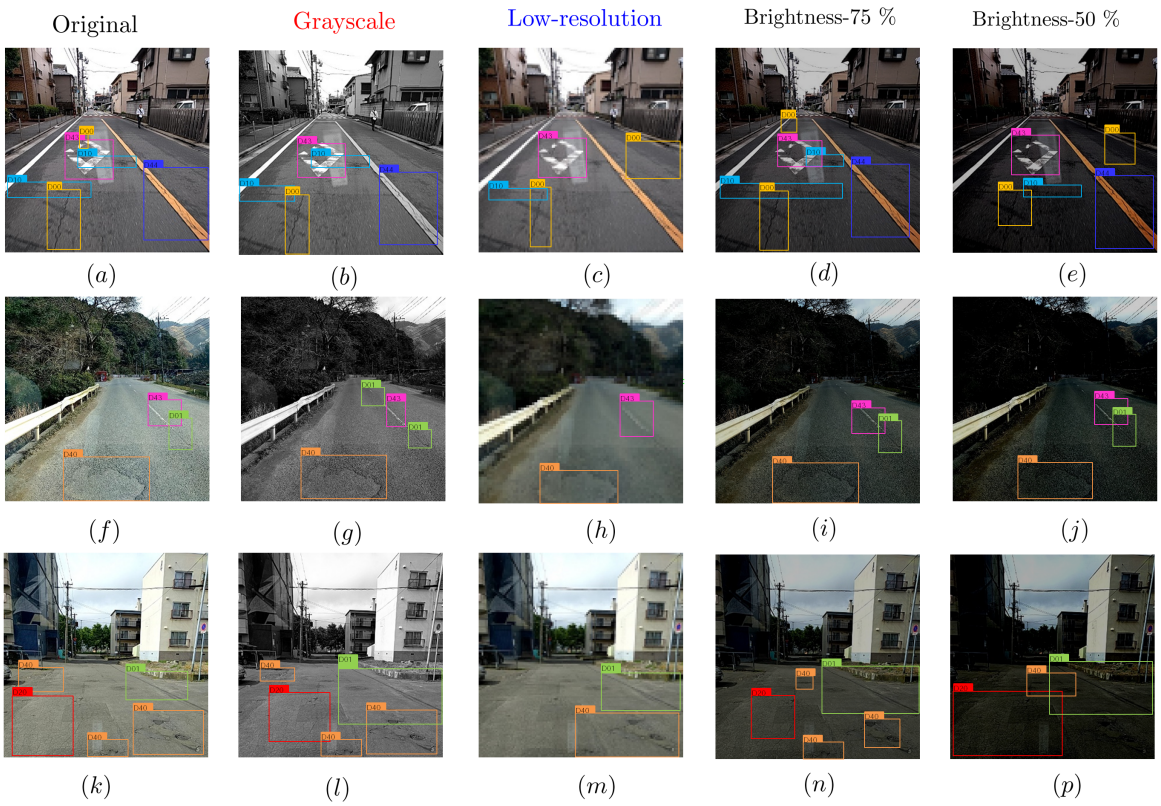

图12：DenseSPH-YOLOv5 的检测结果：（a、f、k）原始RGB图像；（b、g、l）对应的灰度图像；（c、h、m）低分辨率图像；（d、i、n）亮度强度为75%；（e、j、p）亮度强度为50%。详细检测结果见表6。

l）。对于像素化低分辨率图像，DenseSPH-YOLOv5 的性能略有下降，但仍能以较高的准确率检测出大多数目标类别。类似地，如图 12-(d, i, n) 和图 12-(e, j, p) 所示，所提出的模型在不同光照强度下均表现出更优的检测结果。测试结果表明，与 TPH-YOLOv5 和 Dense-YOLOv4 模型相比，所提出的模型在更复杂的环境中具有更强的适应性。

5. 讨论：

从总体检测结果来看，与当前最先进模型相比，DenseSPH-YOLOv5在复杂环境和严峻条件下具有更强的适应性，也更善于定位各种损伤类别。此外，该模型在边界框检测方面表现出更高的准确率，因此能够有效避免其他检测网络中出现的误检和漏检问题，这证明了所提出模型在实际现场损伤检测任务中的实用价值。未来的工作可着眼于进一步优化检测准确率和检测速度，以便在移动计算平台上实现便携式现场损伤检测。当前模型还可通过引入一些最先进的算法来进一步改进，例如 YOLO-X（Ge 等人，2021）、YOLOv7（Wang 等人，2022年）等。此外，还可将其与无人机视频监控系统集成，以实现实时、准确的损伤检测。此外，重叠和遮挡可能导致预测错误，可通过缩小特征步幅来改善定位和边界框预测。为避免边界框重叠，可考虑用点（Ribera 等人，2019年）或掩码（Xu 等人，2020年）替代边界框。此外，一种潜在应用是将目标检测框架与语义分割方法相结合，例如 Mask R-CNN（Bharati 和 Pramanik，2020年）、U-Net（Esser 等人，2018），以提取各种损伤类别的形态学信息。尽管如此，当前研究阐明了将尖端的基于深度学习的计算机视觉方法应用于商业应用中的多类别自动化损伤检测流程的可行性。

6. 结论：

总而言之，本研究基于改进版 YOLOv5网络开发了一种高效、稳健的目标定位模型 DenseSPH-YOLOv5，用于准确分类和定位损伤。所提出的 DenseSPH-YOLOv5 已用于检测八种不同的道路损坏类别，在各种复杂且具有挑战性的环境下均能实现优异且准确的检测。在 RDD-2018 数据集上进行评估后发现，DenseSPH-YOLOv5 的检测速度为 62.40 帧/秒，平均精度均值（mAP）、F1值和精确率（P）分别达到 85.25%、81.18% 和 89.51%，在检测所有损伤类别时，其分类准确率和边界框预测定位性能均优于现有最先进的损伤检测模型。本研究有效弥补了现有基于深度学习的损伤检测模型的不足，并展现了其在实时现场应用中的卓越潜力。简言之，当前研究为实现完全自动化、高效的基于深度学习的计算机视觉多类别损伤检测流程迈出了重要一步。

致谢：谨此感谢航空研究与发展委员会的支持（资助编号：DARO/08/1051450/M/I）。

利益冲突：作者声明，他们不存在任何已知的竞争性经济利益或个人关系，且这些利益或关系可能会影响本文所报告的研究。

参考文献

Agarwal, S. 和 Singh, D.（2015）。一种用于工业应用的自适应统计方法，利用毫米波成像雷达对陶瓷砖进行无损地下裂缝检测。*IEEE Sensors Journal*，15(12)：7036--7044。

Alfarrarjeh, A., Trivedi, D., Kim, S. H., 和 Shahabi, C.（2018）。一种基于智能手机图像的道路损伤检测深度学习方法。载于*2018 IEEE*国际大数据会议（大数据），第5201--5204 页。IEEE。

Anand, S., Gupta, S., Darbari, V., 和 Kohli, S.（2018）。Crack-pot：自动道路裂缝和坑洞检测。载于*2018* 数字图像计算：技术与应用（*DICTA*），第 1--6 页。IEEE。

Angulo, A., Vega-Fern´andez, J. A., Aguilar-Lobo, L. M., Natraj, S., 和 Ochoa-Ruiz, G.（2019年）。一种用于实体资产管理、基于深度神经网络的道路损伤检测采集系统。载于墨西哥国际人工智能会议，第 3--14 页。Springer。

Anitha, M.、Hemalatha, R. 和 Radha, S.（2021年）。混凝土结构裂缝检测算法综述。智能系统技术进展，第639--654页。

Arya, D.、Maeda, H.、Ghosh, S. K.、Toshniwal, D.、Mraz, A.、Kashiyama, T. 和 Sekimoto, Y.（2021a）。基于深度学习的多个国家道路损伤检测与分类。《建筑自动化》，132：103935。

Arya, D.、Maeda, H.、Ghosh, S. K.、Toshniwal, D.、Omata, H.、Kashiyama, T. 和 Sekimoto, Y.（2020年）。全球道路损伤检测：先进解决方案。载于*2020*年 *IEEE*国际大数据会议（大数据），第5533--5539页。IEEE。

Arya, D.、Maeda, H.、Ghosh, S. K.、Toshniwal, D. 和 Sekimoto, Y.（2021b）。RDD2020：用于基于深度学习的自动道路损伤检测的标注图像数据集。*Data in Brief*，36：107133。

Azimi, M., Eslamlou, A. D., and Pekcan, G.（2020年）。通过深度学习实现数据驱动的结构健康监测与损伤检测：最新进展综述。*Sensors*，20(10):2778。

Bang, S., Park, S., Kim, H., Yoon, Y., and Kim, H.（2018）。采用迁移学习进行像素级道路裂缝检测的深度残差网络。网络，93(84.90):89--03。

Bharati, P. 和 Pramanik, A.（2020年）。深度学习技术——从 R-CNN 到 Mask R-CNN：综述。《模式识别中的计算智能》，第 657--668 页。

Bi̇cici, S. 和 Zeybek, M.（2021年）。一种从无人机衍生点云中自动提取路面病害的方法。《建筑自动化》，122:103475。

Bochkovskiy, A.、Wang, C.-Y. 和 Liao, H.-Y. M.（2020年）。YOLOv4：目标检测的最佳速度与准确率。

Bose, R. 和 Roy, A.（2022年）。用于大涡模拟的高精度深度学习亚网格尺度模型。《美国物理学会通报》。

Cao, M.-T.、Tran, Q.-V.、Nguyen, N.-M. 和 Chang, K.-T.（2020年）。利用多种行车记录仪图像资源检测道路损伤的深度学习模型性能综述。*Advanced Engineering Informatics*, 46:101182。

Chandio, A., Gui, G., Kumar, T., Ullah, I., Ranjbarzadeh, R., Roy, A. M., Hussain, A., 和 Shen, Y.（2022年）。精准单阶段检测器。*arXiv*预印本 *arXiv:2210.04252*。

Chen, J. 和 Cho, Y. K.（2022年）。CrackEmbed：利用异常检测对灾害现场点云进行裂缝分割的点特征嵌入。*Advanced Engineering Informatics*，52：101550。

Chen, Q., Huang, Y., Sun, H. 和 Huang, W.（2021年）。利用 Hessian结构传播进行路面裂缝检测。*Advanced Engineering Informatics*，49：101303。

Chow, J. K., Su, Z., Wu, J., Tan, P. S., Mao, X. 和 Wang, Y.-H.（2020年）。利用卷积自编码器检测混凝土结构中的缺陷。*Advanced EngineeringInformatics*，45：101105。

Davis, J. 和 Goadrich, M.（2006年）。精确率-召回率曲线与ROC曲线之间的关系。载于第*23*届国际机器学习大会论文集，第233--240页。

Dong, J.、Meng, W.、Liu, Y. 和 Ti, J.（2021年）。基于物联网和大数据的路面管理系统框架。*Advanced Engineering Informatics*，47：101226。

Dorafshan, S., Thomas, R. J., and Maguire, M. (2018). SDNET2018：一个用于基于深度卷积神经网络进行非接触式混凝土裂缝检测的标注图像数据集。*Data in Brief*, 21:1664--1668.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 等（2020）。一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer。*arXiv* 预印本 *arXiv:2010.11929*。

Du, Y., Pan, N., Xu, Z., Deng, F., Shen, Y., and Kang, H. (2021)。基于 YOLO网络 的路面病害检测与分类。*International Journal of Pavement Engineering*, 22(13):1659--1672.

Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., Ebersbach, D., Stoeckert, U. 和 Gross, H.-M.（2017）。如何为深度学习做好路面病害检测的准备？一种系统性方法。载于*2017*年国际神经网络联合会议（*IJCNN*），第2039--2047页。IEEE。

Esser, P., Sutter, E. 和 Ommer, B.（2018）。用于条件外观与形状生成的变分U-Net。载于*IEEE*计算机视觉与模式识别会议论文集，第8857--8866页。

Fan, Z., Wu, Y., Lu, J. 和 Li, W.（2018）。基于卷积神经网络结构化预测的自动路面裂缝检测。*arXiv*预印本 *arXiv:1802.02208*。

Fang, F., Li, L., Gu, Y., Zhu, H. 和 Lim, J.-H.（2020年）。一种用于裂缝检测的新型混合方法。模式识别，107:107474。

Ferri, C., Hern´andez-Orallo, J. 和 Modroiu, R. (2009). 分类性能度量的实验比较。模式识别快报，30(1):27--38。

Ge, Z., Liu, S., Wang, F., Li, Z., 和 Sun, J.（2021）。YOLOX：2021年超越YOLO系列。*arXiv*预印本 *arXiv:2107.08430*。

Girshick, R.（2015）。快速区域卷积神经网络，载于 IEEE国际计算机视觉会议论文集（第 1440--1448 页）。新泽西州皮斯卡特维：*IEEE*。*[Google* 学术搜索*]*。

Gong, H., Mu, T., Li, Q., Dai, H., Li, C., He, Z., Wang, W., Han, F., Tuniyazi, A., Li, H., 等（2022年）。采用 Swin Transformer 和注意力机制的 YOLOv5，用于卫星图像小目标检测。遥感，14(12):2861。

Gopalakrishnan, K.（2018）。数据驱动路面图像分析与自动病害检测中的深度学习：综述。数据，3(3):28。

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., 和 Agrawal, A.（2017）。结合迁移学习的深度卷积神经网络：基于计算机视觉的数据驱动路面病害检测。《建筑与建材》，157:322--330。

Guan, J., Yang, X., Ding, L., Cheng, X., Lee, V. C., 和 Jin, C.（2021年）。基于立体视觉和深度学习的自动像素级路面病害检测。《建筑自动化》，129:103788。

Guo, G. 和 Zhang, Z.（2022）。改进版 YOLOv5 道路损伤检测算法。科学报告，12(1):1--12。

Hacıefendio˘glu, K. 和 Ba¸sa˘ga, H. B.（2022）。使用基于深度学习的 Faster R-CNN 方法进行混凝土道路裂缝检测。伊朗科技期刊：土木工程分刊，46(2):1621--1633。

Han, J.、Zhang, D.、Cheng, G.、Liu, N. 和 Xu, D.（2018）。用于显著目标和类别特定目标检测的先进深度学习技术：综述。*IEEE*信号处理杂志，35(1):84--100。

Hanzaei, S. H.、Afshar, A. 和 Barazandeh, F.（2017）。陶瓷砖表面缺陷的自动检测与分类。模式识别，66:174--189。

Hartmann, T. 和 Trappey, A.（2020）。Advanced engineering informatics——哲学与方法论基础：土木与建筑工程实例。建成环境的发展，4:100020。

He, K.、Gkioxari, G.、Doll´ar, P. 和 Girshick, R.（2017）。Mask R-CNN。载于 IEEE国际计算机视觉会议论文集。

He, K., Zhang, X., Ren, S., and Sun, J. (2015). 用于视觉识别的深度卷积网络中的空间金字塔池化。*IEEE*模式分析与机器智能汇刊, 37(9):1904--1916.

Hendrycks, D. and Gimpel, K. (2016). 高斯误差线性单元（GELU）。*arXiv*预印本 *arXiv:1606.08415*.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). 密集连接卷积网络。载于*IEEE*计算机视觉与模式识别会议论文集，第 4700--4708 页。

Jamil, S., Abbas, M. S., and Roy, A. M. (2022). 使用视觉Transformer区分恶意无人机。人工智能, 3(2):260--273.

Jamil, S. and Roy, A. M. (2022). 基于深度卷积神经网络、仿生算法和视觉Transformer的稳健PCG检测模型。*SSRN* 上提供：*http://dx.doi.org/10.2139/ssrn.4316752*, 第 41 页。

Jamil, S. 和 Roy, A. M.（2023）。一种基于心音图（pcg）的高效且稳健的瓣膜性心脏病（vhd）检测框架，采用视觉Transformer（ViT）。《生物学与医学中的计算机》，第 106734 页。

Jocher, G.、Stoken, A.、Borovec, J.、Chaurasia, A.、Changyu, L.、Laughing, A.、Hogan, A.、Hajek, J.、Diaconu, L.、Marc, Y. 等（2021）。Ultralytics/YOLOv5：v5. 0-YOLOv5-P6 1280 模型、亚马逊云科技、Supervise.ly 和 YouTube 集成。*Zenodo*，11。

´Kapela, R.、Sniata-la, P.、Turkot, A.、Rybarczyk, A.、Po˙zarycki, A.、Rydzewski, P.、Wycza-lek, M. 和 B-loch, A.（2015）。基于方向梯度直方图的沥青路面裂缝检测。载于*2015 22nd* 集成电路与系统混合设计国际会议（*MIXDES*），第 579--584 页。IEEE。

Karaaslan, E., Bagci, U., and Catbas, F. N. (2021年). 使用半监督深度学习对基础设施损伤进行注意力引导分析. 《建筑自动化》, 125:103634.

Khan, W., Kumar, T., Cheng, Z., Raj, K., Roy, A. M., and Luo, B. (2022a). SQL 和 NoSQL 数据库软件架构的性能分析与评估——系统文献综述. *arXiv* 预印本 *arXiv:2209.06977*.

Khan, W., Raj, K., Kumar, T., Roy, A. M., 和 Luo, B.（2022b）。介绍乌尔都语数字数据集，并展示一种高效且稳健的基于噪声解码器的伪样本生成器。*Symmetry*，14(10)：1976。

Kluger, F., Reinders, C., Raetz, K., Schelske, P., Wandt, B., Ackermann, H., 和 Rosenhahn, B.（2018）。用于目标检测的基于区域的循环一致数据增强。载于*2018 IEEE*国际大数据会议（大数据），第 5205--5211 页。IEEE。

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., 和 Fieguth, P.（2015）。基于计算机视觉的混凝土和沥青土木基础设施缺陷检测与状况评估综述。*Advanced Engineering Informatics*，29(2)：196--210。

LeCun, Y., Bengio, Y., 和 Hinton, G.（2015）。深度学习。*Nature*，521(7553):436--444。

Li, D., Xie, Q., Gong, X., Yu, Z., Xu, J., Sun, Y., 和 Wang, J.（2021年）。使用基于视觉的检测系统自动检测地铁隧道表面的缺陷。*AdvancedEngineering Informatics*，47:101206。

Li, P., Xia, H., Zhou, B., Yan, F., 和 Guo, R.（2022年）。一种结合语义分割与边缘检测模型以提高路面裂缝识别准确率的方法。*Applied Sciences*，12(9):4714。

Li, S. 和 Zhao, X.（2021年）。使用更快的基于区域的卷积神经网络和形态学特征提取，对混凝土裂缝进行像素级检测与测量。*Measurement Science and Technology*，32(6):065010。

Li, Y., Li, H., 和 Wang, H.（2018）。使用深度局部模式预测器进行逐像素裂缝检测的机器人应用。*Sensors*，18(9):3042。

Lin, T.-Y.、Goyal、Girshick、Girshick, R.、He, K. 和 Doll´ar, P.（2017a）。用于密集目标检测的焦点损失。载于*IEEE*国际计算机视觉会议论文集，第 2980--2988 页。

Lin, T.-Y.、Goyal、Girshick、Girshick, R.、He, K. 和 Doll´ar, P.（2017b年）。用于密集目标检测的焦点损失。载于*IEEE*国际计算机视觉会议论文集，第 2980--2988 页。

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll´ar, P., and Zitnick, C. L. (2014). Microsoft COCO：上下文中的常见物体。载于欧洲计算机视觉会议，第740--755页。Springer。

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). 用于实例分割的路径聚合网络。载于*IEEE*计算机视觉与模式识别会议论文集，第8759--8768页。

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., and Berg, A. (2016). SSD：单次多框检测器，‖载于欧洲计算机视觉会议（ECCV）。

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin Transformer：使用移位窗口的分层视觉Transformer。载于*IEEE/CVF*国际计算机视觉会议论文集，第10012--10022页。

Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., 和 Omata, H. (2021). 用于道路损伤检测的生成对抗网络。《计算机辅助土木与基础设施工程》, 36(1):47--60.

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., 和 Omata, H. (2018). 使用智能手机图像和深度神经网络进行道路损伤检测与分类。《计算机辅助土木与基础设施工程》, 33(12):1127--1141.

Majidifard, H., Jin, P., Adu-Gyamfi, Y., 和 Buttlar, W. G. (2020年). 路面图像数据集：用于路面病害分类和细化的新基准数据集。《交通研究记录》, 2674(2):328--339.

Mandal, V., Mussah, A. R., 和 Adu-Gyamfi, Y. (2020年). 路面病害分类的深度学习框架：比较分析。载于*2020*年 *IEEE*国际大数据会议（大数据），第 5577--5583 页。IEEE。

Misra, D.（2020年）。Mish：一种自正则化非单调激活函数。

Mohan, A. 和 Poobal, S.（2018）。使用图像处理进行裂缝检测：批判性综述与分析。《亚历山大工程学报》，57(2)：787--798。

Naddaf-Sh, S., Naddaf-Sh, M.-M., Kashani, A. R. 和 Zargarzadeh, H.（2020年）。一种高效且可扩展的道路损伤检测深度学习方法。载于*2020*年 *IEEE*国际大数据会议（大数据），第 5602--5608 页。IEEE。

Nath, N. D., Cheng, C.-S. 和 Behzadan, A. H.（2022年）。GPS拒止灾害现场损伤信息的无人机测绘。*Advanced Engineering Informatics*，51：101450。

Nhat-Duc, H., Nguyen, Q.-L. 和 Tran, V.-D.（2018）。使用元启发式优化边缘检测算法和卷积神经网络自动识别沥青路面裂缝。《建筑自动化》，94：203--213。

Ni, F., He, Z., Jiang, S., Wang, W. 和 Zhang, J.（2022年）。一种用于增强轻量级裂缝描绘网络的生成对抗学习策略。*Advanced EngineeringInformatics*，52：101575。

Oliveira, H. 和 Correia, P. L. (2009). 使用熵和图像动态阈值化进行自动道路裂缝分割。载于*2009*年第*17*届欧洲信号处理会议，第622--626页。IEEE。

Patra, S.、Middya, A. I. 和 Roy, S. (2021年). Potspot：一种基于参与式感知、利用深度学习进行坑洞检测的监测系统。《多媒体工具与应用》，80(16):25171--25195。

Quintana, M.、Torres, J. 和 Men´endez, J. M. (2015). 用于路面检测和养护的简化计算机视觉系统。《IEEE智能交通系统汇刊》，17(3):608--619。

Rawat, W. 和 Wang, Z. (2017). 用于图像分类的深度卷积神经网络：全面综述。《神经计算》，29(9):2352--2449。

Redmon, J., Divvala, S., Girshick, R., 和 Farhadi, A.（2016年）。You Only Look Once：统一的实时目标检测。载于*IEEE*计算机视觉与模式识别会议论文集，第 779--788 页。

Redmon, J. 和 Farhadi, A.（2017）。Yolo9000：更好、更快、更强。载于*IEEE*计算机视觉与模式识别会议论文集，第 7263--7271 页。Redmon, J. 和 Farhadi, A.（2018）。Yolov3：渐进式改进。

Ren, S., He, K., Girshick, R., and Sun, J.（2016年）。Faster R-CNN：利用区域建议网络实现实时目标检测。*IEEE*模式分析与机器智能汇刊，39(6):1137--1149。

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S.（2019年）。广义交并比：一种用于边界框回归的度量和损失。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，第658--666页。

Ribera, J., Guera, D., Chen, Y., and Delp, E. J.（2019年）。无需边界框定位目标。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，第6479--6489页。

Roy, A. M.（2022a）。基于自适应迁移学习的多尺度特征融合深度卷积神经网络，用于脑机接口中的脑电图运动想象多分类。*EngineeringApplications of Artificial Intelligence*, 116:105347。

Roy, A. M.（2022b）。一种高效的多尺度卷积神经网络模型，通过内在特征整合对脑机接口中的运动想象脑电图受试者进行分类。《生物医学信号处理与控制》，74:103496。

Roy, A. M.（2022c）。一种基于自适应迁移学习的多尺度融合卷积神经网络模型，用于脑机接口系统中的多类别运动想象分类。*BioRxiv*。

Roy, A. M. 和 Bhaduri, J.（2021年）。一种基于计算机视觉的深度学习多类别植物病害检测模型。人工智能，2(3):413--428。

Roy, A. M. 和 Bhaduri, J.（2022年）。一种用于高遮挡场景的实时生长阶段检测模型，采用 DenseNet 融合 YOLOv4。《农业中的计算机与电子学》，193:106694。

Roy, A. M., Bhaduri, J., Kumar, T., 和 Raj, K.（2022a）。一种基于计算机视觉的濒危野生动物检测目标定位模型。《生态经济学》，即将出版。

Roy, A. M., Bhaduri, J., Kumar, T., 和 Raj, K.（2022b）。Wildect-yolo：一种高效、稳健且精准的基于计算机视觉的濒危野生动物自动检测目标定位模型。*Ecological Informatics*，第 101919 页。

Roy, A. M. 和 Bose, R.（2023a）。基于物理信息神经网络的线弹性求解深度学习加速计算框架。*Neural Networks*，2:2。

Roy, A. M. 和 Bose, R.（2023b）。用于线弹性的物理感知深度学习框架。*arXiv*预印本 *arXiv:2302.09668*。

Roy, A. M.、Bose, R. 和 Bhaduri, J.（2022年c）。基于YOLOv4深度神经网络的快速、精确细粒度目标检测模型。*Neural Computing and Applications*，第1--27页。

Roy, A. M. 和 Guha, S.（2022年）。用于J2塑性的弹塑性物理信息深度学习方法。可从 *SSRN*获取：*https://ssrn.com/abstract=4332254*，第48页。

Roy, A. M. 和 Guha, S.（2023）。用于求解von Mises塑性的数据驱动物理约束深度学习计算框架。*Engineering Applications of Artificial Intelligence*，122:106049。

Shang, H., Sun, C., Liu, J., Chen, X., 和 Yan, R. (2023). 用于智能视觉表面缺陷检测的缺陷感知Transformer网络。*Advanced Engineering Informatics*, 55:101882。

Silva, W. R. L. d. 和 Lucena, D. S. d. (2018). 基于深度学习图像分类的混凝土裂缝检测。载于论文集，第2卷，第489页。MDPI AG。

Singh, A., Raj, K., Kumar, T., Verma, S., 和 Roy, A. M. (2023a). 用于自闭症治疗的基于深度学习、经济高效且响应灵敏的机器人。*Drones*, 7(2):81。

Singh, A., Ranjbarzadeh, R., Raj, K., Kumar, T., 和 Roy, A. M. (2023b). 理解用于按受试者定义Armoni活动的脑电信号。*arXiv*预印本 *arXiv:2301.00948*。

Stricker, R.、Eisenbach, M.、Sesselmann, M.、Debes, K. 和 Gross, H.-M.（2019年）。通过在扩展GAPS数据集上进行广泛实验改进视觉道路状况评估。载于*2019*年国际神经网络联合会议（*IJCNN*），第1--8页。IEEE。

Varadharajan, S.、Jose, S.、Sharma, K.、Wander, L. 和 Mertz, C.（2014）。道路检测视觉。载于*IEEE*冬季计算机视觉应用会议，第115--122页。IEEE。

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). 注意力就是一切。神经信息处理系统进展, 30。

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). 面向计算机视觉的深度学习：简要综述。计算智能与神经科学, 2018。

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2021a). Scaled-YOLOv4：扩展跨阶段局部网络。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 13029--13038 页。

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022年). YOLOv7：可训练的免费赠品包为实时目标检测器树立了新的最先进水平。*arXiv*预印本 *arXiv:2207.02696*。

Wang, C.-Y.、Liao, H.-Y. M.、Wu, Y.-H.、Chen, P.-Y.、Hsieh, J.-W. 和 Yeh, I.-H.（2020）。Cspnet：一种能够增强 CNN 学习能力的新型骨干网络。载于*IEEE/CVF*计算机视觉与模式识别会议研讨会论文集，第 390--391 页。

Wang, W.、Li, L. 和 Han, Y.（2021b）。基于移动窗口中的灰度偏差及连通分量之间的距离偏差检测阴影图像中的裂缝。《建筑与建材》，271:121885。

Wang, W.、Wu, B.、Yang, S. 和 Wang, Z.（2018a）。使用 Faster R-CNN 进行道路损伤检测与分类。载于*2018*年 *IEEE*国际大数据会议（大数据），第 5220--5223 页。IEEE。

Wang, Y. J., Ding, M., Kan, S., Zhang, S., 和 Lu, C. (2018b). 用于道路损伤检测与分类的深度提议与检测网络。载于*2018 IEEE*国际大数据会议（大数据），第 5224--5227 页。IEEE。

Woo, S., Park, J., Lee, J.-Y., 和 Kweon, I. S. (2018). CBAM：卷积块注意力模块。载于欧洲计算机视觉会议论文集（*ECCV*），第 3--19 页。

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., and Schneider, D.(2020). 使用 Mask R-CNN 在四旋翼无人机视觉系统中自动统计牛只数量。《农业中的计算机与电子学》，171:105300。

Xu, R., Hao, R., and Huang, B. (2022年). 利用自监督学习策略和分割网络高效检测表面缺陷。《Advanced Engineering Informatics》，52:101566。

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., and Yoo, Y. (2019年). CutMix：通过可定位特征训练强分类器的正则化策略。载于《IEEE/CVF国际计算机视觉会议论文集》，第6023--6032页。

Zhang, A., Wang, K. C., Li, B., Yang, E., Dai, X., Peng, Y., Fei, Y., Liu, Y., Li, J. Q., and Chen, C. (2017a). 使用深度学习网络在三维沥青路面上自动进行像素级路面裂缝检测。《计算机辅助土木与基础设施工程》，32(10):805--819。

Zhang, H., Cisse, M., Dauphin, Y. N., 和 Lopez-Paz, D.（2017b年）。MixUp：超越经验风险最小化。*arXiv*预印本 *arXiv:1710.09412*。

Zhang, L., Yang, F., Zhang, Y. D., 和 Zhu, Y. J.（2016年）。使用深度卷积神经网络进行道路裂缝检测。载于*2016*年 *IEEE* 国际图像处理会议（*ICIP*），第3708--3712页。IEEE。

Zhao, H., Qin, G., 和 Wang, X.（2010年）。基于路面边缘检测改进Canny算法。载于*2010*年第三届国际图像与信号处理大会，第2卷，第964--967页。IEEE。

Zhao, Z.-Q., Zheng, P., Xu, S.-t., 和 Wu, X.（2019a年）。深度学习目标检测：综述。《*IEEE*神经网络与学习系统汇刊》，30(11)：3212--3232。

Zhao, Z.-Q., Zheng, P., Xu, S.-t. 和 Wu, X.（2019b年）。基于深度学习的目标检测：综述。《*IEEE*神经网络与学习系统汇刊》，30(11):3212--3232。

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R. 和 Ren, D.（2020年）。距离交并比损失：更快、更好地学习边界框回归。载于人工智能促进协会人工智能会议论文集，第34卷，第12993--13000页。

Zhu, X., Lyu, S., Wang, X., and Zhao, Q.（2021年）。TPH-YOLOv5：基于Transformer预测头改进的YOLOv5，用于无人机拍摄场景中的目标检测。载于*IEEE/CVF*国际计算机视觉会议论文集，第2778--2788页。