工程与管理科学
Engineering and Management Science
第 7 卷◆第 6 期◆版本 1.0◆2025 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

人工智能预训练模型：为自动驾驶带来新突破

周文丽　母崇铭
桂林电子科技大学　广西桂林　541004

DOI:10.12238/ems.v7i6.13812

［摘　要］自动驾驶技术推动智能交通发展，但复杂场景下的目标检测困难重重。近年来，自监督学习在深度学习领域取得显著进展，在缺乏大规模标注数据的情况下展现出潜力。本文聚焦基于掩码自编码器的预训练方法，结合多视图对比学习应用于自动驾驶领域。在自然图像上进行预训练，可将优质特征迁移至目标检测任务。测试表明，我们的方法提升了检测效果，增强了自动驾驶的感知能力。

［关键词］自动驾驶；深度学习；目标检测；预训练模型

1 引言

自动驾驶技术正深刻变革交通领域，成为科技发展的热点。环境感知是自动驾驶的核心环节，它能让车辆理解周围环境，准确识别障碍物、道路标志和其他车辆等，为自动驾驶决策提供基础。而目标检测作为环境感知的关键技术，如同车辆的 “眼睛”，在图像和视频中精准识别、定位并分类各种交通元素，帮助车辆实时掌握交通状况，做出正确行驶决策，是实现自动驾驶的重要支撑。

2 相关研究

2.1 基于深度学习的目标检测算法

深度学习广泛应用于自动驾驶目标跟踪检测。CNN 被大量用于分析摄像头和激光雷达所采集的数据，以便能够快速、精准地识别和分类出环境中的各种物体。YOLO 算法是典型代表，它能够以较高的精度实时识别和分类物体。RNN 用于随时间推移分析传感器数据，进而预测未来事件和行为。不过，自动驾驶领域面临获取大规模高质量标注数据的难题，大量数据未被有效利用，目标检测的自监督学习发展滞后。

2.2 基于掩码自编码器的预训练

掩码自编码器（MAE）是自监督学习中无需人工标注的方法。它通过随机遮盖图像大部分区域，再重建被遮盖部分，从而高效学习图像深层次特征。MAE 能在无标签数据上训练，且学习到的特征具有良好迁移性，可应用于目标检测、语义分割等多种下游任务。本文基于 MAE 对自然图像进行预训练，结合多视图对比学习，验证其在自动驾驶场景的应用效果。

3 方法与模型

本文提出一种基于 MAE 结构的多视图对比学习自监督预训练网络，先在大型数据集上预训练学习通用特征，再在公开的自动驾驶行人检测数据集上微调。

网络结构图如图 1 所示。首先对输入图片进行数据增强，如添加随机噪声、随机裁剪等，得到增强视图。将原始视图和增强视图切分成图像块并序列化，进行随机掩蔽操作，之后分别输入两个不同的编码器。编码阶段，主视图序列直接编码成特征向量，通过反向传播更新主编码器；增强视图编码器借助主编码器的指数平均移动（EMA）更新，这种双编码器设计让主视图和增强视图从不同视角学习一致的特征表示，提升模型在下游任务的表现。

为进一步优化特征表示，引入记忆库（Memory Bank）机制，通过对比学习强化编码器学习效果。新批次数据经编码器处理后，主编码器和增强编码器生成的特征向量分别存入对应的 Memory Bank，保持特征表示更新。当前批次样本中，主、增强编码器生成的特征向量。

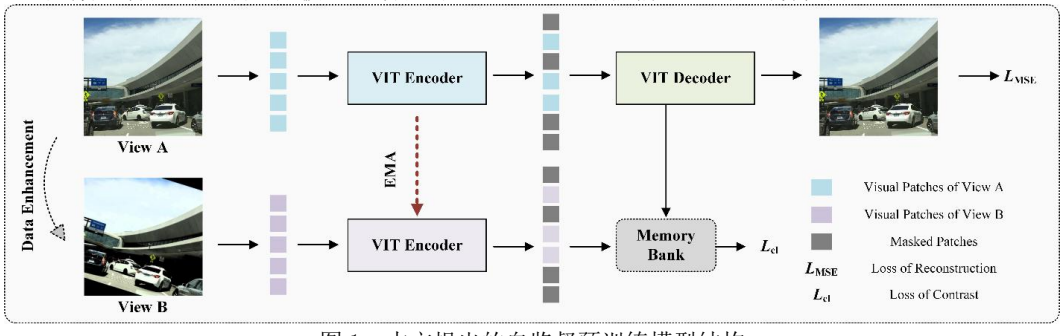

图 1　本文提出的自监督预训练模型结构

作为正样本对，从 Memory Bank 随机采样的特征向量作为负样本。通过最大化正样本对相似度，最小化与负样本的相似度，使模型学习到更具区分性的特征。对比损失计算公式为：

$$\mathcal{L}_{\text{contrast}} = -\ln\left(\frac{\exp\left(\frac{\text{sim}(z_i, z_i^+)}{\tau}\right)}{\exp\left(\frac{\text{sim}(z_i, z_i^+)}{\tau}\right) + \sum_{k=1}^{K}\exp\left(\frac{\text{sim}(z_i, z_k^-)}{\tau}\right)}\right) \quad (1)$$

其中，Z_i 和 Z_i^+分别是主编码器和增强编码器生成的同一图像的特征向量，Z_k^-是从 Memory Bank 中采样的负样本特征，$\text{sim}(\cdot,\cdot)$表示余弦相似度，τ 是温度参数，K 是负样本的数量。

解码阶段，主视图解码器对被掩蔽图像块重构，结合掩蔽标记生成完整图像表示。通过最小化重构损失，模型学习图像潜在结构和语义信息，重构损失函数为：

$$\mathcal{L}_{\text{MSE}} = \frac{\sum_{i=1}^{n}\left(f(x_i) - y_i\right)^2}{n} \quad (2)$$

其中，f （x_i）指的是模型对于第 i 个样本的预测结果，

文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

y_i 指的是第 i 个输入样本的实际目标值。网络的总损失则由上述的对比损失和重建损失加权得到，如公式 3 所示。

$$\mathcal{L}_{\text{all}} = \alpha \cdot \mathcal{L}_{\text{contrast}} + \mathcal{L}_{\text{MSE}} \quad (3)$$

预训练完成后，将预训练的编码器迁移到下游目标检测任务中，由于已在大型数据集上学习了通用特征，微调时能针对性地学习数据特征编码。

4 实验与结果分析

4.1 实验平台与数据集

实验硬件配置为 13th Gen Intel （R） Core （TM） i5 - 13500HX CPU、16GB 内存和 Tesla P100 - SXM2 - 16GB 显卡，足以处理复杂图像和深度学习任务。

实验使用 BDD100K 数据集，这是自动驾驶领域常用的大型数据集。它涵盖多种驾驶场景，包括不同天气（晴天、雨天、雪天等）、时间段（白天、夜晚）和路况（城市街道、高速公路、乡村小道等）。数据类型丰富，有高分辨率图像数据和详细标注信息，涉及目标检测、语义分割、实例分割等任务类别标注，为自动驾驶感知算法研究提供有力支持。在目标检测任务中，其清晰标注各类常见目标物体，助力模型学习在实际场景中准确识别目标，提升自动驾驶系统可靠性和安全性。

图 2　经过预训练后下游目标检测任务上的检测图

4.2 实验结果与分析

将基于 MAE 和多视图对比学习的预训练模型与传统监督学习目标检测模型对比，在相同测试数据集上，新模型准确率和召回率显著提升。通过设置不同参数实验，探究模型对参数的敏感性，并测试模型在不同天气和光照条件下的性能。结果表明，新模型在雨天、雾天、雪天以及强光、逆光、弱光等条件下，鲁棒性更强。预训练后迁移到下游目标检测任务的检测图展示了模型良好的检测效果。

5 讨论

5.1 方法的优点与创新点

基于 MAE 和多视图对比学习的预训练方法，解决自动驾驶目标检测标注数据难题，利用海量未标注数据降低成本，提升模型泛化能力。双编码器机制使模型从不同视角学习一致特征，增强特征学习能力，优化下游检测任务表现。大规模自然图像预训练覆盖多天气和光照条件，提升模型环境鲁棒性，保障自动驾驶安全。输出头通过 4 层卷积和 AnchorGenerator 处理特征，结合 QualityFocalLoss 等多损失函数加权优化，平衡分类与回归性能，提高检测精度。

5.2 方法的局限性与改进方向

5.2.1 计算资源需求大

预训练依赖大规模数据和复杂结构（双编码器、记忆库），对 GPU 资源要求高。未来可通过模型量化、剪枝技术压缩参数量，降低计算成本，推动实际部署。

5.2.2 多模态融合不足

当前方法聚焦图像自监督，缺乏多模态（图像、点云、雷达）整合方案。需探索多模态自监督架构，设计联合训练策略，早期融合多源数据特征，提升复杂场景检测准确性。

5.2.3 对复杂场景的适应性不足

模型在极端交通场景（严重拥堵、施工区）泛化能力有限。需构建包含极端场景的数据集，结合对抗训练、强化学习，增强模型对复杂情况的识别与应对能力。

6 总结

针对自动驾驶目标检测的标注数据瓶颈，本文提出基于 MAE 和多视图对比的预训练方法，通过自监督网络学习通用特征并迁移至检测任务，显著提升准确率、召回率及环境鲁棒性。研究存在计算资源需求高、多模态融合不足、复杂场景适应性待优化等问题。未来将通过模型压缩、多模态架构探索及极端场景数据增强，推动技术落地与性能突破。

[参考文献]

[1]侯佩玉，徐淼，张明，等.基于 YOLOv5s 的自动驾驶车辆行人检测方法[J].北华大学学报（自然科学版），2025，26（01）：107-114.

[2]陈其怀，林添良，马荣华，等.基于多模态的履带式挖掘机自动驾驶决策网络[J/OL].机械工程学报，1-9[2025-01-09].

[3]于康鸿，张军，刘元盛.基于 ACVAE-MPPO 算法的端到端自动驾驶算法研究[J/OL].计算机工程与应用，1-16[2025-01-09].

[4]He K，Chen X，Xie S，et al.Masked autoencoders are scalable vision learners[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.2022： 16000-16009.

[5]Jawahar G，Sagot B，Seddah D.What does BERT learn about the structure of language[C]//ACL 2019-57th Annual Meeting of the Association for Computational Linguistics.2019.

[6]Hart K.The memory bank： Money in an unequal world[M].London： Profile Books，2000.

[7]Tian Y，Sun C，Poole B，et al.What makes for good views for contrastive learning[J].Advances in neural information processing systems，2020，33： 6827-6839.

作者简介：

周文丽（2003—），本科，研究方向：计算机视觉；

母崇铭（2004—），本科，研究方向：计算机视觉。