

## 摘要

随着以深度学习为代表的机器学习技术的快速发展，人工智能的应用呈现出迅猛发展的增长趋势。但现有大部分深度学习方法都建立在闭集假设基础上，即训练数据与测试数据满足“独立同分布”假设。然而，现实场景往往是开放的且复杂多变的，包含众多未知的对象与分布，例如自动驾驶车辆在道路运行阶段可能会遇到未见过的场景或对象。这也对人工智能系统在开放视觉场景中的应用提出了新的挑战：一方面，可靠安全的人工智能系统能够检测出测试场景中与训练数据不同的未知对象；另一方面，人们希望人工智能系统能够自主对未知类进行区分与学习，从而具备提取新概念的能力。因此本文将聚焦未知类检测与新类别发现问题，主要研究开放场景下的未知类建模与学习方法。

本文从鲁棒未知类检测、未知类特征空间建模和增量新类别发现三个角度出发，探究了深度神经网络易将未知类识别为已知类的原因，并构建了面向未知类检测的优化学习方法，还研究了如何通过机器学习方法在无标注未知类数据中增量发现新类别，进而构建未知类检测与新类别发现的交互优化方法。本文的主要创新成果包括：

第一，针对深度神经网络易将未知类识别为已知类问题，提出了一种面向未知类检测的振幅相位重组优化方法。围绕深度神经网络对未知类过度自信现象，本文对深度神经网络与人类视觉系统在图像频域感知范围差异进行了系统性分析，首次推断出这种过度自信来源于神经网络对图像频域振幅谱的过度依赖。因此，提出了一种基于振幅相位重组的训练优化方法，通过引导深度神经网络对振幅相位的独立感知提升对未知类检测准确度。该方法通过基于振幅相位重组的数据增广策略分离图像的振幅和相位信息，并采用分离训练引导深度神经网络学习振幅或相位的独立特征。与现有方法相比，该方法可以有效缓解图像部分频域主导的网络过度自信现象。在国际基准数据集上的实验结果表明，本方法在未知类检测任务中将基础方法性能提升6%以上。

第二，针对开集空间风险难以评估问题，提出了一种面向未知特征空间建模的互惠点学习方法。闭集假设下未知类数据的缺失导致了未知类分布空间相似且有限。为了挖掘每种未知类的特点来建模潜在的未知类特征空间，本文提出了一种全新分类概念“互惠点”，即每个已知类别类外空间的代表性特征，并构建了基于对抗互惠点学习的未知类深度特征空间建模方法。该方法通过由互惠点和可学习半径构建的多个类外空间交集来近似建模潜在的未知类深度特征空间，并进一步在特征空间分离已知类和未知类。同时，设计了一种互惠点与已知类别间的对抗机制，生成分布在已知类特征边缘的混淆训练样本，用于优化模型对未知类检测的辅助训练。与现有算法相比，

该方法通过互惠点建模未知类的深度特征空间，从而在已知类训练中分离已知类与未知类特征。在国际基准数据集上的实验结果表明，本方法在未知类检测任务上比前沿方法性能提升了4%，并提供了更好的可解释性。

第三，针对无标注新类别发现与过拟合问题，提出了一种面向增量新类别发现的融合先验学习优化方法。增量新类别发现问题需要同时解决无标注数据中的新类别聚类以及已知类的灾难性遗忘问题。因此，本文提出了基于前序模型先验指导的可融合学习优化方法，将增量新类别发现任务划分为新类别发现与增量模型融合两阶段。在新类别发现阶段基于前序模型的先验知识，引入包含基础模型与新模型双分支结构，提高对未标注未知类数据中新类别聚类的准确性。在此基础上，提出了门控线性重参数化机制，在新类别学习阶段抑制新模型对基础模型的干扰，并在增量模型融合阶段实现双分支的无损融合。与现有算法相比，该方法充分利用了前序模型的先验信息指导，并通过新旧模型的无损融合抑制灾难性遗忘的产生。在国际基准数据集上的实验结果表明，本方法在增量新类别发现任务上比国际前沿方法性能提升了约10%，并进一步提升多种增量学习方法性能。

第四，构建了一种未知类检测与新类别发现的交互优化方法，用于未知类的自主检测与识别。该方法将未知类到新类别的转换分为未知类检测、新类别发现学习与增量模型融合三阶段。针对未知类检测问题，提出基于互惠点的分离学习机制，增强深度神经网络对未知类别的识别能力。在新类别发现阶段通过基于互惠点的新类别伪标签标注策略，提升新类别发现准确度，形成新类别发现与新类别学习结合的交互优化。在增量模型融合阶段，进一步通过门控线性重参数化机制实现新模型与基础模型无损融合。在国际基准数据集上的实验结果表明，本方法在增量新类别发现任务上比国际前沿方法性能提升了约12%，验证了未知类检测对新类别发现的促进作用。

综上所述，本文针对开放场景中任务经验风险、开集空间风险与遗忘风险对未知类建模与学习方法进行研究，并对未知类检测与新类别发现的交互优化作出了初步探索，不仅实现了对未知类的精准检测，也实现了对新类别有效学习，并在自动驾驶等实际应用场景验证了其有效性。此外，本文的研究工作为开放场景下人工智能自主探索领域开展更为广泛深入的研究奠定了基础，具有重要的理论意义和应用价值。

**关键词：**开放世界学习，开集识别，分布外检测，增量学习，新类别发现