摘要

图像分割是计算机视觉领域的一个基础任务，通过对图像和视频中的物体形状进行分割来对视觉场景进行深入的理解。其中实例分割任务融合了图像分类、目标检测和语义分割等多种基础识别算法的任务目标，被广泛应用于自动驾驶、医疗影像、工业机械臂等多个实际场景中。然而，现有的实例分割算法主要关注目标物体可见部分的形状，对被遮挡部分的形状关注较少。在实际场景中，物体之间相互遮挡的情况十分普遍，例如汽车与行人之间的相互遮挡以及医疗影像中肿瘤被器官遮挡等。在这些场景中，如果能够同时获得被遮挡物体的可见区域和被遮挡区域的形状，即目标物体的完整形状，则可以更加深刻和全面地理解和分析场景中的内容。预测目标物体完整形状的任务被称为非模态实例分割任务，已有的工作通过多种途径对该任务进行了一定的挖掘，但仍存在许多问题亟待探索。本论文首先基于现有的非模态实例分割任务的研究现状归纳出三个仍未被解决的重要问题，包括遮挡信息感知问题、形状多样性问题和单视角病态问题。通过对这三个十分具有挑战性的问题进行分析与探索，本文分别从遮挡信息建模、形状先验建模以及多视角融合建模三个方面递进地提出解决方法。本文取得的主要创新成果包括：

第一，针对遮挡信息感知问题，提出通过对遮挡信息进行建模以提高对目标物体的被遮挡情况的感知能力。在本研究中，首先分析了遮挡混淆问题对于非模态实例分割任务的影响。然后，通过对目标物体的被遮挡程度以及被遮挡形状两个方面进行理解和学习，提出了从实例信息和空间信息两个层面分别对遮挡信息从全局角度和局部角度进行建模的方法。具体来说，该方法结合 Transformer 网络的结构，提出了遮挡判别网络以学习具有遮挡程度感知能力的初始特征，以及具有遮挡形状感知能力的注意力掩码。实验结果表明，该方法在非模态实例分割任务的常用公开数据集 D2SA 和 COCOA-cls 上能够获得优异的性能。这一结果证明了通过对遮挡信息进行建模的方式能够提高对模型目标物体被遮挡情况的感知能力。

第二，针对形状多样性问题中的二维仿射变换问题，提出基于二维形状先验的非模态实例分割算法。该方法提出一个新的框架将形状先验的学习与形状的变换进行解耦，以降低二维仿射变换对非模态形状学习的影响，提高对不同二维仿射变换下的被遮挡形状的泛化性能。基于这一新框架，该方法能够形成对被遮挡物体的完整形状的理解与认知，即使对于训练集中没有的样例也能够较为准确的进行预测。实验结果表明，该方法在三个公开的非模态实例分割任务数据集（D2SA、COCOA-cls 和 KINS）上均表现出不错的性能，证明了该方法在解决形状多样性问题中的二维仿射变换问题上

的有效性。

第三，为了解决形状多样性问题中的三维视角变换问题，提出一种基于三维形状先验的非模态实例分割算法。该方法将被遮挡的二维实例和完整的三维模型联系起来，通过理解物体在三维空间的真实形状以应对三维视角变换所带来的二维形状变化问题。此外，为了应对三维物体的形状多样性问题，该方法在大型三维重建数据集上进行预训练，以获得高质量的预测结果。该方法使用了无监督的单视角三维重建方法来避免对三维模型真实数据的依赖。在非模态实例分割任务公开数据集上的实验证明了该方法所提出的三维形状先验模型对于新视角具有很好的泛化性能，对于非模态实例分割问题也能够表现出不错的性能。

第四，为了解决单视角病态问题，提出了一种多视角信息融合建模的方法，以利用多视角信息来减少被遮挡区域的形状不确定性。由于现有的非模态实例分割研究都是基于单视角的设定的，缺乏基于多视角下的任务设定，因此首先提出了一个新的任务——多视角非模态分割，并针对此任务提出了一个基于合成方法生成的多视角非模态分割数据集。该数据集是第一个用于探索多视角实例分割任务的数据集。此外，针对上述所提出的新任务，还提出了一种新的方法，用于解决如何高质量地融合不同视角和不同实例之间特征的问题。实验结果表明，该方法能够有效利用多视角信息来提升非模态分割的性能。此外，还通过实验验证了该数据集和该方法在真实场景中的非模态分割任务中的有效性。

综上所述，本文系统地研究了面向遮挡场景的非模态实例分割问题，通过针对单视角下的遮挡信息建模问题和形状多样性问题，以及多视角下的特征融合问题进行研究来提升非模态实例分割算法的性能。本文提出的方法能够在真实场景中具有不错的泛化性能，具有重要的理论意义与实用价值。

关键词：图像识别，实例分割，遮挡场景，非模态感知

Amodal Instance Segmentation for Occluded Images

Zhixuan Li (Computer Application Technology)
Directed by: Prof. Tiejun Huang and Prof. Tingting Jiang

ABSTRACT

Segmentation is a fundamental task in computer vision, which aims to deeply understand visual scenes by segmenting objects in images and videos. Instance segmentation, which integrates various basic recognition algorithms such as image classification, object detection, and semantic segmentation, has been widely applied in practical scenarios such as autonomous driving, medical imaging, and industrial robotic arms. However, current instance segmentation algorithms mainly focus on the visible shape of the target object, paying little attention to the shape of the occluded parts. In actual scenes, object occlusion is very common due to various reasons, such as occlusion between cars and pedestrians or the tumor being occluded by organs in medical imaging. In these scenes, obtaining the complete shape of the target object, including both visible and occluded regions, can lead to a more comprehensive understanding of the content in the scene. The task of predicting the complete shape of the target object is known as amodal instance segmentation, and existing research has explored this task through various approaches, but many issues still need to be solved. This thesis first summarizes three important challenges that have not yet been solved based on the current researches of amodal instance segmentation, including occlusion information perception, shape diversity, and single-view ill-posed problems. This thesis proposes solutions from three aspects: occlusion information modeling, shape prior modeling, and multi-view information fusion. The main contributions of this thesis include:

The first study analyzes the impact of occlusion on amodal instance segmentation tasks and proposes a method to model occlusion information in order to improve the perception of occluded objects. This is achieved by understanding and learning the degree and shape of occlusion from both global and local perspectives using a Transformer network. The proposed method uses an occlusion discrimination network to learn an occlusion-aware initial query vector and an occlusion-aware attention shape mask. Experimental results on the D2SA and COCOA-cls benchmarks show that the proposed method effectively handles the occlusion confusion problem in amodal instance segmentation tasks.

The second study deals with the problem of shape diversity, specifically the effect of affine transformations on amodal shape learning. The proposed method introduces a new framework that decouples the learning of shape priors from the effects of affine transformations, which improves the generalization performance of amodal shape learning under different affine transformations. Experimental results on the D2SA, COCOA-cls, and KINS benchmarks demonstrate the effectiveness of the proposed method in addressing the problem of affine transformations in shape diversity.

The third study proposes a 3D shape prior model that is learned through unsupervised single-view reconstruction to handle the problem of shape diversity caused by 3D viewpoint changing. The proposed method links occluded 2D instances with complete 3D models, which enables the understanding of the shape of objects in 3D space. Experimental results demonstrate that the proposed 3D shape prior model has good generalization performance for new viewpoints and can perform well on amodal instance segmentation tasks.

The fourth study proposes a method that uses multi-view information to reduce shape uncertainty in occluded regions and introduces a new task of multi-view amodal segmentation, along with a corresponding dataset. The proposed dataset uses a synthetic approach to create a multi-view amodal segmentation dataset for accurate annotations. Besides, a new method to fuse features from different views and instances is proposed. Experimental results demonstrate that the proposed method effectively uses multi-view information to improve amodal instance segmentation performance.

In summary, this thesis systematically investigates the problem of amodal instance segmentation in occlusion scenarios. By progressively addressing the modeling of occlusion information and shape diversity issues in a single view, as well as the feature fusion problem in multiple views, the performance of amodal instance segmentation algorithms is improved. The proposed method in this thesis demonstrates well generalization performance in real-world scenarios, therefore it has important theoretical significance and practical value.

KEY WORDS: Image Recognition, Instance Segmentation, Occluded Scenario, Amodal Perception