# **A Tri-Layer Plugin to Improve Occluded Detection**

## Guanqi Zhan, Weidi Xie, Andrew Zisserman Visual Geometry Group, University of Oxford

## Introduction

### Occlusion is very common in the 3D world

- One object is in front of another
- A portion of the scene disappears behind the non-transparent object that is closer to the viewer

**Detecting/Segmenting occluded objects** still remains a challenge for state-of-the-art object detectors



## **Data Preparation**

Amodal Completion & Occlusion Reasoning



Original Image



Depth Map



A's Modal Mask



Object A: Umbrella

A's Amodal Completion



Dataset

Separated COCO

Occluded COCO

A's Modal Mask on Depth Map



# Total Object

3522

5550



Dataset

Occluder Masks

Occludee Masks

Object B: Person

B's Modal Mask

B's Amodal Completion

Generated Training Datasets: Occluder & Occludee Masks Target Object Occluder Mask **Original Image** 

Occludee Mask



### Generated Evaluation Datasets: Occluded COCO v.s. Separated COCO Original Image **Original Image** GT Mask





GT Mask







B's Modal Mask on Depth Map





### **Experiment Results**

### Comparison with State-of-the-Art

Detector	Backbone	Plugin	Recall	Recall	val mAP		test-dev mAP	
		U	Occluded	Separated	BBox	Mask	BBox	Mask
Mask R-CNN	Swin-T [1]	-	3264(58.81%)	1125(31.94%)	46.0	41.6	46.3	42.0
Mask R-CNN	Swin-T	bi-layer	3315(59.73%)	1147(32.57%)	46.3	42.0	46.5	42.3
Mask R-CNN	Swin-T	ours	3441(62.00%)	1223(34.72%)	48.5	43.0	48.7	43.4
Mask R-CNN	Swin-S [1]	-	3393(61.14%)	1186(33.67%)	48.5	43.3	49.0	44.1
Mask R-CNN	Swin-S	ours	3473(62.58%)	1261 (35.80%)	50.3	44.2	50.6	44.9
Cascade Mask R-CNN	Swin-B [1]	_	3491(62.90%)	1279(36.31%)	51.9	45.0	52.6	45.6
Cascade Mask R-CNN	Swin-B	ours*	3532(63.64%)	1299 (36.88%)	52.1	45.4	52.7	45.9

For evaluation, we introduce two extra measures - **Recall on Occluded COCO** and Recall on Separated COCO to evaluate model's capability of detecting partially occluded or separated objects. The plugin can always improve the number of recalled objects for both Occluded COCO and Separated COCO, which demonstrates the effectiveness of our plugin. The improvement on occluded objects could be transfered to improve the final mAP. We can see the overall detection performance reflected by bbox OpenImages and mask mAP is consistently boosted for different architectures.

### Ablation Study

Model	Tri-Layer Modelling	BBox Adjustment	RoI Feature Re-weighting	Fine-tuning Whole Network?	Recall Occluded	Recall Separated	BBox mAP	Mask mAP
B1					3264(58.81%)	1125(31.94%)	46.0	41.6
B2		$\checkmark$			3296(59.39%)	1141(32.40%)	47.9	42.2
B3	$\checkmark$				3339(60.16%)	1157(32.85%)	46.0	41.9
B4	$\checkmark$	$\checkmark$			3400(61.26%)	1187(33.70%)	48.1	42.5
B5	$\checkmark$	$\checkmark$	$\checkmark$		3410(61.44%)	1208(34.30%)	48.2	42.8
C1		$\checkmark$		$\checkmark$	3367(60.67%)	1170(33.22%)	48.3	42.5
C2	$\checkmark$			$\checkmark$	3360(60.54%)	1159(32.91%)	46.3	42.2
C3	$\checkmark$	$\checkmark$		$\checkmark$	3434(61.87%)	1208(34.30%)	48.3	42.9
C4	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	3441(62.00%)	1223 (34.72%)	48.5	43.0

Only fine-tuning the head could already contribute the majority of the improvement, validating the effectiveness of our proposed module as a general 'plugin', which can be inserted into pre-trained detectors, and give quick performance improvement

GΤ Our Model Swin-T + Mask R-CNN



COCO

COCO

OVIS

KINS