

Humans need not *label* more humans:

Occlusion Copy & Paste for Occluded Human Instance Segmentation

Evan Ling, Dezhao Huang, Minhoe Hur

AIR Center, Hyundai Motor Group Innovation Center in Singapore



A Difficult Problem: Same-class Occluded Instance Segmentation

0

We look at this harder problem of same-class occluded instance segmentation that remains unsolved by modern models. Good model-centric approaches to tackle occlusions have been introduced over the years, but they are supervised methods. We can collect and label relevant occluded data, but these hard scenarios are extremely costly to acquire and label.

Predictions by Mask R-CNN fine-tuned on MS COCO - Person dataset

Current methods tackling Occluded Instance Recognition

Model-Centric Approaches

OCFusion [1]

Pose2Seg [3]

Supervised methods decline we have labelled & relevant data

BCNet [2]

PoSeg [4]

Data-Centric Approaches

Occlusion Copy & Paste (ours)

[1] Lazarow et al. CVPR, 2020. [2] Ke et al. CVPR, 2021. [3] Zhang et al. CVPR, 2019. [4] Zhou et al. IEEE Access, 2020.

Starting from our simple approach:

1

With these considerations in mind, we introduce a data-centric approach, Occlusion Copy & Paste to directly induce occluded scenarios and feed relevant examples to model during training as an online augmentation approach. This allow us to tackle the problem for "free" without additional data or labels.

Basket of Images

Pasting Instances

Copy

Paste

Training Dataset

Training Image + Annotations

Train

2

Example image augmented with our Copy & Paste augmentation

Blue: Original | Pink: Pasted

Baseline results right off the bat

Training Approach	OCHuman		OCHuman ^{FL}		COCO
	<i>AP^{val}</i>	<i>AP^{test}</i>	<i>AP^{val}</i>	<i>AP^{test}</i>	<i>AP^{val}_{person}</i>
Pre-trained from [5]	14.9	14.9	24.5	24.9	47.5
Baseline vanilla training	16.5	16.6	27.0	27.4	48.7
+ Basic Copy & Paste (ours)	18.6	17.8	29.3	28.5	49.2

[5] Kai et al. Mmdetection: Open mmlab detection toolbox and benchmark. arXiv:1906.07155, 2019.

Add-ons

Are our synthesized images realistic?

- (1) Photo-Realism: We got around this by copy & pasting
- (2) Semantic Realism: Relatively lacking here

But do we really need perfect realism?

Realism Enhancements:

Min Size Filter

Scale-aware Pasting

Better quality masks

Blending

We experiment if we require perfect realism in our augmentation in other to train generalisable model. As it turns out, imposing such realism controls actually restricts scope of variability and does not improve performance.

Realism isn't favoured over variability

copy & paste add-ons	<i>AP^{val}</i>		<i>AP^{test}</i>	
Basic Copy & Paste	18.6		17.8	
Minimum Pasting Size	18.8		18.5	
Minimum Pasting Size + Scale Aware	18.2	-0.6	18.2	-0.3
Minimum Pasting Size + Better Quality Mask	18.4	-0.4	18.0	-0.5
Minimum Pasting Size + Blend (Fixed)	18.6	-0.2	17.9	-0.6
Minimum Pasting Size + Blend (Random)	19.0	+0.2	18.4	-0.1

Increasing Efficiency & Variability

copy & paste add-ons	<i>AP^{val}</i>		<i>AP^{test}</i>	
Basic C&P, $R_{paste} = [1, 10]$	18.6		17.8	
+ Targeted	18.6	+0.0	18.2	+0.4
Basic C&P, $R_{paste} = [1, 3]$	17.9		17.5	
+ Targeted	19.1	+1.2	18.0	+0.5
+ Targeted & Augm. Paste	19.2	+1.3	18.4	+0.9
+ Targeted, Augm. Paste & Min. Size	19.5	+1.6	18.6	+1.1

Basic Copy & Paste

+

Targeted Pasting

Augmented Instance Pasting

Realism Enhancements: Min Size Filter

Occlusion Copy & Paste

Our eventual Occlusion Copy & Paste on the SOTA

Model	External Pose Model	Modelled for Occlusion	OCHuman		OCHuman ^{FL}	
			<i>AP^{val}</i>	<i>AP^{test}</i>	<i>AP^{val}</i>	<i>AP^{test}</i>
Pose2Seg [§] [3]			-	-	22.8 ⁺	22.9 ⁺
+ Occlusion C&P (ours)	✓	✓	-	-	25.3 ⁺	25.1 ⁺
Mask R-CNN [§] [6]			14.9	14.9	24.5	24.9
Mask R-CNN [†]	✗	✗	16.5	16.6	27.0	27.4
+ Occlusion C&P (ours)			19.5	18.6	30.6	29.9
PoSeg (JoPoSeg) [4]	✗	✓	25.8 [*]	26.4 [*]	-	-
PoSeg (ExPoSeg)	✓	✓	26.4 [*]	26.8 [*]	-	-
Mask2Former [§] [7]			25.9	25.4	43.2	44.7
Mask2Former [†]	✗	✗	26.7	26.3	45.2	46.4
+ Simple Copy-Paste [8]			28.0	27.7	48.9	50.2
+ Occlusion C&P (ours)			28.9	28.3	49.3	50.6

[3] Zhang et al. Pose2seg. CVPR, 2019. [6] He et al. Mask r-cnn. ICCV, 2017. [4] Zhou et al. Poseg. IEEE Access, 2020.

[7] Cheng et al. Mask2Former. CVPR, 2022. [8] Ghiasi et al. Simple copy-paste. CVPR, 2021.

★ Easily interoperable with any models, demonstrating strong potential of data-centric approaches

MaskRCNN

MaskRCNN w/ OC&P (ours)

Mask2Former

Mask2Former with Occlusion Copy & Paste (ours)

6

