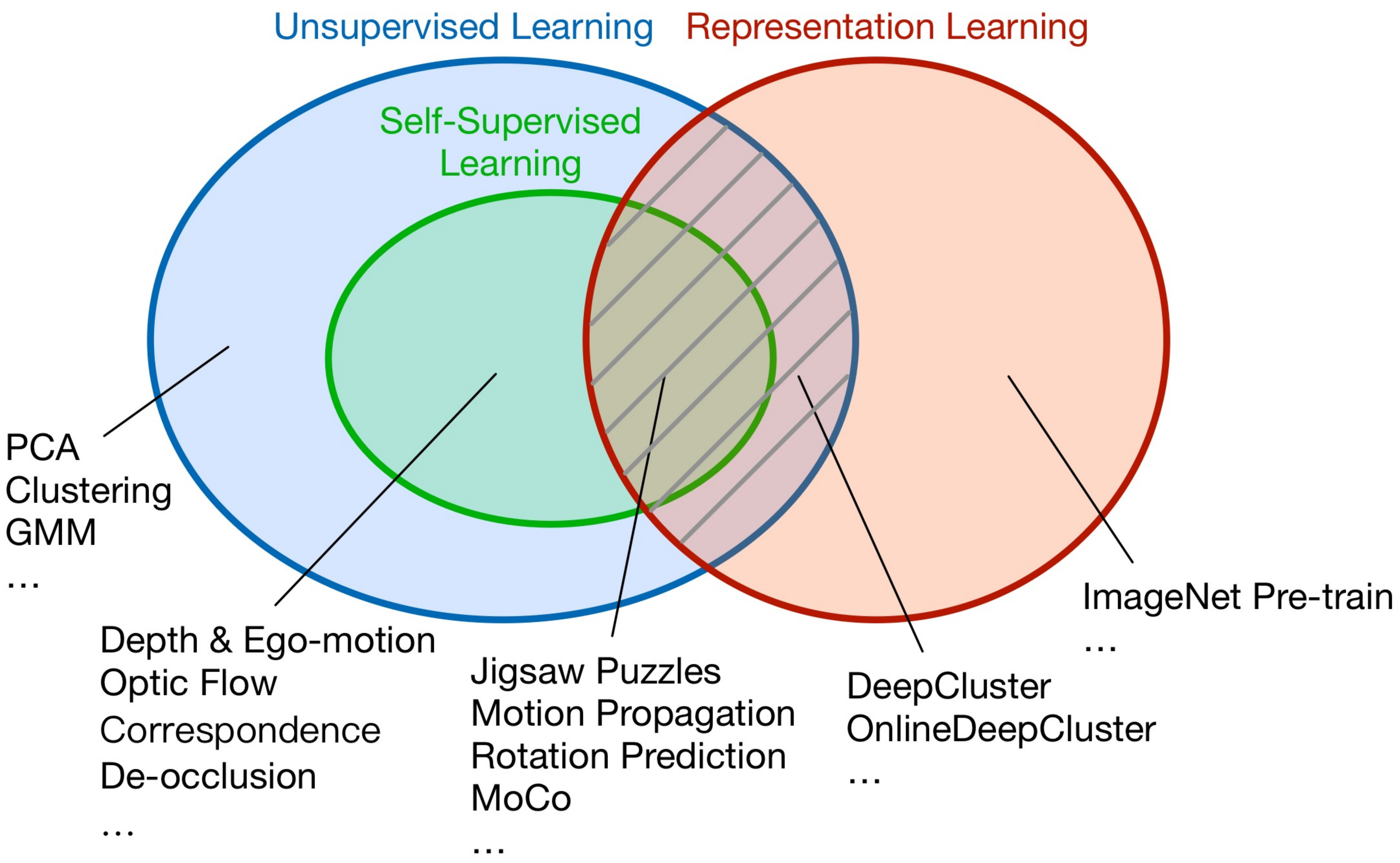
Self-Supervised Learning

Xiaohang Zhan
MMLab, The Chinese University of Hong Kong

June 2020

What is Self-Supervised Learning?



What is Self-Supervised Learning?

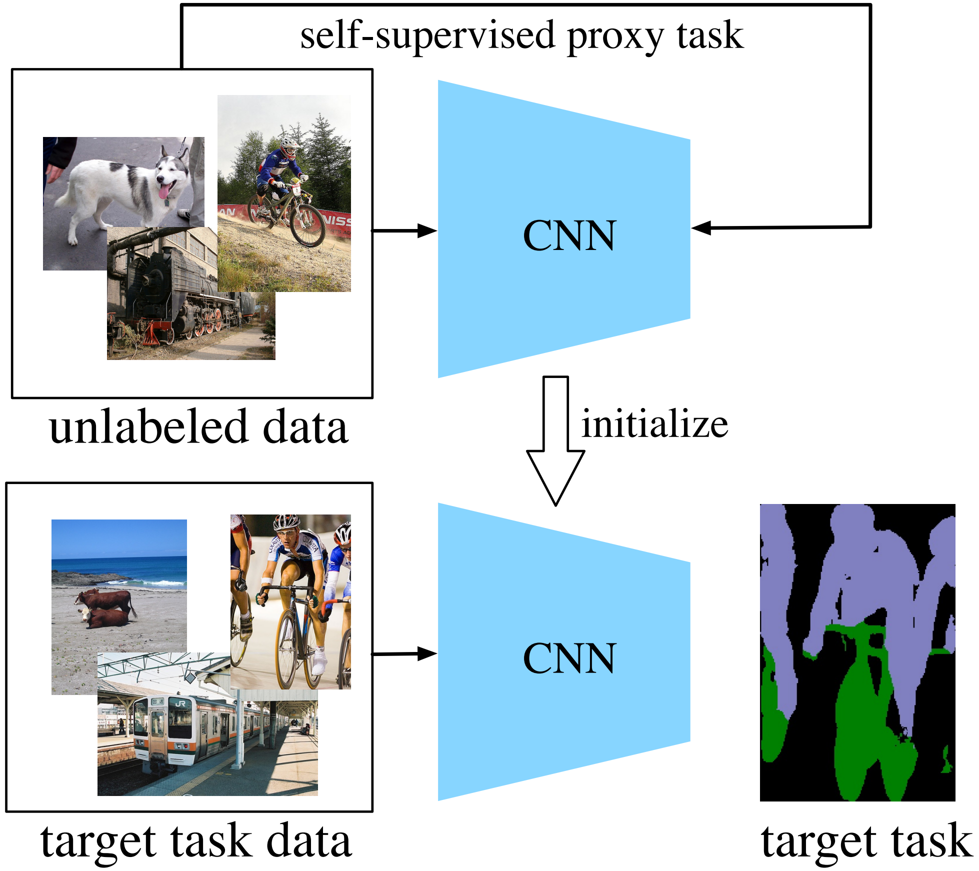
Self-Supervised

Annotation free

Learning

To learn something new

Does image in-painting belong to self-supervised learning?



A typical pipeline

Self-Supervised Proxy/Pretext Tasks

Image Colorization

Solving Jigsaw Puzzles

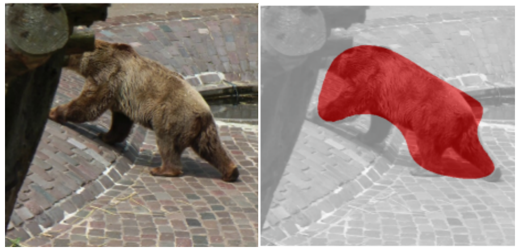
Image In-painting

Rotation Prediction

Instance Discrimination

Counting

Motion prediction



Moving foreground segmentation

Motion propagation

Why does SSL learn new information?

Prior

- Appearance prior

Image Colorization

Image In-painting

- Physics prior

Rotation Prediction

- Motion tendency prior

Motion prediction
(Fine-tuned for seg: 39.7% mIoU)

- Kinematics prior

Motion propagation
(Fine-tuned for seg: 44.5% mIoU)

Low-entropy priors are more predictable.

Coherence

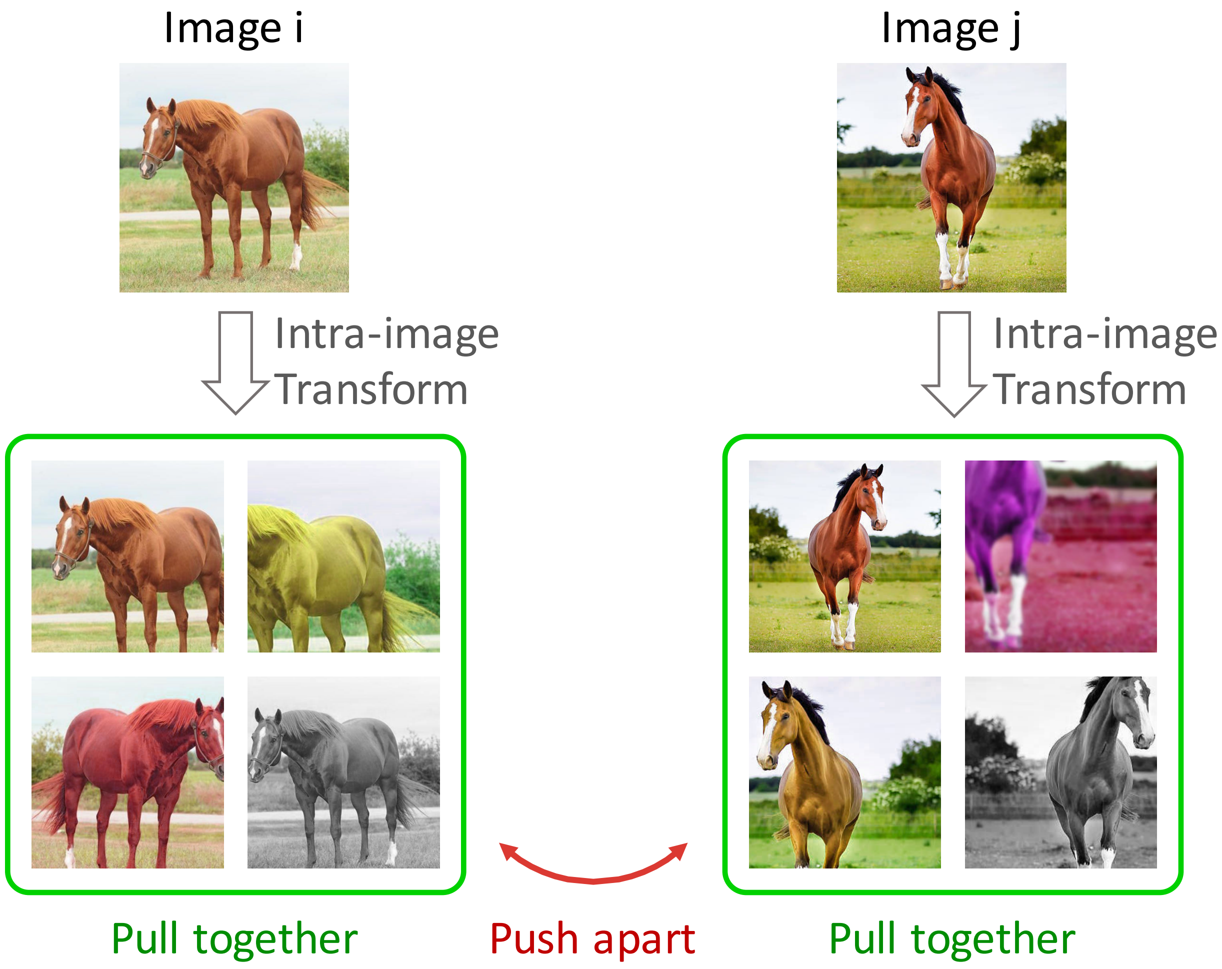
- Spatial coherence

Solving Jigsaw Puzzles

- Temporal coherence

Temporal order verification

Structure



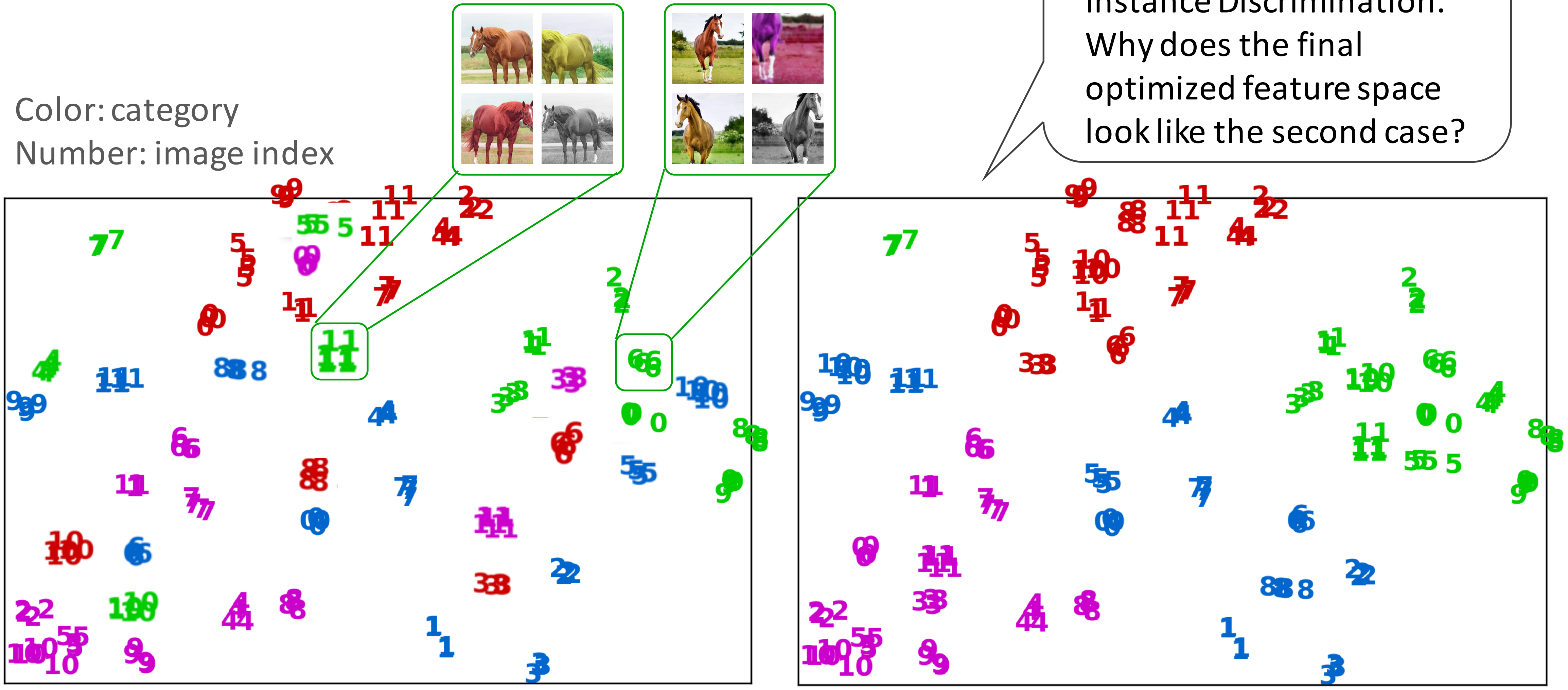
Instance Discrimination
(Contrastive Learning)

- NIPD
- CPC
- MoCo
- SimCLR
- ...

Structure

Color: category
Number: image index

Both are optimal for Instance Discrimination. Why does the final optimized feature space look like the second case?

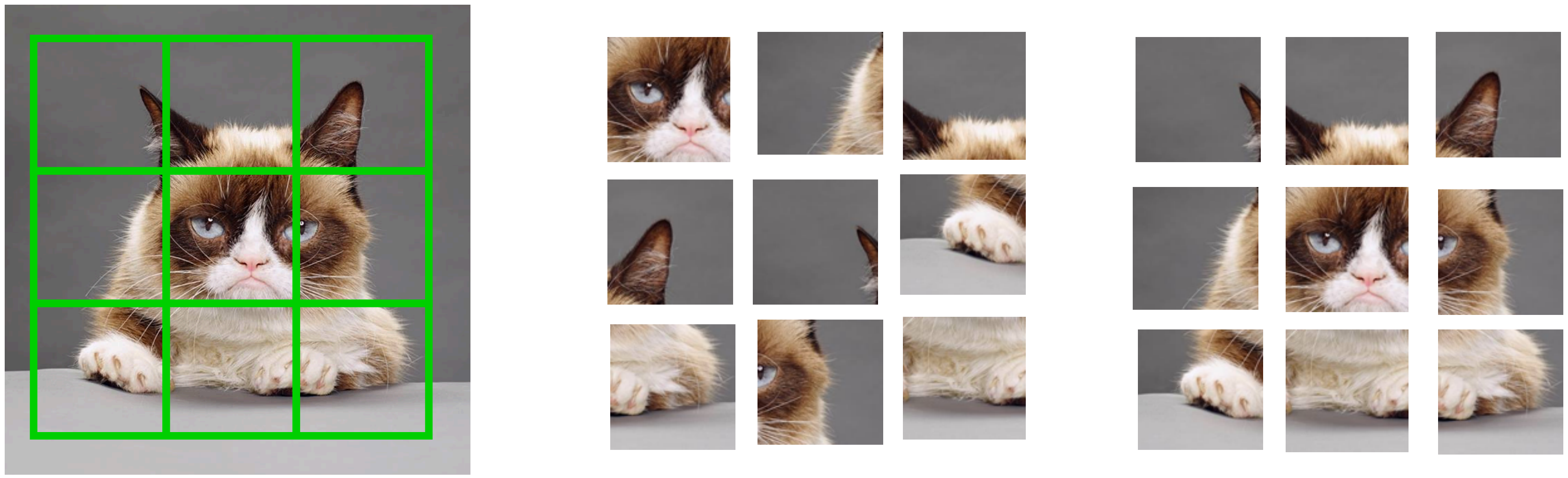


Optimal solution #1 Optimal solution #2

What to consider in proxy task design?

Shortcuts

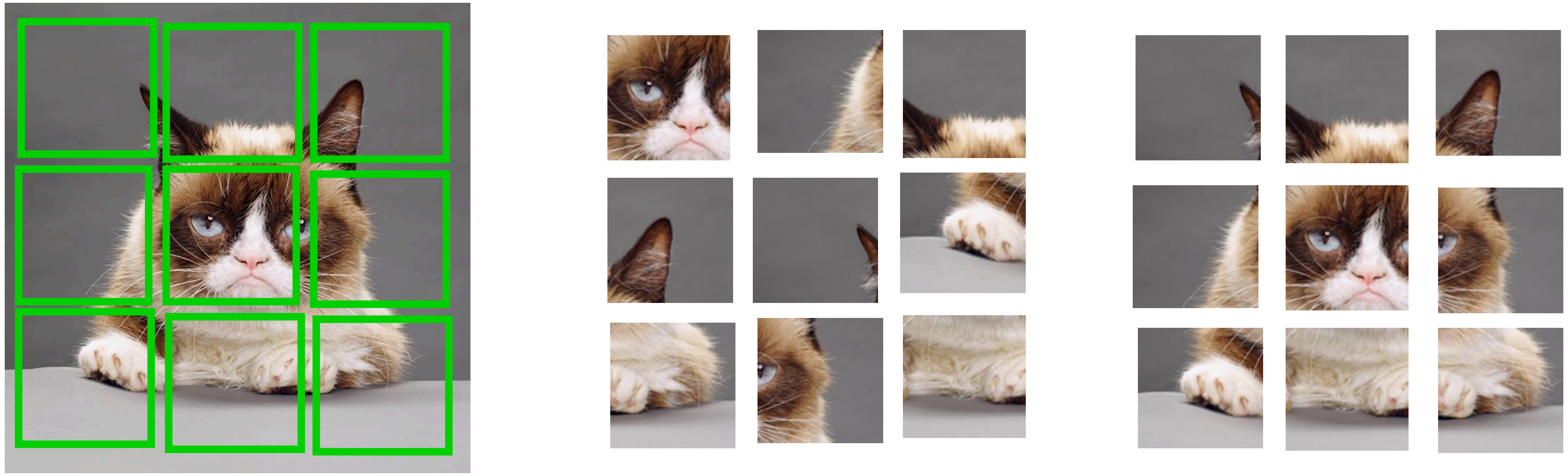
- Continuity



Solving Jigsaw Puzzles

Shortcuts

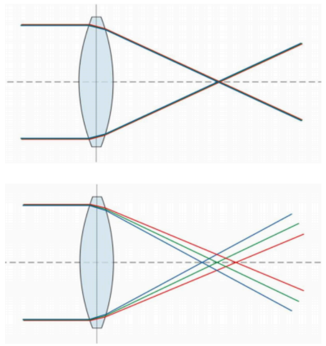
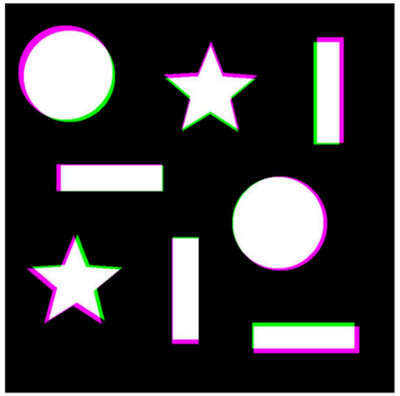
- Solution regarding continuity



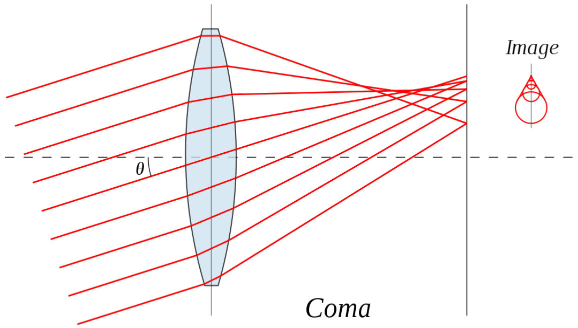
Solving Jigsaw Puzzles

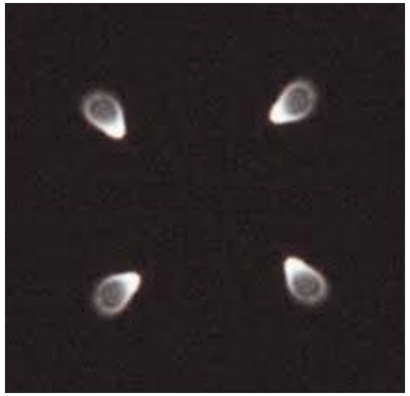
Shortcuts

- Chromatic Aberration

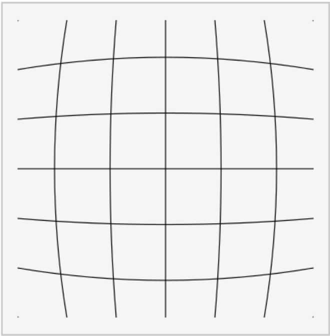



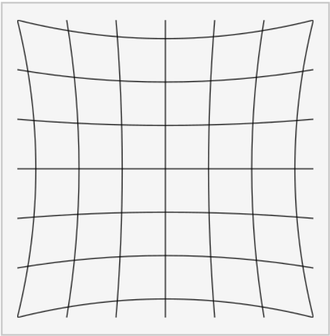
- Coma





- Distortion


Barrel-type

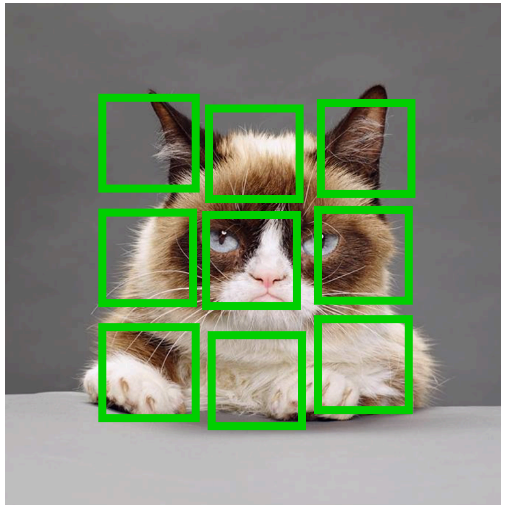

Pincushion-type

- Vignetting

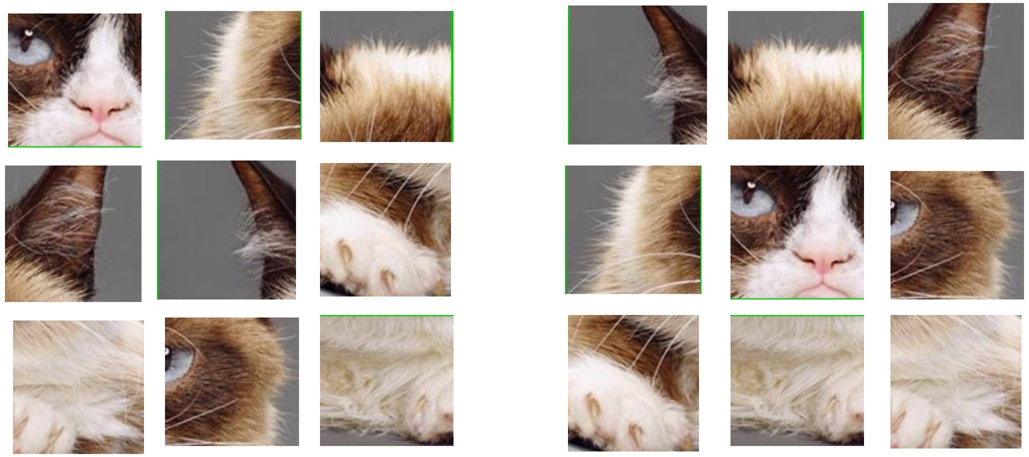

Do not apply heavy vignetting effects in your photos!!!

Shortcuts

- Solution regarding aberration



After aberration correction



Solving Jigsaw Puzzles

Ambiguity

- Appearance prior

Image Colorization

Image In-painting

- Physics prior

Rotation Prediction

- Motion tendency prior

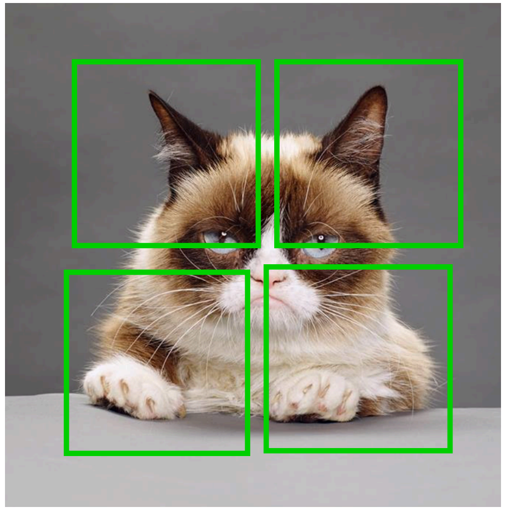
Motion prediction
(Fine-tuned for seg: 39.7% mIoU)

- Kinematics prior

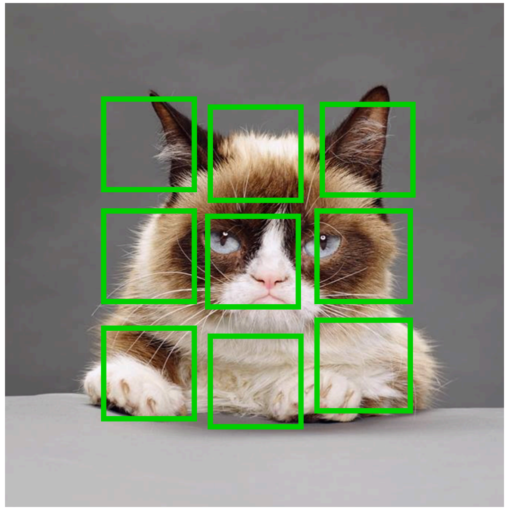
Motion propagation
(Fine-tuned for seg: 44.5% mIoU)

1. Low-entropy priors are less ambiguous.
2. Any other solutions?

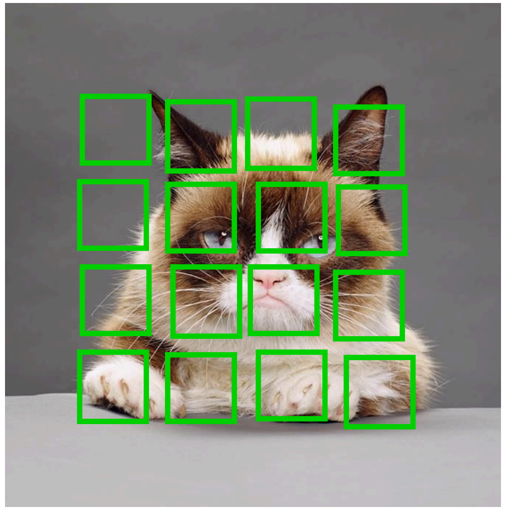
Difficulty



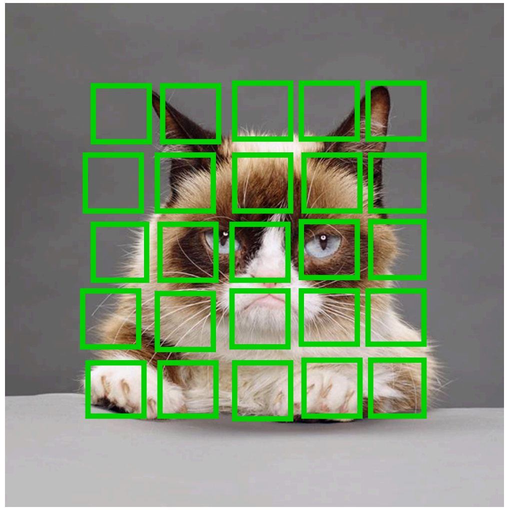
Easy mode



Normal mode



Difficult mode



Hell mode

How to design the difficulty of the task?

Summary

1. Learning from unlabeled data is feasible through:
   a) prior
   b) coherence
   c) structure

2. In designing proxy tasks, you have to consider:
   a) shortcuts
   b) ambiguity
   c) difficulty

Self-Supervised Scene De-occlusion

Xiaohang Zhan[1], Xingang Pan[1], Bo Dai[1], Ziwei Liu[1], Dahua Lin[1], Chen Change Loy[2]

[1]MMLab, The Chinese University of Hong Kong

[2]Nanyang Technological University

CVPR 2020 Oral

What We Have

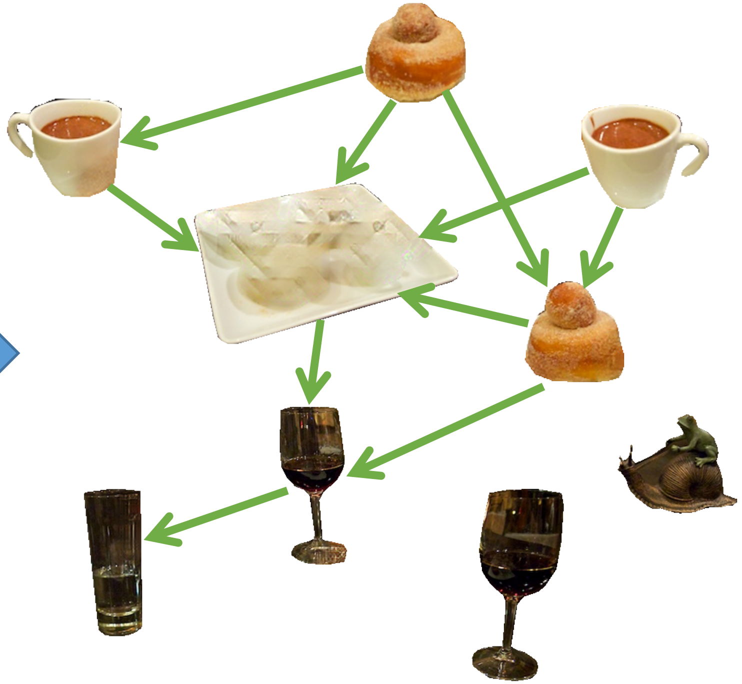
- A typical instance segmentation dataset:

RGB image

Modal masks & Category labels

Scene De-occlusion

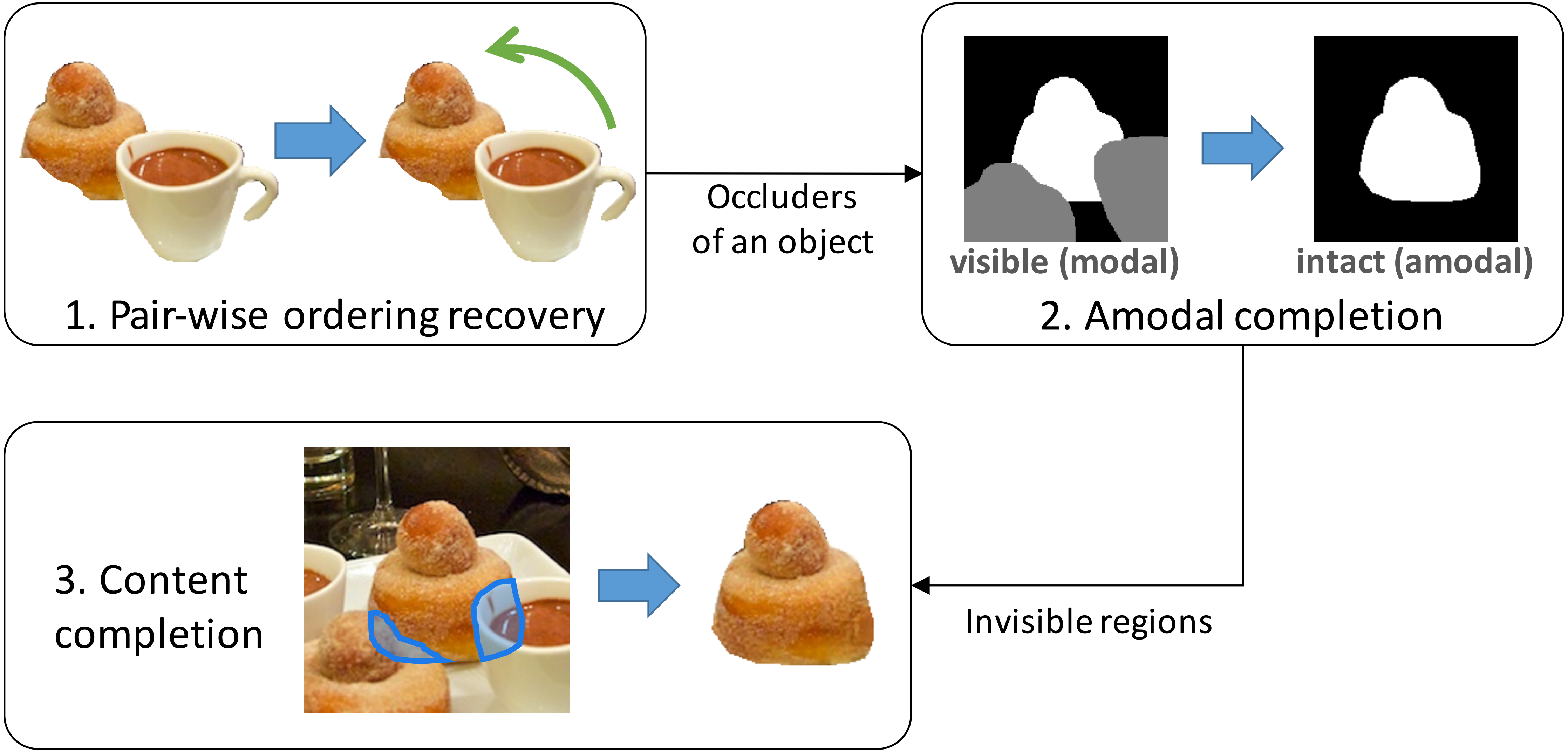
Real-world scene



Intact objects with invisible parts
+ ordering graph

Background

Tasks to Solve



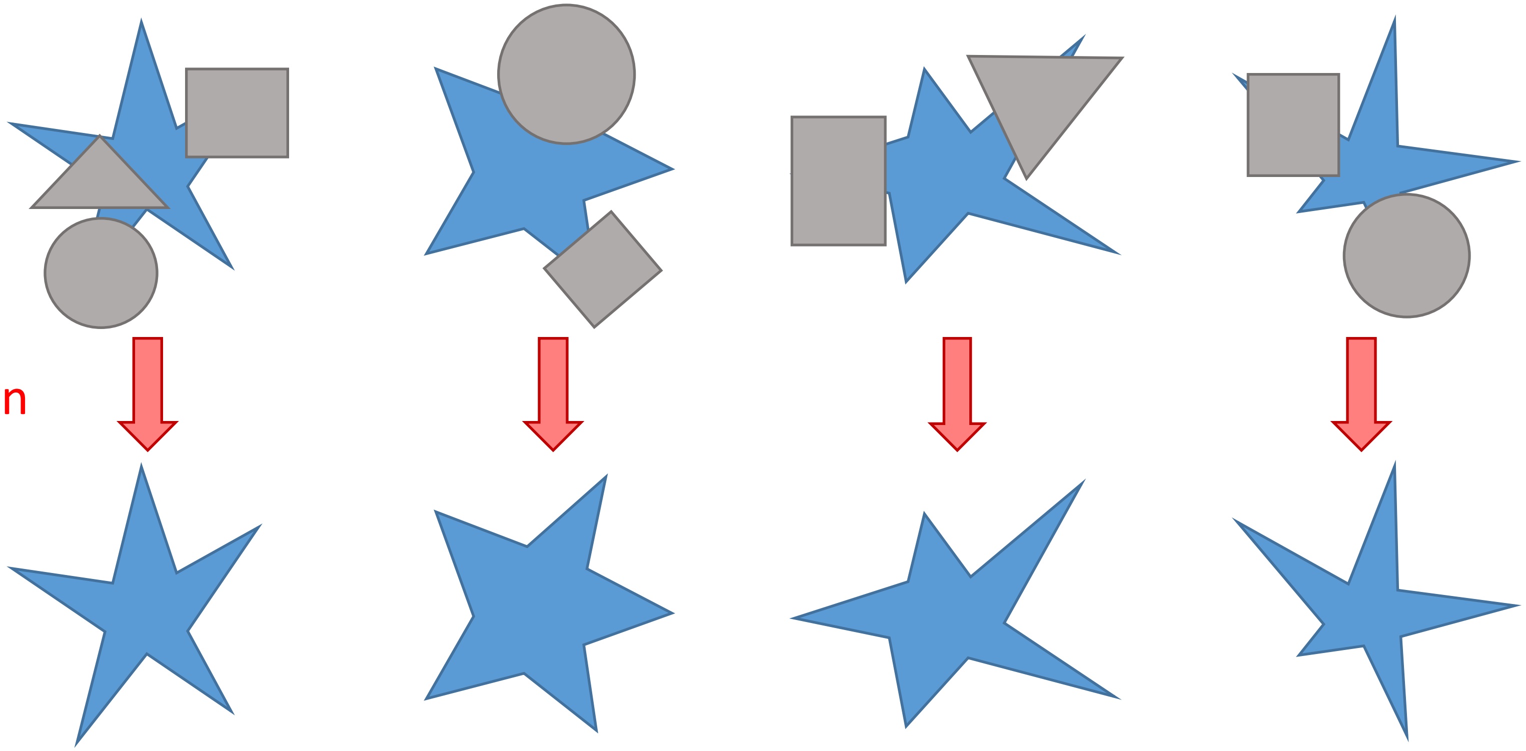
香港中文大學
The Chinese University of Hong Kong
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Amodal Completion

What's the shape of the blue objects?

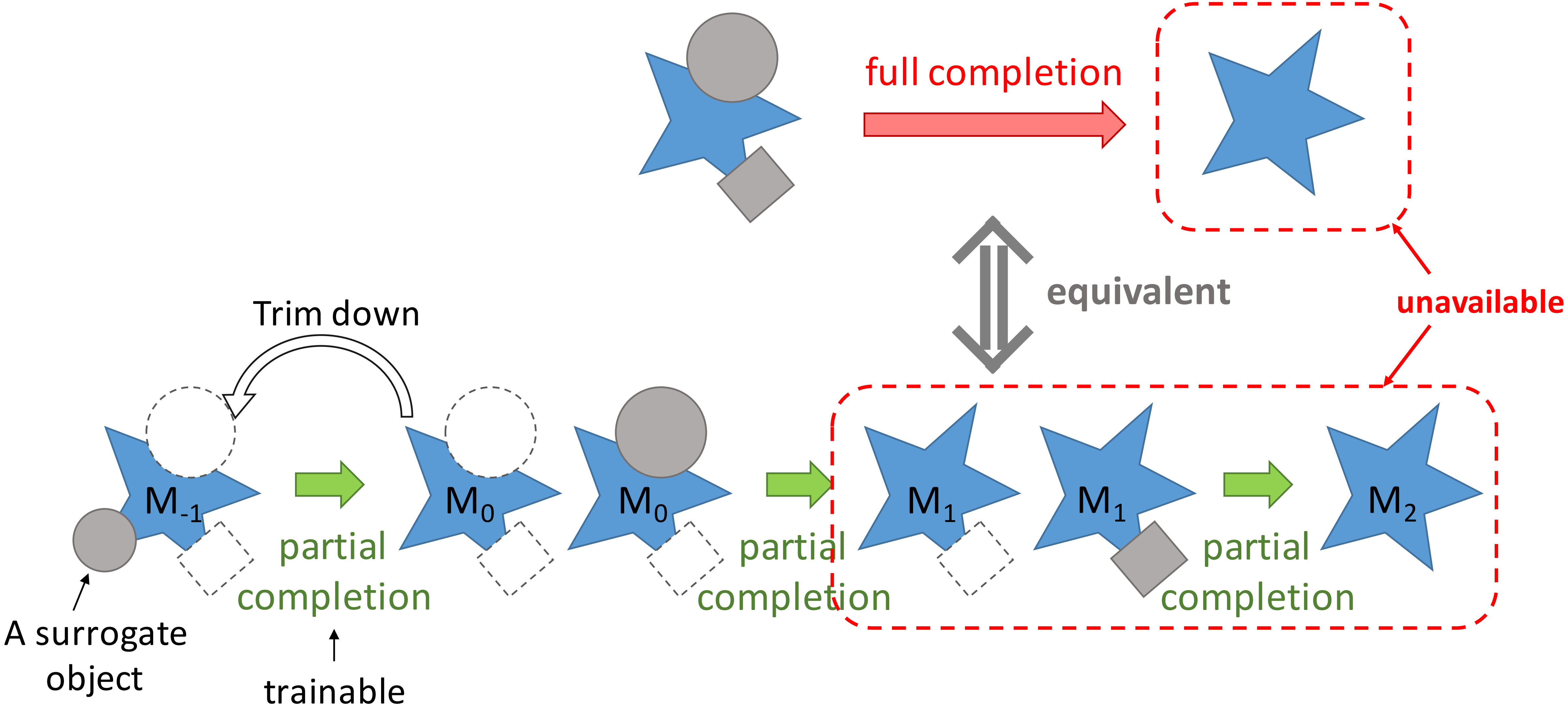
Full completion

Ground truth amodal masks as supervision



What if we do not have the ground truth?

Partial Completion



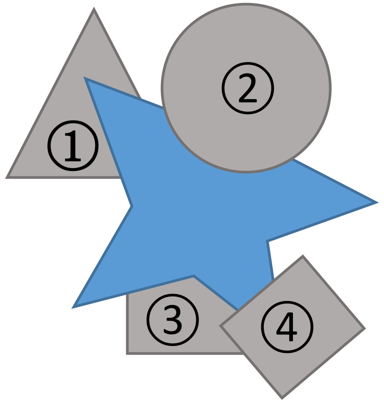
To Do

✓ Partial completion mechanism

- Complete part of an object occluded by a given occluder, without amodal annotations.

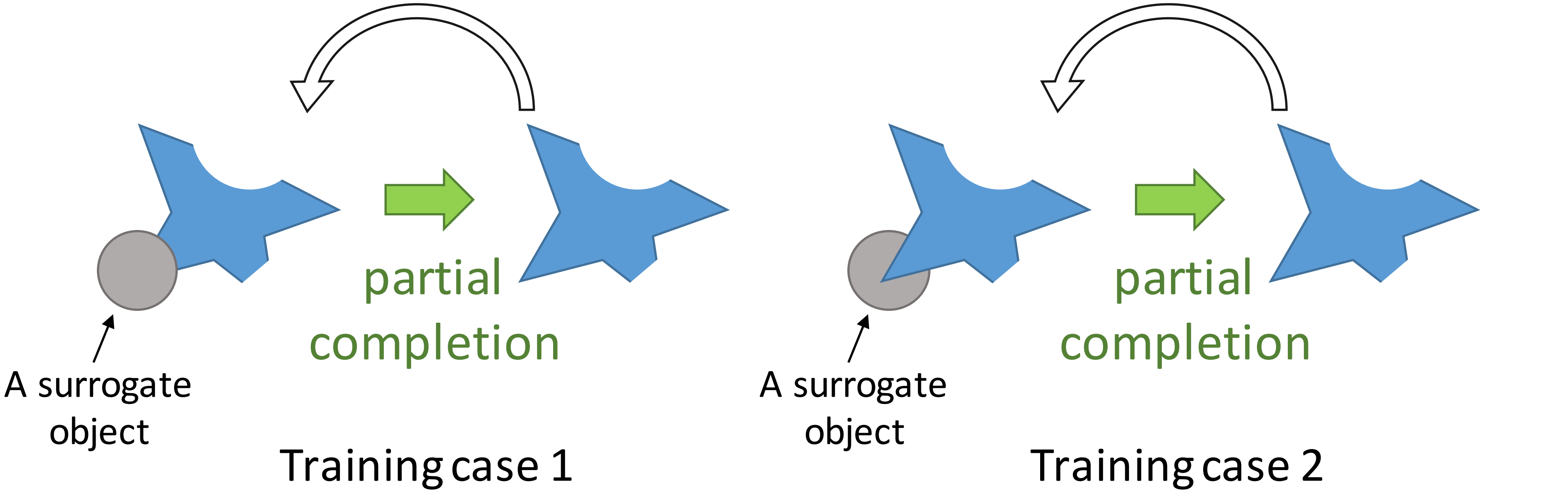
? Ordering recovery

- Predict the occluders of an object.



Among objects ①,②,③,④, who are its occluders?

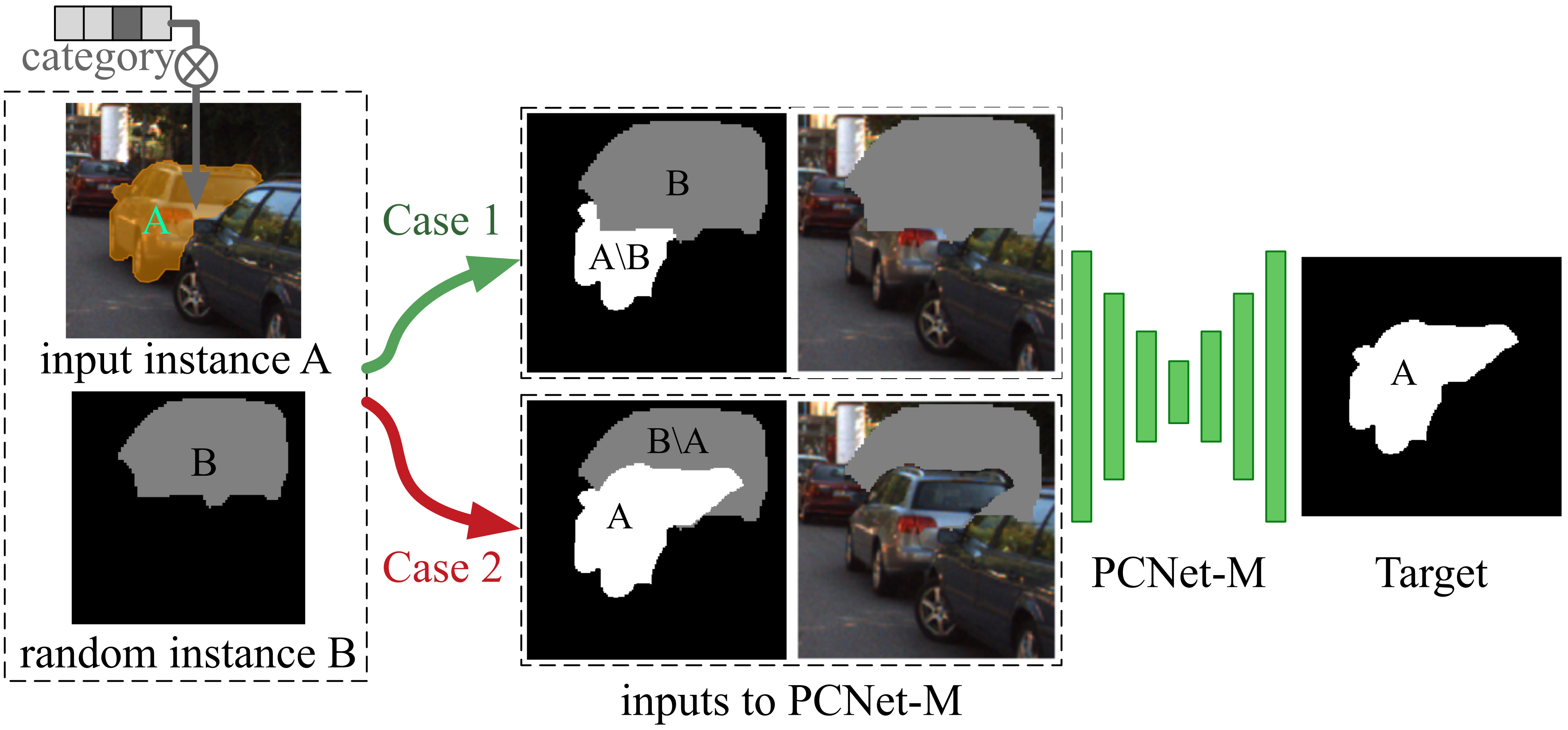
Partial Completion Regularization



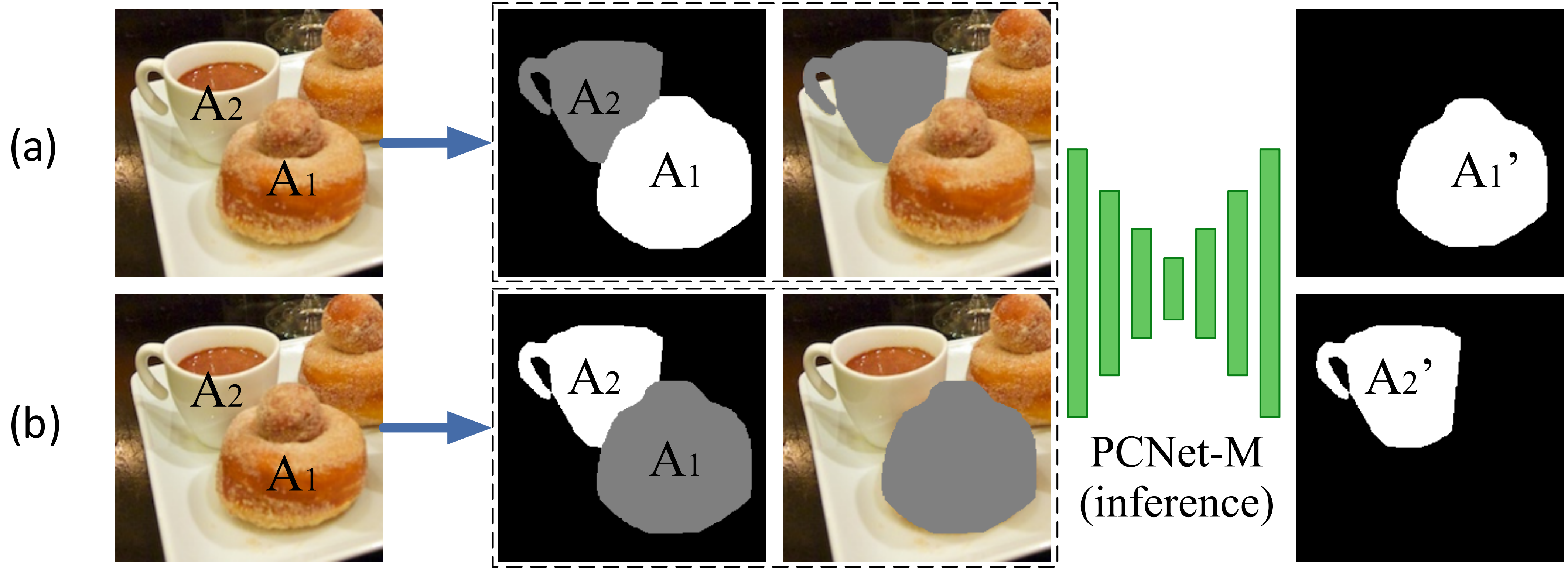
Trained with case 1:
always encourages increment of pixels

Trained with case 1 & 2:
if the target object looks like to be occluded by the surrogate object:
complete it
else:
keep unmodified

Train Partial Completion Net-Mask (PCNet-M)



Dual-Completion for Ordering Recovery



(a) Regarding A1 as the target and A2 as the surrogate occluder, the incremental area of A1: $\Delta A_1' | A_2$
(b) Regarding A2 as the target and A1 as the surrogate occluder, the incremental area of A2: $\Delta A_2' | A_1$

Decision: $\Delta A_1' | A_2 < \Delta A_2' | A_1 \Rightarrow$ A1 is above A2

To Do

✓ Partial completion

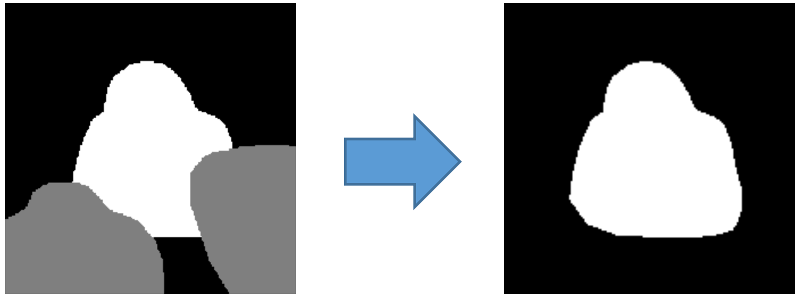
- Complete part of an object occluded by a given occluder, without amodal annotations.

✓ Ordering recovery

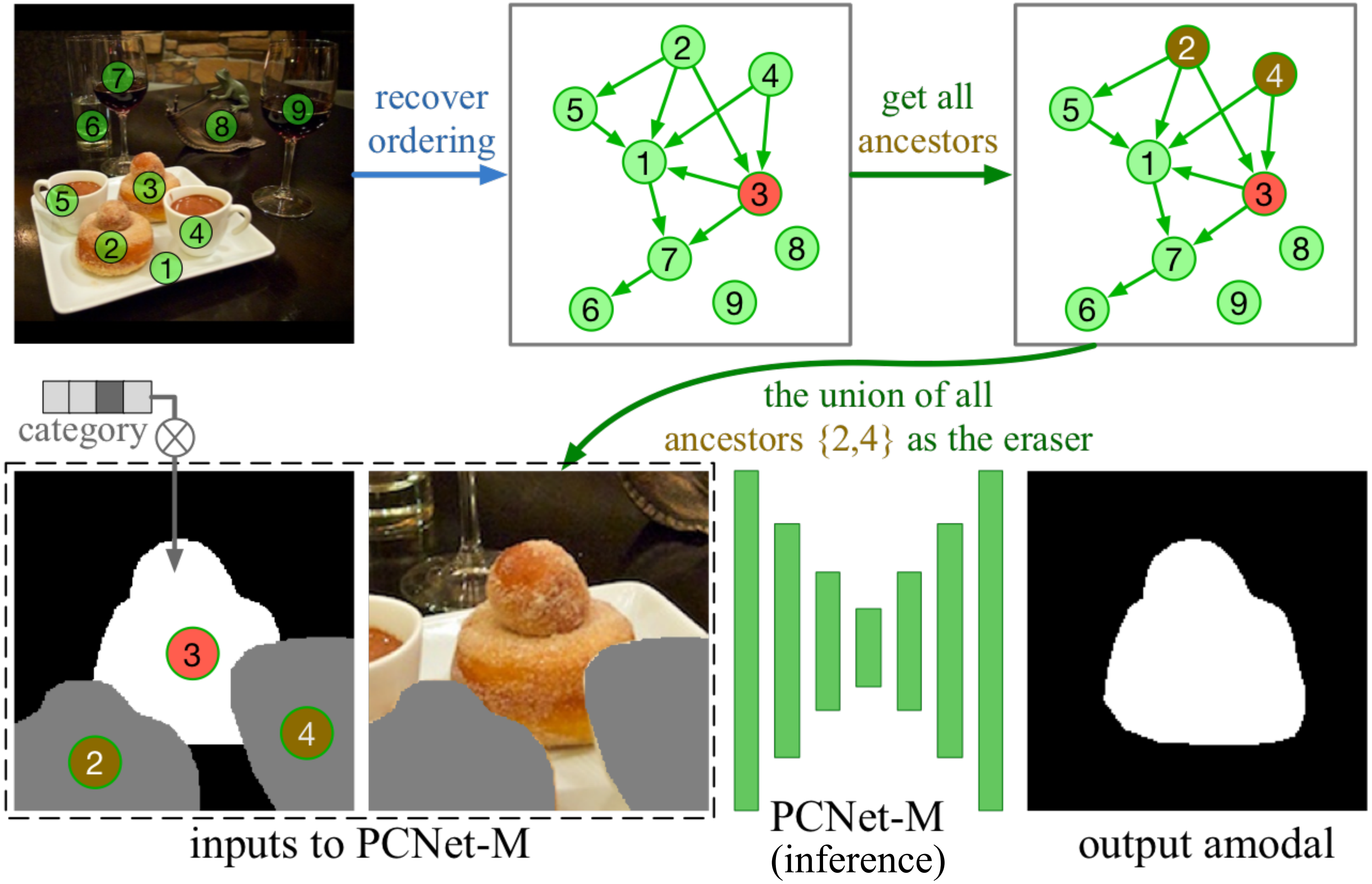
- Predict the occluders of an object.

? Amodal completion

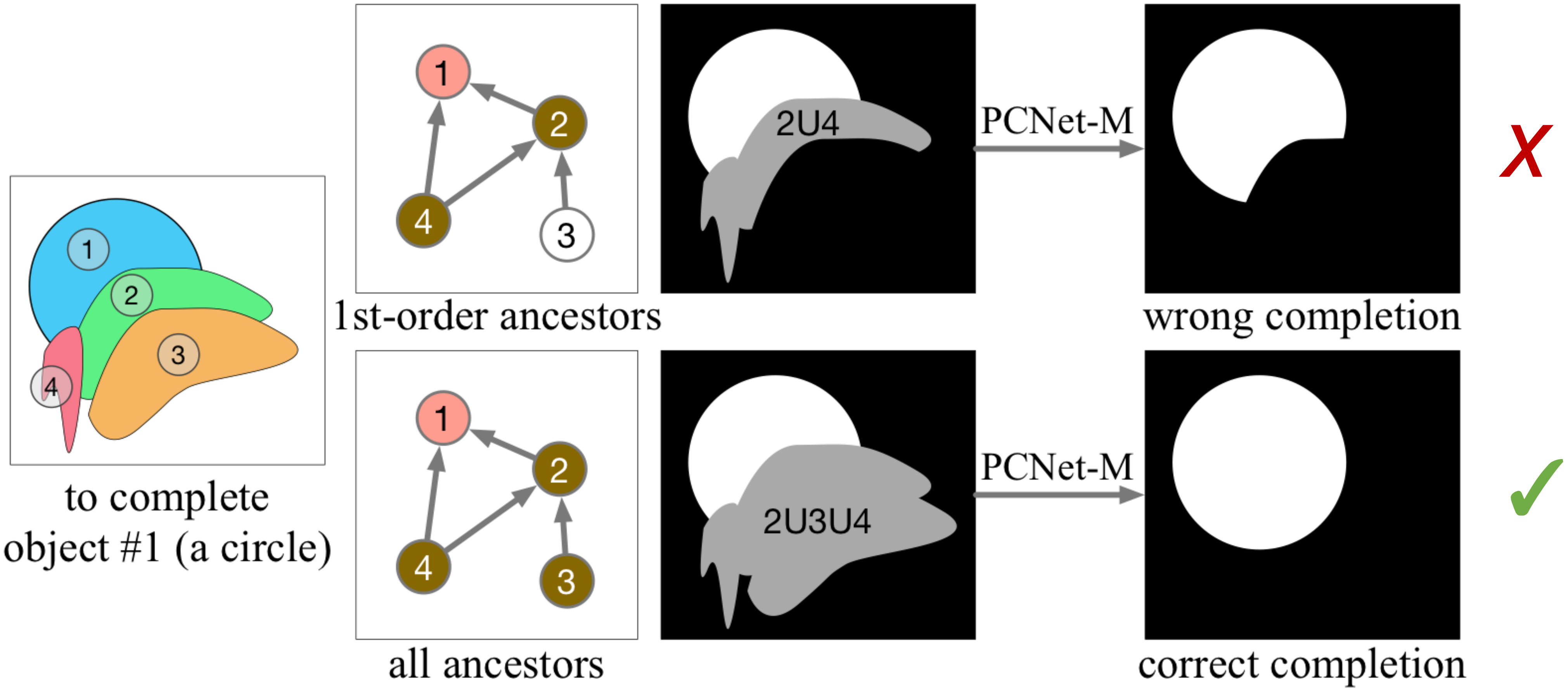
- Predict the amodal mask of each object given its occluders.



Ordering-Grounded Amodal Completion



Why All Ancestors?



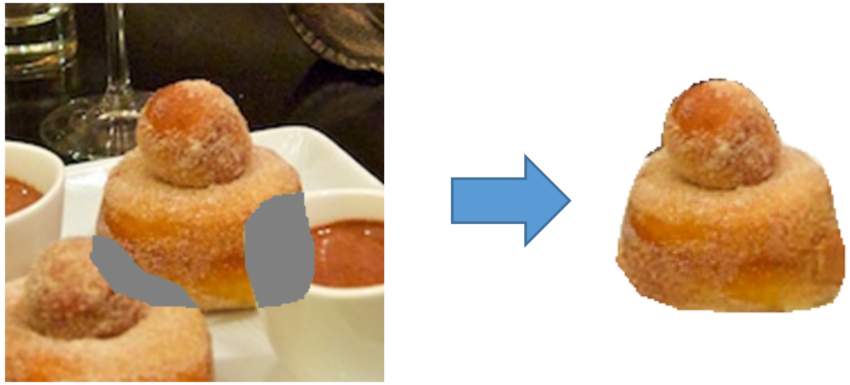
To Do

✓ Partial completion
- Complete part of an object occluded by a given occluder, without amodal annotations.

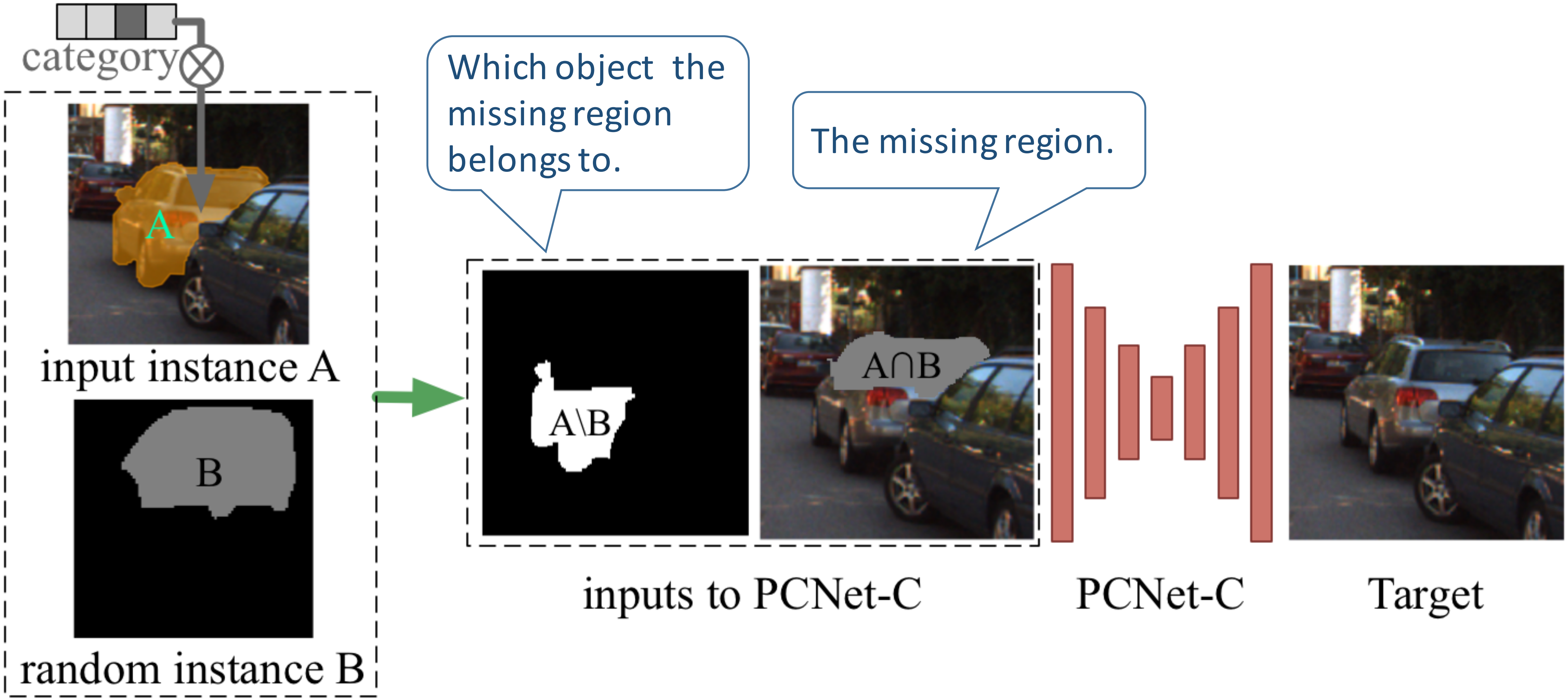
✓ Ordering recovery
- Predict the occluders of an object.

✓ Amodal completion
- Predict the amodal mask given occluders.

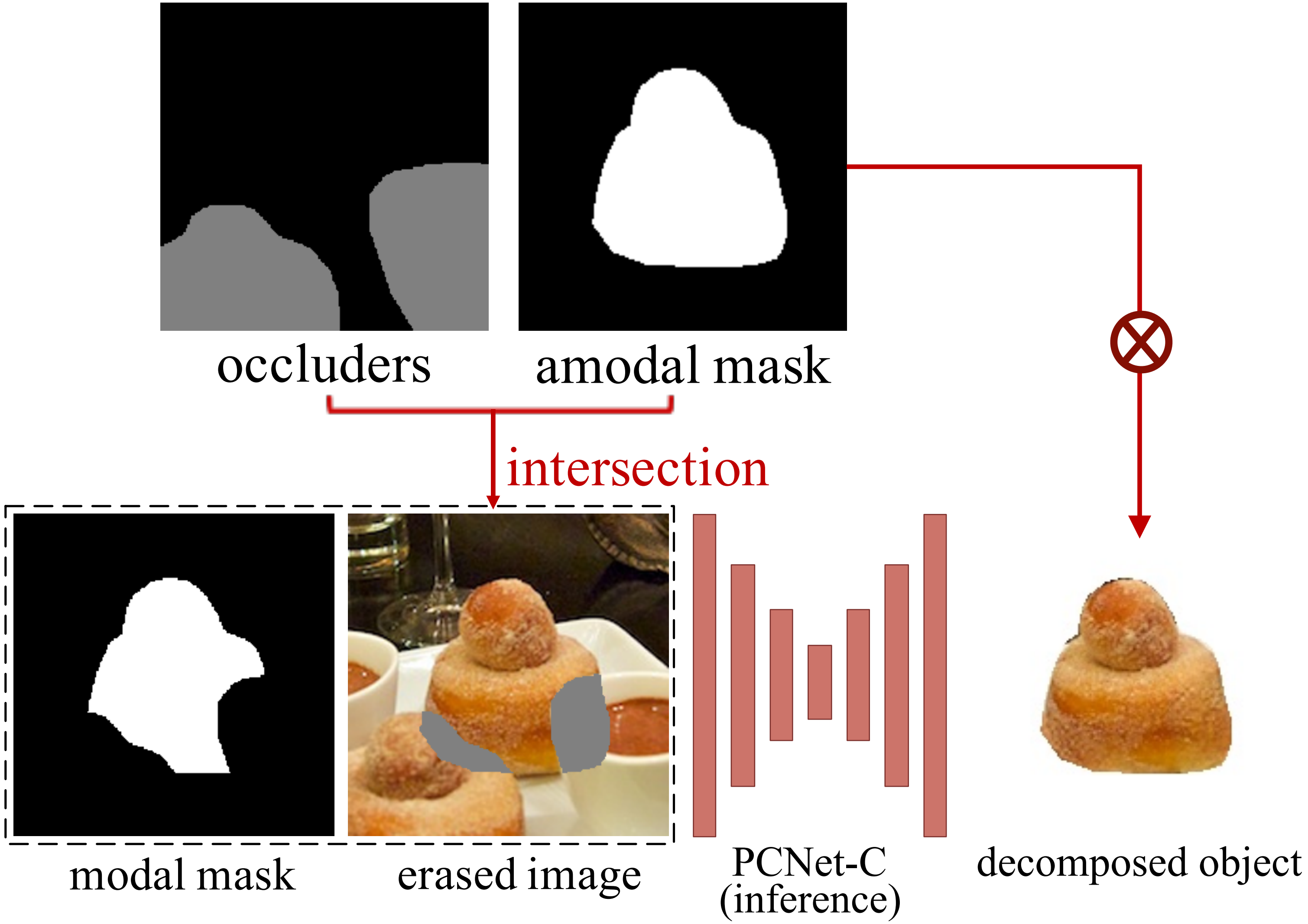
? Content completion
- Is it the same as image inpainting?



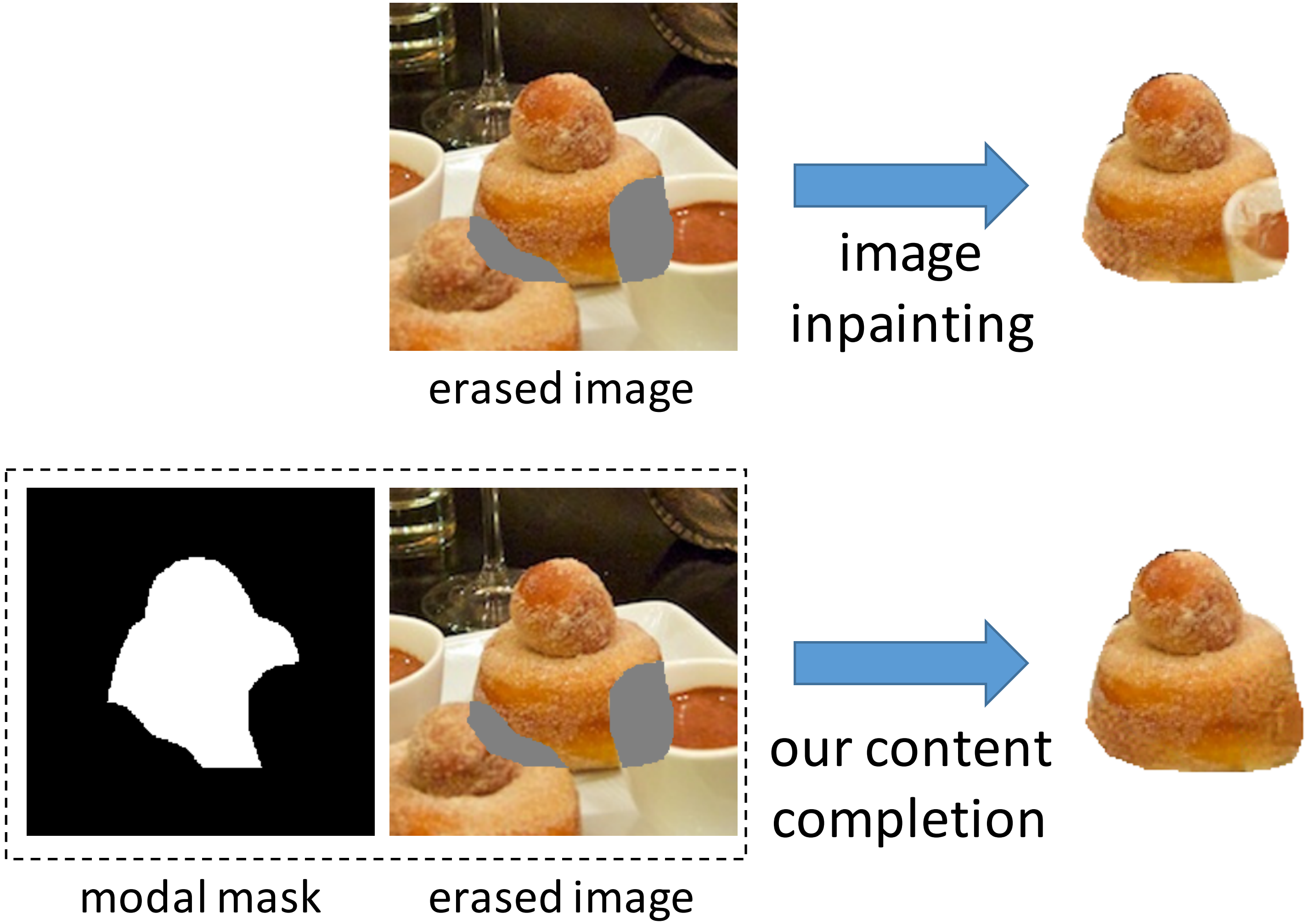
Train Partial Completion Net-Content (PCNet-C)



Amodal-Constrained Content Completion

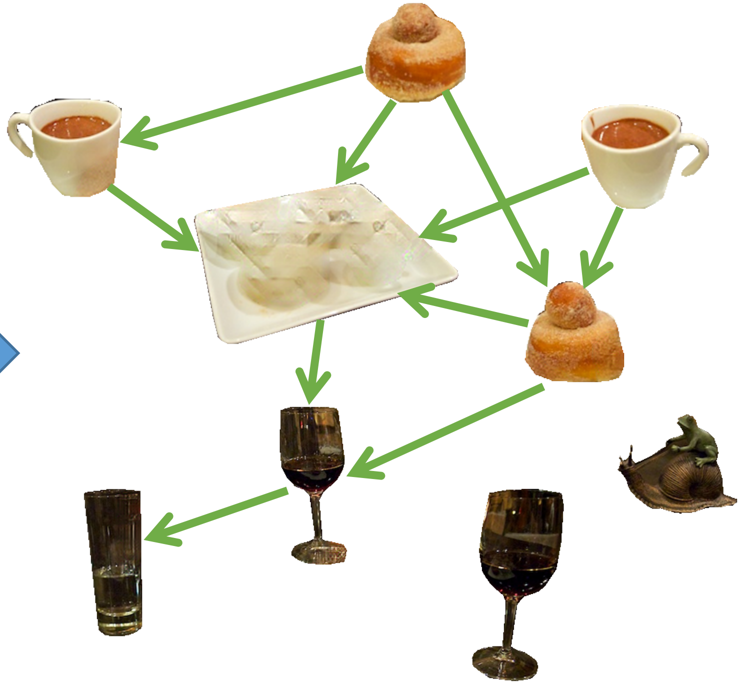


Compared to Image Inpainting



Scene De-occlusion

Real-world scene



Objects with invisible parts
+ ordering graph

Background

Todo list

✓ Partial completion

- Complete part of an object occluded by a given occluder, without amodal annotations.

← self-supervised training framework

✓ Ordering recovery

- Predict the occluders of an object.

✓ Amodal completion

- Predict the amodal mask given occluders.

✓ Content completion

- Slightly different from image inpainting.

progressive inference scheme

Evaluations

Table 1: Ordering estimation on COCOA validation and KINS testing sets, reported with pair-wise accuracy on occluded instance pairs.

method	gt order (train)	COCOA	KINS
Supervised			
OrderNet^{M} [16]	✔	81.7	87.5
OrderNet^{M+I} [16]	✔	88.3	94.1
Unsupervised			
Area	✘	62.4	77.4
Y-axis	✘	58.7	81.9
Convex	✘	76.0	76.3
Ours	✘	**87.1**	**92.5**

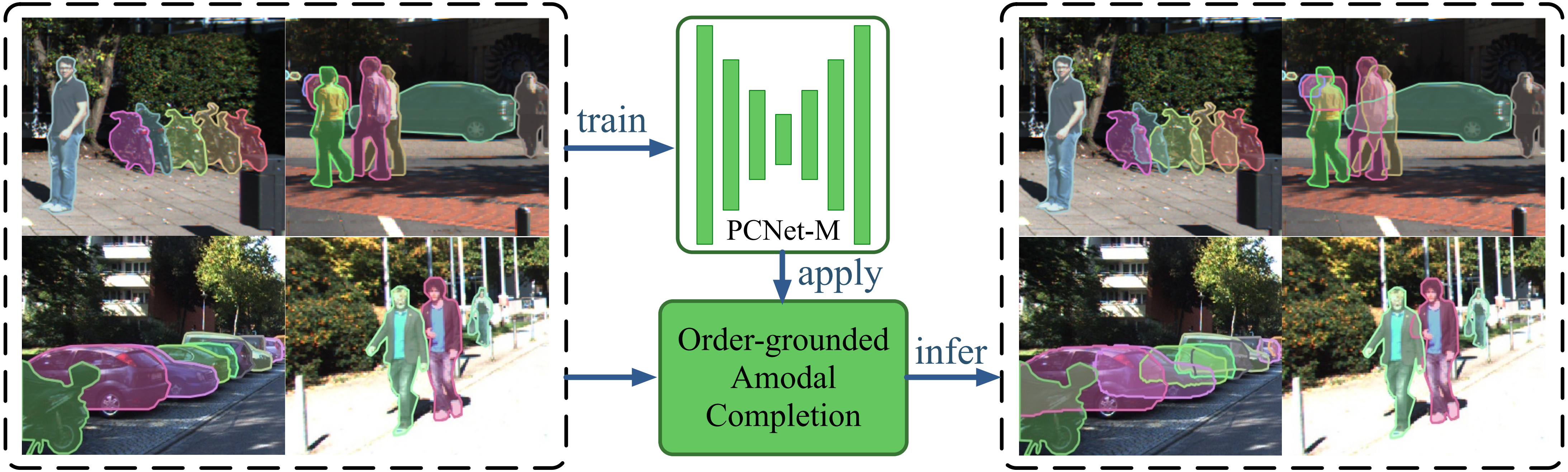
Ordering Recovery

Table 2: Amodal completion on COCOA validation and KINS testing sets, using ground truth modal masks.

method	amodal (train)	COCOA %mIoU	KINS %mIoU
Supervised	✔	82.53	94.81
Raw	✘	65.47	87.03
Convex^{R}	✘	74.43	90.75
Ours (NOG)	✘	76.91	93.42
Ours (OG)	✘	**81.35**	**94.76**

Amodal Completion

Modal Dataset to Amodal Dataset



dataset with
modal instances

dataset with
pseudo amodal instances

Pseudo Amodal Masks v.s. Manual Annotations

Table 4: Amodal instance segmentation on KINS testing set. Convex^{R} means using predicted order to refine the convex hull. In this experimental setting, all methods detect and segment instances from raw images. Hence, modal masks are not used in testing.

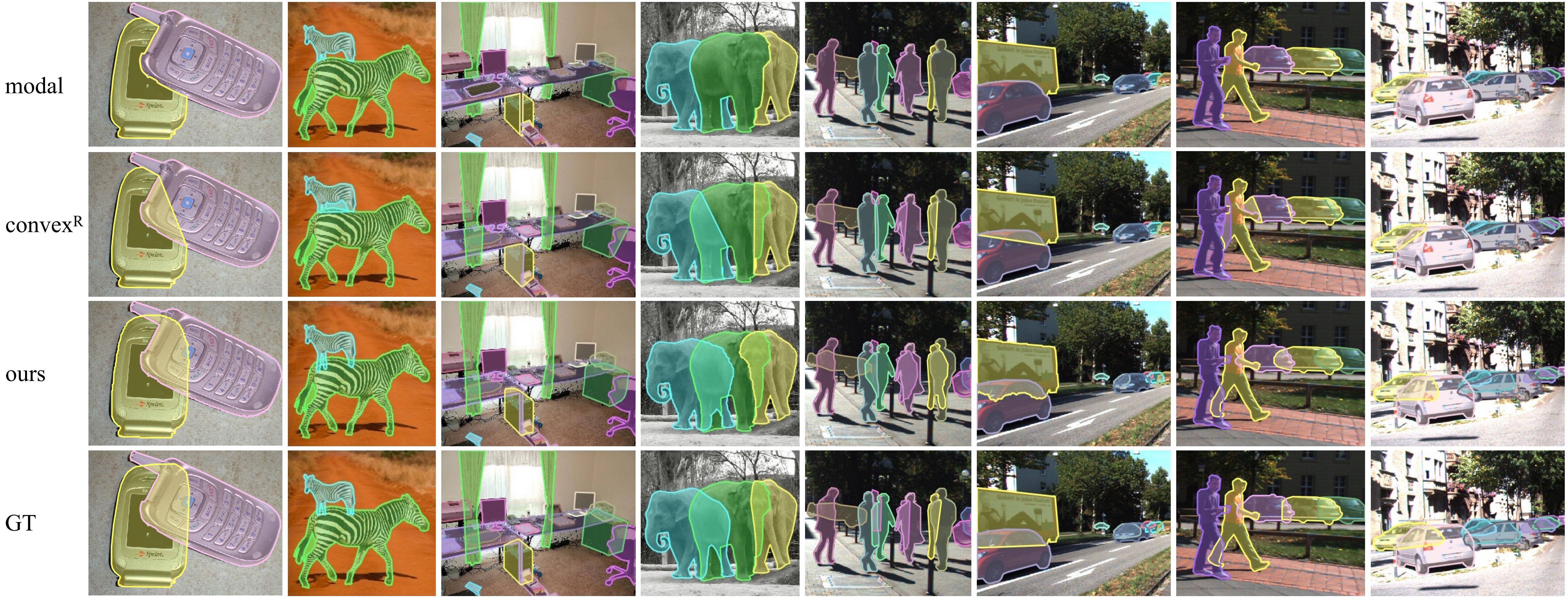
Ann. source	modal (train)	amodal (train)	%mAP
GT [17]	✘	✔	29.3
Raw	✔	✘	22.7
Convex	✔	✘	22.2
Convex^{R}	✔	✘	25.9
Ours	✔	✘	**29.3**

Using manual annotations

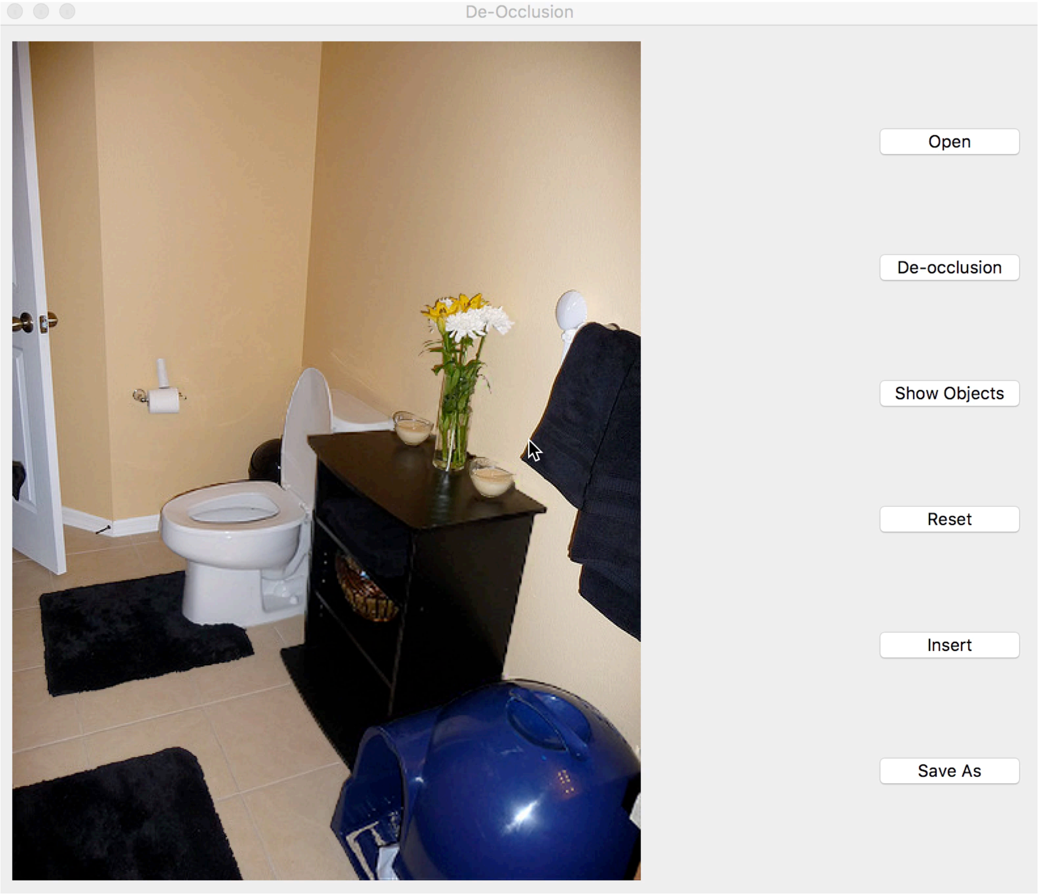
Using our pseudo annotations

Maybe in the future, we do not need to annotate amodal masks anymore!

Qualitative Amodal Completion Results



Demo: Scene Re-organization



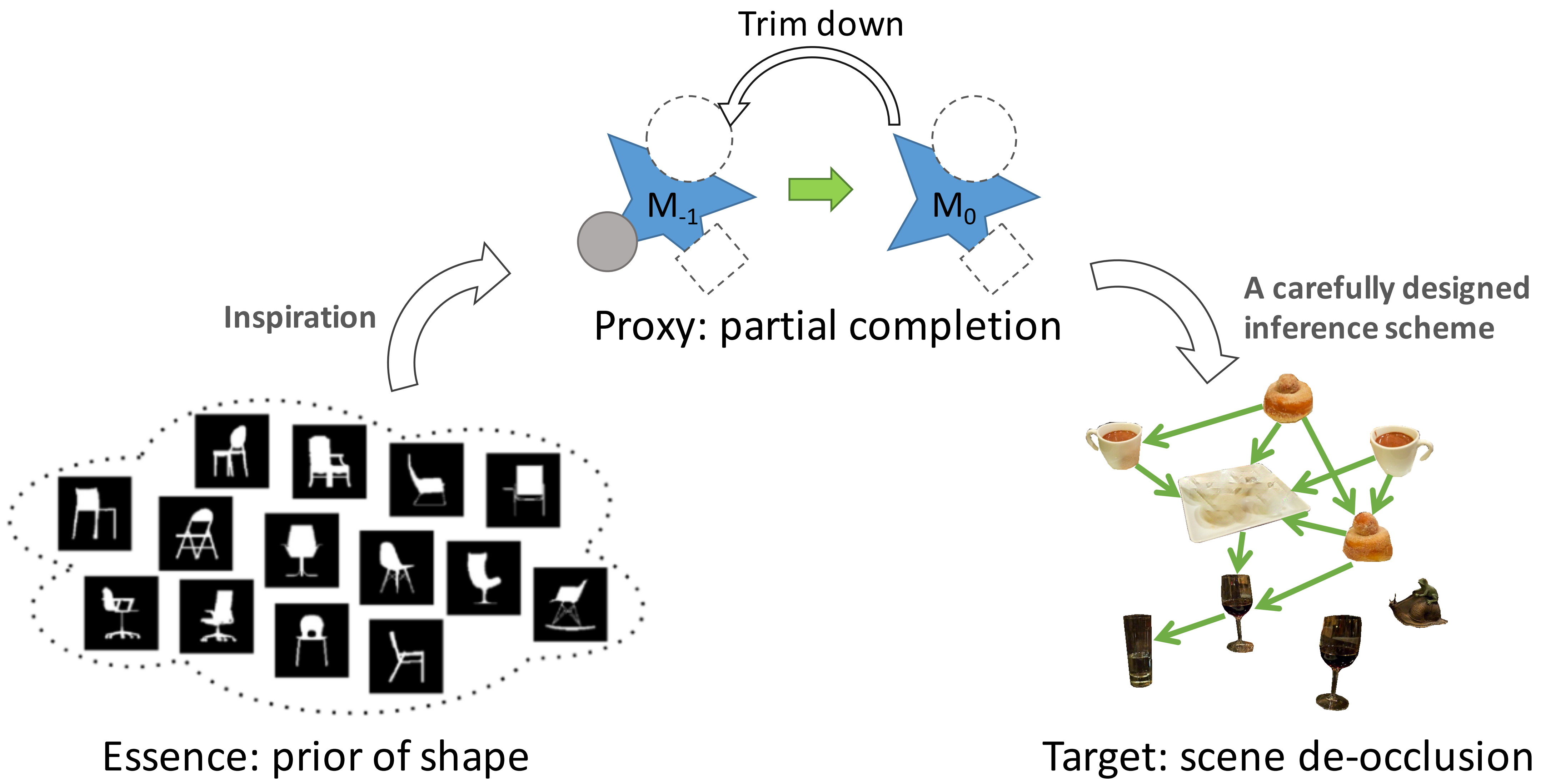
Watch the video here: https://xiaohangzhan.github.io/projects/deocclusion/

Future Directions with Self-Supervised Scene De-occlusion

- Data augmentation / re-composition for instance segmentation.
  - Previous: InstaBoost [ICCV'2019]
- Ordering prediction for mask fusion in panoptic segmentation.
- Occlusion-aware augmented reality.

No need for extra annotations!

What's the Intrinsic Methodology?



Messages to take away

1. Our world is low-entropy, working in rules.

2. The visual observations reflect the intrinsic rules.

3. Deep learning is skilled in processing visual observations.

Thank you!

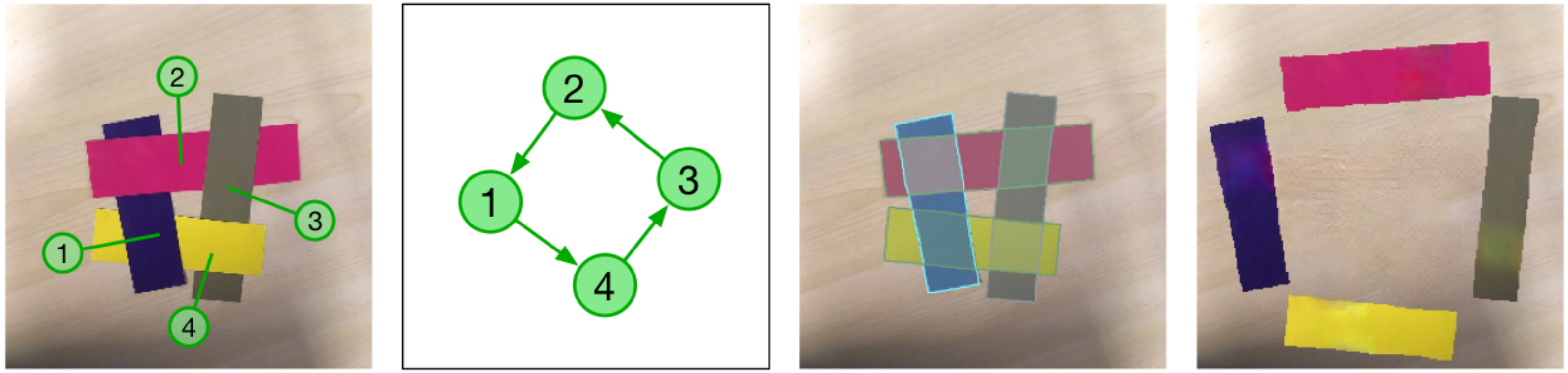
香港中文大學 The Chinese University of Hong Kong

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Discussions

Can it solve mutual occlusion? **No.**

Can it solve cyclic occlusion? **Yes.**



circularly occluded case | recovered ordering | amodal completion | content completion