# LEARNING CONTINUOUSLY WITHOUT FORGETTING FOR CONTINUAL SEMANTIC SEGMENTATION

CVPR 2021

Arthur Douillard, @ar_douillard
Yifu Chen
Arnaud Dapogny
Matthieu Cord

Machine Learning & Deep Learning for Information Access

# Classification

Image classification → one global label $D^t = \{(I, y), \dots\}, y \in \mathbb{N}$

Fish

Frog

Bird

# Segmentation

Semantic Segmentation → each pixel is labeled $D^t = \{(I, S), \dots\}, S \in \mathbb{N}^{W \times H}$

# Multiple labels

Semantic Segmentation → each pixel is labeled $D^t = \{(I, S), \dots\}, S \in \mathbb{N}^{W \times H}$

- Also for **multi-labels** classification $D^t = \{(I, Y), \dots\}, Y \in \mathbb{N}^C$
- Or **object detection**

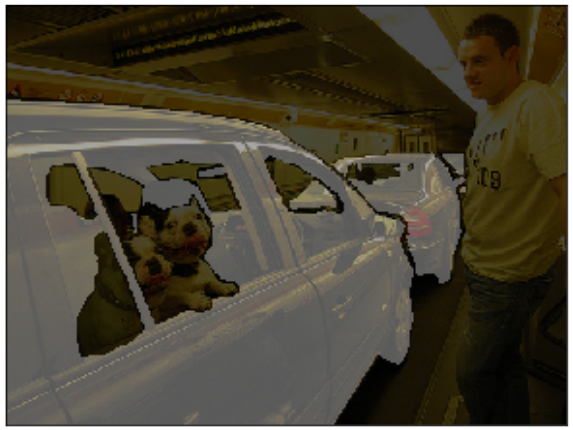
# What happens in Continual Learning?

# Background shift

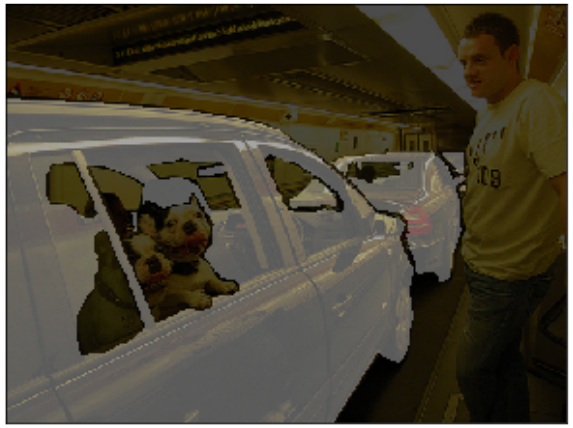
Learning **car → dog → person**

# Background shift

Task Ground-truth



Predicted Mask



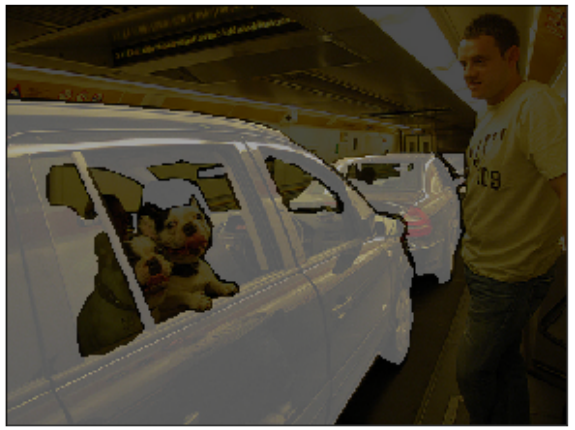
car

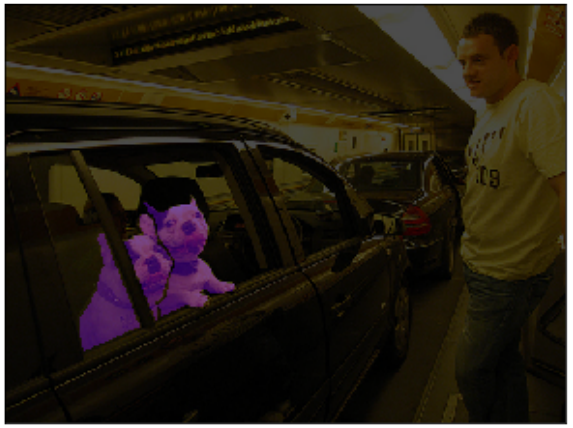
Task 1

[13]: Cermelli et al., Modeling the Background for Incremental in Semantic Segmentation, 2020

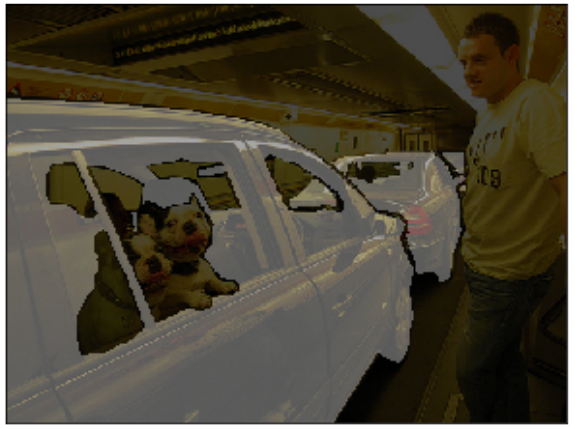
# Background shift

Task Ground-truth





Predicted Mask



car

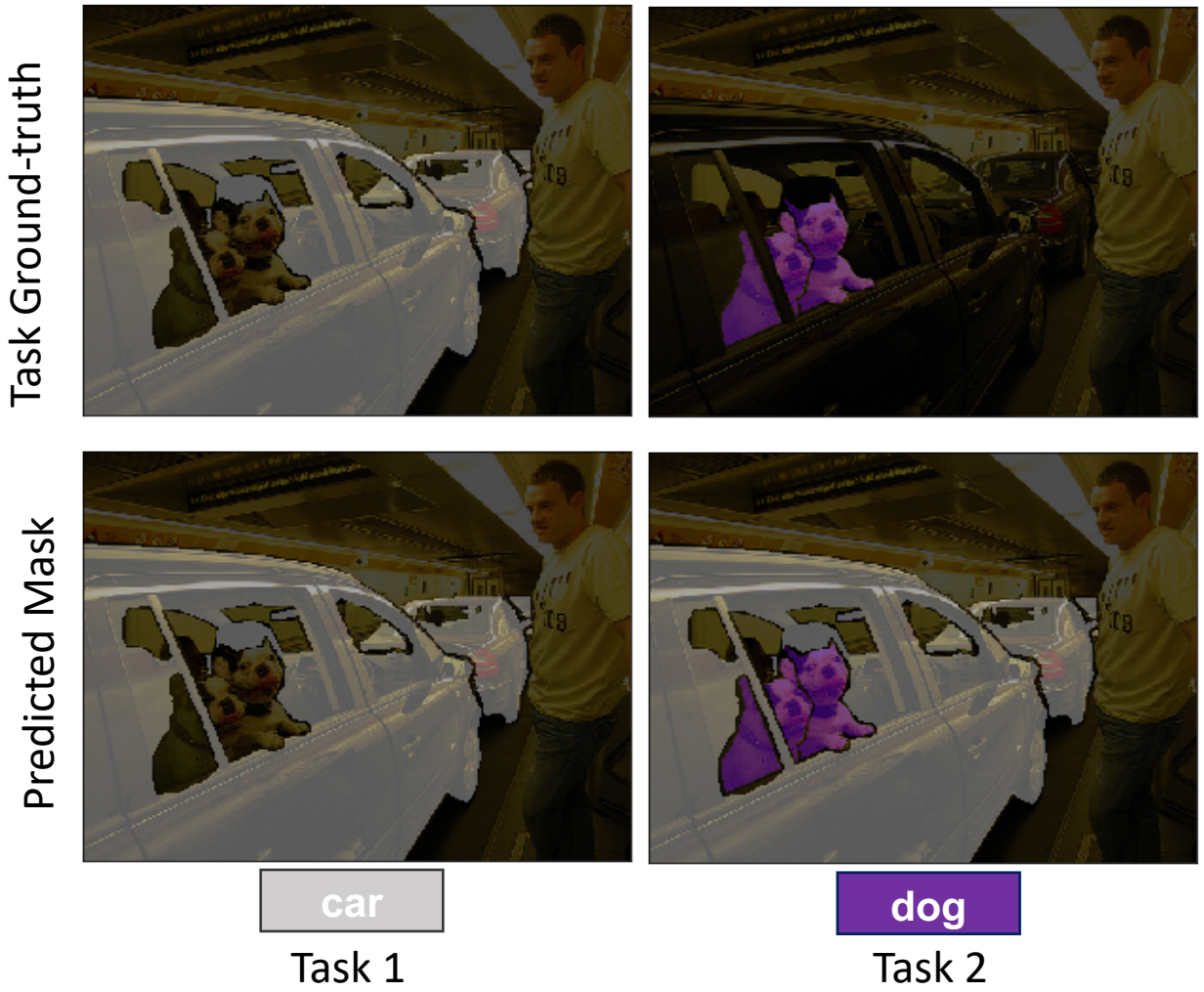
Task 1

dog

Task 2

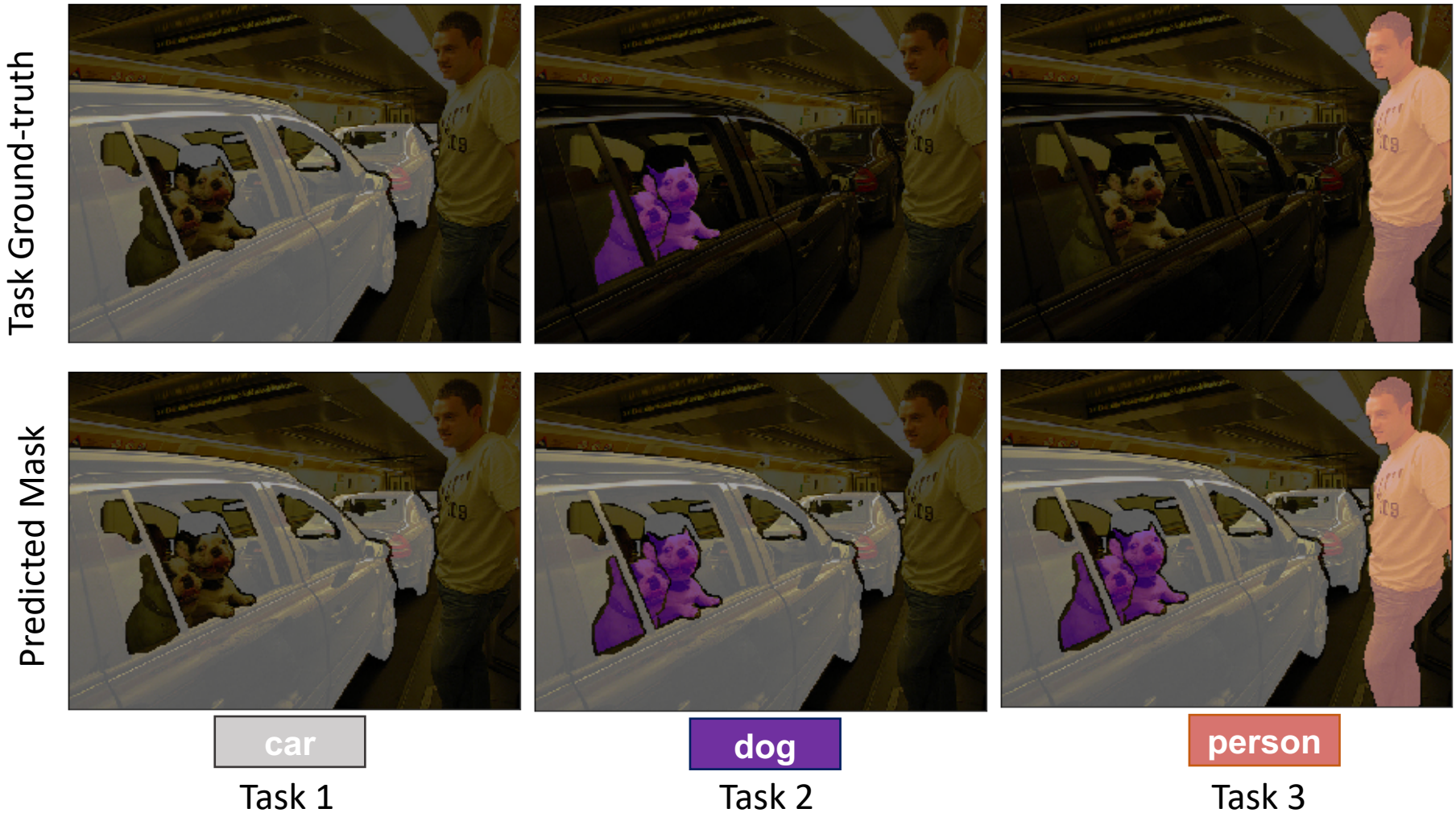
[13]: Cermelli et al., Modeling the Background for Incremental in Semantic Segmentation, 2020

# Background shift



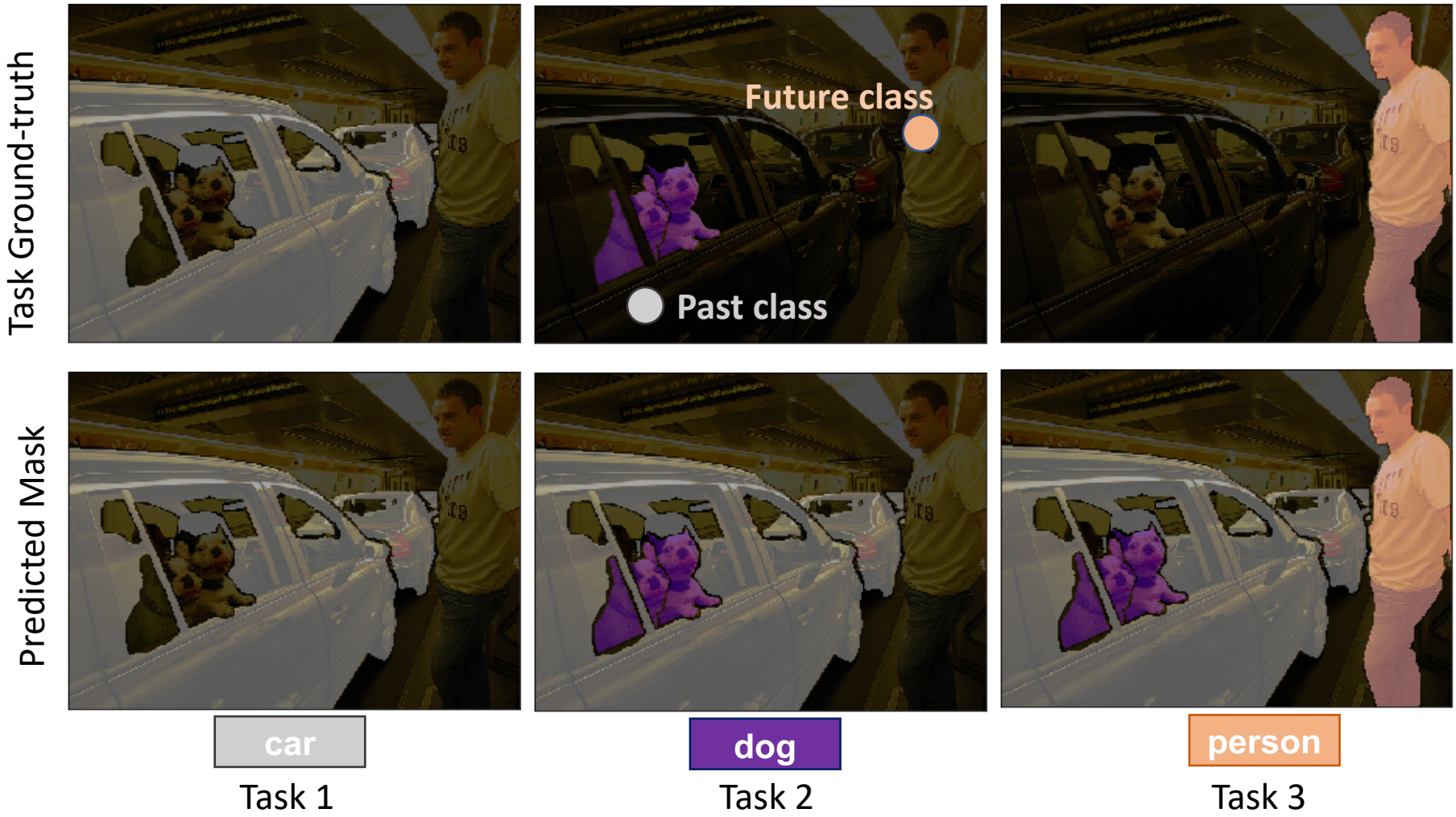

[13]: Cermelli et al., Modeling the Background for Incremental in Semantic Segmentation, 2020

# Background shift




[13]: Cermelli et al., Modeling the Background for Incremental in Semantic Segmentation, 2020

# Background shift

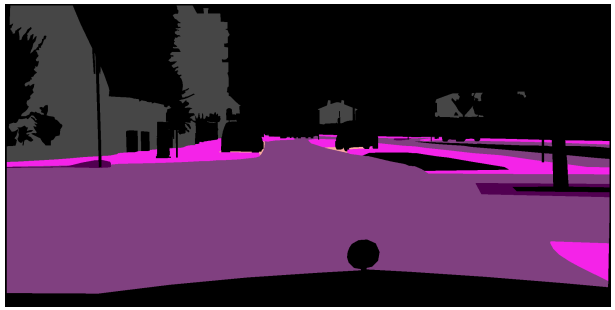


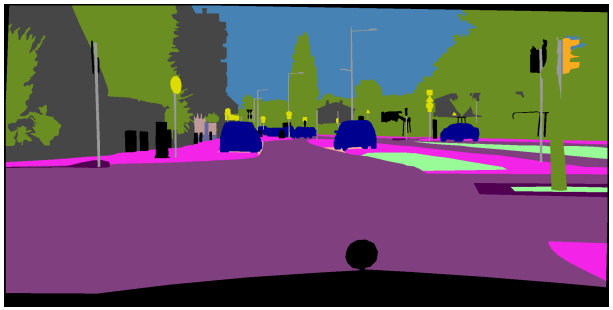

[13]: Cermelli et al., Modeling the Background for Incremental in Semantic Segmentation, 2020

# Realist setting?

The unsupervision is realist for real-life application.

Take **autonomous driving**,
- first learn well **basic classes** (road, sidewalk).
- Then **specialized classes** (car, tuktuk).



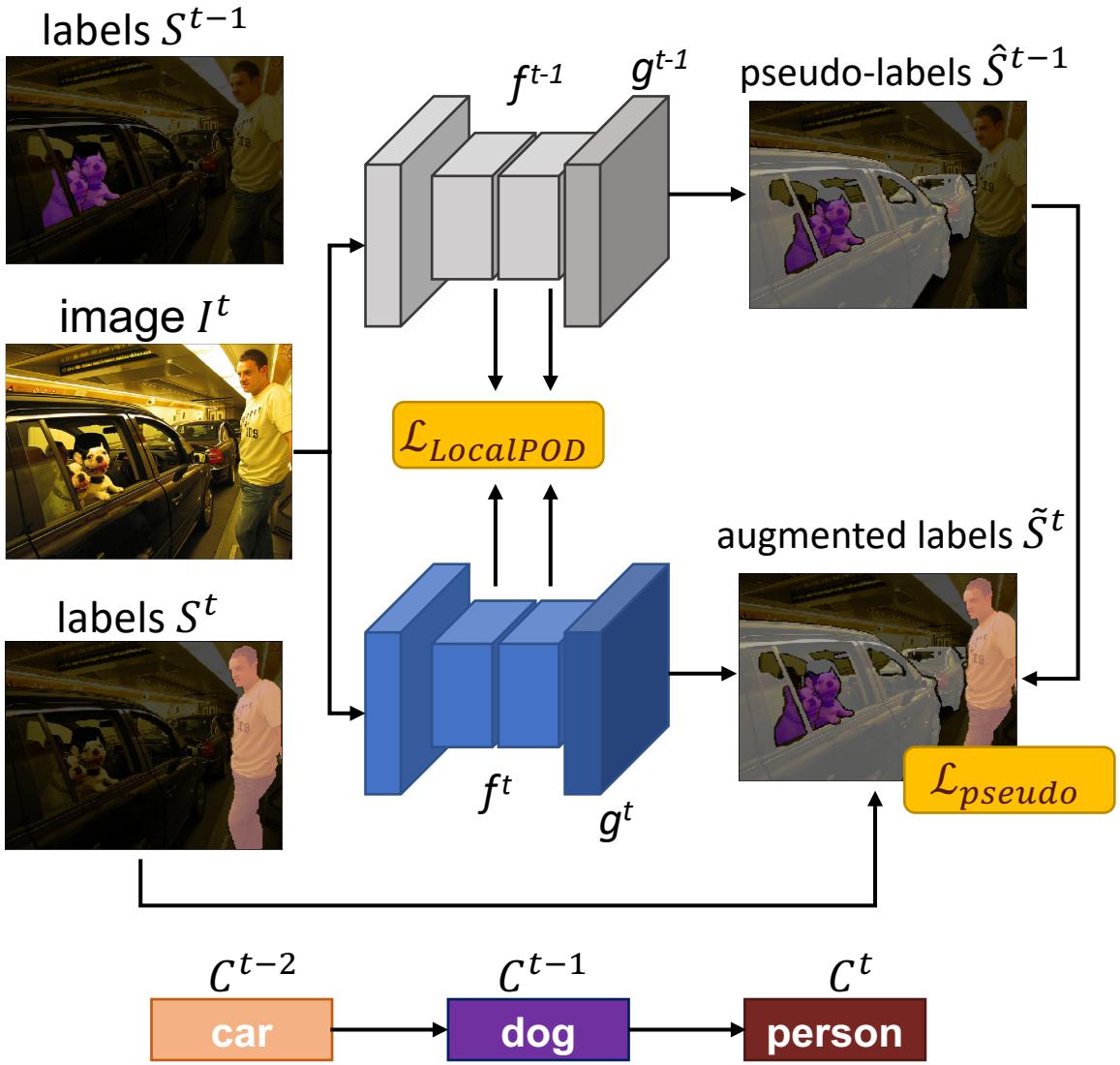


Less cost in human labeling.

Image classification: **~5s / image**
Semantic Segmentation: **5 – 30min / image**

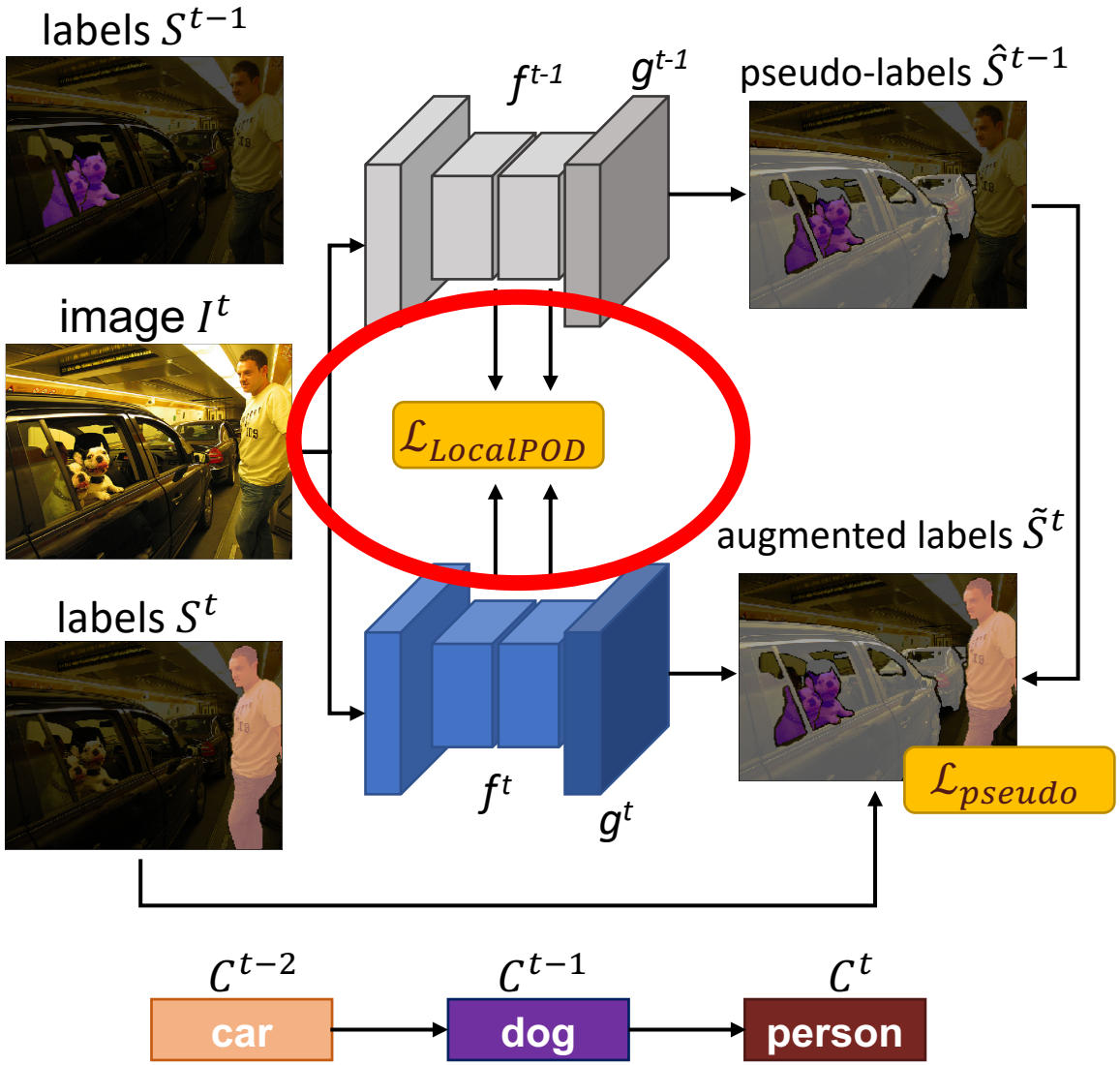
# PLOP Strategy

# PLOP Strategy




(a) PLOP strategy with pseudo-labeling and local POD

# PLOP Strategy




(a) PLOP strategy with pseudo-labeling and local POD

# POD (ECCV20)

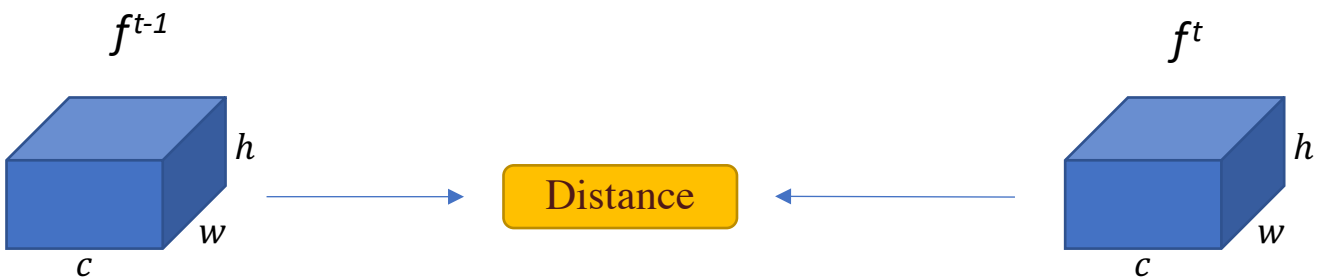
Naive distance between features doesn't work

$c \times w \times h$ constraints

→ too **rigid**

→ sensitive to outliers

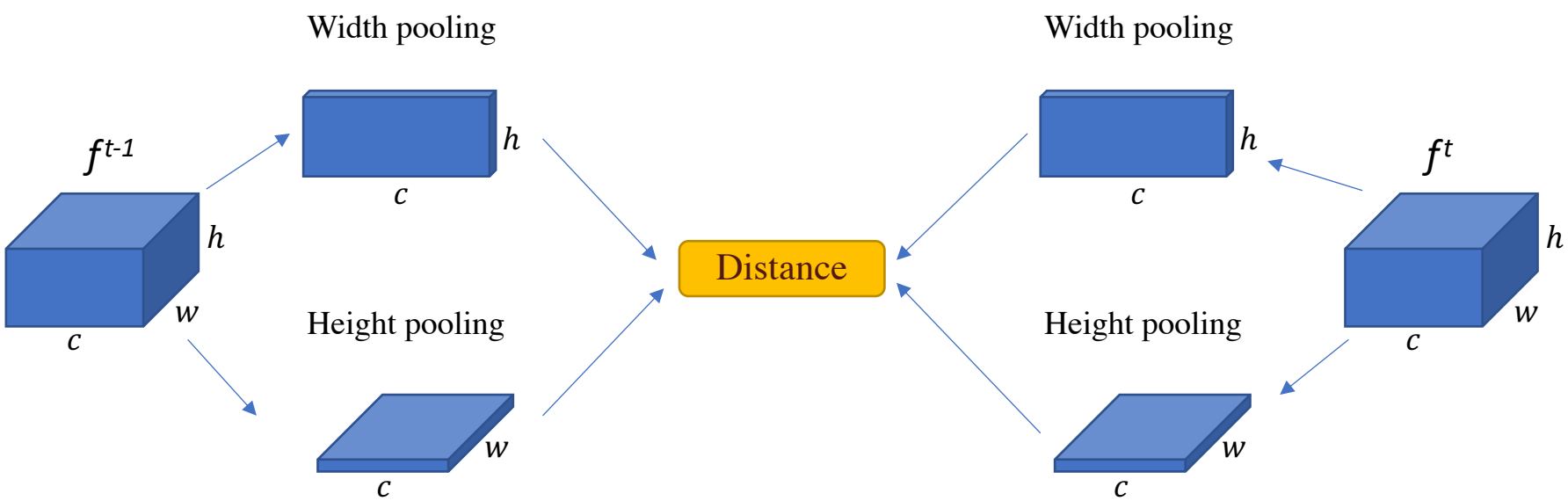
→ no spatial prior



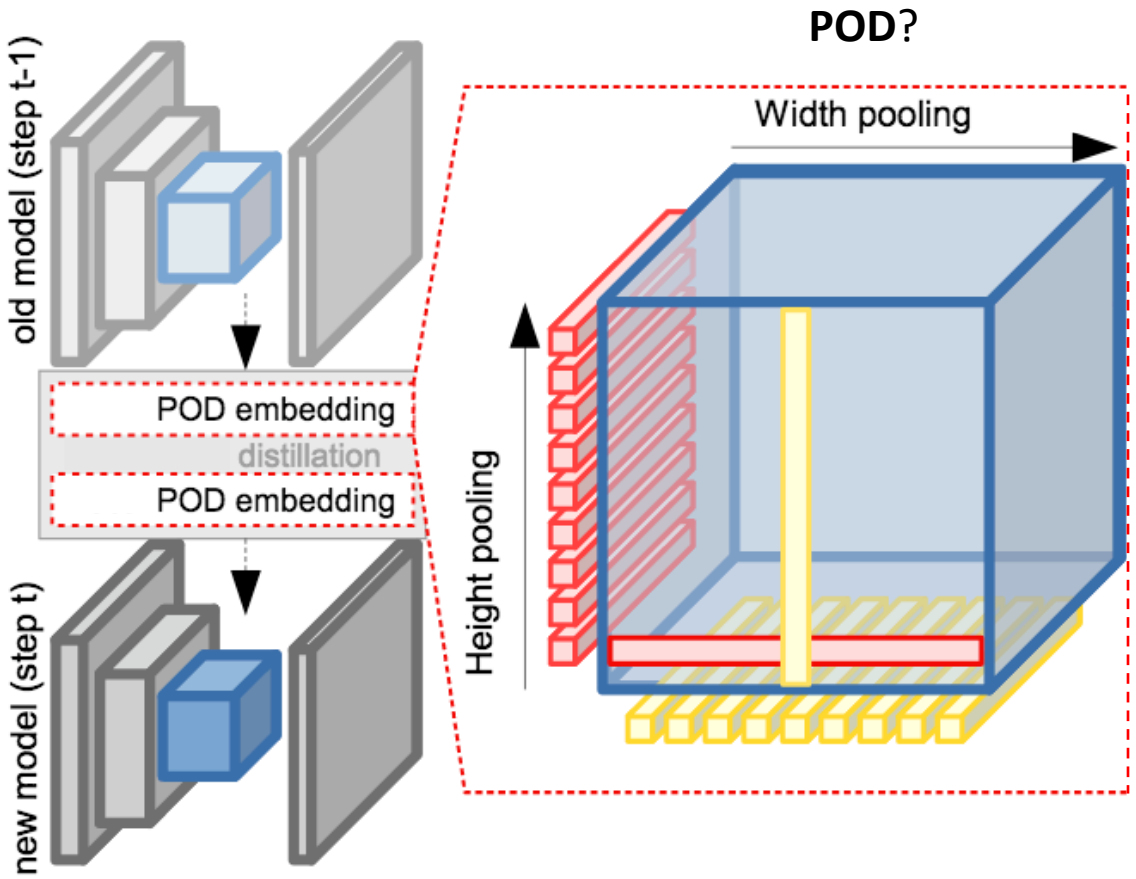
# POD (ECCV20)

Distance between **spatial statistics**

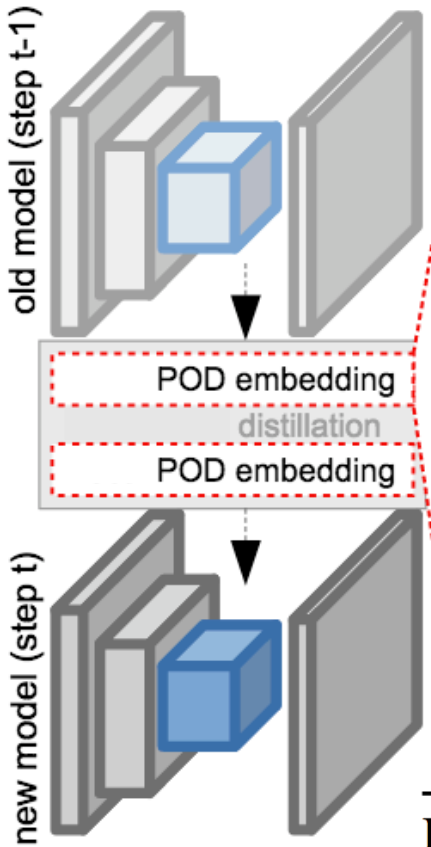
Balancing **rigidity** (not forgetting) and **plasticity** (learning)



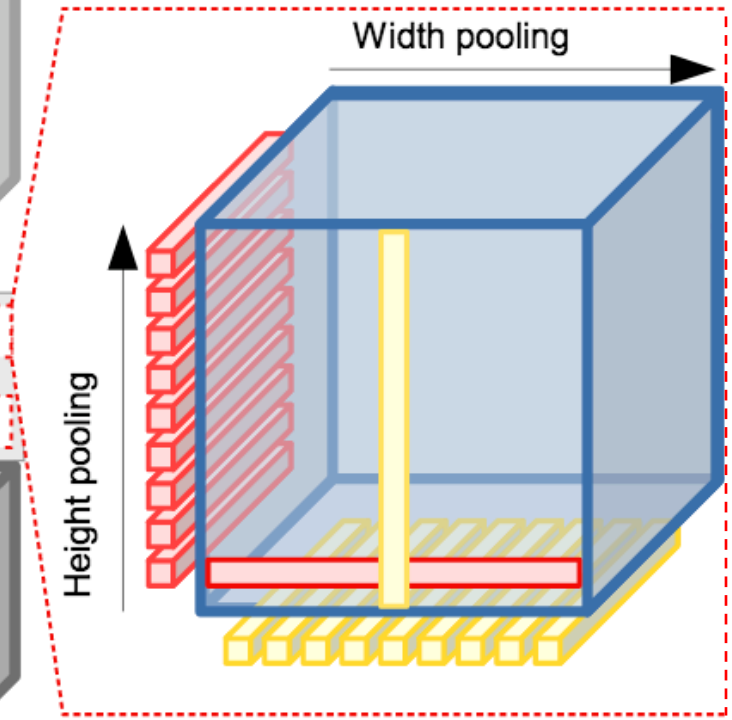
# Local POD


POD?
old model (step t-1)
POD embedding
distillation
POD embedding
new model (step t)
Width pooling
Height pooling

# Local POD

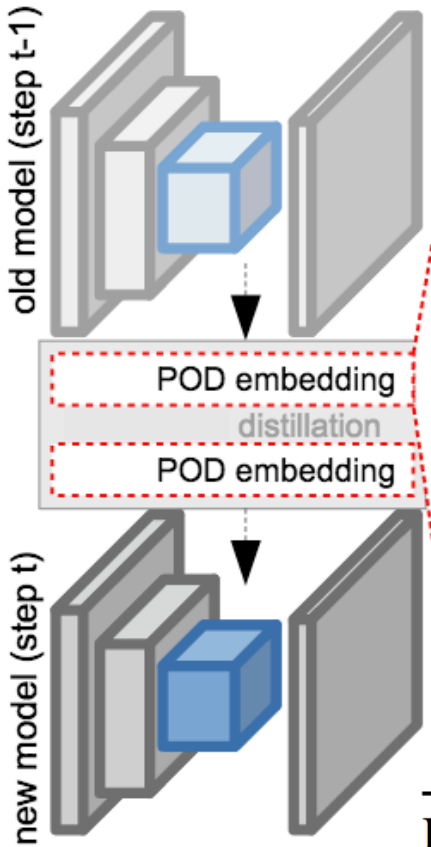



**POD?**

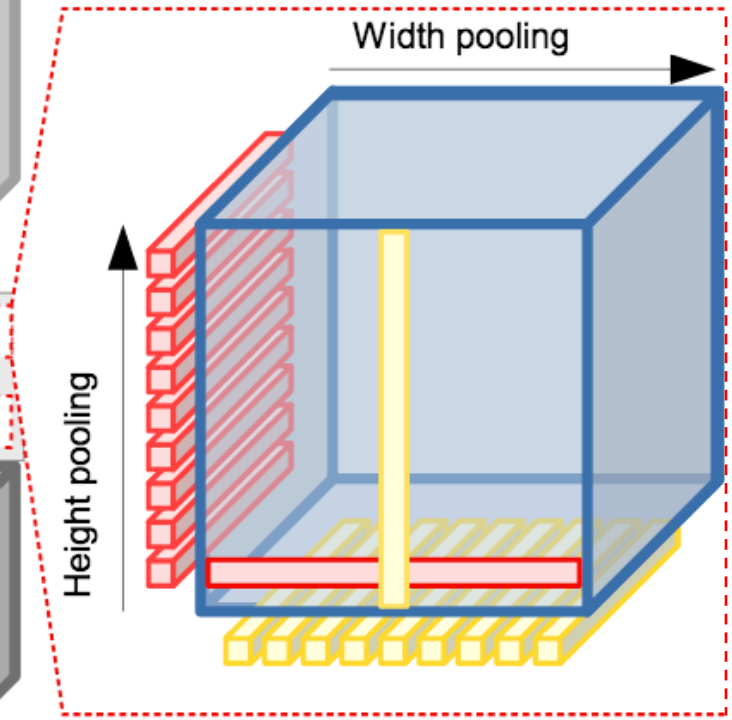



| Distillation loss | 0-15 | 16-20 | *all* | *avg* |
|---|---|---|---|---|
| Knowledge Distllation | 29.72 | 4.42 | 23.69 | 49.18 |
| UNKD | 34.85 | 5.26 | 27.80 | 46.39 |
| POD | 43.94 | 4.82 | 34.62 | 53.35 |

# Local POD




**POD?**

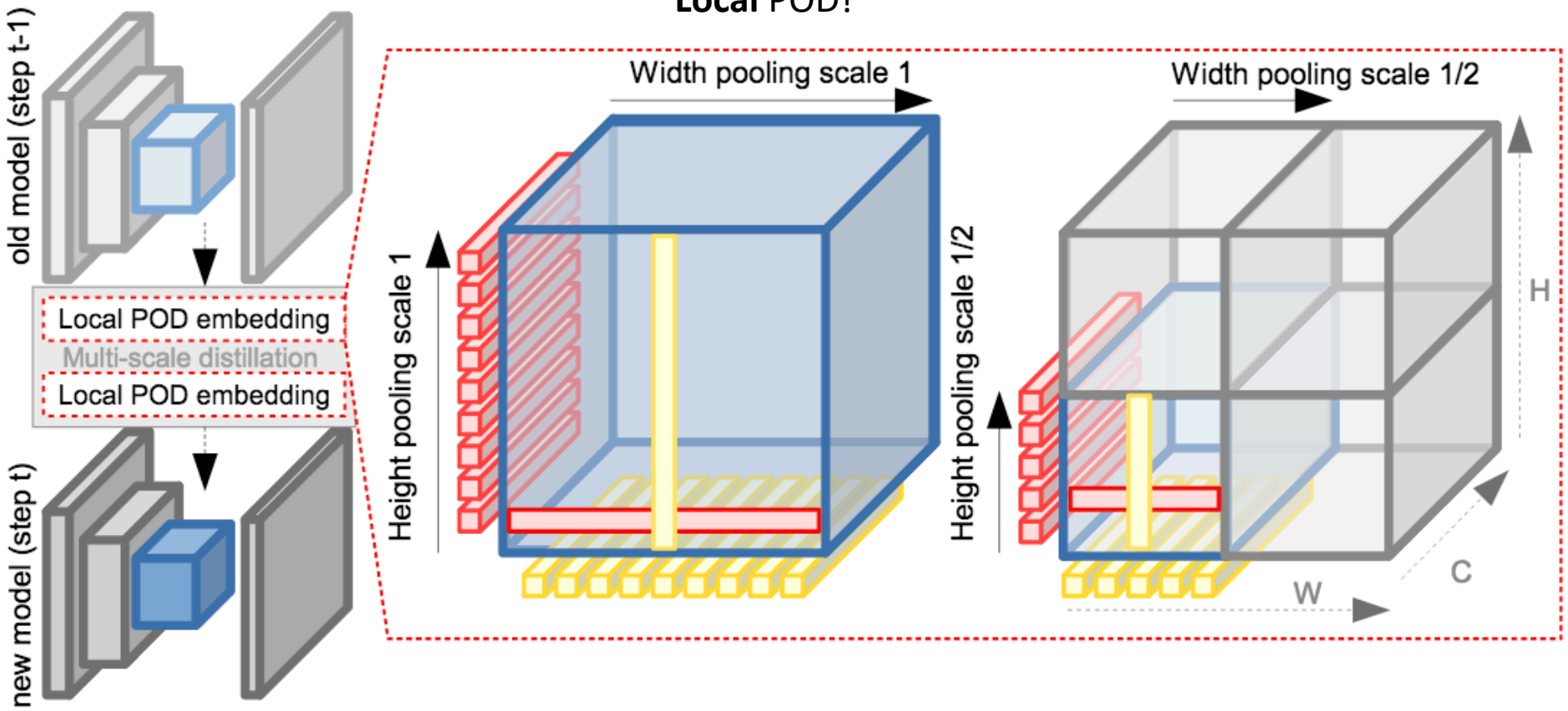



Segmentation
≠
Classification

| Distillation loss | 0-15 | 16-20 | *all* | *avg* |
|---|---|---|---|---|
| Knowledge Distllation | 29.72 | 4.42 | 23.69 | 49.18 |
| UNKD | 34.85 | 5.26 | 27.80 | 46.39 |
| POD | 43.94 | 4.82 | 34.62 | 53.35 |

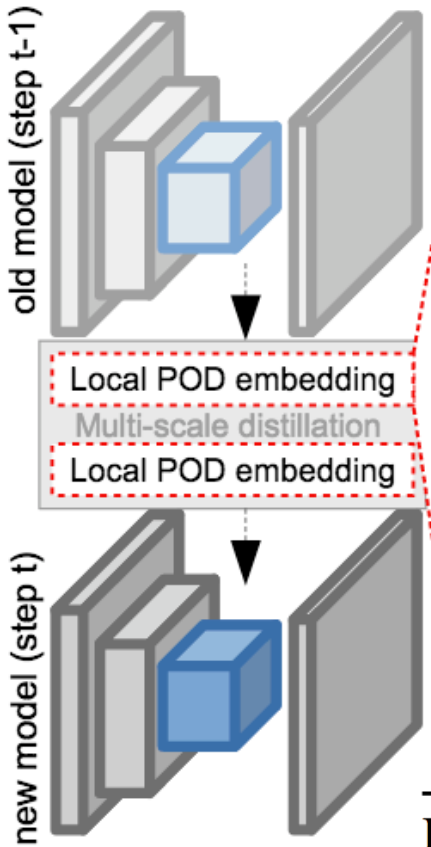
# Local POD

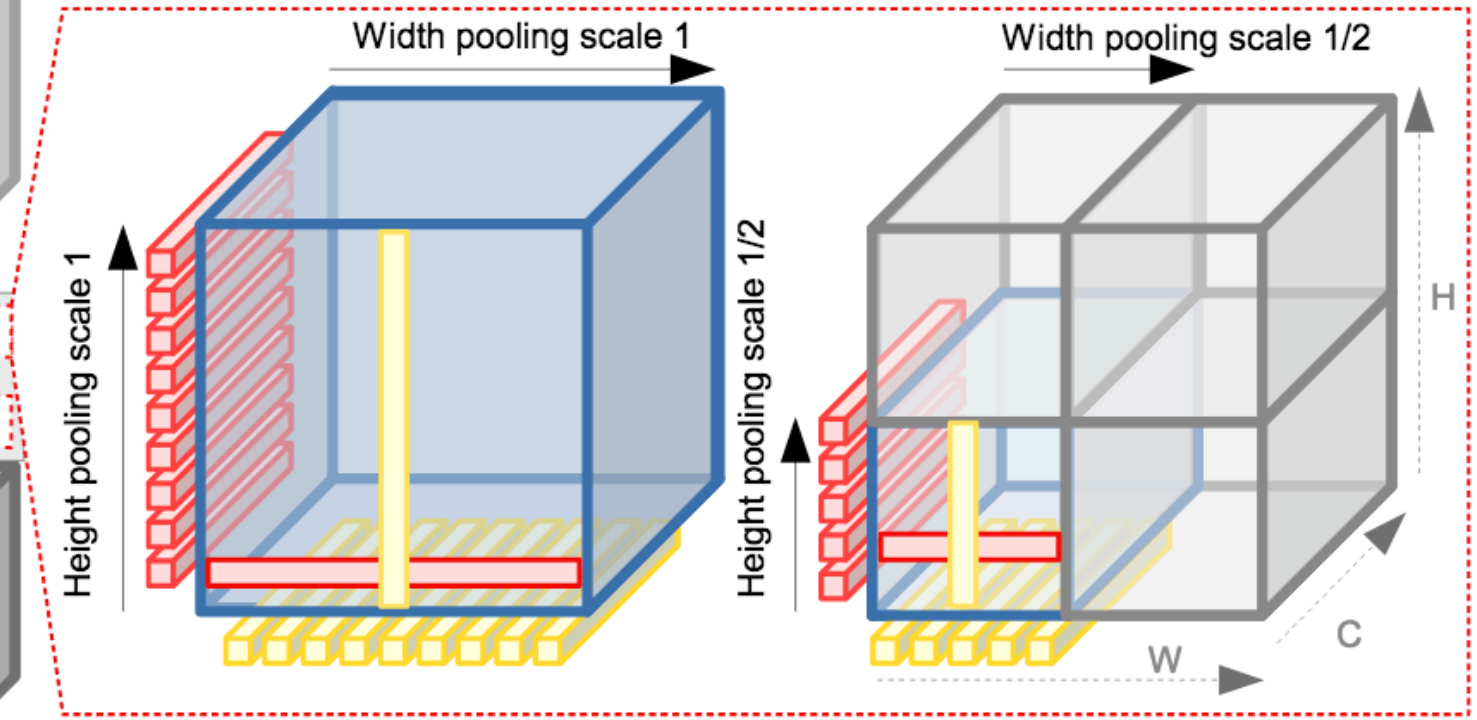
**Local** POD!



# Local POD

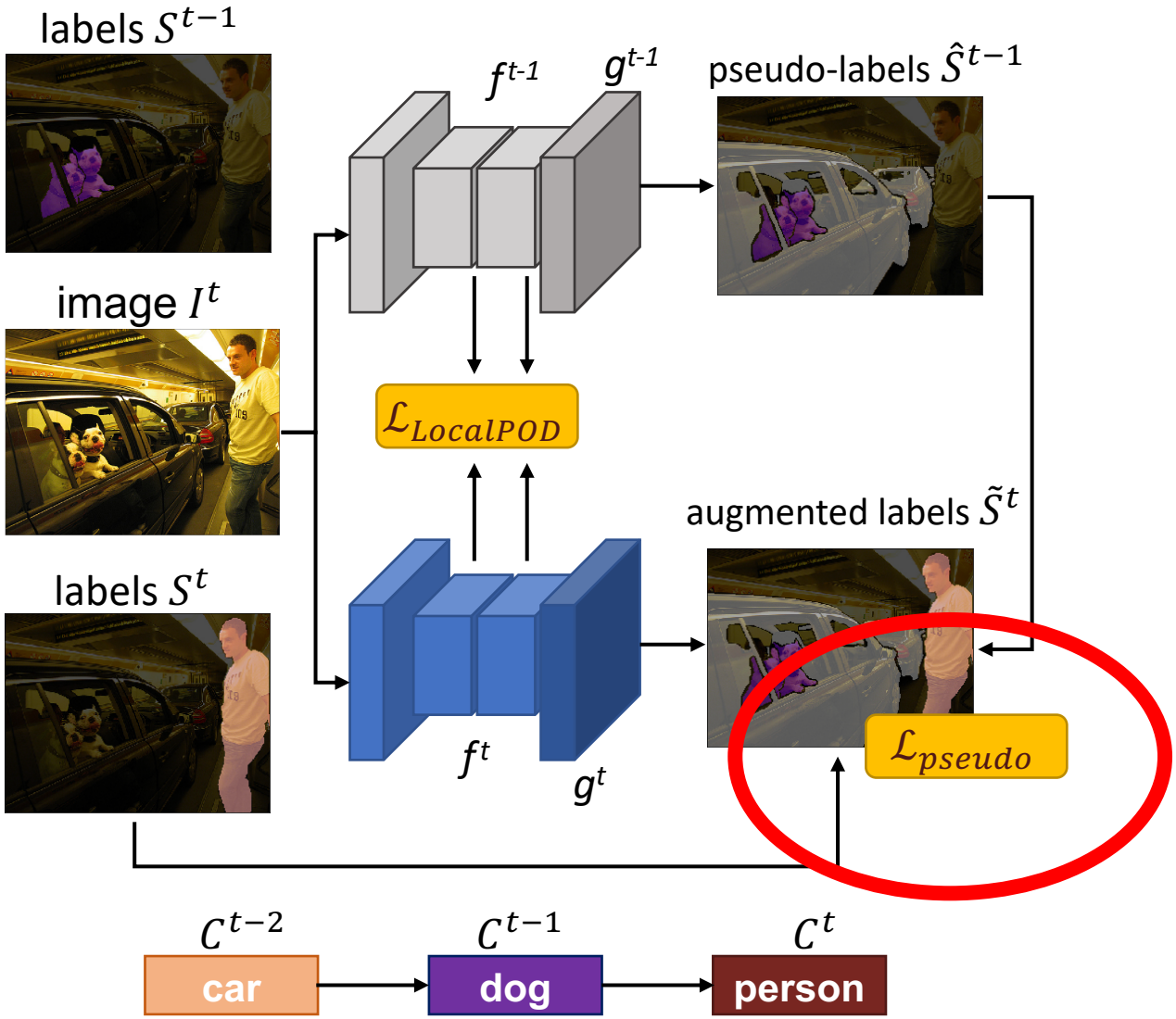
**Local** POD for the win!




| Distillation loss | 0-15 | 16-20 | *all* | *avg* |
|---|---|---|---|---|
| Knowledge Distllation | 29.72 | 4.42 | 23.69 | 49.18 |
| UNKD | 34.85 | 5.26 | 27.80 | 46.39 |
| POD | 43.94 | 4.82 | 34.62 | 53.35 |
| Local POD (Eq. 5) | **63.06** | **17.92** | **52.31** | **65.71** |

# PLOP Strategy




(a) PLOP strategy with pseudo-labeling and local POD

# Problem 1: Background shift

Step 1

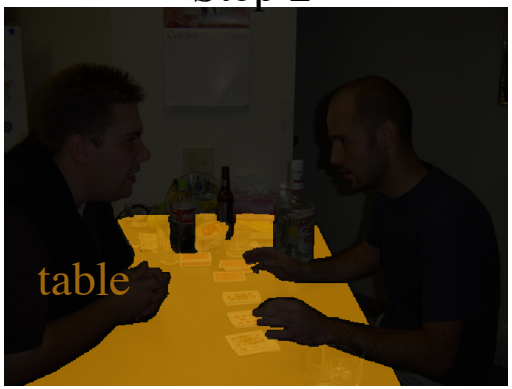
GT


Current Predictions

# Problem 1: Background shift

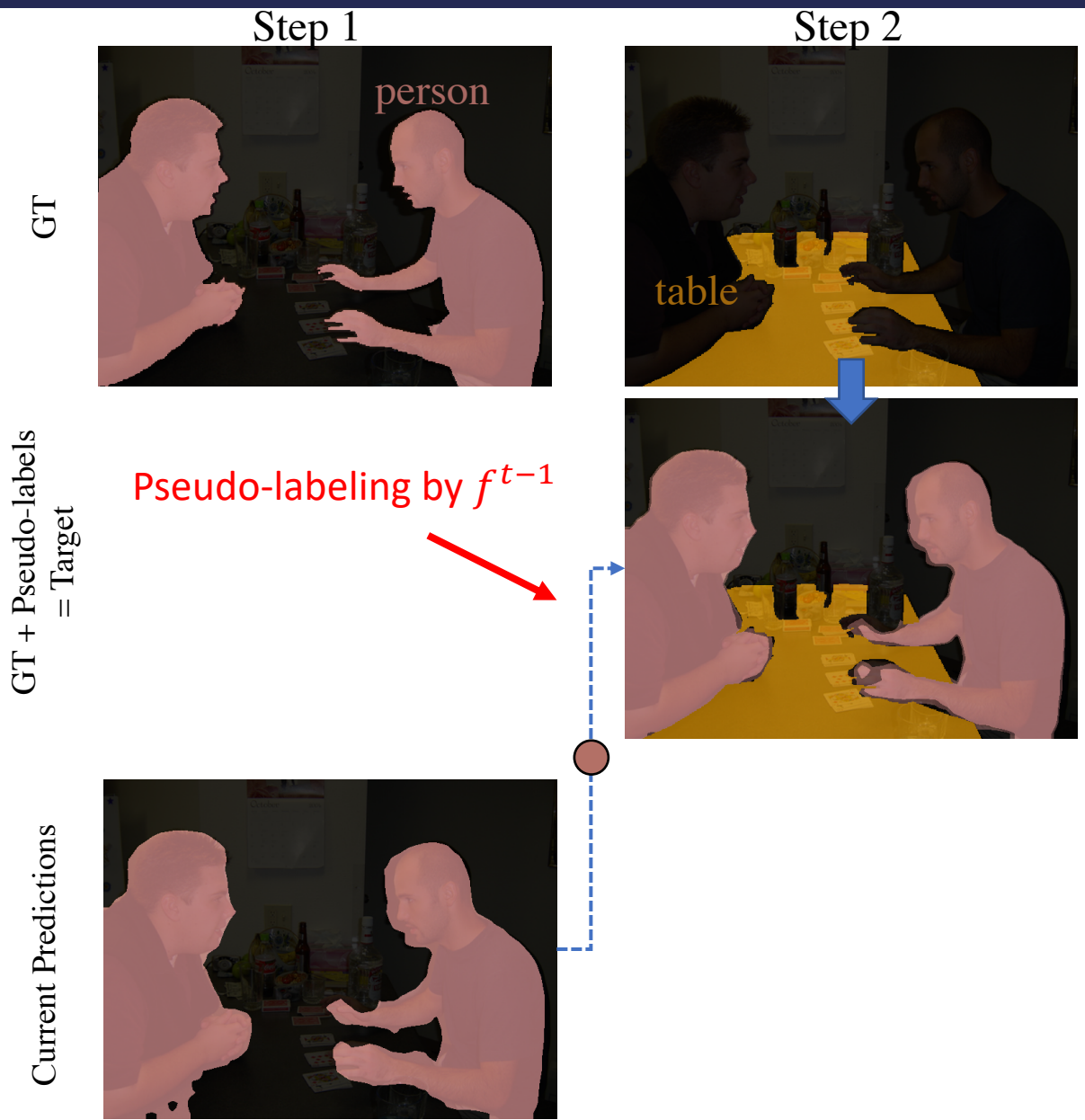
Step 1 Step 2

GT

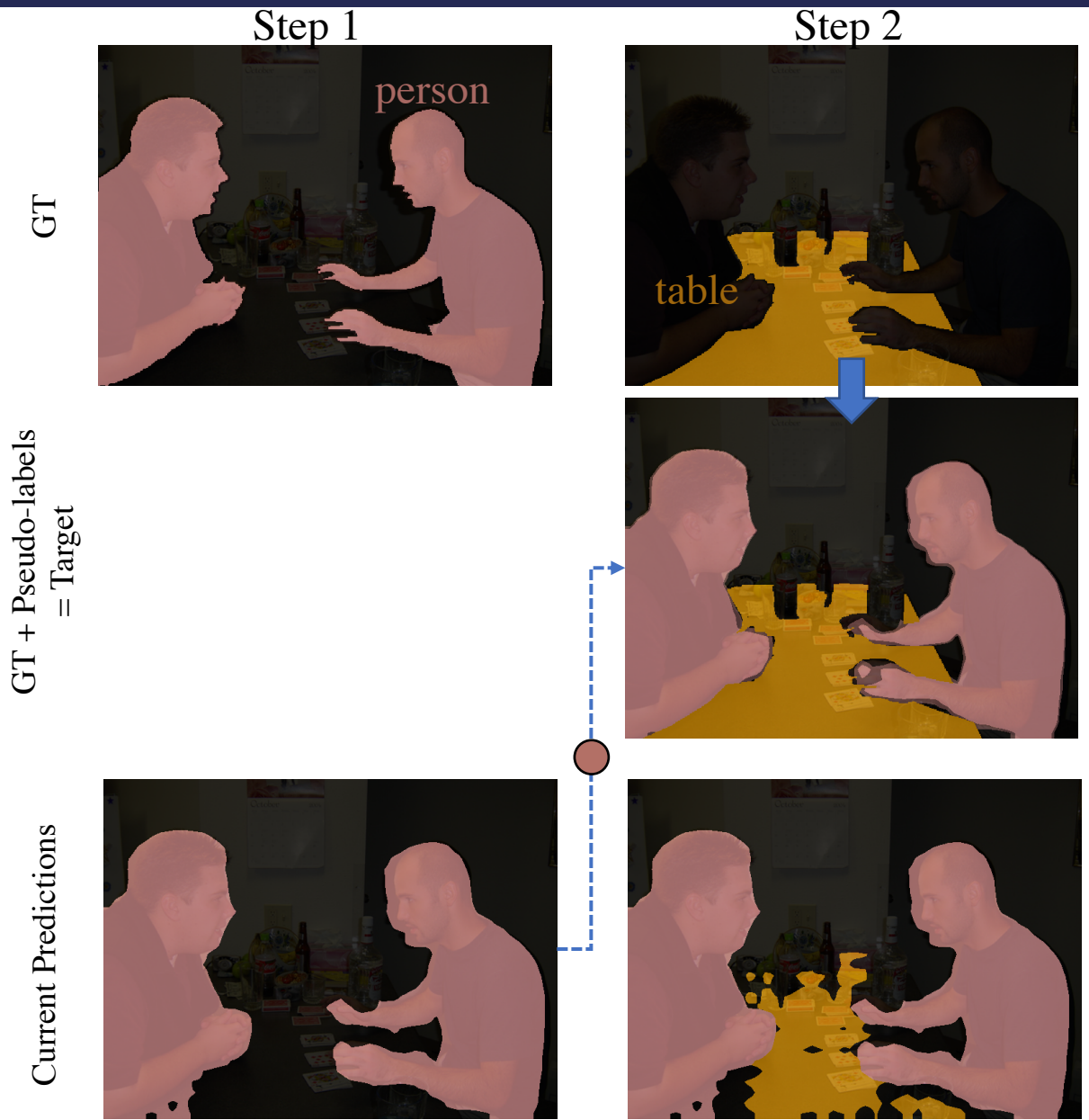




Current Predictions

# Problem 1: Background shift

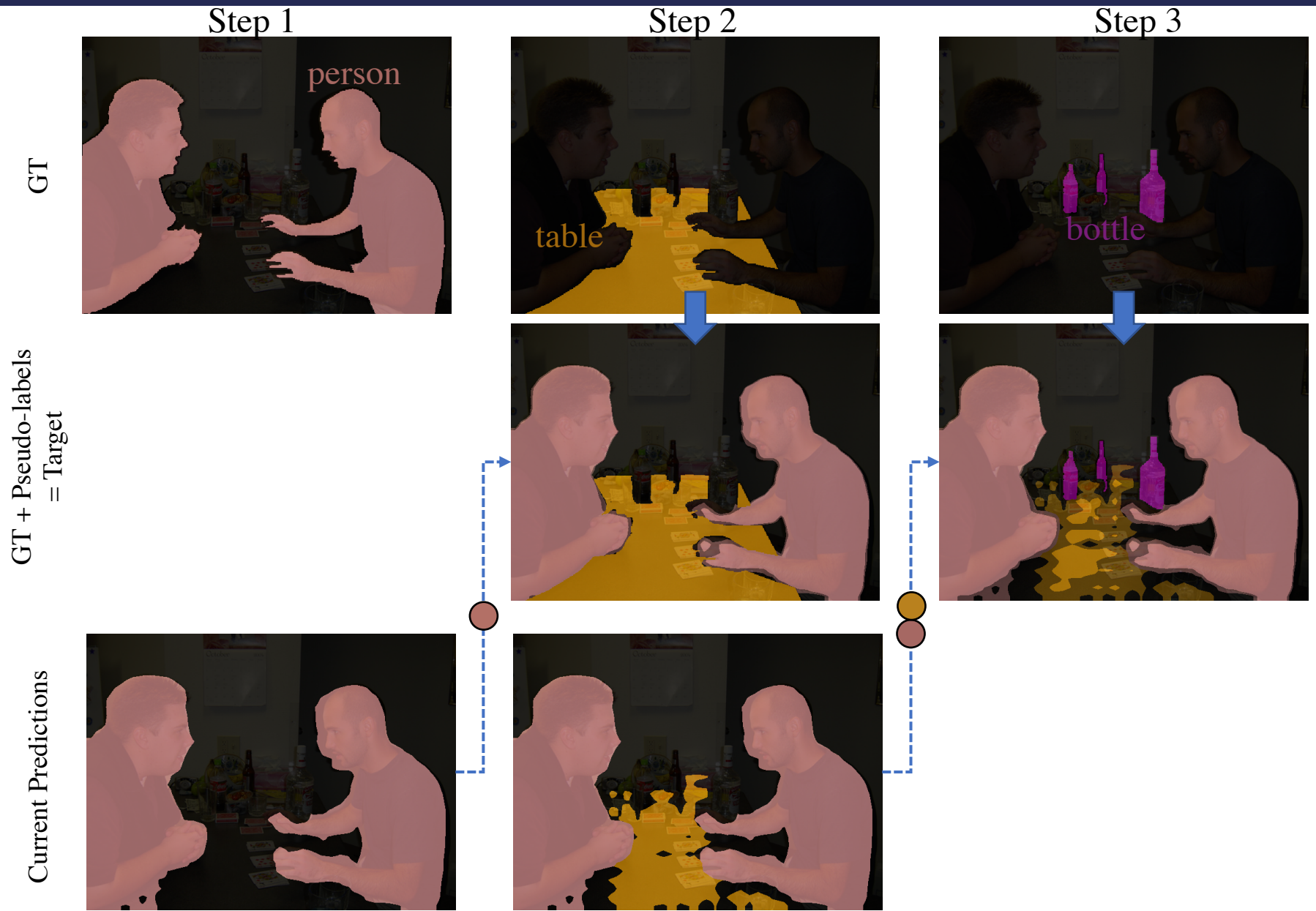


# Problem 1: Background shift

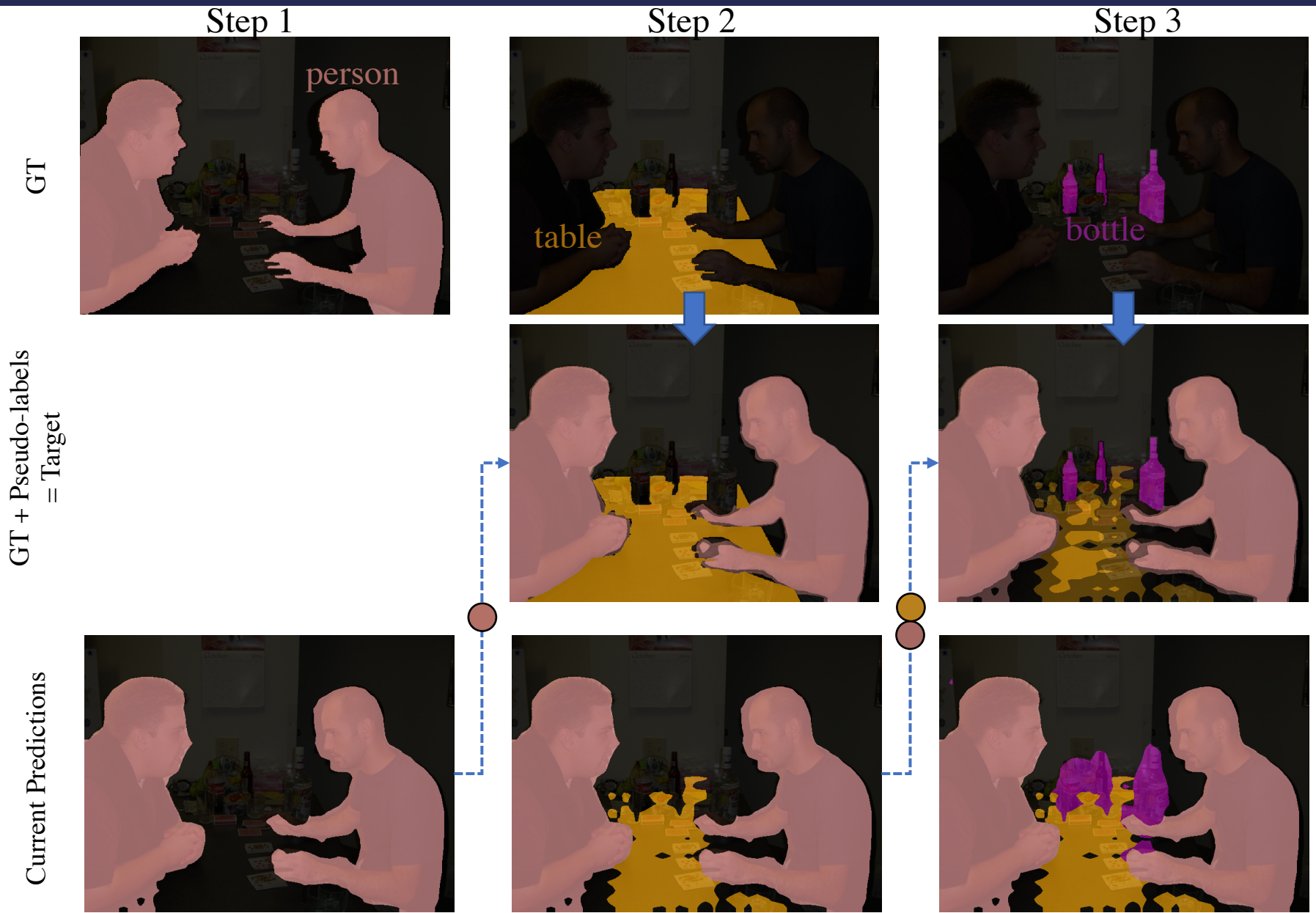


# Problem 1: Background shift

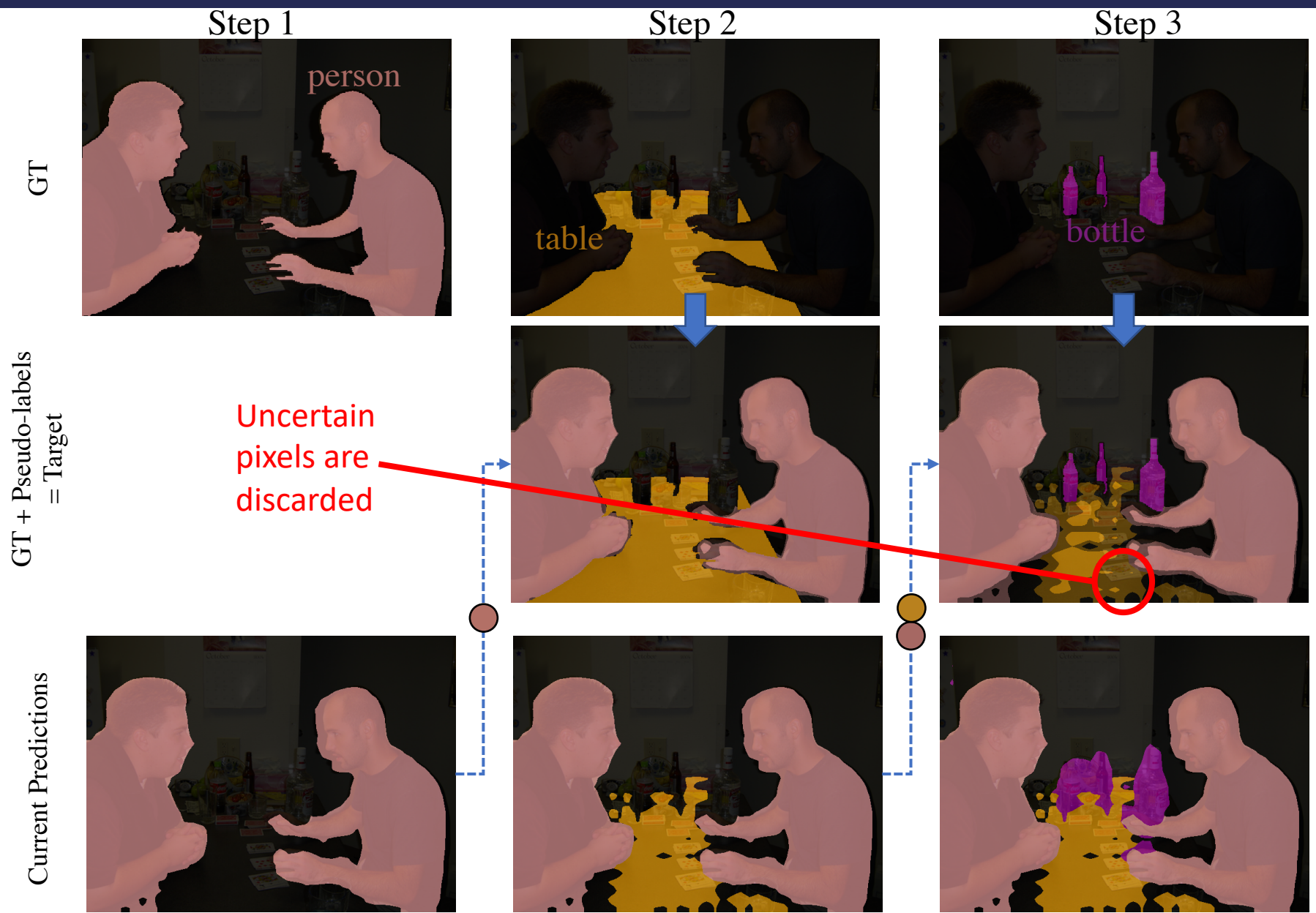
# Problem 1: Background shift



# Problem 1: Background shift



# Problem 1: Background shift



# Problem 1: Background shift

UNCE (CVPR 2020) merges predictions of old classes with background

| Classification loss | 1-15 | 16-20 | *all* | *avg* |
|---|---|---|---|---|
| CE only on new | 12.95 | 2.54 | 10.47 | 47.02 |
| CE | 33.80 | 4.67 | 26.87 | 50.79 |
| UNCE | 48.46 | 4.82 | 38.62 | 53.19 |
| Pseudo (Eq. 8) | **63.06** | **17.92** | **52.31** | **65.71** |
| *Pseudo-Oracle* | *63.69* | *23.35* | *54.09* | *66.05* |

# Experiments

Pascal-VOC (20 classes) experiments

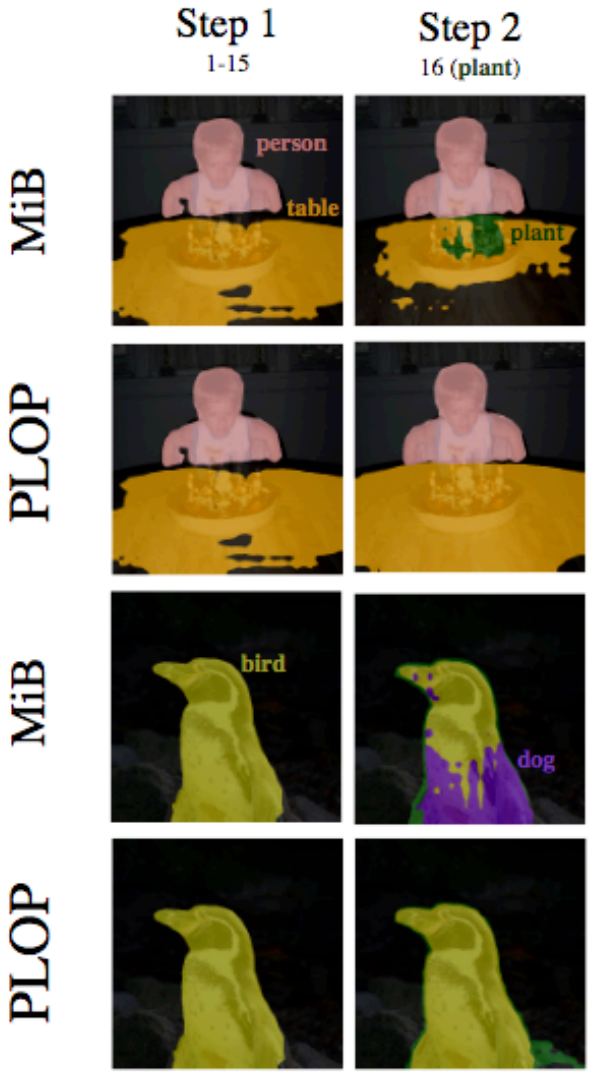
| Method | 19-1 (2 tasks) | | | | 15-5 (2 tasks) | | | | 15-1 (6 tasks) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1-19 | 20 | *all* | *avg* | 1-15 | 16-20 | *all* | *avg* | 1-15 | 16-20 | *all* | *avg* |
| EWC† [36] | 26.90 | 14.00 | 26.30 | | 24.30 | 35.50 | 27.10 | | 0.30 | 4.30 | 1.30 | |
| LwF-MC† [54] | 64.40 | 13.30 | 61.90 | | 58.10 | 35.00 | 52.30 | | 6.40 | 8.40 | 6.90 | |
| ILT† [49] | 67.10 | 12.30 | 64.40 | | 66.30 | 40.60 | 59.90 | | 4.90 | 7.80 | 5.70 | |
| ILT [49] | 67.75 | 10.88 | 65.05 | 71.23 | 67.08 | 39.23 | 60.45 | 70.37 | 8.75 | 7.99 | 8.56 | 40.16 |
| MiB† [7] | 70.20 | 22.10 | 67.80 | | 75.50 | 49.40 | 69.00 | | 35.10 | 13.50 | 29.70 | |
| MiB [7] | 71.43 | 23.59 | 69.15 | 73.28 | **76.37** | 49.97 | **70.08** | **75.12** | 34.22 | 13.50 | 29.29 | 54.19 |
| PLOP | **75.35** | **37.35** | **73.54** | **75.47** | 75.73 | **51.71** | **70.09** | **75.19** | **65.12** | **21.11** | **54.64** | **67.21** |

| Method | VOC 10-1 (11 tasks) | | | |
|---|---|---|---|---|
| | 1-10 | 11-20 | *all* | *avg* |
| ILT [55] | 7.15 | 3.67 | 5.50 | 25.71 |
| MiB [8] | 12.25 | 13.09 | 12.65 | 42.67 |
| PLOP | **44.03** | **15.51** | **30.45** | **52.32** |

# Visuals

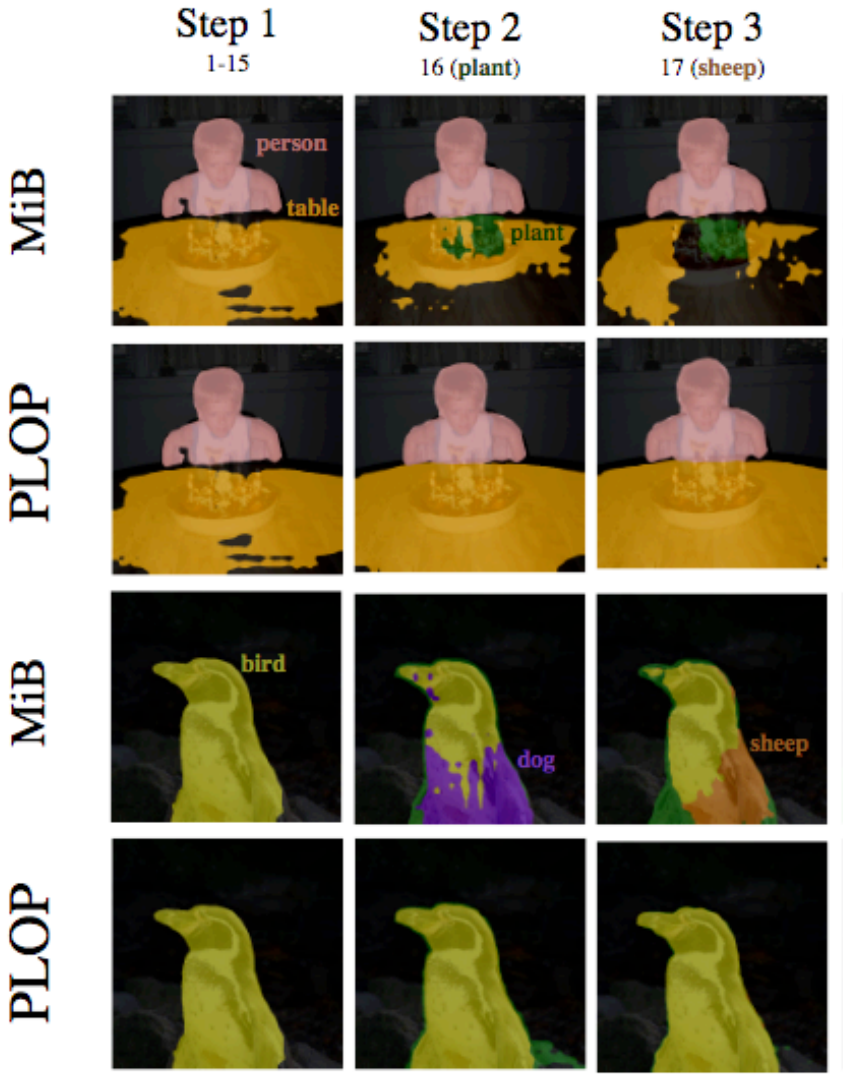

First, learn 15 classes

# Visuals



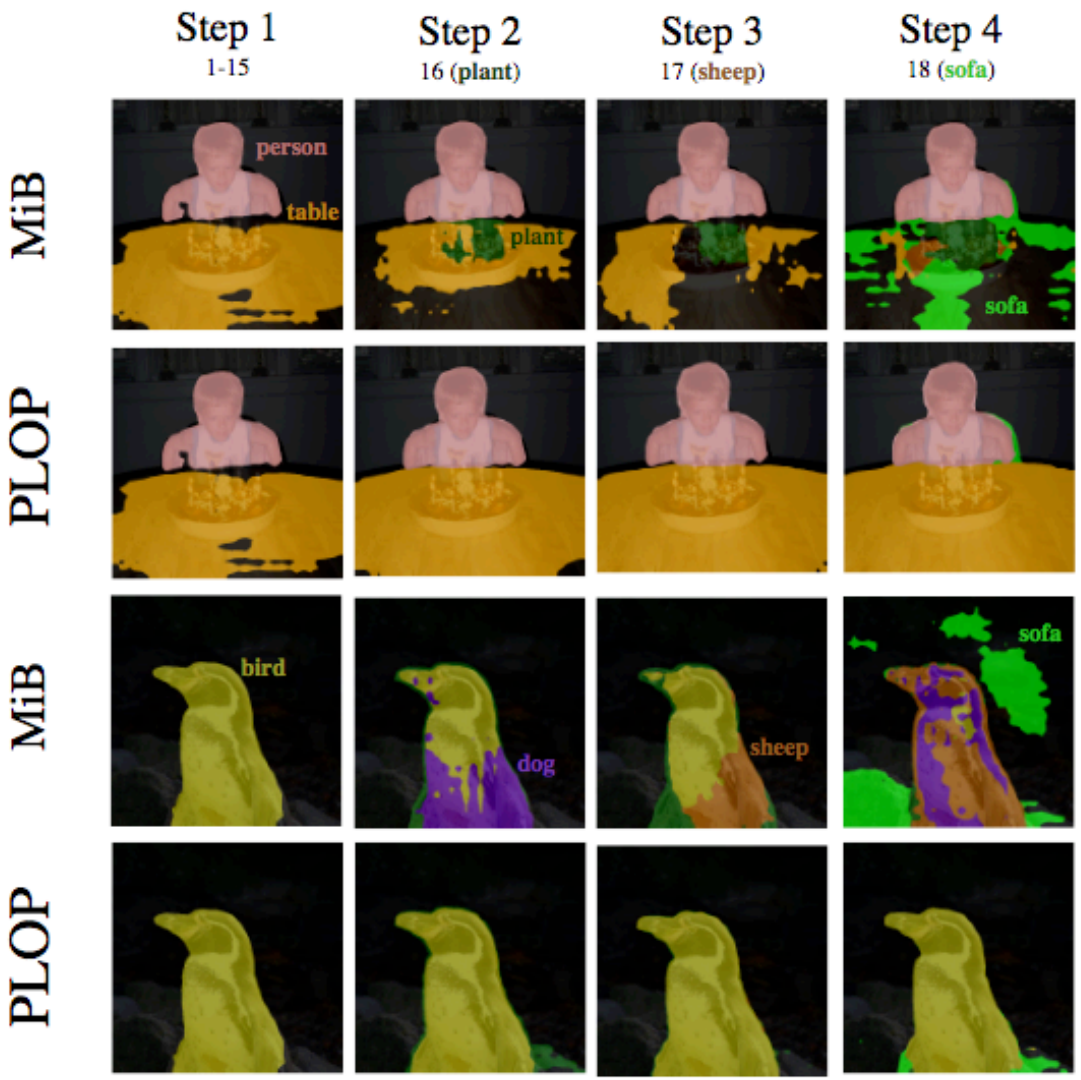

Learn the “plant” class

# Visuals




So far, it's still OK

# Visuals




Catastrophic forgetting

# Visuals

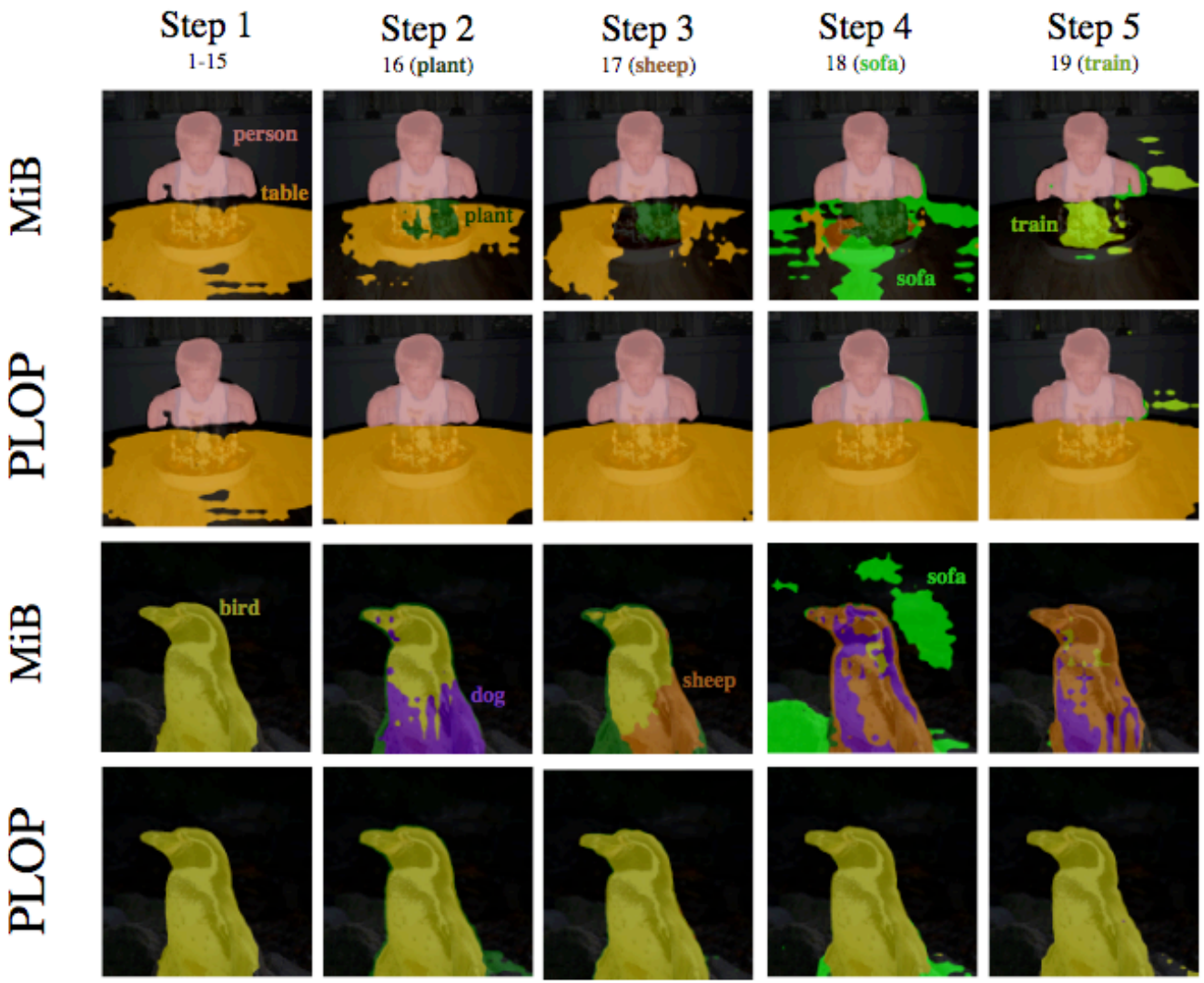
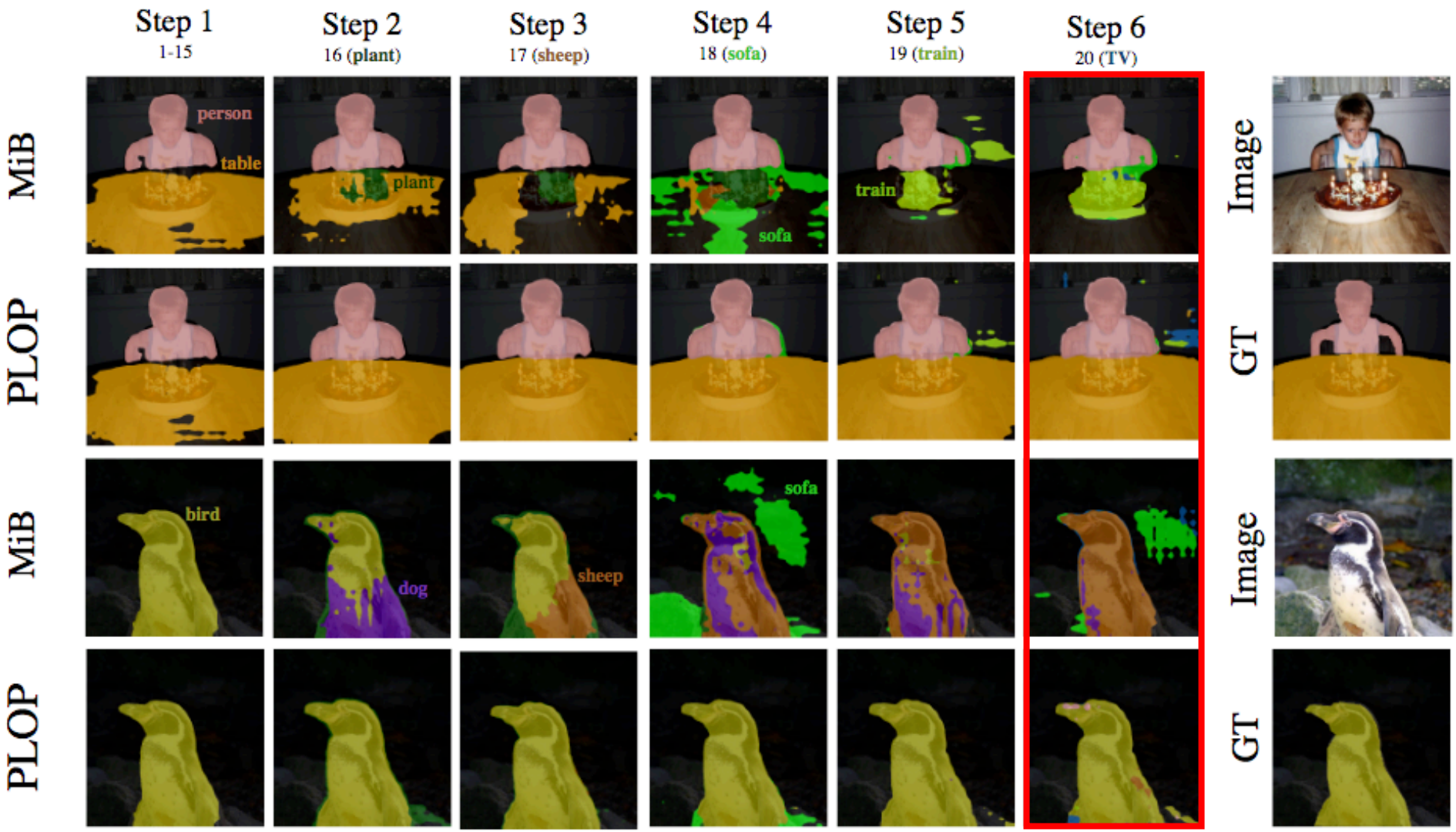

Step 1
1-15
Step 2
16 (plant)
Step 3
17 (sheep)
Step 4
18 (sofa)
Step 5
19 (train)
MiB
PLOP
MiB
PLOP
person
table
plant
sofa
train
bird
dog
sheep
sofa

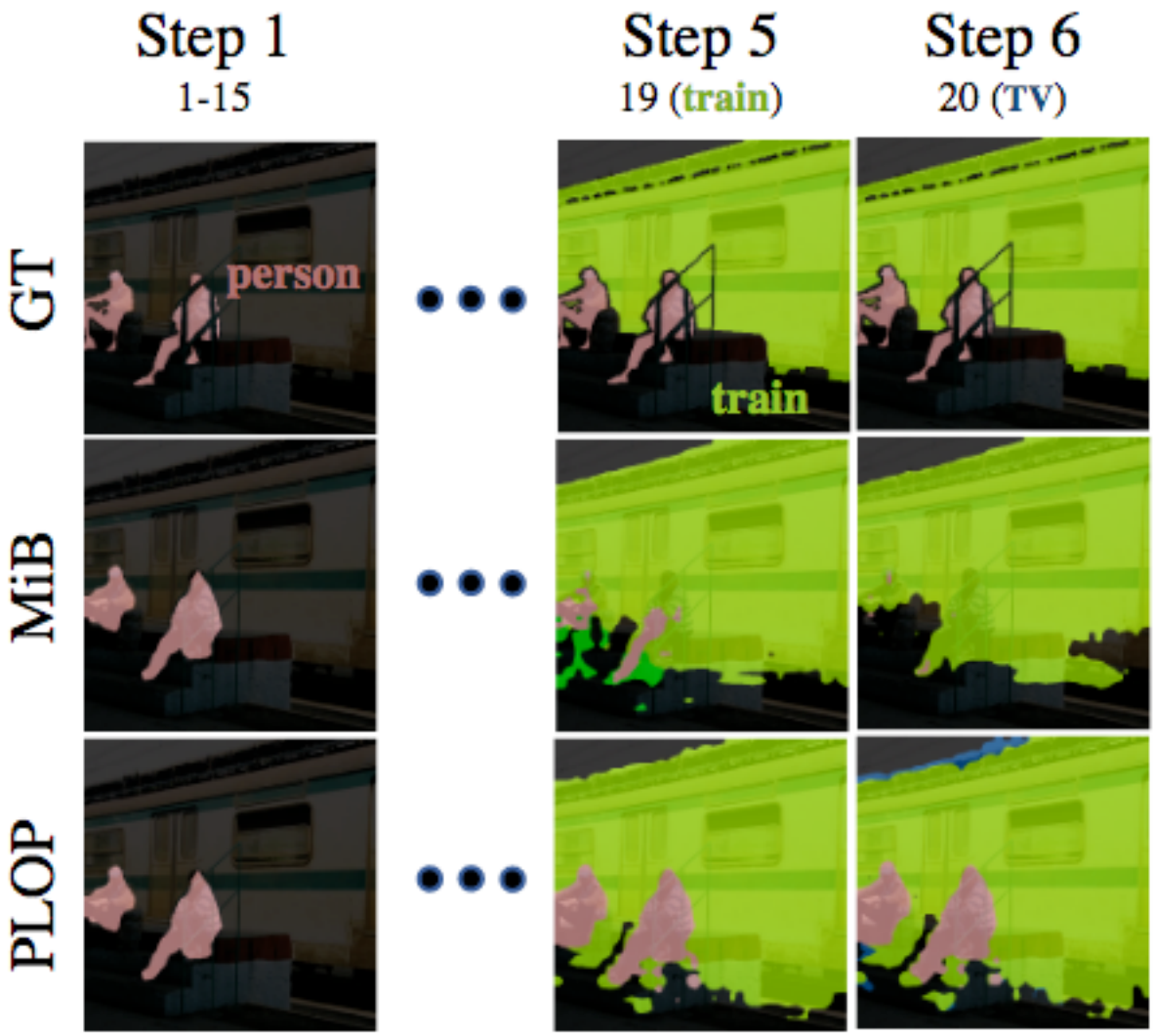

Image
GT
Image
GT

# Visuals


Step 1
1-15
Step 2
16 (plant)
Step 3
17 (sheep)
Step 4
18 (sofa)
Step 5
19 (train)
Step 6
20 (TV)
MiB
PLOP
MiB
PLOP
person
table
plant
sofa
train
bird
dog
sheep
sofa
Image
GT
Image
GT

# Visuals

When a class appear only latter in the image



# PLOP Strategy

**Class-agnostic** distillation loss

Pseudo-labeling to **complete partially-labeled images**