# ADAPT: Attentive Self-Distillation and Dual-Decoder Prediction Fusion for Continual Panoptic Segmentation

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

ICLR International Conference On Learning Representations

Ze Yang [1], Shichao Dong [1], Ruibo Li [1], Nan Song [1], Guosheng Lin [1]

[1] College of Computing and Data Science, Nanyang Technological University

## Introduction

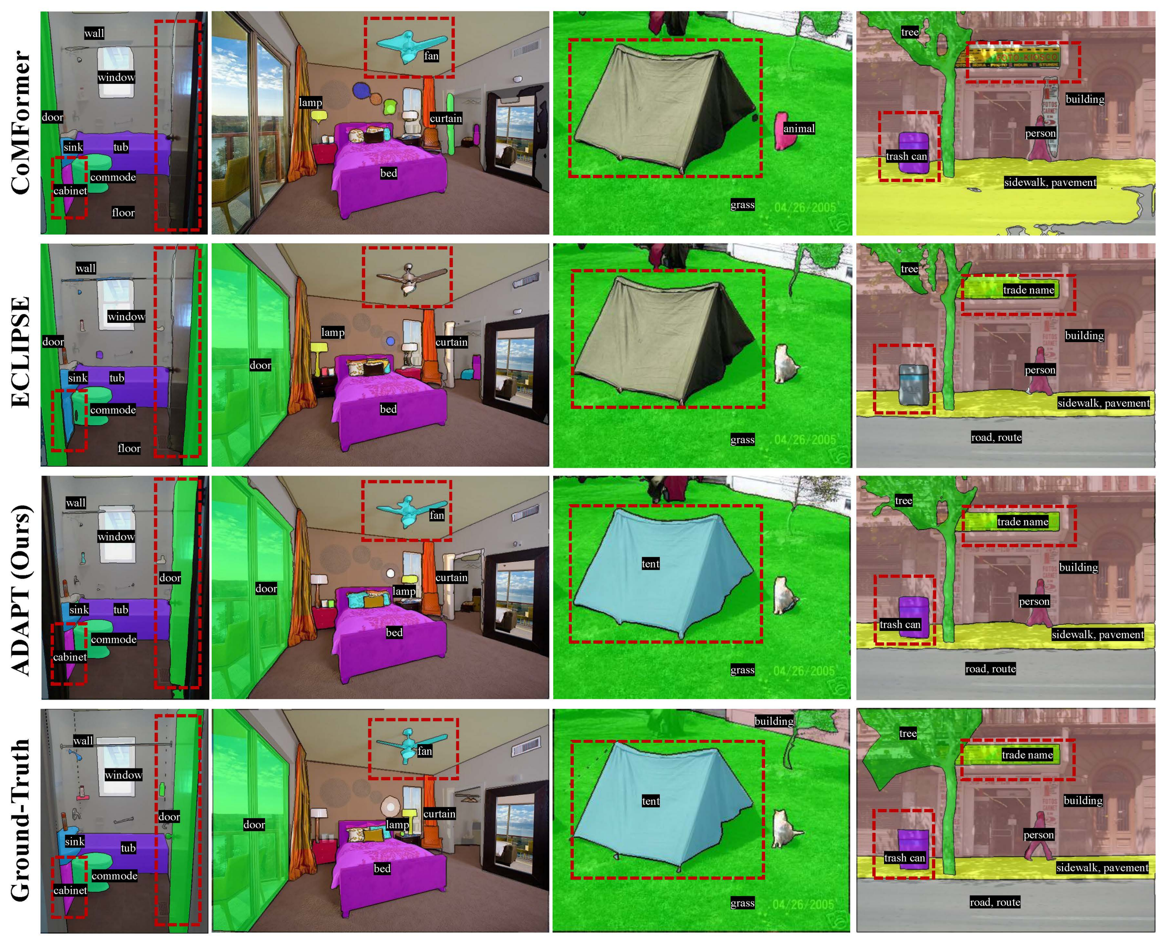
Continual panoptic segmentation (CPS) aims to extend an existing model to tackle unseen tasks while retaining its old knowledge. However, existing CPS methods generally suffer from:

- Catastrophic forgetting
- High computation (e.g., dual-model distillation)
- Scalability issues (e.g., additional prompts)

### Our contributions:

- Efficient adaptation via **frozen shared weights**
- **Attentive self-distillation** to retain old knowledge
- **Query-Level Fusion (QLF)** avoids scale mismatch
- State-of-the-art on **ADE20K** and **COCO**

## Visualization & Experiment Results




Qualitative visualization for ADE20K 100-10

### Training Time

CoMFormer: 3.36 hrs

ECLIPSE: 3.48 hrs

**ADAPT: 2.27 hrs** (−35%)

### Reference

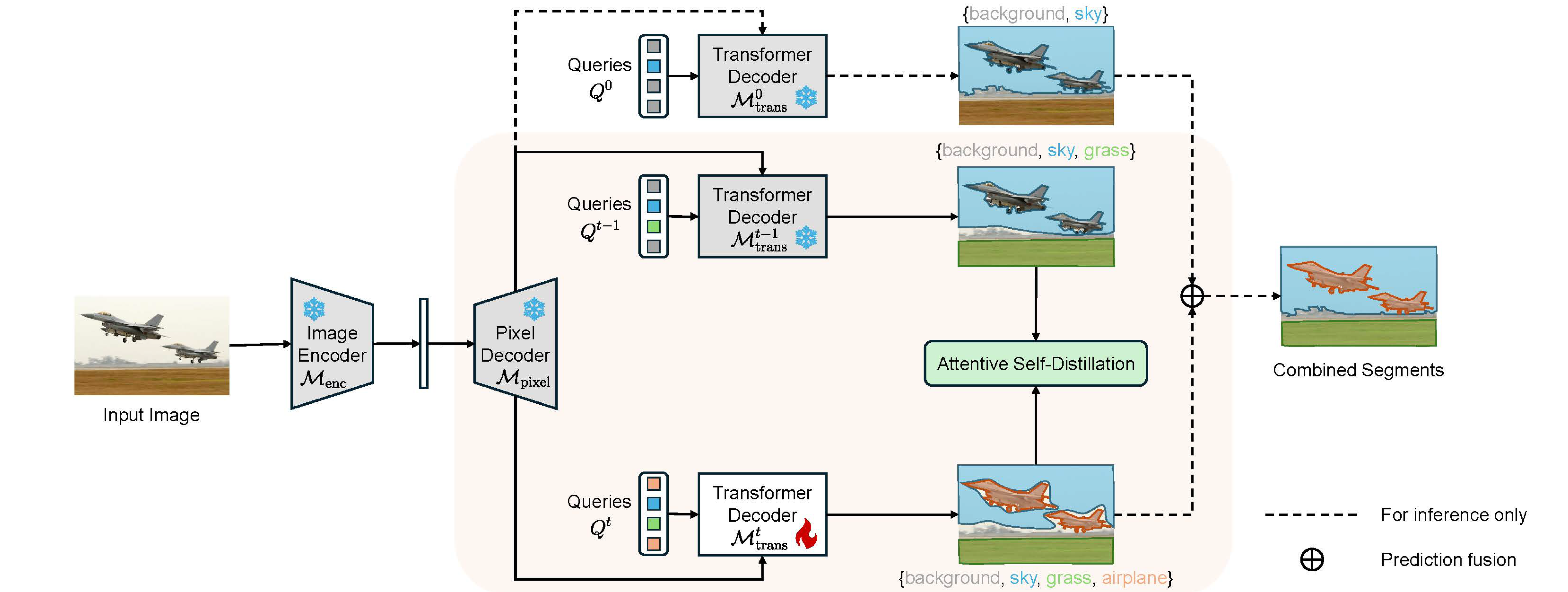
[CoMFormer] Cermelli, Fabio, Matthieu Cord, and Arthur Douillard. "Comformer: Continual learning in semantic and panoptic segmentation." CVPR 2023.

[ECLIPSE] Kim, Beomyoung, Joonsang Yu, and Sung Ju Hwang. "Eclipse: Efficient continual learning in panoptic segmentation with visual prompt tuning." CVPR 2024.

| Method | 100-5 (11 tasks) 1-100 | 100-5 101-150 | 100-5 all | 100-10 (6 tasks) 1-100 | 100-10 101-150 | 100-10 all | 100-50 (2 tasks) 1-100 | 100-50 101-150 | 100-50 all |
|---|---|---|---|---|---|---|---|---|---|
| FT† | 0.0 | 1.3 | 0.4 | 0.0 | 2.9 | 1.0 | 0.0 | 25.8 | 8.6 |
| MiB† Cermelli et al. (2020) | 24.0 | 6.5 | 18.1 | 27.1 | 10.0 | 21.4 | 35.1 | 19.3 | 29.8 |
| PLOP† Douillard et al. (2021) | 28.1 | 15.7 | 24.0 | 30.5 | 17.5 | 26.1 | 41.0 | 26.6 | 36.2 |
| CoMFormer† Cermelli et al. (2023) | 34.4 | 15.9 | 28.2 | 36.0 | 17.1 | 29.7 | 41.1 | 27.7 | 36.7 |
| ECLIPSE‡ Kim et al. (2024) | 41.1 | 16.6 | 32.9 | 41.4 | 18.8 | 33.9 | 41.7 | 23.5 | 35.6 |
| ECLIPSE* Kim et al. (2024) | 39.5 | 14.7 | 31.2 | 41.2 | 16.5 | 33.0 | 39.9 | 21.1 | 33.6 |
| ADAPT (Ours) | **42.3** | **19.8** | **34.8** | **42.2** | **26.3** | **36.9** | **42.5** | **27.8** | **37.6** |
| joint | 43.4 | 32.9 | 39.9 | 43.4 | 32.9 | 39.9 | 43.4 | 32.9 | 39.9 |

Continual panoptic segmentation results (PQ) on ADE20K

## Method




### 1. Efficient Adaptation

- Freeze **image encoder** & **pixel decoder** → **shared forward pass**
- Only update **cross-attention** & **FFN** in transformer decoder

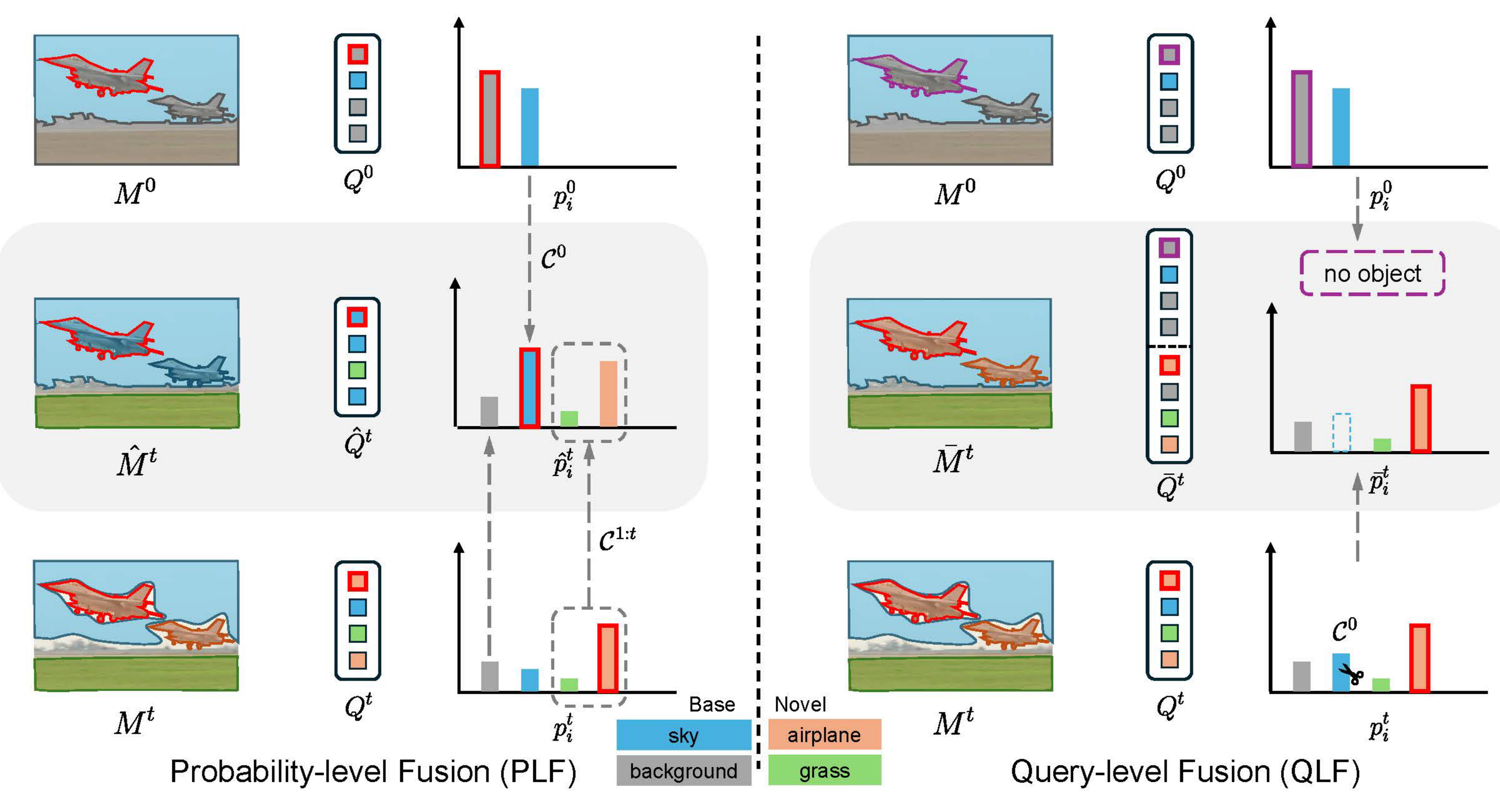
### 2. Attentive Distillation

- Distills only from **informative queries** (object regions)
- Weighted by background confidence using $\alpha(1-\hat{p}_x^{t-1}(i,\emptyset))^\gamma$

### 3. Query-Level Fusion (QLF)

- Use **fixed decoder** for base classes
- Use **adapted decoder** for novel classes
- Avoid probability-level scale mismatch

## Query-Level Fusion




Probability-level Fusion (PLF) vs. Query-level Fusion (QLF)

**Query-Level Fusion** avoids probability-level conflicts by assigning **distinct responsibilities** to the two decoders:

- The **base decoder** only predicts **base classes**.
- The **adapted decoder** only predicts **novel classes**.

### During inference

- **All queries** (e.g., 8 in total) from both decoders are combined.
- A **standard panoptic fusion** (like in Mask2Former) is applied to select final segments.
- Overlapping or redundant predictions are suppressed, similar to **non-maximum suppression** (NMS).