Bridging Domains with Minimal Supervision: Domain Adaptation and Generalization for Semantic Segmentation

Yasser Benigmim
Institut Polytechnique de Paris

Outline

- [CVPR-W’23] DATUM : One-shot Unsupervised Domain Adaptation with Personalized Diffusion Models

- [CVPR’24] CLOUDS : Collaborating Foundation models for Domain Generalized Semantic Segmentation

DATUM : One-shot Unsupervised Domain Adaptation with Personalized Diffusion Models

Yasser Benigmim [1 2], Subhankar Roy [1], Slim Essid [1], Vicky Kalogeiton [2], Stéphane Lathuilière [1]

[1] LTCI, Télécom-Paris, Intitute Polytechnique de Paris
[2] LIX, Ecole Polytechnique, CNRS, Intitute Polytechnique de Paris

https://github.com/yasserben/DATUM

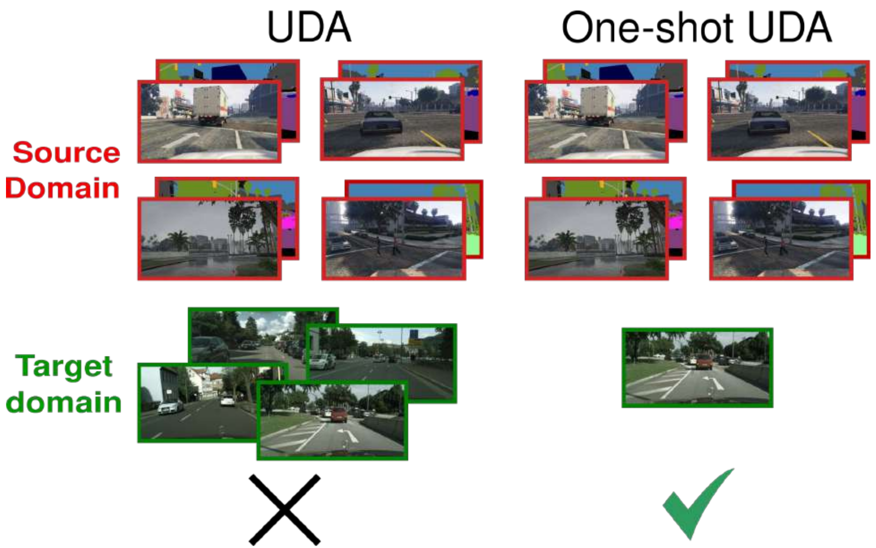
CVPR-W'23

TELECOM Paris
IP PARIS

l'X ÉCOLE POLYTECHNIQUE
IP PARIS

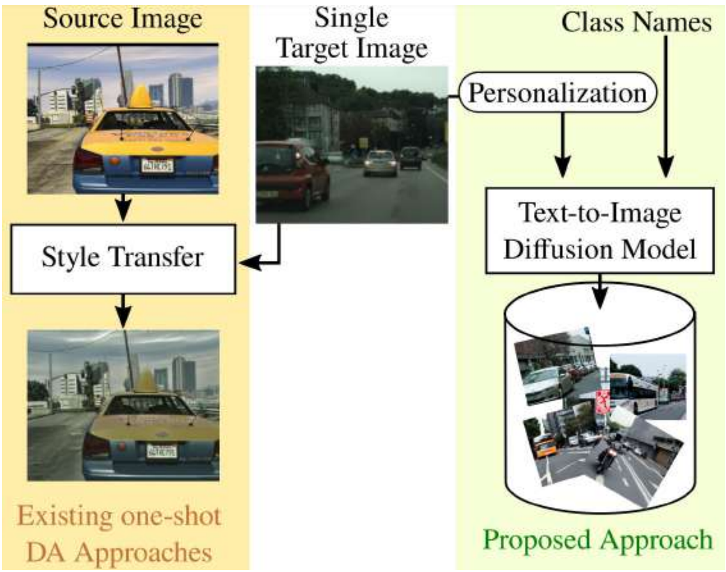
One-shot UDA

- We have access to a **labeled source domain** and **one unlabeled image** from target domain



Previous works

- Previous works [1,2], rely on **style transfer** to adapt the source images to the target



1 : Luo, Yawei, et al. "Adversarial style mining for one-shot unsupervised domain adaptation." NeurIPS 2020
2: Wu, Xinyi, et al. "Style mixing and patchwise prototypical matching for one-shot unsupervised domain adaptive semantic segmentation." AAAI 2022

Dreambooth

- The key idea behind Dreambooth is to **associate a unique identifier to the concept** we want to inject in a diffusion model

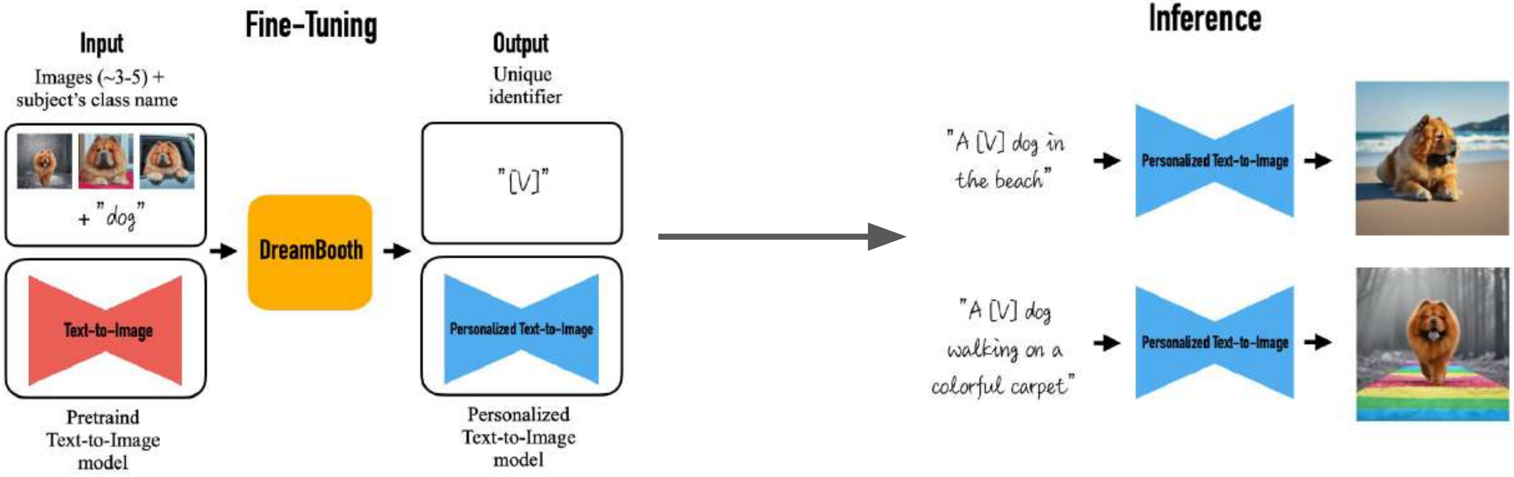
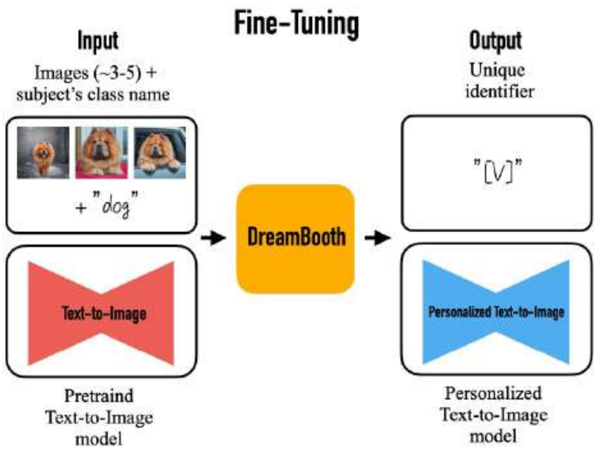


Image credits : Ruiz, Nataniel, et al. "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation." CVPR 2023

Step 1 : Personalization

- We **finetune** our diffusion model with the single target image using **Dreambooth**



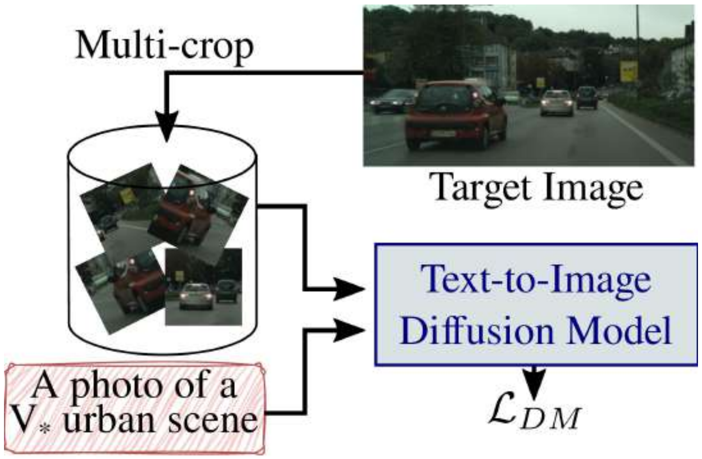
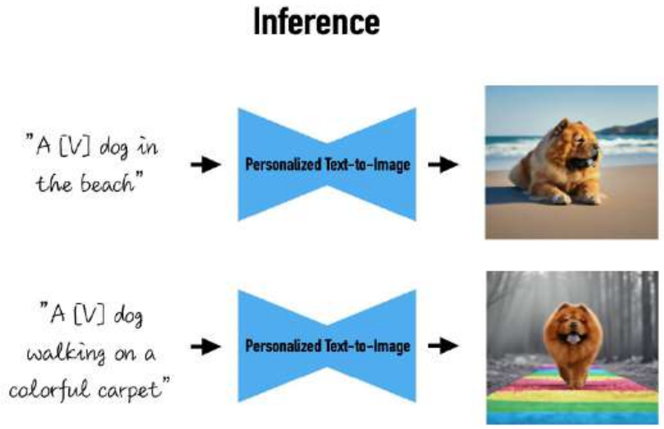


Image credits : Ruiz, Nataniel, et al. "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation." CVPR 2023

Step 2 : Generation

- We **generate new images** using the unique identifier associated with the target image
- We use **class-specific prompts + unique identifier** to increase image diversity



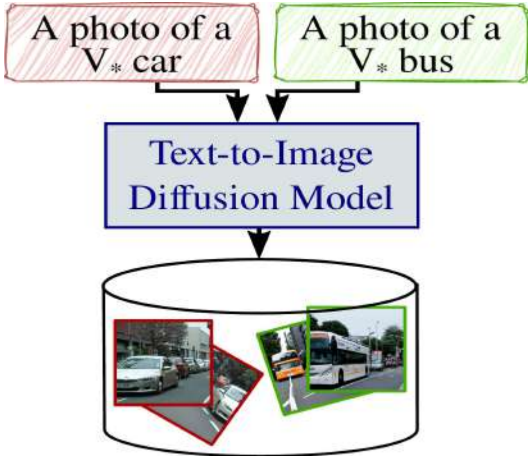
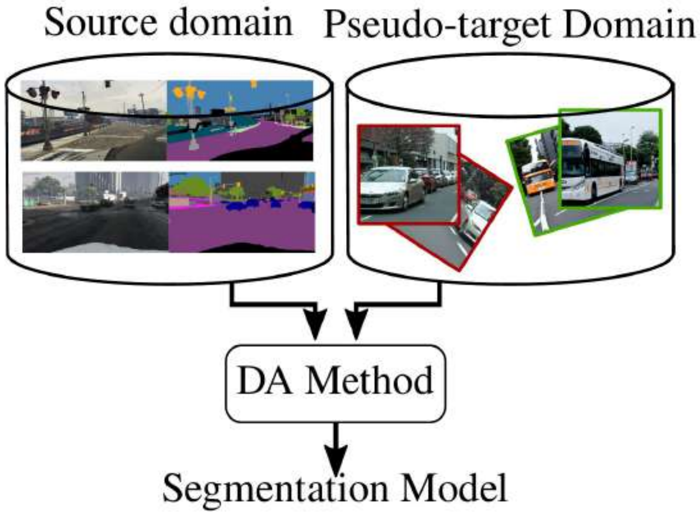


Image credits : Ruiz, Nataniel, et al. "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation." CVPR 2023

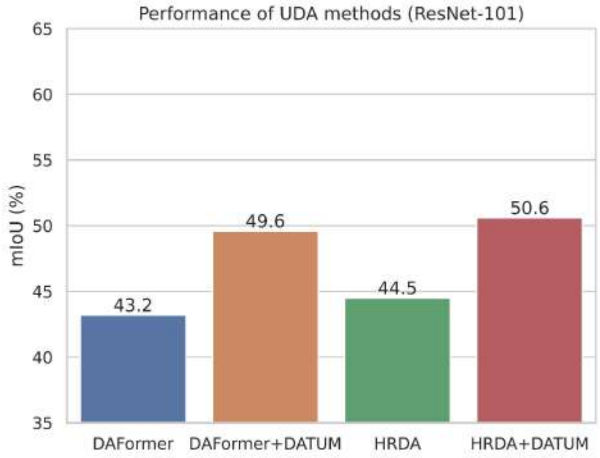
Step 3 : Adaptation

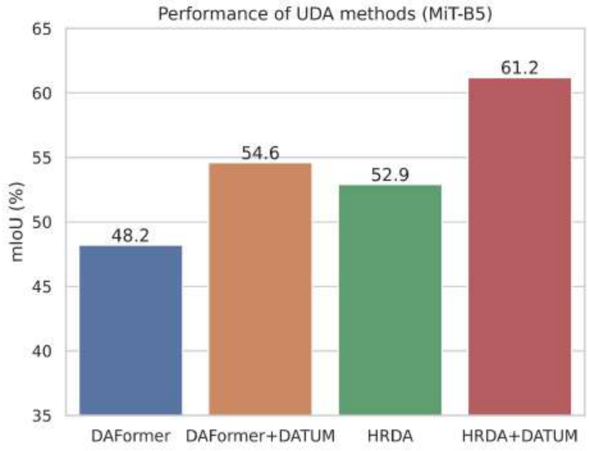
- We **inject the generated dataset** into any previous UDA framework



Results

- DATUM is **plug-and-play** method making any UDA method work in a data-scarce scenario





CLOUDS : Collaborating Foundation models for Domain Generalized Semantic Segmentation

Yasser Benigmim [1 2], Subhankar Roy [3], Slim Essid [1], Vicky Kalogeiton [2], Stéphane Lathuilière [1]

[1] LTCI, Télécom-Paris, Intitute Polytechnique de Paris
[2] LIX, Ecole Polytechnique, CNRS, Intitute Polytechnique de Paris
[3] University of Aberdeen

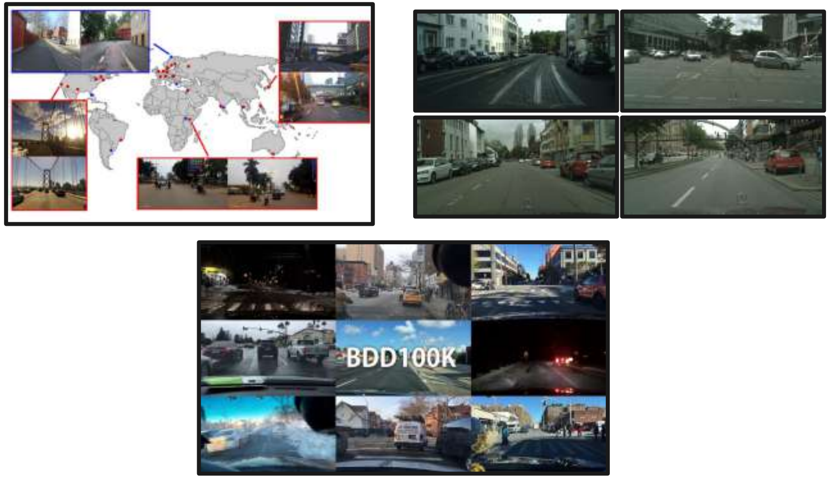
https://github.com/yasserben/CLOUDS

CVPR’24

Domain Generalization

- in DG, we only have access to a **labeled source domain** during training
- At inference, we test our model on **unseen target domains**

Source domain



Unseen target domains

Previous works

- in DG, most methods rely on **Domain Randomization** or **tailor-made modules** to eliminate **domain specific features**

- They also relied on **vision-centric models** pretrained on **ImageNet** like ResNet-50, 101

Style-Hallucinated Dual Consistency Learning for Domain Generalized Semantic Segmentation

Texture Learning Domain Randomization for Domain Generalized Segmentation

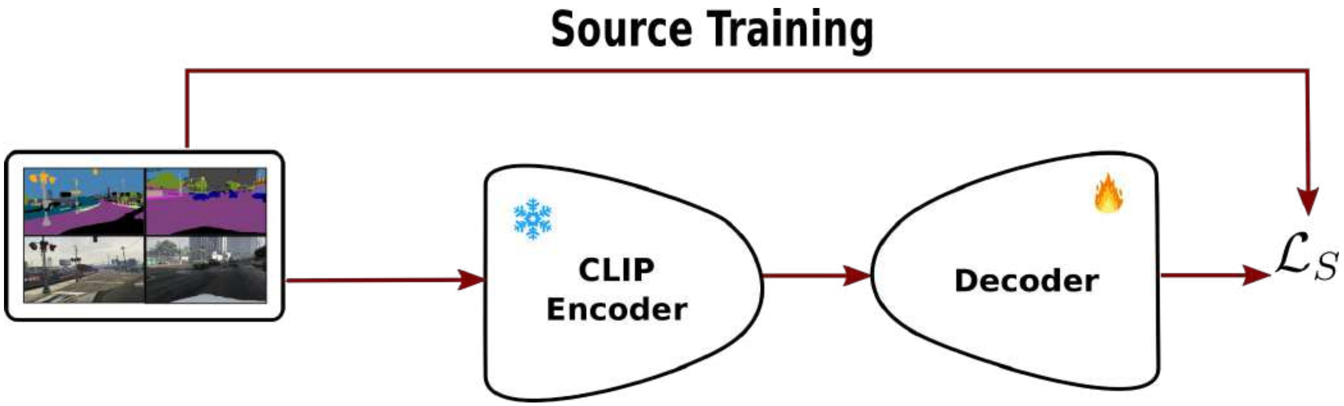
Adversarial Style Augmentation for Domain Generalized Urban-Scene Segmentation

RobustNet: Improving Domain Generalization in Urban-Scene Segmentation via Instance Selective Whitening

Two at Once: Enhancing Learning and Generalization Capacities via IBN-Net

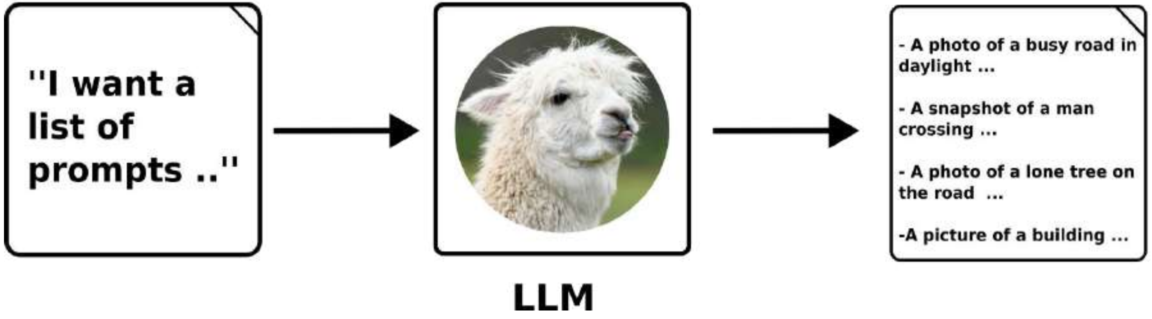
Usage of CLIP

- We first decided to use the feature representations of **CLIP** for their **strong generalizability**



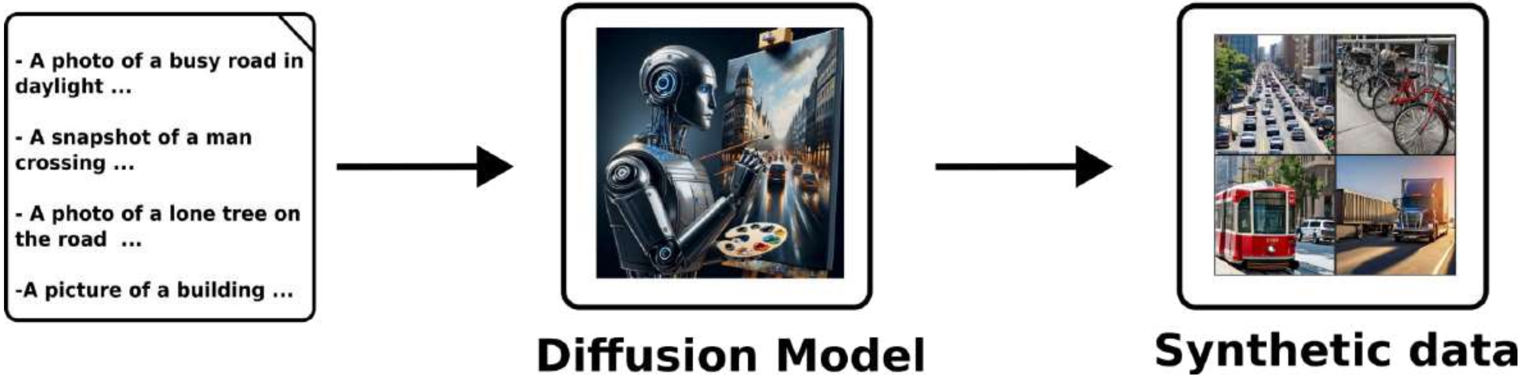
Usage of an LLM

- We want to generate **synthetic data** that is **as diverse as possible**
- We increase **diversity** of prompts using a **Large Language Model**



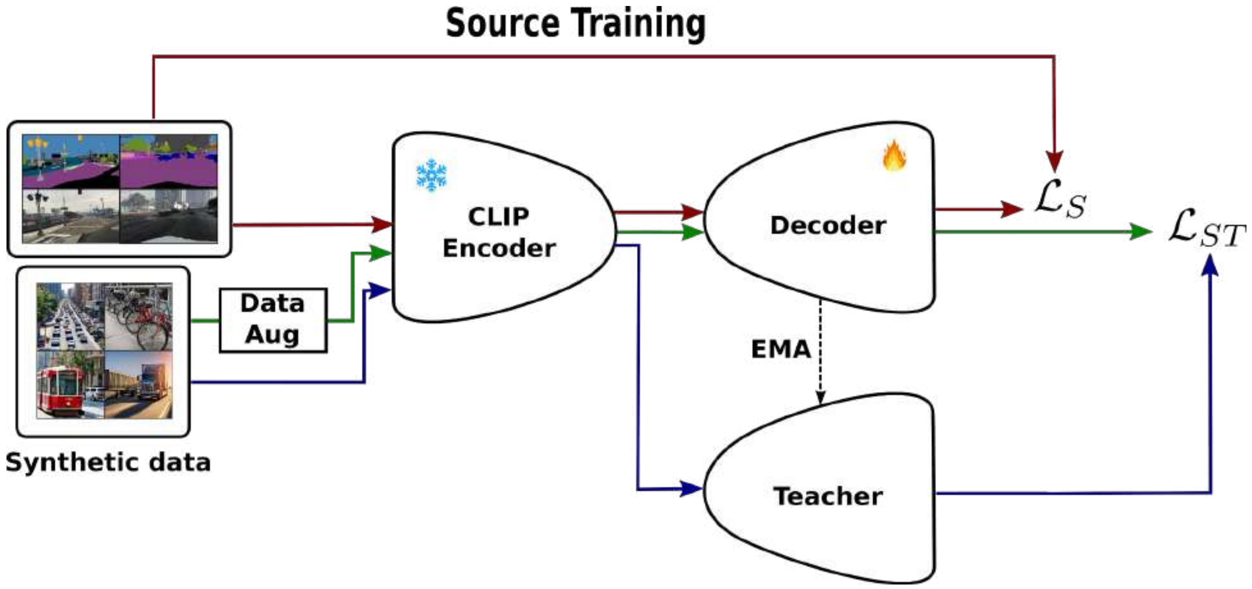
Usage of a Diffusion Model

- We generate **synthetic data** that is **as diverse as possible**



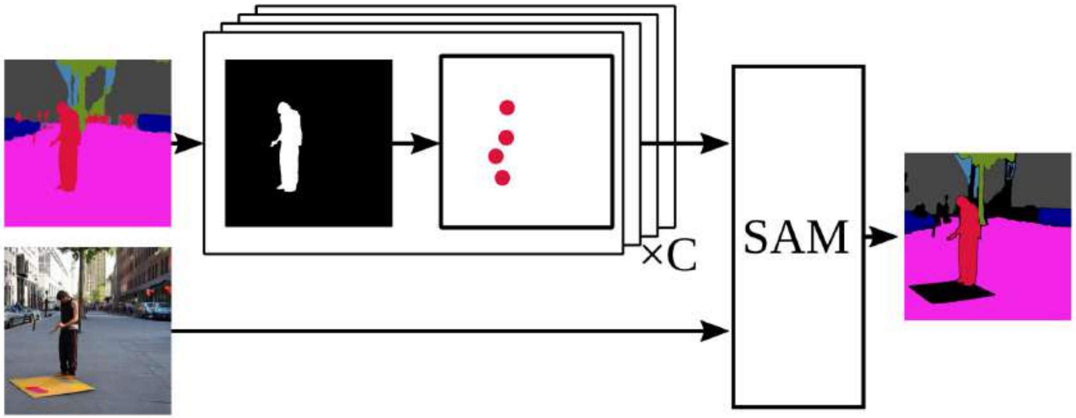
Self-Training

- We **self-train** the model on the generated data using **pseudo-labels (PLs)**



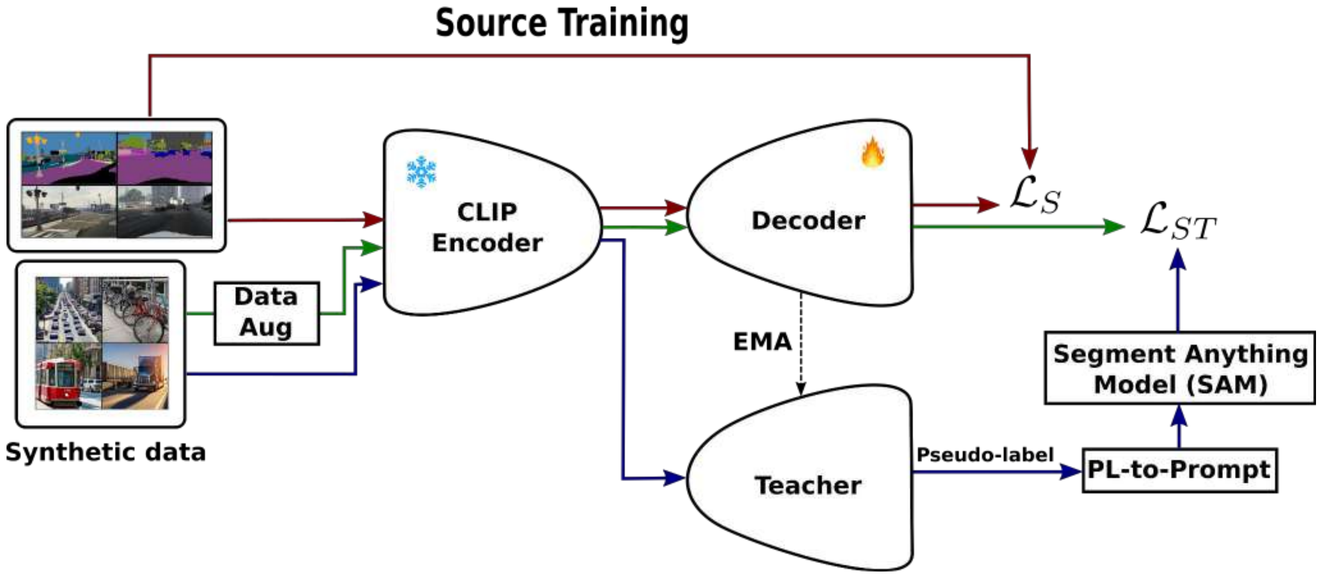
PL-refinement

- To improve the **noisy PLs**, we use the **Segment Anything Model (SAM)** to refine them

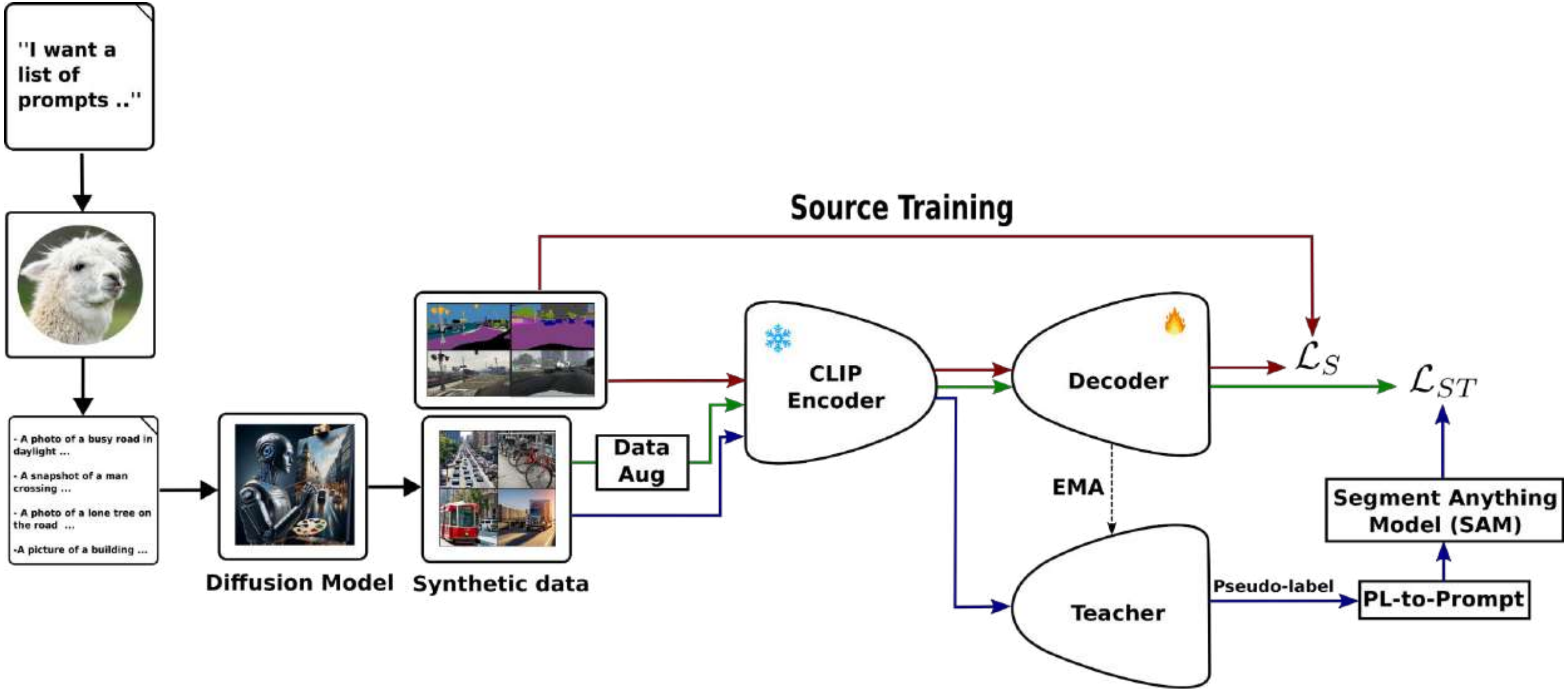


Training pipeline

- We incorportate our PL-refinement module during training
- We use MobileSAM for faster inference

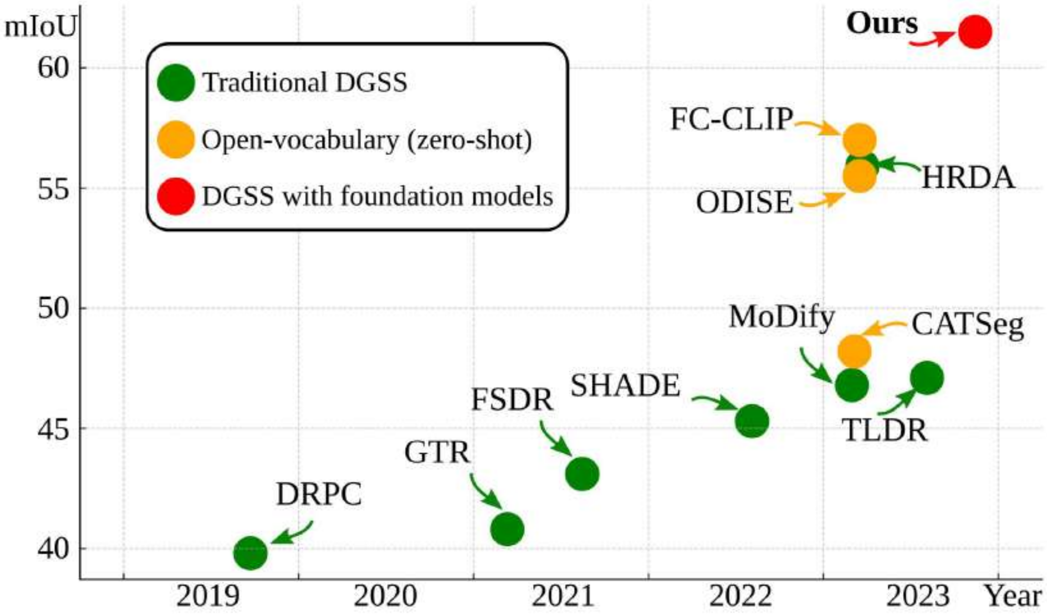


CLOUDS : System of Foundation models



Results

- CLOUDS outperforms previous **traditional** DGSS methods by a **large margin**



Thank you !