

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS519 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS519.Q11

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



Semi-supervised Semantic Segmentation

Trần Vạn Tân - 23521407

Lại Khánh Hoàng - 23520519

Tóm tắt

- Nội dung thuyết trình: Giới thiệu - Mục tiêu - Nội dung và phương pháp - Kết quả dự kiến



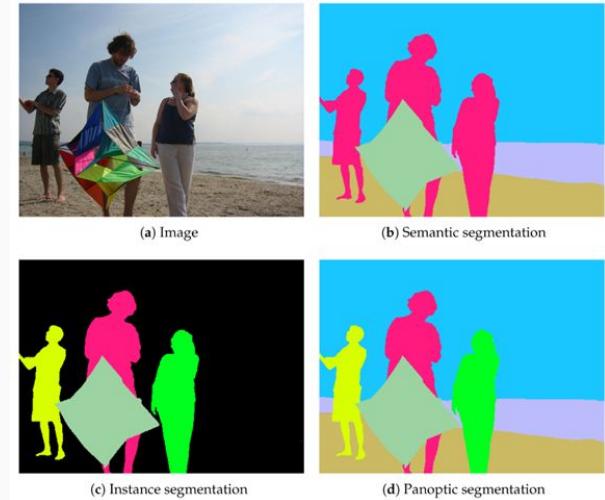
Trần Vạn Tấn
23521407

Lại Khánh Hoàng
23520519

- [Link Github của nhóm](#)
- [Link Youtube của nhóm](#)

Giới thiệu

- **Semantic Segmentation** là bài toán trong thị giác máy tính nhằm **gán một nhãn cho từng pixel** trong ảnh. Mục tiêu là mô hình phải **hiểu nội dung ảnh ở mức chi tiết nhất**, không chỉ nhận dạng object mà còn xác định chính xác **ranh giới và vùng** mà object chiếm giữ.
 - Input: ảnh RGB
 - Output: mặt nạ (mask) cùng kích thước ảnh, mỗi pixel thuộc một lớp như *road, sky, building, car...*
- Mặc dù segmentation rất quan trọng, việc **gán nhãn mask cho từng pixel cực kỳ tốn công và tốn thời gian**. Vì vậy trong thực tế thường có rất **ít ảnh có nhãn đầy đủ** và rất **nhiều ảnh không nhãn**.
- **Semi-Supervised Semantic Segmentation** là bài toán segmentation trong đó **chỉ một phần nhỏ dữ liệu có mask**, phần còn lại là ảnh **unlabeled**, nhưng mô hình vẫn phải học phân đoạn tốt.



Giới thiệu

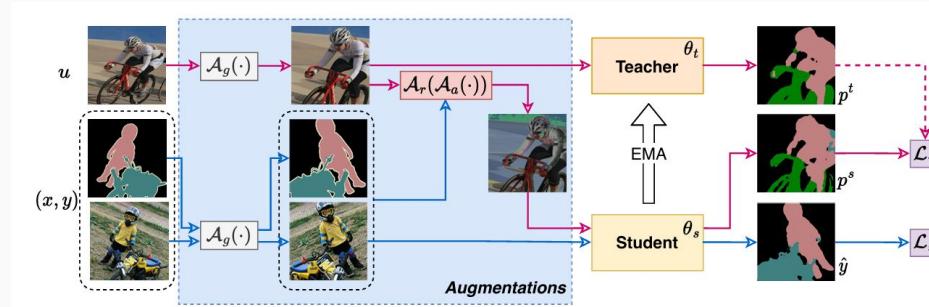
- Các model **SOTA** hiện nay, theo paper “**Training Vision Transformer for Semi-supervised Semantic Segmentation**” - CVPR 2024:

Dataset Split		1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)
Backbone	Methods				
DeepLabV3+ (ResNet101)	Sup-Only	66.3	72.8	75.0	78.0
	U ² PL [45]	74.9	76.5	78.5	79.1
	PS-MT [31]	—	76.9	77.6	79.1
	AugSeg [56]	75.2	77.8	79.6	80.4
	UniMatch [53]	76.6	77.9	79.2	79.5
(Dual Back.)	CVT	67.2	73.1	75.1	78.6
	SemiCVT [20]	72.2	75.4	77.2	79.6
SegFormer (MiT-B4)	Sup-Only	73.3	77.0	79.1	80.4
	S ⁴ Former-Base	76.7	79.2	80.1	80.6
	+ Ours	78.5	79.9	80.6	80.9

- Các paper đều thực hiện thử nghiệm trên hai dataset chính là Pascal VOC 2012 và Cityscapes.

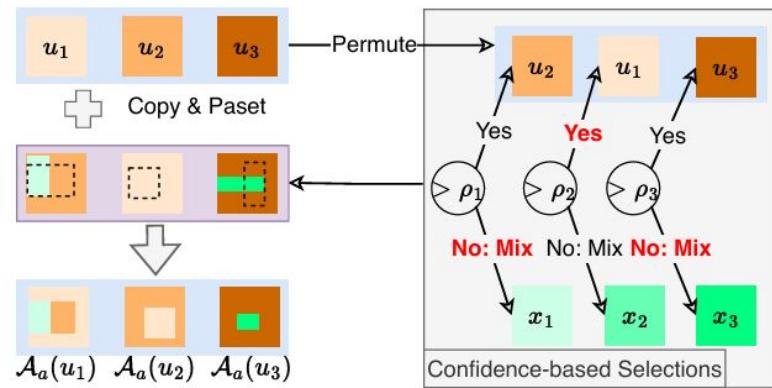
Giới thiệu

- AugSeg tập trung vào sự đơn giản và hiệu quả của việc tăng cường dữ liệu thay vì làm phức tạp hóa kiến trúc mạng và quy trình huấn luyện, nhưng vẫn đạt hiệu suất ngang với các mô hình SOTA khác.
- Cơ chế:
 - Hoạt động giữa trên khuôn khổ Teacher-Student tiêu chuẩn với hai thành phần chính được huấn luyện song song, thường là các mạng pretrained Resnet101, ViT,...
 - Sử dụng các chiến lược tăng cường dữ liệu \mathcal{A}_g , \mathcal{A}_r , \mathcal{A}_a thay vì làm phức tạp các mô hình backbone.



Mục tiêu

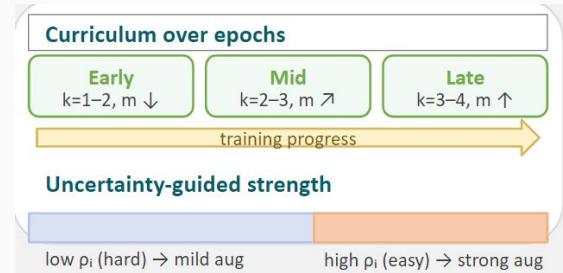
- Hạn chế của \mathcal{A}_r gốc:
 - Mức độ biến đổi (magnitude) **cố định**, không thay đổi theo epoch
 - Không thích ứng với độ tự tin ρ_i của pseudo-label từng ảnh → dễ làm nhiễu ảnh khó, chưa tận dụng tốt ảnh dễ.
- Mục tiêu của nghiên cứu:
 - Xây dựng cơ chế \mathcal{A}_r mới giúp tăng hiệu quả augmentation trên ảnh unlabeled trong SSS.
 - Cải thiện hiệu năng mIoU trên VOC/Cityscapes so với AugSeg gốc.



Nội dung và Phương pháp

- **Độ đo:** $mIoU = \frac{1}{N_c} \sum_{k=1}^{N_c} IoU_k$
- **Phương pháp cải thiện A_r đề xuất:**
 - **Xây dựng Curriculum A_r :**
 - Chia huấn luyện thành các giai đoạn (early–mid–late).
 - Tăng dần số phép k và biên độ magnitude theo epoch.
 - **Thiết kế Uncertainty-guided A_r :**
 - Tính độ tự tin ρ_i từ pseudo-label của teacher
 - Ảnh có ρ_i cao → áp dụng augmentation mạnh. Ảnh có ρ_i thấp → giảm cường độ augmentation để tránh làm nhiễu
 - **Kết hợp hai cơ chế thành CAUG- A_r (Curriculum + Uncertainty Guided A_r):** A_r được điều chỉnh theo thời gian huấn luyện và độ khó của từng ảnh.
 - Tích hợp vào pipeline AugSeg - với dữ liệu chưa được gán nhãn: $u \rightarrow A_g \rightarrow A_a \rightarrow A_r^{\text{proposed}} \rightarrow \text{Student}$

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



Kết quả dự kiến

- **Cải thiện mIoU** trên VOC 2012 và Cityscapes so với AugSeg gốc nhờ A_r thích ứng tốt hơn với quá trình huấn luyện.
- **Pseudo-label ổn định hơn** ở giai đoạn đầu (giảm nhiễu do augmentation quá mạnh).
- **Khả năng tổng quát hóa cao hơn** ở giai đoạn sau nhờ augmentation tăng dần theo curriculum.
- **Hiệu suất nhất quán hơn giữa các ảnh** nhờ Uncertainty-guided A_r , giúp ảnh khó được xử lý “nhẹ tay”, ảnh dễ được tăng cường mạnh.
- **Tăng độ ổn định của training**, giảm dao động loss và hạn chế confirmation bias.
- **Tính mở rộng tốt**: A_r mới có thể tích hợp vào các framework SSS khác ngoài AugSeg.

Tài liệu tham khảo

- Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, Jingdong Wang. Augmentation Matters: A Simple-yet-Effective Approach to Semi-supervised Semantic Segmentation. In CVPR, 2023.
- Xinting Hu, Li Jiang, Bernt Schiele. Training Vision Transformer for Semi-supervised Semantic Segmentation. In CVP2, 2024.
- Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, Yinghuan Shi. Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation. In CVPR, 2023.

**Cảm ơn thầy cô và
các bạn đã lắng nghe !**

Môn học: CS519 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS519.Q11

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM

