

# BÁO CÁO ĐỒ ÁN CUỐI

KỲ

Môn học

**CS519 - PHƯƠNG PHÁP LUẬN  
NGHIÊN CỨU KHOA HỌC**

Lớp học

**CS519.P11**

Giảng viên

**PGS.TS. LÊ ĐÌNH DUY**

Thời gian

**09/2025 - 12/2025**

----- Trang này có tình để trống -----

# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://www.youtube.com/watch?v=vxrN24QPeng>
- Link slides (dạng .pdf đặt trên Github của nhóm):  
<https://github.com/laikhanhhoang/CS519.Q11.KHTN/blob/main/Semi-Supervise%20Semantic%20Segmentation.pdf>
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none"><li>• Họ và Tên: Trần Vạn Tân</li><li>• MSSV: 23521407</li></ul> 	<ul style="list-style-type: none"><li>• Lớp: CS519.P11</li><li>• Tự đánh giá (điểm tổng kết môn): 9.0/10</li><li>• Số buổi vắng: 0</li><li>• Số câu hỏi QT cá nhân: 4</li><li>• Số câu hỏi QT của cả nhóm: 2</li><li>• Link Github: <a href="https://github.com/laikhanhhoang/CS519.Q11/">https://github.com/laikhanhhoang/CS519.Q11/</a></li><li>• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none"><li>○ Lên ý tưởng tối ưu bài báo(bài báo nghiên cứu hiện tại)</li><li>○ Làm slides báo cáo</li><li>○ Làm poster</li></ul></li></ul>
--	--

- Họ và Tên: Hoàng Sơn Kim
- MSSV: 12345



- Lớp: CS519.P11
- Tự đánh giá (điểm tổng kết môn): 9.0/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 4
- Số câu hỏi QT của cả nhóm: 2
- Link Github:  
<https://github.com/laikhanhhoang/CS519.Q11/>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
  - Làm docs báo cáo
  - Làm video YouTube

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

CAUG-A\_R: TĂNG CƯỜNG DỮ LIỆU THEO GIÁO TRÌNH VÀ ĐỘ BẤT ĐỊNH  
CHO PHÂN ĐOẠN NGỮ NGHĨA BÁN GIÁM SÁT

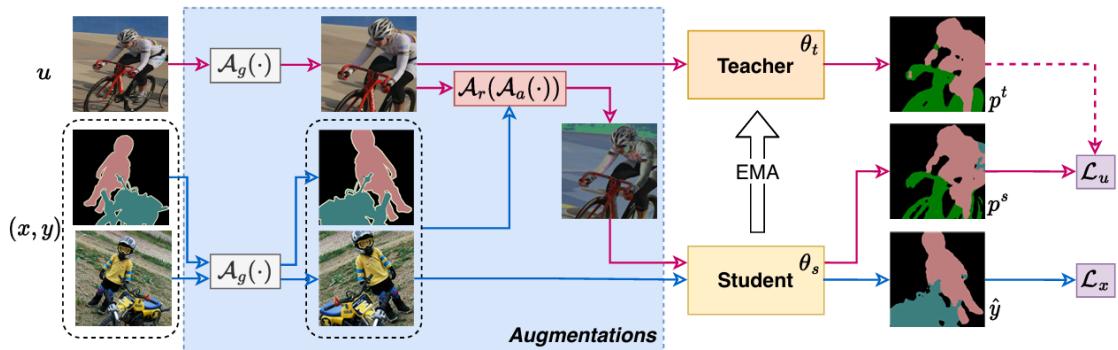
## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CAUG-A\_R: CURRICULUM AND UNCERTAINTY-GUIDED AUGMENTATION  
FOR SEMI-SUPERVISED SEMANTIC SEGMENTATION

## TÓM TẮT (Tối đa 400 từ)

Phân đoạn ngữ nghĩa (semantic segmentation) là bài toán gán nhãn cho từng pixel để tạo mặt nạ (mask) cùng kích thước ảnh. Tuy nhiên, gán nhãn mức pixel rất tốn kém nên trong thực tế thường chỉ có ít ảnh labeled và nhiều ảnh unlabeled.

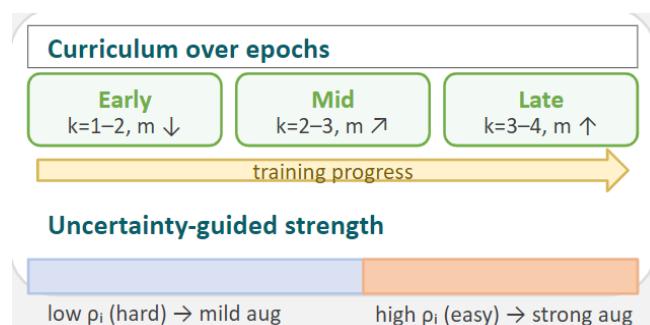
Semi-supervised Semantic Segmentation (SSS) khai thác đồng thời hai nguồn dữ liệu này để đạt hiệu năng gần với huấn luyện có giám sát hoàn toàn.



Trong công trình “Augmentation Matters” (AugSeg), Zhao et al. đề xuất khung Teacher–Student đơn giản tập trung vào perturbation dữ liệu, gồm chuỗi augmentation cho nhánh unlabeled:  $\mathcal{A}_g$  (hình học yếu),  $\mathcal{A}_r$  (tăng cường cường độ/intensity) và  $\mathcal{A}_a$  (tiêm nhãn thích ứng). Tuy vậy,  $\mathcal{A}_r$  trong AugSeg sử dụng các

siêu tham số (số phép biến đổi và mức độ biến đổi) gần như cố định trong suốt quá trình huấn luyện, chưa phản ánh (i) tiến trình học theo thời gian và (ii) độ tự tin khác nhau của pseudo-label trên từng ảnh. Điều này có thể gây nhiễu ở giai đoạn đầu (teacher chưa ổn định) và làm giảm khả năng khai thác các ảnh “dễ”.

Đồ án đề xuất CAUG- $\mathcal{A}_r$  (Curriculum kết hợp Uncertainty-guided) để thay thế  $\mathcal{A}_r$  cho dữ liệu unlabeled. Thành phần Curriculum chia giai đoạn early–mid–late, tăng dần số phép biến đổi  $k$  và magnitude theo epoch. Thành phần Uncertainty-guided ước lượng độ tự tin  $\rho_i$  từ pseudo-label của teacher:  $\rho_i$  cao (ảnh dễ) thì augmentation mạnh hơn;  $\rho_i$  thấp (ảnh khó) thì augmentation nhẹ hơn nhằm hạn chế khuếch đại sai nhầm. CAUG- $\mathcal{A}_r$  được tích hợp trực tiếp vào pipeline AugSeg mà không cần thêm mạng phụ.



Hiệu quả được đánh giá trên Pascal VOC 2012 và Cityscapes theo chuẩn mIoU, so sánh với AugSeg gốc và phân tích ablation cho từng thành phần. Kỳ vọng CAUG- $\mathcal{A}_r$  giúp giảm nhiễu pseudo-label ở giai đoạn đầu và tăng tổng quát hóa giai đoạn sau, từ đó cải thiện mIoU trên cả hai benchmark.

## GIỚI THIỆU (Tối đa 1 trang A4)

Semantic Segmentation gán nhãn cho từng pixel của ảnh, vì vậy đầu ra là một mask cùng kích thước ảnh đầu vào. Bài toán này là nền tảng cho nhiều ứng dụng như xe tự hành, robot, và phân tích ảnh y tế. Khó khăn lớn nhất nằm ở chi phí gán nhãn mức

pixel: việc tạo mask chính xác đòi hỏi nhiều thời gian và nhân lực, dẫn đến tình huống “ít labeled – nhiều unlabeled” trong dữ liệu thực tế.

Semi-supervised Semantic Segmentation (SSS) giải quyết bài toán bằng cách huấn luyện từ tập nhỏ dữ liệu có nhãn D\_L và tập lớn dữ liệu không nhãn D\_U. Một hướng tiếp cận phổ biến là Teacher–Student: teacher sinh pseudo-label cho dữ liệu unlabeled, student học từ pseudo-label kết hợp loss có nhãn. Để giảm overfitting và tăng tính bất biến, augmentation cho nhánh unlabeled (đặc biệt strong augmentation) đóng vai trò quyết định.

AugSeg (Zhao et al., CVPR 2023) nhấn mạnh “đơn giản nhưng hiệu quả” bằng cách tập trung vào thiết kế augmentation. Trong pipeline gốc,  $\mathcal{A}_r$  là intensity-based augmentation kiểu RandAugment: chọn ngẫu nhiên k phép biến đổi và sampling magnitude để tạo nhiễu mạnh. Tuy nhiên, cường độ augmentation  $\mathcal{A}_r$  trong thực tế vẫn thường được đặt cố định (không thay đổi theo epoch) và áp dụng gần như đồng đều cho mọi ảnh unlabeled. Điều này bỏ qua hai quan sát: (1) teacher tốt dần theo thời gian nên mức regularization có thể tăng dần; (2) ảnh unlabeled có độ khó khác nhau, thể hiện qua độ tự tin  $p_i$  của pseudo-label, vì vậy nên điều chỉnh augmentation theo từng ảnh.

Đò án đặt câu hỏi nghiên cứu: liệu một cơ chế điều chỉnh augmentation đơn giản, vừa theo tiến trình huấn luyện (curriculum) vừa theo độ tin cậy pseudo-label (uncertainty-guided), có thể cải thiện mIoU so với AugSeg gốc trên Pascal VOC 2012 và Cityscapes hay không?

- Đầu vào của bài toán:
  1. Tập ảnh labeled kèm mask
  2. Tập ảnh unlabeled.
- Đầu ra kỳ vọng là một mô hình phân đoạn ngữ nghĩa được huấn luyện theo khung SSS cùng module CAUG- $\mathcal{A}_r$  có thể tích hợp “plug-and-play” vào pipeline Teacher–Student.

## MỤC TIÊU (Viết trong vòng 3 mục tiêu)

1. Tái hiện và xây dựng baseline AugSeg trong thiết lập Teacher–Student cho SSS; chạy thử nghiệm trên Pascal VOC 2012 và Cityscapes theo các tỉ lệ labeled phổ biến (ví dụ 1/8, 1/4).
2. Thiết kế và cài đặt CAUG- $\mathcal{A}_r$ :
  - a. Curriculum cho k và magnitude theo epoch (early–mid–late);
  - b. Uncertainty-guided điều chỉnh strength theo độ tự tin  $\rho_i$  của pseudo-label từng ảnh.
  - c. Đánh giá định lượng và định tính: so sánh mIoU với AugSeg gốc, thực hiện ablation cho từng thành phần và phân tích độ ổn định huấn luyện.
3. Đánh giá và kiểm định: Thực nghiệm trên bộ dữ liệu Pascal VOC 2012 và Cityscapes, so sánh với các mô hình SOTA khác ở thời điểm hiện tại

## NỘI DUNG VÀ PHƯƠNG PHÁP

1. Khảo cứu và tái hiện AugSeg
  - Đọc và tổng hợp cơ chế Teacher–Student, EMA teacher, supervised loss  $\mathcal{L}_x$  và unsupervised consistency loss  $\mathcal{L}_u$ .
  - Chuẩn hóa pipeline augmentation gốc  $\mathcal{A}_g-\mathcal{A}_r-\mathcal{A}_a$  cho dữ liệu unlabeled.
  - Thiết lập thực nghiệm trên Pascal VOC 2012 và Cityscapes; đánh giá bằng mIoU.

## 2. Đè xuất CAUG- $\mathcal{A}_r$ (Curriculum + Uncertainty-guided)

### 2.1. Curriculum $\mathcal{A}_r$ (theo epoch)

- Chia tổng số epoch E thành 3 giai đoạn: early (0,E/3), mid (E/3,2E/3), late (2E/3,E).
- Ở mỗi giai đoạn, tăng dần:
  - (i) số phép biến đổi k (hoặc k\_max)
  - (ii) magnitude tối đa M của các phép intensity-based augmentation.
- Ví dụ (sẽ hiệu chỉnh khi chạy thực nghiệm): early ( $k \leq 2$ , M nhỏ), mid ( $k \leq 3$ , M trung bình), late ( $k \leq 4$ , M lớn).

### 2.2. Uncertainty-guided $\mathcal{A}_r$ (theo từng ảnh)

- Với ảnh unlabeled  $u_i$ , teacher dự đoán phân phối xác suất  $p_i$ ; tính độ tự tin  $\rho_i$  (ví dụ: trung bình max-probability trên các pixel được pseudo-label).
- Thiết kế hàm ánh xạ strength:  $\rho_i$  cao  $\rightarrow (k_i, M_i)$  lớn hơn;  $\rho_i$  thấp  $\rightarrow (k_i, M_i)$  nhỏ hơn.
- Mục tiêu là “không làm hỏng” ảnh khó ở giai đoạn teacher còn yếu, đồng thời tăng regularization cho ảnh dễ.

### 2.3. Tích hợp CAUG- $\mathcal{A}_r$ vào pipeline Teacher–Student

- Nhánh weak:  $u^w = \mathcal{A}_g(u)$ ; teacher sinh pseudo-label  $\hat{y} = \text{argmax } f_T(u^w)$ .
- Nhánh strong:  $u^s = \mathcal{A}_a(\text{CAUG-}\mathcal{A}_r(u^w, \rho_i, \text{epoch}))$ .
- Student tối ưu  $\mathcal{L} = \mathcal{L}_x + \lambda \cdot \mathcal{L}_u(u^s, \hat{y})$ ; teacher cập nhật bằng EMA từ student.

### 3. Thực nghiệm và đánh giá

- - Datasets: Pascal VOC 2012 và Cityscapes theo chuẩn split của SSS; tỉ lệ labeled 1/8 và 1/4 (và/hoặc theo cấu hình gốc của AugSeg).
- - So sánh: phiên bản AugSeg với thuật toán cải tiến CAUG- $\mathcal{A}_r$  với các model SOTA khác.
- - Chỉ số: mIoU trên tập validation/test; phân tích độ ổn định (loss curve, phân bố  $\rho_i$  theo epoch).

## KẾT QUẢ MONG ĐỢI

- Một phiên bản AugSeg tái hiện được (baseline) kèm hướng dẫn chạy và log thực nghiệm.
- Module CAUG- $\mathcal{A}_r$  dạng “plug-and-play” (không thêm network phụ), có thể bật/tắt từng thành phần Curriculum và Uncertainty-guided.
- Kết quả thực nghiệm trên Pascal VOC 2012 và Cityscapes: kỳ vọng mIoU tăng so với AugSeg gốc trong cùng thiết lập (backbone, split, số epoch).
- Phân tích ablation chứng minh đóng góp của từng thành phần, kèm giải thích cơ chế giảm nhiễu pseudo-label giai đoạn đầu và tăng tổng quát hóa giai đoạn sau.

## TÀI LIỆU THAM KHẢO (Định dạng DBLP)

[1]. Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, Jingdong Wang:

Augmentation Matters: A Simple-Yet-Effective Approach to Semi-Supervised Semantic Segmentation. CVPR 2023: 11350-11359.

[2]. Ekin Dogus Cubuk, Barret Zoph, Jonathon Shlens, Quoc Le: RandAugment:

Practical Automated Data Augmentation with a Reduced Search Space. NeurIPS 2020.

[3]. Yoshua Bengio, Jérôme Louradour, Ronan Collobert, Jason Weston: Curriculum

learning. ICML 2009: 41-48.

[4]. Antti Tarvainen, Harri Valpola: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. NIPS 2017: 1195-1204.

[5]. Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, Yang Gao: ST++: Make Self-training Work Better for Semi-supervised Semantic Segmentation. CVPR 2022: 4258-4267.

[6]. Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, Andrew Zisserman: The Pascal Visual Object Classes (VOC) Challenge. Int. J. Comput. Vis. 88(2): 303-338 (2010).

[7]. Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, Bernt Schiele: The Cityscapes Dataset for Semantic Urban Scene Understanding. CVPR 2016: 3213-3223.

----- Trang này có tình để trống - Các nhóm copy & paste bài làm  
của mình vào trang tiếp theo -----

