

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):

<https://youtu.be/mxq5nPe2m6s>

- Link slides (dạng .pdf đặt trên Github của nhóm):

https://github.com/KhaLee2307/CS519/blob/main/STR_GDA_slides.pdf

- Họ và Tên: Lê Nhật Kha

- MSSV: 20520208

- Lớp: KHTN2020



- Lớp: CS519.O11

- Tự đánh giá (điểm tổng kết môn): 10/10

- Số buổi vắng: 1

- Số câu hỏi QT cá nhân: 3

- Số câu hỏi QT của cả nhóm: 15

- Link Github:

<https://github.com/mynameuit/CS519.O11/>

- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:

- Chọn đề tài và xác định bài toán
- Tìm hiểu nội dung tài liệu liên quan
- Quan sát problem và đặt ra hypothesis
- Thiết kế các thực nghiệm
- Lập kế hoạch để phát triển dự án
- Viết proposal
- Làm slides
- Thuyết trình và quay video

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

THÍCH ỨNG MIỀN TĂNG TIẾN TRONG BÀI TOÁN NHẬN DIỆN VĂN BẢN

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

GRADUAL DOMAIN ADAPTATION IN SCENE TEXT RECOGNITION

TÓM TẮT (Tối đa 400 từ)

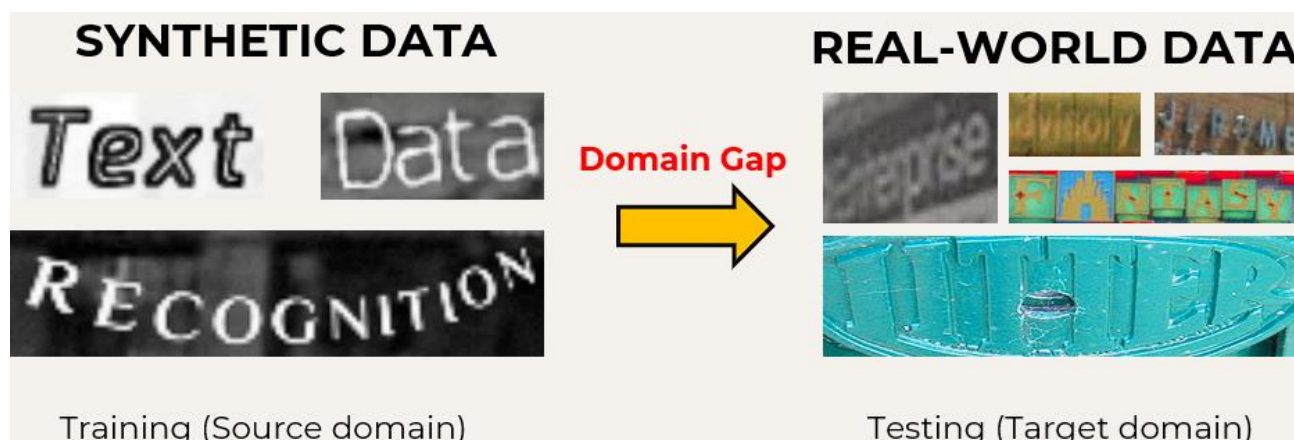
Nhận diện văn bản môi trường tự nhiên (**STR**) là một bài toán quan trọng trong lĩnh vực Computer Vision. Các mô hình SOTA hiện nay yêu cầu số lượng lớn dữ liệu có gán nhãn để huấn luyện. Tuy nhiên việc thu thập dữ liệu gán nhãn đòi hỏi chi phí cao và tốn thời gian. Do đó, các tác giả ưu tiên sử dụng dữ liệu synthetic (dữ liệu được tạo tự động) để huấn luyện mô hình. *Baek và các cộng sự* [4] đã chỉ ra rằng có sự khác biệt lớn giữa dữ liệu real-world và dữ liệu synthetic (được gọi là *domain gap*). Để giải quyết vấn đề này, một số tác giả đã đề xuất các phương pháp sử dụng kỹ thuật

Unsupervised Domain Adaptation (UDA), với mục tiêu giúp cho mô hình được huấn luyện trên source domain (*synthetic data*) có thể hoạt động tốt trên target domain (*real-world data*). Hướng tiếp cận phổ biến của UDA là self-training với dữ liệu không có nhãn. Áp dụng vào bài toán STR, UDA có thể tận dụng tốt dữ liệu synthetic có nhãn và dữ liệu real-world không có nhãn. Tuy nhiên hạn chế của phương pháp là khi domain gap giữa source domain và target domain ngày càng lớn, độ hiệu quả của UDA càng giảm, theo *Kumar* [1]. Do đó để giải quyết vấn đề này, nhóm quyết định tìm hiểu phương pháp **Gradual Domain Adaptation (GDA)** với ý tưởng chính là chia nhỏ domain gap và tiến hành progressive self-training.

GIỚI THIỆU (Tối đa 1 trang A4)

Nhận diện văn bản trong môi trường tự nhiên thường chia làm 2 nhiệm vụ chính: xác định vùng chứa văn bản, được gọi là Scene Text Detection (STD) và nhận diện nội dung văn bản chứa trong vùng đó, gọi là Scene Text Recognition (STR). Trong đồ án này, nhóm tập trung giải quyết bài toán STR với input là word box và output là text. Vì là ảnh chụp trong môi trường tự nhiên nên bài toán STR thường gặp phải rất nhiều thách thức về font chữ khó (chữ viết tay, chữ viết nghệ thuật), chữ bị cong, bị che khuất, hình ảnh bị mờ, có độ phân giải thấp, ...

Một thách thức lớn khác của bài toán là sự thiếu hụt dữ liệu real-world có gán nhãn (bởi chi phí gán nhãn lớn). Do đó các mô hình SOTA thường sử dụng dữ liệu synthetic (được tạo tự động) để huấn luyện. Tuy nhiên giữa synthetic và real-world data có khoảng cách phân phối dữ liệu lớn (gọi là *domain gap*).



Các công trình trước đây sử dụng kỹ thuật Domain Adaptation để giải quyết bài toán với hướng tiếp cận self-training (pseudo-label). Dữ liệu real-world không có nhãn sẽ được source-trained STR model dự đoán nhãn giả. Sau đó sẽ retrain lại mô hình kết hợp dữ liệu synthetic có nhãn và dữ liệu real-world không có nhãn. Hướng tiếp cận này sẽ cải thiện hiệu suất của mô hình mà không tốn thêm chi phí gán nhãn nào. Tuy nhiên khi domain gap càng lớn thì nhãn giả sẽ không chính xác dẫn tới UDA kém hiệu quả. Do đó nhóm nghiên cứu hướng tiếp cận Gradual Domain Adaptation.

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

- Đánh giá việc chia nhỏ domain gap ảnh hưởng như thế nào đến hiệu quả của phương pháp Unsupervised Domain Adaptation (UDA).
- Đề xuất phương pháp tổng quát nhằm cải thiện kỹ thuật UDA dựa trên hướng tiếp cận Gradual Domain Adaptation.
- Cải thiện hiệu suất của các mô hình STR (synthetic-trained model) thông qua phương pháp đề xuất.

NỘI DUNG VÀ PHƯƠNG PHÁP

Nội dung:

- Gradual domain adaptation là phương pháp kết hợp nhiều bước UDA.
- Thay vì self-training một lần duy nhất cho toàn bộ dữ liệu target domain, GDA sẽ chia dữ liệu thành các nhóm sub-domain sao cho domain gap của từng nhóm so với source domain có xu hướng *tăng dần*.
- Sau khi đã có các nhóm sub-domain được chia, tiến hành progressive self-training sử dụng pseudo-label để retrain lại mô hình STR.
- Với phương pháp này, mô hình sẽ thích ứng dần từ source domain sang target domain và cải thiện dần performance của nó.
- Phương pháp có thể được xem là progressive self-training từ dễ đến khó.

Phương pháp:

- Với hypothesis là “chia nhỏ domain gap sẽ giúp domain adaptation hiệu quả hơn”, nhóm tiến hành thiết kế các thực nghiệm chia nhỏ dữ liệu để self-training. Đầu tiên là chia nhỏ theo một cách ngẫu nhiên, sau đó là chia nhỏ nhưng có tính thứ tự (với các phương pháp sắp xếp tự đề xuất). So sánh kết quả thực nghiệm với nhau để kiểm chứng hypothesis và cả phương pháp đề xuất.
- Áp dụng phương pháp lên bài toán cụ thể là Scene Text Recognition.

KẾT QUẢ MONG ĐỢI

- Đề xuất phương pháp cải thiện độ hiệu quả của kỹ thuật UDA.
- Phương pháp đề xuất tăng hiệu suất của các mô hình STR baseline mà không tốn thêm chi phí gán nhãn.
- Đánh giá trên các tập scene-text benchmark, phương pháp chứng tỏ độ hiệu quả hơn các phương pháp UDA khác [5], và đạt được *state-of-the-art* domain adaptation cho bài toán nhận diện văn bản (STR).
- Tổng hợp, viết paper và được chấp nhận đăng tại hội nghị lớn như CVPR2024

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

- [1]. Kumar, A., Ma, T. and Liang, P., 2020, November. Understanding self-training for gradual domain adaptation. In *International Conference on Machine Learning* (pp. 5468-5479). PMLR.
- [2]. Chen, H.Y. and Chao, W.L., 2021. Gradual domain adaptation without indexed intermediate domains. *Advances in Neural Information Processing Systems*, 34, pp.8201-8214.
- [3]. Wang, H., Li, B. and Zhao, H., 2022, June. Understanding gradual domain adaptation: Improved analysis, optimal path and beyond. In *International Conference on Machine Learning* (pp. 22784-22801). PMLR.
- [4]. Baek, J., Matsui, Y. and Aizawa, K., 2021. What if we only use real datasets for scene text recognition? toward scene text recognition with fewer labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3113-3122).
- [5] Tien, H.T. and Ngo, T.D., 2023, October. Unsupervised Domain Adaptation with Imbalanced Character Distribution for Scene Text Recognition. In *2023 IEEE International Conference on Image Processing (ICIP)* (pp. 3493-3497). IEEE.