

# NGHIÊN CỨU PHƯƠNG PHÁP NHẬN DẠNG VĂN BẢN TRONG ẢNH CÓ CHẤT LƯỢNG THẤP

Lê Xuân Tùng - 240101031

# Tóm tắt

- Lớp: CS2205.CH183
- Link Github của nhóm:  
[https://github.com/coderaohong/low\\_quality\\_STR](https://github.com/coderaohong/low_quality_STR)
- Link YouTube video:  
<https://www.youtube.com/watch?v=3ngyIVMAKro>
- Ảnh + Họ và Tên:

Lê Xuân Tùng - 240101031



# Giới thiệu

Nhận dạng văn bản trong ảnh ngoại cảnh (Scene Text Recognition - STR) là một bài toán quan trọng trong lĩnh vực *thị giác máy tính*. Khác với nhận dạng ký tự quang học (OCR) truyền thống vốn hoạt động trên văn bản được quét từ tài liệu, STR đòi hỏi mô hình có khả năng nhận diện văn bản xuất hiện trong các bối cảnh thực tế, nơi điều kiện ánh sáng, góc nhìn, phong nền và chất lượng ảnh có thể thay đổi đáng kể.



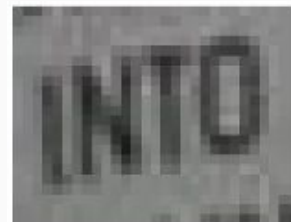
*Hình 1. Minh họa định nghĩa bài toán*

# Giới thiệu

Văn bản trong ảnh có chất lượng thấp:

→ Bao gồm các đặc điểm:

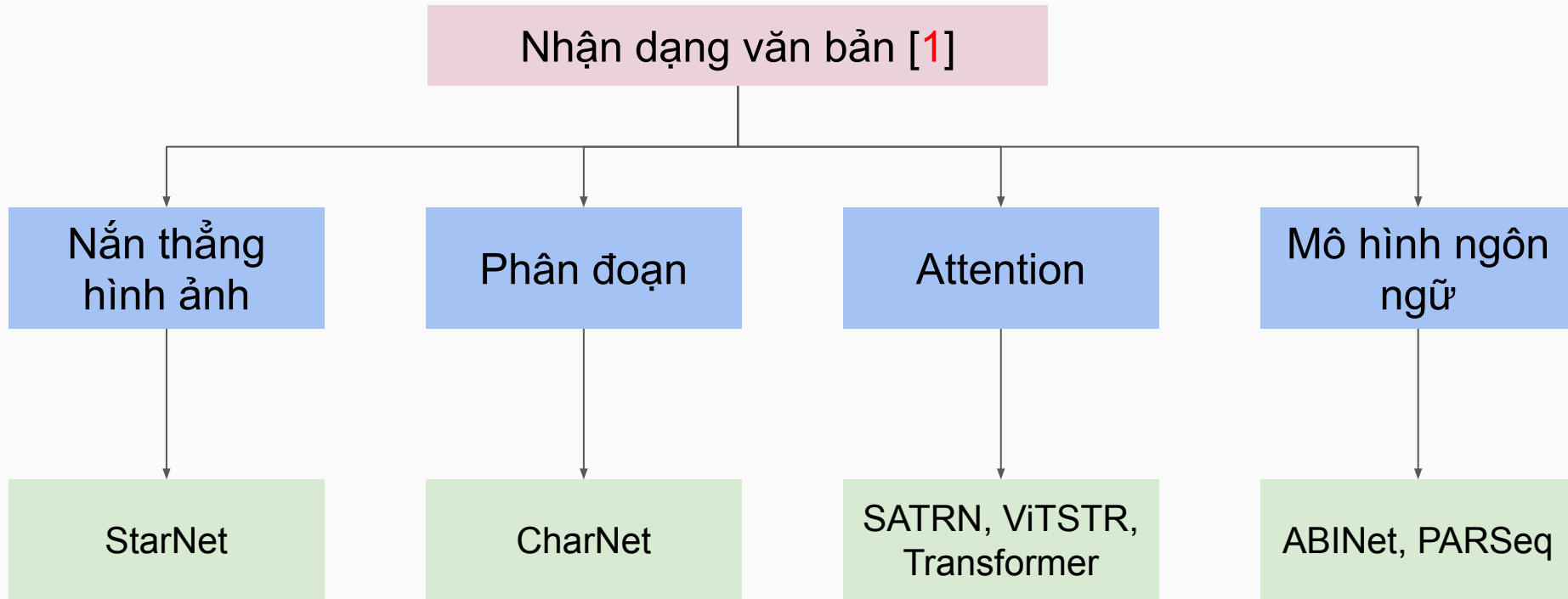
- ◆ Ảnh nhỏ, độ phân giải thấp
- ◆ Nhiều trong quá trình trình chụp
- ◆ Văn bản gốc có nhiều khiếm khuyết



# Mục tiêu

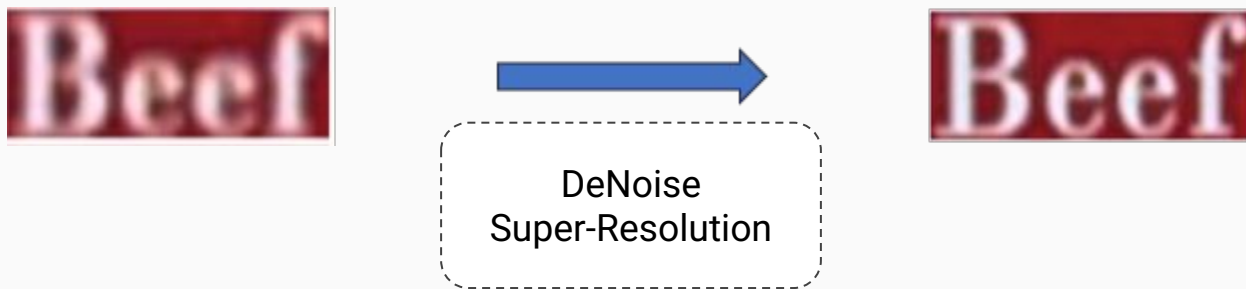
- **Xây dựng tập dữ liệu** dành riêng cho bài toán nhận dạng văn bản trong ảnh chất lượng thấp.
- **Xây dựng mô hình** nhận dạng văn bản tích hợp module tiền xử lý với các kiến trúc tiên tiến.
- **Kết quả đánh giá và so sánh** của mô hình vừa xây dựng được trên các tập dữ liệu nhận dạng văn bản với các phương pháp tiên tiến hiện nay.

# Nội dung và Phương pháp



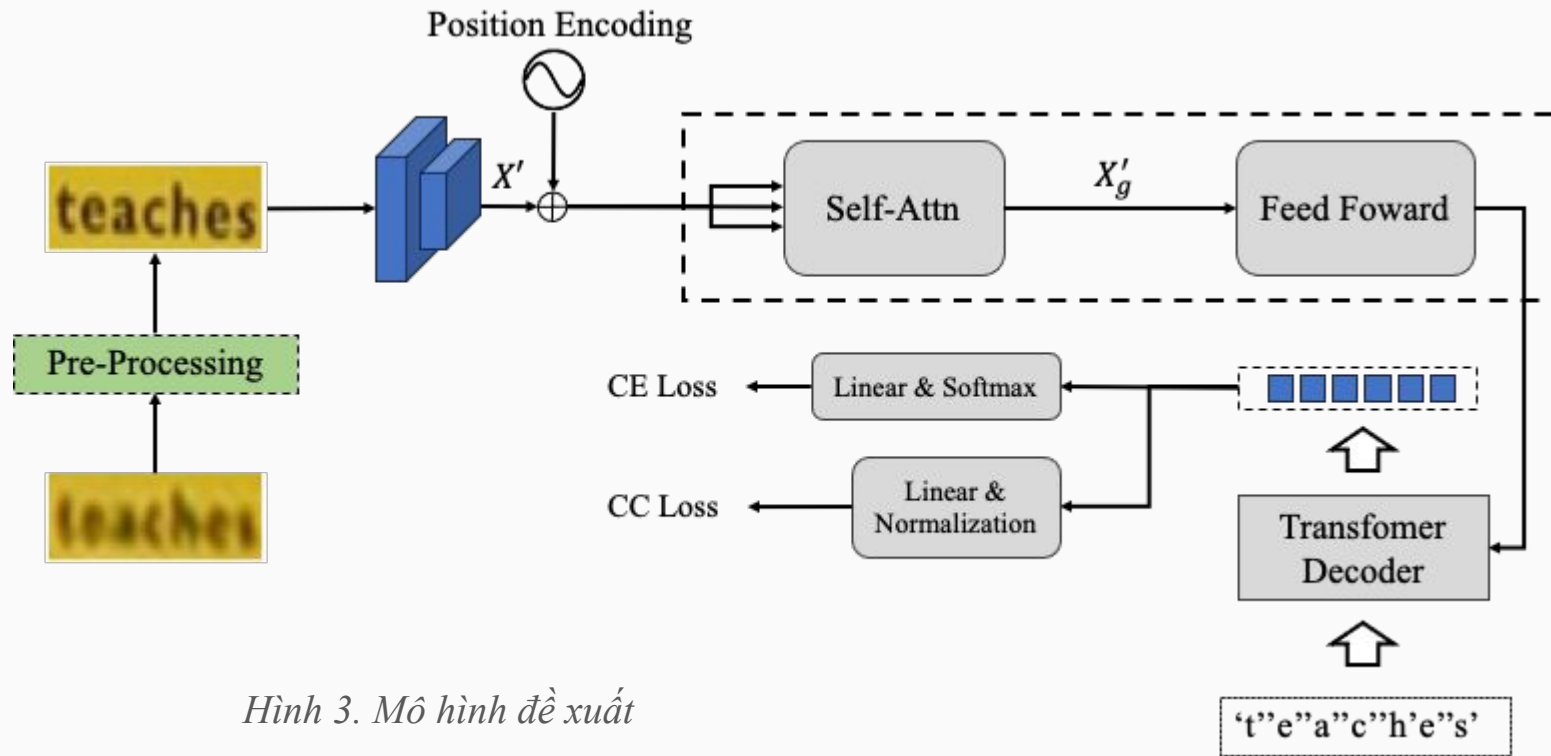
# Nội dung và Phương pháp

Module tiền xử lý giúp **cải thiện chất lượng ảnh** trước khi đưa vào mô hình nhận dạng, giảm tác động của nhiễu, tăng độ sắc nét của văn bản, giúp mô hình nhận dạng chính xác hơn.



*Hình 2. Minh họa đầu vào đầu ra của module tiền xử lý*

# Nội dung và Phương pháp



Hình 3. Mô hình đề xuất



# Kết quả dự kiến

- Có được một **bộ dữ liệu mới** dành riêng cho vấn đề nhận dạng văn bản trong ảnh có chất lượng thấp.
- **Một mô hình nhận dạng văn bản mới** có tích hợp module tiền xử lý ảnh chất lượng thấp
- Mô hình mới có **độ chính xác vượt trội** trên bộ dữ liệu ảnh có chất lượng thấp và cho kết quả tốt trên các bộ dữ liệu chuẩn khi so sánh với các mô hình tiên tiến hiện nay

# Tài liệu tham khảo

- [1]. Xudong Xie, Ling Fu, Zhifei Zhang, Zhaowen Wang, & Xiang Bai (2022). Toward Understanding WordArt: Corner-Guided Transformer for Scene Text Recognition. In Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXVIII (pp. 303–321). Springer.
- [2]. Yoshua Bengio, Li Yao, Guillaume Alain, & Pascal Vincent (2013). Generalized Denoising Auto-Encoders as Generative Models. In Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States (pp. 899–907).
- [3]. Baoguang Shi, Xiang Bai, & Cong Yao (2017). An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. IEEE Trans. Pattern Anal. Mach. Intell., 39(11), 2298–2304.
- [4]. Rowel Atienza (2021). Vision Transformer for Fast and Efficient Scene Text Recognition. In 16th International Conference on Document Analysis and Recognition, ICDAR 2021, Lausanne, Switzerland, September 5-10, 2021, Proceedings, Part I (pp. 319–334). Springer.
- [5]. Wenjia Wang, Enze Xie, Xuebo Liu, Wenhai Wang, Ding Liang, Chunhua Shen, & Xiang Bai (2020). Scene Text Image Super-Resolution in the Wild. In Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part X (pp. 650–666). Springer.