


THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
<https://youtu.be/tzBKA6vTt-s?si=FIHGxroBqyEwAWNl>
- Link slides (dạng .pdf đặt trên Github của nhóm):
<https://github.com/imhwy/CS519.O11/blob/main/slides.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">● Họ và Tên: Ngô Phúc Danh● MSSV: 21521924 	<ul style="list-style-type: none">● Lớp: CS519.O11● Tự đánh giá (điểm tổng kết môn): 9.5/10● Số buổi vắng: 1● Số câu hỏi QT cá nhân: 11● Link Github: https://github.com/imhwy/CS519.O11
--	---

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

PHÂN ĐOẠN ĐỐI TƯỢNG NGUY TRANG BẰNG CÁC MÔ HÌNH HỌC SÂU

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CAMOUFLAGED INSTANCE SEGMENTATION USING DEEP LEARNING MODELS

TÓM TẮT (Tối đa 400 từ)

Hiện nay, bài toán phân đoạn các vùng ảnh chi tiết đến từng đối tượng trong mỗi nhãn (instance segmentation) đã trở nên vô cùng phổ biến kể từ sau sự ra đời của mô hình FCN và U-Net. Thế nhưng vào những năm gần đây, một bài toán mới về phân đoạn đối tượng nguy trang đã trở thành một thách thức vô cùng lớn đối với các mô hình segmentation bởi đặc điểm giống nhau giữa đối tượng nguy trang và môi trường xung quanh dẫn tới khó khăn trong việc tái tạo lại segmentation mask. Để giải quyết vấn đề trên, đề tài này tập trung vào việc nghiên cứu sự hiệu quả giữa các phương pháp tiếp cận state-of-the-art hiện nay trong bài toán Camouflaged Instance Segmentation (CIS) là one-stage và two-stage. Phương pháp one-stage tập trung vào việc thực hiện phân đoạn và dự đoán đối tượng nguy trang trong một lượt duy nhất. Ngược lại, phương pháp two-stage phân thành hai giai đoạn: giai đoạn đầu tiên để xác định đối tượng và giai đoạn thứ hai để thực hiện phân đoạn chi tiết. Thông qua quá trình huấn luyện và đánh giá trên hai bộ dữ liệu COD10K và NC4K với độ đo mAP để so sánh hiệu quả của hai phương pháp tiếp cận trên. Nghiên cứu này sẽ không chỉ cung cấp cái nhìn sâu sắc về thách thức của bài toán CIS mà còn đề xuất và so sánh sự hiệu quả giữa phương pháp one-stage và two-stage. Kết quả của nghiên cứu có thể đóng góp vào sự hiểu biết chung về cách tiếp cận bài toán segmentation trong những điều kiện khó khăn như đối tượng nguy trang.

GIỚI THIỆU (Tối đa 1 trang A4)

Quan sát trong quá trình tiến hóa, khả năng nguy trang đã trở thành một chiến lược sinh tồn quan trọng đối với nhiều loài động vật trên khắp thế giới. Tính năng này giúp chúng tồn tại và phát triển trong môi trường tự nhiên, nơi mà khả năng nguy trang có thể cung cấp lợi thế lớn trong việc tránh bị săn mồi hoặc làm cho chúng trở nên khó nhận biết đối với kẻ săn

môi. Sự phát triển của khả năng nguy trang không chỉ là một kết quả của sự chọn lọc tự nhiên mà còn là sự thích ứng linh hoạt của các loài động vật với môi trường sống đa dạng và thay đổi. Con người cũng học hỏi từ tự nhiên khi biết áp dụng vào nhiều lĩnh vực và nổi bật nhất chính là quân sự.

Đối diện với thách thức làm thế nào để giải quyết bài toán (CIS) khi những mô hình segmentation truyền thống [1, 2, 3, 4] tỏ ra đuối sức trong việc phân đoạn đối tượng nguy trang và môi trường? Để giải quyết vấn đề trên, nghiên cứu này tập trung vào việc làm nổi bật các đặc điểm của đối tượng nguy trang từ đó phân đoạn chi tiết từng vùng ảnh của các đối tượng bằng cách tiếp cận one-stage [5, 6] (nhận thông tin quan trọng tại tất cả các vị trí) và two-stage [7, 8] (khử đi khả năng nguy trang của đối tượng sau đó tiến hành phân đoạn). Từ đó giúp xây dựng nên một hệ thống phân đoạn mạnh mẽ, khám phá và phân loại một cách chi tiết những thực thể ẩn mình tồn tại trong thế giới phong phú của môi trường tự nhiên. Điều này là một bước quan trọng trong việc hiểu rõ sâu sắc về sự liên kết phức tạp giữa các đối tượng và bối cảnh tự nhiên.

Cụ thể đầu vào và đầu ra của bài toán CIS:

Input: một bức ảnh chứa đối tượng nguy trang.

Output: segmentation mask mô tả vị trí và hình dạng của đối tượng nguy trang.

Minh họa:



input



output

Giải quyết bài toán Camouflaged Instance Segmentation chính là giải quyết các vấn đề về nhu cầu mở rộng tri thức của con người thông qua việc quan sát được tập tính và hành vi của các loài động vật có cơ chế nguy trang trong môi trường tự nhiên, phát hiện các loài động vật nguy hiểm ẩn mình tấn công con người hay trong quân sự là phát hiện kẻ thù đang nguy trang trên chiến trường.

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

- Nghiên cứu các mô hình dựa theo phương pháp one-stage và two-stage cho bài toán Camouflaged Instance Segmentation (CIS).
- So sánh sự hiệu quả về độ chính xác và tốc độ inference giữa các phương pháp. Từ đó lựa chọn ra phương pháp phù hợp với yêu cầu deployment.
- Xây dựng hệ thống demo để kiểm thử với những dữ liệu thực tế bên ngoài tự nhiên.

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Nội dung:

- Tìm hiểu về phương pháp tiếp cận one-stage với mô hình OSFormer [5].
- Tìm hiểu về phương pháp tiếp cận two-stage với mô hình DCNet [7].
- Tiến hành huấn luyện và đánh giá hai mô hình dựa trên các tiêu chí về độ chính xác và tốc độ inference.
- Xây dựng hệ thống ứng dụng trên dữ liệu thực tế.

Phương pháp:

- Đối với phương pháp one-stage, nghiên cứu mô hình OSFormer được phát triển theo hướng Transformer [9] bằng cách decompose thành các vấn đề con. Trong đó, tìm hiểu từng components của kiến trúc có nhiệm vụ xử lý đặc trưng, kết hợp các thông tin high và low levels để tái tạo segmentation mask. Tìm hiểu các multi-scale features được kết hợp với nhau thông qua một quá trình encoder-decoder sử dụng cross-attention. Tìm hiểu về quá trình giám sát các đặc trưng cạnh trong quá trình training.
- Đối với phương pháp two-stage, nghiên cứu mô hình DCNet được phát triển theo hai quá trình xử lý bằng cách decompose thành hai module. Tìm hiểu module đầu tiên đảm nhận nhiệm vụ khử đối tượng nguy trang ở nhiều tỉ lệ khác nhau và module còn lại đảm nhận nhiệm vụ tái tạo segmentation mask bằng cách kết hợp với output của module đầu tiên. Tìm hiểu cách các multi-scale features khác nhau kết hợp với nhau thông qua kiến trúc FPN [10]. Tìm hiểu các prototype được học để kết hợp với output của module đầu tiên nhằm tăng cường khả năng tái tạo segmentation mask.

- Cài đặt và tiến hành huấn luyện trên bộ dữ liệu COD10K.
- Tìm hiểu phương pháp và độ đo để đánh giá sự hiệu quả của các hướng tiếp cận trên phương diện về độ chính xác cũng như tốc độ inference trên bộ dữ liệu COD10K và NC4K.
- Tìm hiểu cách deploy một hệ thống có thể áp dụng vào dữ liệu thực tế.

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

- Phương pháp huấn luyện mô hình one-stage sử dụng kiến trúc Transformer kết hợp các multi-scale features giúp nâng cao sự chính xác trong việc phát hiện đối tượng nguy trang, đồng thời đáp ứng được tốc độ inference tốt.
- Phương pháp two-stage không những kết hợp các đặc trưng ở multi-scale mà còn sử dụng các prototype được đào tạo trước đó để tăng khả năng tái tạo segmentation mask, điều này mang lại sự chính xác cao nhưng sẽ đánh đổi lại tốc độ inference chậm hơn do cần xử lý hai quá trình.
- Báo cáo về phương pháp, kỹ thuật áp dụng, cách cài đặt, kết quả thực nghiệm và demo.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

- [1] Jonathan Long, Evan Shelhamer, Trevor Darrell: Fully convolutional networks for semantic segmentation. CVPR 2015: 3431-3440
- [2] Olaf Ronneberger, Philipp Fischer, Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI (3) 2015: 234-241
- [3] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross B. Girshick: Mask R-CNN. ICCV 2017: 2980-2988
- [4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40(4): 834-848 (2018)
- [5] Jialun Pei, Tianyang Cheng, Deng-Ping Fan, He Tang, Chuanbo Chen, Luc Van Gool: OSFormer: One-Stage Camouflaged Instance Segmentation with Transformers. ECCV (18) 2022: 19-37

- [6] Bo Dong, Jialun Pei, Rongrong Gao, Tian-Zhu Xiang, Shuo Wang, Huan Xiong: A Unified Query-based Paradigm for Camouflaged Instance Segmentation. ACM Multimedia 2023: 2131-2138
- [7] Naisong Luo, Yuwen Pan, Rui Sun, Tianzhu Zhang, Zhiwei Xiong, Feng Wu: Camouflaged Instance Segmentation via Explicit De-Camouflaging. CVPR 2023: 17918-17927
- [8] Tuan-Anh Vu, Duc Thanh Nguyen, Qing Guo, Binh-Son Hua, Nhat Minh Chung, Ivor W. Tsang, Sai-Kit Yeung: Leveraging Open-Vocabulary Diffusion to Camouflaged Instance Segmentation. CoRR abs/2312.17505 (2023)
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is All you Need. NIPS 2017: 5998-6008
- [10] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, Serge J. Belongie: Feature Pyramid Networks for Object Detection. CVPR 2017: 936-944