# ĐỀ CƯƠNG NGHIÊN CỨU

**TÊN ĐỀ TÀI (IN HOA)**

EMPIRICAL STUDY OF REPPOINTS-V2 WITH A DECOUPLED HEAD FOR OBJECT DETECTION IN AERIAL IMAGES

**TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**

THỰC NGHIỆM TINH CHỈNH MÔ HÌNH REPPOINT-V2 VỚI DECOUPLED HEAD TRÊN BÀI TOÁN PHÁT HIỆN ĐỐI TƯỢNG TRONG KHÔNG ẢNH

**TÓM TẮT** *(Tối đa 400 từ)*

Trong bài toán Phát hiện đối tượng của lĩnh vực Thị giác máy tính, các phương pháp anchor-free nổi lên như cách tiếp cận mới thay thế cho sự thống trị của anchor-based, đặc biệt với nhiệm vụ phát hiện phương tiện giao thông trong dữ liệu không ảnh [4,5]. Một trong số đó là thuật toán dựa trên phương thức biểu diễn đối tượng Reppoints (*Representative points).* Tận dụng hiệu quả của biểu diễn RepPoints tiền nhiệm, trước tiên chúng tôi lựa chọn phiên bản cải tiến ***RepPoints-v2***[1], đề xuất kết hợp phương pháp vertification và regression trong bước dự đoán định vị vật thể (*localization prediction*). Tiếp đó, chúng tôi mở rộng thực nghiệm cài đặt một số kiến trúc backbone, gồm: ResNeSt-50, ResNeSt-101, ResNeXt-50, ResNeXt-101 với mục tiêu tối ưu hoá hiệu suất và thời gian tính toán của mô hình. Cuối cùng, chúng tôi giới thiệu cách tiếp cận ***Decoupled Head*** (Đầu tách rời) trên mô hình RepPoints-v2, nhằm hỗ trợ giải quyết xung đột xảy ra do khác biệt bản chất của hai tác vụ classification và regression. Toàn bộ thực nghiệm của nghiên cứu sẽ được huấn luyện trên bộ dữ liệu không ảnh giao thông VisDrone-DET [2]. Ngoài ra, chúng tôi cũng xây dựng chương trình ứng dụng chạy trên nền web minh hoạ kết quả nghiên cứu được trở thành một sản phẩm ứng dụng thực tế.

**GIỚI THIỆU** *(Tối đa 1 trang A4)*

Những năm gần đây, phương tiện bay không người lái (UAV) (hay còn gọi là Drone) được ứng dụng rộng rãi trong nhiều lĩnh vực, đặc biệt là giải pháp cho các vấn đề của

giao thông đô thị. Việc tận dụng các hình ảnh từ trên không sẽ đem đến cái nhìn toàn cảnh về hiện trạng giao thông, nhằm hỗ trợ cho các hệ thống phân tích, giám sát giao thông. Đó chính là động lực chúng tôi lựa chọn tìm hiểu bài toán *Phát hiện đối tượng tham gia giao thông trong không ảnh*.

Các phương pháp phát hiện đối tượng được nghiên cứu và áp dụng hiện nay phần lớn đều ứng dụng anchor (anchor-based) xuất phát từ độ phổ biến của hộp giới hạn (bounding box). Tuy dù đơn giản và thuận tiện khi sử dụng, hộp giới hạn hai chiều bộc lộ nhiều hạn chế, như: 1/ Chỉ biểu diễn được phạm vi không gian cục bộ của đối tượng; 2/ Không phản ánh hình dạng, tư thế hay thông tin ngữ nghĩa quan trọng. Do đó, các mô hình không ứng dụng anchor (anchor-free) được xem là hướng nghiên cứu hứa hẹn, nhờ loại bỏ được lượng lớn siêu tham số của quá trình tính toán anchor mà vẫn đạt được hiệu suất cao tương đương các mô hình achor-based tiền nhiệm.

Mặt khác, trong các mô hình phát hiện đối tượng, sự mâu thuẫn giữa nhiệm vụ phân loại và hồi quy khi tính toán đồng thời là một vấn đề phổ biến. Nhằm giảm thiểu sự nhập nhằng giữa hai tác vụ này, khái niệm về Decoupled Head ra đời và được sử dụng rộng rãi trong nhiều mô hình một giai đoạn và hai giai đoạn, hỗ trợ cải thiện tốc độ hội tụ và hiệu suất mô hình.

Trong đề tài này, chúng tôi quyết định tìm hiểu và thực nghiệm mô hình phát hiện đối tượng anchor-free dựa trên phương thức biểu diễn cải tiến **RepPoints-v2** [1], đồng thời giới thiệu cách tiếp cận Decoupled Head trên mô hình RepPoints-v2, đánh giá hiệu suất mô hình trên bộ dữ liệu không ảnh VisDrone-DET [2].

a) Đầu vào

b) Đầu ra

Mô tả đầu ra đầu vào của bài toán được như sau:

- **Input:** Hình ảnh có chứa phương tiện giao thông được chụp từ thiết bị bay không người lái.
- **Output:** Vị trí và nhãn của từng đối tượng có trong hình ảnh được thể hiện qua hộp giới hạn tối tiểu.

**MỤC TIÊU**

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

- Đánh giá hiệu suất của phương thức biểu diễn đối tượng cải tiến- ***Reppoints-v2*** [1] đến kết quả của mô hình phát hiện đối tượng, so với phiên bản Reppoints tiền nhiệm và hộp giới hạn truyền thống.
- Tối ưu hiệu suất của mô hình RepPoints-v2 hiện có, bằng việc thực nghiệm tinh chỉnh backbone với một số kiến trúc state-of-the-art, gồm: ResNeSt-50[6], ResNeSt-101[6], ResNeXt-50[7], ResNeXt-101[7].
- Giới thiệu hướng tiếp cận Decoupled Head (Đầu tách rời) trên mô hình RepPoints-v2, tiến hành thực nghiệm trên bộ dữ liệu không ảnh giao thông VisDrone-DET[2].

**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 tổng quan về hướng phát triển các mô hình anchor-free với bài toán Phát hiện đối tượng, cũng như các phương thức biểu diễn đối tượng mới thay thế cho hộp giới giạn truyền thống.
- Nghiên cứu và đánh giá phương thức biểu diễn đối tượng mới Repoints-v2, với cải tiến kết hợp phương pháp vertification và regression trong bước dự đoán định vị vật thể.
- Nghiên cứu và áp dụng các kỹ thuật Decoupled Head (Đầu tách rời) vào mô hình anchor-free RepPoints-v2, đây là phương thức tách riêng tác vụ phân loại và tác vụ hồi quy thành các đầu khác nhau.

- Thực nghiệm cài đặt một số kiến trúc backbone ResNeSt-50, ResNeSt-101, ResNeXt-50, ResNeXt-101 vào mô hình Repoints-v2 nhằm tìm ra mô hình phù hợp nhất và tối ưu hoá hiệu suất của phương pháp.
- Huấn luyện mô hình Repoints-v2 tinh chỉnh trên bộ dữ liệu VisDrone-DET[2] để so sánh và đánh giá các kỹ thuật đã sử dụng.

**Phương pháp:**

- Khảo sát các công trình nghiên cứu mới nhất về mô hình phát hiện đối tượng anchor-free, cũng như các phương thức biểu diễn đối tượng mới kèm theo theo các tiêu chí điểm mạnh, điểm yếu, độ linh hoạt mô hình, kết quả đạt được.
- Nghiên cứu kiến trúc mô hình Repoints-v2, đặc biệt là cải tiến kết hợp phương pháp vertification và regression trong bước dự đoán định vị vật thể.
- Tìm hiểu về kỹ thuật Decoupled Head đã được cài đặt trên các mô hình một giai đoạn và hai giai đoạn trước đó; tiến hành triển khai cài đặt trên mô hình Repoints-v2.
- Tìm hiểu và khảo sát các bộ dữ liệu không ảnh thu thập từ drone hiện có, xem xét ưu nhược điểm, kết quả thực nghiệm trong các công trình mới nhất trên bộ dữ liệu không ảnh VisDrone-DET.
- Thực nghiệm mô hình Repoints-v2 với các kiến trúc backbone: ResNeSt-50, ResNeSt-101, ResNeXt-50, ResNext-101, so sánh và đánh giá ứng với từng trường hợp áp dụng kiến trúc.
- Xây dựng chương trình ứng dụng trên nền Web cho phép người dùng nhập đầu vào một hình ảnh giao thông (ưu tiên ảnh chụp từ trên cao) và xem đầu ra là các phương tiện giao thông được phát hiện và gán nhãn.

**KẾT QUẢ MONG ĐỢI**

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

Báo cáo kết quả thực nghiệm mô hình Repoints-v2 với bộ dữ liệu VisDrone-DET trên từng kiến trúc backbone: ResNeSt-50, ResNeSt-101, ResNeXt-50, ResNext-101;
đánh giá, so sánh các kiến trúc với nhau và với baseline ban đầu.

- Báo cáo về kỹ thuật áp dụng Decoupled Head trên mô hình Repoints-v2 mà chúng tôi triển khai. Kết quả thực nghiệm, đánh giá, so sánh phương pháp Decoupled head so với mô hình Coupled Head truyền thống.
- Chương trình minh họa phát hiện phương tiện giao thông trong hình ảnh.

**TÀI LIỆU THAM KHẢO** *(Định dạng DBLP)*

[1] Chen, Y., Zhang, Z., Cao, Y., Wang, L., Lin, S., & Hu, H. (2020). Reppoints v2: Verification meets regression for object detection. Advances in Neural Information Processing Systems, 33, 5621-5631.

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.

[3] Du, D., Zhu, P., Wen, L., Bian, X., Lin, H., Hu, Q., ... & Zhang, L. (2019). VisDrone-DET2019: The vision meets drone object detection in image challenge results. In Proceedings of the IEEE/CVF international conference on computer vision workshops (pp. 0-0).

[4] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

[5] Dai, L., Chen, H., Li, Y., Kong, C., Fan, Z., Lu, J., & Chen, X. (2022). TARDet: Two-stage Anchor-free Rotating Object Detector in Aerial Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4267-4275).

[6] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., ... & Smola, A. (2022). Resnest: Split-attention networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2736-2746).

[7] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).