# **THÔNG TIN CHUNG CỦA NHÓM**

* Link YouTube video của báo cáo (tối đa 5 phút):   
  *https://youtu.be/hKSkUQnPxd4*
* Link slides (dạng .pdf đặt trên Github của nhóm):   
  *(ví dụ: https://github.com/mynameuit/CS519.M1.KHCL/TenDeTai.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*

|  |  |
| --- | --- |
| * Họ và Tên: Nguyễn Đức Anh Phúc * MSSV: 20520276 | * Lớp: CS519.M11 * Tự đánh giá (điểm tổng kết môn): 9.5/10 * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: 10 * Số câu hỏi QT của cả nhóm: 3 * Link Github: [https://github.com/PhucNDA/CS519.N11](https://github.com/PhucNDA/HA-RDet) * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Đề xuất bài toán   + Lên ý tưởng cải tiến   + Viết chính phần phương pháp   + Đóng góp ý tưởng làm slide Powerpoint/youtube |
| * Họ và tên: Huỳnh Viết Tuấn Kiệt * MSSV: 20521494 | * Lớp: CS519.M11 * Tự đánh giá (điểm tổng kết môn): 9/10 * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: 10 * Số câu hỏi QT của cả nhóm: 3 * Link Github: [https://github.com/hvtkiet/CS519.N11](https://github.com/HiImKing1509/CS519.N11) * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Khảo sát các nghiên cứu liên quan   + Thử nghiệm với các module khác nhau và chạy thực nghiệm   + Viết chính báo cáo   + Đóng góp ý tưởng làm Slide/Youtube/Powerpoint |
| * Họ và tên: Trần Văn Lực * MSSV: 20521587 | * Lớp: CS519.M11 * Tự đánh giá (điểm tổng kết môn): 9/10 * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: 10 * Số câu hỏi QT của cả nhóm: 3 * Link Github: <https://github.com/VanlucCS/CS519.N11> * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Khảo sát các nghiên cứu liên quan   + Tìm kiếm và thu thập dữ liệu   + Làm chính slide/poster/edit video youtube   + Đóng góp ý tưởng làm báo cáo |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

|  |
| --- |
| **TÊN ĐỀ TÀI (IN HOA)**  CÔNG CỤ PHÁT HIỆN KẾT HỢP ANCHOR CHO BÀI TOÁN PHÁT HIỆN ĐỐI TƯỢNG CÓ HƯỚNG |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  HYBRID-ANCHOR ROTATION DETECTOR FOR ORIENTED OBJECT DETECTION |
| **TÓM TẮT** *(Tối đa 400 từ)*  Phát hiện đối tượng trong không ảnh đã trở thành một bài toán phổ biến trong những năm gần đây trong lĩnh vực thị giác máy tính (computer vision) và nhận diện mẫu (pattern recognition ), bài toán này còn nhiều thách thức bởi vì định hướng tùy ý của các đối tượng có trong không ảnh khác các đối tượng trong ảnh tự nhiên. Đối mặt với thách thức về hướng, các phương pháp phát hiện đối tượng yêu cầu có nhiều tham số để mã hóa thông tin về hướng, điều này dư thừa và cực kì không hiệu quả. Các phương pháp SOTA một giai đoạn hoặc hai giai đoạn hiện nay thường áp dụng các chiến lược dựa trên Anchor để tạo các khu vực đề xuất (Region Proposals) bằng cách sử dụng số lượng anchor được tạo dư thừa để huấn luyện, điều này không hiệu quả đối với các nguồn lực hạn chế. Đồng thời, các cách tiếp cận không có Anchor nhanh hơn nhiều nhưng thường làm giảm một số lượng lớn các mẫu huấn luyện, làm giảm độ chính xác của tác vụ phát hiện. Trong bài báo này, chúng tôi trình bày phương pháp phát hiện kết hợp Anchor (Hybrid-Anchor Rotation Detector) nhằm mục đích thu hẹp khoảng cách giữa các cách huấn luyện Anchor-free và Anchor-based để phát hiện đối tượng có hướng. Cụ thể, chúng tôi chỉ sử dụng một Anchor đặt trước cho từng vị trí trong feature map và triển khai thêm các thành phần định hướng bổ sung giúp tăng đáng kể hiệu suất phát hiện của HA-RDet. Các thực nghiệm mở rộng của HA-Rdet được thực hiện trên nhiều bộ dữ liệu định hướng nổi tiếng như DOTA, DIOR-R và HRSC2016 và so sánh một cách cạnh tranh với các phương pháp dựa trên Anchor hiện tại và tốc độ suy luận xấp xỉ với các phương pháp Anchor-free. |
| **GIỚI THIỆU** *(Tối đa 1 trang A4)*  Những nghiên cứu về phát hiện đối tượng trong không ảnh gần đây gặp phải những khó khăn do sự khác nhau do những đối tượng trong không ảnh thường được phân phối với hướng tùy ý khác với các đối tượng trong ảnh tự nhiên. Để giải quyết những thử thách này, các phương pháp phát hiện đối tượng trong không ảnh được xây dựng như một công cụ thực hiện tác vụ phát hiện đối tượng có hướng bằng cách biểu diễn đối tượng với Oriented Bounding Boxes (OBB) thay vì Horizontal Bounding Boxes (HBB).  Nghiên cứu của chúng tôi đóng góp một mô hình học sâu dùng để phát hiện các đối tượng có hướng trong không ảnh. Mô hình của chúng tôi kết hợp sử dụng cơ chế Ancho-free và Anchor-based để tăng cường hiệu năng của mô hình. Ngoài ra, chúng tôi chỉ sứ dụng 1 anchor trên mỗi vị trí trên feature map giúp giảm thiểu số lượng tài nguyên tính toán của mô hình. Việc chỉ sử dụng 1 anchor trên 1 scale và 1 ratio làm cho việc huấn luyện mô hình vô cùng khó khăn vì nó dẫn đến tình trạng thiếu mẫu đữ liệu positive để huấn luyện. Phương pháp của chúng tôi đề xuất kết hợp 2 cơ chế huấn luyện thông qua bước cải thiện các proposals có thể làm cho kết quả của mô hình được cải thiện hơn rất nhiều và đạt được state-of-the-art cho bài toán phát hiện đối tượng có hướng trong không ảnh.  Chúng tôi thực nghiệm mô hình của chúng tôi trên ba bộ dữ liệu nổi tiếng về phát hiện đối tượng có hướng là DOTA, DIOR-R và HRSC2016 và đạt đưuọc kết quả vô cùng khả quan.  Ngoài ra, chúng tôi còn đề xuất ba đầu R-CNN thực nghiệm là Oriented Cascade Head, Oriented Dynamic Head và Oriented Double Head và kết hợp chúng cho mô hình của chúng tôi với mong đợi đạt được kết quả cao so với các phương pháp tiền nhiệm.  Thêm vào đó, chúng tôi còn thực nghiệm phương pháp của chúng tôi trên 3 backbone ResNet50, ResNeXt101-DCN và ReResNet50-ReFPN và bổ sung các phân tích về sự tác động của các backbone lên hiệu năng của mô hình chúng tôi đề xuất trên cả ba bộ dữ liệu thực nghiệm. |
| **MỤC TIÊU**  *(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)*   1. Khảo sát các hướng tiếp cận phổ biến cho bài toán phát hiện đối tượng có hướng trong không ảnh 2. Đề xuất mô hình phát hiện đối tượng trong không ảnh sử dụng kết hợp cơ chế Anchor-Free và Anchor-based. 3. Đề xuất 3 đầu thực nghiệm R-CNN: Oriented Double Head, Oriented Cascade R-CNN, Oriented Dynamic R-CNN 4. Đề xuất đánh giá 3 backbones: ResNet50, ReResNet50, ReXtNet101\_DCN. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  *(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)*   1. **Kiến trúc thành phần của phương pháp đề xuất**   Trong đề tài nghiên cứu này, chúng tôi dự kiến sẽ nghiên cứu và tìm hiểu thiết kế kiến trúc mô hình bao gồm 2 phần chính: Hybrid Alignment RPN và R-CNN head  Diagram  Description automatically generated   1. **Nội dung 1: Hybrid-Anchor alignment RPN**   Hầu hết các phương pháp phát hiện đối tượng được định hướng tuân theo cơ chế dựa trên anchor để dự đoán các hộp giới hạn có hướng được mã hóa nhiều, chúng thường tạo ra các anchor dự phòng có tỷ lệ khác nhau để phân bổ “tham lam” nhiều như các phiên bản có hình dạng và kích cỡ khác nhau. Xu hướng này dẫn đến số lượng anchor cần thiết theo cấp số nhân để các mô hình đạt được hiệu suất đầy hứa hẹn, dẫn đến nhu cầu về tài nguyên tính toán cao. Mặt khác, anchor-free chỉ yêu cầu một neo duy nhất cho mỗi vị trí của feature map; tuy nhiên nó hoạt động nhanh nhưng độ chính xác sẽ giảm đi do thiếu các mẫu huấn luyện. Có nhiều nghiên cứu cải thiện hiệu suất của một trong hai phương pháp thông qua tinh chỉnh sau, label assignment mới, định nghĩa anchor mới, v.v. các chiến lược bế tắc và không được khai thác triệt để trong lĩnh vực. Chúng tôi đề xuất một sự kết hợp giữa Anchor-free và Anchor-based để phát hiện đối tượng có hướng, cụ thể là Bộ phát Hybrid kết hợp (HA-RDet).  **Anchor Generator**: Chúng tôi sử dụng duy nhất một preset anchor với 1 scale và 1 ratio cho mỗi vị trí của feature map sử dụng chơ chế Anchor-free. Các phương pháp tiền nhiệm sử dụng nhiều anchor với nhiều kích thước và tỉ lệ khác nhau để đảm bảo số lượng các anchor đủ để mô hình học một cách tốt nhất. Bằng thực nghiệm, chúng tôi sử dụng anchor có scale là 4.0 và ratio là 1.0 cho đầu generator bởi vì mô hình được huấn luyện dưới hyper parameter này cho kết quả tốt nhất.  **Bridged Extracted Features**: Sau khi extracted features được đưa qua layer ra để predict ra các H.Proposals. Trên mỗi location của feature map, một preset anchor với cùng một scale và ratio được sinh ra cùng với kernel của conv trượt trên feature map. Các feature cùng với các H.Proposals đưa qua H.Alignment Conv và cuối cùng được refined bởi Anchor-based head   1. **Nội dung 2: R-CNN heads**   Trong báo cáo này, chúng tôi chỉ thực nghiệm mô hình đề xuất trên R-CNN head đơn giản: “kiến trúc oriented double head được đề xuất”. Rotation R-CNN head được dùng để hồi quy và phân lớp sử dụng RiRoI Align.  **RiRoI Align** bao gồm 2 phần:   * Spatial Alignment (RRoI Align): Với một RRoI (𝑥,𝑦,𝑤,ℎ,𝜃), spatial alignment warps RRoI đó từ các feature maps để produce ra được rotation-invariant region features trong spatial dimension. * Orientation Alignment (Rotation-invariant): Đảm bảo rằng RRoIs với các hướng khác nhau đều sẽ produce ra được rotation-invariant features, chúng ta tiến hành thực hiện orientation alignment trong orientation dimension.  1. **Các kiến trúc thử nghiệm khác:**   Backbones:   * ResNet50\_FPN * ReResNet50\_FPN * ReXtNet101\_DCN   Heads:   * Oriented Double Heads * Oriented Dynamic Heads * Oriented Cascade Heads |
| **KẾT QUẢ MONG ĐỢI**  *(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)*   1. Xây dựng mô hình hoản chỉnh theo kiến trúc kết hợp Anchor. Thử nghiệm thành công mô hình đã xây dựng với sự kết hợp các backbone và các đầu RCNN khác nhau. 2. Mô hình cuối cùng đạt độ chính xác tốt hơn trong khi đạt tốc độ xử lý xấp xỉ tương đồng so với các mô hình hiện nay thực hiện trên bài toán phát hiện đối tượng có hướng trong không ảnh 3. Phương pháp HA-Rdet nếu thành công, chúng tôi dự kiến công bố:  * 01 bài báo tạp chí Q1 – ISI * 01 bài hội nghị rank A  1. Mô hình triển khai được trong các công cụ thực tế phục vụ cho việc trực quan hóa nghiên cứu (drone, điện thoại) |
| **TÀI LIỆU THAM KHẢO** *(Định dạng DBLP)*  [1]. Jiaming Han, Jian Ding, Nan Xue, Gui-Song Xia: ReDet: A Rotation Equivariant Detector for Aerial Object Detection. CVPR 2021: 2786-2795 CVPR 2021: 1-11.  [2]. Jiaming Han, Jian Ding, Jie Li, Gui-Song Xia: Align Deep Features for Oriented Object Detection. IEEE Trans. Geosci. Remote. Sens. 60: 1-11 (2022).  [3]. Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, Qikai Lu: Learning RoI Transformer for Oriented Object Detection in Aerial Images. CVPR 2019: 2849-2858.  [4]. Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, Junwei Han: Oriented R-CNN for Object Detection. ICCV 2021: 3500-3509.  [5]. Yue Wu, Yinpeng Chen, Lu Yuan, Zicheng Liu, Lijuan Wang, Hongzhi Li, Yun Fu: Rethinking Classification and Localization for Object Detection. CVPR 2020: 10183-10192.  [6]. Zhaowei Cai, Nuno Vasconcelos: Cascade R-CNN: High Quality Object Detection and Instance Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 43(5): 1483-1498 (2021).  [7]. Hongkai Zhang, Hong Chang, Bingpeng Ma, Naiyan Wang, Xilin Chen: Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training. ECCV (15) 2020: 260-275.  [8]. Wentong Li, Yijie Chen, Kaixuan Hu, Jianke Zhu: Oriented RepPoints for Aerial Object Detection. CVPR 2022: 1819-1828.  [9]. Xue Yang, Junchi Yan, Ziming Feng, Tao He: R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. AAAI 2021: 3163-3171.  [10]. Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge J. Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, Liangpei Zhang: DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. CVPR 2018: 3974-3983. |