

# **YOLACT:**

## **Phân vùng Thực thể**

## **Thời gian thực**

**Nguyễn Tấn Phúc - 220101035**



# Tóm tắt

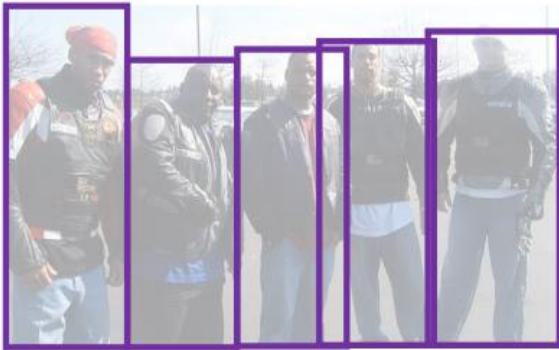


- Họ và Tên: Nguyễn Tấn Phúc
- MSHV: 220101035
- Lớp: CS2205.APR2023
- Link Github:  
[phucnguyen250300/CS2205](https://github.com/phucnguyen250300/CS2205)  
[github.com](https://github.com)
- Link YouTube video:  
<https://youtu.be/9odfH74w4Hw>

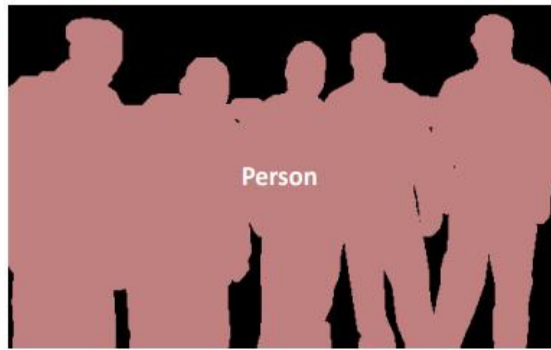


# Giới thiệu

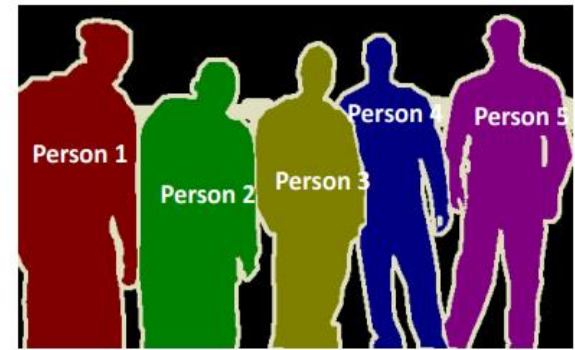
- Bài toán Phân vùng Thực thể:
  - Định nghĩa



Object Detection



Semantic Segmentation



Instance Segmentation

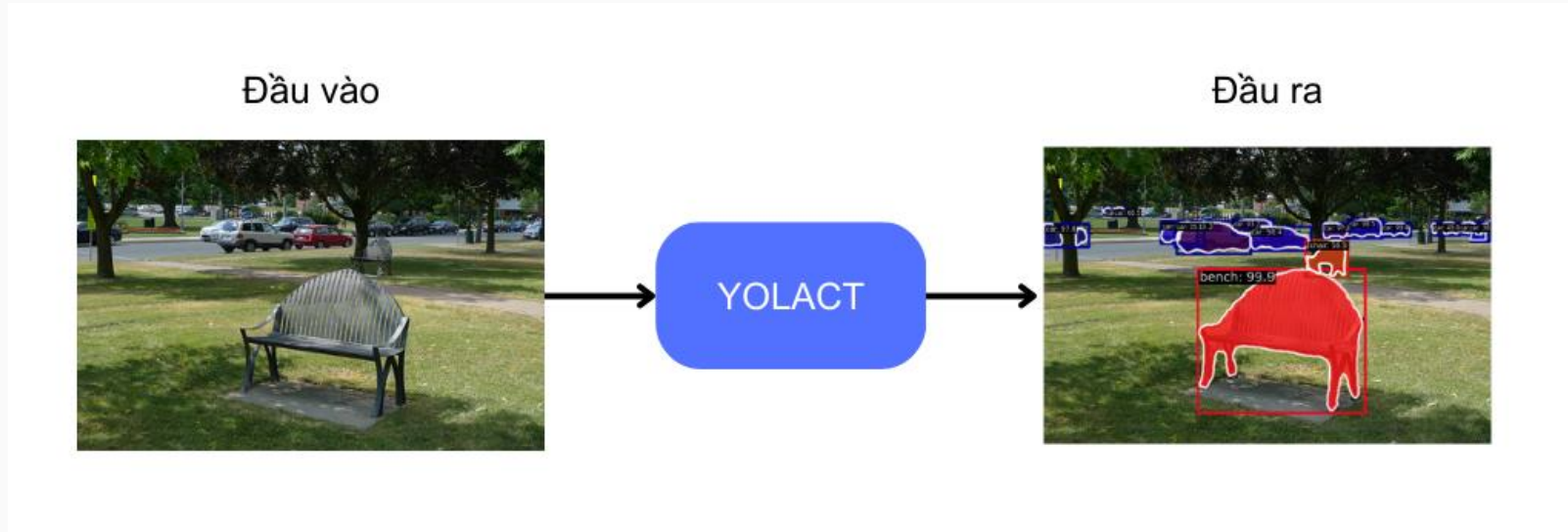
Nguồn: Towards Data Science

[<https://towardsdatascience.com/single-stage-instance-segmentation-a-review-1eeb66e0cc49>]



# Giới thiệu

- Bài toán Phân vùng Thực thể: YOLACT



**Hình 1.** Đầu vào và đầu ra của YOLACT



# Mục tiêu

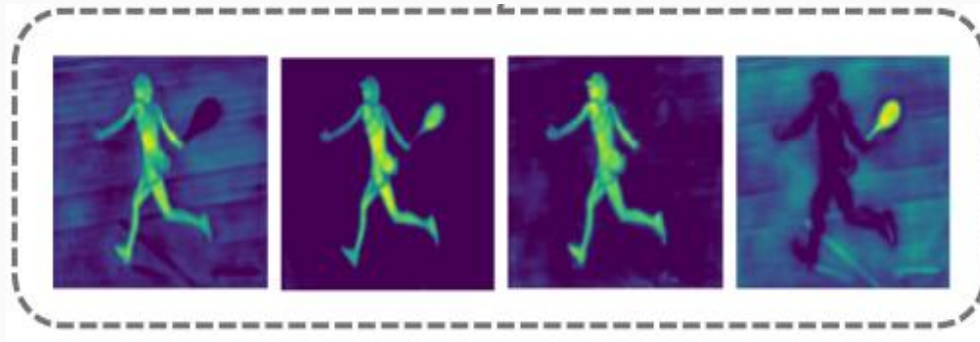
- **Giảm thời gian thực thi:** Tiến hành thử nghiệm song song 2 tác vụ sinh mặt nạ mẫu và dự đoán hệ số mặt nạ. Sau đó, thử nghiệm các cách kết hợp 2 đầu ra kể trên sao cho tối ưu hóa cho quá trình thực thi thời gian thực.
- **Cải thiện độ chính xác:** Tiến hành nghiên cứu lại cách bảo toàn lượng thông tin mất mát ở các lớp pooling mà những phương pháp SOTA sử dụng.
- **Mang tính khái quát:** Ý tưởng sinh mặt nạ mẫu và hệ số mặt nạ có thể được áp dụng vào đa số các phương pháp phân vùng thực thể.



# Nội dung và Phương pháp

## 1. Sinh mẫu

- Sinh ra một tập các “mặt nạ mẫu” có kích thước bằng với ảnh đầu vào.



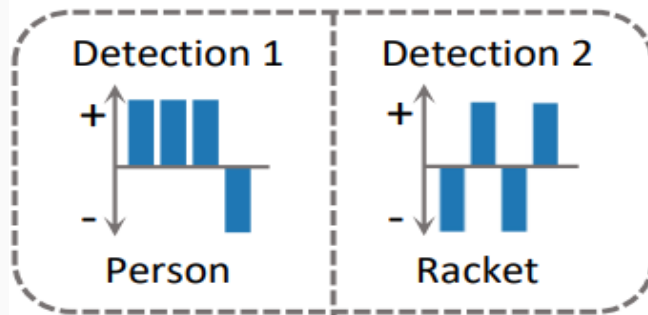
Hình 2. Sinh mẫu



# Nội dung và Phương pháp

## 2. Hệ số mặt nạ

- Dự đoán hệ số mặt nạ cho mỗi thực thể.



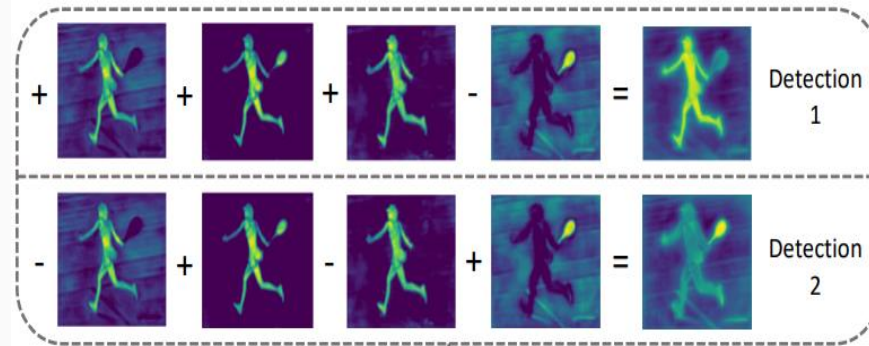
Hình 3. Hệ số mặt nạ



# Nội dung và Phương pháp

## 3. Kết hợp

- Tạo ra mặt nạ kết quả cho mỗi thực thể bằng cách kết hợp tuyến tính mặt nạ mẫu và hệ số mặt nạ.



Hình 4. Kết hợp





# Kết quả dự kiến

- **Thời gian thực thi được rút ngắn:** Các tác vụ được thực hiện song song cùng với bước kết hợp tuyến tính đơn giản.
- **Mặt nạ kết quả bao phủ đúng thực thể:** Thông tin không gian bị mất mát ở các tầng pooling đã được bảo toàn.
- **Có tính khái quát:** Ý tưởng triển khai được trên hầu hết những phương pháp thuộc phạm vi bài toán nhận dạng đối tượng.



# Tài liệu tham khảo

- [1] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask R-CNN." arXiv, Jan. 24, 2018. Accessed: Jul. 16, 2023. [Online]. Available: <http://arxiv.org/abs/1703.06870>
- [2] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei, "Fully Convolutional Instance-aware Semantic Segmentation." arXiv, Apr. 10, 2017. Accessed: Jul. 16, 2023. [Online]. Available: <http://arxiv.org/abs/1611.07709>
- [3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." arXiv, Jan. 06, 2016. Accessed: Jul. 16, 2023. [Online]. Available: <http://arxiv.org/abs/1506.01497>
- [4] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2016. Accessed: Jul. 16, 2023. [Online]. Available: [https://papers.nips.cc/paper\\_files/paper/2016/hash/577ef1154f3240ad5b9b413aa7346a1e-Abstract.html](https://papers.nips.cc/paper_files/paper/2016/hash/577ef1154f3240ad5b9b413aa7346a1e-Abstract.html)
- [5] Joseph Redmon and Ali Farhadi, "YOLO9000: Better, Faster, Stronger." arXiv, Dec. 25, 2016. Accessed: Jul. 16, 2023. [Online]. Available: <http://arxiv.org/abs/1612.08242>
- [6] Joseph Redmon and Ali Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018. Accessed: Jul. 16, 2023. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [7] Wei Liu et al., "SSD: Single Shot MultiBox Detector," 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0\_2.
- [8] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature Pyramid Networks for Object Detection," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, Jul. 2017, pp. 936–944. doi: 10.1109/CVPR.2017.106.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." arXiv, Dec. 10, 2015. Accessed: Jul. 17, 2023. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice: IEEE, Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.
- [11] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL: IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

