TỐI ƯU PHÂN ĐOẠN ĐỐI TƯỢNG TRONG VIDEO DỰA TRÊN KIẾN TRÚC BIẾN HÌNH ĐA HÌNH

Võ Huy Khôi - 220101042

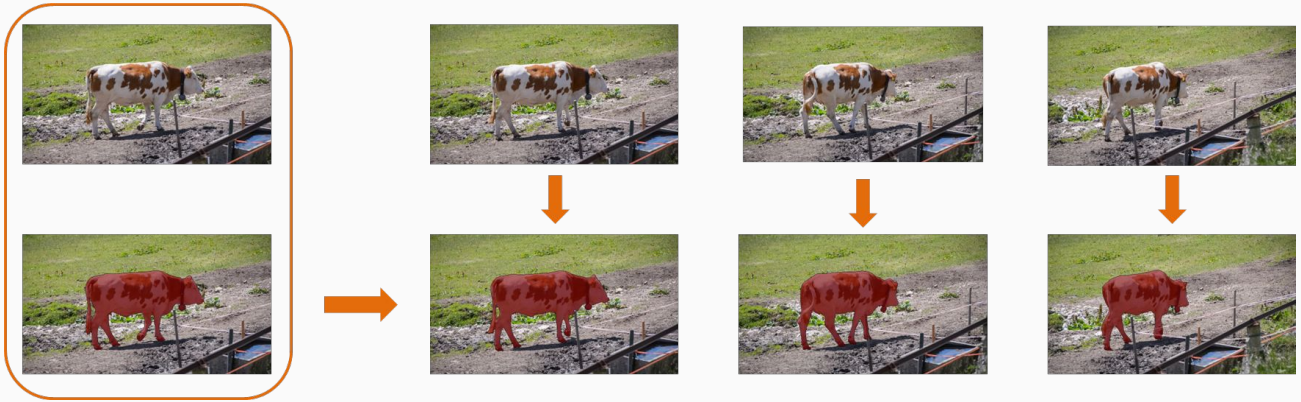
Tóm tắt

- Họ và tên: Võ Huy Khôi
- Lớp: CS2205.APR2023
- Link Github: https://github.com/HuyKhoiGrad
- Link YouTube video: https://youtu.be/NDB7PZ1zC68

Giới thiệu

Phân đoạn vật thể trong video (Video Object segmentation hay VOS) theo phương pháp học **bán giám sát.**

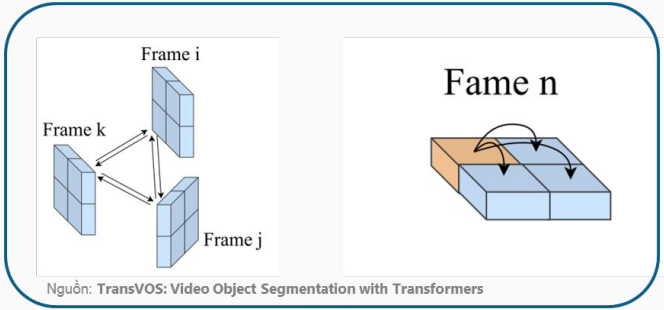
- **Đầu vào:** Một chuỗi các video frame và objects mask của frame đầu tiên.
- **Đầu ra:** Object mask được phân đoạn cho từng frame còn lại trong chuỗi video.



Giới thiệu

Một trong những yêu cầu tiên quyết cho VOS là phải khai phá được **các mối quan hệ phụ thuộc giữa các pixels trong không-thời gian**.

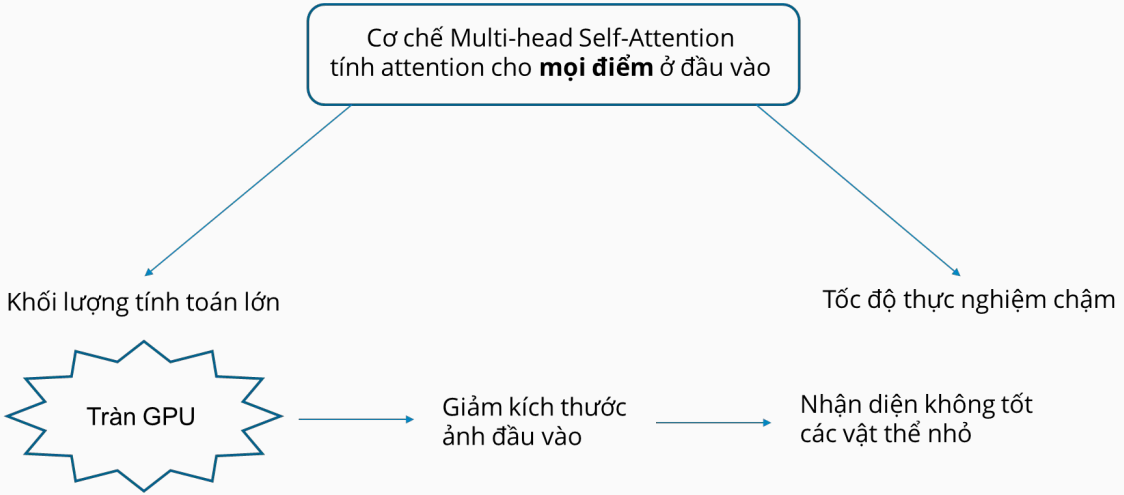
Mô hình TransVOS với kiến trúc Transformer đã cho thấy khả năng áp dụng cơ chế Multi-head Self-attention để giải quyết bài toán Phân đoạn vật thể trong video.



Nguồn: **TransVOS: Video Object Segmentation with Transformers**

Cơ chế Attention

Giới thiệu



Cơ chế Deformable Attention ra đời nhằm giải quyết cho vấn đề tính toán của Multi-head Self-Attention trong kiến trúc Transformer

=> Nhóm đề xuất mô hình DeformableTransVOS với kiến trúc của Deforrmable Transformer

Mục tiêu

- Tìm hiểu về cơ chế hoạt động của Attention trong bài toán Phân đoạn vật thể trong video.
- Nghiên cứu và ứng dụng kiến trúc của Deformable Transformer để giải quyết những hạn chế của mô hình TransVOS. Nhóm tạm gọi mô hình mới là DeformableTransVOS.
- Xây dựng ứng dụng web cho người dùng truyền tải video và sử dụng mô hình DeformableTransVOS để tự động phân tách vật thể.

Nội dung và Phương pháp

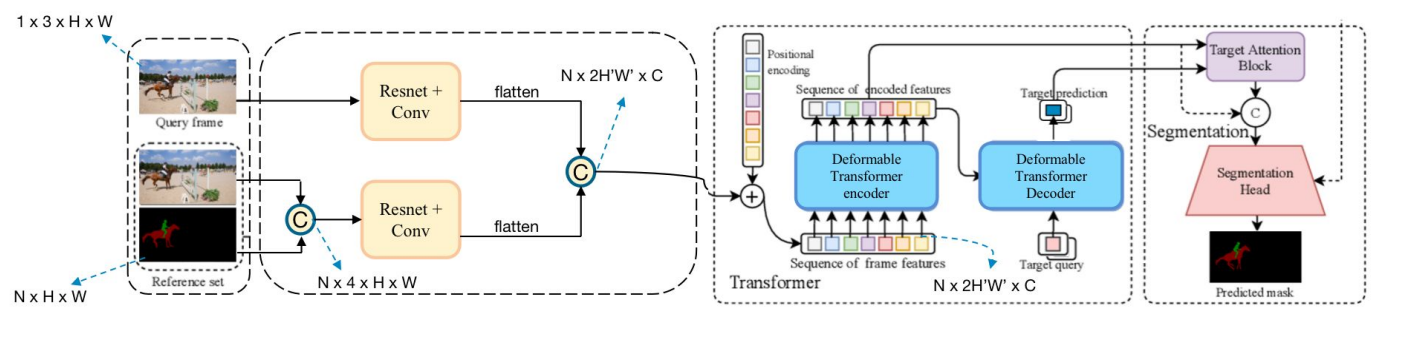
Nội dung:

- Nghiên cứu cơ chế phát hiện vật thể trong các bài toán Nhận diện vật thể trong ảnh của hai kiến trúc Transformer và Deformable Transformer.
- Xây dựng mô hình DeformableTransVOS.
- Huấn luyện và đánh giá các mô hình trên bộ dữ liệu đã được benchmark dành cho bài toán Phân đoạn đa vật thể là DAVIS-2017.
- Xây dựng ứng dụng web cho phép người dùng đăng tải video và thực hiện Phân tách vật thể bằng mô hình DeformableTransVOS và TransVOS

Nội dung và Phương pháp

Phương pháp:

- Xây dựng mô hình Deformable TransVOS bằng cách thay thế kiến trúc của Transformer trong mô hình TransVOS bằng Deformable Transformer
- Thực hiện huấn luyện hai mô hình TransVOS và DeformableTransVOS và đánh giá dựa trên bộ dữ liệu DAVIS-2017



Kết quả dự kiến

- Chỉ số đánh giá:
  - **J&F mean:** Trung bình cộng của 2 giá trị J và F
  - **FPS:** Số lượng frame mà mô hình xử lý trong một giây
- DeformbleTransVOS có khả năng tối ưu, giải quyết về nhược điểm tính toán lớn và tốc độ dự đoán chậm (fps) của mô hình TransVOS mà vẫn giữ được độ chính xác (J&F mean) trên bộ dữ liệu DAVIS-2017.
- Ứng dụng web hoàn chỉnh dễ sử dụng, cho phép thực hiện đăng tải video, gán nhãn đối tượng và tự động phân tách vật thể.

Tài liệu tham khảo

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention Is All You Need. CoRR abs/1706.03762 (2017)

[2] Jianbiao Mei, Mengmeng Wang, Yeneng Lin, Yong Liu: TransVOS: Video Object Segmentation with Transformers. CoRR abs/2106.00588 (2021)

[3] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, Jifeng Dai: Deformable DETR: Deformable Transformers for End-to-End Object Detection. ICLR 2021

[4] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbelaez, Alexander Sorkine-Hornung, Luc Van Gool: The 2017 DAVIS Challenge on Video Object Segmentation. CoRR abs/1704.00675 (2017)