

THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):
https://youtu.be/l9_dJlo5k6U
- Link slides (dạng .pdf đặt trên Github):
(ví dụ: <https://github.com/mynameuit/CS519.O21.KHTN/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

<ul style="list-style-type: none">• Họ và Tên: Dương Phạm Hoàng Anh• MSSV: 22520042 	<ul style="list-style-type: none">• Lớp: CS519.O21.KHTN• Tự đánh giá (điểm tổng kết môn):• Số buổi vắng: 1• Số câu hỏi QT cá nhân: 8• Link Github: https://github.com/mynameuit/CS519.O21.KHTN/
---	--

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

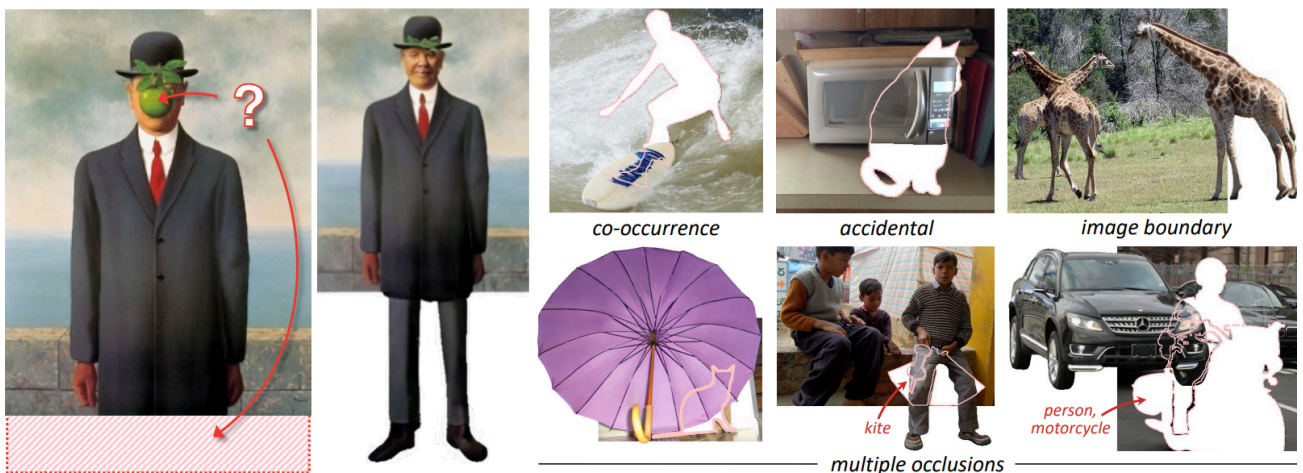
AMODAL COMPLETION CHO VIDEO: ÁP DỤNG PROGRESSIVE MIXED
CONTEXT DIFFUSION [1]

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

ENHANCING AMODAL COMPLETION FOR VIDEO: PROGRESSIVE MIXED
CONTEXT DIFFUSION APPROACH

TÓM TẮT (Tối đa 400 từ)

Amodal Completion là kỹ thuật quan trọng trong thị giác máy tính, giúp nhận diện và hoàn thiện các đối tượng bị che khuất, có ứng dụng trong robot, xe tự lái và AR/VR. Hiện tại, các phương pháp tập trung vào hình ảnh tĩnh, trong khi ứng dụng thực tế yêu cầu xử lý video liên tục. Chúng tôi phát triển phương pháp mới dựa trên "Amodal Completion via Progressive Mixed Context Diffusion" để hoàn thiện các đối tượng bị che khuất trong video một cách mượt mà và liên tục. Quy trình bao gồm phân tích và phân đoạn video, khuếch tán ngữ cảnh hỗn hợp, và kiểm tra đánh giá chất lượng video. Phương pháp này sẽ cải thiện khả năng nhận diện và hoàn thiện trong video phức tạp, có ứng dụng rộng rãi, và đạt chất lượng cao theo các chỉ số định lượng.



GIỚI THIỆU (Tối đa 1 trang A4)

Bối Cảnh và Tầm Quan Trọng của Đề Tài: *Amodal Completion* là một khái niệm quan trọng trong thị giác máy tính, tập trung vào việc nhận diện và hoàn thiện các đối tượng bị che khuất một phần trong hình ảnh

Bài Toán Cần Giải Quyết: Hiện tại, các phương pháp *Amodal Completion* chủ yếu tập trung vào hình ảnh tĩnh, trong khi đó, các ứng dụng thực tế thường yêu cầu xử lý chuỗi video liên tục. Bài toán đặt ra là làm thế nào để hoàn thiện các đối tượng bị che khuất trong chuỗi video một cách mượt mà và liên tục giữa các khung hình, đảm bảo tính nhất quán thời gian và không làm mất đi các chi tiết quan trọng của đối tượng. Điều này đòi hỏi một phương pháp tiếp cận mới, kết hợp giữa kỹ thuật xử lý hình ảnh tĩnh và khả năng phân tích ngữ cảnh động trong video.

Động Lực và Lý Do Chọn Đề Tài: Đề tài này được lấy cảm hứng từ phương pháp "*Amodal Completion via Progressive Mixed Context Diffusion*," một phương pháp tiên tiến trong việc hoàn thiện các đối tượng bị che khuất trong hình ảnh tĩnh. Phương pháp này không sử dụng cách tiếp cận hai bước truyền thống mà thay vào đó, nó mở rộng từ hộp giới hạn của đối tượng bị che khuất để sử dụng ngữ cảnh hình ảnh, sau đó sử dụng mô hình khuếch tán đã được huấn luyện trước để từng bước hoàn thiện đối tượng bị che khuất và loại bỏ nền thừa. Nhìn thấy tiềm năng và hiệu quả của phương pháp này, chúng tôi quyết định mở rộng và áp dụng nó cho chuỗi video, nơi mà các đối tượng cần được hoàn thiện một cách liên tục và mượt mà giữa các khung hình.

Đầu Vào và Đầu Ra Của Đề Tài:

- **Đầu Vào:** Chuỗi video có các đối tượng bị che khuất, mô hình phân đoạn và phân tích độ sâu, các thông tin ngữ cảnh liên quan.
- **Đầu Ra:** Chuỗi video với các đối tượng bị che khuất được hoàn thiện một cách mượt mà và liên tục, giữ nguyên tính nhất quán thời gian và các chi tiết quan

trọng.

MỤC TIÊU

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

Hoàn thiện amodal: Áp dụng mô hình khuếch tán để hoàn thiện amodal cho từng khung hình.

Duy trì liên tục: Đảm bảo tính nhất quán và liên tục của đối tượng qua chuỗi khung hình.

Giảm co-occurrence bias: Sử dụng kỹ thuật Mixed Context Diffusion Sampling để giảm thiểu đối tượng không mong muốn.

NỘI DUNG VÀ PHƯƠNG PHÁP

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

1. Phân Tích và Phân Đoạn Video:

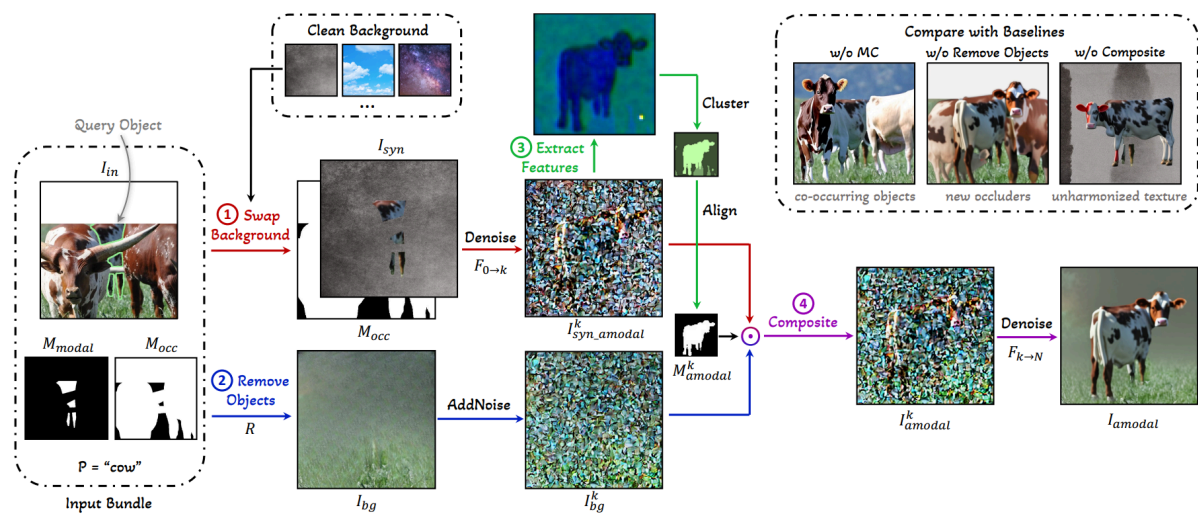
- **Instance Segmentation:** Áp dụng mô hình phân đoạn như Mask R-CNN để phân tích và xác định các đối tượng trong từng khung hình của chuỗi video.
- **Depth Analysis:** Thực hiện phân tích độ sâu để xác định các đối tượng bị che khuất và các ngữ cảnh xung quanh, giúp phân biệt giữa các lớp đối tượng khác nhau trong video.

2. Khuếch Tán Ngữ Cảnh Hỗn Hợp (Mixed Context Diffusion): Áp dụng kỹ thuật này để từng bước hoàn thiện các đối tượng bị che khuất trong từng khung hình của chuỗi video.

- **Swap Background:** Thay thế nền gốc của khung hình bằng một nền sạch tạm thời để giảm thiểu sự thiên vị ngữ cảnh. Tạo ra một hình ảnh tổng hợp (Isyn) và áp dụng khuếch tán đến một bước khuếch tán k nhất định (Iksyn amodal).
- **Create Object-removed Background Image:** Tạo một hình ảnh nền sạch bằng

cách loại bỏ các đối tượng chính và occluders từ khung hình gốc, sau đó thêm noise để đạt đến bước khuếch tán k (I_{kbg}).

- *Segment Query Object in Noisy Image*: Sử dụng các đặc trưng từ mô hình khuếch tán để tạo ra các cluster và chọn mask amodal cho đối tượng ở bước khuếch tán k ($M_{k\text{ amodal}}$).
- *Composite*: Sử dụng $M_{k\text{ amodal}}$ để gắn đối tượng vào hình ảnh nền sạch, tạo ra hình ảnh tổng hợp $I_{k\text{ amodal}}$. Tiếp tục khuếch tán cho các bước còn lại để hoàn thiện hình ảnh cuối cùng (I_{amodal})



3. Duy Trì Tính Liên Tục và Nhất Quán:

- Thực hiện hoàn thiện đối tượng từng bước, sử dụng đầu ra của bước trước làm đầu vào.
- Kiểm tra và tiếp tục khuếch tán nếu đối tượng còn bị che khuất.
- Đảm bảo tính nhất quán và liên tục của đối tượng qua các khung hình.

4. Progressive Context Control:

- Giảm thiểu thiên vị ngữ cảnh bằng cách thay nền sạch.
- Xác định và loại bỏ occluders mới để đảm bảo đối tượng không bị méo mó.

5. Kiểm Tra và Đánh Giá:

- Sử dụng PSNR, SSIM, và LPIPS để đo lường chất lượng video hoàn thiện.

KẾT QUẢ MONG ĐỢI

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

Hệ thống hoàn thiện amodal cho video:

- *Xây dựng một mô hình có khả năng phân tích và xử lý các video đầu vào, hoàn thiện các đối tượng bị che khuất trong từng khung hình.*

Khả năng duy trì tính liên tục và nhất quán:

- *Đối tượng liên tục: Hệ thống đảm bảo các đối tượng trong video được hoàn thiện một cách liên tục và không bị đứt đoạn qua các khung hình.*
- *Chất lượng cao: Sử dụng các kỹ thuật tiên tiến như Mixed Context Diffusion để đảm bảo tính chân thực và giảm thiểu sự thiên vị không mong muốn.*

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

[1]. Katherine Xu, Lingzhi Zhang, Jianbo Shi: Amodal Completion via Progressive Mixed Context Diffusion. CoRR abs/2312.15540 (2023)