Ứng dụng Xử lý ảnh số và video số
Seminar

Lưu Nam Đạt
22127062
lndat22@clc.fitus.edu.vn

Nguyễn Bá Công
22127046
nbcong22@clc.fitus.edu.vn

Nguyễn Huỳnh Hải Đăng
22127052
nhhdang22@clc.fitus.edu.vn

Đặng Trần Anh Khoa
22127024
dtakhoa22@clc.fitus.edu.vn

Lời Giới Thiệu

None

Mục lục

1 NHẬN DIỆN HÀNG HOÁ BÁN LẺ 1
1.1 Giới thiệu 1
1.2 Phát biểu bài toán 2
1.3 Phương pháp 2
1.3.1 Truyền thống 2
1.4 Nhận xét 2
1.4.1 phương pháp dựa theo đặc trưng 2
1.4.2 Deep learning 3

2 HỆ THỐNG TRUY VẾT ĐỐI TƯỢNG DỰA VÀO CÂU MÔ TẢ 3

3 Phát hiện bất thường trong giao thông 3

4 GRAPH OCR 4

5 View Synthesis using NeRF 4

6 BONE DISEASE VQA BASED ON MULTIMODAL TRANSFORMER ... 4
6.1 4
6.2 Phương pháp 4
6.3 Nhận xét 5

Tham khảo 5

1 NHẬN DIỆN HÀNG HOÁ BÁN LẺ

1.1 Giới thiệu

Nhận diện hàng hoá bán lẻ là quá trình ứng dụng Thị giác máy tính để tự động xác định, phân loại và theo dõi sản phẩm trong các siêu thị, cửa hàng.

Ví dụ là Amazon Go, mô hình cửa hàng không thu ngân, ứng dụng nhiều công nghệ hiện đại để nhận diện người dùng, tính tiền tự động theo đơn hàng.

1.2 Phát biểu bài toán

Đầu vào:
- Ảnh hoặc Video của sản phẩm
- Ảnh hoặc Video của kệ hàng

Đầu ra: Thông tin sản phẩm, bao gồm vị trí trên kệ, tên sản phẩm, giá cả, hạn sử dụng, v.v.

1.3 Phương pháp

1.3.1 Truyền thống

1. Template Matching: So khớp đặc trưng là một phương pháp để truy tìm vùng ảnh có chứa đặc điểm hoặc vật thể cụ thể bằng cách so sánh những đặc điểm của ảnh đầu vào với ảnh mục tiêu.

Nhược điểm: Không chống chọi được với phép biến đổi, tức là nếu hình ảnh bị xoay ngang, dọc, chéo, phóng to, thu nhỏ thì cũng không thể nhận dạng được sản phẩm. Nếu vật bị che khuất thì phương pháp này cũng hoạt động kém.

Phương pháp này không phù hợp với bán lẻ.

2. Đặc trưng SIFT (Scale-Invariant Feature Transform): Phương pháp này dựa vào trích xuất đặc trưng, chuyên xác định những điểm đặc trưng (keypoint) không bị ảnh hưởng khi phóng to, thu nhỏ, xoay, cũng như biến đổi affine.

Phương pháp này có 4 bước chính:
- Phát hiện các điểm đặc trưng trong không gian
- Định vị điểm đặc trưng (Keypoint Localization)
- Gán hướng (Orientation Assignment)
- Tạo mô tả về đặc trưng (Keypoint Descriptor)

Ưu điểm:
- Bất biến 1 phần trước phép quay, độ sáng, góc nhìn
- Có thể hoạt động cả khi bị che khuất một phần

Nhược điểm:
- Độ phức tạp tính toán cao
- Hoạt động kém với sản phẩm có ít đặc trưng

1.4 Nhận xét

- Nhóm vẫn chưa phát biểu được về cách thức phân loại trong những tình huống cụ thể, như phân biệt các sản phẩm cùng loại, khác nhãn hiệu (Coca-Cola với Pepsi, Sữa Vinamilk và sữa TH, …)
- Trong phần 2, nhóm không nêu rõ được mình sẽ mục tiêu thực hiện của công trình là gì,

(cần xếp thành category, cần có những tác vụ gì)

1.4.1 phương pháp dựa theo đặc trưng

SIFT - scale invariant feature extraction

1.4.2 Deep learning

A deep learning pipeline for product recognition on store shelves

Detection

2 HỆ THỐNG TRUY VẾT ĐỐI TƯỢNG DỰA VÀO CÂU MÔ TẢ

1. GIỚI THIỆU

1.1. BỐI CẢNH CHUNG

- xe tự hành, giao thông, an ninh

thách thức: hạn chế ngôn ngữ phân biệt đối tượng mục tiêu theo vết trong điều kiện phức tạp

2. PHÁT BIỂU BÀI TOÁN
3. CÁC CÔNG TRÌNH LIÊN QUAN

phải nói rõ về cách thức theo vết đối tượng

- TP-GMOT: Tracking Generic Multiple Object by Textual Prompt with Motion Appearance Cost SORT
- DTLLM-VLT:

tại 1 frame bất kỳ, có 2 trường hợp: 1 là đối tượng đang theo vết bị biến mất, 2 là đối tượng xuất hiện; khi đó câu mô tả phát huy như thế nào?

3 Phát hiện bất thường trong giao thông

Input: một đoạn video từ camera hành trình / camera an ninh

Output: Xác suất xảy ra tai nạn trong frame đang xét

Threshold: Một ngưỡng cảnh báo mức độ nguy hiểm

- MEDAVET: Traffic Vehicle Anomaly Detection Mechanism based on

spatial and temporal structures in vehicle traffic

- chưa giải thích được cơ chế tìm chiều di chuyển và vận tốc của phương tiện
- cần nói rõ ý chung trước khi đi sâu vào những biểu đồ và thuật toán, tuy có rất nhiều những neural network nhưng việc giải thích chưa đáng kể
- trong khung cảnh mà camera bắt được,
- dữ liệu không gian - thời gian (spetial - temporal)

cần hiểu "thế nào là tai nạn"

- đối với mỗi frame, cần quan tâm đến object nào để tính toán ra xác suất?

$\longrightarrow$ liệt kê 11 vật thể nó quan tâm:

- từ hình ảnh, rút ra đối tượng ra sao, từ đối tượng rút ra xác suất thế nào
- tại sao khi sắp có tai nạn thì xác suất được tăng lên?

Dùng YOLOv7 để phát hiện

- Dữ liệu đến từ những xe đã bị tai nạn, nhưng công tác gán nhãn diễn ra thế nào?

4 GRAPH OCR

Nhận diện đồ thị bằng OCR

- Chưa nhận diện rõ ứng dụng
- chưa sử dụng đồ thị viết tay
-

có những luận văn làm rất tốt, nhưng chatgpt có thể thừa sức đánh bại luận văn đó, gây điểm thấp

5 View Synthesis using NeRF

Tổng hợp góc nhìn

Từ một vài ảnh có góc nhìn hữu hạn, tạo thành một video với góc nhìn vô hạn

- bao nhiêu góc?

Neural Radiance Field

trong một không gian ảnh có điểm (x, y, z) Hàm 5d trả ra color, density, qua đó render trên một mặt phẳng 2D.

hàm lỗi là độ

- bản chất là tạo ra ảnh mới,
- ví dụ có 5 ảnh, cần tạo ảnh thứ 6 có view mới, từ góc alphabeta, thì lấy màu từ đâu? Có trước là tập hợp
- Cần giải thich cụ thể về radiance
- Cần xem các biểu thức toán
- không rõ input output: với mỗi r(t), suy ra được c(r)
- Cần giải thích vì sao phải lấy nhiều điểm
- $\hat{C}$?
- positional encoding cần tính 1 feature vector có sự biến thiên cao?
- Hierarchical volume sampling
- coorse network vs fine network: không hiểu
- Hàm lỗi của mô hình:

6 BONE DISEASE VQA BASED ON MULTIMODAL TRANSFORMER

6.1

6.2 Phương pháp

Decoder:

- Decoder giải mã và tìm cách liên kết với encoder

Encoder:

- Ảnh chụp y khoa được đưa vào Vision Encoder là SWIN
- Câu hỏi của bác sĩ được đưa vào Text Encoder là ViHealthBERT
- Kết quả của 2 encoder được đưa vào Fusion, gọi là CMAN.
- Chuyển tiếp qua Decoder có Learnable Answer

- MLP: có Sigmoid, Cross Entropy, AdamW (có weight decay để tránh làm ảnh hưởng đến Gradient khi loss thay đổi nhiều)
- Output là Class ID.

Vấn đề là cơ chế Generation tốn quá nhiều tài nguyên, nên chọn cơ chế Classification.

Dataset kết hợp hình ảnh xét nghiệm và

Training:
- Giai đoạn 1: Train 6 epoch, trọng số không đổi
- Giai đoạn 2: Train 3 epoch, có cho thay đổi trọng số

Thời gian train là cho cả 2 giai đoạn là hơn 10 tiếng.

⇒

6.3 Nhận xét

- Cần chắt lọc dữ liệu lại, nếu dữ liệu y khoa quá lớn
- Cần tự bổ sung thêm dữ liệu bằng cách đặt câu hỏi tương ứng.
- Nên sắp xếp câu hỏi theo category: What?, Where?
- Chưa giải thích rõ được cách đưa dữ liệu học vào mô hình: tức là 1 bảng, các cột là hình ảnh - câu hỏi - câu trả lời. Việc đưa raw data vào mô hình là vô lý.
- Cần trình bày câu hỏi, kết quả theo từng nhóm bệnh
-

Tham khảo