

BÁO CÁO ĐỒ ÁN MÔN HỌC

Bài toán phân đoạn ảnh với giám sát yếu

Ngày 14 tháng 5 năm 2024

Nội dung

- 1 DẪN NHẬP
- 2 CÔNG TRÌNH LIÊN QUAN
 - Thuật toán dựa trên nhãn toàn ảnh
 - Thuật toán dựa trên nhãn điểm
 - Thuật toán dựa trên nhãn vẽ tay
 - Thuật toán dựa trên nhãn là khung bao quanh
- 3 PHƯƠNG PHÁP HỌC BÁN GIÁM SÁT DỰA VÀO KHÔNG GIAN NHÚNG GIỮA CÁC ĐIỂM ẢNH
 - Ý tưởng
 - Mối quan hệ tương phản nhóm giữa điểm ảnh và phân vùng
 - Hàm đánh giá mất mát cho đặc trưng điểm ảnh
- 4 THỰC NGHIỆM VÀ KẾT QUẢ
- 5 TỔNG KẾT

Xác định bài toán

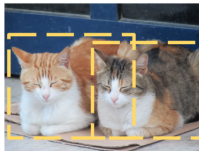
Có 4 loại nhiệm vụ chính liên quan đến ảnh trong thị giác máy tính là:

Classification



Is there a cat?

Detection



Where is the cat?

Semantic labeling



Which pixel is cat ?

Instance segmentation



Which are the cat's pixels?

- Phân loại ảnh
- Nhận diện đối tượng

- Gán nhãn theo ngữ nghĩa
- Gán nhãn theo thể hiện của lớp

Xác định bài toán

Classification



Is there a cat?

Detection



Where is the cat?

Semantic
labeling

Which pixel is cat ?

Instance
segmentation

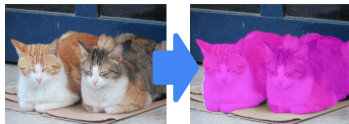
Which are the
cat's pixels?

Ở nhiệm vụ nhóm em xét sẽ là gán nhãn theo ngữ nghĩa

Xác định bài toán

Trong học giám sát, với mỗi mẫu ta có nhãn ngữ nghĩa với mỗi điểm ảnh trong tập huấn luyện còn đối với giám sát yếu thì nhãn chỉ có thưa thớt, không trực tiếp.

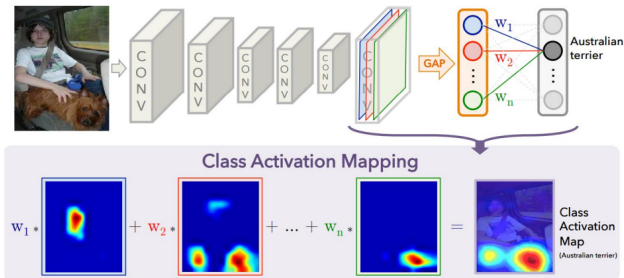
Fully supervised → **Directly** supervised



Weakly supervised → **Indirectly** supervised



Thuật toán dựa trên nhãn toàn ảnh

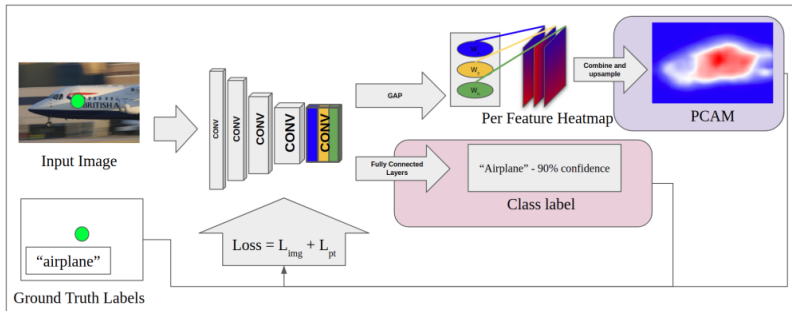


(Approach is known as GAP or CAM)

[Zhou et al. 2016]

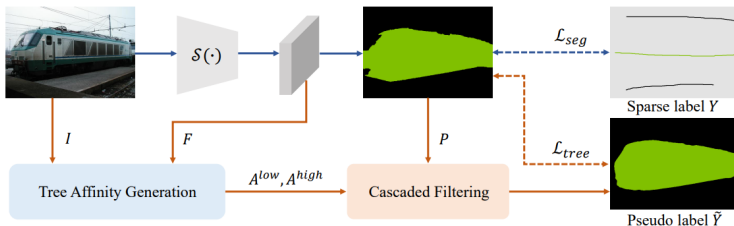
CAM - Nó xác định các vùng ảnh có thể sử dụng để phân biệt bằng cách chiếu các trọng số đầu ra thu được bởi phép pooling toàn cục trong mạng nơ-ron tích chập lên map đặc trưng, và tạo ra class activation map.

Thuật toán dựa trên nhãn điểm



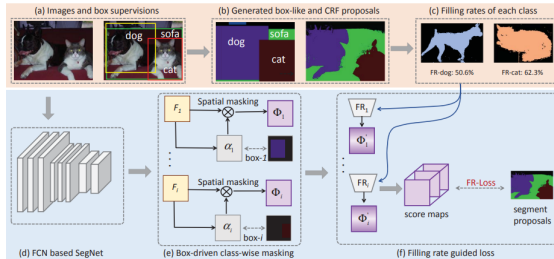
PCAM - Phương pháp để tạo ra nhãn giả bằng việc kết hợp giữa CAM và lan truyền ngược nhãn

Thuật toán dựa trên nhãn vẽ tay



TEL - Tree energy loss dùng các thuật toán tìm cây khung nhỏ nhất để tìm ra thông tin ngữ nghĩa giữa các điểm ảnh.

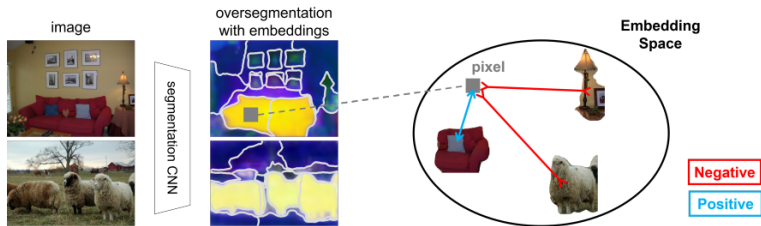
Thuật toán dựa trên nhãn là khung bao quanh



Box-driven class-wise masking model (BCM) phương pháp loại bỏ các vùng không liên quan đến đối tượng bên trong khung bao và được huấn luyện với hàm mất mát thích ứng được hướng dẫn bởi tỷ lệ lấp đầy bên trong khung bao cho mỗi lớp.

Ý tưởng

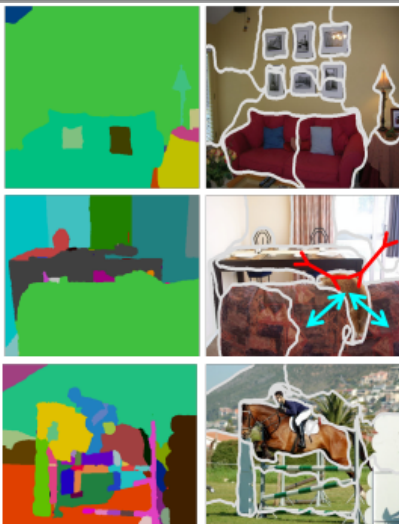
Tạo ra không gian nhúng tốt giữa các điểm ảnh thông qua việc học tương phản giữa các điểm ảnh và phân vùng



Mối quan hệ tương phản nhóm giữa điểm ảnh và phân vùng

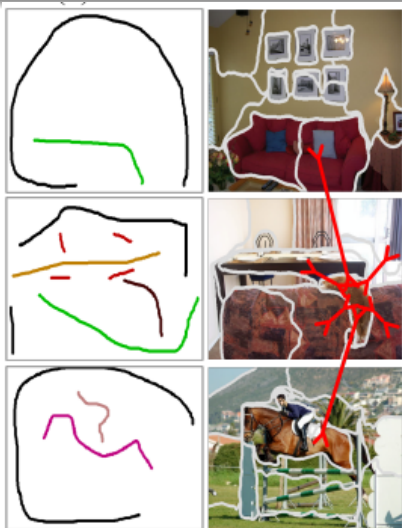
- Tương đồng hình ảnh cấp thấp
- Chú thích ngữ nghĩa
- Đồng xuất hiện ngữ nghĩa
- Tương thích đặc trưng

Tương đồng hình ảnh cấp thấp



- Tác giả sử dụng bộ phát hiện biên cạnh HED để tạo ra một phân vùng không có thông tin ngữ nghĩa.
- $V+$ và $V-$ lần lượt là các phân vùng dương và âm

Chú thích ngữ nghĩa



- Phải suy ra các nhãn giả từ hình ảnh hoặc bản đồ kích hoạt lớp và căn chỉnh chúng với các phân vùng được tạo ra bởi đặc trưng điểm ảnh.
- Các phân vùng dương hoặc âm của điểm ảnh i là các vùng cùng hoặc khác danh mục ngữ nghĩa, ký hiệu lần lượt là $C+$ và $C-$.

Đồng xuất hiện ngữ nghĩa

sofa



cat
sofa
table
chair

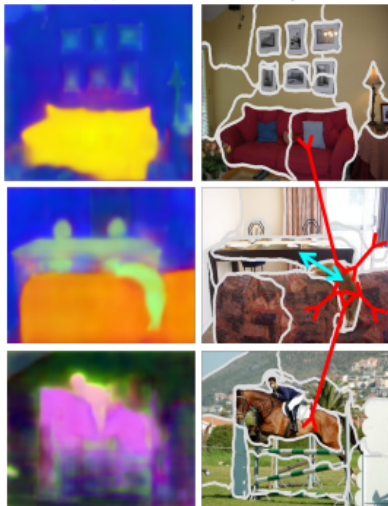


person
horse



- Ngữ cảnh ngữ nghĩa là hợp của các lớp đối tượng trong mỗi hình ảnh. Tận dụng ngữ cảnh ngữ nghĩa để áp dụng quy chuẩn toàn cầu cho việc học đặc trưng ẩn.
- Cho $O+$ ($O-$) là tập hợp các phân vùng trong các hình ảnh có (không có) các danh mục chồng lên nhau như hình ảnh của điểm ảnh i

Tương thích đặc trưng



- Phương pháp này gán một nhãn ngữ nghĩa cho mỗi đoạn không được gán nhãn bằng cách tìm đoạn gần nhất của nó trong không gian đặc trưng.
- Với điểm ảnh i , định nghĩa phân vùng dương, âm của nó lần lượt là \hat{C}^+ và \hat{C}^- tùy thuộc vào việc một đoạn có cùng nhãn với i hay không.

Hàm đánh giá mất mát cho đặc trưng điểm ảnh

Công thức hàm mất mát là :

$$\begin{aligned} L(i) = & \lambda_{IL} L_{SegSort^+}(i, V^+, V^-) + \lambda_C L_{SegSort^+}(i, C^+, C^-) \\ & + \lambda_{OL} L_{SegSort^+}(i, O^+, O^-) + \lambda_{AL} L_{SegSort^+}(i, \hat{C}^+, \hat{C}^-) \end{aligned}$$

Tập dữ liệu

Dataset **Pascal VOC** của Everingham

- 20 loại object cần học và 1 class nền
- 10582 ảnh trên tập training và 1449 ảnh trên tập validation.

Kết quả

Bảng: Kết quả so sánh với nhãn Image Tags

Image Tags	Saliency	val	test
DSRG	✓	61.4	63.2
FickleNet	✓	64.9	65.3
RRM	-	66.3	66.5
SGAN	✓	67.1	67.2
CAM+SE	-	66.1	65.9
SPML	-	69.5	71.6

Kết quả

Bảng: Kết quả so sánh với nhãn Bounding Box

Bounding Box	val	test
MCG+GrabCut+	69.4	-
BCAM	70.2	-
SPML	73.5	74.7

Kết quả

Bảng: Kết quả so sánh với nhãn Scribble

Scribble	CRF	Full	Weak	WvF
NCL	-	75.6	72.8	96.3
NCL	✓	76.8	74.5	97.0
RL	-	75.6	73.0	96.6
RL	✓	76.8	75.0	97.7
BPG	-	75.6	73.2	96.8
BPG	✓	75.6	73.2	96.8
SPML	✓	77.3	76.1	98.4

Kết quả

Bảng: Kết quả so sánh với nhãn Points

Points	val
What's the point[?]	46.1
BCAM	70.5
TEL	74.2
SPML	73.2

Tổng kết

- Mô hình Semisupervised Pixel-wise Metric Learning trong việc học bán giám sát cho bài toán phân vùng ảnh với giám sát yếu đạt được kết quả khá là tốt so với phương pháp khác.
- Bài toán được đặt ra còn cần được cải thiện nhiều hơn nữa trong các nghiên cứu tới.
- Kết quả về độ chính xác của mô hình như trên là hoàn toàn hợp lý và có thể chấp nhận được trong khả năng nghiên cứu của chúng em cũng như theo yêu cầu của đề án môn học.