

ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ
THÔNG TIN



NGÀNH: KHOA HỌC MÁY TÍNH

MÔN HỌC: PHƯƠNG PHÁP LUẬN NCKH

Áp dụng thị giác máy tính để
đếm mọi thứ từ hình ảnh

Sinh viên:

Phạm Trung Hiếu
Võ Khoa Nam
Trịnh Minh Hoàng

Giảng viên:

PGS.TS. Lê Đình Duy

1 Tóm tắt

Con người có thể đếm mọi vật một cách dễ dàng, nhưng để thực hiện nó một cách thủ công thì rất mất thời gian. Đặc biệt việc đếm hàng trăm, hàng nghìn đối tượng rất dễ xảy ra sai sót và ảnh hưởng trực tiếp đến doanh thu của doanh nghiệp. Hiện nay trong lĩnh vực thị giác máy tính có nhiều thuật toán có thể đếm số lượng vật thể thông qua hình ảnh phương pháp có một lợi thế lớn là không quá phụ thuộc vào việc lắp đặt thiết bị đếm. Nhưng đa phần các thuật toán ấy chỉ tập chung vào một lớp đối tượng cụ thể tại một thời điểm. Trong nghiên cứu này chúng tôi hướng đến phương pháp đếm tổng quát, tức là có thể đếm bất kì đối tượng thuộc lớp đối tượng nào. Để làm được việc này chúng tôi thực hiện ba công việc chính là nghiên cứu ba thuật toán mới được công bố trong những năm gần đây có khả năng đếm mọi thứ, xây dựng bộ dữ liệu và xây dựng chương trình minh họa.

2 Giới thiệu đề tài

Gần như tất cả các phương pháp đếm hiện có được thiết kế cho một lớp đối tượng cụ thể. Tuy nhiên trong bài báo này, chúng tôi quan tâm đến việc đếm mọi thứ, nghĩa là đếm các đối tượng từ bất kỳ lớp đối tượng nào chỉ đưa ra một vài trường hợp được chú thích từ lớp đối tượng đó. Có hai thách thức lớn ngăn cản cộng đồng Thị giác Máy tính:

1. Hầu hết các phương pháp đếm hiện đại coi việc đếm như một nhiệm vụ hồi quy có giám sát, yêu cầu hàng nghìn hình ảnh được gắn nhãn để học một hồi quy chấp hoàn toàn ánh xạ hình ảnh đầu vào tới bản đồ mật độ tương ứng của nó, từ đó số lượng ước tính thu được bằng cách tính tổng tất cả các giá trị mật độ.

2. Không có bất kỳ bộ dữ liệu đếm không bị giới hạn nào đủ lớn với nhiều danh mục trực quan để phát triển một phương pháp đếm tổng quát.

Một số thuật toán (GMN[1], MAML[2], FamNet[3]) được công bố gần đây đã có những thành công trong việc đếm bất kì đối tượng nào.

Trong đề tài này, chúng tôi nghiên cứu các thuật toán (GMN, MAML, FamNet) trong việc đếm đối tượng, và áp dụng thuật toán tốt nhất trong quá trình nghiên cứu để tạo ra một chương trình có thể đếm bất kì đối tượng nào. Cụ thể..

input: 1 bức ảnh chứa các đối tượng người dùng cần đếm, chỉ đưa ra một vài trường hợp được chú thích từ lớp đối tượng đó(tọa độ bounding box bao

quanh đối tượng cần đếm)

output: 1 bức ảnh người dùng đưa vào kèm theo chú thích bằng bounding box xung quanh mỗi đối tượng mà người dùng cần đếm và in ra số lượng đối tượng đếm được ở dạng văn bản.

3 Mục tiêu

- Nghiên cứu một số thuật toán giải quyết bài toán đếm mọi thứ hiện có.
- Xây dựng một bộ data mới
- Xây dựng ứng dụng minh họa có thể đếm được các đối tượng quan tâm từ hình ảnh đầu vào

4 Nội dung

- Nghiên cứu các thuật toán Few Shot Adaptation and Matching Network (FamNet), Generic Matching Network (GMN), Model Agnostic Meta Learning (MAML) trong việc đếm số lượng đối tượng.
- Xây dựng bộ dữ liệu mới gồm 6000 ảnh và 147 lớp để mô hình có thể học từ dữ liệu và đánh giá.
- Xây dựng các mô hình, so sánh kết quả và đưa ra đánh giá.
- Xây dựng chương trình minh họa.

5 Phương pháp

- Nghiên cứu về cấu trúc, nguyên lý hoạt động của các thuật toán Few Shot Adaptation and Matching Network (FamNet), Generic Matching Network (GMN), Model Agnostic Meta Learning (MAML) trong việc đếm số lượng đối tượng.
- Xây dựng bộ dữ liệu: Chúng tôi thu thập data bằng cách tìm kiếm từ khóa. Sau đó kiểm tra thủ công để lọc ra các hình ảnh đáp ứng yêu cầu: Chất lượng hình ảnh cao, độ phân giải cao để dễ dàng phân

biệt các đối tượng; số lượng đối tượng đủ lớn, số lượng đối tượng quan tâm ít nhất phải lớn hơn 7; giống nhau về ngoại hình, các cá thể đối tượng có kết cấu, tư thế, diện mạo hơi giống nhau; không có sự chồng chất vật thể nghiêm trọng. Đối với mỗi danh mục đối tượng chúng thu thập từ 300-3000 hình ảnh tìm kiếm trên web. Công cụ tìm kiếm Flickr, Google, Bing.

- Xây dựng các mô hình, so sánh kết quả và đưa ra đánh giá:
 - Xây dựng và đánh giá mô hình FamNet, GMN, MAML trên bộ dữ liệu mới tạo
 - Đánh giá trên bộ dữ liệu đã xây dựng trên độ đo (MAE, RMSE)
 - Xây dựng chương trình minh họa.
- Xây dựng chương trình minh họa trên web, người dùng có thể chỉ định đối tượng cần đếm, bằng cách vẽ bounding box vào một số đối tượng cần đếm để làm mẫu. Kết quả đầu ra là số lượng đối tượng.

6 Kết quả dự kiến

- Có bản báo cáo về các phương pháp nghiên cứu. Kết quả, đánh giá các phương pháp ưu và nhược điểm.
- Có bộ dữ liệu mới gồm 6000 ảnh và 147 lớp cho bài toán đếm vật thể.
- Có chương trình minh họa xuất ra số lượng đối tượng đầu vào từ hình ảnh đầu vào.

7 Tài liệu Tham khảo

- [1] Erika Lu, Weidi Xie, and Andrew Zisserman. Class-agnostic counting. In Proceedings of the Asian Conference on Computer Vision, 2018.
- [2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Modelagnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning, 2017.
- [3] Viresh Ranjan, Udbhav Sharma, Thu Nguyen, Minh Hoai. Learning To Count Everything, CVPR 2021.

- [4] Del Riccardo Chiaro. `python-flickr-image-downloader`.
- [5] Hardik Vasa. `Google images download`.