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

KHOA: KHOA HỌC MÁY TÍNH

MÔN HỌC: CS117.N21.KHTN - TƯ DUY TÍNH TOÁN

Crowd Counting

Thành viên:
Nguyễn Tuấn Anh – 21520142
Hà Văn Hoàng – 21520033
Phan Trường Trí – 21520117
Lương Toàn Bách – 21521845
Nguyễn Trường Thịnh – 21520110

Giảng viên:
TS. Ngô Đức Thành

Mục lục

1 Đặt vấn đề **2**

2 Giới thiệu bài toán **2**

3 Lý do chọn bài toán **3**

4 Phương pháp tiếp cận **3**

4.1 Density Map . . . 3

4.1.1 Ground-truth Density Map . . . 3

4.1.2 Density Map Estimation . . . 4

4.2 Multi-Column Convolutional Neural Network . . . 4

5 Abstraction **4**

6 Decomposition **5**

7 Pattern Recognition **5**

8 Giải thuật **6**

9 Độ đo **6**

10 Thực nghiệm **7**

10.1 Bộ dữ liệu . . . 7

10.2 Kết quả thực nghiệm . . . 7

11 Kết luận và hướng phát triển **8**

11.1 Kết luận . . . 8

11.2 Hướng phát triển . . . 9

1. Đặt vấn đề

Hiện nay, việc kiểm soát lưu lượng người trong một khoảng không gian là rất quan trọng. Các cửa hàng tiện lợi, các siêu thị cần biết được số lượng người ra vào để có thể phục vụ tốt. Ở các sân bay, người quản lý cần theo dõi những thay đổi về mật độ khách trên những con đường quan trọng, để có thể điều phối nhân viên giảm bớt mật độ khi cần thiết.

Kể từ khi dịch COVID-19 bùng phát, các giải pháp đếm người ngày càng trở nên phổ biến, giúp thực thi các quy tắc giãn cách xã hội và cải thiện an toàn.

Như vậy, việc có một hệ thống có khả năng kiểm soát được lưu lượng người phân bố là rất cần thiết.

2. Giới thiệu bài toán

Crowd Counting là bài toán ước tính số lượng người trong đám đông.

- **Input:** Ảnh
- **Output:** Số người được ước tính.

Ví dụ:

Hình 1: Ví dụ bài toán

Số người được ước tính: 23.

3. Lý do chọn bài toán

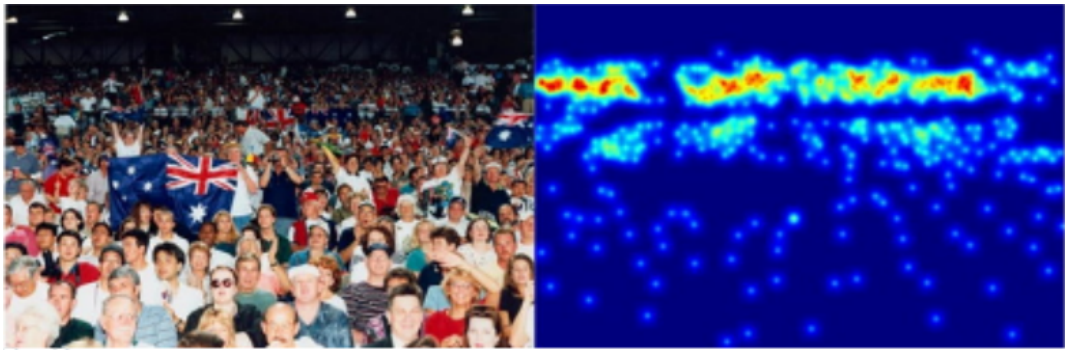
Một số lý do khả quan xung quanh bài toán.

- **Ứng dụng thực tế:** Giám sát an ninh, quản lý lưu lượng ở các cửa hàng, các sân bay, các khu vực diện rộng khác và nhiều ứng dụng khác.
- **Tiềm năng phát triển:** Mở rộng thành các hệ thống trong lĩnh vực sinh học như ước lượng số lượng tế bào,...

4. Phương pháp tiếp cận

4.1. Density Map

Density map là một ảnh hai chiều thể hiện mật độ người phân bố. Thông qua Density map có thể ước tính được số người. Density map bảo toàn nhiều thông tin hơn, ngoài ra còn cho biết sự phân bố không gian của đám đông trong hình ảnh, thông tin phân phối như vậy rất hữu ích trong nhiều ứng dụng.



Hình 2: Density Map

4.1.1. Ground-truth Density Map

Với một ảnh đám đông được đánh dấu tại các vị trí có đầu người, Ground-truth density map là một Density Map thể hiện **chính xác** mật độ.

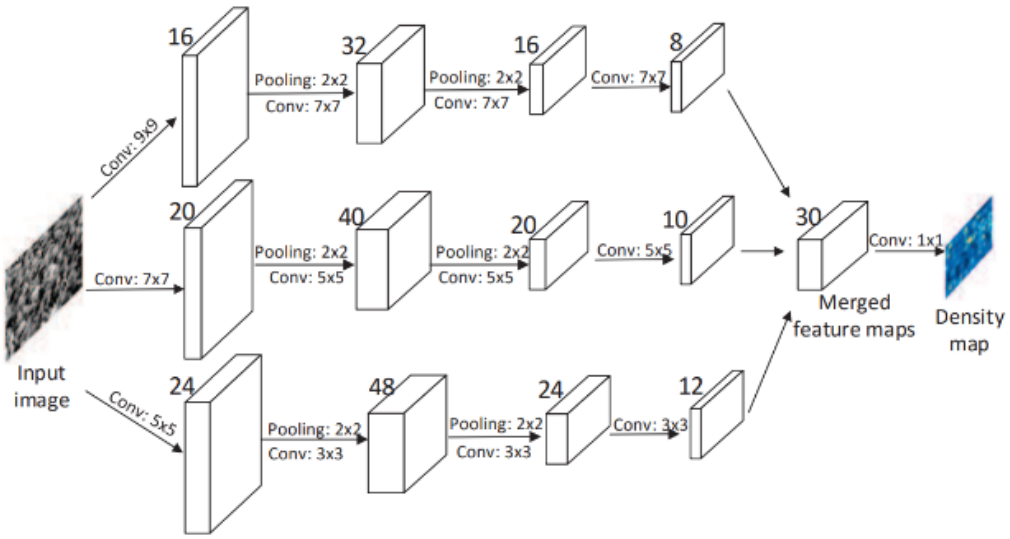
Ground-truth Density Map có thể được sinh bằng cách đặt Gaussian Kernel tại vị trí chứa đầu người. Giả định rằng mỗi đầu sẽ có mối liên hệ tới k đầu lân cận theo phân phối Gauss. Phương sai được tính dựa trên trung bình khoảng cách từ điểm đầu đang xét đến k đầu lân cận.

4.1.2. Density Map Estimation

Density Map Estimation là một Density Map được sinh bằng các model trong quá trình ước tính. Density Map Estimation thường có độ chính xác kém hơn Ground-truth Density Map.

4.2. Multi-Column Convolutional Neural Network

Sử dụng kiến trúc M-CNN để sinh Density Map. Kiến trúc được thiết kế như sau:



Hình 3: Multi-Column Convolutional Neural Network

Do biến dạng phối cảnh, hình ảnh thường chứa các phần đầu có kích thước rất khác nhau, do đó các bộ lọc có trường tiếp nhận có cùng kích thước khó có thể nắm bắt được các đặc điểm của mật độ đám đông ở các tỷ lệ khác nhau. Còn các bộ lọc với các trường tiếp nhận (Receptive Fields) khác nhau (lớn, trung bình, nhỏ), các đặc trưng mỗi column học được tương ứng với sự thay đổi lớn về kích thước đầu do phối cảnh hoặc độ phân giải (Spatial Resolution) khác nhau.

Lớp Fully Connected bị loại bỏ bởi vì không phù hợp với bài toán. Thay vào đó, để ánh xạ các Feature Map ở các lớp tích chập vào Density Map, sử dụng bộ lọc kích thước 1×1.

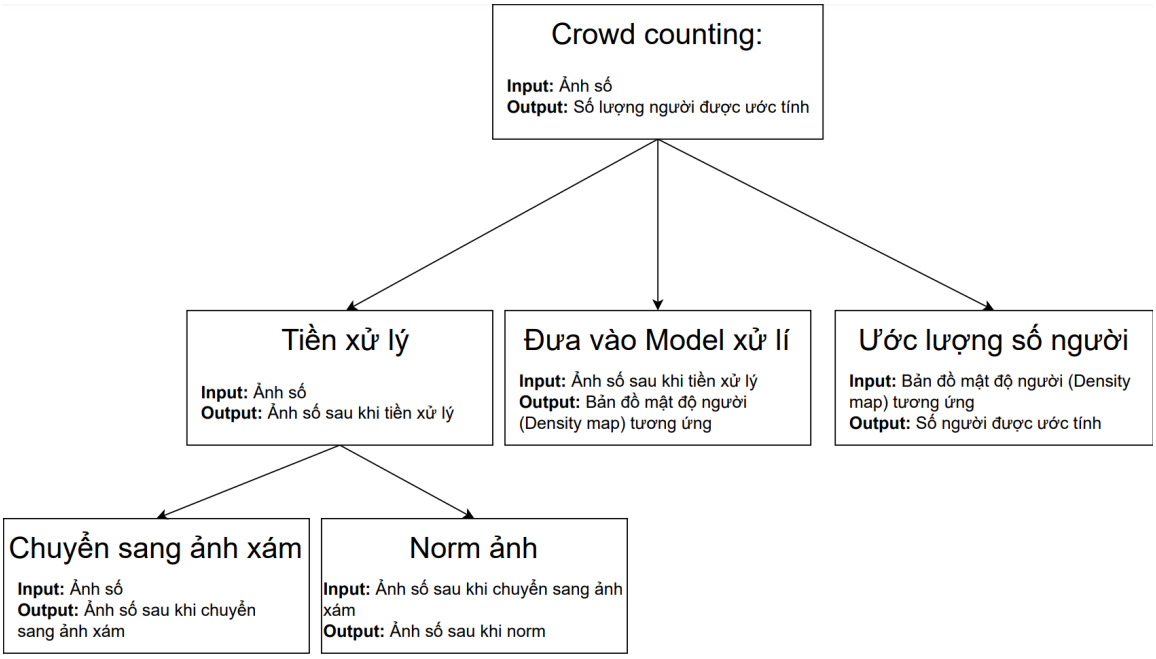
5. Abstraction

Đầu vào là một hình ảnh bất kì, có thể chứa đám đông hoặc không. Ngoài ra, một số ràng buộc cho đầu vào và đầu ra cũng được thiết lập như sau:

- Kích thước input bất kì, chiều cao nằm trong khoảng từ 128 đến 768, chiều rộng nằm trong khoảng từ 128 đến 1024.
- Vấn đề này có thể giải quyết cho nhiều bối cảnh khác nhau, điều kiện thời tiết, mức độ sáng khác nhau, nhưng độ chính xác sẽ bị giảm xuống.

6. Decomposition

Dưới đây là phần Decomposition mà nhóm đã phân chia ra các task để giải quyết vấn đề.



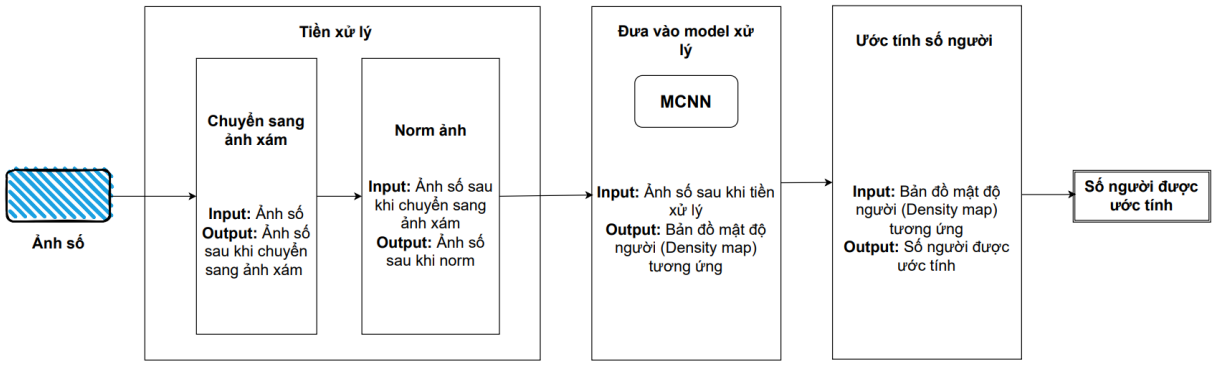
Hình 4: Decomposition

7. Pattern Recognition

Dựa vào input cũng như yêu cầu của output, có thể xác định vấn đề cốt lõi chính là **Density Map Estimation**.

- **Input:** Ảnh xám.
- **Output:** Density map.

8. Giải thuật



Hình 5: Algorithm

9. Độ đo

Đặc trưng của bài toán cần đo độ lệch giữa số lượng đầu người ước tính và thực tế trên tập dữ liệu nên Mean Absolute Error (MAE) và Mean Squared Error (MSE) được sử dụng đồng thời để xem xét tổng thể.

Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Trong đó, y_i là số lượng người thực tế trong ảnh thứ i, $\hat{y}_i$ là số người được ước tính trong ảnh thứ i, N là tổng số lượng ảnh.

10. Thực nghiệm

10.1. Bộ dữ liệu

Các bộ dữ liệu được sử dụng để thực nghiệm là: ShanghaiTech_PartB và Mall. Trong đó:

- Với bộ dữ liệu **ShanghaiTech_PartB**:
  - Bao gồm 716 ảnh, được lấy từ những con phố đông đúc ở các khu đô thị ở Thượng Hải, trong đó có 400 bức ảnh cho thư mục train, 316 bức ảnh cho thư mục test.
  - Với mỗi bức ảnh sẽ có tệp ground-truth định dạng matlab tương ứng có đuôi là .mat. Trong mỗi tệp đó những tâm ở phần đầu mỗi người được đánh dấu theo dạng danh sách các tọa độ (x, y).
- Với bộ dữ liệu **Mall**:
  - Bao gồm 2000 ảnh được trích xuất từ video giám sát của một trung tâm mua sắm.
  - Có một tệp ground-truth định dạng matlab có đuôi là .mat, chứa các danh sách các toạ độ (x, y) là vị trí của các đầu người được đánh dấu của các bức ảnh tương ứng.

Dưới đây là bảng chi tiết hơn cho từng bộ dữ liệu:

Dataset	Kích thước	Số ảnh	Số người tối đa	Số người tối thiểu	Tổng số người
ShanghaiTech Part B	1024×768	716	578	9	88488
Mall	640×480	2000	53	13	62325

10.2. Kết quả thực nghiệm

Với dataset **ShanghaiTech_PartB**, tập train sẽ được chia thành 350 ảnh cho việc train và 50 ảnh cho việc cải tiến mô hình trong quá trình học, còn 316 ảnh ở tập test dùng để kiểm tra và đánh giá mô hình.

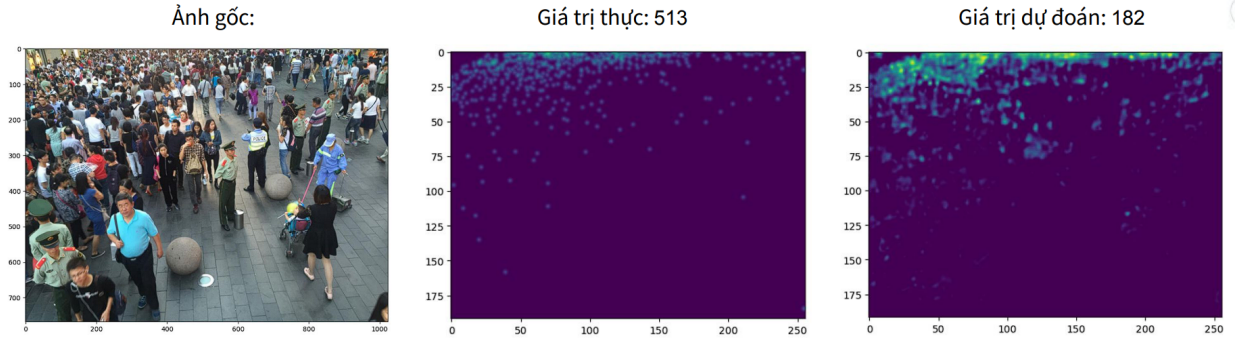
Với dataset **Mall**, 2000 ảnh sẽ được chia như sau: 1400 ảnh được dùng làm tập train, 300 ảnh được dùng làm tập để cải tiến mô hình trong quá trình học, 300 ảnh còn lại được dùng làm tập kiểm tra và đánh giá mô hình.

Về phần cứng, mô hình được chạy thông qua phần mềm Jupyter Notebook được tích hợp trong phần mềm Anaconda trên máy tính Acer Nitro 5 AN515-45 với CPU là AMD Ryzen 5 5600H 512MB và GPU là NVIDIA Geforce GTX 1650 4GB, RAM có dung lượng là 16GB, sử dụng hệ điều hành Windows 11 Home Single Language và trong Anaconda có cài đặt sẵn CUDA để có thể chạy mô hình trên GPU.

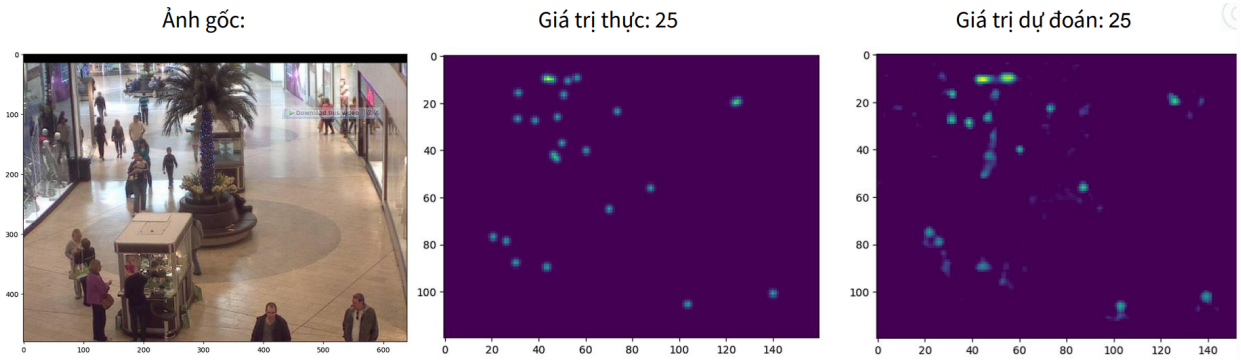
Kết quả thử nghiệm như sau:

Dataset	MAE	MSE
SanghaiTech PartB	45.5	5509.57
Mall	2.09	6.99

Hai ảnh dưới đây được lấy ra từ hai bộ test và được chạy thực nghiệm như sau:



Hình 6: Ảnh từ tập test của bộ dữ liệu ShanghaiTech và kết quả



Hình 7: Ảnh từ tập test của bộ dữ liệu Mall và kết quả

Link code mô hình cho hai bộ dữ liệu: GitHub(nAuTahn).

Trong đó, tệp **Crowd_Counting_MCNN-Mall.ipynb** để chạy trên bộ dữ liệu **Mall**, tệp **Crowd_Counting_MCNN-PartB.ipynb** để chạy trên bộ dữ liệu **ShanghaiTech Part B**.

11. Kết luận và hướng phát triển

11.1. Kết luận

Nhóm đã thực nghiệm với mô hình M-CNN trên hai tập dữ liệu là Mall và ShanghaiTech_PartB, kết quả cho ra đều tạm chấp nhận được. Ngoài ra phương pháp cũng giải

quyết được trong trường hợp ảnh có độ phân giải thấp. Tuy nhiên trường hợp các đầu bị che khuất thì mô hình lại kém hiệu quả. Việc tập dữ liệu chưa đủ nhiều, phần cứng không đủ mạnh cũng khiến cho việc train model khó khăn, tốn nhiều thời gian.

11.2. Hướng phát triển

Trong tương lai, nếu nhóm có đủ tiềm lực về dữ liệu cũng như hệ thống train hiệu suất hơn, thì nhóm có thể hoàn thiện model này tốt hơn.

Tài liệu

[1] M. P. Deisenroth, A. A. Faisal, C. S. Ong, *Mathematics for Machine Learning*, Cambridge University Press, 2020.

[2] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, Yi Ma, *Single-Image Crowd Counting via Multi-Column Convolutional Neural Network*, Institute of Electrical and Electronics Engineers, 2016.

[3] GitHub nhóm: nAuTahn.