

THÔNG TIN CHUNG

- Link YouTube video của báo cáo
(ví dụ: <https://www.youtube.com/watch?v=AWq7uw-36Ng>)
- Link slides
https://github.com/lilalinhlan/ReasearchMethodology/blob/main/250104012_DeCuong_Slide_PPNCCKH.pdf
- Họ và tên: Nguyễn Thùy Linh
- MSSV: 250104012
- Lớp: CS2205.SEP2025
- Tự đánh giá (điểm tổng kết môn): 8/10
- Số buổi vắng: 1



ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI

NGHIÊN CỨU TỔNG QUÁT HÓA TỪ NGUỒN DỮ LIỆU ĐƠN MIỀN CHO BÀI TOÁN ĐẾM NGƯỜI TRONG ẢNH BẰNG PHƯƠNG PHÁP MPCOUNT

TÊN ĐỀ TÀI TIẾNG ANH

EVALUATING SINGLE DOMAIN GENERALIZATION FOR CROWD COUNTING USING MPCOUNT

TÓM TẮT

Bài báo “*Single Domain Generalization for Crowd Counting*” (CVPR 2024) nghiên cứu bài toán đếm số lượng người trong ảnh đám đông dựa trên phương pháp hồi quy bản đồ mật độ (density map regression). Mặc dù đây là hướng tiếp cận phổ biến và hiệu quả, các mô hình dựa trên hồi quy mật độ thường gặp hiện tượng suy giảm hiệu suất nghiêm trọng khi áp dụng cho các miền dữ liệu chưa từng xuất hiện trong quá trình huấn luyện, do ảnh hưởng của hiện tượng dịch chuyển miền dữ liệu (Domain Shift). Trước thực tế đó, nghiên cứu tập trung vào phương pháp tổng quát hóa từ nguồn dữ liệu đơn miền (Single Domain Generalization – SDG), trong đó mô hình chỉ được huấn luyện trên một miền nguồn duy nhất nhưng vẫn phải đảm bảo khả năng hoạt động tốt trên các miền đích khác nhau.

Để giải vấn đề trên, tác giả đề xuất khung phương pháp MPCount với hai thành phần cốt lõi là Attention Memory Bank (AMB) - có nhiệm vụ lưu trữ các biểu diễn mật độ đa dạng, hỗ trợ tái cấu trúc các đặc trưng mang tính bất biến theo miền (domain-invariant features). Và Patch-wise Classification (PC) - một luồng phụ trợ chia hình ảnh thành các vùng nhỏ để thực hiện phân loại có người hoặc không có người, nhằm giảm thiểu hiện tượng mơ hồ trong gán nhãn (label ambiguity) ở cấp độ điểm ảnh.

Kết quả thực nghiệm trên các bộ dữ liệu chuẩn như ShanghaiTech và JHU-Crowd++ cho thấy phương pháp MPCount cải thiện đáng kể độ chính xác so với các phương pháp tiên tiến hiện nay. Đặc biệt, mô hình thể hiện ưu thế rõ rệt trong các kịch bản có phân phối dữ liệu nguồn hẹp, chẳng hạn như điều kiện thời tiết tuyết hoặc sương mù, qua đó khẳng định hiệu quả của hướng tiếp cận SDG trong bài toán đếm người.

GIỚI THIỆU

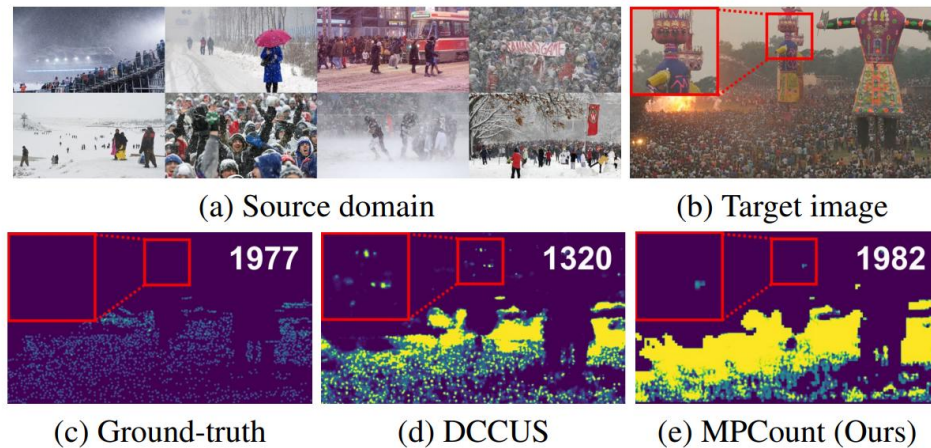
Đếm đám đông là bài toán nhằm ước lượng số lượng người xuất hiện trong một hình ảnh, được ứng dụng rộng rãi cho camera giám sát an ninh công cộng, tổ chức sự kiện quản lý giao

thông và quy hoạch đô thị thông minh. Trong những năm gần đây, các phương pháp tiếp cận dựa trên học sâu, đặc biệt là hồi quy bản đồ mật độ (density map regression), đã trở thành hướng nghiên cứu chủ đạo. Các phương pháp này xây dựng một bản đồ mật độ trong đó tổng giá trị các điểm ảnh tương ứng với số lượng người trong ảnh, từ đó đạt được độ chính xác cao trong điều kiện dữ liệu huấn luyện và dữ liệu kiểm tra có phân phối tương đồng.

Tuy nhiên, phần lớn các mô hình hiện nay đều được xây dựng dựa trên giả định rằng dữ liệu kiểm tra tuân theo cùng một phân phối với dữ liệu huấn luyện. Trong môi trường thực tế, giả định này thường không còn phù hợp do sự thay đổi về góc quay camera, điều kiện thời tiết, ánh sáng hoặc bối cảnh quan sát. Những khác biệt này dẫn đến hiện tượng dịch chuyển miền dữ liệu (Domain shift), khiến hiệu suất của mô hình suy giảm đáng kể khi được triển khai ngoài môi trường huấn luyện. Mặc dù các phương pháp như Domain Adaptation đã được đề xuất nhằm khắc phục vấn đề trên, chúng thường yêu cầu dữ liệu từ miền đích trong quá trình huấn luyện, điều vốn rất khó đáp ứng trong nhiều kịch bản thực tế.

Câu hỏi đặt ra là làm sao để tổng quát hóa từ nguồn dữ liệu đơn miền (Single Domain Generalization – SDG) trở thành một thách thức lớn đối với crowd counting.

Để giải quyết thách thức này, phương pháp MPCount đã được đề xuất như một giải pháp tiếp cận mới cho bài toán SDG trong crowd counting mà không cần chia nhỏ miền. Phương pháp này khai thác cơ chế Content Error Mask (CEM) nhằm loại bỏ các đặc trưng không ổn định liên quan đến domain-related style, kết hợp với hàm mất mát Attention Consistency Loss (ACL) để đảm bảo bộ nhớ chỉ lưu trữ các đặc trưng mang tính bất biến. Nhờ đó, mô hình có khả năng học được biểu diễn tổng quát hơn và cải thiện hiệu suất trên các miền dữ liệu chưa từng xuất hiện trong quá trình huấn luyện.



Hình 1. Áp dụng phương pháp tổng quát hóa từ nguồn dữ liệu đơn miền để xử lý bài toán đếm đám đông

MỤC TIÊU

1. Tổng hợp các kỹ thuật domain generalization hiện có và đánh giá hạn chế của chúng khi áp dụng cho crowd counting.
2. Xây dựng và triển khai khung mô hình MPCount tích hợp cơ chế Attention Memory Bank (AMB) và phân loại Patch-wise Classification (PC) để tối ưu hóa việc trích xuất đặc trưng bất biến và giảm thiểu sai số do gán nhãn không đồng nhất..
3. Đánh giá định lượng hiệu quả của mô hình trên các tập dữ liệu tiêu chuẩn (SHA, SHB, JHU-Crowd++) và so sánh với các phương pháp hiện có để chứng minh khả năng tổng quát hóa trong các điều kiện môi trường thực tế phức tạp

NỘI DUNG VÀ PHƯƠNG PHÁP

1. Nội dung nghiên cứu

Nội dung nghiên cứu được xây dựng bám sát các mục tiêu đề ra, tập trung vào việc giải quyết hai thách thức lớn nhất của đếm đám đông là hồi quy mật độ (density regression) và sự mơ hồ của việc gán nhãn (label ambiguity).



(a) Density Map

(b) PCM

Hình 2. Minh họa về Label ambiguity và cách Patch-wise Classification Map (PCM).

- **Phân tích và hệ thống hóa các kỹ thuật Domain Generalization (DG):**
 - Nghiên cứu các nhóm phương pháp SDG phổ biến hiện nay như: tạo dữ liệu đối nghịch (adversarial data generation), chuẩn hóa đặc trưng (feature normalization) và thiết kế mạng chuyên biệt
 - Đánh giá hạn chế của phương pháp phân chia tiểu miền (sub-domain division) khi đối mặt với phân phối dữ liệu nguồn hẹp (narrow source distribution)
- **Xây dựng và triển khai khung mô hình MPCount:**
 - **Cấu trúc Encoder-Decoder:** Sử dụng VGG16-BN làm bộ trích xuất đặc trưng nền tảng
 - **Cơ chế Attention Memory Bank (AMB):** Triển khai AMB gồm 1024 vector để tái cấu trúc đặc trưng dưới dạng tổ hợp tuyến tính, giúp mô hình biểu diễn được các giá trị mật độ liên tục thay vì các phân loại rời rạc.

- **Content Error Mask (CEM):** Loại bỏ các thành phần đặc trưng bị ảnh hưởng bởi domain-related style bằng cách so sánh sự sai khác giữa các đặc trưng đã được chuẩn hóa (Instance Normalization)
- **Patch-wise Classification (PC):** Xây dựng luồng phụ trợ chia ảnh thành lưới 16x16 để phân loại vùng có người/không có người, từ đó lọc nhiễu cho bản đồ mật độ
- **Đánh giá định lượng và phân tích hiệu năng:**
 - Thực hiện các thí nghiệm so sánh chéo giữa các tập dữ liệu (ví dụ: Huấn luyện trên SHA, kiểm tra trên SHB hoặc QNRF)
 - Thực hiện **Ablation Study** (nghiên cứu thành phần) để làm rõ đóng góp của từng module AMB, CEM, và PC đối với độ chính xác tổng thể.

2. Phương pháp nghiên cứu

Đề tài kết hợp giữa nghiên cứu lý thuyết và thực nghiệm trên máy tính dựa trên phương pháp học sâu (Deep Learning).

- **Phương pháp nghiên cứu tài liệu:**
 - Thu thập và phân tích các bài báo từ các hội nghị hàng đầu như CVPR, ICCV, AAAI trong 5 năm gần đây để xây dựng nền tảng lý thuyết
- **Phương pháp thực nghiệm:**
 - **Cài đặt:** Sử dụng thư viện PyTorch, trình tối ưu hóa AdamW và chiến lược học tập OneCycleLR.
 - **Dữ liệu:** Sử dụng các bộ dữ liệu công khai (Public datasets) gồm ShanghaiTech Part A & B, UCF-QNRF và JHU-Crowd++. Đặc biệt chú trọng vào các nhãn instance-level như "Snow" và "Fog" để kiểm tra khả năng chịu đựng sai lệch miền.
- **Phương pháp đánh giá:**
 - Sử dụng hai chỉ số thống kê tiêu chuẩn để đo lường sai số giữa giá trị dự đoán (c^{\wedge}) và giá trị thực tế (c):

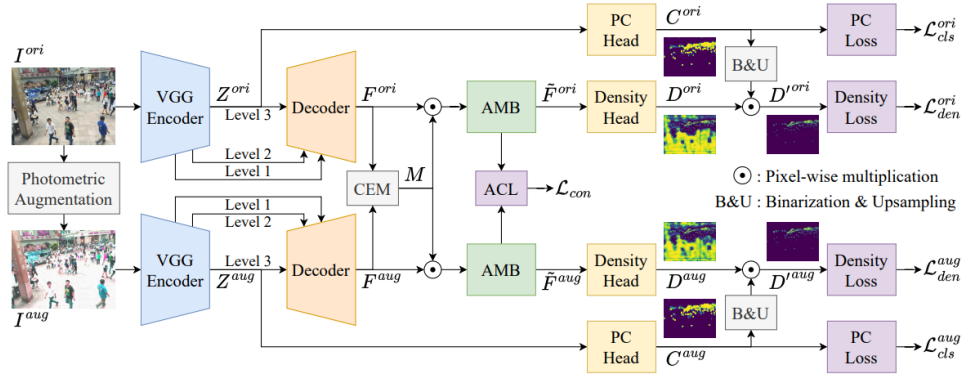
MAE (Mean Absolute Error):

$$MAE = \frac{1}{N} \sum_{i=1}^N |c_i - \hat{c}_i|$$

MSE (Mean Squared Error):

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (c_i - \hat{c}_i)^2}$$

- Đánh giá khả năng tổng quát hóa bằng cách so sánh kết quả với các mô hình SOTA như DCCUS hoặc các phương pháp Domain Adaptation để xác định vị thế của MPCount



Hình 3. Quy trình huấn luyện tổng thể MPCount

- **Photometric Augmentation:** Hình ảnh gốc được đi qua các phép biến đổi hình học và màu sắc để tạo ra phiên bản tăng cường
- **VGG Encoder:** Bộ mã hóa trích xuất các đặc trưng từ hình ảnh
- **CEM:** Lọc bỏ các thông tin liên quan đến đặc điểm riêng biệt của từng miền, chỉ giữ lại các đặc trưng nội dung bất biến
- **AMB:** Tái cấu trúc các đặc trưng để dự đoán bản đồ mật độ một cách ổn định.
- **PC Head:** Nhánh phụ trợ thực hiện phân loại theo từng mảng ảnh để xác định vùng nào có người, giúp loại bỏ các dự đoán sai ở vùng nền
- **ACL:** Hàm mất mát đảm bảo sự nhất quán về sự chú ý giữa ảnh gốc và ảnh tăng cường

KẾT QUẢ MONG ĐỢI

Sau khi hoàn thành quá trình phân tích và thực nghiệm lại bài báo này, nghiên cứu dự kiến đạt được các kết quả cụ thể như sau:

1. Về mặt làm chủ công nghệ và lý thuyết:

- Làm rõ và giải thích được cơ chế vận hành của AMB trong việc lưu trữ các đại diện đặc trưng bất biến thay vì các giá trị mật độ đơn thuần.
- Hiểu rõ quy trình xây dựng CEM để phân tách giữa thông tin domain-related style và content trong ảnh đám đông.

2. Về mặt thực nghiệm và kiểm chứng:

- Tái lập thành công khung mô hình MPCount trên môi trường thực nghiệm cá nhân dựa trên mã nguồn công khai của tác giả.

- Kiểm chứng được tính đúng đắn của các kết quả mà bài báo đã công bố trên các bộ dữ liệu tiêu chuẩn như ShanghaiTech và JHU-Crowd++.
- Xác nhận được vai trò của nhiệm vụ phụ trợ phân loại mảng (PC) trong việc giúp mô hình vận hành ổn định hơn khi loại bỏ các vùng nền gây nhiễu.

3. Về mặt phân tích và đánh giá:

- Xây dựng được các biểu đồ so sánh MAE và MSE để thấy rõ sự khác biệt giữa mô hình đầy đủ và các mô hình bị lược bỏ thành phần (Ablation Study).
- Đưa ra được các nhận xét khách quan về ưu điểm và hạn chế của mô hình khi xử lý các trường hợp dữ liệu có mật độ cực cao hoặc điều kiện ánh sáng thay đổi đột ngột.
- Báo cáo tổng kết đầy đủ về khả năng ứng dụng thực tế của MPCount trong bài toán giám sát đám đông tại Việt Nam dựa trên các kịch bản miền chưa từng thấy (unseen scenarios)

TÀI LIỆU THAM KHẢO

[1] Zhuoxuan Peng, S.-H. Gary Chan.

"Single Domain Generalization for Crowd Counting."

In proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

[2] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma.

"Single-Image Crowd Counting via Multi-Column Convolutional Neural Network."

In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[3] Vishwanath A. Sindagi, Rajeev Yasarla, and Vishal M. Patel.

"JHU-CROWD++: A Large-Scale Benchmark for Crowd Counting."

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

[4] Karen Simonyan and Andrew Zisserman.

"Very Deep Convolutional Networks for Large-Scale Image Recognition."

In International Conference on Learning Representations (ICLR), 2015.