**1. Nó làm gì? (What does it do?)**

Bài báo cung cấp một tổng quan toàn diện về các phương pháp **phát hiện người, theo dõi, nhận dạng cá nhân và nhận diện khuôn mặt**, tập trung vào các kỹ thuật học máy (ML) và học sâu (DL). Nó tổng hợp **142 bài báo** nổi bật trong lĩnh vực này và sử dụng phương pháp **PRISMA** để chọn lọc tài liệu.

Các xu hướng chính trong lĩnh vực này, bao gồm:

* **Chuyển dịch từ phương pháp truyền thống (HOG, SVM, DPM, Viola-Jones)** sang các mô hình học sâu như **Faster R-CNN, YOLO, Mask R-CNN**.
* **Những thách thức hiện tại** như độ chính xác trong điều kiện môi trường thay đổi, tổng quát hóa mô hình, và vấn đề đạo đức & quyền riêng tư.

**2. Công nghệ sử dụng (Technologies Used)**

Bài báo chia thành các kỹ thuật chính theo từng nhiệm vụ:

**Phát hiện người (Person Detection)**

* **Phương pháp truyền thống**: HOG-SVM, Viola-Jones, DPM
* **Mô hình học sâu (DL)**:
  + **Faster R-CNN** (độ chính xác cao nhưng chậm)
  + **SSD (Single Shot MultiBox Detector)** (tốc độ nhanh hơn nhưng độ chính xác thấp hơn)
  + **YOLO (You Only Look Once)**
  + **Mask R-CNN** (phát hiện chính xác hơn bằng segmentation)
* **Kỹ thuật bổ trợ**: Pose estimation (OpenPose, UniPose+), Instance & Semantic Segmentation.

**Theo dõi người (Person Tracking)**

* **Các kỹ thuật cũ**: Kalman filter, Particle filter, Mean Shift
* **Học sâu**:
  + **DeepSORT**: Kết hợp giữa **YOLO + Kalman filter**
  + **FairMOT**: Kết hợp giữa phát hiện & theo dõi trong một mạng duy nhất
  + **ByteTrack**: Theo dõi hiệu quả ngay cả khi có nhiều người

**Nhận dạng cá nhân (Person Identification)**

* **Phương pháp truyền thống**: Ensemble RankSVM, SDALF
* **Học sâu**:
  + **ResNet, Inception, MobileNet** (trích xuất đặc trưng)
  + **Triplet Loss, Contrastive Loss** (học metric cho nhận dạng)
  + **Transformer-based Multi-Grained Feature (TMGF)**

**Nhận diện khuôn mặt (Face Recognition)**

* **Phương pháp truyền thống**: Eigenfaces, Fisherfaces, LBPs
* **Học sâu**:
  + **DeepFace** (Facebook, 2014)
  + **FaceNet** (Google, sử dụng Triplet Loss)
  + **ArcFace** (sử dụng Angular Margin Loss)
  + **VGGFace** (VGGNet-based)

**3. Dataset Process**

Bài báo liệt kê nhiều bộ dữ liệu được sử dụng cho từng tác vụ:

**Phát hiện người**

* **COCO (200,000 ảnh)**: Đánh giá theo mAP@50
* **PeopleSansPeople (500,000 ảnh 3D)**: Dữ liệu tổng hợp để bảo vệ quyền riêng tư
* **AI City Challenge**: Nhận diện người lái xe vi phạm

**Theo dõi người**

* **MOTChallenge (17,757 frames)**: Chuẩn benchmark cho multiple object tracking
* **SportsMOT (150,000 frames)**: Theo dõi vận động viên
* **PoseTrack (66,374 frames)**: Dành riêng cho pose estimation

**Nhận dạng cá nhân**

* **Market-1501**: 32,668 hình ảnh từ 1,501 danh tính
* **DukeMTMC-reID**: Re-ID giữa các camera khác nhau

**Nhận diện khuôn mặt**

* **LFW (13,233 ảnh)**: Benchmark phổ biến nhất
* **CelebA (200,000 ảnh)**: Nhận diện khuôn mặt người nổi tiếng
* **VGGFace (2.6 triệu ảnh)**: Dữ liệu lớn phục vụ deep learning

**4. Kết quả đạt được (Results)**

**📍 Đọc trong phần: Section 6 (Trang 24-25)**

* **Các mô hình học sâu đã đạt độ chính xác cao hơn so với phương pháp truyền thống**. Ví dụ:
  + **YOLOv4**: 43.0% AP trên COCO
  + **Faster R-CNN**: 27.2% AP trên COCO
  + **DeepSORT**: MOTA = 61.4% trên MOT16
  + **FaceNet**: 97% Accuracy trên LFW
  + **ArcFace**: 98% Accuracy trên LFW
* **Các ứng dụng thực tế**:
  + **Giám sát giao thông & thành phố thông minh** (camera thông minh)
  + **Nhận diện khuôn mặt trong thanh toán & an ninh** (ví dụ: Face Pay ở Almaty Metro)
  + **Hệ thống phát hiện xâm nhập** (giám sát nhà ở bằng AI)

👉 **Kết luận:** Các mô hình hiện tại hoạt động tốt nhưng vẫn có nhiều thách thức về **tính công bằng**, **khả năng tổng quát hóa**, và **độ bền trước tấn công đối nghịch**.

**5. Khoảng trống của bài báo (Research Gaps)**

**📍 Đọc trong phần: Section 7 (Trang 25-26)**

* **Khả năng giải thích mô hình (Explainability & Interpretability)**: DL vẫn là hộp đen.
* **Tổng quát hóa kém (Domain Adaptation)**: Mô hình chưa hoạt động tốt ở môi trường mới.
* **Khả năng chống lại tấn công đối nghịch (Adversarial Robustness)**: Hệ thống dễ bị tấn công giả mạo.
* **Hạn chế về dữ liệu (Bias & Privacy Issues)**: Dữ liệu có bias về chủng tộc, giới tính.
* **Chi phí tính toán cao**: Các mô hình deep learning yêu cầu GPU mạnh, khó triển khai thực tế.

**6. Kết luận**

**Học sâu đã thống trị lĩnh vực nhận dạng & theo dõi người**, nhưng vẫn còn **nhiều thách thức về công bằng AI, tổng quát hóa và chi phí tính toán**.

Cần nghiên cứu thêm về **học không giám sát, học liên tục và AI có thể giải thích**.

Các ứng dụng như **thành phố thông minh, an ninh, giao thông** đang hưởng lợi từ công nghệ này.