**Detection and Tracking of People in a Dense Crowd through Deep Learning Approach**

*Muhammad Firdaus bin Mohamed Badauraudine*

ABSTRACT

There are many incidents involving crowds such as crowd crushing and stampedes. These incidents include the Hillsborough Disaster, the stampede at Kanjuruhan Stadium and an incident knows as a "crush and stampede" that either suffocated or crushed over 2000 pilgrimages during the annual Hajj pilgrimage. These incidents mainly occur due to poor crowd management, which is done manually most of the time, increasing room for human error. Among the human errors are more people in a specified area compared to the allowed people in that area. Hence, this research studies upon the approach of deep learning to detect and track people in a crowd to observe the number of people in a crowd at a given time. To automate the detection and tracking of people in a crowd, artificial intelligence was implemented. The current research reflects upon the detection and tracking of people in a dense crowd through a topological method. It also reflects upon the accuracy of the detection and tracking of people in a crowd by taking re-identification of the individual as a main factor. To ease the process of tracking and detection, auto annotation of the targets in the datasets is also being researched. To further understand how these methods are implemented for dense crowd tracking and detection, a systematic literature review (SLR) is carried out. This paper presents an SLR to further analyse similar works of this research as well as what can be improved upon for the current research to fill in the gaps of the previous research. The review covers articles published from 2019 to 2024. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used to carry out the systematic review where the guidelines are implemented for the data collection and analysis of 4384 articles studied across five different databases. From 1889, 13 are the primary studies that met the inclusion criteria. The primary studies will be used as the staple to be analysed and categorised to answer related questions relating to the accuracy of the tracking and detection of people in a dense crowd. This SLR can be used to provide references for research in artificial intelligence and those seeking to implement it alongside auto annotation and re-identification in the tracking and detection of people in a dense crowd.

KEYWORDS *topology, dense crowd, track, detect, deep learning, re-identification, annotation*

Introduction

Crowd detection by Multiple Head Tracking (MHT) is currently one of the most essential and challenging research topics in the computer vision community. Due to the common availability of high-quality low-cost video cameras and considering the inefficiency of manual surveillance and monitoring systems, automated video surveillance is now essential for today’s crowded and complex environments. Accurate information about numbers plays a vital role in operational and security efficiencies for monitoring, controlling, and protecting crowds. The counting and tracking of many persons pose as a challenge due to several reasons that are occlusions, the constant displacement of people, different perspectives and behaviours, varying illumination levels, and because, the bigger the crowd, the lower the allocation of pixels per person.

However, the counting and tracking of people in crowds is important to be implemented. This is due to public domains such as stadiums, airports, and, and even religious gathering areas tend to be difficult to track the number of people in crowds due to its density. This difficulty has proven to be deadly such as the Hillsborough disaster on 15th April 1989, the human “crush and stampede” event that occurred for the Hajj pilgrims on 24th September 2015, and the human crush at Kanjuruhan Stadium on 1st October 2022. Due to these incidents, a question that usually pops up is how this can be avoided or lesson. One of the solutions that was being researched upon was crowd detection using MHT.

Crowd detection by MHT is currently one of the most essential and challenging research topics. However, methods developed by different researchers only produce satisfactory results in sparse crowd setting for crowd detection and tracking. Hence, this area of research needs further elaboration to implement this area of research in dense crowd benchmarks.

It is important to analyse the methods that can be used for tracking and detecting people in a crowd. It is also important to take a step back and understand the flow of how the tracking and detection was originally to how much it has evolved so that the current research can understand more on how the gaps were filled with each new evolution. Similar works to this research need to be researched to understand the methods used by different research to detect and track people in crowds and their gaps for the current research to fill them to be novel research.

There have been many review studies focusing on topology to track or detect people in a crowd. However, some of those studies sometimes only detect and does not track. Another limitation is the accuracy of tracking and detection. The output of the studies can determine the accuracy of the detection and tracking of people in a dense crowd. For example, a topological output will have a higher accuracy in the study compared to a density map output. Since this area of study deals with the physical security of humans, it is important to maintain a high accuracy of tracking and detection. Current architectures implemented in several papers such as YOLO was developed to detect objects in normal situations and not in a highly-densed environment. Another contribution to the accuracy of the output is the re-identification of people in a dense crowd. This is because if the same person exited the tracking parameter and rejoined, they can be counted as another person and increase the crowd count. Hence, it is important to further filter the papers to review research that are more similar to the current research.

In this paper, our aim is to help other researchers by making an SLR of related studies to using deep learning in tracking and detection of people in a dense crowd. The SLR’s purpose is to research and compare how different studies approach tracking and detection involving dense crowds using deep learning approaches to find out which approach is best suited to answer the study’s research question. The SLR encompass studies in the last five years that are from 2019 to 2024. The main contributions of this systematic review can be summarised as follows:

1. Searching articles from various databases with specific keywords that are broadly related to current research. Mainly, the research must include tracking in a crowd using deep learning. The reason why only tracking is used as the keyword and not detection is because to track, the experiment must detect the people first. Hence, both tracking and detection goes together. There are 4384 papers collected using the specified keywords from seven databases that are ACM Digital Library, Wiley Online Library, SpringerLink, ScienceDirect, IEEExplore, and PubMed. The research carried out must have a published date within the past five years' timeframe.

2. A large-scale literature review is completed using the PRISMA process methodology. The current research proposes three research questions. Firstly, the research questions were formulated. Then, the related studies were collected, the inclusion and exclusion criteria were defined, and from there, the primary studies focused on the related topics were identified with an acceptable quality score.

3. A thorough analysis has been implemented in the primary studies. Each reviewed paper is based on research questions, involving the current methods of detection and tracking of people in a dense crowd using deep learning approach and whether it is applicable in public places. The aim of this is to provide researchers with methods to track and detect people in dense crowds and how their output will be to assist them more in their research.

The remainder of this paper is structured as follows:

1. Section 2: The research methodology involving research questions and research protocols

2. Section 3: The results of the literature review to provide answers and discussions to the initial research questions.

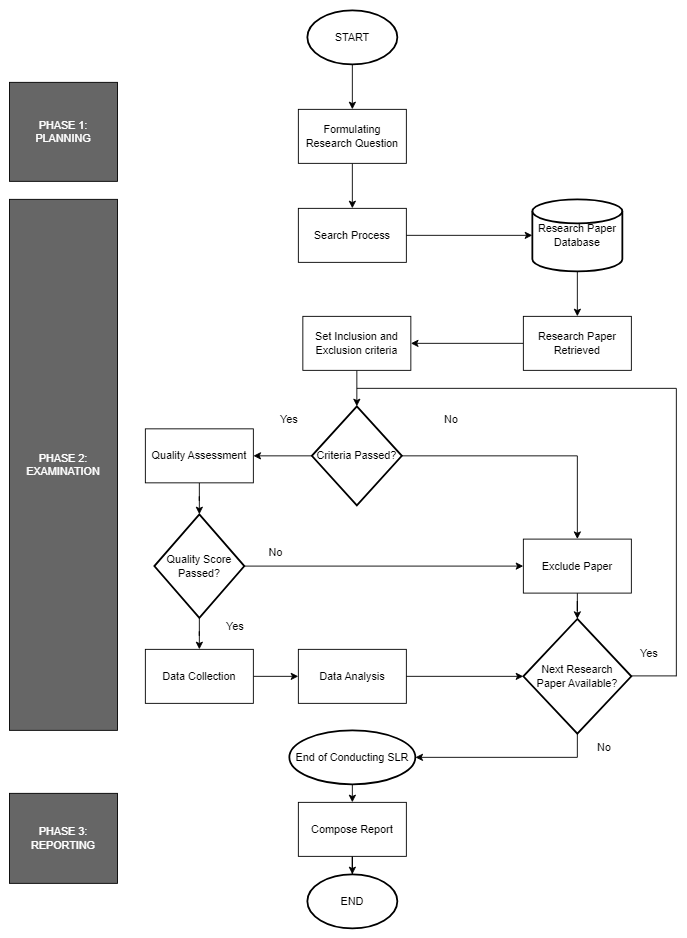
3. Section 4: The limitations of the study

4. Section 5: The conclusion of the research

2. Methodology

The methodology used for the current SLR follows the original guidelines proposed by S. Keele in 2007 in the paper “Guidelines for Performing Systematic Literature Reviews in Software Engineering”. The figure below shows the methodology’s flowchart.

FIGURE 1. SLR’s Methodology



Based on the flowchart, there are 6 main phases for the methodology

A. Planning

The planning phase is where the actions required to achieve the research objectives are decided. In this study, the objectives are related to the topic of the methods and algorithms used in the tracking and detection of people in a dense crowd as well as the accuracy of the output and what datasets are used for the experiment. The keywords to be used to search for papers in the five databases were also planned during this stage. Hence, this step is the foundation of the SLR’s methodology.

B. Formulation of Research Questions

The research questions were discussed and laid out during this phase. The main questions proposed were related to the research study’s output and what datasets were used for the research study’s experiment. The method and architecture used for the detection and tracking of people in a dense crowd also serves as questions needed to be answered as different research provides different methods and algorithms that might not be as accurate as other studies and not suitable for dense crowds. The annotation method is also considered to further research on auto annotation of the individuals in the video frames of the dataset used. Some research focuses on detection only and not tracking. Due to the reasons stated, the research questions formulated are:

RQ1: What method will be used for detecting and tracking people in a dense crowd

RQ2: What method is used for the annotation of individuals in a dense crowd

RQ3: How accurate is the detection and tracking of people in a dense crowd

C. Search Process

In this subsection, each article to be used for comparing similar works is identified. The purpose of this phase is to extract relevant experimental studies on the tracking and detection of people in a dense crowd and if they can be implemented in public places. The databases used to get these articles are IEEExplore, ACM Digital Library, Scopus, Wiley Online Library, and PubMed. The original keywords planned during the planning phase is used to search for related articles in these databases but with a few adjustments. For example, instead of using “DENSE CROWD”, the keyword used will be “CROWD“. This is due to the number of papers being retrieved if the query is too strict. Some research papers might not use “DENSE CROWD” but they use different names to signify the same meaning. The papers will be more accurate to the current study in the future steps, especially the quality assessment step. Publish or Perish is used to obtain the papers and compile them as a CSV file. These papers will be filtered in the further stages. Table 1 shows the keywords used for the search query and articles gotten from each repository.

TABLE 1. Search query used for each database and number of papers gotten from them.

|  |  |  |
| --- | --- | --- |
| Database | Search Query | Number of Papers |
| IEEExplore | *("CROWD") AND ("DEEP LEARNING") AND ("TRACK" OR "TRACKING")* | 99 |
| ACM Digital Library | *("CROWD") AND ("DEEP LEARNING") AND ("TRACK" OR "TRACKING")* | 1170 |
| Scopus | *("CROWD") AND ("DEEP LEARNING") AND ("TRACK" OR "TRACKING")* | 195 |
| Wiley Online Library | *("CROWD") AND ("DEEP LEARNING") AND ("TRACK" OR "TRACKING")* | 420 |
| PubMed | *("CROWD") AND ("DEEP LEARNING") AND ("TRACK" OR "TRACKING")* | 5 |

D. Inclusion and Exclusion Criteria.

The studies included in the SLR were based on specific criteria. This is to filter out articles into papers that complement the current research. If the papers don’t meet the criteria, they will be excluded and filtered out of the study. Among the most important criteria is that it must be within the appropriate timeframe, which is the past five years of the current studies, to make sure that they are still relevant to the current day and age. Another important factor is that the paper must be written in English. Papers that do not meet these criteria are filtered out. The filtering occurs step by step to achieve higher accuracy in filtering the papers. Table 2 refers to the inclusion and exclusion criteria and the number of papers excluded.

TABLE 2. Inclusion and Exclusion criteria and papers excluded.

|  |  |  |
| --- | --- | --- |
| Inclusion Criteria | Exclusion Criteria | Number of Papers Excluded |
| Articles must be unique | There are duplicated articles | 140 |
| Paper must be published within 5 years of when the SLR is written (2019 to 2024). | Paper is not published within 2019 to 2024 | 52 |
| Have an abstract related to the current research | No abstract or abstract not related to current research | 71 |
| Paper must be in English | Paper is in a language other than English | 2 |
| Articles’ titles must relate to the current research | Articles not related to current research. Focus is on crowds. If crowd is not present, the article is excluded | 1497 |
| Have a DOI or URL | Doesn’t have both URL and DOI | 11 |
| The same article could still be obtained | The same article cannot be found through the search engine used | 2 |
| The article can be fully accessed | The article is not open-accessed even when using tools such as Zotero and Sci-Hub | 26 |

E. Quality Assessment

Table 3 shows the quality assessment questions.

TABLE 3. Quality Assessment Questions

|  |  |
| --- | --- |
| Quality Assessment | Respective RQ |
| QA 1: The research uses topological localisation as its deep learning approach | Q1 |
| QA 2: The crowd involved in the study is a dense crowd | Q1 |
| QA 3: The research uses auto annotation for tracking the individuals in a dense crowd | Q2 |
| QA 4: One trained dataset is used on different test datasets | Q2 |
| QA 5: Re-identification of the same person is implemented in the tracking | Q3 |

These questions are to find research papers that can help answer the research questions proposed at the start of the SLR. The quality assessment questions are formulated similarly to the current study such as finding papers that implement dense crowds and topological method.

The studies filtered from the previous step were further filtered through the quality assessment criteria. The main reason for this is although the papers gathered are somewhat related to the current topic, they are not able to help answer the research questions proposed during the research questions’ formulation phase. The quality assessment phase is executed to filter out the best of the best papers that can help answer the research questions proposed in this SLR and extract data from the remaining research studies. The quality assessment checklist is as shown in Table 3.

The quality criteria are as follows in Table 4.

TABLE 4. Quality Assessment Scoring Criteria

|  |  |
| --- | --- |
| Quality Assessment Scoring | Score |
| The author(s) demonstrated a clear, detailed and explicit explanation on the answers for the specific RQ | HIGH = H = 1 |
| The author(s) provided a general, non-detailed, and non-explicit explanation on the answers for the specific RQ | MEDIUM = M = 0.5 |
| The author(s) either provided no or very few technical details on answering the specific RQ | LOW = L = 0 |

The remaining research papers were evaluated against the quality assessment questions shown in Table 3. If the score is at least 3, then the paper is accepted in this SLR. Table 5 shows the current papers, the QAs, and their score.

TABLE 5. Papers, the QAs they answer, and their score

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Paper Title | QA answered (Score) | QA Score | Total Score | Include/Exclude |
| Pedestrian Tracking Algorithm for Dense Crowd based on Deep Learning | 1  2  3  4  5 | 0  1  0  0  0.5 | 1.5 | Exclude |
| Simulation of Evacuating Crowd Based on Deep Learning and Social Force Model | 1  2  3  4  5 | 0  1  0  1  0.5 | 2.5 | Exclude |
| Deep Learning and One-class SVM based Anomalous Crowd Detection | 1  2  3  4  5 | 0  0  0  0  0.5 | 0.5 | Exclude |
| Pedestrian Detection and Feedback Application Based on YOLOv5s and DeepSORT | 1  2  3  4  5 | 0  1  0  0.5  1 | 2.5 | Exclude |
| Real-Time Pedestrian Detection using YOLO | 1  2  3  4  5 | 0  1  0  0  1 | 2 | Exclude |
| A Review on Unconstrained Real-Time Rotation-Invariant Face Detection | 1  2  3  4  5 | 0.5  0  0  0  0 | 0.5 | Exclude |
| Selective face de-identification scheme using multiple face recognition and classification techniques | 1  2  3  4  5 | 0.5  0  0  0  1 | 1.5 | Exclude |
| Learning how to analyse crowd behaviour using synthetic data | 1  2  3  4  5 | 0  1  1  1  0 | 3 | Include |
| An Overview of Crowd Counting on Traditional and CNN-based Approaches | 1  2  3  4  5 | 0  0.5  0  0.5  0.5 | 1.5 | Exclude |
| People Counting Based on Multi-scale Region Adaptive Segmentation and Depth Neural Network | 1  2  3  4  5 | 0  0  0  1  0 | 1 | Exclude |
| AcousticID: Gait-based Human Identification Using Acoustic Signal | 1  2  3  4  5 | 0  0  0.5  0  0.5 | 1 | Exclude |
| Person Tracking Using Ankle-Level LiDAR Based on Enhanced DBSCAN and OPTICS | 1  2  3  4  5 | 0.5  0.5  0  0  0.5 | 1.5 | Exclude |
| Deep Learning-Based Crowd Scene Analysis Survey | 1  2  3  4  5 | 0  0.5  0  0  0 | 0.5 | Exclude |
| Tracking Hundreds of People in Densely Crowded Scenes With Particle Filtering Supervising Deep Convolutional Neural Networks | 1  2  3  4  5 | 0  1  1  1  0 | 3 | Include |
| People Tracking System Using DeepSORT | 1  2  3  4  5 | 0  1  0  0.5  0.5 | 2 | Exclude |
| Performance Comparison and Analysis for Large-Scale Crowd Counting Based on Convolutional Neural Networks | 1  2  3  4  5 | 0  1  0  0  0.5 | 2 | Exclude |
| Application of Cognitive Computing for Smart Crowd Management | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| REGROUP: A Robot-Centric Group Detection and Tracking System | 1  2  3  4  5 | 0  0  0.5  0  1 | 1.5 | Exclude |
| DenseCAvoid: Real-time Navigation in Dense Crowds using Anticipatory Behaviors | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| The Indoor People Tracking and Counting System | 1  2  3  4  5 | 1  0  0  0.5  1 | 2.5 | Exclude |
| Informative scene decomposition for crowd analysis, comparison and simulation guidance | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Pedestrian motion in simulation applications using deep learning | 1  2  3  4  5 | 0  1  0.5  0  0 | 1.5 | Exclude |
| Crowd Abnormal Behavior Detection Combining Movement and Emotion Descriptors | 1  2  3  4  5 | 0  1  1  1  0 | 3 | Include |
| Informative scene decomposition for crowd analysis, comparison and simulation guidance | 1  2  3  4  5 | 0  1  1  0.5  0 | 2.5 | Exclude |
| 3DGCN: 3-Dimensional Dynamic Graph Convolutional Network for Citywide Crowd Flow Prediction | 1  2  3  4  5 | 0.5  0  0  1  0 | 1.5 | Exclude |
| A sound-based crowd activity recognition with neural network based regression models | 1  2  3  4  5 | 0  0  1  1  0 | 2 | Exclude |
| A Survey on Device-free Indoor Localization and Tracking in the Multi-resident Environment | 1  2  3  4  5 | 0.5  1  0  0  0 | 1.5 | Exclude |
| Crowd activity recognition in live video streaming via 3D-ResNet and region graph convolution network | 1  2  3  4  5 | 0.5  1  0  0.5  0.5 | 2.5 | Exclude |
| Enriched and discriminative convolutional neural network features for pedestrian re-identification and trajectory modeling | 1  2  3  4  5 | 0  0  0.5  1  1 | 2.5 | Exclude |
| Algorithms for Microscopic Crowd Simulation: Advancements in the 2010s | 1  2  3  4  5 | 0.5  0.5  0.5  1  0 | 2 | Exclude |
| Transfer learning-based online multiperson tracking with Gaussian process regression | 1  2  3  4  5 | 0  1  0  1  0 | 2 | Exclude |
| Fast intensive crowd counting model of Internet of Things based on multi-scale attention mechanism | 1  2  3  4  5 | 0  1  0.5  0.5  1 | 3 | Include |
| Passenger Flow Statistics Algorithm of Scenic Spots Based on Multi-Target Tracking | 1  2  3  4  5 | 0  1  0  0.5  1 | 2.5 | Exclude |
| UUCT - HyMP: Towards Tracking Dispersed Crowd Groups from UAVs | 1  2  3  4  5 | 0.5  0  1  1  1 | 3.5 | Include |
| Multi-Scale Occluded Pedestrian Detection Based on Deep Learning | 1  2  3  4  5 | 1  1  0.5  1  0.5 | 4 | Include |
| Crowd Evacuation Simulation Using Hierarchical Deep Reinforcement Learning | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Automated Pedestrian Tracking Based on Improved ByteTrack | 1  2  3  4  5 | 0  1  0  0  1 | 2 | Exclude |
| Group Tracking for Video Monitoring Systems: A Spatio-Temporal Query Processing Approach | 1  2  3  4  5 | 0  1  0  1  0 | 2 | Exclude |
| STP-TrellisNets: Spatial-Temporal Parallel TrellisNets for Metro Station Passenger Flow Prediction | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| Scene invariant crowd counting using multi-scales head detection in video surveillance | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| Practical model with strong interpretability and predictability: An explanatory model for individuals' destination prediction considering personal and crowd travel behavior | 1  2  3  4  5 | 0  0.5  0  0  0 | 0.5 | Exclude |
| Research of Pedestrian Object Tracking Algorithms | 1  2  3  4  5 | 0  0.5  0  0  1 | 1.5 | Exclude |
| People Tracking and Counting using Jetson TX2 Kit with Tracking Algorithm | 1  2  3  4  5 | 0  0.5  0.5  0  0 | 1 | Exclude |
| Locating people in Real-World for Assisting Crowd Behaviour Analysis Using SSD and Deep SORT Algorithm | 1  2  3  4  5 | 0  0.5  0.5  0.5  0 | 1.5 | Exclude |
| Dynamic Gallery for Real-Time Multi-Target Multi-Camera Tracking | 1  2  3  4  5 | 0  0.5  0  0.5  1 | 2 | Exclude |
| Covid-19 crowd detection and alert system using image processing | 1  2  3  4  5 | 0  0  0  1  0 | 1 | Exclude |
| Collision-Line Counting Method Using DeepSORT to Count Pedestrian Flow Density and Hungary Algorithm | 1  2  3  4  5 | 0  0  0  0  0 | 0 | Exclude |
| Deep Learning in Smart Video Surveillance for Crowd Management: A Systematic Literature Review | 1  2  3  4  5 | 1  0  0  0  0 | 1 | Exclude |
| Two-stream Adaptive Convolutional Neural Network for Crowd Counting | 1  2  3  4  5 | 1  0  0  0  0 | 1 | Exclude |
| Cooperative Hierarchical Framework for Group Activity Recognition: From Group Detection to Multi-activity Recognition | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Development and testing of a pedestrian traffic monitoring system on a convolutional neural network platform | 1  2  3  4  5 | 0  0  0  0  0 | 0 | Exclude |
| Pedestrian Trajectory Prediction in Crowd Scene Using Deep Neural Networks | 1  2  3  4  5 | 0  0.5  0  0.5  0 | 1 | Exclude |
| Deep Person Detection in Two-Dimensional Range Data | 1  2  3  4  5 | 0  0  1  1  0 | 2 | Exclude |
| Deep neural network-based Wi-Fi/pedestrian dead reckoning indoor positioning system using adaptive robust factor graph model | 1  2  3  4  5 | 0  0  0  0  0 | 0 | Exclude |
| Automated Physical Distance Estimation and Crowd Monitoring Through Surveillance Video | 1  2  3  4  5 | 0  1  0.5  0  0 | 1.5 | Exclude |
| Deep Learning Based Crowd Monitoring And Person Identification System | 1  2  3  4  5 | 0  0  0  0  1 | 1 | Exclude |
| Context Aware Crowd Tracking and Anomaly Detection via Deep Learning and Social Force Model | 1  2  3  4  5 | 0  1  0.5  0  0 | 1.5 | Exclude |
| People Counting in Public Spaces using Deep Learning-based Object Detection and Tracking Techniques | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Implementation of Realtime design of crowd Enumeration via tracking using AI system | 1  2  3  4  5 | 0  0  0  0  0 | 0 | Exclude |
| Realtime Crowd Monitoring—Estimating Count, Speed and Direction of People Using Hybridized YOLOv4 | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Anomaly Detection Approach for Human Detection in Crowd Based Locations | 1  2  3  4  5 | 0  1  0.5  0.5  0 | 2 | Exclude |
| Rapid Detection of Pilgrims Whereabouts During Hajj and Umrah by Wireless Communication Framework : An application AI and Deep Learning | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| AerialMPTNet: Multi-Pedestrian Tracking in Aerial Imagery Using Temporal and Graphical Features | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| DensePeds: Pedestrian Tracking in Dense Crowds Using Front-RVO and Sparse Features | 1  2  3  4  5 | 0  1  0  1  0 | 2 | Exclude |
| An Aerial Crowd-Flow Analyzing System for Drone Under YOLOv5 and StrongSort | 1  2  3  4  5 | 0  1  1  0  1 | 3 | Include |
| On the Fine-Grained Crowd Analysis Via Passive WiFi Sensing | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| Density-based clustering with fully-convolutional networks for crowd flow detection from drones | 1  2  3  4  5 | 0  1  0.5  1  0 | 2.5 | Exclude |
| Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces | 1  2  3  4  5 | 0.5  1  0.5  0.5  1 | 3.5 | Include |
| Deep Reinforcement Learning with Pedestrian Trajectory Prediction Model for Service Robot Navigation in Crowded Environments | 1  2  3  4  5 | 0  0  0  0  0 | 0 | Exclude |
| Person Monitoring by Full Body Tracking in Uniform Crowd Environment | 1  2  3  4  5 | 0  0  0  0  0 | 0 | Exclude |
| Development of a Realistic Crowd Simulation Environment for Fine-Grained Validation of People Tracking Methods | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Realtime Crowd Monitoring - Estimating Count, Speed and Direction of People Using Hybridized YOLOv4 | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Multiple Pedestrian Tracking in Dense Crowds Combined with Head Tracking | 1  2  3  4  5 | 0  1  0  0  1 | 2 | Exclude |
| Deep Learning driven automated person detection and tracking model on surveillance videos | 1  2  3  4  5 | 0  0.5  0  1  1 | 2.5 | Exclude |
| Research on Pedestrian Detection and DeepSort Tracking in Front of Intelligent Vehicle Based on Deep Learning | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| Intelligent Deep Learning Enabled Crowd Detection and Classificai Model in Real Time | 1  2  3  4  5 | 0  0.5  0  0.5  0 | 1 | Exclude |
| Deep Transfer Learning Enabled Intelligent Object Detection for Crowd Density Analysis on Video Surveillance Systems | 1  2  3  4  5 | 0  1  0  0  0 | 1 | Exclude |
| Multi-person multi-camera tracking for live stream videos based on improved motion model and matching cascade | 1  2  3  4  5 | 0  0  0  0  1 | 1 | Exclude |
| A Real-time improved pedestrian dead reckoning trajectory tracking algorithm | 1  2  3  4  5 | 0  1  0  0.5  0 | 1.5 | Exclude |
| Enhancing Real-Time Human Tracking using YOLONAS-DeepSort Fusion Models | 1  2  3  4  5 | 0  1  0  1  1 | 3 | Include |
| Less Is More: Learning from Synthetic Data with Fine-Grained Attributes for Person Re-Identification | 1  2  3  4  5 | 0  0.5  1  1  1 | 3.5 | Include |
| A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance | 1  2  3  4  5 | 0  0.5  0  0.5  1 | 2 | Exclude |
| Handling Heavy Occlusion in Dense Crowd Tracking by Focusing on the Heads | 1  2  3  4  5 | 0  1  0.5  0.5  1 | 3 | Include |
| Multi-objective deep reinforcement learning for crowd-aware robot navigation with dynamic human preference | 1  2  3  4  5 | 1  1  0.5  0.5  0.5 | 3.5 | Include |
| Topology and channel affinity reinforced global attention for person re-identification | 1  2  3  4  5 | 1  1  0  0  1 | 3 | Include |
| A convolutional neural-network-based pedestrian counting model for various crowded scenes | 1  2  3  4  5 | 0  1  0.5  0.5  0 | 2 | Exclude |

F. Data Collection

After the quality assessment phase, data extraction is implemented on the remaining papers which are now used as the primary studies. The reason is to find the strong point of each research paper that can answer a specific research question. Table 6 shows the research paper along with their most respected research question and how it helped to answer it.

TABLE 6. Primary Studies and its most respected RQ

|  |  |
| --- | --- |
| Primary Study | Respected RQ |
| Learning how to analyse crowd behaviour using synthetic data | RQ2 |
| Tracking Hundreds of People in Densely Crowded Scenes With Particle Filtering Supervising Deep Convolutional Neural Networks | RQ2 |
| Crowd Abnormal Behavior Detection Combining Movement and Emotion Descriptors | RQ2 |
| Fast intensive crowd counting model of Internet of Things based on multi-scale attention mechanism | RQ3 |
| UUCT - HyMP: Towards Tracking Dispersed Crowd Groups from UAVs | RQ2 |
| Multi-Scale Occluded Pedestrian Detection Based on Deep Learning | RQ1 |
| An Aerial Crowd-Flow Analyzing System for Drone Under YOLOv5 and StrongSort | RQ2 |
| Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces | RQ1 |
| Enhancing Real-Time Human Tracking using YOLONAS-DeepSort Fusion Models | RQ3 |
| Less Is More: Learning from Synthetic Data with Fine-Grained Attributes for Person Re-Identification | RQ2 |
| A Survey on Multi-Target Multi-Camera Tracking Methods | RQ3 |
| Handling Heavy Occlusion in Dense Crowd Tracking by Focusing on the Heads | RQ1 |
| Topology and channel affinity reinforced global attention for person re-identification | RQ1 |

Figure 2 shows the PRISMA process implemented during the SLR from the papers gotten until the papers used as primary studies.

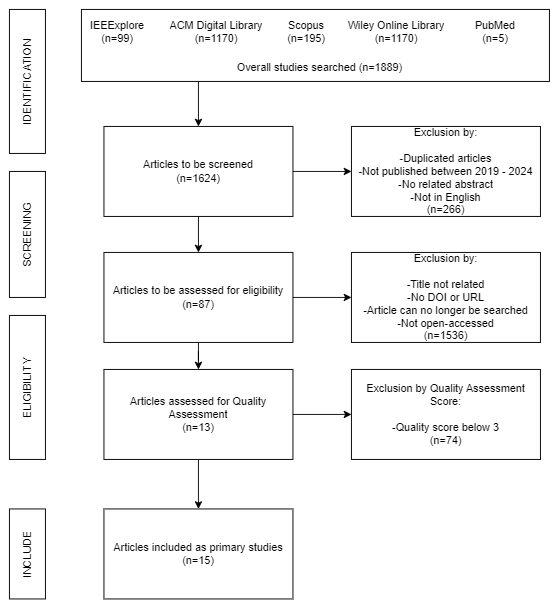


FIGURE 2. PRISMA Diagram for the current SLR

The final part of the process is cleaning the data. This is because Publish or Perish mixes up the data between different databases.

The screening process from getting the papers in the database to getting the primary studies where the important details are fixed can be found in the Excel sheet, <https://github.com/firdauskotp/SLR-DETECTION-AND-TRACKING-OF-PEOPLE-IN-A-CROWD-THROUGH-DEEP-LEARNING-APPROACH/blob/main/Publish%20or%20Perish%20Excel%20Sheets/PoP%20Primary%20Studies%20Screening%20Process.xlsx>

G. Data Analysis

This is the final step of the SLR methodology where the primary studies and the data extracted from them are analysed to answer the research questions according to the strong points of each paper.

3. Result and Analysis

In this SLR, there are 13 primary studies obtained after the PRISMA process from 1889 studies collected from five different databases. These primary studies will be used to answer the research questions proposed at the start of the SLR. A summary of the primary studies can be found in Table 7.

TABLE 7. Summary of Primary Studies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ID | TITLE | REFERENCE | PUBLISHER | TYPE | RESPECTED RQ | YEAR |
| PS1 | Learning how to analyse crowd behaviour using synthetic data | (A.R. Khadka et al., 2019) | ACM Digital Library | Conference Paper | RQ2 | 2019 |
| PS2 | Tracking Hundreds of People in Densely Crowded Scenes With Particle Filtering Supervising Deep Convolutional Neural Networks | (Gianni Franchi et al., 2020) | IEEE | Conference Paper | RQ2 | 2020 |
| PS3 | Crowd Abnormal Behavior Detection Combining Movement and Emotion Descriptors | (Xiao Li et al., 2020) | ACM Digital Library | Conference Paper | RQ2 | 2020 |
| PS4 | Fast intensive crowd counting model of Internet of Things based on multi-scale attention mechanism | (Dong Liu et. al., 2022) | The Institute of Engineering Technology | Article | RQ3 | 2022 |
| PS5 | UUCT - HyMP: Towards Tracking Dispersed Crowd Groups from UAVs | (Tonmoay Deb et. al., 2021) | IEEE | Conference Paper | RQ2 | 2021 |
| PS6 | Multi-Scale Occluded Pedestrian Detection Based on Deep Learning | (Fang Li et. al., 2022) | IEEE | Journal Article | RQ1 | 2022 |
| PS7 | An Aerial Crowd-Flow Analyzing System for Drone Under YOLOv5 and StrongSort | (Kuan-Hao Yeh et. al., 2022) | IEEE | Conference Paper | RQ2 | 2022 |
| PS8 | Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces | (Vivian W. H. Wong et. al., 2023) | MDPI | Article | RQ1 | 2023 |
| PS9 | Enhancing Real-Time Human Tracking using YOLONAS-DeepSort Fusion Models | (Athilakshmi R. et. al., 2024) | ResearchGate | Conference Paper | RQ3 | 2024 |
| PS10 | Less Is More: Learning from Synthetic Data with Fine-Grained Attributes for Person Re-Identification | (Suncheng Xiang et. al., 2021) | arXiv | Article | RQ2 | 2021 |
| PS11 | A Survey on Multi-Target Multi-Camera Tracking Methods | (Temitope Ibrahim Amosa et. al., 2023) | ScienceDirect | Survey Paper | RQ3 | 2023 |
| PS12 | Handling Heavy Occlusion in Dense Crowd Tracking by Focusing on the Heads | (Yu Zhang et. al., 2023) | arXiv | Article | RQ1 | 2023 |
| PS13 | Topology and channel affinity reinforced global attention for person re-identification | (Xile Wang et. al., 2021) | Wiley Online Library | Research Article | RQ1 | 2021 |

VOSviewer was used to conduct a brief data analysis on the keywords found in the primary studies’ title and abstract to determine the relationships and key aspects between them. As Publish or Perish gives a CSV file and it couldn’t be converted to a RIS file which is accepted by VOSViewer, Zotero is used to get the DOIs and convert them into a RIS file. The RIS file is then used by VOSViewer to obtain the key aspects and relationships between the primary studies. Figure 3 shows the output of VOSviewer.

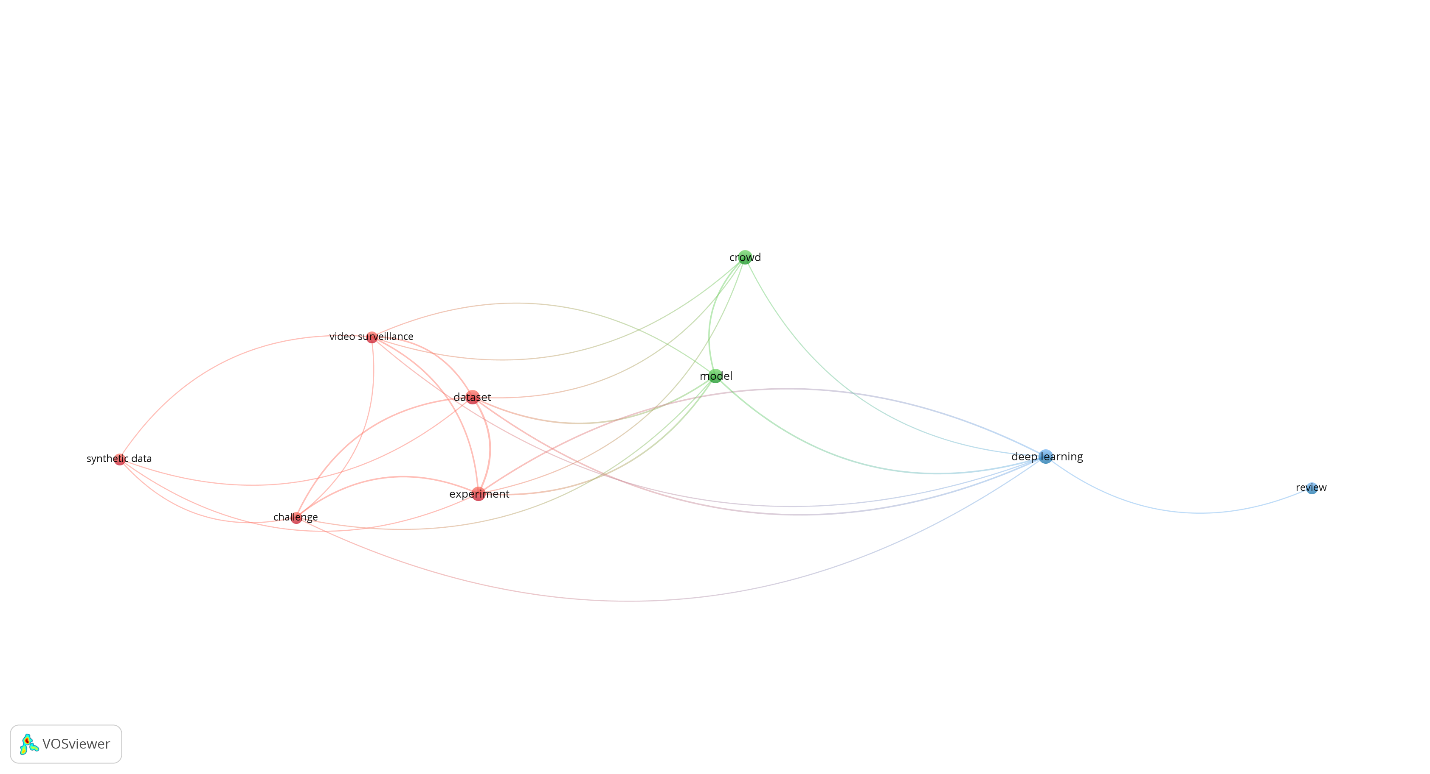


FIGURE 3. Key items of the primary studies visualised by VOSviewer

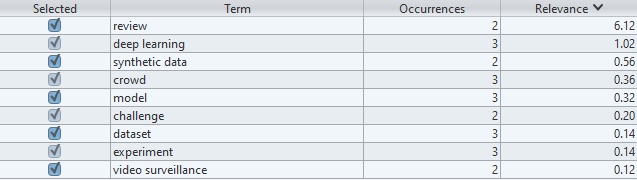


FIGURE 4. Terms of key items from primary studies that are visualised by VOSViewer

Based on Figure 3 and Figure 4, technical terms such as deep learning and crowd are repeated among the primary studies. This is to show that the primary studies involved are correlated with the current research. However, terms such as dense are not shown as not all of the primary studies involve dense crowds, and those that do uses different names across the studies such as high-density crowds. Re-identification is not shown here as well as they are written in the primary studies in different spellings such as re-indentification, re-identification, and also shorten to Re-ID. An important term is the synthetic data, which plays a role in auto annotation of some of the primary studies.

RQ1: What method will be used for detecting and tracking people in a dense crowd

There were not many papers that used the topological localisation method. Hence, if they were incorporating any kind of topological method, they were researched further and counted towards the quality assessment. Most of the primary studies involve dense crowd but not many use any kind of topological methods. Hence, three primary studies were mainly used to answer RQ1. PS13’s topology method was one of the closest to the current research as it uses the method with modules for re-identification of people with the help of spatial topology information (STI) and channel affinity information (CAI). Multi-informant Fusion reinforced Global Attention (MIFGA) is a module proposed by the researchers of PS13 in their research for more comprehensive information about STI and CAI. Although the study doesn’t deal with a dense crowd, it helps further the research on involving topology with re-identification. PS12 also shows how to deal with occlusions when tracking people in a dense crowd. The research doesn’t use topological methods. However, it shows how to re-identify a person from the dense crowd. The research also focuses on head tracking of the people in the crowd. By avoiding occlusions, which are overlapping of the boundary boxes or something blocking the annotated target for detection, we can further improve the accuracy for RQ3. PS8 solves the average for all three RQs. It uses a topology constraint as information for the spatial connectivity of egresses. It also involves re-identification and uses a method that the current research’s experiment is doing that is training unlabeled datasets with different models to improve their annotation. If we were to combine the research of all three primary studies, we could achieve the current research objectives through the experiment conducted.

RQ2: What method is used for the annotation of individuals in a dense crowd

Most of the primary studies use manual annotation. However, there are primary studies like PS8 that train different models on unlabeled datasets to improve the accuracy. However, most of the auto annotation from these primary studies come from synthetic data. Let’s take PS1 for example where there was a discussion on exploiting abundantly available unlabeled crowd imagery in a learning-to-rank framework. PS2 overcomes the issue of manual annotation for videos by only annotating the first frame and allowing the last 200 frames to train the diffusion-convolutional neural network (DCNN), which produced a very high accuracy of 98%. PS5 also has an interesting approach to the annotation of their models. The researchers created their own generator to generate the annotations. By combining this primary study with the others, it is highly possible to create an accurate annotation generator for future researchers in this field so that they can reduce the labour on annotating and can focus more on other areas to improve the quality of tracking people in dense crowds.

RQ3: How accurate is the detection and tracking of people in a dense crowd

All the primary studies that involve a dense crowd show a high accuracy in detection and tracking of people in it. Most of the high accuracy tests come from testing different architectures such as YOLO and GCN. However, there are primary studies that managed to further improve accuracy with the help of re-identification to reduce the rate of false positives. In the primary studies that respect RQ3, it is shown that most of the research uses a tracking algorithm called Deep Association Metric (DeepSORT). DeepSORT uses a simple person Re-ID model architecture, making it possible to easily re-identify a person from a dense crowd. Due to this research, what seemed to be a difficult task in re-identifying a person in the high-density crowd becomes an easier approach.

4. Limitation

The SLR about the tracking and detection of people through a deep learning approach is conducted based on 13 primary studies between 2019 to 2024. However, there is a possibility that the result of the SLR is affected by a few factors. These factors are the coverage of the search strategy and quality assessment inaccuracy. These factors have been discussed and explained further in this section.

The coverage of the search strategy influences the number of papers gotten across different online databases. The keywords used need were related to the current research such as topology and dense crowd. The Boolean expression used can also influence the number of papers retrieved from the databases. While using the AND Boolean expression, that keyword must be available in the topic.

The filtering of the papers also affects the number of papers gotten. This is due to how EndNote files and papers gotten from Scopus don't give abstracts, influencing the step where the papers without abstract. Due to them not getting the abstract from the beginning, these papers were not removed.

Publish or Perish and Zotero were used to retrieve the data. However, Publish or Perish mixes up the data in different column headers as the SLR retrieves data from different databases. Publish or Perish also causes some of the papers dates to be wrong and having the wrong DOI. Hence, it was needed to manually check some of the papers that were screened to make sure that they can be counted as a primary study.

The quality assessment might also affect the SLR. This is due to how the final exclusion of the papers before the current study gets the primary studies relies on the quality for the papers. If the papers do not meet the quality score set during the section, the paper will be excluded. There are only 5 research questions that are closely related to the current study that the quality score is based on. The strictness of the quality set and the number of research questions will influence the final number of primary studies. Since there are not many papers that use topological localisation, the research that incorporated any kind of topological method was counted.

5. Conclusion

Detection and tracking of people in a dense crowd is essential research as there have been many accidents that have happened due to bad crowd management. To further the current study that is to track and detect people in a dense crowd through a deep learning approach, a SLR must be carried out to understand the methods and architectures used for tracking and detection of people in a crowd by similar studies. Even when we have the data for tracking the people in a dense crowd, we will need to manually annotate the people to be detected. Hence, the research also needed to cover on how to auto annotate the targets in our dataset. Research questions were formulated for the current study to be solved by the primary studies of the SLR. 1889 articles were originally obtained when five online repositories were searched using specific keywords. The number of articles was filtered down to 13 primary studies published from 2019 to 2024. The filtering includes exclusion of papers that are not related to the topic, not English, and doesn’t achieve the target of the quality score set. The primary studies are then used to answer the three research questions where each primary study is researched more to see which research question is most suitable to fill in the gap. Finally, the limitations of the studies were researched and the future directions for researchers who are keen on this area of research were discussed. The aim of this paper is to serve as a guide for research in this area of study through the help of a systematic review of the methods and architectures of detection and tracking of people in a dense crowd and improve the productivity of research in terms of annotating the targets for labelling.

6. Reference

Jing Li (2023) ‘A systematic literature review on feature selection for machine learning-based attack classification for IoT security’, *Int. J. Sensor Networks*, Vol. X, No. Y4, pp.000---000.

Amosa, T. I., Sebastian, P., Izhar, L. I., Ibrahim, O., Ayinla, L. S., Bahashwan, A. A., Bala, A., & Samaila, Y. A. (2023). Multi-camera multi-object tracking: A review of current trends and future advances. *Neurocomputing*, *552*, 126558. <https://doi.org/10.1016/j.neucom.2023.126558>

Athilakshmi, R., Chandan Sainagakrishna, P. S., Chaitanya Chowdary Kota, S. S., Kiran Teja, M. C., Venkatesh, T., & Prasad, V. J. (2023). Enhancing Real-Time Human Tracking using YOLONAS-DeepSort Fusion Models. *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1118–1125. <https://doi.org/10.1109/ICECA58529.2023.10394864>

Deb, T., Rahmun, M., Bijoy, S. A., Raha, M. H., & Khan, M. A. (2021). UUCT - HyMP: Towards Tracking Dispersed Crowd Groups from UAVs. *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8. <https://doi.org/10.1109/IJCNN52387.2021.9533600>

Franchi, G., Aldea, E., Dubuisson, S., & Bloch, I. (2020). Tracking Hundreds of People in Densely Crowded Scenes With Particle Filtering Supervising Deep Convolutional Neural Networks. *2020 IEEE International Conference on Image Processing (ICIP)*, 2071–2075. <https://doi.org/10.1109/ICIP40778.2020.9190953>

Khadka, A. R., Oghaz, M. M., Matta, W., Cosentino, M., Remagnino, P., & Argyriou, V. (2019). Learning how to analyse crowd behaviour using synthetic data. *CASA ’19: Computer Animation and Social Agents*, 11–14. <https://doi.org/10.1145/3328756.3328773>

Li, X., Yang, Y., Xu, Y., Wang, C., & Li, L. (2020). Crowd Abnormal Behavior Detection Combining Movement and Emotion Descriptors. *ICNSER2020: The 2nd International Conference On Industrial Control Network And System Engineering Research*, 106–110. <https://doi.org/10.1145/3411016.3411166>

Liu, D., Wang, Z., & Meng, X. (2022). Fast intensive crowd counting model of Internet of Things based on multi‐scale attention mechanism. *IET Image Processing*, ipr2.12686. <https://doi.org/10.1049/ipr2.12686>

Wang, X., Gao, C., Xin, M., Zhang, S., & Zhang, M. (2021). Topology and channel affinity reinforced global attention for person re‐identification. *International Journal of Intelligent Systems*, *36*(9), 5136–5160. <https://doi.org/10.1002/int.22506>

Wong, V., & Law, K. (2023). Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces. *Algorithms*, *16*(3), 154. <https://doi.org/10.3390/a16030154>

Xiang, S., You, G., Guan, M., Chen, H., Yan, B., Liu, T., & Fu, Y. (2021). *Less is More: Learning from Synthetic Data with Fine-grained Attributes for Person Re-Identification*. arXiv. <https://doi.org/10.48550/ARXIV.2109.10498>

Yeh, K.-H., Hsu, I.-C., Chou, Y.-Z., Chen, G.-Y., & Tsai, Y.-S. (2022). An Aerial Crowd-Flow Analyzing System for Drone Under YOLOv5 and StrongSort. *2022 International Automatic Control Conference (CACS)*, 1–6. <https://doi.org/10.1109/CACS55319.2022.9969785>

Zhang, Y., Chen, H., Bao, W., Lai, Z., Zhang, Z., & Yuan, D. (2023). *Handling Heavy Occlusion in Dense Crowd Tracking by Focusing on the Heads*. arXiv. <https://doi.org/10.48550/ARXIV.2304.07705>