Jurnal Teknologi Riset Terapan (JATRA)
ISSN: 2986-7169, Vol 4, No 1, 2026, 15-32 https://doi.org/10.35912/jatra.v4i1.7215

Crowd Heatmap Detection at Elephant Park During COVID-19 Using SSD Object Detection

Wahid Salahudin[1*], Ismoyo Aji Sasmita[2], Al Barra Harahap[3], Nike Dwi Grevika Drantantiyas[4]
Institut Teknologi Sumatera, Lampung, Indonesia[1,2,3,4]
wahid.118320015@student.itera.ac.id[1*]

Article History
Received on 5 November 2025
1st Revision on 21 November 2025
2nd Revision on 13 December 2025
3rd Revision on 4 January 2026
Accepted on 15 January 2026

Abstract
Purpose: This study investigates the application of deep learning based object detection to construct a motion heatmap of visitor crowds at Elephant Park, Bandar Lampung, as a monitoring measure during the COVID-19 pandemic. The research aims to determine whether an object detection algorithm can identify crowd density patterns and to evaluate the performance of the detector under varying confidence thresholds.
Research Methodology: The Single Shot MultiBox Detector with a MobileNet V1 backbone, pretrained on the Common Objects in Context dataset, was deployed on an NVIDIA Jetson Nano connected to a USB camera positioned at two intersections inside the park. Background subtraction in OpenCV was combined with detected bounding boxes to generate a motion heatmap. Detector performance was evaluated using a confusion matrix across confidence threshold values ranging from 0 to 0.9.
Results: The detector achieved a mean average precision of 23.2%, with an accuracy of 0.60, precision of 0.84, recall of 0.63, and F1 score of 0.72 at a confidence threshold of 0.3. A confidence threshold of 0.6 produced the most balanced trade-off for generating a usable motion heatmap, yielding a precision of 0.87.
Conclusions: The combined detection and heatmap pipeline successfully visualized visitor flow patterns at both observation points.
Limitations: Detection accuracy degraded under intense lighting variation and small pedestrian scale.
Contributions: This work offers a low cost, embedded surveillance approach that can inform public space design and health protocol enforcement in open air tourist attractions.

Keywords: *Crowd Monitoring, Deep Learning, Motion Heatmap, Object Detection, Single Shot Multi Box Detector*

How to Cite: Salahudin, W., Sasmita, I. A., Harahap, A. B., & Drantantiyas, N. D. G. (2026). Crowd Heatmap Detection at Elephant Park During COVID-19 Using SSD Object Detection. *Jurnal Teknologi Riset Terapan*, 4(1), 15-32.

1. Introduction

The spread of COVID-19 across Indonesia was strongly associated with insufficient adherence to health protocols during public activities. Government agencies attempted to curb transmission through mobility restriction policies, which in turn produced significant declines in tourism activity nationwide, with foreign visitor arrivals dropping sharply compared with previous years (Wang, Tang, Chen, & Zhang, 2022). Reduced mobility affected the tourism sector disproportionately because the sector depends heavily on gatherings and open access to public spaces, and its contraction generated concern given its role as a major contributor to national foreign exchange earnings.

As the pandemic evolved through successive waves driven by population mobility and inconsistent protocol compliance, Indonesia gradually moved toward loosened restrictions once case numbers declined, marking a transition from pandemic response toward endemic management. This loosening created renewed opportunities for the tourism industry to recover, yet field reports consistently indicated that segments of the public remained reluctant to wear masks indoors, maintain hand hygiene, or observe physical distancing even where guidance still recommended these measures. Because the pandemic threat had not fully subsided, continuous monitoring of public compliance remained an important, though logistically demanding, undertaking, particularly across expansive open spaces where manual supervision by government personnel becomes inefficient (Masita, Hasan, & Shongwe, 2020).

Given these constraints, engineered spatial design and monitoring technology present an alternative pathway to reduce transmission risk without relying solely on in-person enforcement. Machine learning, and more specifically deep learning, has become a widely adopted tool for automated surveillance because it enables continuous, remote observation of public behavior. Object detection, a core deep learning task, allows a trained model to localize and classify objects of interest within images or video streams, and its accuracy is typically expressed through standard detection metrics computed against labeled ground truth. Deep learning based detectors have already demonstrated utility in adjacent public health applications, including face mask compliance detection with reported accuracies ranging from 85% to 100% using the Single Shot MultiBox Detector, or SSD, architecture (Pramita, Kurniawan, & Utama, 2020; Muharram, 2021). Similarly, physical distancing surveillance implemented through closed circuit television combined with the You Only Look Once, or YOLO, family of detectors has achieved accuracy figures between 73% and 95.35% in prior evaluations (Marutotamtama & Setyawan, 2021; Rezaei & Azarmi, 2020). Other researchers have applied YOLOv3 to construct indoor crowd heatmaps and reported qualitatively superior performance relative to SSD and histogram of oriented gradients based approaches (Kajabad & Ivanov, 2019).

Building on this body of work, the present study proposes a heatmap based crowd monitoring system for Elephant Park, an open public garden in Bandar Lampung, using the SSD object detector implemented with a MobileNet V1 backbone trained on the Common Objects in Context, or COCO, dataset. The system was designed to operate on an embedded computing platform, the NVIDIA Jetson Nano, to keep deployment costs low while still supporting near real time inference. The resulting motion heatmap synthesizes accumulated detections of the person class over time, offering park managers and public health authorities an interpretable visualization of where and when crowding tends to concentrate.

The research addresses three specific questions. First, whether an object detection algorithm is capable of identifying human presence accurately enough to support crowd mapping. Second, what level of detection performance the chosen algorithm achieves under field conditions specific to Elephant Park. Third, how the resulting crowd map can be visually represented to communicate spatial and temporal patterns of visitor density. The scope of the investigation is deliberately bounded to the person class, to two intersection points within the park selected for their potential to attract concentrated foot traffic, and to the SSD architecture as the sole detection method under evaluation.

By pursuing these questions, the study contributes evidence relevant to two audiences. For policy makers and park administrators, the findings offer a concrete illustration of how automated monitoring can supplement manual health protocol enforcement in open, uncontrolled environments. For the computer vision research community, the study provides an empirical account of SSD MobileNet V1 performance when deployed on constrained edge hardware and applied to a real, non curated crowd scene, complementing prior evaluations that have largely relied on benchmark datasets or indoor settings (Kim, Sung, & Park, 2020; Ahamad, Zaini, & Latip, 2020). The remainder of this article is organized as follows. Section two reviews relevant literature on object detection architectures and prior crowd monitoring research and develops the guiding propositions of the study. Section three details the research methodology, including hardware configuration, algorithm construction, and evaluation procedures. Section four presents the empirical results and discusses their implications. Section five concludes with a summary of findings, acknowledges limitations, and proposes directions for future research.

2. Literature Review and Hypotheses Development

2.1 Cyber Physical Systems and Embedded Artificial Intelligence

A cyber physical system integrates physical hardware with software driven intelligence so that the two layers communicate continuously within a coherent technological ecosystem. Large scale data processing, artificial intelligence, and cloud infrastructure form the backbone that supports the interconnected architecture underlying such systems, and their emergence has accompanied a broader shift toward more complex industrial and civic applications (Zhou, Zhou, Wang, & Zang, 2019). Because cyber physical systems draw together multiple disciplines through shared computational infrastructure, tasks that would otherwise exceed the practical limits of manual human oversight, such as continuous crowd surveillance across an entire public park, become tractable. Commercial cloud providers have accelerated this trend by making storage, computing, and connectivity broadly accessible, while artificial intelligence techniques have simultaneously diversified into increasingly specialized subfields (Jiang et al., 2017; Kumar, Tiwari, & Zymbler, 2019). The architecture underlying the present study, comprising an embedded processor, a camera sensor, and cloud connected messaging, reflects this cyber physical paradigm at a modest, single site scale.

2.2 Convolutional Neural Networks and Object Detection

Convolutional neural networks, a specialized extension of feedforward multilayer perceptrons, underpin most modern image recognition systems by decomposing an image into local receptive fields that are filtered through learned convolutional kernels (Zhao, Zheng, Xu, & Wu, 2019). Successive pooling operations reduce the resulting feature maps to a manageable size before a fully connected layer performs final classification. Object detection extends this classification capability by simultaneously predicting the spatial location of an object, typically expressed as a bounding box, alongside its class label. Early detection pipelines, exemplified by Region based Convolutional Neural Networks, or R-CNN, relied on a separate region proposal stage using selective search before classifying each proposed region, a design that proved computationally expensive (Zhao, Zheng, Xu, & Wu, 2019). Fast R-CNN reduced this overhead by pooling regions of interest directly from convolutional feature maps, and Faster R-CNN further integrated the proposal mechanism into the network itself using a dedicated region proposal network (Ren, He, Girshick, & Sun, 2017). These two stage detectors generally offer strong accuracy but incur processing latency that limits real time deployment on constrained hardware.

Single stage detectors address this limitation by predicting class scores and bounding box offsets in a single forward pass rather than through a separate proposal stage. The Single Shot MultiBox Detector, introduced by Liu et al. (2016), exemplifies this approach by attaching multiple convolutional feature layers of decreasing spatial resolution to a base network, in this case VGG16, so that detections can be generated at several scales simultaneously. Each feature layer applies a small convolutional kernel to a set of default boxes, analogous to the anchor boxes used in Faster R-CNN, producing offset and class confidence predictions that are subsequently filtered through non maximum suppression. Benchmark comparisons indicate that SSD variants can process video frames considerably faster than Faster RCNN while retaining competitive mean average precision, with SSD300 reported to reach 46 frames per second at 77.2% mean average precision on the PASCAL VOC2007 test set (Liu et al., 2016). An alternative single stage family, YOLO, divides an input image into a grid and predicts bounding boxes directly from each grid cell, generally trading some accuracy for additional inference speed relative to SSD (Kim, Sung, & Park, 2020).

2.3 MobileNet Architecture for Resource Constrained Devices

Standard convolutional layers demand considerable computational resources, which restricts their deployment on mobile or embedded hardware lacking dedicated graphics processing capacity. MobileNet addresses this constraint by decomposing standard convolution into depthwise and pointwise operations, substantially reducing the number of multiply accumulate operations required per layer while preserving representational capacity (Howard et al., 2017). Because MobileNet was purpose built for constrained devices, pairing it with SSD as a feature extraction backbone, commonly denoted SSD MobileNet, yields a detector suited to platforms such as the NVIDIA Jetson Nano, which was employed throughout the present study (Cao, Hidalgo, Simon, Wei, & Sheikh, 2021).

2.4 Common Objects in Context Dataset

The Common Objects in Context, or COCO, dataset comprises several hundred thousand annotated images spanning ninety one object categories and has become a standard benchmark for training and evaluating detection, segmentation, and captioning models (Lin et al., 2014). Because COCO includes a person category alongside many object categories captured at varied scales and contexts, pretrained COCO models offer a convenient starting point for human detection tasks, although the dataset is known to emphasize medium and large scale objects positioned near the frame center, which can constrain performance on distant, small scale subjects such as pedestrians viewed from an elevated surveillance angle (Chen et al., 2021).

2.5 Motion Heatmap Techniques

A heatmap conveys quantitative information through a two dimensional color gradient, allowing viewers to intuitively perceive density, intensity, or frequency without parsing raw numerical data. Motion heatmaps specifically accumulate detected movement across successive video frames, typically derived through background subtraction, to reveal which regions of a scene experience the most frequent activity (Ihaddadene & Djeraba, 2008). Because human perception more readily interprets visual patterns than tabular statistics, heatmap based crowd visualization has been adopted across retail analytics, museum visitor studies, and public safety monitoring. Kajabad and Ivanov (2019) demonstrated that combining YOLOv3 based person detection with Gaussian mixture background subtraction could reveal areas of high visitor concentration within a museum setting, reporting detection latency below one second per frame. Zhang, Lian, and Xu (2020) similarly combined positional tracking with spatial analysis to characterise visitor temporal and spatial behaviour in a heritage garden, underscoring the versatility of density based visualisation across both indoor and outdoor environments. Recent surveys of deep learning based density estimation confirm that accumulator based visualisation remains a practical alternative to regression based crowd counting when absolute headcounts are not required (Gao, Gao, Liu, Wang, & Wang, 2025).

2.6 Prior Applications in Pandemic Related Monitoring

Several studies have examined deep learning based monitoring specifically in response to COVID-19 containment needs. Pramita, Kurniawan, and Utama (2020) trained a convolutional neural network to classify mask wearing status from facial images, achieving accuracy up to 97% after data augmentation. Ahamad, Zaini, and Latip (2020) developed a MobileNet SSD based person detector coupled with segmented region of interest analysis to flag physical distancing violations, reporting distance estimation accuracy between 56.5% and 68% for outdoor video and up to 100% under controlled indoor conditions. Rezaei and Azarmi (2020) similarly proposed a joint detection and distance estimation pipeline intended for deployment on standard surveillance cameras. These studies collectively suggest that detector performance depends heavily on the specific deployment context, including camera elevation, lighting, and crowd density, motivating the present study's site specific evaluation at Elephant Park rather than reliance on benchmark figures alone.

2.7 Hypotheses and Propositions

Table 1. Proposed research propositions for SSD MobileNet V1 COCO-based human detection and heatmap generation

Code	Research Proposition
P_1	SSD MobileNet V1 COCO is capable of detecting human subjects in Elephant Park surveillance videos with sufficient precision for heatmap generation.
P_2	Confidence threshold variation significantly influences SSD MobileNet V1 COCO detection performance metrics (precision, recall, accuracy, and F1-score).
P_3	Optimal confidence threshold selection improves the quality and visual coherence of visitor density heatmaps.

Table 1 shows the propositions advanced by this study on the basis of the reviewed literature. First, given that SSD MobileNet models pretrained on COCO have previously achieved usable precision for person detection in comparable outdoor surveillance contexts (Ahamad, Zaini, & Latip, 2020; Heredia & Barros-Gavilanes, 2019). It is proposed that the SSD MobileNet V1 COCO detector can identify human subjects within Elephant Park video footage at a precision level sufficient to support heatmap construction. Second, because detection sensitivity and confidence threshold are inversely related in single stage detectors (Zhao, Zheng, Xu, & Wu, 2019). It is proposed that detector performance, expressed through precision, recall, accuracy, and F1 score, will vary systematically as the confidence threshold is adjusted, with an intermediate threshold offering the most favorable balance for heatmap generation. Third, given that background subtraction based heatmaps depend on consistent bounding box output to avoid spurious coloring (Ihaddadene & Djeraba, 2008). It is proposed that a confidence threshold that minimizes false positive detections while retaining a workable number of true positive detections will yield the most visually coherent motion heatmap of visitor crowding.

3. Research Methodology

3.1 Research Design and Location

This study adopted a field based experimental design combining hardware deployment with algorithm evaluation. Data collection took place at Elephant Park, located in the Tanjung Karang Pusat subdistrict of Bandar Lampung, between the third week of January 2022 and June 2022, while algorithm development and physical system assembly were conducted at the Sustainable and Intelligent Built Environment Laboratory of the Institut Teknologi Sumatera.

Because the deployment operated in a public space, the pipeline was configured to minimise the personal information it retained. Detection was restricted to the person class among the ninety one COCO categories, so no facial, demographic, or identity attributes were extracted at any stage of processing. The system stored only the accumulated heatmap frames sampled at fifteen second intervals rather than the continuous video stream, and those frames were transmitted to a private monitoring channel accessible solely to the research team. Recording sessions were scheduled during periods of stable illumination in the late afternoon and early evening, a constraint identified during the preliminary validation described in subsection 3.5, since abrupt changes in ambient light were found to introduce spurious colouring into the accumulated heatmap.

3.2 Hardware Configuration

The physical monitoring unit combined an NVIDIA Jetson Nano Developer Kit with 4 gigabytes of memory, a Logitech C922 Pro HD Stream USB webcam, and a regulated power subsystem comprising a 12 volt, 5 ampere power supply, a 12 volt, 7.5 ampere hour lead acid battery, an automatic transfer switch, and a direct current step down converter. The Jetson Nano was selected over alternatives such as a general purpose laptop or a Raspberry Pi because it integrates a dedicated 128 core Maxwell graphics processing unit capable of approximately 472 giga floating point operations per second, a capability that Raspberry Pi boards lack given the absence of an onboard graphics processor (Süzen, Duman, & Şen, 2020; Jeong, Kim, & Ha, 2022). The USB camera was chosen over a MIPI camera serial interface module primarily for its plug and play compatibility and lower development cost, despite the latter offering marginally higher achievable bandwidth. The camera was configured to capture at 720p resolution and 60 frames per second, a setting selected to balance processing load against image detail on the embedded platform.

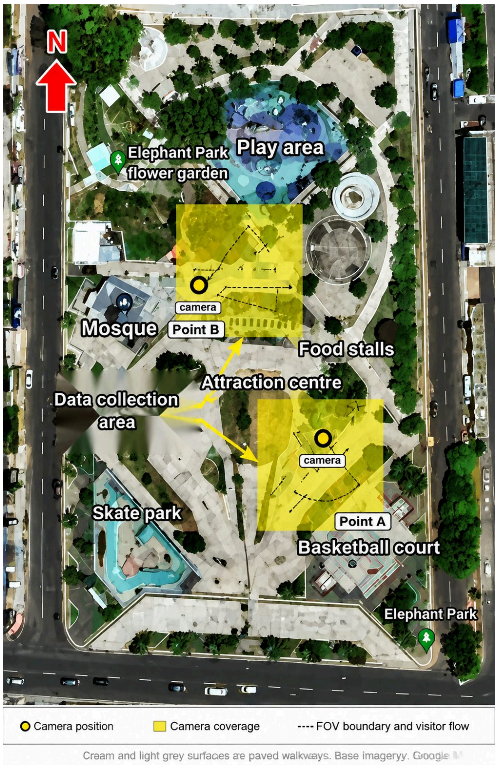

Figure 1. Site plan of Elephant Park showing observation Points A and B

The camera was mounted atop a five meter stainless steel pole combined with a tripod base, producing a bird's eye viewing angle across each observation point. Two locations within the park were selected following a site survey against four criteria, namely an unobstructed line of sight between the camera and the intersection, a mounting location that would not obstruct visitor movement, comparatively higher pedestrian traffic relative to other locations, and the presence of an attraction likely to draw visitors along at least one approach to the intersection. Point A was situated between a basketball court and a cluster of vendor stalls and rental toy cars, while Point B was located near the Ar-Ridho mosque at a four way junction connecting the park entrance, the mosque, the vendor cluster, and a children's play area, as can be seen in Figure 1. Figure 2 Site plan of Elephant Park showing observation Points A and B

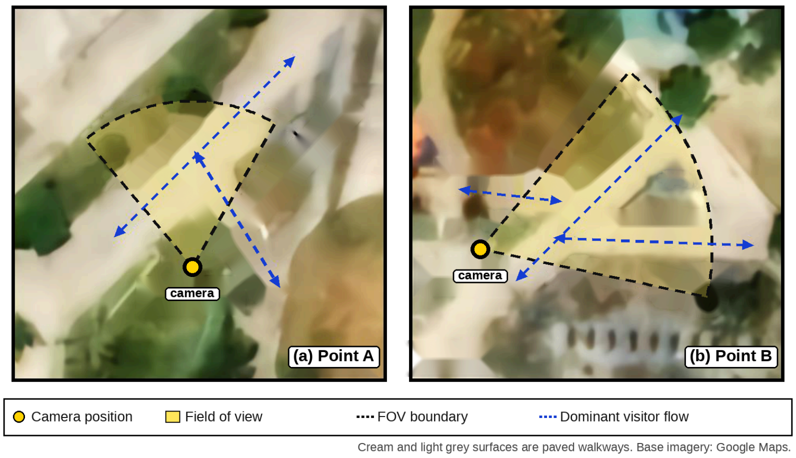

Figure 2. Camera field of view and dominant visitor flow directions at (a) Point A and (b) Point B

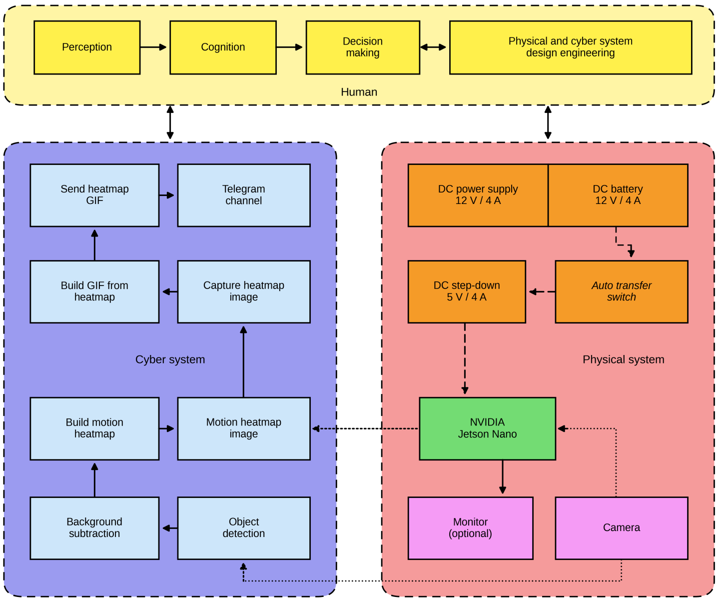

Figure 3. Cyber-physical system architecture of the crowd monitoring unit

Figure 3 shows the cyber physical architecture of the monitoring unit, in which the human, cyber, and physical layers exchange information continuously. Table 2 shows the full hardware and software specification of the deployed system.

Table 2. Hardware and software specification of the monitoring unit

Component	Specification	Role in the system
NVIDIA Jetson Nano Developer Kit	4 GB RAM, JetPack 4.2 (Ubuntu 18.04 LTS)	Edge inference
Logitech C922 Pro	USB, 1080p	Image acquisition
DC power supply	12 V / 5 A	Primary power
Lead-acid battery with automatic transfer switch	12 V	Backup power, uninterrupted switching
DC-DC step-down converter	12 V → 5 V / 4 A	Jetson supply
Software stack	Python 3.6, TensorFlow-GPU, TensorRT, PyCUDA, OpenCV, Protobuf	Detection and heatmap pipeline
Mounting	5 m stainless steel pole with tripod base	Bird-eye camera placement

Table 2 show the system configuration consists of integrated hardware and software components designed to support real-time edge-based image processing. The NVIDIA Jetson Nano serves as the primary edge inference device, while the Logitech C922 Pro camera enables high-resolution image acquisition. A dual power system consisting of a DC supply and backup battery ensures continuous operation, supported by a DC-DC converter for stable voltage regulation. The software stack, including TensorFlow-GPU, TensorRT, PyCUDA, and OpenCV, facilitates object detection and heatmap generation. The camera is mounted on a 5-meter stainless steel pole with a tripod base to achieve optimal bird's-eye-view monitoring.

3.3 Software Configuration and Detection Algorithm

The Jetson Nano ran NVIDIA JetPack 4.2, built on Ubuntu 18.04 Long Term Support Linux, with detection and heatmap scripts written in Python 3.6. The object detection component adapted an existing SSD implementation, modified to run with a MobileNet V1 feature extraction backbone pretrained on the COCO dataset, a configuration denoted throughout this article as SSD MobileNet

V1 COCO. Required dependencies included PyCUDA, TensorRT, TensorFlow GPU, Protobuf, and OpenCV, each configured to versions compatible with the JetPack 4.2 environment. Detection was restricted to the person category among the ninety one COCO classes, since the research scope concerned only human crowding rather than general object recognition.

3.4 Motion Heatmap Construction

The motion heatmap was generated by combining detected bounding boxes with an OpenCV background subtraction routine. For each processed frame, an empty canvas matching the frame dimensions was populated only within regions where a person bounding box had been drawn, then converted to grayscale and blurred with a Gaussian filter to soften transitions. A Gaussian mixture based background subtractor, coupled with a morphological closing operation to suppress noise, was applied to the resulting canvas to accumulate motion intensity over successive frames (Zuo, Jia, Yang, & Kasabov, 2019; Song, Ali, & Bouguila, 2020). Subsequent refinements of this family of methods introduce explicit tracking of motion regions to stabilise the foreground mask under gradual illumination change (Delibaşoğlu, 2023). The accumulated intensity values were normalized, mapped to an eight bit range, and rendered using the OpenCV hot colormap, which transitions from black through orange to near white as accumulated intensity increases.

The resulting heatmap provides a spatial representation of movement frequency by aggregating temporal changes detected across consecutive frames. Unlike conventional object detection outputs that only indicate the presence and location of objects at a single time instance, the accumulated heatmap enables analysis of persistent activity patterns within the monitored area. Regions with higher intensity values indicate locations where human movements occur more frequently or remain active for longer durations, while lower intensity areas represent limited or occasional movements. This approach improves the interpretability of the detection system by transforming individual frame-level detections into a continuous spatial behaviour map. Furthermore, the integration of edge-based inference and motion accumulation allows the system to operate efficiently in real-time environments while reducing the dependency on extensive post-processing or centralized computational resources. Therefore, the generated heatmap serves not only as a visualization tool but also as an analytical representation for identifying movement concentration and activity distribution within the observation area.

The colored heatmap layer was then blended with the original camera frame using a weighting scheme that assigned full opacity to the original frame and reduced opacity to the heatmap layer, producing a semi transparent overlay that preserved visibility of the underlying scene while still conveying accumulated crowd density as can be seen in Figure 4 .

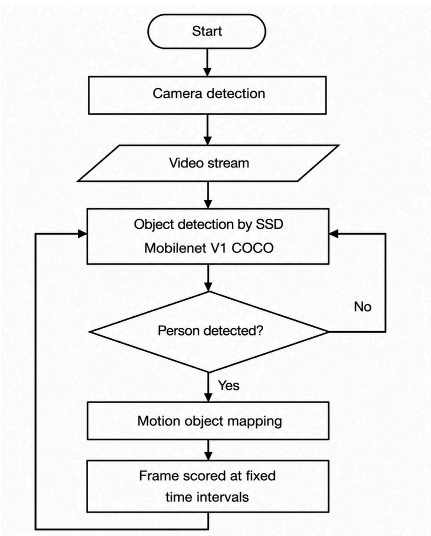

Figure 4. Workflow of the detection and motion heatmap algorithm

3.5 Field Testing and Preliminary Validation

Before formal data collection, the combined detection and heatmap pipeline was validated under varied lighting and motion conditions. A preliminary trial conducted in a building lobby during a one hour morning window demonstrated that the system could correctly capture directional movement patterns toward two separate building entrances, with brighter heatmap coloring correctly corresponding to the more heavily trafficked entrance. Additional trials at Elephant Park under both low light, around 18:30 local time, and bright daylight, around 16:50 local time, indicated that detection and heatmap generation both functioned during low light conditions, whereas bright daylight conditions occasionally produced camera shake or rapid pixel intensity changes that degraded heatmap coherence, a known limitation of background subtraction based motion visualization.

3.6 Confidence Threshold Tuning Procedure

Because SSD assigns a confidence score to each candidate detection, the sensitivity of the overall pipeline depends heavily on the chosen confidence threshold, below which candidate detections are discarded. To identify an appropriate operating threshold, fifteen representative frames were extracted from recorded crowd footage at the park intersections. Detector output was compared against manually annotated ground truth across confidence threshold values ranging from 0 to 0.9 in increments of 0.1, allowing systematic evaluation of how threshold selection affected detection outcomes before selecting a threshold suited to heatmap generation specifically.

3.7 Evaluation Metrics

Detector performance was assessed using standard confusion matrix framework, in which a true positive denotes a correctly localized bounding box around a human subject, a false negative denotes an undetected human subject lacking any bounding box, a false positive denotes a bounding box incorrectly assigned to a non human region, and a true negative denotes correct rejection of a non human region, though no true negative cases arose in this study because no non person class was explicitly evaluated. From these counts, recall was computed as the ratio of true positives to the sum of true positives and false negatives, precision as the ratio of true positives to the sum of true positives and false positives, accuracy as the ratio of true positives to the sum of true positives, false negatives, and false positives, and the F1 score as the harmonic mean of precision and recall, following the Formula 1, Formula 2, Formula 3, and Formula 4 .

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Accuracy = \frac{TP}{TP + FN + FP} \times 100 \tag{3}$$

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

The performance evaluation of the proposed detection system was conducted using several standard classification metrics to provide a comprehensive assessment of its effectiveness. Formula (1) represents the recall metric, which measures the system's ability to correctly identify all relevant instances from the actual positive samples, thereby indicating the detection capability and its sensitivity toward missed objects. Formula (2) presents the precision metric, which evaluates the proportion of correctly detected positive instances among all predicted positive results, reflecting the reliability of the system in minimizing false detections. Formula (3) defines the accuracy metric, which describes the overall correctness of the model by considering both correctly classified positive and negative instances relative to the total number of evaluated samples. Furthermore, Formula (4) illustrates the F1-score, a harmonic mean between precision and recall, which provides a balanced measurement of model performance, particularly when dealing with potential class imbalance and the trade-off between false positives and false negatives. These evaluation metrics were consistently applied throughout the experimental analysis to quantify the robustness, reliability, and overall effectiveness of the proposed detection framework.

3.8 Data Collection Workflow

Once the operating confidence threshold was selected, the system continuously captured video, processed each frame through the SSD MobileNet V1 COCO detector, and, upon detecting a person, updated the accumulating motion heatmap. A single output frame was captured from the heatmap every 15 seconds to allow sufficient time for the color intensity to develop meaningfully between captures. Collected frames were periodically compiled into an animated Graphics Interchange Format file and transmitted to a monitoring channel through a Telegram bot, allowing remote review of crowd patterns without requiring on site access to the Jetson Nano.

4. Results and Discussions

4.1 Physical System Implementation

The completed monitoring unit was housed within a weatherproof panel enclosure containing three integrated subsystems, an input output subsystem comprising the USB camera and an optional monitor used only for initial alignment, a power subsystem comprising the 12 volt supply, battery, automatic transfer switch, and voltage step down converter, and a processing subsystem centered on the Jetson Nano itself. The automatic transfer switch successfully alternated between the primary power supply and the backup battery without interrupting system operation, confirmed through direct observation of the panel mounted volt ampere meter across three switching states, namely full shutdown, primary supply active, and battery active. The assembled unit was installed at both Point A and Point B using the five meter stainless steel pole and tripod combination described in the methodology, producing the intended elevated, bird's eye camera perspective at each site as can be seen in Figure 5

Figure 5. Field installation of the monitoring unit at Elephant Park

4.2 Object Detection Model Characteristics

The SSD MobileNet V1 architecture used in this study comprised a base network extended with six auxiliary convolutional layers, producing a compact trained model file of 28.42 megabytes. Training relied on the pretrained COCO weights, derived from roughly 164000 images spanning 91 object classes, yielding an overall mean average precision of 23.2%. Table 3 summarises the resulting classification metrics obtained at the confidence threshold ultimately selected for reporting, namely 0.3.

@tableindicates that the detector achieved a comparatively high precision of 0.84 alongside a moderate recall of 0.63, suggesting that although the model rarely produced spurious detections, it more frequently failed to localize every human subject present within a crowded frame, a pattern consistent with the small scale, densely packed appearance of pedestrians viewed from an elevated angle.

4.3 Confidence Threshold Analysis

To examine how detection outcomes shifted with confidence threshold, fifteen frames captured at the park intersections were manually annotated and compared against detector output across threshold values from 0 to 0.9. Table 4 reports the resulting confusion matrix counts at each threshold.

Table 4. Confusion matrix outcomes across confidence threshold values

Confidence Threshold	True Positive	False Negative	False Positive	True Negative
0.0	19	9	1	0
0.1	18	10	1	0
0.2	18	10	1	0
0.3	17	11	1	0
0.4	8	20	1	0
0.5	7	21	1	0
0.6	6	22	0	0
0.7	5	23	0	0
0.8	3	25	0	0
0.9	1	27	0	0

As shown in Table 4, raising the confidence threshold consistently reduced both true positive and false positive counts, with false positives eliminated entirely once the threshold reached 0.6. This pattern reflects the expected behavior of a single stage detector, whereby a higher acceptance threshold filters out lower confidence candidate boxes regardless of whether those candidates were correct or incorrect. Table 5 translates these counts into the four standard performance metrics across the same threshold range.

Table 5. Precision, recall, accuracy, and F1 score across confidence threshold values

Confidence Threshold	Precision	Recall	Accuracy	F1 Score
0.0	0.828	0.636	0.604	0.719
0.1	0.842	0.675	0.636	0.749
0.2	0.842	0.595	0.579	0.697
0.3	0.840	0.629	0.602	0.719
0.4	0.898	0.407	0.446	0.560
0.5	0.923	0.283	0.341	0.433
0.6	0.879	0.198	0.266	0.323
0.7	1.000	0.128	0.221	0.227
0.8	1.000	0.067	0.162	0.126

Confidence Threshold	Precision	Recall	Accuracy	F1 Score
0.9	0.500	0.010	0.103	0.021

The values presented in Table 5 show that precision generally increased as the confidence threshold rose, reaching a perfect score of 1.000 at thresholds of 0.7 and 0.8, before declining sharply at 0.9 due to the small number of remaining detections skewing the ratio. Recall, by contrast, declined steadily from 0.675 at a threshold of 0.1 down to 0.010 at 0.9, illustrating the familiar precision recall trade-off inherent to threshold based filtering. The F1 score, which balances both metrics, peaked at 0.749 when the threshold was set to 0.1, then declined progressively as the threshold increased further. These findings partially support the second proposition advanced in section two, confirming that detector performance varies systematically with threshold selection, although the intermediate threshold that maximized the F1 score, namely 0.1, differed from the threshold ultimately favored for heatmap construction, which prioritized precision over recall for reasons discussed in the following subsection.

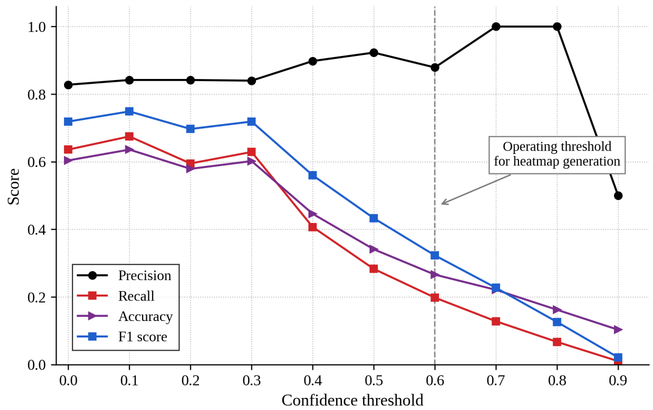

Figure 6. Precision, recall, accuracy, and F1 score across confidence threshold values

Figure 6 presents the performance evaluation of the detection model across different confidence thresholds using precision, recall, accuracy, and F1-score metrics. The results indicate that increasing the confidence threshold improves precision but reduces recall, as stricter filtering decreases false detections while increasing missed objects. The F1-score achieves its highest performance at lower thresholds, showing a better balance between precision and recall. The selected operating threshold at 0.6 provides a practical trade-off, achieving reliable detection confidence while maintaining sufficient performance for heatmap generation.

4.4 Selection of Operating Threshold for Heatmap Generation

Although the F1 score peaked at a low threshold, the requirements of heatmap construction differ from those of general detection benchmarking. Because false positive detections introduce spurious coloring into regions of the frame where no visitor is actually present, a threshold that tolerates excessive false positives can visually mislead viewers into perceiving crowding where none exists. Conversely, an excessively high threshold, such as 0.7 or 0.8, eliminates false positives entirely but also suppresses so many true positive detections that the heatmap fails to register genuine visitor movement, since recall at those thresholds fell to 0.128 and 0.067 respectively. Balancing these competing concerns, a confidence threshold of 0.6 was selected as the operating point for heatmap generation, since it represented the highest threshold at which false positive detections were fully eliminated while still retaining six true positive detections among the sample frames, yielding a precision of 0.879 and an accuracy of 0.266. This selection supports the third proposition, confirming that a threshold minimizing false positives while retaining a workable count of true positives produced the most visually coherent heatmap output in practice. Figure 7 shows a representative detection result obtained at the selected confidence threshold of 0.6.

Figure 7. SSD MobileNet V1 COCO detection output at a confidence threshold of 0.6

4.5 Motion Heatmap Outcomes at Point A and Point B

Applying the selected 0.6 confidence threshold, the combined detection and background subtraction pipeline produced motion heatmaps at both observation points across an extended two-hour recording window. The generated heatmaps were sampled at 40-minute intervals to evaluate the accumulation of detected movements over time. The intensity categories were assigned through visual inspection of the accumulated heatmap overlays using the OpenCV HOT colour ramp. Low intensity represents colouring limited to dark red tones along a single path, moderate intensity indicates orange regions covering one or more complete approach paths, and high intensity corresponds to yellow-to-white regions covering most of the traversable area within the field of view. Since the assessment was conducted by a single observer, the intensity classification is considered ordinal rather than a quantitative measurement.

Table 6. Ordinal assessment of accumulated heatmap intensity by elapsed time and location

Elapsed time (min)	Point A – dominant direction	Point A – intensity	Point B – dominant direction	Point B – intensity
40	Entrance → vendor stalls	Low	Mosque → play area	Moderate
80	Vendor stalls, with secondary branch from sports court	Moderate	Full intersection spread	Moderate–high
120	Vendor stalls sustained	High	All four approaches	High

Table 6 presents the accumulated heatmap intensity levels recorded at both observation points throughout the monitoring period. The table summarizes the temporal variation of movement intensity, allowing comparison of spatial activity patterns between Point A and Point B across different observation intervals. The results provide an overview of how detected movements accumulated and changed during the 120-minute recording session.

(a) Point A (b) Point B

Figure 8. Accumulated motion heatmap after 120 minutes at (a) Point A and (b) Point B

Figure 8 illustrates the final accumulated motion heatmap generated after the complete 120-minute observation period at both locations. The visualization provides a spatial representation of movement distribution, where areas with higher colour intensity indicate regions with more frequent detected activity. The comparison between Point A and Point B highlights differences in movement concentration and the effectiveness of the proposed detection pipeline in capturing activity patterns within the monitored area.

At Point A, the heatmap consistently exhibited its highest intensity along the pathway leading toward the vendor stall cluster, with a secondary but comparatively weaker branch of activity emerging from the direction of the basketball court, while the branch extending toward the general attraction area showed the lowest relative intensity among the three approaches. This pattern is consistent with the observation that visitors arriving from the park entrance tended to converge on the vendor stalls, with a smaller additional stream joining from the sports court side, together concentrating pedestrian density around the food and rental toy vendors. Because data collection coincided with the Sarapan Pagi Festival held between May 28 and June 5, 2022, additional temporary food stalls attracted visitors beyond typical baseline conditions, a contextual factor that should be considered when interpreting the observed intensity levels at both sites.

At Point B, the heatmap displayed its highest intensity near the immediate intersection and along the approach toward the children's play area, with moderate intensity extending toward the food stall direction, while the approach from the park exit showed comparatively lower intensity, indicating that the dominant pedestrian flow moved from the mosque approach toward the play area and food stalls rather than toward the exit. A localized zone of elevated intensity also appeared directly at the intersection itself, attributable to visitors who paused briefly to converse or select a rental toy car positioned nearby, illustrating how stationary lingering behavior, not just directional movement, can register strongly within a motion heatmap constructed from background subtraction.

4.6 Comparative Discussion with Prior Detection Studies

Relative to prior SSD based evaluations, the present detector's precision of 0.84 compares favorably with the 0.68 precision reported by Heredia and Barros-Gavilanes (2019) using a custom dataset, though it falls somewhat below the 0.89 precision the same authors achieved using COCO trained weights. The recall of 0.63 obtained here, however, sits below the 0.87 recall reported in general SSD benchmarking summaries by Kim, Sung, and Park (2020), and the overall accuracy of 0.60 remains below the 0.79 accuracy reported by Kumar, Zhang, and Lyu (2020) for an enhanced SSD variant trained on COCO. These comparisons suggest that although the baseline SSD MobileNet V1 COCO configuration used in this study delivers usable precision, its accuracy and recall could likely be improved through architectural enhancements or dataset fine tuning specific to elevated, small scale pedestrian imagery, consistent with the observation that COCO trained models tend to underperform on small, densely packed objects relative to the medium and large scale objects that dominate the COCO training distribution (Chen, Ding, Zou, Chen, & Li, 2021).

Table 7. Comparison of detection performance with prior SSD based studies

Study	Detector	Dataset	Precision	Recall	Accuracy	Platform
Heredia & Barros-Gavilanes (2019)	SSD-MobileNet	Custom	0.68	–	–	Embedded
Heredia & Barros-Gavilanes (2019)	SSD-MobileNet	COCO	0.89	–	–	Embedded
Kim, Sung, & Park (2020)	SSD	–	–	0.87	–	Desktop
Kumar, Zhang, & Lyu (2020)	Improved SSD	COCO	–	–	0.79	Desktop
Ahamad, Zaini, & Latip (2020)	SSD	COCO	–	–	–	Embedded
This study	SSD-MobileNet V1	COCO	0.84	0.63	0.60	Jetson Nano

Table 7 shows that the precision obtained in this study sits between the two configurations reported by Heredia and Barros-Gavilanes (2019), while recall and accuracy remain below the figures reported for desktop class deployments. The mean average precision of 23.2% obtained in this study also falls within a range comparable to, though on the lower end of, figures reported elsewhere for SSD COCO configurations, including 19.3% reported by Ahamad et al. (2020) and considerably higher figures near 79.8% reported for enhanced SSD variants (Kumar et al., 2020). This variation likely reflects differences in training data composition, backbone architecture depth, and the specific object scale distribution present in each evaluation dataset rather than any single dominant factor. Comparable degradation on small scale targets has been reported for other lightweight detectors, and dedicated architectural modification is typically required before small object performance approaches that achieved on medium and large objects (Amudhan & Sudheer, 2022). Regarding processing speed, the Jetson Nano platform sustained detection at roughly 20 frames per second in this study, a rate broadly consistent with embedded deployments reported elsewhere, though considerably below the frame rates achievable on desktop class graphics processing units exceeding 50 frames per second (Chen et al., 2021).

4.7 Implications for Public Health Monitoring and Space Design

Beyond the technical evaluation, the resulting heatmaps carry practical implications for how Elephant Park might be managed during periods of elevated infection risk. The concentration of visitor density around vendor stalls at both observation points suggests that these locations represent priority zones for reinforced health protocol signage, additional queue management measures, or physical distancing markers, since prolonged, static crowding around food vendors plausibly elevates the risk of airborne or droplet based transmission relative to areas experiencing only transient pedestrian flow (Rowe, Canosa, Drouffe, & Mitchell, 2021). Similarly, the identification of a specific lingering hotspot near the Point B intersection illustrates how motion heatmaps can reveal not only directional traffic patterns but also stationary gathering behavior that might otherwise escape casual observation, information that park administrators could use to inform the placement of seating, vendor stalls, or supervisory staff in future space planning decisions. Comparable monitoring pipelines have since been proposed as inputs to urban planning workflows (Mansouri et al., 2025), while studies of small urban green spaces indicate that the spatial distribution of amenities materially shapes both visitor satisfaction and dwell behaviour (Wang, Han, & Mei, 2022).

5. Conclusions

5.1 Conclusion

This study set out to determine whether an SSD based object detection algorithm could support the construction of a crowd motion heatmap at Elephant Park during the COVID-19 pandemic, and the empirical findings answer each of the three research questions posed at the outset. The SSD MobileNet V1 COCO detector proved capable of identifying human subjects within park footage, achieving a mean average precision of 23.2% and, at a confidence threshold of 0.3, a classification profile of 0.60 accuracy, 0.84 precision, 0.63 recall, and 0.72 F1 score. When the confidence threshold was tuned specifically for heatmap generation rather than for benchmark style detection accuracy, a threshold of 0.6 offered the most favorable balance, eliminating false positive detections while retaining sufficient true positive detections to sustain a coherent heatmap, yielding a precision of 0.87. Applying this threshold, the combined detection and background subtraction pipeline successfully generated motion heatmaps at both Point A and Point B that visibly traced dominant pedestrian flows toward vendor stalls, the children's play area, and the sports court, offering an interpretable visualization of where visitor crowding concentrated across a two hour observation window.

5.2 Research Limitations

Several limitations qualify these conclusions. The evaluation sample used to construct the confusion matrix comprised only fifteen manually annotated frames, a modest sample size that limits the statistical precision of the reported metrics and constrains generalization to the full range of conditions the system might encounter across different times of day, weather, or seasonal events. The motion heatmap technique itself proved sensitive to abrupt lighting changes and physical camera movement, both of which introduced spurious coloring unrelated to genuine pedestrian activity, a constraint that limited data collection to relatively stable lighting periods. The presence of the temporary Sarapan Pagi Festival during the recording window also introduced an atypical surge in vendor stall activity that may not reflect baseline visitor behavior at Elephant Park under ordinary conditions. Finally, the study relied on a single detector architecture, SSD MobileNet V1 COCO, without directly comparing alternative architectures such as YOLO variants or enhanced SSD configurations under identical field conditions, which constrains claims about the relative merit of the chosen approach against competing detectors.

5.3 Suggestions and Directions for Future Study

Building on these limitations, several directions merit further investigation. Future work could expand the manually annotated evaluation sample substantially to strengthen the statistical reliability of reported detection metrics and could incorporate a wider range of lighting and weather conditions to better characterize detector robustness across a full daily and seasonal cycle. Fine tuning the pretrained COCO weights on a dataset specifically curated from elevated, small scale pedestrian imagery similar to the Elephant Park deployment context may also improve recall without sacrificing the precision advantage already demonstrated. Comparative studies pairing SSD MobileNet with YOLO based alternatives, evaluated under identical hardware and site conditions, would help clarify which architecture offers the most favorable trade-off between detection accuracy and processing speed for embedded crowd monitoring applications of this kind. Finally, extending the adaptive heatmap decay mechanism to allow administrators to adjust color fade rates dynamically, alongside clearer heatmap color scale documentation, would improve the interpretability of the visualization for non technical park management staff tasked with translating detection output into actionable health protocol decisions.

Acknowledgement

The author gratefully acknowledges the guidance and support provided throughout this research by the supervising faculty members at the Department of Engineering Physics, Institut Teknologi Sumatera, as well as the technical staff who assisted with equipment preparation and field deployment at Elephant Park, Bandar Lampung.

Author Contributions

WS designed the study, assembled the monitoring hardware, implemented the detection and heatmap pipeline, conducted the field data collection, performed the analysis, and drafted the manuscript. IAS supervised the system design and reviewed the hardware configuration. ABH supervised the algorithm development and evaluation procedure. NDGD contributed to the interpretation of results and critically revised the manuscript. All authors read and approved the final version.

References

Ahamad, A. H., Zaini, N., & Latip, M. F. A. (2020). Person detection for social distancing and safety violation alert based on segmented ROI. Paper presented at In Proceedings of the 2020 10th IEEE International Conference on Control System, Computing and Engineering (pp. 113–118). IEEE.

Amudhan, A. N., & Sudheer, A. P. (2022). Lightweight and computationally faster hypermetropic convolutional neural network for small size object detection. *Image and Vision Computing, 119, 104396*. https://doi.org/10.1016/j.imavis.2022.104396

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172-186. https://doi.org/10.1109/TPAMI.2019.2929257

Chen, L., Ding, Q., Zou, Q., Chen, Z., & Li, L. (2021). Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *22*(6), 3234-3246. https://doi.org/10.1109/TITS.2020.2993926

Delibaşoğlu, İ. (2023). Moving object detection method with motion regions tracking in background subtraction. *Signal, Image and Video Processing*, *17*(5), 2415-2423. https://doi.org/10.1007/s11760-022-02458-y

Gao, G., Gao, J., Liu, Q., Wang, Q., & Wang, Y. (2025). A survey of deep learning methods for density estimation and crowd counting. *Vicinagearth*, *2*(1). https://doi.org/10.1007/s44336-024-00011-8

Heredia, A., & Barros-Gavilanes, G. (2019). Video processing inside embedded devices using SSD-MobileNet to count mobility actors. Paper presented at In Proceedings of the 2019 IEEE Colombian Conference on Applications in Computational Intelligence (pp. 1–6). IEEE.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv*. https://doi.org/10.48550/arXiv.1704.04861

Ihaddadene, N., & Djeraba, C. (2008). Real-time crowd motion analysis. Paper presented at In Proceedings of the 2008 19th International Conference on Pattern Recognition (pp. 1–4). IEEE.

Jeong, E., Kim, J., & Ha, S. (2022). TensorRT-based framework and optimization methodology for deep learning inference on Jetson boards. *ACM Transactions on Embedded Computing Systems*, *21*(5), 1-26. https://doi.org/10.1145/3508391

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230-243. https://doi.org/10.1136/svn-2017-000101

Kajabad, E. N., & Ivanov, S. V. (2019). People detection and finding attractive areas by the use of movement detection analysis and deep learning approach. *Procedia Computer Science, 156, 327–337*. https://doi.org/10.1016/j.procs.2019.08.209

Kim, J., Sung, J. Y., & Park, S. (2020). Comparison of Faster-RCNN, YOLO, and SSD for real-time vehicle type recognition. Paper presented at In Proceedings of the 2020 IEEE International Conference on Consumer Electronics – Asia (pp. 1–4). IEEE.

Kumar, A., Zhang, Z. J., & Lyu, H. (2020). Object detection in real time based on improved single shot multi-box detector algorithm. *EURASIP Journal on Wireless Communications and Networking*, *2020*(1), 204. https://doi.org/10.1186/s13638-020-01826-x

Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: A review. *Journal of Big Data*, *6*(1), 111. https://doi.org/10.1186/s40537-019-0268-2

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). *Microsoft COCO: Common objects in context*.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). *SSD: Single shot MultiBox detector*.

Mansouri, W., Alohali, M. A., Alqahtani, H., Alruwais, N., Alshammeri, M., & Mahmud, A. (2025). Deep convolutional neural network-based enhanced crowd density monitoring for intelligent urban planning on smart cities. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-90430-4

Marutotamtama, J. C., & Setyawan, I. (2021). Physical distancing detection using YOLO v3 and bird's eye view transform. Paper presented at In Proceedings of the 2021 2nd International Conference on Innovative and Creative Information Technology (pp. 50–56). IEEE.

Masita, K. L., Hasan, A. N., & Shongwe, T. (2020). Deep learning in object detection: A review. Paper presented at In Proceedings of the 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (pp. 1–11). IEEE.

Muharram, R. F. (2021). Implementasi artificial intelligence untuk deteksi masker secara realtime dengan TensorFlow dan SSD MobileNet berbasis Python. *Jurnal Rekayasa Komputasi Terapan*, *1*(3), 220-229. https://doi.org/10.30998/jrkt.v1i03.5832

Pramita, M. D., Kurniawan, B., & Utama, N. P. (2020). Mask wearing classification using CNN. Paper presented at In Proceedings of the 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (pp. 1–4). IEEE.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rezaei, M., & Azarmi, M. (2020). DeepSOCIAL: Social distancing monitoring and infection risk assessment in COVID-19 pandemic. *Applied Sciences*, *10*(21), 7514. https://doi.org/10.3390/app10217514

Rowe, B. R., Canosa, A., Drouffe, J. M., & Mitchell, J. B. A. (2021). Simple quantitative assessment of the outdoor versus indoor airborne transmission of viruses and COVID-19. *Environmental Research, 198, 111189*. https://doi.org/10.1016/j.envres.2021.111189

Song, Z., Ali, S., & Bouguila, N. (2020). Background subtraction using infinite asymmetric Gaussian mixture models with simultaneous feature selection. *IET Image Processing*, *14*(11), 2321-2332. https://doi.org/10.1049/iet-ipr.2019.1029

Süzen, A. A., Duman, B., & Şen, B. (2020). Benchmark analysis of Jetson TX2, Jetson Nano and Raspberry Pi using deep-CNN. Paper presented at In Proceedings of the 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (pp. 1–5). IEEE.

Wang, P., Han, L., & Mei, R. (2022). An impact asymmetry analysis of small urban green space attributes to enhance visitor satisfaction. *International Journal of Environmental Research and Public Health*, *19*(5), 2922. https://doi.org/10.3390/ijerph19052922

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212-3232. https://doi.org/10.1109/TNNLS.2018.2876865

Zhou, J., Zhou, Y., Wang, B., & Zang, J. (2019). Human–cyber–physical systems (HCPSs) in the context of new-generation intelligent manufacturing. *Engineering*, *5*(4), 624-636. https://doi.org/10.1016/j.eng.2019.07.015

Zuo, J., Jia, Z., Yang, J., & Kasabov, N. (2019). Moving target detection based on improved Gaussian mixture background subtraction in video images. *IEEE Access, 7, 152612–152623*. https://doi.org/10.1109/ACCESS.2019.2946230