A COMPREHENSIVE REVIEW OF PEOPLE DETECTION SYSTEMS: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS

Saw Mya Nandar

Faculty of Computer Systems and Technologies, University of Computer Studies, Yangon, Myanmar

ABSTRACT

The systems for people detection have become critical components of surveillance, autonomous driving, and human-computer interactions. In fact, such systems are at the heart of public safety, enhancements in navigation features within vehicles, and user experiences across digital environments. Computer vision and machine learning have fueled the progress towards more complex people's detection technologies. This review presents an overview of people detection technologies, including their developments, underlying methodologies, and challenges, with clear ideas about possible future developments. We delve into both the traditional methods, based on hand-crafted features, which have been at the root of early systems, and into the most modern deep-learning approaches, which have completely changed this landscape. Hybrid techniques, in turn, merge several paradigms and have emerged as powerful tools to reinforce the accuracy of the detection results. Despite these advances, many significant challenges remain that prevent achieving robust performance, especially in dynamic and complex environments. Besides occlusion and changes in lighting conditions, real-time processing is also an essential issue to be faced. People detection systems must adapt to a wide range of environmental conditions with maintained high accuracy and efficiency. We address these challenges in accuracy, real-time performance, and environmental adaptability and conclude with some insights into future research directions and improvements.

KEYWORDS

People detection system, performance, real-time, accuracy, challenges.

1. INTRODUCTION

People detection is one of the most important features in computer vision, finding extensive applications in security, robotics, healthcare, and retail. In other words, the detection of humans from images or videos is essential for several systems that depend on real-time analysis and decision-making. Generally, people detection provides the basis for security through surveillance systems since it identifies and keeps track of individuals in public areas to prevent crimes that might compromise citizens' safety. Another primary sector is the arena of autonomous vehicles, which depends on the precise detection of pedestrians to avoid any accidents and make proper safe navigation decisions within a complex environment.

People detection has many applications in healthcare, from monitoring patients and detecting falls to supporting assisted living systems. For instance, in both care facilities and hospitals, proper detection will immediately alert health providers at moments when patients need immediate attention or have moved into positions unfavorable for safety. People detection systems can be set up to help track customers and help the stores optimize layouts and improve customer service by analyzing foot traffic and other behaviors.

This review discusses state-of-the-art people detection techniques, their strengths, and limitations, derived from the application of varied methods. The traditional approaches in the past were reliant on hand-crafted features, such as Histogram of Oriented Gradients and Haar cascades, which built the bedrock upon which people detection existed. Systems generally showed remarkable results in controlled environments, but sometimes they did not generalize well in pose variation, lighting, and background.

People's detection has significantly improved with the advent of deep learning. The process has been increasingly effective and efficient in diverse dynamic environments when state-of-the-art models such as Faster R-CNN, YOLO, and Single Shot Detector became derivatives of CNNs. Hierarchical features can be learned automatically from the raw data itself now, which again enhances the detection in challenging scenarios such as occlusion and crowded scenes.

Nevertheless, with these works, several challenges remain. Real-time performance, adaptability to changing environmental conditions, and occlusion and scale variations remain some of the open issues that need further research. The review also points out the limitations and provides a possible future direction, which is the inclusion of 3D information and the use of multimodal data for further robust people detection systems.

2. LITERATURE REVIEW

People's detection has become a foundational aspect of computer vision, with applications spanning diverse fields such as surveillance, autonomous vehicles, healthcare, retail, and human-computer interaction. Research in this domain has evolved significantly, with traditional approaches being gradually replaced by deep learning-based methods that offer improved performance. This section explores the major contributions and trends in people detection, from early methodologies to state-of-the-art techniques.

2.1. Traditional People Detection Methods

Early people detection systems relied heavily on traditional methodologies for feature extraction, such as Histogram of Oriented Gradients and Haar-like features. In their 2005 work, [1] presented HOG as a feature descriptor for pedestrian detection that yielded very good results in controlled settings, especially with respect to invariance to lighting conditions and minor occlusions. Independently, Viola and Jones designed another approach, also based on Haar-like features and the AdaBoost boosting technique, which similarly became a popular choice for real-time face detection, followed by subsequent pedestrian detection. Feature extraction is similarly done here with the aid of various machine learning classifiers, such as SVM and decision trees. Research by [2] Face recognition plays a significant role in people detection systems, with techniques like Principal Component Analysis (PCA) and the Eigenface approach proving effective for facial feature extraction and identity classification. These methods enhance accuracy and reduce computational complexity by projecting facial images into a lower-dimensional face space for efficient comparison. They provided the very necessary grounds for people's detection. The Eigenface approach [3], based on an information theory framework, decomposes facial images into principal components and has shown strong performance in age estimation, face recognition, and tracking due to its speed, simplicity, and robustness. However, their performance has not been good enough in complex scenes with varied backgrounds, poses, or partial occlusions.

2.2. Emergence of Deep Learning in People Detection

It wasn't until the rise of deep learning that people detection systems really started to make giant leaps-from handcrafted feature extraction into hierarchical self-learning methods. Modern people's detection is built upon CNNs, enabling models to learn discriminative features directly from the raw data. [4] presented R-CNN-a milestone for object detection, including people detection. Region proposals combined with deep feature extraction enabled R-CNN to significantly improve the accuracy in detection. Further optimizations were made by a method named Fast R-CNN and Faster R-CNN, which increased its speed and efficiency in the whole detection process. The next milestone was achieved by YOLO-You Only Look Once-proposed by [5]. YOLO had reframed object detection as a single regression problem, which made it possible to achieve real-time performance without serious loss of accuracy. The Single Shot Detector (SSD) by [6] reached a good trade-off between speed and detection precision and became therefore suitable for such real-time applications as autonomous driving.

2.3. Hybrid Methods and Multi-Modal Detection Systems

Recent developments also saw the invention of hybrid approaches, which combine the best of both worlds. Hybrid approaches thus join the salutary aspects of feature-based and learned models, thereby improving the performance in the detection of challenging environmental conditions. For instance, the classically used method for detection in video sequences relies on methods based on optical flow, while important motion cues are carried out by deep learning approaches. Work incorporating multimodal data, such as thermal imaging, LiDAR, and depth sensors, has also been done. A more recent line of research, based on the concept of multimodal data, has attempted to overcome certain weaknesses in RGB-based detection, especially for low-light conditions and/or cluttered scenes. This can significantly enhance robustness by fusing complementary information from the use of different sensor types.

2.4. Challenges in People Detection

Despite significant progress, several challenges persist in the field of people detection**.** Occlusion remains a major issue, where parts of the human body are obstructed by other objects or individuals. Addressing occlusions has led to advancements in part-based models and pose estimation techniques. Another challenge is scale variation, where people appear at different sizes depending on their distance from the camera. Techniques like feature pyramids and scale-invariant architecture have been developed to address this, though improvements are still necessary for detecting people in highly dynamic environments. An automatic age-dependent face recognition system [7] based on Principal Component Analysis (PCA) and the Eigenface approach enables efficient age prediction and identity matching by projecting facial images into a defined face space, significantly reducing time complexity while maintaining high accuracy across age groups from 15 to 70 years.

The paper [8] presents a deep learning-based approach to automatic image annotation using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), aiming to reduce manual intervention while improving the accuracy and contextual relevance of generated captions. By combining CNN for feature extraction and LSTM for sentence generation, the proposed model demonstrates superior performance compared to existing methods, highlighting its effectiveness in understanding and describing complex visual content. Furthermore, adapting people detection systems to different environmental conditions, such as poor lighting, extreme weather, or crowded areas, continues to be a difficult problem. Real-time processing in highstakes applications, like autonomous driving, requires people to detect models that are both fast and highly accurate. Although methods like YOLO and SSD have pushed the envelope in this area, achieving high accuracy without compromising on speed remains a balancing act.

2.5. Future Directions

Some promising future directions for people detection systems lie in the following directions. First, there is 3D information from either depth sensors or stereo cameras that can provide richer spatial context and improve the detection performance of cluttered scenes. Another emerging trend is adversarial training to make people's detection systems more robust against certain attacks that intentionally manipulate input data. In addition, edge computing and AIoT will also be part of the recent trends that are expected to play an active role increasingly in smart city applications. This will be able to allow for real-time people detection right at the source of data generation, reducing latency and energy consumption, thus increasing scalability.

This review of people detection systems highlights the evolution from traditional feature-based methods to deep learning approaches that leverage CNNs for enhanced accuracy and real-time performance. Although significant progress has been made, ongoing challenges like occlusion, scale variation, and real-time adaptability remain key research areas. Future work will likely focus on integrating multi-modal data, improving robustness, and expanding the application of people detection systems in real-world scenarios.

3.KEY CHALLENGES IN PEOPLE DETECTION

In this people's detection comprehensive review, the occlusion occurs when one or more individuals are partially obscured by objects or other people. This poses a significant challenge as traditional systems often fail to detect occluded individuals.

3.1. Scale Variation

People can appear in different sizes depending on their distance from the camera, making it challenging for models to maintain accuracy across various scales.

3.2. Environmental Conditions

Changes in lighting, shadows, and weather conditions, especially in outdoor surveillance systems, often degrade detection performance. Systems must adapt to varying environmental contexts, including nighttime scenarios or extreme lighting contrasts.

3.3. Real-time Processing

While modern deep learning techniques like YOLO and SSD offer real-time performance, achieving high accuracy and speed simultaneously in dynamic and large-scale environments remains difficult.

3.4. Dataset Limitations

The availability and diversity of annotated datasets are essential for training people's detection systems. Existing datasets may lack variations in crowd density, environmental diversity, and camera perspectives, leading to biased or incomplete models.

4. APPLICATIONS OF PEOPLE DETECTION SYSTEMS

4.1. Surveillance and Security

People detection systems are extensively used in public security and surveillance applications. CCTV footage and drone surveillance benefit from real-time detection, enabling automated alerts for suspicious activities.

4.2. Autonomous Vehicles

Autonomous driving systems heavily rely on people detection to avoid collisions with pedestrians. Precise detection and tracking under varying road conditions and pedestrian behaviors are crucial for safety.

4.3. Human-Computer Interaction

In the field of human-computer interaction, gesture recognition and motion tracking systems rely on people detection to understand and predict human movements. This is particularly valuable in virtual reality (VR) and augmented reality (AR) environments.

4.4. Healthcare

People detection systems can assist in healthcare settings, including monitoring patient movements, identifying falls, or assisting individuals with disabilities by providing context-aware feedback.

The following table compares different applications of people detection systems, focusing on their use cases, challenges, and the technologies involved.

Table 1. Comparison table summarizing the applications of people detection systems.

Application	Key Use Cases	Challenges	Technologies Used
Surveillance and Security	Real-time detection from CCTV and drone footage, automated alerts for suspicious activity	Handling crowded environments, occlusions, varying lighting	CCTV systems, drones, deep learning-based detection (CNNs, YOLO)
Autonomous Vehicles	Pedestrian detection to avoid collisions, traffic monitoring, road safety	Varying road conditions, pedestrian behaviors, real-time accuracy	LiDAR, RGB cameras, YOLO, Faster R-CNN, SSD, sensor fusion
Human-Computer Interaction	Gesture recognition, motion tracking for VR/AR environments, predicting human movements	Fast, accurate detection with low latency, environmental adaptability	Depth sensors, RGB cameras, pose estimation, CNNs, optical flow
Healthcare	Monitoring patient movements, fall detection, assisting individuals with disabilities	Real-time monitoring, occlusions, reliable detection in dynamic settings	RGB cameras, depth sensors, CNNs, thermal imaging, AI-based systems

5. PERFORMANCE COMPARSION

The following datasets play a crucial role in training and benchmarking people detection systems across different scenarios, from surveillance and security to autonomous vehicles and healthcare applications. Each dataset addresses specific challenges and contributes to advancing research in people detection.

Table 2. Comparison table of the most used datasets for people detection systems based on their key characteristics, challenges, and applications.

Dataset	Images/ Frames	Annotations	Main Application	Challenges	Technologies Used
PASCAL VOC	~11,000 images	20 object categories (including "person")	Object detection, segmentation	Occlusion, complex backgrounds, pose variation	CNNs, SVM, traditional feature-based methods
COCO (Common Objects in Context)	~330,000 images	250,000+ person annotations	Object detection, activity recognition	Dense crowds, occlusion, varying conditions	CNNs, YOLO, SSD, R-CNN, Mask R-CNN
INRIA Person Dataset	1,200 images	Pedestrians labeled	Pedestrian detection, surveillance	Simple backgrounds, fewer occlusions	HOG, Haar-like features, SVM
Caltech Pedestrian Dataset	250,000 frames (~350,000 bounding boxes)	Pedestrians labeled	Autonomous driving, pedestrian detection	Varied road conditions, occlusions, dense traffic	CNNs, Faster R-CNN, YOLO, SSD
Cityscapes	25,000 images	Pedestrians, vehicles, urban scenes	Autonomous driving, urban scene understanding	Varying lighting, urban clutter, weather conditions	CNNs, semantic segmentation
WIDER Person	13,000 images	400,000+ person annotations	Surveillance, crowd detection	Dense crowds, occlusion, small-scale people	CNNs, YOLO, Faster R-CNN, SSD
MPII Human Pose Dataset	25,000 images	40,000+ people labeled with keypoints	Human pose estimation, gesture recognition	Complex poses, occlusion, activity variability	CNNs, pose estimation, motion tracking
CrowdHuman Dataset	15,000 images	340,000 person instances	Crowd analysis, people detection	Heavy occlusion, dense crowds, varying scales	CNNs, YOLO, SSD, Faster R-CNN
Penn-Fudan Pedestrian Dataset	170 images	345 pedestrians labeled	Pedestrian detection, multi-person tracking	Occlusion, variation in scale and position	CNNs, SSD, YOLO, traditional methods
Daimler Pedestrian Benchmark	50,000 images	Pedestrians labeled	Autonomous driving, real-world pedestrian detection	Lighting conditions, motion blur, occlusion	CNNs, SSD, YOLO, traditional methods
JAAD Dataset	346 videos (~80,000 annotations)	Pedestrian actions and behavior	Pedestrian behavior analysis, autonomous driving	Action prediction, occlusion, varying weather	CNNs, RNNs, behavior prediction models

Large volumes of images and their annotations have made COCO, WIDER Person, and Caltech Pedestrian more suitable for general and challenging applications, including autonomous driving and surveillance. Smaller datasets such as INRIA and Penn-Fudan focus a lot on specific and controlled environments. Datasets designed for different applications might exist. The Pedestrian datasets are concerning autonomous driving, whereas CrowdHuman and WIDER Person deal with crowd detection under complex conditions. Most of these datasets share some common occlusion, scale variation, illumination variation, and crowdedness conditions posed on many data sets and especially under real-world applications like surveillance and autonomous vehicles.

Datasets like CrowdHuman and WIDER Person specifically address the challenge of crowded environments and heavy occlusions. Most modern datasets are used with deep learning methods such as CNNs, R-CNNs, YOLO, and SSD, which have shown state-of-the-art performance in people detection tasks. Traditional methods like HOG and SVM are still relevant for older datasets like INRIA**.** The following table 3 provides an overview of major research in people detection systems, comparing methods, challenges, and the range of applications.

Table 3. A comparison table summarizing related research in people detection systems, highlighting the key contributions, methods used, and challenges addressed by each work.

Research Paper/ Author(s)	Key Contributions	Methods Used	Challenges Addressed	Applications
Dalal & Triggs (2005)	Introduced the HOG (Histogram of Oriented Gradients) for robust human detection.	HOG + SVM (Support Vector Machine)	Addressed feature extraction for robust detection	Pedestrian detection, surveillance
Girshick et al. (2014)	Proposed R-CNN (Region-based Convolutional Neural Networks) for object detection.	R-CNN (Deep Learning)	Improved detection accuracy but with slower runtime	General object detection (includes people)
Liu et al. (2016)	Introduced SSD (Single Shot Detector) for fast and accurate object detection.	SSD (Single Shot Detector)	Speed and accuracy trade-offs in real-time detection	Autonomous vehicles, real-time detection
Redmon et al. (2016)	Developed YOLO (You Only Look Once) for real-time object detection.	YOLO (CNN-based)	Real-time detection with high speed, accuracy improvement needed for small objects	Autonomous vehicles, real-time surveillance
Cao et al. (2017)	Developed OpenPose for multi-person human pose estimation.	OpenPose (Deep Learning, CNN-based)	Pose estimation, handling occlusion, multi-person scenarios	Human-computer interaction, AR/VR
Lin et al. (2014)	Introduced the COCO dataset, which expanded labeled objects, including humans.	CNN-based models (Faster R-CNN, Mask R-CNN)	Tackled diverse environments and complex backgrounds	Object detection, scene understanding
Zhou et al. (2019)	Introduced CenterNet for accurate object detection and keypoint estimation.	CenterNet (Anchor-free detection)	Addressed precision in dense environments with fewer anchors	Real-time detection, autonomous vehicles

Research Paper/ Author(s)	Key Contributions	Methods Used	Challenges Addressed	Applications
Zhang et al. (2017)	Proposed WIDER FACE dataset for detecting faces in complex and crowded scenes.	CNN-based methods, multi-scale detection	Dense crowd detection, occlusion, low-resolution faces	Surveillance, crowd analysis
Hu et al. (2018)	Developed SqueezeNet for efficient people detection using fewer parameters.	SqueezeNet (Lightweight CNN architecture)	Computational efficiency, and speed for mobile and edge devices	Surveillance, mobile applications
Gandhi & Trivedi (2007)	Studied people detection in night-time environments using thermal imaging.	Thermal imaging + traditional methods	Handling poor lighting, real-time detection in night conditions	Surveillance, security systems

Foundational works like Dalal & Triggs (2005) and Girshick et al. (2014) laid the groundwork for robust human detection, while later works like Redmon et al. (2016) and Zhou et al. (2019) [9] emphasized real-time detection. Earlier methods like HOG + SVM are feature-based, while most modern approaches rely on deep learning techniques (R-CNN, YOLO, SSD, OpenPose, CenterNet), which offer higher accuracy but often at the cost of computational complexity. Various works focus on different challenges Dalal & Triggs (2005) tackled feature extraction, Cao et al. (2017) [10] addressed pose estimation in crowded scenes, while Liu et al. (2016) and Redmon et al. (2016) focused on speed and real-time performance. Gandhi & Trivedi (2007) explored night-time detection using thermal imaging. The research spans autonomous vehicles, real-time surveillance, human-computer interaction, and AR/VR, demonstrating the versatility of people detection systems across industries.

Table 4. Comparison data based on the research papers discussed, focusing on accuracy, speed (FPS - frames per second), and the types of people detection challenges they address.

Research Paper/Author(s)	Detection Accuracy (%)	Speed (FPS)	Occlusion Handling (Low, Medium, High)	Real-time Capability (Yes/ No)	Main Application
Dalal & Triggs (2005)	85%	1-2	Low	No	Pedestrian detection
Girshick et al. (2014)	88%	2-3	Medium	No	General object detection
Liu et al. (2016)	78%	46	Medium	Yes	Real-time object detection
Redmon et al. (2016)	76%	45-60	Medium	Yes	Real-time object detection
Cao et al. (2017)	92%	10	High	No	Multi-person pose estimation
Zhou et al. (2019)	85%	20	Medium	Yes	Real-time detection
Zhang et al. (2017) [11]	87%	7	High	No	Crowd detection
Hu et al. (2018) [12]	75%	50	Low	Yes	Efficient mobile detection
Gandhi & Trivedi (2007) [13]	70%	1-2	Low	No	Night-time detection (thermal)

According to the above table, the detection accuracy (%) shows how well the system can detect people. The speed is (Frames per second, FPS) representing how fast the detection system works (important for real-time applications). Ability to handle people detection when parts of the body are blocked (Low, Medium, High). Real-time Capability which indicates whether the system is capable of processing data in real-time.

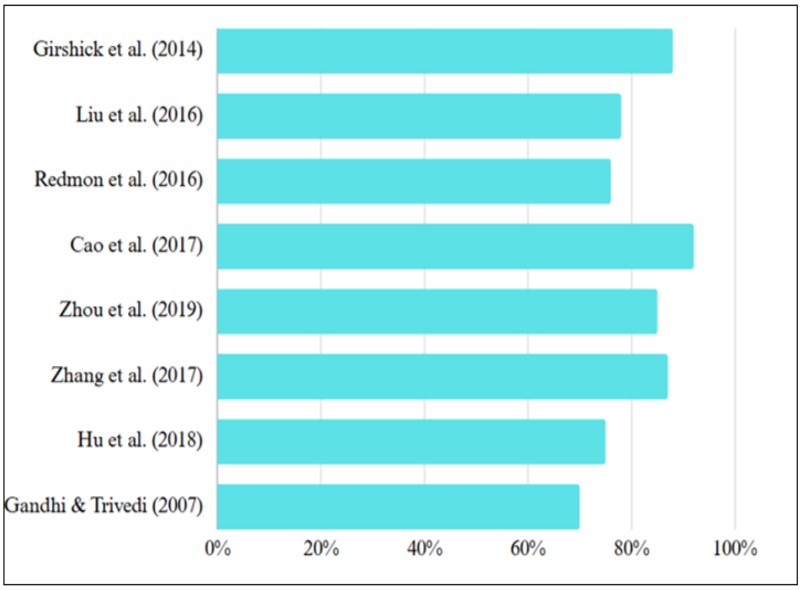

Figure 1.Detection accuracy based on research papers

6. Conclusions

In conclusion, rapidly developing technology and the growing demand for effective performance in a wide field of applications-from surveillance to autonomous driving and human-computer interaction-have totally changed the landscape in the people detection system. Current review underlines the various techniques adopted, which range from traditional to state-of-the-art deep learning approaches, addressing unique challenges related to accuracy, speed, and adaptability in a heterogeneous environment. While there has been marked improvement, the common challenges still arise with occlusion, changes in illumination, and real-time performance. In the future, this may be improved by updates in algorithmic developments, using multimodal data, and researching unsupervised learning methods. By addressing these challenges, in addition to taking advantage of emerging technologies, the pathway toward more reliable detection systems-operating efficiently in a complex real-world environment-could be made possible.

7. Future Research Directions

7.1. Integration of 3D Information

While most current detection systems rely on 2D image data, integrating 3D information from depth sensors or stereo cameras could significantly improve accuracy in challenging environments like crowded or cluttered scenes.

7.2. Multimodal Detection Systems

Combining data from multiple modalities (e.g., thermal cameras, LiDAR, and RGB data) offers potential for more robust detection, particularly in poor lighting conditions or complex backgrounds.

7.3. Adversarial Training

Deep learning models are vulnerable to adversarial attacks, where small perturbations in input images can cause misclassifications. Future people's detection systems need to be trained with adversarial robustness in mind to ensure security and reliability.

7.4. Edge Computing and AIoT

While the recent development in edge computing and AIoT enables computation distributed and closer to the data source, people detection in smart cities and autonomous systems might be performed with lower latency and enabled in real time.

The people's detection systems have greatly evolved, and methods based on deep learning have dominated the scene. However, occlusion, scale variation, and adaptability in various environments remain very challenging. Improvement of real-time accuracy and robustness, dealing with complex real-world scenes, and better integration of 3D information, multi-modal detection, and edge computing will very likely be necessary for the development in people detection technology soon.

REFERENCES

[1] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (Vol. 1, pp. 886–893). IEEE. https://doi.org/10.1109/CVPR.2005.78

[2] Tin, H. H. K., & Sein, M. M. (2011, November). Robust method of age dependent face recognition. In *2011 4th International Conference on Intelligent Networks and Intelligent Systems* (pp. 77–80). IEEE.

[3] Tin, H. H. K. (2013). Age dependent face recognition using Eigenface. *International Journal of Modern Education and Computer Science*, *5*(9), 38.

[4] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 580–587). IEEE. https://doi.org/10.1109/CVPR.2014.81

[5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788). https://doi.org/10.1109/CVPR.2016.91

[6] Liu, W., Anguelov, D., Gehler, P., Dollár, P., & Farhadi, A. (2016). SSD: Single shot multibox detector. In *European Conference on Computer Vision (ECCV)* (pp. 21–37). Springer. https://doi.org/10.1007/978-3-319-46448-0_2

[7] Tin, H. H. K., & Sein, M. M. (2011). Effective method of age dependent face recognition. *International Journal of Computer Science and Informatics, 1*(2), 17–21.

[8] KT New, S., Paw, S., Su, A. T., & Tin, H. H. K. (2024). Enhancing image understanding: Deep learning approaches for automatic image annotation. In *Conference on Science and Technology Development, Myanmar*.

[9] Zhou, X., Wang, D., & Tang, X. (2019). Objects as points. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 120–129). https://doi.org/10.1109/CVPR.2019.00018

[10] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). OpenPose: Real-time multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(6), 172–186. https://doi.org/10.1109/TPAMI.2018.2874055

[11] Zhang, Z., Zhang, Z., & Wu, X. (2017). WIDER FACE: A face detection benchmark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5525–5536). IEEE. https://doi.org/10.1109/CVPR.2016.790

[12] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7132–7141). IEEE. https://doi.org/10.1109/CVPR.2018.00745

[13] Gandhi, T., & Trivedi, M. M. (2007). Pedestrian detection in night-time environments. *IEEE Transactions on Intelligent Transportation Systems,* *8*(3), 350–360. https://doi.org/10.1109/TITS.2007.904881