Open Peer Review on Qeios

RESEARCH ARTICLE

An Intelligent Analytics for People Detection Using Deep Learning

Fatima Isiaka[1]

1 Nasarawa State University

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

People detection has become crucial in various applications, from security systems and surveillance to retail analytics and traffic management. With the advent of deep learning, particularly convolutional neural networks (CNNs), we've witnessed significant advancements in object detection accuracy and efficiency. This paper explores the power of intelligent analytics driven by deep learning for people detection, highlighting its benefits, challenges, and potential applications. The main aim is to build a people behaviour detection framework through body language, events, objects around people and their postures to determine the behaviour of people and environment genuinely based on given attributes like walking (still or moving), sitting (still or fidgeting), running (steady paise or high speed) and standing (still or fidgeting). These attributes contribute to detecting people's behaviour from a given input of video sequence, both in real-time or pre-recorded from MATLAB using three different deep learning algorithms (CNN, You Only Look Once (YOLO) and Faster region CNN). The results obtained were compared to determine which model best suits people's behaviour detection.

1. Introduction: Deep Learning for People Detection, a powerful tool

People detection plays a crucial role in various computer vision applications, enabling tasks such as surveillance, security, and autonomous navigation. Deep learning, a subfield of machine learning, has revolutionised this field by providing powerful models capable of accurately detecting humans and objects around them.

1.1. Traditional Approaches vs. Deep Learning

Traditional people detection methods relied heavily on handcrafted features, such as shape, size, and motion patterns. However, these features can be highly susceptible to noise and variations in illumination. In contrast, deep learning models learn features directly from data, allowing them to capture complex patterns and variations. afddfggf [1][2][3][4]

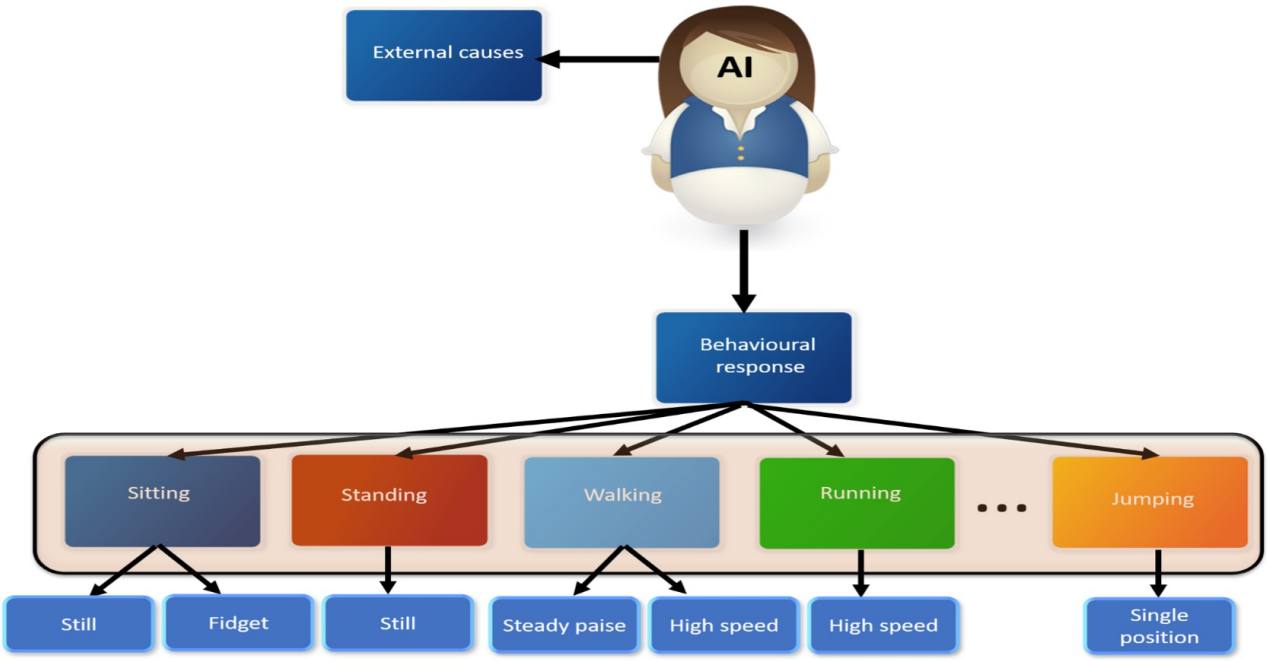


Figure 1. Convolutional Neural network for People detection.

1.2. Deep Learning Architectures

Several deep learning architectures have been developed specifically for people detection. These include:

- Convolutional Neural Networks (CNNs) CNNs are the backbone of many image recognition tasks. They consist of multiple layers of filters and uses Artificial Intelligence (A)I to learn and identify specific features within the image (Figure 1).
- You Only Look Once (YOLO): YOLO is a real-time object detection algorithm that processes the entire image in a single pass. It predicts bounding boxes and class probabilities for each potential object (Figure 2).
- Faster Region-based Convolutional Neural Networks (Faster R-CNN): Faster R-CNN generates region proposals from the image and then uses a CNN to classify and refine the bounding boxes (Figure 3).

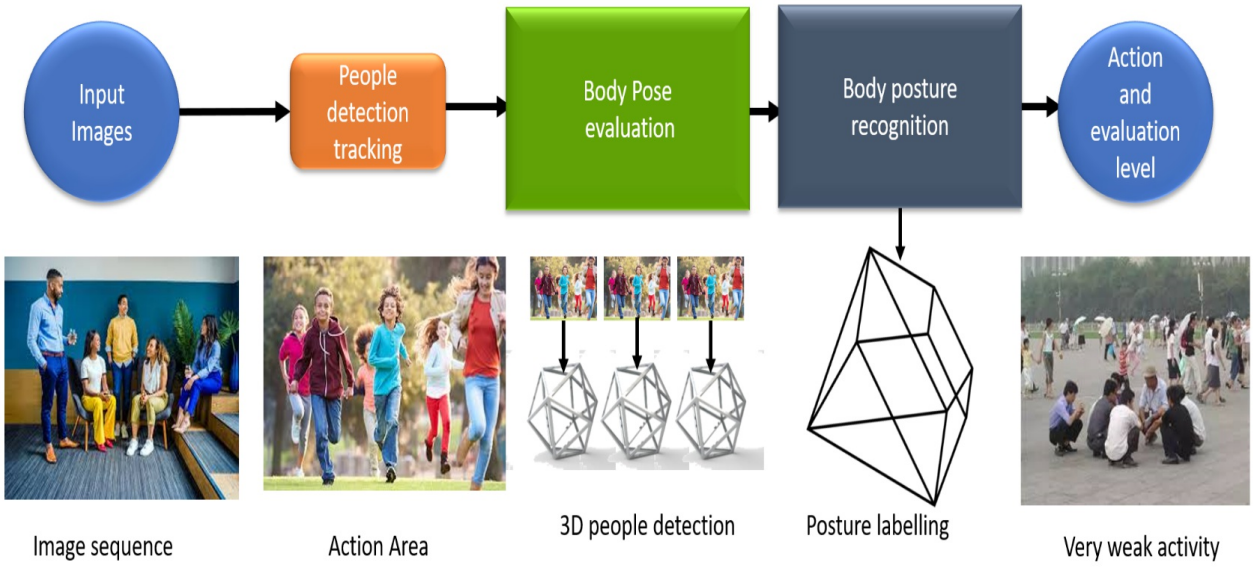


Figure 2. YOLO for People detection framework.

1.3. Advantages of Deep Learning for People Detection

Despite the advantages of people detection, there are its drawbacks such as:

- High Accuracy: Deep learning models can achieve state-of-the-art accuracy in people detection, outperforming traditional methods by a wide margin.
- Robustness: Deep learning models are less sensitive to noise and variations in illumination, making them suitable for real-world applications.
- Real-Time Detection: Architectures like YOLO enable real-time people detection, critical for surveillance and autonomous systems.
- Generalizability: Deep learning models can be trained on large datasets, allowing them to generalize well to different scenarios and environments.

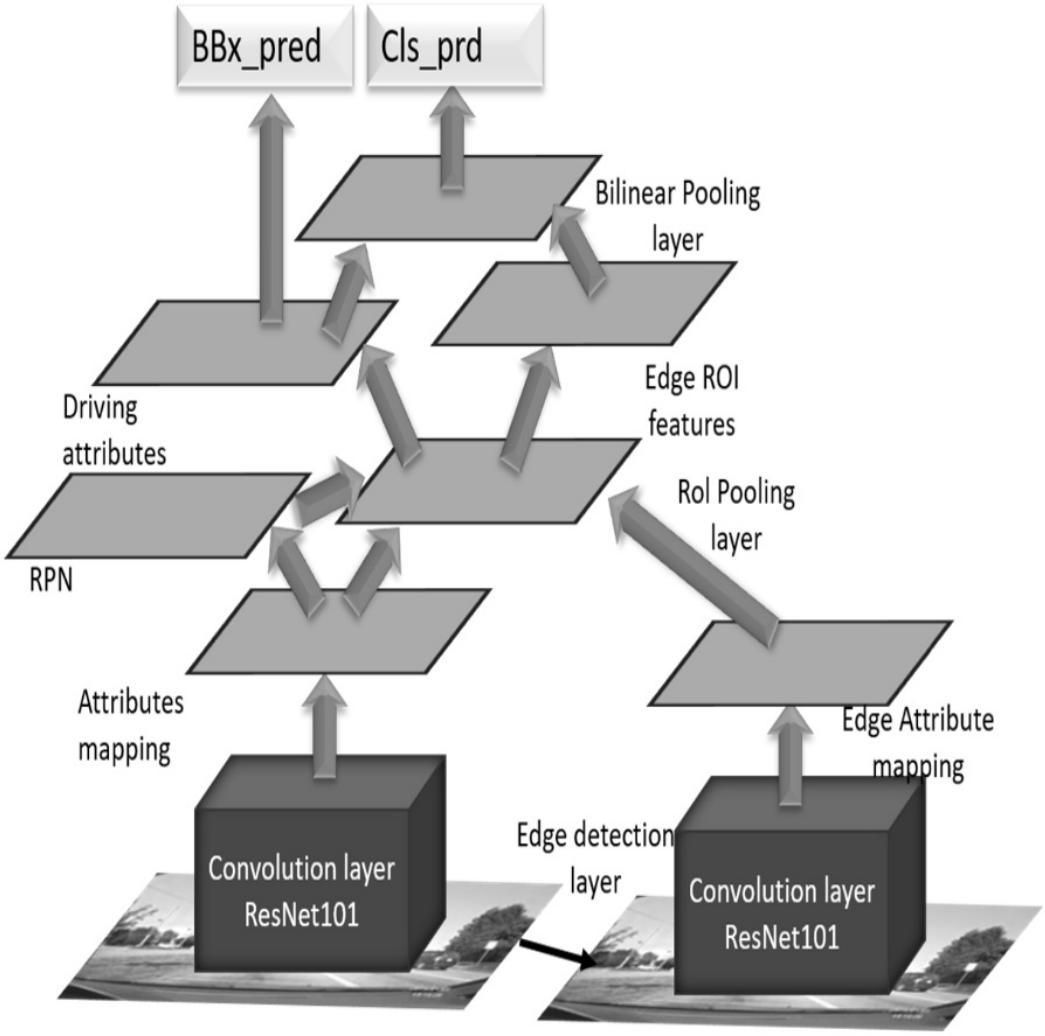


Figure 3. Faster Region Convolutional Neural Network for People Detection.

2. Literature Review: Applications of People Detection with Deep Learning

One of the foremost applications of deep learning in people detection is surveillance and security where people detection is essential for monitoring public spaces, detecting intruders, and preventing crime. The aspect enhances security benefits by keeping crime sequences and prevention protocols [5][6][7][8][9][10][11]. Autonomous vehicles in deep learning models are one of the applications of people detection which are used in autonomous vehicles to detect pedestrians and other road users, ensuring safe navigation. Autonomous vehicles (AVs) are rapidly becoming a reality, promising safer and more efficient transportation. But at the heart of this technological revolution lies a crucial component "people detection". AVs must be able to accurately perceive and interpret their surroundings, especially the presence and behaviour of pedestrians and other road users [12][13][14][15][16][17][18]. From sensors to sophisticated algorithms, AVs rely on a suite of sensors, including cameras, LiDAR (Light Detection and Ranging), and radar, to gather data about their environment. This data is then processed by powerful algorithms that analyze the scene and identify potential threats, including people.

Even if there are advantages in people detection, there are also some challenges such as accurately detecting people in a

dynamic environment which presents several challenges such as:

- Occlusion: People can be partially or fully hidden by objects, making them difficult to detect.
- Variable Appearance: People come in different shapes, sizes, and clothing, and can be moving or stationary.
- Lighting Conditions: Varying light levels can significantly impact sensor performance.
- Weather: Rain, snow, and fog can obscure visibility and hinder detection.

To overcome these challenges, researchers are exploring innovative approaches such as deep learning for Neural networks (NN) which are trained on massive datasets to learn complex patterns and make accurate predictions about people's presence and movements. Multi-sensor fusion (MSF) combines data from multiple sensors, such as cameras and LiDAR, to provide a more comprehensive and robust understanding of the environment with contextual information by leveraging information about the environment, such as road markings and traffic signals, which can help to refine people detection algorithms [19][20][21][22][23][24][25][26]. Going beyond detection requires understanding human behaviour, the future of AVs lies not just in detecting people, but in understanding their intentions and predicting their actions. This involves analyzing body language, facial expressions, and even pedestrian behaviours to anticipate potential risks [27][28][29][30][31].

2.1. The Impact on Safety and Efficiency

Accurate people detection is paramount for the safety and efficiency of AVs. By reducing accidents caused by human error, AVs have the potential to drastically improve road safety. They can also optimize traffic flow by adapting to changing conditions and anticipating potential bottlenecks. The development of advanced people detection systems is transforming the way we interact with our environment. AVs are not only shaping the future of transportation, but also driving innovation in other fields, like robotics, security, and healthcare. Looking ahead, one can envision AV technology continuing to evolve, people detection algorithms will become even more sophisticated and reliable. The integration of artificial intelligence and machine learning will create a future where AVs can navigate complex environments with unprecedented accuracy and safety, paving the way for a more efficient and human-centric future for transportation [32][33][34][35][36].

One of the realistic safety and efficient impacts is the retail analytics for people detection which can provide insights into customer behaviour, track foot traffic, and optimize store layout. The healthcare aspect of deep learning for people detection can assist in patient monitoring, fall detection, and medical image analysis. The crowd control aspect uses algorithms that can help manage crowds, prevent overcrowding, and ensure safety during large events.

Deep learning has emerged as a powerful tool for people detection. Its ability to learn complex features and its robustness to variations make it an ideal choice for a wide range of applications, from surveillance and security to autonomous vehicles and healthcare. As technology continues to advance, we can expect even more sophisticated and accurate people detection models in the future. Deep learning algorithms, specifically CNNs, excel at extracting complex features from images and videos. This ability enables them to accurately identify and localize people in diverse scenarios, even with varying poses, lighting conditions, and occlusions.

2.2. Applications of Intelligent People Detection

There is a vast amount of experience in the convenience of*'Hey Google'* or *'Alexa,'* where devices respond to users' voices. But behind the scenes, much more is happening, and intelligent people detection is playing a crucial role. This technology, fueled by advancements in artificial intelligence (AI) and computer vision, is revolutionizing various industries, enhancing security, streamlining operations, and ultimately improving our lives. While the benefits of intelligent people detection are undeniable, it's crucial to address ethical concerns surrounding data privacy and potential misuse. Implementing robust data security measures, transparent data usage policies, and responsible AI development are essential to ensure this technology is used ethically and for the betterment of society. As AI technology continues to advance, intelligent people detection will become even more powerful and ubiquitous. By harnessing its potential responsibly and addressing ethical concerns, one can unlock a future where this technology enhances our safety, convenience, and overall quality of life {m/35,36,**39?**/}.

Intelligent analytics powered by deep learning has revolutionized people detection, leading to more accurate, efficient, and adaptable solutions across diverse industries. While challenges remain, ongoing research and development are paving the way for even more robust and intelligent systems that will shape our future interactions with technology. By addressing ethical concerns and leveraging the potential of deep learning, we can harness its power to create safer, more efficient, and more insightful environments. These are some of the objectives of this paper; as a contribution to the field of research, this paper has highlighted the disadvantages and provides ways of developing and building people detection strategies that not only detect people in an unguarded environment but also their behavioural qualities such as attributes that make a genuine individual based on body gestures and stance. The preceding sections discuss more on the methods and results.

3. Method: Applying Deep Learning Models for People Detection

People behaviour detection is a challenging task, as it requires the ability to understand the complex interactions between people and their environment. Traditional methods for people behaviour detection, such as hand-crafted features and rule-based systems, are often unable to capture the subtle nuances of human behaviour. CNNs, on the other hand, can learn these nuances from data, making them a more powerful tool for people behaviour detection.

3.1. Using CNN for People Behaviour Detection

There are several different ways to apply CNNs to people's behaviour detection. One common approach is to use a pre-trained CNN model (Figure 1), such as VGGNet or ResNet, and then fine-tune the model on a dataset of people behaviour videos (Matlab videos). This approach can be effective for tasks such as recognising gestures, detecting falls, and tracking people in a scene. The CNN model is pre-trained and can be used to detect people behaviour in new videos. The model was run on a frame-by-frame basis, and it can output a probability distribution over the possible behaviours.

This probability distribution is used to classify the behaviour of the people in the video.

3.2. Using YOLO for People Behaviour Detection

To train the YOLO model for people behaviour detection, a large dataset of images and corresponding labels from MATLAB archive was required. The data include a variety of people performing different actions, such as walking, running, jumping, sitting, and interacting with objects. Preprocessing involves cleaning the data, resizing images, and annotating them with bounding boxes around the people.

3.3. Model Training

The YOLO model is trained on the preprocessed dataset using a deep learning framework such as TensorFlow. The model learns to identify the presence of people and objects in their surroundings and classify their behaviour based on the visual features extracted from the input images. Once the model was trained, it was evaluated on a validation set to assess its accuracy and generalization performance. One of the common evaluation metrics used is the mean average precision (MAP) and intersection over union (IoU). These metrics measure the model's ability to correctly detect and locate people, as well as its precision in classifying their behaviour and behaviour of objects in their suroundings, this is done also for the CNN model.

Applying YOLO for people's behaviour detection is a powerful and versatile approach that can be leveraged across a range of domains. By utilizing preprocessed data, training the model effectively, and evaluating its performance, it is possible to develop accurate and real-time systems that can identify and classify human behaviour with high precision.

3.4. Applying Faster Region Convolutional Neural Network (F-RCNN) for People Behaviour Detection

Faster R-CNN uses a single-shot object detection algorithm that combines the power of deep CNNs with region proposal networks (RPNs). RPNs generate candidate object bounding boxes, which are then refined using a CNN-based classifier and regressor. Faster R-CNN has been widely adopted for human pose estimation, action recognition, and people behaviour detection.

The dataset from the MATLAB archive was collected that includes videos or images capturing people performing various behaviours like the one used for CNN and preprocessing the data by resizing, cropping, and normalizing the inputs. Feature extraction focused on deep features from the input data. The features represent high-level semantic information about the people's appearance, pose, and motion, and generate candidate object bounding boxes for people in the input. The region refinement and classification uses a regressor to refine the candidate bounding boxes and classify the people based on their behaviours such as 'walking,' 'running,' 'sitting,' or 'talking', to predict the best possible attribute of a genuine person.

Applying Faster Region-based Convolutional Neural Networks (Faster R-CNN) for people behaviour detection offers a powerful and efficient approach to accurate and real-time behaviour recognition. Its end-to-end training and high accuracy

make it a promising tool for various computer vision applications in security, human-computer interaction, and healthcare. The three models were then compared on different epochs and analysed, the results for the three models were compared and discussed in the following result section.

4. Result: Comparing the three models

People behaviour detection is a challenging task in computer vision, with applications in surveillance, healthcare, and human-computer interaction. Convolutional Neural Networks (CNNs), You Only Look Once (YOLO), and Faster RCNN are the three popular deep learning architectures that have been successfully applied to this problem in this paper.

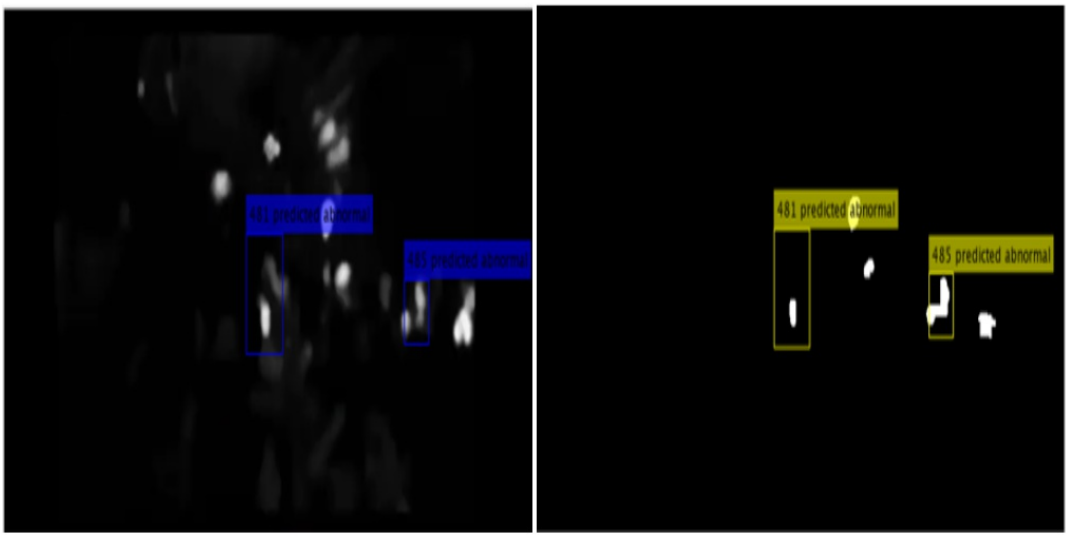
The following table (Table 1) compares the three architectures in terms of accuracy, speed, and complexity:

Table 1.

Architecture	Accuracy	Speed	Complexity
CNN	High	Slow	High
YOLO	Medium	Fast	Low
Faster RCNN	High	Medium	High

CNNs are well-suited for image recognition tasks with convolutional layers, each of which learns to extract specific features from the input image. The output of the convolutional layers is then typically fed into a fully connected layer, which makes the final prediction.

CNNs are very effective for people and object behaviour detection. Figure 4a and 4b show detection objects in a dark scene from a video sequence as predicted abnormal behaivour. Figure 5c shows an image of a person not detected as abnormal or normal but objects and environment around detected as exhibiting predicted abnormal behaviour.



(a) Detection objects in a dark scene as exhibiting predicted abnormal behaviour

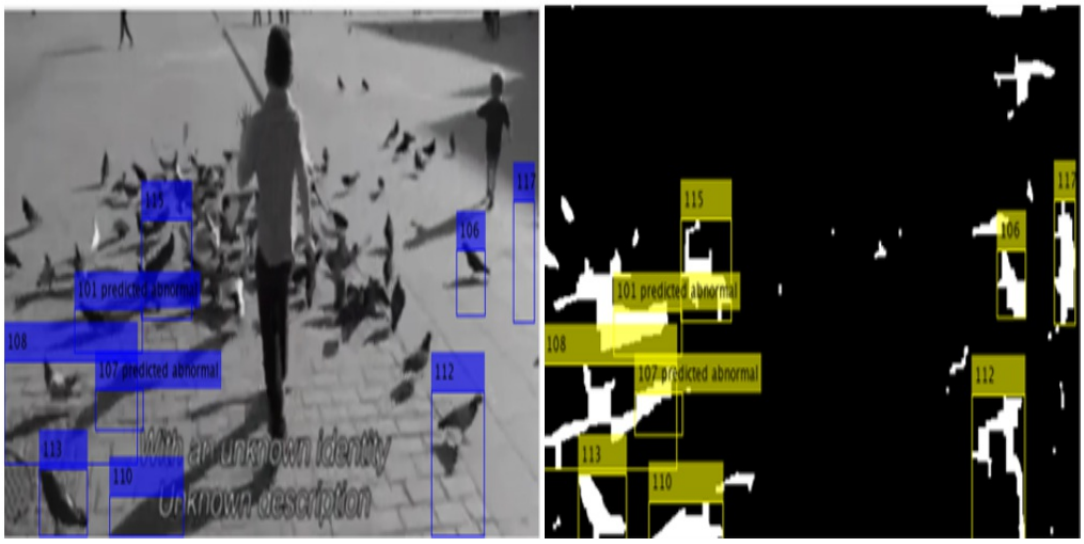
(b) Foreground image detection of objects in a dark scene as exhibiting predicted abnormal behaviour

(c) Image of a person not detected as abnormal or normal but around detected as exhibiting predicted abnormal behaviour

(d) Foreground image detection of objects in a dark scene as exhibiting predicted abnormal behaviour

Figure 4. 1st sequence of detected and predicted images of both foreground and background scene of people and objects' behaviour around them.

The YOLO model is a single-shot object detection algorithm that requires multiple forward passes through the network to make a prediction, YOLO predicts a single pass. This makes YOLO much faster than CNNs, but it can also lead to a decrease in accuracy. YOLO is effective for people detection, but it is not as accurate as CNNs. However, YOLO is much faster than CNNs, making it a good choice for applications where speed is important. Figure 5b and 5b show detected running child in a dark scene as normal behaviour and objects as predicted abnormal, the foreground image detection of objects in a dark scene as exhibiting predicted abnormal behaviour.



(a) Detected running child in a dark scene as normal behaviour and objects as predicted abnormal

(b) Foreground image detection of objects in a dark scene as exhibiting predicted abnormal behaviour

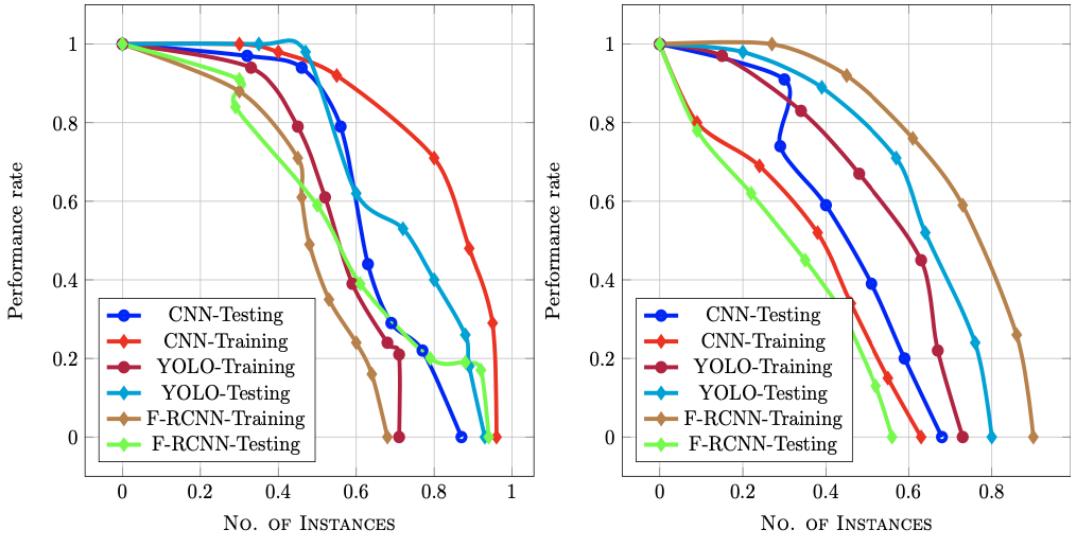
(c) Detected running child playing as normal but object around him detected as exhibiting predicted abnormal behaviour

(d) Foreground image detection of objects in a dark scene as exhibiting predicted abnormal behaviour

Figure 5. 2nd sequence showing detected and predicted images of both foreground and background scenes of people and objects' behaviour around them.

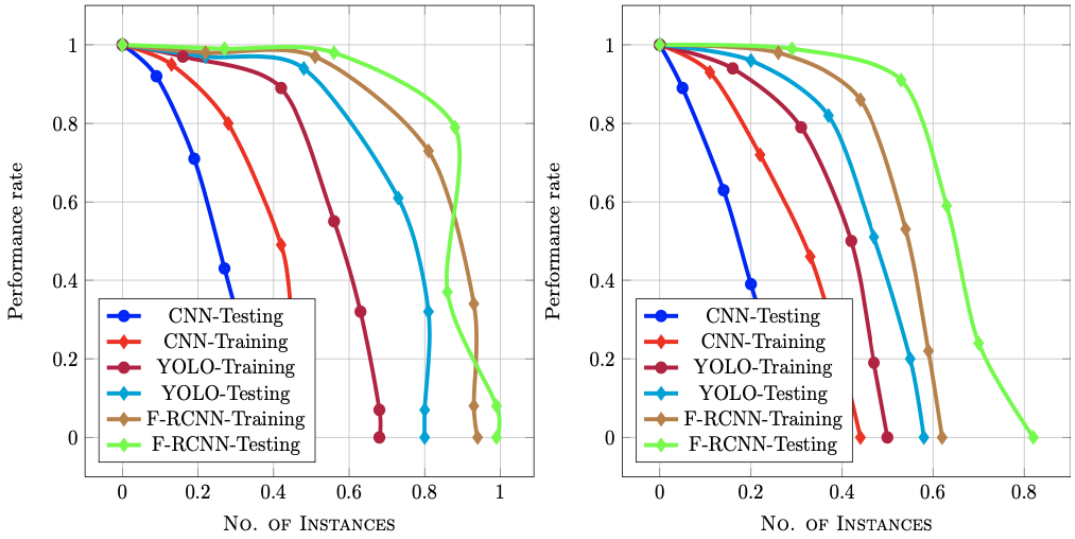
Faster RCNN uses a two-stage object detection module that generates a set of candidate object proposals. The second stage of object detection classifies each proposal and refines its bounding box. Faster RCNN is more accurate than both CNNs and YOLO, but it is also slower. This makes Faster RCNN a good choice for applications where accuracy is important, but speed is not a major concern. Figure 6c and 6d indicate the performance of F-RCNN in both training and

testing sets shows the highest tracking accuracy of 100%, while the performance of YOLO and CNN test sets show the highest tracking accuracy of 100% and 0.80%.



(a) Performance of YOLO test set show highest tracking accuracy of 100%.

(b) Performance of F-RCNN training set shows highest tracking accuracy of 99%



(c) Performance of F-RCNN for testing set shows highest tracking accuracy of 100%.

(d) Performance of F-RCNN for testing set shows highest accuracy of 100%

Figure 6. Performance of Tracking operation characteristics for image detection using CNN, YOLO and Faster RCNN.

5. Conclusion

The choice of which architecture to use for people behavior detection depends on the specific requirements of the application. If accuracy is the most important factor, then a CNN or Faster RCNN is a good choice. If speed is the most

important factor, then YOLO is a good choice.

Acknowledgements

The author would like to thank Nasarawa State University, for the support and sponsor of this paper.

References

1. ^*Qi Guo, Eugene Agichtein. (2012). Beyond dwell time: Estimating document relevance from cursor movements and other post-click searcher behavior. In: Proceedings of the 21st international conference on world wide web. pp. 569–578.*
2. ^*Eugene Agichtein, Eric Brill, Susan Dumais, Robert Ragno. (2006). Learning user interaction models for predicting web search result preferences. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval. pp. 3–10.*
3. ^*Sabina-Cristiana Necula. (2023). Exploring the impact of time spent reading product information on e-commerce websites: A machine learning approach to analyze consumer behavior. Behavioral Sciences. 13(6):439.*
4. ^*Richard Atterer, Monika Wnuk, Albrecht Schmidt. (2006). Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. In: Proceedings of the 15th international conference on world wide web. pp. 203–212.*
5. ^*Aleksandr Chuklin, Ilya Markov, Maarten De Rijke. (2022). Click models for web search. Springer Nature.*
6. ^*Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, Sheelagh Carpendale. (2012). Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. IEEE Transactions on Visualization and Computer Graphics. 18(12):2689–2698.*
7. ^*Liqiong Deng, Marshall Scott Poole. (2010). Affect in web interfaces: A study of the impacts of web page visual complexity and order. Mis Quarterly. :711–730.*
8. ^*Keith S. Vallerio, Lin Zhong, Niraj K. Jha. (2006). Energy-efficient graphical user interface design. IEEE Transactions on Mobile Computing. 5(7):846–859.*
9. ^*Zhicheng Liu, Jeffrey Heer. (2014). The effects of interactive latency on exploratory visual analysis. IEEE transactions on visualization and computer graphics. 20(12):2122–2131.*
10. ^*Majken K. Rasmussen, Esben W. Pedersen, Marianne G. Petersen, Kasper Hornbaek. (2012). Shape-changing interfaces: A review of the design space and open research questions. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 735–744.*
11. ^*Fang Chen, Natalie Ruiz, Eric Choi, Julien Epps, M. Asif Khawaja, et al. (2013). Multimodal behavior and interaction as indicators of cognitive load. ACM Transactions on Interactive Intelligent Systems (TiiS). 2(4):1–36.*
12. ^*Mi Jeong Kim, Mary Lou Maher. (2008). The impact of tangible user interfaces on spatial cognition during collaborative design. Design Studies. 29(3):222–253.*

13. ^Peiling Wang, William B. Hawk, Carol Tenopir. (2000). Users' interaction with world wide web resources: An exploratory study using a holistic approach. Information processing & management. 36(2):229–251.
14. ^Bjorn B. De Koning, Huib K. Tabbers. (2011). Facilitating understanding of movements in dynamic visualizations: An embodied perspective. Educational Psychology Review. 23:501–521.
15. ^Izak Benbasat, Peter Todd. (1993). An experimental investigation of interface design alternatives: Icon vs. Text and direct manipulation vs. menus. International Journal of Man-Machine Studies. 38(3):369–402.
16. ^Jenifer Tidwell. (2005). Designing interfaces: Patterns for effective interaction design. " O'Reilly Media, Inc.".
17. ^Andrew Chan, Karon MacLean, Joanna McGrenere. (2008). Designing haptic icons to support collaborative turn-taking. International Journal of Human-Computer Studies. 66(5):333–355.
18. ^Andrew Chan, Karon MacLean, Joanna McGrenere. (2008). Designing haptic icons to support collaborative turn-taking. International Journal of Human-Computer Studies. 66(5):333–355.
19. ^David Kirsh. (2013). Embodied cognition and the magical future of interaction design. ACM Transactions on Computer-Human Interaction (TOCHI). 20(1):1–30.
20. ^Gyanendra Sharma, Richard J. Radke. (2021). Multi-person spatial interaction in a large immersive display using smartphones as touchpads. In: Intelligent systems and applications: Proceedings of the 2020 intelligent systems conference (IntelliSys) volume 3.: Springer pp. 285–302.
21. ^Christos Kouroupetroglou. (2014). Enhancing the human experience through assistive technologies and e-accessibility. IGI Global.
22. ^Abdulmotaleb El Saddik, Mauricio Orozco, Mohamad Eid, Jongeun Cha. (2011). Haptics technologies: Bringing touch to multimedia. Springer Science & Business Media.
23. ^Stephen Woods. (2013). Building touch interfaces with HTML5: Develop and design speed up your site and create amazing user experiences. Peachpit Press.
24. ^Hari Prasath Palani. (2013). Making graphical information accessible without vision using touch-based devices.
25. ^Martin Hecher, Robert Mostl, Eva Eggeling, Christian Derler, Dieter W. Fellner. (2011). "Tangible culture"–designing virtual exhibitions on multi-touch devices. Information services & use. 31(3-4):199–208.
26. ^Lik-Hang Lee, Tristan Braud, Simo Hosio, Pan Hui. (2021). Towards augmented reality driven human-city interaction: Current research on mobile headsets and future challenges. ACM Computing Surveys (CSUR). 54(8):1–38.
27. ^Weizhi Meng, Yu Wang, Duncan S. Wong, Sheng Wen, Yang Xiang. (2018). TouchWB: Touch behavioral user authentication based on web browsing on smartphones. Journal of Network and Computer Applications. 117:1–9.
28. ^Katharina Reinecke, Abraham Bernstein. (2011). Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. ACM Transactions on Computer-Human Interaction (TOCHI). 18(2):1–29.
29. ^Lars Kayser, Andre Kushniruk, Richard H. Osborne, Ole Norgaard, Paul Turner, et al. (2015). Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: A framework for understanding users' needs. JMIR human factors. 2(1):e3696.
30. ^Marilyn A. Walker, Stephen J. Whittaker, Amanda Stent, Preetam Maloor, Johanna Moore, et al. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. Cognitive Science. 28(5):811–840.
31. ^Christian Crumlish, Erin Malone. (2009). Designing social interfaces: Principles, patterns, and practices for improving

the user experience. " O'Reilly Media, Inc.".

32. ^*Scott W. Ambler. Tailoring usability into agile software development projects. In: Maturing usability: Quality in software, interaction and value.: Springer 2008. pp. 75–95.*
33. ^*Xuan Wang, SK Ong, Andrew Yeh-Ching Nee. (2016). Multi-modal augmented-reality assembly guidance based on bare-hand interface. Advanced Engineering Informatics. 30(3):406–421.*
34. ^*Roope Raisamo. (1999). Multimodal human-computer interaction: A constructive and empirical study. Tampere University Press.*
35. ^*Jamil Hussain, Wajahat Ali Khan, Taeho Hur, Hafiz Syed Muhammad Bilal, Jaehun Bang, et al. (2018). A multimodal deep log-based user experience (UX) platform for UX evaluation. Sensors. 18(5):1622.*
36. ^*Jamil Hussain, Anees Ul Hassan, Hafiz Syed Muhammad Bilal, Rahman Ali, Muhammad Afzal, et al. (2018). Model-based adaptive user interface based on context and user experience evaluation. Journal on Multimodal User Interfaces. 12:1–16.*