Real-Time Billiard Shot Stability Detection Based on YOLOv8

Boning Yang and Wei Qi Yan

Auckland University of Technology, Auckland 1010 New Zealand

ABSTRACT

This book chapter presents a real-time system for detecting a player's billiard shot, based on the YOLOv8 neural network. The system comprises a real-time object detection model and a real-time monitoring system. The model focuses on detecting four classes: The cue ball, hand, cue tip, and bridge hand (hand support point). The project involved iterative model training on a customized dataset, eventually achieving a YOLOv8 model with 95% accuracy. A player's shot is detected by simulating slope change of cue stick during aiming, using the cue stick tip and bridge hand. Overall, the project highlights the immense potential of YOLOv8 in sports applications.

Keywords: YOLOv8, Deep learning, OpenCV, real-time object detection, iterative model training, fine-tunning, stable detection

INTRODUCTION

Despite advances in billiard training, assessing and improving shot stability remain challenging for players and coaches. In billiards, slight changes in stroke angle and force critically impact the trajectory, making stability vital for hit success (Chen, 2023). Traditionally, coaching depends on subjective judgment, struggling to quantify subtle movement changes. Self-assessment by players also lacks precision without professional guidance.

The rise of computer vision and deep learning leads to new opportunities in sports (Zhou, 2023; Can, 2022; Cao & Yan, 2022; Zhu & Yan, 2022), including billiards. Implementing these technologies for real-time shot stability analysis provides instant, objective, and precise feedback. This aids players across all levels in mastering games and refining skills, marking a transformative integration of technology and sports coaching (Herrera, et al., 2008).

This book chapter presents the development of a real-time billiards shot detection by using deep learning, computer vision, and YOLOv8. The effectiveness hinges on several core areas: Adapting YOLOv8 for precise tracking of subtle movements in billiards, evaluating accuracy against other methods, examining processing speed for real-time feedback, integrating the system into training routines, assessing adaptability to various playing conditions and techniques, and gathering user feedback from players and coaches. The key contributions of this book chapter include creating a unique dataset specifically for billiard shot stability, optimizing a customized YOLOv8 model for enhanced detection accuracy and speed, and implementing a practical system that provides instantaneous, objective feedback, proven valuable in real-world training and competitive environments. This book chapter underscores the significance of a technologically advanced, data-driven approach in revolutionizing billiards training.

LITERATURE REVIEW

Table 1 Real-time target detection ranking based on MS COCO dataset.

Rank	Model	Box AP	FPS	References	Year
1	YOLOv6-L6	57.2	46	*YOLOv6 v3.0: A Full-Scale Reloading*	2023
2	PRB-FPN6-MSP	57.2	27	*Parallel Residual Bi-Fusion Feature Pyramid Network for Accurate Single-Shot Object Detection*	2020
3	YOLOv7-E6E	56.8	36	*YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*	2022
4	YOLOv7-D6	56.6	44		2022
5	YOLOv7-E6	56	56		2022
6	YOLOv7-W6	54.9	84		2022
7	PP-YOLOE+_X	54.7	45	*PP-YOLOE: An evolved version of YOLO*	2022
8	PP-YOLOE+_L	54.0	78		2022
9	PRB-FPN-MSP	53.3	94	*Parallel Residual Bi-Fusion Feature Pyramid Network for Accurate Single-Shot Object Detection*	2020
10	Gold-YOLO-L	53.28	116	*Gold-YOLO: Efficient Object Detector via Gather-and-Distribute Mechanism*	2023

In Table 1, the models were trained on the publicly available dataset MS COCO. YOLO variants dominate the leaderboard for real-time object detection on the MS COCO dataset, with YOLOv6-L6 taking the top spot. This algorithm achieves the highest Box Average Precision (AP) at 57.2, with a frame rate of 46 frames per second (FPS), suggesting a balance between accuracy and speed suitable for real-time applications.(Li et al., 2023) The table indicates the trend where more recent YOLO versions, like YOLOv7, offers a variety of trade-offs between precision and speed, with versions attaining higher FPS potentially favoring applications where speed is crucial, even at the expense of some accuracy.(Wang et al., 2023) Notably, PRB-FPN6-MSP matches YOLOv6-L6 in Box AP but at a lower speed, highlighting the efficiency of YOLO architectures.(Chen et al., 2021) These results collectively underscore advancements in YOLO algorithms, particularly in their ability to deliver high-accuracy detection in real-time scenarios, a key factor for applications like the billiards shot stability detection system discussed earlier.

YOLO series, particularly YOLOv6-L6, YOLOv7, PP-YOLOE, and Gold-YOLO, have marked significant advancements in the field of real-time object detection.(Wang et al., 2023) YOLOv6-L6 stands out with its network design enhancements, anchor-assisted training, and self-distillation strategies, boosting accuracy and speed. PRB-FPN distinguishes itself with a unique architecture designed to detect objects of varying sizes efficiently, featuring bi-fusion modules and a residual design for improved precision.

YOLOv7 brings architectural and training optimizations with elements like VoVNet and CSPNet, setting new standards for speed and accuracy in real-time detection. PP-YOLOE innovates on the YOLO architecture with an anchor-free design, a robust backbone, and dynamic label assignment, achieving a balance between detection performance and inference speed.(Wang et al., 2023)

Gold-YOLO introduces the Gather-and-Distribute mechanism, improving information fusion, and employs unsupervised pre-training, showing substantial improvements in accuracy and speed over its predecessors.(Wang et al., 2023)

Collectively, these iterations of the YOLO series showcase a trajectory of continuous improvement (Liang et al., 2022; Yu & Yan, 2020), each bringing novel features and optimizations that enhance the model adaptability and performance in real-time object detection applications. This evolution cements YOLO's position as a leading solution in the domain.

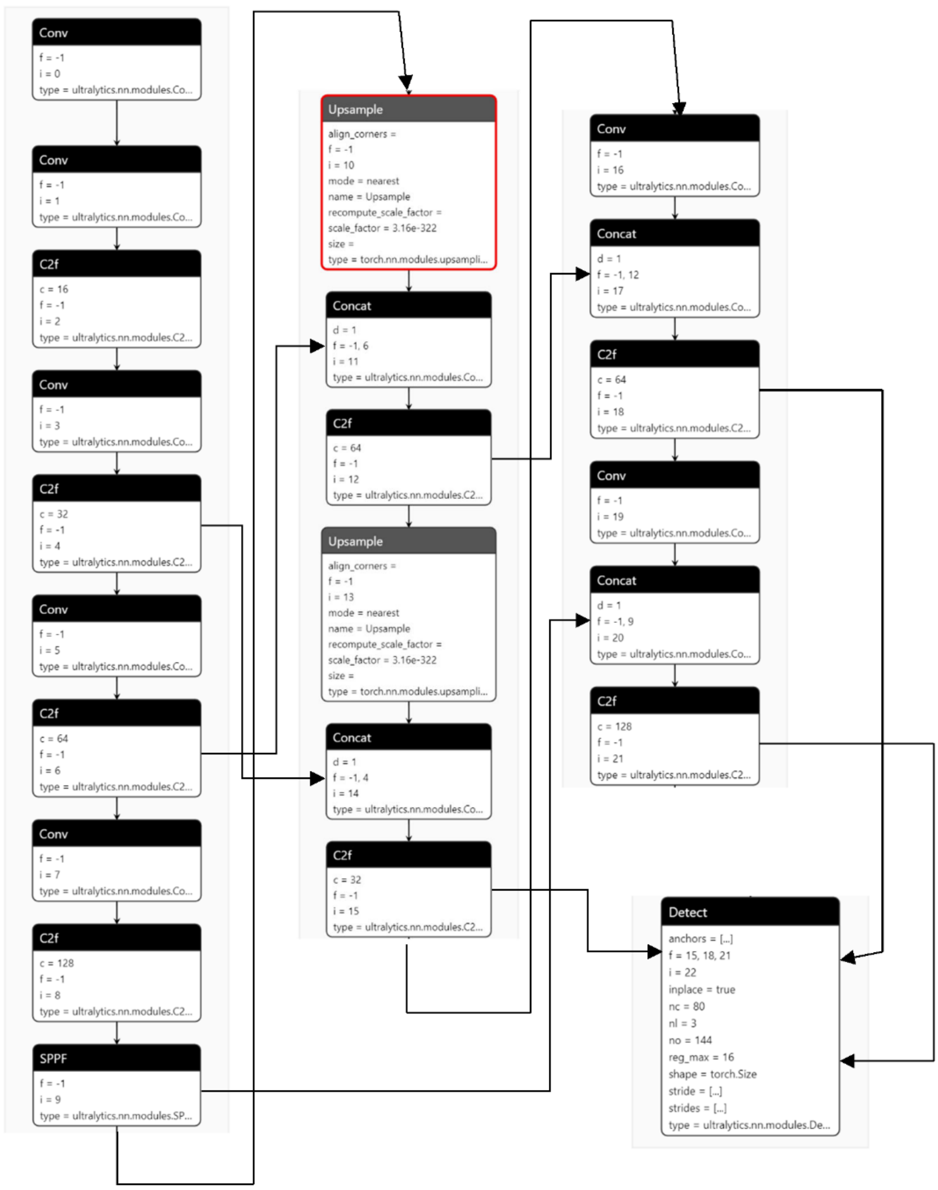

Figure 1 The structure of YOLOv8 model

Figure 1 shows the network structure of the YOLOv8 model in .pt format opened using the

Netron software. By referring to the official diagram provided by the open-source community Open-mmlab, it can be understood that modules 0-9 are the Backbone part of YOLO, 10-21 are the Neck part, and 22 is the Head part.

YOLO, an acronym for "You Only Look Once", is a deep learning framework for visual object detection. It primarily comprises three segments: Backbone, Neck, and Head. The Backbone, typically a deep convolutional neural network such as VGG, ResNet, or DarkNet, is tasked with deriving basic spatial and contextual features from unprocessed images.(Ayob et al., 2021; Demetriou et al., 2023; Sujatha et al., 2023; Zhou et al., 2019) The Neck component, such as FPN or PANet, is added post-Backbone and aimed at integrating and enhancing features of varying depths and resolutions to capture multi-scale information of objects. The Head, on the other hand, directly predicts the object's class, location, and size based on the features obtained from the Neck. In essence, the entire YOLO framework firstly extracts image features through the Backbone, enhances and integrates them via the Neck, and subsequently outputs the final object detection results through the Head.(Huang et al., 2023; Vasanthi & Mohan, 2023; Zhang et al., 2021)

YOLO framework is renowned for its unique approach to object detection. In contrast to traditional methods which often require multiple scans of an image, YOLO captures the entirety of an image in just one forward pass, allowing for real-time detection. To ensure the quality of predictions made by YOLO or any similar object detection systems, there are a series of key metrics commonly employed:

Precision is:

$$\text{Precision} = \frac{\text{True Positives(TP)}}{\text{True Positives(TP)} + \text{False Positives(FP)}} \tag{1}$$

Recall is:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{2}$$

F1 Score is:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

IoU quantifies the overlap between the predicted bounding box and the ground truth bounding box. It is a pivotal metric in object detection, particularly for determining how well the model predicted bounding box aligning with the ground truth. It is expressed as:

$$IoU = \frac{\text{Area of Overlap (Intersection)}}{\text{Area of Union (Union)}} \tag{4}$$

These evaluation metrics are imperative for comprehending the efficacy of object detection models. They not only offer insights into the model's precision and coverage but also guide improvements, allowing for the optimization of the model's performance.

METHODOLOGY

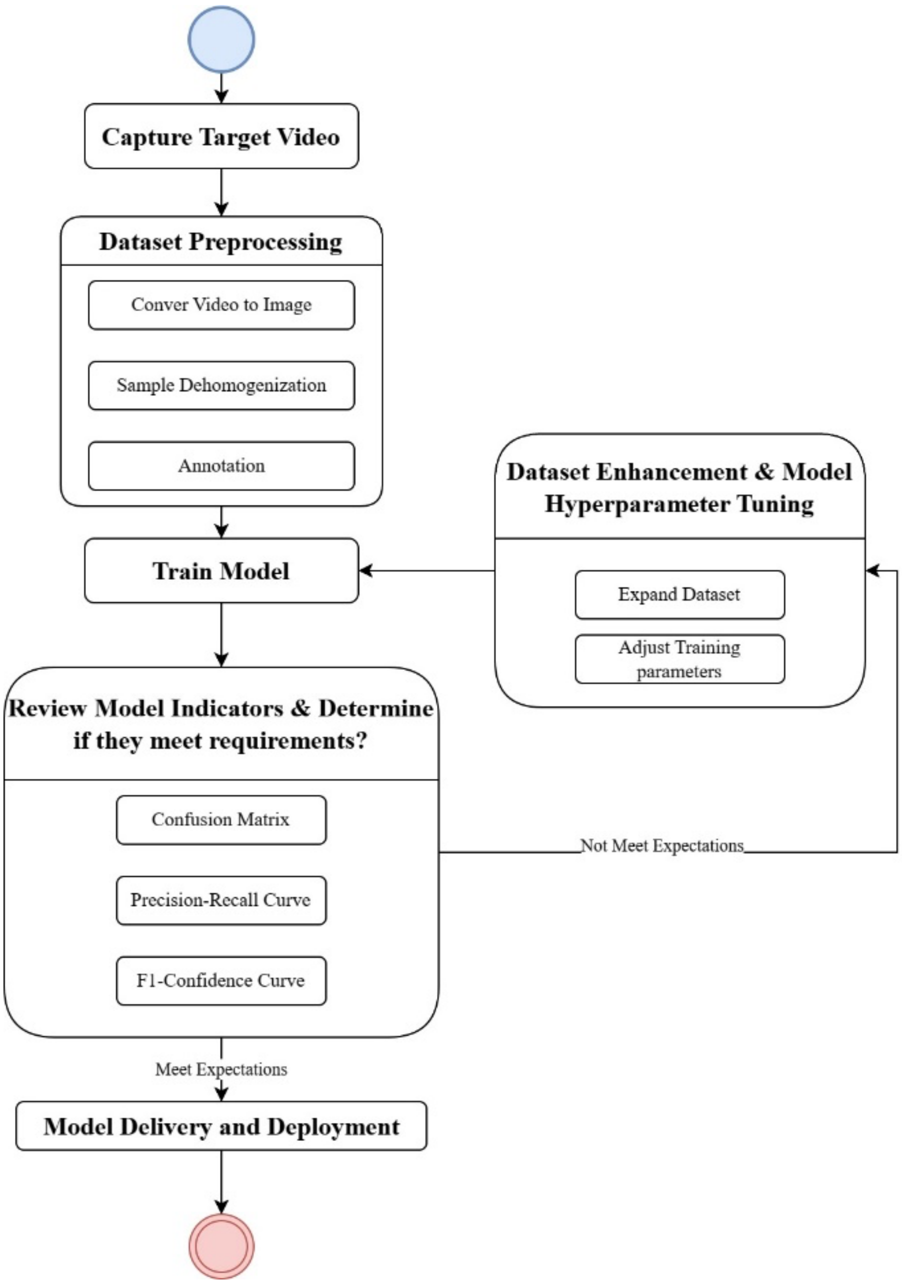

Figure 2 Model iteration and optimization flowchart.

Figure 2 offers a systematic representation of the steps involved in capturing a target video, preprocessing, training a model, and the eventual deployment of the model after evaluating its performance indicators.

The experimental process laid out is iterative in nature, wherein the model undergoes cycles of training and evaluation until it meets the desired threshold. The ultimate goal is for the model to attain a recognition accuracy, quantified by the metric mean Average Precision (mAP), of over 0.95 for all the object classes. This high threshold ensures that the model not only correctly identifies the objects but also does so with high confidence across various scenarios and conditions.

During the model evaluation process, regarding models that do not meet expectations, we analyze the confusion matrix of the model to determine which class has a problematic recognition rate. Then, we increase the proportion of this class in the dataset through sample augmentation to improve its recognition rate. Sample augmentation involves rotating the original samples of this class, the original sample set is,

$$S = \{s_1, s_2, \dots, s_{472}\}$$

The rotation operation is,

$$s_i' = \text{Rotate}(s_i, -15°)$$

The sample after rotation is

$$S' = \{s_1', s_2', \dots, s_{472}'\}$$

The combined sample is

$$S_{\text{total}} = S \cup S'$$

For the classes that are highly homogenous, we are use of the Structural Similarity Index Measure (SSIM) algorithm to reduce these samples, as the original samples are created by converting videos into images, a process that generates many redundant samples. (Fuentes-Hurtado et al., 2022)

The samples were manually labeled to create the dataset used for model training. Four classes were marked within the samples: “0”, “sp”, “hp”, and “hand”. “0” means the cue ball in billiards. “sp” is the bridge hand of the player, which is the hand formation used to support and guide the cue stick. “hp” represents the hitting point on the cue stick (also referred to as the tip). “Hand” stands for the player's hand. The annotation details are shown in Figure 3.

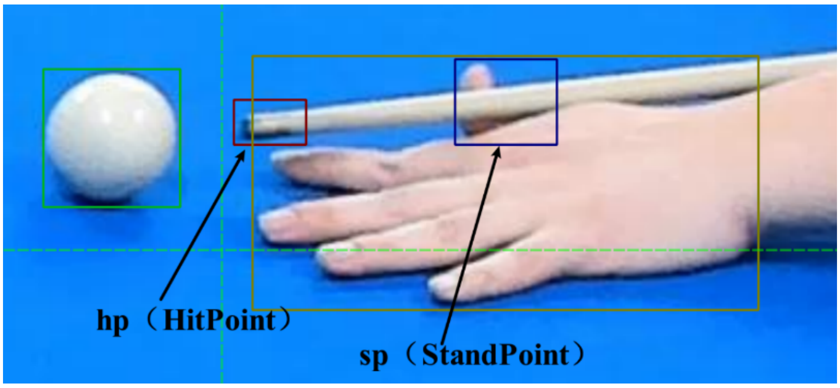

Figure 3 Sample Annotation Explanation

In addition to modify the dataset, a few adjustments were also made to the model training parameters (Lu et al, 2016; Lu et al., 2017; Lu et al., 2018; Lu et al., 2020). One such parameter is *imgsz*, which controls the initial image size at the input layer of the model. By increasing this parameter from the default value of 640 to 1280, a significant improvement in the mAP (mean Average Precision) value for all classes was observed.

MODEL DEPLOYMENT

After obtaining a high-accuracy model, we normalized the prediction boxes for “tip” and “sp” to obtain key points (Li, et al., 2016). These two points correspond to the contact point of the cue stick in a billiards game, also known as the “tip”, and the “sp”, where the player supports the cue stick. By connecting these two points, we form a white line, as shown in Figure 4. When the player strokes the ball, the stability of the slope of this line is monitored to determine whether the player's shot is stable.

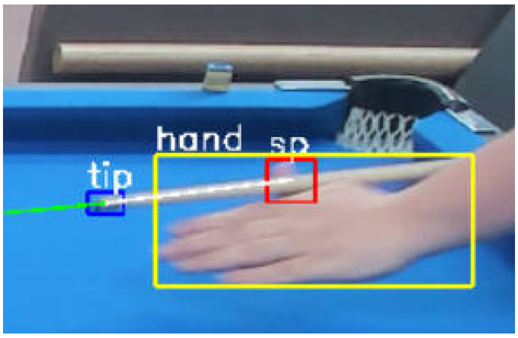

Figure 4 Details of the demo

Figure 5 Billiards Shooting DEMO

The slope detection is expressed as:

$$\mathrm{Slope}(m) = \frac{\Delta y}{\Delta x} \tag{5}$$

where,

$$\begin{aligned} \Delta y &= y_2 - y_1 \\ \Delta x &= x_2 - x_1 \end{aligned} \tag{6}$$

where, $(x_1,\ y_1)$ and $(x_2,\ y_2)$ are the coordinates of two consecutive points.

$$Jittering = \left|\arctan(m) - \arctan(m_{\mathrm{prev}})\right| > threshold \tag{7}$$

where m is the current calculated slope, m_{prev} is the previously calculated slope. To detect significant variation between two consecutive slopes, we compute their difference (transforming the slope into angles using the arctangent function) and see if this difference surpasses a set threshold. If it exceeds the threshold, we consider jitter to be detected.

As shown in Figure 5, after we design the slope detection, the captions will be displayed on the prediction screen and sound will be emitted to prompt the player whether their aiming and shooting are stable.

CONCLUSION

In this book chapter, a new model was trained to track four classes in real time with 95% accuracy, which offers a novel method to assess players' stability during billiard shots. It focuses on two key points: “Hand bridge” (sp) in a player's hand and the cue-stick contact point (“hp”). Analyzing the steadiness and relationship of these points provides insights into a player's cue alignment and shot consistency. This tool, beneficial in both real-time and post-game reviews, can act as an alternative to traditional coaching, aiding players to identify and improve their techniques, leading to enhanced gameplay (Yan, 2019; Yan, 2023).

REFERENCE

Ayob, A. F., Khairuddin, K., Mustafah, Y. M., Salisa, A. R., & Kadir, K. (2021). Analysis of Pruned Neural Networks (MobileNetV2-YOLO v2) for Underwater Object Detection. Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019 (Vol. 666, pp. 87–98). Springer Nature.

Cao, X. (2022) Pose Estimation of Swimmers from Digital Images Using Deep Learning. Master’s Thesis, Auckland University of Technology.

Cao, X. and Yan, W. (2022) Pose estimation for swimmers in video surveillance. Multimedia Tools and Applications, Springer.

Chen, P.-Y., Chang, M.-C., Hsieh, J.-W., & Chen, Y.-S. (2021). Parallel Residual Bi-Fusion Feature Pyramid Network for Accurate Single-Shot Object Detection. IEEE Transactions on Image Processing, 30, 9099–9111.

Chen, Z. (2023) Real-Time Pose Recognition for Billiard Players Using Deep Learning. Research Report, Auckland University of Technology, New Zealand.

Demetriou, D., Mavromatidis, P., Robert, P. M., Papadopoulos, H., Petrou, M. F., & Nicolaides, D. (2023). Real-Time Construction Demolition Waste Detection Using State-of-The-Art Deep Learning Methods; Single – Stage vs Two-Stage Detectors [Preprint].

Fuentes-Hurtado, F., Delaire, T., Levet, F., Sibarita, J.-B., & Viasnoff, V. (2022). MID3A: Microscopy Image Denoising meets Differentiable Data Augmentation. International Joint Conference on Neural Networks (IJCNN), 1–9.

Herrera, A., Beck, A., Bell, D., Miller, P., Wu, Q., Yan, W. (2008) Behavior analysis and prediction in image sequences using rough sets. *International Machine Vision and Image Processing Conference* (pp.71-76)

Huang, M., Liu, Z., Liu, T., & Wang, J. (2023). CCDS-YOLO: Multi-Category Synthetic Aperture Radar Image Object Detection Model Based on YOLOv5s. Electronics, 12(16), 3497.

Li, C., Li, L., Geng, Y., Jiang, H., Cheng, M., Zhang, B., Ke, Z., Xu, X., & Chu, X. (2023). YOLOv6 v3.0: A Full-Scale Reloading. https://doi.org/10.48550/ARXIV.2301.05586

Li, F., Zhang, Y., Yan, W., Klette, R. (2016) Adaptive and compressive target tracking based on feature point matching. *International Conference on Pattern Recognition* (ICPR), (pp.2734-2739).

Liang, C., Lu, J., Yan, W. (2022) Human action recognition from digital videos based on deep learning. *ACM ICCCV.*

Lu, J. (2016) Empirical Approaches for Human Behavior Analytics. Master's Thesis. Auckland University of Technology, New Zealand.

Lu, J., Shen, J., Yan, W., Boris, B. (2017) An empirical study for human behaviors analysis. *International Journal of Digital Crime and Forensics* 9 (3), 11-17.

Lu, J., Nguyen, M., Yan, W. (2018) Pedestrian detection using deep learning. IEEE AVSS.

Lu, J., Nguyen, M., Yan, W. (2020) Human behavior recognition using deep learning. *International Conference on Image and Vision Computing New Zealand*.

Lu, J., Nguyen, M., Yan, W. (2020) Comparative evaluations of human behavior recognition using deep learning. *Handbook of Research on Multimedia Cyber Security*, 176-189.

Sujatha, K., Amrutha, K., & Veeranjaneyulu, N. (2023). Enhancing Object Detection with Mask R-CNN: A Deep Learning Perspective. International Conference on Network, Multimedia and Information Technology (NMITCON), 1–6.

Vasanthi, P., & Mohan, L. (2023). Multi-Head-Self-Attention based YOLOv5X-transformer for multi-scale object detection. Multimedia Tools and Applications.

Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Han, K., & Wang, Y. (2023). Gold-YOLO: Efficient Object Detector via Gather-and-Distribute Mechanism.

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7464–7475.

Yan, W. (2019) Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics. Springer Nature.

Yan, W. (2023) Computational Methods for Deep Learning: Theory, Algorithms, and Implementations. Springer Nature.

Yu, Z., Yan, W. (2020) Human action recognition using deep learning methods. *International Conference on Image and Vision Computing New Zealand.*

Zhang, Z., Lu, X., Cao, G., Yang, Y., Jiao, L., & Liu, F. (2021). ViT-YOLO: Transformer-Based YOLO for Object Detection. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2799–2808.

Zhou, L., Wei, S., Cui, Z., & Ding, W. (2019). YOLO-RD: A lightweight object detection network for range doppler radar images. IOP Conference Series: Materials Science and Engineering, 563(4), 042027.

Zhou, H., Nguyen, M., Yan, W. (2023) Computational analysis of table tennis matches from real-time videos using deep learning. PSIVT 2023.

Zhu, Y., Yan, W. (2022) Ski fall detection from digital images using deep learning. *ACM ICCCV.*