Moar: A Swimmer Motion Swimming Style Identification Model using Deep Learning

Atheer Al-Majnoni
Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
atheeraedh2002@gmail.com

Jumana Al-Sahli
Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
jumanamoqbelalsehli@outlook.com

Dana Al-Ahmady
Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
danah_137@hotmail.com

Amani Al-Mutairi
Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
amanialmutere1@gmail.com

Areej Alsini
Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
aosini@uqu.edu.sa

Manal Alharbi
Department of Computer Science and Artificial Intelligence, Umm Al-Qura University, Makkah, Saudi Arabia
mhharbi@uqu.edu.sa (corresponding author)

Received: 17 October 2024 | Revised: 17 November 2024 | Accepted: 21 November 2024

 DOI: https://doi.org/10.48084/etasr.9309

ABSTRACT

Athletes in various sports, such as swimming, are increasingly using motion capture to identify and optimize their movement techniques. However, traditional motion capture systems tend to be expensive and limited. Computer vision-based methods have emerged as alternatives to identify four swimming styles: freestyle, backstroke, breaststroke, and butterfly. However, previous models did not identify flaws in swimmer movement. A significant challenge is the lack of labeled swimming video datasets that indicate these flaws. To overcome this challenge, this study collected and labeled a dataset of swimmer flaws and integrated them with the publicly available dataset SwimXYZ. Then, YOLO models were trained on the generated data. The YOLOv8s model demonstrated an impressive mean average precision (mAP@0.50) of 98% in the detection of swimming style and 95% in the simultaneous detection of swimming style and the identification of incorrect movements. This model can be used in real-time applications to help swimmers evaluate and improve the accuracy of their techniques.

Keywords-human motion capture; artificial intelligence; computer vision; YOLOv8; swimming styles; swimming flaws; incorrect movement

I. INTRODUCTION

The emphasis on sports and youth initiatives to improve quality of life and healthy lifestyles is essential. Promoting aquatic sports and increasing the availability of modern swimming pools are crucial steps in spreading aquatic sports culture. Swimming is a recognized sport throughout the world in organized sports events. In these events, swimmers compete on their skills and speed in various races. Swimming competitions adopt four main swimming styles: freestyle, backstroke, breaststroke, and butterfly. During training, swimmers often struggle to accurately assess their swimming tactics. Many seek professional advice or resort to instructional videos. However, challenges such as limited time, cost, and difficulty in accessing facilities and finding professional coaching hinder the level of physical activity and professionalism.

With the rapid advance of technology, motion capture has been applied in the field of computer vision. Motion capture is a technique to record and convert the movements of objects or people into digital data, which can then be analyzed and manipulated [1], and is used in various fields, such as entertainment [2-4], autonomous driving [5], medical applications [6-9], and sports [10-14]. It is also used to validate computer vision systems [15–18]. The use of motion capture in sports helps to analyze and optimize athlete skills [10, 19-21], injury prevention [22-25], and detection of drowning [26]. Despite traditional human motion capture systems that involve manual capture of images and videos, automatic motion capture systems have greatly improved accuracy and reliability, making them useful in different sports contexts, from analyzing individual techniques to assessing group tactics. However, computer vision-based motion capture systems for swimming face limitations due to the lack of labeled data. This study proposes Moar, a YOLOv8-based deep learning model that is trained to simultaneously identify swimming styles and flaws in them from videos. To train Moar, data for incorrect swimming style movements were collected and integrated with the publicly available SwimXYZ dataset [27].

II. RELATED WORKS

In general, four swimming style forms are used in swimming competitions, such as freestyle, backstroke, breaststroke, and butterfly. These styles encourage swimmers to adhere to unique sets of guidelines and tactics. Freestyle involves alternating arm motions and a flutter kick while positioning the body to face down in the water and breathing to the side. The backstroke style is much like the freestyle, but the swimmers move their bodies side to side to create propulsion using a flutter kick and are allowed to breathe continuously. Breaststroke involves a simultaneous arm motion and a frog-like kick, with a glide phase between strokes. The butterfly style is characterized by an undulating dolphin-like body movement and simultaneous arm motion, and it is considered a challenging swimming style.

Many studies have focused on monitoring and analyzing swimmer movements. Some used electronic devices known as Inertial Measurement Units (IMUs), which measure and report data related to certain forces, angular rates, and occasionally magnetic fields. Accelerometers, gyroscopes, and occasionally magnetometers are common components found in IMUs. They are used when comprehension of motion and orientation is required, especially in the detection of swimming styles [28-30]. Other studies used various sensors for swimming movement tracking such as sonar sensors [31], mobile phones [32], smart watches [33], biosensors and wearable sensors [34], and wrist wearable assistants [35].

In [28], a single waterproof inertial sensor attached to the swimmer's back waist was used to collect 3-axis acceleration and gyro data and detect the start time of the swimming stroke using a Convolutional Neural Network (CNN), achieving a precision of 85% and a recall of 90%. In [29], a deep learning-based Swimming Stroke Recognition (SSR) system was developed using Inertial Measurement Units (IMU) and a hybrid DCNN-BiLSTM model, achieving a balanced accuracy of 96.27% in recognizing four swimming strokes. In [30], four different multiclass classification models were proposed and evaluated, using Long Short-Term Memory (LSTM) networks to capture temporal dependencies in swimming movements, achieving a high classification accuracy of 95% and acceptable loss values. In [31], sonar and a deep CNN were used to identify swimming styles, achieving 93.7% training accuracy. In [32], an aqua-tracker system was developed, achieving a real-time classification accuracy of 95% with a CNN model. In [33], a CNN was used to recognize swimming styles and transitions, achieving an average F1-score of 97%. In [34], a real-time framework was proposed to analyze swimming performance using wearable sensors and biosensors, with a Random Forest (RF) classifier achieving a macro-averaged F1-score of 95%.

Although previous studies have shown impressive results, acquiring these sensors might be a challenge. Thus, several studies focused solely on computer vision using machine learning and deep learning techniques [29, 36-41]. In [36], computer vision was used for pose detection, achieving 67% accuracy in distinguishing between efficient and inefficient pulling poses. In [37], a performance prediction model was developed for young swimmers, using a feedforward neural network that achieved 80% precision. In [38], a CNN-based approach was proposed to automatically detect swimming strokes in continuous video, achieving an F1-score of 92%. In [29], a hybrid Deep CNN and Bidirectional LSTM (DCNNBiLSTM) model was proposed to recognize four swimming strokes with an accuracy of 96%. In [39], bidirectional LSTM was used to classify swimming activities with an F1-score of 96% and high precision in calculating lap times. In [40], a system was proposed that used machine learning and computer vision to analyze freestyle swimming strokes. This system classified strokes as "good" or "bad," offering corrective feedback to help swimmers improve their technique and reduce the risk of injuries, achieving an accuracy of more than 90%.

However, earlier studies did not identify flaws or incorrect tactics in swimmer movement, while the lack of labeled datasets that can identify these errors is a major obstacle.

III. METHODOLOGY

A. Datasets and Data Cleaning

Data on swimming style flaws were collected and integrated with the publicly accessible SwimXYZ dataset [27]. Training models on the integrated data enables effective and simultaneous identification of the style and flaws through the swimmer's stroke. Before integrating the two datasets, each dataset was cleaned, annotated, and augmented.

1) SwimXYZ Dataset [27]

This is a synthetic dataset containing swimming motions and videos, consisting of 11,520 videos, amounting to a total of 3.4 million frames. Each 5-second video, recorded at 60 fps, is annotated in both 2D and 3D using three different formats. The dataset consists of 240 swimming motion sequences represented in SMPL parameters. It includes 60 sequences for each of the four swimming strokes, capturing variations in both movement and body shape. It is important to ensure the quality of the dataset for reliable training and analysis. This was achieved by following several key steps. First, videos showing a swimmer from above (located on the side above the water folder in the SwimXYZ dataset) were removed and excluded because they do not clearly explain the swimmer's movement. This can mislead or distract the model, affecting its predictions, as shown in Figure 1. Moreover, data redundancy was reduced by identifying and removing videos that do not capture the swimmer's movement. Furthermore, the dataset was transformed from WEBM to MP4 format to ensure compatibility with the model. After cleaning, the dataset consisted of 1,110 videos, divided into 395 videos in Butterfly, 288 videos in Backstroke, 288 videos in Breaststroke, and 139 videos in Freestyle. All the videos in the dataset were converted into a frame of six images. As not all images show the swimmer's movement, images in which the swimmer's movement is not clear or the swimmer is not present were removed. This process generated 1,030 individual images that were used to train the model, 339 images in Butterfly, 200 images in Backstroke, 269 images in Breaststroke, and 222 images in Freestyle. This process was essential to ensure that the model had a diverse range of visual data for accurate predictions.

B. SwimMistakes Dataset

Several videos and photos were collected that showcased different types of incorrect swimming movements in Butterfly, Backstroke, and Breaststroke swimming styles. Videos depicting mistakes in freestyle swimming were excluded due to the wide variation in this style among different swimmers, which could potentially mislead or distract the model, thus affecting its predictions. Cleaning datasets enhance data quality and highlight their relevance. Videos that depict a swimmer from above were excluded, as this pose does not provide a clear understanding of the swimmer's movement. This step resulted in a dataset of 236 images divided as follows: 39 images of incorrect movement in Butterfly, 99 images of incorrect movement in Backstroke, and 98 images of incorrect movement in Breaststroke. Each image captures different differences in camera angle, subject appearance, water conditions, lighting, and movement.

C. Data Annotation

This is an important step in the data preparation process to make it understandable and usable for deep learning models. Annotation was implemented using Roboflow, starting by defining the annotation class with the names Butterfly, Backstroke, Breaststroke, Freestyle, and Mistake. This class presents the type of swimming movement along with the mistake within the frame of uploaded videos. The annotation process involves manual frame labeling. Swimmer labels and bounding boxes were assigned to accurately illustrate the swimmer's movement and the place of the mistake. This step ensures that the dataset contains detailed information to be used by the deep learning model to identify the style and flaws in the swimming style.

D. Data Augmentation

Data augmentation is used to improve the quality and diversity of a dataset. Roboflow was used to apply some augmentation strategies, including the ability to accept images horizontally, apply grayscale to 10% of them, blur up to 0.7 pixels, and add noise up to 0.14% of pixels. After these steps, the dataset consisted of 2,720 images, ensuring that it includes many variations and circumstances found in the actual world. Figure 2 shows the total number of labels in each class.

Fig. 1. Sample of side above water frame from the SwimXYZ dataset.

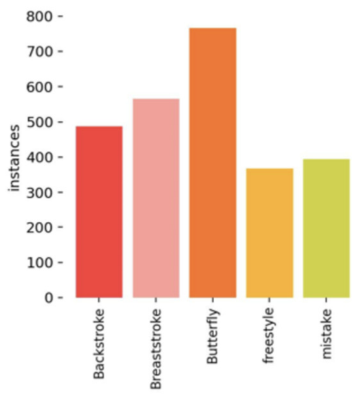

Fig. 2. Number of labels in each class.

E. Data Splitting

To ensure the reliability of the results, the dataset was divided into 80% for training (2,181 images) and 20% for validation (539 images).

F. The Proposed Models and Training Process

The aim was to develop a model that identifies swimming styles and detects common technique flaws. Among the various object detection models that identify objects within images or videos, analyze postures, and categorize movements, the YOLO models (You Only Look Once) are particularly effective and come in various versions [42]. YOLOv5n and YOLOv8n are more complex and high-performance options for large-scale tasks, while YOLOv5s and YOLOv8s are lightweight and optimized for speed on resource-limited devices. Each variant may differ in the architecture of the backbone, the number of layers, and the design of the classifier component to suit specific requirements in terms of object detection accuracy, speed, and computational resources. The backbone of YOLOv5 and YOLOv8 is a modified version of the CSPDarknet53 architecture.

YOLOv5 [43, 44] uses a modified CSP-Darknet53 backbone, which contains 53 convolutional layers. It utilizes Cross Stage Partial (CSP) connections to enhance the flow of information and reduce computational load. The head of YOLOv5 consists of convolutional layers that predict bounding boxes, abjectness scores, and class probabilities.

- The YOLOv5n model has a total of 47-119 layers. The total number of layers can vary depending on the specific configuration.
- The YOLOv5s model has 19 convolutional layers, the total number of layers in the architecture is 157, encompassing various operations that transform the data through the network.

YOLOv8 [43-44] introduces several improvements over previous versions, including more advanced convolutional layers and potentially different structures for the head. The YOLOv8 head is composed of several convolutional layers, followed by a set of fully connected layers. The three heads correspond to the cv2, cv3, and dfl layers in the Detect module.

- YOLOv8n has a total of 225 layers, The number of layers in the C2f (Cross-stage Partial Connections) blocks and bottleneck blocks is determined by the depth multiplier (d) of the model. For the YOLOv8n model, the depth multiplier (d) is 0.33. This means that the number of repeats in the C2f and bottleneck blocks is scaled down by a factor of 0.33 compared to the base model.
- The YOLOv8s model has a total of 225 layers.

G. Training Process

These models were fine-tuned, establishing the following hyperparameters. The number of epochs, each representing a full pass over the entire dataset, was set at 80, and the minimum confidence threshold for detections was set at 0.25. The number of images per batch was set to -1, indicating AutoBatch for YOLOv8n and YOLOv8s, and set to 16 for YOLOv5n and YOLOv5.

H. Evaluation Measures

Several evaluation metrics are used in the context of object detection tasks with bounding boxes and considering IoU thresholds. These metrics calculate the Intersection over the Union (IoU) threshold to determine if a predicted bounding box is True Positive (TP) or False Negative (FN), where TP are correctly predicted instances, and FP are instances incorrectly predicted as positive. This study used the recall and mean Average Precision (mAP). Recall is calculated as the ratio of TP to the total number of relevant instances:

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

Precision is calculated as the ratio of TP to the total number of predicted positive instances.

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

mAP is also a commonly used performance metric to evaluate object detection models. This study used mAP@0.50 and mAP@0.50:0.95. mAP@0.50 measures the average precision of object detection models at a specific IoU threshold of 0.50, ensuring at least 50% overlap between predicted and ground truth bounding boxes. mAP@0.50:0.95 is a more comprehensive evaluation metric than mAP@0.50, since it considers the average precision across IoU thresholds, ranging from 0.50 to 0.95. However, both metrics are widely used in object detection to evaluate model accuracy and robustness. Higher recall and precision values indicate better object detection.

IV. RESULTS AND DISCUSSION

This section presents the results and a comprehensive analysis of the performance of the trained models. The evaluation of the model includes its ability to identify the swimming style and flaws in the style tactics. Table I shows the results of the four YOLO models used in this study in terms of precision (P), recall (R), and mAP at various IoUs. From the results, the following can be observed:

- In general, YOLOv5s and YOLOv8s perform better in both the All and the Mistake classes compared to YOLOv5n and YOLOv8n, respectively. YOLOv5s and YOLOv8s have higher mAP@0.50 values in identifying the Mistake class, indicating better accuracy in detecting instances in this class. YOLOv5s consistently outperforms YOLOv5n, while YOLOv8s tends to have a slight edge over YOLOv8n in both the identification of All and Mistake classes.
- The results show varying levels of performance across different IoU thresholds. For example, the model demonstrates strong performance at an IoU threshold of 0.50, with mAP values ranging from 0.78 to 0.99 in YOLOv5n, 0.82 to 0.99 in YOLOv5s, 0.79 to 0.99 in YOLOv8n and 0.81 to 0.99 in YOLOv8s across different classes. This variability suggests that the model's accuracy and precision levels can fluctuate based on the IoU threshold considered.
- When comparing the values of mAP@0.50:0.95, they range from 0.35 to 0.82 in YOLOv5n, 0.37 to 0.80 in YOLOv5s, 0.36 to 0.81 in YOLOv8n, and 0.36 to 0.84 in YOLOv8s across different classes. This variability suggests that the models exhibit different levels of accuracy and precision in detecting objects, depending on the degree of overlap

between predicted and ground-truth bounding boxes. However, YOLOv8n and YOLOv8s outperform YOLOv5n and YOLOv5s in terms of average mAP @0.50:0.95, achieving 0.66 and 0.68, respectively.

- All four models achieved the highest recall and mAP in the Freestyle and Breaststroke classes.
- YOLOv5s and YOLOv8s show slightly better mean performance compared to YOLOv5n and YOLOv8n, respectively.

TABLE I. COMPARATIVE VALIDATION RESULTS OF FINE-TUNED YOLO MODELS ON THE INTEGRATED DATASET

Class	Ref.	Dataset	Model	Evaluation Measures			
				R	P	mAP@ 0.50	mAP@ 0.50:0.95
Mistake	This study	Swim-Mistakes	YOLOv5n	0.76	-	0.78	0.35
			YOLOv5s	0.74	-	0.82	0.37
			YOLOv8n	0.75	-	0.82	0.38
			YOLOv8s	0.76	-	0.82	0.37
All	[39]	Collected	Bi-LSTM	0.96	0.96	-	-
	This study	SwimXYZ	YOLOv5n	0.91	-	0.94	0.66
			YOLOv5s	0.92	-	0.95	0.66
			YOLOv8n	0.93	-	0.94	0.68
			YOLOv8s	0.92	-	0.95	0.68
Butterfly	[28]	Collected	CNN	0.904	0.855	-	-
	[29]		DCNN-BiLSTM	0.915	0.886	-	-
	[33]		CNN	0.91	1	-	-
	[39]		Bi-LSTM	0.88	0.99	-	-
	This study	SwimXYZ	YOLOv5n	0.95	-	0.96	0.69
			YOLOv5s	0.95	-	0.96	0.71
			YOLOv8n	0.95	-	0.97	0.72
			YOLOv8s	0.96	-	0.97	0.72
Back-stroke	[29]	Collected	DCNN-BiLSTM	.0.98	0.996	-	-
	[33]		CNN	0.98	0.97	-	-
	[39]		Bi-LSTM	0.99	0.98	-	-
	This study	SwimXYZ	YOLOv5n	0.89	-	0.98	0.67
			YOLOv5s	0.91	-	0.98	0.67
			YOLOv8n	0.94	-	0.98	0.7
			YOLOv8s	0.95	-	0.98	0.7
Breast-stroke	[29]	Collected	DCNN-BiLSTM	0.93	0.89	-	-
	[33]		CNN	0.95	1	-	-
	[39]		Bi-LSTM	0.99	0.98	-	-
	This study	SwimXYZ	YOLOv5n	0.96	-	0.99	0.75
			YOLOv5s	0.98	-	0.99	0.74
			YOLOv8n	0.99	-	0.99	0.78
			YOLOv8s	0.99	-	0.99	0.77
Front-crawl	[29]	Collected	DCNN-BiLSTM	0.917	0.94	-	-
	[33]		CNN	0.99	0.99	-	-
	[39]		Bi-LSTM	0.99	0.99	-	-
Freestyle	This study	SwimXYZ	YOLOv5n	1.00	-	0.99	0.82
			YOLOv5s	1.00	-	0.99	0.8
			YOLOv8n	1.00	-	0.99	0.81
			YOLOv8s	1.00	-	0.99	0.84

A. *Analyzing the Performance of YOLOv8s Model*

As YOLOv8s slightly outperformed other models, Figure 3 shows the precision and recall confidence curves for each of the five classes. These curves provide information on the trade-off between classification performance and the confidence level of the predicted results. The precision-confidence curve shows how a classifier's precision varies as the confidence threshold is adjusted, while the recall-confidence curve illustrates the relationship between a classifier's recall and the confidence threshold. The precision and recall confidence curves provide insight into how precise and sensitive the model is individually for each class. The model reaches high precision and recall in detecting the Freestyle class followed by Breaststrokes, Butterfly, Backstroke, and finally the Mistake class. To investigate the over- and under-fitting of the YOLOv8s fine-tuned model, the loss functions were compared in the training and validation sets. Figure 4 shows the training and validation box losses associated with the model predictions of the location and size of the bounding boxes during the training and validation processes. The training and validation class losses focus on the model's ability to classify objects within these boxes on the training and validation. These graphs show that the model does not suffer from overfitting or underfitting, since both the training and validation curves exhibit decreasing trends, while the accuracy metrics demonstrate upward curves.

(a)

(b)

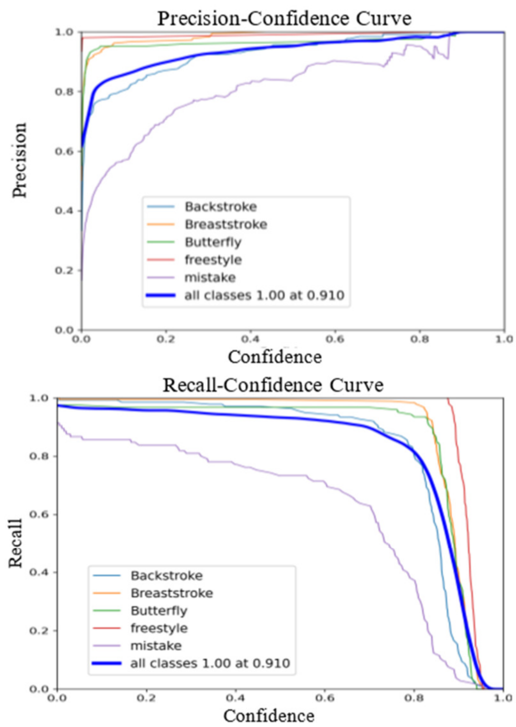

Fig. 3. Precision and recall confidence curves for the YOLOv8s model.

B. *Investigating the Results in a Real-World Scenario*

Tests were carried out to evaluate the precision of the proposed model in detecting swimming styles and incorrect movements in real-world scenarios. The model was subjected to various sizes and colors of 105 images and videos to assess its ability to draw bounding boxes and correctly label the swimmer's movements. The model showed impressive performance in detecting incorrect swimming tactics and movements, drawing bounding boxes around the swimmer's movements with precise labeling, as shown in Figure 5. The tests showed that high-confidence detection with scores greater than 0.7 was consistently accurate. However, low-confidence detection with scores less than 0.4 sometimes resulted in false alarms due to possible FP or inconclusive results.

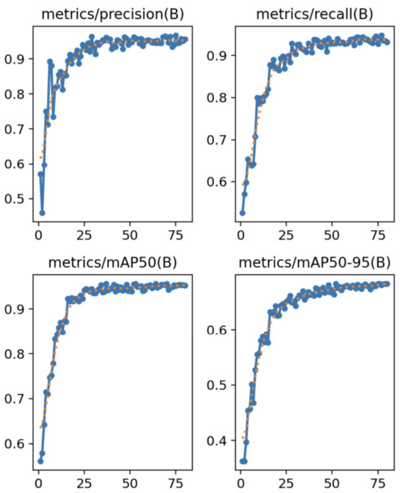

Fig. 4. The train/validation loss and accuracy curves for the YOLOv8s model.

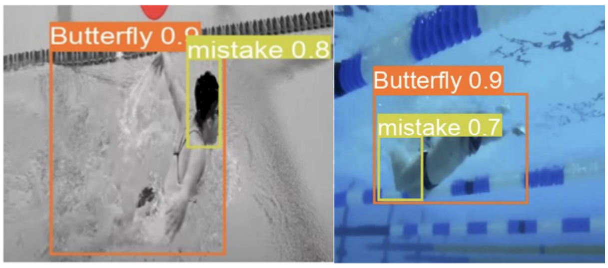

Fig. 5. Testing using the YOLOv8s model.

C. Comparisons with Prior Studies

Table I compared the performance of YOLO with other models such as Bi-LSTM, DCNN-BiLSTM, and CNN, indicating several key points. YOLO models, particularly YOLOv8, consistently show strong performance across various classes and datasets, achieving high recall values and indicating their effectiveness in detecting instances of the target classes. For example, in the SwimXYZ dataset, the YOLOv8 models achieved recall values up to 0.99 for the Breaststroke and Freestyle classes. Additionally, the mAP@0.50:0.95 values for YOLO models were generally high, demonstrating their precision across different IoU thresholds. YOLO models performed well in both the SwimMistakes and SwimXYZ datasets, showcasing their adaptability to different types of data. This versatility makes them suitable for a wide range of applications in object detection and classification tasks.

However, the Bi-LSTM and DCNN-BiLSTM models excelled in a collected dataset, achieving high precision and recall values. For instance, the Bi-LSTM model achieved a recall of 0.99 and a precision of 0.98 for the Backstroke class. However, their performance in the SwimXYZ dataset was not provided, so their generalizability to different datasets remains to be fully assessed. CNN models also performed well in collected datasets, with high recall and precision values, such as a recall of 0.99 and precision of 0.99 for the Front Crawl class. However, their performance in the SwimXYZ dataset was not provided, limiting the ability to compare their generalizability and robustness against YOLO models.

V. CONCLUSION

Traditional human motion capture systems are not practical for large-scale use due to their setup, preprocessing, and calibration requirements. This study used four YOLO models to detect and analyze swimming movements and detect errors. The models were fine-tuned on an integrated dataset of SwimXYZ and Swim-Mistakes. Fine-tuned models show promising results with high mAP scores and good performance across IoU thresholds, indicating robustness in the detection of swimming techniques and incorrect movements. YOLOv5s and YOLOv8s exhibited stronger and more consistent performance on various metrics compared to YOLOv5n and YOLOv8n, especially in the detection of swimming techniques and movements. Models could distinguish between the various classes effectively, with high precision and recall values. The reliability of the model was further strengthened by the high mAP values. The rigorous testing process on the dataset provided robust results, demonstrating the precision of the YOLOv8s model in identifying swimming techniques and incorrect movements.

It is important to note that datasets in the literature rarely include evaluations of movement error detection. This means that while the models are evaluated on their ability to classify and detect various swimming strokes, there is no direct evaluation of their performance in identifying movement errors. This study addressed this gap by incorporating error detection evaluations through the Swim-Mistakes dataset. This work highlights a potential area for future research and dataset development, which could enhance the models' applicability in real-world scenarios where precise movement detection is crucial.

REFERENCES

[1] A. Menache, Understanding Motion Capture for Computer Animation. Elsevier, 2011.

[2] T. Baker, "The History of Motion Capture Within The Entertainment Industry," B.S. Thesis, Metropolia University of Applied Sciences, Helsini, Finland, 2020.

[3] C. Bregler, "Motion Capture Technology for Entertainment [In the Spotlight]," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 160–158, Aug. 2007, https://doi.org/10.1109/MSP.2007.906023.

[4] S. Sharma, S. Verma, M. Kumar, and L. Sharma, "Use of Motion Capture in 3D Animation: Motion Capture Systems, Challenges, and Recent Trends," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, Feb. 2019, pp. 289–294, https://doi.org/10.1109/COMITCon.2019.8862448.

[5] A. A. Alsuwaylimi, R. Alanazi, S. M. Alanazi, S. M. Alenezi, T. Saidani, and R. Ghodhbani, "Improved and Efficient Object Detection Algorithm based on YOLOv5," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14380–14386, Jun. 2024, https://doi.org/10.48084/etasr.7386.

[6] L. Mündermann, S. Corazza, and T. P. Andriacchi, "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications," *Journal of NeuroEngineering and Rehabilitation*, vol. 3, no. 1, Mar. 2006, Art. no. 6, https://doi.org/10.1186/1743-0003-3-6.

[7] I. Takeda, A. Yamada, and H. Onodera, "Artificial Intelligence-Assisted motion capture for medical applications: a comparative study between markerless and passive marker motion capture," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 24, no. 8, pp. 864–873, Jun. 2021, https://doi.org/10.1080/10255842.2020.1856372.

[8] G. Bleser, B. Taetz, M. Miezal, C. A. Christmann, D. Steffen, and K. Regenspurger, "Development of an Inertial Motion Capture System for Clinical Application: Potentials and challenges from the technology and application perspectives," *i-com*, vol. 16, no. 2, pp. 113–129, Aug. 2017, https://doi.org/10.1515/icom-2017-0010.

[9] W. W. T. Lam, Y. M. Tang, and K. N. K. Fong, "A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, May 2023, Art. no. 57, https://doi.org/10.1186/s12984-023-01186-9.

[10] B. Pueo Ortega and J. M. Jiménez Olmedo, "Application of motion capture technology for sport performance analysis," *Retos: nuevas tendencias en educación física, deporte y recreación*, no. 32 (2º semestre), pp. 241–247, 2017.

[11] E. van der Kruk and M. M. Reijne, "Accuracy of human motion capture systems for sport applications; state-of-the-art review," *European Journal of Sport Science*, vol. 18, no. 6, pp. 806–819, Jul. 2018, https://doi.org/10.1080/17461391.2018.1463397.

[12] G. R. D. Bernardina, T. Monnet, H. T. Pinto, R. M. L. de Barros, P. Cerveri, and A. P. Silvatti, "Are Action Sport Cameras Accurate Enough for 3D Motion Analysis? A Comparison With a Commercial Motion Capture System," *Journal of Applied Biomechanics*, vol. 35, no. 1, pp. 80–86, https://doi.org/10.1123/jab.2017-0101.

[13] S. Barris and C. Button, "A Review of Vision-Based Motion Analysis in Sport," *Sports Medicine*, vol. 38, no. 12, pp. 1025–1043, Dec. 2008, https://doi.org/10.2165/00007256-200838120-00006.

[14] S. Noiumkar and S. Tirakoat, "Use of Optical Motion Capture in Sports Science: A Case Study of Golf Swing," in *2013 International Conference on Informatics and Creative Multimedia*, Kuala Lumpur, Malaysia, Sep. 2013, pp. 310–313, https://doi.org/10.1109/ICICM.2013.58.

[15] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System," *Sports Medicine - Open*, vol. 4, no. 1, Jun. 2018, Art. no. 24, https://doi.org/10.1186/s40798-018-0139-y.

[16] R. J. Aughey *et al.*, "Comparison of a computer vision system against three-dimensional motion capture for tracking football movements in a stadium environment," *Sports Engineering*, vol. 25, no. 1, Jan. 2022, Art. no. 2, https://doi.org/10.1007/s12283-021-00365-y.

[17] G. Nagymáté, T. Tuchband, and R. M. Kiss, "A novel validation and calibration method for motion capture systems based on micro-triangulation," *Journal of Biomechanics*, vol. 74, pp. 16–22, Jun. 2018, https://doi.org/10.1016/j.jbiomech.2018.04.009.

[18] Q. Fathima and R. Fathima, "Motion Capture Technology in Animation," *Kristu Jayanti Journal of Computational Sciences (KJCS)*, pp. 44–57, Dec. 2023, https://doi.org/10.59176/kjcs.v3i1.2312.

[19] K. Maduwantha *et al.*, "Accessibility of Motion Capture as a Tool for Sports Performance Enhancement for Beginner and Intermediate Cricket Players," *Sensors*, vol. 24, no. 11, Jan. 2024, Art. no. 3386, https://doi.org/10.3390/s24113386.

[20] M. Jacobsson, J. Willén, and M. Swarén, "A Drone-mounted Depth Camera-based Motion Capture System for Sports Performance Analysis," in *Artificial Intelligence in HCI*, Copenhagen, Denmark, 2023, pp. 489–503, https://doi.org/10.1007/978-3-031-35894-4_36.

[21] M. Talha, "Research on the use of 3D modeling and motion capture technologies for making sports training easier," *Revista de Psicología del Deporte (Journal of Sport Psychology)*, vol. 31, no. 3, pp. 1–10, Oct. 2022.

[22] S. A. Rawashdeh, D. A. Rafeldt, T. L. Uhl, and J. E. Lumpp, "Wearable motion capture unit for shoulder injury prevention," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Cambridge, MA, USA, Jun. 2015, pp. 1–6, https://doi.org/10.1109/BSN.2015.7299417.

[23] A. N. Belova *et al.*, "State-of-the-art possibilities of motion capture technologies in sport injuries research (review)," *Russian Journal of Biomechanics*, vol. 26, no. 2, pp. 74–86, Dec. 2022, https://doi.org/10.15593/RZhBiomeh/2022.2.07.

[24] Z. Ang, "Application of IoT technology based on neural networks in basketball training motion capture and injury prevention," *Preventive Medicine*, vol. 175, Oct. 2023, Art. no. 107660, https://doi.org/10.1016/j.ypmed.2023.107660.

[25] Z. Chen and G. Zhang, "CNN sensor based motion capture system application in basketball training and injury prevention," *Preventive Medicine*, vol. 174, Sep. 2023, Art. no. 107644, https://doi.org/10.1016/j.ypmed.2023.107644.

[26] N. Alharbi, "Exploring Advance Approaches for Drowning Detection: A Review," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 16032–16039, Aug. 2024, https://doi.org/10.48084/etasr.7804.

[27] G. Fiche, V. Sevestre, C. Gonzalez-Barral, S. Leglaive, and R. Séguier, "SwimXYZ: A large-scale dataset of synthetic swimming motions and videos," in *ACM SIGGRAPH Conference on Motion, Interaction and Games*, Rennes, France, Nov. 2023, pp. 1–7, https://doi.org/10.1145/3623264.3624440.

[28] Y. Omae, M. Kobayashi, K. Sakai, T. Akiduki, A. Shionoya, and H. Takahashi, "Detection of swimming stroke start timing by deep learning from an inertial sensor," *ICIC Express Letters, Part B: Applications*, vol. 11, no. 3, pp. 245–251, 2020.

[29] L. Chen and D. Hu, "An effective swimming stroke recognition system utilizing deep learning based on inertial measurement units," *Advanced Robotics*, vol. 37, no. 7, pp. 467–479, Apr. 2023, https://doi.org/10.1080/01691864.2022.2160274.

[30] N. Dirani and G. Pelletier, "Comparaison between the performance of distinct models in classification of swimming types using rolling windows and compression methods," Sep. 2023.

[31] H. Kulhandjian, N. Ramachandran, M. Kulhandjian, and C. D'Amours, "Human Activity Classification in Underwater using Sonar and Deep Learning," in *Proceedings of the International Conference on Underwater Networks & Systems*, Atlanta, GA, USA, Oct. 2019, pp. 1–5, https://doi.org/10.1145/3366486.3366509.

[32] L. C. Powell, "The Evaluation of Recognizing Aquatic Activities Through Wearable Sensors and Machine Learning," Ph.D. dissertation, Texas AM University, USA, 2019.

[33] G. Brunner, D. Melnyk, B. Sigfússon, and R. Wattenhofer, "Swimming style recognition and lap counting using a smartwatch and deep learning," in *Proceedings of the 23rd International Symposium on Wearable Computers*, London, UK, Sep. 2019, pp. 23–31, https://doi.org/10.1145/3341163.3347719.

[34] J. Costa, C. Silva, M. Santos, T. Fernandes, and S. Faria, "Framework for Intelligent Swimming Analytics with Wearable Sensors for Stroke Classification," *Sensors*, vol. 21, no. 15, Jan. 2021, Art. no. 5162, https://doi.org/10.3390/s21155162.

[35] M. Ehab, H. Mohamed, M. Ahmed, M. Hammad, N. ElMasry, and A. Atia, "ISwimCoach: A Smart Coach guiding System for Assisting Swimmers Free Style Strokes: ISwimCoach," in *Companion Publication of the 2020 International Conference on Multimodal Interaction*, Oct. 2020, pp. 265–269, https://doi.org/10.1145/3395035.3425314.

[36] H. Fani, A. Mirlohi, H. Hosseini, and R. Herperst, "Swim Stroke Analytic: Front Crawl Pulling Pose Classification," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, Oct. 2018, pp. 4068–4072, https://doi.org/10.1109/ICIP.2018.8451756.

[37] A. J. Silva *et al.*, "The Use of Neural Network Technology to Model Swimming Performance," *Journal of Sports Science & Medicine*, vol. 6, no. 1, pp. 117–125, Mar. 2007.

[38] B. Victor, Z. He, S. Morgan, and D. Miniutti, "Continuous Video to Simple Signals for Swimming Stroke Detection with Convolutional

Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 122–131, https://doi.org/10.1109/CVPRW.2017.21.

[39] E. Delhaye, A. Bouvet, G. Nicolas, J. P. Vilas-Boas, B. Bideau, and N. Bideau, "Automatic Swimming Activity Recognition and Lap Time Assessment Based on a Single IMU: A Deep Learning Approach," *Sensors*, vol. 22, no. 15, Jan. 2022, Art. no. 5786, https://doi.org/10.3390/s22155786.

[40] A. Jayawardene and P. Kalansooriya, "Swimming Stroke Analysis and Feedback System using Machine Learning," presented at the 1st International Conference on Advanced Computing Technologies (ICACT 2024), Oct. 2024.

[41] A. Hall *et al.*, "The detection, tracking, and temporal action localisation of swimmers for automated analysis," *Neural Computing and Applications*, vol. 33, no. 12, pp. 7205–7223, Jun. 2021, https://doi.org/10.1007/s00521-020-05485-3.

[42] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics YOLO." Jan. 2023, [Online]. Available: https://github.com/ultralytics/ultralytics.

[43] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, https://doi.org/10.3390/make5040083.

[44] Ultralytics, "Comprehensive Guide to Ultralytics YOLOv5." https://docs.ultralytics.com/YOLOv5.