

Dronaquatics: Real-time Swimming Analytics Using Drone Captured Imagery

Thu Tran¹ Harold Abraham Joseph¹ Kichang Lee² Kenny Tsu Wei Choo³
Dong Ma¹ Shaohui Foong³ Thivya Kandappu¹ Jeonggil Ko² Rajesh Balan¹

¹ Singapore Management University, Singapore, ² Yonsei University
³ Singapore University of Technology and Design

Abstract

Accurate swimming performance monitoring has traditionally relied on wearable sensors, which can disrupt natural technique and are impractical in competitive settings. In this paper, we present a fully vision-based system for automatic swimmer analysis using overhead drone footage, removing the need for any wearable device or underwater equipment. By fine-tuning pose estimation models for aerial aquatic conditions, our approach robustly extracts full-body swimmer skeletons even under challenging scenarios such as splashes and partial occlusions. From these poses, we classify swimming strokes, compute instantaneous speed, estimate lap times, and count individual strokes. Unlike existing methods, our system provides scalable, unobtrusive, and infrastructure-free tracking. Evaluated on real-world drone-captured swimming competition data, our method achieves a median speed estimation error below 4% (under 0.05 m/s), a median lap time error of just 0.03s, and stroke count errors typically under one stroke per lap.

1. Introduction

Precise, continuous performance analytics are critical in elite swimming, where races are often decided by fractions of a second [24]. Minor improvements in technique, stroke efficiency, and pacing require detailed biomechanical feedback on timing, rhythm, and posture provided by coaches [16]. Traditional coaching, based on visual assessment and selective video review, provides valuable insights but is constrained by subjectivity, human perceptual limits, and the time demands of manual analysis [2]. This makes real-time tracking for multiple athletes with high accuracy impossible. As coaching methodologies evolve toward data-driven training optimization, there is an urgent need for systems that deliver high-fidelity, real-time analytics in a manner that does not interfere with natural swimming behavior. These methodologies should enhance,

rather than disrupt, coaching practice. They must be practical for real-world environments by supporting the simultaneous monitoring of multiple athletes without interrupting training flow or requiring time-intensive, individualized setups.

Various technologies have been explored for swimmer monitoring, each offering partial solutions but presenting critical limitations for real-world application. Wearable sensors, with inertial measurement units (IMUs), deployed to capture swimmer motion characteristics like stroke counts, turn detection, and velocity estimation [5, 6, 18]. While these systems provide quantitative data, they are invasive by nature and introduce added drag, which may interfere with the swimmer's natural movement patterns, and can be perceived as uncomfortable or distracting during high-intensity training. Moreover, data from wearables often requires complex post-processing and is commonly limited to point measurements without providing holistic visual feedback about body posture or stroke form [15].

In addition, whether mounted on the pool or underwater, vision-based approaches (using stationary cameras) have also been widely investigated. Underwater camera setups can yield detailed information about body position and propulsion mechanics; however, they require infrastructure installation, precise calibration, and are restricted to narrow fields of view [7, 25]. Tracking swimmers over longer distances is challenging, and occlusions caused by splashing, water turbulence, and lane lines are common [22]. Poolside camera systems offer easier deployment but often suffer from severe perspective distortion, making accurate pose estimation and velocity calculation difficult unless the swimmer remains within a narrow tracking corridor.

In this paper, we propose *Dronaquatics*, a drone-based swimmer monitoring framework explicitly tailored for elite swimming coaching applications. Our key idea is to leverage lightweight aerial drones to capture continuous overhead video of swimmers during training, providing a mobile and adaptive vantage point that overcomes many of the

inherent limitations of previous approaches. By employing aerial drones to capture video from an overhead perspective, we overcome the field-of-view and occlusion limitations of stationary systems and the biomechanical disruption caused by wearable devices. Our system integrates fine-tuned deep learning pipelines for robust swimmer pose estimation, temporal sequence modeling for stroke classification and stroke count estimation, and optical flow-based velocity tracking. Importantly, all analytics are derived automatically (with velocity in real-time; stroke count and lap time every lap), from raw drone footage without requiring manual video segmentation or swimmer instrumentation, enabling continuous, scalable, and immediate feedback during regular training sessions. Moreover, we explicitly address the visual challenges posed by aquatic environments, enhancing pose robustness against water distortions.

This work makes the following key contributions:

- We developed a set of algorithms to accurately extract stroke type, instantaneous speed, stroke count and lap time, from the drone-captured imagery.
- *Dronaquatics*, a complete end-to-end solution for swimming analytics was deployed and tested with 14 videos (each video is a different race) collected from two different competitive swimming and one training events, at an indoor and an outdoor pool.

2. Related Work

Recent advancements in swimmer performance monitoring have focused on vision-based systems and wearable inertial sensors (IMUs). These technologies aim to reduce coaching workload while enabling detailed, scalable analysis of biomechanics. Vision-based methods provide non-invasive, real-time insights from video, while IMUs offer high-frequency, personalized data across varied conditions. However, vision systems can struggle with occlusion, lighting, and setup complexity, while IMUs may be intrusive, affect technique, and are less practical in competition.

2.1. Video-Based Pose Estimation in Swimming

Some studies have explored camera-based methods for swimming analysis. Driscoll et al. [7] used a fixed pool-side camera to detect stroke rates for all four strokes, while Woinoski and Bajić [25] extracted stroke rates from overhead race footage. Drone-based approaches have also emerged: for example, Tran et al. [22] tracked swimmers' movements, limb angles, stroke times, and speeds with good accuracy (0.3s error for stroke duration, 0.35 m/s for speed), but their evaluation was limited to post-hoc analysis rather than live feedback. Similarly, Woinoski et al. [26] demonstrated a YOLO-based swimmer analytics pipeline, highlighting the dependence on carefully annotated datasets and the challenges of generalizing across pools and angles. Commercial systems such as ezML's Swim AI [8]

and Phlex Swim Coach [19] provide useful stroke metrics or training plans, often combining wearable sensors with AI, but they do not provide real-time comprehensive measures such as instantaneous speed, cycle counts, or stroke phases. Other research has also focused on swimmer tracking or pose estimation in controlled or underwater settings [27]. In contrast, our system combines the flexibility of drones with real-time processing to deliver a broader set of metrics that can be used by coaches for immediate feedback and not demonstrated by prior drone-based or fixed-camera platforms.

2.2. Wearable Motion Analysis in Swimming

In previous swimming analytics research, IMU sensors are attached to the body, which can be intrusive and are usually not allowed in competitions. For example, Delhay et al. [6] used a single IMU on the sacrum with a deep learning model to classify swim activities and estimate lap times precisely. Costa et al. [5] built a framework using inertial and biosensors to detect strokes and turns, giving real-time feedback. While these methods are accurate, they depend on body-worn devices, which limits their use in real races.

Magalhaes et al. [15] reviewed inertial sensors for swimming, noting their value for continuous monitoring but also challenges like sensor placement and lack of standards. Morais et al. [18] found that wearables can be useful for real-time feedback, but accuracy varies across devices and studies. Mooney et al. [17] reviewed video-based methods, highlighting that video can provide rich qualitative and quantitative insights. However, they also noted issues like camera placement and complex analysis steps, which can make video systems hard to use widely.

3. Swim Data Collection and Processing

3.1. Data Characteristics

We collected video footage of national-level swimmers during their training sessions and publicly accessible competitive events between December 2023 and September 2025, at an indoor and an outdoor pool. These videos were collected on behalf of the swimming association responsible for the swimmers, and the data remains their property for their review of their athletes' performance. The athletes have also given their consent to the sports association to take photos and videos of their activities. The sports association kindly shared this institutional data with us to conduct research that would benefit their programs. The athletes were not identifiable in the footage and the only demographic information associated with the video footage was age and gender. While no IRB approval was required, we nevertheless handled the data in accordance with established best practices for data privacy and research ethics. Relevant permission was obtained from facility owners, competition organizers

and regulatory authorities for filming and drone operations.

These videos were recorded using a human-operator flying a DJI Mavic 3 or DJI Mini 4 Pro drone at 4K resolution at 30 and 60 fps over the swimmer of interest at heights between 8-10 meters. The videos are then resized to 900×600 and 30 fps for more intensive processing and analysis.

We captured 100 videos, ranging from 4.5 to 638 seconds, totalling 264 minutes (see Table 1). Our videos cover various swimming strokes and distances. It covers backstroke, breaststroke, butterfly, freestyle, and medley (see Table 2b)—with distances between 50-200 m (see Table 2a). The swimmers in the videos comprised 33 females and 60 males (see Table 2c), and were subdivided into three age categories: 13-14, 15-17, and 18+, of which seven swimmers belonged to an unknown age category. As our focus was on analytics and not on identifying the swimmers, swimmers in the videos may not be unique (i.e., a swimmer doing backstroke could also be doing the butterfly in a different video).

	No. of Videos	No. of Frames	Time (s)
Train	86	1937	11843.8
Test	14	695	4173.7
Total	100	2632	16017.5

Table 1. Overview of the dataset used for pose estimation.

3.2. Data Annotation

We extracted one frame from each video every 6 seconds, resulting in frames containing between one and five swimmers and at least one lane divider. We also divided the data into train-test splits (see Table 1), using the competition data as the test data, since this would provide the most challenging data to detect, as swimmers would be performing at their best and cause a significant amount of visual artifacts. Our annotators labelled body landmarks for all swimmers present in each frame using the 17-keypoint skeleton format defined by MSCOCO [13]. Representative samples from the dataset are illustrated in Figure 1 in different pools, lighting conditions, lane dividers and swimmers.

In our dataset, all keypoints are labelled as "visible", despite some being occluded by splashes or movements. This forces the model to learn and estimate the keypoints' positions even under uncertainty.

4. System Overview

As illustrated in Figure 2, Dronaquatics, comprises three components: (1) a pose estimation module, (2) a stroke classification module, and (3) a speed estimation module.

4.1. Swimmer Pose Estimation

The YOLO-Pose model, trained on large human keypoint datasets like COCO [13], works well for general pose tasks.

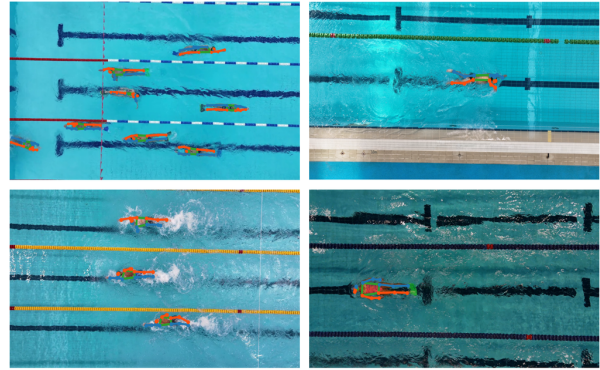


Figure 1. Samples of the collected dataset and their annotations in different pools, lighting conditions, lane dividers and swimmers.

But estimating a swimmer's pose in water is very challenging due to glare, reflections, submersion, and motion blur. When we apply the pre-trained YOLO-Pose [23] to our swimming frames, it often fails to detect key joints well.

To solve this, we fine-tuned YOLO-Pose on our swimming images. We kept the backbone and detection heads and trained the whole model with a small learning rate of 0.01 using the standard OKS loss [20]. We used 1937 frames from real competitions covering different strokes and swimmers. After fine-tuning, the model can reliably detect the swimmer's full pose, even in tough scenes.

4.2. Stroke Classification

Taking the pose estimation model further, to classify swimming stroke types from drone footage, we design a lightweight temporal model that processes sequences of 2D skeletons extracted by the pose estimation module. To account for movements outside of these standard strokes (such as transitional gestures, drifting, or preparatory motions), we include an additional class labeled as "other," capturing any activity that does not correspond to a canonical stroke.

4.2.1. Training Data

The model is trained on the subset of the dataset used to fine-tune the pose estimation model. Each video segment in the dataset is annotated with one of four predefined stroke types: freestyle, backstroke, breaststroke, and butterfly. These labels represent distinct and representative strokes commonly observed in competitive swimming [4]. There are 81,391 samples for training and 27,034 samples for testing.

Our model takes as input a 3-second sequence of 2D keypoints, representing the swimmer's pose trajectory over time. To ensure robustness across different swimmer orientations and camera viewpoints, we apply a series of normalization and data augmentation steps prior to classification:

- **Centering:** Each skeleton is centered by translating the nose keypoint to the origin (0, 0).
- **Scaling:** The skeleton is normalized by the shoulder

(a) Videos per stroke type by distance.

Dist. (m)	50	100	200	Total
Back stroke	7	7	6	20
Breast stroke	10	10	3	23
Butterfly	2	5	7	14
Freestyle	4	26	9	39
Medley	0	0	4	4
Total	23	48	29	100

(b) Videos per stroke type by age group.

Age (years)	13-14	15-17	18+	Total
Back stroke	4	8	4	16
Breast stroke	6	6	11	23
Butterfly	1	4	6	11
Freestyle	8	15	16	39
Medley	0	0	4	4
Total	20	34	39	93

(c) Videos by sex and age group.

Age (years)	13-14	15-17	18+	Total
Female	8	7	18	33
Male	12	27	21	60
Total	20	34	39	93

Table 2. Summary statistics of the annotated swimming video dataset. Age metadata is available for 93 out of 100 videos.

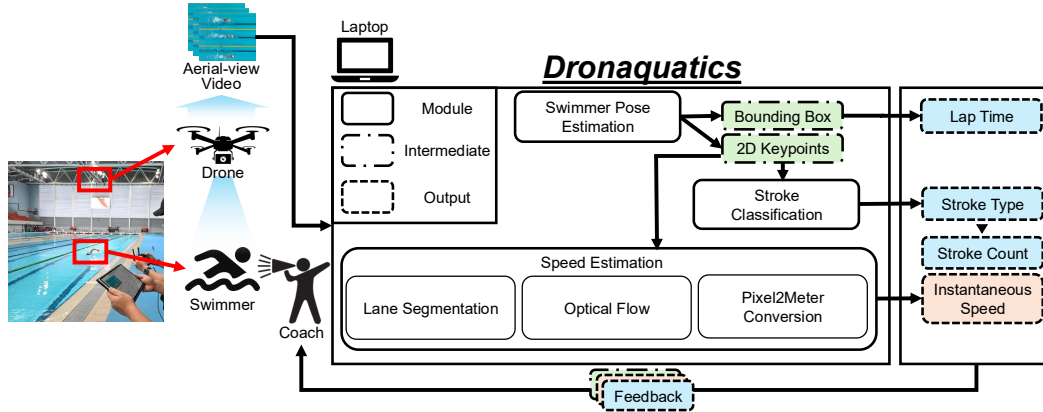


Figure 2. Overall workflow of Dronaletics

width, defined as the Euclidean distance between the left and right shoulder keypoints.

- **Rotation Augmentation:** To enhance generalization to varying swimmer alignments and overhead perspectives, we apply random rotations of 90° , 180° , and 270° during training.

4.2.2. Model

After preprocessing, the normalized keypoint sequences are fed into a Recurrent Neural Network (RNN)[21] to model how body movements change over time[12]. The RNN is well-suited to capture rhythmic stroke patterns and joint dynamics, allowing accurate stroke classification for different swimmers and styles.

The model uses a sequence of 90 frames as input. Each frame has a flattened vector of 17 keypoints with (x, y) and confidence, giving 51 values per frame. As shown in Figure 3, the input goes into a 3-layer Gated Recurrent Unit (GRU) [3] with a hidden size of 32. The GRU starts with zeros as hidden states and outputs hidden states for all time steps, but only the final step’s output is used. This output goes through Layer Norm, ReLU, and dropout (rate 0.3) before a fully connected layer maps it to 4 class logits.

All GRU and linear weights use Xavier initialization [10] for stable training. The model is trained with Cross Entropy Loss using Adam (learning rate 0.001) and a scheduler that lowers the rate after 5 epochs. Training runs for 20 epochs, and performance is measured by classification accuracy.

This stroke classification module is vital for the analysis pipeline, as it provides input for stroke count metric.

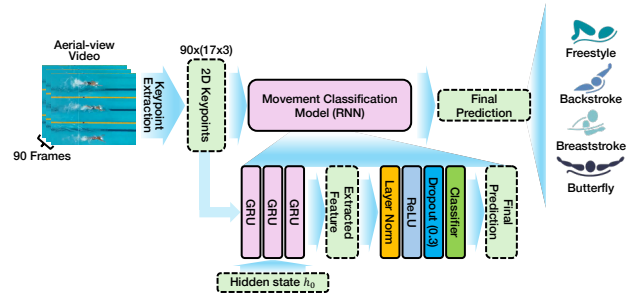


Figure 3. Architecture of RNN used in the stroke classification

4.3. Instantaneous Speed Calculation

To measure performance, we define swimming speed as the head’s velocity along the lane’s length. Using the standard lane width (2.5 meters), we convert pixel movement in the video to real-world distance through camera calibration.

A key challenge with drone video is motion artifacts from water, swimmer movement, and drone drift. To handle this, we use lane dividers as stable references to correct for camera motion. Lane dividers work well because (1) their movement mostly comes from camera shake and small water waves, and (2) their colored, textured surfaces are easy to track with optical flow. The full speed estimation process is shown in Figure 4.

4.3.1. Lane Segmentation

Lane detection and its segmentation are essential in supporting camera motion compensation and head projection. For this purpose, we train a YOLOv11-based lane segmentation

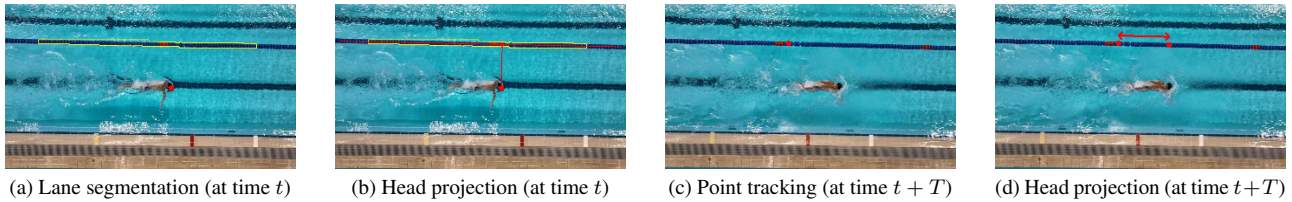


Figure 4. Illustration of instantaneous speed calculation steps. (a) Lane segmentation and swimmer’s head coordinates are obtained from the preceding models. (b) The swimmer’s head position is projected orthogonally onto the lane divider at time t . (c) This projected point on the lane divider is tracked from time t to $t + T$ (d) At time $t + T$, the swimmer’s head is again projected onto the lane divider. The instantaneous swimming speed is computed based on the traveled distance between these two projected points. In this paper, T is 0.5s.

Precision	Recall	mAP50	mAP50-95
0.902	0.907	0.907	0.596

Table 3. Lane divider segmentation performance

model [11]. The training dataset is drawn from our swimming video corpus (Section 3) and comprises 205 training images and 54 validation images, containing one to four lane dividers per frame in a variety of colors (e.g., blue, green, red). The model is trained for 100 epochs using standard data augmentation techniques. Performance metrics of the lane segmentation model are shown in Table 3.

4.3.2. Optical Flow for Camera Motion Compensation

To correct for camera drift, we use the Lucas-Kanade (LK) optical flow method [14], which is fast because it tracks sparse keypoints instead of the full image. However, when used on water surfaces, LK flow can be distorted by ripples and reflections.

To solve this, we track many points along the segmented lane dividers, assuming most points show true camera motion. We remove outlier vectors caused by water disturbances and average the remaining ones to get a stable motion vector. This helps align frames over time and keeps head tracking accurate. Figure 5 shows how lane-based tracking works better than simple methods.

4.3.3. Pixel-to-Meter Conversion

To convert pixel movement to real-world distance, we estimate the pixel-to-meter ratio using the known lane width. Lane dividers are sorted vertically, and the pixel gap between two adjacent dividers (with a swimmer between them) is measured. Since the real lane width is 2.5m [1, 9], this gives a scale factor for the vertical direction.

However, pixel density can differ between width and height because of the camera sensor’s aspect ratio and resolution. To adjust for this, we calculate pixel density for each axis:

$$\text{WidthPixelDensity} = \frac{900}{17.3/1000} \approx 51023.12$$

$$\text{HeightPixelDensity} = \frac{600}{13.3/1000} \approx 45112.78$$

Using a 4/3 sensor (17.3 mm \times 13.3 mm) and a 900 \times 600 image, the vertical-to-horizontal scaling ratio is about 0.8672. We apply this factor to ensure accurate conversion of displacement in both directions to meters.

5. Evaluation

5.1. Pose Estimation

By fine-tuning the raw YOLO-Pose model on swimming-specific data, Dronaquatics greatly improves pose estimation under aquatic conditions. To measure this, we report four standard metrics: Precision, Recall, mAP50 and mAP50-95. Table 4b shows the average results across all test frames. Dronaquatics achieves strong scores on all metrics, showing it is robust and reliable for swimmer pose estimation in challenging conditions.

(a) YOLO-Pose pretrained on COCO			
Precision	Recall	mAP50	mAP50-95
0.0808	0.0275	0.00539	0.000863
(b) Finetuned Model			
Precision	Recall	mAP50	mAP50-95
0.834	0.783	0.788	0.414

Table 4. Pose estimation performance comparison

5.2. Stroke Detection

Dronaquatics achieves strong performance, with an average precision of 0.92 and recall of 0.93 across the four stroke types and others. The confusion matrix, shown in Figure 6, illustrates the classification performance for each individual stroke, with backstroke exhibits the highest classification accuracy compared to the other 3 strokes.

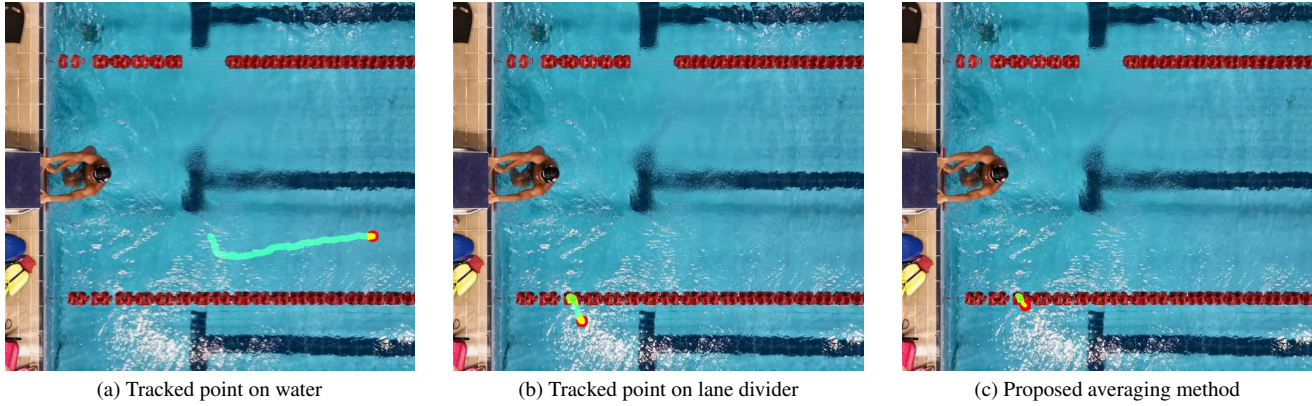


Figure 5. Comparison of LK tracking stability across different regions. The red point is the current position at time t . The trace represents the trajectory of the point from time $t - 5$ to t .

	others	freestyle	backstroke	breaststroke	butterfly
Actual others	5505	163	428	591	217
Actual freestyle	123	4455	72	0	0
Actual backstroke	79	26	6495	0	60
Actual breaststroke	64	0	10	4331	5
Actual butterfly	224	0	4	0	4182
Predicted	others	freestyle	backstroke	breaststroke	butterfly

Figure 6. Confusion matrix of stroke detection.

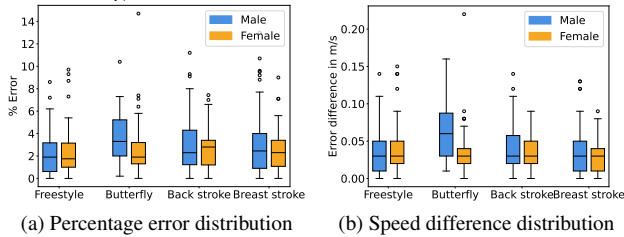


Figure 7. Speed estimation error across stroke types and gender.

5.3. Swimming Velocity Estimation

We evaluate swimmer speed estimation using our vision-based pipeline, which tracks the swimmer’s motion in drone video and converts pixel displacement to real-world speed. For ground truth, we manually track the swimmer’s head across frames and calculate its displacement projected on the lane divider over time, then convert the number of the divider’s buoys to meters.

Figure 7a shows the percentage error relative to ground truth. Our method achieves a median error below 4% for all stroke types and genders. Freestyle, backstroke, and breast-

stroke have tighter errors due to stable head movement and less visual blockage, while butterfly shows higher variance because of vertical head motion and splashes.

Figure 7b shows the absolute error in m/s. For all groups, median deviations are under 0.05 m/s, showing the method is accurate enough for training feedback. The consistent results across genders also show that the system generalizes well to different body shapes and stroke styles.

5.4. Lap time Estimation

We estimate lap times by analyzing how the swimmer interacts with the pool wall in overhead drone video. The swimmer’s bounding box comes from pose keypoints, while the poolside panel is detected using color filtering, assuming it stays yellowish. To improve accuracy, we limit the panel to no more than 50% of the screen and require it to be roughly perpendicular to the lane lines—conditions that hold in most pools.

A lap is marked complete when the swimmer’s bounding box overlaps the panel or comes within 30 pixels. If contact lasts over several frames, it’s marked as a stop; otherwise, it’s a turn. These points split the video into laps and help calculate lap durations using the frame rate.

Getting precise lap times depends on detecting the exact start and end of each lap. Figure 8 compares errors for two methods: (i) a direction-based method that finds lap boundaries by detecting trajectory changes, and (ii) our vision-based method, which uses stroke state changes from pose data. Results show our vision-based method has lower start and end time errors.

Figure 9 breaks this down by stroke type. Again, the vision-based method performs better for most strokes. The only exception is freestyle, where the direction-based method does slightly better, likely because freestyle has a stable body orientation that suits trajectory heuristics better than discrete stroke phases. Table 5 summarizes the resulting error metrics.

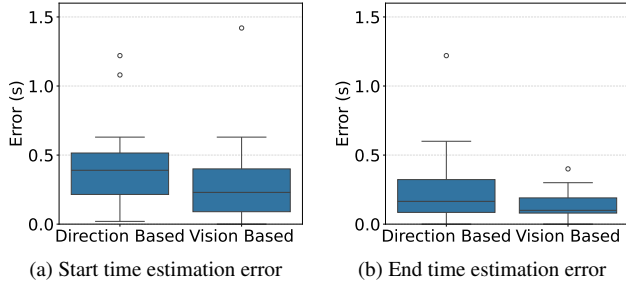


Figure 8. Start and end time estimation error distributions for direction-based and vision-based algorithms.

Stroke	Male Median Error (s)	Female Median Error (s)
Freestyle	0.03	0.03
Butterfly	0.06	0.03
Backstroke	0.03	0.03
Breaststroke	0.03	0.03

Table 5. Error in lap time estimation by gender and stroke type.

5.5. Stroke Count Estimation

Stroke detection is performed by analyzing the temporal variation in wrist-to-head distances across frames, leveraging periodicity in arm motion. For strokes characterized by alternating arm movements, such as freestyle and backstroke, stroke events are detected independently for the left and right wrists. In contrast, for strokes characterized by bilateral symmetry, such as butterfly and breaststroke, as well as for underwater gliding phases, the stroke detection process leverages the averaged wrist-to-head distance signal from both arms. This averaging strategy accounts for the synchronized nature of these strokes, where both arms move synchronously and generate nearly identical kinematic signatures.

To mitigate the impact of noise and minor fluctuations introduced by pose estimation jitter, we apply a Savitzky-Golay smoothing filter to each distance signal prior to peak analysis. The smoothed signals are then subjected to peak detection, with adaptive thresholds for peak height and prominence, which are tuned according to the expected range and periodicity of motion within each segment. This adaptive approach improves the system’s sensitivity to genuine stroke cycles while reducing susceptibility to spurious peaks caused by noise or transient errors in pose tracking. Stroke counts are aggregated over individual lap segments, as determined by the lap segmentation module. This enables lap-wise reporting of stroke counts, which can be further used to derive metrics such as stroke rate and efficiency.

As a baseline for stroke count estimation, we implemented a Fast Fourier Transform (FFT)-based method using the same input as our primary approach—wrist-to-head distances from both arms. Instead of temporal filtering and

peak detection, this method estimates stroke frequency by identifying the dominant frequency component in the signal, under the assumption that typical stroke rates lie between 0.2 and 1.2 Hz. To minimize spectral leakage and improve frequency resolution, a Hann window is applied during preprocessing. The accuracy of the stroke counting pipeline was evaluated against manually annotated ground truth counts on a per-lap basis. Figure 10a shows the FFT-based error, while Figure 10b shows the vision-based error. The peak detection approach consistently outperforms FFT, showing considerably lower median errors and lower error dispersion across all stroke styles. Notably, freestyle and butterfly benefit the most, with error reductions of several strokes per lap, highlighting the peak detection method’s robustness to noise and non-sinusoidal motion patterns.

5.6. Real-time Performance

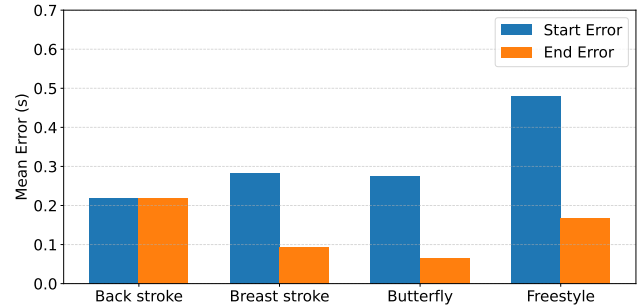
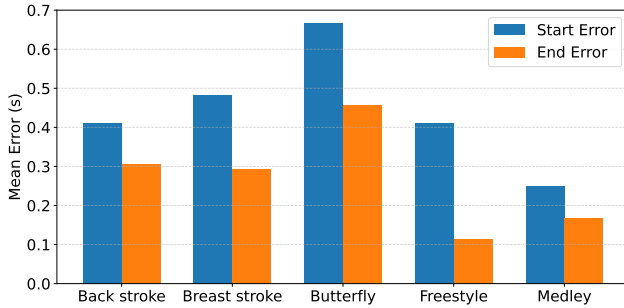
We tested the system on a NVIDIA GeForce RTX 4090 Laptop GPU with batch size 1. It runs at 30 FPS. Each frame takes about 4.5 - 10 ms for pose inference, 0.47 ms for lane divider segmentation, 1 ms for stroke classification and 1.5 ms for optical flow ($\approx 7.47 - 12.97$ ms total), which is fast enough for real-time tracking and speed calculation. The segmentation model can achieve that speed because it is called every 0.5 second (15 frames), only when the head needs to be projected onto the lane divider.

6. Discussion

In this work, we showed that a drone-based system can successfully monitor swimmers and provide useful feedback on their performance. One of the biggest strengths of our system is that it can automatically collect detailed movement data without needing swimmers to wear any extra equipment. It also reduces the need for coaches to manually analyze videos, making it easier to track progress over time. A key advantage over fixed-camera setups is flexibility: drones can cover any lane, adjust viewpoints dynamically, and be deployed quickly without the need for permanent installations. This avoids occlusions and coverage gaps common in static cameras, making the system suitable for different training environments and continuous monitoring. The drones used are off-the-shelf commercial models operated by certified drone pilots, adhere to safety standards, and have been rigorously tested for stability in commercial use. Battery life lasts up to two hours, and hot-swapping batteries takes under a minute, minimizing disruption. While safety precautions remain necessary, our findings suggest that the benefits in adaptability and data capture outweigh the constraints compared to static setups.

6.1. Limitation: Small Test Set Size

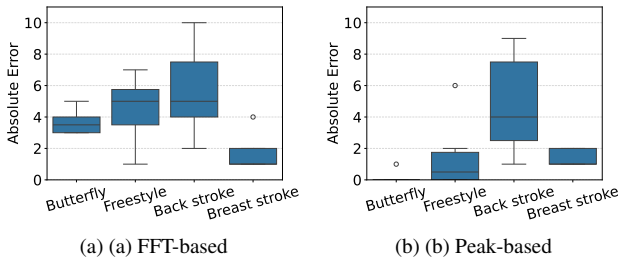
The test set includes only 14 videos, which is a limitation in terms of scale. However, the set was intentionally selected



(a) Estimation error for direction-based algorithm

(b) Estimation error for vision-based algorithm

Figure 9. Direction-based and vision-based mean start and end errors by stroke type.



(a) (a) FFT-based

(b) (b) Peak-based

Figure 10. Stroke count error distributions across stroke types.

to be diverse: recordings were made on different days, at different times with different lighting conditions, in both indoor and outdoor pools, and with athletes of varying ages, skin tones, genders, and heights. This diversity helps ensure that, despite its size, the test set captures a broad range of real-world conditions. Future work should incorporate larger-scale test sets to further strengthen the generalizability of the results.

6.2. Limitation: Multi-Swimmer Scenarios

Currently, although it is possible to extract the analyses of all the swimmers at a time, the system is focusing on tracking a single swimmer. In training environments where multiple swimmers are present, some individuals may move out of the camera’s field of view, making consistent tracking difficult. This restricts the applicability of the system in real-world practice settings where coaches often supervise several athletes simultaneously. Addressing this limitation, through improvements in drone coordination, tracking algorithms, or camera coverage, is essential for scaling the system to team-based coaching contexts.

6.3. Toward Real-Time Feedback in Training

We aim to integrate real-time analytics into live training settings. This transition will involve close collaboration with coaches and swimmers to understand how feedback can be presented in real-time without disrupting practice, and how it can most effectively improve technique and performance. Realizing such a vision will require optimizing the process-

ing pipeline for low-latency pose estimation and trajectory analysis onboard the drone or via edge computing. Investigating lightweight models that balance accuracy and inference speed is crucial for deployment in dynamic training environments.

Further, we plan to enhance the system with additional metrics that can provide deeper insight into swimmer performance, such as:

- Evaluating bilateral symmetry of arm movements
- Assessing stroke-to-stroke stability
- Measuring lateral deviation during laps
- Enabling tracking of multiple swimmers concurrently

7. Conclusion

In conclusion, Dronaquatics demonstrates that accurate, large-scale swimmer performance analysis is achievable using a fully vision-based system with overhead drone footage, removing the need for intrusive wearables or underwater sensors. By fine-tuning pose estimation models for aquatic environments, Dronaquatics robustly extracts swimmer skeletons under challenging conditions and enables reliable stroke classification, speed estimation with median errors below 4% (under 0.05 m/s), lap time estimation with a median error of just 0.03 s, and stroke counting with errors typically under one stroke per lap. These results show that Dronaquatics provides a practical, unobtrusive, and scalable solution for real-world swimming analytics, paving the way for natural, competition-ready athlete monitoring and feedback.

References

- [1] World Aquatics. Swimming pool certification olympic games and world championships. <https://shorturl.at/boylR>, 2024. 5
- [2] Victoria Brackley, Sian Barris, Elaine Tor, and Damian Farrow. Coaches’ perspective towards skill acquisition in swimming: What practice approaches are typically applied in training? *Journal of sports sciences*, 38(22):2532–2542, 2020. 1

- [3] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014. 4
- [4] Cecil M Colwin. *Breakthrough swimming*. Human kinetics, 2002. 3
- [5] Joana Costa, Catarina Silva, Miguel Santos, Telmo Fernandes, and Sérgio Faria. Framework for intelligent swimming analytics with wearable sensors for stroke classification. *Sensors*, 21(15):5162, 2021. 1, 2
- [6] Erwan Delhaye, Antoine Bouvet, Guillaume Nicolas, João Paulo Vilas-Boas, Benoît Bideau, and Nicolas Bideau. Automatic swimming activity recognition and lap time assessment based on a single imu: a deep learning approach. *Sensors*, 22(15):5786, 2022. 1, 2
- [7] Heather Driscoll, Chris Hudson, Marcus Dunn, and John Kelley. Image based stroke-rate detection system for swim race analysis. 2(6):286, 2018. 1, 2
- [8] ezML. Swim ai: Real-time swimming stroke rate counting and auto split tracking. <https://ezml.io/blog/ai-swimming-stroke-rate-counting-tracking>, 2023. Accessed April 2025. 2
- [9] FINA. FINA FACILITIES RULES. <https://shorturl.at/nxYO9>, 2021. 5
- [10] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010. 4
- [11] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024. 5
- [12] Kichang Lee, Jaeho Jin, JaeYeon Park, Songkuk Kim, and JeongGil Ko. Tazza: Shuffling neural network parameters for secure and private federated learning. *arXiv preprint arXiv:2412.07454*, 2024. 4
- [13] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 3
- [14] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81: 7th international joint conference on Artificial intelligence*, pages 674–679, 1981. 5
- [15] Fabricio Anicio de Magalhaes, Giuseppe Vannozzi, Giorgio Gatta, and Silvia Fantozzi. Wearable inertial sensors in swimming motion analysis: A systematic review. *Journal of sports sciences*, 33(7):732–745, 2015. 1, 2
- [16] Katie E McGibbon, DB Pyne, ME Shephard, and KG Thompson. Pacing in swimming: A systematic review. *Sports medicine*, 48(7):1621–1633, 2018. 1
- [17] Robert Mooney, Gavin Corley, Alan Godfrey, Conor Osborough, L Quinlan, and Gearóid ÓLaighin. Application of video-based methods for competitive swimming analysis: a systematic review. *Sports and Exercise Medicine*, 1(5):133–150, 2015. 2
- [18] Jorge E Morais, João P Oliveira, Tatiana Sampaio, and Tiago M Barbosa. Wearables in swimming for real-time feedback: A systematic review. *Sensors*, 22(10):3677, 2022. 1, 2
- [19] Phlex. Phlex swim coach. <https://www.phlexswim.com/coach>, 2023. Accessed April 2025. 2
- [20] Matteo Ruggero Ronchi and Pietro Perona. Benchmarking and error diagnosis in multi-instance pose estimation. In *Proceedings of the IEEE international conference on computer vision*, pages 369–378, 2017. 3
- [21] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986. 4
- [22] Thu Tran, Kenny Tsu Wei Choo, Shaohui Foong, Hitesh Bhardwaj, Shane Kyi Hla Win, Wei Jun Ang, Kenneth Goh, and Rajesh Krishna Balan. Analyzing swimming performance using drone captured aerial videos. In *Proceedings of the 10th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications (DroNet)*. ACM, 2025. 1, 2
- [23] Ultralytics. Github - ultralytics yolov11. <https://github.com/ultralytics/ultralytics>, 2025. 3
- [24] Daniel Ward. The effect of stroke type, stage of competition and final race position on pacing strategy in 200m swimming performance. 2018. 1
- [25] Timothy Woinoski and Ivan V Bajić. Swimmer stroke rate estimation from overhead race video. pages 1–6, 2021. 1, 2
- [26] Timothy Woinoski, Alon Harell, and Ivan V Bajić. Towards automated swimming analytics using deep neural networks. *arXiv preprint arXiv:2001.04433*, 2020. 2
- [27] Dan Zecha, Thomas Greif, and Rainer Lienhart. Swimmer detection and pose estimation for continuous stroke-rate determination. In *Multimedia on Mobile Devices 2012; and Multimedia Content Access: Algorithms and Systems VI*, pages 282–294. SPIE, 2012. 2