Universität Rostock

Traditio et Innovatio

visual & analytic computing

Drowning Detection with Computer Vision

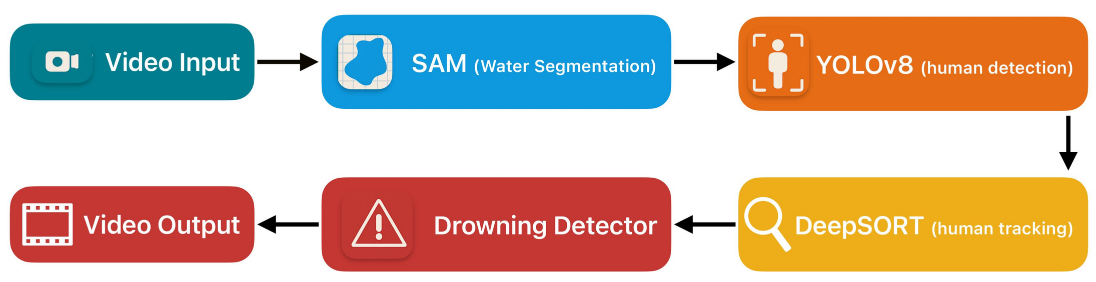
Modular Pipeline with YOLOv8, SAM, DeepSort

Motivation

Drowning is a major global health issue and, according to the WHO, ranks among the **three most common causes** of unintentional deaths. Every year, about **236,000 people** lose their lives due to drowning, with children and young adults being most affected [1, 2]. A key challenge is that drowning often occurs silently and without obvious warning: victims show almost no visible signals, cannot call for help, and frequently submerge within seconds [3]. Traditional preventive measures such as education, swimming lessons, or the deployment of lifeguards are important but often insufficient, especially at crowded beaches or poorly monitored coastal areas. This is where research comes in: **AI-powered systems** based on computer vision and machine learning offer the possibility of detecting at-risk individuals early and initiating rescue measures. This opens new perspectives for beach and coastal safety.

Methodology

After processing beach video recordings, a total of 18 video clips were available: 13 videos with an average duration of 20 s containing **drowning scenes**, and 5 videos with an average duration of 20 s showing only **normal swimming**. The videos contained scenes with one to three people simultaneously.



1. **Video Input:** Raw camera recordings.
2. **SAM – Segment Anything Model:** On the first frame of each video, the water area is segmented using SAM [4].
3. **YOLOv8 – Person Detection:** YOLOv8 is used to detect people in the water for each frame [5].
4. **DeepSORT – Tracking:** Detected persons are tracked over time using DeepSORT, producing stable IDs [6].
5. **Drowning Detector:** For each DeepSORT-tracked individual, the system checks whether they remain invisible for an extended period. DeepSORT retains the **ID** for up to **90 frames** (approx. 3 seconds), even if the person is temporarily lost due to occlusion or camera motion. This reduces false alarms. However, if a person remains invisible after this period, an **alarm** is triggered after another **1.5 seconds**.
6. **Video Output:** Results are visualized with bounding boxes, IDs, and alarm signals.

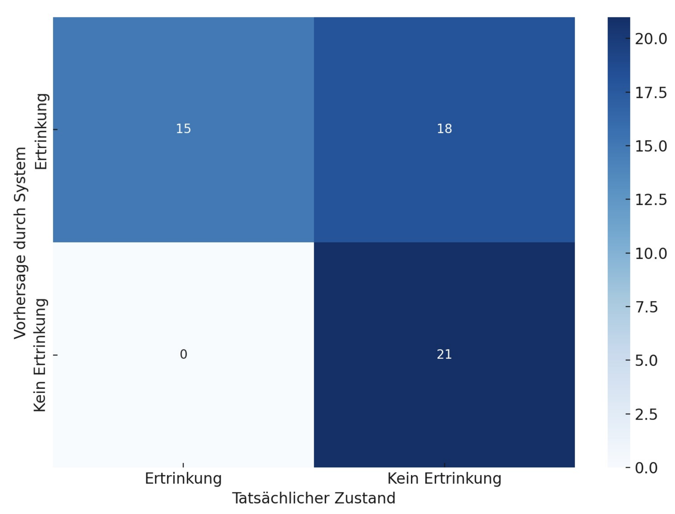
These figures illustrate that with **SAM**, individuals sitting or walking on the beach are not tracked. This ensures that no alarms are raised when they leave the frame.

This example shows two swimmers who were correctly detected and assigned unique **IDs**. When both disappeared underwater for about three seconds, the system detected this – triggering **alarm messages** in the output video.

Evaluation and Results

The detected events were compared against annotated ground-truth data. Key metrics are presented below:

- **Recall (Sensitivity):** 1.00 → All actual drowning cases were detected by the system. This is crucial since in safety-critical applications **no case must be missed**.
- **Precision:** 0.45 → Of all cases detected as drowning by the system, only 45 % were correct. This indicates that there are still **many false alarms** that need to be reduced.
- **F1-Score:** 0.63 → The F1-score combines precision and recall into a single value. It shows a **reasonable balance**, but also highlights the need for improvement.
- **Accuracy:** 0.67 → Two-thirds of all decisions (drowning vs. no drowning) were correct. However, in a safety-critical context, accuracy is less meaningful since recall is more important than overall correctness.



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Discussion

The analysis shows that the high number of false alarms is mainly due to technical limitations: unstable camera handling and zooming make robust detection difficult, while the small dataset of only 18 videos limits generalizability. Additionally, interactions between multiple swimmers occasionally lead to tracking errors in DeepSORT, causing individuals to lose their IDs. Overall, these findings emphasize the need for more data and improved stabilization to sustainably reduce false alarms.

Outlook

In the future, the system will be made more robust by using stabilized cameras and adjusted perspectives (e.g., overhead view). Furthermore, advanced models will be tested, and segmentation will be enhanced by automatically generating SAM prompts. Finally, a standardized evaluation methodology and an extended dataset will enable more reliable assessment and improved generalizability.

References

[1] Jalalifar, S., Belford, A., Razmjou, A., Abbassi, R., Mohseni-Dargah, M., & Asadnia, M. (2024). Enhancing Water Safety: Exploring Recent Technological Approaches for Drowning Detection. ETASR, 7804.

[2] Zhang, Y., et al. (2023). AI-based Drowning Detection Systems: A Review. BMC Medical Imaging, 23(549).

[3] Jalalifar, S., Belford, A., Erfani, E., Razmjou, A., Abbassi, R., Mohseni-Dargah, M., & Asadnia, M. (2024). Enhancing Water Safety: Exploring Recent Technological Approaches for Drowning Detection. Sensors, 24(331). https://doi.org/10.3390/s24020331

[4] Kirillov, A. et al. (2023). Segment Anything. Meta AI Research. https://github.com/facebookresearch/segment-anything.

[5] Ultralytics. YOLOv8 (2023). https://github.com/ultralytics/ultralytics.

[6] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple Online and Realtime Tracking with a Deep Association Metric (DeepSORT). https://github.com/nwojke/deep_sort.

Author Rowshanak Hosseinzadehattar

supervised by Supervisor Nr. One Dr. rer. nat. Sebastian Bader, Supervisor Nr. Two Jun.-Prof. Dr. rer. nat. Stefan Oehmcke

VAC | HAIML | COMPUTER SCIENCE | **University of Rostock**, A-Einstein-Str. 22 | Rostock, Germany