Mark Shperkin

Travis Shuler II

CSCE585 Machine Learning Systems

November 25, 204

Underwater Swimmer Pose Estimation Using HRNet-W32

Introduction to Results

Our initial project idea was to create a 3D pose estimation model that would allow the swimmer, after going a couple of laps in the pool, to receive instant feedback, allowing the swimmer to see key joint points that affect their swim style and stroke pattern for a better outcome in times. We condensed our initial project scope to a pose estimation model using HRNet-W32 architecture to identify 13 anatomically key points. The 13 anatomically key points consisted of the head, shoulders, elbows, hands, hips, knees, and ankles. With enough training data that is diverse, we would be able to estimate poses and generalize swimmer movements accurately. Our Milestone 3 report discusses and focuses on the results obtained, evaluating the promising outcomes and limitations identified through the process and proposing improvements for future iterations.

Positive Outcomes

The HRNet-W32 architecture has demonstrated its effectiveness as a pose estimation model, particularly for tasks requiring precision in spatial feature learning. One of the critical successes of this implementation was the model's ability to accurately predict key point locations in familiar swimming poses present in the training data. This accuracy is attributed to HRNet-W32's unique multi-resolution design, which allows for retaining high-resolution representations throughout the network. By progressively fusing multi-scale features, the model effectively combines detailed spatial features from higher resolutions with abstracted information from lower resolutions.

The performance of the 411-frame model underscores the importance of dataset size and diversity. Compared to the 84-frame model, the 411-frame model exhibited better generalization to poses and strokes not explicitly represented in the training dataset. This improvement suggests that even modest expansions in dataset size can significantly enhance the model's ability to learn robust representations. Furthermore, quantitative metrics such as Hamiltonian distances between predicted and ground truth key points confirmed that the model performed consistently well on training data, reaffirming the architecture's suitability for underwater pose estimation tasks.

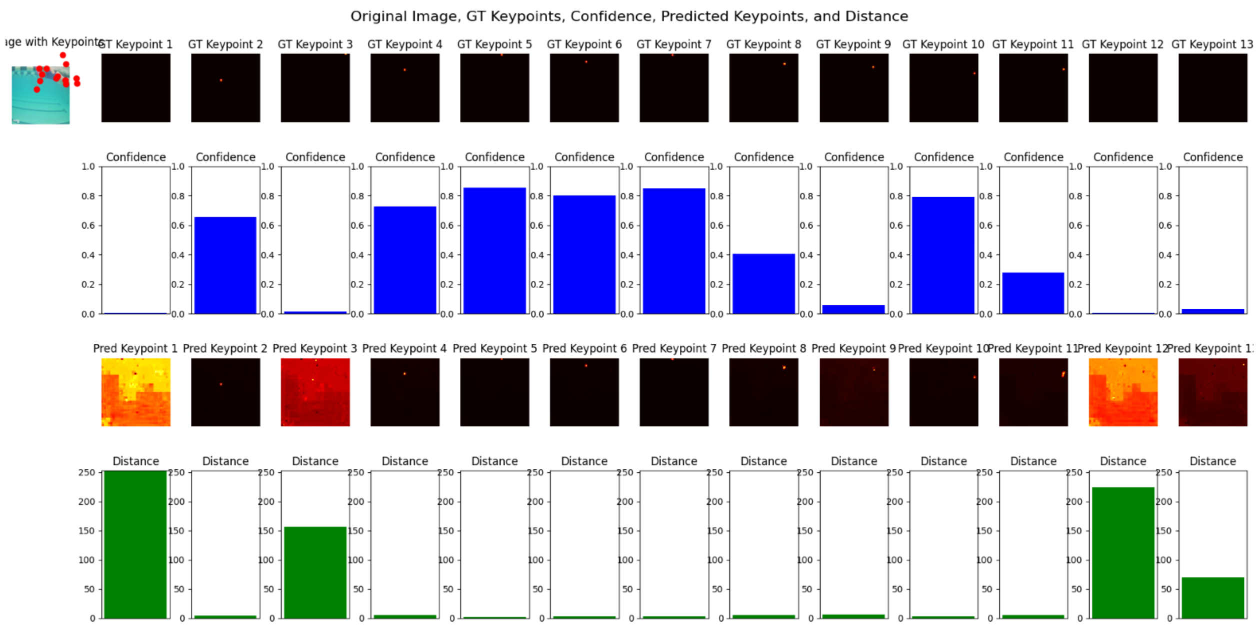

Figure 1.a: Model-411 Performance on one frame.

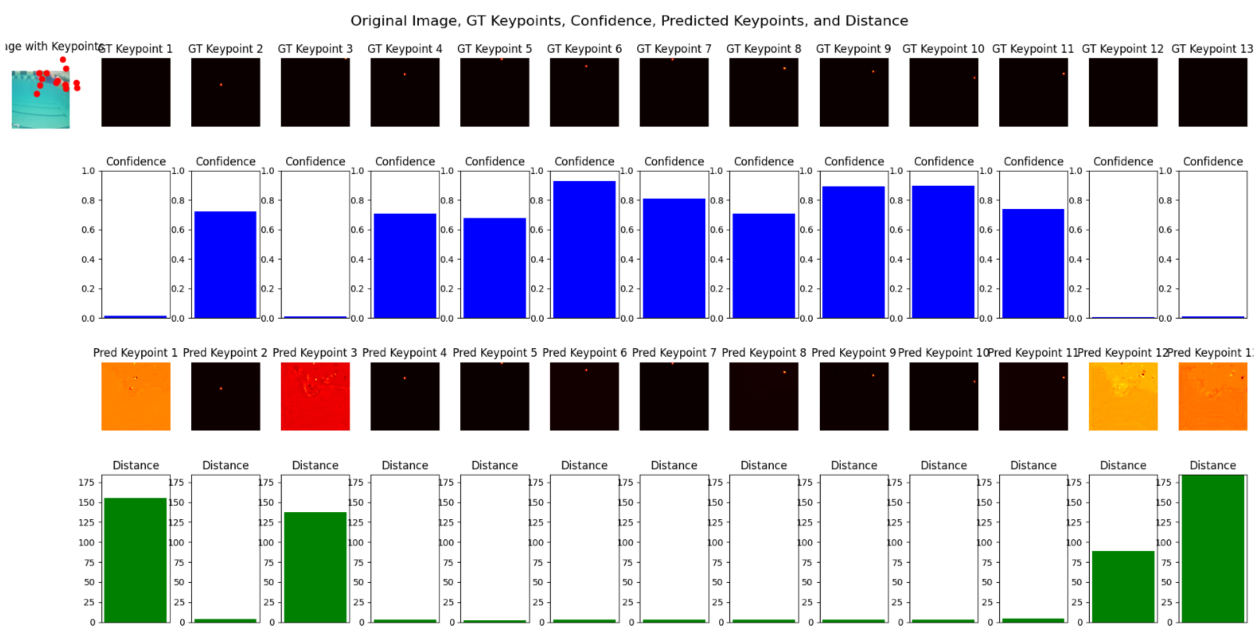

Figure 1.b: Model 84 Performance on one frame from the same dataset.

These results validate the project's hypothesis that HRNet-W32 can be a robust backbone for underwater pose estimation, particularly in controlled environments where the dataset aligns closely with real-world scenarios.

Negative Outcomes

While the model exhibited strength in pose estimation for familiar scenarios, several limitations became apparent when evaluating its generalization ability. The most significant challenge arose from the need for more diversity in the training dataset. Both models struggled when presented with swimmer poses or strokes not adequately represented during training. The 84-frame model demonstrated a pronounced inability to generalize, performing well only on poses identical to those it had seen. This limitation highlights the susceptibility of machine learning models to overfitting when trained on small or narrowly scoped datasets.

Another major shortcoming was the need for data augmentation techniques. Augmentation strategies, such as random cropping, rotation, flipping, and brightness adjustments, are commonly employed to introduce variability into training data and improve model robustness. The model's ability to handle real-world variations—such as lighting conditions or swimmer orientation—was significantly reduced without these techniques.

Evaluation constraints also posed a limitation. The absence of dedicated validation and test datasets prevented a rigorous assessment of model performance across varied scenarios. At the same time, qualitative analyses provided insights into general trends and quantitative benchmarks for robustness and accuracy under diverse conditions that needed to be improved. This gap in the evaluation framework reduced the reliability of conclusions drawn about the model's generalization capabilities.

Conclusion and Next Steps

The HRNet-W32 model demonstrates significant potential for addressing the challenges of underwater pose estimation. Its ability to accurately estimate swimmer key points in familiar scenarios establishes a solid foundation for further development. However, the limitations observed—particularly regarding generalization and robustness—underscore the need for strategic enhancements in future iterations.

To address these challenges, the priority is to expand the dataset, incorporating more frames that capture diverse swimming strokes, poses, and environmental conditions. By increasing dataset variability, the model will have a broader range of scenarios to learn from, improving its ability to generalize unseen data. In parallel, implementing data augmentation techniques will further enhance robustness by simulating variations in input data, such as changes in lighting, orientation, and occlusions.

Additionally, introducing a structured evaluation framework that includes separate validation and testing datasets is critical. Such a framework will provide a more rigorous assessment of the model's performance and ensure that its capabilities extend beyond the confines of the training dataset.

Hyperparameter tuning and architectural refinements may also be explored in future iterations to optimize the model's learning process and computational efficiency. By systematically addressing these areas, the model can evolve into a more robust and generalizable tool for underwater pose estimation, with applications extending to real-world swimming analysis and training optimization.