AI Coach for Divers

Xiaohang Yu
Tsinghua University
China
yuxh21@mails.tsinghua.edu.cn

Lekang Yuan
Tsinghua University
China
yuanlk21@mails.tsinghua.edu.cn

***Abstract*—Neutral buoyancy skill is a primary skill to be mastered by scuba divers and can be challenging for beginners. In this paper, we built a smart AI Coach to help scuba divers to master the skill. We extracted motion and breath features from videos recorded in a deep-diving pool, established a simulation environment for training coaching agents, and investigated the application in the actual environment. We hope such methods can facilitate the training of divers' skills, enhancing human's ability to explore the rest 70% of the earth underwater.**

***Index Terms*—neutral buoyancy, motion perception, reinforcement learning**

I. INTRODUCTION

For most sports, it's of great significance for beginners to have a one-to-one coach to evaluate the action and give guidance during training. As employing a human coach is expensive, a low-cost virtual AI coach is desirable. With the development of deep learning methods, such AI coach is becoming available, and examples can be found in areas such as body building [1], [2] and yoga [3].

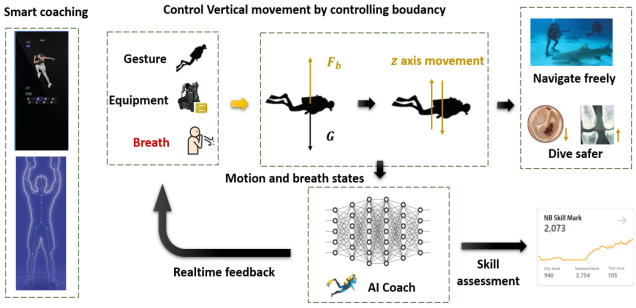

Fig. 1. Smart coaching for the neutral buoyancy problem

In this work, we plan to explore the application of artificial intelligence in the area of diving, focusing on the problem of neutral buoyancy (Fig.1). While underwater, a diver controls the vertical position by controlling buoyancy, and neutral buoyancy means keeping the buoyancy equal to the gravity to hold the vertical position still. It is an extremely important skill for navigating freely underwater, saving gas to dive longer, and avoiding decompression disease to dive safer. This requires the diver to adjust the counterweight, body gesture, balance control device (BCD), and breath properly, which make it a very challenging task for beginners. We believe an AI coach may facilitate the training of such skills by recording training data, providing live feedback, and visualizing the training progress.

To build an AI Coach for Divers, three problems must be solved:

Perception of the diver's state: How to extract the motion and breath states of the diver? To provide guidance, a coach must first observe the state of the trainee. There are two technology pathways to enable an AI coach the ability to observe human motion: IoT based approach and Computer Vision-based approach. Lots of researchers have applied the Internet of things (IoT) to fitness [4], [5], which extracts human motion and other features like heart rate by placing sensors on the body. However, applying such methods to the area of diving can be quite challenging due to the strict water-proofing requirement and the difficulties of signal transmission underwater. On the other hand, obtaining visual signals in deep-diving pools is relatively easy with motion cameras, and most deep-diving pools even have several cameras already installed. Visual observations can provide a lot of useful information for coaching, and with the development of deep learning, we do have methods for extracting that information, such as keypoint extraction methods [1], [3], [6]. Therefore, we plan to use CV based approach to observe the state of divers.

Behavior optimization: How to choose the right action based on observed motion and breath states? After seeing the state of the trainee, a coach must also know how to provide the right feedback. Feedback on adjusting counterweight, body gesture, and BCD can be computed directly based on the extracted key points, as there are clear rules to follow. However, how to control the breath rate and switch between breathing in and breathing out is not that clear. To teach an AI coach for controlling the breath, we built a simulation environment for the problem based on insights from actual data and used reinforcement learning to train an agent to master the skill of neutral buoyancy.

Application: How to apply the smart coaching methods underwater in an actual pool environment? Building an AI coach for divers can be quite different from that for other sports, as states perception and behavior feedback is performed underwater. Besides designing the coaching software, we must choose proper hardware to tackle the challenge, so we investigated the environment of a typical deep-diving pool and proposed device setups for recording videos and providing feedback underwater both for individual divers and the diving

club.

II. METHODS

A. Device setup and data collection underwater

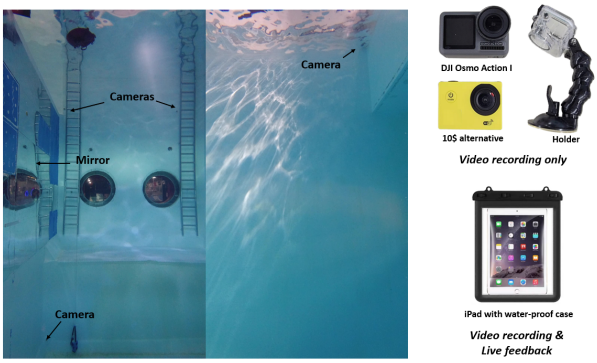

Fig. 2. Pool environment and hardware setup

To record videos underwater, individual divers can use a motion camera with water-proofing cases mounted onto the flat surfaces of the diving pool. Besides individual usage, diving pool owners can also use the cameras already installed in the pool (Fig.2) to record customers' motions and provide a performance analysis report as a service.

To provide real-time feedback, there must be a device to obtain video input, run the coaching program, and display feedback information underwater. We found this is available now with a tablet such as an iPad: a tablet has a front camera for video recording and a screen for display, the efficient network can run on mobile devices at video rate, and water-proofing cases for tablet can be bought easily. This can be used individually (one can just install the app on his tablet) or provided by the diving pool as a service (rent such configured tablet to the customer). For better display, the diving pool can also install smart mirrors like FITURE's on the way, building a "smart diving pool".

The data used in this project was recorded using the DJI Osmo Action I at 1080p, 30fps in a typical deep-diving pool in Shenzhen. From the video, the motion and breath states of the diver can be extracted.

B. Perception of motion and breath

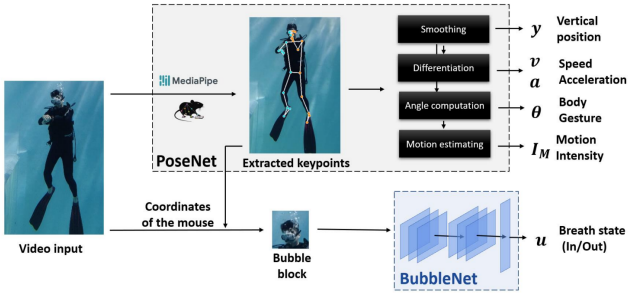

Fig. 3. Perception of motion and breath

PoseNet: Extracting the motion of divers is the first thing to do to build an AI coach. This can be done with deep-learning-based keypoint extraction methods. We choose to use Mediapipe [6], a pre-trained pose tracking framework developed by Google and evaluated its performance on the training videos.

Post-processing and motion estimation: The outputs of the PoseNet are 33 key points for each frame. To extract the dynamic of the diver's position, we averaged key points with little relative body motion, which is then differentiated to get v and a. As there is some level of high-frequency noise in the extracted key points, performing differentiation on y directly will lead to corrupted v and a, so low-pass filtering was performed on y first. Besides the overall position, other useful information can also be obtained from the extracted key points, such as the posture information for determining if the diver is in the "trim" state, or the intensity of the leg's motion for extra assessment of the diver's skill.

Bubble Net: Besides the posture and the position, the breath state of the diver is also an important feature to be extracted. To determine the breath state, for each frame an image block is cropped to the same size around the mouth of the diver with key points extracted with the PoseNet. The image block is used as the input to the Bubble Net, and the Bubble Net returns a binary classification output indicating if the diver is breathing in (no bubbles are generating) or breathing out (bubbles are generating). We adopt EfficientNet [7] to be the bubble classifier. EfficientNet uniformly scales all dimensions of depth/width/resolution using a simple yet highly effective compound coefficient and yields better performance than scaling up MobileNets and ResNet. The training settings are as follows: RMSProp optimizer with decay 0.9 and momentum 0.9; batch norm momentum 0.99; weight decay 1e-5; initial learning rate 0.256 that decays by 0.97 every 2.4 epochs; drop out rate 0.2.

C. Behavior optimization based on reinforcement learning

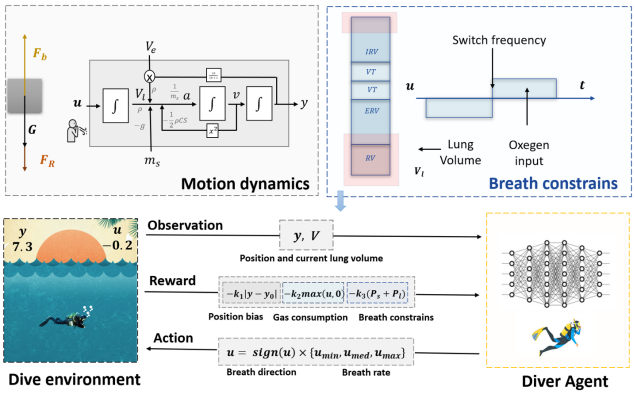

Fig. 4. The reinforcement learning environment for the neutral buoyancy problem

The optimization problem: To simplify the problem, we abstracted the diver to be a mass point with motion dynamic

shown in Fig.4. The main input of the system is the breath rate u, and the main output of the system is the vertical position z, and the control target is to minimize the distance between z and a given depth z for any time point.

Breath constraints: Besides the motion dynamics, constraints on the breath rate input U must be addressed to fit a real human's behavior. There are several constraints here: 1) One can't breathe too much to exceed the lung's capacity. 2) One can't breathe too slowly or he can't get enough oxygen. 3) One won't switch between breathing in and breathing out too frequently.

Diver agent training with reinforcement learning: We choose to use reinforcement learning to obtain the optimized pattern of input U. A simulation model was built following the dynamics in Fig.4, providing standard interfaces like a gym environment. To make it easy to generate a proper breath pattern, action U is replaced with breath direction (in or out) and breath rate (choosing between minimum rate u_min, medium rate u_med, and maximum rate u_max), and the reward is designed to penalize three values:

- the offset from y_0.
- the gas consumption.
- abnormal breathing action, including breathing too much and switching too frequently.

After establishing the environment, diver agents were trained with DQN to see if an agent can solve the environment if it can learn to breathe like a human, and what breath pattern it can generate.

III. RESULTS

A. Motion and breath extraction

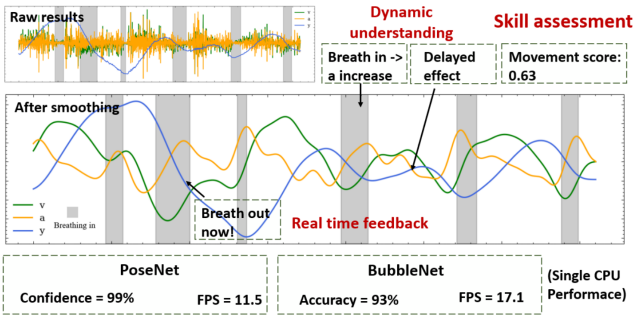

Fig. 5. Motion and breath extraction results

For PoseNet, the confidence score of motion extraction is 99%. For Bubble Net, the weighted average precision, recall rate, and f1-score is 0.94, 0.93, and 0.94, respectively, and accuracy is 0.93. In practice, our model is expected to be real-time. Therefore, the frame rate of image processing is also taken into consideration. The frame rate of PoseNet and Bubble Net is 11.5 and 17.1. We found that PoseNet and Bubble Net can extract the motion and breath states of the diver with high accuracy and speed, which enables its application even on mobile devices.

Combining the observed motion and breath states, the AI coaching framework can provide an understanding of the dynamics such as the delayed effect, assessment on the divers' buoyancy control skill, and real-time feedback, as shown in Fig.5 and the supplementary video.

B. Agents training

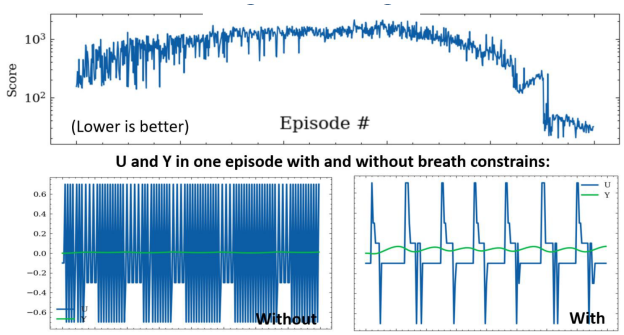

Fig. 6. Training record and the breath pattern

We trained diver agents with 3 layers full-connection network with DQN on the established environment. After training, the agent can learn to solve the environment, hold at a fixed vertical position by controlling neutral buoyancy. By adding extra reward terms for the breath contains, the agent can also learn to generate breath patterns like a human, as shown in Fig.6.

IV. DISCUSSION

After doing the project, we think the question "how can an AI coach benefits trainees" can be answered from the relationship between observability and controllability, similar to that in the control science: you can control your movement better if you can get better observation.

First, an AI coach can remind you to follow the given instructions at any moment with an accurate measurement of how well you are doing. For example, divers are encouraged to breathe slowly but beginners often breath fast unconsciously, and an AI coach can remind you to slow down when the measured duration of last breath is too short. Besides providing feedback in real-time, an AI coach can also record your overall performance of one training session in a digitalized manner, making it easy to visualize your progress in the long term and compare it with others.

Second, it may reveal some inner states that can not be observed well by humans. One of the major challenges of the neutral buoyancy problem is the delay effect: there are three times of integration between breath input u and position output z. Imagine a case when the position is below the given point, but you have an upwards velocity and an upwards acceleration. As human can't sense v and a accurately, he may choose to continue breathing in according to z, so a and v will continue to increase and becomes quite large, leading to a large overshoot of z above the given point (which we can observe from Fig.5). With the v and a acutely sensed by an AI

coach, a diver can get a sense of when to change the breath in advance to combat the delayed effect.

In the project, we used reinforcement learning to obtain optimized breath input. The problem is an interesting RL problem itself, and training RL agents will help us get a deeper understanding of the neutral buoyancy skill. However, after experiments, we don't actually think it is the most suitable approach in practical usage, as it is not explainable and there are complex simu2real problems to consider. Some simple and understandable methods might be preferred; for example, we can compute an "overall upward trend" from $k_1 z + k_2 v + k_3 a$, and switch the breath direction when it's above a certain threshold; k_1, k_2, k_3 and the threshold can be optimized with a genetic algorithm. We may try those approaches and test them in the real application later.

V. CONCLUSION

In this paper, we explored the application of artificial intelligence in the area of diving by analyzing the motion perception, coaching training, and application problem focusing on the neutral buoyancy skill. We hope such a smart AI coach can help improve scuba divers' skills, enhancing human's ability to explore the rest 70% of the earth underwater.

REFERENCES

[1] "Fiture," https://www.fiture.com/cn/, 2021.

[2] B. Yong, Z. Xu, W. Xin, L. Cheng, L. Xue, W. Xiang, and Q. Zhou, "Iot-based intelligent fitness system," *Journal of Parallel and Distributed Computing*, vol. 118, no. PT.1, pp. 14–21, 2017.

[3] "Otari– an interactive workout mat with display," https://gforgadget.com/cool-gadgets/otari-interactive-workout-mat/, 2020.

[4] C. C. Lin, Y. S. Liou, Z. Zhou, and S. Wu, "Intelligent exercise guidance system based on smart clothing," 2019.

[5] A. Jl, B. Qh, A. Tg, A. Sw, and A. Ys, "What and how well you exercised? an efficient analysis framework for fitness actions," *Journal of Visual Communication and Image Representation*, 2021.

[6] C. Lugaresi, J. Tang, H. Nash, C. Mcclanahan, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," 2019.

[7] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2019.