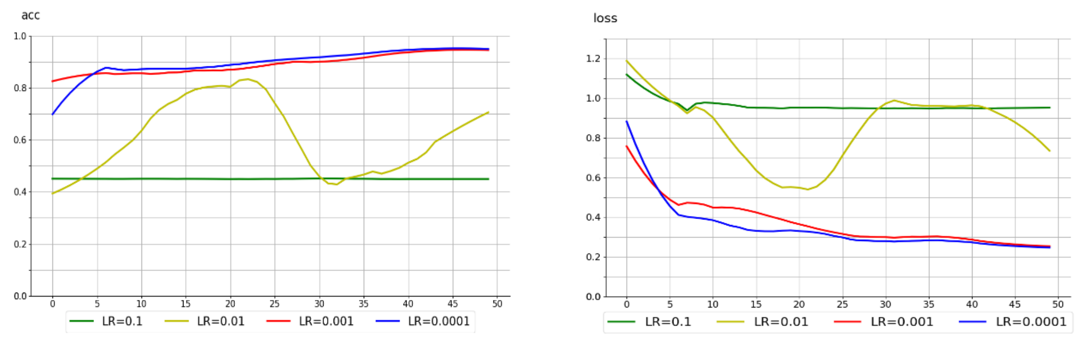
# Supplementary Figures and Tables

**Table S1.** The architectures of models in References [14-16].

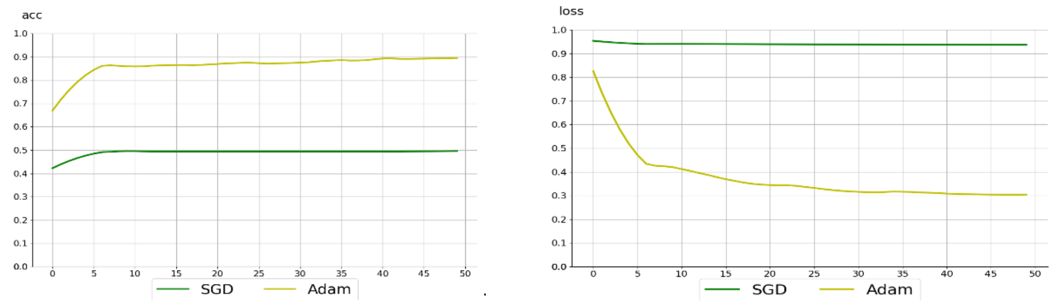
| Reference [14] | Reference [15] | Reference [16] |
|---|---|---|
| Conv1+Relu1 | Conv2d, 3×3 | C1(convolutional layer) |
| Conv2+Relu2 | Bneck, 3×3 | S2(subsampling layer) |
| Conv3+Relu3 | Bneck, 3×3 | C3(convolutional layer) |
| Pooling | Bneck, 3×3 | S4(subsampling layer) |
| Conv4+Relu4 | Bneck, 5×5 | C5(first fully connected layer) |
| Conv5+Relu5 | Bneck, 5×5 | F6 (fully connected layer) |
| Conv6+Relu6 | Bneck, 5×5 | Softmax |
| Pooling | Bneck, 5×5 | |
| FC1 | Bneck, 5×5 | |
| FC2 | Bneck, 5×5 | |
| FC3 | Bneck, 5×5 | |
| Softmax | Bneck, 5×5 | |
| | Conv2d, 1×1 | |
| | Pool, 16×16 | |
| | Conv2d, 1×1, NBN | |
| | Conv2d, 1×1, NBN | |




(**a**) (**b**)

**Figure S1.** The evaluation result of different learning rate settings: (a) accuracy curve; (b) loss curve.

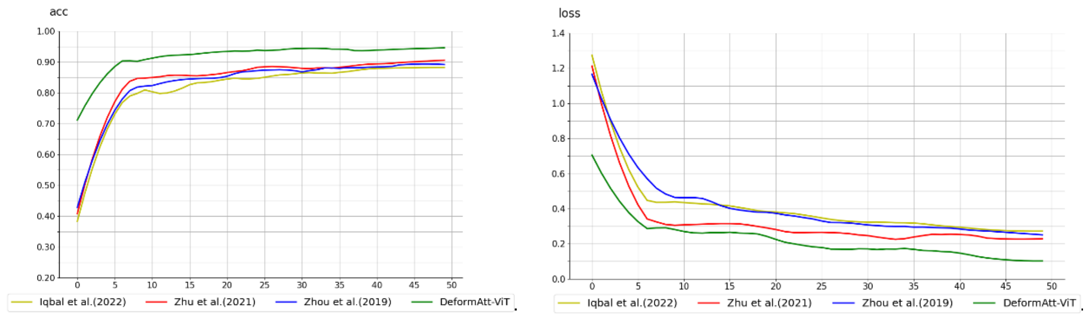
Note: When the learning rate was set to 0.1, it led to oscillations in the early stages, resulting in persistently high loss values. A relatively large initial learning rate (e.g., 0.01) also caused the model to fail to converge, indicating the need for further reduction in the learning rate. When the learning rate was too small (e.g., 0.0001), it resulted in slow convergence speed and requires longer training time. Therefore, we chose the learning rate of 0.001 as the optimal value in this study.




(**a**) (**b**)

**Figure S2.** The evaluation result of different optimizers: (a) accuracy curve; (b) loss curve.

Note: Adam achieved a better performance. This is because Adam combined the momentum method to avoid oscillation and can quickly converge to the vicinity of local minima.




(**a**) (**b**)

**Figure S3.** The evaluation result of four models: (a) the accuracy change curve of Iqbal et al. (2022), Zhu et al. (2021), Zhou et al. (2019), and DeformAtt-ViT; (b) the loss value change curve of Iqbal et al. (2022), Zhu et al. (2021), Zhou et al. (2019), and DeformAtt-ViT.

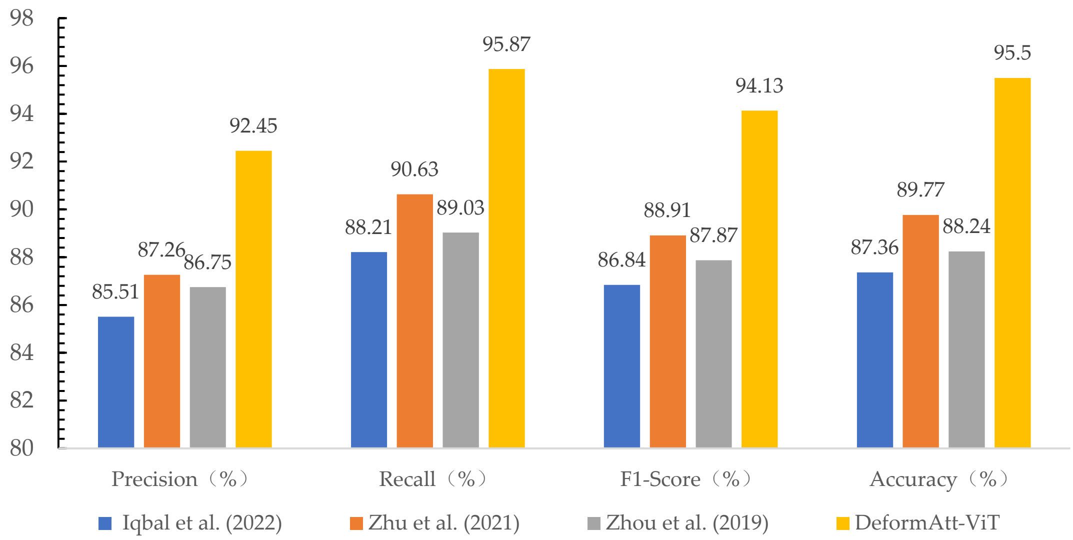
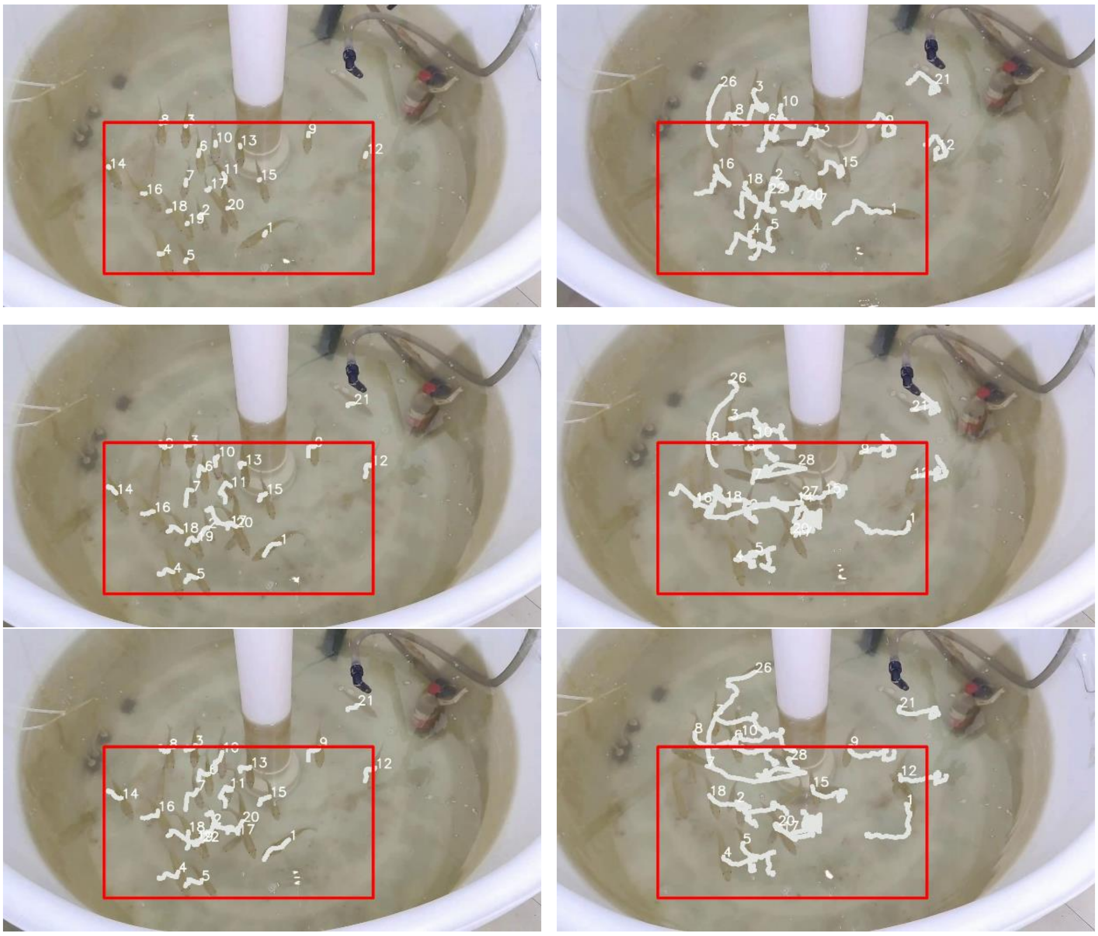



**Figure S4.** Comparison of accuracy, recall, F1-score, and accuracy between Iqbal et al. (2022), Zhu et al. (2021), Zhou et al. (2019) and DeformAtt-ViT.



(a) (b)

**Figure S5.** Trajectories in different states: (a) hungry state; (b) swimming state

**References**


Iqbal, U.; Li, D.; Akhter, M. Intelligent Diagnosis of Fish Behavior Using Deep Learning Method. Fishes 2022, 7, 201. https://doi.org/10.3390/fishes7040201. [Reference 14 in the paper]

Zhu, M.; Zhang, Z.; Huang, H.; Chen, Y.; Liu, Y.; Dong, T. Classification of perch ingesting condition using light-weight neural network MobileNetV3-Small. Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng. 2021, 37, 165–172. https://doi.org/10.11975/j.issn.1002-6819.2021.19.019. [Reference 15 in the paper]

Zhou, C.; Xu, D.; Chen, L.; Zhang, S.; Sun, C.; Yang, X.; Wang, Y. Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. Aquaculture 2019, 507, 457–465. https://doi.org/10.1016/j.aquaculture.2019.04.056. [Reference 16 in the paper]