

Title: End-to-End Multiple Object Tracking Using MOTRv3 with Fish Videos

Research question: Can the proposed end-to-end multiple object tracking approach using MOTRv3 be effectively extended and adapted to address the challenges of fish tracking in datasets characterized by small and visually similar fishes with rapidly changing appearances due to complex 3D motion, thereby enhancing fish tracking accuracy for ecological studies? It is worth noting that while the MOTRv3 code is currently unavailable, adaptation from MOTR code is possible.

Relevant context (methodology): Multiple object tracking (MOT) presents challenges in effectively modeling object behavior over time, primarily involving two aspects: detection and association. Typically, the association part is not learned, and learned object id embeddings are combined using the Hungarian matching algorithm. However, MOTR offers a trainable end-to-end solution that incorporates object detection and association within the same context of detection queries and track queries, effectively extending deformable DETR. Nevertheless, MOTR encounters an imbalance between detection and track queries, resulting in lower performance when compared to non end-to-end methods. Addressing this issue, MOTRv3 introduces the release-fetch supervision strategy, which balances the label assignment process. Moreover, strategies such as label distillation and track group denoising appear to enhance supervision for both detection and association.

References:

Zeng, Fangao, et al. "Motr: End-to-end multiple-object tracking with transformer." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.
Yu, En, et al. "MOTRv3: Release-Fetch Supervision for End-to-End Multi-Object Tracking." arXiv preprint arXiv:2305.14298 (2023).