aquaculture journal

MDPI

Article

Markerless Individual Re-Identification of Visually Homogeneous Zebrafish for Aquaculture Monitoring Using Hybrid Spatial–Visual Tracking and Angular-Margin Deep Embeddings

Georgi Atanasov [1,*], Martin Ilarionov [2], Kapka Mancheva [3,*] and Zlatko Zahariev [2]

[1] Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, 25 Georgi Bonchev Str., 1113 Sofia, Bulgaria
[2] Faculty of Applied Mathematics and Informatics, Technical University of Sofia, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria; milarionov@tu-sofia.bg (M.I.); zlatko@tu-sofia.bg (Z.Z.)
[3] Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 23 Georgi Bonchev Str., 1113 Sofia, Bulgaria
* Correspondence: gatanassov@gmail.com (G.A.); kapka_mancheva@abv.bg (K.M.)

Abstract

Persistent markerless identification of visually similar fish is difficult when occlusion interrupts tracking. This study evaluates a hybrid spatial–visual pipeline combining YOLO26n detection, ByteTrack association and a Swin-T re-identification (Re-ID) model. Experiments were carried out using two Aalborg University (AAU) zebrafish recordings containing six known individuals. The detector and embedding model were retrained on consecutive temporal blocks separated by ten omitted frames; 426 frames were held out from retraining and checkpoint selection. A 400 px vertical mask simulated occlusion, and predictions were evaluated against ground-truth AAU Object IDs. The added association stage restored the prior identity in three of eighteen eligible episodes after full emergence (16.7%), compared with none for ByteTrack. All three successes bridged changed ByteTrack IDs, with successes in both recordings; one assignment was incorrect and fourteen episodes remained unresolved. Under a separate partial-reappearance definition, Re-ID recovered six of twenty-four episodes (25.0%), versus none for ByteTrack, with one incorrect identity, one different identifier for the same fish and sixteen unresolved episodes. These verified recoveries demonstrate that the added association stage can restore identities after tracker interruption. Limited assignment coverage and testing on previously seen individuals constrain conclusions about reliability and transfer to independent recordings.

Keywords: aquaculture monitoring; zebrafish; *Danio rerio*; markerless tracking; fish re-identification; computer vision; deep learning; Swin Transformer; ArcFace; ByteTrack

Check for updates

Academic Editor: Aires Oliva-Teles

Received: 27 June 2026
Revised: 14 October 2026
Accepted: 21 October 2026
Published: 24 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Aquaculture increasingly depends on automated monitoring technologies capable of observing fish health, welfare, growth and behavior without disturbing the animals. Manual visual inspection is time-consuming, subjective and difficult to scale in dense or visually complex aquatic environments. Computer vision and artificial intelligence have therefore become important tools for smart aquaculture, including fish recognition, counting and behavioral monitoring [1–9].

Recognizing an individual fish, and not only detecting that a fish is present, is important because aquaculture decisions are increasingly temporal and individual-specific. If the same fish can be recognized repeatedly, the system can accumulate a history for that individual, including feeding response, growth trajectory, body condition changes, activity level, social position, abnormal movement, visible disease signs and recovery after treatment. Such individual histories can support more precise automatic feeding, earlier disease control decisions, adjustment of thermal and photoperiod regimes, welfare assessment and selective management. Without persistent identity, each video frame is only a separate observation; with identity, observations become a longitudinal record that can be used for prediction and management.

Individual-level tracking is important because longitudinal information can reveal behavior, abnormal movement, social interaction, growth variation and early welfare deterioration. In research and production environments, persistent identity is often maintained using physical markers, RFID tags, dyes, surgical marks or manual observation. Such approaches may increase handling stress, affect behavior and limit scalability. Markerless identity preservation is therefore an attractive alternative for aquaculture and laboratory fish studies.

There is a significant problem regarding individual-level tracking for visually homogeneous fish. Zebrafish (*Danio rerio*), a widely used model species, are small, reflective, fast-moving and highly similar in appearance. Their bodies change shape with swimming motion, and their apparent texture varies with viewpoint and illumination. During occlusion or overlap, geometry-only tracking can lose identity continuity. Tracking-by-detection methods such as SORT and ByteTrack preserve short-term trajectories [10,11], but identity continuity may fail after complete disappearance or ambiguous crossings. Re-identification methods address this limitation by comparing visual embeddings [12–14].

Recent animal-tracking tools, including DeepLabCut and SLEAP, have enabled markerless multi-animal pose estimation and identity-aware tracking [15,16]. However, pose tracking and individual re-identification are not equivalent. Pose estimation localizes body landmarks, while re-identification must decide whether two visual fragments belong to the same individual. For visually similar fish, this requires learning subtle identity cues that may not be visible to human observers.

This study proposes and experimentally evaluates a hybrid spatial–visual system for markerless individual re-identification of zebrafish. The central question is whether transformer-based embeddings trained with angular-margin and metric learning objectives can help recover individual identities after track interruption. The evaluated task is closed-set identification of previously enrolled fish within a bounded recording processed offline; unseen-individual, cross-session and cross-dataset generalization are outside the scope of the present experiment. This is consistent with the recent computer vision literature in relation to fish, which emphasizes that controlled imaging, limited datasets, color-space dependence, occlusion, real-world transfer and explainability remain recurring challenges [1–3,8,9,17]. We describe a Triplet-only development baseline and a BNNeck model trained using ArcFace, Triplet and Center loss, and evaluate the final model against ByteTrack alone. The paper makes the following contributions: (i) a detector–tracker–re-identification pipeline for unmarked fish; (ii) an angular-margin embedding design for few-identity fish Re-ID; (iii) an occlusion-oriented evaluation protocol; and (iv) a ground-truth evaluation of identity recovery on temporally held-out segments from both recordings. Rank-1 retrieval consistency, retrieval mAP, cosine similarity, Principal Component Analysis (PCA) and separation gap are retained only as historical embedding diagnostics. The experimental validation is based on the publicly available AAU Zebrafish Re-Identification

Dataset [18,19], which provides side-view color recordings and identity-level annotations for zebrafish re-identification.

2. Related Work

Computer vision has been used in aquaculture and fisheries for recognition, monitoring, measurement and management [1]. Barbedo highlighted the gap between controlled academic experiments and operational deployment, emphasizing challenges such as underwater imaging, occlusion, low image quality and dataset bias [1]. Petrellis showed that image processing and deep learning can estimate morphological features such as fish length, height, area and fins [3]. Kumar et al. further showed that bioimage analysis methods are increasingly used to detect skeletal deformities in biomedical and aquaculture fish, connecting image analysis with welfare-relevant morphology rather than only category labels [8]. Recent studies have also used deep learning for salinity/freshwater fish classification [17] and transformer-based species recognition [20].

Fish disease detection is one of the most active areas of AI-enabled aquaculture. Reviews by Pauzi et al. and Li et al. describe image acquisition, preprocessing, segmentation, feature extraction and classification workflows [2,4]. Li et al. noted that aquatic diagnosis is more complex than many two-dimensional plant disease tasks because fish move in three-dimensional water environments and are affected by overlap, illumination, turbidity and rapid motion [2]. Recent studies and datasets, including SalmonScan, feature-ranking disease recognition, and CNN-based disease classification, show the growth of automated disease detection [5,21,22]. Classical image-based machine learning work has also used color and texture features for disease recognition [6], while deep learning parasite detection and recent explainable color-space approaches show the same movement toward faster and more interpretable automated diagnosis [7,9]. However, disease classification usually depends on visible pathological signs and does not preserve individual identity.

Tracking-by-detection pipelines combine object localization with frame-to-frame association. SORT introduced a simple real-time tracker based on Kalman filtering and Hungarian assignment [10]. ByteTrack improved multi-object tracking by associating almost every detection box, including low-confidence detections, which can help when objects are partially occluded [11]. In animal behavior analysis, DeepLabCut and SLEAP provide markerless pose estimation and multi-animal tracking [15,16]. These systems are powerful for behavior analysis, but full identity recovery among visually homogeneous fish remains difficult when geometry and pose are ambiguous.

Re-identification maps image fragments to an embedding space where same-identity samples are close and different-identity samples are far apart. Triplet loss is a common metric-learning objective [23], but it can be unstable in few-class settings and depends strongly on hard-sample mining. ArcFace introduces an additive angular margin in a normalized hyperspherical space, strengthening inter-class separation [13]. BNNeck separates the feature used for metric learning from a batch-normalized feature used for classification, reducing conflict between Euclidean and angular optimization [14]. Transformer-based ReID methods such as TransReID show that attention-based representations are effective for identity-sensitive recognition [12]. The present work transfers these principles to visually homogeneous fish [18].

Haurum et al. [18] introduced the AAU dataset and demonstrated crop-level zebrafish re-identification using handcrafted color and texture descriptors with metric and subspace learning. Recent work has also applied Swin-T to fish re-identification in conveyor-belt electronic monitoring imagery [24]. The present study uses the AAU dataset for a different engineering integration: learned Swin-T embeddings are combined with YOLO detection, ByteTrack short-term association and a persistent spatial–visual gallery for reconnecting

previously enrolled identities after track interruption. This reduces task-specific feature engineering, but no direct accuracy or computational superiority is claimed because the evaluation protocols differ [25].

3. Materials and Methods

3.1. Study Design

This study was designed as an experimental computer vision investigation of markerless individual re-identification in visually homogeneous zebrafish. The primary experimental objective was to evaluate whether a transformer-based embedding model can support recovery of a previously enrolled identity after simulated occlusion. V1 and V2 are sequential development configurations under the same detection and tracking framework; because several settings changed simultaneously, their comparison is not a controlled component ablation.

The final evaluation uses a newly trained V2 model and a newly fine-tuned detector. The earlier V1/V2 embedding diagnostics are retained as development results and are not compared numerically with the temporal test as if they measured the same endpoint.

3.2. Video Data and Annotation

The experimental material was obtained from the public AAU Zebrafish Re-Identification Dataset [18,19], released by Aalborg University, Denmark. The dataset contains side-view color video recordings of zebrafish (Danio rerio) acquired in a controlled laboratory setup. Six fish were recorded in total, in two sessions, with three fish at a time. The released material contains 2224 annotated frames at 2056 × 1542 pixels and 6672 bounding boxes: Vid1 contains 1086 frames and Object IDs 1–3, while Vid2 contains 1138 frames and Object IDs 4–6. The recordings were acquired in a 32 × 32 × 32 cm fish tank, with the fish constrained to swim within the first 3.5 cm of the tank. The original dataset includes video frames and annotation files with bounding-box positions, temporally consistent fish identity labels and four Boolean metadata tags: swimming direction, turning, occlusion and image glitch.

The source publication describes a variable acquisition frame rate [18]. We used the released annotated image sequences in filename order, without additional fixed-interval subsampling. A frame index therefore denotes sequence order rather than a verified constant time interval. The 30 fps setting used to encode the demonstration videos is an output playback setting, not a measurement of the source acquisition rate.

Training and validation fragments were cropped from complete ground-truth fish bounding boxes; no restricted body region was selected. Each recording was divided into consecutive training, validation and test blocks. Vid1 used frames 1–760 for training, 771–868 for validation and 879–1086 for testing; Vid2 used frames 1–796, 807–910 and 921–1138, respectively. Ten frames were omitted at each train/validation and validation/test boundary. The resulting totals were 1556 training frames (4668 crops), 202 validation frames (606 crops), 426 test frames and 40 omitted frames. All six identities occur in training and testing, so the task remains closed-set and within-recording. The test frames were excluded from this retraining and checkpoint selection. The gap prevents adjacent retained frames from crossing a split boundary; it does not establish statistical independence between observations of the same fish in the same recording.

3.3. Rationale for the Controlled Small-Cohort Design

The occlusion-oriented tracking experiment used both recording sessions of the AAU dataset, each containing three active fish identities. This small cohort supports a bounded feasibility study rather than population-scale validation. The task investigated here is not ordinary fish detection, species classification or disease-state classification; it is individual

biometric separation among visually homogeneous animals. In such a setting, the central scientific question is whether the embedding space can separate individuals whose differences are weak, local and not reliably visible to human observers.

A small controlled cohort reduces confounding factors such as lighting variation, tank heterogeneity, density effects, age differences and camera changes. It permits a bounded study of identity association under similar imaging conditions, but does not establish population-scale or cross-session performance; those require larger datasets and independent recordings.

The choice is also aligned with the long-term goal of the research: reliable identification is a prerequisite for longitudinal analysis of each fish. Once identity is preserved, subsequent modules can quantify individual growth, morphology, movement patterns, social interactions, abnormal behavior and welfare-related traits over time. Thus, the current experiment addresses the first layer of a broader pipeline for individual-level aquaculture analytics.

3.4. Methodological Components: YOLO, ByteTrack and Swin Transformer

The proposed pipeline uses three complementary computer vision components. The first component is YOLO-based detection. YOLO (You Only Look Once) is a family of one-stage object-detection models that localize objects in an image by predicting bounding boxes and confidence scores in a single forward pass through a neural network. In this study, the detector is used only to find fish locations in each video frame and to produce cropped fish images for later identity analysis.

The second component is ByteTrack short-term association. ByteTrack is a tracking-by-detection algorithm that links detections across consecutive frames. Unlike trackers that discard low-confidence detections immediately, ByteTrack uses both high-confidence and lower-confidence detections during association. This makes it useful when fish are partially visible, blurred or temporarily detected with lower confidence. In the present system, ByteTrack provides short-term trajectory continuity, but it is not expected to solve identity recovery after complete occlusion by itself.

The third component is a Swin Transformer [26] used as the visual re-identification backbone. A Swin Transformer is a hierarchical vision transformer that processes images through shifted local attention windows. Compared with ordinary convolutional networks, transformer-based models can capture relationships between distant image regions while still preserving local texture information. Here, the Swin Transformer converts each cropped fish image into a numerical feature embedding. These embeddings are then optimized so that images assigned to the same enrolled identity are close together and images assigned to different enrolled identities are far apart in the learned identity space. The three components are integrated into a sequential identity-recovery workflow. YOLO first localizes fish and generates cropped image fragments, ByteTrack preserves short-term spatial tracklets, and the Swin Transformer module compares quality-gated visual embeddings against stored identity galleries to recover identity after occlusion.

3.5. Simulated Occlusion Protocol

To evaluate identity recovery after disappearance and reappearance, a simulated occlusion stress test was used. A black vertical obstacle with a nominal width of 400 px was placed in the central region of the video frame (x = 828–1228 pixels; approximately 19.5% of the 2056-pixel frame width). When a fish passed through this region, visual evidence was removed. The primary retrospective evaluation defines an episode as a maximal consecutive run of frames in which the ground-truth box overlaps the mask and is fully contained within it for at least one frame. Both entry and subsequent full emergence must fall within the test

segment. The decision window starts at the first frame in which the entire box is outside the mask and ends immediately before the next mask overlaps or at the end of the test segment. We separately report episodes lacking a qualifying pre-occlusion identity and events censored by a test boundary. As a sensitivity analysis, each run of full containment is instead treated as an event ending at partial reappearance. In this alternative definition, the decision window ends before the next run of full containment or at test end. Neither definition assumes that a fish exits on the opposite side of the mask. Events were determined from AAU ground-truth boxes and the fixed mask geometry, independently of whether the pipeline returned an identity. This protocol does not reproduce all physical properties of real occlusion.

3.6. Detection Module

The first stage localizes fish in each frame using an Ultralytics YOLO26n detector (Ultralytics Inc., London, UK) ine-tuned for 20 epochs at an input size of 640 × 640 pixels [27]. For the temporal experiment, training started from the pretrained detector weights and used only the training blocks specified above. An interrupted run resumed from its own temporal-training checkpoint; the old fish-fine-tuned detector was not used as the starting model. For each frame, the detector outputs bounding boxes and confidence scores. The detections serve two purposes: (i) they provide spatial observations for ByteTrack, and (ii) they define cropped fish fragments for the re-identification network. Transfer learning was used because the available fish-specific dataset is smaller than the large-scale datasets typically used for detector pretraining.

The frozen temporal detector checkpoint selected during training was additionally evaluated on all 426 original, unmasked test frames (1278 annotated fish boxes), using the same temporal test lists as the tracking experiment. We used Ultralytics 8.4.37 validation at an image size of 640, with rectangular preprocessing, a confidence floor of 0.001 for the precision–recall curves, a maximum of 300 detections per image and no test-time augmentation. Precision and recall follow the evaluator's maximum-F1 operating point; mAP is reported at IoU 0.50 and averaged over IoU 0.50–0.95. No weights or association settings were changed. This measures detector localization on the original test images, separately from the downstream pipeline under the artificial mask.

3.7. Short-Term Tracking Module

The second stage applies ByteTrack to associate detections across adjacent frames [11]. ByteTrack performs two-stage association: high-confidence detections are matched first, and lower-confidence detections are then considered to preserve partially visible objects. This is useful for fish video because motion blur, reflections and partial overlap can reduce detector confidence. Nevertheless, ByteTrack is primarily a spatial association method. When a fish disappears behind an occlusion or crosses another fish, the tracker may lose identity continuity. The re-identification module is introduced to recover identity in such cases.

3.8. Baseline Re-Identification Model: V1

The V1 baseline uses an ImageNet-pretrained Swin-T backbone with a 512-unit linear layer, BatchNorm, ReLU, dropout (0.3) and a 128-dimensional feature embedding. The objective operates on triplets consisting of an anchor, a positive sample from the same identity and a negative sample from another identity. TripletMarginLoss with hard mining encourages the anchor-positive distance to be smaller than the anchor-negative distance by a margin of 0.2 [23]. The V1 training configuration specifies five epochs with a batch size of 32, AdamW (learning rate 1×10^{-4}; weight decay 1×10^{-4}), and an M-per-class sampler

($m = 4$). Images were resized to 224 × 224 pixels and augmented by horizontal/vertical flips, rotation, perspective transformation and color jitter. A longer development run appeared to lose separation, but this was not a matched stability experiment.

3.9. Proposed Re-Identification Model: V2

The proposed V2 model keeps the ImageNet-pretrained Swin-T backbone but replaces the output head and objective function. The backbone output is reduced from 768 feature dimensions to a 256-dimensional raw embedding. A BNNeck layer then produces a batch-normalized embedding; its bias parameter is present but frozen during training [14]. The pre-BN embedding is passed to Triplet and Center loss, while the BNNeck output is passed to ArcFace [13,14,28]. The TripletMarginLoss implementation applies its default L2-normalized Euclidean distance, and inference applies L2 normalization before cosine comparison. The architecture of the V2 re-identification head is summarized in Figure 1.

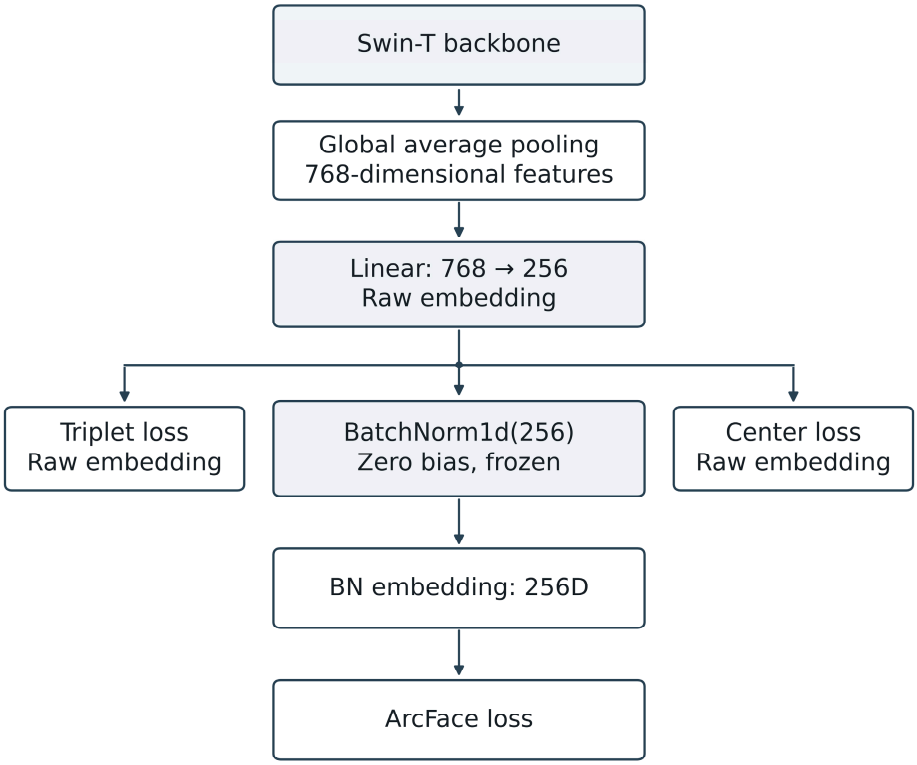


Figure 1. V2 re-identification head. The raw 256-dimensional embedding is used for Triplet and Center loss; the BNNeck output is used for ArcFace. The BatchNorm bias is initialized to zero and frozen. Inference separately L2-normalizes the BNNeck output.

The total training objective is

$$L_{Total} = 1.0L_{ArcFace} + 1.0L_{Triplet} + 0.005L_{Center} \quad (1)$$

where the coefficients are dimensionless loss weights.

ArcFace uses a dimensionless scale parameter, $s = 30$. In PyTorch metric learning, the configured margin ($m = 0.30$) is interpreted in degrees (approximately 0.0052 rad). Triplet loss uses a dimensionless embedding distance margin of 0.3 with hard mining. Center loss learns identity centroids, encourages within-class compactness and is weighted by 0.005. V2 was configured for up to 30 epochs with early-stopping patience of 10, a batch size of 32, AdamW (base learning rate 3.5×10^{-4}; weight decay 1×10^{-4}), three warm-up epochs followed by cosine decay, an M-per-class sampler ($m = 4$), and gradient clipping

at 5.0. Its augmentations add random erasing to the V1 transforms. The temporal V2 run started from ImageNet-pretrained backbone weights, completed 16 epochs and selected the epoch-6 checkpoint using the minimum validation objective, which comprised ArcFace and Triplet losses and excluded Center loss. The run stopped after ten further epochs without improvement. The individual contributions of these settings were not isolated in the present experiment.

3.10. Hybrid Spatial–Visual Identity Association

The final association layer combines spatial and visual information. ByteTrack maintains short-term trajectories, while the re-identification model provides cosine similarities between current fish crops and stored identity galleries. Candidate galleries exclude Bio IDs already mapped to visible tracks at the start of the current frame. A non-negative spatial bonus then favors galleries last observed near the current position; it does not impose a hard spatial rejection rule. Before a crop was admitted to the identity gallery, quality filters were applied to reduce unreliable identity updates: V2 rejects crops whose boxes overlap another detection, requires an aspect ratio of 1.5–3.5 and Laplacian variance above 150, and aggregates three accepted crops (V1 uses five); accepted crops need not be consecutive. Candidate crops were assigned to an existing biometric identity only when the hybrid similarity score exceeded 0.75. This score adds cosine similarity to a spatial bonus that decreases linearly from 0.2 at zero displacement to 0 at 300 pixels. When a newly observed track was matched to an existing gallery, the gallery embedding was updated by a normalized exponential moving average, 0.9 old plus 0.1 new. These thresholds are development settings and were not calibrated on an independent set.

The association settings listed above were retained from development. The architecture and these settings were developed using the same two recordings before the present temporal experiment. Accordingly, the new test isolates frames from final retraining and checkpoint selection, but is not an independent dataset or a test of previously unseen identities.

3.11. Evaluation Metrics

The primary evaluation compares the ByteTrack IDs and Bio IDs stored for the same detections on the same test frames. The Re-ID stage does not feed its identity assignments back into ByteTrack, so the raw ByteTrack column provides the matched tracker-only reference. Within each frame, detections and AAU boxes were matched one-to-one, maximizing the number of pairs above an intersection-over-union (IoU) threshold and then their total IoU. The primary threshold is 0.3; an additional analysis uses a threshold of 0.5. Each predicted identity is assigned a fixed ground-truth owner at its first qualifying match, rather than being remapped to the most frequent fish after observing the full sequence.

For whole-sequence identification, the first correctly owned Bio ID for each AAU Object ID defines its reference identifier. Every ground-truth-matched output is classified as the correct reference ID, a different ID owned by the same fish (fragmentation), an ID owned by another fish, or Scanning. Percentages use all matched outputs for that fish, including Scanning and initial enrollment. Because ownership is initialized from the first ground-truth match, this is an evaluation of identity consistency after label alignment, not independent recognition of an externally named fish. Ground-truth observations without a qualifying pipeline match are reported separately; these include intentionally hidden fish and cannot be interpreted as detector misses alone.

A shared recovery opportunity requires a correctly owned numeric identity before occlusion in both outputs. The reference is the most recent such identity for each output. We score the first qualifying numeric identity within the post-emergence decision window as

follows: correct recovery of the reference, an incorrect identity belonging to another fish, or a different identity for the same fish (fragmentation). An event is classified as unresolved if no qualifying numeric identity is produced within its decision window. Missing detections, nonmatching boxes and "Scanning..." remain failures to provide a recoverable identity; they are not removed from the eligible-event denominator. Rates equal correct recoveries divided by all shared eligible opportunities. The event rules and IoU thresholds were specified retrospectively for the saved runs, without changing model weights. An earlier correctly owned reference is retained even when the latest pre-entry output has already switched to a wrong owner; these contaminated pre-event states are explicitly marked in the event table.

An identity-switch count here is the number of changes between successive numeric predicted labels matched to the same AAU Object ID, in chronological order, retaining the preceding label across unmatched or "Scanning..." observations. These label-transition counts are reported with assignment coverage and are not presented as a complete IDF1, HOTA or MOTA evaluation. Unassigned outputs marked Scanning are reported explicitly. Counts and rates are given by recording, by AAU Object ID and overall.

For failure analysis, we replayed the original crop-quality gates and three-crop buffer using the original test images with the same mask and the archived tracked boxes, without rerunning or changing the tracker or Re-ID model. This reproduces which stored outputs reach an association decision or remain in Scanning mode. Logged hybrid scores and track-to-Bio-ID assignments were then inspected to distinguish rejected crops, insufficient accepted crops, duplicate enrollment and propagation of an already assigned identity. It is a retrospective trace of the saved run, not a counterfactual test of alternative thresholds. Supplementary Material S2 provides the frame-level trace and detector test evaluation, performed on CPU with PyTorch 2.11.0 and Ultralytics 8.4.37.

Historical embedding diagnostics summarized the original re-identification models using Rank-1 retrieval consistency (the first rank of the Cumulative Matching Characteristic, CMC), retrieval mean Average Precision (mAP), cosine similarity matrices, PCA visualization and within-/between-folder similarity distributions. Let S_intra be the set of cosine similarities between embeddings from the same predicted Bio ID folder, and let S_inter be the set of cosine similarities between embeddings from different predicted Bio ID folders. The separation gap is a dimensionless cosine similarity difference and is defined as

$$G = mean(S_{intra}) - mean(S_{inter}) \quad (2)$$

A larger gap indicates stronger separation between the predicted folders, which need not be correct fish identities. For normalized high-dimensional embeddings, a between-folder cosine similarity close to zero indicates that different predicted groups occupy nearly orthogonal directions. The V2 retrieval diagnostic followed a leave-one-out protocol on 44 detector/tracker-derived fragments: each fragment was used once as a query, while the remaining fragments formed the gallery. Rank-1 measured whether the closest retrieved fragment belonged to the same predicted Bio ID folder, and retrieval mAP measured the ranked list across all same-folder matches. The V1 diagnostic used 29 fragments from a separately generated proof directory, so the V1 and V2 summaries are descriptive development diagnostics rather than a paired evaluation on a shared query set. Moreover, the fragments were selected and grouped by the same pipeline subsequently assessed; the analysis is therefore subject to circularity and selection bias. It measures internal assignment consistency, not ground-truth identity accuracy.

The original development experiments were run under Ubuntu 24.04 using Python 3.12.3, PyTorch 2.11.0+cu130, torchvision 0.26.0, Ultralytics 8.4.37, PyTorch metric learning 2.9.0 and OpenCV 4.13.0 on an NVIDIA GeForce GTX 1650 GPU with 4 GB VRAM and 6.7 GB system RAM. The original training process on the combined data from both

recordings took approximately 3 h. Generating the original two demonstration videos took approximately 10–30 min in total on this system. These are approximate wall clock observations rather than controlled benchmarks, and the current use case is an offline analysis of recorded video.

Reported numerical results were computed from archived experimental outputs and ground-truth annotations, supplemented by evaluation of the frozen detector checkpoint on the original test frames and replay of crop-quality decisions; no synthetic experimental observations were introduced.

4. Results

On the 202-frame temporal validation set, the final detector training log reported a precision of 0.98679, a recall of 0.98613, an mAP@0.50 of 0.99458 and an mAP@0.50:0.95 of 0.92264. On the 426-frame held-out test set, precision, recall, mAP@0.50 and mAP@0.50:0.95 were 0.98878, 0.96515, 0.98384, and 0.93119, respectively. These new detector test results use unmasked frames and therefore do not establish detection performance behind the artificial obstacle. The temporal V2 training reached its minimum ArcFace-plus-Triplet validation objective of 0.0420 at epoch 6 and stopped at epoch 16.

4.1. Historical Development Diagnostics

The separately generated V1 and V2 diagnostics produced different predicted-folder separation and internal retrieval summaries. Table 1 reports these historical descriptive development results, not performance on the new temporal test. Because the configurations differ in several settings and were not evaluated on a shared query set, the comparison does not isolate the effect of any single component.

Table 1. Historical descriptive results for the original V1 and V2 development configurations; retrieval diagnostics used separate predicted Bio ID sets and are not temporal test results. Cosine similarities are reported as mean ± standard deviation across distinct fragment pairs, not variability across training runs.

Metric	V1 Baseline	V2 BNNeck	Interpretation
Embedding dimension	128-D feature vector	256-D feature vector	Different configurations
Training schedule	5 epochs	Up to 30 epochs; patience 10	Different schedules
Training objective	Triplet loss	ArcFace + Triplet + Center loss	Loss values are not comparable
Internal Rank-1 consistency	72.4%	97.7%	Separate predicted Bio ID sets
Internal retrieval mAP	78.6%	96.0%	Separate predicted Bio ID sets
Within-folder cosine similarity (dimensionless)	0.9105 ± 0.10	0.8956 ± 0.19	Descriptive single-run values
Between-folder cosine similarity (dimensionless)	0.7114 ± 0.12	0.0365 ± 0.24	Descriptive single-run values
Separation gap (dimensionless)	0.1991	0.8592	Predicted-folder separation
Gallery entries in historical runs	3 gallery entries	3 gallery entries	Not ground-truth-verified identities

Across the separately generated diagnostics, the most notable numerical difference was the reduction in between-folder similarity. V1 produced high within-folder similarity, but different predicted Bio ID folders also remained similar. Mean between-folder cosine similarity was 0.7114 for V1 and 0.0365 for V2, producing separation gaps of 0.1991 and

0.8592, respectively. These are descriptive single-run values from different fragment sets and do not identify which simultaneous V2 design change produced the difference.

4.2. Retrieval Performance

V2 produced 97.7% Rank-1 consistency using predicted Bio ID folder labels. Out of 44 image-query fragments, 43 fragments had a first retrieved match from the same folder. The retrieval mAP was 96.0%, showing strong ranked consistency within this diagnostic set. The separately generated V1 diagnostic produced 72.4% Rank-1 consistency and 78.6% retrieval mAP. For three balanced identities, chance-level Rank-1 is approximately 33.3%; for the V2 folder sizes of 12/15/17 fragments, the empirical random same-folder probability is 32.5%. Because the labels were generated by the same system and the two configurations used different fragment sets, these values are exploratory self-consistency diagnostics rather than paired or ground-truth identity-recovery accuracy.

4.3. Embedding-Space Structure

The results of the embedding-space analysis are important diagnostic evidence for this experiment. V2 produced a discriminative structure among the predicted Bio ID folders, with high within-folder similarity and low between-folder similarity. The cosine similarity statistics and separation gap analysis indicate that the numerical difference reflects the geometry of these internal groups; they do not by themselves establish correspondence between Bio IDs and AAU Object IDs.

To make this embedding geometry explicit, three complementary visual diagnostics are shown in Figures 2–4. The cosine similarity matrices (Figure 2) show both V1 and V2 on one common color scale, indicating similarity within and between predicted Bio ID folders that remain separated. The PCA projection (Figure 3) provides a two-dimensional view of the embedding spaces, while the within-/between-folder similarity distributions (Figure 4) show the corresponding ranges. Together, these diagnostics describe the internal structure of separately generated V1 and V2 fragment collections rather than ground-truth identity accuracy.

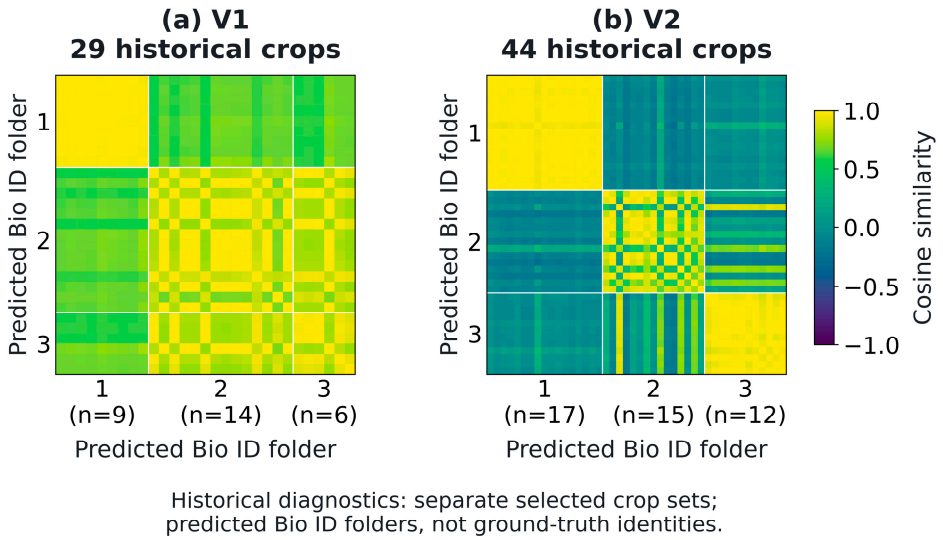


Figure 2. Historical cosine similarity matrices for V1 (29 fragments) and V2 (44 fragments), grouped by their separately predicted Bio ID folders. Both panels use the same color scale from −1 to 1. These internal diagnostics are not a paired ground-truth identity evaluation.

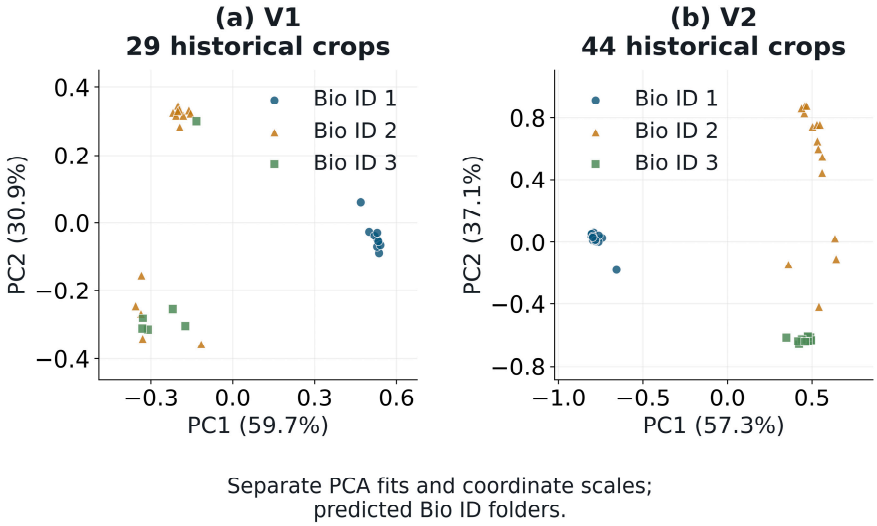


Figure 3. Historical PCA projections of V1 and V2 embeddings, colored by the predicted Bio ID folder. PCA is fitted separately in each embedding space; the axes are not a common coordinate system. Sample sets contain 29 and 44 fragments, respectively.

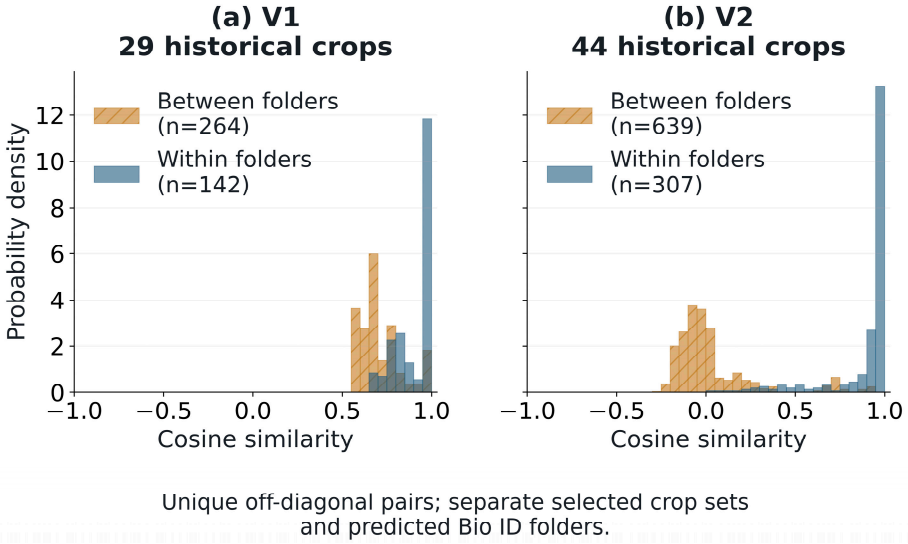


Figure 4. Historical within-folder and between-folder cosine similarity distributions for V1 and V2, using common bins and axes. Group labels are predicted Bio IDs. The hybrid association threshold is not a pure-cosine threshold and is not shown as a decision boundary.

4.4. Training Stability

The historical V1 configuration specifies five epochs, whereas V2 allows up to 30 epochs with early-stopping patience of ten. In the temporal V2 run, training stopped after 16 epochs, and the checkpoint selected by validation was from epoch 6. These observations do not establish that any individual V2 component improves stability: embedding size, objective, optimization and augmentation also differ, and matched multi-seed ablations were not performed.

4.5. Whole-Sequence Individual Identification

Across both test segments, 712 of 1278 ground-truth fish observations had a qualifying match to a stored pipeline output at IoU ≥ 0.30. Of these matched outputs, 298 (41.9%) retained the fish's reference Bio ID, six (0.8%) used a different Bio ID for the same fish, 18 (2.5%) used an ID owned by another fish and 390 (54.8%) remained in Scanning mode

(Table 2). Thus, the 45.2% numeric-ID coverage is distinct from correct identification. Results for IoU ≥ 0.50 and the complete Bio-ID-to-ground-truth correspondence are provided in Supplementary Material S2.

Table 2. Whole-sequence Bio ID outcomes at IoU ≥ 0.30. AAU IDs 1–3 belong to Vid1 and IDs 4–6 to Vid2.

AAU ID	Matched (N)	Correct ID N (%)	Different ID N (%)	Wrong Fish N (%)	Scanning N (%)	No GT Match (N)
1	173	149 (86.1)	0 (0.0)	0 (0.0)	24 (13.9)	35
2	118	50 (42.4)	6 (5.1)	0 (0.0)	62 (52.5)	90
3	147	3 (2.0)	0 (0.0)	18 (12.2)	126 (85.7)	61
4	131	52 (39.7)	0 (0.0)	0 (0.0)	79 (60.3)	87
5	47	25 (53.2)	0 (0.0)	0 (0.0)	22 (46.8)	171
6	96	19 (19.8)	0 (0.0)	0 (0.0)	77 (80.2)	122
Total	712	298 (41.9)	6 (0.8)	18 (2.5)	390 (54.8)	566

Percentages use the matched count in each row. Correct ID denotes the first correctly owned Bio ID for that fish; Different ID denotes another Bio ID with the same fixed owner. No GT match is a separate count across all test frames, including artificial occlusion. Initial owner-defining observations are included; labels are not reassigned by majority vote over the test sequence.

4.6. Identity Recovery on Temporally Held-Out Segments

The final evaluation covered 208 test frames from Vid1 and 218 from Vid2. The strict episode definition yielded 23 complete episodes containing full disappearance. Five lacked a qualifying pre-occlusion identity for at least one output, leaving 18 shared recovery opportunities: six in Vid1 and twelve in Vid2. Re-ID correctly restored the prior identity in 1/6 episodes in Vid1 and 2/12 in Vid2, or 3/18 overall (16.7%). All three correct recoveries bridged changed ByteTrack IDs, demonstrating successful identity reconnection in both recordings. It made one incorrect assignment and left 14 opportunities unresolved. ByteTrack recovered none of the 18 reference identities; its outcomes were two incorrect identities, eleven different identities for the same fish and five unresolved opportunities. Tables 3 and 4 report overall and per-identity outcomes. Changing the IoU threshold from 0.3 to 0.5 did not alter these primary counts. Under the alternative partial-reappearance definition, Re-ID succeeded in 6/24 recovery episodes (25.0%) versus 0/24 for ByteTrack; its other outcomes were one incorrect identity, one different identifier for the same fish and sixteen unresolved episodes. This sensitivity result is kept separate from the primary denominator. The corresponding identity-label transition counts were 11 for ByteTrack versus 2 for Bio IDs in Vid1, and 10 versus 0 in Vid2. Numeric Bio IDs were available for only 226/438 (51.6%) and 96/274 (35.0%) GT-matched detections, respectively. The remaining matched detections were Scanning outputs. These lower switch counts coexist with low identity-assignment coverage and do not imply near-perfect recovery. ByteTrack and ByteTrack plus Re-ID were compared on identical detection outputs and ground-truth events. The stored test outputs created four Bio ID galleries for Vid1 and three for Vid2; these counts are reported separately from identity correctness. A smaller number of output identities or switches is not sufficient evidence of correct recovery when outputs remain unassigned or two fish are merged into one identity. Table 5 reports identity switches together with assignment coverage, and Figure 5 summarizes the event-level outcomes.

Table 3. Ground-truth recovery outcomes on shared pre-enrolled events at IoU ≥ 0.30.

Definition	N	Output	Correct	Wrong Fish	Different ID	No Decision	Correct (%)
Full emergence	18	ByteTrack	0	2	11	5	0.0%
Full emergence	18	+Re-ID	3	1	0	14	16.7%
Partial reappearance	24	ByteTrack	0	2	15	7	0.0%
Partial reappearance	24	+Re-ID	6	1	1	16	25.0%

Correct means restoration of the most recent correctly owned pre-event ID. Wrong fish means an ID first owned by another AAU Object ID. Different ID means another identifier for the same fish. All outcomes use the same event denominator for both outputs; unresolved cases are retained.

Table 4. Primary full-emergence recovery by ground-truth fish identity.

AAU ID	N	ByteTrack Correct	Re-ID Correct	Re-ID Wrong Fish	Re-ID Different ID	Re-ID No Decision
1	1	0/1 (0%)	1/1 (100.0%)	0	0	0
2	2	0/2 (0%)	0/2 (0.0%)	0	0	2
3	3	0/3 (0%)	0/3 (0.0%)	1	0	2
4	4	0/4 (0%)	1/4 (25.0%)	0	0	3
5	4	0/4 (0%)	1/4 (25.0%)	0	0	3
6	4	0/4 (0%)	0/4 (0.0%)	0	0	4

Table 5. Identity switches and Bio ID assignment coverage at IoU ≥ 0.30.

Video	GT-Matched Detections	Numeric Bio ID	Scanning	ByteTrack Switches	Bio ID Switches
Vid1	438	226 (51.6%)	212	11	2
Vid2	274	96 (35.0%)	178	10	0

The exported test logs also recorded seven operations that reattached a new ByteTrack tracklet to an existing Bio ID. Six operations had a qualifying one-to-one ground-truth match at the assignment frame, and all six agreed with the fixed identity ownership. The seventh did not meet the primary IoU threshold at that frame and was not counted as a verified correct assignment. This conditional check of executed reattachments is separate from event-level recovery: it excludes opportunities in which no reattachment was made.

Table 3 summarizes both event definitions. For the partial-reappearance analysis, 26 complete full-containment runs yielded 24 shared eligible events after excluding two initial-registration cases. The outcome categories were unchanged at IoU 0.5 for both definitions.

Table 4 separates the primary outcomes by AAU Object ID. Initial-registration cases are counted separately and are not represented as failures to recover a previously established identity.

AAU Object IDs 1–3 occur in Vid1 and 4–6 in Vid2. Vid1 contributes six eligible episodes and Vid2 contributes twelve. Detailed ByteTrack failure categories and all initial-registration cases are included in Supplementary Material S1.

Table 5 reports chronological identity-label changes and the fraction of matched detections carrying a numeric Bio ID. Coverage is conditional on a GT-matched detection and is not a whole-sequence detection recall.

Switches count changes between successive numeric labels matched to the same AAU Object ID, carrying the preceding label across unmatched or Scanning observations. This definition yields the same switch totals at IoU ≥ 0.50. No complete tracking benchmark or 100% recovery claim is inferred from these counts.

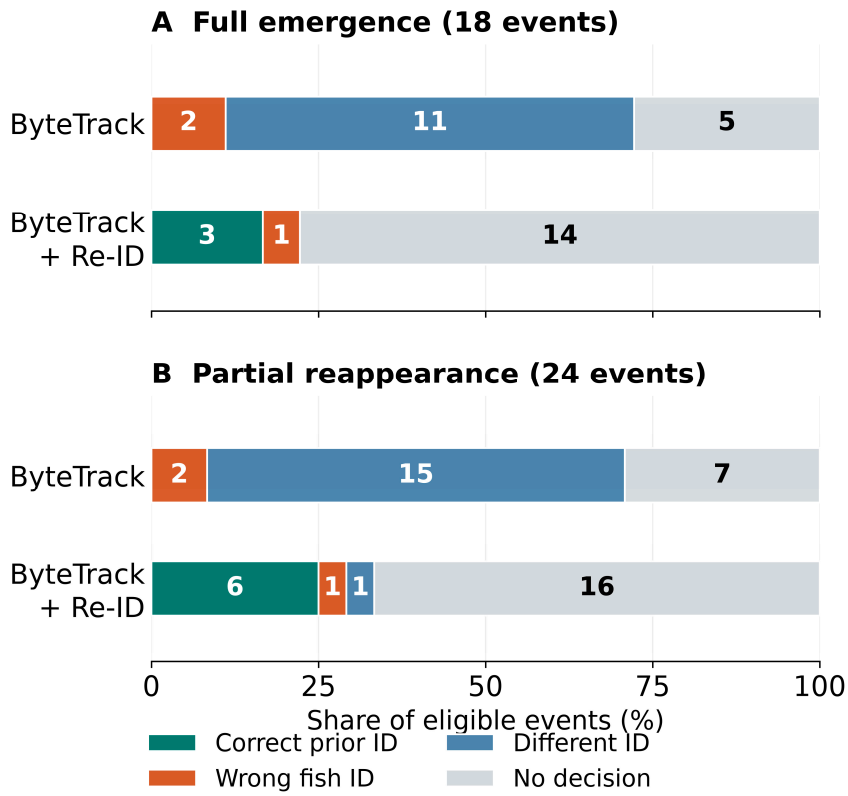


Figure 5. Ground-truth event outcomes for ByteTrack and ByteTrack plus Re-ID. Panel A uses the primary full-emergence definition (18 shared eligible episodes); panel B uses partial reappearance after each fully hidden run (24 shared eligible episodes). Labels are event counts. Different ID denotes fragmentation of an already enrolled fish. No decision includes Scanning and missing qualifying detections. Initial-registration episodes are reported separately in Supplementary Material S1.

4.7. Failure Analysis of the Temporal Test

The quality-gate replay reproduced the numeric-versus-Scanning state for all 749 logged outputs without discrepancies. Among the 390 ground-truth-matched Scanning outputs, 355 (91.0%) were rejected by at least one crop-quality gate and 35 (9.0%) passed the gates but had not yet accumulated three accepted crops for that track. Low sharpness occurred in 328 of the 355 rejected crops, sometimes together with overlap or an invalid aspect ratio; these overlapping causes are tabulated in Supplementary Material S2. A failed hybrid-score threshold is not itself a Scanning outcome: once three crops are available, the implementation creates a new gallery if no eligible score exceeds 0.75.

Of the 14 unresolved full-emergence recovery opportunities, five had no qualifying post-emergence pipeline match, six had matched Scanning outputs throughout the decision window, and three combined Scanning with unmatched frames. The saved post-tracker outputs cannot separate a detector miss from tracker suppression or inadequate box overlap in the unmatched cases. Five further complete episodes lacked a qualifying pre-occlusion identity in at least one output and were reported separately from the 18 shared recovery opportunities.

The fourth gallery in Vid1 was duplicate enrollment of AAU Object ID 2: Bio ID 4 was created at frame 1052 with a logged best eligible hybrid score of 0.4516, below 0.75, although that fish already had reference Bio ID 3. It accounts for the six different ID outputs in Table 2. Conversely, all 18 wrong-fish assignments occurred when ByteTrack ID 3 carried Bio ID 1 from AAU Object ID 1 to Object ID 3 at frames 956–973. These outputs inherited the existing track-to-Bio-ID mapping rather than executing a fresh gallery match. The failure therefore includes propagation of an incorrect tracker association as well as duplicate

enrollment; counting galleries or executed reattachments alone would result in these errors being missed.

4.8. Baseline Failure Modes and Model-Selection Rationale

Two preliminary observations motivated the transition from the Triplet-only baseline to the BNNeck angular-margin model. First, grayscale preprocessing appeared to reduce available chromatic and texture cues and encourage pose-oriented grouping rather than identity-oriented separation. Second, a longer Triplet-only V1 development run appeared to lose separation. These observations were not obtained from matched ablations and do not establish the effect of an individual component. RGB-versus-grayscale comparisons and component-specific experiments under otherwise identical settings are planned for subsequent work.

5. Discussion

The temporal evaluation shows that the added hybrid association stage can reconnect known zebrafish after tracker interruption: Re-ID restored the prior identity in three of eighteen shared full-emergence opportunities, whereas ByteTrack alone restored none. Each success bridged a changed ByteTrack ID, with successes in both recordings. However, fourteen opportunities remained unresolved and one was assigned to the wrong identity. The whole-sequence analysis further shows that low switch counts coexist with extensive Scanning, duplicate enrollment and propagation of an incorrect track identity. The comparison evaluates the association stage as a whole and does not isolate embeddings from the spatial bonus, gallery or quality gates. Recovery rates also depend on the event boundary and decision window. The historical V1/V2 diagnostics remain descriptive results from different fragment sets; they establish neither matched V2 superiority nor a causal benefit of an individual component.

The aquaculture relevance of the study lies in the transition from image-level observation to identity-linked longitudinal monitoring. Disease-detection, species-recognition and morphology-estimation systems address valuable but different tasks and need not preserve individual identity [1–5,17,20–22]. In contrast, a reliable re-identification layer allows observations to be accumulated for the same fish after occlusion, temporary disappearance or ambiguous crossings. This matters because the practical value of recognizing one specific fish is not limited to the act of recognition itself. Persistent identity can become the indexing mechanism for unique-fish counting and later records of feeding response, growth trajectory, body-size change, swimming activity, preferred tank regions, social interactions, visible disease signs, welfare, treatment response and selective breeding. It could also link visual phenotype to physiological or microbiome measurements [29]. In this sense, the present system should be viewed as the first layer of a broader individual-level monitoring framework for aquaculture and laboratory fish research.

The relationship between the proposed framework and existing tracking methods also clarifies the contribution of the study. Geometry-based trackers such as SORT and ByteTrack are efficient and useful for short-term continuity [10,11], especially when detections are available in consecutive frames. However, they are not designed to recover identity from visual evidence after complete disappearance or strong ambiguity. Multi-animal pose frameworks such as DeepLabCut and SLEAP provide powerful markerless tools for behavior analysis [15,16], but pose estimation and individual re-identification are not equivalent tasks. In visually homogeneous fish, body landmarks and short-term motion may remain ambiguous after occlusion. The proposed detector–tracker–re-identification pipeline therefore complements spatial tracking by adding a visual identity signal that can be used when trajectory continuity alone is insufficient.

The failure trace suggests concrete priorities for subsequent work. The fixed sharpness and geometry gates should be calibrated on independent validation data, and quality-weighted crop aggregation or preservation of accepted crops across short track interruptions should be tested against the current three-crop rule. Periodic checks of already assigned track identities could address the observed propagation error; duplicate-gallery checks could address repeated enrollment of one fish. These changes should be evaluated for both coverage and wrong-fish assignments, since simply relaxing rejection thresholds could trade abstention for identity errors. They are proposed interventions, not tested improvements. Matched component ablations, repeated training seeds and standard tracking metrics such as IDF1 and HOTA remain further work.

The study remains a controlled feasibility experiment rather than a population-level validation. The occlusion was simulated rather than created by a physical object in the tank, and generalization to different cameras, lighting conditions, fish ages, group sizes, water turbidity and aquaculture species remains to be tested. The original demonstration videos were generated in approximately 10–30 min in total on the GTX 1650 system; this approximate development observation was not repeated as a controlled temporal test benchmark. The exact visual cues learned by the model are also not yet fully explained. Future work will extend validation to larger identity sets, independent recordings, real occlusion, cross-tank testing and explainability analysis such as attention visualization or occlusion sensitivity. These limitations define the proper scope of the present results and the next stage of individual-level aquaculture analytics.

6. Conclusions

This study presents an experimental workflow for markerless re-identification of visually homogeneous zebrafish. The pipeline combines YOLO26n detection, ByteTrack short-term tracking and a Swin-T embedding model trained with BNNeck, ArcFace, Triplet and Center loss. The final detector and V2 model were retrained on temporally ordered blocks and evaluated on 426 held-out frames covering both recordings and all six known identities. Under the primary full-emergence definition, Re-ID restored the prior identity in 3 of 18 eligible recovery episodes (16.7%), compared with none for ByteTrack alone, but made one incorrect assignment and left fourteen opportunities unresolved. All three correct recoveries bridged changed ByteTrack IDs, with successes in both recordings. The partial-reappearance analysis gave 6/24 (25.0%) versus 0/24, with one incorrect identity, one different identifier for the same fish and sixteen unresolved episodes. These verified cases establish that the added association stage can restore identity after tracker interruption in this controlled setting; improving the frequency of successful recovery remains necessary. The ground-truth evaluation distinguishes successful recovery, incorrect assignments and unresolved identities; the historical predicted-folder Rank-1 and retrieval mAP values are auxiliary diagnostics. The present results concern identity recovery within two controlled recordings and do not establish recognition of unseen fish or transfer to other sessions and aquaculture environments.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/aquacj6040052/s1, Supplementary Material S1 contains frame split lists, temporal test predictions, one-to-one ground-truth matches, complete event tables, evaluation code, historical embedding arrays and figure files. Supplementary Material S2 adds frozen-checkpoint detector test results, per-fish whole-sequence identity outcomes, crop-quality replay records and reproducible analysis scripts.

Author Contributions: Conceptualization, G.A. and Z.Z.; methodology, Z.Z. and M.I.; software, M.I.; validation, G.A., M.I. and Z.Z.; formal analysis, G.A., M.I. and Z.Z.; investigation, G.A., K.M., M.I. and Z.Z.; resources, G.A.; data curation, G.A. and M.I.; writing—original draft preparation, G.A. and K.M.; writing—review and editing, Z.Z., G.A. and K.M.; visualization, M.I.; supervision, G.A. and Z.Z.; project administration, G.A. and K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: As this manuscript does not involve any research on animals or human subjects, ethical approval from an Institutional Review Board was not required.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source video frames and annotations are available from the AAU Zebrafish Re-Identification Dataset on Kaggle: https://www.kaggle.com/datasets/aalborguniversity/aau-zebrafish-reid (accessed on 26 June 2026). The source code is available at https://github.com/martonikon/fish-recognition. The original V1 and V2 development demonstration videos are available at https://drive.google.com/file/d/1RvDP7i9uyYpyFAMh_mRg1RaAYnzXXeO4/view (accessed on 26 June 2026) and https://drive.google.com/file/d/1dZNDSClTywWWITT0SD5JeiWuIr0KJDAI/view (accessed on 26 June 2026), respectively. The frame lists, temporal test predictions and original evaluation outputs are included in Supplementary Material S1. The additional analyses reported during the peer review process are provided in Supplementary Material S2.

Acknowledgments: The authors acknowledge Aalborg University and the creators of the AAU Zebrafish Re-Identification Dataset for making the annotated zebrafish video material publicly available. During preparation of this manuscript, ChatGPT with Codex (OpenAI; GPT-6.1 Sol; accessed September–October 2026) was used to assist with evaluation and visualization code. The authors reviewed and edited the assisted material and take responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CMC	Cumulative Matching Characteristic
CNNs	Convolutional Neural Networks
mAP	Mean Average Precision
PCA	Principal Component Analysis
RFID	Radio-Frequency Identification
SLEAP	Social LEAP Estimates Animal Poses
V1	Triplet-Only Baseline
V2	BNNeck Angular-Margin Model
YOLO	You Only Look Once

References

1. Barbedo, J.G.A. A Review on the Use of Computer Vision and Artificial Intelligence for Fish Recognition, Monitoring, and Management. *Fishes* **2022**, *7*, 335. [CrossRef]
2. Li, D.; Li, X.; Wang, Q.; Hao, Y. Advanced Techniques for the Intelligent Diagnosis of Fish Diseases: A Review. *Animals* **2022**, *12*, 2938. [CrossRef]
3. Petrellis, N. Measurement of Fish Morphological Features through Image Processing and Deep Learning Techniques. *Appl. Sci.* **2021**, *11*, 4416. [CrossRef]
4. Pauzi, S.N.; Hassan, M.G.; Yusoff, N.; Harun, N.H.; Abu Bakar, A.H.; Kua, B.C. A Review on Image Processing for Fish Disease Detection. *J. Phys. Conf. Ser.* **2021**, *1997*, 012042. [CrossRef]
5. Ahmed, M.S.; Jeba, S.M. SalmonScan: A Novel Image Dataset for Machine Learning and Deep Learning Analysis in Fish Disease Detection in Aquaculture. *Data Brief.* **2024**, *54*, 110388. [CrossRef] [PubMed]
6. Ahmed, M.S.; Aurpa, T.T.; Azad, M.A.K. Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 5170–5182. [CrossRef]

7. Li, J.; Lian, Z.; Wu, Z.; Zeng, L.; Mu, L.; Yuan, Y.; Bai, H.; Guo, Z.; Mai, K.; Tu, X.; et al. Artificial Intelligence-Based Method for the Rapid Detection of Fish Parasites (*Ichthyophthirius multifiliis*, *Gyrodactylus kobayashii*, and *Argulus japonicus*). *Aquaculture* **2023**, *563*, 738790. [CrossRef]
8. Kumar, N.; Maree, R.; Geurts, P.; Muller, M. Recent Advances in Bioimage Analysis Methods for Detecting Skeletal Deformities in Biomedical and Aquaculture Fish Species. *Biomolecules* **2023**, *13*, 1797. [CrossRef]
9. Pati, N.; Sharma, A.; Jena, J.J.; Gourisaria, M.K.; Bandyopadhyay, A.; Castro, E.L. XAI Based Enhanced Fish Disease Detection Using Multi-Feature Extraction in Varied Color Spaces. *Procedia Comput. Sci.* **2025**, *258*, 2909–2919. [CrossRef]
10. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple Online and Realtime Tracking. In Proceedings of the IEEE International Conference on Image Processing (ICIP 2016), Phoenix, AZ, USA, 25–28 September 2016; pp. 3464–3468. [CrossRef]
11. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In Proceedings of the European Conference on Computer Vision (ECCV 2022), Tel Aviv, Israel, 23–27 October 2022; pp. 1–21. [CrossRef]
12. He, S.; Luo, H.; Wang, P.; Wang, F.; Li, H.; Jiang, W. TransReID: Transformer-Based Object Re-Identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021), Montreal, QC, Canada, 11–17 October 2021; pp. 15013–15022. [CrossRef]
13. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA, USA, 15–20 June 2019; pp. 4690–4699.
14. Luo, H.; Gu, Y.; Liao, X.; Lai, S.; Jiang, W. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops 2019), Long Beach, CA, USA, 16–20 June 2019. [CrossRef]
15. Lauer, J.; Zhou, M.; Ye, S.; Menegas, W.; Schneider, S.; Nath, T.; Rahman, M.M.; Di Santo, V.; Soberanes, D.; Feng, G.; et al. Multi-Animal Pose Estimation, Identification and Tracking with DeepLabCut. *Nat. Methods* **2022**, *19*, 496–504. [CrossRef] [PubMed]
16. Pereira, T.D.; Tabris, N.; Matsliah, A.; Turner, D.M.; Li, J.; Ravindranath, S.; Papadoyannis, E.S.; Normand, E.; Deutsch, D.S.; Wang, Z.Y.; et al. SLEAP: A Deep Learning System for Multi-Animal Pose Tracking. *Nat. Methods* **2022**, *19*, 486–495. [CrossRef] [PubMed]
17. Rahman, W.; Rahman, M.M.; Mozumder, M.A.I.; Sumon, R.I.; Chelloug, S.A.; Alnashwan, R.O.; Muthanna, M.S.A. A Deep CNN-Based Salinity and Freshwater Fish Identification and Classification Using Deep Learning and Machine Learning. *Sustainability* **2024**, *16*, 7933. [CrossRef]
18. Haurum, J.B.; Karpova, A.; Pedersen, M.; Bengtson, S.H.; Moeslund, T.B. Re-Identification of Zebrafish Using Metric Learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1–11. [CrossRef]
19. Aalborg University. AAU Zebrafish Re-Identification Dataset. Kaggle. 2020. Available online: https://www.kaggle.com/datasets/aalborguniversity/aau-zebrafish-reid (accessed on 26 June 2026).
20. Ergun, E. SwinFishNet: A Swin Transformer-Based Approach for Automatic Fish Species Classification Using Transfer Learning. *PLoS ONE* **2025**, *20*, e0322711. [CrossRef] [PubMed]
21. Rajbongshi, A.; Shakil, R.; Akter, B.; Lata, M.A.; Joarder, M.M.A. A Comprehensive Analysis of Feature Ranking-Based Fish Disease Recognition. *Array* **2024**, *21*, 100329. [CrossRef]
22. Haddad, M.H.; Mohammed, F.H. A Convolutional Neural Network Approach for Precision Fish Disease Detection. *Evol. Stud. Imaginative Cult.* **2024**, *8.2*, 1018–1033. [CrossRef]
23. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
24. Thilakarathna, M.W.S.N.; Avsar, E.; Nielsen, M.M.; Pedersen, M. Towards Visual Re-Identification of Fish Using Fine-Grained Classification for Electronic Monitoring in Fisheries. *Proc. 7th North. Light. Deep Learn. Conf.* **2026**, *307*, 428–438. [CrossRef]
25. Ziaei, M.M.; Urban, J.; Cisar, P. Image-Based Individual Fish Identification as a Substitute for Invasive Methods: State-of-the-Art and Research Gaps. *Rev. Aquac.* **2025**, *17*, e70078. [CrossRef]
26. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 10012–10022. [CrossRef]
27. Jocher, G.; Qiu, J.; Liu, M.; Lyu, S.; Akyon, F.C.; Kalfaoglu, M.E. Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models. *arXiv* **2026**, arXiv:2606.03748. [CrossRef]

28. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A Discriminative Feature Learning Approach for Deep Face Recognition. In Proceedings of the European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, 11–14 October 2016; pp. 499–515. [CrossRef]
29. Van Vu, S.; Kundu, S.; Woo, K.H.; Uttarotai, T.; Van Doan, H. Impact of Body Weight on the Intestinal Microbiome of Cage-Cultured Oyster Pompano (*Trachinotus anak*). *Anim. Genet.* **2026**, *57*, e70092. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.