scientific reports

Check for updates

OPEN

Zebrafish identification with deep CNN and ViT architectures using a rolling training window

Jason Puchalla[1,3]✉, Aaron Serianni[2,3] & Bo Deng[1]

Zebrafish are widely used in vertebrate studies, yet minimally invasive individual tracking and identification in the lab setting remain challenging due to complex and time-variable conditions. Advancements in machine learning, particularly neural networks, offer new possibilities for developing simple and robust identification protocols that adapt to changing conditions. We demonstrate a rolling window training technique suitable for use with open-source convolutional neural networks (CNN) and vision transformers (ViT) that shows promise in robustly identifying individual maturing zebrafish in groups over several weeks. The technique provides a high-fidelity method for monitoring the temporally evolving zebrafish classes, potentially significantly reducing the need for new training images in both CNN and ViT architectures. To understand the success of the CNN classifier and inform future real-time identification of zebrafish, we analyzed the impact of shape, pattern, and color by modifying the images of the training set and compared the test results with other prevalent machine learning models.

Keywords Machine learning, Convolutional neural networks, Vision transformer, Rolling window, Time evolving

Zebrafish (*Danio rerio*) are a primary model organism in a variety of biological and preclinical studies[1,2]. Their transparent embryos, rapid development, genomic similarity to humans, and ability to regenerate are just a few of the reasons for their growing use in laboratory research[3–8]. Several types of studies on zebrafish require identification of individuals. Though often possible, labeling individuals manually, either through surface markings or clipping tool patterns, presents substantial difficulty due to tissue regeneration, fish interactions, or changes in animals during maturation[7,9,10]. Electronic RF tagging is associated with a potentially significant loss of microtags and high mortality after several months[11,12]. In these cases, zebrafish are often removed from their storage tank, anesthetized[13] to provide a limited window of time for imaging, and moistened until reintroduced into the storage tank. Stripe color analysis using high-resolution, small field-of-view ($\approx 4mm^2$) images captured against a black plastic background has previously been reported as an effective classification method for adult male zebrafish[14]. Thus, there is a significant benefit to a simple and non-invasive identification system that requires minimal data and can be adapted to a variety of research situations and standard aquarium geometries.

In the past decade, the rise of machine learning and high-speed processing algorithms have opened the door to new and tunable identification approaches that have been applied to the challenging task of zebrafish tracking where occlusion and variable backgrounds can confound re-identification algorithms. Several open-source and commercially available tracking systems have been previously reported and reviewed[15–19]. These approaches leverage video motion to track the movement of individual zebrafish within a tank, often through capturing high-speed grayscale video from a top view, sometimes in limited water depth. Haurum et al. (2020) applied and evaluated statistical re-identification techniques to two side-view RGB videos of zebrafish captured on the same day from two tanks (3 fish per tank), using classical feature extractors[20]. The primary goal of this study was to demonstrate the potential of a custom statistical learning metric for restricted-volume, side-view zebrafish identification. In all tracking studies, substantial video data and continuous wide-field imaging are critical. However, tracking systems are designed primarily for identification tasks that require continuous imaging.

Some types of studies using zebrafish require the use of a procedure to identify individuals over an extended period of time in an experiment-specific environment. In this regard, classification and not re-identification is often a primary goal. Re-identification procedures typically do not require a fixed number of classes because the goal is not to classify into a set number of categories, but rather to identify if two images from different times,

[1]Department of Physics, Princeton University, Princeton, NJ 08544, USA. [2]Department of Mathematics, Princeton University, Princeton, NJ 08544, USA. [3]These authors contributed equally: Jason Puchalla and Aaron Serianni. ✉email: puchalla@princeton.edu

scenes, and even perspectives show the same individual[21–23]. On the other hand, in many laboratory settings, well-defined animal *classes* can be defined throughout the experiment using a fixed number of individuals and viewing angles. In general, the classification task is more robust to outliers and better utilizes all training data[24], two highly advantageous qualities when working with limited amounts of data. Oftentimes, the classification problem is used as part of the re-identification process[25].

Using classification methods, previous work has explored temporal changes within zebrafish that occur more slowly than swimming time scales. In 2017, Ishaq et al. reported results of a CNN using the AlexNet architecture that identified specific fish morphological changes during early development (e.g., missing head, partial tail, dark regions). The trained output was used to monitor changes in the induced morphology after neuronal damage in response to camptothecin, a chemical compound known to inhibit certain DNA enzymes[26]. Training on as few as 84 microwell plate images (before augmentation) achieved a test accuracy of 92.8%. Environmental temperature variations have also been noted as driving adaptation mechanisms that can affect zebrafish development and appearance over days or weeks[27]. More recently, Jones et al. (2023) described the classification of developmental delays in zebrafish embryos using a CNN where the number of training images used had a significant impact on the outcome[28]. However, the continuum of potential developmental states meant that some differences between populations could not be easily detected. There are potentially other areas of research that have yet to be explored, such as embryonic and adult zebrafish xenographic studies[29–32] that may benefit from deep learning techniques that can follow individual temporal changes. Although the relative importance of these changes will be specific to the type of study that is being carried out, in general, the ability to adapt to changes in data set features is an important subset of the emerging area of continuous learning[33,34].

Notably, transfer learning is one technique[35] that has been exploited[28] to facilitate adaptation and improve performance in classifying zebrafish embryos. Transfer learning involves adopting a set of architecture parameters developed for one task for a related but different task with different output classes. The method is particularly valuable when the available data for the new task is limited or when training time can be significantly reduced compared to starting from scratch. However, in some circumstances, it is possible that a rolling window[36] technique may be beneficial. This additional method allows for retraining on a small subset of data advancing over time to keep the model relevant as new data come in and old data become progressively outdated. This technique is better suited to maintain accuracy as data and training classes evolve in time but do not clearly involve new classes or require new tasks. Although both methods use new data to update model parameters, rolling windows stress temporal relevance, while transfer learning focuses on cross-domain applicability. Beyond feature drift due to maturation, studio lighting, environmental background, water conditions, and even camera setting can all contribute to significant dataset-specific variations over time that affect classification accuracy but will not alter the number of classes.

Computational neural networks (CNN) and vision transformers (ViT) architectures offer two distinct machine learning approaches and have become the preferred backbone for many classification studies. Here, we explore the benefit of applying a rolling window approach as a form of continuous learning to each model after pretraining using ImageNet[37]. We demonstrated that both InceptionV3[38] (CNN) and Vision Transformer[39] (ViT) architectures can be used as part of a robust identification protocol to classify free-swimming juvenile and adult zebrafish. Our study used a standard 38-liter commercial aquarium with a plastic imaging studio to help reduce the complexity inherent in 3D swimming environments, eliminate direct handling of fish, limit changes in perspective during imaging, and help normalize lighting conditions. This proof-of-concept study demonstrates the feasibility of achieving high-accuracy, individual, late-juvenile[2,40] zebrafish identification over three weeks using publicly available and compact deep learning libraries. To explore the effectiveness of the technique and better understand the nature of our CNN model, we also investigate the relative importance of pattern, color, and shape in this analysis.

Results

In situ zebrafish imaging studio

To collect zebrafish images used in the training data sets, an acrylic aquarium insert was built to allow the partitioning and photography of individual fish. The insert acted as an imaging studio, allowing fish to remain free-swimming while being sequentially cycled through a staging area by means of manually operated sliding partitions. To begin an imaging study, all aquarium decorations were removed from a portion of the tank and the studio was inserted close to one side of the tank. The studio insert was then moved to the other side of the tank to confine the whole fish population near a side wall. A single zebrafish was then allowed to pass into the studio by temporarily sliding open the studio panel nearest the side wall, while the other side panel remained closed. Once a fish swam into the studio, the open side panel was closed and the middle panel (oriented perpendicular to the side panels) was moved forward to limit the imaging depth of field to 5 mm and establish a uniform white photographic background. The photographs included views of each side of each fish, as presented during the imaging session. Figure 1A shows a single zebrafish isolated for imaging while the remaining population was sequestered on the far right. Both the camera and diffuse light source remained outside the tank and were stationary during an imaging session. After collecting sufficient images for the experiment on that day, the imaged fish was released to the far side of the tank through the other sliding side partition (Fig. 1B). The middle panel was then slid back, and the process was repeated until all the fish were imaged. The loading of a single fish for imaging typically took about 60-120 seconds.

For this proof-of-concept study, we collected 5 image data sets over 19 days (denoted as Days 1, 8, 12, 13, and 19). The five zebrafish used in this study were estimated to be between 8 and 10 weeks old when data collection began. Based on variations in size and coloration, the group included both male and female fish. A total of $N_{images} = 3441$ images were collected over five days, with a roughly equal number of images per fish. The 100×100 mm camera field of view allowed for capturing multiple free-swimming orientations without

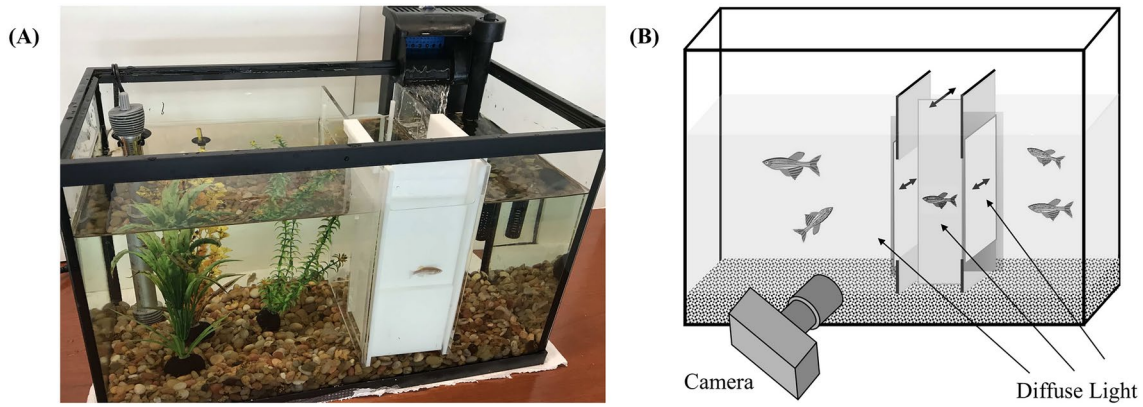

Fig. 1. The free-swimming zebrafish studio for acquiring study images. (**A**) The studio was inserted into a standard 38-liter aquarium and could be either removed or left in place post-imaging. (**B**) A schematic drawing showing the sliding direction of the studio panels and the placement of the camera and diffuse light source external to the tank.

the need to reposition the camera or studio. The raw images were cropped and padded using a custom semi-automated algorithm that roughly centered the fish in the cropped image. Lastly, the resulting images were scaled to 320×180 pixels before being added to the experiment image library (Fig. 2A). To illustrate the basic features displayed by the five fish during the course of the experiment, Fig. 2B shows each fish on Day 1 and Day 19 in a similar random pose. Relative length cannot be directly inferred from these images.

Deep learning approaches

Given the changing features of maturing zebrafish, there was an expectation that a model trained on one day would produce a lower accuracy when applied to images acquired on a subsequent day. In addition, the unknown rate of feature evolution meant that there was an unclear relationship between the number of images collected on a day and the fidelity of class matching from one day to the next. We hypothesized that regardless of the deep learning architecture used, a rolling window training method could help reduce the number of new images per class required after Day 1 to achieve high-fidelity cross-day class matching while still maintaining the desired 95% (or more) accuracy. We developed the following protocol and tested its performance using both a CNN and a vision transformer model.

CNN model with rolling window training

To assess the efficacy of a rolling-window analysis, we began by building a CNN classification scheme based on a modified version of Google's Inception V3[38] (see Methods). The first and last layers of the model were altered to fit our preprocessed image size of 320×180 pixels with an output vector of length 5. We initialized the network to a prior weighting matrix available from ImageNet[41] for InceptionV3[42]. This technique allowed for additional fine-tuning of the model in minutes rather than hours.

The preprocessed images were randomly split into training, validation, and testing sets using a 70%/15%/15% split. We used a Keras augmentation function (see Methods) to increase the size of the training set[43]. The same cropping and augmentation pipeline was followed for the ViT analysis. The model output for a given input image is the predicted class of a zebrafish individual. We measure the performance of the models using binary accuracy, comparing the model's predictions on the test set to the ground-truth class labels. Since neither the number of images needed to train the model and achieve $\geq 95\%$ same-day average classification accuracy nor the rate of time evolution was known *a priori*, our protocol required a best-guess number of images to collect on the first two days. The commonly cited rule of ten hueristic[44] suggested that there should be ten times the number of training images than the number of distinct classification features: at least 10 images per class. For this proof-of-concept study, we collected a minimum of 50 images per class for each of the five days to ensure that there were sufficient images to demonstrate the generality of the protocol and investigate the reproducibility of these results between any of the five days. The analysis described below was used to set the size of the rolling window and the minimum number of images to collect in the following days.

Figure 3A shows the classification accuracy of the same-day testing versus the number of training images per class for each of the five experiment days. Only 10 of the 50 training images per class (50 images total) on a typical day were needed to achieve the 95% average test accuracy for that day (Fig. 3A dashed). According to our protocol, only the first-day line (Fig. 3A; Day 1 line) was used to establish the baseline size of the first day training set on 10 images.

Next, we estimated the rate of evolution of the model features by measuring the cross-day test accuracy when trained using the 10 images of Day 1 and tested on all of the images of Day 8. The general evolution is demonstrated in Fig. 3B, where 10 images per class on a given day were used to train the model, after which images from other days were tested against that model. Train-test combinations using all five image data sets led to the measured cross-day average fractional accuracies and experimental uncertainties in Fig. 3B and underscored the temporal evolution of the fish. The average cross-day test fractional accuracy measured for one

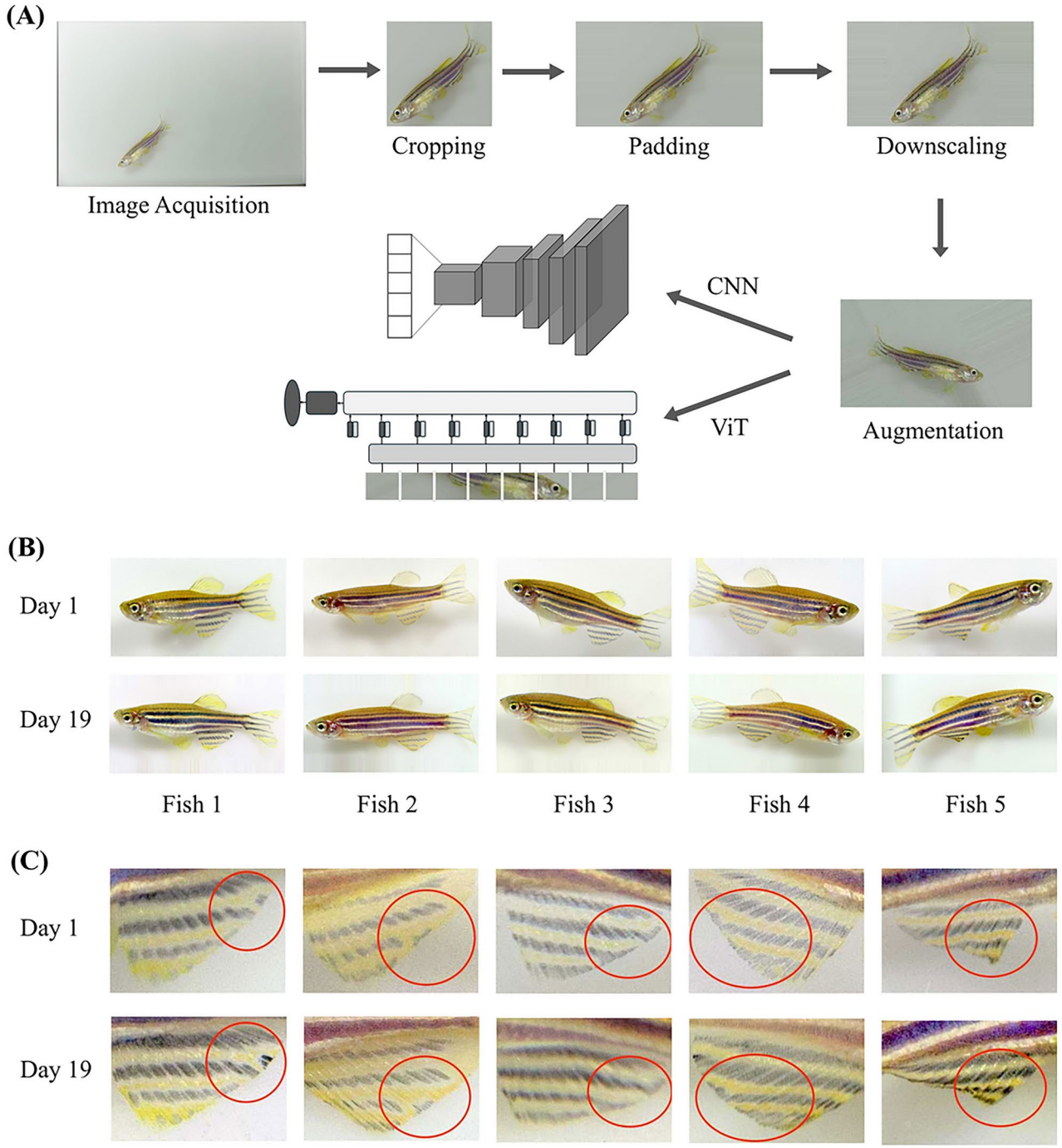

Fig. 2. (**A**) There were six components in the analysis pipeline. After automated cropping and padding, images with incorrect cropping or rotation were manually corrected. The image augmentation subroutine expanded the training set by generating additional images with alternate scale and rotation angles. The images were split into training, validation, and testing datasets as needed for analysis using either a computational neural network or vision transformer architecture. (**B**) Cropped images showing the five zebrafish in similar poses on Day 1 and Day 19. The Day 1 images were randomly selected from those images that showed a full side profile, and a similar pose was manually found in the Day 19 dataset. The length of each fish was estimated as 1.4 cm to 2.0 cm over the study. (**C**) Representative images showing a by-eye manual validation of the cross-day matching of fish groups using time-stable patterns of the anal fin. The large number of excess images in our proof-of-concept dataset enabled this independent matching.

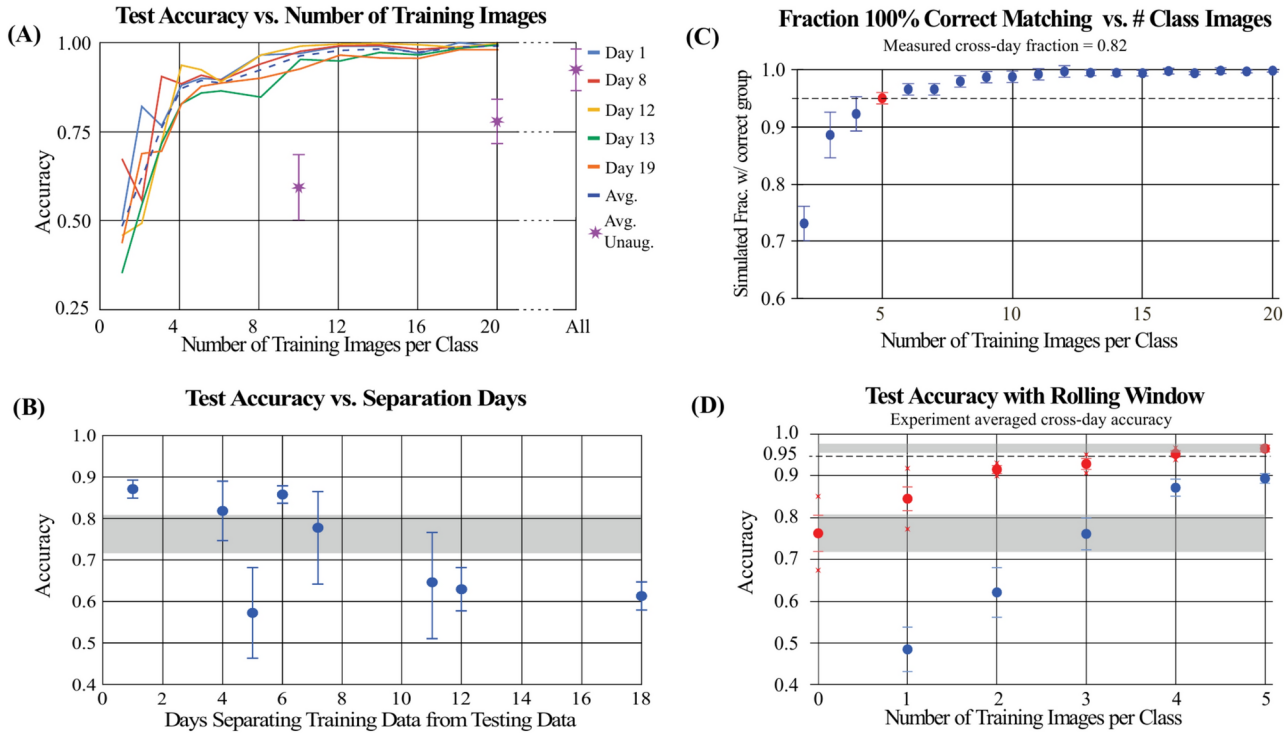

Fig. 3. (**A**) CNN same-day test accuracy dependence on the number of training images per class for each day's dataset. The dashed (-) line indicates the day-averaged behavior, and the asterisk (*) indicates the diminished training accuracy caused by training without image augmentation. (**B**) The measured dependence of the accuracy of the cross-day testing without a rolling window as a function of the number of days of separation between the training and testing data. Twenty fish per class on the test day were used to establish a putative high-accuracy class matching to the first day where no class inconsistencies occurred. Ten images per class were used for training. The gray bar shows the uncertainty on the mean fractional accuracy of all cross-day measurements and error bars indicate the minimum and maximum fractional accuracy. (**C**) The simulated accuracy of all of the five classes in a subsequent day matching without conflict classes used on the first day, given that it was known that the same groups exist on the testing day. Using the measured value $p_{crossday} = 0.82$, 5 images per class (red) led to 100% correct matching in $\geq 95\%$ (dashed) of the runs. Uncertainty bars show the standard deviation about the mean. (**D**) The experiment averaged test accuracy of the rolling window method using 10 images per class on the initial day and adding 0 to 5 images per class as a rolling window for each subsequent training day. The red data points show the average accuracy enhancement with a rolling window relative to training without data from the prior day (blue data points from the dashed shown in **A**). A rolling window of size 0 corresponds to no rolling window. Data sets were averaged both forward (starting on Day 1 with 10 images per class and rolling forward in time) and backward (starting on Day 19 with 10 images per class and rolling backward in time). The same wide gray bar as shown in B has been included for reference. The narrow gray bar demonstrates the increased mean test accuracy and reduced uncertainty on the mean accuracy when a rolling window of five images was used. Error bars depict the standard error (SE) about the data point mean values, and red asterisks indicate the forward and backward rolling window accuracies.

day of separation was 0.87 ± 0.02, while after 7 days separation decreased to near 0.80. The gray bar indicates the uncertainty of the cross-day tests averaged over the entire 18-day experiment. Based on our protocol, the seven-day separation between the first two data sets collected, Day 1 and Day 8, set $p_{crossday} = 0.82$ as the estimate of the cross-day test accuracy.

We then estimated a minimum the number of rolling window images needed after Day 1 to: (1) accurately match the fish classes of a new day to those of the first day, while (2), facilitating a model that achieves at least 95% test accuracy on the new day after retraining with the new day images. Importantly, since the rate of feature evolution was unknown, the number of images required for sufficient same-day test accuracy did not inform the number needed to secure high-fidelity cross-day class matching.

Setting the size of the rolling window: In this experiment, it was known that all days had the same number of classes (zebrafish individuals) and all collected images must belong to a Day 1 class, although membership was not explicitly verifiable. To establish class matching between days, we employed a *majority rule* matching protocol (see Methods). In brief, all of the images in a new day's group were classified using the model trained only on the previous days' image data. A new group was paired with the group from the previous day that was most often matched to the new group. In the case where an equal number of images from a new group matched to two or more prior-day groups, the new group was flagged as uncommitted. After repeating the procedure for

all five groups, if there was only a single uncommitted new group, it was matched to the remaining unmatched prior-day group.

To assist in choosing the number of images for future days, we developed a Monte Carlo simulation that determined the fraction of times 100% correct class matching (i.e., no error flag) for the five groups would occur when following our class matching method. The fractional success simulation depended on the measured average cross-day likelihood of correctly classifying a single image the next day when using a model trained only on the images from the previous day (Fig. 3B). Based on the seven-day span between Day 1 and Day 8, we set $p_{crossday} = 0.82$. For five classes, our simulation predicted that five images per class were sufficient to generate 100% class matching between days $\geq 95\%$ of the time. Using 10 images per class increased the success above 98%. Based on this simulation (Fig. 3C), we chose to evaluate the accuracy of the rolling window using a maximum of five images per class combined with 10 images from the first day of the experiment.

Figure 3D shows the improvement in the experiment average test accuracy when the rolling window images were included in the training data of the previous day. After each new data set was acquired, up to five images per class were added to the 10 images per class from the first day of the experiment for training. In summary, no more than five new images were needed for each new day after the second day to achieve high-fidelity, label-free, cross-day class matching, with an average test accuracy of $\geq 95\%$ when the rolling window training approach was used.

ViT model with rolling window training

The same rolling-window approach was also tested using a Vision Transformer (ViT) architecture. The cross-day class matching and the rolling window implementation followed the same procedure. The baseline ViT model chosen was ViT-B/32, which includes 32×32 pixel patches. The "B" indicates that this model is the base version of the Vision Transformer architecture, which has 12 layers in total. Figure 4A shows the results of the first step in our procedure, where images from the first day of data collection were used to train a model to classify additional images from the same day. In this case, around six images were required for 95% cross-day average test accuracy. The time evolution of the cross-day average test accuracy followed a trajectory similar to the CNN model, but demonstrated a modestly improved experiment average test accuracy (Fig. 4B; wide gray bar). Importantly, the inclusion of the rolling window once again significantly improved the experiment averaged test accuracy (Fig. 4B; narrow gray bar) to more than 95% when using a rolling window, although only two new images were required to exceed 95% test accuracy.

Feature analysis

We used class activation heatmap (CAM) analysis[45] as one method to explore the potentially complex features most relevant to our CNN model. A CAM was formed by applying the final filter layers of the trained CNN to an image to compute a weighted global average based on the predicted class. However, as the spatial resolution of each successive filter layer decreases, the final heatmap corresponded to a 4×8 grid in the input image (Fig. 5A). As expected, the heatmaps were centered on the zebrafish rather than spurious features in the image background. In addition, the CAMs demonstrated that our model is primarily activated by the fish's bodies rather than their heads or tails. Figure 5B shows activation maps associated with an earlier-level convolution layer of resolution 20×37 and is consistent with the dominant role of the whole body in classification.

Digital alteration of the zebrafish images in lieu of ablation or genetic manipulation provided additional insight into the CNN feature selection. To investigate specific feature changes, the standard CNN model was first trained on 70% of the Day 8 dataset (409 images) and validated on 15% of the dataset (72 images), with an equal

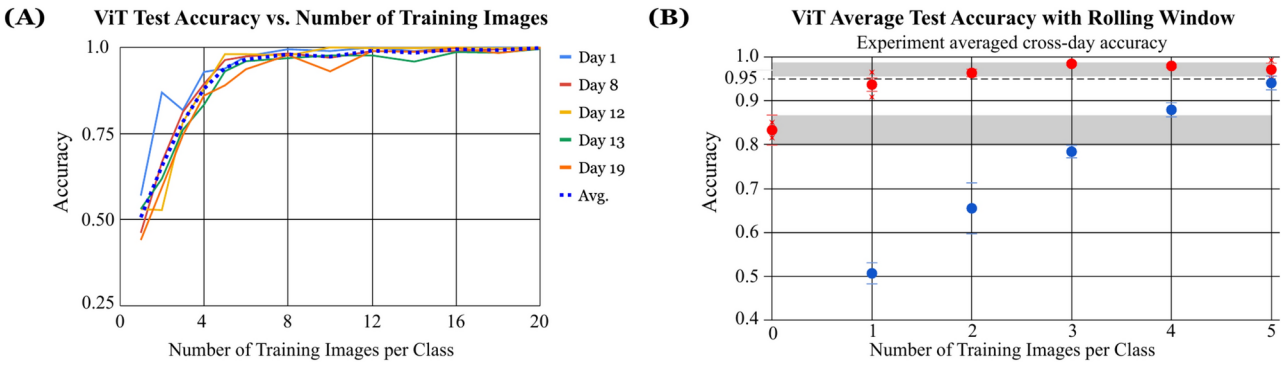

Fig. 4. (**A**) ViT same-day test accuracy dependence on the number of training images per class for each day's dataset. The dashed (-) line indicates the day-averaged behavior. (**B**) Average test accuracy of the ViT-B/32 model, averaged forward and backward as before, using 10 images per class on the initial day and adding 0 to 5 images per class as a rolling window for each subsequent training day. Similar to the Inception-V3 model, there is an increase in test accuracy as the number of training images increases, with diminishing returns as the number of training images increases. Once again, inclusion of the rolling window significantly improves the average test accuracy (narrow gray band) compared to the analysis without the rolling window (wide gray bar). Error bars depict the standard error (SE) about the data point mean values, and red asterisks indicate the forward and backward rolling window accuracies.

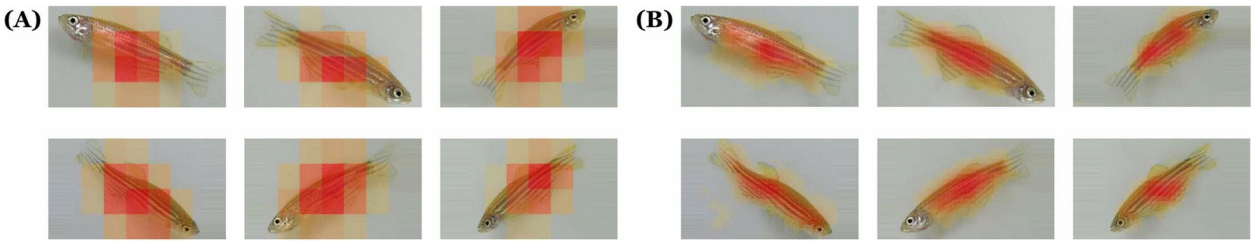

Fig. 5. Representative class activation maps (**A**) at the resolution the final convolutional layer, and (**B**) at 5x higher resolution using an earlier convolutional layer.

number of images for each class. This trained model was then tested on three modified test datasets formed from the remaining 15% of the Day 8 data (75 images): (1) contrast blurring of the head, body, and fins, (2) spatially unchanged but grayscale converted, and (3) B/W elliptical shapes matched to the measured aspect ratio of the five fish in this study (Fig. 6A,C,E). Blurring of image body regions was performed manually using the GIMP image manipulation program.

Only the wide-scale blurring of the entire body of the zebrafish produced any significant effect on the model's test accuracy (Fig. 6B). We interpret this to mean that the CNN as trained on the original images relied heavily on the pattern of stripes, arguably the most visually apparent feature of zebrafish, to guide its output, rather than the head, tail, or fin of the zebrafish. These results follow other findings about CNN models pretrained on the ImageNet databases being biased towards pattern over shape[46].

We found color to be an important feature when part of the training image dataset. The testing of gray-converted images against a color-image trained model was consistent with random accuracy despite the presence of body stripes highlighted previously. In contrast, when trained on gray-converted images, color was not needed for high-accuracy test results (Fig. 6D) and led to significant misclassification when part of the test images. Together, the implication was that, when available, color may have dominated the model's classification feature space, but was not strictly necessary for high-accuracy classification.

To further demonstrate the ability to explore salient cues sensed by these trained models, we also explored the relative importance of the overall shape of the body. The length-to-height ratio (Fig. 6E,F) of the five fish was measured and used to create five shape-representative black ellipses on a white background. Then a group of testing images was created using these five ellipses, mimicking the Day 8 fish shapes. The CNN model, trained on gray-converted images from Day 8 without augmentation, demonstrated a 90% classification accuracy when tested on the simulated dataset. When augmentation was included in training, only 41% test accuracy was achieved. Since augmentation was a standard part of the high-accuracy CNN and ViT training above, the results suggested that overall body shape alone was not a sufficient feature to achieve the previous classification results.

Discussion

In this study, we addressed some of the challenges of minimally invasive tracking and identification of individual zebrafish in laboratory settings using advanced machine learning techniques. Leveraging convolutional neural networks (CNNs) and vision transformers (ViTs), we developed a rolling window training technique that adapts to changing conditions over time. Our method demonstrated robust identification of maturing zebrafish in groups over several weeks with high fidelity, significantly reducing the need for new training images. By analyzing the impact of shape, pattern, and color on classification outcomes, we highlight the potential of this approach to improve real-time identification protocols and improve the understanding of CNN classifier success in this context. Our findings suggest that this technique provides a reliable and efficient solution for tracking temporally evolving zebrafish classes in research settings. Notably, the procedure is general enough to be applied to studies beyond zebrafish, though it should be noted that efficient scaling of the presented methods to studies with a significantly greater number of individuals has not been demonstrated.

Given the limited size of the image set and the distinct, though not visually obvious, features of individual zebrafish, it is possible that this study was not a particularly challenging classification task. To place our accuracy results in a historical context and confirm the significant advantage of recent deep learning approaches, images from a single day were used to explore the typical testing accuracy under a range of commonly reported models from scikit-learn[47] (Fig. 7) and the widely used UNet CNN architecture[48]. The models were trained and tested on images from Day 19. All classifiers except UNet require eigenvector inputs without image augmentation. In these cases, we computed 12 vectors for input. The UNet CNN used the same image augmentation as InceptionV3[47]. While these image classification approaches are generally considered obsolete in light of current deep learning techniques, it is striking to compare these numbers to the >95% testing accuracy achieved with both the CNN and transformer models.

For any specific segmentation tasks, it is generally an open question as to which method, convolutional neural network (CNN) or vision transformer (ViT), offers the best performance[49,50]. Although the vision transformer model allowed for a smaller rolling window size with the same accuracy performance, this result may be experiment-specific and warrants further study. There are reasons to expect a performance difference. CNNs use convolutional layers that apply filters to local patches of an image, capturing spatial hierarchies and local features through successive layers[51]. This local processing is efficient for handling image data due to its inherent translation invariance. In contrast, ViTs[39,52] divide an image into fixed-size patches and treat these patches as

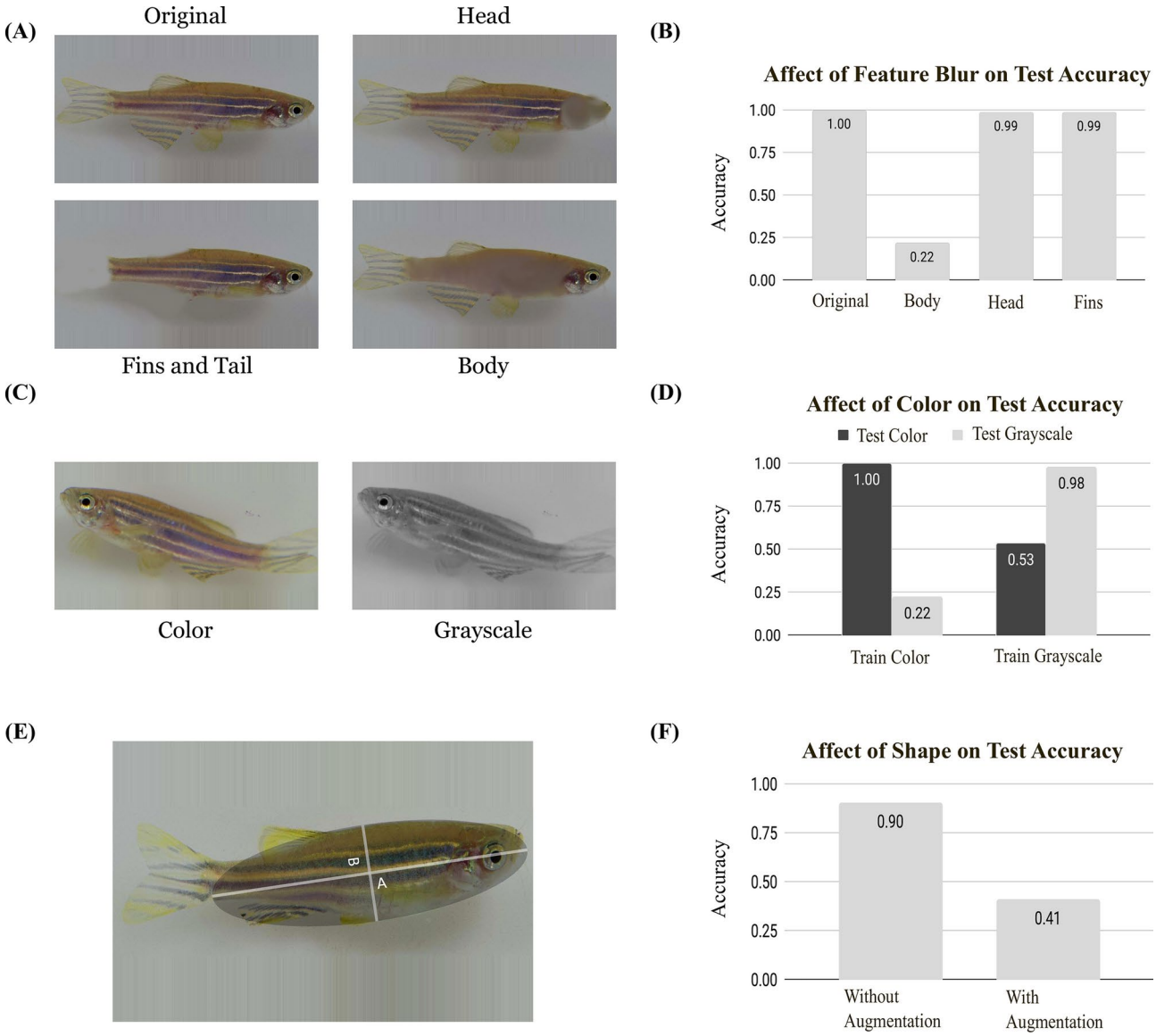

Fig. 6. Feature analysis. (**A**) Fish images were blurred by hand in GIMP to remove particular features. (**B**) Results of testing on blurred images (trained on unblurred images). Each type of blurring (unmodified, head, fins and tail, and body) was performed on the same 386 images drawn from Day 8. (**C**) Example of grayscale conversion. (**D**) Accuracy test results using grayscale images. (**E**) Measuring the major and minor axes of zebrafish. (**F**) Testing of ellipse images when trained on Day 8 grayscale images with and without augmentation.

sequences, similar to words in a sentence, using self-attention mechanisms to model relationships between patches, irrespective of their position. In CNNs, the receptive field grows layer-by-layer, creating early layers that capture fine details and deeper layers that address broader areas. ViTs, with their attention mechanisms[53], can theoretically consider all patches at once, offering a global perspective from the very beginning of the processing pipeline. Lastly, CNNs typically require fewer data and computational resources than ViTs, which often need extensive data and training time to effectively learn the global dependencies without overfitting[39]. Given the potential for significant performance differences on a limited dataset, we explored both approaches using the same training method. Despite the differences in architecture, both models demonstrated a significant benefit from the use of rolling window training for these evolving images. The relative enhancement of the test accuracy and the computational cost of the Vision Transformer were modest but may become significant factors in larger studies.

In general, there will be more images collected on Day 1 than are required to achieve a same-day test accuracy of 95% (or whatever threshold was specified). Although not included in the analysis presented here, including these images in the first-day baseline training dataset may prove beneficial to cross-day test accuracy at least initially. However, if the feature evolution of the fish is sufficiently rapid, including the extra images from Day 1 may be deliterious to long-term performance. Optimizing these effects will be of interest in future studies.

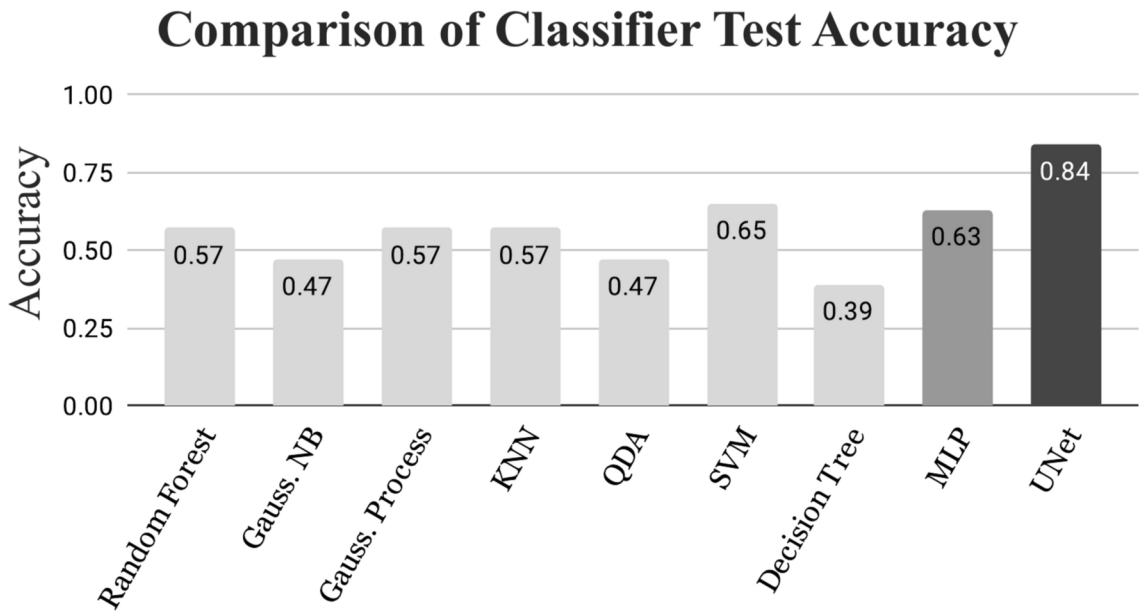

Fig. 7. Various classifiers run on the Day 19 dataset. All classifiers except UNeT require eigenvector inputs without image augmentation. We computed 12 vectors for input. The light gray bars indicate statistical machine learning models, medium gray bar is a multilayer perceptron, and the dark gray bar corresponds to the UNet CNN architecture.

A consideration of our rolling-window approach and protocol lies in situations where the objects under study begin to change more rapidly than is reflected in the first two measurements. The rolling-window technique retains the ability to assess feature drift by monitoring the test accuracy of the deep learning model over time. Should the testing accuracy for the most recent data decrease below the expected level, additional data are to be gathered.

Our feature alteration studies (Fig. 6) indicated that zebrafish body-specific stripes and, similar findings reported by Haurum et al. (2020), color are the two most informative characteristics to classify juvenile zebrafish. The significance of color in this classification work has an analog reported in behavioral and physiological studies. Abreu et al. reviewed several studies that support the important role of color in a variety of animal behaviors and extended these concepts to zebrafish models and their neurobehavioral attributes[54]. Beyond this behavioral response, zebrafish adjust their body coloration in response to chemical and visual cues in their environment. For example, both larvae and adult zebrafish can adapt to become lighter or darker by melanosome translocation in response to lighting conditions[55]. Furthermore, the complex layering of blue iridophores, black melanophores, and yellow xanthophores in the skin of zebrafish provides multiple pathways for color change[56]. In a more recent behavioral study,[57] reported that wild zebrafish, similar to other fish, showed a general preference for red and a preference for green over blue.

Although the representative average color of an individual zebrafish is not uniquely specified, it is worth considering the potential role of body color in these models. The body-blurring method resulted in a fish-specific averaged color over the majority of the fish body. However, Fig. 6D implies that the average color without the finer scale stripe structure effectively led to the prediction of the CNN model. If the spatial structure of the fish stripes was the predominant feature required for high model accuracy, then one may expect that training with either color or grayscale images would lead to high accuracy test results independent of the coloration of the test images. In 6D, the difference between the two left bars corresponding to the test results when trained on color images implies that color plays a significant role in the CNN test accuracy. However, the right two bars corresponding to the test results when trained on grayscale images imply that without color, other unknown features of the fish images provided *some* ability to correctly classify the color images. From these observations, we posit that a significant feature associated with the training of this CNN model on color images may best be thought of as a composite of color and stripe structure rather than two independent features. This type of combinatoric learning is reminiscent of the observed effect of multiple intrasensory cues in studies of child category learning[58]. Furthermore, the result implies that the relative average body intensity, which was likely preserved through both body blurring and grayscale conversion, was not a key factor in classification. A more rigorous analysis of the role of local color and stripe pattern is an area that would benefit from future study.

Methods

Aquarium and imaging

The five zebrafish imaged for this proof-of-concept study were part of an existing office pet aquarium. The fish were sourced from a local pet store (PetSmart, NJ) and were estimated to be between 8 and 12 weeks old when the study began. The fish were not handled directly or removed from the aquarium during this study. The studio

tank insert was designed to allow all fish to temporarily sequester near a side wall of the aquarium. Using a sliding panel, a single fish was allowed to enter the studio region, after which the sliding back wall was moved forward to restrict the fish to a $100 \times 100 \times 5$ mm volume. Acrylic spacers set a 5 mm depth between the panel and the aquarium glass. The limited depth sagittally biased the orientation of the nominally 3 to 4 mm thick fish while allowing free swimming. After the images were acquired, a third sliding panel was used to release the fish into the remaining portion of the tank, and the process was repeated as needed. A single training image data set for the five danios typically took 30 minutes to acquire. The same lighting and camera positions were maintained throughout the study.

Raw image acquisition

Based on size and coloration, the group consisted of four males and one female. The 38-liter aquarium was temperature controlled[59] at $25.6^{\circ}\mathrm{C}$ +/- $1^{\circ}\mathrm{C}$. Imaging studies began after 2 weeks of acclimatization and were carried out over 19 days. The images were cropped using a color analysis algorithm so that the body of the zebrafish was centered within the image and occupied the majority of pixels. They were rotated, padded, and resized to a 320×180 pixel image. The images were taken with a Canon 7D camera and converted to JPEG before use. After photographing the fish, we compile the data into a dataset containing 200-500 pictures of each fish. Approximately 10% of the original images were discarded due to blur or convoluted body orientation. The pictures were then cropped to roughly fit the shape of the body of the fish. To preserve the aspect ratio of the images, we also added constant color padding that was similar to that of the background, so that all images were 16:9 when inputted into our model. The images were then divided into training, validation, and testing sets with a 70% / 15% / 15% split. Keras' built-in image augmentation function was applied to the training set to generate additional training images through random rotations, shear, and translation. Each data set contained roughly the same number of images per fish. The data set on Day 1 consisted of 201 images, while Days 8 and 12 contained approximately 550 images. Days 13 and 19 contained 1005 and 1133 images, respectively. Augmentation increased the size of a training dataset to approximately 10,000 images.

Data processing

Models were created using Keras on a Tensorflow backend. The CNN model was developed on the Keras API in Python (https://keras.io), running on top of the popular open-source framework TensorFlow (https://www.tensorflow.org). For training, images were augmented using Keras's built-in image generator, including rescale (1/255), channel shift range (5), shear range (0.05), zoom range (0.1), vertical flip, horizontal flip, rotation range (360), width shift range (0.1), and height shift range (0.1). To allow the model to adapt to variations in image intensity, random brightness shifting was also included during augmentation using the default shift range of 0.3 to 1.3. Details can be found at https://keras.io/api/layers/preprocessing_layers/image_augmentation/.

We based our initial model on a modified version of Google's Inception V3 (https://keras.io/applications/#inceptionv3). The first and last layers of the Inception V3 model were altered to fit our cropped image size of 320 x 180 pixels, and we used an output vector of size 5, one scalar for each class. Typically, the training phase was the most time-intensive part of the modeling. Since developing robust network weights could have required days or even weeks, we chose to initialize to a prior weighting matrix available from ImageNet for InceptionV3[38,42]. We then retrained our model as described above. This technique allowed model retraining to complete in minutes to hours rather than days. All computations were run on a Princeton Research Computing cluster node with one Nvidia A100 GPU and 16 CPU cores. The full Inception V3 model has 11 modules. Varying the number of modules when trained and tested on Day 19 showed that using 5 or more modules leads to a test accuracy of 1.0 for the same day. We chose to use the full number of modules throughout this study. The ViT[39] modeling was carried out using the same cluster.

We evaluated training time on a Nvidia A100 GPU when using 10,000 images (after augmentation) from a range of standard variants of the ViT model (Table 1) and compared these with our Inception V3 CNN. The ViT-S is the smallest and most lightweight variant, the ViT-B is a mid-size variant with a balance between capacity and efficiency, and the ViT-L is the largest and most powerful variant.

Rolling window and majority rule protocols

The following protocol outlines our procedure for estimating the number of rolling window images per class needed to maintain the cross-day class matching accuracy $\geq 95\%$ and provide sufficient data for cross-day test accuracy $p_{rw} \geq 95\%$:

1. Set up an experiment for data collection and assign the number of classes n_c to be used. Make a best guess at the maximum number of measurements per class N_{mc} needed to achieve the desired testing accuracy within the same day's dataset (in our case, $p_{sameday} \geq 0.95$). A suggested value for our type of experiment in which animal morphology changed on the time scale of days is between 10 and 30 images per class.
2. After image collection, train, validate, and then test the image classification model. Plot the test accuracy of classifying an image $p_{sameday}$ as a function of the number of images per class (e.g., Fig. 1A). Ideally, this accuracy curve should plateau above $p_{sameday}$ for the largest number of images per class. Additional images for each class should be acquired as needed to achieve this result. From this test accuracy curve, determine the number of images per class N_{sd95} needed for the test accuracy to meet or exceed a value of $p_{sameday} \geq 0.95$ (or one's preferred threshold) for training and testing on the same day. Typically, $N_{sd95} \leq N_{mc}$.
3. Estimate the image feature drift rate $p_{crossday}$ by acquiring a second data set a time Δt later with the same number N_{mc} of images per class as in the first data set. In our case, we chose $\Delta t = 7$ days; however, it was later determined that measurable change occurred within 24 hours (e.g., Fig. 3B). In general, a larger value of Δt will lead to a need for more images.

Model	Train FLOPs	Inference FLOPs	Params	Batch size	Train time
ViT-S/16	1.34e14	2.24e10	22M	50	3.3h
ViT-B/8	2.35e15	1.56e11	86M	20	11.7h
ViT-B/16	4.95e14	3.30e10	86M	20	13.8h
ViT-B/32	1.32e14	8.82e9	86M	20	38m
ViT-L/16	1.85e15	1.23e11	303M	20	18h
InceptionV3	2.45e13	6.59e10	22M	60	7m

Table 1. Comparison of model computational costs, assuming 10,000 image training dataset. Total training time on a A100 GPU assumes 15 epochs of training for ViT models, and 30 epochs for InceptionV3. Training FLOPs includes forwards and backwards pass. These values are estimates, depending on specific dataset configurations and hardware setups, and FLOPs do not include overhead operations outside the model, including moving data into and out of GPU memory. FLOPs were calculated using TensorFlow's built-in profiler.

Assign cross-day class matching using the following *majority rule* matching protocol:
Use the model trained on the first-day images to classify the groups of the second day where there are N_{mc} images per class. For each class, the assigned match is the first-day class corresponding to the class most frequently assigned for the N_{mc} fish in a group. After repeating for all second-day groups, there are then n_c match fractions f_i. The average cross-day match fraction is $p_{crossday} = \frac{\sum(f_i)}{n_c}$. If $p_{crossday} \approx p_{sameday}$, there was no significant evolution of features during the time Δt and a rolling window is not needed.

4. Handling special majority rule cases:
   - If all classes are equally assigned, a conflict flag is set for that group and it remains unassigned.
   - If, after cycling through all groups, a single second day class remains unassigned, the group is paired to the unmatched first day group and the conflict flag is removed.
   - If two or more second day groups have conflict flags set, N_{mc} is declared too small, more images are collected, and the matching process is repeated.
5. Use the Matlab simulator to determine an initial rolling window size N_{rw} such that the likelihood of cross-day class matching is $\geq 95\%$ (or the desired value). Our Monte Carlo simulation uses the average measured successful matching probability between two image data sets collected at different times where there are some feature changes in the classes with no change in the number of classes. The probability of successful cross-day matching of each class is treated as independent.
6. The last step is to confirm that a rolling window of size N_{rw} sufficiently increases the cross-day fraction p_{rw} above $p_{crossday}$. Retrain the model using the N_{sd95} images from the first day plus N_{rw} rolling window matched images from the second day. Confirm the test accuracy for the second day images is $p_{rw} \geq 0.95$. If $p_{rw} < 0.95$, increase N_{rw} until the condition is met. If $N_{rw} = N_{mc}$ still does not meet this condition, additional second-day images should be collected and a shorter delay between imaging days Δt should be considered.

   In this reported experiment, $N_{rw} = 5$ satisfied both the class matching accuracy and the cross-day test accuracy conditions, although, in general that did not need to be the case.

Data availability

Long-term access to the raw zebrafish image dataset is possible at https://www.doi.org/10.34770/pz36-j044. The deep learning code, Matlab simulation code, and the preprocessed zebrafish images (before augmentation) are available at https://github.com/backfire42/Zebra_Code_And_Data.

Received: 12 September 2024; Accepted: 10 January 2025
Published online: 12 March 2025

References

1. Chakraborty, M., Kalyan, R., Sedhain, A., Dhakal, P. & Karunakaran, G. Zebrafish an emerging model for preclinical drug discovery. *Int. J. Res. Pharm. Sci.* **11**, 1638–1648 (2020).
2. Singleman, C. & Holtzman, N. G. Growth and maturation in the zebrafish, *Danio rerio*: A staging tool for teaching and research. *Zebrafish* **11**, 396–406. https://doi.org/10.1089/zeb.2014.0976 (2014).
3. Choi, T.-Y., Choi, T.-I., Lee, Y.-R., Choe, S.-K. & Kim, C.-H. Zebrafish as an animal model for biomedical research. *Exp. Mol. Med.* **53**, 310–317. https://doi.org/10.1038/s12276-021-00571-5 (2021).
4. Lidster, K., Readman, G. D., Prescott, M. J. & Owen, S. F. International survey on the use and welfare of zebrafish *Danio rerio* in research. *J. Fish Biol.* **90**, 1891–1905. https://doi.org/10.1111/jfb.13278 (2017).
5. Giraldez, A. J. et al. Micrornas regulate brain morphogenesis in zebrafish. *Science* **308**, 833–838 (2005).
6. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498EP (2013).

7. Sehring, I. M. & Weidinger, G. Recent advancements in understanding fin regeneration in zebrafish. *WIREs Dev. Biol.* **9**, https://doi.org/10.1002/wdev.367 (2019).
8. Poss, K. D., Wilson, L. G. & Keating, M. T. Heart regeneration in zebrafish. *Science* **298**, 2188–2190 (2002).
9. Wehner, D. et al. Signaling networks organizing regenerative growth of the zebrafish fin. *Trends Genet.* **31**, 336–343. https://doi.org/10.1016/j.tig.2015.03.012 (2015).
10. Oliveira, R. F., Silva, J. F. & Simoes, J. M. Fighting zebrafish: characterization of aggressive behavior and winner–loser effects. *Zebrafish* **8**, 73–81. https://doi.org/10.1089/zeb.2011.0690 (2011).
11. Delcourt, J. et al. Individual identification and marking techniques for zebrafish. *Rev. Fish Biol. Fish.* **28**, 839–864 (2018).
12. Cousin, X. et al. Electronic individual identification of zebrafish using radio frequency identification (rfid) microtags. *J. Exp. Biol.* **215**, 2729–2734 (2012).
13. Martins, T., Diniz, E., Félix, L. M. & Antunes, L. Evaluation of anaesthetic protocols for laboratory adult zebrafish (danio rerio). *PLoS ONE* **13**, 1–12 (2018).
14. Al-Jubouri, Q., Al-Azawi, R., Al-Taee, M. & Young, I. Efficient individual identification of zebrafish using hue/saturation/value color model. *Egypt. J. Aquat. Res.* **44**, 271–277. https://doi.org/10.1016/j.ejar.2018.11.006 (2018).
15. Guilbeault, N. C., Guerguiev, J., Martin, M., Tate, I. & Thiele, T. R. Bonzeb: open-source, modular software tools for high-resolution zebrafish tracking and analysis. *Sci. Rep.* https://doi.org/10.1038/s41598-021-85896-x (2021).
16. Xu, Z. & Cheng, X. E. Zebrafish tracking using convolutional neural networks. *Sci. Rep.* **7**, 42815 https://doi.org/10.1038/srep42815 (2017).
17. Romero-Ferrero, F., Bergomi, M. G., Hinz, R. C., Heras, F. J. H. & de Polavieja, G. G. idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. *Nat. Methods* **16**, 179–182. https://doi.org/10.1038/s41592-018-0295-5 (2019).
18. Panadeiro, V., Rodriguez, A., Henry, J., Wlodkowic, D. & Andersson, M. A review of 28 free animal-tracking software applications: current features and limitations. *Lab Anim.* **50**, 246–254. https://doi.org/10.1038/s41684-021-00811-1 (2021).
19. Franco-Restrepo, J., Forero, D. & Vargas, R. A review of freely available, open-source software for the automated analysis of the behavior of adult zebrafish. *Zebrafish* **16**, 223–232. https://doi.org/10.1089/zeb.2018.1662 (2019).
20. Haurum, J. B., Karpova, A., Pedersen, M., Bengtson, S. H. & Moeslund, T. B. Re-identification of zebrafish using metric learning. In *IEEE Winter Applications of Computer Vision Workshops (WACVW)* 1–11, https://doi.org/10.1109/WACVW50321.2020.9096922 (2020).
21. Schneider, S., Taylor, G. W., Linquist, S. & Kremer, S. C. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods Ecol. Evol.* **10**, 461–470. https://doi.org/10.1111/2041-210X.13133 (2019).
22. Ravoor, P. C. & Sudarshan, T. S. B. Deep learning methods for multi-species animal re-identification and tracking - a survey. *Comput. Sci. Rev.* **38**, 100289. https://doi.org/10.1016/j.cosrev.2020.100289 (2020).
23. Ye, M. et al. Transformer for object re-identification: a survey. *Int. J. Comput. Vis.* https://doi.org/10.1007/s11263-024-02284-4 (2024).
24. Zhai, Y., Guo, X., Lu, Y. & Li, H. In defense of the classification loss for person re-identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1526–1535, https://doi.org/10.1109/CVPRW.2019.00194 ISSN: 2160-7516 (2019).
25. Ye, M. et al. Deep learning for person re-identification: a survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 2872–2893. https://doi.org/10.1109/TPAMI.2021.3054775 (2022).
26. Ishaq, O., Sadanandan, S. K. & Wählby, C. Deep fish: Deep learning-based classification of zebrafish deformation for high-throughput screening. *SLAS Discov.* **22**, 102–107. https://doi.org/10.1177/1087057116667894 (2017).
27. Toulany, N. et al. Uncovering developmental time and tempo using deep learning. *Nat. Methods* **20**, 2000–2010. https://doi.org/10.1038/s41592-023-02083-8 (2023).
28. Jones, R. A., Renshaw, M. J. & Barry, D. J. Automated staging of zebrafish embryos with deep learning. *Life Sci. Alliance* https://doi.org/10.26508/lsa.202302351 (2023).
29. Frantz, W. & Ceol, C. From tank to treatment: Modeling melanoma in zebrafish. *Cells* https://doi.org/10.3390/cells9051289 (2020).
30. Gamble, J., Elson, D., Greenwood, J., Tanguay, R. & Kolluri, S. The zebrafish xenograft models for investigating cancer and cancer therapeutics. *Biology* **10**, 252. https://doi.org/10.3390/biology10040252 (2021).
31. Huang, E. et al. Establishment of a zebrafish xenograft model for in vivo investigation of nasopharyngeal carcinoma. *EBioMedicine (Lancet)* **31**, https://doi.org/10.1177/09636897221116085 (2022).
32. Yan, C., Yang, Q., Do, D., Brunson, D. & Langenau, D. Adult immune compromised zebrafish for xenograft cell transplantation studies. *EBioMedicine (Lancet)* **47**, 24–26. https://doi.org/10.1016/j.ebiom.2019.08.016 (2019).
33. van de Ven, G., Tuytelaars, T. & Tolias, A. Three types of incremental learning. *Nat. Mach. Intell.* **4**, 1185–1197. https://doi.org/10.1038/s42256-022-00568-3 (2022).
34. Zhu, D., Bu, Q., Zhu, Z., Zhang, Y. & Wang, Z. Advancing autonomy through lifelong learning: a survey of autonomous intelligent systems. *Front. Neurorobotics* https://doi.org/10.3389/fnbot.2024.1385778 (2024).
35. Iman, M., Arabnia, H. & Rasheed, K. A review of deep transfer learning and recent advancements. *Technologies* https://doi.org/10.3390/technologies11020040 (2023).
36. Shen, L., Wei, Z. & Wang, Y. Determining the rolling window size of deep neural network based models on time series forecasting. *J. Phys. Conf Ser.* **2078** (2021).
37. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255, https://doi.org/10.1109/CVPR.2009.5206848 (2009).
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. CoRR **abs/1512.00567** (2015).
39. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations* (2021).
40. Bradford, Y. et al. Zebrafish information network, the knowledgebase for *Danio rerio* research. *Genetics* https://doi.org/10.1093/genetics/iyac016 (2022).
41. Russakovsky, O. et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252. https://doi.org/10.1007/s11263-015-0816-y (2015).
42. Johnson, J. M. & Khoshgoftaar, T. M. Survey on deep learning with class imbalance. *J. Big Data* **6**, 27 (2019).
43. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019).
44. Alwosheel, A., van Cranenburgh, S. & Chorus, C. G. Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. *J. Choice Model.* **28**, 167–182. https://doi.org/10.1016/j.jocm.2018.07.002 (2018).
45. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2921–2929 (2016).
46. Geirhos, R. et al. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness, https://doi.org/10.48550/arXiv.1811.12231 (2022).
47. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
48. Olaf Ronneberger, P. F., Thomas Brox", e. N., Hornegger, J., Wells, W. & Frangi", A. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer International Publishing, Cham, 2015).

49. Maurício, J., Domingues, I. & Bernardino, J. Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Appl. Sci.* https://doi.org/10.3390/app13095521 (2023).
50. Takahashi, S. et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. *J. Med. Syst.* https://doi.org/10.1007/s10916-024-02105-8 (2024).
51. Zhao, X., Wang, L., Zhang, Y., Han, X. & Deveci, M. A review of convolutional neural networks in computer vision. *Artif. Intell. Rev.* https://doi.org/10.1007/s10462-024-10721-6 (2024).
52. Li, X. et al. Transformer-based visual segmentation: A survey. arXiv:2304.09854 (2024).
53. Naseer, M. et al. Intriguing properties of vision transformers. arXiv:2105.10497 (2021).
54. Abreu, M. S. D. et al. Color as an important biological variable in zebrafish models: Implications for translational neurobehavioral research. *Neurosci. Biobehav. Rev.* **124**, 1–15. https://doi.org/10.1016/j.neubiorev.2020.12.014 (2021).
55. Gerlai, R., Lahav, M., Guo, S. & Rosenthal, A. Drinks like a fish: zebra fish (*Danio rerio*) as a behavior genetic model to study alcohol effects. *Pharmacol. Biochem. Behav.* **67**, 773–782. https://doi.org/10.1016/S0091-3057(00)00422-6 (2000).
56. Singh, A. & Nüsslein-Volhard, C. Zebrafish stripes as a model for vertebrate colour pattern formation. *Curr. Biol.* **25**, R81–R92. https://doi.org/10.1016/j.cub.2014.11.013 (2015).
57. Roy, T. et al. Color preferences affect learning in zebrafish, *Danio rerio*. *Sci. Rep.* **10**, 5. https://doi.org/10.1038/s41598-019-51145-5 (2021).
58. Broadbent, H., Osborne, T., Mareschal, D. & Kirkham, N. Are two cues always better than one? the role of multiple intra-sensory cues compared to multi-cross-sensory cues in children's incidental category learning. *Cognition* **199**, 104202. https://doi.org/10.1016/j.cognition.2020.104202 (2020).
59. Aleström, P. et al. Zebrafish: Housing and husbandry recommendations. *Lab. Anim.* **54**, 213–224. https://doi.org/10.1177/0023677219869037 (2020).

Acknowledgements

AS and BD were part of the Princeton Laboratory Learning Program in 2019. The authors thank Rick Soden for lending the camera used in this experiment and for his photography advice. The author(s) are pleased to acknowledge that the work reported on in this paper was substantially performed using the Princeton Research Computing resources at Princeton University which is consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology's Research Computing. The simulations presented in this article were performed on computational resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

Author contributions

JP conceived and conducted the experiments. AS developed the deep learning models and carried out related numerical tests. AS and BD performed the feature analysis. JP developed and performed the class matching simulation. All authors contributed and reviewed the manuscript.

Declarations

Ethics statement

Princeton University's Institutional Animal Care and Use Committee (IACUC) provides supervision, coordination, review, and approval of university vertebrate animal projects consistent with ARRIVE guidelines. This zebrafish imaging study was conducted and reported in accordance with the relevant guidelines and regulations. The fish were not handled directly, anesthetized, euthanized, or removed from the aquarium during this study.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.