Keypoint Annotation for Electrocommunication Source Separation with PIKAChU and RAIChU

Kaden Zheng Harvard University
Sonja Johnson-Yu Harvard University
Satpreet Harcharan Singh Harvard University
Denis Turcu Columbia University
Federico Pedraja Columbia University
Pratyusha Sharma MIT
Naomi Saphra Harvard University
Nathaniel Sawtell Columbia University
Kanaka Rajan * Harvard University

Abstract

Source separation is a key step in understanding animal communication, particularly in acoustic and electric modalities. Often, positional information is used to determine which animal emitted a given signal. Automatic keypoint tracking algorithms can provide this information, but they sometimes require a prohibitively large number of hand-labeled examples. We introduce a methodology for automatically labeling keypoints in weakly electric fish, which sense and communicate by emitting electric pulses. Specifically, we invert a physics-based electrogeneration model to reconstruct the position of the fish which emitted a recorded pulse. This approach allows us to transform inexpensive electric-signal recordings into pose estimates. Our algorithm makes feasible the study of social interactions and *communication* in schools of weakly electric fish by increasing the efficacy of source separation in electrocommunication settings. More broadly, we provide a general framework for augmenting pose estimation training data to improve source separation, with potential applications in acoustic modalities and beyond.

1 Introduction

When deciphering animal communication, it is crucial to identify which animal produced a given signal. In some modalities [9, 1, 3], this **source separation** problem must be solved to conduct further experimental analysis. Often, source separation relies on positional tracking generated from videos of animal behavior [6, 4]. Automated positional tracking algorithms label *keypoints* for each animal in the video, but annotating keypoint training data can be time-consuming and labor-intensive. Without sufficient annotations, the predicted keypoints are more likely to be inaccurate.

Our work focuses on electrocommunication in weakly electric fish, specifically *G. petersii*. To study group behavior, we must correctly attribute each pulsatile **electric organ discharge (EOD)** to an emitting fish [2, 8, 5]. This can prove challenging when fish are shoaling close together; their similar locations produce similar electric fields, so even a small keypoint labeling error can cause EOD misattributions. To improve keypoint accuracy, trackers require multitudes of high-quality training data labeled by humans. Can we expand this training set without relying on human annotators?

We propose the **Physics-Informed Keypoint Annotator for Channel Unmixing (PIKAChU)**, a method for *automatically* labeling keypoints without any manual annotation. PIKAChU exploits

*Correspondence: kanaka_rajan@hms.harvard.edu

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

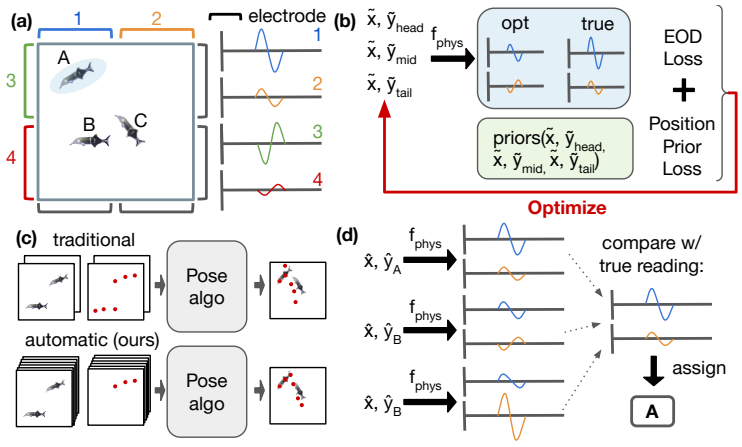

Figure 1: **(a) Experimental setup.** Weakly electric fish are recorded in a square 60cm^2 tank surrounded by electrode pairs. When a fish emits an EOD, the electric field is recorded by the electrode array. **(b) Our Physics-Based Optimization Pipeline.** Randomly initialized keypoint locations are run through the physics model to produce an estimate of the electrode trace corresponding to those locations. This is compared with the true EOD trace to yield the EOD Loss. Additionally, keypoint locations are compared to a series of priors to calculate a Positional Prior Loss. Optimization is run over these two losses. **(c) How Automatic Labeling Improves Pose Estimation**. Traditionally, pose estimation algorithms like SLEAP are trained with a small set of hand annotations. If an insufficient number of labels are provided, there is high error at inference time. Our approach automates the creation of labels to improve pose estimation training and inference. **(d) EOD Assignment Pipeline**. Estimates of fish positions are fed into the physics model to generate the expected electrode traces, repeated for each fish. Then the electrode traces are compared with the true electrode reading. The fish with the highest similarity is assigned as the source. This pipeline enables automatic keypoint labeling and improves assignment accuracy downstream (see Fig. 2).

the well-defined physics of electric field generation [10] to guide an optimizer towards the source of an EOD. Given the electric field recorded by an array of electrodes, our algorithm accurately identifies the position of the pulsing animal and labels its keypoints. We improve further on PIKAChU by introducing the **Regularizer-Augmented Inverter for Channel Unmixing (RAIChU)**, which incorporates positional priors during optimization. Using these automatic labelers, we can scalably improve the animal tracker by further training.

2 Methods

We aim to automatically label fish keypoints from synchronized video and electrode recordings using a physics-guided optimization pipeline. We formulate EOD assignment and pose recovery as matching between a physics-based prediction and the measured electrode pattern (Fig. 1).

2.1 Problem formulation

Each discharge induces an electric potential measured by E electrode pairs. We summarize each event as an E-dimensional vector of signed peak-to-peak (PTP) amplitudes. Concretely, for a frame t we have $(\text{frame}_t, \mathbf{y}_t^{\text{obs}})$, where $\mathbf{y}_t^{\text{obs}} \in \mathbb{R}^E$ is the recorded PTP (with $E = 10$). The forward physics model maps a pose $\mathbf{p} = (x_{head}, y_{head}, x_{middle}, y_{middle}, x_{tail}, y_{tail})$ to a predicted PTP $\mathbf{y} = f_{\text{phys}}(\mathbf{p})$. We invert by optimizing pose to match the measured pattern:

$$\mathbf{p}_t^* \;=\; \arg\min_{\mathbf{p}} \; \mathcal{J}(\mathbf{p}) \;=\; -\cos\big(\tilde{\mathbf{y}}(\mathbf{p}), \tilde{\mathbf{y}}_t^{\text{obs}}\big),$$

after an optional residual correction $\mathbf{y}' = \tilde{\mathbf{y}} + f_\phi(\tilde{\mathbf{y}})$ that compensates systematic mismatches between physics and measurements. We apply max-absolute normalization to both vectors before computing cosine: $\tilde{\mathbf{y}} = \mathbf{y}/\|\mathbf{y}\|_\infty$ and $\tilde{\mathbf{y}}^{\text{obs}} = \mathbf{y}^{\text{obs}}/\|\mathbf{y}^{\text{obs}}\|_\infty$. This scales each vector to unit max magnitude, removing overall amplitude so cosine compares only the spatial pattern (relative ratios across electrodes). The resulting pairs $\{(\text{frame}_t, \mathbf{p}_t^*)\}$ constitute pairs of a dataset for downstream pose trackers to be trained on (position, frame) data. Our methods label the emitting fish per frame.

The underlying assumption is that high cosine implies proximity to the true pose. For our purposes, we used a two-fish dataset with one silenced (non-emitting). Thus, we treat one fish, arbitrarily named Fish A, as the emitter and use the provided fixed electrode configuration across experiments.

Baseline: Trained neural inverter To test whether a compact data-driven mapping can recover pose from the 10-D pattern, we include a neural baseline: an MLP (see App. A) that maps the 10-D EOD to per-fish center, angle (sin/cos), and length.

2.2 PIKAChU: Inference via optimization

Our method inverts the physics-based forward model from Turcu et al. [10] which maps a position onto an electric field. Following this prior work, we model the fish as 101 point currents along its body axis (tail→head). Electrode potentials follow an inverse-distance law with first-order reflections for tank walls, and pairwise differences yield the predicted 10-D PTP. We score predictions with cosine similarity to the recorded signal.

Having defined our forward model, we now seek to invert it and identify a fish's position from its recorded electric field. We consider two complementary optimizers: we use L-BFGS-B (BFGS) as a fast local method when initialized near an accurate pose, and Differential Evolution (DE) as a global, derivative-free search to escape local minima. We observed that in practice, DE locates a basin and BFGS optionally polishes for speed. Our first optimization technique was **local center–orientation.** We minimize negative cosine similarity with L-BFGS-B over (x, y, θ), while holding length fixed. With a good initial guess, this method converges quickly, but it is susceptible to local minima when multiple poses result in similar signal patterns. We therefore also run DE over (x, y, θ) followed by some optional polishing, with the length as the median length of the emitting fish over all recorded frames. We test our inversion model on a synthetic position–orientation grid (details in App. A; see Fig. 3). In both cases the objective is:

$$\min_{\mathbf{p}} \ \mathcal{J}(\mathbf{p}) \ = \ -\cos\big(\tilde{\mathbf{y}}(\mathbf{p}), \tilde{\mathbf{y}}^{\,\text{obs}}\big)\,, \quad \mathbf{p} \in \{(x, y, \theta) \text{ or } (\mathbf{h}, \mathbf{m}, \mathbf{t})\}.$$

2.3 RAIChU: Matching on real data with priors

However, real measurements deviate slightly from the idealized model. We therefore add two complementary components to stabilize inference and reduce mismatch.

Residual normalizer. Real electrode measurements deviate systematically from idealized physics predictions due to hardware calibration, model simplifications (e.g., 2D approximations, wall reflection assumptions), etc. We learn a small residual map f_ϕ (denoting the function with parameters ϕ applied to input x) on normalized PTP that applies per-electrode corrections to physics predictions before scoring, compensating for these systematic mismatches. The **Normalizer+Inv** variant applies this correction before inversion, but we found that geometric priors (as in RAIChU) provide more robust improvements than calibration corrections alone. We theorize our inversion model converges on this position which produces the normalized EOD.

Geometric priors. The EOD pattern alone does not uniquely determine a pose. For instance, different configurations (e.g., head–tail flips or wall-related symmetries) can produce similar electrode patterns. To resolve these ambiguities, we regularize the search with weak anatomical and geometric cues on a head–middle–tail (abbreviated HMT) skeleton. The priors softly penalize implausible shapes and encourage consistency through: (i) the total length remains within a plausible range; (ii) the middle point lies about 0.4 of the head–tail axis from the tail; (iii) the three points are nearly colinear; and (iv) when available, the center and orientation stay near coarse estimates. These are soft penalties (not hard constraints), allowing the optimizer to deviate when the data demands it. We use a two-stage schedule: first an exploratory DE search at lower fidelity to locate a good basin, then a local refinement in HMT space at higher fidelity. Our composite objective is:

$$\min_{\mathbf{h},\mathbf{m},\mathbf{t}} \ \mathcal{J}_{\text{data}} + \lambda\Big[\big(\tfrac{L-L_0}{\sigma_L}\big)^2 + \big(\tfrac{p_\parallel - 0.4}{\sigma_{\text{ratio}}}\big)^2 + \big(\tfrac{p_\perp}{\sigma_{\text{col}}}\big)^2 + \big(\tfrac{\|\mathbf{c}-\mathbf{c}_0\|}{\sigma_c}\big)^2 + \big(\tfrac{\text{ang}(\mathbf{u},\mathbf{u}_0)}{\sigma_\theta}\big)^2\Big],$$

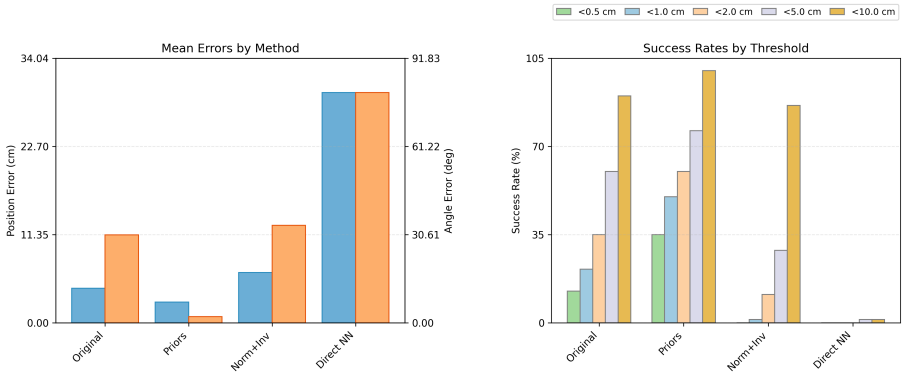

Figure 2: Our priors-augmented inverter RAIChU consistently outperforms the physics-only optimizer PIKAChU and other baselines, yielding lower mean position/angle errors and higher success rates across distance thresholds. The Normalizer+Inv variant provides smaller gains, while the Direct NN baseline underperforms. Exact metrics are in Appendix (Table 1).

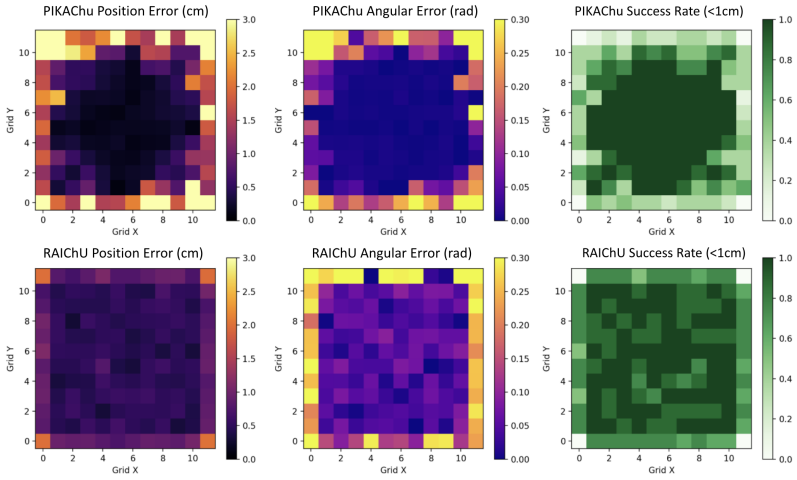

Figure 3: Synthetic consistency test on a 12×12 position grid (8 orientations). Heatmaps compare PIKAChU (top) versus RAIChU (bottom) for position/angle error and < 1 cm success. By injecting HMT priors, RAIChU improves accuracy near tank boundaries.

where $\mathbf{u} = (\mathbf{h} - \mathbf{t})/L$, $L = \|\mathbf{h} - \mathbf{t}\|$, $\mathbf{c} = (\mathbf{h} + \mathbf{t})/2$, $p_{\parallel} = \langle \mathbf{m} - \mathbf{t}, \mathbf{u} \rangle / L$ is the normalized axial position of the middle point, $p_{\perp} = \|(\mathbf{m} - \mathbf{t}) - \langle \mathbf{m} - \mathbf{t}, \mathbf{u} \rangle \mathbf{u}\|/L$ is its normalized off-axis distance, and $\mathrm{ang}(\cdot, \cdot)$ is the wrapped angular difference.

3 Results

We evaluate all methods on our `2-fish-1-silent` set: recordings of two fish in which only one of the fish is capable of discharging electricity, ensuring a ground truth EOD attribution. As seen in Fig. 2 and Table 1, the priors-augmented inverter RAIChU decisively outperforms the physics-only optimizer PIKAChU, the normalizer-based variant, and a direct neural baseline. The normalizer variant, which applies learned calibration corrections before inversion, provides smaller gains than RAIChU because it addresses only systematic mismatches without resolving geometric ambiguities (head–tail flips, wall symmetries) that require anatomical priors. Thus, the key difference is that RAIChU injects weak geometric priors (HMT) and a two-stage global-to-local search that break wall-induced symmetries and reduce implausible configurations. On a synthetic 12×12 grid (Fig. 3), these priors suppress large boundary errors from wall reflections and head–tail flips, yielding markedly higher success near tank edges.

4 Discussion

Physics-guided optimization with lightweight geometric priors yields reliable pose labels from electrode recordings. As Figs. 2–3 show, priors-aware HMT reduces error and boosts success with minor runtime overhead. Importantly, boundary improvements are ethologically relevant: weakly electric fish tend to swim along tank walls and objects, so gains near edges are likely to matter more in naturalistic settings than synthetic uniform grids.

In contrast to our approach of automatically annotation keypoint training data, other work aims to make manual keypoint annotation more data-efficient. Active learning approaches can partially mitigate the number of frames that need to be labeled, but this approach still requires careful human annotation, and its iteration speed is slow due to its human-in-the-loop approach [7].

4.1 Limitations and future directions

Our approach currently requires a known tank geometry and fixed electrode array configuration, as the physics model relies on boundary reflections and precise electrode positions. Extending to naturalistic environments would require: (i) simultaneous estimation of environmental geometry from EOD patterns, (ii) adaptive electrode calibration, and (iii) handling of dynamic obstacles or boundaries. However, the physics-guided framework provides a foundation for future extensions where learned boundary models could replace fixed geometries, and uncertainty-aware optimization could handle sensor calibration drift.

Our next steps fall into three directions. First, we will refine the forward model with better three-dimensional corrections. Second, we plan to extend from single-emitter inference to joint multi-fish inference. Third, we will incorporate uncertainty estimates so the system accepts only high-confidence labels from a baseline pose estimation tool, SLEAP. We expect this same pipeline to transfer to other bioacoustic/electric source-separation problems.

Acknowledgements:

Funded by NIH (RF1DA056403), James S. McDonnell Foundation (220020466), Simons Foundation (Pilot Extension-00003332-02), McKnight Endowment Fund, CIFAR Azrieli Global Scholar Program, NSF (2046583), Harvard Medical School Dean's Innovation Award, Harvard Medical School Neurobiology Lefler Small Grant Award, Alice and Joseph Brooks Fund Postdoctoral Fellowship (S.H.S.), Shanahan Family Foundation Fellowship at the Interface of Data and Neuroscience at the Allen Institute and University of Washington, supported in part by the Allen Institute (D.T.).

References

[1] P. C. Bermant. BioCPPNet: automatic bioacoustic source separation with deep neural networks. *Scientific Reports*, 11(1):23502, Dec. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-02790-2. URL https://www.nature.com/articles/s41598-021-02790-2. Publisher: Nature Publishing Group.

[2] A. A. Caputi. The electric organ discharge of pulse gymnotiforms: the transformation of a simple impulse into a complex spatio-temporal electromotor pattern. *Journal of Experimental Biology*, 202(10):1229–1241, May 1999. ISSN 0022-0949, 1477-9145. doi: 10.1242/jeb.202.10.1229. URL https://journals.biologists.com/jeb/article/202/10/1229/8157/The-electric-organ-discharge-of-pulse-gymnotiforms.

[3] I. Chrtkova, V. Koudelka, V. Langova, J. Hubeny, P. Horka, K. Vales, R. Cmejla, and J. Horacek. Unsupervised Approach for Electric Signal Separation in Gnathonemus petersii: Linking Behavior and Electrocommunication, Mar. 2025. URL https://www.biorxiv.org/content/10.1101/2025.03.04.641376v1. Pages: 2025.03.04.641376 Section: New Results.

[4] T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, and M. W. Mathis. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7):2152–2176, July 2019. ISSN 1750-2799. doi: 10.1038/s41596-019-0176-0. URL https://www.nature.com/articles/s41596-019-0176-0. Number: 7 Publisher: Nature Publishing Group.

[5] F. Pedraja and N. B. Sawtell. Collective sensing in electric fish. *Nature*, 628(8006):139–144, Apr. 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07157-x. URL https://www.nature.com/articles/s41586-024-07157-x. Publisher: Nature Publishing Group.

[6] T. D. Pereira, N. Tabris, A. Matsliah, D. M. Turner, J. Li, S. Ravindranath, E. S. Papadoyannis, E. Normand, D. S. Deutsch, Z. Y. Wang, G. C. McKenzie-Smith, C. C. Mitelut, M. D. Castro, J. D'Uva, M. Kislin, D. H. Sanes, S. D. Kocher, S. S.-H. Wang, A. L. Falkner, J. W. Shaevitz, and M. Murthy. SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19(4):486–495, Apr. 2022. ISSN 1548-7105. doi: 10.1038/s41592-022-01426-1. URL https://www.nature.com/articles/s41592-022-01426-1. Number: 4 Publisher: Nature Publishing Group.

[7] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang. A Survey of Deep Active Learning. *ACM Comput. Surv.*, 54(9):180:1–180:40, Oct. 2021. ISSN 0360-0300. doi: 10.1145/3472291. URL https://doi.org/10.1145/3472291.

[8] N. B. Sawtell, A. Williams, and C. C. Bell. From sparks to spikes: information processing in the electrosensory systems of fish. *Current opinion in neurobiology*, 15(4):437–443, 2005.

[9] I. Tolkova and H. Klinck. Source separation with an acoustic vector sensor for terrestrial bioacoustics. *The Journal of the Acoustical Society of America*, 152(2):1123, Aug. 2022. ISSN 0001-4966. doi: 10.1121/10.0013505. URL https://doi.org/10.1121/10.0013505.

[10] D. Turcu, A. N. Zadina, L. F. Abbott, and N. B. Sawtell. An end-to-end model of active electrosensation. *Current Biology*, 35(10):2295–2306.e4, May 2025. ISSN 0960-9822. doi: 10.1016/j.cub.2025.03.074. URL https://www.cell.com/current-biology/abstract/S0960-9822(25)00390-2. Publisher: Elsevier.

A Hyperparameters and configuration

A.1 Differential evolution testbed

Grid: 12×12 centers spanning the tank interior; 8 orientations $\{0, \pi/4, \ldots, 7\pi/4\}$. Bounds: $x, y \in [0, 60]$ cm with 2 cm margin; $\theta \in [-\pi, \pi]$. SciPy DE defaults unless stated: strategy=best1bin, popsize=50, maxiter=500, tol=10^{-10}, recombination=0.7, mutation in [0.5,1.0], workers=1; optional local L-BFGS-B polishing (maxiter=100). Objective: negative cosine similarity between max-abs–normalized predicted and recorded PTP.

A.2 HMT with priors

Penalty terms (weights in code units): length bounds/prior ($\sigma_L \in [0.3, 0.6]$ fraction of length), middle ratio ≈ 0.4 ($\sigma_{\text{ratio}} \in [0.03, 0.05]$), colinearity ($\sigma_{\text{col}} \in [0.03, 0.06]$), center prior ($\sigma_c \in [5, 10]$ cm), angle prior ($\sigma_\theta \in [12^\circ, 20^\circ]$). Two-stage schedule: stage 1 DE at reduced current count (e.g., 51) with search radius 6–10 cm; stage 2 local refinement at higher fidelity (e.g., 201 currents) with temporary combined loss then cosine for final scoring.

A.3 Neural inverse

MLP architecture: layers [128, 256, 512, 512, 256, 128] with BatchNorm and dropout 0.2; outputs per fish $(x, y) \in [0, 1]^2$, $(\sin\theta, \cos\theta)$ normalized, and length in $[0.25, 1]$ scaled to cm (tank=60, max length=20). Input normalization: per-electrode standardization if specified.

A.4 Results table

Table 1: Aggregate metrics on 500 emitting-fish frames. Success is percentage of frames with position error below threshold.

Method	Mean Pos (cm)	Mean Angle (deg)	< 0.5 cm	< 1 cm	< 2 cm	< 5 cm	< 10 cm
PIKAChU	4.43	30.49	12.50%	21.25%	35.00%	60.00%	90.00%
Priors HMT	**2.66**	**2.13**	**35.00%**	**50.00%**	**60.00%**	**76.25%**	**100.00%**
Normalizer+Inv	6.46	33.78	0.00%	1.25%	11.25%	28.75%	86.25%
Direct NN	29.60	79.86	0.00%	0.00%	0.00%	1.25%	1.25%

A.5 Detailed results for Fig. 2

Inversion on `2-fish-1-silent` (500 frames). Left: mean errors—position [cm] and angle [deg] per method. RAIChU: 2.66 cm / 2.13° vs. PIKAChU: 4.43 cm / 30.49° (Normalizer+Inv: 6.46 cm / 33.78°; Direct NN: 29.60 cm / 79.86°). Right: success = % frames with center error below threshold; at < 1 cm: 50.00% vs. 21.25% (RAIChU vs. PIKAChU), at < 2 cm: 60.00% vs. 35.00%, at < 5 cm: 76.25% vs. 60.00%. RAIChU lowers error and raises success; Normalizer+Inv helps less; Direct NN underperforms.

A.6 Tank screenshot & data collection

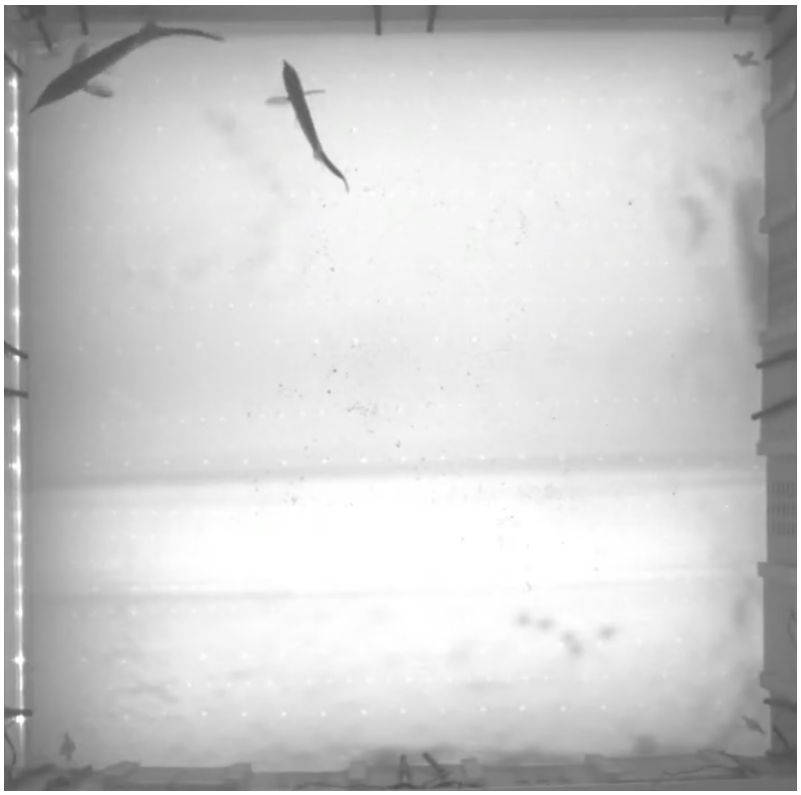

Figure 4: A screenshot from the tank from which our data was collected. The dark vertical slabs along the wall are the electrode sensors.

Experiment setup: We recorded from *Gnathonemus petersii* (elephantfish, 5–15 cm body length) in a square acrylic tank (60 × 60 × 10 cm) filled with freshwater (conductivity: 100–300 μS/cm). Ten electrode pairs positioned around the tank perimeter measured potential differences during EOD discharges, sampled at 25 time points per electrode to capture the biphasic waveform. For ground-truth EOD attribution, we used a 2-fish-1-silent dataset where one fish (Fish A) emitted EODs while the other was silenced. Synchronized video recordings (912 × 912 pixels, 15.2 pixels/cm) provided 2D position tracking via SLEAP [6], yielding head, middle, and tail keypoints. The evaluation dataset consisted of 500 emitting-fish frames.