Global Underwater Geolocation from Time-Lapse Polarization Imagery

Sara Aghajanzadeh[1] Xiaoyang Bai[2] Zhongmin Zhu[1] David Forsyth[1] Viktor Gruev[1]
[1]University of Illinois Urbana-Champaign [2]The University of Hong Kong

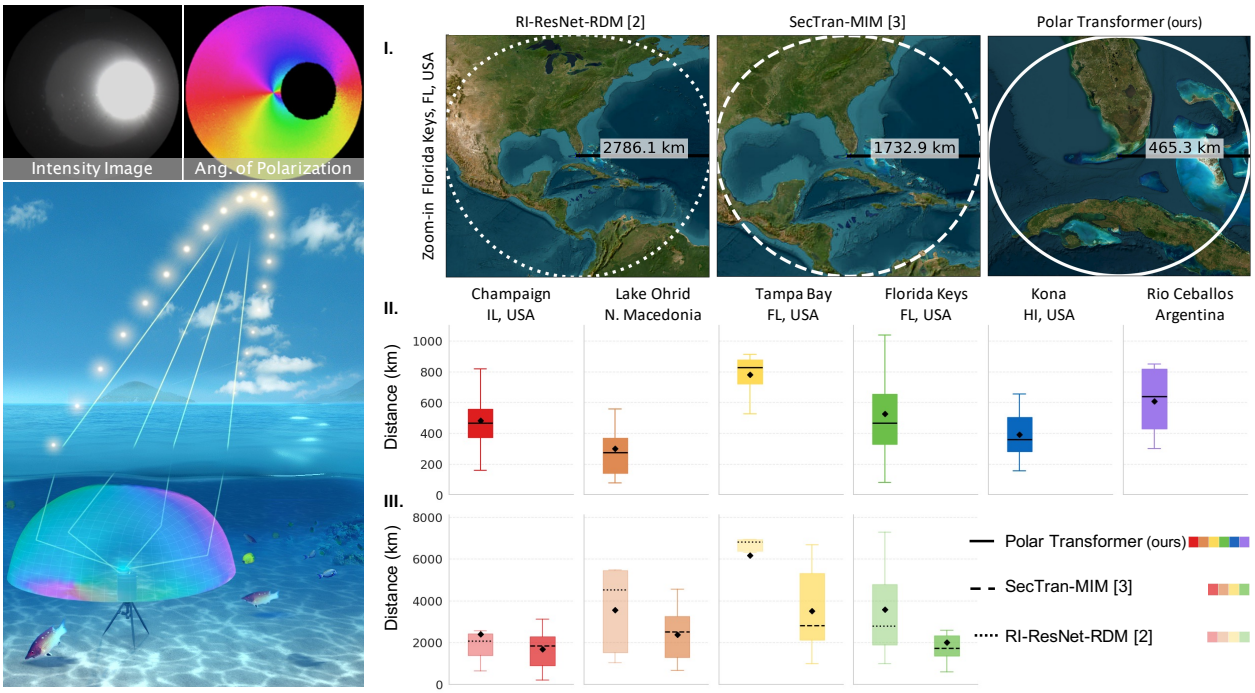

Figure 1. **Left:** Our method takes a sequence of underwater images of the sky, using the polarization pattern, computes the solar elevation curve (extremely challenging, Fig. 4) and from it, computes geolocation. **Right:** Our method exhibits very substantial improvements over SOTA in **cross-site geolocation accuracy**. Colors denote test sites; line style denotes methods (solid: ours; dashed: SecTran-MIM [3]; dotted: RI-ResNet-RDM [2]) **I**, Florida Keys zoom (other sites in Supplementary Fig. S12); circle radii equal each method's median geodesic error (kilometers along the Earth's surface); all maps share a common geographic scale. The full distribution of errors for our method in **II** and the baselines in **III**; note the horizontal axis is expanded *8-fold* to accommodate their larger errors. Black diamonds mark site-averaged performance. Across the six sites, our model yields median errors of 300–800 km, whereas SecTran-MIM and RI-ResNet-RDM range from 1,700–6,800 km across the four sites they report. Baseline outliers beyond the axis limits are omitted for clarity.

Abstract

It is extremely hard for an underwater agent to know where it is. Satellite signals disappear within centimeters of the surface; acoustic baselines require heavy infrastructure to instrument small regions. The polarization of the sky, visible underwater, reveals the elevation of the sun. The pattern of elevation over the day reveals location to an agent with a clock. However, recovering elevation from polarization images is very difficult. State-of-the-art (SOTA) geolocalization methods can localize well for locations where they have seen data*, but accuracy collapses when the data comes from a new location. Our physics-guided synthesis pipeline expands a huge library of polarization images from a small set of sites to 2.8 million solar-elevation–matched training sequences spanning latitudes, seasons, and water types. A compact two-stage transformer reconstructs the solar-elevation curve and predicts geolocation. Under leave-one-site-out tests, the site-averaged median geodesic error is ∼500 km—about an eightfold improvement over previous deep-learning baselines (Fig. 1); with limited target-site data, the median error contracts to single-digit kilometers.*

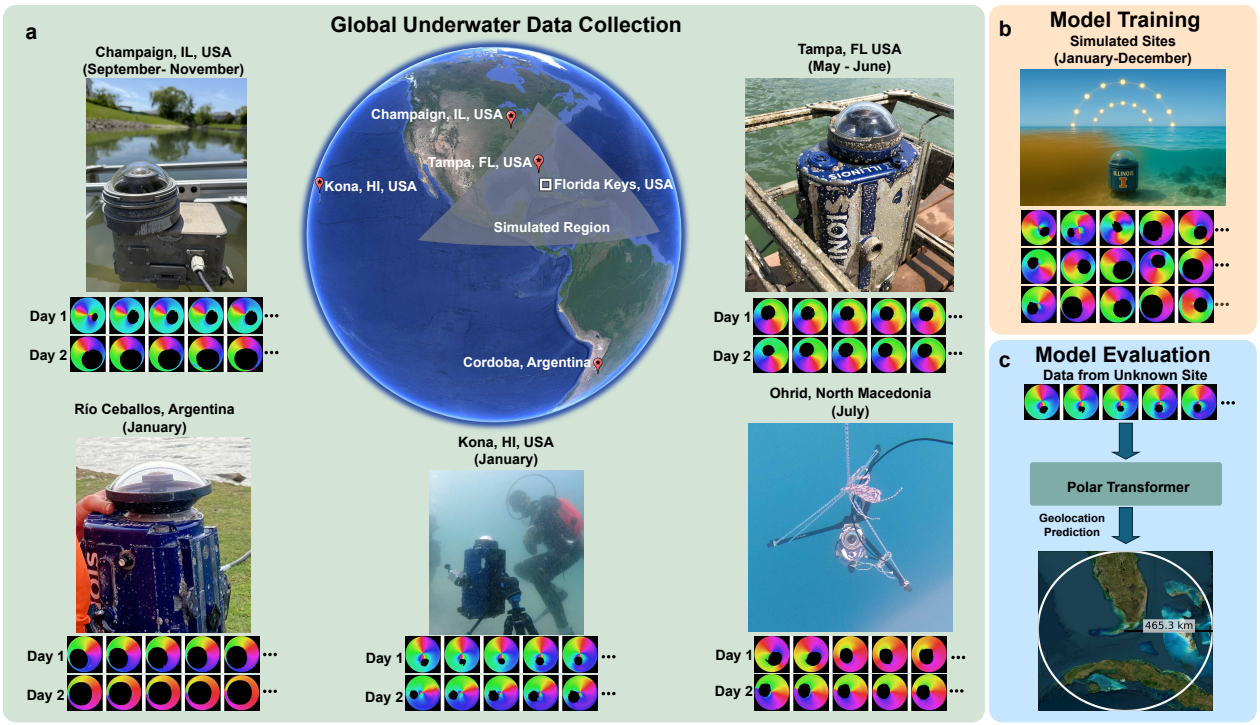

Figure 2. **Physics-guided simulation and transformer model for polarization-based underwater geolocalization. a,** Time-lapse polarization videos from five optically diverse sites—Río Ceballos (Argentina), Champaign (USA), Tampa Bay (USA), Lake Ohrid (North Macedonia), and Kona (USA)—are converted to angle-of-polarization (AoP) images and sorted by solar elevation. **b,** For each random target point on a global simulation grid, the synthesizer concatenates 64 AoP frames that share successive solar elevations while varying heading and water type, yielding a realistic sequence annotated with Coordinated Universal Time (UTC). Simulated sequences are used to train the Polar Transformer which first reconstructs the solar elevation curve and then regresses latitude and longitude. **c,** At inference, the model ingests real sequences (AoP frames, UTC timestamps, date) from an unseen site (example: Florida Keys) and estimates its geographical coordinates.

1. Introduction

It is hard to know where you are if you are underwater—GPS signals are not available. One important cue is in the light from the sky, the polarization pattern can yield solar elevation. A sequence of solar elevations and time can reveal geolocation [32]. SOTA [2, 3] geolocations have uncertainties in the large 3,000 km error. We describe a method that offers an order of magnitude improvement over SOTA.

A core technical challenge is that the data are both dense and sparse. We have a vast number of polarization frames captured underwater—dense in volume—but sparse in location. Our data is collected at few locations for practical reasons (Fig. 2). SOTA methods [2, 3] have great difficulty predicting geolocation at new sites. Our dataset is simultaneously very dense in observations, and very sparse in locations. Sparsity in locations matters because effects that make predicting solar elevation from the frames challenging are somewhat linked to location. The same elevation can produce very different images, as a result of weather, marine animals, water turbidity and the like (Fig. 4). We show how to manage the sparsity in locations using a simulation procedure that exploits an important physical property: the prime predictor of a polarization image of the sky obtained underwater is solar elevation (Fig. 3). Our procedure uses polarization time-lapse videos captured at a handful of optically diverse sites to synthesize millions of sequences for locations and times never visited. Simulations mix observations across sites and so each synthetic sequence in our training corpus corresponds to a sequence of elevations that would be observed at some site – a physically valid solar trajectory – but mixes observations across source sites and so treats water, marine animals and weather as nuisance parameters which are then averaged out by the predictor. As a result, our trained predictor exhibits geolocation accuracy and generalization very significantly higher than current SOTA.

Our key contributions are: (1) Our data-driven synthesis leverages real sequence statistics to produce diverse, realistic training data. (2) We describe a model that estimates solar elevations from angle-of-polarization (AoP) frames, enabling accurate geolocalization. To enhance robustness, we design a sequence dropout regularization strategy that

trains the model to operate effectively on incomplete or short time-lapse sequences, as often encountered in real-world deployments. (3) We demonstrate substantial performance improvements over existing SOTA polarization-based geolocation methods. (4) We release a new benchmark dataset to support future research.

2. Background

Geolocation for an agent underwater in the ocean is very hard. Satellite-based navigation signals vanish within centimeters of the surface [25]. Marine exploration agents – with applications ranging from carbon-cycle monitoring to defense – must use terrain-matching sonars or acoustic baseline networks that are costly to deploy and can localize accurately only within instrumented zones that are typically less than 10 km across and are costly to deploy [29, 31, 36].

Animals use natural polarization patterns —most prominently Rayleigh-scattered skylight— to geolocate and navigate [12]. Animals (from dung beetles to greater mouse eared bats) sense the sky's e-vector pattern to maintain headings and calibrate multisensory compasses using specialized neural circuitry [13, 18, 20, 23, 28, 34, 41]. Engineering has paralleled biology: bio-inspired skylight compasses and robots achieve Global Navigation Satellite System (GNSS)-free heading and homing with polarization sensors. Geodesy methods can even recover the celestial pole from the sky's polarization field [17, 19, 21, 27, 37].

Distinctive polarization patterns are visible underwater because refraction and multiple scattering generate a partially polarized field throughout the photic zone. Many marine animals exploit this pattern—cephalopods, stomatopods, crustaceans, and pelagic fish use polarization for orientation, foraging, and camouflage [4, 22, 24, 30]. Because the pattern encodes the Sun's daily arc, it also contains information about geographic position. Sunlight provides an infrastructure-free alternative for underwater orientation, geolocation, and navigation as well [35, 39, 40].

A simple, effective model of visible polarization underwater rests on a first-order optical rule: for a camera that samples an entire radial polarization field, the AoP pattern is governed primarily by the Sun's elevation; local water optics (e.g., turbidity and color) scale contrast, and solar heading contributes only a rigid in-plane rotation. Parametric modeling and multi-site measurements across latitudes, headings, and elevations confirm this behavior (Supplementary Notes 2-3). There are two steps to our model of underwater polarization. First, nearly parallel sunlight refracts at the air–water interface; the transmitted field's polarization is set by Snell's law and the Fresnel transmission coefficients, $t_s(\theta_i)$ and $t_p(\theta_i)$, and therefore depends only on the angle of incidence θ_i—i.e., solar elevation—and the refractive indices of air and water. Second, the transmitted beam undergoes predominantly single-Rayleigh scattering from suspended particles; the scattering angle χ directs photons toward the lens, and the AoP is perpendicular to the scattering plane. In short, geometry fixes orientation while water optics scale contrast.

Every pixel in an underwater sky image contains elevation information in our model. The camera looks largely along the normal to the earth (upwards; at the sky). The refraction plane contains the incident and refracted rays; the scattering plane contains the scattered ray and the camera's optical axis. With the image plane parallel to the surface (in-plane x–y axes, z-axis upward) and a fisheye lens that records the full 360° azimuth, every scattered direction maps somewhere on the sensor (Fig. 1). Changing solar elevation alters the incidence angle at the interface and therefore the polarization state impressed on every scattered photon, so the entire AoP pattern changes. This effect means that every pixel in the image has information about the solar elevation. Changing solar heading simply rotates the refraction plane about the optical axis, producing a rigid in-plane rotation of the AoP map without altering its radial structure.

Experimental observations confirm that (a) the whole AoP pattern depends on elevation and (b) changing solar heading just rotates the image. Dawn-to-dusk time-lapses were captured at turbid Río Ceballos (Córdoba, Argentina) and ultra-clear Lake Ohrid (North Macedonia). Fig. 3 shows three elevations on the morning ascent ($25^\circ, 45^\circ, 65^\circ$) and the same three on the afternoon descent (more in Supplementary Fig. S1). Each morning–afternoon pair shares elevation but differs markedly in heading. After rigidly rotating each AoP frame by the solar azimuth computed from the frame's geodetic coordinates and UTC timestamp, the maps recorded at the same elevation collapsed onto a single radial pattern. AoP traces along the outer ring coincided when elevation matched, whereas curves from different elevations remained distinct. Patterns from the two sites overlapped closely, with differences confined to turbidity-driven contrast variations that altered intensity and DoLP but not AoP orientation (Supplementary Figs. S1–S3).

Related work demonstrates optical compasses [21], but geolocation and navigation have been harder [7–10, 32, 38, 42]. Powell *et al* demonstrate a model that can locate an observer to within about 2,000 km, but which degrades sharply under low solar elevation or in turbid waters [32]. Deep neural networks trained on ten-million-frame datasets reduced median error to roughly 400 km *at the training sites* [2]. The key challenge is that it is wholly impractical to collect data from a very large number of sites (say, on a 100km grid over the ocean). But current methods generalize poorly to unseen locations. SOTA median error balloons to about 3,000 km when evaluated in unvisited waters [3].

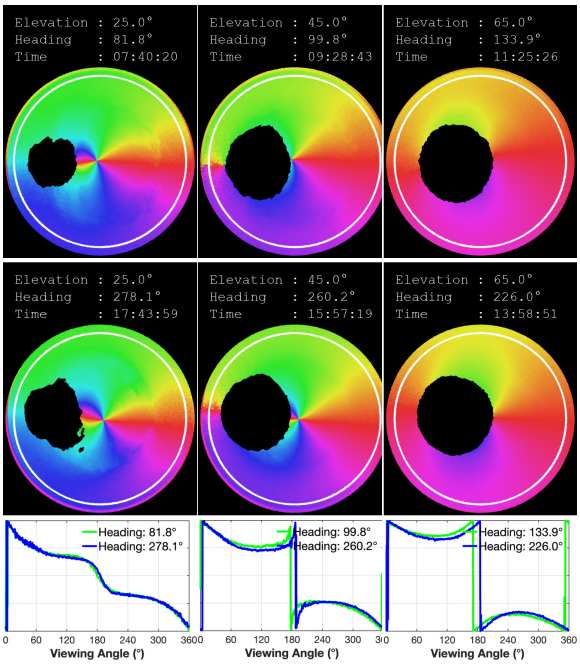

Figure 3. **Solar elevation—not heading—shapes the underwater angle-of-polarization.** AoP maps and profiles from a clear-water site in Lake Ohrid are shown. **Top:** AoP maps at three morning solar elevations (25°, 45°, 65°). **Middle:** the same elevations recorded hours later, after the Sun's azimuth (heading) has changed. All frames are rotated by the measured solar heading to share a common reference direction. **Bottom:** AoP profiles sampled along the white ring: profiles taken at the same elevation (matching colors) coincide, whereas profiles from different elevations remain distinct.

3. Methodology

There is strong evidence that inferring solar elevation from underwater images is very difficult (Fig. 4). Instead of pushing direct improvements on this task, we hypothesize that training on sequences from many scattered locations should yield significant geolocation gains. Since collecting such real data is impractical, we simulate it.

3.1. Physics-guided synthesis of a global training set

Rather than rendering scenes with full radiative-transfer solvers, we start from polarization time-lapses collected at a handful of optically diverse waters and splice frames to form new sequences (Fig. 2). To simulate a sequence, we choose a location on earth and a day of the year, then use first-order radiative-transfer constraints and a software ephemeris, to construct a sequence of observations appropriate for that location and that day. The simulator (i) controls elevation explicitly by selecting frames at desired θ_i; (ii) breaks nuisance correlations by randomizing heading and applying rigid in-plane rotations; (iii) recombines frames from optically distinct sites at matched elevations, so treating water physics as a nuisance parameter; and (iv) varies daylight coverage (morning-only, afternoon-only) to train robustness to partial arcs.

Base data for the simulator consists of dawn-to-dusk time-lapse videos from six widely separated and optically diverse sites: Lake Ohrid, North Macedonia (July); Champaign, Illinois, USA (September–November); Florida Keys, USA (December–January); Tampa Bay, USA (May–June); Río Ceballos, Córdoba, Argentina (January); and Kona, Hawaii, USA (January). These locations span more than an order of magnitude in water clarity—visibility exceeds 10 m in Lake Ohrid but can fall to $\sim$0.3 m during the most turbid hours in Champaign—providing the optical diversity needed for generalization. Tampa Bay exhibits daily swings in visibility (mean $\approx$0.9 m), Córdoba remains near 2 m, and the Florida Keys and Kona fluctuate between roughly 0.3 m and 5 m with the tides. Every frame is tagged with its UTC timestamp and binned by solar elevation; these annotated clips become the building blocks the simulator recombines into synthetic sequences that cover unvisited places and dates.

We acquired time-lapse polarization videos with a FLIR Blackfly S monochrome polarization camera and a Fujinon FE185C057HA-1 fisheye lens, mounted on a rigid aluminum rig or tripod, recording 20–40 s clips. To limit redundancy, we retained one frame per clip. Raw $2{,}048 \times 2{,}448$ sensor images were demosaiced into four analyzer channels $(0°, 45°, 90°, 135°)$, downsampled $4\times$, and stacked. AoP images were computed from Stokes S_0, S_1, S_2 and then radially calibrated to remove the fisheye offset (Supplementary Note 1).

Training locations are sampled uniformly over Earth's surface. Because the planet is an oblate spheroid, one degree of longitude spans $d_\lambda \approx 111.32 \cos\phi$ km at latitude ϕ, whereas one degree of latitude is nearly constant at about 111 km; a simple latitude–longitude grid would therefore overweight high latitudes. We instead adopt the area-preserving sampler of Arvo [1], which maps stratified points from the unit square onto an arbitrary spherical triangle through an area-preserving bijection.

Seasonal coverage is important because the sun drifts over the year (Supplementary Note 5, Fig. S5), so we augment spatial sampling with a temporal dimension. The year is divided into four-day bins; from each bin we pick one random day, ensuring that successive trajectory samples differ by no more than seven days—a span over which the solar path changes only slightly.

Ephemeris: For every sampled location and selected day we compute the solar-elevation angle at one-minute resolution from sunrise to sunset using Astropy [11], which implements IAU-standard solar algorithms.

Assembly of training sequences proceeds by pairing

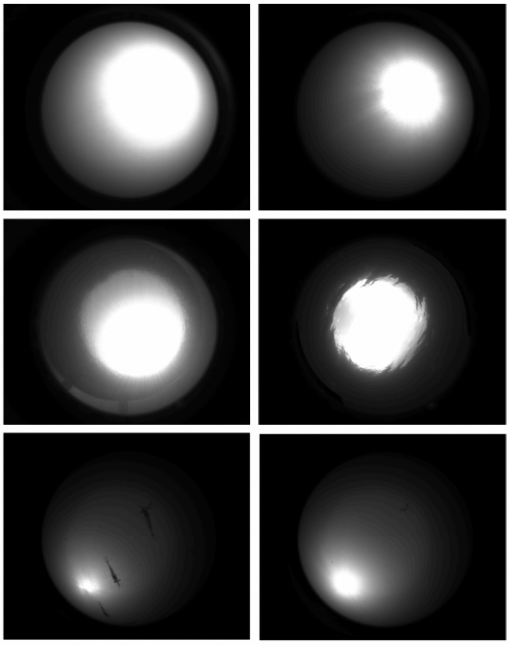

Figure 4. **Predicting independent solar elevations from individual observations is challenging. Top:** Two observations with different exposures but identical solar elevations. **Middle:** Two observations at 2 PM in Lake Ohrid, Macedonia—left: July 26 (sunny), right: July 28 (cloudy, turbid). **Bottom:** Observations at 10:22 AM and 10:35 AM with nearly identical solar elevations, one is partially occluded by marine animals.

each point in a solar trajectory (elevation angle against time) with an AoP frame. We maintain a pool of real polarization images from the empirical library. For every time step we search this pool for the five frames whose solar elevations most closely match the desired value and randomly pick one. Repeating this match-and-select procedure across the entire trajectory yields a candidate sequence. Trajectories are stratified to 64 evenly spaced samples, adjusting the spacing so that every daylight span contributes 64 points regardless of length (mean spacing ≈10 minutes). Because the pool of images is large, a single trajectory can be instantiated many times by re-drawing the random matches, providing useful stochastic augmentation while preserving physical consistency and diversity in heading and water type. The resulting 64-frame AoP sequences span latitudes, seasons, and optical regimes without requiring exhaustive field collection. **When a site served as the cross-site test set, its entire library was excluded from synthesis, training and validation.**

The synthesis procedure preserves the geometry identified in Sec. 2 while eliminating two key biases. Across sites, the radial AoP pattern evolves smoothly with solar elevation, yet its geometry is unchanged by heading or sensor orientation. Accordingly, for a given elevation the simulator chooses frames from random headings, so the masked solar patch drifts around the image. It also mixes water types, preventing the network from keying onto a single optical regime (Supplementary Fig. S4 for a visual explanation of the simulation). Supplementary analyses quantify these claims: simulated sequences reproduce target elevations to within 0.01°, and the final dataset contains a balanced number of images from each site, delivering the diversity and uniformity required for robust, unbiased training (Supplementary Note 6; Supplementary Figs. S6–S11).

Our aim is to train a model to recover an entire solar trajectory from a short sequence of polarization images rather than from isolated frames. Sequences offer three complementary advantages. First, temporal context makes the trajectory estimate more accurate and more tolerant of outliers. Second, redundancy across successive frames suppresses sensor noise. Third, temporally ordered data align naturally with transformer architectures—the state of the art for sequence learning across natural language [5, 14], audio [15], and, increasingly, computer vision and robotics [6, 26, 33].

3.2. Polar Transformer

Our model estimates the solar curve, leverages temporal context, and the attention mechanism in transformers [16, 43] that allows every point to attend to every other, so the network can reason about global properties of the daily arc—peak height, symmetry, and slope—rather than making independent frame-by-frame guesses as in prior deep models [2, 3], enforcing physical consistency and improving generalization.

Objective. Given a sequence of 64 AoP frames $\mathcal{X} = \{x_i\}_{i=1}^{64}$ with UTC timestamps $\mathcal{T} = \{t_i\}_{i=1}^{64}$ acquired on calendar day d, the model first predicts the solar-elevation curve $\hat{\mathbf{s}} = (\hat{s}_1, \ldots, \hat{s}_{64})$ and then regresses geographic coordinates $(\hat{\phi}, \hat{\lambda})$, where $\hat{\phi}$ and $\hat{\lambda}$ denote estimated latitude and longitude.

Input encoding. For each frame x_i we extract a spatial descriptor

$$h_i = \mathrm{CNN}(x_i) \in \mathbb{R}^H,$$

where the CNN is a shallow convolutional neural network. We encode the intra-day UTC timestamps as seconds past midnight and normalized,

$$\tilde{t}_i = \frac{\text{seconds since midnight}(t_i)}{86{,}400} \in [0, 1],$$

and the day-of-year (season) as a periodic two-vector,

$$e(d) = \big[\sin(2\pi\,\tilde{d}),\ \cos(2\pi\,\tilde{d})\big], \quad \tilde{d} = \frac{\text{day index}(d)}{D} \in [0, 1].$$

where $D \in [1, 365]$ (366 in leap years). We concatenate these with the spatial feature to form an input token

$$z_i = \big[h_i\ ;\ \tilde{t}_i\ ;\ e(d)\big] \in \mathbb{R}^{H+3},$$

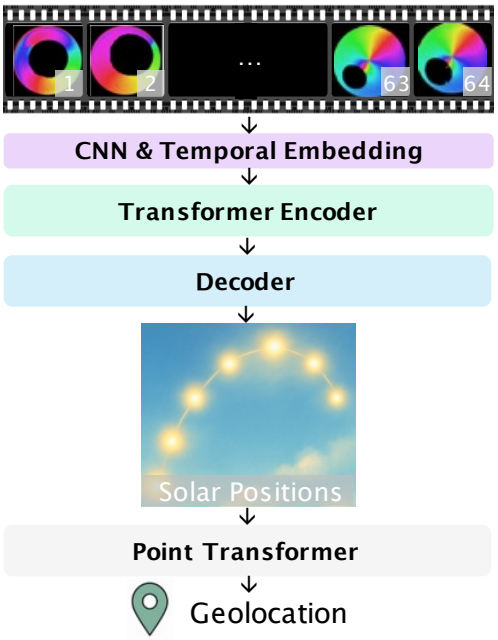

Figure 5. **Model.** A sequence of AoP frames and temporal data are embedded and processed by a transformer encoder to summarize the sequence. The input and the summary token are decoded by an MLP head to predict the solar-elevation curve. The estimated curve and temporal context are treated as a "point cloud" and fed to a point transformer, which regresses the geographical coordinates (ϕ, λ) of the camera.

add a learned 64-position embedding p_i, and feed the sequence $Z = \{z_i + p_i\}_{i=1}^{64}$ to the transformer encoder.

Solar-elevation module. A vision transformer encoder [16] $\mathcal{T}_\theta$ processes the 64 tokens and outputs a global state, summarizing the sequence. A multilayer perceptron (MLP) decodes the input features and summarizing state into a smooth solar elevation curve $\hat{\mathbf{s}} = (\hat{s}_1, \ldots, \hat{s}_{64})$.

Geolocation module. We treat the predicted curve $\hat{\mathbf{s}}$ and temporal context $\left[\, \tilde{t}_i \;;\; e(d) \right] \in \mathbb{R}^3$, as a "64-point cloud" and process it with a point transformer [43] $\mathcal{P}_\phi$ to regress latitude and longitude coordinates.

Loss function. We train end-to-end using an elevation MSE term

$$\mathcal{L}_{\text{MSE}} = \tfrac{1}{64} \sum_{i=1}^{64} \| s_i^{\text{gt}} - \hat{s}_i \|_2^2,$$

and a cosine loss on predicted vs. true location vectors,

$$\mathcal{L}_{\cos} = 1 - \langle \hat{c}, c \rangle,$$

where $\hat{c}$ and c are unit-norm Cartesian coordinates. The total loss is

$$\mathcal{L} = \lambda_{\text{elev}} \mathcal{L}_{\text{MSE}} + \lambda_{\text{geo}} \mathcal{L}_{\cos}.$$

Model and optimization details appear in Supplementary Note 7.

Augmentation strategy. Since real timelapse sequences used at test time are often incomplete (i.e., they do not cover full period from dawn to dusk), while our simulated training sequences are complete, a model trained only on complete sequences would fail to generalize well. To mitigate this mismatch, we introduce a trajectory dropout regularization strategy that randomly removes continuous portions of the input sequence during training. This encourages robustness to missing temporal segments and improves generalization to real-world incomplete sequences. Details about the real test sequences are provided in Supplementary Note 10 and Fig. S16.

4. Result

Setup. At each site, the empirical library is ordered chronologically and split 85%/15% into training and validation; when a site serves as the cross-site test set, its entire library is excluded from synthesis, train and val for every method. We compare against strongest published methods using their reported results, obtained under the same dataset and experimental setting. Note baseline numbers for Río Ceballos and Kona are not available.

Cross-site geolocation. Fig. 1 summarizes cross-site performance with each of the six locations held out in turn. Panel **I** zooms into the Florida Keys so the extent of errors (i.e. geodesic circle of error) is visible. Panels **II-III** show ours and baselines' error distributions.

With the Florida Keys withheld, our model attains a median geodesic error of 465 km, compared with 1,733 km for SecTran-MIM and 2,786 km for RI-ResNet-RDM. Averaged over the six leave-one-out trials, the *site-averaged* median geodesic error is 513 km. By contrast, the strongest published baselines report 2,394 km and 3,971 km on the four sites they evaluate—an ∼eightfold gap. Zoomed views for all test sites are provided in Supplementary Fig. S12. Directional-error analysis appears in Supplementary Fig. S13 for North-South error and Fig. S14 for East-West error.

Same-site geolocation. We next examine the simpler—yet operationally important—case in which the model is trained and tested at the same location and season, shown in Fig. 6.

For the same-site study we examined three variants. Variant *a* trained on the real 85% split and was evaluated on the held-out 15%. Variant *b* trained on simulator sequences that could include frames from the target site, while variant *c* excluded all target-site frames during synthesis. Both simulator-trained models perform on-par with variant *a*, indicating that physics-guided synthesis supplies sufficient optical diversity, enough to replace on-site collection. Our model (variant *a*) attains a *mean* geodesic error of 9 km averaged across Lake Ohrid, Champaign, Tampa Bay, the Florida Keys, Río Ceballos, and Kona. By con-

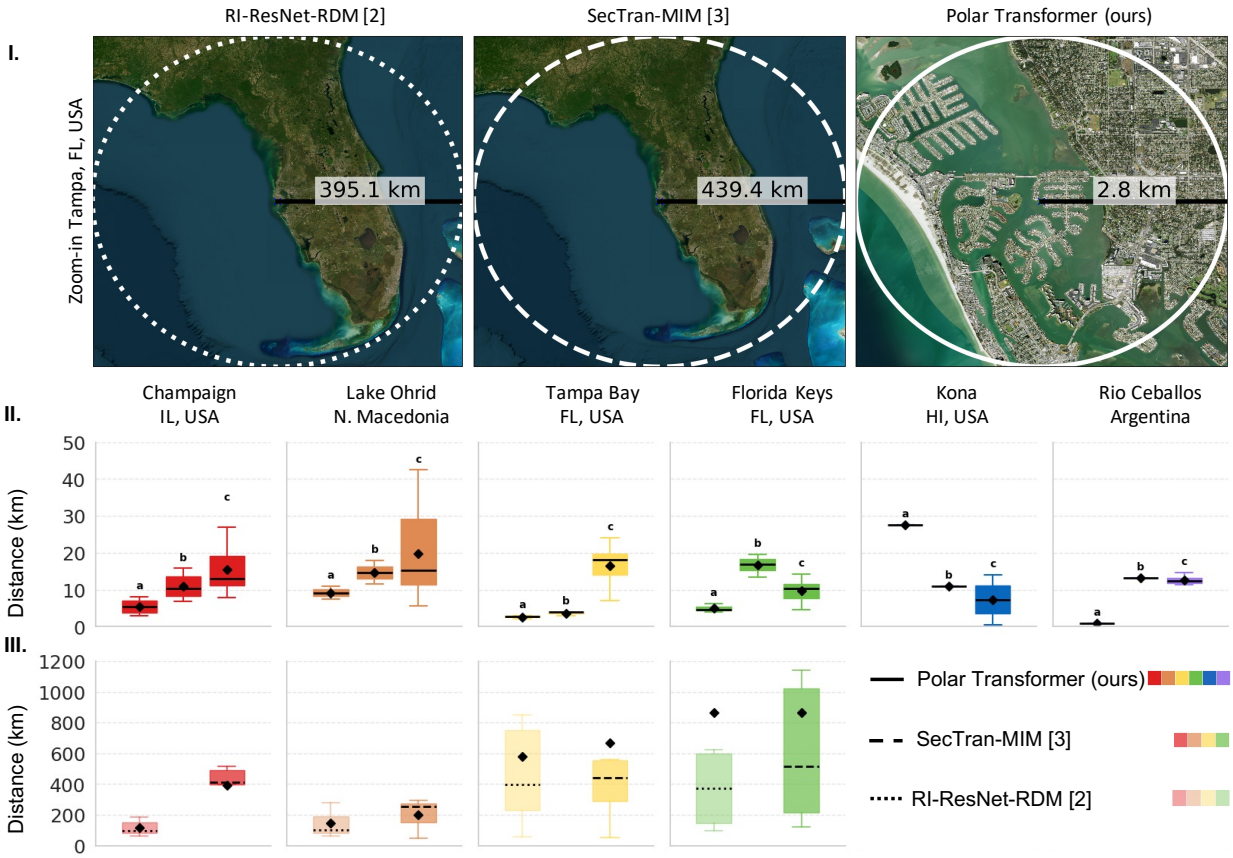

Figure 6. **Same-site geolocation accuracy.** **I**, Tampa Bay zoom comparing RI-ResNet-RDM [2], SecTran-MIM [3], and our model; each dashed or solid circle shows that method's median geodesic error (zoomed maps for all sites in Supplementary Fig. S15). **II**, error distributions for our model three variants; variants b and c—trained on synthetic sequences with and without target-site frames—perform on par with variant a. Black diamonds mark site means. **III**, corresponding distributions for the baselines; the horizontal axis is expanded *24-fold* to accommodate their larger errors. Across all six waters, our method averages 9 km, whereas SecTran-MIM and RI-ResNet-RDM reach 427 km and 530 km on the four sites they report. Baseline outliers beyond the axis limits are omitted for clarity.

trast, the strongest published networks—RI-ResNet-RDM and SecTran-MIM—report *median* errors of 427 km and 530 km, respectively, and only on four of the six sites. The negligible gap between variants *b* and *c* suggests little benefit from target-site exposure, though sites with distinctive water or weather might still gain from it.

Solar-elevation accuracy. Fig. 7 benchmarks how well each model estimates the solar-elevation curve that underpins geolocation. In the same-site setting (top panel), all three variants remain in the single-degree regime: the *median* RMSE ranges from 0.3° at Champaign to 3.6° at Río Ceballos, with a six-site mean of 1.3°. The strongest published baselines are less precise, averaging 3.8° for RI-ResNet-RDM and 4.7° for SecTran-MIM. Notably, the two models trained purely on synthetic data (variants *b* and *c*) slightly outperform the real-data model (variant *a*), indicating that physics-guided synthesis provides a richer diversity of elevation–pattern pairings than any single site.

Cross-site evaluation (bottom panel) is more demanding, yet our method degrades gracefully: RMSE peaks at 9° in turbid Tampa Bay but still averages 4.5° across the six unseen waters. By comparison, RI-ResNet-RDM rises to 18° on average, while SecTran-MIM exceeds 13° and reaches $\approx 24°$ at ultra-clear Lake Ohrid. Thus, even when trained exclusively on other locations, our model maintains sub-5-degree accuracy—tight enough to bound global position within a $\sim$500 km geodesic circle—whereas existing models can miss the solar trajectory by tens of degrees, precluding useful geolocation.

Ablation study. To isolate the contribution of each design choice, we present ablation study on the *cross-site* task in Champaign test site (Fig. 8). Everything is fixed; only the component under test is modified. Starting from a baseline (Index 1) that ingests AoP frames with time embeddings, adding a day-of-year token (Index 2) lowers RMSE from 6.5° to 5.2° and narrows the geodesic error distribution, indicating more consistent predictions.

Replacing *relative* with *absolute* attention in Point

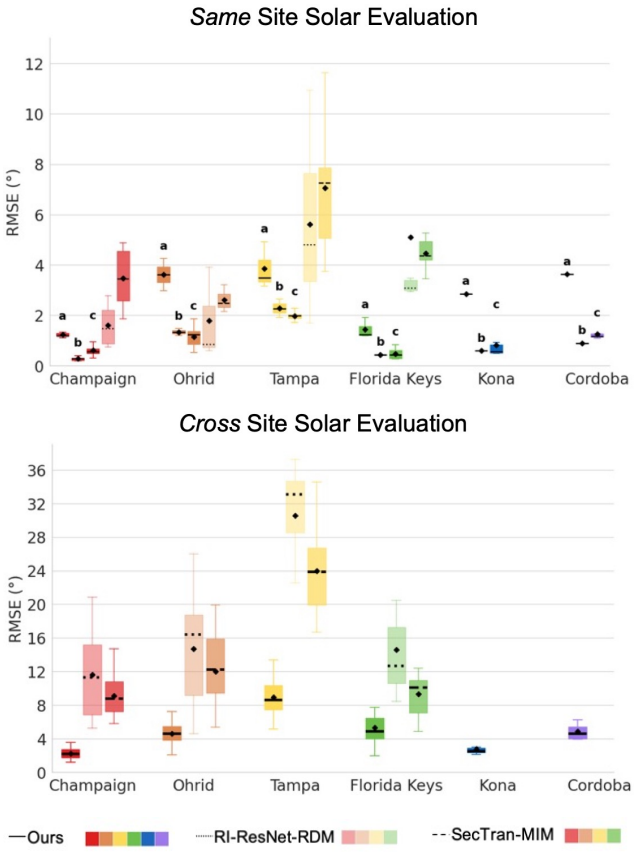

Figure 7. **Solar-elevation accuracy.** Root-mean-square error (RMSE, degrees) of estimated solar elevations—lower values indicate better fits. **Top:** same-site evaluation. **Bottom:** cross-site evaluation. In same site evaluation, our model is shown in three training modes: (a) real train sequences from the target site; (b) simulated sequences that may include the target site frames; and (c) simulated sequences that omit target site frames. Our model consistently outperforms the strongest published baselines (RI-ResNet-RDM and SecTran-MIM). Black diamonds indicate site-averaged errors.

Transformer blocks (Index 3) improves all metrics: RMSE drops to $2.47°$ and median distance from $845\,\mathrm{km}$ to $567\,\mathrm{km}$ ($\sim 33\%$ gain), showing that absolute attention offers a more stable global reference when point ordering matters. Introducing trajectory-dropout regularization (Index 4) further reduces the median distance to $466\,\mathrm{km}$ and RMSE to $2.2°$. Qualitative improvements are shown in Supp. Fig. S17.

Doubling the sequence length (Index 5) from 64 to 128 samples slightly *increases* RMSE ($3°$) and yields marginal geolocation gains while doubling cost; thus, we retain 64 samples in all experiments.

Removing the ViT encoder (Index 6) deprives the decoder of global context (i.e. summary token), producing jagged trajectories (Supplementary Fig. S17) and higher RMSE, and although the *median* geodesic error improves,

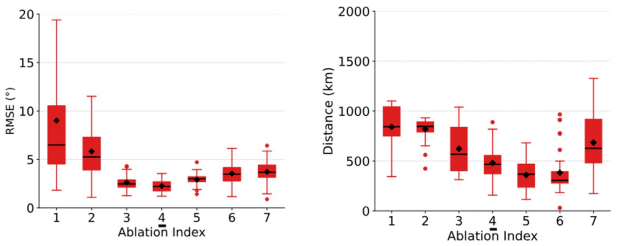

Figure 8. **Ablation study.** Impact of individual components on cross-site generalization (Champaign test site). Box-and-whisker plots summarizing model performance with each configuration in the ablation study.

the distribution develops outliers. Replacing the Point Transformer with an MLP (Index 7) degrades all metrics, raising geolocation error beyond $620\,\mathrm{km}$. These confirm that self-attention across both input sequence and point cloud (estimated solar curve and temporal context) is critical.

5. Conclusion

We present a camera-only framework that achieves practical underwater geolocation at both local and global scales. Trained entirely on physics-guided synthetic sequences—and requiring at inference only polarization time-lapse imagery and UTC timestamp—our approach dispenses with auxiliary sensors (e.g., inertial units, Doppler logs) while outperforming published baselines. With the same-site model variants, we show comparable precision persists when training uses only simulated sequences that exclude frames from the target site, indicating that the physics engine captures the elevation-driven structure and that moderate differences in local optics are second-order. Ablations show that day-of-year encoding, absolute attention across elevation samples, and dropout-style trajectory masking are all critical; removing any one increases error.

Limitations and opportunities remain. Performance tapers once the prior simulation region exceeds $2 \times 10^6\,\mathrm{km}^2$ (Supplementary Note 11; Fig. S18), suggesting that denser sampling and larger models will be required for still wider domains. Very low ($< 10°$) and very high ($> 80°$) solar elevations are under-represented in our library, so performance at those extremes warrants further study. Future work should target dense sampling of equatorial noon and polar-winter conditions, include waters with unusual scattering (e.g., glacial flour, phytoplankton blooms), and explore larger transformers that ingest raw four-channel Stokes imagery end-to-end. With these extensions—and continued reliance on inexpensive polarization cameras—global, infrastructure-free navigation for autonomous underwater platforms moves within reach.

References

[1] James Arvo. Stratified sampling of spherical triangles. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, page 437–438, New York, NY, USA, 1995. Association for Computing Machinery.

[2] Xiaoyang Bai, Zuodong Liang, Zhongmin Zhu, Alexander Schwing, David Forsyth, and Viktor Gruev. Polarization-based underwater geolocalization with deep learning. *eLight*, 3(15), 2023.

[3] Xiaoyang Bai, Zhongmin Zhu, Alexander Schwing, David Forsyth, and Viktor Gruev. Learning a global underwater geolocalization model with sectoral transformer. *Optics Express*, 32(12):20706–20718, 2024.

[4] Parrish C. Brady, Alexander A. Gilerson, George W. Kattawar, James M. Sullivan, Kort Travis, Sheila David, Evan Accorsi, Robert Foster, Antonio Tonizzo, and Molly E. Isaac, Robert and(// additional authors omitted for brevity //) Cummings. Open-ocean fish reveal an omnidirectional solution to camouflage in polarized environments. *Science*, 350(6263): 965–969, 2015.

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[6] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Computer Vision – ECCV 2020*, page 213–229, Berlin, Heidelberg, 2020. Springer-Verlag.

[7] Mingzhi Chen, Yuan Liu, Daqi Zhu, Wen Pang, and Jianmin Zhu. Underwater polarized light navigation: Current progress, key challenges, and future perspectives. *Robotics*, 14(8):104, 2025.

[8] Haoyuan Cheng, Qi Chen, Xiangwei Zeng, Haoxun Yuan, and Linjie Zhang. The polarized light field enables underwater unmanned vehicle bionic autonomous navigation and automatic control. *Journal of Marine Science and Engineering*, 11(8):1603, 2023.

[9] Haoyuan Cheng, Xiaoqing Guo, Haozhe Bai, Guanghao Li, and Chuanyun Su. The research on ocean polarized light fields at different depths for polarization navigation. *IEEE Access*, 11:137702–137708, 2023.

[10] Hao-yuan Cheng, Shi-min Yu, Hao Yu, Jin-chi Zhu, and Jinkui Chu. Bioinspired underwater navigation using polarization patterns within snell's window. *China Ocean Engineering*, 37(4):628–636, 2023.

[11] The Astropy Collaboration, Adrian M. Price-Whelan, Pey Lian Lim, Nicholas Earl, Nathaniel Starkman, Larry Bradley, David L. Shupe, Aarya A. Patil, Lia Corrales, C. E. Brasseur, Maximilian Nöthe, Axel Donath, Erik Tollerud, Brett M. Morris, Adam Ginsburg, Eero Vaher, Benjamin A. Weaver, James Tocknell, William Jamieson, Marten H. van Kerkwijk, Thomas P. Robitaille, Bruce Merry, Matteo Bachetti, H. Moritz Günther, Paper Authors, Thomas L. Aldcroft, Jaime A. Alvarado-Montes, Anne M. Archibald, Attila Bódi, Shreyas Bapat, Geert Barentsen, Juanjo Bazán, Manish Biswas, Médéric Boquien, D. J. Burke, Daria Cara, Mihai Cara, Kyle E Conroy, Simon Conseil, Matthew W. Craig, Robert M. Cross, Kelle L. Cruz, Francesco D'Eugenio, Nadia Dencheva, Hadrien A. R. Devillepoix, Jörg P. Dietrich, Arthur Davis Eigenbrot, Thomas Erben, Leonardo Ferreira, Daniel Foreman-Mackey, Ryan Fox, Nabil Freij, Suyog Garg, Robel Geda, Lauren Glattly, Yash Gondhalekar, Karl D. Gordon, David Grant, Perry Greenfield, Austen M. Groener, Steve Guest, Sebastian Gurovich, Rasmus Handberg, Akeem Hart, Zac Hatfield-Dodds, Derek Homeier, Griffin Hosseinzadeh, Tim Jenness, Craig K. Jones, Prajwel Joseph, J. Bryce Kalmbach, Emir Karamehmetoglu, Mikołaj Kałuszyński, Michael S. P. Kelley, Nicholas Kern, Wolfgang E. Kerzendorf, Eric W. Koch, Shankar Kulumani, Antony Lee, Chun Ly, Zhiyuan Ma, Conor MacBride, Jakob M. Maljaars, Demitri Muna, N. A. Murphy, Henrik Norman, Richard O'Steen, Kyle A. Oman, Camilla Pacifici, Sergio Pascual, J. Pascual-Granado, Rohit R. Patil, Gabriel I Perren, Timothy E. Pickering, Tanuj Rastogi, Benjamin R. Roulston, Daniel F Ryan, Eli S. Rykoff, Jose Sabater, Parikshit Sakurikar, Jesús Salgado, Aniket Sanghi, Nicholas Saunders, Volodymyr Savchenko, Ludwig Schwardt, Michael Seifert-Eckert, Albert Y. Shih, Anany Shrey Jain, Gyanendra Shukla, Jonathan Sick, Chris Simpson, Sudheesh Singanamalla, Leo P. Singer, Jaladh Singhal, Manodeep Sinha, Brigitta M. Sipőcz, Lee R. Spitler, David Stansby, Ole Streicher, Jani Šumak, John D. Swinbank, Dan S. Taranu, Nikita Tewary, Grant R. Tremblay, Miguel de Val-Borro, Samuel J. Van Kooten, Zlatan Vasović, Shresth Verma, José Vinícius de Miranda Cardoso, Peter K. G. Williams, Tom J. Wilson, Benjamin Winkel, W. M. Wood-Vasey, Rui Xue, Peter Yoachim, Chen Zhang, Andrea Zonca, and Astropy Project Contributors. The astropy project: Sustaining and growing a community-oriented open-source project and the latest major release (v5.0) of the core package*. *The Astrophysical Journal*, 935(2):167, 2022.

[12] Thomas W Cronin and Justin Marshall. Patterns and properties of polarized light in air and water. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366 (1565):619–626, 2011.

[13] Marie Dacke, Dan-Eric Nilsson, Clarke H Scholtz, Marcus Byrne, and Eric J Warrant. Insect orientation to polarized moonlight. *Nature*, 424(6944):33–33, 2003.

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *NAACL*, 1, 2019.

[15] Linhao Dong, Shuang Xu, and Bo Xu. Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5884–5888, 2018.

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

[17] Julien Dupeyroux, Julien R. Serres, and Stéphane Viollet. Antbot: A six-legged walking robot able to home like desert ants in outdoor environments. *Science Robotics*, 4(27): eaau0307, 2019.

[18] C Evangelista, P Kraft, Marie Dacke, T Labhart, and MV Srinivasan. Honeybee navigation: critically examining the role of the polarization compass. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1636): 20130037, 2014.

[19] KONG Fang, GUO Yingjing, FAN Xiaojing, GUO Xiaohan, et al. Review on bio-inspired polarized skylight navigation. *Chinese Journal of Aeronautics*, 36(9):14–37, 2023.

[20] Stefan Greif, Ivailo Borissov, Yossi Yovel, and Richard A. Holland. A functional role of the sky's polarization pattern for orientation in the greater mouse-eared bat. *Nature Communications*, 5:4488, 2014.

[21] Chuanlong Guan, Jinkui Chu, Yuanyu Ji, Jinshan Li, and Ran Zhang. A micro/nano-integrated polarization solar compass: Solar position and geographical position. *ACS Photonics*, 12 (6):3032–3041, 2025.

[22] Roger T Hanlon and John B Messenger. *Cephalopod behaviour*. Cambridge University Press, 2018.

[23] Ben J Hardcastle, Jaison J Omoto, Pratyush Kandimalla, Bao-Chau M Nguyen, Mehmet F Keleş, Natalie K Boyd, Volker Hartenstein, and Mark A Frye. A visual pathway for skylight polarization processing in drosophila. *Elife*, 10: e63225, 2021.

[24] Gábor Horváth and Dezsö Varjú. *Polarized light in animal vision: polarization patterns in nature*. Springer Science & Business Media, 2004.

[25] Hemani Kaushal and Georges Kaddoum. Underwater optical wireless communication. *IEEE Access*, 4:1518–1547, 2016.

[26] Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

[27] Thomas Kronland-Martinet, Léo Poughon, Marcel Pasquinelli, David Duché, Julien R. Serres, and Stéphane Viollet. Skypole—a method for locating the north celestial pole from skylight polarization patterns. *Proceedings of the National Academy of Sciences*, 120(30):e2304847120, 2023.

[28] Michael F Land and Dan-Eric Nilsson. *Animal eyes*. Oxford University Press, 2012.

[29] Guorui Li, Xiangping Chen, Fanghao Zhou, Yiming Liang, Youhua Xiao, Xunuo Cao, Zhen Zhang, Mingqi Zhang, Baosheng Wu, Shunyu Yin, Yi Xu, Hongbo Fan, Zheng Chen, Wei Song, Wenjing Yang, Binbin Pan, Jiaoyi Hou, Weifeng Zou, Shunping He, Xuxu Yang, Guoyong Mao, Zheng Jia, Haofei Zhou, Tiefeng Li, Shaoxing Qu, Zhongbin Xu, Zhilong Huang, Yingwu Luo, Tao Xie, Jason Gu, Shiqiang Zhu, and Wei Yang. Self-powered soft robot in the mariana trench. *Nature*, 591(7848):66–71, 2021.

[30] N Justin Marshall. A unique colour and polarization vision system in mantis shrimps. *Nature*, 333(6173):557–560, 1988.

[31] I. Masmitja, J. Navarro, S. Gomariz, J. Aguzzi, B. Kieft, T. O'Reilly, K. Katija, P. J. Bouvet, C. Fannjiang, M. Vigo, P. Puig, A. Alcocer, G. Vallicrosa, N. Palomeras, M. Carreras, J. del Rio, and J. B. Company. Mobile robotic platforms for the acoustic tracking of deep-sea demersal fishery resources. *Science Robotics*, 5(48):eabc3701, 2020.

[32] Samuel B Powell, Roman Garnett, Justin Marshall, Charbel Rizk, and Viktor Gruev. Bioinspired polarization vision enables underwater geolocalization. *Science Advances*, 4(4): eaao6841, 2018.

[33] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. Real-world humanoid locomotion with reinforcement learning. *Science Robotics*, 9(89):eadi9579, 2024.

[34] Steven M Reppert, Haisun Zhu, and Richard H White. Polarized light helps monarch butterflies navigate. *Current Biology*, 14(2):155–158, 2004.

[35] Nadav Shashar, Sönke Johnsen, Amit Lerner, Shai Sabbah, Chuan-Chin Chiao, Lydia M. Mäthger, and Roger T. Hanlon. Underwater linear polarization: physical limitations to biological functions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1565):649–654, 2011.

[36] KL Smith Jr, AD Sherman, PR McGill, RG Henthorn, J Ferreira, TP Connolly, and CL Huffard. Abyssal benthic rover, an autonomous vehicle for long-term monitoring of deep-ocean processes. *Science Robotics*, 6(60):eabl4925, 2021.

[37] Sirimuvva Tadepalli, Joseph M Slocik, Maneesh K Gupta, Rajesh R Naik, and Srikanth Singamaneni. Bio-optics and bio-inspired optical materials. *Chemical Reviews*, 117(20): 12705–12763, 2017.

[38] Shanpeng Wang, Zhenbing Qiu, Panpan Huang, Xiang Yu, Jian Yang, and Lei Guo. A bioinspired navigation system for multirotor uav by integrating polarization compass/magnetometer/ins/gnss. *IEEE Transactions on Industrial Electronics*, 70(8):8526–8536, 2022.

[39] Talbot H Waterman. Polarization patterns in submarine illumination. *Science*, 120(3127):927–932, 1954.

[40] Rudiger Wehner. Polarization vision–a uniform sensory capacity? *Journal of Experimental Biology*, 204(14):2589–2596, 2001.

[41] Rüdiger Wehner. *Desert navigator: the journey of an ant*. Harvard University Press, 2020.

[42] Teng Zhang, Jian Yang, Lei Guo, Pengwei Hu, Xin Liu, Panpan Huang, and Chenliang Wang. A bionic point-source polarisation sensor applied to underwater orientation. *The Journal of Navigation*, 74(5):1057–1072, 2021.

[43] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16239–16248, 2021.