SIAMESE LEARNING-BASED MONARCH BUTTERFLY LOCALIZATION

Sara Shoouri[⋆] *, Mingyu Yang*[⋆] *, Gordy Carichner*[⋆] *, Yuyang Li*[†] *, Ehab A.Hamed*[†] *, Angela Deng*[⋆] *, Delbert A. Green II*[⋆] *, Inhee Lee*[†] *, David Blaaw*[⋆] *, Hun-Seok Kim*[⋆]

[⋆] University of Michigan, Ann Arbor, MI, USA , [†] University of Pittsburgh, Pittsburgh, PA, USA

2022 IEEE Data Science and Learning Workshop (DSLW) | 978-1-6654-5426-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/DSLW53931.2022.9820497

ABSTRACT

A new GPS-less, daily localization method is proposed with deep learning sensor fusion that uses daylight intensity and temperature sensor data for Monarch butterfly tracking. Prior methods suffer from the location-independent day length during the equinox, resulting in high localization errors around that date. This work proposes a new Siamese learning-based localization model that improves the accuracy and reduces the bias of daily Monarch butterfly localization using light and temperature measurements. To train and test the proposed algorithm, we use 5658 daily measurement records collected through a data measurement campaign involving 306 volunteers across the U.S., Canada, and Mexico from 2018 to 2020. This model achieves a mean absolute error of 1.416° in latitude and 0.393° in longitude coordinates outperforming the prior method.

***Index Terms*—** Light-level geolocalization, Siamese learning, Contrastive learning

1. INTRODUCTION

GPS-less multi-sensor data geolocators [1, 2] are miniaturized tracking devices that periodically collect sunlight, temperature, and other data to enable tracking of animals and even insects that migrate long (>1000 km) distances. Determining the daily location and trajectory of small animal and insect migrations is essential for understanding ecology, species interactions, and the impact of climate change on animals. For Monarch butterfly tracking, where a GPS receiver is not feasible because of its excessive power, size, and weight of the data logger, sunlight and temperature-based localization techniques have been investigated as a potential solution [2, 3].

Most prior work using light and temperature-based localization typically estimates the daily geo-coordinate based on a 'threshold method' [4] or 'template-fit' model [5]. In the threshold method, sunset and sunrise times are computed based on the moment when solar irradiance crosses a predefined threshold level [6]. Longitude is then estimated by the time of local noon and latitude by the measured day length. However, since latitude depends on the sun elevation angle, the estimation error is date- and location-dependent. The template-fit approach uses an analytical or data-driven template function for the location-dependent light variation to map the input data into latitude and longitude coordinates for a particular day. However, these methods typically have significant latitude ambiguity around equinox days due to the low day length variation everywhere on Earth.

To alleviate these issues, Yang [3] presents a deep neural network (DNN)-based localization algorithm in which light and temperature data are fed into neural network models that estimate the daily position by approximating a likelihood probability. This algorithm can offer reduced estimation errors by combining the temperature and light-based probability estimation/heatmaps. However, it still has a relatively high latitude error, especially around the equinox when Monarch butterflies are actively migrating.

In this work, we propose a Siamese learning-based localization approach to estimate the location of the data loggers attached to Monarch butterflies collecting light intensity and temperature data. Our framework attempts to learn a general pattern representation of data and produce a pairwise similarity score for two given inputs, which quantifies the probability of proximity of their data collection locations. By evaluating the similarity between the sensor collected data (from unknown location) and the reference data (with known location) in the database, the proposed method estimates the daily butterfly location more reliably outperforming the prior art [3]. We implement our algorithm for daily localization of Monarch butterflies that migrate from Canada to Mexico during September – December each year. For training and testing the proposed network models, we use the data collected by a measurement campaign with 306 volunteers [7] to record the sunlight intensity and temperature from 2018 to 2020. Our method is applicable to the mSAIL [2] data logger, which is customized for Monarch butterfly attachment with a system size of $8\times8\times2.6\text{mm}^3$ and the weight of 62mg (Fig. 1 top left). Our experimental results performed on the volunteer data show that the proposed algorithm can significantly reduce the localization error around the equinox (from 3.52° to 1.74°) and increase the robustness of the results compared to the state-of-the-art [3]. Our code is available at `https://github.com/sarashoouri/Siamese_Monarch`.

2. METHOD

Our aim is to estimate the probability that a particular pair of light intensity L_t and temperature T_t measurements belongs to a $2D$ coordinate x (latitude and longitude) on a

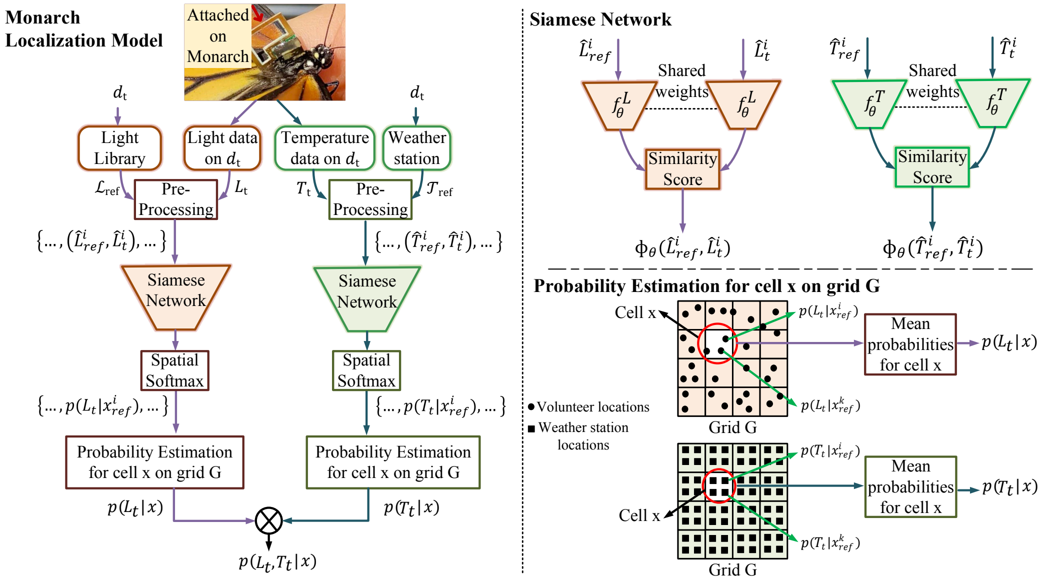

Fig. 1: **Left:** Structure of the proposed algorithm. The sensor image is from [2]. **Top Right:** The Siamese network estimates the similarity between a pair of data. **Bottom Right:** $p(L_t|x)$ and $p(T_t|x)$ calculation for a given cell center x on grid G.

specific date d_t. Thus, given a search grid G containing all possible $2D$ coordinates, the estimated position x_{est} on date d_t can be computed using the maximum likelihood, $x_{est} = \arg\max_{x\in G}\, p(L_t, T_t|x)$. The likelihood value is estimated using two different Siamese neural network models: a light intensity model to compute $p(L_t|x)$, and a temperature model to compute $p(T_t|x)$. Similar to [3], we make a simplifying assumption that light and temperature generate conditionally independent probabilities for a given coordinate such that $p(L_t, T_t|x) \approx p(L_t|x)p(T_t|x)$ holds. We recognize that this assumption is not generally valid, but this simplification is necessary for our model training because of the data availability imbalance between the light and temperature data (i.e., it is difficult to fully characterize the correlation between them). While the temperature data is available from relatively densely populated weather stations, light intensity data is not measured/reported by weather stations. Hence, light intensity data is entirely obtained through volunteer data collections, which are only available for particular dates and on coarser coordinates.

Our proposed algorithm has two main stages, 1) Data pre-processing, and 2) Contrastive Siamese learning [8, 9, 10]-based localization. Fig. 1 shows the overall algorithm.

2.1. Data pre-processing

Monarchs rest during the night without moving. Thus, we perform daily localization utilizing the light and temperature data centered around (±9 hours) the night center. The proposed data pre-processing method consists of two functions: a night center computing function N and a time-shift function r. The night center computing function is designed separately for light (N_L) and temperature data (N_T). The function $r(I, N)$ time-shifts the data I such that it is centered around the night center N.

To compute the night center for a given light intensity record L for a day, we divide L into two parts based on a hypothetical center point and calculate the cross-correlation between the two separated parts. Then, the hypothetical center point is adjusted until the cross-correlation reaches the maximum value, meaning that the divided parts have the highest symmetry. The time point that maximizes the cross-correlation value is assigned as the night center for L. Hence, the night center computing function N_L takes L as input and estimates the night center $n_c = N_L(L)$.

Applying a similar method to the temperature measurement is unreliable since the temperature is often asymmetric around the night center. Hence, to align a temperature record T of one day to its night center, we use an astronomical equation MATLAB function [11] to calculate the night center n_c for the location x and date d. The night center n_c for T obtained at a (hypothetical) location x and date d is calculated by $n_c = N_T(x, d)$, where N_T is the astronomical equation-based night center calculation function.

We collect temperature data T with a 1-hour interval to match the weather station's time resolution. We then produce a pre-processed temperature measurement $\hat{T}$ by time-shifting it (via r) to be centered around the night center: $\hat{T} = r(T, N_T(x, d))$ given a (hypothetical) location x and a measurement date d. On the other hand, light intensity L data for each day is collected from the sensor with a 1-minute interval, and is converted to log scale to emphasize the low light level variations around the sunrise and sunset. Note that the light intensity data collected from a sensor attached to a butterfly is prone to environmental variations that can lead to incorrect localization results. Thus, we first apply a denoising adversarial autoencoder (DAAE) [12] to L to construct a denoised light record. The pre-processed light data $\hat{L}$, is thus obtained through the denoising DAAE (Ψ), night center (N_L), and time-shifting (r) calculation as follows: $\hat{L} = r(\Psi(L), N_L(\Psi(L)))$.

We now explain the denoising DAAE for the light intensity L. The goal is to estimate the clean data $\tilde{L}$ by an autoencoder Ψ that consists of an encoder Ψ_E and a decoder Ψ_D pair that is trained to minimize a reconstruction loss. We treat this denoising as a distribution alignment task. The encoder Ψ_E takes the noisy light L to generate its latent representation z with a smaller dimension, and the decoder Ψ_D produces the denoised light $\tilde{L}$ from z. We desire to align the latent representations z of noisy original data and $\tilde{z}$ of clean data by establishing a suitable discriminator D to classify the latent vectors into original or denoised (clean) data. The autoencoder $\Psi(L) = \Psi_D(\Psi_E(L))$ and the discriminator D are trained in an adversarial setting where each tries to outperform the other playing a two-player minimax game as initially introduced in [13] using the loss:

$$\mathscr{L}_{dis} = \mathbb{E}_{\tilde{z}}[log(D(\tilde{z}))] + \mathbb{E}_z[log(1 - D(z))]. \tag{1}$$

2.2. Siamese learning-based localization

The prior published 'template-fit' method [4] observed that the slope of light intensity has variations around sunrise and sunset that match at nearby locations on the same day. For example, the sun sets more slowly at high latitude vs. low latitude locations. We extend this idea to pattern matching and apply these slope variations to both light and temperature measurements. The primary assumption of our Siamese learning-based algorithm is that two closely located data on approximately the same day of the year (but in different years) should have highly correlated patterns.

We first build a reference library of light intensities from the reference volunteer data with their known positions to perform the pattern matching. However, the constructed reference library is sparsely populated with only the locations where volunteers collected data, making it infeasible to perform pattern matching at arbitrary locations. We mitigate this issue by populating 'synthesized' light data along with the longitude coordinate, significantly expanding the available light data for matching. The longitude coordinate mainly affects the time of the night center, and simply time-shifting the night center generates a synthesized light intensity at a new longitude coordinate. Thus, all of the volunteer data are time-shifted to cover our desired longitude range and construct the synthesized light intensity reference library. The amount of time-shift is 4 minutes per one degree of longitude. Note that we cannot apply a similar light intensity shifting to create synthesized data along the latitude coordinate as the light intensity pattern (after the night center alignment) per day depends on the latitude, and we cannot have a reliable model to 'synthesize' it.

To describe the proposed Siamese learning-based localization, suppose L_t is the light data collected on a known date d_t at an unknown ground-truth 2D (latitude and longitude) location x_t. Let $\mathscr{L}_{ref}$ be a reference library (with volunteer collected and synthesized data). The reference light data subset L_{ref} contains light data collected from ± 5 days around d_t at known 2D coordinates x_{ref}, and the size of this subset is denoted by B. Our goal is to quantify the similarity between each entry in L_{ref} and L_t. Thus, for the ith element in L_{ref}, we obtained the pre-processed version $\hat{L}^i_{ref} = r(\Psi(L^i_{ref}), N_L(\Psi(L^i_{ref})))$ after calculating its night center. Then, we generate a pre-processed version of the target light data $\hat{L}^i_t = r(\Psi(L_t), N_L(\Psi(L^i_{ref})))$, using the same night center obtained from the reference $\Psi(L^i_{ref})$. In this way, when the locations of the L^i_{ref} and L_t are matched, the estimated night centers align to each other, resulting in more accurate similarity scores.

We then use a mapping function to convert the difference between $\{\hat{L}^i_{ref}, \hat{L}^i_t\}$ into a distance between their locations $\{x^i_{ref}, x_t\}$. For instance, if $\hat{L}^i_{ref}$ has a similar pattern as $\hat{L}^i_t$, then the distance between their locations $||x^i_{ref} - x^i_t||_2$ should be relatively small. We implement a Siamese neural network to construct this mapping function (transformer encoder) ϕ_θ, parameterized by θ. Siamese networks are twin neural networks that share the identical weights [8, 9, 10]. The function ϕ_θ maps the pattern representation of the tuple $\{\hat{L}^i_{ref}, \hat{L}^i_t\}$ into a similarity score which indicates how close they are located. To construct ϕ_θ, we apply the same convolutional neural network f^L_θ to both $\hat{L}^i_{ref}$ and $\hat{L}^i_t$ to generate the feature vectors $f^L_\theta(\hat{L}^i_t)$ and $f^L_\theta(\hat{L}^i_{ref})$ in the latent space. Euclidean distance between the latent vectors quantifies the similarity score between $\hat{L}^i_{ref}$ and $\hat{L}^i_t$ such that $\phi_\theta(\hat{L}^i_{ref}, \hat{L}^i_t) = ||f^L_\theta(\hat{L}^i_{ref}) - f^L_\theta(\hat{L}^i_t)||_2$.

To train the Siamese networks, positive (matching) and negative (non-matching) samples are required. Positive samples are the pairs of closely located data whose distance differences are less than $55km$ or $0.5°$ in longitude and latitude, and the negative samples are the pairs of data whose distances are longer than $55km$. The model is trained using a contrastive loss function [14] as described in Eq. (2). The y value is 1 for the positive samples and 0 for the negative samples, and m is the threshold margin. The contrastive loss aims to maximize the similarity score for the positive samples while minimizing it for the negative samples.

$$\mathscr{L}_{CL} = \frac{1-y}{2}||\hat{L}^i_{ref} - \hat{L}^i_t||^2_2 + \frac{y}{2}\{max(0, m - ||\hat{L}^i_{ref} - \hat{L}^i_t||_2)^2\}. \tag{2}$$

The spatial softmax function (3) is then applied to the similarity score between $\hat{L}^i_{ref}$ and $\hat{L}^i_t$ to convert it to a probability representation in the range of $(0, 1)$. This represents the probability that L_t is obtained at the location x^i_{ref}.

$$p(L_t|x^i_{ref}) = exp\left(-\frac{\phi^2(\hat{L}^i_{ref}, \hat{L}^i_t)}{2\sigma^2}\right) / \sum_{j=1}^{B} exp\left(-\frac{\phi^2(\hat{L}^j_{ref}, \hat{L}^j_t)}{2\sigma^2}\right). \tag{3}$$

In (3), σ is the standard deviation of the similarity scores for positive samples. The estimated probability (3) is evaluated with all $\{\hat{L}^i_{ref}, \hat{L}^i_t\}$ for $i = \{1, \cdots, B\}$ to create a set I_{ref} containing the probabilities for given reference locations:

$$I_{ref} = \{(p(L_t|x^1_{ref}), x^1_{ref}), \cdots, \quad (p(L_t|x^B_{ref}), x^B_{ref})\}. \quad (4)$$

After generating I_{ref}, the position x_t of L_t can be estimated by performing a coarse-to-fine grid search. A coarse search grid G has $[27° : 1° : 48°]$ latitude coordinate grids, and $[-122° : 1° : -66°]$ longitude coordinate grids. This search range covers our study area of the U.S, Canada, and Mexico. Each point on G has a 2D coordinate x, and we compute $p(L_t|x)$ for all points on G.

To estimate $p(L_t|x)$, we use the reference positions in I_{ref} which are located around the position x. Thus, a subset of probabilities, called I^{cell}_{ref}, from I_{ref} is created such that all points in I^{cell}_{ref} satisfy $|x_{ref} - x| < 1°$. Finally, $p(L_t|x)$ is calculated by taking an average of the probabilities in the subset I^{cell}_{ref}. Although the reference library is densely populated along the longitude coordinate (due to the data 'synthesis' procedure with time-shifting), the population along the latitude may be coarse for some regions where fewer volunteers were available. Thus, it is possible that with the constraint of $|x_{ref} - x| < 1°$, some grid cells centered at x may have an empty I^{cell}_{ref} set. For these empty cells, $p(L_t|x)$ is estimated based on the probabilities from nearest neighboring cells using linear interpolation.

The spatial resolution of the likelihood estimation for the grid G is refined by upsampling and linear interpolating the heatmap of $p(L_t|x)$ with a 0.1° resolution of x on the refined $\tilde{G}$. The localization given L_t (light-only localization) is completed through the maximum likelihood estimation on the refined $\tilde{G}$ such that $x_t \approx \arg\max_{\tilde{G}} p(L_t|x)$.

The proposed method for the light data-based likelihood probability $p(L_t|x)$ estimation was extended to the temperature likelihood probability $p(T_t|x)$ estimation in a straightforward manner. Unlike the light data that is only available at sparse volunteer locations, the reference temperature data is available at all weather station locations which are densely distributed. We construct the temperature reference library $\mathcal{T}_{ref}$ by accessing the weather station data through WeatherBit API [15]. The pre-processed reference (weather station) data $\hat{T}^i_{ref}$ and sensor temperature data $\hat{T}^i_t$ on date d_t are obtained by time-shifting both data using the night-center calculated by the reference temperature data, such that $\hat{T}^i_{ref} = r(T^i_{ref}, N_T(d_t, x^i_{ref}))$ and $\hat{T}^i_t = r(T_t, N_T(d_t, x^i_{ref}))$ hold. Note that denoising is not used for temperature data. A Siamese neural network f^T_θ is trained for temperature matching, then computes a similarity score between the pre-processed weather station temperature data $\hat{T}^i_{ref} \in \mathcal{T}_{ref}$ and the sensor data $\hat{T}^i_t$. This similarity score is then converted to a probability $p(T_t|x^i_{ref})$ using Eq. (3) by replacing L with T. The remaining steps to generate $p(T_t|x)$ are identical to the $p(L_t|x)$ estimation.

The final likelihood probability that jointly considers light and temperature data is approximated by the product: $p(L_t, T_t|x) \approx p(L_t|x)p(T_t|x)$ based on the simplifying assumption of conditional independence between the light and temperature measurements.

3. EXPERIMENTS

3.1. Data collection

We use real-world data collected through a data measurement campaign with 306 volunteers across the U.S., Canada, and Mexico from 2018 to 2020 [7]. Volunteers recorded light and temperature data using HOBO [16] sensors to emulate the mSAIL platform [2] (Fig. 1, top left) from September to early December. The collected dataset contains 5658 daily records with a time resolution of 10 sec for light and 15 sec for temperature. We use the year 2018 – 2019 volunteer data as the training dataset (size of 3834) and the year 2020 data as the testing set (size of 1824). For temperature data, we use the weather station data with a time resolution of 1 hour accessed through WeatherBit API [15]. Although the weather station data is much more densely populated than the volunteer data, it does not necessarily cover the entire study area. When the temperature data is not available for a particular location x, we apply the Kriging spatial interpolation [17] to the nearby weather station data to obtain the temperature at that location.

3.2. Network structure

The Siamese network f^L_θ for light data consists of 4 convolutional (conv) layers containing convolution, batch normalization, ReLU, and max pooling, followed by 2 fully connected layers (FCLs). A dropout layer of $p = 0.30$ is applied after the first FCL. The size of the first conv layer is $128 \times 1 \times 9$ and the sizes of the other conv layers are $128 \times 1 \times 5$. The Siamese network for temperature data f^T_θ consists of 2 conv layers with the size of $32 \times 1 \times 3$, and 2 FCLs. A dropout layer of $p = 0.30$ is used after the first FCL. The denoising encoder Ψ_E (and decoder Ψ_D) consists of 3 FCLs of size 480×200 (50×100), 200×100 (100×200), and 100×50 (200×480), each followed by ReLU. The discriminator contains 3 FCLs of size 50×500, 500×500, and 500×1, followed by Sigmoid. We use an ADAM optimizer with a learning rate of 0.001 and a StepLR scheduler with a step size of 1000 to train the models. At each epoch of the training for light data, we choose a batch size of two light records, whose date difference is less than 5 days (but across different years). The temperature model is trained in a similar way, except that one input to the Siamese network is from the weather station.

3.3. Localization Results

We evaluate our test data localization on a $2D$ grid covering southern Canada, the U.S., and Mexico for Monarch butterfly localization. Fig. 2 provides the performance comparison between our proposed method and the state-of-the-art [3] using the mean absolute error (MSE) evaluated bi-

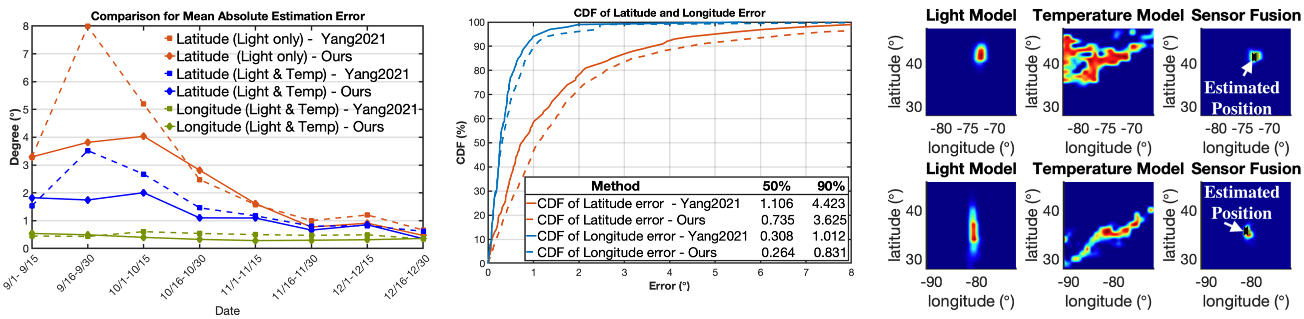

Fig. 2: **Left**: MSE comparison evaluated biweekly for proposed method vs. Yang [3]. **Middle**: Error CDF. **Right**: Example likelihood probability heatmaps for two different days. Red/yellow/blue indicates high/mid/low probability. First row: Dec 7, Second row: Nov 11. The ground-truth and estimated locations are marked with black and green points, respectively.

weekly. All results in this figure are based on the volunteer data (2018 - 2019 data for training and 2020 data for testing). The CDF of error in longitude and latitude degree is shown in Fig. 2 middle. Our algorithm outperforms the baseline [3], especially around the equinox (Sep. 22), proving that the pattern matching technique can significantly compensate for the low day length variation issue around the equinox. Fig. 2 shows that using the temperature data is also critical to compensate for the latitude estimation error from the light-only method around the equinox day when the night length is the same everywhere. Fig. 2 (right subplot) visualizes the estimated probabilities $p(L_t|x)$, $p(T_t|x)$, and $p(L_t|x)p(T_t|x) \approx p(L_t, T_t|x)$.

We also evaluate our method using the data collected by an mSAIL [2] sensor attached to a wild Monarch butterfly on September 17, 2021 in Leamington, Ontario, near Lake Erie. The Monarch was released into the wild for a day, and then its sensor data was retrieved wirelessly while the butterfly was resting on a tree before flying across Lake Erie. Fig. 3 shows the localization result using the obtained data (the first hour of the data was extrapolated because the mSAIL sensor did not record it). The measured localization accuracy error is 0.032 and 0.1 degrees in latitude and longitude.

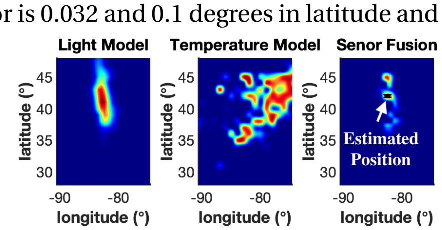

Fig. 3: Localization using butterfly-attached mSAIL [2] on Sep 17, 2021. Estimated point: $(42.0°, -82.50°)$. Ground-truth: $(42.032°, -82.60°)$

3.4. Bias Evaluation

For a reliable unbiased localization, the error should not strongly depend on the number of available reference data near the ground-truth position. Learning such an unbiased method can be challenging as the training dataset is not uniformly populated and depends on the volunteer locations. To evaluate the robustness and bias of the localization model, we first quantify how densely the reference data is populated around a given ground-truth position by computing the average euclidean distance between the neighboring reference points and the ground-truth position. This is named the 'Isolation score'. Then, we evaluate the correlation between the localization error and the computed Isolation score. Ideally, the error should be independent of the Isolation score, and the accuracy around isolated points should be as reliable as in more densely populated areas. We use three formulas, Pearson Correlation coefficient [18], Distance Correlation [19], and Mutual Information [20] to measure the correlation between the error and Isolation score. Low numbers in these measures indicate that the method has less correlation and results have less bias towards low Isolation score locations. Table 1 compares the measured correlation (or bias) between our model and the prior work [3]. Our model has significantly less bias in all measures while producing lower errors (Fig. 2). Low correlation/bias scores from our model imply that the localization error depends less on the density of reference (training/testing) data around the ground-truth location.

Table 1: Measured bias of Yang 2021 [3] and our model.

Method	Pearson Correlation (latitude, longitude)	Distance Correlation (latitude, longitude)	Mutual Information (latitude, longitude)
Yang2021 [3]	(0.26, 0.324)	(0.23, 0.292)	(0.33, 0.397)
Our model	**(0.070, 0.172)**	**(0.0744, 0.167)**	**(0.245, 0.308)**

4. CONCLUSION

We developed a Siamese learning-based localization using light intensity and temperature measurements for miniaturized data loggers to study Monarch butterfly migration. The proposed method significantly outperforms the prior method, especially around the equinox. Moreover, it has less bias towards more densely populated areas and achieves lower accuracy errors. The proposed algorithm demonstrates the successful localization of a wild butterfly using real-world data. The presented model exhibits a mean absolute error of 1.416° in latitude and 0.393° in longitude for the volunteer collected test dataset.

ACKNOWLEDGMENTS

This work was in part funded by NSF IIBR Award #2045017 and National Geographic Society Grant. We thank all the volunteers who participated in the data measurement campaign for this research.

5. REFERENCES

[1] Eli S Bridge, Jeffrey F Kelly, Andrea Contina, Richard M Gabrielson, Robert B MacCurdy, and David W Winkler, "Advances in tracking small migratory birds: a technical review of light-level geolocation," *Journal of Field Ornithology*, vol. 84, no. 2, pp. 121–137, 2013.

[2] Inhee Lee, Roger Hsiao, Gordy Carichner, Chin-Wei Hsu, Mingyu Yang, Sara Shoouri, Katherine Ernst, Tess Carichner, Yuyang Li, Jaechan Lim, et al., "msail: milligram-scale multi-modal sensor platform for monarch butterfly migration tracking," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 517–530.

[3] Mingyu Yang, Roger Hsiao, Gordy Carichner, Katherine Ernst, Jaechan Lim, Delbert A Green, Inhee Lee, David Blaauw, and Hun-Seok Kim, "Migrating monarch butterfly localization using multi-modal sensor fusion neural networks," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1792–1796.

[4] Steven LH Teo, Andre Boustany, Susanna Blackwell, Andreas Walli, Kevin C Weng, and Barbara A Block, "Validation of geolocation estimates based on light level and sea surface temperature from electronic tags," *Marine Ecology Progress Series*, vol. 283, pp. 81–98, 2004.

[5] A Philip et al., "An advance in geolocation by light," *Memoirs of National Institute of Polar Research. Special issue*, , no. 58, pp. 210–226, 2004.

[6] Simeon Lisovski, Chris M Hewson, Raymond HG Klaassen, Fränzi Korner-Nievergelt, Mikkel W Kristensen, and Steffen Hahn, "Geolocation by light: accuracy and precision affected by environmental factors," *Methods in Ecology and Evolution*, vol. 3, no. 3, pp. 603–612, 2012.

[7] mSAIL, "Volunteer data for light intensity and temperature measurement across the US, Canada, and Mexico," https://github.com/sarashoouri/Monarch_Butterfly_Tracking, [Online].

[8] Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah, "Signature verification using a "siamese" time delay neural network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 04, pp. 669–688, 1993.

[9] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al., "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*. Lille, 2015, vol. 2.

[10] Eric Lei, Oscar Castañeda, Olav Tirkkonen, Tom Goldstein, and Christoph Studer, "Siamese neural networks for wireless positioning and channel charting," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 200–207.

[11] François Beauducel (2021), "Sunrise: sunrise and sunset times," https://github.com/beaudu/sunrise/\releases/tag/v1.4.1, GitHub. Retrieved September 15, 2021.

[12] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[14] Chao-Yuan Wu, R Manmatha, Alexander J Smola, and Philipp Krähenbühl, "Sampling matters in deep embedding learning," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 2859–2867.

[15] Weatherbit API, ," https://www.weatherbit.io/, [Online].

[16] ONSET, "Hobo pedant mx temperature/light data logger," https://www.onsetcomp.com/products/data-loggers/mx2202/, [Online; accessed 20-March-2021].

[17] Jialin Zhang, Xiuhong Li, Rongjin Yang, Qiang Liu, Long Zhao, and Baocheng Dou, "An extended kriging method to interpolate near-surface soil moisture data measured by wireless sensor networks," *Sensors*, vol. 17, no. 6, pp. 1390, 2017.

[18] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, pp. 1–4. Springer, 2009.

[19] Gábor J Székely, Maria L Rizzo, and Nail K Bakirov, "Measuring and testing dependence by correlation of distances," *The annals of statistics*, vol. 35, no. 6, pp. 2769–2794, 2007.

[20] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, pp. 066138, 2004.

A. SUPPLEMENTAL MATERIAL

Fig. 4 displays the slope of light intensity variations around the sunset and sunrise around the equinox day. The light measurements are located at the same longitude coordinate with different latitude coordinates. The night length (time duration when the log of light intensity is below 0) is equal regardless of the latitude. It can be observed that the light records with close latitude coordinates (red and green light data) have matching slopes and patterns before the sunset and after the sunrise, whereas the slopes are different when latitude coordinates are mismatched. Our algorithm exploits this similarity/difference to improve the localization performance near equinox days.

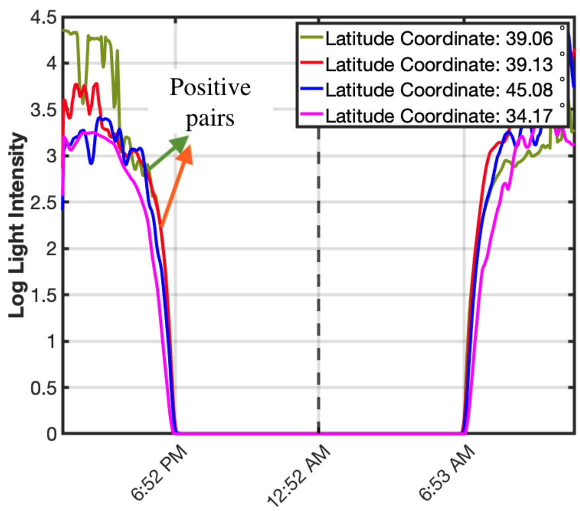

Fig. 4: Light intensities at various latitude coordinates with the same longitude coordinate.