Fighting Poaching Through Targeted Deep Learning and Sensor Integration

Naveen Dhar
High Tech High Mesa
San Diego, CA 92111
naveendhar8030@gmail.com

Irina Tolkova
K. Lisa Yang Center for Conservation Bioacoustics
Cornell University
Ithaca, NY 14850
ivt2@cornell.edu

Abstract

Passive acoustic monitoring (PAM) has become a crucial and widespread tool for conservation monitoring, aiding the protection of species threatened by gun-based poaching through detecting calls and vocalizations. However, real-time detection of gun-based poaching activity remains an unsolved challenge despite its large ecological implications. Existing methodologies face high false positive rates and utilize computationally intensive models unsuitable for real-time field deployment. This research developed a lightweight deep neural network suitable for on-board processing and a sensor integration layer to address these limitations. The developed model achieved a 0.91 validation F1 at 935k parameters, retaining 94% performance (F1 @ 95% recall) of existing literature while reducing size by over 87%. Statistical evaluation across acoustic array simulations demonstrated consistent false positive reduction through the proposed sensor integration function, presenting a promising approach for cost-effective real-time poaching detection and wildlife conservation.

1 Introduction

Passive acoustic monitoring (PAM) with autonomous recording units (ARUs) enables long-term, cost-effective, large-scale studies of vocal wildlife in remote environments [1]. In addition to the study of non-human animal communication, PAM has become an indispensable tool for conservation: enabling estimates of animal abundance, evaluation of ecosystem health, and assessment of anthropogenic impacts [2, 3, 4, 5]. Moreover, acoustic monitoring can aid in detecting major threats to biodiversity, such as illegal logging or poaching [6, 7, 8]. In particular, gun-based poaching drives far-reaching species decline, from the illicit ivory trade [9, 10] to unsustainable bushmeat hunting [11, 10]. Yet while the detection of animal vocalizations has been successfully performed across a broad range of studies and taxa [12, 13, 14], accurate real-time detection of gunshots is still lacking despite its large ecological implications. Immediate on-the-ground intervention of gun-based poaching can dramatically reduce poaching rates, improve ecosystem stability, and protect targeted species [15], at a potentially global scale, due to the benefits of acoustic monitoring [16]. A majority of gunshot detection research has been applied for urban contexts, utilizing curated datasets, reducing applicability for field applications [17, 18]. Existing literature on acoustic gunshot detection using real field datasets (from rainforest environments) consistently mentions the inherent challenges with false positives and generalization. Additionally, most methodologies utilize large fine-tuned models better suited for retrospective studies [19, 20, 21]. To address these bottlenecks and to enable real-time gun-based poaching detection, this research developed a lightweight deep neural network suitable for on-board processing and a sensor integration function as a novel approach to reduce false positive rates.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Non-Human Animal Communication.

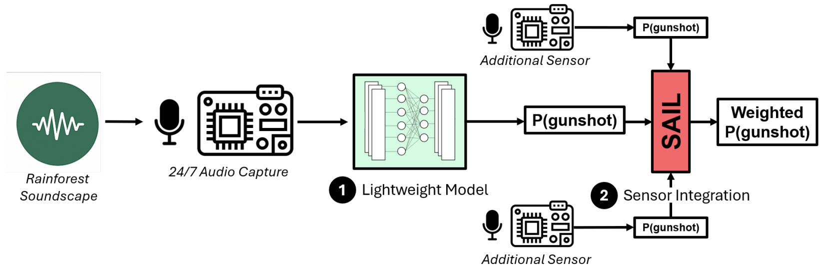

Figure 1: Pipeline for detecting poaching activity through real-time acoustic gunshot detection. This research develops and validates a lightweight model suitable for on-board processing (1) and SAIL, or the Sensor Analysis and Integration Layer (2). SAIL is proposed as a novel method to reduce false positive detections by integrating the predictions from spatially separate sensors positioned in an array. The data used, the developed model, and the SAIL function are available with CLI functionality at https://github.com/sail-gunshot-detect/sail-gunshot-detect-repo.

2 Methodology

2.1 Datasets

A pre-partitioned 35,980 4-second waveform dataset collected by Katsis et al. [19] was used for training. The dataset contains a 50:1 class imbalance with gunshots in the minority. The same partitioning used by Katsis et al. [19] was employed to ensure direct comparison. For evaluation and acoustical simulation on a spatially distinct dataset, the test partition of a Vietnamese rainforest dataset collected by Thinh Tien Vu et al. [20] was used, containing 129 background noise waveforms and 19 gunshots. Both datasets hold the CC BY 4.0 license. The possibility of performing the simulation on larger datasets is acknowledged; however, limited computing resources constrained the size of the simulation dataset.

2.2 Preprocessing and Augmentation

Log-mel spectrograms were chosen as model input for greater feature representation and dimensionality reduction. A window size of 256 samples with 50% overlap maintained temporal resolution, while 64 mel bins covered frequencies from 100Hz to 4kHz.

To enhance generalizability, extensive waveform-based augmentations were employed during training. As seen in Fig. 2, augmentations included pitch shifting, time shifting, Gaussian noise addition, negative sample overlay, and spectral/temporal masking inspired by SpecAugment [22]. Each augmentation had an occurrence probability of 0.3, except negative sample overlay, which always occurred.

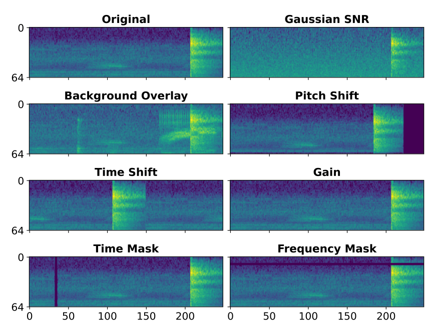

Figure 2: Augmentations applied during training. On the x and y axes are the 250 time bins spanning 4 seconds and the 64 mel bins spanning 0-4kHz.

2.3 Lightweight Supervised Classification

Our model architecture incorporated depthwise-separable 2D convolutions, 1D convolutions, and a GRU layer for computational efficiency while capturing gunshot features. The architecture comprised three sequential components: three depthwise-separable 2D convolutional blocks with descending kernel sizes from (7,7) to (3,3) for spatial feature extraction; three 1D convolutional layers (kernel size 3) and a GRU layer for temporal modeling; and a classification block with global max pooling and two fully connected layers for binary classification. The final model contained 935k parameters.

Models were trained for 40 epochs using the Adam optimizer and binary cross-entropy loss across five seeds. Training lasted two hours per model on a personal computer with a GTX 1060 GPU. The epoch with the lowest validation loss was selected for serialization.

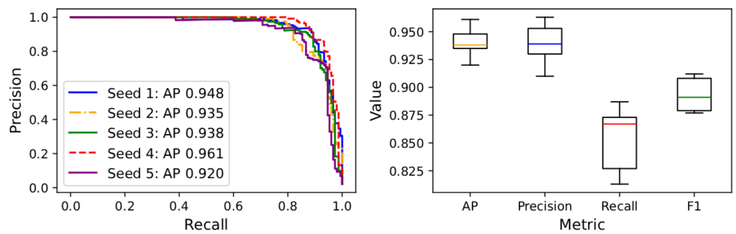

Figure 3: Precision-recall curves and boxplots of model performance across five seeds. AP denotes average precision.

2.4 Sensor Integration

The simple spectral structure of a gunshot can resemble commonplace percussive sounds such as snapping vegetation, often causing high false positive rates (FPRs). Yet unlike these sound sources, gunshots propagate for much farther distances, giving the opportunity to improve detection performance by considering predictions across multiple sensors.

A confidence weighting function mapping three or more sensor predictions to a more accurate final prediction was designed with the following criteria: (i) penalizing confident and isolated predictions, (ii) boosting sensor agreement, and (iii) producing probabilities from 0 to 1.

$$P_f(p_1, \ldots, p_n) = \sigma\left(1 - \frac{\displaystyle\sum_{1 \le i < j \le n} (1 - p_i)(1 - p_j)}{\displaystyle\sum_{k=1}^{n} p_k}\right) \tag{1}$$

Equation 1 describes the Sensor Analysis and Integration Layer (SAIL) function. It considers all unique sensor pairs and calculates the product of their negative confidences, measuring joint agreement that no event occurred. Summing across pairs yields an overall negative consensus, normalized by the total positive confidence and inverted before applying a sigmoid operation for the final probability of gunshot presence. Here, $\sigma(x) = \frac{1}{1+e^{k(x-m)}}$. The constant k determines the steepness of the sigmoid function slope, and the constant m determines the centering. To ensure smooth and centered mapping of probabilities to outputs between zero and one, values of $k = 10$ and $m = 0.5$ were chosen.

2.4.1 Simulation Design

Acoustic propagation simulation used waveforms from the spatially distinct dataset. Three augmented copies per waveform underwent gain, Gaussian noise, and temporal shift transformations simulating distance-dependent attenuation, emulating a three-sensor array. Inference produced confidence scores for each augmented waveform, which were fed into the SAIL function. To compare SAIL to a stochastic rather than deterministic baseline, the method of averaging the confidence scores was evaluated as well.

Statistical analysis was performed at the per-run and per-file scales, yielding dataset-level FPRs and mean FPRs of individual files across runs, respectively. Both scales were chosen to evaluate the stability of SAIL across a large number of simulated sensor arrays as well as the file-conditional effect of SAIL. Across $K = 1000$ simulation runs, Wilcoxon signed-rank tests evaluated whether median differences $d = \text{FPR}_{\text{inference}} - \text{FPR}_{\text{SAIL/AVG}}$ exceeded zero.

3 Results

Performance evaluation addressed (1) model performance versus larger networks, (2) the effect of domain shift and model generalization to a spatially distinct dataset, and (3) SAIL's effect on FPRs.

Classifier	$F1_{Best}$	F1 @ 95% Recall	$F1_{Distinct}$	Model Size	Parameters
Katsis et al. (ResNet18)	Unk.	0.89	Unk.	87MB	≈ 11.7M
Proposed	0.91	0.85	0.80	11MB	935K

Table 1: A table comparing the performance and model file size of the leading deep-learning classifier for gunshot detection on the Belizean dataset with the best-performing model developed in this study. $F1_{Distinct}$ denotes the best F1 on the spatially-distinct Vietnamese dataset.

	Per-Run			Per-File		
FPR Difference	Mean	95% CI	p-value	Mean	95% CI	p-value
INF-SAIL	0.0233	0.023–0.023	<1e-5	0.0233	0.0–0.0543	0.04
INF-AVG	0.0035	0.0032–0.0039	<1e-5	0.0035	-0.0076–0.0203	0.57
AVG-SAIL	0.0197	0.0198–0.0201	<1e-5	0.0198	0.0022–0.0441	9e-3

Table 2: Summary of Wilcoxon statistical testing the difference in false positive rates (FPRs) between inference, SAIL, and averaging across the 1000 simulations for both per-run and per-file scales. INF-SAIL denotes the FPR difference between inference and SAIL, and similarly INF-AVG and AVG-SAIL denote the respective FPR differences between inference, averaging, and SAIL.

3.1 Gunshot Detection

Fig. 3 presents precision-recall curves for the Belizean dataset validation partition. Models achieved 0.94 mean average precision and 0.99 average specificity at optimal F1 thresholds. Table 1 compares our best-performing model against the fine-tuned ResNet18 from Katsis et al. [19]. Our model is 11MB, compressing to 964KB via TensorflowLite serialization. Additionally, our approach obtains F1 scores of 0.91 (optimal threshold) and 0.85 (95% recall threshold), achieving significant model size and parameter reduction with only a six-percent reduction in F1 @ 95% Recall.

3.2 SAIL Simulation

Table 2 summarizes statistical testing, reporting mean FPR differences, Wilcoxon p-values, and 95% confidence intervals (bootstrapped with $n = 5000$). SAIL achieved statistically significant FPR reductions at both scales, eliminating FPRs on average, whereas averaging produced a small, 0.35 percent-point per-run FPR reduction and a non-significant per-file FPR reduction.

Part **(a)** of Fig. 4 displays per-run metrics across 1000 simulations, comparing baseline inference with SAIL. Simple inference achieved an FPR of 0.0233. However, even this relatively low FPR generates 42 false alarms per hour with a 4-second, 50% overlap sliding window detection at a threshold of 0.2. Averaging produced a mean FPR of 0.0197, slightly lower than simple inference, while also producing a much wider spread, including some outliers of higher FPR and lower precision and F1. Averaging produced a mean recall, or true-positive rate, of 0.72, higher than SAIL's mean recall of 0.61 or inference's mean recall of 0.68. SAIL consistently reduced FPR to zero and produced a mean

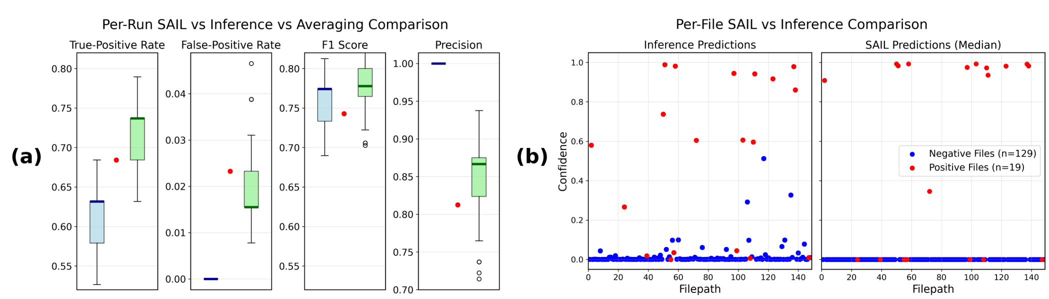

Figure 4: From left to right: **(a)** Distribution of SAIL metrics (boxplots) on a per-run scale across simulations alongside the performance of simple inference (red dots). **(b)** Scatterplot of median SAIL predictions for each filepath across simulations alongside inference predictions.

recall of 0.61. This reduction in recall may be attributable to already low-SNR or far-away gunshots undergoing substantial distance augmentations, resulting in more than one spectrogram copy having no meaningful gunshot features. SAIL's penalization system would treat this simulated array as a negative.

Part **(b)** of Fig. 4 shows confidence score scatterplots comparing inference and median SAIL per-file predictions. SAIL produces near-zero predictions for negatives and low-confidence positives while boosting high-confidence positives. Noting the filepath position of the low-confidence (below 0.5) positives on the left-hand scatterplot, we can see that these same filepaths constitute the false negatives produced by SAIL.

4 Conclusion

In future work, this research will be adapted for use in Cornell's Elephant Listening Project to detect poaching activity threatening elephant populations via the detection of gunshots. Evaluation of the lightweight model and the SAIL function on distance-labeled datasets is recommended to infer suitable sensor array distances in field deployment. Additionally, the SAIL function can be compared against other variants of itself, i.e., SAIL with simple output thresholding instead of a sigmoid operation, or SAIL without normalization by total positive event confidence. While not directly extendable to FPR reduction in field deployments, statistical evaluation of the 1000 acoustic simulation runs demonstrates that SAIL consistently reduces false positive rates compared to simple inference and averaging across a wide range of simulated sensor arrays.

The developed gunshot detector achieved a validation F1 score of 0.91 at 935k parameters, resulting in less than 964KB of storage when serialized. Compared to existing literature, the model retains 94% of performance while reducing size by over 87%. Overall, efficient deep learning and SAIL present a promising approach towards cost-effective real-time poaching detection and wildlife conservation.

Acknowledgements

We thank Bobbi Estabrook and Anahita Verahrami from the Elephant Listening Project for their domain expertise and discussion.

References

[1] Peter Prince, Andrew Hill, Evelyn Piña Covarrubias, Patrick Doncaster, Jake L Snaddon, and Alex Rogers. Deploying acoustic detection algorithms on low-cost, open-source acoustic sensors for environmental monitoring. *Sensors*, 19(3):553, 2019.

[2] Larissa S. M. Sugai, Thiago S. F. Silva, José W. Jr. Ribeiro, and Diego Llusiá. Terrestrial passive acoustic monitoring: review and perspectives. *BioScience*, 69(1):15–25, 2019. doi: 10.1093/biosci/biy147. URL `https://academic.oup.com/bioscience/article/69/1/15/5193506`.

[3] Rory Gibb, Ella Browning, Paul Glover-Kapfer, and Kate E. Jones. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2):169–185, 2019. doi: 10.1111/2041-210X.13101. URL `https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13101`.

[4] Daniel T. Blumstein, Louise M. Turner, Esther B. Reynolds, Adam W. Barlow, and et al. Terrestrial passive acoustic monitoring: Review and perspectives. *BioScience*, 61:203–214, 2011. doi: 10.1525/bio.2011.61.3.8. URL `https://academic.oup.com/bioscience/article/61/3/203/246016`.

[5] Connor M Wood, Jacob Socolar, Stefan Kahl, M Zachariah Peery, Philip Chaon, Kevin Kelly, Robert A Koch, Sarah C Sawyer, and Holger Klinck. A scalable and transferable approach to combining emerging conservation technologies to identify biodiversity change after large disturbances. *Journal of Applied Ecology*, 61(4):797–808, 2024.

[6] Christos Astaras, Joshua M. Linder, Peter Wrege, Robinson Orume, Paul J. Johnson, and David W. Macdonald. Boots on the ground: The role of passive acoustic monitoring in evaluating anti-poaching patrols. *Environmental Conservation*, 47(3):213–216, 2020. doi: 10.1017/S0376892920000193. URL `https://www.cambridge.org/core/journals/environmental-conservation/article/boots-on-the-ground-the-role-of-passive-acoustic-monitoring-in-evaluating-antipoaching-patr` `DF7099C1489D362F21C80A5C017140C9`.

[7] Peter H. Wrege, Elizabeth D. Rowland, Sara Keen, and Yu Shiu. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods in Ecology and Evolution*, 8(10):1292–1301, 2017. doi: 10.1111/2041-210X.12730. URL `https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12730`.

[8] Iosif Mporas, Isidoros Perikos, Vasilios Kelefouras, and Michael Paraskevas. Illegal logging detection based on acoustic surveillance of forest. *Applied Sciences*, 10(20):7379, 2020. doi: 10.3390/app10207379. URL `https://www.mdpi.com/2076-3417/10/20/7379`.

[9] George Wittemyer, Joseph M. Northrup, Julian Blanc, Iain Douglas-Hamilton, Patrick Omondi, and Kenneth P. Burnham. Illegal killing for ivory drives global decline in african elephants. *Proceedings of the National Academy of Sciences*, 111(36):13117–13121, 2014. doi: 10.1073/pnas.1403984111. URL `https://www.pnas.org/doi/10.1073/pnas.1403984111`.

[10] Sean L. Maxwell, Richard A. Fuller, Thomas M. Brooks, and James E. M. Watson. Biodiversity: The ravages of guns, nets and bulldozers. *Nature*, 536:143–145, 2016. doi: 10.1038/536143a. URL `https://doi.org/10.1038/536143a`. 11 August 2016.

[11] William J. Ripple, Thomas M. Newsome, Christopher Wolf, Rodolfo Dirzo, Kristoffer T. Everatt, Matthew Gale, Mike Hayward, and et al. Bushmeat hunting and extinction risk to the world's mammals. *Royal Society Open Science*, 3:160498, 2016. doi: 10.1098/rsos.160498. URL `https://royalsocietypublishing.org/doi/10.1098/rsos.160498`.

[12] Dan Stowell. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, 10:e13152, 2022. doi: 10.7717/peerj.13152. URL `https://peerj.com/articles/13152/`.

[13] Ankit Chouhan and Nitin Gupta. A review of automated bioacoustics and general acoustics classification. *International Journal of Scientific Research in Science, Engineering and Technology*, 9:94–99, 2022. doi: 10.32628/IJSRSET229520. URL `https://www.entomoljournal.com/archives/2023/vol11issue5/PartA/11-4-33-915.pdf`.

[14] K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology. Elephant listening project, n.d. URL `https://www.elephantlisteningproject.org/`.

[15] World Wildlife Fund. Wildlife crime technology project. `https://www.worldwildlife.org/projects/wildlife-crime-technology-project`, 2022.

[16] Timothy Lynam, Van-Truong Nguyen, Adam Zbyryt, Zachary Wendt, Maria Rulli, Eoin R. White, Hoai-Schmidt Nguyen, Brenda Low, Katie Leggett, Stephanie Brittain, and Anna Nekaris. The rising tide of conservation technology: Empowering the fight against poaching and unsustainable wildlife harvest. *Frontiers in Ecology and Evolution*, 12:1527976, 2025. doi: 10.3389/fevo.2025.1527976.

[17] Jakub Bajzik, Jiri Prinosil, and Dusan Koniar. Gunshot detection using convolutional neural networks. In *2020 24th International Conference Electronics*, pages 1–5. IEEE, 2020.

[18] Alex Morehead, Lauren Ogden, Gabe Magee, Ryan Hosler, Bruce White, and George Mohler. Low cost gunshot detection using deep learning on the raspberry pi. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 3038–3044. IEEE, 2019.

[19] Lydia KD Katsis, Andrew P Hill, Evelyn Pina-Covarrubias, Peter Prince, Alex Rogers, C Patrick Doncaster, and Jake L Snaddon. Automated detection of gunshots in tropical forests using convolutional neural networks. *Ecological Indicators*, 141:109128, 2022.

[20] Dena Jane Clink Thinh Tien Vu, Thai Son Le et al. Investigating hunting in a protected area in southeast asia using passive acoustic monitoring with mobile smartphones and deep learning. *Ecological Indicators*, 167:112501, 2024.

[21] NP García-de-la Puente, Félix Fuentes-Hurtado, Laura Fuster, Valery Naranjo, and Gema Piñero. Deep learning models for gunshot detection in the albufera natural park. In *2023 31st European Signal Processing Conference (EUSIPCO)*, pages 206–210. IEEE, 2023.

[22] Daniel S Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D Cubuk, and Quoc V Le. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*, 2019.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction claim (i) a lightweight gunshot detection model suitable for on-board processing and (ii) a sensor integration function to reduce false positives. This accurately reflects the paper's contributions and scope, as reflected by Sec. 2 and 3, where a lightweight model is proposed, architecture and training is described, results validated against existing literature using the same dataset is provided, and for the sensor integration, the SAIL function is proposed and evaluated on a spatially distinct dataset through acoustic simulation. The results (0.91 F1, 935k params, consistent FPR reduction through SAIL across simulations) provide evidence for and back up the claims made in the abstract and introduction, as it was motivated that the bottlenecks to real-time gunshot detection in field environments were model file size and false positive rates.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In the conclusion, the author discusses that the results of the acoustic simulations do not directly extend to FPR reduction in real field deployments, identifying the limitations of the evaluation of SAIL. Additionally, in Sec. 2.1 (Datasets), it is acknowledged that larger datasets exist for the acoustic simulation, but limited compute resources constrained that size.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.

- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: There are no theorems or mathematical proofs in this paper. While the acoustic simulations could be considered a theoretical result, the limitations are acknowledged in the conclusion, identifying that the results do not directly carry over to real field deployments.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Key details appear in Methods: data partitions, spectrogram parameters, augmentations, model composition, training epochs, optimizer/loss, seeds, and selection criteria, acoustic simulation design, function description, and statistical testing description. While more detail could be given to the exact model architecture and the exact code used for the acoustic simulation, we were unable to provide them within the paper due to page (four pages max) constraints. The developed model and code is available in the linked repository.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The manuscript states the developed model, data-processing code, and SAIL are available at the linked repository given in Fig. 1; datasets used are public and cited (Katsis et al., Vu et al.) with instructions to access them in the repo.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Key details appear in Methods: data partitions (train/validation/test described in Datasets), spectrogram parameters, augmentations, model composition, training epochs, optimizer/loss, seeds, and selection criteria; additional exact hyperparameters and scripts are provided in the linked repository.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Statistical testing (Wilcoxon signed-rank) is reported in Table 2 with p-values and 95% CIs, with the bootstrap n provided, in Sec. 3.2, and the method of performing the Wilcoxon signed-rank provided in Sec. 2.4.1; model variation across five seeds and boxplots/PR curves (Fig. 3) are also provided.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: In Sec. 2.3, the paper reports training time and computer resources ("two hours on a personal computer with a GTX 1060 GPU"). Neither the experiments nor the datasets were large enough that memory or storage were relevant.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The work uses public datasets, aims to aid conservation efforts, does not involve human or animal subjects in its experiments, and no actions appear to conflict with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The manuscript discusses positive conservation impacts (Introduction, Future Work) but does not explicitly discuss potential negative societal impacts or misuse (e.g., surveillance or dual-use risks) nor mitigation strategies because the scope and content of the work was identified to not require explicit discussion of potential negative societal impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The scope, content, and supplementary material pose no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: In Sec. 2.1 (Datasets), the datasets used in the paper are properly cited, credited, and the license for each (CC BY 4.0) is explicitly stated.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Assets (SAIL function, developed model) are introduced, well documented, and provided through an anonymous GitHub repository link.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.