Toward Multimodal Models of Animal Communication: Insights from New York City Rats

Anonymous Author(s)
Affiliation
Address
email

Abstract

Urban rats thrive in cities worldwide, yet little is known about the communicative behaviors that support their survival. We conducted fieldwork on free-ranging rats in New York City, combining thermal imaging and ultrasonic audio recordings with artificial intelligence techniques for movement and acoustic analysis. Our approach captures fine-scale locomotion, reconstructs the 3D geometry of foraging environments such as subways, streets, and parks, and quantifies vocalizations across ecological contexts. This work sets the stage for a larger scale species-agnostic proposal to use integrated behavioral analysis systems and multimodal modeling frameworks to study bioacoustics in the wild.

1 Introduction

Understanding animal communication in real-world environments requires tools that can bridge laboratory science, field biology, and artificial intelligence. While rodents and other species have long been studied in controlled settings, the dynamics of social signaling, collective behavior, and vocal communication in natural environments are poorly characterized. Urban rats present a compelling model for the study of vocal communication: they are abundant, highly social, produce rich repertoires of vocalizations, and are closely embedded in human-dominated landscapes [8, 16, 23, 10, 5, 31, 20, 4]. This work is crucial for informing rodent mitigation efforts, city design, and controlling disease spread. Moreover, an understanding of behavior in a rat's natural urban ecosystem is essential for understanding the true biological relevance of rodent vocalizations, which remains mysterious. Advances in AI for movement tracking, 3D environmental reconstruction, and bioacoustic analysis now make it possible to study how communication and group behavior unfold in these naturalistic contexts [21, 15, 18, 34, 7, 27, 28, 9, 26].

We propose to use wild rats in New York City as a testbed for developing a multimodal analysis system that integrates locomotion, vocalization, and 3D environmental reconstructions to uncover how communication is shaped by ecological context. By combining thermal imaging and ultrasonic audio recordings, we quantify locomotion and vocalization within urban environments such as subways, streets, and parks. This approach highlights that communication as a key component of rat behavior in the wild and reveals that wild rat vocalizations are used more flexibly than previously noted in laboratory studies. Our overarching goal is to establish a methodological blueprint for studying animal communication under natural conditions, with applications across neuroscience, ecology, and artificial intelligence. Although the work presented herein applies to rats, the framework is inherently species-agnostic and stands to generalize easily to other biological domains.

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

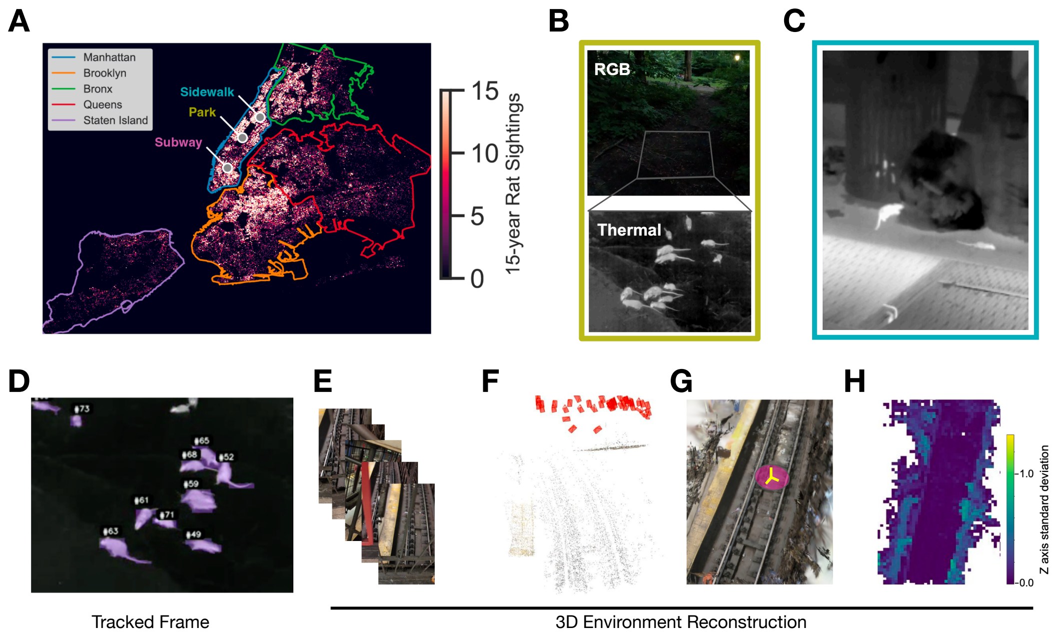

Figure 1: **New York City fieldwork, rat tracking, and 3D environment reconstruction (A)** Spatial heatmap of rat sightings over a 15-year period generated from NYC311 reports. Recording locations from the study are labeled with gray points (Subway, Park, Sidewalk). **(B)** (top) RGB image of an area where rats were observed foraging in Central Park, New York, NY. Gray outline indicates field of view for thermal imaging camera. (bottom) Frame from a thermal video of rats foraging in the area outlined above. **(C)** Thermal image of a rat exploring a trash bag on the sidewalk. **(D)** Thermal video frame overlayed with bounding masks of rats detected with YOLO. **(E)** Example RGB images of a subway scene from different angles. **(F)** Inferred camera positions and points in the 3D scene, from the COLMAP algorithm. **(G)** 3D scene composed of gaussians inferred by the gaussian splat algorithm. **(H)** Top-down heatmap of this 3D scene, showing the standard deviation in the Z axis of gaussian locations.

2 Results and Methods

2.1 Quantifying rat behavior and environment in New York City

We mined NYC311 reports of rat sightings to target field recording sessions to high-density sites in Manhattan (Figure 1A). There is a seasonality to rat sightings, with more rats reported in summer months, thus we collected data throughout July 2024 from three distinct urban contexts — subway, park, and sidewalk. Rats prefer to forage in the evening, and in shadowed areas, which poses a challenge for standard videography methods, but thermal videography (FLIR E54) made it possible to visually resolve rodents, even in dark shadowed areas (Figure 1B-C). Using these methods, it was possible to observe groups of rats foraging in a variety of urban environments, including even areas with some occlusion due to fencing and underbrush (data not shown). While multi-object tracking is a well-researched area in AI and computer vision [14], and in particular in the context of animal tracking [22, 15, 34, 13, 25, 30], applying the different existing approaches in our field study to extract accurate rat tracks from thermal videos (Figure 1D) proved to be a nontrivial task. In outdoor real-world environments, specific challenges include varying number of animals per frame, a wide distribution of animal sizes, multiple animal types (e.g. rats and squirrels), occlusions, and inaccurate and missing detections. Our detection and tracking pipeline included the most recent version of the YOLO (You Only Look Once) model [24, 12], fine-tuned on 50 hand-labeled frames from our thermal videos of rats foraging, combined with the ByteTrack tracking algorithm [35], which is robust to occlusions and missing track segments, and utilizes low-score detections and Kalman filtering for predicting new locations.

Recent advances in computer vision make it possible to reconstruct 3D models of a scene from a sparse collection of camera angles [17, 11]. Urban environments where rodents forage pose some potential challenges, because they can be highly dynamic spaces, with shadowed areas, and limited

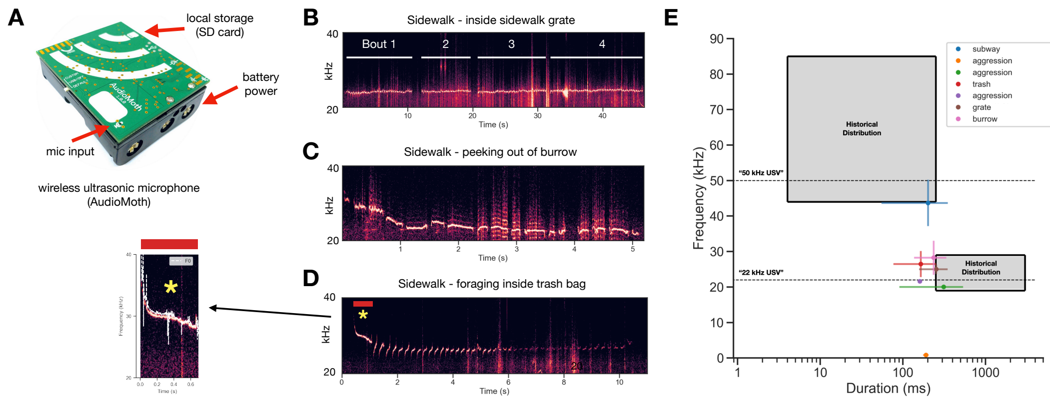

Figure 2: **Ultrasonic recordings of NYC rat vocalizations.** **(A)** Field recordings were taken using an Audiomoth wireless ultrasonic microphone. **(B)** Multiple bouts of vocalizations recorded from a sidewalk grate where rats were frequently seen entering and exiting. **(C)** Rat cautiously poking its head out of a burrow hole from a sidewalk tree lawn. **(D)** Bout of calls recorded from a rat foraging inside of a trashbag on the sidewalk. (inset) First syllable of bout, showing the time-varying fundamental frequency of the vocalization (F0, white dashed line). **(E)** Projection of the vocalizations bouts recorded in this study into duration-frequency space detailed in [33]. Gray shaded regions denote the historical distribution of durations and frequencies reported in an expansive literature search from [33]. Horizontal dashed lines denote the two predominant vocalization types studied in rats. Legend indicates which context the vocalizations were observed in.

access. We wondered whether it was possible to reconstruct 3D models of such environments using data acquired with a single handheld RGB camera (GoPro). Multiple camera views were collected of a subway scene (Figure 1E), and run through the gaussian splatting model, which first involves estimating camera positions using the COLMAP algorithm (Figure 1F), then constructing a 3D model in the form of collection of 3D gaussians that best explain the data. These models capture environmental geometry at a high level of detail (Figure 1G). The models also allow for quantification of aspects of the environmental statistics that may be relevant to rodent foraging, such as the degree of shelter versus open-field. We found that this could be captured by analyzing the standard deviation of gaussian centers in the z-axis (Figure 1H).

2.2 Ultrasonic recordings of rat vocalizations in New York City

There have been few attempts to document the social vocal lives of rats in their natural urban habitat. Here, we used a wireless ultrasonic microphone (Figure 2A) to record vocalizations emitted by rats in different urban environments during social interactions. A session was acquired if rat activity was detected or suspected to be observed in a given area. The device was either handheld or placed on a surface pointed towards the area of interest at a distance of approximately 2 meters. Next, we extracted vocalization annotations from raw audio using a deep neural network (Deep Audio Segmenter, [29]). Vocalizations mostly occurred in bouts (sequences of vocalizations separated by less than 250 ms of silence) and were observed in various types of social interactions (Figure 2B-D, see legend of E).

Next, we calculated acoustic features of vocalizations and compared them to a large-scale meta-analysis of previously published rat vocalization features. Figure 2D shows an example vocalization bout with a DAS-annotated onset and offset of the first syllable in the bout (red). Next, we estimated the fundamental frequency at each time point in the spectrogram (Figure 2D left, using VocalPy [19]) and calculated the median for downstream analysis. We computed the median frequency +/- 1 standard deviation for all bouts recorded in the study, then projected their measurements into a duration vs. frequency feature space detailed by [33] (Figure 2E). We find that wild NYC rat vocalizations are consistently shorter duration and lie outside of the historical frequency-duration range reported in the meta-analysis.

3 Discussion

This study validates a computational toolkit to quantify high-resolution movement and acoustic behavior of rats living in New York City. In this unconstrained urban environment, we can track large groups of rats (Figure 1B-D), reconstruct 3D environments in which rats operate (Figure 1E-H), and record ultrasonic vocalizations (Figure 2). The computational techniques are all open-source, and the recording technology is non-invasive with widely available hardware, which will make reproducing this study in other animals quite straightforward. We were able to map variations in foraging speeds and coordination of movements for rats of different sizes (data not shown), compare 3D environmental statistics across different places where rats forage (data not shown), and classify ultrasonic vocalizations rats use in different contexts, finding vocal structure distinct from the distribution of rat vocalizations commonly studied in lab environments. It is generally thought that 22 and 50 kHz vocalizations signal aversive and appetitive contexts, respectively [4]. Here, we observe that 22 kHz vocalizations are used in diverse contexts, some of which are seemingly not aversive. For example, a long bout of near-22 kHz USVs was emitted while a single rat foraged inside of a trash bag (Figure 2D). Rats have not been reported to emit 22 kHz vocalizations while foraging in laboratory settings; instead, studies have shown that 22 kHz calls actually suppress feeding behavior [3]. Given that these data are mostly proof-of-concept for the toolkit at large, the observations are relatively underpowered, which is a limitation of this study. Future work will include more vocalization examples from more sites.

By situating vocal behavior within a multimodal framework, our study advances the idea that communication must be interpreted relative to the surrounding environment and concurrent behavior. Foraging inside trash bags or navigating crowded sidewalks involves different environmental affordances than standard laboratory cages, and our data show that vocal patterns vary accordingly. Multimodal analysis — integrating ultrasonic calls, body movements, and environmental structure — is therefore essential for disentangling the meaning of vocalizations in natural settings. This perspective moves beyond acoustic categorization alone, framing animal communication as a multimodal phenomenon that encodes context-specific information about the animal's social and ecological world.

4 Conclusion and Future Directions

The computational tools we applied — open-source models for tracking, acoustic segmentation, and 3D reconstruction — provide a generalizable pipeline for studying communication across species. The fact that these tools could be deployed with relatively lightweight hardware in urban fieldwork underscores their potential to democratize studies of natural communication.

Rats demonstrate an impressive ability to survive in rapidly changing urban environments, but the question of what cognitive strategies they use remains open. With sufficient data across a range of environmental conditions, it may be possible to infer cognitive strategies from unconstrained rodent foraging behavior, using a variety of recent computational techniques [32, 1, 2, 6]. A natural next step is to develop multimodal "world models" that integrate locomotion, vocalizations, and environmental mapping into unified representations. Such models could allow us to infer not only when and where calls occur but also how their function is shaped by the affordances of the environment and the dynamics of group behavior. For example, embedding vocal features alongside measures of shelter, openness, and movement coordination could reveal whether calls serve to signal threat, recruit conspecifics, or coordinate foraging.

Future studies should prioritize longitudinal and large-scale data collection, ideally with individual tagging, to follow the same animals across contexts and over time. Integrating environmental soundscapes — including anthropogenic noise and biological sounds from other species — will further enrich these models, enabling us to ask how animals filter relevant signals from complex acoustic backgrounds. The ultimate goal is to build species-agnostic frameworks where AI systems infer communicative meaning by linking signals to multimodal context. Such approaches could not only transform our understanding of rodent vocalizations but also provide a roadmap for studying natural communication across a wide range of taxa.

References

[1] Zoe C Ashwood et al. "Mice alternate between discrete strategies during perceptual decision-making". In: *Nature Neuroscience* 25.2 (2022), pp. 201–212.

[2] Bence Bagi, Michael Brecht, and Juan Ignacio Sanguinetti-Scheck. "Unsupervised discovery of behaviorally relevant brain states in rats playing hide-and-seek". In: *Current Biology* 32.12 (2022), pp. 2640–2653.

[3] Robert J Blanchard and D Caroline Blanchard. "Antipredator defensive behaviors in a visible burrow system." In: *Journal of comparative psychology* 103.1 (1989), p. 70.

[4] Stefan M Brudzynski. "Biological functions of rat ultrasonic vocalizations, arousal mechanisms, and call initiation". In: *Brain Sciences* 11.5 (2021), p. 605.

[5] Kaylee A Byers et al. "Rats about town: a systematic review of rat movement in urban ecosystems". In: *Frontiers in Ecology and Evolution* 7 (2019), p. 13.

[6] Pablo Samuel Castro et al. "Discovering Symbolic Cognitive Models from Human and Animal Behavior". In: *bioRxiv* (2025). DOI: 10.1101/2025.02.05.636732. eprint: https://www.biorxiv.org/content/early/2025/02/06/2025.02.05.636732.full.pdf. URL: https://www.biorxiv.org/content/early/2025/02/06/2025.02.05.636732.

[7] Alex Hoi Hang Chan et al. "YOLO-Behaviour: A simple, flexible framework to automatically quantify animal behaviours from videos". In: *Methods in Ecology and Evolution* 16.4 (2025), pp. 760–774.

[8] Kevin J Gaston. *Urban ecology*. Oxford University Press, 2010.

[9] Jack Goffinet et al. "Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires". In: *Elife* 10 (2021), e67855.

[10] Chelsea G Himsworth et al. "Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers". In: *Vector-Borne and Zoonotic Diseases* 13.6 (2013), pp. 349–359.

[11] Bernhard Kerbl et al. *3D Gaussian Splatting for Real-Time Radiance Field Rendering*. 2023. arXiv: 2308.04079 [cs.GR]. URL: https://arxiv.org/abs/2308.04079.

[12] Rahima Khanam and Muhammad Hussain. "YOLOv11: An Overview of the Key Architectural Enhancements". In: arXiv:2410.17725 (Oct. 2024). DOI: 10.48550/arXiv.2410.17725. arXiv: 2410.17725 [cs]. (Visited on 07/16/2025).

[13] Jessy Lauer et al. "Multi-animal pose estimation, identification and tracking with DeepLabCut". In: *Nature Methods* 19.4 (2022), pp. 496–504.

[14] Wenhan Luo et al. "Multiple Object Tracking: A Literature Review". In: *Artificial Intelligence* 293 (Apr. 2021), p. 103448. ISSN: 0004-3702. DOI: 10.1016/j.artint.2020.103448. (Visited on 07/16/2025).

[15] Mackenzie Weygandt Mathis and Alexander Mathis. "Deep learning tools for the measurement of animal behavior in neuroscience". In: *Current opinion in neurobiology* 60 (2020), pp. 1–11.

[16] Mark J McDonnell and J Niemelä. "The history of urban ecology". In: *Urban ecology: Patterns, processes, and applications* (2011), pp. 5–13.

[17] Ben Mildenhall et al. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. 2020. arXiv: 2003.08934 [cs.CV]. URL: https://arxiv.org/abs/2003.08934.

[18] Victoria Ngo et al. "Active vision during prey capture in wild marmoset monkeys". In: *Current Biology* 32.15 (2022), pp. 3423–3428.

[19] David Nicholson. *vocalpy*. Version 0.10.0.post1. Nov. 2024. URL: https://github.com/vocalpy/vocalpy.

[20] Michael H Parsons, Ronald J Sarno, and Michael A Deutsch. "Jump-starting urban rat research: conspecific pheromones recruit wild rats into a behavioral and pathogen-monitoring assay". In: *Frontiers in Ecology and Evolution* 3 (2015), p. 146.

[21] Talmo D Pereira, Joshua W Shaevitz, and Mala Murthy. "Quantifying behavior to understand the brain". In: *Nature neuroscience* 23.12 (2020), pp. 1537–1549.

[22] Talmo D. Pereira et al. "SLEAP: A deep learning system for multi-animal pose tracking". In: *Nature Methods* (2022). DOI: 10.1038/s41592-022-01426-1.

[23] Cristina E Ramalho and Richard J Hobbs. "Time for a change: dynamic urban ecology". In: *Trends in ecology & evolution* 27.3 (2012), pp. 179–188.

[24] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 779–788. URL: https://arxiv.org/abs/1506.02640.

[25] Tianhe Ren et al. *Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks*. 2024. arXiv: 2401.14159 [cs.CV]. URL: https://arxiv.org/abs/2401.14159.

[26] David Robinson et al. "NatureLM-audio: an Audio-Language Foundation Model for Bioacoustics". In: *arXiv preprint arXiv:2411.07186* (2024).

[27] Tim Sainburg, Marvin Thielk, and Timothy Q Gentner. "Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires". In: *PLoS computational biology* 16.10 (2020), e1008228.

[28] Sarab S Sethi et al. "Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set". In: *Proceedings of the National Academy of Sciences* 117.29 (2020), pp. 17049–17055.

[29] Elsa Steinfath et al. "Fast and accurate annotation of acoustic signals with deep neural networks". In: *Elife* 10 (2021), e68837.

[30] Jennifer J. Sun et al. "Video Foundation Models for Animal Behavior Analysis". In: *bioRxiv* (2024). DOI: 10.1101/2024.07.30.605655. eprint: https://www.biorxiv.org/content/early/2024/07/31/2024.07.30.605655.full.pdf. URL: https://www.biorxiv.org/content/early/2024/07/31/2024.07.30.605655.

[31] Stephen Takács, Pawel Kowalski, and Gerhard Gries. "Natural and synthetic vocalizations of brown rat pups, Rattus norvegicus, enhance attractiveness of bait boxes in laboratory and field experiments". In: *Pest management science* 72.10 (2016), pp. 1873–1882.

[32] Rafal Urbaniak, Marjorie Xie, and Emily Mackevicius. "Linking cognitive strategy, neural mechanism, and movement statistics in group foraging behaviors". In: *Scientific Reports* 14.1 (2024), p. 21770.

[33] Reo Wada et al. "Beyond dichotomy: diversity of rat's ultrasonic vocalizations". In: *bioRxiv* (2024), pp. 2024–12.

[34] Tristan Walter and Iain D. Couzin. "TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields". In: *eLife* 10 (2021). DOI: 10.7554/eLife.64000.

[35] Yifu Zhang et al. "ByteTrack: Multi-object Tracking by Associating Every Detection Box". In: *Computer Vision – ECCV 2022*. Ed. by Shai Avidan et al. Cham: Springer Nature Switzerland, 2022, pp. 1–21. ISBN: 978-3-031-20047-2. DOI: 10.1007/978-3-031-20047-2_1.

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 results in this paper are relatively under-powered and prelimninary, therefore no strong claims are made. Rather, this paper — as a submission to the "proposals" track — outlines a framework and demonstrates proof-of-concept for that framework.

   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: See Section 3 for discussion of limitations.

   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: The paper does not include theoretical results.

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: We will provide code (open-sourced for camera-ready version). Section 2 describes how the fieldwork data were acquired.

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: [No]

Justification: Upon camera-ready submission, we will provide open source access to our code and data, with detailed instructions for reproducing the reported experimental results.

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: [No]

Justification: These details were omitted due to space constraints for the initial submission. All model training details will be included in the supplement in the camera-ready submission.

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: [NA]

Justification: This is N/A for most of the paper which provides qualitative proof-of-concept results of technical advances.

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: [No]

   Justification: These details were omitted due to space constraints for the initial submission. All compute resource details will be included in the supplement in the camera-ready submission.

   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:

   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: [Yes]

    Justification: Yes, this is discussed Section 1.

    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 paper does not pose risks of misuse.

    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: [NA]

    Justification: The paper does not use existing assets.

    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: [NA]

    Justification: The paper does not release new assets

    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: Our paper involves neither crowdsourcing experiments 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: Our paper involves neither crowdsourcing experiments 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: We do not use LLMs in any important, original, or non-standard component of our research.

    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.