# Probing the lighting sensitivity of image encoders with repeat drone imagery: A case study of plant height estimation

**Kyle Doherty**[1], **Max Gurinas**[2], **Brandon Trabucco**[3], **Ruslan Salakhutdinov**[3]

[1] MPG Ranch, Aerial Survey Program, Missoula, MT, USA (`kdoherty@mpgranch.com`)
[2]Harvard University, Department of Mathematics and Computer Science, Cambridge, MA, USA
[3]Carnegie Mellon University, Department of Machine Learning, Pittsburgh, PA, USA

## Abstract

Image encoders provide a strong backbone for tasks such as retrieval, classification, and depth estimation, and recent releases tailored to remote sensing, such as DINOv3 with SAT pretraining [Siméoni et al., 2025], promise improved performance on ecologically important applications. It is uncertain whether such encoders yield robust features under variable illumination, where shadows and hue shifts can obscure relevant plant structure. To address this, we developed a drone imagery dataset with high-resolution RGB captures of the same site at three time points in a single day, paired with plant canopy height models. Using these data we identified the subspace of embeddings dominated by lighting variation and progressively projected embeddings away from lighting subspace components. Across both DINOv2 and DINOv3, canopy height prediction remained stable until more than 80% of the lighting variance was removed, after which performance degraded sharply, with a pronounced error spike when the full lighting subspace was eliminated. These results suggest that while much of the lighting variance is nuisance, the final fraction contains useful textural and chromatic cues. DINOv3-SAT consistently outperformed the general-purpose DINOv2, maintaining ~1 cm lower error until complete removal of the lighting subspace. We release the Drone-LSR dataset on Hugging Face under a Creative Commons 4.0 license to facilitate further exploration of lighting sensitivity in image encoders for remote sensing.

## 1 Introduction

Foundational image encoders are widely used in computer vision for tasks such as classification, retrieval, depth estimation, and segmentation. Because they are trained on very large and diverse image datasets, they can be adapted to many different domains. Well-known examples include CLIP [Radford et al., 2021], DINOv2 [Oquab et al., 2023], and DINOv3 [Siméoni et al., 2025], all of which achieve strong results across a range of applications. Importantly, Siméoni Siméoni et al. [2025] introduced versions of DINOv3 specifically designed for aerial imagery, a form of earth observation data that differs substantially from the internet-based collections typically used for pretraining.

Remotely sensed earth observation data are predominantly of top-down perspective gathered by satellites, manned aircraft, and unmanned aerial vehicles (UAV's; drones). The most common form is optically sensed visible light imagery, though multi- and hyper-spectral sensors capturing non-visible wavelengths are also abundant and strongly correlate with physical and ecological phenomena [Rouse et al., 1973, Qi et al., 1994, Hall et al., 1995, McFeeters, 1996]. Some ecological targets for prediction with remotely sensed data include land cover [Justice et al., 1998], plant species

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Imageomics: Discovering Biological Knowledge from Images Using AI.

composition [Feilhauer et al., 2017], chlorophyll content [Rouse et al., 1973], plant canopy height [Harris et al., 2021, Tolan et al., 2024], and others.

In aerial imagery, lighting conditions may vary dramatically, which can impact performance on downstream tasks. In particular, cast shadows from clouds, terrain, vegetation, and built structures can obscure critical features within imagery and harm biophysical indices, and identification of shadows is a critical post-processing step in remote sensing [Hagolle et al., 2010, Zhu and Woodcock, 2012, Coleman et al., 2020, Alavipanah et al., 2022]. Relatedly, there is a rich literature base targeting removal of shadows, including neural architectures, to address this challenge [Li et al., 2022, Liu et al., 2022, Shao et al., 2025]. In some instances, shadows may enhance, e.g., tree detection, as trees are likely to cast longer shadows than lower-lying vegetation [Hung et al., 2011]. However, because shadows vary greatly with solar angle and atmospheric conditions, systems dependent on these features are brittle. How sensitive aerial-specific foundation models like DINOv3 are to lighting variation is largely unknown, despite illumination being one of the main factors that can alter the quality of remote sensing products.

We contribute the Drone-LSR dataset [Doherty et al., 2025] to explore the sensitivity of image encoders to variable lighting conditions. In the dataset, high resolution aerial images of the same natural scenes were captured at three time points in a single day. The scenes were identical except for dramatic differences in illumination driven by change in solar position. With these data we studied the lighting sensitivity of DINO encoders [Oquab et al., 2023, Siméoni et al., 2025] by incrementally removing the variance attributed to lighting components of the feature space, and evaluated the effects on downstream error in plant canopy height prediction. We expected models to have rich representations for lighting and that a plant height decoder would exploit these features, as taller plants cast longer shadows. Furthermore, we expected performance would degrade as we removed increasing amounts of variance in the lighting subspace. Our approach provides a controlled framework to test the robustness of foundation models to nuisance illumination and their capacity to retain ecologically relevant signal.

## 2 Methods

### 2.1 Site Description and Spatial Data Acquisition

The field site was located at MPG Ranch, Montana, USA. Previously this area served as a working cattle ranch, but the land is now managed for wildlife conservation. Our survey site (**Fig. 1**) was 12 hectares in area and ranged in elevation from 972 to 1009 meters. Vegetation types are varied and include reclaimed agricultural fields, restored native grasslands, and conifer woodlands with interspersed dirt access roads.

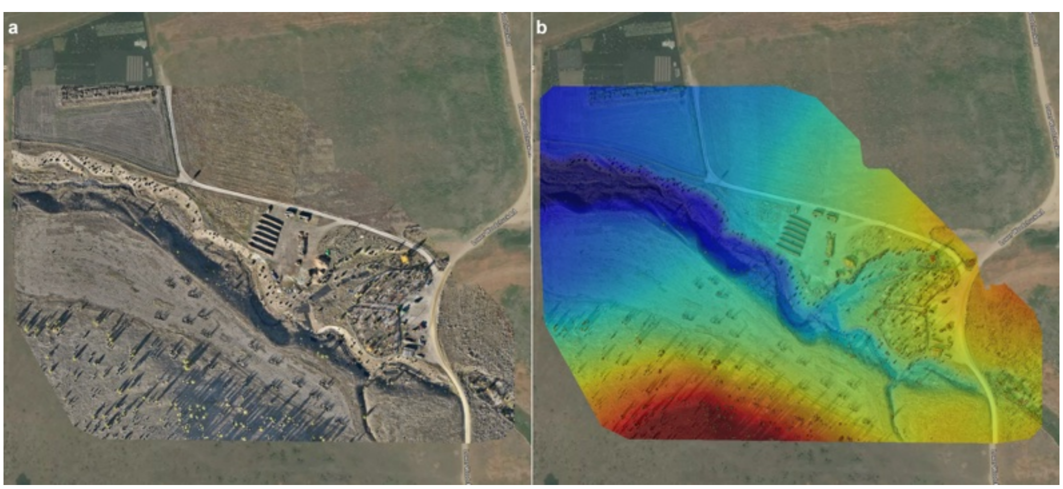


Figure 1: An orthoimage (**a**) of the 12-hectare survey site at MPG Ranch, MT, USA and accompanied elevation model (**b**).

We conducted drone surveys on November 9th, 2024, capturing the site with a DJI Mavic 3M with onboard real-time kinematic global navigation satellite systems. We gathered data at three time points: 10:00 am, 12:00 pm, and 3:00 pm, each survey lasting 25 minutes. We flew the drone at 50 meters above ground level, resulting in a ground sampling distance of 1.4 cm/pixel. We generated 2D orthomosaics and 3D point clouds with Drone Deploy (www.dronedeploy.com) using ground control to improve georeferencing, and this process resulted in three-dimensional spatial error of less than two centimeters.

To generate canopy height models for our study area, we combined drone-based point cloud data derived from photogrammetirc processes with a LiDAR-derived, bare-earth digital terrain model from the USGS 3D Elevation Project (3DEP; public domain; Sugarbaker et al. [2014]). For each survey time point we classified drone point cloud data as ground if the nearest neighbor in the 3DEP model differed by less than 1.4 centimeters along the z-axis (Z-error of drone survey). Then for each time point we produced canopy height models following normalization and rasterization methods described in [Roussel et al., 2020]. We then averaged canopy height across the three time points to generate a single composite model which we used as the target for prediction in subsequent analyses.

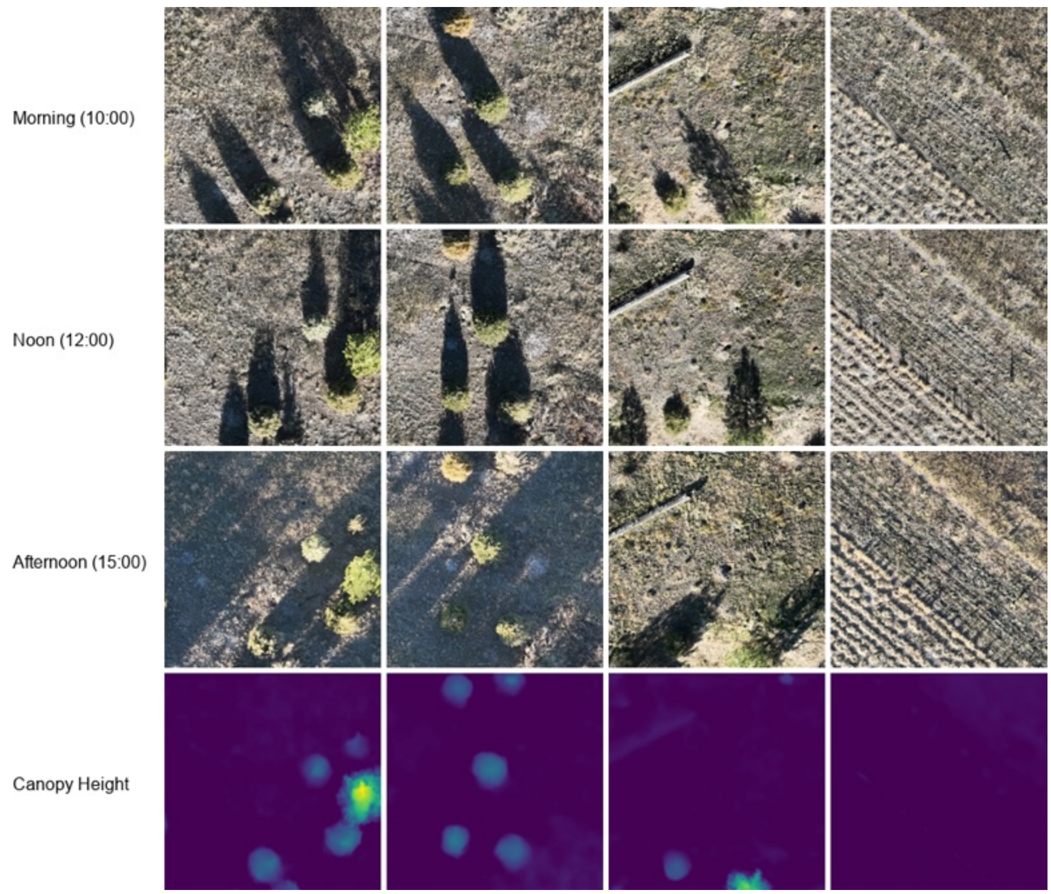


Figure 2: A sample of four tiles from the Drone-LSR dataset. Rows one through three contain images from the morning, noon, and afternoon captures, while the fourth row depicts the corresponding canopy height model targeted for prediction.

### 2.2 Dataset Preparation

We tiled the orthomosaics from each time point and canopy height model into squares that were 1024 pixels on a side (~14 meters; **Fig. 2**). A key concern in this work was maintaining scene consistency across time points, excepting changes to lighting. Therefore, an ecologist familiar with the site manually reviewed each scene in the dataset to identify any changes to elements other than lighting. We excluded any tile from further analysis if it contained transient objects, such as animals or cars present in the morning and absent in the afternoon. We found only 2% of scenes contained transient items that would alter scene semantics. After removing scenes with transient items, a total of 609 scenes remained, each with three RGB images corresponding to morning, noon, and afternoon scans and a corresponding canopy height model.

For each scene and time point, we pre-encoded RGB images using two vision transformer models: DINOv2-Base (ViT-B/14; `facebook/dinov2-base`; Oquab et al. [2023]; Apache 2.0) and DINOv3-Large with SAT pretraining (ViT-L/16; `facebook/dinov3-vitl16-pretrain-sat493m`; Siméoni et al. [2025]; custom license). Images were resized to $224 \times 224$ pixels and normalized using ImageNet statistics for DINOv2 [Russakovsky et al., 2015], and SAT-493M dataset statistics for DINOv3 [Siméoni et al., 2025]. Each encoding produced a class token and a set of patch tokens: class token dimensionality was 768 (DINOv2) and 1024 (DINOv3); patch tokens formed 2D arrays of shape [256, 768] and [196, 1024], respectively.

We serve the Drone-LSR Dataset on Hugging Face with a Creative Commons 4.0 license in three configurations: 1) `default` – containing RGB images for three time points and canopy height model, 2) `dinov2_base`, and 3) `dinov3_sat`, each containing their respective patch and class tokens. The dataset is split by tile index, with an 80/20, train/test regime. For further dataset details, refer to the Drone-LSR Hugging Face repository.

### 2.3 Lighting Subspace Removal Experiment

To test the sensitivity of encoders to illumination, we identified the variance associated with lighting and progressively removed it from embeddings. We pooled and mean-centered patch tokens from all three time points and then applied singular value decomposition to the resulting embedding matrix to extract orthogonal components ordered by explained variance. The leading components captured the dominant variation shared across time, which we interpret as being largely driven by lighting and shadows. We defined this set of components as the lighting subspace.

Each patch token was projected away from the first $k$ components of this subspace, yielding modified features that progressively excluded lighting-related variance (**Fig. 3**). These projected embeddings were then used to train a decoder to predict canopy height. This procedure allowed us to measure how prediction error changed as lighting variance was increasingly filtered from the representation. Because DINOv2 and DINOv3 differ in patch token dimensionality, we report the percentage of lighting-related variance removed rather than the raw number of components, for ease of interpretation.

### 2.4 Decoder Architecture, Training Procedure, and Evaluation Regime

We developed a lightweight convolutional decoder to map projected token embeddings to canopy height, following standard encoder–decoder designs [Ronneberger et al., 2015, Badrinarayanan et al., 2017] with GroupNorm [Wu and He, 2018] and GELU activations [Hendrycks and Gimpel, 2016]. We trained for 50 epochs with AdamW (learning rate 1e-3) and mean squared error loss.

We conducted five-fold cross-validation where folds were established randomly by scene indices and maintained across experiment configurations. Splitting at the level of scene ensured no spatial leakage among in-sample and holdout sets. We computed the best cross-validated epoch for each configuration and report performance as the cross-validated mean RMSE of canopy height (cm) and associated 95% confidence interval. In total there were 110 configurations: 2 Models × 11 lighting subspace conditions × 5 folds. This consumed an estimated 55 hours of compute across eight nodes, each with an Nvidia A6000 GPU, 32G RAM, and 8 CPU's.

## 3 Results and Discussion

Across both DINO models, error in canopy height prediction remained consistent while removing large fractions of the lighting subspace (**Fig. 4**). At 0% removal, canopy height predictions achieved mean RMSEs of 14.2[12.7,15.7] cm (95% CI's in square brackets; DINOv2-base) and 13.4[11.7,15.1] cm (DINOv3-sat), remaining stable through 80% removal (14.9[13.5,16.3] cm and 13.8[12.3,15.3] cm, respectively) before sharply degrading at 100%, where errors rose to 23.4[19.06,27.70] cm and 24.7[22.49,26.97] cm. The error spike at 100% suggests that critical semantic information was

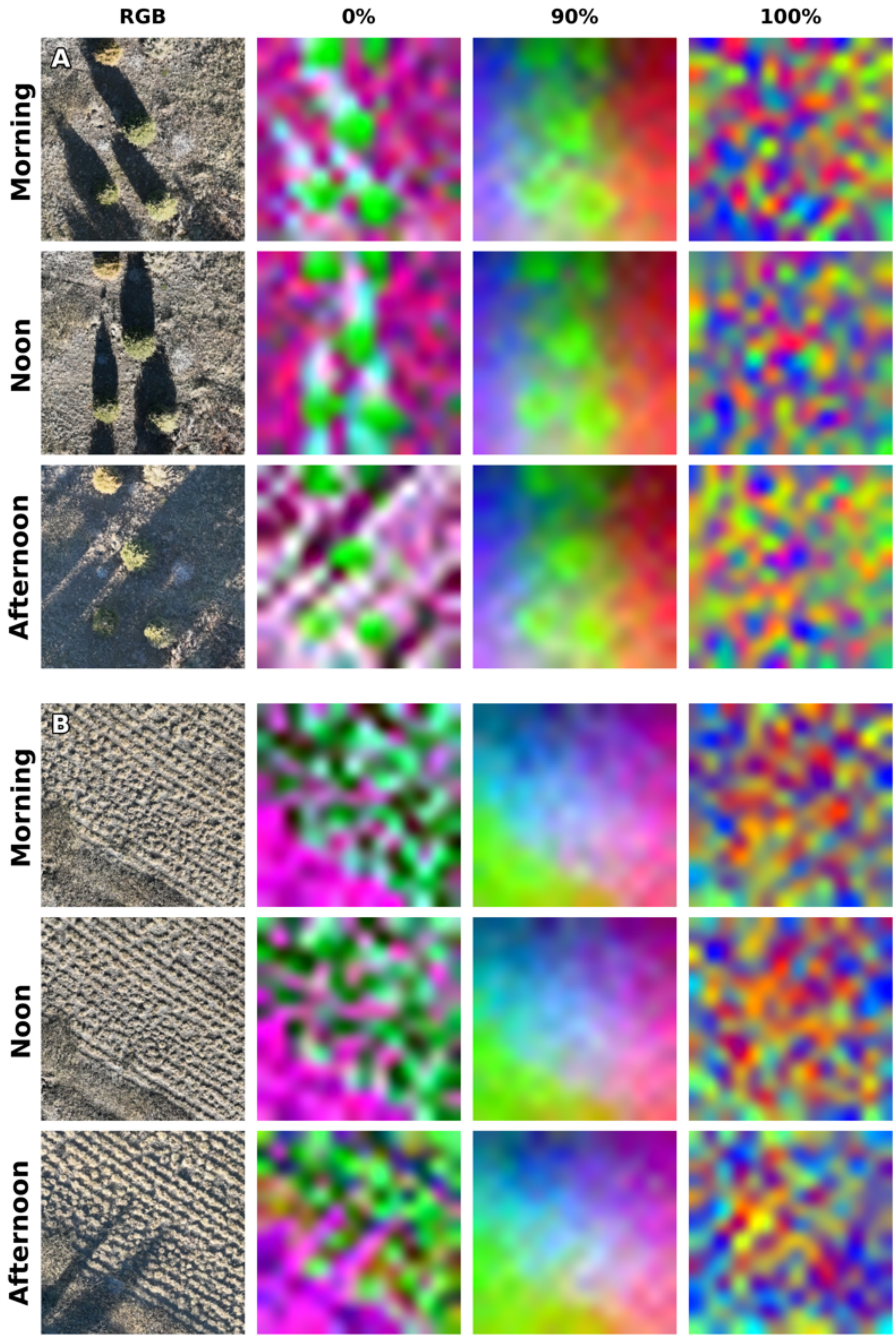


Figure 3: Two example scenes at three points with depictions of the DINOv3 feature space remaining after removal of $k$ components representing 0, 90, and 100% of the variance attributed to lighting. The first three components after lighting subspace removal are mapped to RGB colorspace. Note how shadows are apparent in feature space with 0% removal, yet difficult to resolve at 90%.

eliminated alongside lighting effects. The satellite-specialized DINOv3-Large (SAT pretraining; [Siméoni et al., 2025]) consistently, though only marginally, outperformed DINOv2 [Oquab et al., 2023], maintaining approximately 1 cm lower error until complete removal of the lighting subspace.

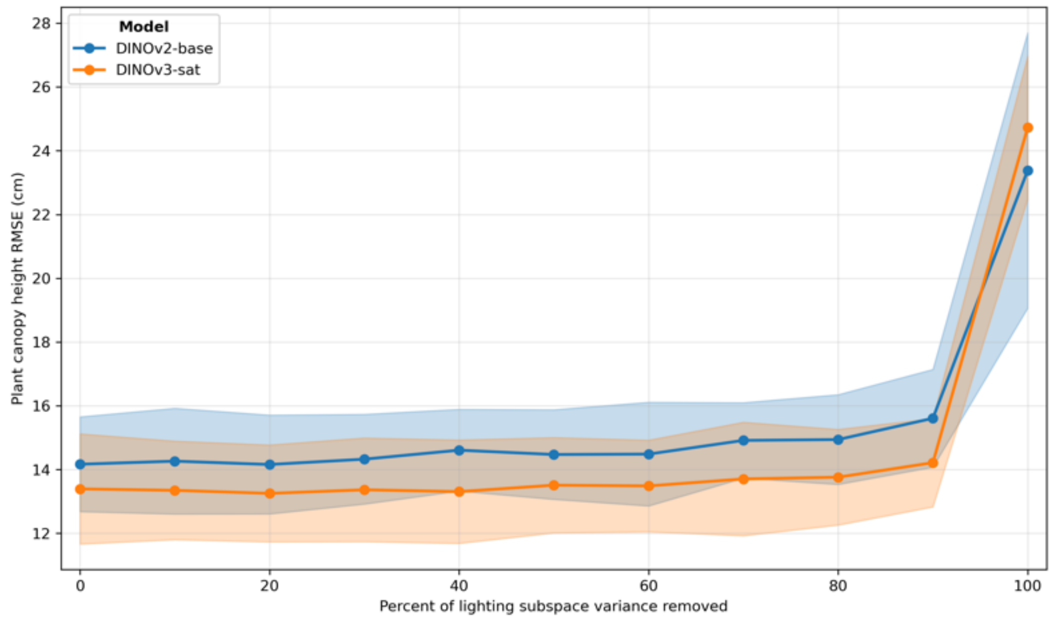


Figure 4: Performance of DINOv2-base and DINOv3-sat models on canopy height estimation from drone imagery under progressive removal of lighting subspace components. The x-axis shows the cumulative percentage of lighting-related variance removed (via SVD projection), and the y-axis shows canopy height prediction error (RMSE, cm).

We suspect that the final portion of the lighting subspace, while still explaining variance across time, contained ecologically relevant texture and hue information. Fine-scale shadow patterns cast by grass stems, broad leaves, or conifer needles likely carry distinct morphological signatures that aid discrimination among plant functional types. Similarly, because imagery was collected in autumn, senesced grasses presented muted hues that contrasted with evergreen vegetation; removing the full complement of lighting-related variance may have suppressed these subtle chromatic cues, thereby reducing the model's ability to distinguish evergreen plants from a chlorotic background.

Our analysis is limited to a single grassland–woodland site on a single calendar day, so results may differ in other contexts such as densely forested regions, open arid environments, or seasons with more uniform green foliage. Nevertheless, the Drone-LSR dataset demonstrates the value of rapidly repeated drone imagery for probing the behavior of image encoders. Our findings further indicate that both current and prior generations of encoders are robust to substantial variation in lighting, and that the DINO family in particular mitigates nuisance illumination effects that have previously challenged remote sensing analyses. These properties suggest strong potential for predicting other ecologically relevant traits beyond canopy height.

## 4 Data and Code Availability

We serve the Drone-LSR dataset on Hugging Face with a Create Commons 4.0 License. Please find code for lighting subspace removal experiments in the in the sub-directory of same name in our GitHub repository.

## 5 Acknowledgments

We would like to thank the owners of MPG Ranch for funding this work and supporting applied ecology and conservation research.

## References


Seyed Kazem Alavipanah, Mohammad Karimi Firozjaei, Amir Sedighi, Solmaz Fathololoumi, Saeid Zare Naghadehi, Samiraalsadat Saleh, Maryam Naghidazegan, Zinat Gomeh, Jamal Jokar Arsanjani, Mohsen Makki, Salman Qureshi, Qihao Weng, Dagmar Haase, Biswajeet Pradhan, Asim Biswas, and Peter M. Atkinson. The shadow effect on surface biophysical variables derived from remote sensing: A review. *Land*, 11(11):2025, 2022. doi: 10.3390/land11112025. URL https://doi.org/10.3390/land11112025.

Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017. doi: 10.1109/TPAMI.2016.2644615.

R.W. Coleman et al. A simplified framework for high-resolution urban vegetation classification: Reclassifying shadowed pixels with median-corrected sentinel-2 imagery. *Remote Sensing*, 12(15): 2399, 2020. doi: 10.3390/rs12152399. URL https://doi.org/10.3390/rs12152399.

Kyle Doherty, Max Gurinas, Brandon Trabucco, and Ruslan Salakhutdinov. Drone-lsr - light stable representations dataset, 2025. URL https://huggingface.co/datasets/mpg-ranch/drone-lsr. Aerial orthomosaic tiles with DINOv2 and DINOv3 embeddings for light-stable representation vision encoder training.

Hannes Feilhauer, Ben Somers, and Sebastian van der Linden. Optical trait indicators for remote sensing of plant species composition: Predictive power and seasonal variability. *Ecological Indicators*, 73:825–833, 2017. doi: 10.1016/j.ecolind.2016.11.003. URL https://doi.org/10.1016/j.ecolind.2016.11.003.

Olivier Hagolle, Mireille Huc, Daniel V. Pascual, and Gerard Dedieu. A multi-temporal method for cloud detection, applied to formosat-2, venμs, landsat and sentinel-2 images. *Remote Sensing of Environment*, 114(8):1747–1755, 2010. doi: 10.1016/j.rse.2010.03.002. URL https://doi.org/10.1016/j.rse.2010.03.002.

Dorothy K. Hall, George A. Riggs, and Vincent V. Salomonson. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, 54(2):127–140, 1995. ISSN 0034-4257. doi: https://doi.org/10.1016/0034-4257(95)00137-P. URL https://www.sciencedirect.com/science/article/pii/003442579500137P.

Nancy L. Harris, Elizabeth Goldman, Sarah Gibbes, Peter Potapov, Xinyi Li, Heeyeon Koo, ..., and Yasuo Uryu. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change*, 11:234–240, 2021. doi: 10.1038/s41558-020-00976-6. URL https://doi.org/10.1038/s41558-020-00976-6.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. URL https://arxiv.org/abs/1606.08415.

C. Hung, M. Bryson, and S. Sukkarieh. Vision-based shadow-aided tree crown detection and classification algorithm using imagery from an unmanned airborne vehicle. In *ISPRS / ISRSE-34*, 2011. URL https://www.isprs.org/proceedings/2011/isrse-34/211104015final00292.pdf. UAV imagery, use of shadow projection via solar angle + pose.

C. O. Justice, E. Vermote, J. R. G. Townshend, R. Defries, D. P. Roy, D. K. Hall, V. V. Salomonson, J. L. Privette, G. Riggs, A. Strahler, W. Lucht, R. B. Myneni, Y. Knyazikhin, S. W. Running, R. R. Nemani, Z. Wan, A. R. Huete, W. van Leeuwen, R. E. Wolfe, L. Giglio, J. P. Muller, P. Lewis, and M. J. Barnsley. The moderate resolution imaging spectroradiometer (modis): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4): 1228–1249, 1998. doi: 10.1109/36.701075. URL https://doi.org/10.1109/36.701075.

Fan Li, Zhiyi Wang, and Guoliang He. Ap shadow net: A remote sensing shadow removal network based on atmospheric transport and poisson's equation. *Entropy*, 24(9):1301, 2022. doi: 10.3390/e24091301. URL https://doi.org/10.3390/e24091301.

Xiaoxia Liu, Fengbao Yang, Hong Wei, and Min Gao. Shadow removal from uav images based on color and texture equalization compensation of local homogeneous regions. *Remote Sensing*, 14 (11):2616, 2022. doi: 10.3390/rs14112616. URL https://doi.org/10.3390/rs14112616.

S. K. McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996. doi: 10.1080/01431169608948714. URL https://doi.org/10.1080/01431169608948714.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. doi: 10.48550/arXiv.2304.07193. URL https://doi.org/10.48550/arXiv.2304.07193.

J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian. A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2):119–126, 1994. doi: 10.1016/0034-4257(94)90134-1. URL https://doi.org/10.1016/0034-4257(94)90134-1.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.

J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering. Monitoring vegetation systems in the great plains with erts. In *Third Earth Resources Technology Satellite-1 Symposium*, volume 1, pages 309–317. NASA SP-351, NASA Goddard Space Flight Center, Greenbelt, MD, 1973.

Jean-Romain Roussel, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R.H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian de Boissieu, and Alexis Achim. lidr: An r package for analysis of airborne laser scanning (als) data. *Remote Sensing of Environment*, 251:112061, 2020. doi: 10.1016/j.rse.2020.112061. URL https://doi.org/10.1016/j.rse.2020.112061.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Chenglin Shao, Huifang Li, Huanfeng Shen, et al. Generative shadow synthesis and removal for remote sensing images through embedding illumination models. *IEEE Transactions on Geoscience and Remote Sensing*, 2025. doi: 10.1109/TGRS.2025.3561307. URL https://doi.org/10.1109/TGRS.2025.3561307.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3, 2025. URL https://arxiv.org/abs/2508.10104.

L. J. Sugarbaker, E. W. Constance, H. K. Heidemann, A. L. Jason, V. Lukas, D. L. Saghy, and J. M. Stoker. The 3d elevation program initiative—a call for action. U.S. Geological Survey Circular 1399, U.S. Geological Survey, 2014. URL https://doi.org/10.3133/cir1399.

Jamie Tolan, Hung-I Yang, Benjamin Nosarzewski, Guillaume Couairon, Huy V. Vo, John Brandt, Justine Spore, Sayantan Majumdar, Daniel Haziza, Janaki Vamaraju, Théo Moutakanni, Piotr Bojanowski, Tracy Johns, Brian White, Tobias Tiecke, and Camille Couprie. Very high resolution canopy height maps from rgb imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300:113888, 2024. doi: 10.1016/j.rse.2023.113888.

Yuxin Wu and Kaiming He. Group normalization. In *European Conference on Computer Vision (ECCV)*, pages 3–19. Springer, 2018. doi: 10.1007/978-3-030-01261-8_1.

Zhe Zhu and Curtis E. Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012. doi: 10.1016/j.rse.2011.10.028. URL https://doi.org/10.1016/j.rse.2011.10.028.

## 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: We discuss and demonstrate the sensitivity of image encoders to varying lighting conditions. Objectives and results in the abstract reflect those in the main body of the paper.

   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: We discuss the limitations of this work starting here 3.

   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: We do not present 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 describe our experimental approach here 2.3, and publicy serve our dataset and code to reproduce analyses here 4.

   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: All data and code to reproduce our analyses are housed in a GitHub repository linked here 4 in the paper.

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: We provide training details here 2.4.

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: Yes, we conducted five-fold cross validation, and report 95% confidence intervals described here 2.4

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: We provide details of compute resources here 2.4

   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: We adhered to the code of ethics in our work.

   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: [NA]

    Justification: We do not expect our work to have societal impacts beyond advancing basic and applied remote sensing research.

    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: Our work poses no such risk.

    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: We cite all external models and data sources, further respecting their licenses as employed in our work. The external models and data sources are Oquab et al. [2023], Siméoni et al. [2025], and Sugarbaker et al. [2014], with Apache 2.0, custom license, and public domain licensure, respectively (described in text).

    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: Yes we describe pre-processing of the data and provide a currently anonymized repo with dataset card illustration usage, linked in text here 4.

    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: No research was crowdsourced.

    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 work does not involve crowdsourcing.

    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: Human researchers developed the core methods in this work.

    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.