14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

ENHANCING COASTAL RESILIENCE: RAPID POST-HURRICANE DEBRIS DETECTION

KOOSHAN AMINI[1], YUHAO LIU[2], JAMIE ELLEN PADGETT[1,3], GUHA BALAKRISHNAN[2,3] AND ASHOK VEERARAGHAVAN[2,3]

[1] Department of Civil and Environmental Engineering, Rice University
Houston, TX, USA

[2] Department of Electrical and Computer Engineering, Rice University
Houston, TX, USA

[3] Ken Kennedy Institute, Rice University
Houston, TX, USA
Email (corresponding): Kooshan.Amini@rice.edu

Key words: Debris Detection, Remote Sensing, Community Resilience, Image Segmentation, Hurricane-Induced Debris

Abstract. Timely and accurate detection of hurricane-induced debris is critical for effective disaster response and enhanced community resilience. While post-disaster aerial imagery is widely available, robust debris segmentation methods applicable across diverse disaster regions remain limited. Developing a generalized solution is challenging due to variations in environmental and imaging conditions that alter the visual signatures of debris, further complicated by the scarcity of high-quality labeled training data. This study addresses these challenges by fine-tuning pre-trained foundational vision models, achieving robust performance despite using a relatively small yet high-quality dataset. Specifically, this work introduces an open-source dataset comprising approximately 1,200 manually annotated aerial RGB images collected after Hurricanes Ian, Ida, and Ike. To mitigate human biases and enhance labeling quality, annotations from multiple experts are strategically aggregated, complemented by visual prompt engineering. The resulting fine-tuned model, named *fCLIPSeg*, achieves a Dice score of 0.70 on imagery from Hurricane Ida—a disaster event completely excluded from training—with virtually no false positives in debris-free regions. This research provides the first event-agnostic debris segmentation model deployable with standard RGB imagery, making it particularly well-suited for rapid, large-scale post-disaster impact assessments and recovery planning.

1 INTRODUCTION

Hurricanes and coastal storms inflict widespread damage, challenging community resilience and incurring significant costs; US tropical cyclones alone have caused over $1.5 trillion in damages since 1980 [1]. A critical consequence of these events is the massive accumulation of debris. For instance, Hurricane Katrina generated over 100 million cubic yards of debris [2], and debris removal often constitutes a major portion of recovery costs, approximately 36% [3]. Furthermore, this accumulated debris significantly impedes emergency response and subsequent recovery efforts by obstructing critical

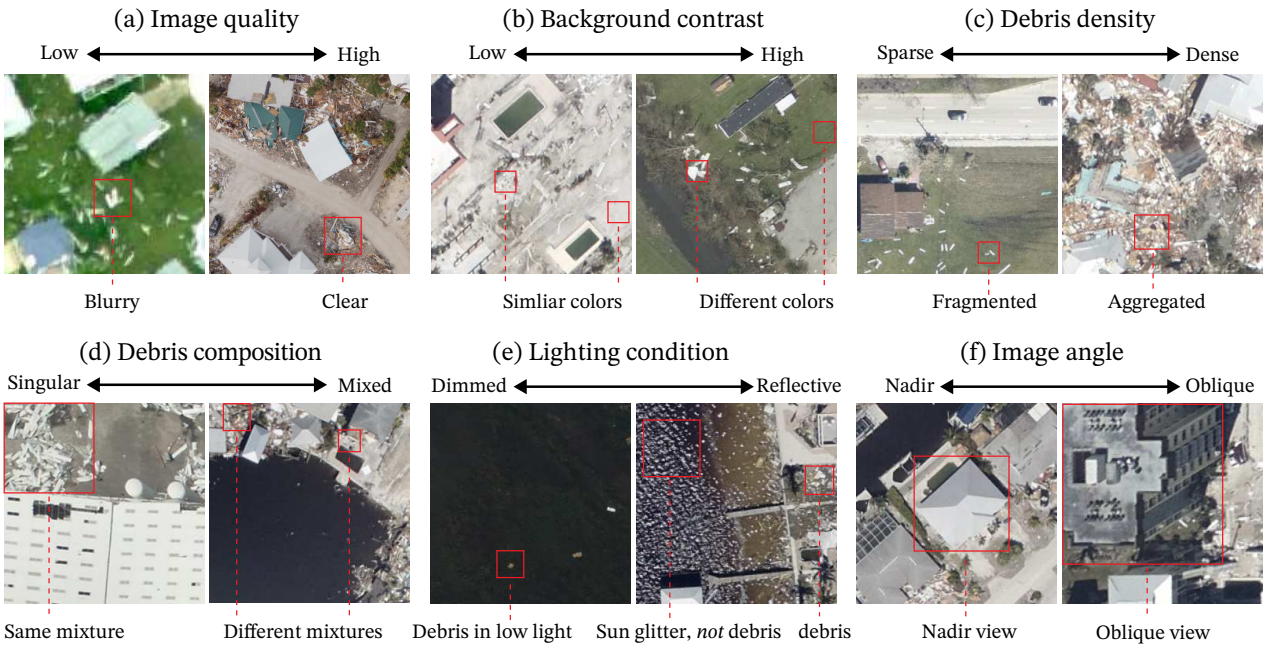

Figure 1: **Challenges in accurate debris segmentation across diverse conditions.** This figure highlights factors altering debris visual signatures (e.g., image quality, lighting, density) and the difficulty in distinguishing fragmented debris from similar objects.

routes [4]. Therefore, rapid post-storm debris detection and quantification are essential for prioritizing clearance operations, optimizing resource allocation, and mitigating secondary losses [5].

High-resolution aerial imagery, often rapidly provided by agencies like NOAA [6], is a crucial resource for post-disaster assessment [7]. However, manual analysis is too slow for timely decision-making, necessitating automated image segmentation techniques. While AI-assisted workflows show potential for debris volume estimation [8], a robust, generalizable debris segmentation model operating efficiently on standard aerial imagery is urgently needed for effective recovery planning.

Accurate debris detection is challenging due to variations in visual characteristics influenced by image quality, background, debris density/composition, lighting, and view angle (Fig. 1). These factors vary significantly across events. Furthermore, debris is often fragmented, occupies small image areas, and lacks context, making it difficult to distinguish from visually similar objects (e.g., water reflections, pavement).

Previous debris detection studies often focused on specific types like *marine* or *vegetative* debris, frequently relying on multi-spectral or hyper-spectral data and spectral signatures using methods like SVMs [9–12]. This study, however, targets hurricane-induced non-vegetative debris on land using standard RGB imagery. Existing methods for this often use single-pixel spectral classification [13] or hand-crafted textural filters [14]. These approaches struggle with the heterogeneity of urban debris, lack sufficient spectral contrast in non-vegetative backgrounds, are sensitive to varying conditions, and have often been validated only on single regions or events [11, 13, 14], limiting their generalizability. Consequently, no robust, event-agnostic debris segmentation model using standard RGB imagery currently exists.

Addressing these challenges likely requires complex models, but training them from scratch is hampered by the scarcity of large, high-quality labeled debris datasets [15], despite the availability of raw post-disaster imagery [16]. This motivates leveraging pre-trained *foundational models* [17–

19]. These models, trained on vast datasets, offer robust, generalizable representations. Among relevant architectures [18–23], CLIPSeg [24] emerged as a promising candidate. Its CLIP encoder [21] captures rich semantics, but its performance can degrade for specialized concepts like debris, likely under-represented in its original training.

Therefore, we propose fine-tuning the pre-trained CLIPSeg model for post-hurricane debris segmentation using only standard aerial RGB imagery, enhancing transferability by avoiding specialized sensors. This leverages the foundational model's power while requiring a relatively small fine-tuning dataset. We curated a dataset of 1,200 images from Hurricanes Ian, Ike, and Ida, balancing debris-present and debris-free examples. To improve data quality, we aggregated labels from multiple annotators into consensus annotations and employed visual prompt engineering. Images from Hurricane Ida were entirely withheld during training to evaluate domain generalization.

Our evaluations demonstrate that this fine-tuned model, *fCLIPSeg*, significantly improves debris segmentation accuracy and robustness across varying environmental and imaging factors, including effective performance on the unseen Hurricane Ida test set. This work presents the first event-agnostic debris segmentation model applicable across multiple disaster regions using only RGB imagery.

Ultimately, our goal is to enable near-real-time debris detection supporting rapid response and long-term resilience planning using publicly available data (e.g., NOAA). This framework offers immediate operational benefits and lays groundwork for future predictive modeling and multi-scale analysis. By open-sourcing the model and dataset [1], we aim to foster collaboration and enhance global disaster preparedness and recovery.

Figure 2: **Three hurricane-affected regions used in this research.** Key details for each storm and location are provided, alongside visualizations highlighting impacted areas and image counts. The diversity across regions is crucial for training and evaluation.

2 DEBRIS SEGMENTATION DATASET

This section details the construction of our dataset for fine-tuning CLIPSeg, involving data collection, classification, and annotation. We sampled 50×50 m ground resolution crops from publicly

[1] Source code and dataset will be made available upon publication.

available aerial RGB images. This size balances compatibility with CLIPSeg's encoder [24] and practicality for manual annotation. These procedures are designed to be region-agnostic.

2.1 Data Collection

Aiming for a generalizable framework, we release our model and dataset openly via the NHERI DesignSafe Cyberinfrastructure [25]. We selected imagery sources based on prompt post-event collection, sufficient image quality, and public availability. After evaluation, the *NOAA Emergency Response Imagery* dataset [16] was chosen. Operational since 2003, it provides timely, high-quality aerial photos post-hurricane, enabling rapid assessments. While NOAA is our primary source, our methodology is applicable to other aerial imagery sources (Sec. 5.1).

2.1.1 Case studies

To ensure diversity and test generalizability, we sourced images from three major hurricanes spanning different years, regions, and imaging instruments: *Ian* (2022), *Ida* (2021), and *Ike* (2008) (Fig. 2). Specifically, we trained using images from Hurricanes Ian (Florida, 15-30 cm resolution, severe impact) and Ike (Texas, 50 cm resolution, older instruments), reserving Hurricane Ida images (Louisiana, 15-30 cm resolution) exclusively for the test set. This design rigorously evaluates the model's ability to generalize to unseen events and locations. While Fig. 2 highlights key impacted areas, our dataset includes surrounding regions for completeness, providing diverse testbeds for assessing variations in resolution, sensors, and debris conditions.

2.2 Debris Classification

To prevent model bias from training only on debris-laden areas, we included negative samples (debris-free crops from various environments like coasts, forests, unaffected built zones). A multi-annotator classification filtered images into *positive* (containing debris) and *negative* samples. Including challenging negative samples (e.g., visually similar water reflections, rooftops) facilitates contrastive learning [26]. Positive samples proceeded to segmentation labeling (Sec. 2.3).

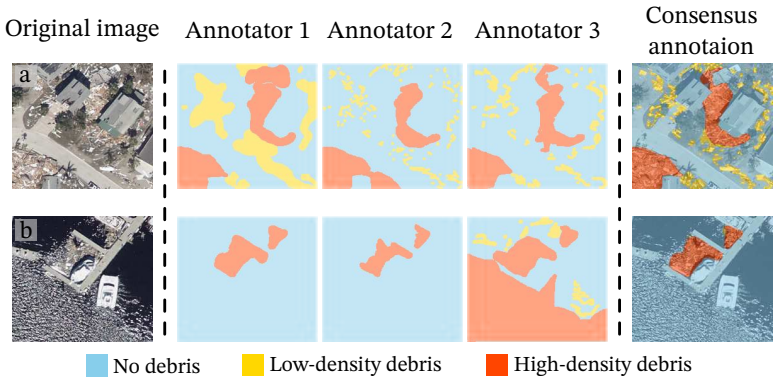

Figure 3: **Debris segmentation challenges and annotation aggregation.** Input from multiple annotators is aggregated into a consensus annotation to mitigate individual errors and biases, yielding more reliable ground truth labels.

2.3 Debris Segmentation Annotation

For an RGB image Q, annotators produced dense segmentations $S_a \in \{0,1,2\}^{H\times W}$, labeling pixels as 0 (no debris), 1 (low-density: scattered fragments), or 2 (high-density: aggregated piles). This distinction aids response planning and leverages CLIP's pattern recognition, though classes can be merged if density information is unneeded. To mitigate individual annotator bias (illustrated in Fig. 3), annotations from N annotators $\{S_{a_1}, ..., S_{a_N}\}$ for each image were aggregated into a *consensus annotation* $\bar{S}$ via per-pixel averaging:

$$\bar{S}(:,:) = \lceil \frac{1}{N} \sum_{a=1}^{N} S_N(:,:,a) \rceil, \tag{1}$$

where $\lceil \cdot \rceil$ is the ceiling function. This $\bar{S}$ served as the ground truth for fine-tuning, proving more reliable than single annotations (Fig. 3).

3 DEBRIS SEGMENTATION MODEL

This section details our fine-tuning of the CLIPSeg model for improved debris segmentation performance using the labeled dataset from Sec. 2. We first outline the base CLIPSeg segmentation process, then describe our visual prompt engineering and fine-tuning procedures.

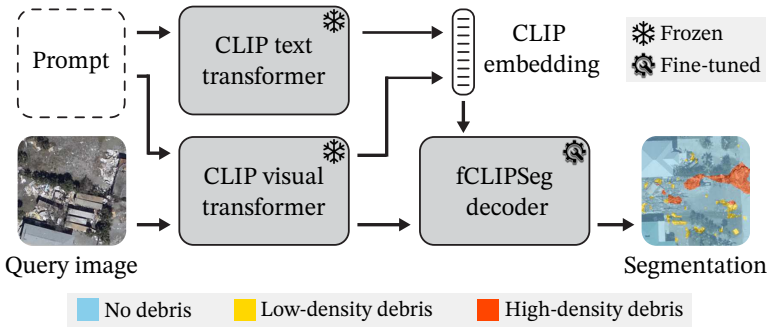

Figure 4: **Overview of fCLIPSeg model.** Based on CLIPSeg, it uses pre-trained CLIP encoders and a fine-tuned decoder for debris segmentation from RGB images.

3.1 Image Segmentation Using CLIPSeg

Given an RGB query image Q, the model produces a dense segmentation map S, where each pixel is labeled as 0 (no debris), 1 (low-density debris), or 2 (high-density debris) based on corresponding text prompts (see Sec. 2.3).

Our model follows the CLIPSeg architecture [24], shown in Fig. 4. It employs frozen pre-trained CLIP ViT-B/16 visual and text encoders [21] to generate 512-dimensional embeddings $\mathbf{e}$ for image inputs and text prompts. A transformer-based decoder then processes these embeddings along with intermediate activations extracted from the visual encoder to produce segmentation maps. Specifically, activations from selected visual encoder layers are projected and combined with the decoder's internal states. A final linear projection yields single-channel segmentation maps S_l for each density level l, where pixel values indicate the match score to the input prompt.

To obtain the final multi-class segmentation S, we generate S_l for all density levels $l \in \{0, 1, 2\}$ and assign the label with the highest score at each pixel (x, y):

$$S(x, y) = \arg\max_{l} S_l(x, y). \tag{2}$$

While applicable to debris, the standard CLIPSeg struggles, likely due to debris being under-represented in its original training data. This motivates fine-tuning on our specialized debris dataset (Sec. 2).

3.2 Visual Prompt Engineering

To improve training efficiency with limited data, we use *visual prompt engineering* [27], modifying input RGB images to highlight target debris features. This improves alignment between visual and text prompts, boosting segmentation performance [24]. We create engineered prompts by degrading the background (non-target pixels defined by ground truth labels) through brightness reduction and Gaussian blur (Fig. 5). This is applied separately for low-density ($\mathcal{P}_1$) and high-density ($\mathcal{P}_2$) debris prompts. Prompts for the 'no debris' class ($\mathcal{P}_0$) use the original, non-engineered negative images from Sec. 2.2. These engineered prompts are used during fine-tuning.

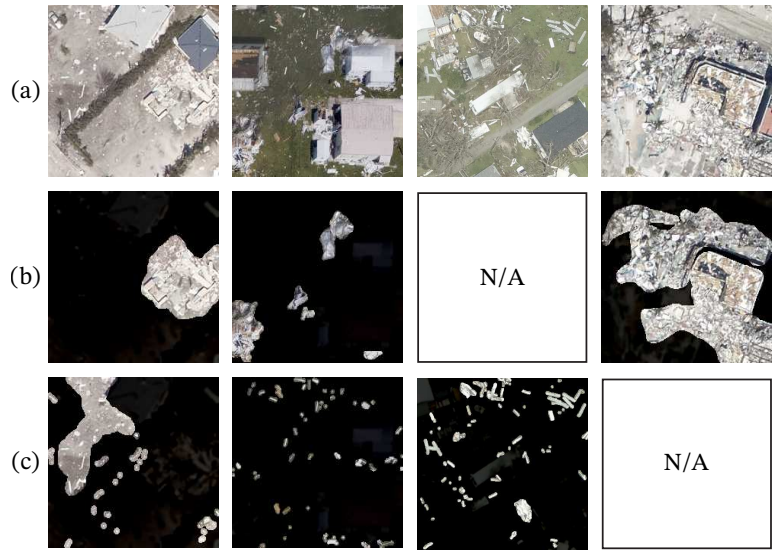

Figure 5: **Visual prompt engineering.** Salient debris is highlighted by blurring and darkening the background, improving training efficiency. Examples show (a) original RGB, and engineered prompts for (b) high-density and (c) low-density debris.

3.3 Fine-tuning CLIPSeg

We fine-tune the CLIPSeg transformer decoder (Fig. 4), initializing from the official checkpoint [24] while keeping the CLIP encoders frozen. Fine-tuning uses our labeled dataset where each image Q has corresponding engineered visual prompts $\mathcal{P}$ and consensus annotation $\bar{S}$.

During training, for each image Q, we randomly sample a density level l, its text prompt (e.g., "debris at low-density"), and a corresponding engineered visual prompt from $\mathcal{P}_l$. The ground truth is the binary segmentation for level l derived from $\bar{S}$. We generate CLIP embeddings for the text prompt ($\mathbf{e}_\text{t}$) and visual prompt ($\mathbf{e}_\text{v}$). As a data augmentation strategy, we interpolate these embeddings:

$$\mathbf{e}_\text{c} = \alpha \mathbf{e}_\text{t} + (1 - \alpha)\mathbf{e}_\text{v}, \tag{3}$$

with α uniformly sampled from $[0, 1]$. This combined embedding $\mathbf{e}_c$ conditions the decoder. We use binary cross-entropy loss between the model's output for density l and the binary ground truth. The resulting fine-tuned model is named *fCLIPSeg*.

4 RESULTS

This section presents the experimental setup (Sec. 4.1), quantitative evaluation (Sec. 4.2), and qualitative assessment (Sec. 4.3) of our proposed fCLIPSeg model.

4.1 Experimental Setup

We used PyTorch Lightning, fine-tuning CLIPSeg [24] (checkpoint *rd64-uni-refined*) for 2 000 epochs (batch 64). Optimization employed AdamW [28] with a cosine-decayed learning rate (0.001 $\rightarrow$ 0.0001) and mixed precision. Training finished in 5 h on one Nvidia A40, and the best model was chosen by validation Dice (debris classes).

Inference throughput on an Nvidia RTX 3090 is 9.6 images/sec (effective batch 3) while using only 1.2 GB VRAM. Segmenting the entire Estero Island, FL ($\approx$10 km^2, 4 016 images) took 2 min 19 s on an Intel® Core™ i9-13900K CPU and an Nvidia RTX 4090 (24 GB).

4.2 Quantitative Evaluation

We evaluated performance against consensus annotations (Sec. 2.3) on the test set (Hurricane Ida images, unseen during training), split into debris-positive (125 images) and debris-free (148 images) subsets (Sec. 2.2). Key metrics include Precision (accuracy of positive predictions), Recall (completeness of positive predictions), Dice Coefficient (harmonic mean of precision/recall), and Intersection over Union (IoU, overlap ratio).

Table 1: **Segmentation performance on the Hurricane Ida test set** (273 images unseen during training). ↑: Higher is better.

Subset	Metric	CLIPSeg	fCLIPSeg (Ours)
Debris-positive	Dice ↑	0.30	**0.70**
	IoU ↑	0.27	**0.65**
	Recall [low-density] ↑	0.00	**0.33**
	Recall [high-density] ↑	**0.92**	0.82
	Precision [low-density] ↑	**0.86**	0.60
	Precision [high-density] ↑	0.10	**0.87**
Debris-free	Dice ↑	0.74	**0.99**
	IoU ↑	0.74	**0.99**
	Recall [no debris] ↑	0.93	**1.00**

Results (Table 1) compare our fCLIPSeg against the baseline CLIPSeg [24]. Baseline CLIPSeg shows some capacity but is highly biased: insensitive to low-density debris (low recall) and inaccurate for high-density debris (low precision). Fine-tuning significantly boosts performance: fCLIPSeg achieves much higher Dice (0.70 vs 0.30) and IoU (0.65 vs 0.27) on the debris-positive set, showing improved detection of low-density debris (better recall) and more accurate identification of high-density debris (much higher precision). On the debris-free set, fCLIPSeg approaches perfect Dice/IoU (0.99), indicating very few false positives, whereas baseline CLIPSeg frequently errs. These results confirm the value of fine-tuning for improved accuracy and reduced bias.

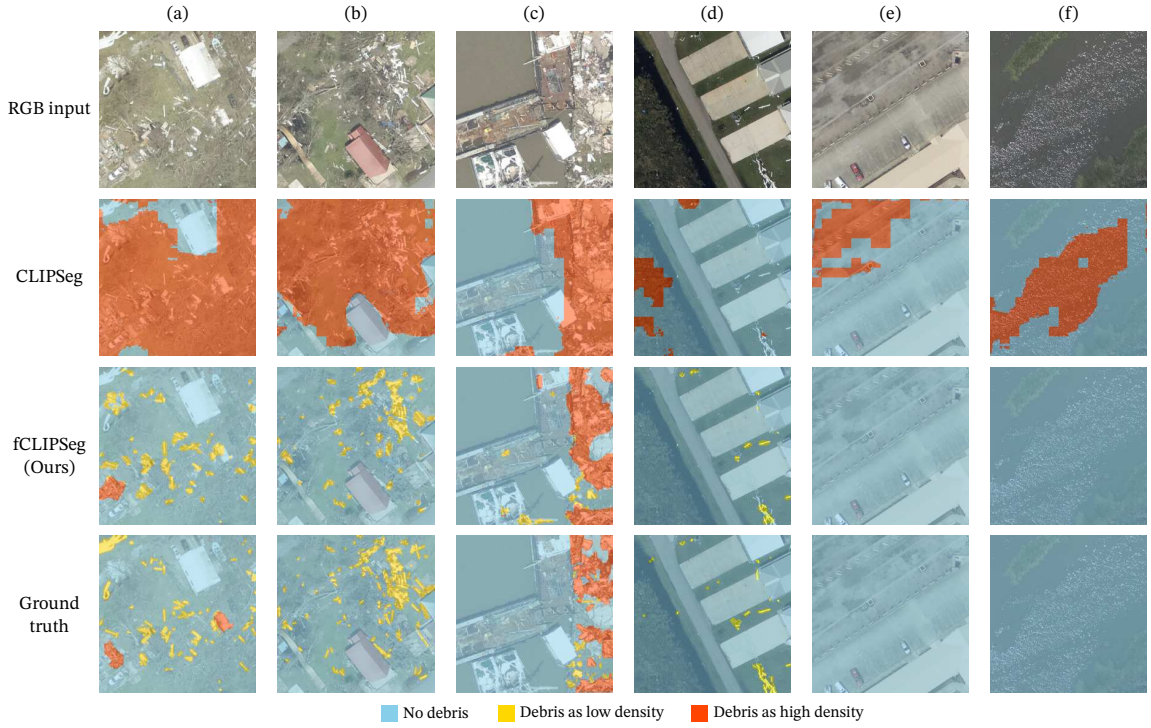

Figure 6: **Visual comparison:** Six examples from the Hurricane Ida test set ((a)–(d) debris-positive, (e)–(f) debris-free). fCLIPSeg accurately segments debris unlike the baseline CLIPSeg.

4.3 Qualitative Evaluation

Visual comparison (Fig. 6) confirms the quantitative findings. Baseline CLIPSeg often mislabels non-debris objects (vegetation, markings, reflections) and produces coarse segmentations, reflecting its low precision for high-density and low recall for low-density debris. In contrast, fCLIPSeg generally provides accurate segmentation across various scenes. It correctly identifies most low-density debris in (a) and (b) and handles heavily damaged areas well (c). It appears to rely on high-level features rather than simple brightness or shape cues, correctly classifying structured objects (c) and complex non-debris textures like water reflections (f). Its high accuracy on non-debris objects aligns with the high recall score for the debris-free subset (Table 1).

fCLIPSeg still makes minor errors, such as occasionally missing small debris fragments or assigning density levels inconsistently with ground truth (examples from the test set shown in Fig. 6(d) and Fig. 6(a), respectively), though their downstream impact is likely small; see Sec. 5.1 for discussions on fCLIPSeg's ability to generalize across a variety of conditions.

5 DISCUSSION AND CONCLUSION

This section synthesizes fCLIPSeg's debris segmentation performance, acknowledges limitations, and outlines future research directions.

5.1 Generalizability and Transferability

Evaluations in Sec. 4 used a test set from Hurricane Ida regions entirely unseen during training or validation, demonstrating model generalizability. Fig. 7 further illustrates fCLIPSeg's robustness

across varied conditions: severe impact with sandy backgrounds (Ian, Fig. 7(a)), less damage with more vegetation (Ida, Fig. 7(b)), poor resolution and color calibration from older instruments (Ike, Fig. 7(c)), and high-resolution drone imagery unseen during training (Fig. 7(d)). Despite challenges like low contrast or domain shift, fCLIPSeg provided accurate segmentation, suggesting resilience to factors like debris density, background, altitude, and sensor type.

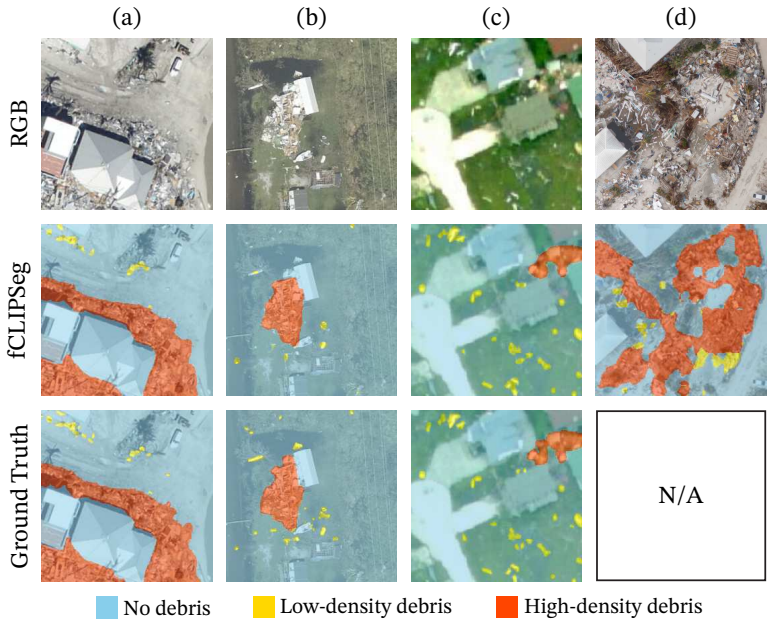

Figure 7: **Segmentation generalizability demonstration.** fCLIPSeg produces accurate results across diverse conditions from Hurricanes Ian, Ida, and Ike, despite variations in impact, resolution, and imaging instruments.

5.2 Limitation and Challenges

Despite promising results, fCLIPSeg has limitations. It may miss very small or occluded debris fragments and occasionally misclassify visually similar textures (e.g., vegetation). While relatively rare, these errors indicate potential for refinement, especially for ambiguous visual cues. Another challenge is the diversity of post-hurricane conditions globally. While trained on several US hurricanes, incorporating data from more diverse geographic regions (unique construction, environments, climates) could enhance generalizability and robustness.

5.3 Future Opportunities

Building on this work, several opportunities exist. First, developing *predictive* models to estimate debris fields pre-landfall, potentially using our model's outputs as ground truth for lower-resolution forecasting, could enhance proactive planning. Second, integrating 3D data (LiDAR, photogrammetry) could enable volumetric debris estimation, crucial for logistics and impact assessment. Lastly, adapting the approach for lower-resolution satellite imagery could facilitate rapid, global-scale debris mapping, using our high-resolution results as reference points to create a more equitable, widely accessible tool.

5.4 Conclusion

This paper introduced *fCLIPSeg*, a fine-tuned CLIPSeg model for robust, generalizable post-hurricane debris segmentation using standard aerial RGB imagery. Incorporating multi-annotator consensus, visual prompt engineering, and contrastive learning yielded substantial performance improvements over baseline CLIPSeg. The model adapts effectively to diverse debris densities, resolutions, and environments.

Evaluations confirmed significant gains in precision and recall, validating the fine-tuning approach. While occasional errors remain, motivating further work on generalizability (e.g., broader datasets), fCLIPSeg provides a strong foundation. This work can enable near-real-time analysis and serves as a stepping stone for predictive modeling, 3D volume estimation, and global debris mapping. The public release of our model, code, and dataset aims to foster collaboration and enhance disaster preparedness and recovery.

Acknowledgments

Kooshan Amini and Jamie E. Padgett were partially supported by NSF Award 2429680 and by the Ken Kennedy Institute's inaugural Research Clusters Initiative. Additionally, the contributions of Yuhao Liu and Ashok Veeraraghavan were supported by NSF Award 2107313. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors. The authors sincerely thank Narges Saeednejad, Andres Calvo, Jainish Patel, and Benjamin Barria Martinez for their valuable contributions as annotators.

References

[1] NOAA National Centers for Environmental Information. U.s. billion-dollar weather and climate disasters. Accessed: 2025-03-01, https://www.ncei.noaa.gov/access/billions, 2025. DOI: 10.25921/stkw-7w73.

[2] U.S. Government Accountability Office (GAO). Hurricane katrina: Continuing debris removal and disposal issues. Technical Report GAO-08-985R, U.S. Government Accountability Office, 2008.

[3] Department of Homeland Security Office of Inspector General. Observations of fema's debris monitoring efforts for hurricane irma. OIG-18-85 Report, Accessed: 2025-03-01, 2018. URL: https://www.oig.dhs.gov/sites/default/files/assets/Mga/2018/oig-18-85-sep18.pdf.

[4] Kooshan Amini and Jamie E. Padgett. Probabilistic risk assessment of hurricane-induced debris impacts on coastal transportation infrastructure. *Reliability Engineering & System Safety*, 240: 109579, December 2023. ISSN 09518320. doi: 10.1016/j.ress.2023.109579. URL https://linkinghub.elsevier.com/retrieve/pii/S0951832023004933.

[5] Navid Nickdoost, Hiba Jalloul, and Juyeong Choi. An integrated framework for temporary disaster debris management sites selection and debris collection logistics planning using geographic information systems and agent-based modeling. *International Journal of Disaster Risk Reduction*, 80:103215, October 2022. ISSN 22124209. doi: 10.1016/j.ijdrr.2022.103215. URL https://linkinghub.elsevier.com/retrieve/pii/S2212420922004344.

[6] NOAA. National geodetic survey damage assessment imagery (hurricane michael). Accessed: 2025-03-01, https://oceanservice.noaa.gov/news/oct18/michael-storm-imagery.html, 2018.

[7] Chih-Shen Cheng, Amir H. Behzadan, and Arash Noshadravan. Deep learning for post-hurricane aerial damage assessment of buildings. *Computer-Aided Civil and Infrastructure Engineering*, 36(6):695–710, 2021. doi: https://doi.org/10.1111/mice.12658. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12658.

[8] Chih-Shen Cheng, Linchao Luo, Sean Murphy, Yu-Chen Lee, and Fernanda Leite. A framework to enhance disaster debris estimation with ai and aerial photogrammetry. *International Journal of Disaster Risk Reduction*, 107:104468, 2024. doi: 10.1016/j.ijdrr.2024.104468. URL https://www.sciencedirect.com/science/article/pii/S2212420924002309.

[9] Chuanmin Hu. Remote detection of marine debris using satellite observations in the visible and near infrared spectral range: Challenges and potentials. *Remote Sensing of Environment*, 259: 112414, June 2021. ISSN 0034-4257. doi: 10.1016/j.rse.2021.112414.

[10] Ali Jamali and Masoud Mahdianpari. A Cloud-Based Framework for Large-Scale Monitoring of Ocean Plastics Using Multi-Spectral Satellite Imagery and Generative Adversarial Network. *Water*, 13(18):2553, September 2021. ISSN 2073-4441. doi: 10.3390/w13182553.

[11] Alican Karaer, Mehmet Baran Ulak, Tarek Abichou, Reza Arghandeh, and Eren Erman Ozguven. Post-Hurricane Vegetative Debris Assessment Using Spectral Indices Derived from Satellite Imagery. *Transportation Research Record: Journal of the Transportation Research Board*, 2675(12):504–523, December 2021. ISSN 0361-1981, 2169-4052. doi: 10.1177/03611981211029921.

[12] Jiyeon Kim, Sorin C. Popescu, Roel R. Lopez, X. Ben Wu, and Nova J. Silvy. Assessing hurricane impact on vegetation and endangered deer habitat using airborne lidar and multispectral images. *Global Ecology and Conservation*, 53:e03007, September 2024. ISSN 23519894. doi: 10.1016/j.gecco.2024.e03007.

[13] Leanne Hauptman, Diana Mitsova, and Tiffany Roberts Briggs. Hurricane Ian Damage Assessment Using Aerial Imagery and LiDAR: A Case Study of Estero Island, Florida. *Journal of Marine Science and Engineering*, 12(4):668, April 2024. ISSN 2077-1312. doi: 10.3390/jmse12040668.

[14] Shasha Jiang and Carol J. Friedland. Automatic urban debris zone extraction from post-hurricane very high-resolution satellite and aerial imagery. *Geomatics, Natural Hazards and Risk*, 7(3):933–952, May 2016. ISSN 1947-5705, 1947-5713. doi: 10.1080/19475705.2014.1003417.

[15] Zeyad Emam, Andrew Kondrich, Sasha Harrison, Felix Lau, Yushi Wang, Aerin Kim, and Elliot Branson. On the state of data in computer vision: Human annotations remain indispensable for developing deep learning models, 2021. URL https://arxiv.org/abs/2108.00114.

[16] National Geodetic Survey. National geodetic survey - emergency response imagery index, 2025. URL https://storms.ngs.noaa.gov/. Accessed: 2025-01-17.

[17] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting Diffusion Prior for Real-World Image Super-Resolution, November 2023.

[18] Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Xianzheng Ma, Hao Dong, Peng Gao, and Hongsheng Li. Personalize Segment Anything Model with One Shot, October 2023.

[19] Lucas Prado Osco, Qiusheng Wu, Eduardo Lopes de Lemos, Wesley Nunes Gonçalves, Ana Paula Marques Ramos, Jonathan Li, and José Marcato. The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*, 124:103540, November 2023. ISSN 1569-8432. doi: 10.1016/j.jag.2023.103540.

[20] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[21] 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, February 2021.

[22] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection, March 2023.

[23] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, Zhaoyang Zeng, Hao Zhang, Feng Li, Jie Yang, Hongyang Li, Qing Jiang, and Lei Zhang. Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks, January 2024.

[24] Timo Lüddecke and Alexander S. Ecker. Image Segmentation Using Text and Image Prompts, March 2022.

[25] Ellen M. Rathje, Clint Dawson, Jamie E. Padgett, Jean-Paul Pinelli, Dan Stanzione, Ashley Adair, Pedro Arduino, Scott J. Brandenberg, Tim Cockerill, Charlie Dey, Maria Esteva, Fred L. Haan, Matthew Hanlon, Ahsan Kareem, Laura Lowes, Stephen Mock, and Gilberto Mosqueda. Designsafe: New cyberinfrastructure for natural hazards engineering. *Natural Hazards Review*, 18(3):06017001, 2017. doi: 10.1061/(ASCE)NH.1527-6996.0000246.

[26] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[27] Aleksandar Shtedritski, Christian Rupprecht, and Andrea Vedaldi. What does clip know about a red circle? visual prompt engineering for vlms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11987–11997, 2023.

[28] I Loshchilov. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.