A Framework for Detecting and Quantifying Hallucinations in Remote Sensing Image Inpainting

Konstantinos Zafeirakis
Faculty of Science
University of Amsterdam
Amsterdam, The Netherlands
konstantinos.zafeirakis@student.uva.nl

Grigorios Tsagkatakis
Institute of Computer Science
Foundation for Research and Technology - Hellas (FORTH)
Heraklion, Greece
greg@ics.forth.gr

***Abstract*—This work addresses the critical issue of hallucinations in deep learning-based image inpainting within remote sensing applications. Remote sensing images are often degraded due to sensor malfunctions or adverse atmospheric conditions. As such, they require inpainting to restore missing information accurately. This restoration is vital for enabling downstream tasks such as classification, detection, and segmentation. Despite the advancements, deep learning models for inpainting face multiple challenges including hallucinations, where the model incorrectly introduces non-existent elements in the image. This study introduces a novel framework for detecting hallucinations using an image inpainting generator coupled with a two-class discriminator and a class activation mapping (Grad-CAM) model. The experimental setup involves diverse masking techniques and analyzes the inpainting results across different image classes. Our findings reveal significant impacts of mask type and size on hallucination metrics, with rectangular masks generally yielding better results than irregular and random masks. Additionally, each class-specific generator exhibited unique inpainting behaviors, influenced by mask size. The study identifies the in-distribution Dice metric and out-of-distribution prediction value as effective measures for hallucination detection, with the FID metric proving optimal for reconstruction quality.**

***Index Terms*—Image Inpainting, Hallucination Detection, Remote Sensing, Generative Adversarial Networks (GANs), Explainable AI (XAI), Grad-CAM.**

I. Introduction

Remote sensing images are invaluable assets for environmental monitoring, civil engineering [1], disaster management, and diverse applications like crop yield modeling [2] and mineral exploration [3]. However, their quality is often compromised by sensor malfunctions, atmospheric conditions like cloud cover, or data transmission errors, resulting in missing or corrupted regions [4]. Image inpainting, the task of filling in these missing regions, is a critical preprocessing step to ensure data completeness for subsequent analysis [5].

In remote sensing, deep learning-based inpainting methods are the current gold standard [6], preferred for their ability to handle the high resolution and geographical variations that render conventional methods less effective [7]. These models excel at producing plausible textures and structures, but their generative nature introduces a significant risk: hallucinations. Hallucination occurs when a model generates content that is plausible in appearance but factually incorrect or inconsistent with the surrounding ground truth. In remote sensing, this could manifest as fabricating a building in a field or inpainting a river with forest texture, thereby corrupting the dataset for downstream applications.

Despite the critical importance of data integrity in remote sensing, there is a lack of systematic methods to detect, localize, and quantify such hallucinations. Most evaluations of inpainting models focus on pixel-wise reconstruction error (e.g., MSE) or perceptual similarity (e.g., LPIPS, FID), which do not explicitly measure factual correctness. This paper addresses this gap by proposing a dedicated framework for hallucination detection, conceptually illustrated in Fig. 1.

Our main contributions are:

- A novel framework combining an inpainting generator, a two-class discriminator, and an explainability model (Grad-CAM) to systematically detect and quantify hallucinations.
- A thorough investigation into how mask type (rectangular, random, irregular) and size influence the emergence and severity of hallucinations in remote sensing imagery.
- An extensive experimental evaluation that validates our proposed metrics—the out-of-distribution (OOD) prediction score and an in-distribution (ID) Dice score—as effective indicators of hallucinatory content.

The remainder of this paper is organized as follows: Section II reviews related work. Section III details our proposed framework. Section IV describes the experimental setup, and Section V presents and analyzes the results. Finally, Section VI concludes the paper.

II. Related Work

A. Image Inpainting

Image inpainting has evolved significantly from traditional methods to modern deep learning-driven approaches. Early techniques included diffusion-based methods that propagated information from boundaries [8] and exemplar-based synthesis that copied patches from known regions of the image [9]. While effective for small gaps, these methods often struggle to generate semantically plausible content for larger missing areas.

The current state-of-the-art is dominated by deep learning, particularly Generative Adversarial Networks (GANs) [10],

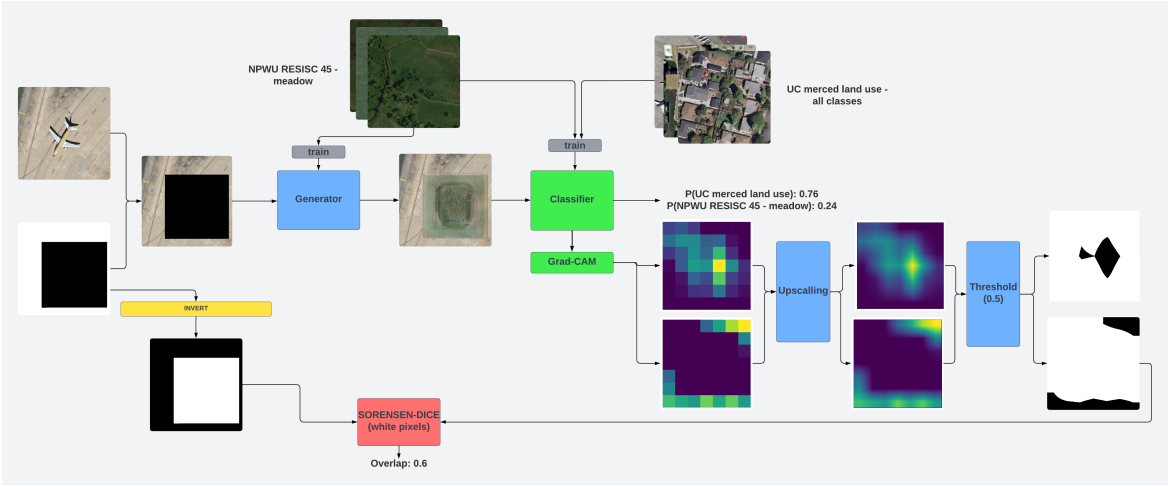


Fig. 1. Overview of the Proposed Hallucination Detection Framework. An out-of-distribution ground truth image (e.g., 'airport') is masked and fed to an inpainting generator trained on an in-distribution class (e.g., 'meadow'). The completed image is passed to a two-class discriminator, which outputs a prediction score (hallucination likelihood) and a Grad-CAM heatmap highlighting the regions responsible for the OOD classification.

[11], which learn the underlying distribution of natural images to generate highly realistic completions. More recently, transformer-based models like the Mask-Aware Transformer (MAT) [12] have pushed performance further, leveraging self-attention mechanisms to capture long-range dependencies and excel at inpainting large, irregular holes. Our work builds upon this progress by utilizing the MAT architecture as the core inpainting generator to study its failure modes.

B. Hallucination and Out-of-Distribution Detection

The concept of "hallucination," where a model generates plausible but factually incorrect content, is a known risk in generative models. This issue has been explored in various domains. For instance, in inverse problems involving galaxy image reconstruction, [13] identified regions prone to hallucinations by analyzing the Fisher information matrix. In medical imaging, studies have measured hallucinations in reconstructed tomographic images by comparing them against a ground truth, using metrics like MSE and SSIM to quantify the error [14]. These works establish the importance of detecting such artifacts but often rely on having a complete ground truth for comparison.

Our work is therefore closely related to Out-of-Distribution (OOD) detection, which aims to identify inputs that differ from a model's training distribution without a direct ground truth comparison. Recent methods have successfully used generative models for OOD detection. For example, [15] proposed using diffusion models to map a corrupted image back to its in-domain manifold, using reconstruction error (LPIPS) as an OOD score. Similarly, [16] used the multi-level reconstruction error from Denoising Diffusion Probabilistic Models (DDPMs) to identify OOD inputs. Our framework adapts this core concept: instead of detecting a foreign image, we aim to detect a foreign patch within an image. We achieve this by training a dedicated discriminator to explicitly distinguish between in-distribution (ID) and OOD content generated by the inpainter, and we use explainability methods to localize it.

III. Methodology

Our proposed framework, illustrated in Fig. 1, is designed to systematically induce, detect, and quantify hallucinations. It comprises an image inpainting generator, a hallucination discriminator, and a class activation-based detection module.

A. Image Inpainting Generator

The core of our system is a generator, G, responsible for filling masked image regions. We employ the Mask-Aware Transformer (MAT) [12], a state-of-the-art transformer-based model for large-hole inpainting. Given a ground truth image I and a binary mask M (where 0 denotes missing pixels), the generator takes the masked image $I_M = I \odot M$ as input and produces a reconstructed image I_G. The MAT architecture consists of a convolutional head, a transformer body, a convolutional tail, a style manipulation module, and a Conv-U-Net for refinement.

The generator is trained on a single class of images, which we define as the in-distribution (ID) class. The training objective is to minimize a composite loss function that balances realism and perceptual quality:

$$L = L_G + \gamma R_1 + \lambda L_p \tag{1}$$

where L_G is the non-saturating adversarial loss, $R_1 = \mathbb{E}_x[\|\nabla D(x)\|]$ is an R1 gradient penalty for stabilizing training, and $L_p = \sum_i \eta_i \|\phi_i(x) - \phi_i(\hat{x})\|_1$ is the perceptual loss calculated from layer activations ϕ_i of a pre-trained VGG-19 network. Following [12], we set $\gamma = 10$ and $\lambda = 0.1$.

B. Hallucination Discriminator

To detect hallucinations, we use a two-class classifier, C, which acts as a discriminator. Its task is to determine if an inpainted image I_G belongs to the generator's ID training class

(class 0) or an out-of-distribution (OOD) class (class 1). We build the classifier via transfer learning using a MobileNetV2 [17] model pre-trained on ImageNet. The original classification head is replaced by a Global Average Pooling (GAP) 2D layer, a Dropout layer ($p = 0.5$), and a final Dense layer with two outputs.

The classifier is trained on two sets: images from the generator's ID class and images from all other classes in the dataset, which form the OOD set. This setup trains the classifier to be an expert at recognizing the generator's expected output style versus anything else.

C. Classifier Output-Based Detection

The primary hallucination metric is the Out-of-Distribution Prediction (OODP) value, derived from the classifier's logits. The softmax function converts the classifier's raw logit outputs for the ID class (z_0) and OOD class (z_1) into probabilities. The OODP is the probability assigned to the OOD class:

$$\text{OODP} = \frac{e^{z_1}}{e^{z_0} + e^{z_1}} \tag{2}$$

A high OODP value indicates high confidence from the classifier that the inpainted image contains features inconsistent with the generator's training data, thus signaling a hallucination.

D. Class Activation-Based Detection

To localize hallucinations and derive a spatial metric, we use Gradient-weighted Class Activation Mapping (Grad-CAM) [18]. Grad-CAM leverages the gradients flowing into the final convolutional layer of the classifier to produce a heatmap highlighting the image regions most influential for a given class prediction.

The neuron importance weights α_k^c for a class c and feature map k are computed by global average pooling the gradients of the class score y^c with respect to the feature map activations A^k:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{3}$$

The final Grad-CAM heatmap is a weighted combination of the feature maps, passed through a ReLU function to isolate positive contributions:

$$L_{Grad-CAM}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k \right) \tag{4}$$

We generate a heatmap for the in-distribution (ID) class, which highlights regions the classifier recognizes as belonging to the generator's training domain. This heatmap is upscaled and binarized with a threshold of 0.5 to create a mask, ID_{THmap}. Our spatial metric measures the overlap between this identified ID region and the actual unmasked region of the image, using the Sørensen-Dice coefficient.

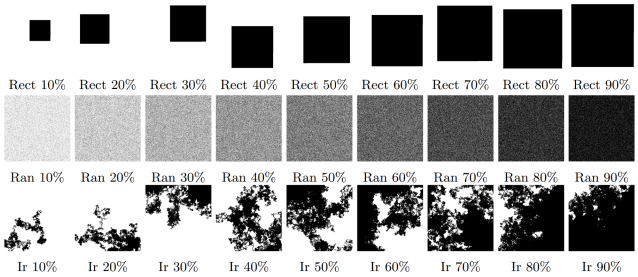


Fig. 2. Examples of the three mask types used in our experiments: Rectangular, Random, and Irregular, shown at increasing coverage percentages.

IV. Experimental Setup

A. Datasets and Implementation

We use the **NWPU-RESISC45** dataset [19] for our primary experiments, with additional testing on the **UC Merced Land Use** dataset [20]. For each experiment, a MAT generator was trained on 300 images (512x512 pixels) from a single class (e.g., 'meadow', 'dense residential'). Generator training was monitored and stopped when the Fréchet Inception Distance (FID) score, evaluated on a validation set, converged. The MobileNetV2 discriminator was trained using images resized to 224x224, with the convolutional base frozen. All models were implemented using PyTorch.

B. Masking Strategies

To simulate varying data loss scenarios, we used the three mask types namely: (i) a single, randomly placed rectangular patch **Rectangular**, (ii) Randomly sampled individual pixels **Random**, (iii) A contiguous mask grown from a random seed **Irregular**, illustrated in Fig. 2: For each type, we varied the masked area from 10% to 90% of the total image area in 10% increments.

C. Evaluation Metrics

In addition to our proposed hallucination metrics, we evaluate the generator's reconstruction quality using three standard perceptual metrics:

- **Mean Squared Error (MSE)**: Calculated between grayscale versions of the ground truth and the inpainted images.
- **Fréchet Inception Distance (FID)** [21]: Measures perceptual quality by comparing feature distributions.
- **Learned Perceptual Image Patch Similarity (LPIPS)** [22]: Quantifies perceptual similarity using deep features.

V. Results and Analysis

We conducted a series of experiments to validate our framework. We first establish the baseline performance of our inpainting generator in an ideal, in-distribution (ID) scenario. We then analyze hallucination detection across scenarios of varying difficulty, defined by the visual similarity between the generator's training class and the out-of-distribution (OOD) target image class.

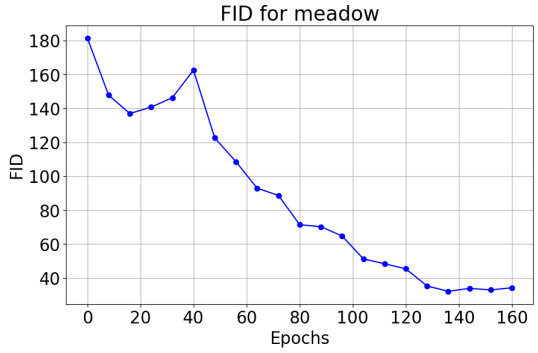
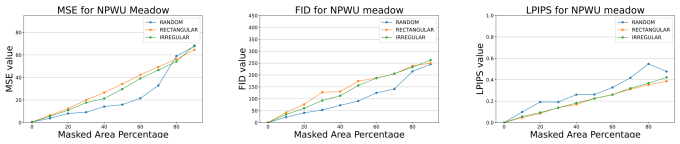


Fig. 3. FID score of the 'meadow' generator during training. The score converges, indicating successful model training.



(a) MSE metrics (b) FID metrics (c) LPIPS metrics

Fig. 4. Baseline reconstruction metrics for the 'meadow' generator on in-distribution 'meadow' images. Error increases with mask size.

A. Baseline Inpainting Performance

Before evaluating hallucinations, it is essential to confirm that our inpainting generator is well-trained and capable of high-quality reconstruction under normal conditions.

Training Convergence: Fig. 3 shows the FID score of the 'meadow' generator during training. The score steadily decreases and converges after approximately 160 epochs, indicating that the model has learned a stable representation of the target class.

Reconstruction Quality: Fig. 4 presents the reconstruction quality for the 'meadow' generator on in-distribution 'meadow' test images. As expected, all error metrics degrade as the masked area increases. Random masks consistently achieve the best reconstruction quality, as they preserve distributed contextual guidance. Conversely, rectangular masks perform the worst, as they create a large, context-free void. Fig. 5 visually confirms this: the random mask inpainting results in a slightly blurry but structurally coherent meadow, while the rectangular mask struggles to recreate fine details. These results confirm the generator's competence and establish a robust performance benchmark.

B. Hallucination Detection: High Dissimilarity Scenario

Having established the generator's baseline, we begin our hallucination analysis with the most straightforward case: applying the 'meadow'-trained generator to the highly dissimilar 'airport' class.

Quantitative Analysis: Fig. 6 shows the performance of our proposed metrics. The OOD Prediction Score (Fig. 6a) for rectangular masks increases sharply with the masked area, exceeding 40% at high coverage. This provides a strong, clear signal that the discriminator is confidently identifying

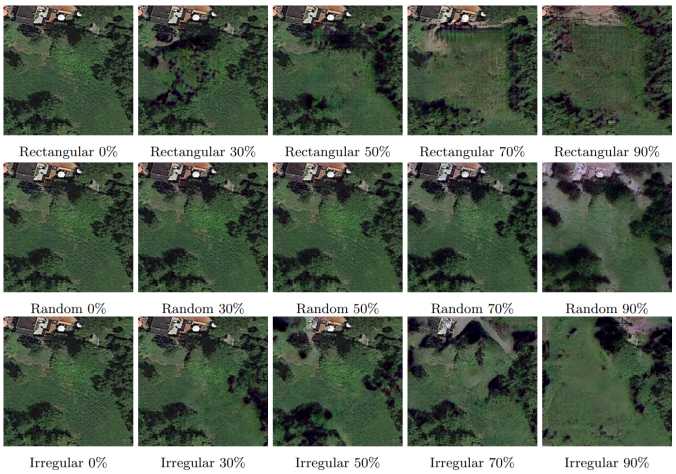
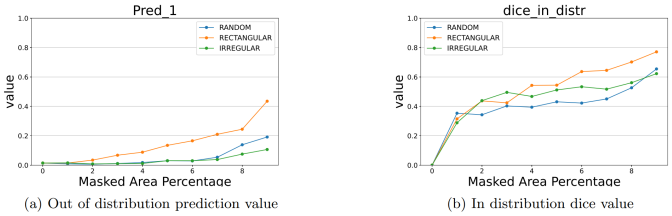


Fig. 5. Qualitative baseline results for 'meadow' inpainting. The inpainted images show the generator's performance under ideal, in-distribution conditions with different mask types and sizes.

the inpainted content as out-of-distribution. The ID Dice Score (Fig. 6b) shows a complementary and stable increase, confirming that the discriminator correctly separates the hallucinated region from the known ground truth context. In contrast, random and irregular masks produce a much weaker and less consistent signal for both metrics, suggesting their resulting artifacts are more subtle and harder for the classifier to distinguish from the noisy background.



(a) Out of distribution prediction value (b) In distribution dice value

Fig. 6. Quantitative hallucination metrics for the 'meadow' generator on 'airport' images. Rectangular masks (orange) yield a strong detection signal.

Qualitative Analysis: The visual results in Fig. 7 confirm the quantitative findings. The inpainted images clearly show the generator filling the masked areas with amorphous green 'meadow' texture, which is visually jarring against the structured airport background. The hallucination is most blatant in the case of large rectangular masks, corresponding to the highest OODP scores.

C. Hallucination Detection: Moderate and Low Dissimilarity

To test the framework's sensitivity, we analyze its performance on moderately dissimilar ('river') and highly similar ('forest') classes.

Quantitative Analysis: Fig. 8 shows a more nuanced response for the 'river' class. The OODP score for rectangular masks remains low until a "tipping point" around 60% mask coverage. At this point, the defining river feature is fully obscured, and the inpainted meadow begins to dominate,

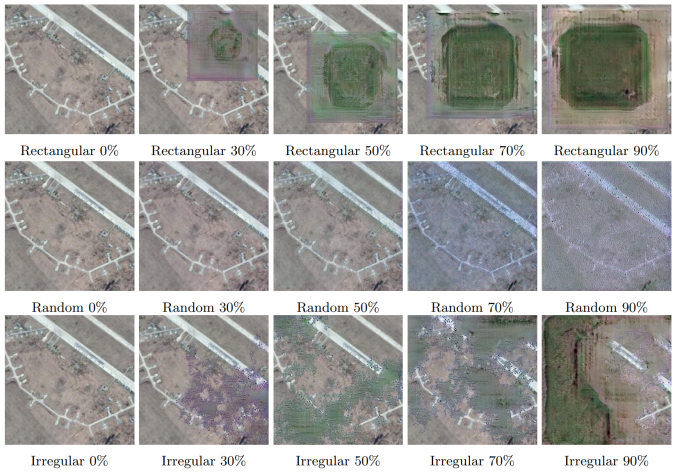
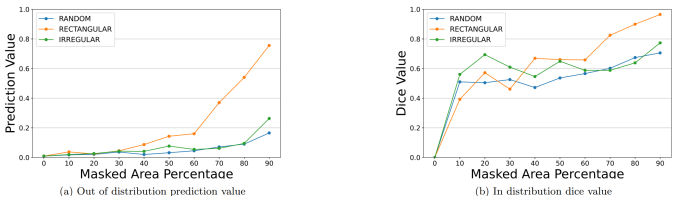


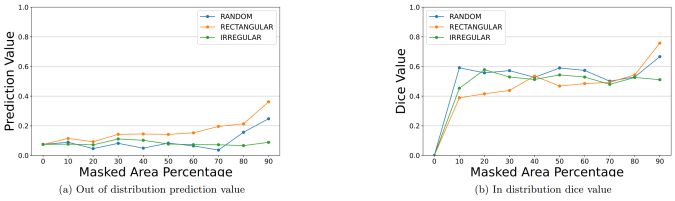
Fig. 7. Qualitative results for 'meadow' generator on 'airport' images. The generator fills the masked areas with out-of-context 'meadow' texture.

causing the OODP score to rise sharply. In the 'forest' scenario (Fig. 9), the OODP score remains low across all mask types and sizes. The discriminator struggles to confidently distinguish the hallucinated 'meadow' texture from the authentic 'forest' context due to their high visual similarity. This highlights an expected limitation: the framework is less effective when the hallucination is semantically very close to the ground truth.



(a) Out of distribution prediction value (b) In distribution dice value

Fig. 8. Quantitative metrics for 'meadow' generator on 'river' images. The OODP score for rectangular masks (orange) shows a distinct "tipping point."



(a) Out of distribution prediction value (b) In distribution dice value

Fig. 9. Quantitative metrics for 'meadow' generator on 'forest' images. The OODP scores are low, indicating the difficulty of detecting near-distribution hallucinations.

D. Analysis of Generator-Specific Behavior

To demonstrate that hallucinations are predictable, generator-specific artifacts, we analyze an alternative scenario using a 'dense residential' generator on an 'airplane' image. The qualitative progression is shown in Fig. 10. At a small mask size, the generator attempts a plausible reconstruction. As the mask grows, it begins to hallucinate building-like textures. By 90% coverage, it fabricates a full high-altitude cityscape, which is accurately localized by our framework. This behavior is entirely different from the 'meadow' generator's output, confirming that hallucinations are structured, class-conditional artifacts, not random noise.

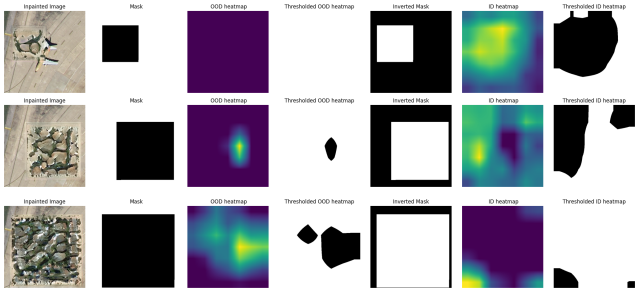


Fig. 10. Qualitative progression of hallucination from a 'dense residential' generator on an 'airplane' image. Hallucination severity and detection accuracy increase with mask size.

VI. Discussion

Our results provide several key insights into the nature of hallucinations in remote sensing inpainting and the effectiveness of our detection framework.

A. Implications of Findings

The strong correlation between hallucination severity and the semantic distance between the generator's training class and the target content is a critical finding. It implies that the risk of inpainting-induced data corruption is not uniform; it is highest when the missing region contains content that is fundamentally different from what the generator has been trained on. This highlights a practical vulnerability: a general-purpose inpainting model trained on a diverse dataset may still fail spectacularly when faced with a rare or highly specific land cover class. Our framework provides a methodology for quantifying this risk.

Furthermore, the predictable, class-conditional nature of hallucinations suggests they are a systematic failure mode, not random noise. The 'dense residential' generator's tendency to produce high-altitude cityscapes reveals a strong learned prior. This predictability is a double-edged sword: while it makes hallucinations detectable by a targeted framework like ours, it also means that the generated artifacts can be deceptively structured and realistic, potentially misleading automated analysis or human interpreters if left unchecked.

B. The Critical Role of Mask Geometry

Our finding that rectangular masks induce the most severe and detectable hallucinations has direct real-world implications. Large, contiguous data loss, such as that caused by cloud cover or sensor striping (dead detector arrays), represents the highest-risk scenario for generating factually incorrect content. In contrast, sparse data loss, like salt-and-pepper noise, is less likely to cause severe semantic hallucinations, though it may result in textural degradation like blurring. Practitioners using

inpainting models should therefore be most cautious when repairing large, continuous gaps in their imagery.

C. Limitations of the Framework

While effective, our framework has several limitations that open avenues for future research. First, its performance degrades when the OOD class is visually very similar to the ID class, as seen in the 'meadow' vs. 'forest' experiment. The two-class discriminator struggles to draw a clear boundary in these near-distribution cases. Second, the current approach requires training a dedicated discriminator for each inpainting generator, which can be computationally intensive. A more universal detection method would be desirable. Finally, our study was limited to a single generator architecture (MAT); the hallucination patterns of other architectures, such as purely CNN-based or diffusion models, may differ.

VII. Conclusion

In this work, we introduced a novel framework for the detection, localization, and quantification of hallucinations in remote sensing image inpainting. By systematically coupling a class-specific inpainting generator with a two-class discriminator and an explainability model (Grad-CAM), we demonstrated that it is possible to reliably identify and measure factually incorrect content generated by the model. Our experiments revealed that hallucination severity is strongly dependent on both the semantic distance between the source and target content and the geometry of the missing region, with large, contiguous masks posing the highest risk. We successfully validated our proposed metrics the OOD Prediction Score and the ID Dice Score—as effective tools for this task, laying the groundwork for building more reliable and trustworthy inpainting systems for critical remote sensing applications.

VIII. Future Work

Building on the insights and limitations identified in this study, future research will pursue several promising directions:

1) **Advanced Discriminator Architectures:** We will investigate more advanced models for the discriminator, particularly one-class classifiers (OCC) and anomaly detection techniques, which may offer improved sensitivity to subtle, near-distribution hallucinations and eliminate the need to define an explicit OOD set.
2) **Universal Hallucination Detectors:** A key goal is to develop a more universal detector that does not require retraining for each new generator. This could involve training a model on the features of real vs. generated patches, independent of the generator that created them.
3) **Analysis of Different Generator Architectures:** We plan to apply our framework to other state-of-the-art inpainting models, including diffusion-based and CNN-based architectures, to compare their characteristic hallucination patterns.
4) **Realistic Corruption Scenarios:** To enhance practical relevance, we will extend our testing to masks that more closely mimic real-world data corruption, such as realistic cloud shapes and sensor-specific artifacts like line noise or striping.

References

[1] D. F. Laefer, "Harnessing remote sensing for civil engineering: Then, now, and tomorrow," in *Applications of Geomatics in Civil Engineering*, J. K. Ghosh and I. da Silva, Eds. Singapore: Springer Singapore, 2020, pp. 3–30.

[2] D. A. Kasampalis, T. K. Alexandridis, C. Deva, A. Challinor, D. Moshou, and G. Zalidis, "Contribution of remote sensing on crop models: A review," *Journal of Imaging*, vol. 4, no. 4, 2018. [Online]. Available: https://www.mdpi.com/2313-433X/4/4/52

[3] H. Shirmard, E. Farahbakhsh, R. D. Müller, and R. Chandra, "A review of machine learning in processing remote sensing data for mineral exploration," *Remote Sensing of Environment*, vol. 268, p. 112750, 2022.

[4] W. Huang, Y. Deng, S. Hui, and J. Wang, "Image inpainting with bilateral convolution," *Remote Sensing*, vol. 14, no. 23, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/23/6140

[5] Z. Qin, Z. Wang, C. Chen, Z. Wang, and Y.-G. Wang, "Image inpainting based on deep learning: A review," *Displays*, vol. 69, p. 102028, 2021.

[6] A. Pondaven, M. Bakler, D. Guo, H. Hashim, M. Ignatov, and H. Zhu, "Convolutional neural processes for inpainting satellite images," *arXiv preprint arXiv:2205.12407*, 2022.

[7] A. Kumar, D. Tamboli, S. Pande, and B. Banerjee, "Rsinet: Inpainting remotely sensed images using triple gan framework," *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 143–146, 2022.

[8] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '00. USA: ACM Press/Addison-Wesley Publishing Co., 2000, p. 417–424. [Online]. Available: https://doi.org/10.1145/344779.344972

[9] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000. [Online]. Available: https://api.semanticscholar.org/CorpusID:308278

[10] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2536–2544.

[11] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, jul 2017. [Online]. Available: https://doi.org/10.1145/3072959.3073659

[12] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, "Mat: Mask-aware transformer for large hole image inpainting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[13] M. L. Sampson and P. M. Melchior, "Spotting Hallucinations in Inverse Problems with Data-Driven Priors," in *Machine Learning for Astrophysics*, Jul. 2023, p. 29.

[14] S. Bhadra, V. Kelkar, F. Brooks, and M. Anastasio, "On hallucinations in tomographic image reconstruction," *IEEE transactions on medical imaging*, vol. PP, 05 2021.

[15] Z. Liu, J. P. Zhou, Y. Wang, and K. Q. Weinberger, "Unsupervised out-of-distribution detection with diffusion inpainting," in *Proceedings of the 40th International Conference on Machine Learning*, ser. ICML'23. JMLR.org, 2023.

[16] M. S. Graham, W. H. L. Pinaya, P.-D. Tudosiu, P. Nachev, S. Ourselin, and M. J. Cardoso, "Denoising diffusion models for out-of-distribution detection," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 2948–2957.

[17] M. Z. A. Z. L.-C. C. Mark Sandler, Andrew Howard, "Mobilenetv2: Inverted Residuals and Linear Bottlenecks," 2019.

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, p. 336–359, Oct. 2019. [Online]. Available: http://dx.doi.org/10.1007/s11263-019-01228-7

[19] Y. Liu, Y. Zhong, S. Shi, and L. Zhang, "Scale-aware deep reinforcement learning for high resolution remote sensing imagery classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 209, pp. 296–311, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271624000224

[20] V. R. Jakkula, “Tutorial on support vector machine (svm),” 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:15115403
[21] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” 2018.
[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” 2018.

ACKNOWLEDGMENT