BEYOND 2D ARTIFACTS: GEOMETRIC CUES FOR AI-GENERATED IMAGE DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

AI-generated image detection remains challenging under unseen-generator shifts, in-the-wild settings, and repeated online dissemination. Existing image-domain cues can be generator-dependent and can degrade under resizing and recompression during online dissemination. We instead examine geometric coherence, motivated by the fact that real photographs arise from physical 3D scenes whereas AI-generated images need not satisfy the same physical constraints. To capture geometric coherence from a single image, we introduce GeoCue. GeoCue first considers point-normal consistency, which measures the agreement between normals implied by predicted 3D points and independently predicted surface normals. With a frozen geometry predictor, we observe that real photographs exhibit a characteristic range of point-normal consistency, whereas AI-generated images deviate from this range in generator-dependent ways. While point-normal consistency retains its discriminative signal under online dissemination, it can miss geometric changes that preserve point-normal agreement. GeoCue therefore further incorporates reference-relative depth and point deviations. GeoCue consistently outperforms prior methods across unseen-generator and in-the-wild benchmarks. Under repeated online dissemination involving resizing and JPEG/WebP re-encoding, GeoCue retains stable detection performance, while many existing detectors degrade substantially or exhibit unstable performance. These results demonstrate the robustness of GeoCue across generator shifts, in-the-wild data, and repeated online dissemination.

1 INTRODUCTION

As image generators continue to improve, visual realism alone is becoming an increasingly unreliable indicator of whether an image is real or AI-generated. A practical detector must remain reliable not only for previously unseen generators and diverse real-world imagery, but also after the repeated resizing and JPEG/WebP re-encoding that often occur during online dissemination. Many existing approaches base this decision on image-domain cues, including local artifacts, frequency-domain patterns, learned visual features, and reconstruction behavior. While these signals can be highly discriminative, they often reflect generator-specific traces or dataset biases and can be unstable under online dissemination. Resizing resamples and interpolates pixel values, changing local pixel relationships, while JPEG/WebP re-encoding suppresses high-frequency and fine-grained information through compression. These changes can weaken or shift the image-domain cues on which many detectors rely. These limitations motivate us to look beyond image-domain cues and examine evidence grounded in the physical process of image formation.

In a real photograph, a camera records an image formed by light interacting with a physical 3D scene. Surface geometry and orientation, material reflectance, and illumination jointly determine the light reaching the camera, producing shading, highlights, and shadows that are physically coupled to scene geometry. AI-generated images, by contrast, are produced by generative models with objectives primarily favoring perceptual plausibility and need not satisfy the same physical constraints that tie appearance to a coherent 3D scene. This difference in image formation motivates us to examine geometric evidence beyond conventional image-domain cues. As illustrated in Fig. 1, both real and AI-generated images can appear visually plausible in RGB, but their geometric predictions can exhibit different patterns.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

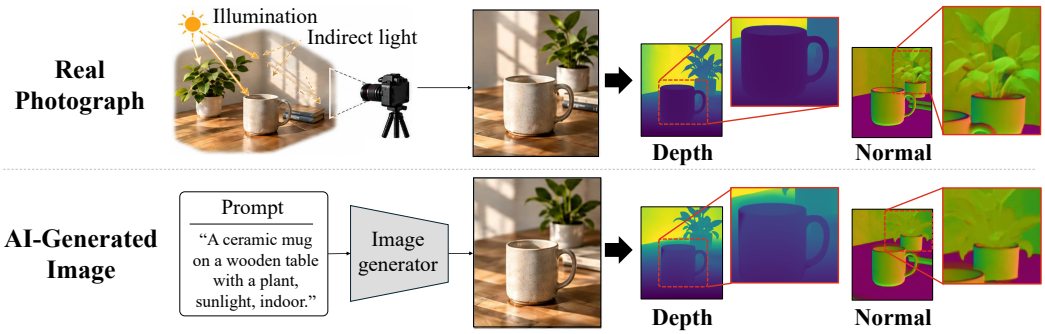


Figure 1: **Motivation for geometric cues beyond image-domain evidence.** Real photographs result from physical image formation in 3D scenes, whereas AI-generated images are synthesized by generative models. Consequently, despite similar RGB appearance, the two can exhibit different structures in their predicted geometry.

To characterize these differences within a single image, we examine the agreement between 3D points and surface normals, which provide complementary descriptions of the same local surface. Predicted 3D points imply a local surface orientation, which should agree with the independently predicted surface normal (Qi et al., 2018; Cao et al., 2021). We measure this agreement as point-normal consistency. As shown in Fig. 2a, a frozen geometry predictor reveals a characteristic range of point-normal consistency responses for real photographs, whereas AI-generated images depart from this range in generator-dependent ways. These response differences remain largely stable after repeated online dissemination, suggesting point-normal consistency as a useful detection cue.

Based on this finding, we introduce GeoCue, which captures complementary geometric cues for AI-generated image detection. Point-normal consistency measures the internal agreement between geometric predictions, but it can overlook cases where point and normal predictions shift together while remaining mutually consistent. We therefore introduce a complementary cue based on depth and point deviations relative to a frozen geometry predictor. GeoCue organizes these geometric responses using a margin-based criterion defined against the relatively stable real-image response region.

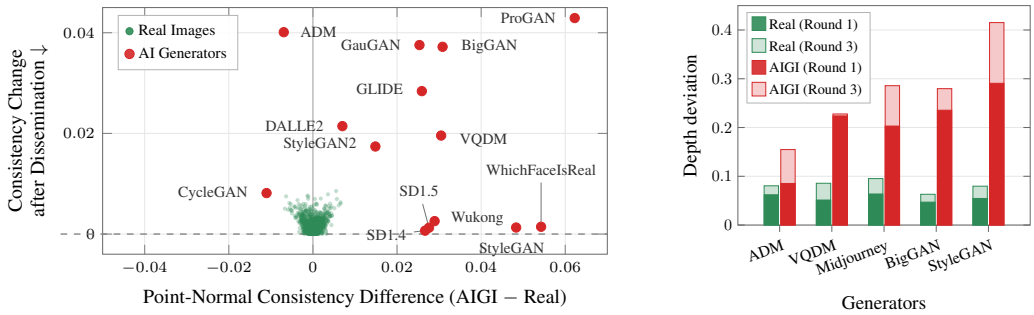
Across unseen-generator and in-the-wild settings, GeoCue achieves strong detection performance. To assess class-wise performance beyond AUC, we additionally report R.Acc. and F.Acc. GeoCue maintains balanced performance across real and AI-generated images. GeoCue retains a high mean AUC with low cross-benchmark variation after repeated online dissemination, while many baselines degrade substantially. These results demonstrate that geometric cues provide robust detection across generator shifts, real-world data, and repeated online dissemination.

The contributions of this work are as follows:

- **Geometric perspective beyond image-domain cues.** Motivated by differences in image formation between real and AI-generated images, we identify distinct geometric prediction patterns that provide detection evidence beyond conventional image-domain cues.
- **Complementary consistency and reference-relative deviation.** We introduce GeoCue, which combines point-normal consistency with reference-relative depth and point deviations to capture complementary geometric evidence.
- **Robust detection across generator shifts and online dissemination.** GeoCue achieves strong performance across unseen generators and in-the-wild data, while remaining robust under repeated online dissemination involving resizing and JPEG/WebP re-encoding.

2 Related Work

AI-Generated Image Detection. AI-generated image detection has evolved alongside advances in generative models. Early work focused on GAN images, exploiting low-level artifacts and frequency



(a) Generator-dependent departures.

(b) Stability of geometric deviation.

Figure 2: **Geometric responses under online dissemination.** All responses are measured with a frozen pretrained geometry predictor. (a) Point-normal consistency shows generator-dependent departures from the real-image response region that remain stable after five rounds of online dissemination. (b) Geometric deviations remain relatively stable for real images but increase more substantially for AI-generated images under dissemination.

fingerprints in generated outputs. CNN-Spot (Wang et al., 2020) showed that ProGAN-trained detectors can transfer to other CNN-based generators. Later methods have sought more transferable image-domain cues. FreqNet (Tan et al., 2024b) and DFFreq (Yan et al., 2026) exploit frequency information, while NPR (Tan et al., 2024a), LaDeDa (Cavia et al., 2024), and FerretNet (Liang et al., 2025) focus on local pixel or patch-level cues. SAFE (Li et al., 2025) further improves generalization by preserving artifact cues and encouraging local awareness.

With the rise of diffusion models and text-to-image models, detection methods increasingly turned to learned visual representations and reconstruction-based cues. Corvi et al. (2023) analyze forensic traces of diffusion-generated images and their robustness to common post-processing. UniversalFD (Ojha et al., 2023), ClipDet (Cozzolino et al., 2024), AIDE (Yan et al., 2025), and C2P-CLIP (Tan et al., 2025) leverage CLIP-based representations for detection, while DIRE (Wang et al., 2023) and AEROBLADE (Ricker et al., 2024) exploit reconstruction-based cues. More recently, DEAR (Kim et al., 2026) improves robustness by suppressing spurious detector features.

Beyond binary detection, recent work has explored source attribution, watermarking, explainability, and real-world evaluation. GAN Fingerprints (Yu et al., 2019) and DE-FAKE (Sha et al., 2023) address generator attribution, while Stable Signature (Fernandez et al., 2023) and Tree-Ring Watermarks (Wen et al., 2023) embed traceable signals during generation. AIGI-Holmes (Zhou et al., 2025) investigates human-verifiable explanations, and benchmarks such as Chameleon (Yan et al., 2025) and WildRF (Cavia et al., 2024) emphasize evaluation under challenging real-world conditions. Despite these broader directions, most binary detectors still rely on appearance- or representation-based evidence. Motivated by differences in real and AI image formation, GeoCue instead explores geometric predictions as detection evidence.

Physical and Geometric Cues for AI-Generated Image Detection. Moving beyond conventional image-domain cues, recent work has explored physically and geometrically motivated evidence for AI-generated image detection. Sarkar et al. (2024) exploit projective geometry through perspective fields, line geometry, and object-shadow relations. SurFake (Ciamarra et al., 2024) uses surface anomalies for face deepfake detection, while Defocus (Jeon & Woo, 2025) exploits defocus blur through estimated defocus maps. These approaches show that physical and geometric properties can provide useful detection evidence, but often target specific phenomena such as projective geometry, surface anomalies, or defocus. In contrast, GeoCue is motivated by the broader difference in how real and AI-generated images are formed, and uses predicted geometry as a general source of detection evidence across diverse scenes and generators.

Geometry Estimation Models. Recent geometry estimation models provide rich geometric predictions from single images or multiple views. MoGe and MoGe-2 (Wang et al., 2025b;c) recover dense point-based geometry from open-domain images, with MoGe-2 extending monocular geometry estimation to metric scale while preserving accurate relative geometry. Metric3D v2 (Hu et al.,

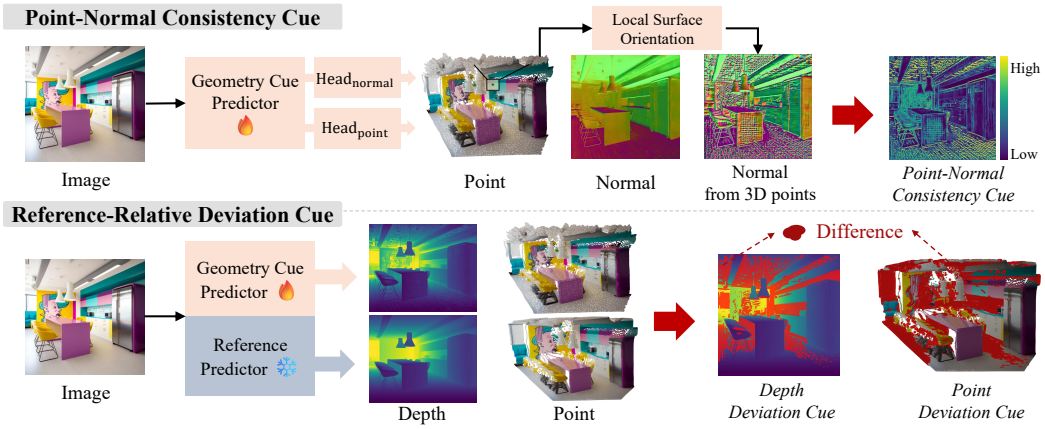


Figure 3: **Overview of GeoCue.** GeoCue combines point-normal consistency with reference-relative depth and point deviations. Point-normal consistency measures the agreement between directly predicted normals and normals derived from 3D points, while reference-relative deviation measures changes in depth and 3D-point predictions relative to a frozen geometry predictor. The resulting geometric cues are used for AI-generated image detection.

2024) and GeoWizard (Fu et al., 2024) estimate depth and surface normals, while Depth Anything V2 (Yang et al., 2024) focuses on monocular depth estimation. Dens3R (Fang et al., 2026) supports both single-view and multi-view geometric prediction, while DUSt3R (Wang et al., 2024) and VGGT (Wang et al., 2025a) address dense 3D reconstruction from image pairs or multiple views. These models are designed for geometry estimation and reconstruction rather than AI-generated image detection. GeoCue instead repurposes geometric predictions as detection evidence for AI-generated image detection.

3 GeoCue: Consistency and Reference-Relative Deviation

3.1 Overview

Motivated by the observed differences in geometric coherence, we examine the agreement between 3D points and surface normals. The local surface orientation implied by predicted 3D points should agree with the independently predicted surface normal. We define this agreement as point-normal consistency. As shown in Fig. 2a, even a frozen geometry predictor shows a characteristic response range for real images, whereas AI-generated images deviate from it in generator-dependent ways. These differences remain largely stable under repeated online dissemination. Given the relative stability of real-image responses, we define the point-normal margin based on their characteristic response range. This formulation associates real images with stronger point-normal agreement while requiring AI-generated responses to lie beyond the margin.

However, point-normal consistency captures the internal agreement among geometric predictions, and similar consistency responses may arise despite changes in depth or 3D-point predictions. To capture such changes, we introduce reference-relative deviation. Reference-relative deviation is motivated by the relative stability of real-image geometric responses under online dissemination. As shown in Fig. 2b, geometric predictions for real images change less across repeated dissemination than those for AI-generated images. We interpret this relative stability as reflecting the coherent physical 3D structure underlying real photographs. Motivated by this stability, GeoCue uses a frozen geometry predictor as a fixed reference and measures depth and 3D-point deviations from it, encouraging real-image predictions to remain close to the reference while AI-generated predictions deviate beyond a margin.

To construct these two cues, GeoCue employs a Geometry Reference Predictor G_{ref} and a Geometry Cue Predictor G_{cue}. Point-normal consistency is computed from the agreement between the point and normal predictions of G_{cue}, while reference-relative deviation is measured from the differences between the depth and point predictions of G_{cue} and G_{ref}.

3.2 POINT-NORMAL CONSISTENCY

Given an input image x, the Geometry Cue Predictor G_{cue} produces a 3D point map $P_{\text{cue}}(x)$ and a surface normal map $N_{\text{cue}}(x)$. From the local neighborhoods in $P_{\text{cue}}(x)$, we estimate local surface tangent directions and derive the implied surface normal by taking and normalizing their cross product, yielding a point-derived normal map $\widetilde{N}_{\text{cue}}(x)$. We then compare $\widetilde{N}_{\text{cue}}(x)$ with the directly predicted normal map $N_{\text{cue}}(x)$. Specifically, we define the point-normal consistency response as

$$c_{\text{pn}}(x) = \frac{1}{|\Omega_x|} \sum_{i \in \Omega_x} \left(1 - \cos(\widetilde{\mathbf{n}}_i(x), \mathbf{n}_i(x))\right), \tag{1}$$

where $\widetilde{\mathbf{n}}_i(x)$ denotes the normal derived from the 3D point map, $\mathbf{n}_i(x)$ denotes the surface normal predicted by G_{cue}, and Ω_x is the set of valid geometric pixels. A smaller c_{pn} indicates stronger agreement between the two geometric descriptions, whereas a larger value indicates greater disagreement.

3.3 REFERENCE-RELATIVE GEOMETRIC DEVIATION

Given an input image x, the frozen Geometry Reference Predictor G_{ref} produces the reference depth and 3D point maps $D_{\text{ref}}(x)$ and $P_{\text{ref}}(x)$, while the Geometry Cue Predictor G_{cue} produces the corresponding maps $D_{\text{cue}}(x)$ and $P_{\text{cue}}(x)$. To account for scene-scale variation, both reference and cue predictions are normalized using a common per-image scale derived from the reference depth, ensuring that their differences are measured on the same scale. We define the depth and point deviations as the average differences between the cue and reference predictions:

$$r_{\text{depth}}(x) = \frac{1}{|\Omega_x|} \sum_{i \in \Omega_x} \left| D^i_{\text{cue}}(x) - D^i_{\text{ref}}(x) \right|, \quad r_{\text{point}}(x) = \frac{1}{3|\Omega_x|} \sum_{i \in \Omega_x} \left\| P^i_{\text{cue}}(x) - P^i_{\text{ref}}(x) \right\|_1. \tag{2}$$

Here, the ℓ_1 norm sums the absolute differences across the three 3D coordinates, and the factor $1/3$ averages over them. The depth deviation captures changes in predicted depth, while the point deviation measures changes across the full 3D coordinates. Smaller r_{depth} and r_{point} indicate predictions closer to the reference, whereas larger values indicate greater geometric deviation.

3.4 MARGIN-BASED GEOMETRIC CRITERION

GeoCue organizes the three geometric responses using a margin-based criterion grounded in the characteristic real-image response range. Using this relatively stable range as a reference, the criterion requires AI-generated responses to lie beyond cue-specific margins. This formulation results in stronger point-normal agreement and smaller reference-relative deviations for real images.

For a real image x, we define

$$\mathcal{L}_{\text{real}}(x) = \lambda^{\text{real}}_{\text{d}} r_{\text{depth}}(x) + \lambda^{\text{real}}_{\text{p}} r_{\text{point}}(x) + \lambda^{\text{real}}_{\text{pn}} c_{\text{pn}}(x), \tag{3}$$

where $\lambda^{\text{real}}_{\text{d}}$, $\lambda^{\text{real}}_{\text{p}}$, and $\lambda^{\text{real}}_{\text{pn}}$ control the contribution of each cue.

For an AI-generated image x, we define

$$\mathcal{L}_{\text{AIGI}}(x) = \lambda^{\text{AIGI}}_{\text{d}} [m_{\text{d}} - r_{\text{depth}}(x)]_+ + \lambda^{\text{AIGI}}_{\text{p}} [m_{\text{p}} - r_{\text{point}}(x)]_+ + \lambda^{\text{AIGI}}_{\text{pn}} [m_{\text{pn}} - c_{\text{pn}}(x)]_+, \tag{4}$$

where m_{d}, m_{p}, and m_{pn} denote the cue-specific margins, and $[z]_+ = \max(z, 0)$. To set the point-normal margin, we measure the point-normal consistency responses of real images in the training set using the frozen geometry predictor. Based on this distribution, we set $m_{\text{pn}} = 0.5$, so that 99.65% of the real-image responses lie below the margin. For the deviation cues, we use a shared margin $m_{\text{d}} = m_{\text{p}} = 0.1875$. Once a response exceeds its corresponding margin, no additional penalty is imposed.

The overall geometric objective is

$$\mathcal{L}_{\text{GeoCue}} = \mathcal{L}_{\text{real}} + \mathcal{L}_{\text{AIGI}}. \tag{5}$$

We train the Geometry Cue Predictor by minimizing $\mathcal{L}_{\text{GeoCue}}$. The resulting geometric responses $c_{\text{pn}}, r_{\text{depth}}, r_{\text{point}}$ are then fed to a linear classifier for final detection.

Table 1: **In-the-wild evaluation.** AUC on Chameleon (Yan et al., 2025), LOKI (Ye et al., 2025), and WildRF (Cavia et al., 2024).

Detector \ Datasets	Chameleon	LOKI	WildRF	Mean
CNNSpot (Wang et al., 2020)	30.8	51.2	43.2	41.7
Corvi (Corvi et al., 2023)	36.9	62.7	66.7	55.4
UnivFD (Ojha et al., 2023)	45.3	53.8	58.7	52.6
NPR (Tan et al., 2024a)	50.1	72.7	67.4	63.4
AEROBLADE (Ricker et al., 2024)	63.6	69.6	70.3	67.8
ClipDet (Cozzolino et al., 2024)	57.4	68.6	84.7	70.2
LaDeDa (Cavia et al., 2024)	48.7	77.9	75.1	67.2
AIDE (Yan et al., 2025)	74.4	82.6	68.7	75.2
C2P-CLIP (Tan et al., 2025)	35.3	69.7	64.6	56.5
FerretNet (Liang et al., 2025)	47.0	77.7	72.2	65.6
DEAR-c (Kim et al., 2026)	72.1	85.9	92.0	83.3
DEAR-r (Kim et al., 2026)	66.1	85.8	90.6	80.8
DFFreq (Yan et al., 2026)	52.1	82.6	65.0	66.6
GeoCue	**88.2**	**81.9**	**94.3**	**88.1**

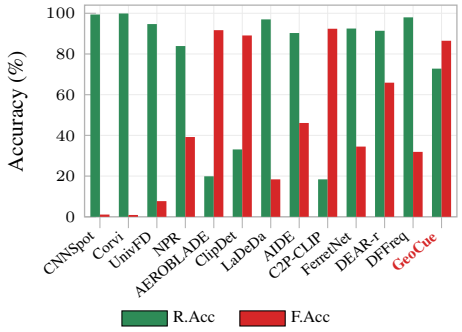


Figure 4: **Class-wise accuracy balance in the wild.** Mean R.Acc. and F.Acc. across the three benchmarks; a smaller gap indicates better balance.

Table 2: **Cross-generator evaluation on AIGCDetectBench (Zhong et al., 2023).** AUC across 16 generators.

Generator	CNNSpot	Corvi	UnivFD	LaDeDa	C2P-CLIP	DEAR-r	GeoCue
GAN							
ProGAN	100.0	92.9	100.0	99.9	83.7	92.0	86.6
CycleGAN	97.0	56.5	99.8	99.7	96.8	57.3	88.2
BigGAN	89.8	80.3	99.3	93.2	84.4	51.2	93.2
StyleGAN	98.0	76.0	97.9	99.9	64.6	72.7	79.6
StyleGAN2	97.7	72.7	97.0	99.9	63.2	67.8	79.2
GauGAN	98.2	90.6	100.0	97.1	77.4	49.3	98.4
StarGAN	94.7	39.8	99.7	100.0	97.6	92.6	59.0
Diffusion							
SD1.4	56.0	91.2	71.7	83.0	94.4	99.9	94.2
SD1.5	55.9	91.0	71.5	83.4	94.2	99.9	94.6
ADM	70.0	86.2	92.5	76.8	83.0	91.4	89.4
GLIDE	72.6	96.5	86.1	67.0	90.0	90.4	99.0
Wukong	52.5	90.2	78.4	85.0	86.9	100.0	93.4
VQDM	73.3	98.9	96.2	92.0	87.7	99.7	91.7
DALLE2	44.5	95.3	63.9	47.7	87.9	98.8	98.4
Others							
Midjourney	49.3	68.3	47.3	61.1	82.1	99.2	95.4
WFIR	83.3	78.7	98.3	91.0	61.2	68.9	87.9
Mean	77.1	81.6	87.5	86.0	83.5	83.2	**89.3**

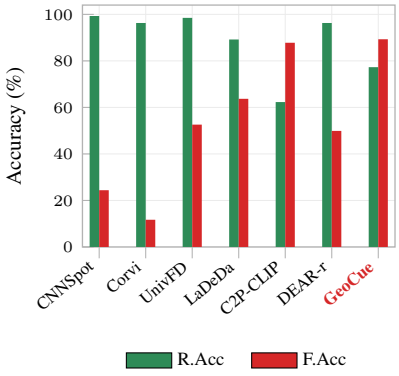


Figure 5: **Class-wise accuracy balance on AIGCDetectBench (Zhong et al., 2023).** Mean R.Acc. and F.Acc. across 16 generators; a smaller gap indicates better balance.

4 Experiments

4.1 Experimental Settings

Training and Evaluation Setup. We use MoGe-2 ViT-L (Wang et al., 2025c) for both geometry predictors, with the Geometry Reference Predictor frozen and the Geometry Cue Predictor trainable. We train on a compact set of 8,000 images from AIGI-Holmes (Zhou et al., 2025), spanning 10 generators: FLUX, Infinity, Janus, Janus-Pro-1B, Janus-Pro-7B, LlamaGen, PixArt-XL, SD35-L, Show-o, and VAR, with 400 real and 400 AI-generated images per generator. The training set does not overlap with any evaluation benchmark. We evaluate on LSUN-Bedroom (Yu et al., 2015) and AIGCDetectBench (Zhong et al., 2023) for cross-generator evaluation, and Chameleon (Yan et al., 2025), LOKI (Ye et al., 2025), and WildRF (Cavia et al., 2024) for in-the-wild evaluation. The same GeoCue model is used for all benchmarks without dataset-specific fine-tuning. Detailed training and evaluation configurations are provided in Appendix A.

Online Dissemination. To evaluate robustness to repeated online sharing, we construct three cumulative dissemination sequences with different combinations and orders of resizing and JPEG/WebP re-encoding. These sequences allow us to assess robustness across diverse dissemination conditions rather than relying on a particular processing trajectory. Each round operates on the previous decoded output and is applied identically to real and AI-generated images without retraining. Detailed configurations are provided in Appendix A.

Table 3: **Cross-generator evaluation on LSUN-Bedroom (Yu et al., 2015).** AUC across 10 generators in a shared semantic domain.

	GAN					Diffusion					
Detector \ Generator	ProGAN	ProjGAN	Diff-ProjGAN	StyleGAN	Diff-StyleGAN2	DDPM	IDDPM	ADM	LDM	PNDM	**Mean**
CNNSpot (Wang et al., 2020)	100.0	90.8	88.2	97.0	99.5	82.5	78.0	67.2	75.7	75.6	85.5
Corvi (Corvi et al., 2023)	89.7	80.6	86.5	89.9	78.7	84.4	84.8	76.3	100.0	77.9	84.9
AEROBLADE (Ricker et al., 2024)	38.9	45.7	46.1	40.0	56.7	66.7	59.1	53.6	65.4	43.4	51.5
ClipDet (Cozzolino et al., 2024)	98.2	97.1	97.5	90.9	95.7	83.5	74.0	60.8	81.6	88.8	86.8
C2P-CLIP (Tan et al., 2025)	97.6	95.7	95.1	90.1	94.7	71.3	66.8	66.9	94.4	96.9	86.9
DEAR-c (Kim et al., 2026)	83.8	78.8	86.2	78.7	75.7	77.4	74.1	75.0	99.9	87.1	81.7
DEAR-r (Kim et al., 2026)	60.9	58.5	67.2	57.4	68.4	60.2	62.4	67.9	98.6	48.7	65.0
GeoCue	97.2	95.2	96.0	78.4	94.1	88.8	80.3	71.8	85.2	89.1	87.6

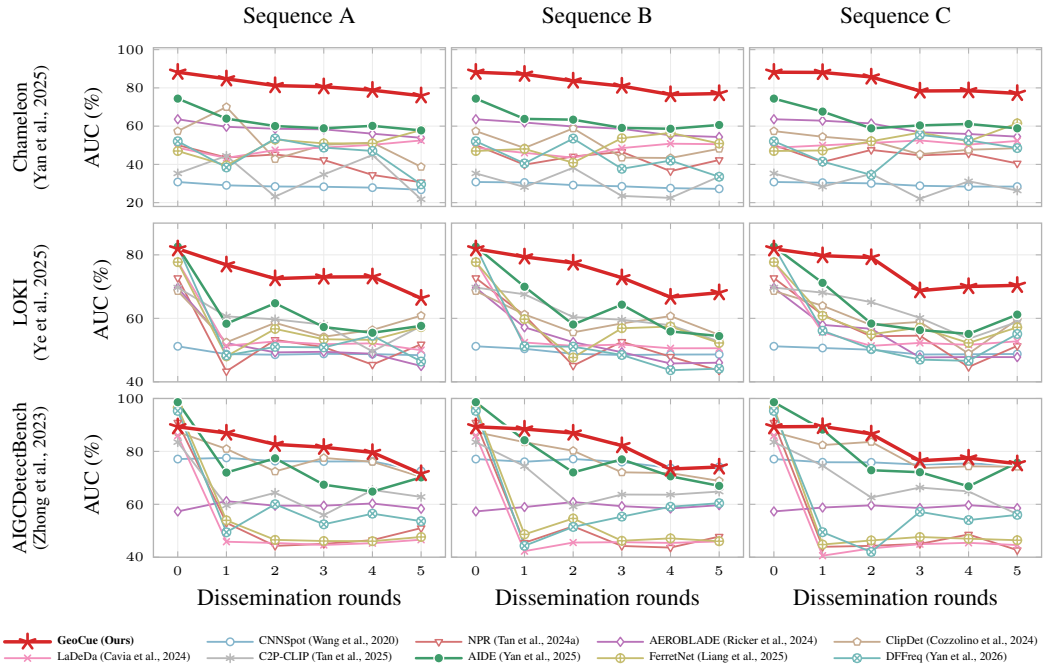


Figure 6: **Robustness to repeated online dissemination.** AUC across successive rounds under three cumulative dissemination sequences on diverse benchmarks, with Round 0 denoting the original images. GeoCue maintains stable performance across rounds and sequences.

4.2 Main Results

In-the-Wild Generalization. We evaluate GeoCue on Chameleon (Yan et al., 2025), LOKI (Ye et al., 2025), and WildRF (Cavia et al., 2024). These benchmarks are particularly challenging because they contain images from unknown or diverse generation sources under uncontrolled real-world conditions. As shown in Table 1, GeoCue achieves the highest mean AUC across the three in-the-wild benchmarks. Fig. 4 further shows that GeoCue maintains relatively balanced R.Acc. and F.Acc. when averaged across the three benchmarks, while several prior methods exhibit substantial class-wise imbalance. Detailed per-benchmark results, including AUC, R.Acc., and F.Acc., are provided in Appendix B.

Cross-Generator Evaluation. We evaluate GeoCue on AIGCDetectBench (Zhong et al., 2023), which contains 16 unseen generators. As shown in Table 2, GeoCue achieves strong mean AUC across diverse unseen generators. Fig. 5 further shows that GeoCue maintains relatively balanced R.Acc. and F.Acc., while several prior methods exhibit substantial class-wise imbalance. We further evaluate on LSUN-Bedroom (Yu et al., 2015), where real and AI-generated images share the same semantic domain. As shown in Table 3, GeoCue achieves the highest mean AUC under this

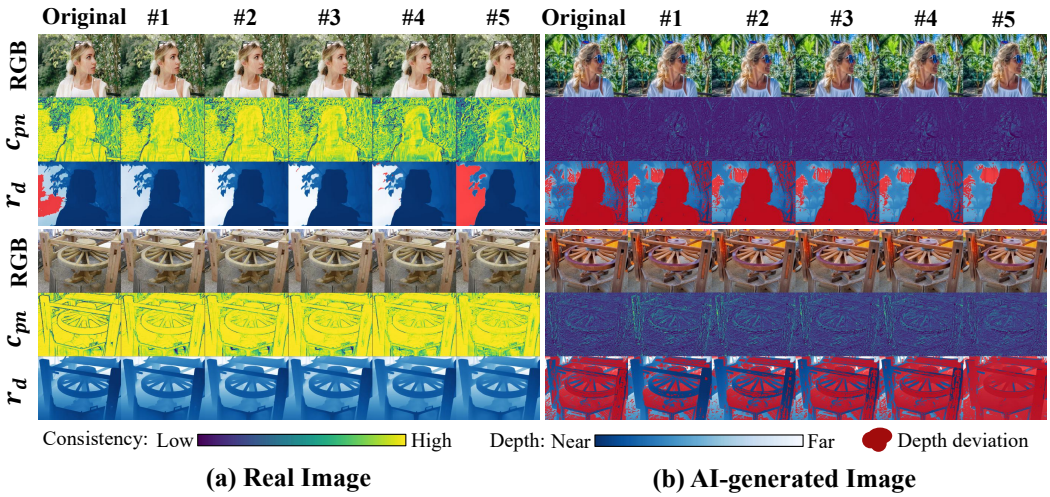


Figure 7: **Point-normal consistency and depth deviation under online dissemination.** The geometric cue responses remain largely preserved across repeated dissemination rounds for representative real and AI-generated images.

Table 4: **Geometric cue complementarity and margin scale.** (a) Complementarity of point-normal consistency and reference-relative deviation, with additional geometric cue variants. (b) Effect of jointly scaling the cue-specific margins.

(a) Complementarity of geometric cues.

Geometric cues					LOKI			WildRF		
c_{pn}	r_{depth}	r_{point}	c_{dn}	r_{normal}	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc
✓					79.4	58.6	80.6	92.2	71.1	92.7
	✓	✓			78.1	87.2	49.1	86.6	80.6	77.4
✓	✓	✓			81.9	65.7	77.6	94.3	83.8	91.3
Additional variants										
✓	✓	✓	✓		79.9	61.7	77.9	90.1	73.4	92.2
✓	✓	✓	✓	✓	79.5	62.7	77.7	89.7	73.1	92.3

(b) Effect of the Margin Scale.

	LOKI			WildRF		
Scale	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc
0.4×	75.4	66.3	67.5	95.0	87.2	91.9
0.8×	80.2	67.0	74.7	91.2	74.2	92.2
1.0×	81.9	65.7	77.6	94.3	83.8	91.3
1.2×	79.7	54.2	84.4	90.1	72.3	93.1
1.4×	80.7	55.3	83.3	90.0	74.3	93.0

controlled semantic setting. Detailed per-benchmark and per-generator results, including AP and class-wise accuracies, are provided in Appendix B.

Robustness to Repeated Online Dissemination. We evaluate robustness under repeated online dissemination across all three dissemination sequences. As shown in Fig. 6, GeoCue maintains stable detection performance across rounds and sequences on diverse benchmarks. In contrast, many existing detectors exhibit substantial performance degradation or fluctuations as dissemination progresses, even when they perform strongly on the original images. This robustness is particularly evident on in-the-wild benchmarks, where the underlying generators are unknown. This further supports the ability of GeoCue to generalize beyond generator-specific characteristics. Fig. 7 further shows that the characteristic point-normal consistency and reference-relative depth deviation patterns remain largely preserved across rounds. Together, these results support the robustness of geometric cues under diverse online dissemination conditions.

4.3 Analysis

Cue Complementarity and Margin Sensitivity. Table 4a examines the complementarity between point-normal consistency and reference-relative deviation. Each cue provides useful detection evidence on its own, while their combination consistently improves AUC on both benchmarks, supporting their complementarity. Adding depth-normal consistency and normal deviation yields no consistent improvement, indicating that these cues largely overlap with the information already captured by the three GeoCue responses. Table 4b examines the effect of the margin scale. Before training the cue predictor, we use the relatively stable real-image response region observed with the frozen geometry predictor as a reference for defining the point-normal margin. We use

Table 5: **Analysis of the GeoCue objective.** (a) Effect of reversing the objective configuration for point-normal consistency and reference-relative deviation. (b) Comparison with frozen MoGe-2 (Wang et al., 2025c) and direct CLS-based classification.

(a) Effect of the objective configuration.

	PN Consistency		Deviation		Chameleon			LOKI		
Setting	Real	AIGI	Real	AIGI	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc
Reversed (All)	↓	↑	↑	↓	80.6	61.5	89.8	70.0	43.4	84.2
Reversed (Consistency)	↓	↑	↓	↑	85.8	76.8	81.1	78.3	72.2	73.3
GeoCue	↑	↓	↓	↑	88.2	68.8	90.6	81.9	65.7	77.6

For PN consistency, ↑ indicates stronger agreement and ↓ responses beyond the margin;
for deviation, ↓ indicates smaller deviation and ↑ indicates larger deviation.

(b) Comparison with MoGe-2 baselines.

	LOKI			WildRF		
Setting	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc
MoGe-2 (Frozen)	63.6	38.0	81.9	89.4	72.3	89.5
MoGe-2 (CLS FT)	78.5	57.0	83.0	92.2	77.5	90.2
GeoCue	81.9	65.7	77.6	94.3	83.8	91.3

Table 6: **Effect of the geometry predictor.** All predictors use the same geometric cue formulation but differ in their input setting and geometry construction.

	LSUN		Chameleon			LOKI			WildRF			Geometry Construction		
Geometry Predictor	AUC	AP	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc	Input	Point	Depth
Dens3R (Fang et al., 2026)	49.7	49.3	69.9	52.8	77.0	61.9	77.3	43.2	78.3	81.9	62.1	Self-pair	Direct XYZ	From point z
Metric3D v2 (Hu et al., 2024)	71.8	63.2	80.5	50.6	90.7	72.6	59.3	74.6	86.3	63.3	95.8	Single	From depth	Direct
MoGe-2 (Wang et al., 2025c)	87.6	84.6	88.2	68.8	90.6	81.9	65.7	77.6	94.3	83.8	91.3	Single	Direct XYZ	From point z

$(m_d, m_p, m_{pn}) = (0.1875,\ 0.1875,\ 0.500)$ as the predefined $1.0\times$ setting. Across a range of margin scales, this setting provides the most balanced overall performance across benchmarks, further supporting our default margin configuration.

Effect of the GeoCue Objective. Table 5a examines the design of the GeoCue objective. When the point-normal consistency term is reversed, AI-generated images are encouraged toward stronger agreement while real-image responses are pushed beyond the margin, resulting in lower performance. Further reversing reference-relative deviation leads to an additional drop. Together, these results support using the relatively stable real-image response region as a reference. Table 5b compares GeoCue with alternative baseline designs. The CLS baseline directly uses the CLS feature for binary classification. Direct CLS-based classification improves over the frozen MoGe-2 baseline, and GeoCue further achieves higher AUC with a better balance between R.Acc. and F.Acc. on both benchmarks. These results highlight the benefit of using explicit geometric cues beyond direct CLS-based fine-tuning.

Geometry Predictor Comparison. As shown in Table 6, Metric3D v2 (Hu et al., 2024) constructs 3D points by back-projecting predicted depth, making the depth and point deviations largely redundant. Dens3R (Fang et al., 2026) directly predicts 3D points, but using self-paired inputs removes the cross-view geometric information its pair-based design is intended to exploit. In contrast, MoGe-2 (Wang et al., 2025c) is natively designed for single-image geometry prediction and directly predicts an XYZ point map, allowing point deviation to capture lateral x-y variations beyond depth alone. These results highlight that the effectiveness of geometric cues depends on both how the underlying geometry is constructed and how well the predictor aligns with the single-image setting.

5 Conclusion

We introduced GeoCue, a geometry-based framework for AI-generated image detection motivated by fundamental differences in image formation between real photographs and AI-generated images. Starting from this perspective, we showed that geometric coherence can provide useful single-image detection evidence through point-normal consistency and reference-relative geometric deviations. Across unseen-generator, shared-domain, and in-the-wild benchmarks, GeoCue achieves strong and balanced performance. The geometric responses themselves also remain largely stable under repeated online dissemination. Together, these results highlight geometric coherence as a robust source of detection evidence beyond conventional image-domain signals.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation, provide feedback on aspects of the experimental methodology and its presentation, assist with parts of the implementation, and help draft and refine parts of the manuscript. We also used these tools to help create or revise conceptual scientific figures and identify relevant literature. All AI-assisted content incorporated into this work was reviewed and verified by the authors. The authors take full responsibility for the accuracy and integrity of the submitted work.

REFERENCES

Xu Cao, Boxin Shi, Fumio Okura, and Yasuyuki Matsushita. Normal integration via inverse plane fitting with minimum point-to-plane distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2382–2391, 2021.

Bar Cavia, Eliahu Horwitz, Tal Reiss, and Yedid Hoshen. Real-time deepfake detection in the real-world. *arXiv preprint arXiv:2406.09398*, 2024.

Andrea Ciamarra, Roberto Caldelli, Federico Becattini, Lorenzo Seidenari, and Alberto Del Bimbo. Deepfake detection by exploiting surface anomalies: The SurFake approach. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pp. 1024–1033, 2024.

Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. On the detection of synthetic images generated by diffusion models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.

Davide Cozzolino, Giovanni Poggi, Riccardo Corvi, Matthias Nießner, and Luisa Verdoliva. Raising the bar of AI-generated image detection with CLIP. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 4356–4366, 2024.

Xianze Fang, Jingnan Gao, Zhe Wang, Zhuo Chen, Xingyu Ren, Jiangjing Lyu, Qiaomu Ren, Zhonglei Yang, Xiaokang Yang, Yichao Yan, and Chengfei Lyu. Dens3R: A foundation model for 3D geometry prediction. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2026.

Pierre Fernandez, Guillaume Couairon, Hervé Jégou, Matthijs Douze, and Teddy Furon. The Stable Signature: Rooting watermarks in latent diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22466–22477, 2023.

Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. GeoWizard: Unleashing the diffusion priors for 3D geometry estimation from a single image. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 241–258, 2024.

Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3D v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10579–10596, 2024.

Minsun Jeon and Simon S. Woo. Seeing through the blur: Unlocking defocus maps for deepfake detection. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1091–1102, 2025.

Dahye Kim, Jaehyun Choi, Hyun Seok Seong, Seongho Kim, Donghun Lee, Sungwon Yi, and Jang-Ho Choi. Dissect and prune: Enhancing robustness in AI-generated image detection. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2026.

Ouxiang Li, Jiayin Cai, Yanbin Hao, Xiaolong Jiang, Yao Hu, and Fuli Feng. Improving synthetic image detection towards generalization: An image transformation perspective. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 2405–2414, 2025.

Shuqiao Liang, Jian Liu, Renzhang Chen, and Quanlong Guan. FerretNet: Efficient synthetic image detection via local pixel dependencies. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Utkarsh Ojha, Yuheng Li, and Yong Jae Lee. Towards universal fake image detectors that generalize across generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24480–24489, 2023.

Xiaojuan Qi, Renjie Liao, Zhengzhe Liu, Raquel Urtasun, and Jiaya Jia. GeoNet: Geometric neural network for joint depth and surface normal estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 283–291, 2018.

Jonas Ricker, Denis Lukovnikov, and Asja Fischer. AEROBLADE: Training-free detection of latent diffusion images using autoencoder reconstruction error. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9130–9140, 2024.

Ayush Sarkar, Hanlin Mai, Amitabh Mahapatra, Svetlana Lazebnik, David A. Forsyth, and Anand Bhattad. Shadows don't lie and lines can't bend! generative models don't know projective geometry... for now. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28140–28149, 2024.

Zeyang Sha, Zheng Li, Ning Yu, and Yang Zhang. DE-FAKE: Detection and attribution of fake images generated by text-to-image generation models. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 3418–3432, 2023.

Chuangchuang Tan, Huan Liu, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, and Yunchao Wei. Rethinking the up-sampling operations in CNN-based generative network for generalizable deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28130–28139, 2024a.

Chuangchuang Tan, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, and Yunchao Wei. Frequency-aware deepfake detection: Improving generalizability through frequency space domain learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 5052–5060, 2024b.

Chuangchuang Tan, Renshuai Tao, Huan Liu, Guanghua Gu, Baoyuan Wu, Yao Zhao, and Yunchao Wei. C2P-CLIP: Injecting category common prompt in CLIP to enhance generalization in deepfake detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, pp. 7184–7192, 2025.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual geometry grounded transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306, 2025a.

Ruicheng Wang, Sicheng Xu, Cassie Dai, Jianfeng Xiang, Yu Deng, Xin Tong, and Jiaolong Yang. MoGe: Unlocking accurate monocular geometry estimation for open-domain images with optimal training supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5261–5271, 2025b.

Ruicheng Wang, Sicheng Xu, Yue Dong, Yu Deng, Jianfeng Xiang, Zelong Lv, Guangzhong Sun, Xin Tong, and Jiaolong Yang. MoGe-2: Accurate monocular geometry with metric scale and sharp details. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, 2025c. doi: 10.52202/085713-1207.

Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A. Efros. CNN-generated images are surprisingly easy to spot... for now. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8695–8704, 2020.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jérôme Revaud. DUSt3R: Geometric 3D vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709, 2024.

Zhendong Wang, Jianmin Bao, Wengang Zhou, Weilun Wang, Hezhen Hu, Hong Chen, and Houqiang Li. DIRE for diffusion-generated image detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22445–22455, 2023.

Yuxin Wen, John Kirchenbauer, Jonas Geiping, and Tom Goldstein. Tree-Rings watermarks: Invisible fingerprints for diffusion images. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 58047–58063, 2023.

Jiazhen Yan, Ziqiang Li, Fan Wang, Ziwen He, and Zhangjie Fu. Dual frequency branch framework with reconstructed sliding windows attention for AI-generated image detection. *IEEE Transactions on Information Forensics and Security*, 21:2313–2325, 2026. doi: 10.1109/TIFS.2026.3664000.

Shilin Yan, Ouxiang Li, Jiayin Cai, Yanbin Hao, Xiaolong Jiang, Yao Hu, and Weidi Xie. A sanity check for AI-generated image detection. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth Anything V2. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 21875–21911, 2024.

Junyan Ye, Baichuan Zhou, Zilong Huang, Junan Zhang, Tianyi Bai, Hengrui Kang, Jun He, Honglin Lin, Zihao Wang, Tong Wu, Zhizheng Wu, Yiping Chen, Dahua Lin, Conghui He, and Weijia Li. LOKI: A comprehensive synthetic data detection benchmark using large multimodal models. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015.

Ning Yu, Larry S. Davis, and Mario Fritz. Attributing fake images to GANs: Learning and analyzing GAN fingerprints. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7556–7566, 2019.

Nan Zhong, Yiran Xu, Sheng Li, Zhenxing Qian, and Xinpeng Zhang. PatchCraft: Exploring texture patch for efficient AI-generated image detection. *arXiv preprint arXiv:2311.12397*, 2023.

Ziyin Zhou, Yunpeng Luo, Yuanchen Wu, Ke Sun, Jiayi Ji, Ke Yan, Shouhong Ding, Xiaoshuai Sun, Yunsheng Wu, and Rongrong Ji. AIGI-Holmes: Towards explainable and generalizable AI-generated image detection via multimodal large language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 18746–18758, 2025.

A Additional Experimental Details

Geometry Predictor Training. We use MoGe-2 ViT-L (Wang et al., 2025c) as the backbone for both G_{ref} and G_{cue}. G_{ref} remains frozen throughout training, whereas G_{cue} is adapted using LoRA on the fused QKV projections of the last 12 transformer blocks, with rank 16 and scaling factor 16.

We train on 8,000 images from AIGI-Holmes (Zhou et al., 2025) using 10 generator subsets: FLUX, Infinity, Janus, Janus-Pro-1B, Janus-Pro-7B, LlamaGen, PixArt-XL, SD35-L, Show-o, and VAR. For each subset, we use 400 real and 400 AI-generated images, with balanced sampling across generators and classes. The training set does not overlap with any evaluation benchmark. All images are processed at 448×448 resolution, with horizontal flipping applied with probability 0.5. Before computing the geometric responses, the depth and 3D point predictions are normalized using a per-image scale defined by the median valid depth from G_{ref}.

We train G_{cue} for 5 epochs using AdamW with a learning rate of 1×10^{-4}, weight decay of 1×10^{-4}, and a batch size of 12 per GPU.

Classifier Training. The three geometric responses, c_{pn}, r_{depth}, and r_{point}, are standardized using statistics computed from the training set and fed to a linear classifier. The classifier is trained for 30 epochs using binary cross-entropy loss and Adam with a learning rate of 1×10^{-2}, weight decay of 1×10^{-4}, and a batch size of 1024.

Evaluation. All evaluation benchmarks use the same geometry predictors and classifier without dataset-specific fine-tuning. We report AUC and AP from the classifier scores, where AUC measures overall real-AIGI discrimination and AP summarizes precision-recall performance for AI-generated images. We additionally report R.Acc. and F.Acc., denoting classification accuracy on real and AI-generated images, respectively. A fixed probability threshold of 0.5 is used for all benchmarks without dataset-specific threshold tuning.

Baseline Training Data. Table 7 summarizes the task-specific training data used by the compared detectors, providing context for differences in training scale and data sources across methods.

Table 7: **Training data of detectors.** For each detector, we report the real/fake sources and the number of training images.

Detector	Training data (real / fake)	Training size
CNNSpot (Wang et al., 2020)	LSUN / ProGAN	360K real + 360K fake
Corvi (Corvi et al., 2023)	COCO + LSUN / LDM	180K real + 180K fake
UnivFD (Ojha et al., 2023)	LSUN / ProGAN	360K real + 360K fake
NPR (Tan et al., 2024a)	LSUN / ProGAN	72K real + 72K fake
AEROBLADE (Ricker et al., 2024)	Training-free	–
ClipDet (Cozzolino et al., 2024)	COCO / LDM	10K real + 10K fake
LaDeDa (Cavia et al., 2024)	LSUN / ProGAN	360K real + 360K fake
C2P-CLIP (Tan et al., 2025)	LSUN / ProGAN	72K real + 72K fake
FerretNet (Liang et al., 2025)	LSUN / ProGAN	72K real + 72K fake
DEAR-c (Kim et al., 2026)	COCO + LSUN / LDM + diagnostic inpaint	180K real + 180K fake + 90K diagnostic
DEAR-r (Kim et al., 2026)	COCO + LSUN / LDM-VAE recon. + diagnostic inpaint	180K real + 180K fake + 90K diagnostic
DFFreq (Yan et al., 2026)	LSUN / ProGAN	72K real + 72K fake
GeoCue	AIGI-Holmes (Zhou et al., 2025)	4K real + 4K fake (8K total)

Online Dissemination Setup. Repeated online sharing can introduce cumulative resizing and recompression as images are repeatedly uploaded, processed, and redistributed. To model this process, we apply a sequence of relative resizing and JPEG/WebP recompression steps, with each round operating on the decoded output of the previous round.

Before the first round, images whose longer side exceeds 768 pixels are resized to 768 pixels while preserving the aspect ratio. At each dissemination round, the current image is resized by the relative ratio specified in Table 8 and then recompressed using the corresponding codec and quality setting. The resize ratios 0.90, 0.95, and 1.00 are relative to the current image size; a ratio of 1.00 preserves

Table 8: **Online dissemination sequences.** Each sequence consists of five cumulative rounds, with each entry specifying the relative resize ratio and recompression setting applied to the decoded output of the previous round.

Sequence	Round 1	Round 2	Round 3	Round 4	Round 5
A	0.90/JPEG70	0.95/WebP85	1.00/JPEG95	0.90/JPEG85	0.95/WebP70
B	0.95/WebP95	1.00/JPEG70	0.90/WebP85	0.95/WebP70	1.00/JPEG85
C	1.00/WebP95	0.90/JPEG95	0.95/WebP70	1.00/JPEG85	0.90/WebP95

Table 9: **Detailed in-the-wild evaluation.** AUC / R.Acc. / F.Acc. on Chameleon (Yan et al., 2025), LOKI (Ye et al., 2025), and WildRF (Cavia et al., 2024). For WildRF, results are reported for Reddit, Twitter, and Facebook subsets and their mean.

Detector \ Datasets	Chameleon			LOKI			WildRF											
							Reddit			Twitter			Facebook			Mean		
	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc	AUC	R.Acc	F.Acc
CNNSpot (Wang et al., 2020)	30.8	99.6	0.0	51.2	99.0	3.0	41.8	99.1	0.7	42.2	99.7	0.0	45.5	100.0	0.0	43.2	99.6	0.2
Corvi (Corvi et al., 2023)	36.9	99.8	0.0	62.7	100.0	2.0	72.4	100.0	2.0	70.0	100.0	0.3	57.7	100.0	0.0	66.7	100.0	0.8
UnivFD (Ojha et al., 2023)	45.3	98.2	0.6	53.8	87.0	19.5	69.8	96.8	5.6	56.8	99.7	3.8	49.4	100.0	0.0	58.7	98.8	3.1
NPR (Tan et al., 2024a)	50.1	99.5	6.1	72.7	73.1	69.7	69.7	73.1	61.7	50.7	71.6	27.5	81.9	93.1	36.2	67.4	79.2	41.8
AEROBLADE (Ricker et al., 2024)	63.6	9.6	98.1	69.6	21.8	90.5	71.4	35.5	82.9	72.1	26.4	89.5	67.4	23.1	87.5	70.3	28.3	86.6
ClipDet (Cozzolino et al., 2024)	57.4	11.8	91.9	68.6	42.8	81.2	83.9	42.0	96.0	86.7	55.4	93.3	83.5	36.2	93.8	84.7	44.6	94.3
LaDeDa (Cavia et al., 2024)	48.7	100.0	1.8	77.9	94.4	40.7	78.8	94.5	29.3	62.3	95.3	4.7	84.1	100.0	4.4	75.1	96.6	12.8
AIDE (Yan et al., 2025)	74.4	95.6	26.0	82.6	84.8	71.1	80.1	88.3	59.7	64.4	94.1	29.8	61.5	89.4	33.8	68.7	90.6	41.1
C2P-CLIP (Tan et al., 2025)	35.3	12.4	91.9	69.7	29.7	90.1	56.9	9.9	93.3	62.5	12.0	94.2	74.4	17.5	98.1	64.6	13.1	95.2
FerretNet (Liang et al., 2025)	47.0	99.8	5.6	77.7	85.7	67.0	75.4	88.5	56.9	58.8	87.7	14.6	82.5	99.4	21.2	72.2	91.9	30.9
DEAR-r (Kim et al., 2026)	66.1	81.7	54.1	85.8	94.2	68.1	81.5	97.2	58.1	91.9	97.9	76.3	98.4	99.4	92.5	90.6	98.2	75.6
DFFreq (Yan et al., 2026)	52.1	99.6	5.5	82.6	94.9	69.1	82.1	98.7	57.2	57.3	99.7	5.3	55.5	100.0	0.6	65.0	99.5	21.0
GeoCue	**88.2**	**68.8**	**90.6**	**81.9**	**65.7**	**77.6**	94.0	83.5	92.9	95.0	84.8	92.1	94.0	83.1	88.8	**94.3**	**83.8**	**91.3**

the resolution while still applying recompression. Bicubic interpolation is used for resizing. JPEG encoding uses 4:2:0 chroma subsampling, while WebP uses lossy encoding. After each round, the image is decoded to RGB and used as the input to the next round.

B Detailed Evaluation Results

This section provides the full benchmark results corresponding to the main evaluations reported in Sec. 4.2, including class-wise accuracies and per-generator performance. The detailed results further show that GeoCue maintains strong detection performance without exhibiting a strong bias toward either the real or AI-generated class.

In-the-Wild Evaluation. As shown in Table 9, GeoCue maintains strong detection performance across all three in-the-wild benchmarks (Yan et al., 2025; Ye et al., 2025; Cavia et al., 2024). It also exhibits relatively balanced R.Acc. and F.Acc., whereas many existing detectors show substantial class-wise bias toward either real or AI-generated images. These results demonstrate that GeoCue remains effective for both classes under diverse and challenging real-world distributions.

Cross-Generator Evaluation on AIGCDetectBench. As shown in Table 10, GeoCue achieves strong overall generalization across the 16 unseen generators in AIGCDetectBench (Zhong et al., 2023), spanning GAN-based, diffusion-based, and other generative models. GeoCue achieves the highest mean AUC while maintaining relatively balanced R.Acc. and F.Acc. In contrast, several baselines exhibit substantial class-wise imbalance despite strong performance on particular generator families. These results demonstrate that geometric cues provide broadly transferable detection evidence across diverse generator shifts.

Cross-Generator Evaluation in a Shared Semantic Domain. As shown in Table 11, GeoCue maintains strong cross-generator performance on LSUN-Bedroom (Yu et al., 2015), where both real and AI-generated images belong to the same semantic domain. This setting reduces the influence of coarse semantic differences between the two classes, providing a more controlled evaluation

Table 10: **Cross-generator evaluation on AIGCDetectBench (Zhong et al., 2023).** AUC / R.Acc. / F.Acc. across 16 generators.

Generator	CNNSpot (Wang et al., 2020)	Corvi (Corvi et al., 2023)	UnivFD (Ojha et al., 2023)	AEROBLADE (Ricker et al., 2024)	ClipDet (Cozzolino et al., 2024)	LaDeDa (Cavia et al., 2024)	C2P-CLIP (Tan et al., 2025)	DEAR-c (Kim et al., 2026)	DEAR-r (Kim et al., 2026)	**GeoCue (Ours)**
GAN										
ProGAN	100.0 / 100.0 / 100.0	92.9 / 100.0 / 1.9	100.0 / 99.8 / 99.9	39.5 / 1.5 / 96.3	92.6 / 46.8 / 99.2	99.9 / 100.0 / 87.3	83.7 / 29.5 / 97.4	94.9 / 100.0 / 45.9	92.0 / 100.0 / 26.6	86.6 / 36.2 / 100.0
CycleGAN	97.0 / 98.6 / 62.8	56.5 / 91.2 / 2.3	99.8 / 97.5 / 99.2	36.7 / 16.0 / 82.4	92.4 / 77.2 / 97.4	99.7 / 99.6 / 87.4	96.8 / 90.8 / 93.1	34.0 / 84.0 / 9.0	57.3 / 97.2 / 6.1	88.2 / 80.9 / 94.2
BigGAN	89.8 / 99.1 / 18.9	80.3 / 99.0 / 5.3	99.3 / 99.0 / 92.5	32.0 / 4.2 / 85.7	81.9 / 67.1 / 81.0	93.2 / 89.5 / 77.6	84.4 / 63.9 / 91.0	76.8 / 87.6 / 53.2	51.2 / 90.2 / 21.0	93.2 / 83.0 / 97.0
StyleGAN	98.0 / 99.9 / 33.9	76.0 / 95.6 / 24.7	97.9 / 99.7 / 65.5	33.1 / 1.1 / 94.4	81.4 / 53.7 / 90.1	99.9 / 100.0 / 74.4	64.6 / 53.3 / 66.5	75.6 / 73.3 / 40.4	72.7 / 79.7 / 27.3	79.6 / 47.7 / 97.1
StyleGAN2	97.7 / 100.0 / 28.7	72.7 / 99.8 / 3.4	97.0 / 99.8 / 43.2	31.3 / 2.1 / 94.6	80.3 / 53.9 / 89.9	99.9 / 100.0 / 78.7	63.2 / 48.9 / 68.5	73.1 / 87.6 / 24.3	67.8 / 92.1 / 19.5	79.2 / 49.4 / 97.7
GauGAN	98.2 / 99.4 / 59.1	90.6 / 100.0 / 1.7	100.0 / 99.0 / 99.7	28.7 / 3.0 / 88.5	92.1 / 87.7 / 80.1	97.1 / 86.2 / 97.3	77.4 / 57.2 / 85.3	77.3 / 90.1 / 53.3	49.3 / 90.1 / 18.6	98.4 / 94.4 / 95.0
StarGAN	94.7 / 99.3 / 62.7	39.8 / 57.3 / 30.9	99.7 / 94.7 / 99.5	16.0 / 0.0 / 100.0	81.2 / 11.0 / 99.0	100.0 / 100.0 / 88.7	97.6 / 88.3 / 97.2	58.2 / 9.0 / 93.1	92.6 / 99.7 / 30.5	59.0 / 10.3 / 100.0
Diffusion										
SD1.4	56.0 / 99.2 / 1.1	91.2 / 99.8 / 8.2	71.7 / 98.1 / 6.0	81.8 / 21.8 / 99.8	96.4 / 82.3 / 96.5	83.0 / 84.0 / 58.9	94.4 / 72.8 / 96.6	99.9 / 96.6 / 100.0	99.9 / 98.7 / 100.0	94.2 / 94.1 / 82.3
SD1.5	55.9 / 99.1 / 0.9	91.0 / 99.8 / 8.0	71.5 / 98.2 / 5.6	82.1 / 21.7 / 99.8	96.2 / 81.5 / 96.4	83.4 / 84.4 / 59.5	94.2 / 71.6 / 96.8	99.8 / 96.6 / 99.9	99.9 / 98.7 / 99.8	94.6 / 94.6 / 82.8
ADM	70.0 / 99.1 / 3.9	86.2 / 99.7 / 6.8	92.5 / 98.2 / 45.9	79.9 / 21.8 / 94.3	62.6 / 82.6 / 36.3	76.8 / 84.7 / 39.6	83.0 / 73.0 / 78.2	90.8 / 97.0 / 64.6	91.4 / 98.9 / 46.8	89.4 / 93.8 / 64.1
GLIDE	72.6 / 98.9 / 5.7	96.5 / 99.8 / 18.6	86.1 / 98.2 / 27.4	97.5 / 21.2 / 100.0	95.4 / 81.5 / 95.2	67.0 / 84.0 / 23.7	90.0 / 71.4 / 90.5	89.5 / 97.1 / 61.5	90.4 / 98.8 / 46.5	99.0 / 93.9 / 98.0
Wukong	52.5 / 99.1 / 1.0	90.2 / 99.8 / 9.2	78.4 / 98.4 / 11.9	83.7 / 21.2 / 99.8	94.6 / 82.2 / 92.4	85.0 / 84.5 / 62.9	86.9 / 72.0 / 86.9	100.0 / 96.8 / 100.0	100.0 / 98.7 / 100.0	93.4 / 93.7 / 79.2
VQDM	73.3 / 99.1 / 4.7	98.9 / 99.7 / 61.8	96.2 / 98.2 / 69.3	62.6 / 21.4 / 85.9	84.7 / 82.0 / 68.9	92.0 / 84.7 / 82.0	87.7 / 73.8 / 86.2	99.3 / 96.6 / 96.1	99.7 / 98.6 / 95.6	91.7 / 94.0 / 76.1
DALLE2	44.5 / 99.6 / 0.4	95.3 / 100.0 / 3.5	63.9 / 98.8 / 1.3	97.6 / 30.2 / 99.7	91.3 / 84.2 / 81.8	47.7 / 71.2 / 16.9	87.9 / 38.9 / 98.6	96.1 / 99.4 / 67.5	98.8 / 100.0 / 65.2	98.4 / 99.1 / 91.1
Others										
Midjourney	49.3 / 99.2 / 1.6	68.3 / 99.9 / 1.0	47.3 / 98.1 / 1.5	84.9 / 21.2 / 99.0	85.5 / 81.7 / 72.2	61.1 / 83.8 / 18.1	82.1 / 72.3 / 76.3	100.0 / 96.9 / 100.0	99.2 / 98.9 / 94.6	95.4 / 94.1 / 84.5
WFIR	83.3 / 99.8 / 5.5	78.7 / 100.0 / 0.0	98.3 / 99.7 / 72.7	28.6 / 0.0 / 99.1	88.8 / 68.3 / 90.4	91.0 / 90.0 / 65.9	61.2 / 19.4 / 94.8	44.8 / 99.7 / 0.1	68.9 / 100.0 / 0.4	87.9 / 77.1 / 89.7
Mean	77.1 / 99.3 / 24.4	81.6 / 96.3 / 11.7	87.5 / 98.5 / 52.6	57.3 / 13.0 / 95.0	87.3 / 70.2 / 85.4	86.0 / 89.2 / 63.7	83.5 / 62.3 / 87.8	81.9 / 88.0 / 63.1	83.2 / 96.3 / 49.9	**89.3 / 77.3 / 89.3**

Table 11: **Cross-generator evaluation on LSUN-Bedroom (Yu et al., 2015).** AUC / AP across 10 generators in a shared semantic domain.

	GAN					**Diffusion**					
Detector \ Generator	ProGAN	ProjGAN	Diff-ProjGAN	StyleGAN	Diff-StyleGAN2	DDPM	IDDPM	ADM	LDM	PNDM	**Mean**
CNNSpot (Wang et al., 2020)	100.0 / 100.0	90.8 / 90.8	88.2 / 88.5	97.0 / 96.9	99.5 / 99.5	82.5 / 80.8	78.0 / 75.7	67.2 / 64.7	75.7 / 72.3	75.6 / 73.0	85.5 / 84.2
Corvi (Corvi et al., 2023)	89.7 / 90.0	80.6 / 75.9	86.5 / 84.5	89.9 / 88.9	78.7 / 73.1	84.4 / 79.1	84.8 / 79.5	76.3 / 71.5	100.0 / 100.0	77.9 / 75.5	84.9 / 81.8
AEROBLADE (Ricker et al., 2024)	38.9 / 42.3	45.7 / 47.3	46.1 / 48.0	40.0 / 42.9	56.7 / 56.6	66.7 / 64.4	59.1 / 56.3	53.6 / 51.3	65.4 / 63.7	43.4 / 45.3	51.5 / 51.8
ClipDet (Cozzolino et al., 2024)	98.2 / 98.2	97.1 / 96.7	97.5 / 97.2	90.9 / 89.9	95.7 / 94.7	83.5 / 79.2	74.0 / 67.7	60.8 / 58.2	81.6 / 78.0	88.8 / 86.8	86.8 / 84.7
C2P-CLIP (Tan et al., 2025)	97.6 / 97.7	95.7 / 94.6	95.1 / 94.0	90.1 / 89.7	94.7 / 93.6	71.3 / 72.2	66.8 / 68.4	66.9 / 66.7	94.4 / 93.5	96.9 / 97.5	86.9 / 86.8
DEAR-c (Kim et al., 2026)	83.8 / 84.8	78.8 / 74.3	86.2 / 85.0	78.7 / 77.5	75.7 / 72.6	77.4 / 72.1	74.1 / 70.1	75.0 / 71.5	99.9 / 99.9	87.1 / 86.9	81.7 / 79.5
DEAR-r (Kim et al., 2026)	60.9 / 62.2	58.5 / 56.0	67.2 / 66.0	57.4 / 58.3	68.4 / 66.2	60.2 / 54.4	62.4 / 57.6	67.9 / 63.2	98.6 / 98.7	48.7 / 47.9	65.0 / 63.1
GeoCue	97.2 / 96.8	95.2 / 94.5	96.0 / 95.4	78.4 / 71.7	94.1 / 92.8	88.8 / 84.8	80.3 / 73.8	71.8 / 67.1	85.2 / 82.5	89.1 / 86.8	87.6 / 84.6

of the underlying detection cues. In addition to AUC, we report AP to provide a complementary evaluation of detection performance in terms of precision and recall. GeoCue achieves the highest mean AUC while maintaining competitive AP across the 10 unseen generators. These results further support geometric coherence as useful detection evidence even when real and AI-generated images are drawn from the same semantic domain.

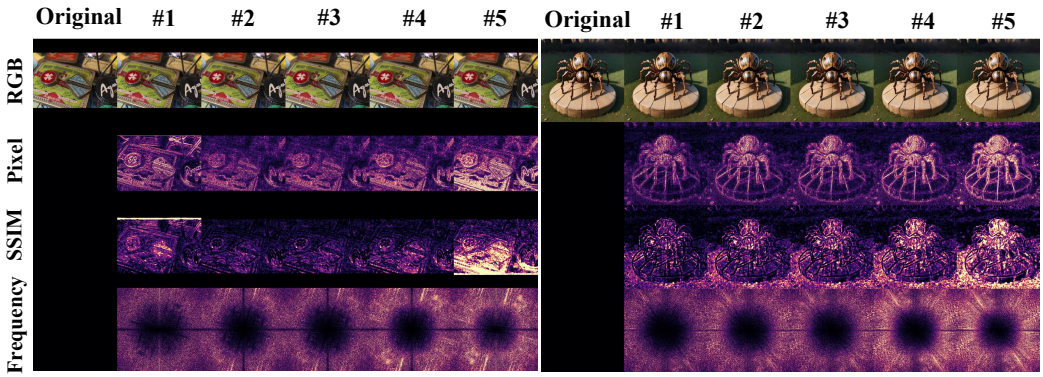


Figure 8: **Changes in image-domain cues under repeated online dissemination.** RGB inputs and pixel, local SSIM, and FFT-magnitude differences relative to the original image across successive dissemination rounds.

Table 12: **Sample-level complementarity between geometric cues.** We compare the point-normal consistency-only and deviation-only models. Each entry reports the percentage of samples correctly classified by both models, by only one model, or by neither model.

	Sample-wise Correctness (%)			
Benchmark	Both correct	Consistency-only	Deviation-only	Neither
Chameleon (Yan et al., 2025)	67.3	7.0	12.8	12.9
WildRF (Cavia et al., 2024)	72.8	10.0	6.4	10.7
LOKI (Ye et al., 2025)	52.2	19.4	12.4	16.0
Overall	**66.7**	**8.1**	**12.3**	**12.9**

C Additional Analysis

Changes in Image-Domain Cues under Repeated Online Dissemination. Fig. 8 visualizes how representative image-domain cues change under repeated online dissemination. Although the RGB appearance remains visually similar, pixel-level, local structural-similarity, and frequency-domain differences relative to the original image become noticeable across successive rounds. For each example and cue type, the same visualization range is used across all rounds, enabling direct comparison across dissemination rounds. These examples illustrate the sensitivity of image-domain cues to repeated resizing and re-encoding.

Sample-Level Complementarity of Geometric Cues. Table 12 further examines the complementarity between point-normal consistency and reference-relative deviation at the sample level. Across all three in-the-wild benchmarks (Yan et al., 2025; Ye et al., 2025; Cavia et al., 2024), each cue correctly classifies samples that the other cue misses. Moreover, which cue provides more unique correct predictions varies across benchmarks. These results further support the complementary roles of the two geometric cues in GeoCue.

D Additional Online Dissemination Results

Extended Baseline Comparison. Fig. 9 extends the online dissemination comparison by including a broader set of baseline detectors. Across all three dissemination sequences, GeoCue maintains relatively stable AUC over successive rounds on both Chameleon and AIGCDetectBench, while many baselines exhibit larger performance degradation or fluctuations.

Qualitative Stability of Geometric Responses. Fig. 10 provides additional qualitative examples of the geometric responses shown in the main paper. Across diverse real and AI-generated images, the characteristic point-normal consistency and reference-relative depth-deviation patterns remain

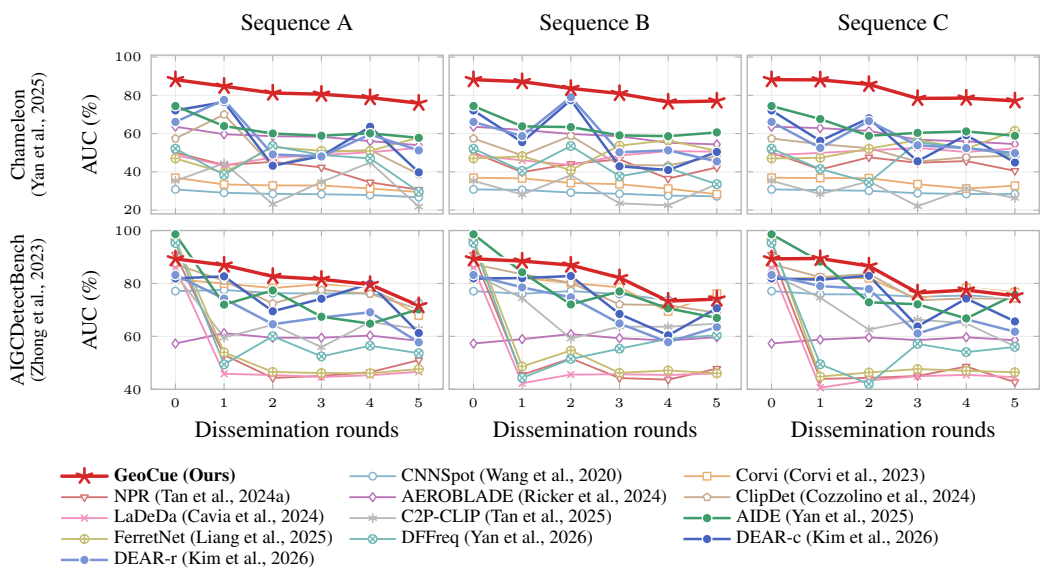


Figure 9: **Extended comparison under repeated online dissemination.** AUC across three dissemination sequences on Chameleon and AIGCDetectBench with additional baselines. Round 0 denotes the original images.

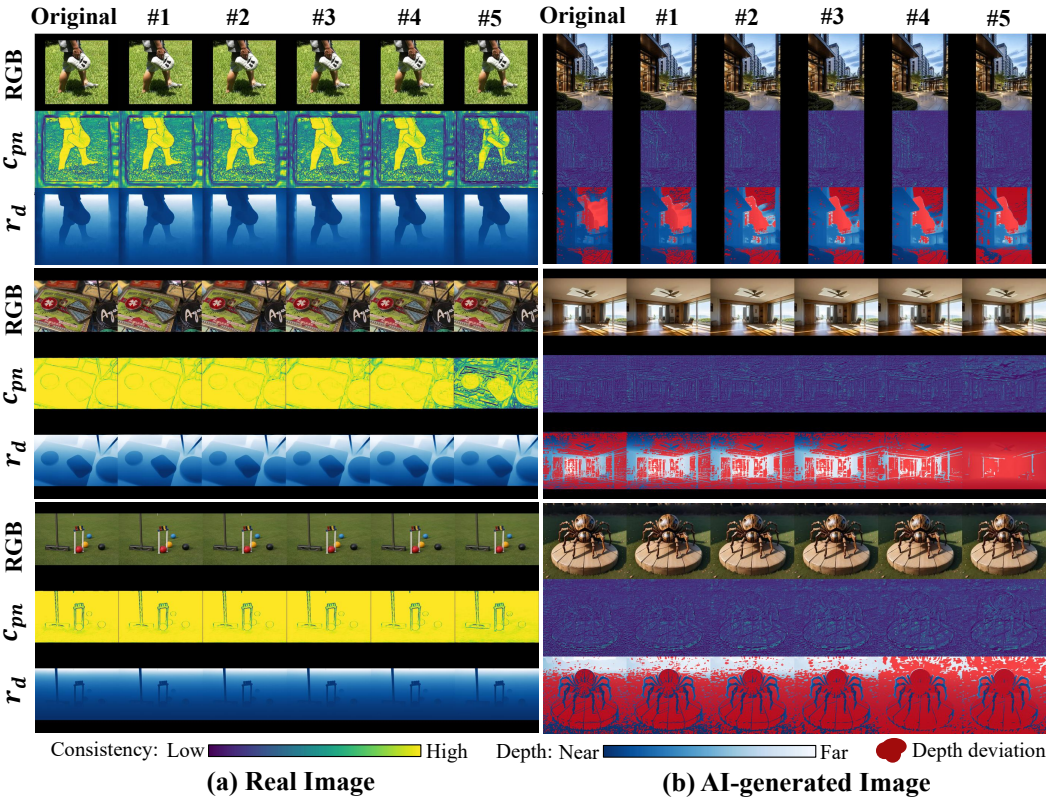


Figure 10: **Geometric cue stability under repeated online dissemination.** Geometric cues remain largely preserved across dissemination rounds.

largely preserved over repeated online dissemination. These examples further illustrate the stability of the geometric cues across different image content and dissemination rounds.