EPIMIRROR: EPIPOLAR CORRESPONDENCE MIRRORED PERTURBATIONS AGAINST STEREO MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

To understand the robustness of stereo-based 3D perception, adversarial perturbations against stereo matching have been increasingly studied. However, existing attacks do not exploit any specific vulnerability in this task, but simply follow conventional attacks (originally developed against image classification) that target final prediction errors or internal feature discrepancies. In this paper, we address this limitation by instead targeting epipolar structure, the task-specific geometric constraint on stereo correspondence. Specifically, we propose ***EpiMirror***, a new attack that constructs **Mirror**ed perturbations along the **Epi**polar line, i.e., perturbations with pixel-wise opposite directions (but identical amplitude) for left vs. right views. In this way, ***EpiMirror*** explicitly disrupts cross-view matching, subsequently promoting disparities along the same epipolar line, and ultimately leading to erroneous depth estimation. Experimental results across six stereo architectures and four datasets show that ***EpiMirror*** improves the D1-error of existing white-box, black-box transfer-based, and black-box query-based attacks by (relative) 16.9%, 20.2%, and 79.4% on average. We further provide a theoretical analysis attributing the success of mirrored perturbations to amplification of correspondence mismatch and collapse of matching margin.

1 INTRODUCTION

Stereo matching (Scharstein & Szeliski, 2002) is a fundamental task in 3D vision that recovers dense depth by establishing correspondences between two rectified views. Despite their strong performance and wide applications, stereo networks remain vulnerable under diverse adversarial settings. To understand the robustness of stereo-based 3D perception, different types of adversarial perturbations against stereo matching have been studied, including prediction-level perturbations (Wong et al., 2021; Berger et al., 2022), stereo-specific geometric constraints (Cheng et al., 2022a), feature-level discrepancy attacks (Wang et al., 2024), and structured or physically realizable attacks (Xing et al., 2025; Liu et al., 2025; Zhao et al., 2026).

However, we identify that existing attacks do not exploit any specific vulnerability in this task, but simply follow conventional attacks (originally developed against image classification) that target final prediction errors or internal feature discrepancies. In this paper, we address this limitation by instead targeting epipolar structure, the task-specific geometric constraint on stereo correspondence. The success of stereo matching largely relies on such epipolar geometry, which constrains candidate matches to the same epipolar line and reduces correspondence search to one dimension. This geometric principle remains central to modern stereo architectures, including cost-volume, recurrent, and geometry-encoding methods (Chang & Chen, 2018; Lipson et al., 2021; Xu et al., 2023; Guo et al., 2025).

Specifically, we propose ***EpiMirror***, a new attack that constructs **Mirror**ed perturbations along the **Epi**polar line, i.e., perturbations with pixel-wise opposite directions (but identical amplitude) for left vs. right views. In this way, ***EpiMirror*** explicitly disrupts cross-view matching, subsequently promoting disparities along the same epipolar line, and ultimately leading to erroneous depth estimation. Our key observation is that the same correspondence structure that reduces ambiguity for stereo matching can also provide structure for adversarial perturbations. Given a perturbation proposal in one view, stereo correspondence directly identifies where its paired perturbation can be placed in the other view. Correspondence therefore provides a natural mechanism for organizing

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.

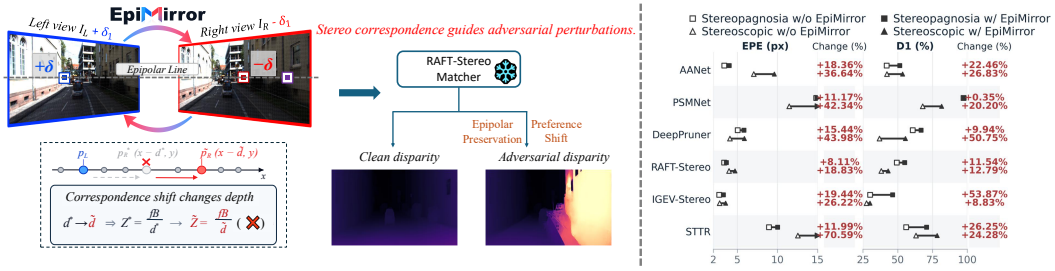


Figure 1: **Left:** Our ***EpiMirror*** constructs **Mirror**ed perturbations along the **Epi**polar line, i.e., perturbations with pixel-wise opposite directions (but identical amplitude) for left vs. right views. In this way, it explicitly disrupts cross-view matching, subsequently promoting disparities along the same epipolar line, and ultimately leading to erroneous depth estimation. **Right:** ***EpiMirror*** consistently improves existing attacks across diverse stereo architectures in both End-Point Error (EPE) and bad-pixel error (D1).

perturbations across the stereo pair. The remaining question is how the paired perturbations should interact. We find that coupling corresponding locations with opposite signs provides an effective adversarial direction, which we call *correspondence mirroring*. This coupling follows the epipolar correspondence structure while weakening the cross-view agreement that supports the clean match, thereby reducing its matching preference and competing disparities along the same epipolar line. Figure 1 illustrates this central mechanism and performance comparison.

The correspondence field serves only to organize cross-view perturbations and *does not require* direct manipulation of intermediate victim representations. We further analyze when opposite-sign coupling amplifies cross-view mismatch and characterize when the resulting reduction in the clean matching margin can change the preferred disparity. In this sense, ***EpiMirror*** preserves where stereo is geometrically allowed to match, but changes what stereo prefers to match.

We evaluate ***EpiMirror*** across six stereo architectures and four datasets, showing that ***EpiMirror*** improves the D1-error of existing white-box, black-box transfer-based, and black-box query-based attacks by (relative) 16.9%, 20.2%, and 79.4% on average. We further provide a theoretical analysis attributing the success of mirrored perturbations to amplification of correspondence mismatch and collapse of matching margin.

Our contributions are summarized as follows:

- **A geometry-specific adversarial perspective.** We identify epipolar structure as a natural, task-specific adversarial prior for structuring cross-view adversarial perturbations, revealing a vulnerability inherent in stereo correspondence.
- **Correspondence mirroring.** We introduce ***EpiMirror***, which couples perturbations at stereo correspondences between left and right views with opposite signs, separating *where* perturbations are paired from *how* they interact.
- **Mechanistic and empirical validation.** We connect opposite-sign coupling to correspondence mismatch amplification and matching-margin collapse, and validate the mechanism across architectures, datasets, white-box and black-box scenarios, as well as controlled ablations.

2 Related Work

Deep stereo matching. Modern stereo methods estimate disparity through learned cross-view matching, including cost volumes, correlation lookup, and iterative geometry encoding (Chang & Chen, 2018; Li et al., 2021; Lipson et al., 2021; Xu et al., 2023). Despite architectural differences, epipolar correspondence remains a shared geometric basis for establishing cross-view matches in modern stereo systems.

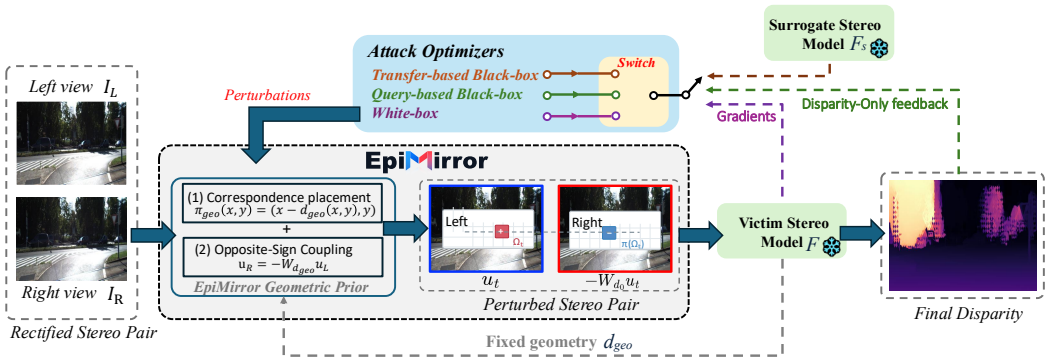


Figure 2: Framework of ***EpiMirror***. The idea of mirrored perturbations can be integrated into optimization in different attack scenarios with their corresponding feedback for perturbation updates.

Adversarial attacks on deep stereo matching. Adversarial attacks have been studied in monocular and stereo depth estimation, as well as multi-view 3D vision (Zhang et al., 2020; Yamanaka et al., 2020; Cheng et al., 2022b; Zheng et al., 2024; Jiang et al., 2024; Wang et al., 2026). Stereo networks are vulnerable to image-dependent, universal, structured, and physical perturbations (Wong et al., 2021; Berger et al., 2022; Xing et al., 2025; Liu et al., 2025; Zhao et al., 2026). Several works further incorporate stereo geometry: Cheng et al. (2022a) constructs photometrically consistent perturbations at corresponding locations, while (Wang et al., 2024) increases discrepancies between geometrically aligned intermediate features. In contrast, ***EpiMirror*** treats epipolar correspondence as a geometry-specific adversarial prior for structuring cross-view perturbations and couples corresponding locations with opposite signs to disrupt clean matching agreement.

3 Methodology

We first formulate the stereo attack setting and distinguish optimization feedback from cross-view perturbation structure (Section 3.1). We then introduce ***EpiMirror***, which constructs **Mirror**ed perturbations along the **Epi**polar line, i.e., perturbations with pixel-wise opposite directions (but identical amplitude) for left vs. right views (Section 3.2). We next describe its optimization in white-box, black-box transfer-based, and black-box query-based attack settings (Section 3.3). Finally, we provide a theoretical analysis attributing the success of mirrored perturbations to amplification of correspondence mismatch and collapse of matching margin (Section 3.4). Figure 2 illustrates the overall framework of ***EpiMirror***.

3.1 Problem Formulation

Given a rectified stereo pair $(I_L, I_R) \in \mathbb{R}^{3\times H\times W}$, a pretrained stereo matching network F estimates a dense disparity map $d = F(I_L, I_R) \in \mathbb{R}^{H\times W}$. We consider an untargeted full-image attack that seeks additive perturbations (δ_L, δ_R) to the left and right views, respectively, subject to a per-view ℓ_∞ budget:

$$\max_{\|\delta_L\|_\infty, \|\delta_R\|_\infty \le \epsilon} S(d_{\text{adv}}, d_0; M), \qquad d_{\text{adv}} = F(\mathcal{T}(I_L + \delta_L), \mathcal{T}(I_R + \delta_R)), \tag{1}$$

where $d_0 = F(I_L, I_R)$ is the frozen clean disparity prediction, $\mathcal{T}$ denotes clipping to $[0, 1]$ followed by input quantization, and M is a scoring region fixed before optimization. ***EpiMirror*** imposes a geometry-specific cross-view prior based on stereo correspondence and opposite-sign coupling, while remaining independent of the particular optimization mechanism. This decoupling allows the same correspondence-mirroring principle to be used across different attack scenarios.

3.2 EpiMirror: Correspondence Mirroring

Motivation. For a rectified stereo pair, the candidate matches of a left-view pixel $p = (x, y)$ are constrained to the same horizontal epipolar line (Scharstein & Szeliski, 2002). Let $C_0(p, d)$ denote

the clean matching cost for disparity candidate d, and let $d^\star(p)$ denote the preferred clean disparity:

$$d^\star(p) \approx \arg\min_{d \in \mathcal{D}(p)} C_0(p, d), \qquad \pi_{d^\star}(p) = \big(x - d^\star(p), y\big). \tag{2}$$

The clean prediction is therefore supported by cross-view agreement between p and its preferred correspondence $\pi_{d^\star}(p)$. We measure how much the preferred clean correspondence is weakened by

$$\Delta C_{\text{corr}}(p) = C_{\text{adv}}(p, d^\star(p)) - C_0(p, d^\star(p)), \tag{3}$$

where $\Delta C_{\text{corr}}(p) > 0$ means that the original clean match becomes less favorable after perturbation. This motivates ***EpiMirror*** to use stereo correspondence to decide *where* perturbations are paired across the two views, and their relative sign to decide *how* the paired perturbations interact. Further details on the matching-cost formulation are provided in Appendix Section A.1.

Correspondence mirroring. Let d_{geo} denote a fixed geometry field used to establish cross-view correspondences. When the victim clean prediction is available, we set $d_{\text{geo}} = \texttt{stopgrad}(d_0)$, where $d_0 = F(I_L, I_R)$; under transfer-based access, a surrogate disparity prediction can be used instead. The geometry field remains fixed during optimization and is used only for correspondence construction. For a left-view pixel $p = (x, y)$, we define

$$\pi_{d_{\text{geo}}}(p) = \big(x - d_{\text{geo}}(p), y\big). \tag{4}$$

Let $\mathcal{W}_{d_{\text{geo}}}$ denote the corresponding left-to-right rendering operator that maps a perturbation proposal to its estimated right-view location. Given a perturbation candidate u, ***EpiMirror*** constructs the correspondence-mirrored pair

$$\mathcal{U}_{\text{mirror}}(u) = \big(u, -\mathcal{W}_{d_{\text{geo}}} u\big). \tag{5}$$

Hence, d_{geo} determines the cross-view placement of the paired perturbations, while the negative sign imposes opposite-sign coupling. ***EpiMirror*** specifies this cross-view structure independently of how the proposal u is optimized; its optimization in different attack scenarios are described in Section 3.3.

3.3 Optimizations in Different Attack Scenarios

White-box. Victim gradients provide the optimization direction for the perturbation candidate u. Rather than modifying the underlying gradient-based objective, ***EpiMirror*** uses stereo correspondence to guide how this candidate is distributed across the two views: the directions of perturbations for the left and right views are opposite at its estimated right-view correspondence according to Equation (5). In this way, the gradient determines *what* direction to optimize, while ***EpiMirror*** imposes a stereo-specific cross-view structure on the resulting update.

Black-box transfer-based. The perturbation optimization follows the same way as the white-box scenario, with the only difference being the gradients from a surrogate model rather than the (unseen) victim model. Once the perturbed views are optimized, they are directly applied to an victim model to get the results.

Black-box query-based. Only the final output of the victim model is available as the feedback for perturbation update. We perform a Square-style `accept-reject` search (Andriushchenko et al., 2020) with our mirrored perturbation constraint. At each iteration, a local support and bounded candidate are generated in the left view, and the bounded candidate is coupled to its estimated right-view correspondence according to Equation (5). The resulting stereo candidate is projected onto the perturbation budget, queried on the victim, and retained if it improves the attack.

3.4 Theoretical Analysis

We next analyze how the sign choice in correspondence coupling affects cross-view mismatch and how the resulting mismatch can alter the preferred disparity.

Why opposite-sign coupling? Consider a clean correspondence (p, p_R) for which the fixed geometry field identifies the preferred clean match:

$$d_{\text{geo}}(p) = d^\star(p), \qquad p_R = \pi_{d^\star}(p) = \pi_{d_{\text{geo}}}(p). \tag{6}$$

For a fixed perturbation direction u, consider the correspondence-conditioned coupling family

$$(\delta_L, \delta_R) = \left(u, \alpha \mathcal{W}_{d_{\text{geo}}} u\right), \qquad \alpha \in [-1, 1]. \tag{7}$$

Here, $\alpha = -1$ corresponds to the *opposite-sign coupling* used by ***EpiMirror***, $\alpha = +1$ to *same-sign coupling*, and $\alpha = 0$ to a *left-view-only perturbation* within this family.

Let ϕ_L and ϕ_R denote differentiable feature representations of the left and right views, respectively. These representations are introduced only for theoretical analysis and are not accessed by the attack itself. Let J_L and J_R denote their local Jacobians at the corresponding locations, with the fixed correspondence mapping absorbed into J_R. We define the clean cross-view feature residual at the correspondence as

$$r = \phi_L(I_L, p) - \phi_R(I_R, p_R). \tag{8}$$

For a perturbation direction u, let

$$a = J_L u, \qquad b = J_R u, \tag{9}$$

denote the corresponding first-order feature responses in the left and right views. Under a first-order approximation, the cross-view residual for coupling coefficient α becomes

$$r_\alpha \approx r + a - \alpha b, \tag{10}$$

yielding the linearized mismatch

$$M(\alpha) = \|r + a - \alpha b\|_2^2. \tag{11}$$

Proposition 1 (Opposite-sign mismatch amplification). The endpoint mismatch difference satisfies

$$M(-1) - M(+1) = 4\langle r + a, b\rangle. \tag{12}$$

Since $M(\alpha)$ is convex in α, its maximum over $\alpha \in [-1, 1]$ occurs at an endpoint. Therefore, if

$$\langle r + a, b\rangle \geq 0, \tag{13}$$

the opposite-sign endpoint $\alpha = -1$ is a maximizer of the local linearized mismatch within this coupling family. Intuitively, when the right-view response b is nonnegatively aligned with the residual-plus-left response $r + a$, opposite-sign coupling increases the cross-view mismatch relative to same-sign coupling. This is a local, fixed-direction result and does not imply global optimality of the attack.

For empirical analysis, we normalize the alignment term as

$$A(p, u) = \frac{\langle r + a, b\rangle}{\|r + a\|_2 \|b\|_2}. \tag{14}$$

For nonzero $\|r + a\|_2$ and $\|b\|_2$, $A \in [-1, 1]$, and $A \geq 0$ is equivalent to the condition in Equation (13). The complete first-order derivation is provided in Appendix Section A.2, A.3.

From mismatch to disparity shift. A larger correspondence mismatch does not necessarily change the preferred disparity. Let $C_0(p, d)$ denote the clean effective matching cost and assume that $d^\star$ is the unique preferred clean disparity. For any competing disparity $d \neq d^\star$, define the clean matching margin

$$m_d = C_0(p, d) - C_0(p, d^\star) > 0. \tag{15}$$

Let $\Delta C(p, d)$ denote the attack-induced change in matching cost:

$$C_{\text{adv}}(p, d) = C_0(p, d) + \Delta C(p, d). \tag{16}$$

Proposition 2 (Matching-margin condition). A competing disparity d' overtakes $d^\star$ if and only if

$$\Delta C(p, d^\star) - \Delta C(p, d') > m_{d'}. \tag{17}$$

Thus, disrupting the preferred correspondence is insufficient by itself: its matching cost must deteriorate relative to at least one competitor by more than the original matching margin. A complete derivation and a sufficient stability condition are provided in Appendix Section A.4.

Together, these results characterize the local mechanism of correspondence mirroring: under the stated alignment condition, opposite-sign coupling can maximize the local linearized mismatch within the correspondence-conditioned coupling family, while changing the preferred disparity additionally requires the induced cost change to overcome the clean matching margin.

4 EXPERIMENTS

In this section, we compare our ***EpiMirror*** to attack baselines in both white-box and black-box (transfer-based and query-based) scenarios. We also conduct ablation studies on the core components to validate their individual contributions.

4.1 EXPERIMENTAL SETUP

Datasets and victim models. We evaluate six stereo matching models using their official implementations: AANet (Xu & Zhang, 2020), DeepPruner (Duggal et al., 2019), PSMNet (Chang & Chen, 2018), RAFT-Stereo (Lipson et al., 2021), IGEV-Stereo (Xu et al., 2023), and STTR (Li et al., 2021). In white-box and black-box transfer-based experiments, all these models are evaluated; in black-box query-based experiments, only RAFT-Stereo, IGEV-Stereo, and STTR are evaluated due to the high computational cost of query search.

Following (Xing et al., 2025), we use 40 image pairs from KITTI 2015 (Menze & Geiger, 2015), 160 pairs from DrivingStereo (Yang et al., 2019), with 40 pairs each under sunny, cloudy, foggy, and rainy conditions. In addition, we also use 15 labeled pairs from Middlebury spanning 12 scene groups (Scharstein et al., 2014) and 27 labeled pairs from ETH3D (Schops et al., 2017). All experiments were conducted on a single NVIDIA A40 GPU.

Evaluation metrics. We report End-Point Error (EPE) and bad-pixel error (D1):

$$\mathrm{D1} = \frac{N_{\mathrm{bad}}}{N_{\mathrm{valid}}} \times 100\%, \qquad \mathrm{EPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{d}_i - d_i^*\right|. \tag{18}$$

A pixel is considered bad if $|\hat{d}_i - d_i^*| > 3$ px and $|\hat{d}_i - d_i^*|/|d_i^*| > 5\%$, where $\hat{d}_i$ and d_i^* denote the predicted and ground-truth disparities. Both EPE and D1 are computed per stereo pair and averaged over all evaluated pairs; higher values indicate stronger degradation.

Attack baselines and parameters. Unless otherwise specified, all attacks use a per-view ℓ_∞ perturbation budget of $\epsilon = 4/255$. For ***EpiMirror***, correspondences are computed from the frozen clean prediction, i.e., $d_{\mathrm{geo}} = \mathrm{stopgrad}(d_0)$, and remain fixed throughout optimization. We use the same clean-reference score parameters across all datasets and victim models.

For the white-box and black-box transfer-based scenarios, we compare ***EpiMirror*** to the FGSM variant of Stereopagnosia (Wong et al., 2021) and Stereoscopic Universal Perturbations (SUP) (Berger et al., 2022). Stereopagnosia perturbs stereo inputs using loss gradients, while the original SUP method learns perturbations shared across image pairs.

For the black-box query-based scenarios, we compare ***EpiMirror*** to *Random Search* (Solis & Wets, 1981), *NES* (Ilyas et al., 2018), *Bandits-TD* (Ilyas et al., 2019), *Square* (Andriushchenko et al., 2020), and *SimBA* (Guo et al., 2019). The maximum query budget is fixed to 600 per stereo pair.

4.2 WHITE-BOX EFFECTIVENESS

We first evaluate the proposed correspondence-mirroring prior under white-box access to determine whether it strengthens existing gradient-based attacks. The prior couples perturbations at estimated stereo correspondences with opposite signs. As shown in Table 1, ***EpiMirror*** increases both EPE and D1 across all 18 attack–victim combinations. For Stereopagnosia, the mean per-model relative gains in EPE and D1 are 14.1% and 20.7%, respectively; for Stereoscopic, they are 39.7% and 23.9%. For patch-based Depthvanish, EPE increases by 8.75-26.47 pixels across the six victim models. As visualized in Figure 3, adding ***EpiMirror*** to either attack produces broader and stronger disparity degradation. These results show that correspondence-aware opposite-sign coupling is an effective adversarial prior across attack families and architectures. Additional white-box results on ETH3D and DrivingStereo are reported in Appendix Section B.1.

Table 1: **White-box full-image and patch-based attacks comparisons with *EpiMirror* on KITTI.** Perturbations are optimized and evaluated on the same model.

Attack on KITTI		AANet		PSMNet		DeepPruner		RAFT-Stereo		IGEV-Stereo		STTR	
		D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑
	Clean	1.38	0.44	3.85	1.74	1.16	0.48	3.59	0.67	3.28	0.66	12.24	2.13
Full image	Stereopagnosia	32.17	3.41	72.33	14.71	48.76	5.08	40.24	3.29	27.97	2.75	56.21	8.87
	Stereoscopic	42.19	7.03	67.88	11.46	36.81	4.08	38.15	3.96	27.51	2.79	63.05	12.53
	EpiMirror	**53.51**	**9.61**	**81.59**	**16.32**	**55.49**	**5.87**	**43.03**	**4.70**	**29.94**	**3.52**	**78.36**	**21.38**
Patch based	Depthvanish	66.42	56.54	55.30	50.71	97.07	67.19	89.31	66.01	98.59	62.73	92.38	69.25
	EpiMirror	**72.44**	**71.95**	**67.69**	**63.81**	**97.73**	**86.62**	**90.33**	**74.76**	**98.78**	**77.02**	**95.23**	**95.72**

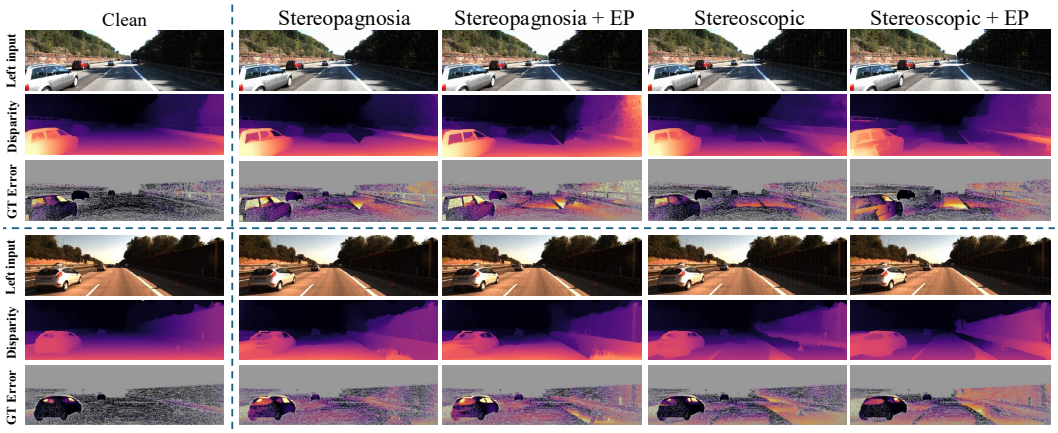


Figure 3: **Visualization of White-box attacks.**

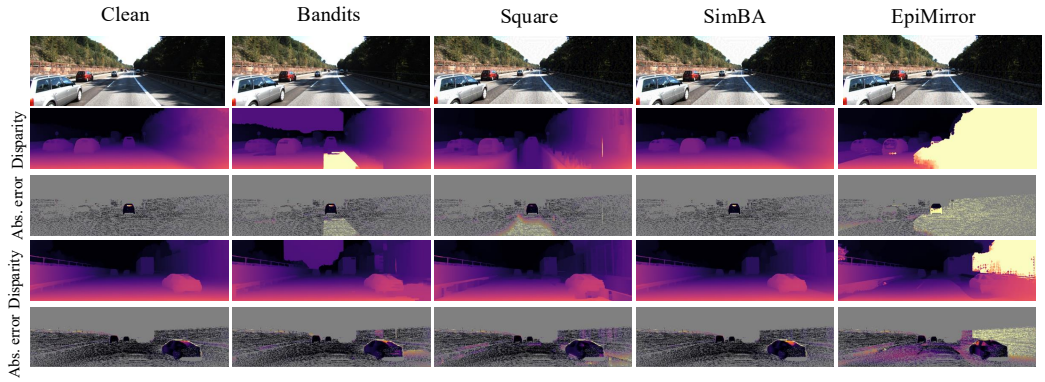


Figure 4: **Visualization of Black-box attacks.**

4.3 Black-box Transfer-based Effectiveness

We evaluate strict cross-model transferability by directly applying adversarial stereo pairs generated on a source model to unseen target models. The target model does not participate in perturbation generation and provides no optimization feedback. For each source model, the same adversarial stereo pairs are reused across different targets without any modification or re-optimization. As shown in Table 2, ***EpiMirror*** perturbations remain effective across different stereo architectures. For IGEV→RAFT, D1 increases from 6.03% to 28.21% and EPE from 1.22 to 9.13 px. Transfers to STTR are also substantial: perturbations generated on RAFT-Stereo and IGEV-Stereo raise D1 from

Table 2: **Black-box transfer-based full-image and patch-based attack comparisons with *EpiMirror* on KITTI.** Results for each target model are averaged over the other five source models.

	Attack on KITTI	AANet		PSMNet		DeepPruner		RAFT-Stereo		IGEV-Stereo		STTR	
		D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑	D1 ↑	EPE ↑
Full image	Stereopagnosia	22.22	3.47	27.03	3.26	24.07	3.64	27.17	3.23	16.74	2.75	52.89	12.28
	Stereoscopic	16.81	2.13	25.49	3.43	24.00	3.67	22.78	2.63	18.73	2.45	47.06	8.47
	EpiMirror	**37.68**	**4.62**	**38.73**	**4.85**	**41.76**	**6.01**	**36.05**	**7.20**	**25.88**	**6.92**	**53.78**	**14.44**
Patch based	DepthVanish	20.35	4.18	29.39	4.85	25.83	6.01	21.86	**7.93**	1.00	1.77	13.53	3.02
	EpiMirror	**25.15**	**4.62**	**66.25**	**19.68**	**66.60**	**18.57**	**27.83**	7.20	**16.89**	**6.92**	**50.64**	**14.44**

25.61% to 49.96% and 49.45%, respectively. Even without any target-side optimization or feedback, all transfer directions substantially degrade the target predictions, indicating that correspondence-mirrored perturbations capture adversarial structure that transfers across stereo architectures.

Table 3: **Black-box query-based attacks on KITTI and Middlebury.**

KITTI				Middlebury			
Method	**RAFT-Stereo**	**IGEV-Stereo**	**STTR**	**Method**	**RAFT-Stereo**	**IGEV-Stereo**	**STTR**
Attack	D1 / EPE	D1 / EPE	D1 / EPE	**Attack**	D1 / EPE	D1 / EPE	D1 / EPE
Clean	6.03 / 1.22	5.68 / 1.24	25.61 / 5.97	Clean	6.85 / 1.11	5.26 / 0.90	19.96 / 4.76
Random Search	6.23 / 1.23	7.45 / 1.39	31.52 / 7.88	Random Search	10.02 / 1.63	10.83 / 1.79	29.94 / 6.56
NES	6.29 / 1.21	6.70 / 1.35	29.64 / 7.46	NES	9.35 / 1.33	10.12 / 1.75	28.21 / 6.20
Bandits	13.59 / 2.58	10.30 / 2.42	46.72 / 15.75	Bandits	20.77 / 3.57	18.23 / 3.19	44.74 / 10.32
Square	13.50 / 2.03	14.80 / 2.32	42.33 / 13.34	Square	16.30 / 2.62	17.61 / 3.02	40.71 / 9.22
SimBA	7.19 / 1.66	6.78 / 1.74	46.62 / **28.81**	SimBA	12.60 / 2.26	11.39 / 3.14	41.48 / **18.68**
EpiMirror	**38.57 / 27.36**	**24.50 / 15.00**	**48.21** / 14.42	*EpiMirror*	**52.59 / 44.30**	**29.67 / 10.38**	**48.29** / 10.33

4.4 Black-box Query-based Effectiveness

As shown in Table 3, ***EpiMirror*** consistently outperforms the query-based baselines in almost all cases. Note that the perturbation initialization in ***EpiMirror*** adopts simple random sampling. EpiMirror achieves the highest D1 across all six victim–dataset settings and the highest EPE on RAFT-Stereo and IGEV-Stereo for both KITTI and Middlebury. On STTR, EpiMirror also achieves the highest D1, reaching 48.21% on KITTI and 48.29% on Middlebury, while its EPE is lower than SimBA. These results indicate that EpiMirror consistently induces a large fraction of bad pixels, although the magnitude of disparity errors can depend on the victim architecture and attack method. Figure 4 visualizes the predicted disparity and absolute error under different query-based attacks, where EpiMirror produces visibly broader errors than the baselines.

A few small regressions are observed, mainly on IGEV-Stereo, indicating that the gain can depend on the optimizer–victim combination. Overall, the consistent improvements across different query-based attacks suggest that correspondence mirroring acts as a reusable geometry-aware search prior rather than being tied to a particular black-box optimizer.

4.5 Ablation Studies

Correspondence and Sign Ablation. As shown in Table 4, we ablate the two key components of ***EpiMirror***: correspondence-aware placement and opposite-sign coupling. **Same** retains stereo correspondences but uses same-sign coupling; **Shuffled** retains opposite signs but randomizes correspondence assignments; **Pixel** applies opposite-sign perturbations at identical image coordinates, ignoring disparity; and **Independent** removes cross-view coupling altogether. Comparing ***EpiMirror*** with **Same** isolates the effect of sign coupling, while **Shuffled** and **Pixel** assess correspondence-aware placement. Across the six model–dataset settings, ***EpiMirror*** achieves the highest EPE and D1 in five settings each, with exceptions on STTR/KITTI (EPE) and IGEV-Stereo/Middlebury (D1).

Table 4: **Ablation of correspondence placement and sign coupling.** Each entry reports EPE / D1 (%), where higher values indicate stronger attacks.

Victim	Dataset	Clean	Same	Shuffled	Pixel	Square	Independent	*EpiMirror*
		EPE / D1	EPE / D1	EPE / D1	EPE / D1	EPE / D1	EPE / D1	EPE / D1
RAFT	KITTI	1.22 / 5.84	1.31 / 7.04	5.78 / 16.37	9.27 / 19.37	1.89 / 13.26	5.43 / 22.10	**14.59 / 31.09**
	Middlebury	1.12 / 6.78	2.30 / 11.37	16.33 / 31.84	15.89 / 36.36	2.72 / 16.83	11.72 / 40.59	**43.16 / 51.09**
IGEV	KITTI	1.28 / 6.27	1.61 / 7.81	5.92 / 13.05	8.73 / 16.69	2.05 / 13.58	6.59 / 18.23	**10.65 / 18.30**
	Middlebury	0.92 / 5.39	2.25 / 11.17	5.01 / 19.64	8.58 / 25.53	3.19 / 18.62	6.53 / **31.02**	**10.94** / 30.39
STTR (SF)	KITTI	4.88 / 25.33	6.91 / 28.14	12.90 / 37.86	**13.41** / 39.69	11.57 / 40.14	12.60 / 40.38	13.27 / **44.63**
	Middlebury	4.88 / 19.90	6.02 / 23.09	8.63 / 33.47	9.49 / 39.79	9.63 / 41.70	8.91 / 39.88	**10.77 / 49.58**

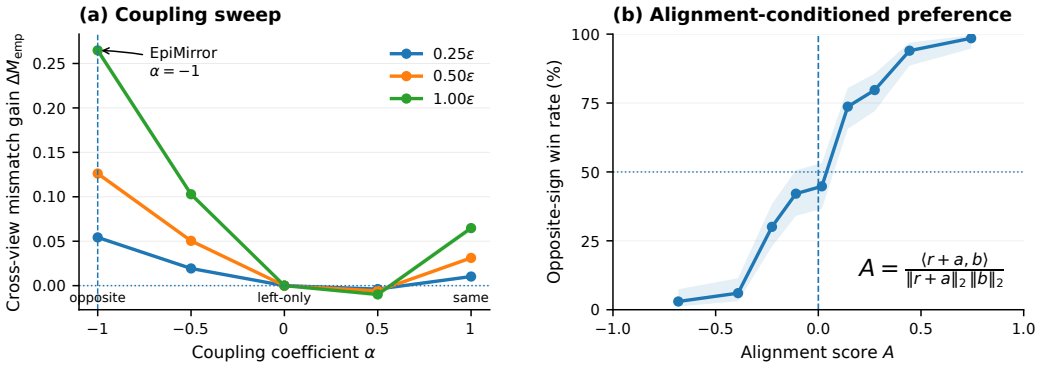


Figure 5: **Empirical validation of opposite-sign correspondence coupling.** (a) Cross-view mismatch under different coupling coefficients α. (b) Opposite-sign win rate versus the normalized alignment score A, with $A = 0$ denoting the Proposition 1 boundary.

Overall, degrading either correspondence placement or sign coupling reduces attack strength, indicating that the two components provide complementary contributions.

Mechanism Validation. We validate Proposition 1 by fixing the stereo correspondence and perturbation direction while varying only the coupling coefficient α, and examine whether the alignment score A predicts the advantage of opposite-sign coupling. Figure 5(a) isolates the effect of α within the correspondence-conditioned coupling family $(u, \alpha \mathcal{W}_{d_{\text{geo}}} u)$. Figure 5(b) further tests whether the theoretical alignment condition predicts when opposite-sign coupling yields a larger empirical mismatch.

5 Conclusion and Outlook

We presented ***EpiMirror***, a geometry-aware adversarial attack for stereo matching that uses stereo correspondence to structure cross-view perturbations and couples corresponding locations with opposite signs. Our analysis shows when this coupling amplifies correspondence mismatch and when the resulting cost change can alter the preferred disparity. Experiments across architectures, datasets, and access settings, together with controlled ablations and mechanism analyses, support the effectiveness of this design. Since ***EpiMirror*** only specifies how perturbations are structured across views, it can be combined with white-box, transfer-based, and query-based optimizers without changing their underlying objectives. These results suggest that stereo correspondence can serve not only as a geometric prior for prediction, but also as an adversarial prior for structuring attacks. Our theoretical analysis is local and first-order, and the correspondence field is fixed from the clean prediction; a fuller account of the iterative optimization dynamics is left for future work. Beyond two-view stereo, similar correspondence-based coupling may also apply to multi-view stereo and other 3D reconstruction systems that rely on cross-view matching. Finally, future work may explore defenses against adversarial perturbations that explicitly account for correspondence-based cross-view coupling.

AI USE STATEMENT

We used generative AI tools only for language polishing, grammar checking, translation assistance, LaTeX formatting, readability improvement, and consistency checks of terminology, equations, tables, figures, and references. We did not use generative AI tools to generate research ideas, design experiments, implement code, process data, produce proofs, conduct experiments, analyze results, or generate reported numbers. The authors reviewed all AI-assisted edits against the manuscript, source files, and experimental outputs, and take full responsibility for the accuracy, originality, and compliance of the final content.

REFERENCES

Maksym Andriushchenko, Francesco Croce, Nicolas Flammarion, and Matthias Hein. Square attack: A query-efficient black-box adversarial attack via random search. In *European Conference on Computer Vision*, pp. 484–501, 2020.

Zachary Berger, Parth Agrawal, Tian Yu Liu, Stefano Soatto, and Alex Wong. Stereoscopic universal perturbations across different architectures and datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15180–15190, 2022.

Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 5410–5418. IEEE, 2018.

Kelvin Cheng, Tianfu Wu, and Christopher Healey. Revisiting non-parametric matching cost volumes for robust and generalizable stereo matching. *Advances in Neural Information Processing Systems*, 35:16305–16318, 2022a.

Zhiyuan Cheng, James Liang, Hongjun Choi, Guanhong Tao, Zhiwen Cao, Dongfang Liu, and Xiangyu Zhang. Physical attack on monocular depth estimation with optimal adversarial patches. In *European conference on computer vision*, pp. 514–532. Springer, 2022b.

Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. Deeppruner: Learning efficient stereo matching via differentiable patchmatch. In *ICCV*, 2019.

Chuan Guo, Jacob Gardner, Yurong You, Andrew Gordon Wilson, and Kilian Weinberger. Simple black-box adversarial attacks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2484–2493. PMLR, 2019.

Xianda Guo, Chenming Zhang, Youmin Zhang, Wenzhao Zheng, Dujun Nie, Matteo Poggi, and Long Chen. LightStereo: Channel boost is all you need for efficient 2d cost aggregation. In *IEEE International Conference on Robotics and Automation*, 2025.

Andrew Ilyas, Logan Engstrom, Anish Athalye, and Jessy Lin. Black-box adversarial attacks with limited queries and information. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2137–2146. PMLR, 2018.

Andrew Ilyas, Logan Engstrom, and Aleksander Madry. Prior convictions: Black-box adversarial attacks with bandits and priors. In *International Conference on Learning Representations*, 2019.

Wenxiang Jiang, Hanwei Zhang, Xi Wang, Zhongwen Guo, and Hao Wang. Nerfail: Neural radiance fields-based multiview adversarial attack. In *AAAI*, 2024.

Zhaoshuo Li, Xingtong Liu, Nathan Drenkow, Andy Ding, Francis X Creighton, Russell H Taylor, and Mathias Unberath. Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6177–6186. IEEE, 2021.

Lahav Lipson, Zachary Teed, and Jia Deng. RAFT-Stereo: Multilevel recurrent field transforms for stereo matching. In *International Conference on 3D Vision*, 2021.

Hangcheng Liu, Xu Kuang, Xingshuo Han, Xingwan Wu, Haoran Ou, Shangwei Guo, Xingyi Huang, Tao Xiang, and Tianwei Zhang. Pixel-optimization-free patch attack on stereo depth estimation. *arXiv preprint arXiv:2506.17632*, 2025. URL https://arxiv.org/abs/2506.17632.

Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3061–3070, 2015.

Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1–3):7–42, 2002.

Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, pp. 31–42, 2014.

Thomas Schops, Johannes L Schonberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3260–3269, 2017.

Francisco J Solis and Roger J-B Wets. Minimization by random search techniques. *Mathematics of operations research*, 6(1):19–30, 1981.

Pengfei Wang, Xiaofei Hui, Beijia Lu, Nimrod Lilith, Jun Liu, and Sameer Alam. Left-right discrepancy for adversarial attack on stereo networks. *arXiv preprint arXiv:2401.07188*, 2024.

Weijie Wang, Songlong Xing, Zhengyu Zhao, Nicu Sebe, and Bruno Lepri. Poinit-of-view: Poisoning initialization of views transfers across multiple 3d reconstruction systems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20670–20679, June 2026.

Alex Wong, Mukund Mundhra, and Stefano Soatto. Stereopagnosia: Fooling stereo networks with adversarial perturbations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2879–2888, 2021.

Yun Xing, Yue Cao, Nhat Chung, Jie M. Zhang, Ivor Tsang, Ming-Ming Cheng, Yang Liu, Lei Ma, and Qing Guo. DepthVanish: Optimizing adversarial interval structures for stereo-depth-invisible patches. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Gangwei Xu, Xianqi Wang, Xiaohuan Ding, and Xin Yang. Iterative geometry encoding volume for stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21919–21928, 2023.

Haofei Xu and Juyong Zhang. Aanet: Adaptive aggregation network for efficient stereo matching. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1956–1965. Ieee, 2020.

Koichiro Yamanaka, Ryutaroh Matsumoto, Keita Takahashi, and Toshiaki Fujii. Adversarial patch attacks on monocular depth estimation networks. *IEEE Access*, 8:179094–179104, 2020.

Guorun Yang, Xiao Song, Chaoqin Huang, Zhidong Deng, Jianping Shi, and Bolei Zhou. Drivingstereo: A large-scale dataset for stereo matching in autonomous driving scenarios. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 899–908. IEEE, 2019.

Ziqi Zhang, Xinge Zhu, Yingwei Li, Xiangqun Chen, and Yao Guo. Adversarial attacks on monocular depth estimation. *arXiv preprint arXiv:2003.10315*, 2020.

Kangqiao Zhao, Shuo Huai, Xurui Song, and Jun Luo. Cheating stereo matching in full-scale: Physical adversarial attack against binocular depth estimation in autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 13190–13198, 2026. doi: 10.1609/aaai.v40i16.38320. URL https://ojs.aaai.org/index.php/AAAI/article/view/38320.

Junhao Zheng, Chenhao Lin, Jiahao Sun, Zhengyu Zhao, Qian Li, and Chao Shen. Physical 3d adversarial attacks against monocular depth estimation in autonomous driving. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24452–24461. IEEE, 2024.

A Derivation of Correspondence Mirroring

A.1 Effective Matching-Cost Formulation

The main paper uses $C_0(p, d)$ as an effective clean matching cost for candidate disparity d. Here we make this notation explicit.

Let $\phi_L(I_L, p)$ and $\phi_R(I_R, \pi_d(p))$ denote feature representations at a candidate correspondence. We write

$$C(p, d) = \rho(\phi_L(I_L, p), \phi_R(I_R, \pi_d(p))) , \tag{19}$$

where $\rho(\cdot, \cdot)$ measures cross-view discrepancy. A simple example is the squared feature distance

$$C(p, d) = \|\phi_L(I_L, p) - \phi_R(I_R, \pi_d(p))\|_2^2 . \tag{20}$$

We use C_0 and C_{adv} to denote the corresponding clean and perturbed costs. This effective-cost view is used only for theoretical analysis; the output-only attack does not access these features or matching costs.

A.2 First-order Correspondence Analysis

Under the local condition in Equation (6), consider the coupling family in Equation (7). A first-order expansion gives

$$\phi_L(I_L + u, p) \approx \phi_L(I_L, p) + J_L u, \tag{21}$$

$$\phi_R(I_R + \alpha \mathcal{W}_{d_{\mathrm{geo}}} u, p_R) \approx \phi_R(I_R, p_R) + \alpha J_R u. \tag{22}$$

Using the definitions of r, a, and b in Equations (8) and (9), the perturbed residual is

$$r_\alpha \approx [\phi_L(I_L, p) + a] - [\phi_R(I_R, p_R) + \alpha b] \tag{23}$$

$$= r + a - \alpha b, \tag{24}$$

which gives Equation (10). Therefore,

$$M(\alpha) = \|r + a - \alpha b\|_2^2, \tag{25}$$

as stated in Equation (11).

A.3 Opposite-sign Mismatch Amplification

At the two endpoints of the coupling family,

$$M(-1) = \|r + a + b\|_2^2, \tag{26}$$

$$M(+1) = \|r + a - b\|_2^2. \tag{27}$$

Their difference is

$$M(-1) - M(+1) = \|r + a + b\|_2^2 - \|r + a - b\|_2^2 \tag{28}$$

$$= 4\langle r + a, b\rangle, \tag{29}$$

which proves Equation (12).

More generally,

$$M(\alpha) = \|r + a\|_2^2 - 2\alpha\langle r + a, b\rangle + \alpha^2 \|b\|_2^2. \tag{30}$$

Since the coefficient of α^2 is nonnegative, $M(\alpha)$ is convex in α. Its maximum over the closed interval $[-1, 1]$ therefore occurs at an endpoint. When

$$\langle r + a, b\rangle \geq 0, \tag{31}$$

we have

$$M(-1) \geq M(+1), \tag{32}$$

which proves Equation (13).

The normalized alignment score in Equation (14) has the same sign as $\langle r + a, b\rangle$ whenever its denominator is nonzero.

Table 5: **White-box full-image and patch-based attacks on ETH3D and DrivingStereo.** Each entry reports D1 (%) / EPE (px). Higher values indicate stronger attacks.

ETH3D

Method	**RAFT-Stereo**	**IGEV-Stereo**	**STTR**
Attack	D1 / EPE	D1 / EPE	D1 / EPE
Clean	0.72 / 0.27	0.84 / 0.32	6.31 / 1.33
Stereopagnosia	44.80 / 2.15	28.70 / **1.95**	52.40 / 6.72
Stereoscopic	35.60 / 2.62	25.90 / 2.06	58.20 / 9.48
EpiMirror	**51.00 / 2.48**	**43.20** / 1.90	**66.80 / 7.81**
DepthVanish	84.60 / 42.80	**94.30** / 47.50	89.70 / 55.10
EpiMirror	**87.10 / 51.90**	93.90 / **59.60**	**93.00 / 73.40**

DrivingStereo

Method	**Sunny**	**Cloudy**	**Foggy**	**Rainy**
Attack	D1 / EPE	D1 / EPE	D1 / EPE	D1 / EPE
Clean	11.27 / 2.35	13.44 / 2.35	11.28 / 1.96	8.64 / 2.48
Stereopagnosia	52.80 / 4.20	56.40 / 4.80	54.20 / 4.50	58.60 / 5.10
Stereoscopic	**43.70** / 5.10	48.50 / 5.80	46.10 / 5.40	50.20 / 6.30
EpiMirror	**58.90 / 4.55**	**62.10 / 5.15**	**60.00 / 4.82**	**64.90 / 5.52**
DepthVanish	90.40 / 58.30	92.10 / 64.70	91.30 / **61.20**	94.00 / 72.60
EpiMirror	**91.80 / 68.10**	**93.20 / 74.80**	**92.40** / 60.70	**95.10 / 83.40**

A.4 Matching-margin Stability and Disparity Shift

A competing disparity d' overtakes the clean preferred disparity $d^\star$ when

$$C_{\mathrm{adv}}(p, d') < C_{\mathrm{adv}}(p, d^\star). \tag{33}$$

Using Equation (16),

$$C_0(p, d') + \Delta C(p, d') < C_0(p, d^\star) + \Delta C(p, d^\star). \tag{34}$$

Rearranging and using Equation (15) gives

$$\Delta C(p, d^\star) - \Delta C(p, d') > m_{d'}, \tag{35}$$

which proves Equation (17).

We additionally obtain a sufficient stability condition. Let

$$m = \min_{d \neq d^\star} m_d. \tag{36}$$

If

$$\max_d |\Delta C(p, d)| < \frac{m}{2}, \tag{37}$$

then, for every $d \neq d^\star$,

$$\begin{aligned} \Delta C(p, d^\star) - \Delta C(p, d) &\leq |\Delta C(p, d^\star)| + |\Delta C(p, d)| && (38)\\ &< m && (39)\\ &\leq m_d. && (40) \end{aligned}$$

Therefore, no competing disparity can satisfy Equation (17), and the preferred disparity remains unchanged.

B Experiments

B.1 White-box attack

Table 8 shows that the effectiveness of ***EpiMirror*** generalizes to ETH3D and DrivingStereo. In the full-image setting, ***EpiMirror*** achieves the highest D1 for all three ETH3D victims and all four DrivingStereo weather conditions, while EPE varies across attacks, indicating different error-magnitude distributions. In the patch-based setting, ***EpiMirror*** outperforms DepthVanish on most model and weather combinations, with only metric-specific exceptions on IGEV-Stereo/ETH3D (D1) and foggy DrivingStereo (EPE). Since the full-image and patch-based groups follow different perturbation protocols, we compare results only within each group.

B.1.1 Additional White-box Results

ETH3D. We further evaluate correspondence and sign coupling on ETH3D using RAFT-Stereo, IGEV-Stereo, and STTR. The results are reported in Table 6.

Across all three architectures, EpiMirror consistently outperforms the Same, Pixel, and Shuffled variants in both D1 and EPE. In particular, the large gap over Same confirms the importance of

Table 6: **White-box ablation on ETH3D.** Each entry reports D1 (%) / EPE (px), where higher values indicate stronger attacks. Bold marks the strongest result for each victim and metric.

Victim	Clean	Same	Pixel	Shuffled	Independent	*EpiMirror*
	D1 / EPE	D1 / EPE	D1 / EPE	D1 / EPE	D1 / EPE	D1 / EPE
RAFT-Stereo	0.72 / 0.27	28.70 / 49.64	85.46 / 308.37	78.68 / 293.17	89.61/ **376.93**	**93.70/ 416.32**
IGEV-Stereo	0.84 / 0.32	39.26 / 30.74	79.51 / 98.74	81.12 / 109.16	86.51 / 108.83	**87.77 / 112.89**
STTR	6.31 / 1.33	15.40 / 8.92	91.46 / 58.72	89.52 / 58.65	95.08 / **70.07**	**97.60** / 64.64

Table 7: **Cross-model transfer attacks across six stereo models on KITTI.** Adversarial pairs generated on each source are reused across targets, which are used only for evaluation. Entries report D1 (%) / EPE over the same 40 KITTI 2015 pairs. The rightmost column reports the mean over the five off-diagonal target models for each source. Stereopagnosia, Stereoscopic, ***EpiMirror***, and DepthVanish follow different input or attack protocols, so results across method groups are not directly comparable. Diagonal entries omit within-model evaluation.

Transfer attack on KITTI		Target Model						Average
Attack	Source	AANet D1 / EPE	PSMNet D1 / EPE	DeepPruner D1 / EPE	RAFT-Stereo D1 / EPE	IGEV-Stereo D1 / EPE	STTR D1 / EPE	5 Targets D1 / EPE
Stereopagnosia	AANet	–	35.68 / 4.38	32.01 / 4.20	33.51 / 3.56	18.09 / 2.67	**52.37** / 10.00	34.33 / 4.96
Stereoscopic	AANet	–	48.04 / **8.22**	44.09 / 6.44	33.15 / 3.89	24.21 / 3.14	45.64 / 9.15	39.03 / 6.17
EpiMirror	AANet	–	**51.09** / 6.33	**50.11 / 7.92**	**41.30 / 8.09**	**26.07 / 7.28**	48.16 / **14.72**	**43.35 / 8.87**
Stereopagnosia	PSMNet	27.35 / **6.10**	–	27.66 / 6.63	28.08 / 4.42	18.16 / 3.61	50.77 / **15.53**	30.40 / 7.26
Stereoscopic	PSMNet	24.09 / 3.21	–	35.74 / **6.86**	17.62 / 2.34	12.73 / 2.00	44.75 / 10.43	26.99 / 4.97
EpiMirror	PSMNet	**63.04** / 5.27	–	**73.10** / 6.59	**51.47 / 7.28**	**34.84 / 8.07**	**64.61** / 14.37	**57.41 / 8.32**
Stereopagnosia	DeepPruner	28.29 / 4.31	32.65 / 4.51	–	29.84 / 3.44	17.90 / 2.44	**50.63** / 11.85	31.86 / 5.31
Stereoscopic	DeepPruner	27.39 / 2.87	34.83 / 3.57	–	27.17 / 2.80	18.82 / 2.46	41.09 / 7.87	29.86 / 3.91
EpiMirror	DeepPruner	**43.52 / 5.24**	**51.68 / 5.32**	–	**39.79 / 6.89**	**22.25 / 7.21**	50.56 / **14.16**	**41.56 / 7.76**
Stereopagnosia	RAFT-Stereo	26.44 / 3.51	30.12 / 3.49	26.41 / 3.62	–	16.58 / 3.13	49.96 / 12.45	29.90 / 5.24
Stereoscopic	RAFT-Stereo	19.62 / 2.20	21.76 / 2.30	19.80 / 2.17	–	23.91 / 2.75	50.35 / 7.75	27.09 / 3.43
EpiMirror	RAFT-Stereo	**38.17 / 4.79**	**42.41 / 4.81**	**42.54 / 5.93**	–	**32.52 / 6.93**	**56.13 / 14.43**	**42.35 / 7.38**
Stereopagnosia	IGEV-Stereo	18.93 / 1.94	22.37 / 2.17	22.43 / 2.19	**31.75** / 2.93	–	**60.71** / 11.58	31.24 / 4.16
Stereoscopic	IGEV-Stereo	9.25 / 1.36	15.95 / 1.79	15.16 / 1.76	24.18 / 2.49	–	53.48 / 7.13	23.60 / 2.91
EpiMirror	IGEV-Stereo	**26.12 / 5.29**	**28.91 / 5.13**	**24.92 / 6.19**	28.21 / **9.13**	–	49.45 / **14.51**	**31.52 / 8.05**
Stereopagnosia	STTR	10.09 / 1.47	14.32 / 1.75	11.83 / 1.56	12.68 / 1.79	12.97 / 1.91	–	12.38 / 1.70
Stereoscopic	STTR	3.72 / 1.00	6.87 / 1.26	5.19 / 1.12	11.80 / 1.62	**14.00** / 1.90	–	8.32 / 1.38
EpiMirror	STTR	**17.53 / 2.49**	**19.58 / 2.67**	**18.13 / 3.42**	**19.49 / 4.62**	13.73 / **5.09**	–	**17.69 / 3.66**
DepthVanish	AANet	–	**61.91 / 20.08**	**59.45 / 16.36**	1.39 / 1.83	1.15 / 1.78	7.04 / 2.77	26.19 / 8.56
EpiMirror	AANet	–	35.68 / 6.33	32.01 / 7.92	**33.51 / 8.09**	**18.09 / 7.28**	**52.37 / 14.72**	**34.33 / 8.87**
DepthVanish	PSMNet	**28.84 / 5.57**	–	**37.11 / 9.82**	4.52 / 2.84	0.81 / 2.08	8.42 / 3.11	15.94 / 4.68
EpiMirror	PSMNet	27.35 / 5.27	–	27.66 / 6.59	**28.08 / 7.28**	**18.16 / 8.07**	**50.77 / 14.37**	**30.40 / 8.32**
DepthVanish	DeepPruner	20.24 / 4.31	**70.68 / 21.38**	–	11.71 / 4.91	0.69 / 1.60	13.37 / 2.61	23.34 / 6.96
EpiMirror	DeepPruner	**28.29 / 5.24**	32.65 / 5.32	–	**29.84 / 6.89**	**17.90 / 7.21**	**50.63 / 14.16**	**31.86 / 7.76**
DepthVanish	RAFT-Stereo	**26.44 / 4.79**	30.12 / 4.81	26.41 / 5.93	–	**16.58 / 6.93**	**49.96 / 14.43**	29.90 / 7.38
EpiMirror	RAFT-Stereo	15.92 / 3.44	**67.53 / 19.32**	**80.64 / 23.11**	–	1.19 / 1.55	18.00 / 3.15	**36.66 / 10.11**
DepthVanish	IGEV-Stereo	**26.12 / 5.29**	28.91 / 5.13	24.92 / 6.19	28.21 / 9.13	–	**49.45 / 14.51**	31.52 / 8.05
EpiMirror	IGEV-Stereo	16.57 / 3.60	**68.21 / 19.83**	**70.94 / 18.21**	**40.68 / 14.62**	–	20.83 / 3.48	**43.45 / 11.95**
DepthVanish	STTR	17.53 / 2.49	19.58 / 2.67	18.13 / 3.42	19.49 / 4.62	**13.73 / 5.09**	–	17.69 / 3.66
EpiMirror	STTR	**20.18 / 3.97**	**62.93 / 17.78**	**84.87 / 25.36**	**51.00 / 15.47**	1.17 / 1.85	–	**44.03 / 12.89**

opposite-sign coupling, while the gains over Pixel and Shuffled show that opposite signs alone are not sufficient and that correspondence-aware placement also matters. Independent optimization can be stronger on some metrics, especially under RAFT-Stereo, indicating that the proposed coupling is not necessarily the globally strongest white-box parameterization. Nevertheless, the controlled variants consistently support the complementary roles of correspondence placement and opposite-sign coupling.

B.2 Black-box Transfer-based

As shown in Table 7, ***EpiMirror*** exhibits consistent cross-model transfer in the full-image setting. For all six source models, ***EpiMirror*** obtains the numerically largest five-target mean D1 and EPE among the reported full-image attacks. For example, when PSMNet is used as the source, ***EpiMirror*** reaches an average D1/EPE of $57.41/8.32$, compared with $30.40/7.26$ for Stereopagnosia and $26.99/4.97$ for Stereoscopic. With RAFT-Stereo as the source, the corresponding average is $42.35/7.38$, compared with $29.90/5.24$ and $27.09/3.43$, respectively. These results indicate that the correspondence-mirrored perturbations can remain effective after transfer to architectures that are not involved in attack generation.

Table 8: **White-box full-image and patch-based attacks on ETH3D and DrivingStereo.** Each entry reports D1 (%) / EPE (px). Higher values indicate stronger attacks.

ETH3D

Method	RAFT-Stereo	IGEV-Stereo	STTR
Attack	D1 / EPE	D1 / EPE	D1 / EPE
Clean	0.72 / 0.27	0.84 / 0.32	6.31 / 1.33
Stereopagnosia	44.80 / 2.15	28.70 / **1.95**	52.40 / 6.72
Stereoscopic	35.60 / 2.62	25.90 / 2.06	58.20 / 9.48
EpiMirror	**51.00 / 2.48**	**43.20** / 1.90	**66.80 / 7.81**
DepthVanish	84.60 / 42.80	**94.30** / 47.50	89.70 / 55.10
EpiMirror	**87.10 / 51.90**	93.90 / **59.60**	**93.00 / 73.40**

DrivingStereo

Method	Sunny	Cloudy	Foggy	Rainy
Attack	D1 / EPE	D1 / EPE	D1 / EPE	D1 / EPE
Clean	11.27 / 2.35	13.44 / 2.35	11.28 / 1.96	8.64 / 2.48
Stereopagnosia	52.80 / 4.20	56.40 / 4.80	54.20 / 4.50	58.60 / 5.10
Stereoscopic	**43.70** / 5.10	48.50 / 5.80	46.10 / 5.40	50.20 / 6.30
EpiMirror	**58.90 / 4.55**	**62.10 / 5.15**	**60.00 / 4.82**	**64.90 / 5.52**
DepthVanish	90.40 / 58.30	92.10 / 64.70	91.30 / **61.20**	94.00 / 72.60
EpiMirror	**91.80 / 68.10**	**93.20 / 74.80**	**92.40** / 60.70	**95.10 / 83.40**

Table 9: **Black-box query-based attack results across four stereo benchmarks.** Each entry reports EPE (px) and D1 (%). Higher attacked errors indicate stronger attacks. Bold marks the numerically larger result between Independent and ***EpiMirror*** for each model, dataset, and metric, except when the paired interval crosses zero.

Model / Condition	Clean		Independent		*EpiMirror*	
	EPE ↓	D1 ↓	EPE ↑	D1 ↑	EPE ↑	D1 ↑
KITTI 2015						
RAFT-Stereo	1.22	5.84	5.43	22.10	**14.59**	**31.09**
IGEV-Stereo	1.28	6.27	6.59	18.23	**10.65**	18.30
STTR	4.88	25.33	12.60	40.38	**13.27**	**44.63**
FoundationStereo	1.49	6.89	7.38	21.80	**9.62**	**25.96**
Middlebury						
RAFT-Stereo	1.12	6.78	11.72	40.59	**43.16**	**51.09**
IGEV-Stereo	0.92	5.39	6.53	**31.02**	**10.94**	30.39
STTR	4.88	19.90	8.91	39.88	**10.77**	**49.58**
FoundationStereo	0.80	4.52	3.40	**26.17**	**3.96**	25.79
DrivingStereo *(RAFT-Stereo)*						
Sunny	1.81	8.23	14.59	47.73	**59.30**	**51.32**
Cloudy	2.83	15.95	17.40	**55.16**	**34.88**	52.37
Foggy	1.94	10.69	13.37	**49.29**	**19.02**	47.90
Rainy	2.27	9.40	10.36	**41.04**	**13.66**	39.74
ETH3D						
RAFT-Stereo	0.26	0.77	11.18	33.01	**40.96**	**35.46**
IGEV-Stereo	0.34	1.34	40.53	44.53	**56.76**	**49.36**
LightStereo-S	65.69	65.01	**87.22**	86.93	87.05	**86.95**

All metrics are computed against ground truth. All attacks use $\epsilon = 4/255$, at most 600 optimization queries, and seed 1. Bold compares Independent and ***EpiMirror*** only. For IGEV-Stereo D1 on KITTI 2015, the paired interval crosses zero; therefore, neither value is bolded, and no clear improvement is detected.

Additional results on DrivingStereo. Table 8 shows that the effectiveness of ***EpiMirror*** generalizes to ETH3D and DrivingStereo. In the full-image setting, ***EpiMirror*** achieves the highest D1 for all three ETH3D victims and all four DrivingStereo weather conditions, while EPE varies across attacks, indicating different error-magnitude distributions. In the patch-based setting, ***EpiMirror*** outperforms DepthVanish on most model and weather combinations, with only metric-specific exceptions on IGEV-Stereo/ETH3D (D1) and foggy DrivingStereo (EPE). Since the full-image and patch-based groups follow different perturbation protocols, we compare results only within each group.

We evaluate the EpiMirror prior with NES, Bandits-TD, Square, SimBA, and Random Search on RAFT-Stereo, IGEV-Stereo, and STTR using the same 40 KITTI 2015 image pairs. Each paired comparison uses seed 1, $\epsilon = 4/255$, and exactly 600 optimization queries, including initialization, with early stopping disabled.