WHEN CORRUPTIONS DISRUPT CROSS-VIEW GEOMETRY IN LIGHT FIELD DEPTH ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

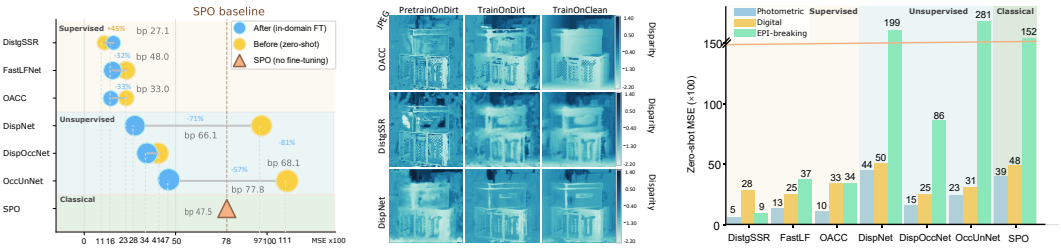
Light field (LF) depth estimation recovers scene geometry from correspondences across sub-aperture views, yet its robustness to corruptions remains largely unexplored despite strong performance on clean benchmarks. Unlike single-image vision, LF corruptions can affect both individual-view appearance and the cross-view geometry carrying disparity, making visual severity alone insufficient to characterize robustness. We introduce LFDepth-C, a corruption-robustness benchmark built on the HCI 4D Light Field Dataset, covering 15 corruption types across five categories, two severity levels, and seven representative classical, supervised, and unsupervised methods. Beyond conventional depth metrics, we quantify corruption-induced changes in epipolar plane image (EPI) structure using a cross-view residual ratio and slope-recovery error. Results show that visual severity poorly predicts depth degradation: corruptions that preserve EPI geometry cause limited errors, whereas those that disrupt cross-view correspondences are substantially more harmful. Supervised methods are generally the most robust under zero-shot evaluation, while in-domain adaptation substantially improves several unsupervised methods, with gains varying across architectures and metrics. These findings identify cross-view geometry preservation as a key factor in LF depth robustness and provide a systematic basis for studying corruption-induced failures.

1 INTRODUCTION

LF imaging records scene radiance across both spatial and angular dimensions, providing multiple sub-aperture views of the same scene (Wu et al., 2017; Lam, 2015).LF depth estimation exploits correspondences across these views to recover scene geometry and supports applications such as 3D reconstruction (Kim et al., 2013), view synthesis (Guo et al., 2021; Jin et al., 2022), refocusing (Wang et al., 2019), and occlusion-aware perception (Wang et al., 2015; Schilling et al., 2018).In EPIs, cross-view correspondences form structured lines whose slopes encode disparity (Wanner & Goldluecke, 2014; Zhang et al., 2016; Sheng et al., 2018), making cross-view geometry central to LF depth estimation. Although recent methods perform well on clean benchmarks, their robustness to corrupted observations remains much less studied.

Corruption poses a particular challenge for LFs because it can affect not only the appearance of individual views, but also the consistency across them (Wanner & Goldluecke, 2014; Jin et al., 2020; Wang et al., 2022a). Photometric changes can substantially alter appearance while leaving cross-view correspondences largely intact, whereas occlusion, geometric distortion, and view-dependent degradation may directly disrupt the EPI structure that encodes disparity (Wanner & Goldluecke, 2012; 2014; Zhang et al., 2016; Sheng et al., 2018). Thus, visual severity alone is insufficient to characterize robustness, and a central question for LF depth estimation is whether the cross-view geometry required for disparity recovery remains reliable. Yet current evaluations offer limited insight into this distinction. Most LF depth estimation methods are developed and compared primarily on clean benchmarks (Wang et al., 2022b; Huang et al., 2021; Wang et al., 2023b; Lin et al., 2022; Zhang et al., 2024; Jin & Hou, 2022), while existing corruption benchmarks mainly target single images (Hendrycks & Dietterich, 2019) or stereo pairs (Agnihotri et al., 2025), without explicitly characterizing corruption effects on the richer angular structure of an LF. It therefore remains unclear which corruptions are most detrimental, how robustness varies across methods, and how much of the lost performance can be recovered through adaptation.

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.



(a) (b) (c)

Figure 1: (a) MSE ($\times 100$) and bad-pixel rate for seven methods under zero-shot and in-domain adaptation. (b) Qualitative disparity maps for a representative subset of the benchmark: three representative methods on JPEG compression at the severe level, comparing three training protocols side by side in each cell. (c) Average impact of three corruption groups on seven methods, grouped by whether they break EPI structure: Photometric includes brightness, contrast, saturation, as well as camera, defocus, Gaussian, PSF blur. Digital includes bit error, JPEG, and pixelate. EPI-breaking includes fog, frost, snow and elastic transform, glass blur.

To answer these questions, we introduce LFDepth-C, a corruption-robustness benchmark built on the HCI 4D Light Field Dataset (Honauer et al., 2017). LFDepth-C covers 15 corruption types across five categories and two severity levels, and evaluates seven representative LF depth estimation methods. In addition to MSE and bad-pixel rate, we quantify corruption-induced EPI degradation using a cross-view residual ratio at the ground-truth geometry and a slope-recovery error. We evaluate zero-shot robustness, in-domain adaptation, and post-adaptation clean performance.

Overall, our experiments reveal several consistent trends, which we summarize in Figure 1. Corruptions that largely preserve EPI geometry generally cause smaller errors, whereas those that occlude, warp, or misalign cross-view correspondences tend to be more harmful, indicating that visual severity alone is an incomplete proxy for LF depth robustness. Zero-shot robustness also varies considerably across estimation paradigms, with supervised methods performing better overall than the classical baseline and unsupervised methods in our evaluation. In-domain adaptation recovers substantial performance for several methods, particularly some unsupervised models, but the gains are neither uniform nor fully captured by a single metric: lower MSE does not always correspond to improved bad-pixel accuracy. Taken together, these results characterize LF depth robustness along three related aspects: the extent to which corruption damages cross-view geometry, the sensitivity of a method to such damage, and the extent to which the resulting errors can be recovered through adaptation. This perspective suggests that preserving usable cross-view geometry, rather than appearance quality alone, is an important consideration for robust LF depth estimation.

Our main contributions are summarized as follows. (1) We introduce LFDepth-C, a systematic benchmark for LF depth corruption robustness, covering 15 corruptions, five categories, two severity levels, seven representative methods, and complementary evaluation settings. (2) We augment standard depth evaluation with two EPI-based measures that quantify corruption-induced damage to cross-view consistency and disparity recoverability. (3) We provide a unified empirical analysis of LF depth robustness in terms of cross-view geometry preservation, estimator sensitivity, and adaptation recoverability, revealing consistent trends across methods and metrics.

2 Related Work

Traditional light field depth estimation Early LF depth estimation recovers geometry from handcrafted cross-view consistency. Multi-view stereo methods combine correspondence and defocus cues (Tao et al., 2013), incorporate shading (Tao et al., 2017), estimate disparity in the Fourier domain (Jeon et al., 2015), or improve robustness to occlusion and noise through angular consistency, entropy, adaptive defocus, and occlusion-aware voting (Wang et al., 2015; Williem & Park, 2016; Han et al., 2022). EPI-based methods instead infer disparity from oriented line structures using structure tensors (Wanner & Goldluecke, 2014), the spinning parallelogram operator (Zhang et al., 2016), multi-orientation slopes (Sheng et al., 2018), and explicit occlusion handling (Schilling et al., 2018). These approaches make the geometric basis of LF depth estimation particularly explicit, since disparity is encoded by the persistence and slope of EPI structures across views. However,

existing evaluations largely focus on clean data or specific nuisances such as occlusion and noise, rather than systematic common corruptions. How hand-crafted cross-view cues degrade under common corruptions remains underexplored and LFDepth-C evaluates this behavior with representative classical methods and EPI degradation measures.

Deep learning-based light field depth estimation Recent LF depth methods improve cross-view matching through several complementary designs. Occlusion-aware approaches construct more reliable matching costs by suppressing unreliable views, as in OACC (Wang et al., 2022b) and the photometric-consistency formulation of Wang et al. (2023a). FastLFNet, MANet, and FPattNet strengthen matching through multi-scale or multi-receptive-field aggregation (Huang et al., 2021; Li et al., 2020; Xiao et al., 2023). Other methods encode angular or epipolar structure more explicitly: DistgSSR disentangles spatial and angular information (Wang et al., 2023b), EPINET exploits epipolar geometry priors (Shin et al., 2018), and OALFGAN combines spatial-angular attention with objectives for occlusion and disparity discontinuities (Yan et al., 2024). Unsupervised methods such as DispNet, DispOccNet, and OccUnNet instead learn from photometric consistency or reliability estimation across views (Lin et al., 2022; Zhang et al., 2024; Jin & Hou, 2022). Despite their different designs, these methods all depend on some form of cross-view consistency, which may itself be altered by corruption. Yet evaluation remains dominated by clean HCI data or occlusion-specific settings, with little analysis of how common corruptions affect the underlying angular and EPI structure. LFDepth-C addresses this gap by evaluating representative classical, supervised, and unsupervised methods under 15 corruptions, two severity levels, and three evaluation protocols, together with conventional depth metrics and EPI-based measures.

Robustness and adaptation under common corruptions Common-corruption robustness has been extensively studied in conventional vision. ImageNet-C (Hendrycks & Dietterich, 2019) established standardized corruption types and severity levels, with later extensions to object detection (Liu et al., 2024) and stereo disparity estimation (Agnihotri et al., 2025). These benchmarks, however, operate on single images or image pairs, whereas LF depth estimation relies on a 4D light field whose disparity cues are encoded in cross-view correspondences. A corruption may therefore affect not only individual-view appearance but also the geometric consistency required for depth recovery. Consequently, visual degradation alone does not fully characterize LF robustness.

Adaptation introduces a complementary question. Many LF methods, particularly unsupervised ones, rely on cross-view photometric consistency, which corruption can directly violate. In-domain adaptation may recover performance by adapting features or reliability estimates to the corrupted distribution, but it may also affect clean-domain behavior. Existing LF studies do not systematically compare zero-shot robustness, in-domain adaptation, and post-adaptation clean performance under a unified corruption setting. LFDepth-C jointly evaluates these three protocols and explicitly measures corruption-induced EPI degradation through cross-view residual and slope-recovery metrics, providing a unified view of corruption sensitivity and adaptation behavior.

3 LFDepth-C: Benchmark and Evaluation

Data and corruption types LFDepth-C is built on the HCI 4D Light Field Dataset (Honauer et al., 2017) and comprises 15 corruption types across five categories and two severity levels (mild and severe). The five categories follow the conventional corruption taxonomy and describe what signal is degraded: Photometric (brightness, contrast, saturation), Blur (defocus, Gaussian, glass, camera motion, PSF), Weather (fog, frost, snow), Digital (JPEG compression, pixelate, bit error), and Geometric (elastic transform). Photometric corruptions rescale image intensity while preserving line geometry. Blur corruptions leave line positions intact but introduce smearing that erodes sub-pixel detail. Weather corruptions operate through two distinct mechanisms: fog reduces image contrast, whereas frost and snow cause partial occlusion. Digital corruptions introduce block-level and pixel-level artifacts that displace or fragment line structures. Geometric corruption distorts angular and spatial correspondence, violating the underlying EPI model. Each corruption is applied to all 81 sub-aperture images in 8-bit sRGB space with uniform severity, while the ground-truth disparity maps remain unmodified.

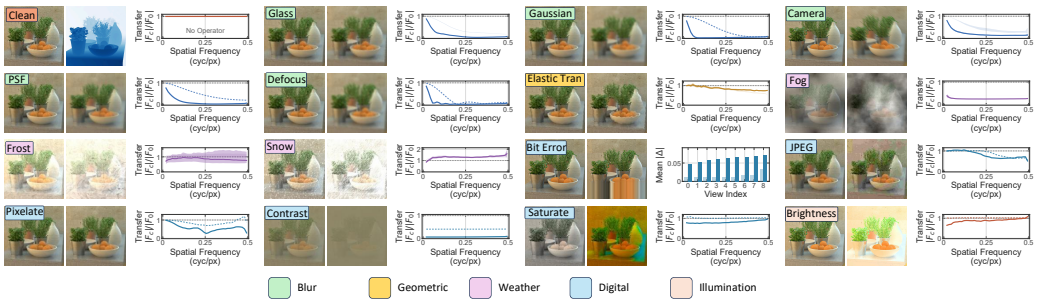


Figure 2: Visual comparison of the clean herbs scene after 15 corruption types, each at two severity levels, grouped by corruption category. Each corruption is illustrated with a pair of sub-aperture images (left) and a diagnostic plot (right) that quantifies its effect on cross-view consistency. For deterministic shared transforms and per-view random fields, the plot shows the transfer ratio $|F_c|/|F_0|$ of the angular frequency spectrum as a function of spatial frequency. A lower curve indicates stronger disruption of cross-view geometry. For coded-stream bit error, the plot shows the mean absolute deviation from clean data as a function of view index, revealing error accumulation along the view index. These plots correspond to the three cross-view coupling groups introduced in the text: deterministic shared transforms preserve angular frequency, per-view random fields introduce view-dependent inconsistencies, and coded-stream bit error accumulates along the view index.

Cross-view coupling taxonomy For LF depth estimation, the visual taxonomy is insufficient, because corruption in a LF affects not only individual-view appearance but also the consistency across all 81 sub-aperture views where disparity cues are encoded. Two corruptions in the same visual category can therefore produce different effects on depth estimation, depending on how they manifest across views. We thus introduce a second, complementary taxonomy that classifies corruptions by their cross-view coupling behavior. This cross-view taxonomy is a contribution of LFDepth-C: it characterizes how each corruption interacts with the angular structure of the LF and predicts the resulting EPI degradation. We partition the 15 corruption types into three groups. The first group consists of corruptions that are deterministic once the severity is specified: brightness, contrast, saturation, JPEG, pixelate, and defocus, Gaussian, and PSF blur. These are applied identically across all views, preserving cross-view geometry and altering only the intensity signal. The second group comprises corruptions where each view receives an independent realization: fog, frost, snow, camera motion blur, glass blur, and elastic transform. Adjacent views are degraded by distinct patterns, introducing cross-view inconsistency on top of the physical corruption. This behavior is realistic for occluders and sensor noise, which are not shared across viewpoints, and it means that elastic transform introduces inter-view geometric misalignment that would not occur in a static scene. The third group contains only bit error. Its coupling pattern operates along the view index rather than through a shared random field: the 81 views are encoded as a single video stream, corrupted, and decoded back, so a corrupted frame distorts neighboring frames through the temporal prediction loop and errors accumulate along the view index. This taxonomy is reflected in the EPI damage patterns: the residual ratio ρ is well above unity for corruptions that occlude views (fog, frost, snow), warp scene geometry (elastic transform), or are independently sampled per view (glass blur, camera motion blur), while it stays near or below unity for deterministic photometric and blur-type corruptions, and falls between these extremes for bit error. These results confirm that the mode of cross-view coupling, rather than the visual category alone, determines whether the EPI structure remains intact.

EPI damage metrics Conventional depth metrics including MSE and $\mathrm{BadPix}(0.07)$ measure aggregate depth estimation error (Honauer et al., 2017), but they cannot characterize whether corruption impairs the cross-view geometric structure that forms the foundation of LF depth estimation. To overcome this drawback, we design two EPI-derived metrics that directly quantify the degradation of epipolar structure introduced by corruption. Figure 2 shows the visual effect of each corruption type together with the resulting disruption of cross-view consistency that these two metrics quantify.

For each spatial pixel x with ground-truth disparity d, we sample the 9×9 sub-aperture views along the EPI line defined by d at position $x + d \cdot (u - u_0, v - v_0)$, and take the variance across the 81 samples as the cross-view residual $\mathrm{cost}(d)$. When the EPI model holds, intensity along this line remains constant, so $\mathrm{cost}(d)$ is small. We report two complementary metrics, each normalized against the clean LF of the same scene and aggregated by the median across all evaluation scenes.

Table 1: Measured EPI damage per corruption type. ρ is the cross-view residual ratio at the ground-truth geometry and s the slope-recovery error. Both are normalized by the clean data of the same scene and aggregated as a median over the 12 evaluation scenes, so 1.0 means "no change".

		Photometric			**Blur**					**Weather**			**Digital**			**Geometric**
		Brightness	Contrast	Saturation	Defocus	Gaussian	Glass	Camera Motion	PSF	Fog	Frost	Snow	JPEG	Pixelate	Bit Error	Elastic
ρ	mild	1.04	0.21	1.30	0.16	0.22	1.47	3.11	0.20	520	270	160	8.80	2.59	4.32	14
	severe	0.99	0.12	0.75	0.17	0.12	2.36	13.23	0.14	980	1000	270	18.31	3.09	5.77	42
s	mild	1.00	1.00	0.98	0.73	0.77	0.94	1.05	0.79	1.58	1.02	1.67	1.00	1.00	0.96	2.28
	severe	0.99	1.00	1.00	1.15	1.17	1.99	2.86	0.88	3.14	2.05	2.09	1.58	2.60	1.33	3.33

The first is the residual ratio

$$\rho = \frac{\mathrm{cost}_{\mathrm{corrupt}}(d)}{\mathrm{cost}_{\mathrm{clean}}(d)}, \tag{1}$$

which measures the degree to which the EPI line retains its correct geometric structure. Values close to 1.0 indicate that the EPI line remains structurally intact at the ground-truth disparity. The second is the slope-recovery error

$$s = \underset{c\in\mathcal{C}}{\mathrm{median}}\ \underset{p\in\Omega_{\mathrm{tex}}}{\mathrm{median}} \frac{|\hat{m}_c(p) - m_c(p)|}{|m_c(p)| + \epsilon}, \tag{2}$$

where $\mathcal{C}$ denotes the set of evaluation scenes, Ω_{tex} denotes the set of high-texture pixels, $m_c(p)$ is the ground-truth EPI slope, $\hat{m}_c(p)$ is the slope estimated using a parameter-free EPI slope estimator, and ϵ is a small constant to avoid division by zero. This metric captures how reliably disparity can be recovered from the EPI structure. These two metrics capture distinct EPI degradation mechanisms and failure modes in depth estimation. Occlusions, geometric warps, and compression artifacts break or displace EPI lines, impairing cross-view consistency and elevating ρ. Blur, by contrast, preserves line geometry but attenuates fine spatial texture, degrading slope estimation and raising s. Since ρ is a ratio of residual energies, it is sensitive to global intensity rescaling: a value below 1.0 does not necessarily imply that the EPI line is better preserved than in clean data, as blur reduces the residual value by suppressing high-frequency content. We therefore interpret ρ within each corruption family, using s as a complementary, detail-sensitive metric. Table 1 reports both metrics for all 15 corruption types at both severities.

Evaluation Protocols We design three hierarchical evaluation protocols to disentangle three distinct properties: zero-shot robustness, in-domain adaptation capability, and post-adaptation performance on clean data. The first protocol, named PretrainOnDirt, trains models on clean data and evaluates them directly on corrupted LFs to quantify zero-shot robustness. The second protocol, TrainOnDirt, fine-tunes models on a specific corruption type and severity, then tests them on held-out views with identical corruption settings to assess in-domain adaptation performance. The third protocol, TrainOnClean, evaluates the same fine-tuned checkpoints on clean test views to measure post-adaptation clean performance, and serves as a sanity check that in-domain adaptation does not come at the cost of clean accuracy. Together they assess whether a model degrades, whether it can adapt, and whether adaptation preserves clean accuracy. As a classical method with no learnable parameters, SPO (Zhang et al., 2016) is only evaluated under its standard deployment configuration, which corresponds to the zero-shot evaluation protocol.

4 Experimental Setup

Implementation details Corruption perturbations follow an established optical flow corruption benchmark (Hendrycks & Dietterich, 2019). All LFs have an angular resolution of 9×9 and a spatial resolution of 512×512. Fine-tuning and testing on identical scenes risks data leakage, and several released baseline checkpoints are pre-trained on parts of the HCI 4D dataset, which can inflate measured zero-shot robustness. To mitigate both issues, we strictly separate adaptation and evaluation scenes across all protocols. For in-domain adaptation, we use 16 scenes from the "Additional" split. For evaluation, we reserve 12 held-out scenes never seen during fine-tuning: 8 from the "Stratified" set (backgammon, boxes, cotton, dino, dots, pyramids, sideboard, stripes) and 4 from the "Test" set (bedroom, bicycle, herbs, origami). Since the released checkpoints may cover four of these 12 scenes, we additionally report conservative averages over the 8 fully unseen scenes, as shown in

Table 2: Average MSE ($\times 100$) over the 12 evaluation scenes and the 8 fully unseen scenes, for each method and evaluation protocol. The 8-scene subset excludes the four scenes that may appear in the pre-training data of the released checkpoints. Excluding these in-distribution scenes generally increases the error, confirming that the 12-scene averages are conservative. Best in red, second-best in blue; "–" excluded.

Method	DistgSSR	OACC	FastLFNet	DispNet	DispOccNet	OccUnNet	SPO
PretrainOnDirt (12)	11.1	22.7	23.1	97.0	40.8	110.8	78.3
PretrainOnDirt (8)	12.5	26.5	25.6	106.6	40.6	118.0	72.4
TrainOnClean (12)	10.9	8.2	17.7	23.5	6.6	19.8	–
TrainOnClean (8)	11.7	11.2	22.3	30.9	6.1	23.0	–
TrainOnDirt (12)	14.4	14.3	15.3	27.7	34.4	44.0	–
TrainOnDirt (8)	16.0	17.8	18.1	32.3	33.4	47.4	–

Table 2, while all other results average over all 12 scenes. This split ensures that PretrainOnDirt measures generalization, TrainOnDirt measures in-domain adaptation, and TrainOnClean measures clean-domain retention, while eliminating confounding from scene memorization or pre-training exposure. All models retain their original architectures and training paradigms: FastLFNet, OACC, and DistgSSR are supervised, while OccUnNet, DispNet, and DispOccNet are unsupervised. Full training runs take approximately 25 days on a server with two 32 GB vGPUs.

Compared methods We select seven representative methods spanning the dominant design paradigms, supervision regimes, and degrees of reliance on cross-view geometry in LF depth estimation. All have public implementations and are widely benchmarked on HCI, ensuring comparability and reproducibility. Supervised methods follow three design routes for handling cross-view geometry. OACC follows an occlusion-aware cost construction paradigm that explicitly models occlusions during matching cost computation. FastLFNet adopts a multi-scale cost aggregation strategy tailored for large disparity ranges and memory-constrained scenarios. DistgSSR uses explicit EPI geometry encoding to disentangle spatial and angular information and preserve cross-view geometric structure. For unsupervised paradigms, methods instead rely on photometric consistency across views, so their depth cues depend directly on cross-view agreement. DispNet acts as the foundational baseline driven purely by photometric consistency through multi-view feature matching. DispOccNet builds on this framework by adding occlusion prediction and loss down-weighting. OccUnNet further introduces a sub-LF partitioning strategy paired with reliability fusion. As a classical non-learning baseline, SPO provides parameter-free EPI slope estimation, which enables direct comparison between hand-crafted geometric priors and learned deep features. Together, these methods cover classical, supervised, and unsupervised paradigms with varying reliance on cross-view geometry, allowing us to test whether robustness correlates with design philosophy, supervision type, or dependence on intact EPI structure.

5 Results and Analysis

On the LFDepth-C dataset, we evaluate all seven methods under 30 corruption scenarios and three evaluation protocols, using MSE ($\times 100$) and $\mathrm{BadPix}(0.07)$ as metrics. The full results are summarized in Table 3. Our analysis follows a mechanism-oriented progression. We first establish that corruption harmfulness is determined by whether it preserves the EPI structure, then analyze zero-shot paradigm stratification, in-domain adaptation and clean retention, metric disagreement, and method-specific mechanisms. Finally, we examine scene-level difficulty.

Observation 1 Corruption harmfulness is determined by whether it preserves the EPI structure, not by visual severity. Structure preserving degradations are nearly harmless across all methods and severities: DistgSSR stays at 1.2 (Brightness), 1.4 (Contrast), and 1.1 (Saturation) at the mild level, compared with 1.0 clean, and never exceeds 6.6 even at the severe level. In contrast, structure breaking degradations produce the largest errors and severity gaps. For example, DispNet jumps from 38.1 to 340.0 on Elastic Transform, and the same split holds for SPO and OccUnNet. The ranking is thus dominated by geometry, not intensity statistics.

Observation 2 In the zero-shot setting, robustness is stratified by learning paradigm: supervised methods are the most robust, while the unsupervised methods spread across the middle and the

Table 3: We report the MSE ($\times 100$) of seven methods on 15 corruption types at two severity levels, averaged over the 12 evaluation scenes, with rows grouped by learning paradigm (classical, supervised, and unsupervised). Supervised and unsupervised models are evaluated under PretrainOnDirt, TrainOnClean, and TrainOnDirt settings, while the classical optimizer SPO is evaluated only under its zero-shot protocol. Within each paradigm group, the best result on each corruption is highlighted.

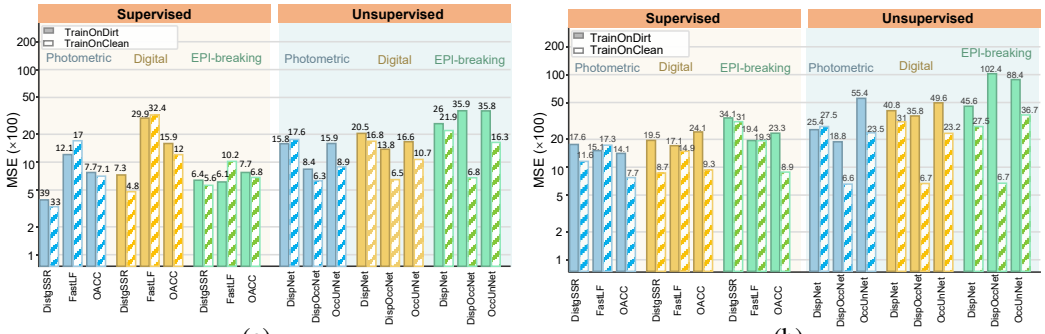
Method	Bit Error		JPEG		Pixelate		Saturation		Fog		Snow		Gaussian Blur	
	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe
SPO														
PretrainOnDirt	22.6	41.1	43.5	124.4	14.4	42.9	14.8	10.4	240.2	224.8	17.1	205.6	18.1	69.4
Supervised														
DistgSSR														
PretrainOnDirt	6.0	18.0	22.2	63.4	4.9	52.1	1.1	1.8	1.2	2.2	1.6	35.5	3.0	16.8
TrainOnClean	1.3	1.6	7.3	17.2	5.8	7.2	4.9	8.9	5.4	94.0	5.4	10.0	4.5	29.0
TrainOnDirt	4.8	21.6	10.7	26.3	6.3	10.5	4.8	6.1	5.3	86.3	5.5	14.6	5.0	40.0
FastLFNet														
PretrainOnDirt	11.9	21.9	29.0	49.3	8.4	28.2	7.1	8.0	7.5	15.8	18.8	61.5	9.3	26.8
TrainOnClean	73.2	11.3	14.0	18.2	10.0	15.0	10.5	12.2	8.0	8.1	9.7	15.3	8.2	16.3
TrainOnDirt	74.9	16.8	11.1	21.6	3.7	12.8	5.4	8.8	3.0	4.6	5.3	18.7	3.4	14.8
OACC														
PretrainOnDirt	11.5	26.8	35.5	70.6	11.1	44.5	6.5	6.9	6.6	8.0	6.7	61.1	7.9	23.1
TrainOnClean	22.0	10.9	7.3	9.5	6.8	7.7	6.5	6.6	6.7	6.8	6.6	7.6	7.9	10.3
TrainOnDirt	21.4	25.1	18.9	34.6	7.4	12.5	6.5	6.8	6.7	7.1	6.7	12.4	9.1	29.5
Unsupervised														
DispNet														
PretrainOnDirt	20.2	143.3	19.5	57.3	14.9	46.8	14.8	44.1	301.4	370.9	112.8	148.7	15.2	32.1
TrainOnClean	18.9	21.8	16.4	35.1	15.2	36.1	15.9	16.5	30.4	23.7	22.9	26.1	15.7	30.2
TrainOnDirt	24.6	34.7	22.2	49.0	14.6	38.6	16.7	21.9	34.9	46.2	34.2	37.6	15.2	31.3
DispOccNet														
PretrainOnDirt	11.9	19.6	21.9	55.8	7.3	33.6	7.4	8.8	65.4	130.3	56.0	144.8	7.3	21.4
TrainOnClean	6.4	6.7	6.6	6.7	6.4	6.6	6.4	6.2	6.4	6.4	6.6	7.2	6.3	6.2
TrainOnDirt	12.0	19.9	21.9	51.5	7.5	36.1	7.8	8.5	22.8	48.2	74.2	151.6	7.4	20.7
OccUnNet														
PretrainOnDirt	9.3	27.5	14.6	61.4	6.5	67.1	4.1	33.7	582.6	695.9	88.7	140.2	6.6	33.8
TrainOnClean	6.5	16.9	19.0	30.7	6.5	22.0	5.9	13.5	42.8	61.4	8.3	34.5	10.8	22.9
TrainOnDirt	10.3	39.1	29.3	60.4	10.2	49.3	7.0	10.3	68.6	74.7	30.9	73.4	17.2	52.7

Method	Defocus Blur		Glass Blur		Camera Motion		PSF Blur		Frost		Contrast		Elastic Transform		Brightness	
	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe
Classical																
SPO																
PretrainOnDirt	28.6	59.5	24.4	119.4	26.6	115.4	23.4	43.4	174.4	200.1	22.3	85.0	55.2	253.5	9.6	18.0
Supervised																
DistgSSR																
PretrainOnDirt	7.0	15.9	4.2	14.7	2.1	5.6	3.0	6.7	4.1	11.5	1.4	3.6	2.9	12.1	1.2	6.6
TrainOnClean	5.2	18.8	6.3	18.3	1.1	1.3	5.2	14.6	5.8	16.6	1.2	7.0	5.2	16.3	1.0	1.4
TrainOnDirt	7.6	36.7	8.1	33.0	1.6	4.0	5.9	19.6	7.1	15.4	1.3	11.8	5.8	20.9	1.0	5.1
FastLFNet																
PretrainOnDirt	11.9	21.0	10.1	55.3	9.0	28.1	9.2	12.7	20.6	62.6	7.5	10.4	33.7	79.8	7.0	11.0
TrainOnClean	12.8	21.9	14.4	27.7	28.8	19.2	18.5	14.9	8.5	22.5	32.0	23.0	10.4	23.0	8.5	13.6
TrainOnDirt	7.5	17.9	9.1	31.1	24.4	17.2	10.2	10.3	5.4	15.9	30.4	25.8	7.8	26.7	3.3	10.8
OACC																
PretrainOnDirt	11.1	18.8	9.4	43.4	7.4	10.6	8.0	11.3	7.0	15.0	6.6	7.9	9.6	170.1	6.5	10.8
TrainOnClean	7.4	9.0	7.2	8.1	6.7	7.1	6.9	7.7	6.7	6.9	6.6	6.7	7.0	14.8	7.3	6.6
TrainOnDirt	8.9	18.0	10.0	32.6	7.2	9.0	7.8	10.8	7.0	8.2	7.2	8.0	8.3	56.1	7.4	16.6
Unsupervised																
DispNet																
PretrainOnDirt	17.0	41.4	15.8	235.7	22.2	289.6	15.0	27.7	130.8	295.6	21.2	40.4	38.1	340.0	15.3	23.4
TrainOnClean	20.6	30.6	17.5	34.4	16.6	36.7	18.5	31.6	23.1	24.2	21.1	30.4	15.8	29.1	14.6	16.4
TrainOnDirt	16.3	26.0	15.4	23.4	17.3	37.5	15.2	16.2	28.5	65.8	15.2	19.6	17.2	54.8	14.8	25.4
DispOccNet																
PretrainOnDirt	9.4	17.7	9.8	56.3	14.6	68.3	8.1	13.8	33.6	176.5	6.9	16.4	36.0	150.4	6.2	8.7
TrainOnClean	6.3	6.3	6.7	6.8	6.3	8.0	6.3	6.4	6.4	6.5	6.4	6.5	7.8	6.7	6.0	6.2
TrainOnDirt	9.2	17.4	9.6	38.5	13.1	48.3	8.1	14.5	22.3	129.6	7.1	13.5	50.6	144.0	6.0	8.9
OccUnNet																
PretrainOnDirt	11.6	30.0	9.0	40.0	11.9	118.2	7.9	21.7	307.4	703.1	5.8	26.0	24.5	222.8	4.3	8.5
TrainOnClean	8.2	13.7	5.2	21.2	16.0	53.2	10.2	13.9	18.8	43.0	7.0	41.2	6.6	23.3	4.2	6.1
TrainOnDirt	15.5	32.4	8.3	55.1	37.8	181.5	18.1	32.3	47.0	109.7	11.2	61.8	24.1	129.0	4.5	17.1

bottom of the leaderboard. The three supervised SOTA models occupy the top of the leaderboard, all clearly ahead of the classical SPO. DispOccNet is the strongest unsupervised method under zero-shot evaluation and even surpasses the classical SPO, whereas DispNet and OccUnNet remain the most fragile, with zero-shot errors of roughly $9\times$ and $10\times$ that of DistgSSR. The per-corruption view in the appendix shows the same stratification spoke by spoke: the supervised curve stays at the rim on almost every corruption, whereas the classical optimizer and the two weakest unsupervised methods fall to the floor on the Weather spokes.

Table 4: Average MSE ($\times 100$) over all 30 corruption scenarios under the three protocols. Change is the relative change of TrainOnDirt over PretrainOnDirt (negative values indicate improvement). For each row, the best result is highlighted in red and the second-best in blue. TrainOnClean shows that the fine-tuned models retain low error on clean inputs.

Method	PretrainOnDirt	TrainOnClean	TrainOnDirt	Change
DistgSSR	11.1	10.9	14.4	$+30.4\%$
OACC	22.7	8.2	14.3	-37.1%
FastLFNet	23.1	17.7	15.3	-33.8%
DispNet	97.0	23.5	27.7	-71.4%
OccUnNet	110.8	19.8	44.0	-60.3%
DispOccNet	40.8	6.6	34.4	-15.8%



(a) (b)

Figure 3: MSE ($\times 100$) on a log scale, averaged across 12 evaluation scenes. Panel (a): mild corruption severity, panel (b): severe severity. Solid bars denote TrainOnDirt: models fine-tuned on the corruption family shown and its severity, tested on held-out views of the same corruption. Hatched bars (TrainOnClean) indicate the retained clean-domain performance after adaptation.

Observation 3 In-domain adaptation on the corrupted distribution substantially improves the robustness of unsupervised methods, lifting all three past the classical optimizer. In relative terms, DispNet improves the most and OccUnNet follows closely. DispOccNet improves least, though it too benefits. Fine-tuning also helps most supervised methods: OACC and FastLFNet both gain. The sole exception is the strongest zero-shot model, DistgSSR. Fine-tuning thus narrows the gap across paradigms and lifts nearly every method, at the cost of only the strongest model's zero-shot advantage. At the same time, as shown in Figure 3 and Table 4, fine-tuned models retain strong clean-domain performance: TrainOnClean is the lowest-error protocol for five of the six methods, indicating that in-domain adaptation does not cause catastrophic forgetting on clean inputs, and the models remain far more accurate on clean inputs than on corrupted ones.

Observation 4 Comparing TrainOnClean and TrainOnDirt reveals that clean-domain retention and corrupted-domain adaptation are largely independent properties. DispOccNet achieves the lowest clean-domain error among all methods (6.6), yet its corrupted-domain error under TrainOnDirt is the highest (34.4), a clean-to-corrupted ratio of roughly $5.2\times$. DistgSSR exhibits the opposite profile: its clean-domain error is slightly higher (10.9), but its corrupted-domain error rises only to 14.4, a ratio of roughly $1.3\times$, the smallest gap among supervised methods. OACC and FastLFNet fall between these extremes, preserving moderate clean accuracy while keeping corrupted-domain error close to DistgSSR. These results indicate that strong clean-domain performance is a poor predictor of corrupted-domain robustness, and that the clean-to-corrupted gap provides a more faithful measure of adaptation quality than either protocol in isolation.

Observation 5 The two metrics agree on most settings but diverge in two cases: structured-noise corruptions and in-domain adaptation. Under mild JPEG, OccUnNet ranks first in MSE yet second-worst in $\mathrm{BadPix}(0.07)$, while DispNet shows the opposite pattern; under mild Pixelate, OccUnNet has the second-lowest MSE but the worst $\mathrm{BadPix}(0.07)$ among all seven methods. We attribute this to its occlusion-aware design: it stays accurate over most pixels but produces sparse, severe errors at occlusion boundaries. MSE's squared averaging dilutes these localized errors, so MSE alone misranks its robustness. SPO behaves differently: on Pixelate, despite only mid-range MSE, it attains the lowest $\mathrm{BadPix}(0.07)$ at both severities, suggesting that explicit EPI-line constraints

Table 5: Average PretrainOnDirt MSE ($\times 100$) and bad-pixel rate over 30 corruption scenarios, per scene and method. Red marks the most severe scene (highest mean MSE) and blue the second.

	bedroom		**bicycle**		**herbs**		**origami**	
Method	MSE	BadPix	MSE	BadPix	MSE	BadPix	MSE	BadPix
DispNet	72.54	59.4	43.53	57.6	158.45	75.0	66.79	67.3
DistgSSR	4.56	21.1	8.23	27.0	22.52	26.3	18.81	28.3
FastLFNet	15.83	38.7	14.09	43.2	52.77	47.3	24.27	49.8
OACC	9.49	23.8	65.33	94.6	28.80	25.7	20.26	30.2
DispOccNet	34.92	60.1	28.33	61.2	65.70	72.0	36.02	71.2
SPO	84.35	43.8	65.81	46.5	198.27	54.7	136.56	55.8
OccUnNet	111.97	65.8	89.26	68.1	210.20	87.2	107.52	79.0

are inherently robust to blocky, structured errors. This advantage is specific to Pixelate, however, since SPO does not achieve the lowest $\mathrm{BadPix}(0.07)$ on JPEG and thus cannot be read as a general digital-artifact advantage. Under in-domain adaptation, several unsupervised methods reduce MSE substantially without a proportional drop in $\mathrm{BadPix}(0.07)$. The extreme case is OccUnNet under TrainOnClean: its MSE remains a modest $\sim$20, yet $\mathrm{BadPix}(0.07)$ stays between 67% and 100% across all 30 corruption scenarios. In sum, fine-tuning reduces error magnitude but does not restore sub-pixel accuracy: MSE obscures this mismatch through averaging, while $\mathrm{BadPix}(0.07)$ exposes the localized large errors.

Observation 6 DistgSSR consistently delivers the best performance across both metrics for corruptions that primarily degrade appearance, which aligns with the EPI-structure principle established in Observation 1. Across most weather and photometric corruption types, it achieves the lowest MSE and the lowest BadPix simultaneously. We attribute this consistent advantage to its disparity-selective angular feature extractor (DS-AFE) (Wang et al., 2023b), which encodes EPI geometry directly into the convolution dilation operation and is therefore inherently invariant to appearance changes. That said, this advantage does not hold universally. The main exception lies in structured noise corruptions, specifically JPEG compression and pixelation, where its performance edge disappears entirely. Blocky artifacts from these corruptions disrupt the precise geometric correspondence that DistgSSR's design relies on, and SPO's global EPI line fitting emerges as the better-performing approach in terms of bad-pixel rate. There is also a marginal exception within the weather group: under severe snow corruption, OACC obtains a slightly lower BadPix score of 29.04 vs. DistgSSR's 29.39, while DistgSSR still retains the lowest MSE. Overall, DistgSSR's dual-metric leadership originates from its explicit EPI prior, and its distinct performance weakness further corroborates that this geometric prior serves as the binding constraint for its robustness.

Observation 7 Table 5 reports the zero-shot MSE and bad-pixel rate of all seven methods on the four test scenes. This difficulty is already present on clean data: herbs is the hardest scene for every method, and corruption amplifies rather than creates the gap. This is consistent with the scene content of herbs: its dense thin leaves create frequent occlusions and many depth discontinuities, which weaken cross-view consistency even without corruption. Corruption therefore amplifies an existing geometric difficulty rather than creating a new one.

6 Conclusion

This work introduces LFDepth-C, a benchmark for evaluating the corruption robustness of LF depth estimation methods. Our experiments reveal that the harm caused by corruptions is primarily determined by whether the EPI structure remains intact, rather than by their perceived visual severity. Supervised methods exhibit the strongest robustness under zero-shot evaluation, whereas in-domain adaptation substantially restores the performance of unsupervised methods without materially compromising accuracy on clean data. Since MSE and $\mathrm{BadPix}(0.07)$ can yield inconsistent conclusions under structured noise and domain adaptation, robustness should be assessed with complementary metrics. These findings suggest that preserving cross-view geometric consistency, strengthening supervised design, and adapting unsupervised models to specific corruption types are promising directions for real-world deployment.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing, for formatting assistance (e.g., generating the LaTeX framework of this manuscript), and for drafting and refining parts of the analysis text. We did not use generative AI tools to fabricate experimental results; all reported numbers are produced by our own evaluation pipeline. We have reviewed all AI-assisted work, including the experimental claims and the data tables, and verified them against the raw evaluation outputs. We take responsibility for the final content of this work, including all text and claims produced with the aid of generative AI.

ETHICS STATEMENT

This work introduces a synthetic corruption benchmark built on a publicly available dataset and does not involve human subjects, private data, or potentially harmful applications. The released benchmark and evaluation protocols follow standard academic data-release practices and comply with the license of the original dataset.

REPRODUCIBILITY STATEMENT

We describe the benchmark construction, the corruption pipeline (Table 1), the evaluation protocols and metrics, and the training schedules (Sec. 4) in detail. All methods use officially released checkpoints and the official evaluation toolkit. The complete per-scenario results are reported in Appendix B. The benchmark, the corruption-generation code, and all experimental results will be publicly released to support reproduction.

REFERENCES

Shashank Agnihotri, Amaan Ansari, Annika Dackermann, Fabian Rösch, and Margret Keuper. DispBench: Benchmarking disparity estimation to synthetic corruptions. In *CVPRW*, 2025.

Mantang Guo, Jing Jin, Hui Liu, and Junhui Hou. Learning dynamic interpolation for extremely sparse light fields with wide baselines. In *ICCV*, 2021.

Kang Han, Wei Xiang, Eric Wang, and Tao Huang. A novel occlusion-aware vote cost for light field depth estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(11):8022–8035, 2022.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *ICLR*, 2019.

Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4D light fields. In *ACCV*, 2017.

Zhicong Huang, Xuemei Hu, Zhou Xue, Weizhu Xu, and Tao Yue. Fast light-field disparity estimation with multi-disparity-scale cost aggregation. In *ICCV*, 2021.

Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon. Accurate depth map estimation from a lenslet light field camera. In *CVPR*, 2015.

Jing Jin and Junhui Hou. OccUnNet: Occlusion-aware unsupervised learning of depth from 4-D light fields. *IEEE Trans. Image Process.*, 31:2216–2228, 2022.

Jing Jin, Junhui Hou, Jie Chen, and Sam Kwong. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization. In *CVPR*, 2020.

Jing Jin, Junhui Hou, Jie Chen, Huanqiang Zeng, Sam Kwong, and Jingyi Yu. Deep coarse-to-fine dense light field reconstruction with flexible sampling and geometry-aware fusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(4):1819–1836, 2022.

Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4), 2013.

Edmund Y. Lam. Computational photography with plenoptic camera and light field capture: tutorial. *J. Opt. Soc. Am. A*, 32 11:2021–32, 2015.

Yan Li, Lu Zhang, Qiong Wang, and Gauthier Lafruit. MANet: multi-scale aggregated network for light field depth estimation. In *ICASSP*, 2020.

Lili Lin, Qiujian Li, Bin Gao, Yuxiang Yan, Wenhui Zhou, and Ercan Engin Kuruoglu. Unsupervised learning of light field depth estimation with spatial and angular consistencies. *Neurocomputing*, 501:113–122, 2022.

Jiawei Liu, Zhijie Wang, Lei Ma, Chunrong Fang, Tongtong Bai, Xufan Zhang, Jia Liu, and Zhenyu Chen. Benchmarking object detection robustness against real-world corruptions. *Int. J. Comput. Vis.*, 132(10):4398–4416, 2024.

Hendrik Schilling, Maximilian Diebold, Carsten Rother, and Bernd Jähne. Trust your model: Light field depth estimation with inline occlusion handling. In *CVPR*, 2018.

Hao Sheng, Pan Zhao, Shuo Zhang, Jun Zhang, and Da Yang. Occlusion-aware depth estimation for light field using multi-orientation EPIs. *Pattern Recognit.*, 74:587–599, 2018.

Changha Shin, Hae-Gon Jeon, Youngjin Yoon, In So Kweon, and Seon Joo Kim. EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images. In *CVPR*, 2018.

Michael W. Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *ICCV*, 2013.

Michael W. Tao, Pratul P. Srinivasan, Sunil Hadap, Szymon Rusinkiewicz, Jitendra Malik, and Ravi Ramamoorthi. Shape estimation from shading, defocus, and correspondence using light-field angular coherence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(3):546–560, 2017.

Ting-Chun Wang, Alexei A. Efros, and Ravi Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *ICCV*, 2015.

Xinya Wang, Jiayi Ma, Peng Yi, Xin Tian, Junjun Jiang, and Xiao-Ping Zhang. Learning an epipolar shift compensation for light field image super-resolution. *Inf. Fusion*, 79:188–199, 2022a.

Xuechun Wang, Wentao Chao, Liang Wang, and Fuqing Duan. Light field depth estimation using occlusion-aware consistency analysis. *Vis. Comput.*, 39(8):3441–3454, 2023a.

Yingqian Wang, Jungang Yang, Yulan Guo, Chao Xiao, and Wei An. Selective light field refocusing for camera arrays using bokeh rendering and superresolution. *IEEE Signal Process. Lett.*, 26(1): 204–208, 2019.

Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Wei An, and Yulan Guo. Occlusion-aware cost constructor for light field depth estimation. In *CVPR*, 2022b.

Yingqian Wang, Longguang Wang, Gaochang Wu, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Disentangling light fields for super-resolution and disparity estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):425–443, 2023b.

Sven Wanner and Bastian Goldluecke. Globally consistent depth labeling of 4d light fields. In *CVPR*, 2012.

Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(3):606–619, 2014.

Williem and In Kyu Park. Robust light field depth estimation for noisy scene with occlusion. In *CVPR*, 2016.

Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE J. Sel. Top. Signal Process.*, 11(7):926–954, 2017.

Min Xiao, Chen Lv, and Xiaomin Liu. FPattNet: A multi-scale feature fusion network with occlusion awareness for depth estimation of light field images. *Sensors*, 2023.

Wenbin Yan, Xiaogang Zhang, and Hua Chen. Occlusion-aware unsupervised light field depth estimation based on multi-scale gans. *IEEE Trans. Circuits Syst. Video Technol.*, 34(7):6318–6333, 2024.

Shansi Zhang, Nan Meng, and Edmund Y. Lam. Unsupervised light field depth estimation via multi-view feature matching with occlusion prediction. *IEEE Trans. Circuits Syst. Video Technol.*, 34 (4):2261–2273, 2024.

Shuo Zhang, Hao Sheng, Chao Li, Jun Zhang, and Zhang Xiong. Robust depth estimation for light field via spinning parallelogram operator. *Comput. Vis. Image Underst.*, 145:148–159, 2016.

A Corruption application and grouping

We apply the 15 corruptions in three ways based on cross-view consistency: (1) deterministic shared transforms (brightness, contrast, saturation, JPEG, pixelate, defocus/Gaussian/PSF blur) apply the same operator to all 81 views; (2) per-view random fields (fog, frost, snow, glass blur, camera motion blur, elastic transform) draw an independent realization per view, adding cross-view inconsistency; and (3) coded-stream bit error encodes the 81 views as a single video stream, so errors propagate and accumulate along the view index.

Figure 2 shows the visual effect of all 15 corruptions, grouped by colour into five categories: Blur, Geometric, Weather, Digital, and Illumination. These categories reflect the dominant physical property of each corruption, while the three application modes above describe how it interacts with the LF's angular dimension: Blur and Illumination are mostly shared across views, Weather and Geometric are view-dependent, and Digital mixes shared transforms with the coded-stream bit error.

B Full per-scenario results

The complete results are reported in Tables 6–8 for average MSE ($\times 100$) over the 12 scenes, and in Table 9 for $\mathrm{BadPix}(0.07)$ under the same evaluation scenes, grouped by learning paradigm with values above 100% clamped to 100%. Figures 4 and 5 restate the zero-shot (PretrainOnDirt) columns of these tables as radial profiles, which makes the per-corruption pattern easier to compare across methods than the tables alone.

C Zero-shot robustness profiles

Tables 6 and 9 report the zero-shot numbers one cell at a time; Figures 4 and 5 recast the same PretrainOnDirt values as radial profiles, so that the shape of a method's failure pattern can be read directly. Each figure shows the 15 corruptions as spokes, ordered by family along the rim, and the seven methods as curves; the two figures differ only in severity, so they can be compared spoke by spoke. The radius is a relative score within each spoke: the best method on a corruption is placed at 100 and the score drops by 50 points per decade of larger error, clipped at 0. A curve closer to the rim is therefore better, and a score of 0 means an error at least two orders of magnitude above the best method on that spoke. Because the score is relative, the figures compare methods with each other rather than reporting absolute error, which is what the tables are for.

The profiles show that zero-shot robustness is organized by corruption family much more than by severity, and that the family ordering is stable across the two levels. On the Digital spokes every method stays in the outer quarter (74.6–96.9 at the mild level and 80.2–98.6 at the severe level), so compression artifacts separate the methods only weakly. The Weather spokes spread the methods the farthest apart (6.2–100 and 26.9–100), followed by the single Geometric spoke; on Fog at the mild level, SPO, DispNet and OccUnNet are pinned to the floor, more than two orders of magnitude behind DistgSSR. Within the Blur family the methods occupy a comparatively narrow band, with the classical SPO as the outermost outlier and OACC consistently second to DistgSSR.

Table 6: Average MSE ($\times 100$) over 12 scenes under PretrainOnDirt, for each of the 30 corruption scenarios. Red: best per row; Blue: second best.

Corruption	SPO	DistgSSR	FastLFNet	OACC	DispNet	DispOccNet	OccUnNet
Bit (mild)	22.61	5.99	11.88	11.47	20.16	11.86	9.31
Bit (severe)	41.14	18.00	21.85	26.84	143.28	19.62	27.46
Jpeg (mild)	43.45	22.23	28.97	35.51	19.46	21.91	14.55
Jpeg (severe)	124.40	63.37	49.26	70.60	57.30	55.75	61.35
Pixelate (mild)	14.43	4.86	8.38	11.09	14.91	7.28	6.47
Pixelate (severe)	42.92	52.14	28.17	44.53	46.78	33.60	67.15
Brightness (mild)	9.60	1.22	7.00	6.45	15.34	6.25	4.35
Brightness (severe)	17.98	6.62	11.03	10.77	23.35	8.71	8.53
Contrast (mild)	22.34	1.39	7.52	6.62	21.18	6.93	5.77
Contrast (severe)	85.01	3.56	10.40	7.89	40.35	16.43	26.02
Saturate (mild)	14.84	1.09	7.10	6.51	14.75	7.36	4.11
Saturate (severe)	10.36	1.76	7.98	6.90	44.08	8.81	33.67
Fog (mild)	240.16	1.16	7.51	6.61	301.36	65.42	582.58
Fog (severe)	224.83	2.25	15.78	8.03	370.87	130.31	695.87
Frost (mild)	174.38	4.09	20.56	7.05	130.78	33.60	307.38
Frost (severe)	200.11	11.47	62.60	14.96	295.56	176.52	703.08
Snow (mild)	17.13	1.56	18.77	6.73	112.82	56.05	88.72
Snow (severe)	205.59	35.48	61.54	61.15	148.72	144.82	140.22
Camera blur (mild)	26.65	2.11	8.97	7.40	22.21	14.55	11.93
Camera blur (severe)	115.38	5.60	28.05	10.59	289.55	68.28	118.24
Defocus blur (mild)	28.59	6.96	11.93	11.05	16.97	9.42	11.64
Defocus blur (severe)	59.55	15.94	21.04	18.83	41.42	17.74	30.02
Gaussian blur (mild)	18.06	2.96	9.29	7.90	15.18	7.34	6.57
Gaussian blur (severe)	69.38	16.78	26.80	23.07	32.11	21.41	33.75
Glass blur (mild)	24.40	4.25	10.12	9.41	15.79	9.79	9.02
Glass blur (severe)	119.42	14.69	55.33	43.42	235.72	56.34	40.05
Psf blur (mild)	23.45	2.95	9.18	7.96	14.96	8.07	7.94
Psf blur (severe)	43.43	6.71	12.68	11.27	27.72	13.81	21.66
Elastic transform (mild)	55.22	2.94	33.67	9.58	38.06	36.03	24.46
Elastic transform (severe)	253.49	12.08	79.81	170.05	340.02	150.45	222.82

DistgSSR is the best method on 14 of the 15 spokes at the mild level and on 13 of 15 at the severe level. The exceptions are informative: at the mild level only JPEG escapes it, where OccUnNet is best, while at the severe level both JPEG and Pixelate go to FastLFNet, the only method besides DistgSSR that reaches the rim anywhere. Most methods move outward from the mild to the severe figure (FastLFNet's mean score rises from 69.9 to 80.2), which means the relative gap to the best method narrows as the corruption becomes stronger; the two exceptions are DistgSSR, which is already at the rim at the mild level, and OccUnNet, the only method whose mean score falls (64.4 to 62.0), reflecting the same Weather weakness that dominates its TrainOnDirt row. Consistently with Observation 2 in the main text, the separation between the supervised and the unsupervised curves is largest exactly on the spokes where the EPI structure is destroyed rather than merely degraded.

D EPI DAMAGE METRICS PER CORRUPTION

Table 1 in Sec. 3 presents the two EPI damage metrics for all 15 corruption types across both severity levels. Below we describe the full computation procedure used to generate these values. Both metrics are computed on the same set of probed pixels per scene. For each of the 12 evaluation scenes, we start from the 9×9 sub-aperture views and corresponding ground-truth disparity map. We retain pixels with valid depth ($d > 0$) on a stride-2 grid, subsampled to a maximum of 192 pixels along the longer dimension.

Residual ratio ρ For each probed pixel with ground-truth disparity d, we bilinearly sample all 81 views at position $x + d\,(u - u_0, v - v_0)$. The variance across these 81 samples gives the cross-view residual $\mathrm{cost}(d)$, which is the quantity minimized by EPI-consistency methods. When the EPI line model holds, intensity remains constant along the line and the resulting cost is low.

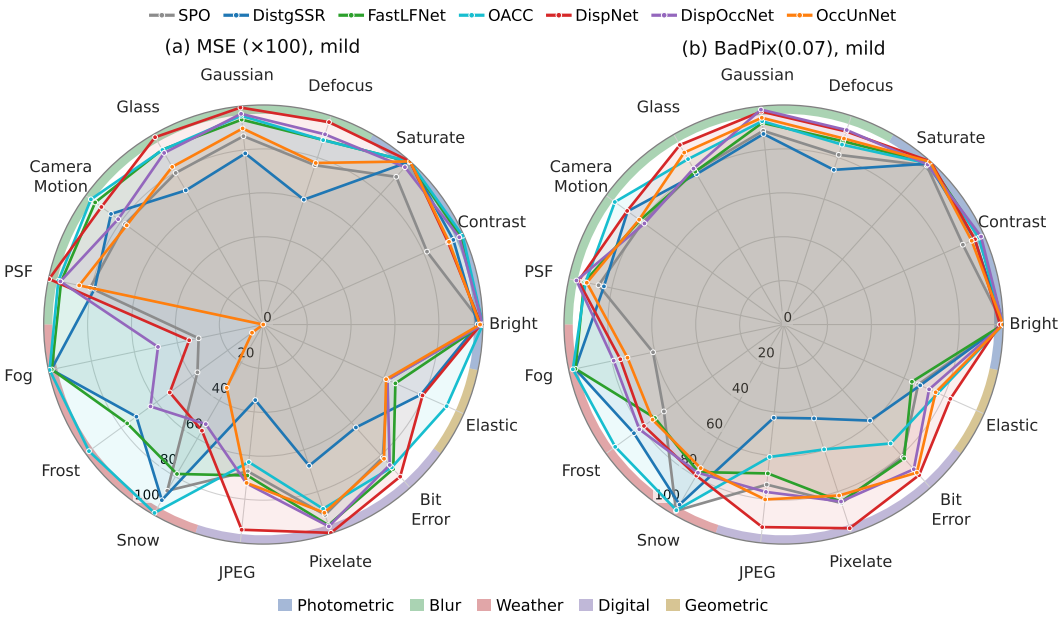


Figure 4: PretrainOnDirt robustness profiles at the mild severity level. The 15 spokes are the corruption types, ordered by family along the rim (family colour band, lower legend). The seven curves are the methods (upper legend). The radius is a relative score within each spoke: 100 for the best method on that corruption, falling by 50 points per decade of larger error and clipped at 0, so that a curve closer to the rim is better and 0 marks an error at least $100\times$ above the best method. The left and right circles show MSE ($\times 100$) and $\mathrm{BadPix}(0.07)$, respectively. The same layout at the severe level is shown in Figure 5.

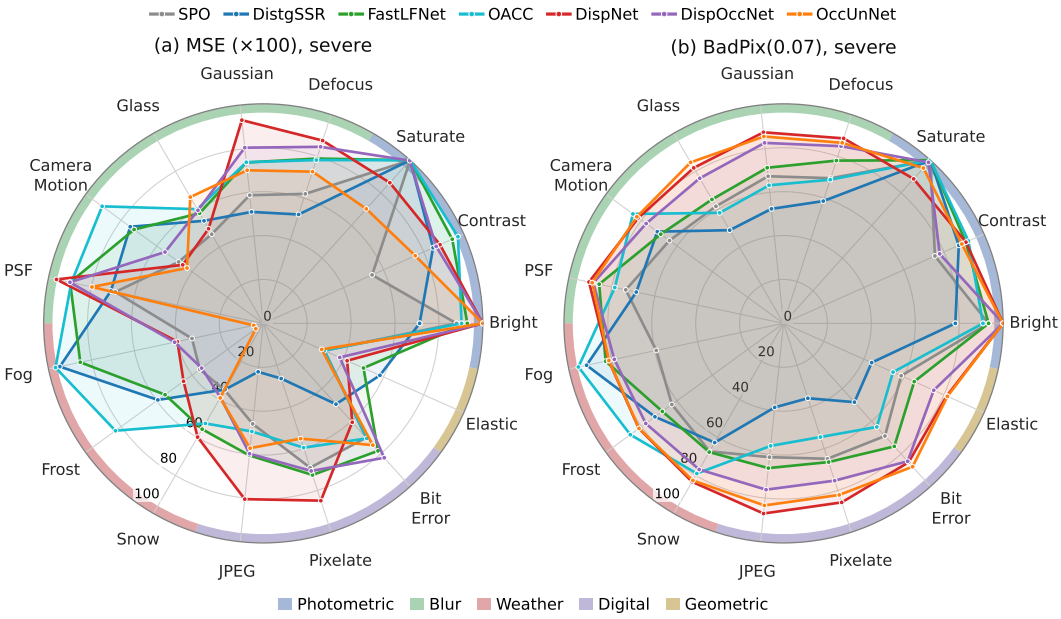


Figure 5: PretrainOnDirt robustness profiles at the severe severity level. The two figures are directly comparable spoke by spoke. The family structure of the ranking is preserved from the mild level, while the spread between the strongest and the weakest method is wider on the Weather and Geometric spokes and slightly narrower on the Digital ones.

Table 7: Average MSE ($\times 100$) over 12 scenes under fine-tuned models evaluated on TrainOnClean, for each of the 30 corruption scenarios. For each row, the best result is highlighted in red and the second-best in blue.

Corruption	DistgSSR	FastLFNet	OACC	DispNet	DispOccNet	OccUnNet
Bit (mild)	1.32	73.20	22.02	18.89	6.38	6.53
Bit (severe)	1.57	11.35	10.86	21.76	6.71	16.89
Jpeg (mild)	7.31	13.96	7.34	16.39	6.60	19.03
Jpeg (severe)	17.23	18.24	9.47	35.10	6.66	30.73
Pixelate (mild)	5.75	10.01	6.77	15.15	6.44	6.52
Pixelate (severe)	7.18	15.04	7.69	36.06	6.63	22.04
Brightness (mild)	1.03	8.46	7.35	14.62	6.01	4.21
Brightness (severe)	1.37	13.57	6.63	16.45	6.25	6.10
Contrast (mild)	1.22	32.00	6.60	21.05	6.44	7.05
Contrast (severe)	7.03	22.98	6.72	30.41	6.51	41.18
Saturate (mild)	4.89	10.48	6.52	15.91	6.42	5.85
Saturate (severe)	8.87	12.18	6.57	16.46	6.23	13.53
Fog (mild)	5.36	8.04	6.66	30.35	6.40	42.76
Fog (severe)	93.96	8.13	6.80	23.66	6.43	61.44
Frost (mild)	5.77	8.48	6.72	23.15	6.42	18.82
Frost (severe)	16.60	22.48	6.94	24.22	6.48	43.04
Snow (mild)	5.39	9.65	6.56	22.95	6.57	8.26
Snow (severe)	9.99	15.34	7.63	26.06	7.17	34.45
Camera blur (mild)	1.08	28.76	6.72	16.58	6.35	16.01
Camera blur (severe)	1.30	19.24	7.13	36.74	7.97	53.24
Defocus blur (mild)	5.15	12.81	7.43	20.63	6.29	8.22
Defocus blur (severe)	18.76	21.90	8.98	30.62	6.32	13.72
Gaussian blur (mild)	4.55	8.21	7.88	15.72	6.27	10.75
Gaussian blur (severe)	28.98	16.29	10.35	30.23	6.24	22.93
Glass blur (mild)	6.31	14.36	7.24	17.53	6.67	5.21
Glass blur (severe)	18.29	27.71	8.11	34.41	6.78	21.15
Psf blur (mild)	5.19	18.47	6.88	18.47	6.30	10.16
Psf blur (severe)	14.60	14.89	7.71	31.62	6.43	13.93
Elastic transform (mild)	5.25	10.40	6.98	15.77	7.82	6.62
Elastic transform (severe)	16.34	23.04	14.81	29.14	6.71	23.34

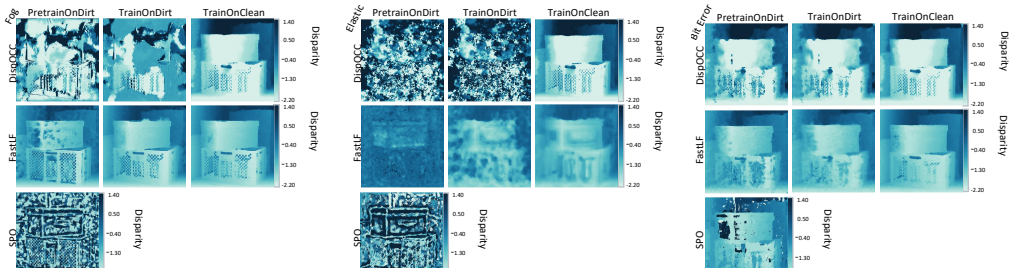


Figure 6: Per-view effect of three representative corruptions from the view-dependent families, one realisation per sub-aperture view: fog (Weather), elastic transform (Geometric), and coded-stream bit error (Digital). Each column shows the disparity maps of the three adaptation protocols PretrainOnDirt, TrainOnDirt, and TrainOnClean, illustrating the cross-view inconsistency introduced by these corruptions across adaptation protocols.

The sign of the ground-truth disparity is fixed per scene to the value that minimizes cost on the clean LF. The exact same set of pixels is used for both the clean and corrupted versions of a given scene. For each scene, we compute the median cost across all probed pixels. The final ρ value is the median over all 12 scenes of the ratio $\mathrm{cost}_{\mathrm{corrupt}}(d)/\mathrm{cost}_{\mathrm{clean}}(d)$. Values close to 1 indicate the EPI line remains intact at its correct geometric position.

Slope-recovery error s We use a parameter-free local slope estimator operating in the epipolar plane. Within a 9-pixel window, we compute angular and spatial intensity gradients, then derive disparity as the least-squares slope of the fitted line, given by $-\langle \partial_u I \, \partial_s I \rangle / \langle (\partial_s I)^2 \rangle$. We take the median value across angular views, and resolve the global sign ambiguity of the estimator once per

Table 8: Average MSE ($\times 100$) over 12 scenes under TrainOnDirt, for each of the 30 corruption scenarios. For each row, the best result is highlighted in red and the second-best in blue.

Corruption	DistgSSR	FastLFNet	OACC	DispNet	DispOccNet	OccUnNet
Bit (mild)	4.78	74.93	21.39	24.58	11.98	10.31
Bit (severe)	21.63	16.80	25.14	34.72	19.87	39.07
Jpeg (mild)	10.73	11.12	18.90	22.16	21.86	29.33
Jpeg (severe)	26.32	21.62	34.56	49.04	51.45	60.39
Pixelate (mild)	6.30	3.68	7.40	14.62	7.50	10.24
Pixelate (severe)	10.51	12.80	12.46	38.64	36.08	49.31
Brightness (mild)	1.04	3.26	7.37	14.84	6.04	4.55
Brightness (severe)	5.10	10.85	16.63	25.42	8.88	17.14
Contrast (mild)	1.29	30.37	7.22	15.19	7.09	11.15
Contrast (severe)	11.84	25.80	7.96	19.55	13.52	61.80
Saturate (mild)	4.83	5.41	6.52	16.72	7.77	6.98
Saturate (severe)	6.08	8.79	6.84	21.93	8.50	10.25
Fog (mild)	5.32	2.96	6.72	34.92	22.80	68.63
Fog (severe)	86.35	4.57	7.14	46.19	48.21	74.66
Frost (mild)	7.14	5.42	6.95	28.49	22.30	47.05
Frost (severe)	15.44	15.91	8.25	65.83	129.62	109.70
Snow (mild)	5.52	5.27	6.70	34.22	74.20	30.87
Snow (severe)	14.60	18.73	12.41	37.59	151.60	73.41
Camera blur (mild)	1.65	24.35	7.22	17.34	13.06	37.77
Camera blur (severe)	3.95	17.22	8.97	37.54	48.27	181.46
Defocus blur (mild)	7.60	7.49	8.90	16.29	9.25	15.55
Defocus blur (severe)	36.68	17.87	18.00	26.00	17.36	32.44
Gaussian blur (mild)	5.01	3.35	9.12	15.15	7.43	17.21
Gaussian blur (severe)	40.03	14.85	29.50	31.31	20.68	52.73
Glass blur (mild)	8.06	9.07	9.98	15.38	9.63	8.28
Glass blur (severe)	32.95	31.09	32.60	23.44	38.51	55.13
Psf blur (mild)	5.94	10.23	7.77	15.16	8.08	18.07
Psf blur (severe)	19.56	10.34	10.85	16.24	14.47	32.28
Elastic transform (mild)	5.84	7.83	8.33	17.17	50.59	24.10
Elastic transform (severe)	20.95	26.74	56.12	54.77	144.03	128.96

scene. Estimates are computed on the 20% most textured pixels, where the slope is well-defined, and compared against ground-truth disparity to yield a relative error. The final s value is the median across scenes of the per-scene mean relative error, normalized by the clean value of the same scene. One scene with no well-defined clean slope is excluded from this calculation. Both metrics are expressed as ratios relative to clean data, so a value of 1.0 corresponds to no change from the clean baseline. They capture distinct failure modes. The ρ metric responds to occlusions, geometric warps, and compression artifacts that break or displace the EPI line. The s metric responds to the same factors, and additionally captures blur, which leaves the EPI line in place but removes the high-frequency detail required for accurate sub-pixel localization. Because ρ is a ratio of residual energies, it is also affected by any global rescaling of intensity. A value below 1 does not mean the EPI line is better preserved than in clean data. Blur reduces the residual simply by removing the high-frequency content used to compute it. For this reason, we interpret ρ only within a given corruption family, and use s as a complementary, detail-sensitive metric. Per-scene values and their corresponding spreads are included with the publicly released results.

Limitations and Scope LFDepth-C is built on HCI 4D to obtain dense ground-truth disparity and controllable corruptions. We mitigate checkpoint exposure with 12 held-out scenes and an additional 8-scene unseen subset, but real captured LFs and capture-time degradations remain outside scope. We cover classical, supervised, and unsupervised methods, but not recent Transformer- or diffusion-based estimators, and we do not perform per-method architecture search. Adaptation uses early stopping, so fine-tuning epochs are model-dependent, with most runs converging around 200 epochs. This avoids overfitting but does not enforce an identical optimization budget. TrainOnDirt is in-domain adaptation to the same corruption type and severity, and the EPI metrics require ground truth and complement standard depth metrics. These choices bound the generality of our rankings while leaving the main EPI-preservation finding intact.

Table 9: BadPix(0.07) over 30 corruption scenarios, averaged over the 12 evaluation scenes, with rows grouped by learning paradigm. Values above 100% are clamped to 100%. Within each paradigm group, the best result on each corruption is highlighted.

Method	Bit Error		JPEG		Pixelate		Saturation		Fog		Snow		Gaussian Blur	
	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe
Classical														
SPO														
PretrainOnDirt	32.46	58.14	46.62	83.15	26.60	70.50	15.32	13.97	83.10	91.30	14.97	62.93	22.90	62.72
Supervised														
DistgSSR														
PretrainOnDirt	26.66	57.10	55.92	87.87	50.62	97.23	4.28	5.27	4.60	7.73	5.12	29.39	7.21	47.09
TrainOnClean	4.71	5.59	34.57	69.54	29.79	38.91	22.53	32.68	21.95	96.37	50.72	65.27	22.62	71.36
TrainOnDirt	23.26	48.56	41.83	77.20	31.99	53.64	23.48	29.07	21.86	95.30	30.50	40.27	23.15	80.40
FastLFNet														
PretrainOnDirt	35.39	56.74	67.09	86.36	31.18	85.79	16.08	18.28	17.89	40.24	43.48	80.92	22.01	68.62
TrainOnClean	91.01	45.01	35.74	54.31	26.69	40.50	33.81	36.01	25.87	26.01	29.36	50.82	22.18	40.55
TrainOnDirt	90.89	64.17	42.58	71.41	20.48	55.45	28.70	32.54	21.70	23.78	29.07	60.52	16.90	52.72
OACC														
PretrainOnDirt	34.21	59.22	60.13	83.18	62.53	89.89	10.46	11.02	10.72	13.54	10.95	29.04	13.47	59.90
TrainOnClean	35.56	20.83	13.17	17.70	11.28	12.50	10.58	10.65	10.68	11.07	10.66	11.60	16.52	17.59
TrainOnDirt	51.14	59.85	39.53	73.23	13.22	34.80	10.78	11.26	10.93	13.11	11.07	25.52	21.51	68.37
Unsupervised														
DispNet														
PretrainOnDirt	45.94	82.23	44.92	75.18	36.04	80.03	32.22	70.26	97.24	97.67	82.92	89.48	36.04	73.73
TrainOnClean	38.70	43.84	59.96	84.62	62.12	89.99	35.03	39.03	46.01	59.34	49.88	48.48	32.84	48.89
TrainOnDirt	49.24	78.44	65.88	90.41	63.95	91.53	33.68	48.31	60.08	75.04	60.63	62.37	32.73	65.37
DispOccNet														
PretrainOnDirt	46.68	68.65	81.12	100.00	55.95	100.00	30.27	35.72	72.54	89.30	76.46	97.64	29.78	74.71
TrainOnClean	28.18	30.27	28.07	29.24	28.20	29.72	26.83	28.10	28.35	27.86	28.20	27.43	28.08	29.85
TrainOnDirt	46.26	68.77	79.83	100.00	57.88	100.00	29.62	35.84	57.59	79.41	81.13	100.00	30.46	77.20
OccUnNet														
PretrainOnDirt	47.51	82.06	77.33	98.33	71.18	99.97	31.41	58.09	100.00	100.00	94.97	99.09	39.64	90.92
TrainOnClean	76.43	76.32	85.73	98.61	88.35	98.89	76.45	89.97	99.95	96.29	83.01	98.91	76.06	97.86
TrainOnDirt	64.41	97.59	85.93	99.78	82.20	100.00	43.38	47.56	98.14	96.15	73.48	96.28	67.70	94.78

Method	Defocus Blur		Glass Blur		Camera Motion		PSF Blur		Frost		Contrast		Elastic Transform		Brightness	
	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe	mild	severe
Classical																
SPO																
PretrainOnDirt	32.30	56.83	33.17	82.03	35.45	72.65	25.82	47.42	61.37	77.25	22.20	43.21	61.95	93.79	13.64	18.94
Supervised																
DistgSSR																
PretrainOnDirt	13.24	35.55	10.49	55.18	7.05	19.98	8.53	22.43	8.30	18.81	4.97	9.43	17.40	69.75	4.01	14.35
TrainOnClean	27.90	53.57	30.86	55.15	4.22	7.25	24.62	47.16	39.39	66.17	4.56	43.51	34.83	71.84	4.11	4.83
TrainOnDirt	31.94	70.18	33.61	69.75	6.45	14.07	26.87	54.87	36.86	56.58	4.85	53.38	37.35	74.10	4.12	10.15
FastLFNet																
PretrainOnDirt	27.23	50.44	38.20	90.83	37.46	75.87	21.57	35.08	55.25	78.97	17.32	27.81	81.28	90.84	15.51	24.77
TrainOnClean	36.47	60.60	35.13	73.10	39.85	64.23	39.52	46.70	24.89	54.83	84.98	59.39	38.52	70.26	23.22	26.09
TrainOnDirt	36.06	55.99	32.56	81.00	47.87	73.56	29.02	40.74	24.80	49.16	83.50	62.15	59.31	85.00	15.92	23.94
OACC																
PretrainOnDirt	18.55	46.33	17.87	75.75	12.38	22.18	14.04	29.56	12.51	20.75	11.05	15.77	28.45	89.18	9.82	16.80
TrainOnClean	11.99	14.04	11.73	14.06	10.96	11.72	11.02	12.42	10.72	11.71	10.88	11.01	11.45	38.30	14.60	10.76
TrainOnDirt	16.87	54.73	18.01	86.60	12.27	18.31	13.85	31.62	12.28	15.94	12.31	15.06	19.54	88.75	14.73	17.05
Unsupervised																
DispNet																
PretrainOnDirt	45.58	70.13	41.59	96.00	55.87	96.11	39.62	64.00	86.39	97.45	39.71	62.95	69.90	97.94	34.32	41.50
TrainOnClean	36.82	48.31	38.58	58.26	39.86	69.63	33.27	46.67	44.49	55.45	34.62	46.17	53.20	67.99	32.19	32.89
TrainOnDirt	34.88	58.63	42.25	77.32	57.39	86.49	35.26	44.34	54.19	86.17	35.06	41.86	72.53	91.08	33.05	38.91
DispOccNet																
PretrainOnDirt	38.01	67.51	63.38	100.00	75.17	95.41	32.67	56.91	65.88	94.49	29.71	59.81	97.94	100.00	27.71	33.63
TrainOnClean	29.52	30.60	29.15	31.31	28.54	35.13	27.49	28.54	27.77	28.68	27.50	28.20	30.77	31.40	27.12	27.51
TrainOnDirt	40.51	70.55	60.43	95.82	73.35	96.52	32.48	54.88	65.20	94.95	29.37	48.60	99.07	100.00	27.10	33.48
OccUnNet																
PretrainOnDirt	51.38	87.62	48.23	94.76	73.51	99.67	45.73	77.94	98.43	100.00	41.31	76.95	93.76	100.00	31.07	47.13
TrainOnClean	97.36	87.24	67.41	99.28	78.76	99.45	68.20	93.70	95.77	99.02	93.05	96.76	83.02	95.52	94.11	96.76
TrainOnDirt	62.89	86.27	58.36	98.82	83.00	100.00	69.08	87.14	83.32	98.08	49.07	92.75	96.30	100.00	31.32	55.90