# NTIRE 2026 Challenge on High-Resolution Depth of non-Lambertian Surfaces

Pierluigi Zama Ramirez Fabio Tosi Luigi Di Stefano Radu Timofte
Alex Costanzino Matteo Poggi Samuele Salti Stefano Mattoccia Junhong Min
Jiachen Tu Guoyi Xu Yaoxin Jiang Jiajia Liu Yaokun Shi Zuolingling Zuo
Zhihao Guan Yang Yang Mykola Lavreniuk Xinglin Li Jing Cao Zihao Zhang
Shiyu Liu Kui Jiang Qianming Yan Ziyi Wang Peishuai Zha
Divyanshu Kumawat Anshika Singh Utkarsh Yadav Karanbir Singh Dhindsa
Vinit Jakhetiya Jiajun Tian Junhyeong Kwon Youngjin Oh Nam Ik Cho
Hatem Ibrahem Ahmed Salem Hyun-Soo Kang Guanghui Wang

## Abstract

*This paper reports on the NTIRE 2026 challenge on HR Depth From images of Specular and Transparent surfaces, held in conjunction with the New Trends in Image Restoration and Enhancement (NTIRE) workshop at CVPR 2026. This challenge aims to advance the research on depth estimation, specifically to address two of the main open issues in the field: high-resolution and non-Lambertian surfaces. The challenge proposes two tracks on stereo and single-image depth estimation, attracting about 109 registered participants. In the final testing stage, 10 and 6 participating teams submitted their models and fact sheets for the two tracks.*

## 1. Introduction

Reversing the image formation process to recover the 3D structure of a scene represents one of the fundamental problems in computer vision. Depth estimation from images lies at the core of this endeavor and serves as a critical building block for a wide range of downstream applications, including augmented reality, autonomous driving, robotics, and more. Compared to *active* sensing technologies – such as LiDARs, Radars, and Time-of-Flight sensors – image-based depth estimation offers a cheaper and more flexible alternative, free from the physical constraints and deployment limitations that characterize dedicated hardware. The rapid progress brought by deep learning, and more recently by the emergence of *foundation models* for depth and 3D vision, has made this strategy increasingly competitive. Yet, despite the steady advances observed over the past decade, depth estimation remains far from solved under certain conditions. We identify two challenges that cut across virtually all image-based approaches – and even affect active sensors.

The first is spatial resolution. Modern color cameras can capture frames at tens of Megapixels: processing such high-resolution images with deep networks, however, introduces significant computational demands and requires dedicated training strategies to fully exploit the available spatial detail. The second is the ambiguity introduced by *non-Lambertian* surfaces. Transparent and reflective materials violate the basic assumptions underlying both active sensors – where LiDAR beams may refract or pass through glass – and image-based methods – where stereo matching tends to lock onto the content behind a transparent object rather than its surface. Ground-truth depth for such materials is consequently scarce in existing training sets, leading most models to fail precisely where robust perception is most needed.

This NTIRE 2026 Challenge on HR Depth from Images of Specular and Transparent Surfaces directly targets both of these open problems. Building on the success of previous editions [84–86], the challenge is organized around the Booster dataset [140, 143], whose 12 Mpx stereo imagery and abundant non-Lambertian objects make it uniquely suited to this purpose. Two tracks are proposed: *Stereo*, where depth is recovered through disparity estimation between rectified image pairs, and *Mono*, where a single frame is the only input. The challenge attracted approximately 109 registered participants, with 10 and 6 teams submitting models and fact sheets for the Stereo and Mono tracks respectively in the final testing phase. The methods, results, and findings of this edition are reported and discussed in detail in Section 4.

*Pierluigi Zama Ramirez (pier.zamaramirez@unive.it), Alex Costanzino, Fabio Tosi, Matteo Poggi, Samuele Salti, Stefano Mattoccia, Luigi Di Stefano and Radu Timofte are the NTIRE 2026 HR Depth from Images of Specular and Transparent Surfaces challenge organizers. The other authors participated in the challenge.
Sections B and C contains the authors' team names and affiliations.
The NTIRE website: https://cvlai.net/ntire/2026/

## 2. Related Work

**Deep Stereo Matching.** Initial end-to-end architectures split into two main families: 2D networks operating directly on feature maps [57, 64, 71, 78, 90, 97, 104, 107, 132, 138] and 3D cost-volume architectures that aggregate matching evidence across disparity levels [8, 12, 13, 18, 31, 44, 46, 96, 112, 124, 131, 146].

The field shifted significantly in the 2020s [111] with the introduction of iterative optimization frameworks and attention-based matching. Recurrent refinement strategies, initiated by RAFT-Stereo [105], rapidly became the dominant approach due to their flexibility and accuracy [40, 52, 113, 128, 129, 145], while transformer-based methods offered complementary advances in long-range context modeling [29, 56, 63]. Together, these developments drove steady performance saturation on established benchmarks including KITTI 2012 [24], KITTI 2015 [65], ETH3D [93], and Middlebury 2014 [92]. Parallel efforts have also targeted high-resolution disparity refinement [1, 110].

More recently, the first foundation models for stereo matching have emerged [4, 11, 39, 50, 123, 127, 135, 150], marking a qualitative leap in zero-shot generalization and robustness to non-Lambertian surfaces — properties that prove directly relevant to the Booster benchmark [140, 143] evaluated in this report.

**Monocular Depth Estimation.** Inferring depth from a single image is an inherently ill-posed problem that deep learning has made tractable, leveraging both large-scale annotated datasets [9, 20, 48, 82, 119] and self-supervised paradigms [25, 26, 30, 41, 49, 77, 108, 109, 126, 141, 147, 149] that reframe training as an image reconstruction problem via multi-view geometry, removing the need for ground-truth annotations entirely.

A major shift came with affine-invariant formulations designed for cross-domain generalization. MiDaS [88] pioneered this direction by mixing heterogeneous depth sources under a scale-and-shift invariant loss, followed by DPT [87], which leveraged transformer encoders to further boost zero-shot transfer. Complementary efforts targeted geometric fidelity, addressing shape distortions in monocular point clouds [137] and recovering high-frequency detail through multi-scale or patch-based inference [55, 66].

This trajectory converged into the current generation of *foundation models* for depth. Depth Anything [133] scaled the training pipeline with tens of millions of unlabeled images via a student-teacher protocol, while Depth Anything v2 [134] refined this with synthetic-to-real distillation to improve fine-grained detail and boundary accuracy. In parallel, generative approaches repurposed large-scale diffusion priors for dense prediction: Marigold [43] adapted Stable Diffusion for affine-invariant depth, GeoWizard [22] extended this to joint depth-normal estimation, and Lotus [33] simplified the inference protocol for efficiency. These methods were subsequently extended to temporally consistent video depth estimation [37, 42, 95]. Metric depth estimation has followed a parallel path: ZoeDepth [5] combined relative and metric heads in a two-stage framework; Metric3D [36, 136] conditioned predictions on camera intrinsics to resolve scale ambiguity; UniDepth [75, 76] jointly estimated depth and camera parameters; and Depth Pro [6] achieved sharp metric maps at high resolution. More recently, several works have reformulated the problem as dense point map estimation, directly regressing per-pixel 3D coordinates from a single image [120–122].

The ability to perceive *transparent and reflective surfaces* has emerged as an important open challenge, supported by dedicated benchmarks [140]. Proposed solutions include annotation pipelines that combine pre-trained depth models with material segmentation [17] or diffusion-based inpainting [114] to generate pseudo-labels for non-Lambertian objects, as well as depth completion strategies targeting transparent surface holes [14, 91]. Notably, recent foundation models such as Depth Anything v2 [134] exhibit surprising robustness on reflective and transparent regions without explicit supervision.

**Competitions and Challenges on Depth Estimation.** Depth estimation has been the subject of numerous dedicated challenges in recent years, covering both stereo and monocular settings. These include the Robust Vision Challenge (ROB) [144]; the Dense Depth for Autonomous Driving challenge (DDAD) [23]; the Fast and Accurate Single-Image Depth Estimation on Mobile Devices Challenge (MAI) [38]; the Argoverse Stereo Challenge [47]; and the long-running Monocular Depth Estimation Challenge (MDEC) [98–100], which held its fourth edition at CVPR 2025 [70]. The present challenge builds on five previous editions: NTIRE 2023 [84] and NTIRE 2024 [85] at CVPR, the TRICKY 2024 challenge [139] at ECCV 2024, NTIRE 2025 [86] at CVPR 2025, and the TRICKY 2025 challenge [83] at ICCV 2025.

**NTIRE 2026 Challenges.** This challenge is one of the challenges associated with the NTIRE 2026 Workshop [1] on: Deepfake detection [35], high-resolution depth [142], multi-exposure image fusion [81], AI flash portrait [28], professional image quality assessment [79], light field super-resolution [125], 3D content super-resolution [118], bitstream-corrupted video restoration [151], X-AIGC quality assessment [62], shadow removal [116], ambient lighting normalization [115], controllable Bokeh rendering [94], rip current detection and segmentation [19], low light image enhancement [15], high FPS video frame interpolation [16], Night-time dehazing [2, 3], learned ISP with unpaired data [73], short-form UGC video restoration [53], raindrop removal for dual-focused images [54], image super-resolution (x4) [10], photography retouching transfer [21],

[1] https://www.cvlai.net/ntire/2026/

mobile real-word super-resolution [51], remote sensing infrared super-resolution [60], AI-Generated image detection [32], cross-domain few-shot object detection [80], financial receipt restoration and reasoning [27], real-world face restoration [117], reflection removal [7], anomaly detection of face enhancement [148], video saliency prediction [69], efficient super-resolution [89], 3d restoration and reconstruction in adverse conditions [61], image denoising [101], blind computational aberration correction [103], event-based image deblurring [102], efficient burst HDR and restoration [72], low-light enhancement: 'twilight Cowboy' [45], and efficient low light image enhancement [130].

## 3. NTIRE Challenge on HR Depth from Images of Specular and Transparent Surfaces

We organize the NTIRE 2026 Challenge on HR Depth from Images of Specular and Transparent Surfaces to further advance the development of state-of-the-art solutions capable of handling high-resolution imagery and non-Lambertian surfaces – such as mirrors and transparent objects.

**Tracks.** The challenge is articulated into two tracks: *Stereo*, targeting methods that estimate disparity from rectified image pairs, and *Mono*, using single frame inputs.

- **Track 1: Stereo.** Participants are required to produce high-quality, dense disparity maps from 12 Mpx stereo pairs. The extreme resolution alone places this track beyond the operational range of most existing stereo networks, while the presence of non-Lambertian surfaces further challenges the underlying assumptions of correspondence-based matching.
- **Track 2: Mono.** This track requires depth estimation from a single 12 Mpx frame. Beyond the inherent ambiguity of the monocular setting, the frequent occurrence of transparent objects and mirrors – underrepresented in standard training corpora – adds further challenges.

**Datasets.** Both tracks are built upon the Booster dataset [140, 143], comprising 419 high-resolution stereo pairs captured across 64 scenes and split into 228 training and 191 testing pairs, with 38 and 26 scenes respectively. A subsequent release [140] augmented the benchmark with a dedicated monocular testing split of 187 single frames across 21 environments.

Following the protocol established in previous editions [84, 85], the original 228 training stereo pairs serve as the shared *training split* for both tracks. Two separate *validation splits* are defined by selecting frames with varying illumination conditions from 3 scenes in each testing partition: *Microwave, Mirror1, Pots* for the Stereo track and *Desk, Mirror3, Sanitaries* for the Mono track, yielding 15 validation samples per track out of the 26 and 28 available from the selected scenes. The remaining images in the two original testing splits constitute the official *testing splits*, amounting to 169 and 159 samples for the Stereo and Mono tracks respectively.

**Evaluation Protocol.** For each track we adopt metrics consistent with the official Booster benchmark [140, 143].

For the *Stereo* track, we report the percentage of pixels with disparity error exceeding a threshold $\tau$ (bad-$\tau$, with $\tau \in [2, 4, 6, 8]$), along with Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) in pixels.

For the *Mono* track, this edition introduces a key change with respect to previous ones [84–86]: submissions are now evaluated in the **metric depth** regime, without any post-hoc scale-and-shift alignment. We therefore report the threshold-based accuracy ($\delta < i$, with $i \in \{1.05, 1.15, 1.25\}$), Absolute Relative error (Abs Rel.), MAE, and RMSE.

Following [17, 85], all metrics are computed on three complementary pixel subsets: *All* pixels, *ToM* pixels (those belonging to transparent or mirror surfaces), and *Others* (the remaining pixels, excluding *ToM*). This decomposition allows to explicitly quantify the impact of non-Lambertian regions on model performance.

The primary ranking metric is bad-2 for the Stereo track and RMSE for the Mono track, both computed over *All* pixels and highlighted in **red** in the result tables. As in previous editions, two parallel rankings are defined based on performance over *ToM* and *All* regions respectively.

## 4. Challenge Results

For each track, four teams participated in the final evaluation phase, with their outcomes detailed in Sections 4.1 and 4.2. A brief explanation of each approach for both stereo and mono tracks is provided in Section 5.1 and Section 5.2, while the team composition is detailed in Sections B and C.

### 4.1. Track 1: Stereo

Table 1 reports the results for this first track. At the bottom, we report the baseline – i.e., CREStereo [52]. From left to right, we report bad-$\tau$ metrics, MAE, and RMSE metrics for *ToM*, *All*, and *Other* pixels, respectively. On the right of the team's name, we report their overall rank, computed according to bad-2 errors – the most restrictive metric – on *ToM* and *All* regions.

The top three submitted methods outperformed the baseline on both *ToM* and *All* pixels, with **GTR-robot** achieving the best results across both rankings, securing first place on *ToM* and *All* pixels alike. **NTR** ranks second on both, while **NJUST-KMG** completes the podium. Notably, **NTR** and **GTR-robot** achieve the lowest bad-2 errors on *Other* pixels, outperforming even the baseline in that region. The **FengFans** and **Depthforge_iitj** submissions fall below the baseline, with the latter producing degenerate predictions reflected in extremely high MAE and RMSE values across

| | ToM | | | | | | | All | | | | | | | Other | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Team | Rank | bad-2 | bad-4 | bad-6 | bad-8 | MAE | RMSE | Rank | bad-2 | bad-4 | bad-6 | bad-8 | MAE | RMSE | bad-2 | bad-4 | bad-6 | bad-8 | MAE | RMSE |
| **GTR-robot** | **#1** | **32.22** | **17.98** | **13.91** | **12.03** | **4.10** | **7.46** | **#1** | **16.10** | **6.13** | **4.06** | **3.30** | **1.76** | **4.57** | **12.98** | **3.16** | **1.50** | **1.02** | **1.18** | **3.41** |
| **NTR** | **#2** | **44.47** | **32.44** | **27.66** | **25.10** | **19.69** | **27.09** | **#2** | **16.86** | **8.47** | **6.55** | **5.60** | **5.17** | **12.82** | **12.13** | **3.23** | **1.65** | **0.93** | **1.11** | **3.34** |
| **NJUST-KMG** | **#3** | **56.79** | **33.76** | **24.85** | **19.22** | **11.39** | **22.88** | **#3** | **37.51** | **18.77** | **12.34** | **9.03** | **5.67** | **21.71** | 33.90 | 15.29 | 9.33 | 6.37 | 3.76 | 17.57 |
| **[Baseline]** | #4 | 59.64 | 47.26 | 40.27 | 35.41 | 24.69 | 42.28 | #5 | 35.75 | 23.51 | 18.98 | 16.42 | 12.13 | 28.46 | **28.34** | **13.93** | **7.91** | **4.64** | **2.95** | **7.59** |
| **FengFans** | #5 | 77.02 | 62.32 | 52.44 | 45.41 | 26.07 | 35.26 | #4 | 57.73 | 44.94 | 38.88 | 34.87 | 19.86 | 37.94 | 54.00 | 40.26 | 34.41 | 30.71 | 15.97 | 30.57 |
| **Depthforge_iitj** | #6 | 100.00 | 100.00 | 100.00 | 100.00 | 455.80 | 458.00 | #5 | 100.00 | 100.00 | 100.00 | 100.00 | 405.90 | 414.93 | 100.00 | 100.00 | 100.00 | 100.00 | 397.42 | 406.80 |

Table 1. **Stereo Track: Evaluation on the Challenge Test Set.** Results are reported at full resolution (4112×3008) across three pixel subsets: All, ToM (Transparent or Mirror), and Other materials. **Gold**, **silver**, and **bronze** indicate first, second, and third place respectively. Rankings are determined by **bad-2**, computed separately on ToM and All pixels.

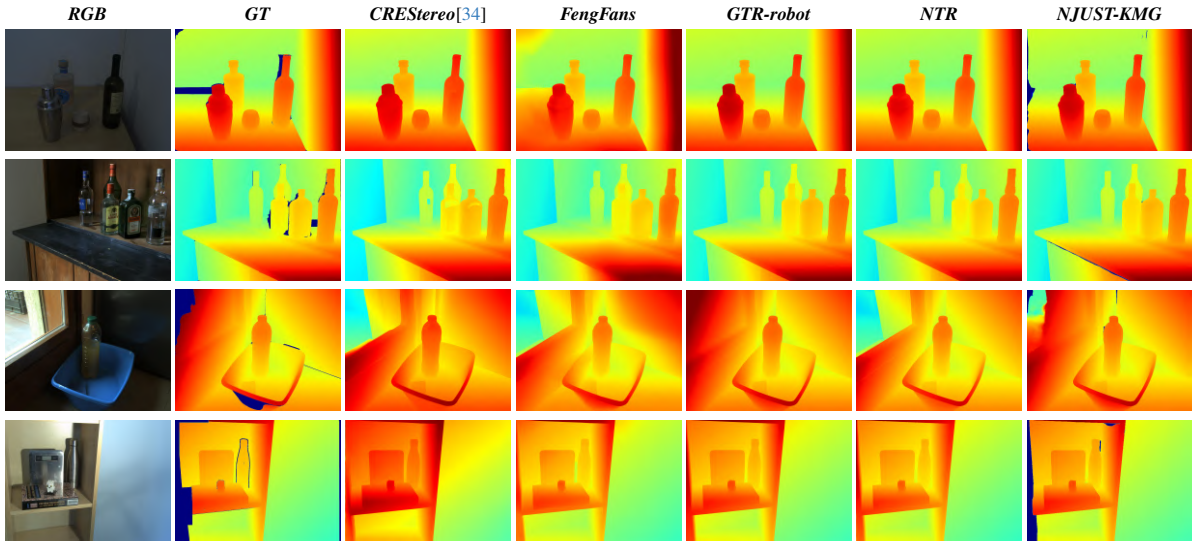


Figure 1. **Qualitative results – Stereo track.** From left to right: RGB reference image, ground-truth disparity, predictions by CREStereo [52], FengFans, GTR-robot, NTR, and NJUST-KMG.

all pixel categories. Fig. 1 depicts some qualitative results from the stereo testing set.

### 4.2. Track 2: Mono

Table 2 shows the results for the second track. At the bottom, we report the results achieved by the baseline method – i.e., the Unik3D [74] model using the weights provided by the authors. From left to right, we report deltas, Abs Rel., MAE, and RMSE metrics for *ToM*, *All*, and *Other* pixels, respectively. We report two different rankings, according to the performance observed on RMSE – the primary metric for this edition – computed over *ToM* and *All* pixels. The majority of submitted methods outperformed the Unik3D baseline, with the exception of **hello_w** and **DepthForgeI-ITJ**, the latter producing degenerate predictions reflected in extremely high error values across all pixel categories. Unlike the previous edition, no single method dominates across all pixel categories: **Lavreniuk** achieves the best RMSE on *ToM* pixels, while **AIIA-Lab** leads on *All* and *Other* pixels, also achieving the highest $\delta < 1.05$ accuracy across all categories. The top three methods – **Lavreniuk**, **AIIA-Lab**, and **MiVideo** – consistently outperform the baseline across all pixel categories, though absolute accuracy on *ToM* regions remains significantly lower than on *Other* pixels, confirming that transparent and mirror surfaces continue to represent the primary open challenge for monocular depth estimation. Fig. 2 shows some qualitative examples from the mono testing set.

## 5. Challenge Methods

### 5.1. Track 1: Stereo

#### 5.1.1 Baseline - CREStereo [52]

For the first track, we set the state-of-the-art CREStereo architecture [52] as our baseline. This model consists of a hierarchical network employing a recurrent refinement process, designed to update the predicted disparity map in coarse-to-fine manner. This process is implemented based on an adaptive group correlation layer (AGCL), where an alternate 2D-1D local search strategy with deformable windows is employed for robust matching even in the presence of imperfect rectification. The AGCL module computes correlations between pixels in local search windows, in contrast to what the all-pairs correlation module from RAFT-Stereo [59] does, reducing the computational requirements. To obtain the final predictions, we process images at quarter resolution using the original weights released by the authors, then we upsample predicted disparity maps to the

| | ToM | | | | | | | All | | | | | | | Other | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Team | Rank | $\delta < 1.05$ | $\delta < 1.15$ | $\delta < 1.25$ | Abs Rel. | MAE | RMSE | Rank | $\delta < 1.05$ | $\delta < 1.15$ | $\delta < 1.25$ | Abs Rel. | MAE | RMSE | $\delta < 1.05$ | $\delta < 1.15$ | $\delta < 1.25$ | Abs Rel. | MAE | RMSE |
| **Lavreniuk** | #1 | 38.04 | 74.65 | 90.91 | 8.33 | 7.98 | 8.22 | #2 | 34.74 | 77.98 | 89.02 | 8.97 | 10.96 | 11.97 | 32.91 | 78.69 | 89.15 | 9.12 | 11.29 | 12.36 |
| **AIIA-Lab** | #2 | 40.14 | 83.62 | 94.62 | 9.06 | 8.65 | 9.05 | #1 | 51.29 | 89.92 | 95.22 | 7.25 | 8.19 | 9.38 | 51.08 | 89.80 | 95.91 | 6.83 | 8.35 | 9.56 |
| **MiVideo** | #3 | 30.53 | 71.66 | 85.33 | 10.53 | 10.85 | 11.24 | #3 | 27.30 | 68.40 | 85.69 | 10.99 | 13.05 | 13.87 | 26.51 | 67.93 | 85.31 | 11.22 | 13.54 | 14.26 |
| **IIT JAMMU** | #4 | 34.96 | 66.68 | 85.97 | 12.87 | 11.95 | 13.04 | #4 | 21.20 | 58.18 | 83.06 | 14.45 | 15.87 | 18.46 | 19.05 | 59.86 | 86.04 | 13.39 | 15.55 | 17.81 |
| **tianjiajun** | #5 | 23.94 | 52.80 | 80.27 | 13.24 | 13.71 | 14.01 | #5 | 21.27 | 52.87 | 76.44 | 13.71 | 17.27 | 18.96 | 18.94 | 50.54 | 76.09 | 14.11 | 18.22 | 19.75 |
| **SNU-ISPL-B** | #6 | 17.28 | 56.67 | 83.76 | 13.46 | 13.48 | 14.44 | #6 | 20.19 | 57.55 | 75.66 | 14.11 | 17.51 | 20.49 | 18.44 | 55.85 | 74.02 | 14.69 | 18.96 | 21.83 |
| **NTR** | #7 | 22.15 | 48.71 | 67.72 | 16.34 | 16.98 | 18.72 | #7 | 14.70 | 42.06 | 63.82 | 18.69 | 24.47 | 28.49 | 15.31 | 43.06 | 62.19 | 19.83 | 27.08 | 30.44 |
| **CVIS-MIP** | #8 | 19.06 | 37.59 | 47.13 | 35.57 | 32.98 | 39.91 | #8 | 19.73 | 41.89 | 57.78 | 24.16 | 27.19 | 31.60 | 18.83 | 41.04 | 61.12 | 21.80 | 25.56 | 27.77 |
| **[Baseline]** | #9 | 7.19 | 41.30 | 64.32 | 25.00 | 26.03 | 26.67 | – | 7.53 | 41.50 | 73.51 | 20.00 | 23.73 | 25.60 | 7.27 | 46.21 | 79.29 | 16.00 | 18.72 | 19.36 |
| **hello_w** | #10 | 4.99 | 15.36 | 26.34 | 45.05 | 42.17 | 47.37 | #9 | 8.55 | 23.74 | 35.00 | 37.42 | 40.07 | 46.64 | 9.57 | 24.69 | 34.50 | 37.85 | 43.20 | 47.66 |
| **DepthForgeITJ** | #11 | 1.31 | 3.75 | 6.07 | 417.74 | 361.27 | 387.54 | #10 | 2.43 | 7.00 | 11.20 | 320.39 | 290.59 | 377.10 | 2.17 | 6.24 | 9.98 | 350.89 | 328.15 | 408.91 |

Table 2. **Mono Track: Evaluation on the NTIRE 2026 Challenge Test Set.** Metric monocular depth predictions assessed at full resolution (4112×3008) across three pixel subsets: All, ToM (Transparent or Mirror), and Other materials. Depth values are in centimeters; MAE and RMSE are expressed in cm, while Abs Rel. is reported as percentage (×100). **Gold**, **silver**, and **bronze** indicate first, second, and third place respectively. Rankings are determined by **RMSE** (cm), computed separately on ToM and All pixels.

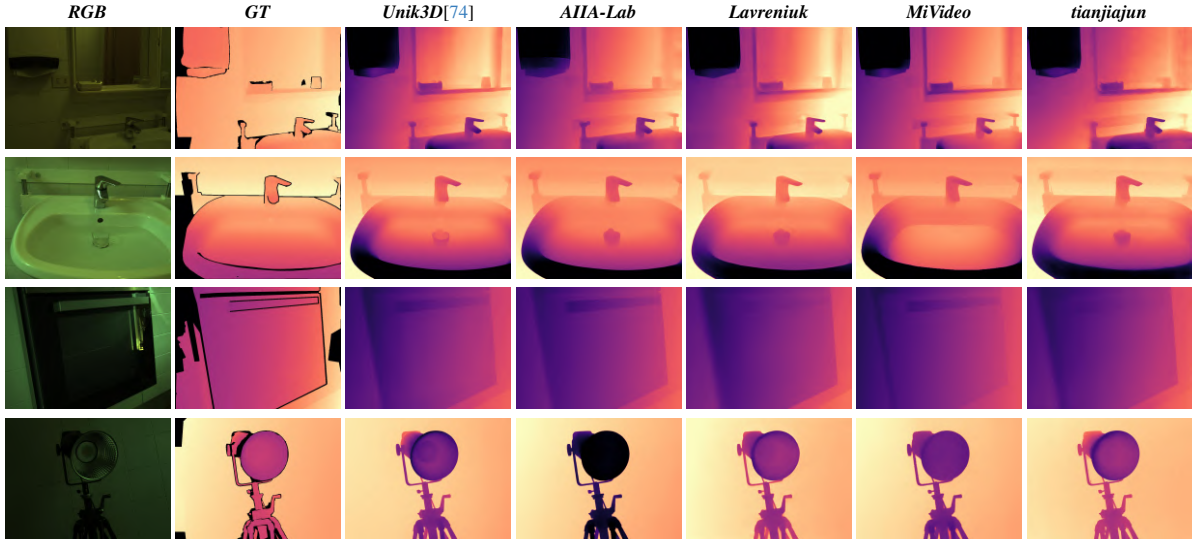


Figure 2. **Qualitative results – Mono track.** From left to right: RGB reference image, ground-truth disparity, predictions by Unik3D [74], AIIA-Lab, Lavreniuk, MiVideo, and tianjiajun.

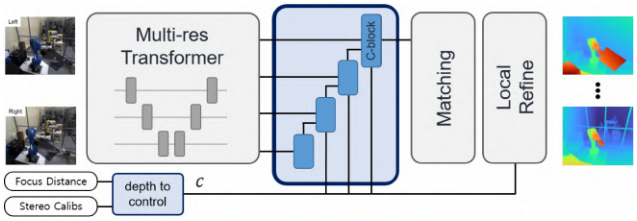


Figure 3. **Network Architecture – Team *GTR-robot*.**

original resolution through bilinear interpolation.

### 5.1.2 Team 1 - GTR-robot

The GTR-robot team (CodaLab: *Junhong*) presents DepthFocus [68], a steerable multi-layer stereo framework designed to address depth ambiguity in transparent and reflective scenes. DepthFocus builds upon the $S^2M^2$ baseline [67] and introduces an intent-driven steering mechanism controlled by a normalized scalar signal c ∈ [0, 1], derived from focus distance and stereo calibration parameters. This signal modulates a Multi-resolution Transformer backbone through Conditional Blocks (C-blocks), dynamically emphasizing features at the targeted depth layer while suppressing out-of-focus signals, before passing through cost-volume matching and local refinement. To bridge the Sim2Real gap, the team employs large-scale photorealistic synthetic data with overlapping transmissive layers and an auxiliary stereo segmentation task to provide structural priors to distinguish highly transmissive or highly reflective object boundaries from complex background textures. During inference, c is set to 0 to prioritize the nearest transmissive or reflective surface, and images are processed at half the resolution.

### 5.1.3 Team 2 - NTR

The NTR team (CodaLab: *miketjc0316*) applies FoundationStereo [127], a zero-shot stereo matching foundation model, directly to the high-resolution Booster stereo pairs without any fine-tuning. FoundationStereo employs a ViT-Large backbone pretrained with DINOv2 and a side-tuning adapter bridging monocular depth priors to stereo matching,

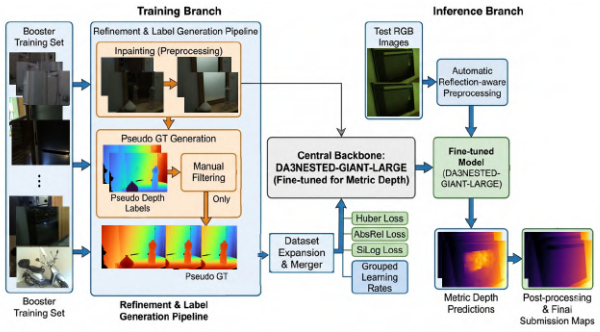


Figure 4. **Network Architecture – Team *AIIA-Lab*.**

with disparity iteratively refined over 32 GRU iterations. To handle the full 3008×4112 resolution within GPU memory, inputs are resized to a scale factor of 0.3 with hierarchical inference enabled, and the resulting disparity is upsampled back to the original resolution via bilinear interpolation.

#### 5.1.4 Team 3 - NJUST-KMG

The NJUST-KMG team (CodaLab: *lllllll07*) presents MonSter-DAv2, an improved MonSter architecture integrating Depth Anything V2 for high-resolution stereo depth estimation on specular and transparent surfaces. The method adopts a dual-branch design combining a monocular branch built on Depth Anything V2 for global structural priors with a stereo branch for fine-grained disparity estimation. Matching reliability in challenging regions is improved through a hybrid fusion of DPT-style transformer features and CNN-based local features. Inference is performed at quarter resolution to handle 12Mpx inputs efficiently, with further refinement to recover high-resolution detail.

### 5.2. Track 2: Mono

#### 5.2.1 Baseline - UniK3D [74]

For this second track, we set the UniK3D model as our baseline, a universal monocular 3D estimation framework. It predicts scene geometry by decomposing 3D structure into a per-pixel ray field and a metric radial distance, supporting arbitrary camera models through a dedicated angular branch. As for the Stereo track, we use the original weights made available by the authors.

#### 5.2.2 Team 1 - AIIA-Lab

The AIIA-Lab team (CodaLab: *jing27*) presents Fake It Till You Depth It (FITD), a reflection-aware fine-tuning approach built upon a pretrained Depth Anything v3 NESTED-GIANT-LARGE-1.1 model [58] for metric monocular depth estimation in reflective and transparent indoor scenes. Starting from the Booster training split, difficult scenes dominated by severe highlights or reflection

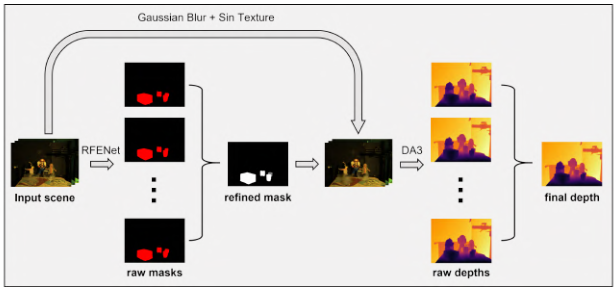


Figure 5. **Network Architecture – Team *MiVideo*.**

artifacts are identified and preprocessed via inpainting to construct an expanded training set. Pseudo ground truth generated from these enhanced samples is further manually filtered to retain only visually consistent and geometrically reasonable labels, and the final model is fine-tuned on the combination of original and curated augmented data. The optimization objective combines Huber loss, AbsRel loss, and SILog loss, with grouped learning rates applied to different parts of the network to stabilize fine-tuning and preserve the pretrained backbone's prior knowledge. At inference, an automatic robust preprocessing step is applied to each test image before feeding it into the fine-tuned model to produce the final metric depth map.

#### 5.2.3 Team 2 - Lavreniuk

The Lavreniuk team (CodaLab: *mlavreniuk*) presents DeepBlend, a method built upon Depth Anything V2 (DAv2) as the primary monocular depth estimator. For training, pseudo-depth labels are generated per surface type: transparent objects are handled via a modified blending procedure fusing RGB images with object masks, extending Costanzino et al. [17], while mirror regions undergo inpainting with a fast Fourier convolution-based model before pseudo-label generation. Depth Anything [133] and DAv2 are used as pseudo-label predictors, with their averaged outputs yielding more stable depth maps. The final model is fine-tuned in two stages: first on a mixture of datasets including TransCG, ClearGrasp, MIDepth, Hammer, HouseCat6D, MSD, Trans10K, and Booster; then for 10 additional epochs on Booster alone. The architecture follows the standard DAv2 design with a DINOv2 ViT encoder and DPT decoder, trained with SILog loss, augmented for illumination robustness, with horizontal flipping and color jittering applied at test time.

#### 5.2.4 Team 3 - MiVideo

The MiVideo team (CodaLab: *yanqm*) presents TextureDepth, a mask-guided preprocessing and ensemble strategy that improves monocular depth estimation on ToM surfaces without retraining the depth model. ToM masks predicted by RFENet are aggregated across illumination captures per

scene, retaining only consistently identified regions, and refined with lightweight manual correction. Masked regions are preprocessed with Gaussian smoothing followed by low-amplitude sinusoidal texture perturbations — zero-mean by design — to introduce directional structural cues while preserving global color statistics. The processed images are fed into a pretrained Depth Anything V3 model for metric depth prediction, refined with bilateral filtering, and aggregated across illumination captures via pixel-wise median ensembling.

#### 5.2.5 Team 4 - IIT JAMMU

The IIT JAMMU team (CodaLab: *divyanshu190*) proposes an ensemble framework combining Depth Anything V2 (ViT-L), fine-tuned on Booster for 50 epochs, with pretrained ZoeDepth NK. During inference, multi-scale test-time augmentation across 3 scales and 4 flips is applied to both models, and outputs are combined via a weighted ensemble (0.6/0.4). The ensembled depth map is aligned to metric scale through scale-shift estimation, with transparent regions further refined via inpainting and object boundaries sharpened using SAM2 segmentation masks.

#### 5.2.6 Team 5 - tianjiajun

The tianjiajun team (CodaLab: *divyanshu190*) adapts UniK3D into a calibration-aware metric depth estimator for specular and transparent surfaces. The DINOv2 ViT-L encoder and geometry-aware decoder predict scene geometry as a combination of a per-pixel ray field and a metric radial distance. Since the NTIRE benchmark provides calibrated pinhole cameras, the angular branch is bypassed and ground-truth rays condition the decoder directly, removing camera ambiguity. The model is fine-tuned for 50,000 iterations using AdamW at a base learning rate of $5\times10^{-6}$, batch size 4, with a ContextCrop strategy at 1504×2056 training resolution and full 3008×4112 at inference.

#### 5.2.7 Team 6 - SNU-ISPL-B

The SNU-ISPL-B team (CodaLab: *Kakarot*) presents FrozenZoe-HB, a fine-tuned ZoeDepth model adapted to Booster via a frozen-backbone strategy. The MiDaS DPT-BEiT-L-384 backbone is kept entirely frozen to preserve large-scale pretraining representations, while only the metric bins module is optimized: 64 adaptive bin centers are initialized, refined through four attractor layers, and decoded via a Conditional Log Binomial head. Training uses a single SILog loss with $\beta$ raised from 0.15 to 0.5, amplifying the global scale error penalty by $\sim$3.3×, for 10,000 iterations under a OneCycleLR schedule at a peak learning rate of $1.61\times10^{-4}$, batch size 16, on a single 3090 GPU.

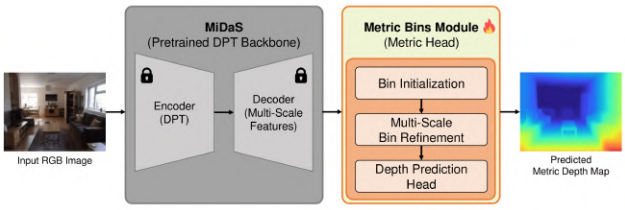


Figure 6. **Network Architecture – Team *SNU-ISPL-B*.**

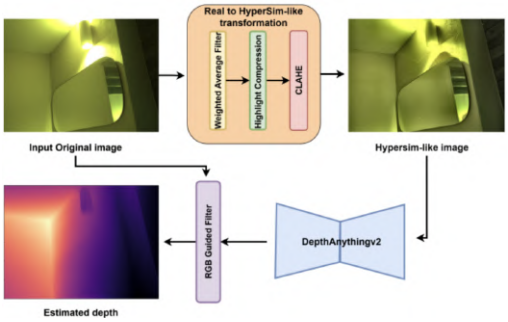


Figure 7. **Network Architecture – Team *CVIS-MIP*.**

#### 5.2.8 Team 7 - NTR (mono)

The NTR team (CodaLab: *miketjc*) combines Depth Anything V2 Large as an offline relative depth estimator with a diffusion-pretrained regression network for metric depth prediction. DAv2 pseudo-depth maps — scale- and shift-ambiguous but structurally robust on ToM surfaces — are concatenated with the RGB image to form a 6-channel input to a TimeDiffiT-ResNet U-Net ($\sim$142M parameters). The network is pretrained via Masked Diffusion Autoencoding (MDAE) with dual spatial masking and VE-SDE noise injection, then fine-tuned on Booster with material-aware MSE upweighting ToM pixels by 5×. At inference, overlapping 512×512 tiles with a stride of 448 are processed and refined through 8-fold geometric self-ensemble [106] and bilateral filtering.

#### 5.2.9 Team 8 - CVIS-MIP

The CVIS-MIP team presents Hypersimify-DA2, a training-free domain adaptation pipeline for monocular metric depth estimation on specular and transparent surfaces. The core idea is to map real input images to the HyperSim synthetic distribution, exploiting a Depth Anything v2 model fine-tuned on HyperSim where perfect ground-truth depth is available for transparent and specular objects. The preprocessing pipeline applies bilinear filtering for texture smoothing, highlight compression to remove saturated values (235–255) via inpainting, and CLAHE with 8×8 patches for local contrast enhancement. Predictions are upsampled to 3008×4112 via bilinear interpolation followed by an RGB-guided filter to preserve edge sharpness.

## Acknowledgments

This work was partially supported by the Humboldt Foundation. We thank the NTIRE 2026 sponsors: OPPO, Kuaishou, and the University of Wurzburg (Computer Vision Lab).

## A. NTIRE 2026 Organizers

Pierluigi Zama Ramirez[1] (pier.zamaramirez@unive.it), Alex Costanzino[2], Fabio Tosi[2], Matteo Poggi[2], Samuele Salti[2], Stefano Mattoccia[2], Luigi Di Stefano[2], Radu Timofte[3]
***Affiliations***:
[1] Ca' Foscari University of Venice, Italy
[2] University of Bologna, Italy
[3] Computer Vision Lab, University of Würzburg, Germany

## B. Track 1: Teams and Affiliations

### GTR-robot

***Members:***
Junhong Min[1] (junhong1.min@samsung.com)
***Affiliations:***
[1] Samsung Electronics, Global Technology Research

### NTR (stereo and mono)

***Members:***
Jiachen Tu[1] (jtu9@illinois.edu), Guoyi Xu[1] (ericx3@illinois.edu), Yaoxin Jiang[1] (yaoxinj2@illinois.edu), Jiajia Liu[1] (ciciliu2@illinois.edu), Yaokun Shi[1] (yaokuns2@illinois.edu)
***Affiliations:***
[1] University of Illinois Urbana-Champaign

### NJUST-KMG

***Members:***
Zuolingling Zuo[1] (zuolingling@njust.edu.cn), Zhihao Guan[1] (zhguan@njust.edu.cn), Yang Yang[1] (yyang@njust.edu.cn)
***Affiliations:***
[1] Nanjing University of Science and Technology, China

## C. Track 2: Teams and Affiliations

### AIIA-Lab

***Members:***
Xinglin Li[1] (xinglin.li@stu.hit.edu.cn), Jing Cao[1] (caojing@stu.hit.edu.cn), Zihao Zhang[1] (25S103633@stu.hit.edu.cn), Shiyu Liu[1] (24S003009@stu.hit.edu.cn), Kui Jiang[1] (jiangkui@hit.edu.cn)
***Affiliations:***
[1] Harbin Institute of Technology, China

### Lavreniuk

***Members:***
Mykola Lavreniuk[1] (nick_93@ukr.net)
***Affiliations:***
[1] Space Research Institute NASU-SSAU, Kyiv, Ukraine

### MiVideo

***Members:***
Qianming Yan[1] (yanqianming@xiaomi.com), Ziyi Wang[1] (wangziyi5@xiaomi.com), Peishuai Zha[1] (zhapeishuai@xiaomi.com)
***Affiliations:***
[1] Xiaomi Inc.

### IIT JAMMU

***Members:***
Divyanshu Kumawat[1] (iitjdivyanshu@gmail.com), Anshika Singh[1] (2024uch0006@iitjammu.ac.in), Utkarsh Yadav[1] (2024umt0189@iitjammu.ac.in), Karanbir Singh Dhindsa[1] (2024uch0021@iitjammu.ac.in), Vinit Jakhetiya[1] (vinit.jakhetiya@iitjammu.ac.in)
***Affiliations:***
[1] Indian Institute of Technology Jammu, India

### tianjiajun

***Members:***
Jiajun Tian[1] (tianjiajun1@insta360.com)
***Affiliations:***
[1] Insta360

### SNU-ISPL-B

***Members:***
Junhyeong Kwon[1] (gjh8760@snu.ac.kr), Youngjin Oh[1] (yjymoh0211@snu.ac.kr), Nam Ik Cho[1] (nicho@snu.ac.kr)
***Affiliations:***
[1] Seoul National University, South Korea

### CVIS-MIP

***Members:***
Hatem Ibrahem[1] (hatem.hosam1991@gmail.com), Ahmed Salem[2] (ahmeddiefy@gmail.com), Hyun-Soo Kang[3] (hskang@chungbuk.ac.kr), Guanghui Wang[1] (wangcs@torontomu.ca)
***Affiliations:***
[1] School of Computer Science, Toronto Metropolitan University
[2] Faculty of Engineering, Assiut University
[3] School of Electrical Engineering, Chungbuk National University

## References


[1] Filippo Aleotti, Fabio Tosi, Pierluigi Zama Ramirez, Matteo Poggi, Samuele Salti, Stefano Mattoccia, and Luigi Di Stefano. Neural disparity refinement for arbitrary resolution stereo. In *International Conference on 3D Vision*, 2021. 3DV.

[2] Radu Ancuti, Codruta Ancuti, Radu Timofte, and Cosmin Ancuti. NT-HAZE: A Benchmark Dataset for Realistic Night-time Image Dehazing . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[3] Radu Ancuti, Alexandru Brateanu, Florin Vasluianu, Raul Balmez, Ciprian Orhei, Codruta Ancuti, Radu Timofte, Cosmin Ancuti, et al. NTIRE 2026 Nighttime Image Dehazing Challenge Report . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[4] Luca Bartolomei, Fabio Tosi, Matteo Poggi, and Stefano Mattoccia. Stereo anywhere: Robust zero-shot deep stereo matching even where either stereo or mono fail. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[5] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023.

[6] Aleksei Bochkovskii, AmaÃĢl Delaunoy, Hugo Germain, Marcel Santos, Yichao Zhou, Stephan R Richter, and Vladlen Koltun. Depth pro: Sharp monocular metric depth in less than a second. *arXiv preprint arXiv:2410.02073*, 2024.

[7] Jie Cai, Kangning Yang, Zhiyuan Li, Florin Vasluianu, Radu Timofte, et al. NTIRE 2026 Challenge on Single Image Reflection Removal in the Wild: Datasets, Results, and Methods . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[8] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5410–5418, 2018.

[9] Weifeng Chen, Zhao Fu, Dawei Yang, and Jia Deng. Single-image depth perception in the wild. In *Proc. NeurIPS*, 2016.

[10] Zheng Chen, Kai Liu, Jingkai Wang, Xianglong Yan, Jianze Li, Ziqing Zhang, Jue Gong, Jiatong Li, Lei Sun, Xiaoyang Liu, Radu Timofte, Yulun Zhang, et al. The Fourth Challenge on Image Super-Resolution (×4) at NTIRE 2026: Benchmark Results and Method Overview . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[11] Junda Cheng, Longliang Liu, Gangwei Xu, Xianqi Wang, Zhaoxing Zhang, Yong Deng, Jinliang Zang, Yurui Chen, Zhipeng Cai, and Xin Yang. Monster: Marry monodepth to stereo unleashes power. *arXiv preprint arXiv:2501.08643*, 2025.

[12] Xinjing Cheng, Peng Wang, and Ruigang Yang. Learning depth with convolutional spatial propagation network. *IEEE transactions on pattern analysis and machine intelligence*, 42(10):2361–2379, 2019.

[13] Xuelian Cheng, Yiran Zhong, Mehrtash Harandi, Yuchao Dai, Xiaojun Chang, Hongdong Li, Tom Drummond, and Zongyuan Ge. Hierarchical neural architecture search for deep stereo matching. *Advances in Neural Information Processing Systems*, 33, 2020.

[14] Jaehoon Choi, Dongki Jung, Yonghan Lee, Deokhwa Kim, Dinesh Manocha, and Donghwan Lee. Selfdeco: Self-supervised monocular depth completion in challenging indoor environments. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 467–474. IEEE, 2021.

[15] George Ciubotariu, Sharif S M A, Abdur Rehman, Fayaz Ali, Rizwan Ali Naqvi, Marcos Conde, Radu Timofte, et al. Low Light Image Enhancement Challenge at NTIRE 2026 . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[16] George Ciubotariu, Zhuyun Zhou, Yeying Jin, Zongwei Wu, Radu Timofte, et al. High FPS Video Frame Interpolation Challenge at NTIRE 2026 . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[17] Alex Costanzino, Pierluigi Zama Ramirez, Matteo Poggi, Fabio Tosi, Stefano Mattoccia, and Luigi Di Stefano. Learning depth estimation for transparent and mirror surfaces. In *The IEEE International Conference on Computer Vision*, 2023. ICCV.

[18] Shivam Duggal, Shenlong Wang, Wei-Chiu Ma, Rui Hu, and Raquel Urtasun. Deeppruner: Learning efficient stereo matching via differentiable patchmatch. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4384–4393, 2019.

[19] Andrei Dumitriu, Aakash Ralhan, Florin Miron, Florin Tatui, Radu Tudor Ionescu, Radu Timofte, et al. NTIRE 2026 Rip Current Detection and Segmentation (RipDetSeg) Challenge Report . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[20] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Proc. NeurIPS*, 2014.

[21] Omar Elezabi, Marcos V. Conde, Zongwei Wu, Yeying Jin, Radu Timofte, et al. Photography Retouching Transfer, NTIRE 2026 Challenge: Report . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[22] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In *ECCV*, 2024.

[23] Adrien Gaidon, Greg Shakhnarovich, Rares Ambrus, Vitor Guizilini, Igor Vasiljevic, Matthew Walter, Sudeep Pillai, and Nick Kolkin. Dense depth for autonomous driving (DDAD) challenge (https://sites.google.com/view/mono3d-workshop), 2021.

[24] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Asian conference on computer vision*, pages 25–38. Springer, 2010.

[25] Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proc. CVPR*, 2017.

[26] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J Brostow. Digging into self-supervised monocular depth estimation. In *Proc. ICCV*, 2019.

[27] Bochen Guan, Jinlong Li, Kangning Yang, Chuang Ke, Jie Cai, Florin Vasluianu, Radu Timofte, et al. NTIRE 2026 Challenge on End-to-End Financial Receipt Restoration and Reasoning from Degraded Images: Datasets, Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[28] Ya-nan Guan, Shaonan Zhang, Hang Guo, Yawen Wang, Xinying Fan, Jie Liang, Hui Zeng, Guanyi Qin, Lishen Qu, Tao Dai, Shu-Tao Xia, Lei Zhang, Radu Timofte, et al. NTIRE 2026 The 3rd Restore Any Image Model (RAIM) Challenge: AI Flash Portrait (Track 3) . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[29] Weiyu Guo, Zhaoshuo Li, Yongkui Yang, Zheng Wang, Russell H Taylor, Mathias Unberath, Alan Yuille, and Yingwei Li. Context-enhanced stereo transformer. In *European Conference on Computer Vision*, pages 263–279. Springer, 2022.

[30] Xiaoyang Guo, Hongsheng Li, Shuai Yi, Jimmy Ren, and Xiaogang Wang. Learning monocular depth by distilling cross-domain stereo networks. In *Proc. ECCV*, 2018.

[31] Xiaoyang Guo, Kai Yang, Wukui Yang, Xiaogang Wang, and Hongsheng Li. Group-wise correlation stereo network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3273–3282, 2019.

[32] Aleksandr Gushchin, Khaled Abud, Ekaterina Shumitskaya, Artem Filippov, Georgii Bychkov, Sergey Lavrushkin, Mikhail Erofeev, Anastasia Antsiferova, Changsheng Chen, Shunquan Tan, Radu Timofte, Dmitriy Vatolin, et al. NTIRE 2026 Challenge on Robust AI-Generated Image Detection in the Wild . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[33] Jing He, Haodong Li, Wei Yin, Yixun Liang, Leheng Li, Kaiqiang Zhou, Hongbo Zhang, Bingbing Liu, and Ying-Cong Chen. Lotus: Diffusion-based visual foundation model for high-quality dense prediction. *arXiv preprint arXiv:2409.18124*, 2024.

[34] Ruozhen He, Jiaying Lin, and Rynson WH Lau. Efficient mirror detection via multi-level heterogeneous learning. *arXiv preprint arXiv:2211.15644*, 2022.

[35] Benedikt Hopf, Radu Timofte, et al. Robust Deepfake Detection, NTIRE 2026 Challenge: Report . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[36] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10579–10596, 2024.

[37] Wenbo Hu, Xiangjun Gao, Xiaoyu Li, Sijie Zhao, Xiaodong Cun, Yong Zhang, Long Quan, and Ying Shan. Depthcrafter: Generating consistent long depth sequences for open-world videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[38] Andrey Ignatov, Grigory Malivenko, David Plowman, Samarth Shukla, and Radu Timofte. Fast and accurate single-image depth estimation on mobile devices, mobile ai 2021 challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2545–2557, June 2021.

[39] Hualie Jiang, Zhiqiang Lou, Laiyan Ding, Rui Xu, Minglang Tan, Wenjie Jiang, and Rui Huang. Defom-stereo: Depth foundation model based stereo matching. *arXiv preprint arXiv:2501.09466*, 2025.

[40] Junpeng Jing, Jiankun Li, Pengfei Xiong, Jiangyu Liu, Shuaicheng Liu, Yichen Guo, Xin Deng, Mai Xu, Lai Jiang, and Leonid Sigal. Uncertainty guided adaptive warping for robust and efficient stereo matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3318–3327, October 2023.

[41] Adrian Johnston and Gustavo Carneiro. Self-supervised monocular trained depth estimation using self-attention and discrete disparity volume. In *Proc. CVPR*, 2020.

[42] Bingxin Ke, Dominik Narnhofer, Shengyu Huang, Lei Ke, Torben Peters, Katerina Fragkiadaki, Anton Obukhov, and Konrad Schindler. Video depth without video models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[43] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[44] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[45] Aleksei Khalin, Egor Ershov, Artem Panshin, Sergey Korchagin, Georgiy Lobarev, Arseniy Terekhin, Sofiia Dorogova, Amir Shamsutdinov, Yasin Mamedov, Bakhtiyar Khalfin, Bogdan Sheludko, Emil Zilyaev, Nikola Banić, Georgy Perevozchikov, Radu Timofte, et al. NTIRE 2026 Low-light Enhancement: Twilight Cowboy Challenge . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[46] Sameh Khamis, Sean Fanello, Christoph Rhemann, Adarsh Kowdle, Julien Valentin, and Shahram Izadi. Stereonet: Guided hierarchical refinement for real-time edge-aware depth prediction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 573–590, 2018.

[47] Henrik Kretzschmar, Alex Liniger, Jose M. Alvarez, Yan Wang, Vincent Casser, Fisher Yu, Marco Pavone, Bo Li, Andreas Geiger, Peter Ondruska, Li Erran Li, Dragomir Angelov, John Leonard, and Luc Van Gool. Argoverse stereo competition (https://cvpr2022.wad.vision/), 2021, 2022.

[48] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, and Nassir Navab. Deeper depth prediction with fully convolutional residual networks. In *2016 Fourth international conference on 3D vision (3DV)*, pages 239–248. IEEE, 2016.

[49] Mykola Lavreniuk and Alla Lavreniuk. Spidepth: Strengthened pose information for self-supervised monocular depth estimation. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR) Workshops*, pages 874–884, 2025.

[50] Jiahao Li, Xinhong Chen, Zhengmin Jiang, Qian Zhou, Yung-Hui Li, and Jianping Wang. Global regulation and excitation via attention tuning for stereo matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 25539–25549, 2025.

[51] Jiatong Li, Zheng Chen, Kai Liu, Jingkai Wang, Zihan Zhou, Xiaoyang Liu, Libo Zhu, Radu Timofte, Yulun Zhang, et al. The First Challenge on Mobile Real-World Image Super-Resolution at NTIRE 2026: Benchmark Results and Method Overview . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[52] Jiankun Li, Peisen Wang, Pengfei Xiong, Tao Cai, Ziwei Yan, Lei Yang, Jiangyu Liu, Haoqiang Fan, and Shuaicheng Liu. Practical stereo matching via cascaded recurrent network with adaptive correlation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16263–16272, 2022.

[53] Xin Li, Jiachao Gong, Xijun Wang, Shiyao Xiong, Bingchen Li, Suhang Yao, Chao Zhou, Zhibo Chen, Radu Timofte, et al. NTIRE 2026 Challenge on Short-form UGC Video Restoration in the Wild with Generative Models: Datasets, Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[54] Xin Li, Yeying Jin, Suhang Yao, Beibei Lin, Zhaoxin Fan, Wending Yan, Xin Jin, Zongwei Wu, Bingchen Li, Peishu Shi, Yufei Yang, Yu Li, Zhibo Chen, Bihan Wen, Robby Tan, Radu Timofte, et al. NTIRE 2026 The Second Challenge on Day and Night Raindrop Removal for Dual-Focused Images: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[55] Zhenyu Li, Shariq Farooq Bhat, and Peter Wonka. Patchfusion: An end-to-end tile-based framework for high-resolution monocular metric depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[56] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6197–6206, 2021.

[57] Zhengfa Liang, Yiliu Feng, Yulan Guo, Hengzhu Liu, Wei Chen, Linbo Qiao, Li Zhou, and Jianfeng Zhang. Learning for disparity estimation through feature constancy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[58] Haotong Lin, Sili Chen, Jun Hao Liew, Donny Y. Chen, Zhenyu Li, Yang Zhao, Sida Peng, Hengkai Guo, Xiaowei Zhou, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth anything 3: Recovering the visual space from any views. In *The Fourteenth International Conference on Learning Representations*, 2026.

[59] Lahav Lipson, Zachary Teed, and Jia Deng. RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching. *arXiv preprint arXiv:2109.07547*, 2021.

[60] Kai Liu, Haoyang Yue, Zeli Lin, Zheng Chen, Jingkai Wang, Jue Gong, Radu Timofte, Yulun Zhang, et al. The First Challenge on Remote Sensing Infrared Image Super-Resolution at NTIRE 2026: Benchmark Results and Method Overview . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[61] Shuhong Liu, Ziteng Cui, Chenyu Bao, Xuangeng Chu, Lin Gu, Bin Ren, Radu Timofte, Marcos V. Conde, et al. 3D Restoration and Reconstruction in Adverse Conditions: RealX3D Challenge Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[62] Xiaohong Liu, Xiongkuo Min, Guangtao Zhai, Qiang Hu, Jiezhang Cao, Yu Zhou, Wei Sun, Farong Wen, Zitong Xu, Yingjie Zhou, Huiyu Duan, Lu Liu, Jiarui Wang, Siqi Luo, Chunyi Li, Li Xu, Zicheng Zhang, Yue Shi, Yubo Wang, Minghong Zhang, Chunchao Guo, Zhichao Hu, Mingtao Chen, Xiele Wu, Xin Ma, Zhaohe Lv, Yuanhao Xue, Jiaqi Wang, Xinxing Sha, Radu Timofte, et al. NTIRE 2026 X-AIGC Quality Assessment Challenge: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[63] Jieming Lou, Weide Liu, Zhuo Chen, Fayao Liu, and Jun Cheng. Elfnet: Evidential local-global fusion for stereo matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17784–17793, 2023.

[64] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[65] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[66] S Mahdi H Miangoleh, Sebastian Dille, Long Mai, Sylvain Paris, and Yagiz Aksoy. Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9685–9694, 2021.

[67] Junhong Min, Youngpil Jeon, Jimin Kim, and Minyong Choi. $S^2M^2$: Scalable stereo matching model for reliable depth estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

[68] Junhong Min, Jimin Kim, Minwook Kim, Cheol-Hui Min, Youngpil Jeon, and Minyong Choi. DepthFocus: Controllable depth estimation for see-through scenes. In *Proceed-*

*ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[69] Andrey Moskalenko, Alexey Bryncev, Ivan Kosmynin, Kira Shilovskaya, Mikhail Erofeev, Dmitry Vatolin, Radu Timofte, et al. NTIRE 2026 Challenge on Video Saliency Prediction: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[70] Anton Obukhov, Matteo Poggi, Fabio Tosi, Ripudaman Singh Arora, Jaime Spencer, Chris Russell, Simon Hadfield, Richard Bowden, Shuaihang Wang, Zhenxin Ma, Weijie Chen, Baobei Xu, Fengyu Sun, Di Xie, Jiang Zhu, Mykola Lavreniuk, Haining Guan, Qun Wu, Yupei Zeng, Chao Lu, Huanran Wang, Guangyuan Zhou, Haotian Zhang, Jianxiong Wang, Qiang Rao, Chunjie Wang, Xiao Liu, Zhiqiang Lou, Hualie Jiang, Yihao Chen, Rui Xu, Minglang Tan, Zihan Qin, Yifan Mao, Jiayang Liu, Jialei Xu, Yifan Yang, Wenbo Zhao, Junjun Jiang, Xianming Liu, Mingshuai Zhao, Anlong Ming, Wu Chen, Feng Xue, Mengying Yu, Shida Gao, Xiangfeng Wang, Gbenga Omotara, Ramy Farag, Jacket Demby's, Seyed Mohamad Ali Tousi, Guilherme N. DeSouza, Tuan-Anh Yang, Minh-Quang Nguyen, Thien-Phuc Tran, Albert Luginov, and Muhammad Shahzad. The fourth monocular depth estimation challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2025.

[71] Jiahao Pang, Wenxiu Sun, Jimmy SJ. Ren, Chengxi Yang, and Qiong Yan. Cascade residual learning: A two-stage convolutional neural network for stereo matching. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[72] Hyunhee Park, Eunpil Park, Sangmin Lee, Radu Timofte, et al. NTIRE 2026 Challenge on Efficient Burst HDR and Restoration: Datasets, Methods, and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[73] Georgy Perevozchikov, Daniil Vladimirov, Radu Timofte, et al. NTIRE 2026 Challenge on Learned Smartphone ISP with Unpaired Data: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[74] Luigi Piccinelli, Christos Sakaridis, Mattia Segu, Yung-Hsu Yang, Siyuan Li, Wim Abbeloos, and Luc Van Gool. UniK3D: Universal camera monocular 3d estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[75] Luigi Piccinelli, Christos Sakaridis, Yung-Hsu Yang, Mattia Segu, Siyuan Li, Wim Abbeloos, and Luc Van Gool. Unidepthv2: Universal monocular metric depth estimation made simpler. *arXiv preprint arXiv:2502.20110*, 2025.

[76] Luigi Piccinelli, Yung-Hsu Yang, Christos Sakaridis, Mattia Segu, Siyuan Li, Luc Van Gool, and Fisher Yu. Unidepth: Universal monocular metric depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10106–10116, 2024.

[77] Matteo Poggi, Filippo Aleotti, Fabio Tosi, and Stefano Mattoccia. On the uncertainty of self-supervised monocular depth estimation. In *Proc. CVPR*, 2020.

[78] Matteo Poggi and Fabio Tosi. Federated online adaptation for deep stereo. In *CVPR*, 2024.

[79] Guanyi Qin, Jie Liang, Bingbing Zhang, Lishen Qu, Ya-nan Guan, Hui Zeng, Lei Zhang, Radu Timofte, et al. NTIRE 2026 The 3rd Restore Any Image Model (RAIM) Challenge: Professional Image Quality Assessment (Track 1) . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[80] Xingyu Qiu, Yuqian Fu, Jiawei Geng, Bin Ren, Jiancheng Pan, Zongwei Wu, Hao Tang, Yanwei Fu, Radu Timofte, Nicu Sebe, Mohamed Elhoseiny, et al. The Second Challenge on Cross-Domain Few-Shot Object Detection at NTIRE 2026: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[81] Lishen Qu, Yao Liu, Jie Liang, Hui Zeng, Wen Dai, Ya-nan Guan, Guanyi Qin, Shihao Zhou, Jufeng Yang, Lei Zhang, Radu Timofte, et al. NTIRE 2026 The 3rd Restore Any Image Model (RAIM) Challenge: Multi-Exposure Image Fusion in Dynamic Scenes (Track2) . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[82] Michael Ramamonjisoa, Yuming Du, and Vincent Lepetit. Predicting sharp and accurate occlusion boundaries in monocular depth estimation using displacement fields. In *Proc. CVPR*, 2020.

[83] Pierluigi Zama Ramirez, Alex Costanzino, Fabio Tosi, Matteo Poggi, Luigi Di Stefano, Jean-Baptiste Weibel, Doris Antensteiner, Markus Vincze, Benjamin Busam, Guangyao Zhai, Weihang Li, Junwen Huang, Hyunjun Jung, Mykola Lavreniuk, Pihai Sun, Yijun Luo, Hongtao Wang, Manying Gao, Kui Jiang, and Junjun Jiang. Tricky 2025 challenge on monocular depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 3311–3322, October 2025.

[84] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, Jun Shi, Dafeng Zhang, et al. Ntire 2023 challenge on hr depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1384–1395, 2023.

[85] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, Yangyang Zhang, Cailin Wu, Zhuangda He, Shuangshuang Yin, Jiaxu Dong, Yangchenxu Liu, Hao Jiang, Jun Shi, Yong A, Yixiang Jin, Dingzhe Li, Bingxin Ke, Anton Obukhov, Tinafu Wang, Nando Metzger, Shengyu Huang, Konrad Schindler, Yachuan Huang, Jiaqi Li, Junrui Zhang, Yiran Wang, Zihao Huang, Tianqi Liu, Zhiguo Cao, Pengzhi Li, Jui-Lin Wang, Wenjie Zhu, Hui Geng, Yuxin Zhang, Long Lan, Kele Xu, Tao Sun, Qisheng Xu, Sourav Saini, Aashray Gupta, Sahaj K. Mistry, Aryan Shukla, Vinit Jakhetiya, Sunil Jaiswal, Yuejin Sun, Zhuofan Zheng, Yi Ning, Jen-Hao Cheng, Hou-I Liu, Hsiang-Wei Huang, Cheng-Yen Yang, Zhongyu Jiang, Yi-Hao Peng, Aishi Huang, and Jenq-Neng Hwang. Ntire 2024 challenge on hr depth from images of specular

and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6499–6512, June 2024.

[86] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, Zhe Zhang, Yang Yang, et al. Ntire 2025 challenge on hr depth from images of specular and transparent surfaces. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 987–1001, 2025.

[87] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. *ICCV*, 2021.

[88] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3), 2022.

[89] Bin Ren, Hang Guo, Yan Shu, Jiaqi Ma, Ziteng Cui, Shuhong Liu, Guofeng Mei, Lei Sun, Zongwei Wu, Fahad Shahbaz Khan, Salman Khan, Radu Timofte, Yawei Li, et al. The Eleventh NTIRE 2026 Efficient Super-Resolution Challenge Report . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[90] Tonmoy Saikia, Yassine Marrakchi, Arber Zela, Frank Hutter, and Thomas Brox. Autodispnet: Improving disparity estimation with automl. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1812–1823, 2019.

[91] Shreeyak Sajjan, Matthew Moore, Mike Pan, Ganesh Nagaraja, Johnny Lee, Andy Zeng, and Shuran Song. Clear grasp: 3d shape estimation of transparent objects for manipulation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3634–3642. IEEE, 2020.

[92] 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*, pages 31–42. Springer, 2014.

[93] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3260–3269. IEEE, 2017.

[94] Tim Seizinger, Florin-Alexandru Vasluianu, Marcos V. Conde, Jeffrey Chen, Zhuyun Zhou, Zongwei Wu, Radu Timofte, et al. The First Controllable Bokeh Rendering Challenge at NTIRE 2026 . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[95] Jiahao Shao, Yuanbo Yang, Hongyu Zhou, Youmin Zhang, Yujun Shen, Vitor Guizilini, Yue Wang, Matteo Poggi, and Yiyi Liao. Learning temporally consistent video depth from video diffusion priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[96] Zhelun Shen, Yuchao Dai, and Zhibo Rao. Cfnet: Cascade and fused cost volume for robust stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13906–13915, June 2021.

[97] Xiao Song, Xu Zhao, Hanwen Hu, and Liangji Fang. Edgestereo: A context integrated residual pyramid network for stereo matching. In *ACCV*, 2018.

[98] Jaime Spencer, C. Stella Qian, Chris Russell, Simon Hadfield, Erich Graf, Wendy Adams, Andrew J. Schofield, James H. Elder, Richard Bowden, Heng Cong, Stefano Mattoccia, Matteo Poggi, Zeeshan Khan Suri, Yang Tang, Fabio Tosi, Hao Wang, Youmin Zhang, Yusheng Zhang, and Chaoqiang Zhao. The monocular depth estimation challenge. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 623–632, January 2023.

[99] Jaime Spencer, C. Stella Qian, Michaela Trescakova, Chris Russell, Simon Hadfield, Erich Graf, Wendy Adams, Andrew J. Schofield, James Elder, Richard Bowden, Ali Anwar, Hao Chen, Xiaozhi Chen, Kai Cheng, Yuchao Dai, Huynh Thai Hoa, Sadat Hossain, Jianmian Huang, Mohan Jing, Bo Li, Chao Li, Baojun Li, Zhiwen Liu, Stefano Mattoccia, Siegfried Mercelis, Myungwoo Nam, Matteo Poggi, Xiaohua Qi, Jiahui Ren, Yang Tang, Fabio Tosi, Linh Trinh, S M Nadim Uddin, Khan Muhammad Umair, Kaixuan Wang, Yufei Wang, Yixing Wang, Mochu Xiang, Guangkai Xu, Wei Yin, Jun Yu, Qi Zhang, and Chaoqiang Zhao. The second monocular depth estimation challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023.

[100] Jaime Spencer, Fabio Tosi, Matteo Poggi, Ripudaman Singh Arora, Chris Russell, Simon Hadfield, Richard Bowden, GuangYuan Zhou, ZhengXin Li, Qiang Rao, YiPing Bao, Xiao Liu, Dohyeong Kim, Jinseong Kim, Myunghyun Kim, Mykola Lavreniuk, Rui Li, Qing Mao, Jiang Wu, Yu Zhu, Jinqiu Sun, Yanning Zhang, Suraj Patni, Aradhye Agarwal, Chetan Arora, Pihai Sun, Kui Jiang, Gang Wu, Jian Liu, Xianming Liu, Junjun Jiang, Xidan Zhang, Jianing Wei, Fangjun Wang, Zhiming Tan, Jiabao Wang, Albert Luginov, Muhammad Shahzad, Seyed Hosseini, Aleksander Trajcevski, and James H. Elder. The third monocular depth estimation challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2024.

[101] Lei Sun, Hang Guo, Bin Ren, Shaolin Su, Xian Wang, Danda Pani Paudel, Luc Van Gool, Radu Timofte, Yawei Li, et al. The Third Challenge on Image Denoising at NTIRE 2026: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[102] Lei Sun, Weilun Li, Xian Wang, Zhendong Li, Letian Shi, Dannong Xu, Deheng Zhang, Mengshun Hu, Shuang Guo, Shaolin Su, Radu Timofte, Danda Pani Paudel, Luc Van Gool, et al. The Second Challenge on Event-Based Image Deblurring at NTIRE 2026: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[103] Lei Sun, Xiaolong Qian, Qi Jiang, Xian Wang, Yao Gao, Kailun Yang, Kaiwei Wang, Radu Timofte, Danda

Pani Paudel, Luc Van Gool, et al. NTIRE 2026 The First Challenge on Blind Computational Aberration Correction: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[104] Vladimir Tankovich, Christian Hane, Yinda Zhang, Adarsh Kowdle, Sean Fanello, and Sofien Bouaziz. Hitnet: Hierarchical iterative tile refinement network for real-time stereo matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14362–14372, June 2021.

[105] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision*, pages 402–419. Springer, 2020.

[106] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1865–1873, 2016.

[107] Alessio Tonioni, Fabio Tosi, Matteo Poggi, Stefano Mattoccia, and Luigi Di Stefano. Real-time self-adaptive deep stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 195–204, 2019.

[108] Fabio Tosi, Filippo Aleotti, Matteo Poggi, and Stefano Mattoccia. Learning monocular depth estimation infusing traditional stereo knowledge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[109] Fabio Tosi, Filippo Aleotti, Pierluigi Zama Ramirez, Matteo Poggi, Samuele Salti, Luigi Di Stefano, and Stefano Mattoccia. Distilled semantics for comprehensive scene understanding from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[110] Fabio Tosi, Filippo Aleotti, Pierluigi Zama Ramirez, Matteo Poggi, Samuele Salti, Stefano Mattoccia, and Luigi Di Stefano. Neural Disparity Refinement . *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 46(12):8900–8917, 2024.

[111] Fabio Tosi, Luca Bartolomei, and Matteo Poggi. A survey on deep stereo matching in the twenties. *International Journal of Computer Vision*, pages 1–32, 2025.

[112] Fabio Tosi, Yiyi Liao, Carolin Schmitt, and Andreas Geiger. Smd-nets: Stereo mixture density networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[113] Fabio Tosi, Alessio Tonioni, Daniele De Gregorio, and Matteo Poggi. Nerf-supervised deep stereo. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 855–866, June 2023.

[114] Fabio Tosi, Pierluigi Zama Ramirez, and Matteo Poggi. Diffusion models for monocular depth estimation: Overcoming challenging conditions. In *European Conference on Computer Vision (ECCV)*, 2024.

[115] Florin-Alexandru Vasluianu, Tim Seizinger, Jeffrey Chen, Zhuyun Zhou, Zongwei Wu, Radu Timofte, et al. Learning-Based Ambient Lighting Normalization: NTIRE 2026 Challenge Results and Findings . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[116] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Radu Timofte, et al. Advances in Single-Image Shadow Removal: Results from the NTIRE 2026 Challenge . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[117] Jingkai Wang, Jue Gong, Zheng Chen, Kai Liu, Jiatong Li, Yulun Zhang, Radu Timofte, et al. The Second Challenge on Real-World Face Restoration at NTIRE 2026: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[118] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Sida Peng, Ye Zhang, Radu Timofte, Minglin Chen, Yi Wang, Qibin Hu, Wenjie Lei, et al. NTIRE 2026 Challenge on 3D Content Super-Resolution: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[119] Lijun Wang, Jianming Zhang, Yifan Wang, Huchuan Lu, and Xiang Ruan. CLIFFNet for monocular depth estimation with hierarchical embedding loss. In *Proc. ECCV*, 2020.

[120] Ruicheng Wang, Sicheng Xu, Cassie Dai, Jianfeng Xiang, Yu Deng, Xin Tong, and Jiaolong Yang. Moge: Unlocking accurate monocular geometry estimation for open-domain images with optimal training supervision. *arXiv preprint arXiv:2410.19115*, 2024.

[121] Ruicheng Wang, Sicheng Xu, Yue Dong, Yu Deng, Jianfeng Xiang, Zelong Lv, Guangzhong Sun, Xin Tong, and Jiaolong Yang. Moge-2: Accurate monocular geometry with metric scale and sharp details. *arXiv preprint arXiv:2507.02546*, 2025.

[122] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20697–20709, 2024.

[123] Xianqi Wang, Hao Yang, Hangtian Wang, Junda Cheng, Gangwei Xu, Min Lin, and Xin Yang. Promptstereo: Zero-shot stereo matching via structure and motion prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[124] Yan Wang, Zihang Lai, Gao Huang, Brian H Wang, Laurens Van Der Maaten, Mark Campbell, and Kilian Q Weinberger. Anytime stereo image depth estimation on mobile devices. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5893–5900, 2019.

[125] Yingqian Wang, Zhengyu Liang, Fengyuan Zhang, Wending Zhao, Longguang Wang, Juncheng Li, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2026 Challenge on Light Field Image Super-Resolution: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[126] Jamie Watson, Michael Firman, Gabriel J. Brostow, and Daniyar Turmukhambetov. Self-supervised monocular depth hints. In *Proc. ICCV*, 2019.

[127] Bowen Wen, Matthew Trepte, Joseph Aribido, Jan Kautz, Orazio Gallo, and Stan Birchfield. Foundationstereo: Zero-shot stereo matching. *arXiv*, 2025.

[128] 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*, pages 21919–21928, 2023.

[129] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofighi, Fisher Yu, Dacheng Tao, and Andreas Geiger. Unifying flow, stereo and depth estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[130] Jiebin Yan, Chenyu Tu, Qinghua Lin, Zongwei WU, Weixia Zhang, Zhihua Wang, Peibei Cao, Yuming Fang, Xiaoning Liu, Zhuyun Zhou, Radu Timofte, et al. Efficient Low Light Image Enhancement: NTIRE 2026 Challenge Report . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[131] Gengshan Yang, Joshua Manela, Michael Happold, and Deva Ramanan. Hierarchical deep stereo matching on high-resolution images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5515–5524, 2019.

[132] Guorun Yang, Hengshuang Zhao, Jianping Shi, Zhidong Deng, and Jiaya Jia. Segstereo: Exploiting semantic information for disparity estimation. In *ECCV*, pages 636–651, 2018.

[133] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*, 2024.

[134] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *Advances in Neural Information Processing Systems*, 37:21875–21911, 2024.

[135] Chengtang Yao, Lidong Yu, Zhidan Liu, Jiaxi Zeng, Yuwei Wu, and Yunde Jia. Diving into the fusion of monocular priors for generalized stereo matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14887–14897, 2025.

[136] Wei Yin, Chi Zhang, Hao Chen, Zhipeng Cai, Gang Yu, Kaixuan Wang, Xiaozhi Chen, and Chunhua Shen. Metric3d: Towards zero-shot metric 3d prediction from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9043–9053, 2023.

[137] Wei Yin, Jianming Zhang, Oliver Wang, Simon Niklaus, Long Mai, Simon Chen, and Chunhua Shen. Learning to recover 3d scene shape from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 204–213, 2021.

[138] Zhichao Yin, Trevor Darrell, and Fisher Yu. Hierarchical discrete distribution decomposition for match density estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6044–6053, 2019.

[139] Pierluigi Zama Ramirez, Alex Costanzino, Fabio Tosi, Matteo Poggi, Luigi Di Stefano, Jean-Baptiste Weibel, Dominik Bauer, Doris Antensteiner, Markus Vincze, Jiaqi Li, Yachuan Huang, Junrui Zhang, Yiran Wang, Jinghong Zheng, Liao Shen, Zhiguo Cao, Ziyang Song, Zerong Wang, Ruijie Zhu, Hao Zhang, Rui Li, Jiang Wu, Xian Li, Yu Zhu, Jinqiu Sun, Yanning Zhang, Pihai Sun, Yuanqi Yao, Wenbo Zhao, Kui Jiang, Junjun Jiang, Mykola Lavreniuk, and Jui-Lin Wang. Tricky 2024 challenge on monocular depth from images of specular and transparent surfaces. In *European Conference on Computer Vision Workshops*, 2024. ECCVW.

[140] Pierluigi Zama Ramirez, Alex Costanzino, Fabio Tosi, Matteo Poggi, Samuele Salti, Luigi Di Stefano, and Stefano Mattoccia. Booster: a benchmark for depth from images of specular and transparent surfaces. *arXiv preprint arXiv:2301.08245*, 2023.

[141] Pierluigi Zama Ramirez, Matteo Poggi, Fabio Tosi, Stefano Mattoccia, and Luigi Di Stefano. Geometry meets semantics for semi-supervised monocular depth estimation. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, pages 298–313. Springer, 2019.

[142] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, et al. NTIRE 2026 Challenge on High-Resolution Depth of non-Lambertian Surfaces . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[143] Pierluigi Zama Ramirez, Fabio Tosi, Matteo Poggi, Samuele Salti, Stefano Mattoccia, and Luigi Di Stefano. Open challenges in deep stereo: The booster dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21168–21178, June 2022.

[144] Oliver Zendel, Angela Dai, Xavier Puig Fernandez, Andreas Geiger, Vladen Koltun, Peter Kontschieder, Adam Kortylewski, Tsung-Yi Lin, Torsten Sattler, Daniel Scharstein, Hendrik Schilling, Jonas Uhrig, and Jonas Wulff. The robust vision challenge (http://www.robustvision.net/), 2018, 2020, 2022.

[145] Jiaxi Zeng, Chengtang Yao, Lidong Yu, Yuwei Wu, and Yunde Jia. Parameterized cost volume for stereo matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 18347–18357, October 2023.

[146] Feihu Zhang, Victor Prisacariu, Ruigang Yang, and Philip HS Torr. GA-Net: Guided aggregation net for end-to-end stereo matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[147] Chaoqiang Zhao, Matteo Poggi, Fabio Tosi, Lei Zhou, Qiyu Sun, Yang Tang, and Stefano Mattoccia. Gasmono: Geometry-aided self-supervised monocular depth estimation for indoor scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16209–16220, 2023.

[148] Yan Zhong, Qiufang Ma, Zhen Wang, Tingting Jiang, Radu Timofte, et al. NTIRE 2026 Challenge Report on Anomaly Detection of Face Enhancement for UGC Images . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.

[149] Chao Zhou, Hong Zhang, Xiaoyong Shen, and Jiaya Jia. Unsupervised learning of stereo matching. In *The IEEE International Conference on Computer Vision (ICCV)*. IEEE, October 2017.

[150] Jingyi Zhou, Haoyu Zhang, Jiakang Yuan, Peng Ye, Tao

Chen, Hao Jiang, Meiya Chen, and Yangyang Zhang. All-in-one: Transferring vision foundation models into stereo matching. In *Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence (AAAI 2025)*, Philadelphia, Pennsylvania, USA, February 2025.

[151] Wenbin Zou, Tianyi Liu, Kejun Wu, Huiping Zhuang, Zongwei Wu, Zhuyun Zhou, Radu Timofte, et al. NTIRE 2026 Challenge on Bitstream-Corrupted Video Restoration: Methods and Results . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2026.