GIFT: Geometry-Invariant Fine-Tuning for Non-Lambertian Monocular Depth Estimation

Xianghui Fan[1,2], Zhaoyu Chen[3], Bingqian Wu[4], Dayu Li[4], Xin Zeng[1,2], Cui Huanran[1,2], Guangzhen Xu[1,2], Xiangru Huang[4], Hang Yang[1,2]*

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]Fudan University
[4]Westlake University

Abstract

Monocular depth foundation models, benefiting from large-scale synthetic training data, have demonstrated strong generalization. However, they often hallucinate depth on non-Lambertian surfaces, estimating reflected content in mirrors or transmitted content behind glass rather than the physical surface itself. Adapting these models with real-world data is challenging because conventional depth sensors are also unreliable in such regions. We observe that while the appearance of a non-Lambertian surface varies with its reflected or transmitted environment, its underlying geometry remains unchanged. Based on this observation, we propose GIFT (Geometry-Invariant Fine-Tuning), a parameter-efficient post-training framework that requires no measured depth labels. We collect groups of RGB images under controlled appearance changes while keeping the camera and target geometry fixed. GIFT exploits geometric invariance across these observations to suppress non-Lambertian depth hallucinations while retaining general depth estimation capability. We further construct a controlled benchmark that evaluates non-Lambertian depth recovery, robustness to appearance changes, and performance retention in other regions. Experiments on our benchmark and an independent real-world dataset demonstrate that GIFT improves depth prediction for mirrors and transparent objects while largely preserving the base model's performance, providing a practical and low-cost approach for adapting monocular depth foundation models to non-Lambertian scenes.

Introduction

Monocular depth estimation has progressed from dataset-specific predictors to models that transfer across diverse scenes and, in some cases, recover metric depth without camera metadata (Ranftl et al. 2022; Yang et al. 2024b; Bochkovskiy et al. 2025). This progress nevertheless relies on regularities between image appearance and physical geometry. Mirrors and transparent surfaces violate these regularities: the observed radiance may describe a reflected or transmitted scene rather than the surface that generated the image. A monocular model may therefore place a reflected room behind a wall or recover the background behind a glass pane instead of the pane itself. The same materials also cause substantial errors in commonly used depth measurements, making both inference and supervision difficult (Tan et al. 2021; Zama Ramirez et al. 2024; Jiang et al. 2024).

Recent methods mainly replace unreliable measured depth with constructed supervision. Some rely on synthetic or generated non-Lambertian data (Zhang et al. 2025; Tosi, Zama Ramirez, and Poggi 2024; Wen et al. 2025); DKT further adapts a video diffusion model using synthetic RGB–depth pairs for transparent-object depth estimation (Xu et al. 2025). Others modify the input appearance: Depth4ToM fills masked non-Lambertian regions to construct virtual labels, while generative opacification methods replace transparent objects with opaque counterparts (Costanzino et al. 2023; Wang et al. 2026). However, these approaches may suffer from synthetic-to-real gaps or geometry distortions introduced by image generation and editing, and dedicated generation modules can add training or inference cost.

A simple physical observation motivates our approach: although the appearance of a non-Lambertian object can vary substantially with its environment, its underlying geometry remains unchanged. Accordingly, a depth foundation model should produce stable geometry predictions for the same object under such appearance variations. We therefore intervene on the reflected or transmitted content to alter the target appearance while preserving its geometry, and minimize the discrepancy between the resulting depth predictions. This cross-appearance disagreement provides a depth-GT-free self-supervised signal for adapting the model to non-Lambertian surfaces.

Based on this principle, we introduce GIFT (Geometry-Invariant Fine-Tuning), a fine-tuning framework for non-Lambertian depth estimation, together with the GIFT-Intervention dataset (GIFT-I) that enables its depth-GT-free training. GIFT-I contains 353 training groups with 3,268 RGB images and a disjoint validation set of 41 groups. Within each group, the camera pose and target geometry remain fixed while the reflected or transmitted content is varied, and target masks identify the regions in which geometric invariance should hold.

GIFT combines two complementary objectives during fine-tuning. A geometry-invariance loss constrains target-

*Corresponding author.

region depth variation across all valid image pairs within an intervention group. Because consistency alone admits group-shared degenerate solutions, such as predicting a constant depth map, a frozen-model self-distillation loss further constrains predictions outside the target and preserves the pretrained depth prior. For efficient adaptation, we employ LoRA (Hu et al. 2022) together with prediction reuse and group-wise micro-batching, enabling rapid fine-tuning on a single consumer-grade GPU. We further introduce geometry-consistency error (GCE), a depth-GT-free metric that directly measures the invariance of predicted geometry under controlled appearance changes and partially reflects the model's non-Lambertian depth estimation performance.

We apply GIFT to three representative depth foundation models: Depth Anything V2 (Yang et al. 2024b), VGGT-1B (Wang et al. 2025a), and Metric3D V2 (Hu et al. 2024). Experiments on the GIFT-I validation set and the Booster benchmark demonstrate clear improvements in non-Lambertian depth estimation.

Our contributions are:

- We construct the grouped GIFT-I dataset, enabling the training and validation of non-Lambertian depth estimation without measured depth labels.
- We propose GIFT, which combines complete-group geometry-invariance supervision with frozen-model self-distillation to improve non-Lambertian depth estimation while preventing prediction collapse and preserving the pretrained depth prior.
- We design an efficient LoRA-based group-wise optimization strategy with prediction reuse and micro-batching, enabling rapid adaptation on a single consumer-grade GPU.
- We validate GIFT on Depth Anything V2, VGGT-1B, and Metric3D V2, demonstrating improvements on GIFT-I and Booster.

Related Work

Monocular depth foundation models. MiDaS and DPT established robust cross-dataset relative depth (Ranftl et al. 2022; Ranftl, Bochkovskiy, and Koltun 2021). Depth Anything and its second version scale this direction with large unlabeled corpora (Yang et al. 2024a,b), while Metric3D, UniDepth, and Depth Pro target metric geometry (Yin et al. 2023; Piccinelli et al. 2024; Bochkovskiy et al. 2025). Diffusion-based estimators provide another strong geometric prior (Ke et al. 2024; Fu et al. 2024). VGGT jointly predicts cameras, depth, point maps, and tracks with a feed-forward transformer (Wang et al. 2025a). GIFT is not a new backbone, but a post-training procedure designed to adapt depth foundation models with different geometric priors.

Non-Lambertian depth completion and estimation. Depth completion methods typically recover transparent-object geometry from RGB-D observations with missing or corrupted sensor depth. ClearGrasp predicts transparent masks, surface normals, and occlusion boundaries to refine the raw depth, while A4T combines hierarchical affordance detection with progressive depth reconstruction for robotic manipulation (Sajjan et al. 2020; Jiang et al. 2022). TDCNet employs parallel CNN and Transformer branches to extract complementary features from partial depth and RGB-D inputs (Fan et al. 2025b). Fan et al. further introduce a self-supervised strategy that simulates transparent-like depth defects on non-transparent regions and uses the original valid depth as supervision (Fan et al. 2025a). ReMake combines instance masks with monocular depth priors to guide RGB-D depth completion and improve generalization to real-world grasping scenarios (Cheng et al. 2026).

Beyond depth completion, Mirror3D refines mirror regions using semantic and planar cues (Tan et al. 2021). Booster provides stereo ground truth for specular and transparent surfaces (Zama Ramirez et al. 2024), while other methods exploit contextual cues, layered representations, diffusion priors, or additional modalities (Liang et al. 2023; Shi et al. 2023; Tosi, Zama Ramirez, and Poggi 2024; Wen et al. 2025). Depth4ToM constructs virtual depth supervision by filling masked non-Lambertian regions with random colors (Costanzino et al. 2023). In contrast, GIFT preserves the real RGB observations and adapts pretrained monocular depth models by exploiting the geometry invariance shared within physical intervention groups, without measured training depth.

Method

Figure 1 summarizes the motivation and overall pipeline of GIFT. A pretrained depth foundation model may incorrectly recover reflected or transmitted content instead of the physical non-Lambertian surface. To address this problem, GIFT first collects intervention groups in which the camera pose and target geometry remain fixed while the target appearance varies with the reflected or transmitted environment. It then adapts the pretrained model using complete-group geometry-invariance supervision, together with frozen-model self-distillation that preserves the pretrained depth prior outside the target region. The entire training process requires no measured depth labels.

GIFT-Intervention Dataset

To provide supervision without measured depth labels, we construct the **GIFT-Intervention Dataset** (**GIFT-I**), a real-world grouped RGB dataset for adapting monocular depth models to non-Lambertian surfaces. As illustrated in Fig. 2, each image group is captured using a rigidly mounted camera while the camera pose and target geometry remain fixed. For transparent objects, we move or replace the content behind the target; for mirrors, we change the content reflected by the surface. These interventions produce substantial variations in target appearance while preserving its physical depth and pixel correspondence. A target-region mask is manually annotated for each image.

To maintain the fixed-geometry assumption, we apply lightweight image registration and quality filtering to remove captures with noticeable target misalignment. Detailed capture settings, intervention procedures for transparent and mirror targets, registration, filtering, mask annotation, and proxy-reference construction are provided in the supplemen-

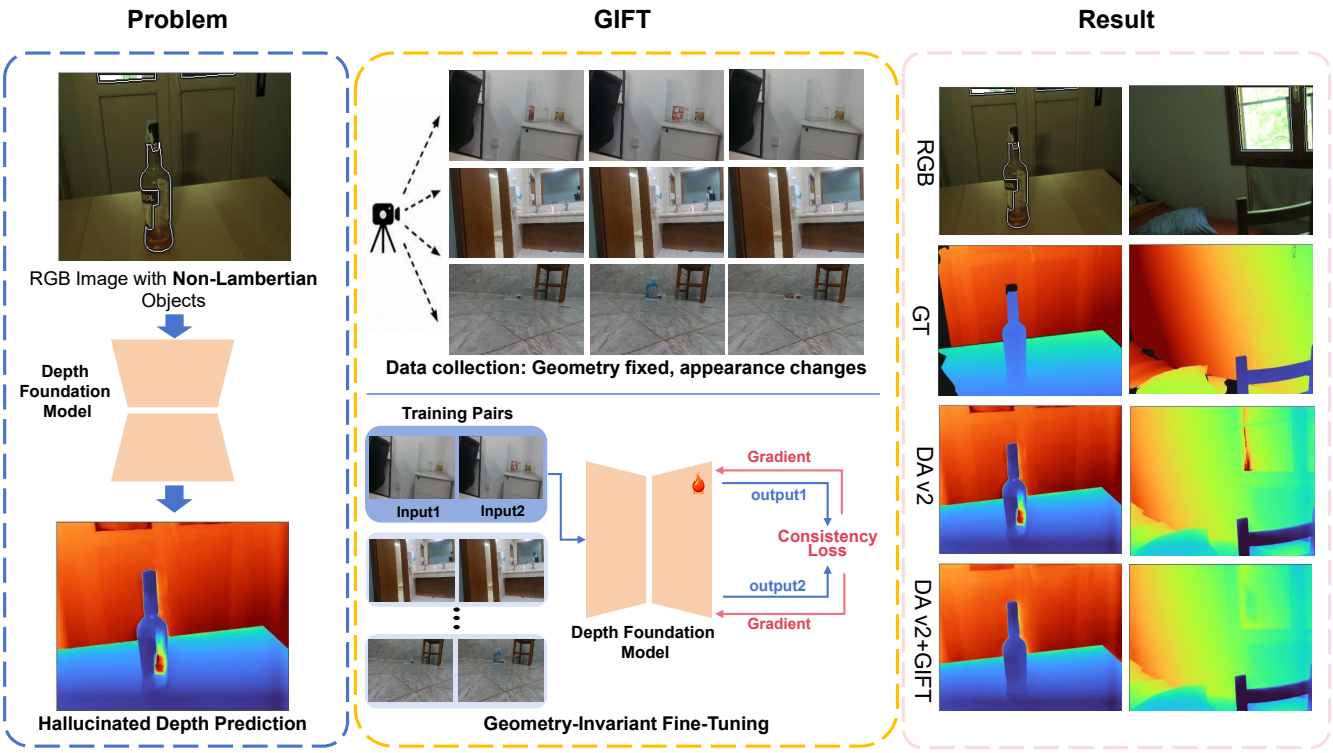

Figure 1: Overview of GIFT. A pretrained depth foundation model may interpret reflected or transmitted content as physical geometry, producing hallucinated depth on non-Lambertian surfaces. GIFT collects grouped observations under fixed camera pose and target geometry while varying the reflected or transmitted content, and adapts the model by minimizing target-region depth disagreement across observations. The adapted model suppresses appearance-dependent depth hallucinations and recovers more stable surface geometry.

tary material, together with additional GIFT-I groups illustrating the diversity of objects, materials, and scenes.

The training split contains 353 groups and 3,268 RGB images, with group sizes ranging from 2 to 55. A disjoint validation split contains 41 groups with 350 non-Lambertian input images. Each validation group additionally includes one surface-treated reference RGB image, resulting in 41 reference images in total.

The reference images are excluded from training and are not used as inputs to the evaluated model. Instead, predictions produced by the frozen depth foundation model on these reference images are used only to construct proxy depth references for assessing target-region recovery. Thus, GIFT-I requires no measured depth labels for either model adaptation or validation.

Geometry-Invariant Fine-Tuning

Geometry-invariance principle. Observations from the same intervention group share the camera viewpoint and physical target geometry but differ in reflected or transmitted appearance. Their target-region depth predictions should therefore remain consistent. As illustrated in Fig. 3, given two observations x_i and x_j, let $d_i = f_\theta(x_i)$ and $d_j = f_\theta(x_j)$ denote their predictions. Using the shared target mask M in this simplified pairwise illustration, we define the geometry-consistency discrepancy as

$$\ell_M(d_i, d_j) = \frac{\|M \odot (d_i - d_j)\|_1}{\|M\|_1}, \tag{1}$$

where $\odot$ denotes element-wise multiplication. A lower value indicates that the predicted target geometry is less sensitive to changes in non-Lambertian appearance. This shared-mask formulation presents the core idea; the complete-group objective below uses per-image masks and their pairwise intersections.

Complete-group geometric consistency. GIFT extends the pairwise principle to all valid unordered pairs in an intervention group. Given $G = \{(x_i, m_i)\}_{i=1}^N$, let $d_i = f_\theta(x_i)$ and $R_{ij}^T = m_i \cap m_j$ denote the prediction of observation i and the common target region of pair (i, j), respectively. The mask intersection reduces the influence of small registration and annotation-boundary errors. We define the valid pair set as

$$\mathcal{U}_T = \left\{(i,j) \,\middle|\, 1 \le i < j \le N,\ |R_{ij}^T| > 0\right\}, \tag{2}$$

and the complete-group consistency loss as

$$\mathcal{L}_T = \frac{1}{|\mathcal{U}_T|} \sum_{(i,j)\in\mathcal{U}_T} \ell_{R_{ij}^T}(d_i, d_j), \tag{3}$$

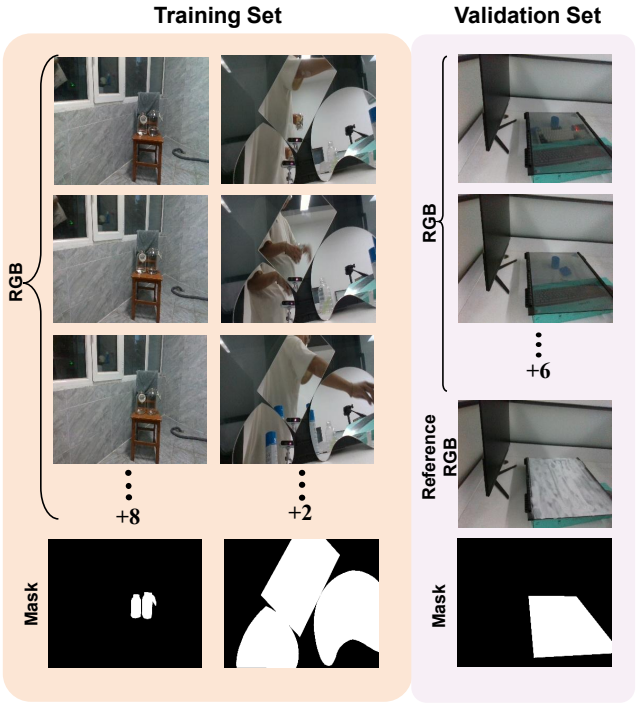

Figure 2: Overview of the GIFT-Intervention Dataset. Each training group contains RGB images captured under fixed camera pose and target geometry but varying non-Lambertian appearance, together with the corresponding target-region masks. Each validation group additionally contains a surface-treated reference RGB image, obtained using matte spray paint or opaque matte stickers, for constructing a proxy depth reference. The "+N" notation indicates that N additional RGB images from the same intervention group are not shown.

where ℓ is the normalized masked L_1 discrepancy defined in Eq. 1. Each prediction is computed once and reused across all pairs involving that image.

Preventing collapse with frozen-model self-distillation. Consistency alone does not uniquely determine the correct physical depth: any group-shared prediction, including a constant depth map, can minimize $\mathcal{L}_T$ and cause the predicted geometry to collapse. To prevent this degeneracy while preserving the pretrained depth prior, we constrain the adapted prediction outside the annotated target using the frozen base model. Let

$$q_i = \mathrm{sg}\left[f_{\theta_0}(x_i)\right] \tag{4}$$

denote the fixed base-model prediction, where $\mathrm{sg}[\cdot]$ stops gradients. These targets can be precomputed before fine-tuning.

The self-distillation loss is

$$\mathcal{L}_{\mathrm{SD}} = \frac{2}{|\mathcal{I}_{\mathrm{SD}}|} \sum_{i \in \mathcal{I}_{\mathrm{SD}}} \ell_{\neg m_i}(d_i, q_i), \tag{5}$$

where $\mathcal{I}_{\mathrm{SD}}$ contains images with valid non-target pixels. The discrepancy follows Eq. 1 with the region replaced by $\neg m_i$.

Algorithm 1 Efficient GIFT training via group-wise optimization

Require: Dataset $\mathcal{D}$, adapted model f_θ, micro-batch size K

for each intervention group (X, M) with frozen targets Q **do**
  $D \leftarrow \varnothing$
  for each micro-batch X_S of size at most K in X **do**
    $D_S \leftarrow f_\theta(X_S)$
    $D \leftarrow \mathrm{Concat}(D, D_S)$
  end for
  Compute $\mathcal{L}_T$ from D and M ▷ Eq. 3
  Compute $\mathcal{L}_{\mathrm{SD}}$ from D, Q, and M ▷ Eq. 5
  $\mathcal{L}_{\mathrm{GIFT}} \leftarrow \lambda_c \mathcal{L}_T + \lambda_d \mathcal{L}_{\mathrm{SD}}$ ▷ Eq. 6
  Update θ using $\nabla_\theta \mathcal{L}_{\mathrm{GIFT}}$
end for

Because it compares adapted and frozen predictions for the same input, this term does not assume that non-target regions are identical across different observations. It limits global prediction drift and preserves general depth estimation outside the target. The factor of two retains the relative contribution of the two endpoints in the original pair-based formulation.

The complete GIFT objective is

$$\mathcal{L}_{\mathrm{GIFT}} = \lambda_c \mathcal{L}_T + \lambda_d \mathcal{L}_{\mathrm{SD}}, \tag{6}$$

where λ_c controls target-region adaptation and λ_d controls retention of the pretrained depth prior.

Efficient group-wise training. Algorithm 1 summarizes the optimization procedure. For each intervention group, GIFT forwards all observations once using micro-batches and reuses the collected predictions to compute the complete-group consistency loss and frozen-model self-distillation loss. This avoids repeated forward passes for different image pairs and reduces the model-forward cost from $N(N-1)/2$ pair evaluations to N image evaluations per group.

Experiments

Experimental Setup

Datasets. We evaluate GIFT on the validation split of GIFT-I and on the independent Booster benchmark. The GIFT-I validation split contains 41 disjoint intervention groups with 350 non-Lambertian input images. Each group additionally provides one excluded surface-treated reference RGB image, which is used only to construct a proxy depth reference and is never used as input to the evaluated model.

We additionally evaluate on the balanced subset of the official Booster training split (Zama Ramirez et al. 2024). This subset contains 38 scenes and 228 left-camera images. Among them, 188 images from 32 scenes contain valid transparent-or-mirror (ToM) pixels. Booster provides stereo depth annotations and is not used to train GIFT. We use *ToM* to denote the annotated transparent and mirror regions and *All* to denote all valid pixels in an image.We use *Other* to denote valid pixels outside the annotated ToM region.

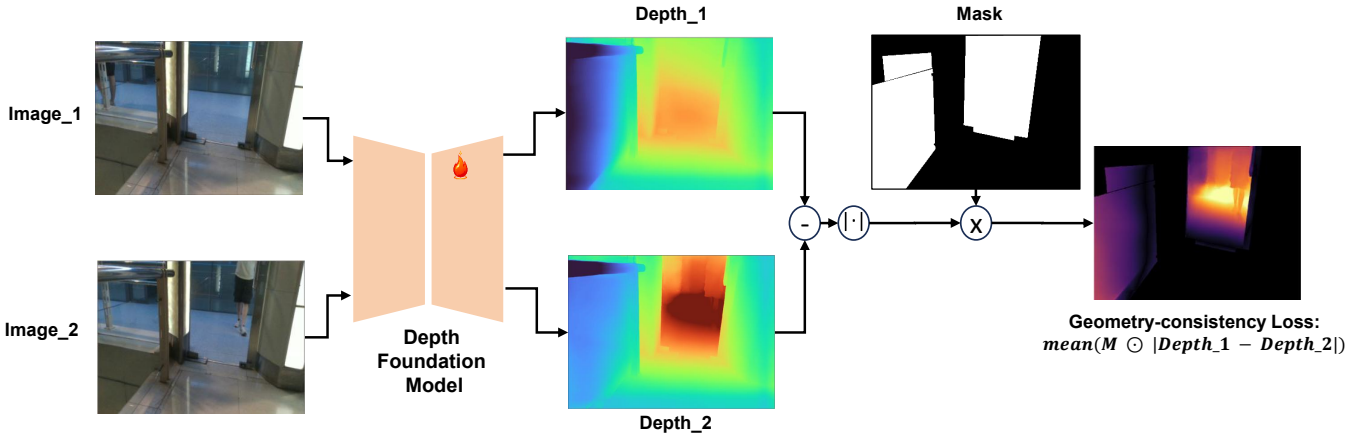

Figure 3: Pairwise illustration of geometry invariance. Observations with different non-Lambertian appearances should produce consistent target-region depth, measured by a normalized masked L_1 discrepancy.

Metrics and affine-aligned evaluation. We evaluate geometry invariance using geometry-consistency error (GCE), which measures the variation of predicted target geometry under controlled appearance changes. For a validation set $\mathcal{V}$ of intervention groups, we define

$$\mathrm{GCE} = \frac{1}{|\mathcal{V}|} \sum_{g \in \mathcal{V}} \frac{1}{|\mathcal{U}_T^{(g)}|} \sum_{(i,j) \in \mathcal{U}_T^{(g)}} \ell_{R_{ij}^{T,(g)}} \left(d_i^{(g)}, d_j^{(g)} \right), \quad (7)$$

where $\mathcal{U}_T^{(g)}$ is the valid unordered pair set of group g, and $R_{ij}^{T,(g)}$ is the common target region of observations i and j. Lower GCE indicates stronger invariance to reflected or transmitted appearance changes.

Because relative-depth predictions may differ in global scale and shift, we affine-align each prediction to its reference using all valid pixels:

$$(s^\star, t^\star) = \arg\min_{s,t} \sum_{p \in \Omega_{\text{valid}}} (s\hat{y}_p + t - y_p)^2, \qquad \tilde{y}_p = s^\star \hat{y}_p + t^\star. \quad (8)$$

The same transformation is applied to both the ToM and All regions. We then report MAE, AbsRel, RMSE, and δ_1, which evaluate aligned relative geometry rather than direct metric-scale prediction.

For GIFT-I, y is a proxy relative-depth label produced by the corresponding frozen teacher model from the excluded surface-treated reference image. For Booster, y is the stereo-derived ground-truth depth used to evaluate the monocular predictions. GIFT-I errors are reported in proxy units, whereas Booster MAE and RMSE are reported in millimeters.

Implementation details. We adapt DAV2, VGGT, and Metric3D V2 using rank-8 LoRA while keeping the majority of their pretrained parameters frozen. The corresponding depth prediction head of each backbone is jointly optimized during fine-tuning. All three backbones use the same input resolution and follow the group-wise optimization procedure described in Algorithm 1.

All models are trained for 50 epochs using Adam with BF16 precision, a learning rate of 2×10^{-5}, and no image augmentation. Each intervention group produces one optimizer update, and validation is performed after every epoch. The frozen version of each backbone is used to precompute the self-distillation targets outside the annotated non-Lambertian regions.

Quantitative Results

Results on GIFT-I. Table 1 shows that GIFT improves non-Lambertian depth recovery for all three adapted backbones. ToM RMSE is reduced by 67.2% for VGGT-1B, 22.9% for DA V2-Large, and 9.6% for Metric3D V2-Giant. VGGT-1B and Metric3D V2-Giant also improve across all All-region metrics, while DA V2-Large exhibits mixed All-region errors, indicating a stronger trade-off between target adaptation and retention. Because the proxy references are backbone-specific, comparisons on GIFT-I are made only between each backbone and its GIFT-adapted counterpart.

Results on Booster. Table 2 reports results on the Booster benchmark. GIFT reduces ToM RMSE by 18.2% for VGGT-1B and 37.7% for DA V2-Large, while Metric3D V2-Giant shows mixed ToM results. All-region RMSE decreases by 12.2%, 7.1%, and 11.8% for VGGT-1B, DA V2-Large, and Metric3D V2-Giant, respectively. Although MoGe-2-Large achieves the strongest absolute results, the paired comparisons show model-dependent gains rather than uniform improvement across all metrics.

Qualitative comparison. Figure 4 presents selected examples from GIFT-I and Booster. Relative to their corresponding base predictions, the GIFT-adapted models recover target-region shapes and depth transitions that more closely resemble the references. Additional zero-shot qualitative comparisons and analysis are provided in the supplementary appendix.

Method	ToM				All			
	MAE ↓	AbsRel ↓	RMSE ↓	δ_1 ↑	MAE ↓	AbsRel ↓	RMSE ↓	δ_1 ↑
VGGT-1B (Wang et al. 2025a)	184.667	0.1476	227.309	84.43	141.443	0.0779	230.101	93.67
DA V2-Large (Yang et al. 2024b)	66.728	0.0737	81.810	93.21	35.786	0.0307	58.634	98.25
Metric3D V2-Giant (Hu et al. 2024)	234.962	0.1958	301.349	73.68	167.029	0.1092	254.395	88.20
DA V3-Giant (Lin et al. 2026)	79.416	0.0717	99.924	94.77	85.197	0.0523	140.604	97.12
MoGe-2-Large (Wang et al. 2025b)	106.236	0.0946	121.600	93.39	80.492	0.0515	133.929	97.44
VGGT-1B + GIFT	62.465 **(-66.2%)**	0.0900 **(-39.1%)**	74.546 **(-67.2%)**	90.76 **(+7.5%)**	30.803 **(-78.2%)**	0.0354 **(-54.6%)**	49.782 **(-78.4%)**	97.56 **(+4.2%)**
DA V2-Large + GIFT	50.542 **(-24.3%)**	0.0481 **(-34.7%)**	63.086 **(-22.9%)**	98.83 **(+6.0%)**	38.897 (+8.7%)	0.0253 **(-17.6%)**	69.767 (+19.0%)	99.58 **(+1.4%)**
Metric3D V2-Giant + GIFT	217.195 **(-7.6%)**	0.1752 **(-10.5%)**	272.389 **(-9.6%)**	80.21 **(+8.9%)**	141.615 **(-15.2%)**	0.0871 **(-20.3%)**	236.046 **(-7.2%)**	92.77 **(+5.2%)**

Table 1: Affine-aligned proxy-depth results on GIFT-I (backbone-specific proxy units). Bold method names denote GIFT variants, bold percentages mark improvements.

Method	ToM				All			
	MAE ↓ (mm)	AbsRel ↓	RMSE ↓ (mm)	δ_1 ↑ (%)	MAE ↓ (mm)	AbsRel ↓	RMSE ↓ (mm)	δ_1 ↑ (%)
VGGT-1B (Wang et al. 2025a)	74.391	0.0740	94.121	94.55	33.198	0.0287	52.115	98.78
DA V2-Large (Yang et al. 2024b)	76.755	0.0798	100.060	95.29	41.852	0.0373	58.771	99.37
Metric3D V2-Giant (Hu et al. 2024)	70.922	0.0641	94.553	96.14	43.241	0.0344	62.634	98.41
DA V3-Giant (Lin et al. 2026)	61.048	0.0527	76.492	96.62	28.423	0.0234	43.906	99.40
MoGe-2-Large (Wang et al. 2025b)	34.120	0.0387	44.410	99.36	24.155	0.0211	35.383	99.71
VGGT-1B + GIFT	61.585 **(-17.2%)**	0.0625 **(-15.6%)**	77.022 **(-18.2%)**	96.78 **(+2.4%)**	29.057 **(-12.5%)**	0.0257 **(-10.7%)**	45.764 **(-12.2%)**	99.39 **(+0.6%)**
DA V2-Large + GIFT	51.450 **(-33.0%)**	0.0527 **(-33.9%)**	62.376 **(-37.7%)**	99.31 **(+4.2%)**	41.591 **(-0.6%)**	0.0367 **(-1.5%)**	54.581 **(-7.1%)**	99.59 **(+0.2%)**
Metric3D V2-Giant + GIFT	70.376 **(-0.8%)**	0.0636 **(-0.7%)**	99.109 (+4.8%)	96.15 **(+0.01%)**	36.194 **(-16.3%)**	0.0300 **(-12.7%)**	55.262 **(-11.8%)**	99.10 **(+0.7%)**

Table 2: Affine-aligned relative-depth results on Booster. MAE and RMSE are in mm and δ_1 is in percent. Bold method names denote GIFT variants, and bold percentages mark improvements.

Effect of Geometry-Consistency Training

Figure 5 compares smoothed trajectories after normalizing each metric by its initial value. Training GCE, GIFT-I validation GCE, and Booster ToM RMSE all decrease overall. The aligned trends indicate that the learned invariance generalizes to held-out groups and is accompanied by improved target-region accuracy on an independent dataset.

Ablation of the Combined Loss Weights

Figure 6 shows that $\lambda_d = 0$ causes Other-region RMSE to diverge because consistency alone admits degenerate group-shared predictions. Among non-collapsed runs, weaker self-distillation also produces larger fluctuations. Thus, self-distillation is necessary both to prevent collapse and to preserve ordinary-region predictions.

Table 3 reports the minimum-GCE checkpoint from each trajectory. The 2:1 ratio gives the lowest GCE, while 1:1.5 gives the best Booster values; no setting dominates all criteria. We use 1:1 because it achieves the lowest GIFT-I proxy ToM RMSE with competitive GCE, without using Booster for selection. The collapsed $\lambda_d = 0$ run is shown only in Figure 6.

	GIFT-I validation		Booster	
$\lambda_c : \lambda_d$	GCE ↓	ToM RMSE ↓ ($\times 10^3$)	ToM RMSE ↓ (mm)	Other RMSE ↓ (mm)
1:1	0.0450	**113.2**	246.4	288.7
1:1.5	0.0476	154.9	**236.1**	**284.7**
1:2	0.0509	123.9	257.1	304.4
2:1	**0.0381**	126.5	249.6	298.9

Table 3: DAV2 loss-weight ablation at the minimum-GCE checkpoint. Booster is evaluation-only; bold marks the best value.

Limitations

GIFT has two main limitations. First, although collecting grouped RGB images is easier than acquiring reliable non-Lambertian depth, controlled capture, registration, quality

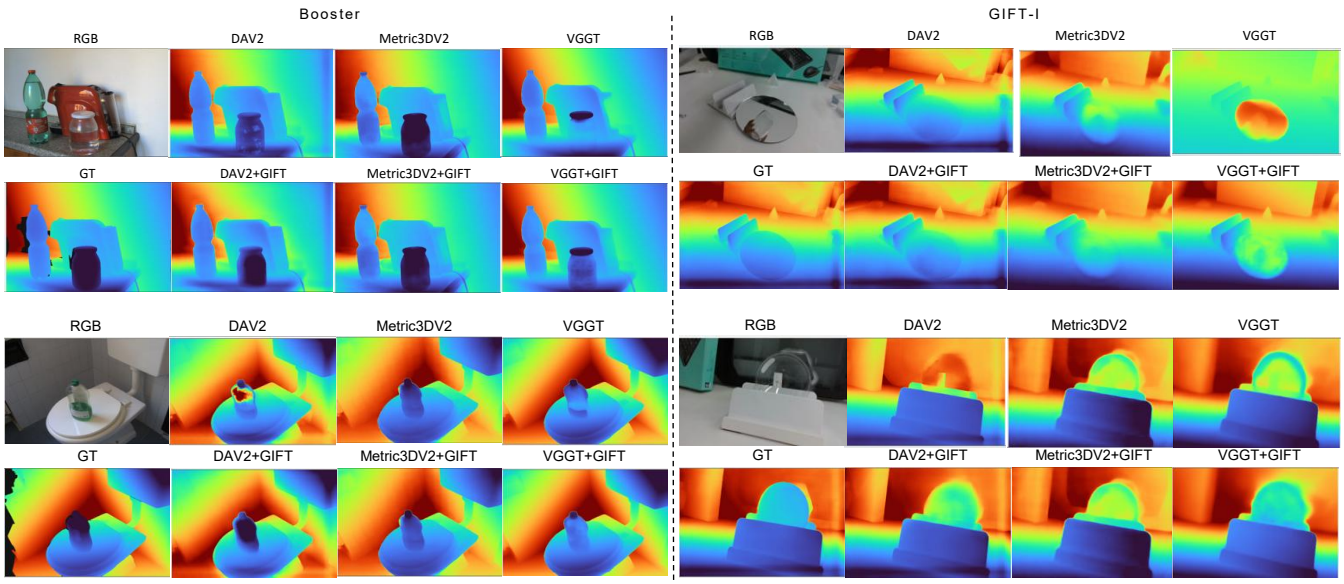

Figure 4: Qualitative comparison on Booster (left) and GIFT-I (right). For each example, the upper row shows RGB and base-model predictions, while the lower row shows the reference and corresponding GIFT-adapted predictions. GT denotes stereo ground truth on Booster and the surface-treated proxy reference on GIFT-I.

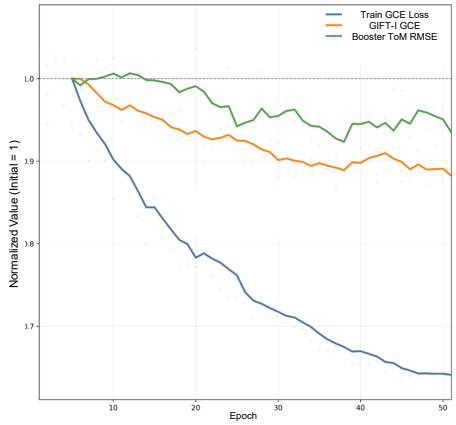

Figure 5: Smoothed, initial-normalized training GCE, GIFT-I GCE, and Booster ToM RMSE. Booster is evaluation-only.

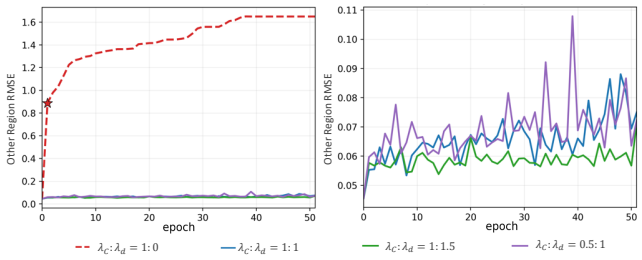

Figure 6: Other-region retention: removing self-distillation causes collapse (left), while weaker retention increases fluctuations (right).

filtering, and mask annotation still require manual effort, limiting the current scale and diversity of GIFT-I. Second, GIFT suppresses appearance-dependent depth hallucinations without measured depth supervision. Its accuracy therefore remains dependent on the pretrained backbone, as reflected by the smaller gains obtained with Metric3D V2. Future work will focus on automated data construction and combining geometry invariance with additional depth supervision.

Conclusion

We presented GIFT, a geometry-invariant fine-tuning framework for non-Lambertian monocular depth estimation. Using grouped observations from GIFT-I, GIFT combines complete-group geometry-invariance supervision with frozen-model self-distillation to reduce appearance-dependent depth hallucinations while preserving the pretrained depth prior. We also introduced GCE for depth-GT-free invariance evaluation. Experiments on GIFT-I and Booster demonstrate improvements across Depth Anything V2, VGGT-1B, and Metric3D V2, showing that controlled appearance intervention provides a practical self-supervised signal for adapting depth foundation models.

References

Bochkovskiy, A.; Delaunoy, A.; Germain, H.; Santos, M.; Zhou, Y.; Richter, S. R.; and Koltun, V. 2025. Depth Pro: Sharp Monocular Metric Depth in Less Than a Second. In *International Conference on Learning Representations (ICLR)*.

Cheng, Y.; Gao, X.; Zhang, S.; Zeng, C.; Zha, F.; Sun, L.; and Yang, C. 2026. Rethinking Transparent Object Grasping: Depth Completion With Monocular Depth Estimation and Instance Mask. *IEEE Robotics and Automation Letters*, 11(5): 5510–5517.

Costanzino, A.; Zama Ramirez, P.; Poggi, M.; Tosi, F.; Mattoccia, S.; and Di Stefano, L. 2023. Learning Depth Estimation for Transparent and Mirror Surfaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9244–9255.

Fan, X.; Chen, Z.; Pan, M.; Deng, A.; and Yang, H. 2025a. Self-Supervised Learning for Transparent Object Depth Completion Using Depth from Non-Transparent Objects. In *2025 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Fan, X.; Ye, C.; Deng, A.; Wu, X.; Pan, M.; Luo, S.; and Yang, H. 2025b. TDCNet: Transparent Objects Depth Completion With CNN-Transformer Dual-Branch Parallel Network. *IEEE Sensors Journal*, 25(19): 36629–36641.

Fu, X.; Yin, W.; Hu, M.; Wang, K.; Ma, Y.; Tan, P.; Shen, S.; Lin, D.; and Long, X. 2024. GeoWizard: Unleashing the Diffusion Priors for 3D Geometry Estimation from a Single Image. In *Computer Vision – ECCV 2024*, 241–258.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations (ICLR)*.

Hu, M.; Yin, W.; Zhang, C.; Cai, Z.; Long, X.; Chen, H.; Wang, K.; Yu, G.; Shen, C.; and Shen, S. 2024. 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.

Jiang, J.; Cao, G.; Deng, J.; Do, T.-T.; and Luo, S. 2024. Robotic Perception of Transparent Objects: A Review. *IEEE Transactions on Artificial Intelligence*, 5(6): 2547–2567.

Jiang, J.; Cao, G.; Do, T.-T.; and Luo, S. 2022. A4T: Hierarchical Affordance Detection for Transparent Objects Depth Reconstruction and Manipulation. *IEEE Robotics and Automation Letters*, 7(4): 9826–9833.

Ke, B.; Obukhov, A.; Huang, S.; Metzger, N.; Daudt, R. C.; and Schindler, K. 2024. Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9492–9502.

Liang, Y.; Deng, B.; Liu, W.; Qin, J.; and He, S. 2023. Monocular Depth Estimation for Glass Walls with Context: A New Dataset and Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12): 15081–15097.

Lin, H.; Chen, S.; Liew, J. H.; Chen, D. Y.; Li, Z.; Shi, G.; Feng, J.; and Kang, B. 2026. Depth Anything 3: Recovering the Visual Space from Any Views. In *International Conference on Learning Representations (ICLR)*.

Piccinelli, L.; Yang, Y.-H.; Sakaridis, C.; Segu, M.; Li, S.; Van Gool, L.; and Yu, F. 2024. UniDepth: Universal Monocular Metric Depth Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10106–10116.

Ranftl, R.; Bochkovskiy, A.; and Koltun, V. 2021. Vision Transformers for Dense Prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 12179–12188.

Ranftl, R.; Lasinger, K.; Hafner, D.; Schindler, K.; and Koltun, V. 2022. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3): 1623–1637.

Sajjan, S. S.; Moore, M.; Pan, M.; Nagaraja, G.; Lee, J.; Zeng, A.; and Song, S. 2020. ClearGrasp: 3D Shape Estimation of Transparent Objects for Manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 3634–3642.

Shi, X.; Dikov, G.; Reitmayr, G.; Kim, T.-K.; and Ghafoorian, M. 2023. 3D Distillation: Improving Self-Supervised Monocular Depth Estimation on Reflective Surfaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9133–9143.

Tan, J.; Lin, W.; Chang, A. X.; and Savva, M. 2021. Mirror3D: Depth Refinement for Mirror Surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15990–15999.

Tosi, F.; Zama Ramirez, P.; and Poggi, M. 2024. Diffusion Models for Monocular Depth Estimation: Overcoming Challenging Conditions. In *Computer Vision – ECCV 2024*, 236–257.

Wang, J.; Chen, M.; Karaev, N.; Vedaldi, A.; Rupprecht, C.; and Novotny, D. 2025a. VGGT: Visual Geometry Grounded Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5294–5306.

Wang, R.; Xu, S.; Dong, Y.; Deng, Y.; Xiang, J.; Lv, Z.; Sun, G.; Tong, X.; and Yang, J. 2025b. MoGe-2: Accurate Monocular Geometry with Metric Scale and Sharp Details. In *Advances in Neural Information Processing Systems*.

Wang, X.; He, Y.; Shi, J.; Lu, J.; Yang, Y.; Jiang, Y.; and Jiang, C. 2026. SeeClear: Reliable Transparent Object Depth Estimation via Generative Opacification. *arXiv preprint arXiv:2603.19547*.

Wen, H.; Zuo, Y.; Subramanian, V.; Chen, P.; and Deng, J. 2025. Seeing and Seeing Through the Glass: Real and Synthetic Data for Multi-Layer Depth Estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6715–6725.

Xu, S.; Wei, S.; Wei, Q.; Geng, Z.; Li, H.; Shen, L.; Sun, Q.; Han, S.; Ma, B.; Li, B.; Ye, C.; Zheng, Y.; Wang, N.; Zhang, S.; and Zhao, H. 2025. Diffusion Knows Transparency: Repurposing Video Diffusion for Transparent Object Depth and Normal Estimation. arXiv:2512.23705.

Yang, L.; Kang, B.; Huang, Z.; Xu, X.; Feng, J.; and Zhao, H. 2024a. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10371–10381.

Yang, L.; Kang, B.; Huang, Z.; Zhao, Z.; Xu, X.; Feng, J.; and Zhao, H. 2024b. Depth Anything V2. In *Advances in Neural Information Processing Systems*, volume 37, 21875–21911.

Yin, W.; Zhang, C.; Chen, H.; Cai, Z.; Yu, G.; Wang, K.; Chen, X.; and Shen, C. 2023. Metric3D: Towards Zero-Shot

Metric 3D Prediction from a Single Image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9043–9053.

Zama Ramirez, P.; Costanzino, A.; Tosi, F.; Poggi, M.; Salti, S.; Mattoccia, S.; and Di Stefano, L. 2024. Booster: A Benchmark for Depth from Images of Specular and Transparent Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1): 85–102.

Zhang, J.; Li, J.; Huang, Y.; Wang, Y.; Zheng, J.; Shen, L.; and Cao, Z. 2025. Towards Robust Monocular Depth Estimation in Non-Lambertian Surfaces. In *Computer Vision – ECCV 2024 Workshops*, volume 15623 of *Lecture Notes in Computer Science*, 175–189. Springer.

GIFT: Geometry-Invariant Fine-Tuning for Non-Lambertian Monocular Depth Estimation

— Supplementary Material —

This supplementary material complements the main paper with further details on the construction and preprocessing of GIFT-I, the implementation of group-wise GIFT adaptation, and additional qualitative comparisons. Section A describes the capture equipment, appearance interventions, manual mask annotation, image registration, and dataset organization. Section B reports implementation and optimization details omitted from the main paper. Section C presents additional qualitative results on GIFT-I and Booster.

Contents

A GIFT-I Dataset

A.1 Data Collection and Annotation

Capture setup. Each GIFT-I group is captured using a fixed camera–target configuration. An Intel RealSense D435i camera is rigidly mounted on a tripod, and the non-Lambertian target remains stationary throughout the capture process. Only the RGB stream of the D435i is used.

The D435i depth stream is deliberately excluded. Active depth measurements are often unreliable on reflective and transparent surfaces. More importantly, GIFT is designed to suppress appearance-dependent depth hallucinations under fixed physical geometry while preserving the geometric prior of a pretrained monocular model. It is not designed to eliminate the general model-to-real-world domain gap or to calibrate a monocular model to the metric output, noise characteristics, and biases of a particular RGB-D sensor. Using the D435i depth stream for training or evaluation would therefore conflate appearance-invariance adaptation with a separate sensor-domain adaptation problem.

Figure S1 shows the equipment used for RGB collection and surface-treated reference acquisition, including the D435i camera, a tripod, AESUB scanning spray, and frosted glass film.

Figure S1: Equipment used for GIFT-I data collection and reference acquisition: an Intel RealSense D435i camera, a tripod, AESUB scanning spray, and frosted glass film.

Appearance interventions. For transparent objects, the transmitted content behind the target is moved or replaced. For mirrors, the reflected scene content is changed. The camera pose and target geometry remain fixed, while the non-Lambertian appearance of the target changes with its reflected or transmitted environment. Other parts of the scene are kept approximately unchanged within the same group.

These controlled interventions provide observations of the same physical surface geometry under different appearances, forming the geometry-invariance signal used by GIFT.

Surface-treated reference RGB. Each validation group additionally contains one surface-treated reference RGB image captured under the same camera pose and target geometry. The treatment directly changes the surface material and its image formation, replacing the original reflection- or transmission-dominated appearance with a more diffuse surface appearance.

AESUB scanning spray and frosted glass film provide two practical treatment options. The choice primarily depends on the area and shape of the non-Lambertian surface. Scanning spray is suitable for relatively small or irregularly shaped targets. For large, approximately planar transparent surfaces, frosted glass film is preferred because it is more economical and material-efficient than coating the entire surface with scanning spray.

The surface-treated reference RGB image is not included among the evaluated non-Lambertian inputs. It is used to construct the proxy reference for the corresponding validation group.

Manual target-mask annotation. The target region in every non-Lambertian RGB image is manually annotated with a binary mask. Masks are annotated independently for individual images rather than copied across a group, because small residual displacements may remain despite the fixed capture setup. Independent annotation also allows pairwise mask

Set	Groups	Inputs	References	Group size
Training	353	3,268	0	2–55
Validation	41	350	41	3–23†

Table S1: Composition of GIFT-I. †The validation group size includes one surface-treated reference RGB image.

intersections to conservatively exclude uncertain boundary pixels after registration.

A.2 Preprocessing and Dataset Overview

Image registration. All non-reference observations within each capture group are registered to the last naturally ordered non-reference RGB image. SIFT features are extracted using a contrast threshold of 0.08 and an edge threshold of 10. Candidate correspondences are first filtered using Lowe's ratio test with a threshold of 0.70.

A displacement-consistency filter then removes matches whose displacement differs from the median displacement by at least 20 pixels. A homography is estimated using USAC-MAGSAC, with RANSAC used as a fallback. The reprojection threshold is 2 pixels, the confidence is 0.9995, and the maximum number of iterations is $5{,}000$.

A registered frame is retained only when at least 90% of the displacement-filtered correspondences are geometric inliers. The same homography is applied to the RGB image, target mask, and stored proxy depth. RGB images use linear interpolation, whereas masks and depth maps use nearest-neighbor interpolation.

Mask processing and quality control. All masks are binarized again after resizing and geometric transformation. For an image pair (i, j), the common target region is defined as $m_i \cap m_j$. This intersection conservatively excludes pixels affected by residual registration errors or uncertain annotation boundaries.

Dataset paths are stored relative to a configurable dataset root. During loading, images are grouped according to their capture directory, duplicate entries are removed, and group members are sorted using natural ordering. Every training RGB image must have a corresponding manually annotated mask and frozen proxy depth. Missing required files cause an explicit error rather than silent sample replacement. Groups containing fewer than two valid observations are rejected because they cannot provide an inter-image consistency target.

Dataset composition. GIFT-I contains 353 training groups with 3,268 RGB images. Group sizes range from 2 to 55, with a mean of 9.26 images. The validation set contains 41 groups with 350 non-Lambertian RGB inputs. Each validation group additionally contains one surface-treated reference RGB image, resulting in 41 reference images.

Table S1 summarizes the composition of GIFT-I.

Frozen proxy generation. For every training RGB image, the frozen version of the corresponding backbone predicts a proxy depth q_i. These proxies are generated once before adaptation and supervise only pixels outside the manually annotated target mask. No proxy depth is applied inside the non-Lambertian target region.

For validation, the same frozen backbone predicts the depth of the surface-treated reference RGB image. The resulting prediction is used as the proxy reference for the non-Lambertian observations in the corresponding group. The GIFT-I reference is therefore model-derived rather than measured ground truth. Consequently, GIFT-I errors are used for before–after comparisons within the same backbone rather than for directly ranking different backbone families.

All RGB images are resized to 476×644 using bicubic interpolation. Proxy inference is performed independently for each backbone in FP32 with a batch size of two. The resulting proxy predictions are retained in the native output scale of each backbone and are not interpreted as metric depth. GIFT-I evaluation follows the affine-alignment protocol defined in the main paper; the proxy values are therefore treated as backbone-specific relative depth.

Figure S2 presents representative complete groups from GIFT-I. For every displayed group, all captured RGB observations are shown. The camera pose and target geometry remain fixed within each group, while the reflected or transmitted content is deliberately changed. Red boxes identify the regions affected by the appearance intervention.

B Training and Implementation Details

B.1 Group-Wise Implementation

The pseudocode in the main paper presents the group-wise objective at a conceptual level. The implementation below realizes the same objective under bounded GPU memory using one snapshot forward and one replay forward per image.

We use the complete-group geometry-consistency and frozen-model self-distillation objective defined in the main paper, with $\lambda_c : \lambda_d = 1 : 1$. No additional loss is introduced.

Each optimizer update consumes one complete intervention group. To limit memory usage, we implement the group-wise objective using snapshot-and-replay. Group predictions are first cached without gradient tracking in micro-batches of two.

The images are then replayed with gradients enabled. For each micro-batch, its complete-group consistency contribution is computed against the cached group predictions, while self-distillation is evaluated against the precomputed frozen-model targets. Loss gradients are first computed with respect to the current depth predictions and then propagated through the current model computation graph.

After each micro-batch backward pass, its computation graph is released, while parameter gradients are accumulated over the complete group. One optimizer update is applied after all group members have been processed. Thus, only detached group-level depth snapshots and accumulated parameter gradients are retained across micro-batches. If a group contains an odd number of images, the final single-image micro-batch is processed normally.

Each epoch consists of one shuffled pass over all 353 training groups, resulting in 353 Adam updates. Frozen self-distillation targets are precomputed and loaded from disk during adaptation.

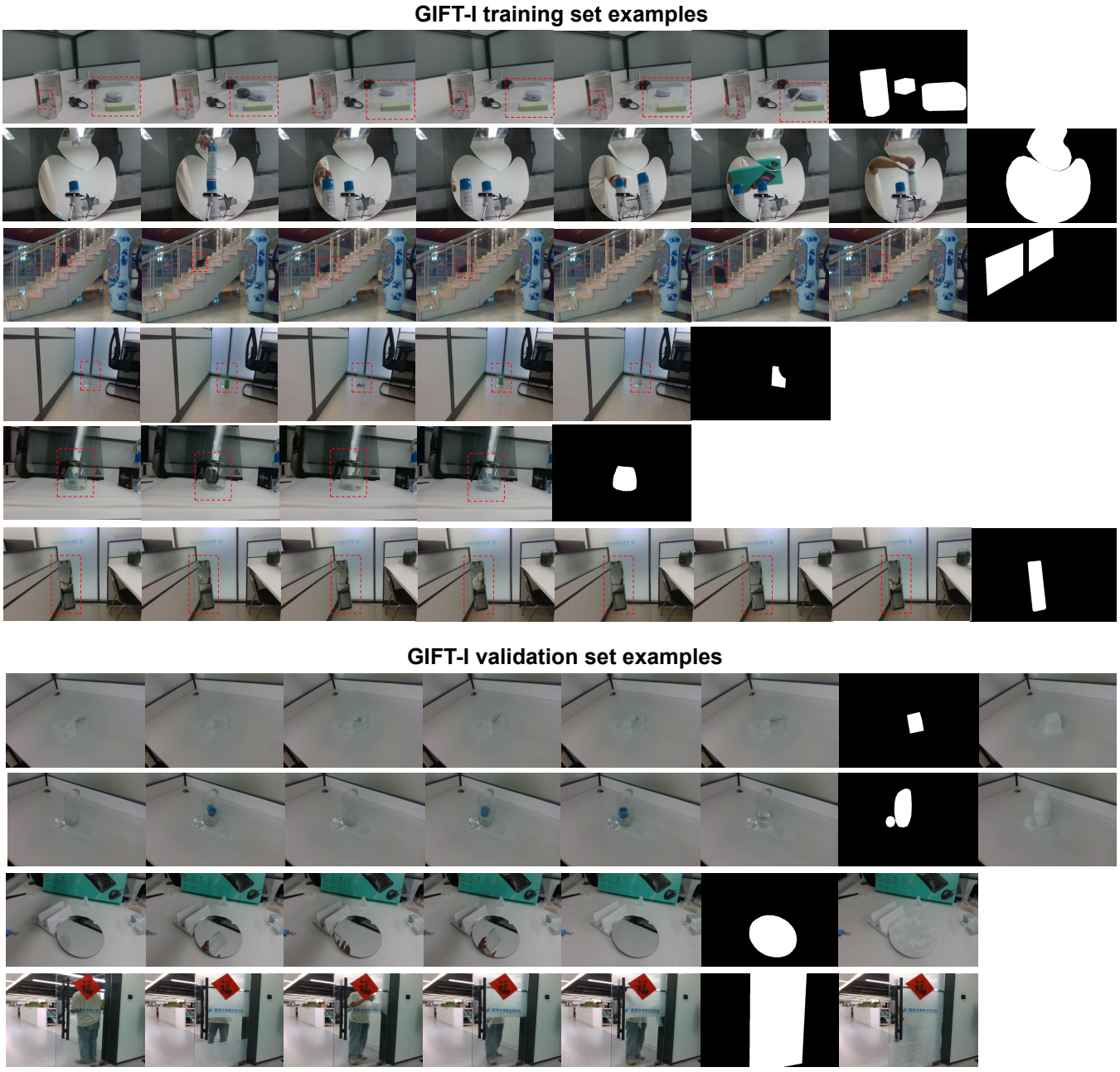

Figure S2: Representative complete groups from GIFT-I. Within each group, the camera pose and target geometry remain fixed while the reflected or transmitted content changes. Red boxes indicate the regions affected by the appearance intervention. All RGB observations belonging to each displayed group are shown. The surface-treated reference RGB image is used to construct the corresponding proxy reference.

B.2 Model and Optimization Settings

We adapt DA V2-Large, VGGT-1B, and Metric3D V2-Giant using rank-8 LoRA modules. For each backbone, the corresponding depth prediction head is jointly optimized with the LoRA parameters, while the remaining pretrained parameters are frozen.

All three models use the same optimization configuration. Input images are resized to 476×644, and each model is trained for 50 epochs using Adam with BF16 precision. The learning rate is fixed at 2×10^{-5}, with zero weight decay and no image augmentation. The geometry-consistency and self-distillation terms are equally weighted, with $\lambda_c : \lambda_d = 1 : 1$. Each intervention group produces one optimizer update.

Implementation	Optimizer updates / epoch	Image forwards / epoch	Time / epoch	Relative speed
Naive pair traversal	20,339	40,678	122.4 min†	$1.0\times$
Snapshot-and-replay group-wise	353	6,536	4.53 min	$27.0\times$

Table S2: Efficiency comparison between naive pair traversal and the snapshot-and-replay group-wise implementation for DA V2-Large on the 353-group GIFT-I training set. The group-wise implementation uses one detached snapshot forward and one gradient-enabled replay forward per image, while reusing each snapshot across all associated pairwise comparisons. †The pair-traversal time is estimated from a measured average of 0.3612 seconds per pair update.

B.3 Runtime and Efficiency

Runtime environment. Training time is measured on one NVIDIA GeForce RTX 4090 with 24 GB of GPU memory. The software environment uses Python 3.10.6, PyTorch 2.5.1 with CUDA 12.1, cuDNN 9.1, and xFormers 0.0.28.post3. The reported runtime covers only the 353 complete-group training updates in one epoch; evaluation and model I/O are excluded.

Under the snapshot-and-replay group-wise implementation, one epoch takes 4.53 minutes for DA V2-Large, 5.90 minutes for VGGT-1B, and 6.94 minutes for Metric3D V2-Giant.

Pair traversal versus group-wise optimization. The 353 training groups contain 20,339 unordered image pairs. A naive pair-traversal implementation processes every pair independently, requiring 20,339 optimizer updates and 40,678 image forward passes per epoch.

The snapshot-and-replay group-wise implementation performs one no-gradient snapshot forward and one gradient-enabled replay for each of the 3,268 training images. It therefore requires 6,536 image forward passes and 353 optimizer updates per epoch. The detached snapshots are reused across all pairwise comparisons involving the corresponding image.

Table S2 summarizes the comparison. Relative to naive pair traversal, the group-wise implementation reduces the number of optimizer updates by $57.6\times$ and the number of image forward passes by $6.22\times$. The reported $27.0\times$ wall-clock speedup is an engineering estimate because the pair-traversal runtime is extrapolated from its measured per-update throughput. The comparison includes both prediction reuse and the reduced number of optimizer updates.

C Additional Qualitative Comparisons

Figure S3 presents additional qualitative comparisons on GIFT-I and Booster. The visualization follows the affine alignment protocol defined in the main paper. For each sample, the base-model and GIFT-adapted predictions are aligned to the corresponding reference and visualized using the same depth range.

For compact presentation, the reference-depth row is labeled "GT" for both datasets in the figure. Its meaning is dataset-dependent. On Booster, GT denotes stereo-derived ground-truth depth. On GIFT-I, GT denotes the corresponding backbone-specific proxy reference generated from the surface-treated reference RGB image; it is not measured ground truth.

Invalid stereo pixels on Booster are displayed only in the GT image and are excluded from evaluation. They are not copied into, removed from, or overlaid on the dense monocular predictions.

C.1 Zero-Shot Comparison

Figure S4 compares the frozen DA V2-Large model with models adapted for 10,000 and 20,000 optimization steps on a set of zero-shot images that are not included in GIFT-I or Booster. These images are used only for qualitative visualization and do not participate in training or quantitative evaluation. The 20,000-step checkpoint is obtained from an extended diagnostic run beyond the 50-epoch schedule used for the quantitative experiments. Neither diagnostic checkpoint is used in the reported quantitative results.

As the adaptation duration increases, the predicted geometry becomes progressively more conservative. Compared with the frozen model, the 10,000-step model reduces appearance-dependent depth structures in the non-Lambertian regions, while the 20,000-step model further suppresses such variations and produces smoother, less appearance-sensitive surface predictions. This trend suggests that longer geometry-invariance adaptation increases the model's preference for stable surface geometry over depth details inferred from reflected or transmitted content.

However, a more conservative prediction is not necessarily uniformly better. Excessive adaptation may also suppress plausible geometric details. Therefore, the comparison illustrates the effect of adaptation duration rather than establishing that more optimization steps always improve depth accuracy.

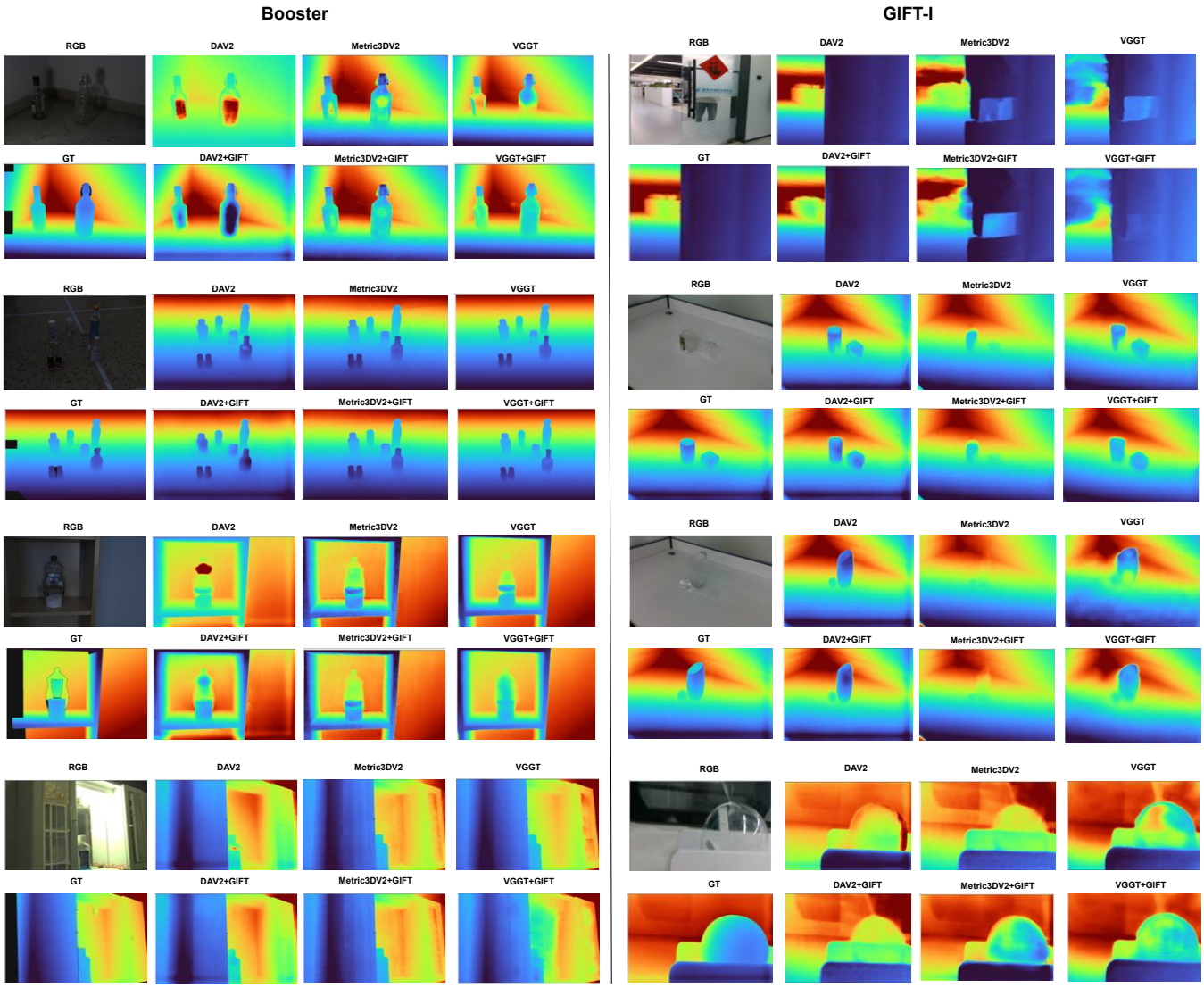

Figure S3: Additional qualitative comparisons on GIFT-I and Booster. For compact presentation, the reference depth is labeled "GT" in both datasets. On Booster, GT denotes stereo-derived ground-truth depth. On GIFT-I, GT denotes the corresponding backbone-specific proxy reference generated from the surface-treated reference RGB image and is not measured ground truth. For each sample, the base-model and GIFT-adapted predictions are affine-aligned to the corresponding reference and visualized using the same depth range. Invalid stereo pixels are shown only in the Booster GT image and are not transferred to the monocular predictions.

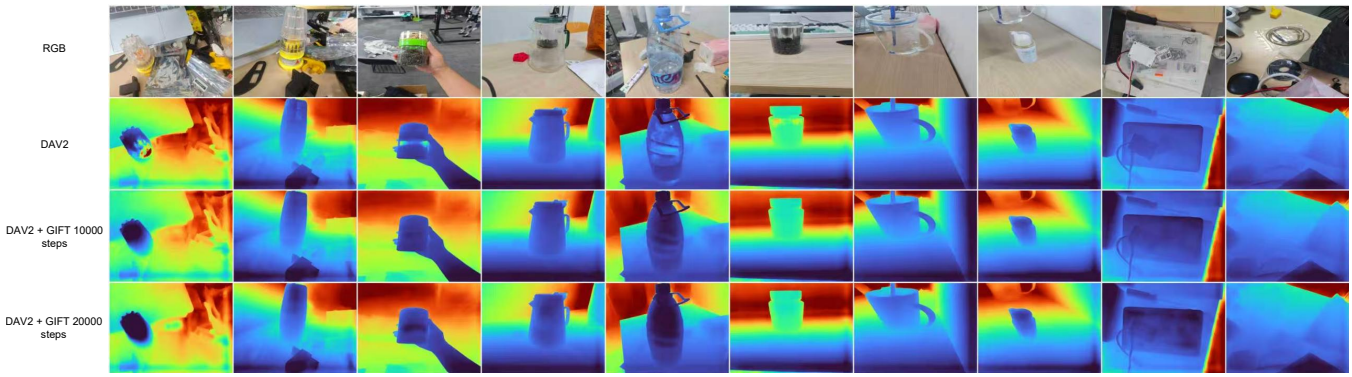

Figure S4: Zero-shot comparison of DA V2-Large before adaptation and after 10,000 and 20,000 GIFT optimization steps. With longer adaptation, the predictions become progressively more conservative: appearance-dependent depth structures are increasingly suppressed, resulting in smoother and less appearance-sensitive estimates of the non-Lambertian surfaces. These examples are used only for qualitative analysis and are not included in GIFT-I or Booster.