MirrorVerse: Pushing Diffusion Models to Realistically Reflect the World

Supplementary Material

Contents

A. Dataset Generation

Grounding Object. We describe the algorithm to ground an object in Algorithm 2. If the minimum value of the bounding box along the z-dimension is greater than the ground level, we adjust the object's position to rest it on the ground. This adjustment enhances the photorealism of the generated dataset, ensuring objects appear naturally grounded in their environment.

Sampling Region For random positions in front of the mirror, we first define a region in the x-y plane where the object and its reflection are visible in the camera view. To determine this region, we compute the intersection of the camera's viewing frustum on the x-y plane with the extent of the mirror. This process is repeated for all camera locations used in the dataset generation. The resulting sampling region, $\mathcal{S}$, ensures that any position within it allows the visibility of both the object and its reflection in the camera view.

Object Placement We provide the procedure to place an object in a scene in Algorithm 3. First, we scale an object to fit inside a unit cube, ensuring uniformity across objects. Next, we randomly sample a position in front of the mirror and update the position of the object accordingly. Finally, we apply a random rotation around the vertical axis. These steps guarantee that the object and its reflection remain visible in the camera's field of view. Further, these steps introduce greater diversity to the dataset.

Algorithm 2 Procedure to Ground an Object

Require: Input 3D model $\mathcal{M}$, Ground-Level $\underline{Z}$

Function GROUNDOBJECT($\mathcal{M}, \underline{Z}$)
  $bbox \leftarrow$ GETBOUNDINGBOX3D($\mathcal{M}$)
  if $bbox.z > \underline{Z}$ **then**
    $z \leftarrow bbox.z - \underline{Z}$
    $\Delta t \leftarrow [0, 0, -z]$
    $\mathcal{M}.position \leftarrow \mathcal{M}.position + \Delta t$
  end if

Algorithm 3 Procedure to Place an Object in a scene

Require: Input 3D model $\mathcal{M}$, Ground-Level $\underline{Z}$, Sampling region $\mathcal{S}$

Function NORMALIZEOBJECT($\mathcal{M}$)
  $bbox \leftarrow$ GETBOUNDINGBOX3D($\mathcal{M}$)
  $d_{max} \leftarrow$ GetMaxDimension($bbox$)
  $s \leftarrow \frac{1}{d_{max}}$
  $S_{matrix} \leftarrow$ GetScaleMatrix(s)
  $\mathcal{M}.scale \leftarrow S_{matrix}$
  $bbox \leftarrow$ GETBOUNDINGBOX3D($\mathcal{M}$)
  return $bbox$

Function SAMPLEPOSITION($\mathcal{M}, S$)
  $t \leftarrow$ SamplePosition(S)
  $\mathcal{M}.translation \leftarrow t$
  $bbox \leftarrow$ GETBOUNDINGBOX3D($\mathcal{M}$)
  return $bbox$

Function RANDOMROTATION($\mathcal{M}$)
  $\theta \leftarrow$ RandomAngle(S)
  $R_{matrix} \leftarrow$ GetRotationMatrix(θ)
  $\mathcal{M}.rotation \leftarrow R_{matrix}$
  $bbox \leftarrow$ GETBOUNDINGBOX3D($\mathcal{M}$)
  return $bbox$

Main Algorithm
  $bbox \leftarrow$ NORMALIZEOBJECT($\mathcal{M}$)
  $bbox \leftarrow$ SAMPLEPOSITION($\mathcal{M}, \mathcal{S}$)
  $bbox \leftarrow$ RANDOMROTATION($\mathcal{M}$)

Additional Samples. We provide additional samples for scenes with single object in Fig. 13 and multiple objects in

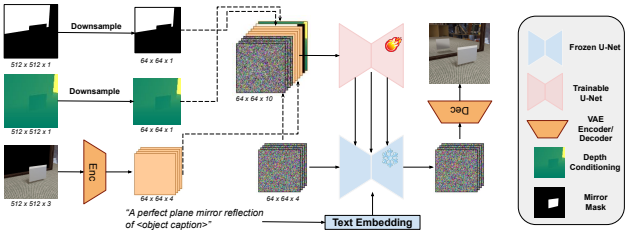

Figure 11. **Overview of architecture used for the experiments.**

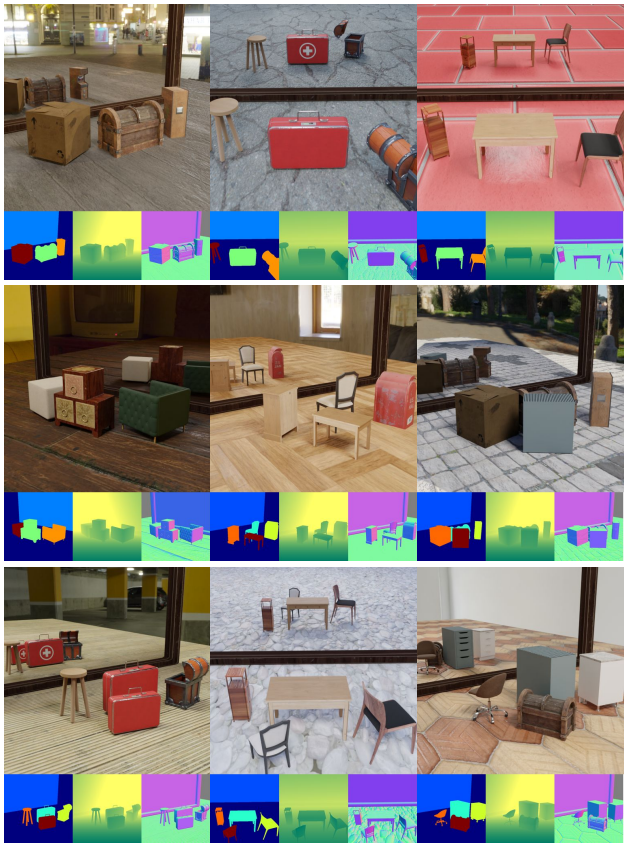

Figure 12. **Samples of scenes containing more than two objects.**

Figs. 12 and 14. Further, we also provide additional samples for visualization from SynMirrorV2 in folder *"dataset-samples"* in the attached supplementary material.

B. Architecture Details

Our work builds upon the MirrorFusion framework [12], which employs a conditioning network that leverages depth information to guide the generation process of a pre-trained generative network. This model is trained using a three-stage curriculum learning strategy, as detailed in Sec. 4.

Figure 13. Samples of scene containing single object from SynMirrorV2

Figure 14. Samples of scene containing multiple objects from SynMirrorV2

Table 4. Comparison of the proposed dataset, SynMirrorV2 , with SynMirror

Augmentations	Grounding	Random Rotations	Random Positions	Multiple Objects	Occlusion Scenarios
SynMirror	✗	✗	✗	✗	✗
SynMirrorV2	✓	✓	✓	✓	✓

C. Additional Details

C.1. Characteristic Comparison with SynMirror

We present a characteristic comparison of the proposed dataset with SynMirror in Tab. 4. This variety aids in the generalization of the MirrorFusion 2.0 to complex scenarios and real-world scenes.

C.2. Text prompts used in the experiments

We provide the text prompts used in the main paper for image generation.

Figure 1. Each row in this figure uses the same text prompt. Text prompts are as follows:

- **First row.** *"A perfect plane mirror reflection of (A cylindrical bottle with a spherical top and bottom, featuring a neck and spout.) and (A cylinder with a conical bottom and a spherical top.)"*
- **Second row.** *"A perfect plane mirror reflection of a yellow mug with a flower design is placed on a desk in front of a mirror. The reflection of the mug can be seen in the mirror, creating an interesting visual effect where the mug appears to be floating."*

Figure 4. Text prompts are as follows:

- **(a)** *"A perfect plane mirror reflection of a chair with a high, rounded back and blue upholstery."*
- **(b)** *"A perfect plane mirror reflection of (3D model of a Chesterfield sofa with cylindrical and spherical elements.) and (3D model of a two-seater sofa with backrest and armrests.)"*

Figure 5. Text prompts are as follows:

- **(a)** *"A perfect plane mirror reflection of a black color furniture in stair shape with multiple drawers."*
- **(b)** *"A perfect plane mirror reflection of a yellow and white mug on a grey surface."*

Figure 6. Text prompts for "Column-1" are as follows:

- **First row.** *"A perfect plane mirror reflection of a curved and slatted 3D chair with backrest, seat, armrests, and footrest."*
- **Second row.** *"A perfect plane mirror reflection of a croissant that is chocolate and covered in nuts on one side and plain on the other."*
- **Third row.** *"A perfect plane mirror reflection of (A wooden stool with a backrest and a seat.) and (3D model of a lamp with a cylindrical body, spherical base, conical bottom, spherical top, and a cylindrical shade with a spherical accent.)"*
- **Fourth row.** *"A perfect plane mirror reflection of (King size bed with a slatted base, tufted headboard and footboard, curved backrest and armrest, and slanted top and bottom edges.) and (A 3D object with a truncated octagonal base and a spherical top.)"*

Text prompts for "Column-2" are as follows:

- **First row.** *"A perfect plane mirror reflection of an orange vat sitting on a grey metal frame with a light gray colored control box attached."*
- **Second row.** *"A perfect plane mirror reflection of a dark blue baby buggy with no one in it."*
- **Third row.** *"A perfect plane mirror reflection of (A 3D model of a single-seater Chesterfield sofa with a tufted back, curved backrest, and slanted seat.) and (Rectangular cabinet with a slanted roof and base, featuring a door and a drawer, 3D modeled as a TV stand.)"*
- **Fourth row.** *"A perfect plane mirror reflection of (3D model of a kitchen cart with three shelves, two drawers on each side, and a two-tiered bunk bed with slatted bases.) and (Three-tiered wheeled cart with shelves and a handle.)"*

Figure 7. Text prompts for "Column-1" are as follows:

- **First row.** *"The dynamics of the mirror and its reflections involve the use of a cardboard box. The box is placed on top of a table, and the mirror is positioned in such a way that it reflects the surrounding environment."*
- **Second row.** *"The mirror's reflection in the cord creates an interesting visual effect, making it appear as if the cord is coming out of the mirror itself. This setup can be useful for individuals who need to charge their electronic devices while working."*
- **Third row.** *'A pink portable charger is placed on a wooden table next to a mirror. The mirror reflects the portable charger, creating an interesting dynamic between the object and its reflection in the mirror."*
- **Fourth row.** *"The mirror reflects a pile of dirty clothes or towels, which appear to be wrinkled and disheveled. This phenomenon is caused by the way light bounces off the surface of the mirror and interacts with the objects in front of it"*

Text prompts for "Column-2" are as follows:

- **First row.** *"Anamorphosis is a technique used to create distorted images that appear normal when viewed from a specific vantage point. In this case, the viewer needs to be positioned directly in front of the mirror to see the full effect of the toy"*
- **Second row.** *"A toy mouse-shaped mirror is placed on a*

Table 5. Ablation studies on mixed-training for multiple objects

Metrics	Reflection Generation Quality			Text Alignment
Models	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**	**CLIP Sim ↑**
Joint Training	17.41	0.615	0.153	**26.37**
Ours 50k	**18.00**	**0.744**	**0.119**	26.09

table in front of a yellow cylindrical object. The mirror reflects the environment around it, including the yellow cylinder and other objects in the room."

- **Third row.** *"The mirror is reflecting a small button with a cartoon character on it, which is placed on a white surface. The reflection of the button in the mirror creates an interesting visual effect, as the character appears to be floating or hovering"*
- **Fourth row.** *"The mirror in the image is reflecting two toy figures, a red one and a blue one, as they interact with each other. This dynamic creates a playful and imaginative scene, as the toys appear to be having a conversation."*

Figure 8. Text prompts are as follows:

- **(a)** *"A perfect plane mirror reflection of (A swivel chair with a mesh seat, backrest, and swivel base.) and (A wooden cuboid stool with a square base, slanted squarish seat, and slanted backrest.)"*
- **(b)** *"A perfect plane mirror reflection of (L-shaped sectional sofa with U-shaped backrest, 3D model.) and (A king size bed with a tufted headboard, footboard, slatted base, and a single seater sofa with a backrest and seat cushion.)"*

Figure 9. Text prompts are as follows:

- **(a)** *"A perfect plane mirror reflection of a red and white striped round life buoy surrounded in a cord."*
- **(b)** *"A perfect plane mirror reflection of a rug"*

D. Additional Results

D.1. Affect of Joint training with single and multiple objects

An ablation study in Tab. 5 reveals the significant impact of staged training on generalization. Training on single and multiple splits simultaneously yielded inferior results, highlighting the importance of our staged approach. Curriculum training further allows us to fine-tune on real-world data such as the MSD dataset, providing better results than direct single-stage training as shown in Appendix D.4 and Fig. 16.

D.2. Additional results from single object scenes from MirrorBenchV2

We present additional results for single objects in Fig. 15. Observe that the baseline method produces several inaccuracies: the object **(Column 1, Row 2)** appears to be floating in mid-air with incorrect orientation, and the bullets **(Column 2, Row 2)** are also misaligned. Additionally, the reflection of the wooden table **(Column 1, Row 5)** has distorted legs. In contrast, our results accurately capture the geometry and appearance of the object in its reflection.

Text prompts for **"Column-1"** in Fig. 15 are as follows:

- **First row.** *"A perfect plane mirror reflection of a shiny dark wooden pepper mill, stood upright, with a silver ornament on the top."*
- **Second row.** *"A perfect plane mirror reflection of a white gravy boat with designs of pink flowers on the side and front with green folliage."*
- **Third row.** *"A perfect plane mirror reflection of a orange and black two wheeled hoverboard."*
- **Fourth row.** *"A perfect plane mirror reflection of a 3D swivel chair model with a curved backrest, armrests, and a swivel base."*
- **Fifth row.** *"A perfect plane mirror reflection of a swivel chair with a slender, curved backrest and armrests, featuring a slanted seat."*

Text prompts for **"Column-2"** in Fig. 15 are as follows:

- **First row.** *"A perfect plane mirror reflection of a swivel chair with a slender, curved backrest and armrests, featuring a slanted seat."*
- **Second row.** *"A perfect plane mirror reflection of two metal bullets for either a gun or cannon."*
- **Third row.** *"A perfect plane mirror reflection of a two-seater sofa with curved backrest, slanted seat, and armrests."*
- **Fourth row.** *"A perfect plane mirror reflection of a 3D lamp with a cylindrical metal arm, spherical metal base, and spherical glass shade."*
- **Fifth row.** *"A perfect plane mirror reflection of a rectangular table with a slatted, slanted top, hairpin legs, and a metal frame."*

D.3. Additional results from multiple object scenes from MirrorBenchV2

We present additional results for multiple objects in Fig. 19. The baseline method struggles to generate accurate reflections in scenes with multiple objects compared to its performance in single-object scenes. Notably, reflections of two sofas **(Column 1, Row 4)** and a sofa-table **(Column 2, Row 4)** pair are incorrectly merged, and in some cases, only a single object is rendered in the reflection. This poor performance is primarily due to the limited diversity of the dataset used to train the baseline method. In contrast, our approach preserves the original geometry of the objects, accurately

captures their spatial relationships, and maintains their appearance, resulting in significantly more realistic and consistent reflections for scenes with multiple objects.

Text prompts for **"Column-1"** in Fig. 19 are as follows:

- **First row.** *"A perfect plane mirror reflection of a (rug) and (Rectangular cabinet with a slanted roof and base, featuring a door and a drawer, 3D modeled as a TV stand.)."*
- **Second row.** *"A perfect plane mirror reflection of a (3D model of a chair with a backrest, armrests, and seat.) and (Spherical table with a round top, square slanted base, and two slender legs.)."*
- **Third row.** *"A perfect plane mirror reflection of (A 3D model of a chair with a curved, tufted backrest, padded seat, armrests, and a squarish bowl with a matching base and lid.) and (rug)."*
- **Fourth row.** *"A perfect plane mirror reflection of (A 3D model of a two to three-seater sofa with a curved, tufted backrest, armrests, and a footrest.) and (3D model of a three-seater sofa with a curved backrest and armrests.)."*
- **Fifth row.** *"A perfect plane mirror reflection of (Cylindrical stool with a spherical top, square base, slanted seat, and backrest.) and (A cuboid with a base, spherical lid, and slanted top and bottom.)."*
- **Sixth row.** *"A perfect plane mirror reflection of (A cylinder with a spherical base and a spherical shade.) and (3D object: Slatted swivel chair with a curved, slanted X-shaped backrest and seat, featuring armrests and legs.)."*
- **Seventh row.** *"A perfect plane mirror reflection of (3D model of a chaise lounge featuring a curved backrest, cushioned seat, armrests, and footrest, made from a single piece of foam.) and (A 3D object resembling a book with a convex spine, slanted top and bottom edges, and stacked pages.)."*
- **Eigth row.** *"A perfect plane mirror reflection of (Two-seater couch with backrest and armrests, and a tetrahedral box with lid.) and (3D model of a rectangular coffee table with a slanted shelf on top, supported by a slanted frame.)."*

Text prompts for **"Column-2"** in Fig. 19 are as follows:

- **First row.** *"A perfect plane mirror reflection of (3D model of stacked cylindrical objects with spherical tops, resembling trash cans or water cisterns, featuring a tetrahedral cuboid and a truncated octahedral.) and (Three-drawer dresser with a slanted top, rectangular base, and rectilinear design.)."*
- **Second row.** *"A perfect plane mirror reflection of a (3D model of a chair with a backrest, armrests, and seat.) and (Spherical table with a round top, square slanted base, and two slender legs.)."*
- **Third row.** *"A perfect plane mirror reflection of (rug) and (3D model of a slanted rectangular coffee table in 3ds Max, available for download.)."*
- **Fourth row.** *"A perfect plane mirror reflection of (Three-seater grey sofa with a curved backrest and armrests in 3D.) and (A 3D tetrahedral desk with a slanted top, two drawers, a shelf, and a truncated triangular base.)."*
- **Fifth row.** *"A perfect plane mirror reflection of (A 3D model of a rectangular table with a pair of legs and a top.) and (Two-seater sofa with curved backrest and armrests, 3D model.)."*
- **Sixth row.** *"A perfect plane mirror reflection of (Swivel bar stool with a cylindrical seat, curved backrest, and swivel functionality.) and (Tall cabinet with a triangular base, slanted roof, flat top, and a retractable banner stand.)."*
- **Seventh row.** *"A perfect plane mirror reflection of (A wooden stool with a backrest and a seat.) and (3D model of a lamp with a cylindrical body, spherical base, conical bottom, spherical top, and a cylindrical shade with a spherical accent.)."*
- **Eigth row.** *"A perfect plane mirror reflection of (A king-size platform bed with a box-shaped, curved-top headboard, footboard, side rails, slatted base, and curved backrest.) and (U-shaped sectional sofa with multiple L-shaped sections, featuring backrests and armrests.)."*

D.4. Comparison by fine-tuning only on the MSD dataset

To highlight the impact of the proposed dataset SynMirrorV2 and stage-wise training more profoundly, we compare our model (which is trained in 3 stages with the first two involving SynMirrorV2) with the model only finetuned directly on the MSD dataset for $10k$ iterations (i.e only stage 3) from the Stable Diffusion v1.5 checkpoint. We call the fine-tuned model "MSD-10k-FT". We compare "MSD-10k-FT" with results from our method in Fig. 16. Note the orientation of the power bank is incorrect **(Second Row)**, a brown toy is not generated in the reflection **(Third Row)** and a pink cup is generated instead of the white teapot **(Fourth Row)** in the results from "MSD-10k-FT". This shows the importance of the proposed synthetic dataset for incorporating the priors of accurate mirror reflections and the importance of stage-wise finetuning to bridge the generalization gap.

Text prompts used in Fig. 16 are as follows:

- **First row.** *"A perfect plane mirror reflection of a cardboard box placed on top of a table"*
- **Second row.** *"A perfect plane mirror reflection of a pink portable charger is placed on a wooden table."*
- **Third row.** *"A perfect plane mirror reflection of a toy poop emoji figurine placed along with a blue cuboid and a green cylindrical object."*
- **Fourth row.** *"A perfect plane mirror reflection of a pink and white ceramic mug with a smiling face on it."*

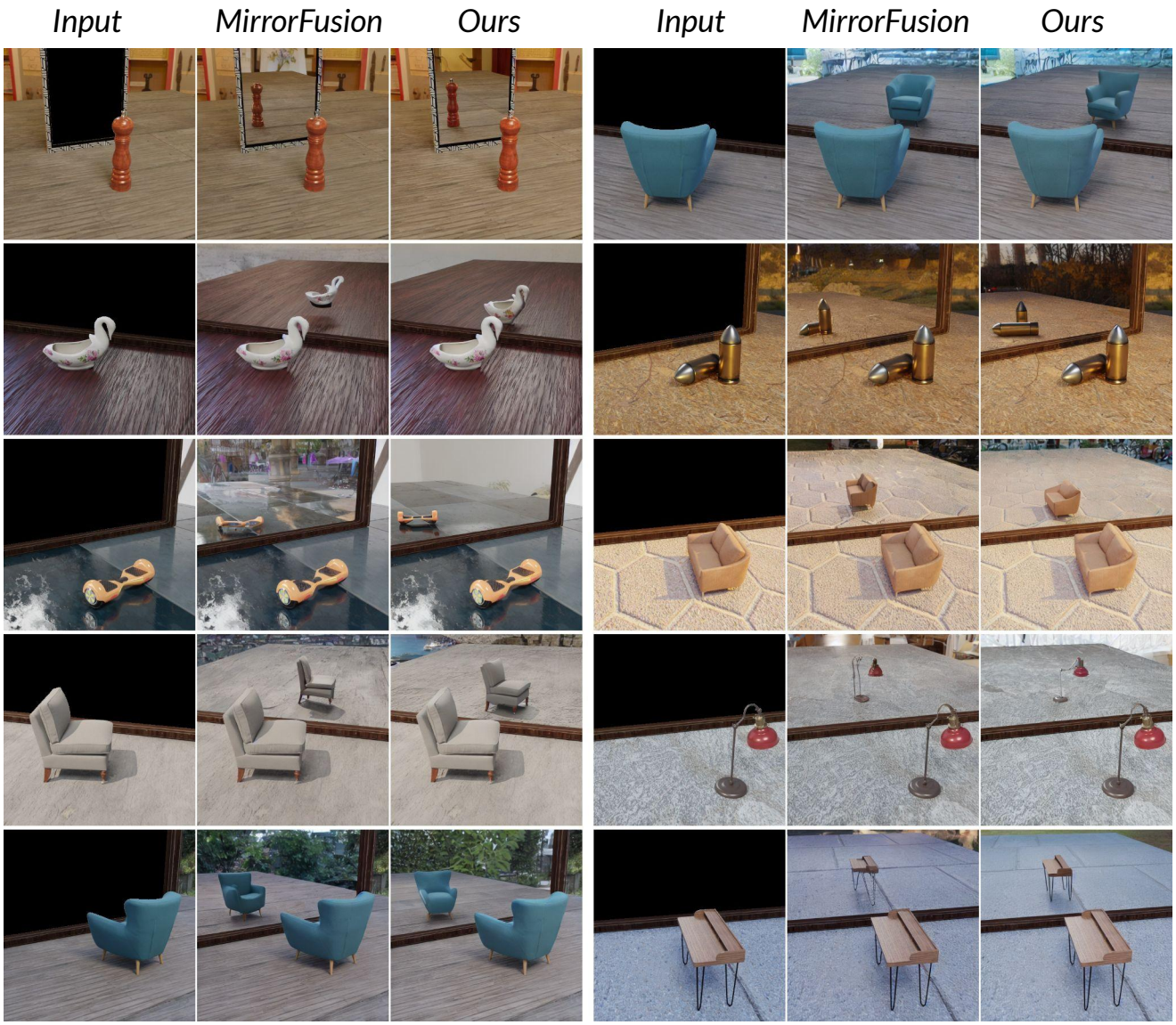

Figure 15. **Results on scenes with single objects.** More discussion in Appendix D.2.

D.5. User Study Details

We provide details of the user study described in Section 4 of the main paper. We selected 40 samples, including single-object and multi-object scenes, from MirrorBenchV2, GSO [13], and real-world scenes from MSD [52]. These samples were generated by the baseline method MirrorFusion and our method MirrorFusion 2.0.

We invited 29 participants (aged 18–50) to compare results based on the realism and plausibility of mirror reflections. Each task involved evaluating and selecting the best result among the outputs from both models with instructions to assess factors such as:

- Apparent distance and alignment of objects in the reflection.
- Geometry consistency and subtle details in reflections.
- Floor reflections and shadow orientations, if present.

Fig. 17 shows that our method was preferred in 84% of cases.

D.6. Comparison with Commercial Product

We compare our method with Adobe Firefly [1] in Fig. 18. Our method demonstrates superior performance compared to a widely used commercial product, Adobe Firefly, in generating accurate reflections on a mirror plane. Notably, the commercial product places reflections incorrectly for multiple objects in the scene (top row) and mispositions the reflection of the yellow mug (bottom row). Additionally, the

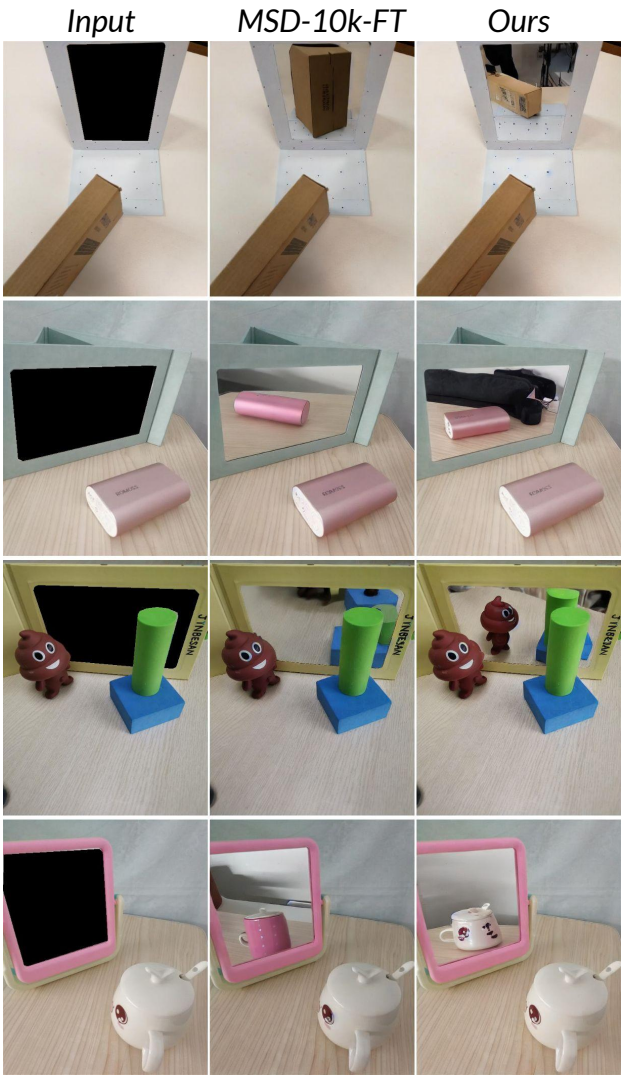

Figure 16. **Comparison with "MSD-10k-FT".** We finetune our model directly on the MSD dataset [52] for $10k$ iterations and compare the results with our full 3-stage finetuning approach. More discussion in Appendix D.4.

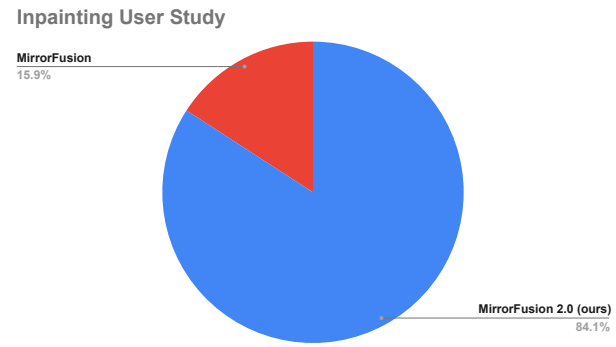

Figure 17. Visual comparison of outputs from our method and the baseline. We discuss in detail in Appendix D.5

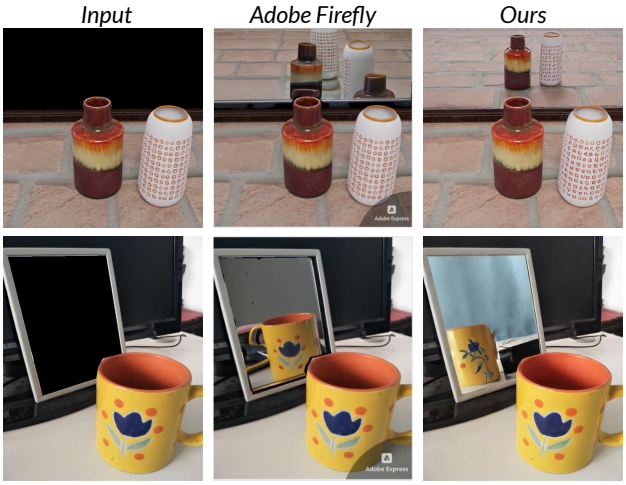

Figure 18. **Comparison with Adobe Firefly [1].** We discuss in detail in Appendix D.6

orientation of objects in the inpainted region is inaccurate. In contrast, our method consistently generates reflections in the correct positions while preserving their proper orientation and appearance, resulting in more realistic and visually accurate outcomes. Text prompts used in Fig. 18 are as follows:

- **First row.** *"A perfect plane mirror reflection of (A cylindrical bottle with a spherical top and bottom, featuring a neck and spout.) and (A cylinder with a conical bottom and a spherical top.)"*
- **Second row.** *"A perfect plane mirror reflection of a yellow mug with a flower design is placed on a desk in front of a mirror. The reflection of the mug can be seen in the mirror, creating an interesting visual effect where the mug appears to be floating."*

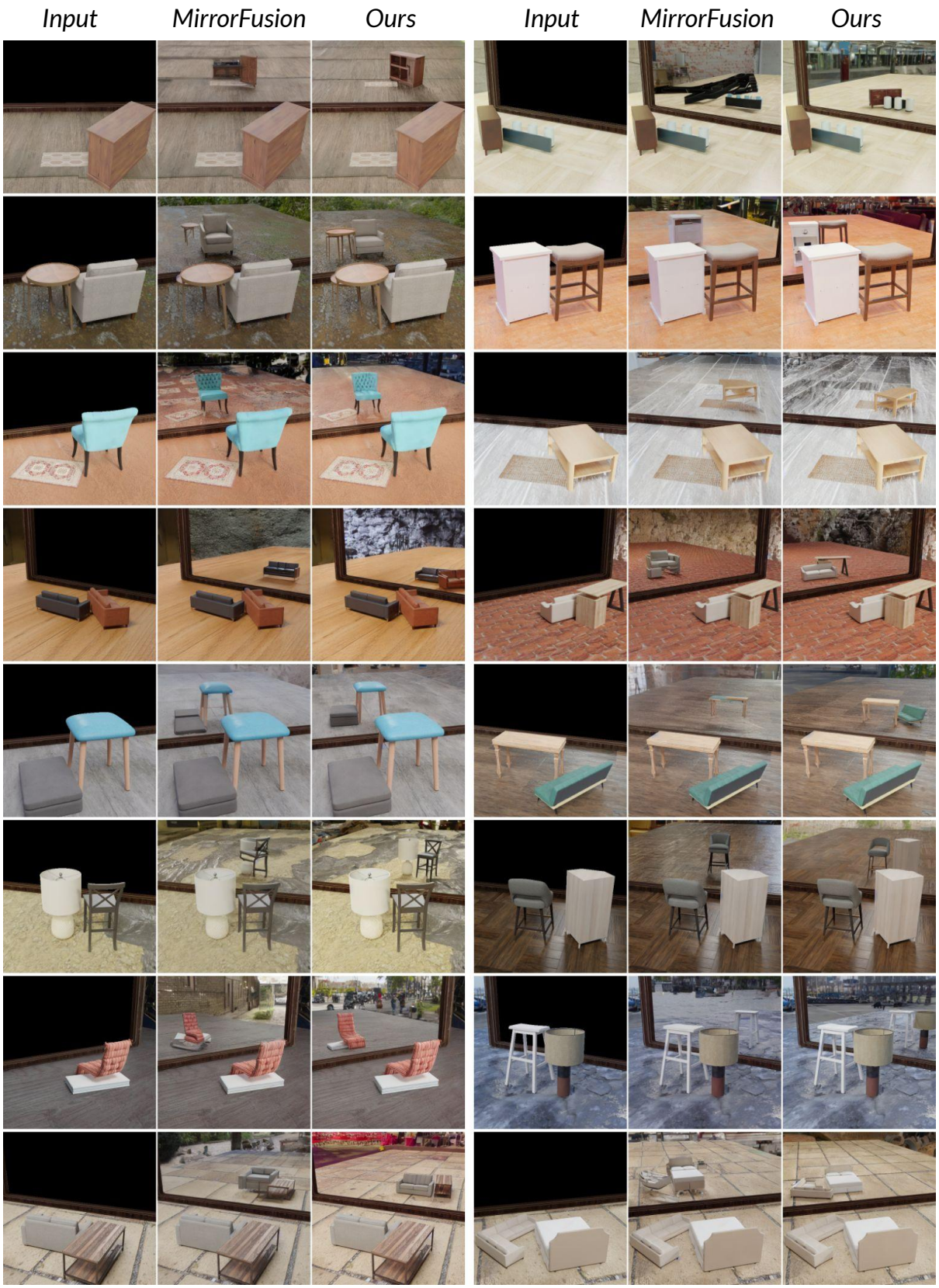

Figure 19. **Results on scenes with single objects.** More discussion in Appendix D.3.