VLM-DREAMER: VLM-IMAGINED BI-DIRECTIONAL INPAINTING FOR SINGLE-IMAGE 360 SCENE GENERATION (SUPPLEMENTARY MATERIALS)

Ting-Wei Huang[1] *Fu-En Yang*[2] *Min-Hung Chen*[2] *Yen-Yu Lin*[1] *Yu-Lun Liu*[1]

[1]National Yang Ming Chiao Tung University [2]NVIDIA

1. ALGORITHM DETAILS

In this section, we provide pseudocode for our framework. The following algorithms illustrate the key steps in each stage.

Algorithm 1 360-Aware Scene Description Generation

Input: Single image I_0, prompt p (text prompt fed to VLM for scene imagination)
Output: 360-aware scene descriptions $\{T_i\}$
Generate multi-view images $\{P_i\}$ using diffusion model
for each i **do**
    $T_i = \text{VLM}(P_{i-1}, p)$
end for
return $\{T_i\}$

Algorithm 2 Uni-Directional Coarse Scene Generation

Input: Input image I_0, 360-aware descriptions $\{T_i\}$
Output: Coarse 360 3DGS scene S
Initialize coarse 3DGS scene S from I_0
for each novel viewpoint i **do**
    Render novel view I_i from S
    Retrieve description T_i from $\{T_i\}$
    $I_i = \text{Inpaint}(I_i, T_i)$
    Update S with I_i
end for
return S

Algorithm 3 Bi-Directional 3DGS Refinement

Input: Coarse 360 3DGS scene S
Output: Refined 360 3DGS scene S
for each target view I_i **do**
    Render (I_{i-1}, I_i, I_{i+1}) from S
    Mask central region of I_i for refinement
    $I_i = \text{BiDirectionalInpainting}(I_{i-1}, I_i, I_{i+1})$
    Update S with refined I_i
end for
return S

2. QUALITATIVE RESULTS

We provide qualitative results of scene generation across different scenes in Figure 1, 2 and 3, comparing our method with several baselines including LucidDreamer [1], Invisible Stitch [2], DreamScene360 [3], Vistadream [4], and FlexWorld [5]. Existing approaches face various challenges. For example, VistaDream often has inconsistent brightness across different views, which hurts overall visual coherence. DreamScene360 generates entire panoramas in a single step, which promotes global consistency but at the cost of reduced details and repetitive structures. FlexWorld occasionally produces noticeable artifacts, possibly due to limitations in video diffusion models. In contrast, our method generates visually and semantically rich scenes with consistent appearance across all viewpoints, effectively reconstructing complex environments with high fidelity.

3. RADAR CHART OF ABLATION STUDY RESULTS

Figure 4 shows a radar chart summarizing our ablation study, highlighting the contribution of each component to overall performance.

4. BI-DIRECTIONAL INPAINTING VISUAL RESULTS

Figure 5 and 6 provide additional visual results of bi-directional inpainting. We compare results obtained using only the left- or right-side prompt, direct averaging of the two prompts, and our proposed linear interpolation weighted sum.

5. USER STUDY

We conducted a user study with 26 participants to assess the perceptual quality of our method. Participants compared our results against FlexWorld [5], VistaDream [4], LucidDream [1], DreamScene360 [3], and Invisible Stitch [2] along three aspects: **Aesthetic Quality (Aes)**, **Diversity**, and **Alignment**. The aggregated win rates (%) across all participants are reported in Table 1, showing that our method consistently outperforms the baselines across all three aspects. The corresponding bar charts illustrating the user

study results are presented in Figures 7, 8, and 9. These figures provide a visual comparison between our method and the competing baselines.

Table 1: **Win rates (%) from our User Study.** We compare our method against five baselines across three aspects: Aesthetic Quality (Aes), Diversity, and Alignment.

Method	Aes (%)	Diversity (%)	Alignment (%)
Ours vs. FlexWorld [5]	84.62	88.46	69.23
Ours vs. VistaDream [4]	76.92	67.31	65.38
Ours vs. LucidDream [1]	76.92	76.92	78.85
Ours vs. DreamScene360 [3]	76.92	88.46	73.08
Ours vs. Invisible Stitch [2]	86.54	76.92	78.85

6. REFERENCES

[1] Jaeyoung Chung, Suyoung Lee, Hyeongjin Nam, Jaerin Lee, and Kyoung Mu Lee, "LucidDreamer: Domain-free generation of 3d gaussian splatting scenes," *arXiv preprint arXiv:2311.13384*, 2023.

[2] Paul Engstler, Andrea Vedaldi, Iro Laina, and Christian Rupprecht, "Invisible stitch: Generating smooth 3d scenes with depth inpainting," *arXiv preprint arXiv:2404.19758*, 2024.

[3] Shijie Zhou, Zhiwen Fan, Dejia Xu, Haoran Chang, Pradyumna Chari, Tejas Bharadwaj, Suya You, Zhangyang Wang, and Achuta Kadambi, "Dreamscene360: Unconstrained text-to-3d scene generation with panoramic gaussian splatting," in *ECCV*, 2024.

[4] Haiping Wang, Yuan Liu, Ziwei Liu, Wenping Wang, Zhen Dong, and Bisheng Yang, "VistaDream: Sampling multiview consistent images for single-view scene reconstruction," in *ICCV*, 2025.

[5] Luxi Chen, Zihan Zhou, Min Zhao, Yikai Wang, Ge Zhang, Wenhao Huang, Hao Sun, Ji-Rong Wen, and Chongxuan Li, "FlexWorld: Progressively expanding 3d scenes for flexiable-view synthesis," *arXiv preprint arXiv:2503.13265*, 2025.

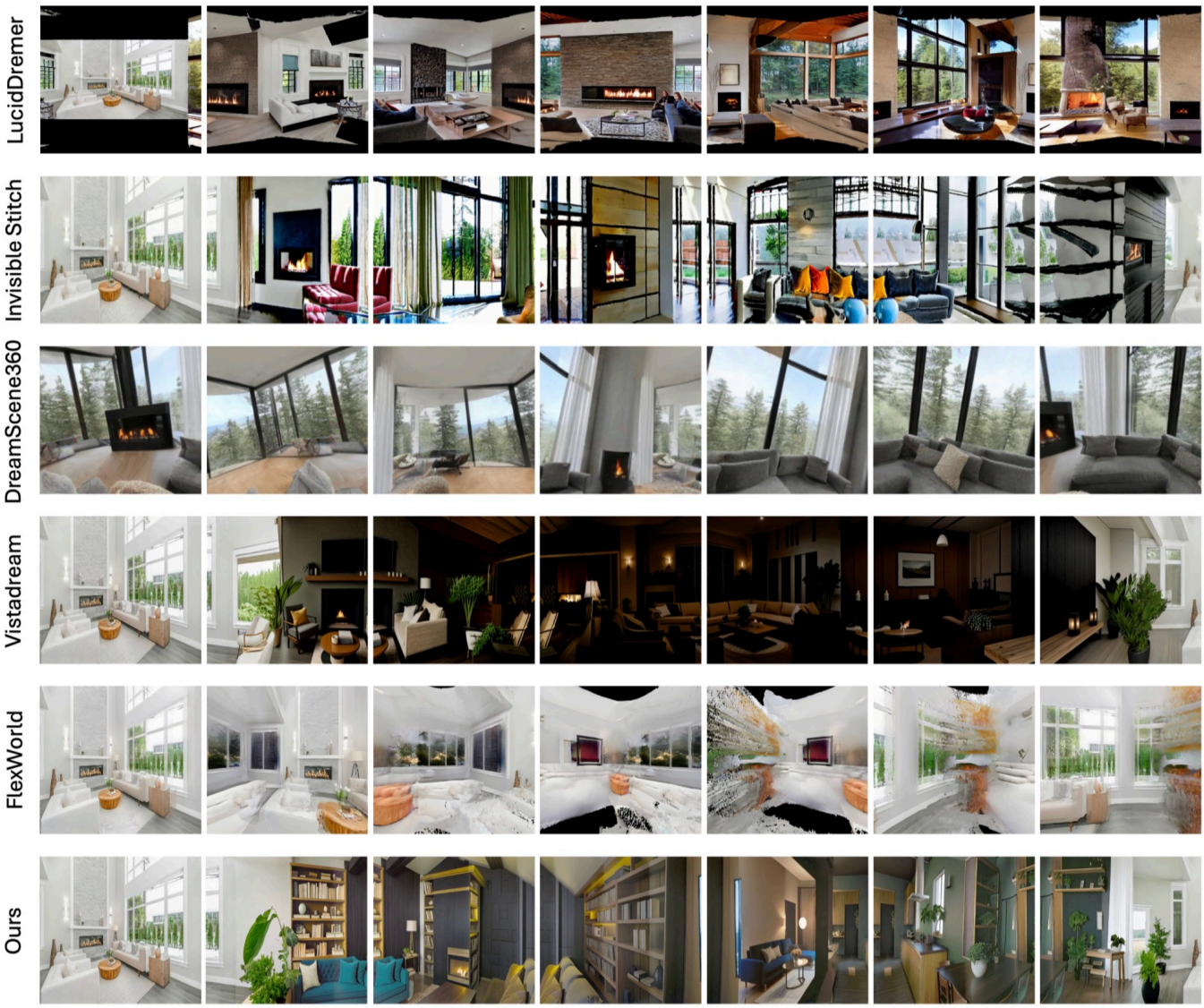


Fig. 1: **Qualitative results of 360 scene generation by different methods.** For each method, we display the input image on the left, followed by six synthesized images from different viewpoints. The used text prompt is provided at the bottom. Note that DreamScene360 is a text-to-3D method that relies solely on the text prompt, so the initial synthesized image may lack consistency with subsequent generations.

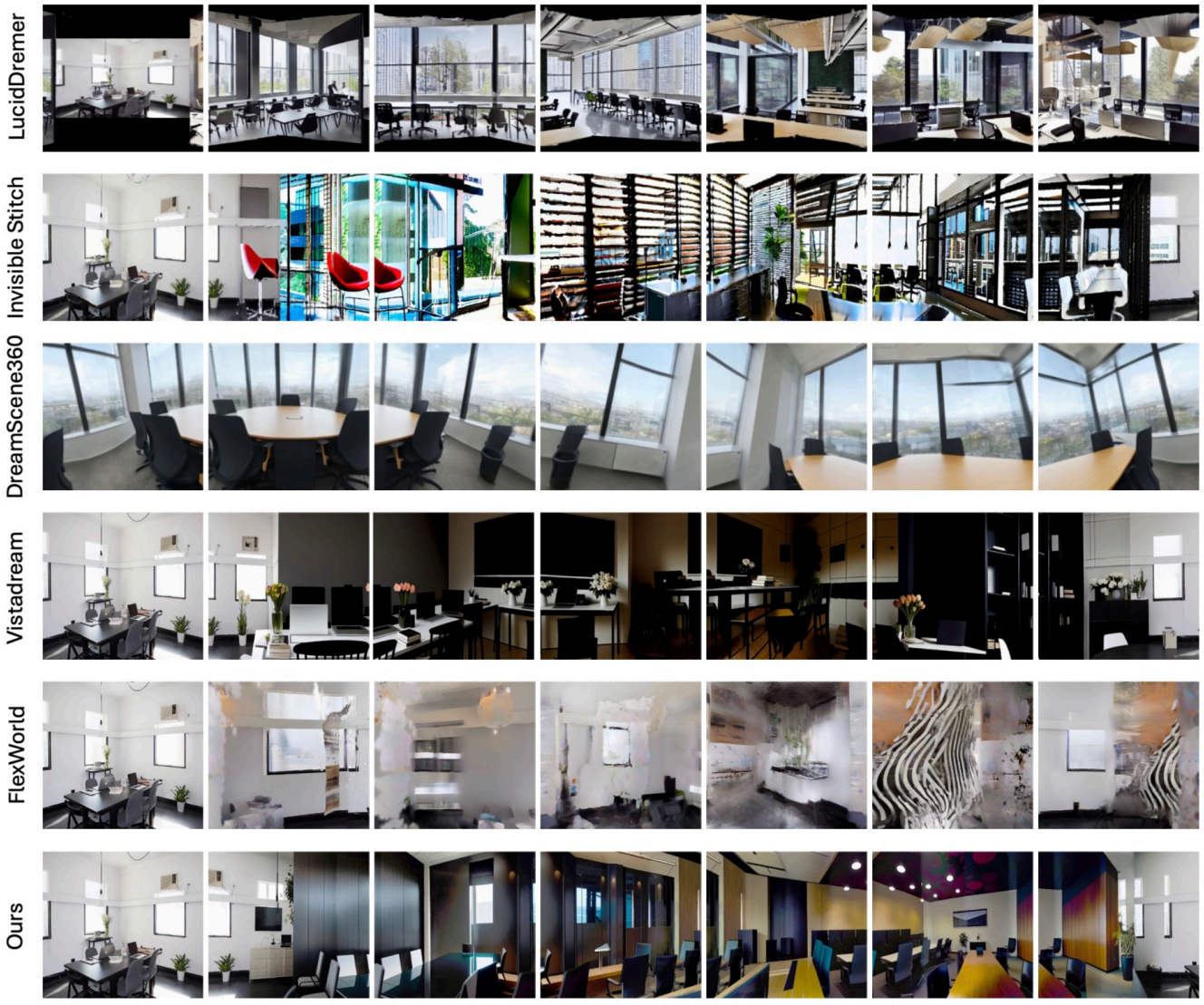


Fig. 2: **Qualitative results of 360 scene generation by different methods.** For each method, we display the input image on the left, followed by six synthesized images from different viewpoints. The used text prompt is provided at the bottom. Note that DreamScene360 is a text-to-3D method that relies solely on the text prompt, so the initial synthesized image may lack consistency with subsequent generations.

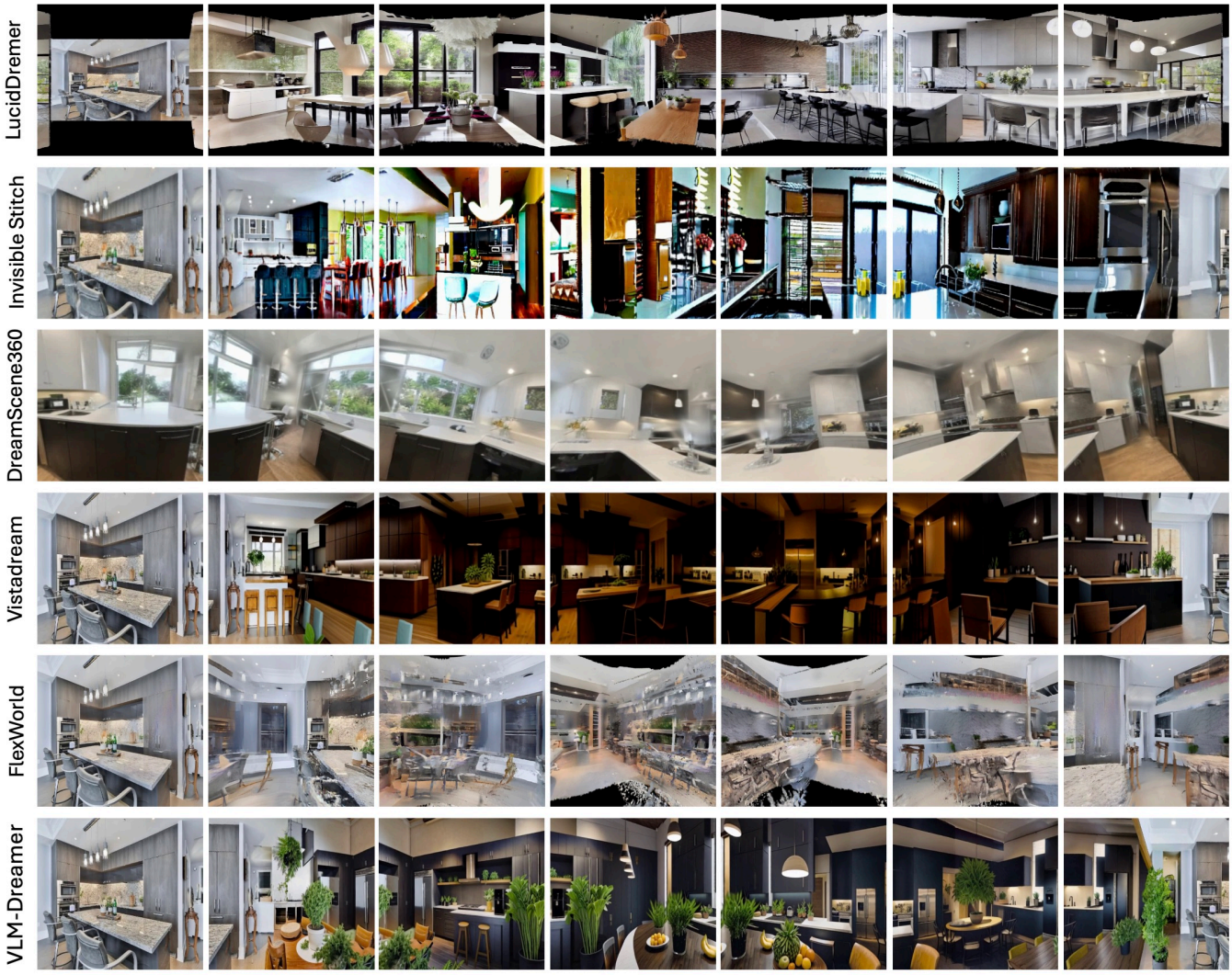


Fig. 3: **Qualitative results of 360 scene generation by different methods.** For each method, we display the input image on the left, followed by six synthesized images from different viewpoints. The used text prompt is provided at the bottom. Note that DreamScene360 is a text-to-3D method that relies solely on the text prompt, so the initial synthesized image may lack consistency with subsequent generations.

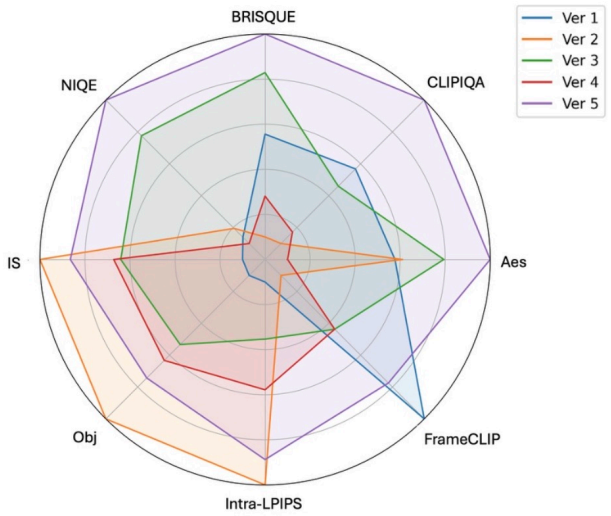
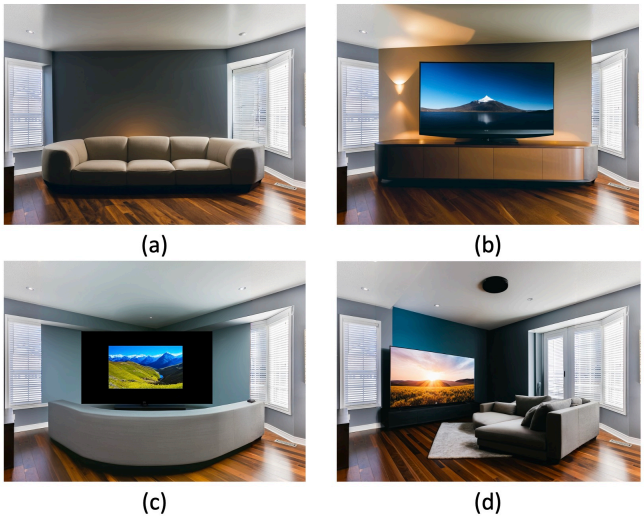
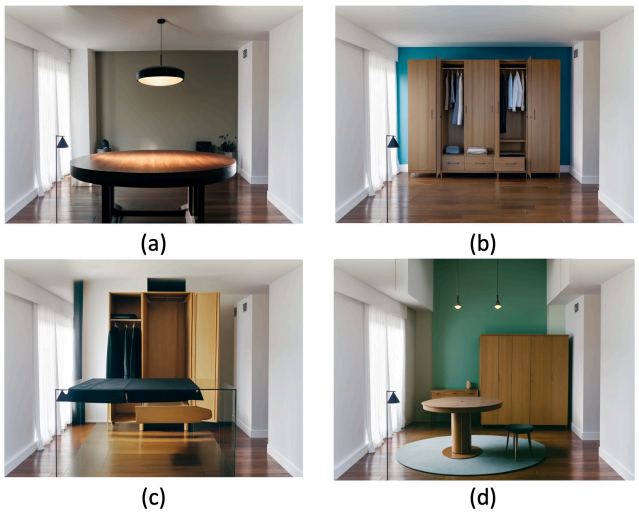


Fig. 4: **Radar chart comparing the performance of the ablation study.** For visualization clarity, each metric is linearly normalized to the range [0.1, 1.0], where a higher value indicates better performance. Metrics where lower is better (e.g., BRISQUE, NIQE) are inverted before normalization to maintain consistency.



Left-side prompt: “ A TV in the living room. ”
Right-side prompt: “ A sofa in the living room. ”

Fig. 5: **Visual Results of bi-directional inpainting.** (a) & (b) Two images are generated with guidance merely from the left-side and right-side prompts, respectively. (c) The image is synthesized by a direct average for two prompt integrations, showing mixed objects, i.e., the dining table and the refrigerator. (d) The image is produced by using our proposed linear interpolation weighted sum for directional integration of the two prompts, displaying the dining table on the left and the refrigerator on the right.



Left-side prompt: “ A Table in the room. ”
Right-side prompt: “ A Wardrobe in the room. ”

Fig. 6: **Visual Results of bi-directional inpainting.** (a) & (b) Two images are generated with guidance merely from the left-side and right-side prompts, respectively. (c) The image is synthesized by a direct average for two prompt integrations, showing mixed objects, i.e., the dining table and the refrigerator. (d) The image is produced by using our proposed linear interpolation weighted sum for directional integration of the two prompts, displaying the dining table on the left and the refrigerator on the right.

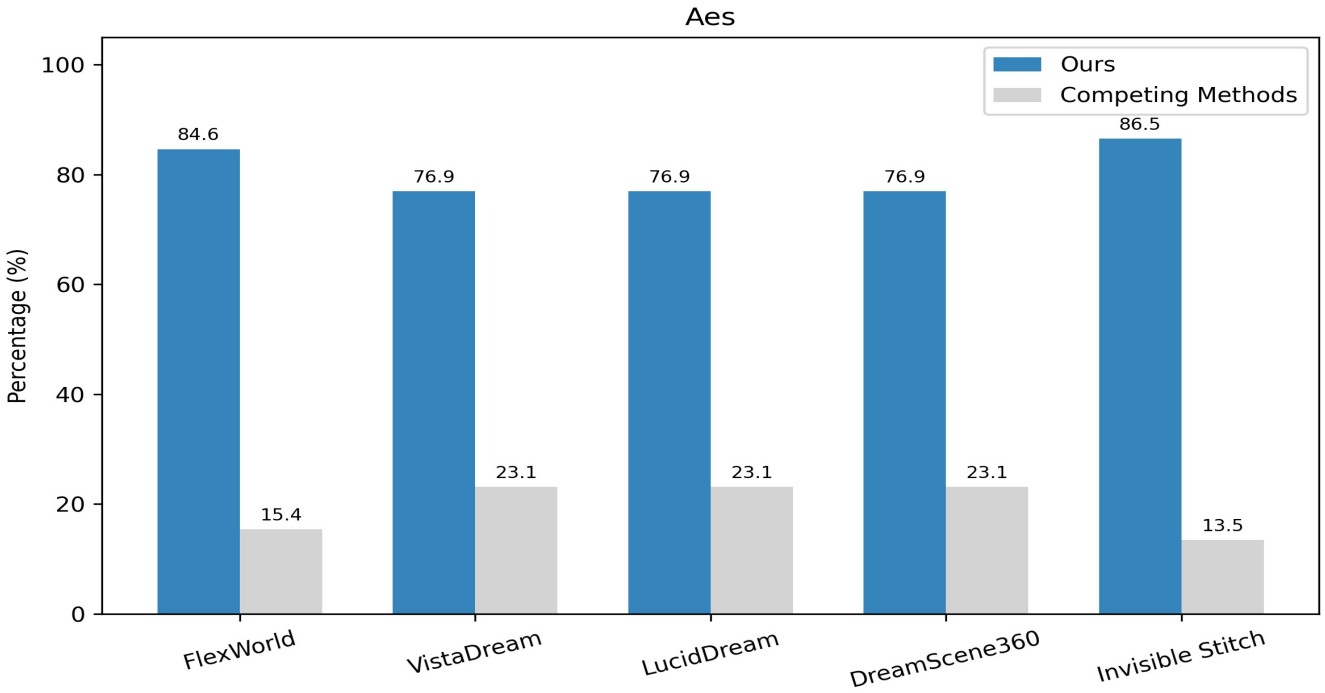


Fig. 7: **User study results – Aesthetic Quality (Aes).** Participants compared our method against FlexWorld, VistaDream, LucidDream, DreamScene360, and Invisible Stitch. The bar chart shows the aggregated win rates for Aesthetic Quality.

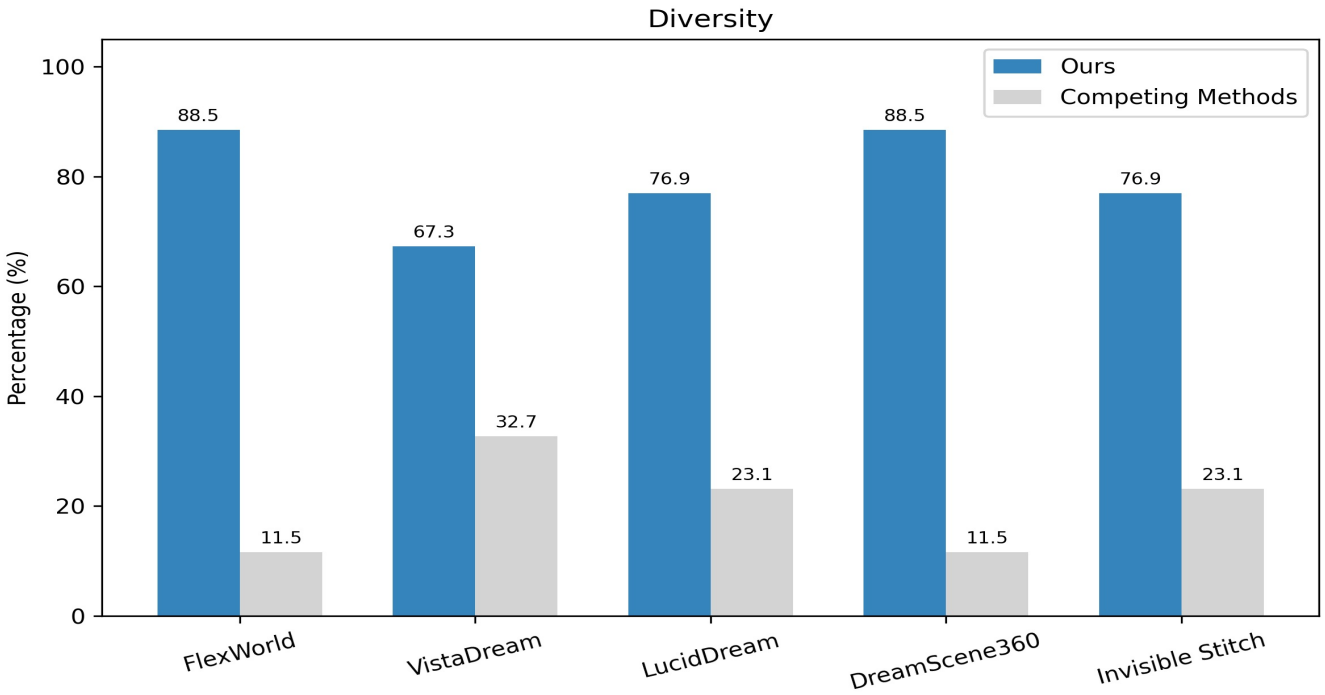


Fig. 8: **User study results – Diversity.** Participants compared our method against FlexWorld, VistaDream, LucidDream, DreamScene360, and Invisible Stitch. The bar chart shows the aggregated win rates for Diversity.

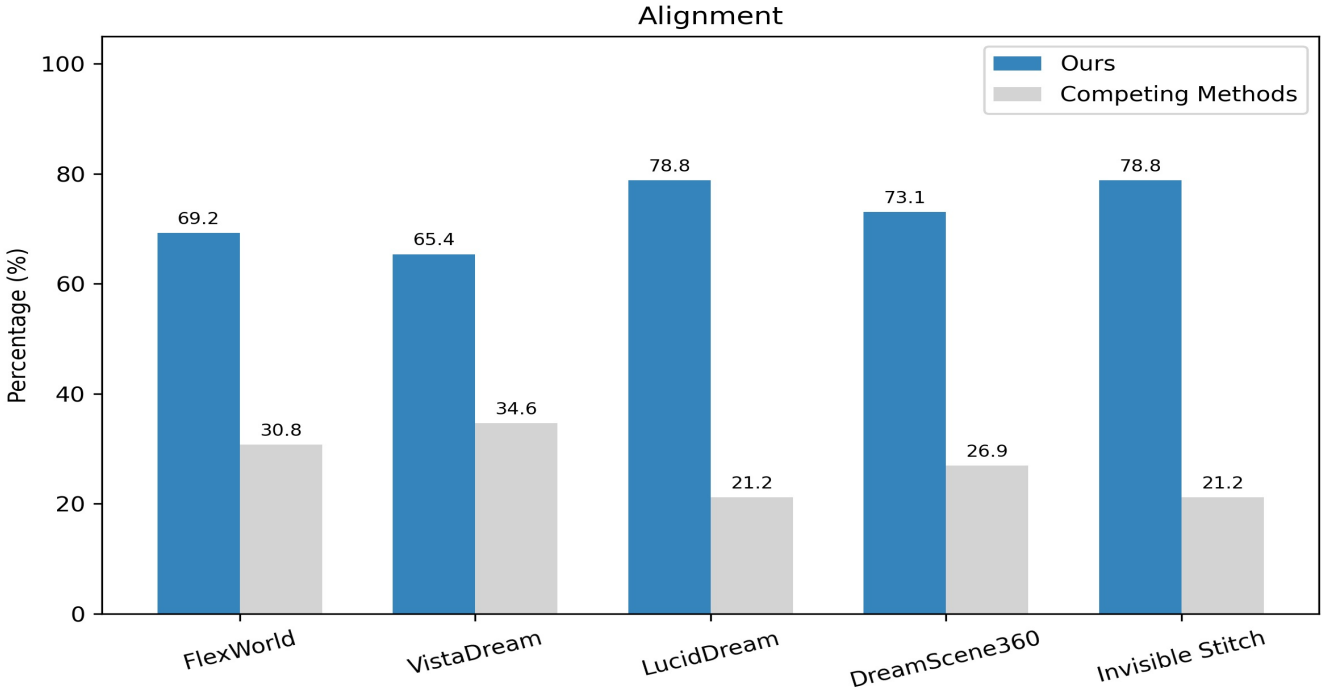


Fig. 9: **User study results – Alignment.** Participants compared our method against FlexWorld, VistaDream, LucidDream, DreamScene360, and Invisible Stitch. The bar chart shows the aggregated win rates for Alignment.