FlipSketch: Flipping Static Drawings to Text-Guided Sketch Animations

Supplementary Material

The squirrel uses its dexterous front paws to hold and manipulate nuts, displaying meticulous and deliberate motions while eating.

The wine in the wine glass sways from side to side.

The dragon is flying slowly by flapping its wings.

A galloping horse.

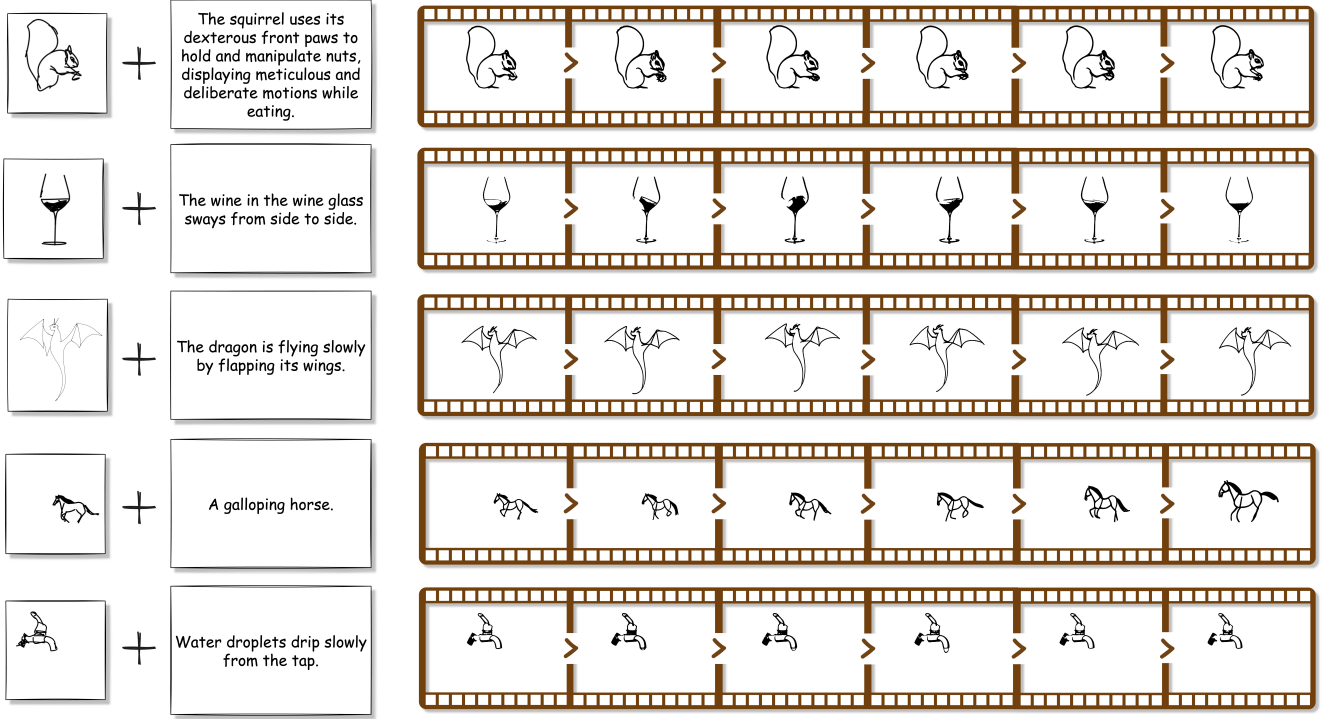Water droplets drip slowly from the tap.

Figure S1. Additional results for qualitative analysis

S1. Accompanying GIFs and videos

We find that some motions (like "a flower swaying in the wind") are too subtle to depict in the form of frame sequences (like Fig. S2). We include such animations along with GIFs for the rest of the animations from the paper in accompanying "supplementary_videos" folder. We also include a ∼ 1 minute teaser video, which we create with the help of "keynote" (a slide-show application) using results from our method. We include the constituent GIFs of the teaser video in the "teaser_props" folder.

S2. Additional results and comparisons

We include comparisons with Animated Drawings [2] in Fig. S2 for completeness. Animated Drawings performs animations from motion templates hence doesn't need a text prompt. We display an animation from the "dance" template here, demonstrating that the algorithm expects humanoid sketches. We also include more results on sketch animation (Fig. S1) including animations for sketches stylised differently (see "dragon" example) than rounded bezier strokes in CLIPasso sketches [3].

The wine in the wine glass sways from side to side

Figure S2. Qualitative comparisons with Animated Drawings [2]

S3. Faithfulness to input prompt

To analyse faithfulness to input prompt, we animate the same sketch with two different prompt configurations. We find the resulting animations (Fig. S4) align closely with their respective input prompts, demonstrating a high degree of faithfulness.

S4. User study

We include an example from the user study in Fig. S5. Each user gets 10 sets of such examples. For each example, they perform three tasks: (i) judging alignment to text prompt, (ii) judging consistency with input sketch and (iii) a general scoring of every GIF (based on text+sketch input). The
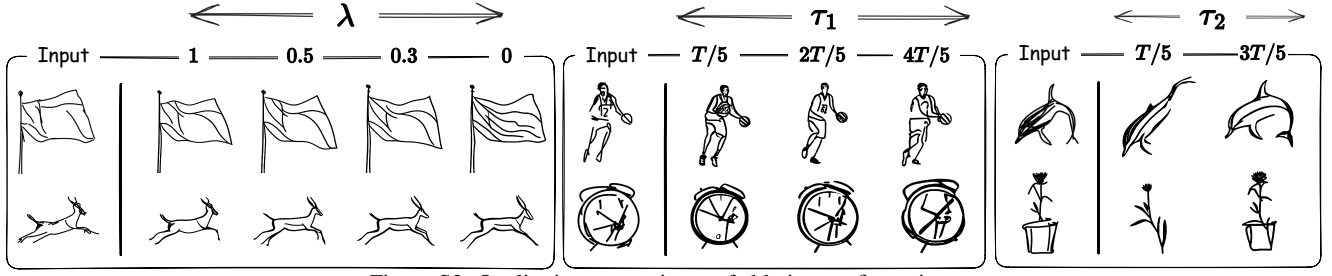
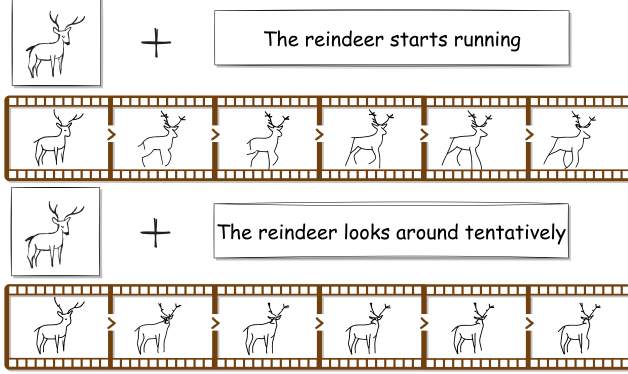Figure S3. Qualitative comparisons of ablative configurations

Figure S4. Qualitative analysis of faithfulness to input prompt

GIFs are randomly arranged and have to be dragged to form the desired rankings for the first two tasks. The final task consists of a slider to adjust the score from 0 to 1.
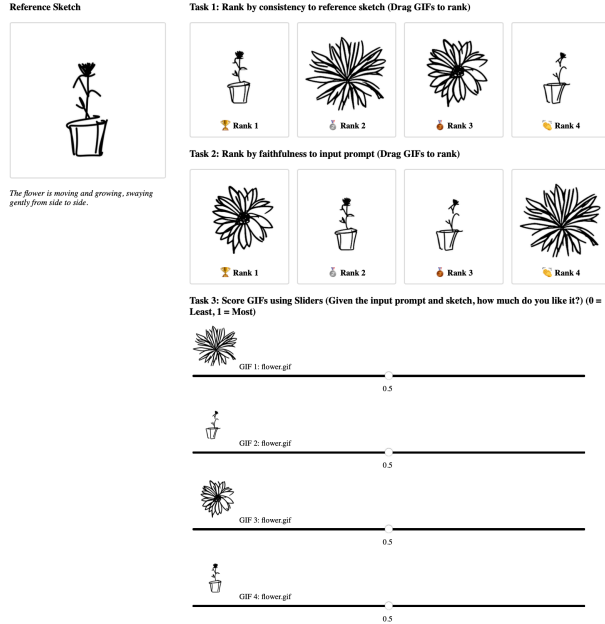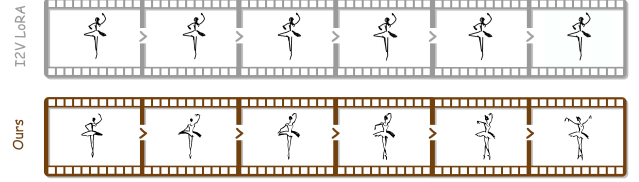

Figure S5. Setup for user study

Figure S6. Qualitative comparison with I2V LoRA model

S5. Ablative studies (Contd.)

We include additional ablative studies for various levels of λ, demonstrating significant control over range of motion in Fig. S3. We also include ablations with changing values for hyperparameters τ_1, τ_2 as fractions of total time-steps T.

S6. Choice of backbone

T2V v/s I2V adaptation: We adapt a T2V model for sketch animation rather than an I2V model, as I2V models [4] strive for high correlation with the input image in their animations. For the image space, this correlation is crucial for preserving input identity, but comes at the cost of limited and restricted motion (see Fig. S6 for comparison with a LoRA of SVD [1]). Sketches, however, are more flexible than images when it comes to identity. We find that we can impose sketch identity on existing T2V models, and explore the identity v/s motion trade-off better with T2V generation.

References

[1] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023. 2

[2] Harrison Jesse Smith, Qingyuan Zheng, Yifei Li, Somya Jain, and Jessica K Hodgins. A method for animating children's drawings of the human figure. *ACM TOG*, 2023. 1

[3] Yael Vinker, Ehsan Pajouheshgar, Jessica Y Bo, Roman Christian Bachmann, Amit Haim Bermano, Daniel Cohen-Or, Amir Zamir, and Ariel Shamir. Clipasso: Semantically-aware object sketching. *ACM TOG*, 2022. 1

[4] Yudian Zheng, Xiaodong Cun, Menghan Xia, and Chi-Man Pun. Sketch video synthesis. In *CGF*, 2024. 2