

Probability Density Geodesics in Image Diffusion Latent Space

Supplementary Material

In this appendix, we provide the full derivations for the mathematical results presented in the main paper and additional (especially qualitative) experimental results.

A. Derivations

In this section, we provide the derivations deferred from the main paper. We formulate the problem in more general terms—as a weighted path length—before returning to the specifics used in the main paper.

Equation for path length. Let $\gamma : [a, b] \rightarrow \mathbb{R}^n$ be a path such that $\gamma(a) = \mathbf{x}_a$ and $\gamma(b) = \mathbf{x}_b$, and $S : \{\gamma_i\} \rightarrow \mathbb{R}$ be the action functional on the set of such paths, defined by

$$S[\gamma] = \int_a^b L(t, \gamma(t), \dot{\gamma}(t)) dt, \quad (15)$$

where L is the Lagrangian given by

$$L(t, \gamma(t), \dot{\gamma}(t)) = \|\dot{\gamma}(t)\| w(\gamma(t)). \quad (16)$$

$$L(t, \gamma(t), \dot{\gamma}(t)) = \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle}_{K(\gamma(t))} \quad (17)$$

$$K(\gamma(t)) = w(\gamma(t))^2 I. \quad (18)$$

Then $S[\gamma]$ is the weighted path length for path γ .

Euler–Lagrange equations. Given this definition, a path γ is a stationary point of S iff it satisfies the Euler–Lagrange equations, viz.,

$$\frac{\partial}{\partial \gamma} L(t, \gamma(t), \dot{\gamma}(t)) - \frac{d}{dt} \frac{\partial}{\partial \dot{\gamma}} L(t, \gamma(t), \dot{\gamma}(t)) = 0, \quad (19)$$

where $\frac{\partial L}{\partial \gamma}$ stacks the partial derivatives w.r.t. the components of γ and $\frac{\partial L}{\partial \dot{\gamma}}$ stacks the partial derivatives w.r.t. the components of $\dot{\gamma}$. We obtain

$$0 = \frac{\partial}{\partial \gamma} L(t, \gamma(t), \dot{\gamma}(t)) - \frac{d}{dt} \frac{\partial}{\partial \dot{\gamma}} L(t, \gamma(t), \dot{\gamma}(t)) \quad (20)$$

$$= \frac{\partial}{\partial \gamma} (\|\dot{\gamma}(t)\| w(\gamma(t))) - \frac{d}{dt} \frac{\partial}{\partial \dot{\gamma}} (\|\dot{\gamma}(t)\| w(\gamma(t))) \quad (21)$$

$$= \|\dot{\gamma}\| \frac{dw}{d\gamma} - \frac{d}{dt} \left(\frac{\dot{\gamma}}{\|\dot{\gamma}\|} w(\gamma(t)) \right) \quad (22)$$

$$= \|\dot{\gamma}\| \frac{dw}{d\gamma} - \frac{d}{dt} \left(\frac{\dot{\gamma}}{\|\dot{\gamma}\|} \right) w(\gamma) - \frac{\dot{\gamma}}{\|\dot{\gamma}\|} \frac{d}{dt} (w(\gamma(t))) \quad (23)$$

$$= \|\dot{\gamma}\| \frac{dw}{d\gamma} - \frac{1}{\|\dot{\gamma}\|} \left(I - \frac{\dot{\gamma} \dot{\gamma}^\top}{\|\dot{\gamma}\|^2} \right) \ddot{\gamma} w - \frac{\dot{\gamma}}{\|\dot{\gamma}\|} \left(\frac{dw^\top}{d\gamma} \frac{d\gamma}{dt} \right). \quad (24)$$

Multiplying both sides by $\|\dot{\gamma}\|/w$, we obtain

$$0 = \|\dot{\gamma}\|^2 \frac{1}{w} \frac{dw}{d\gamma} - \left(I - \frac{\dot{\gamma} \dot{\gamma}^\top}{\|\dot{\gamma}\|^2} \right) \ddot{\gamma} - \dot{\gamma} \frac{1}{w} \frac{dw^\top}{d\gamma} \dot{\gamma} \quad (25)$$

$$= \|\dot{\gamma}\|^2 \nabla \log w - \left(I - \hat{\gamma} \hat{\gamma}^\top \right) \ddot{\gamma} - \langle \nabla \log w, \dot{\gamma} \rangle \dot{\gamma}, \quad (26)$$

where we use that $\nabla \log w = \frac{1}{w} \frac{dw}{d\gamma}$. Rearranging, we obtain

$$\underbrace{\left(I - \hat{\gamma} \hat{\gamma}^\top \right)}_{\perp \dot{\gamma}} \frac{\ddot{\gamma}}{\|\dot{\gamma}\|^2} = \underbrace{\left(I - \hat{\gamma} \hat{\gamma}^\top \right)}_{\perp \dot{\gamma}} \nabla \log w(\gamma), \quad (27)$$

where the unit velocity is given by $\hat{\gamma} = \dot{\gamma}/\|\dot{\gamma}\|$. In other words, we obtain a relationship between quantities that are both perpendicular to the velocity, one a component of the scaled acceleration and the other a component of the gradient of the log weight.

For a constant speed parameterization, we observe that acceleration in the direction of the path must be zero, and so $(I - \hat{\gamma} \hat{\gamma}^\top) \ddot{\gamma} = \ddot{\gamma}$. We therefore obtain

$$\ddot{\gamma} = \|\dot{\gamma}\|^2 \left(I - \hat{\gamma} \hat{\gamma}^\top \right) \nabla \log w(\gamma). \quad (28)$$

Functional derivative. This second-order ODE expresses the relationship at optimality, i.e., given an initial position and velocity we can obtain the associated optimal path. However, we can also derive the functional derivative $\frac{\delta S}{\delta \gamma}$ of the path length functional S by approximating the curve by a polygonal line with n segments, as n grows arbitrarily large. We obtain, for any (potentially sub-optimal) path γ ,

$$\frac{\delta S}{\delta \gamma} = \frac{w(\gamma)}{\|\dot{\gamma}\|} \left(I - \hat{\gamma} \hat{\gamma}^\top \right) \left(\nabla \log w(\gamma) - \frac{\ddot{\gamma}}{\|\dot{\gamma}\|^2} \right), \quad (29)$$

or, for a constant speed parameterization,

$$\frac{\delta S}{\delta \gamma} = \frac{w(\gamma)}{\|\dot{\gamma}\|} \left(\left(I - \hat{\gamma} \hat{\gamma}^\top \right) \nabla \log w(\gamma) - \frac{\ddot{\gamma}}{\|\dot{\gamma}\|^2} \right). \quad (30)$$

High-probability geodesics. In our case of interest, the weight is inversely proportional to the probability density, that is,

$$w(\gamma) = p(\gamma)^{-1} \text{ and } \nabla \log w(\gamma) = -\nabla \log p(\gamma), \quad (31)$$

giving us, for a constant speed parameterization of the path, the following second-order ODE expressing the optimality condition

$$\ddot{\gamma} + \|\dot{\gamma}\|^2 \left(I - \hat{\gamma} \hat{\gamma}^\top \right) \nabla \log p(\gamma) = 0 \quad (32)$$

and the functional derivative

$$\frac{\delta S}{\delta \gamma} = \frac{-1}{p(\gamma)\|\dot{\gamma}\|} \left((I - \hat{\gamma}\hat{\gamma}^\top) \nabla \log p(\gamma) + \frac{\ddot{\gamma}}{\|\dot{\gamma}\|^2} \right). \quad (33)$$

B. Further Implementation Details

For the score function ϕ in Eq. (12), we use a uniform weight function $w(\tau) = 1$, and the output is normalized by $1 + \sigma$. For the negative text prompt, we used ‘‘A doubling image, unrealistic, artifacts, distortions, unnatural blending, ghosting effects, overlapping edges, harsh transitions, motion blur, poor resolution, low detail’’ for all the experiments. In the inference process of image interpolation, we applied the same perceptually-uniform sampling strategy as Zhang et al. [53] to produce an image sequence with a more homogeneous transition rate, using histogram equalization. In both the deterministic DDIM forward (DDIM-F) and backward (DDIM-B) processes on BVP and IVP, we set the classifier-free guidance scale (CFG) to 1 and use the same positive conditional embedding as the one used in ϕ (text-inverted). After optimizing each point x in the path, we project the point back to the sphere by scaling the norm of x to the radius of the sphere. For IVP, we aim to generate an initial velocity that points towards the distribution of the target prompt. Given a source and target text embedding z_0 and z_1 , we compute a pseudo target x_{tgt} by optimizing the initial latent vector x_0 using the score function $\phi(x|mz_1 + (1 - m)z_0, \tau)$ with $m = 0.8$, learning rate as 1 and number of iteration as 300. Then the initial velocity is set as $x_{\text{tgt}} - x_0$ projected to the tangent space of the sphere.

C. Further Details on Evaluation Metrics

For the TOPIQ score, we weight it to emphasize the quality of the middle frames of the generated sequence, as they tend to be farther from the source images and more indicative of the overall perceptual quality. Instead of a simple average, we compute a weighted TOPIQ score as:

$$\text{TOPIQ}(\{I_\lambda\}_{\lambda \in [0,1]}) = \frac{\sum_\lambda w(\lambda)\text{TOPIQ}(I_\lambda)}{\sum_\lambda w(\lambda)}, \quad (34)$$

where $w(\lambda) = \lambda$ for $\lambda \leq 0.5$ and $w(\lambda) = 1 - \lambda$ for $\lambda > 0.5$.

D. Additional Ablation Study on the Conditioning Signal

We compare two types of conditioning signals—constant versus linearly varying along the path as discussed in Sec. 4.1. For constant conditioning, the text embedding is initialized as $p_0 + p_1$ and then fine-tuned using text inversion. For time-linear conditioning, we apply text inversion separately to both prompts of the image pair and interpolate

Table 3. Ablation study on the validation dataset that ablates the time-dependence of the conditioning signal (z_t) and the geodesic optimization.

z_t	Opt.	FID↓	PPL↓	PDV↓	TOPIQ↑
		45.30	0.853	0.037	0.516
	✓	55.80	0.931	0.020	0.541
✓		49.39	0.917	0.035	0.526
✓	✓	63.28	1.000	0.017	0.553

their embeddings as $z_t = (1 - t)z_0 + tz_1$, where we use the shorthand $z_t = \zeta(t)$ in this section. As shown in Tab. 3, constant conditioning results in a distribution closer to the input images (as measured by FID), while the time-linear conditioning yields higher image quality (TOPIQ) and a more homogeneous transition rate (PDV). The main paper reports results using time-linear conditioning (see Tab. 1).

E. Sensitivity Analysis of the Hyperparameters

We analyze several key hyperparameters of our method, as illustrated in Fig. 8. The parameter β controls the trade-off between path directness and alignment with high-probability regions. A larger β encourages the path to move toward regions of higher probability density at the expense of directness, while a smaller β keeps the path more direct. The diffusion timestep τ influences the level of detail in the generated images. A higher τ tends to morph the high-level image features but may result in a loss of fine details, whereas a very low τ can degrade image quality due to insufficient denoising. A properly chosen sampling range $\Delta\tau$ can help the path escape local minima compared to using a zero range. However, if $\Delta\tau$ is too large, the guidance signal gets weaker, resulting in smoother paths but lower FID and TOPIQ scores. These parameters exhibit interpretable behavior, allowing users to make choices based on their specific needs.

F. Further Qualitative Results

In this section, we present additional qualitative results and failure cases, as shown in Figs. 9 to 16.

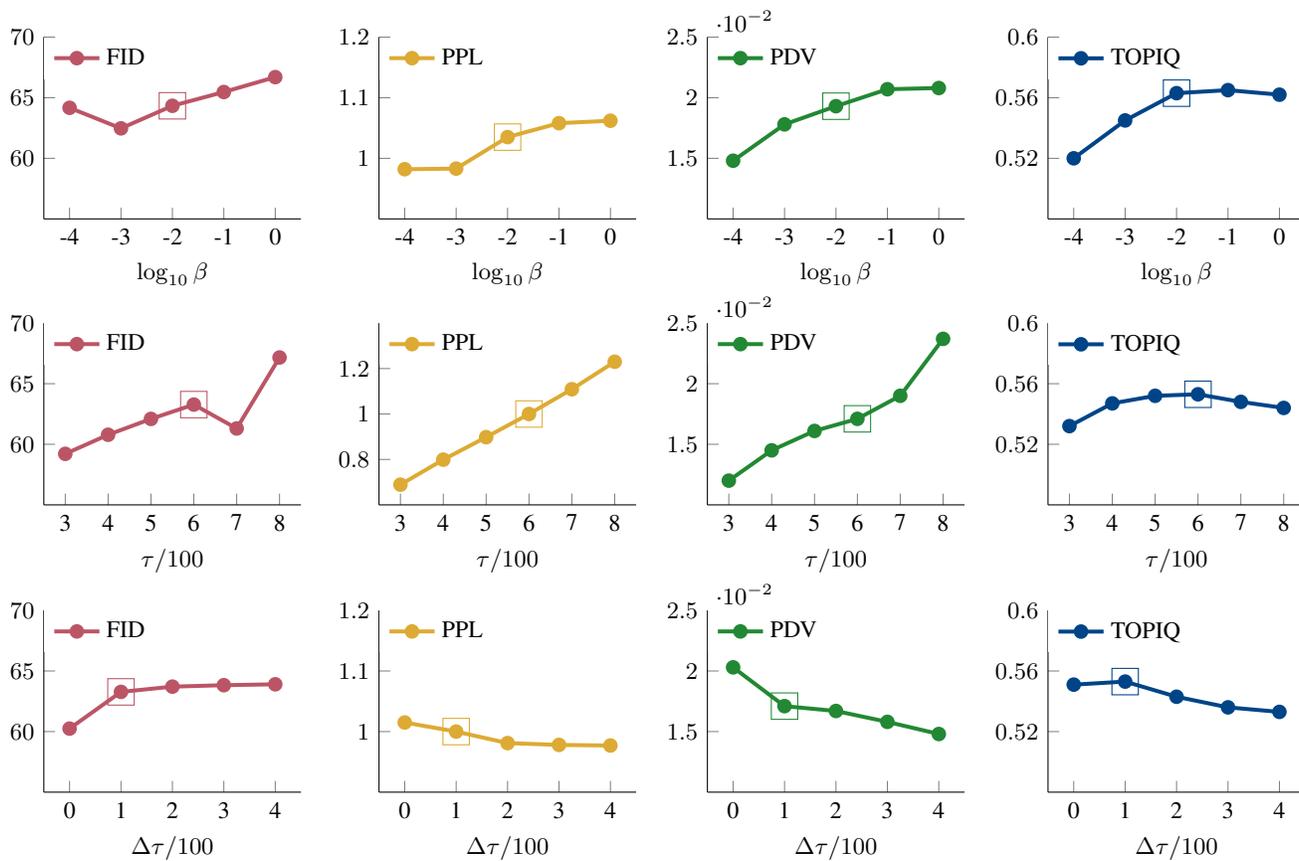


Figure 8. The quantitative analysis of selecting hyperparameters $\beta, \tau, \Delta\tau$. The default settings are $\beta = 0.002, \tau = 0.6, \Delta\tau = 100$, which are marked in the plots as squares.

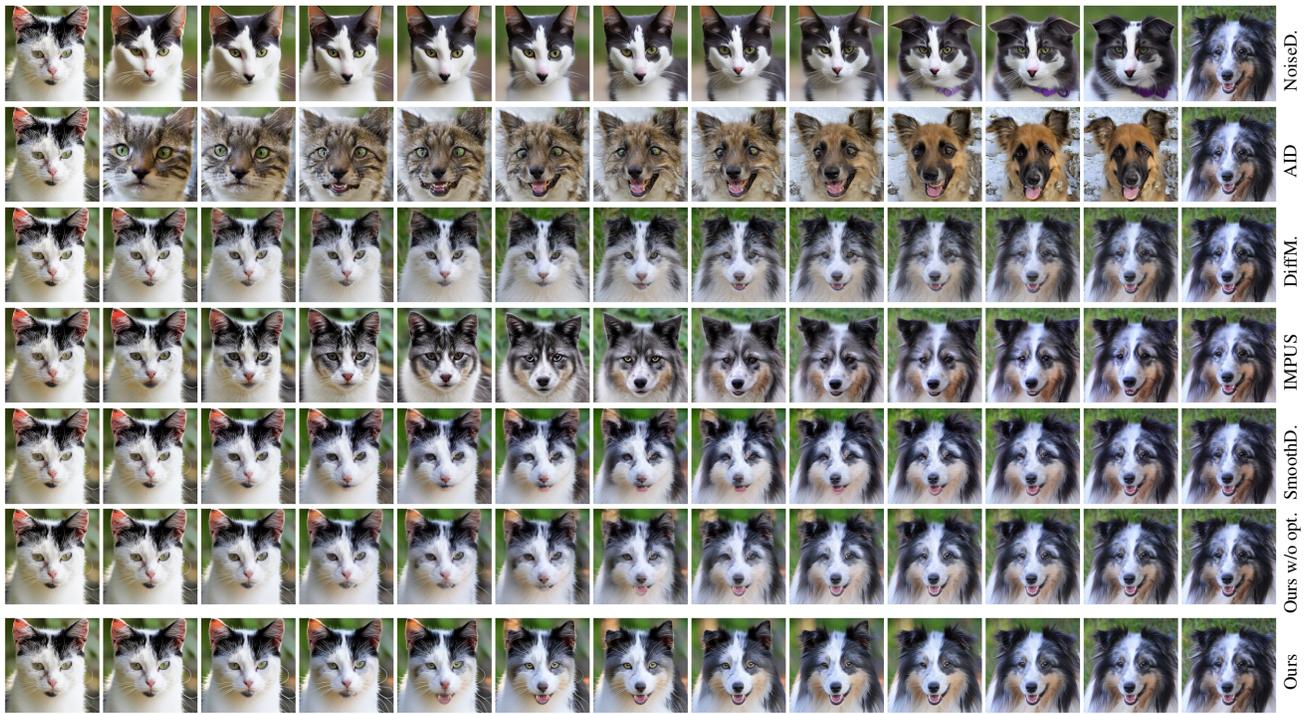


Figure 9. Qualitative image interpolation results, comparing all methods.

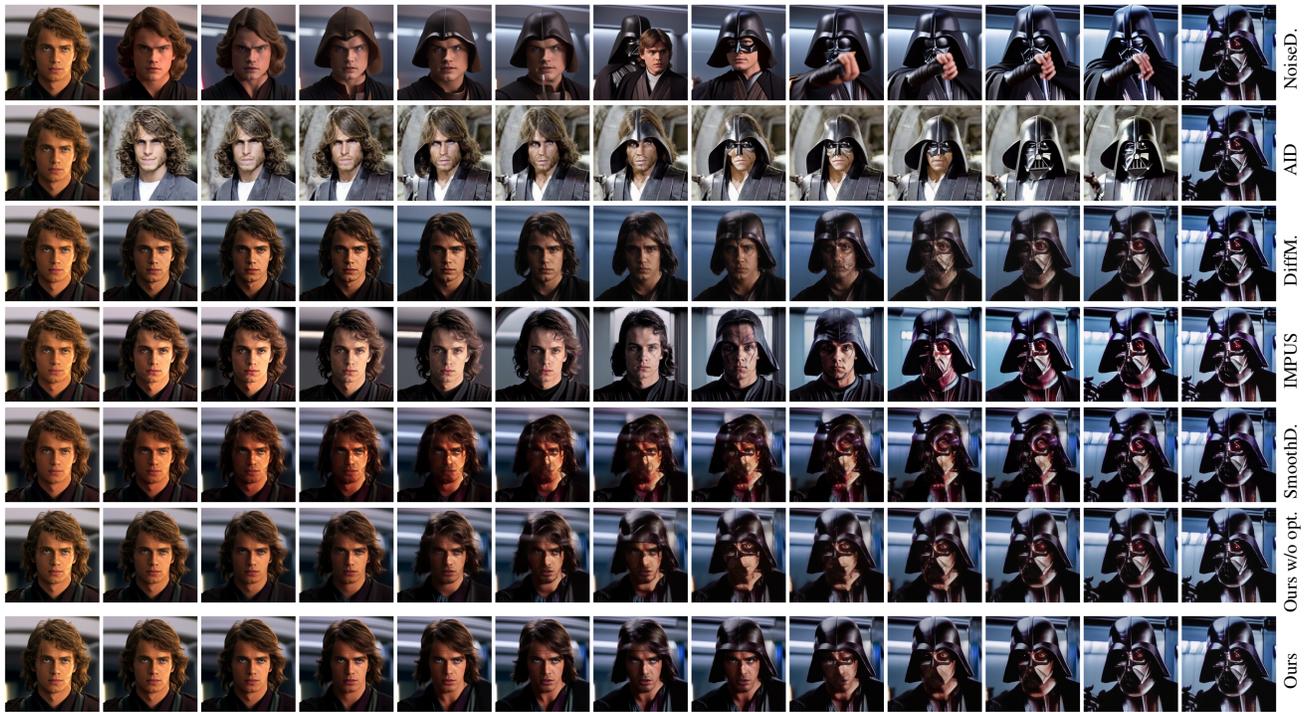


Figure 10. Qualitative image interpolation results, comparing all methods.

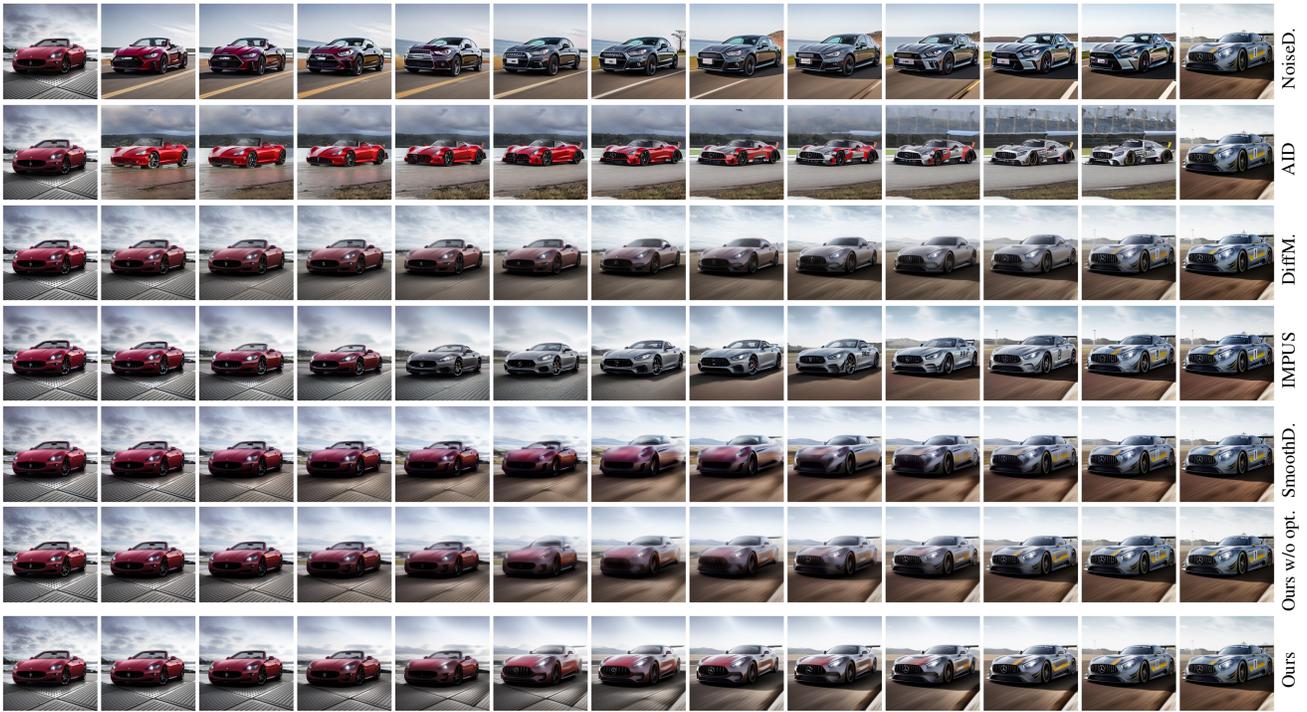


Figure 11. Qualitative image interpolation results, comparing all methods.

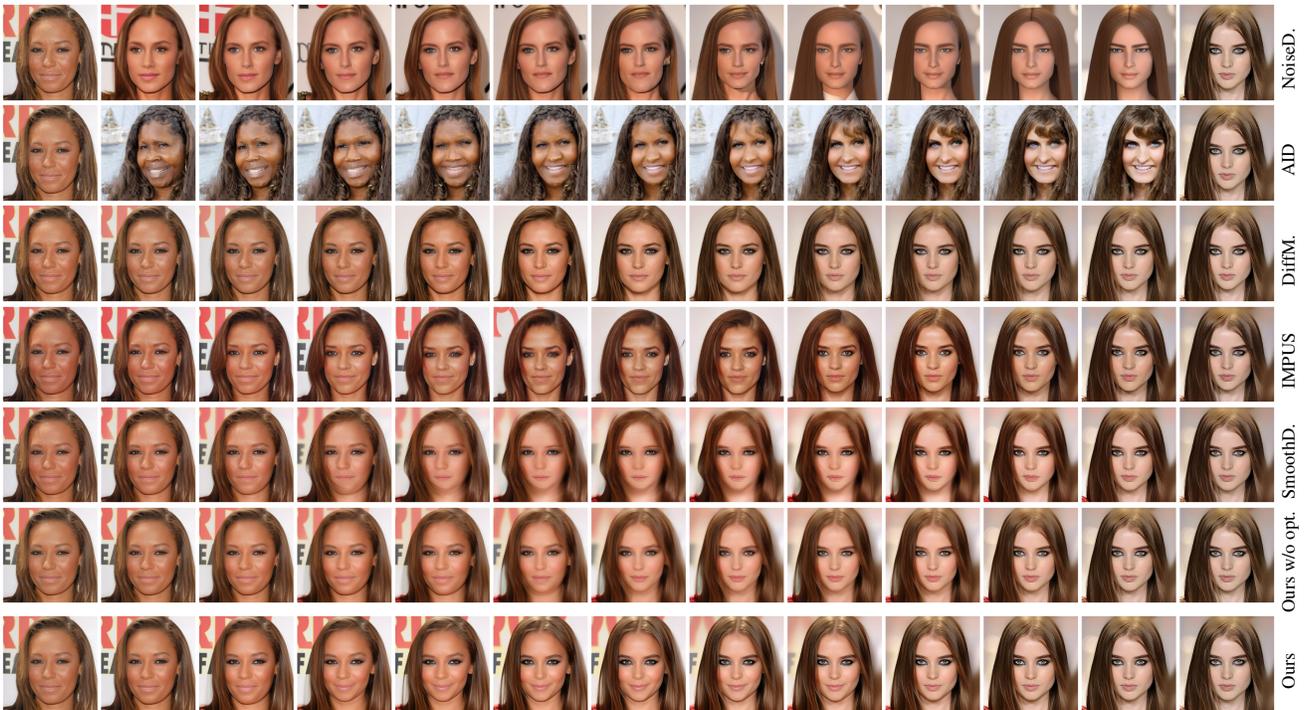


Figure 12. Qualitative image interpolation results, comparing all methods.

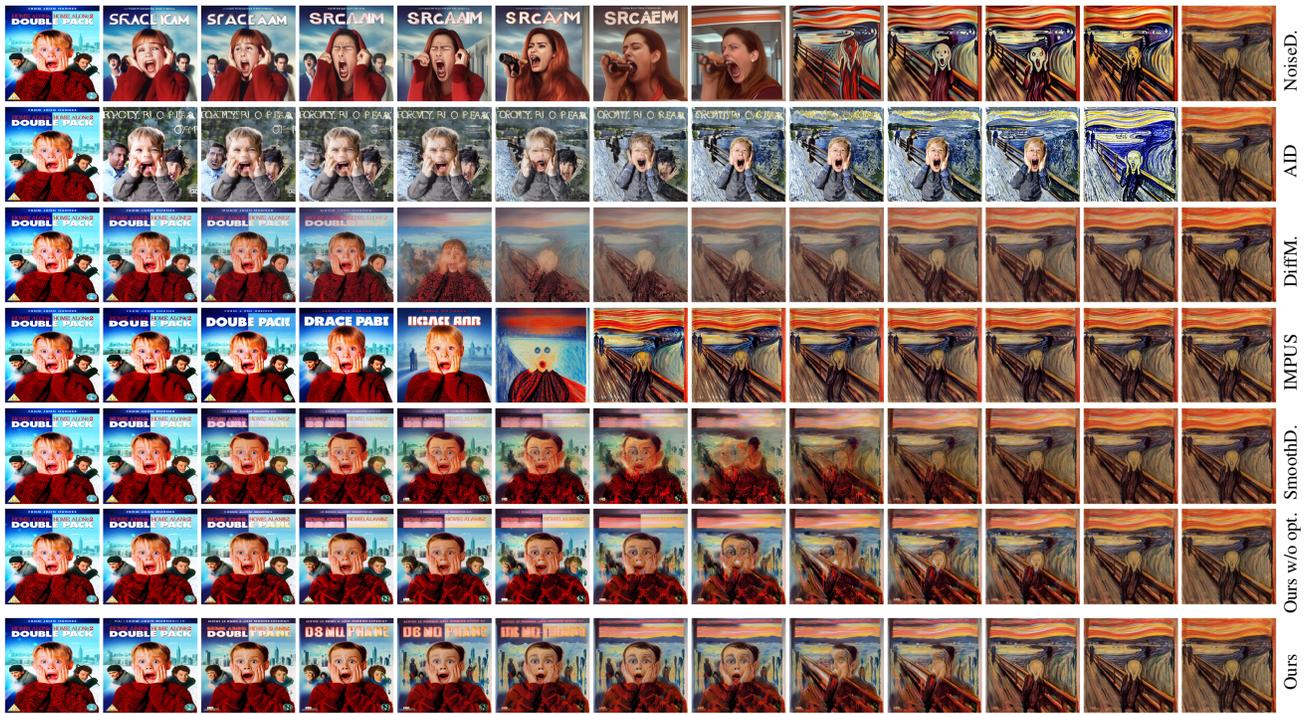


Figure 13. Qualitative image interpolation results, comparing all methods.



Figure 14. Qualitative image interpolation results, comparing all methods.

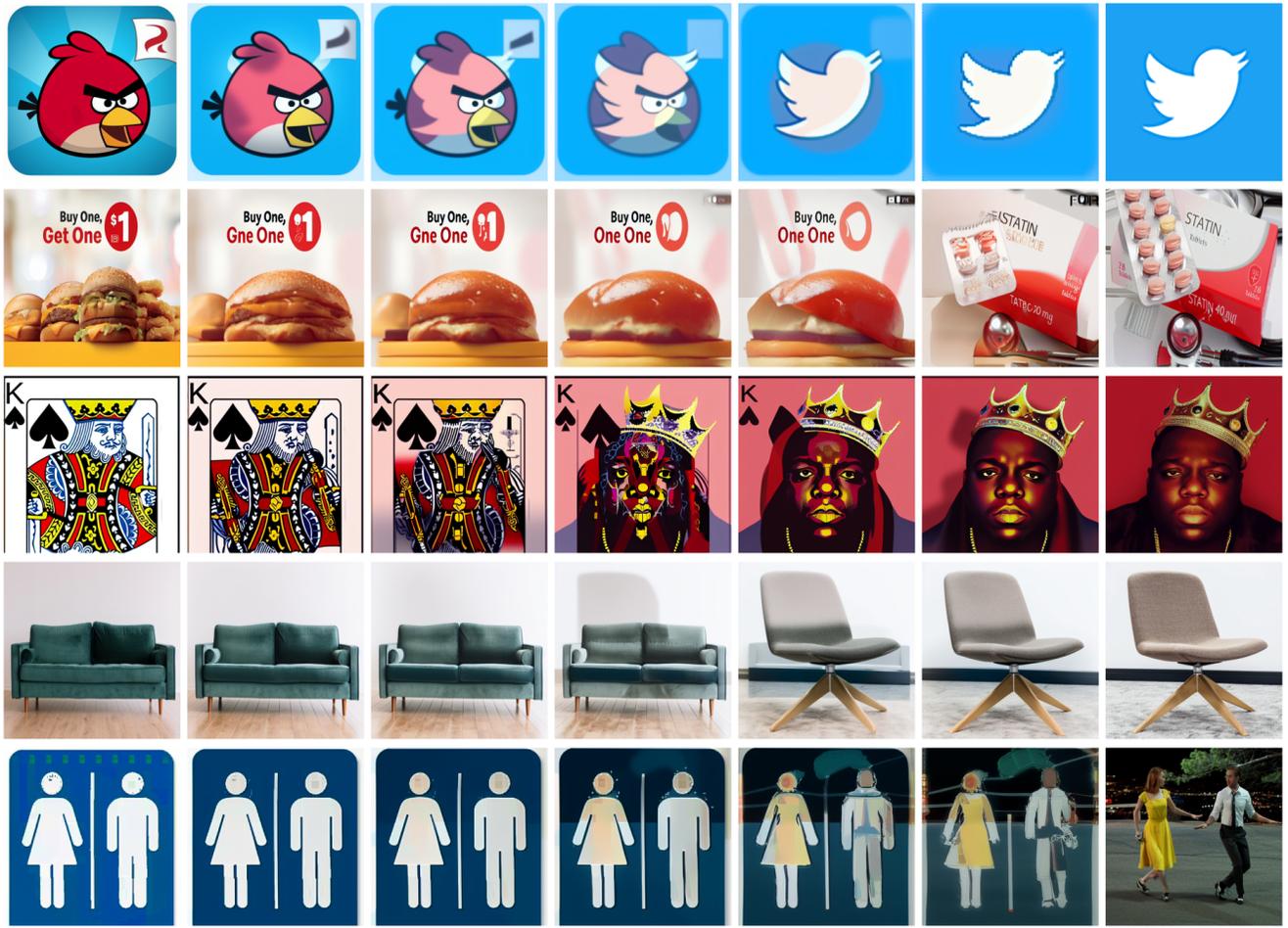


Figure 15. Image interpolation failure cases. Here we show examples with a significant appearance or semantic gap between the image pairs, where the computed geodesic is unable to smoothly connect the two.



Figure 16. Image interpolation failure cases. Here we show how smoothness in image space does not necessarily correspond to smoothness in the projected 3D world. For example, in the top row we see a shadow boundary become a ridge line, and in the bottom row we see a mountain become a roof line.