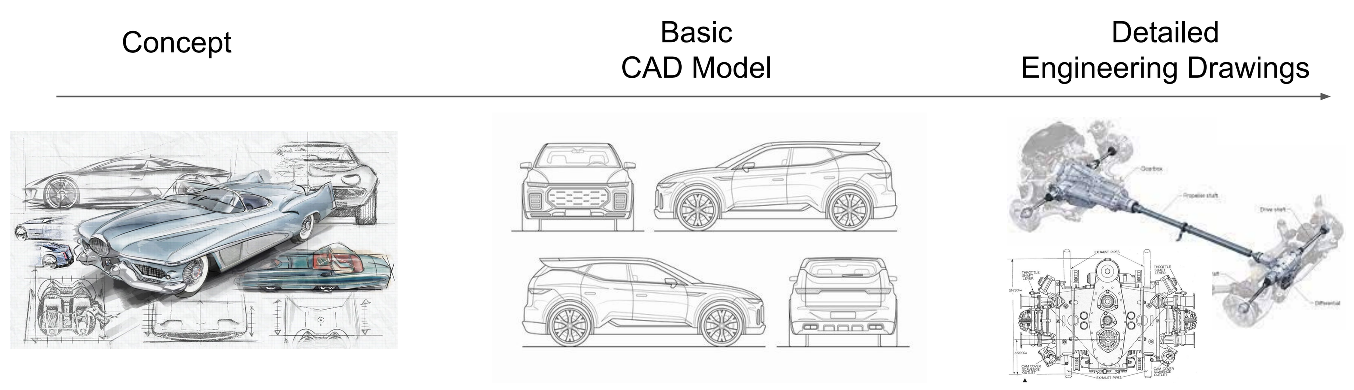
Drag-guided diffusion models for vehicle image generation

Nikos Arechiga[1], Frank Permenter[1], Binyang Song[2], Chenyang Yuan[1]

[1]Toyota Research Institute [2]MIT

Introduction and motivation

We want to integrate Generative AI tools in all stages of design



Can use AI tools today!

Need to incorporate complex engineering design constraints

Existing GAI tools understand artistic guidance… "A car that is sleek and sporty" →

…but not *quantitative metrics* and *constraints* essential to engineering. "A car with a drag coefficient of 0.5" → ???

Our contribution

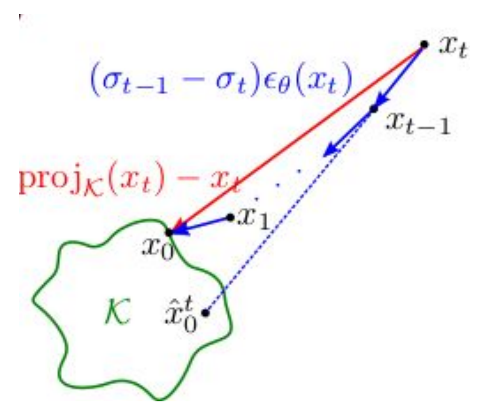
We design and implement a diffusion guidance technique that simultaneously estimates and optimizes an engineering metric.

Diffusion training: $L(\theta) := \mathop{\mathbb{E}}_{x,\sigma,\epsilon} \|\epsilon_\theta(x+\sigma\epsilon,\sigma)-\epsilon\|^2$

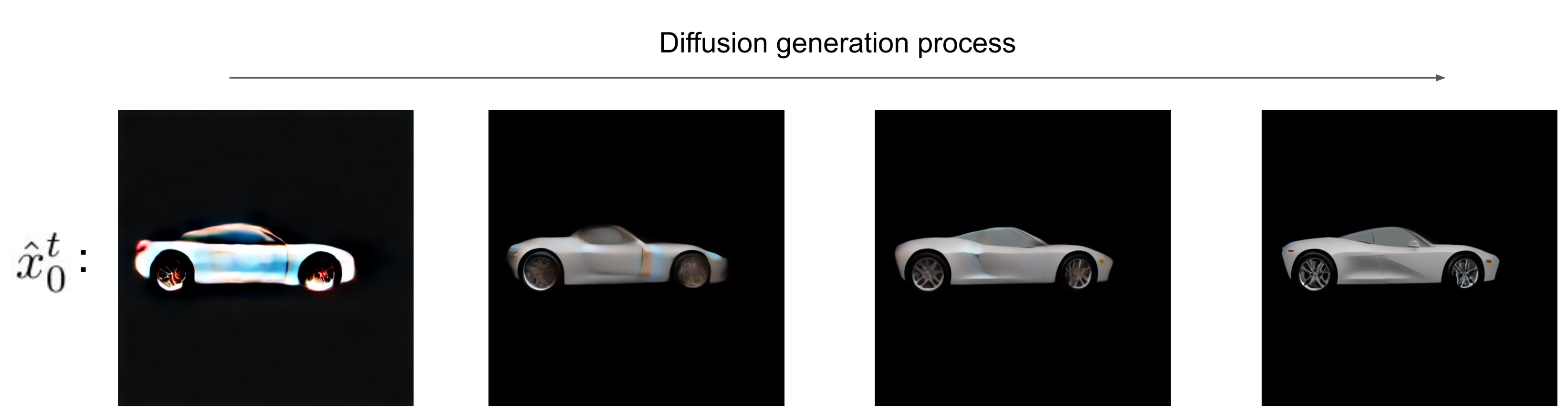
DDIM sampling update: $x_{t-1} = x_t - (\sigma_t - \sigma_{t-1})\epsilon_\theta(x_t, t).$

Diffusion approximates a projection onto data manifold at each step (Permenter and Yuan 2023)

$$\hat{x}_0^t = x_t - \sigma_t \epsilon_\theta(x_t, t)$$



We demonstrate a proof-of-concept on drag optimization of vehicles:



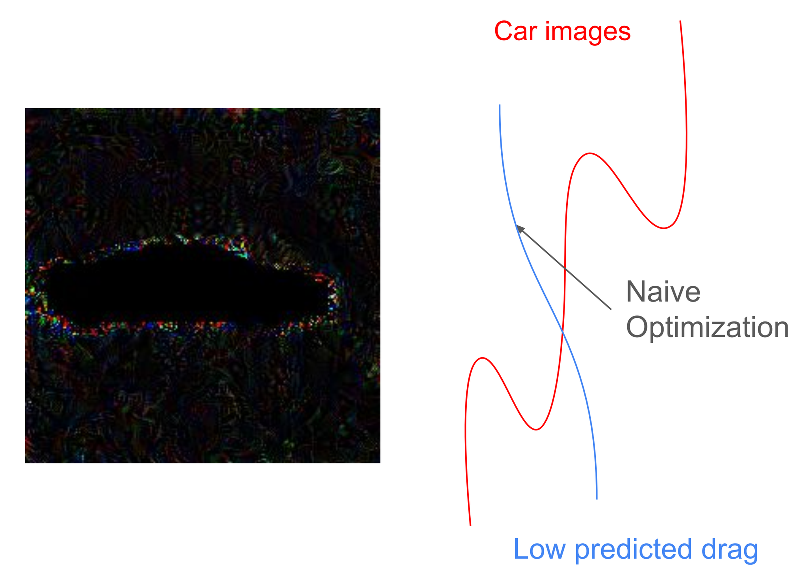
Drag minimization update:

$$x_{t-1} = x_t - (\sigma_t - \sigma_{t-1})(\epsilon_\theta(x_t,t) + \eta_t \nabla\phi_{\text{drag}}(\hat{x}_0^t))$$

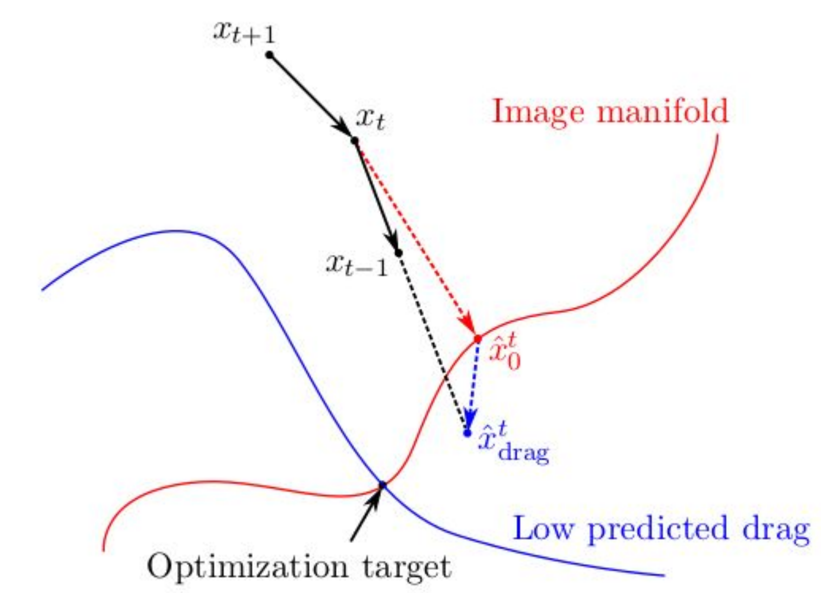
Our technique uses a surrogate model of the engineering metric Φ(x) that:

1. Takes as input an image produced by a diffusion process
2. Is robust to distributional shifts in the input
3. Is differentiable, allowing for gradient-based optimization of the output

Why our method works



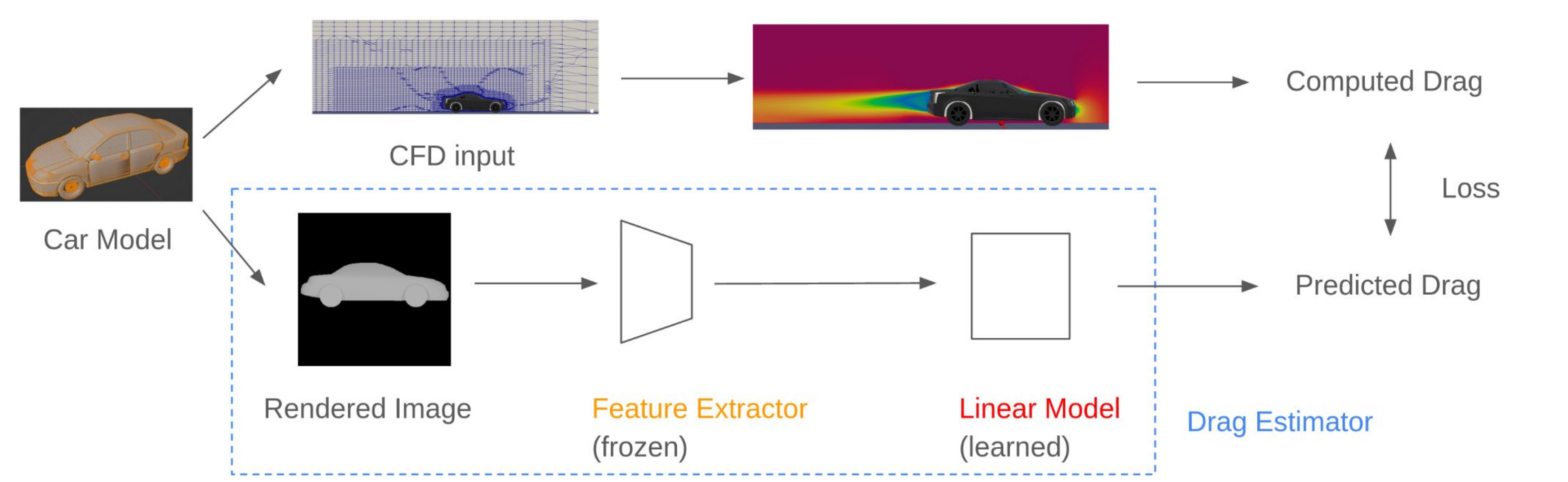
A naive implementation that only optimize the output of the drag surrogate model would produce noise, finding an adversarial example for the surrogate model.



Our approach alternate steps of optimizing the estimated drag and projecting to the data manifold, preventing the generation from devolving into noise.

Training the surrogate drag model

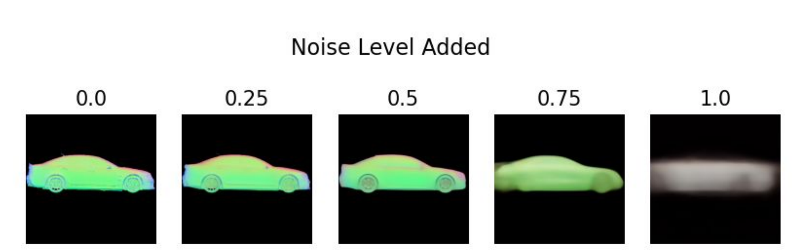
Our drag-guided diffusion pipeline relies on a differentiable drag estimator

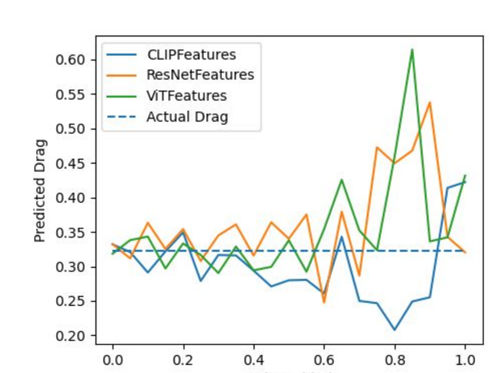


We use pretrained vision models (CLIP, ViT, ResNet) as feature extractors, then fine tune a linear layer using ~9k (car model, drag coefficient) dataset in (Song et. al. 2023)

Experiments and results: surrogate model

Drag predictions on out-of-distribution images

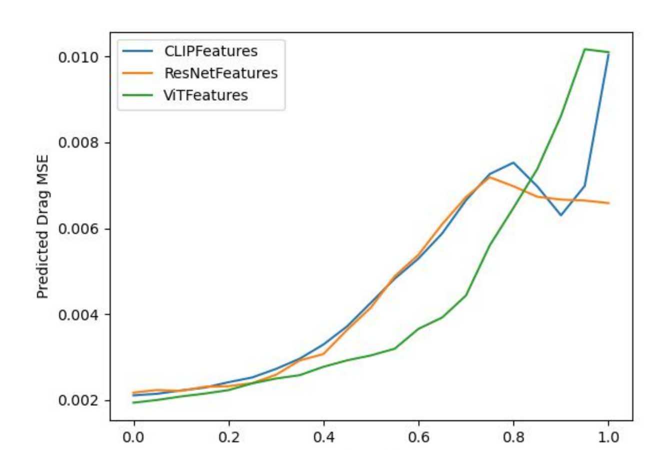




Accuracy of learned drag estimator

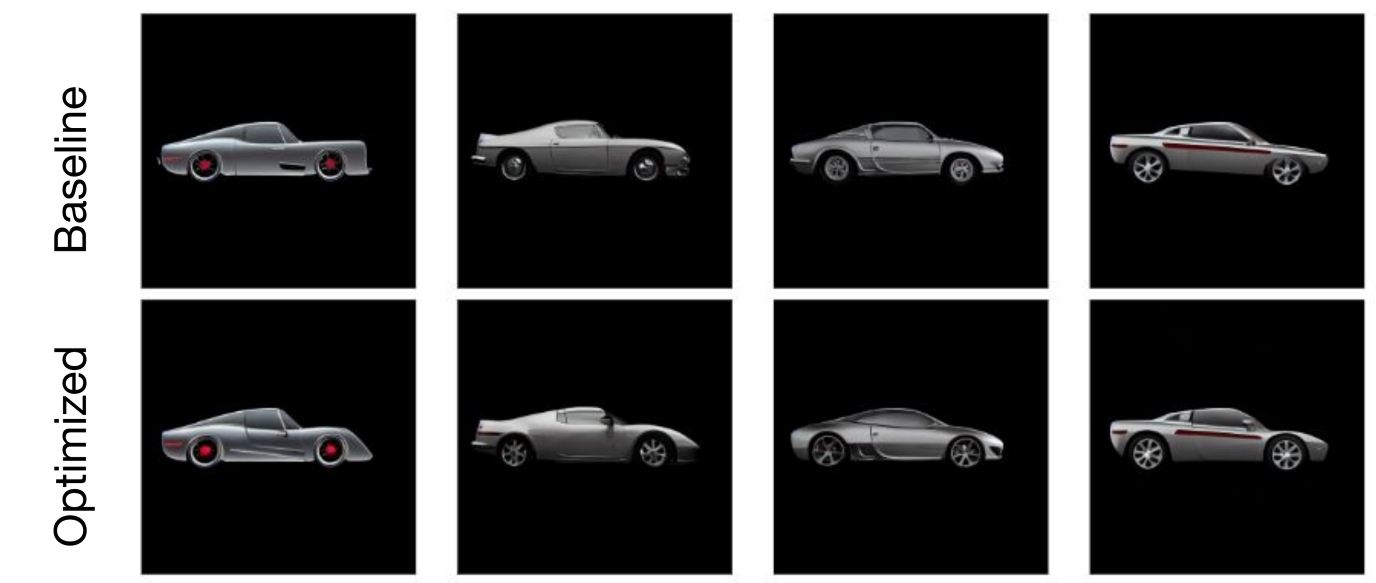
Features	Dimension	Train R^2	Test R^2	Test MSE
CLIP	512	0.639	0.545	0.00251
ViT	768	0.639	0.549	0.00249
ResNet	2048	0.630	0.509	0.00271

Robustness of drag estimator to added noise

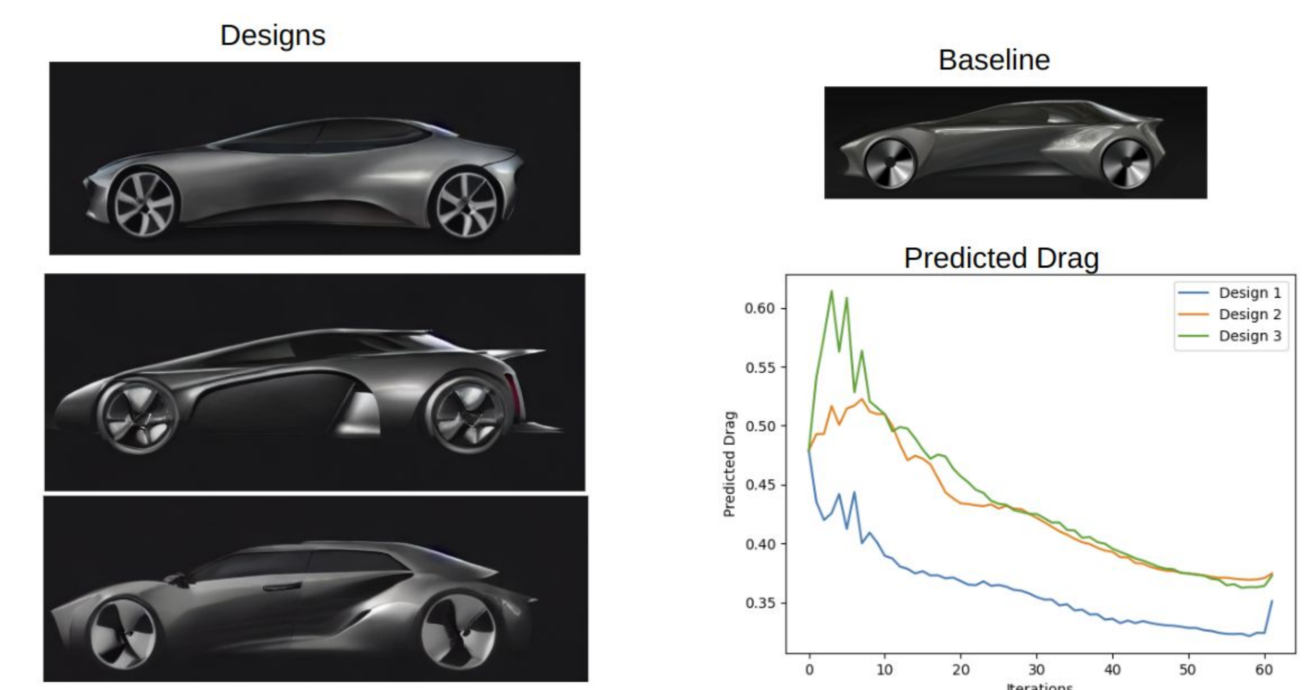


Experiments and results: drag guidance

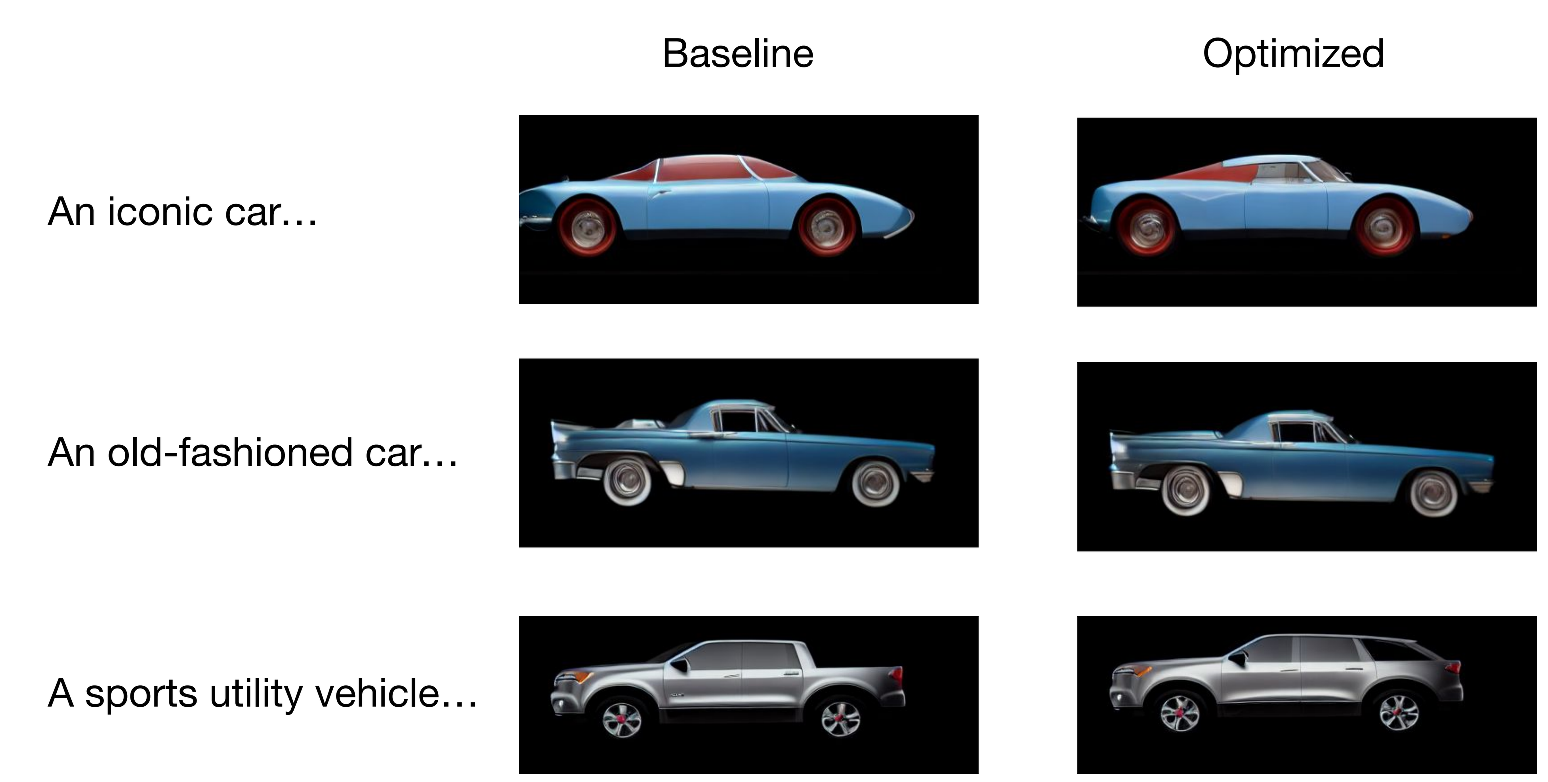
Our end-to-end pipeline is able to generate vehicles with low predicted drag



We can create variations of a baseline design while optimizing drag



We can combine different text prompts with drag guidance



References

F. Permenter and C. Yuan (2023). "Interpreting and improving diffusion models using the Euclidean distance function". Preprint: https://arxiv.org/abs/2306.04848

B. Song, C. Yuan, N. Arechiga, F. Permenter, and F. Ahmed (2023). "Surrogate modeling of car drag coefficient with depth and normal rendering". In: Proceedings of the ASME, IDETC/CIE

Link to our paper: https://arxiv.org/pdf/2306.09935.pdf