

Edicho: Consistent Image Editing in the Wild

Qingyan Bai^{1,2} Hao Ouyang² Yinghao Xu^{3,2} Qiuyu Wang²
Ceyuan Yang⁴ Ka Leong Cheng^{1,2} Yujun Shen² Qifeng Chen¹

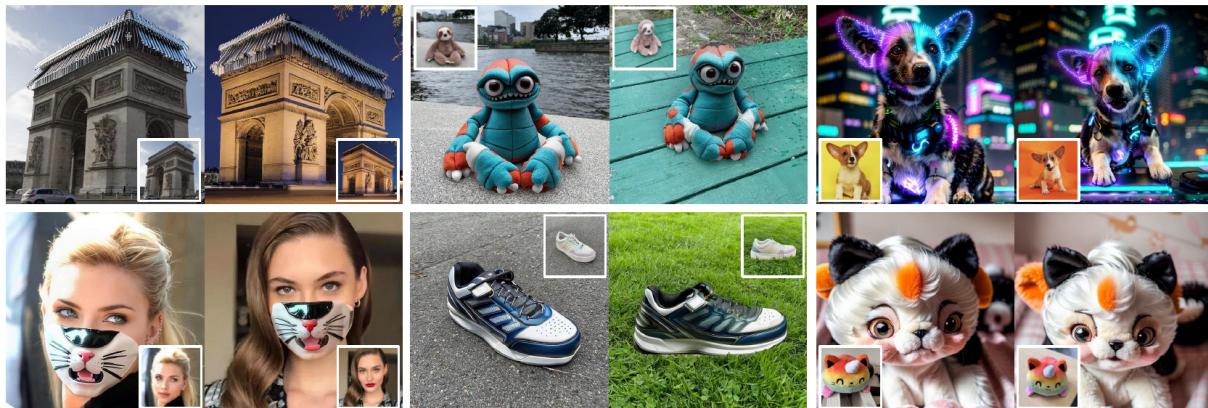
¹HKUST

²Ant Group

³Stanford University

⁴CUHK

Introduction



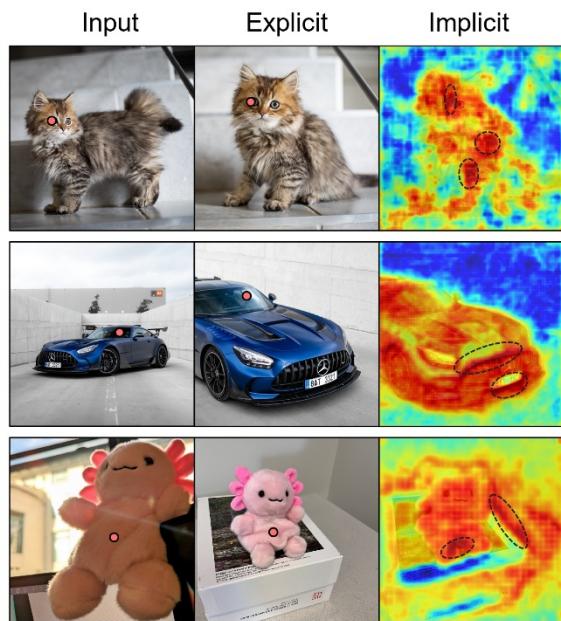
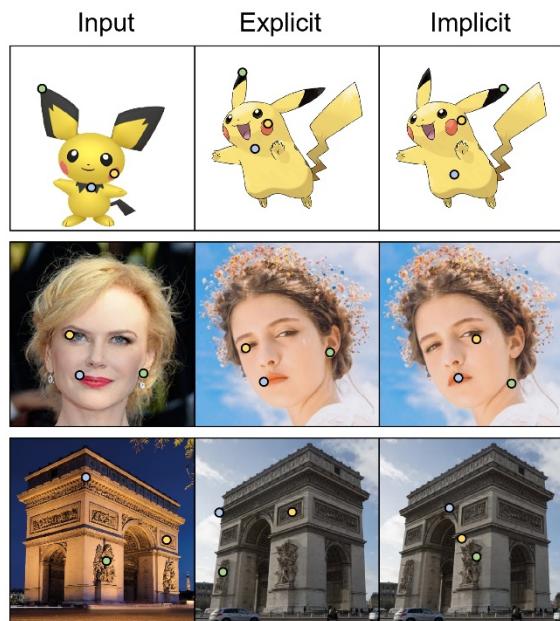
Paper



Page

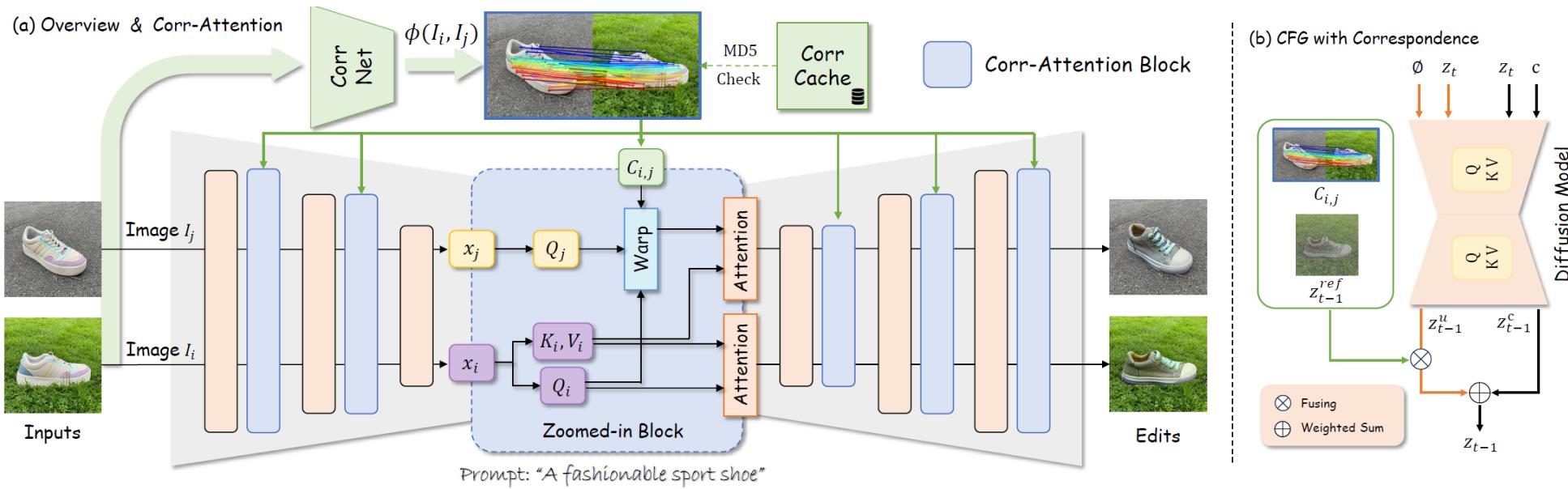
- Consistent editing across real-world images remains challenging due to variations in object pose, lighting, and background.
- We propose Edicho, a training-free, plug-and-play method that leverages explicit semantic correspondence to guide diffusion-based editing models.
- Our approach introduces two novel mechanisms: Correspondence-Guided Attention (Corr-Attention) and Correspondence-Guided Classifier-Free Guidance (Corr-CFG).

Motivation



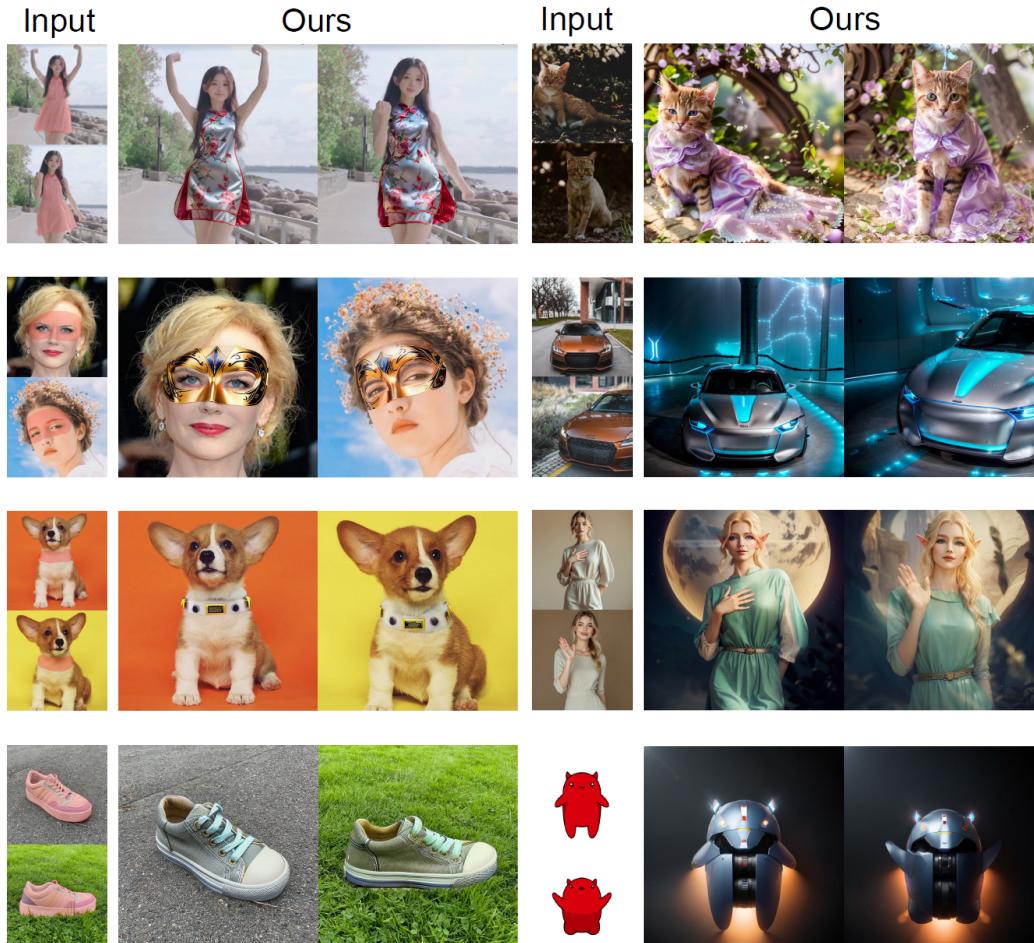
- In the task of in-the-wild image editing, learning-based methods often lack proper regularization, resulting in inconsistent edits due to the difficulty of obtaining high-quality training data and enforcing uniformity constraints.
- Non-optimization methods rely on implicit correspondence from attention features for appearance transfer, but struggle with unstable predictions and intrinsic image variations, leading to inconsistent or distorted edits.
- We visualize the correspondence predicted respectively by explicit and attention-based implicit methods in the left figure, accompanied by the attention maps for correspondence prediction.

Method



- Our method is built upon a pre-trained single-image editing model (e.g., BrushNet) and explicit correspondence extractor.
- We modify the self-attention mechanism to incorporate features from a source image based on explicit correspondence.
- We also align unconditional noise predictions between images using correspondence for more fine-grained consistency.
- This inference-time algorithm enjoys a plug-and-play nature and is compatible to most diffusion-based editing methods.

Results



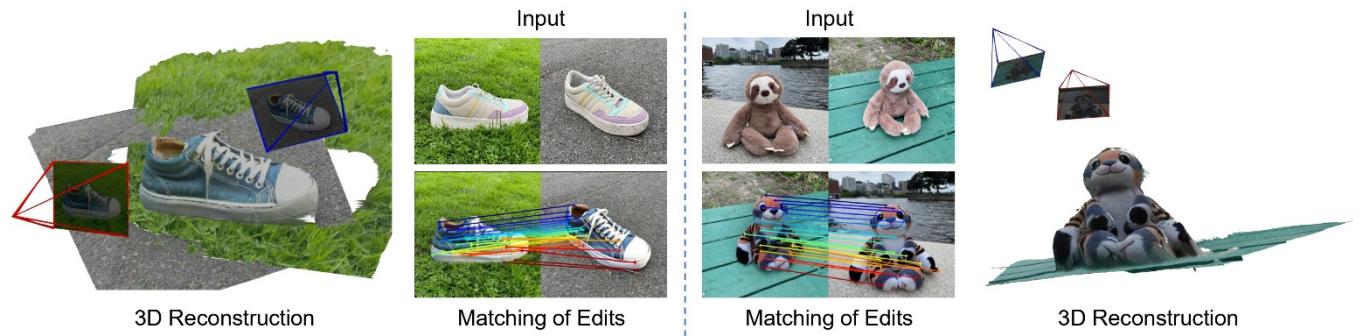
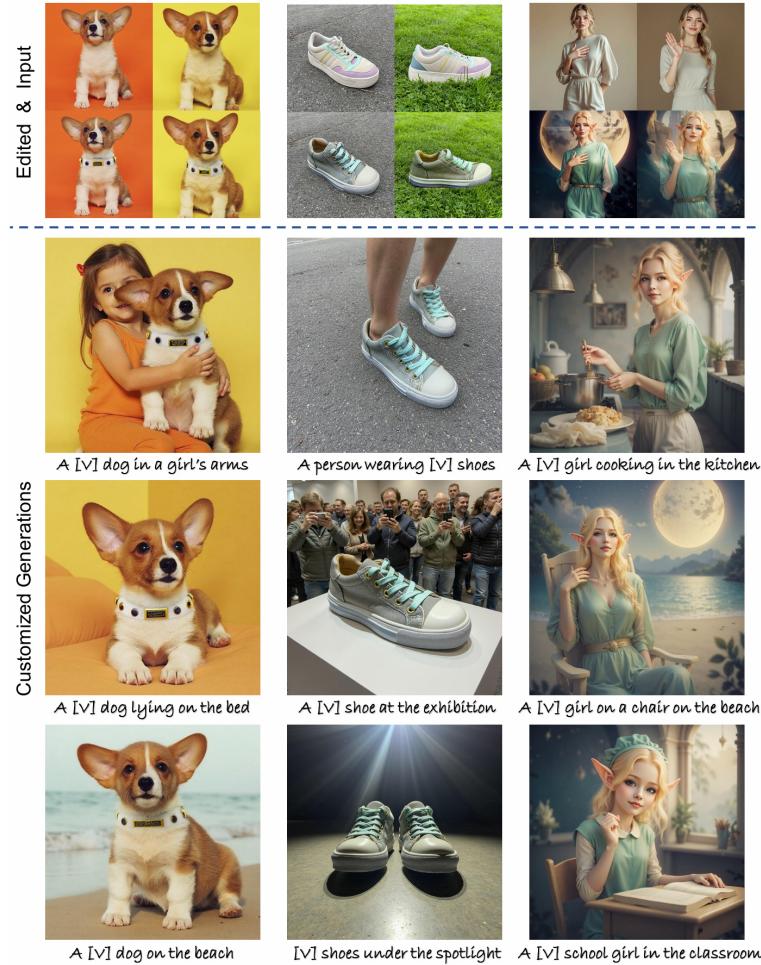
- We provide results by Edicho for consistent image editing in the left figure.
- Our framework works for various editing tasks including local and global editing.

Ablation Studies



- Removing either the Corr-Attention or Corr-CFG would drop the consistency of the editing results, validating the importance of our proposed correspondence-guided mechanisms.

Applications



We showcase two practical applications:

- Customized generation via fine-tuning with consistently edited images, enabling personalized concept injection;
- 3D matching and reconstruction by adopting the neural regressor based on the edits by matching the 2D points.

Summary

- We propose Edicho introducing explicit correspondence into the diffusion denoising process and achieving the consistent image editing without retraining. It is a versatile and plug-and-play solution for various editing tasks.
- Future directions may include integrating learned matchers to further improve correspondence accuracy and extending the method to video and 3D-aware generation to address more complex and diverse editing scenarios.

Thanks!
