# HIVE: Harnessing Human Feedback for Instructional Visual Editing

Shu Zhang*[1], Xinyi Yang*[1], Yihao Feng*[1], Can Qin[1], Chia-Chih Chen[1], Ning Yu[1], Zeyuan Chen[1], Huan Wang[1], Silvio Savarese[1,2], Stefano Ermon[2], Caiming Xiong[1], Ran Xu[1]

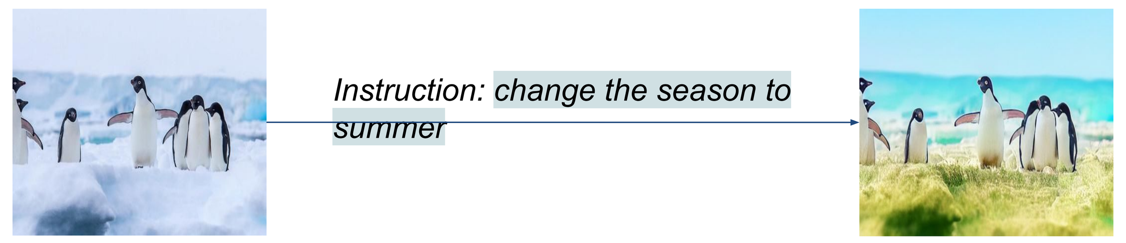
*Equal Contribution [1]Salesforce AI Research [2]Stanford University

CVPR SEATTLE, WA JUNE 17-21, 2024

Code

## Background

- Instructional image editing has emerged as one of the most promising application scenarios for content generation. We hypothesize that **instructional image editing could benefit from human feedback**, as their outputs may not adhere to the correct instructions and preferences of users.



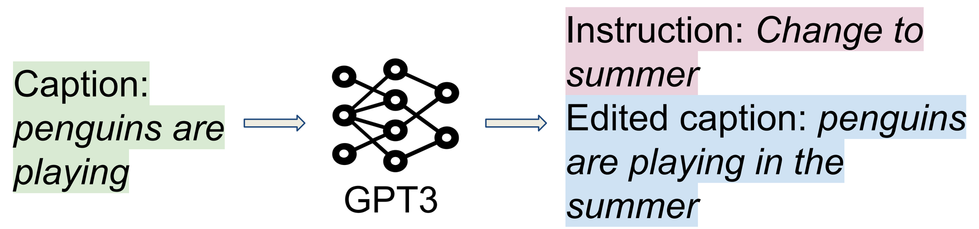

## Contribution

- To tackle the technical challenge of fine-tuning diffusion models using human feedback, we introduce **two scalable fine-tuning approaches**, which are computationally efficient and offer similar costs compared with supervised fine-tuning. Moreover, we empirically show that **human feedback is an essential component** to boost the performance of instructional image editing models.
- We create a new dataset for HIVE including three subdatasets: a new **1.1M training dataset,** a **3.6K reward dataset** for rewards learning, and a **1K evaluation dataset**.
- We introduce **cycle consistency** augmentation based on the inversion of editing instruction. Our dataset has been enriched with one pair of data for bi-directional editing.

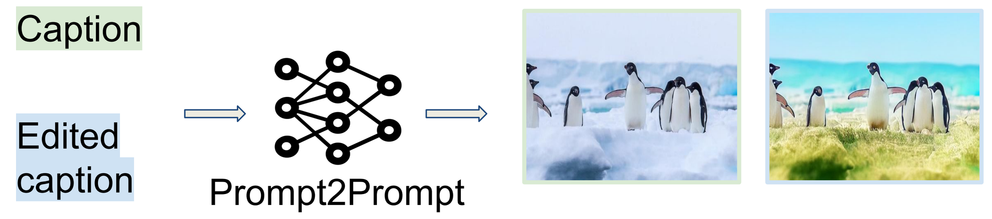
## Proposed Methods

### Step 1: instructional supervised training

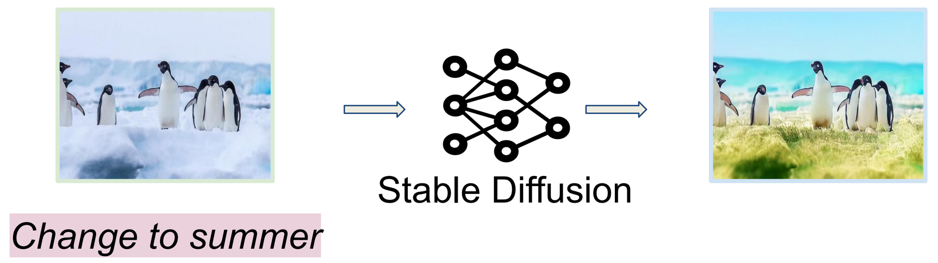
- Collect data to fine-tune GPT-3 and use fine-tuned GPT-3 to generate text edits




- Use Prompt-to-Prompt to generate paired images

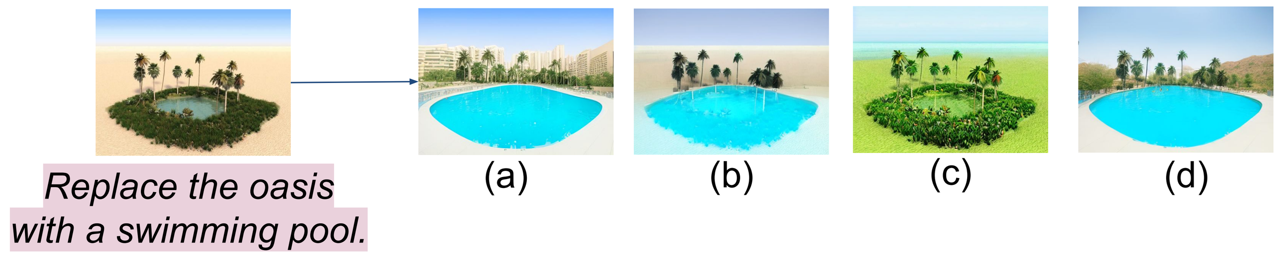



- Fine-tune stable diffusion with the paired images and instructions

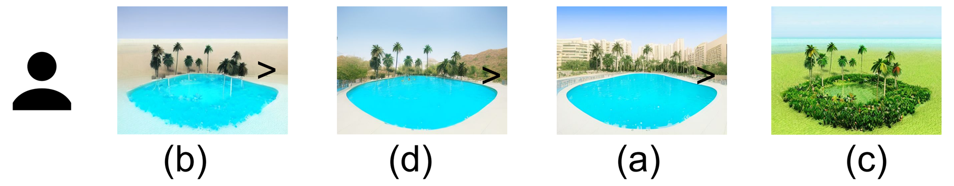



### Step 2: collect comparison data, and train a reward model

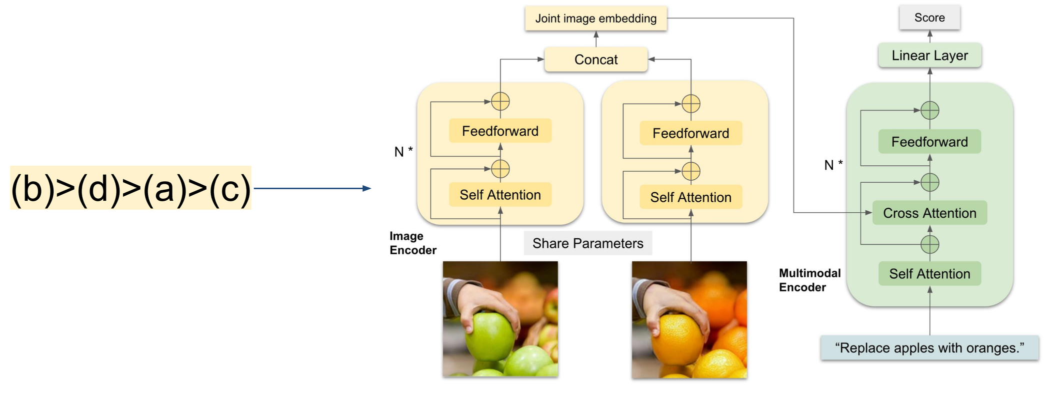
- Collect a reward dataset and generate sampled outputs from Step 1




- Let annotators rank outputs from best to worst




- Train a reward model on the reward dataset




BLIP based RM

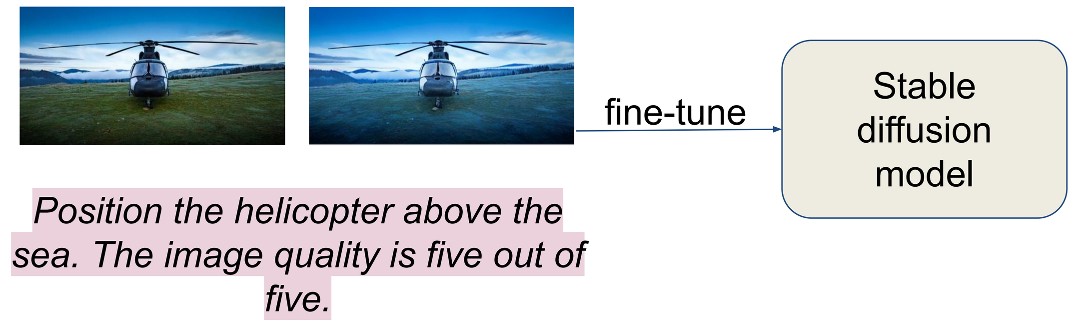
### Step 3: fine-tune diffusion model with learned rewards

- Use the learned reward model to calculate reward values for each training pair

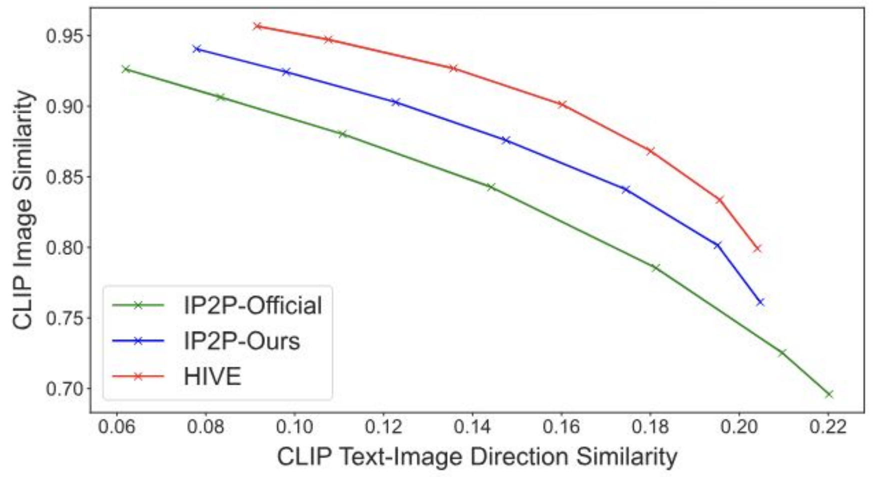

- Convert reward value to text prompt and use it as a condition to fine-tune diffusion model

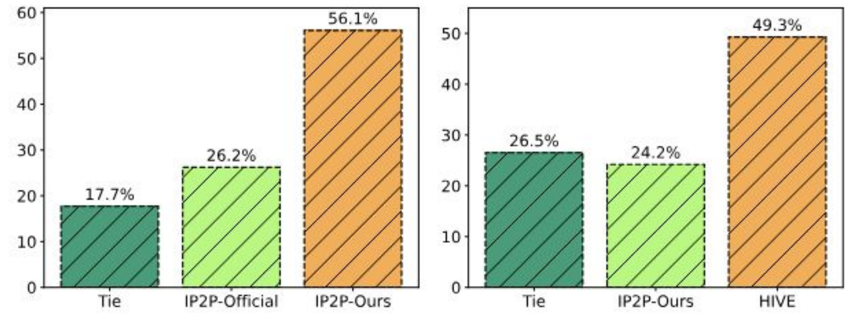
(1.3, 0.5, …) → (5, 1, ...) → <original instruction> +The image quality is <reward label> out of five.

rewards values | rewards labels | modified instruction




## Quantitative Results

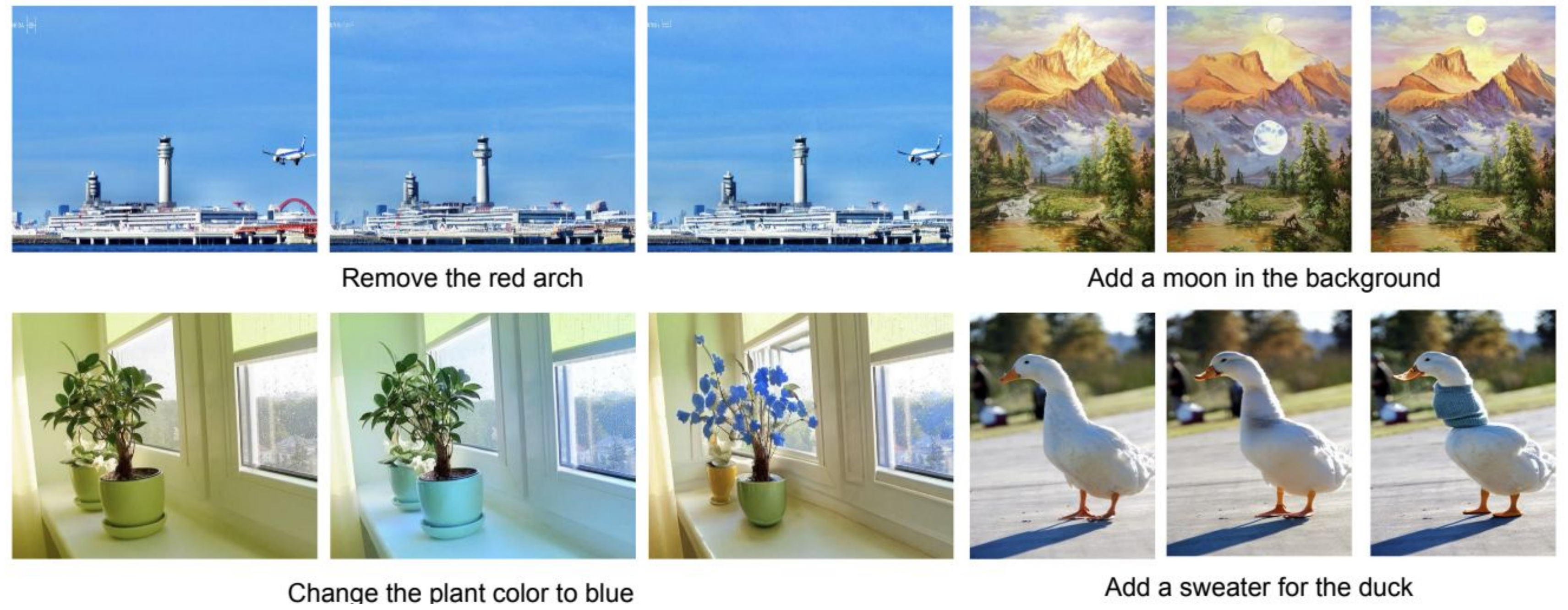


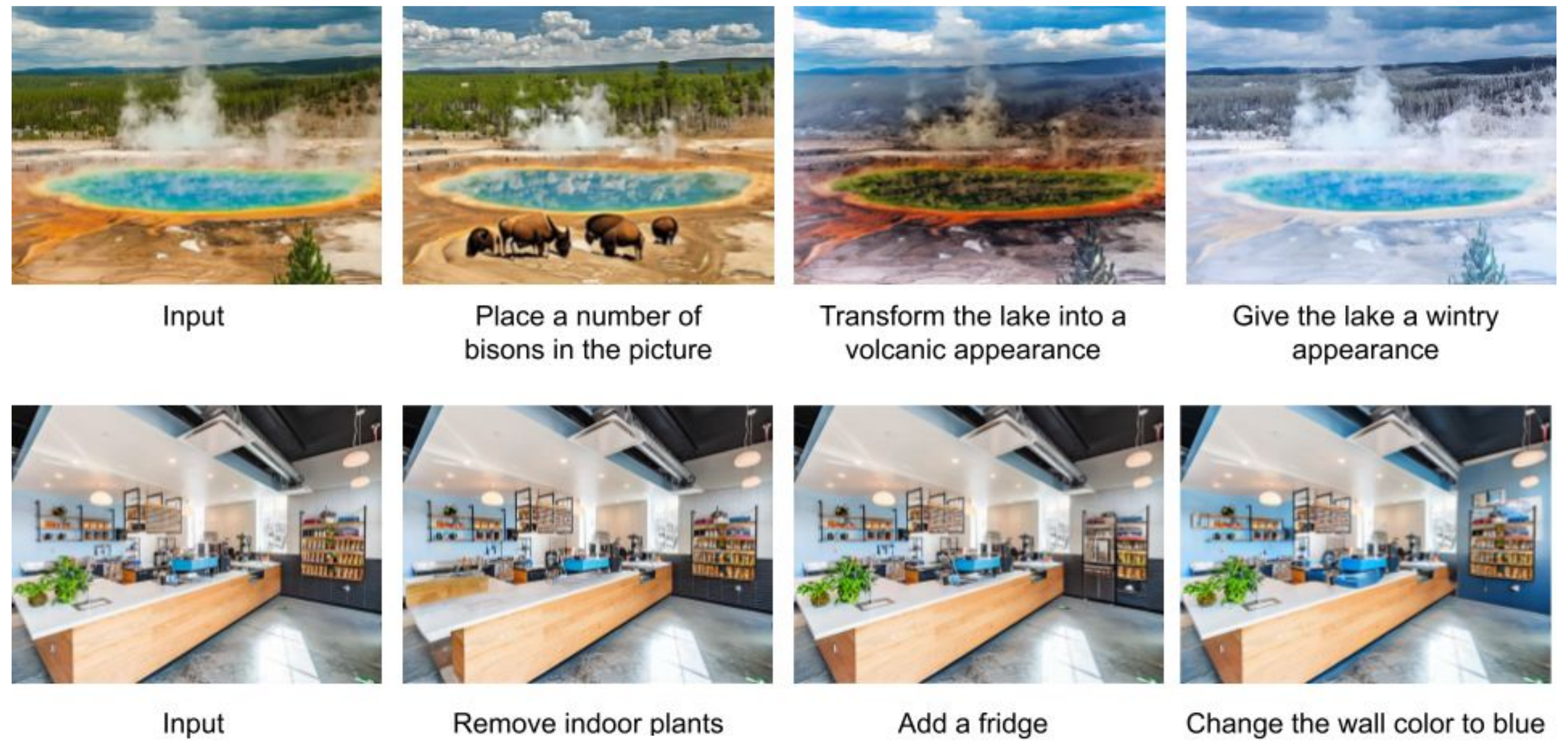

IP2P-Official vs IP2P-Ours IP2P-Ours vs HIVE

- Comparisons between InstructPix2Pix(IP2P) and HIVE. Illustration of tradeoffs between consistency with the input image and with the edit. **HIVE achieves higher similarity on both metrics.**
- User study of comparison between InstructPix2Pix(IP2P) and HIVE. **HIVE obtains 25% more votes.**

## Qualitative Results




- From left to right: **Input image, HIVE without human feedback, HIVE with human feedback.**




- More examples of **HIVE with human feedback.**