

RealCustom: Narrowing Real Text Word for Real-Time Open-Domain Text-to-Image Customization

What is the problem?

The primary problem addressed by the RealCustom paper is the challenge in achieving high-quality text-to-image customization, where both the similarity to the given subject and the controllability of the generated text must be optimized. Existing approaches, which rely on a pseudo-word paradigm to represent subjects, struggle to balance these two aspects simultaneously. This creates a "dual-optimum paradox," where improving the resemblance to the subject compromises the control over the text, and vice versa. Moreover, real-time generation without additional test-time optimization steps is a major obstacle, preventing efficient content creation.

What has been done earlier?

Previous work in text-to-image synthesis typically uses a pseudo-word to represent the given subject, such as a portrait or object, and then combines it with text descriptions. While this method provides a way to generate customized images, it entangles the influence of the subject with the influence of the text, leading to suboptimal results. The similarity of the generated image to the original subject is often compromised, and the flexibility of the text to control other elements is limited. Additionally, these methods are not designed for real-time performance and require test-time optimization, which is time-consuming and inefficient for practical use.

put the elephant in the refrigerator



Amisha Nayak, B421004

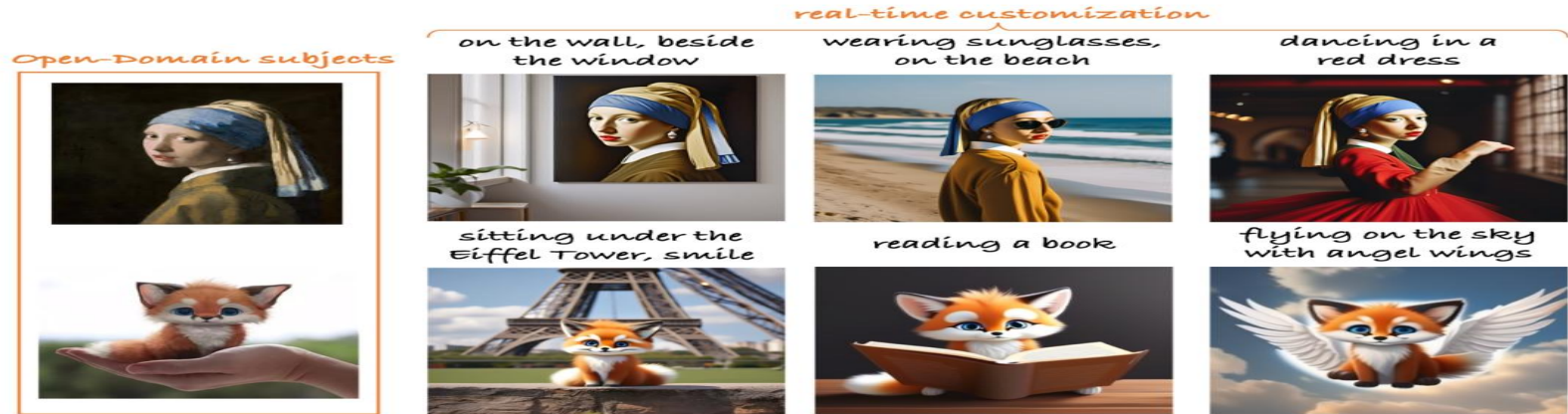
What are the remaining challenges? What novel solution proposed by the authors to solve the problem?

Remaining Challenges:

- **Balancing similarity and controllability:** Existing methods cannot optimize both at the same time, leading to low-quality images.
- **Influence control:** Difficulty in precisely limiting the subject's influence to the relevant parts of the image.
- **Real-time customization:** Current models often require test-time optimization, making them inefficient for real-time usage. are physically plausible.

Novel Solution Proposed by the Authors:

The authors propose RealCustom, a solution that disentangles the influence of the subject from the text. It uses an adaptive mask guidance strategy to precisely control the subject's impact on the image, while maintaining control over the text. RealCustom introduces a "train-inference decoupled framework," where training aligns visual and text inputs, and inference updates subject influence in real-time. This approach improves both subject similarity and text controllability without needing extra optimization steps.



Given a **single** image representing the given subject in the **open domain** (any subjects, portrait painting, favorite toys, etc.), RealCustom could generate realistic images that consistently adhere to the given text for the given subjects in **real-time** (without any test-time optimization steps).