

CS445 Final Project Proposal

Motivation

We chose to focus on implementing the pix2pix-zero^[1] paper due to its novel approach to zero-shot image-to-image translation, leveraging pre-trained text-to-image diffusion models without the need for task-specific training or manual text prompting. Our interest lies in understanding and assessing the effectiveness of its unique mechanism for automatic editing direction discovery and cross-attention guidance for content preservation. Through this project, we hope to delve deep into the mechanics of pix2pix-zero, validating its claimed capacities and exploring its potential limitations. We also want to expand our understanding of diffusion models and their practical applications in real-world image editing.

Milestones

- Week1 (04/01 - 04/07):
 - Literature review and environment setup
- Week2 (04/08 - 04/14):
 - Implementing automatic editing direction discovery mechanisms
 - Implementing cross-attention guidance for content preservation
 - Perform initial image-to-image translation tasks as outlined in the paper
- Week3 (04/15 - 04/21):
 - Perform quantitative and qualitative evaluations
 - Ablation study
- Week4 (04/22 - 04/28):
 - Model acceleration with conditional GANs
 - Explore potential limitations
- Week5 (04/29 - 05/05):
 - Gradio demo
 - Wrap up final report and code
 - Final submission

Evaluation

- Quantitative evaluation
 - Edit similarity rate: measuring whether the edit was applied successfully by using CLIP Acc to assess alignment with target attributes.
 - Content preservation: measuring whether the structure of the input and edited image is consistent by using Structure Dist.
 - Background preservation: measuring if the background regions of the image stay unchanged after edits by calculating the background LPIPS error.
- Qualitative evaluation

- Create a diverse set of test images, including both real and synthetic images, to evaluate the model's robustness and versatility across various domains and editing tasks.
- Ablation study
 - Perform ablation studies to understand the contribution of key components of the methods, such as the automatic discovery of editing directions and cross-attention guidance.

Resources

- Data
 - Images can be retrieved from the [LAION 5B dataset](#), which is mentioned in Appendix D of the pix2pix-zero paper.
- Computation
 - GPU compute environment can be accessed via Google Colab.
- GitHub repo
 - Project's source code: <https://github.com/pix2pixzero/pix2pix-zero>
 - Stable diffusion model: <https://github.com/CompVis/stable-diffusion>
 - BLIP: <https://github.com/salesforce/LAVIS?tab=readme-ov-file>

Group Contribution

We plan on assigning 1-2 members to each milestone. The specific assignments will depend on members' background, interest, and expertise. This will be determined after performing the Literature Review milestone in Week 1.

| | |
|-----------------|-----|
| Xiaoran Du | TBD |
| Simon Liu | TBD |
| Mohul Varma | TBD |
| Pascal Adhikary | TBD |

Reference

1. Parmar, Gaurav, et al. "Zero-shot image-to-image translation." ACM SIGGRAPH 2023 Conference Proceedings. 2023.