Dynamic Prompt Learning: Addressing Cross-Attention Leakage for Text-Based Image Editing

Kai Wang[1], Fei Yang[1*], Shiqi Yang[1],
Muhammad Atif Butt[1], Joost van de Weijer [1,2]

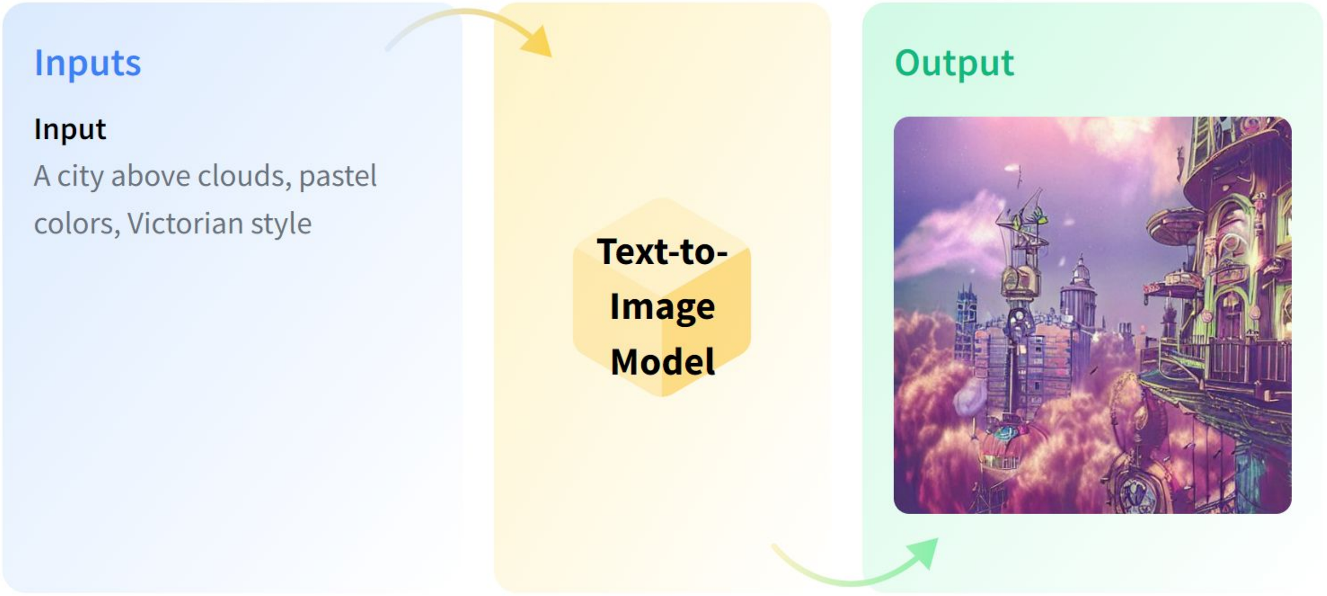
[1] Computer Vision Center, Barcelona, Spain.
[2] Universitat Autònoma de Barcelona, Barcelona, Spain.

kwang, fyang, syang, mabutt, joost @ cvc.uab.es

Learning and Machine Perception Team (LAMP)

Text-to-Image (T2I)

advancing at revolutionary pace



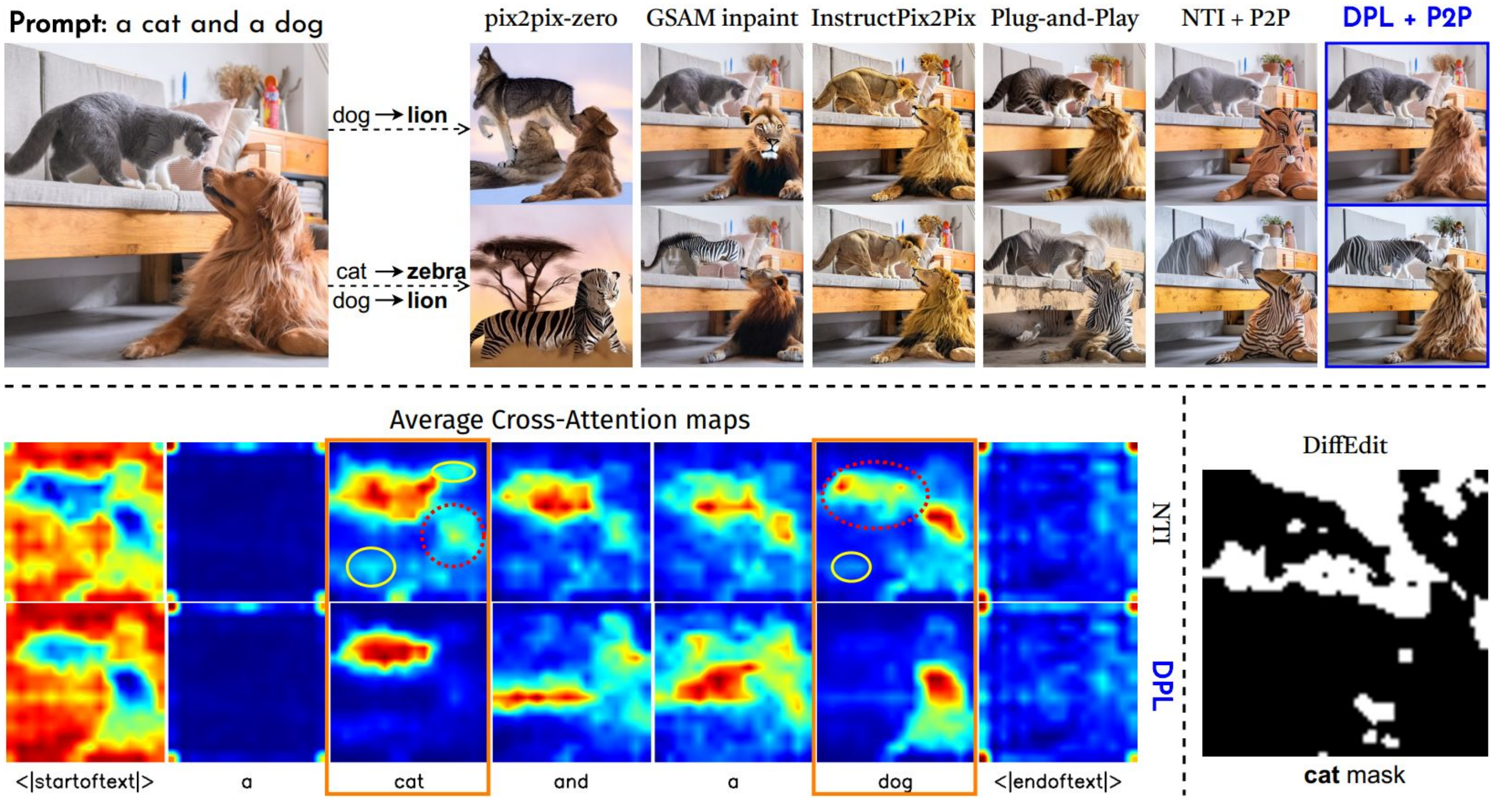
DALL-E, Imagen, Stable Diffusion, Mid-Journey, and many others ...

Current Problems

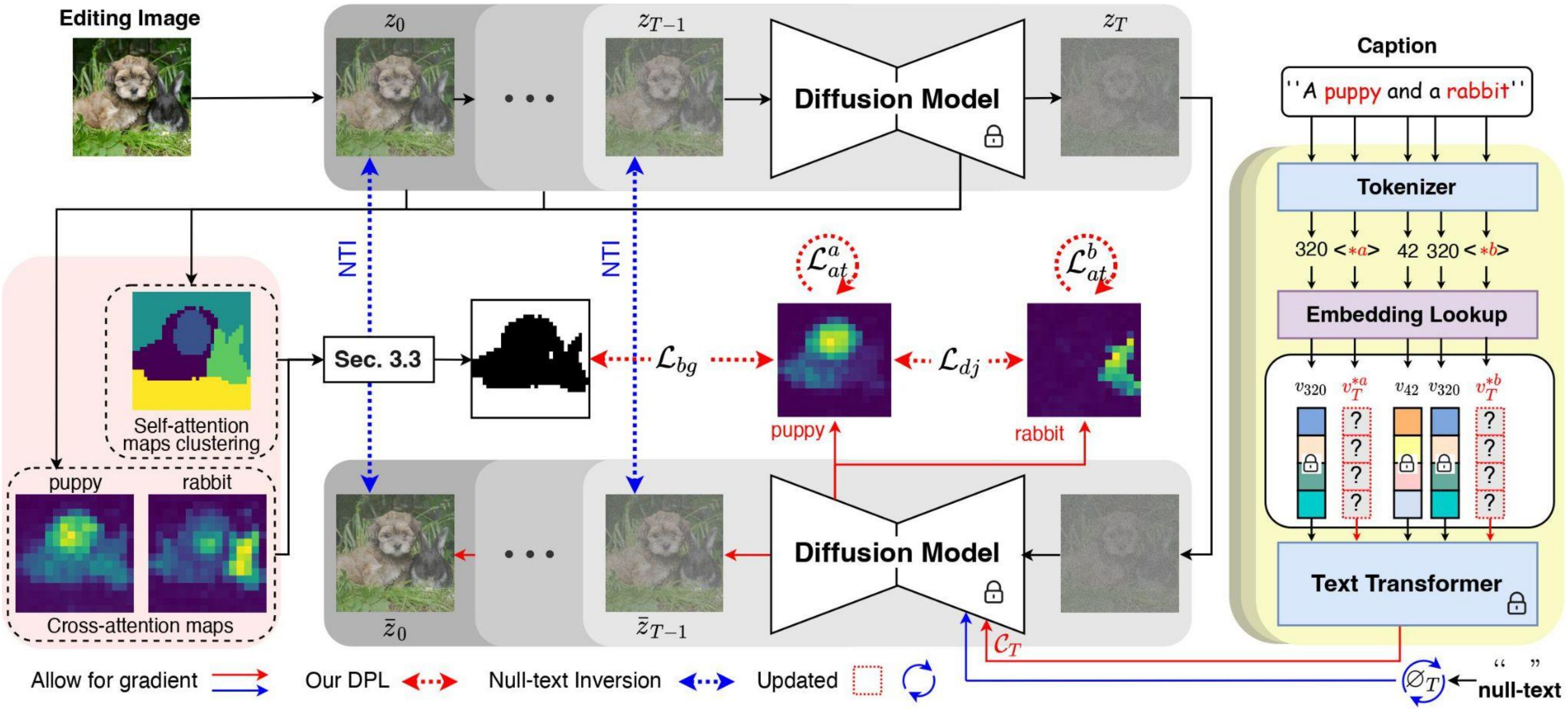
- Deep Generative Learning
  - Requires huge computational resources to train
  - Does not have image editing ability
- Text-guided Image Editing
  - Allows users to easily manipulate an image using text prompts
  - Not able to deal with complex scenarios

Limitation in Existing Text-guided Image Editing

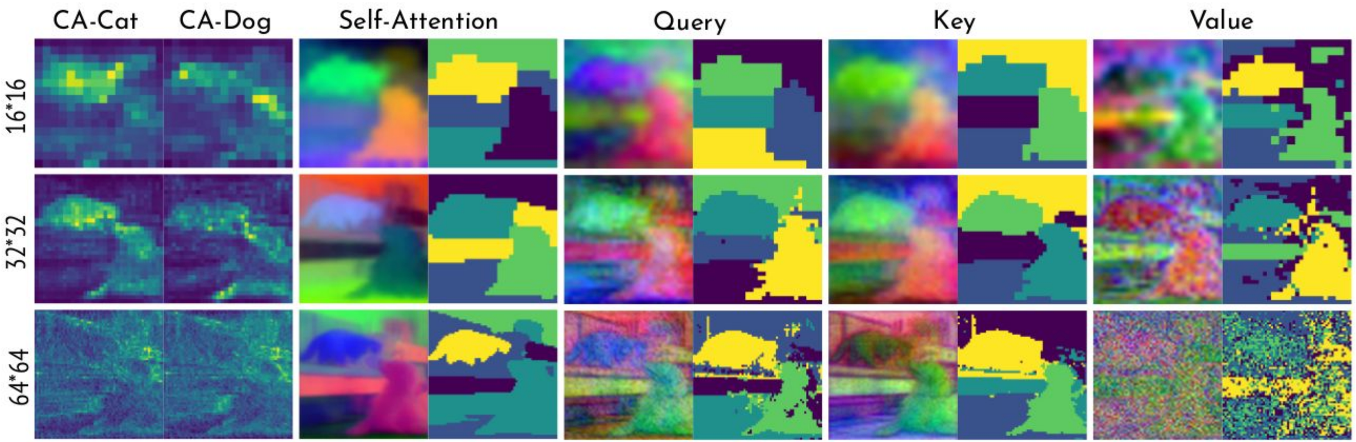
Lacks the capability to control specific regions of given image

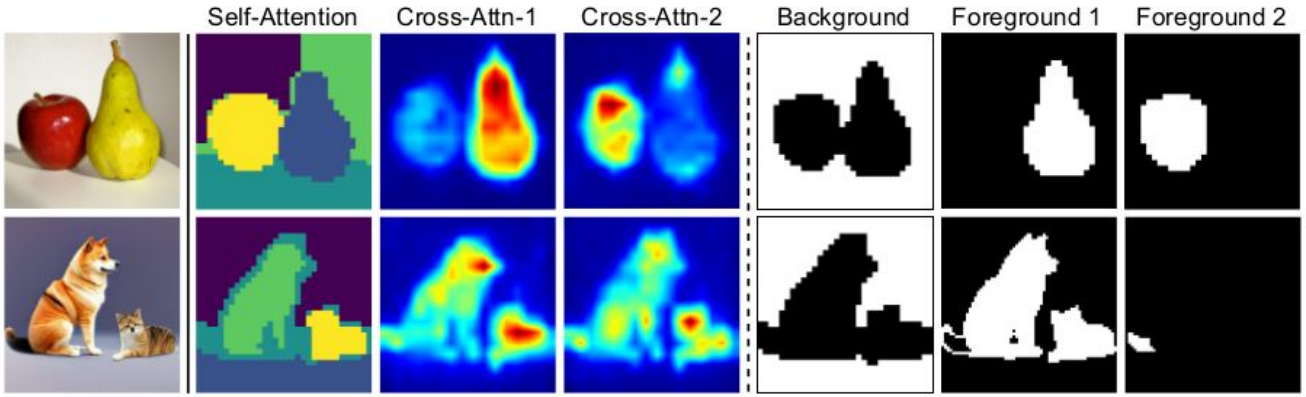


Dynamic Prompt Learning (DPL)



Feature visualization and background estimation





Addressing Cross Attention Leakage Loss Functions

- Disjoint Object Attention Loss

$$\mathcal{L}_{dj} = \sum_{i=1}^{K} \sum_{\substack{j=1 \\ i \neq j}}^{K} \cos(\mathcal{A}_t^{v_t^i}, \mathcal{A}_t^{v_t^j})$$

- Background Leakage Loss

$$\mathcal{L}_{bg} = \sum_{i=1}^{K} \cos(\mathcal{A}_t^{v_t^i}, \mathcal{B})$$

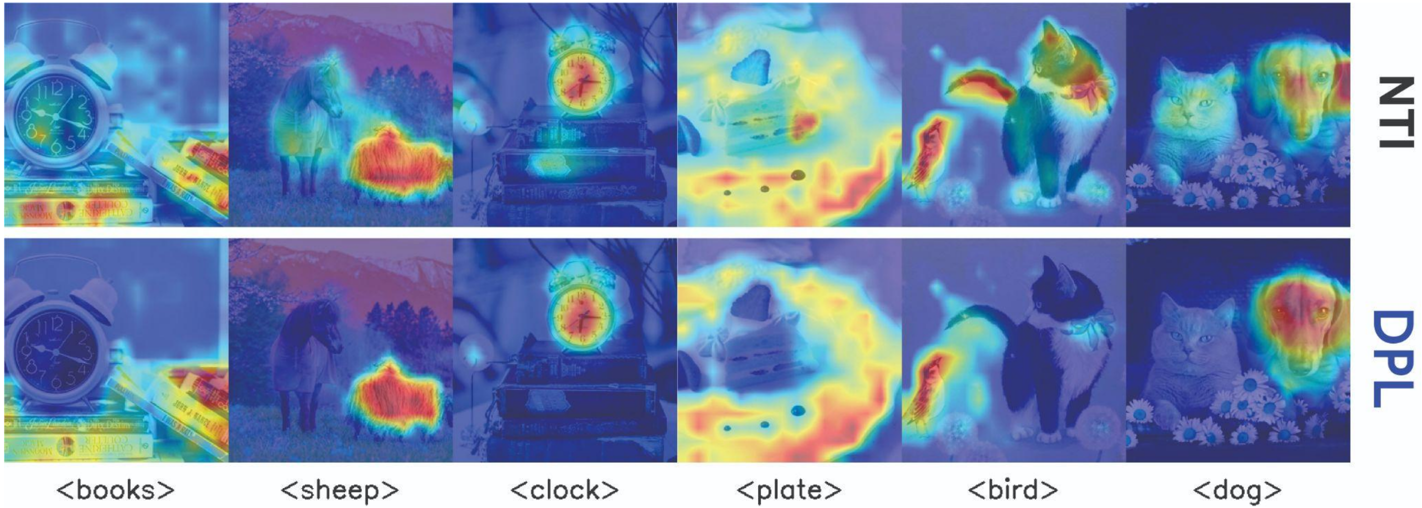
- Attention Balancing Loss

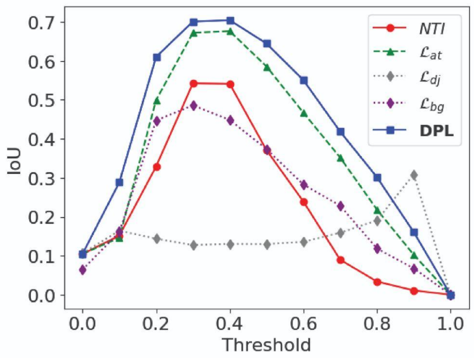
$$\mathcal{L}_{at} = \max_{v_t^k \in \mathcal{V}_t} \mathcal{L}_{v_t^k} \quad \text{where} \quad \mathcal{L}_{v_t^k} = 1 - \max[\mathcal{F}(\mathcal{A}_t^{v_t^k})]$$

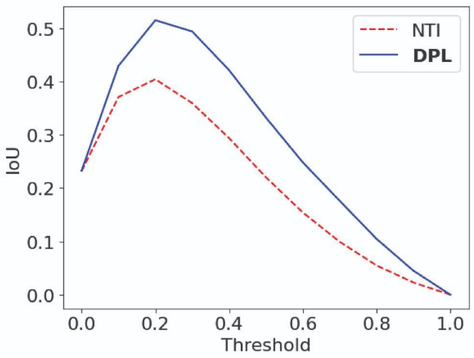
- Updating all token

$$\arg\min_{\mathcal{V}_t} \mathcal{L} \quad \text{where} \quad \mathcal{L} = \lambda_{at} \cdot \mathcal{L}_{at} + \lambda_{dj} \cdot \mathcal{L}_{dj} + \lambda_{bg} \cdot \mathcal{L}_{bg}$$

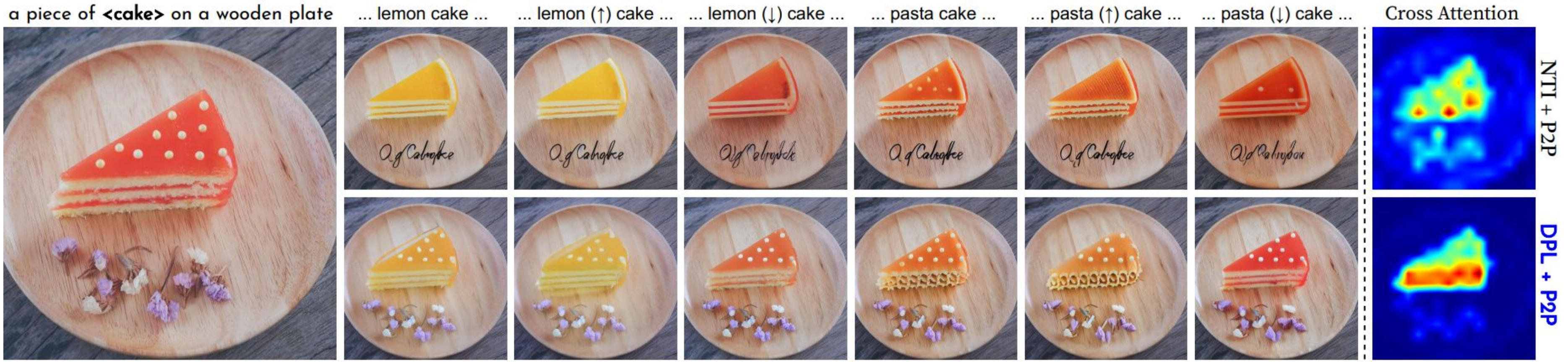
Cross-attention Maps Comparison



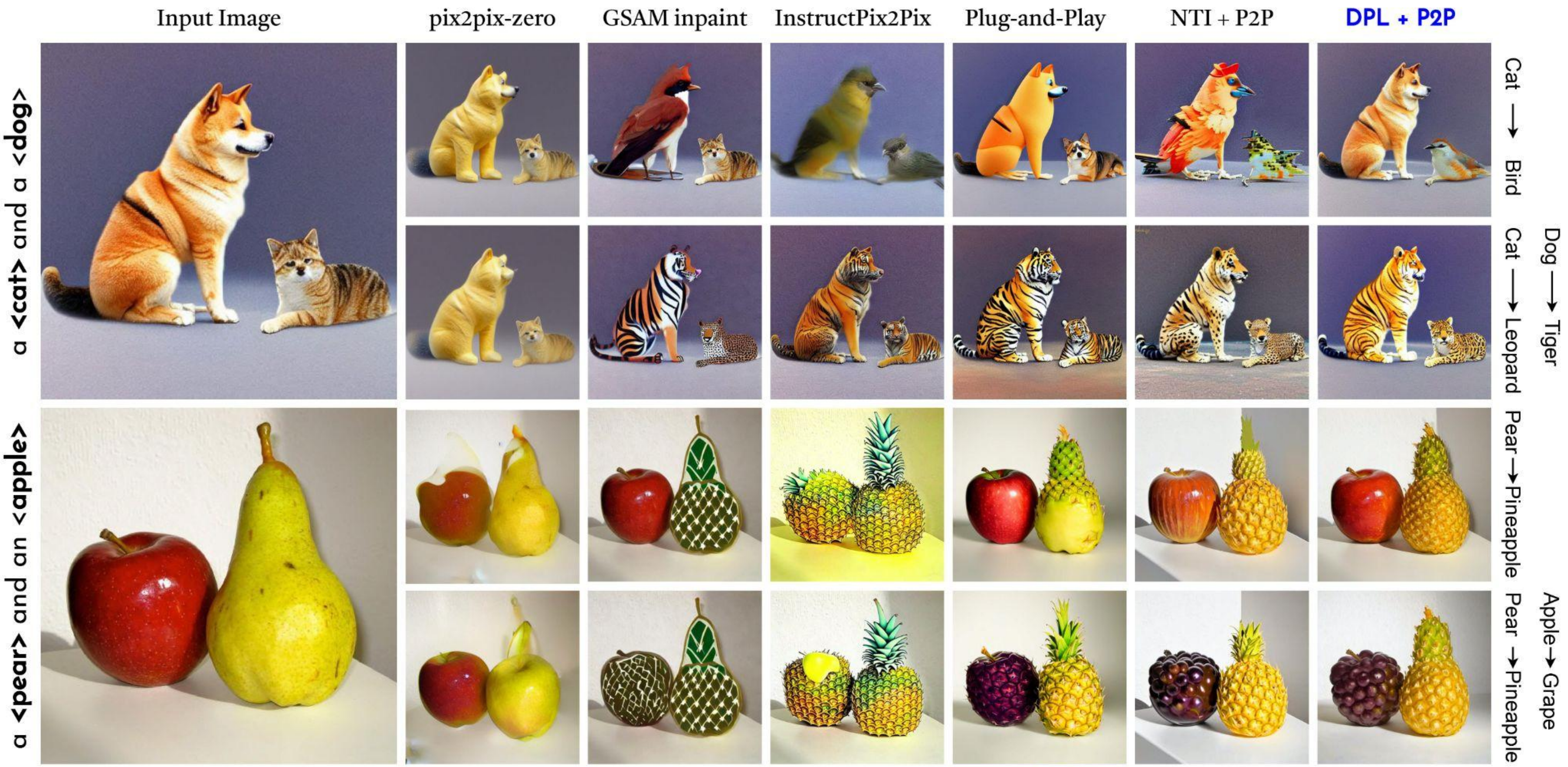




Attention Refinement and Reweighting

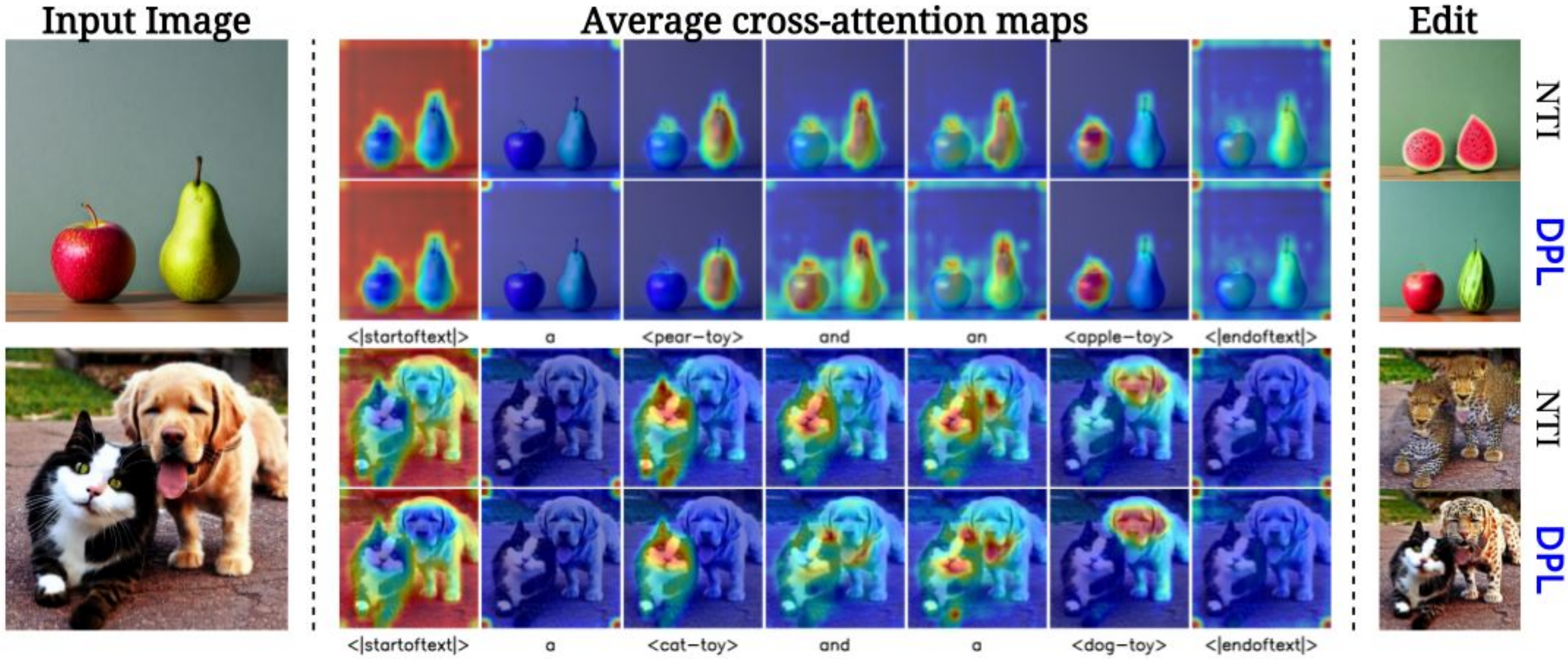


Some Qualitative Results



Some Qualitative Results

Some Qualitative Results (Generated Images)



Thanks for your attention!

Code, Arxiv, Neurips 2023