ARIA

APPLIED RESEARCH IN ACTION

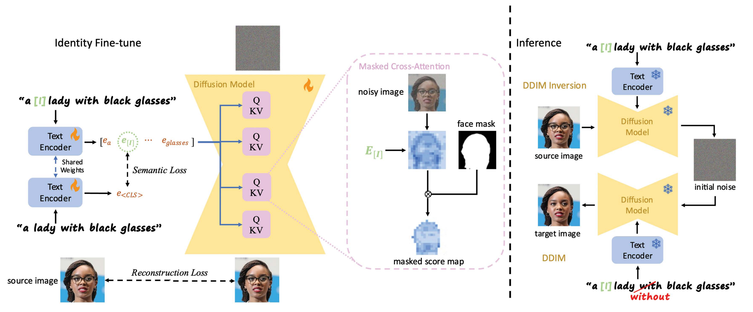
Detailed Cross-Attention Control in Image Editing

Refining Image Editing: Achieve Personalized, Pinpoint Adjustments While Preserving Identity and Quality.

Xudong Liu

Igor Gilitschenski
ACADEMIC SUPERVISOR

Shelly Chen
INDUSTRY SUPERVISOR



PROJECT SUMMARY

Recent advances in Diffusion Models enable high-quality image synthesis from text prompts and visual concepts learning from a given set of images. However, naively applying existing frameworks to editing tasks that require precise control of local details such as face editing often results in failure cases with deteriorated image quality, e.g., loss of subtle identity information and high-frequency details. Moreover, irrelevant image regions are often affected due to entangled learned concepts. In this work, we propose a novel framework based on a pre-trained text-to-image model that enables personalized disentangled editing with precise conceptual and regional control. We first fine-tune our model to extract and embed the relevant concept information in the textual latent space. During fine-tuning, we achieve conceptual disentanglement between the learned embedding and attributes the user wants to edit by enforcing an orthogonality constraint loss. Moreover, we apply masks on the attention maps to ensure the learned embedding only affects regions of interest. At inference time, our method can faithfully preserve the original identity while achieving highly localized editing results. Our method outperforms state-of-the-art methods both qualitatively and quantitatively with better visual quality and editing accuracy. Finally, we explore more practical applications of our model including attribute mixing and makeup transfer.

MODIFACE

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing