ARIA
APPLIED RESEARCH IN ACTION

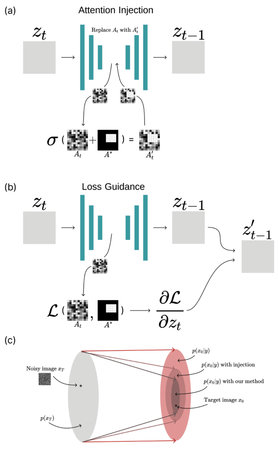
Enhancing Image Composition Control with Loss-Guided Diffusion Models

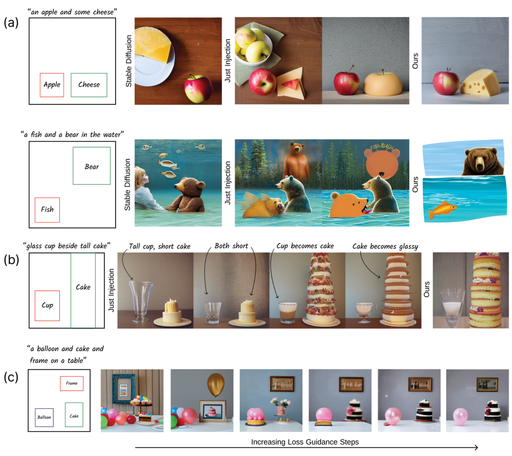
Leveraging the synergy between loss guidance and attention injection to achieve precise layout and composition control.

Zakaria Patel

Kirill Serkh
ACADEMIC SUPERVISOR

Timur Luguev
INDUSTRY SUPERVISOR





PROJECT SUMMARY

Diffusion models are a powerful class of generative models capable of producing very high-quality images from pure noise. In particular, conditional diffusion models allow one to specify the contents of the desired image using a simple text prompt. However, conditioning on a text prompt does not allow one fine-grained control over the composition and layout of the final image, which is instead highly dependent on the initial noise sample. An earlier attempt [1] to solve this problem was to modify the cross-attention maps of the diffusion model, which control the influence of the individual text prompt tokens on the diffusion process, in order to approximately localize each token's influence to specific regions of interest in the final image. We observe that, while it is possible to introduce and manipulate concepts in this way, the concepts may still mix in unexpected ways that are difficult to control. This project proposes a joint approach which leverages both attention injection and diffusion loss guidance to allow for more robust and fine-grained control over the composition of the final image. We show that this approach results in images which are both semantically and compositionally faithful.

REFERENCES

[1] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, Bryan Catanzaro, et al. ediffi: Text-to-image diffusion models with an ensemble of expert denoisers. arXiv preprint arXiv:2211.01324, 2022

Ecomtent

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing