



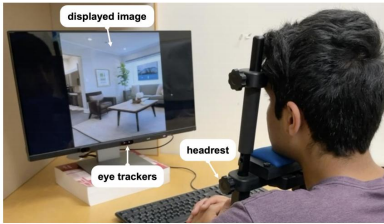
User Study of Saliency-Guided Image Generation Model

Nan Wu
Stanford University

Introduction

Diffusion-based generative models like Stable Diffusion lack control over directing visual attention, which is vital for applications needing targeted viewer focus. Earlier methods like ControlNet variants fall short in manipulating visual saliency, which involves complex multi-level semantics like depth, color, and context. Yunxiang Zhang et al. Developed a saliency-conditioned image diffusion model that produces images that guide viewer attention to specific areas. This project focuses on validating the effectiveness of the model on guiding viewers' attention through a series of user studies, comparing observer saliency patterns against desired saliency conditions.

Experiment Setup



We conducted a user study with eye trackers to record 20 participants' eye gaze patterns while they browse through 150 generated image samples, 50 from each of the Stable Diffusion v2.1 (SD2.1) model, the GLIGEN BBox-guided model, and the saliency-guided model.

Experiment Result Analysis

	Saliency similarity metrics:					
	AUC ↑	NSS ↑	CC ↑	KL ↓	SIM ↑	
TEXT	0.65	0.71	0.21	4.75	0.34	Area Under ROC Curve (AUC)
BBOX	0.78	1.21	0.47	2.67	0.48	Normalized Scanpath Saliency (NSS)
OURS	0.84	1.82	0.78	0.79	0.66	Correlation Coefficient (CC)
						Kullback–Leibler Divergence (KL)
						Histogram Intersection (SIM)

Experiment Images and Saliency Maps

