

MOHAMED BIN ZAYED
UNIVERSITY OF
ARTIFICIAL INTELLIGENCE



GaussCtrl: Multi-View Consistent Text-Driven 3D Gaussian Splatting Editing

Jing Wu*, Jia-Wang Bian*, Xinghui Li, Guangrun Wang, Ian Reid, Philip Torr, Victor Adrian Prisacariu

* denotes equal contribution

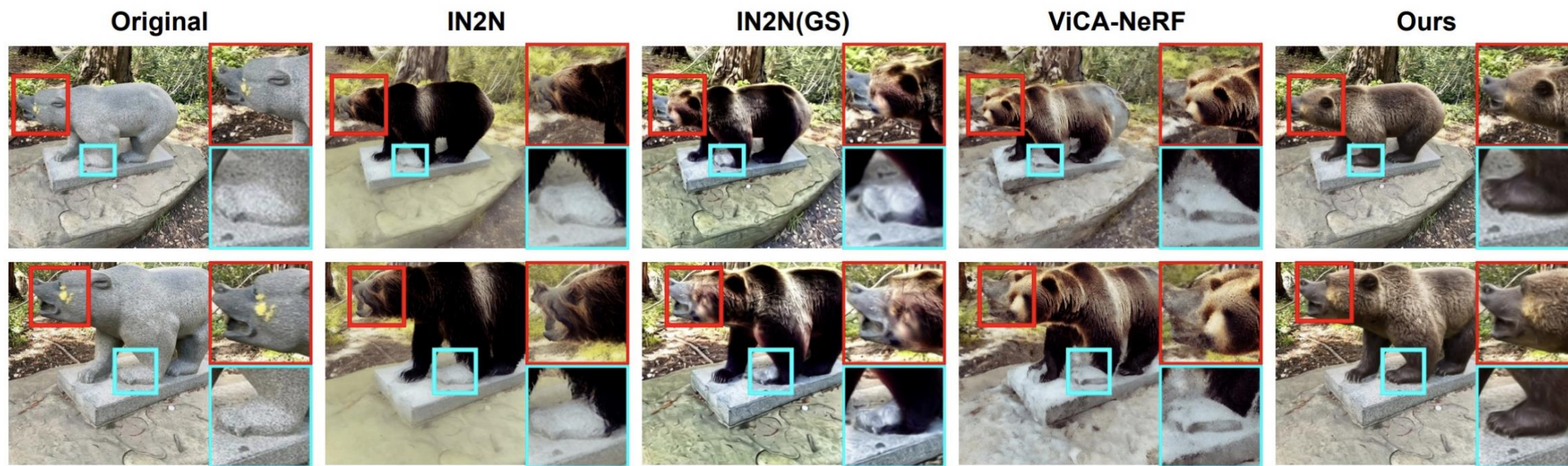
Introduction

What is this task?

Text Prompt Guided 3D Editing

Motivation

To improve the multi-view consistency of 3D editing, reducing visual artefacts such as blurring and inconsistent appearance at different viewpoints.



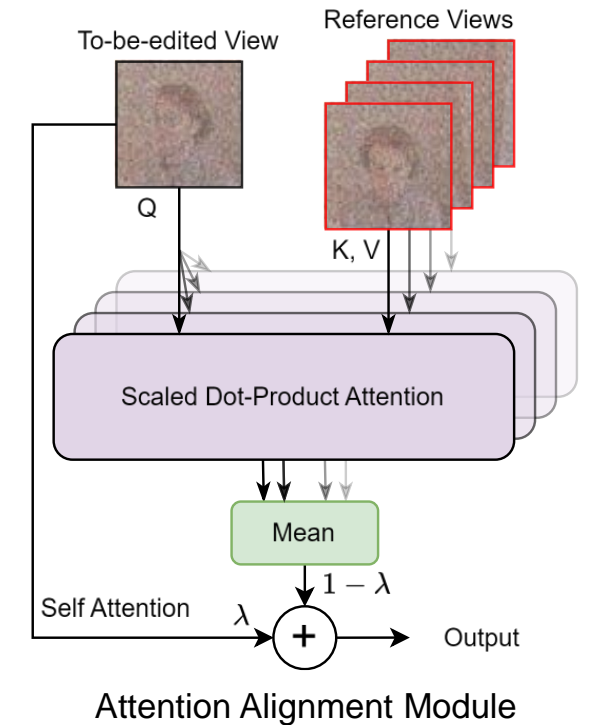
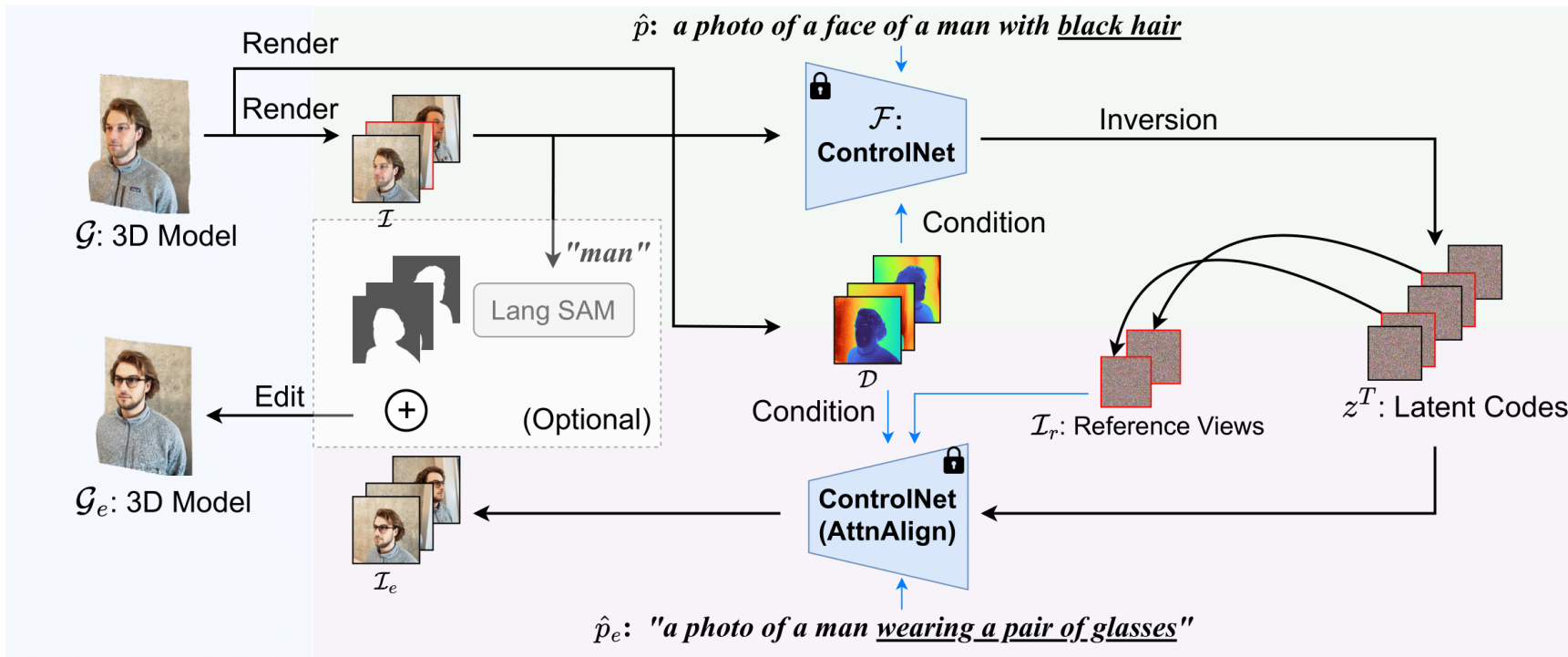
Proposed Method

Depth-conditioned Image Editing:

We utilize ControlNet as our editing backbone and follow the DDIM Inversion approach for 2D editing, where we condition the editing on depths information to encourage multi-view consistency.

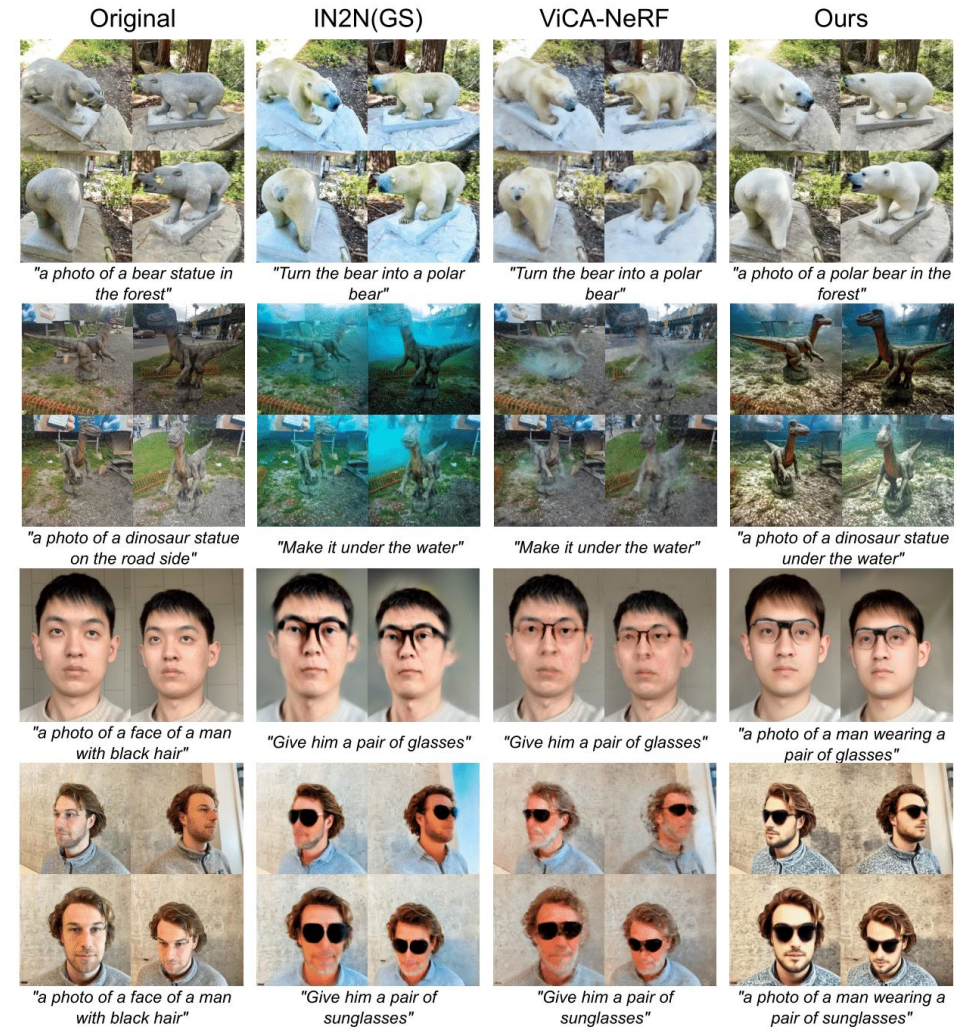
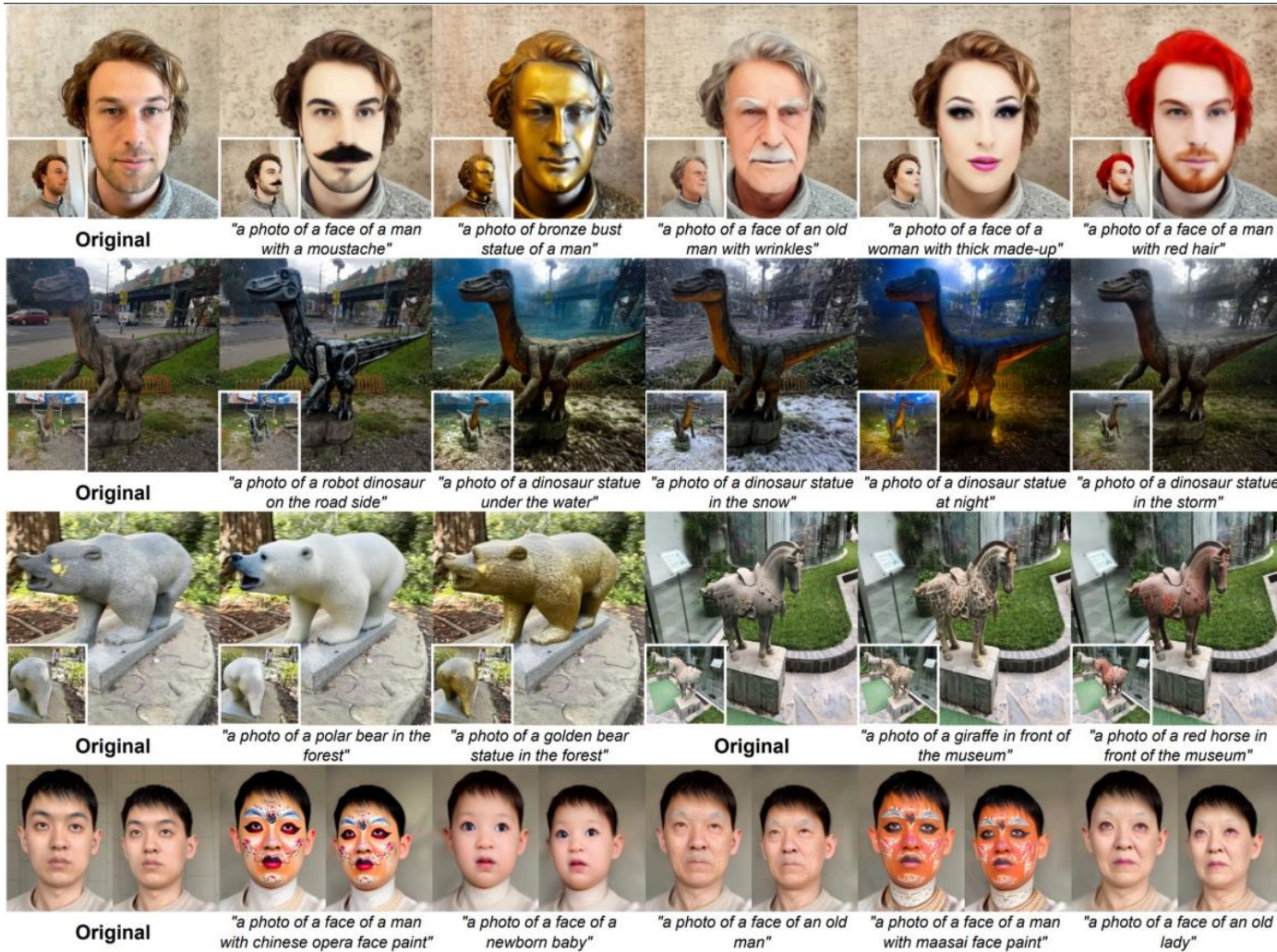
Attention-based Latent Code Alignment

In addition to the depth-conditioned module, we also propose an Attention-based Latent Code Alignment module to unify the appearances of images to a common standard. This ensures greater consistency across edited images and mitigates issues related to appearance discrepancies.



Experiments

Qualitative Results



Experiments

Qualitative Results (Consistency Comparison)



Experiments

Quantitative Results

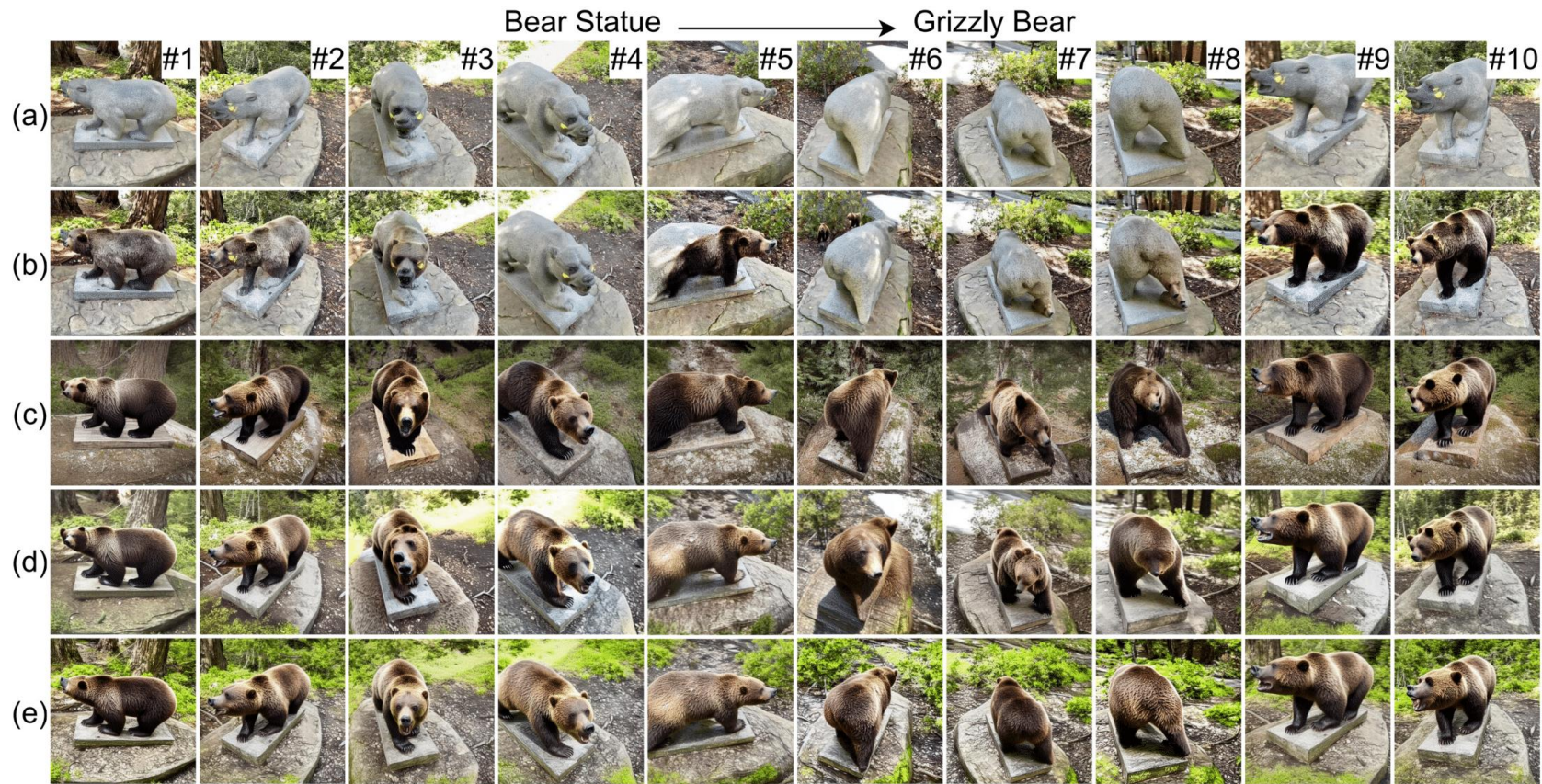
Experiments are conducted on one NVIDIA RTX A5000 24G.

Scenes		IN2N		IN2N(GS)		ViCA-NeRF		Ours	
		$CLIP_{dir}$	Time	$CLIP_{dir}$	Time	$CLIP_{dir}$	Time	$CLIP_{dir}$	Time
360 Scenes	Bear Statue	0.1019	~1.5h	0.1165	~13.5mi n	0.1104	~38.5mi n	0.1388	~9min
	Dinosaur	0.1466		0.1490		0.0732		0.1584	
	Garden	0.3027		0.1663		0.2903		0.2891	
	StoneHorse	0.1654		0.1947		0.1926		0.2268	
Face Forward	Fangzhou	0.1598		0.2032		0.1809		0.1887	
	Face	0.1332		0.1357		0.1119		0.1503	

Experiments

Ablation Study

- (b) One-time Instruct Pix2Pix Edit
- (c) ControlNet with Random Noise
- (d) ControlNet with Inverted Latent Codes (w/o AttnAlign)
- (e) ControlNet with Inverted Latent Codes & AttnAlign



Conclusion

- Propose an efficient 3D-aware consistency control editing method
- The method greatly mitigates the artefacts and blurry results caused by the inconsistency in 2D editing
- Propose the Attention-based Latent Code Alignment module to align the multi-view appearance, increasing the multi-view consistency
- Limitation: Our method struggles to perform significant geometry editing
- Experiments are conducted on diverse scenes, text prompts, and objects
- Project Page: <https://gaussctrl.active.vision/>



🌟 Code, Data and More Results Available Here!



UNIVERSITY OF
OXFORD

Thank you