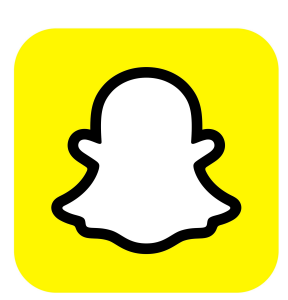CVPR
JUNE 17-21, 2024
SEATTLE, WA

3D Paintbrush: Localized Stylization of 3D Shapes with Cascaded Score Distillation

Dale Decatur[1]   Itai Lang[1]   Kfir Aberman[2]   Rana Hanocka[1]

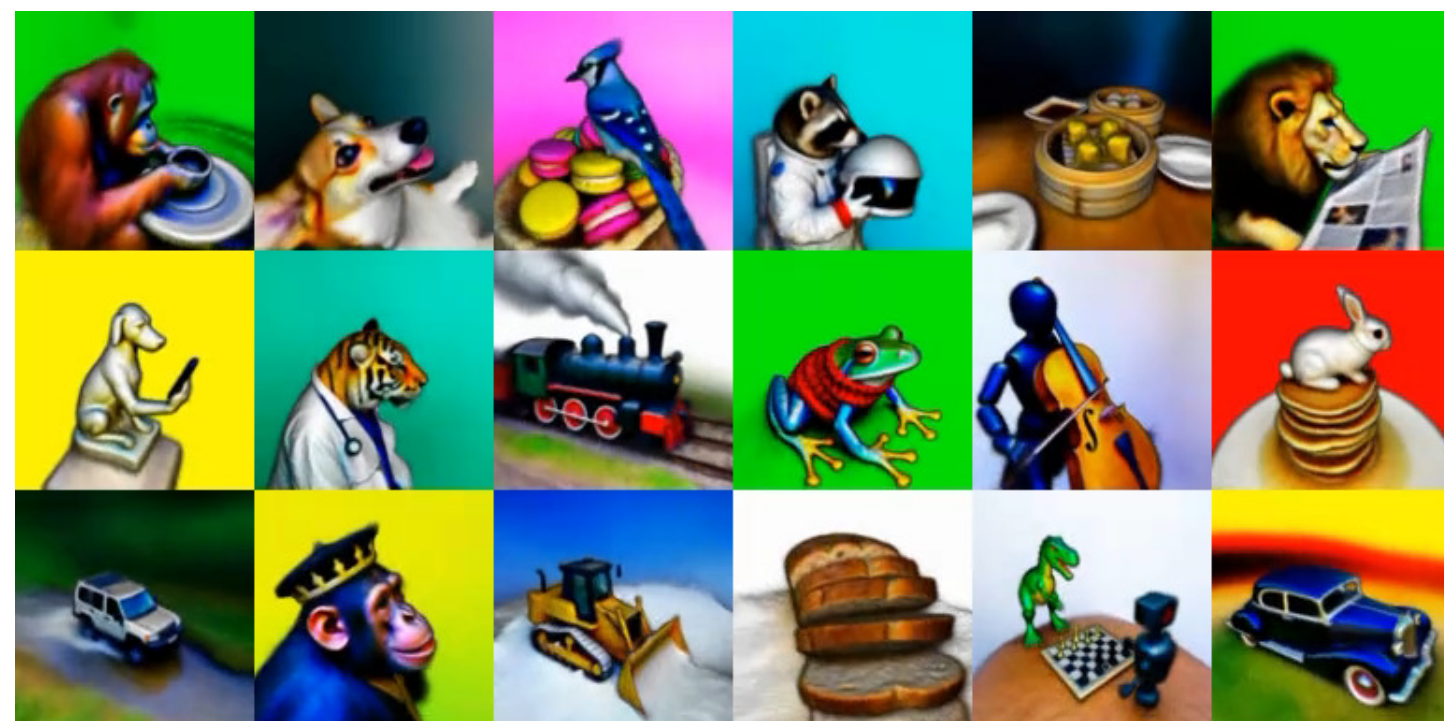University of Chicago[1], Snap Research[2]

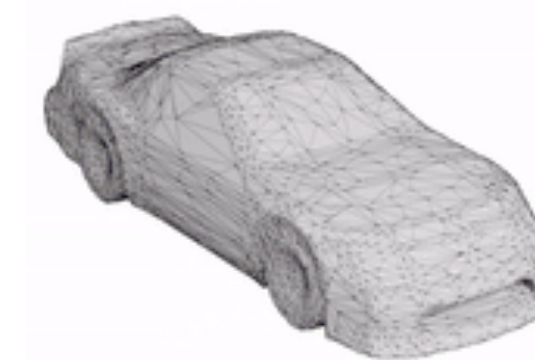# Text-Driven 3D Generation and Editing

Text-Driven 3D Generation and Editing

DreamFusion, Poole et al., 2023

Text2Mesh, Michel et al., 2022

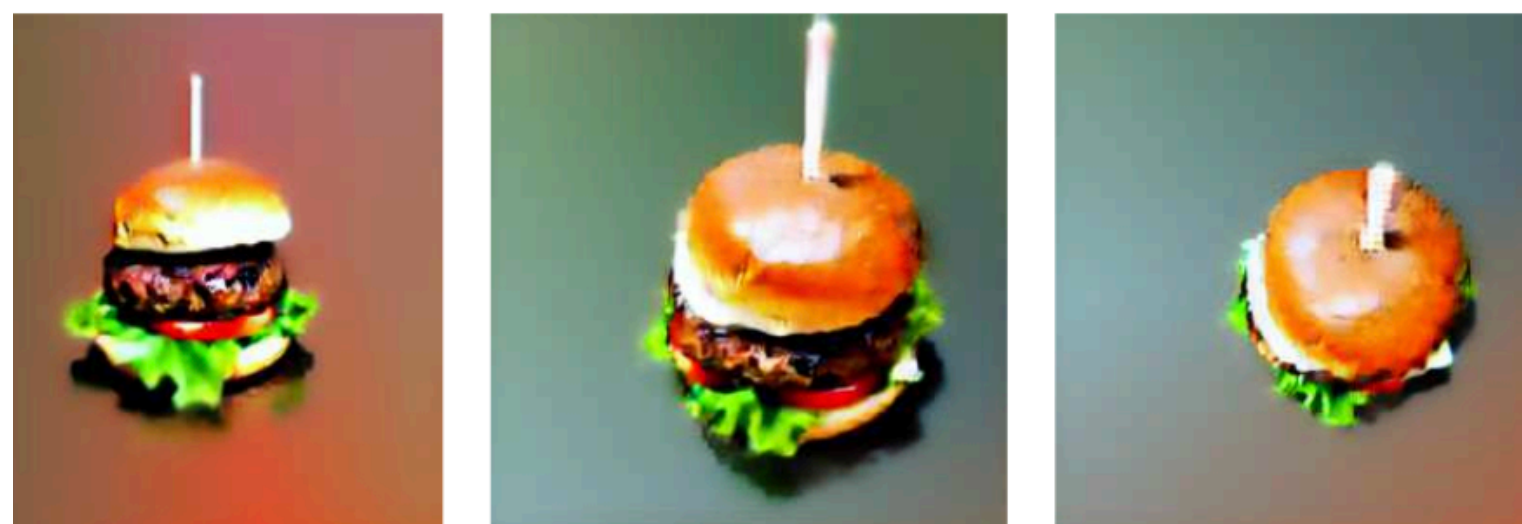
Base mesh   "Nascar car"

Latent-NeRF, Metzer et al., 2023

A high quality photo of a delicious burger

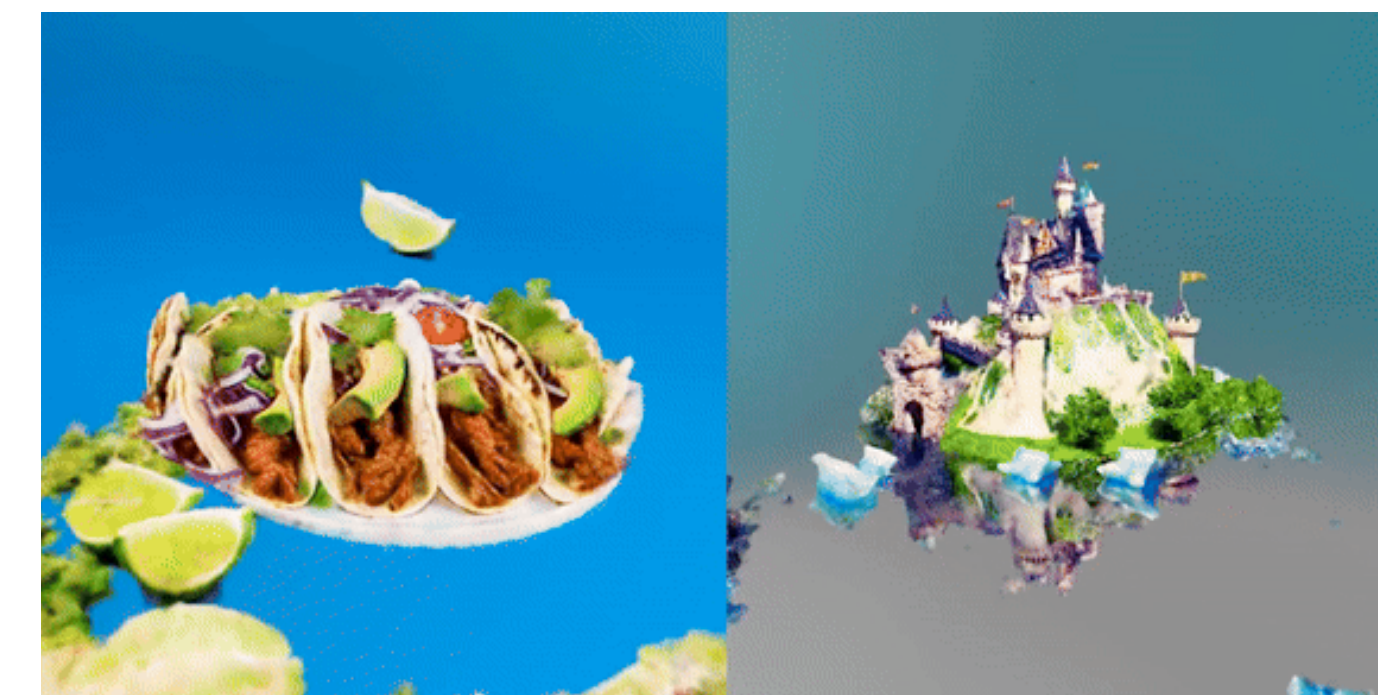Score Jacobian Chaining, Wang et al., 2023

Michelangelo style statue of dog reading news on a cellphone.
A chimpanzee dressed like Henry VIII king of England.
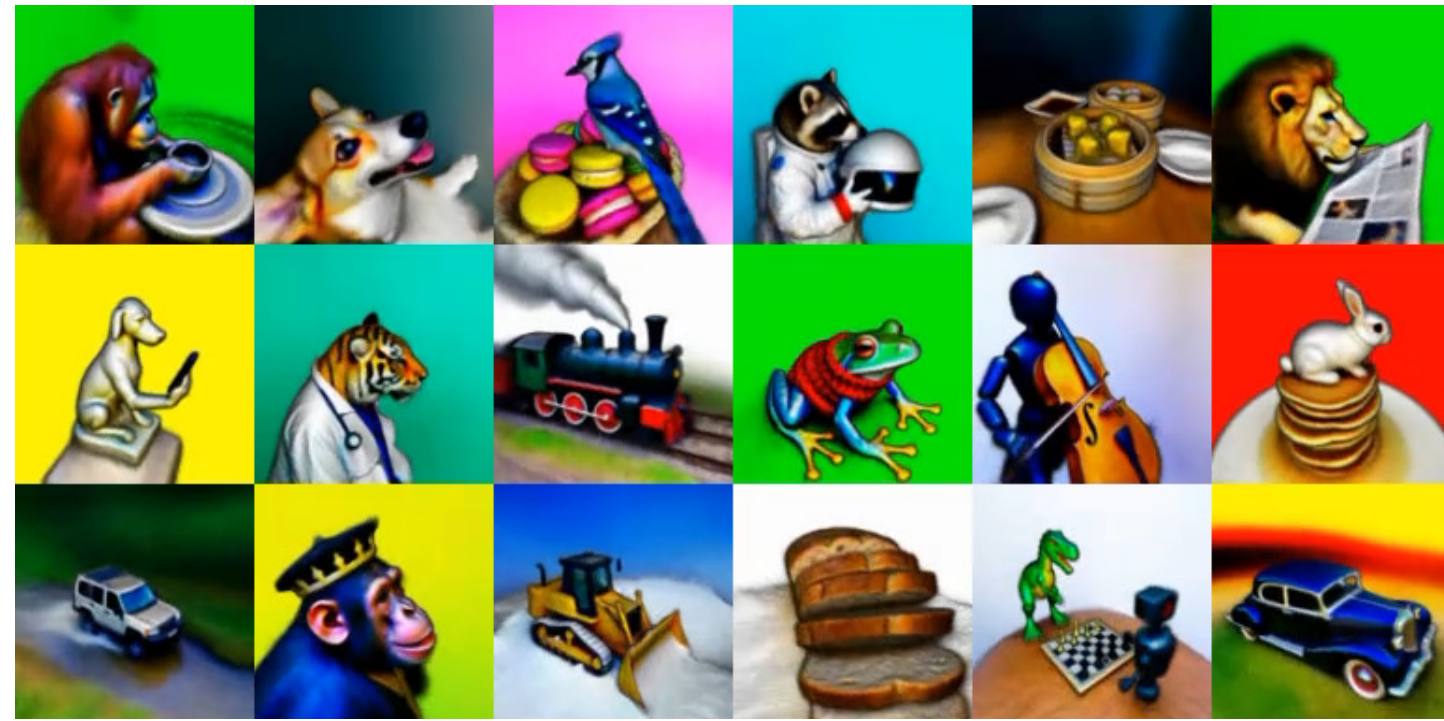An elephant skull.

Prolific Dreamer, Wang et al., 2023
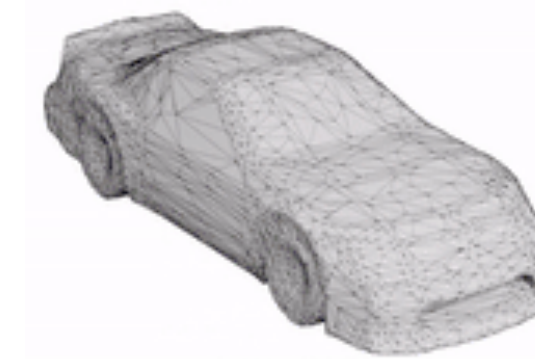

HiFA, Zhu et al., 2023

Text-Driven 3D Generation and Editing

DreamFusion, Poole et al., 2023

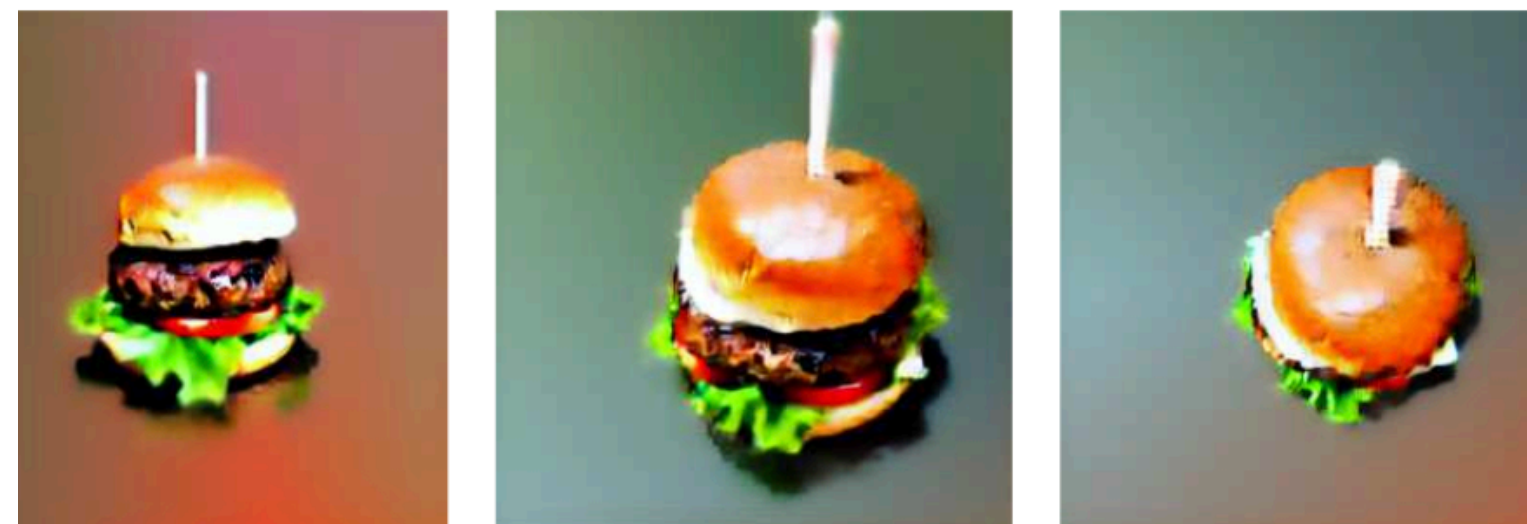Text2Mesh, Michel et al., 2022

Base mesh   "Nascar car"

Latent-NeRF, Metzer et al., 2023

Only Global Edits!

A high quality photo of a delicious burger

Score Jacobian Chaining, Wang et al., 2023

Michelangelo style statue of dog reading news on a cellphone. | A chimpanzee dressed like Henry VIII king of England. | An elephant skull.

Prolific Dreamer, Wang et al., 2023

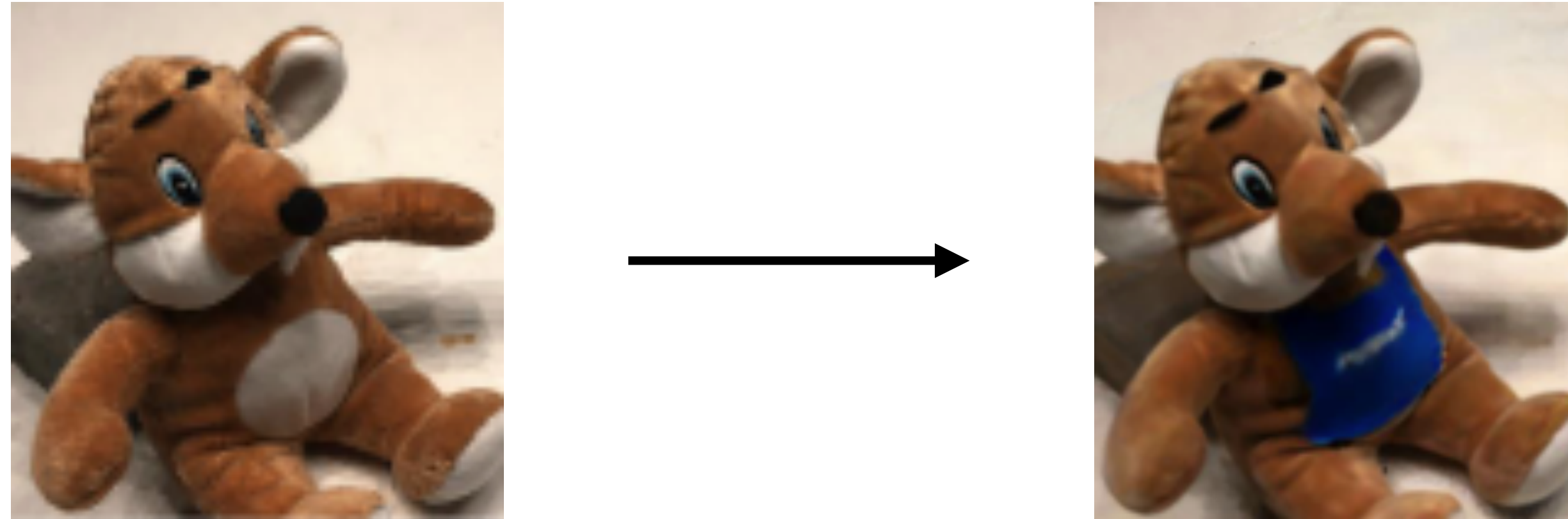HiFA, Zhu et al., 2023

Motivation for Local Editing

DreamEditor [SIGGRAPH Asia 2023]
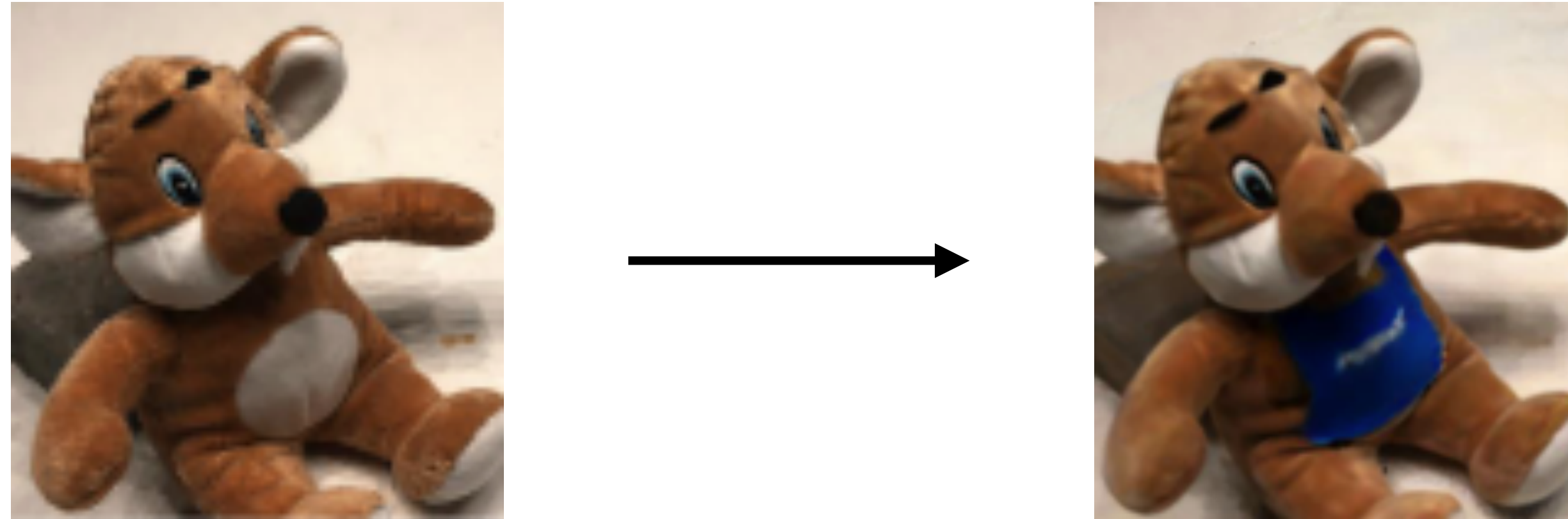
Want to…

Motivation for Local Editing

DreamEditor [SIGGRAPH Asia 2023]

Want to…

- Change **only** local components

Motivation for Local Editing

DreamEditor [SIGGRAPH Asia 2023]

Want to…

- Change **only** local components
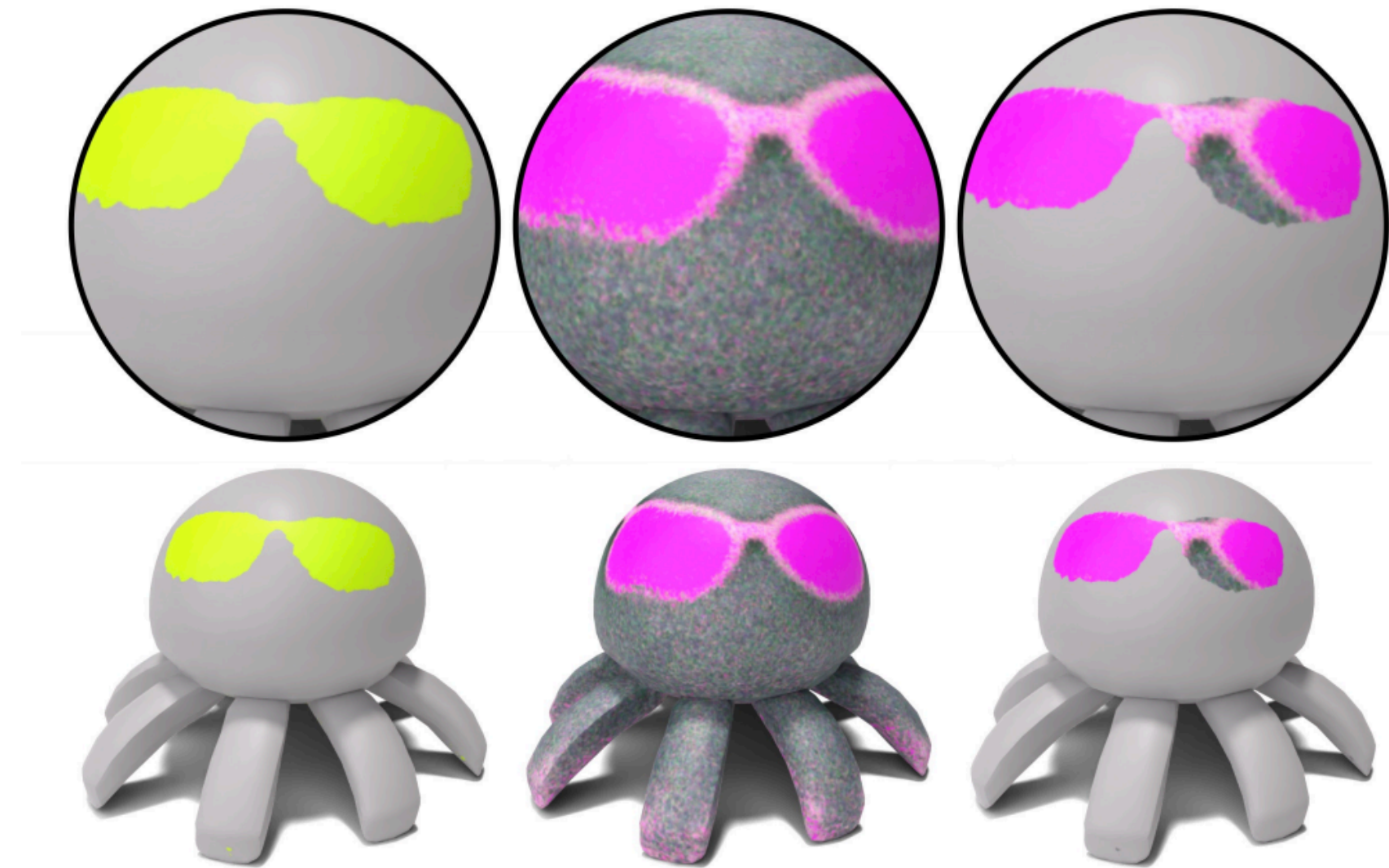
- Preserve the overall identity

Challenges of Local Editing
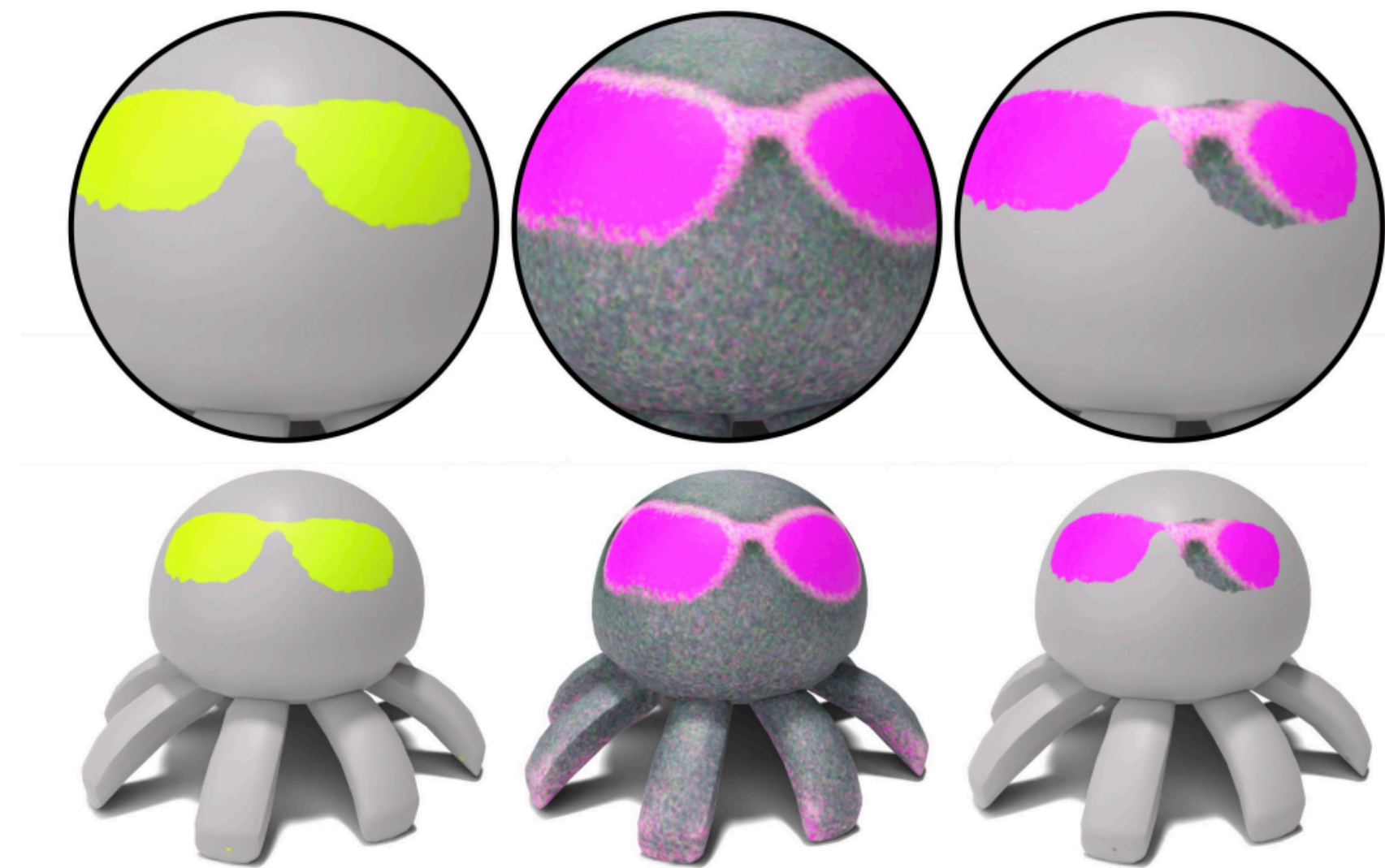
Challenges of Local Editing

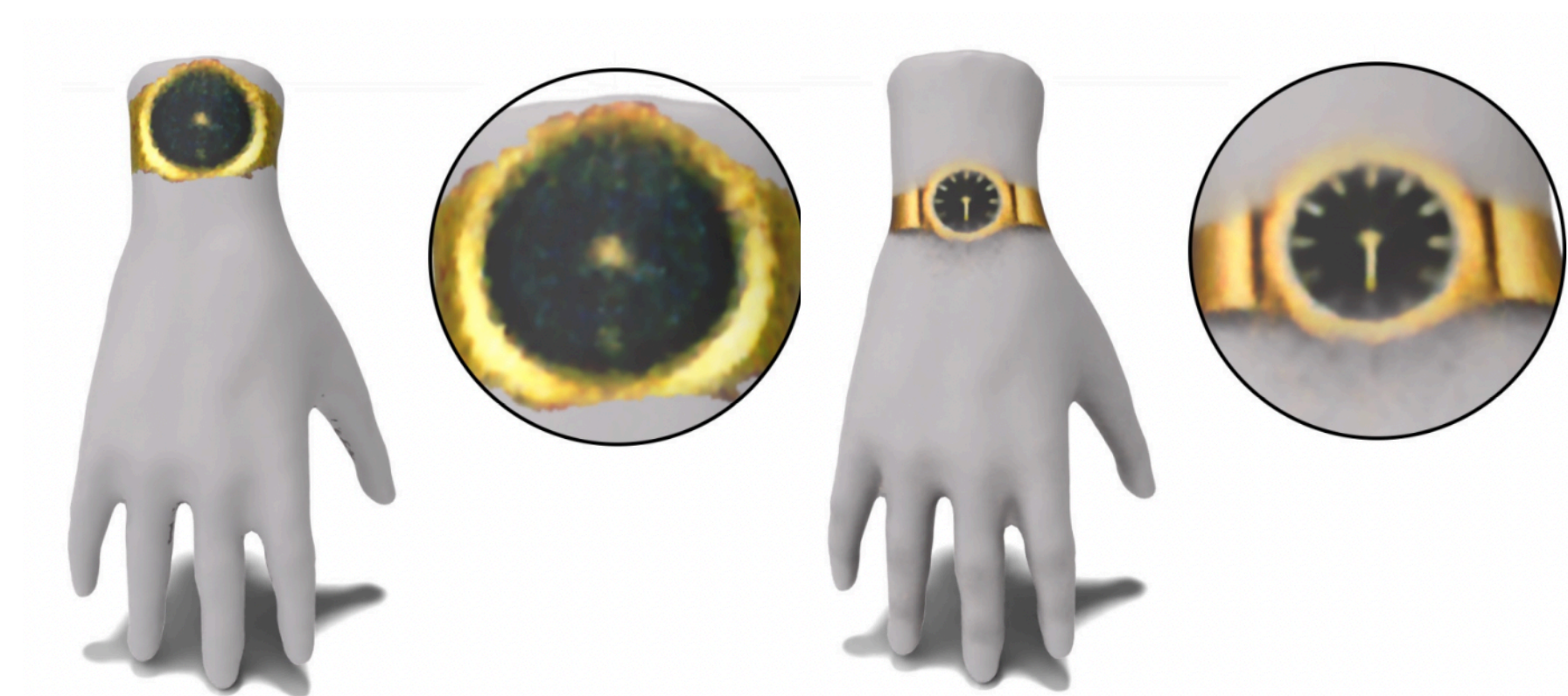- Requires explicit edit region identification

Challenges of Local Editing

- Requires explicit edit region identification

- Requires high resolution supervision

3D Paintbrush

Input 3D Object

3D Paintbrush

"A gold chain necklace"

Input 3D Object

Locally Edited 3D Object

Joint Optimization of Localization and Texture

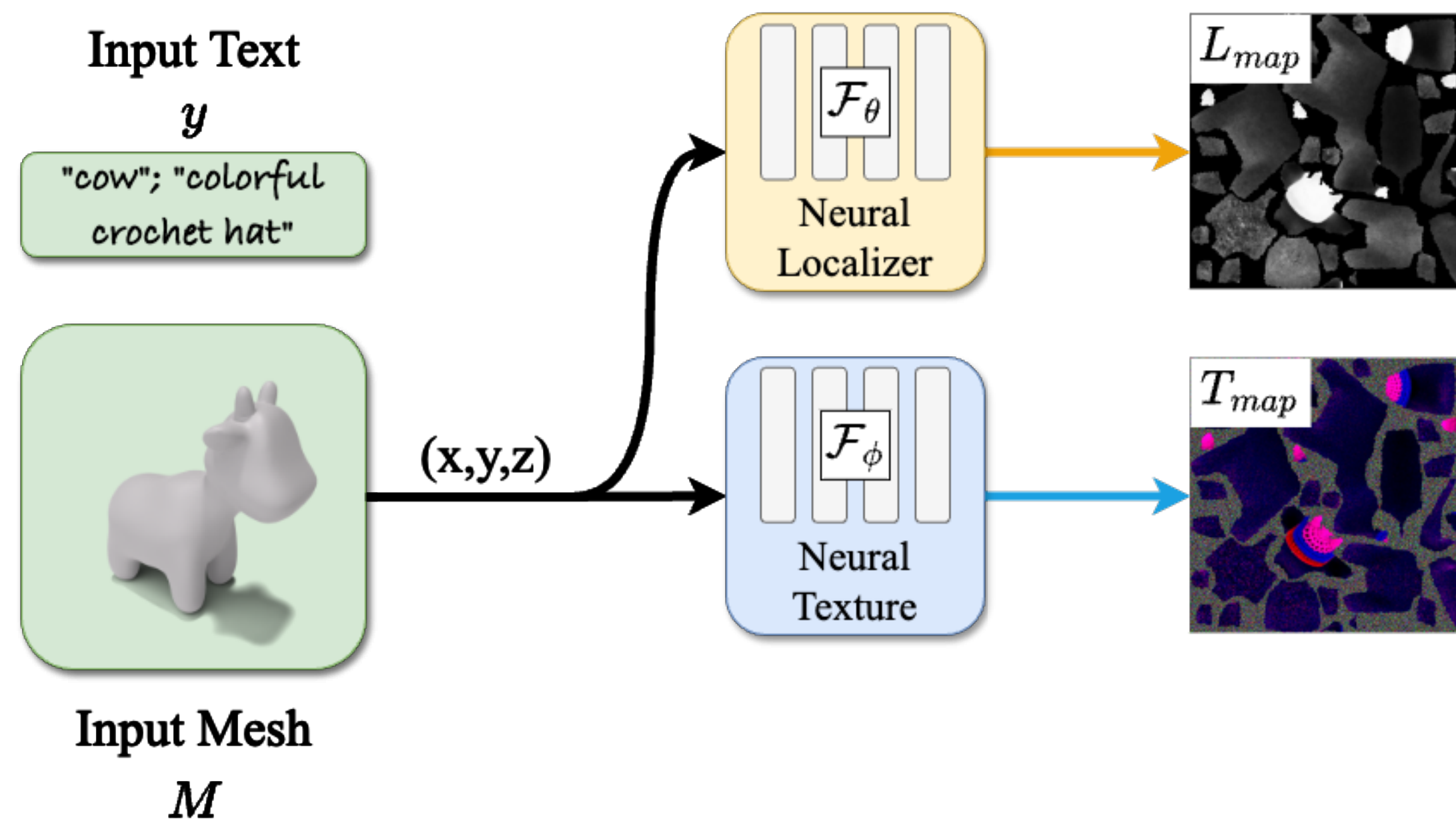## Input Text
y

"cow"; "colorful crochet hat"

Input Mesh
M

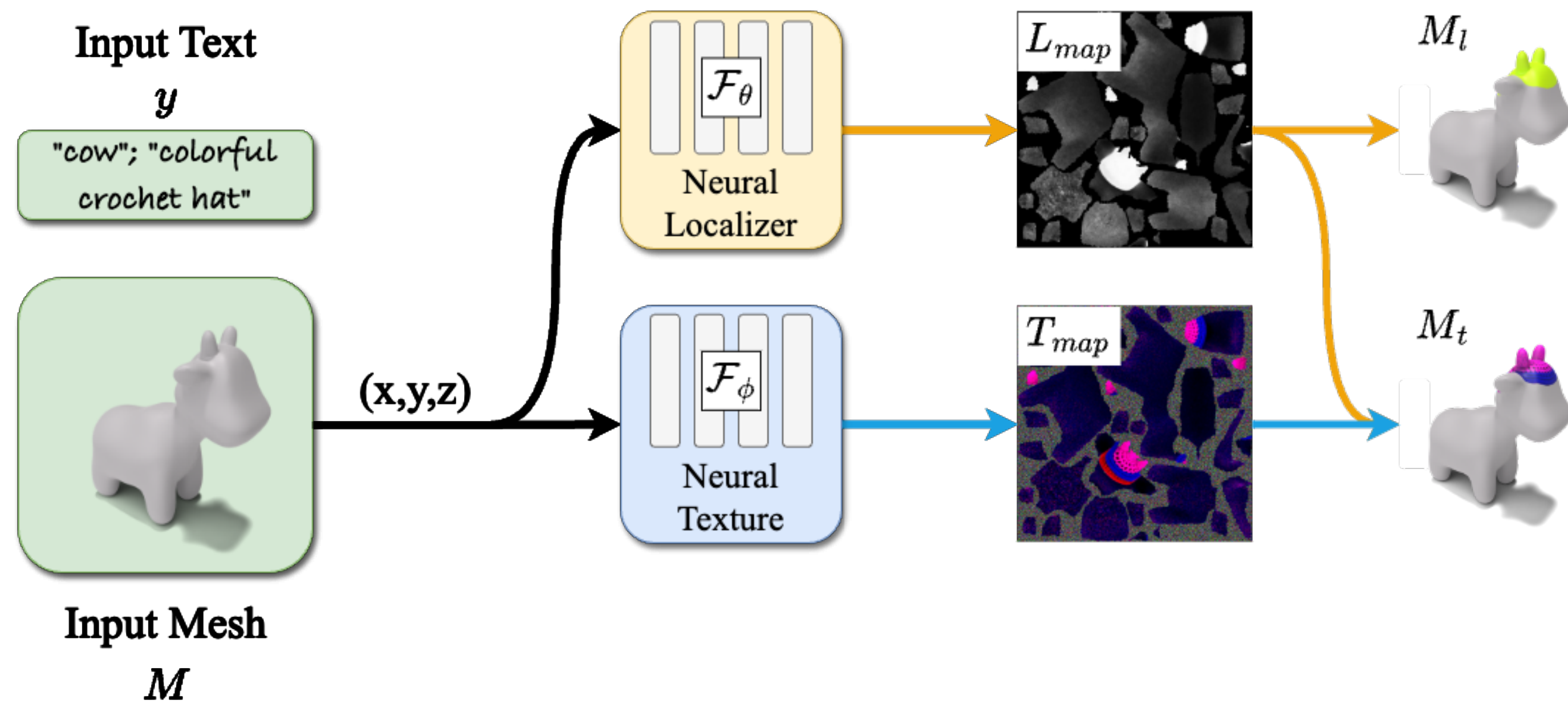# Joint Optimization of Localization and Texture

Joint Optimization of Localization and Texture

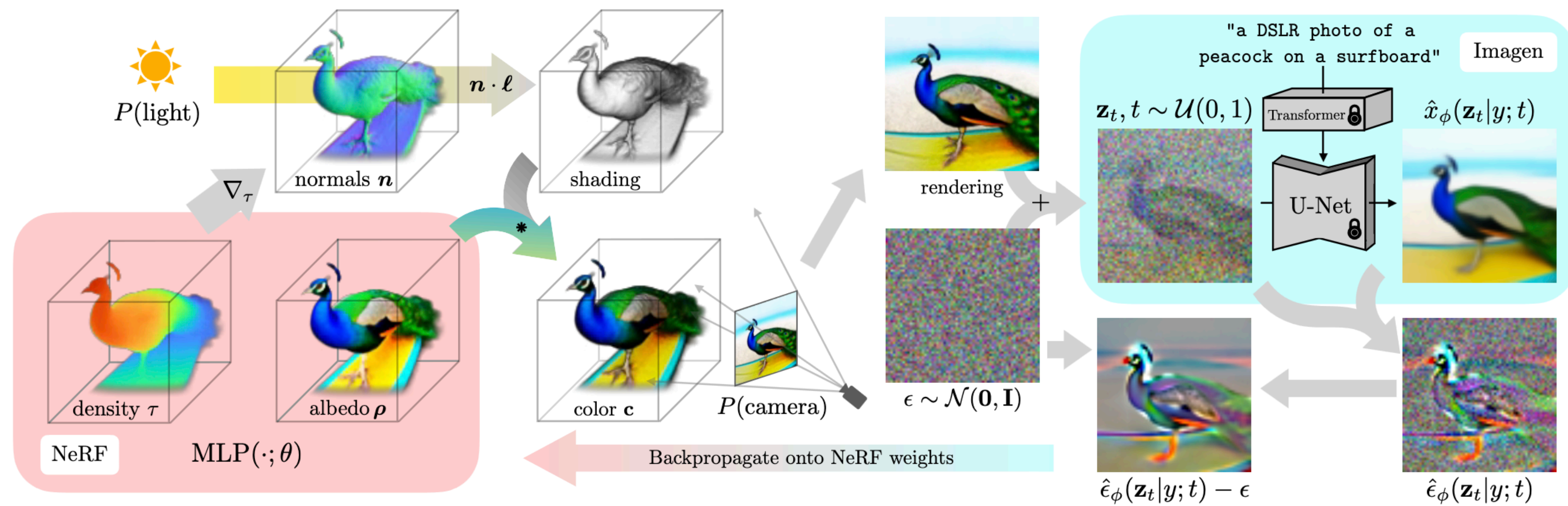# Joint Optimization of Localization and Texture

- Score Distillation Sampling (SDS) enables 3D supervision using text-to-image diffusion models

DreamFusion, Poole et al., 2023
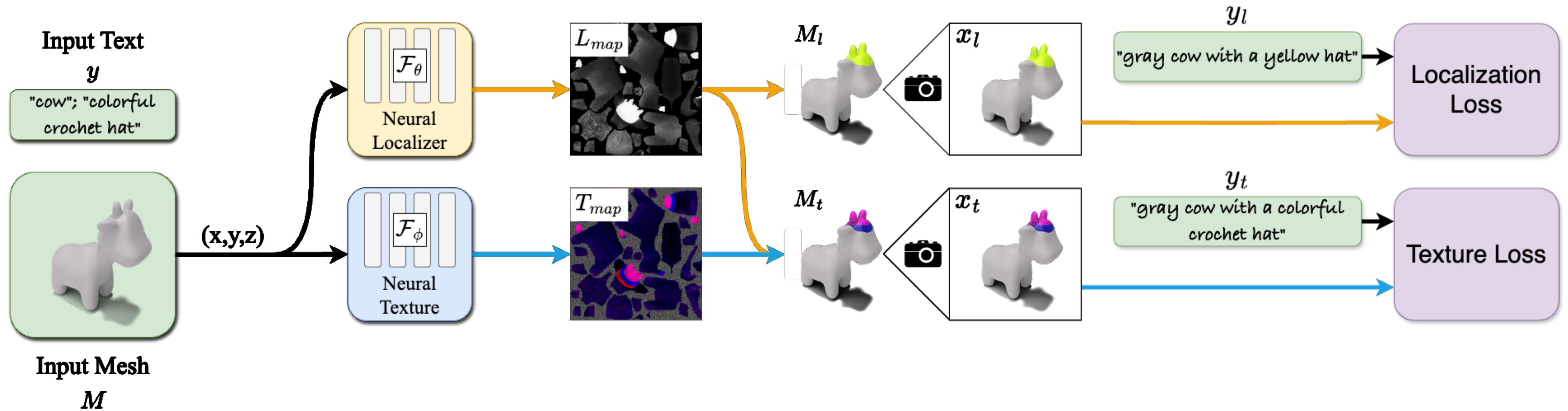
3D Supervision from 2D Models

- Cascaded diffusion models combine multiple stages to achieve higher resolution
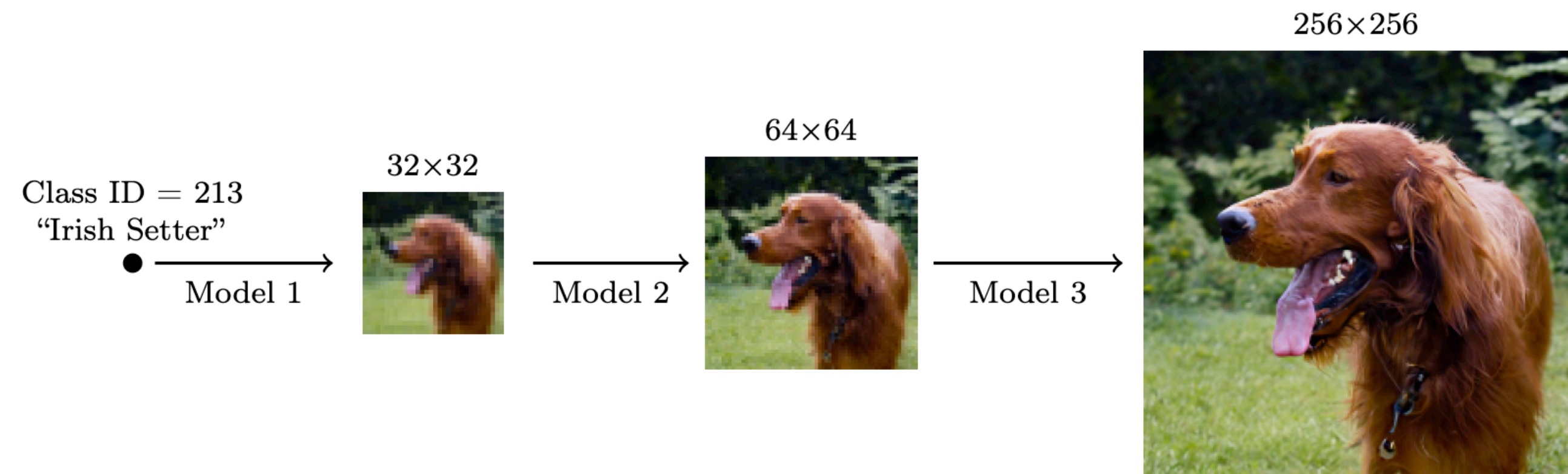

Cascaded Diffusion Models for High Fidelity Image Generation, Ho et al., 2022

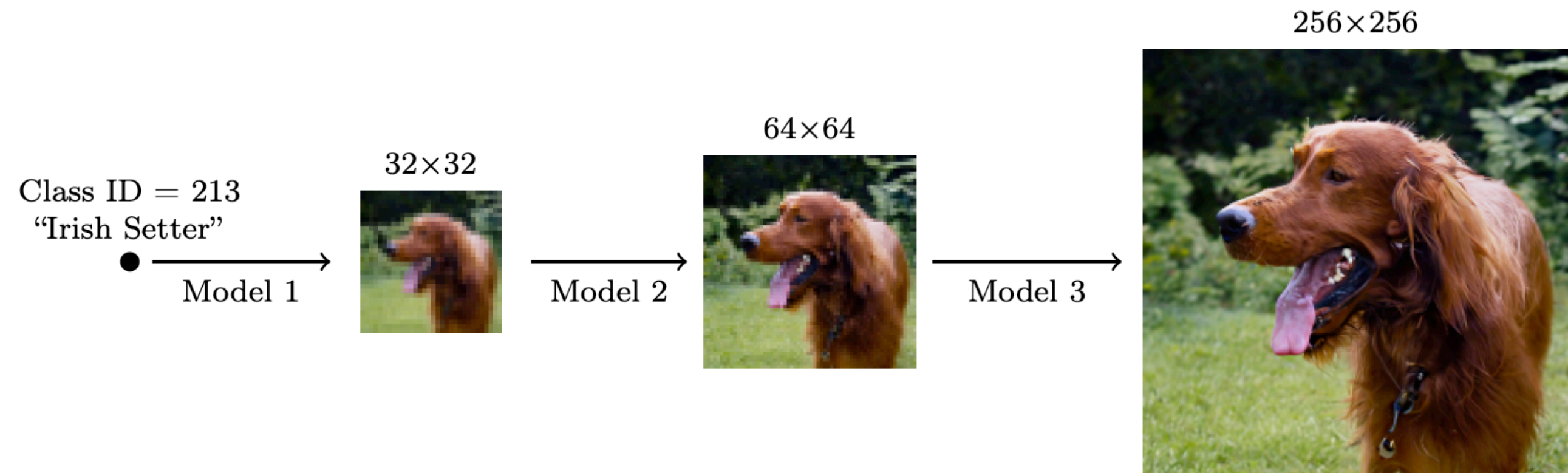- Cascaded diffusion models combine multiple stages to achieve higher resolution

Cascaded Diffusion Models for High Fidelity Image Generation, Ho et al., 2022
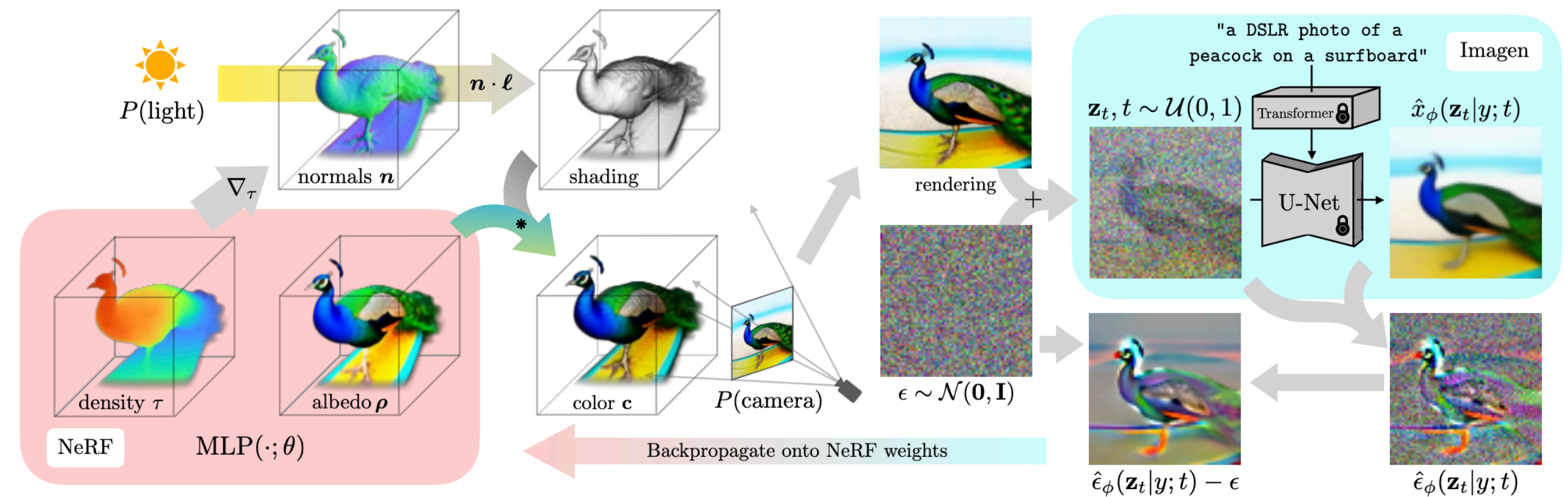
- Standard SDS only utilizes the base stage

DreamFusion, Poole et al., 2023

Cascaded Score Distillation (CSD)
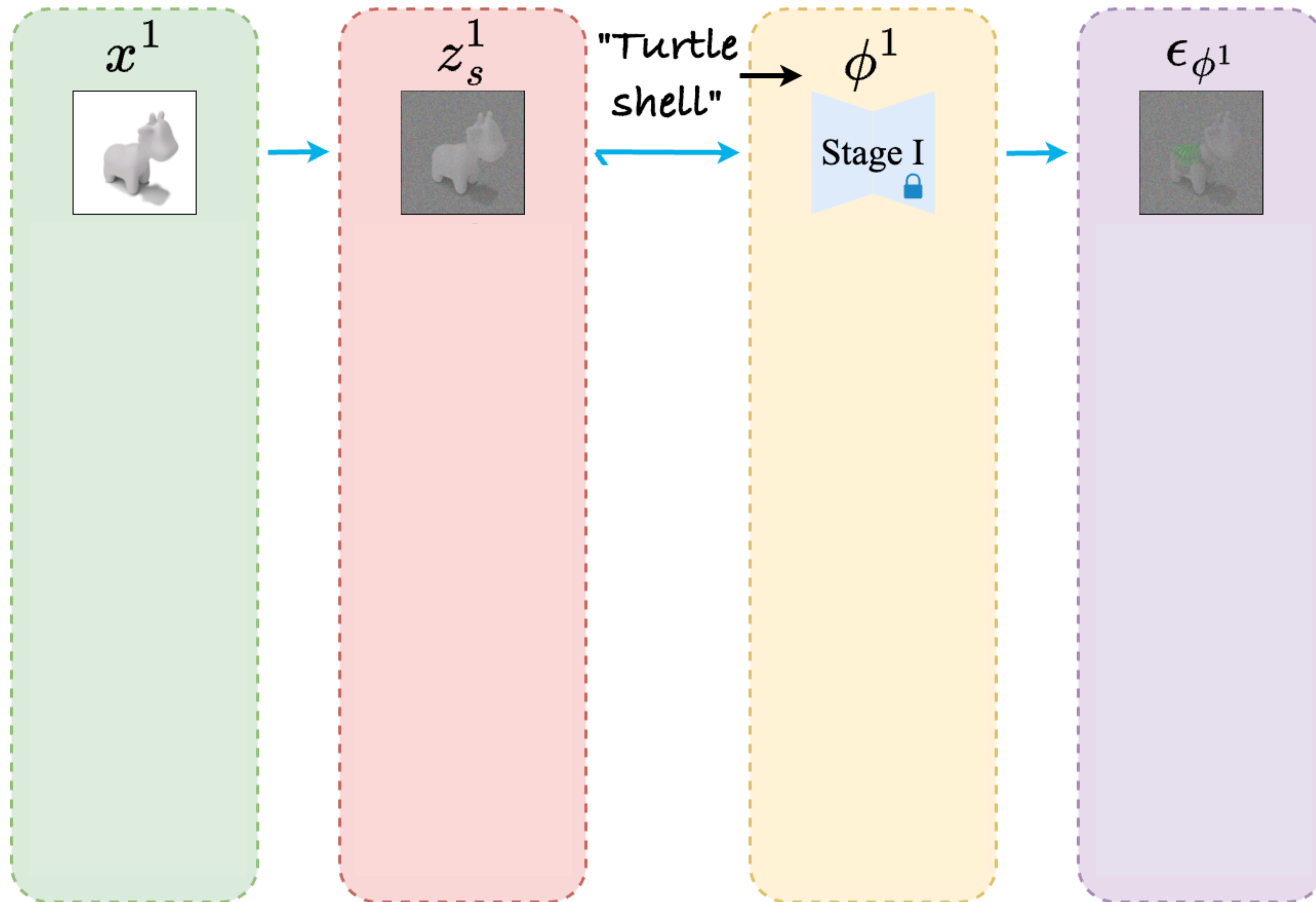
Cascaded Score Distillation (CSD)
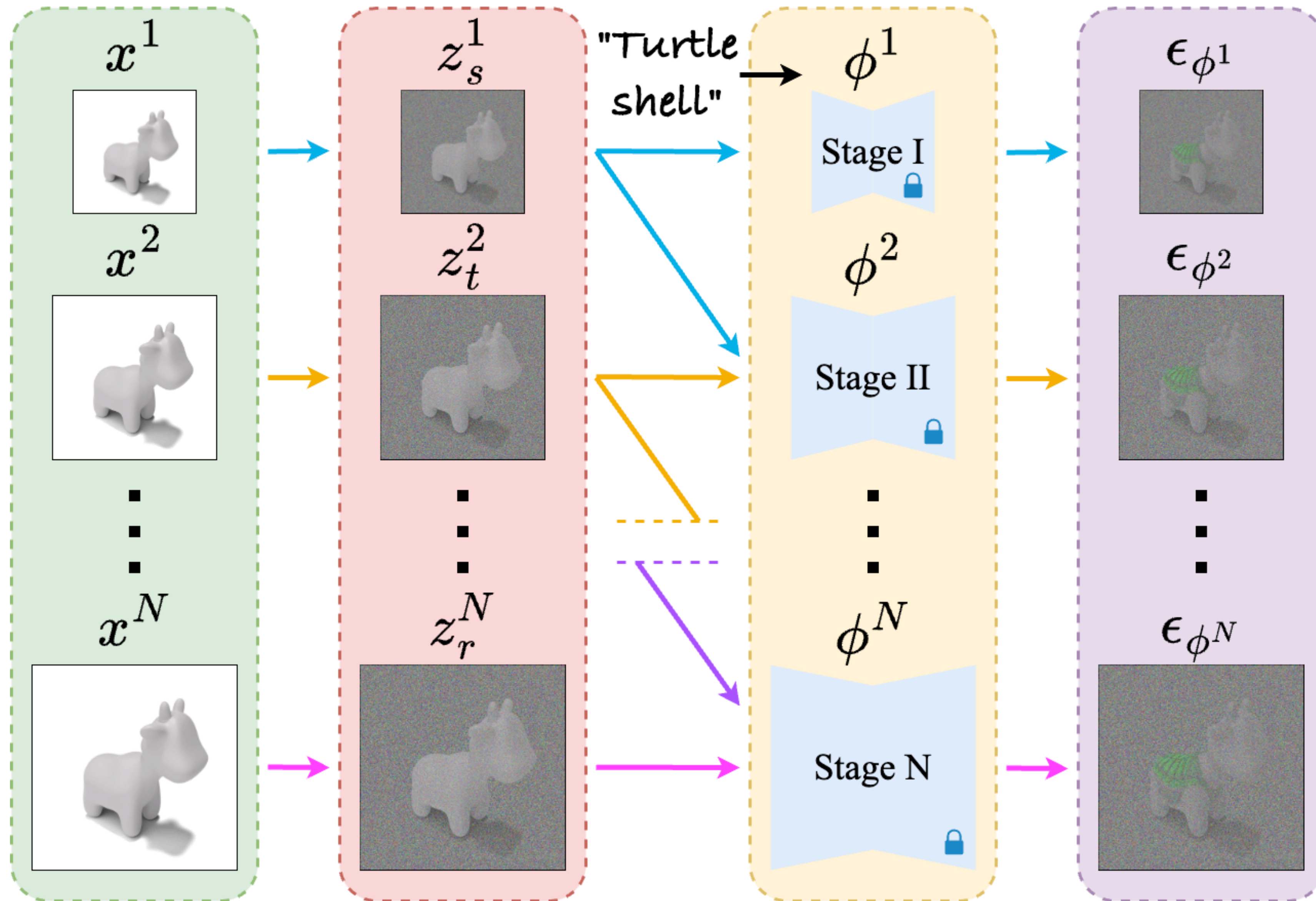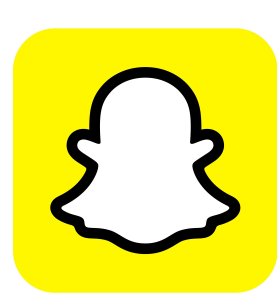
CSD vs SDS

CSD vs SDS

SDS

CSD vs SDS

SDS

CSD

CSD vs SDS

SDS
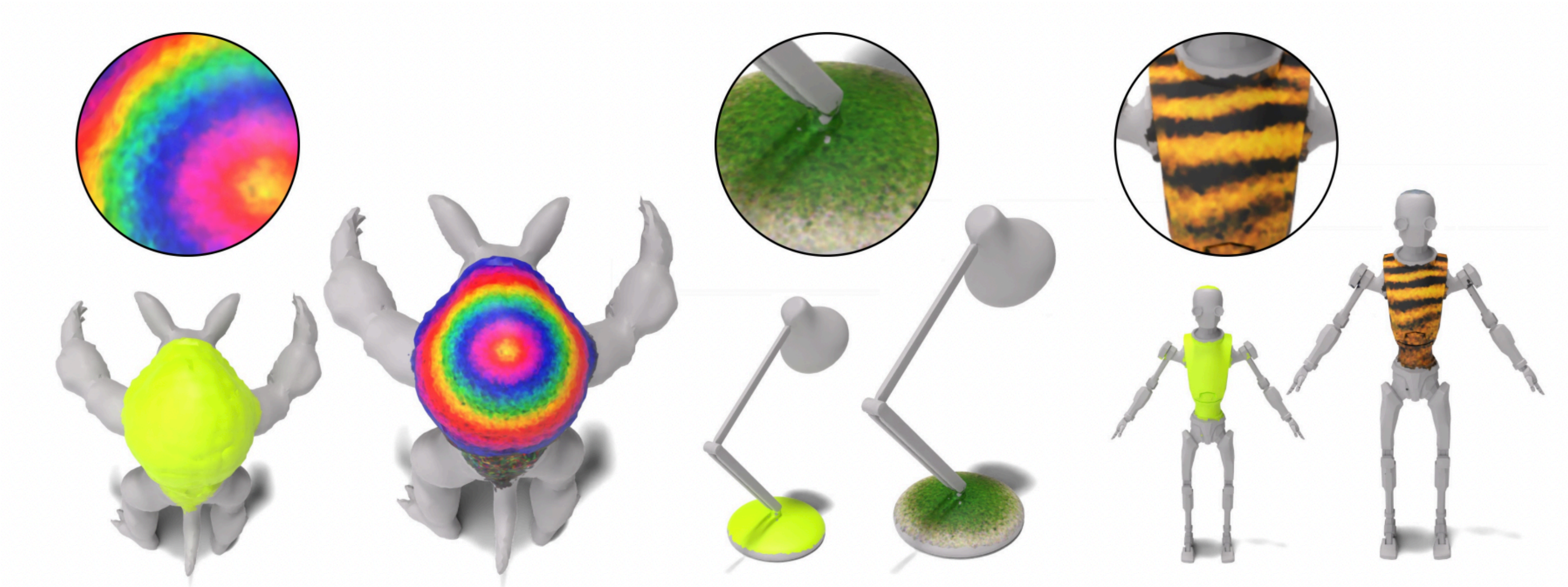
CSD
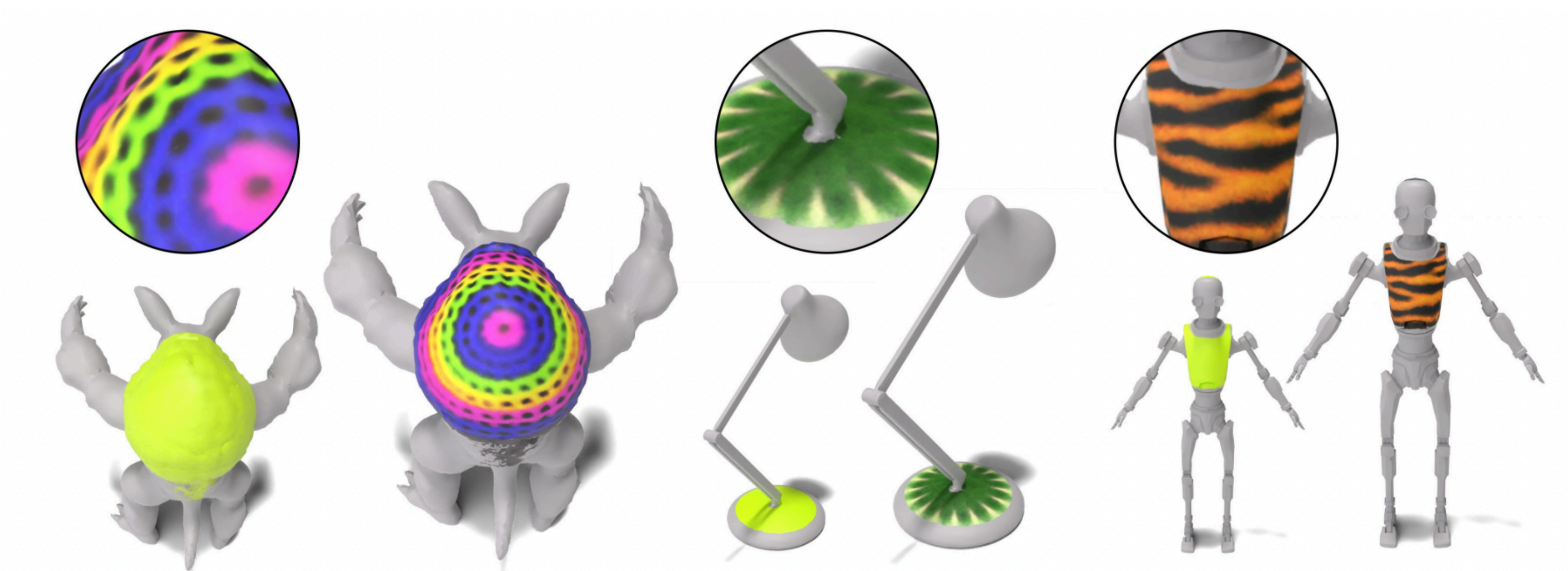
3D Paintbrush Results
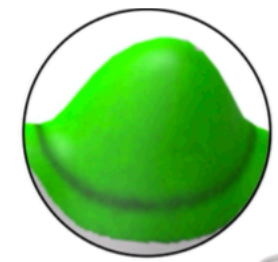
Red bow tie   Turtle shell   Beautiful rose hump   Red bow tie   Barcelona jersey   Denim overalls   Camo poncho   Superhero cape   Tiger stripe hat
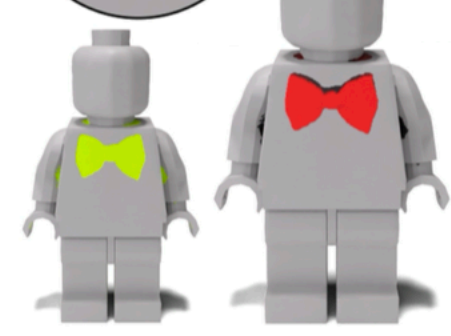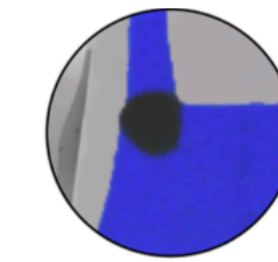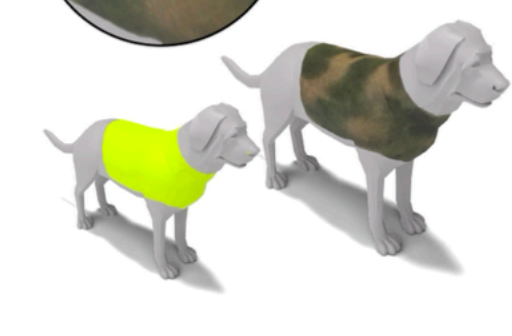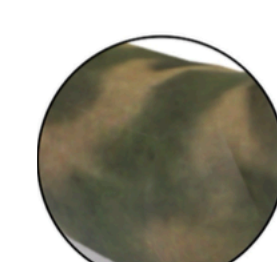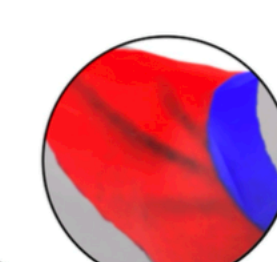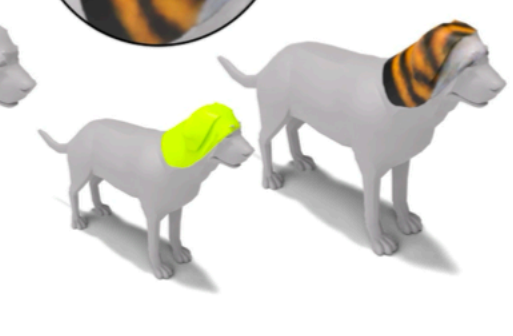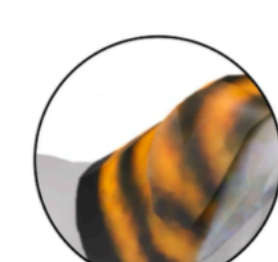
Camo poncho   Beautiful rose   Colorful doily seat cushion   Batman emblem   Rainbow headband

Tiger stripe belt   Rainbow shinguards   Tie-dye shirt   Red flip flops   Green ridged turtle shell

Ours

Qualitative Evaluation

Ours

Latent Paint

Latent Paint, Metzer et al., 2023

Vox-E

Vox-E, Sella et al., 2023

SATR, Abdelreheem et al., 2023; 3D Highlighter, Decatur et al., 2023; Latent Paint, Metzer et al., 2023; Vox-E, Sella et al., 2023;

Quantitative Evaluation

Perceptual Study

Localization			SATR	3D Highlighter	**Ours**
Average Score ↑			1.89	2.03	**4.80**
Local Edits	Latent Paint	Vox-E		Ours (SDS)	**Ours**
Average Score ↑	2.14	2.15		4.06	**4.88**

SATR, Abdelreheem et al., 2023; 3D Highlighter, Decatur et al., 2023; Latent Paint, Metzer et al., 2023; Vox-E, Sella et al., 2023;

Quantitative Evaluation

Perceptual Study

Localization			SATR	3D Highlighter	**Ours**
Average Score ↑			1.89	2.03	**4.80**

Local Edits	Latent Paint	Vox-E	Ours (SDS)	**Ours**
Average Score ↑	2.14	2.15	4.06	**4.88**

	Clip R-Precision ↑					
Method	CLIP B/32		CLIP B/16		CLIP L/14	
	Loc	Style	Loc	Style	Loc	Style
3DH	13.3	n/a	7.0	n/a	23.7	n/a
LatentPaint	n/a	20.0	n/a	20.0	n/a	20.0
Ours	**26.7**	**60.0**	**13.0**	**40.0**	**46.7**	**73.3**

SATR, Abdelreheem et al., 2023; 3D Highlighter, Decatur et al., 2023; Latent Paint, Metzer et al., 2023; Vox-E, Sella et al., 2023;

Summary

Heart-shaped sunglasses

Summary

Heart-shaped sunglasses

Summary

- Our approach to local editing **produces better localizations and local textures** than existing works

Heart-shaped sunglasses

Summary

- Our approach to local editing **produces better localizations and local textures** than existing works

- Introduce Cascaded Score Distillation (CSD) a **new loss** which **enables higher resolution supervision** for 3D generative tasks

Heart-shaped sunglasses

Summary

- Our approach to local editing **produces better localizations and local textures** than existing works

- Introduce Cascaded Score Distillation (CSD) a **new loss** which **enables higher resolution supervision** for 3D generative tasks

- **CSD** can be used for **general purpose text-to-3D** applications to increase supervision resolution

Heart-shaped sunglasses

Summary

- Our approach to local editing **produces better localizations and local textures** than existing works

- Introduce Cascaded Score Distillation (CSD) a **new loss** which **enables higher resolution supervision** for 3D generative tasks

- **CSD** can be used for **general purpose text-to-3D** applications to increase supervision resolution

Code is available!

Heart-shaped sunglasses