



ATT3D: Amortized Text-to-3D

Jonathan Lorraine, Kevin Xie, Xiaohui Zeng, Chen-Hsuan Lin, Towaki Takikawa,
Nicholas Sharp, Tsung-Yi Lin, Ming-Yu Liu, Sanja Fidler, James Lucas

Progress in Text-to-3D

DreamFusion



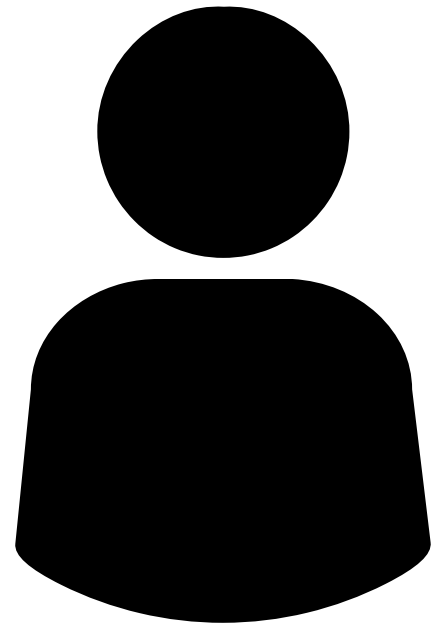
a squirrel wearing
an elegant ballgown
playing the
saxophone

Magic3D

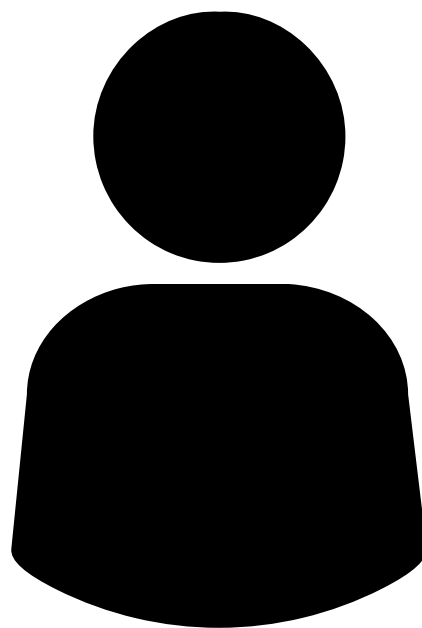


a brightly colored
mushroom growing on
a log

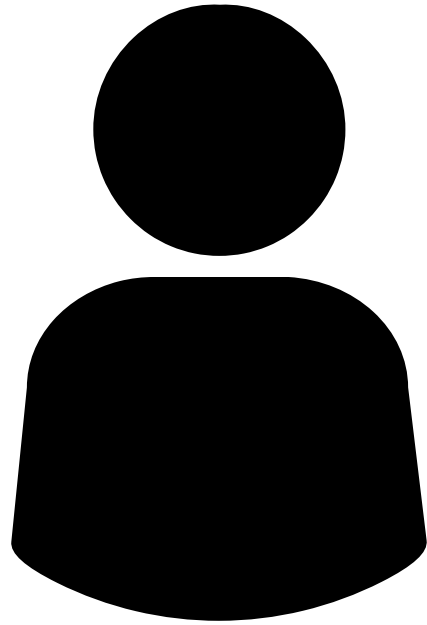
Existing text-to-3D methods



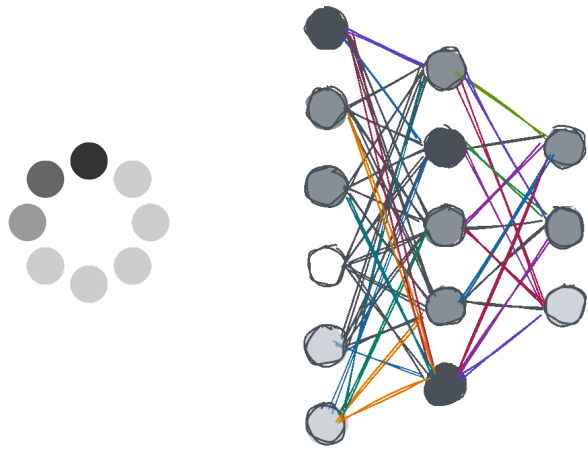
A pig wearing a
tophat holding
a blue balloon



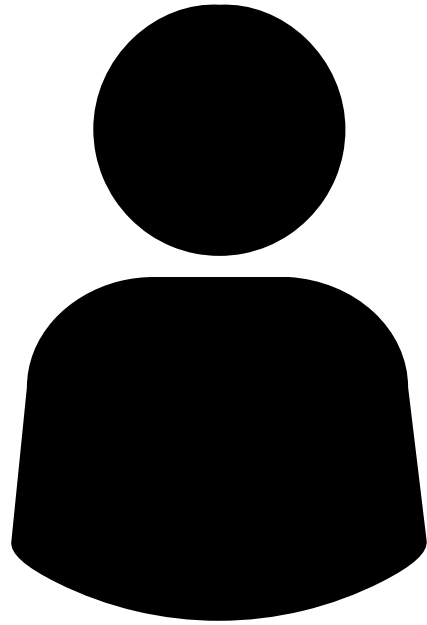
A pig wearing a
tophat holding
a blue balloon



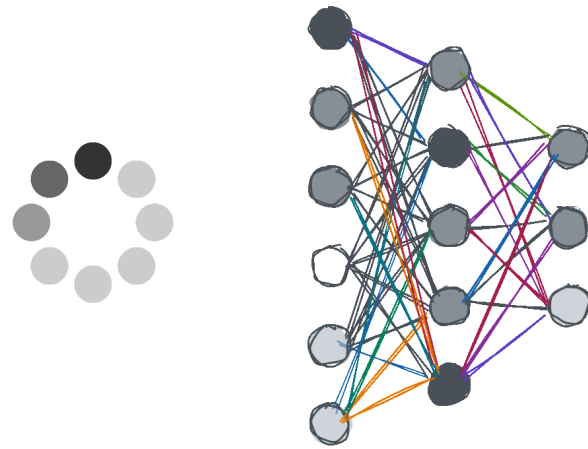
Coarse optimization stage



A pig wearing a
tophat holding
a blue balloon



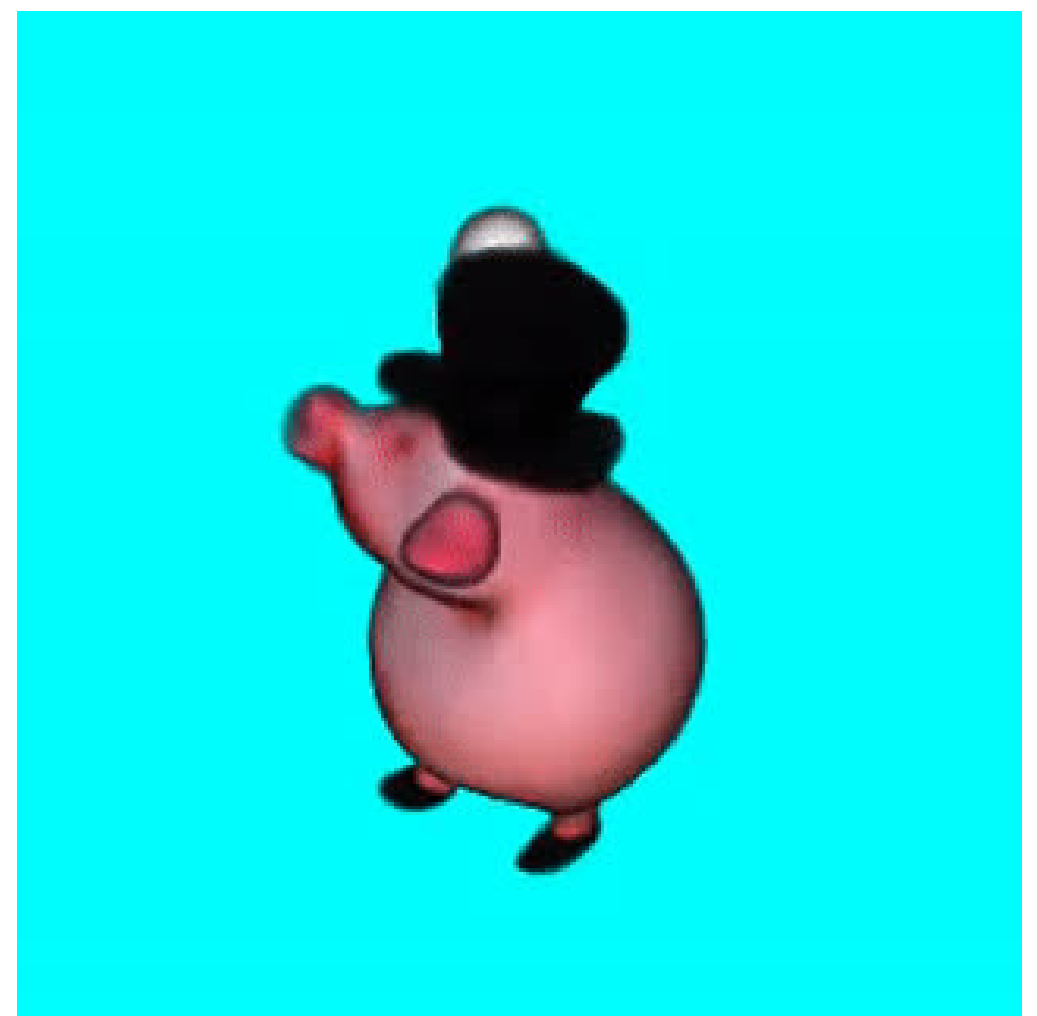
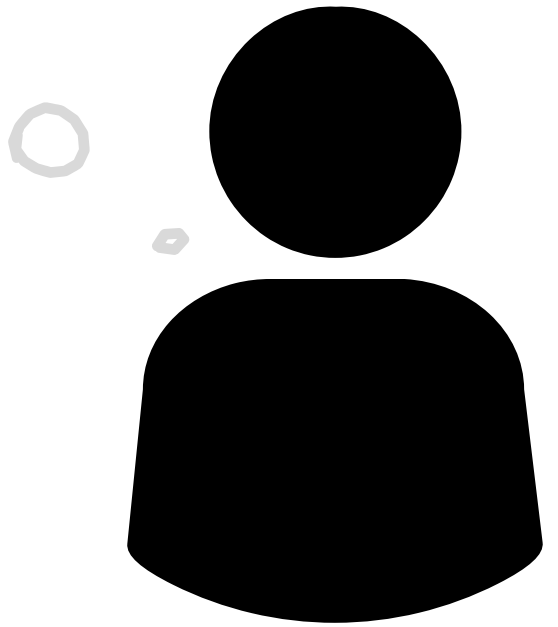
Currently 10-15 mins



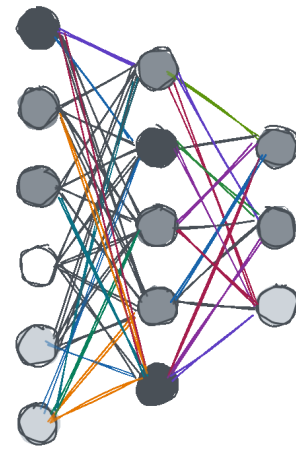
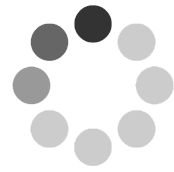
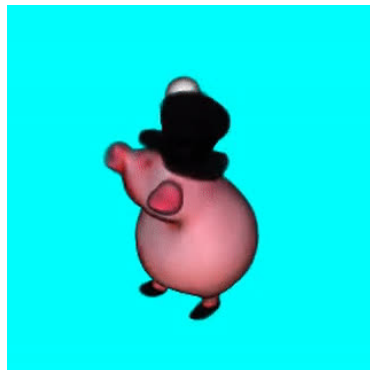
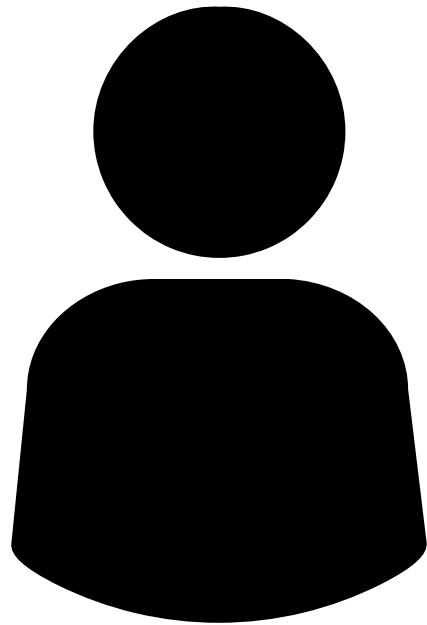
Uses 8+ GPUs

Hmm...
I like the pig,
but not the
clothing.

A pig wearing a
tophat holding
a blue balloon

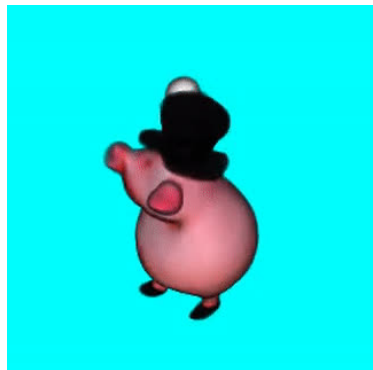
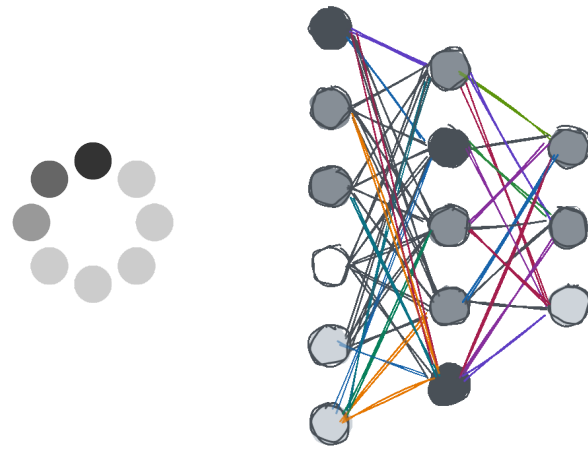


A pig wearing a
red cape sitting
on a chair

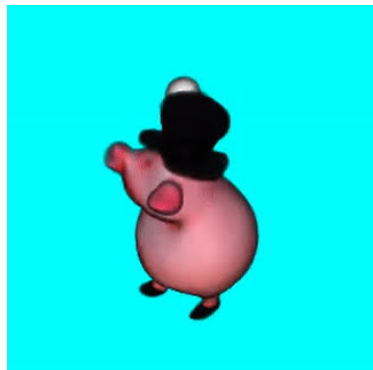
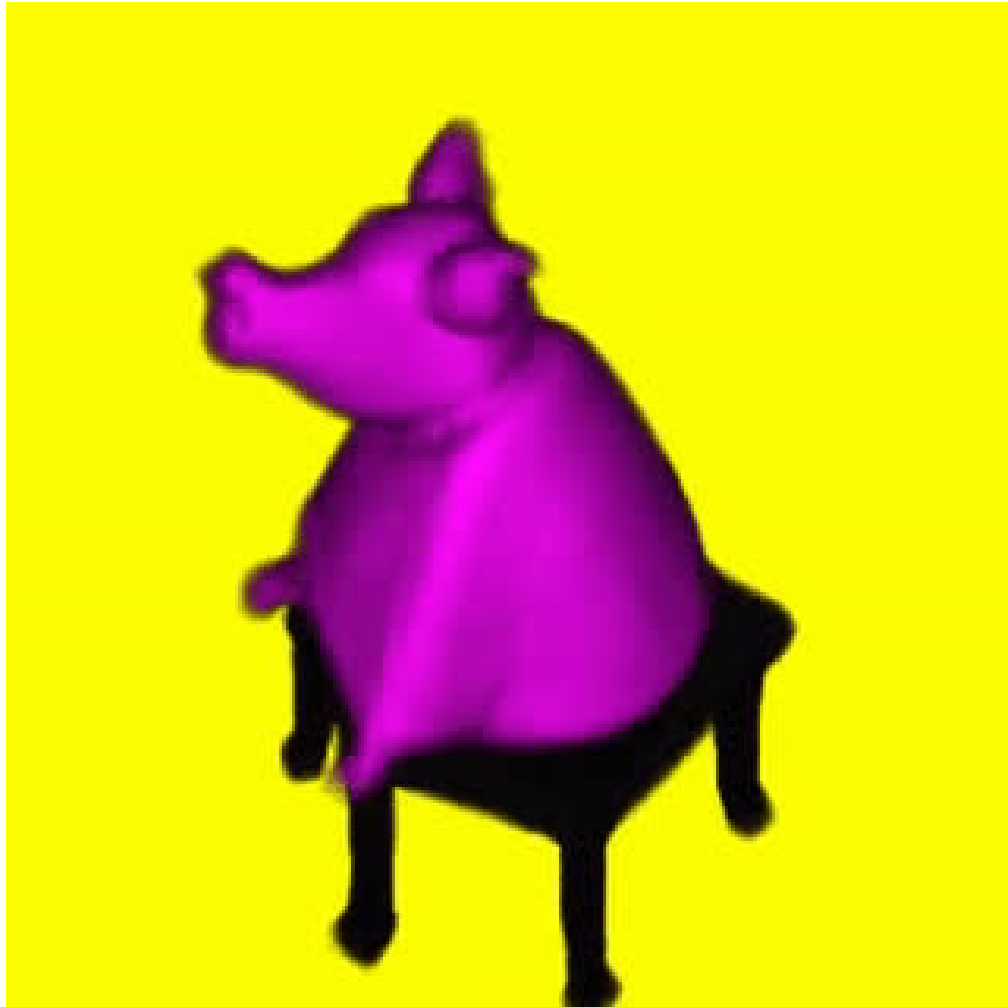
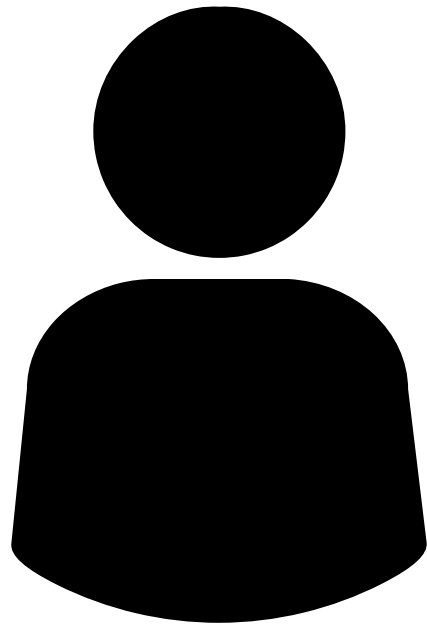


A pig wearing a
red cape sitting
on a chair

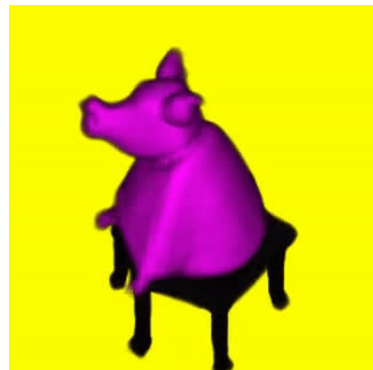
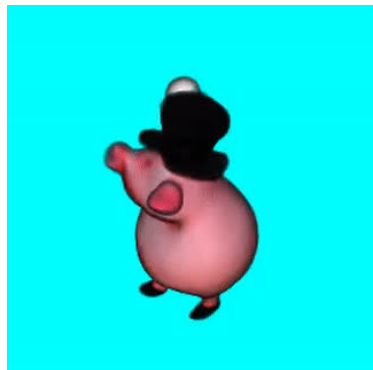
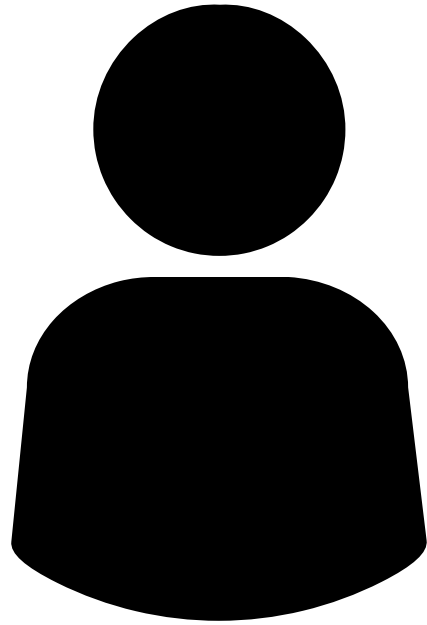
Waiting again...



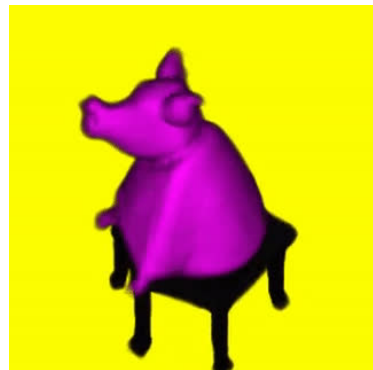
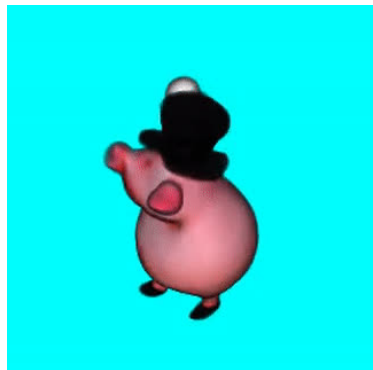
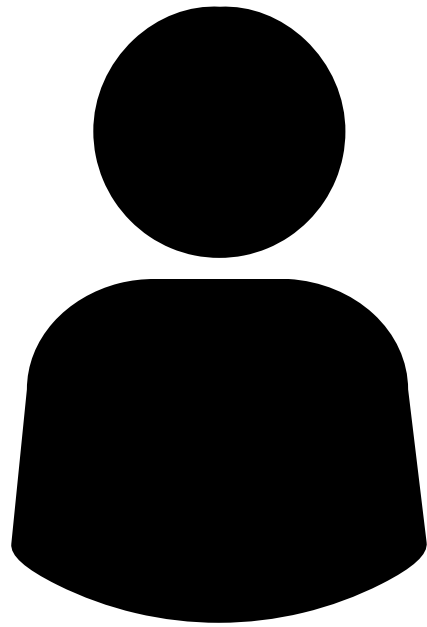
A pig wearing a red cape sitting on a chair



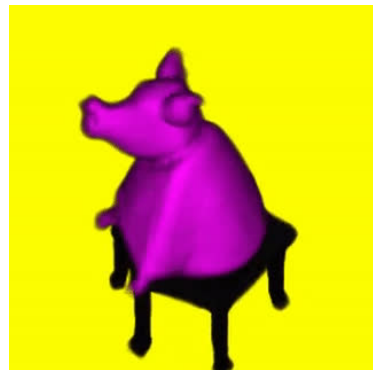
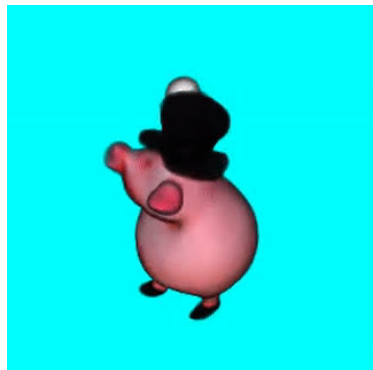
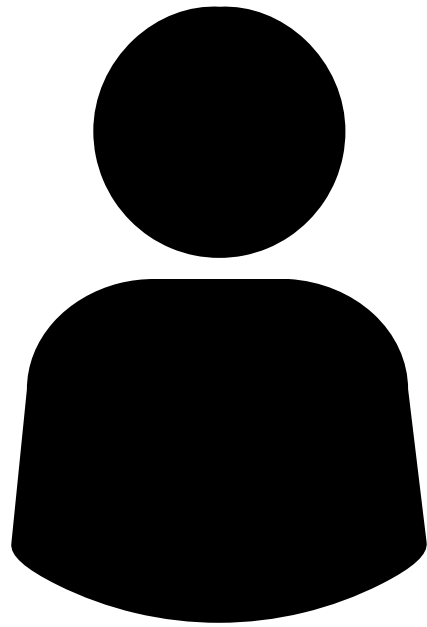
A pig wearing a
tophat and tuxedo
riding a bike



A pig wearing a
tophat and tuxedo
riding a bike

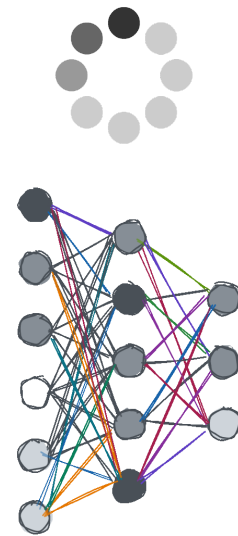


A pig wearing a
tophat and tuxedo
riding a bike

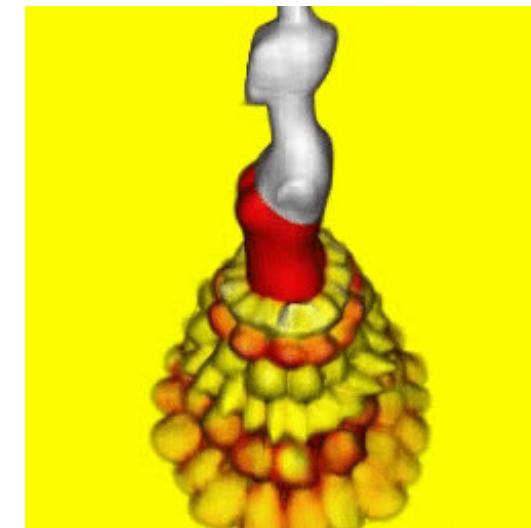
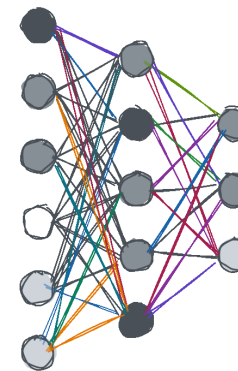


Existing Methods

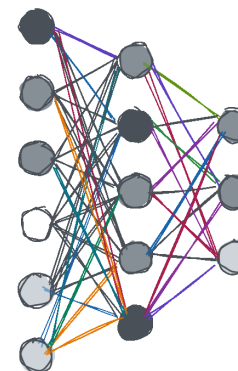
a frog wearing a sweater



a beautiful dress made out of fruit, on a mannequin.



...



...a NeRF for the prompt...

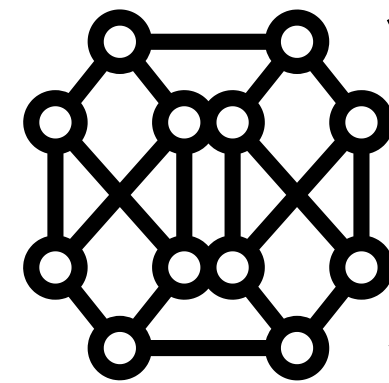
ATT3D: Amortized Text-to-3D

a frog wearing a sweater

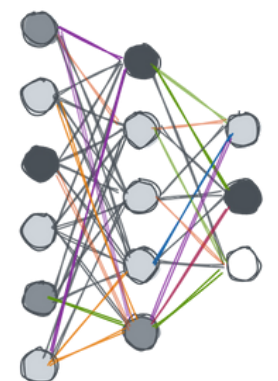
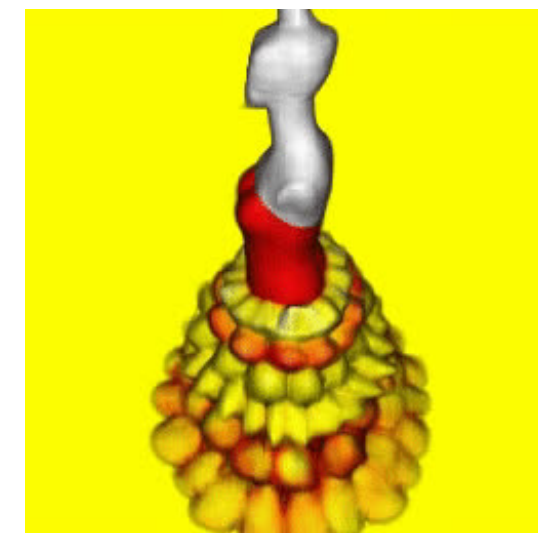
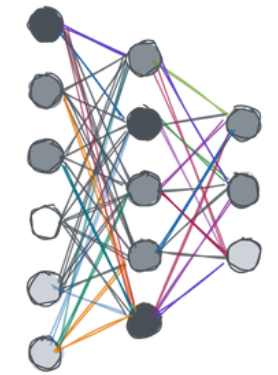
a beautiful dress made out of fruit, on a mannequin.

...

1 Mapping net:
Text \rightarrow NeRF



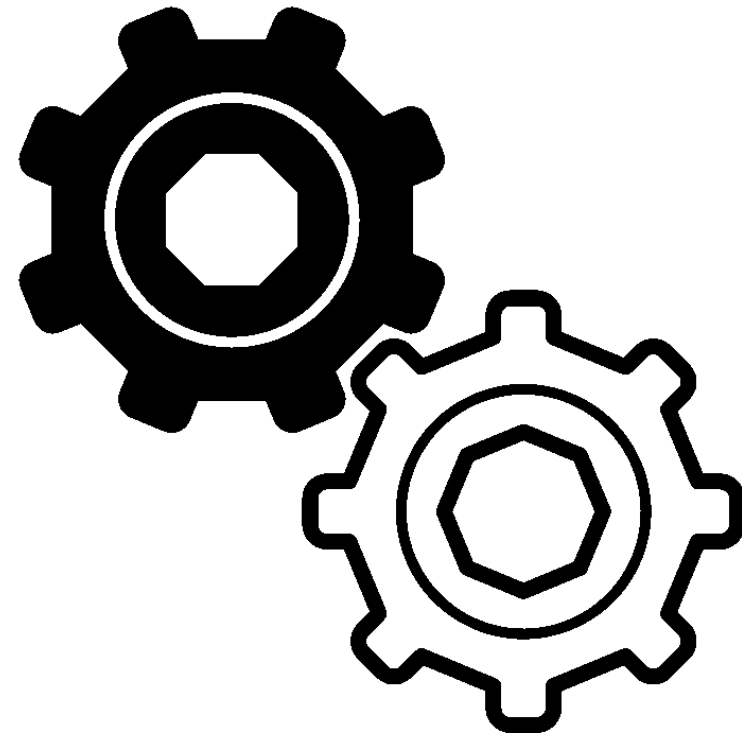
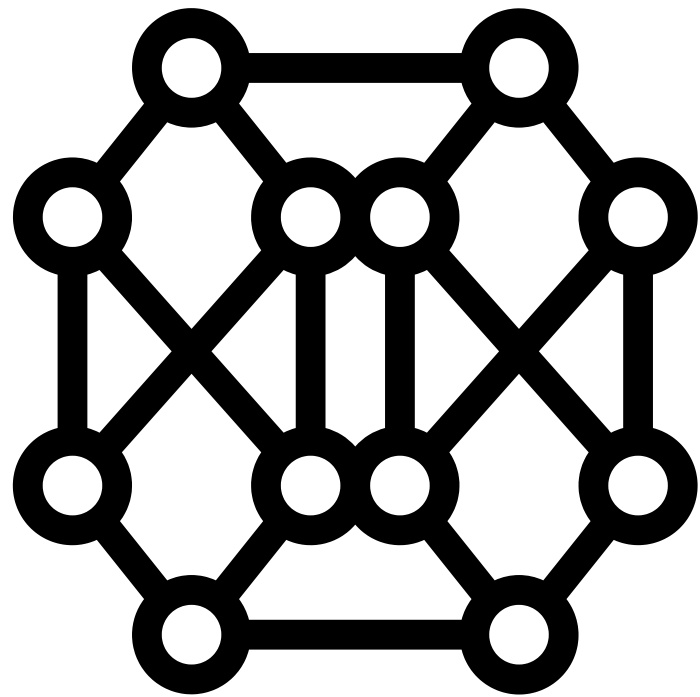
Trained offline



...a NeRF for the prompt...

ATT3D: Amortized Text-to-3D

Training happens offline



A peacock riding a surfboard

An orangutan eating icecream

A corgi wearing a top hat

A bear wearing a sweater vest

A frog sitting on a lily pad

A car driving through a river

A humanoid playing chess

A castle made out of cake

A ceramic ornamental flower

A chimpanzee making a clay bowl

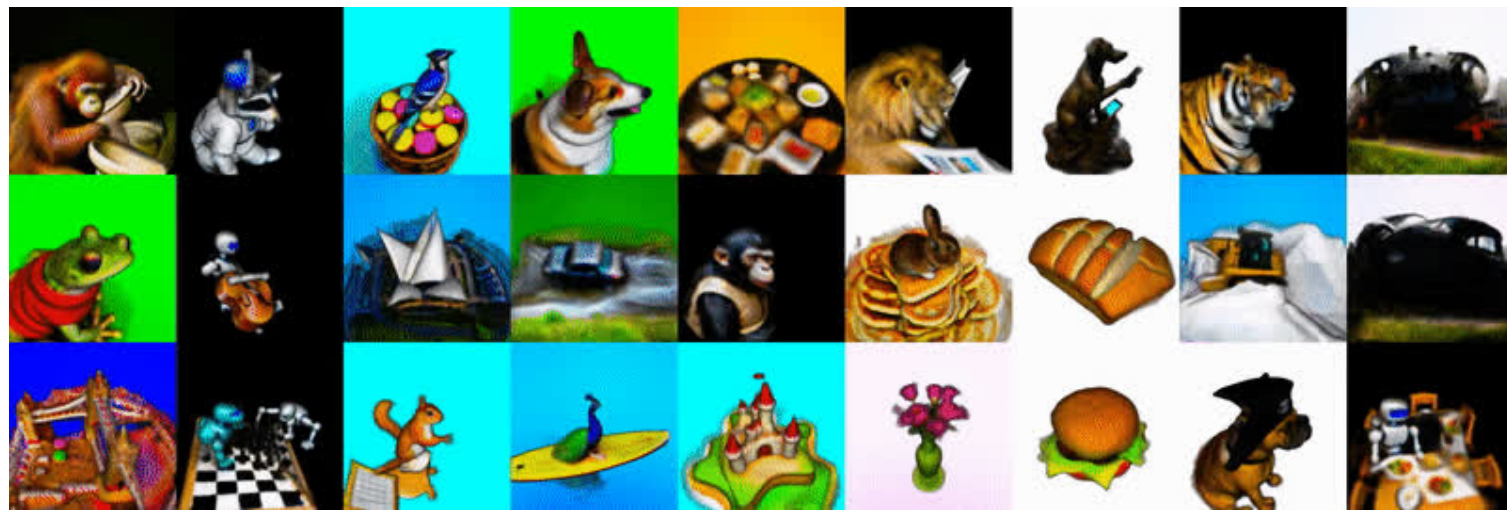
A tall skyscraper with a viewing deck

• • •

ATT3D: Amortized Text-to-3D

Benefits

Reduction in training time

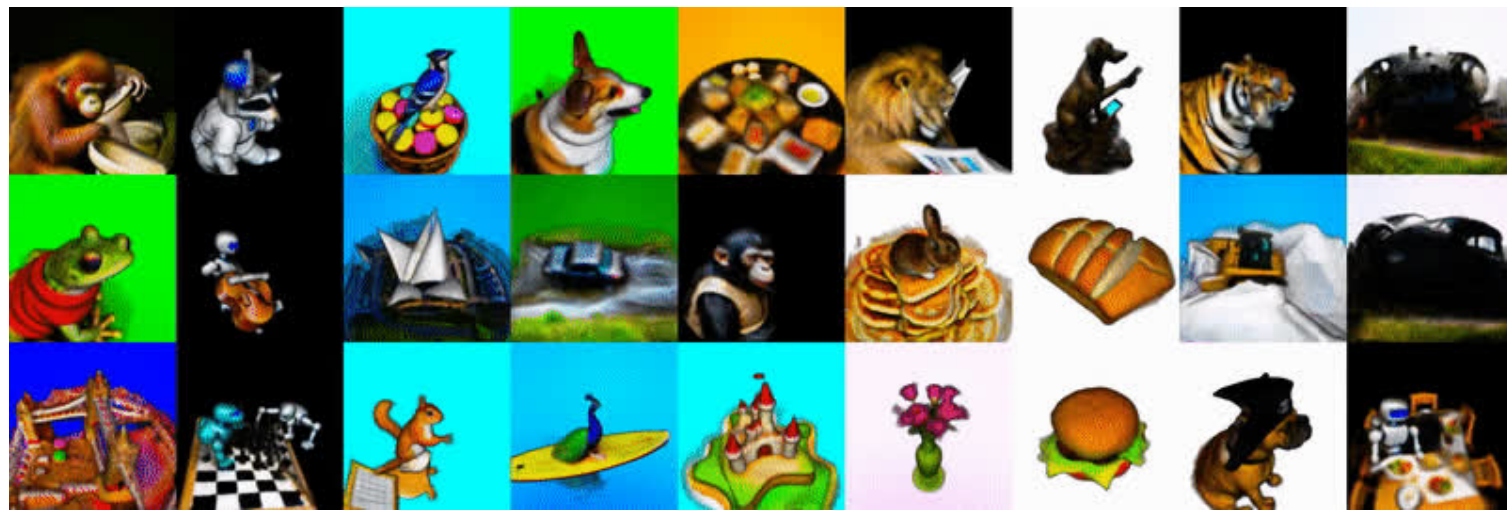


Successfully trained on 2400 prompts

ATT3D: Amortized Text-to-3D

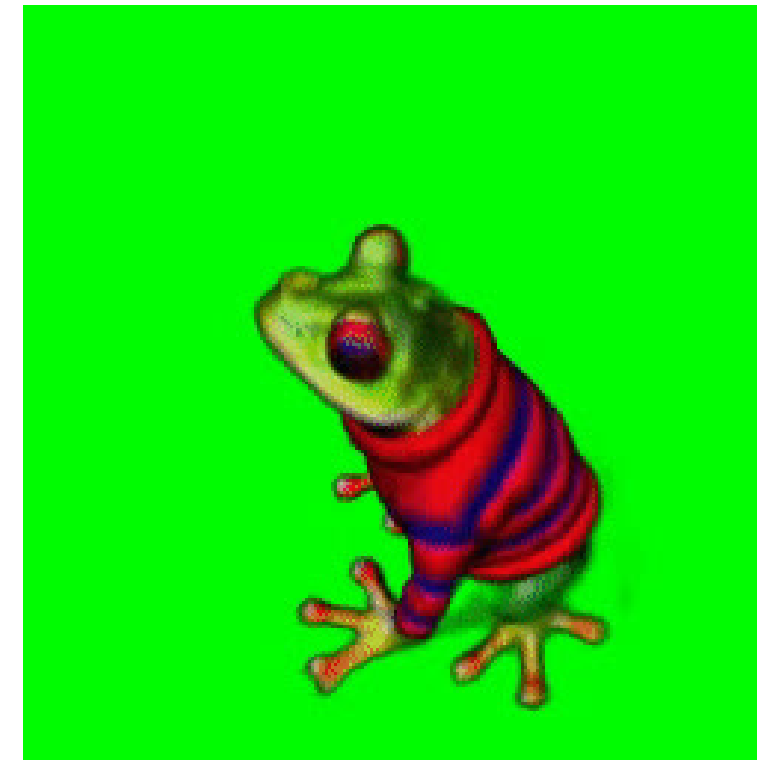
Benefits

Reduction in training time

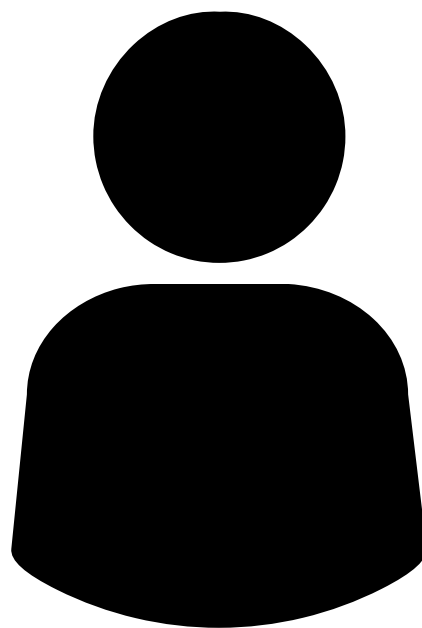


Successfully trained on 2400 prompts

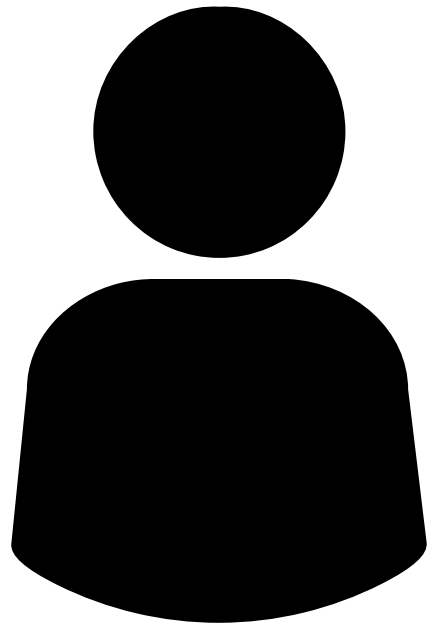
Mixing prompts for free



A pig wearing a
tophat holding
a blue balloon



A pig wearing a
tophat holding
a blue balloon

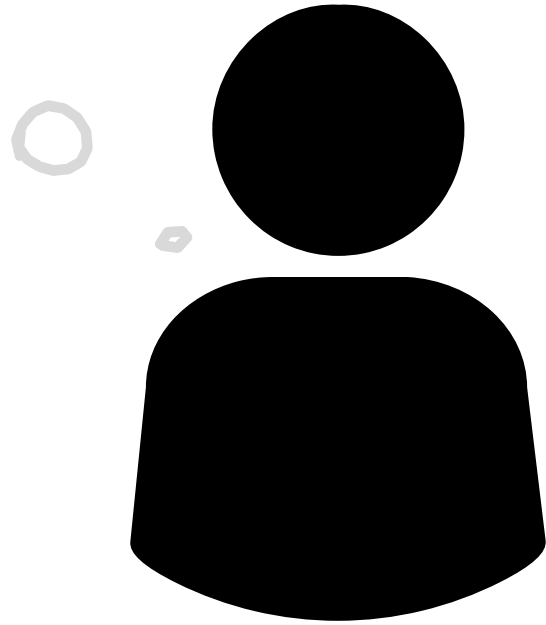


Results in <1 second



Hmm...
I like the pig,
but not the
clothing.

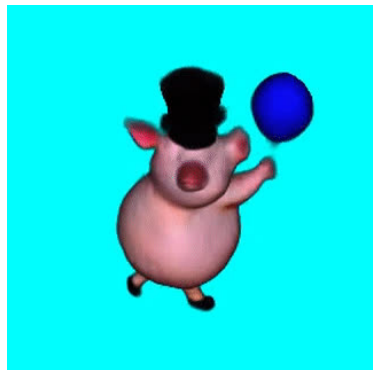
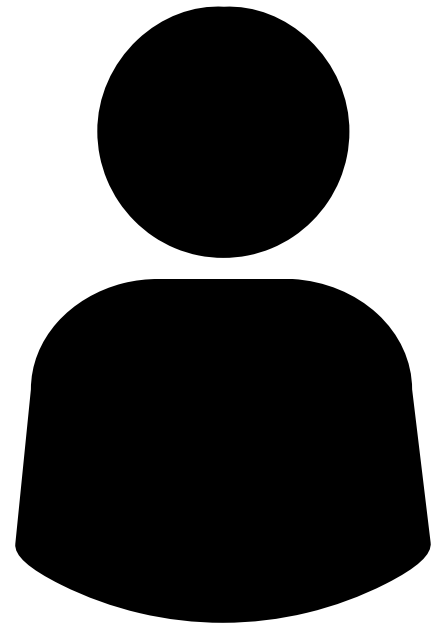
A pig wearing a
tophat holding
a blue balloon



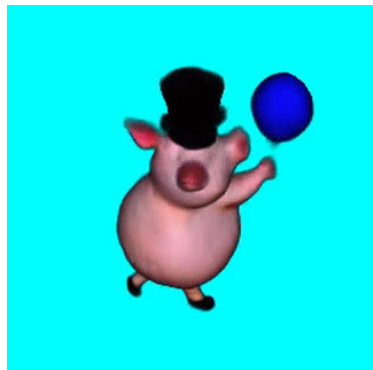
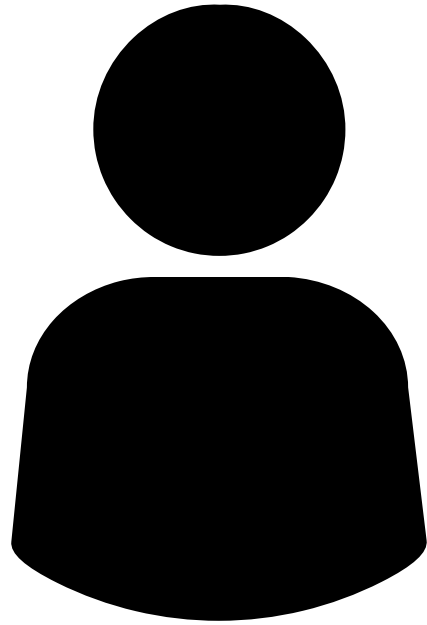
Uses only a single GPU



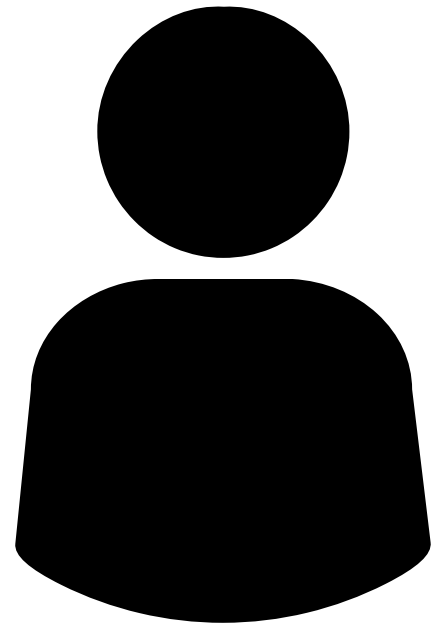
A pig wearing a
red cape sitting
on a chair



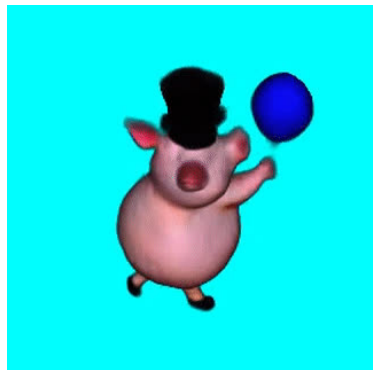
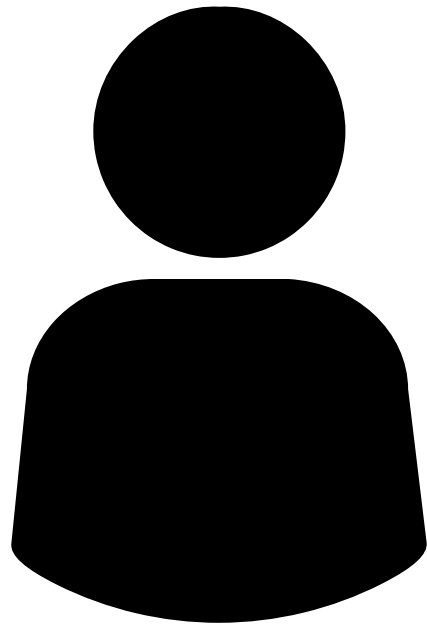
A pig wearing a red cape sitting on a chair



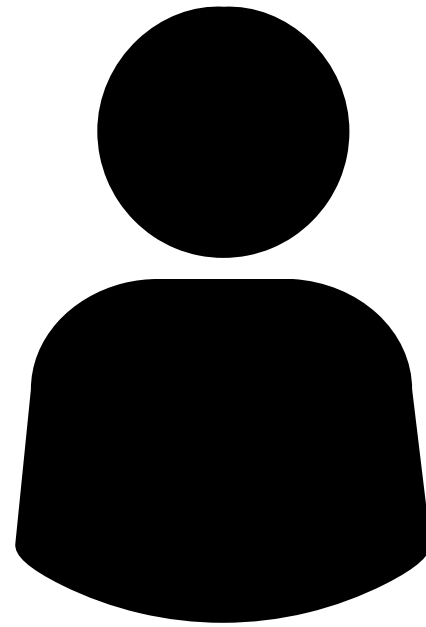
A pig wearing a
tophat and tuxedo
riding a bike

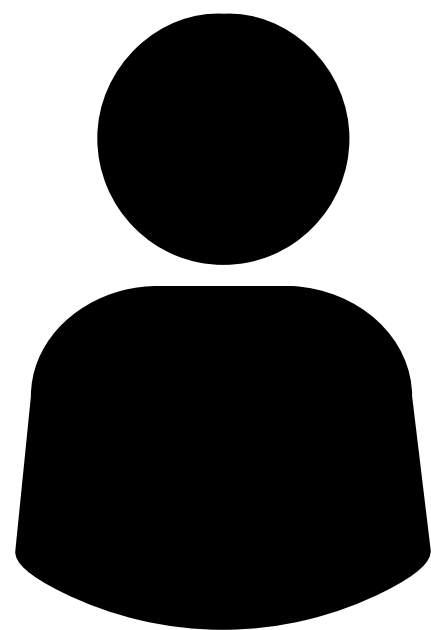


A pig wearing a
tophat and tuxedo
riding a bike



Can use Magic3D's finetuning





Capabilities of ATT3D

Generate novel combinations

We can train on combinatorial prompt sets

A **pig** wearing a **helmet**, wearing a **cape**, **riding a bike**

Capabilities of ATT3D

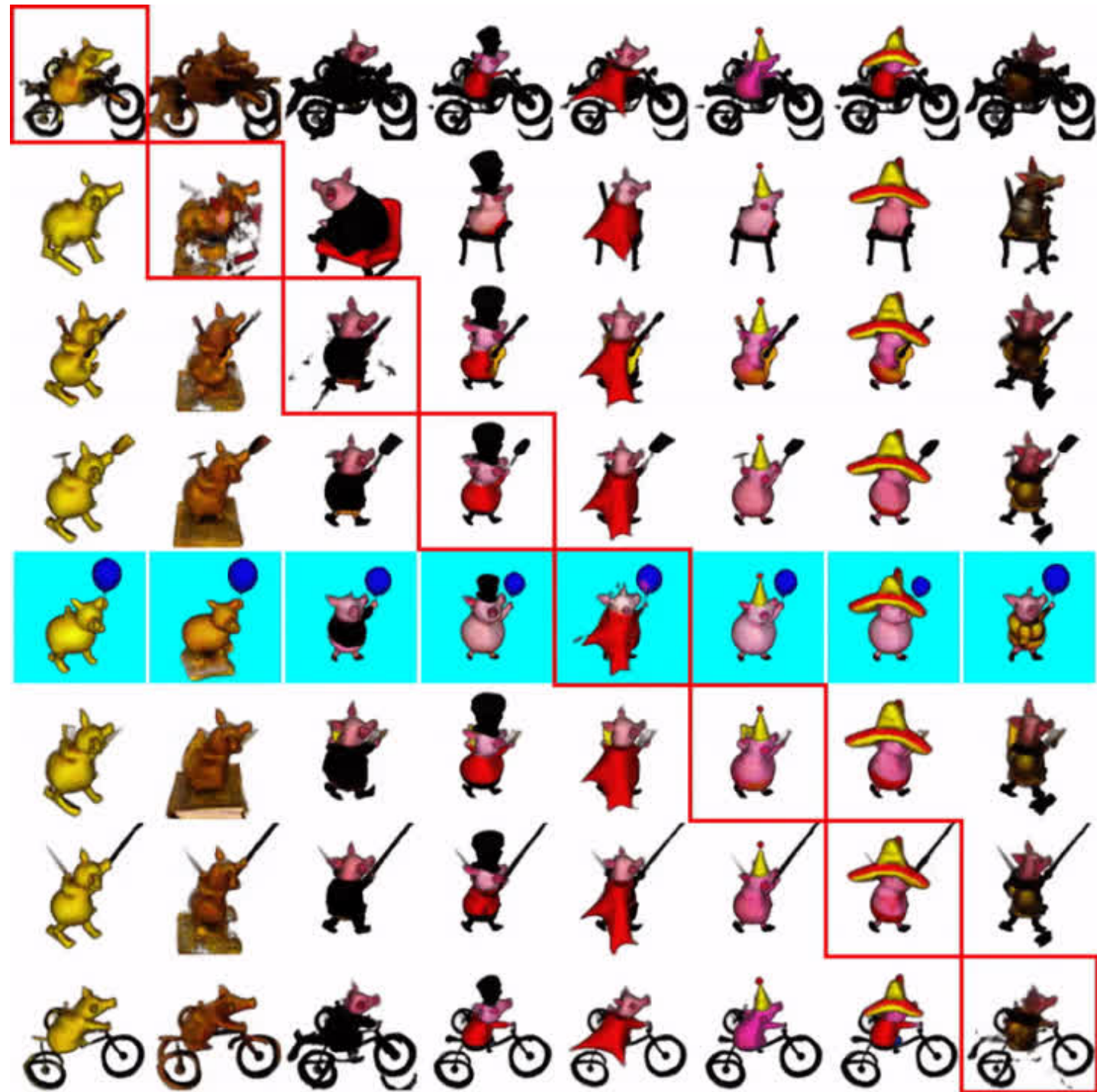
Generate novel combinations

We can train on combinatorial prompt sets

A **lion** wearing a **tophat**, wearing a **backpack**, holding a **balloon**

Capabilities of ATT3D

Generate novel combinations



ATT3D

Capabilities of ATT3D

Generate novel combinations



ATT3D



Prior method

Capabilities of ATT3D

Improved consistency with prompt



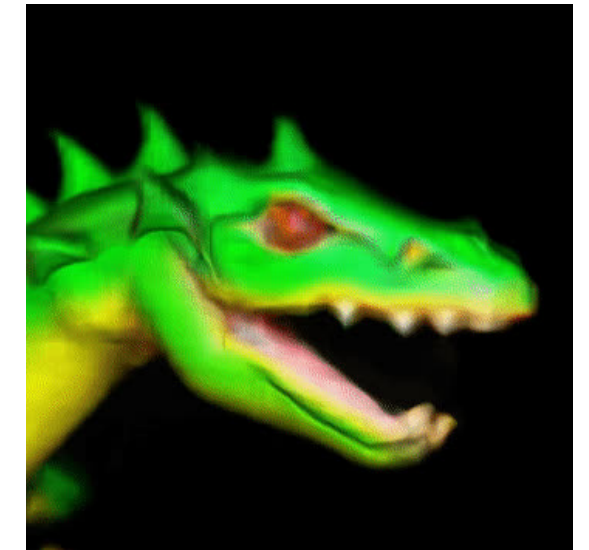
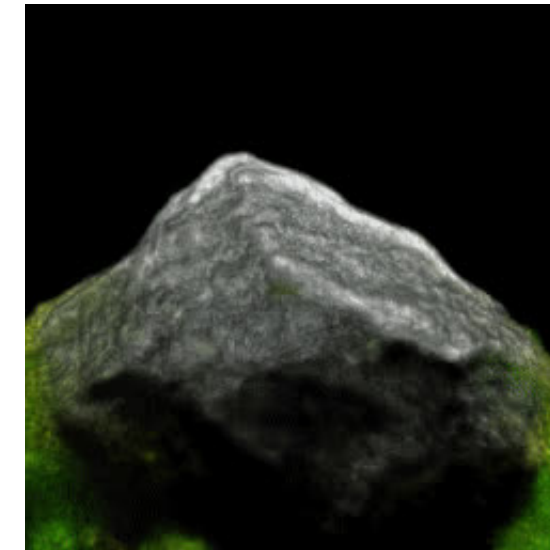
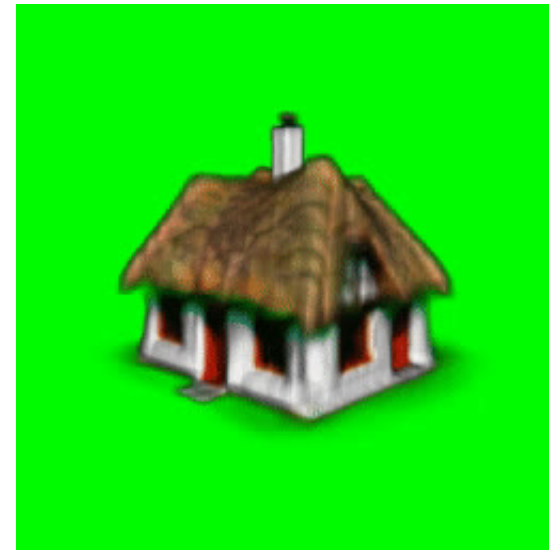
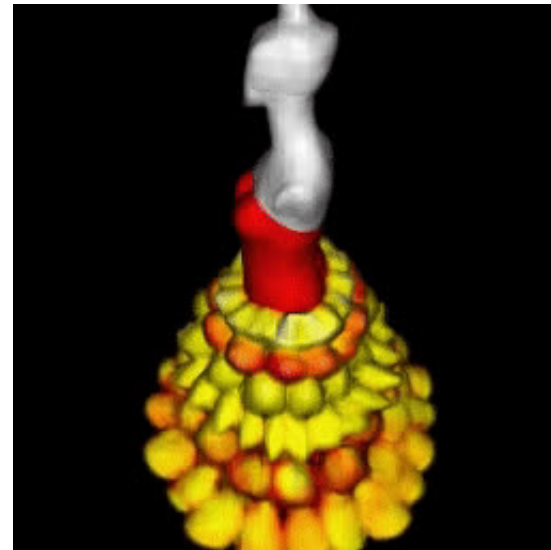
ATT3D



Prior method

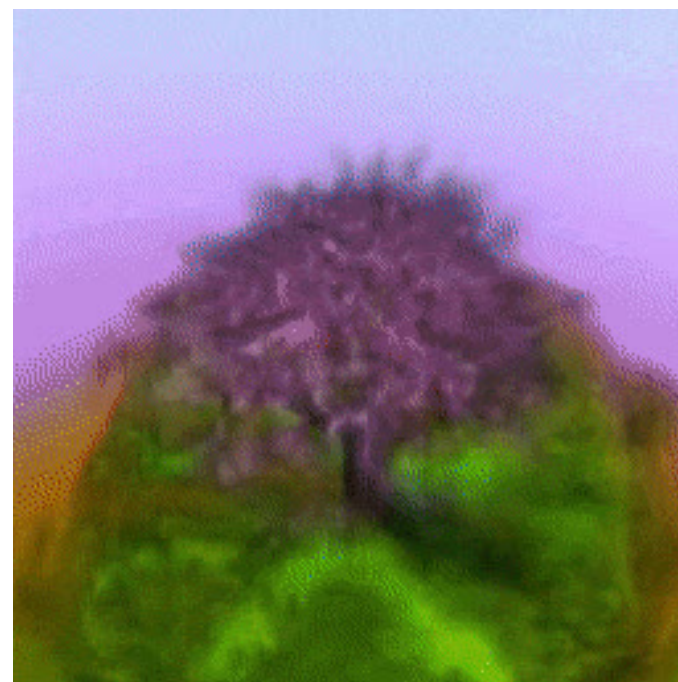
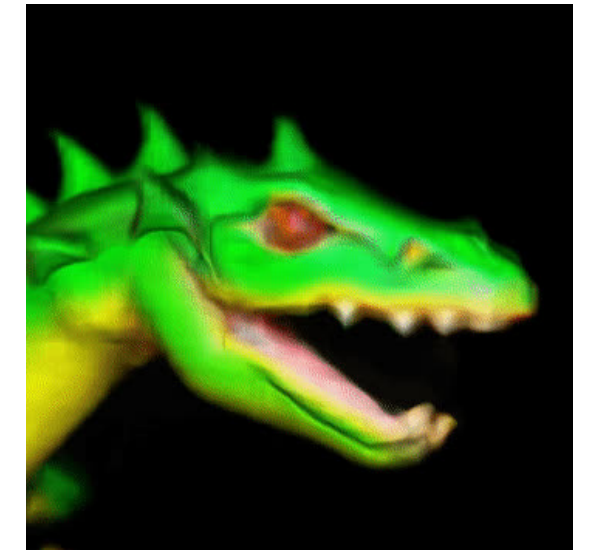
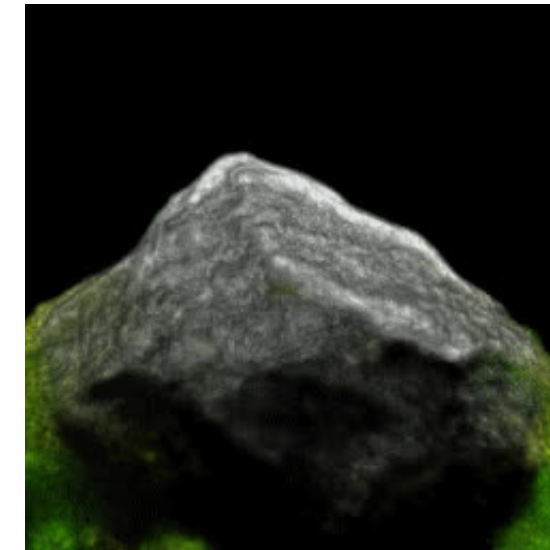
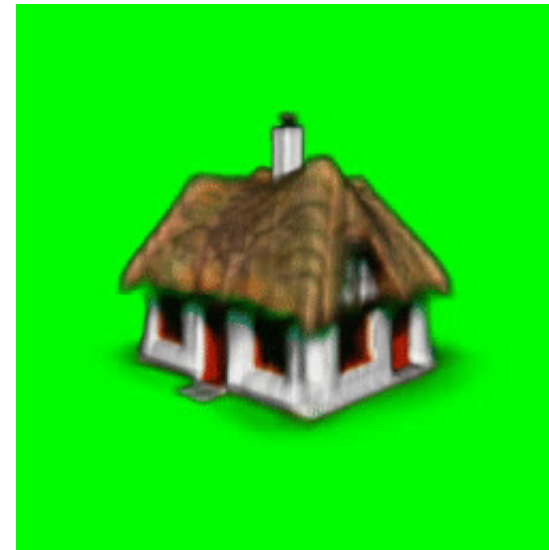
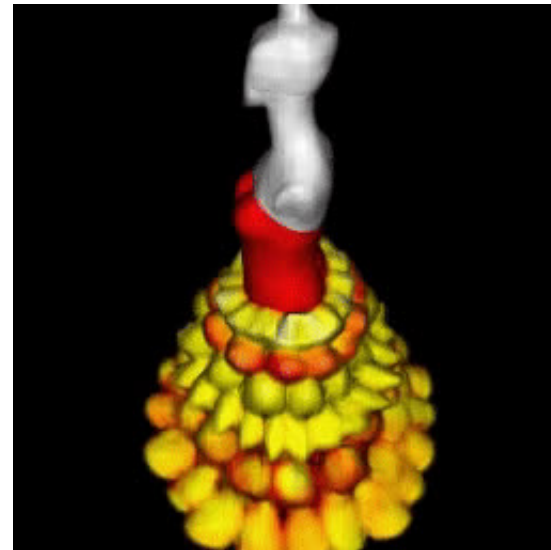
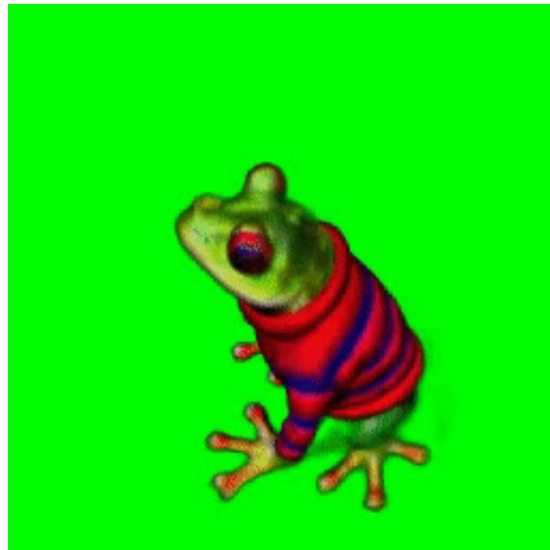
Capabilities of ATT3D

Smooth mixing of prompts

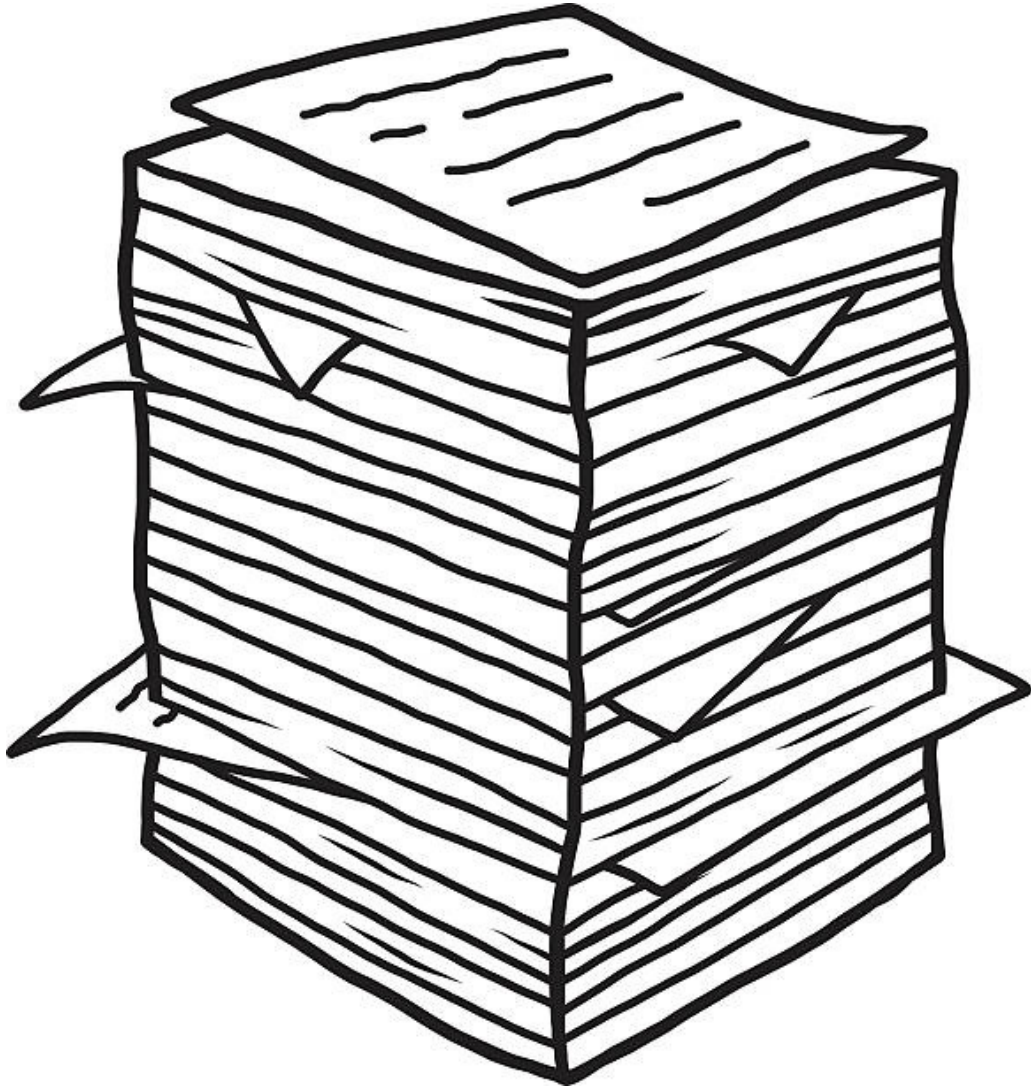


Capabilities of ATT3D

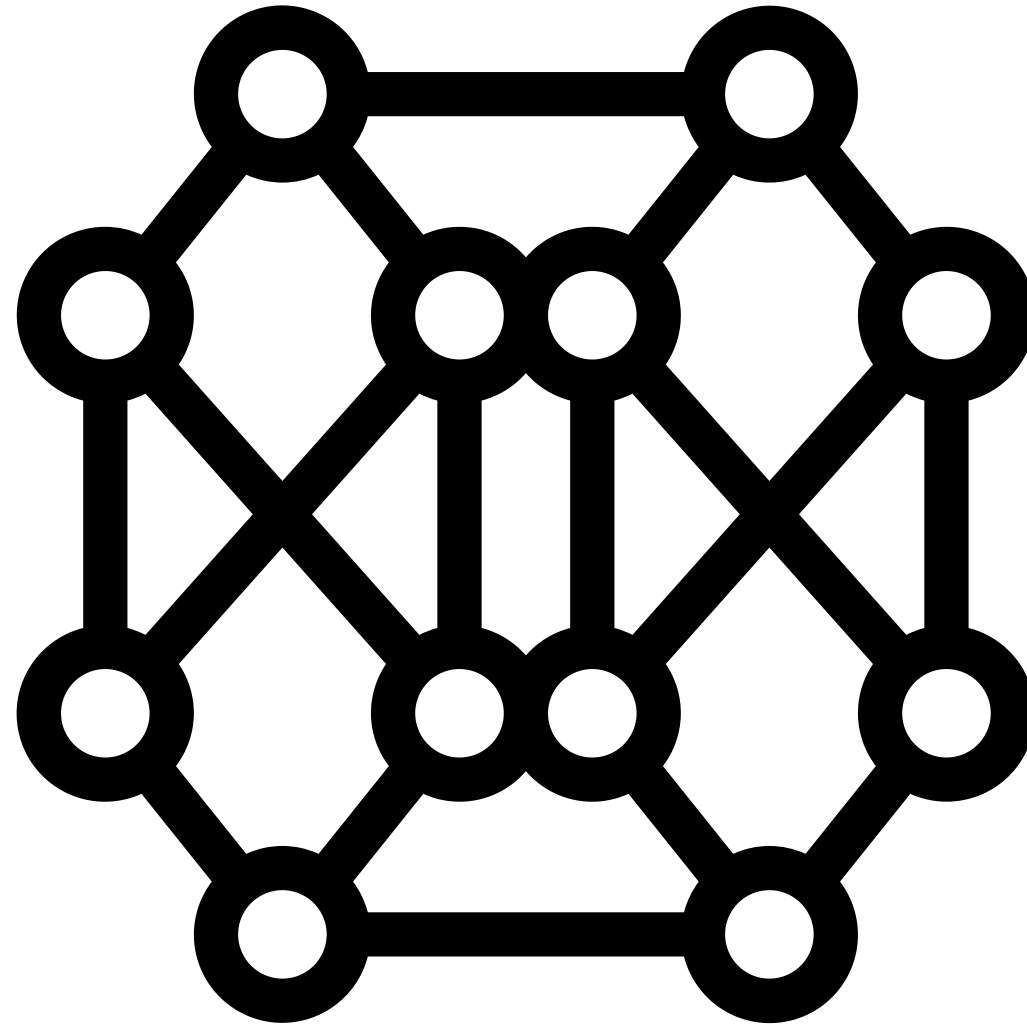
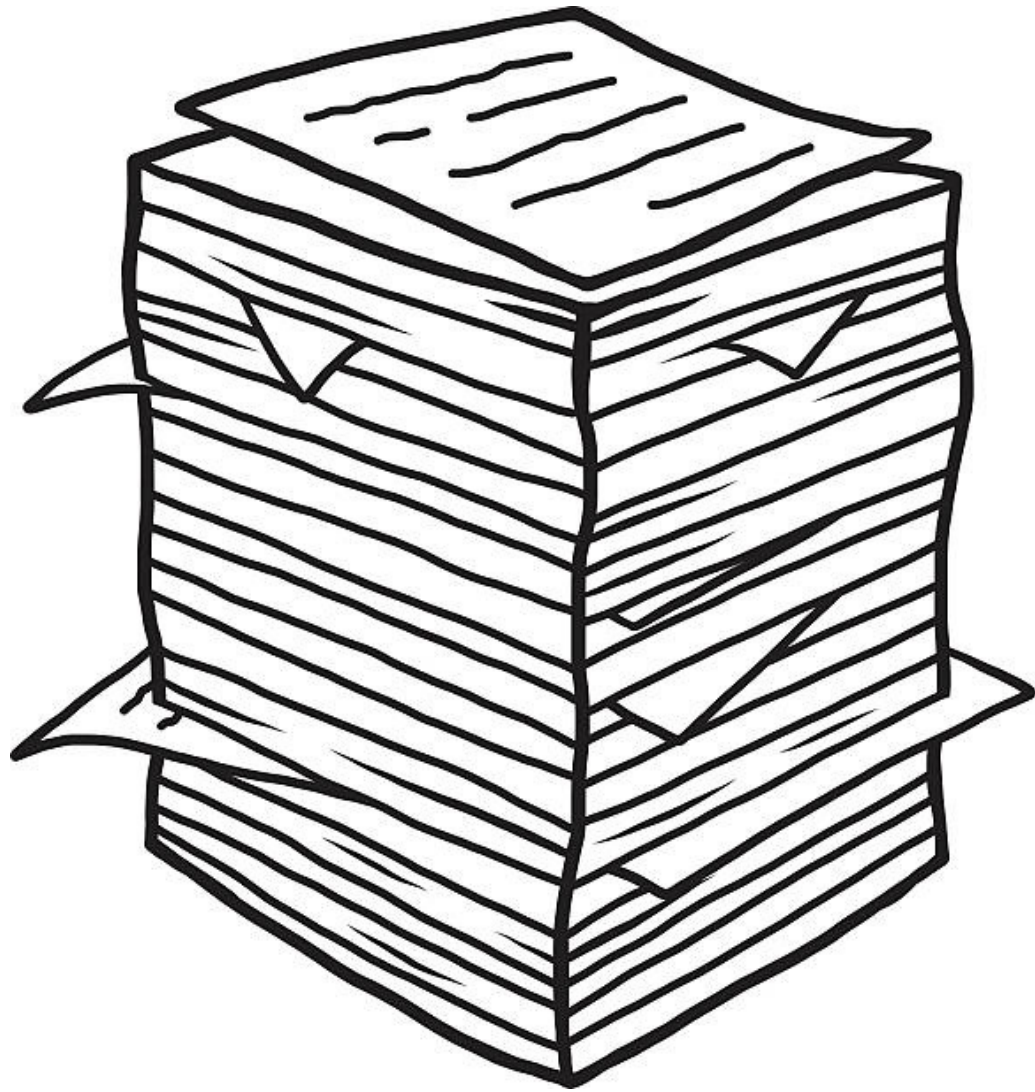
Smooth mixing of prompts



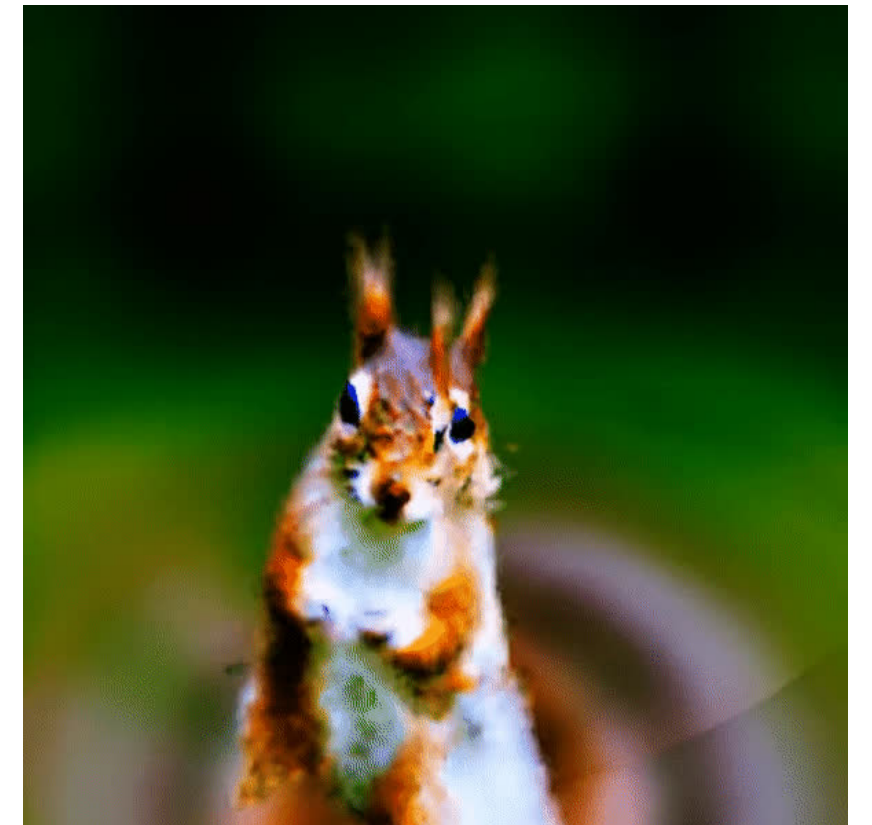
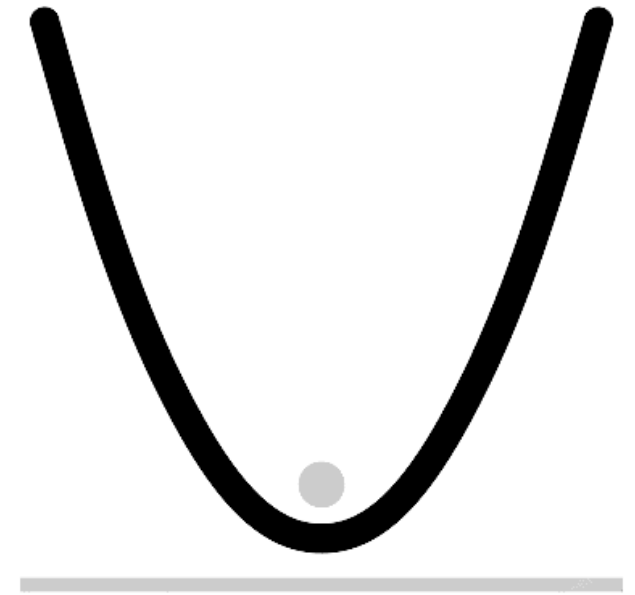
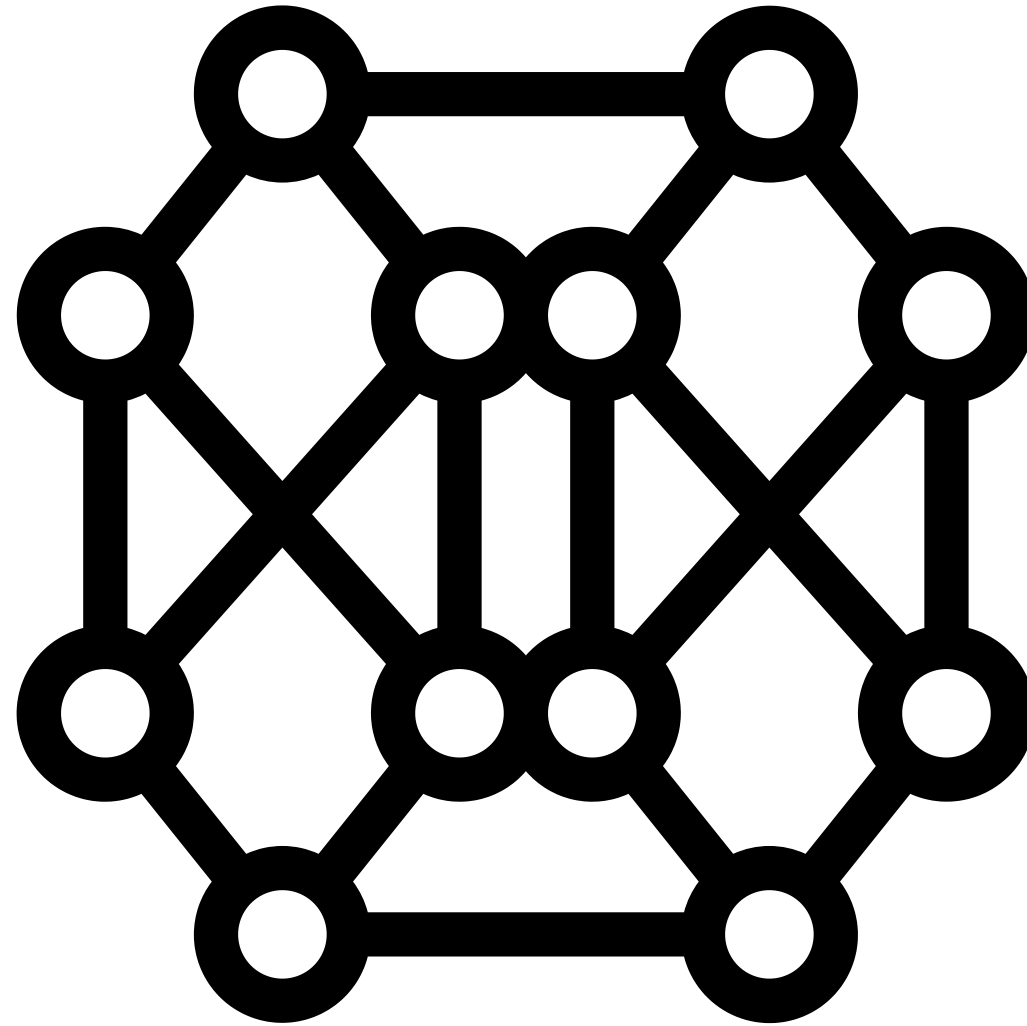
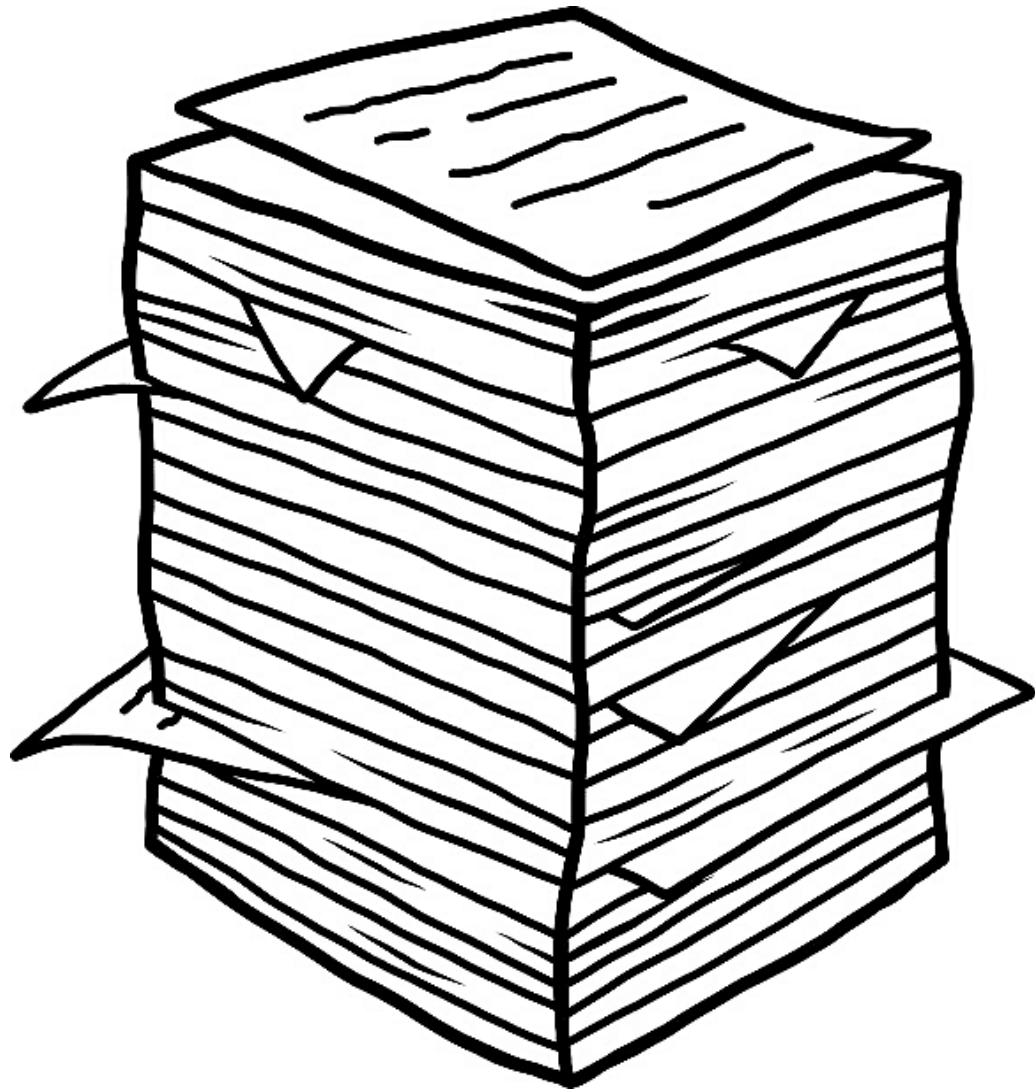
Limitations and next steps



Limitations and next steps



Limitations and next steps





ATT3D: Amortized Text-to-3D

Jonathan Lorraine, Kevin Xie, Xiaohui Zeng, Chen-Hsuan Lin, Towaki Takikawa,
Nicholas Sharp, Tsung-Yi Lin, Ming-Yu Liu, Sanja Fidler, James Lucas

Thanks for watching!