Generative Adversarial Transformers

Drew A. Hudson & Larry Zitnick

ICML 2021

facebook AI

Stanford

*Special Thanks to Christopher D. Manning 🌻

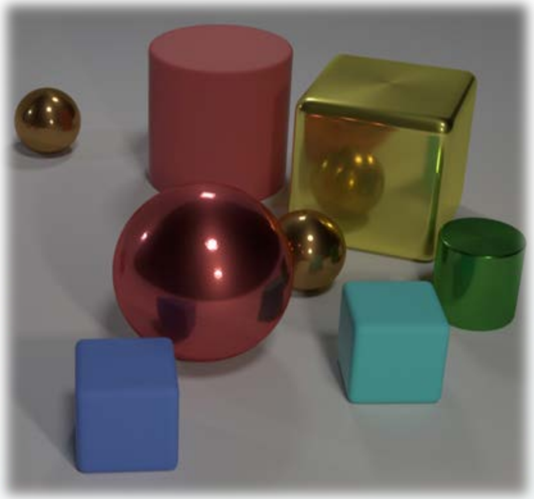
Most GANs use Convolution

fake

Convolutional GANs struggle with long-range dependencies

Consistency

Global Structure



Compositionality

Transformers

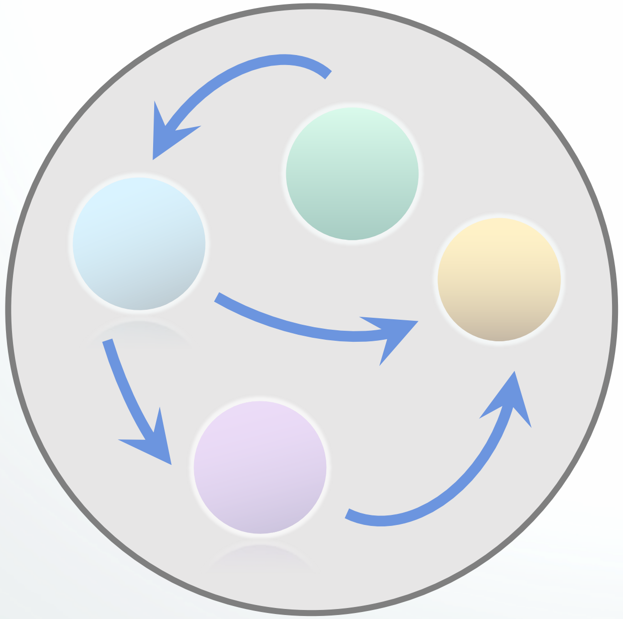
$O(n^2)$

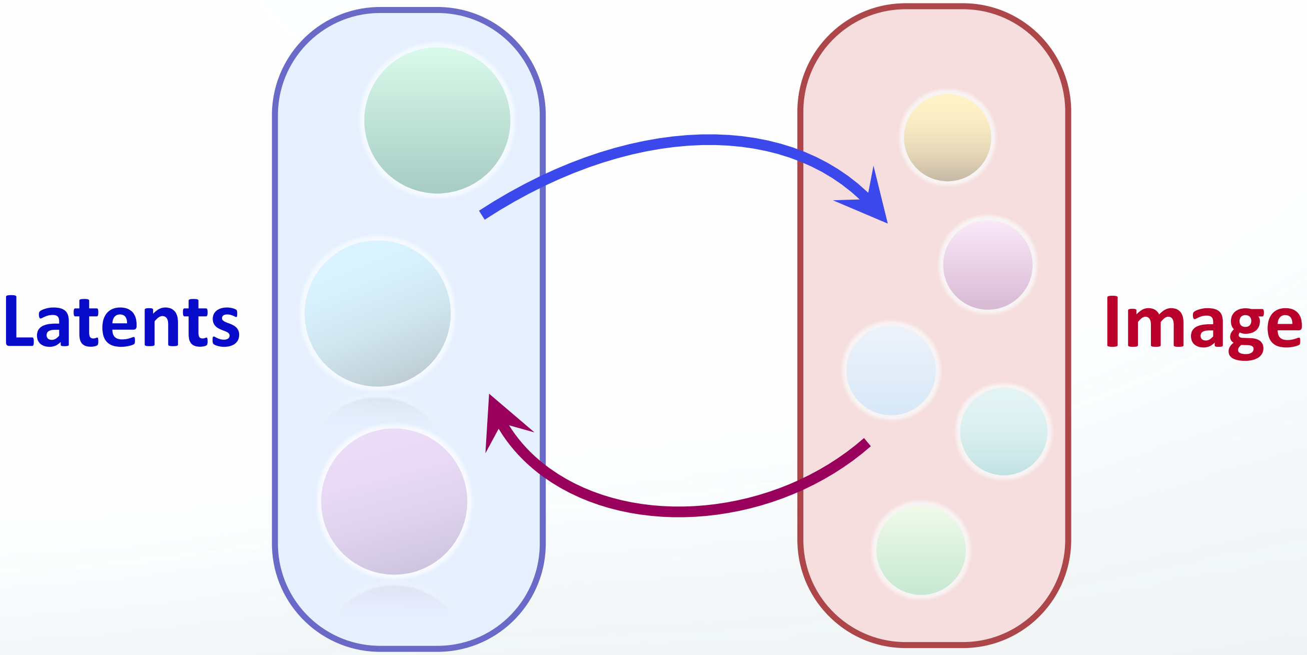
use Self-Attention to model long-range interactions

but they don't scale

Can we efficiently apply transformers for image generation?

Generative Adversarial Transformers

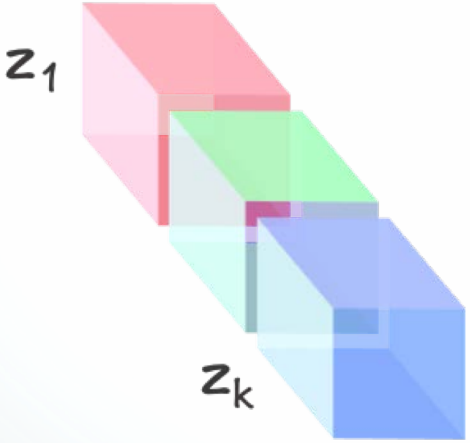




Self-Attention

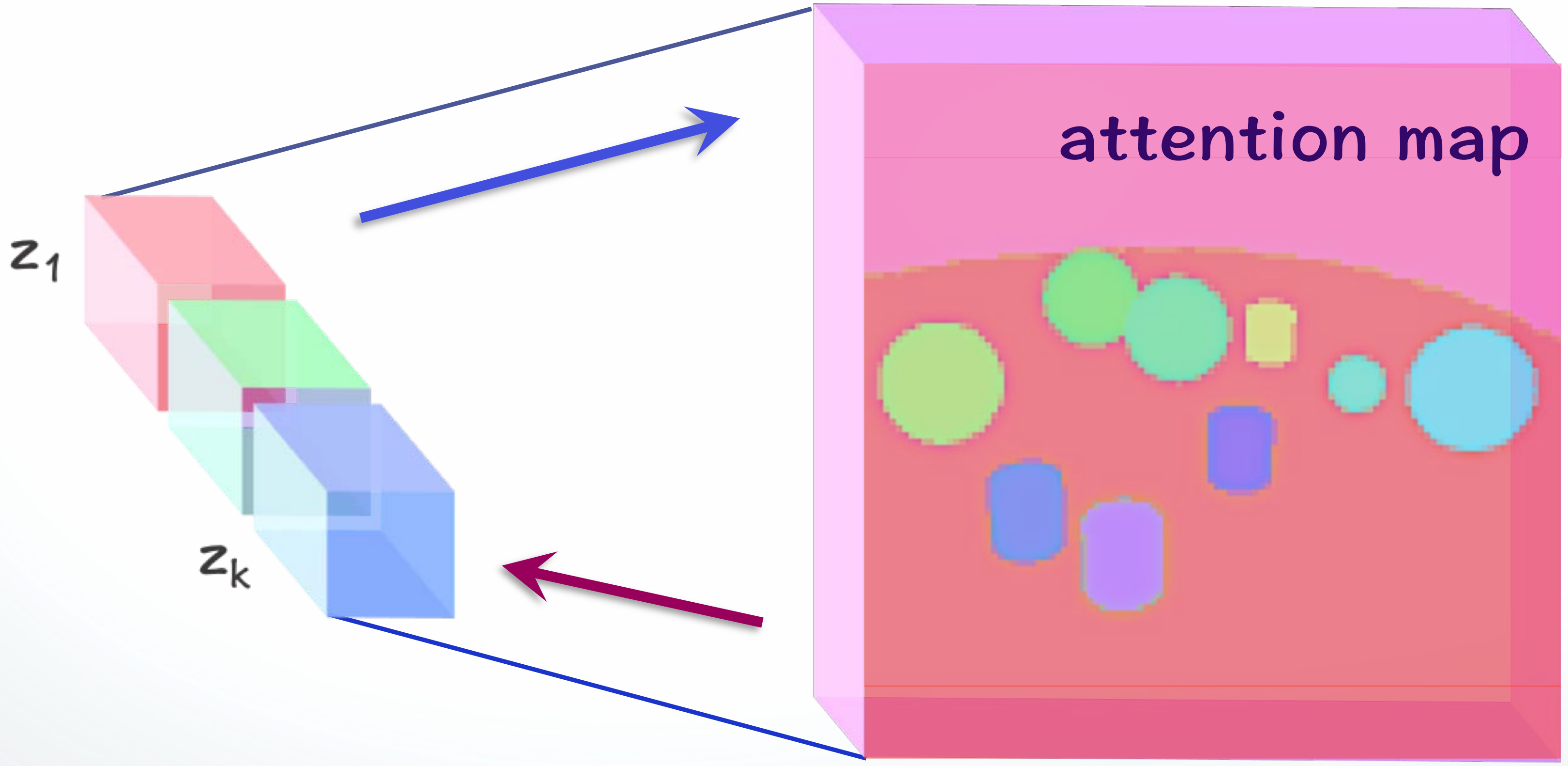
Bipartite Attention

From Transformer to GANformer



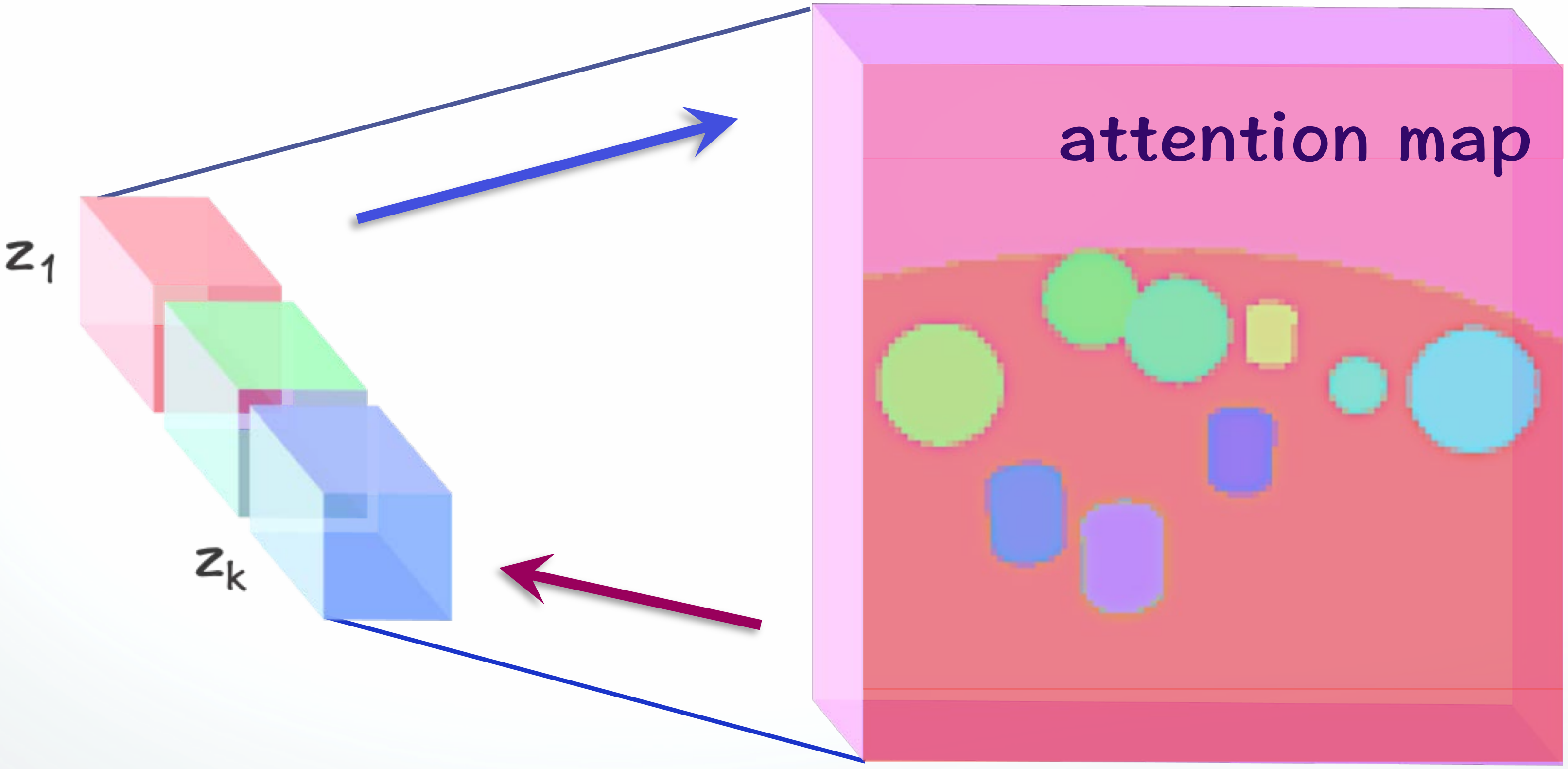
Compositional latent space with multiple variables

From Transformer to GANformer



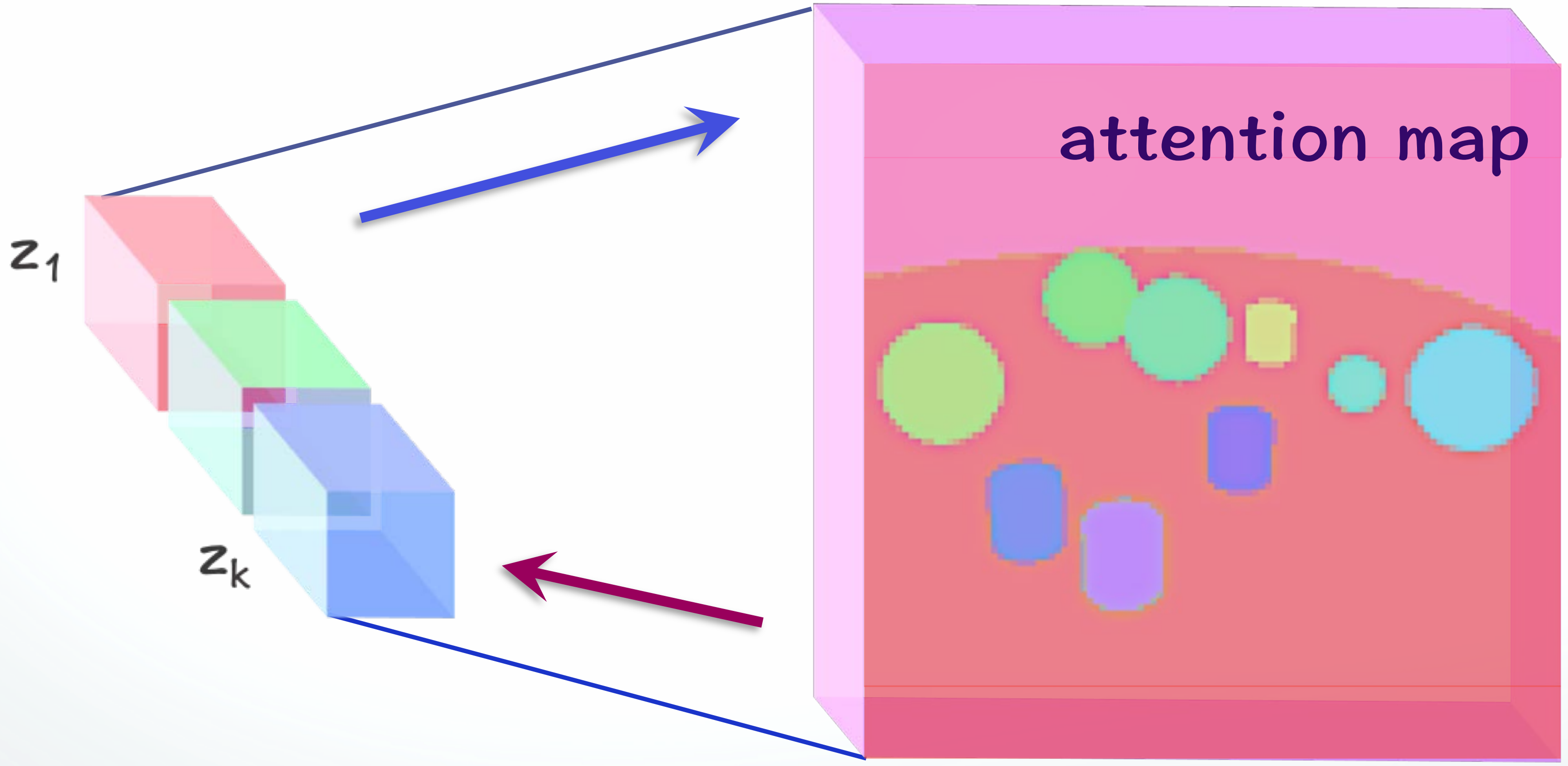
Compositional latent space with multiple variables

From Transformer to GANformer



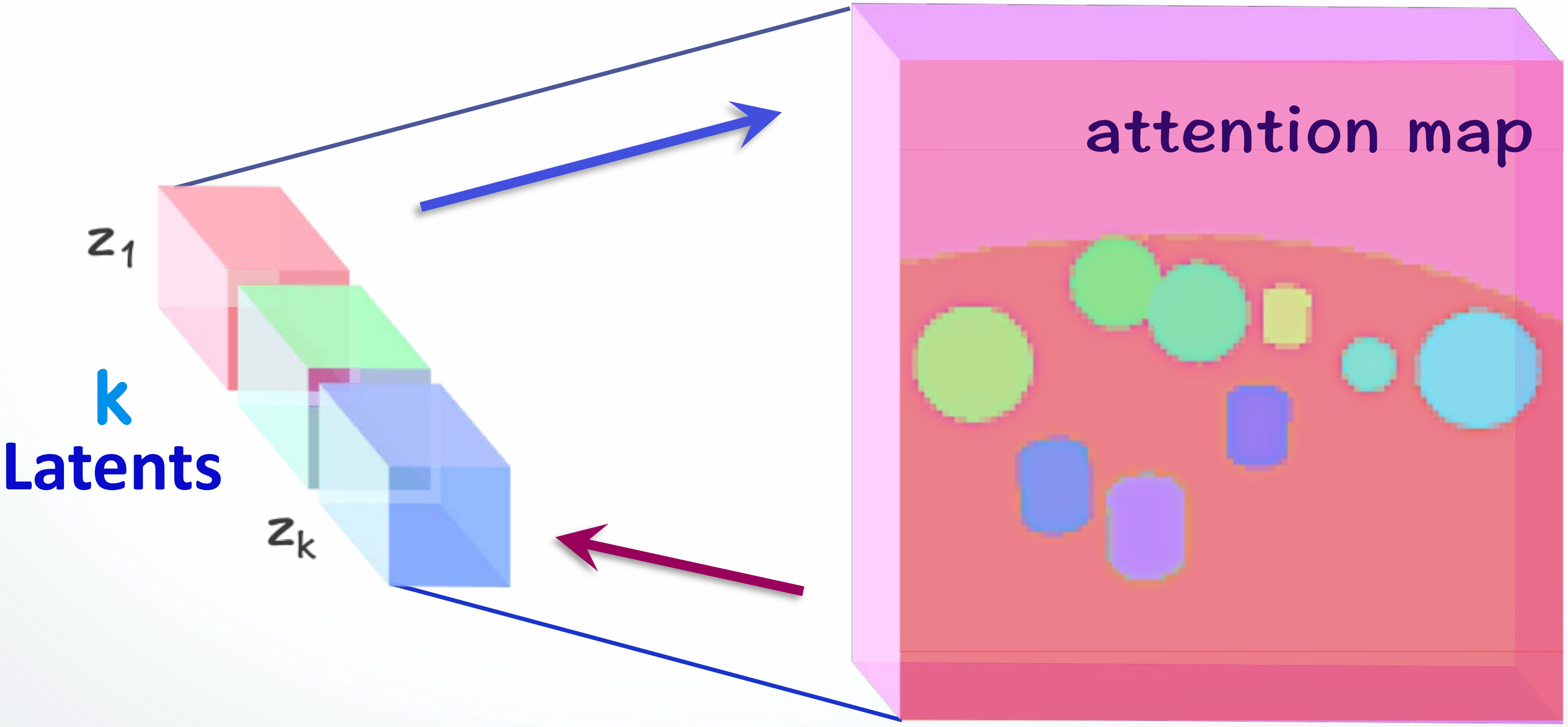
Bidirectional interaction between the latents and the image enables bottom-up & top-down processing.

From Transformer to GANformer



The latents guide the image synthesis cooperatively, attending and modulating different objects or entities.

From Transformer to GANformer



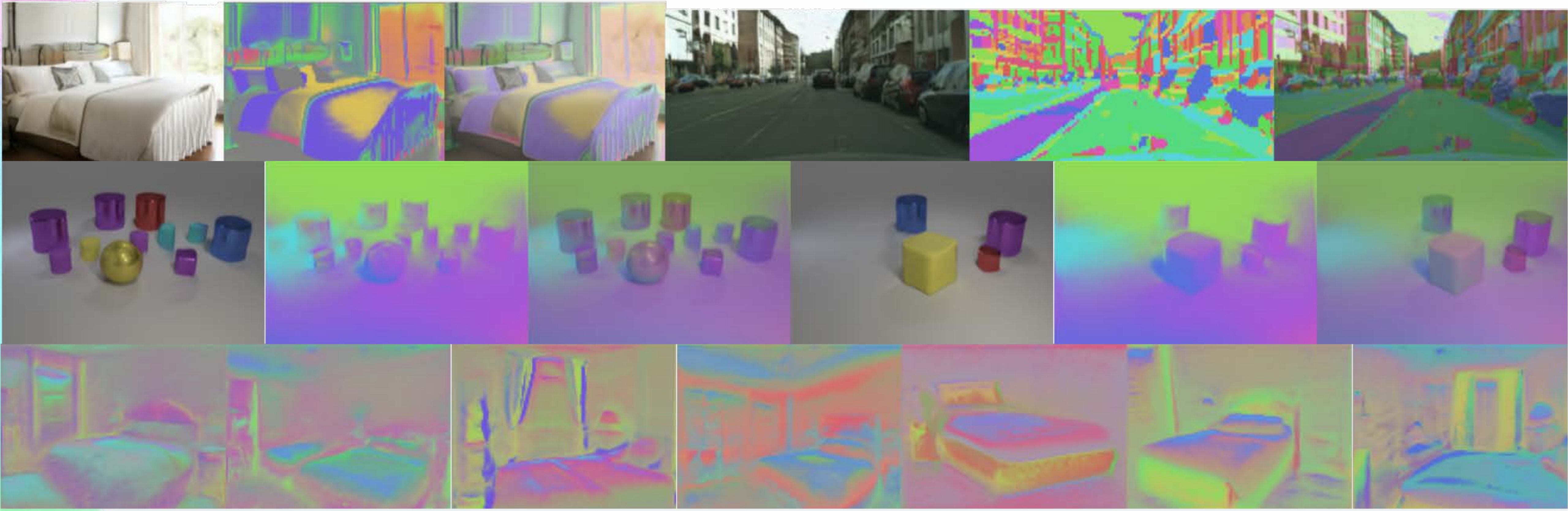
$O(kn)$

n
Image

Has linear efficiency that scales to high-resolutions while capturing long-range dependencies.

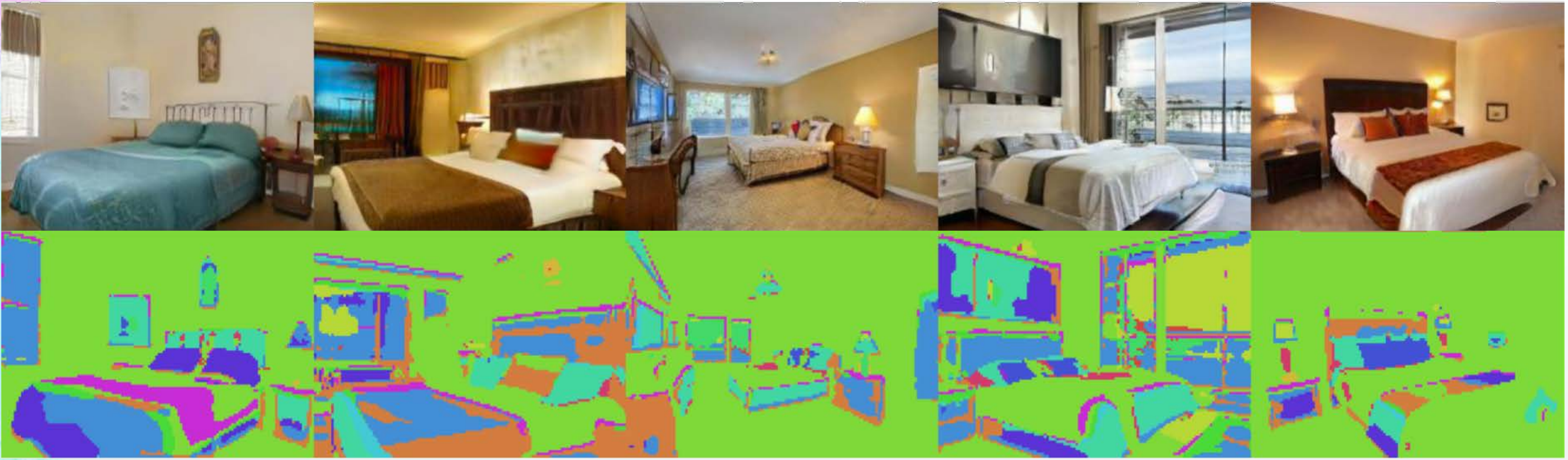
GANformer's Attention Maps

The **latents attend** to **semantic entities** and **cooperatively** generate **compositional** scenes.



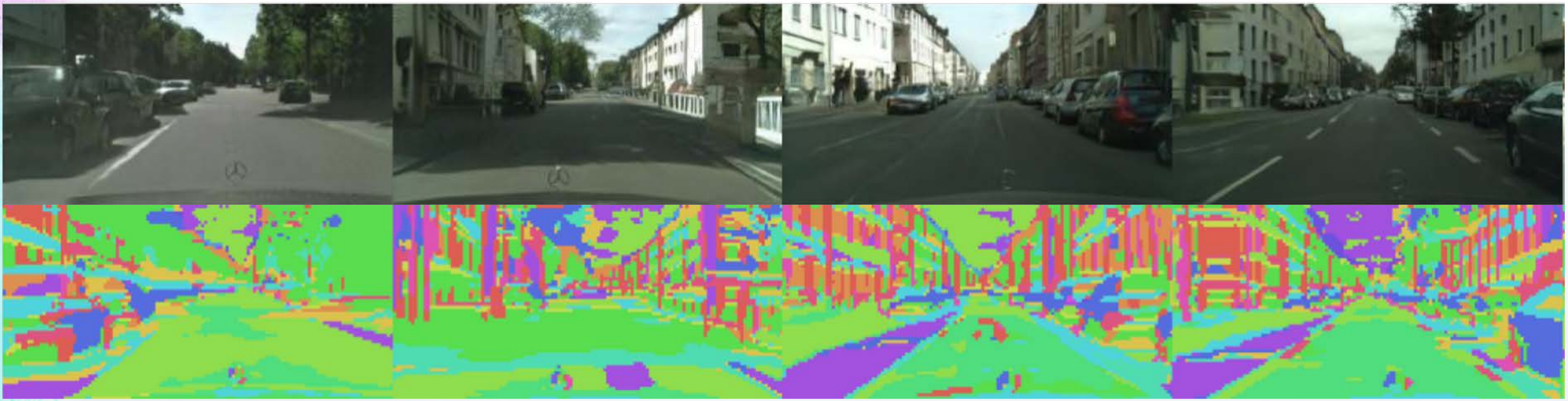
GANformer's Attention Maps

The latents attend to semantic entities and cooperatively generate compositional scenes.



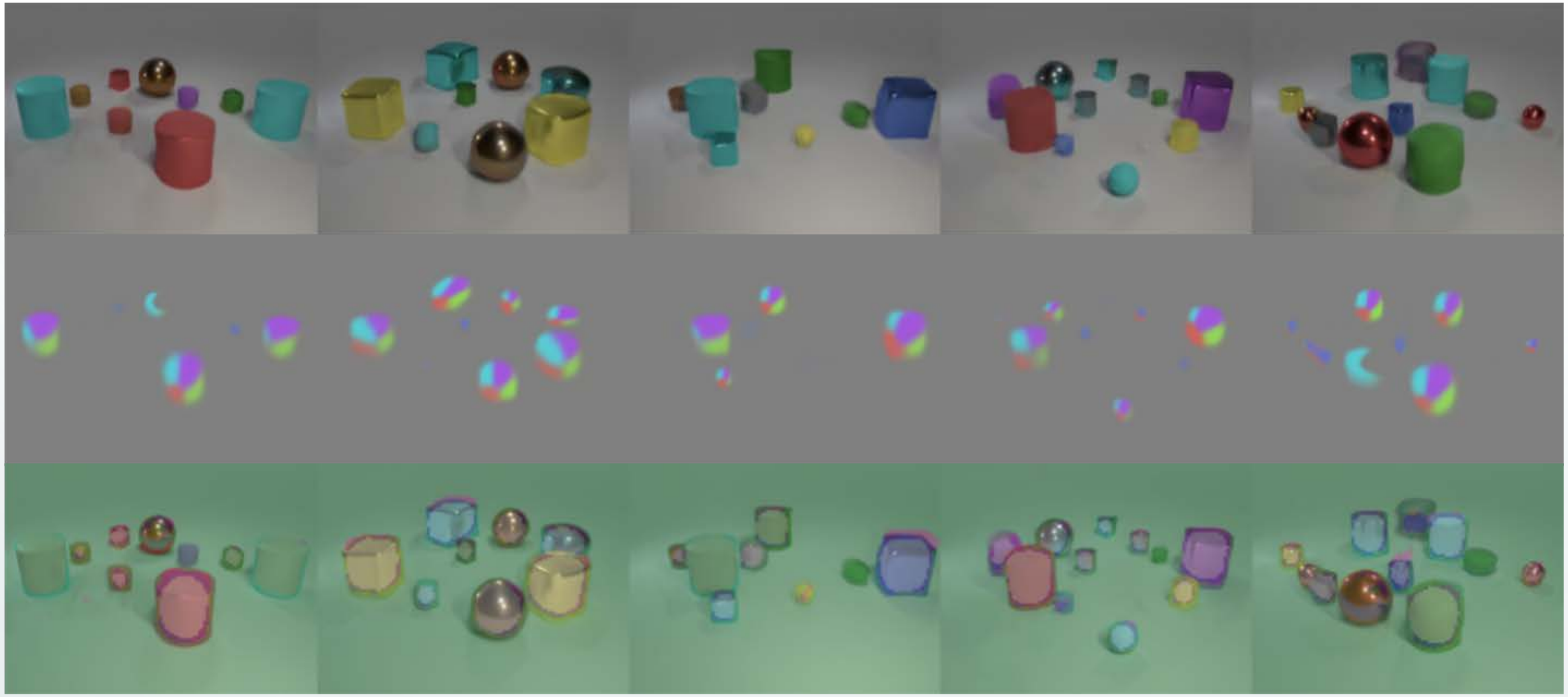
GANformer's Attention Maps

The latents attend to semantic entities and cooperatively generate compositional scenes.

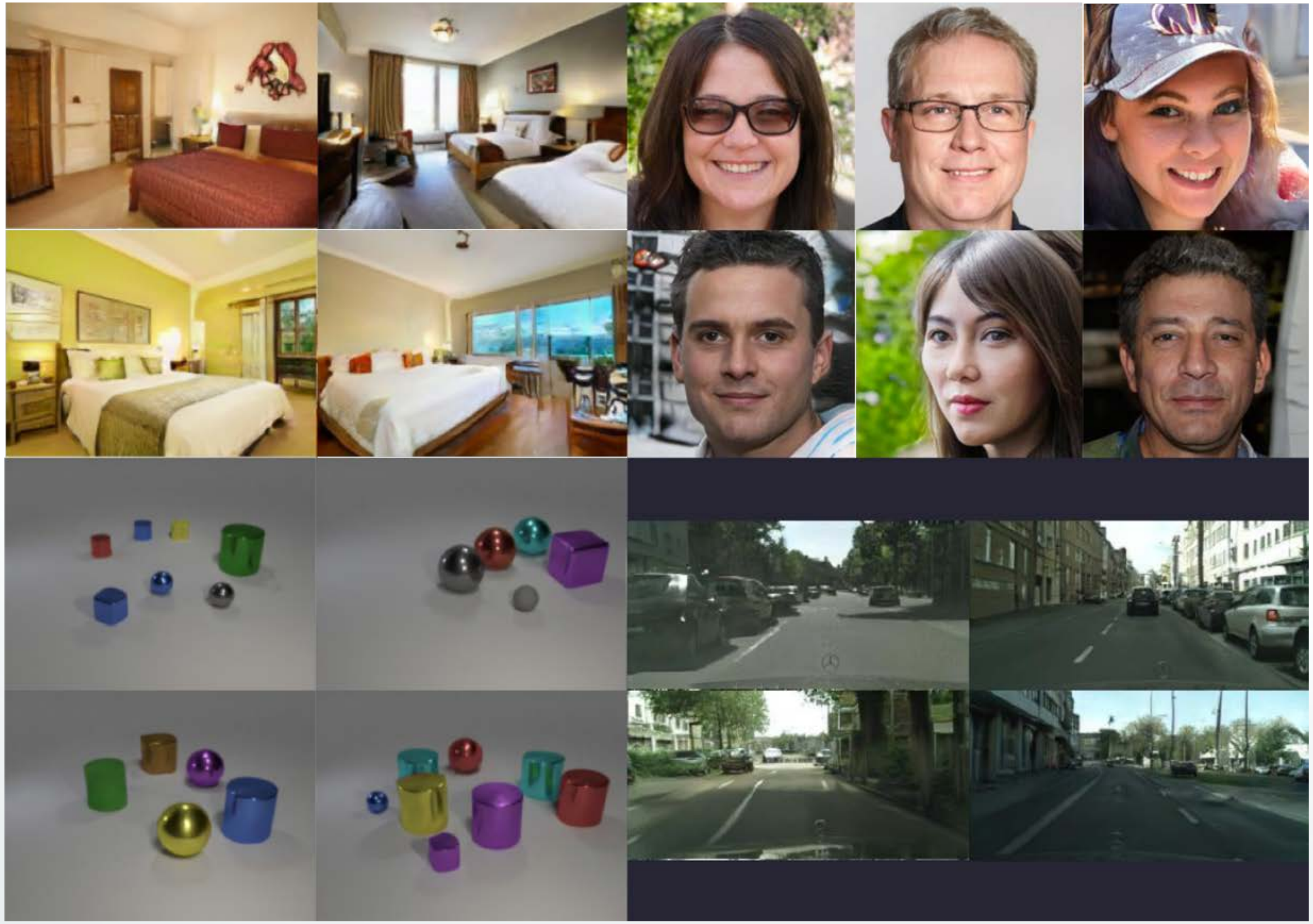


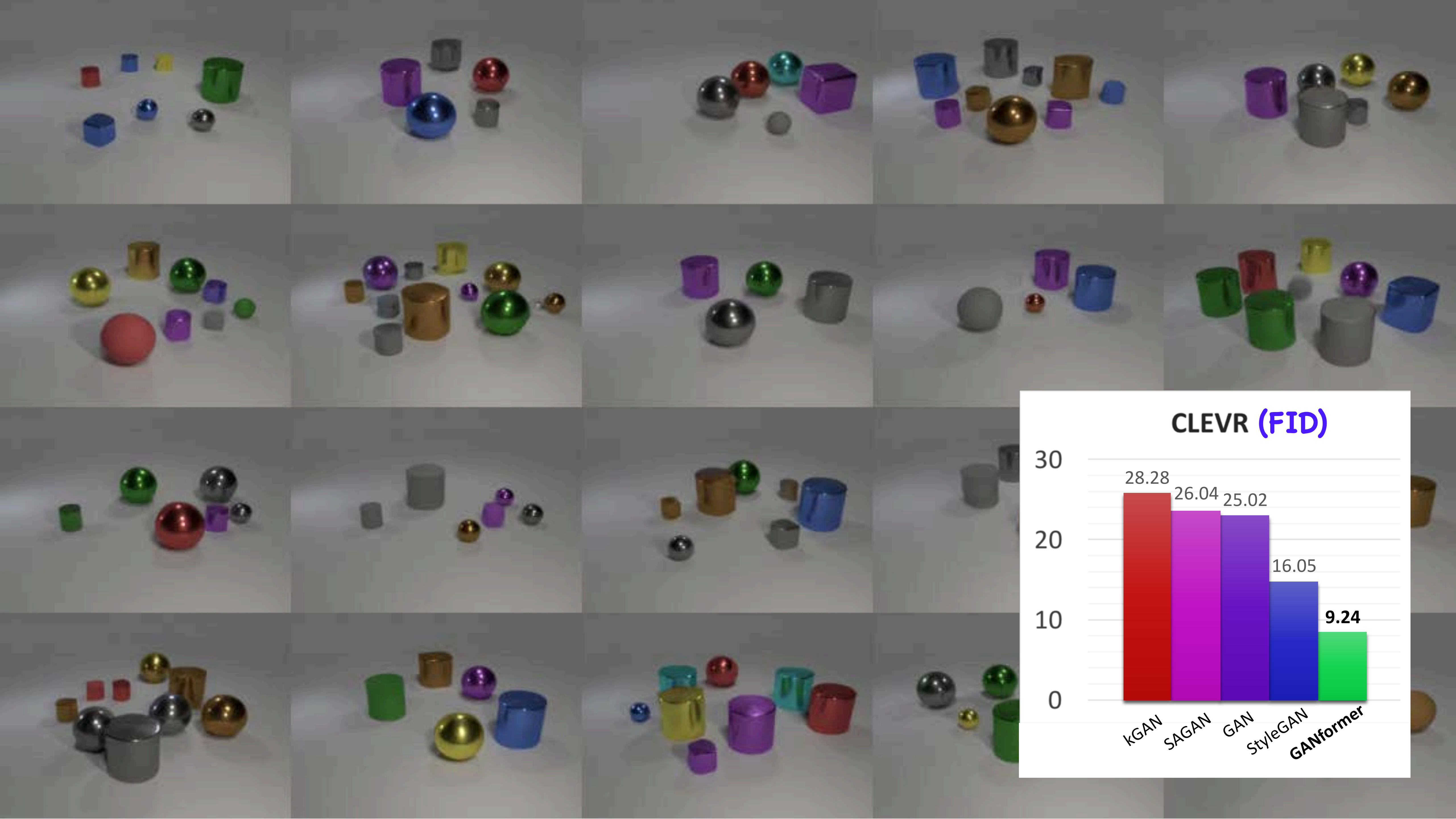
GANformer's Attention Maps

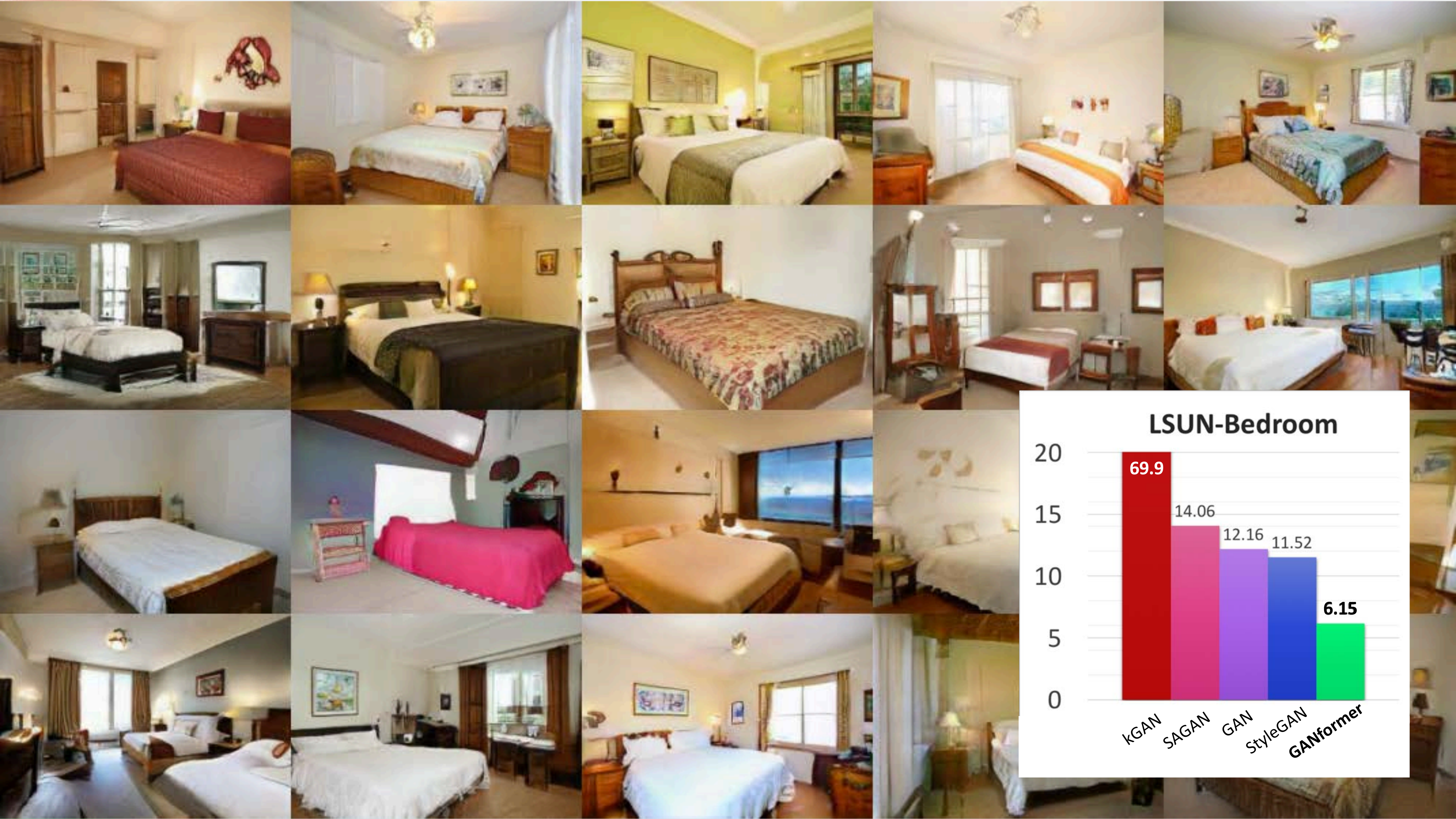
The latents attend to semantic entities and cooperatively generate compositional scenes.

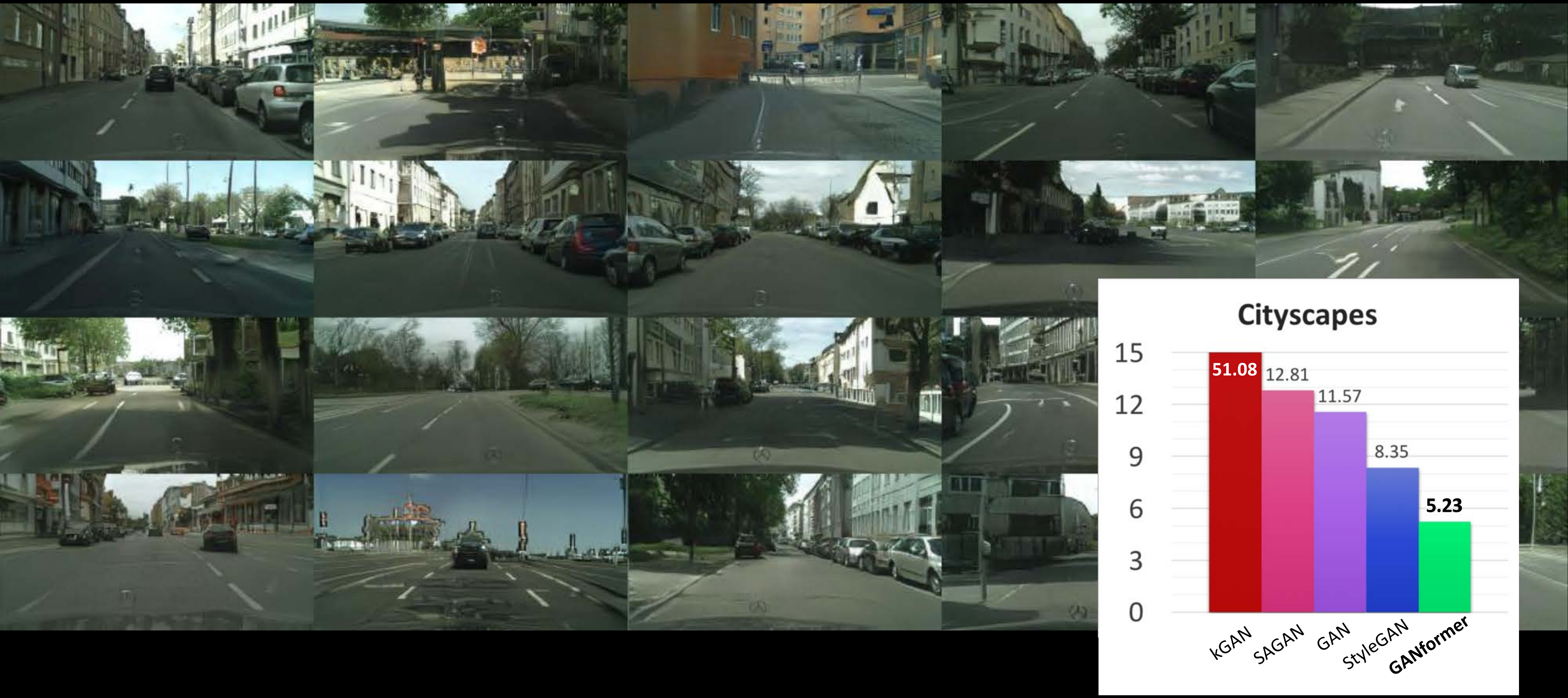


Excels on highly-structured scenes




CLEVR (FID)
30
20
10
0
28.28
26.04
25.02
16.05
9.24
kGAN
SAGAN
GAN
StyleGAN
GANformer

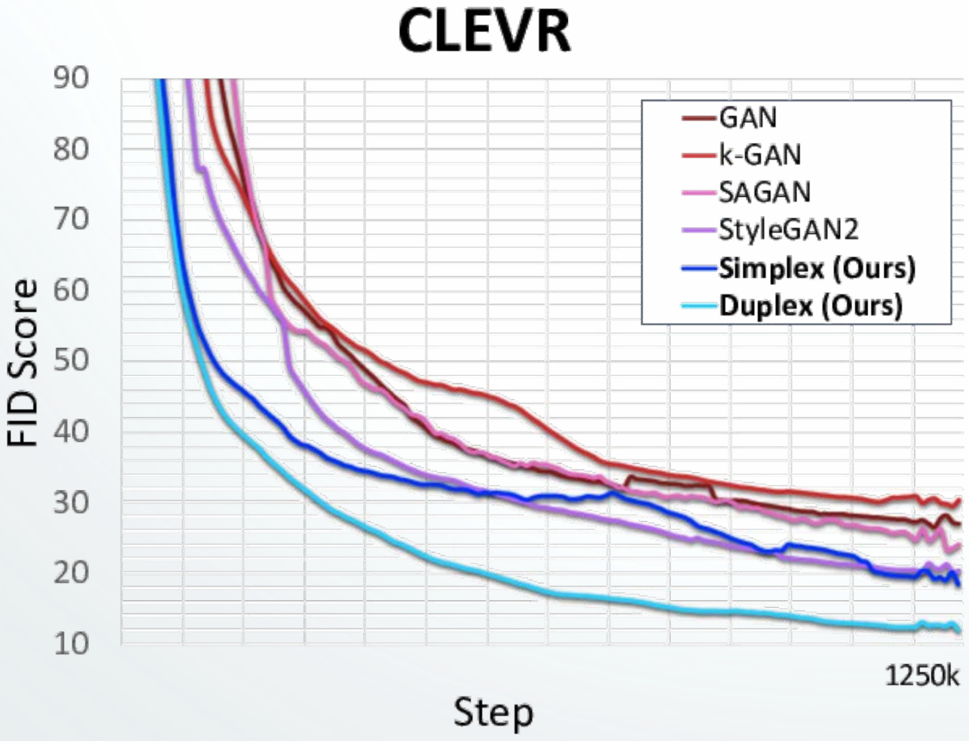

LSUN-Bedroom
20
15
10
5
0
69.9
14.06
12.16
11.52
6.15
kGAN
SAGAN
GAN
StyleGAN
GANformer

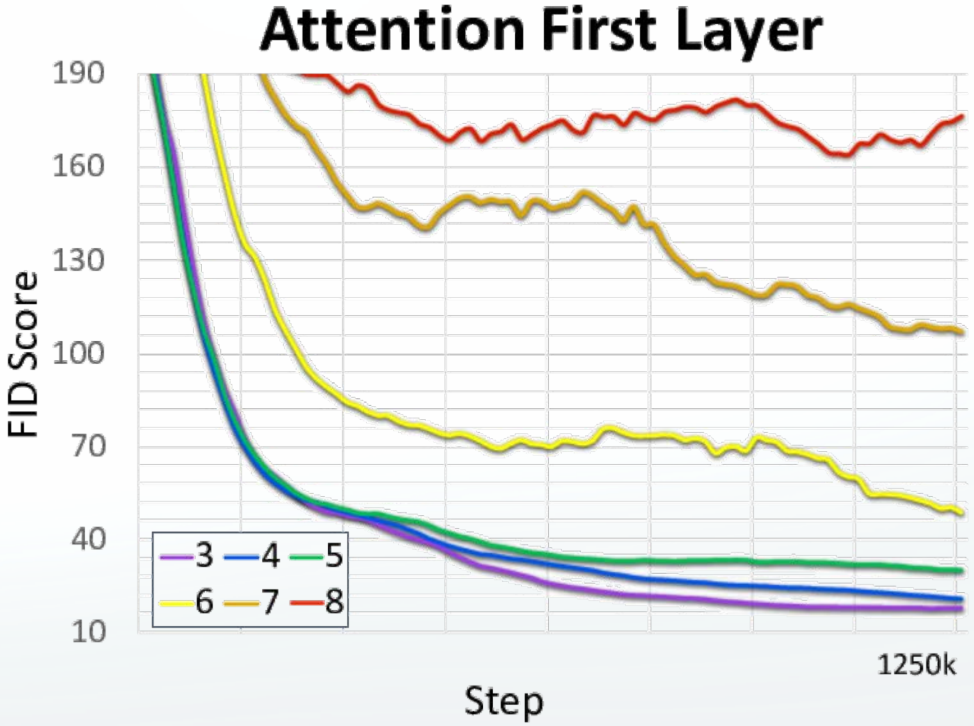

Cityscapes
15
12
9
6
3
0
51.08
12.81
11.57
8.35
5.23
kGAN
SAGAN
GAN
StyleGAN
GANformer

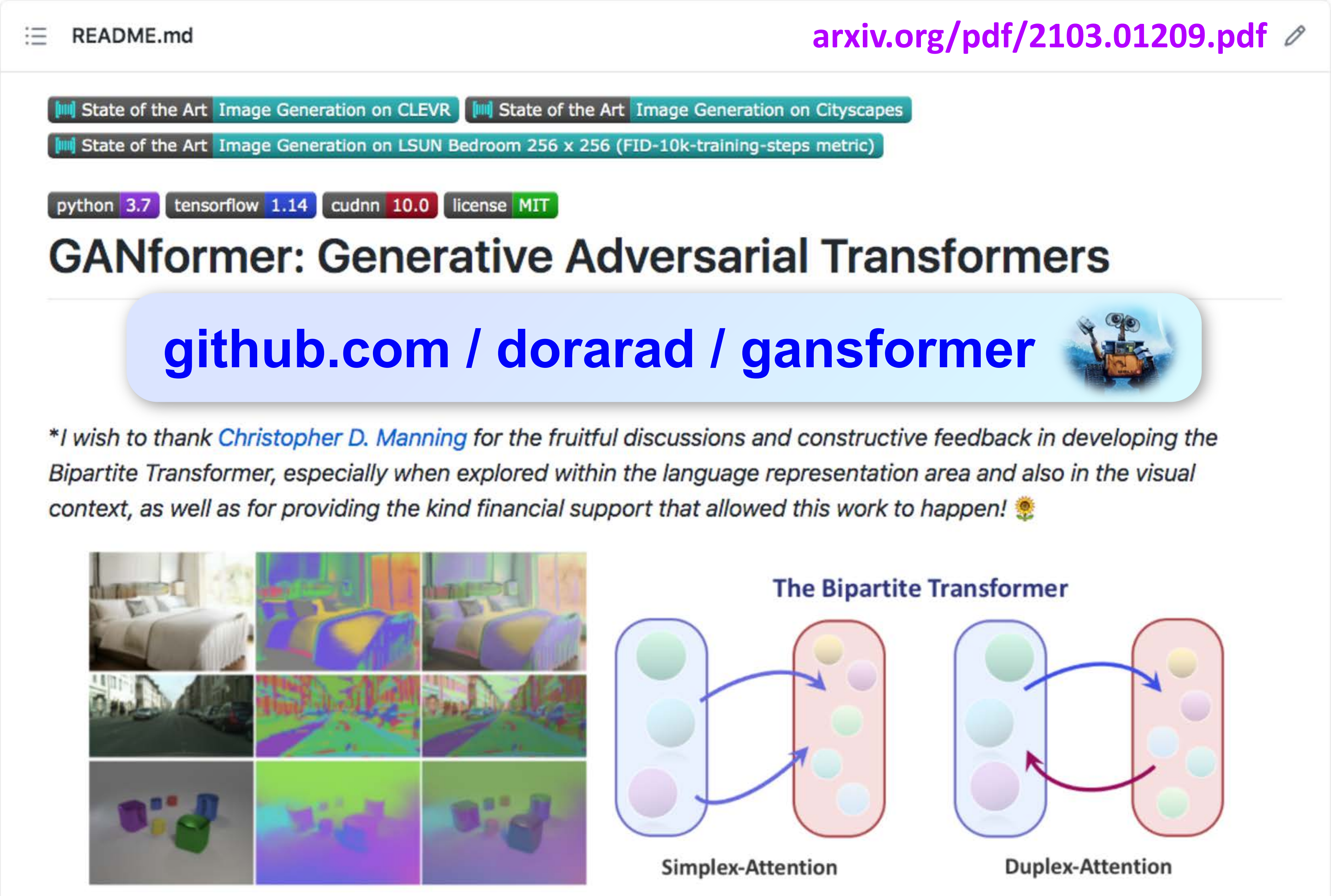
GANformer Model Analysis

The model learns faster and enjoys higher data-efficiency.

Latents are more disentangled; Images are more diverse.







README.md

arxiv.org/pdf/2103.01209.pdf

State of the Art | Image Generation on CLEVR

State of the Art | Image Generation on Cityscapes

State of the Art | Image Generation on LSUN Bedroom 256 x 256 (FID-10k-training-steps metric)

python 3.7 | tensorflow 1.14 | cudnn 10.0 | license MIT

GANformer: Generative Adversarial Transformers

github.com / dorarad / gansformer

**I wish to thank Christopher D. Manning for the fruitful discussions and constructive feedback in developing the Bipartite Transformer, especially when explored within the language representation area and also in the visual context, as well as for providing the kind financial support that allowed this work to happen!* 🌻

github.com / dorarad / gansformer

Thank you! ☺