Diffusion Models for Visual Content Creation

Niloy Mitra, Duygu Ceylan, Paul Guerrero,

Daniel Cohen-Or, Or Patashnik, Chun-Hao Huang, Minhyuk Sung

Part 2: Guidance and Conditioning

https://geometry.cs.ucl.ac.uk/courses/diffusion4ContentCreation_sigg24/

Presentation Schedule

Introduction to Diffusion Models

Guidance and Conditioning

Attention

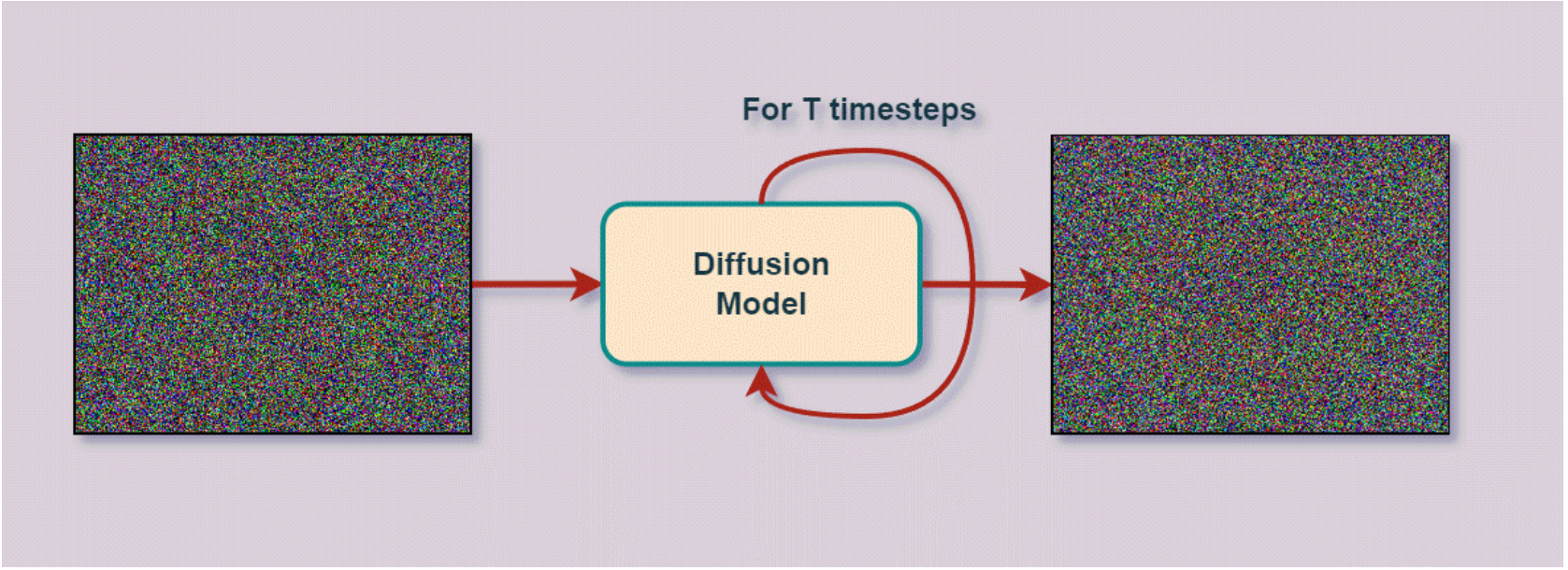
Break

Personalization and Editing

Beyond RGB Images

Diffusion Models for 3D Generation

Diffusion Models: Unconditional Sampling

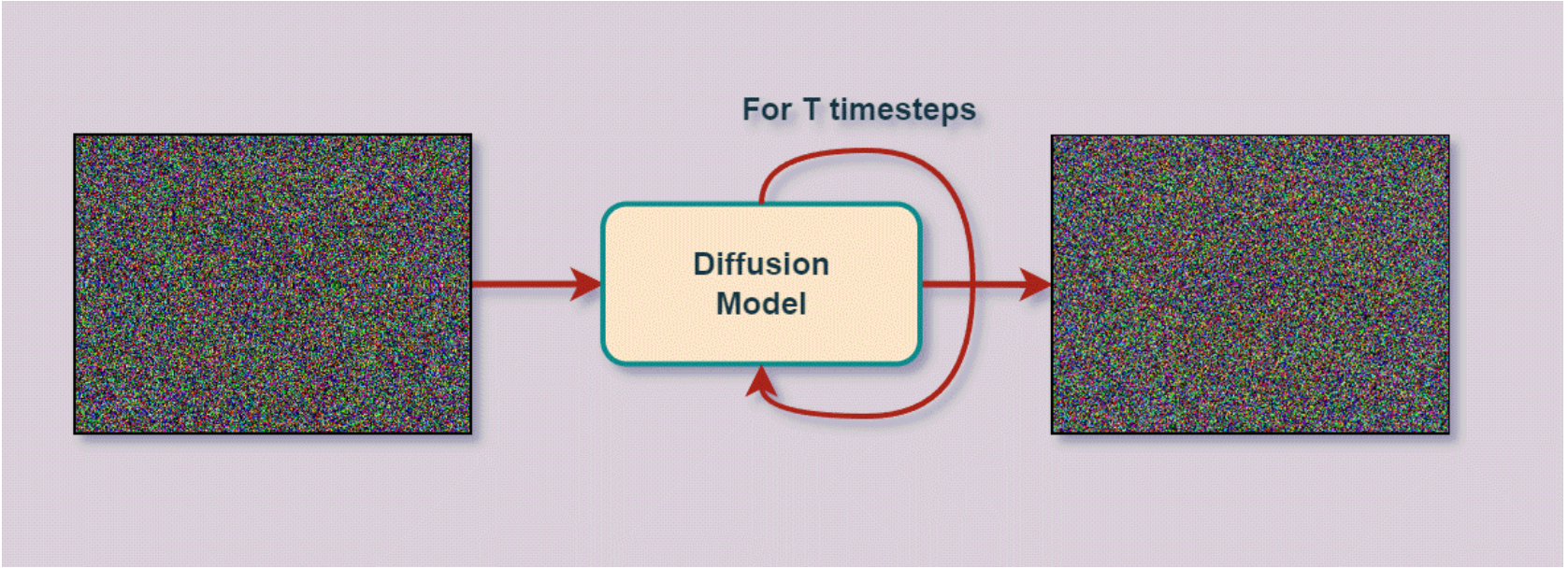


$$p(x) \approx \nabla_x \log p(x)$$

score function

https://learnopencv.com/image-generation-using-diffusion-models/
https://sander.ai/2022/05/26/guidance.html

Diffusion Models: Unconditional Sampling



$$p(x|y) \approx \nabla_x \log p(x|y) = \nabla_x \log p(y \mid x) + \nabla_x \log p(x)$$

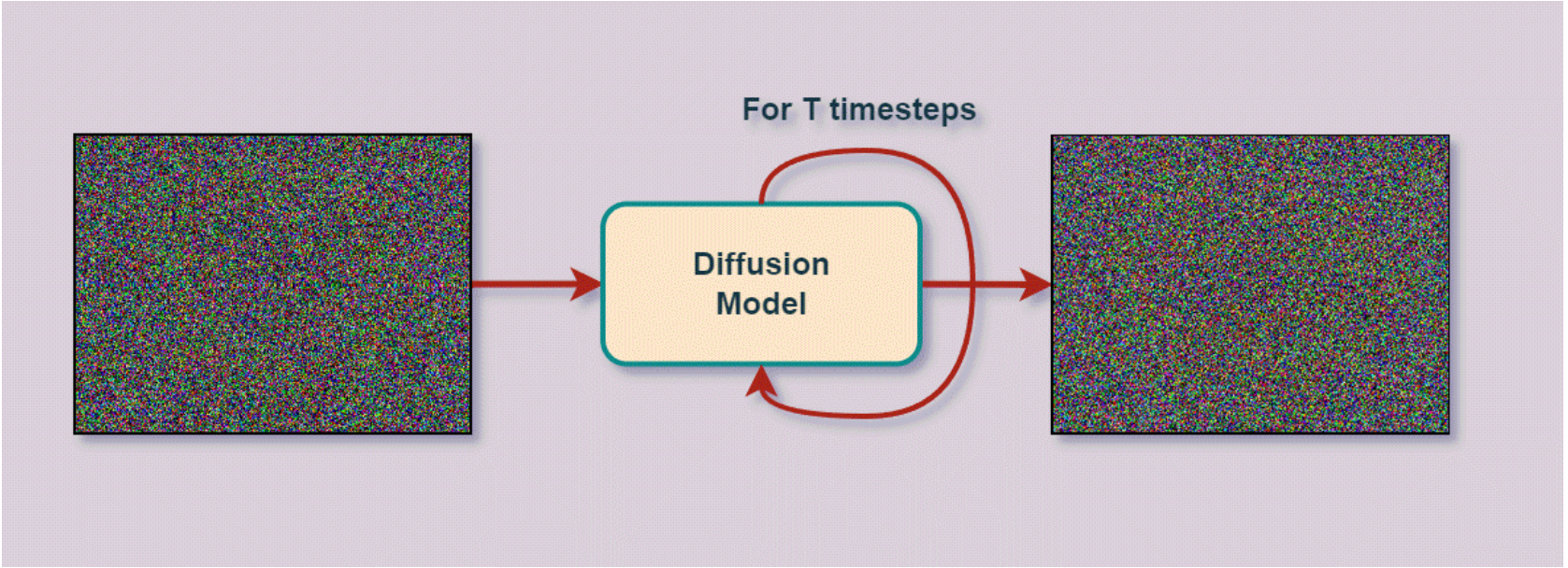
condition

unconditional score function

conditioning term

https://learnopencv.com/image-generation-using-diffusion-models/
https://sander.ai/2022/05/26/guidance.html

Diffusion Models: Conditional Sampling



$$p(x|y) \approx \nabla_x \log p(x|y) = \nabla_x \log p(y \mid x) + \nabla_x \log p(x)$$

classifier

https://learnopencv.com/image-generation-using-diffusion-models/
https://sander.ai/2022/05/26/guidance.html

Classifier Guidance

$$\nabla_x \log p(x|y) = \gamma \nabla_x \log p(y \mid x) + \nabla_x \log p(x)$$

guidance scale

$\gamma = 1$

$\gamma = 10$

Diffusion Models Beat GANs on Image Synthesis, Dhariwal et al., 2021

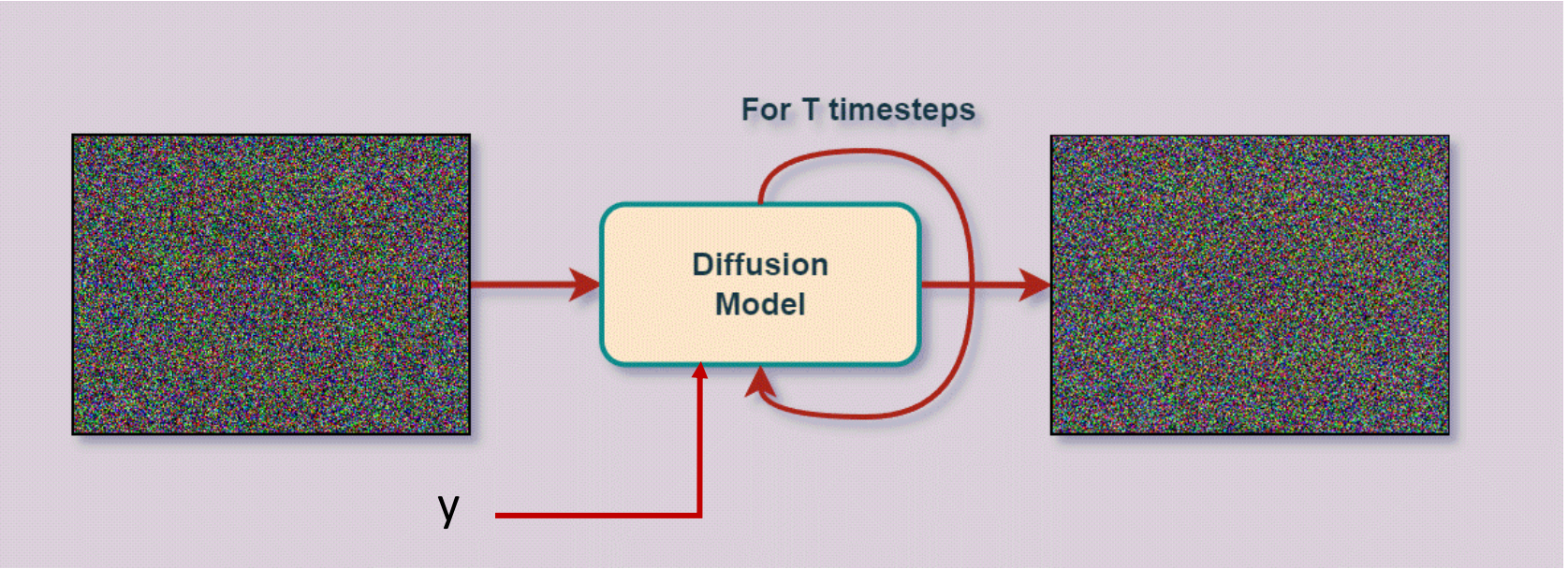
Classifier Guidance

✓ Effectively turns an unconditional diffusion model to a conditional one via a differentiable classifier

✗ Requires the classifier to cope with inputs with high noise levels (e.g., retraining the classifier)

✗ Most information in x may not be relevant to the classifier so the gradient can yield arbitrary directions

Classfier-free Guidance (CFG)



Dropout strategy: 10-20% of the time y is replaced with a special input (e.g., a null text in case of text conditioning)

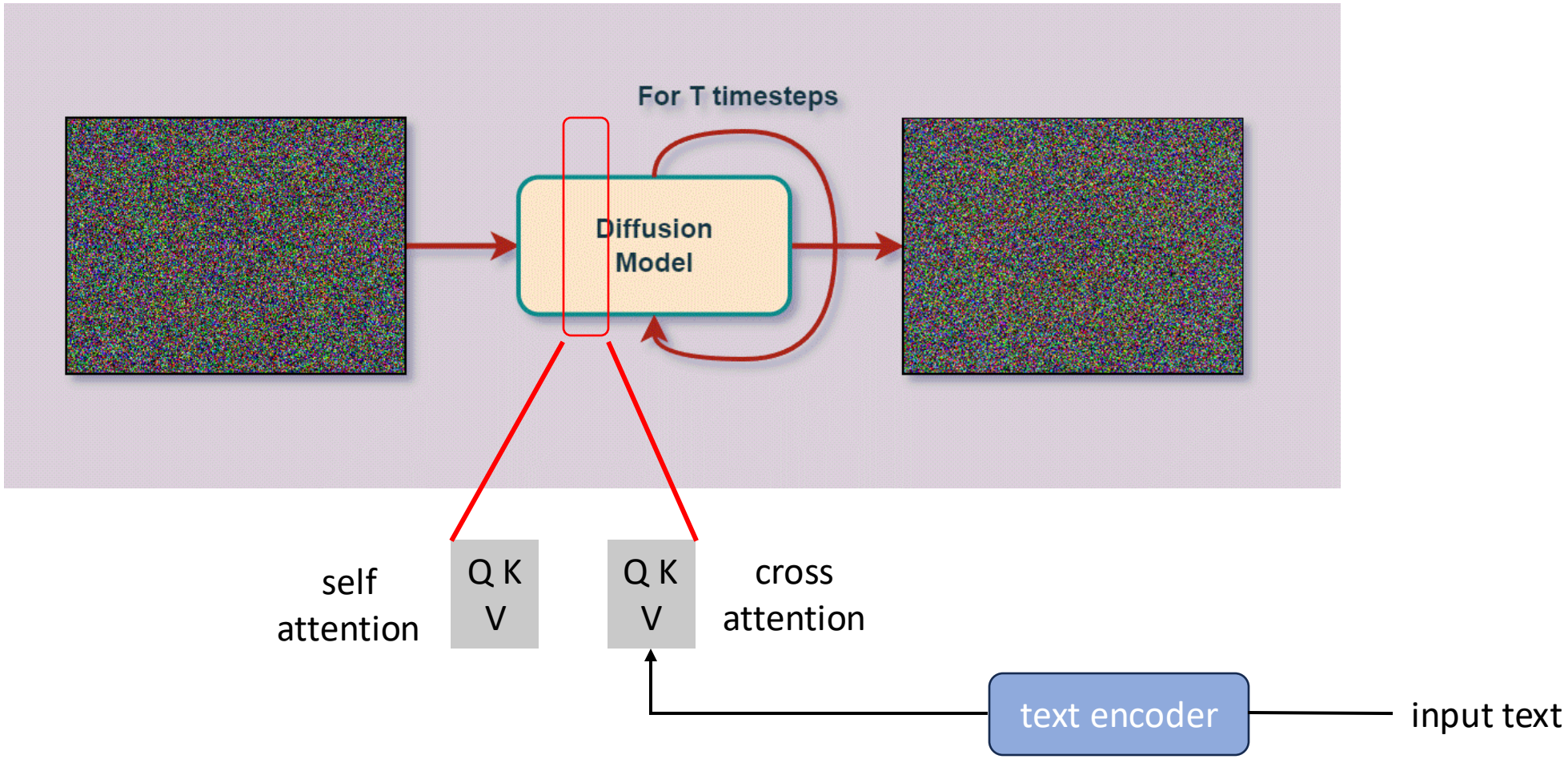
https://learnopencv.com/image-generation-using-diffusion-models/
https://sander.ai/2022/05/26/guidance.html

Classifier-free Guidance

✓ Improves adherence to the conditioning signal

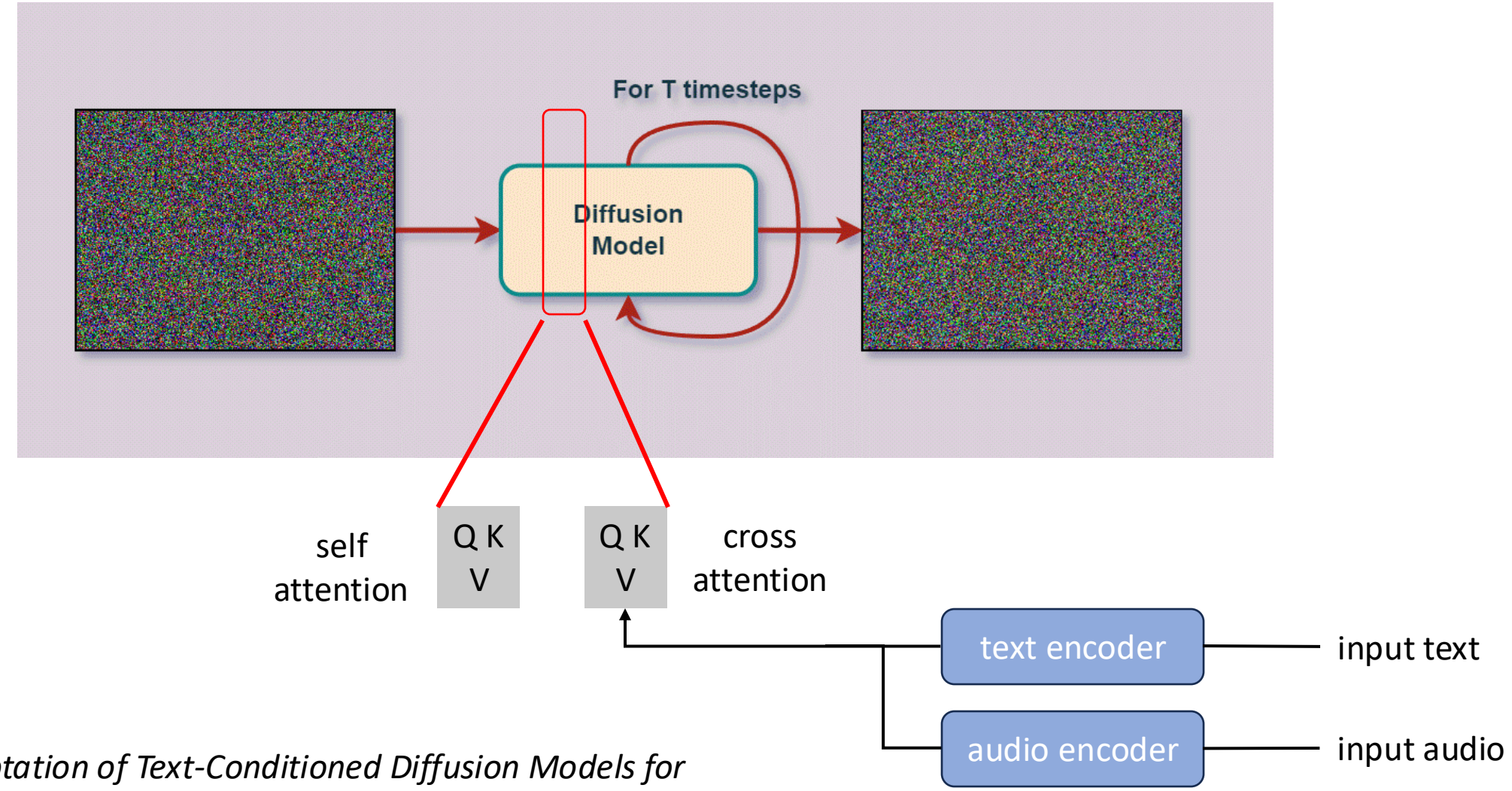
✗ Less diversity (may not be an issue for conditional generation)

How are Conditioning Signals Provided?



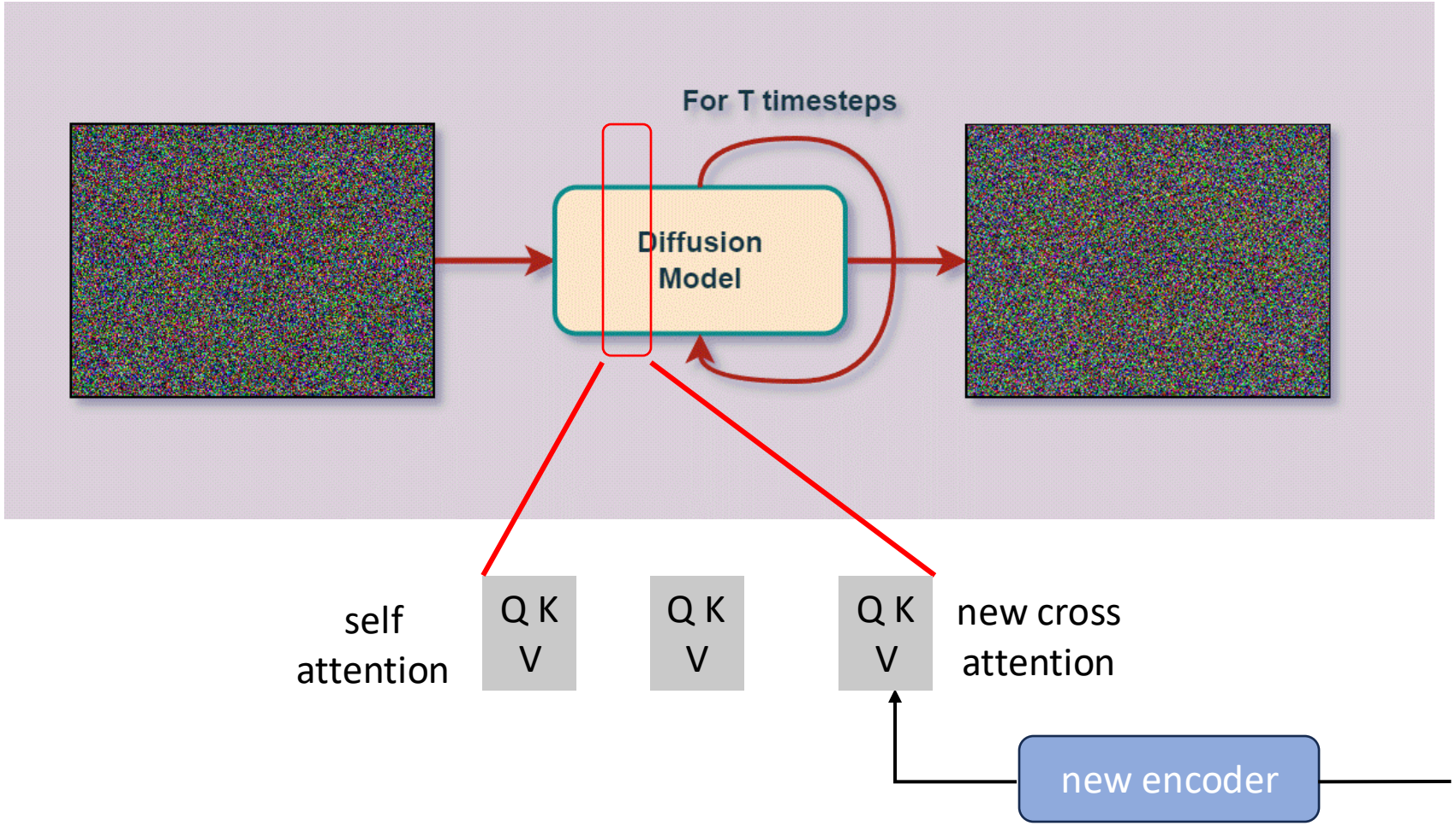
https://learnopencv.com/image-generation-using-diffusion-models/

Different Conditioning Signals

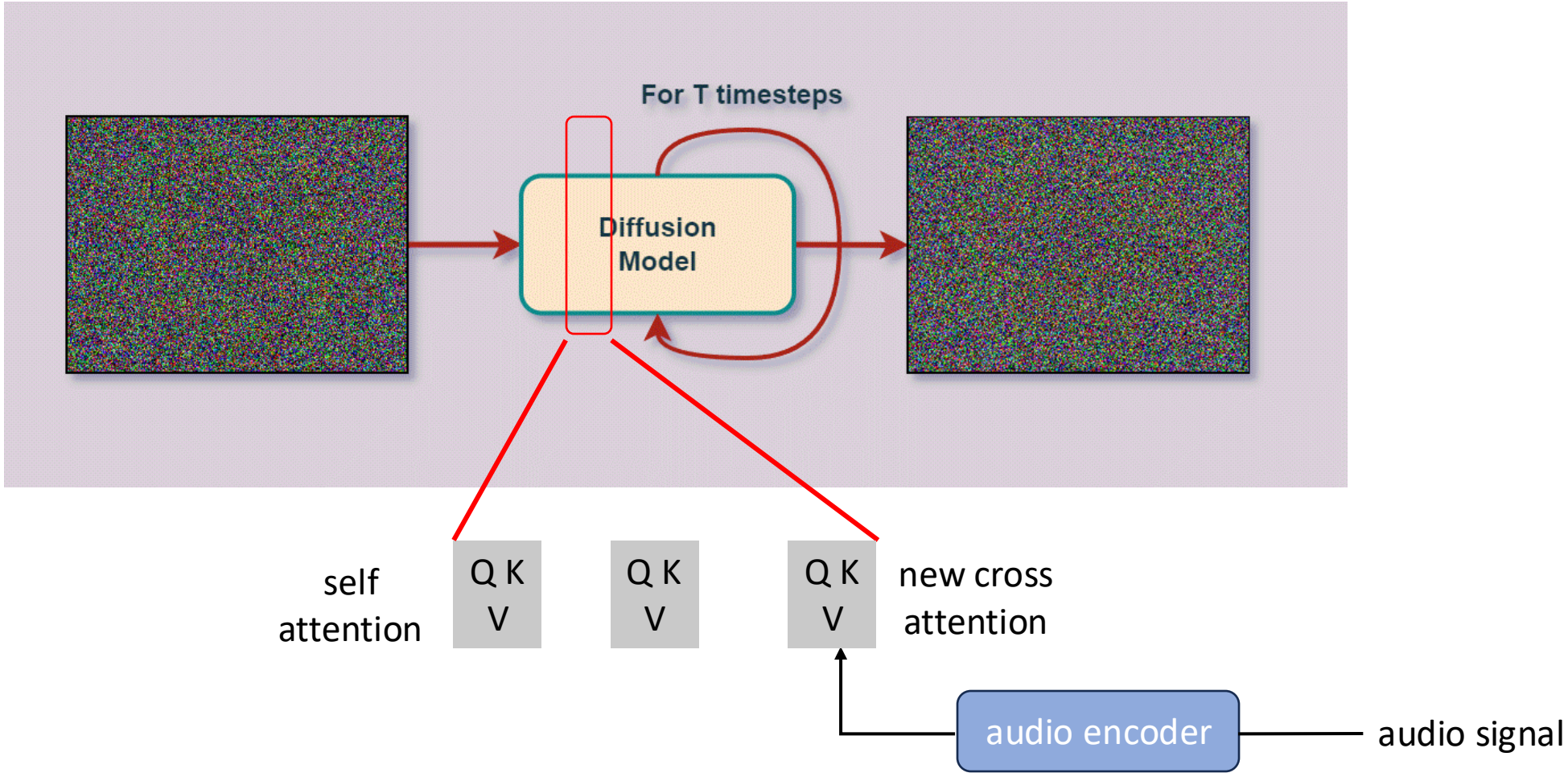


AudioToken: Adaptation of Text-Conditioned Diffusion Models for Audio-to-Image Generation, Yariv et al. 2023

Different Conditioning Signals



Different Conditioning Signals



under review

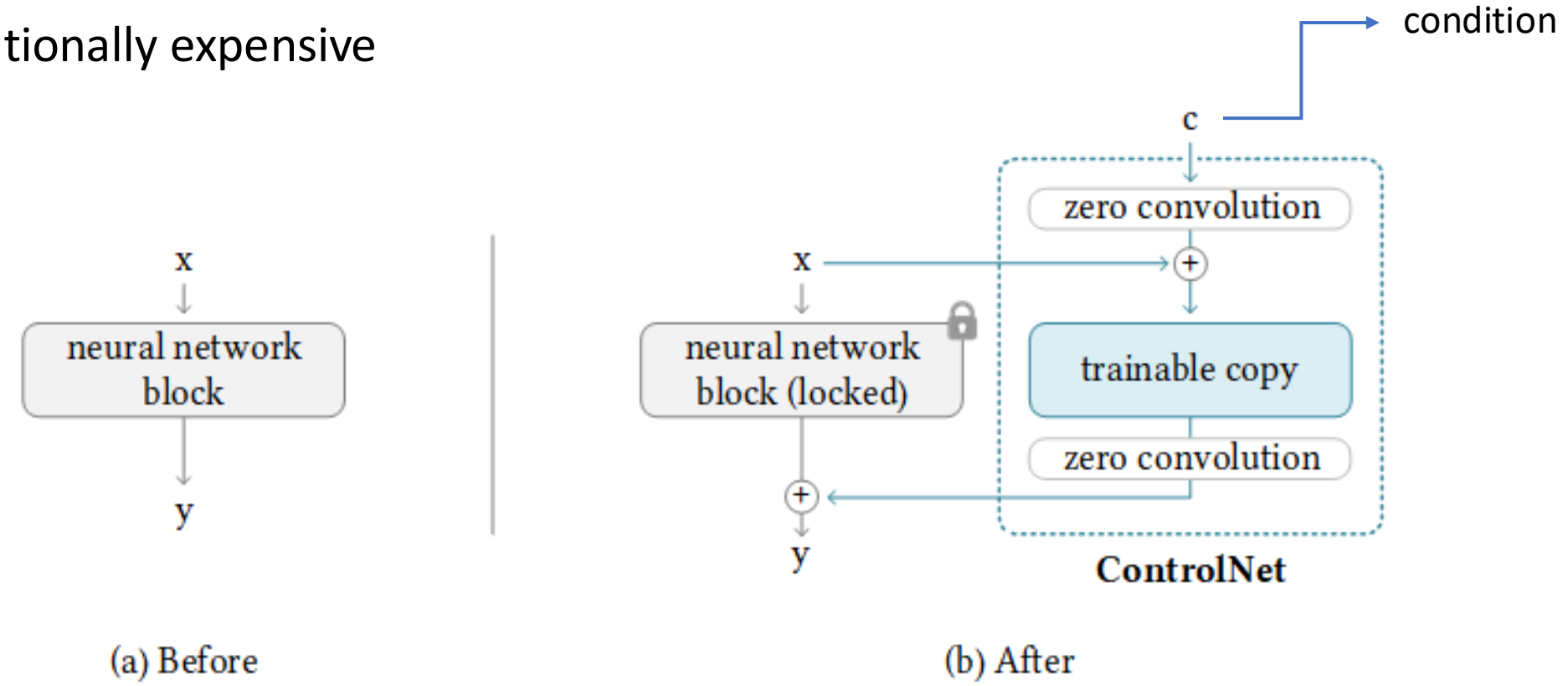
Audio-conditioned Generation

under review

Other Conditioning Methods: ControlNet

✓ Can be trained on smaller amount of data.

✗ Computationally expensive



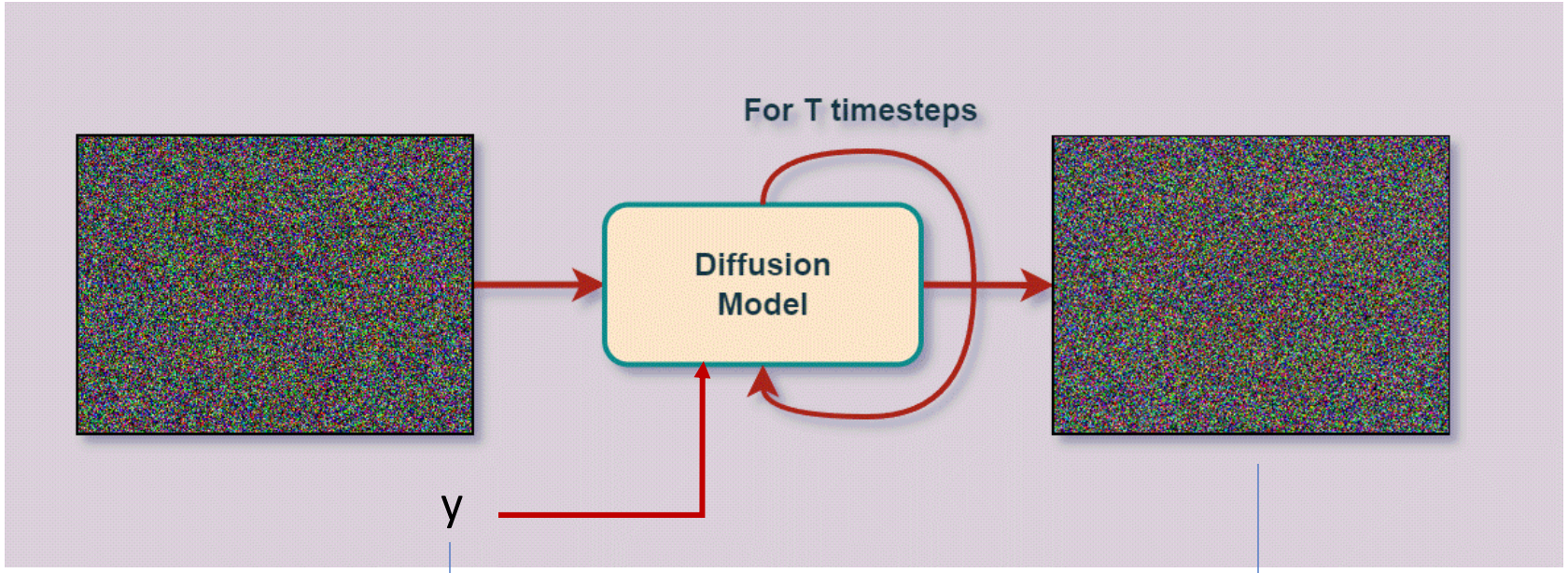
Adding Conditional Control to Text-to-Image Diffusion Models, Zhang et al. 2023

Other Conditioning Methods

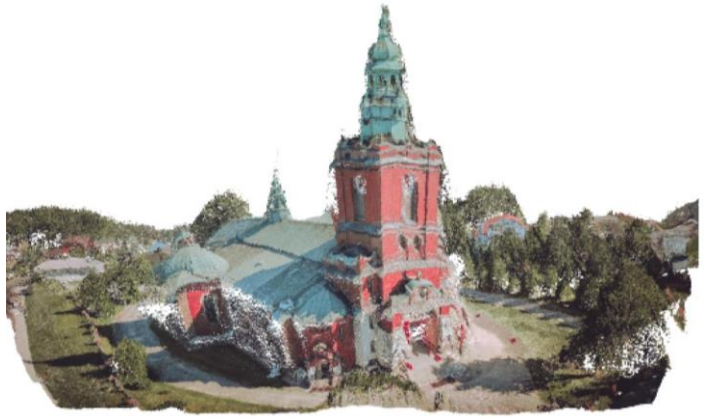
- Concatanate spatial conditioning signal with the input noise

Stable Diffusion v2 Depth-to-Image model

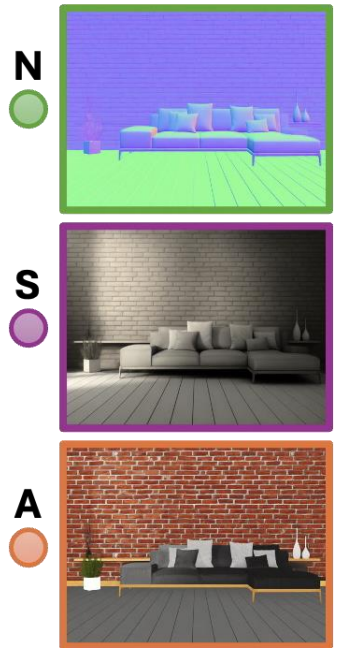
Conditioning to Convert Generators to Regressors



Generation -> Analysis

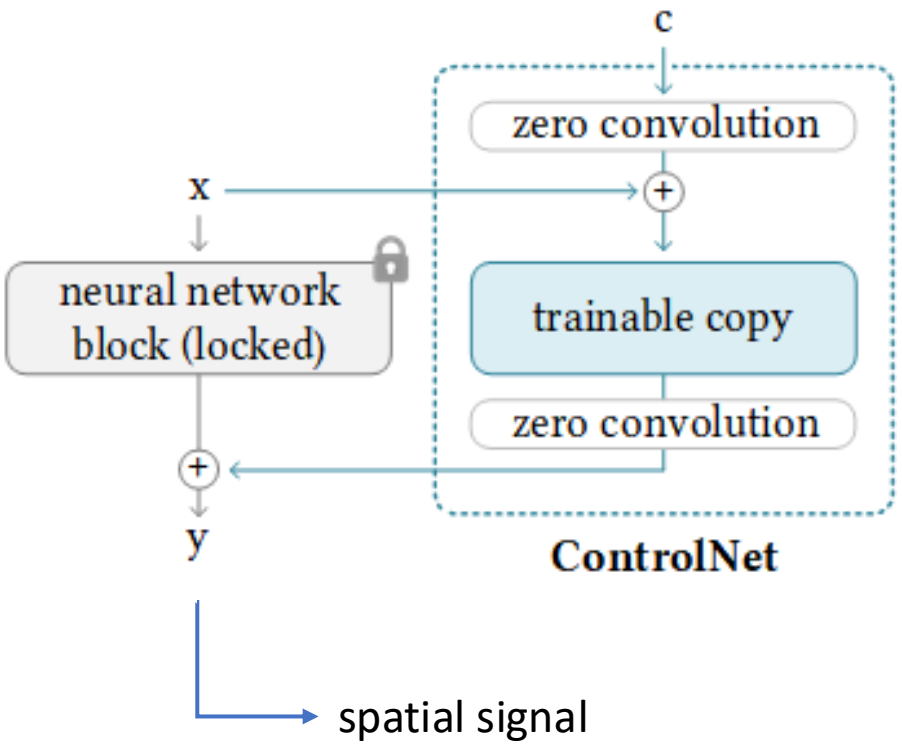


Marigold Depth, CVPR 2024
(Best paper award candidate)



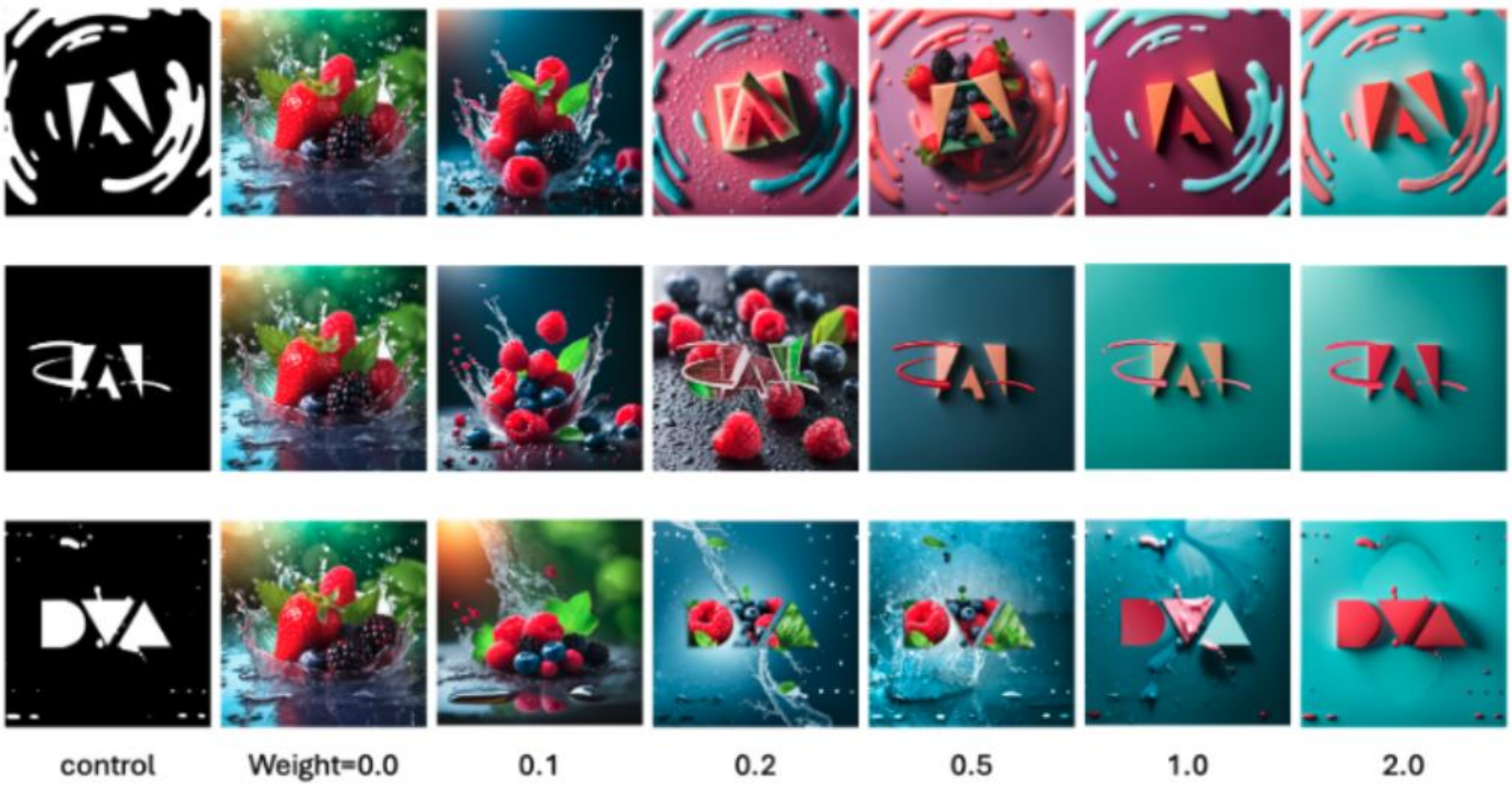
RGB-X, Intrinsic Diffusion
SIGGRAPH 2024

Adherence to the Conditioning Signal



Weighting can be applied!

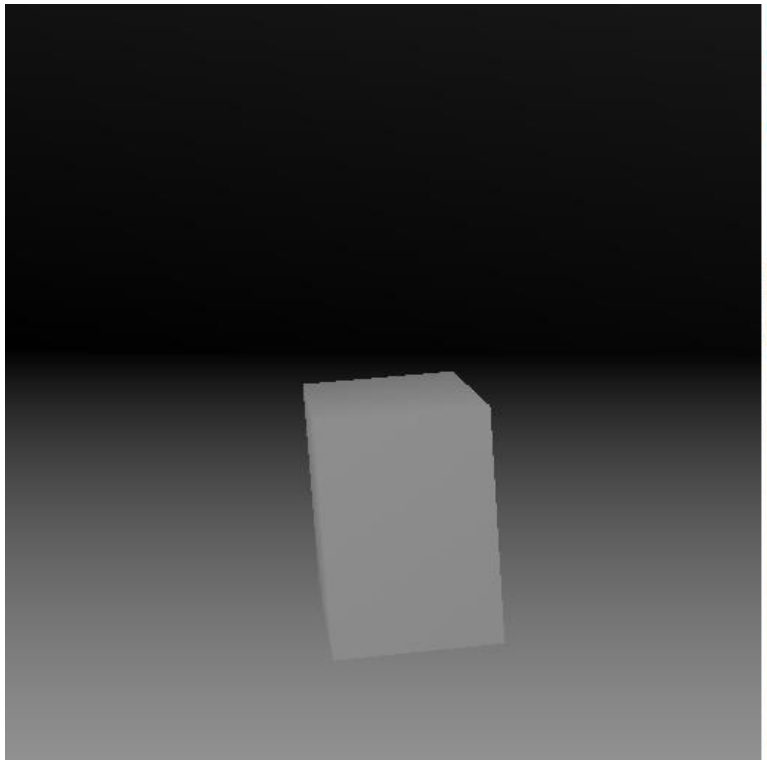
Adherence to the Conditioning Signal



Adherence to the Conditioning Signal

control | spatial weight | w/ spatial weight | w/ global weight

Adherence to the Conditioning Signal

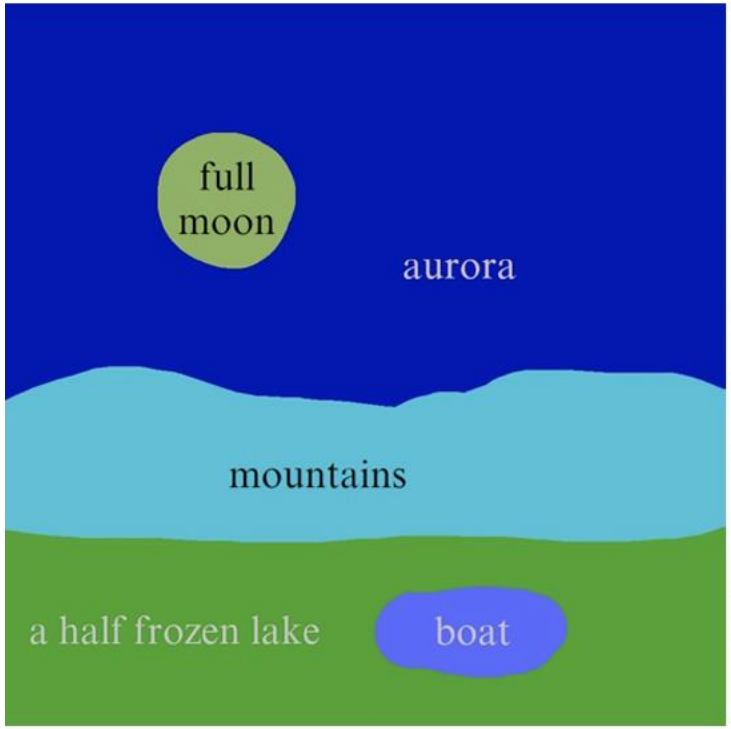


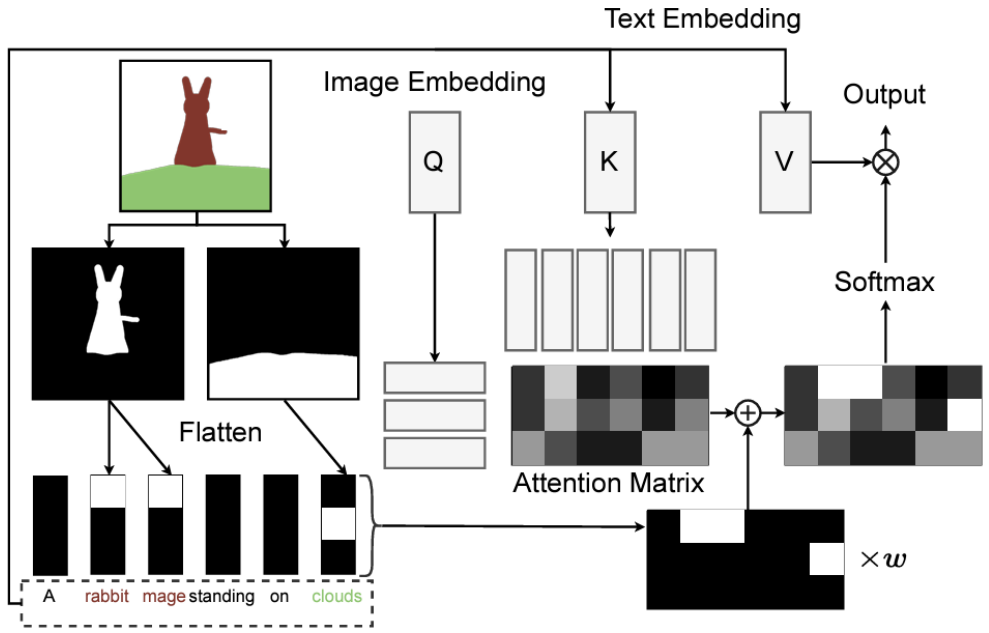
original ControlNet

LooseControl

LooseControl: Lifting ControlNet for Generalized Depth Conditioning, Bhat et al. 2023

Regional Control of Conditioning Signal

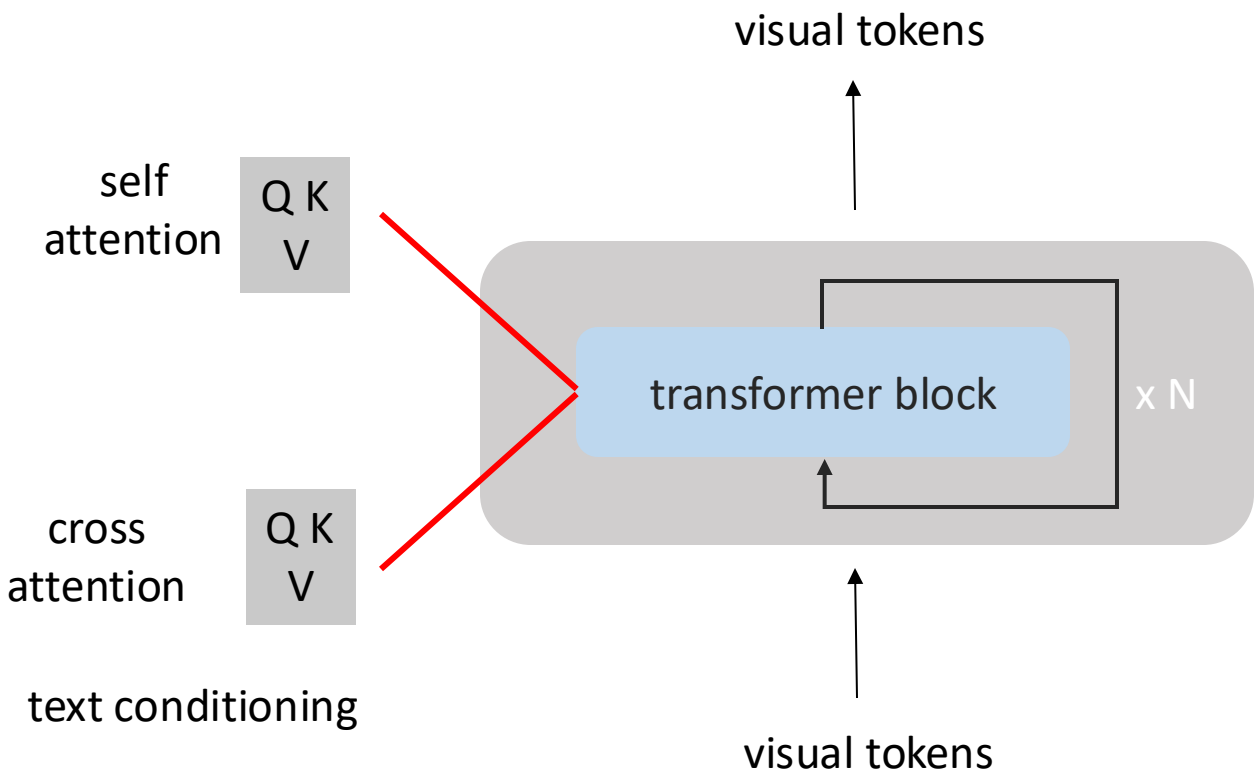




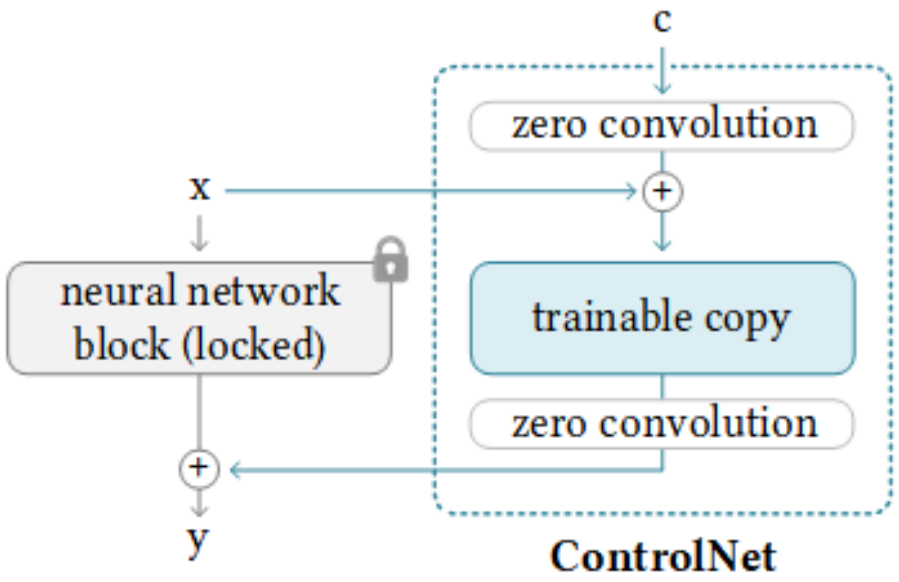
cross attention layer

eDiff-I: Text-to-Image Diffusion Models with an Ensemble of Expert Denoisers, Balaji et al. 2023

How About Diffusion Transformers?



ControlNet for DiT



farm workers in the style of photorealistic detailing

PIXART-δ: Fast and Controllable Image Generation with Latent Consistency Models

Conclusion

- Various options for providing conditioning signals
- Interesting research questions:
  - Fine-grained control on adherence to conditioning signals
  - Regional control over conditioning signals
  - Multi-modal generation

Presentation Schedule

Introduction to Diffusion Models

Guidance and Conditioning

Attention

Break

Personalization and Editing

Beyond RGB Images

Diffusion Models for 3D Generation