STAR

<u>ST</u>ochastic <u>A</u>stronomical TU<u>R</u>BO

M. Drozdova, V. Kinakh, Y. Belousov, O. Bait, D. Piras, E. Lastufka, M. Dessauges-Zavadsky, T. Holotyak, D. Schaerer, and S. Voloshynovskiy

SKACH Winter Meeting
January 22, 2024

Problem formulation

Challenges: Accurate reconstruction and localization of astronomical sources under limited observational data.

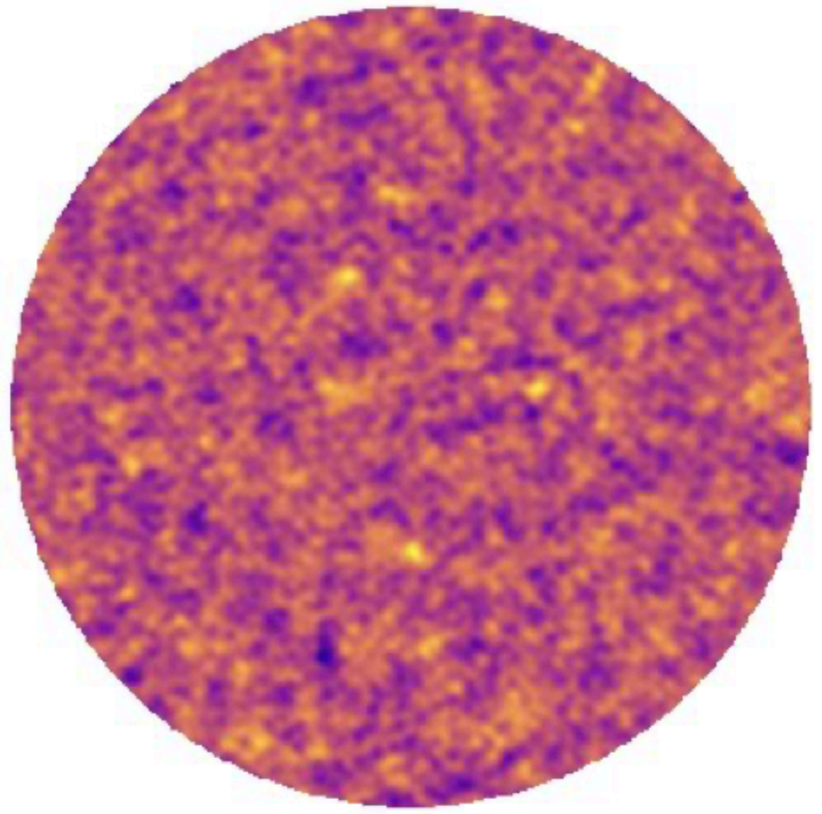
Solution: Framework separating **content** (astronomical sources) from **style** (imaging artifacts and noise) - STAR.

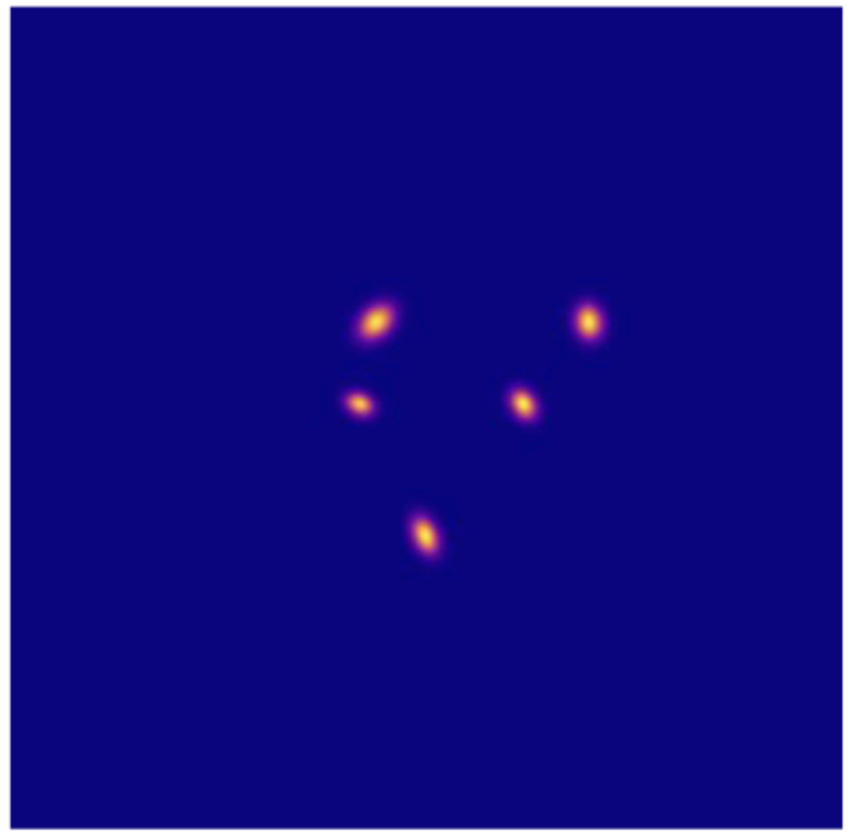
Improved sources localization while keeping the complexity low due to the changes in training procedure.

Possibility to generate synthetic data to overcome limitations in training data availability and quality.

Dataset

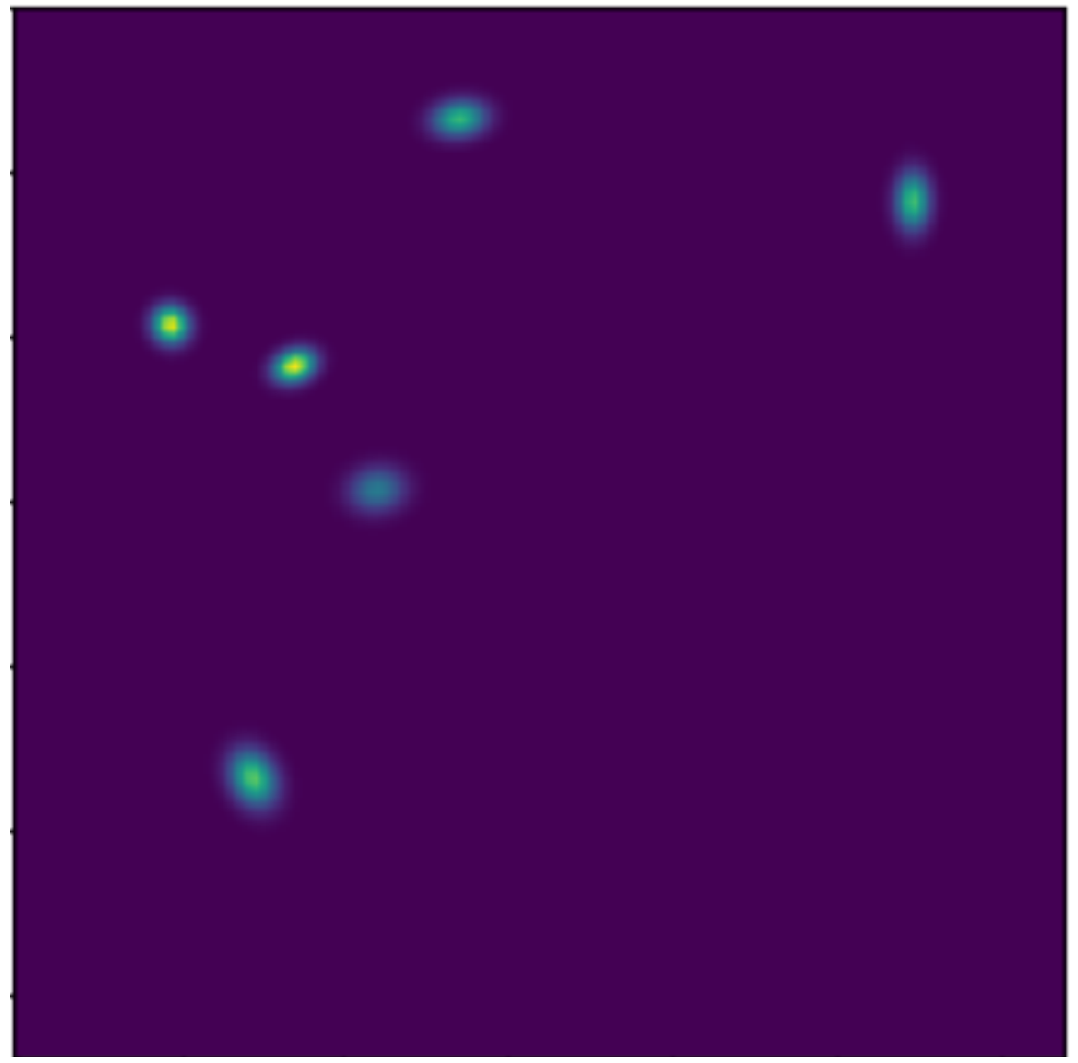
- simulated with CASA data processing software v6.2
- fixed ALMA configuration cycle 5.3
- image size 512x512 corresponding to 51.2'×51.2′ across
- 9164 simulated sky models and the corresponding dirty images (1-5 sources per image)
- 1000 dirty images without sources
- a total of 27632 sources
- the root-mean-squared of the noise is 50 μJy





Gap between observation and ground truth

Sky model

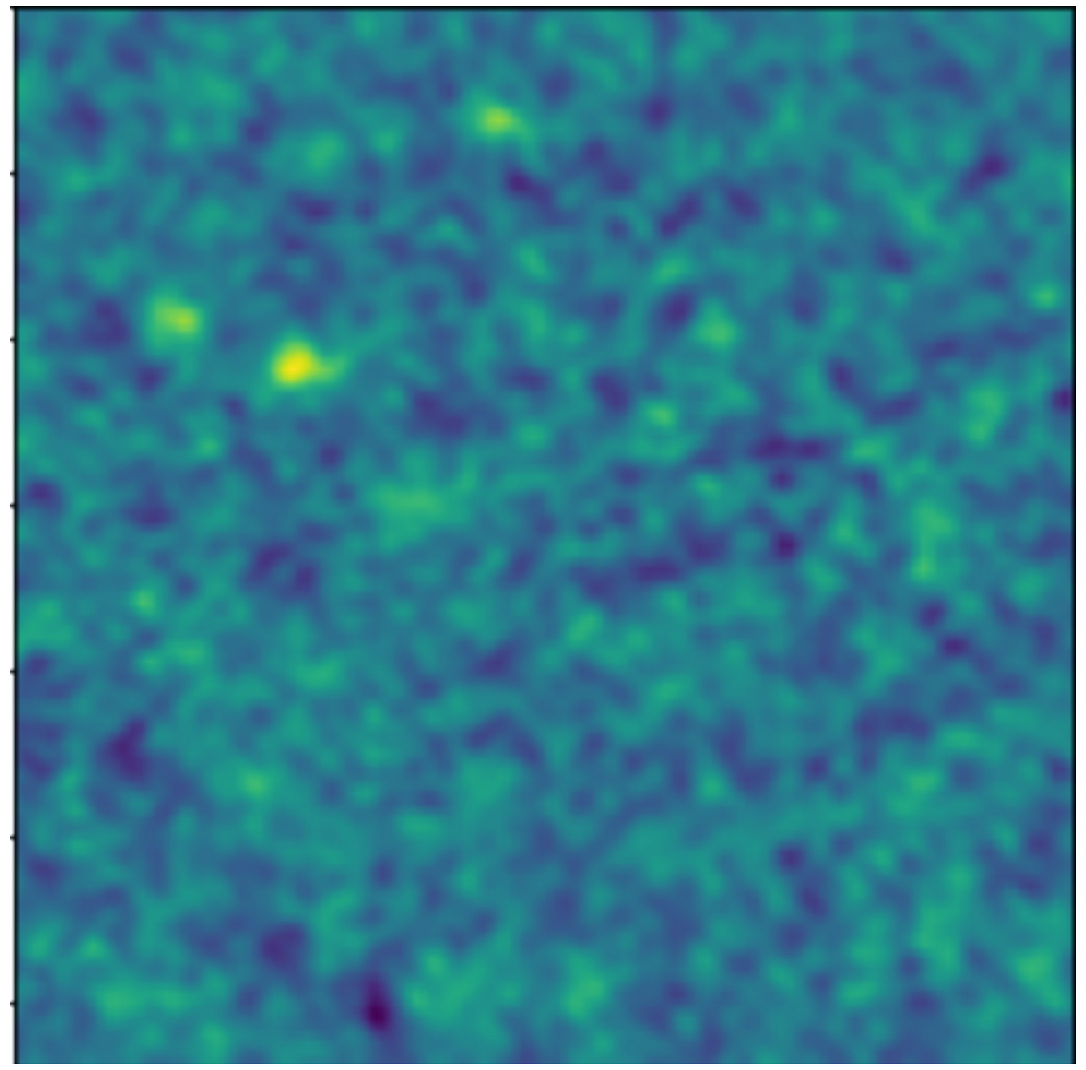


$\mathbf{c}$

$$\mathbf{c} \xrightarrow{\mathbf{x} = \phi(H\mathbf{c}) + \epsilon} \mathbf{x}$$

$$\tilde{\mathbf{c}} \xleftarrow[?]{} \mathbf{x}$$

Dirty image



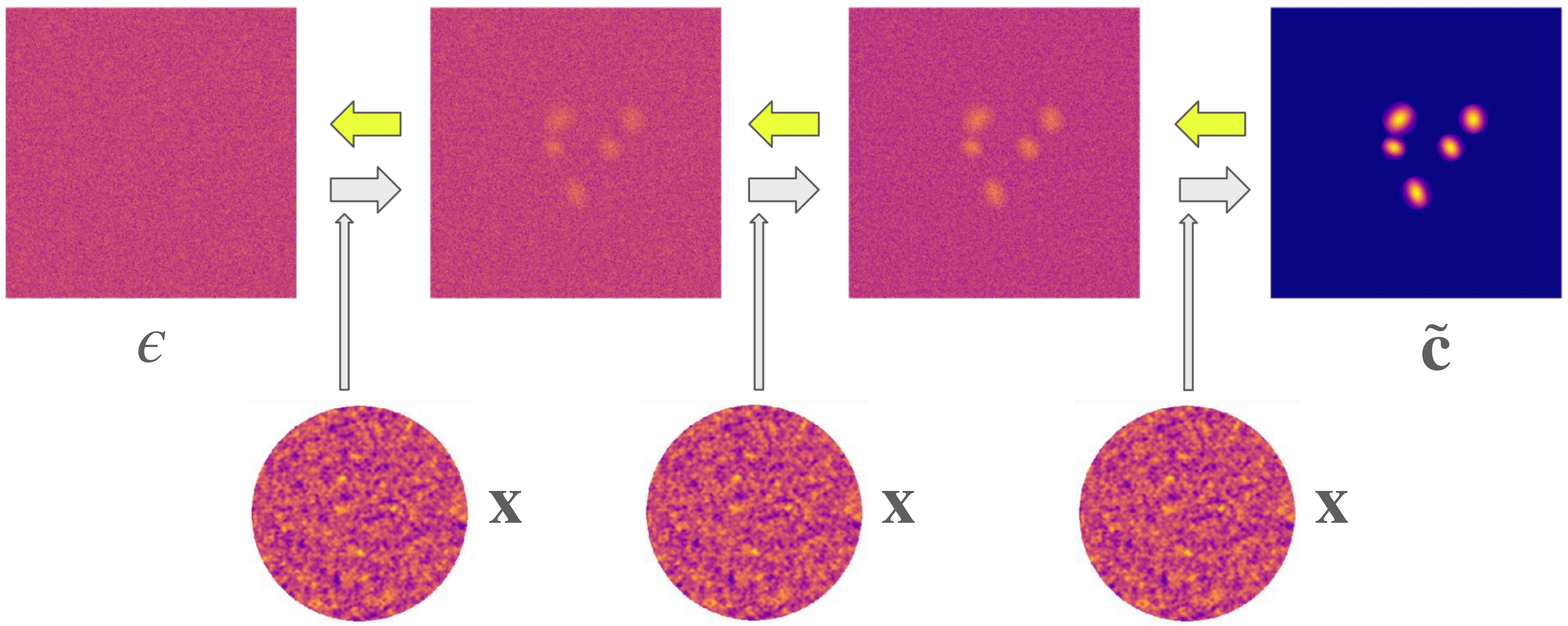
$\mathbf{x}$

$\mathbf{x} \rightarrow \tilde{\mathbf{c}}$ is an inverse ill-posed problem with many solutions => we need (a)regularizations or priors and (b)stochasticity

Denoising Diffusion Probabilistic Model approach

Previously: multi-stage conditional image generation via DDPM (Palette)

+ State-of-the-art accuracy
+ Stochasticity
- Very high complexity



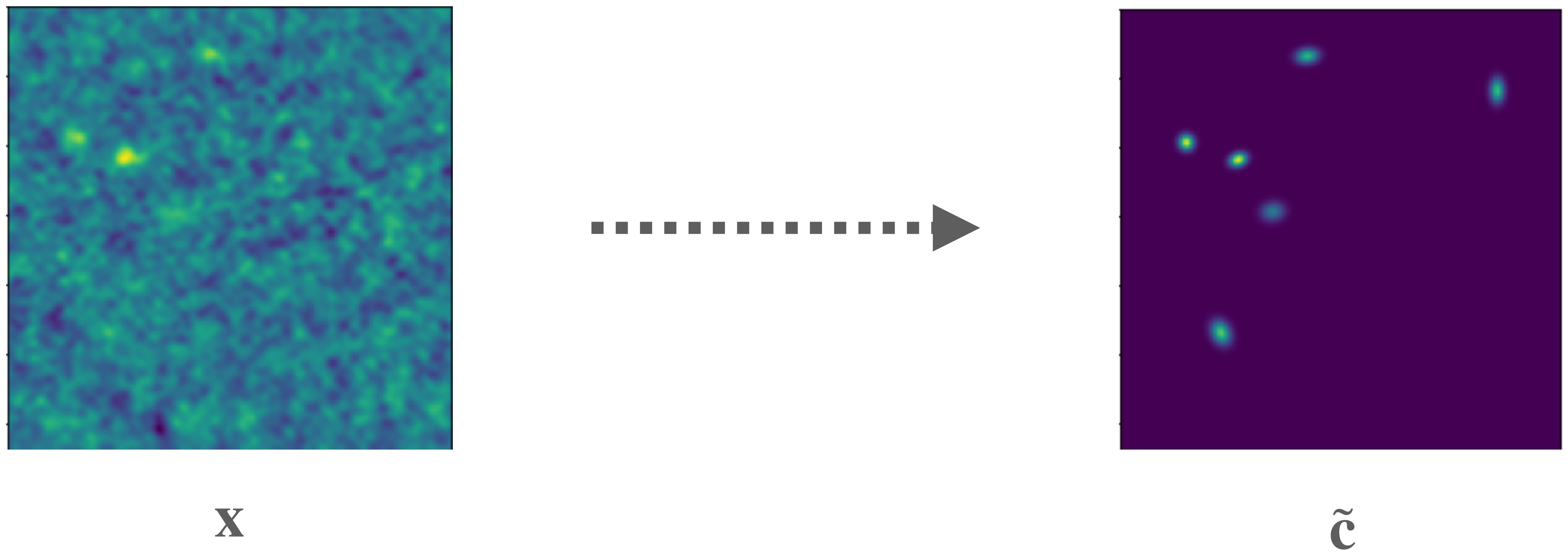
Palette: Image-to-Image Diffusion Models, C. Saharia et al. 2022 SIGGRAPH
Radio-astronomical image reconstruction with a conditional denoising diffusion model, M. Drozdova et al. 2024, Astronomy&Astrophysics

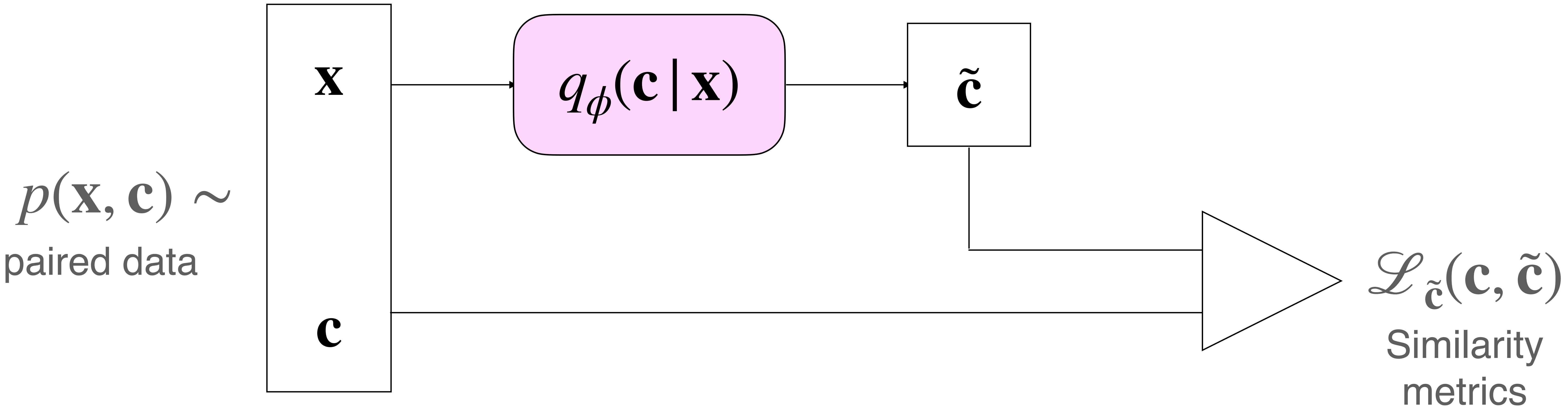
Concept of TURBO

Given: Paired dataset

Goal: Learn mapping $\mathbf{x} \rightarrow \tilde{\mathbf{c}}$

Solution: Encoder $q_\phi(\mathbf{c} \mid \mathbf{x})$ with MSE loss.



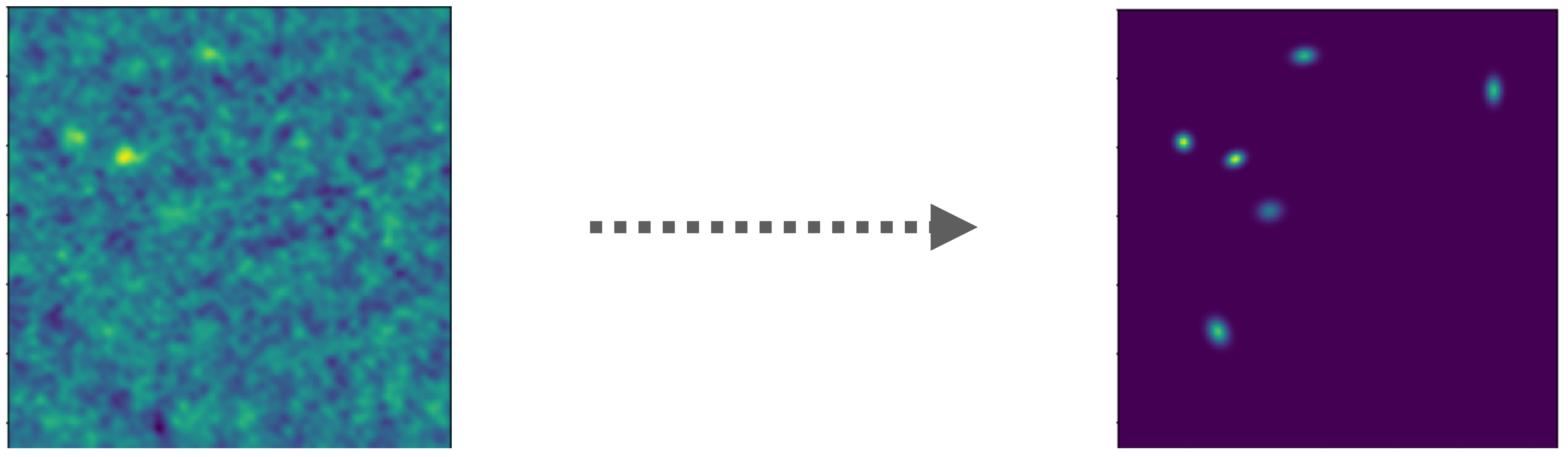


Concept of TURBO

Given: Paired dataset

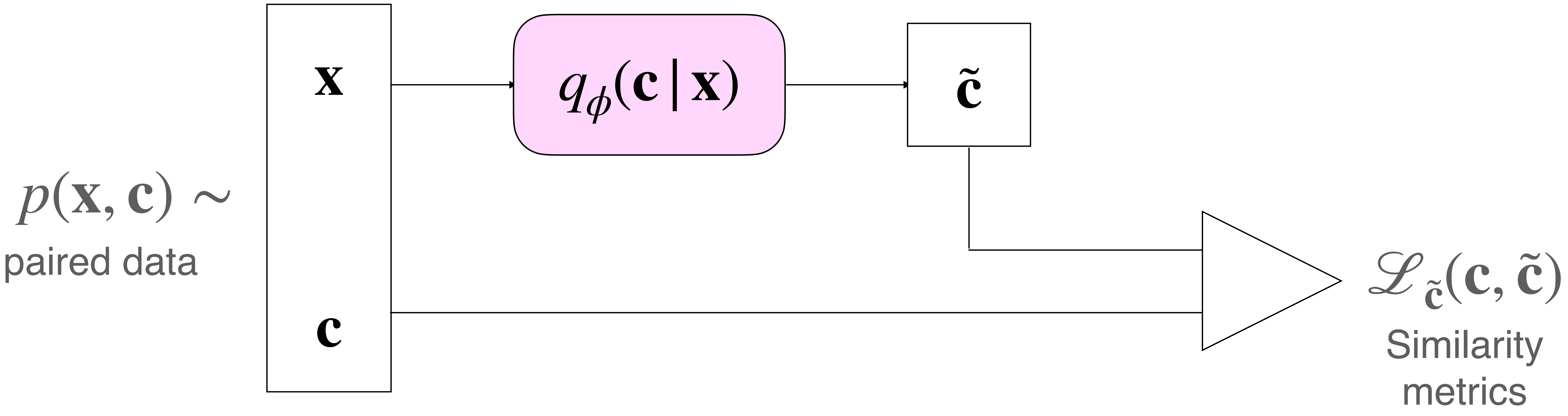
Goal: Learn mapping $\mathbf{x} \rightarrow \tilde{\mathbf{c}}$

Solution: Encoder $q_\phi(\mathbf{c}\,|\,\mathbf{x})$ with MSE loss.



$\mathbf{x}$ $\tilde{\mathbf{c}}$

Can we do better?

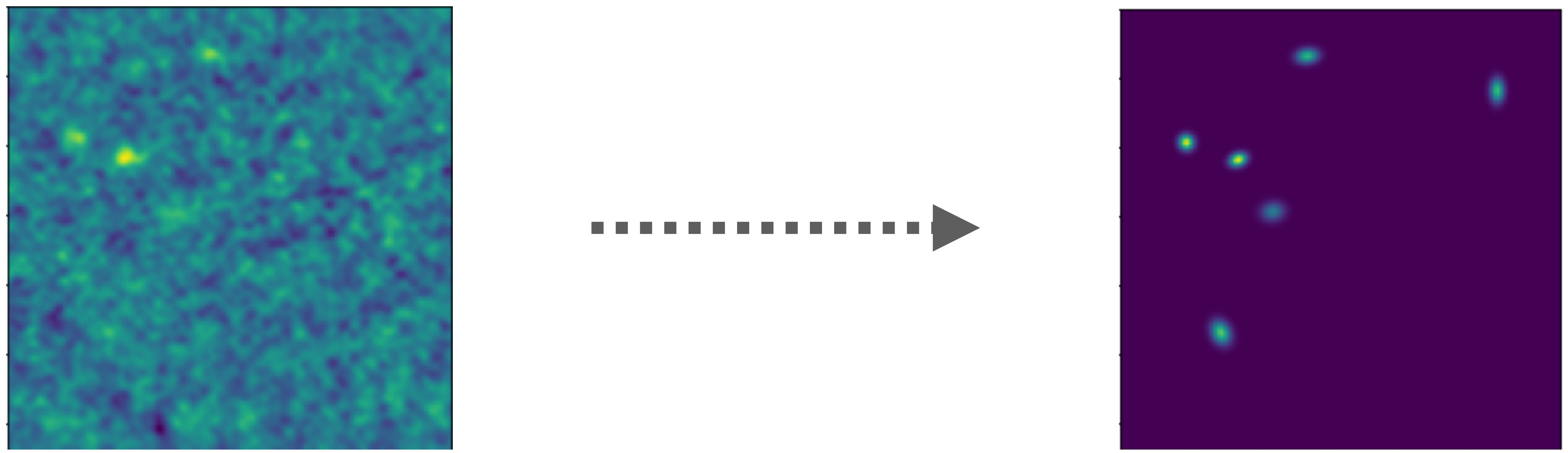


Concept of TURBO

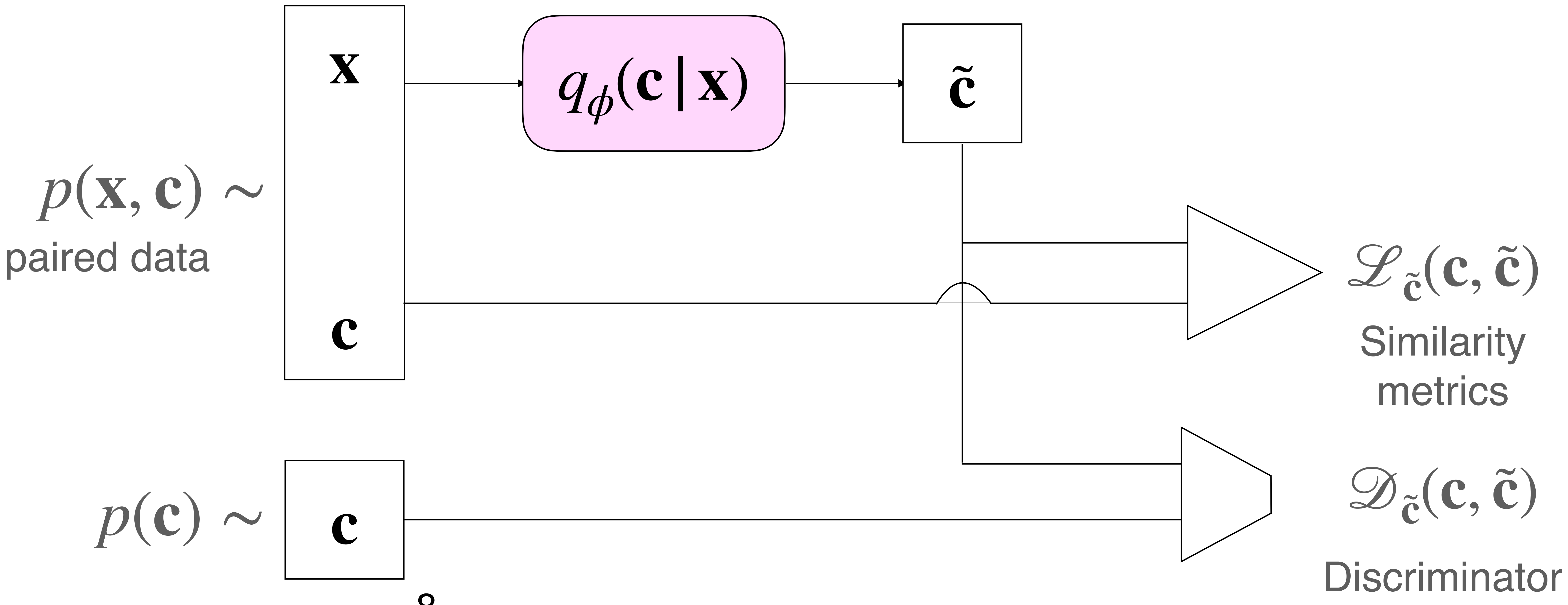
Given: Paired dataset

Goal: Learn mapping $\mathbf{x} \rightarrow \tilde{\mathbf{c}}$

Solution: Encoder $q_\phi(\mathbf{c} \,|\, \mathbf{x})$ with MSE loss and discriminator. (Pix2pix)



$\mathbf{x}$ $\tilde{\mathbf{c}}$



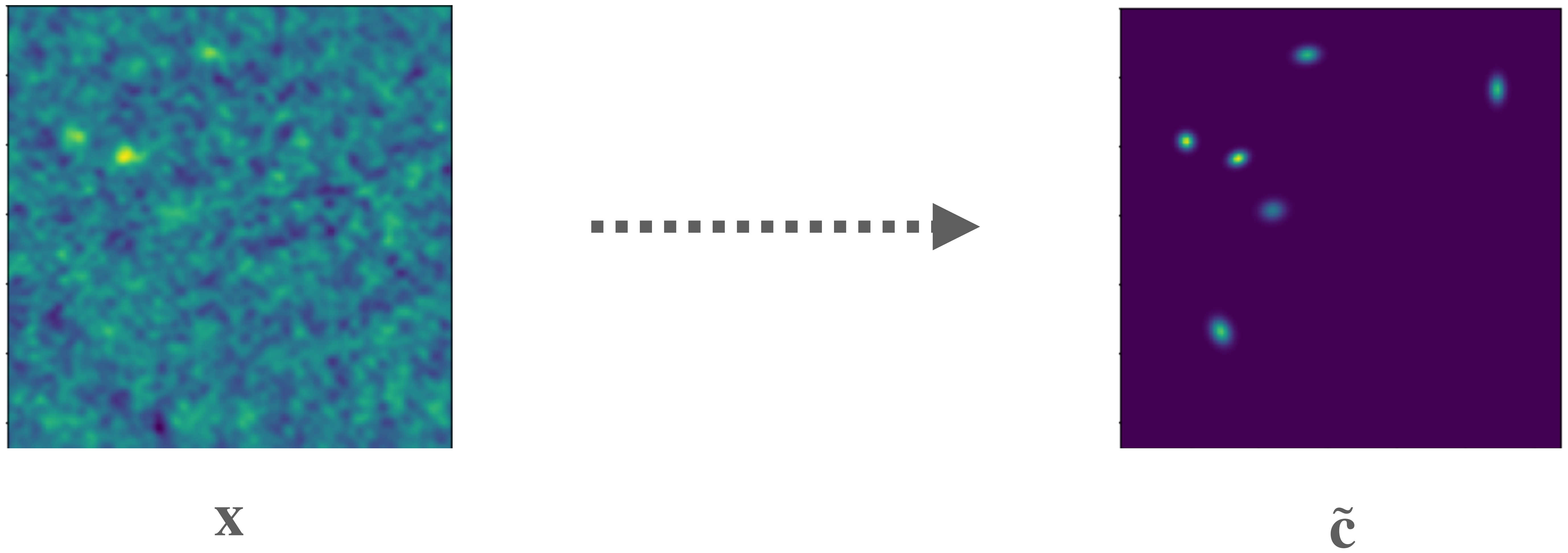
Concept of TURBO

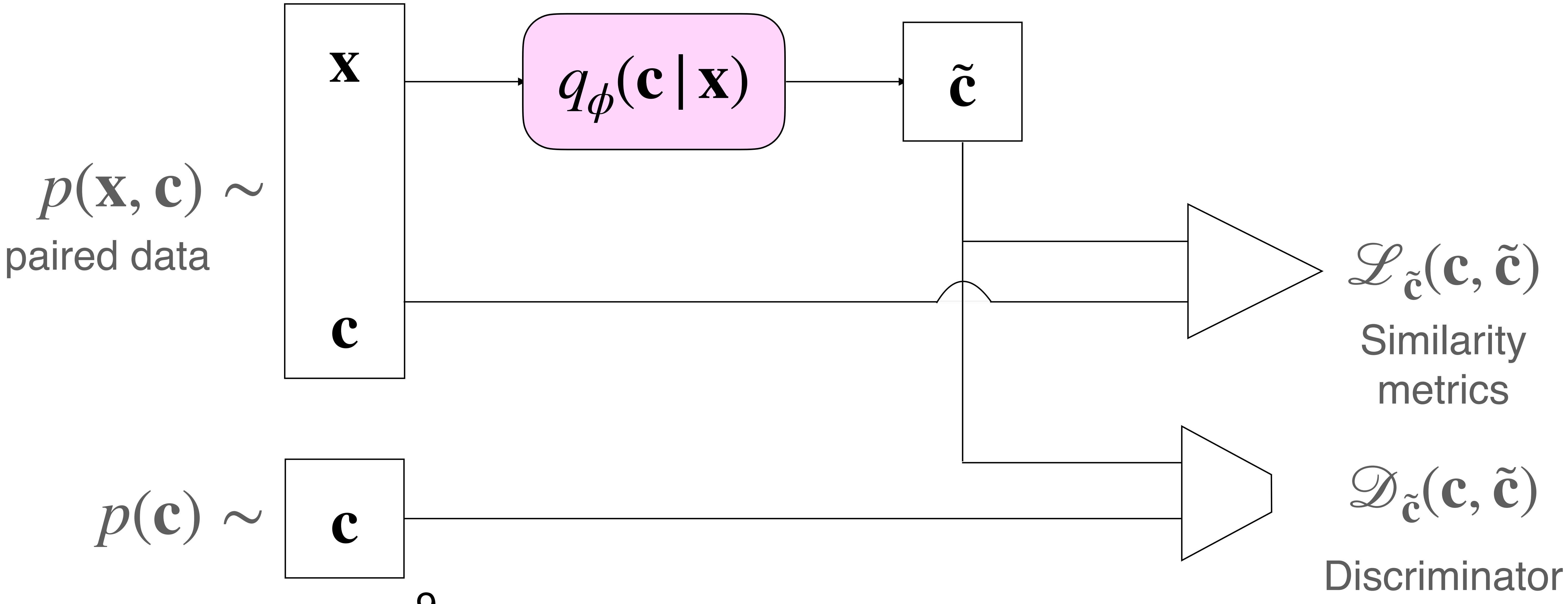
Given: Paired dataset

Goal: Learn mapping $\mathbf{x} \rightarrow \tilde{\mathbf{c}}$

Solution: Encoder $q_\phi(\mathbf{c} \mid \mathbf{x})$ with MSE loss and discriminator. (Pix2pix)

Can we do better?



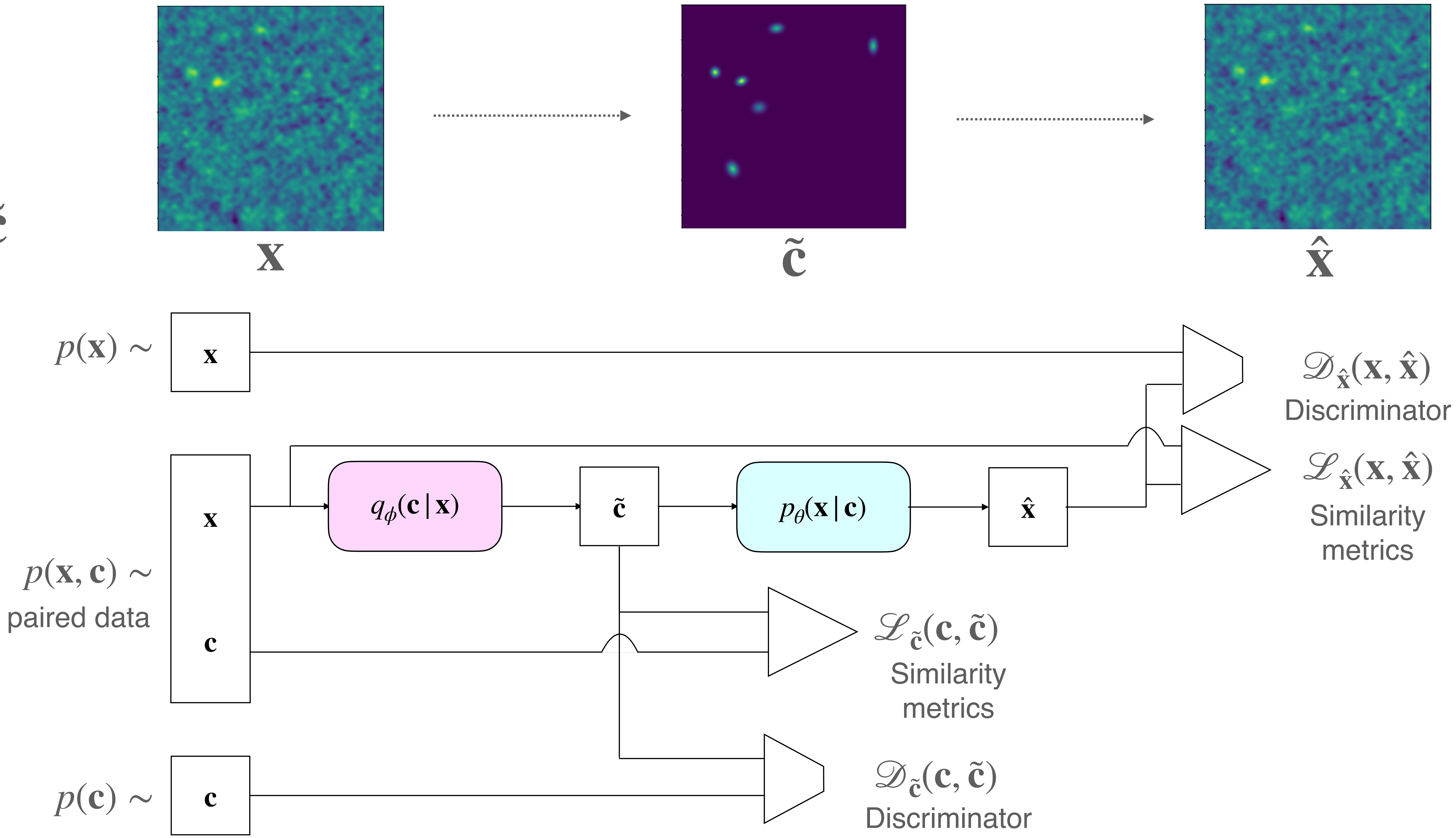


Concept of TURBO

Given: Paired dataset

Goal: Learn mapping $\mathbf{x} \rightarrow \tilde{\mathbf{c}}$

Solution: Two encoders $q_\phi(\mathbf{c}\,|\,\mathbf{x})$ and $p_\phi(\mathbf{x}\,|\,\mathbf{c})$ with MSE loss and discriminator.



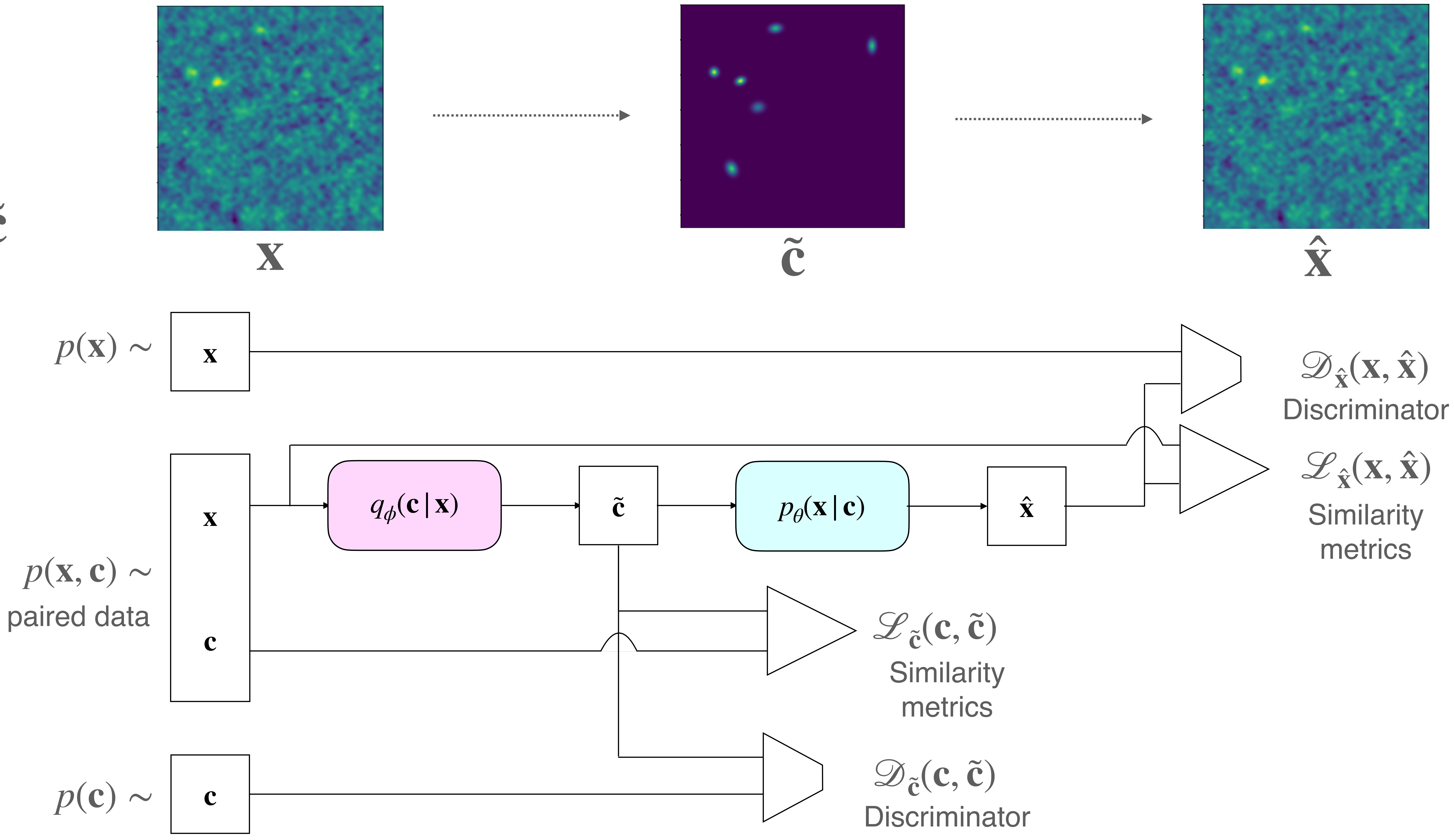
Concept of TURBO

Given: Paired dataset

Goal: Learn mapping $\mathbf{x} \rightarrow \tilde{\mathbf{c}}$

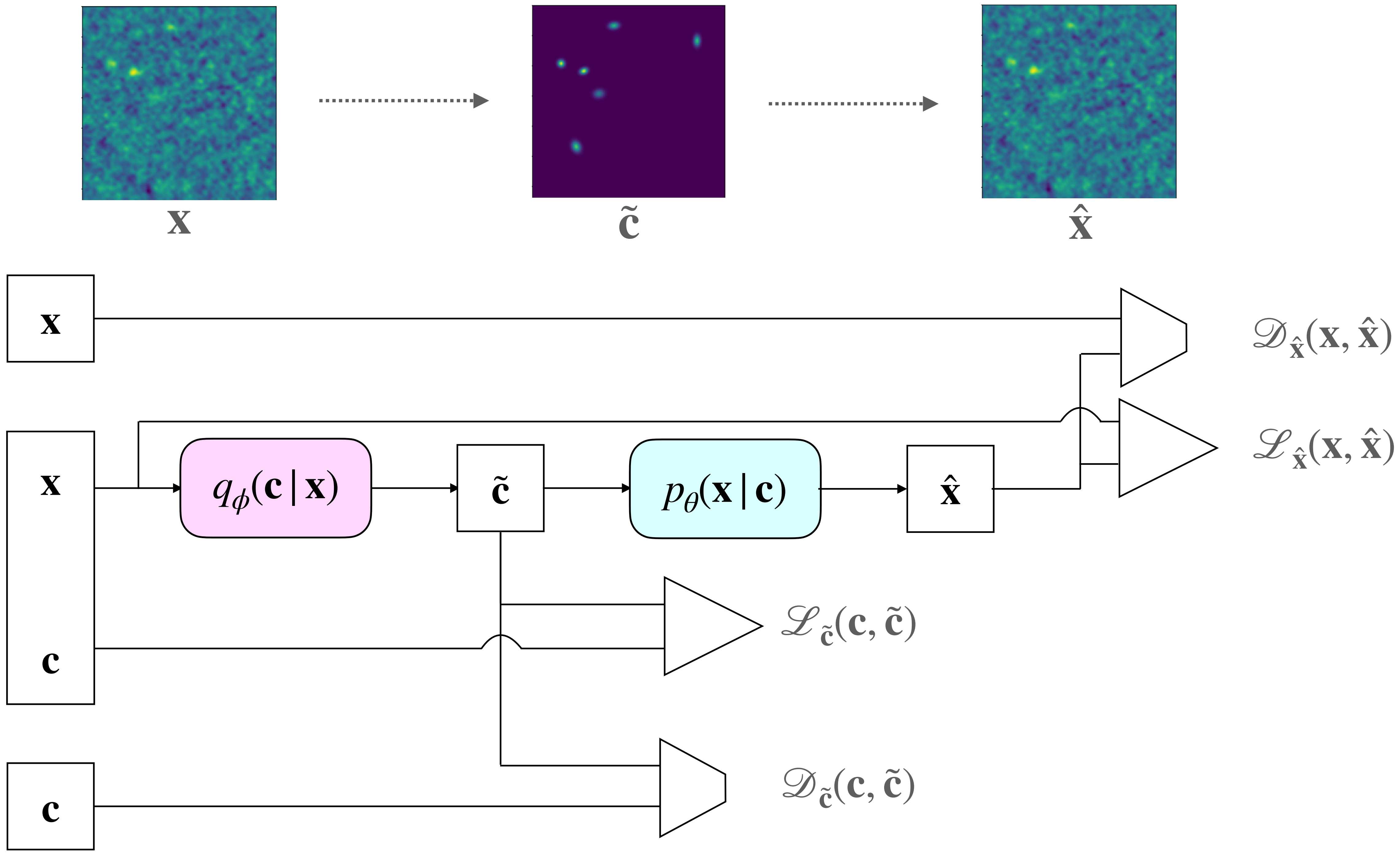
Solution: Two encoders $q_\phi(\mathbf{c}\,|\,\mathbf{x})$ and $p_\phi(\mathbf{x}\,|\,\mathbf{c})$ with MSE loss and discriminator.

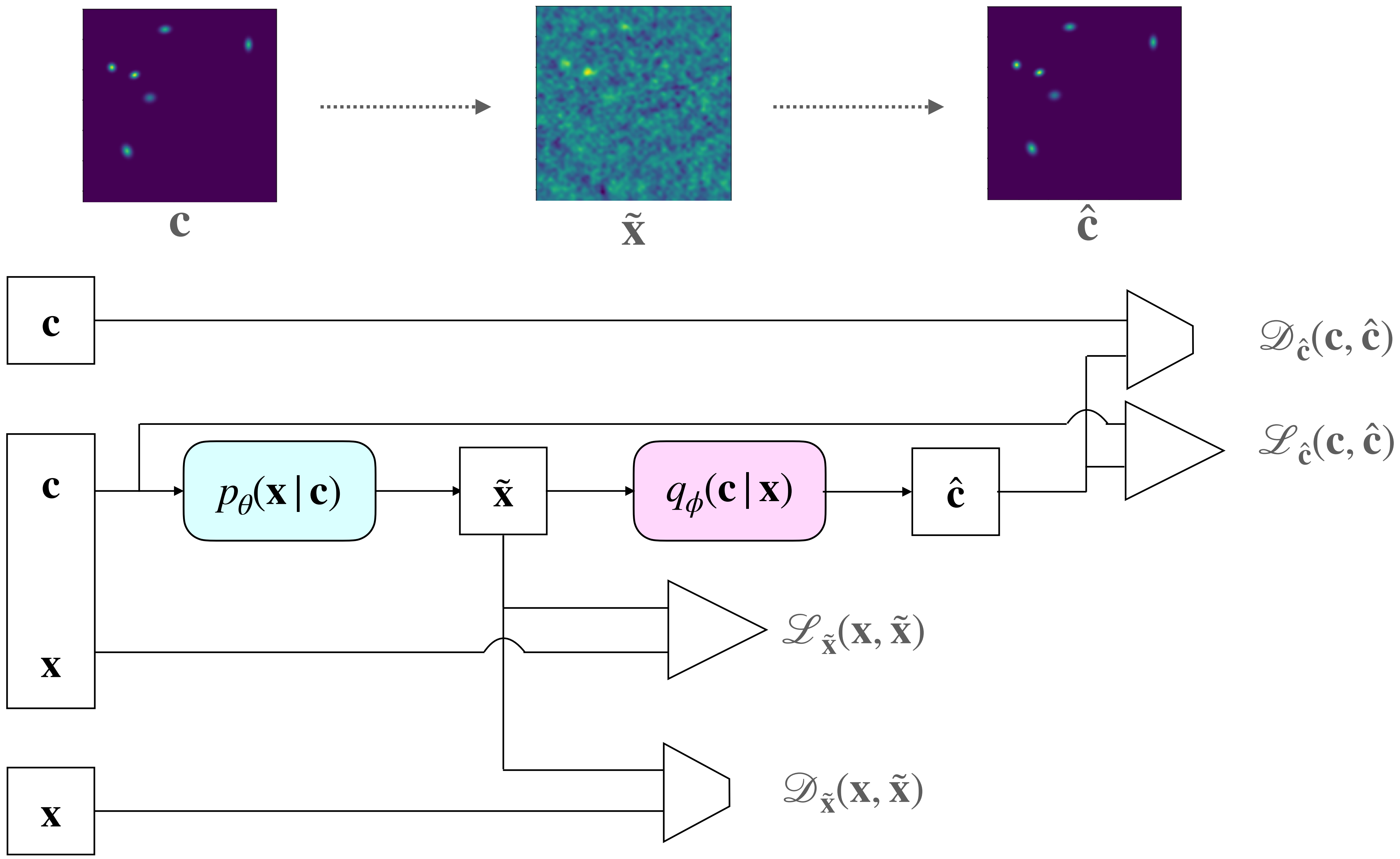
Can we do better?



TURBO

Two branches





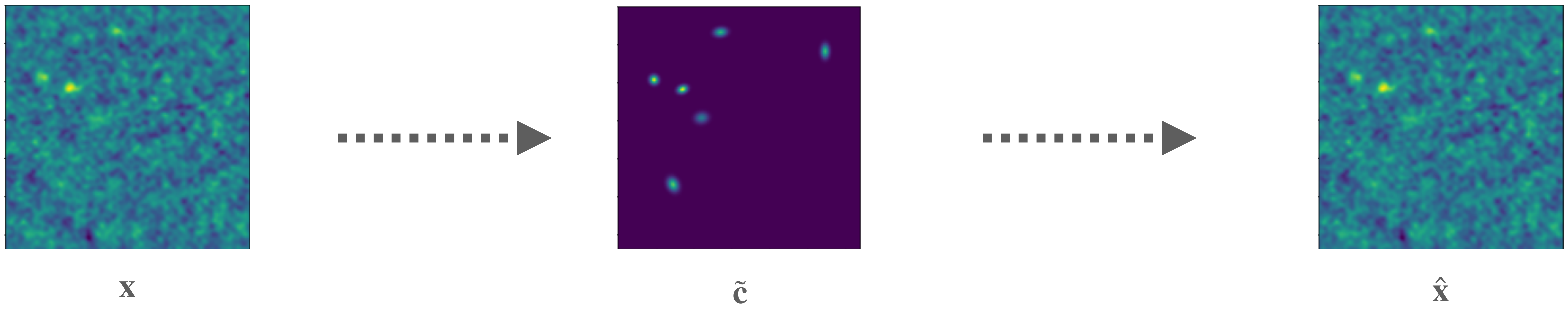
A powerful framework that generalizes pix2pix and CycleGAN.

However, it is **deterministic**.

[Turbo-Sim: a generalised generative model with a physical latent space, G.Quetant, M.D. et al., 2021 NeurIPS Workshop](#)

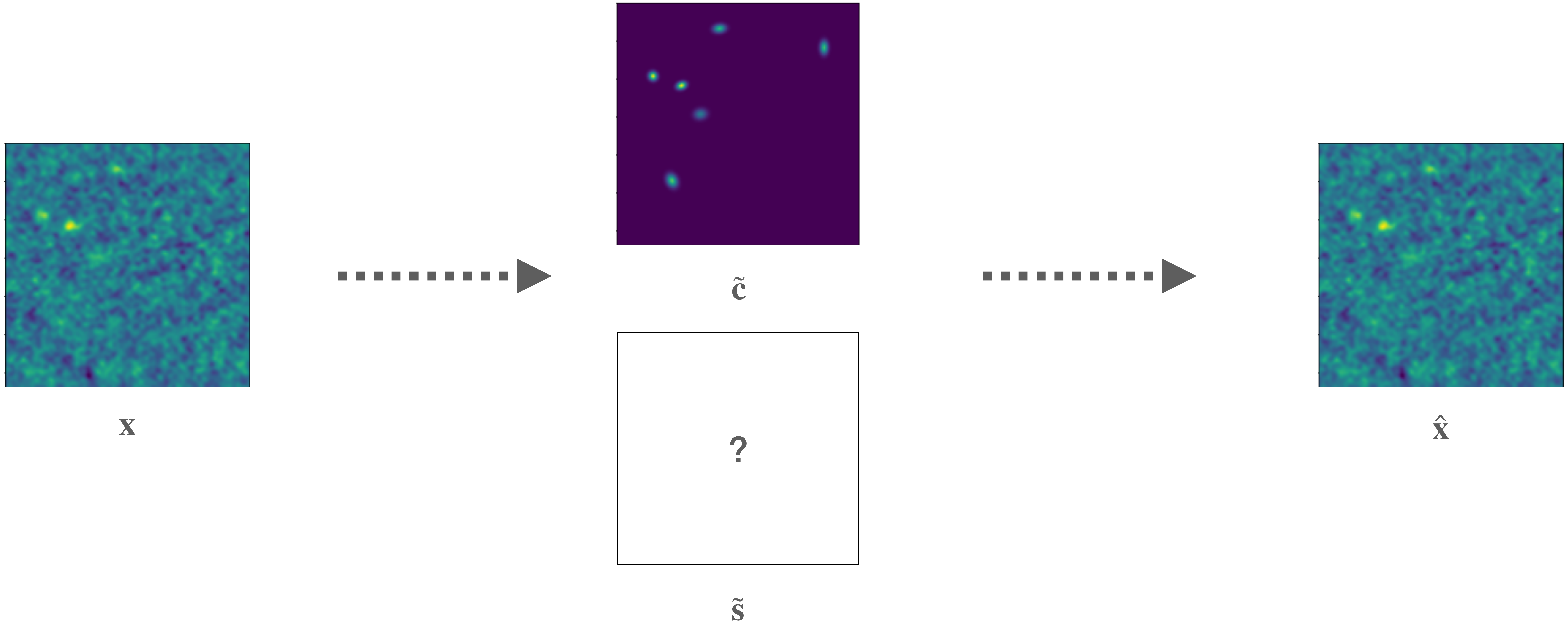
Stochastic TURBO

Suggestion: Disentangling content and style.



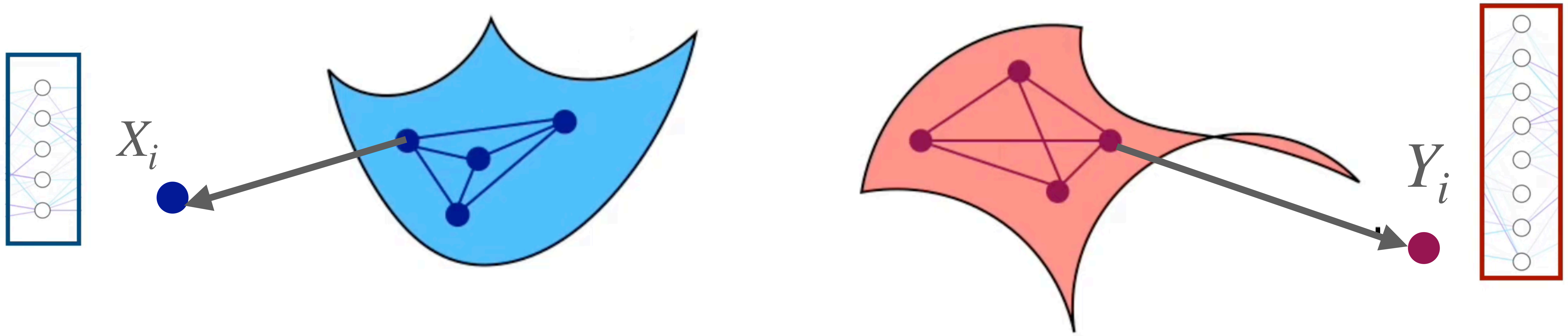
Stochastic TURBO

Suggestion: Disentangling content and style.



Goal: the content $\tilde{\mathbf{c}}$ and style $\tilde{\mathbf{s}}$ should be disentangled ($\tilde{\mathbf{c}} \perp\!\!\!\perp \tilde{\mathbf{s}}$)
Question: how to achieve it in practice without adversarial games?

Distance correlation: a basic concept



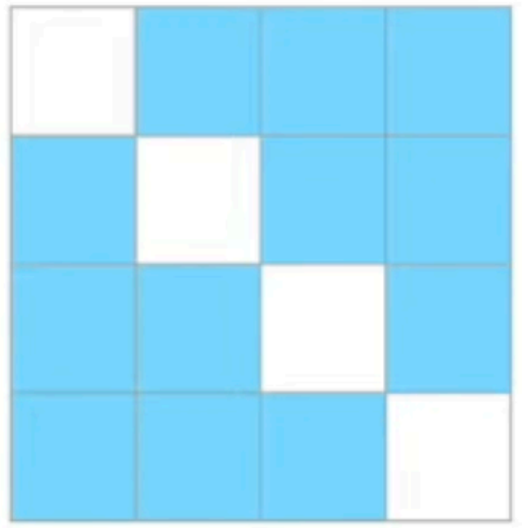
$(X_i, Y_i), 1 \leq i \leq m$

Distance matrix: $a_{k,l} = ||X_k - X_l||$

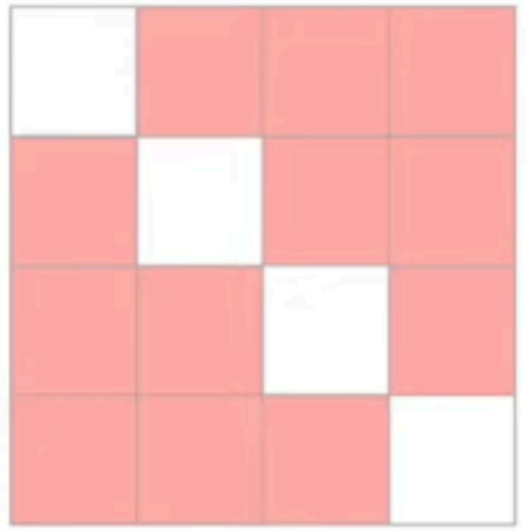
Normalization: $A_{k,l} = a_{k,l} - \bar{a}_{k,\cdot} + \bar{a}_{\cdot,\cdot}$

$b_{k,l} = ||Y_k - Y_l||$:

$B_{k,l} = b_{k,l} - \bar{b}_{k,\cdot} + \bar{b}_{\cdot,\cdot}$

$A =$ 

Distance correlation

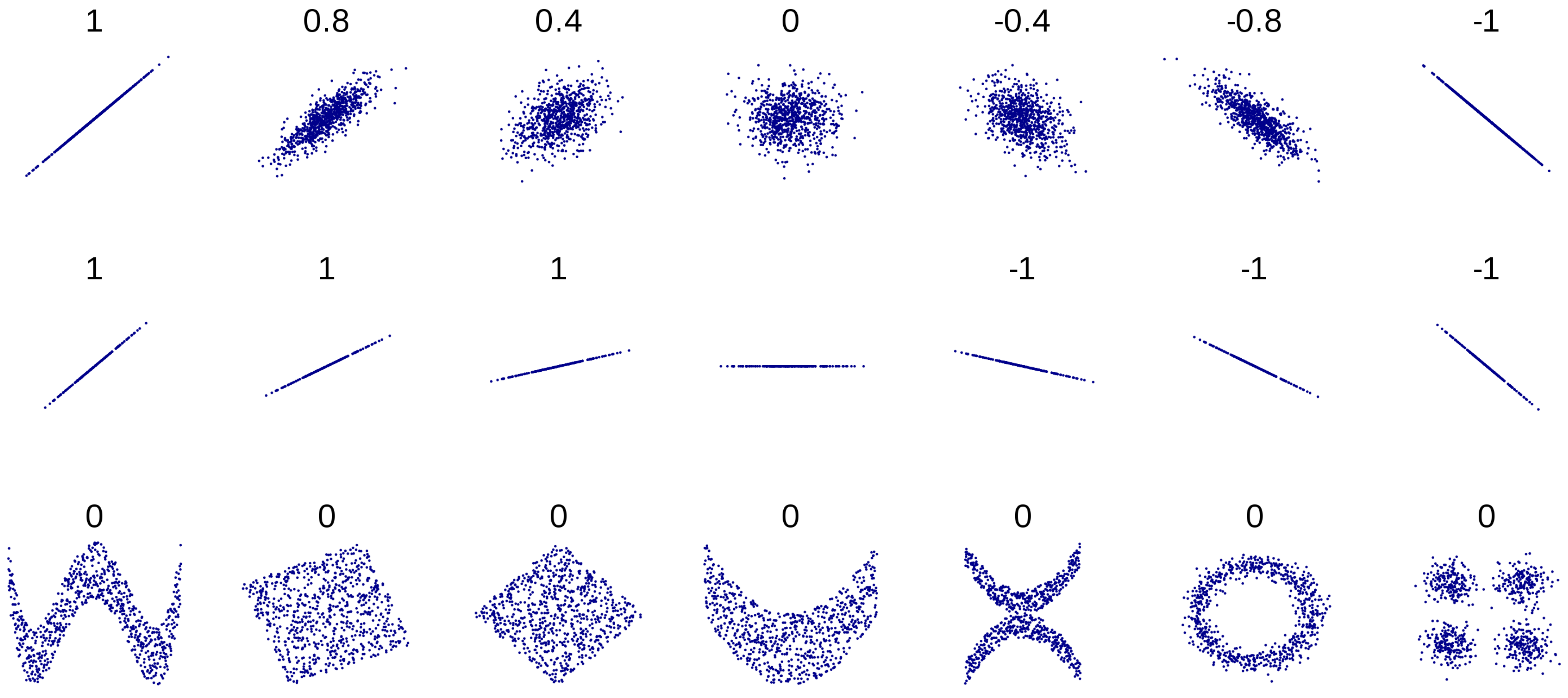
$B =$ 

$$\text{dcor}_{X,Y}(X, Y) = \frac{\sum_{k,l} A_{k,l} B_{k,l}}{\sqrt{(\sum_{k,l} A_{k,l} A_{k,l})(\sum_{k,l} B_{k,l} B_{k,l})}}$$

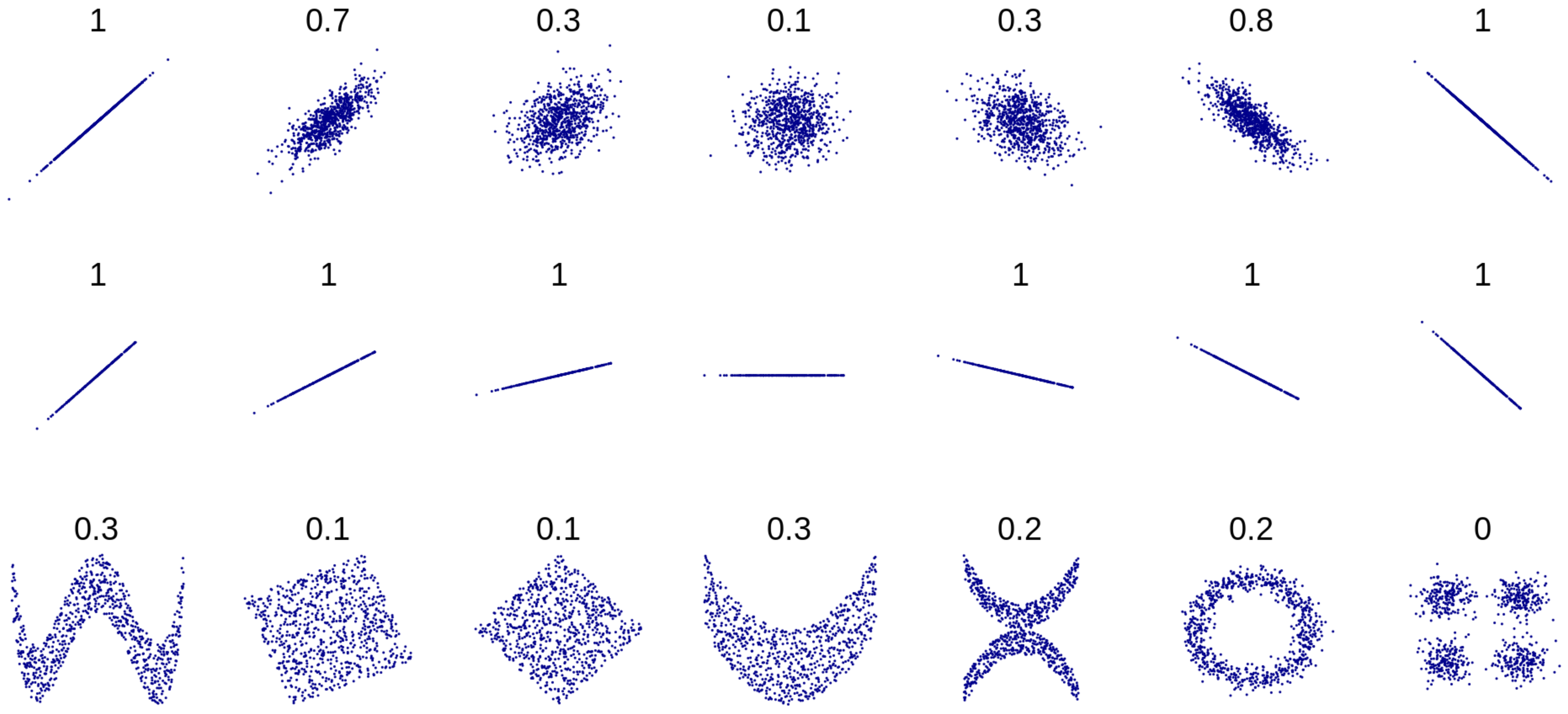
On the Versatile Uses of Partial Distance Correlation in Deep Learning, Zhen et al. 2022, Best Award Paper 2022

Distance correlation: basic properties

- Dependence measure between two vectors (linear correlation)
- Vectors can be of **different dimensions**
- Between 0 and 1



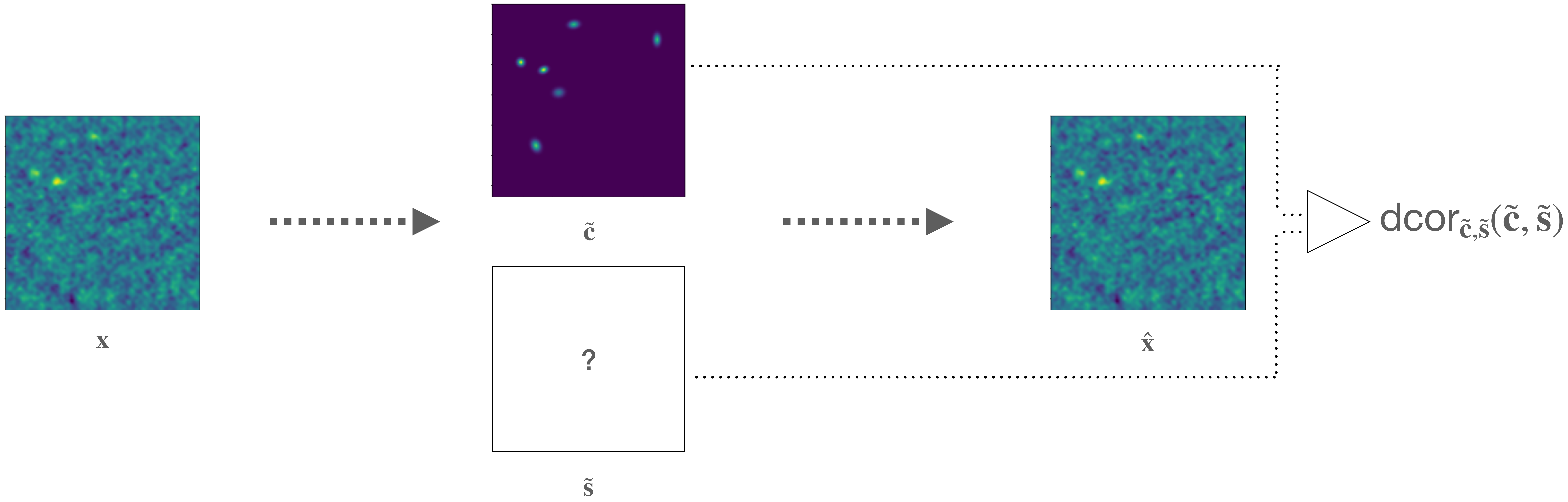
Pearson correlation



Distance correlation

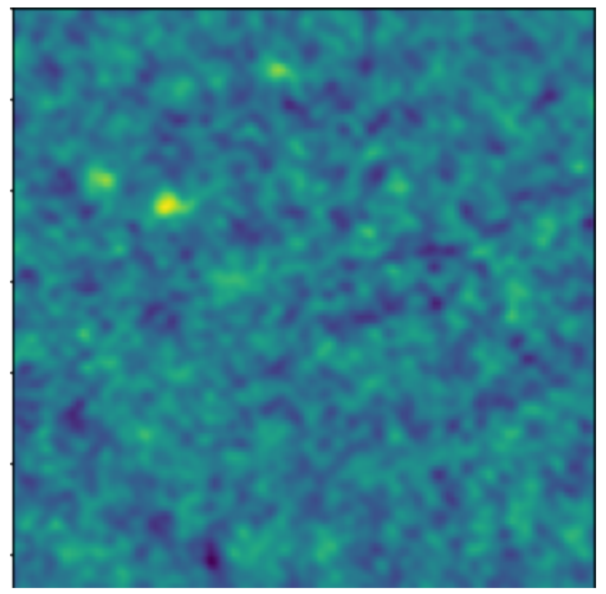
Stochastic TURBO

Suggestion: Disentangling content and style based on dCor.

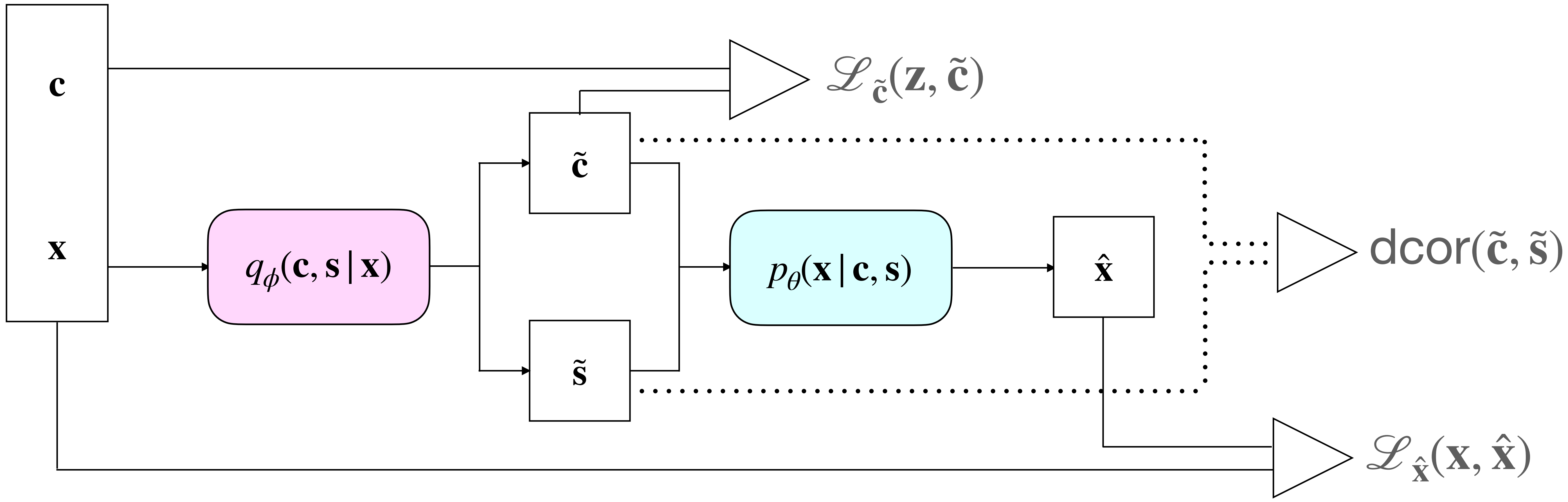


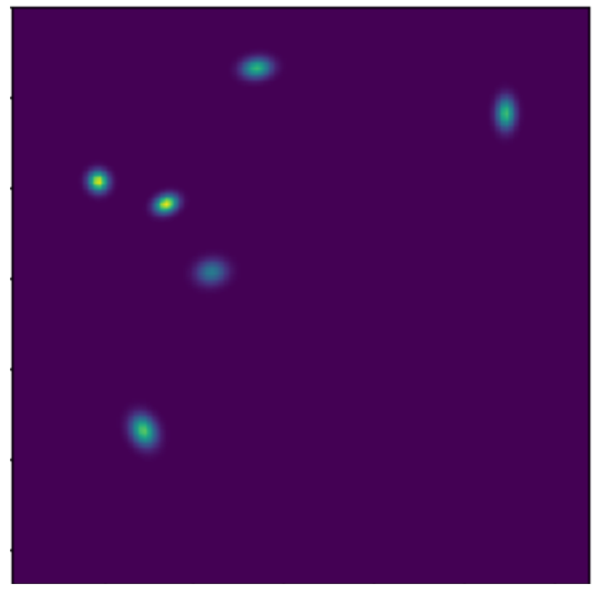
This framework overcomes limitations of deterministic Turbo and DDPM by combining their strengths.

STAR model (simplified without discriminators)

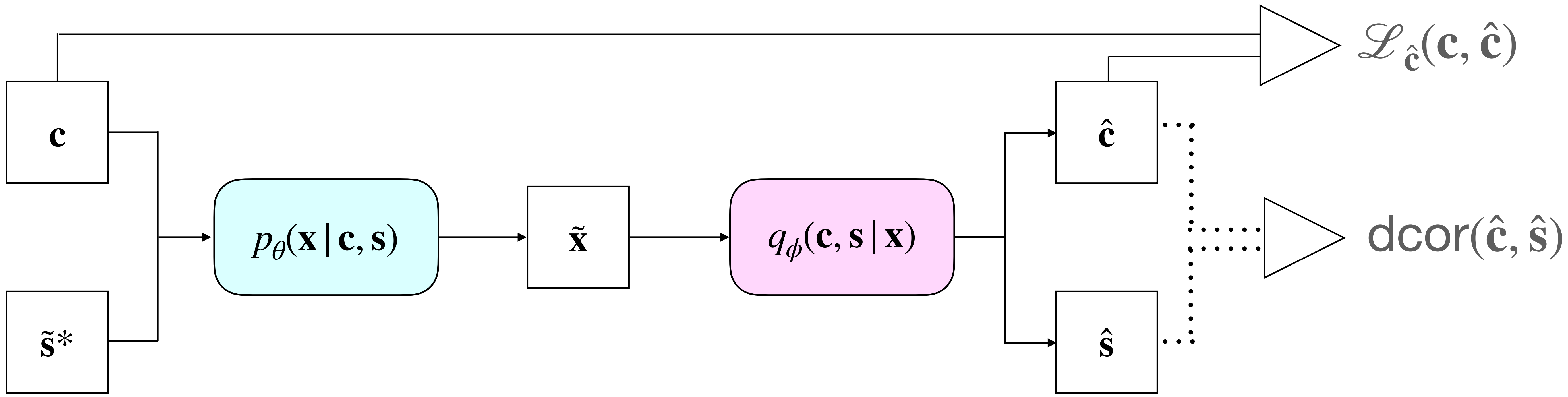


x

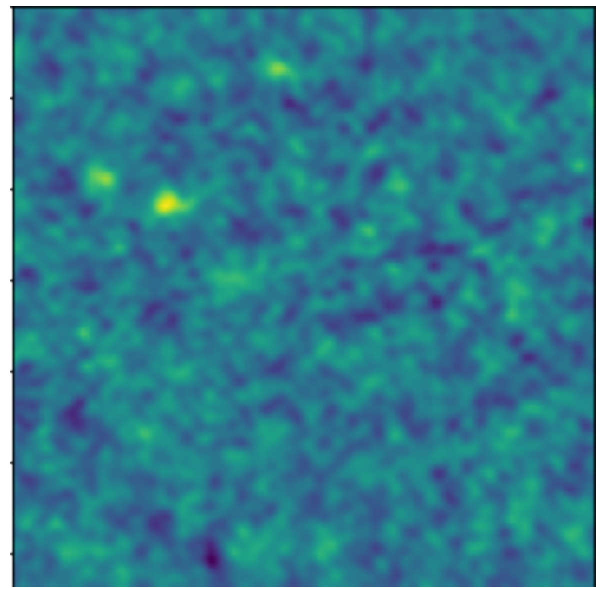




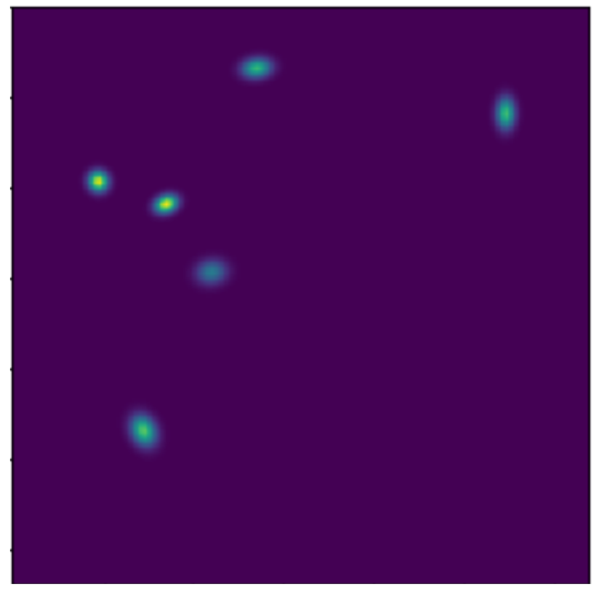
c



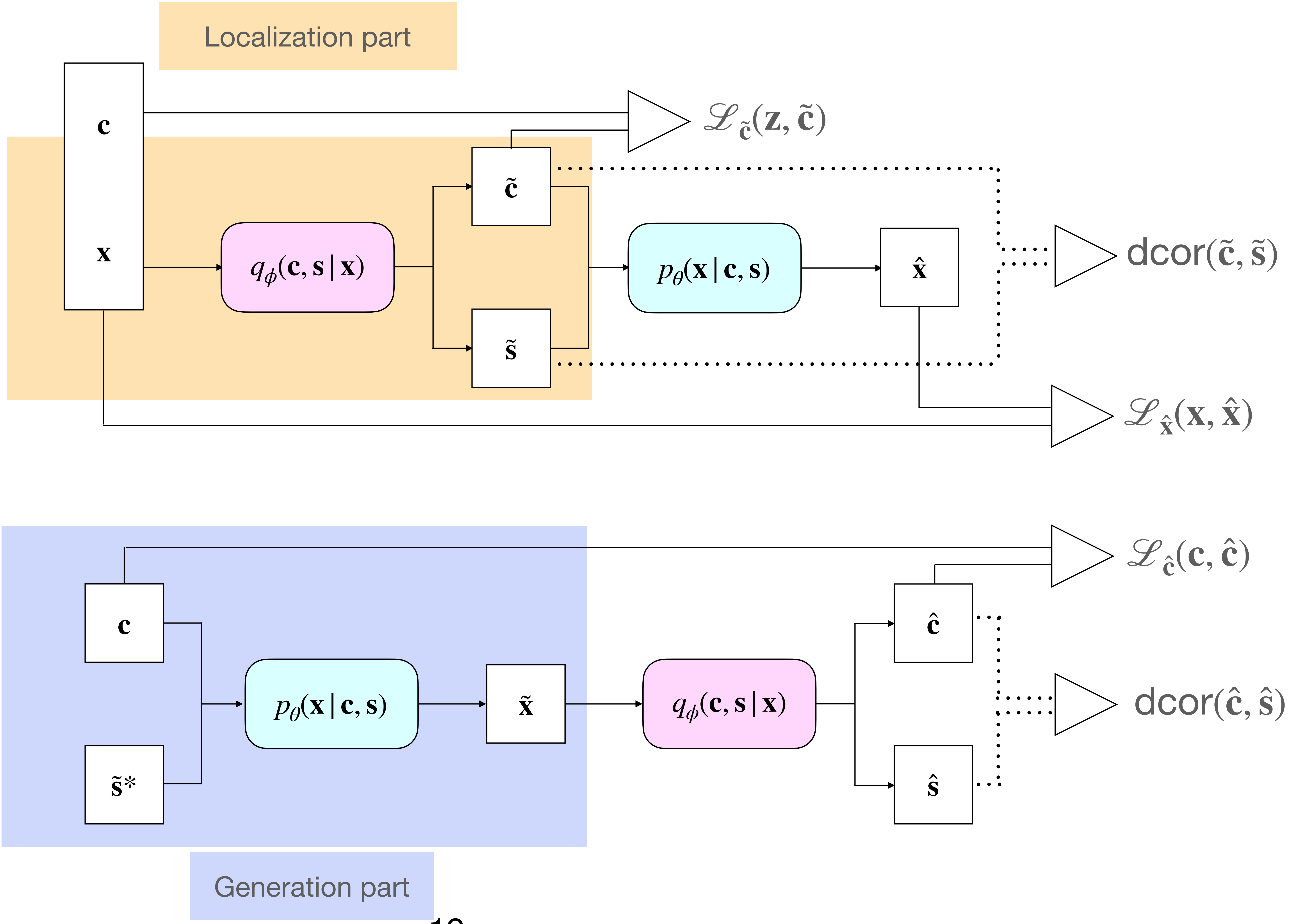
STAR model



$\mathbf{x}$



$\mathbf{c}$



Results: Source detector

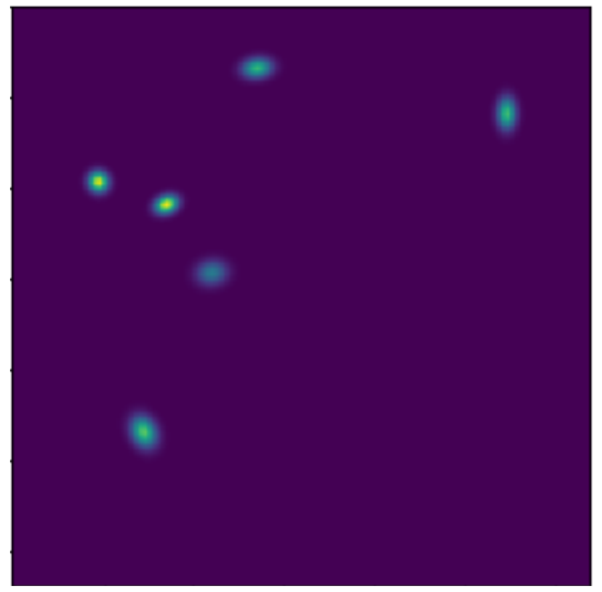
	Purity	Completeness	F1 Score	L2	Network runs
Only MSE	93.23	93.60	93.41	1.61E-05	1
DDPM	99.30	97.05	**98.16**	1.19E-05	250*20
STAR (MSE, without dcor)	91.55	93.63	92.58	1.69E-05	1
STAR (MSE, dcor)	93.98	94.09	94.03	1.56E-05	1
STAR (MSE, dcor, 2 branches)	**97.16**	**93.39**	**95.23**	**1.62E-05**	**1**

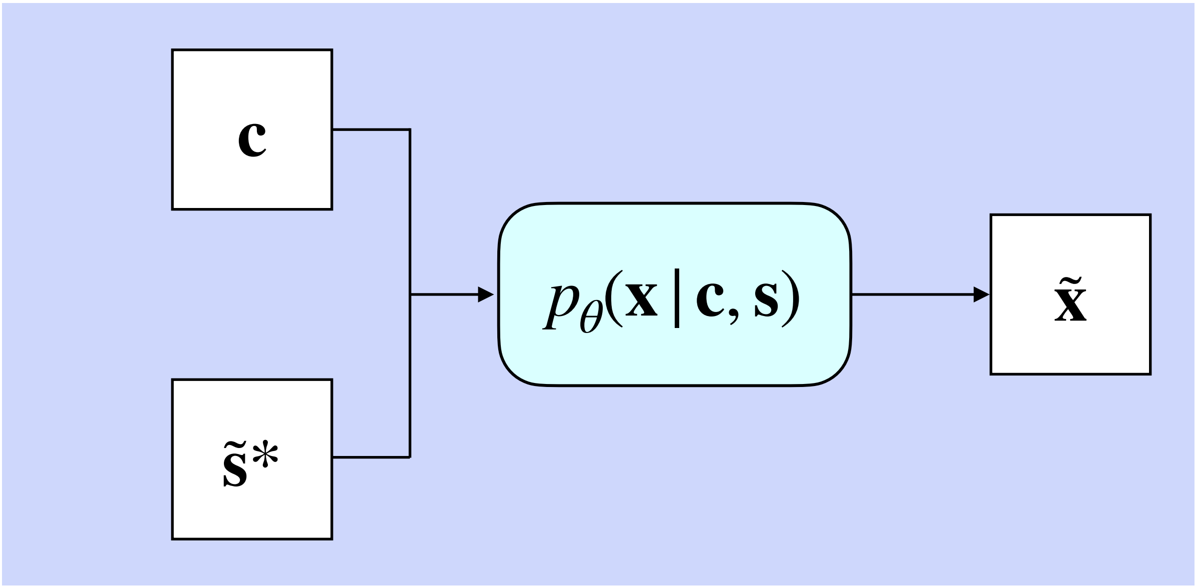
STAR:
- encoder 13.85M parameters
- decoder 10.96M parameters
- one run

DDPM:
- UNet 36M parameters
- 250 steps
- 20 runs

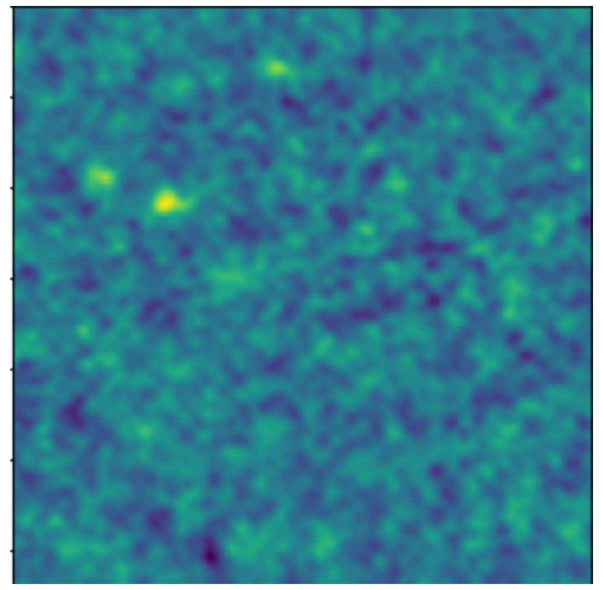
Generator

Sky model (CASA synthetic)

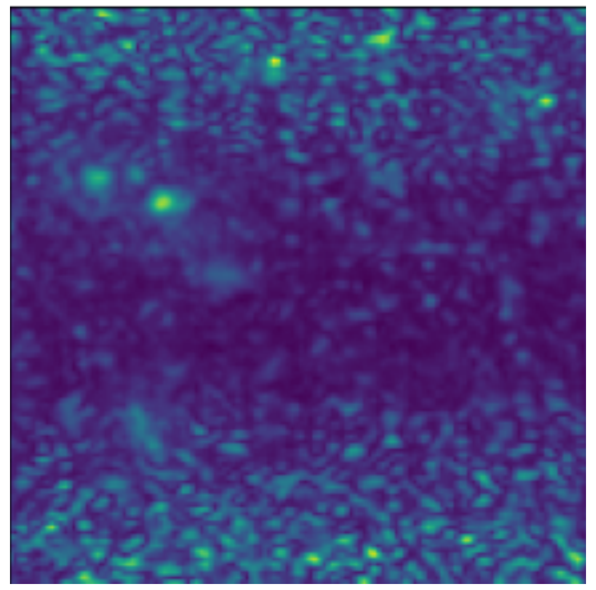




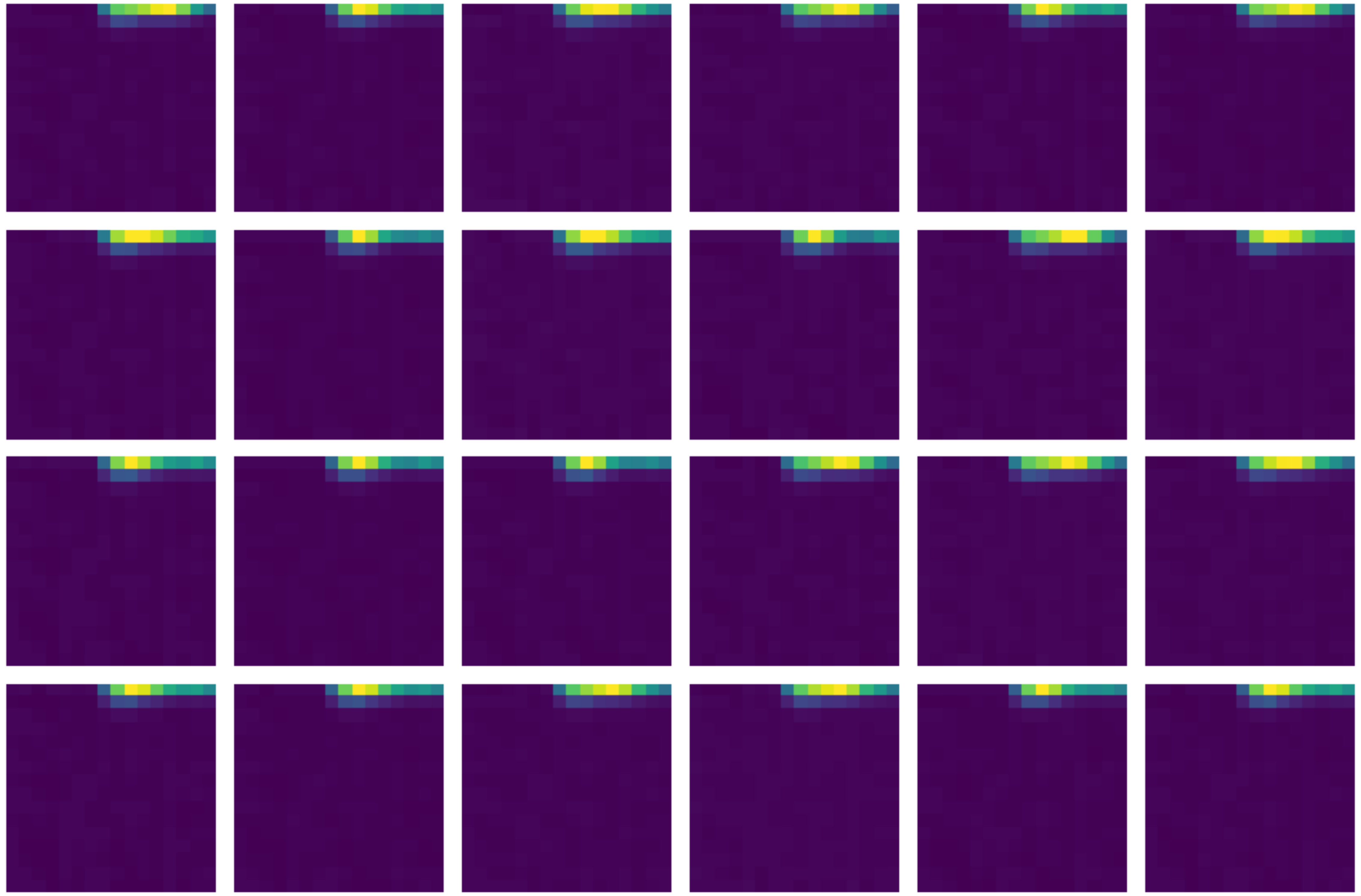
Mean of generated



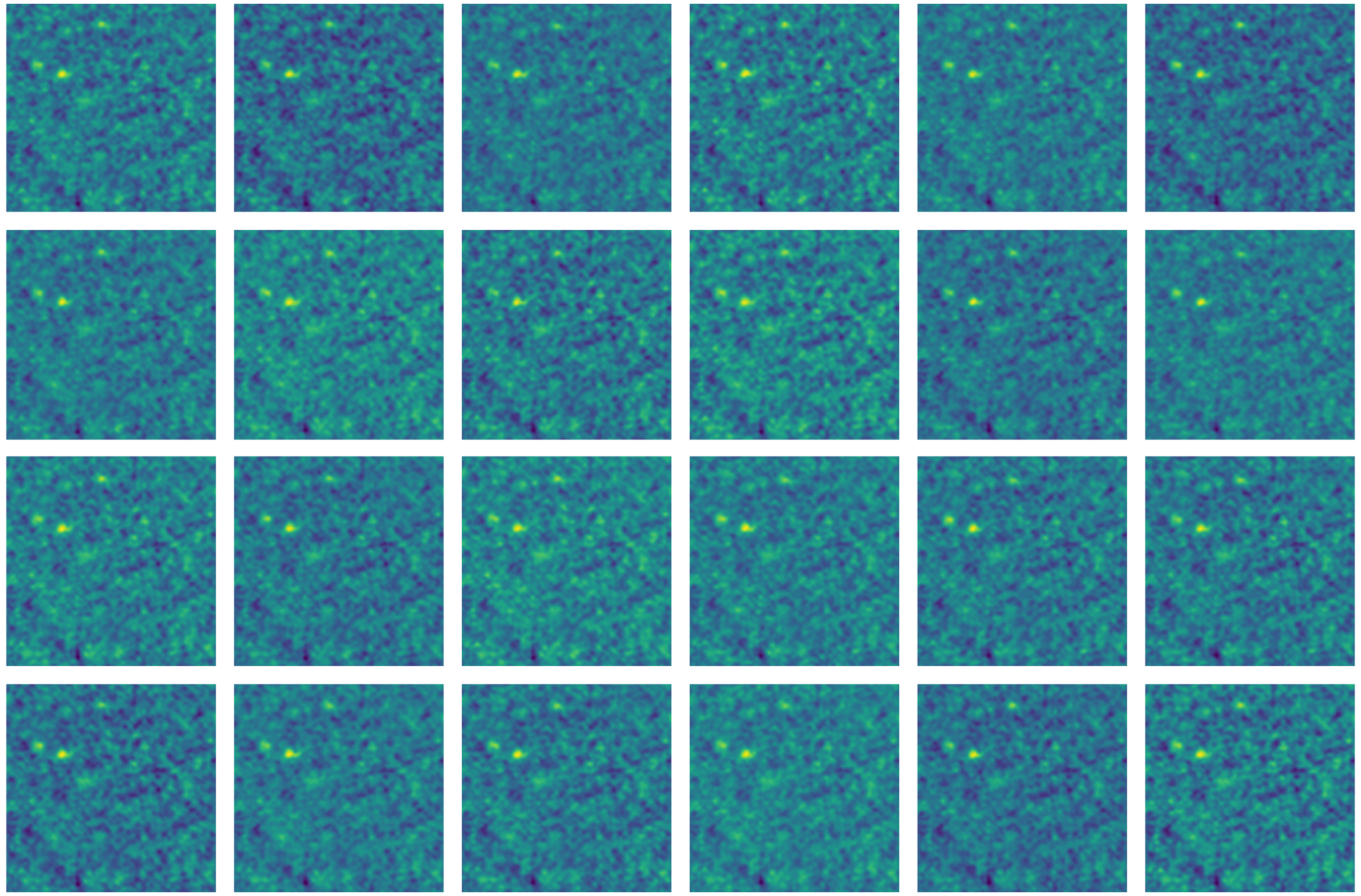
Std



Styles

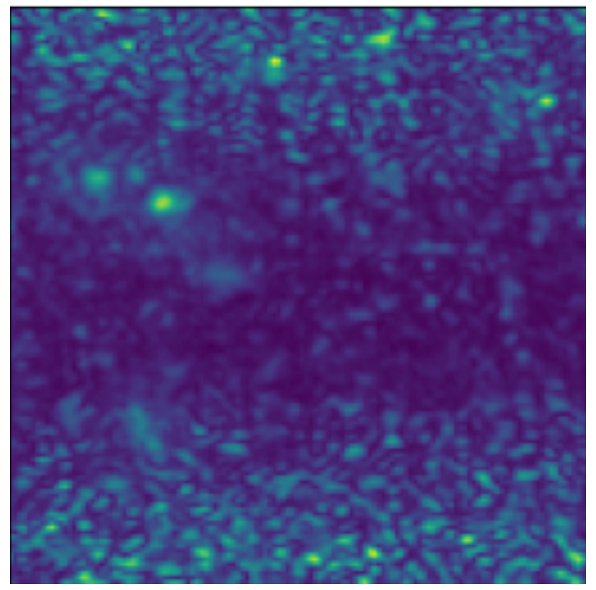


Dirty Images (generated)

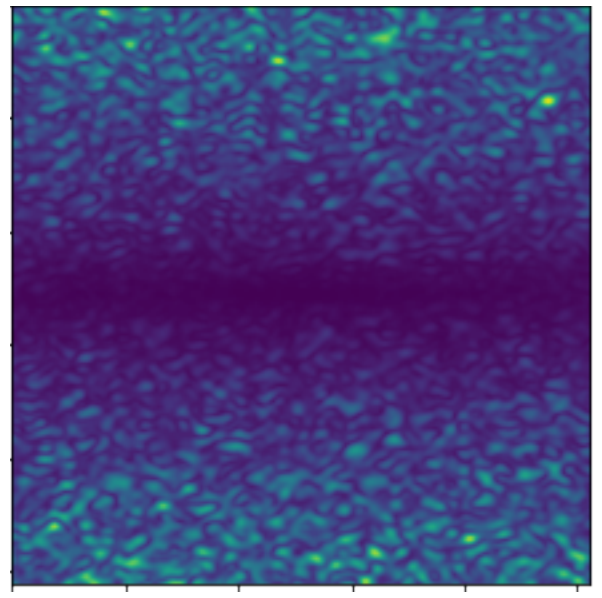


CASA generated data statistics

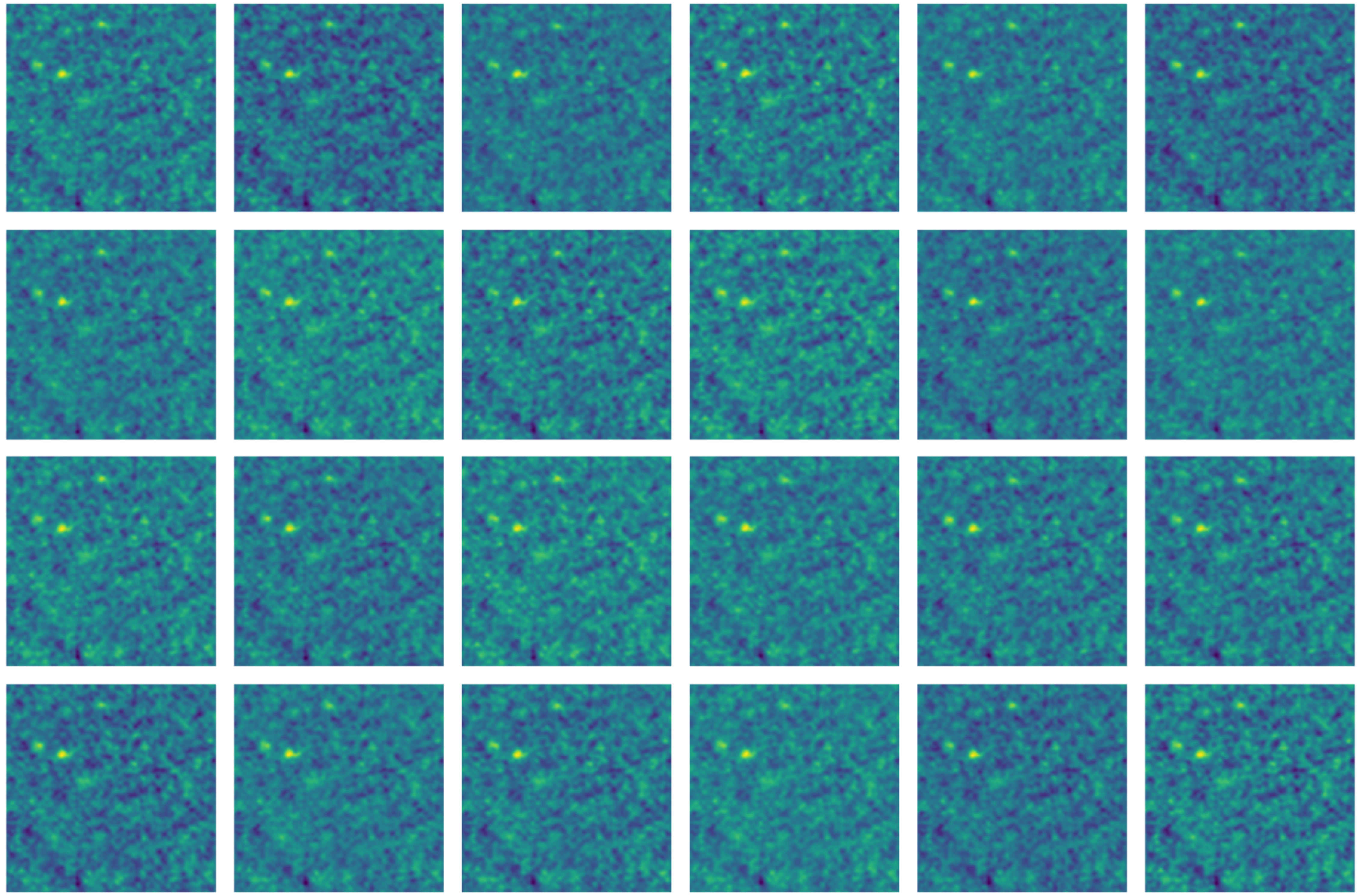
Std



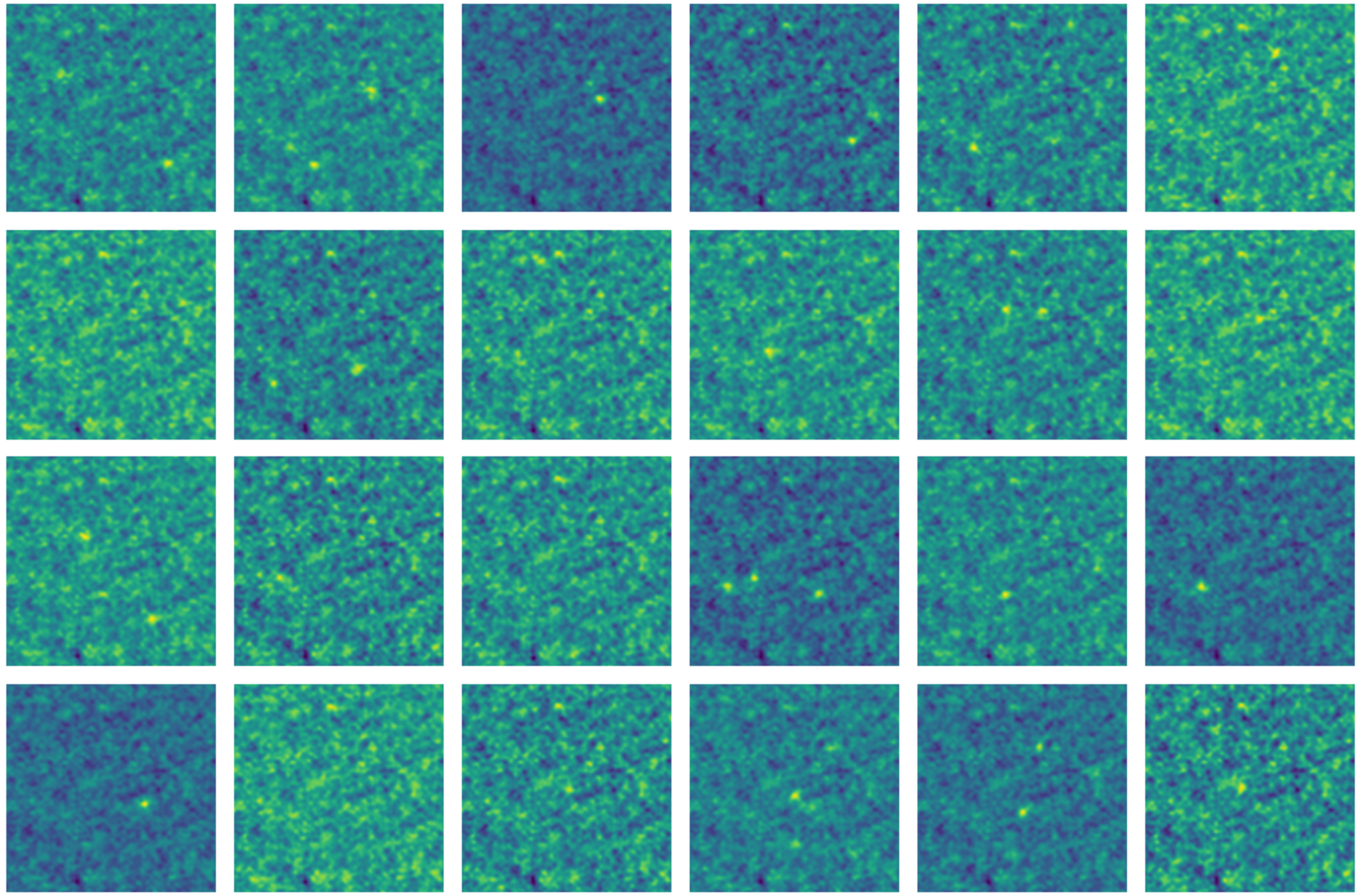
Std (CASA generated data)



Dirty Images (generated)



Dirty Images (CASA generated data)



Conclusions and future work

- We introduced a stochastic generative model with a low complexity comparing to DDPM while maintaining high localization performance.
- STAR showed the ability to generate additional synthetic data.
- This hybrid model can be improved further by studying combinatorics for different losses
- We plan to add a generative network for styles (Implicit Autoencoder, FLOWs etc.)
- Investigate other methods for disentanglement between content and style.

Thank you! Questions?