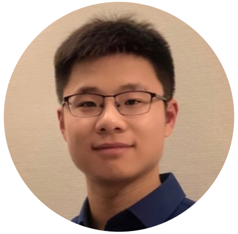
ICLR
International Conference On Learning Representations

How I Warped Your Noise:
A Temporally-correlated Noise Prior for Diffusion Models

ICLR 2024 (Oral)


Pascal Chang


Jingwei Tang

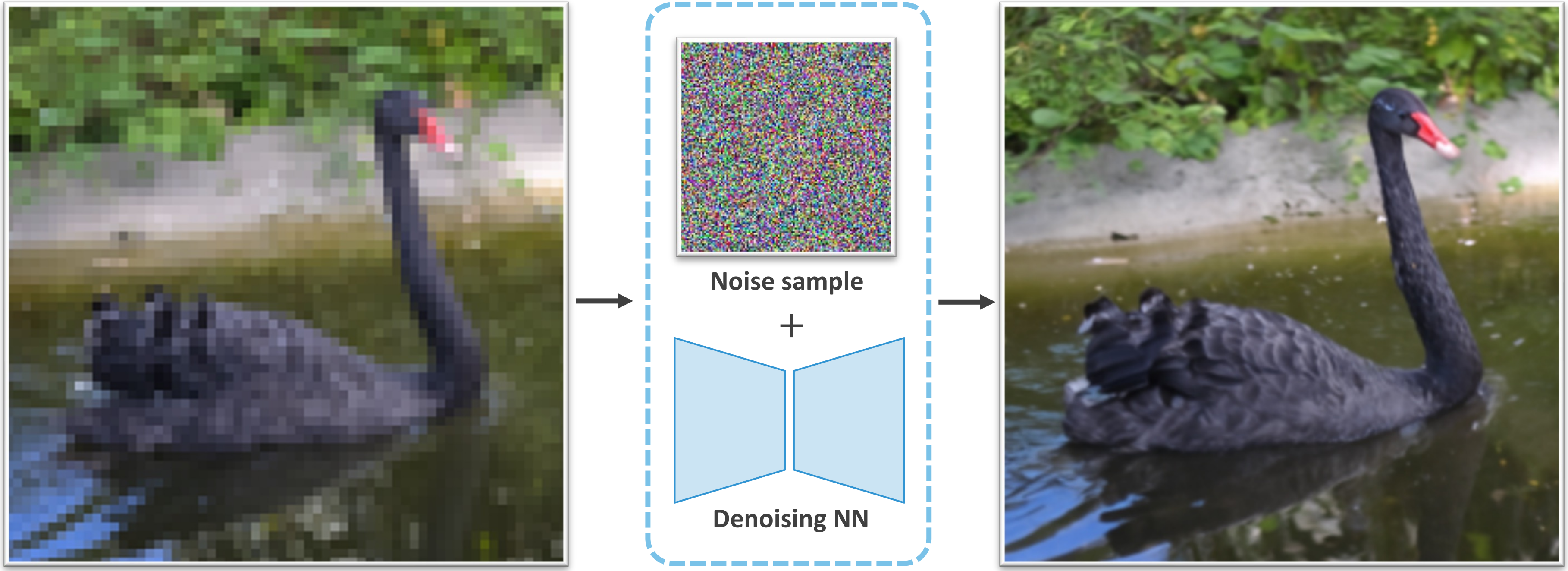

Markus Gross


Vinicius C. Azevedo

ETH zürich

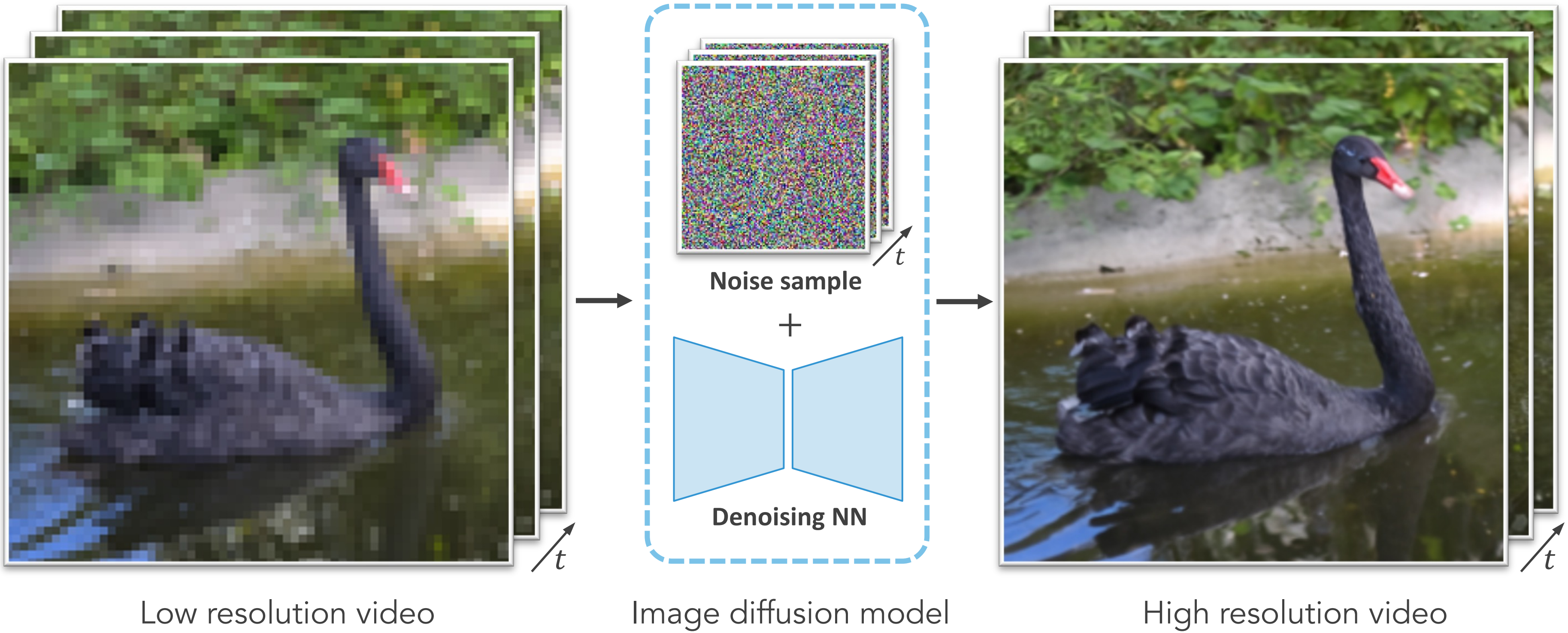
cgl computer graphics laboratory

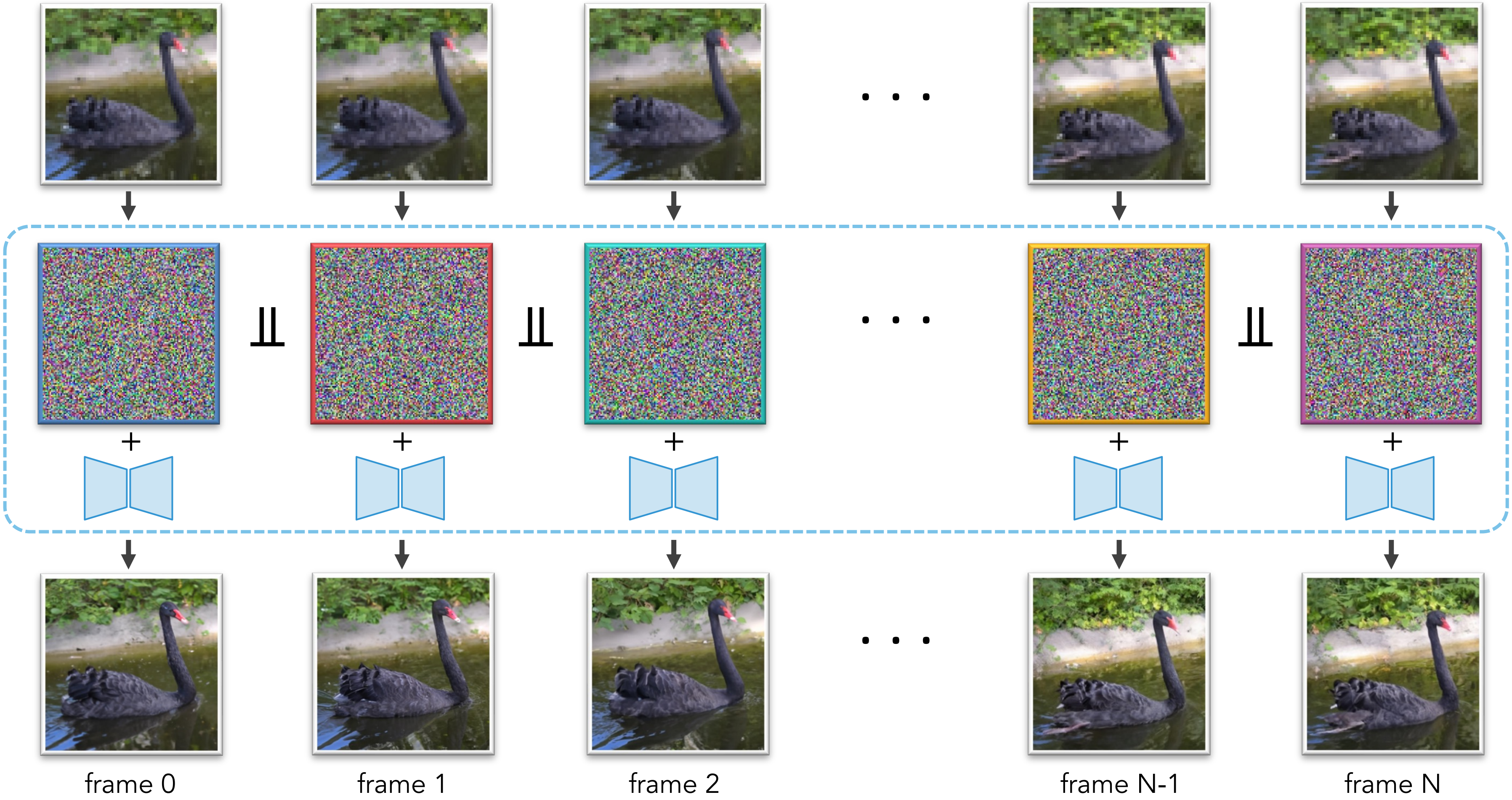
From diffusion image translation…



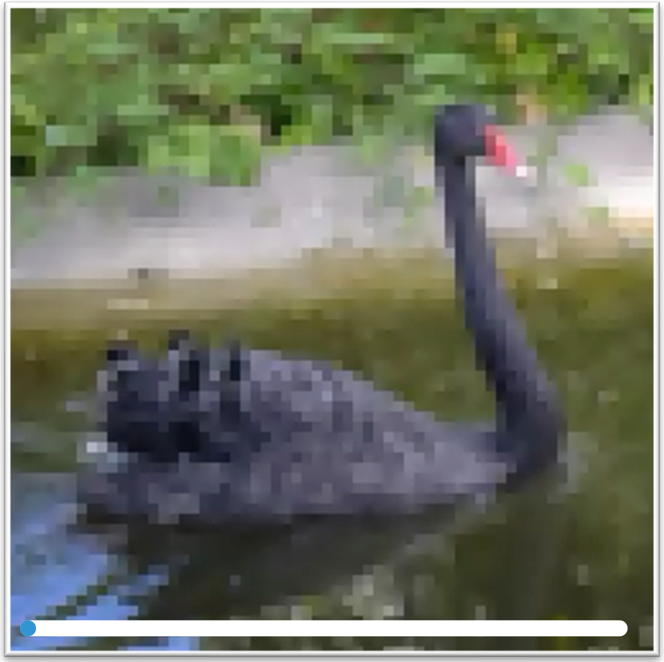
Low resolution image Image diffusion model High resolution image

...To zero-shot video translation




frame 0
frame 1
frame 2
frame N-1
frame N

…To zero-shot video translation



Low resolution video

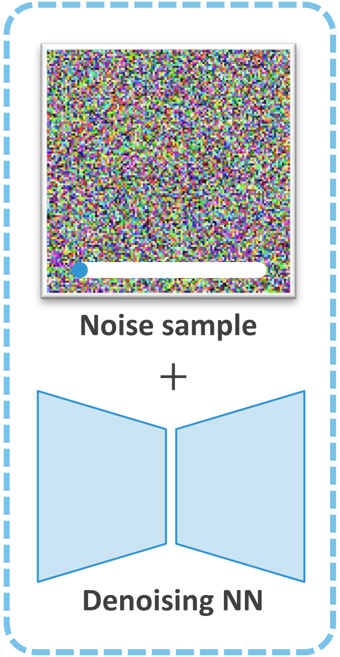
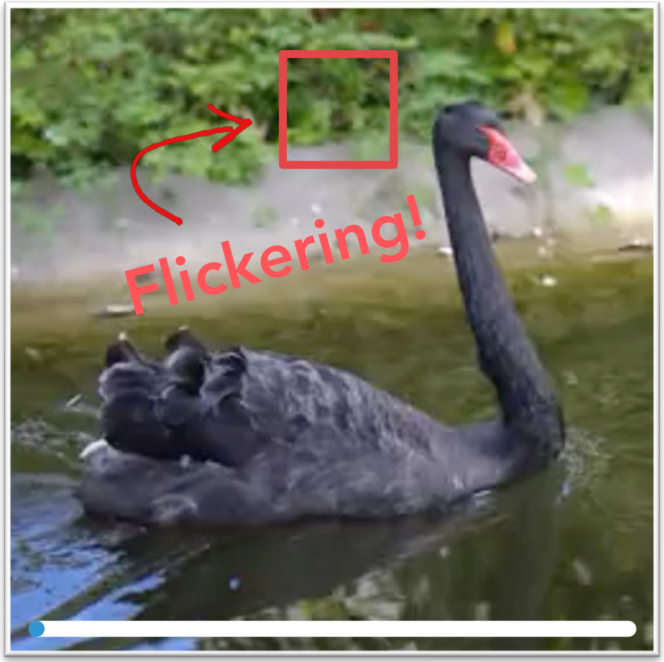
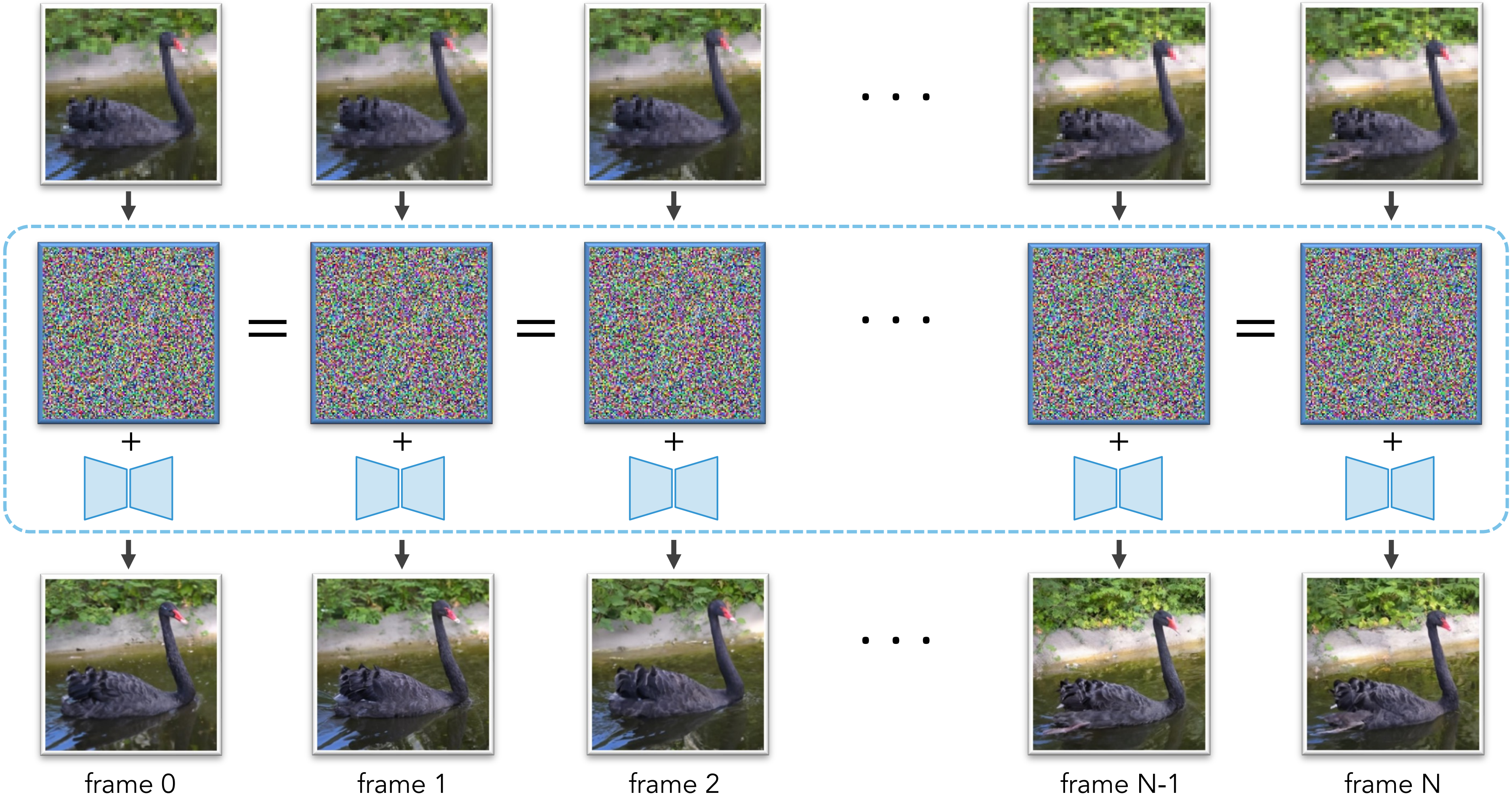


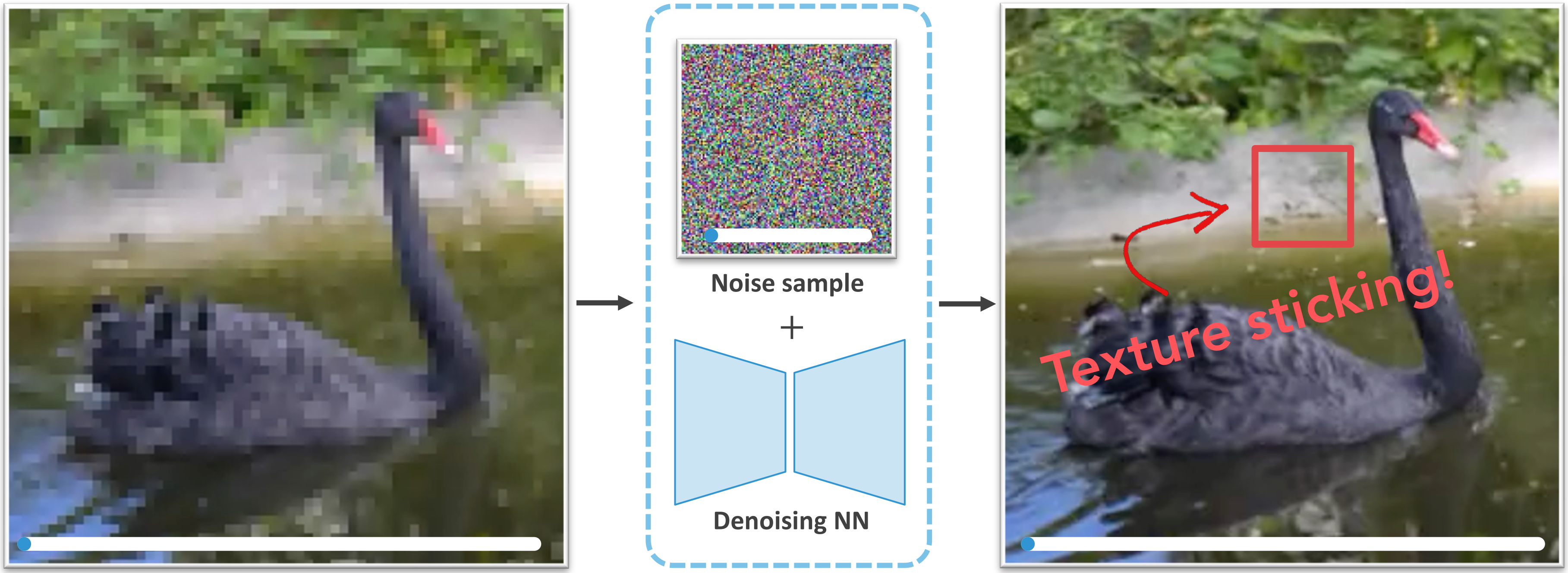
Image diffusion model



High resolution video


frame 0
frame 1
frame 2
frame N-1
frame N

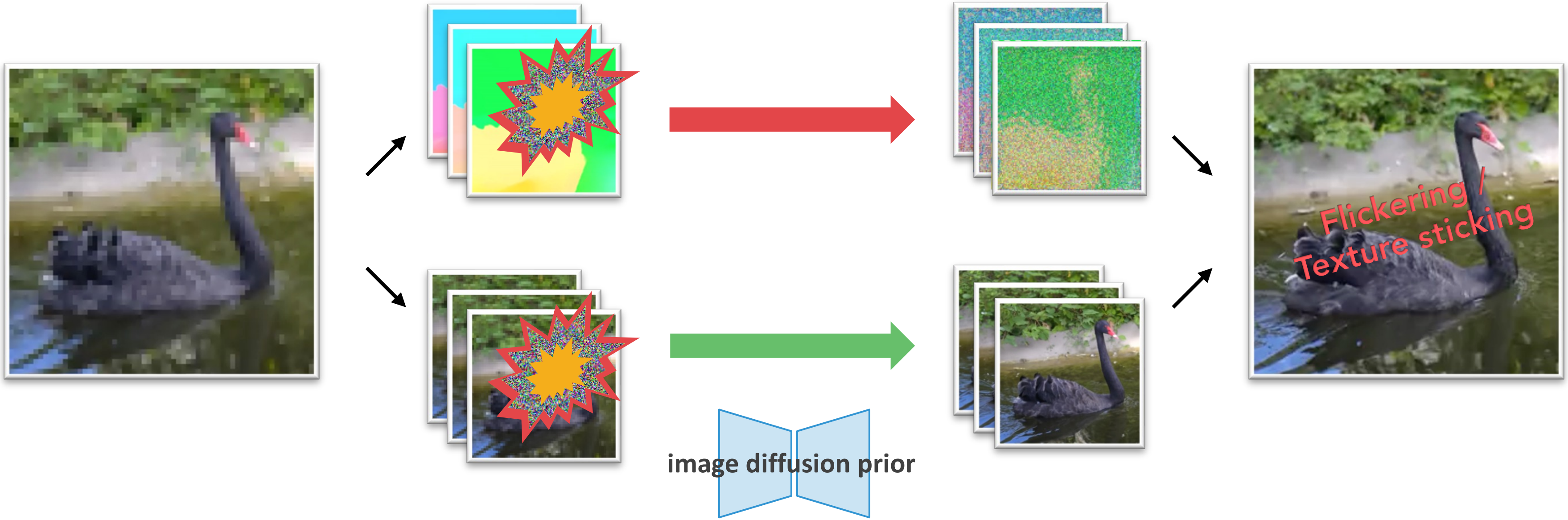
...To zero-shot video translation



Low resolution video | Image diffusion model | High resolution video

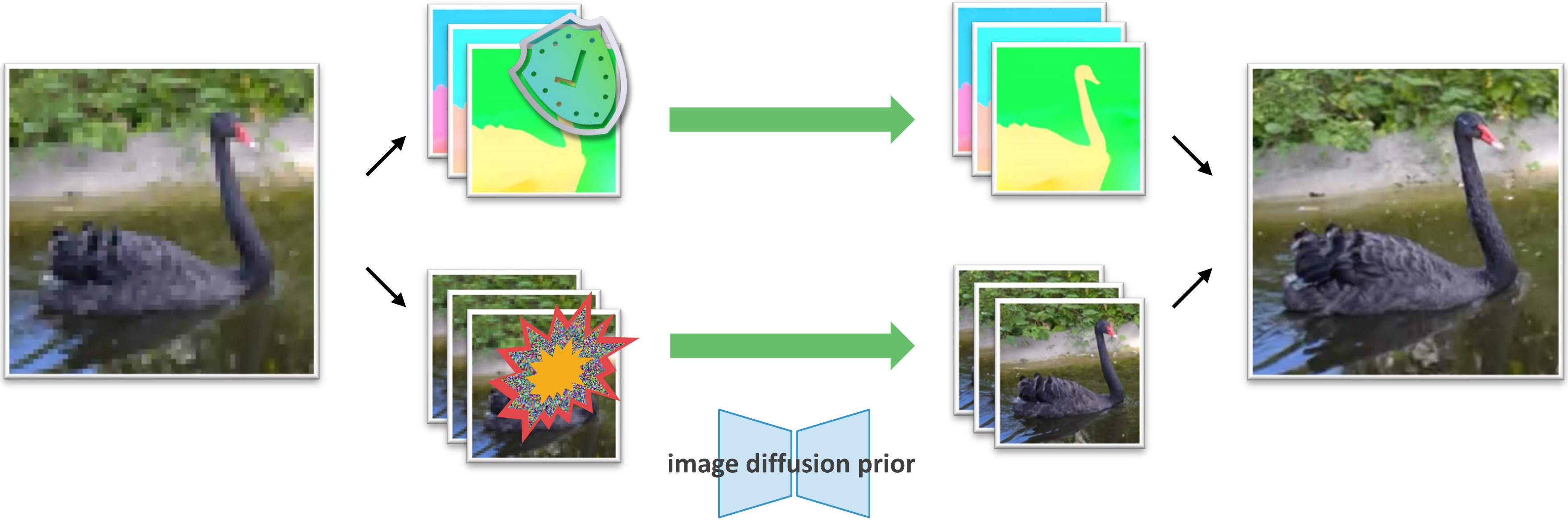
Zero-shot video translation

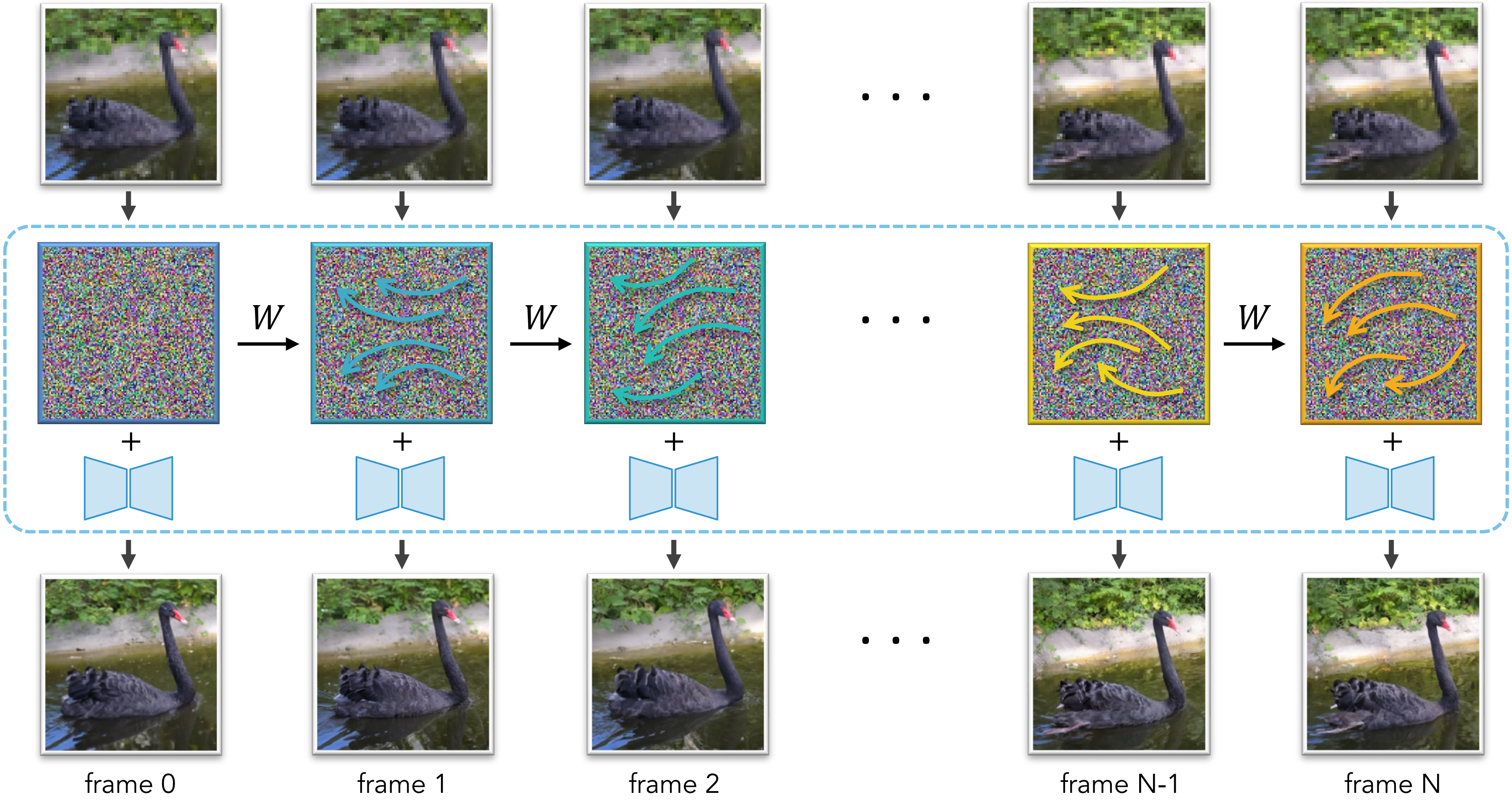
Issue: adding noise destroys both frame content and inter-frame correlation but pretrained image diffusion models only learns to recover the content.



Zero-shot video translation

Issue: adding noise destroys both frame content and inter-frame correlation but pretrained image diffusion models only learns to recover the content.

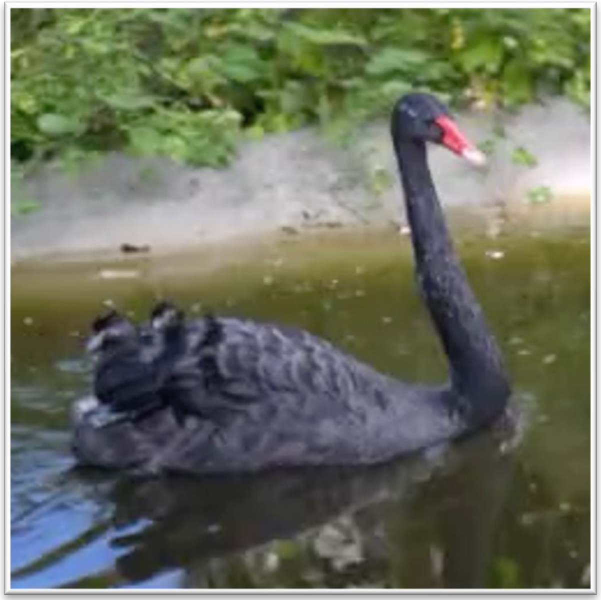




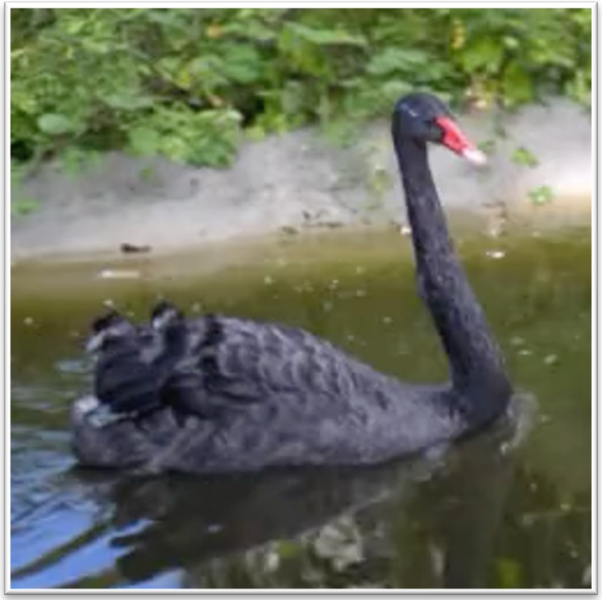
. . .
W
W
W
+
+
+
+
+
frame 0
frame 1
frame 2
frame N-1
frame N

Zero-shot video translation

Random noise



Fixed noise



Warped noise (ours)

Discrete Gaussian noise

Pixels are:

- normally distributed:

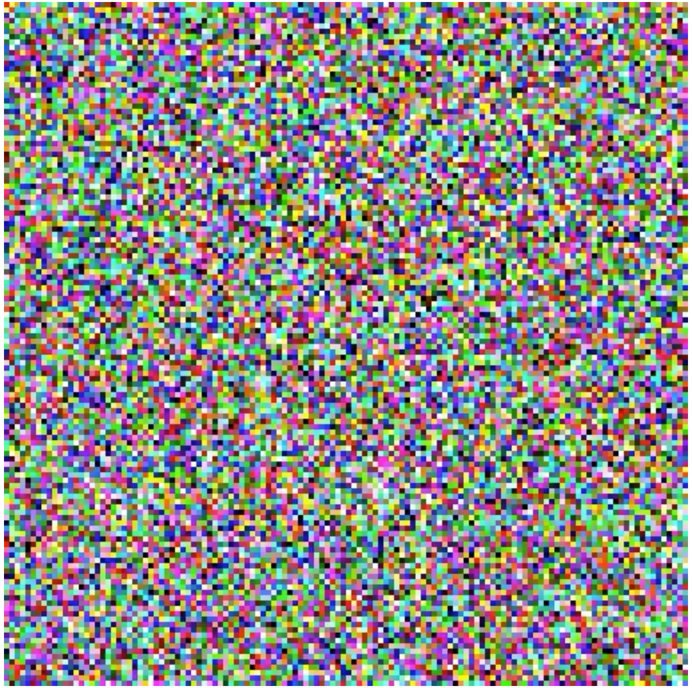
$$p_{ij} \sim \mathcal{N}(0, I)$$

⇒ **unit variance**

- independently sampled:

$$p_{ij} \perp\!\!\!\perp p_{kl}$$

⇒ **spatially uncorrelated**



Discrete Gaussian noise

Pixels are:

- normally distributed:

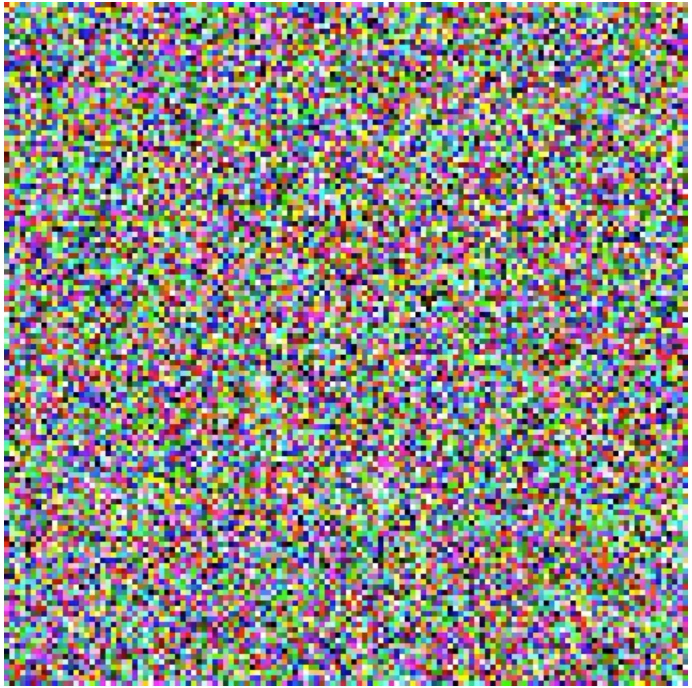
$$p_{ij} \sim \mathcal{N}(0, I)$$

~~⇒ unit variance~~

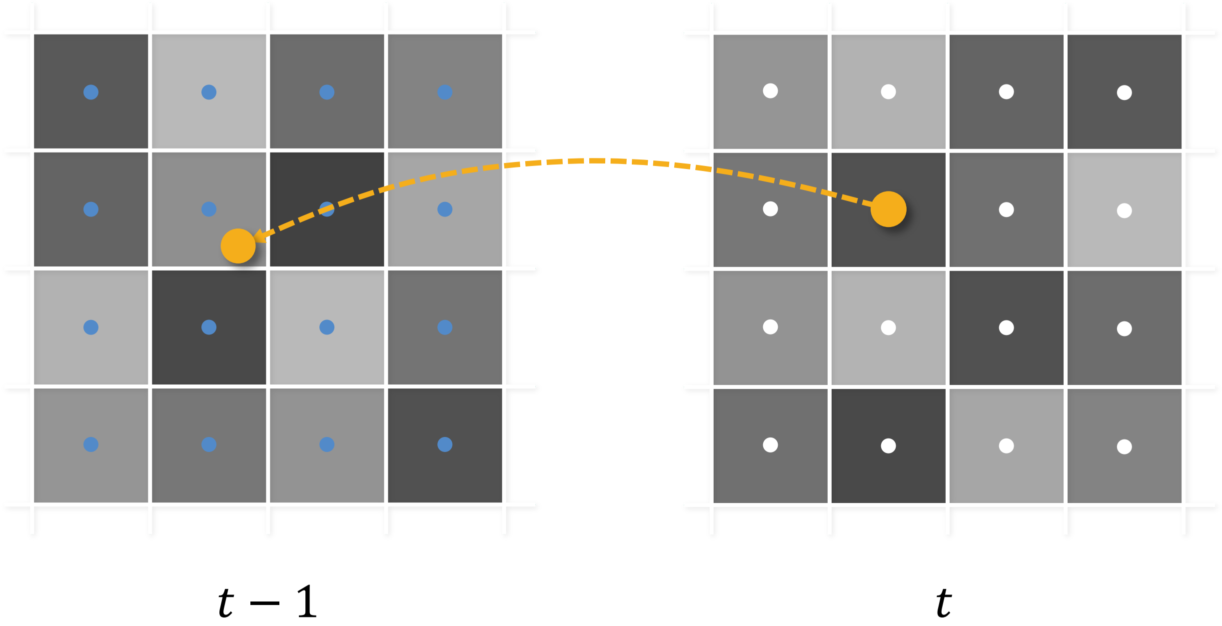
- independently sampled:

$$p_{ij} \perp\!\!\!\perp p_{kl}$$

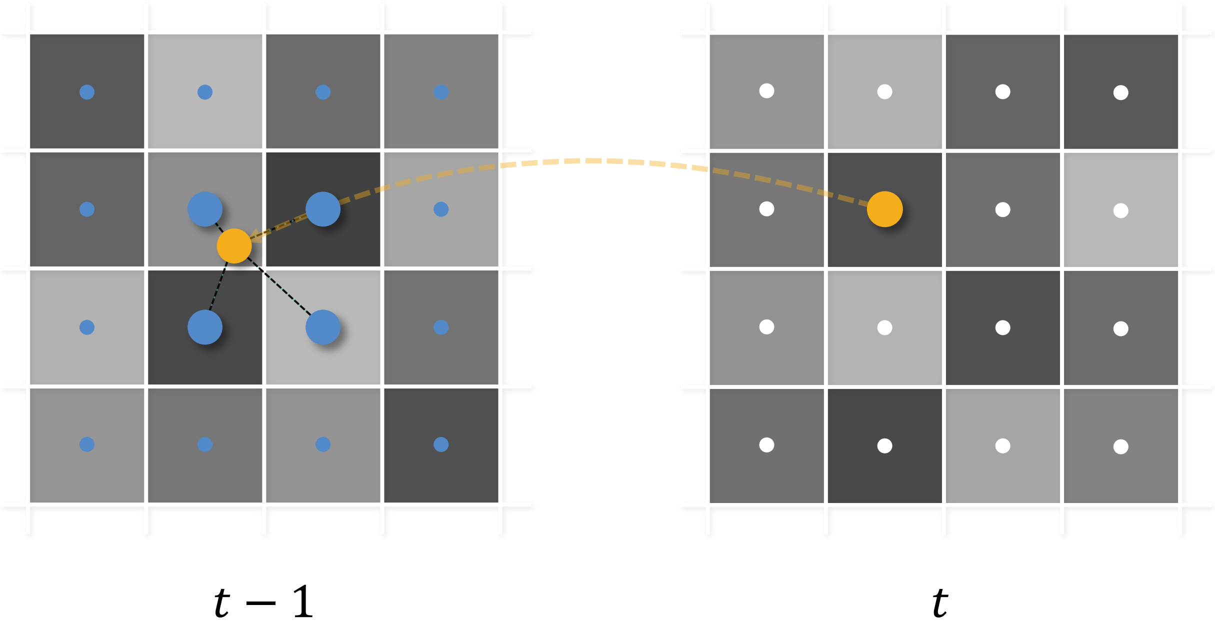
~~⇒ spatially uncorrelated~~



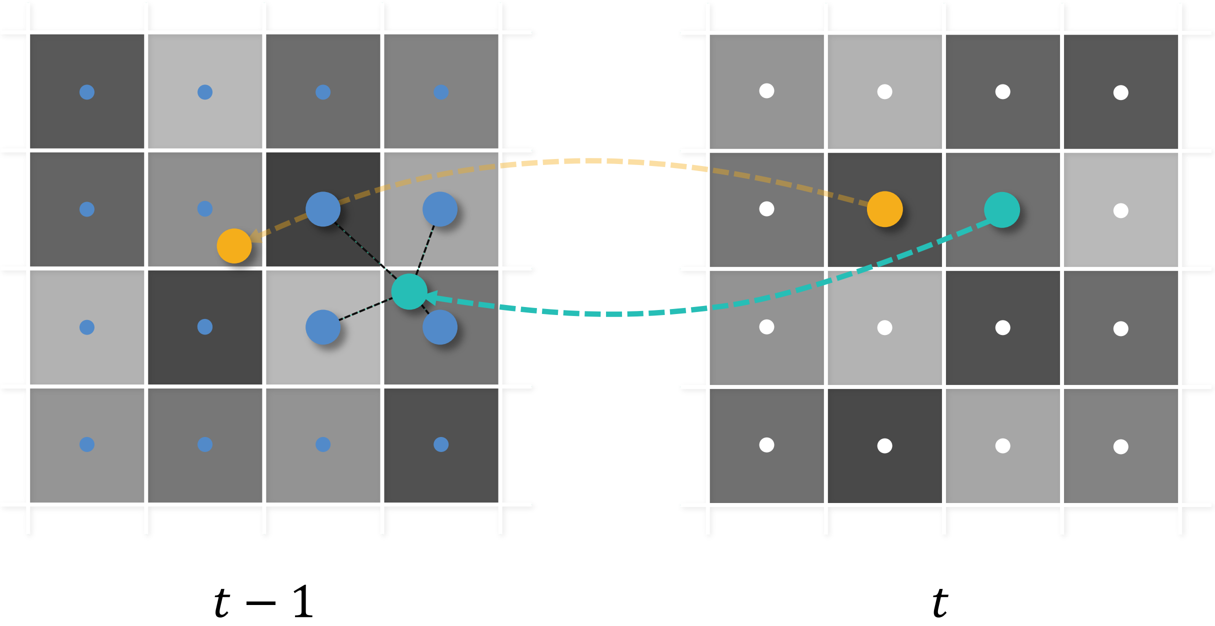
Spatial correlation in warping



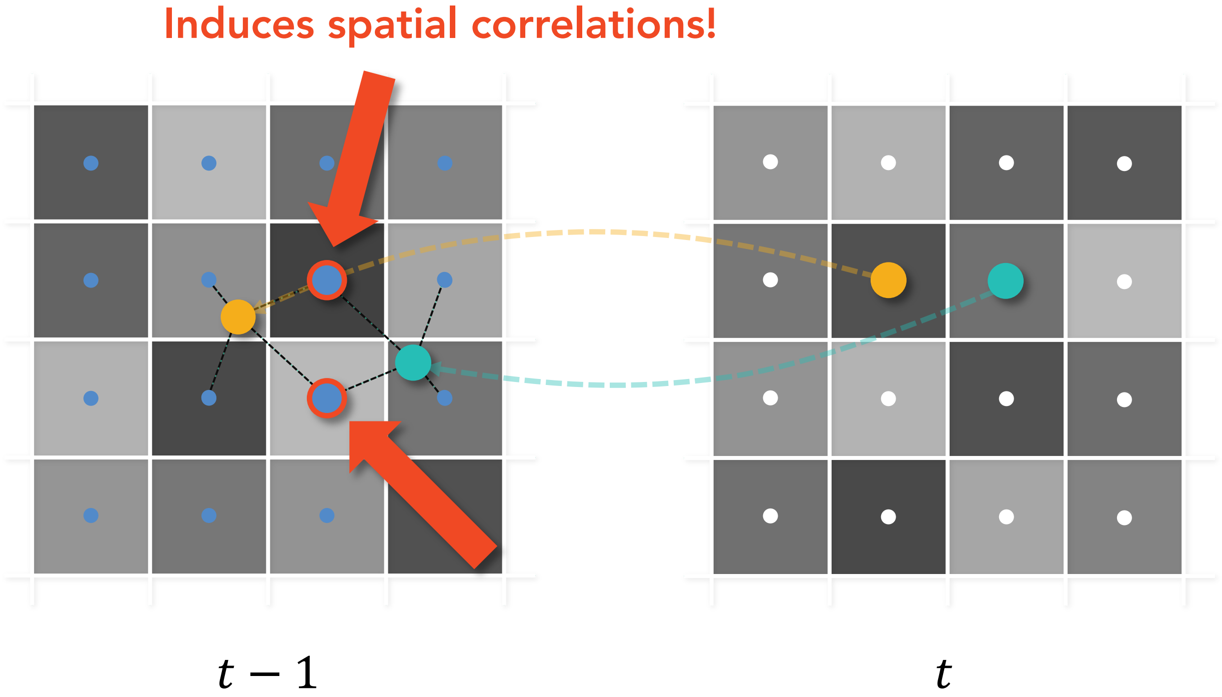
Spatial correlation in warping



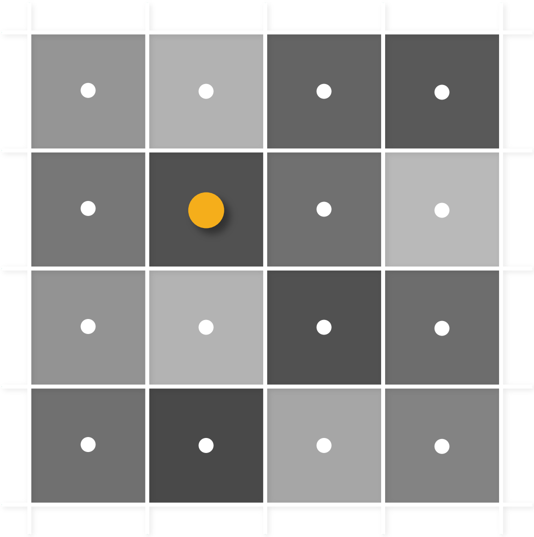
Spatial correlation in warping



Spatial correlation in warping



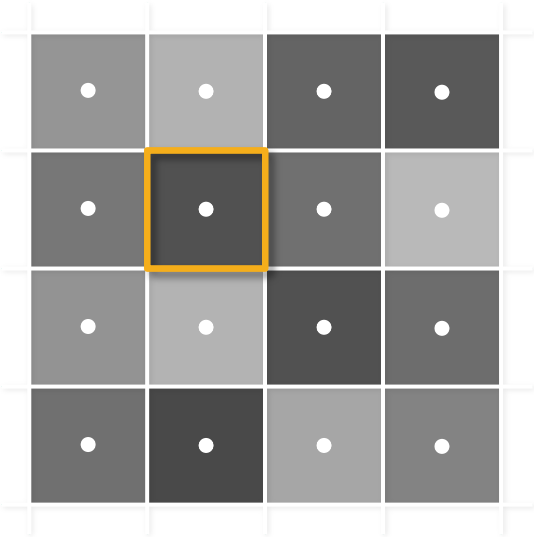
Main insight: pixels are *areas*



A pixel as a *discrete point:*

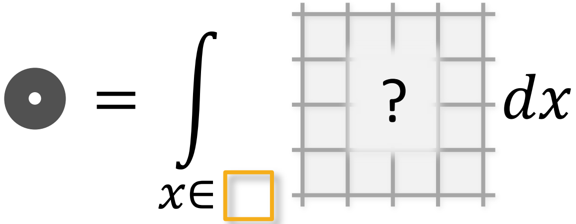
$$= x \sim \mathcal{N}(0,1)$$

Main insight: pixels are *areas*

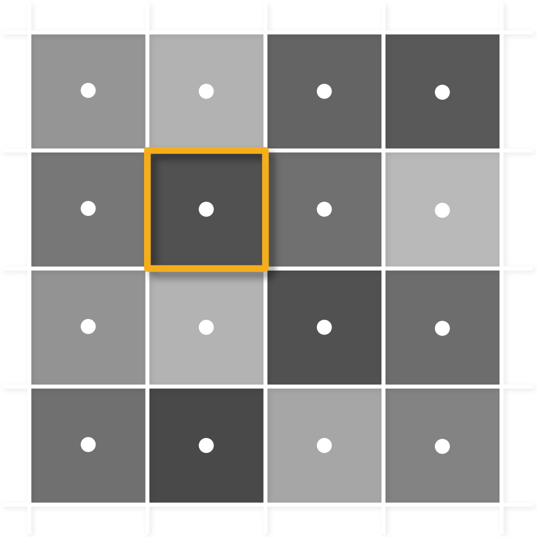


A pixel as a *discrete point:*

A pixel as an *area integral:*

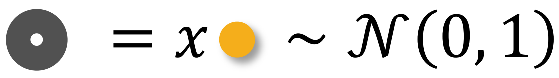


Main insight: pixels are *areas*

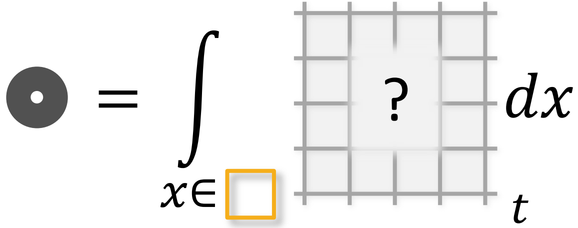


t

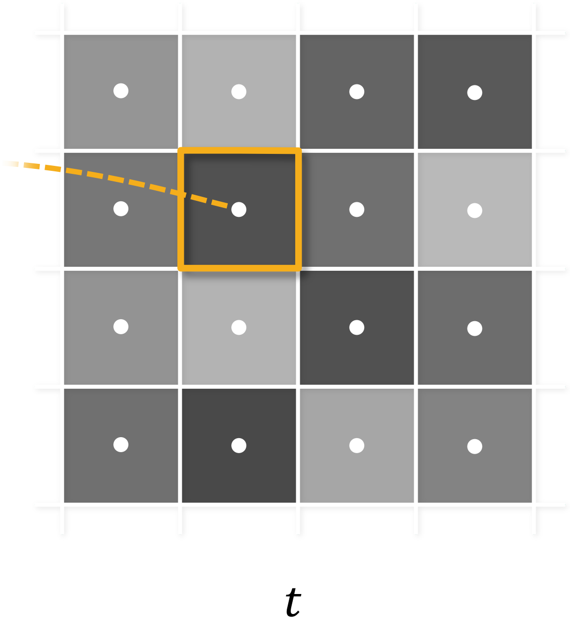
A pixel as a *discrete point*:



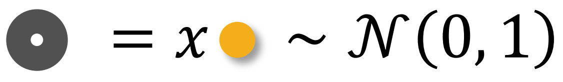
A pixel as an *area integral*:



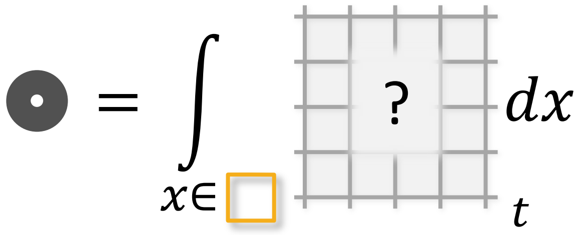
Main insight: pixels are *areas*



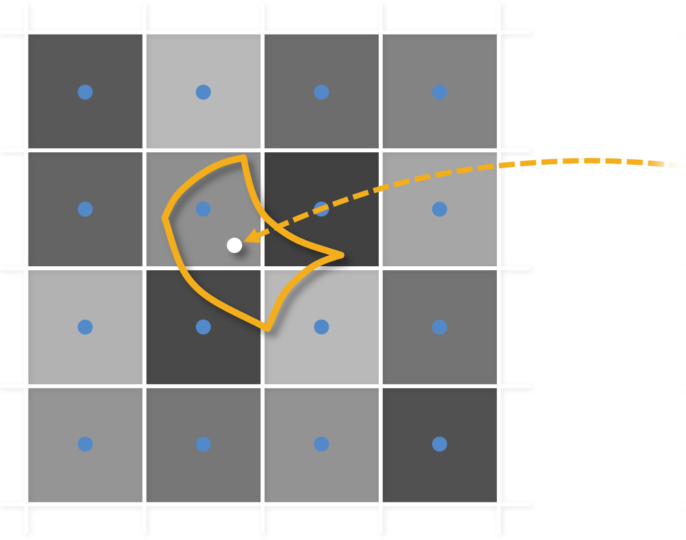
A pixel as a *discrete point:*



A pixel as an *area integral:*



Main insight: pixels are *areas*



$t - 1$

A pixel as a *discrete point:*

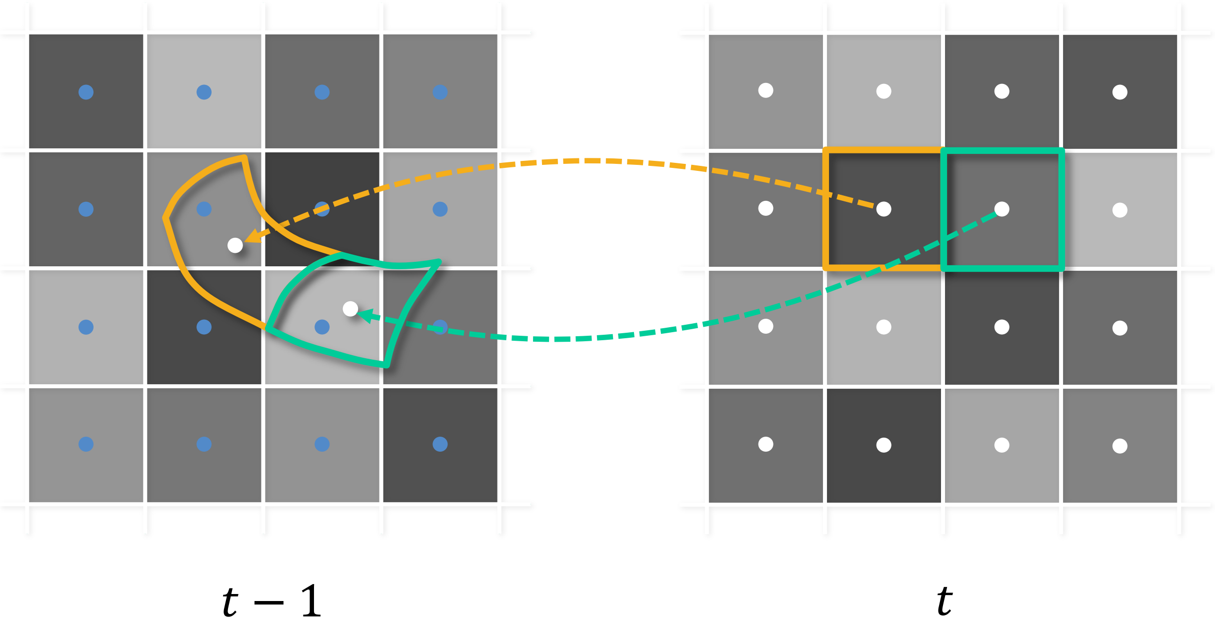
$$= x \sim \mathcal{N}(0, 1)$$

A pixel as an *area integral:*

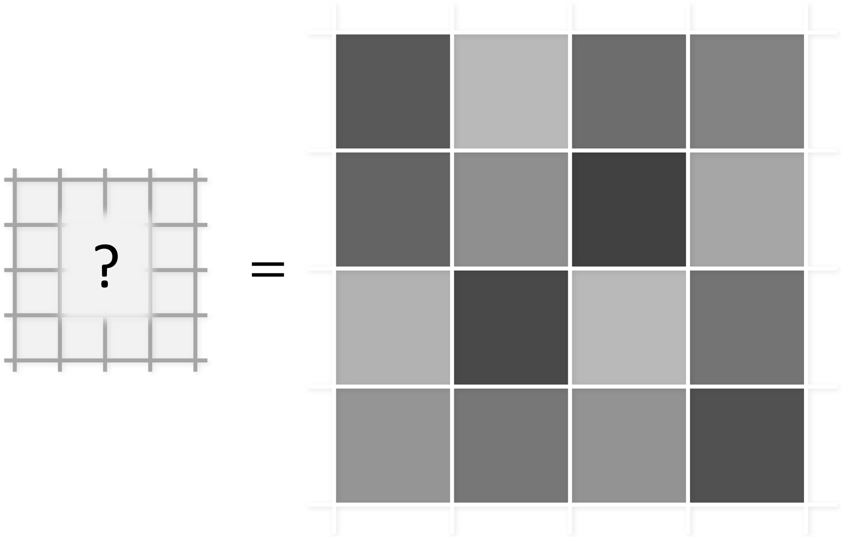
$$= \int_{x \in} ? \, dx$$

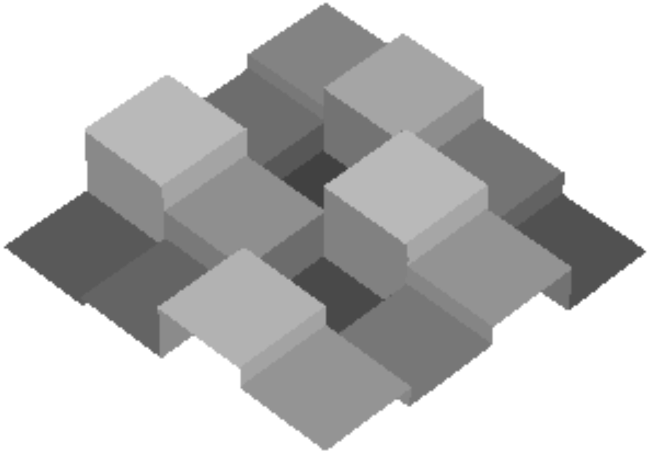
$t - 1$

Main insight: pixels are *areas*

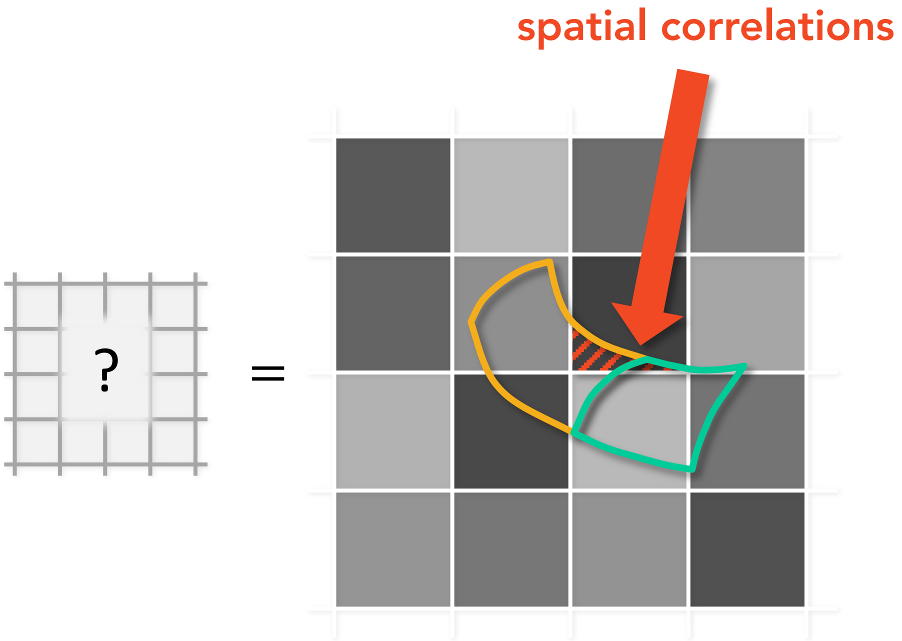


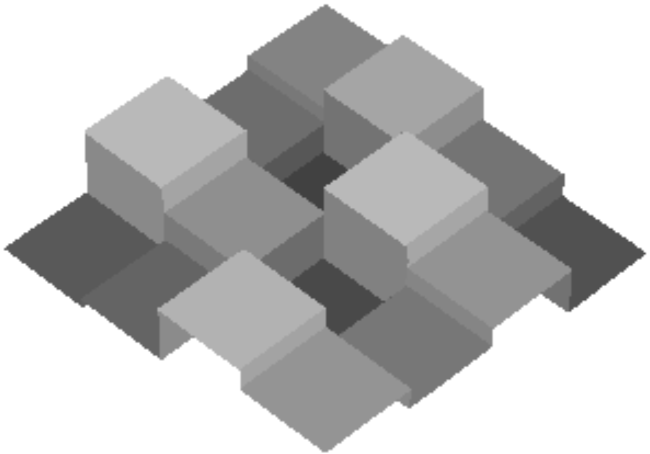
Choice of the underlying signal



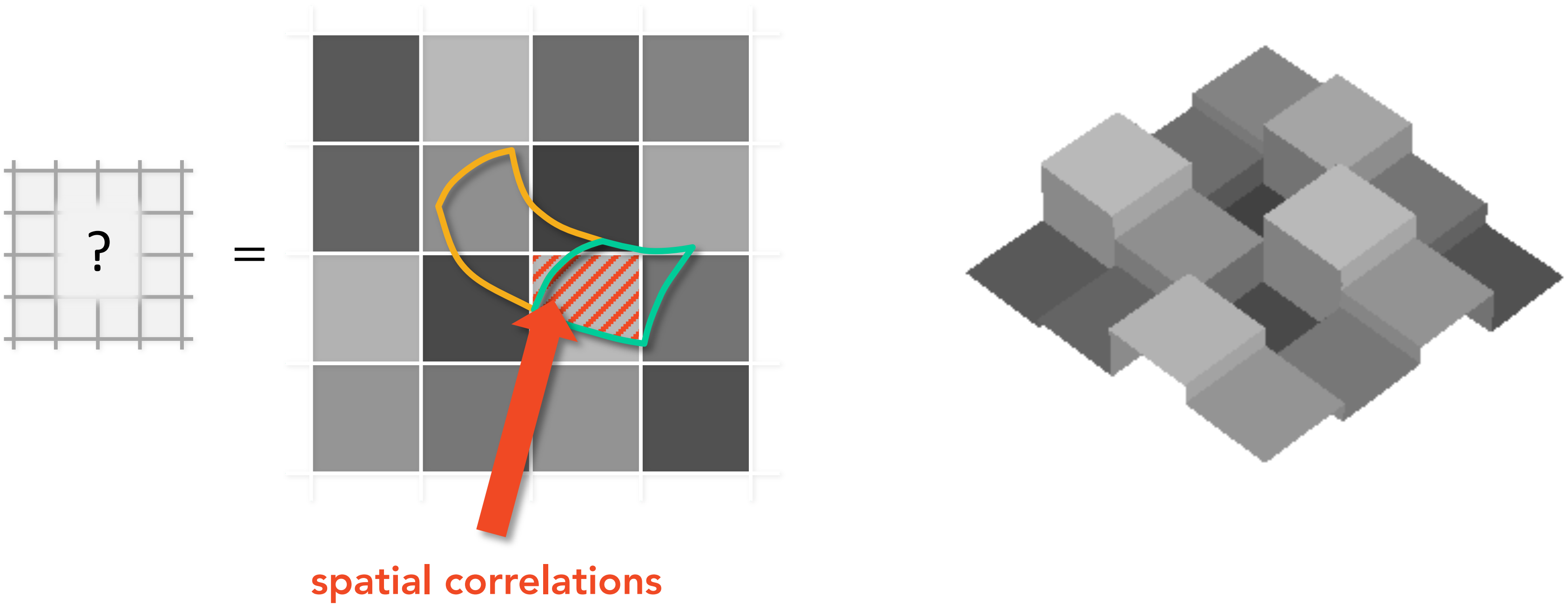


Choice of the underlying signal

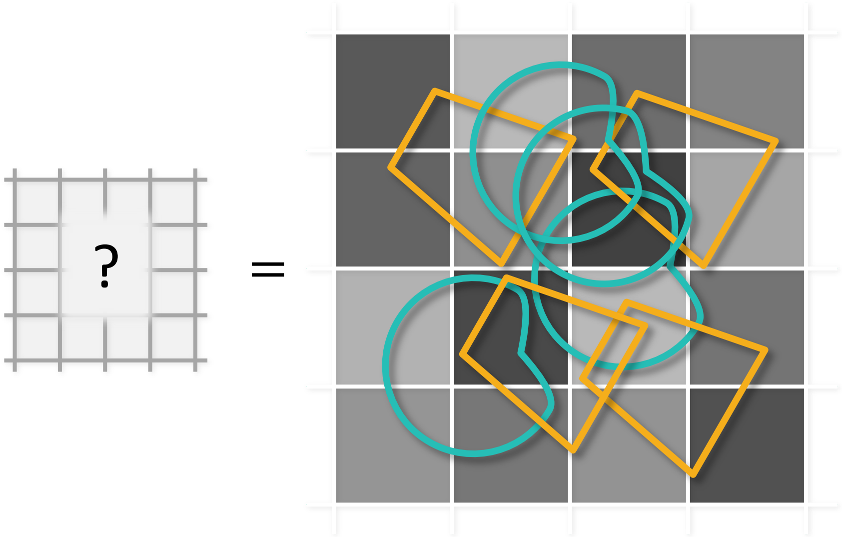




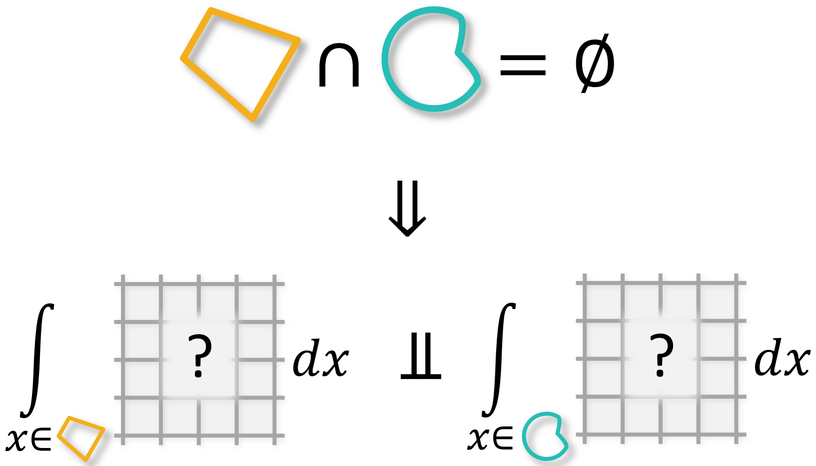
Choice of the underlying signal



Choice of the underlying signal

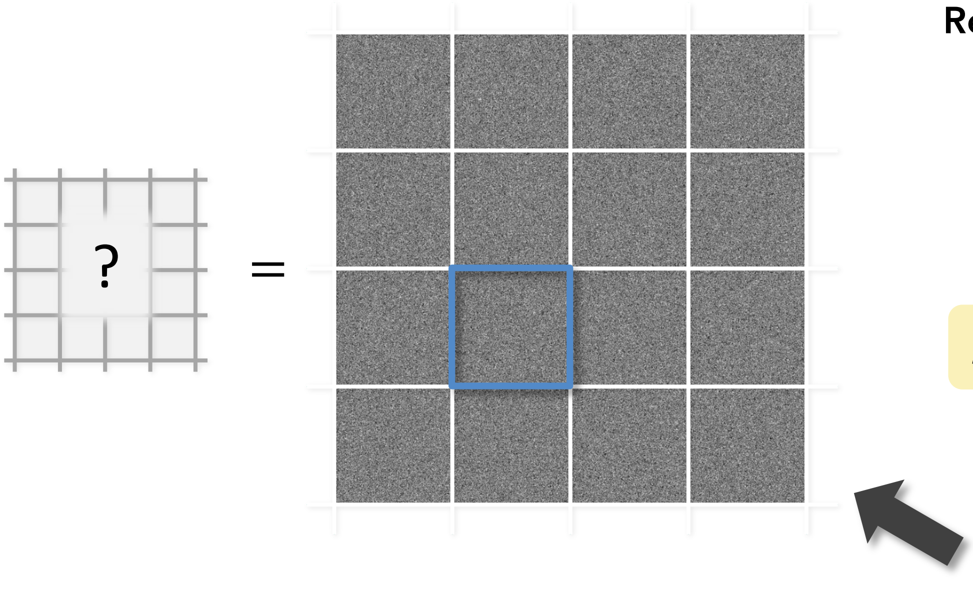


Desired property:



independent

Choice of the underlying signal

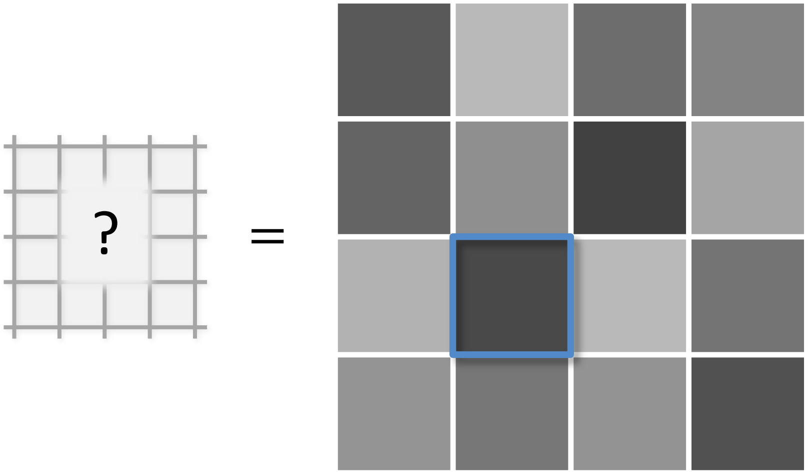


Reproductive property of normal distribution:

$$x_1 \sim \mathcal{N}(0, \sigma_1^2), \qquad x_2 \sim \mathcal{N}(0, \sigma_2^2)$$

$$x_1 \perp\!\!\!\perp x_2 \;\Rightarrow\; x_1 + x_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$$

Choice of the underlying signal



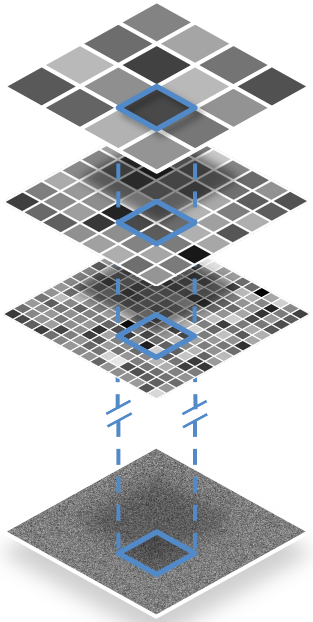
Reproductive property of normal distribution:

$$x_1 \sim \mathcal{N}(0, \sigma_1^2), \qquad x_2 \sim \mathcal{N}(0, \sigma_2^2)$$

$$x_1 \perp\!\!\!\perp x_2 \;\Rightarrow\; x_1 + x_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$$

Choice of the underlying signal

∫-noise representation

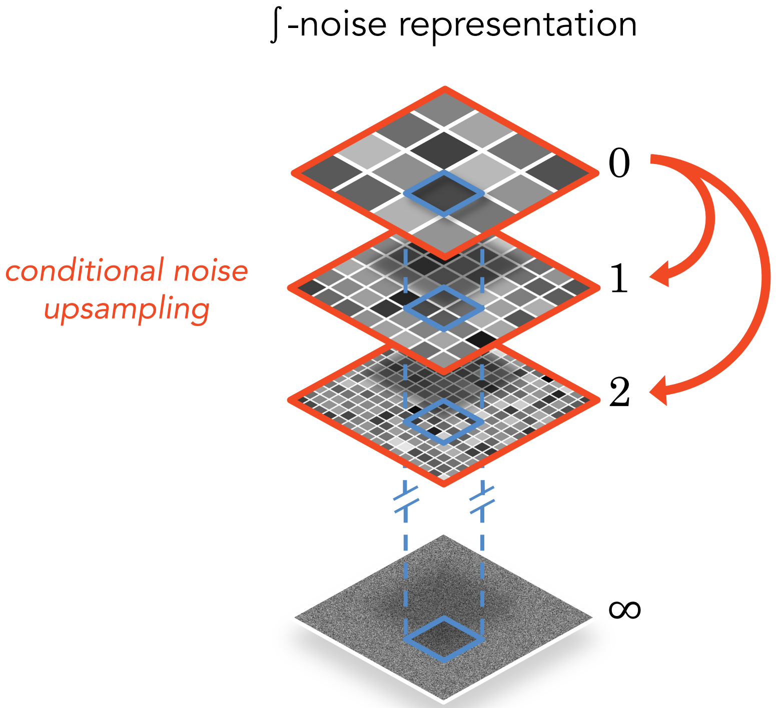


Reproductive property of normal distribution:

$$x_1 \sim \mathcal{N}(0, \sigma_1^2), \qquad x_2 \sim \mathcal{N}(0, \sigma_2^2)$$

$$x_1 \perp\!\!\!\perp x_2 \;\Rightarrow\; x_1 + x_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$$

Choice of the underlying signal

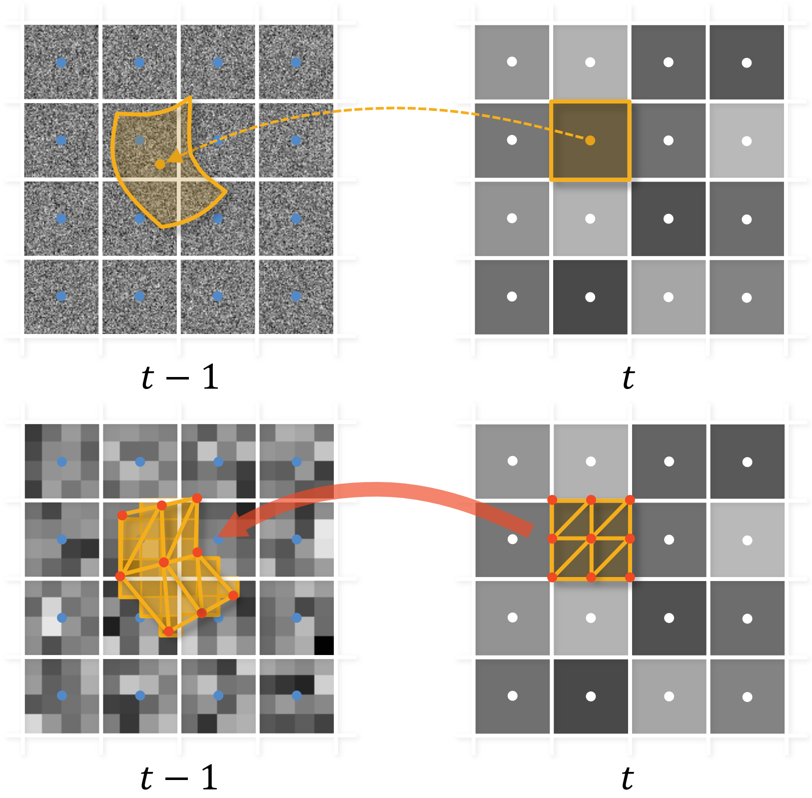


Reproductive property of normal distribution:

$$x_1 \sim \mathcal{N}(0, \sigma_1^2), \qquad x_2 \sim \mathcal{N}(0, \sigma_2^2)$$

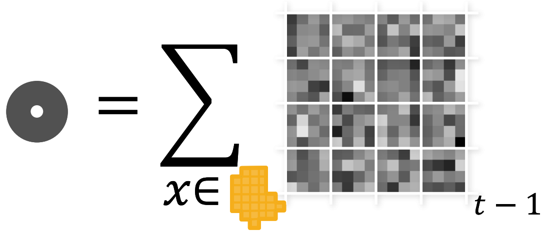
$$x_1 \perp\!\!\!\perp x_2 \;\Rightarrow\; x_1 + x_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$$

Noise warping: from theory to practice

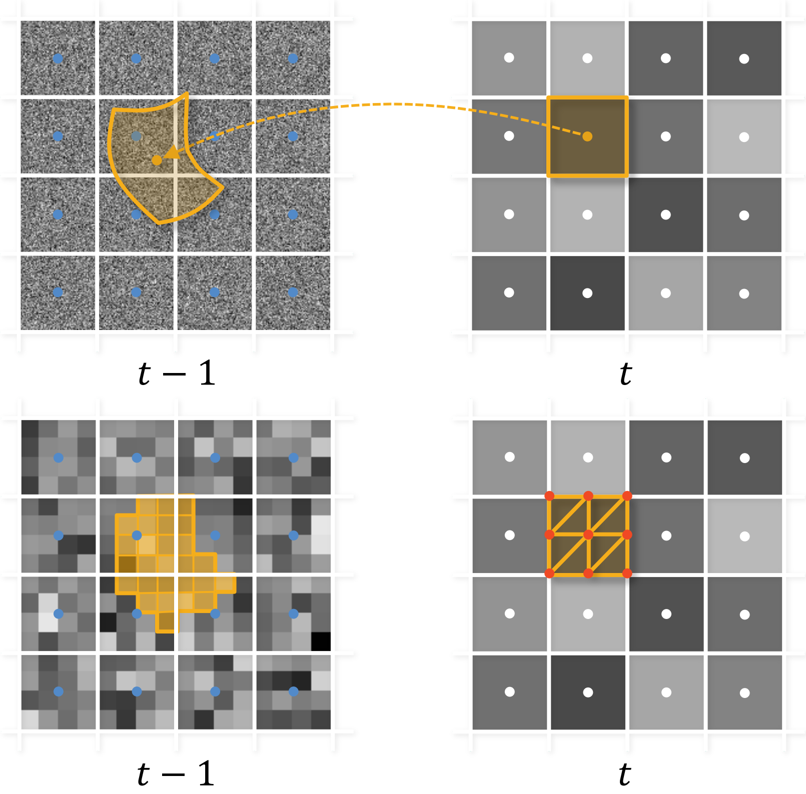


Continuous form:

Discrete implementation:

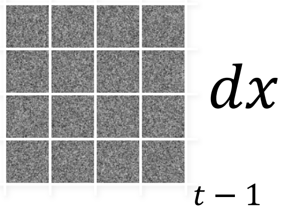


Noise warping: from theory to practice



Continuous form:

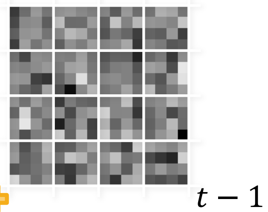
$$\odot = \int_{x \in \text{(region)}} \sqrt{|\nabla T(x)|}\; [\text{noise}]_{t-1}\, dx$$



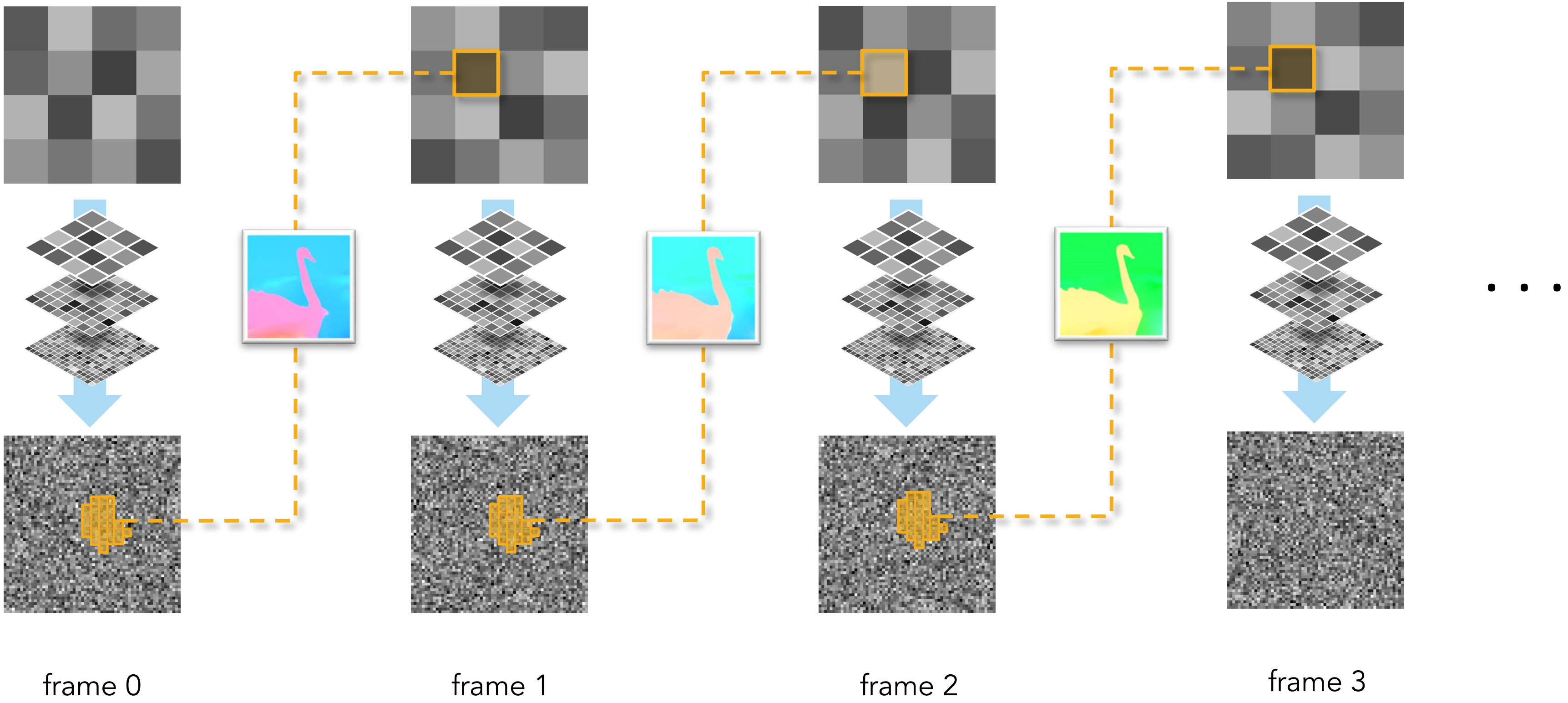
(noise transport equation)

Discrete implementation:

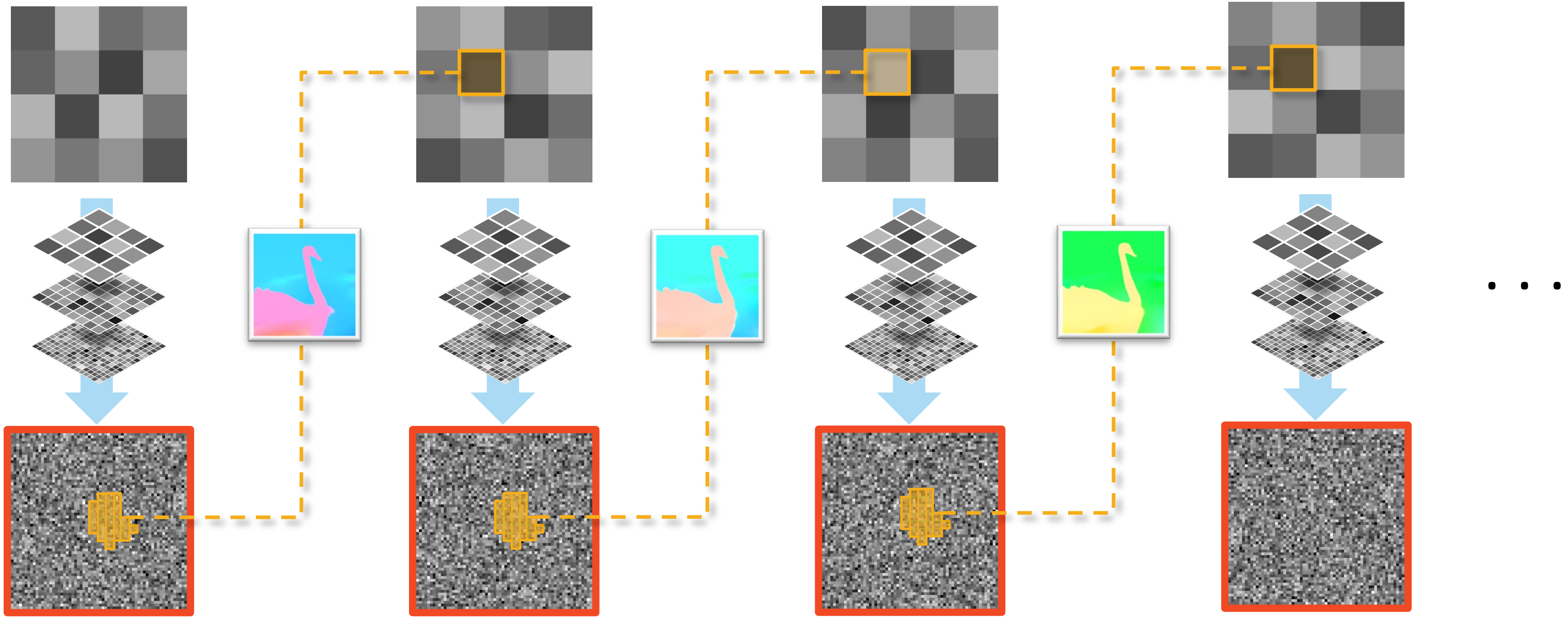
$$\odot = \frac{1}{\sqrt{\#\text{(region)}}} \sum_{x \in \text{(region)}} [\text{noise}]_{t-1}$$



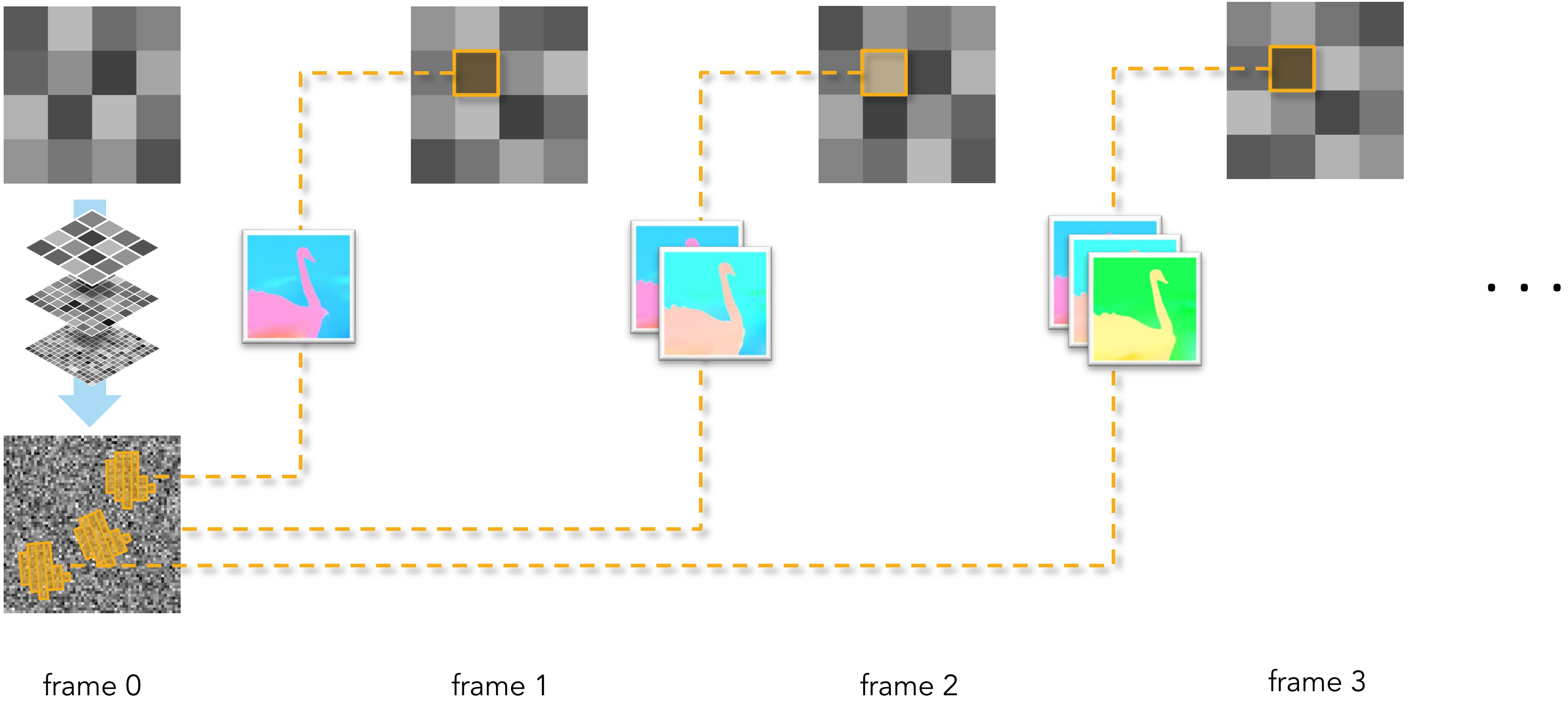
Noise warping: extending to long sequences



Noise warping: extending to long sequences

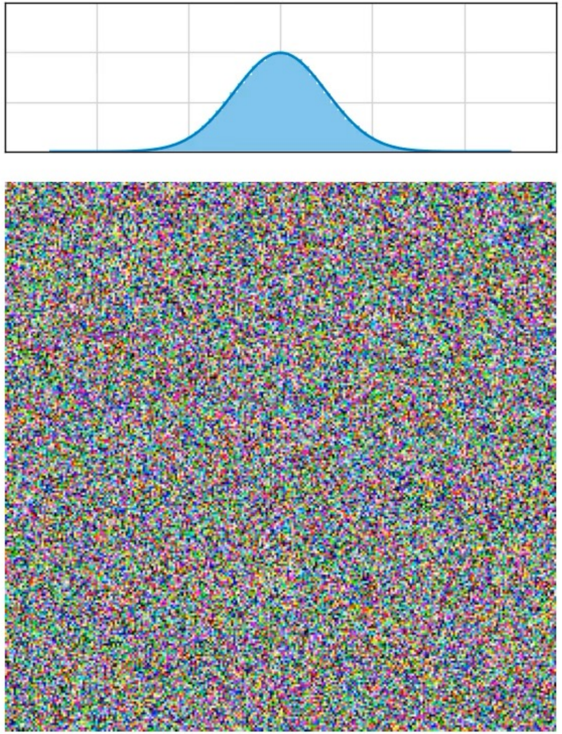


Noise warping: extending to long sequences

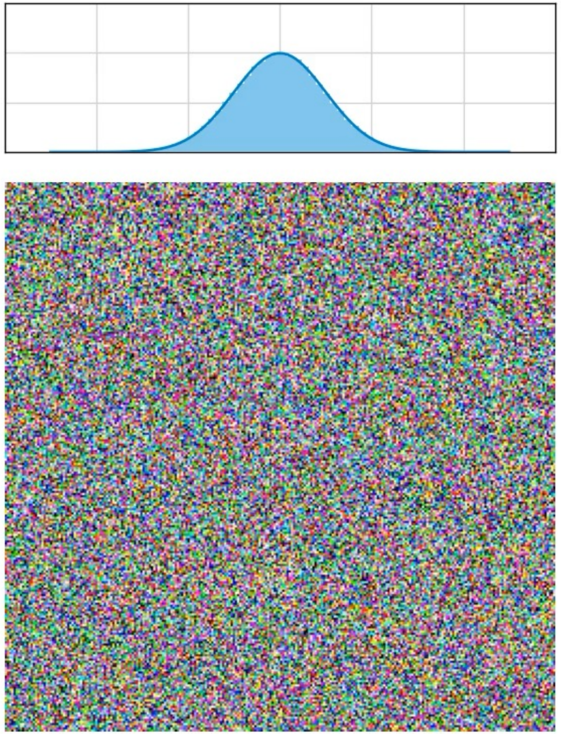


Results: visualizing the noise

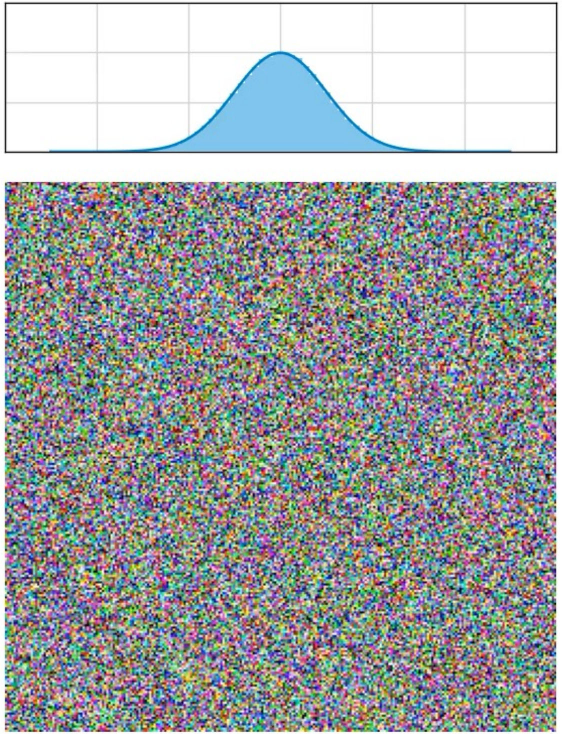
Bilinear interpolation



Nearest interpolation

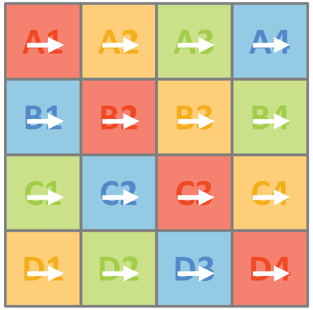


$\int$-noise warping (ours)

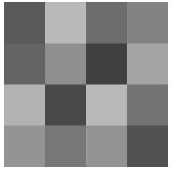


Results: visualizing spatial & temporal correlation

$W =$

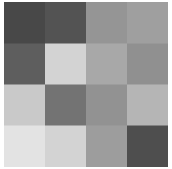


$G_0 =$



$\sim \mathcal{N}(0, I)$

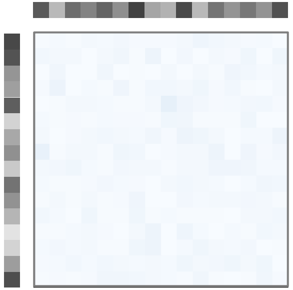
$G_1 =$



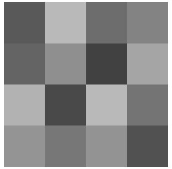
$\perp\!\!\!\perp \; G_0$

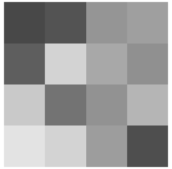
Results: visualizing spatial & temporal correlation

$W =$ 



$\longrightarrow$ temporal correlation

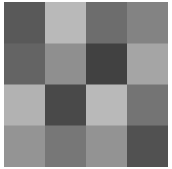
$G_0 =$  $\sim \mathcal{N}(0, I)$

$G_1 =$  $\perp\!\!\!\perp \; G_0$

Results: visualizing spatial & temporal correlation

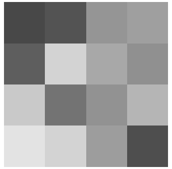
$W =$

$G_0 =$

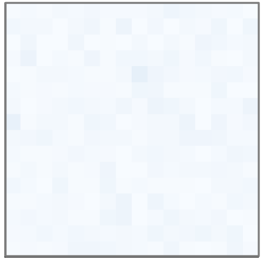


$\sim \mathcal{N}(0, I)$

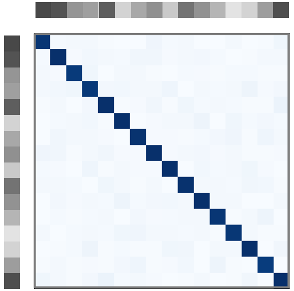
$G_1 =$



$\perp\!\!\!\perp \; G_0$

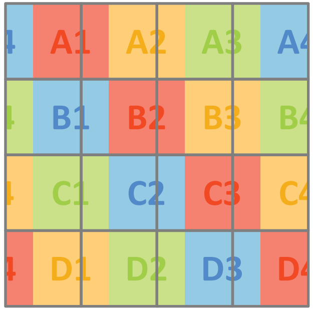


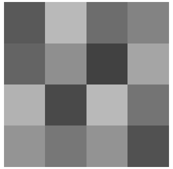
⟶ temporal correlation

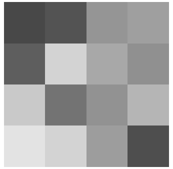


⟶ spatial correlation

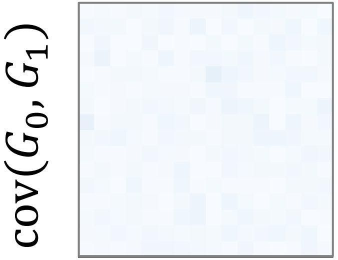
Results: visualizing spatial & temporal correlation

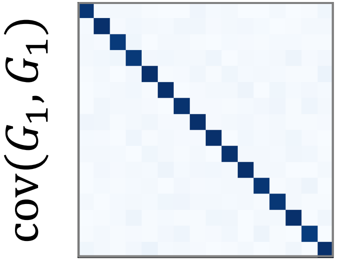
$W =$ 

$G_0 =$  $\sim \mathcal{N}(0, I)$

$G_1 =$  $\perp\!\!\!\perp \; G_0$

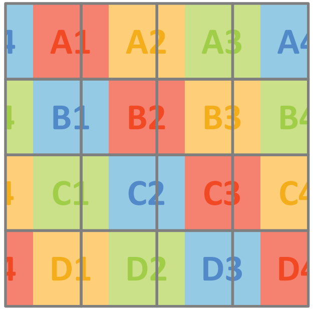
$G_1 \perp\!\!\!\perp G_0$



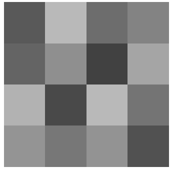


Results: visualizing spatial & temporal correlation

$W =$

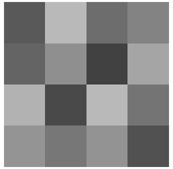


$G_0 =$

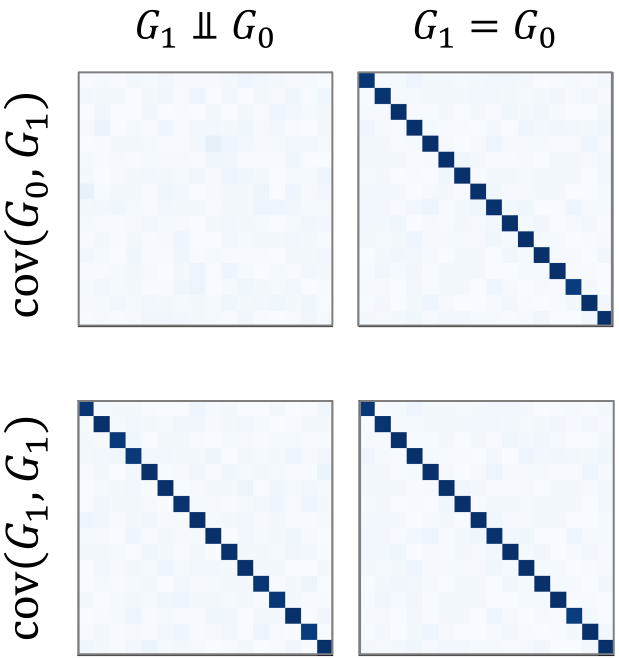


$\sim \mathcal{N}(0, I)$

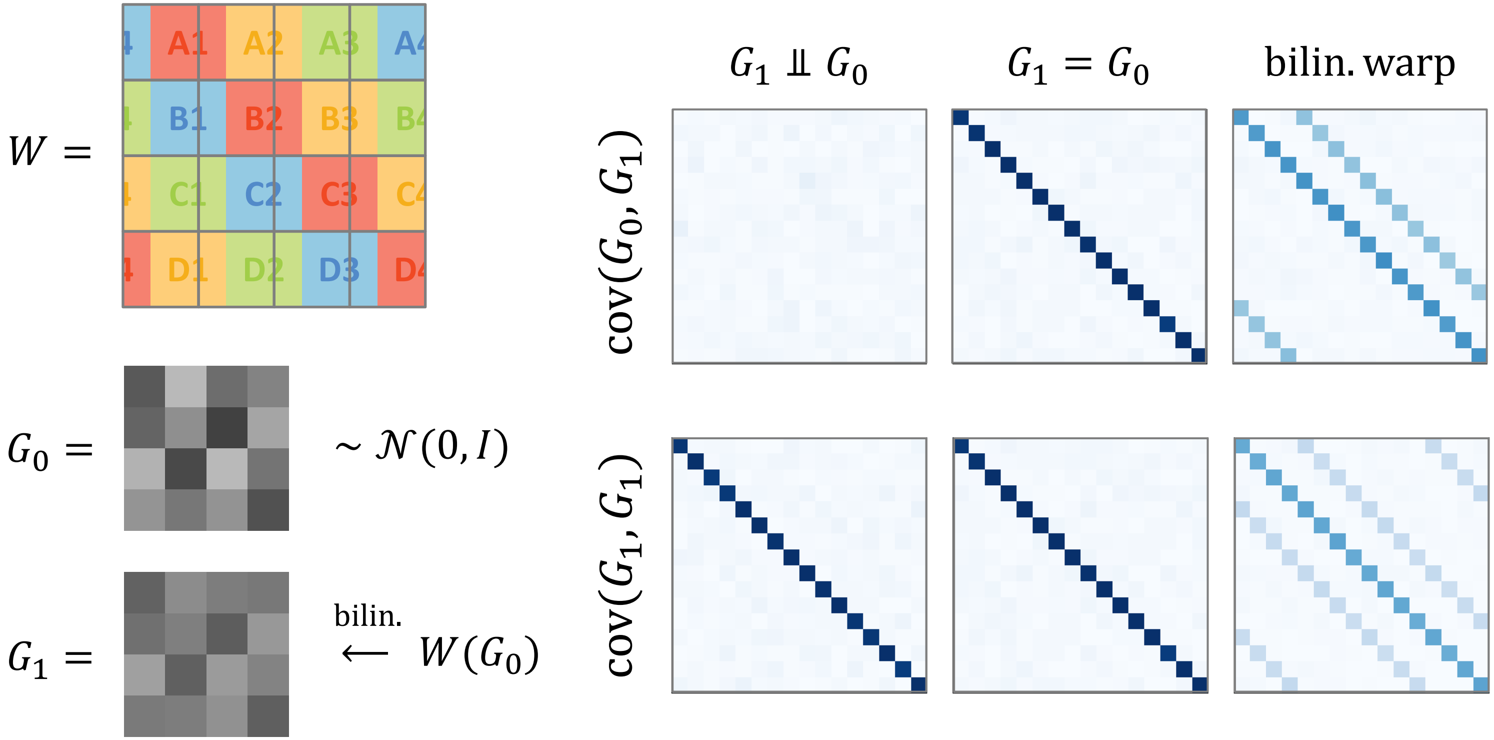
$G_1 =$



$= \; G_0$

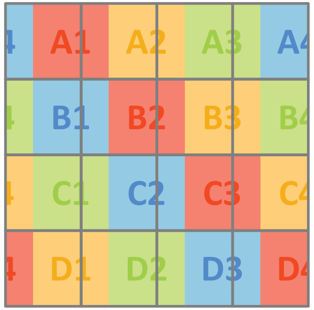


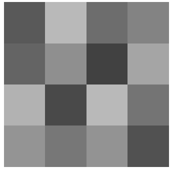
Results: visualizing spatial & temporal correlation

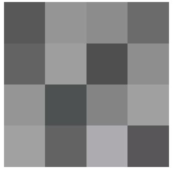


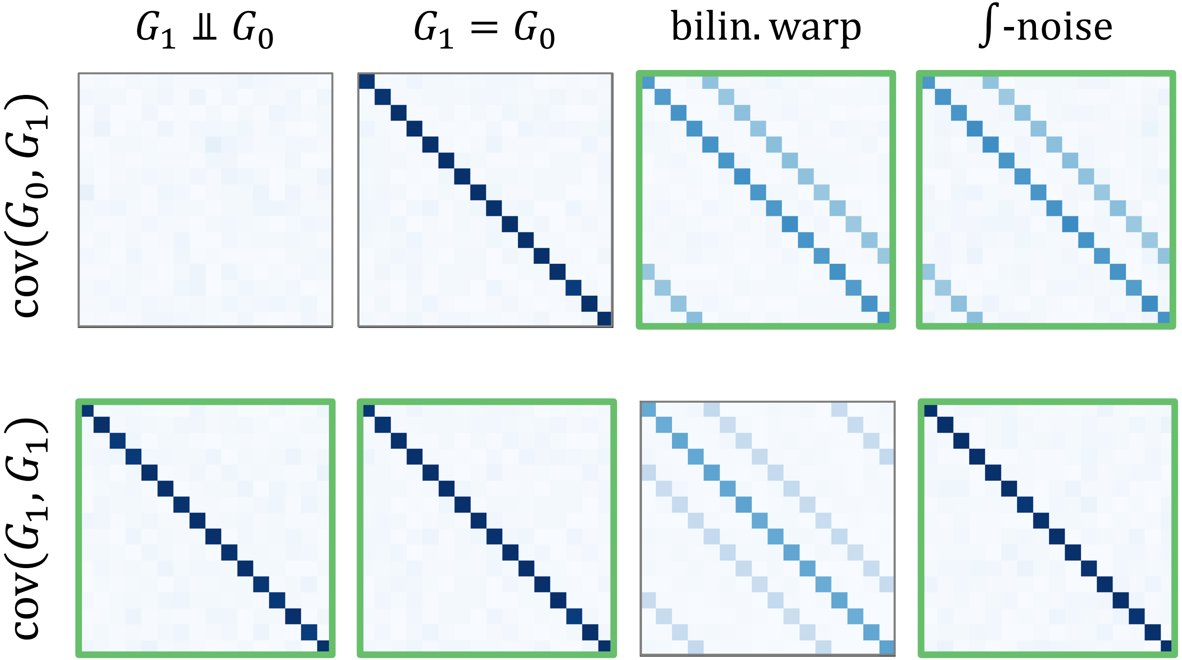
ETH zürich | cgl computer graphics laboratory

Results: visualizing spatial & temporal correlation

$W =$ 

$G_0 =$  $\sim \mathcal{N}(0, I)$

$G_1 =$  $\overset{\int\text{-noise}}{\longleftarrow} W(G_0)$



Results: photorealistic rendering w/ SDEdit [Meng et al. 2021]

Random noise

Fixed noise

∫-noise warping (ours)

Results: photorealistic rendering w/ SDEdit [Meng et al. 2021]

Bilinear interpolation

Nearest interpolation

∫-noise warping (ours)

Results: photorealistic rendering w/ SDEdit [Meng et al. 2021]

PYoCo (progressive)
[Ge et al. 2023]

Control-A-Video
[Chen et al. 2023]

$\int$-noise warping (ours)

Conclusion

- A **novel noise representation** dubbed $\int$-noise, that interprets a discrete noise sample as a discretized view of an underlying continuous noise field;
- A theoretically-grounded **noise transport equation** tailored to Gaussian noise and a **practical implementation** for distribution-preserving noise warping;
- Showcased **improved temporal coherency** in zero-shot video translation tasks using pre-trained diffusion image models.

Limitations & Future Work

- Can be sensitive to the accuracy of the flow;
- The importance of the noise prior depends on the pipeline chosen;
- Extend the noise prior to larger classes of noises, e.g. latent diffusion;
- Explore new applications of the method.

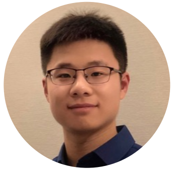
Thank you!

How I Warped Your Noise:

A Temporally-correlated Noise Prior for Diffusion Models

Project page:
warpyournoise.github.io

Pascal Chang



Jingwei Tang

Markus Gross

Vinicius C. Azevedo

ETH zürich | cgl computer graphics laboratory