SinGAN:

Learning a Generative Model from a Single Natural Image

Tamar Rott Shaham, Tali Dekel, Tomer Michaeli

[Technion, Google Research]

ICCV 2019

Sungman, Cho.

Introduction

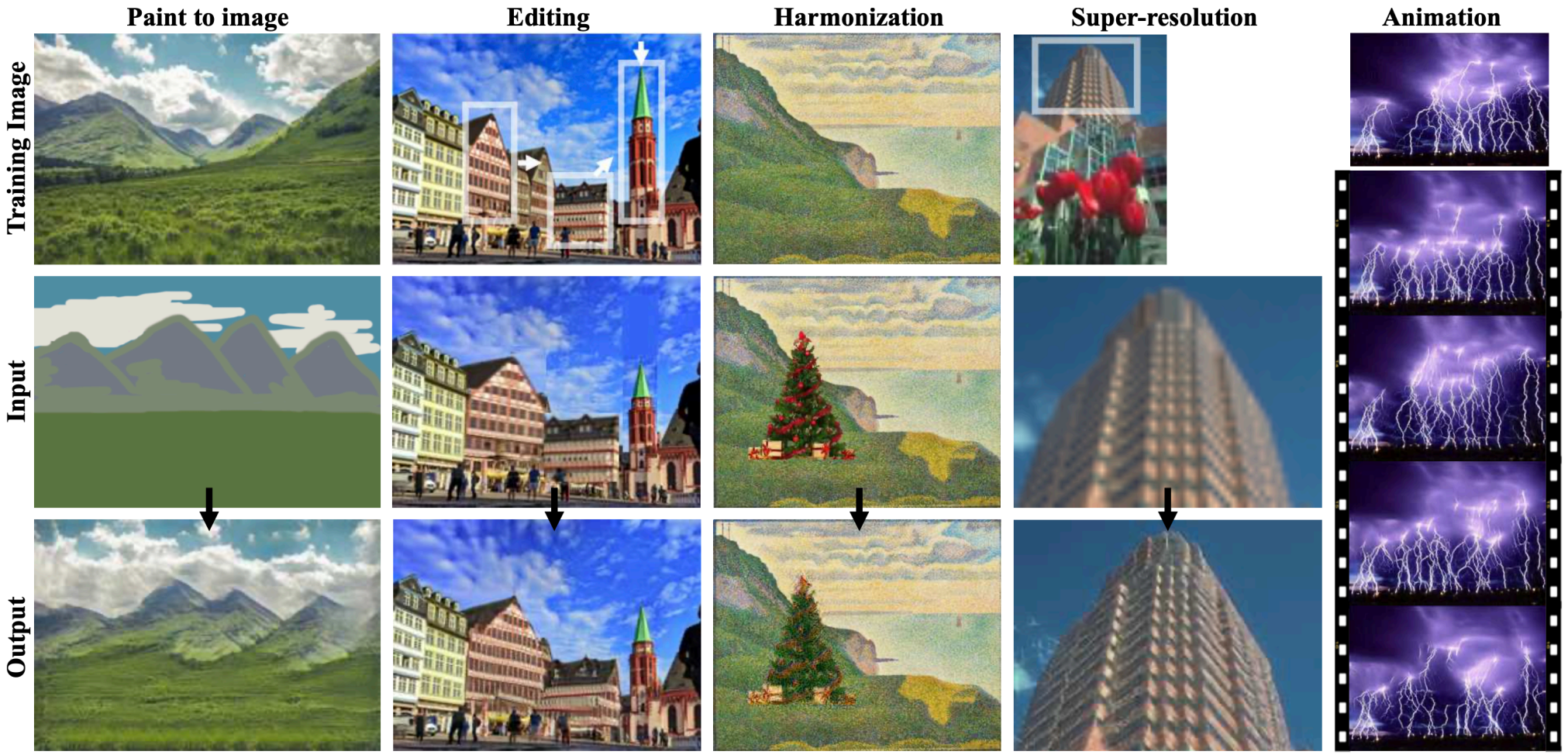
Challenges

- Previous single image GAN schemes, **limited to texture images, and conditional**
- **Capturing the distribution of highly diverse datasets** with multiple object classes is still considered a **major challenge**.

Contribution

- **Unconditional generation** learned from a **single natural image.**
- Can **produce diverse high-quality image** samples, which semantically resemble the training image, yet **contain new object configurations and structures.**

Image manipulation



Related Work

- **Single image deep models**
  - proposed to “overfit” a deep model to a single training image
  - designed for specific tasks (super resolution, texture expansion)
  - generation is not used to draw random samples.

D.Ulyanov, “Deep image prior”, CVPR 2018
A.Shocher, “Zero-Shot Super-Resolution using Deep Internal Learning”, CVPR 2018
A.Shocher, “InGAN: Capturing and Remapping the “DNA” of a Natural Image”, ICCV 2019

Related Work

- **Generative models for image manipulation**
  - image editing, sketch2image, image-to-image translation tasks are trained on class specific datasets.

    (CycleGAN, SketchyGAN, Scribbler)

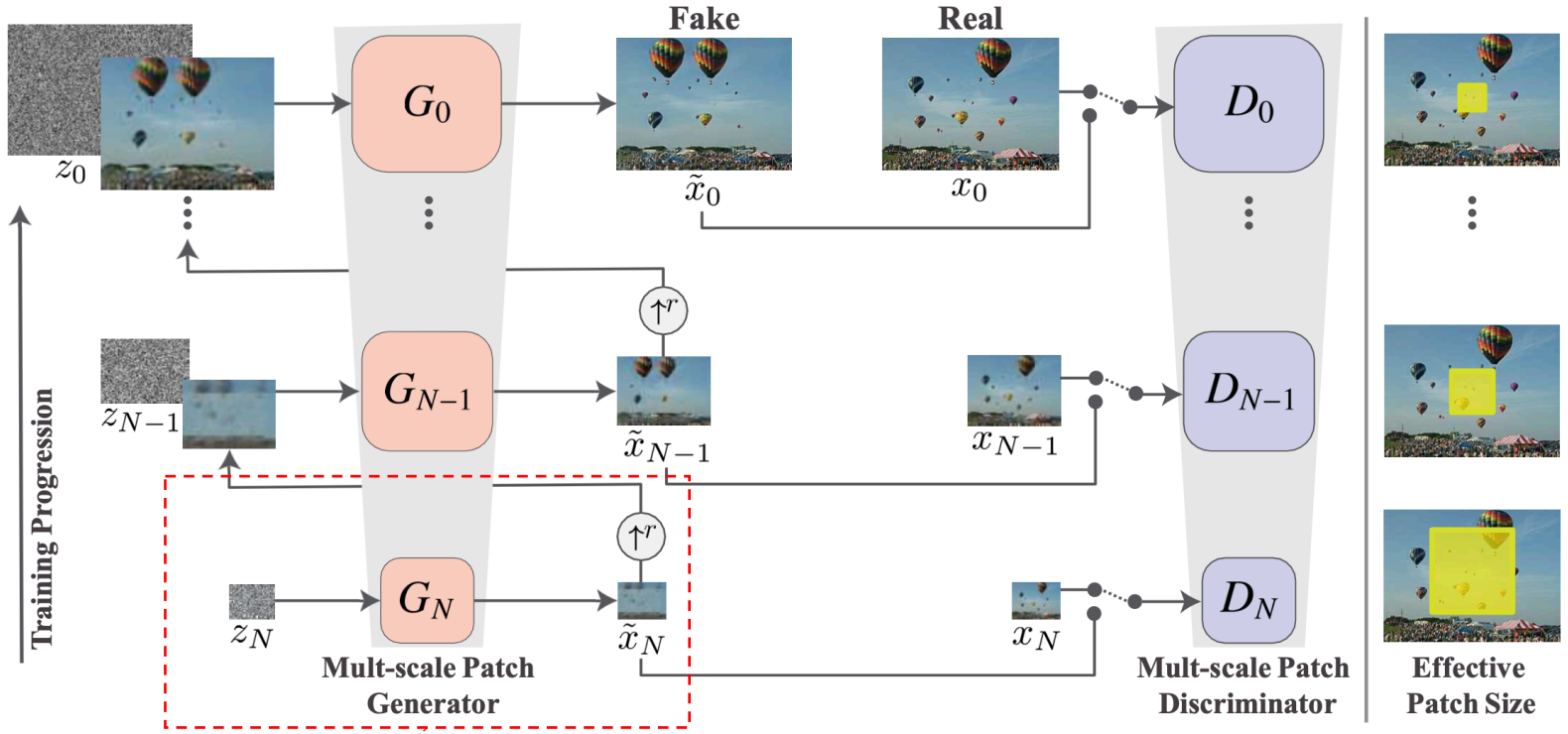
J.Zhu, "Unpaired image-to-image translation using cycle-consistent adversarial networks", ICCV 2017.
W.Chen, "Sketchygan: towards diverse and realistic sketch to image synthesis", CVPR 2018.
P.Sangkloy, "Scribbler: Controlling deep image synthesis with sketch and color", CVPR 2017.

Methodology

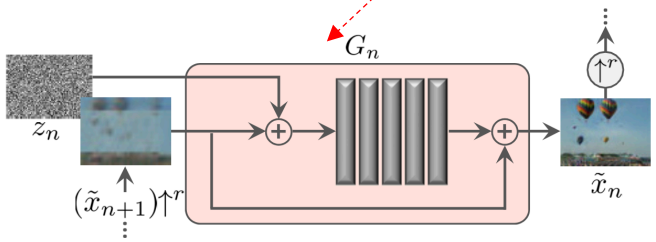
Architecture

: Captures the internal statistics of a single training image x

Multi-scale architecture



Single-scale generation



Multi-scale architecture

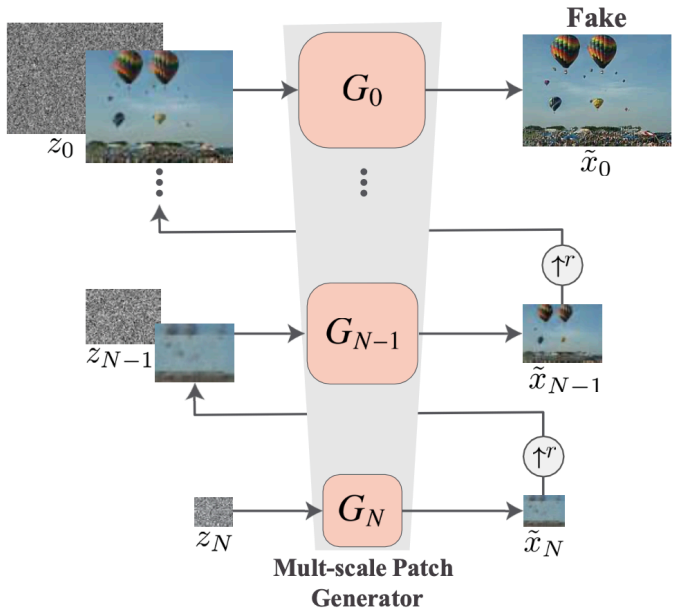
Pyramid of generators : $\{G_0, \ldots, G_N\}$

Image pyramid : $\{x_0, \ldots, x_N\}$

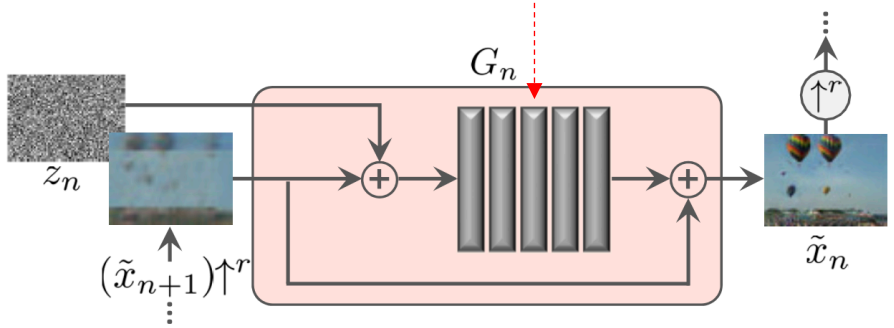
At coarsest scale : $\tilde{x}_N = G_N(z_N).$

$$\tilde{x}_n = G_n\left(z_n,\ (\tilde{x}_{n+1})\uparrow^r\right), \qquad n < N.$$

$$\tilde{x}_n = (\tilde{x}_{n+1})\uparrow^r + \psi_n(z_n + (\tilde{x}_{n+1})\uparrow^r),$$



5 conv-blocks (Conv(3x3) – BN – LeakyReLU)



Loss function

- **Adversarial Loss**

$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n).$$

- ✓ Use **WGAN-GP loss**, which we found to increase training stability, where final discrimination score is the average over the patch discrimination map.
- ✓ Define the loss over the whole image rather than over random crops.
  (This allows the net to learn boundary conditions)

Loss function

- **Reconstruction Loss**

$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n).$$

want to ensure that **there exists a specific set of input noise maps**, which generates the original image x

$\{z_N^{\text{rec}}, z_{N-1}^{\text{rec}} \ldots, z_0^{\text{rec}}\} = \{z^*, 0, \ldots, 0\}$, where z^* : some fixed noise map

$$\mathcal{L}_{\text{rec}} = \|G_n(0, (\tilde{x}_{n+1}^{\text{rec}}) \uparrow^r) - x_n\|^2 \qquad \text{for } n = N, \text{ we use } \mathcal{L}_{\text{rec}} = \|G_N(z^*) - x_N\|^2$$

$\tilde{x}_n^{\text{rec}}$: determine the std. σ_n of the noise z_n in each scale

Loss function

- **Reconstruction Loss**

$$\mathcal{L}_{\text{rec}} = \|G_n(0, (\tilde{x}_{n+1}^{\text{rec}}) \uparrow^r) - x_n\|^2 \quad \text{for } n = N, \text{ we use } \mathcal{L}_{\text{rec}} = \|G_N(z^*) - x_N\|^2$$

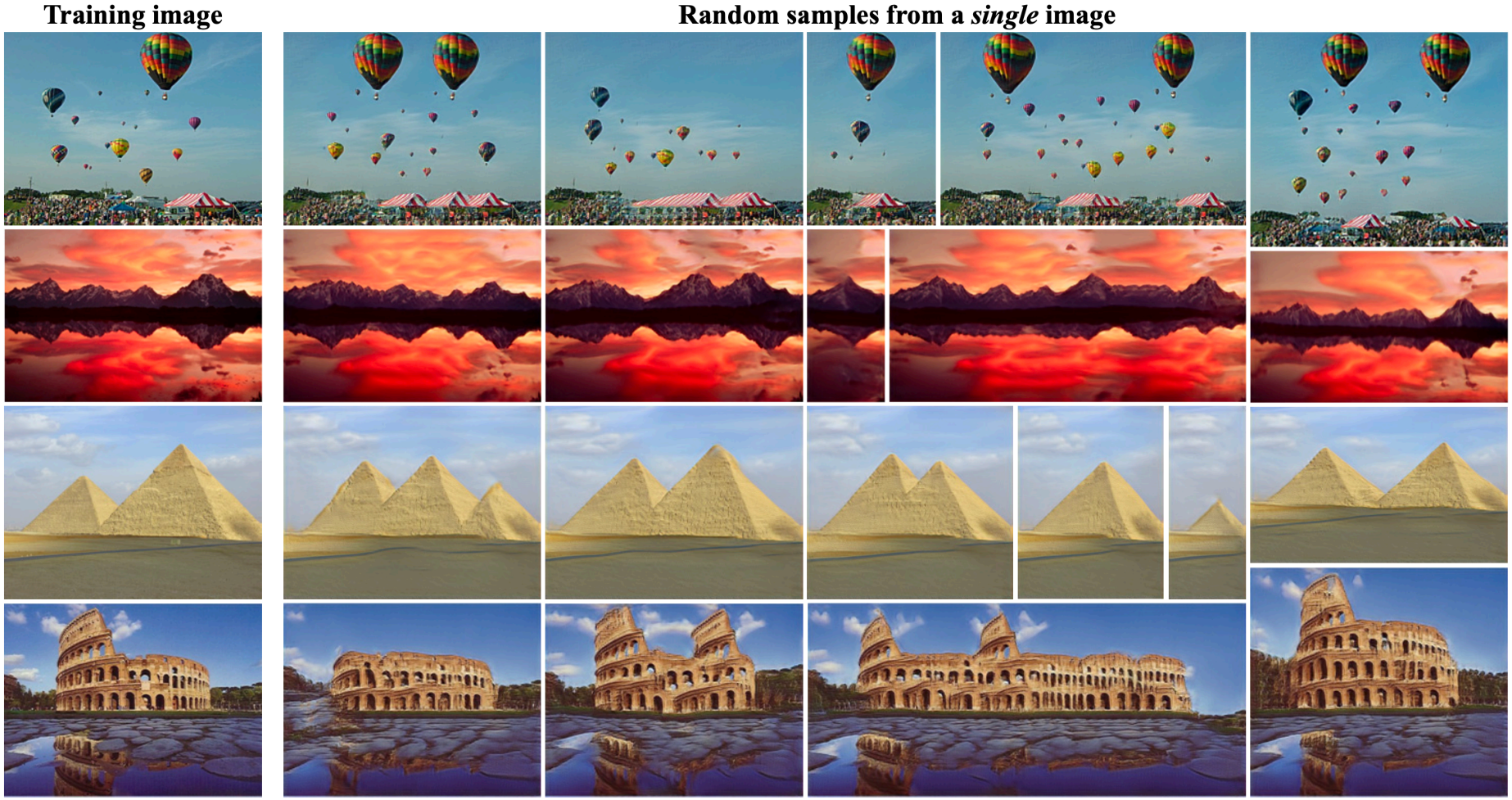
During training…

$\tilde{x}_n^{\text{rec}}$: determine the std. σ_n of the noise z_n in each scale

take σ_n to be proportional to the RMSE between $\tilde{x}_{n+1}^{rec} \uparrow^r$ and x_n

Results

Qualitative examples (random samples)



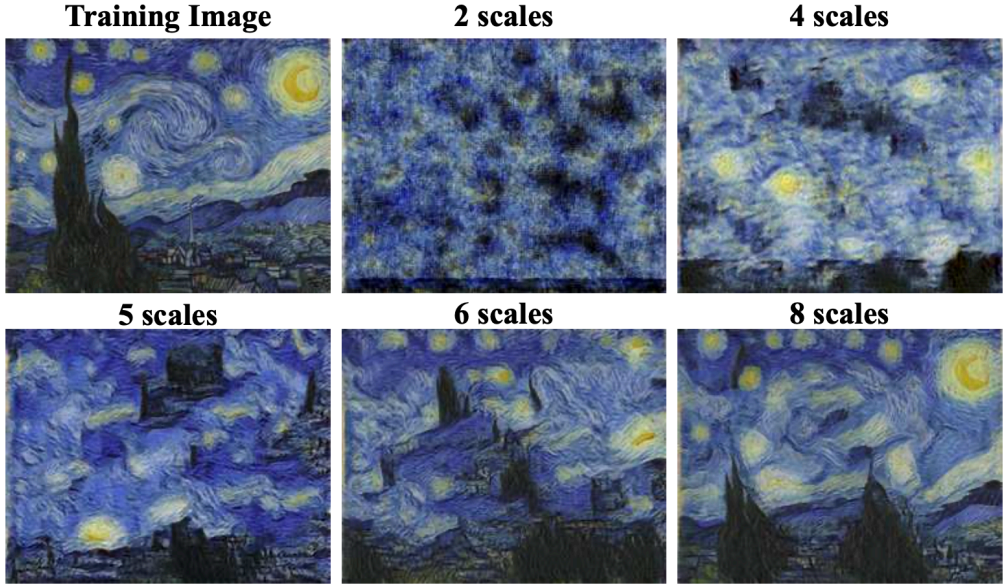
SinGAN successfully preservers global structure of objects, as well as fine texture

Qualitative examples (high resolution)

SinGAN's architecture is resolution agnostic and can thus be used on high resolution images.

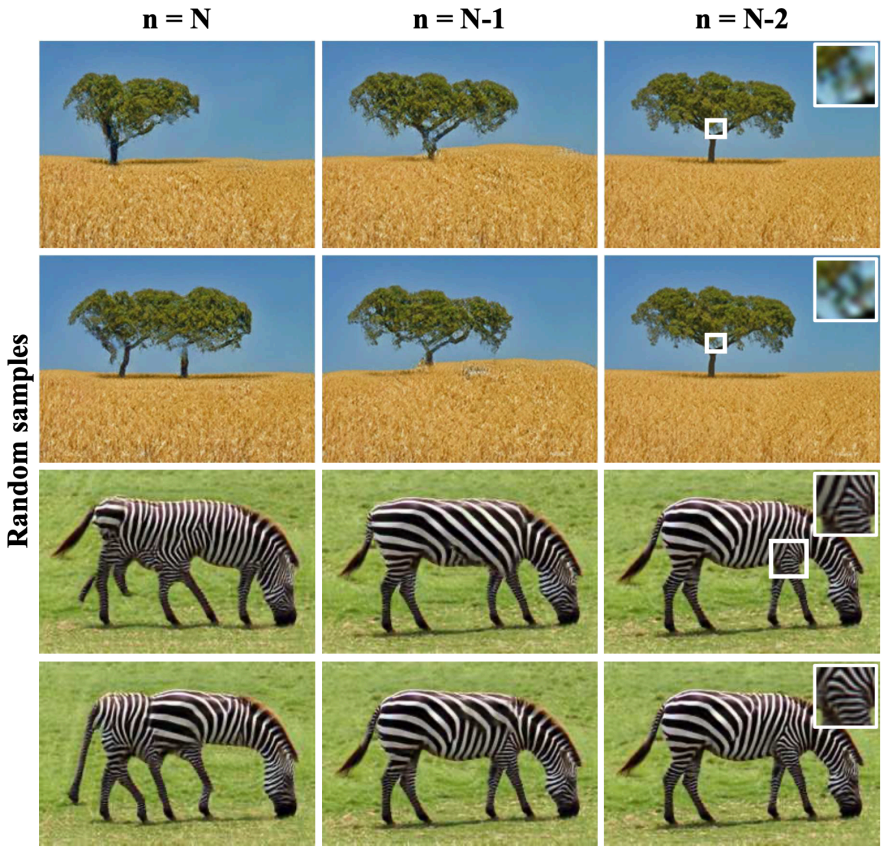
Qualitative examples (different scales)

Training



As the number of scales increases, SinGAN manages to capture larger structures as well as the global arrangement of objects in the scene

Inference



Can preserve the shape and pose of the Zebra and only change its stripe texture

Quantitative Evaluation

Two metrics.

- Amazon Mechanical Tur (AMT) "Real/Fake" user study
- Single Image Frechet Inception Distance

1st Scale	Diversity	Survey	Confusion
N	0.5	paired	$21.45\% \pm 1.5\%$
		unpaired	$42.9\% \pm 0.9\%$
$N-1$	0.35	paired	$30.45\% \pm 1.5\%$
		unpaired	$47.04\% \pm 0.8\%$

paired (real vs fake) : 50trials, 1 second, workers were asked to pick the fake image.

Unpaired (either real or fake) : single image for 1 second, were asked if it was fake.

Diversity : diversity of the generated images, std of the intensity values of each pixel over 100 generated images.

Real Images were randomly picked from the "places" databases

Quantitative Evaluation

Two metrics.

- Amazon Mechanical Tur (AMT) "Real/Fake" user study
- Single Image Frechet Inception Distance (SIFID)

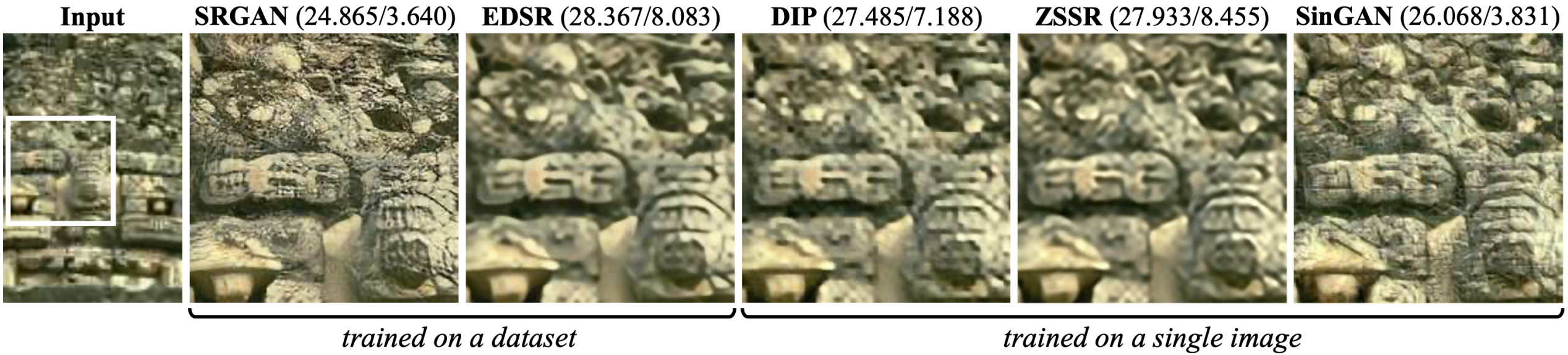
1st Scale	SIFID	Survey	SIFID/AMT Correlation
N	0.09	paired	-0.55
		unpaired	-0.22
$N-1$	0.05	paired	-0.56
		unpaired	-0.34

SIFID : average score for 50 images

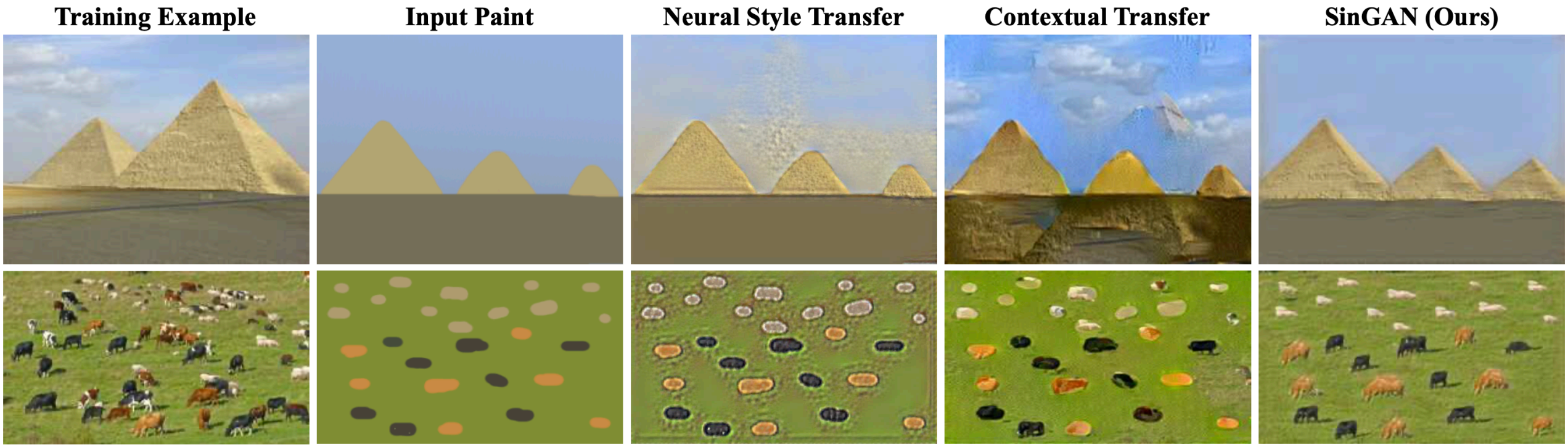
Small SIFID is typically a good indicator for a large confusion rate.

Applications

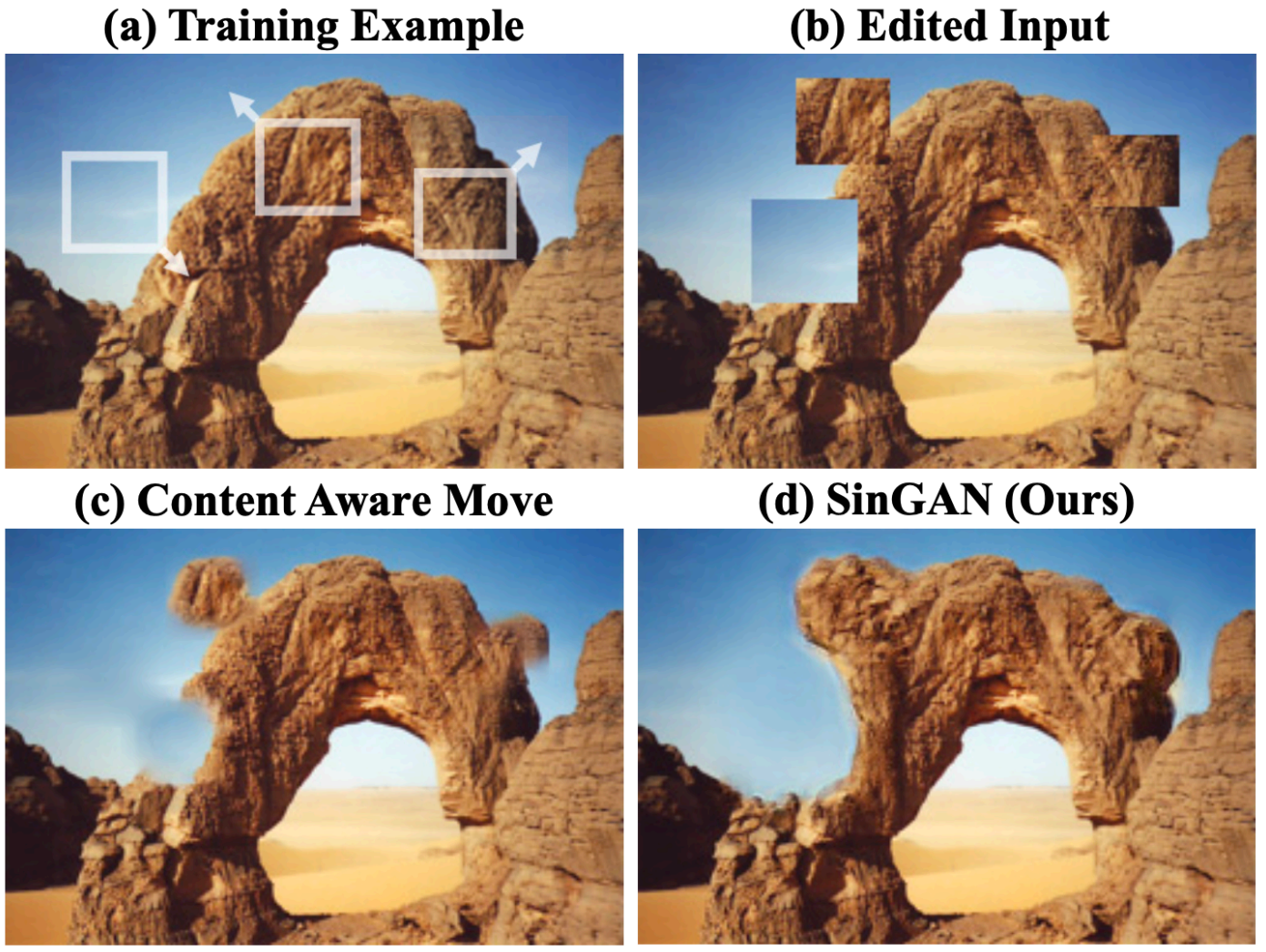
Super-Resolution



Paint-to-Image



Editing


(a) Training Example
(b) Edited Input
(c) Content Aware Move
(d) SinGAN (Ours)

Harmonization

Conclusion

Conclusion

- Introduce a new **unconditional generative scheme** that is learned from a single natural image.

- **Internal learning is inherently limited in terms of semantic diversity**. But, SinGAN can provide a very powerful tool for a wide range of image manipulation tasks.

Appendix

- Maximum Likelihood Estimation (MLE) as minimizing Kullback-Leibler Divergence (KLD)

MLE as minimizing KLD

Minimizing KL Divergence

$$\begin{aligned}\theta_{\text{min KL}} &= \arg\min_{\theta} D_{KL}(\hat{p}_{\text{data}}||p_{\text{model}}) \\ &= \arg\min_{\theta} E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log \hat{p}_{\text{data}}(\mathbf{x}) - \log p_{\text{model}}(\mathbf{x})]\end{aligned}$$

independent of model parameters

Maximum Likelihood Estimation

$$\theta_{ML} = arg\max_{\theta} P_{model}\,(Y|X;\theta)$$

MLE as minimizing KLD

Minimizing KL Divergence

$$\theta_{\text{min KL}} = \arg\min_{\theta} D_{KL}(\hat{p}_{\text{data}} || p_{\text{model}})$$
$$= \arg\min_{\theta} E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log \hat{p}_{\text{data}}(\mathbf{x}) - \log p_{\text{model}}(\mathbf{x})]$$

$$\theta_{\text{min KL}} = \arg\min_{\theta} -E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log p_{\text{model}}(\mathbf{x}|\theta)]$$

Maximum Likelihood Estimation

$$\theta_{ML} = arg\max_{\theta} P_{model}(Y|X;\theta)$$

MLE as minimizing KLD

Minimizing KL Divergence

$$\theta_{\text{min KL}} = \arg\min_{\theta} D_{KL}(\hat{p}_{\text{data}} || p_{\text{model}})$$
$$= \arg\min_{\theta} E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log \hat{p}_{\text{data}}(\mathbf{x}) - \log p_{\text{model}}(\mathbf{x})]$$

$$\theta_{\text{min KL}} = \arg\min_{\theta} -E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log p_{\text{model}}(\mathbf{x}|\theta)]$$

-argmin → argmax

Maximum Likelihood Estimation

$$\theta_{ML} = arg\max_{\theta} P_{model}(Y|X;\theta)$$

MLE as minimizing KLD

Minimizing KL Divergence

$$\begin{aligned}\theta_{\text{min KL}} &= \arg\min_{\theta} D_{KL}(\hat{p}_{\text{data}} || p_{\text{model}}) \\ &= \arg\min_{\theta} E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log \hat{p}_{\text{data}}(\mathbf{x}) - \log p_{\text{model}}(\mathbf{x})]\end{aligned}$$

$$\theta_{\text{min KL}} = \arg\min_{\theta} -E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log p_{\text{model}}(\mathbf{x}|\theta)]$$

$$\theta_{\text{min KL}} = \arg\max_{\theta} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \log(p(\mathbf{x_i}|\theta))$$

If the datapoints x are i.i.d, by the Law of Large Numbers

Maximum Likelihood Estimation

$$\theta_{ML} = arg\max_{\theta} P_{model}(Y|X;\theta)$$

MLE as minimizing KLD

Minimizing KL Divergence

$$\theta_{\text{min KL}} = \arg\min_{\theta} D_{KL}(\hat{p}_{\text{data}} || p_{\text{model}})$$
$$= \arg\min_{\theta} E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log \hat{p}_{\text{data}}(\mathbf{x}) - \log p_{\text{model}}(\mathbf{x})]$$

$$\theta_{\text{min KL}} = \arg\min_{\theta} -E_{\mathbf{x}\sim\hat{p}_{\text{data}}}[\log p_{\text{model}}(\mathbf{x}|\theta)]$$

$$\theta_{\text{min KL}} = \arg\max_{\theta} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \log(p(\mathbf{x_i}|\theta))$$

Maximum Likelihood Estimation

$$\theta_{ML} = arg\max_{\theta} P_{model}(Y|X;\theta)$$

Appendix

- WGAN GP LOSS

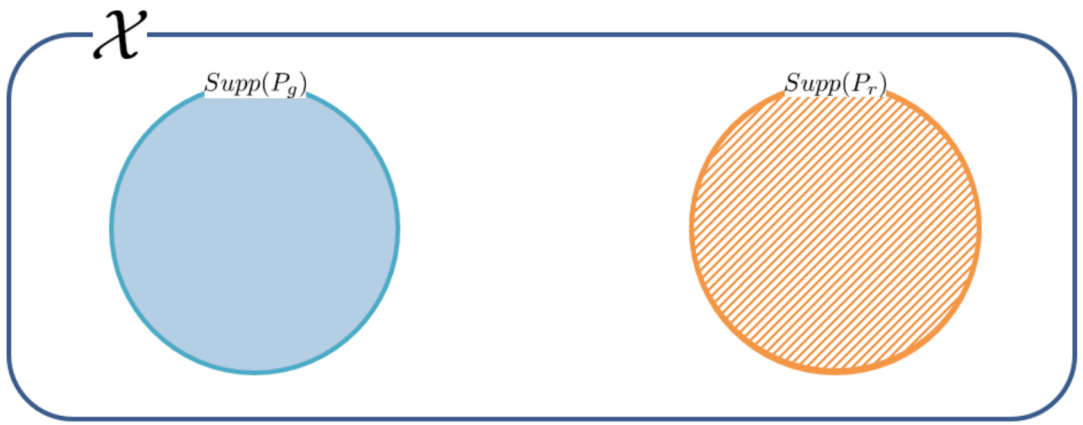
Wasserstein Distance

Problems of KLD

KLD : $D_{KL}(P_A||P_B) = \int \log \frac{p_A(x)}{p_B(x)} p_A(x)dx$

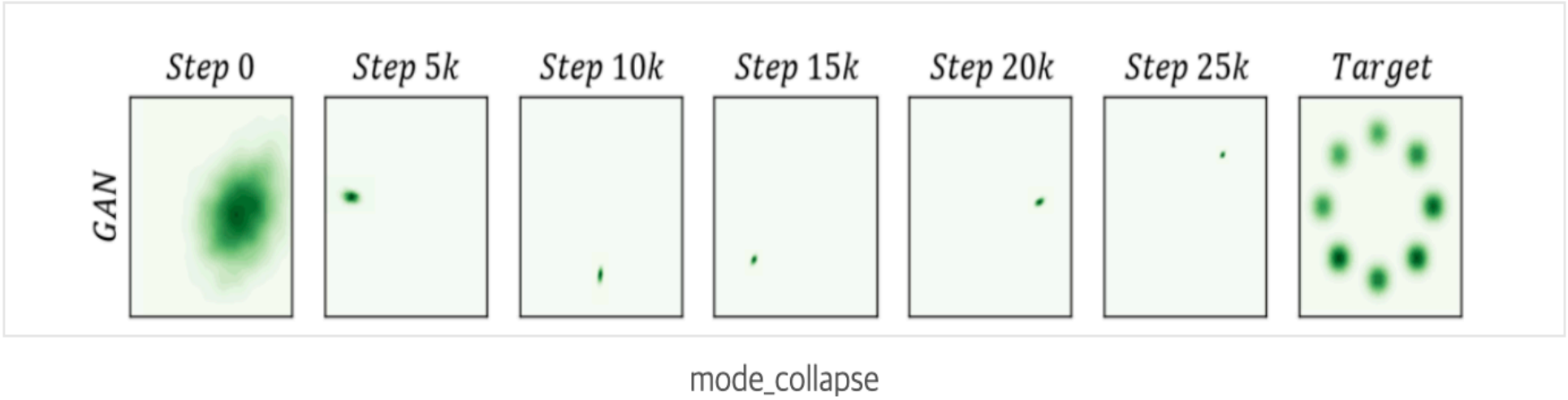
$if\ p_A(x)\ \neq 0\ , p_B(x) = 0$ → divergence

As Manifold hypothesis



Wasserstein Distance

Problems of KLD → Mode Collapse

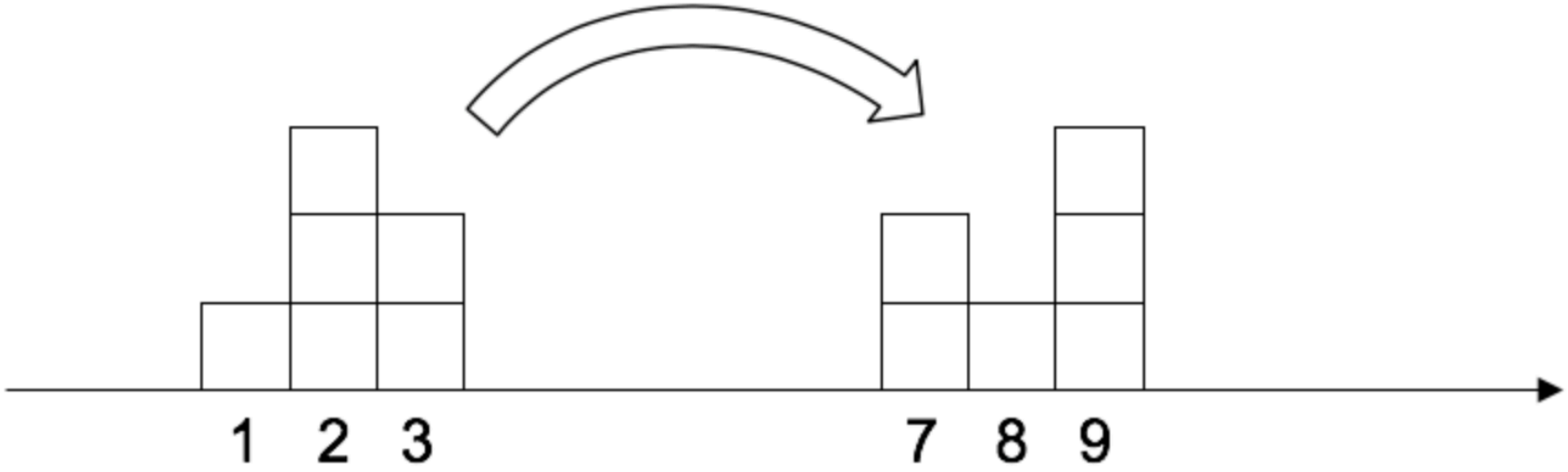


Have to find a new distance measure

Wasserstein Distance

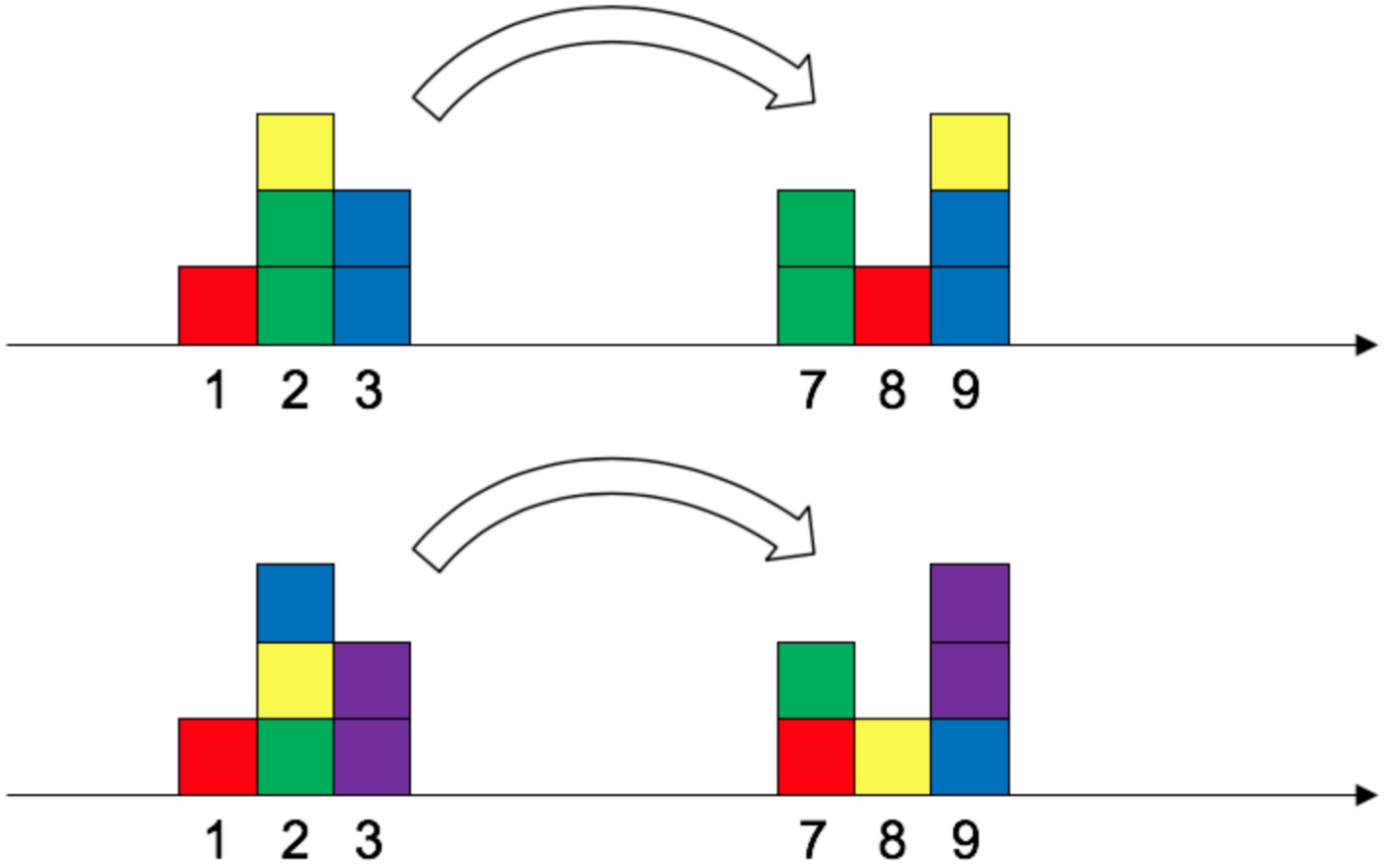
Wasserstein distance (Earth Moving distance)

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma} \left[\|x - y\| \right]$$



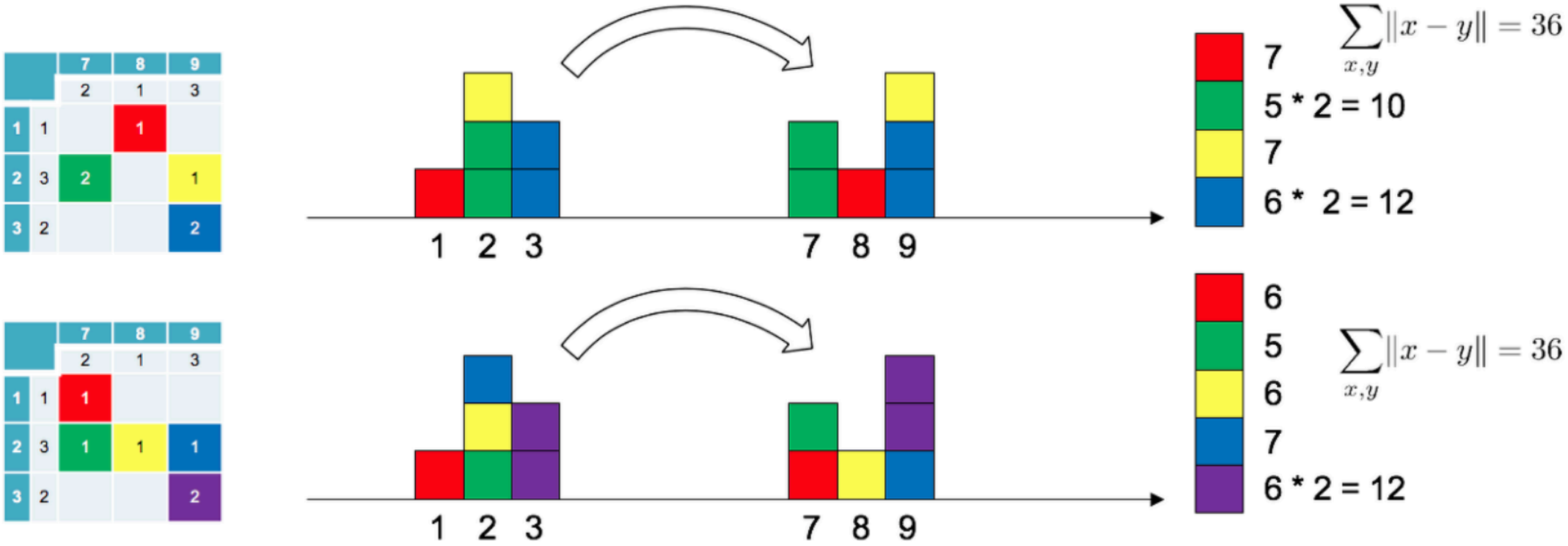
Wasserstein Distance

Wasserstein distance (Earth Moving distance)



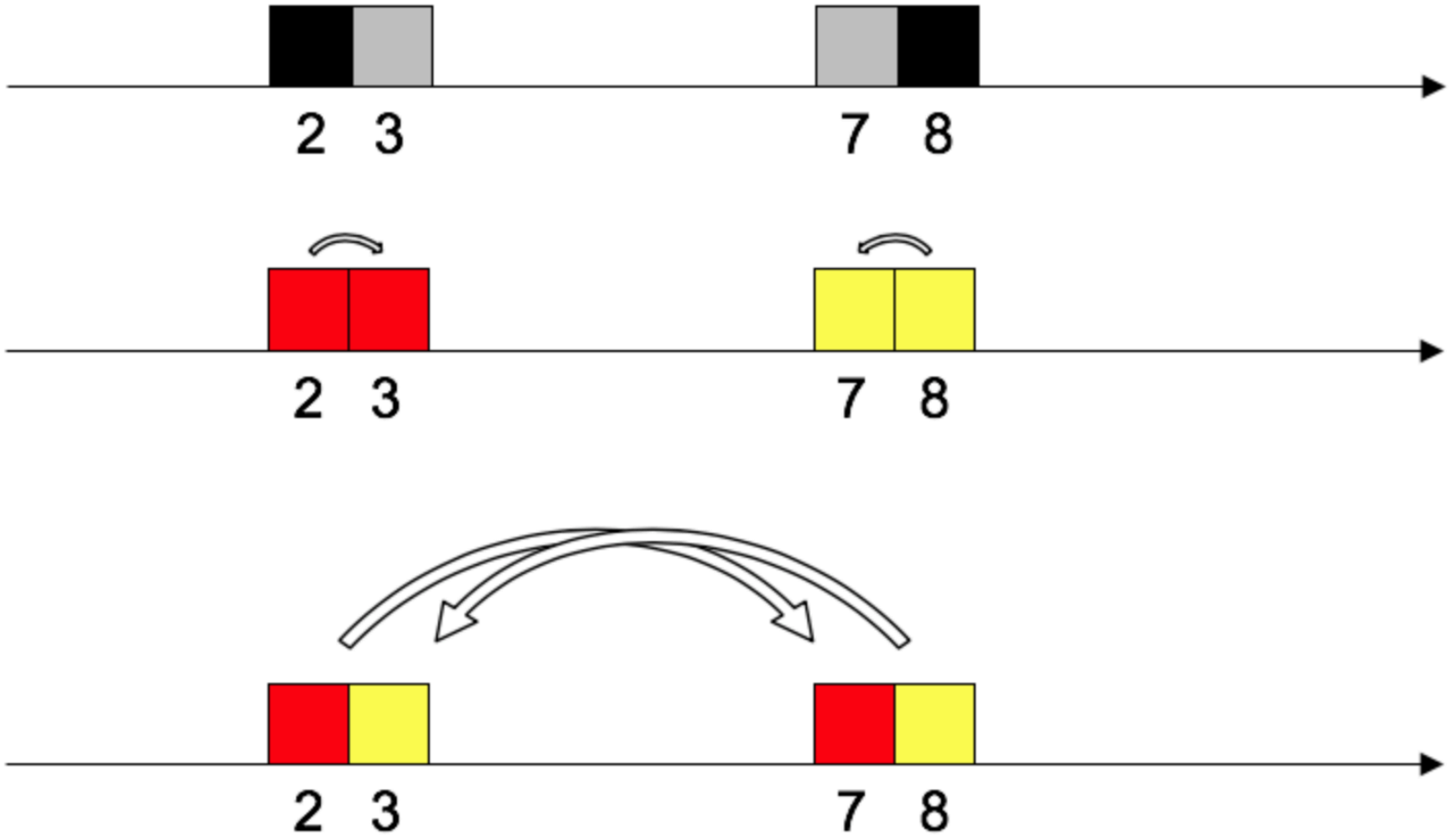
Wasserstein Distance

Wasserstein distance (Earth Moving distance)



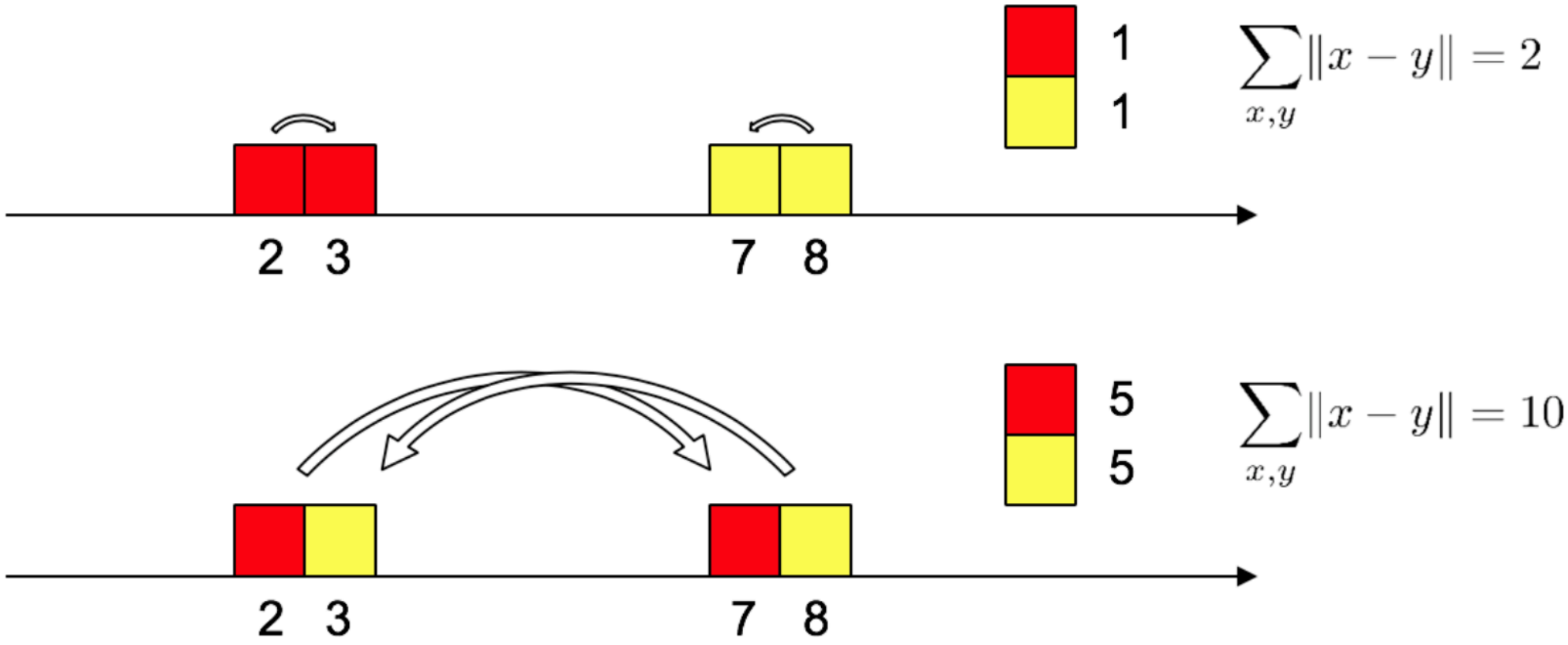
Wasserstein Distance

Change the block (black <-> gray)



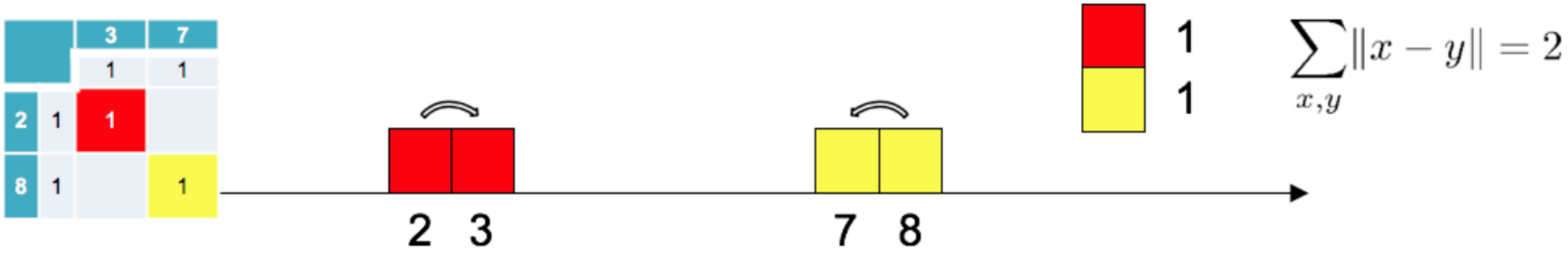
Wasserstein Distance

Change the block (black <-> gray)



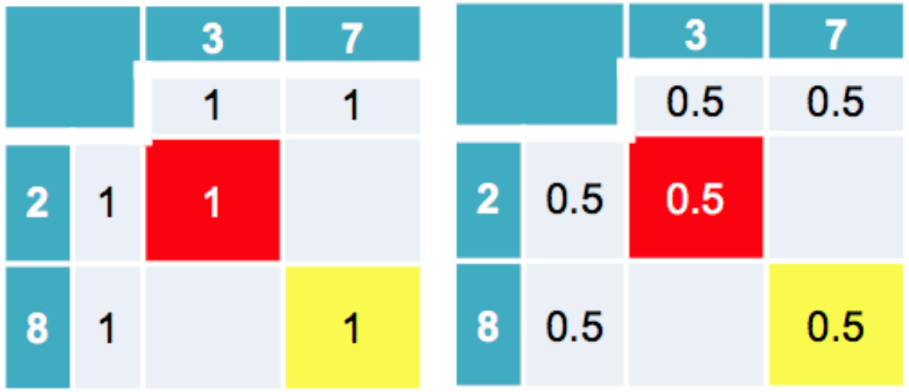
Wasserstein Distance

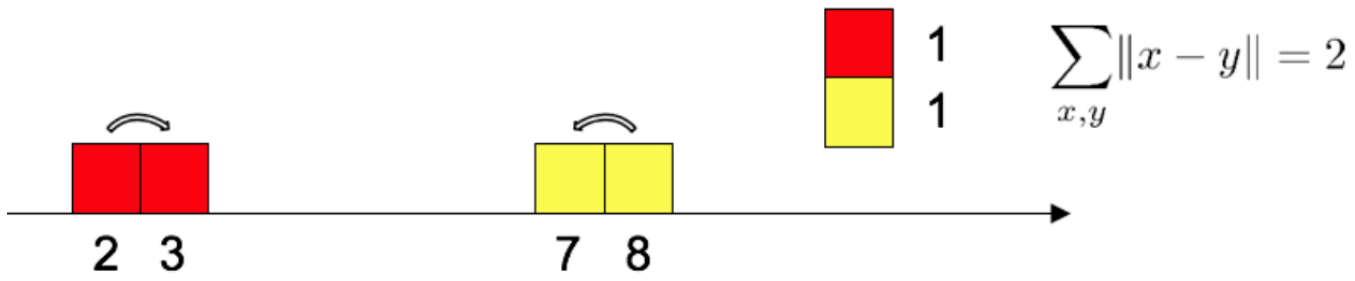
Change the block (black <-> gray)



Wasserstein Distance

Change the block (black <-> gray)





$$\sum_{x,y} \|x-y\| = 2$$

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma} \left[\|x-y\|\right]$$

$$\gamma_{X,Y}(2,3) \times |2-3| + \gamma_{X,Y}(8,7) \times |8-7| = 0.5 \times 1 + 0.5 \times 1 = 0.5$$

References

- https://haawron.tistory.com/21
- https://kionkim.github.io/2018/06/01/WGAN_1/
- https://ratsgo.github.io/statistics/2017/09/23/MLE/
- https://www.jessicayung.com/maximum-likelihood-as-minimising-kl-divergence/
- https://math.stackexchange.com/questions/1888364/when-arg-max-fx-arg-min-fx

Thank You.