Image Inpainting

Bulat Shelkhonov

MIPT

Moscow, Jan 2023

telegram: boolean17

girafe
ai

Outline

1. Problem statement
2. Metrics
3. Datasets
4. Architectures
5. My experiments

girafe ai

Problem statement

girafe
ai

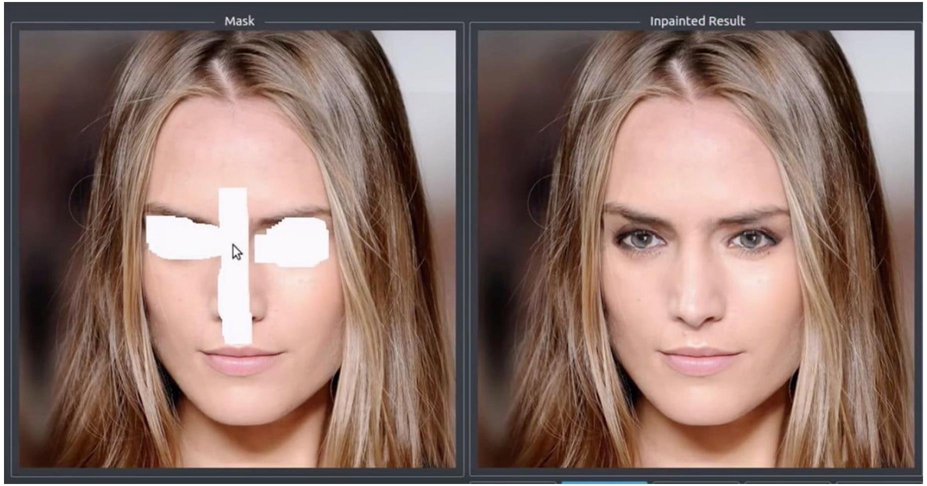
01

What is image inpainting?

Image Inpainting is a task of reconstructing missing regions in an image.

Applications:

- object removal
- image restoration
- manipulation
- compositing
- etc



Examples

Metrics

girafe
ai

02

Metrics

- Peak signal-to-noise ratio (PSNR)
- Structural simularity (SSIM)
- Learned Perceptual Image Patch Similarity (LPIPS)
- Frechet inception distance (FID)

Metrics: PSNR

MAX_I – maximum possible pixel value of the image (2^bits - 1)

For color images MSE is the sum over all squared value differences

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

girafe ai

Metrics: SSIM

Based on 3 components:

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$ – luminance

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$$ – contrast

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$ – structure

- μ_x the pixel sample mean of x;
- μ_y the pixel sample mean of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;

c_1 c_2 – to stabilize division

girafe ai

https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf

Metrics: SSIM

- usually applied on sliding window 8x8
- multichannel: calculate it per channel and take the average

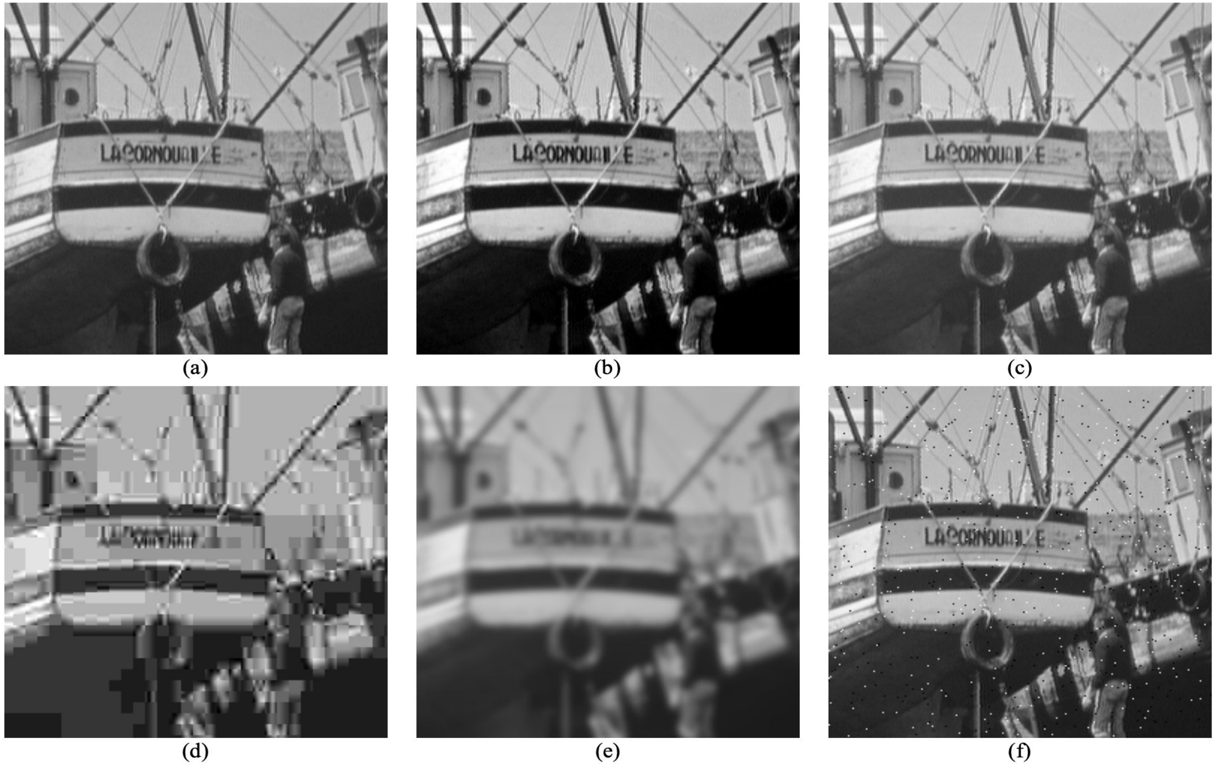
$$\mathrm{SSIM}(x, y) = l(x, y)^{\alpha} \cdot c(x, y)^{\beta} \cdot s(x, y)^{\gamma}$$

usually $\alpha = \beta = \gamma = 1$

https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf

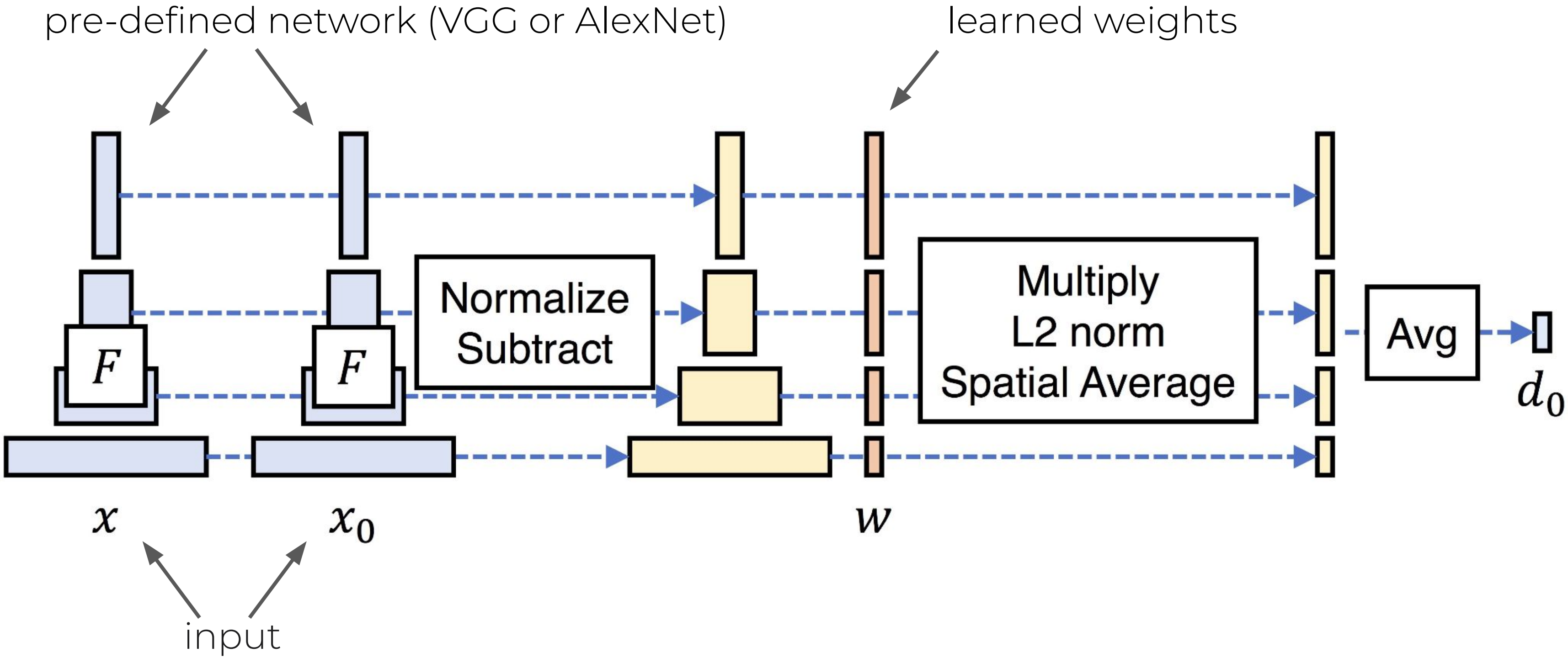
Same PSNR, different SSIM

Original | 0.9168 | 0.9900



0.6949 | 0.7052 | 0.7748

Metrics: LPIPS



CVPR 2018, https://richzhang.github.io/PerceptualSimilarity/

Metrics: FID

$$FID = ||\mu_X - \mu_Y||^2 - Tr(\sum_X + \sum_Y - 2\sqrt{\sum_X \sum_Y})$$

X, Y – real and fake embeddings (activation from the Inception model)

Authors recommend evaluate on dataset with size at least 10k images

girafe ai

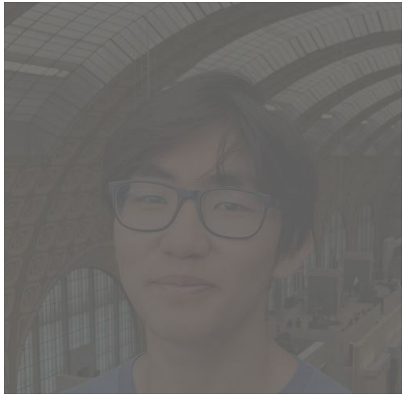
NeurIPS 2017, https://arxiv.org/pdf/1706.08500.pdf

Metrics: Comparison

Reference

Blurred
SSIM=0.75 (best)
PSNR=31 (best)
LPIPS=0.3

Low Contrast
SSIM=0.41
PSNR=27
LPIPS=0.57



Deformation
SSIM=0.40
PSNR=29
LPIPS=0.21 (best)

Datasets

girafe
ai

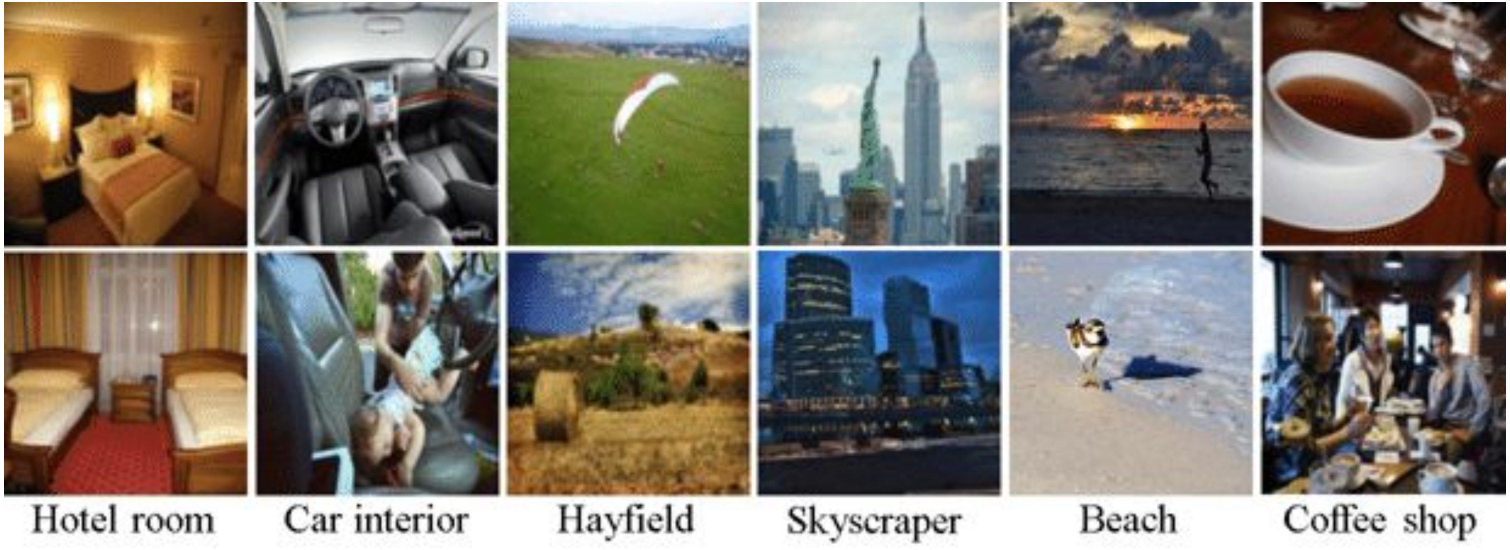
03

Datasets

Most popular datasets:

- Places2
- CelebA-HQ
- Paris StreetView
- ImageNet
- Waymo Open Dataset

Datasets: Places2



girafe ai

http://places2.csail.mit.edu/

Datasets: Places2

- **year**: 2017
- **images**: 10m
- **resolution**: >=200x200
- **article**: Places: A 10 million Image Database for Scene Recognition
- contains photos of different places with labels

http://places2.csail.mit.edu/

Datasets: CelebA-HQ

https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

Datasets: CelebA-HQ

- **year**: 2017
- **images**: 30k
- **resolution**: 1024x1024
- **article**: Progressive Growing of GANs for Improved Quality, Stability, and Variation https://arxiv.org/abs/1710.10196
- photos of people, high-quality version of CelebA

https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

Paris StreetView

http://graphics.cs.cmu.edu/projects/whatMakesParis/

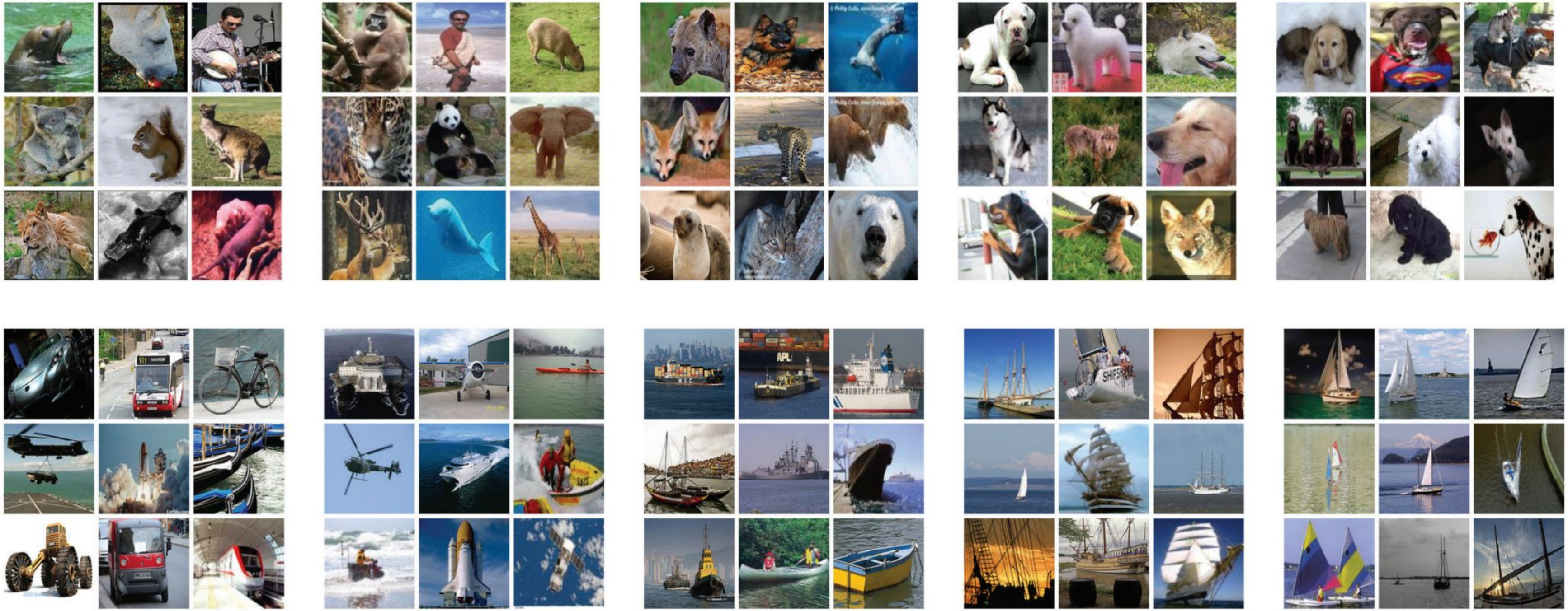
Paris StreetView

- **year**: 2012
- **images**: 15k
- **resolution**: >=936x537
- **article**: What Makes Paris Look like Paris?
- city photos

girafe ai

http://graphics.cs.cmu.edu/projects/whatMakesParis/

ImageNet



https://image-net.org/

ImageNet

- **year**: 2009
- **images**: 14m
- **resolution**: average 469×387
- **article**: ImageNet: A large-scale hierarchical image database
- contains 14m annotated images of more than 20,000 categories

https://image-net.org/

Waymo Open Dataset

girafe ai

Waymo Open Dataset

- **year**: 2020
- **images**: >12m
- **resolution**: high
- **article**: Scalability in Perception for Autonomous Driving: Waymo Open Dataset
- high resolution sensor data collected by autonomous vehicles operated by the Waymo Driver in a wide variety of conditions

girafe ai

https://arxiv.org/pdf/1912.04838v7.pdf

Architectures

girafe
ai

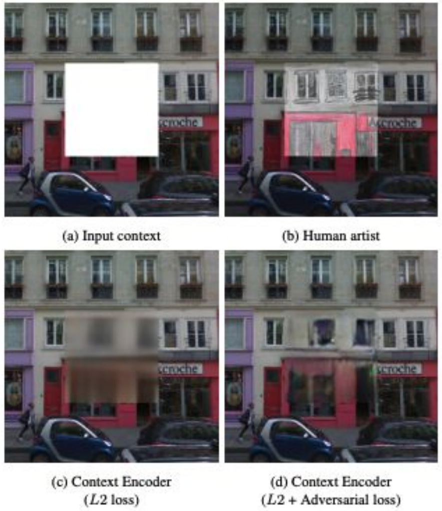
04

Architectures

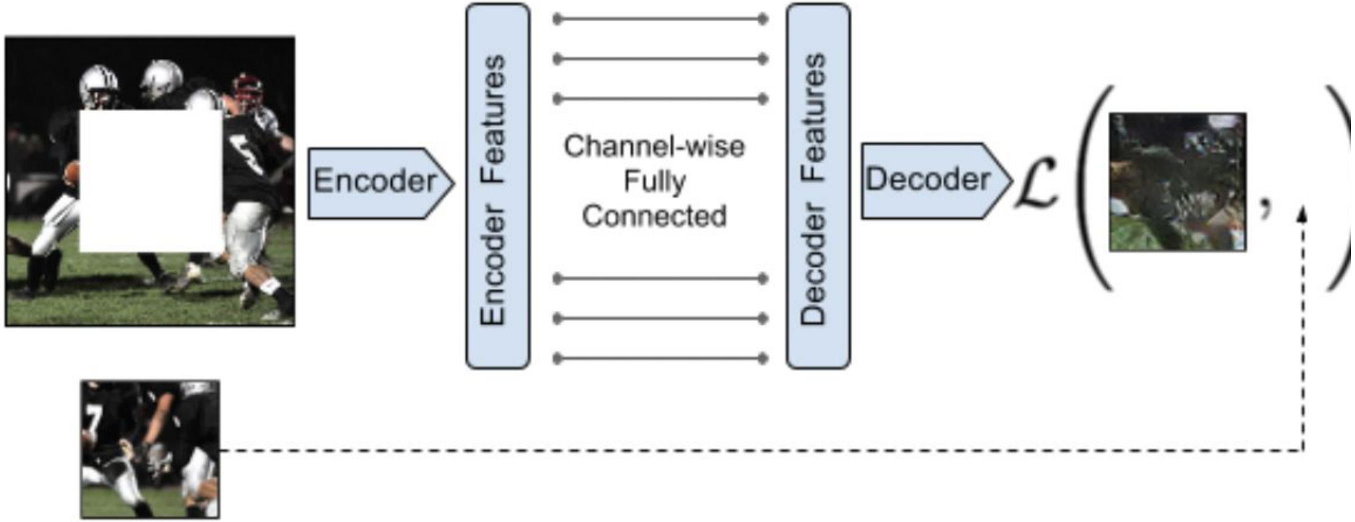
1. Context Encoders
2. Generative Image Inpainting
3. Deep Fusion Network for Image Completion (DFNet)
4. Resolution-robust Large Mask Inpainting with Fourier Convolutions (LaMa)
5. RePaint: Inpainting using Denoising Diffusion Probabilistic Models

Context Encoders: Feature Learning by Inpainting

- CVPR 2016
- University of California, Berkeley
- Paris StreetView, ImageNet
- L1, L2, PSNR
- code: https://github.com/pathak22/context-encoder



https://arxiv.org/pdf/1604.07379.pdf



https://arxiv.org/pdf/1604.07379.pdf

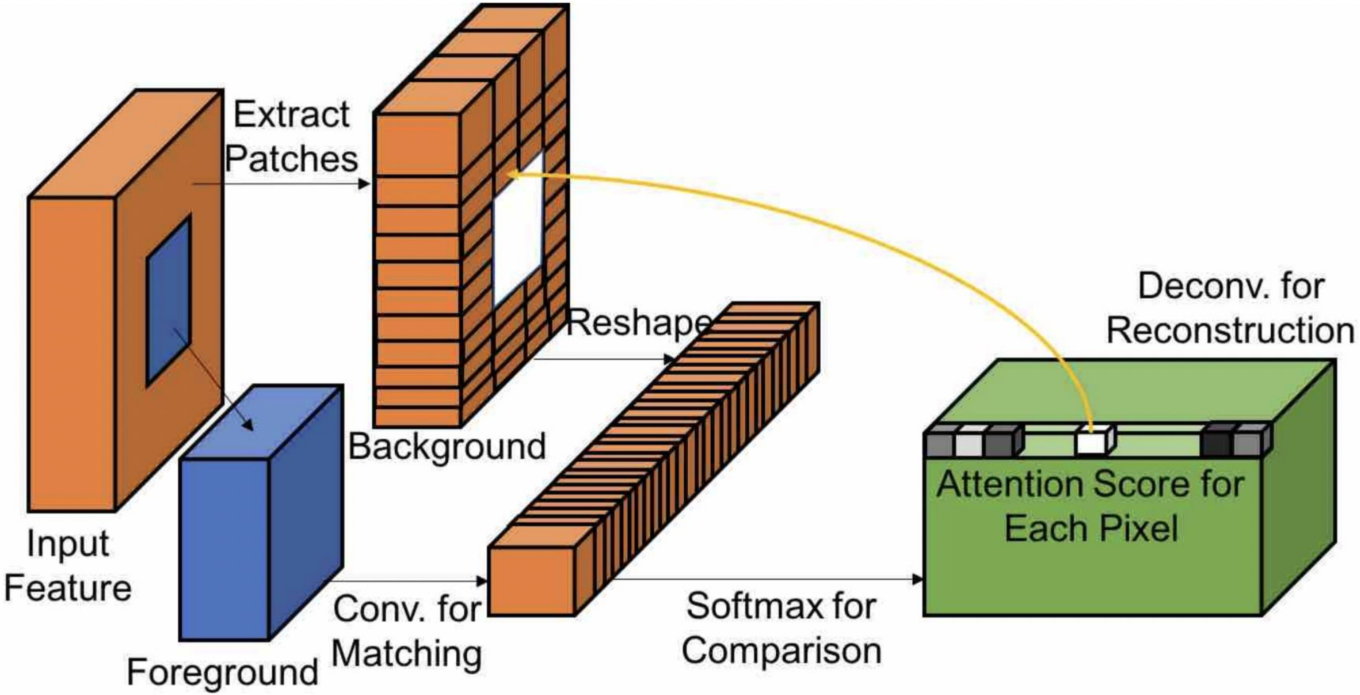
Generative Image Inpainting with Contextual Attention

- CVPR 2018
- University of Illinois at Urbana-Champaign, Adobe Research
- CelebA, CelebA-HQ, DTD, ImageNet, Places2
- L1, L2, PSNR, TV
- code: https://github.com/JiahuiYu/generative_inpainting

girafe ai

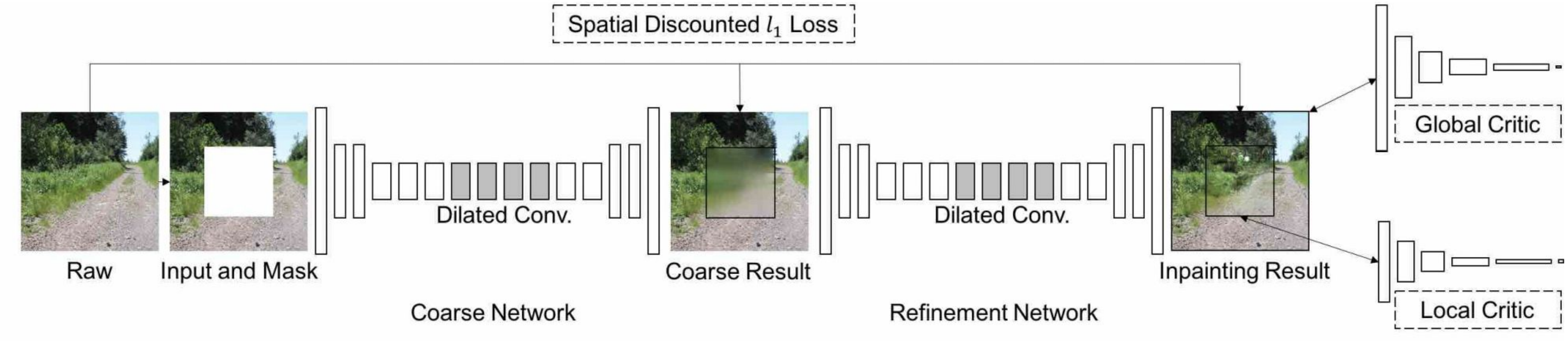
https://arxiv.org/pdf/1801.07892.pdf

Contextual Attention



https://arxiv.org/pdf/1801.07892.pdf

Coarse-to-refinement



girafe ai

https://arxiv.org/pdf/1801.07892.pdf

Deep Fusion Network for Image Completion

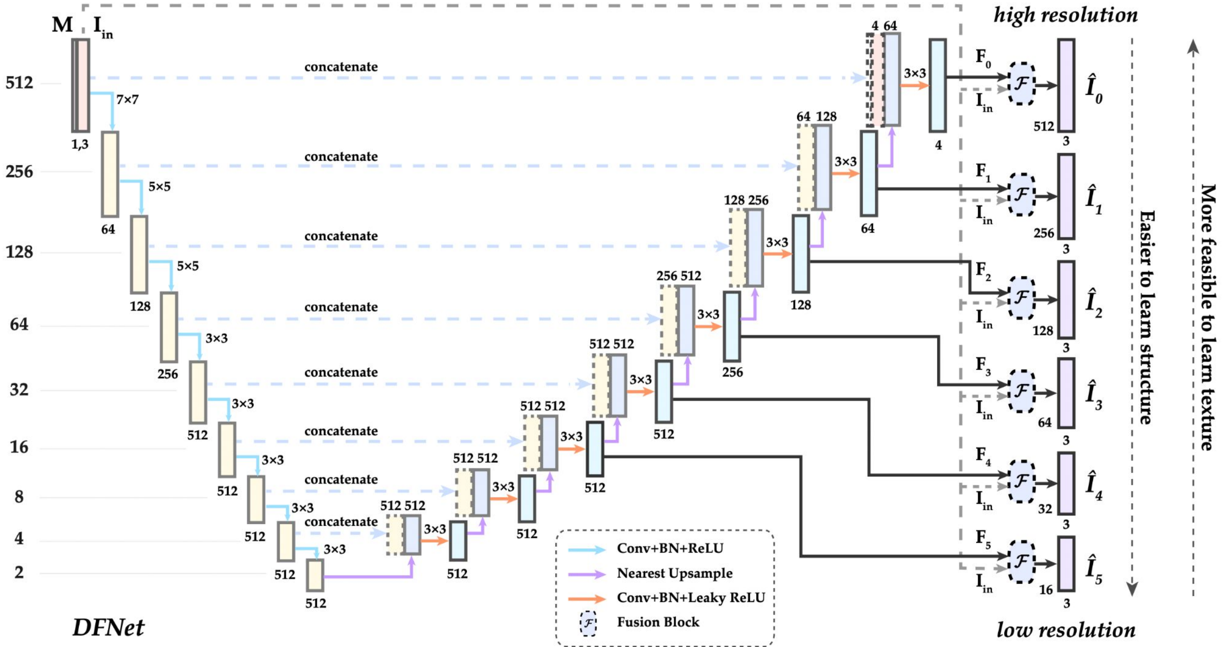
- ACMMM 2019
- Institute of Computing Technology, Chinese Academy of Sciences, Megvii Technology
- Places2, CelebA-HQ
- L1, BPE, FID
- code: https://github.com/hughplay/DFNet

$$BPE = \frac{\|\mathbf{b} \odot (\mathbf{I} - \hat{\mathbf{I}})\|_1}{\|\mathbf{b}\|_1}$$

girafe ai

https://arxiv.org/pdf/1904.08060.pdf

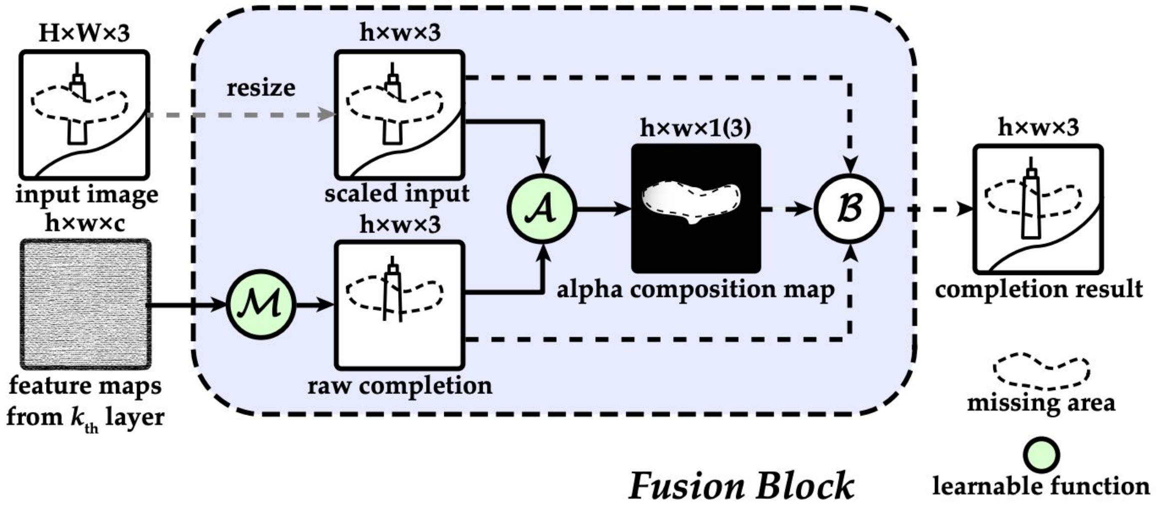
Deep Fusion Network for Image Completion



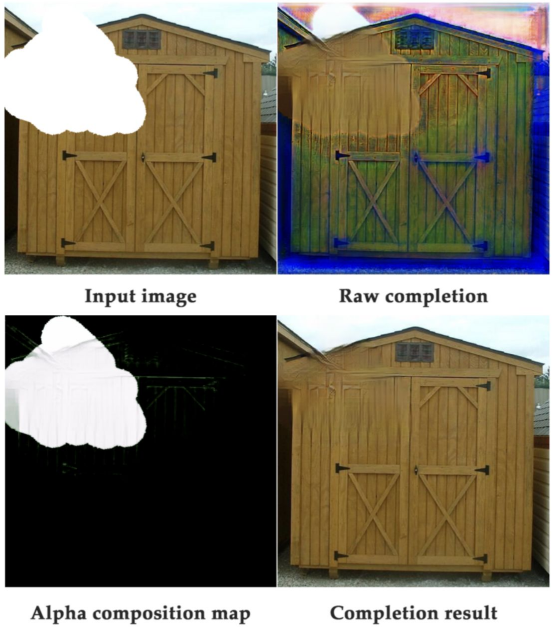
girafe ai

https://arxiv.org/pdf/1904.08060.pdf

Fusion Block



$$\hat{\mathbf{I}}_k = \mathcal{B}(\boldsymbol{\alpha}_k, \mathbf{C}_k, \mathbf{I}_k) = \boldsymbol{\alpha}_k \odot \mathbf{C_k} + (1 - \boldsymbol{\alpha}_k) \odot \mathbf{I}_k$$



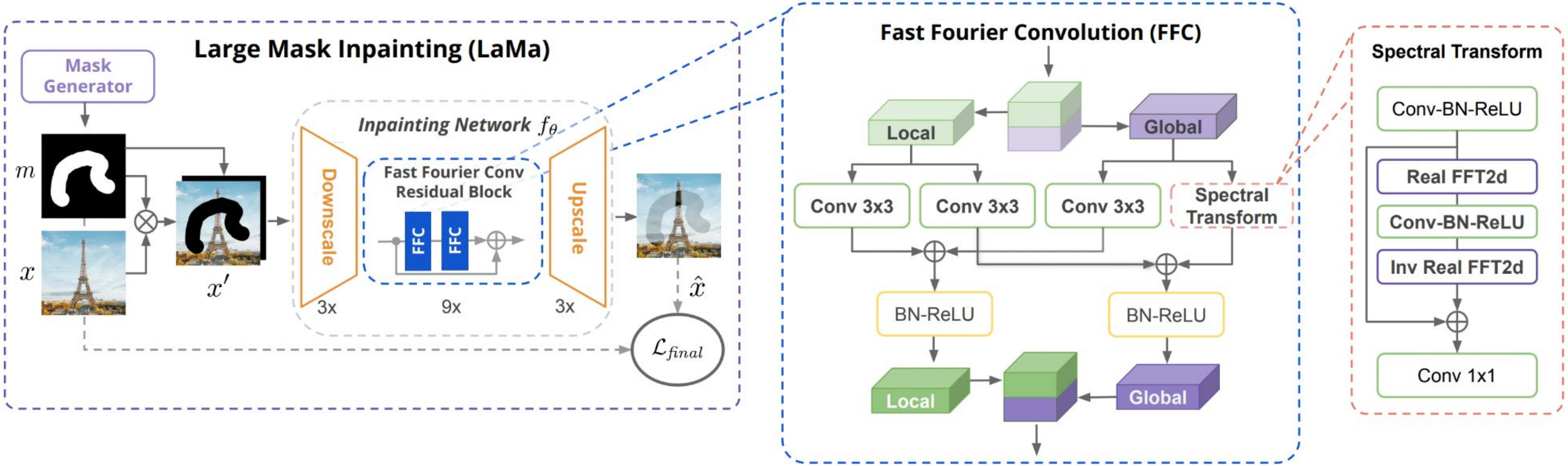
https://arxiv.org/pdf/1904.08060.pdf

Resolution-robust Large Mask Inpainting with Fourier Convolutions

- WACV 2022
- Samsung AI Center Moscow, School of Computer and Communication Sciences, Skolkovo Institute of Science and Technology
- Places2, CelebA-HQ
- FID, LPIPS
- code: https://github.com/saic-mdal/lama

https://arxiv.org/pdf/2109.07161.pdf

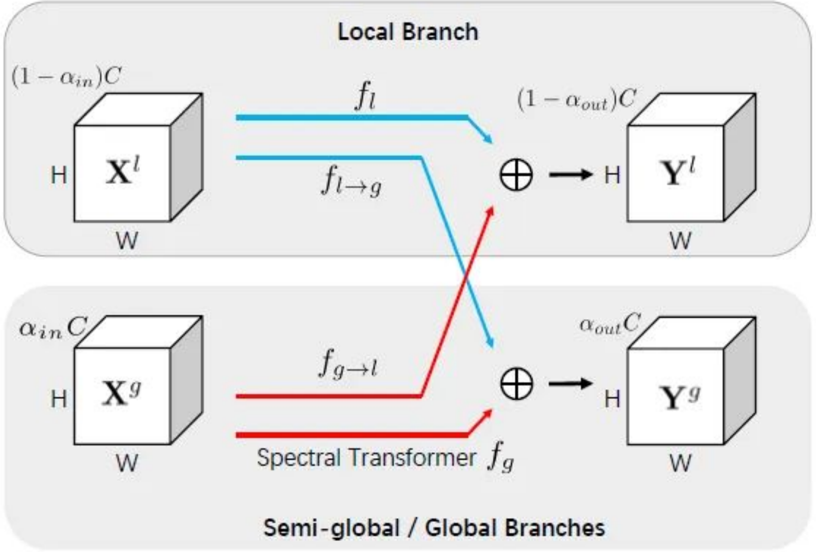
Resolution-robust Large Mask Inpainting with Fourier Convolutions



girafe ai

https://arxiv.org/pdf/2109.07161.pdf

Fast Fourier Convolution



$$\mathbf{Y}^l = \mathbf{Y}^{l\to l} + \mathbf{Y}^{g\to l} = f_l(\mathbf{X}^l) + f_{g\to l}(\mathbf{X}^g),$$
$$\mathbf{Y}^g = \mathbf{Y}^{g\to g} + \mathbf{Y}^{l\to g} = f_g(\mathbf{X}^g) + f_{l\to g}(\mathbf{X}^l).$$

https://medium.com/mlearning-ai/fast-fourier-convolution-a-detailed-view-a5149aae36c4

RePaint: Inpainting using Denoising Diffusion Probabilistic Models

- CVPR 2022
- Computer Vision Lab, ETH Zurich
- CelebA-HQ, ImageNet
- Votes, LPIPS
- code: https://github.com/andreas128/RePaint

https://arxiv.org/pdf/2201.09865.pdf

RePaint: Inpainting using Denoising Diffusion Probabilistic Models

Input | Denoising 0% | Denoising 60% | Denoising 75% | Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5

https://arxiv.org/pdf/2201.09865.pdf

About DDPMs

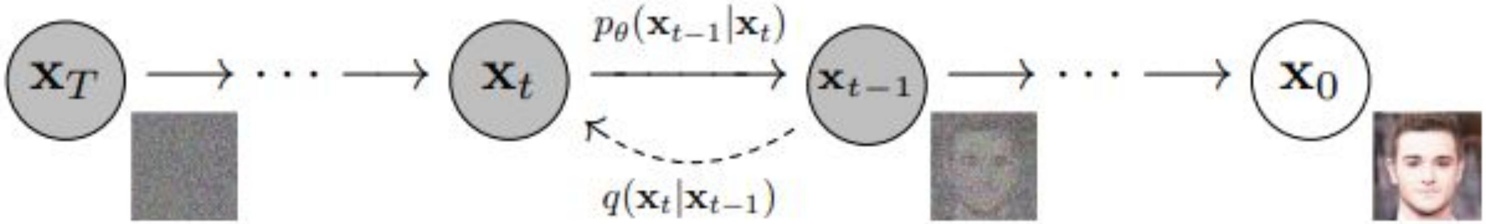


Figure 2: The directed graphical model considered in this work.

About DDPMs

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on

$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$

6: **until** converged

RePaint: Inpainting using Denoising Diffusion Probabilistic Models

Algorithm 1 Inpainting using our RePaint approach.

$x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \ldots, 1$ **do**
  for $u = 1, \ldots, U$ **do**
    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\epsilon = \mathbf{0}$
    $x_{t-1}^{\text{known}} = \sqrt{\bar{\alpha}_t} x_0 + (1 - \bar{\alpha}_t)\epsilon$
    $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
    $x_{t-1}^{\text{unknown}} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(x_t, t) \right) + \sigma_t z$
    $x_{t-1} = m \odot x_{t-1}^{\text{known}} + (1 - m) \odot x_{t-1}^{\text{unknown}}$
    if $u < U$ and $t > 1$ **then**
      $x_t \sim \mathcal{N}(\sqrt{1 - \beta_{t-1}} x_{t-1}, \beta_{t-1}\mathbf{I})$
    end if
  end for
end for
return x_0

https://arxiv.org/pdf/2201.09865.pdf

Experiments

girafe ai

05

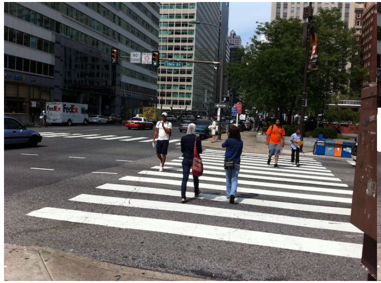
Visual Comparison: My Face

Original Masked Contexual Attention

DFNet LaMa Repaint

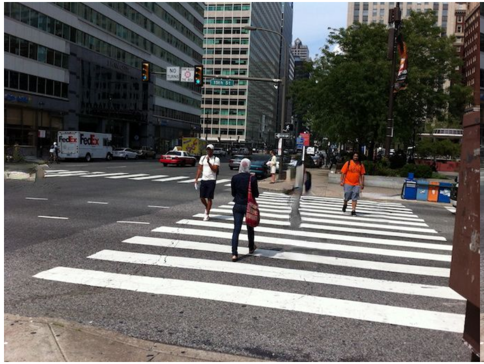
Visual Comparison: Street

Original



Masked

Contexual Attention



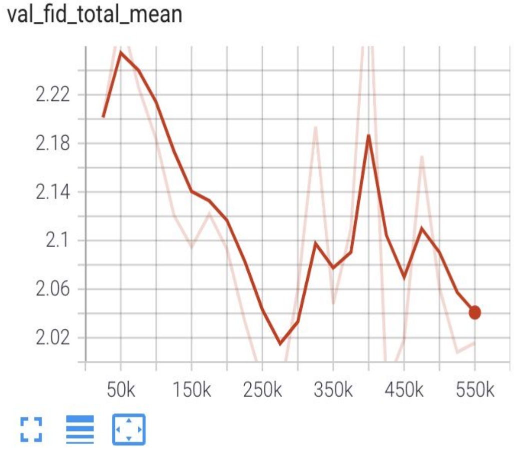
DFNet

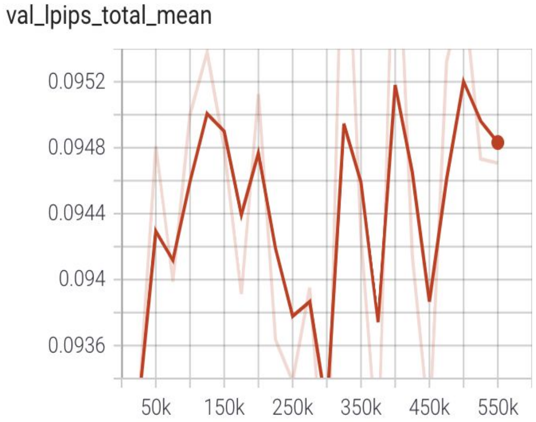
LaMa

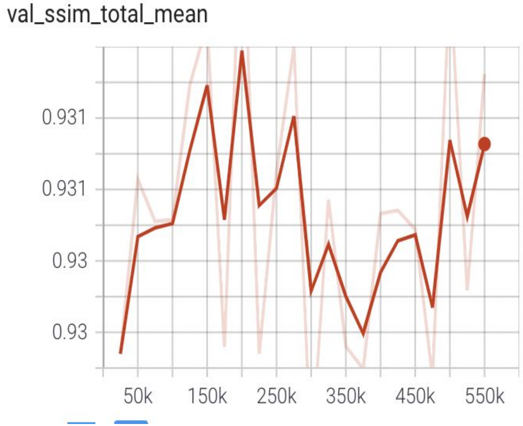
Repaint

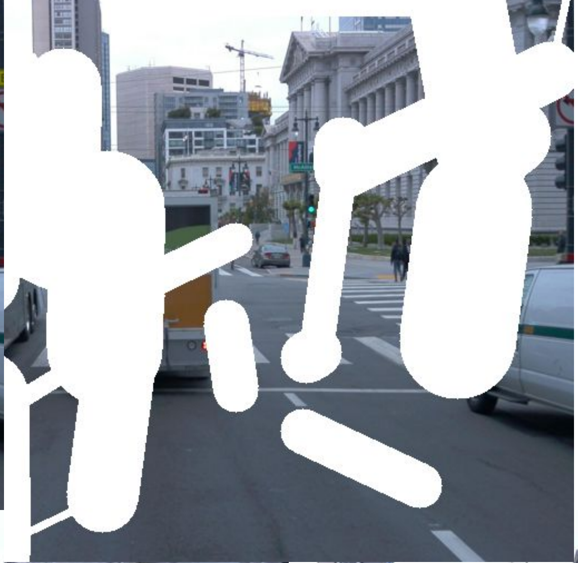
Finetune: Config

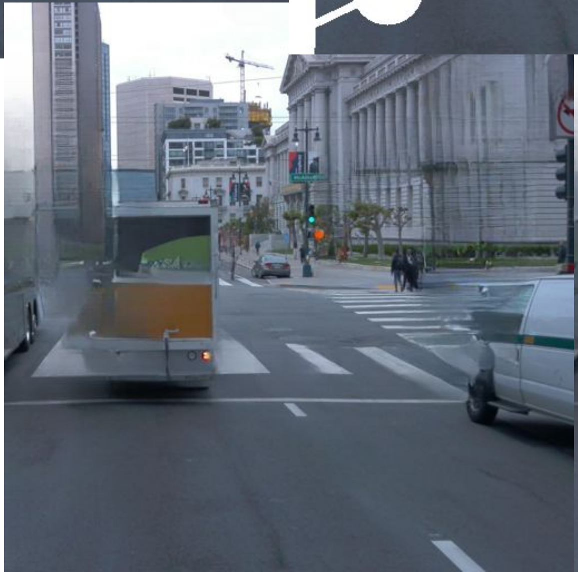
- **train**: 380k images from Waymo
- **val**: 24k images from Waymo
- **epochs**: 18
- **batch size**: 8
- **gpu**: 1x RTX 2080 Ti


val_fid_total_mean
2.22
2.18
2.14
2.1
2.06
2.02
50k
150k
250k
350k
450k
550k

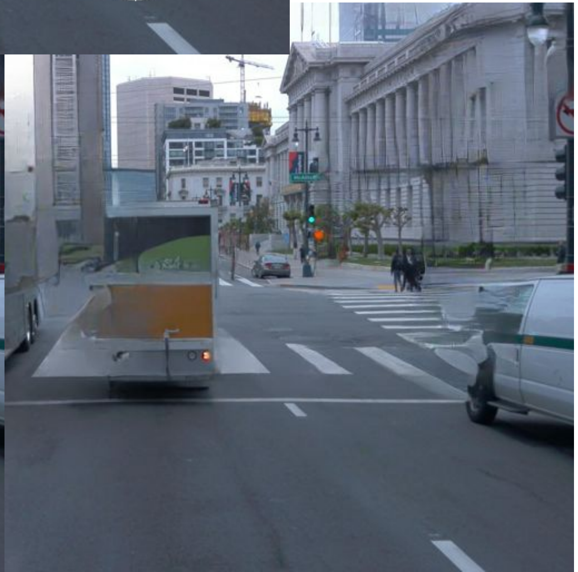

val_lpips_total_mean
0.0952
0.0948
0.0944
0.094
0.0936
50k
150k
250k
350k
450k
550k


val_ssim_total_mean
0.931
0.931
0.93
0.93
50k
150k
250k
350k
450k
550k

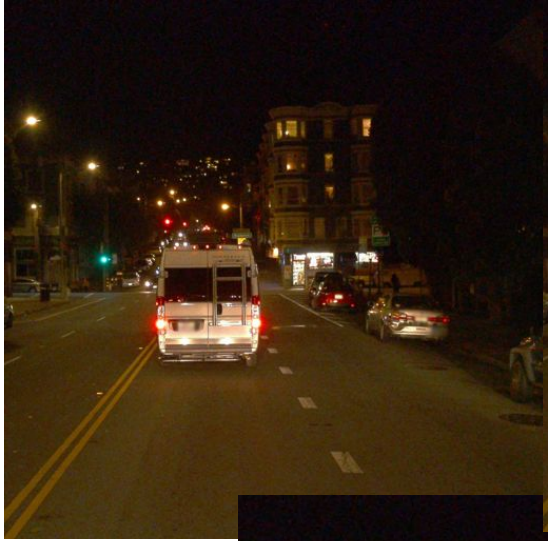




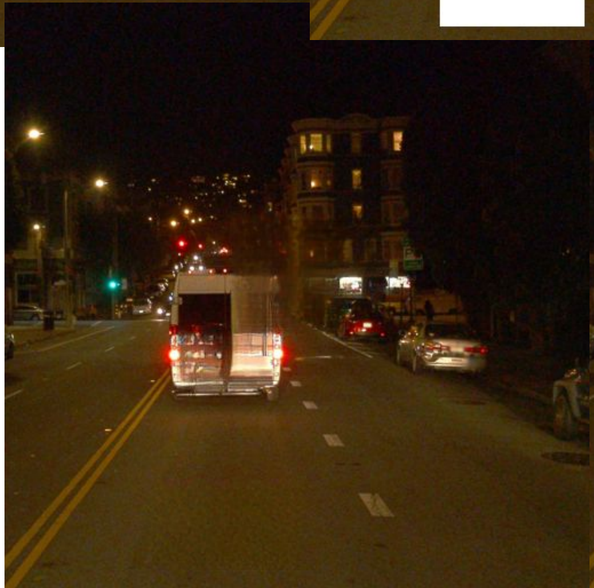
Pretrain



Fine-tune



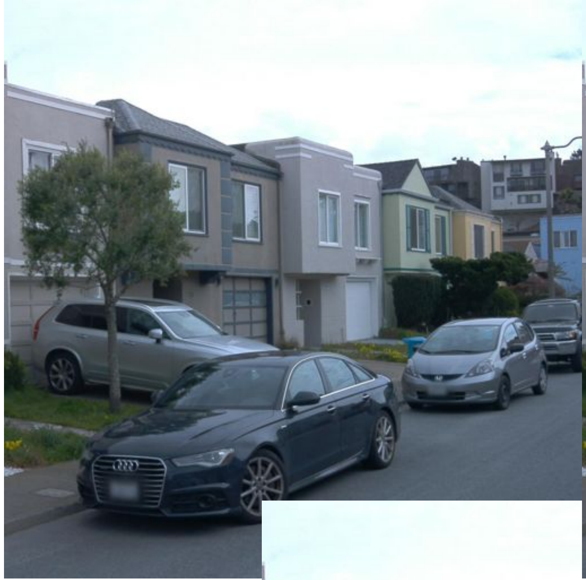
Pretrain



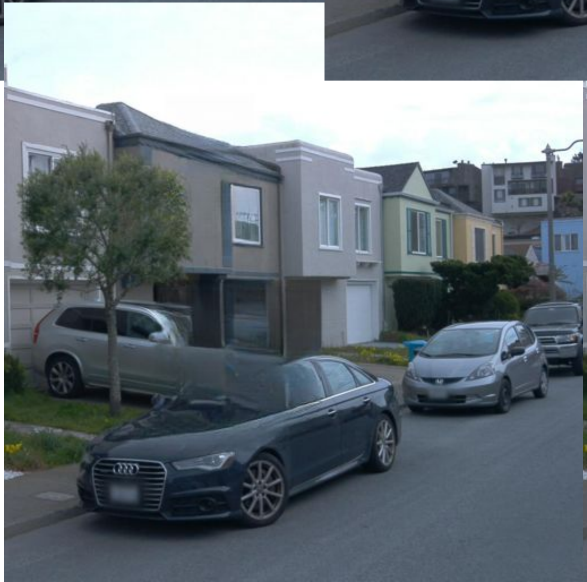
Fine-tune

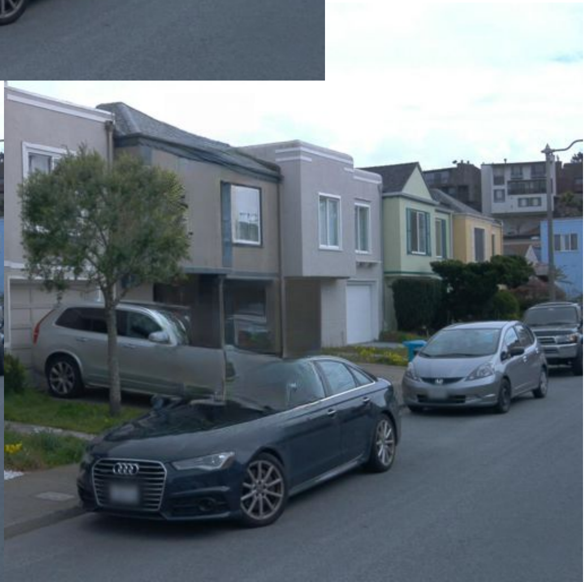
Pretrain

Fine-tune



Pretrain





Fine-tune

Metrics

	FID	LPIPS	SSIM
Pretrain	2.4985	0.0931	0.9304
Fine-tune	2.0990	0.0936	0.9298

Thanks for attention!

Questions?

girafe
ai