Spatially-Adaptive Pixelwise Networks for Fast Image Translation

Skoltech

Dawit Sefiw

Oluwafemi Adejumobi

Mekan Hojayev

Melaku Getahun

Motivation

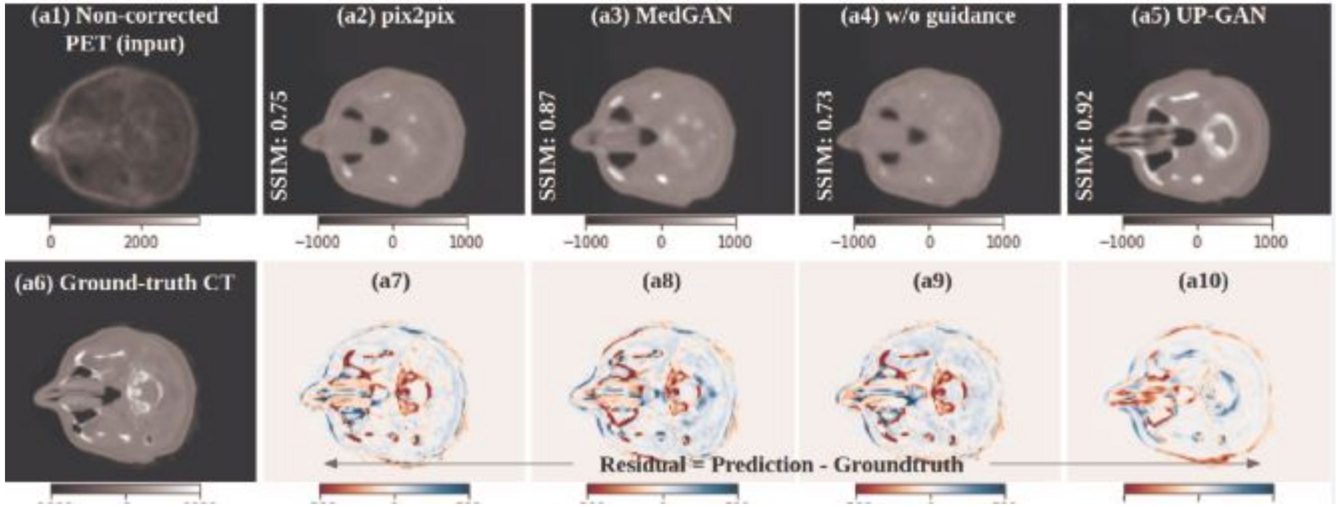
Translating image from one domain to another

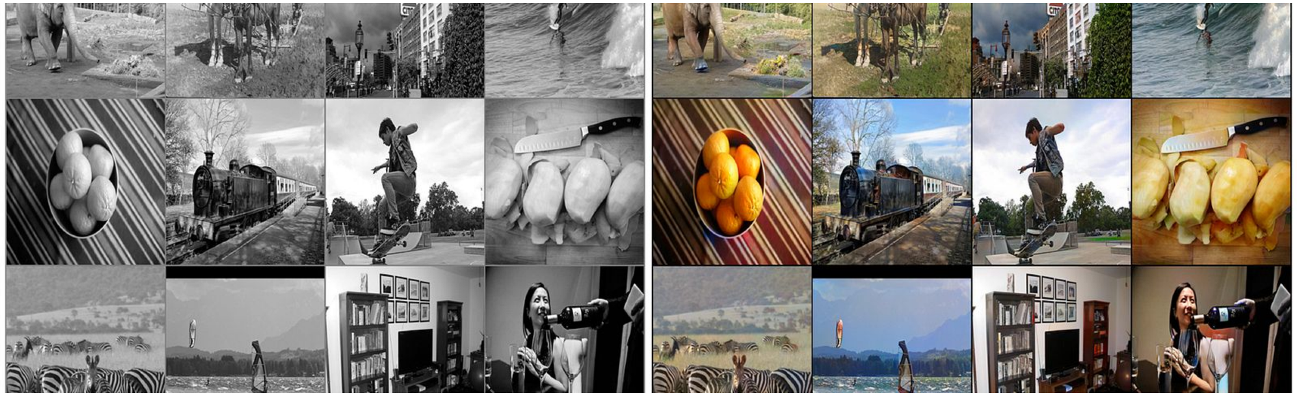
https://www.google.com/url?sa=i&url=https%3A%2F%2Ftheailearner.com%2F2019%2F11%2F09%2Fimplementation-of-image-to-image-translation-using-conditional-gan%2F&psig=AOvVaw37K65-3RMrAZtPoSDJj6OG&ust=1679721346785000&source=images&cd=vfe&ved=0CBEQjhxqFwoTCICDu-Pn8_0CFQAAAAAdAAAAABAE

https://www.google.com/url?sa=i&url=https%3A%2F%2Fanalyticsindiamag.com%2Fhand-on-implementation-of-cyclegan-image-to-image-translation-using-pytorch%2F&psig=AOvVaw3-5uplrKD2Udyw2_ylvZ00&ust=1679721511656000&source=images&cd=vfe&ved=0CBEQjhxqFwoTCIiZobPo8_0CFQAAAAAdAAAAABAD

Motivation

Improving quality





https://www.google.com/url?sa=i&url=https%3A%2F%2Flink.springer.com%2Fchapter%2F10.1007%2F978-3-030-87199-4_58&psig=AOvVaw3p9C_0QXnmUAmJpo4iVwbx&ust=1679721605983000&source=images&cd=vfe&ved=0CBEQjhxqFwoTCPDWiN_o8_0CFQAAAAAdAAAAABAb

https://www.google.com/url?sa=i&url=https%3A%2F%2Ftowardsdatascience.com%2Fcolorizing-black-white-images-with-u-net-and-conditional-gan-a-tutorial-81b2df111cd8&psig=AOvVaw3YkpUqFrRVtmHvX2cVT_WQ&ust=1679721737026000&source=images&cd=vfe&ved=0CBAQjRxqFwoTCLDCup3p8_0CFQAAAAAdAAAAABAE

Problem Statement

Current approaches use conditional Generative Adversarial Networks to learn direct mapping from one domain to the other.

Although reaching substantial visual quality, model size and inference time have also grown significantly.

Computational cost surges when operating on high resolution images under real world settings.

Skoltech

ASAPNet

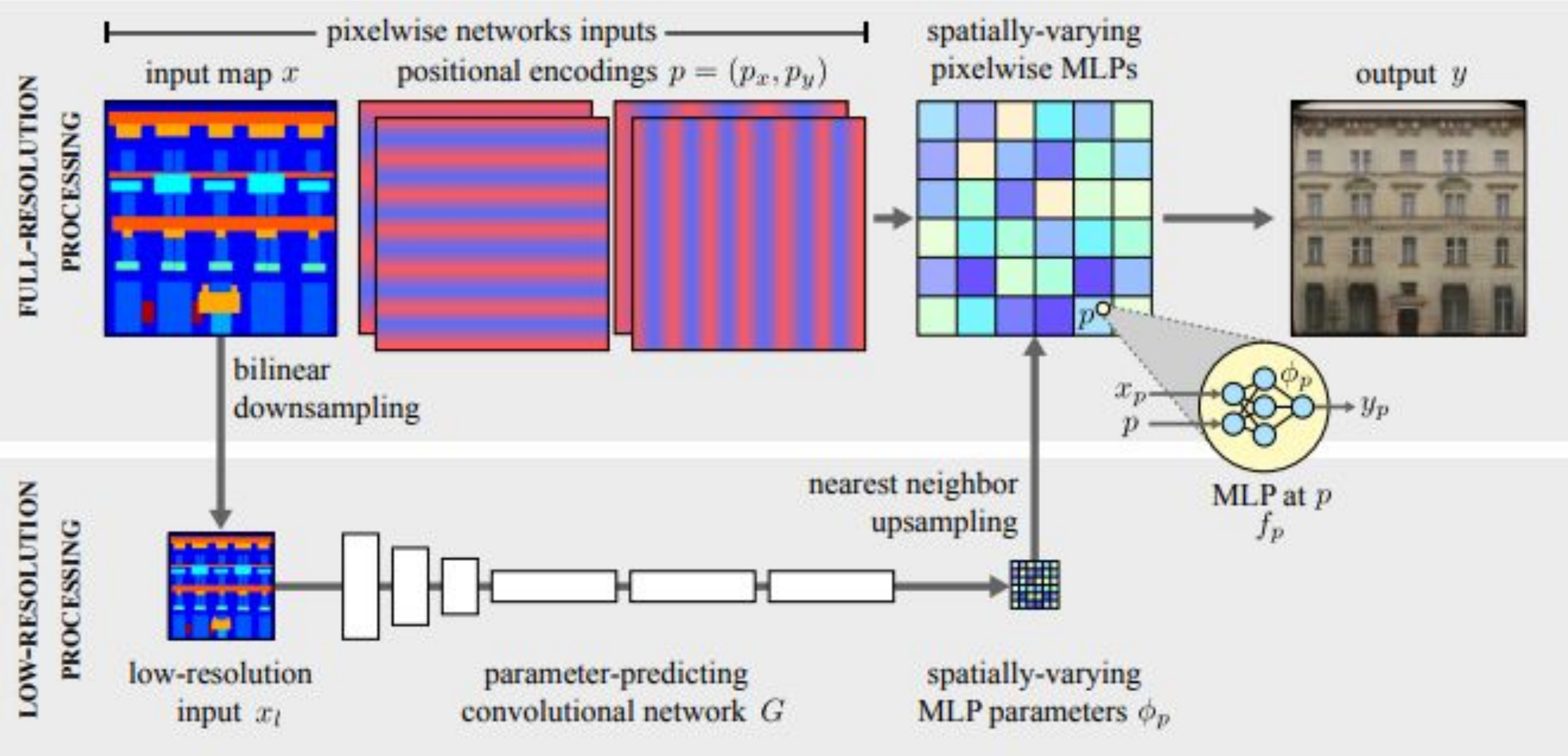
Spatially-Adaptive Pixel-wise Networks for fast image translation

A generator that operates pixel-wise using pixel specific MLP.

Three key features:

1. Each pixel is effectively transformed by a different function
2. Parameters are predicted at low-resolution representation
3. MLP consume a sinusoidal encoding of the pixel’s spatial position

ASAPNet Architecture



Skoltech

Related Work

Conditional GANs

➔ aim to model the conditional distribution of real images given the input semantic label maps via the following minimax game.

$$\mathbb{E}_{(\mathbf{s},\mathbf{x})}[\log D(\mathbf{s},\mathbf{x})] + \mathbb{E}_{\mathbf{s}}[\log(1 - D(\mathbf{s}, G(\mathbf{s}))].$$

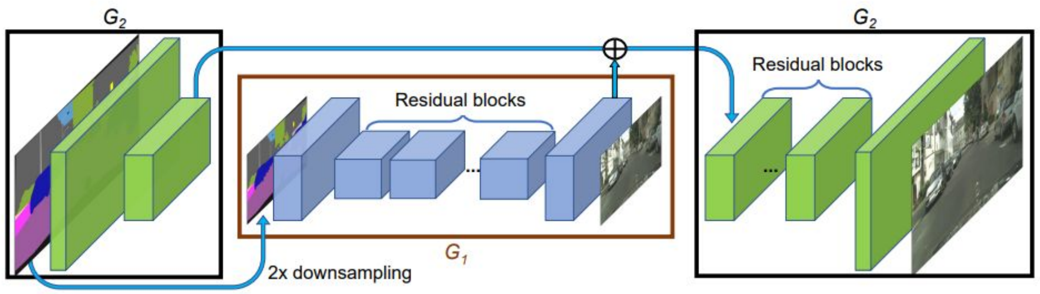
Related Work

Pix2PixHD

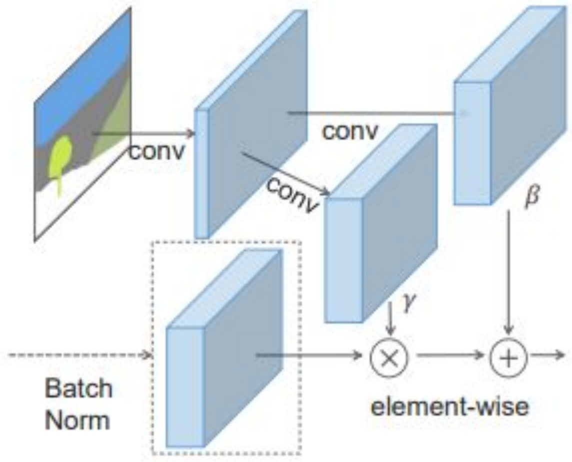
➔ a novel adversarial loss, as well as new multi-scale generator and discriminator architectures is introduced

SPADE (Spatially Adaptive De(normalization))

➔ proposed a conditional normalization method through a spatially-adaptive, learned transformation



https://arxiv.org/pdf/1711.11585v2.pdf



https://arxiv.org/pdf/1903.07291v2.pdf

Goal

- To reproduce some key results of this paper for segmentation problem.
  - ➔ Inference time
  - ➔ Frechet Inception Distance (FID)
  - ➔ Mean Intersection over Union (meanIoU)
- Researching the possibility of the usage of such network for other image-to-image translation task: denoising problem.
  - ➔ Peak Signal to Noise Ratio (PSNR)
  - ➔ Structural Similarity Index Measure (SSIM)

Conducted Experiment

Replication:

Dataset - Cityscape - images of urban scenes and semantic label maps

Train/Validation data size - 3000/500 images of size 256x512

ASAPNet is trained

Baseline Model: Pix2PixHD

Research:

Dataset **-** Berkeley segmentation dataset + noise

Training/Validation data size: 432/68

Baseline Model: DnCNN (Denoising CNN)

https://www.cityscapes-dataset.com/downloads/

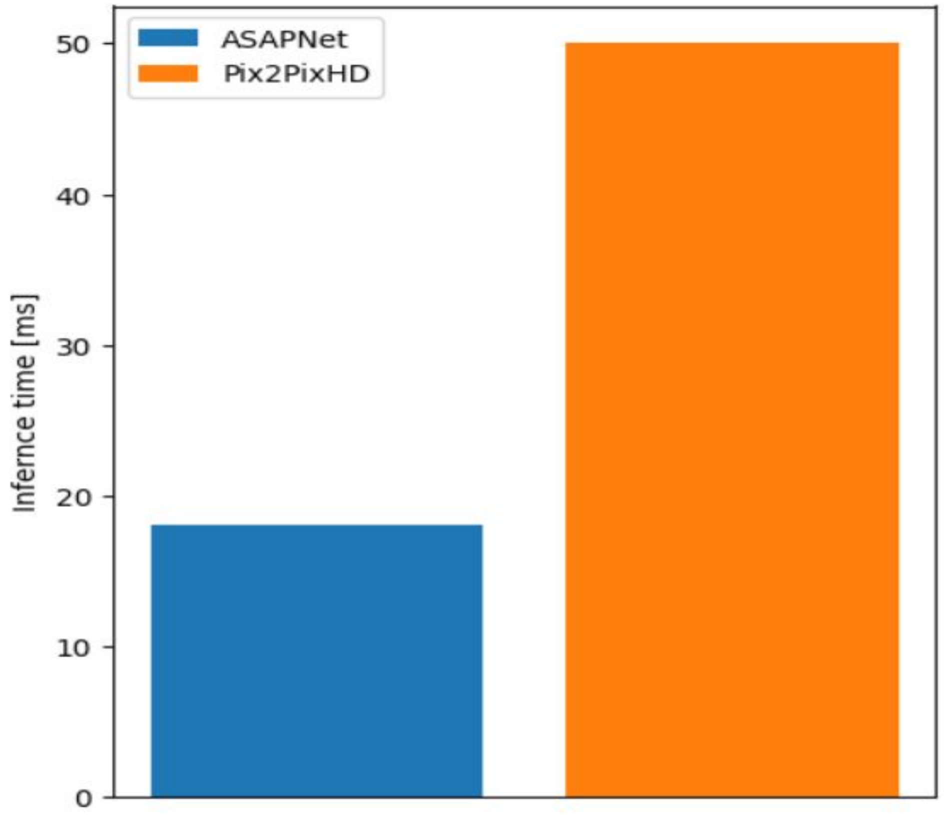
Obtained Result

Synthesized Images

Label ASAPNet Pix2PixHD

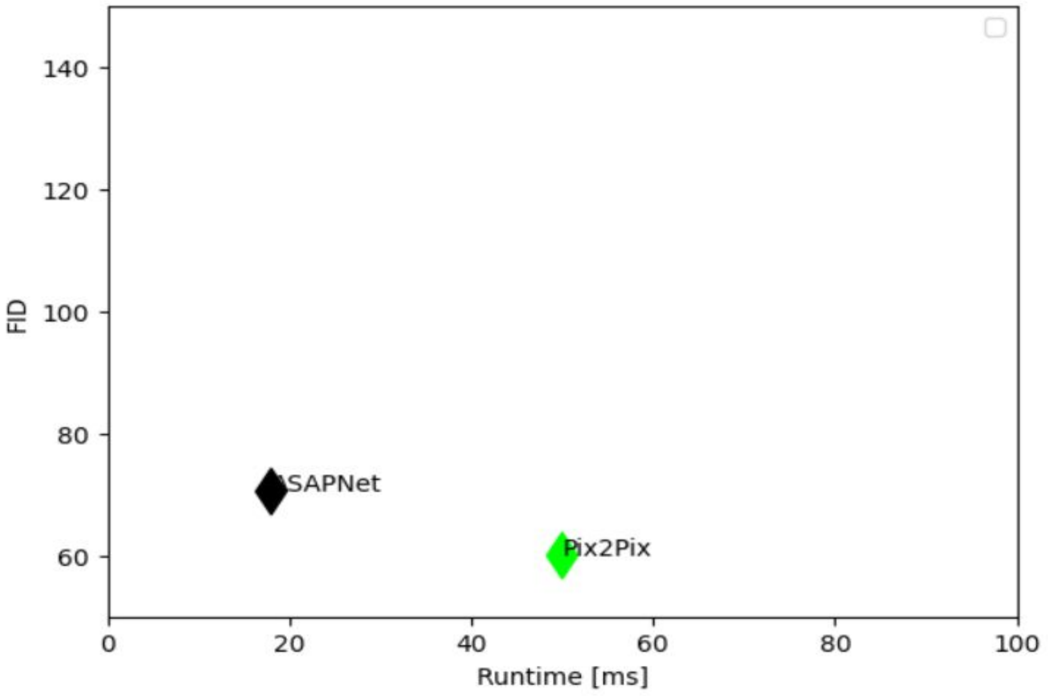
Obtained Result

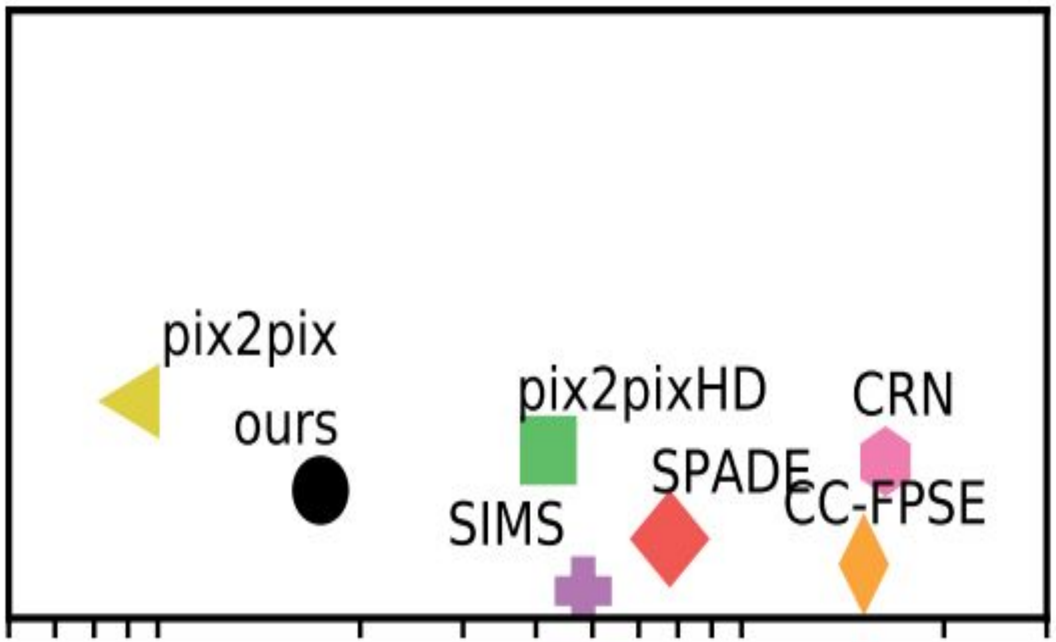
Inference time 2-6x faster



Obtained Result

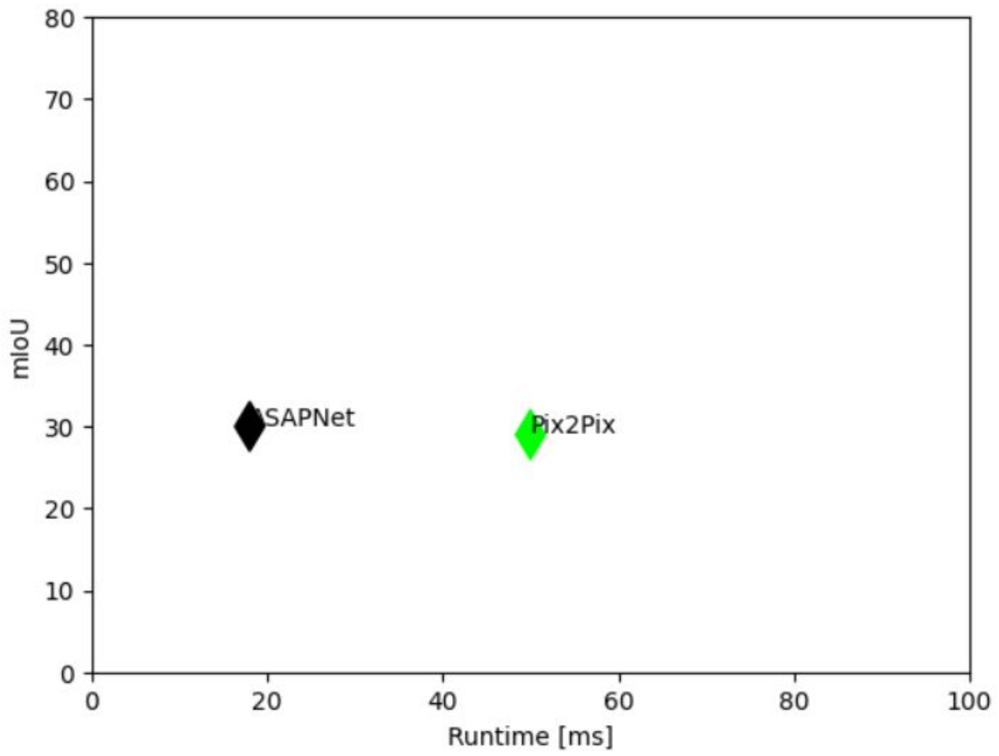
Inference time vs Frechet distance

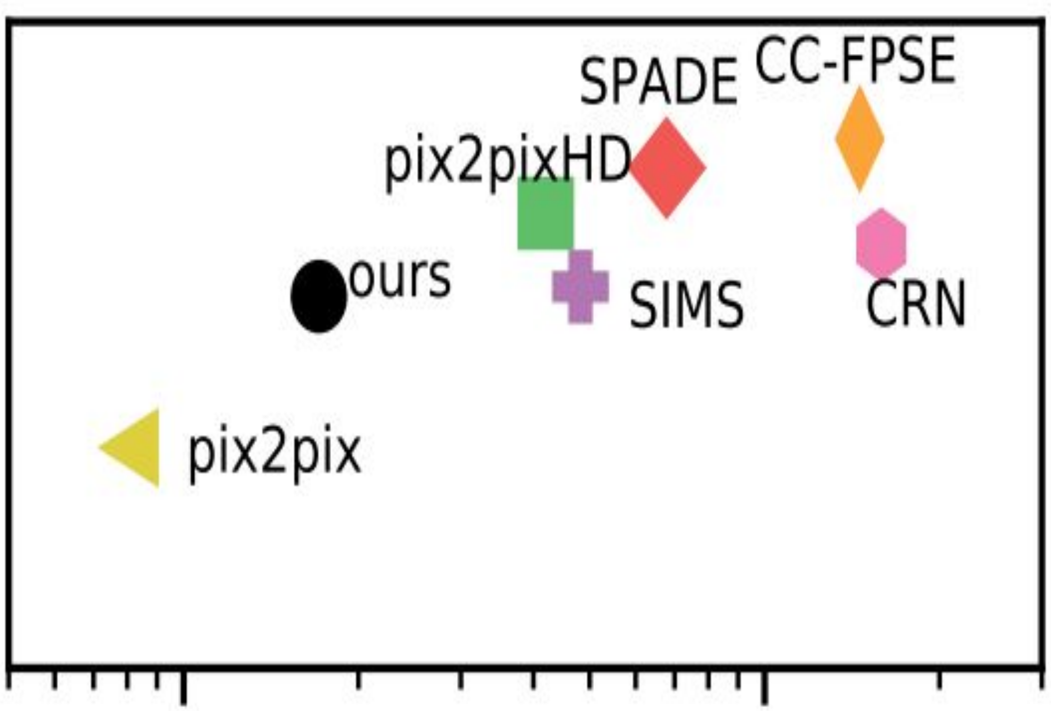




Obtained Result

Inference time vs mean Intersection over Union





Obtained Result

- For training synthetic noisy image is created with $y = x+\sigma.n, n \sim N(0,I)$, with different σ of noise level ranging from 10 to 55.
- And for validation the same type of synthetic image is created with σ of 15, 25, and 50.
- Peak signal to noise ratio (**PSNR**) of 28.55 is achieved.

To Do

We will experiment on as to how the ASAPNet model performs for the denoising task

We will run the same metrics for performance comparison with DnCNN

We would suggest architectural changes for ASAPNet to better suit the denoising task (if needed)

https://www.cityscapes-dataset.com/downloads/

Thank you!