Nvidia’s GauGAN AI turns rough sketches into photorealistic images in real-time

Semantic Image Synthesis with Spatially-Adaptive Normalization

R.A.J. Wacanno

03-04-2019

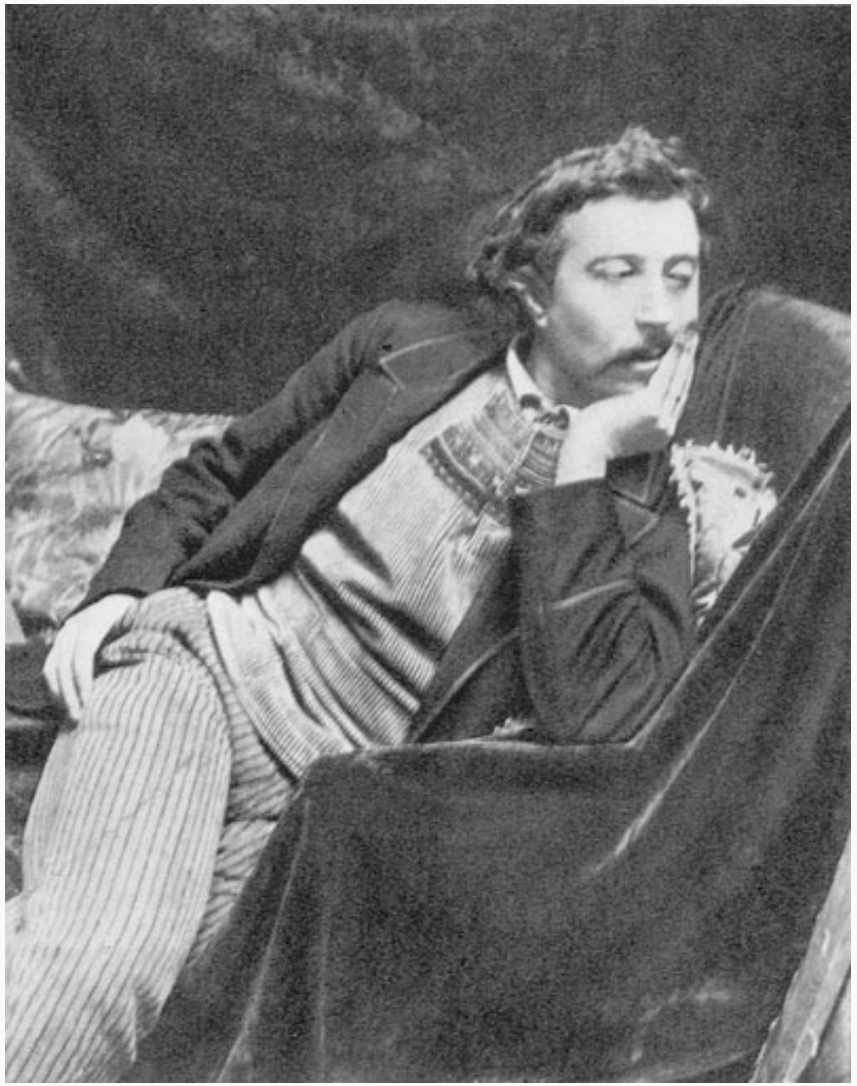
Informatica in het Nieuws
Academische Vaardigheden Informatica 2 - Groep A

Table of contents

Introduction

Introduction

- Published March 19^{th} 2019
- Techspot
- Nvidia, UC Berkely and MIT
- GauGAN
- Generating photo-realistic images



The Article

Nvidia's GauGAN AI turns rough sketches into photorealistic images in real-time

Anyone can be an artist with this MS Paint for the AI era

By **Cal Jeffrey** on March 19, 2019, 12:41 PM

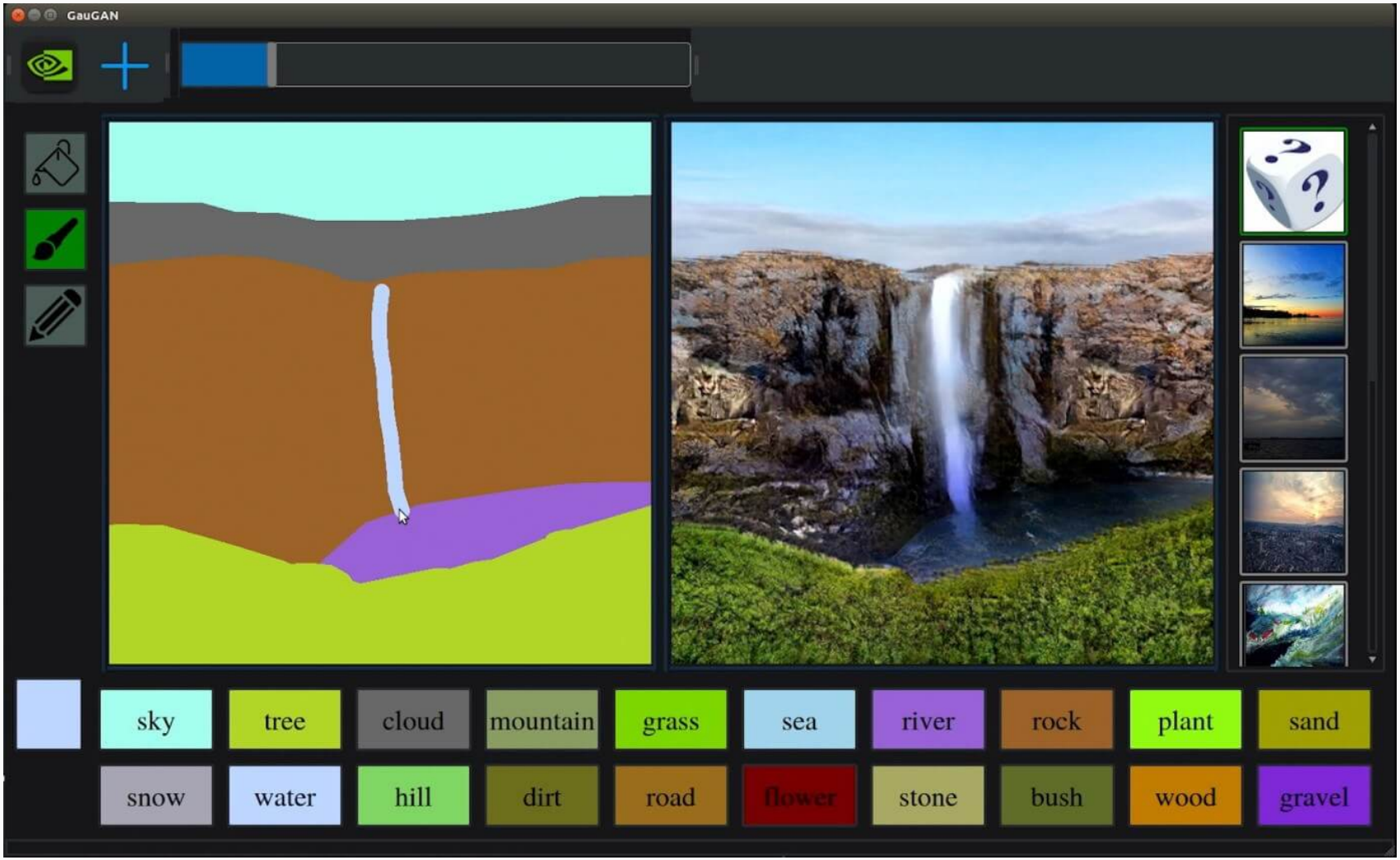
The Article

A novice painter might set brush to canvas aiming to create a stunning sunset landscape — craggy, snow-covered peaks reflected in a glassy lake — only to end up with something that looks more like a multi-colo(u)red inkblot." Nvidia has created a tool that can turn the simplest drawings into photo-realistic images.

The Article

- Introducing GauGAN
- Brief technical details
- Describing usage

The Interface

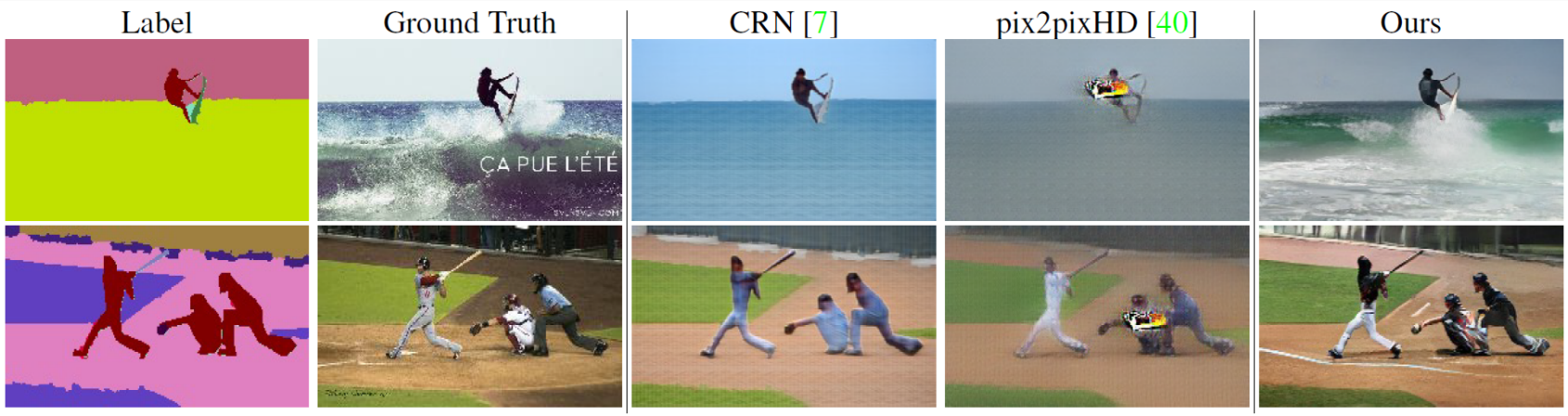


The Paper

The Paper

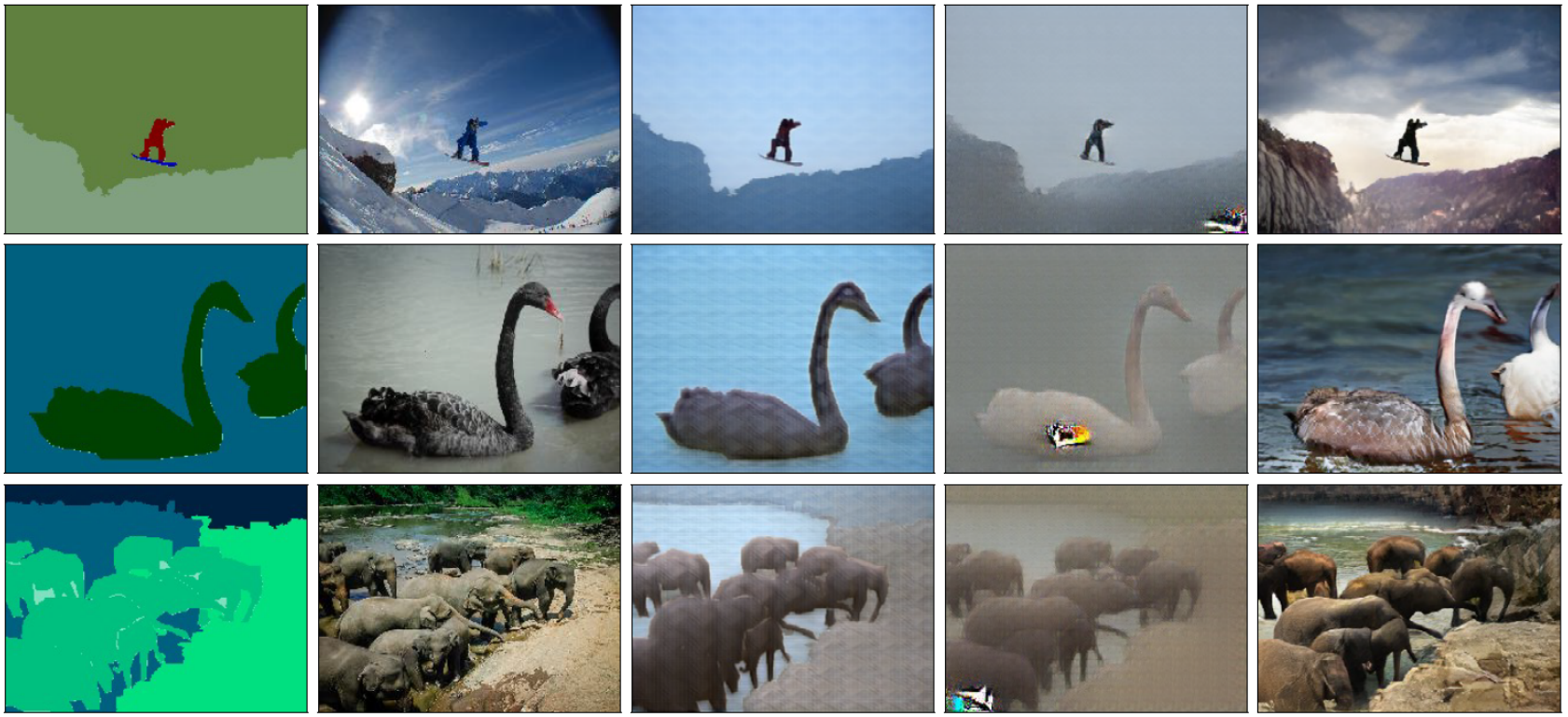
- Introduce problem/solution
- Outline earlier work and used research
    - Earlier methods
- Describing the SPADE algorithm
- Testing and evaluating

Testing the Network

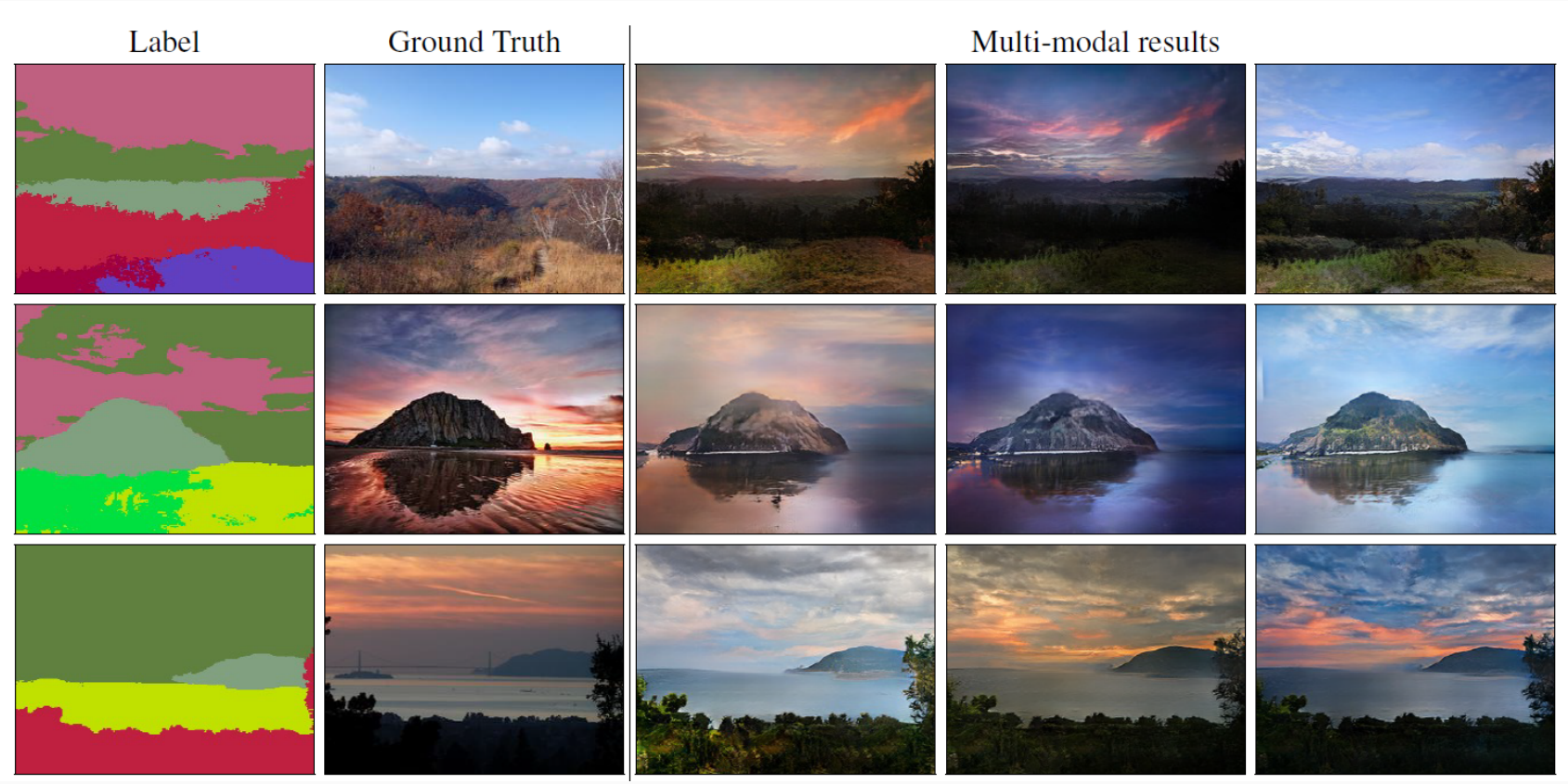


Testing the Network

Testing the Network



Multiple Styles



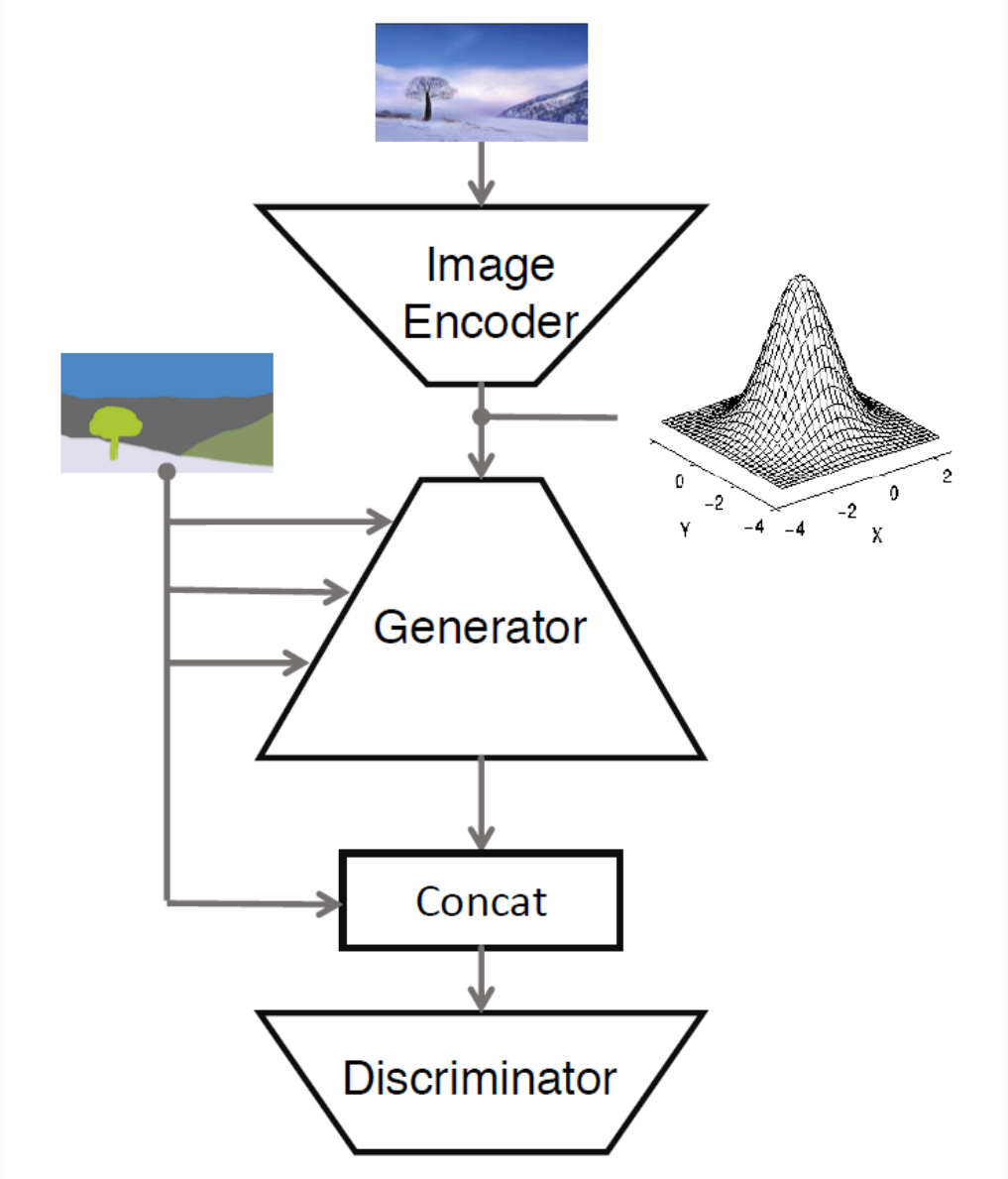
Article vs. Paper

Article vs. Paper

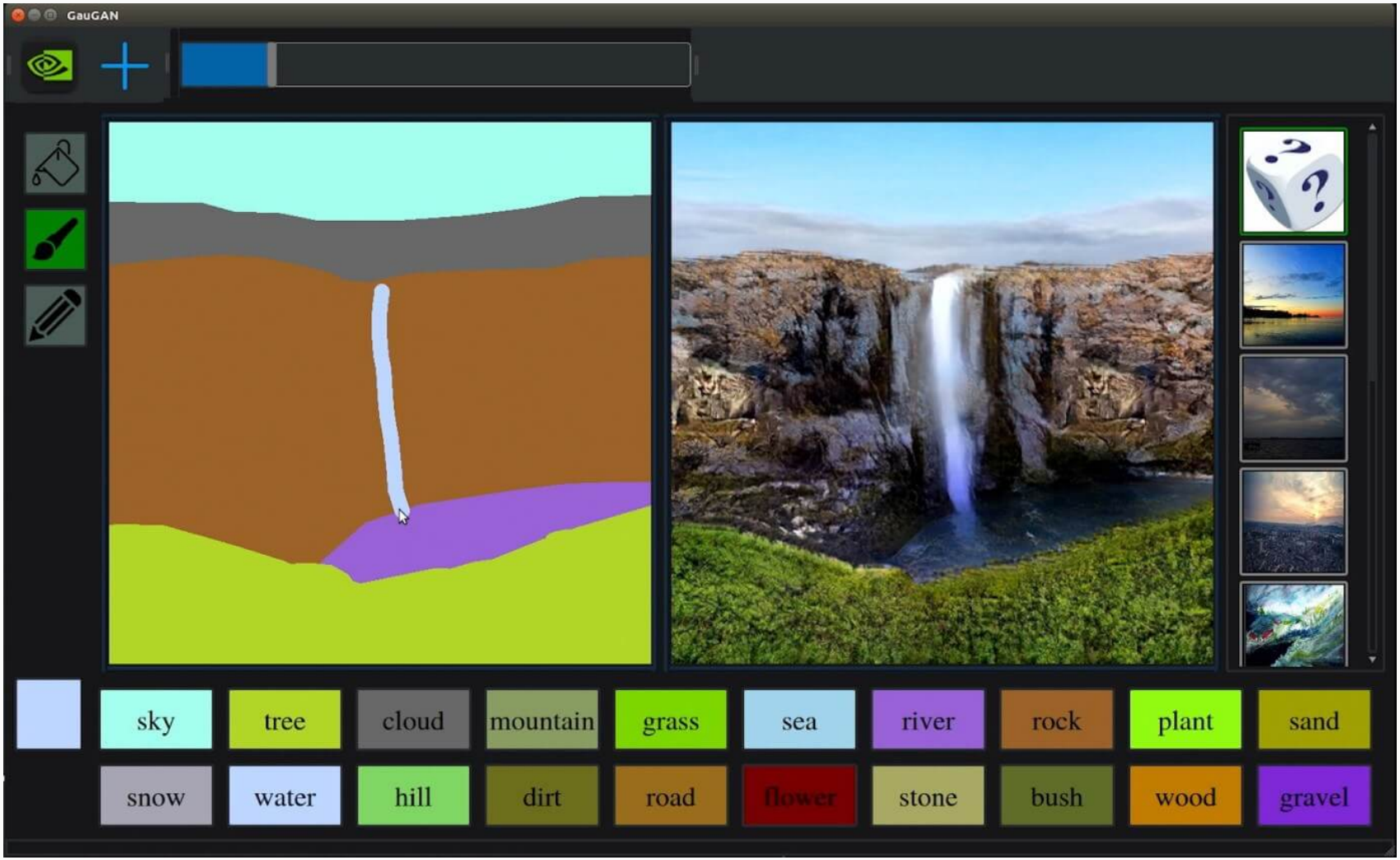
- Article:
  - All important details
  - Skims over uninteresting details
  - No exaggeration
- Paper:
  - More in depth
- Usage vs. implementation

Explanation of SPADE

Generative Adversary Networks



Semantic Mask



Semantic Mask

$$m = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

$$x_{nm} \in \mathbb{L}$$

for example:

$$\mathbb{L} = \{0\ (\text{sky}), 1\ (\text{grass}), \dots\}$$

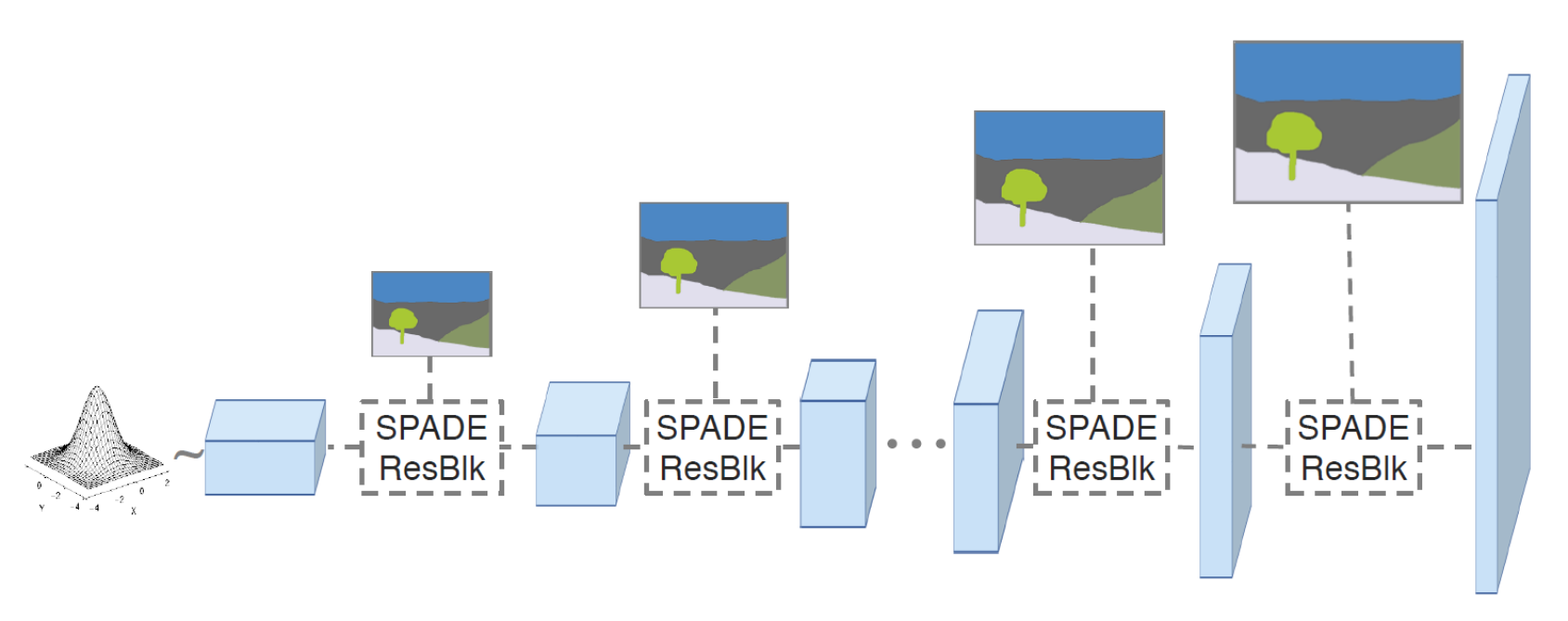
Normalisation

$$\gamma \frac{h_i - \mu}{\sigma} + \beta$$

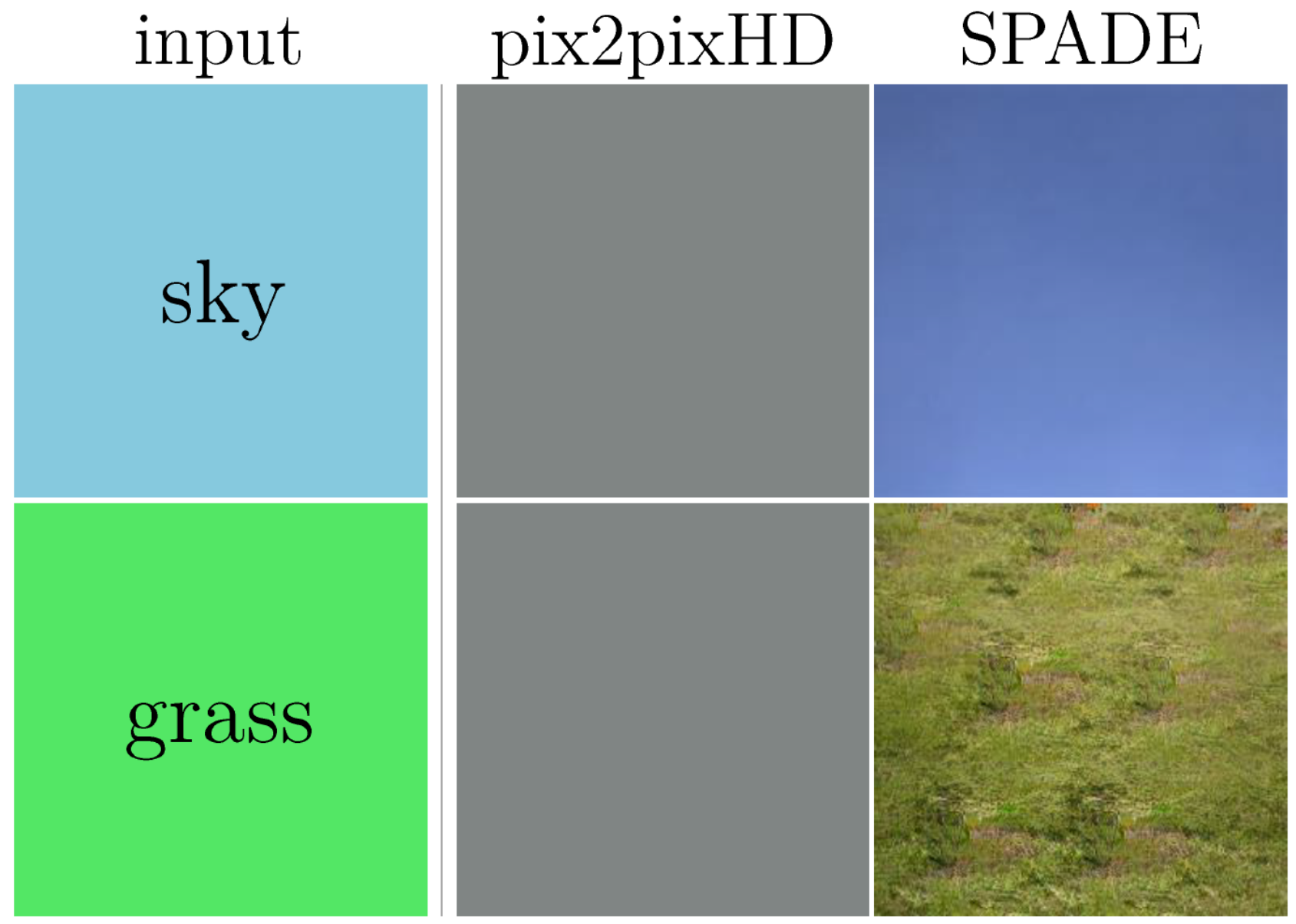
With SPADE:

$$\gamma(\boldsymbol{m}) \frac{h_i - \mu}{\sigma} + \beta(\boldsymbol{m})$$

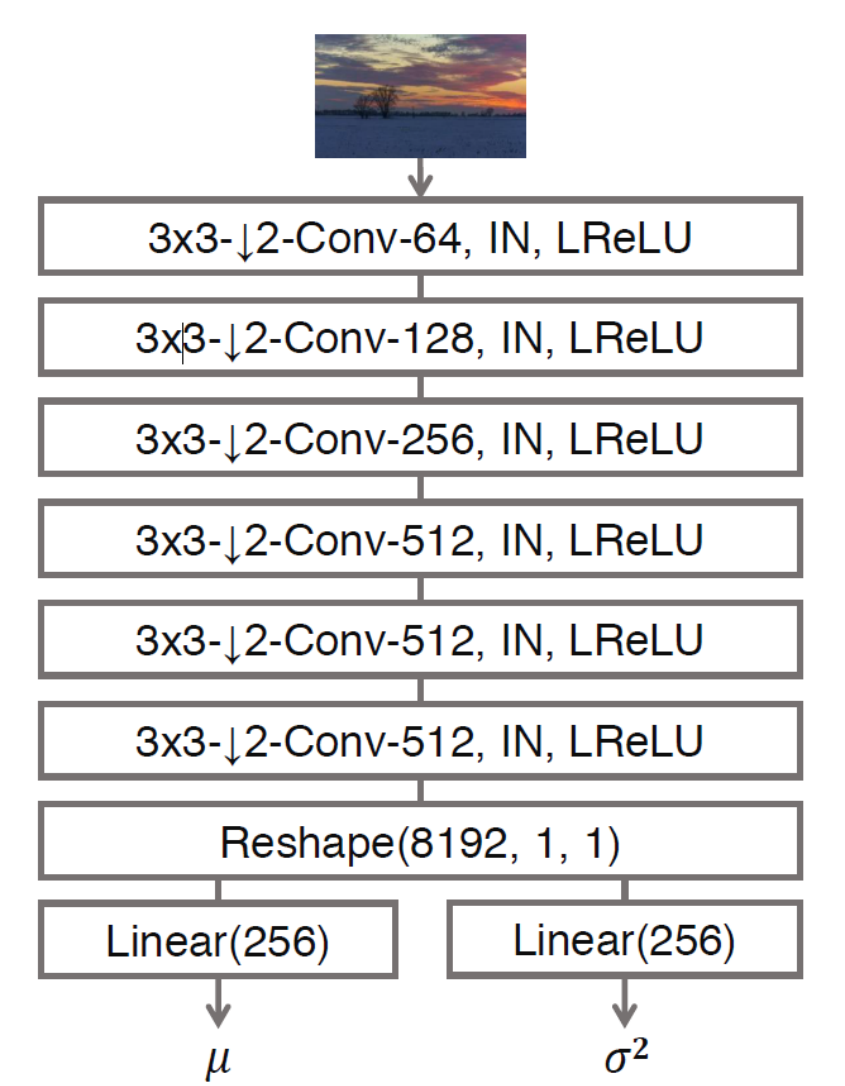
Generator



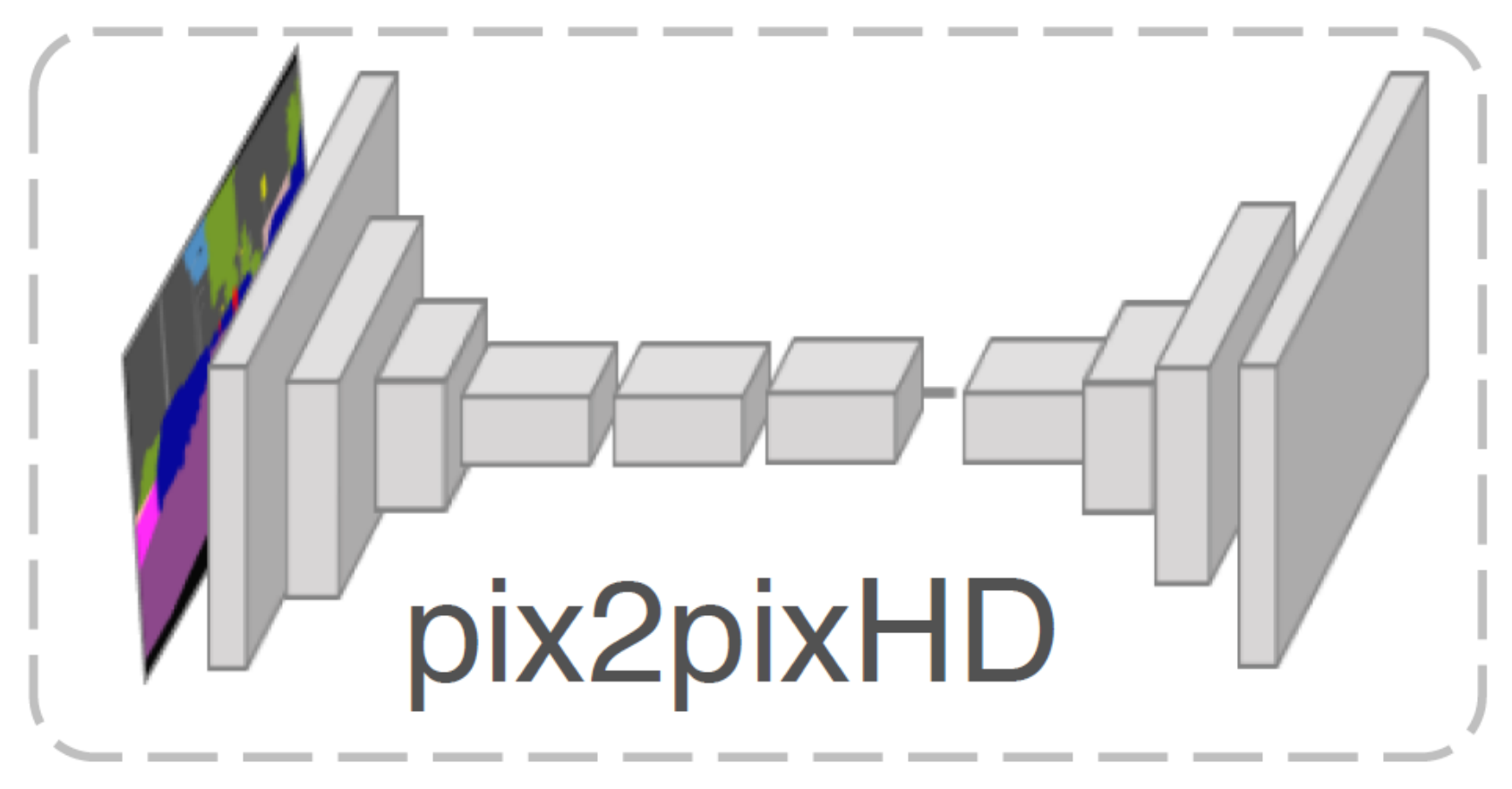
Comparison after Normalisation



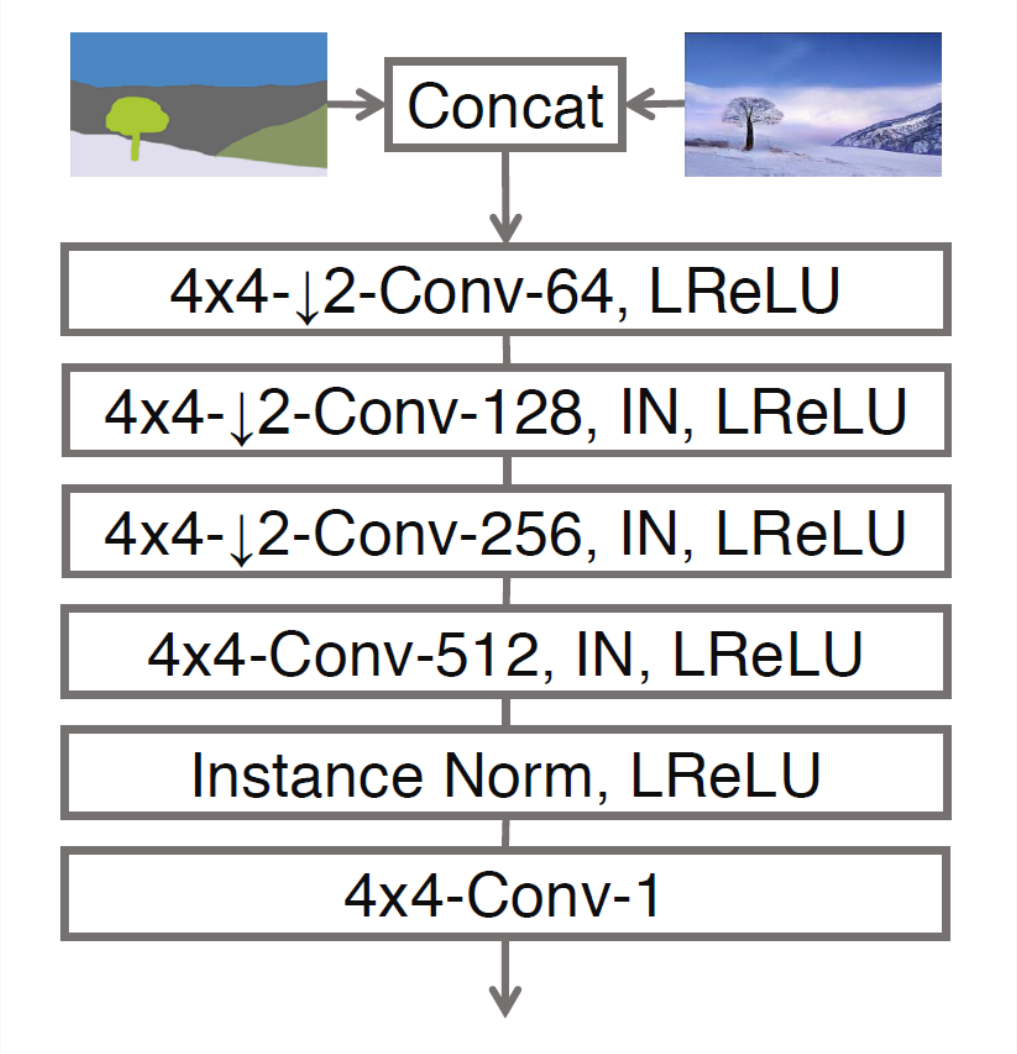
Encoder


3x3-↓2-Conv-64, IN, LReLU
3x3-↓2-Conv-128, IN, LReLU
3x3-↓2-Conv-256, IN, LReLU
3x3-↓2-Conv-512, IN, LReLU
3x3-↓2-Conv-512, IN, LReLU
3x3-↓2-Conv-512, IN, LReLU
Reshape(8192, 1, 1)
Linear(256)
Linear(256)
μ
σ^2

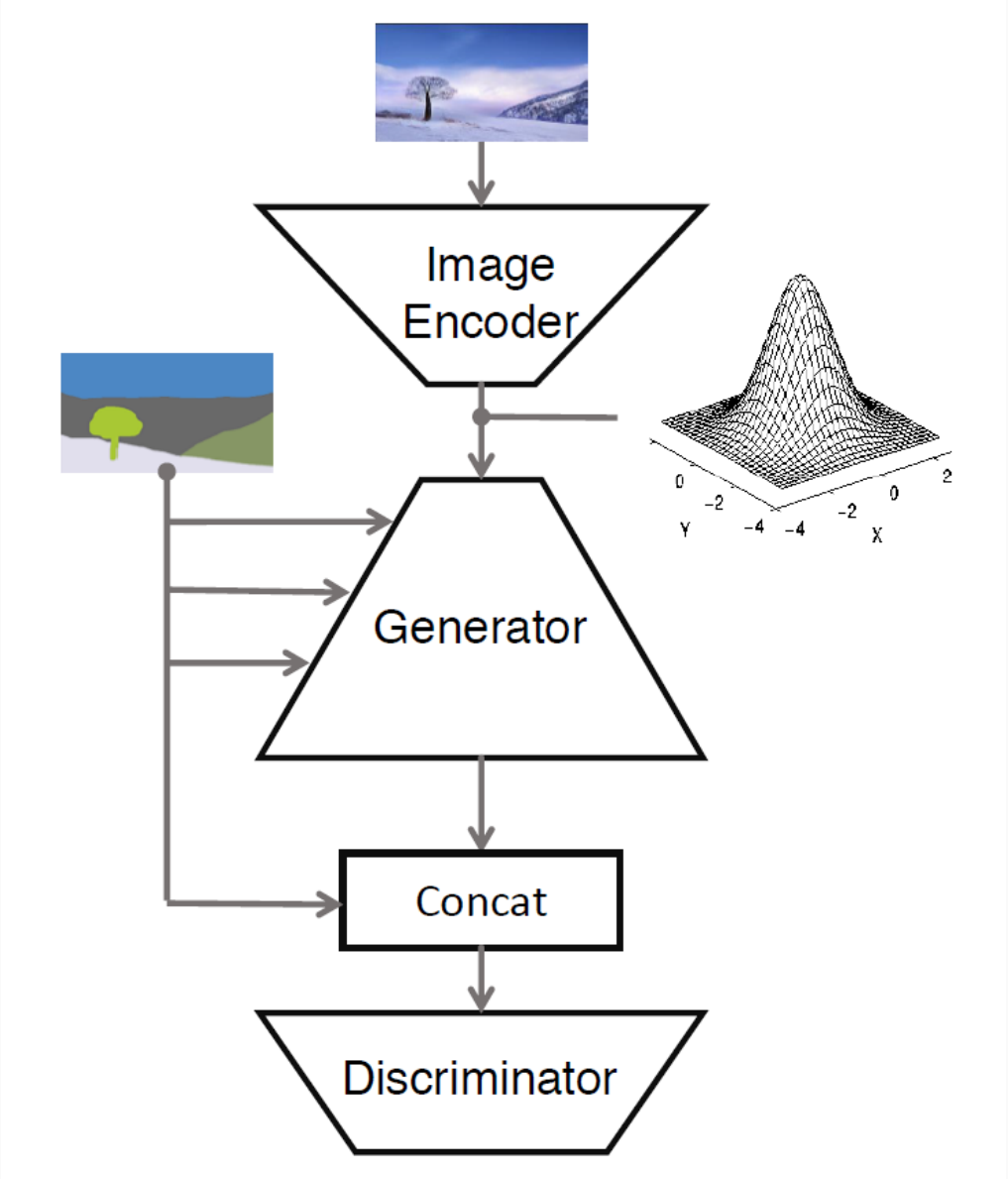
Vs. Pix2pixdh



Discriminator



Full Algorithm



Impact of these Developments

- Use cases:
  - Architecture
  - Urban planning
  - Landscape design
  - Game development
  - Data sets for other NN applications
- Further research *spatial adaptive* GAN's

Discussion

Discussion

Someone using tools like these is not a real artist.

Questions?