

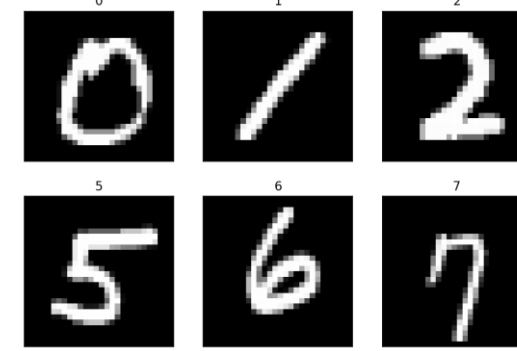
Proof of concept on MNIST

Idea: In order to prove that our idea is valid we began our investigation on an easy to train dataset like MNIST

Goal: Train a VAE aiming to create a network that can increase the resolution of the input image

The dataset

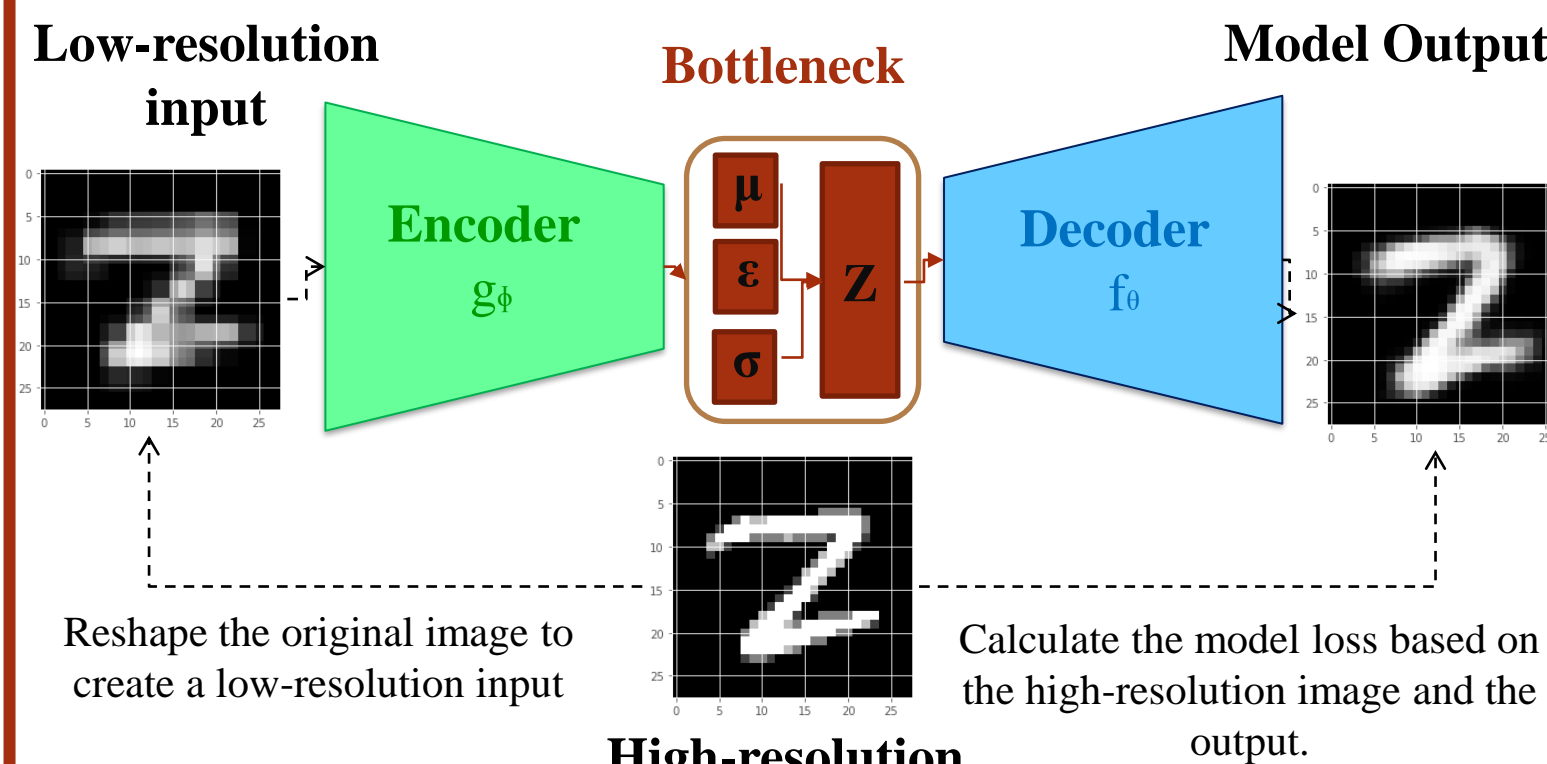
- Handwritten digits from 0-9
- Consists of 1x28x28 black and white image



Linear VAE

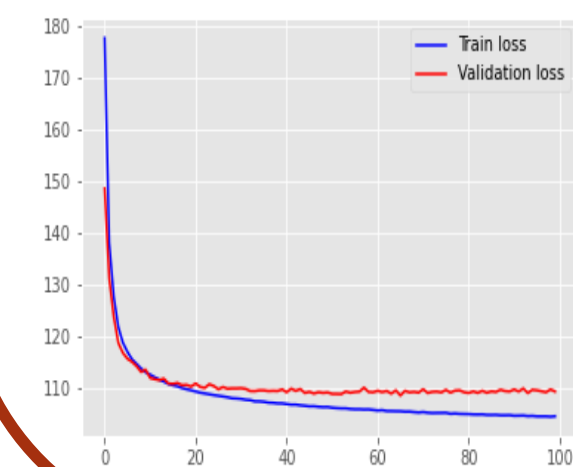
- Initial VAE network trained with linear layers
- Encoder with **4 hidden linear layers**
- Decoder with **4 hidden linear layers**
- Trained for **150 epochs** with **0.001 lr** and **64 batch**

VAE architecture



Results

Images/ metrics	PSNR	MSE	SSIM
Input-target	64.8	0.022	0.74
Output-target	65.9	0.02	0.79

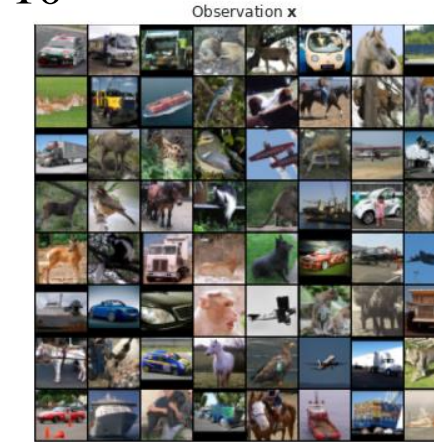


STL-10 Dataset

Idea: After proving that our initial idea is working on the MNIST dataset we tried to follow the same test in a more complex dataset like STL-10

The dataset

- Consists of 100k 3x96x96 images from ImageNet splitting into 10 classes: airplane, bird, car, cat, deer, dog, horse, monkey, ship, truck



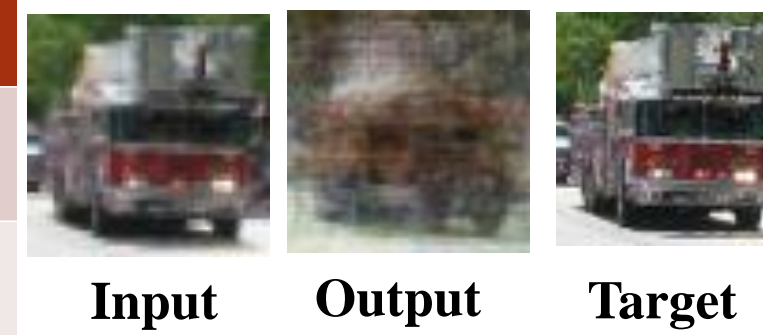
Linear VAE

- Trained a VAE model with linear layers on STL
- Rescale the images to 3x64x64 due to lack of computational power.
- Encoder with **5 hidden linear layers**
- Decoder with **5 hidden linear layers**
- Trained for **150 epochs** with **0.001 lr** and **64 batch**

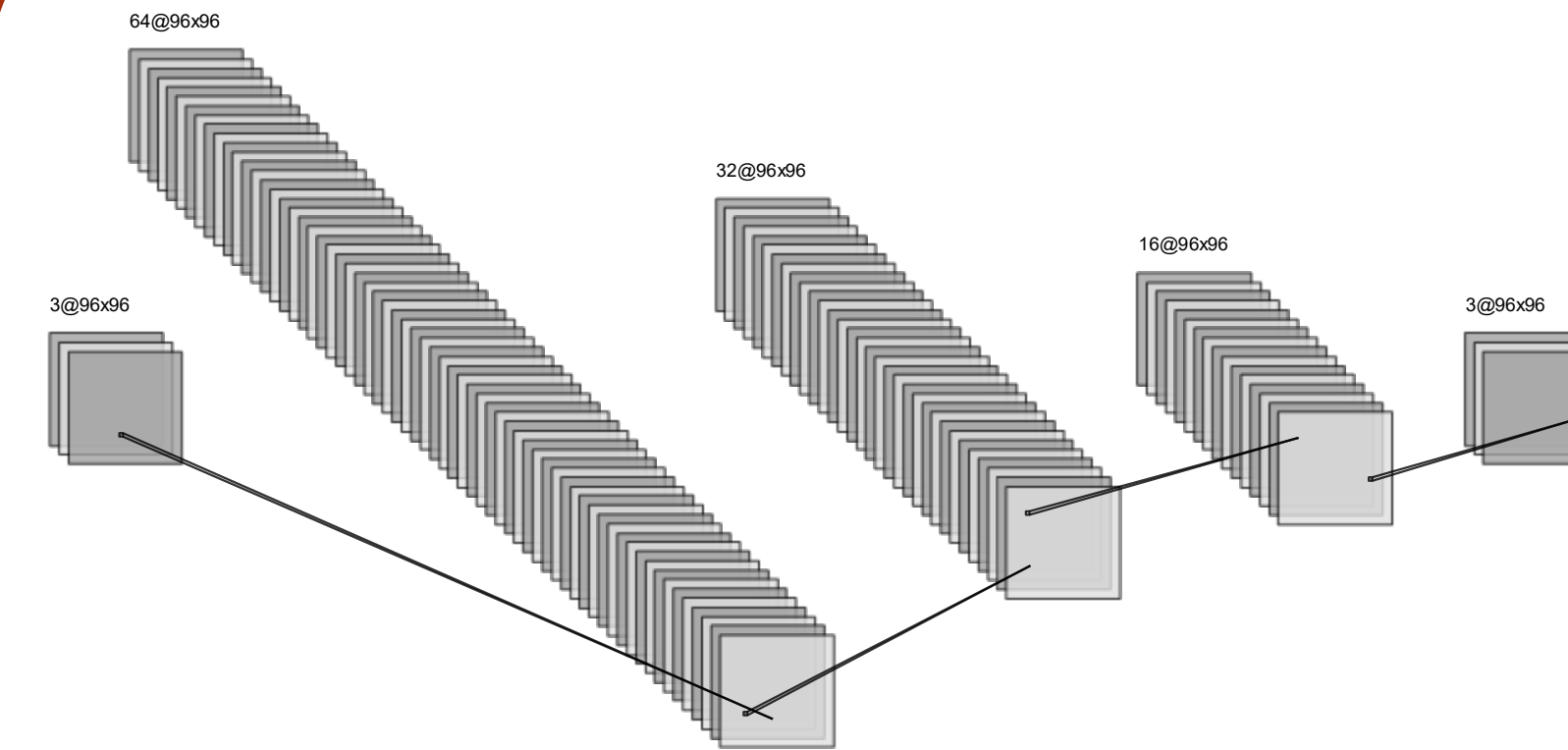
The architecture of the VAE is the same as the one we use for the MNIST dataset.

Results

Images/ metrics	PSNR	MSE	SSIM
Input-target	72.03	0.004	0.90
Output-target	63.07	0.036	0.36



Convolution Network



Architecture

Idea: By keeping the shape of the image constant and expanding the channels, the model would be able to accommodate larger detail in the structures of the image

Results

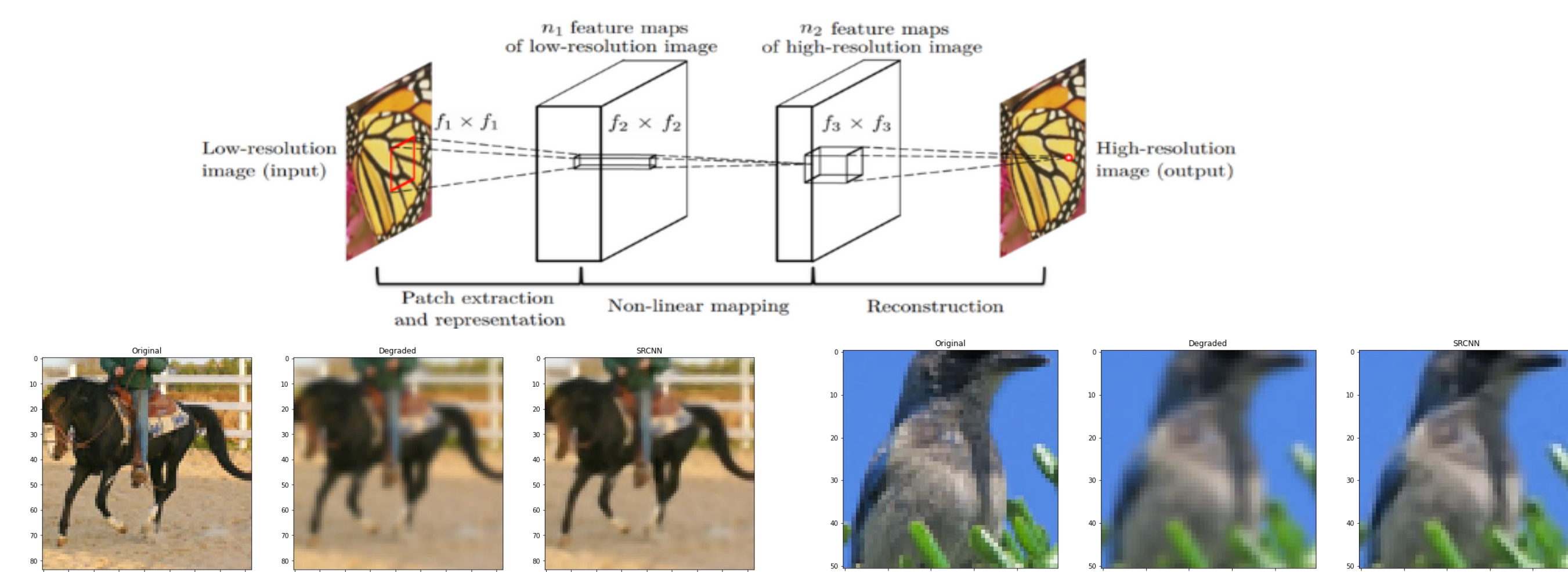


Images/ metrics	PSNR	MSE	SSIM
Input-target	66.47	0.02	0.78
Output-target	69.23	0.008	0.82

Pre-trained SR-Convolution Network

Idea: SRCNN is a type of convolution network that tries to convert a low-resolution input image to a high-resolution output. The network consists of 3 main layers: the patch extraction layer, the Non-linear mapping layer and the reconstruction layer. The network is targeting only the Y channel of each image.

SRCNN architecture



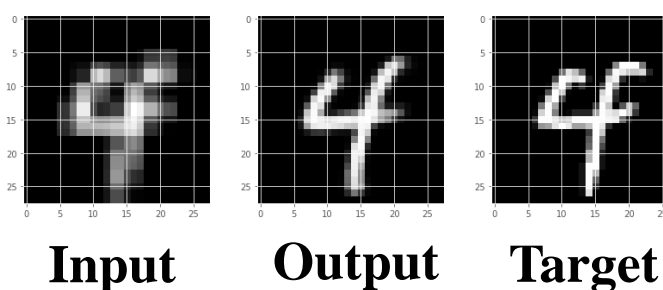
Images/ metrics	PSNR	MSE	SSIM
Input-target	24.90	239.30	0.82
Output-target	26.94	157.47	0.89

Conv VAE

- VAE with convolutional layers
- Encoder with **5 hidden convolutional layers**
- Bottleneck with **3 linear layers (894 features)**
- Decoder with **5 hidden Trans-convolutional layers**
- Trained for **150 epochs** with **0.001 lr** and **64 batch**

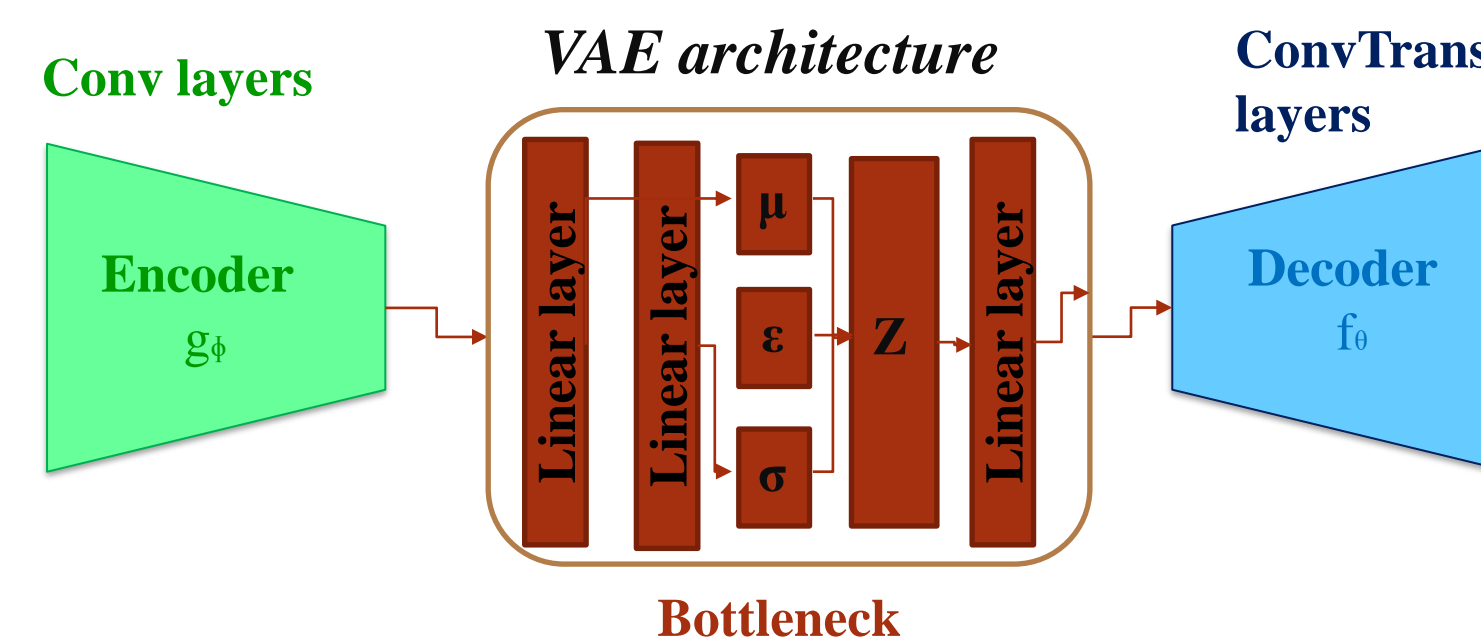
Results

Images/ metrics	PSNR	MSE	SSIM
Input-target	64.8	0.022	0.74
Output-target	68.8	0.009	0.90



Conv VAE

- VAE with convolutional layers on STL-10
- Encoder with **5 hidden convolutional layers**
- Bottleneck with **3 linear layers (894 features)**
- Decoder with **5 hidden Trans-convolutional layers**
- Trained for **150 epochs** with **0.001 lr** and **64 batch**



Bottleneck

Results

Images/ metrics	PSNR	MSE	SSIM
Input-target	72.03	0.004	0.90
Output-target	66.2	0.01	0.61

