THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING

SIPL Signal and Image Processing Lab

TECHNION Israel Institute of Technology

Work in Progress

SinGAN for Temporal Super-Resolution

Tomer Arama, Itay Shemer Supervised by Tamar Rott Shaham

Introduction

- SinGAN is an architecture that can generate new samples based on a single image
- Among SinGAN's applications is spatial super resolution
- Recent work has shown that single sample temporal super resolution is possible

Samples generated from a single image by SinGAN

Goals

Use a single SinGAN scale to perform temporal super resolution on a single video

Challenges

- GPU memory limit
- Quality and reconstruction metrics

Related Work – ZSSR,TSR

ZSSR – Zero Shot Super Resolution

- Paper by Shocher et.al.
- CNN based architecture that performs spatial super resolution from a single image
- Using data augmentation, ZSSR creates Low-res / High-res pairs of images to train on
- Statistical assumption – patch statistics of low-resolution image is similar to that of the high-resolution image for spatial patches

Low resolution image (left) and patches of it's high-resolution version,

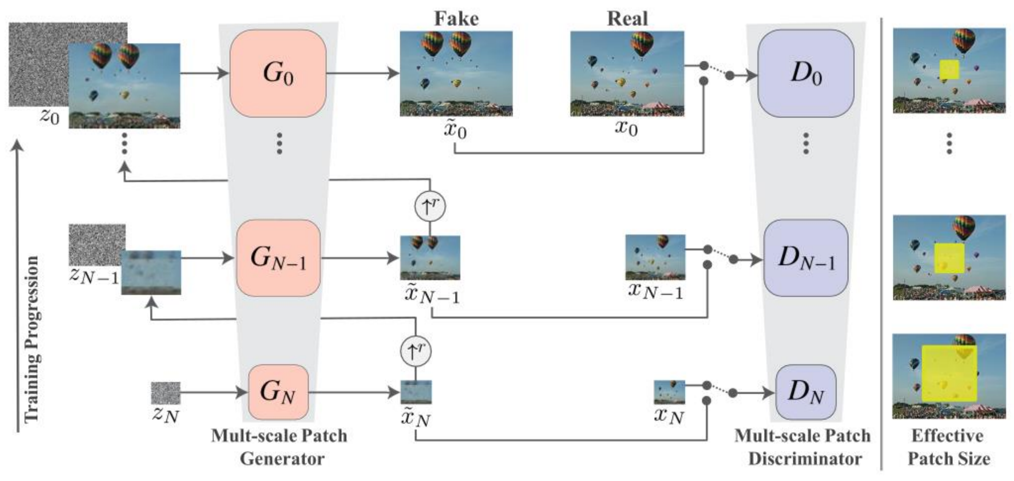
Comparison with EDSR

TSR – Temporal Super Resolution

- Paper by Pollak Zuckerman et.al.
- Adaptation of ZSSR's principal for video super resolution
- CNN based architecture
- Showed that patch statistics of low-resolution image is similar to that of the high-resolution image for temporal - spatial patches

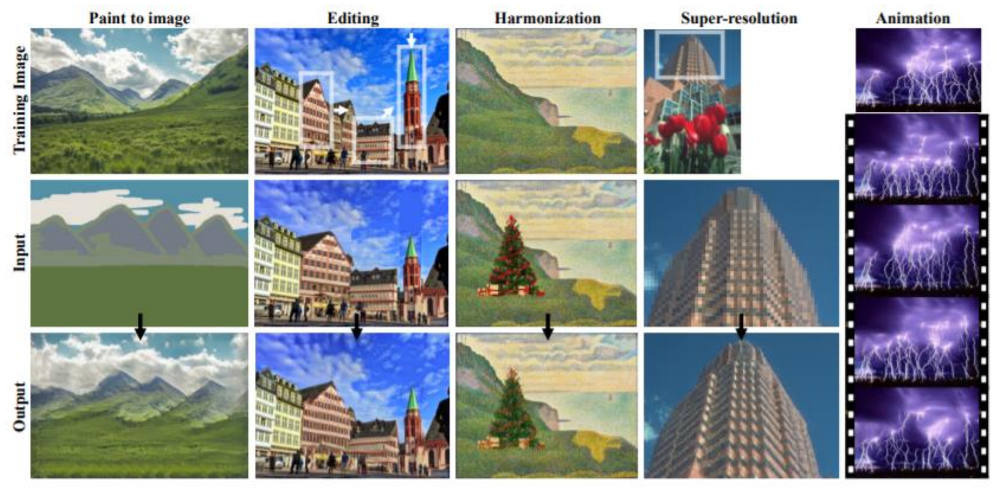
SinGAN

- SinGAN is a generative model that trains on a single image
- The architecture comprises of multiple GANs in different scales



The Mult-scale Patch Generator Architecture

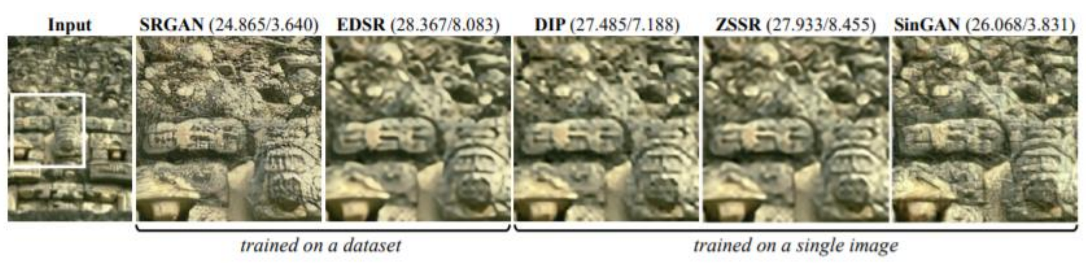
- SinGAN's different applications include
  - Generation of new samples in the image's domain
  - Domain Adaptation
  - Image Editing
  - Image Harmonization
  - Spatial Super Resolution
  - Image To Animation



SinGAN's Different Applications

Spatial SR In SinGAN

- In order to perform spatial super resolution in SinGAN, a single generator scale is trained
- Data augmentation creates low-resolution patches of the image, and the generator learns to map the low-resolution to the input image
- At inference, the input image is treated as the low-resolution image as it passes through the generator and is super resolved
- The image passes through the trained scale to reach its desired size
- SinGAN showed SOTA super resolution results in image naturalness



SinGAN's Spatial Super Resolution (right) On An Input Image (left) Compared To Other SR Architectures

Architecture Adaptation

In order to support temporal super resolution we

- Expanded SinGAN's generator and discriminator to support 3D patches
- Tuned the model's new architecture and hyper parameters
- Implemented a video resize function
- Implemented video quality metrics

A Frame From Side-By-Side Comparison To TSR's Temporal Super Resolution

Quality Metrics

Generative adversarial models perform well in naturalness metrics, in our case we wanted to test naturalness as well as video reconstruction and smoothness.

Absolute Error Metrics:

- PSNR

Perceptual Metrics:

- SSIM, NIQE, LPIPS

Temporal Metrics (Compared To GT):

- Temporal LPIPS (tLP) – difference in LPIPS between following frames
- Temporal Optical Flow (tOF) Distance – distance of optical flow score of following frames between 2 videos

Data Set

- The architecture was developed and tested on the WAIC TSR Dataset

WAIC TSR Dataset Example Frames

Intermediate Results

- The architecture outperforms TSR on naturalness metrics (NIQE)
- We suspect that increasing the networks temporal receptive field will improve results

Winter 2021