Learning to Recover Folds and Wrinkles in Images and Surfaces

Introduction

The goal of the project is to apply deep learning models to recover high frequency details in images and surfaces, in particular, folds and wrinkles in clothes and other deformable surfaces. Motivated by recent successes in generic natural image super-resolution, we formulate this as an upsampling problem and leverage existing architectures to tackle these tasks.

There are three main components to the project:

- **Replicate SRResNet and SRGAN [1], and obtain pre-trained models on subset of ImageNet.**
- **Apply transfer learning to upsample clothes and accessories in the fashion dataset Chictopia10K [2].**
- **Extend model to upsample deformable surface patches in a synthetic ocean water dataset.**

Replicate and Pre-train Networks

We follow the reference paper [1] and use the same generator architecture as shown in Figure 1 below.

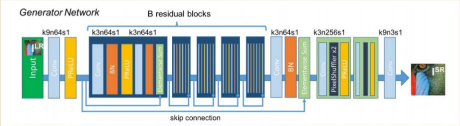

Figure 1: SRResNet architecture. We use B=16 throughout.

We trained the SRResNet models with both MSE and VGG feature content loss. We are able to match the best quantitative results of the reference paper for SRRes-MSE. See Table 1.

	PSNR		SSIM	
	Reference	Ours	Reference	Ours
Set5	32.05	32.08	0.9019	0.9016
Set14	28.49	28.66	0.8184	0.8203
bsd100	27.58	27.58	0.7620	0.7617

Table 1: SRRes-MSE evaluation comparison.

The qualitative results also look reasonable, see Figure 2.

Figure 2: Left: GT. Right: 4x SRRes-MSE output.

References

- C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016.
- X. Liang, C. Xu, X. Shen, J. Yang, S. Liu, J. Tang, L. Lin, and S. Yan, "Human parsing with contextualized convolutional neural network," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1386–1394, Dec 2015.
- J. Tessendorf, C. C, and J. Tessendorf, "Simulating ocean water," 1999.

Recover Folds and Wrinkles in Clothes Images

Since we are interested mostly in clothes, we mask out the background during training and evaluation. See Figure 3.

Figure 3: Top: Sample fashion images. Bottom: Background removed.

Unsurprisingly, the SRGAN-VGG54 model gives the best visual result. Figure 4 shows some examples from the test set. One can see that the model has learned to recover fine folds and wrinkles in clothes - it is able to fill in the missing high frequency details.

Figure 4: Left: GT. Middle: 4x bicubic. Right: 4x model output.

Table 2 shows the quantitative performance.

	Bicubic	SRRes-MSE	SRRes-VGG22	SRGAN-VGG54
PSNR	23.05	24.4	23.6	22.58
SSIM	0.7198	0.7833	0.7523	0.7087

Table 2: Quantitative evaluation on the fashion dataset (masked).

Upsample surfaces

We generated a synthetic dataset of 20K surface patches from a ocean water spectrum [3]. These height field surfaces contain rich features at different frequencies. See Figure 5.

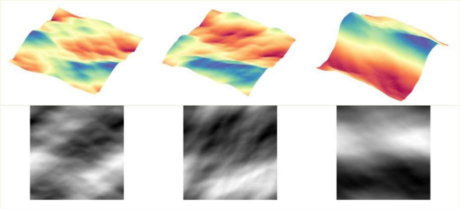

Figure 5: Top: Sample surfaces. Bottom: Depth images.

First we show the result on 4x upsampling in Figure 6. Model has improvement over the bicubic result on high frequency details.

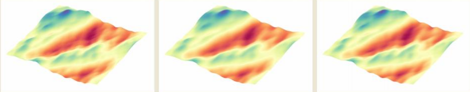

Figure 6: Left: GT. Middle: 4x bicubic. Right: 4x fine-tuned model output.

Next we show the result on 8x upsampling in Figure 7. Here the improvement over the bicubic result is a bit more obvious.

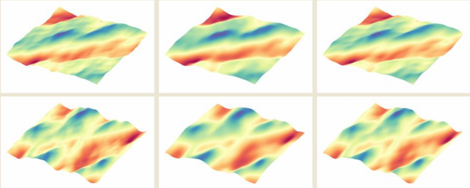

Figure 7: Left: GT. Middle: 8x bicubic. Right: 8x model output.

Table 3 shows the quantitative performance.

	4x		8x	
	Bicubic	Model	Bicubic	Model
PSNR	42.0	46.2	32.8	39.2
SSIM	0.983	0.993	0.917	0.970
Height Field MSE	6.22e-4	2.34e-4	5.30e-3	1.1e-3

Table 3: Quantitative evaluation on the synthetic surface dataset.

Future Work

We would like to 1) gather a cleaner dataset that focuses on folds and wrinkles in clothes, 2) gather a more challenging dataset for surfaces, 3) explore different cost functions for surface upsampling.

CS230 Project, Stanford University

Ning Jin (Jenny)