Deep Learning for image restoration

Deep Learning - Miniproject 1

Ali Benabdallah, João Correia, Gil Tinde
EPFL

Spring 2022

1 Abstract

We compared different image denoising networks trained without a reference clean image trained on 50000 pairs of noisy RGB 32x32 images and tested on 10000 pairs of a noisy image with a clean reference. We used the peak-signal-to-noise ratio (PSNR) to evaluate our predictions. Our best model, which uses 8 convolutional neural networks with residual information, achieves a performance of **25.60** dB when trained for 3 epochs. We compare this model to other architectures, such as a UNet proposed in the original Noise2Noise paper [1], a denoising autoencoder and other networks which leverage deeper architectures or multiple paths from the input to the output.

2 Introduction

Progress in deep learning and computer vision led to major advances in signal reconstruction. Major works ([2–11]) achieved groundbreaking results for image deblurring or denoising on GoPro or Smartphone Image Denoising Dataset. However, most of these tasks consisting of learning a mapping from a corrupted image to a clean one. The authors of [1] show that is also possible to perform signal reconstruction with pairs of corrupted signal without using clean data.

In this work, we compare neural networks trained to perform signal reconstruction with noisy pairs of images and evaluate our models using clean data under a constraint of time. In section 3, we present the experimental settings and architectures that we compare. In section 4, we discuss our results before concluding in section 5.

3 Experimental settings

Each model is evaluated using two different metrics: PSNR and Structured similarity Index Mesure (SSIM) [12]. We train and evaluate each model 10 times in order to have consistent results.

They are all trained using Adam optimizer [13] with default settings from PyTorch [14] using a batch size of 100 images and trained during 3 epochs on Google Colab using a Tesla P100 GPU. Since training samples come in pairs, we swap the (input, target) order of each sample and consider it as additional training data, effectively doubling the size of the training set. All models train for less than 10 minutes.

We also compare two loss functions. Let N be the number of samples, x_i be the input, $f(x_i)$ be the prediction of our model and y_i be the target, we aim to minimize :

- $L_1 = \frac{1}{N} * \sum_{i=1}^{N} ||f(x_i) - y_i||_1$
- $L_2 = \frac{1}{N} * \sum_{i=1}^{N} ||f(x_i) - y_i||_2^2$

3.1 Models

We compare four different architectures : - An 8 convolution layer network with residual paths. We call this model **ConvRes**. - An **AutoEncoder** composed of 5 convolution layers and 5 transposed convolution layers. - A **MultiPath** model composed of an autoencoder path an a path containing 5 convolutional layers. - The **UNet** model presented in the Noise2Noise article [1]

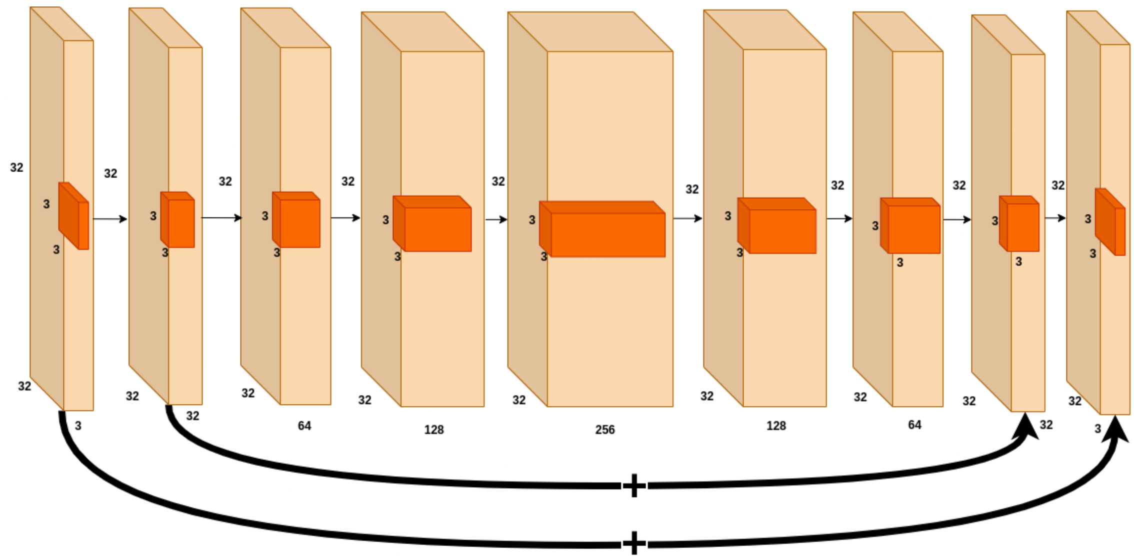

Figure 1: Convres model

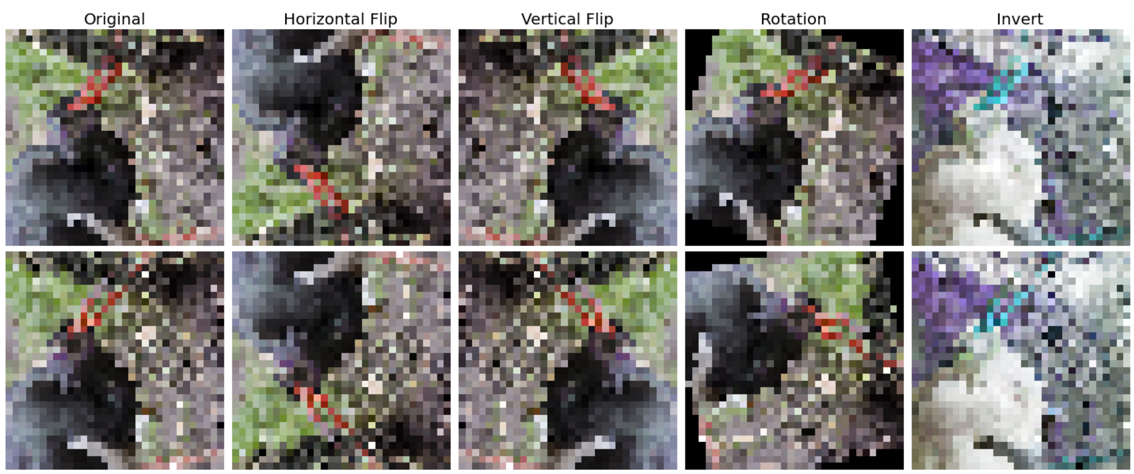

Figure 2: Data augmentation on a pair of noisy images. The two different rows represent the same image with different noise. The represent in order the original sample, a vertically flipped version, a horizontally flipped, an inverted and a rotated version.

Figure 1 better depicts the Convres model and its skip connections. For this model, the activation function used was LeakyRelu.

3.2 Data Augmentation

We augment our dataset with random vertical and horizontal flip, inversion and rotation as presented in figure 2.

4 Results

The results are presented in table 1. The AutoEncoder has a high SSIM score with data augmentation but always has the lowest PSNR whereas our ConvRes achieves better denoising following the PSNR metric. The standard deviation is also low, meaning that this model gives consistent results. Data augmentation does

Table 1: Results

Loss	L_1		L_2	
Metric	SSIM	PSNR	SSIM	PSNR
Without Data Augmentation				
AutoEncoder	0.790 ± 0.059	19.90 ± 0.19	0.646 ± 0.016	19.69 ± 0.17
UNet	0.837 ± 0.001	24.48 ± 0.06	0.844 ± 0.002	25.25 ± 0.06
MultiPath	0.820 ± 0.010	23.81 ± 1.01	0.826 ± 0.008	24.68 ± 0.50
ConvRes	0.841 ± 0.001	24.70 ± 0.04	$\mathbf{0.850 \pm 0.001}$	$\mathbf{25.60 \pm 0.04}$
With Data Augmentation				
AutoEncoder	$\mathbf{0.879 \pm 0.075}$	20.10 ± 0.38	0.698 ± 0.009	20.60 ± 0.23
UNet	0.838 ± 0.002	24.56 ± 0.18	0.843 ± 0.004	$\mathbf{25.30} \pm 0.07$
MultiPath	0.832 ± 0.001	24.47 ± 0.18	0.826 ± 0.007	24.50 ± 0.80
ConvRes	0.839 ± 0.003	24.73 ± 0.09	0.836 ± 0.009	24.23 ± 0.37

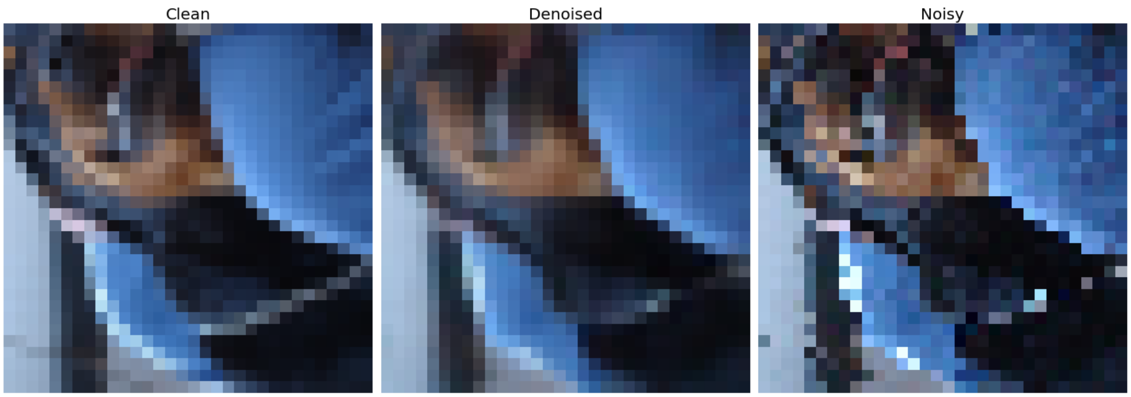

Figure 3: Denoising result

not increase the performances and using the L_2 loss gives better results.

5 Conclusion

Our best performing model performed achieves a very reasonable performance on the denoising task, with a PSNR value of 25.60. This model highlights the importance of convolutional layers on denoising tasks but also of residual/skip connections. This fact is further highlighted by the second best performing mode, the UNet, which also features both these types of mechanisms.

Possible improvements could be the use of custom loss functions specific for the task of image denoising as well as increasing the time limit for the training of the model.

References

[1] Jaakko Lehtinen et al. "Noise2Noise: Learning Image Restoration without Clean Data". In: (2018). DOI: `10.48550/ARXIV.1803.04189`. URL: `https://arxiv.org/abs/1803.04189`.

[2] Liangyu Chen et al. "Simple Baselines for Image Restoration". In: *arXiv preprint arXiv:2204.04676* (2022).

[3] Liangyu Chen et al. "HINet: Half instance normalization network for image restoration". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 182–192.

[4] Shen Cheng et al. "Nbnet: Noise basis learning for image denoising with subspace projection". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 4896–4906.

[5] Sung-Jin Cho et al. "Rethinking coarse-to-fine approach in single image deblurring". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 4641–4650.

[6] Xiaojie Chu et al. "Revisiting Global Statistics Aggregation for Improving Image Restoration". In: *arXiv preprint arXiv:2112.04491* (2021).

[7] Xintian Mao et al. "Deep Residual Fourier Transformation for Single Image Deblurring". In: *arXiv preprint arXiv:2111.11745* (2021).

[8] Zhengzhong Tu et al. "MAXIM: Multi-Axis MLP for Image Processing". In: *arXiv preprint arXiv:2201.02973* (2022).

[9] Zhendong Wang et al. "Uformer: A general u-shaped transformer for image restoration". In: *arXiv preprint arXiv:2106.03106* (2021).

[10] Syed Waqas Zamir et al. "Multi-stage progressive image restoration". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 14821–14831.

[11] Syed Waqas Zamir et al. "Restormer: Efficient Transformer for High-Resolution Image Restoration". In: *arXiv preprint arXiv:2111.09881* (2021).

[12] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. "Multiscale structural similarity for image quality assessment". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 2. Ieee. 2003, pp. 1398–1402.

[13] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[14] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.