ATML Assignment 03 Report

Lukas Zbinden
lukas.zbinden@unifr.ch
University of Bern - ATML, Spring 2018, 14/05/18

1 Denoising Autoencoder

We train and evaluate the performance of different denoising autoencoders (DAE) based on three noise data augmentation techniques, namely additive Gaussian noise, Salt & Pepper noise and masking noise on images from the MNIST dataset.

1.1 Evaluation of 3 denoising tasks

Table 1 depicts the best test accuracy for each of the three denoising autoencoders for the cases of fixed feature representation and finetuning. The numbers were achieved with 2 training epochs for the DAE and 7 epochs for the classifier (due to time constraints), respectively, and a transfer dataset size of 5000, on which the classifier was trained. The finetuned model trained on masked image data (masking) outperforms the others. However, for the fixed DAEs the Salt & Pepper noise originates the most challenging training data such that 'its' DAE performs best. In all but one case the DAE outperform the standard autoencoder (SAE), based on which we can conclude that the noise data augmentation technique is effective in creating a more accurate model at test time.

DAE type	Fixed	Finetuning
None (SAE)	68.6	85.7
Additive Gaussian	67.8	87.9
Salt & Pepper	**73.0**	88.3
Masking	72.2	**88.7**

Table 1: Test accuracy for each type of DAE

1.2 Sample Reconstructions vs. Inputs

Following we show sample noisy inputs generated by the respective noise function along with their reconstructions as they are output by the DAE's decoder.

1.2.1 Additive Gaussian noise

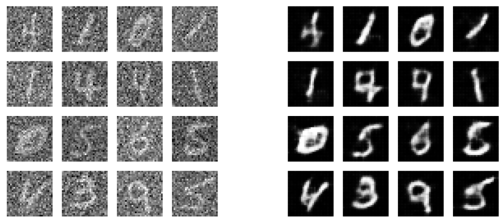
Figure 1 depicts samples of additive Gaussian noise along with their reconstructions.

1.2.2 Salt & Pepper noise

Figure 2 depicts samples of Salt & Pepper noise along with their reconstructions.

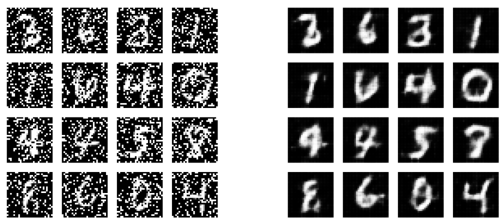
1.2.3 Masking noise

Figure 3 depicts samples of Masking noise along with their reconstructions.



(a) Noisy input (b) Decoder output

Figure 1: Additive Gaussian noise samples and their reconstructions



(a) Noisy input (b) Decoder output

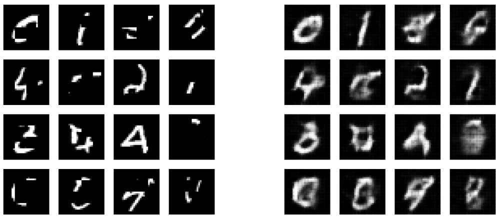
Figure 2: Salt & Pepper noise samples and their reconstructions

1.3 Challenging and easy forms of noise

Masking noise appears to be the most challenging as some of the images are perturbated to a degree to which the number becomes hardly recognizable to the human eye. Also some of the respective reconstructions remain unclear what number they represent. Here the DAE is challenged the most to reconstruct the image and in inpainting the missing regions, respectively.

Also challenging but to a lesser degree seems the Salt & Pepper noise as it basically erases the original image pixel wise at some rate. However the reconstructions seem better recognizable. Lastly the additive Gaussian noise seems the least distorting by considering the noisy input samples and the reconstructions, respectively. It is most likely the easiest for the DAE to reconstruct as the pixel values are not erased completely but only changed according to a normal distribution.

The more challenging a noise is to the DAE the more it is forced to learn the distinctive features for correct reconstruction of the original images and eventually the better it's test time performance will be thanks to the well developed features. This consideration is reflected in table 1 where the DAEs trained on masking and Salt & Pepper noise, respec-



(a) Noisy input (b) Decoder output

Figure 3: Masking noise samples and their reconstructions

tively, outperform the other types.

2 Transfer Learning

Here we evaluate the quality of learnt representations in transfer learning experiments for classification of MNIST images.

2.1 Transfer Performance

Table 2 shows the transfer performance for two transfer types: 'fixed feature' in which the trained encoder is fixed for the training of the classifier and 'finetuning' in which the trained encoder is continued to be trained along with the classifier (i.e. the new task). For each type the performances are collected for different transfer dataset sizes, i.e. the dataset used to train the classifier (and in case of 'finetuning' also the encoder), and finally for each of the four autoencoders. The last column then shows the average test accuracy across all the latter. It stands out that the average performance in the finetuning case is much better than in the fixed case.

2.1.1 Type of noise to yield the best features

The best features are achieved by the masking DAE and finetuning transfer with 88.7% accuracy. Yet all noise types achieve comparable highest scores for the largest transfer dataset size with 'finetuning' whereas they differ more significantly as the dataset size decreases.

2.1.2 Comparing the features to features of a standard autoencoder

The performance of the standard AE improves with increasing dataset size in a similar fashion as the DAEs do. The standard AE however delivers in no case the very best results, however some numbers are very close to the highest. In general, the DAEs, in particular those based on S & P and masking noises, outperform the standard AE especially as the dataset size increases.

2.2 Impact of dataset size on performance of finetuning vs. fixed feature representation

Figure 4 depicts the impact of the transfer dataset size (i.e. on which the classifer is learnt) on the performance of both finetuning and fixed feature representation transfer learning. It shows that finetuning outperforms the fixed variant with respect to all sizes and further continues to deliver results in an almost linear way as the size increases whereas the fixed representation lessens its performance as the dataset gets much bigger with a size of 5000.

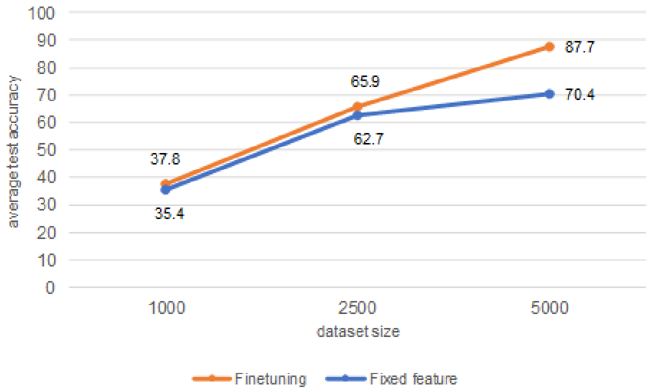


Figure 4: Impact of dataset size on transfer learning type

3 Conclusion

The experiments show that using noise as a form of data augmentation yields notable improvements over the standard dataset as the autoencoders are forced to learn to process harder training examples and thus develop more elaborate feature representations. Further, the experiments demonstrate that transfer learning is a complementary and also effective technique to improve the performance of autoencoders especially in the case of finetuning. Transfer learning is complementary because in the combination of the two techniques lies the best performance as seen in table 2 in the case of masking and S & P noise, respectively.

Transfer type	size	Gaussian	S & P	masking	none	∅
Fixed Feature	**1000**	43.3	26.9	32.0	39.2	**35.6**
	2500	60.6	62.8	63.5	63.8	**62.7**
	5000	67.8	73.0	72.2	68.6	**70.4**
Finetuning	**1000**	32.6	35.7	41.5	41.4	**37.8**
	2500	75.9	59.6	73.8	54.4	**65.9**
	5000	87.9	88.3	88.7	85.7	**87.7**

Table 2: Transfer performance of four autoencoders for different transfer dataset sizes and two transfer types. The last column shows the average test accuracy.