INPAINTING EXPERIMENTS

SUPP. MATERIAL FOR DIRECT EVOLUTIONARY OPTIMIZATION OF VARIATIONAL AUTOENCODERS WITH BINARY LATENTS

Anonymous authors
Paper under double-blind review

The TVAE approach can be applied to a zero-shot inpainting task using a procedure similar to the one presented in the main text for denoising (also see Sec. B.2 in the appendix): a single image with missing pixels (Fig. 1, center) is divided into square patches to form the training set; during training, missing pixels can be treated as unknown observables when evaluating log-joint probabilities of a data-point; Eq. (15) is then used to estimate likely values of the missing pixels, providing "inpainted" datapoints that can be used for DNN backpropagation as usual. As TVAE is a non-amortized approach, missing values can directly be treated in a grounded probabilistic way. Amortized approaches will have to specify how an encoder network should treat missing values.

When evaluating TVAE on standard inpainting benchmarks, we observe competitive performance compared to other approaches. Tab. 1 shows a comparison of inpainting performance (in terms of PSNR) with previous state-of-the-art systems that like TVAE do not require large, clean training data nor information, e.g., on the noise level. As can be observed, TVAE outperforms approaches such as BPFA (Zhou et al., 2012) or Papyan et al. (2017). TVAE performance is lower than for DIP (Ulyanov et al., 2018). TVAE, like BPFA and Papyan et al. (2017), is a permutation-invariant approach, however. That is, the TVAE model is itself not using information about the 2D nature of images. DIP results rely, on the other hand, on a large dedicated DNN with LeakyReLU as activation functions, a U-net / hourglass architecture with skip connections, and convolutional units with reflection padding (see supplement of Ulyanov et al. (2018)). The convolutional stages do explicitly assume the 2D image structure. We also remark that DIP uses in total 2 million parameters (and many more hyperparameters) compared to about 0.5 million parameters of the standard multi-layer perceptron used in TVAE.

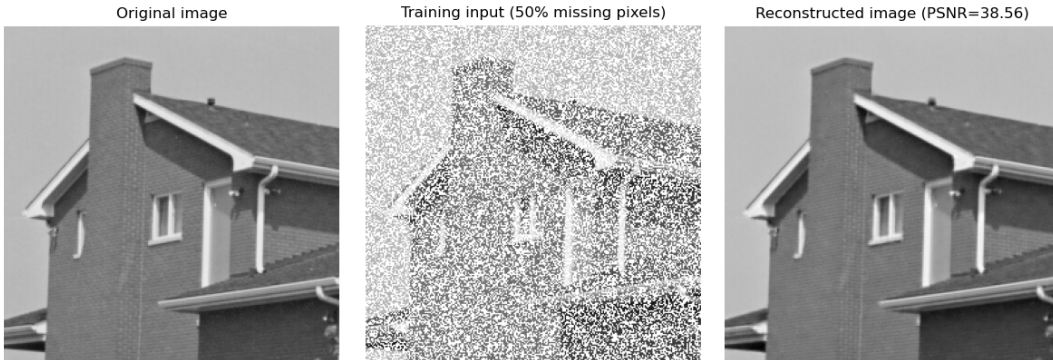Tab. 2 provides a list of the hyper-parameters used for these experiments.

Figure 1: Inpainting of the 'house' image with TVAE.

Table 1: Inpainting performance in PSNR (dB) for the 'house' image with 50% of missing pixels.

Papyan et al.	34.58
BPFA	38.02
TVAE	38.56
DIP	**39.16**

Table 2: Hyper-parameters used to produce the results of Fig. 1. Training lasted 500 epochs taking around 60 seconds/epoch on a single NVIDIA Titan Xp GPU.

Neural network units	
Input (H)	512
Middle	512
Output (D)	144
Cyclic Learning Rates	
Min l.r.	0.0001
Max l.r.	0.01
Epochs/cycle	20
Batch size	32
Evolutionary parameters	
Parents	5
Children	4
Generations	1
Size of $\Phi^{(n)}$	64

REFERENCES

Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad. Convolutional dictionary learning via local processing. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5296–5304, 2017.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Mingyuan Zhou, Haojun Chen, John Paisley, Lu Ren, Lingbo Li, Zhengming Xing, David Dunson, Guillermo Sapiro, and Lawrence Carin. Nonparametric Bayesian Dictionary Learning for Analysis of Noisy and Incomplete Images. *IEEE Transactions on Image Processing*, 21(1):130–144, 2012.