The Deep Capsule Prior – advantages through complexity?

Maximilian Schmidt¹⁺, Alexander Denker¹, and Johannes Leuschner¹
¹ Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 5, 28359 Bremen, Germany

In inverse problems, an extensive number of ground truth samples for the training of supervised deep learning models is seldom available. Unsupervised approaches, like the Deep Image Prior, offer a valuable alternative in this case. In our work, we combine the idea of the Deep Image Prior with recently proposed capsule networks. The new model is tested against a standard convolutional Deep Image Prior on different image processing tasks and computed tomography reconstruction.

1 Introduction

Deep learning approaches have become state-of-art in image classification, translation, denoising and a wide range of other applications. In many cases, so-called supervised methods are used, requiring pairs of input data and ground truth for the training of the network. Depending on the problem, the number of necessary data pairs can be extensive. Inverse problems are an example where the amount of available ground truth data is often limited. Therefore, unsupervised deep learning models have great relevance for these applications. A prominent example of an unsupervised model is the Deep Image Prior (DIP) [1]. For image applications, convolutional neural networks (CNN) are usually used as DIP and the parameters are determined individually for each measurement. Recently, capsules [2] have been introduced as an extension of regular convolutional layers. In our work, we investigate the use of a capsule network for the Deep Image Prior. We compare this Deep Capsule Prior (DCP) model against a classical CNN in tasks from image processing and computed tomography reconstruction.

2 Inverse problems

In inverse problems, one is interested in finding the solution \( x^\dagger \) from noisy measurement \( y^\delta \). The relationship \( Ax^\dagger + \xi = y^\delta \) between solution and measurement can be described by a forward operator \( A : X \to Y \). In addition, the measurements are disturbed by noise. Therefore, the forward mapping is combined with a realization \( \xi \) from a noise generating process and the noise is assumed to be bounded \( \mathcal{D}_Y (Ax^\dagger, y^\delta) \leq \delta \) w.r.t. some discrepancy \( \mathcal{D}_Y : Y \times Y \to \mathbb{R}_+ \). At best, there exists a unique solution \( x^\dagger \) and the inverse mapping \( A^{-1} : Y \to X \) is continuous, i.e. stable against small perturbations. Otherwise, the inverse problem is called ill-posed in the sense of Hadamard. To determine a candidate \( \hat{x} \) for the solution, one can solve an optimization problem

\[
\hat{x} = \arg \min_{\dot{x}} \mathcal{D}_Y (Ax, y^\delta) + \alpha \mathcal{R}(x). \tag{1}
\]

Since the majority of inverse problems are ill-posed, it is advisable to stabilize the problem by including prior knowledge in form of a weighted regularization term \( \mathcal{R} : X \to \mathbb{R}_+ \).

3 Deep Capsule Prior

The Deep Capsule Prior is based on the idea of the Deep Image Prior to convert the optimization problem (1) into a parameter identification problem for a function \( \varphi_\Theta : Z \to X \)

\[
\hat{\Theta} = \arg \min_\Theta \mathcal{D}_Y (A\varphi_\Theta(z), y^\delta) + \alpha \mathcal{R}(x), \quad \hat{x} = \varphi_\Theta(z). \tag{2}
\]

In general, \( \varphi \) is a deep neural network with parameters \( \Theta \) and the final parameter configuration \( \hat{\Theta} \) is determined by gradient descent. Note, that the parameter problem (2) is usually non-convex and, therefore, the existence of a unique configuration \( \hat{\Theta} \) is not guaranteed. The idea of the DIP is combined with the recently proposed capsule networks. Capsules are multi-dimensional extensions of regular neurons (cf. Sabour et al. [2]). In addition, the activation function, called routing, acts on all inputs simultaneously and can be understood as a clustering step. This additional complexity was shown to be beneficial in supervised tasks like classification and semantic segmentation. Its effect on unsupervised problems is investigated in the next section.

4 Numerical experiments

For all DIP & DCP experiments we use a U-Net network architecture and replace the convolutional layers with their capsule equivalent in case of the DCP. For the routing mechanism, the direct routing as proposed by Sabour et al. [2] is employed. Both networks are regularized by total variation (TV), i.e. \( \mathcal{R}(x) = \| \nabla x \|_1 \). We investigate the same image processing tasks...
as in the original publication by Ulyanov et al. [1] and show the results for the inpainting experiments. The other results and the code for all experiments can be found at https://github.com/mschmidt25/deep-capsule-prior.

### 4.1 Inpainting

In inpainting, we are given an image \( y = M \odot x \) in which pixels are missing as measurement data. The mask \( M \) is a binary matrix specifying which pixels are missing. The L2 norm is used as the discrepancy. As the discrepancy is independent of the masked pixels, the prior has a large influence in inpainting. If minimized directly over the pixel values of the image, as in equation (1), the optimization would have no effect on the masked pixels.

![Masked Image DIPDCP](Masked Image DIPDCP)

**Fig. 1:** Inpainting a large region. The white pixels in the masked image correspond to the missing pixels. The results for the DIP were created using the code from the official Github https://github.com/DmitryUlyanov/deep-image-prior.

### 4.2 Computed Tomography

In computed tomography (CT) the forward operator \( A \) can be modeled by the Radon transform. For the numerical experiments we use two different datasets with 100 test images each. The first consists of synthetic ellipses and simulated sparse-angle, parallel beam measurements with added Gaussian noise. The measurements are undersampled, such that there exist multiple possible solutions. The reconstructed images have a size of 128 px × 128 px. The second dataset contains simulated low-intensity measurements from real, normal-dose thoracic CT images of size 362 px × 362 px. In this case, the measurements are oversampled and the existence of a solution is thereby not guaranteed. The DCP reconstructions are compared against results from Baguer et al. [3], which include a CNN DIP and classical reconstruction methods. We refer the reader to their publication for more information on the two datasets and the included methods. The results are displayed in Table 1.

| Method                  | Ellipses  | LoDoPaB   |
|-------------------------|-----------|-----------|
|                         | PSNR  | SSIM       | PSNR  | SSIM       |
| DCP + TV (ours)         | 28.03  | 0.8656     | 33.96  | 0.8127     |
| DIP + TV                | 28.94  | 0.8855     | 34.44  | 0.8143     |
| TV                      | 27.84  | 0.8495     | 32.95  | 0.8034     |
| Filtered Back Projection| 24.18  | 0.5939     | 30.37  | 0.7386     |

**Table 1:** Results on the first 100 test images on the Ellipses and LoDoPaB datasets. Higher values are better.

![DIP and DCP reconstruction on a LoDoPaB test image](DIP-DCP-LoDoPaB)

**Fig. 2:** DIP and DCP reconstruction on a LoDoPaB test image.

5 Conclusion

The Deep Capsule Prior showed promising results in the numerical experiments. Still, further research is necessary to surpass the performance of a classical CNN-based Deep Image Prior in the future and, thereby, justify the additional computational cost. Potential improvements lie in the choice and design of the routing function and the network architecture.

Acknowledgements

M.S. A.D. and J.L. acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) within the framework of GRK 2224/1 “π³: Parameter Identification – Analysis, Algorithms, Applications”. M.S. and J.L. additionally acknowledge support by the project DELETO funded by the Federal Ministry of Education and Research (BMBF, project number 05M20LBB). A.D. further acknowledges support by the Klaus Tschira Stiftung via the project MALDISTAR (project number 00.010.2019). Open access funding enabled and organized by Projekt DEAL.

References

[1] D. Ulyanov, A. Vedaldi, V. Lempitsky, *Deep image prior*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9446–9454.

[2] S. Sabour, N. Frosst, and G. E. Hinton, *Dynamic routing between capsules*, Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, (Curran Associates Inc., Red Hook, NY, USA, 2017), pp. 3859–3869.

[3] D. O. Baguer, J. Leuschner, M. Schmidl, *Computed tomography reconstruction using deep image prior and learned reconstruction methods*, Inverse Problems 36 (9), 2020, p. 094004, DOI: http://dx.doi.org/10.1088/1361-6420/aba415

© 2021 The Authors. *Proceedings in Applied Mathematics & Mechanics* published by Wiley-VCH GmbH. www.gamm-proceedings.com