Seismic data interpolation using an anti-over-fitting mixed-scale dense convolutional neural network

Zhang, D.; Verschuur, E.

DOI
10.3997/2214-4609.2023101026

Publication date
2023

Document Version
Final published version

Citation (APA)
Zhang, D., & Verschuur, E. (2023). *Seismic data interpolation using an anti-over-fitting mixed-scale dense convolutional neural network*. Paper presented at 84th EAGE ANNUAL Conference and Exhibition 2023, Vienna, Austria. https://doi.org/10.3997/2214-4609.2023101026

Seismic data interpolation using an anti-over-fitting mixed-scale dense convolutional neural network

D. Zhang[1], E. Verschuur[1]

[1] Delft University of Technology

Summary

Seismic data interpolation is a topic well suited for deep learning (DL) applications. Scaling operation-based DL neural networks, e.g., U-Net, have been popular since its booming development in the field of seismic data processing. Although many successful studies using U-Net on seismic data, scientists start to realize the downside of its implementation, i.e., large trainable parameters (normally larger than 1 million), the potential risks of over-fitting, and tedious hyper-parameter selection. Therefore, in this abstract, we introduce a mixed-scale dense convolutional neural network (MS-DCNN) for seismic data interpolation with relatively few trainable parameters to reduce the risk of over-fitting. This MS-DCNN was originally developed for biomedical image processing. In addition, this neural network can be trained with relatively small training set. Via a field data case study, the different behavior of U-Net and MS-DCNN is analyzed and compared for a specific interpolation problem, where 9 consecutive shot records were missing from a 2D line of marine seismic data.

Seismic data interpolation using an anti-over-fitting mixed-scale dense convolutional neural network

Introduction

Many interpolation approaches have been developed over the years since most seismic processing and imaging methods require data to be regularly and densely sampled in the spatial direction. With recent developments of deep learning (DL), new options for seismic data interpolation are explored (Kaur et al., 2020; Fang et al., 2021; Qu et al., 2021; Saad et al., 2022). The U-Net (Ronneberger et al., 2015) gradually becomes one of the most widely used and stable DL neural networks in the field of seismic data processing. Qu et al. (2021) shows a DL approach based on a U-Net trained on only synthetic and applied to field data, specifically for the case of missing near offsets. This approach outperforms the traditional parabolic Radon-based near-offset interpolation (Kabir and Verschuur, 1995). Although many successful cases using U-Net on seismic data, scientists start to realize the downside of its implementation, i.e., large trainable parameters (normally larger than 1 million), the potential risks of over-fitting, tedious hyper-parameter selection, and non-flexibility to adapt to other seismic applications. Therefore, in this abstract, we introduce a mixed-scale dense convolutional neural network (MS-DCNN) for seismic data interpolation with relatively few trainable parameters to reduce the risk of over-fitting.

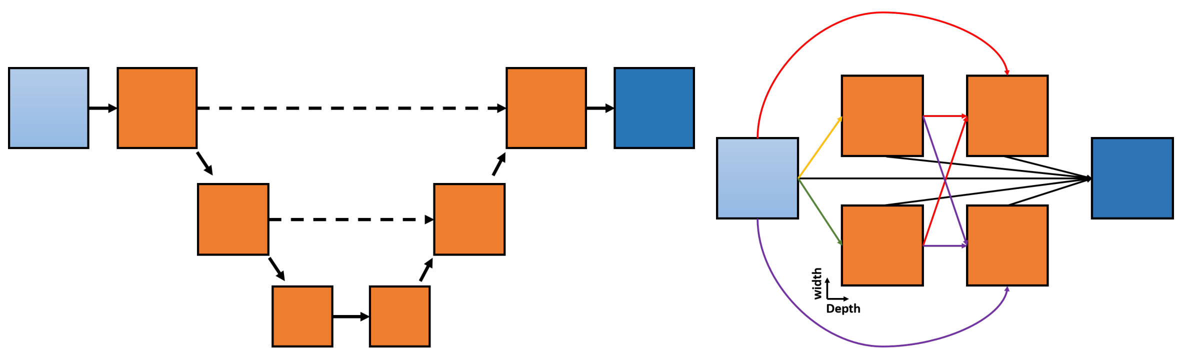
Mixed-scale dense convolutional neural network (MS-DCNN)

MS-DCNN is proposed by Pelt and Sethian (2017) for biomedical image analysis and tomographic image reconstruction (Pelt et al., 2018). The fundamental differences between conventional scaling operation-based DL neural networks, e.g., U-Net, and MS-DCNN can be observed from their simplified equations as follows:

$$\mathbf{x}_i^j = \sigma(\mathscr{C}_{ij}(\mathbf{x}_{i-1}) + b_{ij}) \tag{1}$$

$$\mathbf{x}_i^j = \sigma(\mathscr{D}_{ij}(\{\mathbf{x}_0, \cdots, \mathbf{x}_{i-1}\}) + b_{ij}) \tag{2}$$

where $\mathbf{x}$, b, and σ represents the feature map, bias, and activation function, respectively. Layer and channel index are indicated by i and j. $\mathscr{C}_{ij}$ represents a 2D convolution operation. $\mathscr{D}_{ij}$ indicates dilated convolutions (Yu and Koltun, 2015), which are able to capture additional large features without the traditional scaling operations. Another advantage of MS-DCNN is all previously calculated feature maps, i.e., $\{\mathbf{x}_0, \cdots, \mathbf{x}_{i-1}\}$, including the training input, are combined together to generate the output as shown in Figure 1. In this way, we can reuse all previously computed feature maps to the maximum. In the end, much fewer parameters are required to achieve the similar performance of standard scaling operation-based neural networks, which can boost its anti-over-fitting ability. Note that the number of trainable parameters for MS-DCNN is controlled by its depth and width as indicated in Figure 1(b).



(a) (b)

Figure 1 *(a) A simplified demonstration of common scaling operation-based DCNN with 2 downscaling operations and 2 upscaling operations, e.g., U-Net. Downward arrows indicate encoding or downscaling operations, upward arrows indicate decoding or upscaling operations, and dashed lines indicate skip connections. (b) A simplified representation of MS-DCNN with depth 2 and width 2. Black lines indicate* 1×1 *convolutions while colored lines represent* 3×3 *dilated convolutions.*

Field data case study using U-Net and MS-DCNN

To better verify and compare the performance of different DL neural networks regarding the interpolated results, we consider a field data case study from the Voring area in the Norweigian Sea. It is a deep-water area (water depth approximately 1.3 km), and the 2D line under consideration consists of about 400 shot records with maximum offset of 4625 m, with each shot having 180 receivers at 25 m spacing after pre-processing from the original data (Davydenko and Verschuur, 2018). However, 9 consecutive shots were missed during the acquisition, so there is a 9-shot gap in the data. The near-offset section is shown in Figure 2, where the gap is clearly visible on the right part.

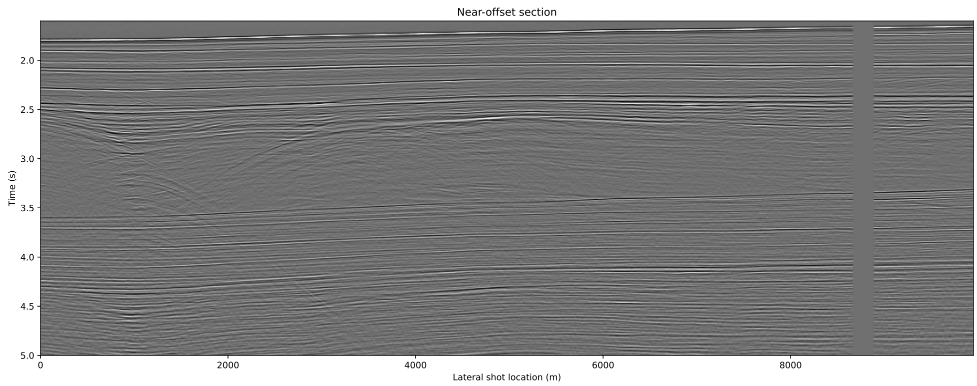


Figure 2 *The near-offset section of the voring field data, showing the missing gap of 9 shots.*

For our problem at hand, we can use the well-sampled part of shot records as training data, which is often used in DL to see if the data themselves provide the required training data. An example from literature is given by Yeeh et al. (2020) who applies seismic cross-line interpolation via learning from densely sampled inline direction.

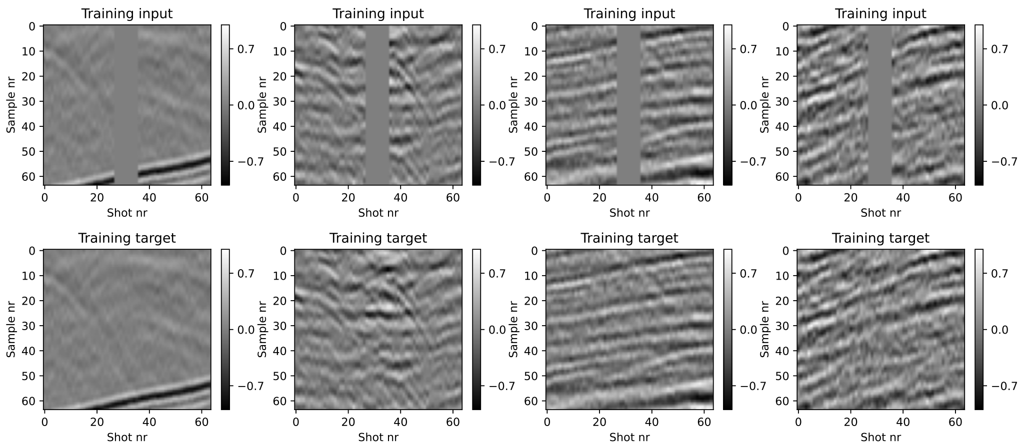


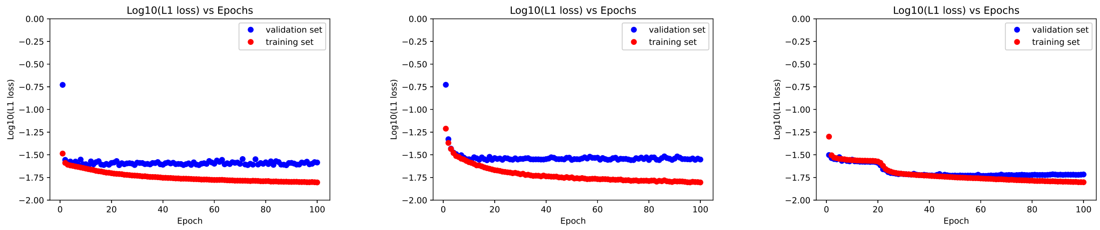
Figure 3 *Four random selected training pairs that are offset-domain panels with* 64×64 *panel size from the well-sampled part of data. The first row indicates the training input with the missing gap of 9 consecutive shots while the second row is the corresponding ground truth panel without gap.*

Next, we will apply the missing shot interpolation in three different settings:

1. DL trained via a U-Net with approximately 1.545 million trainable parameters in 5000 common offset panels (U-Net 1).
2. DL trained via a U-Net with approximately 0.387 million trainable parameters in 500 common

offset panels (U-Net 2).

3. DL trained via a MS-DCNN with approximately 0.016 million trainable parameters in 500 common offset panels (MS-DCNN).



(a) (b) (c)

Figure 4 *Loss function of (a) U-Net with approximately 1.545 million trainable parameters, i.e., U-Net 1, on 5000 training samples, (b) U-Net with approximately 0.387 million trainable parameters, i.e., U-Net 2, on 500 training samples, and (c) MS-DCNN with approximately 0.016 million trainable parameters on 500 training samples.*

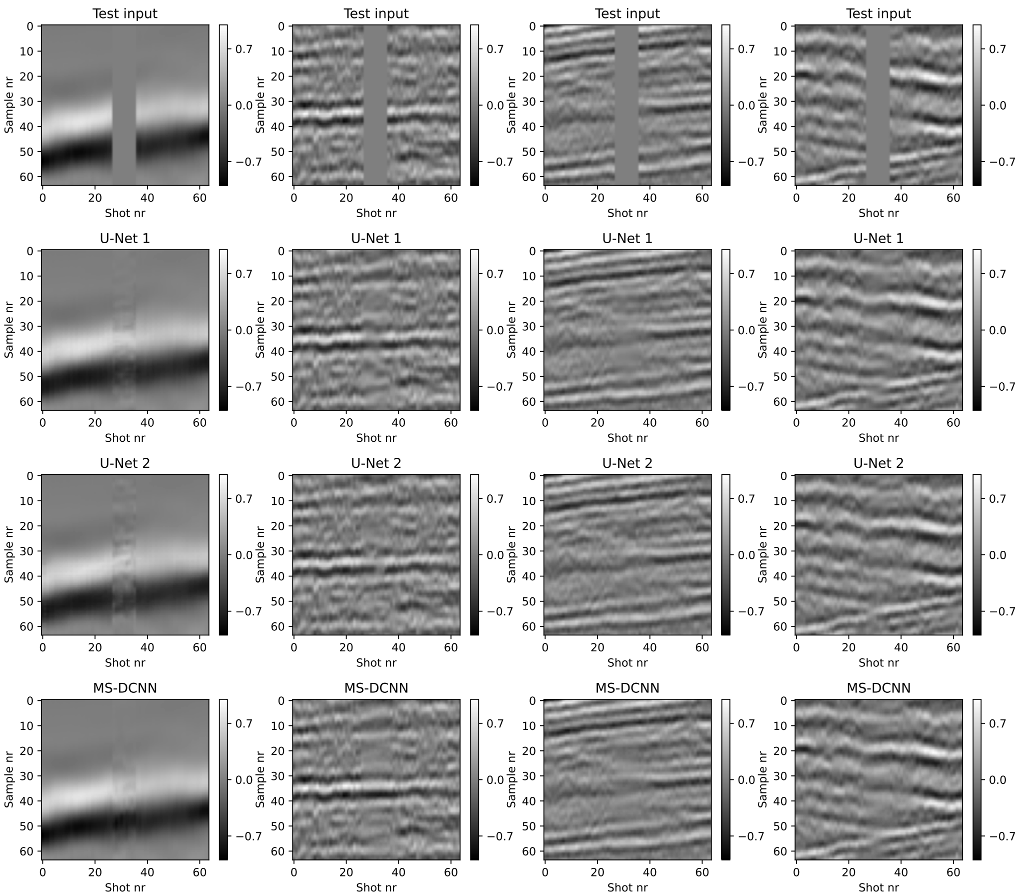


Figure 5 *Four random selected test panels after DL in the common offset domain. The first row is the input test panel with missing gap of 9 consecutive shots. The rest three rows corresponds to the interpolated results after U-Net 1, U-Net 2, and MS-DCNN, respectively.*

During the training, all three settings are trained on 64×64 data panels. These panels/windows are selected along common offset sections in the normal moveout-corrected seismic data. As for the training pairs shown in Figure 3, the missing gap of 9 consecutive shots is always put at the same location for convenience. Note that U-Net 1 and 2 have the same general structure (3 downscalings and 3 upscalings) but different number of channels, which leads to different number of trainable parameters. For U-Net 2 we reduced the number of parameters because of the smaller training size. In addition, we use depth 30 and width 2 for MS-DCNN, which requires much less trainable parameters. In Figure 4, we see the development of the loss function, plotted at logarithmic scale for all three settings. Although the overall loss values do not differ that much for U-Net 1 and U-Net 2, we observe more over-fitting for U-Net 2 due to much less number of training samples, i.e., 500 samples compared to 5000 used in U-Net 1. However, it is clear that the loss function of MS-DCNN is significantly better with only slightly over-fitting after 60 epochs. Besides, the loss value of the validation data set (blue curve) is lower than both U-Nets, which demonstrates its good performance and anti-over-fitting ability.

Furthermore, Figure 5 shows four interpolated test panels using DL. At visual inspection, it seems that the MS-DCNN gives more consistent results with less noisy imprints. Note that all interpolated test panels can be merged to reconstruct the complete data. Please keep in mind that MS-DCNN achieves better results than U-Net with 10 times fewer training samples and approximately 100 times fewer parameters, which verifies again its anti-over-fitting capability. In terms of the computational training cost for the previous three different settings, the MS-DCNN is similar to U-Net 2, and approximately 8 times faster than U-Net 1. The costs of MS-DCNN heavily rely on its depth and width.

Conclusions

We have introduced an anti-over-fitting MS-DCNN for seismic data interpolation, which can achieve at least similar or often better results with much fewer training samples and trainable parameters than the conventional U-Net. A field data case study demonstrate the good performance of MS-DCNN in terms of its anti-over-fitting capability and a more consistent interpolated result.

Acknowledgements

The authors thank the Delphi Consortium members for their support and Equinor for providing the Voring field data.

References

Davydenko, M. and Verschuur, D.J. [2018] Including and using internal multiples in closed-loop imaging — Field data examples. *Geophysics*, **83**(4), R297–R305.

Fang, W., Fu, L., Zhang, M. and Li, Z. [2021] Seismic data interpolation based on U-net with texture loss. *Geophysics*, **86**(1), V41–V54.

Kabir, M.N. and Verschuur, D. [1995] Restoration of missing offsets by parabolic Radon transform. *Geophysical Prospecting*, **43**(3), 347–368.

Kaur, H., Pham, N. and Fomel, S. [2020] Seismic data interpolation using deep learning with generative adversarial networks. *Geophysical Prospecting*, **69**(2), 307–326.

Pelt, D.M., Batenburg, K.J. and Sethian, J.A. [2018] Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks. *Journal of Imaging*, **4**(11), 128.

Pelt, D.M. and Sethian, J.A. [2017] A mixed-scale dense convolutional neural network for image analysis. *Proceedings of the National Academy of Sciences*, **115**(2), 254–259.

Qu, S., Verschuur, E., Zhang, D. and Chen, Y. [2021] Training deep networks with only synthetic data: Deep-learning-based near-offset reconstruction for (closed-loop) surface-related multiple estimation on shallow-water field data. *Geophysics*, **86**(3), A39–A43.

Ronneberger, O., Fischer, P. and Brox, T. [2015] U-Net: Convolutional networks for biomedical image segmentation. Springer International Publishing, 234–241.

Saad, O.M., Fomel, S., Abma, R. and Chen, Y. [2022] Unsupervised deep learning for 3D interpolation of highly incomplete data. *Geophysics*, **88**(1), WA189–WA200.

Yeeh, Z., Byun, J. and Yoon, D. [2020] Crossline interpolation with the traces-to-trace approach using machine learning. *90th SEG Annual International Meeting*, Expanded Abstracts, 1656–1660.

Yu, F. and Koltun, V. [2015] Multi-scale context aggregation by dilated convolutions. *arXiv.org*.