Improved Automating Seismic Facies Analysis Using Deep Dilated Attention Autoencoders

Zengyan Wang, Fangyu Li, Thiab R. Taha, Hamid R. Arabnia
The University of Georgia, Athens, GA, USA
zengyan@cs.uga.edu

Abstract

With the dramatic growth and complexity of seismic data, manual annotation of seismic facies has become a significant challenge. The encoder-decoder neural network architecture has been widely used in image segmentation. In recent years, the same architecture has also been used in seismic surveys for facies classification applications.

In this paper, a modified U-Net architecture with trainable soft attention mechanism and dilated convolution is proposed to improve the automatic seismic facies analysis. This proposed framework generates more accurate results in a more efficient way. The dilated convolution achieves more accurate results with less computation than the CNN with pooling in U-Net. With the attention mechanism, the dilated U-Net model further improves classification accuracy. Our experiments show that the dilated attention autoencoder model is less prone to overfitting and at the same time, it achieves a smoother increasing validation accuracy.

1. Introduction

Seismic facies analysis plays a key role in interpreting geological patterns. With the dramatic growth and complexity of 3D seismic surveys, human experts manually interpret a seismic volume, which is more and more challenging [13]. Previous experiences show that convolutional neural network (CNN) based models provide improved results over traditional methods such as support vector machines (SVMs) and self-organizing maps (SOMs) for identifying geologic features in 3D volumes [3, 9, 10, 12].

The main advantage of CNN is the end-to-end automatic processing of extracting features from original seismic data to generate geological classification results, which increases the efficiency of interpretation and reduces the dependency on hand-crafted seismic attributes [2]. Seismic data classification can be formulated as an image segmentation problem. The encoder-decoder neural network is a commonly used architecture, which usually uses CNN and pooling to produce low level representation followed by trainable filters to producing multi-dimensional features for each pixel for segmentation. Shi [9] uses an encoder-decoder model for salt-body classification. The results show that the model captures subtle salt features but the segmentation maps have crossing-shape noises. Zhao [12] compares seismic facies classification results between 3D patch-based CNN model and 2D whole seismic slice convolution encoder-decoder model. The experiments indicate that the 2D model performs higher accuracy on the classification. Wang [10] involves the spectral decomposition in the 2D segmentation model and increases the model interpretability.

In this study, we develop a dilated attention U-Net to perform seismic facies classification. Seismic data inherently have high resolution and noises. So we use dilated convolution to replace convolution-pooling operation to reduce computation costs. Meanwhile we use a soft-attention mechanism [8] to highlight the salient features while suppressing the irrelevant interference. We also conduct experiments on the dilation factor to expand receptive field efficiently and properly. Furthermore, the resulting attention maps can be used to provide more interpretation insights linking seismic amplitudes and geological structures [8].

2. Method

2.1. Model Development

Our models build upon the U-Net [7]. We employ the soft trainable attention mechanism from Attention-Gated Networks architectures [4, 6, 8]. The attention is used through skip connections to guide the network focus on salient regions during the training process. The overview of the model architectures are shown in Figure 1, where the top branch is the U-Net architecture and the bottom branch used dilated convolution to replace the max pooling operation in U-Net, which effectively enlarges the field of view of filters without increasing the amount computation [1].

To segment objects from noisy data, one of the important elements is the context information. In U-Net, the context information is acquired by the pooling layers after CNN to

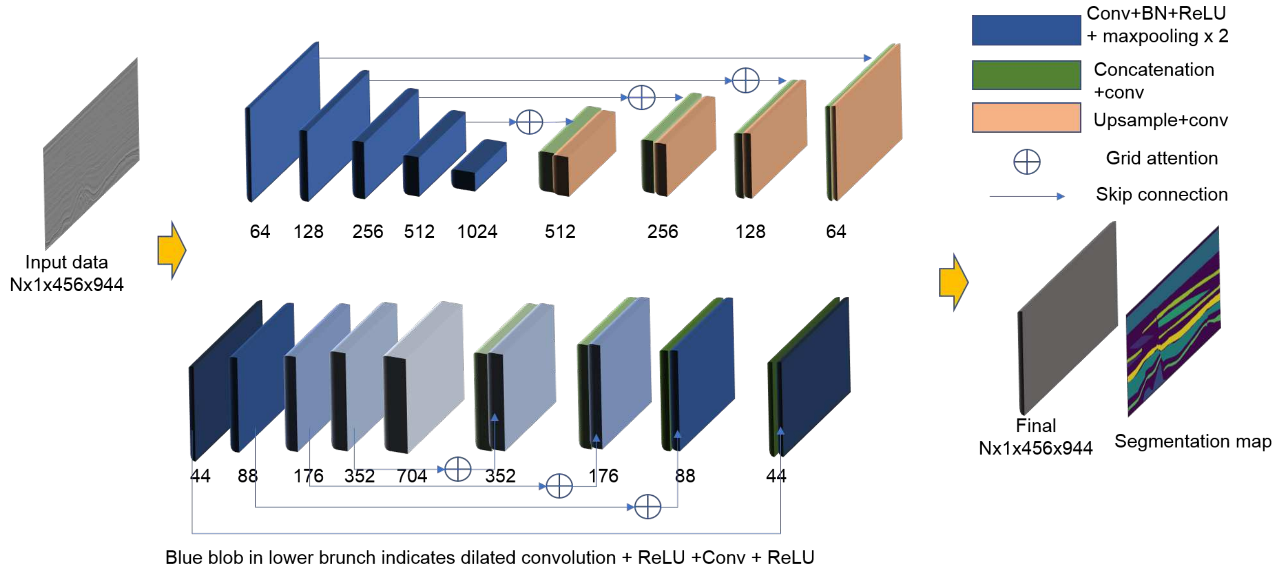

Figure 1. The overview of our studied auto-encoder and decoder architecture with trainable soft attention mechanism. The input data are converted to advanced features by convolution (or dilated convolution) at the encoder path. The extracted features are recovered to original resolution by interpolation and convolution at the decoder path. The attention mechanism filters the features passed by the skip connection. The number under each blob indicates the number of channels after each layer operation.

expand the receptive field. But the global information and resolution are gradually lost through layers of a network. The resulting coarse features may not recover subtle information for segmentation [11]. While dilated convolution can expand the receptive field without losing global information. The dilation factor increases the space between the kernel values to expand the receptive field without increasing too much computation and loss of context information, which is a desirable method for large seismic data.

The soft attention mechanism computes attention probabilities over intermediate features which highlight the relevant parts of the input image for segmentation task [4]. The mechanism performs right before the concatenation operation through skip connections to merge only relative features at different stages [6]. Seismic data are intuitively noisy, thus attention can enforce the model to identify salient features and suppress confusing information.

2.2. Input preparation

In this study, we use the F3 seismic survey acquired in the North Sea, offshore Netherlands [5]. We shift the original seismic data towards the positive direction to eliminate negative values. We do this because the rectified linear unit (ReLU) in the model will suppress negative values - we need to keep the phase in the data to leave magnitude response unchanged. The shifted $3D$ volume is cut into 651 $2D$ amplitude slices. The ground truth in each slice was determined manually by an independent expert with nine seismic facies including: Varies amplitude steeply dipping, Random, Low coherence, Low amplitude deformed, Low amplitude dipping, High amplitude deformed, Moderate amplitude continuous, Chaotic and others as shown in Figure 2. Our final dataset consists of 41 slices from position 101 to 501 with interval 10. The initial size of each slice, 462×951, was reduced to 456×944, for dilated convolution to avoid the overhead introduced by upsampling the features to the original resolution.

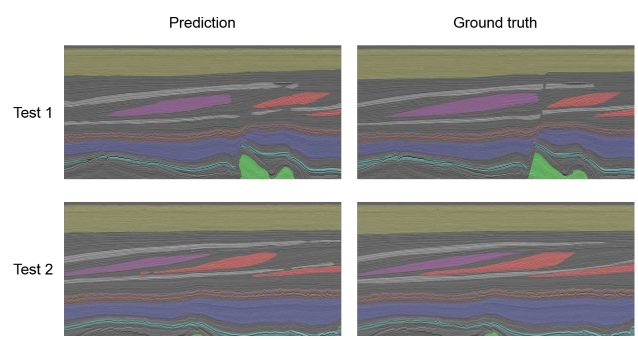

Figure 2. Two seismic facies classification examples using the proposed model. The prediction results on the test set are shown and compared with ground truth labels.

2.3. Experiments

In this study, we compare different combination strategies discussed above: U-Net, U-Net with attention, dilated U-Net and dilated U-net with attention. The dilation factor is set to 1 for dilated convolution. The proposed models are compared in term of Intersection over Union (IoU) metric. All models are trained from scratch on an Nvidia GTX 1080Ti GPU. We divide the dataset into a training and test sets in $9:1$ proportion. A validation dataset is prepared using 10% data from the training dataset.

The training process is applied with a mini-batch size of 4, an initial learning rate of 5×10^{-4} with step decay 1×10^{-6}, a weight decay of 4×10^{-4} and a momentum factor of 0.9 with Adam optimization for 500 epochs.

3. Results

Figure 2 shows examples of the prediction results from Dilated U-Net with attention on the test set. This model provides a clean facies distribution. Figure 3 shows the curves of validation accuracy at each epoch. The models using CNN with pooling have more volatile curves than dilated convolution models.

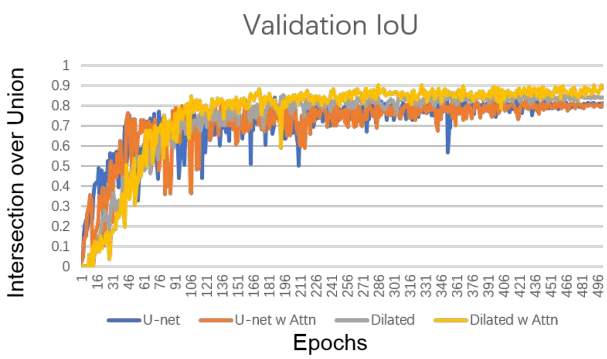

Figure 3. The validation accuracy curves of different models at each epoch during training.

Table 1 lists the experimental results on four models on test set. The dilated U-Net with attention mechanism yields promising results.

Method	IoU on test data	Number of parameters
U-Net	0.813	$31,042,889$
Dilated	0.822	$7,856,869$
U-Net+Attn	0.820	$34,424,652$
Dilated+Attn	0.883	$8,074,406$

Table 1. Performance in IoU on four models.

4. Conclusion

In this study, we proposed a dilated U-Net with soft attention mechanism to perform automatic seismic facies analysis. We compared four models of different combinations of dilated convolution and attention mechanism. The proposed model generates superior classification result over the other methods. The generated attention maps can also be used for explaining the effect of seismic amplitude in the geological interpretation by visualizing the attention maps according to each class.

References

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.

[2] Haibin Di, Zhen Wang, and Ghassan AlRegib. Why using cnn for seismic interpretation? an investigation. In *SEG Technical Program Expanded Abstracts 2018*, pages 2216–2220. Society of Exploration Geophysicists, 2018.

[3] Lei Huang, Xishuang Dong, and T Edward Clee. A scalable deep learning platform for identifying geologic features from seismic attributes. *The Leading Edge*, 36(3):249–256, 2017.

[4] Saumya Jetley, Nicholas A Lord, Namhoon Lee, and Philip HS Torr. Learn to pay attention. *arXiv preprint arXiv:1804.02391*, 2018.

[5] Fangyu Li, Bo Zhang, Rui Zhai, Huailai Zhou, and Kurt J Marfurt. Depositional sequence characterization based on seismic variational mode decomposition. *Interpretation*, 5(2):SE97–SE106, 2017.

[6] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[8] Jo Schlemper, Ozan Oktay, Liang Chen, Jacqueline Matthew, Caroline Knight, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention-gated networks for improving ultrasound scan plane detection. 2018.

[9] Yunzhi Shi, Xinming Wu, and Sergey Fomel. Automatic salt-body classification using a deep convolutional neural network. In *SEG Technical Program Expanded Abstracts 2018*, pages 1971–1975. Society of Exploration Geophysicists, 2018.

[10] Zengyan Wang, Fangyu Li, Thiab R. Taha, and Hamid R. Arabnia. 2d multi-spectral convolutional encoder-decoder model for geobody segmentation. In *International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1183–1188, 2018.

[11] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *ICLR*, 2016.

[12] Tao Zhao. Seismic facies classification using different deep convolutional neural networks. In *SEG Technical Program Expanded Abstracts 2018*, pages 2046–2050. Society of Exploration Geophysicists, 2018.

[13] Tao Zhao, Vikram Jayaram, Atish Roy, and Kurt J Marfurt. A comparison of classification techniques for seismic facies recognition. *Interpretation*, 3(4):SAE29–SAE58, 2015.