Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

A comparison of deep learning methods for seismic impedance inversion

Si-Bo Zhang [a], Hong-Jie Si [a], Xin-Ming Wu [b, *], Shang-Sheng Yan [b]

[a] Huawei Cloud EI Product Department, Xi'an, Shaanxi, 710077, China
[b] School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui, 230026, China

ARTICLE INFO

Article history:
Received 25 August 2021
Accepted 11 January 2022
Available online 22 January 2022

Edited by Jie Hao

Keywords:
Seismic inversion
Impedance
Deep learning
Network architecture

ABSTRACT

Deep learning is widely used for seismic impedance inversion, but few work provides in-depth research and analysis on designing the architectures of deep neural networks and choosing the network hyperparameters. This paper is dedicated to comprehensively studying on the significant aspects of deep neural networks that affect the inversion results. We experimentally reveal how network hyperparameters and architectures affect the inversion performance, and develop a series of methods which are proven to be effective in reconstructing high-frequency information in the estimated impedance model. Experiments demonstrate that the proposed multi-scale architecture is helpful to reconstruct more high-frequency details than a conventional network. Besides, the reconstruction of high-frequency information can be further promoted by introducing a perceptual loss and a generative adversarial network from the computer vision perspective. More importantly, the experimental results provide valuable references for designing proper network architectures in the seismic inversion problem.

1. Introduction

Seismic impedance inversion has been studied for decades as it is one of the most effective methods for reservoir characterization in seismic exploration. It aims at reconstructing impedance sequences from their corresponding seismic traces which is based on the forward model:

$$\mathbf{s} \approx \mathbf{w} * \mathbf{r}, \tag{1}$$

where $\mathbf{s}$ is a seismic trace which is approximated as the convolution of a wavelet $\mathbf{w}$ and a reflectivity sequence $\mathbf{r}$. Impedance $\mathbf{i}$ and reflectivity $\mathbf{r}$ have the following relationship:

$$\mathbf{r}[k] = \frac{\mathbf{i}[k] - \mathbf{i}[k-1]}{\mathbf{i}[k] + \mathbf{i}[k-1]} \tag{2}$$

where $\mathbf{i}[k]$ represents the value of vertical impedance sequence at depth k. Solving $\mathbf{i}$ from $\mathbf{s}$ is an underdetermined problem (Jackson, 1972), so traditional methods (Hu et al., 2009; Zhang and Castagna, 2011; Zhang et al., 2013) use different regularization to constrain solution space, and also many structure-guided methods are proposed (Ma et al., 2012; Zhang and Revil, 2015; Zhou et al., 2016; Wu, 2017). Although these methods are widely used in the industry, some drawbacks still remain as the regularization terms need to be pertinently designed which may limit the generalization of the model. Besides, these model-driven methods usually need to solve an optimization problem which is time-consuming and often yields a smooth result. In addition, the wavelet $\mathbf{w}$ in Equation (1) is typically unknown and might be hard to estimate as it is often varying in time and space. In practice, the relationship between the recorded seismogram and the true impedance is much more complicated than the simple convolution model described in Equations (1) and (2). The acquisition limitations, potential measurement errors, processing errors, and noise make the impedance estimation from seismograms a highly nonlinear problem with large uncertainties. Therefore, a data-driven deep learning method is expected to estimate the complicated and nonlinear relationship between the seismic traces and impedance sequences.

In recent years, Deep Learning (DL) has an explosive development in Computer Vision (CV) (Krizhevsky et al., 2012). Various architectures and skills (Szegedy et al., 2015; He et al., 2016; Huang et al., 2017) are proposed to promote the benchmarks in this area.

* Corresponding author.
E-mail address: xinmwu@ustc.edu.cn (X.-M. Wu).

https://doi.org/10.1016/j.petsci.2022.01.013

Other fields, such as medicine, meteorology, remote sensing as well as seismic exploration, also benefit from the development and make significant breakthroughs (Zhao, 2018, 2019; Di et al., 2018, 2019; Wu et al., 2019, 2020). Compared with traditional model-driven methods, the advantage of DL mainly lies in that feature learning, extraction and prediction are all included in an end-to-end process, which avoids tedious manual design and achieves less errors by jointly optimizing all parameters. Consequently, many DL-based methods (Das et al., 2018; Wang et al., 2019a,b; Phan and Sen, 2018; Biswas et al., 2019; Alfarraj and AlRegib, 2019b,a; Zheng et al., 2019) are put forward to solve the seismic impedance inversion. The critical technology is Convolutional Neural Network (CNN) whose basic idea is to hierarchically extract features using stacked convolution layers and nonlinear activations. Due to the strong feature representation ability of CNN, DL-based methods are able to more accurately approximate the relationship between the seismograms and impedance sequences and therefore can generate more accurate inversion results. However, in most cases, CNNs are used as black boxes, few work performs in-depth research on how to appropriately design an effective and efficient DL mechanism for the inversion problem.

In this paper, we focus on further research on DL-based inversion methods. The in-fluence of various network hyperparameters and architectures on the inversion results is explored. Specifically, we carry out comparative experiments on three basic hyperparameters (*i.e.*, kernel sizes, number of channels and layers) and two multi-scale architectures, and make comprehensive analysis of the experimental results. In addition, we design a series of methods inspired by perceptual losses (Johnson et al., 2016) and Generative Adversarial Network (GAN) (Goodfellow et al., 2014) to promote the high-frequency information. The contributions of this paper can be summarized as follows:

- We provide important bases for inversion network design by revealing the influence of network hyperparameters and structures on inversion performances.
- We show a clear clue to address the inversion of high-frequency details by borrowing ideas from CV, and achieve the desired results.

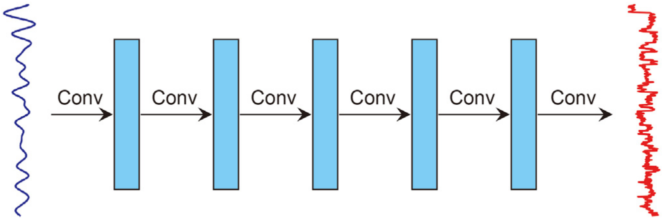

Fig. 1. A conventional CNN which contains 6 convolutional layers, the input is a seismic trace and the output is an impedance sequence.

2. Methods

As a data-driven technology, DL-based methods learn mapping functions from seismograms to impedance sequences. We use the conventional CNN as the baseline, based on which other architectures and techniques are developed step by step to improve the inversion performance.

2.1. Conventional CNN

2.1.1. Architecture

CNN consists of stacked convolutional layers as shown in Fig. 1. A convolutional layer can be defined as follows:

$$\mathbf{x}^l = \sigma\left(\mathbf{x}^{l-1} * \mathbf{w}^l + \mathbf{b}^l\right), \tag{3}$$

where $\mathbf{w}^l$ and $\mathbf{b}^l$ represent the kernel and bias at the l −th layer, and $\mathbf{x}^{l-1}$ and $\mathbf{x}^l$ are the input and output, respectively. In addition, we use Parametric Rectified Linear Unit.

(PReLU) (He et al., 2015) as the nonlinear activation $\sigma(x)$ which is formulated as

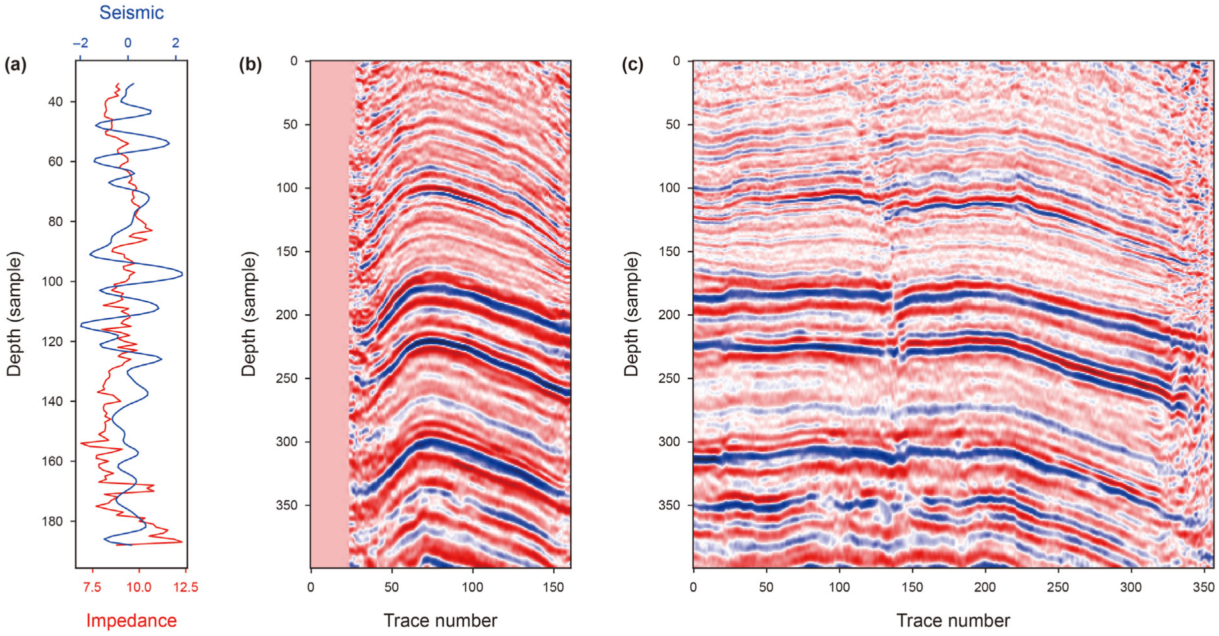

Fig. 2. (**a**) a seismic-impedance training pair. (**b**) a crossline seismic section. (**c**) an inline seismic section.

$$\sigma(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.25x & \text{if } x \le 0 \end{cases}. \tag{4}$$

The most intuitive inversion method is using 1-dimensional CNN as a mapping function from a seismic trace to an impedance sequence. Unlike model-driven methods, training the CNN requires a lot of training data.

2.1.2. Dataset

The seismic data and well logs that we use in this paper are extracted from the freely available Teapot Dome dataset (Anderson, 2009). The seismic data is already converted to depth domain and matched with the well logs. Hundreds of wells are provided along with the seismic data, however, in our experiments of seismic impedance estimation, we choose only the wells that contain both velocity and density logs with significant long depth ranges. Consequently, we choose totally 27 wells and extract the seismic traces near the wells to obtain 27 pairs of impedance sequences and seismograms, in which 22 paris are randomly selected to train our DL networks for the impedance estimation and the remaining 5 pairs are used as validation set. Fig. 2a shows one of the training data pairs where the smooth blue curve represents a seismogram while the red curve with more details denotes a target impedance sequence that we expect to estimate from the seismogram. Fig. 2b and c shows a crossline and inline seismic sections that are extracted from the original 3D seismic volume. These two seismic sections are used in this paper to demonstrate effectiveness of our trained neural networks for the impedance estimation.

2.1.3. Experiments

Hyperparameters have a great impact on the CNN performance. In order to figure out the network design principles for inversion problem, we study three key parameters including kernel size, number of layers and channels which are related to network structure. In the experiments, we adopt the Adadelta (Zeiler, 2012) optimizer with initial learning rate of 0.1, and the learning rate decays to 0.9 times every 50 epochs. The batch size is set to 8. The Mean Squared Error (MSE) is used as loss function, whose formula is as follows:

$$\ell_{\text{MSE}} = \frac{1}{NK}\sum_{n=1}^{N} \|\mathbf{i}_n - f(\mathbf{s}_n)\|_2^2 \tag{5}$$

where $\mathbf{i}_n$ and $f(\mathbf{s}_n)$ are the true and predicted impedances of the n-th training pair, $\|\cdot\|_2$ is the ℓ_2 −norm operator, N is the number of training pairs, K is the signal length. Note that all the input seismic traces and target impedance sequences are normalized by subtracting mean and being divided by standard deviation. All experiments adopt the above settings by default.

First, we fix the number of convolution layers to 5 and channels of each layer to 16, and observe the effect of the kernel size on the inversion result. The kernel size increases from 5 to 23 with steps of 6. As shown in Fig. 4a and b, the larger the kernel size, the better the network convergence. We also observe that larger kernel size brings more high-frequency information as shown in Fig. 3a.

We then adjust the output channel of each layer from 8 to 64, and fix the kernel size to 11 and the number of layers to 5. We observe a trend similar to the kernel size experiment shown in Fig. 4c and d. Networks with more channels can converge to lower training losses, but they converge to the same level of validation loss as the epoch increases. Despite this, their visual effects on the predicted impedance sequences are quite different as shown in

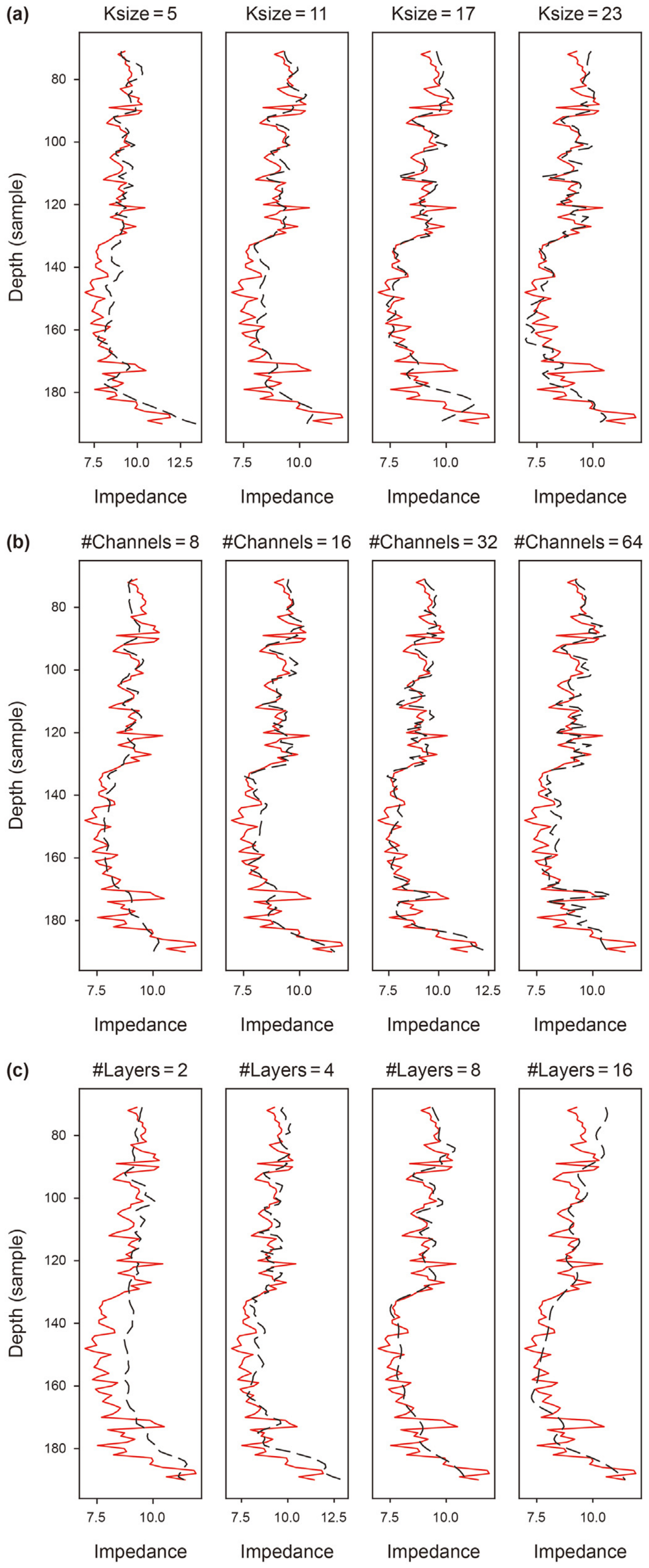

Fig. 3. Inversion results of a seismic trace with different hyperparameters. The red solid curve and black dashed curve represent the true and predicted values respectively. (**a**) results with different kernel sizes. (**b**): results with different number of channels. (**c**) results with different number of layers.

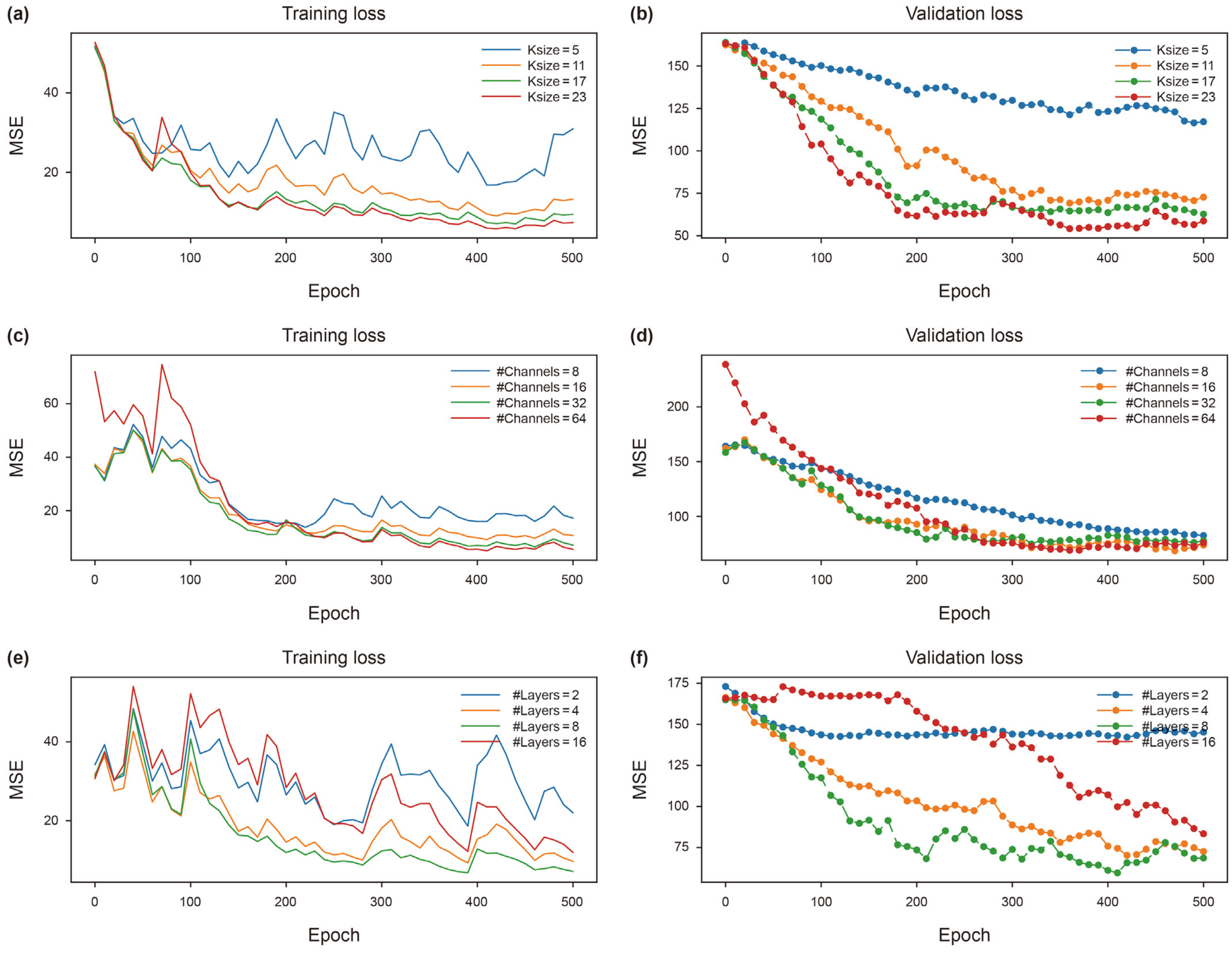

Fig. 4. Training and validation loss curves. Left column: training loss. Right column: validation loss. Top row: loss curves with different kernel sizes. Middle row: loss curves with different number of channels. Bottom row: loss curves with different number of layers.

Fig. 3b. We can see that high frequency is getting richer, as the number of channels increases, especially within the depth window between 160 and 180.

Furthermore, we study the effect of the number of layers on the inversion results. In this experiment, kernel size and channels of each layer are fixed to 11 and 16 respectively, and the number of layers arranges from 2 to 16 in multiples of 2. Fig. 4e and f shows that shallow network with 2 layers is underfitting. When the number of layers increases to 8, the network achieves the best convergence. It is worth noting that the performance of the network with 16 layers has a great degradation. But this is not caused by overfitting since both the training and validation losses are degraded. The reason is that deeper architecture brings huge challenges to the gradient back-propagation (He et al., 2016). From the visual effects in Fig. 3c, the results of $\#layers = 2$ and 16 are underfitting, and result of $\#layers = 4$ yields more details than that of $\#layers\ = 8$.

In general, increasing the complexity of the architecture can improve the network's representation ability, but such improvement is limited. Different hyperparameters may lead to different visual effects. Therefore, it is necessary to consider various factors when designing the network. In order to compare all the methods designed in later chapters, we use a conventional CNN with kernel size of 13, channels of 16 and layers of 6 as a baseline model.

2.2. Multi-scale architecture

A conventional CNN uses a fixed kernel size to extract seismic features at a specific scale, which limits the feature representation. To improve the multi-scale representation capability of the network, we propose two methods in this chapter.

2.2.1. Multi-scale CNN

Inspired by the inception module (Szegedy et al., 2015), a Multi-Scale CNN (MSCNN) is designed as shown in Fig. 5. It is composed of a stacked multi-scale block which is marked by a red frame, where the input feature is parallelly feeded into three conventional layers with different kernel sizes, the three-way output features are then concatenated in channel dimension to form the final output of the multi-scale block. The MSCNN can extract multi-scale features of seismic traces block by block, and uses a normal conventional layer to calculate the impedance in the end of the network.

2.2.2. UNet

UNet (Ronneberger et al., 2015) is another multi-scale architecture which is originally proposed to solve the image segmentation. As shown in Fig. 6, the UNet has two basic components: encoder and decoder. The encoder is similar to the backbone of a

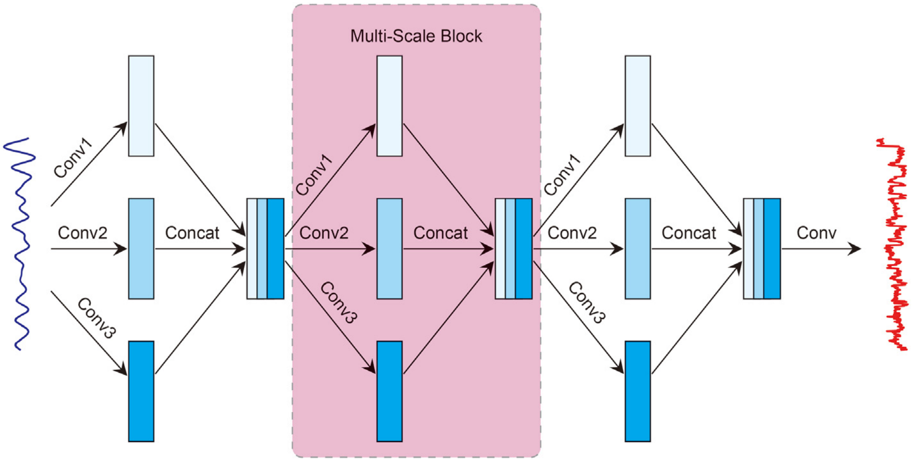

Fig. 5. MSCNN with three stacked multi-scale blocks. Three shades of blue rectangle stand for features extracted by convolutional layers with three different kernel sizes, and the bottleneck containing the three rectangles represents the concatenated features.

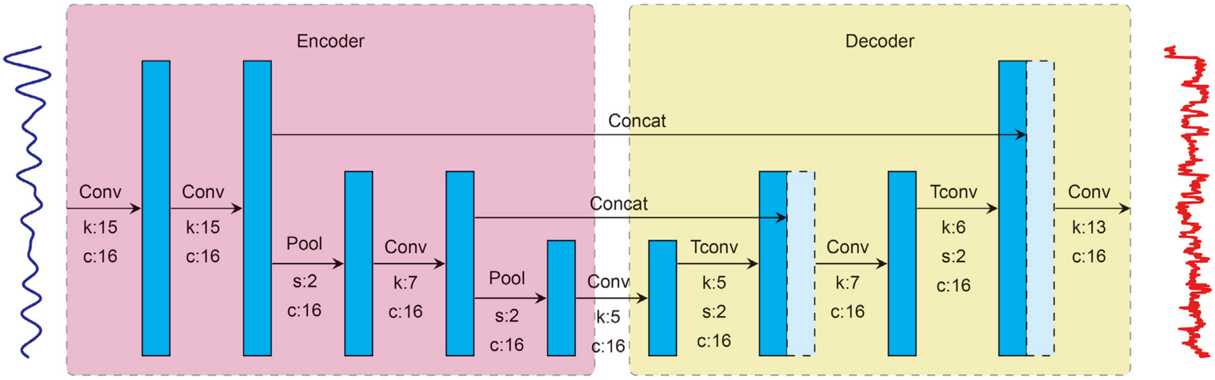

Fig. 6. UNet Architecture. The k, c, s stand for kernel size, number of output channels and stride respectively. The pool and tconv represent max pooling and transpose convolutional layers respectively.

classification network, which consists of conventional layers and max pooling layers. A max pooling layer down samples features with stride of 2 and obtains larger-scale seismic representations. The decoder acts as an upsampling process which uses transpose convolution as the upsampling operator. In the decoder, each upsampled feature is concatenated with the feature of the same scale in the encoder. This concatenation contributes to high-resolution information reconstruction.

2.2.3. *Experiments*

To make a relatively fair comparison of the baseline CNN, MSCNN and UNet, we keep their parameter amounts at the same level. For MSCNN, we use 5 multi-scale blocks whose kernel sizes of three ways are 7, 13 and 19, and each way has 5 output channels. The kernel size of the final conventional layer is 11. For UNet, the hyperparameters are shown in Fig. 5. Parameter amounts of the three methods are given in Table 1.

The same inversion experiments are executed on the three methods. Fig. 7 shows that the three methods converge to almost the same level on the training set, but MSCNN and UNet perform better on the validation set. This means the three networks have the same learning ability since they consist of the similar number of parameters, but multi-scale architectures show better generalization.

The first column of Fig. 8 shows the trace inversion results of the three methods. We can observe that MSCNN and UNet obtain relatively better results than the conventional CNN, especially within the depth window between 140 and 180 where the CNN yields highly smooth predictions. The same observation can be found in the first column of Fig. 13 and Fig. 14, where the layers, especially those thin ones, can be hardly resolved as the conventional CNN yields smooth predictions with limited details in the vertical dimension.

Table 1
Parameter amounts of the three methods.

Parameter amounts			
Method	Baseline	MSCNN	UNet
# of parameters	13814	12151	13600

2.3. *Perceptual loss*

Even though the multi-scale methods achieve better results than the baseline model, they all lose much high-frequency information. This is because they all trained by using MSE loss function which is easy to produce smoothness. From the perspective of the CV, MSE only penalizes Euclidean distance between two images, but ignores the image content. To overcome this problem, we introduce the perceptual loss (Johnson et al., 2016), which measures content similarity, into the networks.

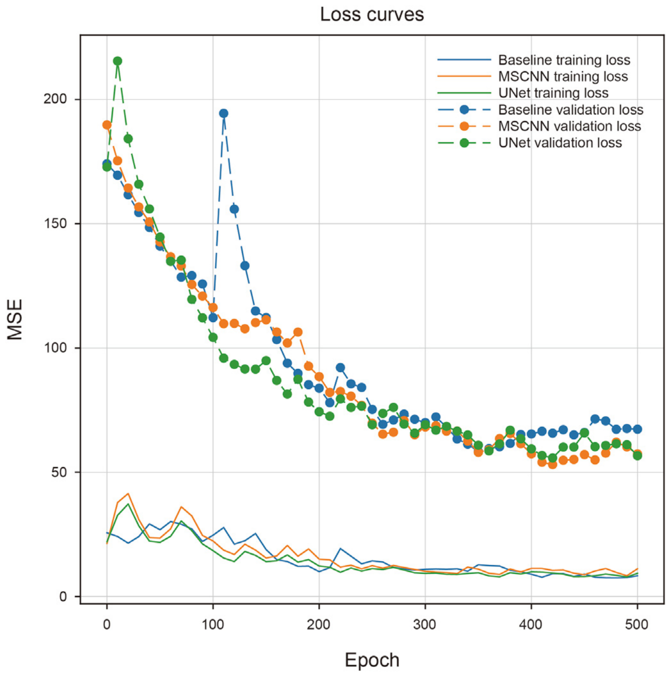

Fig. 7. Training and validation loss curves of the baseline, MSCNN and UNet.

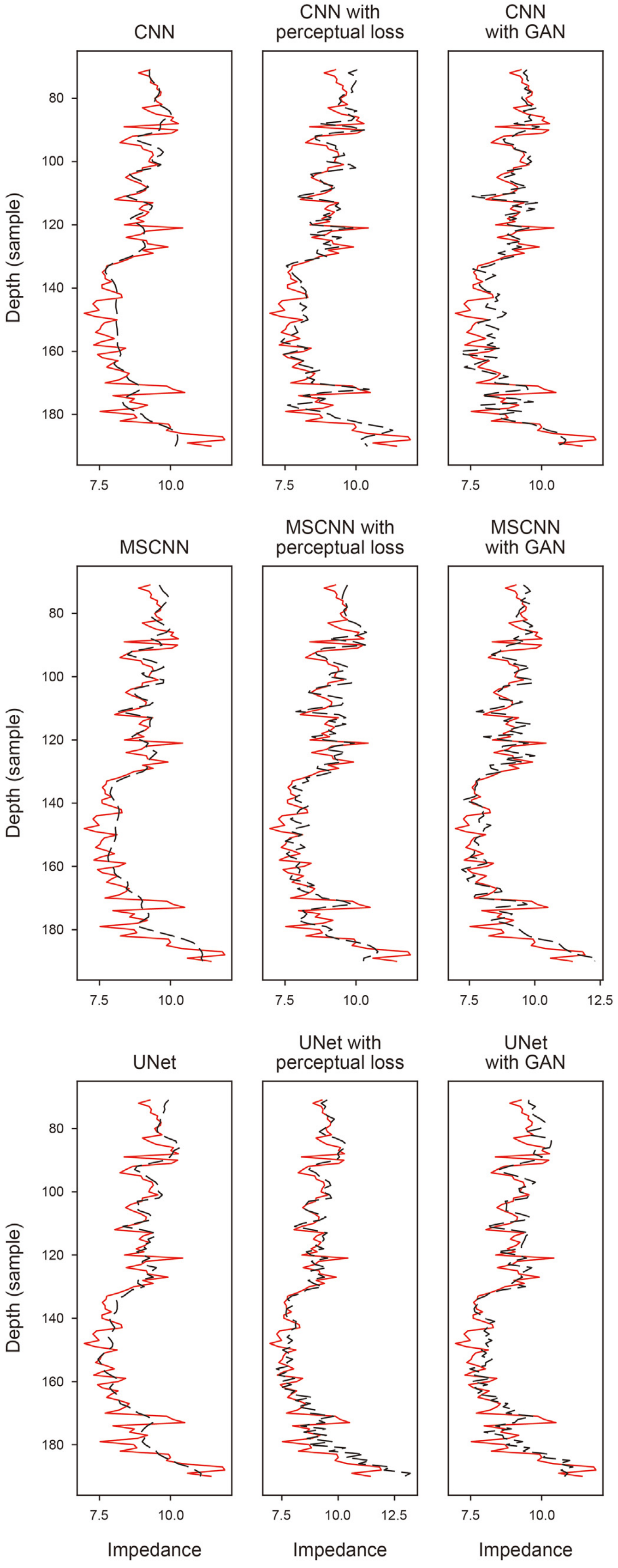

Fig. 8. Trace inversion results by different methods. Top row: The baseline CNN. Middle row: MSCNN. Bottom row: UNet. First column: The pure networks. Second column: Networks with perceptual loss. Third column: Networks with GAN.

2.3.1. *Definition*

Seismic impedance inversion can be considered as a signal reconstruction problem which is similar to the image super-resolution, where it recovers the high-frequency impedance from a low-frequency seismic trace. The perceptual loss states that the reconstructed image should be similar to the ground truth not only in pixels, but also in the feature domain. The common idea is using a pre-trained network, *e.g.*, VGG-16 (Simonyan and Zisserman, 2014), as a loss network to extract the features in different layers, and calculating the Euclidean distance between the true and predicted features to measure the content difference. The perceptual loss experimentally proves the effectiveness of reconstructing high-frequency information.

Inspired by the above ideas, we design a simple autoencoder as the loss network as shown in Fig. 9. The autoencoder has the same structure with the UNet, but there are no links from encoder to decoder. Its hyperparameters of each layer are displayed in the figure. The autoencoder learns a mapping function from the impedance to itself. In other words, the input and output of the network are the same impedance. The main purpose is to extract proper features at different scales as shown in Fig. 9, then we can use the features to calculate the perceptual loss which is defined as follows:

$$\ell_P^l = \frac{1}{NK}\sum_{n=1}^{N}\|\varphi_l(\mathbf{i}_n) - \varphi_l(f(\mathbf{s}_n))\|_2^2 \quad (6)$$

where $\varphi_l(\mathbf{i})$ is the $l-$ th layer's feature of the impedance $\mathbf{i}$ extracted by the loss network.

2.3.2. *Experiments*

We train the autoencoder using the impedance sequences of training set with the same implementation as the previous experiments. Fig. 10 shows the reconstruction performance of the autoencoder, we can see that all the curves are well fitted. It should be noted that we add small Gaussian noise on the training samples

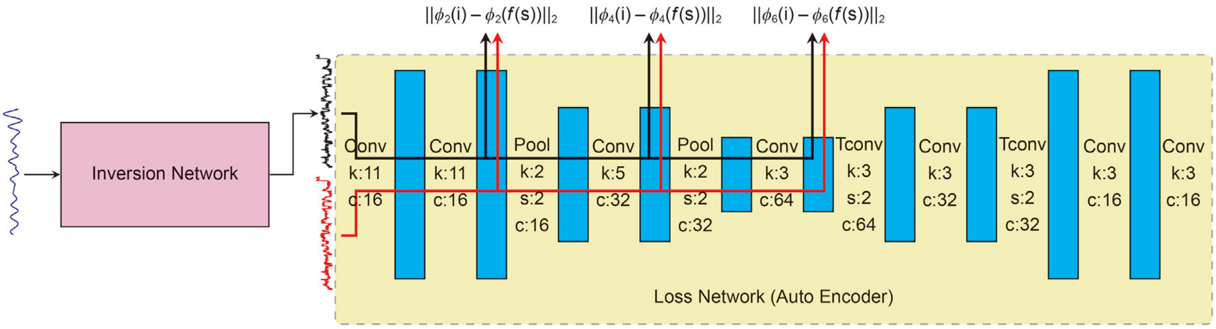

Fig. 9. Training with perceptual loss. The inversion network in the dashed box can be any architecture. The black and red traces are predicted impedance $f(\boldsymbol{s})$ and true impedances $\mathbf{i}$ respectively. Three layers are used as the endpoints to export features which is represented by $\varphi_l(\cdot)$, where the subscript l represents the layer index.

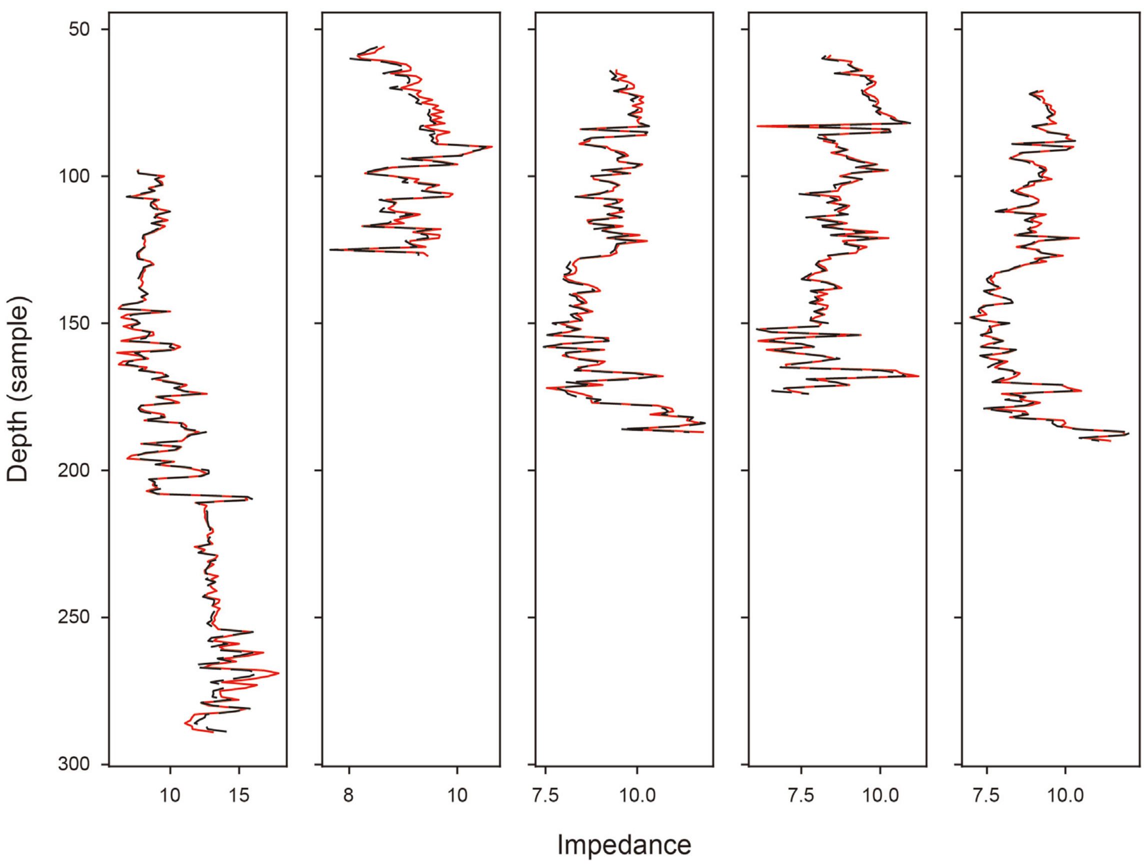

Fig. 10. The recovered impedance sequences by the autoencoder for the validation set. The red and black curves are the input ground truth and output predictions respectively.

at each step to release the overfitting in all experiments, since the size of training set is small.

By combing the MSE and perceptual losses, we can train the networks by the following loss function:

$$\mathscr{L} = \ell_{\text{MSE}} + \lambda_p + \ell_P^l \tag{7}$$

where λ_p is the weight factor of the perceptual loss. We conduct a series of experiments to study how to select λ_p and l. The baseline model is used as the inversion network. First, the λ_p is fixed to 1.0, and we use different endpoints (*i.e.*, $l = 2, 4, 6$) as the feature extraction layers. The results in Fig. 11a shows that when l reaches 6, the ability to reconstruct high-frequency information is limited. The results of $l = 2, 4$ obtain relatively better details for depth around 170. Then we set $l = 2$, and increase λ_p from 0.01 to 10.0 by a factor of 10. We can see from Fig. 11b that as the weight of the perceptual loss increases, more details are reconstructed, but some peak values may exceed the ground truth, *e.g.*, for depth at 100 and 120 with $\lambda_p = 10.0$.

The above observations validate the effectiveness of the perceptual loss. But we need to make a balance between detail reconstruction ability and amplitude fitting stability. We use $l = 4$ and $\lambda_p = 1.0$ as the default setting to make comparisons with other methods. The first two columns of Fig. 8 are the inversion results of the three architectures with or without perceptual loss, which illustrates that perceptual loss improves the reconstruction of high-frequency information a lot. Besides, we have a consistent observation on the inversion planes in Figs. 13 and 14, inversion planes by using perceptual loss have more and clearer horizons than that using the pure MSE loss.

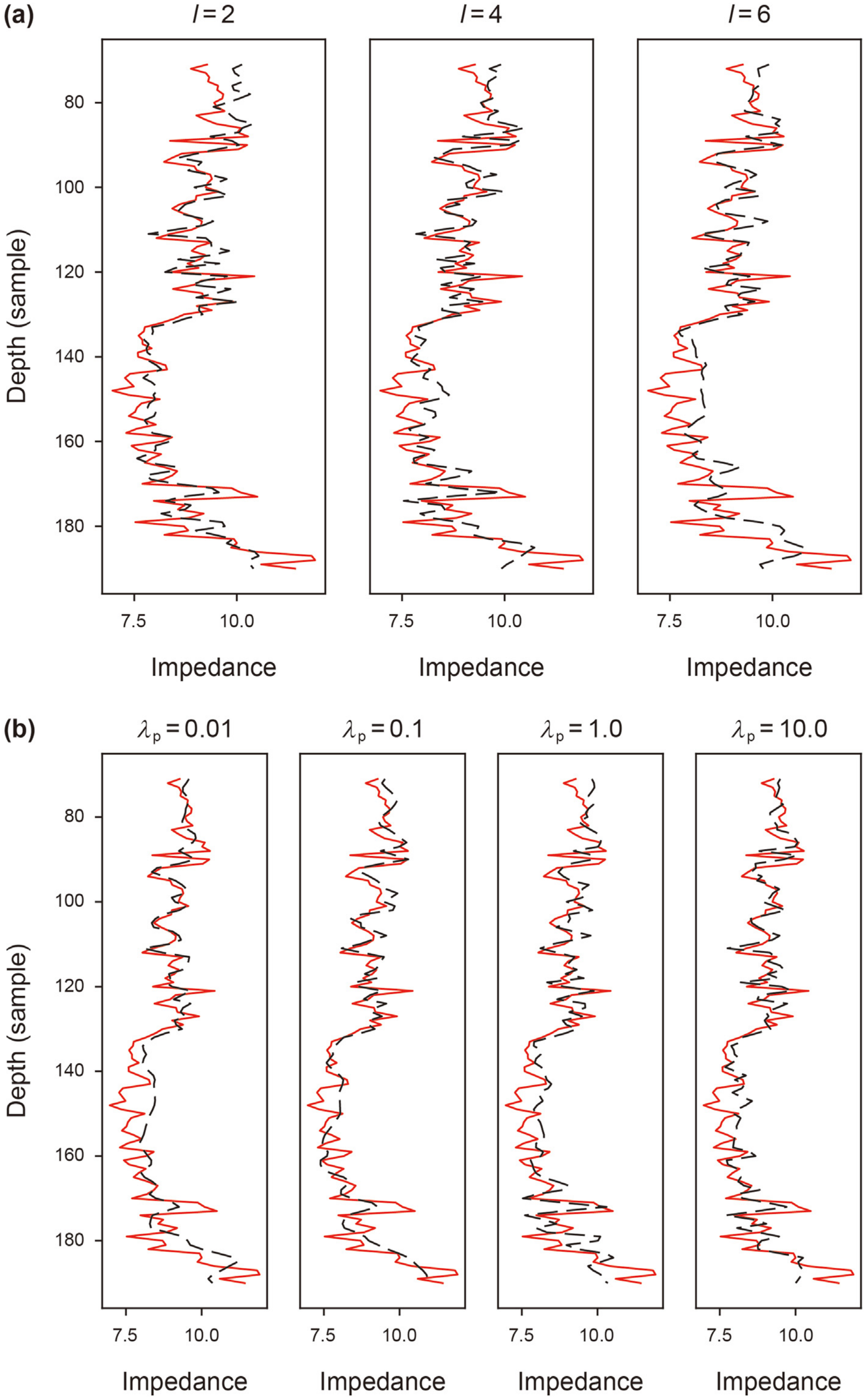

Fig. 11. Inversion results by using different endpoint layers l and weight λ_p.

2.4. GAN

The previous methods focus on the design of backbones and loss functions to achieve desired results, which demonstrates that the developments and techniques in CV field can be used to tackle the seismic inversion problem. Following this clue, we further explore how to estimate more realistic impedance sequences by using GAN which achieves great success in image generation.

2.4.1. Architecture

GAN has two basic modules: generator and discriminator as shown in Fig. 12. The generator can be any inversion network and it aims at fooling the discriminator. The discriminator is a classification network, and it should distinguish between the real traces and traces produced by the generator as much as possible. The two modules form an adversarial mechanism during the training process, as a result the generator produce realistic impedance

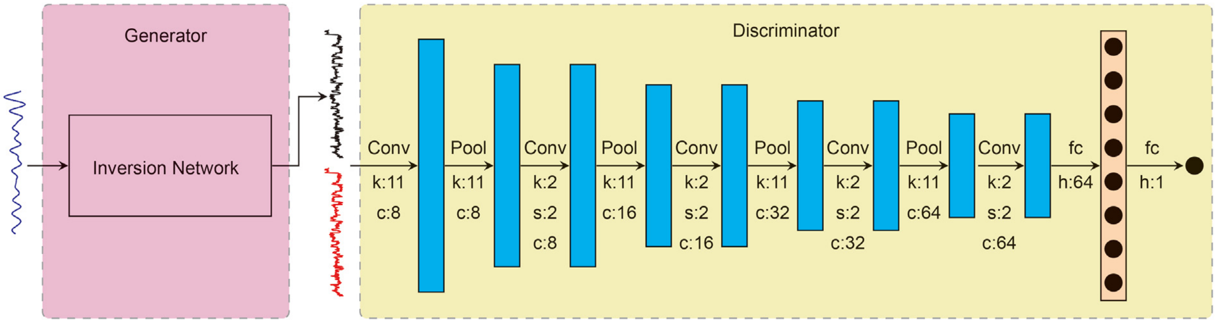

Fig. 12. GAN Architecture. The generator is an inversion network that generates impedance sequences from seismic traces. The discriminator distinguish between generated and real impedance sequences. The hyperparameters of the discriminator are given in the yellow frame, where h and fc stand for number of hidden node and fully connected layer respectively.

sequences, and the discriminator could not distinguish between true and generated sequences.

There is a strong correlation between seismic inversion and image super-resolution problems, as both of them reconstruct high-frequency signals from low-frequency signals. So we refer to Enhanced Super-Resolution GAN (ESRGAN) (Wang et al., 2018) as a reference to design an inversion GAN. The discriminator architecture and hyperparameters are given in Fig. 12. The final fully connected layer has only one hidden node as it is a binary classification network. Different from the standard classification, we use the Relativistic average Discriminator (RaD) (Jolicoeur-Martineau, 2019) to predict the realistic degree of generated impedance relative to true impedance. The RaD is formulated as follows:

$$D(\mathbf{i}_r,\mathbf{i}_g)=\delta\Big(f_D(\mathbf{i}_r)-\mathbb{E}_{\mathbf{i}_g}\left[f_D(\mathbf{i}_g)\right]\Big), \tag{8}$$

where $\mathbf{i}_r$ and $\mathbf{i}_g$ are the real and generated impedances, $f_D(\mathbf{i})$ represents the output of final fully connected layer for impedance $\mathbf{i}$, $\mathbb{E}_{\mathbf{i}_g}[\,\boldsymbol{\cdot}\,]$ represents the average value of the generated impedances in a mini-batch. $\delta(\,\boldsymbol{\cdot}\,)$ is the sigmoid function. An ideal discriminator makes $D_{\mathrm{Ra}}(\mathbf{i}_r,\mathbf{i}_g)=1$ and $D_{\mathrm{Ra}}(\mathbf{i}_g,\mathbf{i}_r)=0$. Then the discriminator loss is defined as:

$$\mathscr{L}_\mathrm{D}=-\mathbb{E}_{\mathbf{i}_r}\left[\log(D(\mathbf{i}_r,\mathbf{i}_g))\right]-\mathbb{E}_{\mathbf{i}_g}\left[1-\log(D(\mathbf{i}_g,\mathbf{i}_r))\right]. \tag{9}$$

The adversarial loss of generator is defined as:

$$\ell_\mathrm{GAN}=-\mathbb{E}_{\mathbf{i}_r}\left[1-\log(D(\mathbf{i}_r,\mathbf{i}_g))\right]-\mathbb{E}_{\mathbf{i}_g}\left[\log(D(\mathbf{i}_g,\mathbf{i}_r))\right]. \tag{10}$$

The total loss for generator is then defined as follows:

$$\mathscr{L}_\mathrm{G}=\ell_\mathrm{MSE}+\lambda_p\ell_P^l+\lambda_g\ell_\mathrm{GAN} \tag{11}$$

where λ_p and λ_g are weight factors for perceptual and adversarial losses respectively. In the training process, the generator and discriminator are alternately updated by minimizing $\mathscr{L}_\mathrm{G}$ and $\mathscr{L}_\mathrm{D}$.

2.4.2. Experiments

In order to speed up convergence, we first train an inversion network with MSE loss using the default setting, and then use the pre-trained model as an initial generator. The parameters λ_p, λ_g, l are empirically set to 1.0, 7e-3, 4. The initial learning rates of generator and discriminator are 0.7 and 0.9, and they decays by factors 0.95 with decay steps of 50 and 100 respectively. The GAN is trained for 1000 epochs. We adopt the three networks, *i.e.*, CNN, MSCNN, UNet, as the generator of the GAN. The trace inversion results are shown in Fig. 8, we can see that GANs recover more details than pure networks and they have the similar visual effect to the networks with perceptual loss. But according to the plane inversion results in Figs. 13 and 14, GANs generate finer layers than other two methods especially within the depth window between 50 and 250. In addition, GANs produce some dark layers (with low impedances) near the depth of 200 which can not be observed in the results by the other methods.

3. Discussion

The hyperparameter experiments on the conventional CNN demonstrate that networks with more parameters show stronger fitting ability from the two perspectives of the number of channels and kernel size. But this promotion tends to disappear as the amount of parameters increase as shown in Fig. 4a, b, 4c, and 4d. This is because the ability of each network to fit small dataset is saturated. From the layer number perspective, as shown in Fig. 4e and f, excessive increase in the number of layers leads to degradation of convergence. A common view is that deeper network makes gradient back propagation difficult, and even produce the vanishing gradient problem (He et al., 2016). Fig. 3a, b and 3c indicate that the curve fitting performance varies a lot with the hyperparameters, which is mainly reflected in the reconstruction of high-frequency details.

Using conventional CNN to solve the inversion problem is an intuitive way. However it is hard to choose the proper hyperparameters, since the inversion result is hyperparameter-sensitive. Multi-scale architecture can extract features at different scales and therefore is able to recover more details than the conventional CNN with the same number of parameters. As a result, multi-scale architecture relieves the cost of hyperparameter selection. But we note that even though the three methods converge to the same level as shown in Fig. 7, they yield quite different visual effects in the inversion sections as shown in the first columns in Figs. 13 and 14. Overall, multi-scale inversion sections show more thin layers, but MSCNN and UNet produce different high impedance areas. Therefore, it is important to adopt appropriate architectures.

From the inversion experiments, the key point is the reconstruction of the high-frequency information. In the CV field, the MSE loss is proven to produce smoothness, which can be improved by the perceptual loss in Equation (6). In order to build an impedance feature space which is used to calculated the perceptual loss, we design an autoencoder that learns a mapping function from impedance to itself as shown in Fig. 9, and then extract the features by the endpoints of the autoencoder. Figs. 8, 13 and 14 show that perceptual loss provides a great contribution to reconstructing high-frequency information. On the other hand, the endpoint layer l and weight factor λ_p in Equation (6) also impose an effect on the inversion results as shown in Fig. 11a and b. So a trade-off must be

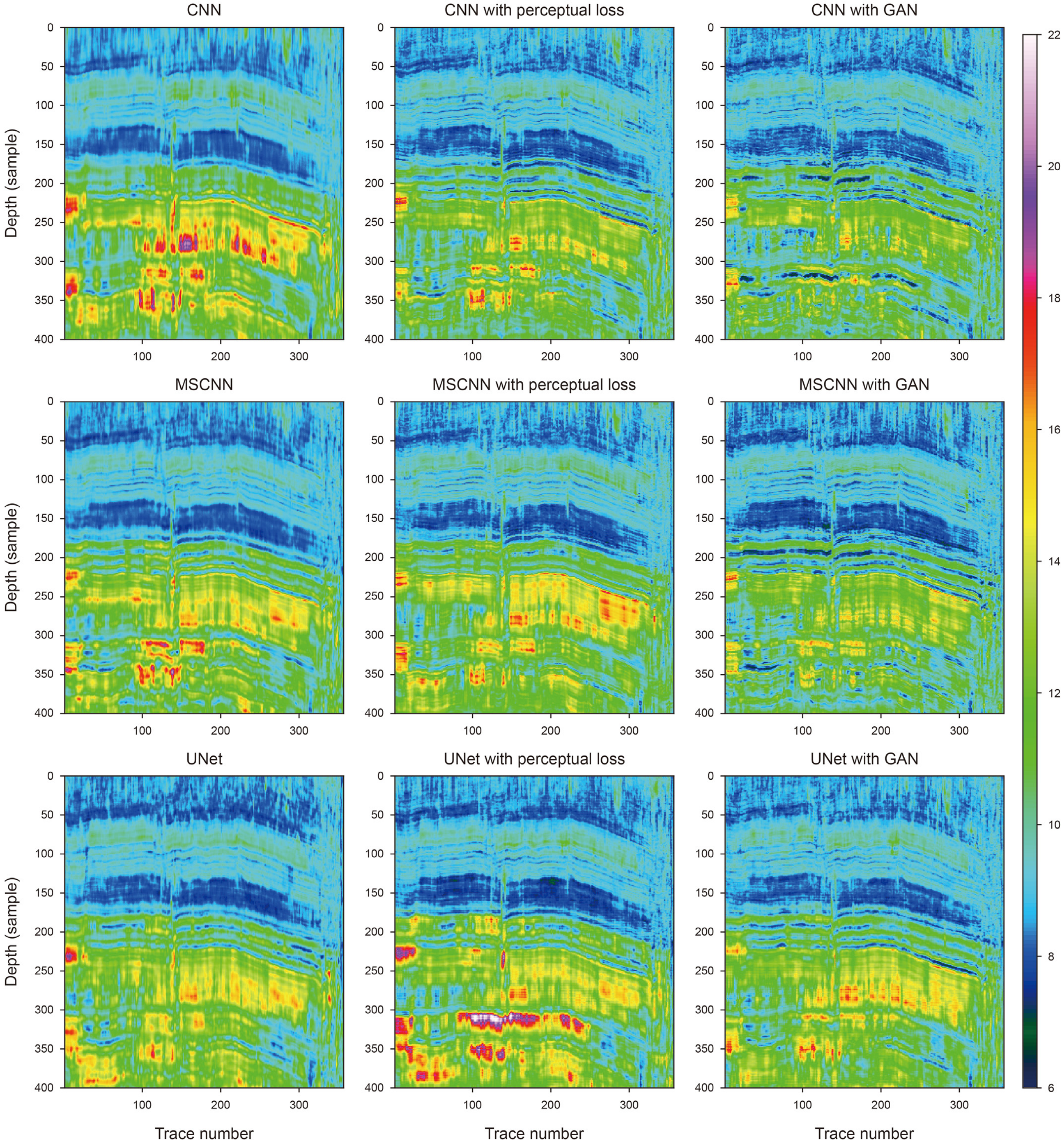

Fig. 13. Inline inversion results by different methods. Top row: The baseline CNN. Middle row: MSCNN. Bottom row: UNet. First column: The pure networks. Second column: Networks with perceptual loss. Third column: Networks with GAN.

made between detail reconstructing and fitting stability. The GAN experiments demonstrate that adversarial training mechanism further promotes the reconstruction of details, and it generates finer layers as shown in Figs. 13 and 14. Besides, some dark layers with low impedance values appear in the GAN inversion results, which may indicate its ability in recovering high-frequency information.

DL-based methods achieve promising results, but also show some limitations. Different architectures produce various vision effects, which may bring confusion to practical applications. However, there is no objective evaluation index to indicate which network should be used. The widely used MSE can provide a reference to the fitting performance, but it produces much smoothness. So it is necessary to build an evaluation function

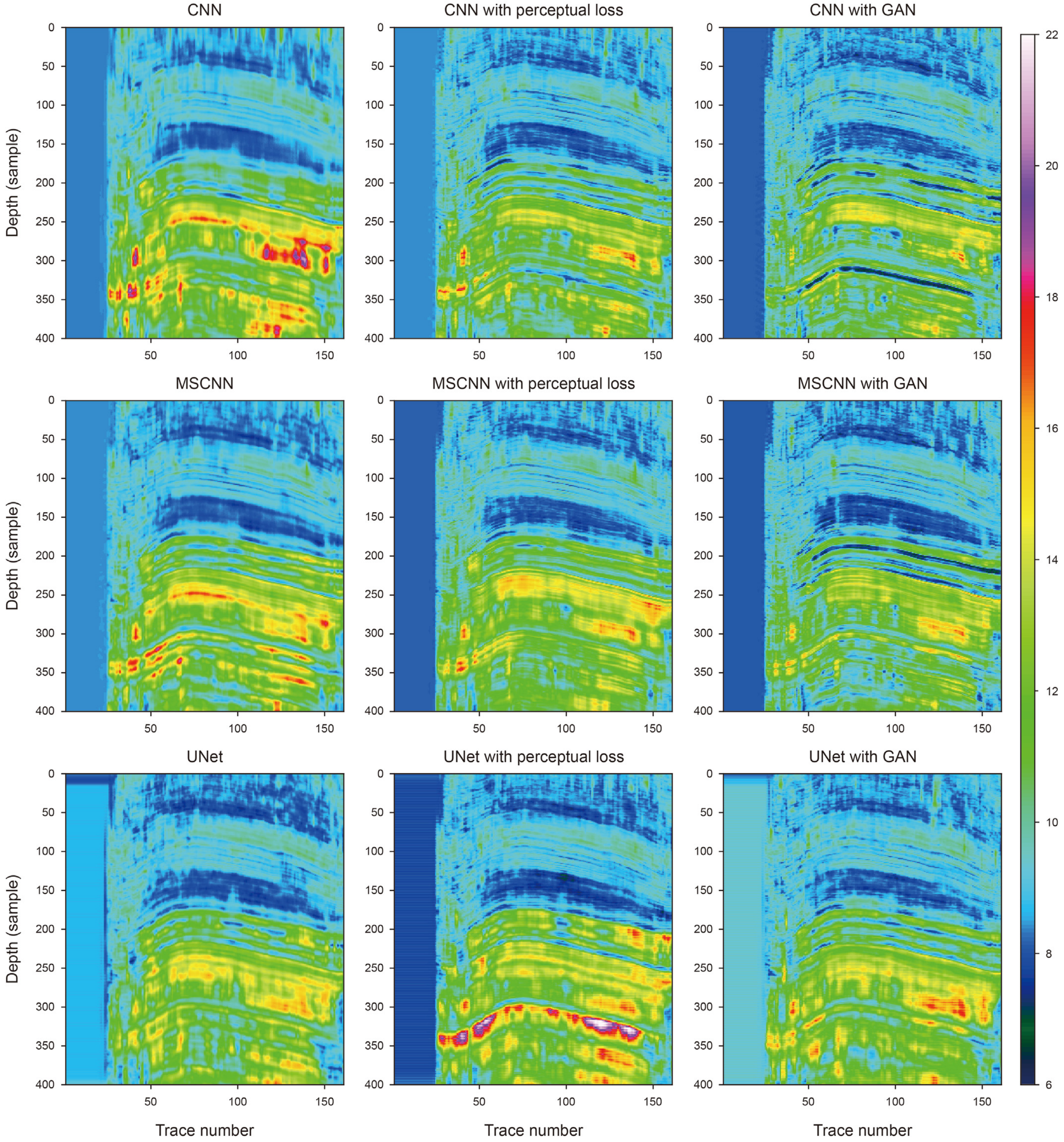

Fig. 14. Crossline inversion results by different methods. Top row: The baseline CNN. Middle row: MSCNN. Bottom row: UNet. First column: The pure networks. Second column: Networks with perceptual loss. Third column: Networks with GAN.

related to structure and content of the impedance. The other obvious problem is the lack of training data. In practice, the number of well logs is highly limited, which often results in network overfitting. Using some tricks, such as adding Gaussian noise, cannot completely avoid the risk of overfitting. One way to address this is building realistic structure models (Wu et al., 2020) to simulate more seismic and impedance pairs. The other meaningful way is introducing the physical mechanism of Equation (1) into the network architecture to make full use of seismic traces that do not correspond to any true impedance sequences, which performs a semi-supervised learning.

4. Conclusion

This paper comprehensively studies the DL-based methods for seismic impedance inversion problem. A series of networks are designed to improve the reconstruction of high-frequency information. Through experiments, we reveal the influence of the network hyperparameter and architecture on the inversion performance. The difference between conventional CNN and multiscale architecture in convergence, trace fitting and vision effect are well studied. Inspired by the developments in the CV field, we adopt perceptual loss and GAN mechanism which are proven to be effective for enhancing high-frequency details. In spite of the success of DL-based methods, they still show the aforementioned limitations of objective evaluation index and training data. We plan to solve these two issues in the future.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant No. 42050104.

References

Alfarraj, M., AlRegib, G., 2019a. Semi-supervised learning for acoustic impedance inversion. In: SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, pp. 2298–2302. https://doi.org/10.1190/segam2019-3215902.1.

Alfarraj, Motaz, AlRegib, Ghassan, 2019b. Semisupervised sequence modeling for elastic impedance inversion. SE237–SE249 Interpre- tation 7. https://doi.org/10.1190/INT-2018-0250.1.

Anderson, T., 2009. History of Geologic Investigations and Oil Operations at Teapot Dome: Presented at the Wyoming: Presented at the 2009 AAPG Annual Convention. https://www.searchanddiscovery.com/pdfz/documents/2013/20215anderson/ndx_anderson.pdf.html.

Biswas, R., Sen, M.K., Das, V., Mukerji, T., 2019. Prestack and poststack inversion using a physics-guided convolutional neural network. SE161–SE174 Interpretation 7. https://doi.org/10.1190/INT-2018-0236.1.

Das, V., Pollack, A., Wollner, U., Mukerji, T., 2018. In: Convolutional Neural Network for Seismic Impedance Inversion: 2071–2075. https://doi.org/10.1190/geo2018-0838.1.

Di, H., Gao, D., AlRegib, G., 2019. Developing a seismic texture analysis neural network for machine-aided seismic pattern recognition and classification. Geophys. J. Int. 218, 1262–1275. https://doi.org/10.1093/gji/ggz226.

Di, H., Wang, Z., AlRegib, G., 2018. Deep Convolutional Neural Networks for Seismic Salt-Body Delineation. Presented at the AAPG Annual Convention and Exhibition. https://doi.org/10.1306/70630Di2018.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems, 27. Curran Associates, Inc., pp. 2672–2680. In: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human- Level Performance on Imagenet Classification. Presented at the The IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV.2015.123.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, Sun, Jian, 2016. Deep Residual Learning for Image Recognition: Presented at the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.90.

Hu, W., Abubakar, A., Habashy, T.M., 2009. Joint electromagnetic and seismic inversion using structural constraints. R99–R109 Geophysics 74. https://doi.org/10.1190/1.3246586.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely Connected Convolutional Networks: Presented at the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.243.

Jackson, D.D., 1972. Interpretation of inaccurate, insufficient and inconsistent data. Geo- Physical J. Int. 28, 97–109. https://doi.org/10.1111/j.1365-246X.1972.tb06115.x.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual Losses for Real-Time Style Transfer and Super-resolution: Computer Vision – ECCV 2016. Springer International Publishing. https://doi.org/10.1007/978-3-319-46475-6_43, 694–711.

Jolicoeur-Martineau, A., 2019. The Relativistic Discriminator: a Key Element Missing from Standard GAN: Presented at the International Conference on Learning Representations. https://arxiv.org/abs/1807.00734.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Curran Associates, Inc. In: Advances in Neural Information Processing Systems 25, 1097–1105. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Ma, Y., Hale, D., Gong, B., Meng, Z., 2012. Image-guided sparse-model full waveform inversion. Geophysics 77, R189–R198. https://doi.org/10.1190/geo2011-0395.1.

Phan, S., Sen, M.K., 2018. Hopfield networks for high-resolution prestack seismic inversion. In: SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists, pp. 526–530. https://doi.org/10.1190/segam2018-2996244.1.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional Networks for Biomed- Ical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28, 234–241.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition cite arxiv:1409.1556). https://arxiv.org/abs/1409.1556.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going Deeper with Convolutions: Presented at the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2015.7298594.

Wang, K., Bandura, L., Bevc, D., Cheng, S., DiSiena, J., Halpert, A., Osypov, K., Power, B., Xu, E., 2019a. In: End-to-end Deep Neural Network for Seismic Inversion. https://doi.org/10.1190/segam2019-3216464.1.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C., 2018. Esrgan: Enhanced Super-resolution Generative Adversarial Networks: Presented at the the European Conference on Computer Vision (ECCV) Workshops. https://doi.org/10.1007/978-3-030-11021-5_5.

Wang, Y., Ge, Q., Lu, W., Yan, X., 2019b, 2498–2502. In: Seismic Impedance Inversion Based on Cycle-Consistent Generative Adversarial Network. https://doi.org/10.1016/j.petsci.2021.09.038.

Wu, X., 2017. Structure-, stratigraphy- and fault-guided regularization in geophysical in- version. Geophys. J. Int. 210, 184–195. https://doi.org/10.1093/gji/ggx150.

Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., Caumon, G., 2020. Building realistic struc- ture models to train convolutional neural networks for seismic structural interpretation. Geophysics 85, WA27–WA39. https://doi.org/10.1190/geo2019-0375.1.

Wu, X., Liang, L., Shi, Y., Fomel, S., 2019. Faultseg3d: using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation. GEO- PHYSICS 84, IM35–IM45. https://doi.org/10.1190/geo2018-0646.1.

Zeiler, M., 2012. Adadelta: an Adaptive Learning Rate Method: **1212**. https://arxiv.org/abs/1212.5701.

Zhang, J., Revil, A., 2015. 2d joint inversion of geophysical data using petrophysical clustering and facies deformation. Geophysics 80, M69–M88. https://doi.org/10.1190/geo2015-0147.1.

Zhang, R., Castagna, J., 2011. Seismic sparse-layer reflectivity inversion using basis pursuit decomposition. Geophysics 76, R147–R158. https://doi.org/10.1190/geo2011-0103.1.

Zhang, R., Sen, M.K., Srinivasan, S., 2013. A prestack basis pursuit seismic inversion. Geophysics 78, R1–R11. https://doi.org/10.1190/geo2011-0502.1.

Zhao, T., 2018. Seismic facies classification using different deep convolutional neural net- works, 2046–2050. In: SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists. https://doi.org/10.1190/segam2018-2997085.1.

Zhao, Tao, 2019. 3d convolutional neural networks for efficient fault detection and orientation estimation. In: SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, pp. 2418–2422. https://doi.org/10.1190/segam2019-3216307.1.

Zheng, Y., Zhang, Q., Yusifov, A., Shi, Y., 2019. Applications of supervised deep learning for seismic interpretation and inversion. Lead. Edge 38, 526–533. https://doi.org/10.1190/tle38070526.1.

Zhou, J., Revil, A., Jardani, A., 2016. Stochastic structure-constrained image-guided inversion of geophysical data. Geophysics 81, E89–E101. https://doi.org/10.1190/geo2014-0569.1.