Article

Three-Dimensional Velocity Field Interpolation Based on Attention Mechanism

Xingmiao Yao [1,2,*], Mengling Cui [1], Lian Wang [1], Yangsiwei Li [1], Cheng Zhou [1], Mingjun Su [3] and Guangmin Hu [1,2]

[1] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China; cml@std.uestc.edu.cn (M.C.); cheng_zhou@std.uestc.edu.cn (C.Z.); hgm@uestc.edu.cn (G.H.)

[2] Center for information Geoscience, University of Electronic Science and Technology of China, Chengdu 611731, China

[3] Research Institute of Petroleum Exploration and Development-Northwest (NWGI), PetroChina, Lanzhou 730020, China

[*] Correspondence: yxm@uestc.edu.cn

Abstract: The establishment of a three-dimensional velocity field is an essential step in seismic exploration, playing a crucial role in understanding complex underground geological structures. Accurate 3D velocity fields are significant for seismic imaging, observation system design, precise positioning of underground geological targets, structural interpretation, and reservoir prediction. Therefore, obtaining an accurate 3D velocity field is a focus and challenge in this field of study. To achieve intelligent interpolation of the 3D velocity field more accurately, we have built a network model based on the attention mechanism, JointA 3DUnet. Based on the traditional U-Net, we have added triple attention blocks and channel attention blocks to enhance dimension information interaction, while adapting to the different changes of geoscience data in horizontal and vertical directions. Moreover, the network also incorporates dilated convolution to enlarge the receptive field. During the training process, we introduced transfer learning to further enhance the network's performance for interpolation tasks. At the same time, our method is a deep learning interpolation algorithm based on an unsupervised model. It does not require a training set and learns information solely from the input data, automatically interpolating the missing velocity data at the missing positions. We tested our method on both synthetic and real data. The results show that, compared with traditional intelligent interpolation methods, our approach can effectively interpolate the three-dimensional velocity field. The SNR increased to 36.22 dB, and the pointwise relative error decreased to 0.89%.

Keywords: three-dimensional interpolation; attention mechanism; transfer learning; dilated convolution

1. Introduction

The establishment of a three-dimensional velocity field is a significant step in seismic exploration [1], and essential for understanding complex underground geological structures. An accurate velocity model is vital for most seismic imaging methods. The precise 3D velocity field model is a prerequisite for reverse time migration and other seismic imaging technologies. Moreover, it is crucial for observation system design, precise positioning of underground geological target bodies, structural interpretation, and reservoir prediction. The accuracy of velocity directly affects all aspects of seismic exploration and the final results. However, due to economic and natural constraints, only a small amount of real data can be obtained [2], and the complete 3D velocity field is often established by interpolating sparse data.

Traditional three-dimensional spatial interpolation methods rely on inferring the relationships between known data points to generate estimates for the entire spatial domain. For example, R. Dumas [3] and others successfully reconstructed soft-tissue pseudoshadow

data using three-dimensional linear interpolation methods, providing additional geometric information helpful for subsequent medical research. Qinghang He [4] and others applied inverse distance weighting (IDW) to three-dimensional fluorescence spectroscopy data interpolation, expressing more details of the fluorescence spectrum. J. Guo [5] and others applied three-dimensional spline interpolation to transmissivity estimation in the South China Sea, obtaining reasonable and accurate results. In the field of Earth Sciences, Ray Abma and Nurul Kabir [6] and others used the convex set projection (POCS) algorithm of Fourier transform as a simple iterative method for interpolating irregular filling grids of seismic data, producing high-quality results. Parisa Bagheripour [7] and others used the ordinary Kriging method to double the resolution of nuclear magnetic resonance (NMR) well logging data, enabling more accurate predictions of the free fluid porosity and permeability of carbonate reservoir rocks in the South Pars gas field, facilitating a better understanding of the reservoir. Ilozobhie A. J. [8] and others used Kriging and co-Kriging methods to analyze well logging data from five oil wells, estimating the porosity in the Bornu Basin area of northeastern Nigeria, laying the foundation for subsequent exploration strategies. Traditional interpolation methods usually rely on mathematical functions for fitting or interpolation functions for data completion, which may not effectively capture complex geological or geographical features in some scenarios, especially when dealing with the nonlinearity of real underground scenarios, such as heterogeneous underground media or terrain undulations. Therefore, more effective techniques are often needed for three-dimensional spatial interpolation.

As the demand continues to grow, traditional 3D interpolation methods are gradually showing their limitations when facing complex, large-scale, or high-uncertainty data. Under such circumstances, introducing deep learning methods brings new possibilities for 3D interpolation. The application of three-dimensional interpolation technology is also used in medical imaging [9–14], hydrological analysis [15], atmospheric sciences [16], hydrodynamics [17], virtual reality [18], and other fields. Mikhailiuk A et al. [19] first achieved the restoration of the full picture of seismic data from 20% of actual data through a deep autoencoder. Wang G [20], Liu Z [21], Zhao T [22], Fisher P F [23], and others have also used three-dimensional interpolation methods to carry out related work in the field of geoscience. Araya Polo [24] and others used deep neural network (DNN) and feature extraction steps to interpolate and establish velocity models from seismic trace sets, reducing computational costs. Wang et al. [25] are dedicated to solving the same problem, using an improved fully convolutional network (FCN) [26] with fewer parameters, thus more efficient than DNN. Kazei et al. [27] used a large synthetic dataset and established velocity models directly from seismic trace sets interpolation based on the VGG (Visual Geometry Group) [28] network. These deep learning 3D interpolation methods typically rely on standard convolution operations with fixed kernel sizes and limited receptive fields, meaning the model can only capture local geological features and is unable to understand global spatial relationships effectively, presenting limitations for handling large-scale geological data.

Most of the current deep learning interpolation methods are based on the training of large datasets or specialized training datasets, which is contrary to the reality of interpolation tasks. In real scenarios, the acquisition of velocity field data will consume a great amount of manpower and financial resources and is easily hindered by natural conditions, making the data very sparse [29]. Secondly, the network finds it difficult to transfer in the face of different underground features in various scenarios. More importantly, some methods are based on two-dimensional spatial processing, which does not conform to the three-dimensional features of underground data, and will also lose some inter-dimensional related information. For this reason, we have built a 3D velocity field intelligent interpolation method based on a hybrid triple attention mechanism, the JointA 3DUnet network model. It selectively enhances information exchange between dimensions by introducing triple and channel attention mechanisms based on the traditional U-Net structure. In addition, dilated convolution is introduced to increase the receptive field of the convolution

kernel. Through transfer learning, we have enhanced the performance of the algorithm in interpolation tasks. It is worth noting that our method operates in an unsupervised mode, learning only from input data, and can effectively interpolate missing velocity field data. We have verified our method using synthetic and real data. The JointA 3DUnet network model demonstrates superiority in 3D velocity field interpolation, showing better performance compared with traditional intelligent interpolation methods.

2. Materials and Methods

In this section, we will provide a detailed introduction to our intelligent interpolation method, including the network architecture we use, detailed information about the related attention modules, and the training method of the network.

2.1. Network Architecture

The three-dimensional velocity field interpolation of geoscientific data is different from three-dimensional spatial interpolation in other fields due to the multidimensional correlation, irregular spatial distribution, and the large-scale nature of the interpolated data in geoscientific datasets. Geoscientific data typically involve multiple dimensions with complex interrelationships among them. Additionally, the irregularity in the horizontal and vertical distribution of geoscientific data, caused by geological structures and features, implies that the velocity exhibits inconsistent variations horizontally and vertically.

The U-Net [30], as a type of encoder–decoder structure, was initially applied in the medical image segmentation domain. Due to its outstanding performance, it was later extended to various fields such as semantic segmentation, remote sensing image processing, fault detection, data interpolation, and more, boasting a wide range of applications. To realize the interpolation of geoscience-related three-dimensional velocity fields, we introduced a triple attention mechanism combined with a channel attention mechanism, as shown in Figure 1. By adding the TA Block and CA Block attention modules, based on the U-Net framework, the two mechanisms are interlinked to allocate weights dynamically to information from different dimensions. This enables the network to understand the relationships between various dimensions better. Furthermore, before interpolating geoscience data, the distribution is irregular, and, after interpolation, it typically exhibits large scale. Therefore, performing three-dimensional interpolation requires not only efficient computational and storage resources but also adjustments to the network training strategy to enhance processing efficiency and accuracy. In this paper, we update network parameters using a transfer learning training method, effectively addressing the challenges of large-scale and unevenly distributed geoscience data. By dividing geoscience data into multiple blocks or subsets and prioritizing training on areas with dense data points, the inaccuracies in training sparse data regions are reduced, making the model easier to handle.
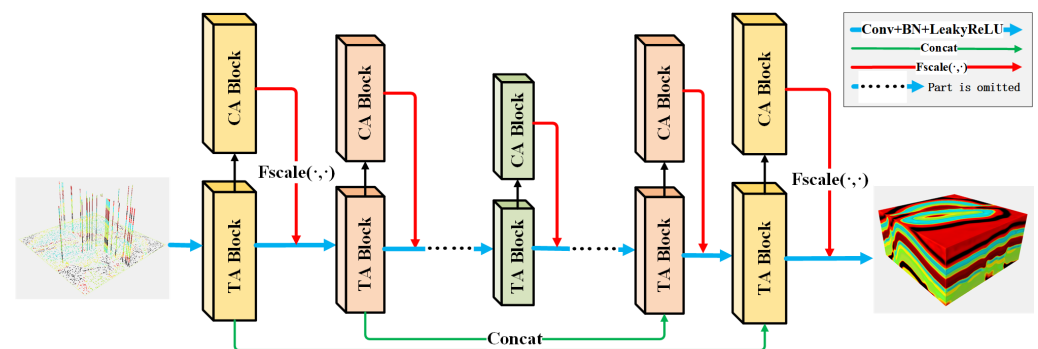
Figure 1. Joint attention mechanism architecture diagram.

TA Block: in extracting feature maps from geoscience data using deep learning models, to establish weight relationships between different dimensions explicitly and focus effectively on the connection between different dimensions in the data, we introduce a triple

attention mechanism. Through the replication and rotation of data, we realize attention interaction between inline, xline, and depth dimensions, making information interaction between different dimensions more compact, and enhancing the model's understanding and expression of spatial information. The architecture of the TA Block will be detailed in Section 2.2.1.

CA Block: during the interaction among inline, xline, and depth dimensions, we interact two dimensions at a time, resulting in three joint channels: inline–xline, inline–depth, and xline–depth. To conform to the irregular lateral and longitudinal distribution of geoscience data more, emphasizing inline–xline attention while weakening inline-depth and xline–depth attention, we introduce the channel attention mechanism. Through the squeeze and excitation modules, we learn the weight of each joint channel, allowing the model to adapt better to the differences and correlations between different channel data. The architecture of the CA Block will be detailed in Section 2.2.2.

2.2. Joint Attention Mechanism

In our network architecture, we place particular emphasis on the correlations among different dimensions and the inconsistency in the horizontal and vertical variations of geoscientific data. Traditional deep learning methods often overlook these interactions and data characteristics among dimensions, leading to insufficient feature extraction and lower accuracy when dealing with multidimensional geoscientific data. Attention mechanisms, on the other hand, can effectively overcome this limitation. There has been substantial research on convolutional attention mechanisms [31–34], and, in this paper, we introduce the combined attention mechanism consisting of the TA Block and CA Block.

2.2.1. TA Block

The three-dimensional velocity field data is composed of inline, xline, and depth dimensions. Effectively extracting key information from these three dimensions and allowing interactions between each dimension are crucial for interpolation tasks. Therefore, we introduce a triple attention mechanism [35]. This mechanism aims to realize the interaction between the three dimensions, namely the H, C, and W dimensions. It is implemented through rotation and parallel computing without reducing the data dimensions. Each branch first rotates a C × H × W input tensor by 90°, changing its shape to W × H × C. Then, through a pooling module, it is reduced to 2 × H × C, and then passes through a convolution layer and batch normalization layer, ultimately becoming a 1 × H × C intermediate representation. This intermediate representation generates an attention weight representation through a sigmoid activation layer, used for weighting the rotated input vector. Finally, the weighted vector is rotated back to its original direction. The pooling module is responsible for reducing the size of one dimension of the tensor to 2 by average-pooling and max-pooling, and then concatenating them.

Our constructed TA Block, as shown in Figure 2, utilizes the triple attention mechanism to establish connections between the inline, xline, and depth dimensions, enhancing the network model's extraction of spatial information and better integrating underground geological structural features. In Figure 2, taking the first branch as an example, a copy of the input is first made, and then the copied data are rotated 90 degrees along the depth direction. The rotated data are then divided into two branches. The first branch performs global and average pooling in the inline direction, then extracts features through a 7 × 7 convolution, and passes the extracted features through a sigmoid layer to become attention weight values. Multiplying the attention weight values with the rotated data generates attention for the depth and xline dimensions, allowing interaction between these two dimensions.

The remaining two branches follow the same approach, but each is rotated 90 degrees along the "inline" and "xline" directions before subsequent operations. This completes the joint processing among the three dimensions, namely, our "inline–xline", "inline–depth", and "xline–depth" joint channels. Next, the outputs of these three joint channels are fed

into the CA Block, where weights are learned for each channel individually. These weights are then used for weighted summation, completing end-to-end triple attention weighting. This enables attention interaction between every pair of dimensions, effectively extracting three-dimensional information from the spatial data.
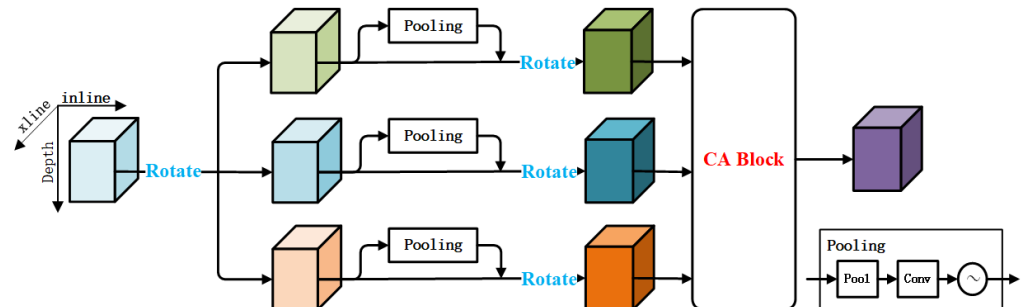
Figure 2. TA Block.

2.2.2. CA Block

In the TA Block, we calculate the attention weights between dimensions pairwise, involving three joint channels: inline–xline, inline–depth, and xline–depth. However, the patterns of changes in the horizontal and vertical directions in geoscience data are constrained by geological structures and features, and are not the same. This is because the "inline" and "xline" dimensions correspond to the planar distribution of geological structures, while the "depth" dimension corresponds to the temporal geological changes in geological ages. Therefore, the network needs to differentiate its focus on the three joint channels. To address this, we introduce a channel attention mechanism to enable the model to allocate and adjust attention weights more accurately. This helps in capturing the characteristics of horizontally distributed geological structures more effectively, allocating resources and attention more efficiently, and reducing interference from unrelated information, given that data exhibit rapid changes or discontinuities along the "depth" axis.

The CA Block mainly includes two sub-modules [32]: the squeeze module and the excitation module. The squeeze module aims to pool the feature map globally to obtain global information for each channel. This extraction allows for capturing each channel's statistical characteristics, providing input for the subsequent excitation module. In this paper, a joint channel is pooled as a feature map. Based on the information provided by the squeeze module, the excitation module uses a small, fully connected network to learn the weights of each channel. Through a sigmoid activation function, the output range is limited between 0 and 1, representing the importance of each channel. These weights are then used to weight the original feature map, realizing the emphasis and de-emphasis of different channels.

In the architecture of the CA Block, as shown in Figure 3, it can be seen that, after the TA Block outputs the feature data of the three joint channels, a squeeze operation must first be performed. This operation encodes the entire spatial feature on one channel as a global feature value, i.e., a scalar, achieved using global average pooling. Subsequently, through concatenation, an excitation operation captures the relationships between channels, learning the attention weights between channels. Finally, multiplying the channel attention weights with the output feature data of the joint channels completes the differentiation of attention to the joint channels.

2.3. Network Training

Geoscientific data typically exhibit nonuniform distribution before interpolation and become large-scale after interpolation. This data characteristic requires the network to have efficient computation methods and training strategies. To address this challenge, we introduce the concept of transfer learning. Transfer learning is a machine learning method that allows us to apply the knowledge and model parameters learned in one task

to another related task without retraining the model from scratch. In the interpolation of three-dimensional velocity fields, we cannot directly train the data for the entire work area, so it is necessary to divide the work area into blocks. The irregularity and unevenness of the data require us to train the network strategically. We introduce the fine-tuning idea of transfer learning to help the network better adapt to the specific requirements of the three-dimensional velocity field interpolation task.
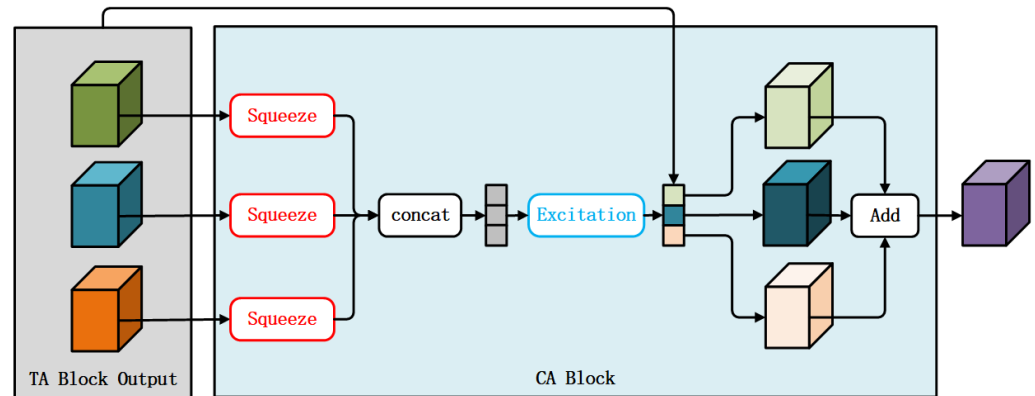
Figure 3. CA Block.

2.3.1. Feasibility Analysis of Transfer Learning

In many real-world applications, the act of gathering the necessary training data and reconstructing models is either costly or unfeasible. Minimizing the need and effort to recollect training data would be advantageous [36]. The task of three-dimensional velocity field interpolation is to recover complete data by interpolating sparse data. Known data are limited, and the data to be recovered are large-scale. The distribution of geoscience datasets is often uneven. This means that data in some areas may be very dense, while other areas may be very sparse. Transfer learning can obtain prior knowledge from data-dense geoscience areas and transfer it to data-sparse areas. This approach can reduce the training complexity in sparse data point work areas. Random initialization of network parameters on large-scale, high-dimensional, and irregular geoscience data may make the model difficult to converge. This paper introduces a fine-tuning strategy of transfer learning to obtain more key features about geological or geoscience phenomena in data-dense areas, so that the model can interpolate in sparse areas more accurately. Once the model gains more feature information in data-dense areas, it can generalize better to sparse areas, thereby improving the convergence of interpolation.

At the same time, this paper introduces the attention mechanism, which is catastrophic for the training of network model parameters. On the one hand, the amount of data is too small to support the network model to learn useful feature representations. On the other hand, the method based on implicit prior information adopted in this paper transforms the interpolation solution of the three-dimensional velocity field from the model space to the parameter space, and it is particularly difficult for the parameters of the network model to find the correct iterative update direction for convergence. Therefore, we use the concept of transfer to solve these two problems.

2.3.2. Network Training Based on Transfer Learning

The key idea of transfer learning is that we can learn some general features and knowledge from a related but different task, and then transfer these features and knowledge to our target task. In our research, by first learning in data-dense areas, we prioritize obtaining general features and knowledge related to the overall area, which helps to improve the initial performance of the model. During the training process, the deep network parameters are kept unchanged, and only the shallow data are fine-tuned. This ensures that the model retains the general knowledge learned from the source task and

only fine-tunes the shallow network layers to adapt to the specific target task. This method effectively balances the irregularity of geoscience data sampling.

In the past, random initialization of network model parameters, if the initialized parameters are poor, leads to difficulties in network model convergence, or even no response, resulting in no training results. To solve this situation, Glorot et al. [37] invented the method of Xavier initialization, which randomly samples parameters from a normally distributed parameter, making the network's initial parameters satisfy a similar distribution, thus enhancing the stability and efficiency of the network model. This paper adopts this initialization method to improve the performance, stability, and efficiency of the deep learning model.

The specific transfer learning strategy of this paper involves dividing the overall survey data into blocks. We first select areas with high data density for training. During the training process, the deep network parameters are kept unchanged, and only the shallow data are fine-tuned. After the training is completed, interpolation is performed on the sparse survey areas. Based on the fine-tuning strategy, the network can benefit from previous learning experiences each time it trains a new area, thus improving training speed and effects. At the same time, it enhances the stability of the model, avoiding problems that may be brought about by random initialization. More importantly, this strategy allows the network to adapt better to various different data distributions and features, retain deep features, and fine-tune shallow features, which is consistent with the characteristics of geoscience data.
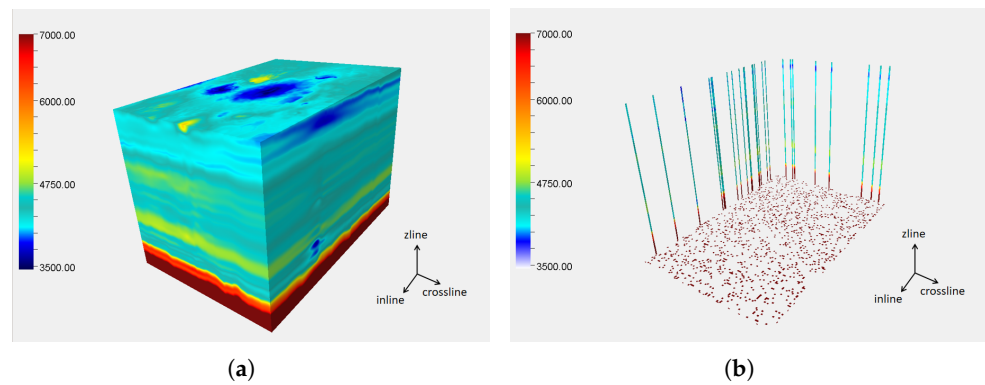
3. Experiments

To verify the effectiveness of the attention-mechanism-based network model JointA 3DUnet on the three-dimensional velocity field interpolation task, we conducted a series of experiments and analyses. This section will discuss our experimental design and results in detail.

3.1. Synthetic Data

Since it is impossible to obtain the real three-dimensional velocity field data for the entire work area, we need to evaluate the network's interpolation results through synthetic data. Synthetic data can be generated according to specific patterns and rules, ensuring that all variables are controllable during the testing process while also reducing the complexity of network pretraining. After generating synthetic data for the entire work area, we simulate sparse data by sampling and inputting them into the network, and then compare the residuals of the interpolation results and synthetic data. By using synthetic data, we can test and verify the algorithms of this paper under known geological conditions.

Figure 4a shows synthesized three-dimensional velocity field data. Subsequently, sampling of the synthesized simulated three-dimensional velocity field is conducted, as shown in Figure 4b. We use the sampling results as actual observed data. Moreover, we can set the sampling rate to evaluate the network's three-dimensional interpolation capability by comparing different existing data volumes. This is then applied to subsequent real data.

To compare the effects of basic 3DUnet and the constructed JointA 3DUnet network's three-dimensional interpolation, we use both networks to interpolate the sampled synthetic data. To control the singularity of the dependent variables, the inputs, loss functions, and training methods of the two networks are the same, and the input data are a randomly sampled 10% of the synthetic data. The obtained three-dimensional velocity field is shown in Figure 5. From Figure 5, neither 3DUnet nor JointA 3DUnet has been able to achieve an ideal interpolation effect. In the areas with layer undulations and low values, the interpolation results of both networks are uneven. The reason for this phenomenon is that the receptive field of the convolution kernel is too small, and the network cannot learn useful features in positions with large data loss. Therefore, this paper introduces dilated convolution to increase the receptive field of the convolution kernel.

Figure 4. 3D velocity field synthetic data schematic diagram. (**a**) Synthetic 3D velocity field. (**b**) Observed data obtained through sampling.
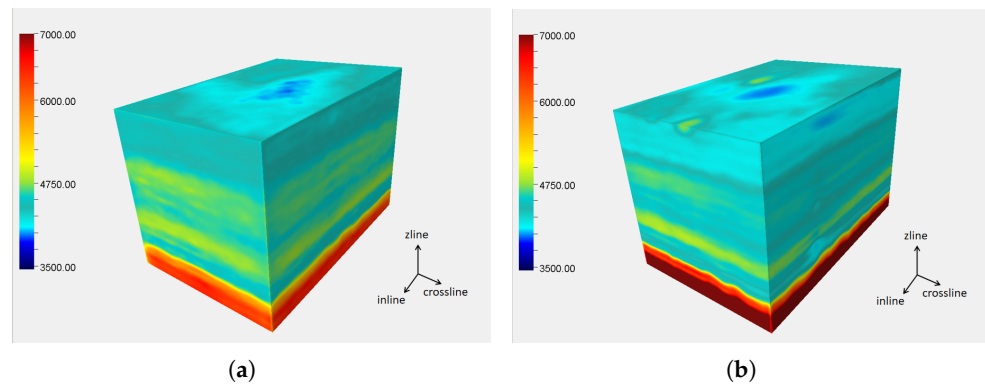
Figure 5. Interpolation results of simulated data. (**a**) 3dUnet. (**b**) JointA 3DUnet.

In the later experiment, we chose the expansion rate as 1,2,5, and the interpolation results are shown in Figure 6. From the figure, the continuity of the undulating layers has been improved. However, the interpolation effect of 3DUnet in the low-value areas is still not ideal after combining with dilated convolution. The network JointA 3DUnet in this paper can overcome this problem better after combining with dilated convolution.

As shown in Figure 7, we conducted experiments on a synthetic dataset. We visualized the overall relative error, and observation point relative error for the interpolation results obtained from both methods as well as after adding dilated convolution.

From Figure 7, it can be observed that JointA 3DUnet outperforms the traditional 3DUnet in terms of observation point relative error and overall relative error. Compared with 3DUnet, the JointA 3DUnet without adding dilated convolution reduces the overall relative error by 22.31%, and the relative error of observation points by 28%. After adding dilated convolution, the errors of both JointA 3DUnet and 3DUnet decrease, but the error of JointA 3DUnet is still the smallest. Compared with 3DUnet, the overall relative error of JointA 3DUnet decreases by 22.77%, and the relative error of observation points decreases by 21.43%. Therefore, the interpolation results of JointA 3DUnet are more accurate and better than those of 3DUnet.

Although the observation point relative error of 3DUnet is also lower, the overall effect is not ideal. This is due to the training method used. The loss function defined in this paper fits the speed data of the observation point positions. As the number of training times increases, even the traditional 3DUnet can fit the observation points. However, the missing parts of the non-observation points need to be generated by network interpolation. The CNN module in the network implicitly undertakes the role of using the correlation in the data to learn its internal structure's prior information. The construction of the CNN model is crucial for the interpolation results of the missing positions. A large receptive field, multi-scale information, and three-dimensional spatial dimension interaction can help the CNN

model to interpolate more suitably the missing speed data around while fitting the speed data of the observation points, thus generating a more accurate three-dimensional velocity field. Overall, the interpolation effect of JointA 3DUnet combined with dilated convolution is better, being able to recover the real speed values of the synthesis more accurately.
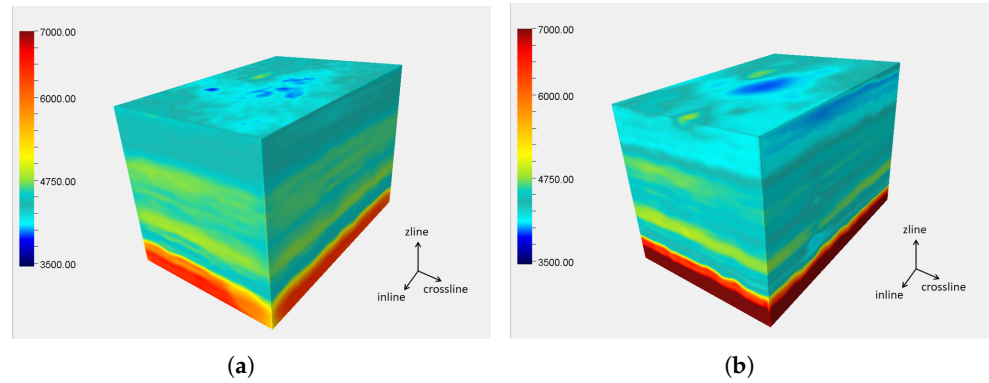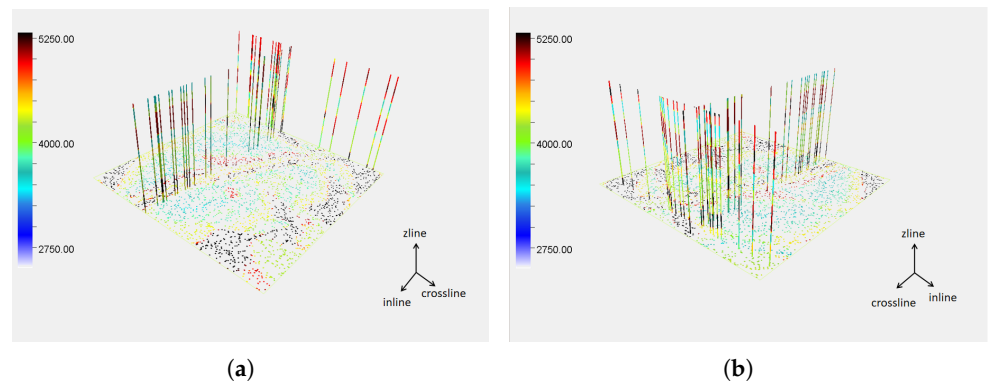

Figure 6. Interpolation results of simulated data with the addition of dilated convolution. (**a**) 3DUnet with dilated convolution. (**b**) JointA 3DUnet with dilated convolution.
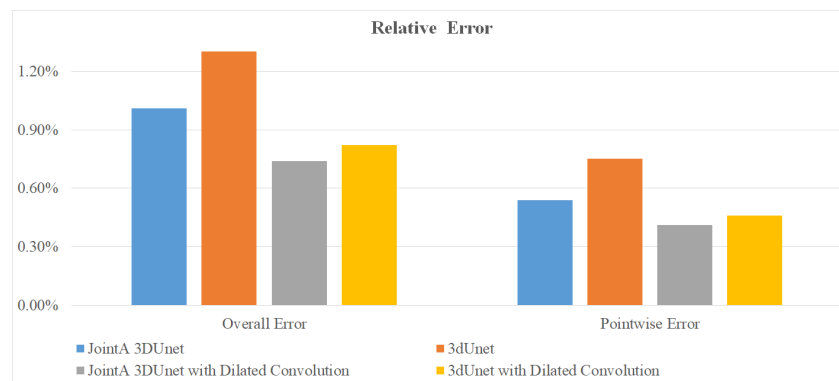
Figure 7. Comparison of interpolation results for simulated data between 3DUnet and JointA 3DUnet.
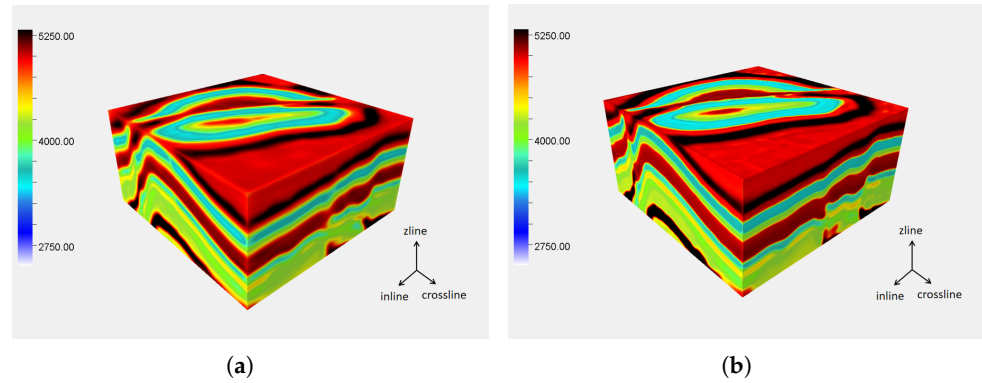
3.2. Real Data

For real data, our objective is to validate the performance of the network constructed in this paper further. Due to the fact that real data typically consist of only a limited amount of data and are unevenly distributed, as shown in Figure 8, the network represents different perspectives of sampled data. The combination of 3DUnet and JointA 3DUnet with dilated convolutions has shown promising interpolation results on synthetic data. Therefore, we test the interpolation performance of these two networks on real data directly.

Figure 8. Different slices of real data are displayed. (**a**) Slice A. (**b**) Slice B.

Figure 9 displays the interpolated result images in real data. It can be observed that, in the interpolation process of real data, the combination of 3DUnet and dilated convolutions leads to discontinuous layers in high value areas, while the integration of JointA 3DUnet and dilated convolutions can overcome this difficulty.

(**a**)   (**b**)

Figure 9. Interpolation results of real data with the addition of dilated convolution. (**a**) 3DUnet with dilated convolution. (**b**) JointA 3DUnet with dilated convolution.

The comparison of interpolation results is shown in Figure 10, where the JointA 3DUnet network has a lower observation point error and overall higher performance, compared with 3DUnet combined with dilated convolution. The interpolation results obtained from both methods have high SNR values, which is due to the inherent smoothing effect of Unet when used for interpolation tasks, resulting in high SNR values for both interpolation results. After adding inflation convolution, JointA 3DUnet reduces the relative error at the observation points by 47.34% compared with 3DUnet. It can be observed that, for more complex measured data, the method mentioned in this paper has precise interpolation effects, especially for more delicate geological structures such as faults, where JointA 3DUnet can restore their shapes better.

We selected a channel of data for plotting the line chart, as shown in Figure 11, It can be seen that the obtained interpolation results are very close to the real curve, and the errors are also found to be within a smaller range after the difference is taken.

In order to compare and evaluate the interpolation ability of JointA 3DUnet combined with dilated convolution for three-dimensional velocity field data with different sampling rates, we conducted interpolation experiments on the original real dataset, with sampling rates of 20% and 30%. The 20% sampling is based on resampling the data sampled at 30%, and the distribution of the two sets of data is the same. The obtained three-dimensional velocity field interpolation results are shown in Figure 12.

Figure 10. Comparison of interpolation results for real data between 3DUnet and JointA 3DUnet.
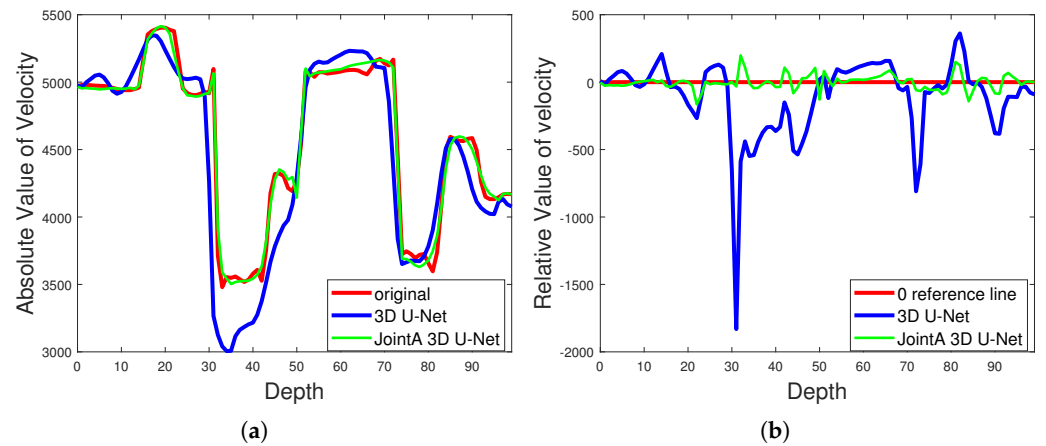
Figure 11. Displaying a graph of one trace extracted from the interpolation results of real data. (**a**) Three values of velocity. (**b**) The output value differs from the true value.
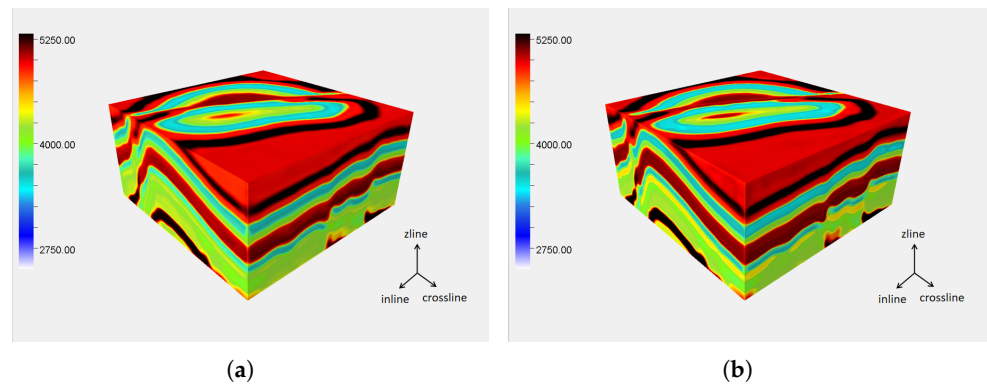

Figure 12. Interpolation Results of real data with different levels of sparsity. (**a**) Sparse 80%. (**b**) Sparse 70%.

The figure demonstrates that the interpolation effects of 30% and 20% sampling are highly similar, and the 3D velocity field data are well recovered. Although the introduction of the attention mechanism improves the model's performance, it also increases the complexity and training difficulty. To address this issue, we decided to adopt a transfer learning strategy to enhance the training efficiency and performance. Figure 13 compares the network convergence curves before and after the transfer learning implementation. It can be observed that, after utilizing transfer learning, the convergence speed is accelerated, and the network performs well in reconstructing velocity field data in sparse areas.
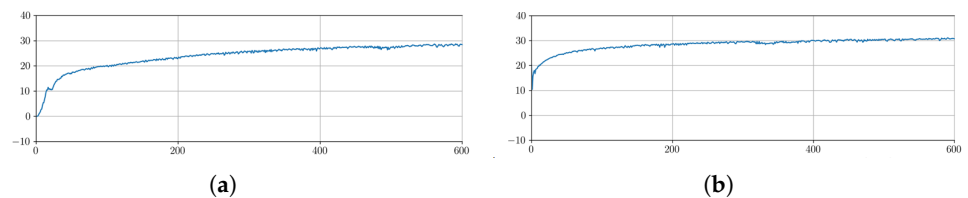
Figure 13. Convergence curve comparison. (**a**) Without transfer learning. (**b**) With transfer learning.

The interpolation effect diagram is shown in Figure 14. From the figure, it can be seen that, after adding the idea of transfer to the JointA 3DUnet network, its three-dimensional interpolation ability has been improved in terms of speed and quality.

| (**a**) | (**b**) |
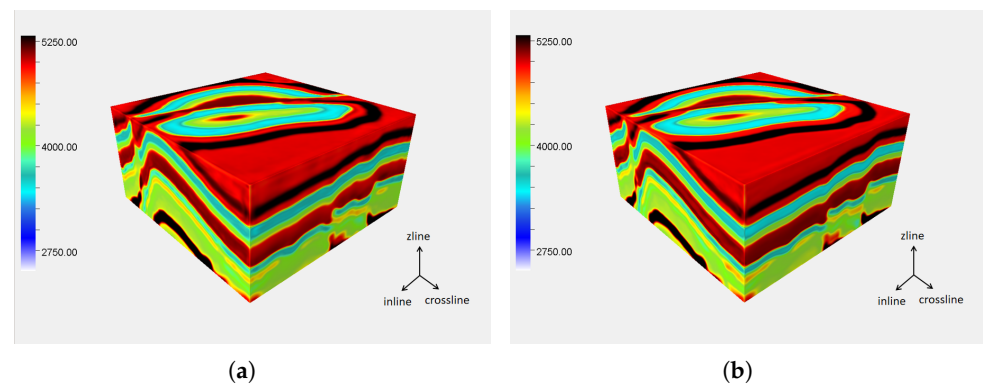
Figure 14. Figure of results for transfer learning. (**a**) Without transfer learning. (**b**) With transfer learning.

4. Discussion

In the experimental process of this study, we first compared the performance of the 3DUnet and JointA 3DUnet network models for synthetic data. Preliminary experiments did not introduce dilated convolution. At this stage, we mainly observed and compared the interpolation effects of these two models under baseline conditions. The experimental results provided us with a preliminary performance evaluation and understanding, which was conducive to our further optimization and improvement of the model structure and strategy. In the next stage, we introduced dilated convolution into the two models and conducted comparison experiments again. The results of this stage showed that the inclusion of dilated convolution indeed enhanced the model's interpolation performance, especially when dealing with real data.

For the real data experiments, we first compared the effects of 3DUnet and JointA 3DUnet with dilated convolution. The experiment proved that JointA 3DUnet performs better than 3DUnet on real data. At the same time, we also conducted interpolation experiments under different data extraction ratios (20% and 30%), further confirming the robustness and efficiency of JointA 3DUnet.

In the final stage, we compared the speed and interpolation effects of the network after adding transfer learning. Through this experiment, we observed that transfer learning effectively enhanced the performance of JointA 3DUnet, and significantly optimized the network's operating speed.

5. Conclusions

The precise construction of three-dimensional velocity fields occupies an indispensable position in seismic exploration, playing an irreplaceable role in insight and interpretation of underground geological structures. This paper constructs a network model based on a joint attention mechanism, JointA 3DUnet, which innovatively integrates technologies such as dilated convolution, triple attention, and channel attention. This effectively expands the receptive field, enhances information interaction between dimensions, and adapts well to the irregular horizontal and vertical distribution of geoscience data. Additionally, the introduction of transfer learning further optimizes the network's interpolation performance.

Through a series of detailed and in-depth experimental verifications, this paper extensively compares and evaluates the performance of the 3DUnet and JointA 3DUnet network models under different conditions and strategies. The experimental results consistently show that the JointA 3DUnet model demonstrates superior and robust performance compared with 3DUnet at every experimental stage and condition. Particularly after the introduction of dilated convolution and transfer learning strategies, JointA 3DUnet not only achieves a significant improvement in interpolation accuracy but also shows evident advantages in computational speed and efficiency. These series of experiments prove that JointA 3DUnet possesses broad application potential and value, representing an efficient seismic data interpolation model worthy of further promotion and application. How to

capture the characteristics of sparse data and reconstruct them more accurately has always been our research goal. We will also explore the application of JointA 3DUnet in other domains and data, continually refining the network architecture, and building a better interpolation network.

In general, the JointA 3DUnet model not only enhances the accuracy of three-dimensional velocity field interpolation but also opens up new possibilities and perspectives for research and application in related fields, demonstrating the tremendous potential and broad prospects of deep learning in the field of earth sciences.

Author Contributions: Conceptualization, X.Y. and G.H.; Methodology, X.Y. and C.Z.; Validation, L.W. and Y.L.; Data curation, M.S.; Writing—original draft, M.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to confidentiality concerns related to the data.

Conflicts of Interest: Author Mingjun Su was employed by the company PetroChina. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received funding from PetroChina. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Geng, Z.; Zhao, Z.; Shi, Y.; Wu, X.; Fomel, S.; Sen, M. Deep learning for velocity model building with common-image gather volumes. *Geophys. J. Int.* **2022**, *228*, 1054–1070. [CrossRef]
2. Chen, G.; Zhu, J.; Qiang, M.; Gong, W. Three-dimensional site characterization with borehole data—A case study of Suzhou area. *Eng. Geol.* **2018**, *234*, 65–82. [CrossRef]
3. Dumas, R.; Cheze, L. Soft tissue artifact compensation by linear 3D interpolation and approximation methods. *J. Biomech.* **2009**, *42*, 2214–2217. [CrossRef] [PubMed]
4. He, Q.; Zhang, Z.; Yi, C. 3D fluorescence spectral data interpolation by using IDW. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2008**, *71*, 743–745. [CrossRef] [PubMed]
5. Guo, J.; Nie, Y.; Li, S.; Lv, X. Application of three-dimensional interpolation in estimating diapycnal diffusivity in the South China Sea. *J. Mar. Sci. Eng.* **2020**, *8*, 832. [CrossRef]
6. Abma, R.; Kabir, N. 3D interpolation of irregular data with a POCS algorithm. *Geophysics* **2006**, *71*, E91–E97. [CrossRef]
7. Bagheripour, P.; Asoodeh, M.; Nazarpour, A. Vertical resolution enhancement of petrophysical Nuclear Magnetic Resonance (NMR) log using ordinary kriging. *Cent. Eur. J. Geosci.* **2013**, *5*, 450–456. [CrossRef]
8. Ilozobhie, A.J.; Obi, D.A.; Okwueze, E.E. Geostatistical analysis of porosity distribution from well log data, in part of Bornu Basin, North-eastern part, Nigeria using Kriging and Co-Kriging methods. *Adv. Appl. Sci. Res.* **2015**, *6*, 83–95.
9. Kang, S.K.; Shin, S.A.; Seo, S.; Byun, M.S.; Lee, D.Y.; Kim, Y.K.; Lee, D.S.; Lee, J.S. Deep learning-based 3D inpainting of brain MR images. *Sci. Rep.* **2021**, *11*, 1673. [CrossRef]
10. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Comput-er-Assisted Intervention—MICCAI 2016, Proceedings of the 19th International Conference, Athens, Greece, 17–21 October 2016*; Proceedings, Part II 19; Springer International Publishing: Cham, Switzerland, 2016; pp. 424–432.
11. Yan, K.; Wang, X.; Lu, L.; Summers, R.M. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J. Med. Imaging* **2018**, *5*, 036501. [CrossRef]
12. Wang, Y.; Yu, B.; Wang, L.; Zu, C.; Lalush, D.S.; Lin, W.; Wu, X.; Zhou, J.; Shen, D.; Zhou, L. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage* **2018**, *174*, 550–562. [CrossRef] [PubMed]
13. Yu, L.; Yang, X.; Chen, H.; Qin, J.; Heng, P.A. Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31.
14. Pham, C.H.; Ducournau, A.; Fablet, R.; Rousseau, F. Brain MRI super-resolution using deep 3D convolutional networks. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 197–200.

15. Gao, Y.; Guo, J.; Wang, J.; Lv, X. Assessment of Three-Dimensional Interpolation Method in Hydrologic Analysis in the East China Sea. *J. Mar. Sci. Eng.* **2022**, *10*, 877. [CrossRef]

16. Le V.D.; Bui, T.C.; Cha, S.K. Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 55–62.

17. Kim, B.; Azevedo, V.C.; Thuerey, N.; Kim, T.; Gross, M.; Solenthaler, B. Deep fluids: A generative network for pa-rameterized fluid simulations. *Comput. Graph. Forum* **2019**, *38*, 59–70. [CrossRef]

18. Sitzmann, V.; Thies, J.; Heide, F.; Nießner, M.; Wetzstein, G.; Zollhöfer, M. Deepvoxels: Learning persistent 3D feature embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2437–2446.

19. Mikhailiuk, A.; Faul, A. Deep learning applied to seismic data interpolation. In Proceedings of the 80th EAGE Conference and Exhibition 2018, Copenhagen, Denmark, 11–14 June 2018; European Association of Geoscientists & Engineers: Utrecht, The Netherlands, 2018; Volume 2018, pp. 1–5.

20. Wang, G.; Huang, L. 3D geological modeling for mineral resource assessment of the Tongshan Cu deposit, Heilongjiang Province, China. *Geosci. Front.* **2012**, *3*, 483–491. [CrossRef]

21. Liu, Z.; Zhang, Z.; Zhou, C.; Ming, W.; Du, Z. An adaptive inverse-distance weighting interpolation method considering spatial differentiation in 3D geological modeling. *Geosciences* **2021**, *11*, 51. [CrossRef]

22. Zhao, T.; Wang, Y. Statistical interpolation of spatially varying but sparsely measured 3D geo-data using compressive sensing and variational Bayesian inference. *Math. Geosci.* **2021**, *53*, 1171–1199. [CrossRef]

23. Fisher, P.F.; Ledoux, H.; Gold, C. An efficient natural neighbour interpolation algorithm for geoscientific modelling. In *Developments in Spatial Data Handling, Proceedings of the 11 th International Symposium on Spatial Data Handling, Leicester, UK, 23–25 August 2004*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 97–108.

24. Araya-Polo, M.; Jennings, J.; Adler, A.; Dahlke, T. Deep-learning tomography. *Lead. Edge* **2018**, *37*, 58–66. [CrossRef]

25. Wang, W.; Ma, J. VMB-Net: A deep learning network for velocity model building in a cross-well acquisi-tion geometry. In Proceedings of the SEG International Exposition and Annual Meeting, San Antonio, TX, USA, 15–20 September 2019; p. 5407.

26. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 640–651.

27. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

28. Kazei, V.; Ovcharenko, O.; Alkhalifah, T. Velocity model building by deep learning: From general synthetics to field data application. In *SEG Technical Program Expanded Abstracts 2020*; Society of Exploration Geophysicists: Houston, TX, USA, 2020; pp. 1561–1565.

29. Morten, J.P.; Roth, F.; Karlsen, S.A.; Timko, D.; Pacurar, C.; Olsen, P.A.; Nguyen, A.K.; Gjengedal, J. Field appraisal and accurate resource estimation from 3D quantitative interpretation of seismic and CSEM data. *Lead. Edge* **2012**, *31*, 447–456. [CrossRef]

30. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*, Proceedings, Part III 18; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

31. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.

34. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 6–11 July 2015; pp. 2048–2057.

35. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2021; pp. 3139–3148.

36. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]

37. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; JMLR Workshop and Conference Proceedings, pp. 249–256.