applied sciences

MDPI

Article

Microseismic Data-Direct Velocity Modeling Method Based on a Modified Attention U-Net Architecture

Yixiu Zhou, Liguo Han, Pan Zhang *, Jingwen Zeng, Xujia Shang and Wensha Huang

College of Geo-Exploration Science and Technology, Jilin University, Changchun 130012, China; yixiu21@mails.jlu.edu.cn (Y.Z.); hanliguo@jlu.edu.cn (L.H.); zengjw19@mails.jlu.edu.cn (J.Z.); shangxj19@mails.jlu.edu.cn (X.S.); huangws23@mais.jlu.edu.cn (W.H.)
* Correspondence: zhangpan@jlu.edu.cn

Abstract: In microseismic monitoring, the reconstruction of a reliable velocity model is essential for precise seismic source localization and subsurface imaging. However, traditional methods for microseismic velocity inversion face challenges in terms of precision and computational efficiency. In this paper, we use deep learning (DL) algorithms to achieve precise and efficient real-time microseismic velocity modeling, which holds significant importance for ensuring engineering safety and preventing geological disasters in microseismic monitoring. Given that this task was approached as a non-linear regression problem, we adopted and modified the Attention U-Net network for inversion. Depending on the degree of coupling among microseismic events, we trained the network using both single-event and multi-event simulation records as feature datasets. This approach can achieve velocity modeling when dealing with inseparable microseismic records. Numerical tests demonstrate that the Attention U-Net can automatically uncover latent features and patterns between microseismic records and velocity models. It performs effectively in real time and achieves high precision in velocity modeling for Tilted Transverse Isotropy (TTI) velocity structures such as anticlines, synclines, and anomalous velocity models. Furthermore, it can provide reliable initial models for traditional methods.

Keywords: microseismic; velocity model; deep learning; Attention U-Net

check for updates

Citation: Zhou, Y.; Han, L.; Zhang, P.; Zeng, J.; Shang, X.; Huang, W. Microseismic Data-Direct Velocity Modeling Method Based on a Modified Attention U-Net Architecture. *Appl. Sci.* **2023**, *13*, 11166. https://doi.org/10.3390/app132011166

Academic Editor: Yoshiyuki Kaneda

Received: 21 September 2023
Revised: 5 October 2023
Accepted: 10 October 2023
Published: 11 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Microseismic technology refers to the research and application of faint seismic signals generated by underground geological deformations or rock fractures. Real-time monitoring and analysis of these microseismic signals provide valuable information about underground structures and material distribution. This technology finds wide applications in various fields, including mining monitoring [1], geotechnical engineering [2], geological hazard prevention [3], and tectonic studies [4]. In recent years, significant advancements in microseismic technology research have primarily focused on seismic source localization techniques [5] and microseismic data processing [6,7]. However, to further enhance the utility of microseismic monitoring for the precise assessment of underground target area structures and timely decision-making, efficient and reliable velocity modeling is essential [8].

Seismic velocity modeling plays a crucial role in geophysical exploration as it involves inferring the underground velocity structure based on observed seismic records. This technique provides valuable insights into the geological structures [9]. There are two main categories of seismic velocity modeling: active source velocity modeling and passive source velocity modeling. Common methods include time-domain Normal Moveout (NMO) velocity analysis [10,11], travel time tomography [12], and full-waveform inversion (FWI) [13]. Currently, these methods are primarily used for active source seismic data velocity modeling. However, when dealing with passive source records such as microseismic data or background noise, a common approach involves reconstructing them into a format similar

to active source records before using conventional methods to obtain velocity information [14,15]. This reconstruction process introduces various optimization parameters and uncontrollable factors, resulting in increased computational costs. Moreover, it may not meet the high-efficiency requirements for velocity model updates in microseismic monitoring. Therefore, in the current phase of research, there is a growing emphasis on the direct inversion of underground velocity structures using microseismic records in the passive source domain [16–18].

In the process of microseismic monitoring, the ability to conduct real-time analysis and make real-time decisions is crucial for ensuring the success of engineering projects and the prevention of geological disasters. This critical capability hinges upon precise seismic source localization, which, in turn, relies on the accuracy of the subsurface velocity model. Therefore, the updating of velocity models should be both efficient and reliable. At present, FWI stands as the most accurate method for constructing velocity models using prestack seismic data. However, FWI heavily relies on low-frequency information, and incorporating high-frequency microseismic data into FWI increases the likelihood of encountering cycle-skipping issues. Traditional algorithms face constraints related to source positions, unknown source wavelets, and high computational costs. To overcome the challenges posed by traditional methods and meet the computational efficiency requirements of microseismic monitoring, this paper has chosen to introduce the method of DL. By using the neural network to directly invert velocity models from microseismic data, dynamically capturing changes in reservoir parameters, which would enhance the accuracy of seismic source localization and help facilitate real-time decision-making. This enables better disaster prediction and safety evaluation [19].

DL, a significant branch of machine learning (ML), has experienced rapid development and extensive application in various fields, with computer vision being particularly prominent. With advancing technology, an increasing number of geophysicists have started applying DL methods to address diverse problems in the field, including lithofacies classification [20], reservoir identification [21], and seismic data processing and interpretation [22], leading to notable progress. Nevertheless, the utilization of DL in seismic velocity modeling is still in its early stages. Pioneering work by Gunter Röth et al. involved using neural networks (NNs) to map synthetic common shot gather data to one-dimensional (1-D) velocity models [23]. Aaron Moya and colleagues used artificial neural networks (ANNs) for seismic inversion to estimate 1D velocity models [24]. Araya-Polo et al. used deep neural networks (DNNs) to reconstruct 2D velocity models from inputs of velocity spectra [25]. Mosser et al. used generative adversarial networks (GANs) for mapping seismic records to velocity models [26]. Li et al. modified a DNN and proposed end-to-end seismic inversion networks (SeisInvNets) to establish spatial correspondences between each seismic trace and the entire velocity model [27]. G. Fabien-Ouellet et al. developed a DNN to estimate root mean square velocity and layer velocities using co-centered shot gather data [28]. Chen et al. introduced Newton machine learning (NML) inversion, which mitigates cycle-skipping issues in conventional FWI by combining forward modeling using Newton wave propagation and the dimensionality reduction capability of ML [29]. Yang et al. modified the U-net architecture to directly reconstruct salt dome velocity models from raw seismic images [9]. Cao et al. proposed a combination of a U-Net and three fully connected networks (UCNN) for multi-parameter pre-stack seismic inversion with constraints on sparse reflection coefficients [30]. Zhang et al. developed a DL-based pre-stack seismic inversion architecture (IHGNN) that improved network accuracy and robustness by introducing a priori loss function and achieved more reasonable inversion results on real data through transfer learning [31].

While DL has made significant progress in active source velocity modeling, the exploration of microseismic velocity inversion based on DL is still in its early stages. Daniel Wamriew et al. used ResNet50, a 50-layer deep residual network, for the inversion of microseismic data recorded by Distributed Acoustic Sensing (DAS) [32]. In the subsequent year, they further modified the AlexNet architecture, enabling real-time localization and

velocity model updates for inter-well microseismic data under horizontally layered (Vertical Transverse Isotropy, VTI) media conditions [19].

In summary, DL has revolutionized the field of geophysics by introducing new tools and methods, leading to significant advancements in active source velocity modeling. Researchers have predominantly employed and adapted various neural networks, or even proposed new ones, to make them more suitable for seismic inversion applications. This progress has seen the evolution from initial 1D models to more complex 2D and even 3D velocity modeling. However, velocity inversion for microseismic events is still in its early stages and encounters several challenges, which highlights the need for further research and exploration in this area.

This paper focuses on velocity modeling in the microseismic field. We adopted and modified the Attention U-Net architecture [33] to automatically extract hidden features and patterns from microseismic records and velocity models. The paper also discusses the degree of coupling of microseismic events individually. Simulate records for single-event and multi-event microseisms with randomly distributed source locations, along with corresponding velocity models, are used for training. Ultimately, real-time reconstruction of microseismic velocity models has been achieved. The effectiveness of this approach is demonstrated through numerical tests using simulated data generated from various complex velocity structures, including anticlines, synclines, and anomalous velocity models.

2. Methods

2.1. U-Net

U-Net is a type of convolutional neural network (CNN) that was initially proposed by Olaf Ronneberger et al. in 2015 [34]. The U-Net's design is inspired by a 'U' shape, featuring symmetric down-sampling and up-sampling paths. The down-sampling path captures image information and generates feature maps at different scales. In contrast, the up-sampling path combines features from different levels and produces high-resolution results. Specifically, it consists of the following components:

- Encoder: This primarily consists of convolutional and pooling layers and captures contextual information in the input images through a series of successive convolution and pooling operations. With an increasing number of layers, the dimensions of the feature maps decrease, enabling the extraction of information at various scales and achieving global awareness. Furthermore, this process substantially reduces computational requirements and memory usage, thereby enhancing training efficiency.
- Skip connections: This is used to retain more input information in the network. After each pooling operation, the results are concatenated with the corresponding layers in the decoder. This approach facilitates the network in capturing features from different levels.
- Decoder: Composed of convolutional layers and up-sampling, systematically restores the fine-grained details that were lost during the encoding process. By mapping the extracted features back to high-resolution representations, the decoder ultimately recovers the original size of the input image.
- Output: The ultimate prediction results are generated by the final layer using a 1×1 convolutional kernel in conjunction with the sigmoid function.

2.2. Attention Gate (AG) Module

The attention mechanism is a simulation of the human visual system, designed to focus on local information. It dynamically adjusts the weights of different positions in feature maps to enhance specific information. This mechanism can be categorized into three main types: spatial attention, channel attention, and spatial and channel mixed attention models. Currently, it finds extensive applications in various domains [35]. In 2018, Ozan Oktay et al. introduced a novel module called AG in the field of medical imaging. This module was integrated with the U-Net architecture, resulting in the Attention U-Net architecture [33].

Specifically, in the Attention U-Net, the AG module is used to adjust the output features of the encoder before concatenating them with the corresponding features from the decoder. The AG module generates a gating signal that automatically calculates the weights of features at different spatial positions, allowing the model to focus on specific regions.

The computational principle within the red box in Figure 1 is as follows: Firstly, the gating tensor controlling encoder features is mapped to the intermediate layer using Formula (1). Secondly, the input tensor representing decoder features is mapped to the intermediate layer using Formula (2):

$$phi_g = W_g^T g \tag{1}$$

$$theta_att = W_x^T x^l \tag{2}$$

where W_g and W_x represent linear transformations, l corresponds to the number of layers. Afterward, phi_g and $theta_att$ undergo correlation operations according to Formula (3) to calculate the query Q_{att}^l. To compute the attention coefficients α using Formula (4), which is finally multiplied with the input tensor x to obtain the weighted encoder feature vector $\widehat{x}$, the following is used:

$$Q_{att}^l = \psi^T(\sigma_1(theta_att + phi_g + b_g)) + b_\psi \tag{3}$$

$$\alpha^l = \sigma_2(Q_{att}^l(x^l, g; \Theta_{att})) \tag{4}$$

where $\sigma_1 = \max(0, x)$ corresponds to ReLU activation function, $\sigma_2 = \frac{1}{1+\exp(-x)}$ corresponds to sigmoid activation function. AG is characterized by a set of parameters Θ_{att}, which includes linear transformation parameters W_g, W_x, ψ and bias parameters b_ψ and b_g [34].

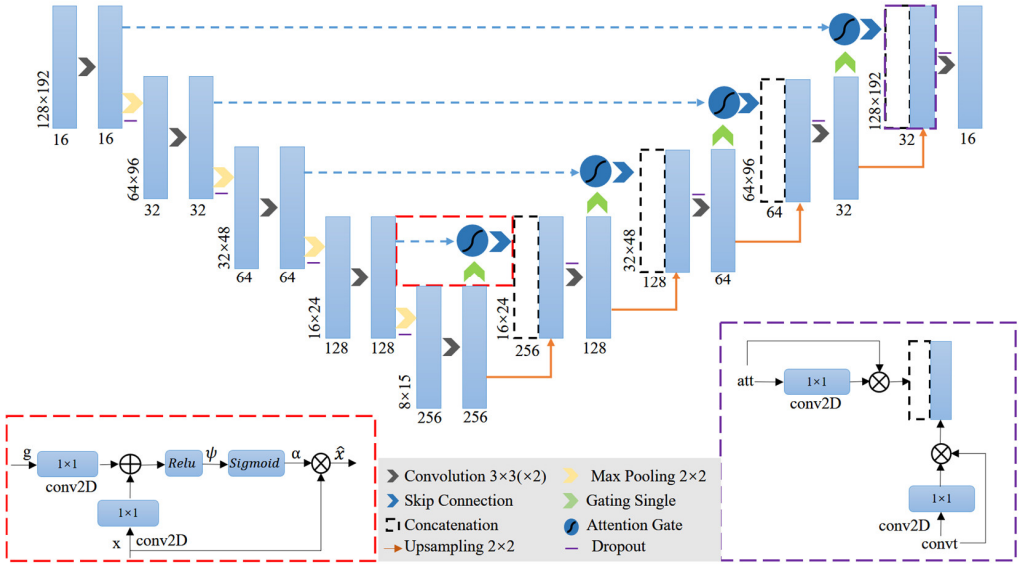


Figure 1. Architecture of the modified Attention U-Net neural network.

2.3. Loss Function

Velocity modeling is approached as a regression problem. As such, the loss function used in this paper is the Mean Squared Error (MSE) function. Its mathematical formula is expressed as follows:

$$m(y, \widetilde{y}) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \widetilde{y}^{(i)}\right)^2 \tag{5}$$

where n represents the sample size, y represents the predicted results, and $\widetilde{y}$ represents the true values. A smaller loss function indicates greater proximity between the predictions and the true values.

2.4. Modified Attention U-Net Neuron Network

The overall architecture of the modified Attention U-Net for velocity modeling in this paper is shown in Figure 1. In a classical Attention U-Net used for image recognition tasks, AGs are typically not applied at the first skip connection due to the argument that they do not capture the characteristics of high-dimensional input data [36]. However, in the context of velocity modeling for seismic data, each layer of feature maps at different scales implicitly contains information and patterns related to the waveforms and velocities. Attention mechanisms can capture the interrelationships between different positions in the input data, aiding in establishing a mapping from the time-domain seismic records to the depth-domain velocity model. Therefore, in this paper, AGs are incorporated into each decoder layer to filter the features propagated through skip connections, thus enhancing the accuracy of velocity modeling. Furthermore, since U-Net's skip connections directly concatenate some of the input with the output, it may add microseismic waveform information into the final velocity model, which can negatively impact the modeling accuracy. Through experiments, it has been observed that the inclusion of AGs alone does not entirely eliminate this influence. As a result, a decision was made to perform weighted concatenation in the final layer. This involves using convolutional operations to automatically learn appropriate weights for the model, followed by weighted concatenation based on the learned weights, as described in Formula (6):

$$\begin{aligned} \lambda_1 &= W_a att \\ \lambda_2 &= W_c convt \\ concat &= concatenate([\tfrac{\lambda_1}{\lambda_1+\lambda_2} att, \tfrac{\lambda_2}{\lambda_1+\lambda_2} convt]) \end{aligned} \tag{6}$$

where λ_1 and λ_2 represent the computed weight coefficients, W_a and W_c as convolution operations, att represents the features after AG filtering, $convt$ represents the up-sampling features, and $concat$ represents the concatenated features.

Training was conducted on a multi-layer random-stratified model dataset using U-Net networks, both with and without the inclusion of attention mechanisms. Experimental testing was performed on test dataset, and the results, including the root mean square error (RMSE) and structural similarity (SSIM), are presented in Table 1. SSIM is designed to simulate human visual perception of image quality by comprehensively assessing the similarity between two images based on brightness, contrast, and structural factors. It provides ranging values [0, 1], with higher values indicating greater similarity. As shown in Table 1, the accuracy of the modified Attention U-Net has improved.

Table 1. The difference between the predictive models before/after adding AGs and the true model.

Neural Network	SSIM	MSE
U-Net	0.9355	1×10^{-3}
Modified Attention U-Net	0.9729	2×10^{-4}

The network ingests microseismic records as input and generates corresponding two-dimensional velocity models as output. Before the network training phase, data preprocessing is conducted. Initially, the microseismic data undergoes normalization. In this paper, maximum value normalization is used to scale all microseismic records within the $[-1, 1]$ range. This normalization process aims to mitigate variations in size among different features, enhance the model's stability during the learning process, improve its resilience to data transformations and perturbations, and expedite convergence. Seismic records typically possess high sampling rates, leading to significant storage requirements for the dataset and increased demands on computational resources. Moreover, the discrep-

ancy in size between inputs and outputs leads to augmented computational complexity. Consequently, down-sampling is applied to the normalized seismic records to enhance the efficiency of model training by reducing the data volume. Additionally, this aligns the input and ground truth to the same spatial positions, simplifying the computation process and enabling the neural network to learn and predict physical parameters more effectively.

The overall workflow of this paper is depicted in Figure 2.

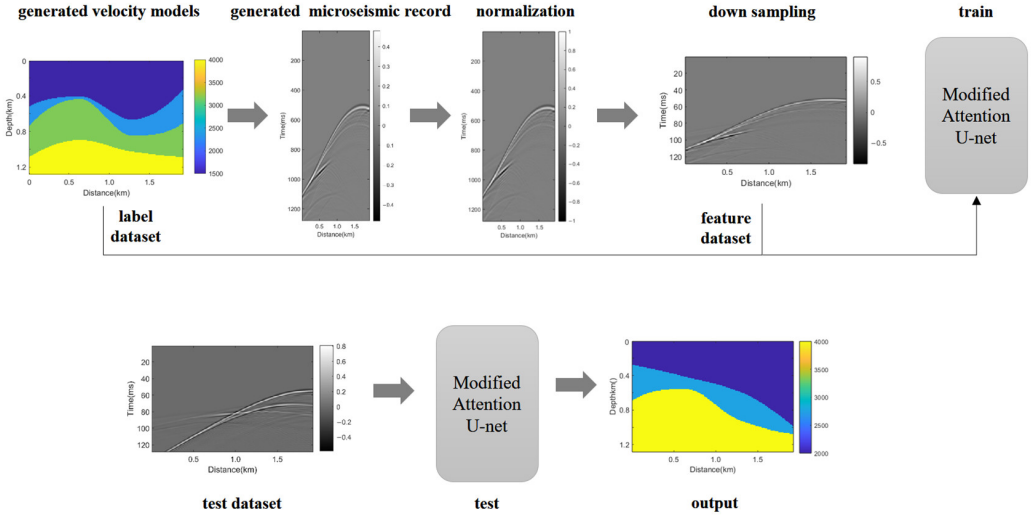


Figure 2. The workflow of this paper.

3. Numerical Tests and Results

For microseismic events, it is a common practice to segment long-term monitoring records into shorter intervals. These segmented records may contain individual or multiple microseismic events. If there is no clear correlation between multiple events, they can be separated into individual microseismic records for further feature extraction and analysis. However, in practical applications, the presence of strong coherence between events, coupled with location offsets or delays caused by measurement or localization errors, can result in signal cross-interference and introduce crosstalk noise. Furthermore, multiple microseismic events may share underground structures and other complex factors, leading to intricate and unpredictable interaction effects among them. As a result, these events become challenging to separate.

To effectively address these scenarios, this paper uses the separability of microseismic events as a foundation. It involves simulating the training of neural networks using both single-event microseismic records and multi-event microseismic records as feature datasets. This approach ensures that even when dealing with heavily coupled and inseparable multiple microseismic signals, the trained model can efficiently and accurately perform velocity modeling.

3.1. Label Datasets

3.1.1. Multi-layer Random Stratigraphic Models

In numerical tests, the dataset creation process involved employing a two-dimensional discretization method to discretize the seismic velocity model into a grid under TTI media condition. The grid had a horizontal and vertical grid spacing of 10 m. We also used Perfectly Matched Layer (PML) that absorbs boundaries to reduce the impact of boundary effects on the simulation results, thereby enhancing numerical accuracy. We have maintained a constant density to simplify the model. To ensure dataset complexity, velocity models were randomly generated using a program. The vertical grid consisted of 128 points, while the horizontal grid had 192 points, so the depth of the individual model obtained is 1.28 km, and the horizontal distance is 1.92 km. The number of layers varied randomly between 2 and 4, with velocity increasing gradually from top to bottom as the

number of layers increased. The minimum velocity was set at 1500 m/s, and the maximum velocity was set at 4000 m/s. A labeled dataset of 3000 velocity models was generated, which included both anticlinal and synclinal structures. Figure 3 shows six of these velocity models, with two models each having two, three, and four layers.

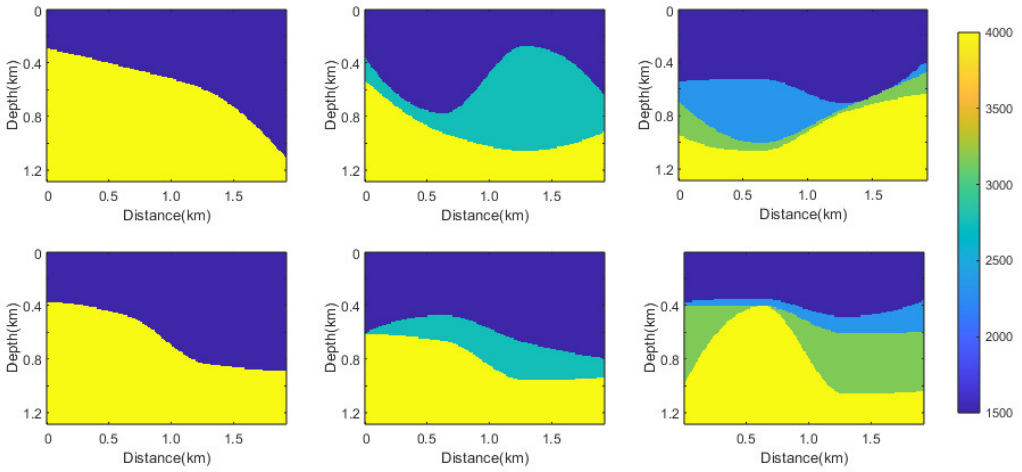


Figure 3. The generated multi-layer random stratigraphic models.

3.1.2. Local Marmousi Models

In this section, a rectangular window measuring 128 × 192 was used to extract velocity models from the Marmousi model, as shown in Figure 4. The maximum velocity of the model was constrained to 4000 m/s. The window was systematically moved from left to right and top to bottom with a horizontal step size of 120 and a vertical step size of 150. Following this process, a total of 1700 velocity models that held practical significance were chosen from the extracted models. Figure 5 shows six representative velocity models selected from the aforementioned set. These models differ in terms of layer numbers, dip angles, and velocity ranges. The velocity distributions demonstrate nonlinearity, contributing to the complexity of the models. To enhance the neural network model's generalization capability and mitigate the risk of overfitting, data augmentation techniques were used. As shown in Figure 6, horizontal and vertical flips were applied to each of the 1700 velocity models. Consequently, the labeled dataset consisted of 5100 velocity models after augmentation. The expanded labeled dataset encompasses a wider range of velocity models, thereby facilitating the neural network's adaptation to diverse samples and application scenarios.

3.2. Feature Datasets

3.2.1. Single-Event Microseismic Records

Based on the completed labeled dataset, we used finite-difference forward modeling to simulate microseismic records corresponding to the features. The detectors were positioned at a depth of 10 m with a horizontal spacing of 10 m, resulting in a total of 192 detectors. The sampling interval was set to 1 ms, and the sampling duration was 1.28 s, incorporating fully matched absorbing boundaries. For the simulation of microseismic signals, we employed a Ricker wavelet with a dominant frequency of 30 Hz as the source function. This source function was applied to each individual underground source location within the velocity model, specifically those situated below 500 m in depth. The forward modeling process yielded a set of single-event microseismic records, with dimensions of 1280 × 192. To align with the dimensions of the labeled dataset, these records were down-sampled as previously described in Section 2.4. The records were down-sampled to dimensions of 128 × 192.

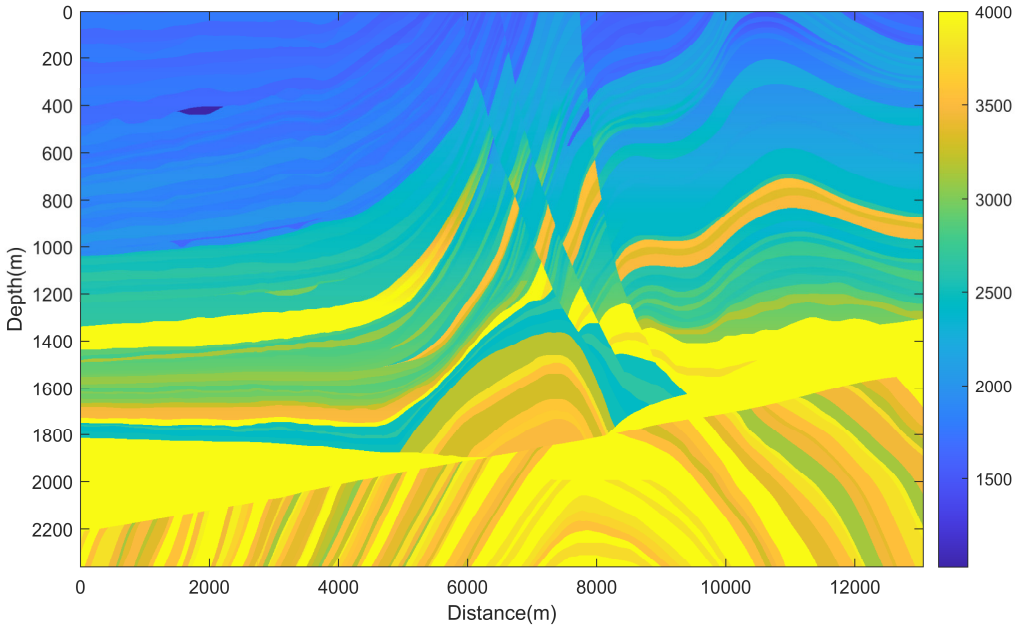


Figure 4. The Marmousi velocity model used for generating the labeled dataset.

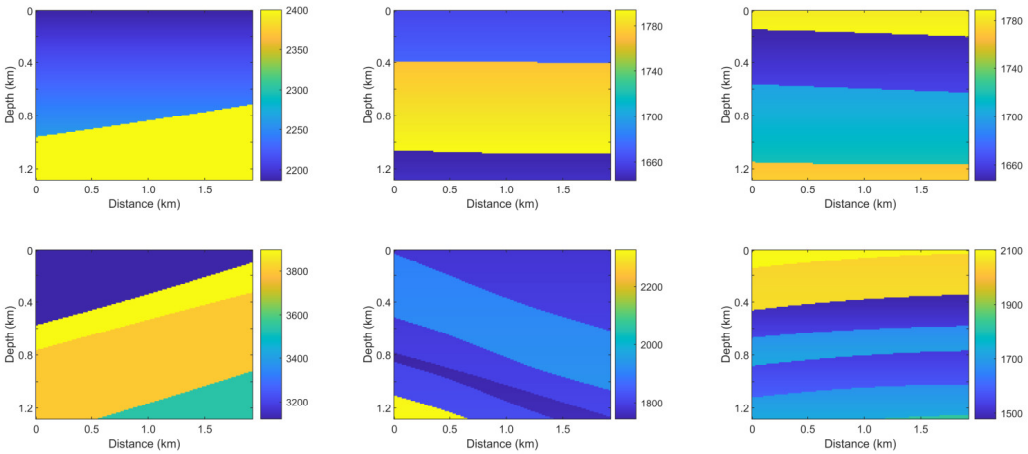


Figure 5. The extracted local Marmousi models obtained from the data subset.

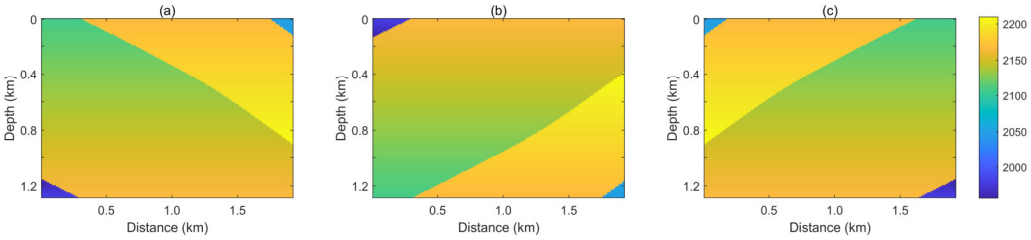


Figure 6. (**a**) The original velocity model. (**b**) The vertically flipped velocity model derived from (**a**). (**c**) The horizontally flipped velocity model derived from (**a**).

3.2.2. Multi-Event Microseismic Records

In our simulation of multi-event microseismic records, we used the same configuration for the observational system as previously described. Regarding the source parameters, we placed an unspecified number of underground sources, ranging from 10 to 15, within a depth range below 500 m. For excitation, we maintained the use of a Ricker wavelet with a dominant frequency of 30 Hz. To enhance the realism of the simulated microseismic signals, this paper incorporated random time delays within the range of (0, 100 ms) for

each sub-wavelet during numerical simulations. By using this methodology, we were able to replicate phase variations and interference phenomena that occur during sub-wavelet propagation, considering factors such as reflection and refraction. This approach facilitated a more accurate reconstruction of vibration signals, thereby improving the representational capabilities of the results. Figure 7 shows an example of a single multi-event microseismic record obtained through this process.

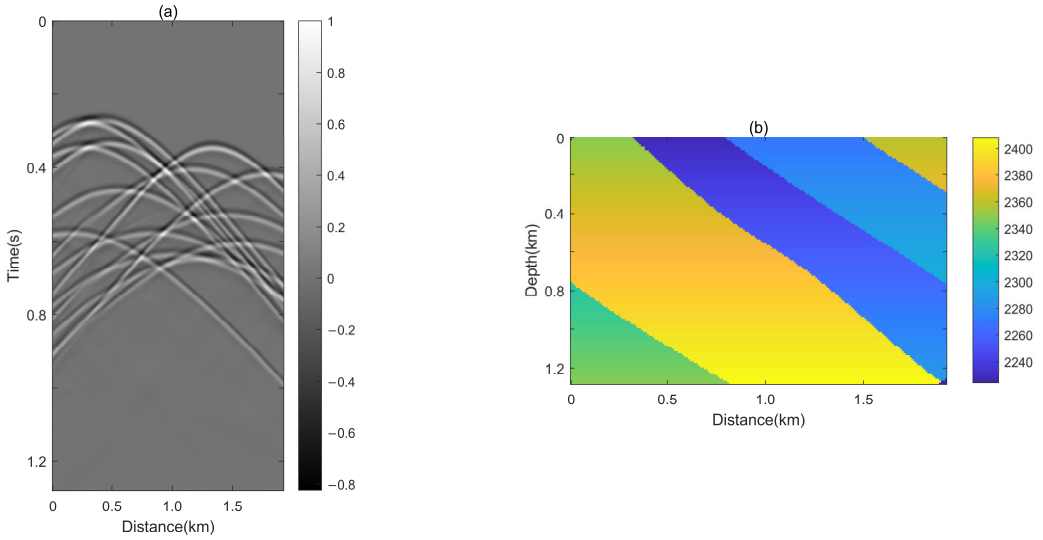


Figure 7. (**a**) An original multi-event microseismic record. (**b**) Corresponding velocity model.

3.3. Results

The datasets are partitioned into three sets using an 8:1:1 ratio, consisting of training sets, validation sets, and test sets. The training set is used to train the parameters and features of the model; the validation set is used to monitor the model's performance and conduct hyperparameter tuning; and the test set is exclusively reserved for evaluating the model's performance. Regarding the model's parameter settings, the optimization algorithm chosen is the Adam algorithm with a learning rate of 0.0001. Dropout layers are incorporated into every layer except the fifth layer to mitigate overfitting, with a dropout rate of 0.25 in the first layer and 0.5 in the subsequent layers. To strike a balance between computational efficiency and dataset size, the batch size is set to eight, and the number of training epochs is adjusted based on real-time monitoring of the loss function curve. Four sets of data are fed into the neural network for training, resulting in the development of four independent models.

3.3.1. Multi-Layer Random Stratigraphic Models

The descent of the loss function during the training process for the multi-layer random stratigraphic models is shown in Figures 8 and 9. Figure 8 shows the loss for the single-event microseismic dataset, while Figure 9 shows the loss for the multi-event microseismic dataset.

Comparing the loss descent curves of the two models reveals distinct patterns for the single-event and multi-event microseismic datasets. The single-event dataset displayed notable oscillations in the early stages of validation during training, gradually stabilizing after approximately 200 iterations and ultimately converging to the optimal solution at around 400 iterations. In contrast, the loss for the multi-event dataset maintained overall stability throughout training, showing smoother descent and reaching the optimal solution after approximately 200 iterations. This disparity can be attributed to the inherent limitations of single-event datasets, which are constrained by the number and positions of seismic sources. These constraints make it challenging to capture all the features of underground velocity structures accurately. As a result, the single-event dataset experiences more pronounced early fluctuations and sharper gradient changes, necessitating a greater number of iterations to find the best parameter configuration compared to multi-event datasets.

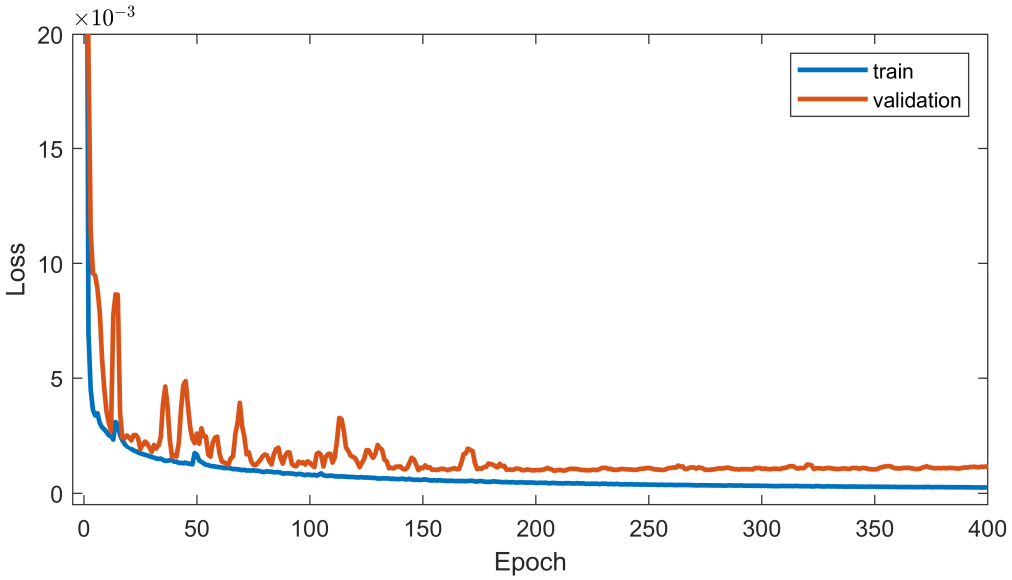


Figure 8. Loss decreases during the training process for single-event dataset of multi-layer random stratigraphic models.

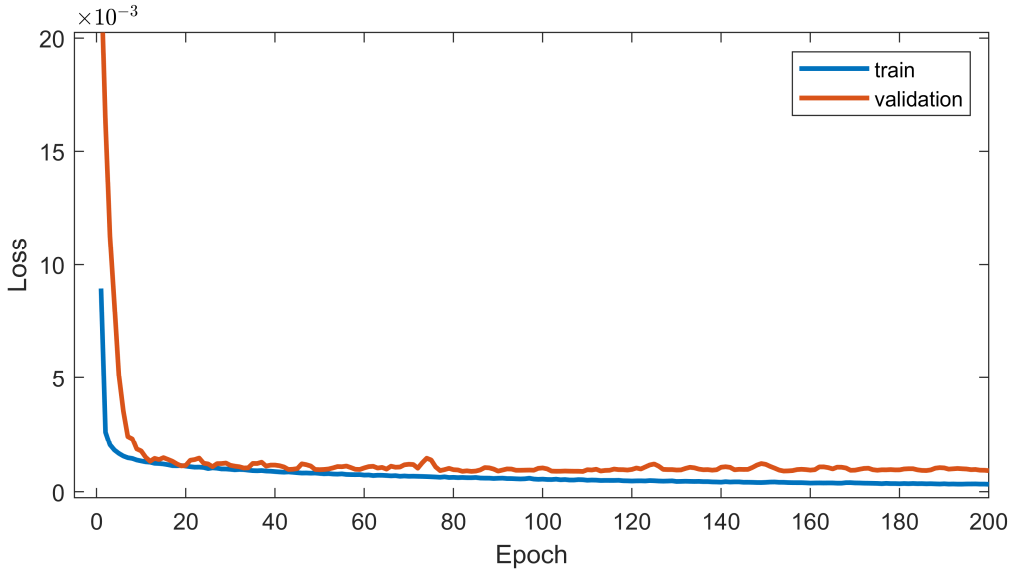


Figure 9. Loss decreases during the training process for multi-event dataset of multi-layer random stratigraphic models.

In the single-event test dataset, we selected one velocity model each with two, three, and four layers for presentation in Figure 10. Additionally, we have plotted the corresponding velocity profiles at the chosen model locations, as shown in Figure 11. This visual representation provides a more intuitive evaluation of the accuracy of the modeling approach proposed.

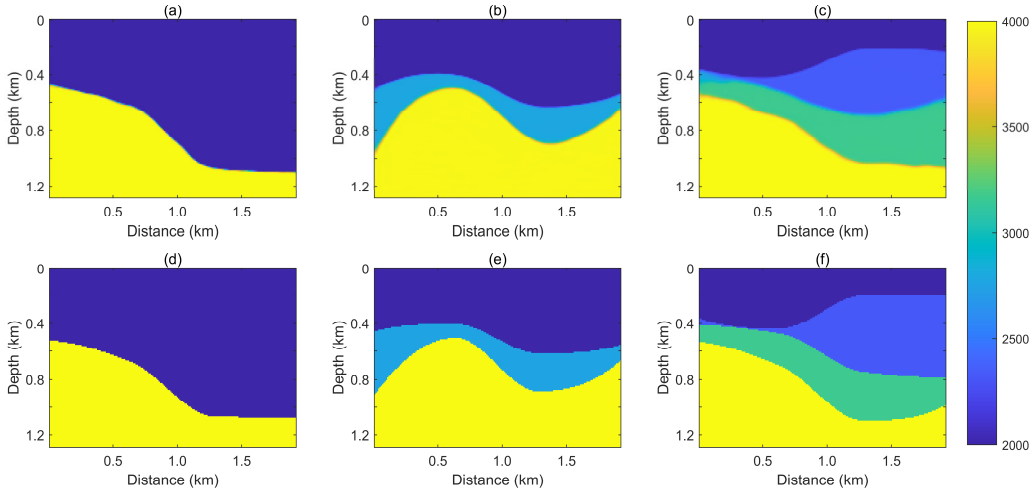


Figure 10. The modeling results for single-event test dataset of multi-layer random stratigraphic models. (**a**–**c**): DL predicted results; (**d**–**f**): real velocity models.

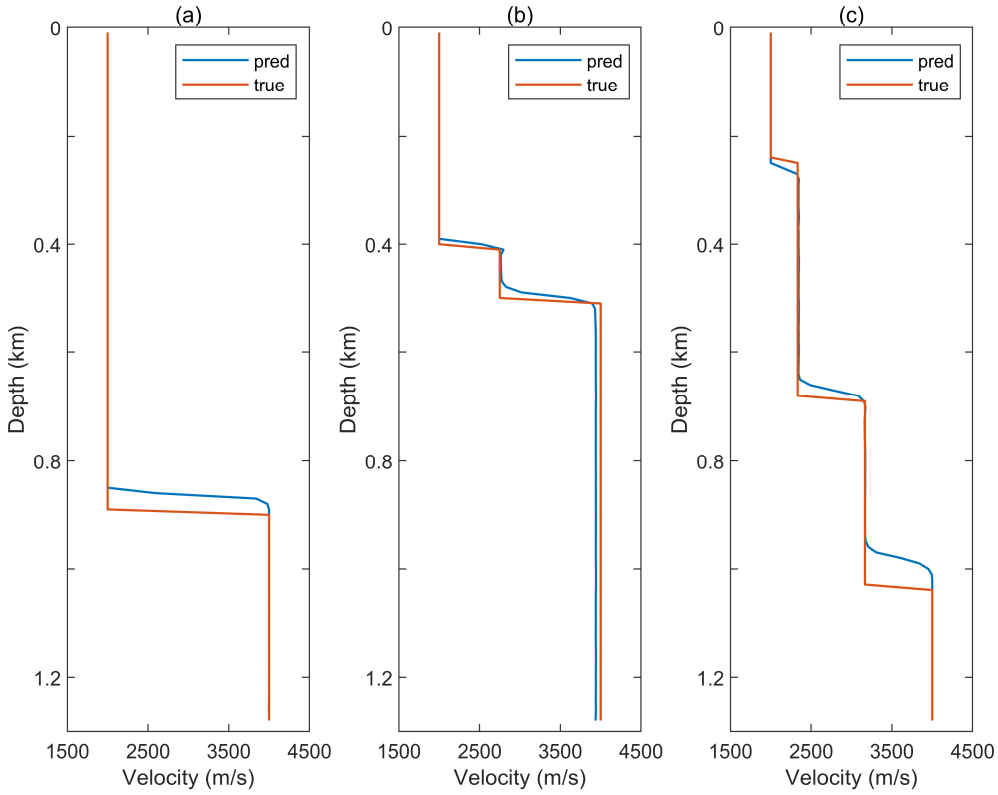


Figure 11. Velocity profile comparisons for single-event test dataset of multi-layer random stratigraphic models: (**a**) at a horizontal distance of 0.96 km; (**b**) at a horizontal distance of 0.6 km; (**c**) at a horizontal distance of 1.1 km.

The comparison of microseismic modeling results with different numbers of layers reveals that for models with fewer layers, such as two or three (Figure 10a,b), the neural network's outputs closely resemble the real models. However, in the case of the four-layer model (Figure 10c), there are discrepancies in depicting the topography and larger errors occur at the model's edges. The profile plots indicate that the modeling results provide reasonably accurate predictions for velocity magnitudes, although there may be some deviations in the positions of the geological layers (Figure 11a,c). Overall, the predictions are reasonably accurate.

Similarly, in the multi-event test dataset, we selected one velocity model each with two, three, and four layers for presentation in Figure 12. The corresponding velocity profiles are shown in Figure 13.

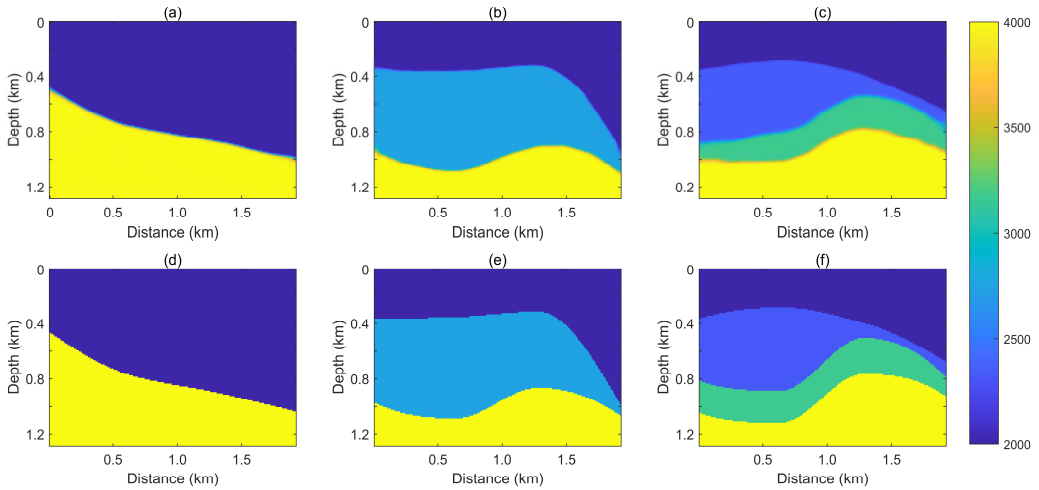


Figure 12. The modeling results for multi-event test dataset of multi-layer random stratigraphic models. (**a**–**c**): DL predicted results; (**d**–**f**): real velocity models.

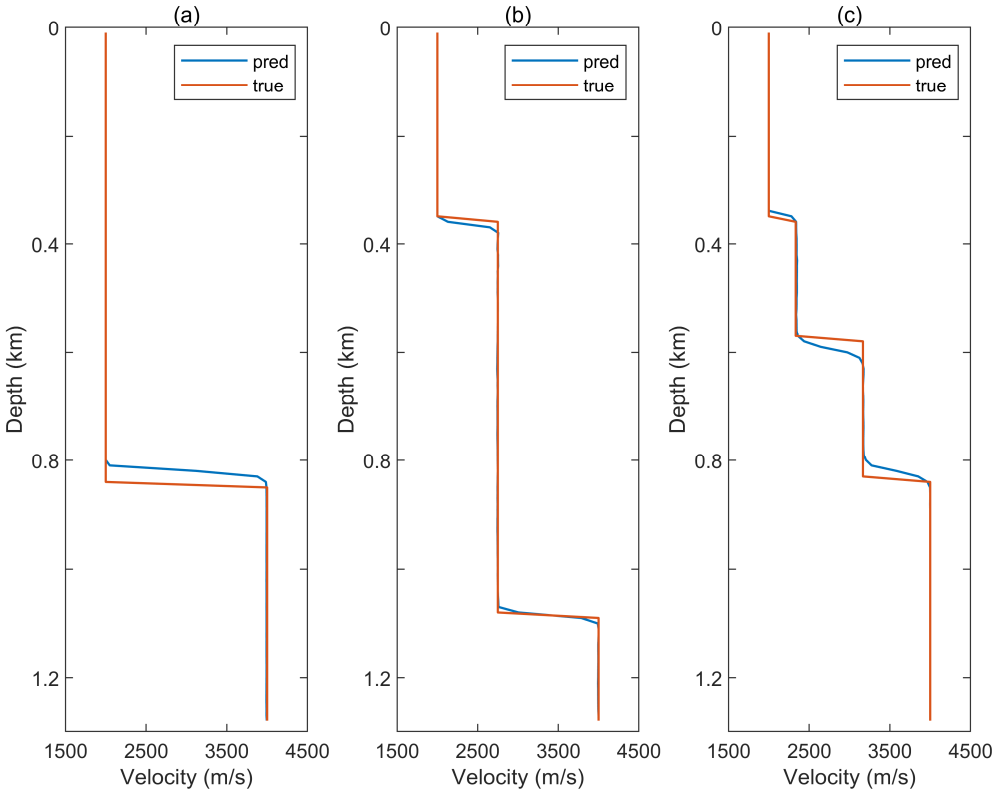


Figure 13. Velocity profile comparisons for multi-event test dataset of multi-layer random stratigraphic models (**a**) at a horizontal distance of 0.96 km; (**b**) at a horizontal distance of 0.6 km; (**c**) at a horizontal distance of 1.1 km.

Similar to the results obtained using single-event microseismic records as the labeled dataset, the multi-event microseismic modeling results are most accurate when dealing with fewer layers (Figure 12a,b). However, when faced with undulating four-layer geological structures (Figure 12c), there are minor inaccuracies in edge characterization. Nevertheless, the overall modeling results remain reliable.

To address the issue of inaccuracies in edge characterization, several approaches can be considered. One option is to train the model with a larger dataset that includes a significant proportion of geological structures with four layers or more. This can enhance the model's ability to represent complex geological structures effectively. Another avenue of improvement is refining our neural network architecture. Additionally, a complementary approach involves combining DL results with traditional methods. For instance, using the outcomes from DL as an initial model for FWI can lead to precise subsurface velocity models.

3.3.2. Local Marmousi Models

The descent of the loss function during the training process for the local Marmousi models is shown in Figures 14 and 15. Figure 14 shows the loss for the single-event microseismic dataset, while Figure 15 shows the loss for the multi-event microseismic dataset.

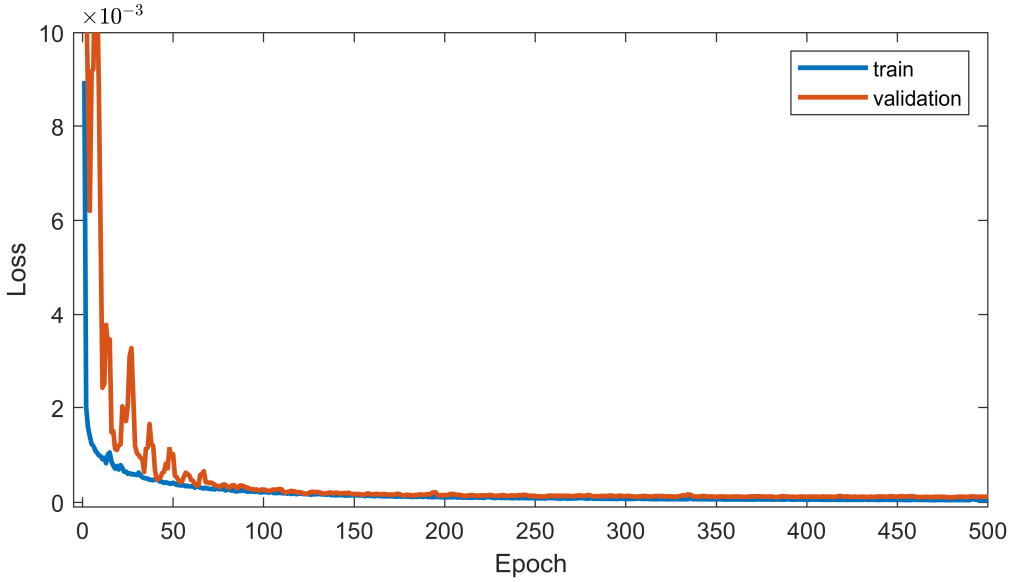


Figure 14. Loss decreases during the training process for single-event dataset of local Marmousi models.

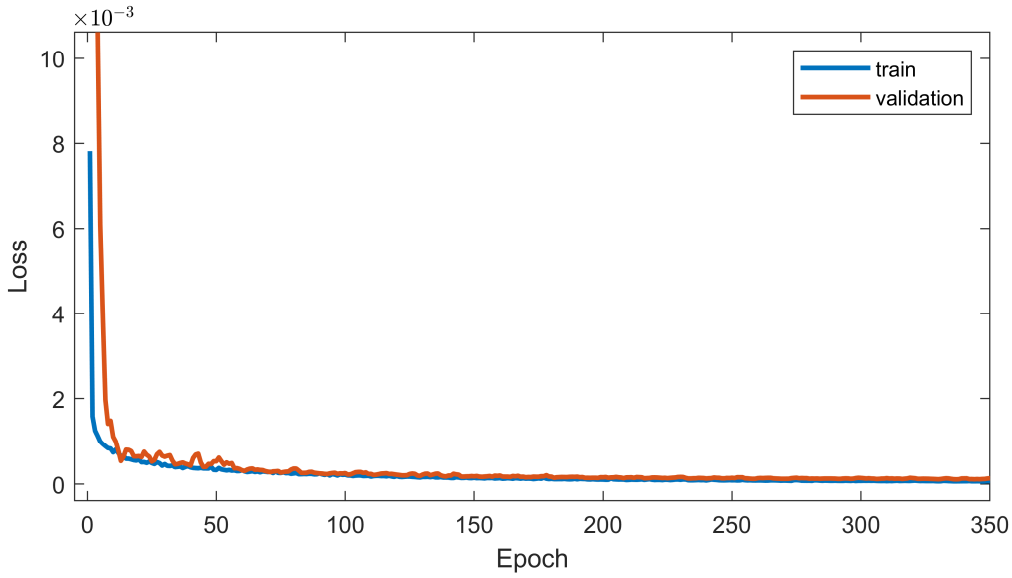


Figure 15. Loss decreases during the training process for multi-event dataset of local Marmousi models.

Similar to the loss curves observed in the multi-layer random stratigraphic models, training with the multi-event dataset results in a smoother descent and faster convergence of the model's loss function compared to the single-event dataset. However, it typically

requires more iterations than the multi-layer random stratigraphic models. This is due to the diverse nature of the extracted Marmousi velocity models in terms of the number of layers, velocity distributions, and angle variations, unlike the linear velocity variations in the former case. Consequently, finding the optimal configuration necessitates additional iterations.

Three randomly selected modeling results from the single-event test set are shown in Figure 16. The velocity model obtained from using single-event microseismic records as the input generally yields accurate results. However, there are discrepancies observed in predicting the locations of geological layer boundaries (Figure 16c) and the magnitude of velocity within these layers (Figure 16a,b). These deviations indicate that although the overall modeling captures the characteristics of the velocity structure, there may be limitations in accurately predicting specific details at geological boundaries and within the layers. This may still be related to the limited ability to capture features in single-event microseismic records.

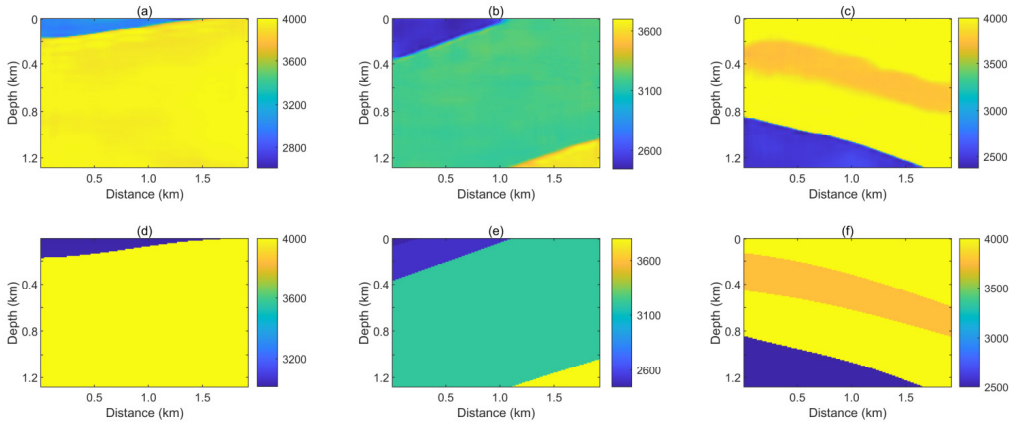


Figure 16. The modeling results for single-event test dataset of local Marmousi models. (**a**–**c**): DL predicted results; (**d**–**f**): real velocity models.

Three randomly selected modeling results from the multi-event test set are shown in Figure 17. The velocity models obtained using multi-event microseismic records as input demonstrate increased reliability. Whether dealing with simpler models that include low-velocity layers (Figure 17a) or more complex models with nonlinear velocity variations across multiple layers (Figure 17b,c), these velocity models accurately depict the angles and distributions of geological layers. This indicates that using multi-event microseismic records as input enhances the modeling capabilities for capturing intricate geological structures and velocity variations.

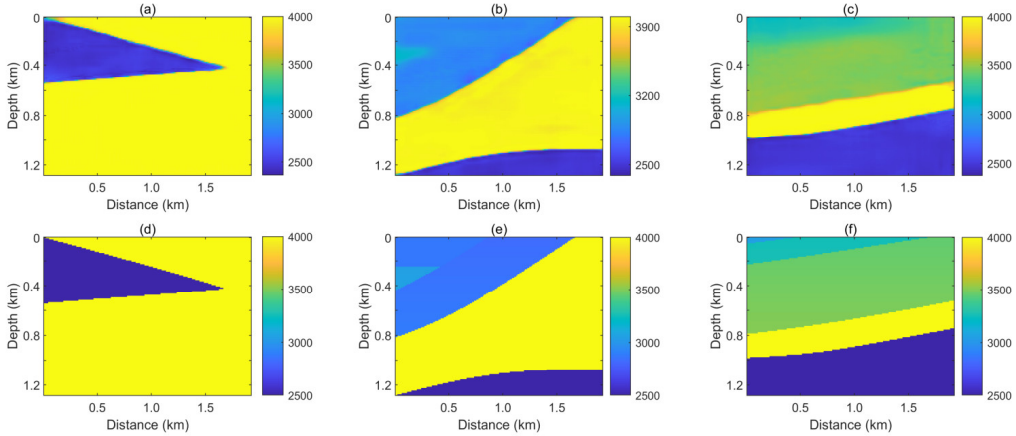


Figure 17. The modeling results for multi-event test dataset of local Marmousi models. (**a**–**c**): DL predicted results; (**d**–**f**): corresponding real velocity models.

FWI is currently considered the most accurate technique for seismic velocity modeling. This approach minimizes the discrepancies between simulated and observed data through gradient optimization techniques, ultimately producing precise velocity models. In this numerical test, a linear initial model was chosen for FWI, with a velocity model (Figure 17f) serving as the true model. Using 17b as the initial model for FWI, the inversion results are shown in Figure 18.

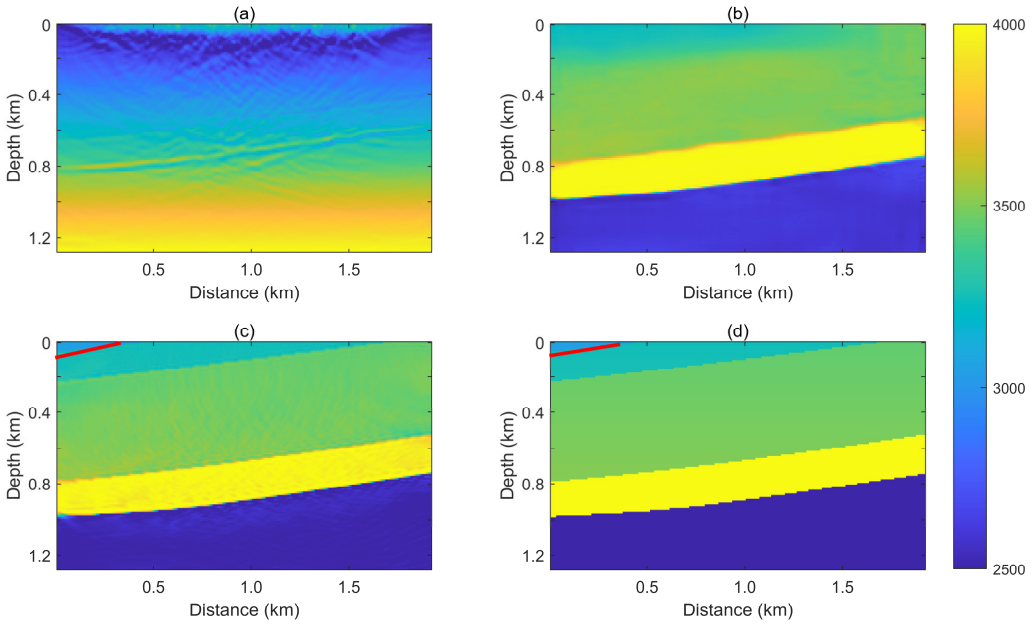


Figure 18. Comparison of modeling results using different methods: (**a**) FWI result with linear initial model; (**b**) modeling result from this paper; (**c**) FWI result with (**b**) as initial model; (**d**) true velocity model.

The results demonstrate that, in the absence of a reliable initial model and when the sub-wavelet frequency is excessively high, FWI can only reconstruct the geological interfaces of high-velocity layers, leading to lower modeling accuracy. In comparison, the modeling results obtained using the method proposed in this paper closely resemble the true model. By using this method as an initial model for FWI, it becomes possible to achieve a more precise characterization of geological interfaces and inter-layer velocities.

Compared to the microseismic inversion research by Daniel Wamriew et al. [19], they used a single network to simultaneously tackle source localization and velocity model updating tasks, while our focus lies solely on microseismic velocity modeling. They trained their network using independent microseismic records separated from multiple microseismic events within the velocity models, whereas we have considered how to update velocity models using coupled records when microseismic events are intertwined and inseparable. In addition, their research focused on VTI models, this paper used TTI models that including anticlines, synclines, and anomalous velocity models. This enables the model to effectively address geological diversity, making microseismic monitoring applicable to a wide range of fields.

3.4. Noise Robustness Test

The main objective of this paper is to showcase the potential of the modified Attention U-Net for seismic velocity modeling. Consequently, all previous numerical simulations were conducted in a noise-free environment. However, in practical applications, recorded microseismic data often contain significant amounts of noise. Therefore, in the final numerical test, Gaussian noise with a signal-to-noise ratio (SNR) of 15 was added to the multi-event test dataset of the multi-layer random stratigraphic models to evaluate the model's robustness to noise. Figure 19 demonstrates that even with the presence of added

noise, the model can still accurately predict simple two-layer velocity models. In the modeling results of multi results of multi-layer velocity models, the velocity distribution generally remains accurate, although there is a noticeable deviation in representing the undulations of deeper geological layers. Future research endeavors will focus on improving microseismic data denoising techniques and fortifying the model against complex geological structures' noise interference.

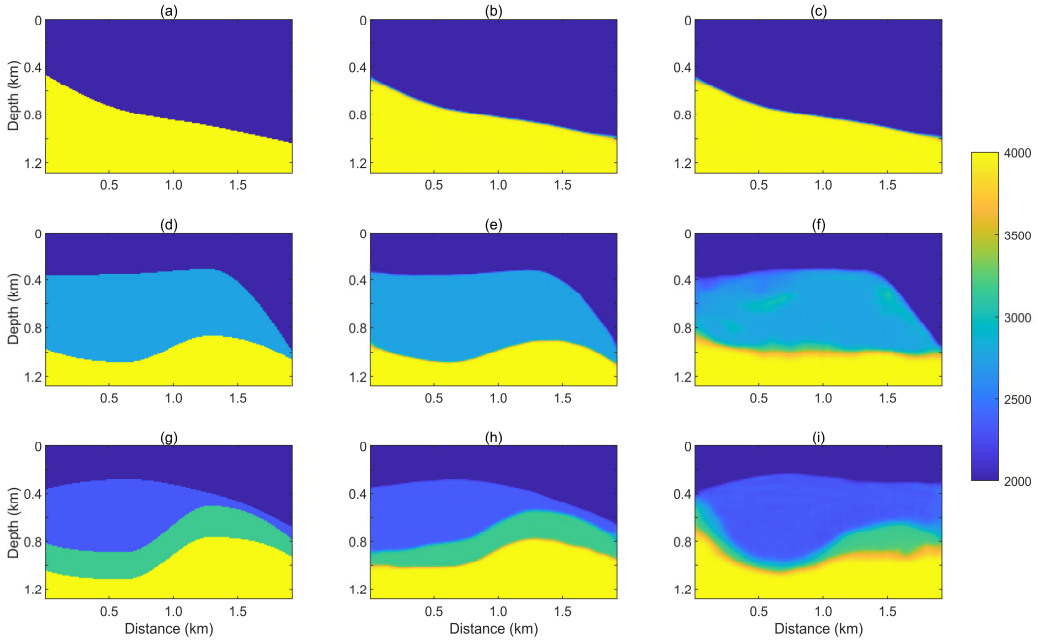


Figure 19. Noise robustness test for multi-layer random stratigraphic multi-event models. (**a**–**c**): True velocity models; (**d**–**f**): modeling results without added noise; (**g**–**i**): results with added gaussian noise at SNR = 15.

4. Conclusions

This paper used finite-difference forward simulations to generate microseismic records, and a modified Attention U-Net was used to successfully predict various complex subsurface velocity fields, including anticlines, synclines, and anomalous velocity models. Numerical tests demonstrated that the modified Attention U-Net effectively captures the hidden spatial correlations and features between microseismic records and velocity models, consistently yielding reliable velocity models for both single-event and coupled multi-event microseismic records. Additionally, the use of multi-event microseismic datasets demonstrates a greater capability for characterizing subsurface features while requiring less training time.

Compared to traditional methods, DL algorithms reduce the reliance on manual feature extraction and data preprocessing, resulting in a more automated and efficient microseismic velocity modeling process. They require only a reasonable amount of training time, and once trained, the model can complete predictions in approximately one second. This addresses the pressing need for real-time decision-making in microseismic monitoring.

Moving forward, our future research will focus on exploring methods to improve the model's noise resistance and investigating effective combinations of DL and traditional techniques such as FWI.

Author Contributions: Methodology, L.H. and P.Z.; Resources, J.Z. and X.S.; Software, Y.Z.; Validation, W.H.; Writing—Original Draft, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (No. 42130805, No. 42074154, No. 42374147, No. 42004106), the Natural Science Foundation of Jilin

Province (No. YDZJ202101ZYTS020), and the Lift Project for Young Science and Technology Talents of Jilin Province (No. QT202116).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shu, L.; Liu, Z.; Wang, K.; Zhu, N.; Yang, J. Characteristics and classification of microseismic signals in heading face of coal mine: Implication for coal and gas outburst warning. *Rock Mech. Rock Eng.* **2022**, *55*, 6905–6919. [CrossRef]
2. Kang, H.; Jiang, P.; Feng, Y.; Gao, F.; Zhang, Z.; Liu, X. Application of large-scale hydraulic fracturing for reducing mining-induced stress and microseismic events: A comprehensive case study. *Rock Mech. Rock Eng.* **2023**, *56*, 1399–1413. [CrossRef]
3. Xu, H.; Lai, X.; Zhang, S.; Shan, P.; Wu, Z.; Xu, H.; Bai, R.; Guo, Q. Precursor information recognition of Rockburst in the coal-rock mass of Meizoseismal area based on multiplex microseismic information fusion and its application: A case study of Wudong coal mine. *Lithosphere* **2022**, *2022*, 7349759. [CrossRef]
4. Chen, J.; Crawford, W.C.; Cannat, M. Microseismicity and lithosphere thickness at a nearly-amagmatic oceanic detachment fault system. *Nat. Commun.* **2023**, *14*, 430. [CrossRef] [PubMed]
5. Wang, H.; Alkhalifah, T. Direct microseismic event location and characterization from passive seismic data using convolutional neural networks. *Geophysics* **2021**, *86*, KS109–KS121. [CrossRef]
6. Li, J.; Stankovic, L.; Pytharouli, S.; Stankovic, V. Automated platform for microseismic signal analysis: Denoising, detection, and classification in slope stability studies. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 7996–8006. [CrossRef]
7. Zhang, C.; van der Baan, M. Microseismic denoising and reconstruction by unsupervised machine learning. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1114–1118. [CrossRef]
8. Gajek, W.; Malinowski, M. Errors in microseismic events locations introduced by neglecting anisotropy during velocity model calibration in downhole monitoring. *J. Appl. Geophys.* **2021**, *184*, 104222. [CrossRef]
9. Yang, F.; Ma, J. Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics* **2019**, *84*, R583–R599. [CrossRef]
10. Garotta, R.; Michon, D. Continuous analysis of the velocity function and of the move out corrections. *Geophys. Prospect.* **1967**, *15*, 584. [CrossRef]
11. Cook, E.E.; Taner, M.T. Velocity spectra and their use in stratigraphic and lithologic differentiation. *Geophys. Prospect.* **1969**, *17*, 433–448. [CrossRef]
12. Chiao, L.-Y.; Kuo, B.-Y. Multiscale seismic tomography. *Geophys. J. Int.* **2001**, *145*, 517–527. [CrossRef]
13. Tarantola, A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **1984**, *49*, 1259–1266. [CrossRef]
14. Claerbout, J.F. Synthesis of a layered medium from its acoustic transmission response. *Geophysics* **1968**, *33*, 264–269. [CrossRef]
15. Schuster, G.T. Theory of daylight/interferometric imaging-tutorial. In Proceedings of the 63rd EAGE Conference & Exhibition, Amsterdam, The Netherlands, 11–15 June 2001. [CrossRef]
16. Zhang, Z.; Rector, J.W.; Nava, M.J. Simultaneous inversion of multiple microseismic data for event locations and velocity model with Bayesian inference. *Geophysics* **2017**, *82*, KS27–KS39. [CrossRef]
17. Grechka, V.I.U.; Heigl, W.M. *Microseismic Monitoring*; Society of Exploration Geophysicists Tulsa: Houston, TX, USA, 2017. [CrossRef]
18. Wang, H.; Guo, Q.; Alkhalifah, T.; Wu, Z. Regularized elastic passive equivalent source inversion with full-waveform inversion: Application to a field monitoring microseismic data set. *Geophysics* **2020**, *85*, KS207–KS219. [CrossRef]
19. Wamriew, D.; Charara, M.; Pissarenko, D. Joint event location and velocity model update in real-time for downhole microseismic monitoring: A deep learning approach. *Comput. Geosci.* **2022**, *158*, 104965. [CrossRef]
20. Liu, J.-J.; Liu, J.-C. Integrating deep learning and logging data analytics for lithofacies classification and 3D modeling of tight sandstone reservoirs. *Geosci. Front.* **2022**, *13*, 101311. [CrossRef]
21. Yang, W.; Xia, K.; Fan, S. Oil logging reservoir recognition based on TCN and SA-BiLSTM deep learning method. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105950. [CrossRef]
22. Mousavi, S.M.; Beroza, G.C. Deep-learning seismology. *Science* **2022**, *377*, eabm4470. [CrossRef]
23. Röth, G.; Tarantola, A. Neural networks and inversion of seismic data. *J. Geophys. Res. Solid Earth* **1994**, *99*, 6753–6768. [CrossRef]
24. Moya, A.; Irikura, K. Inversion of a velocity model using artificial neural networks. *Comput. Geosci.* **2010**, *36*, 1474–1483. [CrossRef]
25. Araya-Polo, M.; Jennings, J.; Adler, A.; Dahlke, T. Deep-learning tomography. *Lead. Edge* **2018**, *37*, 58–66. [CrossRef]
26. Zhang, Z.; Wu, Y.; Zhou, Z.; Lin, Y. VelocityGAN: Subsurface velocity image estimation using conditional adversarial networks. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 705–714. [CrossRef]

27. Li, S.; Liu, B.; Ren, Y.; Chen, Y.; Yang, S.; Wang, Y.; Jiang, P. Deep-learning inversion of seismic data. *arXiv* **2019**, arXiv:1901.07733. [CrossRef]
28. Fabien-Ouellet, G.; Sarkar, R. Seismic velocity estimation: A deep recurrent neural-network approach. *Geophysics* **2020**, *85*, U21–U29. [CrossRef]
29. Chen, Y.; Schuster, G.T. Seismic inversion by Newtonian machine learning. *Geophysics* **2020**, *85*, WA185–WA200. [CrossRef]
30. Cao, D.; Su, Y.; Cui, R. Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints. *J. Pet. Sci. Eng.* **2022**, *209*, 109836. [CrossRef]
31. Zhang, J.; Zhao, X.; Chen, Y.; Sun, H. Domain knowledge-guided data-driven prestack seismic inversion using deep learning. *Geophysics* **2023**, *88*, M31–M47. [CrossRef]
32. Wamriew, D.; Pevzner, R.; Maltsev, E.; Pissarenko, D. Deep neural networks for detection and location of microseismic events and velocity model inversion from microseismic data acquired by distributed acoustic sensing array. *Sensors* **2021**, *21*, 6627. [CrossRef]
33. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999. [CrossRef]
34. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; pp. 234–241.
35. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
36. Jetley, S.; Lord, N.A.; Lee, N.; Torr, P.H. Learn to pay attention. *arXiv* **2018**, arXiv:1804.02391. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.