Disentangling Noise Patterns From Seismic Images:
Noise Reduction and Style Transfer

Haiwen Du, Yu An, Qing Ye, Jiulin Guo, Lu Liu, Dongjie Zhu, Conrad Childs, John Walsh, and Ruihai Dong, Member, IEEE

Abstract—Seismic interpretation is a fundamental approach for obtaining static and dynamic information about subsurface reservoirs, such as geological faults/salt bodies and associated fluid types and distribution. Due to the exponential growth in seismic data volume and considerable uncertainty in manual interpretation, deep learning (DL) algorithms have been introduced to assist seismic interpretation. Our investigation of the trained neural networks suggests that they underperform on seismic data with different noise characteristics. One of the main issues is that the noise patterns of seismic data are highly inconsistent due to many factors, including geological features, sampling parameters, and human intervention. To address this problem, we propose a noise pattern transfer (NPT) framework to transfer or remove seismic noise style between datasets by treating noise patterns as styles of image, which can also improve the generality of automatic seismic interpretation algorithms. Extensive experiments on three synthetic datasets and two field seismic datasets demonstrate the promising performance of our proposed NPT approach. Pairs of clean and stylized seismic data are generated by extending the use of the neural style transfer algorithm beyond the artistic domain. We then demonstrate how our method achieves superior noise pattern transferability between datasets and denoising performance on field datasets. Associated improvements in accuracy and generalization of the neural-network-based fault recognition tasks successfully demonstrate the practicality of our NPT approach. The source code is made publicly available online at https://github.com/Magnomic/npt-code.

Index Terms—Convolutional neural networks (CNNs), deep learning (DL), image processing, neural style transfer (NST), seismic image.

I. INTRODUCTION

SEISMIC data can provide static and dynamic information for subsurface reservoirs, such as their lateral extent and thickness, the location and displacements of faults, and the distribution of reservoir properties, including porosity and clay content [1]. Deep learning (DL) approaches based on signal/image processing are proposed to assist geoscientists in achieving robust seismic interpretation with less reliance on the interpreter’s experience and knowledge [2], [3]. Differences in geological structures, sampling parameters, and artifacts mean that seismic images show a wide range of resolutions and levels of noise corruption [4], [5]. The state-of-the-art approaches using computer-vision-based supervised DL models are now helping to address the above shortcomings of seismic imaging [6]. However, the performance of the supervised DL methods will deteriorate when the seismic features extend beyond those of the training data [7]. In these circumstances, two fundamental challenges must be overcome.

1) The impossibility of acquiring noise-free field seismic data.

2) The unsatisfactory performance of pretrained neural networks when applied to seismic data with different geological and seismic characteristics.

For the first challenge, researchers use synthetic seismic data and simulated noise (e.g., Gaussian white noise) to form the pairwise training data [8]. Although this method can make the denoised data cleaner, it cannot correctly distinguish specific noise patterns from the actual signal. For the other challenge, some articles suggest transfer learning techniques [9], [10], in which additional training on seismic data is used to enrich neural network learning. However, explaining the learning process regarding what features are considered and how they are learned is difficult.

Consequently, this article focuses on noise pattern differences between field seismic datasets. Our objective is to learn noise representations and transfer or remove them, with two allied key questions: 1) how can we successfully isolate geological information and noise patterns from the input seismic patches? and 2) how can we determine whether the generated noise matches the target field seismic dataset?

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
For the first question, we introduce the image-to-image noise transfer (I2I-NT) block, which inherits the concept of neural style transfer (NST) to generate pairwise field seismic data with different noise patterns. To tackle the second question, we develop a dataset-to-dataset reverse noise transfer (D2D-RNT) path that learns datasetwise noise pattern representation from the pairwise seismic data we generated after I2I-NT. To the best of our knowledge, this is the first method that achieves explainable noise transferring between field seismic datasets.

The highlights of this article are as follows.

1) Provision of a novel noise pattern transfer (NPT) approach to disentangle noise patterns from field seismic reflection data.

2) Consideration of the limitations of the Gaussian-noise-based pairwise synthetic data generating methods and the extension of NST technique usage to generate plausible and explainable seismic data with different noise patterns.

3) Demonstration that by comparison to the state-of-the-art methods, the proposed method offers significant improvements in denoising and transfer capability, with near real-time speed and reasonable memory requirements.

4) Highlight how the NPT method provides statistically significant improvements in the accuracy and generalization of neural networks for recognizing geological faults.

The rest of this article is arranged as follows: Section II discusses previous work related to our proposed method, Section III details our model architecture, Section IV outlines our experiments, and Section V provides the principal conclusions of our work.

II. RELATED WORKS

A. Transfer Learning in Seismic Data

Due to considerable variation in seismic data, the DL-based seismic interpretation methods tend to suffer from poor generalization over different field seismic data. Many recent studies have applied transfer learning techniques to alleviate this problem. Transfer learning is a machine learning technology designed to improve the ability of the current machine learning algorithms to generalize to new tasks/domains by transferring knowledge from a known task to another task [11], [12]. It relaxes the demands of machine learning, and DL models in particular, on the amount of labeled training data. Transfer learning on DL methods is also known as deep transfer learning [12].

We reviewed 14 articles that applied deep transfer learning to seismic data, 12 of which involved network-based deep transfer learning approaches [9], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], and two used adversarial-based approaches [10], [24]. The network-based approach is the most popular deep transfer learning approach for seismic data interpretation due to its relative simplicity and good performance. All or part of the pretrained weights are used to initialize the DL models and are subsequently fine-tuned with data pairs from the target domain. The main limitation of this method is the requirement for labeled seismic data. Our proposed approach does not require labeled seismic data.

The adversarial-based approach is less popular than other deep transfer learning approaches because of its complexity [9]. This type of approach, inspired by generative adversarial neural networks (GANs) [25], uses a domain adversarial neural network (DANN) [26] as a backbone to learn generic features from two different domains [27]. The advantage of this method is that no annotation of the target domain is required and a further benefit of our implementation is that it considers disentangling noise patterns to solve the noise reduction problem.

B. Image Style Transfer Methods

Style transfer in computer vision (CV) is the task of changing the style of an image from one domain to another domain [28]. It is derived from visual modeling [29] and image reconstruction [30] tasks. With the development of convolutional neural networks (CNNs), GAN [25] and NST [31] have replaced the traditional nonrealistic rendering (NPR) approach [32] as popular methods for rendering different styles of content images. According to the different style transferring processes of the models, we summarize them into two classes: 1) online neural methods and 2) offline neural methods.

As represented by NST, the online neural methods have become popular recently. NST calculates the style loss with a style image and content loss with a content image, thereby generating the target image by minimizing these losses. We cannot, however, control the position of the style extracted from the style image [33] because the gram matrix ignores the positions of the features [34]. As a result, online methods are widely used by several entertainment applications that offer intuitive image art style transfer, despite the high computational cost of this method. The most important limitations of the online model are, therefore, the issues of efficiency and localization (i.e., location preservation) [28].

For the offline neural methods, the major difference is that they reconstruct the stylized result by a feed-forward network. These methods use the pairwise original images and generated (stylized) images to train the neural network [35]. The complexity of the offline methods is lower than online equivalents because no iterative calculation is required. Like NST, the GAN-based method trains a generator to produce high-quality synthetic images and a discriminator to determine whether the generated images are authentic, showing good universal applicability. Nevertheless, for seismic data, the depth-related noise patterns may cause false and extensive omission in GAN-generated data [36].

C. Denoising Models Based on CNN

1) Computer-Vision-Based Denoising Models for Seismic Data: The CV-based denoising models have been the mainstream seismic denoising approaches in recent years and have attracted more researchers to explore this field. These CV-based models can be classified into two groups: 1) unsupervised self-similarity-based denoising methods and 2) supervised DL-based denoising methods.

The self-similarity methods are proven to be asymptotically optimal under a generic statistical image model [37]. As a
consequence, block-matching and 3D filtering (BM3D) [38] is widely used as a baseline of the state-of-the-art denoising models. It integrates similar blocks into a 3-D matrix, performs filtering in 3-D space, and then inversely transforms the result into two dimensions to get a clean patch. Similarly, tensor decomposition and total variation (TDTV) [39] uses Candecomp/Parafac (CP)-decomposition processed patches to find similar patches within a fixed range. These patches are used to estimate the target patch and construct a loss function to smooth the seismic image. Noise-to-noise [40] is another popular natural image denoising technique that combines the self-similarity and DL methods. Fang et al. [41] have proposed a CNN-based self-similarity method to realize unsupervised seismic denoising using multiple similar noisy data as input–label pairs. These methods suffer from high computing complexity and an assumption that noise should follow the zero-mean distribution and are independent of the signals.

The DL-based denoising methods can learn noise features and remove them from a noisy image by automatically adjusting numerous parameters. Unlike the traditional prediction-based, transformation-based, and low-rank-based models, the DL-based models do not need to redesign the complex hyperparameters according to the new features when the dataset changes. They show good scalability because we can optimize the models using a new dataset to retrain them or applying transfer learning. Most state-of-the-art DL-based denoising models [42] select CNNs because they automatically detect the essential features without any human supervision [43], [44]. Researchers design new network frameworks to optimize DL models’ noise feature extraction ability for better denoising performance. Among them, multigranularity feature fusion CNN (MFFCNN) [6] designed a new CNN layer architecture with different kernel sizes in noise feature extraction, which has been verified by the atrous spatial pyramid pooling (ASPP) architecture proposed by Chen et al. [45]. Generative adversarial network (GAN) [25] is another way to train denoising models, with Wang et al. [46] proposing a GAN-based denoising model to solve the poor continuity of events problem. However, this method needs pairwise noisy and clean data for the supervised training process, which is impossible to obtain from field seismic data.

2) Pairwise Noisy Seismic Data Generating Methods in DL: Although seismic data with different noise levels can be generated by changing shot density differences [47], any such noise is not caused by sensors and cannot be removed because it is contained within both the noisy and clean data. Splitting noise from seismic images is not, therefore, an easy task. Researchers generate clean synthetic seismic data and add noise to them to form pairwise datasets to address the problem, but the quality of the pairwise data significantly impacts the denoising model’s effect. For clean data, there is a consensus that randomized curves with the Ricker wavelet can fit the clean field seismic data well. The principal challenge relates to adding noise that is close to real noise. Here, we summarize the associated methods in two classes: 1) addition of synthetic white noise and 2) extraction of the noise from field seismic data.

Gaussian white noise is a commonly used synthetic noise pattern in both natural and seismic images [48]. The standard methods can be summarized as 1) generating lines; 2) shearing shifts; 3) adding Ricker wavelet; and 4) adding Gaussian noise [8]. It is a popular choice for the recently proposed denoising models to generate pairwise noisy–clean datasets and train supervised denoising models [6], [49]. Besides, Bugge et al. [50] include elements of wave propagation physics to generate pairwise datasets using synthetic data and target noise distribution, which can denoise the target noise well.

For field seismic noise, Dong et al. [51] use the field seismic data under the condition without sources as pure noise to generate noisy synthetic data. Then, the authors train a DnCNN denoising model with these pairwise data. Nonetheless, the noise can be coherent in the seismic signal [52], which prompts us to consider that we should find a noise-adding method that combines both the noise feature and the signal feature of the seismic data. Wang et al. [53] use GAN to generate field noise to synthetic data. Then they train a CNN to perform denoising. GAN uses the generator G, discriminator D, and adversarial loss to generate high-quality synthetic data.

Nevertheless, these approaches only work well on seismic data from desert environments, which show signal-sparse and low SNR [54] and are therefore more amenable to noise extraction. Cycle-GAN [55], [56], [57] is another novel approach that does not need pairwise data to train a denoising network. Cycle-GAN uses two discriminators, two generators, and cycle-consistency loss to achieve plausible image-to-image translation. However, we observed that this method lacks control over the generated seismic images and only runs well on synthetic datasets. Instead, our method generates an intermediate product, a plausible paired seismic dataset, which can be easily quality controlled by geologists.

III. METHODOLOGY

Noise in field seismic data varies with many underlying factors, which together contribute significantly to the diversity of seismic data circumstances and greatly hinder the generalization of different DL-based automatic seismic interpretation models. Inspired by NST, we consider noise patterns as styles of seismic datasets. To minimize the variation between seismic datasets caused by unique noise patterns, we proposed an NPT approach to disentangle the noise patterns from the seismic images and replace specific noise patterns. Specifically, when we use (1) to represent the relationship between field seismic images $Y$, geological signals $X$, and noise $N$ mathematically, the distance between two seismic images $Y = X + N$ and $Y' = X' + N'$ should be smaller if we replace noise pattern $N$ with $N'$

$$Y = X + N.$$  

In response to this assumption, in this section we introduce the overall network structure of our proposed NPT model in Fig. 1. NPT aims to transfer the noise style of field seismic
image in the dataset \( S \) source to the noise style in the dataset \( T \) target. Therefore, it is easy to understand that NPT can be applied to the denoising tasks when dataset \( T \) is set to a synthetic and noise-free dataset. NPT consists of two key components: an I2I-NT block for generating plausible pairs of seismic images with different noise patterns from datasets \( S \) and \( T \) and a D2D-RNT block for implementing dataset-level NPT. We will describe each in detail below.

### A. Image-to-Image Noise Transfer (I2I-NT)

1) **Network Structure:** As discussed in Section II-C, disen-tangling noise from field seismic data is not a straightforward task, as it is impossible to obtain noise-free field seismic data. In this article, we make the novel proposal to use the I2I-NT task, as it is impossible to obtain noise-free field seismic data.

In this article, we make the novel proposal to use the I2I-NT task, as it is impossible to obtain noise-free field seismic data. In our approach, the seismic image (from dataset \( T \)) providing geological information is called the content image, while the other seismic image (from dataset \( S \)), providing another noise pattern, is referred to as the style image. They are fed into VGG-19 in batches together with the synthetic image \( \hat{Y} \) (initialized with the content image), which is back-propagated into a plausible synthetic seismic image using the loss function of (3). \( \mathcal{L}_c \) represents the content loss for retaining the geological information, while \( \mathcal{L}_r \) represents the style loss, which minimizes the noise pattern difference between \( Y' \) and \( \hat{Y} \). \( \alpha \) and \( \beta \) are the weights, and \( l_c \) and \( l_s \) are the layers used to compute \( \mathcal{L}_{cs} \) and \( \mathcal{L}_{cs} \).

\[
\mathcal{L}_c(Y, Y', \hat{Y}, l_c, l_s) = \alpha \mathcal{L}_c(Y, \hat{Y}, l_c) + \beta \mathcal{L}_r(Y', \hat{Y}, l_s).
\] (3)

2) **Content Loss:** As introduced earlier, we can use feature maps and filters to reconstruct the seismic image. Deeper layers, which preserve the geological information of the seismic images, are used. The content loss, therefore, is calculated as (4). Feature maps are represented as \( F(Y) \), the element in the \( i \)th filter at position \( j \) in layer \( l \) of image \( Y \) is represented by \( F_{i \in [0, P_l], j \in [0, M_l]}(Y) \), where \( M_l \) is the size of the feature map \( F_l \), and \( P_l \) is the number of filters in layer \( l \).

\[
\mathcal{L}_c(Y, \hat{Y}, l_c) = \sum_{i,j} \left( F_{ij}^l(Y) - F_{ij}^l(\hat{Y}) \right)^2.
\] (4)

3) **Style Loss:** For the noise style loss \( \mathcal{L}_r \), we introduce the gram matrix proposed by the NST algorithm. The gram matrix was originally designed to extract artistic style features from paintings. It obtains style features independent of position.
DU et al.: DISENTANGLING NOISE PATTERNS FROM SEISMIC IMAGES

Fig. 2. Calculating process of the gram matrix. Here, we take conv_1_1 layer as an example. The number of feature maps changes with the layer we use (i.e., 128 for conv_2_1, 256 for conv_3_1, 512 for conv_4_1, and conv_5_1).

The overall style loss, which is calculated by (6), is the weighted sum of the noise pattern difference $E_l$ between $Y'$ and $\hat{Y}$ in layer $l$. $E_l$ can be calculated by (7)

$$\mathcal{L}_s(Y', \hat{Y}, l_s) = \sum_{i \in l_s} \omega_i E_l$$

$$E_l = \frac{1}{4P^2 l^2 M^2} \sum_{i,j} (G^l_{ij}(Y') - G^l_{ij}(\hat{Y}))^2.$$ (7)

### B. Dataset-to-Dataset Reverse Noise Transfer (D2D-RNT)

While the I2I-NT block can add noise from a style image to the geological signal of a content image, this method is a one-to-one mapping without knowledge of the noise style of the entire dataset. Therefore, we add a reverse noise transfer (D2D-RNT) path that can learn and transfer the noise pattern representation between the raw (i.e., dataset T) and stylized (i.e., noise pattern: S, signal: T) seismic datasets.

1) **Network Structure:** The right part of Fig. 1 illustrates the architecture of the proposed reverse NST path. In this article, we adopt the classic DnCNN model as the backbone of the D2D-RNT path. It is a denoising CNN modified based on the visual geometry group (VGG) network. It consists of 17 convolutional layers and additional batch normalization and rectified linear unit (ReLU) layers. The input to D2D-RNT is the stylized seismic image $\hat{Y} = X+N'$ and the original seismic image $Y = X+N$. DnCNN is trained to learn the noise pattern residuals between the original and stylized seismic images $\mathcal{R}(\hat{Y}, \phi) = N' - N$.

2) **Loss Function:** To train the DnCNN-based D2D-RNT path, we used the same mean square error loss function as the original DnCNN method to minimize the learned noise pattern residuals and the expected residuals. The loss function is summarized in (8), where $Q$ represents the number of seismic images

$$\mathcal{L}_\text{D2D-RNT} = \frac{1}{2Q} \sum_{i=1}^{Q} \| \mathcal{R}(\hat{Y}_i; \phi) - (\hat{Y}_i - Y_i) \|^2.$$ (8)

Ideally, the trained model should learn the datasetwise style representation as many one-to-one paired (stylized and raw) seismic images are being passed into the DnCNN model.

### IV. EXPERIMENTS

#### A. Datasets

In this article, five seismic datasets are involved in the experiments. Two are field seismic datasets (i.e., ThebeFault and Beatrice), and the other three are synthetic datasets [i.e., FaultSeg (FS), FaultSegClean, and faultseg synthetic (FSSynth)].

ThebeFault [58], [59], [60], to our knowledge, is the largest open-source field fault recognition seismic dataset available. The dataset contains raw seismic data from the Thebe gas field on the North West Shelf of Australia and annotations by expert fault interpreters from the University College Dublin Fault Analysis Group. The dataset size is 1803[inline] × 1537[sample] × 3174[crossline].

Beatrice is an oil field located in the Inner Moray Firth, Scotland [61]. The quality of this seismic dataset is relatively good, but there is significant migration noise, particularly in deeper layers. Geological faults exist in this dataset, but no corresponding fault annotations are available. The size of this dataset is 900[inline] × 1200[samples] × 1861[crossline].

FaultSeg [8] is an open-source 3-D synthetic seismic dataset intended for studying fault segmentation. It is generated by first defining a random horizontal reflection model with values ranging from $-1$ to $1$, then adding folding structures via composite Gaussian functions and adding fault elements, and finally using Ricker wavelets and Gaussian white noise to bring it closer to the actual seismic data. Many data pairs can be generated with different combinations of parameters. A model trained using this geologically diverse data should ideally perform relatively well on various real-world seismic datasets. The size of this dataset is 220 seismic cubes of size $128 \times 128 \times 128$, of which 200 are divided into a training set and the other 20 into a test set.

FaultSegClean [5] is a 2-D synthetic seismic dataset designed to achieve seismic data super-resolution and denoising. It involves generated 3-D synthetic seismic cubes following the same workflow as FaultSeg [8] without the last step of adding the Gaussian noise. 2-D patches were then extracted from the 3-D cubes with the size of the dataset comprising 1000 seismic patches of size $256 \times 256$.

The FSSynth dataset is a 2-D noise-free synthetic dataset generated mostly based on the FaultSeg dataset’s description. The generation process involves generating random straight and parabolic lines with values between $-1$ and $1$, adding the Ricker wavelets to the signal, and creating random geological faults. The size of the dataset comprises 1000 seismic patches of size $192 \times 192$. 
Fig. 3. Samples of the I2I-NT model. In the first row, we add Gaussian noise to the FSSynth patch. In the second and third rows, the style and content patches are from the Beatrice and Thebe datasets, respectively.

B. Environments and Configurations

In this section, we detail the environments and configurations of our model, which can help readers reproduce the model. For the hardware environment, all the experiments conducted in this article were performed on a Linux server, with compute unified device architecture (CUDA) version 10.2 using GeForce GTX 1080 Ti Graphics Cards. Codes are written based on the open-source machine learning framework PyTorch.

The parameters used in the experiments are as follows. For the I2I-NT block, $\alpha:\beta$ is set to $1:1e7$ for all the datasets, while $\omega_l = 1$ and $\omega_{l+1} = 0.9 * \omega_l$. $l_c$ includes conv_5_2, and $l_s$ includes conv_1_1, conv_2_1, conv_3_1, conv_4_1, and conv_5_1. We use the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-GBFS) to optimize the loss function, where the max iteration time is set to 20, and the max optimizing step is set to 10. The optimizing steps stop when the loss decreases less than 5% of the previous step. For hyperparameters, we recommend researchers to set $\alpha:\beta$ in a range of $1:1e4$–$1:1e8$. It depends on the noise pattern (e.g., the strength and density of noise) of dataset S. Higher $\beta$ makes the generated images blend in a stronger noise pattern from dataset S and vice versa. It should be noted that if we use an excessively high $\beta$, the generated image contains too much or intense noise, which makes the D2D-RNT model hard to distinguish between noise and signal, which leads to the result that part of the signal is regarded as noise and removed. This parameter recommendation is suggested based on related image style transfer literature [31], [62]. For patch selection, the style image and content image can be randomly selected from dataset S and dataset T. It is recommended to select patches from the area with lower SNR and higher uncertainty.

For the D2D-RNT model, the training batch size is 64 and the number of iterations is 8000. We use Adam optimizer, with the learning rate set to $1e-4$, and the default configurations of the baseline methods. We use the min-max normalization method to normalize data before feeding them into the neural network. Since THE DL models, including MFFCNN, ASPP, and multi-stage progressive image restoration (MPRNet), perform differently on different datasets, the training iteration numbers are modified to achieve the best performance.

C. Style Transfer Abilities

As we achieve noise transfer by suppressing the added noise pattern, I2I-NT and D2D-RNT procedures must be appropriately verified to explain how the noise patterns change. The most exciting part of our model is that we propose an I2I-NT model to disentangle noise patterns from the field seismic dataset (i.e., dataset S). It is, therefore, essential to ensure that the added noise shows the same or similar noise distribution to the target data. Since we cannot define the similarity of noise distribution when we add noise from a field seismic data patch, we select a known noise distribution,
DU et al.: DISENTANGLING NOISE PATTERNS FROM SEISMIC IMAGES 4513214

Fig. 4. Three samples of noise transfer results from our NPT models. The patches in the first row are patches from the three target datasets T, which are provided for reference.

Gaussian white noise, to verify it. We also provide other examples to help readers more easily see the effect of NPT on the field seismic data, as shown in Fig. 3.

From the samples in the first row of Fig. 3, the generated noise shows good normality for the Gaussian noise, which indicates that our model can transfer the target noise pattern to another patch. For the field data patch in the second and third rows, the generated patch looks noisier, and the generated noise shows a similar noise pattern to the style image. It is worth noting that the generated noise is related to the signal of the content patch. Moreover, the position of the generated noise is not the same as the style image. The reason is that the noise generation process is based on the gram matrix, which cannot perceive the position information of the noise. Since the results match our expectations in Section III, the noise pattern depends on the model’s transferring target, and the transferred patches look more like the target style.

Then, we test the style transferability of our NPT method on the Beatrice dataset. Here, we train three models, NPT (Bea. to Thebe) with dataset S: Beatrice, dataset T: Thebe, NPT (Bea. to FS) with dataset S: Beatrice, dataset T: FaultSeg, and NPT (Bea. to FSClean) with dataset S: Beatrice, dataset T: FaultSegClean. We provide examples in Fig. 4. From the results, we can clearly distinguish different structured noise patterns in the disentangled noise difference patches. We can see that the patches transferred to FaultSegClean contain the least noise since FaultSegClean is a noise-free dataset. The patches transferred to FaultSeg are noisier than the patches transferred to Thebe because FaultSeg has artificially added Gaussian noise.

D. Denoising Abilities

By analogy, we can apply our model to the denoising task by setting dataset S as a noisy field dataset and dataset T as a noise-free synthetic dataset. We verify whether DnCNN is a suitable network structure for the D2D-RNT block by testing its performance on seismic denoising tasks. Here, we compare DnCNN with several well-designed seismic denoising methods, including TDTV [39], MFFCNN [6], ASPP [63], BM3D [38], and MPRNet [64], on the synthetic seismic dataset FSSynth with Gaussian noise. Specifically, we use FSSynth with different noise strengths to train these CNN-based denoising models. Then, we feed the patches to denoising models with different Gaussian noise strengths (i.e., the standard deviation of normal distribution, $\sigma$). Our D2D-RNT model uses the same network structure as DnCNN, but the training set in D2D-RNT is generated from I2I-NT. To distinguish them, DnCNN stands for the network model trained with pairwise Gaussian clean patches, while NPT stands for our noise pattern transfer model trained by the I2I-NT and D2D-RNT processes. The peak signal-to-noise ratio (PSNR) results and sample patches are shown in Table I and Fig. 5. These results show that the DnCNN model performs very similar to the MFFCNN model and is the best performing model on most Gaussian denoising tasks.

Our NPT model is then applied to field seismic data as follows. In this experiment, we have chosen Beatrice to be denoised because it is the noisiest dataset we have available. We train the model NPT (Bea. to Thebe) with dataset S: Beatrice, dataset T: Thebe, and NPT (Bea. to FSClean) with dataset S: Beatrice dataset T: FaultSegClean.
We compare the denoising result with the state-of-the-art models in Fig. 6. Because the clean seismic patches were unavailable, we used denoised patches and noise differences to analyze the results. Compared with other methods, our model provides fairly good denoising results. BM3D, MFFCNN, ASPP, MPRNet, and DnCNN treat some noise as signals and retain them in the output. TDTV compensates for signal loss due to noise and heavily enhances the geological signal, which contains too much nonnoise data in the noise images. Our model extracts more accurate noise patterns because it learns from the field datasets, whereas other baseline models learn noise patterns from the Gaussian noise. Since synthetic data cannot simulate all the noise patterns of field data, allowing the denoising model to learn noise from field seismic data enhances the noise learning capability of the model.

### E. Scalability

In this section, we test the scalability of our model by counting the parameters and processing time. We use the torchvision module to monitor the DL models’ CPU usage and running time by feeding them patches of different sizes.

We have derived the following observations from the results in Table II. First, the DL approaches show better performance on running time because they benefit from graphics processing unit (GPU) computation for fast implementation. Second, the processing time of NPT is close to the baseline models, although the parameter number is 20× more than ASPP and 2× more than MFFCNN. Its network structure does not
Fig. 6. Denoising/noise transfer results on field dataset. TDTV is a self-supervised method, while MFFCNN, ASPP, DnCNN, and MPRNet are trained using FaultSeg and FaultSegClean datasets. The parameters are set to the default suggested by their authors.

contain concatenate operations, so it can fully utilize the advantage of GPU parallel processing ability. Third, although MPRNet is the state-of-the-art denoising model for natural image denoising, the running time and denoising performance are unsatisfactory. Therefore, NPT is very competitive for seismic interpretation applications considering denoising performance and scalability.

F. Fault Interpretation

To further examine the practicality of our proposed method in terms of denoising and generalization, in this section, we consider the task of fault interpretation. Robust seismic fault interpretation allows geological faults to be identified and characterized effectively, enabling appropriate evaluation of subsurface risks related to fault sealing capacity in hydrocarbon or CO2 reservoirs. With the rapid development of the DL algorithms, the exponential growth of geological data, and the associated uncertainties of manual interpretation, the research community has started using the DL algorithms to automatically identify faults from seismic images. The realization, however, that the DL algorithms struggle with the vast diversity of seismic data (i.e., sampling rate, noise, artifacts)
TABLE II
NUMBER OF PARAMETERS AND AVERAGE TIME COSTS

| Model Name | Parameters | Running time (ns) |
|------------|------------|-------------------|
|            |            | 256x256 | 512x512 | 1024x1024 |
| BM3D       | —          | 2960.6   | 9534.88 | 36391.76  |
| MPRNet     | 15,741,025 | 144,6419 | 471.5992 | 1850.2380 |
| MPRCNN     | 220,769    | 13,6271  | 45.0712  | 171.5096  |
| ASPN       | 28,097     | 7.4522   | 16.4127  | 62.1624   |
| NPT        | 557,057    | 11,5090  | 40.1513  | 165.3424  |

Fig. 7. Fault predictions of U-NetT on three random ThebeFault test patches before and after our NPT denoising. Columns from left to right: fault interpretation by experts, original seismic patches, fault prediction of original seismic images, denoised seismic image, and fault prediction of the denoised seismic image.

provides the principal motivation behind our examination of the proposed method [65].

We designed two separate experiments testing our proposed method’s denoising and generalization abilities. For the denoising experiment, we hypothesize that our NPT approach can improve the DL-based fault recognition method by denoising the input seismic data. For the generalization ability experiment, we focus on the constraint that fault annotations are difficult to access, and the DL model trained using synthetic seismic datasets shows unsatisfactory performance on field seismic data. We hypothesize that our approach can improve the generalization of the DL-based models on field seismic data even if there are no fault annotations.

U-Net, a classic deep neural network often used for fault recognition, is chosen for the two fault recognition experiments [58], [66]. The network hyperparameters are set consistent for a fair comparison, and more details can be seen in the supplementary materials. In terms of evaluation, 100 seismic patches and 100 corresponding fault annotations from the ThebeFault test set were randomly selected and cropped.

To provide a numerical evaluation of the fault recognition ability, we introduce the average precision (AP) metrics provided by Scikit-learn package [67]. AP is a classic metric in CV that provides a comprehensive evaluation of an algorithm, regardless of the choice of threshold. It is designed explicitly for extremely unbalanced datasets and focuses specifically on minority categories, which in our case are fault pixels. This metric calculates the area under the precision–recall curve by (9)–(11). True positive (TP) stands for pixels that U-Net correctly classified as fault pixels. False positives (FPs) are pixels U-Net falsely classified as fault pixels, while false negatives (FNs) are pixels that U-Net falsely classified as nonfault pixels. The mean AP of the 100 ThebeFault test patches was set as the evaluation metric

\[
AP = \frac{\sum_{n} (\text{Recall}_{n} - \text{Recall}_{n-1}) \cdot \text{Precision}_{n}}{n} \tag{9}
\]

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}
\]

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{11}
\]

In Table III, we summarize the numerical results of the two experiments described above. For the denoising experiment, we test whether a U-Net trained on denoised ThebeFault (U-NetT_denoised) performs better than U-Net trained on the original ThebeFault dataset (U-NetT_original). The denoised ThebeFault was obtained by transferring ThebeFault toward the noise-free FaultSegClean dataset. For the generalization experiment, we investigate whether U-Net trained using FaultSeg (U-NetF) performs better on the transferred ThebeFault.
Fig. 10. Fault probability map of a large ThebeFault seismic section. Image from left to right: expert annotation on the original seismic data, U-NetT prediction on the original seismic data, U-NetT prediction on the denoised seismic data, and U-NetF prediction on the original seismic data, U-NetF prediction on the transferred seismic data. See the electronic version for better visualization.

### Table III

| Model Name         | Mean  | Variance | \( P(T<0) \) two-tail |
|--------------------|-------|----------|------------------------|
| **Experiment 1: denoising ability** |
| U-NetT_original    | 0.7171| 0.0516   |                        |
| U-NetT_denoised    | 0.7300| 0.0512   | 0.0334                 |
| **Experiment 2: generalisation ability** |
| U-NetF_original    | 0.3222| 0.0192   |                        |
| U-NetF_transferred | 0.3530| 0.0212   | 5.8957E-07             |

We provided three visual examples for the denoising and generalization experiments, respectively. In Fig. 7, the migration noise on the denoised seismic patches is attenuated, and the denoised fault predictions are much clearer. For the generalization experiment, due to the large difference between the synthetic FaultSeg dataset and the ThebeFault dataset, U-NetF performs well on the FaultSeg test set (Fig. 8), but gives noisy prediction on ThebeFault (Fig. 9 column 3). Thus, our proposed NPT method was used to make the ThebeFault dataset more like the target FaultSeg dataset while retaining its geological features. The same U-NetF provides a more precise fault interpretation on the transformed seismic images. In Fig. 10, a large seismic section is presented to better illustrate our proposed method. Similarly, cleaner fault probability maps are achieved using our proposed NPT method. It is worth noting that as U-NetF is trained using only the synthetic data FaultSeg and has no prior experience with the Thebe dataset, the model predictions are not as good as U-NetT. However, U-NetF can provide delicate annotations, which, together with U-NetT, can give experts different viewpoints.

An additional Fourier transform experiment was performed to check whether the proposed NPT method retains the necessary geological information. As shown in Fig. 11, geological information that occurs mainly at lower frequencies is safely retained. Higher amplitudes indicate enhancement of some of the weaker geological signals.

### V. Conclusion

This article presents a deep neural network method for decomposing noise patterns in seismic images, intending to achieve seismic data NPT and noise reduction. Unlike traditional methods that use the Gaussian white noise to model seismic data noise, we use a deep neural network to disentangle the noise patterns of seismic reflection data. We extend the use of NST algorithms beyond the realm of art to produce plausible pairings of clean and noisy seismic data. Extensive experimental results show that the proposed method successfully transfers the noise patterns between field seismic data and exhibits excellent denoising performance on synthetic and field datasets. Additional experiments also demonstrate the almost real-time speed achieved using GPU. In addition, our approach achieves promising improvements in the denoising and generalization capabilities of the DL-based fault interpretation model.
However, our proposed method still has some limitations. We found that differences in seismic signal patterns between dataset S and dataset T may affect the performance of the NPT model. If dataset T is a noise-free synthetic dataset, NPT will over-correct unstable or blurry geological signals in dataset S, which leads to hints of geological signals in the extracted noise difference patches. Nevertheless, the problem is negligible when both datasets S and T are field seismic data since they share similar signal patterns. In the future, we would like to use a signal feature alignment method to assess the similarity of signal patterns, making our method insensitive to the effect of geological signal differences.

ACKNOWLEDGMENT

This publication has emanated from research supported in part by a grant from Science Foundation Ireland under Grant number [12/RC/2289_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] R. E. Brackenridge, F. J. Hernández-Molina, D. A. V. Stow, and E. Llave, “A pliocene mixed contourite–turbidite system offshore the Algarve margin, Gulf of Cadiz: Seismic response, margin evolution and reservoir implications,” Mar. Petroleum Geol., vol. 46, pp. 36–50, Sep. 2013.

[2] Y. Ma, X. Ji, N. M. BenHassan, and Y. Luo, “A deep learning method for automatic fault detection,” in Proc. SEG Int. Exp. Annu. Meeting, Houston, TX, USA: Society of Exploration Geophysicists, 2018, pp. 1941–1945.

[3] Z. Wang, H. Di, M. A. Shafiq, Y. Alaudah, and G. AlRegib, “Successful leveraging of image processing and machine learning in seismic structural interpretation: A review,” Lead. Edge, vol. 37, no. 6, pp. 451–461, Jun. 2018.

[4] S. Guillou, F. Joncour, P. Goutorbe, and L. Castanié, “Reducing training dataset bias for automatic fault detection,” in Proc. SEG Int. Exp. Annu. Meeting, Aug. 2019, pp. 2423–2427.

[5] J. Li, X. Wu, and Z. Hu, “Deep learning for simultaneous seismic image super-resolution and denoising,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–11, Nov. 2021.

[6] J. Feng, X. Li, X. Liu, and C. Chen, “Multigranularity feature fusion convolutional neural network for seismic data denoising,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–11, 2021.

[7] H. Meyer and E. Pebesma, “Predicting into unknown space? Estimating the area of applicability of spatial prediction models,” Methods Ecol. Evol., vol. 12, no. 9, pp. 1620–1633, Sep. 2021.

[8] X. Wu, L. Liang, Y. Shi, and S. Fomel, “FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation,” Geophysics, vol. 84, no. 3, pp. IM35–IM45, May 2019.

[9] Y. Ao, W. Lu, P. Xu, and B. Jiang, “Seismic dip estimation with a domain knowledge constrained transfer learning approach,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2022.

[10] R. Zhou, X. Yao, and G. Hu, “Learning from unlabelled real seismic data: Fault detection based on transfer learning,” Geophys. Prospecting, vol. 69, no. 6, pp. 1218–1234, 2021.

[11] F. Zhuang et al., “A comprehensive survey on transfer learning,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 5, pp. 3202–3212, May 2019.

[12] D. Chevaretse, D. Szwarzman, R. M. D. Silva, and E. V. Brazil, “Transfer learning applied to seismic images classification,” in Proc. AAPG Annu. Expo., 2018, pp. 1–14.

[13] Z. Wu, D. Meng, L. Wu, N. Liu, and Y. Wang, “Seismic impedance inversion using fully convolutional residual network and transfer learning,” IEEE Geosci. Remote Sens. Lett., vol. 17, no. 12, pp. 2140–2144, Dec. 2020.

[14] X. Liao, J. Cao, J. Hu, J. You, X. Jiang, and Z. Liu, “First arrival time identification using transfer learning with continuous waveform transform feature images,” IEEE Trans. Geosci. Remote Sens., vol. 17, no. 11, pp. 2002–2006, Nov. 2019.

[15] M. J. Park and M. D. Sacchi, “Automatic velocity analysis using convolutional neural network and transfer learning,” Geophysics, vol. 85, no. 1, pp. V33–V43, Jan. 2020.

[16] J. S. Dromsch and M. Lüttge, “Deep-learning seismic facies on state-of-the-art CNN architectures,” in Proc. SEG Int. Exp. Annu. Meeting, Houston, TX, USA: Society of Exploration Geophysicists, 2018, pp. 2036–2040.

[17] J. You, Y. Xue, J. Cao, and C. Li, “Attenuation of seismic swell noise using convolutional neural networks in frequency domain and transfer learning,” Interpretation, vol. 8, no. 4, pp. T941–T952, Nov. 2020.

[18] L. Wang, D. Meng, and B. Wu, “Seismic inversion via closed-loop fully convolutional residual network and transfer learning,” Geophysics, vol. 86, no. 5, pp. R671–R683, Sep. 2021.

[19] S. Kim, S. Jee Seol, J. Byun, and S. Oh, “Extraction of diffractions from seismic data using convolutional U-net and transfer learning,” Geophysics, vol. 87, no. 2, pp. V117–V120, Mar. 2022.

[20] R. Zhou, X. Yao, Y. Wang, G. Hu, and F. Yu, “Seismic fault detection with progressive transfer learning,” Acta Geophysica, vol. 69, no. 6, pp. 2187–2203, Dec. 2021.

[21] I. Goodfellow et al., “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., vol. 27, 2014, pp. 139–144.

[22] G. Ganin et al., “Domain-adversarial training of neural networks,” J. Mach. Learn. Res., vol. 17, no. 1, pp. 2030–2096, 2016.

[23] S. Misra and H. Li, “Deep neural network architectures to approximate the fluid-filled pore size distributions of subsurface geological formations,” Machine Learning for Subsurface Characterization, vol. 183, USA: Elsevier, 2019.

[24] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, “Neural style transfer: A review,” IEEE Trans. Vis. Comput. Graphics, vol. 26, no. 11, pp. 3365–3385, Nov. 2019.

[25] M. Elad and P. Milanfar, “Style transfer via texture synthesis,” IEEE Trans. Image Process., vol. 26, no. 5, pp. 2388–2391, May 2017.

[26] A. Mahendran and A. Vedaldi, “Understanding deep image representations by inverting them,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 5188–5196.

[27] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” arXiv:1508.06976, 2015.

[28] M. Kumar, B. Poormina, H. Nagendraswamy, and C. Manjunath, “A comprehensive survey on non-photorealistic rendering and benchmark developments for image abstraction and stylization,” Iran J. Comput. Sci., vol. 2, no. 3, pp. 131–165, 2019.

[29] Y. Wang, Y. Wei, X. Qian, L. Zhu, and Y. Yang, “Sketch-guided scenery image outpainting,” IEEE Trans. Image Process., vol. 30, pp. 2643–2655, 2021.

[30] Y. Li, N. Wang, J. Liu, and X. Hou, “Demystifying neural style transfer,” in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 1–7.

[31] M. Lu, H. Zhao, A. Yao, F. Xu, Y. Chen, and L. Zhang, “Decoder network over lightweight reconstructed feature for fast semantic style transfer,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2469–2477.

[32] D. Bai et al., “Seeing what a GAN cannot generate,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 4502–4511.

[33] A. Buades, B. Coll, and J.-M. Morel, “Self-similarity-based image denoising,” Commun. ACM, vol. 54, no. 5, pp. 109–117, May 2011.

[34] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[35] J. Feng, X. Li, X. Liu, C. Chen, and H. Chen, “Seismic data denoising based on tensor decomposition with total variation,” IEEE Geosci. Remote Sens. Lett., vol. 18, no. 7, pp. 1303–1307, Jul. 2021.

[36] J. Lehtinen et al., “Noise2Noise: Learning image restoration without clean data,” in Proc. Int. Conf. Mach. Learn., 2018, pp. 2965–2974.
[41] W. Fang, L. Fu, and H. Li, “Unsupervised CNN based on self-similarity for seismic data denoising,” IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, 2021.

[42] X. Dong, T. Zhong, and Y. Li, “A deep-learning-based denoising method for multiarray surface seismic data,” IEEE Geosci. Remote Sens. Lett., vol. 18, no. 5, pp. 925–929, May 2020.

[43] H. Di, Z. Wang, and G. AlaReghi, “Why using CNN for seismic interpretation? An investigation,” in Proc. SEG Int. Expo. Annu. Meeting, Aug. 2018, pp. 2216–2220.

[44] L.-C. Chen, Y. Zhu, G. Panandreas, F. Schroff, and H. Adam, “Encoder-decoder with Atrous separable convolution for semantic image segmentation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 801–818.

[45] W. Li and J. Wang, “Residual learning of cycle-GAN for seismic data denoising,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 7062–7075, Nov. 2020.

[46] T. Zhong and Y. Li, “Deep convolutional neural network for automatic fault recognition from 3D seismic datasets,” Comput. Geosci., vol. 153, Aug. 2021, Art. no. 104776.

[47] Y. An, J. Guo, Q. Ye, C. Childs, J. Walsh, and R. Dong, “A gigabyte interpreted seismic dataset for automatic fault recognition,” Data Brief, vol. 37, Aug. 2021, Art. no. 107219.

[48] J. J. Walsh, C. Childs, J. Imber, T. Manzocchi, J. Watterson, and P. A. R. Neill, “Strain localisation and population changes during fault system growth within the inner moray firth, northern north sea,” J. Structural Geol., vol. 25, no. 2, pp. 307–315, Feb. 2003.

[49] M. Huzaifah bin Md Shahrin and L. Wyse, “Applying visual domain style transfer and texture synthesis techniques to audio: Insights and challenges,” Neural Comput. Appl., vol. 32, no. 4, pp. 1051–1065, 2020.

[50] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking Atrous convolution for semantic image segmentation,” 2017, arXiv:1706.05587.

[51] A. Mehri, P. B. Ardakani, and A. D. Sappa, “MPRNet: Multi-path residual network for lightweight image super resolution,” in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., Jan. 2021, pp. 2704–2713.
Lu Liu received the Ph.D. degree in civil engineering from the Harbin Institute of Technology, Weihai, China, in 2013. Her research interests focus on performance-based seismic design and seismic analysis of self-centering braced frames, gusset-to-beam, gusset-to-column, and brace-to-gusset interactions and earthquake records' classification.

Dongjie Zhu received the Ph.D. degree in computer architecture from the Harbin Institute of Technology, Weihai, China, in 2015. He is an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include machine learning and distributed systems. He has published more than 30 articles in peer-reviewed journals and conferences.

Conrad Childs is a Lecturer in structural geology with the School of Earth Sciences, University College Dublin (UCD), Dublin, Ireland and a Co-Director of the Fault Analysis Group. His research interests focus on the geometry and evolution of faults and their impact on fluid flow in the subsurface. His main involvement in Irish Centre for Research in Applied Geosciences (iCRAG) is overseeing work on the structural evolution of the Irish offshore basins within the Earth resources' challenge.

John Walsh received the B.Sc. degree from the University College Dublin (UCD), Dublin, Ireland, in 1980, and the Ph.D. degree from the University of Galway, Galway, Ireland, in 1987. In 1985, he co-founded the Fault Analysis Group with the University of Liverpool, Liverpool, U.K., overseeing its relocation to UCD in 2000. The group performs basic research on all aspects of faults and fractures, and applies their research results to practical problems within a variety of geoscience-related industries. He was a Co-principal investigator (PI) and the Founding Director from 2014 to 2018 of the Irish Centre for Research in Applied Geosciences (iCRAG), a National Research Centre comprising 50 academic staff and more than 70 researchers (Post-Docs and PhDs).

Dr. Walsh awarded an Honorary Fellowship in 2015 and the William Smith Medal in 2017 from the Geological Society of London, and he has also won the Royal Irish Academy Gold Medal for Environmental Sciences and Geosciences in 2018.

Ruihai Dong (Member, IEEE) received the M.S. degree from Dublin City University (DCU), Dublin, Ireland, in 2008, and the Ph.D. degree in computer science from the University College Dublin, Dublin, in 2014. He is currently an Assistant Professor with the School of Computer Science, University College Dublin. His research interests lie broadly in machine learning and deep learning and their applications in recommender systems, finance, health, and geoscience. He has published in top peer-reviewed journals and conferences such as International World Wide Web Conference (WWW), ACM Recommender Systems Conference (RECSYS), Annual Conference on Intelligent User Interfaces (IUI), and International Joint Conferences on Artificial Intelligence (IJCAI).