Geoscientific
Model Development
Discussions

DeepISMNet: Three-Dimensional Implicit Structural Modeling with Convolutional Neural Network

Zhengfa Bi[1], Xinming Wu[1], Zhaoliang Li[2], Dekuan Chang[3], and Xueshan Yong[3]

[1]School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui, P.R.China.
[2]China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, Beijing, P.R.China.
[3]Research Institute of Petroleum Exploration & Development-NorthWest(NWGI), PetroChina, Gansu, Lanzhou, P.R.China.
Correspondence: Xinming Wu (xinmwu@ustc.edu.cn)

Abstract. Implicit structural modeling using sparse and unevenly distributed data is essential for various scientific and societal purposes ranging from natural source exploration to geological hazard forecasts. Most advanced implicit approaches formulate structural modeling as least-squares minimization or spatial interpolation problem and solve partial differential equations (PDEs) for a scalar field that optimally fits all the input data under smooth regularization assumption. However, the PDEs in these methods might be insufficient to model highly complex structures in practice and may fail to reasonably fit a global structure trend when the known data are too sparse. In addition, solving the PDEs with iterative optimization solvers could be computationally expensive in 3-D. In this study, we propose an efficient deep learning method using a convolution neural network to predict a scalar field from sparse structural data associated with multiple distinct stratigraphic layers and faults. Our deep learning architecture is beneficial for the flexible incorporation of empirical geological knowledge by training with numerous and realistic structural models that are automatically generated from a data simulation workflow. It also presents an impressive characteristic of integrating various types of structural constraints by optimally minimizing a hybrid loss function to compare predicted and reference structural models, opening new opportunities for further improving geological modeling. Moreover, the deep neural network, after training, is highly efficient to predict implicit structural models in practical applications. The capacity of our approach for modeling highly deformed geological structures is verified by using both synthetic and real-world datasets, where the produced models are geologically reasonable and structurally consistent with the inputs.

1 Introduction

A structurally reasonable geological model is essential to well understand subsurface spatial organization and quantitatively simulate geological processes for a wide variety of earth science applications (Wellmann and Caumon, 2018). Structural modeling intends to accurately represent the geometry of geological structures with a numerical model by using mathematical methods. The traditional modeling approach can be described as explicit or surface modeling (Chaodong et al., 2010). It intends to reproduce the complex geometry of structures by digitizing the interpreted surface elements and their arrangements, and the resultant model typically incorporates a series of geological interfaces derived by a triangulation algorithm. In addition to being time-consuming, the modeling process is also related to each individual geologist's interpretation and might not be replicated by others (Caumon et al., 2009; Chaodong et al., 2010).

25 Recently, more and more implicit structural modeling methods have been proposed for constructing geological models because of their automatic, efficient, and reproducible characteristics (Guillen et al., 2008; Caumon et al., 2012; Hillier et al., 2014; Laurent et al., 2014; Collon et al., 2015). The implicit method is distinguished from the explicit approach because it consists in interpreting field structural observations into a volumetric function that is defined on the entire region of interest to implicitly represent geological structures. In this function, the geological interfaces are embedded as scalar iso-surfaces, while

30 the structural discontinuities are highlighted by discontinuous value jumps of the function. Thus, the scalar function is also called the implicit model. The implicit method benefits from incorporating available geological information into the resultant model from a global respective, providing an effective alternative to reproduce the geometry of the subsurface by integrating the observed data and the empirical rules (Fossen, 2016). File observations and measurements are interpreted by the geologists and geophysicists to obtain the input data of the implicit method that typically included various types of modeling objects, such

35 as spatial points, vectors, polylines, and surfaces. The output model needs to represent geologically reasonable structures while honoring the input structural data. As it is hardly possible to observe a ground truth of subsurface, the geological structures are often sparsely and heterogeneously sampled, or regional available in a limited number of highly developed mining and oil fields. This arises the necessity of adding prior geological rules and simplifications as structural constraints to guide the modeling process.

40 The discrete smooth interpolation (DSI) is one class of implicit methods that compute structural models by discretizing the scalar function on a volumetric mesh (Mallet, 1988, 1992, 1997, 2014; Souche et al., 2014; Renaudeau et al., 2019). In DSI and its variant approaches, structural modeling is performed by solving a least-squares minimization problem with smooth constraint to compute a scalar field compatible with the inputs. This smooth constraint incorporates empirical geological knowledge into the modeling process with a fundamental assumption that the desired model should be as smooth as possible.

45 However, the mesh elements prohibit from crossing structural discontinuities because the scalar function is always continuous on the mesh elements, and the method cannot correctly estimate the gradients of the scalar function near the faults or unconformities (Shewchuk, 2002). In addition to DIS, the potential field method (PFM) is another class of implicit approaches (Lajaunie et al., 1997; Jessell, 2001; McInerney et al., 2007; Phillips et al., 2007). PFM typically formulates structural modeling as a dual cokriging interpolation (Chiles et al., 2004; Calcagno et al., 2008) or as a radial basis function interpolation (Carr et al., 2001).

50 In comparison to DIS, although the modeling results can be evaluated on a volumetric mesh for a visual purpose, PFM does not use any mesh grids in computing the scalar function. Instead, structure interpolation is fully dependent on the distribution of the observed structural data and the influence range of each structural point data is determined by the chosen interpolants. However, PFM usually yields a dense system to scale the influence of the interpolants over the entire volume of interest for acquiring a structurally valid solution, which causes the computational cost quickly increase with the input data size and soon

55 become prohibitive.

The existing approaches exhibit many promising characteristics, however, reproducing structures of highly deformed regions remains a challenging task regarding geological consistency as their modeling reliability depends on the availability and quality of the observed data. Structural interpolation fully guided by mathematical equations might not always produce a geologically valid model given sparse or unevenly distributed data in some complex geological circumstances. Corresponding structural
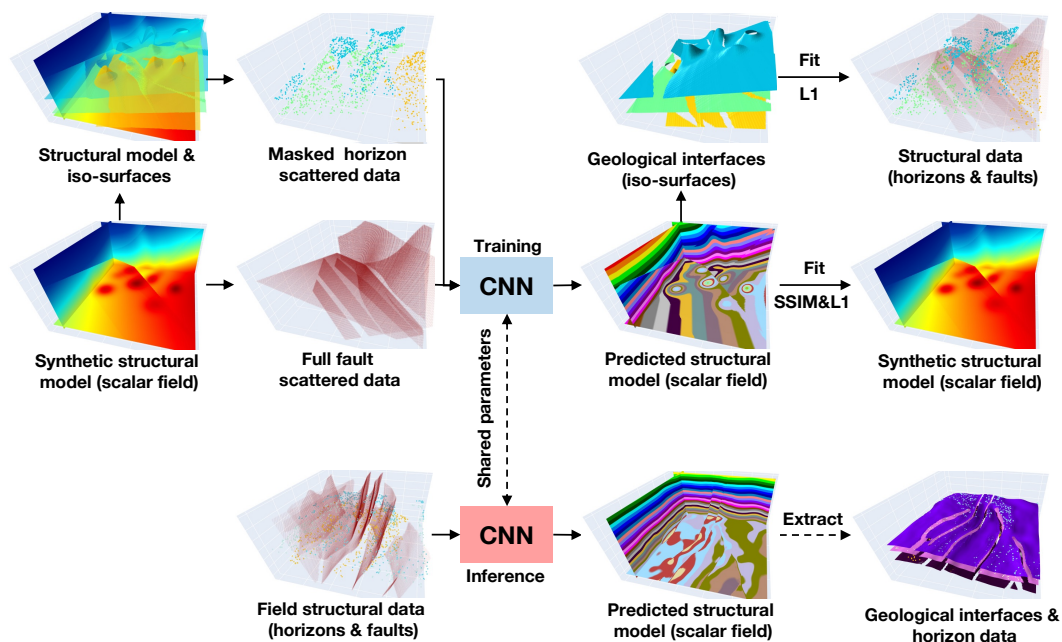
Figure 1. Our implicit modeling method produces a volumetric scalar function as an implicit representation of all the geological structures from input point data by using CNN. Trained with numerous synthetic data, the network can be applied to field structural data to efficiently predict a geologically reasonable model that well match the input structure data.

60 models often have erroneous geometrical features that are inconsistent with geological knowledge and spatial relationships with relevant structures. The problem is mainly attributed to limitations on the implicit constraints permitted in structural interpolants, in which all sorts of the observed data and empirical rules are mathematically represented as linear forms of constraints to compute the scalar function as smoothly as possible. Although this assumption contributes to a unique model, imposing such an explicit smoothness criterion might compromise the influence of local structural variations and negatively

65 impact the modeling accuracy of highly variant structures (de Kemp et al., 2016; Hillier et al., 2021). Because the modeling flexibility is limited to the models that a specific interpolant can generate, the current implicit methods usually suffer from artifacts or geometrical features physically impossible from a geological modeling point of view. Therefore, it is significant to improve implicit modeling by using all available geological information to ensure that we can obtain a structurally reasonable model.

70   In this study, we present a deep learning method using a convolutional neural network (CNN) as an alternative to conventional implicit structural modeling. Deep learning is a type of data-driven and statistical approach that estimate an implicit function that maps inputs to outputs from past experiences or example data by optimizing a quality criteria or loss function (Donmez, 2010). By contrast with the traditional approach, deep learning is beneficial for making prediction without fully defining a specific physical process and solving a sophisticated linear system of equations under prescribed mathematical con-

75 straints at cost of expensive computation. Among current learning-based methods, CNN is essential for its remarkable power
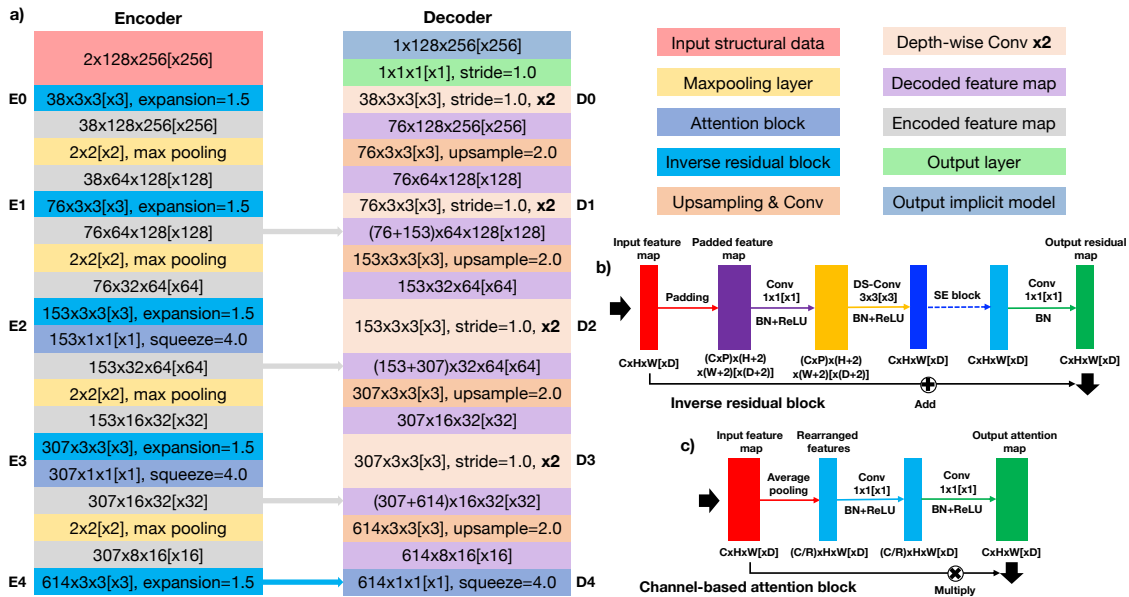
Figure 2. Our network has a U-shaped architecture that consists of encoder and decoder branches shown in (a). The encoder uses an inverse residual block (b) supplemented by a lightweight channel-based attention (c) to deal with the input structural data at each of the 5 different spatial scales. The decoder computes the hidden representations at the corresponding 5 resolution scales to form a sufficiently deep CNN.
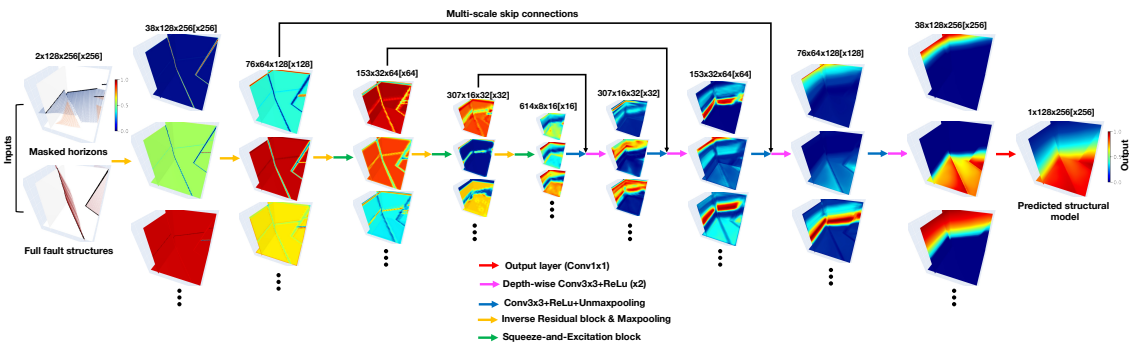
Figure 3. The hidden feature representations computed in each spatial scale of the encoder and decoder networks.

in analyzing geometrical features and capturing complexly nonlinear spatial relations given a sufficiently large training dataset. To find an optimal trade-off between prediction accuracy and efficiency, there exist many convolutional modules available for constructing our CNN's architecture, such as depth-wise separable convolution (Ioffe and Normalization, 2014; Howard et al., 2017), attention mechanism (Iandola et al., 2016; Howard et al., 2019), and residual learning structure (Sandler et al., 2018). It is not a surprise that the CNN-related applications in geosciences have been growing rapidly during the past years, including seismic interpretation (Shi et al., 2019; Wu et al., 2019; Geng et al., 2020; Bi et al., 2021), earthquake location and

detection (Wu et al., 2018; Perol et al., 2018), remote-sensing image classification (Chen et al., 2016; Maggiori et al., 2016), and so on. It is worth noting that a novel learning-based method using Graph Neural Networks (GNN) (Hillier et al., 2021) has been recently developed to integrate structural observations into a graphic mesh encoding all relevant geometrical relations

85   for producing a structural scalar field. This method presents a promising foundation for introducing interpolation constraints that current implicit mathematical methods cannot permit in modeling by comparing the GNN's prediction and the structural observations. It shows an impressive performance to deal with implicit and discrete geological data, opening new opportunities for improving the modeling capacity in many geological applications. However, the method cannot reproduce the modeling results under the same inputs as the network parameters are initialized randomly in each run of computation. By measuring

90   prediction error only on the sparse structural observations, it remains a challenging task to incorporate constraints associated with structural discontinuities into graphic structures, such as representing the spatial relation of the modeling elements across faults. Another potential limitation results from a bottle-necking problem in the current GNN's architecture (Alon and Yahav, 2020), in which further improvement of the modeling capability is restrained by network depth with a few layers. A network with a simple structure might not be sufficient to deal with relatively complex geological structures.

95   As is shown in Figure 1, we formulate implicit modeling as image inpainting with deep learning, in which a full model is interpolated from the sparse and heterogeneously sampled structural data based on the past experiences and knowledge learned from a large training dataset. This characteristic permits a flexible introduction of empirical geometrical relations and structural interpolation constraints by defining an appropriate loss function to measure the structural differences between the CNN predictions and the reference models. Our network, also called DeepISMNet, produces a structural scalar field as an

100   implicit representation of all the geological structures from various types of the input data including horizons that encode the stratigraphic sequence of the sampled interfaces and faults in the presence of the geological boundaries. In preparing for training dataset, we use an automatic data simulation workflow to generate numerous geological models with realistic faulted and folded structures that are not limited to a specific pattern by randomly choosing simulating parameters within reasonable ranges. In addition, we randomly mask the synthetic models for generating the sparsely and unevenly distributed horizon data,

105   which together with the fault data are used as inputs of the network to make structural field predictions. We also demonstrate that the normal vectors sampled on geological interfaces can be used to constrain structural orientations associated with the gradient of the model. In training the network, we use a hybrid loss function combining element-wisely fitting on the known horizons and multi-scale structural similarity over the entire model to guarantee a structurally consistent CNN prediction. Once obtaining an implicit model, the horizon surfaces can be computed by using the iso-surface extraction method, while the faults

110   can be represented by the local jumps in the structural scalar field (Figure 1). In both synthetic and field data applications, our trained CNN can efficiently construct a geologically reasonable and reproducible model, showing promising potential for further leveraging deep learning to improve many geological applications.

We organize this paper as follows. In the Methodology section, we describe the CNN architecture designed for implicit modeling and its associated loss function definition. In the Data Preparation section, we introduce the methodology used

115   to automatically generates training data and simulate the partially missing horizons. The Implementation and Application sections include both synthetic and real-world case studies to verify the performance of our network in representing complex

geological structures. The Discussion section presents the promising characteristics of the proposed CNN approach and its current limitations and possible improvements that we will focus on in future research. Finally, we summarize our work in the Conclusions section.

2 METHODOLOGY

In this section, we describe the CNN architecture and its associated loss function used in training the network to generate implicit structural models.

2.1 Network Architecture

Our developed CNN architecture uses a U-shaped framework modified from UNet and its associated variants (Ronneberger et al., 2015; Zhou et al., 2018), in which we include further improvements based on previous works to find an optimal trade-off between accuracy and efficiency in geological modeling. In many image recognition tasks where inputs and outputs share the same spatial resolution, UNet is typically regraded as a standard principle due to its excellent performance (Lin et al., 2017; Yu et al., 2018a). Its great representational power results from an encoder-decoder architecture, in which features are first downsampled in the encoder and then recombined with their upsampled counterparts through skip connections at each spatial resolution in the decoder. The localized components of the inputs are typically extracted at an early stage of the CNN, while the relatively high-level and global features are obtained when the receptive fields are increasingly large in deep convolutional layers. Thus, because the hidden representations with different spatial resolutions have much distinctive geometrical information, systematically aggregating the multi-scale features provides hierarchical constraints to make a reliable and stable structural field prediction. Furthermore, the low-level features computed from the shallow layers better follow the input structures than the deep high-level features because the structural information might be gradually missing in recursive feature compressions at the downsampled spatial resolutions. The use of skip connections helps to enhance the low-level features throughout the network and produce a model structurally consistent with the inputs.

The encoder branch in our proposed network consists of 5 successive inverse residual blocks dealing with the input structural data at 5 different spatial scales (from $\mathbf{E}0$ to $\mathbf{E}4$) related to 2, 4, 8, 16, and 32 downsampling rates, respectively (Figure 2a). When downsampling the hidden representations at each spatial scale by using the max pool layer, the encoder network simultaneously doubles its channels. As is shown in Figure 2b, we adopt a linear bottleneck and inverted residual architecture to make an efficient convolutional structure by leveraging the low-rank nature of the structural interpolation in each block. This structure is composed of a 1×1 expansion convolutional layer, a 3×3 depth-wise convolutional layer, and another 1×1 projection convolutional layer, and each convolution is followed by a Batch Normalization (BN) and a Rectified Linear Unit (ReLu). The two 1×1 convolutional layers at the ends of the depth-wise convolutional layer are designed to expand the input features to higher-dimensional feature space and project them back to the output channels, such that the block forms a compact feature embedding to improve the expressiveness of the nonlinear transformation at each channel. With a residual connection over the expansion and projection convolutional layers, the block is formulated as an implicit residual learning function to speed up

the backpropagation of gradient responses. Although the deep encoder layers aggregate rich structural features through recur-

150   sive channel expansions, not all the features are informative for making predictions because there exist many invalid features with zeros everywhere due to the spare and heterogeneously distributed characteristics of the inputs. Treating the zero features equally might negatively influence the performance of the network and degrade its representational power. To enhance the CNN's discriminative learning ability, we insert a lightweight channel-based attention module into the bottleneck structure of the residual block in the last three spatial scales of the encoder network. The attention block (Figure 2c) consists of squeeze and

155   citation modules, in which the input features are first compressed into lower-dimensional feature space in the squeeze module and then transformed to the channel-wise attention weights with the same channels as the inputs in the citation module. This module encourages the network to adaptively learn the relations across hundreds of high-level features with relatively global structural information and rescale their importance to stabilize the modeling process by emphasizing the informative features and suppressing the irrelevant ones.

160   The decoder branch includes the 5 spatial scales (from $\mathbf{D}4$ to $\mathbf{D}0$) consistent with the encoder to form a sufficiently deep network. It is responsible to integrate the hidden representations in the encoder branch into the output and compensate for the spatial resolution mismatch between the concatenated features. In each spatial scale, the upsampled decoded features are first combined with their downsampled counterparts from a skip connection of the encoder branch and then fed into the two successive convolutions to further refine these features. We use depth-wise separable convolutional layers (Howard et al., 2017)

165   as an efficient replacement of the traditional convolutional layers. The depth-wise separable convolutional layer factorizes the convolutional operation as two separate layers including a lightweight 3×3 convolutional layer for spatial feature fusion within each channel and a relatively heavy 1×1 convolutional layer across multiple channels for feature combination. By expressing convolution as a two-step process of fusing and combining, we can reduce the number of operations (MAdds) and GPU memory consumption to build a lightweight decoder network. In the final layer, we adopt a simple linear transform

170   implemented by a 1×1 convolutional layer to cross all the decoded features for producing a full geological model.

Figure 3 visualizes the hidden representations at each resolution scale of the encoder and decoder units that the inputs are passed through. Our method is designed to progressively complete structural features layer by layer through sequential non-linear convolutional filters that are conditioned on the previous convolutions. As is displayed in Figure 3, the valid convolutional responses only exist near the input structures in the starting layer of the network. To spread geological structures elsewhere,

175   every convolutional filter collects features from the previous layer outputs within an increasingly expanding receptive region by recursively downsampling the input hidden features. The structural information in the inputs can be used to constrain the modeling process over the entire volume in the bottom layer of our CNN from a global receptive region of view. This characteristic imposes a fundamental assumption that nearby features are more important than distant ones for geologically related predictions. As the modeling result of our method is estimated from the recursive aggregations of the structural information,

180   the more local features exert a greater impact than distant ones, which is consistent with the traditional spatial interpolation scheme.
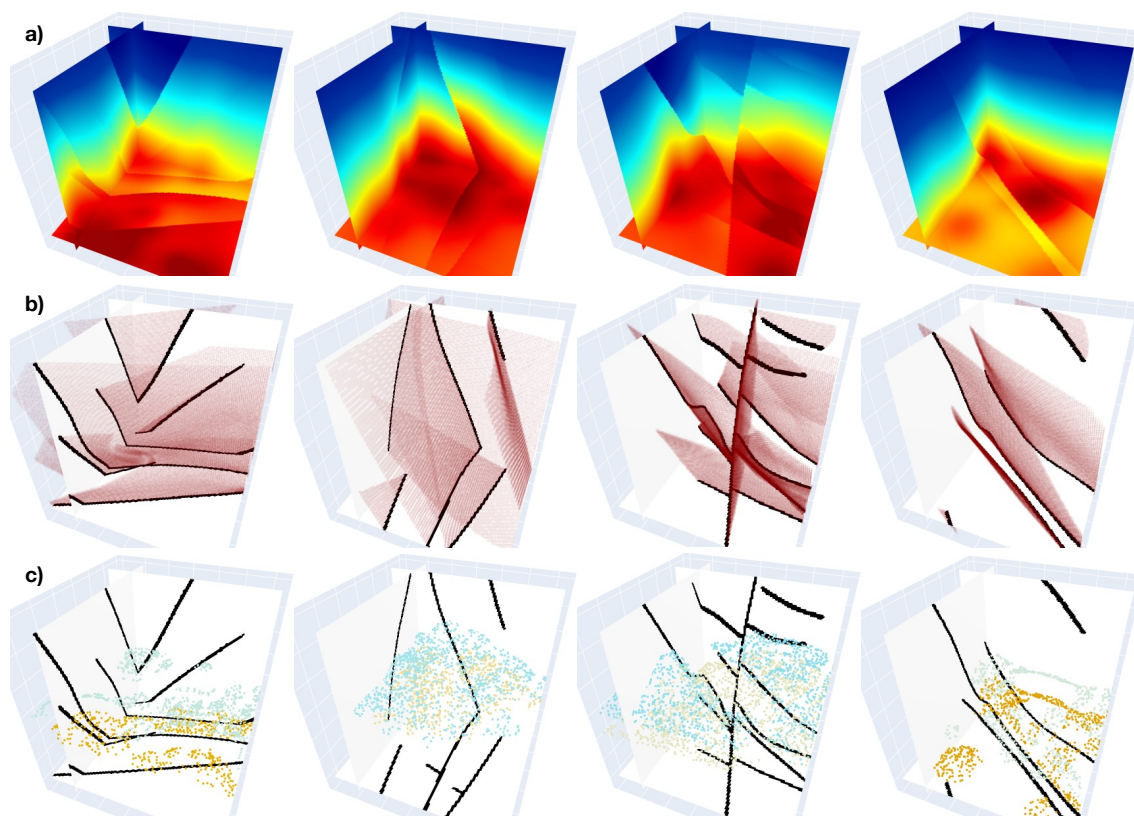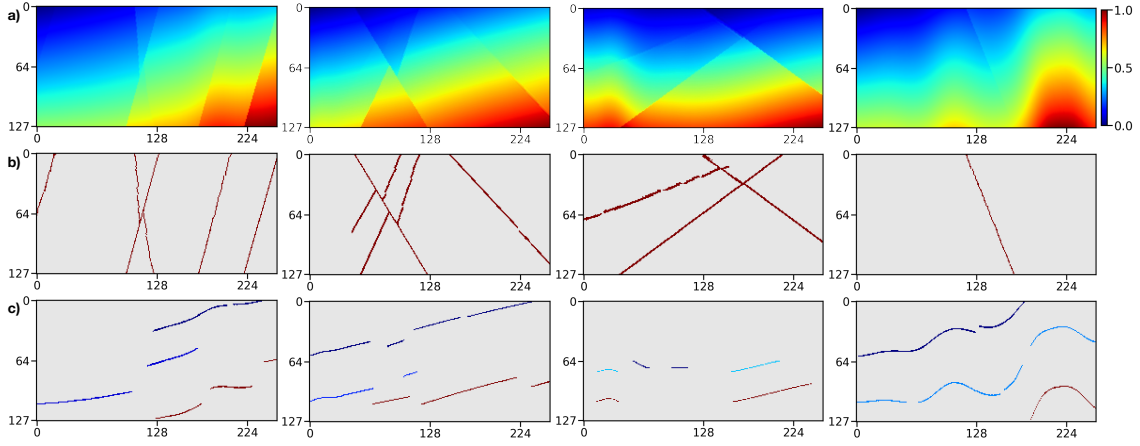
Figure 4. Four pairs of 3-D training data samples. The first row shows 3-D synthetic implicit structural models used as labels in training our 3-D network. The second and third rows, respectively, display the fault volumes and sparse horizon points extracted from the label models (first row), which are together used as inputs of the CNN.

2.1.1 Loss Function

The network provides an attractive characteristic to integrate various structural constraints by minimizing the corresponding errors between the predicted and reference models. For geologically valid CNN predictions, we combine element-wise accuracy with multi-scale structural similarity to define a hybrid loss function. We introduce notations and formal definitions used in this loss function. Let $\mathbf{m}$ be a binary mask, in which the points on the input horizons are set to 1 and the rests are set to 0. For each reference model $\mathbf{x}$, the network f_θ with trainable kernel parameters θ receives horizon data $\mathbf{h} = \mathbf{x} \odot \mathbf{m}$ and fault data $\mathbf{f}$ as inputs to make field prediction $\hat{\mathbf{y}} = f_\theta(\mathbf{h}, \mathbf{f})$. We denote the geological structures interpreted from the inputs as $\mathbf{y} = \hat{\mathbf{y}} \odot (1 - \mathbf{m}) + \mathbf{x} \odot \mathbf{m}$.

In many geologically related regression problems, Mean Square Error (MSE) and Mean Absolute Error (MAE) are commonly used to element-wisely measure the accuracy of the solutions. However, MSE typically emphasizes the elements with larger errors but is more tolerant to smaller ones, regardless of the underlying spatial pattern of the data. In comparison to MSE,

Figure 5. Four pairs of 2-D training data samples. The first row displays 2-D synthetic implicit structural models used as labels in training our 2-D CNN. The second and third rows, respectively, show the fault images and sparse horizon points extracted from the label models (first row), which are together used as inputs if the CNN. It is worth noting that the points denoted by the same color in each image of the third row correspond to the same horizon.

MAE can be more sensitive to the local structural variations and reduce the artifacts caused by excessively penalizing large errors between modeling results and targets. We adopt masked MAE as a point-wise measurement in the hybrid loss function, which is formulated as follows:

$$\mathcal{L}_{\mathrm{mae}}(\mathbf{p}) = \frac{1}{N} \sum_{p \in \mathbf{p}} |\mathbf{x}(p) - \hat{\mathbf{y}}(p)|, \tag{1}$$

where N represents the grid point number within the patch $\mathbf{p}$ cropped from the same spatial location from the two structural models being compared, respectively.

Although MAE can outperform MSE in geological modeling scenarios, the results are still not optimal. CNN trained by using MAE alone might not correctly recover geometrical features that are represented by the distribution of the neighboring points (Wang et al., 2003, 2004), and might blur high-frequency and sharp discontinuous structures. Therefore, the two models with similar MAE might might appear significantly distinct structures, which negatively impact the optimization of the CNN's parameters. To alleviate such smooth effectiveness, we use a hybrid loss function by combining MAE with Structural Similarity (SSIM). By adaptively assigning higher weights on the structural boundaries in which the geological structures present significant contrasts, SSIM can better preserve the high-frequency geometrical features than the other loss functions. SSIM measuring the CNN prediction and the reference model on a single grid point p ($p \in \mathbf{p}$) can be represented as follows,

$$\begin{aligned}
f_{\mathrm{ssim}}(p) &= \left(\frac{2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1}{\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1}\right)^{\beta} \left(\frac{2\sigma_{\mathbf{xy}} + C_2}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2}\right)^{\gamma} \\
&= l(p)^{\beta} \cdot cs(p)^{\gamma},
\end{aligned} \tag{2}$$

where $\mu_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$ represent the mean values of model $\mathbf{x}$ and $\mathbf{y}$ within $\mathbf{p}$, respectively. $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{y}}$ are the corresponding variances, while $\sigma_{\mathbf{xy}}$ denotes the covariance of the two patches being measured. The means, variances and covariance are

210 computed by using a Gaussian filter G_{σ_g} with standard deviation σ_g. Approximately, $\mu_{\mathbf{x}}$ and $\sigma_{\mathbf{x}}$ can be viewed as estimates of the stratigraphic sequences and structural variations in a local patch of model $\mathbf{x}$, and $\sigma_{\mathbf{xy}}$ measures the tendency of the two patches in model $\mathbf{x}$ and $\mathbf{y}$ to vary together, thus an indication of structural similarity. β and γ define the relative significance of the two terms l and cs, which are both set to 1 based on Wang et al. (2003). In addition, we use two small constant factors C_1 and C_2 to avoid the numerically unstable circumstance of dividing by zero.

215 The standard deviation σ_g of the Gaussian filter G_{σ_g} is a super-parameter that requires to be defined before training. However, the choice of σ_g might influence the prediction accuracy of a network trained with SSIM. The network trained by SSIM with a large standard deviation G_{σ_g} might overly emphasize the local variations and generate spurious features in the proximity of edges while blurring sharp structural boundaries for a small standard deviation G_{σ_g} (Zhao et al., 2016). Instead of fine-tuning the parameter G_{σ_g}, we use Multi-scale Structural Similarity (MS-SSIM) (Wang et al., 2003, 2004) with a dyadic pyramid of

220 M scale levels and formulate it as follows,

$$f_{\text{ms-ssim}}(p) = l_S(p)^\beta \cdot \prod_{j=1}^{S} cs_j(p)^{\gamma_j}, \tag{3}$$

in which γ_j are parameters to define the relative importance of every scale level in the variance-related scheme cs_j. MS-SSIM computes a pyramid of patches $\mathbf{p}$ with S spatial scales defined by various σ_g of the used Gaussian filter G_{σ_g}. To save computational cost, we define 5 different values of $\sigma_g = \{0.5, 1, 2, 4, 8\}$, and set each to half of the previous one by recursively

225 downsampling full-resolution patch using 2×2 average pool layer ($S = 5$). According to the recommendation from Wang et al. (2003), we set $\gamma_j = \{0.05, 0.29, 0.3, 0.24, 0.12\}$ for each scale level and make sure the sum of them is equal to 1. Thus, the loss function for MS-SSIM used in our CNN training is formulated as follow:

$$\mathcal{L}_{\text{ms-ssim}}(\mathbf{p}) = \frac{1}{N} \sum_{p \in \mathbf{p}} (1 - f_{\text{ms-ssim}}(p)). \tag{4}$$

SSIM can highlight structural variations focusing on a neighborhood of grid point p as large as the given Gaussian filter G_{σ_g},

230 but its derivatives cannot be correctly estimated near the boundary regions of patch $\mathbf{p}$ in backpropagation and thus introduce artifacts in predictions. This can be alleviated by supplying an element-wise criterion such as MAE to the MS-SSIM as it is only related to the values on a single point in the processed and reference patches. Also, MS-SSIM is not sensitive to uniform biases, which might cause unexpected changes in stratigraphic sequences or shifts of geological interfaces in modeling results. In comparison, although MAE can better preserve stratigraphic sequences by minimizing error at each point equally within the

235 patch, it might not produce quite the same high-frequency contrast as MS-SSIM regardless of local structures. To capture the best characteristics of both loss functions, we thus propose to combine them as,
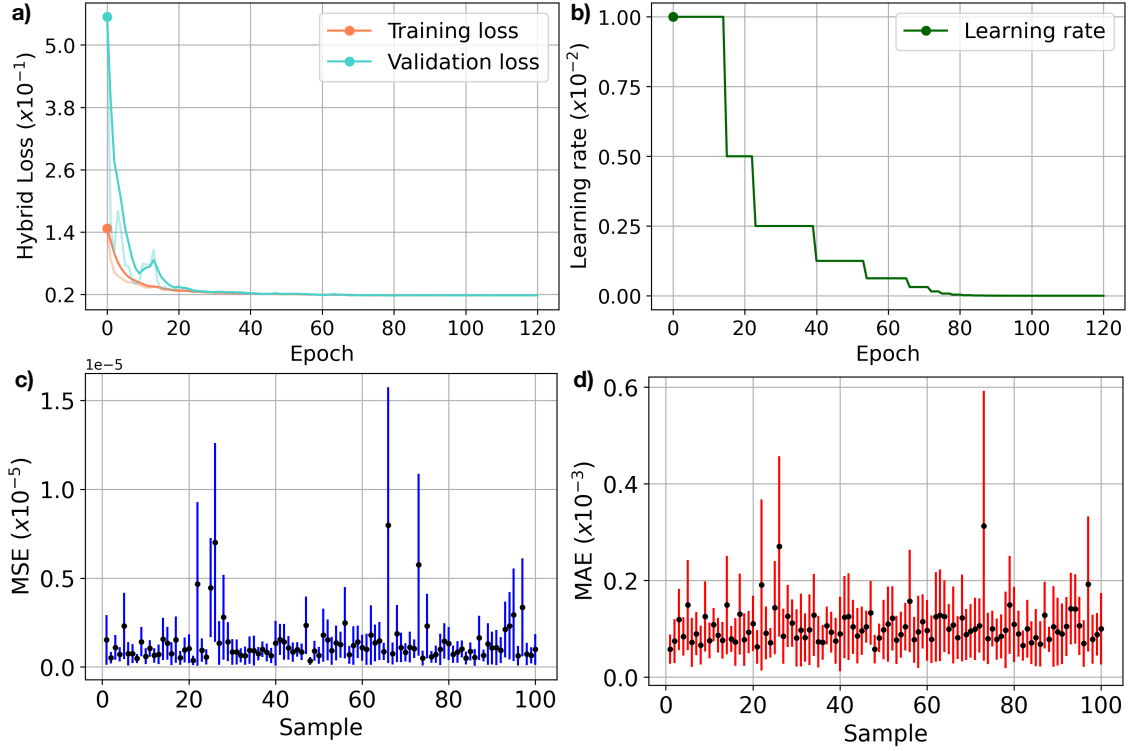
Figure 6. Training (cyan) and validation (orange) curves of using our developed hybrid loss (a), and the adaptive adjustment of the learning rate during the training (b). We run 20 times of the trained CNN to compute the MSE (c) and MAE (d) for the 100 models randomly chosen from the validation dataset, in which the input structural data are regenerated in each computation. The black dot represents the average error while the blue and red lines indicate the error ranges of the MSE and MAE, respectively.

$$\mathcal{L}_{\text{sum}} = \frac{1}{M} \sum_{m \in \mathbf{M}} \left(\lambda \mathcal{L}_{\text{mae}}(\mathbf{p}_m) + \mathcal{L}_{\text{ms-ssim}}(\mathbf{p}_m) \right), \tag{5}$$

where λ is a balancing factor used to adjust the relative significance of different loss schemes, and M denotes the number of all the cropped patches. In training the CNN, we point-wisely crop patches from the reference and processed models being measured and compute the loss function within each patch according to Equation 5, in which we empirically set the size of each patch to 7×7 and the λ to 1.25. These parameters are kept fixed throughout the study to avoid the need for tuning. The total loss $\mathcal{L}_{sum}$ is estimated by averaging all the computed losses in the patches.

11
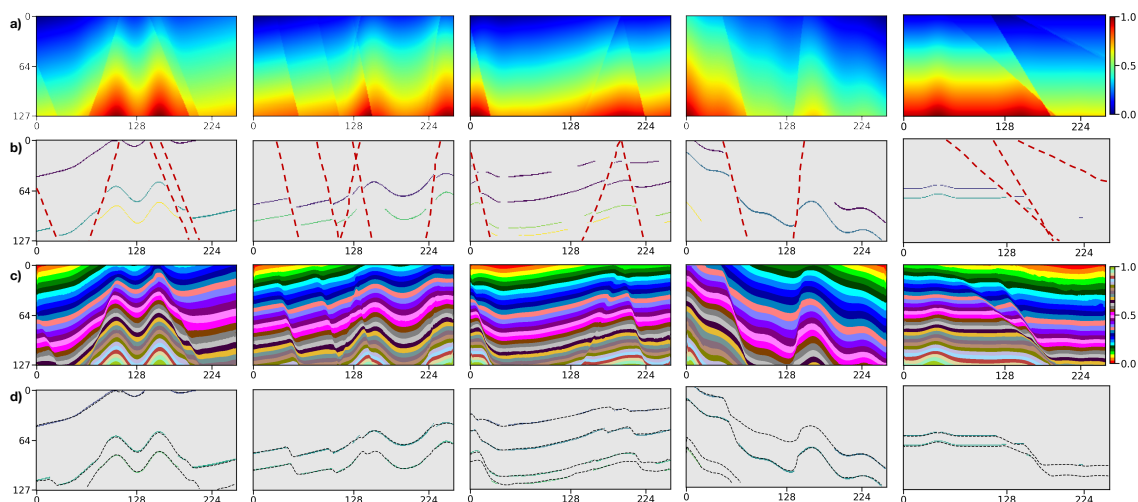
Figure 7. We apply the trained CNN to the 5 geological models (a) not included in the training dataset. We randomly generate horizon and fault structural data (b) from the models as the inputs of our network. By visual comparison, the modeling results (c) are nearly identical to the original models (a), which can be supported by the great consistency between the scalar field iso-lines (dashed curves) and the input horizons (solid curves) in (d).

3 DATA PREPARTION

Our CNN architecture is beneficial for the flexible incorporation of empirical geological knowledge in a supervised learning
245   framework with numerous structural models that are all automatically generated from an automatic data simulation workflow. We randomly delete some segments from the models to obtain the partially missing horizons similar to the modeling objects collected from field observations. In training our network, the incomplete horizons, together with the faults, are used as inputs to predict a structural scalar field under the supervision of the full model.

3.1 Automatic Data Generator

250   A supervised learning method typically requires enough example data with various feature patterns to achieve its reliable generalization in real-world applications. In our implicit modeling problem, the training dataset should incorporate structurally various models as much as possible to guarantee that the CNN can learn representative geological knowledge to interpolate a model from relatively sparse inputs. However, as ground truth of the subsurface is inaccessible, it is hardly possible to acquire sufficient example data by fully labeling all the structures in a field geological survey. To build a large training dataset, we
255   adopt a workflow (Wu et al., 2020) that automatically generates synthetic models with realistic geological structures. In this workflow, we first create a flat layered model with horizontally constant and vertically monotonically increasing values as an initial model, and then sequentially apply folding, dipping, and faulting deformations to further complicate the structures in this model. By randomly choosing deformation parameters within reasonable ranges as recommended by Wu et al. (2020), we

can generate numerous models that are not limited to a specific structural and stratigraphic pattern to enrich the training dataset.
Such a synthetic model can be viewed as a structural scalar function because its iso-surfaces track stratigraphic interfaces while
the structural discontinuities indicate faults. While adding structural discontinuities, we also record the positions of faults to
obtain fault structural data with ones on faults and zeros elsewhere. By using this workflow, we totally obtain 600 3-D structural
models shown in Figure 4a, and each model contains $256 \times 256 \times 128$ grid points. These models are vertically flipped up-down
and horizontally left-right to build an augmented dataset consisting of a total of $2,400$ pairs of models. At the same time as
vertically flipping models, we reverse the sign of their values to ensure that they vertically increase except across a reverse
fault. As is shown in Figure 5a, we extract 4 evenly spaced slices along crossline and inline from the 3-D structural models,
respectively, to further build a training dataset for the 2-D network. We use 90% of the datasets for training and the rest for
validating the trained network.

3.2 Masked Structural Data

The input structural data of our network consists of many horizon and fault points that are scattered into a regular volumetric
mesh with valid annotations on structures and zeros elsewhere. As is displayed in Figure 4b and Figure 5b, we set the scattered
points only on faults to 1 and 0 elsewhere in the fault data. In the horizon data, we set the scattered points on the geologi-
cal interfaces to the iso-values of the model to ensure that the points on the same horizon are labeled with the same value.
To simulate the sparse and unevenly sampled horizon data, we randomly delete some segments from the model to simulate
incomplete horizons in each run of data generation. The geological interfaces are implicitly embedded within the scalar field
with the iso-values and can be obtained by iso-surface extraction methods. Jittered sampling method (Cook, 1986; Hennenfent
and Herrmann, 2008) is adopted to choose iso-values of the model and obtain initial complete horizons, from which we ran-
domly mask some segments to further generate incomplete horizons. This benefits from remedying the deficiency of a regular
sampling grid that introduce a specific pattern in the horizon data while preserving the beneficial properties of randomness.
Specifically, we first divide all the iso-values into uniform intervals and then randomly extract one value within each interval
and the corresponding horizon surface, such that the horizon distribution can be varying instead of being spaced closely.

Numerous strategies can be used to randomly mask the initial horizons to generate partially missing data. The simplest
masking approach is to construct one or more square patches that cross all available horizons in the model. Although this
pattern is commonly used for numerous image inpainting tasks (Yeh et al., 2017; Yu et al., 2018b), which might negatively
impact the CNN to be well generalized in real-world applications where the inaccessible regions are unlikely in the shape
of squares. Thus, we further randomize this process by masking the segments of every individual horizon to prevent the
network from learning a specific pattern that all the horizons are partially missing in the same square regions. Specifically,
we estimate the top and bottom positions of each horizon and divide them into segments based on the depth coordinate of
each point, and randomly mask one or more segments to generate the complete or incomplete horizons. As is displayed in
Figure 4c and Figure 5c, the partially missing horizons are highly similar to the structural elements interpreted by geologists
and geophysicists, allowing the network to learn more effective feature representations. We use this masking strategy for all
the 2-D and 3-D synthetic experiments in this study.

Network architecture		Computational cost		Quality metric							
Name	Backbone	GFLOPs	#Params[MB]	SSIM	EVS	MAE	MSE×10^{-1}	MSLE×10^{-1}	R2S	MDAE	HFA
UNet	-	32.715	34.526	0.989	0.990	0.019	0.009	0.005	0.972	0.017	1.078
AttUNet	-	33.265	34.878	0.981	0.978	0.027	0.035	0.018	0.901	0.025	1.029
NestUNet	-	76.406	39.091	0.839	0.773	0.129	0.250	0.126	0.288	0.115	3.025
DeepLabv3$^+$	Xception	10.328	54.510	0.988	0.990	0.194	0.008	0.004	0.977	0.017	1.634
DeepLabv3$^+$	DRNet54	23.293	40.672	0.989	0.991	0.019	0.008	0.004	0.978	0.018	1.228
DeepLabv3$^+$	ResNet101	11.042	59.226	0.986	0.985	0.024	0.014	0.007	0.956	0.022	1.623
DeepLabv3$^+$	MobileNetv2	4.364	7.555	0.985	0.982	0.027	0.017	0.008	0.949	0.025	1.843
RefineNet	MobileNetv2	1.015	3.250	0.973	0.963	0.031	0.035	0.019	0.887	0.028	1.223
RefineNet	MobileNetv3	**0.937**	**2.600**	0.977	0.981	0.030	0.022	0.011	0.937	0.028	1.735
DeepISMNet*	-	4.711	4.300	**0.993**	**0.996**	**0.016**	**0.004**	**0.002**	**0.988**	**0.015**	**0.331**

Table 1. A quantitative comparison between our network and the widely used powerful networks using various quality metrics. For each of the quality metrics, the best performance is highlighted in **bold**. The proposed network (DeepISMNet) is marked with an asterisk to distinguish from the others.

Loss function	Modeling quality metrics						
	MAE	MSE×10^{-1}	EVS	R2S	MDAE	SSIM	HFA
L1	0.017	0.005	0.994	0.986	0.016	0.991	0.527
L2	0.017	0.005	0.995	0.987	0.016	0.989	1.321
SmoothL1	0.017	0.005	0.995	**0.988**	0.017	0.990	0.511
SSIM	0.017	0.006	0.992	0.978	0.018	0.989	0.643
MS-SSIM	0.017	0.005	0.994	0.986	0.016	0.991	0.630
MS-SSIM&L1	**0.016**	**0.004**	**0.996**	**0.988**	**0.015**	**0.993**	**0.331**
MS-SSIM&L2	**0.016**	**0.004**	0.995	0.987	**0.015**	0.990	1.040

Table 2. A quantitative analysis of our network trained with the distinct loss functions using multiple modeling quality metrics. For each of the quality metrics, the best modeling result is highlighted in **bold**.

4 IMPLEMENTATION

We present the geological structural models derived from our CNN for both synthetic data examples and field data applications in this section to demonstrate its modeling performance.

4.1 Training and Validation

Considering the coordinate ranges of the geological dataset can be much varying from each other, we rescale all the structural data to obtain the normalized training samples. This normalization is implemented by applying a global shift to the entire model to make it range from zero and one, which would not change its geological structures. In training the network, we formulate these normalized training samples in batches and set bath size to 4 based on our computational resources. Within each epoch, the training data are all passed throughout the network to compute the hybrid loss function. We utilize Adam optimization (Kingma and Ba, 2014) with an adaptive learning step length to speed up the network optimization. The initial

learning rate is set to 0.01, which reduces gradually when the criterion performance has stopped further improving. We fold the learning rate by a factor of 0.5 once the loss stagnates within 2 training iterations. As is shown in Figure 6a, the training

305   and validation loss curves gradually converge to low levels when the optimization stops (less than 0.1 after 120 epochs). The convergence of the loss curves demonstrates that the CNN has captured general hidden representations of geological structures from very few structural data and generated a geologically valid model based on these features. The learning rate is adaptively adjusted as is displayed in Figure 6b during the training process.

The model stability is evaluated by using horizon fitting error (HFA) which measures the average nearest distance between

310   the input horizons and the corresponding geological interfaces extracted from the modeling results. In this test, we randomly choose 100 models in the validation dataset and run 20 times using the trained network to compute the MSE and MAE for each model. During the inference step, it is worth noting that the input horizons are regenerated in each computation to ensure that they are different even for the same structural model. We show the variations of MSE and MAE for each model in Figure 6c and 6d, respectively, where the MSE and MAE are represented by black dots while the error ranges are denoted by the blue

315   and red lines. We observe that most MSE and MAE are less than 0.5×10^{-5} and 0.2×10^{-3}, which are considered to be not very significant in geological modeling scenarios. This demonstrates the proposed CNN architecture is beneficial for implicit structural modeling.

4.2   Synthetic Data Examples

To verify the modeling performance, we apply this trained CNN to the 5 synthetic structural models not included in the training

320   dataset. As is shown in Figure 7a, the models are of complexly faulted layered volumes, in which the folded interfaces are reformed by multiple high-angle normal faults. From the original structural models, we generate the incomplete horizon and the fault data (Figure 7b) used as inputs of our network. By visual comparison in Figure 7c, the modeling results with similar geometrical features to the inputs maintain the localized variations of the folded interfaces in spite of no global structural information used to constrain the model. We further overlap scattered horizon points on the iso-lines of the field predictions

325   (Figure 7d), in which the great consistency between the given structures and the interpolated features again supports our observation in Figure 7c.

As is tabulated in Table 1, the CNN's modeling ability is quantitatively measured by using various quality metrics including SSIM, MSE, MAE, Explained Variance Score (EVS), Mean Squared Log Error (MSLE), Median Absolute Error (MDAE), and R2 Score (R2S) for the validation dataset. In addition, we also measure the modeling accuracy of geological interfaces

330   that correspond to the input horizons by using HFA. Table 1 also shows a quantitative comparison of the proposed method (DeepISMNet) and the other powerful networks commonly used in similar scenarios. Our method not only shows better performance on all the metrics but also has a more lightweight architecture with fewer trainable parameters and GFLOPs in comparison to other CNNs. Simplification of the network architecture is mainly associated with the use of inverse residual modules followed by depth-wise separable convolutions in each spatial scale level throughout the network, enabling our CNN to

335   be applied to a large 3-D field modeling task. To guarantee the representational ability of the simplified CNN, the channel-wise dependencies have been explicitly learned by using an attention module that can adaptively highlight more informative features
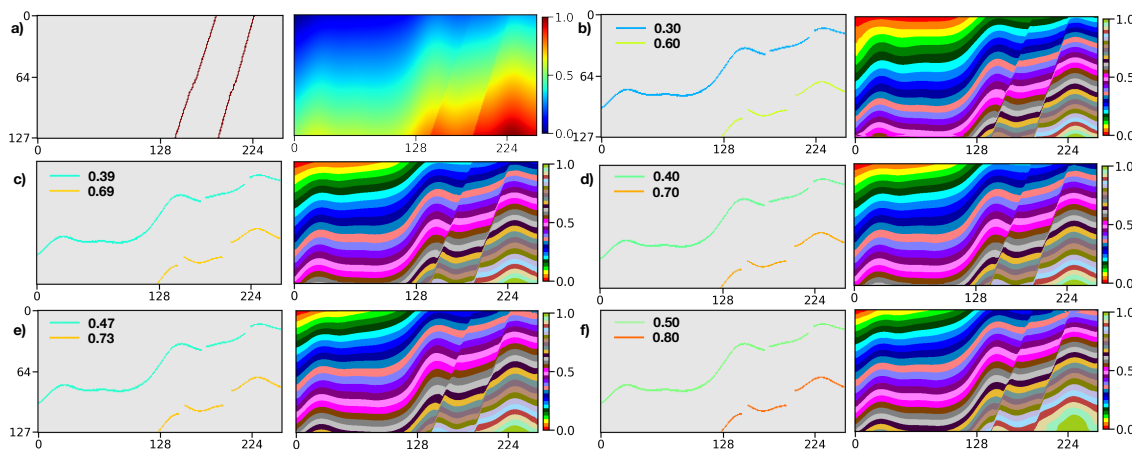
Figure 8. We use a synthetic model (a) to study how the varying horizon values impact the CNN's modeling results. The experiments are used to verify that our network can produce almost identical structural models from the same faults and the horizons with varying values, as shown from (b) to (f). This indicates that our approach is not sensitive to the annotation assigned to each horizon in the input structural data, which facilitates its applications in field surveys.

while suppressing irrelevant ones. Therefore, although the number of hidden representations is less than the conventional CNN architectures such as UNet, our network can still achieve stable structural interpolation and reliable generalization performance.

Another improvement of our approach is attributed to the use of a loss function based on element-wise accuracy and struc-
340 tural similarity in updating the CNN's parameters. To demonstrate the improved modeling performance, we implement a quantitative analysis of our CNN trained with the different loss functions using the multiple quality metrics. The average of these metrics on the validation dataset are tabulated in Table 2. The CNN trained with the hybrid loss function of MS-SSIM and MAE (denoted by MS-SSIM&L1) can outperform the others in Table 2 on all the quality metrics even including the quality metrics which we use as cost function to train the network. This loss function is attributed to a better reconstruction of fault-
345 related features in the resultant model by assigning high weights to regional structural contrasts. Also, a reliable identification of faults is useful to constrain the lateral occurrence of stratigraphic interfaces across structural discontinuities.

4.3 Horizon Annotation Experiment

The scatted points along the input horizons can be assigned to iso-values of the implicit model in the training dataset, but how to annotate each horizon before modeling remains a problem in real-world applications. We supplement a data experiment using a
350 synthetic model (Figure 8a) to study how the varying horizon values impact the CNN's modeling results. In this experiment, we set values ranging from 0.3 to 0.8 with a fixed interval to the scattered points on the two distinct horizons. The value interval is computed by averaging the differences in vertical coordinates of these horizons. As is shown from Figure 8b to 8f, the horizons with varying values, together with the faults, are used as inputs in our network. By visual comparison, the modeling results using different horizon values are nearly identical to each other, which indicates that our method is not sensitive to the iso-value
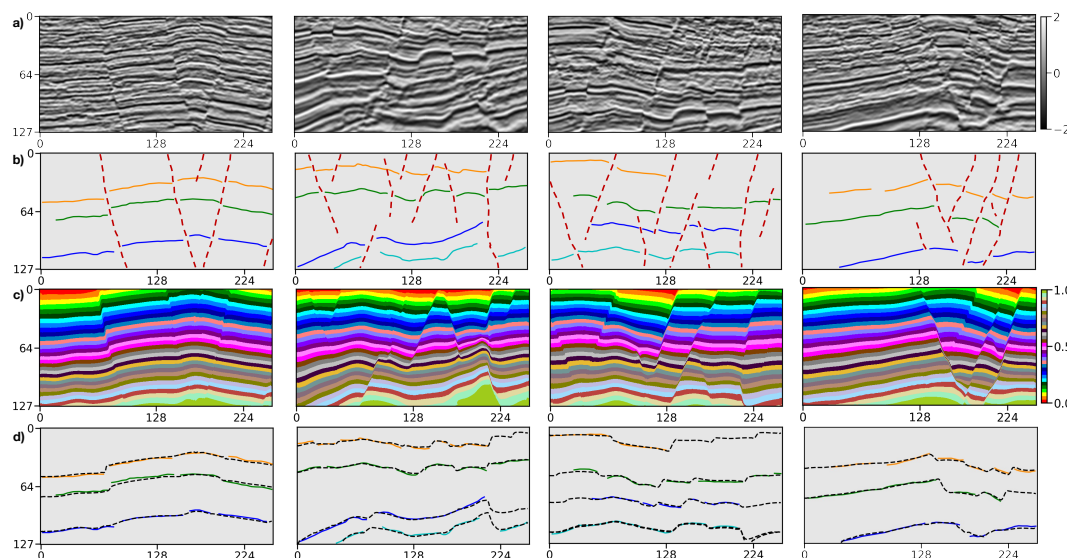
Figure 9. Application in a seismic field dataset. We display seismic images (a), input structural data (b) interpreted from the seismic images, predicted structural models (c), and horizon fitting results (d), respectively.

355   definition of the input horizons, which is what we expect. Nevertheless, we recommend assigning scattered points on each of the input horizons to their average vertical coordinate. The input horizons are required to be normalized for being consistent with the training data.

5   APPLICATION

It might not be surprising that the CNN trained with a synthetic dataset works well to produce a geologically valid and consistent
360   model by using the structural data created from the same workflow for creating the training dataset. In this section, we further present structural modeling results from our trained network for real-world data that are acquired at different geological surveys to demonstrate proof of concept. The modeling objects collected from field observations or seismic data are required to convert into the uniformly sampling grids to obtain the input structural data of our CNN.

5.1   Real World 2-D Case Studies

365   We apply the trained CNN to a field 2-D seismic dataset to interpolate structures from horizon and fault interpretations with geometrical patterns distinct from the training data. The structural interpolations are acquired in regions with closely spaced and complexly crossing faults, in which the seismic images are of low resolution due to insufficient data coverage. As is shown in Figure 9a, the ambiguous reflections are difficult to be continuously tracked from seismic images, which causes the noisy and partially missing horizon data shown by different colors in Figure 9b. The faults (Figure 9c) might not be fully annotated
370   from the seismic images because of data-incoherent noise and stratigraphic features apparent to discontinuous structures. The
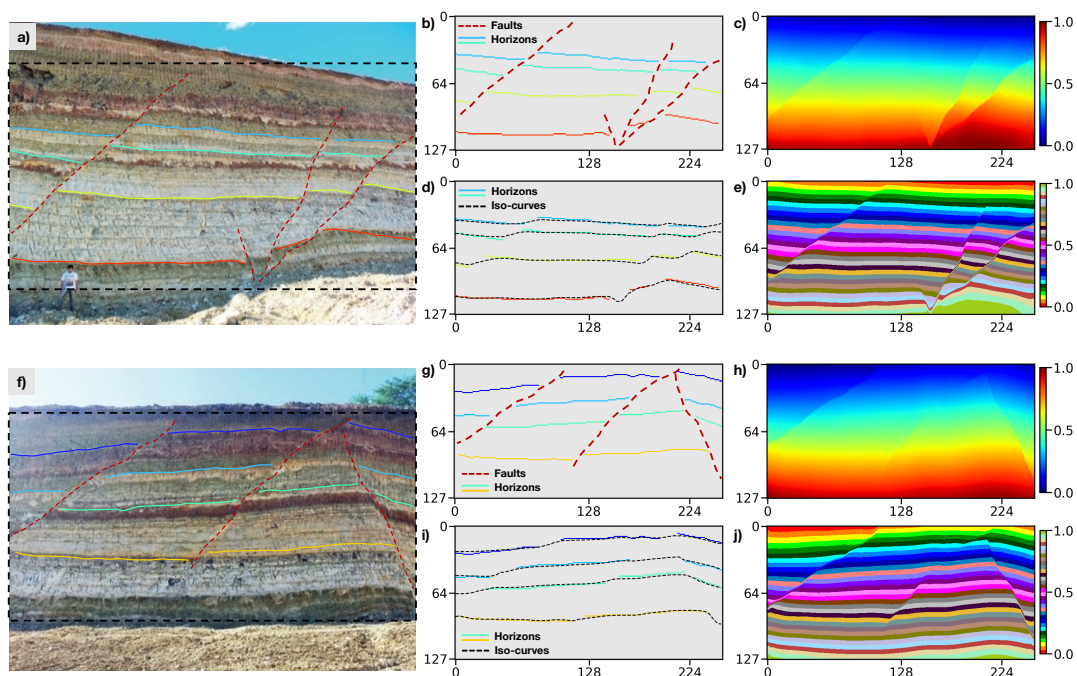
Figure 10. Application in an outcrop field dataset. We display two outcrop images (Fabin et al., 2018) in (a) and (f), input structural data in (b) and (g) manually interpreted from the outcrop images within the dashed boxes, predicted structural models using a continuous color map in (c) and (h), and a discrete color map in (e) and (j), and horizon fitting results in (d) and (i), respectively.

structural contradictions and hard-to-reconcile features in the inputs might negatively impact the modeling quality of geological structures. Therefore, it still remains a challenging task for many traditional methods to obtain a geologically reasonable model that is structurally consistent with the inputs.

As is shown in Figure 9d, our method presents geologically valid models (Figure 9d) with the structural discontinuities and
375 the stratigraphic interface variations that consistently honor the faults and horizons, respectively. From Figure 9e, we observe that the iso-lines (black lines) extracted from the modeling results can accurately follow the horizon interpretations in Figure 9b, which again supports our previous observation. In comparison to the scattered point-sets, a full structural model is more useful to well understand geological structures and qualify reservoir properties of continuity and morphology. Furthermore, the method benefiting from high computational efficiency of deep learning can even produce real-time prediction to correct
380 interpretation errors for improving the geological consistency of the model by taking all the structural interpretations into account.

The second 2-D field data experiment uses the dataset acquired from a geological survey and mineral exploration of Araripe Basin in the region of the Borborema Province of Northeastern Brazil (Fabin et al., 2018). We collect the outcrop observations shown in Figure 10a and 10f from exposures of quarries on the southwestern and northern borders of the basin. There exist a
385 series of moderate to high angle faults caused by local subsidence due to the syn-depositional dissolution of the gypsum in a
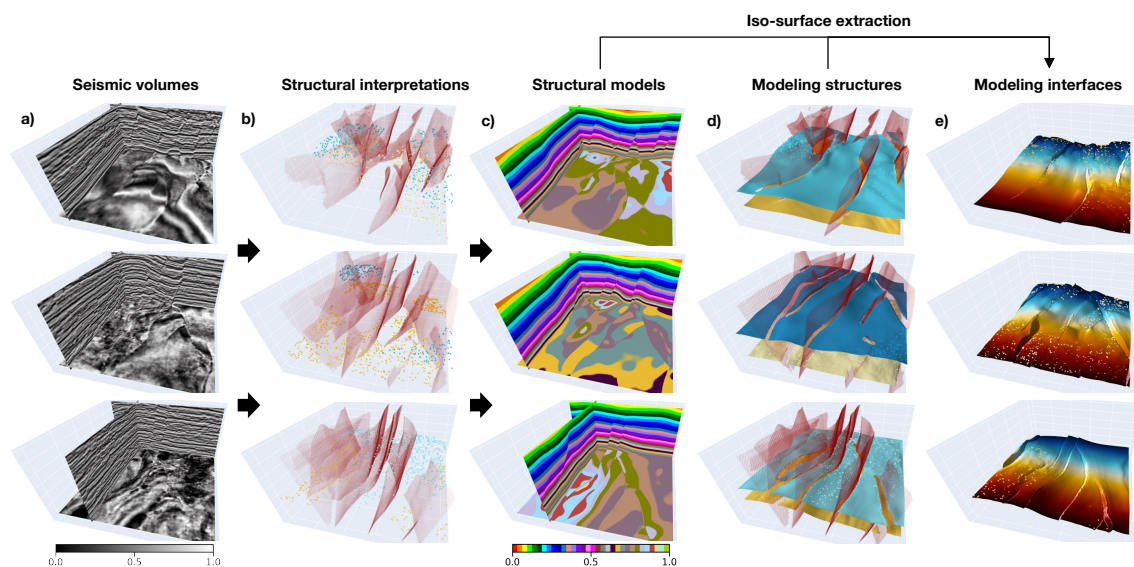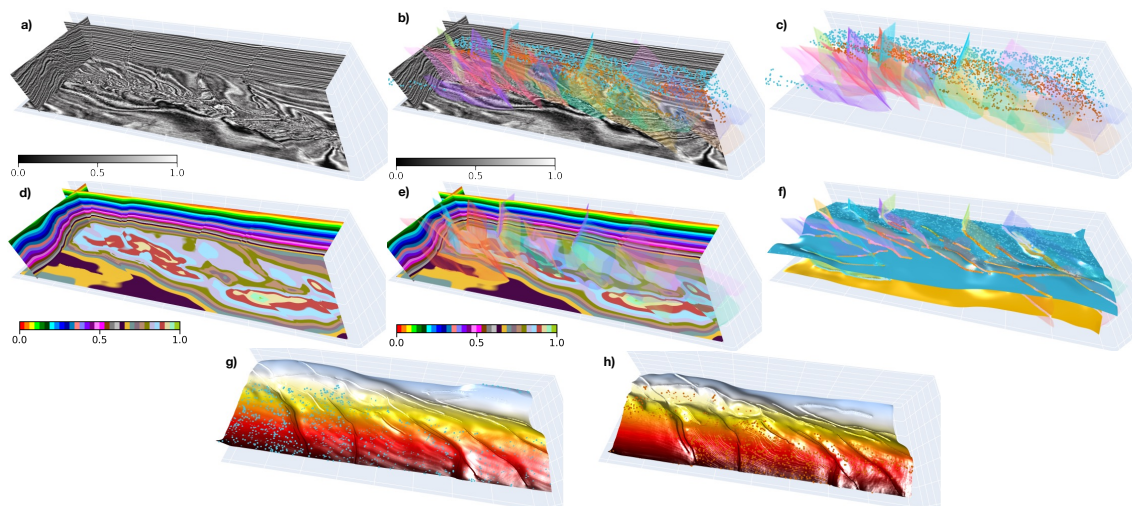
Figure 11. The first real-world data application. We display seismic volumes (a), input horizon and fault data interpreted from the seismic volumes (b), modeling results (c), geological interfaces extracted from the modeling results and overlaid with the faults (d), and one of the recovered full horizon surfaces(e), respectively.

large exposure of deposits of Romualdo Formation. These syn-depositional faults control the lateral thickness variation of the stratigraphic interfaces that the field observations sample from outcrops. The field observations are integrated into the uniform sampling grids and used as inputs of our trained CNN to predict a full structural model. As is displayed in Figure 10b and 10g, the dataset incorporates four stratigraphic interfaces with three and four faults, respectively, exhibiting a structural pattern of
390   syn-depositional deformation that is not included in our training dataset.

The modeling results presented in Figure 10c and 10h are obtained from the trained CNN, in which all the sharp edges are structurally consistent with the faults shown in Figure 10b and 10g, respectively. The same structural models can be displayed in Figure 10e and 10j using a discrete color map to clearly indicate dislocated stratigraphic layers on the opposites of the faults. The iso-lines extracted from the modeling results for the four distinct horizons are shown in Figure 10d and 10i, respectively.
395   Both figures highlight the excellent fitting characteristic of the trained CNN on the input structural data. Therefore, although the CNN is trained with the automatically simulated data, it still provides a promising performance on the real-world dataset acquired at totally different surveys with complex geological structures.

5.2   Real World 3-D Case Studies

We adopt the trained CNN in two 3-D field seismic datasets to validate its modeling capability to construct a full implicit model from spare and unevenly distributed structural points obtained by seismic interpretation. The first seismic data are sampled in
400   regions with complex geological settings and have relatively low resolution and signal-to-noise ratio because of insufficient

Figure 12. The second real-world data application. We display seismic volume (a) overlaid with structural data (b), input horizon and fault data (c) interpreted from the seismic volume, modeling result (d) overlaid with faults (e), geological interfaces extracted from the modeling result (f), and each of the recovered full horizon surfaces in (g) and (h), respectively. Noting that we rotate the horizon surfaces (g) and (h) by $180°$ for a better visual comparison.

coverage and data stacking. As is shown in Figure 11a, the seismic reflections are partially ambiguous and difficult to be continuously tracked across the entire volume of interest. The closely spaced and crossing faults further complicate structures in the presence of seismic data-incoherent noise and stratigraphic features that appear to be structural discontinuities. Therefore,
405   it is difficult for the existing approaches to model the highly deformed structures from the noisy and partially missing horizons and faults (Figure 11b). The modeling results shown in Figure 11c demonstrate that our CNN architecture is beneficial for 3-D structural modeling by predicting a geologically valid model, where the structural discontinuities and the interfaces are consistent with the given point data (Figure 11d). They even maintain the variations of the folded layer structures without global plunge information used to constrain the modeling. By visual comparison, Figure 11c shows that a group of horizon
410   points sampled at the same geological layer can be accurately located on the corresponding iso-surface of the model, which again demonstrates a great fitting characteristic of our network.

The second 3-D real-world case study is of a conformably folded and layered model with numerous faults that are curved and complexly intersected with each other. As is shown in Figure 12a, the available horizon data are manually interpreted on the two stratigraphic interfaces, while the fault data are derived by using the automatic fault detection method from the seismic
415   volume (Figure 12b and 12c). In our CNN's solution shown in Figure 12d, the geological layers represented by iso-values with the same color accurately follow the tendency of seismic structural variations even though we do not input any seismic data in our CNN. We also display the modeling result overlaid with the input fault data in Figure 12e, from which we can observe relative dislocations of geological layer on the opposites of the fault structures. In Figure 12f, we extract the geological
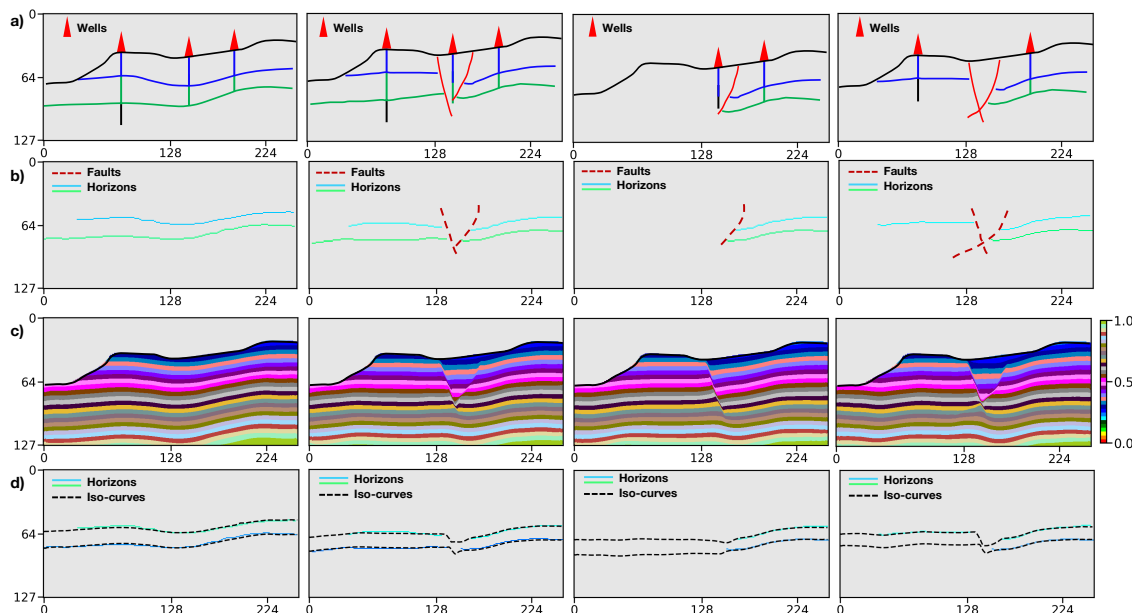
Figure 13. Geological uncertainty analysis. We display multiple sets of modeling elements interpreted from the borehole and the outcrop observations (a), input horizon and fault data (b), modeling result (c) , iso-curves extracted from the prediction (d), respectively.

interfaces that correspond to the input horizons using the iso-surface extraction approach. To show more details, Figure 12g

420   and 12h display each of the horizon surfaces, by which we rotate 180° for a better visual comparison. A great consistency between the input and recovered structures highlights the CNN's fitting characteristics on the given geological knowledge and structural constraints.

6   DISCUSSION

In this section, we discuss the modeling characteristic of our method and its abilities for structural uncertainty analysis, along

425   with the current limitations. We also demonstrate a potential improvement that we will focus on in future research to incorporate structural orientation constraints in the CNN-based structural modeling.

6.1   Structural Uncertainty Analysis

In modeling complex geological structures, the reliability of the implicit methods is heavily dependent on the quality and availability of the input structural data. However, the inputs typically spare and regional available in a geological survey

430   causes an ill-posed problem because there exist multiple plausible resultant models that equally fit the inputs. Therefore, data uncertainty analysis is necessarily critical to looking for an optimal solution, especially for the noisy and hard-to-reconcile structural observations (Viard et al., 2011; Lindsay et al., 2012). Our CNN is beneficial for the flexible interpretation of this
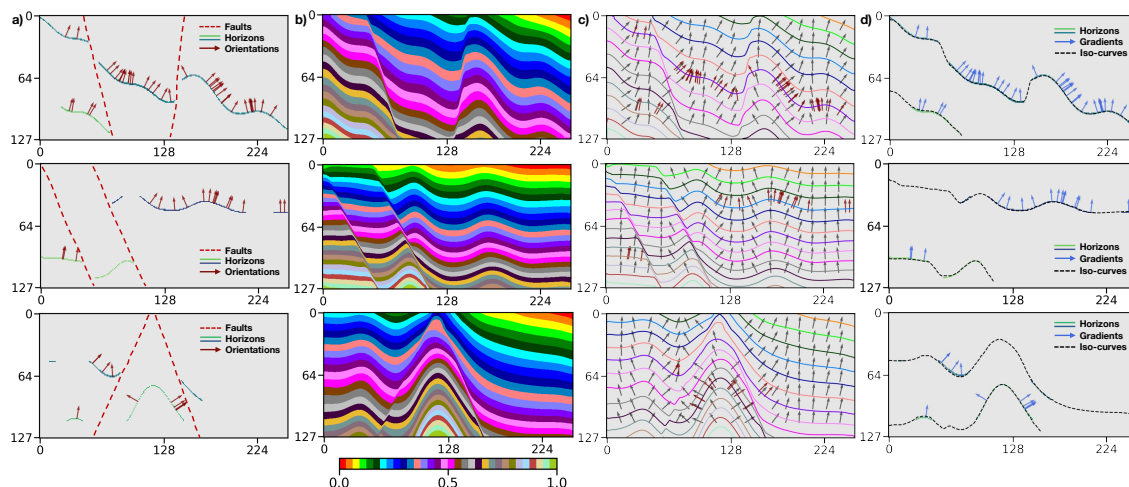
Figure 14. The faults and incomplete horizons, together with the structural orientations, are used as inputs (a) into the network to predict implicit structural models (b). The sets of regularly spaced iso-lines and local orientations are obtained from the modeling results (c) and compared with the given normal vectors (red arrows). The iso-lines and normal vectors (blue arrows) extracted on the predicted stratigraphic interfaces can well match the two distinct input horizons (d).

data uncertainty by generating a diverse of possible modeling realizations instead of one best due to its higher computational efficiency than the conventional approaches.

We use multiple combinations of modeling objects with the horizons and faults interpreted from the borehole and the outcrop observations shown in 13a to study the uncertainties associated with the positions of geological structures at depth. The simplest possible structural model consists of multiple continuous and conformal horizons in the first data example. By contrast, the modeling situations are more complex when considering additional geometrical objects of faults that dislocate the geological layers. In addition, we randomly perturb horizon positions to yield the variations in layer thickness because the stratigraphic interface transition might not be accurately observed from the vertical boreholes. The two distinct horizons and faults displayed in Figure 13b are used as inputs into the network to compute the structural models presented in Figure 13c to demonstrate proof of concept. All the four models are obtained by using a desktop PC with Intel Xeon-5120 CPU (2.20GHz) and a single NVIDIA Tesla V100 GPU. Although we take a few hours in training the CNN, the average time for generating each structural model is approximately 0.2 seconds using a 128×256 image size. Iso-curves extracted from the models (Figure 13c) for the two distinct horizons are displayed in Figure 13d, which verify an excellent fitting characteristic of our method on the given geological structures.

6.2 Structural Orientation Constraint

Our CNN architecture permits a flexible incorporation of varying types of geological information by defining an appropriate loss function to measure the modeling error for each structural constraint. In our method, the input data are not limited to

450   horizons and faults and also include the structural angular observations. The new angular constraints represent local orientations of geological layers and permit geometrical relationships in the gradient of the scalar function to be considered. The loss function for orientation constraints measures the angle differences between the gradient of the predicted model and orientation observations. We adopt a central differences method to compute structural orientations by using the second-order Taylor series approximation of the predicted model $\mathbf{z}$. The cosine similarity in every single point between the gradient vector of the scalar

455   function $\nabla \mathbf{z}$ and a prescribed orientation vector $\mathbf{n}$ can be represented as follows,

$$F_{\text{cs}}^{\text{pred}}(p) = \frac{\mathbf{n}_p \cdot \nabla \mathbf{z}(p)}{\|\mathbf{n}_p\| \|\nabla \mathbf{z}(p)\|}, \tag{6}$$

which is also used to compute the cosine similarity of the reference model gradient $F_{\text{cs}}^{\text{obs}}(p)$. The loss function used to measure the angle errors in comparison to the observed orientation constraints can be formulated as follows,

$$\mathcal{L}_{\text{normal}}(\mathbf{p}) = \frac{1}{N} \sum_{p \in \mathbf{p}} |F_{\text{cs}}^{\text{pred}}(p) - f_{\text{cs}}^{\text{obs}}(p)|. \tag{7}$$

460   Therefore, the total loss function is defined by combining varying types of geological constraints as,

$$\mathcal{L}_{\text{sum}}(\mathbf{p}) = \frac{1}{M} \sum_{i \in \mathbf{M}} (\lambda \mathcal{L}_{\text{mae}}(\mathbf{p}_i) + \mathcal{L}_{\text{ms-ssim}}(\mathbf{p}_i) + \beta \mathcal{L}_{\text{normal}}(\mathbf{p}_i)), \tag{8}$$

in which β is used to balance the relative significance of different constraints. As is displayed in Figure 13a, the faults and the incomplete horizons, together with the structural orientations sparsely distributed on the horizons, are used as inputs in the network to interpolate geological structures. The modeling results shown in Figure 13b exhibit the remarkable performance of

465   our method when using multiple types of geological data inputs. We compute regularly spaced iso-lines and local orientations from every modeling result and compare them with the input normal vectors (red arrows) in Figure 13c, respectively, which presents a great consistency between the predicted and given geological structures. The interpolated structures by using the CNN maintain the large localized geometrical variations even though there is no global geological information to constrain the modeling process. In addition, the iso-lines and normal vectors (blue arrows) extracted along the stratigraphic interfaces of the

470   predictions can well match the two distinct input horizons (Figure 13d), which again highlights the CNN's fitting characteristic on the given structural constraints.

6.3   Current Limitations and Improvements

The CNN trained by using the synthetic dataset presents excellent modeling capacities in real-world cause studies to represent complicated geological structures that are distinct from the simulated structural models. Instead of imposing any explicit math-

475   ematical constraints in the traditional implicit method, our CNN-based structural modeling is implemented by the recursive spatial convolutions with trainable kernel parameters and the loss function related to various geological constraints. The spatial

convolutions in the CNN can be viewed as the implicit interpolants used in the traditional interpolation methods, and the only difference is that the parameterization of their kernel functions can be optimized through training. As structural modeling is dependent on the analysis of the spatial relations of the observed structures to interpolate new geologically valid structures

480   elsewhere, acquiring representative example data is essential for training the CNN to achieve its reliable generalization performance. Therefore, we adopt an automatic workflow to generate numerous models with realistic faulted and folded structures and simulate partially missing horizons in building the training dataset. It is a significant reason why our network could be applied to the real-world datasets acquired in different geological surveys with distinct structural patterns.

Although working well to recover faulted and folded structures, the proposed CNN might not represent other geological

485   structures that are not considered in the training dataset, such as unconformities and igneous intrusions. The trained network also might not correctly construct low dip-angle thrust faults in predicted models because we still do not include this type of fault in the currently used training data generator. Despite the current limitations, the proposed CNN architecture still shows promising potential to compute a geological valid and structurally consistent model honoring the observed structures. Considering the used training dataset is still not sufficiently large to train a 3-D deep network, future works will focus on

490   further complicating the data simulation workflow by adding more complex and diverse geological structural patterns in the synthetic models.

7   Conclusions

A CNN-based deep learning method has been used to represent geological structures over the entire volume of interest from typically spare and hard-to-reconcile structural interpretation data. The network is composed of encoder and decoder branches

495   and supplemented with lightweight depth-wise separable convolution and channel-wise attention to find an optimal trade-off between modeling accuracy and computational efficiency. The developed CNN architecture leverages the low-rank nature of the spare and heterogeneously sampled structural data to adaptively suppress uninformative features by using a linear bottleneck and inverted residual structure in each of the encoded convolutional layers. Our approach is beneficial for the flexible incorporation of empirical geological knowledge constraints in a supervised learning framework using numerous and realistic

500   structural models that are generated from an automatic data simulation workflow. This also provides an impressive characteristic to flexibly integrate multiple types of structural constraints into the modeling by using an appropriate loss function, exhibiting a promising perspective for further improving geological modeling. We verify the effectiveness of the proposed approach by using the case studies acquired in distinct geological surveys, including synthetic examples created by the same workflow for acquiring the training dataset, the randomly created modeling objects without any ground truth of geology, and

505   the structural interpretations obtained from the seismic images. In both synthetic data and real-world data applications, we verify its modeling capacities in representing complex and noisy structures with a model geologically reasonable and structurally consistent with the inputs.

24

Code and data availability. The synthetic structural models, used for training and validating our network, are uploaded to Zenodo and are freely available through the DOI link https://doi.org/10.5281/zenodo.6480165. The source codes for the neural network developed in Pytorch
510   are provided at the Github URL https://github.com/zfbi/DeepISMNet.

Author contributions. XW initiated the idea of this study and advised the research on it. ZB conducted the research and implemented the 2-D and 3-D CNN-based structural modeling algorithms. XW conducted numerical structural simulations to provide synthetic structural models for training. ZB prepared the training datasets from the simulated structural models and carried out the experiments for both synthetic and real-world case studies. ZL, DC, and XY helped design the experiments and advised on result analysis from a geological perspective. ZB
515   and XW prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

References

520 Alon, U. and Yahav, E.: On the bottleneck of graph neural networks and its practical implications, arXiv preprint arXiv:2006.05205, 2020.

Bi, Z., Wu, X., Geng, Z., and Li, H.: Deep relative geologic time: a deep learning method for simultaneously interpreting 3-D seismic horizons and faults, Journal of Geophysical Research: Solid Earth, 126, e2021JB021 882, 2021.

Calcagno, P., Chilès, J.-P., Courrioux, G., and Guillen, A.: Geological modelling from field data and geological knowledge: Part I. Modelling method coupling 3D potential-field interpolation and geological rules, Physics of the Earth and Planetary Interiors, 171, 147–157, 2008.

525 Carr, J. C., Beatson, R. K., Cherrie, J. B., Mitchell, T. J., Fright, W. R., McCallum, B. C., and Evans, T. R.: Reconstruction and representation of 3D objects with radial basis functions, in: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 67–76, 2001.

Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., and Sausse, J.: Surface-Based 3D Modeling of Geological Structures, Mathematical Geosciences, 41, 927–945, 2009.

530 Caumon, G., Gray, G., Antoine, C., and Titeux, M.-O.: Three-dimensional implicit stratigraphic model building from remote sensing data on tetrahedral meshes: theory and application to a regional model of La Popa Basin, NE Mexico, IEEE Transactions on Geoscience and Remote Sensing, 51, 1613–1621, 2012.

Chaodong, F., Peng, Y., and Bo, X.: Rapid geological modeling by using implicit 3D potential field interpolation method, in: 2010 International Conference On Computer Design and Applications, vol. 5, pp. V5–50, IEEE, 2010.

535 Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks, IEEE Transactions on Geoscience and Remote Sensing, 54, 6232–6251, 2016.

Chiles, J.-P., Aug, C., Guillen, A., and Lees, T.: Modelling the geometry of geological units and its uncertainty in 3D from structural data: the potential-field method, in: Proceedings of international symposium on orebody modelling and strategic mine planning, Perth, Australia, vol. 22, p. 24, Citeseer, 2004.

540 Collon, P., Steckiewicz-Laurent, W., Pellerin, J., Laurent, G., Caumon, G., Reichart, G., and Vaute, L.: 3D geomodelling combining implicit surfaces and Voronoi-based remeshing: A case study in the Lorraine Coal Basin (France), Computers & Geosciences, 77, 29–43, 2015.

Cook, R. L.: Stochastic sampling in computer graphics, ACM Transactions on Graphics (TOG), 5, 51–72, 1986.

de Kemp, E. A., Schetselaar, E. M., Hillier, M. J., Lydon, J. W., and Ransom, P. W.: Assessing the workflow for regional-scale 3D geologic modeling: An example from the Sullivan time horizon, Purcell Anticlinorium East Kootenay region, southeastern British Columbia, Interpretation, 4, SM33–SM50, 2016.

545 Donmez, P.: Introduction to Machine Learning, by Ethem Alpaydin, 2010.

Fabin, C. E., Correia Filho, O. J., Alencar, M. L., Barbosa, J. A., MIRANDA, T. S., Neumann, V. H., Gomes, I. F., and SANTANA, F. R.: Stratigraphic relations of the Ipubi formation: siliciclastic-evaporitic succession of the Araripe Basin, Anais da Academia Brasileira de Ciências, 90, 2049–2071, 2018.

550 Fossen, H.: Structural geology, Cambridge University Press, 2016.

Geng, Z., Wu, X., Shi, Y., and Fomel, S.: Deep learning for relative geologic time and seismic horizons, Geophysics, 85, WA87–WA100, 2020.

Guillen, A., Calcagno, P., Courrioux, G., Joly, A., and Ledru, P.: Geological modelling from field data and geological knowledge: Part II. Modelling validation using gravity and magnetic data inversion, Physics of the Earth and Planetary Interiors, 171, 158–169, 2008.

555 Hennenfent, G. and Herrmann, F. J.: Simply denoise: Wavefield reconstruction via jittered undersampling, Geophysics, 73, V19–V28, 2008.

Hillier, M., Wellmann, F., Brodaric, B., de Kemp, E., and Schetselaar, E.: Three-Dimensional Structural Geological Modeling Using Graph Neural Networks, Mathematical Geosciences, 53, 1725–1749, 2021.

Hillier, M. J., Schetselaar, E. M., de Kemp, E. A., and Perron, G.: Three-dimensional modelling of geological surfaces using generalized interpolation with radial basis functions, Mathematical Geosciences, 46, 931–953, 2014.

560   Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314–1324, 2019.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861, 2017.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer

565   parameters and< 0.5 MB model size, arXiv preprint arXiv:1602.07360, 2016.

Ioffe, S. and Normalization, C. S. B.: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167, 2014.

Jessell, M.: Three-dimensional geological modelling of potential-field data, Computers & Geosciences, 27, 455–465, 2001.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

570   Lajaunie, C., Courrioux, G., and Manuel, L.: Foliation fields and 3D cartography in geology: principles of a method based on potential interpolation, Mathematical Geology, 29, 571–584, 1997.

Laurent, G., Aillères, L., Caumon, G., and Grose, L.: Folding and poly-deformation modelling in implicit modelling approach, 34th Gocad Meet. Proc, pp. 1–18, 2014.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S.: Feature pyramid networks for object detection, in: Proceedings

575   of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125, 2017.

Lindsay, M. D., Aillères, L., Jessell, M. W., de Kemp, E. A., and Betts, P. G.: Locating and quantifying geological uncertainty in three-dimensional models: Analysis of the Gippsland Basin, southeastern Australia, Tectonophysics, 546, 10–27, 2012.

Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P.: Convolutional neural networks for large-scale remote-sensing image classification, IEEE Transactions on geoscience and remote sensing, 55, 645–657, 2016.

580   Mallet, J.: Three-dimensional graphic display of disconnected bodies, Mathematical geology, 20, 977–990, 1988.

Mallet, J.-L.: Discrete smooth interpolation in geometric modelling, Computer-aided design, 24, 178–191, 1992.

Mallet, J.-L.: Discrete modeling for natural objects, Mathematical geology, 29, 199–219, 1997.

Mallet, J.-L.: Elements of mathematical sedimentary geology: The GeoChron model, EAGE publications, 2014.

McInerney, P., Goldberg, A., Calcagno, P., Courrioux, G., Guillen, A., and Seikel, R.: Improved 3D geology modelling using an implicit

585   function interpolator and forward modelling of potential field data, in: Proceedings of exploration, vol. 7, pp. 919–922, 2007.

Perol, T., Gharbi, M., and Denolle, M.: Convolutional neural network for earthquake detection and location, Science Advances, 4, e1700 578, 2018.

Phillips, J. D., Hansen, R. O., and Blakely, R. J.: The use of curvature in potential-field interpretation, Exploration Geophysics, 38, 111–119, 2007.

590   Renaudeau, J., Malvesin, E., Maerten, F., and Caumon, G.: Implicit structural modeling by minimization of the bending energy with moving least squares functions, Mathematical Geosciences, 51, 693–724, 2019.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C.: Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520, 2018.

Shewchuk, J.: What is a good linear finite element? interpolation, conditioning, anisotropy, and quality measures (preprint), University of California at Berkeley, 73, 137, 2002.

Shi, Y., Wu, X., and Fomel, S.: SaltSeg: Automatic 3D salt segmentation using a deep convolutional neural network, Interpretation, 7, SE113–SE122, 2019.

Souche, L., Iskenova, G., Lepage, F., Desmarest, D., et al.: Construction of structurally and stratigraphically consistent structural models using the volume-based modelling technology: Applications to an Australian dataset, in: International petroleum technology conference, International Petroleum Technology Conference, 2014.

Viard, T., Caumon, G., and Levy, B.: Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions?, Computers & Geosciences, 37, 511–520, 2011.

Wang, Z., Simoncelli, E. P., and Bovik, A. C.: Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, pp. 1398–1402, Ieee, 2003.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing, 13, 600–612, 2004.

Wellmann, F. and Caumon, G.: 3-D Structural geological models: Concepts, methods, and uncertainties, in: Advances in Geophysics, vol. 59, pp. 1–121, Elsevier, 2018.

Wu, X., Liang, L., Shi, Y., and Fomel, S.: FaultSeg3D: using synthetic datasets to train an end-to-end convolutional neural network for 3D seismic fault segmentation, GEOPHYSICS, 84, IM35–IM45, 2019.

Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., and Caumon, G.: Building realistic structure models to train convolutional neural networks for seismic structural interpretation, Geophysics, 85, WA27–WA39, 2020.

Wu, Y., Lin, Y., Zhou, Z., Bolton, D. C., Liu, J., and Johnson, P.: DeepDetect: A cascaded region-based densely connected network for seismic event detection, IEEE Transactions on Geoscience and Remote Sensing, 57, 62–75, 2018.

Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M., and Do, M. N.: Semantic image inpainting with deep generative models, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5485–5493, 2017.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., and Sang, N.: Learning a discriminative feature network for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1857–1866, 2018a.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S.: Generative image inpainting with contextual attention, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5505–5514, 2018b.

Zhao, H., Gallo, O., Frosio, I., and Kautz, J.: Loss functions for image restoration with neural networks, IEEE Transactions on computational imaging, 3, 47–57, 2016.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J.: Unet++: A nested u-net architecture for medical image segmentation, in: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3–11, Springer, 2018.