

## Responses to comments from reviewers

Dear reviewer,

We sincerely appreciate all your careful reviewing so that we could get the reviewed manuscript promptly. We appreciate all your valuable comments and suggestions, which help a lot to improve our manuscript. We have corrected the figures and spelling mistakes and made modifications in the manuscripts according to your comments. Below we are trying to responses all your comments, suggestions, and questions. Let's discuss more if some of our explanations in the responses are not clear to you. The related modifications are not shown in the responses but are all marked in the manuscript revision history.

Thanks!

### ● **Introduction:**

1) **Line 26: Calcagno et al 2008 is a more appropriate reference than Guillen et al 2008. Guillen focusses on geophysics, not so much on structural modelling.**  
Thanks. We have modified reference citation according to your suggestion.

2) **Line 28: “interpreting” should be “interpolating”?**  
Thanks for your suggestion. We have modified the related words.

3) **Line 33: Please elaborate on what you mean by “empirical rules” – is this the tectonic history or topology (i.e. stratigraphy; fault-fault and fault-stratigraphy relationships)? See Calcagno et al 2008 for one example).**

Thanks for your comments. We have added the corresponding texts to demonstrate “empirical rules” for improving readers’ understanding to the content. The empirical knowledge can be mentally inferred from the structural data to constrain the possible geometrical relationships among the interfaces of geological formations and drive the modeling behavior of the implicit methods. The input structural data of the implicit method typically consists of various types of modeling objects, such as spatial points, vectors, polylines, and surfaces, interpreted by the geologists and geophysicists from field observations. The implicit method benefits from incorporating available geological information into the resultant model by integrating the observed data and the empirical knowledge, providing an effective alternative to reproduce the geometry of the subsurface from a global perspective.

4) **Line 33: “file observation” is “field observation”?**  
Thanks for your comment. Corrected.

5) **Line 37: “or regional available” - this part of the sentence doesn’t make sense.**

Thanks for your suggestion. We have deleted “or regional available” in the correspond content of our manuscript.

- 6) Line 47: “DIS” should be “DSI”.

Thanks, Corrected.

- 7) Lines 56 to 69: The authors make some valid points. Also worth mentioning is implicit methods suffer with dense and clustered data points, which can be addressed through downsampling. The problem is that downsampling brings its own issues, such as how or whether to aggregate data points and what appropriate algorithms from which to produce representative inputs to implicit methods. I suppose this is what you describe Lines 128-130?

Thanks for your comment. We have added the related discussions in section “Introduction” of the manuscript to supplement the potential challenges of “clustered data points” in the existing implicit approaches based on your suggestions. On the one hand, structural interpolation fully guided by mathematical equations might not always produce a geologically valid model given sparse data in some complex geological circumstances. On the other hand, the implicit methods also suffer from the unevenly distributed data, such as dense or clustered points, in which how to appropriately aggregate the structural information and produce representative inputs is still an active area of research.

- 8) Line 61 to 62: Also look at Grose et al 2018; 2021 (10.1029/2017JB015177; 10.5194/GMD-143915-2021) for implicit methods that incorporate greater number of relevant structural data as constraints.

Thanks for your comment. We have added the reference citations associated with the implicit methods that incorporate different relevant structural constraints in our manuscript.

- 9) Line 70 to 73: The authors need to explain how “past experiences” are relevant to structural modelling. What is ‘example data’? Observations, or other models (in which case, not data). Providing examples would help here.

Thanks for your comment. We have added the related texts to demonstrate how ‘example data’ are relevant to structural modeling in the manuscript. In the structural modeling, the ‘example data’ are considered as numerous geological models with full structural annotations, which are used to create a sufficiently large dataset for supervising the optimization of the CNN model parameters. Deep learning is a type of data-driven and statistical approach that estimate an implicit function that maps inputs to outputs based on past experiences or example data by optimizing a given quality criteria or loss function. To build example or training dataset, we first 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. Then we randomly some mask parts of the synthetic models for

generating the sparsely and unevenly distributed horizon data, which together with the fault data are used as inputs of our CNN-based modeling process.

- 10) Line 74: “without defining a physical process”. Implicit modelling doesn’t require a description of a process, but geometrical and topological constraints. The resulting model geometries can be interpreted to represent some process, but that is not quite what is written here. Process models are an entirely different class to structural ones, at least in the context described earlier in the document.

Thanks for your comment. We have deleted the phrase that confuses the readers and corrected the related sentence as, “By contrast with the traditional approaches, deep learning is beneficial for making a prediction without solving a sophisticated linear system of equations with prescribed mathematical constraints at cost of expensive computation”.

- 11) Line 78: citation “Ioffe and Normalization” is Normalization really the name.

Thanks. We have modified the corresponding reference citation.

- 12) Line 88: You imply that GNNs being not repeatable is a problem... I assume this means that your method is? Perhaps state how your NN architecture is able to ensure repeatability (presumably because the parameters are not randomly initialised?)

Thanks for your comment. The CNN can produce a structural scalar field as an implicit representation of all the geological structures from various types of data based on the knowledge learned from a large training dataset. In comparison to GNN, the solutions of our approach can be reproducible as it is not necessary to randomly initialize the parameters of the network at each modeling process. Once trained, we can observe that the CNN can efficiently provide a geologically reasonable structural model in the synthetic and real-world data applications, showing promising potential for further leveraging deep learning to improve many geological applications.

- 13) Line 99: How is the scalar field produced? Automated model building (how?) Masking why? Many questions about where the rendering happens.

Thanks for your comment. We have modified the related texts to improve the readers’ understanding to the whole workflow of our CNN-based method. Our network can produce a structural scalar field as an implicit representation of all the structures from various types of structural data such as horizons that encode the stratigraphic sequence of the sampled interfaces, and faults in the presence of the geological boundaries. To achieve this, we are first required to prepare a training dataset. As discussed in the manuscript, 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. Then we utilize a hybrid loss function that combines element-wisely fitting on the known values and multi-scale structural

similarity over the local sliding windows to train the network. Once trained, the CNN can efficiently provide a geologically reasonable structural model from the heterogeneously sampled data based on the knowledge learned from the training dataset and implicitly embedded in the CNN model parameters.

- **Method:**

- 1) Skip connections are an important feature in your NN implementation. You need to explain how they work and why they are necessary so that the reader understands how they allow what appears to be multiscale modelling.

Thanks for your comment. We have modified the related texts in the subsection “Network Architecture” of our manuscript. Our developed CNN architecture uses a U-shaped framework with the encoder and decoder branches linked by using skip connections. In such a framework, the features are first downsampled at different spatial resolutions in the encoder and then recombined with their upsampled counterparts through skip connections in the decoder. In general, 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. Therefore, as the hidden representations with different spatial resolutions have much distinctive geometrical information, systematically aggregating the multi-scale features via skip connections can provide hierarchical constraints for a reliable and stable structural field prediction. Furthermore, the low-level features computed from the shallow layers can 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 is helpful to enhance the low-level features and produce a model structurally consistent with the inputs.

- 2) Line 168: What is “MAdds” referring to?

Thanks. “MAdds” refers to the number of multiply–accumulate operations. It is typically used to measure the computational complexity of the method.

- 3) \* Line 180: Inverse-distance weighting schemes are well-used in spatial stats, but just because it’s consistent with traditional methods, doesn’t mean it’s appropriate. For example, what about in geological scenarios where there is a large shear zone separates one geological terrane from another? The data on each side of the shear zone may be spatially close, but may be quite distant contextually. So, the assumption of weighting on spatial proximity goes against some of your criticisms of existing methods in the introduction. It’s a known drawback of other implicit geomodelling schemes. Justifying this assumption is important, and please provide supporting references. In fairness, this is quite a difficult problem that I don’t believe you’re attempting to solve directly with this contribution, Nonetheless, it’s worthwhile discussing as a potential limitation.

Thanks for your comment. We have modified the corresponding paragraph in the subsection “Network Architecture” of our manuscript. Our CNN is designed to progressively complete structural features layer by layer through sequential non-linear convolutional filters that are conditioned on the previous convolutions. The valid convolutional responses only exist near the input structures in the starting layer of the network. To spread geological structures elsewhere, each convolutional layer collects information from the previous layer outputs within an increasingly expanding receptive field by recursively stacking convolutions and down-sampling the input hidden features. Therefore, at the bottom layers of the CNN, the structural information in the inputs can be used to constrain the modeling process over the entire volume from a global receptive region of view. This characteristic allows the network to correctly understand the relations of the spatially distant but contextually close features. Although weighting on spatial proximity is typically used in many traditional structural interpolation methods, the nearby features are not necessarily more significant than distinct ones for making geological-related predictions. For example, when the stratigraphic terranes are located on the opposite of a large shear zone or other discontinuous structures, the correlations of distinct data points computed from a global review can be helpful to capture a more accurate structural pattern.

4) [Line 190: provide references for which methods use MSE/MAE.](#)

Thanks for your suggestion. We have added the associated reference citations in the manuscript.

5) [Line 204: provide references for SSIM.](#)

Thanks for your comment. We have added the corresponding reference paper in the manuscript.

6) [Line 210: the term  \$G\_{\sigma\_g}\$  is missing from eq. 2? As is  \$\sigma\_g\$ .](#)

Thanks for your suggestion. In the definition of Structural Similarity (SSIM), variances and covariance are computed by using a Gaussian filter with standard deviation  $\sigma_g$ . Approximately,  $\mu_x$  and  $\sigma_x$  can be viewed as estimates of the stratigraphic sequences and structural variations in a local patch of the structural model, respectively, and  $\sigma_{xy}$  measures the tendency of the two patches in models being compared to vary together, thus an indication of structural similarity. However, in the conventional SSIM, the standard deviation  $\sigma_g$  of the Gaussian filter requires to be defined before training. However, the choice of  $\sigma_g$  might influence the prediction accuracy because the CNN trained by using a large standard deviation might overly emphasize the local variations and generate spurious features while blurring sharp structural boundaries for a small standard deviation. Instead of fine-tuning the parameter, we use Multi-scale Structural Similarity (MS-SSIM) to compute a pyramid of patches with different scale levels defined by various standard deviations of the Gaussian filter.

- 7) Lines 215: I see that  $\sigma_g$  is a ‘superparameter’. It may help to define this as a ‘hyperparameter’ to fit with ML semantic conventions, unless there is a specific reason you’re using ‘super’. If so please explain why ‘super’ and not ‘hyper’.

Thanks. Corrected. We have replaced “super-parameter” with “hyper-parameter” in the related sentence of the manuscript.

- 8) Line 231: ‘artifacts’ are old and interest physical objects from antiquity, ‘artefacts’ are unintended side-effects from numerical processes.

Thanks. Corrected.

- 9) Line 231: what is ‘it’ in the phrase “such as MAE to the MS-SSIM as \*it\* is only related to the values on a single point” the artefacts? Or the MAE... ?

Thanks for your comment. We have modified the related phrases in this paragraph to improve the reader’s understanding to why we use a joint loss function that combines the MAE with the MS-SSIM for training the network. Although the MS-SSIM can highlight structural variations focusing on a neighborhood of grid point as large as the given Gaussian filter, it might produce artefacts in predictions because its derivatives cannot be correctly estimated near the boundary regions of the patch. This can be alleviated by supplying an element-wise criterion that is computed on a single point of the two patches being compared in the loss function. In addition, using MS-SSIM alone might not be sensitive to uniform biases, which might cause unexpected changes in stratigraphic sequences or shifts of geological interfaces in modeling results.

● **Data preparation:**

- 1) Given how important the training data is for your procedure, you need to have at least a two or three sentences explaining the method of Wu et al 2020. Wu et al 2020 generates a training set using a CNN for your CNN... Both methods produce a scalar field from which to render geology – so it would be helpful to make clear how your approach differs.

Thanks for your constructive comment. We have added the related discussions in the subsection “Automatic Data Generator” in our manuscript. As far as I’m concerned, one of the most significant steps of applying the supervised learning method is the preparation of many example data and especially the corresponding labels for training the network. In the structural modeling problem, the training dataset should incorporate structurally various geological models to allow the CNN to learn representative knowledge and achieve its reliable generalization in real-world applications. However, it is hardly possible to manually label all geological structures in a field survey as the ground truth of the subsurface is inaccessible. To overcome this problem, we adopt an automatic workflow to simulate geological structural models with some typical folding and faulting features that are controlled by a set of random parameters. 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 add folding, dipping, and faulting structures to

further complicate the geometrical features of this model. We simulate the folding and dipping structures by vertically shearing the initial flat model using a combination of linear and Gaussian-like shift fields, while creating faulting structures through volumetric vector fields that are defined around the fault surfaces. By randomly choosing the parameters within the reasonable ranges, we can simulate numerous geological models with diverse and realistic features not limited to a specific structural pattern to enrich the training dataset. Based on the models with known features, we can simultaneously obtain the corresponding ground truth of fault data with ones on faults and zeros elsewhere. The generated models can be viewed as a structural scalar function because their iso-surfaces represent the corresponding stratigraphic interfaces while the local value jumps indicate the structural discontinuities. This is important for our next step of constructing an example dataset to train our CNN for geological structural modeling.

- 2) Line 262: This is a question commonly asked, but needs to be addressed. How do you know 600 models is enough given the uncertainties of a result that you admit cannot be easily validated (L253)?

Thanks for your comment. Although the CNN model is trained by using only 600 synthetic structural models, it still works pretty well to model the geological structures from the sparse and unevenly distributed structural data that are recorded at totally different field surveys. We cannot make sure the current number of the training samples is the best one, but further tuning might be very time-consuming and hardly obtain further improvements. Considering the used training dataset is still not sufficiently large to train a 3-D deep network, future works will focus on further complicating the data simulation workflow by adding more complex and representative geological patterns in the synthetic structural models.

● **Implementation:**

- 1) Section 4.1: Training and validation. I understand the need for normalization, but it's not clear how this is achieved on the entire model. Structural data, at least in most implicit schemes consists of contact observations (X,Y,Z coordinates, or U,V,W and a 'lithology' label with topology, usually a stratigraphic direction or polarity) and orientation data (X,Y,Z, type [contact orientation; fault orientation; fold hinge; etc] a vector; usually expressed as dip direction/dip and the lithology it represents). How do you simultaneously normalize coordinates and vectors? What is method for normalization e.g. min-max?

Thanks for your comment. We have added the related sentence in the subsection of "Training and Validation" in our manuscript. Considering the coordinate ranges of the geological dataset can be much varying from each other, we rescale all the structural data and scalar functions to obtain the normalized training samples. This normalization is implemented by applying a global shift to the entire structural model to make it range from zero and one, which would not change its geological structures. Therefore, we assign the scatter points on the same geological interface to the corresponding iso-values of the normalized model when processing the

structural data. In normalizing orientation vector data, we use the cosine similarity to compare the observed structural orientations and the gradients of the computed model for eliminating their amplitude differences.

- 2) [Lines 328 to 329: You do explain some of these metrics elsewhere in the manuscript, but not all e.g. R2S. You may think to supply these explanations as supplementary material.](#)

Thanks for your constructive suggestion. We have supplied the corresponding explanations of the used regression metrics as supplementary material in Appendix A. To verify the modeling performance of our CNN, we quantitatively measure the differences between the ground-truth structural models and predictions by using various regression metrics including SSIM (Structural Similarity), MSE (Mean Square Error), MAE (Mean Absolute Error), EVS (Explained Variance Score), MSLE (Mean Square Logarithm Error), MDAE (Median Absolute Error), and R Square Score (R<sup>2</sup>S) in the validation dataset. MSLE measures the prediction performance that corresponds to the expected value of the squared logarithmic error. MDAE is computed by taking the median of all absolute differences and thus can be robust to outliers. EVS is used to measure the proportion of the variability of the predictions in a machine learning method, and the score value can range from zero to one. Higher EVS typically indicates a stronger strength of association between regression targets and predictions and thus represents better network performance. R<sup>2</sup>S, also called the coefficient of determination, offers a measure of how well the predictions of the network are based on the proportion of total variation. R<sup>2</sup>S is similar to EVS, with the notable difference that it can account for systematic offsets in the predictions. In addition, EVS and R<sup>2</sup>S can be more robust and informative than MAE and MSE in regression analysis evaluation because the former can be represented as percentage errors.

- 3) [Line 339 to 346: This paragraph belongs in the discussion.](#)

Thanks for your suggestion. We have placed this paragraph into the section “Discussion” of the manuscript.

- 4) [Line 348 to 350: The context of this problem isn't clear. Annotate with what, and do other applications do this, but poorly, or not do it at all? What do you mean horizon values – it would seem that it is the iso-value from the scalar field \(or fields, especially if considering faults.\). Provide some examples with relevant citations.](#)

Thanks for your suggestion. We have modified the related texts in the subsection “Horizon Annotation Experiment” of our manuscript. When training the network, the horizon points along the geological interfaces are assigned to the corresponding iso-values of the implicit model, which is not always the case for real-world modeling applications. Although the annotation of fault points is straightforward by assigning ones on faults and zeros elsewhere, how to annotate appropriate values on the horizon points in real-world applications still remains a problem as the ground-truth model is inaccessible before computation. Therefore, we implement a



data experiment using the horizons with varying annotations and the same faults to study how the different values on the horizon data impact the modeling results of our CNN. By visual comparison in Figure 8, the modeling results are nearly identical to each other, which indicates that our method is not sensitive to the annotations of the input horizon data, which is what we expect. Based on this observation, we recommend to assign scattered points on each horizon to the average vertical coordinate. It is worth noting that the horizon annotations require to be normalized by the model size for consistency with the training dataset.

- **Application:**

- 1) Section 5.1: Provide a sentence describing the geological scenario and location of study area #1.

Thanks for your comment. We have added the corresponding sentences in the subsection “Real World 2-D Case Studies” of our manuscript. In this subsection, we apply the trained CNN to a field 2-D seismic dataset to verify its modeling performance for the incomplete structural interpretations with geometrical patterns distinct from the training data. The input structural data are manually interpreted from the seismic images that are randomly extracted from the Westcam dataset. This dataset is acquired in regions with closely spaced and complexly crossing faults with large slips, in which the seismic images are of low resolution due to insufficient coverage.

- 2) Section 5.1: 2d case studies are simple and not a great test as the horizons are laterally continuous being essentially flat. The fault displacements seem to be honored, which is good. But existing model packages are very effective at producing models with similar simple geometry and topology.

Thanks for your comment. We have added the corresponding sentences in the subsection “Real World 2-D Case Studies” of our manuscript. In this case study, we apply the trained CNN to a field 2-D seismic dataset to verify its modeling ability in the incomplete structural interpretations with geometrical patterns distinct from the training data. The ambiguous seismic reflections shown in Figure 9a are difficult to be continuously tracked, which causes the noisy and partially missing horizon data displayed in Figure 9b. The faults shown in Figure 9c also might not be fully detected because of the seismic data-incoherent noise and the stratigraphic features that are apparent to structural discontinuities. Moreover, the structural contradictions and hard-to-reconcile features in the structural data might negatively impact the modeling performance of the implicit methods. Therefore, there still remains a challenging task for many traditional approaches to obtain a geologically reasonable model that is structurally consistent with the inputs.

- 3) Line 401: how do you define complex geological settings here? Polydeformed beds? Domains with vastly different tectonic histories? Multiple magmatic intrusion and extrusion events? Overturned bedding?

Thanks for your constructive comment. We have modified the related words to improve reader's understanding to the content of the complex geological settings in the paragraph. The first seismic data sampled in regions with complexly deformed structures have relatively low resolution and signal-to-noise ratio. As is shown in Figure 11a, some seismic reflections are noisy and difficult to be continuously tracked across the entire volume of interest. The closely spaced and curved faults further complicate the structures especially when there exists data-incoherent noise and stratigraphic features that are similar to structural discontinuities. Therefore, it is challenging for the existing approaches to accurately model the structures from the noisy and partially missing horizons and faults.

- 4) Figs 11 and 12 have the necessary content, but are too small. Enlarge each panel in the figure, and use more space in the manuscript. Annotations would also help the reader, such as pointing out where the complex parts of the model are (see previous comment), where “the seismic reflections are partially ambiguous and difficult to be continuously tracked” and “results shown in Figure 11c demonstrate that our CNN architecture is beneficial for 3-D structural modeling by predicting a geologically valid model”.

Thanks for reminding us of this important point. We guess this question is similar to your Question 3). We have enlarged each panel of the figures and added some annotations in the corresponding figure captions to point out where the complex part of the geological models and seismic structures are for improving the reader's comfort and ease his/her understanding of the content in the subsection “Real World 3-D Case Studies” of our manuscript.

- **Discussion:**

- 1) Section 6.1 and Fig 13: I like the uncertainty analysis. One thing worth pointing out is that you deal with both aleatory uncertainty (uncertainties resulting from measurement error) and epistemic uncertainty (relating to missing knowledge or data) citing Pirot et al 2022 in GMD. You remove/add drill holes in your fig 13, which is a nice test, so it is worth highlight that point here, or at the very least in the figure caption.

Thanks for your comment. We have modified the texts and added the related reference citation in the subsection “Structural Uncertainty Analysis” of the manuscript. When 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 heterogeneously distributed inputs might cause an ill-posed problem that multiple plausible structural models exist to 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. Although the existing implicit methods can generate various models by perturbing the inputs to characterize uncertainties, they might not explore a broad range of possible geological patterns and structural relationships in nature by using a single model suit for stochastic simulation. Working on the automating of

modeling workflow, our CNN is beneficial for a flexible interpretation of aleatory and epistemic uncertainties by generating diverse modeling realizations instead of one best due to its high computational efficiency.

- 2) Lines 484 to 487: This is good to acknowledge as modelling these types of structures realistically is challenging, however you can elaborate on why your method is unable to replicate them. I would think the type of geological object is arbitrary given the framework you developed. Is it not possible because the synthetic model generation of Wu et al 2020 cannot recreate unconformities and intrusions? A possible solution maybe to expand the training set to a wider range of objects? Another source could be the Noddyverse (Jessell et al .2022) where dykes, plugs and unconformities are represented. Perhaps you can comment in whether this would work or not. It could be that because you're using a scalar field in the tradition of Lajuane et al 1997, in which intrusions and unconformities (as distinct from onlap relationships) have not yet been solved (at least to my knowledge).

Thanks for your constructive comment. We have added the related discussions in the subsection "Current Limitations and Improvements" of the manuscript. As structural modeling is dependent on the analysis of the spatial relations of the observed structures to interpolate new geologically valid structures elsewhere, acquiring representative example data is essential for training the CNN to achieve its reliable generalization ability. Therefore, we adopt an automatic workflow to generate diverse models with realistic structures and simulate partially missing horizons. Although working well to recover faulted and folded structures, our CNN might not represent other geological structures not considered in our training dataset, such as unconformities and igneous intrusions. It 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 data generator. Considering the used training dataset is still not sufficiently large to train a 3-D deep network, future works will focus on the expansion of the training dataset to a broad range of geological geometries and relationships. For example, we can further complicate the simulation workflow by adding more complex and diverse features in the structural models, or using a recently developed open-access 3-D modeling dataset where dykes, plugs, and unconformities are represented.

- **Code and data availability:**

- 1) I attempted to test the github code on a Windows operating system and had a few issues with conflicting dependencies with installed modules. In particular, I received open MPI errors: "OMP: Error #15: Initializing libiomp5md.dll, but found libiomp5md.dll already initialized." In attempting to fix this, a number of other dependencies now no longer work.

Thanks for your comment. We suggest using Anaconda and Jupyter Notebook to run our code package on a Windows operating system. All the dependent libraries can be easily installed via "pip install -r requirements.txt" by following the instruction.