

Geosci. Model Dev. Discuss., referee comment RC2  
<https://doi.org/10.5194/gmd-2022-117-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on gmd-2022-117**

Anonymous Referee #2

---

Referee comment on "DeepISMNet: three-dimensional implicit structural modeling with convolutional neural network" by Zhengfa Bi et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-117-RC2>, 2022

---

This manuscript presents a new deep learning-based approach for implicit geological modeling using a customized UNet architecture. It's a very interesting approach and shows some promising first results on the synthetic examples and case studies using interpretative inputs. However, there is a large issue with respect to 3D applications (detailed below) that really limits its applicability. Therefore, I suggest developing a workable solution for 3D CNNs to demonstrate it can work with typical 3D implicit inputs, or clearly focusing the manuscript as a method for 2D interpretative inputs. The approach has a lot of potential if the issues can be addressed. Additionally, I have many suggestions and comments to improve the manuscript, with major points given below first. The quality of the writing is generally good.

Issues/Suggestions/Questions:

The deep learning architecture is trained using a 2D network (2D convolutions): Section 2.1, P13 L267, P24 L489-491. This is a flaw for 3D applications. The biggest issue is that applying the 2D trained network on a very sparse 3D dataset. 2D slices of your 3D grid must have sampled data points otherwise input to the network is all zero (your binary mask has all zero elements), yielding a scalar field with all zeros. This greatly reduces its applicability for 3D real-world scatter datasets. The presented 3D "real-world" case study had interpretations that covered the modeling domain such that each 2D slice used for inference had sampled data points. To capture geometrical characteristics of 3D geology, 3D convolutions must be employed.

The manuscript shows this approach produces good modelling results for applications

where there are significant interpretations. The inputted interpretations provide densely sampled points along horizons forcing strong lateral continuity. How would the trained CNN generalize on 'real world' field observation datasets that are very heterogenous? E.g., sampled points are sparse (large varying distances/gaps between points on the same interface) in some regions along with possible dense sampling from exposed outcrop or borehole in localized parts. It is claimed in the discussion "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". This is not substantiated. Most geological survey real world datasets are noisy and very heterogenous, this type of data hasn't been used. Interpretative datasets with near continuous sampling along interfaces (horizons/faults) were used. Manuscript would provide larger impact if it can be shown it can be applied to noisy and very heterogenous 3D datasets.

Fig. 2 Diverges from how UNet (encoder/decoder) architecture is normally illustrated and makes it confusing. Update so that readers can understand the flow of representations and the figure in general. Clearly indicate tensor dimensionalities. The bracket [x256] I presume means for 3D convolutions which are not employed in any results so why is it here? Expansion mean increasing dimension of representations by 1x1 convolutions right? why is 1.5?. Show where the bottleneck is, since you refer to it in the text.

How scattered points are sampled from voxel/pixel grids, e.g., jittering method, how do the 7x7 patches work, how 2<sup>nd</sup> order scalar gradients are computed from the voxel/pixels. Requires figure.

To produce a model for scattered dataset, there is a preprocessing step that requires the associated scattered points to grid cells to obtain the mask m. Correct? If so, it's important to mention.

While authors have their own limited 3D geological model simulation workflow the following work should be referenced.

Jessell, M., Guo, J., Li, Y., Lindsay, M., Scalzo, R., Giraud, J., Pirot, G., Cripps, E., and Ogarko, V.: Into the Noddyverse: a massive data store of 3D geological models for machine learning and inversion applications, *Earth Syst. Sci. Data*, 14, 381–392, <https://doi.org/10.5194/essd-14-381-2022>, 2022.

This could be leveraged for future CNN training, and a potential solution to challenges with training 3D CNN UNet architecture.

Detailed Comments/Edits:

P1 L2-L8. All approaches are formulated in terms of a spatial interpolation problem. Within each approach there are numerous mathematical methods to solving the problem. There are very few that setup the problem as a variation problem in which PDEs are solved. Update this part.

P1 L17: Reword. Suggest not using the term 'reasonable geological' until what this means is defined

P1 L20 Incorrect reference. Given reference is for implicit modeling! There is a suite of good reference for explicit methods.

P1 L20-L21: "It intends"?

P2 L27-L33: Suggest clearly identifying the advantages over explicit: Fast, reproducible, updatable.

P2L31-L32: "resultant model from a global respective"?

P2 L33: spelling "File"

P2 L37: "or regional available in"?

P2 L38: "simplification as structural constraints"? explain

P2 L45-47: To deal with this, discontinuous structures are modeled first, which are used to produce a constrained unstructured mesh. Modeling domains are assigned, for which different scalar fields can operate given the geological history. This can work well. Update this.

P2 L53-55. There are methods to deal with this; each having advantages and disadvantages. Fast rbf, different covariance functions, etc.

P3 L64. Not stated correctly

P3 L65. Incorrect reference de Kemp EA, Jessell MW, Aillères L, Schetselaar EM, Hillier M, Lindsay MD, Brodaric B (2017)

Earth model construction in challenging geologic terrain: designing workflows and algorithms

that make sense. In: Proceedings of exploration 17: sixth decennial international conference on

mineral exploration, pp 419–439

P3 L74. 'sophisticated' remove

P3 L75 'is essential for its' remove

P4 Fig 3: Unclear why features @  $z_{max}$  have large activations on encoder branches (there isn't any data there) and is horizontally consistent (same value at  $\sim z_{max}$ ). How is the output restricted to  $[0, 1]$ ? You use tanh or softmax activation function in final layer? If so, please provide details.

P4 L80 – 83. Please add: Kirkwood, C., Economou, T., Pugeault, N. *et al.* Bayesian Deep Learning for Spatial Interpolation in the Presence of Auxiliary Information. *Math Geosci* **54**, 507–531 (2022). <https://doi.org/10.1007/s11004-021-09988-0>

P5 L88-89. Method can reproduce results, just not exactly/perfectly.

P6 L125-126 trade accuracy for efficiency?

P6 L127 regraded (spelling)

P6 L133 "hierarchical constraints" these are not constraints. They multi-scale features have the capacity to influence structural predictions if the learned multi scale features improve prediction accuracy

P6 L141 "doubles its channels" discrepancy with fig 1 if I interpreted correctly (2->38->76->153->307->614), "linear bottleneck" linear?

P6 L142 "low rank nature"?

P6 L147-149 "the block is ..." ?

P7 L151 spare (spelling)

P7 L152 "might" From empirical tests, what do the results indicate? In many applications, it is fine to some elements of the features to be zero. (resulting from relu – preactivation negative).

P7 L157-159 reference?

P7 L168 "fusing and combining" are the same thing?, define MAdds

P7 L173 "non-linear convolutional filters" convolutions are linear transformations.

P8 L 185-189. Indicate the dimensions of  $m$ , "points" you mean voxels/pixels?, define  $f$ , where is  $y$  used in the paper?

P9 L197 reword.  $N$  represents the total number of points within a patch  $p$ ?

P10 L210 define the gaussian filter

P10 L216 "might" it will influence

P10 Eqn 3. Instead of fine-tuning one parameter  $G_{\sigma_g}$ , many new parameters are added in which you fix according to Wang 2003. They use those parameters for an entirely different application. How the results compare with just fssim

P11 Eqn 5. So  $m$  and  $p_{\text{bold}}$  represent the same thing? A patch?

P11 L241. Unclear on size of each patch to  $7 \times 7$ . Need a figure to know the relations between pixels/voxels, sampled points, and patches surrounding data inputs, jittering, etc. Would "these parameters" require tuning for very different structural settings?

P13 L280-281 unclear

P14 Table 2. How many epochs were performed here? Also, negligible differences, since the input constraints are sampled randomly, you can repeat the experiment and other loss functions could come out first/better

P14 L297 "much varying" ?

P14 L298-299 Re: Normalization. Clearly indicate what is normalized? I assume its scalar values.

P14 L300 "bath" spelling

P15 L306-307 reword "has captured...". "very few structural data" subjective term. 5,10, 100, 1000 points?

P15 L309 stability in terms of what? From iteration to iteration? HFA would be computational expensive if the points from isosurfaces are used. What are the details on how it's computed? Are a sub sample of those points used?

P15 L311 provide the details on the validation dataset ( $2400 * 0.1$ ). Why not use the entire validation dataset instead of running 20 x of the random sample of 240 ? Why

compute MSE since it is mentioned (and is correct) that it's not a useful error metric (over bias outliers, etc)?

P15 L312-313 unclear

P15 L320 "complexly"

P16 Fig 8. The scalar value differences (0.3) between interfaces are the same. What happens when the differences change? I presume that as delta between scalar values adjacent interfaces ( $\Delta s$ ) gets larger, accuracy is improved. While as  $\Delta s$  gets smaller, accuracy is decreased. If 50 horizons are modelled  $\Delta s$  between adjacent surfaces is  $\sim 0.02$ . Does  $\Delta s$  impose/bias constant unit thickness between interfaces?

P16 L348 "scatted" spelling

P17 L369 "partially missing horizon" a tiny % is missing, "might not be"? are they or aren't they fully annotated?

P18 L372-373. For the datasets used in the manuscript, traditional methods can make good geological models consistent with the inputs.

P21 Fig 13. "Geological uncertainty analysis" Multiple realizations are made given different interpretations. What "analysis" was done? How is uncertainty here quantified?

P21 L428-429 And relies on the availability of knowledge and how those methods incorporate that knowledge particularly for data sparsity increases.

P21 L432-434. Given a scattered data set of geological observations, the trained CNN produces 1 model. It doesn't generate a "diverse [set] of possible modeling realization" ([] grammar fix). Fig 13 shows given different interpretative inputs, different realizations are produced. Generating interpretative inputs is extremely time consuming, so the point that the CNN are very computational efficient for generating ensemble realizations is incorrect until the method can generate those realizations without interpretative inputs.

P23 L453-454  $z$  variable? Predicted model was  $\hat{y}$  before. Show formula for  $\text{grad } \hat{y}/z$ . This uses the neighbouring cells surrounding a given pixel/point?

P23 Eqn 8. You manually tune lambda, beta? Apply MTL principles? What values did you use?

P24 L489-491 "the used training dataset is still not sufficiently large to train a 3D deep network" To create 3D model, you slice 3D grid into 2D slices, each of which you perform inference on using the trained 2D network? Critical issues here if this is the case.

P24 L496 "low-rank nature"?

P25 L497 "spare" spelling

P25 L506 "noisy structures" sample interfaces are not noisy. You can clearly see scattered points on sampled interface there is no positional (or orientational) fluctuation in adjacent/nearby points.