Response to Reviewer1 comments

Dear Reviewer:

Thanks for your effort to review our manuscript titled " Three-Dimensional Geological Modeling based on Dual-Task Stratigraphy-Aware Attention Networks(Geo-SAN v1.0) ", and now we have just revised this manuscript according to the good suggestions. The details are as follows, and all the revisions are done using track changes in Word.

Comment 1. The physical basis for the stratigraphic sequence similarity matrix is insufficient. In the learnable lithology similarity matrix defined in Section 2.1.2, the denominator value of σ(x) lacks geological basis and sensitivity analysis, and the impact of this hyperparameter on the distinguishability of the 13 lithology types is not discussed.

Response: Thank you for this comment. The impact of hyperparameters on lithology classification was discussed, as follows (lines 213 to 217):

Where $\sigma(x) = \exp(-d(i,j)/\tau)$, and $d(i,j) = |i-j|$ denotes that the quantitative difference between the ordinal numbers of two lithological categories within the stratigraphic sequence, and τ is a stratigraphic decay length governing how rapidly lithological affinity is lost with increasing stratigraphic separation. Setting τ=3 assigns adjacent units a similarity of $e^{-1/3}$ ≈0.72 and units three steps apart e^{-1} ≈0.37, which is consistent with the initialised similarity structure recovered in Figure 5.

Comment 2. The weight settings for the dual-task loss lack an adaptive mechanism. In the total loss function in Section 2.2.4, the specific values of the four weight coefficients α, β, γ, and δ are not explicitly given in the text, and a fixed weighting strategy is used instead of a dynamic adaptive balancing strategy. In the early stages of training, the scalar field loss and the lithology classification loss differ significantly in magnitude; fixed weights may lead to one task dominating gradient updates. It is recommended to supplement the experimental basis for weight selection.

Response: Thank you for this comment. We have used the GradNorm (gradient normalization) method to optimize the training process (Chen et al., 2018). However, this was not mentioned in the paper, so it is noted here. The revised clarifies that (lines 358 to 362):

In this study, the GradNorm (Gradient Normalization) method is introduced to optimize the training process (Chen et al. 2018). GradNorm adaptively adjusts the weights of each task's loss function to ensure that the gradients of each task have similar magnitudes during training, improving the stability and efficiency of training.

Comment 3. The baseline settings for the ablation experiments are not comprehensive enough. Table 2 only compares four internal variants in the ablation experiments, lacking a horizontal comparison with current mainstream deep learning geological modeling methods, and therefore cannot demonstrate the advantages of Geo-SAN over other methods. Furthermore, whether the accuracy difference between M3 and M4 is entirely attributable to the SAN module or the contribution of the dual-task structure is unclear, lacking ablation experiments that isolate the dual-task module.

Response: Thank you for this comment. We have redesigned the ablation experiments and added comparison experiments with mainstream graph neural networks, as follows (lines 399 to 493):

During model training, various configurations significantly affect performance. To isolate the individual contributions of the stratigraphy-aware attention (SAN) mechanism and the dual-task learning head, four models were designed on an identical graph neighborhood aggregation backbone that alternately integrates GAT and GraphSAGE, as detailed in Table 2. Model M1 (GAT-GraphSAGE) is the baseline, using only the GAT/GraphSAGE embedding with the two prediction branches trained independently. Model M2 (GAT-GraphSAGE-DT) additionally activates the dual-task cross-task constraint, the KL-divergence and compatibility terms ($\mathcal{L}_{cross}$) that couple the scalar-field and lithology branches, while excluding SAN. Model M3 (GAT-GraphSAGE-SAN) adds the SAN mechanism to the baseline while keeping the two branches decoupled. Model M4 (GAT-GraphSAGE-SAN-DT) incorporates both SAN and the dual-task constraint, constituting the complete Geo-SAN framework. In all configurations both prediction branches are retained, R^2 and RMSE or accuracy are reported for every model; the Dual-Task toggle controls only whether the cross-task coupling $\mathcal{L}_{cross}$ is applied.

Table 2. Ablation experiments.

Model	Graph neighborhood aggregation	SAN	Dual-Task	Description
M1	GAT/GraphSAGE	×	×	GAT-GraphSAGE
M2	GAT/GraphSAGE	×	√	GAT-GraphSAGE-DT
M3	GAT/GraphSAGE	√	×	GAT-GraphSAGE-SAN
M4	GAT/GraphSAGE	√	√	GAT-GraphSAGE-SAN-DT

Figure 10 presents scalar field prediction metrics, all four models converge after roughly 300 epochs. Both components improve scalar field interpolation relative to the baseline M1 ($R^2 = 0.931$, RMSE = 0.086): enabling SAN raises R^2 to 0.961 and lowers RMSE to 0.073 (from M1 to M3), while enabling the dual-task constraint raises R^2 to 0.955 and lowers RMSE to 0.072 (from M1 to M2). Their effects are complementary, and the full model M4 attains the best result ($R^2 = 0.971$, RMSE = 0.062). The gain from SAN is slightly larger on R^2, consistent with its explicit injection of stratigraphic age and contact relationship priors, which regularizes the continuous scalar field and yields smoother, more geologically consistent transitions.

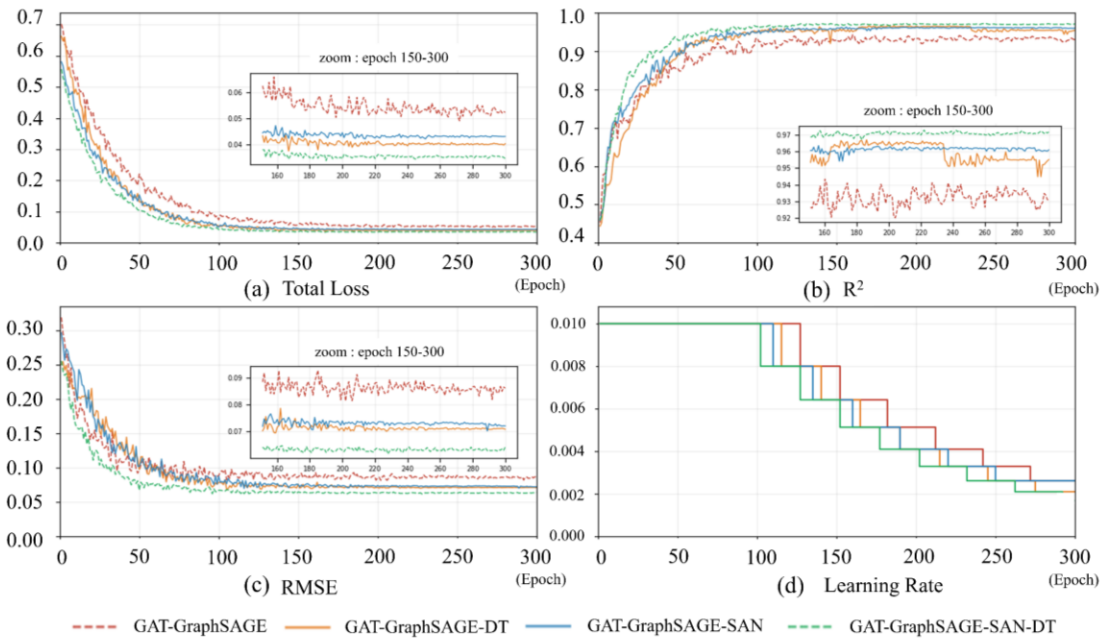

Fig. 10. Model evaluation metrics for scalar field interpolation training processes: (a) total loss, (b) R^2, (c) RMSE, and (d) learning rate.

The lithology classification training curves are shown in Figure 11. Relative to the baseline M1 (accuracy = 0.841), adding the SAN mechanism improves accuracy to 0.879 (from M1 to M3), and with the dual-task constraint, from 0.841 to 0.899 (from M1 to M2), adding the dual-task constraint and SAN mechanism improves accuracy from 0.841 to 0.921 (from M1 to M4). Both factors therefore contribute positively and roughly additively, with the cross-task constraint contributing somewhat more to lithology classification than SAN. The SAN improvement reflects the geological principle that lithologies adjacent in the stratigraphic sequence tend to be similar, explicitly incorporating lithology similarity and stratigraphic sequence constraints. The dual-task improvement reflects the cross-task coupling that regularizes lithological predictions against the stratigraphic scalar field.

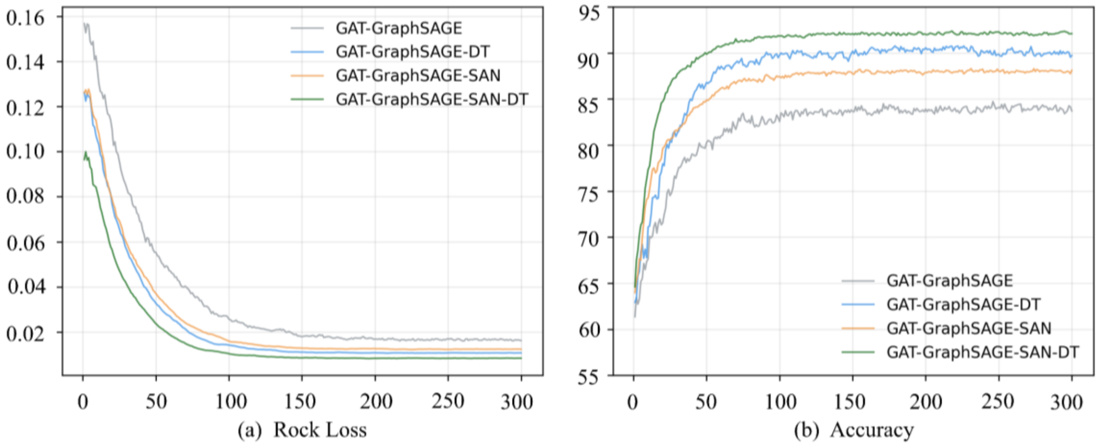

Fig. 11. Model evaluation metrics in lithology classification training processes: (a) rock loss, and (b) accuracy.

To benchmark Geo-SAN against mainstream learning-based interpolators, Table 4 evaluates three standalone GNN backbones, GCN (M5), GAT (M6), and GraphSAGE (M7), each without the SAN prior and without dual-task coupling, on the identical graph, feature set, and sampling split.

Table 4. Comparison experiments.

Model	Graph neighborhood aggregation	SAN	Dual-Task	Description
M5	GCN	×	×	GCN
M6	GAT	×	×	GAT
M7	GraphSAGE	×	×	GraphSAGE

Table 5 summarizes the metrics of the ablation models (M1–M4) and the comparison models (M5–M7). The full model M4 performs best on both tasks (R^2 = 0.971, RMSE = 0.062, and accuracy = 0.921). Relative to the baseline M1, SAN and the dual-task constraint yield consistent gains but favor different tasks: SAN contributes more to scalar field prediction (M3 vs M2, R^2 0.961 vs 0.955), whereas the dual-task constraint contributes more to classification (M2 vs M3, accuracy 0.899 vs 0.879). The declining cross-task loss shows that the two components reinforce each other, and the 8.0% accuracy gain of M4 over M1 stems from both jointly rather than from either alone. Under the baseline setting, the single backbone models rank GraphSAGE > GAT > GCN on both tasks (accuracy of 0.828, 0.810, and 0.805; R^2 of 0.920, 0.904, and 0.879), reflecting the suitability of GraphSAGE's regional aggregation for stratified structures. The alternating GAT+GraphSAGE backbone (M1) surpasses all single backbones (R^2 = 0.931, accuracy = 0.841) at comparable cost, confirming that multi-scale aggregation extracts more representative features from sparse and irregular samples.

Table 5. Model performance evaluation metrics (300 epochs).

Model	$\mathcal{L}_{scalar}$	$\mathcal{L}_{orientation}$	$\mathcal{L}_{litho}$	$\mathcal{L}_{cross}$	R^2	RMSE	Accuracy	Time(s)
M1	0.0056	0.5157	0.3956	/	0.931	0.086	0.841	61.49
M2	0.0052	0.5203	0.4017	0.608	0.955	0.072	0.899	142.37
M3	0.0056	0.5075	0.4156	/	0.961	0.073	0.879	101.78
M4	**0.0053**	**0.4541**	**0.2719**	**0.543**	**0.971**	**0.062**	**0.921**	**162.51**
M5	0.0052	0.5238	0.4103	/	0.879	0.098	0.805	58.12
M6	0.0057	0.5567	0.4237	/	0.904	0.089	0.810	60.24
M7	0.0060	0.5689	0.4297	/	0.920	0.087	0.828	60.76

Comment 4. The coupling mechanism between scalar field prediction and lithological classification is not sufficiently demonstrated. The paper claims that the dual-task framework establishes an "intrinsic coupling" between the scalar field and lithological categories, but Section 2.1.3 only describes a simple shared backbone + dual-branch structure, without explaining the specific interaction between the two tasks at the feature level. The KL divergence term in L_strat transforms the scalar field into lithological prior probabilities, but the mapping function from scalar values to probability distributions in this transformation process is not defined in detail. The confusion between T_1m and P_1m, and D_2d and D_1y, occurs precisely between adjacent stratigraphic units, indicating that cross-task constraints cannot completely resolve the classification ambiguity problem at stratigraphic boundaries.

Response: Thank you for this comment. First, the specific interaction mechanisms between the two tasks at the feature level are explained. Second, the Discussion section addresses the issue of classification ambiguity at stratigraphic boundaries. All have been corrected as summarized below (lines 260 to 268, 625 to 628):

The two branches are coupled at three levels rather than merely sharing a backbone. (1) Reading the same SAN-refined node embedding, so gradients from the two tasks jointly shape the shared representation, forcing features to be simultaneously predictive of the scalar value and of the lithology. (2) The stratigraphy-aware attention injects lithological similarity, geological age and contact similarity, so the aggregation is tied to stratigraphic structure for both tasks. (3) The cross-task loss explicitly links the two predictions , the KL term (soft) aligns the classifier distribution with the distribution implied by the predicted scalar field, and the compatibility term (hard) penalises scalar predictions that fall outside the admissible interval of the predicted lithology.

First, the cross-task constraint narrows but does not eliminate this overlap where some ambiguity is irreducible where lithologies are genuinely similar and contacts are gradational. This is consistent with geological reality and delimits the scope of the coupling.

Comment 5. The research depth on fault handling methods is insufficient. Section 2.1.1 mentions that "faults are encoded as a feature of nodes in relation to fault planes," but only encodes faults as a 0/1 binary feature, without considering key parameters such as fault displacement and fault type (normal fault/reverse fault/strike-slip fault). For left-lateral strike-slip reverse faults like the Nacha fault, it is unclear whether the model can accurately characterize the stratigraphic shift relationship between the two sides of the fault. The magnified view of region A in Figure 15b shows abrupt changes in stratigraphic thickness near the fault, requiring quantitative evaluation of the prediction error near the fault area.

Response: Thank you for this comment. we address the reviewer's concern concisely in the Discussion, as follows (lines 638 to 641,653 to 655):

Fourth, faults are encoded as a one-hot node feature, which marks the presence of a discontinuity but does not represent fault displacement, fault type (normal, reverse, or strike-slip), or the detailed offset across the fault.

and (4) for oblique-slip structures such as the Nacha Fault, encoding the slip vector and displacement magnitude, thereby further improving the prediction of near fault stratigraphic thickness.

Comment 6. The model's generalization ability and sample imbalance issues are not adequately discussed. Figure 13d shows extreme imbalance in the training samples. The paper claims the model has "strong small-sample generalization ability" for sparse classes, but does not provide performance tests under conditions of fewer or zero samples. Furthermore, the experiment uses a fixed 8:2 training-to-test split without cross-validation. This limitation of the modeling method, which often occurs in actual exploration where "new stratigraphic units do not appear in the training set," should be explained in the discussion.

Response: Thank you for this comment. We have added an explicit discussion of the unseen-unit (open-set) limitation. All have been corrected as summarized below (lines 631 to 633, 643 to 645):

a closed set supervised classifier can only predict lithologies present in the training set. For a stratigraphic unit absent from the training data, the current model would then assign it to the nearest known class.

(2) handling this "untrained-unit" case requires open-set recognition (flagging out-of-distribution nodes as "unknown"), semi-supervised or self-training strategies, or continual/transfer learning

Comment 7. Inconsistencies between the claim of "geological interpretability" and empirical evidence. Section 5 mentions that "model interpretability is still limited," but the abstract and conclusion emphasize that the method "reflects prior geological knowledge." It is recommended to supplement the analysis with spatial distribution of attention weights to support the core innovation claim of "stratigraphic awareness."

Response: Thank you for this comment. We have resolved the apparent inconsistency by grounding the interpretability claim directly in the attention formulation (Eqs. 3–7) rather than treating it as a black-box property. All have been corrected as summarized below (lines from 241 to 256):

For any node pair (i, j), where S_{base} is the standard query key feature similarity used by conventional attention, S_{age} is an explicit function of the geological ages of the two nodes' lithologies, S_{seq} is the learnable lithological similarity that decays with ordinal stratigraphic distance, and $S_{contact}$ encodes the conformable or unconformable contact relationship. Because these terms are added before the softmax, each attention weight admits an exact additive attribution into a learned feature-similarity component and named stratigraphic-prior components; the contribution of every prior to any aggregation is therefore transparent and auditable. The two prior terms are monotone in geological proximity. S_{age} and S_{seq} raise the weight for node pairs that are close in the stratigraphic sequence and lower it for distant pairs. After the softmax, aggregation is thus biased toward stratigraphically coherent neighbours irrespective of incidental feature resemblance. When S_{base} is ambiguous (two candidate neighbours with similar features but on opposite sides of a fault or across an unconformity), the prior terms dominate the sum and steer attention to the geologically correct neighbour, which is the mechanism by which cross fault propagation is suppressed and within unit aggregation is sharpened.

We thank the reviewer once again for the constructive and detailed comments, which have substantially improved the rigor and clarity of our manuscript. We hope that the revised manuscript satisfactorily addresses all concerns raised, and we look forward to the reviewer's further feedback.

Sincerely

Zhenxi Fang

Response to Reviewer2 comments

Dear Reviewer:

Thanks for your effort to review our manuscript titled " Three-Dimensional Geological Modeling based on Dual-Task Stratigraphy-Aware Attention Networks(Geo-SAN v1.0) ", and now we have just revised this manuscript according to the good suggestions. The details are as follows, and all the revisions are done using track changes in Word.

Reviewer2 Comments to Author:

Dear authors,

In the manuscript entitled “Three-Dimensional Geological Modeling based on Dual-Task Stratigraphy-Aware Attention Networks (Geo-SAN v1.0)”, the authors propose a three-dimensional geological modeling framework named Geo-SAN. The framework first employs a multi-scale neighborhood aggregation mechanism based on graph neural networks to identify key sampling points near fault planes and aggregate lithological features. It then introduces a stratigraphy-aware attention mechanism to explicitly incorporate stratigraphic sequence similarities into the model. Finally, through a dual-task learning strategy, the framework simultaneously performs lithological classification and scalar field interpolation. The authors use real geological data from the Lingnian–Ningping area in Guangxi Zhuang Autonomous Region, China, including stratigraphic interfaces, lithological data, and attitude points, to construct the geological model. The proposed method is compared with the AdaHRBF method and several ablation variants, namely M1 – M4, and the modeling results demonstrate its effectiveness. In addition, by integrating training data from geological maps and cross-sections, together with confusion matrices and class-wise evaluation metrics, the authors further demonstrate the reliability and generalization capability of the model in multi-lithology classification. Overall, the results sufficiently validate the effectiveness and geological plausibility of the proposed method. However, the manuscript still has the following issues:

Comment 1. The discussion of the limitations of implicit modeling methods in the Introduction lacks a sufficient summary of the relevant literature. The authors are therefore encouraged to supplement this section with a more comprehensive review of related studies.

Response: Thank you for this comment. We have expanded this discussion to provide dedicated support for each of the three limitations mentioned, as follows (lines 56 to 61):

However, existing interpolation-driven implicit methods scale poorly to large, sparse, and irregular datasets and offer no principled way to embed stratigraphic-sequence and lithological priors into the reconstruction. Motivating a learnable and data-driven implicit modeling framework that is capable of aggregating geological sampling data while coupling stratigraphic knowledge with lithology recognition poses challenges.

Comment 2. The discussion of the advantages of deep learning methods in the Introduction lacks a comparative analysis of existing studies. The authors are encouraged to further clarify the limitations of existing methods and explain the specific problems that the proposed method is designed to address.

Response: Thank you for this comment. We have revised the corresponding paragraph in Section 1 and added new references. The revised clarifies that (lines 82 to 88):

GNN-based approaches naturally accommodate the irregular and sparse nature of geological sampling data through graph structure and message passing, but existing GNN-based frameworks still typically rely on similarity-based attention or aggregation, which is not explicitly aware of stratigraphic ordering. This makes them susceptible to information ignoring across stratigraphic discontinuities and confusion between stratigraphically adjacent but lithologically distinct units.

Comment 3. The manuscript further validates the effectiveness of the proposed modules through ablation experiments. However, the current ablation study does not include separate evaluations of the SAN module and the dual-task module. As a result, the performance difference between M3 and M4 cannot be fully attributed to either SAN or dual-task learning alone. This experimental design makes it difficult to clearly distinguish the specific contribution of each component within the proposed network framework. The authors are therefore encouraged to supplement the ablation study by evaluating the SAN module and the dual-task module independently.
Response: Thank you for this comment. We have redesigned the ablation experiments and added comparison experiments with mainstream graph neural networks, as follows (lines 399 to 493):

During model training, various configurations significantly affect performance. To isolate the individual contributions of the stratigraphy-aware attention (SAN) mechanism and the dual-task learning head, four models were designed on an identical graph neighborhood aggregation backbone that alternately integrates GAT and GraphSAGE, as detailed in Table 2. Model M1 (GAT-GraphSAGE) is the baseline, using only the GAT/GraphSAGE embedding with the two prediction branches trained independently. Model M2 (GAT-GraphSAGE-DT) additionally activates the dual-task cross-task constraint, the KL-divergence and compatibility terms ($\mathcal{L}_{cross}$) that couple the scalar-field and lithology branches, while excluding SAN. Model M3 (GAT-GraphSAGE-SAN) adds the SAN mechanism to the baseline while keeping the two branches decoupled. Model M4 (GAT-GraphSAGE-SAN-DT) incorporates both SAN and the dual-task constraint, constituting the complete Geo-SAN framework. In all configurations both prediction branches are retained, R^2 and RMSE or accuracy are reported for every model; the Dual-Task toggle controls only whether the cross-task coupling $\mathcal{L}_{cross}$ is applied.

Table 2. Ablation experiments.

Model	Graph neighborhood aggregation	SAN	Dual-Task	Description
M1	GAT/GraphSAGE	×	×	GAT-GraphSAGE
M2	GAT/GraphSAGE	×	√	GAT-GraphSAGE-DT
M3	GAT/GraphSAGE	√	×	GAT-GraphSAGE-SAN
M4	GAT/GraphSAGE	√	√	GAT-GraphSAGE-SAN-DT

Figure 10 presents scalar field prediction metrics, all four models converge after roughly 300 epochs. Both components improve scalar field interpolation relative to the baseline M1 ($R^2 = 0.931$, RMSE = 0.086): enabling SAN raises R^2 to 0.961 and lowers RMSE to 0.073 (from M1 to M3), while enabling the dual-task constraint raises R^2 to 0.955 and lowers RMSE to 0.072 (from M1 to M2). Their effects are complementary, and the full model M4 attains the best result ($R^2 = 0.971$, RMSE = 0.062). The gain from SAN is slightly larger on R^2, consistent with its explicit injection of stratigraphic age and contact relationship priors, which regularizes the continuous scalar field and yields smoother, more geologically consistent transitions.

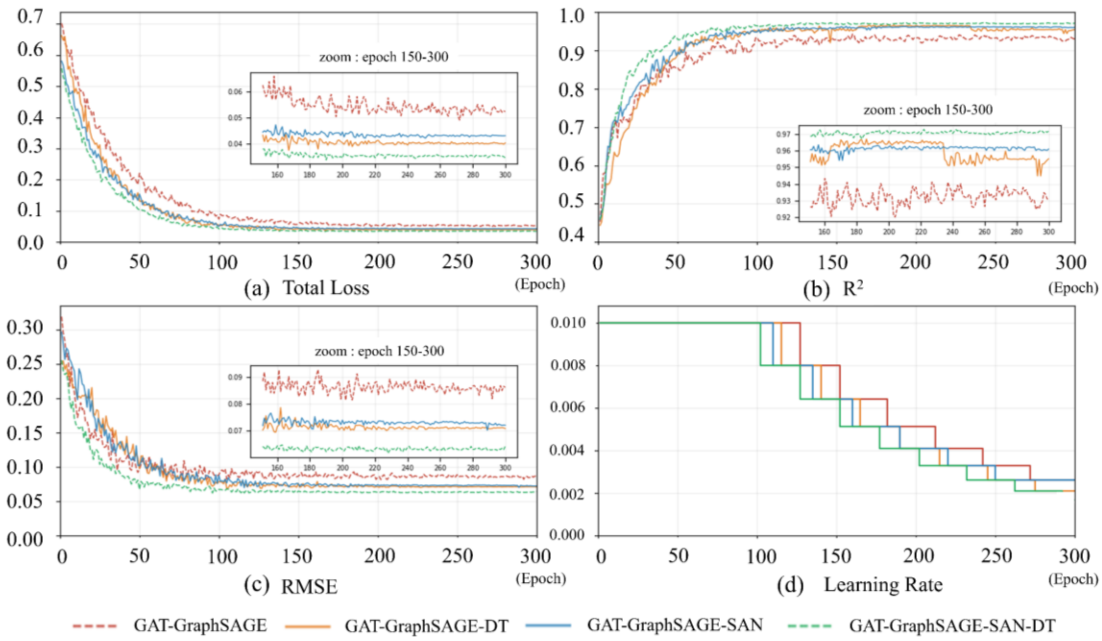

Fig. 10. Model evaluation metrics for scalar field interpolation training processes: (a) total loss, (b) R^2, (c) RMSE, and (d) learning rate.

The lithology classification training curves are shown in Figure 11. Relative to the baseline M1 (accuracy = 0.841), adding the SAN mechanism improves accuracy to 0.879 (from M1 to M3), and with the dual-task constraint, from 0.841 to 0.899 (from M1 to M2), adding the dual-task constraint and SAN mechanism improves accuracy from 0.841 to 0.921 (from M1 to M4). Both factors therefore contribute positively and roughly additively, with the cross-task constraint contributing somewhat more to lithology classification than SAN. The SAN improvement reflects the geological principle that lithologies adjacent in the stratigraphic sequence tend to be similar, explicitly incorporating lithology similarity and stratigraphic-sequence constraints. The dual-task improvement reflects the cross-task coupling that regularizes lithological predictions against the stratigraphic scalar field.

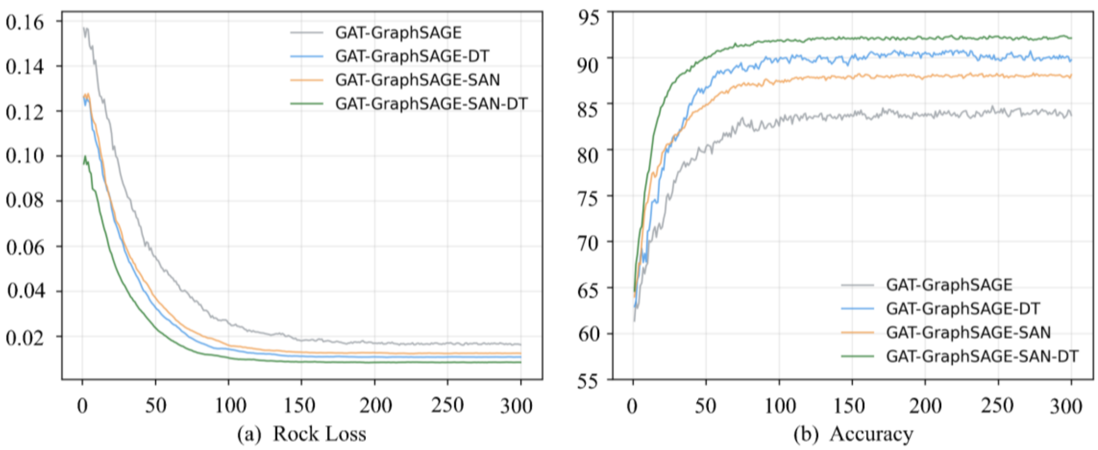

Fig. 11. Model evaluation metrics in lithology classification training processes: (a) rock loss, and (b) accuracy.

To benchmark Geo-SAN against mainstream learning-based interpolators, Table 4 evaluates three standalone GNN backbones, GCN (M5), GAT (M6), and GraphSAGE (M7), each without the

SAN prior and without dual-task coupling, on the identical graph, feature set, and sampling split.

Table 4. Comparison experiments.

Model	Graph neighborhood aggregation	SAN	Dual-Task	Description
M5	GCN	×	×	GCN
M6	GAT	×	×	GAT
M7	GraphSAGE	×	×	GraphSAGE

Table 5 summarizes the metrics of the ablation models (M1–M4) and the comparison models (M5–M7). The full model M4 performs best on both tasks (R^2 = 0.971, RMSE = 0.062, and accuracy = 0.921). Relative to the baseline M1, SAN and the dual-task constraint yield consistent gains but favor different tasks: SAN contributes more to scalar field prediction (M3 vs M2, R^2 0.961 vs 0.955), whereas the dual-task constraint contributes more to classification (M2 vs M3, accuracy 0.899 vs 0.879). The declining cross-task loss shows that the two components reinforce each other, and the 8.0% accuracy gain of M4 over M1 stems from both jointly rather than from either alone. Under the baseline setting, the single backbone models rank GraphSAGE > GAT > GCN on both tasks (accuracy of 0.828, 0.810, and 0.805; R^2 of 0.920, 0.904, and 0.879), reflecting the suitability of GraphSAGE's regional aggregation for stratified structures. The alternating GAT+GraphSAGE backbone (M1) surpasses all single backbones (R^2 = 0.931, accuracy = 0.841) at comparable cost, confirming that multi-scale aggregation extracts more representative features from sparse and irregular samples.

Table 5. Model performance evaluation metrics (300 epochs).

Model	$\mathcal{L}_{scalar}$	$\mathcal{L}_{orientation}$	$\mathcal{L}_{litho}$	$\mathcal{L}_{cross}$	R^2	RMSE	Accuracy	Time(s)
M1	0.0056	0.5157	0.3956	/	0.931	0.086	0.841	61.49
M2	0.0052	0.5203	0.4017	0.608	0.955	0.072	0.899	142.37
M3	0.0056	0.5075	0.4156	/	0.961	0.073	0.879	101.78
M4	**0.0053**	**0.4541**	**0.2719**	**0.543**	**0.971**	**0.062**	**0.921**	**162.51**
M5	0.0052	0.5238	0.4103	/	0.879	0.098	0.805	58.12
M6	0.0057	0.5567	0.4237	/	0.904	0.089	0.810	60.24
M7	0.0060	0.5689	0.4297	/	0.920	0.087	0.828	60.76

Comment 4. The Results section should not be limited to describing changes in the evaluation metrics associated with the GAT+GraphSAGE module. Instead, the authors should provide a more in-depth analysis of the specific effects of each core module from the perspective of the overall model architecture, including their contributions to feature extraction, stratigraphic prior constraints, lithological classification, and scalar field interpolation. Such an analysis would more clearly reveal how different modules contribute to the improvement of model performance.

Response: Thank you for this comment. In Section 4.1, a summary of the strengths of this study from an overall architectural perspective has been added after Table 5, as follows (lines 494 to 499):

Viewed from the overall architecture, the three modules address complementary bottlenecks: the alternating GAT+GraphSAGE backbone strengthens feature extraction from sparse, irregular samples, the SAN module injects stratigraphic constraints as additive attention terms and thus benefits scalar field interpolation most, while the cross-task loss regularizes the classifier against the predicted scalar field and thus benefits lithology classification most.

Comment 5. The Discussion section lacks sufficient analysis and comparison with existing research methods. The authors are encouraged to supplement this section with more detailed and relevant discussion.

Response: Thank you for this comment. We have added the following discussion regarding existing machine learning methods to the Discussion section, as follows (lines 595to 603):

Among NNG approaches, CNNs require resampling onto regular grids (Bi et al., 2022; Zhang et al., 2024), and MLPs discard explicit topology (Hillier et al., 2023; Chu et al., 2024), whereas GNNs retain irregular sampling topology (Hillier et al., 2021; Gao and Wellmann, 2025; Liao et al., 2026); in all of these, however, aggregation weights derive from feature similarity alone and the prediction branches are not explicitly coupled. Under an identical graph, feature set, and sampling split, Geo-SAN improves R^2 from 0.920 to 0.971 and accuracy from 0.828 to 0.921 relative to the strongest single backbone (GraphSAGE, M7), indicating that the gain originates from the stratigraphic priors and the cross-task coupling rather than from the backbone itself.

Comment 6. The manuscript should ensure terminological consistency throughout the text. For example, “lithological classification” and “lithology classification” are used interchangeably. The authors are encouraged to standardize the use of relevant technical terms throughout the manuscript to improve the clarity, consistency, and professionalism of the writing.

Response: Thank you for this comment. We conducted a full terminology audit of the manuscript, including all section headings, figure captions, and table labels. All have been corrected as summarized below:

Replace "lithological classification" / "lithology classification" in this paper with "lithology classification" uniformly.

We thank the reviewer once again for the constructive and detailed comments, which have substantially improved the rigor and clarity of our manuscript. We hope that the revised manuscript satisfactorily addresses all concerns raised, and we look forward to the reviewer's further feedback.

Sincerely

Zhenxi Fang