Supplementary file

Physics-guided self-supervised upscaling of three-dimensional digital rock models from multi-resolution micro-CT data

Batyrkhan Gainitdinov[1,*], Rinat Prochii[1], Denis Orlov[1], Maxim Sharaev[1,2], Yili Ren[3] , Dmitry Koroteev[1]

[1] *Skolkovo Institute of Science and Technology, Moscow 121205, Russia*

[2] *Biomedically Informed Artificial Intelligence Laboratory, University of Sharjah, Sharjah 27272, United Arab Emirates*

[3] *PetroChina Research Institute of Petroleum Exploration Development, Beijing 100083, P. R. China*

E-mail address: Batyrkhan.Gainitdinov@skoltech.ru (Batyrkhan Gainitdinov); Rinat.Prochii@skoltech.ru (Rinat Prochii); D.Orlov@skoltech.ru (Denis Orlov); M.Sharaev@skoltech.ru (Maxim Sharaev); renyili@petrochina.com.cn (Ren Yili); D.Koroteev@skoltech.ru (Dmitry Koroteev).

[*]Corresponding author (ORCID: 0009-0008-1501-0224 (B. Gainitdinov))

Gainitdinov, B., Prochii, R., Orlov, D., Sharaev, M., Ren, Y., Koroteev, D. Physics-guided self-supervised upscaling of three-dimensional digital rock models from multi-resolution micro-CT data. Advances in Geo-Energy Research, 2026, 21(1): 13-28.

The link to this file is: https://doi.org/10.46690/ager.2026.07.04

Appendix A: Data augmentation strategy for supervised training

Digital rock classifier finetune is performed under severe class imbalance and limited labelled data, particularly for rare high-permeability rock types. To improve generalisation and mitigate overfitting, we apply a set of carefully designed intensity-preserving augmentations to the low-resolution μCT minicubes used for supervised learning. The primary goals of the augmentation pipeline are:

• Increasing the effective sample size for minority classes without introducing duplicate samples,

• Preserving key statistical properties of the grayscale images (mean and standard deviation distributions),

• Preserving upscaled petrophysical properties such as porosity and absolute permeability on the augmented fragments.

Augmentations are applied using the Albumentations library. For each 3D minicube we apply a stochastic composition of rigid and photometric transforms:

• random 90° rotations in the axial plane;

• one of horizontal flip, vertical flip, or transpose;

• one of Gaussian noise or random brightness/contrast adjustment, applied jointly with.

The chosen ranges for noise variance and brightness/contrast are deliberately conservative, so that local grayscale statistics and pore–solid interfaces remain close to those observed in the original μCT data. Rotations and flips are restricted to the axial plane in order to avoid unphysical permutations of the geological z-axis and to preserve vertical anisotropy. Augmentation is applied in a class-aware manner to reduce imbalance between digital rock types.

For each class k, the original set of labelled minicubes is expanded by generating a fixed number of augmented variants per sample, resulting in additional factors {0:7, 1:1, 2:1, 3:1} for classes 0 – 3, respectively. Transformation parameters are sampled independently for each realisation, and duplicates in the augmented dataset are explicitly filtered out to avoid oversampling identical fragments.

To verify that the augmentation pipeline does not distort the physical content of the data,

we define an invariance check that is applied to a held-out subset of labelled minicubes. For each original 3D binary model and its augmented counterpart, we compute (i) porosity and (ii) directional single-phase permeabilities along the three axes using the procedures described in Appendix B. The distributions of mean intensity, standard deviation, porosity and axis-averaged permeability are then compared between original and augmented samples. Quantitative results of this analysis, demonstrating that the augmentations preserve petrophysical properties within acceptable tolerances while effectively enriching the training set, are reported in Appendix B.

Appendix B: Effect of data augmentation on petrophysical properties

To assess whether the chosen augmentations alter the physical content of the data, we first quantified their impact on simple grayscale statistics. For each rock type, we selected a representative low-resolution μCT minicube and generated 38 augmented variants using the pipeline described in Appendix A. Table 1 summarises the mean and standard deviation of voxel intensities for original and augmented samples. Across all five classes, the relative change in mean intensity does not exceed 1.2%, while the change in standard deviation remains below 0.5%, indicating that the global grayscale distributions are effectively preserved.

Table 1. Mean and standard deviation (Std) of grayscale intensities for original and augmented minicubes per class. Δ denotes the relative change.

Class	Mean (orig)	Mean (aug)	Std (orig)	Std (aug)
0	75.24	74.67 ± 4.15 (−0.76%)	50.42	50.23 ± 1.31 (−0.38%)
1	56.34	56.90 ± 5.46 (+0.98%)	68.74	69.07 ± 1.48 (+0.48%)
2	92.82	91.75 ± 5.11 (−1.16%)	91.93	91.48 ± 1.93 (−0.48%)
3	192.91	192.64 ± 5.22 (−0.14%)	65.59	65.52 ± 1.40 (−0.10%)
4	226.35	226.52 ± 3.06 (+0.08%)	43.81	43.64 ± 1.05 (−0.38%)

Fig. S1 illustrates a typical example of an original and an augmented CT slice, together with their binarised counterparts. Visual inspection confirms that augmentations primarily modify local texture and noise while preserving the morphology and connectivity of the pore space.

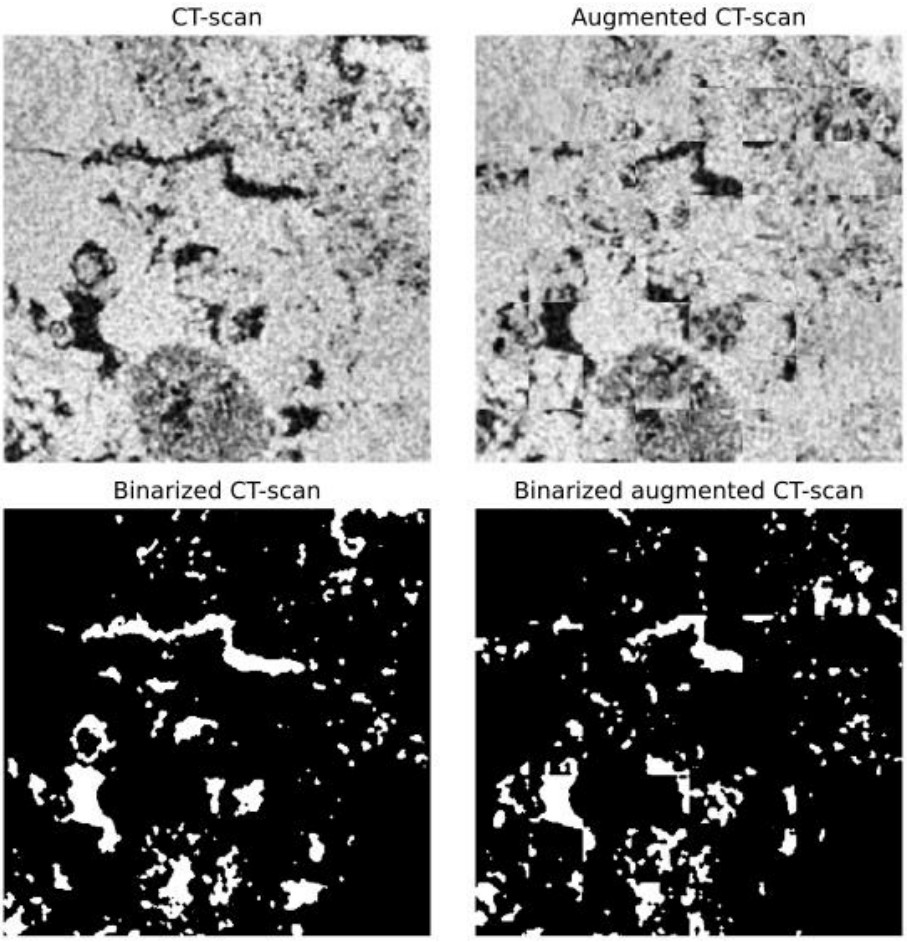

Fig. S1. Original CT-scan and binarized model vs their counterparts assembled from augmented crops.

We further verified that augmentations do not distort upscaled transport properties by comparing single-phase permeabilities computed on original and augmented binary fragments. For each minicube, permeability was estimated along the three principal axes using the procedure in Appendix B, and the axis-averaged values were compared pairwise. As shown in Fig. S2, the points lie close to the identity line, with deviations remaining small over more than two orders of magnitude in permeability. Agreement between original and augmented permeability was quantified (N = 500 fragment pairs). Pearson and Spearman correlations were 0.99 and 0.96 (95% bootstrap CIs [0.97,1.00] and [0.92, 0.99]); for fragments with Korig > 0.1 mD, 82% lay within ±5% margin (median |rel| = 5.1%). Thus augmentation preserves permeability rankings and absolute-scale agreement, with bounded deviations on the mD scale.

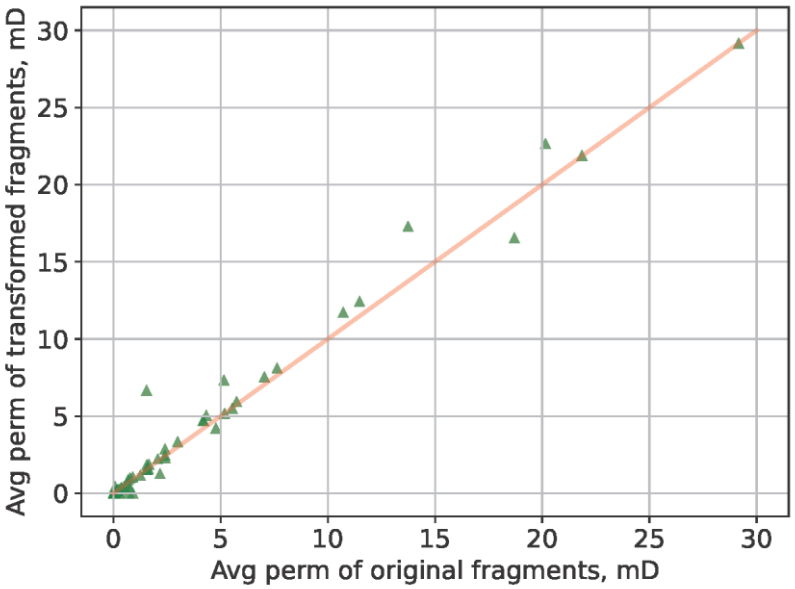

Fig. S2. Comparison of single-phase permeabilities computed on original and augmented binary fragments.

This indicates that the augmentation pipeline preserves both porosity and effective permeability within acceptable tolerances for all rock types.

Finally, class-aware augmentation substantially improves the balance between rock types in the supervised training set. The number of labelled samples in classes 0 – 3 is increased by factors {0:7, 1:1, 2:1, 3:1}, respectively, while explicit duplicate checking ensures that no identical augmented fragments are introduced. This enrichment of minority classes is essential for stabilising the downstream classification task both for supervised and SSL approaches.

Appendix C: Ablation of permeability-guided regularization

In the combined SSL objective,

$$L = L_{pred} + \lambda L_{perm} \quad (1)$$

λ controls the contribution of the permeability-aware regularisation relative to the contextual prediction loss. In the reported implementation, λ was fixed to 0.3 and kept unchanged for all experiments presented in this work. The temperature parameter in the exponential kernel of Eq. (6) in article was likewise fixed to $\tau = 0.3$.

To clarify the contribution of the permeability-guided pretraining term, an ablation study

was performed by varying the λ in the pretraining objective. In particular, we compared three settings: 1) λ = 0, which is equivalent to removing the proxy-permeability regularization entirely, 2) the default setting λ = 0.3 used in the main paper, and 3) a stronger regularization setting λ = 0.6.

The results are summarized in Table 2. Removing the proxy-permeability term (λ = 0) degraded the downstream classification performance compared to the default setting, indicating that the permeability-guided regularization contributes positively to the learned representation. Increasing the regularization weight from λ = 0.3 to λ = 0.6 did not improve downstream classification metrics. Instead, the best balance between pretraining and downstream performance was achieved at λ = 0.3, which yielded the highest macro F1-score and accuracy among the tested settings. This suggests that a moderate contribution of the proxy-permeability term is beneficial, whereas too weak or too strong weighting is suboptimal.

Table 2. Ablation study for the proxy-permeability loss weight

Pretraining setting	λ	Accuracy	F1-macro	F1-micro
Without proxy-permeability loss	0.0	0.7672	0.7702	0.7695
Default setting	0.3	0.7810	0.7840	0.7810
Stronger proxy regularization	0.6	0.7790	0.7780	0.7800

When λ is too small, the permeability-aware term becomes too weak and the latent space is dominated by purely contextual or appearance-based similarity. When λ is too large, the optimization overemphasizes the proxy-permeability constraint and degrades the contextual discrimination learned by the original VoCo objective. Fixing λ and τ across all reported experiments ensured that the comparison between the supervised and SSL-based pipelines was not affected by run-specific retuning of the loss.

Appendix D: Pseudocode of the pretraining and fine-tuning procedures

For clarity, the overall two-stage learning procedure is summarised below in algorithmic form.

Algorithm A1: Physics-guided VoCo pretraining.

1. Initialise the student and teacher encoders with identical weights.

2. Sample a batch of unlabeled 3D sub-volumes and extract (i) a fixed set of base crops and (ii) additional random/query crops.

3. Forward all crops through the student encoder; forward the same crops through the teacher encoder without gradient updates.

4. Compute the contextual prediction loss by comparing student–teacher similarities with soft overlap-based pseudo-labels for the query/basecrop pairs.

5. For the subset of base crops with available proxy permeability values, compute the embedding similarity matrix S and the hydrodynamic similarity matrix G.

6. Evaluate the permeability-aware regularisation term

$$L_{perm} = S + G_F^2$$

and combine it with the contextual loss:

$$L = L_{pred} + \lambda L_{perm}$$

7. Update the student encoder by backpropagation and update the teacher encoder by exponential moving average.

8. Repeat until the prescribed number of optimisation steps is reached.

Algorithm A2: Fine-tuning for rock typing.

1. Load the pretrained student weights into the downstream Swin-based classifier.

2. Replace the SSL projection head with a dropout-plus-linear classification head.

3. Optionally freeze the backbone for the first few epochs and train only the classification head.

4. Unfreeze the full network and continue supervised training on labelled minicubes using cross-entropy loss.

5. Use separate learning rates for the backbone and the classification head, validate after

each epoch, and retain the checkpoint with the best validation macro-F1.

6. Use the resulting classifier to assign digital rock types to low-resolution minicubes and assemble the Darcy-scale model.

Appendix E: Practical note on computational overhead

The additional physics-guided computations are required only during SSL pretraining. In particular, proxy permeability, percolation statistics, and the similarity matrices entering L_{perm} are computed only in the representation - learning stage. At inference time, the trained classifier operates as a standard patch-wise forward model and does not require online evaluation of permeability proxies or percolation analysis. Thus, the main computational overhead of the proposed approach is concentrated in the one-time offline pretraining stage rather than in the Darcy-scale inference workflow.

Appendix F: Training setup and hyperparameters

For reproducibility, the exact optimisation settings used for the three training stages analysed in this work are summarised in (Table 5): 1) Self-supervised VoCo pre-training of the Swin-based encoder, 2) supervised fine-tuning of the pre-trained encoder for five-class rock typing, and 3) the purely supervised DenseNet-121 baseline. The labelled supervised dataset is the one described in Section 2.3.

Several implementation details are important for interpreting the comparison. First, the self-supervised pipeline is genuinely two-stage: The Swin-based encoder is first trained on unlabeled carbonate volumes with the combined contextual and permeability-guided VoCo objective (Section 3.1.2), after which the resulting encoder is fine-tuned for the downstream five-class rock-typing task. Second, the supervised DenseNet-121 baseline and the SSL fine-tuning stage used the same labelled data protocol and the same train/validation evaluation cadence, which makes the comparison in Section 3.2.1 attributable primarily to the benefit of self-supervised pre-training rather than to differences in downstream optimisation.

Finally, the same supervised fine-tuning protocol summarised above was reused in the label-scarcity experiments of Section 3.2.2, where only the amount of labelled training data was reduced to 50%, 25%, and 10% of the original set, while the validation/test protocol and the

optimisation settings were kept fixed.

Table 3. Training parameters for the three models used in this study

Model	Optimisation setup
VoCo pre-training (Swin-based VoCoHead)	AdamW with AMSGrad, learning rate 10^{-4}, batch size 2. Training length: 250,000 optimisation steps. Learning-rate schedule: **WarmupCosineSchedule** with 5,000 warmup steps and total length 250,000 steps. Mixed precision enabled.
SSL fine-tuning (Swin classifier initialised from VoCo student encoder)	AdamW with two parameter groups: Learning rate 3×10^{-4} for the classification head and 3×10^{-5} for the backbone; weight decay 10^{-5}. Batch size 16, 60 epochs, validation every epoch. Learning-rate schedule: **WarmupCosineSchedule** with 5 warmup epochs. Loss: Cross-entropy with label smoothing 0.1. Gradient clipping threshold 1.0. Mixed precision enabled.
Supervised baseline (3D DenseNet-121)	AdamW, learning rate 3×10^{-4}, weight decay 10^{-4}. Batch size 16, 50 epochs, validation every epoch. Learning-rate schedule: **WarmupCosineSchedule** with zero warmup epochs (cosine decay from the start). Loss: Standard cross-entropy. Gradient clipping threshold 1.0. Mixed precision was inactive in the reproduced run.