MIXSIM3D : Une Nouvelle Méthode d'Apprentissage Curriculum Contrastif 3D Appliquée à la Physique des Roches

Van Thao NGUYEN[1,2] Dominique FOURER[2] Jean-François LECOMTE[1] Souhail YOUSSEF[1] Désiré SIDIBÉ[2]

[1]IFP Énergies nouvelles, Ruel Malmaison, France

[2]Laboratoire IBISC (EA 4526), Université d'Evry Paris-Saclay, Évry-Courcouronnes, France

Résumé – Dans cet article, nous introduisons une nouvelle méthode appelée MixSim3d, une approche d'apprentissage profond auto-supervisée conçue pour apprendre des représentations latentes pour des tâches de régression, à partir d'images 3D. Elle combine l'apprentissage curriculum (CL) et l'apprentissage contrastif pour améliorer la robustesse de la représentation latente apprise sur les images 3D en entrée. Ici, nous appliquons cette méthode dans le cadre de la Physique des Roches Numérique (PRN) pour prédire des propriétés telles que la porosité et la perméabilité à partir des ensembles de données 3D observées. Notre évaluation montre que la méthode MixSim3d proposée peut obtenir des résultats prometteurs, surpassant certaines approches auto-supervisées de référence existantes dans des scénarios particuliers.

Abstract – In this paper, we introduce a new method named MixSim3d, a deep learning self-supervised approach designed to learn meaningful representations and extract relevant parameters from 3D images. It combines curriculum learning and contrastive learning (CL) to improve the robustness of the learned embedded representation of input 3D images. Here, we apply this method in the context of Digital Rocks Physics (DRP) to predict properties, such as porosity and permeability, from the observed 3D datasets. Our evaluation shows that the proposed Mixsim3d method can obtain promising results, outperforming existing baseline self-supervised approaches in particular scenarios.

1 Introduction

L'analyse d'images 3D est omniprésente dans divers domaines tels que l'imagerie médicale, les géosciences et la science des matériaux. En particulier, des tâches comme la segmentation, la classification et la prédiction de paramètres à partir de données 3D nécessitent une représentation latente pertinente qui capture les caractéristiques discriminantes de l'entrée. Extraire des représentations efficaces à partir de données 3D reste un défi dans les scénarios où les ensembles de données annotées sont rares ou inexistants. L'apprentissage auto-supervisé ou *Self-Supervised Learning* (SSL) [8] a émergé comme une approche prometteuse pour surmonter cette limitation, en exploitant des données non annotées afin d'apprendre des caractéristiques utiles à travers des tâches prétextes.

Plusieurs techniques SSL ont été proposées ces dernières années, posant les bases de l'apprentissage de représentations robustes à partir d'images. Contrairement aux approches supervisées qui reposent fortement sur de grands ensembles de données annotées, SSL exploite les données non annotées pour générer des pseudo-étiquettes via des tâches prétextes innovantes. Ces méthodes ont permis des avancées remarquables, établissant de nouveaux standards en matière de performance dans diverses tâches de vision [2, 4, 7].

Les premiers travaux comme SimCLR [2] et MoCo [10] ont démontré l'efficacité de l'apprentissage contrastif, qui rapproche les paires positives (par exemple, des vues augmentées d'une même image) dans l'espace d'encodage tout en éloignant les paires négatives. Des extensions ont été proposées pour adapter ces approches aux données 3D afin de mieux gérer leur nature volumétrique, mais elles causent une augmentation de la complexité computationnelle [12]. Malgré ces succès dans l'imagerie naturelle, l'SSL rencontre des défis spécifiques en Physique des Roches Numérique (PNR), où l'analyse d'images 3D en μCT nécessite des tâches prétextes adaptées afin de capturer les géométries complexes et irrégulières des structures rocheuses. De plus, les données en PNR sont exigeantes en termes de calcul, et la transférabilité des représentations apprises vers des tâches comme la prédiction de perméabilité reste peu explorée.

Dans le contexte de l'analyse d'images 3D, le Curriculum Learning (CL) [1] montre également des résultats prometteurs. En introduisant les tâches de manière progressive, le CL permet aux modèles d'apprendre d'abord des représentations simples avant d'aborder des motifs plus complexes. Lorsqu'il est combiné à l'apprentissage contrastif, il ouvre la voie à de nouvelles représentations robustes et interprétables. Cependant, les travaux existants exploitent rarement pleinement la synergie entre ces deux paradigmes, en particulier pour les applications aux données 3D. Des recherches récentes visent à pallier ces limites en développant des stratégies adaptatives et automatisées pour le CL, offrant ainsi un potentiel d'amélioration de l'efficacité et de la robustesse de l'apprentissage dans diverses tâches d'apprentissage automatique.

Dans cet article, nous introduisons *MixSim3d*, une nouvelle méthode d'apprentissage auto-supervisé conçue pour l'apprentissage de représentations d'images 3D, combinant l'apprentissage par curriculum et l'apprentissage contrastif afin de répondre aux défis mentionnés précédemment. En augmentant progressivement la complexité des tâches prétextes et en utilisant une fonction de coût contrastive, MixSim3d apprend des représentations robustes à partir de données 3D. MixSim3d est appliqué à la Physique des Roches Numérique (PNR), où la prédiction précise des propriétés matérielles, telles que la porosité et la perméabilité, est cruciale.

Cet article est structuré comme suit. La section 2 présente

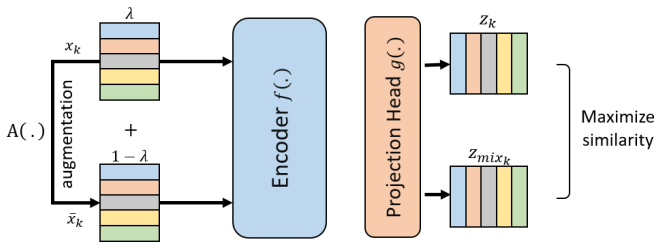


FIGURE 1 : MixSim3D.

la méthode MixSim3d, avec son architecture neuronale et la stratégie d'entraînement utilisée. Les résultats numériques sont exposés et discutés dans la section 3. Enfin, la section 4 conclut l'article avec des perspectives de travaux futurs.

2 Méthode Proposée

Algorithme 1 : MixSim3D Pseudo-Code

Input : $f(\cdot)$: Encodeur
$g(\cdot)$: Fonction de Projection
τ : Température
T : Nombre d'époques
`loader` : Calcule les mini-lots à partir des échantillons
Output : Réseaux entrainés $f(\cdot)$ and $g(\cdot)$
for $x_k \in$ `loader` **do**
`// Itère sur les mini-lots`
$\tilde{x}_k \leftarrow A(x_k)$ `// Échantillon Augmenté`
$h_k \leftarrow f(x_k)$ `// Représentation`
$z_k \leftarrow g(h_k)$ `// Projection`
$\lambda \leftarrow \frac{1}{2}\left(1 - \cos\left(\frac{\pi \cdot t}{T}\right)\right)$ à l'époque $t \in [1, T]$
$x_{\text{mix}_k} \leftarrow (1-\lambda)x_k + \lambda\tilde{x}_k$
$h_{\text{mix}_k} \leftarrow f(x_{\text{mix}_k})$ `// Représentation`
$z_{\text{mix}_k} \leftarrow g(h_{\text{mix}_k})$ `// Projection`
Calcul de la similarité :

$$L_{\text{sim}} \leftarrow \frac{-1}{N}\sum_{k=1}^{N}\log\left(\frac{\exp\left(\frac{\text{sim}(z_k, z_{\text{mix}_k})}{\tau}\right)}{\sum_{\substack{m=1 \\ m\neq k}}^{N}\exp\left(\frac{\text{sim}(z_k, z_{\text{mix}_m})}{\tau}\right)}\right)$$

Actualise $f(\cdot)$ et $g(\cdot)$ pour minimiser L_{sim}
end
return réseaux entrainés $f(\cdot)$ et $g(\cdot)$

Nous proposons de calculer des représentations discriminantes pour des entrées 3D volumétriques, de manière à exploiter pleinement leur contexte spatial. La figure 1 illustre l'architecture de notre méthode formalisée dans l'algorithme 1.

MixSim3D repose sur trois composants fondamentaux :

1. **Module d'augmentation de données stochastique** : Ce module, noté $A(\cdot)$, applique des transformations à un échantillon d'entrée x pour produire un échantillon augmenté aléatoire $\tilde{x}$. Chaque volume est transformé à l'aide d'une séquence d'augmentations successives. Plus précisément, quatre transformations sont appliquées :

- **Flou gaussien** : Simule le lissage de l'image et réduit le bruit.
- **Bruit gaussien** : Introduit des variations en imitant le bruit naturel.
- **SobelFilter3D** : Accentue les contours et les détails structuraux du volume 3D.
- **Cutout** : Masque aléatoirement certaines parties du volume, obligeant le modèle à se concentrer sur le contexte global.

2. **Encodeur basé sur un réseau de neurones** : Une fonction $f(\cdot)$ mappe une entrée 3D x vers un vecteur de représentation $h = f(x) \in \mathbb{R}^{d_e}$ dans un espace latent de dimension d_e. Ici, nous utilisons une version 3D de ResNet18 [6].

3. **Réseau de projection** : Une fonction $g(\cdot)$ projette le vecteur de représentation h vers un espace de dimension inférieure $z = g(h) \in \mathbb{R}^{d_p}$ pour le calcul de la perte contrastive. Ce vecteur z est normalisé sur l'hypersphère unité, permettant ainsi d'utiliser le produit scalaire comme mesure de similarité dans l'espace de projection.

Mixup : Le concept de *mixup* a été initialement introduit comme une technique d'augmentation visant à améliorer la généralisation des modèles de classification d'images [3]. Cette approche entraîne un réseau neuronal profond sur des combinaisons convexes de paires d'exemples et de leurs étiquettes. Plus précisément, *mixup* construit des exemples d'entraînement virtuels $x_{\text{mix}}, y_{\text{mix}}$ selon la formule suivante :

$$x_{\text{mix}} = \lambda x_i + (1-\lambda)x_j, \tag{1}$$
$$y_{\text{mix}} = \lambda y_i + (1-\lambda)y_j, \tag{2}$$

où x_i et x_j sont deux échantillons distincts, et y_i et y_j leurs étiquettes correspondantes. Le coefficient de mélange $\lambda \in [0, 1]$ est généralement échantillonné aléatoirement à partir d'une distribution Beta.

Stratégie *Curriculum Mixup* : Pour résoudre le problème de discrimination des instances en apprentissage contrastif, notre approche s'inspire des principes du *mixup* et de l'apprentissage contrastif. Contrairement au *mixup* traditionnel, qui combine deux échantillons indépendants, MixSim3D génère des échantillons mixtes x_{mix} en fusionnant l'échantillon original x_{ref} avec sa version augmentée x_{aug}, selon la formule :

$$x_{\text{mix}} = (1-\lambda)x_{\text{ref}} + \lambda x_{\text{aug}}. \tag{3}$$

Cette fusion progressive est contrôlée par une fonction de planification cosinus qui met à jour λ à chaque époque, augmentant progressivement la contribution de l'échantillon augmenté. En débutant avec l'échantillon original puis en intégrant progressivement sa version transformée, MixSim3D s'aligne sur le paradigme de l'apprentissage *curriculum*, permettant au modèle d'apprendre d'abord des motifs simples avant de s'adapter à des variations plus complexes.

3 Résultats Expérimentaux

3.1 Jeu de Données

Bien que les étiquettes pour les images naturelles puissent être facilement obtenues par traitement des jeux de données,

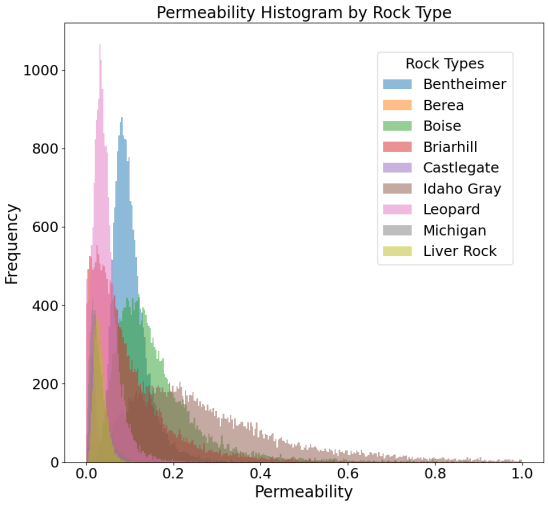


FIGURE 2 : Histogramme de perméabilité pour les 9 types de roche considérés.

cette approche est limitée dans le cas de la PNR en raison des lourdes exigences en termes de temps de calcul et de stockage. Les milieux poreux désignent des matériaux possédant une matrice solide et des espaces vides interconnectés permettant le passage des fluides. Comprendre, caractériser et mesurer les propriétés influençant l'écoulement des fluides à travers ces matériaux est essentiel dans divers domaines, notamment les applications géologiques et d'ingénierie telles que la gestion des eaux souterraines, le stockage du CO2 et la gestion de l'énergie souterraine.

Dans cette étude, neuf types de roches distincts sont analysés contenant un total de 50 images 3D μCT. Chaque volume a une dimension de $1100 \times 1100 \times 2800$ voxels et est encodé en 16 bits. Nous calculons les cartes de chaleur de perméabilité de référence de taille $1000 \times 1000 \times 2700$ en utilisant la méthode de Boltzmann sur réseau *Lattice Boltzmann Method* (LBM) [13]. Pour chaque type de roche, 20 000 sous-échantillons de dimensions $100 \times 100 \times 100$ sont extraits aléatoirement. Chaque sous-échantillon est associé à une valeur de perméabilité correspondante dérivée de la carte de chaleur. Afin d'assurer la stabilité de l'apprentissage, toutes les valeurs sont normalisées en les divisant par la valeur maximale de perméabilité. La distribution de perméabilité des neuf types de roches est illustrée dans la Fig. 2.

3.2 Détails d'implémentation

La méthode proposée[1] est entraînée via l'algorithme d'optimisation SGD, avec des mini-batch de taille 128. Le taux d'apprentissage est fixé à $lr = 0.0001$ qui est réduit de moitié toutes les 10 époches avec un momentum de 0.9.

L'entraînement se fait sur 30 époques sur un serveur de calcul a 8 noeuds dont chacun dispose de 8 GPUs A100 de 32GB avec un paramètre de température $\tau = 0.1$.

La parallélisation des données a été utilisée pour répartir la charge de calcul entre les noeuds. Ainsi, l'entraînement requiert environ 3 heures par époque. Plusieurs techniques d'augmentation de données ont été utilisées avec une probabilité de 0.5 (flou gaussien, bruit blanc gaussien, filtrage de Sobel 3D et troncatures, afin d'améliorer la robustesse du modèle. Après l'entraînement, les paramètres résultants du réseau sont utilisés pour initialiser le modèle utilisé pour les tâches principales.

Ainsi, nous entraînons d'abord le modèle en utilisant une approche SSL, sur un jeu de données composé de 20 000 images 3D μCT non labélisées, de taille $100 \times 100 \times 100$, extraites à partir de 9 types distincts de roche. Ensuite, nous optimisons le modèle de manière supervisée avec 6 types de roches : "Bentheimer," "Boise," "Castlegate," "Idaho Gray," "Leopard," et "Liver Rock", pour l'entraînement et la validation avec une répartition (80% pour l'entraînement et 20% pour la validation). Les 3 roches restantes ("Berea," "Briahill," et "Michigan") sont utilisées comme base de test pour valider la capacité de généralisation du modèle.

3.3 Résultats pour la classification du type de roche

Pour réaliser la tâche de classification, seul l'encodeur pré-entraîné $f(\cdot)$ est utilisé sans activation non-linéaire. La sortie de l'encodeur est récupéré dans une couche linéaire qui est entraînée de manière supervisée pour la tâche souhaitée. Notre méthode est comparée à plusieurs techniques state-of-the-art (SOTA) qui sont respectivement Resnet18 [5], CCT-2 [9], SimCLR [2], BYOL [7] et MoCo v2 [10] qui ont été adaptées au traitement de données 3D [11].

Nous considérons 2 expériences considérant respectivement 1% et 10% des données labélisées pour l'entraînement. Les Tables 1 et 2 présentent nos résultats comparatifs reposant sur les métriques usuelles : F1 Score, Rappel, Précision, et Top-1 Exactitude. Ainsi, notre modèle MixSim3D fournit des résultats compétitifs sur tous les jeux de données. Pour le petit jeu de données, MixSim3D atteint de bonnes performances équilibrées sur toutes les métriques. MixSim3D fournit de meilleurs résultats que tous les autres modèles sauf MoCo-v2 sur le plus petit jeu de données. Les métriques obtenues montrent que le modèle est capable de généraliser avec un nombre limité d'échantillons. Sur un jeu de données de plus grande taille, MixSim3D atteint parmi les meilleurs résultats. Dans tous les cas, MixSim3D est prométteur et suggère que les données latentes ne sont pas limitées par la faible taille du jeu de données d'entraînement.

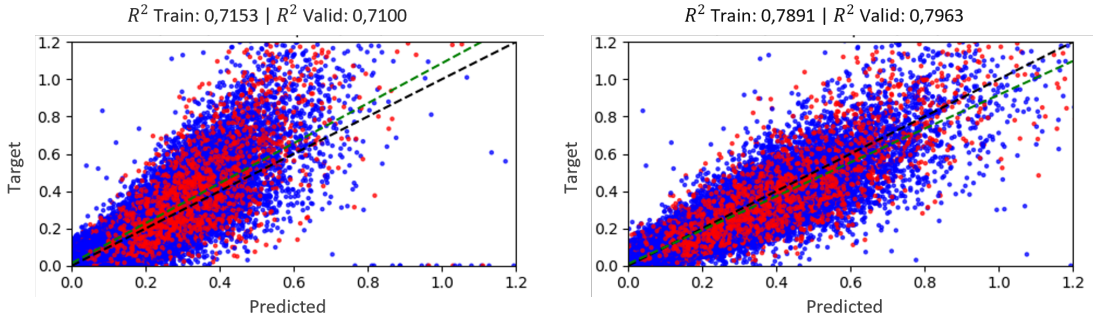
TABLE 1 : Résultats utilisant 1% de la base d'entraînement.

Modèle	F1 Score	Rappel	Précision	Top1 Exact.
ResNet18[5]	73.75	71.98	75.61	71.85
CCT[9]	75.40	72.96	78.02	72.92
SimCLR [2]	74.58	72.91	76.33	72.85
MoCo-v2 [10]	**81.18**	**81.04**	**81.32**	**81.04**
Byol [7]	80.30	80.19	80.46	80.23
SimSiam [4]	75.45	75.01	75.64	75.56
MixSim3D	<u>80.65</u>	<u>80.64</u>	<u>80.65</u>	<u>80.62</u>

3.4 Résultats d'estimation de la perméabilité des roches.

Nous évaluons notre méthode MixSim3d sur une tâche de régression pour prédire la perméabilité à partir de l'observation d'une image 3D. La qualité de l'estimation est mesurée par

[1] Codes PyTorch 2.1 disponibles sur `https://github.com/nguyenva04/mixsim3d_gretsi`



(a) 3D ResNet18 (b) MixSim3D

FIGURE 3 : Graphiques de dispersion des résultats après ajustement des modèles 3D ResNet18 (a) et MixSim3D (b). Les points bleus correspondent aux échantillons du jeu d'entraînement, et les points rouges représentent les échantillons du jeu de validation. La ligne verte est la ligne de régression estimée et la ligne noire est l'identité idéale $(y = x)$.

TABLE 2 : Résultats utilisant 10% de la base d'entraînement.

Modèle	**F1 Score**	**Rappel**	**Précision**	**Top1 Exact.**
ResNet18[5]	93.80	93.72	93.89	93.74
CCT[9]	93.62	93.78	93.46	93.01
SimCLR [2]	94.25	94.23	94.27	94.26
MoCo-v2 [10]	95.10	95.03	95.18	95.05
Byol [7]	**95.47**	**95.46**	**95.48**	**95.47**
SimSiam [4]	93.91	93.85	93.98	93.87
MixSim3D	95.26	95.24	95.29	95.33

le coefficient de détermination R^2 et la norme L_2 de l'erreur. Nous observons dans la Table 3 que MixSim3D fournit de meilleurs résultats que les autres méthodes avec un coefficient R^2 supérieur et une erreur L2 plus faible, aussi bien pour le jeu d'entraînement que pour l'ensemble de validation. La Figure. 3 montre cette amélioration de l'amélioration avec un graphique de dispersion comparant ResNet18-3D et MixSim3D pour la prédiction de la perméabilité.

TABLE 3 : Résultats de prédiction de la perméabilité.

Modèle	**R^2 (Entraîn.)**	**R^2 (Valid.)**	**L_2 (Entraîn.)**	**L_2 (Valid.)**
ResNet18 [5]	0.7153	0.7100	7.4465	7.8241
SimCLR [2]	0.7456	0.7478	7.0958	6.5604
SimSiam [4]	0.7387	0.7465	6.8354	6.6257
BYOL [7]	0.7795	0.7860	6.3073	5.5659
MoCo-v2 [10]	0.7648	0.7714	6.4059	6.3526
MixSim3D	**0.7819**	**0.7963**	**5.8450**	**5.3613**

4 Conclusion

Nous avons introduit MixSim3D, une nouvelle technique d'apprentissage auto-supervisé (SSL) appliquée aux images 3D. Notre approche combine l'apprentissage contrastif et l'apprentissage curriculum afin d'améliorer la transition entre les données originales et leurs versions augmentées. Cela favorise l'émergence de représentations plus robustes et, surtout, plus pertinentes. Par ailleurs, MixSim3D est particulièrement adapté aux données 3D, où les structures volumétriques offrent un contexte riche pour l'apprentissage. Nos expériences menées sur un jeu de données de roches, ainsi que la comparaison avec plusieurs méthodes de l'état de l'art, montrent un avantage pour notre approche en termes de robustesse et de qualité des représentations apprises. Pour les tâches principales de classification et de régression, MixSim3D fournit également des résultats prometteurs en comparaison avec l'état de l'art. Nous espérons que cette approche inspirera de nouvelles techniques pour l'apprentissage auto-supervisé.

Références

[1] Yoshua BENGIO, Jérôme LOURADOUR, Ronan COLLOBERT et Jason WESTON : Curriculum learning. *In Proc. 26th annual international conference on machine learning (ICML)*, pages 41–48, 2009.

[2] Ting CHEN, Simon KORNBLITH, Mohammad NOROUZI et Geoffrey HINTON : A simple framework for contrastive learning of visual representations. *In Proc. International conference on machine learning (ICML)*, pages 1597–1607. PMLR, 2020.

[3] Xinlei CHEN, Haoqi FAN, Ross GIRSHICK et Kaiming HE : Improved baselines with momentum contrastive learning. *arXiv preprint arXiv :2003.04297*, 2020.

[4] Xinlei CHEN et Kaiming HE : Exploring simple siamese representation learning. *In Proc. IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758, 2021.

[5] Jia DENG, Wei DONG, Richard SOCHER, Li-Jia LI, Kai LI et Li FEI-FEI : Imagenet : A large-scale hierarchical image database. *In Proc. IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009.

[6] Christoph FEICHTENHOFER, Haoqi FAN, Jitendra MALIK et Kaiming HE : Slowfast networks for video recognition. *In Proc. IEEE/CVF international conference on computer vision*, pages 6202–6211, 2019.

[7] Jean-Bastien GRILL, Florian STRUB, Florent ALTCHÉ, Corentin TALLEC, Pierre RICHEMOND, Elena BUCHATSKAYA, Carl DOERSCH, Bernardo AVILA PIRES, Zhaohan GUO, Mohammad GHESHLAGHI AZAR *et al.* : Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[8] Jie GUI, Tuo CHEN, Jing ZHANG, Qiong CAO, Zhenan SUN, Hao LUO et Dacheng TAO : A survey on self-supervised learning : Algorithms, applications, and future trends. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2024.

[9] Ali HASSANI, Steven WALTON, Nikhil SHAH, Abulikemu ABUDUWEILI, Jiachen LI et Humphrey SHI : Escaping the big data paradigm with compact transformers. *arXiv preprint arXiv :2104.05704*, 2021.

[10] Kaiming HE, Haoqi FAN, Yuxin WU, Saining XIE et Ross GIRSHICK : Momentum contrast for unsupervised visual representation learning. *In Proc. IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[11] Van Thao NGUYEN, Dominique FOURER, Desiré SIDIBÉ, Jean-François LECOMTE et Souhail YOUSSEF : A comparative evaluation of self-supervised methods applied to rock images classification. *In Proc. 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2023)*, 2024.

[12] Adrian SPURR, Aneesh DAHIYA, Xi WANG, Xucong ZHANG et Otmar HILLIGES : Self-supervised 3d hand pose estimation from monocular rgb via contrastive learning. *In Proc. IEEE/CVF international conference on computer vision*, pages 11230–11239, 2021.

[13] Laurent TALON, Daniela BAUER, Nicolas GLAND, Souhail YOUSSEF, Harold AURADOU et Irina GINZBURG : Assessment of the two relaxation time lattice-boltzmann scheme to simulate stokes flow in porous media. *Water Resources Research*, 48(4), 2012.