

# A Deep Metric Learning Approach for Content Based Image Retrieval in Rock Tomography: A Triplet Loss Study

Anderson Silva\*, Marcos Farias\*, João Souza\*, Daniel Pinto\*, Filipe Belfort\*, Mario Freitas\*, Alexandre Araújo\*, Alexandre Pessoa\*, Aristófanés Silva\*, Andrey Rodrigues†, Marcelo Albuquerque†

\*Applied Computing Group (NCA), Federal University of Maranhão (UFMA), Av. dos Portugueses, São Luís, 65.085-580, MA, Brazil

Emails: {anderson.silva, marvinfar852, joao.souza, daniel.pinto, cpbelfort, mario.freitas, alexandrearaujo, alexandre.pessoa, ari}@nca.ufma.br

†Tecgraf Institute, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), R. Marquês de São Vicente, Rio de Janeiro, 38097, RJ, Brazil

Emails: arodrigues@tecgraf.puc-rio.br, mralbuquerque@petrobras.com.br

**Abstract**—The automatic analysis of images in rock tomography datasets still is an unexplored task despite its high relevance in geological studies and characterization of materials. Traditional Content-Based Image Retrieval (CBIR) methods face difficulties when dealing with the structural complexity of these images while approaches based on deep learning still lack research specifically at this domain. In this paper we propose a CBIR approach using Deep Metric Learning with Triplet Loss function and different convolutional backbones. The results on the sandstone samples dataset has shown superior performance than the method found in the literature. The highlight was the DenseNet121 architecture, which obtained  $99.23\% \pm 0.14$  of F1-Score and  $99.09\% \pm 0.16$  of mAP@10. These results show the potential of the proposed approach to efficiently structure the similarity space in rock tomography images, demonstrating the relevance of the choice of a deep encoder when retrieving tomographic images in geological contexts.

## I. INTRODUCTION

The analysis of physical and microstructural features of images is a relevant task to various segments of the materials industry, including mining, construction and energy [1], [2]. From this perspective, the analysis of computed tomography (CT) allows three-dimensional visualization of the internal structure of rock samples, taking into account their structural integrity [3], [4]. The automation of these processes has a direct impact on the industry's production, guiding the procedure of exploration and reducing inspection times, cutting costs and providing assistance to human experts [4].

Content-Based Image Retrieval (CBIR) is a consolidated approach in the field of computer vision, aimed at automatically searching for images that are visually similar to a query image in large-scale datasets [5]. Historically, CBIR applications have used manual descriptors, such as texture histograms, LBP, HOG or SIFT, followed by comparison in vector space using a distance metric [5], [6]. Albeit they work in general domains, these methods have limitations in detecting heterogeneous patterns and capturing nuances of

shared features in the structure as a whole, such as those observed in CTs [6].

Although some initiatives have already explored the use of CBIR for other industries, the literature still lacks research focused on CT rock samples [7]. Due to the growing importance of analyzing structural characteristics in this type of image, the complexity of the internal patterns, the high textural variability between samples and the need for approaches adapted to the geological context, there is a demand for solutions that involve retrieval by similarity in this scenario.

Consequently, advances in deep learning techniques have significantly boosted the performance of image retrieval systems. Convolutional Neural Networks (CNNs) in particular have proven to be highly effective in automatically extracting discriminative representations from images, replacing traditional manual descriptors [8]. In this context, Deep Metric Learning (DML) techniques propose not only the extraction of representations, but also the direct learning of a metric space in which the proximity between vectors encodes semantic similarity relationships between images [9]. In these approaches, neural networks are trained to minimize distances between similar samples and maximize separation between distinct samples, according to a specific loss function [10].

In this paper, we propose an experimental CBIR pipeline for rock sample tomography, using Deep Metric Learning with Triplet Loss in conjunction with convolutional networks as feature extractors. The goal is to evaluate the combination of CNNs with Deep Metric Learning in capturing structural variations that reflect relevant geological patterns to the rock domain. We expect that our findings will provide evidence for real-world applications in the materials industry, such as reducing costs and providing assistance to human experts. The main contributions generated by this work were:

- Proposal and validation of a CBIR approach using Deep Metric Learning with Triplet Loss for rock tomography;

- Comparative analysis of multiple CNN backbones, demonstrating superior performance over prior work, without relying on metadata of physical attributes.

## II. BACKGROUND

### A. Content-based Image Retrieval

CBIR techniques enable the direct comparison of visual features extracted from images for retrieval purposes, and have been successfully applied in fields such as medicine, security, and e-commerce [11], [12]. In the geological context, the use of CBIR has emerged as an alternative for dealing with the growing volume of data generated by the industry. This is especially relevant when the efficient retrieval of rock sections with similar internal structures can facilitate the comparison of related regions, the selection of representative samples for study, or quality control processes [7].

In CBIR tasks, CNNs are widely used as feature extractors across various domains, producing high-dimensional feature vectors that represent the content extracted from images [13]. This approach has enabled significant gains in retrieval precision, especially in scenarios where the definition of visual similarity depends on subtle aspects of the image’s texture and composition, as in rock CTs.

### B. Triplet Loss

Triplet Loss [14] is a loss function widely used in the context of DML. Unlike traditional supervised approaches, which are trained to classify examples into discrete categories, DML with Triplet Loss seeks to structure the feature space so that samples considered similar are closer together, while dissimilar samples are further apart. Triplet Loss works by using input tuples made up of an anchor image, a positive image (of the same class or concept as the anchor) and a negative image (of a different class to the anchor) [15]. Its objective is to ensure that the distance between the anchor and the positive example is smaller than the distance between the anchor and the negative example, respecting a predefined margin of separation. This structure allows the model to capture relative proximity relationships, which are especially useful in scenarios where the boundaries between classes are ambiguous or where there is an interest in modeling similarity rather than rigid classifications [15], [16].

When applied to CBIR, these techniques have shown promising results in several applications in different specialized domains, allowing images to be organized in a vector space that is more coherent with the expected visual relationships [16], [17]. In domains such as rock tomography, where class concepts are often ambiguous, metric learning offers a more flexible and effective alternative for modeling the notion of similarity between samples [7].

## III. MATERIALS AND METHOD

This section introduces the utilized dataset to evaluate the proposed method, the description of the network architecture, the evaluation procedure, and the inference process of the CBIR approach. Figure 1 illustrates the overall flow of the

proposed method, with each stage detailed in the following subsections.

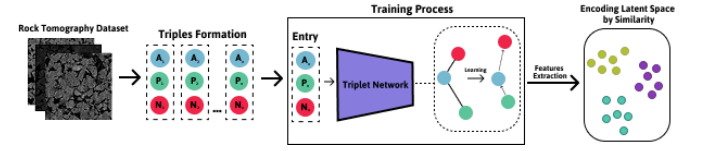


Fig. 1. Proposed method based on the formation of triples, training process, and encoding of images in latent space

### A. Rock Samples Dataset

The experiments presented in this paper were conducted with the 11 Sandstones [18] dataset, publicly accessible through the Digital Porous Media Portal [19]. This dataset brings together 11 CT volumes of sandstone rock samples: Bentheimer, Bandera Brown, Bandera Gray, Buff Berea (BB), Berea, Berea Sister Gray (BSG), Berea Upper Gray (BUG), Castlegate, Kirby, Leopard and Parker. Each volume has a resolution of  $1000^3$  voxels in 16-bit and is composed of 1000 slices. This dataset was chosen for its variety of samples and solid foundation in previous studies of image and physical attribute analysis (such as analyzing pore connectivity, geometric properties, and rock heterogeneity) [20], [21], which reinforces the dataset’s relevance for geological and petrophysical characterization tasks. It also exhibits significant variations in terms of internal structure, texture, and pore connectivity, which makes it suitable for studying image retrieval methods based on similarity.

### B. Preprocessing

Due to the computational limitations of our training environment, a preprocessing stage was performed to overcome them. An initial quantization step transformed the voxels to grayscale with a depth of 8 bits, followed by a resize of each volume slice to  $128 \times 128$  pixels. A representative view of one slice per volume is shown in Figure 2. Each volume from the dataset represents a labeled class, since the labels allow quantitative evaluation of the model’s performance based on supervised measures of similarity.

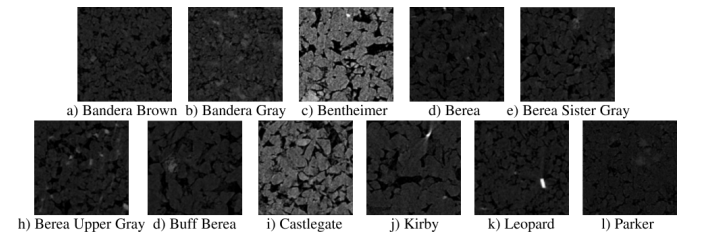


Fig. 2. Images example of each rock sample from 11 Sandstones dataset

### C. CNNs for Feature Extraction

Convolutional Neural Networks are known for their wide application in the field of image processing and pattern recognition. This architectural model is defined by its automatic

learning capacity, lower possibility of overfitting and the quality of the extracted features [13]. These features allow a better generalization of feature extraction combined with lower computational costs. In the context of CBIR, this characteristic has proven particularly advantageous, as it eliminates the need for manual definition of descriptors and better adapts to the visual variations present in images. Wangi and Makandar (2024) achieved higher results on benchmark datasets than other works in the literature that used different feature extractors. This reinforces the better adaptability of convolutional methods for the image domain. Similarly, dos Anjos et al. (2021) conducted experiments on the retrieval of rock images acquired by micro-CT using CNNs. The authors achieved great results for a visually homogeneous dataset, highlighting the benefits of CNNs on such cases.

Besides that, there are indications that the use of deeper or more specialized architectures can provide additional gains in the quality of the extracted representations. These architectures incorporate mechanisms such as residual connections, composite depth and width scaling or efficiency optimizations, making them more suitable for capturing structural nuances present in complex domains like tomographic images of rocks [23] [24]. Thus, the choice of a convolutional backbone becomes an important step in constructing CBIR approaches based on deep learning.

#### D. Triplet Loss

Triplet Loss [14] is calculated from image triplets made up of an anchor, a positive image belonging to the same concept or class as the anchor, and a negative image associated with a different class or concept. The goal is to reduce the distance between the embeddings generated by the convolutional network of the anchor and the positive sample while increasing the distance between the anchor and the negative sample, respecting an  $\alpha$  separation margin. This  $\alpha$  hyperparameter defines the minimum interval necessary to consider that the network has correctly learned to separate the samples in representation space. The Triplet Loss formulation is formally presented in Equation 1.

$$L_{\text{Triplet}} = \max \left( \|G_W(X) - G_W(X_p)\|_2^2 - \|G_W(X) - G_W(X_n)\|_2^2 + \alpha, 0 \right) \quad (1)$$

In Equation 1,  $X$  represents the input anchor image,  $X_p$  corresponds to the positive image, and  $X_n$  represents the negative image. Then,  $G_W$  represents the encoding function learned by the neural network for  $W$ , which transforms the input images (anchor, positive and negative) into a vector representation in the embedding space. The function  $\|\cdot\|_2^2$  calculates the squared Euclidean distance between the embeddings, indicating the degree of proximity in the learned vector space.

This loss function advantage lies in its ability to organize the feature space based on similarity relationships rather than relying exclusively on class labels. This feature is particularly relevant for domains such as rock tomography images, where the visual and structural differences between samples can be subtle and complex to capture using traditional methods.

#### E. Triplet Network

The Triplet Network architectural model, as introduced by Hoffer and Ailon (2015), consists of three identical instances of the same neural network sharing the same weights. These encoding structures are responsible for transforming inputs into feature vectors in a latent space. The network receives a trio of samples - an anchor, a positive sample and a negative sample - and calculates the distances between the embeddings of the anchor and the other two samples. This enables the model to learn discriminative representations based solely on relative similarity relationships, eliminating the need for direct classification. Figure 3 presents the structure, the encoding process and the approximation or distancing between embeddings in the latent space in relation to the anchor.

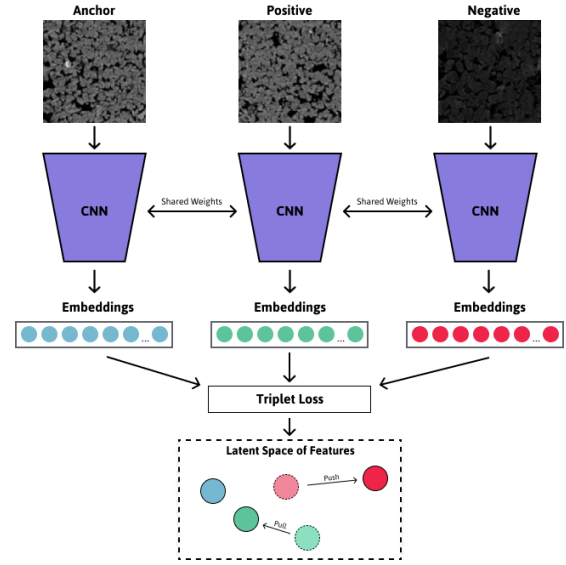


Fig. 3. Model Architecture using Triplet Loss

#### F. Metrics

Quantitative metrics, already widely used in deep learning studies, were employed to evaluate different aspects of the reported retrievals. The F1-Score, presented in Equation 2, is defined as the harmonic mean between Precision and Recall. Precision measures the proportion of relevant images among the results returned, while Recall indicates the proportion of relevant images retrieved from the total number of relevant images in the dataset. Therefore, F1-Score represents the ratio of relevant returns in relation to the total of returned instances and the ratio of relevant returns among the entire set of also relevant data [25].

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2)$$

In addition, mAP is characterized by measuring the quality of the ranking of retrievals, quantifying the information relevance returned at the top of the list. The AP (Average Precision), defined in Equation 3, computes the precision for each top  $N$ , where  $P(k)$  represents the precision returned

for each slice in position  $k$  and  $r(k)$  the relevance function, returning 1 for relevant slices and 0 otherwise. The sum of these values divided by the number of relevant images  $R$  results in the AP. Then, the mAP is obtained as the average of the APs of all queries, as illustrated in Equation 4.

$$AP = \frac{1}{R} \sum_{k=1}^n P(k) * r(k) \quad (3)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (4)$$

### G. Inference

During the inference stage of the DML-based CBIR solution presented in this paper, all the images in the dataset utilized for training are previously processed by the trained model, which transforms them into feature vectors in a latent space. These embeddings are stored for future queries. When a new query image is provided, it undergoes the same encoding process, utilizing the weights learned by the network. The resulting vector is then compared with the vectors in the encoded dataset using a distance function such as the Euclidean distance. The most similar images are those with the smallest distance in latent space, being returned as the most relevant results.

## IV. RESULTS AND DISCUSSION

This section presents the results obtained for the image retrieval task using Triplet Loss to train models based on the Triplet Network. It is also described the procedures adopted to prepare the data, the configuration of the experiments, the methods used to evaluate the similarity searches and, finally, an analysis of the results obtained from the retrieved images.

### A. Experiments Settings and Dataset Preparation

Before the experiments, each sample from *11 Sandstones*, comprising 11 sandstone volumes measuring 1000x1000x1000, was split in half, with the first half used for training and the second for evaluation. Finally, the training set comprises the first half of each volume in the dataset, while the test set comprises the second half of each volume. The final sets consist of 5500 slices for training and 5500 slices for testing.

All experiments were conducted on a machine equipped with an Intel Core i7-14700 processor and an NVIDIA RTX 3060 GPU (12GB). For the triplet formation process, we adopted a strategy based on random sampling. Each image was used as an anchor, then a positive sample was randomly selected from the same class, and a negative sample was randomly chose from a different class. This procedure follows the traditional protocol for triplet generation. The model was implemented with Keras 3.4.1 [26] and trained for 10 epochs with the Triplet Loss function using different backbones of CNNs pre-trained on ImageNet [27]: ResNet50 [23], MobileNetV2 [28], DenseNet201 [29], EfficientNetB0 [24], VGG16 [30] and ConvNeXtTiny [31]. A batch size of 32 and the Adam [32] optimizer with a learning rate of 0.0001 were used. Also,

the features extracted by the CNN are passed through a fully connected layer, which performs the final transformation of the extracted information into a 32-dimensional array. The values of all hyperparameters adopted were defined based on preliminary experiments to balance performance and computational cost. Each experiment with the different backbones was run five times independently. The values reported in the metrics correspond to the mean and standard deviation of the results obtained on these runs. Additionally, Euclidean distance was used in both the calculation of loss during training and the inference step. The evaluation metrics were based on retrieving the 10 closest images, considering each volume as a distinct class. The choice of a cutoff of 10 was based on previous CBIR studies, where this value is widely adopted as a performance metric [7], [33].

### B. Results Analysis

The results obtained in the experiments demonstrate the positive impact of using consolidated backbones for extracting features in similarity-based image retrieval tasks. These results can be seen in Table I. All models based on pre-trained CNNs performed significantly better than the baseline CNN used as a reference, built with two convolutional layers interleaved by pooling, followed by a dense layer.

TABLE I  
CBIR PERFORMANCE BY CNN ARCHITECTURE

Backbone	F1-Score (%)	mAP@10 (%)	Parameters
Baseline CNN	85.71 ± 1.45	84.89 ± 1.70	16.8M
ResNet50	98.22 ± 0.48	97.96 ± 0.57	25.6M
MobileNetV2	94.35 ± 1.02	93.35 ± 1.15	3.5M
DenseNet121	<b>99.23 ± 0.14</b>	<b>99.09 ± 0.16</b>	8.1M
EfficientNetB0	98.99 ± 0.42	98.77 ± 0.49	5.3M
VGG16	98.40 ± 1.13	98.22 ± 1.27	138.4M
ConvNeXtTiny	98.57 ± 1.06	98.33 ± 1.24	28.6M

The baseline CNN obtained the worst results among the models evaluated, with F1-Score of 85.71% and mAP@10 of 84.89%. In contrast, architectures such as DenseNet121, EfficientNetB0 and VGG16 achieved more expressive results, with DenseNet121 standing out as the best overall performance with F1-Score of 99.23% and mAP@10 of 99.09%. These values indicate not only greater precision in retrieving the most similar images but also superior stability across runs, as evidenced by the lower standard deviations.

The MobileNetV2 also demonstrated good performance stability despite its lightweight architecture, which is also a characteristic of EfficientNetB0, and even achieved results close to those of models with a higher number of parameters. ResNet50 and VGG16 likewise showed solid performance, with F1-Scores above 98% and good consistency across runs, reinforcing the effectiveness of well-established architectures in the literature. Additionally, the ConvNextTiny architecture also delivered competitive results, achieving a mean F1-Score of 98.57% and a mAP@10 of 98.33% ± 1.24.

Thus, pre-trained models proves to be a more effective solution when compared to deep models that do not use

any type of transfer learning. This is because these pre-trained CNNs are capable of extracting better features, which directly impacts the model’s ability to differentiate the features of distinct examples [13], [34]. These findings also support the importance of selecting a good encoder in the context of metric learning applied to CBIR. Pre-trained models on large datasets such as ImageNet enable more effective generalization, even in specific domains like rock tomography images, contributing to the formation of more discriminative and task-appropriate latent representations. Thus, the results demonstrate that adopting well-established backbones is a beneficial choice for similarity-based retrieval, both in terms of performance and stability.

Figure 4 presents examples of image retrieval performed using the model trained with the DenseNet121 architecture, which achieved the best results in the experiments. Green borders represent correct matches, while red borders indicate retrieval errors. It can be observed that for different query images, the proposed method mostly returns samples belonging to the same original volume, indicating that the network has learned to structure the similarity space consistently with the intended objective. This behavior demonstrates that training with Triplet Loss was effective in bringing similar slices closer together in the latent space, validating the approach for volume-based retrieval.

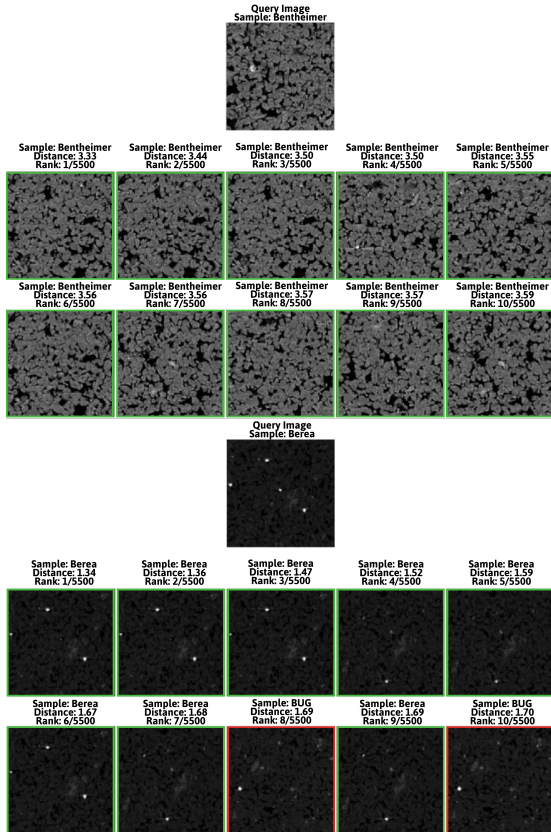


Fig. 4. Retrieval examples in 11 Sandstones dataset

### C. Discussion

After collecting the results, a comparison was made with the only method proposed for the CBIR task within the same domain of rock sample CT scans. [7] proposed an approach based on a Double Siamese Neural Network (DSNN), which combines visual similarity learning with the incorporation of physical rock properties directly into the loss function. The model consists of two parallel Siamese networks: one focused on comparing visual features using Contrastive Loss, and the other dedicated to the regression of physical properties (permeability and porosity) between samples. This approach aims to enhance image retrieval not only based on visual similarity but also on the physical-geological compatibility between volumes.

However, since the dataset used by [7] is private, it was necessary to use their model in the context of the 11 Sandstones dataset, which also provides physical properties information (porosity and permeability), making it possible to compare with the described work. In this way, the DSNN model was replicated in a scenario comparable to ours, following the same evaluation criteria applied to our proposed method.

The results are presented in Table II, which shows that the approach using Triplet Loss and DenseNet201 delivers considerably superior performance, achieving 99.06% F1-Score and 98.91% mAP@10, compared to 92.96% and 90.91%, respectively, obtained with the DSNN model. This difference highlights the performance gain achieved by using a Triplet Network with Triplet Loss and a deep CNN-based backbone. Our approach was able to structure the latent space more effectively and capture more discriminative representations, even without incorporating external physical variables.

TABLE II  
COMPARISON WITH LITERATURE

Models	F1-Score	mAP@10
[Shim et al. 2023]	92.96 ± 0.62	90.91 ± 0.78
Triplet Loss + DenseNet121	<b>99.06 ± 0.27</b>	<b>98.91 ± 0.26</b>

The results obtained, along with the consistently superior performance compared to the existing literature, demonstrate the effectiveness of the proposed method for retrieving tomographic rock images. These findings not only validate the methodological proposal but also reinforce its applicability in real-world scenarios, establishing a solid foundation for future research in the field. Although the experiments in this work were performed on a dataset consisting of only one material, this choice provided a reliable scenario for the initial validation of the study proposal. Furthermore, it is understood that the importance of expanding this evaluation in future experiments with datasets from other types of materials, aiming to provide a more comprehensive analysis of the generalization capacity of the proposed solution.

### V. CONCLUSION

Based on the experiments and results, this work demonstrated the effectiveness of applying Triplet Networks with

Triplet Loss for the retrieval task of rock sample tomography images. Using different CNN-based backbones yielded significant improvements over simpler architectures, emphasizing the importance of encoder choice for the quality of latent representations. The proposed approach outperformed the literature method even without incorporating explicit physical properties, reinforcing its potential as a robust and scalable CBIR solution in geological domains.

The best performance was achieved with DenseNet121, yielding an F1-Score of 99.23% and mAP@10 of 99.09%. As future work, this approach will be tested on other geological and industrial materials, extending its applicability to diverse structures. Ultimately, this study provides practical evidence of Deep Metric Learning for image retrieval from rock tomographic datasets, with potential for industrial applications such as structural analysis and informed decision-making.

#### ACKNOWLEDGEMENTS

The authors acknowledge the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and Fundação de Amparo à Pesquisa Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Brazil).

#### REFERENCES

- [1] R. Subramani, M. A. Mustafa, G. K. Ghadir, H. M. Al-Tmimi, Z. K. Alani, D. Haridas, M. A. Rusho, N. Rajeswari, A. J. Rajan, and A. P. Kumar, "Advancements in 3d printing materials: A comparative analysis of performance and applications," *Applied Chemical Engineering*, vol. 7, no. 2, pp. ACE-3867, 2024.
- [2] X. Liu and C. Aldrich, "Deep learning approaches to image texture analysis in material processing," *Metals*, vol. 12, no. 2, p. 355, 2022.
- [3] A. Teles, A. Machado, A. Pepin, N. Bize-Forest, R. Lopes, and I. Lima, "Analysis of subterranean pre-salt carbonate reservoir by x-ray computed microtomography," *Journal of Petroleum Science and Engineering*, vol. 144, pp. 113-120, 2016.
- [4] N. Saxena, A. Hows, R. Hofmann, F. O. Alpak, J. Dietderich, M. Appel, J. Freeman, and H. De Jong, "Rock properties from micro-ct images: Digital rock transforms for resolution, pore volume, and field of view," *Advances in Water Resources*, vol. 134, p. 103419, 2019.
- [5] A. S. Ahmed and I. N. Ibraheem, "Recent advances in content based image retrieval using deep learning techniques: A survey," in *AIP Conference Proceedings*, vol. 3219, no. 1. AIP Publishing, 2024.
- [6] S. S. Sadiq, "Improving cbir techniques with deep learning approach: An ensemble method using nasnetmobile, densenet121, and vgg12," *Journal of Robotics and Control (JRC)*, vol. 5, no. 3, pp. 863-874, 2024.
- [7] M. S. Shim, C. Thiele, J. Vila, N. Saxena, and D. Hohl, "Content-based image retrieval for industrial material images with deep learning and encoded physical properties," *Data-Centric Engineering*, vol. 4, p. e21, 2023.
- [8] J. I. Bhat, R. Yousuf, Z. Jeelani, and O. Bhat, "An insight into content-based image retrieval techniques, datasets, and evaluation metrics," in *Intelligent Signal Processing and RF Energy Harvesting for State of art 5G and B5G Networks*. Springer, 2024, pp. 127-146.
- [9] G. Hoxha, G. Sumbul, J. Henkel, L. Möllenkamp, and B. Demir, "Annotation cost-efficient active learning for deep metric learning driven remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
- [10] K. K. Suresh, S. Sundaresan, R. Nishanth, and K. T. Ananth, "Optimization and deep learning-based content retrieval, indexing, and metric learning approach for medical images," *Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications*, pp. 79-106, 2021.
- [11] A. Anand, K. Singh, and A. Saxena, "A detailed survey on content based image retrieval (cbir)," 2024.
- [12] M. Bano, P. Matta, and S. Chandel, "Content based image retrieval: A study of approaches and techniques," in *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*. IEEE, 2024, pp. 16-22.
- [13] K. Wangi and A. Makandar, "Cnn pre-trained model using the fusion of features for cbir framework," in *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*. IEEE, 2024, pp. 1-5.
- [14] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *Similarity-based pattern recognition: third international workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3*. Springer, 2015, pp. 84-92.
- [15] G. Kertész, "Different triplet sampling techniques for lossless triplet loss on metric similarity learning," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMII)*. IEEE, 2021, pp. 000 449-000 454.
- [16] K. Rahbar and F. Taheri, "Enhancing image retrieval through entropy-based deep metric learning," *Multimedia Tools and Applications*, pp. 1-27, 2024.
- [17] H. Lin, Y. Fu, P. Lu, S. Gong, X. Xue, and Y.-G. Jiang, "Tc-net for isbir: Triplet classification network for instance-level sketch based image retrieval," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1676-1684.
- [18] R. Neumann, M. Andreetta, and E. Lucas-Oliveira, "11 sandstones: raw, filtered and segmented data," <https://www.digitalrockportal.org/projects/317>, 2020.
- [19] Digital Porous Media Portal, "Digital porous media portal," <https://digitalporousmedia.org>, 2024, accessed: 2024-05-23.
- [20] M. Haq, I. N. Yulita, and I. Dharmawan, "A study of transfer learning in digital rock properties measurement," *Machine Learning: Science and Technology*, vol. 4, no. 3, p. 035034, 2023.
- [21] S. Zhong, X. Ge, H. R. Thomas, and C. Li, "Investigation of strain-sensitive properties of porous media through micro-ct imaging and numerical modelling," *Computers and Geotechnics*, vol. 174, p. 106560, 2024.
- [22] C. E. dos Anjos, M. R. Avila, A. G. Vasconcelos, A. M. Pereira Neta, L. C. Medeiros, A. G. Evsukoff, R. Surmas, and L. Landau, "Deep learning for lithological classification of carbonate rock micro-ct images," *Computational Geosciences*, vol. 25, pp. 971-983, 2021.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [24] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105-6114.
- [25] S. Kumar, M. K. Singh, and M. Mishra, "Efficient deep feature based semantic image retrieval," *Neural Processing Letters*, vol. 55, no. 3, pp. 2225-2248, 2023.
- [26] F. Chollet, "Keras," 2015. [Online]. Available: <https://keras.io>
- [27] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211-252, 2015.
- [28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520.
- [29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [31] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976-11 986.
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [33] Z. Cheng and J. Shen, "On very large scale test collection for landmark image search benchmarking," *Signal Processing*, vol. 124, pp. 13-26, 2016.
- [34] A. Ahmed, A. O. Almagrabi, and A. H. Osman, "Pre-trained convolution neural networks models for content-based medical image retrieval," *Int. J. Adv. Appl. Sci.*, vol. 9, p. 12, 2022.