8th Global Stone Congress 2025, Drama, Greece, June 16 – 20, 2025

NATURAL STONE IMAGE CLASSIFICATION USING ONLINE DATABASES AND CONVOLUTIONAL NEURAL NETWORKS

J. H. Brito[1,2*], D. M. Morim[1], C. Carvalho[3], R. Alves[4]

(1) 2Ai - School of Technology, IPCA, Barcelos, Portugal *jbrito@ipca.pt
(2) LASI – Associate Laboratory of Intelligent Systems, Guimarães, Portugal
(3) LNEG – National Laboratory for Energy and Geology
(4) Lab2PT – Universidade do Minho

Summary: *In this paper, we discuss the relevance of two distinct types of online natural stone databases (technical-institutional repositories and logistics-commercial e-platforms) for extracting (training and testing) input images and creating an automatic visual inspection system for natural stone classification. Using convolutional neural networks (CNNs) and images from a selected online data repository, a Deep Learning (DL) system was developed to estimate the class of the natural stone in a given image. The DL models were developed through transfer learning from existing image classifiers, as pre-trained classifiers were retrained on our dataset. Our best model achieved an Accuracy of 70.3% and an F-score of 0.67 for 70 classes.*

Key words: *Online Natural Stone Databases, Visual inspection, Stone classification, Computer Vision, Deep Learning*

1. Introduction

Over the last decade, many online inventories of natural stone (blocks and especially slabs) have been developed and promoted as part of the marketing and e-commerce strategies of companies in the sector (e.g. LongaritoSG, DraenertSS, BrachotLS). Additionally, there has been a significant improvement in the digital databases of public institutions, universities, museums, archives, and even private collections is also remarkable (e.g. LNEG ROP, OxfordCC, GreekMA).

Institutional databases represent reference collections, with a large number of stones, with good resolution images, associated geometric scales and relevant metadata, such as petrographic classification, quarry location, and physical-mechanical properties. Conversely, commercial databases are updated daily with new products, but they feature variable resolution images, lack standardisation and only contain strict commercial information (denomination and size).

The first premise of this work is the potential complementarity between these two types of digital databases for automatic identification/classification with AI, e.g. training with images from institutional databases and automatically classifying images from commercial databases, using a standardised taxonomy. The second premise is the definition of a stone classification system for digital images, covering the main taxonomic classification presented in EN12670: 2019. This facilitates product quality assurance and may avoid trade legal litigation (Bastida *et al.* 2010), especially for construction products (whose applicable standards and regulations are currently under review (Reg. (EU) 2024/3110).

The objective of this work is to develop a system to automatically identify the stone class from a stone image. To develop the system, a dataset was created and used to train a DL classifier. The dataset was created with data from the Deutsches Naturstein Archiv (DNSA) and comprises stone images and their respective classes.

The conventional classification system is framed in a high-level genetic class set (igneous, metamorphic and sedimentary) based on compositional and/or mineralogical, petrographic and/or textural attributes (Gillespies & Styles, 1999 a, b and Hallsworth & Knox, 1999). These attributes have different hierarchical weights in class designation, resulting in hundreds of low-level classes, forming a very large class set. To reduce the number of classes and make them compatible with computer-based query systems, a multi-hierarchical classification system was proposed by Struik *et al.* (2002). This system is applicable to geological rock databases in a stratigraphic cartography domain, but it also could be tested with natural stones. The data from DNSA comprises images of natural stone products that have been standardised by a similar denomination criteria (EN 12440) based on petrographic groups (amongst other criteria) and also by a standard classification (EN 12670, specifically its Annex A), with an exemplar set of 147 classes (96 igneous, 27 sedimentary and 23 metamorphic). To fulfil the objective defined in this work, a subset of classes in each genetic group was selected (29 igneous, 24 sedimentary and 14 metamorphic), consistent with the DNSA class set. For the classifier, several pretrained DL models were retrained with the data, and their performance was evaluated and compared.

2. Previous Work

Previous works have addressed the automatic classification of stone images (Perez *et al.* 2015, Deng *et al.* 2017, Kaynar *et al.* 2018), using visual feature extractors and classical Machine Learning classifiers. More recent approaches employ DL models (Ferreira *et al.* 2017, Pascual *et al.* 2019, Canayaz *et al.* 2020, Xu *et al.* 2021). These approaches propose a new CNN or repurpose existing DL networks to classify stone images, demonstrating the superior performance of DL-based solutions. Authors emphasise the importance of a large and diverse set of examples with good quality images. However, they typically focus on a specific type of stone (Ferreira *et al.* 2017, Canayaz *et al.* 2020), or use very few classes (Pascual *et al.* 2019). They propose datasets tailored to the needs of a particular application which are not reusable in other settings, due to their reduced class sets, and lack of a general classification framework. The work by Xu *et al.* (2021) proposes a significantly large dataset (8,974 images) with a relatively large number of classes (30) and shows very promising results. However, the authors use a taxonomy that remains somewhat restrictive.

3. Methodology

The first step in building a DL system is acquiring data with abundant, diverse, and good quality examples. Our approach was to search online repositories that met specific criteria: a) a large number of images b) consistent size, scale, and image quality, and c) classification data with significant genetic and petrographic diversity. DNSA meets these criteria, providing a large number of images of 70 classes from several countries around the world. The images have consistent acquisition conditions (camera pose, lighting conditions, geometric scale). This data was cleaned, resulting in a dataset size of 4,338 images.

The classification taxonomy used by DNSA was minimally adjusted to follow EN 12670, and the classification of some images was corrected. The dataset was split into a training set and a test set (75%-25%), following the standard supervised learning approach, with a stratified split. Fig. 1 shows the class histogram in the training set, i.e. the number of images for each class in the dataset. From Fig. 1 it is evident that the dataset is quite unbalanced, as most examples belong to 5 classes, and around a third of the classes have only 1 or 2 examples. Augmentation techniques were used to mitigate these issues and generate several equally valid training samples. The original images were randomly cropped in 128x128 sample windows, randomly rotated, flipped horizontally and vertically.

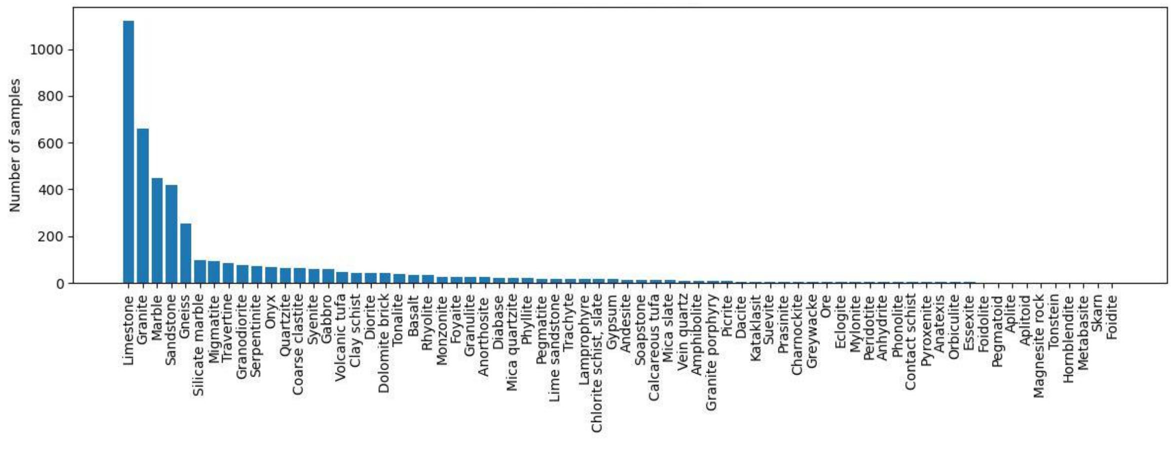

Figure. 1. Dataset histogram with the number of images for each DNSA 70' classes (n=4338 images).

To develop the classifier, the approach was to reuse existing CNNs. These CNNs have achieved good results in other image classification tasks and were adapted and retrained on the images in the dataset, in a process commonly referred to as transfer learning. Following other works (Su *et al.* 2018), the selected pretrained models include popular variants of VGG, Inception, ResNet, DenseNet, MobileNet, NASNet, EfficientNet.

The pretrained models were adapted and retrained to build DL classifiers. All adapted models were created by taking the original network pretrained on the ImageNet dataset, removing the last classification layer, freezing the network parameters, and adding a fully-connected layer with 512 neurons, a classification layer with a number of neurons equal to the number of classes in the dataset, and a SoftMax activation. Only the parameters of the new fully-connected and classification layers were trained. The images were normalised according to the requirements of each of the different pretrained CNNs. Each training session was performed for 100 epochs, with a learning rate of 1e^{-4}, and a batch size of 16. For each training session, the model with the lowest loss on the test set was saved at the end of each epoch.

Following training, each model was evaluated using the usual classification metrics of Precision, Recall, F1-score, and Accuracy. To evaluate the performance, one test sample was extracted from each test image, by cropping a 128x128 area from the top left corner without further augmentation or transformation.

4. Results

Fig. 2 shows typical classification results, where the probability of each class, as estimated by the classifier, is presented, alongside with the correct (ground truth) stone class. Table 1 shows the performance of the various classifiers. Given the class imbalance, the most relevant metric is F1-score, but Accuracy is also presented. The best performing model was DenseNet201, although other models also showed good results.

Table 1. Classification metrics (F1-score/Accuracy) for each classifier

Models	F1-score/Accuracy
DenseNet121	0.65/68.2%
DenseNet169	0.66/69.1%
DenseNet201	**0.67/70.3%**
EfficientNetB0	0.63/66.5%
EfficientNetB7	0.55/60.1%
InceptionV3	0.58/61.1%
MobileNetV2	0.61/64.8%
MobileNetV3Small	0.60/64.0%
MobileNetV3Large	0.61/63.5%
NASNetMobile	0.60/63.3%
NASNetLarge	0.51/56.0%
ResNet152V2	0.57/61.9%
VGG16	0.63/66.7%
VGG19	0.65/68.2%

The models were then tested on an additional dataset of images from Carvalho *et al.* (2023), featuring stones from Portugal. This dataset includes 61 images of 7 classes (46 granite, 7 schist, 3 granodiorite, 2 slate, 1 syenite, 1 migmatite, and 1 quartzite). The classification metrics achieved in Table 3 are comparable to those in Table 2. However, there are stark disparities in performance, for the individual classes. Most stones in the additional dataset are granites, for which the classifier was well trained, as the granite class is very well represented in the training set. Using the DenseNet201 model, the granite class achieves a Precision of 0.98, Recall of 0.91, F1-score of 0.94, and Accuracy of 89.4%. On the other hand, the schist images are all miss-classified, as this class is extremely underrepresented in the training set. Furthermore, these images are visually very different to the training samples, as the stones in the training set were polished and cut in the direction of the main grain alignment (*crosscut*), whereas the stones in the additional dataset are not polished and were cut perpendicular to the main grain alignment (*vein cut*) (Ouzounis *et al.* 2021). This highlights the need for very good quality datasets for training.

Table 2. Classification metrics (F1-score/Accuracy) for a dataset of Portuguese Stones (Carvalho et al. 2023)

Models	F1-score/Accuracy
DenseNet121	0.73/**70.5%**
DenseNet169	0.67/59.0%
DenseNet201	0.73/**70.5%**
EfficientNetB0	**0.75/70.5%**
EfficientNetB7	0.74/**70.5%**
InceptionV3	0.67/59.0%
MobileNetV2	0.69/65.5%
MobileNetV3Small	0.73/68.9%
MobileNetV3Large	0.70/63.9%
NASNetMobile	0.72/67.2%
NASNetLarge	0.73/68.9%
ResNet152V2	0.67/62.3%
VGG16	0.70/65.6%
VGG19	0.72/68.9%

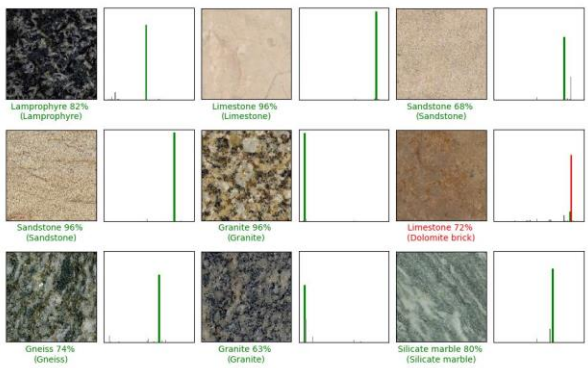

Fig. 2. Stone Classification for DenseNet201

5. Conclusion

This work presents a system to automatically classify stone images within a generic classification taxonomy, based on the retraining of pretrained DL networks. The taxonomy was designed to be applicable to generic applications, as the class set is general enough to be specialized for specific settings. The metrics are reasonably good, considering the number of supported classes and the limitations of the dataset.

Future work includes expanding the dataset to include more examples of underrepresented classes, which should improve results for these classes and the overall metrics. This should be accompanied by modifying the training process so that underrepresented classes are given more importance than overrepresented classes. Another short-term objective is to apply the system to other specific settings, applications and test datasets with class subsets, and evaluate the performance. The results on the dataset of Portuguese stones are encouraging and motivates further experiments to apply the system to commercial datasets of online inventories.

Acknowledgments: This work was partially funded by national funds, through the FCT - Fundação para a Ciência e Tecnologia and FCT/MCTES under the scope of the projects UIDB/05549/2020, UIDP/05549/2020 and LASILA/P/0104/2020.

References

Bastida, J.; López-Buendía, A.; Urquiola, M. (2010). Natural stone classification and specifications review, GSC2010.

BrachotLS, Brachot Live Stock [online], https://www.brachot.com/en/live-stock

Canayaz, M.; Uludag, F. (2020). Marble Classification using Deep Neural Networks. European Journal of Technique (EJT), 10(1), 52-63.

Carvalho, C. *et al.* (2023). Granitos e Xistos Ornamentais de Portugal. Carvalho, C. *(coord)*. ASSIMAGRA, 351 p.

Deng, C.; Pan, H.; Fang, S.; Konaté, A. A.; Qin, R. (2017). Support vector machine as an alternative method for lithology classification of crystalline rocks, Journal of Geophysics and Engineering, vol. 14, no. 2, pp. 341–349.

DNSA, Deutsches Naturstein-Archiv [online], https://www.steinzentrum-wunsiedel.de/natursteinarchiv.

DraenertSS. Draenert Stone Selection [online], https://www.draenert.de/en/materials/stone.html.

EN 12440:2017 — Natural stone - Denomination criteria. (CEN/TC 246, 2017).

EN 12670:2019 — Natural stone – Terminology. (CEN/TC 246, 2019).

Ferreira, A.; Giraldi, G. (2017). Convolutional Neural Network approaches to granite tiles classification, Expert Systems with Applications, Volume 84, Pages 1-11.

GreekMA. Greek Marble Atlas, https://gaia.igme.gr/portal/apps/dashboards/4c4d358ac56846b8ab34cc1c4e8f90ae

Kaynar, O.; Torun, Y.; Temiz, M.; Görmez, Y. (2018). Automatic classification of natural stone tiles with computer vision, 3rd International Conference on Computer Science and Engineering (UBMK). IEEE, pp. 527–532.

LNEG ROP. Rochas Ornamentais Portuguesas [online], https://geoportal.lneg.pt/pt/bds/rop.

LongaritoSG. Antonio Longarito Stone Gallery [online], https://antoniolongarito.com/stock.

Ouzounis, A.; *et al.* (2021). Marble Quality Assessment with Deep Learning Regression, Fifth International Conference on Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, pp. 1-5.

OxfordCC. Oxford Corsi Collection of Decorative Stones [online], https://www.oum.ox.ac.uk/corsi.

Pascual, A. *et al.* (2019). Towards Natural Scene Rock Image Classification with Convolutional Neural Networks," IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, pp. 1-4.

Perez, C. A.; *et al.* (2015). Rock lithological classification using multi-scale gabor features from sub-images, and voting with rock contour information," International Journal of Mineral Processing, vol. 144, pp. 56–64.

Struik, L.; Quat, M.; Davenport, P.; Okulitch, A. (2002) A preliminary scheme for multihierarchical rock classification for use with thematic computer-based query systems, Current research 20021701-43872002-D10, Geological Survey of Canada, ISBN 0-662-31680-0.

Su, D.; Zhang, H.; Chen, H.; Yi, J.; Chen, P.-Y.; Gao, Y. (2018). Is robustness the cost of accuracy? –a comprehensive study on the robustness of 18 deep image classification models, Proceedings of the European Conference on Computer Vision (ECCV), pp. 631–648.

Xu, Z.; Ma, W.; Lin, P.; Shi, H.; Pan, D.; Liu, T. (2021). Deep learning of rock images for intelligent lithology identification, Computers & Geosciences, vol. 154, p. 104799.