A convolutional neural network model for marble quality classification

İdris Karaali¹ · Mete Eminağaoğlu¹

Received: 9 April 2020 / Accepted: 16 September 2020 / Published online: 24 September 2020
© Springer Nature Switzerland AG 2020

Abstract

The fundamental policy of marble industries is to establish sustainable high-quality products in a standardized manner. Identification and classification of different types of marbles is a critical task that is usually carried out by human experts. However, marble quality classification by humans can be time-consuming, error-prone, inconsistent, and subjective. Automated and computerized approaches are required to obtain faster, more reliable, and less subjective results. In this study, a deep learning model is developed to perform multi-classification of marble slab images with six different quality types. Blur filter, 5×5 low-pass 2D linear separable convolution filter using Gaussian kernel, and erosion filter were applied to the images for data augmentation, and a special convolutional neural network (CNN) architecture was designed and implemented. It has been observed that the data augmentation approach for marble image samples has significantly improved the accuracy of the CNN model ranging between 0.922 and 0.961.

Keywords Convolutional neural networks · Marble images · Marble quality classification · Machine learning · Data augmentation

1 Introduction

Marbles or rock pieces from marbles homogeneous in color, texture, and construction are not usually available on earth (except for the hardest stones). Discontinuities, color differences, presence of extrinsic particles or elements, block shapes, and the levels of roughness and glossiness are the major factors that affect the quality of marbles. The marble blocks are evaluated according to the marble type, general geological characteristics, internal parameters such as cohesion, pressure, and internal friction angle, processing line, and usage characteristics. Marble quality classification is mostly carried out with the human workforce and intervention today, despite the advances in hardware and software technologies.

Marble quality usually depends on the density and size of marbles, and the fossils and stains on the marble [1]. There are several classical approaches or methods for marble quality classification, which are discussed in a literature

survey [2]. In most of the related works, the pictures are preprocessed by different filters and other techniques, such as Gabor, percentile, or chromatic features before they are fed into the classification models or systems [3]. Bianconi et al. used these image preprocessing techniques for the classification of granite tiles through color and texture features [3]. The number of samples for each quality group (from one to 12) is 48, comprising a data set of 576 samples in total. Images were cropped to 544×544 pixels to discard the distortion effect. They have obtained a 8.5% accuracy using SVM (support vector machine) algorithm [3]. K-means clustering algorithm was used in one of these studies, and an industrial application was developed for marble classification [4]. In that study, the image sizes of the marbles were set to 315×310 (width $\times$ height). The number of samples for each quality group (from one to four) was 172, 388, 411, and 187, respectively, comprising a data set of 1158 samples in total [4]. Textural features, which are extracted using sum and difference histograms

✉ Mete Eminağaoğlu, mete.eminagaoglu@deu.edu.tr | ¹Department of Computer Science, Dokuz Eylül University, Buca, İzmir 35390, Turkey.

(SDH) [5] and used for computing features of mean, variance, energy, correlation, entropy, contrast, and homogeneity, have been implemented for marble (limestone) classification [6]. In [7], wavelet transform provides the energy decomposition level of each sample and the mean, median, and variance of each level of decomposition are computed to generate the feature vector. The methodology achieves 83.60% of accuracy with the K-means clustering algorithm.

Doğan and Akay proposed a study with high classification accuracy scores by using an image dataset with four distinct marble types [8]. They proposed a system for automatic classification of marble (limestone) slabs based on sum and difference histograms and AdaBoost algorithm. AdaBoost provided accuracy scores higher than all the other methods, which were reported previously in another study [9]. They obtained the best accuracy scores as 96.52%, 91.79%, 94.53%, and 99.25%, respectively, for each quality group.

Ferreria and Giraldo proposed a method for granite tiles classification by using CNN (convolutional neural networks) [10]. They used 28×28 Gy-scale images and 32×32 color images for the training datasets in their CNN. They used the same granite tiles dataset applied in the work of Bianconi et al. [11]. The granite tiles dataset contains 1000 RGB images with 1500×1500 dimensions, which is divided into 25 different classes with 40 images for each class. The first 100 images were acquired by scanning with a specific device. The remaining 900 samples were created by rotating the original images in nine different angles that are between 10 and 90 degrees. They converted these images into small gray-scale images with 2,809,000 samples, which were used in a pre-trained neural network for MNIST [12]. Similarly, they also prepared 2,116,000 small colored images, which were used in a pre-trained neural network for CIFAR [13]. They used a 5×2 cross-validation method for performance evaluation. The MNIST-based network was trained using batches of images containing 100 images, with a constant learning rate as 0.001. They used the stochastic gradient descent method for weights' update with a momentum coefficient of 0.9, and the weighting decay was set to 0.0005, where no dropout was used in their model. They also used almost the same parameters for the pre-trained CIFAR model, where the only difference was the decay rate that was set to 0.0001. They have observed a perfect accuracy rate of 100.00% with the CIFAR model [10].

One of the recent studies that use CNN to classify marble slab images is Pençe and Çeşmeli's research article [14]. They used several CNN architectures and hyper-parameters among 80 marble images. However, they did not use any methodologies for data augmentation, and they implemented their model only for binary classification of

marbles (first and second quality). The highest accuracy they could observe was around 75% [14].

It has been shown in another recent study that deep convolutional neural networks could be effectively implemented and used for the determination of rock strength parameters [15]. They used the images of several types of rocks to train the CNN and to estimate the rock strength parameters where they obtained very high R^2 scores such as 0.998 [15]. It should also be mentioned that CNN can be adapted and used successfully for very difficult problems with complex image analysis such as quantifying image distortions caused by strong gravitational lenses [16].

This study aims to propose a novel marble quality classification model that might be implemented and used in the marble industry as autonomous agents that could minimize human intervention and human workforce requirements in the near future. To the best of our knowledge, the utilization of CNN for marble quality classification with the aid of image preprocessing techniques for data augmentation is a novel approach in the related literature.

2 Design and implementation

The volume of marble slab images used in this study is the actual production data provided by "Haz Marble–Marble, Natural Stone, Fixing System Industry and Trade Inc.," which is an international company founded in İzmir, Turkey. The number of samples for each quality class is 350 images, comprising a data set of 2100 samples in total. There are two distinct marble types and three different classes for each type, namely Q3-A, Q3-B, Q3-C, Q4-A, Q4-B, Q4-C. Some sample images from each of these different quality classes are shown in Figs. 1 and 2. It should be noted that marble quality classification is a challenging task because most of the marble samples that belong to different classes are very similar in appearance and they are difficult to distinguish even by a human expert. These

Fig. 1 Marble slab images from two different quality classes. **a** Quality class: Q3-B, **b** quality class: Q3-C

Fig. 2 Marble slab images from two different quality classes. **a** Quality class: Q4-A, **b** quality class: Q4-B

challenges and problems are also true for the dataset used in this study. As it could be seen from Fig. 3a–c, it is very difficult both for human and computer vision to perceive and categorize these marble images in the same class. In other words, both the risks of false positive and false negative categorizations of marble samples are considered high. The original images with 2400×2400 pixels were cropped, and the final original dataset comprised RGB marble images with the size of 300×300 pixels. It should be noted that the image cropping is usually preferred and recommended in CNN implementations whenever the original sizes of images are not small [17, 18]. Thus, the images with the original sizes (2400×2400) were not used in this study.

The architecture of the CNN model proposed in this study is presented in Fig. 4. Batch normalization was used for the input data. Two convolution layers with 64 and 128 filters of (3×3) sizes were designed and used. After the convolution layers, dropout was used with a rate of 0.20. Then, a flattening layer was used to convert the data into a one-dimensional array with 139,392 elements. A fully connected layer was constructed where 139,392 nodes were fully connected with 512 nodes, and dropout was used with a rate of 0.50. After that, the output layer was implemented with 6 nodes using Softmax as the activation function [17, 18]. ReLU (Rectified Linear Unit) [17] was used as the activation function within all the other layers in our

Fig. 3 Three marble slab images from Q4-C quality class

Fig. 4 The architecture of our CNN model used in this study

CNN model. It should be noted that in most of CNN models, max or average pooling is used, however, in our model; no pooling was used within the CNN. This is because the accuracy scores were slightly improved for our dataset when we did not use the pooling method.

The CNN model was implemented and coded with Python programming language version 3.6.7 using the frameworks and libraries as follows: Keras version 2.2.4, Tensorflow version 1.13.1, PyCharm Pro. version 2018.1.4, and Atom version 1.34.0. It should be mentioned that we first used the original marble images without data augmentation. However, even with many changes in the architecture and hyper-parameters in our CNN model, we could not get any successful results and there was an overfitting problem. This was also true for the results obtained by the alternative machine learning algorithms that were used in this study.

In order to improve the accuracy of test results and avoid the overfitting problem, data augmentation methods were used in our work. It should be noted that we used dropout in our architecture, which also helps to decrease overfitting, but it was not enough in our case. We first tried L1 and L2 regularization [17, 18], but they did not have any positive impact on the results. We also experimented with various image processing methods and techniques that are commonly used for data augmentation in CNN such as noise injection, image rotation, color jittering, horizontal, and vertical flips, and random cropping [17]. We also tried some special image processing filters such as Gabor, bilateral filter, median filter, Retinex, Msrnp, and Automated-Retinex [19, 20]. However, none of these improved the accuracy of test results or helped us to

avoid the overfitting problem. Thus, we combined some different filters and applied them to the cropped images. We finally achieved the best accuracy scores after various experiments by deriving a novel method for image processing of marble samples. Blur filter, 2D linear separable filter, and erosion filter were applied to the marble images in order.

Image blurring is achieved by convolving the image with a low-pass normalized box filter using a 5×5 kernel, which is useful for removing noise and distortion in images [17]. Since the original marble images were composed of a lot of pits, veins, and fossils, the blur filter helped us to create more smooth images. After that, a 2D linear separable filter [20] was applied to make vein patterns in marbles more visible. A Gaussian filter with a kernel size of 15 and a sigma value of 2 was used for the 2D linear separable filter to convolve the source image. In other words, a 15×2 Gaussian kernel was used to obtain each of the row and column coefficients, and then these coefficients were used as the parameters for the 5×5 low-pass 2D linear separable filter. All the other parameters were set to their default values in Python's image processing library. The desired depth of the destination image is equivalent to the source image, where the depth term indicates the precision of each pixel.

Finally, we used erosion with a 4×4 kernel [21] to make pits and fossils more distinguishable and visible. Dilation and erosion are two fundamental morphological processes. Dilation adds pixels to the boundaries of objects in an image, and erosion removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image [21].

After these data augmentation operations, the marble image dataset size was increased up to 6300, where 5400 samples were used for training and 900 samples were used for testing. Also, all the 6300 samples were used in additional experiments with 10-folds cross-validation method. The number of samples for each of the six different marble quality classes was 1050.

3 Results

The CNN model proposed in this study is comparatively analyzed with some other types of machine learning algorithms where the average classification accuracy scores over the marble images are given in Tables 1 and 2. The machine learning algorithms used in this study are shortly described in the next paragraph.

Naïve Bayes is a probabilistic classifier based on Bayes' theorem, and it assumes strong independence between features [22]. The probability of each categorical attribute

Table 1 Performance comparison of several classifier algorithms versus our CNN model using the original dataset with 2100 images

Algorithm name and parameters	Average accuracy (train)	Average accuracy (test)	Average accuracy (tenfolds cross-validation)
Our CNN model	1.000	0.714	0.701
k-NN	0.915	0.662	0.651
MLP 1	0.908	0.695	0.687
MLP 2	0.910	0.711	0.699
C4.5	0.877	0.672	0.625
Naïve Bayes	0.849	0.609	0.674

Table 2 Performance comparison of several classifier algorithms versus our cnn model using the augmented dataset with 6300 images

Algorithm name & parameters	Average accuracy (train)	Average accuracy (test)	Average accuracy (tenfolds cross-validation)
Our CNN model	1.000	0.922	0.961
k-NN	0.906	0.739	0.703
MLP 1	0.893	0.748	0.757
MLP 2	0.917	0.708	0.702
C4.5	0.827	0.684	0.685
Naïve Bayes	0.753	0.665	0.651

is calculated by using the frequency ratios, and the probabilities of continuous attributes are calculated by the probability density function based on Gaussian probability distribution. k-NN (k-nearest neighbors) is a type of instance-based learner that implements instances as vectors among a hyperspace [23]. A set of labeled training examples is used to predict the new example's label within the range of its nearest k neighbors that are identified using a similarity or distance measure such as Euclidean distance, Cosine similarity, and so on [23]. MLP (multilayer perceptron) is an artificial neural network with a feed-forward learning model that has one or more hidden layers and the differential error and updates of the weights are achieved by backpropagation with gradient descent methodology [24]. C4.5 is a type of decision tree algorithm that uses information gain scores for splitting and constructing the tree [25]. Decision tree induction constructs a flowchart-like structure where each non-leaf node denotes a test on an attribute, each branch corresponds to an outcome of the test, and each external leaf node denotes a class prediction [22]. C4.5 is implemented with various pruning methods like most of the other decision tree algorithms to minimize overfitting.

We used Weka software version 3.9.3 [26] for training and testing our dataset with the alternative machine learning algorithms mentioned in the previous paragraph. The

parameter settings for some of these machine learning algorithms were different than their default values to obtain the best accuracy performances for those algorithms by using the Auto-Weka tool [27, 28]. For instance, the highest accuracy scores by k-NN were obtained when k was set to 1 and Euclidean distance was used as the distance measure. The best accuracy scores for C4.5 Decision Tree and Naïve Bayes were observed when they were set to their default values in Weka. On the other hand, we experimented with two different MLP models with different parameter values automatically set by Auto-Weka. The highest accuracy values for one of these models, MLP 1 namely, were obtained when one hidden layer with 64 nodes was used. The learning rate was set to 0.1, and momentum was set to 0.2. The highest accuracy scores with the MLP 2 were obtained when two hidden layers with 256 and 128 nodes were used. The learning rate was set to 0.03, and momentum was set to 0.1. Sigmoid was used as the nonlinear activation function, and stochastic gradient descent was used [17, 24] for the updates of the weights in both of the MLP models.

Our CNN model was trained for 30 epochs using a batch size of 16 samples. Adam algorithm [17, 18] is used for adaptive learning rate optimization with an initial learning rate of 0.001. It should be noted that the results in Table 1 were obtained by using the original marble image samples where 1800 of them were used for training, 300 of them for testing, and all of them (2100 samples) for 10-folds cross-validation method.

As discussed in the previous section, the results obtained using only the original marble images were not satisfactory because of low accuracies in test and validation, as well as the overfitting problem, which can be seen in Table 1. However, when the augmented data set with 6300 images were used in the experiments, our CNN model provided outstanding average accuracy scores that can be observed in Table 2.

It should be mentioned that in some of the previous studies such as [9, 10], the accuracy scores are higher than the ones obtained by our model. However, the data types, conditions, and situations in these studies are different than our study. Hence, it is would not be an objective and acceptable approach to make a direct comparison between their results and our results. For instance, the limestone samples in the Manisa region were used in [9] and the distinctiveness of these samples is easier than the marble samples in our study due to the patterns and discernible areas in the limestone. Although their chemical structure is almost similar to each other, there are many differences between limestone and marble in terms of their origin and physical properties. Marble has a variety of colors compared to limestone, and marble slabs of different qualities are more difficult to distinguish

than limestone samples for human experts and computer systems. A similar issue is also true for the study in [10]. In that study, the dataset is composed of granite tiles. In contrast to the hardly distinguishable marble slab images in our study, granite tiles' images were used. It is known that granites contain more crystalline substances than marbles. Distortions in the shape of clumps are generally observed in granites. However, veins, fossils, or holes occur in marbles, which makes it much more difficult to classify. It could be seen in [10] that even by the human eye, each of the different granite classes is much easier to distinguish than the marble samples in our study.

4 Conclusion

In this study, we have proposed a novel approach and model for marble quality classification using marble slab images. The contributions of this study to the relevant literature can be summarized as follows; using a specific convolutional neural network for marble quality classification, the use of three image processing methods (Blur filter, 2D linear separable filter, and erosion filter) for data augmentation, and the observation of results with no overfitting and very high accuracy. The average accuracy scores obtained by our CNN model with augmented dataset were 0.922 for the test and 0.961 for 10-folds cross-validation, which are both much higher than the results obtained by other algorithms. The second-best accuracy scores were obtained by multilayer perceptron neural network model 0.748 for the test and 0.757 for 10-folds cross-validation), which were much lower than the results obtained by our CNN model. These results are considered being at least as good as (or even exceeding) the human experts' manual classification, which has also been confirmed by the experts and the executives of the "Haz Marble" company. It can be concluded that this proposed model can be implemented and used in marble companies as a fully automated alternative for manual operations carried out by humans.

One of our plans for future work is to implement this model as a product for industrial applications in the marble business. Our pre-trained CNN model and image pre-processing for data augmentation can be implemented in a single software package that can be executed on hardware devices with cameras or on smartphones. These devices will be placed at suitable locations within the marble operation benches that can take photographs of each new marble slab and automatically assign each marble to the correct quality class with no human intervention. It should be noted that since a pre-trained CNN with the marble images will be used in this system, there will not be any time lag for the training process; however,

there will be a need for image preprocessing operations in the live system. Hence, the image preprocessing must be implemented with highly efficient and fast algorithms to execute the entire automatic marble classification process in an acceptable amount of time for real-time business operations.

Acknowledgements The authors would like to thank the “Haz Marble Inc.” for supplying the marble slab images and valuable managerial feedback and comments throughout this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Karaca Z (2003) Quality control of marble blocks. In: Proceedings of MERSEM 2003 IV. Marble symposium, pp 497–503
- Yavuz AB, Türk N, Koca MY (2003) The use of micritic limestone as building stone: a case study of Akhisar beige marble in western Turkey. In: Proceedings of industrial minerals and building stones, pp 277–281
- Bianconi F et al (2012) Automatic classification of granite tiles through colour and texture features. *Expert Sys Appl* 39(12):11212–11218
- Selver MA et al (2011) An automated industrial conveyor belt system using image processing and hierarchical clustering for classifying marble slabs. *Robot Comp Integr Man* 27(1):164–176
- Unser M (1986) Sum and difference histograms for texture classification. *IEEE Trans Pattern Anal Mach Int* 8(1):118–125
- Martinez-Alajarin J, Luis-Delgado JD, Tomas-Balibrea LM (2005) Automatic system for quality based classification of marble textures. *IEEE Trans Sys Man Cyber Part C* 35(4):488–497
- Martinez-Alajarin J, Luis-Delgado JD, Tomas-Balibrea LM (2003) Classification of marble surfaces using wavelets. *Electron Lett* 39(9):714–715
- Doğan H, Akay O (2010) Using AdaBoost classifiers in a hierarchical framework for classifying surface images of marble slabs. *Expert Syst Appl* 37(12):8814–8821
- Selver MA et al (2009) Cascaded and hierarchical neural networks for classifying surface images of marble slabs. *IEEE Trans Syst Man Cyber Part C* 39(4):426–439
- Ferreira A, Giraldi G (2017) Convolutional neural network approaches to granite tiles classification. *Expert Syst App* 84:1–11
- Bianconi F et al (2015) On comparing colour spaces from a performance perspective: application to automated classification of polished natural stones. *New Trends Image Anal Process* 9281:71–78
- The MNIST Database of handwritten digits. <https://yann.lecun.com/exdb/mnist/>. Accessed 8 Nov 2019
- The CIFAR-10 dataset. <https://www.cs.toronto.edu/~kriz/cifar.html>. Accessed 21 Sep 2019
- Pençe İ, Çeşmeli MŞ (2019) Deep learning in marble slabs classification. *Techno-Science* 2(1):21–26
- He M et al (2019) Deep convolutional neural network for fast determination of the rock strength parameters using drilling data. *Int J Rock Mech Min Sci* 123(104084):1–9
- Hezaveh YD et al (2017) Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature* 548:555–561
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, London
- Buduma N, Locascio N (2017) *Fundamentals of deep learning: designing next-generation machine intelligence algorithms*. O’Reilly, USA
- Matin F et al (2018) Color image enhancement using multiscale retinex based on particle swarm optimization method. *J Phys Conf Ser* 960(1):012026–12036
- Smith SW (1997) *The scientist and engineer’s guide to digital signal processing*. Cal Tech Pub, Pasadena
- Maragos P (2000) *Morphological signal and image processing*. CRC Press, Boca Raton
- Han J, Pei J, Kamber M (2011) *Data mining: concepts and techniques*, 3rd edn. Morgan Kaufmann Pub, Burlington
- Aha DW et al (1991) Instance-based learning algorithms. *Mach Learn* 6(1):37–66
- Haykin S (2009) *Neural networks and learning machines*. Pearson Education Inc., New York
- Quinlan RJ (1993) *C4.5: programs for machine learning*. Morgan Kaufmann Pub, Burlington
- Weka: The workbench for machine learning. <http://www.cs.waikato.ac.nz/ml/weka/>. Accessed 15 Oct 2019
- Kotthoff L et al (2017) Auto-WEKA 2.0: automatic model selection and hyperparameter optimization in WEKA. *J Mach Learn Res* 18(25):1–5
- Auto-Weka. <https://www.cs.ubc.ca/labs/beta/Projects/>. Accessed 29 Oct 2019

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.