Classification of Paintings Authorship Using Convolutional Neural Network

Richardson Santiago Teles de Menezes
Department of Computer Engineering
and Automation
UFRN, Natal-RN, Brazil

Angelo Marcelino Cordeiro
Department of Computer Engineering
and Automation
UFRN, Natal-RN, Brazil

Rafael M. Magalhães
Federal University of Paraiba, Brazil
UFPB, Brazil

Helton Maia
School of Science and Technology
UFRN, Natal-RN, Brazil
helton.maia@ufrn.br

***Abstract*—In this paper, state-of-the-art architectures of Convolutional Neural Networks (CNNs) are explained and compared concerning authorship classification of famous paintings. The chosen CNNs architectures were VGG-16, VGG-19, Residual Neural Networks (ResNet), and Xception. The used dataset is available on the website Kaggle, under the title "Best Artworks of All Time". Weighted classes for each artist with more than 200 paintings present in the dataset were created to represent and classify each artist's style. The performed experiments resulted in an accuracy of up to 95% for the Xception architecture with an average F1-score of 0.87, 92% of accuracy with an average F1-score of 0.83 for the ResNet in its 50-layer configuration, while both of the VGG architectures did not present satisfactory results for the same amount of epochs, achieving at most 60% of accuracy.**

I. Introduction

Artistic creation is amongst the highest forms of human expression and imagination. The ability to communicate our vision sets us apart from all other beings. Painting, being an expression of visual language, has attracted and connected brilliant human minds since the beginning of civilization. After a long way, a stage has finally been reached where not only humans but also computers, another brilliant creation of human minds, are creating paintings [1]–[3].

With its efficiency in identifying patterns, the human brain can quickly identify abstract concepts such as style, trait, and intensity, which are as related and unique to the body of the artist's work as a kind of personal signature. Artificial neural networks are mathematical structures that can learn and later recognize patterns [4]. In this way, this paper aims to use convolutional neural networks as a technique for identifying and classifying the authorship of artists in different images of their paintings [5], [6].

This work used the dataset called *Best Works of All Time* [7] available on the Kaggle platform, which is composed 2.16 GB of separate data on paintings and artists. This dataset features paintings by the 50 most influential artists of all time, according to Icaro (2019) [7]. Names like Frida Khalo, Andy Warhol, Claude Monet, Francisco Goya, among others, are present in this dataset.

In this context, this paper aims to further compare and discuss differences in the application and design of state-of-the-art CNNs architectures, such as VGG [8], Residual Neural Networks (ResNets) [9], and Xception [10]. Applying them in solving authorship classification: given a set of paintings of some artist, can the CNNs correctly identify the paintings authorship based on the author's style and overall feel of their paintings?

This paper is structured as follows. In Section II, we describe the research method characterization. In Section III, we present the results and discussion. Finally, in Section IV, we present the conclusions.

II. METHODS

The approach chosen in this work was to compare and evaluate the performance of transfer learning with different architectures of convolutional neural networks, pre-trained in robust databases such as ImageNet [11].

It is essential to highlight that collecting and labeling data typically requires a lot of time and energy cost. One solution is to apply transfer learning, where a pre-trained model from some database that has characteristics similar to the object of interest, such as texture, border, and shape, will reduce the training time and may even improve the generalization [12], [13].

Therefore, pre-trained models are advantageous for several reasons, mainly in time savings. Some models used spent weeks being trained in cutting-edge hardware. Still, as their results are only weights and biases, these results become very accessible and applicable to other works [13].

This section goes on to present the overall computational development and specify the three models of convolutional neural networks used.

A. Computation Development

All the necessary steps in this work are described in the UML activity diagram shown in Figure 1.

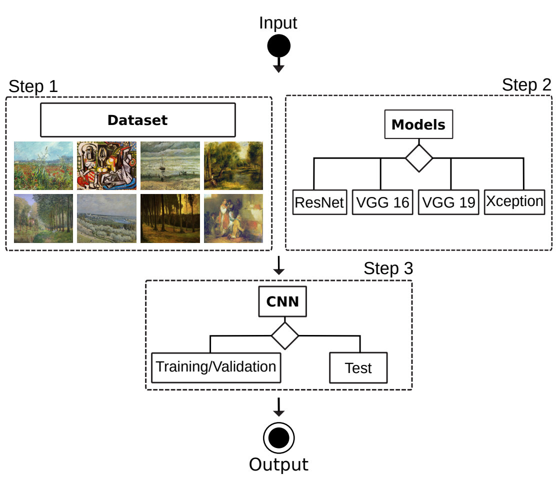

Fig. 1. UML diagram of the system developed to classify the paintings authorship.

- Step 1: The dataset *Best Works of All Time* [7] is a collection of artworks of 50 of the most influential artists of all time under the CC BY-NC-SA 4.0 license. For the analysis of this work, the dataset was sampled to 11 artists with more than 200 indexed works to have enough data to train the classification algorithms. Table I shows the artists that have been selected, and to balance the classes, a weight feature has been added so that authors with fewer paintings have the same representation in the training phase. This metric can be calculated as shown in Equation 1.

$$Weight = \frac{T_P}{T_C \cdot T_{AP}} \tag{1}$$

where, $T_P = 4299$ is the total number of paintings, T_{AP} is the number of paintings by a specific author, and $T_C = 11$ is the total number of artists in the dataset.

TABLE I
DATASET EXAMPLE DISTRIBUTION
WITH THE RESPECTIVE BALANCING FACTOR.

Artist Name	**# Paintings**	**Class weight**
Vincent van Gogh	877	0.445631
Edgar Degas	702	0.556721
Pablo Picasso	439	0.890246
Pierre-Auguste Renoir	336	1.163149
Albrecht Dürer	328	1.191519
Paul Gauguin	311	1.256650
Francisco Goya	291	1.343018
Rembrandt	262	1.491672
Alfred Sisley	259	1.508951
Titian	255	1.532620
Marc Chagall	239	1.635223
Total	4299	-

The values obtained from the Equation 1 are used to weight the loss function during the training phase, which effectively makes the loss function a weighted average of each sample and its corresponding class weight [14]. In other words this technique is used to tell the model to "pay more attention" to samples from an under-represented class.

- Step 2: Selection a state-of-the-art image classifications algorithm amongst the three here analysed, namely, ResNet [9], VGG [8], and Xception [10]. A series of comparative tests are going to be performed to show the advantages and disadvantages of each technique.
The comparisons presented in the paper are going to really on classifications metrics such as precision which represents the relationship between positive predictions

correctly predicted and the total observations, recall that is the proportion of positive observations correctly predicted for all observations, and the F1 score defined as the weighted average of accuracy and recall, so the maximum value assigned to it is 1, which means a perfect relationship between precision and recall.

- Step 3: The dataset containing the images was randomly divided into three parts, known as the training, testing, and validation sets, using the following proportions: 60% for the training set amounting to 2579 images, 20% for validation amounting to 865 images, and finally, the remaining 855 images were allocated in the test set.

B. ResNet

ResNet [9], short for Residual Networks, is a neural network used primarily for computer vision tasks. This model was the winner of the 2015 ImageNet challenge [11], which consists of a software competition to classify and detect objects in scenes. The main innovation of this network was that it allows to successfully train deep neural networks with more than 150 layers, something that previously suffered a lot with the famous problem of the vanishing gradient problem.

Previous ImageNet winners, such as GoogleNet [15] in 2014, had significantly fewer layers, being the largest so far, with 22 layers. However, working with multiple layers for deep neural networks is not as trivial as simply adding n-layers to the network and hoping for the best. Differently, deep networks suffer from the notorious problem of the vanishing gradient problem. As the gradient is propagated through the layers, operations tend to decrease its value considerably. Thus, the deeper the network, the faster its performance is saturated, or degradation begins to be perceived.

The new concept introduced by the ResNet was a “shortcut”, or “skip connections”, which allows the gradient to be propagated directly to the layers that suffered from the mentioned problem. The implementation is relatively simple, with layers that were previously sequential, now have shortcuts between them as shown in Figure 2.

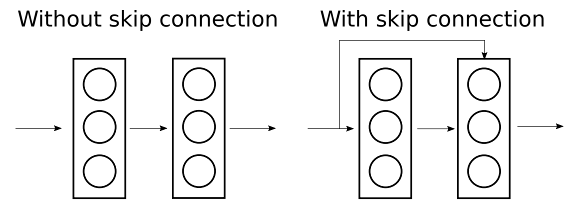

Fig. 2. Diagram of the shortcut concept introduced by ResNet [16].

More specifically, the model addressed in this work is the ResNet-50 variant, a more compact version of the ResNet 152. This model consists of 5 stages with a 3-layer block of convolutions and an identity block, also with three convolutional layers. ResNet has more than 23 million trainable parameters.

C. VGG

VGG [8] is a model of deep convolutional neural networks for large-scale image recognition, developed by the Visual Geometry Group (VGG). Like ResNet, it is a model considered state-of-the-art that participated and won the 2014 ImageNet competition.

According to Wei, 2019 [17], the VGG model is an evolution of the famous AlexNet [18] model, focused on convolutional layers of small windows (3×3). The input is an RGB image of fixed size (224×224), which is passed to a series of convolutional layers. Next some of these layers, we have a 2×2 max-pooling layer, with step 2. Finally, three layers are fully connected, the first with 4096 channels, and the last one with 1000, allowing the classification of up to 1000 classes. The complete architectural diagram can be seen in Figure 3.

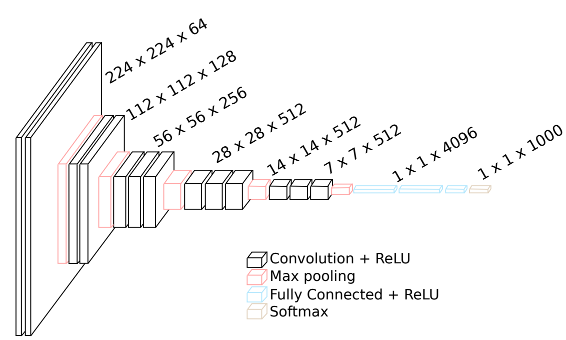

Fig. 3. VGG architecture diagram [19].

There is another variant of the architecture shown in the

Figure 3, with 19 hidden layers instead of 16, the diagram in Figure 4 illustrates such architecture. The models are called VGG-16, for the model with 16 layers and VGG-19 for 19 layers.

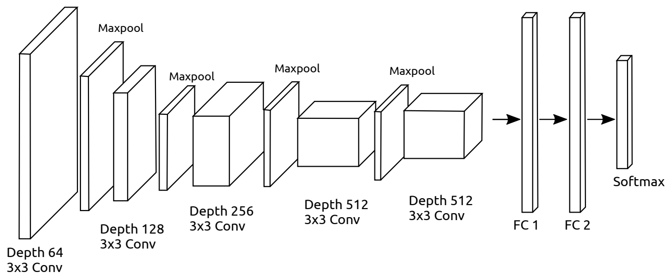

Fig. 4. VGG-19 architecture diagram [20].

D. Xception

The Xception [10] architecture published in 2016, is result of the evolution in the scenario of convolutional neural networks, being developed by Google, also based on the architecture Inception-V3 [21].

This architecture uses a modified depthwise separable convolution technique, which separates the image channels and applies specific convolutional filters for each one, flowed by a point-to-point convolution. That is, a 1×1 convolutional filter is applied after the deep convolution. The Figure 5 illustrates this convolutional process.

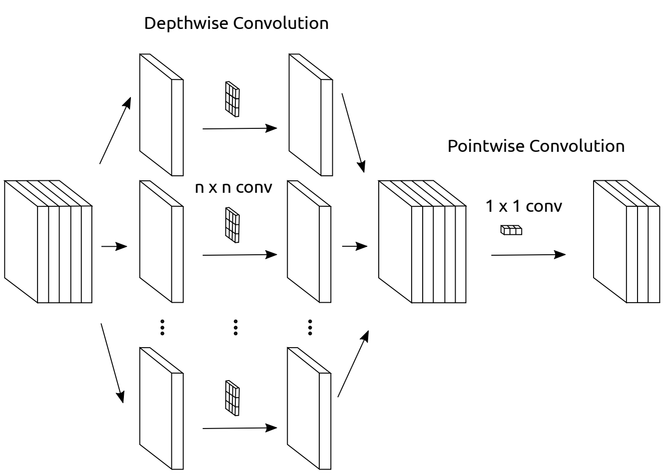

Fig. 5. Depthwise separable convolution [22].

In addition, its most striking feature is the presence of residues, the same concept introduced by ResNet [9], in its layers of separable convolutions. The presence of these modifications is aimed at reducing the problem of the disappearance of the gradient.

III. Results and Discussion

The approach presented in this work used four state-of-the-art image classification algorithms to define the authorship of paintings by 11 artists belonging to different artistic movements. The results here present were based on the analysis of 4299 images, organized according to the dataset described in Section II-A.

ResNet50: The first network trained was ResNet50 [9], which, as the name suggests, has an architecture composed of 50 convolutional layers. This model was trained using transfer learning from a pre-trained model in the Imagenet [23] dataset. The training process for this network was divided into two stages to maximize the benefits of using transfer learning.

Initially, all layers were trained, and their weights were updated in a typical process of training convolutional networks that lasted 10 epochs. For the next training stage, the weights referring to the convolutional layers originating from ResNet50 were frozen, and only the totally connected network added to the end of the convolutional layers continued to be trained for another 50 epochs, the Adam [24] optimizer was chosen due to its computational efficiency and small memory requirements, a initial learning rate of 0.0001 was set and a callback function was used to reduce this value by a factor of 0.1 if there were no improvement to the validation loss on the last 5 epochs.

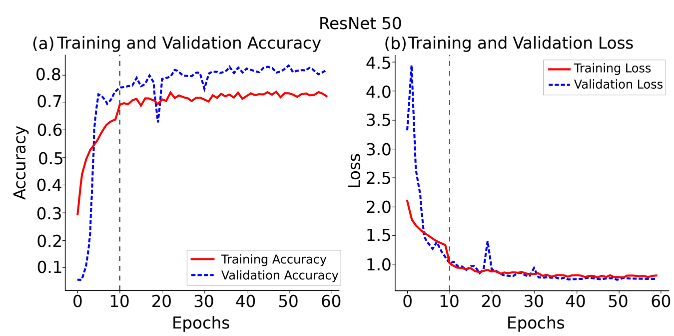

Fig. 6. Optimization of accuracy and minimization of the loss function during the training process of the ResNet50 network.

This training process has been designed to make it possible to freeze the already optimized feature extractor represented by the convolutional network while the fully connected classifier continues to learn and generalize the features that have been extracted.

Figure 6 shows the precision of the ResNet50 architecture during the training process with small oscillations in its precision curve. The loss function was quickly optimized during this process, successfully leaving local minimums and approaching zero after the training.

Table II shows the classification metrics for ResNet50 when classifying the test set. These metrics are used to assess the quality of the classifier output. Also, the high rates in the metrics for most classes can be verified, obtaining a good average of precision, recall, and f1-score, the latter with 0.83, which indicates a good relationship between precision and recall for all classes.

TABLE II
CLASSIFICATION METRICS IN THE TEST SET FOR THE RESNET50 ARCHITECTURE.

Artist	**Precision (%)**	**Recall (%)**	**F1-score**
Vincent van Gogh	89	69	0.78
Edgar Degas	92	88	0.90
Pablo Picasso	76	*73*	*0.75*
Pierre Renoir	**93**	85	0.89
Albrecht Dürer	85	**98**	0.91
Paul Gauguin	81	84	0.83
Francisco Goya	*72*	81	0.76
Rembrandt	82	87	0.84
Alfred Sisley	64	90	*0.75*
Titian	*72*	84	0.77
Marc Chagall	88	96	**0.92**
Average	**81**	**85**	**0.83**

To conclude the analysis of the ResNet50 results, Figure 7 shows the confusion matrix generated from the network predictions in the test data set. As expected from the high score of the metrics, the classifier made relatively few prediction errors reaching the correct classification of 99% of Albrecht Dürer's works, and as further evidence of the generalization achieved by the network, it still manage to correctly classify 73% of the paintings in the worst performing class.

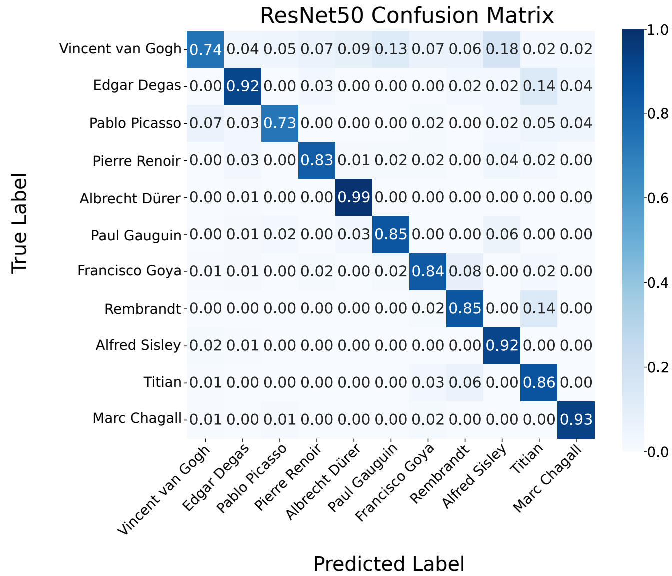

Fig. 7. Confusion matrix generated with the ResNet50 predictions.

VGG: This stage of the experiments used two variations of the VGG [8] architecture that differs in the number of convolutional layers applied to the input, so the VGG-16 version uses 16 convolutional layers, while the VGG-19 makes use of 19. The same strategy of training previously described was used, adapting the frozen weights depending on the number of convolutional layers in the respective architecture.

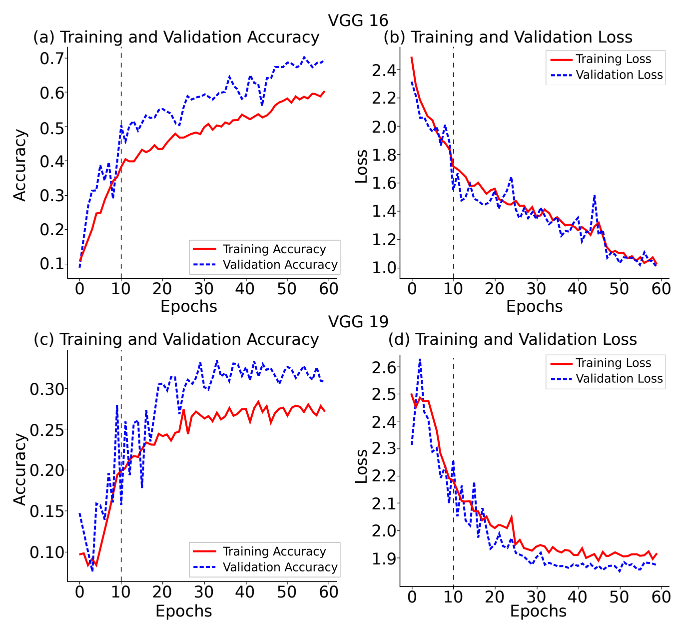

Fig. 8. Optimization of accuracy and minimization of the loss function during the training process of the VGG architectures.

Figure 8 illustrates the training process described. Both networks did not achieve satisfactory performance, and the network with fewer convolutional layers was surprisingly doing better during training but obtaining only 60% accuracy. However, in both networks, the validation accuracy responded better to the training obtaining a higher precision in both cases, indicating that a longer training could make this architecture perform better, a hypothesis reinforced by the continuous decline of the loss function.

The under-performance of the VGG architecture can be confirmed by the confusion matrices shown in Figure 9, where the errors in the classifications for both architectures are clear. It is worth highlighting the inferior performance of the VGG-19 for authors like Vincent van Gogh and Edgar Degas, where the network got very little or no hits.

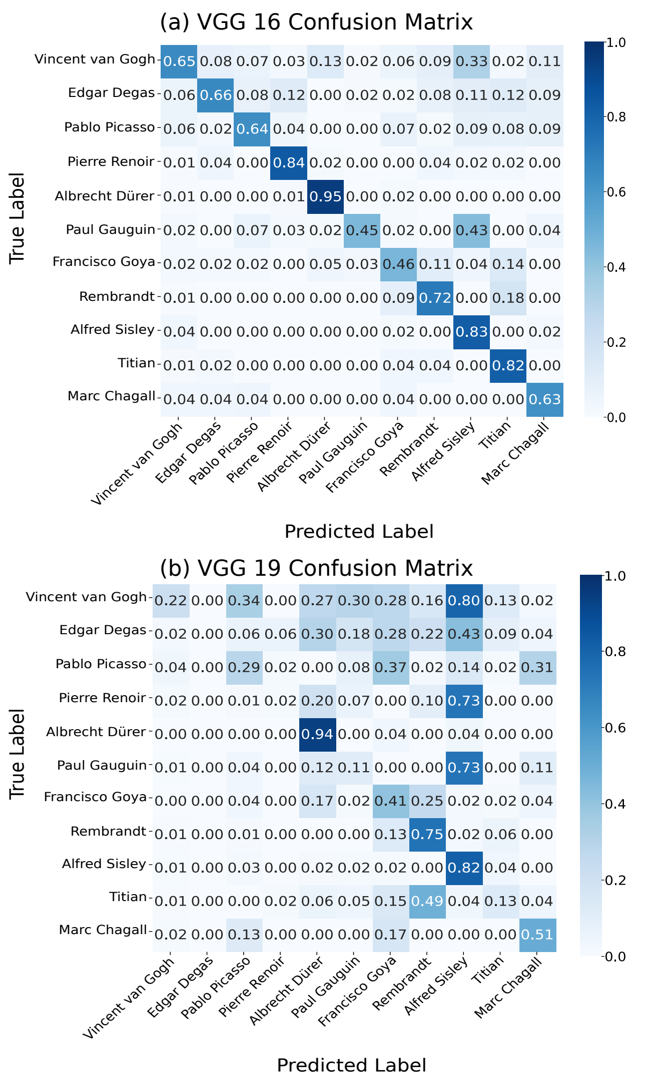

Fig. 9. Confusion matrix generated with the predictions of both VGG architectures.

Convolutional networks with more convolutional layers suggest that its feature extractor would be capable of extracting more characteristics from the inputs as compared to smaller networks [25], however, still on Figure 9, the confusion matrices further confirms the surprising result that the VGG-16 achieves better classification results despite theoretically being capable of less feature extraction. And, as stated previously, the curve of its loss function seems to have been cut-short from its convergence, suggesting that expanding the time-variable on the training process should yield even better results for both architectures.

Xception: Finally, the Xception [10] network was trained, which is a convolutional neural network architecture that depends exclusively on depthwise separable convolution layers. Again, the training strategy used in the other architectures was repeated this time, freezing the first 50 convolutional layers.

Figure 10 illustrates the training process of the architecture where once again the validation accuracy can be seen growing faster, proving the efficiency of the pre-trained features extractor. However, the training accuracy has reached higher levels in the analyzed training interval.

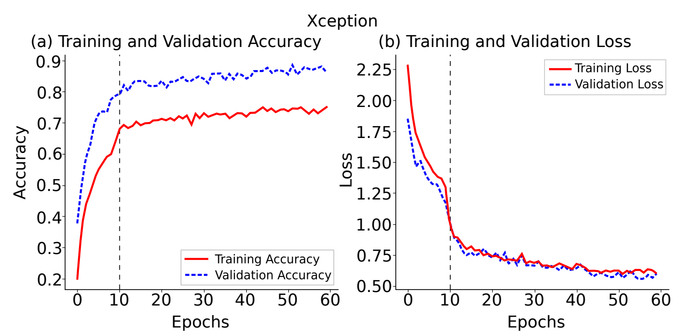

Fig. 10. Optimization of accuracy and minimization of the loss function during the training process of the Xception network.

The efficiency of this architecture using the dataset is shown through the metrics in Table III. High rates in the metrics in the metrics can be seen for all classes, scoring an average of precision of 0.87, which is the best result encountered during the experiments.

TABLE III
CLASSIFICATION METRICS IN THE TEST SET FOR THE XCEPTION ARCHITECTURE.

Artist	Precision (%)	Recall(%)	F1-score
Vincent van Gogh	85	92	0.88
Edgar Degas	92	91	0.91
Pablo Picasso	90	*71*	*0.79*
Pierre Renoir	84	89	0.87
Albrecht Dürer	**93**	**98**	**0.95**
Paul Gauguin	86	84	0.85
Francisco Goya	91	74	0.82
Rembrandt	85	87	0.86
Alfred Sisley	82	94	0.88
Titian	*81*	84	0.82
Marc Chagall	**93**	89	0.91
Average	**87**	**87**	**0.78**

Figure 11 shows the confusion matrix generated from the classifications in the test set, and as indicated by the metrics shown previously in the network, had an excellent performance in the task of classifying the dataset presented, being able to classify almost all the examples correctly.

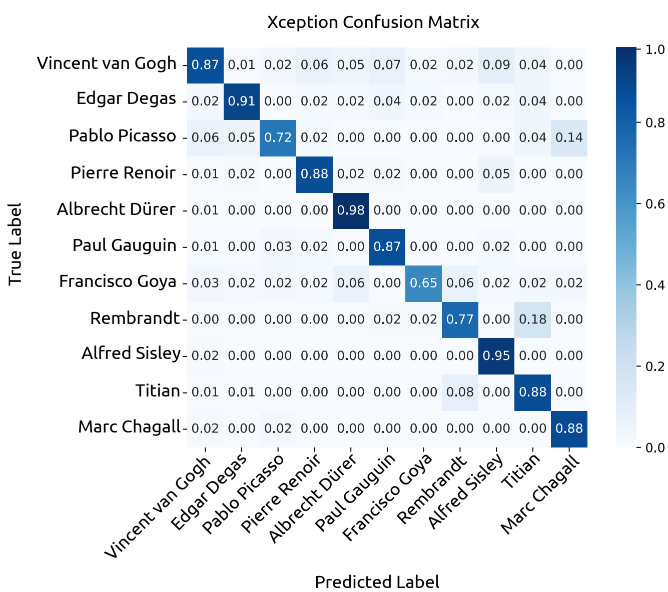

Fig. 11. Confusion matrix generated from the predictions of the Xception architecture.

To exemplify the application of the classifiers presented in this work, Figure 12 shows examples of classification using the Xception architecture, this being the approach that presented the best performance in the tests performed. In this example, it is clear the great performance of this classifier, correctly predicting all authorship with good confidence.

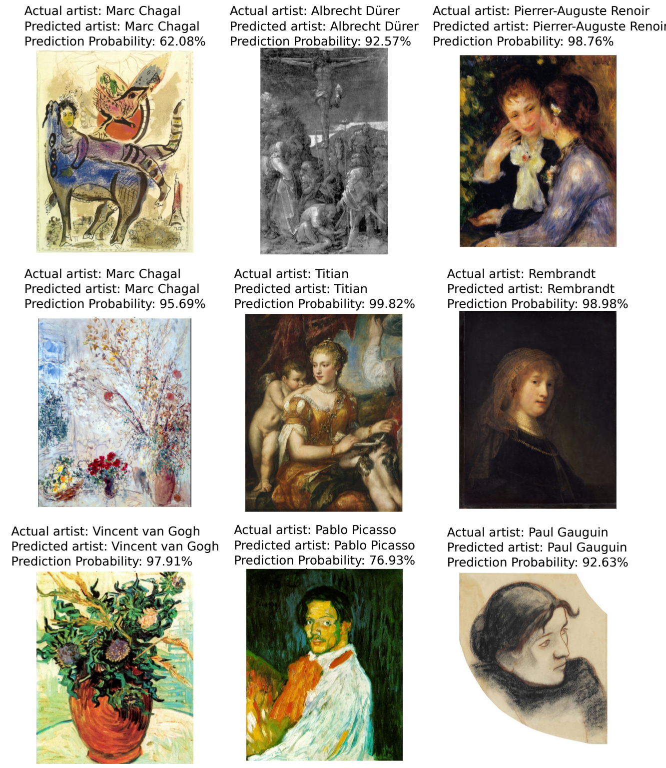

Fig. 12. Output examples of the Xception network.

IV. CONCLUSION

From the results obtained in this work, it is possible to observe the difficulties and advantages of using convolutional neural network architectures already present in the literature outside of their original sample space. The use of these architectures is advantageous when the time for development is taken into account since all the architectures used were exceptional in their specific tasks and have already been extensively studied. However, it was also seen that this is not a guarantee that they will perform well in any task. Although image classification is a specific area of problem solving, each application that falls into said area has its set of difficulties and characteristics that some well-established state-of-the-art architectures may not excel in.

During the experiments described in this paper, all networks were submitted to the same training conditions in order to compare them in solving the proposed task, without favouring any specific network. However, when thoroughly analyzing the construction of each of the architectures, it was seen that perhaps the VGG architectures used would have benefited from a longer training cycle since its original training lasted for weeks [17], whilst the Xception architecture excelled, achieving up to 95% of accuracy. This also indicates that the latter network can perform well on tasks with limited training time. The ResNet configuration also demonstrated its problem-solving generalization, achieving 92% of accuracy in this application of image classification, even though the application differs from object detection, the problem area that awarded this architecture the 2015 ImageNet challenge [11].

Finally, this work explored the adaptability of prominent convolutional neural networks architectures present in the literature when applied to a different problem domain. These types of explorations demonstrate how intense the development scenario in this area is, and that the current scenario is far from being a optimized, and it is full of evolution and new knowledge.

Acknowledgment

This work was supported by the School of Sciences and Technology at the Federal University of Rio Grande do Norte (ECT-UFRN).

References

[1] D. Lopes, *A philosophy of computer art*. Routledge, 2009.

[2] M. A. Boden, "Computer models of creativity," *AI Magazine*, vol. 30, no. 3, pp. 23–23, 2009.

[3] K. Zhang, S. Harrell, and X. Ji, "Computational aesthetics: on the complexity of computer-generated paintings," *Leonardo*, vol. 45, no. 3, pp. 243–248, 2012.

[4] R. S. T. de Menezes, R. M. Magalhaes, and H. Maia, "Object recognition using convolutional neural networks," in *Recent Trends in Artificial Neural Networks-from Training to Prediction*. IntechOpen, 2019.

[5] A. Puthenputhussery, Q. Liu, and C. Liu, "Color multi-fusion fisher vector feature for fine art painting categorization and influence analysis," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–9.

[6] M. Čuljak, B. Mikuš, K. Jež, and S. Hadjić, "Classification of art paintings by genre," in *2011 Proceedings of the 34th International Convention MIPRO*. IEEE, 2011, pp. 1634–1639.

[7] Icaro, "Best artworks of all time," Mar 2019. [Online]. Available: https://www.kaggle.com/ikarus777/best-artworks-of-all-time

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions, 2016," *arXiv preprint arXiv:1610.02357*, 2016.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[12] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*, 2020, https://d2l.ai.

[13] P. Marcelino, "Transfer learning from pre-trained models," *Towards Data Science*, 2018.

[14] G. King and L. Zeng, "Logistic regression in rare events data," *Political analysis*, vol. 9, no. 2, pp. 137–163, 2001.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Computer Vision and Pattern Recognition (CVPR)*, 2015. [Online]. Available: http://arxiv.org/abs/1409.4842

[16] P. Dwivedi, "Understanding and coding a resnet in keras," Mar 2019. [Online]. Available: https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33

[17] J. Wei, "Vgg neural networks: The next step after alexnet," Jul 2019. [Online]. Available: https://towardsdatascience.com/vgg-neural-networks-the-next-step-after-alexnet-3f91fa9ffe2c

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[19] "A brief report of the heuritech deep learning meetup," Feb 2016. [Online]. Available: https://heuritech.wordpress.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/

[20] R. Hewage, "Extract features, visualize filters and feature maps in vgg16 and vgg19 cnn models," May 2020. [Online]. Available: https://towardsdatascience.com/extract-features-visualize-filters-and-feature-maps-in-vgg16-and-vgg19-cnn-models-d2da6333edd0

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[22] S.-H. Tsang, "Review: Xception - with depthwise separable convolution, better than inception-v3," Mar 2019. [Online]. Available: https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] M. Jogin, M. Madhulika, G. Divya, R. Meghana, S. Apoorva *et al.*, "Feature extraction using convolution neural networks (cnn) and deep learning," in *2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*. IEEE, 2018, pp. 2319–2323.