

Evaluation of CNN Models with Transfer Learning in Art Media Classification in Terms of Accuracy and Class Relationship

Juan Manuel Fortuna-Cervantes¹, Carlos Soubervielle-Montalvo^{*,2},
Cesar Augusto Puente-Montejano², Oscar Ernesto Pérez-Cham³,
Rafael Peña-Gallardo²

¹ Instituto Tecnológico de San Luis Potosí,
Tecnológico Nacional de México,
Mexico

² Universidad Autónoma de San Luis Potosí,
Facultad de Ingeniería,
Mexico

³ Universidad del Mar,
Instituto de Industrias,
Mexico

juan.fc@slp.tecnm.mx, {carlos.soubervielle, cesar.puente,
rafael.pena}@uaslp.mx, operezcham@zicatela.umar.mx

Abstract. The accuracy obtained in Art Media Classification (AMC) using CNN is lower compared to other image classification problems, where the acceptable accuracy ranges from 90 to 99%. This article presents an analysis of the performance of three different CNNs with transfer learning for AMC, to answer the question of what challenges arise in this application. We proposed the Art Media Dataset (ArtMD) to train three CNNs. ArtMD contains five classes of art: Drawing, Engraving, Iconography, Painting, and Sculpture. The analysis of the results demonstrates that all the tested CNNs exhibit similar behavior. Drawing, Engraving, and Painting had the highest relationship, showing a strong relationship between Drawing and Engraving. We implemented two more experiments, removing first Drawing and then Engraving. The best performance with 86% accuracy was achieved by removing Drawing. Analysis of the confusion matrix of the three experiments for each CNN confirms that Drawing and Painting have the lowest accuracy, showing a strong misclassification with the other classes. This analysis presents the degree of relationship between the three CNN models and details the challenges of AMC.

Keywords. Art media classification, convolutional neural networks, transfer learning.

1 Introduction

Art restorers and collectors frequently classify art media by evaluating their physical features, subjective characteristics, and historical periods [16]. However, this classification process can be challenging because specific attributes may need to fit neatly into predefined styles, genres, or art periods, leading to potential misclassification.

A favorable solution to this challenge involves the utilization of Convolutional Neural Networks (CNNs). These deep learning algorithms have garnered recognition in the scientific community for their prowess in image classification and object detection tasks [2, 17, 22].

Although there is growing interest in CNNs for Art Media Classification (AMC), limited research delves deeply into their classification performance and class relationship [12, 20]. Furthermore, there is a growing inclination towards pre-trained models over traditional computer vision methods, demonstrating the potential for achieving more accurate dataset classification [7].

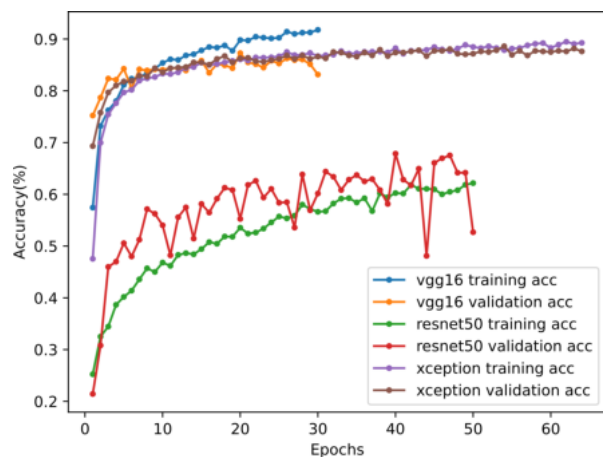


Fig. 1. Image distribution and composition for the Art Media Dataset (ArtMD)

In a primary study, serving as a basis for this work [8], an assessment of the performance accuracy was conducted on three well-established CNN architectures in AMC.

The principal objective is to significantly emphasize the resilience of CNN learning models in art media classification when leveraging transfer learning. This study of the three proposed CNN architectures seeks to determine the optimal choice for future applications.

Based on the insights gained from previous work, this study presents a comprehensive evaluation and performance analysis of three well-known CNN architectures in the context of AMC, aiming to address the challenges that arise when using CNNs with transfer learning [14].

In addition, it investigates the relationship between classes to shed light on poor classification performance and how dataset characteristics influence CNN learning. The main contributions of this study are as follows:

1) Introduction of an experimental approach to evaluate CNN performance in the Art Media Classification (AMC) context and to demonstrate that AMC represents a problem in the accuracy of the classifier, being an area of opportunity in the development of CNN.

2) Creation of the Art Media Dataset (ArtMD), used for training and evaluating the classification model.

The dataset combines digitized artworks sourced from diverse repositories, including the Kaggle website, the WikiArt database, and institutional archives from the Prado National and the Louvre National Museum. The proposal can be considered a standard for evaluating CNN models in AMC.

3) Evaluation of three state-of-the-art CNN models in AMC highlights that accurate inferences can be drawn for most classes of art media, with a notable finding that Drawing and Engraving exhibit a strong relationship with each other.

4) Conducting additional experiments by removing Drawing and Engraving, which accentuates a slight relationship with the Painting class across all remaining classes (Iconography, Sculpture, and Engraving).

Furthermore, a high relationship is observed between the predicted class and the original label for Iconography and Sculpture classes. These relationship effects can be seen in these experiments for all CNN models, as presented in the Experiments and Results section. This article unfolds as follows: Section 2 briefly overviews the work related to AMC.

Section 3 delves into the materials and methods. Section 4 contains experimental details, presents results, and analyzes the classification outcomes. We showcase the accuracy and interclass relationships of the devised image classifiers, which remain unexplored in the current state-of-the-art. Finally, in Section 5, we end the paper with some conclusions and ideas for future work.

2 Related Work

Computer vision has become an intriguing approach for recognizing and categorizing objects across various applications. It is an auxiliary tool that mimics human visual perception, opening doors to various practical applications. One of these applications pertains to safeguarding data against adversarial attacks. Deep Genetic Programming (DGP) employs a hierarchical structure inspired by the brain's behavior to extract image features and explore the transfer of adversarial attacks within artwork databases.



Fig. 2. Art Image training set exhibits the five art categories: (a) Drawings produced using a pencil, pen, or similar tools on paper or another medium; (b) Engravings, images crafted through cutting or etching into a surface; (c) Iconography, encompassing religious images or symbols; (d) Paintings, artworks generated by applying pigments onto a surface; and (e) Sculptures, representing three-dimensional art forms shaped or modeled from materials to achieve a specific form

In this context, the application focuses on adversarial attacks in categorization [20]. The paper [11] presents a comparative study on the impact of these attacks within the art genre categorization, involving feature analysis and testing with four Convolutional Neural Networks (AlexNet, VGG, ResNet, ResNet101) alongside brain-inspired programming.

Deep learning algorithms have significantly advanced image classification, particularly in [18], where pre-trained networks like VGG16, ResNet18, ResNet50, GoogleNet, MobileNet, and AlexNet are utilized on the Best Artworks of all Time dataset.

After adjusting training parameters, the study selects the best model, finding that ResNet50 achieves the highest accuracy among all other deep networks.

In [15], the focus shifts to style classification using the Painter by Numbers dataset, encompassing five classes: impressionism, realism, expressionism, post-impressionism, and romanticism. The model is based on a pre-trained ResNet architecture from the ImageNet dataset and is refined by different transformations, such as random affine transform, crop, flip, color fluctuations and normalization.

Additionally, the papers [6, 5] explore further the correlation between feature maps, which effectively describe the texture of the images. These correlations are transformed into style vectors, surpassing the performance of CNN features from fully connected layers and other state-of-the-art deep representations.

Furthermore, the introduction of inter-layer correlations is proposed to enhance classification efficiency. In [21], a novel approach is presented to improve the classification accuracy of fine art paintings. This approach combines transfer learning with subregion classification, utilizing the weighted sum of individual patch classifications to obtain the final statistical label for a given painting.

The method offers computational efficiency and is validated using standard artwork classification datasets with six pre-trained CNN models. Further, [1] employs two machine learning algorithms on an artwork dataset to demonstrate that features derived from the artwork play a significant role in accurate genre classification.

These features encompass information about the nationality of the artists and the era in which they worked. Finally, in [9], VGG19 and ResNet50 are applied to classify artworks based on their style. The study compares their performance in recognizing underlying features, including aesthetic elements.

The dataset is derived from The Best Artworks in the World, selecting five subsets from artists with distinct styles. The results indicate that CNNs can effectively extract and learn these underlying features, with VGG19 showing preference for subjective items and ResNet50 with favoring objective markers. In summary, our work has two main differences from related works: Firstly, this work presents an in-depth study of CNN models in AMC, which can be used to

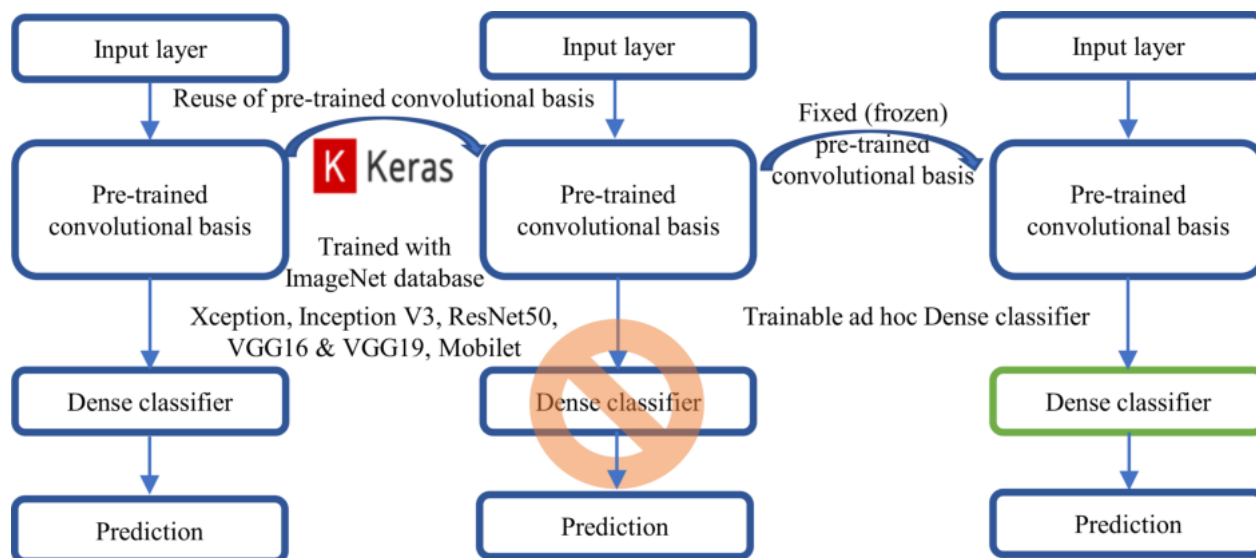


Fig. 3. Process for the reuse of convolutional pre-trained networks (feature-based transfer learning)

understand the difficulties in this task and find new alternatives to improve the performance. Secondly, a detailed analysis of accuracy and class relationship is presented using a proposed dataset consisting of the Art Image dataset, the WikiArt database, and digital artworks from The Louvre and Prado Museums.

3 Materials and Methods

3.1 Dataset

Information is the paramount element in deep learning tasks, particularly in the Art Media Classification (AMC) domain. The Art Image dataset [20] assumes significance. This dataset includes training and validation images sourced from the Kaggle website's repository of digitized artworks.

The dataset contains five art media categories: Drawing, Painting, Iconography, Engraving, and Sculpture. We opted to formulate the Art Media Dataset (ArtMD)¹, as illustrated in Fig. 1. This decision was prompted by the existence of corrupted or preprocessed images within the original dataset.

¹github.com/JanManuell/Art-Media-Classification---Dataset.git

The dataset consists of the same five classes, each comprising 850 images for training and 180 for validation, originating from the Art Image dataset. For the test set, 180 images per category were curated from the WikiArt database and digital artworks from the Louvre National Museum² for Painting and the Prado National Museum³ for Engraving. A notable characteristic of this dataset is the RGB format, each with a size of 224×224 , ideal for the input requirements of the proposed architectures. Fig. 2 showcases a selection of random images from the training set.

3.2 CNN Architecture and Transfer Learning

Several Convolutional Neural Network (CNN) architectures are available for addressing real-world challenges associated with image classification, detection, and segmentation [3, 10, 24]. However, each architecture has distinct advantages and limitations concerning training and implementation. Choosing the most suitable architecture involves experimentation and relies on the specific performance requirements and intended application.

²collections.louvre.fr/en/

³www.museodelprado.es/coleccion/obras-de-arte



Fig. 4. Process to improve the classification model

When trading with limited datasets in deep learning, Transfer Learning emerges as a popular approach [4]. The idea behind Transfer Learning is that a Convolutional Neural Network (CNN) previously trained on a large and diverse dataset, such as ImageNet, has already acquired knowledge about general and useful features present in the images, such as edges, textures, and shapes.

These features can be reused in a specific task without the need to train a network from scratch. The CNN architecture proposed contains two elements: the feature extraction stage and the classification stage. Feature extraction involves the use of previously learned representations during the original training.

The pre-trained network is taken in this stage, and the output layers designed for the original task are removed. The convolutional layers in charge of feature extraction are retained, which will process the images of the new task.

Then, in the classification stage, additional layers, such as fully connected and output layers, are added at the end of the network to adapt it to the new features of the specific dataset (feature-based transfer learning). After that, the complete network is trained with the dataset, and its performance is evaluated using task-relevant metrics, as shown in Fig. 3.

3.3 Improving Model Classification

The proposed methodology for improving the learning model's performance can be summarized in three key stages. In the first stage, the integration of the dataset is carried out.

It is essential that this dataset presents a balance between classes and contains images representative of the problem being addressed. In the second stage, the images are processed. The pixel values are normalized to ensure that the model converges efficiently during training. The third stage focuses on model validation. In this stage, the training parameters are adjusted and updated, allowing the learning model to be retrained to perform better, as shown in Fig. 4.

3.3.1 Model Evaluation

The model's classification accuracy improvement process involves iterative testing, selecting initial training parameters, and automatic feature extraction through optimal kernel filters. This enables subsequent model adjustments. Evaluation relies on Accuracy, measuring the percentage of correct predictions, while the confusion matrix, an $N \times N$ table (N being the number of classes), analyzes patterns of prediction errors by revealing the relationship between predicted and actual labels.

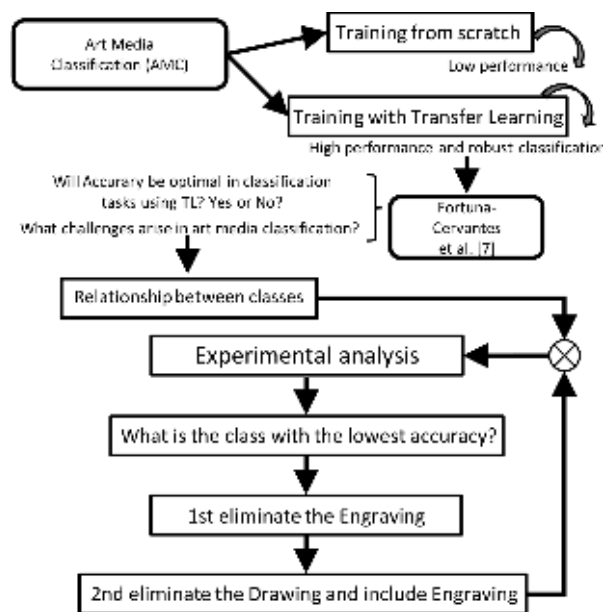
3.3.2 Network Training and Parameter Settings

The models are implemented using the Python programming language and the Keras API with Tensorflow as the backend. The training was conducted utilizing an NVIDIA Tesla K80 GPU within the Google Colaboratory⁴ (Colab) environment. Colab's GPU, a graphics processor in the system, accelerates the result epoch.

⁴colab.research.google.com/

Table 1. Training parameters of the proposed model

Hyperparameter	Value
Learning rate	0.0001
Minibatch	16 or 32
Loss function	'categorical_crossentropy'
Metrics	'acc','loss'
Epochs	500
Optimizer	Adam
Callbacks API	
ModelCheckpoint	Monitor = 'val_loss', save_best_only = True, mode='min'
EarlyStopping	Monitor = 'val_acc', patience = 15, mode = 'max'
CVLogger	'model_history.csv', append = True
ReduceLROnPlateau	Monitor = 'val_loss', factor=0.2, patience=10, min_lr = 0.001

**Fig. 5.** Workflow to analyze the performance of CNNs and the relationship between classes

Notably, Colab determines itself by offering free GPU and TPU support during runtime, extending up to 12 hours in some instances, unlike other cloud systems. The base architectures used are the VGG16, ResNet50, and Xception networks, renowned for their early success in large-scale visual recognition challenges such as ILSVRC [24].

Before training each CNN, it is essential to define the loss function-indicating how the network measures its performance on the training data and guides itself in the desired direction (also known as the objective function) and the optimizer-dictating how the network updates itself based on the observed data and its loss function.

These parameters control the adjustments to the network weights during training. Additionally, regularization techniques, including DropOut (DO) [25], Data Augmentation [19], and Batch Normalization (BN) [23], are incorporated.

A Callback, serving as an object capable of executing actions at different stages of training (e.g., ModelCheckpoint for saving the Keras model, EarlyStopping to halt training when a metric plateau, CSVLogger for logging epoch results in a CSV file, and ReduceLROnPlateau to decrease learning rate on metric stagnation), is integrated.

This holistic approach yields a learning model capable of predicting art media in dataset (test) images with enhanced Accuracy. The training parameters for the proposed models are detailed in Table 1.

4 Experiments and Results

In a previous study [8], three CNN architectures were evaluated for classifying art media, demonstrating the robustness of CNN learning models with a focus on transfer learning. This current work builds on those results, and a detailed evaluation of the same architectures in the context of AMC is performed. The main objective is to address the challenges when employing CNNs with transfer learning in this domain, in addition to analyzing the relationship between ArtMD classes to understand the poor classification performance and how the dataset influences the learning process of CNNs. The workflow for the proposed experimental study is depicted in Fig. 5. As described earlier, the learning models are built upon three foundational architectures: VGG16, ResNet50, and Xception. The models are trained using the ArtMD, incorporating images from the Kaggle website, WikiArt database, and digital artworks sourced from the Louvre Museum in France and the Prado Museum in Spain.

Table 2. Overview of the classification model performance on the Art Media Dataset [8]

Setup 1: Pre-trained CNN base+Dense Classifier (GlobalAveragePooling2D(GAveP2D)+DO(0.2))									
CNN	Params [M]	Epoch	Time [min]	loss	acc	val_loss	val_acc	test_loss	test_acc
VGG16	14.7	91 (90)	173	0.5983	0.7832	0.5699	0.7868	0.8017	0.6911
ResNet50	23.6	50 (49)	84	1.2745	0.4981	1.2295	0.5335	1.5442	0.4122
Xception	20.8	64 (64)	136	0.2920	0.8927	0.3470	0.8761	0.6792	0.7444
Setup 2: Pre-trained CNN base+Dense Classifier (Dense(128)+D0(0.4)+Dense(64)+DO(0.2))									
CNN	Params [M]	Epoch	Time [min]	loss	acc	val_loss	val_acc	test_loss	test_acc
VGG16	17.9	30 (14)	89	0.3313	0.8707	0.4026	0.8527	0.7551	0.7544
ResNet50	36.4	51 (47)	107	1.3590	0.3860	1.2926	0.4275	1.4392	0.3822
Xception	33.7	25 (15)	44	0.3437	0.8654	0.3614	0.8862	0.7967	0.7422
Setup 3: Pre-trained CNN base+Dense Classifier (GAveP2D+Dense(64)+BN()+DO(0.4)+Dense(64)+BN()+DO(0.5))									
CNN	Params [M]	Epoch	Time [min]	loss	acc	val_loss	val_acc	test_loss	test_acc
ResNet50	23.7	50 (40)	100	1.0438	0.6024	0.8845	0.6786	1.3535	0.5422

Table 3. Performance of classification models (Only four classes)

Setup 4 (Engraving class was removed): Pre-trained CNN base+Dense Classifier (Dense(128)+DO(0.3)+Dense(64)+D0(0.2))									
CNN	Params [M]	Epoch	Time [min]	loss	acc	val_loss	val_acc	test_loss	test_acc
VGG16	17.9	35 (18)	57	0.1422	0.9524	0.3070	0.8991	0.6550	0.8208
ResNet50	36.4	26 (4)	68	0.1806	0.9351	0.2454	0.9304	0.7702	0.8514
Xception	33.7	27 (14)	68	0.1318	0.9548	0.2615	0.9056	0.6025	0.8278
Setup 5 (Drawing class was removed): Pre-trained CNN base+Dense Classifier (Dense(128)+DO(0.3)+Dense(64)+D0(0.2))									
CNN	Params [M]	Epoch	Time [min]	loss	acc	val_loss	val_acc	test_loss	test_acc
VGG16	17.9	29 (19)	73	0.0696	0.9747	0.1412	0.9631	0.6434	0.8375
ResNet50	36.4	26 (4)	64	0.1147	0.9649	0.1195	0.9645	0.6549	0.8514
Xception	33.7	17 (4)	36	0.1549	0.9461	0.1343	0.9597	0.4589	0.8611

4.1 Classification Performance Evaluation

Table 2 illustrates a comparison between the reference models' accuracy and loss across different datasets (training, validation, and test) and the proposed setups to the base structure.

This initial investigation delves into the CNNs' performance concerning each dataset class. Notably, the Xception model excels, achieving the highest classification accuracy of 74% in the first setup. Conversely, the VGG16 model attains its peak performance with 75% accuracy in the second setup.

The ResNet50 model exhibits a lower accuracy in the test set compared to the training and validation sets. In a third setup focusing on enhancing classification performance through the dense classifier, the ResNet50 model demonstrates acceptable performance with an accuracy of 54%.

Furthermore, this proposed approach features a reduced number of training parameters compared to its predecessor. The accuracy of the proposed models, in particular, maintains homogeneity when training with the training and validation sets. This is expected because there is a control to avoid model overfitting.

The proposed regularization methods and Callbacks are integrated into the architecture to eliminate overfitting to monitor the learning process. With the test information, the base models achieve an accuracy below the training and validation set.

Interestingly, the models predict images (test) that have never been used for training, meeting the goal of generalization of knowledge in CNNs, but not enough to achieve the optimal performances reported in classification tasks. Fig. 6a, 6d and 6g show the confusion matrix for the test set (with five classes) in the three models.

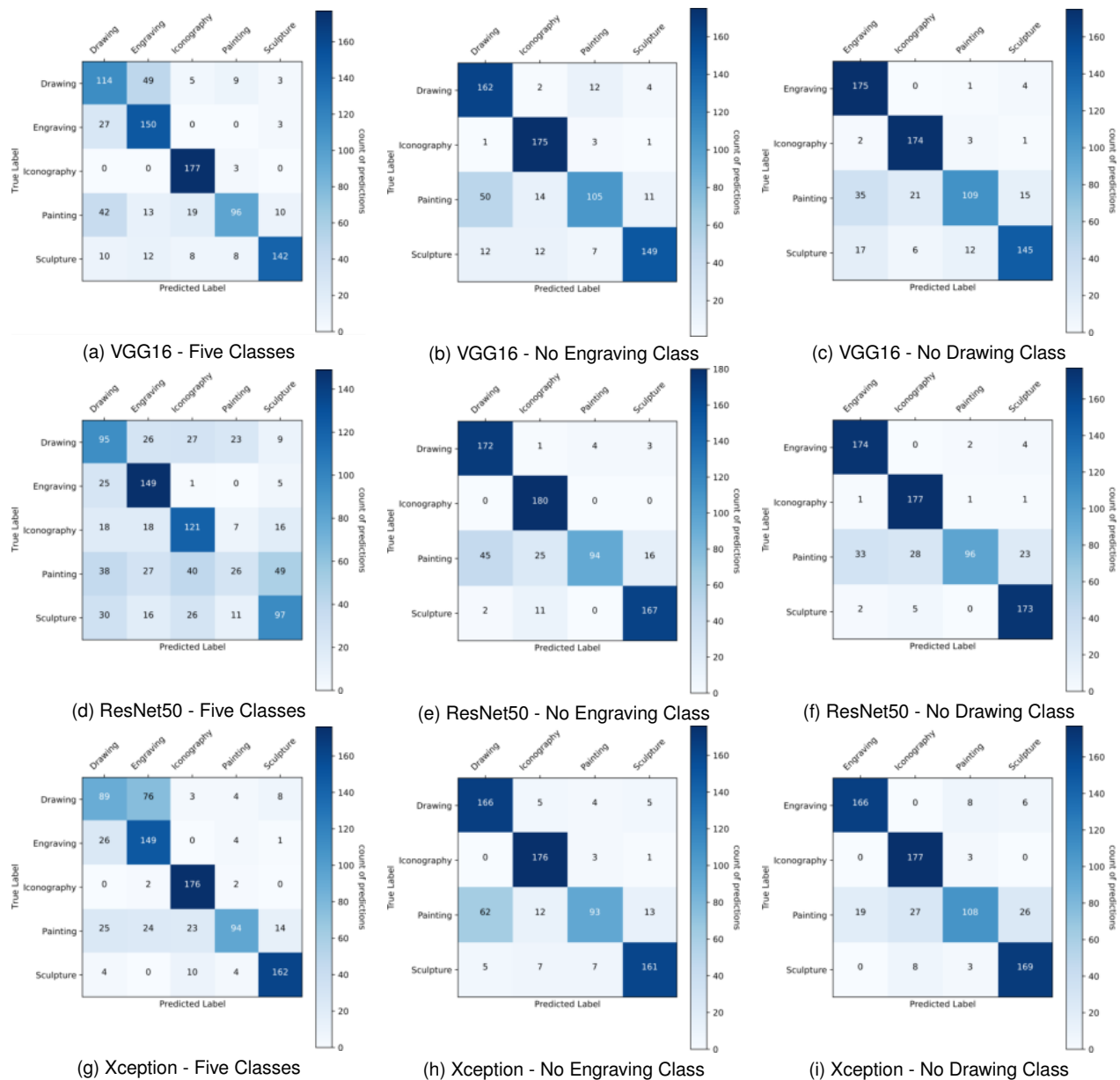


Fig. 6. Confusion matrix for the Art Media Dataset (Test)

As illustrated, the Iconography class has a high classification performance by the VGG16 and Xception model (177 and 176 images correctly classified). Also, the Xception model improves classification performance with respect to the Sculpture class (162 images correctly classified). In both cases (Iconography and Sculpture) with

a classification performance above 90%. Some categories share similarities in color, composition, and texture. Therefore, misclassification errors in the three CNN models, such as the Drawing and Engraving class, are common. On the other hand, the Painting class shows a classification rate of about 95 images in the three CNN models.

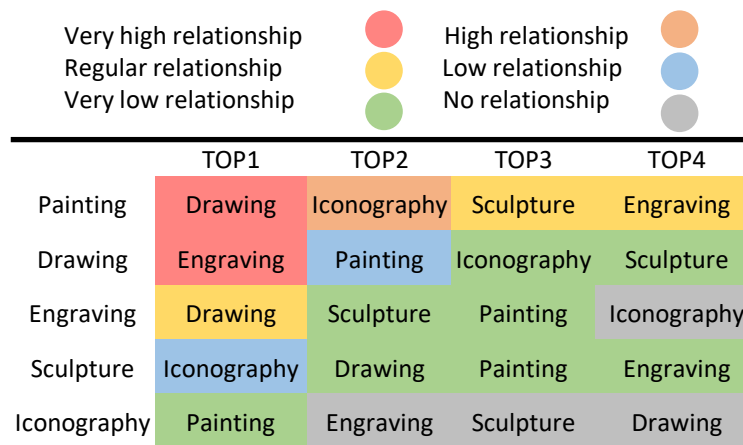


Fig. 7. Summary of the class relationship effect in the CNN models using ArtMD

This means that the class is highly connected with the other classes and that it is difficult for CNN to predict which category it belongs to.

4.2 Classes Relationship Effects in the CNN Models

The relationship between classes refers to the similarity between the characteristics of each class, which can confuse CNN models [13]. In addition, errors in the confusion matrix can occur for various reasons, such as the quality and quantity of training data, the complexity of the classification problem, or the suitability of the learning algorithm used.

Therefore, it is essential to analyze further the nature of the errors and the dataset's characteristics to understand why the three CNN models are making errors and to determine if there is a real relationship between classes or if they are due to other causes. To get an idea of which class (Drawing or Engraving) has fewer characteristics in common, it is proposed to modify the dataset to only four classes.

This involves modifying the dense classifier stage setup of the three models (VGG16, ResNet50, and Xception): Dense(128) + DO(0.3) + Dense(64) + DO(0.2) + Dense(4). In the first additional study (setup 4), the Engraving class was removed, increasing the accuracy of the VGG16, ResNet50, and Xception models, reaching a top accuracy of 85% (ResNet50).

In the second study (setup 5), the Drawing class was removed, and a similar behavior was obtained with a maximum accuracy of 86% (Xception). It should be noted that this increase in accuracy was mainly observed in the test set, while in the training and validation sets, top accuracy exceeded 90%, as detailed in Table 3.

The confusion matrices shown in Figures 6b-c, 6e-6f, and 6h-6i reveal that three of the four classes (Drawing or Engraving, Iconography, and Sculpture) have a classification performance above 90% in the ResNet50 and Xception models in setup 4 and setup 5.

Furthermore, it is noted that in all three CNN models, the Painting class is highly related to the other categories, as they share characteristics of style, period, and techniques. This suggests that the main challenge lies in the complexity of the field of study, particularly in the Drawing and Engraving classes and the Painting class.

The summary of the three CNN models yields the following Fig. 7 In which we observe that the Drawing class presents the most problems for the classification task, with two (Painting and Engraving) of the four remaining classes. The Engraving class shows a very high relationship with the Drawing class. As for the Sculpture class, it has a shallow relationship with the Iconography class. The class with minor problems is the Iconography class, achieving almost null relationships.

The color selection was made based on the miss-classification in the three CNN models: (125 < Very high), (100 < High < 125), (75 < Regular < 100), (50 < Low < 75), (15 < Very low < 50), and (No relationship < 15).

In the setup and implementation of the network, it was decided to use a function of the Keras library, `preprocess_input`, which allows processing the images with the same characteristics as the CNN pre-trained with the ImageNet database. The function is only applied to the ResNet50 architecture due to its low performance.

5 Conclusion and Future Work

This paper proposes an evaluation and performance analysis of three different CNNs applied to Art Media Classification (AMC) in order to answer the question of what challenges arise in AMC using CNNs with transfer learning. The features previously obtained in training the CNNs allow improving the accuracy of each learning model, without the need to start from scratch.

Given the need to evaluate the learning model, the Art Media Dataset (ArtMD) was introduced. The dataset includes the art classes: Drawing, Engraving, Iconography, Painting, and Sculpture. Initially, the VGG16 model obtained the best accuracy with 75%, but when analyzing that the main challenge lies in the dataset and that the CNNs have a difficult field of study, a new configuration is proposed.

Instead of using five classes, it was decided to evaluate only four (Drawing or Engraving, Iconography, Painting, and Sculpture). Therefore, the three proposed models now obtain a top accuracy of 86%. These experiments allow us to analyze miss-classification and discuss the relationship effects in the three CNN models to understand the artwork's composition.

The results show that all the tested CNNs present a high relationship in the classification of Painting due to characteristics of style, period, etc., followed by the relationship between classes of Drawing and Engraving due to the similarities of both classes. Separately, both classes are unrelated and have a classification performance above 90%.

In the case of Iconography and Sculpture (with low or no relationship), it can be inferred that any model will be able to perform a correct classification. In our experimental study, we applied Data Augmentation, DropOut, and Batch Normalization to the dataset to mitigate the overfitting of CNNs.

As future work, we will design a classification system based on the results obtained in this research. To achieve this, a more detailed analysis of different styles of artwork will be carried out to extract additional information that reduces the class relationship effect.

Furthermore, we propose to use wavelet analysis as a preprocessing module to obtain spectral information and improve the accuracy of the proposed CNN architectures. Finally, the results can be used to enhance the design of image classification systems applied in other areas, such as medical, surveillance, aerial robotics, and automation.

Acknowledgments

This work was funded by CONACYT through the grant "Convocatoria de Ciencia Básica y/o Ciencia de Frontera 2022", project ID 320036.

References

1. **Abidin, D. (2021)**. The effect of derived features on art genre classification with machine learning. *Sakarya University Journal of Science*, Vol. 25, No. 6, pp. 1275–1286. DOI: 10.16984/saufenbilder.904964.
2. **Berrouane, N., Benyettou, M., Ibtissam, B. (2022)**. Deep learning and feature extraction for COVID-19 diagnosis. *Computación y Sistemas*, Vol. 26, No. 2, pp. 909–920. DOI: 10.13053/cys-26-2-4268.
3. **Chollet, F. (2016)**. Xception: Deep learning with depthwise separable convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807. DOI: 10.1109/cvpr.2017.195.

4. **Chollet, F. (2021).** Deep learning with Python. Simon and Schuster.
5. **Chu, W. T., Wu, Y. L. (2016).** Deep correlation features for image style classification. Proceedings of the 24th ACM international conference on Multimedia, pp. 402–406. DOI: 10.1145/2964284.2967251.
6. **Chu, W. T., Wu, Y. L. (2018).** Image style classification based on learnt deep correlation features. IEEE Transactions on Multimedia, Vol. 20, No. 9, pp. 2491–2502. DOI: 10.1109/TMM.2018.2801718.
7. **Fortuna-Cervantes, J. M., Ramírez-Torres, M. T., Mejía-Carlos, M., Martínez-Carranza, J., Murguía-Ibarra, J. S. (2021).** Texture classification for object detection in aerial navigation using transfer learning and wavelet-based features. 12th International Micro Air Vehicle Conference, pp. 210–215.
8. **Fortuna-Cervantes, J. M., Soubervielle-Montalvo, C., Perez-Cham, O. E., Peña-Gallardo, R., Puente, C. (2023).** Experimental study of the performance of convolutional neural networks applied in art media classification. Mexican Conference on Pattern Recognition, pp. 169–178.
9. **Gao, J., Zhou, H., Zhang, Y. (2020).** The performance of two CNN methods in artworks aesthetic feature recognition. Proceedings of the 12th International Conference on Machine Learning and Computing, pp. 289–296. DOI: 10.1145/3383972.3383974.
10. **He, K., Zhang, X., Ren, S., Sun, J. (2016).** Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. DOI: 10.1109/cvpr.2016.90.
11. **Ibarra-Vazquez, G., Olague, G., Chan-Ley, M., Puente, C., Soubervielle-Montalvo, C. (2022).** Brain programming is immune to adversarial attacks: Towards accurate and robust image classification using symbolic learning. Swarm and Evolutionary Computation, Vol. 71, pp. 101059. DOI: 10.1016/j.swevo.2022.101059.
12. **Ibarra-Vazquez, G., Olague, G., Puente, C., Chan-Ley, M., Soubervielle-Montalvo, C. (2021).** Automated design of accurate and robust image classifiers with brain programming. Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1385–1393. DOI: 10.1145/3449726.3463179.
13. **Jiang, Y. G., Wu, Z., Wang, J., Xue, X., Chang, S. F. (2018).** Exploiting feature and class relationships in video categorization with regularized deep neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, No. 2, pp. 352–364. DOI: 10.1109/TPAMI.2017.2670560.
14. **Kandel, I., Castelli, M. (2020).** Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review. Applied Sciences, Vol. 10, No. 6, pp. 2021. DOI: 10.3390/app10062021.
15. **Kovalev, V. Y., Shishkin, A. G. (2020).** Painting style classification using deep neural networks. IEEE 3rd International Conference on Computer and Communication Engineering Technology, pp. 334–337. DOI: 10.1109/ccet50901.2020.9213161.
16. **Lombardi, T. E. (2005).** The classification of style in fine-art painting. ETD Collection for Pace University, pp. 1–158.
17. **Lugo-Sánchez, O. E., Sossa, H., Zamora, E. (2020).** Reconocimiento robusto de lugares mediante redes neuronales convolucionales. Computación y Sistemas, Vol. 24, No. 4, pp. 1589–1605. DOI: 10.13053/cys-24-4-3340.
18. **Masilamani, G. K., Valli, R. (2021).** Art classification with pytorch using transfer learning. International Conference on System, Computation, Automation and Networking, pp. 1–5. DOI: 10.1109/icscan53069.2021.9526457.
19. **Mikołajczyk, A., Grochowski, M. (2018).** Data augmentation for improving deep learning in image classification problem. International Interdisciplinary

PhD Workshop, pp. 117–122. DOI: 10.1109/iiphdw.2018.8388338.

20. **Olague, G., Ibarra-Vázquez, G., Chan-Ley, M., Puente, C., Soubervielle-Montalvo, C., Martínez, A. (2020).** A deep genetic programming based methodology for art media classification robust to adversarial perturbations. Proceedings of the 15th International Symposium on Visual Computing. Advances in Visual Computing, pp. 68–79. DOI: 10.1007/978-3-030-64556-4_6.
21. **Rodríguez, C. S., Lech, M., Pirogova, E. (2018).** Classification of style in fine-art paintings using transfer learning and weighted image patches. 12th International Conference on Signal Processing and Communication Systems, pp. 1–7. DOI: 10.1109/icspcs.2018.8631731.
22. **Rojas-Pérez, L. O., Martínez-Carranza, J. (2020).** Autonomous drone racing with an

opponent: A first approach. *Computación y Sistemas*, Vol. 24, No. 3, pp. 1271–1279. DOI: 10.13053/cys-24-3-3486.

23. **Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986).** Learning representations by back-propagating errors. *Nature*, Vol. 323, No. 6088, pp. 533–536. DOI: 10.1038/323533a0.
24. **Simonyan, K., Zisserman, A. (2014).** Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations, pp. 1–14. DOI: 10.48550/ARXIV.1409.1556.
25. **Srivastava, N. (2013).** Improving neural networks with dropout. University of Toronto, Vol. 182, No. 566, pp. 7.

*Article received on 27/08/2023; accepted on 13/10/2023.
Corresponding author is Carlos Soubervielle-Montalvo*.*