Leveraging Computer Vision Application in Visual Arts: A Case Study on the Use of Residual Neural Network to Classify and Analyze Baroque Paintings

Daniel Kvak*
Faculty of Arts
Masaryk University
Brno, Czech Republic

ORCID: 0000-0001-7808-7773

October 27, 2022

Abstract

With the increasing availability of large digitized fine art collections, automated analysis and classification of paintings is becoming an interesting area of research. However, due to domain specificity, implicit subjectivity, and pervasive nuances that vaguely separate art movements, analyzing art using machine learning techniques poses significant challenges. Residual networks, or variants thereof, are one the most popular tools for image classification tasks, which can extract relevant features for well-defined classes. In this case study, we focus on the classification of a selected painting 'Portrait of the Painter Charles Bruni' by Johann Kupetzky and the analysis of the performance of the proposed classifier. We show that the features extracted during residual network training can be useful for image retrieval within search systems in online art collections.

Keywords: computational creativity; deep learning; feature extraction; image analysis; machine perception; painting classification; residual networks; transfer learning.

*Corresponding author: kvak@mail.muni.cz

 doi:10.20944/preprints202210.0448.v1

1 Introduction

Image classification is one of the most widely used computer vision tasks. [Lu and Weng, 2007] In the recent past, deep learning has been very successful in various visual tasks, such as agent-based simulation of autonomous vehicles [Schwarting et al., 2018] or computer-aided detection / diagnosis in the healthcare segment. [Doi, 2007] The extensive digitization that has occurred in the last two decades [Aydoğan, 2019] has led to the question of whether the curation segment can also be automated using machine methods. The conversion of information from physical works of art into digital image format plays a key role in the opening of new research challenges in the interdisciplinary field of computer vision, machine learning, and art history. [Cetinic et al., 2018, Tan et al., 2016, Saleh and Elgammal, 2015]

Different convolutional neural network (CNN) architectures have been proven to work well for image recognition and classification tasks. The basic idea is that neurons in the visual cortex process images into increasingly complex shapes. [Lindsay, 2021] The image is first segmented at edge boundaries using a light / dark interface, then merged into simple shapes, and finally merged into recognizable complex features in subsequent layers. [Albawi et al., 2017] Individual class labels may be based on some low-level features such as color, texture, or shape, but are most often based on higher-level features such as semantic description, activity, or artistic style. [O'Shea and Nash, 2015] CNN tries to mimic this idea using several layers of artificial neurons. The standard architecture includes several convolutional layers that segment the image into small chunks that can be easily processed. [Albawi et al., 2017]

2 Proposed Method

The use of machine learning for automatic classification of fine art collections has received little attention in the literature so far. [Arora and Elgammal, 2012, Rodriguez et al., 2018] In recent years, libraries, museums, galleries, and art centers have been digitizing their collections to promote public interest in the arts and facilitate access to masterpieces from the comfort of home, a trend that has been further reinforced by the ongoing COVID-19 pandemic. [Habsary et al., 2021] These activities create a demand for automated analysis and classification of digitized art. [Khoronko and Mokina, 2021] In this paper, we propose a novel approach to using CNN output to classify visual artwork. Using CNN pre-trained on ImageNet,[1] we consider feature maps computed at the level of several different layers before fully connected layers and compare the perception of artificial intelligence with the analysis of art historians and curators. We show that the extracted features are effective for classifying artists and styles and

[1] ImageNet is a large-scale visual database designed for use in image classification and object recognition research. The project includes more than 14 million images that have been manually annotated to indicate what objects are shown. ImageNet features more than 20,000 categories, with a typical category such as "balloon" or "strawberry" consisting of several hundred images

 doi:10.20944/preprints202210.0448.v1

provide a detailed visualization and discussion of the suitability and effectiveness of the different layers.

2.1 Transfer Learning

In transfer learning, a neural network is first trained on a generic dataset (e.g. ImageNet visual database), and the features learned from the initial task are transferred to a new network that is fine-tuned for a specific task. [Weiss et al., 2016] Deploying pre-trained models on similar data has shown solid results in image classification-related tasks. [Weiss et al., 2016, Zhuang et al., 2020] Several organizations have created models such as VGG [Sengupta et al., 2019], Inception [Szegedy et al., 2016], or ResNet [He et al., 2016] that would take weeks to train on user-accessible hardware. Pre-trained networks can be downloaded and easily fine-tuned to result in lower generalization error while using less computational effort.

2.2 ResNet50V2 Model Architecture

As deep learning evolves, the structure of neural networks deepens; while this helps the network to perform more complex feature extraction, it can also introduce the problem of vanishing or exploding gradients. [Joshi et al., 2019] This can lead to the following drawbacks: (1) Long training time with the convergence of the network becomes very difficult or even non-convergent. (2) The network performance gradually becomes saturated and even starts to decline. [Joshi et al., 2019, Kim et al., 2016]

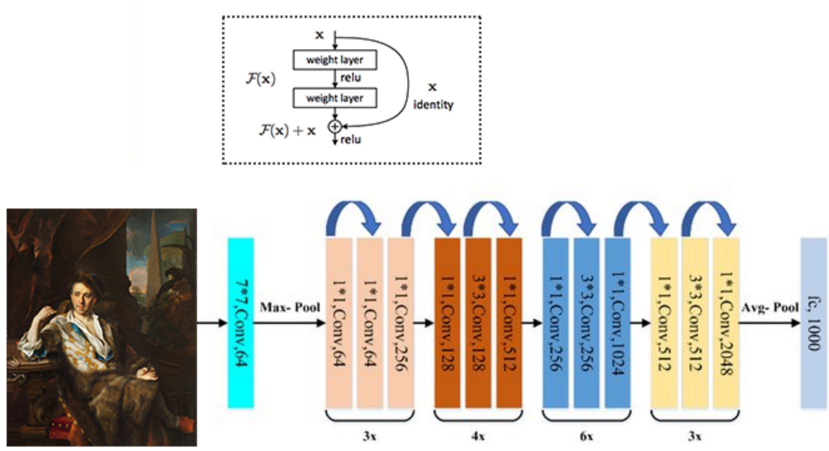

Figure 1: Proposed architecture of ResNet50V2 model.

 doi:10.20944/preprints202210.0448.v1

[He et al., 2016]	Layers	Depth	Parametres	Input matrix	Output activation
ResNet50V2	50	103	25.6M	224, 224, 3	softmax

Table 1: Architecture of the proposed model

Instead of designing own architecture, we leverage insights from a set of existing convolutional neural network architectures that show excellent performance in solving a variety of classification tasks. Residual networks (ResNets) are a unique type of deep convolutional networks whose basic idea is to skip blocks of convolutional layers by using shortcut connections. [He et al., 2016] In this case study, we use a variant of residual neural networks called ResNet50V2 (shown in Figure 1). [Yamazaki et al., 2019]

The basic building blocks follow two simple rules: (i) for the same output feature map size, the layers have the same number of filters; and (ii) if the feature map size is halved, the number of filters is doubled. [He et al., 2016] The down-sampling is performed by convolutional layers that have a stride of 2 and batch normalization is performed right after each convolution operation and before ReLU activation. [He et al., 2016] When the input and output are of the same dimensions, the identity shortcut is used. When dimensions increase, the projection shortcut is used to match dimensions through 1×1 convolutions. In both cases, when the shortcuts go across feature maps of two sizes, they are performed with a stride of 2. [He et al., 2016] The network output ends with a 1,000 fully-connected layer and softmax function. The total number of weighted layers is 50, with 25.6 million trainable parameters. [Yamazaki et al., 2019]

ResNets provide a trade-off between performance and number of parameters. [Huang et al., 2017] The weights used in the proposed model have been pre-trained using the ImageNet database. [Yamazaki et al., 2019] Our hypothesis is that despite the obvious discrepancy between images embedded in ImageNet databbase and fine art collections, ResNet-50V2 comprehensively pre-trained on the ImageNet may still be transferred to perform style classification tasks more effective.

2.3 Feature Extraction

The classification of painting styles is usually done by art historians and curators based on the relationship between subjective attributes, physical characteristics, as shown in Figure 2 (light, lines, colors, textures, shapes, space, etc.), and appropriate historical periods. [Zhao et al., 2021] In many cases, however, significant stylistic differences can be observed, such as seamless transitions between artistic movements over time, stylistic differences between paintings by the same artist, unique personal characteristics that do not belong to any one style or artistic period, the influence of one artist on others, and varying interpretations of abstract and surreal elements. Therefore, these variables make it difficult to apply classification methods in the visual arts. [Sandoval et al., 2019]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 October 2022 doi:10.20944/preprints202210.0448.v1

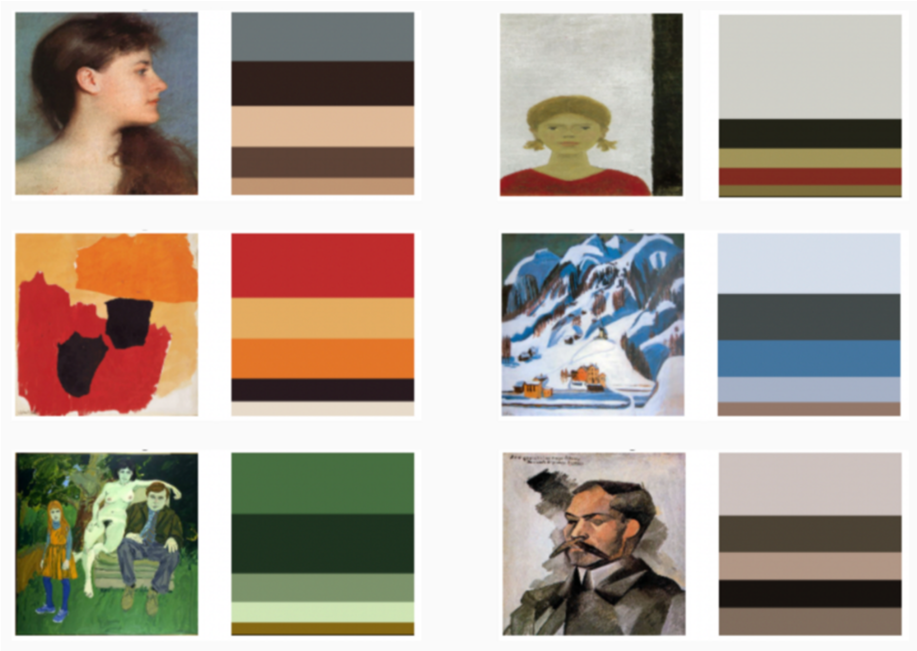

Figure 2: Dominant color palette of represented examples.

Feature extraction is part of the dimensionality reduction process, where the initial raw data set is split and transformed into more manageable groups. [Chen et al., 2016] Feature extraction aims to reduce the number of features in the dataset by creating new features from existing ones. This new reduced feature set should be able to summarize most of the information contained in the original feature set. [Wiatowski and Bölcskei, 2017]

Image processing is one of the domains where feature extraction finds wide application. [Nixon and Aguado, 2019] Feature extraction in CNN uses many techniques that include methods to detect low-level features such as colors, brightness, edges, or textures in order to process them. [Chen et al., 2016] The convolution layer consists of a set of digital filters that perform convolution operations on the input data. [He et al., 2016] The convolutional layer serves as a dimensionality reduction layer and decides the threshold. During backpropagation, a number of parameters need to be adjusted, which in turn minimizes the coupling within the neural network architecture. [He et al., 2016]

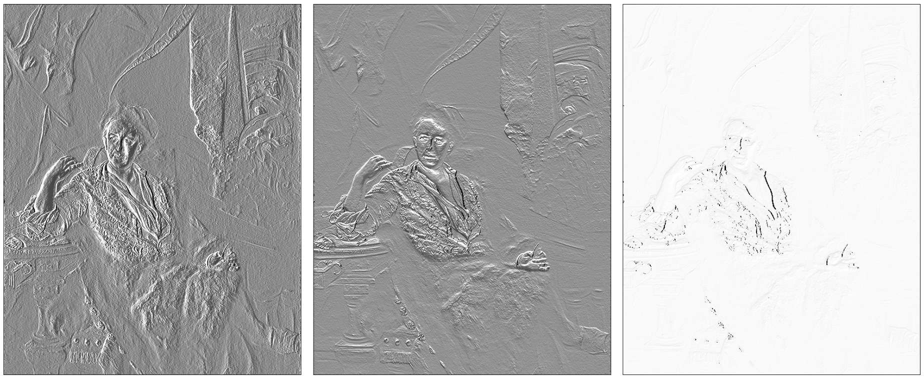

Figure 3: Principal Component Analysis of luminance gradient feature of 'Portrait of the Painter Charles Bruni' by Johann Kupetzky.

3 Experiment

3.1 Dataset

Since each class of styles contains paintings by different artists, training and classification is not simple. [Sandoval et al., 2019] The ability to correctly classify artworks into narrowly defined domains is a difficult task even for curators of existing datasets. [Zhao et al., 2021] The achieved classification results are based on the freely available data set 'Art Snobs Data2' extracted from Kaggle. [Martin, 2019] These paintings were collected from various sources and contain a large collection of art pieces from different eras, with 19 distinct styles, namely: Abstract Expressionism, Art Nouveau, Baroque, Cubism, Early Renaissance, Expressionism, High Renaissance, Impressionism, Mannerism, Naive Art, Neoclassicism, Northern Renaissance, Post-Impressionism, Realism, Rococo, Romanticism, Surrealism, Symbolism and Ukiyo-e. To perform style classification, we used 923 images from each class to train the classifiers and used the remaining 102 as the validation set, whereas according to Fig. 4, it is a perfectly balanced set. Thus, the dataset consists a total of 17,537 train images and 1,938 validation images.

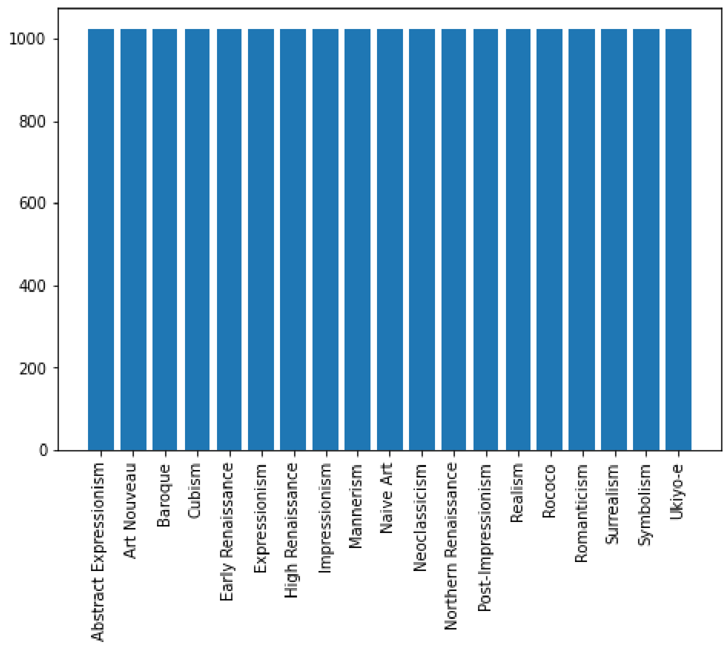

Figure 4: Distribution of train and validation examples for individual classes.

3.2 Data Augmentation

Data augmentation increases the size of the input training data along with the regularization of the model, thus improving the generalization of the training model. [Shorten and Khoshgoftaar, 2019] It also helps to create new train examples by randomly applying different transformations to the available dataset to reflect the noisiness of real-world data. [Perez and Wang, 2017] In our experiment, we used transformations that involve vertical flipping of training images, random rotations, modifications in lighting conditions, zoom, saturation, and JPEG encoding noise. The training data was augmented using five randomly selected variations; the extension of validation dataset was not investigated.

3.3 Training

The neural network learns using the backpropagation method. [He et al., 2016] The weights of fully-connected layer of ResNet50V2 can be fine-tuned, i.e., this layer adapts by backpropagation, while the other layers of the network are invariant after pre-training on ImageNet database. [Yamazaki et al., 2019] Fine-tuning of the top layer in the ResNet50V2 is performed as it is not guaranteed that the mean and variance of these layers will be similar to the mean and variance of our dataset.

An F1 Score becomes a critical evaluation tool to determine False Positive and False Negative rates yielded through a discriminating threshold in a similar situation with unbalanced dataset samples. [Goutte and Gaussier, 2005] The classification performance of our model for multi-class problem was evaluated for each component and the average classification performance of the model was calculated. Table 2 includes the precision, recall, and F1 Score, calculated based on the following equations:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{3}$$

Based on the equations, the table below shows the Precision, Recall, F1 Score results for the multi-class classification:

Class	**Precision**	**Recall**	**F1 Score**
Abstract Expressionism	0.88	0.89	0.88
Art Nouveau	0.73	0.66	0.69
Baroque	0.63	0.45	0.53
Cubism	0.87	0.80	0.84
Early Renaissance	0.83	0.66	0.74
Expressionism	0.56	0.69	0.62
High Renaissance	0.41	0.72	0.52
Impressionism	0.86	0.55	0.67
Mannerism	0.59	0.60	0.60
Naive Art	0.86	0.73	0.79
Neoclassicism	0.70	0.70	0.70
Northern Renaissance	0.68	0.61	0.64
Post-Impressionism	0.64	0.66	0.65
Realism	0.47	0.63	0.54
Rococo	0.61	0.79	0.69
Romanticism	0.74	0.32	0.45
Surrealism	0.76	0.75	0.75
Symbolism	0.65	0.69	0.67
Ukiyo-e	0.88	0.95	0.91

Table 2: ResNet50V2 performance on test data.

 doi:10.20944/preprints202210.0448.v1

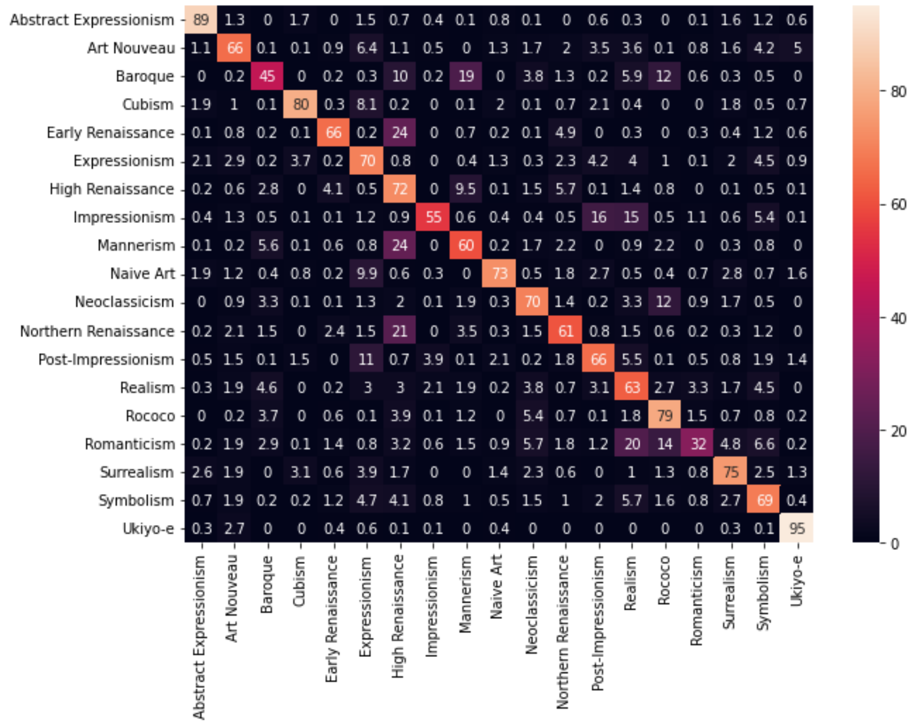

Figure 5: Confusion matrix for multi-class classification.

4 Results

The results obtained from the experiment can be analyzed and discussed from several perspectives. From the confusion matrix in Figure 5, we can observe the internal logic of the misclassified classes. The common visual features of the different styles explain the high misclassification rate (=lower model accuracy) between classes such as Northern Renaissance and High Renaissance [Nash, 2008] or Baroque and Rococo [Wittkower et al., 1999] compared to other tasks, the lower classification of styles corresponds to the high overlap of visual features between classes and also to the large variety of content displayed in a single style. In contrast, the classes of the genre classification task are more uniform in content, and CNNs show a high ability to discriminate classes with radically different visual cues. [Hagtvedt and Patrick, 2011]

This example of inaccurate classification, which can also be observed in the prediction for the analyzed work 'Portrait of the Painter Charles Bruni' by Johann Kupetzky, highlights the fact that style is not only associated with the mere visual characteristics and content of an artwork, but is often a subtly distinguishable and context-dependent concept, e.g. Baroque and Rococo suggested considerable confusion, which can be attributed to the fact that these

 doi:10.20944/preprints202210.0448.v1

Class	Confidence
Rococo	0.2942
Neoclassicism	0.2503
Baroque	0.1205
Realism	0.1201
Romanticism	0.1173

Table 3: Top 5 class predictions for 'Portrait of the Painter Charles Bruni' by Johann Kupetzky.

styles are historically related. [Ruth and Kolehmainen, 1974]

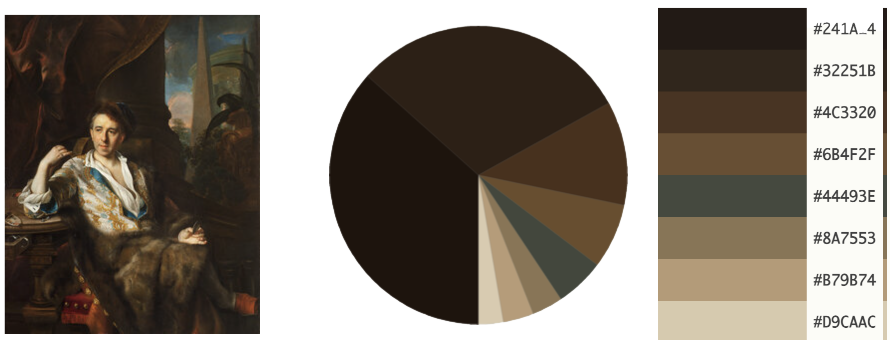

Figure 6: Dominant color palette of 'Portrait of the Painter Charles Bruni' by Johann Kupetzky.

5 Conclusions

This study investigated the design and construction of a CNN-based model for fine art classification using transfer learning and feature extraction utilizing the freely available dataset 'Art Snobs Data2' from the Kaggle repository and evaluated its performance on the example of 'Portrait of the Painter Charles Bruni' by Johann Kupetzky. The classification results demonstrated that the proposed approach provides a computationally efficient way to classify artworks by fine art style without the necessity of defining and fully training new CNN model structures or increasing the size of existing databases. Empirical results based on the confusion matrix demonstrate that CNN has learned relevant features for individual classes, which can be proven by misclassification when trying to classify a work of art across historically related styles.

Although the classifier has achieved interesting results and proposed possible

 doi:10.20944/preprints202210.0448.v1

Figure 7: Grad-CAM class activation visualization from the proposed model.

applications, in the future we would like to extend this research to the context of contemporary curatorial practice. While the CNN only predicts the outcome on the basis of visual input (individual shapes, edges, or colors) without any knowledge of the sociocultural context, the human expert interprets the works according to complex relevant features and characteristics that the CNN cannot observe. It is difficult to assess whether it is possible to simply define individual art movements, to evaluate a work using visual interpretation alone, or whether such an interpretation is even relevant to existing curatorial practice; nevertheless, this research has outlined possible pathways for future research.

References

Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

Ravneet Singh Arora and Ahmed Elgammal. Towards automated classification of fine-art painting style: A comparative study. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3541–3544. IEEE, 2012.

D Aydoğan. Reflections of digitalization on painting. *CTC 2019*, 2019.

Eva Cetinic, Tomislav Lipic, and Sonja Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, 114: 107–118, 2018.

Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.

Kunio Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5):198–211, 2007.

Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer, 2005.

Dwiyana Habsary, Agung Kurniawan, and Indra Bulan. Digitalization of arts. In *Proceedings of the Tenth International Conference on Languages and Arts (ICLA 2021)*, pages 246–250. Atlantis Press, 2021.

Henrik Hagtvedt and Vanessa M Patrick. Turning art into mere illustration: Concretizing art renders its influence context dependent. *Personality and Social Psychology Bulletin*, 37(12):1624–1632, 2011.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7310–7311, 2017.

Soumya Joshi, Dhirendra Kumar Verma, Gaurav Saxena, and Amit Paraye. Issues in training a convolutional neural network model for image classification. In *International Conference on Advances in Computing and Data Sciences*, pages 282–293. Springer, 2019.

Lubov Khoronko and Anna Mokina. Museum practice in the developing of applied artists' professional competencies in the context of digitalization of education. In *E3S Web of Conferences*, volume 273, page 12066. EDP Sciences, 2021.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

Grace W Lindsay. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, 33(10): 2017–2031, 2021.

Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.

Parker Martin. Art snobs data2, Nov 2019. URL `https://www.kaggle.com/datasets/parkerzmartin/art-snobs-data2`.

Susie Nash. *Northern renaissance art*. OUP Oxford, 2008.

Mark Nixon and Alberto Aguado. *Feature extraction and image processing for computer vision*. Academic press, 2019.

Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

Catherine Sandoval Rodriguez, Margaret Lech, and Elena Pirogova. Classification of style in fine-art paintings using transfer learning and weighted image patches. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–7. IEEE, 2018.

Jan-Erik Ruth and Kyösti Kolehmainen. Classification of art into style periods; a factor-analytical approach. *Scandinavian Journal of Psychology*, 15(1):322–327, 1974.

Babak Saleh and Ahmed Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.

 doi:10.20944/preprints202210.0448.v1

Catherine Sandoval, Elena Pirogova, and Margaret Lech. Two-stage deep learning approach to the classification of fine-art paintings. *IEEE Access*, 7:41770–41781, 2019.

Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, 1:187–210, 2018.

Abhronil Sengupta, Yuting Ye, Robert Wang, Chiao Liu, and Kaushik Roy. Going deeper in spiking neural networks: Vgg and residual architectures. *Frontiers in neuroscience*, 13:95, 2019.

Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

Wei Ren Tan, Chee Seng Chan, Hernán E Aguirre, and Kiyoshi Tanaka. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE international conference on image processing (ICIP)*, pages 3703–3707. IEEE, 2016.

Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.

Thomas Wiatowski and Helmut Bölcskei. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory*, 64(3):1845–1866, 2017.

Rudolf Wittkower, Joseph Connors, and Jennifer Montagu. *Art and Architecture in Italy, 1600–1750: Volume 3: Late Baroque and Rococo, 1675–1750*, volume 3. Yale University Press, 1999.

Masafumi Yamazaki, Akihiko Kasagi, Akihiro Tabuchi, Takumi Honda, Masahiro Miwa, Naoto Fukumoto, Tsuguchika Tabaru, Atsushi Ike, and Kohta Nakashima. Yet another accelerated sgd: Resnet-50 training on imagenet in 74.7 seconds. *arXiv preprint arXiv:1903.12650*, 2019.

Wentao Zhao, Dalin Zhou, Xinguo Qiu, and Wei Jiang. Compare the performance of the models in art classification. *Plos one*, 16(3):e0248414, 2021.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.