Research Inventy: International Journal of Engineering And Science
Vol.11, Issue 5 (May 2021), PP 18-26
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Deep Neural Networks based Classical Painting Style Prediction

Rabeeah Ahmed Meelad* and Marina Maranovic
Department Advanced Security System, Faculty of Informatics and Computing
Singidunum University, Danijelova 32, Beograd 160622, Serbia
**Corresponding author: Rabeeah Ahmed Meelad*

Abstract: *How does deep convolution neural network classify painting style? Since the deep Convolution Neural Network(CNN or ConvNet) have been introduced in early of 90's, from that date to until now, the convolution neural network undergone many improvements, and it have consistently been competitive with other techniques for image classification and recognition,Several studies have shown the ability of the deep convolution neural network to learn and predict painting style. In this paper we are going to representing* deep neural network,and use different deep neural networks methods, such as *(AlexNet,VGG16,ResNet,LeNet5,Inception), and we are going to built in our classifier to classification Art Style of different art works for the famous Painters,such as [Abstract Expression-ism,Analytical Cubism,Art Nouveau Modern,Color Field Paint-ing, Cubism,Early Renaissance,Expressionism,Fauvism, High Re-naissance, Impressionism].*

Keywords: *Painting style prediction, deep neural networks, image processing, colour information, painting evaluation.*

Date of Submission: 27-05-2021 Date of Acceptance: 09-06-2021

I. INTRODUCTION

In this paper we are going to representing how to implement convolution neural networks (CNNs), in python by using Keras, tensorflow, where we are going to assess our architecture on the given data set, to achieve higher accuracy, and at the end we will compare our results with group of selected works, to see if we can get better results. In these days, you might have already heard about image or facial classification, or self-driving cars. These are real-life implementations of convolution neural network (CNNS). As we mentioned before, we will present and understand how to implement these deep architectural, so by that we can see if these architectures can work very well to classification Artwork of different art styles.

II. RELATED WORKS

Lecoutre et al. (2017) have studied the use of Deep Residual Neural Network in order to solve the problem of detecting the artistic style of a painting was investigated. It is shown that this method retraining about 20 layers outperforms existing approaches by almost 10 % on the Wikipaintings dataset (Wikiart dataset). The network was first pre-trained on ImageNet (Deng et al. 2009). Using this approach, the accuracy of 62% was achieved in recognition of 25 artistic styles. https://www.overleaf.com/project/5e335dd06a5fb60001d4dd10. Hicsonmez et al. (2017) carried out that Deep Neural Networks such as: AlexNet, VGG 19 and GoogLeNet with fine tuning were exploited in order to categorise 24 artists from 6500 collected pages from children books. It is shown that fine tuned GoogLeNet outperformed other architectures, achieving the accuracy of 94%.

Dushkoff et al. (2016) reported that a VGG style Convolutional Neural Network (CNN) which was pre-trained on the ImageNet dataset in order to be used for the frame-by-frame stylization. In this stylization, style loss and content loss were computed at each layer. It is demonstrated that this spatiotemporal framework was robust enough to automatically produce temporally coherent videos representing a given artistic style without distracting jittering effects. A recent advance in texture synthesis that were motivated by visual neuroscience and have led to a substantial advance in image synthesis and manipulation in computer vision using CNNs were presented (Gatys et al. 2017).

Study by (Mohammad and Kiritchenko 2018), shows a dataset of more than 4,000 pieces of art from WikiArt.org's collection that has annotations for emotions evoked in the observer was created. Additionally, the art is also annotated for whether it includes the depiction of a face and how much the observers like the art. They showed that the title often impacts the effectual response to art and that the paintings that depict faces draw more consistent emotional responses than those that do not. Additionally, it was shown for each art category and emotion combination, the average agreements on the emotions evoked and the average art ratings.

Vilalta (2016) conducted research that studied the possibilities that the DNN used for the purpose of creation of artwork were analysed. They were using Japanese anime images for the training. A special CNN that encodes a plain colours effect while keeping sharp edges was used originally designed for super-resolution. As As investigated by Dushkoff et al (2017) Art500k, a large-scale visual arts dataset is presented that contains over 500,000 art works, annotated with detailed labels of artist, art movement, genre, etc. In order to achieve better representation of visual arts, they proposed an efficient two stage triplet sampling method. They compared the joint representations with hand-craft visual features and learned visual features. The hand-craft features are Color visual features, GIST visual features (Oliva and Torralba 2001), SIFT-like visual features (Lowe 1999) and SIFT-like Fisher visual features (Perronnin et al. 2010). They were using VGG-16 architecture (Simonyan and A. Zisserman 2014) with the ImageNet pre-trained weights. The best achieved result was 39,1%.

III. THE PROPOSED METHODOLOGY

Deep convolutional neural networks have recently played a transformative role in advancing artificial intelligence. We evaluated a large number of state-of-the-art deep convolutional neural network models, and variants of them, trained to classify styles. We focused on increasing the interpretability of the learned presentation by forcing the machine to achieve classification with a reduced number of variables without sacrificing classification accuracy. We then analyzed the achieved representations through linear and nonlinear dimensionality reduction of the activation space. We used a collection of 25K digitized paintings to train, and validate, also we used collection of 10k digitized paintings to test model . We utilized two sets of digitized paintings for visualization and correlation analysis of the achieved representations. In particular, we used variants of AlexNet , VGG16, ResNet, Inception,which were originally developed for the task of object categorization for the our datasets and each of them raised the state of the art in that task when they were introduced. We adapted these networks for classifying 10 style classes. Our study included varying the training strategies (training from scratch on art data vs. using pre-trained models and fine-tuning them on art data), varying the network architecture, and data augmentation strategies.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Painting style database

The images of visual art were mainly scraped from four web-sites: Wikiart[2] Training-testing Set: We trained, validated, and tested the networks using paintings from the publicly available WikiArt data-set1. This collection (as downloaded in 2015) has images of 81,449 paintings from 10 artists ranging from the fifteenth century to contemporary artists. Several prior studies on style classification used subsets of this dataset

Originally WikiArt has 27 style classes. For the purpose of our study we reduced the number of classes to 10 classes by merging fine-grained style classes with small number of images, as shown in Figureure 1.

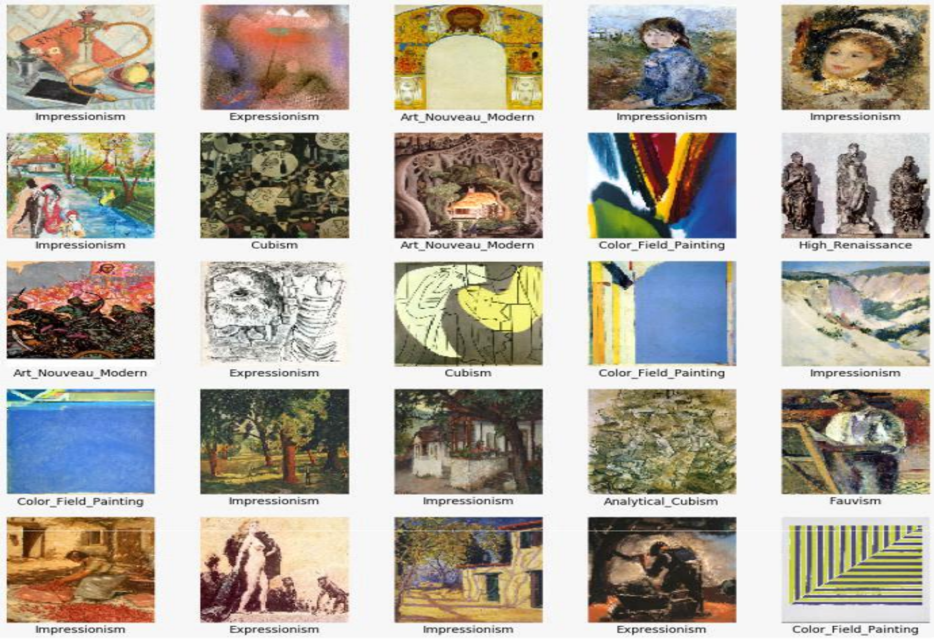
We excluded from the collections images. The total number of images used for training, validation, and testing are 35k images. We split the data into training, validation and test sets.

B. Experimental results

In this paper, we want to answer the following questions. Q1: How different neural networks architectures do to recog-nize image style? Q2: What proportion of the network should be retrained to obtain the best performance? Q3: What is the impact of the optimizations and how does it perform in our data sets?

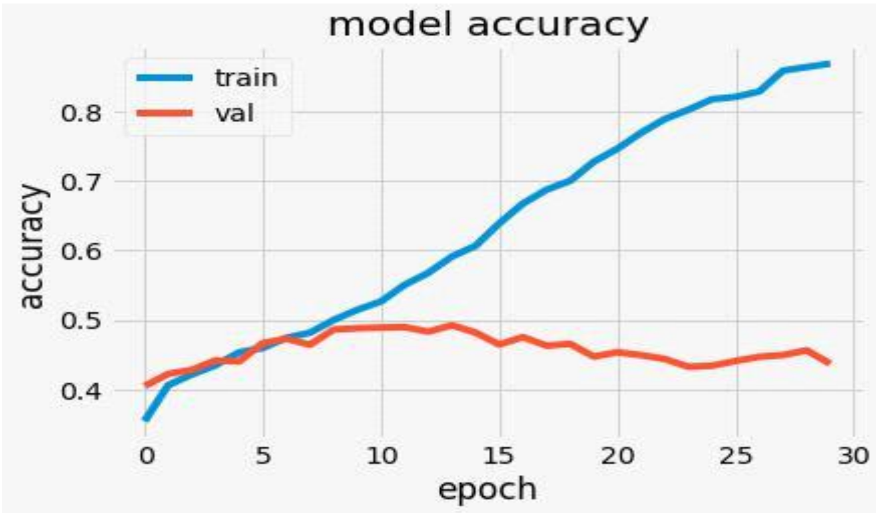
We used different CNNs architectures , first CNN based on AlexNet architecture which used in (Karayev et al., 2014), accuracy for train and test data of this CNN, we are going to mention it later, also the rest of CNNs architectures that we used such as, the state-of-the-art residual neural network as described in (He et al., 2016) (ResNet),VGG16 (also called OxfordNet) is named after the Visual Geometry Group from Oxford, then Inception and so on. Then we will camper there accuracy with our built in models. Our implement for dataset follows VGG16, AlexNet, Inception, LeNet5,ResNet, The size of images resized [100,100]. the Figure 1 shows samples of different art style that we used in our Dataset.

Figure1. Samples of different art style that we used in dataset



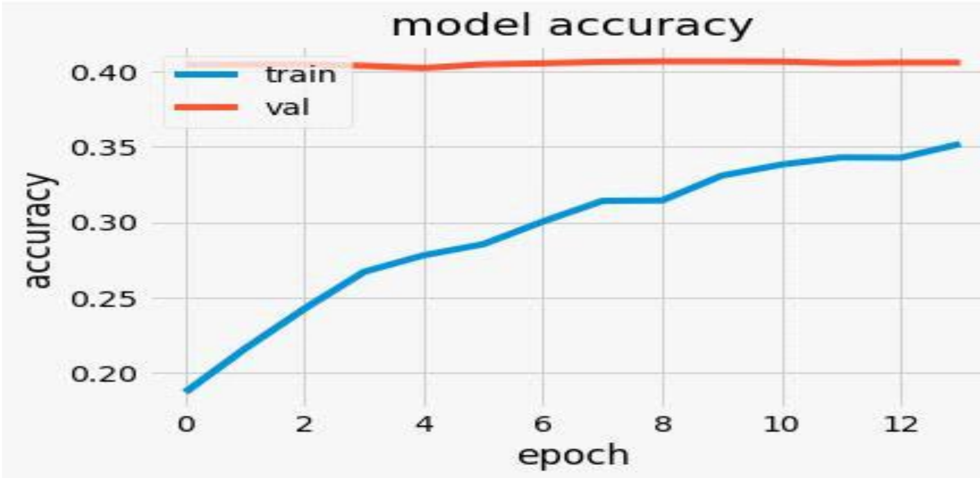
AlexNet: The input of the AlexNet network is a 100 100 3 matrix. Its structure is made of five convolutional layers and three fully connected layers. Convolutional layers have respectively filter of size 11 11, 5 5, 3 3, and 3 3. They each respectively generate 96, 256,384, 384, and 256 feature maps.Three max-pooling layers follow the first, third and fifth convolutional layer. Then there are 4096 neurons in the first fully connected layer, then there are 1000 neurons in the second fully connected layer. Dropout was not implemented after each of those two fully-connected layers.because we found that we got better results than if we implemented it.The accuracy result of AlexNet CNN shown in Figureure 2.

Figure2. Accuracy of train and validation data for AlexNet CNN



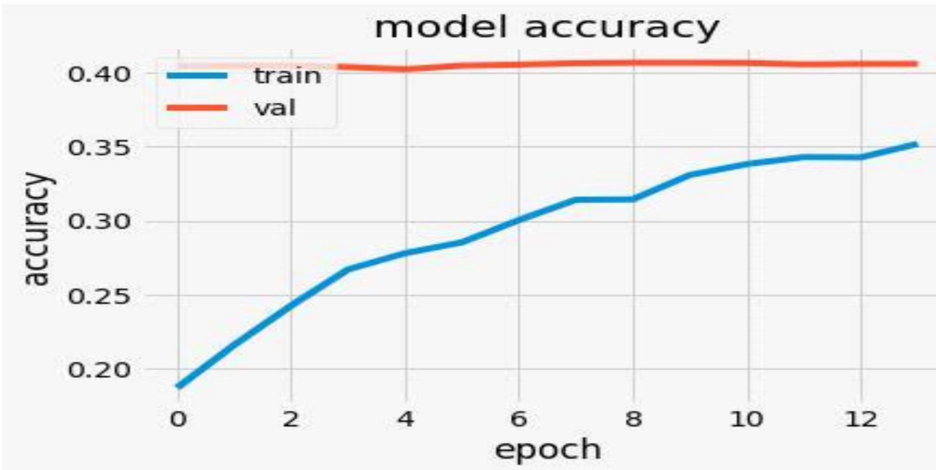
ResNet:The ResNet has input dimensions of 100 100 3.The ResNet starts with a convolutional layer with a filter size of 3x3, generating 32 filters, followed by a batchnormalization layer, an activation layer and a max-pooling layer. Then they followed by Dropout, rectified Linear Unit (ReLU) is used as the activation function for all weight layers, except for the last layer that uses softmax regression. A fully-connected.layer ends each network, of which the number of neurons corresponds to the number of classes. The accuracy result of RestNet CNN shown in Figure3.

Figure3. Accuracy of train and validation data for ResNet CNN



VGG16: is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition”.The input to cov1 layer is of fixed size 100 x 100 RGB image. The image is passed through convolutional layers, where the filters were used 33. The convolution stride is fixed to 1 pixel. the padding is 1-pixel for 33 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 22 pixel window, with stride 2. Three Fully-Connected (FC) layers (which has a different depth in different architectures): the first two have 4096 channels each, the third has 1000 channels. The final layer is the soft-max layer.the accuracy result that we got by applying our data-set on VGG16 model shown in Figure 4.

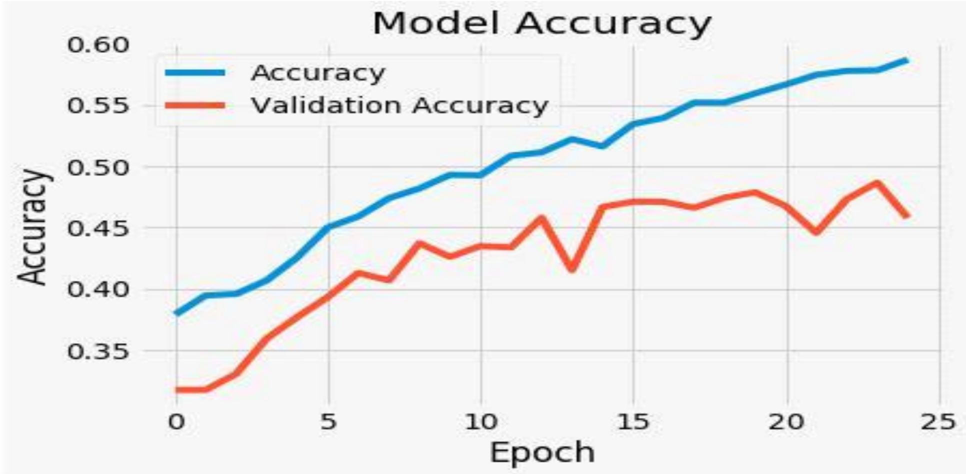
Figure4. Accuracy of train and validation data for VGG16 CNN



VGG16 is over 533MB. This makes VGG16 is annoying task.VGG16 is used in many deep learning image style classi-fication problems,A modification to the module was made in order to reduce the amount of computation required; therefore, smaller network architectures are sometimes more sufficient way (such as LeNet5, etc.). But it is a magnificent building block for deep learning objectives. Inseption:The inception CNN was described and used in the GoogLeNet model in the 2015 paper by Christian Szegedy, et al. titled “Going Deeper with Convolutions.”

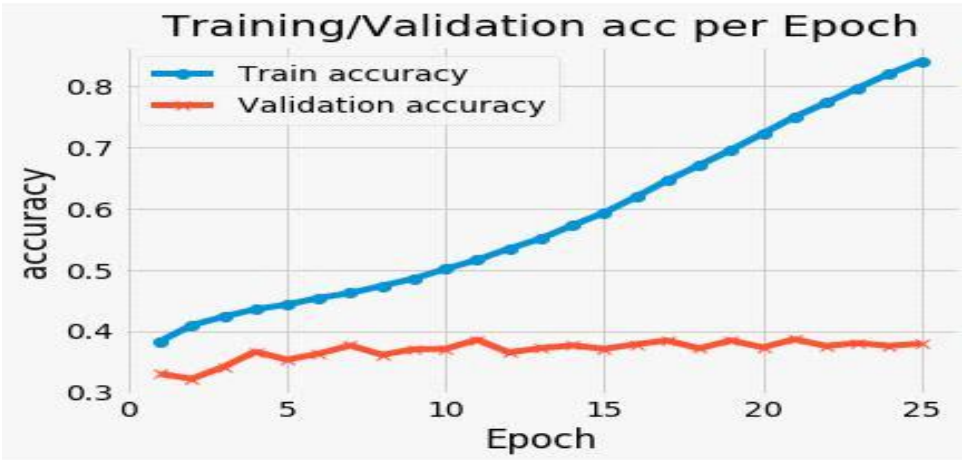
The inception model is called the inception module. This is a block of parallel convolutional layers with different sized filters (e.g. 11, 33, 55) and a and 33 max pooling layer, the results of which are then concatenated.pooling operations have been essential for the success of current convolutional networks.This is a very simple and powerful architectural.T he accuracy of train and validation data of Inseption are shown in Figure 5.

Figure5. Accuracy of train and val data for Inseption CNN



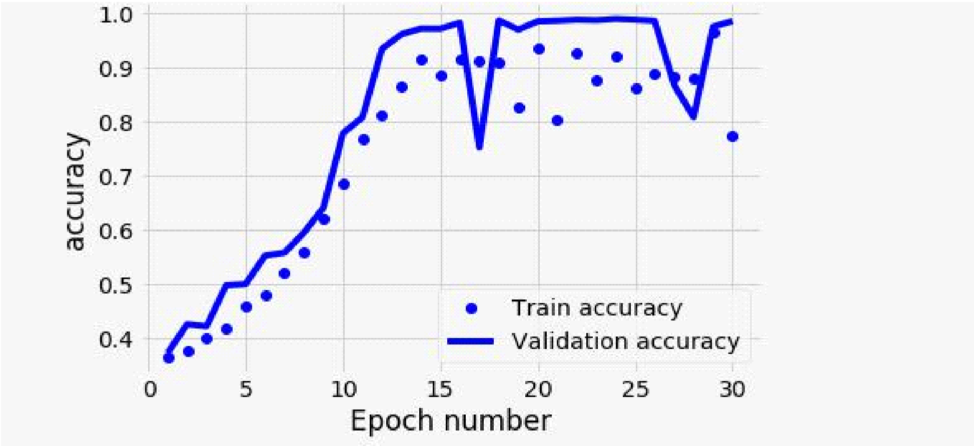
LeNet5:The LeNet-5 architecture consists of two sets of convolutional and average pooling layers, followed by a flat-tening convolutional layer, then two fully-connected layers and finally a softmax classifier.First Layer (C1) input is a 3232 gray-scale image, but we motivated to 100 x 100 x3 RGB image,to make it work with our data-set,filter size 55 and a stride of one.Second Layer applies(S2) average pooling layer with a filter size 22 and a stride of two.Third Layer (C3), there is a second convolutional layer with 16 feature maps AND filter size 55 and a stride of 1. The fourth layer (S4) is an average pooling layer with filter size 22 and a stride of 2.The fifth layer(C5) is a fully connected convolutional layer with 120 feature maps.The sixth layer is a fully connected layer (F6) with 84 nodes.Finally output Layer is a fully connected softmax output layer with 10 possible values corresponding to the art style.After we applied it with our data-set we got an accuracy of train and validation data set as shown in Figure 6.

Figure 6. Accuracy of train and val data for LeNet5 CNN



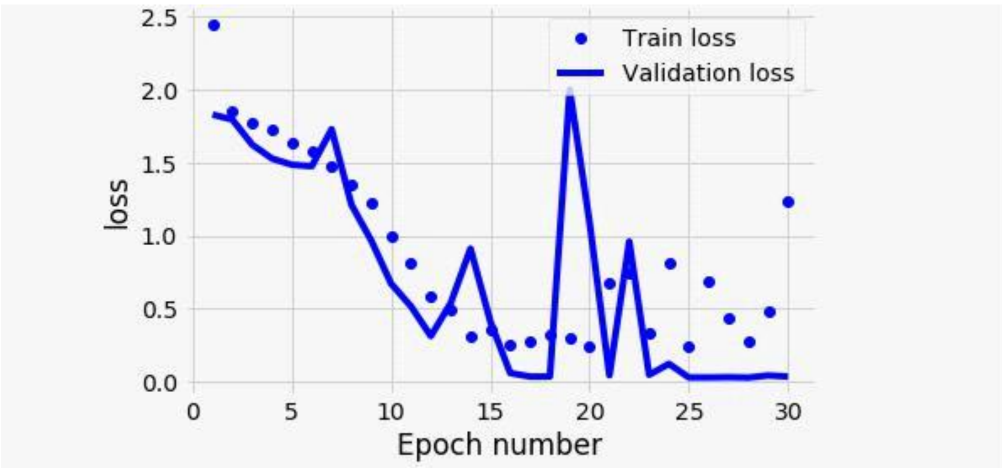
Here we going to take about deep convolution neural networks architectures of our model, from all models of CNNs that we used before,The objective of these model is to have a final model that performs well both on the train data and test data (the new data which the model will be used to make predictions).during our implementation we got problems with overfitting.to reduce overfitting by training the network on more examples.but Each training example requires heavy computations, which this lead us to another big problem. these models. The problems that we got unfortunately, there are two major drawbacks with any of CNNs modules used: 1.It is painfully slow to train 2.The network archi-tecture weights themselves are quite large (concerning disk/bandwidth). Therefore, in these deep convolution neural networks ar-chitectures,the reason of we used different architectures, to see if we can get better performance by using different methods and solving the problems that we got by implementation (VGG16,Inseption, LeNet5, AlexNet),such as reducing the number of filters, Applying regularization to our model, dropout layers, and optimal learning rate. The following Figureures show the performance on train and test data, after applying these methods. The next Figureure (Figure 7) show up the performance of our model before we applied all methods that we had already mentioned before

Figure 7. Accuracy of train and val data for Base model



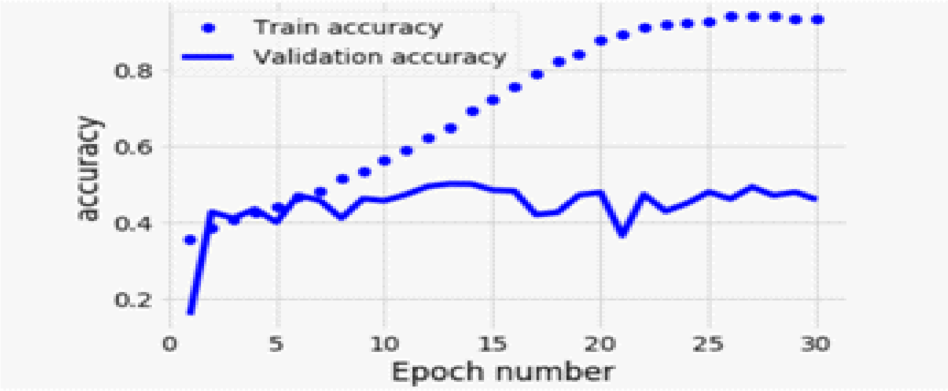
The next Figure (Figure 8) show up the loss of our model in training and validation dataset, before we applied all methods that we had already mentioned before.

Figure 8. Accuracy of train and val data for Base model



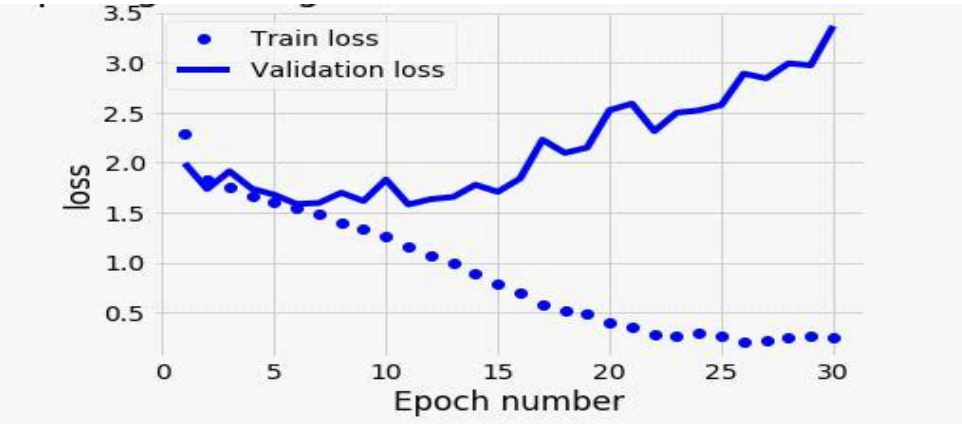
Next Figure (Figure 9) shows the performance of the base model after reducing the number of filters.

Figure 9. Accuracy of train and val data for reduce model



Next Figure (Figure 10) shows the loss of the base model after reducing the number of filters.

Figure 10. Loss of train and val data for reduce model



The next two Figures (Figure 11 and Figure 12) show the per-formance of the base model after applying regularization method.Figure 11 for the accuracy of training and validation dataset.and Figure12 for the loss of training and validation dataset.

Figure 11. Accuracy of train and val data for regularization model

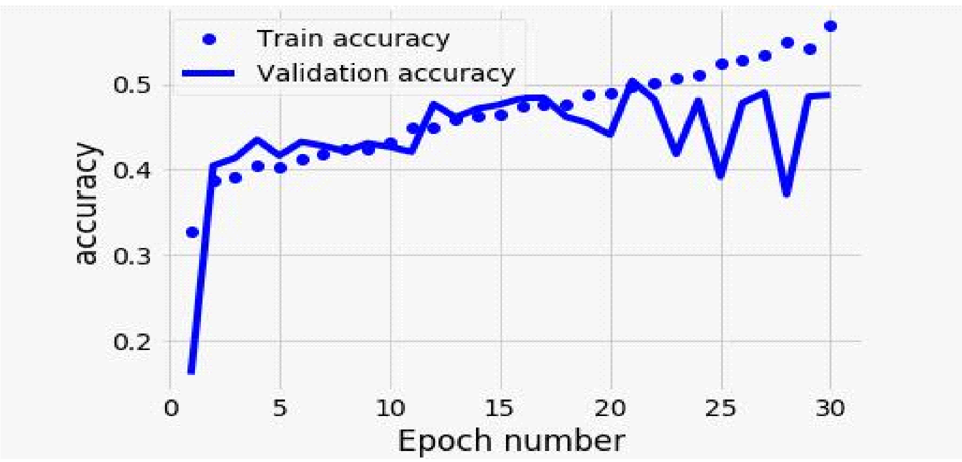
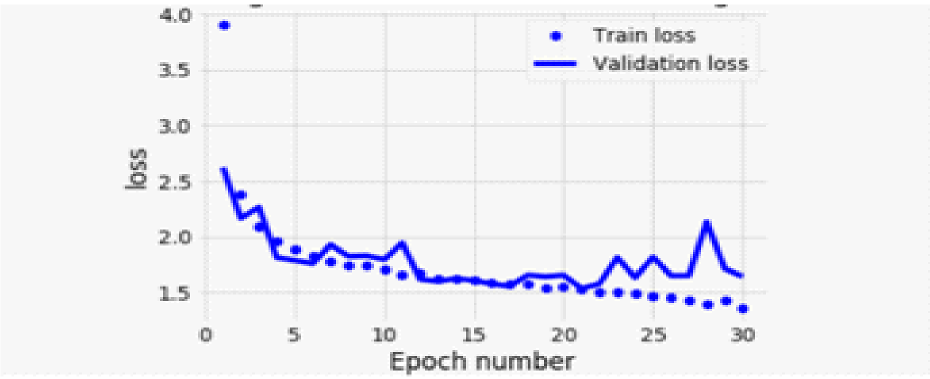


Figure 12. Loss of train and val data for regularization model



The next two Figures (Figure 13 and Figure 14) show the per-formance of the base model after Adding dropout layers method.Figure 13 for the accuracy of training and validation dataset.and Figure14 for the loss of training and validation dataset.

Figure 13. Accuracy of train and val data for Dropout model

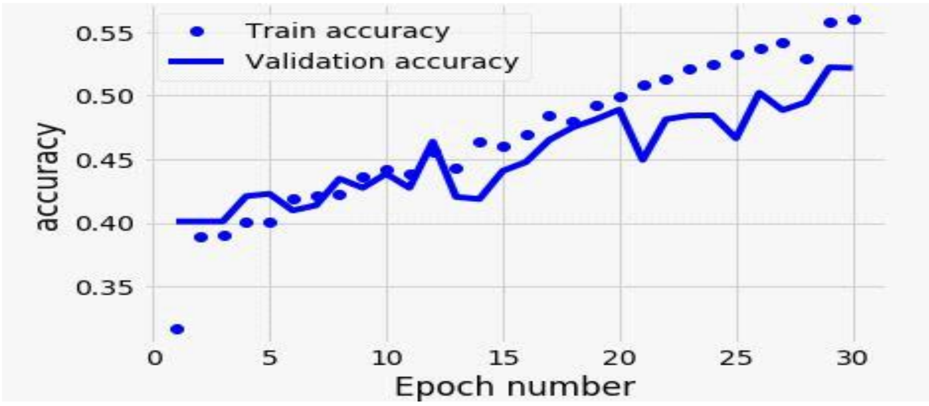
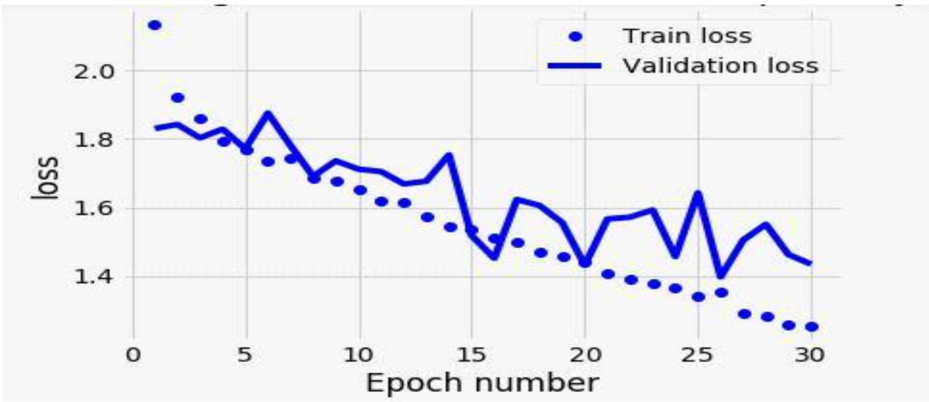


Figure 14. Loss of train and val data for Dropout model



Now we are going to present Table 1 that shows the classification performance for the train and test data of all CNNs that we used in this paper.

TABLE 1 State-of-the-art art style prediction methods

Method		Database		# of Classes	Contains Special Classes (y/n)		Accuracy (training,test data)(%)	
AlexNet		Our database		10	n		72.75%_38.90%	
Inception$_{v}$3		Our database		10	n		45.25%_31.74%	
LeNet5		Our database		10	n		83.97%_37.97%	
ResNet		Our database		10	n		35.20%_31.84%	
VGG16		Our database		10	n		97.70%_46.48%	
Base-Model		Our database		10	n		96.51%_40.01%	
Reduced-Model		Our database		10	n		94.34%_39.49%	
Regl2-Model		Our database		10	n		56.9%5_40.06%	
Drop-Model		Our database		10	n		56.58%_44.27%	

V. CONCLUSION

As you can see, knowledge of machine learning is not for a simple image classification task. Nowadays, we have got to have huge amounts of data and huge computational resources for painting style classification tasks.

By the time, Deep learning methodologies are improving not only on image classification, but also on other machine learning applications such as speech recognition, natural language processing, and so on. However, deep learning still has some problems, such as the huge computation time, training difficulties, and huge number of parameters. If we could represent deep learning by fewer parameters, it will be even more efficient. So far, one day we will be able to solve those problems and provide a better user experience through machine learning!

ACKNOWLEDGEMENT

This work has been partially supported the Scientific and Technological Research of Singidunum University of Belgrade Serbia. The principal researcher would also like to thank the Libyan Ministry of Higher Education and Tripoli University for the PhD sponsorship and study leave respectively.

REFERENCES

[1]. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[2]. Dushkoff, M., McLaughlin, R., & Ptucha, R. (2016, December). A temporally coherent neural algorithm for artistic style transfer. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 3288-3293). IEEE.

[3]. Gatys, L. A., Ecker, A. S., & Bethge, M. (2017). Texture and art with deep neural networks. *Current opinion in neurobiology*, *46*, 178-186.

[4]. Hicsonmez, S., Samet, N., Sener, F., & Duygulu, P. (2017, June). DRAW: Deep networks for Recognizing styles of Artists Who illustrate children's books. In *Proceedings of the 2017 ACM on international conference on multimedia retrieval* (pp. 338-346).

[5]. Lecoutre, A., Negrevergne, B., & Yger, F. (2017, November). Recognizing art style automatically inpainting with deep learning. In *Asian conference on machine learning* (pp. 327-342). PMLR.

[6]. Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision* (Vol. 2, pp. 1150-1157). Ieee.

[7]. Mohammad, S., & Kiritchenko, S. (2018, May). Wikiart emotions: An annotated dataset of emotions evoked by art. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.

[8]. Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145-175.

[9]. Perronnin, F., Liu, Y., Sánchez, J., & Poirier, H. (2010, June). Large-scale image retrieval with compressed fisher vectors. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3384-3391). IEEE.

[10]. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[11]. Vilalta Arias, A. (2016). Generation of art works using deep neural networks. Master thesis. [Online]. Available: http://digital.csic.es/handle/10261/ 155244

[12]. "Wikiart dataset," https://www.wikiart.org/.

Rabeeah Ahmed Meelad, et. al. "Deep Neural Networks based Classical Painting Style Prediction." *International Journal of Engineering and Science*, vol. 11, no. 5, 2021, pp. 18-26.