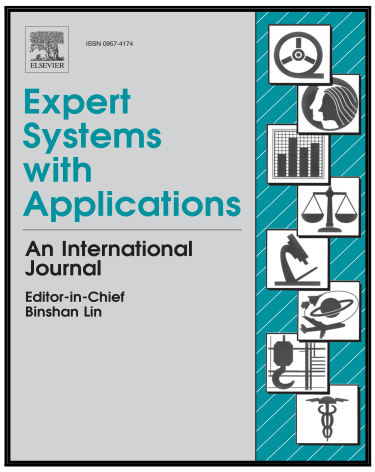

Fine-tuning Convolutional Neural Networks for Fine Art Classification

Eva Cetinic, Tomislav Lipic, Sonja Grgic

PII: S0957-4174(18)30442-1
DOI: 10.1016/j.eswa.2018.07.026
Reference: ESWA 12079

To appear in: *Expert Systems With Applications*

Received date: 20 February 2018
Revised date: 9 July 2018
Accepted date: 10 July 2018

Please cite this article as: Eva Cetinic, Tomislav Lipic, Sonja Grgic, Fine-tuning Convolutional Neural Networks for Fine Art Classification, *Expert Systems With Applications* (2018), doi: 10.1016/j.eswa.2018.07.026

Highlights

- We achieve state-of-the-art results for five fine art-related classification tasks
- Different convolutional neural network fine-tuning strategies are explored
- Impact of various source domain-dependent weight initialization is studied
- Networks pre-trained for scene and sentiment recognition perform best for art tasks
- Fine-tuned models can be used to retrieve images similar in style or content

Fine-tuning Convolutional Neural Networks for Fine Art Classification

Eva Cetinic[a,*], Tomislav Lipic[a], Sonja Grgic[b]

[a] *Rudjer Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia*
[b] *University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia*

Abstract

The increasing availability of large digitized fine art collections opens new research perspectives in the intersection of artificial intelligence and art history. Motivated by the successful performance of Convolutional Neural Networks (CNN) for a wide variety of computer vision tasks, in this paper we explore their applicability for art-related image classification tasks. We perform extensive CNN fine-tuning experiments and consolidate in one place the results for five different art-related classification tasks on three large fine art datasets. Along with addressing the previously explored tasks of artist, genre, style and time period classification, we introduce a novel task of classifying artworks based on their association with a specific national artistic context. We present state-of-the-art classification results of the addressed tasks, signifying the impact of our method on computational analysis of art, as well as other image classification related research areas. Furthermore, in order to question transferability of deep representations across various source and target domains, we systematically compare the effects of domain-specific weight initialization by evaluating networks pre-trained for different tasks, varying from object and scene recognition to sentiment and memorability labelling. We show that fine-tuning networks pre-trained for scene recognition and sentiment prediction yields better results than fine-tuning networks pre-trained for object recognition. This novel outcome of our work suggests that the semantic correlation between different domains could be inherent in the CNN weights. Additionally, we address the practical applicability of our results by analysing different aspects of image similarity. We show that features derived from fine-tuned networks can be employed to retrieve images similar in either style or content, which can be used to enhance capabilities of search systems in different online art collections.

Keywords: painting classification, deep learning, convolutional neural networks, fine-tuning strategies

1. Introduction

Large-scale digitization efforts which took place in the last two decades led to a significant increase of online accessible fine art collections. The availability of those collections makes it possible to easily explore and enjoy artworks which are scattered within museums and art galleries all over the world. The increased visibility of digitized artworks is particularly useful for art history education and research purposes. Apart from the advantages of the visibility boost, the very translation of information, from the domain of the physical artwork into the digital image format, plays a key role in opening new research challenges in the interdisciplinary field of computer vision, machine learning and art history.

The majority of available online collections include some particular metadata, usually in the form of annotations done by art experts. Those annotations mostly contain information about the artist, style, genre, technique, date and location of origin, etc. Art experts can easily identify the artist, style and genre of a painting using their experience and knowledge of specific features. However, a great current challenge is to automate this process using computer vision and machine learning techniques. Generating metadata by hand is time consuming and requires the expertise of art historians. Therefore, automated recognition of artworks' characteristics would enable not only a faster and cheaper way of generating already existing categories of metadata such as style and genre in new collections, but also open the possibility of creating new types of metadata that relate to the artwork's content or its specific stylistic properties.

Stylistic properties of paintings are abstract attributes inherent to the domain of human perception. Analysing artworks is a complex task which involves understanding the form, expression, content and meaning. All those components originate from the formal elements of paintings such as line, shape, colour, texture, mass and composition (Barnet, 2011). The translation of those semantically charged features into meaningful numerical descriptors remains a great challenge. Most of the research done in the field of computational fine art classification is based on extracting various low-level image features and using them for training different types of classifiers. However, recent

*Corresponding author
Email addresses: ecetinic@irb.hr (Eva Cetinic), tlipic@irb.hr (Tomislav Lipic), sonja.grgic@fer.hr (Sonja Grgic)

breakthroughs in computer vision achieved by deep convolutional neural networks, demonstrate the dominance of learned features in comparison to engineered features for many different image classification tasks (Krizhevsky et al., 2012).

One of the main arguments for the recent success of deep CNNs in solving computer vision tasks is the availability of large hand-labelled datasets such as the ImageNet dataset ([dataset] Deng et al., 2009), which consists of over 15 million hand-labelled high-resolution images, covering approximately 22,000 different object categories. If we aggregated all the digitized paintings in all available online collections, the number of images would still be considerably smaller than the number of images in the ImageNet dataset and not adequate to train a deep CNN from scratch without over-fitting. However, many different image-related classification tasks (Reyes et al., 2015; Tajbakhsh et al., 2016), which deal with datasets of limited size, managed to achieve state-of-the-art classification performance by fine-tuning CNNs pre-trained on the ImageNet dataset to the new target dataset and/or task. This motivated us to explore how CNNs pre-trained on photographic images can be fine-tuned for fine art specific tasks such as style, genre or artist recognition.

In our work we explore how different fine-tuning strategies can be used for various art-related classification tasks. Knowing that a smaller distance between the source and target domains leads to a better performance on the new task (Yosinski et al., 2014), we investigate the impact of different weight initializations by using CNNs of the same architecture, but pre-trained on different source domains and for different tasks. By changing the transfer learning source domain, we are trying to explore how different task- and data-driven weight initializations influence the performance of fine-tuned CNNs for art-specific tasks and whether this can indicate a semantic correlation between domains. Besides weight initialization, we also address several other aspects of the fine-tuning process such as the number of layers being re-trained. Furthermore, we show how models fine-tuned for solving a particular classification task can be used to broaden the possibilities of content-based search across art datasets.

2. Related work

The topic of fine art classification has been addressed with continuous interest in a number of different studies over the last few years. One of the first attempts to classify paintings was done by Keren (2002), applying a naive Bayes classifier to local features derived from discrete cosine transformation coefficients. The task of classifying paintings by artist has later been addressed in different studies (Cetinic & Grgic, 2013), as well as the challenge of visualizing similarities (Bressan et al., 2008; Shamir et al., 2010; Shamir & Tarakhovsky, 2012) and exploring influential connections among artists (Saleh et al., 2016). Most of the earlier studies that addressed the topic of artist and other art-related tasks such as style (Lombardi, 2005; Arora & Elgammal, 2012; Falomir et al., 2018) and genre classification (Zujovic et al., 2009), share one similar methodology. Their approach usually includes extracting a set of various image features and using them to train different classifiers such as support vector machines (SVM), multilayer perceptron (MLP) or k-nearest neighbours (k-NN). The features used for training the classifiers commonly include low-level features that capture shape, texture, edge and colour properties. A comprehensive overview of these earlier studies and other uses of computational methods in art history is given in Brachmann & Redies (2017).

The fine art classification challenge faced several common issues, most notably the lack of a large commonly accepted dataset to adequately compare results. Studies addressing the same classification task used different small to medium-sized collections of paintings, as well as arbitrary chosen and different sets of classification classes. Recently the fine art classification research progress was induced by two parallel streams: the appearance of large, well-annotated and openly available fine art datasets on one side; and significant advancements in computer vision related tasks achieved with the adoption of convolutional neural networks on the other side.

In the context of fine art classification, CNNs were firstly introduced as feature extractors. The approach to use layers' activations of a CNN trained on ImageNet as features for artistic style recognition was introduced by Karayev et al. (2014), where authors showed how features derived from the layers of a CNN trained for object recognition on non-artistic images, achieve high performance on the task of painting style classification and outperform most of the hand-crafted features. The efficiency of CNN-based features, particularly in combination with other hand-crafted features, was confirmed for style (Bar et al., 2014), artist (David & Netanyahu, 2016) and genre classification (Cetinic & Grgic, 2016), as well as for other related tasks such as recognizing objects in paintings (Crowley & Zisserman, 2014). Even better performance for a variety of visual recognition tasks has been achieved by fine-tuning a pre-trained network on the new target dataset as shown by Girshick et al. (2014), as opposed to using CNNs just as feature extractors. The superiority of this approach has also been confirmed on artistic datasets for different classification tasks (Hentschel et al., 2016; Tan et al., 2016), as well as for retrieving visual links in painting collections (Seguin et al., 2016) or distinguishing illustrations from photographs (Gando et al., 2016).

Although fine-tuning does not require as much data as training a deep CNN from scratch, a relatively large corpus of images is still considered a necessary prerequisite. Fortunately the appearance of large, annotated and online available fine art collections such as the WikiArt[1] dataset,

[1] www.wikiart.org

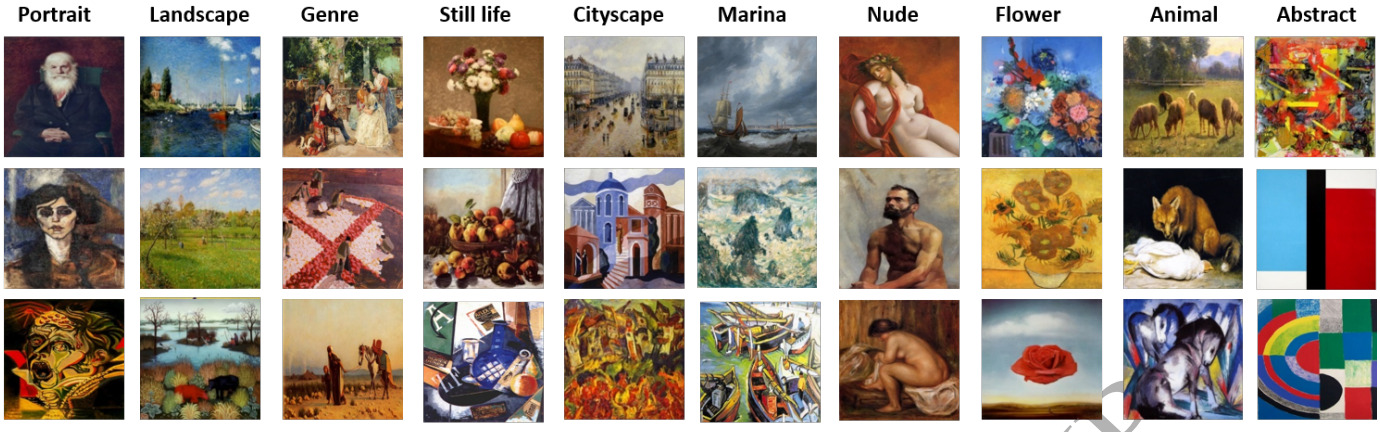

Figure 1: Examples of paintings from ten different categories included in the Wikiart genre classification dataset

which contains more than 130k artwork images, enabled the adoption of deep learning techniques, as well as helped shaping a more uniform framework for method comparison. To the best of our knowledge, the WikiArt dataset is currently the most commonly used dataset for art-related classifications tasks (Karayev et al., 2014; Bar et al., 2014; David & Netanyahu, 2016; Girshick et al., 2014; Hentschel et al., 2016; Seguin et al., 2016; Chu & Wu, 2016; Saleh & Elgammal, 2016), even though other online available sources are also being used such as the Web Gallery of Art[2] (WGA) with more than 40k images (Seguin et al., 2016); or the Rijksmuseum challenge dataset (van Noord et al., 2015; Mensink & Van Gemert, 2014). Furthermore, there were several initiatives for building painting datasets dedicated primarily to fine art image classification such as Painting-91 (Khan et al., 2014), which consists of 4266 images from 91 different painters; the Pandora dataset consisting of 7724 images from 12 art movements (Florea et al., 2016) and the recently introduced museum-centric OmniART dataset with more than 1M photographic reproductions of artworks (Strezoski & Worring, 2017).

Based on the datasets used, as well as the methodology and results, we identify several particularly interesting works for comparison. Saleh & Elgammal (2016) initiated the use of WikiArt for creating data sub-collections for the tasks of artist, style and genre classification, as well as identified the classes for each task based on the number of available images. In their work they explore how different image features and metric learning approaches influence the classification performance. Regarding the used features, their best result is achieved with the feature fusion method which included also CNN-based features. Consequently, based on the same dataset and class distribution, Tan et al. (2016) fine-tuned an ImageNet pre-trained CNN and achieved significant performance improvement, as well as showed that fine-tuning a CNN not only outperforms using only CNN-based features, but also exceeds the results achieved by training a CNN from scratch with fine art images. Similarly, Hentschel et al. (2016) also showed that fine-tuned CNNs yield best results for the task of style classification on the WikiArt dataset, in comparison to other approaches such as linear classifiers applied on Improved Fisher Encodings. In both of these works the results were obtained by fine-tuning an AlexNet model (Krizhevsky et al., 2012). More recently, Lecoutre et al. (2017) managed to achieve a higher performance for the style classification task by fine-tuning the deeper ResNet50 model (He et al., 2016). This indicates that further classification improvement on other tasks might be also be achieved using deeper architectures. Apart from fine-tuning, an approach for learning scale-variant and scale-invariant representations from high-resolution images of the TICC[3] dataset was presented by van Noord & Postma (2017). By designing a multi-scale CNN architecture consisting of multiple single-scale CNNs, they managed to achieve very high performance for the task of artist classification and present a method that is particularly useful for tasks involving image structure at varying scales and resolutions.

The main methodological novelty of our approach is comprised in our attempt to not only outperform current classification results, but also investigate the semantic correlation between art-related tasks and other domain-specific tasks. To achieve this, we limit our choice of CNN architecture and concentrate on investigating the different domain-specific weight initialization and fine-tuning scenarios impact. The results show that our fine-tuning approach outperforms the current state-of-the-art achieved with this particular CNN architecture. A detailed comparison of experimental results regarding the tasks, number of classes and classification accuracy is presented in section 5.3.

3. Datasets and classification tasks

With the aim to include the largest possible number of paintings, as well as to cover a wide range of classi-

[2]www.wga.hu

[3]https://auburn.uvt.nl/

fication tasks, we use three different sources for creating our datasets and identifying the classification tasks. Our first source is WikiArt, the currently largest online available collection of digitized paintings. WikiArt is a well-organized collection which integrates a broad set of metadata such as artist, style, genre, nationality, technique, etc. It includes artworks from a wide time period, with a particular focus on 19th and 20th century, as well as contemporary art. Because of its extensiveness, WikiArt is a frequent choice for creating datasets in many of the recent studies that addressed the question of painting classification and is therefore suitable for results comparison. The dataset is continuously growing and includes different types of artworks such as paintings, sculptures, illustrations, sketches, photos, posters, etc. At the time of our data collection process, the WikiArt dataset contained 133220 artworks in total. However, to be consistent regarding the type of artwork and therefore more eligible for exploring the challenge of different classification tasks, we included only paintings and drawings when creating our data subsets. Therefore when building the classes subsets, we made sure to remove artworks that are classified as architecture, photography, poster, graffiti, installation, etc. Particularly, when choosing the classes for the task of genre classification, we also made sure to be consistent with what the term "genre" refers to in the traditional division of paintings, namely the type of content depicted. Therefore we focus exclusively on genre categories which correspond to specific objects or scenes, rather than including categories such as illustration or sketch and study which are included in the WikiArt genre set of annotations and included in the genre classification task performed by Tan et al. (2016). Examples of different images for the selected genre classes can be seen in Figure 1.

In total we defined four classification tasks performed on the WikiArt dataset: genre, style, artist and artist's nationality. Recognizing the artist, genre and style of paintings are three commonly addressed tasks, but the task of classifying paintings by the artist's nationality has, as far as we know, not yet been undertaken and represents an interesting challenge. It explores an underlying interrelationship between artworks from different artists, genres and time periods, but belonging to the same national artistic context.

Based on the number and distribution of images, as well as the number of classes used in previous works (Saleh & Elgammal, 2016), we define the subset of classes for each task. In particular, for artist classification we use a subset of 23 artists, where each artist is represented with at least 500 paintings. For style we use a subset of 27 classes where each class has more than 800 paintings, for genre a subset of 10 classes where each class has more than 1880 paintings and for nationalities we use a subset of 8 classes with at least 3200 samples per class. The distribution of number of images per class can be seen in Figure 2 (left) for the WikiArt style subset and in Figure 2 (right) for the WikiArt genre subset. The complete list of the classes and number of samples per class for all prepared datasets is given in the Supplementary material.

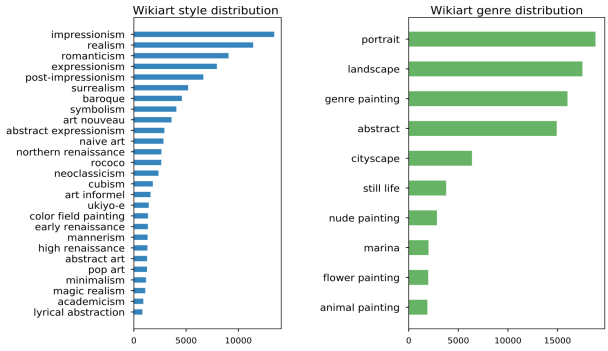

Figure 2: Class distribution of the WikiArt style (left) and WikiArt genre (right) datasets

Furthermore, we explore another online source of paintings – the Web Gallery of Art (WGA). This collection is not as commonly used as the WikiArt dataset and has a different historical distribution of paintings, covering fine arts from the 8th to 19th century, with a notably extensive selection of mediaeval and renaissance artworks. Similarly as in the WikiArt dataset, paintings are labelled with genre, art historical period, school and timeframe (in 50 years steps) in which the artists were active. The collection contains various types of artworks and for our purpose we used a subset of 28952 paintings. Based on the available metadata, we identified the following tasks for classification: artist, genre, nationality (school) and timeframe. The timeframe classification task can be considered most similar to the task of style classification because style is usually linked with an artistic movement active in a specific time period. However, the WGA timeframe distribution is specified by a 50 years time step which might include overlapped artistic movements and can therefore not be considered as a strict equivalent to the task of style classification. A detailed distribution of images per timeframe within the WGA collection can be seen in Figure 3.

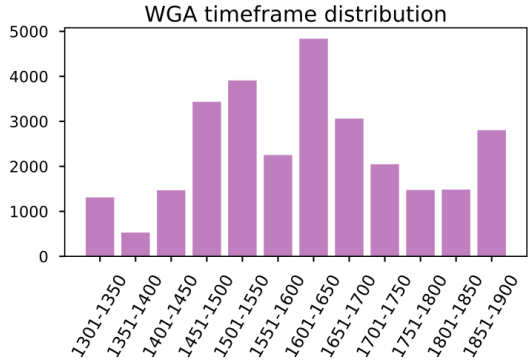

Figure 3: Class distribution of the WGA timeframe dataset

The WGA timeframe subset consists of 12 classes with more than 500 images per class. For the WGA genre subset

we selected 6 classes with more than 1000 images per class; for the WGA artist subset we took 23 classes with more than 170 images per class and for the nationality subset 8 classes with more than 500 images per class.

WikiArt and WGA both include digitized paintings from different sources which vary in size and quality. As the third data source, we used the TICC Printmaking Dataset (van Noord & Postma, 2017), which is essentially different from the other two datasets because it contains high-resolution digital photographic reproductions of prints made on paper from the online collection of the Rijksmuseum, the Netherlands state museum. The dataset includes 58630 reproductions of artworks made by 210 different artists where each artist is represented with at least 96 artworks. Having only prints included, this dataset is more uniform in terms of colour and physical size and therefore suitable for addressing the task of artist classification on a less style-dependent level. Examples of representative images from the three different data sources are shown in Figure 4, while the total number of images per task and dataset is given in Table 1.

Table 1: Number of images and classes for different tasks and data sources

Task	Source	# of classes	# of images
artists	TICC	210	58,630
	WikiArt	23	20,320
	WGA	23	5,711
genre	WikiArt	10	86,087
	WGA	6	26,661
style	WikiArt	15	96,014
timeframe	WGA	12	28,605
nationality	WikiArt	8	80,428
	WGA	8	27,460

For each data source and task we split the total number of images in order to keep 70 % of the images for training, 10 % for validation and 20 % for testing. This distribution is kept consistent within all classes. All images are resized to 256 × 256 pixels.

4. Experimental setup

4.1. CNN architecture

The main CNN architecture used in our experiments is CaffeNet (Jia et al., 2014), which is a slightly modified version of the AlexNet model (Krizhevsky et al., 2012). This CNN contains five convolutional layers and three fully connected layers. Three max-pooling layers follow after the first, second and fifth layer, while dropout is implemented after the first two fully connected layers. The activation function for all weight layers is the rectification linear unit (ReLU). The output of the last fully connected layer is connected to a softmax layer that determines the probability for each class. The input of the network is a 227 × 227 crop of the resized RGB image.

Besides the aim to maximize classification performance for art-related tasks, we explore the transferability of deep representations across different source/target domains. For this purpose, we narrow our choice of architecture to one well-studied architecture such as AlexNet, rather than expanding our fine-tuning experiments to deeper architectures such as VGG (Simonyan & Zisserman, 2014), GoogleLeNet (Szegedy et al., 2015) or ResNet (He et al., 2016). All our experiments are implemented using the open-source deep learning framework Caffe (Jia et al., 2014).

4.2. Fine-tuning scenarios

The transferability of internal deep representations makes pre-trained CNNs useful for solving a variety of visual recognition tasks. Pre-trained CNNs can be used either as feature extractors or as weight initializers for fine-tuning towards new tasks. Generally, better performance is achieved if the pre-trained network is fine-tuned rather than only used as a feature extractor which fails to capture some discriminative information of the new dataset. The earlier layers of the network extract generic features such as edges or blobs, while features from later layers correspond more to the specific image details of classes included in the source dataset (Yosinski et al., 2014). When fine-tuning, a common practice is to copy all layers of the pre-trained model except the last layer, which is specific for the source classification task, and replace it with a new layer in which the number of neurons corresponds to the number of different classes in the new target domain. Because early layers extract features that are relevant to diverse image recognition tasks, fine-tuning only a part of the network, usually the last or last few layers, makes the network adapt to the specifics of the target domain and results in a boost of performance for many different classification problems.

Based on the target dataset size and the similarity between target and source domain, different fine-tuning scenarios are considered in order to find the most efficient solution, as well as avoid over-fitting. These scenarios include variations of the extent to which the error from the new task is being back propagated within the network or, in other words, how many of the transferred layers are kept frozen. In our work we test five different scenarios:

- all - upon each iteration, the weights of all layers are being modified
- skip first - the weights of the first convolutional layer (conv1) are kept frozen
- skip first 2 - the weights of the first two convolutional layers (conv1 and conv2) are kept frozen
- only last 3 - only the weights of the last three fully connected layers (fc6, fc7 and fc8) are being modified
- only last - only the weights of the last fully connected layer (fc8) are being modified

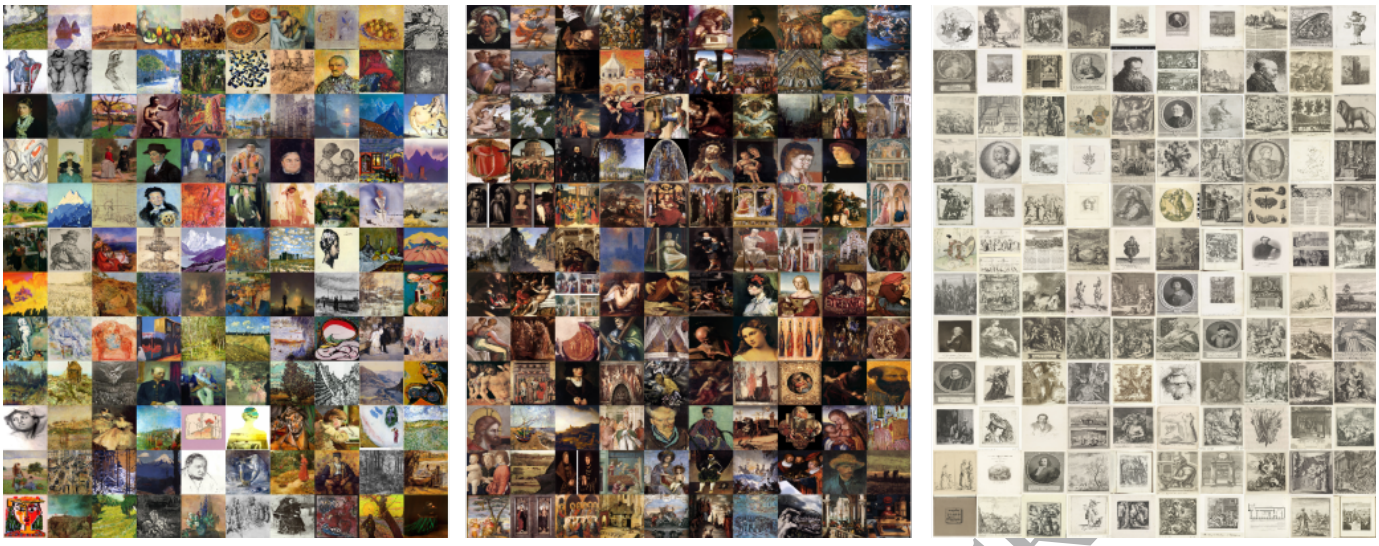

Figure 4: Examples of images from the three different data sources: WikiArt, WGA and TICC

4.3. Domain-specific weight initializations

It is known that the distance between the source and target domain influences the transferability of features. However, Yosinski et al. (2014) showed that initialization with transferred features improves the fine-tuned network's generalization performance even when domains are distant and enough training data is available to avoid over-fitting when training from scratch. To gain further insights in the impact of weight initialization, we explore the impact of various domain-specific initializations for different art-related classification tasks.

In this regard, we evaluate five different pre-trained networks in order to explore how changing the source domain influences the fine-tuning performance:

- CaffeNet is the BVLC reference model (Jia et al., 2014) trained on the subsets of ImageNet used in the ILSVRC-2012 competition (Deng et al., 2012), consisting of 1.2 million images with 1000 categories, where the goal was to identify the main objects present in images.

- Hybrid-CNN network (Zhou et al., 2014) is a CaffeNet model trained to classify categories of objects and scenes. The training set consists of 3.5 million images from 1183 categories, obtained by combining the Places database and ImageNet.

- MemNet network (Khosla et al., 2015) is pre-trained Hybrid-CNN model fine-tuned on the LaMem dataset, a large memorability dataset of 60 000 images annotated with human memory scores conducted through a memory game experiment using Amazon Mechanical Turk. Because the memorability score is a single real value in the range [0, 1], the Euclidean loss layer is used to fine-tune the Hybrid-CNN.

- Sentiment network (Campos et al., 2017) is a fine-tuned CaffeNet model for visual sentiment prediction on the DeepSent dataset (You et al., 2015), a set of 1269 Twitter images manually annotated as reflecting either positive or negative sentiment. The output of the fine-tuned network is the probability for the positive and negative sentiment evoked by the image.

- Flickr network (Karayev et al., 2014) is a CaffeNet model trained on the Flickr Style dataset, which consists of 80,000 photographic images labelled with 20 different visual styles comprising different stylistic concepts such as composition, mood or atmosphere of the image.

The concepts addressed within these five different models cover different domains, from the straightforward challenge of object recognition to more abstract ideas such as exploring the sentiment or memorability of images. By using learned weights of these five task-specific networks as different weight initializations for our fine-tuning experiments, we aim to explore if the initialization influences the performance in such a way that it reflects some inherent relatedness of those concepts with art-related concepts of genre, style and artist. It is worth mentioning that all those networks were developed before the introduction of batch normalization (Ioffe & Szegedy, 2015), which potentially reduces the dependence on model initialization by normalizing the input of each layer for each training mini-batch.

4.4. Training settings

During the training process, we employ simple data augmentation by horizontal mirroring and random cropping of input images. All the networks are fine-tuned using stochastic gradient descent with L2 regularization, momentum of 0.9, weight decay of 0.0005, with training

Table 2: Comparison of task-wise classification test accuracies achieved with different initializations

Dataset	Test accuracy					Variance of accuracies
	hybrid	memnet	sentiment	caffe	flickr	
TICC_artist	**0.762**	0.666	0.738	0.719	0.678	12.9×10^{-4}
wikiart_artist	**0.791**	0.725	0.787	0.763	0.714	9.92×10^{-4}
wikiart_style	**0.563**	0.526	0.558	0.542	0.507	4.27×10^{-4}
wikiart_genre	**0.776**	0.759	0.774	0.772	0.755	0.72×10^{-4}
wikiart_nationality	**0.583**	0.534	0.571	0.551	0.513	6.31×10^{-4}
wga_artist	**0.696**	0.551	0.686	0.655	0.569	36.2×10^{-4}
wga_timeframe	**0.527**	0.482	0.526	0.506	0.469	5.24×10^{-4}
wga_genre	0.796	0.779	**0.801**	0.787	0.765	1.59×10^{-4}
wga_nationality	**0.656**	0.612	0.655	0.635	0.603	4.70×10^{-4}

batch size of 256 and with unchanged dropout probability of 0.5. Changing those chosen parameters' values, using grid search or Bayesian optimization with SigOpt[4], does not give any significant qualitative and quantitative differences to our results. We perform numerous experiments for each classification task in order to determine the optimal number of training epochs, as well as the initial value of the learning rate and its reduction factor and frequency. Depending on the size of the dataset, a different number of epochs is needed for different tasks in order to achieve training convergence.

5. Results and discussion

The experimental results obtained can be analysed and discussed from several viewpoints. Firstly, we focus on the fine-tuning setup and analyse the impact of domain-specific weight initialization, as well as the influence of the extent to which the network is being re-trained. Furthermore, we address the overall classification results for each dataset and task, particularly in comparison to related works, as well as discuss the applicability of the fine-tuned models for image similarity analysis and visual link retrieval purposes.

5.1. Impact of domain-specific weight initialization

In order to evaluate how domain-specific weight initialization influences the fine-tuning performance, we fine-tune the differently initialized models under same conditions: re-training all the layers for 100 epochs with a fixed learning rate of 10^{-4}. The results for the weight initialization impact, which are compared in Table 2, show that for most of the tasks the highest test accuracy is achieved with the Hybrid-CNN initialization. The performance of differently initialized models on the validation set, which is consistent with the test set accuracy results, is shown in Figure 5 (left) for the WikiArt artist task when re-training all the layers and in Figure 5 (right) when re-training only the last layer. Similarly, validation accuracy curves for WGA artist classification task can be seen in Figure 6.

Because Hybrid-CNN was trained on a large dataset which combines the Places dataset and of ImageNet, it

[4]https://sigopt.com

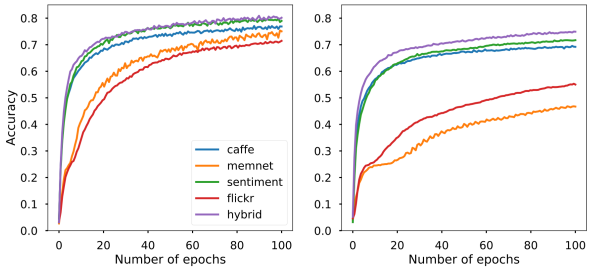

Figure 5: Validation accuracy curves of differently initialized models for the WikiArt artist classification task when fine-tuning all layers (left) and only the last layer (right)

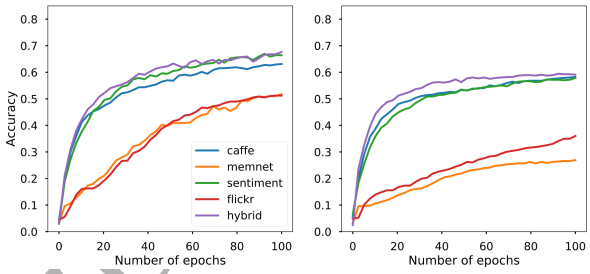

Figure 6: Validation accuracy curves of differently initialized models for the WGA artist classification task when fine-tuning all layers (left) and only the last layer (right)

outperforms CaffeNet trained only on the ImageNet objects dataset. We could interpret this boost in performance as a result of expanding the scope of recognizable image content from objects to scenes.

However, a very high performance achieved with the Sentiment network initialization represents an interesting finding. We might presume that the differentiation between emotionally positive and negative image content serves as a good starting point for differentiating art-related content. Besides acknowledging the universal entanglement of art and emotions, in order to gain a better understanding of the Sentiment network behaviour, a deeper layer-wise output analysis is needed.

The lower accuracy rates of the MemNet network indicates that the image memorability counteracts the learning convergence towards art-related tasks. Similarly, the Flickr network underperforms in comparison to other networks for all the datasets and task. Although the Flicker model addresses the concept of image style and should therefore be considered as a good basis for fine-tuning towards artistic style recognition, its lower performance might be a result of the discrepancy between the Flickr style concept and the art history style concept, as well as of the lack of distinctiveness between the initial 20 Flickr style classes, which most likely explains the network's rather low performance (39% accuracy) on the original style recognition task by the conducted fine-tuning setting.

Furthermore, based on the variance of accuracies ob-

tained for the same tasks using differently initialized networks (Table 2), we can see that the variance is high for the artist task (in all three datasets) and low for the genre task (in both Wikiart and WGA datasets). This leads us to conclude that weight initialization has a higher impact on the overall performance in tasks with many classes and fewer examples per class (the artist classification task), than with tasks with fewer classes and more images per class (the genre task).

5.2. Influence of different fine-tuning scenarios

In order to conclude what is the optimal relation of frozen/trainable layers when fine-tuning towards art-related classification tasks, we tested five different scenarios. We keep the other training setup properties fixed by using the same best performing weight initialization (Hybrid-CNN) and re-training the model for 100 epochs with a constant learning rate. The performance results for each scenario and each task is given in Table 3.

Table 3: Comparison of task-wise classification test accuracies for different fine-tuning scenarios

scenario	Test accuracy								
	TICC	WikiArt				WGA			
	artist	artist	style	genre	nationality	artist	timeframe	genre	nationality
all	0.762	0.791	0.563	0.776	0.584	0.696	0.526	**0.796**	**0.656**
skip_first	**0.767**	**0.798**	0.564	0.774	**0.585**	**0.704**	**0.537**	0.792	0.651
skip_first2	0.765	0.795	**0.570**	**0.777**	0.583	0.689	0.524	0.790	0.652
only_last3	0.668	0.762	0.537	0.762	0.554	0.665	0.505	0.791	0.634
only_last	0.583	0.740	0.516	0.754	0.532	0.646	0.488	0.772	0.619

Based on the accuracy results, we can conclude that the best scenario in most cases is to re-train all except the first convolutional layer. The correlation between the original tasks of object and scenes recognition and various art-related tasks is sufficient enough to confirm that the first, and in many cases the second, convolutional layer extracts mutually relevant features. From the results and the validation accuracy curves of the WikiArt style classification task presented in Figure 7, we can observe that very similar performance is achieved by fine-tuning all layers or skipping the first one or two layers.

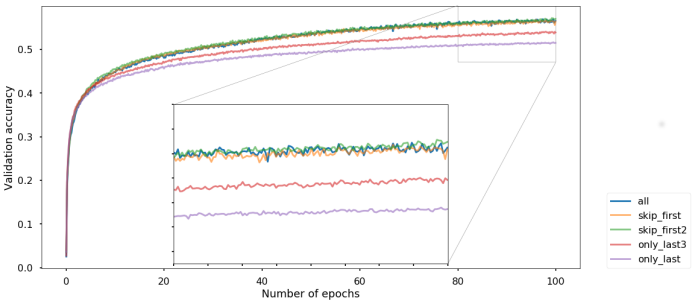

Figure 7: Comparison between validation accuracy learning curves of different scenarios for the WikiArt style task

By comparing the accuracy results (Figure 8) and the training time for 100 epochs (Figure 9) when fine-tuning

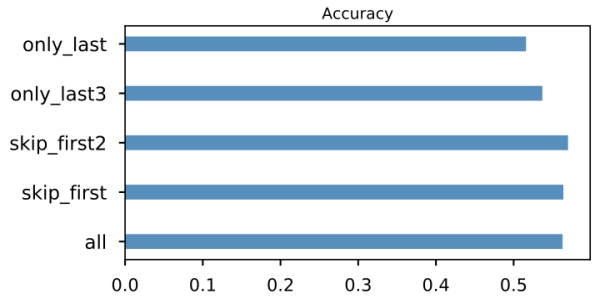

Figure 8: WikiArt style accuracy for different fine-tuning scenarios

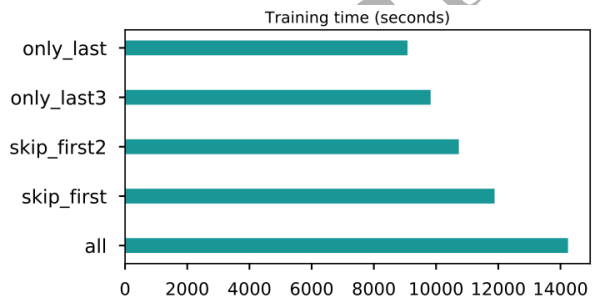

Figure 9: Training time of different fine-tuning scenarios for the WikiArt style task

for the WikiArt style task, we can conclude that by freezing the first two convolutional layers we gain the best performance in significantly less time compared to re-training all layers.

5.3. Best overall classification performance

In order to compare the weight initialization and fine-tuning scenario impacts, we fine-tuned all the models for 100 epochs with a fixed learning rate. However, to identify the best classification performance for each task, we performed a large number of experiments with different training settings. The best results for each task are summarized in Table 4, together with results of previous studies.

In most cases the best accuracy is achieved by training for a large number of epochs with a constant learning rate of 10^{-4}. On the other hand, if we start training with a higher learning rate and decrease it over time, we can achieve relatively high classification performance within a smaller number of epochs. For instance, if we train the model for the WikiArt style classification task for only 20 epochs, starting with a learning rate 10^{-3} and reduce it by factor of 10 after 5 epochs, we achieve 53.02% accuracy which is deterioration of only $\sim 3\%$ in comparison to the best result achieved by training for 100 epoch with a fixed smaller learning rate.

The results show that our simple and conventional fine-tuning approach outperforms the current state-of-the-art reported for the WikiArt dataset in (Tan et al., 2016). In this work the authors achieved the best results with fine-tuning of an AlexNet network pre-trained on the Im-

Table 4: Comparison of results for all tasks and datasets

Reference	Method	Dataset	Style/Time-frame		Genre		Artist		Nationality	
			# of classes	acc. (%)	# of classes	acc. (%)	# of classes	acc. (%)	# of classes	acc. (%)
Our results	CNN fine-tuning (CaffeNet)	TICC	-	-	-	-	210	80.42 (80.26 F-score)	-	-
		WGA	12	53.75	6	80.1	23	70.42	8	65.20
		WikiArt	27	56.43	10	77.6	23	81.94	8	58.35
B. Saleh et al.	Feature fusion	WikiArt	27	45.97	10	60.28	23	68.25	-	-
Tan et al.	CNN fine-tuning (AlexNet)	WikiArt	27	54.50	10	74.14	23	76.11	-	-
Hentschel et al.	CNN fine-tuning (CaffeNet)	WikiArt	22	55.9 (MAP)	-	-	-	-	-	-
Lecoutre et al.	CNN fine-tuning (ResNet50)	WikiArt	25	62.8	-	-	-	-	-	-
Noord et. al	multi-scale CNN (All-CNN)	TICC	-	-	-	-	210	77.01 (F-score)		

ageNet dataset. However, with our implementation of different domain-specific weight initializations and different training settings, we show that the model performance can be further improved. On the other hand, using a deeper model such as ResNet50 can lead to a boost of performance as shown for the WikiArt style classification task by Lecoutre et al. (2017). Although they use a smaller number of classes (25 instead of 27), the results achieved by fine-tuning an ImageNet pre-trained ResNet50 model, together with applying data augmentation methods such as bagging and distortion, represent the currently highest result for the task of style classification. Regarding the TICC dataset, our approach surpasses the results achieved with ensembling multi-scale CNNs by van Noord & Postma (2017). Regarding the WGA dataset, to the best of our knowledge, there are currently no other works available for comparing classification results.

5.4. Interpretation of classification results

After determining the best performing training setup for each task, a further exploration of the task-specific classification can be carried out by looking into the per-class classification performance. Figure 10 shows the confusion matrix for the WikiArt style classification task. From it we can observe that the most distinctively categorized style is Ukiyo-e (84%), which refers to a style of Japanese woodblock print and paintings from the 17th through 19th centuries. This observation is in line with the results presented by Tan et al. (2016). The poorest results are achieved for the style of academism, which is being misclassified most commonly as realism. This however is due to the fact that both rely on using precise illusionistic brushwork, but academism aims at emphasising intellectual messages and high-minded themes, whereas realisms relates to an artistic movement that emerged with the aim to portray everyday subjects and ordinary situations, as shown in examples in Figure 11.

This misclassification example demonstrates the fact that style is not only associated with mere visual characteristics and content of an artwork, but is often a subtly differentiable and contextually depended concept. The common visual properties of different styles explain the high misclassification rate between classes such as abstract expressionism and lyrical abstraction, as well as impressionism and post-impressionism or rococo and baroque.

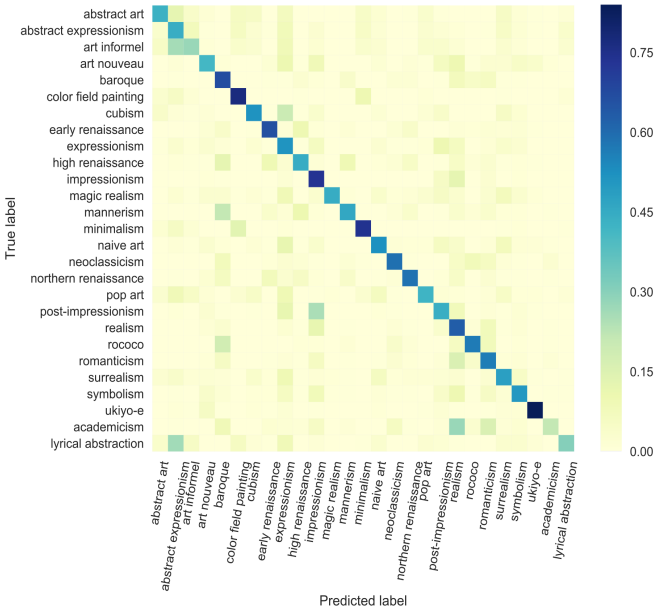

Figure 10: Confusion matrix for WikiArt style classification task

Figure 11: Examples of paintings belonging to the styles of academism and realism

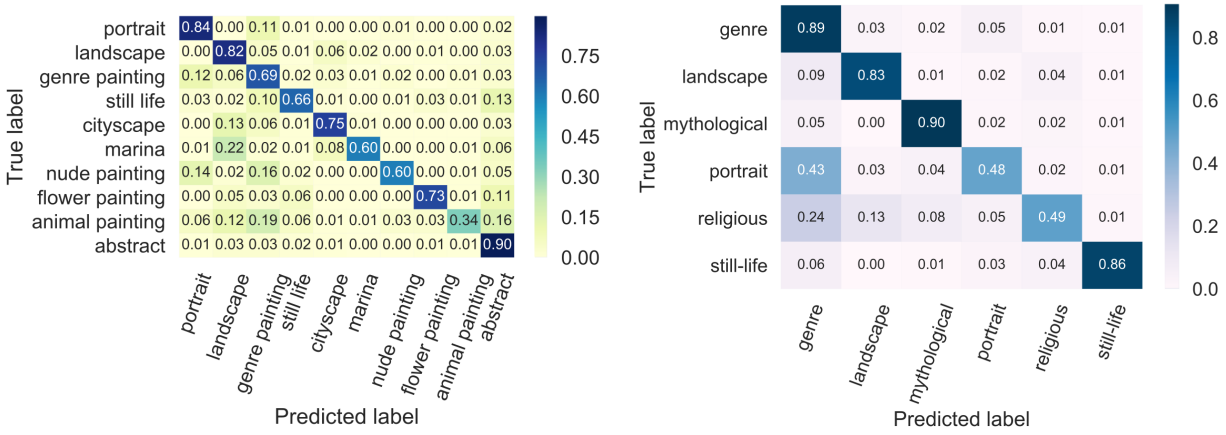

Figure 12: Confusion matrix for WikiArt (left) and WGA (right) genre classification

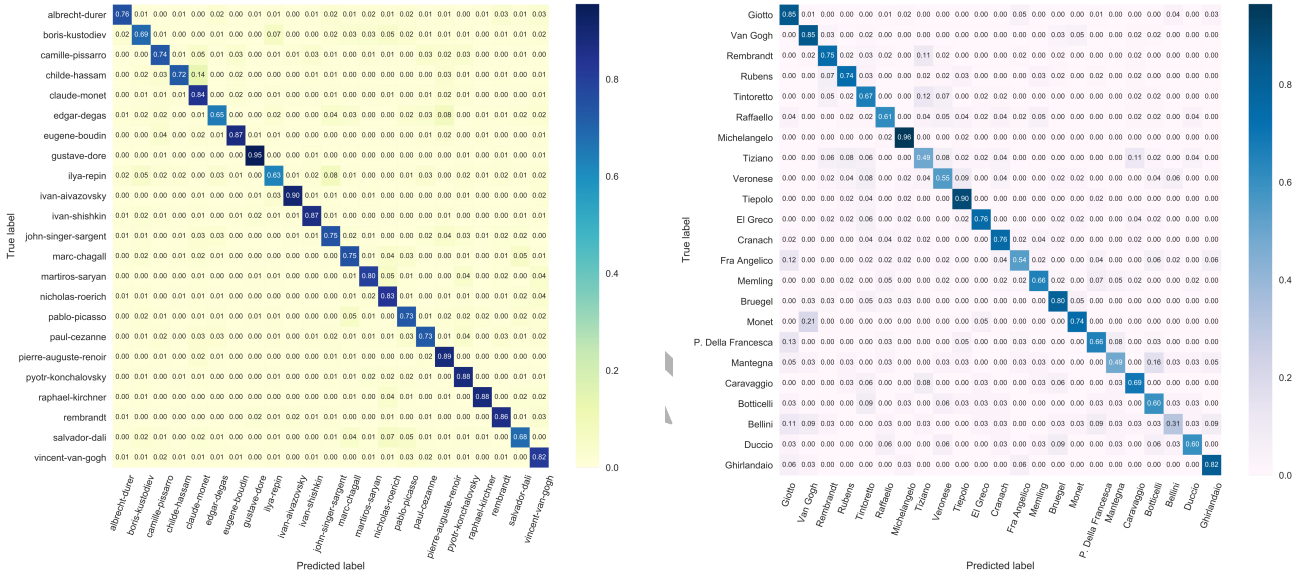

Figure 13: Confusion matrix for WikiArt (left) and WGA (right) artist classification

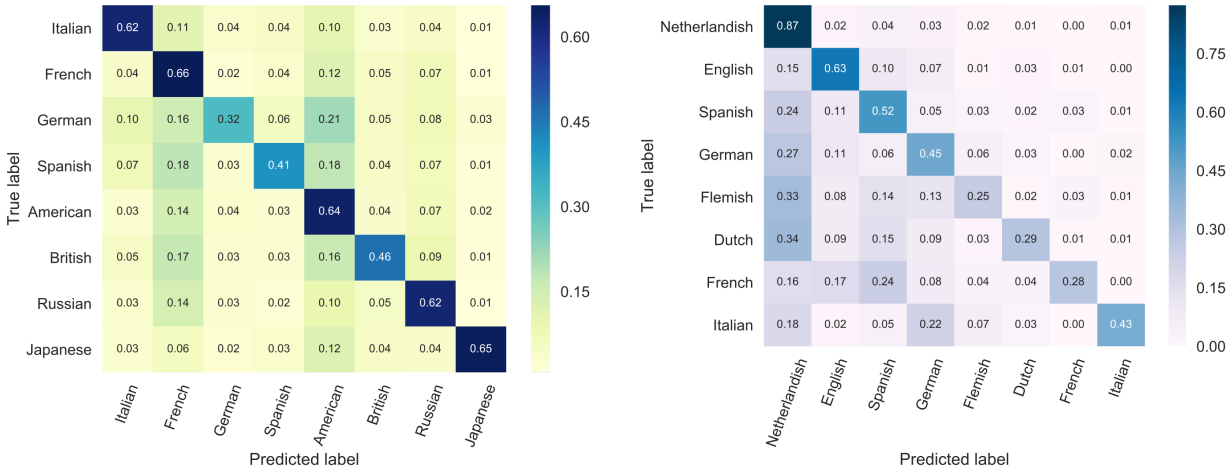

Figure 14: Confusion matrix for WikiArt (left) and WGA (right) nationality classification

In comparison to other tasks, the lower style classification corresponds to the high level of visual properties overlapping between classes, as well as to the great diversity of content depicted in the same style. On the other hand, the classes of the genre classification task are more uniform in terms of content and CNNs show a high ability to distinguish scenes and objects in paintings, regardless of the various artistic techniques and styles. From the confusion matrices in Figure 12 for WikiArt genre (left) and WGA genre (right), we can observe the inner logic of misclassified classes. For example, the high rate of cityscape and marina paintings being misclassified as landscape because they include outdoor scenes; or genre and nude paintings being confused for portraits because they depict faces.

For the task of artist classification, the overall accuracy is quite high for all three datasets, particularly in respect to the high number of classes and lower number of images per class. Based on the confusion matrices (Figure 13), the interpretation of the misclassified paintings indicates a general similarity between the works of different artist, for instance impressionist painters Childe Hassam and Cloude Monet in the WikiArt dataset; or 16th century Italian painters Tintoretto, Veronese and Tiziano in the WGA dataset.

The results obtained for the task of recognizing artworks belonging to the same national artistic context present an interesting finding. Having in mind that the only common baseline in this task is that the artist of an artwork is associated with a particular national artistic circle, the overall accuracy for WikiArt being 58.4% and for WGA 65.2% can be considered a surprisingly high result. Figure 14 (left) shows the confusion matrix for the WikiArt and Figure 14 (right) for the WGA nationality classification task.

The result of this experiment can be considered as preliminary test for addressing the task of classifying artworks by nationality in a more thorough manner. The existing correlations between classes could potentially explain artistic influences and patterns within different national artistic heritages. However, an in-depth analysis of the dataset in collaboration with art history experts is needed before drawing any meaningful conclusion.

5.5. *Fine-tuned CNNs as feature extractors for image similarity*

In addition to exploring the best training setting and task-specific classification performance, we aim to address the usability of the fine-tuned models. One apparent applicability is to enhance the search capabilities within online art collections by enabling image content search and visual similarity retrieval. Therefore we want to explore if CNN models fine-tuned for genre and style recognition can be used for retrieving images of similar genre or style.

For this purpose we use the fine-tuned models as feature extractors and calculate the similarity between feature vectors. Concretely, we use the outputs of the penultimate layer (fc7 layer) for representing the image with a 4096 feature vector. As a distance metric for calculating the image similarity based on the extracted feature vectors we use the cosine distance measure. Figure 15 presents examples of images retrieved as most similar to the input image when using the best performing CNN models fine-tuned for the WikiArt genre task or fine-tuned for the WikiArt style task as feature extractors.

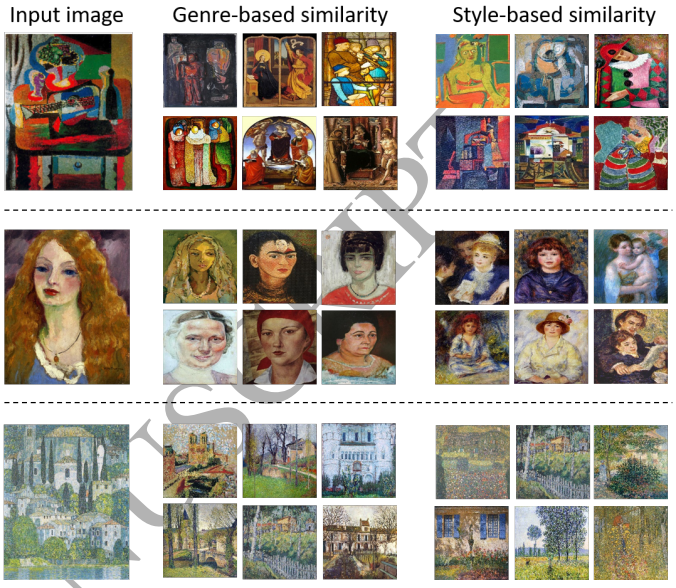

Figure 15: Examples of paintings retrieved as most similar to the input image when using the genre-tuned CNN model and the style-tuned CNN model as feature extractors

From those examples we can see that the CNN fine-tuned for the genre recognition task retrieves images that are more similar in terms of content, by including specific objects and similar compositions.

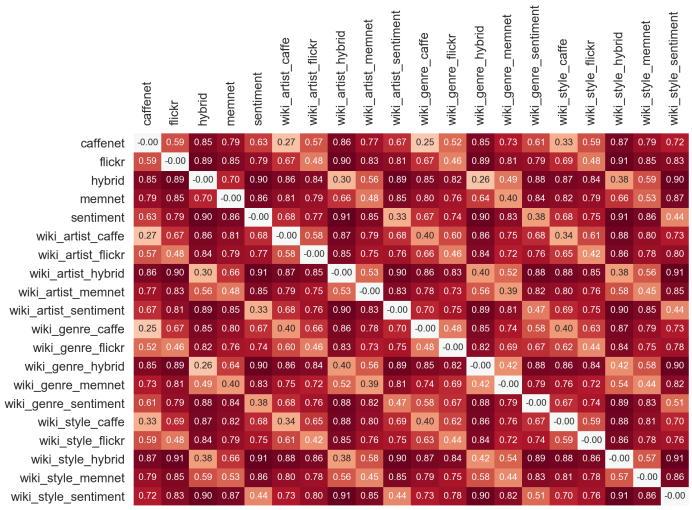

Figure 16: Cosine distance matrix of image features extracted from different fine-tuned models

On the other hand, the CNN fine-tuned for style recognition focuses more on style properties such as brushwork or level of details. We presume that a further improvement

of task-specific classification performance would lead to a higher level of distinctiveness between genre-similar and style-similar images.

Additionally, we use this approach of calculating image similarity to explore the distance of features extracted by differently initialized models. For this purpose we created a sub-collection of 100 randomly chosen art images for which we extracted the fc7 features with differently initialized models and calculated the mean of the overall cosine distance between images. The distance matrix of various models is shown in Figure 16. Knowing that for most input images, the first 1000 most similar images have a distance smaller than 0.4, we can conclude that the domain-specific initialization, as well as the task-specific fine-tuning, can highly influence the performance of retrieving similar images.

6. Conclusion

This paper presents the results of extensive CNN fine-tuning experiments performed on three large art collections for five different art-related classification tasks. We compared different fine-tuning strategies in order to identify the best training setup for different art-related tasks and datasets, with a particular focus on exploring the impact of domain-specific weight initialization. We showed that the pre-trained model initialization influences the fine-tuning performance, particularly when the target dataset consists of many classes with fewer images per class. Moreover, we showed that fine-tuning networks pre-trained for scene recognition and sentiment prediction yields better results than fine-tuning networks pre-trained only for object recognition. This indicates that the semantic correlation between different domains could be inherent in the CNN weights. However, in order to draw definite conclusions about the semantic implications of weight initialization, further exploration is necessary. In particular, ground-truth labelling of different image properties on the same dataset is a prerequisite for investigating the perceptual correlation of domains such as sentiment and memorability. Having conclusions established on the psychological level, would enable a stronger evaluation of the CNN behaviour. However, collecting ground-truth labels for attributes related to subjective perception of images requires complex experimental surveys. Nevertheless, pre-trained CNN models can be used in order to shape inceptive hypotheses about the relation of different domain-specific image features.

This constitutes the central direction of our future research. In particular, in our future work we aim to investigate the applicability of CNN beyond classification and towards understanding perceptually relevant image features and their relation to different artistic concepts. Furthermore, we aim to strengthen our interdisciplinary collaboration and investigate the relevance of our findings to concrete art history-related research topics. Specifically, we aim to explore how deep neural networks can be used for extracting high-level and semantically relevant features that can serve as a basis for discovering new knowledge patterns and meaningful relations among specific artworks or artistic oeuvres. Besides using CNN to gain a new perspective of fine art, we also aim to advance our understanding and interpretability of deep learning models by utilizing CNN representations visualization techniques (such as activation maximization, saliency maps and class activation maps) and other interpretability concepts such as semantic dictionaries. Furthermore, the fine-tuned models presented in this work outperform the current state-of-the-art classification results for most of the tasks and datasets used. However, we plan to investigate if further improvement can be achieved by using deep models of different architectures. Finally, in this work we address the practical applicability of task-specific fine-tuned models for visual image similarity analysis. Our findings suggest that the proposed approach can serve as a basis for implementing a novel framework for refined retrieval of fine art images, as well as enhancing capabilities of search systems in existing online art collections.

Acknowledgement

This research has been partly supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

References

References

Arora, R. S., & Elgammal, A. (2012). Towards automated classification of fine-art painting style: A comparative study. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 3541–3544). IEEE.

Bar, Y., Levy, N., & Wolf, L. (2014). Classification of artistic styles using binarized features derived from a deep neural network. In *ECCV Workshops (1)* (pp. 71–84). doi:10.1007/978-3-319-16178-5_5.

Barnet, S. (2011). *A short guide to writing about art*. Pearson Education.

Brachmann, A., & Redies, C. (2017). Computational and experimental approaches to visual aesthetics. *Frontiers in computational neuroscience*, *11*, 102. doi:10.3389/fncom.2017.00102.

Bressan, M., Cifarelli, C., & Perronnin, F. (2008). An analysis of the relationship between painters based on their work. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on* (pp. 113–116). IEEE. doi:10.1109/ICIP.2008.4711704.

Campos, V., Jou, B., & Giro-i Nieto, X. (2017). From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction. *Image and Vision Computing*, *65*, 15–22. doi:10.1016/j.imavis.2017.01.011.

Cetinic, E., & Grgic, S. (2013). Automated painter recognition based on image feature extraction. In *ELMAR, 2013 55th International Symposium* (pp. 19–22). IEEE.

Cetinic, E., & Grgic, S. (2016). Genre classification of paintings. In *ELMAR, 2016 International Symposium* (pp. 201–204). doi:10.1109/ELMAR.2016.7731786.

Chu, W.-T., & Wu, Y.-L. (2016). Deep correlation features for image style classification. In *Proceedings of the 2016 ACM on Multimedia Conference* (pp. 402–406). ACM. doi:10.1145/2964284.2967251.

Crowley, E. J., & Zisserman, A. (2014). In search of art. In *ECCV Workshops (1)* (pp. 54–70). Springer. doi:10.1007/978-3-319-16178-5_4.

David, O. E., & Netanyahu, N. S. (2016). Deeppainter: Painter classification using deep convolutional autoencoders. In *International Conference on Artificial Neural Networks* (pp. 20–28). Springer. doi:10.1007/978-3-319-44781-0_3.

Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., & Fei-Fei, L. (2012). Ilsvrc-2012, 2012. *URL http://www.image-net.org/challenges/LSVRC*, .

[dataset] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 248–255). IEEE. doi:10.1109/CVPRW.2009.5206848.

Falomir, Z., Museros, L., Sanz, I., & Gonzalez-Abril, L. (2018). Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (qart-learn). *Expert Systems with Applications*, *97*, 83–94. doi:10.1016/j.eswa.2017.11.056.

Florea, C., Condorovici, R., Vertan, C., Butnaru, R., Florea, L., & Vrânceanu, R. (2016). Pandora: Description of a painting database for art movement recognition with baselines and perspectives. In *Signal Processing Conference (EUSIPCO), 2016 24th European* (pp. 918–922). IEEE. doi:10.1109/EUSIPCO.2016.7760382.

Gando, G., Yamada, T., Sato, H., Oyama, S., & Kurihara, M. (2016). Fine-tuning deep convolutional neural networks for distinguishing illustrations from photographs. *Expert Systems with Applications*, *66*, 295–301. doi:10.1016/j.eswa.2016.08.057.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587). doi:10.1109/CVPR.2014.81.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). doi:10.1109/CVPR.2016.90.

Hentschel, C., Wiradarma, T. P., & Sack, H. (2016). Fine tuning cnns with scarce training data adapting imagenet to art epoch classification. In *Image Processing (ICIP), 2016 IEEE International Conference on* (pp. 3693–3697). IEEE. doi:10.1109/ICIP.2016.7533049.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448–456).

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675–678). ACM. doi:10.1145/2647868.2654889.

Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., & Winnemoeller, H. (2014). Recognizing image style. In *Proceedings of the British Machine Vision Conference*. BMVA Press. doi:10.5244/C.28.122.

Keren, D. (2002). Painter identification using local features and naive bayes. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on* (pp. 474–477). IEEE volume 2. doi:10.1109/ICPR.2002.1048341.

Khan, F. S., Beigpour, S., Van de Weijer, J., & Felsberg, M. (2014). Painting-91: a large scale database for computational painting categorization. *Machine vision and applications*, *25*, 1385–1397. doi:10.1007/s00138-014-0621-6.

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2390–2398). doi:10.1109/ICCV.2015.275.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lecoutre, A., Negrevergne, B., & Yger, F. (2017). Recognizing art style automatically in painting with deep learning. In *Asian Conference on Machine Learning* (pp. 327–342).

Lombardi, T. E. (2005). *Classification of Style in Fine-art Painting*. Pace University.

Mensink, T., & Van Gemert, J. (2014). The rijksmuseum challenge: Museum-centered visual recognition. In *Proceedings of International Conference on Multimedia Retrieval* (p. 451). ACM. doi:10.1145/2578726.2578791.

van Noord, N., Hendriks, E., & Postma, E. (2015). Toward discovery of the artist's style: Learning to recognize artists by their artworks. *IEEE Signal Processing Magazine*, *32*, 46–54. doi:10.1109/MSP.2015.2406955.

van Noord, N., & Postma, E. (2017). Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognition*, *61*, 583–592. doi:10.1016/j.patcog.2016.06.005.

Reyes, A. K., Caicedo, J. C., & Camargo, J. E. (2015). Fine-tuning deep convolutional networks for plant recognition. In *CLEF (Working Notes)*.

Saleh, B., Abe, K., Arora, R. S., & Elgammal, A. (2016). Toward automated discovery of artistic influence. *Multimedia Tools and Applications*, *75*, 3565–3591. doi:10.1007/s11042-014-2193-x.

Saleh, B., & Elgammal, A. (2016). Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *International Journal for Digital Art History*, *0*. doi:10.11588/dah.2016.2.23376.

Seguin, B., Striolo, C., Kaplan, F. et al. (2016). Visual link retrieval in a database of paintings. In *ECCV Workshops (1)* (pp. 753–767). Springer. doi:10.1007/978-3-319-46604-0_52.

Shamir, L., Macura, T., Orlov, N., Eckley, D. M., & Goldberg, I. G. (2010). Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception (TAP)*, *7*, 8. doi:10.1145/1670671.1670672.

Shamir, L., & Tarakhovsky, J. A. (2012). Computer analysis of art. *Journal on Computing and Cultural Heritage (JOCCH)*, *5*, 7. doi:10.1145/2307723.2307726.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, *abs/1409.1556*. URL: http://arxiv.org/abs/1409.1556. arXiv:1409.1556.

Strezoski, G., & Worring, M. (2017). Omniart: Multi-task deep learning for artistic data analysis. *CoRR*, *abs/1708.00684*. URL: http://arxiv.org/abs/1708.00684. arXiv:1708.00684.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9). doi:10.1109/CVPR.2015.7298594.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, *35*, 1299–1312. doi:10.1109/TMI.2016.2535302.

Tan, W. R., Chan, C. S., Aguirre, H. E., & Tanaka, K. (2016). Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *Image Processing (ICIP), 2016 IEEE International Conference on* (pp. 3703–3707). IEEE. doi:10.1109/ICIP.2016.7533051.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320–3328).

You, Q., Luo, J., Jin, H., & Yang, J. (2015). Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *AAAI* (pp. 381–388).

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Advances in neural information processing systems* (pp. 487–495).

Zujovic, J., Gandy, L., Friedman, S., Pardo, B., & Pappas, T. N. (2009). Classifying paintings by artistic genre: An analysis of features & classifiers. In *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on* (pp. 1–5). IEEE.

doi:10.1109/MMSP.2009.5293271.