

# Transfer Learning with Style Transfer between the Photorealistic and Artistic Domain

**Nikolay Banar; University of Antwerp; Antwerp, Belgium**

**Matthia Sabatelli; University of Liege; Liege, Belgium**

**Pierre Geurts; University of Liege; Liege, Belgium**

**Walter Daelemans; University of Antwerp; Antwerp, Belgium**

**Mike Kestemont; University of Antwerp; Antwerp, Belgium**

## Abstract

*Transfer Learning is an important strategy in Computer Vision to tackle problems in the face of limited training data. However, this strategy still heavily depends on the amount of available data, which is a challenge for small heritage institutions. This paper investigates various ways of enriching smaller digital heritage collections to boost the performance of deep learning models, using the identification of musical instruments as a case study. We apply traditional data augmentation techniques as well as the use of an external, photorealistic collection, distorted by Style Transfer. Style Transfer techniques are capable of artistically stylizing images, reusing the style from any other given image. Hence, collections can be easily augmented with artificially generated images. We introduce the distinction between inner and outer style transfer and show that artificially augmented images in both scenarios consistently improve classification results, on top of traditional data augmentation techniques. However, and counter-intuitively, such artificially generated artistic depictions of works are surprisingly hard to classify. In addition, we discuss an example of negative transfer within the non-photorealistic domain.*

## Introduction: Computer Vision and Art History

In recent years, the scientific community has increasingly acknowledged the considerable potential of computer vision for art history [1]. Several case studies have demonstrated the general feasibility of applying machine learning methods to artistic collections, with impressive results that are as relevant to art history as to wider computer sciences. Most of the recent studies in this field capitalize on the considerable advances which computer vision has witnessed in the last decade, following the popularization of Deep Convolutional Neural Networks (DCNNs) [2, 3]. It is a well-known fact that the “renaissance” of this data-hungry family of models has only been possible because of the open availability of large, annotated benchmark datasets such as ImageNet [4] or MS-COCO [5]. These invaluable resources contain millions of photorealistic images, which enabled scholars to train effective image models, of a complexity and depth that was previously inconceivable.

A commonly heard worry, however, is that the available models remain difficult to transfer to the artistic domain, which has several reasons. First of all, the available (annotated) datasets from art history are typically much smaller; luckily, a number of domain-specific, larger-scale resources have been released that

help mitigate this situation [6, 7, 8, 9, 10, 11, 12, 13, 14]. Secondly, there is often a mismatch between the (contemporary) ontologies used for annotating present-day images and the (historical) concepts that are more relevant to art historians [15]. In the face of such situations of data and label shortage, transfer learning [16] and data augmentation [4] have become a standard practice in computer vision, and these methods have also been successfully applied to the artistic domain in previous work (see Related Work). A third and more salient issue is that artworks typically actively distort the subjects they depict, resulting in a staggering variety, for instance, in the textures attested in image collections. As such, artworks clearly belong to another “domain” than the photorealistic one that is represented in the more established resources in the field. The shift in domain between photorealistic and artistic collections can be described in terms of “style”, a vexed notion that is notoriously hard to define, but which generally relates to individual or group-level variation between artworks, that is less related to the content of a work than to the manner in which this content is depicted.

The notion of style has received considerable attention in recent computer vision studies, especially in the wake of the seminal “style transfer” paper. In this work, Gatys et al. [17] demonstrated the general feasibility of transferring styles across images that depict different contents, leading to convincing, highly publicized results. Follow-up work [18, 19, 20, 21] has fine-tuned the original technique, developed faster algorithms and investigated whether style transfer [22, 23] would be a viable alternative to more conventional approaches of image augmentation. This previous work, however, has remained somewhat indecisive as to the precise benefits of style transfer in this context, which in some cases only seemed incremental, in particular in comparison to more conventional data augmentation techniques.

In this paper, we revisit the effectiveness of style transfer as a data augmentation technique in the context of sparse digital heritage data. We present a small-scale, yet focused case study from music iconography, in which we closely compare several transfer strategies in a classification set-up. The structure of this paper is as follows. We first present the related work and discuss the rationale of our contribution. Then, we describe the datasets and present our theoretical framework in greater detail, before presenting results of the core case study. Finally, we discuss our results and summarize the main contributions of this work with ideas for future work.

## Related Work

### Artistic Image Classification

Image classification in itself is a task that has a rich history of applications in art history. Before the deep learning era, scholars used the following approaches in their attempts to classify artworks. Low-level features representing shape, color or texture were extracted and used as an input of a simple classifier such as naive Bayes, support vector machines, k-nearest neighbours or multilayer perceptron [24]. Machine learning methods were also applied to attribution problems [25, 26, 27], genre classification [28] and other kinds of style-based classification [29, 30, 31]. DCNNs were seminally introduced as feature extractors [10]. They outperformed most of the hand-crafted features or were used in combination with them [32, 33, 34, 35]. Only with the release of larger artistic datasets, the training of DCNNs from scratch has become feasible [36].

### Transfer Learning for Artistic Image Classification

Previous work has confirmed the usefulness of transfer learning in the context of art classification tasks too, for instance for material classification or artist attribution. Tan et al. [37] demonstrated that the fine-tuning of an AlexNet [4], pretrained on ImageNet, yielded state-of-the-art results, even outperforming training from scratch. This result was confirmed by a number of follow-up studies [38, 39, 40, 41, 42, 43]. Sabatelli et al. [42] applied DCCNs to several classification tasks in digital heritage and further demonstrated the benefits of transferring pretrained networks to and across artistic collections. Cetinic et al. [38] found that DCNNs, pretrained for scene recognition and sentiment prediction, outperform DCNNs pretrained for object recognition. Therefore, fine-tuning the entire network proved to be clearly superior to other approaches, demonstrating how the mere initialization of the weights in a DCNNs is crucial for downstream applications in the artistic domain. Nowadays, fine-tuning is widely used in the artwork classification problem [44, 45, 46, 47].

### Rationale of the Present Contribution

Notwithstanding the presented benefits, transfer learning still heavily depends on the size of the available datasets, which, as described above, is a challenge when working with relatively smaller heritage collections. Therefore, the enrichment of these collections is an important challenge which might boost the performance of DCNNs in this area. In computer vision, data augmentation [4] is a standard technique to improve the performance of DCNNs, especially in the case of limited datasets. However, while this approach increases the amount of variance in the dataset, it does not change the original number of objects available. Hence, the usage of additional, *external* data sources is appealing, because it increases the actual number of objects in the dataset. Unfortunately, modern computer vision mostly has large photorealistic collections on offer, such as ImageNet [4], that are far from the artistic domain. Apart from traditional techniques, based on affine transformations, style transfer has been successfully applied as a data augmentation technique [22, 23]. In the present case, style transfer is appealing due to the ability to generate artistic images from any image given one style. Hence, artificial artistic images, especially generated from external photorealistic collections, can enrich small heritage collections in order to improve the classification performance of DCNNs.

In this work we investigate whether style transfer can be successfully applied to the classification of musical instruments. For this purpose, we build two datasets, one with artistic depictions, and one with photorealistic photographs. In the first, we “internally” transfer styles (inner style transfer), i.e. within the artistic dataset. The second dataset is treated as an external photorealistic asset, that is used to enrich the internal dataset via style transfer (outer style transfer). We compare these approaches to traditional data augmentation methods. Additionally, we compare two DCNNs pretrained on the general domain and the heritage domain. Below, we show that both inner and outer style transfer, as well as external photorealistic data are extremely useful for art classification. We shall also conclude that style transfer applied to the test set might worsen the performance of DCNNs. Additionally, we demonstrate that intermediate fine-tuning can hurt the final performance.

### Datasets: MIMO and Minerva

In this paper, we shall build on two data collections from the field of music iconography, a field at the intersection of art history and musicology, which is concerned with the scholarly study of the depiction of musical instruments, performances and artists across the visual arts.

1. **MIMO** (Musical Instrument Museums Online) [48] is an international database of photographs depicting non-fictional musical instruments, aggregated from multiple heritage collections. MIMO uses a standardized (hierarchical) ontology, with an unambiguous code to identify each instrument class.
2. **MINERVA** (Musical Instruments Represented in the Visual Arts) [15] is a benchmark dataset for the detection of musical instruments in artworks, derived from the RIDIM database<sup>1</sup>, a number of museum collections and Flickr<sup>2</sup>. Minerva adopts the same ontology as MIMO, allowing us to identify instruments across both collections. This work only considers classification and is restricted to the instrument patches that can be extracted using the bounding boxes.

Both collections depict similar contents (musical instruments) but come from different domains, displaying clear “stylistic” differences (see the examples in Figures 1 and 2). This makes the confrontation of both collections an ideal case study in transfer learning across different domains.

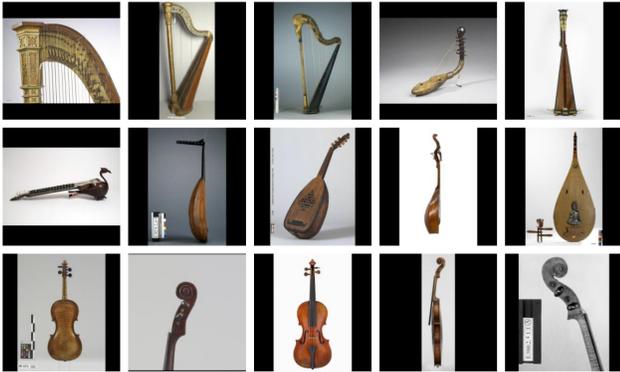
MIMO is entirely photorealistic and depicts real, non-fictional instruments: these are typically historical museum objects, on display in isolation, mostly against a neutral background. Minerva, however, is limited to non-photorealistic depictions of musical instruments in works of art. Apart from the artistic stylization that has gone into the depiction, the instruments are often portrayed as being handled by artists during a live performance. Also, the background is typically much less neutral and the instrument in a single patch might even partially overlap with an artist’s body or other objects in the scene, including other instruments. For these reasons, the instrument classification task in Minerva intuitively seems harder than in MIMO.

Unfortunately, both data collections also show very distinct distributions: especially the presence of individual instruments in

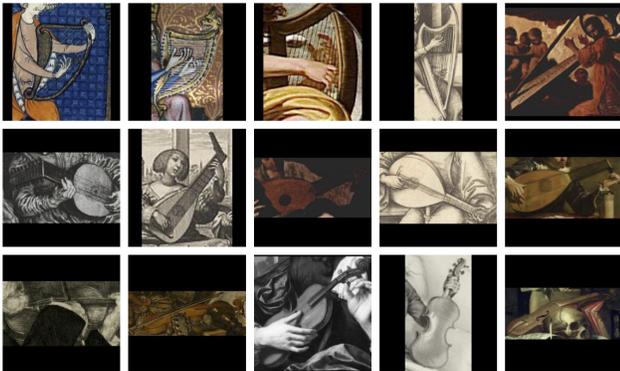
<sup>1</sup><https://ridim.org/ridim-database/>

<sup>2</sup><https://www.flickr.com>

Minerva is highly skewed, suggesting how only a relatively small number of instruments have been favored as artistic subjects in Western cultural history. To mitigate this skewness and allow for a balanced classification setup, we have restricted our experiments to the three the most common instrument labels from Minerva: ‘Lute’, ‘Harp’, and ‘Violin’. This selection is rather severe but enabled us to construct generous train, validation and test splits, that contained enough instances of each instrument to produce reliable learning curves. For each instrument and for each dataset, we included 400 images in the training set, 200 for the development set and 200 for the test set. Every image was padded and resized to 224 by 224 pixels.



**Figure 1.** Random selection of images from MIMO. Consecutive rows display examples for the categories ‘harp’, ‘lute’ and ‘violin’.



**Figure 2.** Random selection of images from MINERVA. Consecutive rows display examples for the categories ‘harp’, ‘lute’ and ‘violin’.

## Methodology

In this section, we present the methods underlying our research. We briefly discuss the type of the transfer learning technique used, as well as the data augmentation techniques applied and the experimental settings, including the training regime.

### Transfer learning

The theoretical background of this paper must be situated in transfer learning, a specific form of machine learning that offers a simple, yet powerful solution to applications where only small amounts of (annotated) in-domain data are available, but re-

searchers have access to (other or larger) out-of-domain datasets. The main intuition supporting this line of research is that the knowledge gained from a large, potentially more general dataset should be at least partially transferable to a smaller, more domain-specific datasets. Ideally, this would not only save computational resources but also manual annotation efforts, which are invariably expensive and error-prone, especially if they require the intervention of human experts. In computer vision, a common approach to transfer learning is to pretrain a DCNN on a large general dataset, before fine-tuning (isolated components of) it on a more domain-specific dataset. The layered architecture of modern DCNNs lends itself particularly well to such an approach and numerous studies have reported significant performance gains for a variety of tasks.

The concept of transfer learning [16] is commonly defined through two high level concepts, namely, domain ( $D$ ) and task ( $T$ ). A domain  $D$  relates to the distribution of the training set through a feature space  $\chi$  and a marginal probability distribution  $P(X)$ , where  $X \in \chi$ . Therefore, two domains  $D_1$  and  $D_2$  are considered as different if  $\chi_1 \neq \chi_2$  or  $P_1(X) \neq P_2(X)$ . A task  $T$ , that can be learned from the training set, relates to the given labels and a model  $F$  (or a function), which predicts corresponding labels based on a given domain  $D$ . Finally, given a source domain-task pair ( $D_S, T_S$ ) and a target domain-task pair ( $D_T, T_T$ ), the transfer learning concept is defined as a process that helps to improve the model  $F_T$  (or function) in the target learning task  $T_T$  based on the knowledge obtained from the source domain-task pair ( $D_S, T_S$ ), where  $D_S \neq D_T$  or  $T_S \neq T_T$ .

Our present task can be cast as an inductive transfer learning problem, where labeled data from the source domain is available. In this setting, the source and target tasks are different, while domains can be similar or not. Some instances of labeled data from the target domain are required to fine-tune the pretrained model. Thus, a pretrained DCNN can get fine-tuned in order to improve the generalization to musical instruments in the artistic target domain. In our experiments, we use Inception-V3 [49], which yielded the best performance among other DCNNs, tested in the baseline experiments [15]. We use the DCNN, pretrained on the following datasets: (i) the ImageNet dataset [4]; (ii) the ImageNet dataset and the Rijksmuseum collection [11] with publicly available weights [42].

### Conventional Data Augmentation

Conventional data augmentation techniques [4] include different types of affine transformations applied to the original images. We use the following techniques, available in a reference implementation from the Keras framework [50], with the parameters given between brackets (see the examples in Figure 3): zoom (0.05, 0.1 and 0.15), rotation (5, 10 and 15 degrees), shear transformation (5, 10 and 15 degrees), vertical and horizontal shift (0.05, 0.1 and 0.15) and horizontal flip. We have 2 sets of experiments where we apply:

1. **One distortion.** In these experiments, we generate datasets up to twice the size of the Minerva subsample. We test augmentation ratios multiplied by a factor of 2 from 16:1 (unaugmented: augmented) to 1:16.
2. **Multiple distortions.** First, we determine the best num-

ber of augmented images per distortion group (based on the development set) that should be added to Minerva. The amount of the augmented images can be up to the size of Minerva with ratios multiplied by a factor of 2 from 16:1 (unaugmented: augmented) to 1:1. Finally, we train the model using Minerva (without distortions) and the sum of the best number of distorted images per group.



**Figure 3.** Example of traditional data augmentation (original image, rotation, shear, shift, zoom, flip)

### Style transfer

Style transfer is a class of image processing techniques that change the visual style of an image, while leaving its semantic content largely unmodified. From a human perspective, these algorithms create artificial artworks. Style transfer based on [17] is formally defined through what is known as the style and content losses. The style concept is represented through the correlation between low-level deep convolutional features and it is calculated using Gram matrices. The content concept is defined as the high-level deep convolutional features. Commonly, these features are extracted from a pretrained VGG network [51].

In this work, we utilize a method proposed by Ghiasi et al. [19] which allows to conduct real-time stylization. The method is based on the prediction of the normalization parameters of the style transfer network. The method consists of two DCNNs. The first DCNN, the style prediction network, predicts an embedding vector from an input style image. The second, the style transfer network, conducts the actual stylization of the content image using the embedding vector, which represents the normalization constants for the latter network.

For style transfer, we utilize a pretrained model [19] with publicly available weights. In our experiments, we vary interpolation weights from 0 to 1 with a stepsize of 0.2. We conduct two types of experiments with one distortion and multiple distortions. We apply the model in the two following settings:

1. **Inner style transfer.** In this setting, we randomly transfer styles within the Minerva dataset in the corresponding data splits (see Figure 4). Thus, we obtain a new dataset with permuted styles.
2. **Outer style transfer.** In this setting, we randomly transfer styles from the Minerva dataset to the MIMO dataset in the corresponding data splits (see Figure 5). Thus, we create artificial artworks from photorealistic images.

### Experimental Setup

We utilize the Keras framework [50] with Tensorflow backend to train the models [52]. The conventional categorical cross-entropy loss function is minimized, using the Adam optimizer[53] with a learning rate of 0.0001, over mini-batches of 32 samples. The training process is interrupted as soon the validation loss does not decrease for five epochs in a row.



**Figure 4.** Example of inner style transfer. The upper images correspond to the style and content images, respectively. Both images are derived from the Minerva dataset. The lower images correspond to stylized images with different degree of transformation.



**Figure 5.** Example of outer style transfer. The upper images correspond to the style image and the content image, respectively. The left-side image is derived from the Minerva dataset and the right-side image is from the MIMO dataset. The lower images correspond to stylized images with different degree of transformation.

## Results

We divide our experimental results into four different sections. We first compare the various data augmentation techniques, applied to different data sources. We then investigate the potential of style transfer at the training and inference phase. Finally, we compare two initialization approaches for DCNNs (Rijksmuseum vs. ImageNet).

### Data Augmentation

Table 1 compares various types of data augmentation techniques obtained from different sources. It is clear that each of the tested techniques is beneficial for the art classification problem. We observe that inner augmentation generally outperforms its outer counterpart. However, the horizontal flip is less beneficial compared to the other tested techniques, which is arguably due to the asymmetric shapes of instruments. In one distortion setting, style transfer is the best outer augmentation technique and it is even competitive to the best inner data augmentation methods. Additionally, we observe that style transfer gains additional accuracy points applied on the top of the traditional augmentation techniques. We obtain the best result through the combined application of all techniques and improve the baseline result up to 7 points in accuracy.

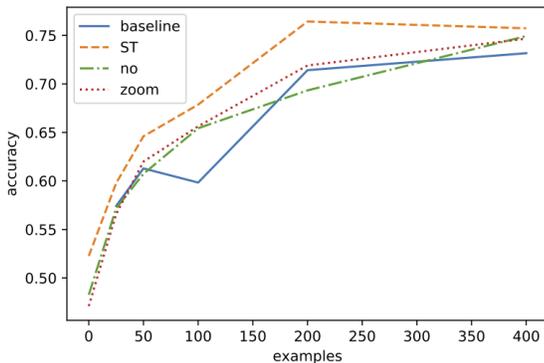
### Training Source

In the previous section, we observed that style transfer is beneficial as a data augmentation technique. However, these results do not show whether artificial artworks can completely substitute authentic ones in the training phase. To answer this question, we now use the MIMO dataset as a main training source and complement it with increasingly large subsamples of the Minerva dataset. From Figure 6, it is clear (when no examples of Minerva are added) that photorealistic images can not substitute the

The results (accuracy) obtained on the Minerva test set in the experiments with different distortions. Minerva is utilized as a main training source and the columns (2, 3, 4) correspond to the source of augmentation. The column Both corresponds to the experiments, where MIMO and Minerva were utilized as the source of augmentation. The row BL corresponds to the baseline experiment, where no distortions were applied. The row ST corresponds to the experiments with style transfer and the row MD corresponds to the multiple distortion setting.

Distortion	Minerva	MIMO	Both
BL	73.17		
Zoom	77.53	73.90	
Rot	76.13	75.07	
Shear	76.27	75.50	
Shift	78.53	75.27	
Flip	75.13	73.00	
ST	77.23	76.40	
MD	77.33	74.43	78.13
MD + ST	78.87	75.50	81.10

artworks in the classification problem, as the accuracy is substantially lower than 73 percent (when 400 examples of Minerva are added and MIMO is not used at all). Additionally, we can observe that the performance of the DCNNs greatly depends on the amount of in-domain data, which is not unexpected. However, style transfer applied to MIMO certainly improves the performance, in comparison to the case when the only training source is Minerva.

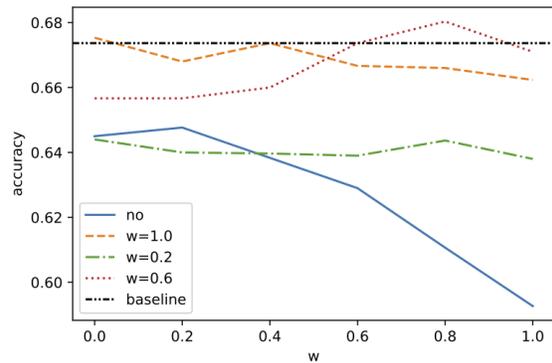


**Figure 6.** The results obtained on the Minerva test set. The model is trained on the distorted MIMO dataset with different amount of examples (undistorted) from the Minerva dataset added. The “no” line corresponds to the case where the MIMO dataset is not distorted. The baseline corresponds to the case where only the Minerva dataset is used in the training set. The line ST corresponds to the experiments with style transfer.

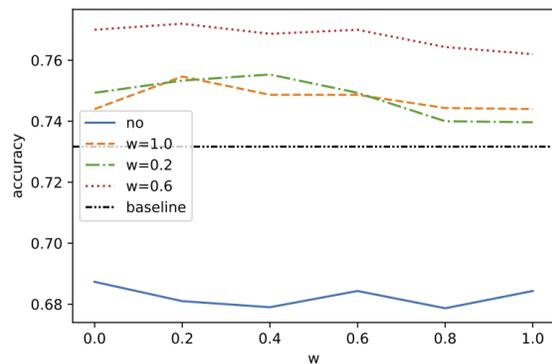
### Inference

Figure 7 and Figure 8 demonstrate how DCNNs trained on artworks react to style transfer distortions. As shown in Figure 7, DCNNs show better results for photorealistic images of

musical instruments than for artificial artworks. When we transfer artistic styles to photorealistic images, we may expect the artificial artworks to be more similar to the artworks. However, we can observe opposite trends. Additionally, we can see that higher degrees of style transfer negatively affects the ability of the DCNNs to recognize photorealistic depictions of musical instruments. From Figure 8, we can observe that the style transfer distortion in the in-domain test set still causes degradation in performance. However, both figures show that if we add any degree of such distortion to the training data, it renders the DCNNs robust to these changes. Therefore, we may additionally observe that the DCNNs perceive artworks and artificial artworks in a very different manner.



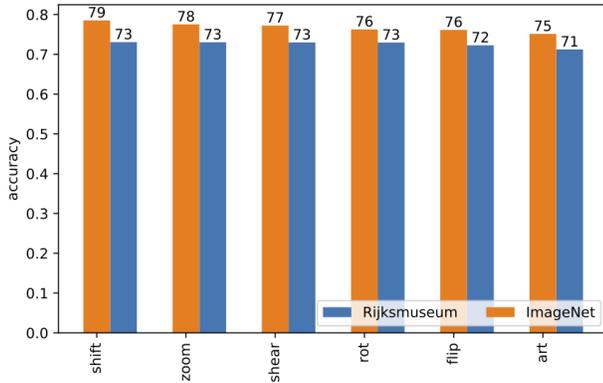
**Figure 7.** The results obtained on the MIMO test set distorted by style transfer. The “w” axis corresponds to different degrees of weights interpolation for style transfer in the test set. The model is trained on Minerva augmented with inner style transfer. The “no” line corresponds to the Minerva training set without distortions. The baseline corresponds to the MIMO test set and the Minerva training set without distortions (in both of them).



**Figure 8.** The results obtained on the Minerva test set distorted by style transfer. The “w” axis corresponds to different degrees of weights interpolation for style transfer in the test set. The model is trained on the Minerva dataset augmented with inner style transfer. The “no” line corresponds to the Minerva training set without distortions. The baseline corresponds to the Minerva test set and training set without distortions (in both of them).

## Negative Transfer From Heritage Domain

Figure 9 compares two weights initialization for DCNNs (Rijksmuseum vs. ImageNet). Pretraining on ImageNet is much more beneficial for our task compared, to the Rijksmuseum dataset initialization.



**Figure 9.** Comparison between two initialization approaches (Rijksmuseum vs. ImageNet) for the experiment where only one distortion is applied. The DCNN pretrained on ImageNet outperforms the DCNN pretrained on Rijksmuseum in all cases. Therefore, we observe how training from in-domain data can lead to worse transferability.

## Discussion

Many GLAM (Galleries, Libraries, Archives, and Museums) institutions rapidly go through a process of digitization of their cultural heritage collections that leads to publicly available datasets of artworks. However, not many institutions can afford the digitization due to high cost of annotations. The digital heritage domain obviously differs from the general domain, where untrained annotators can often be resorted to, and hence, requires high-skilled subject experts that manually annotate large collections. Additionally, DCNNs require large datasets for training from scratch and are highly sensitive to the amount of training data available [54]. Therefore, computational approaches that improve DCNNs when tackling small datasets are in high demand and can help the institutions to significantly speed up large-scale cataloguing campaigns.

Style transfer has proved to be an effective data augmentation applied to in-domain data. In this case, in-domain images are distorted with other styles from the same dataset to make a DCNN more robust. We also tackled this problem from another perspective and tried to create more in-domain-like examples. We investigated the effectiveness of style transfer as a data augmentation. The results demonstrate that both inner and outer style transfer are highly effective and can be applied in combination with conventional data augmentation techniques.

Arguably, the human eye perceives artificial artworks as very similar to photorealistic depictions and artificially generated artworks might help to improve classification performance. Therefore, it raises another question if DCNNs perceive artworks and artificial artworks similarly. In this case, artworks can be substituted by artificial artworks. When we transfer styles to photorealistic images, we may expect the artificial artworks to become more similar to the artworks. However, a DCNN fine-tuned on the

artistic images is a better predictor for photorealistic images than for artificially stylized artworks. Therefore, the DCNN considers artificial artworks less similar to the artworks than photorealistic images. Additionally, we showed that a DCNN fine-tuned on artificial artworks performs substantially worse compared to real artworks. Consequently, art classification is still very much dependent on the availability of in-domain data. Disappointingly, this highly appealing idea of artificial artworks that could have helped in substitute artistic work did not work out well.

In artwork classification, DCNNs pretrained on artworks look appealing due to initialization from ImageNet may seem to be too general. However, similarity of target tasks also matters for the transferability of pretrained networks [55] that may lead to negative transfer [16]. In the previous section, we observed an example of negative transfer, when pretraining on in-domain data does not lead to better results. We are not first who observe this collision. Romero et al. [56] conducted classification of human body parts in the medical domain and found that pretraining on in-domain data from another part of the body has little advantage compared to Imagenet. Cetinic et al. [38] demonstrated that transferability of deep representations for art classification is task dependent. Sabatelli et al. [42] demonstrated transferability from the in-domain initialization across similar tasks is better than from ImageNet. However, we observe that transferability across different tasks even in the same domain may be not preserved.

## Conclusion

This paper investigated the potential of style transfer as well as the usage of external photorealistic data sources for image classification in the artistic domain. As a novel contribution, we have compared inner and outer style transfer to conventional data augmentation techniques. Unsurprisingly, we observed that almost all of the tested data augmentation techniques improve classification performance. Even conventionally augmented, external data from the photorealistic domain consistently helped to gain additional accuracy points. We demonstrated that outer style transfer is the best outer augmentation method and inner style transfer is competitive to other inner augmentation methods. Additionally, we showed both types of style transfer are beneficial to apply in addition to conventional data augmentation. We used style transfer as a method to generate artificial artworks from photorealistic images and investigated if these artworks could substitute authentic depictions from the art domain. However, while we observed that these artificial artworks can help to improve results, they should not be used as the only source of training data. In future work, we shall investigate the effect of style transfer to larger, more skewed classification problems, that are meaningful to art historians and consider a more inclusive range of objects (such as varieties of fruits and mammals in the visual arts). Likewise, the application of style transfer in object detection is high on our agenda.

## References

- [1] Marco Fiorucci, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, and Stuart James. Machine learning for cultural heritage: A survey. *Pattern Recognition Letters*, 133:102–108, 2020.
- [2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

- [3] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [5] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [6] Arran. <https://github.com/ickramer/Arran>.
- [7] Gustavo Carneiro, Nuno Pinho Da Silva, Alessio Del Bue, and João Paulo Costeira. Artistic image classification: An analysis on the printart database. In *European conference on computer vision*, pages 143–157. Springer, 2012.
- [8] Europeana. <https://www.europeana.eu/>.
- [9] Nicolas Gonthier, Yann Gousseau, Said Ladjal, and Olivier Bonfait. Weakly supervised object detection in artworks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.
- [10] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. Recognizing image style. *arXiv preprint arXiv:1311.3715*, 2013.
- [11] Thomas Mensink and Jan Van Gemert. The rijksmuseum challenge: Museum-centered visual recognition. In *Proceedings of International Conference on Multimedia Retrieval*, pages 451–454, 2014.
- [12] Gjorgji Strezoski and Marcel Worring. Omniart: multi-task deep learning for artistic data analysis. *arXiv preprint arXiv:1708.00684*, 2017.
- [13] Web gallery of art. <https://www.wga.hu/>.
- [14] Michael J Wilber, Chen Fang, Hailin Jin, Aaron Hertzmann, John Collomosse, and Serge Belongie. Bam! the behance artistic media dataset for recognition beyond photography. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1202–1211, 2017.
- [15] Matthia Sabatelli, Nikolay Banar, Marie Cocriamont, Eva Coudyzer, Karine Lasaracina, Walter Daelemans, Pierre Geurts, and Mike Kestemont. *Advances in Digital Music Iconography. Benchmarking musical instrument localization in non-photorealistic images*, page to be added. Zenodo, May 2020.
- [16] Ricardo Ribani and Mauricio Marengoni. A survey of transfer learning for convolutional neural networks. In *2019 32nd SIBGRAP Conference on Graphics, Patterns and Images Tutorials (SIBGRAP-T)*, pages 47–57. IEEE, 2019.
- [17] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.
- [18] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016.
- [19] Golnaz Ghiasi, Honglak Lee, Manjunath Kudlur, Vincent Dumoulin, and Jonathon Shlens. Exploring the structure of a real-time, arbitrary neural artistic stylization network. *arXiv preprint arXiv:1705.06830*, 2017.
- [20] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
- [21] Zheng Xu, Michael Wilber, Chen Fang, Aaron Hertzmann, and Hailin Jin. Learning from multi-domain artistic images for arbitrary style transfer. *arXiv preprint arXiv:1805.09987*, 2018.
- [22] Philip T Jackson, Amir Atapour-Abarghouei, Stephen Bonner, Toby P Breckon, and Boguslaw Obara. Style augmentation: Data augmentation via style randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 83–92, 2019.
- [23] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
- [24] Anselm Brachmann and Christoph Redies. Computational and experimental approaches to visual aesthetics. *Frontiers in computational neuroscience*, 11:102, 2017.
- [25] Eva Cetinic and Sonja Grgic. Automated painter recognition based on image feature extraction. In *Proceedings ELMAR-2013*, pages 19–22. IEEE, 2013.
- [26] Daniel Keren. Painter identification using local features and naive bayes. In *Object recognition supported by user interaction for service robots*, volume 2, pages 474–477. IEEE, 2002.
- [27] Lior Shamir, Tomasz Macura, Nikita Orlov, D Mark Eckley, and Ilya G Goldberg. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception (TAP)*, 7(2):1–17, 2010.
- [28] Jana Zujovic, Lisa Gandy, Scott Friedman, Bryan Pardo, and Thrasyvoulos N Pappas. Classifying paintings by artistic genre: An analysis of features & classifiers. In *2009 IEEE International Workshop on Multimedia Signal Processing*, pages 1–5. IEEE, 2009.
- [29] 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.
- [30] Zoe Falomir, Lledó Museros, Ismael Sanz, and Luis Gonzalez-Abril. 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, 2018.
- [31] Thomas Edward Lombardi. *The classification of style in fine-art painting*. Citeseer, 2005.
- [32] Yaniv Bar, Noga Levy, and Lior Wolf. Classification of artistic styles using binarized features derived from a deep neural network. In *European conference on computer vision*, pages 71–84. Springer, 2014.
- [33] Eva Cetinic and Sonja Grgic. Genre classification of paintings. In *2016 International Symposium ELMAR*, pages 201–204. IEEE, 2016.
- [34] Omid E David and Nathan S Netanyahu. Deeppainter: Painter classification using deep convolutional autoencoders. In *International conference on artificial neural networks*, pages 20–28. Springer, 2016.
- [35] 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.
- [36] Nanne Van Noord, Ella Hendriks, and Eric Postma. Toward discovery of the artist’s style: Learning to recognize artists by their artworks. *IEEE Signal Processing Magazine*, 32(4):46–54, 2015.
- [37] 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.
- [38] Eva Cetinic, Tomislav Lipic, and Sonja Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems*

- with Applications, 114:107–118, 2018.
- [39] Ahmed Elgammal, Bingchen Liu, Diana Kim, Mohamed Elhoseiny, and Marian Mazzone. The shape of art history in the eyes of the machine. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [40] Christian Hentschel, Timur Pratama Wiradarma, and Harald Sack. Fine tuning cnns with scarce training data—adapting imagenet to art epoch classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3693–3697. IEEE, 2016.
- [41] Adrian Lecoutre, Benjamin Negrevergne, and Florian Yger. Recognizing art style automatically in painting with deep learning. In *Asian conference on machine learning*, pages 327–342, 2017.
- [42] Matthia Sabatelli, Mike Kestemont, Walter Daelemans, and Pierre Geurts. Deep transfer learning for art classification problems. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.
- [43] Tiancheng Sun, Yulong Wang, Jian Yang, and Xiaolin Hu. Convolution neural networks with two pathways for image style recognition. *IEEE Transactions on Image Processing*, 26(9):4102–4113, 2017.
- [44] Eva Cetinic, Tomislav Lipic, and Sonja Grgic. A deep learning perspective on beauty, sentiment, and remembrance of art. *IEEE Access*, 7:73694–73710, 2019.
- [45] Eva Cetinic, Tomislav Lipic, and Sonja Grgic. Learning the principles of art history with convolutional neural networks. *Pattern Recognition Letters*, 129:56–62, 2020.
- [46] Matthia Sabatelli, Mike Kestemont, and Pierre Geurts. Improving the training of deep convolutional neural networks for art classification: from transfer learning to multi-task learning. In *The 6th Digital Humanities (DH) Benelux Conference*, 2019.
- [47] 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.
- [48] Emily I Dolan. Mimo: Musical instrument museums online, 2017.
- [49] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [50] François Chollet et al. Keras. <https://keras.io>, 2015.
- [51] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [52] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [53] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [54] Leonard Berrada, Andrew Zisserman, and M. Pawan Kumar. Smooth loss functions for deep top-k classification. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.
- [55] Hong Liu, Mingsheng Long, Jianmin Wang, and Michael I Jordan. Towards understanding the transferability of deep representations. *arXiv preprint arXiv:1909.12031*, 2019.
- [56] Miguel Romero, Yannet Interian, Timothy Solberg, and Gilmer Valdes. Training deep learning models with small datasets. *arXiv preprint arXiv:1912.06761*, 2019.

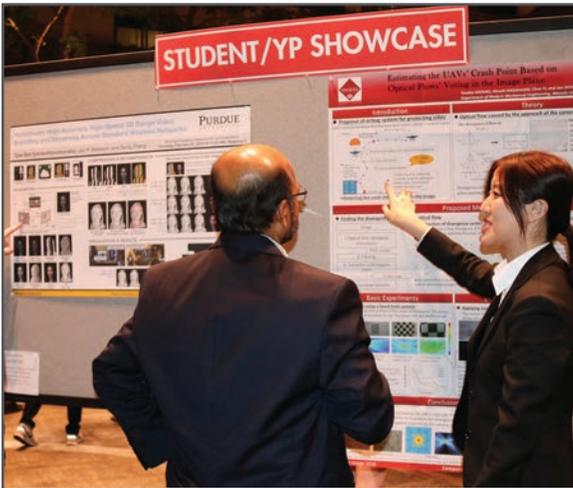
**JOIN US AT THE NEXT EI!**

IS&T International Symposium on

# Electronic Imaging

SCIENCE AND TECHNOLOGY

*Imaging across applications . . . Where industry and academia meet!*



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

[www.electronicimaging.org](http://www.electronicimaging.org)

