Improving the Training of Deep Convolutional Neural Networks for Art Classification: from Transfer Learning to Multi-Task Learning

Matthia Sabatelli[1], Mike Kestemont[2] and Pierre Geurts[1]

[1]Department of Electrical Engineering and Computer Science, University of Liège

[2]Antwerp Center for Digital Humanities and Literary Criticism (ACDC), Universiteit Antwerpen, Belgium

Deep Convolutional Neural Networks have become the most popular algorithm for Computer Vision (CV) problems. However, they are well known to be particularly hard to train. A large set of possible hyperparameters, combined with the need for large amounts of training data, can put serious constraints on their optimization procedure. In this paper we explore several strategies which can facilitate their training when classifying images representing heritage objects, a CV area which is relatively unexplored.

MOTIVATION

The rapid process that has led museums to digitize large parts of their cultural heritage has promoted the creation of datasets that can serve researchers working at the intersection between machine learning and digital heritage. In fact, the research presented hereafter uses a dataset that comes from the Rijksmuseum in Amsterdam[1]. Particularly worth exploring is the field of Deep Learning[2], which has recently seen the development of state of the art algorithms in Computer Vision (CV). Most of such algorithms however, are designed to perform well on natural images[3], yet such a source domain can be very different from images representing heritage objects in terms of resolution, size, and texture. Thus, the potential of Deep Convolutional Neural Networks (DCNNs) as a tool for classifying heritage objects still represents an open research question. In what follows, we explore different strategies that can be used for successfully training such networks. The aim is to train a DCNN so that it will be able to recognize who the *artist* of a heritage object is, which *material* has been used to create it, and in what *artistic category-type* such a sample falls into.

TRAINING APPROACHES

The explored training approaches can be divided into two categories: Transfer Learning and Multi-Task Learning. The first approach explores whether it is helpful to initialize the weights of the DCNNs with the ones that are obtained from training them on natural images in the context of the ImageNet challenge. Furthermore, it investigates if such an initialization can be beneficial enough to simply retrieve `off the shelf' features[4] from the networks. If so, such features can be classified without having to compute any backpropagation operations, resulting in faster training. If this is not the case, in order to yield successful performance such DCNNs need to be `fine-tuned'. The neural architectures used for this set of experiments are: VGG19[5], Inception-V3[6], Xception[7] and ResNet50[8]. In a Multi-Task scenario we no longer train a single architecture for each classification challenge, but all tasks are tackled simultaneously. The motivation behind this approach lies in the fact that there might exist a correlation between the creator of an artwork, and e.g. the material he/she has used in his/her creation. In other words, if the type associated to the heritage object is a sculpture made out of bronze, the chances that its creator is *Van Gogh* are very low, given that *Van Gogh* is a very well known painter. From a statistical point of view, addressing all tasks with a single network might therefore help reducing overfitting problems, while also representing an important advantage from a computational perspective.

TRANSFER LEARNING RESULTS

Figure 1 shows the accuracies that have been obtained at testing time on a separate 10% subsample of the Rijksmuseum dataset. We can see how fine-tuning the DCNNs yields significantly better results in almost all classification tasks. Furthermore, the benefits of an ImageNet initialization are also clear from the results.

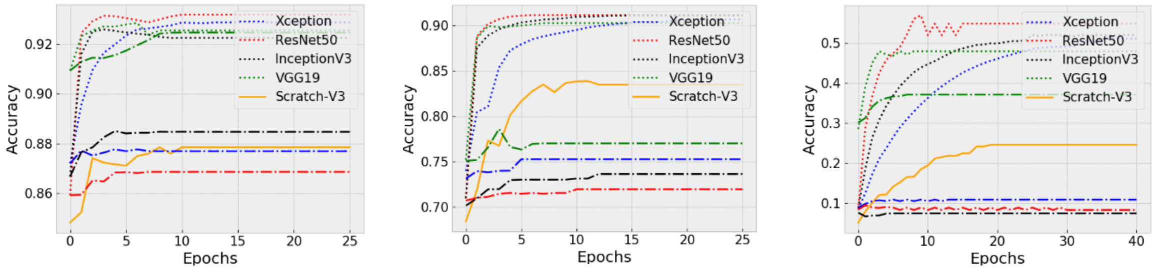

Figure 1 : From left to right the results obtained on the ‘material’, ‘type’ and ‘artist’ classification tasks. The dotted lines report the `fine-tuned' accuracies, the dashed ones the accuracies from the `off the shelf' approach, while the solid orange line corresponds to an Inception-V3 DCNN trained completely from scratch (results coming from[9]).

MULTI-TASK LEARNING RESULTS

The results presented in Figure 2 have been obtained on three smaller subsets of the Rijksmuseum dataset, and show how a Multi-Task learning approach is able to yield even better performance than when a DCNN is `fine-tuned'. The neural architecture used in this set of experiments is ResNet50**.**

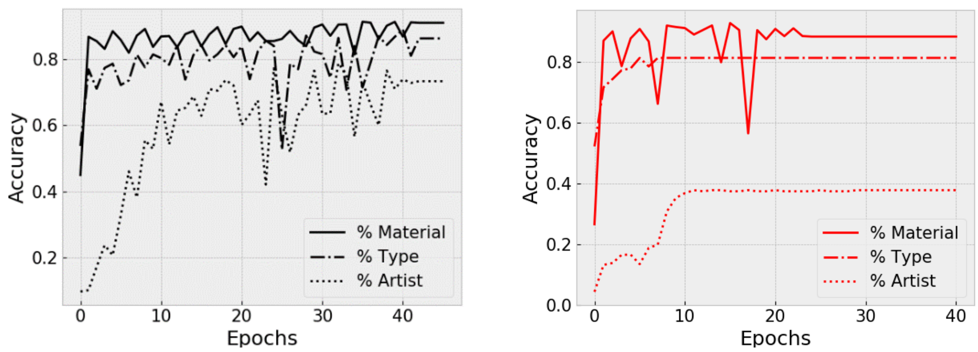

Figure 2 : The benefits obtained from training a DCNN in a Multi-Task Learning setting (left figure) compared to three different individually fine-tuned networks (right figure). We see that the first approach yields significantly higher accuracies for two out of the three classification tasks.

CONCLUSION

In this work we have tackled the problem of successfully training DCNNs for classifying heritage objects. Our results show how both the Transfer Learning approach and the Multi-Task Learning one can yield promising results, with the latter strategy being so far the best performing one.

REFERENCES

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

[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning.*nature*, *521*(7553), 436.

[3] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*,*115*(3), 211-252.

[4] Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 806-813).

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

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[7] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[8] 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).

[9] Sabatelli, M., Kestemont, M., Daelemans, W., & Geurts, P. (2018). Deep transfer learning for art classification problems. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 631-646).