Object Classification in Images of Neoclassical Artifacts Using Deep Learning

Bernhard Bermeitinger, Simon Donig, Maria Christoforaki, André Freitas, Siegfried Handschuh

1. Classifying Aesthetic Forms – a Methodology at the Heart of Art History

The transformation of aesthetic styles has been at the heart of Art History since its inception as a scholarly discipline in the late eighteenth century. Analyzing the single artifact and the carefully curated corpus have been the techniques for crafting hermeneutic understanding for such processes of change. Recently new instruments based on statistical techniques empower us for a fresh take on bodies of sources once disregarded as second tier complementary sources such as for instance very large corpora.

2. The Neoclassica Research Framework

The *Neoclassica* research framework (Donig, Christoforaki, and Handschuh (2016)) was conceived to provide scholars with such new instruments and methods for analyzing and classifying artifacts and aesthetic forms from the era of Classicism (ca. 1760–1860). The Neoclassic movement was of almost global scale—affecting architecture and design from Sidney to New York, and from Athens to the outreach of the Russian Urals—while relating to a common reference in Classical Antiquity, therefore making it an almost ideal topic for studying processes of stylistic transformation.

It accommodates both traditional knowledge representation as a formal ontology and data-driven knowledge discovery, where cultural patterns will be identified by means of algorithms in statistical analysis and machine learning, having in particular the potential to uncover hitherto unknown patterns in the source data. The outcomes of both the top-down and the bottom-up approach will be united in a consistent, unified formal knowledge representation.

Motivated by the need to combine object classification with domain knowledge representation, the ontology at the moment focuses on artifacts (in particular furniture and architecture) and their components. Following the preliminary hypotheses that aesthetic forms in furniture and architecture are in closest communication with each other due to constructional commonalities and their shared reference of the Classic, we decided to start developing the knowledge discovery module of Neoclassica by classifying artifacts in digital images.

3. Knowledge discovery

In this paper, we report on our efforts for using Deep Learning for classifying artifacts in digital visuals as a part of the Neoclassica framework. We chose a Deep Learning approach for our classification method because of its current superiority over other methods and still rising accuracy over the last years in nearly all image classification and object detection challenges.

Initially, we compiled a body of images both from commercial sources such as auction houses and antique dealers as well as public domain images provided by museums, private collectors, and the scholarly community. Due to the complex copyright situation, this corpus can not be redistributed. In order to make our experiment reproducible and since the Metropolitan Museum of Art has released 375,000 images of artifacts in the public domain[1] from February 7th, 2017. we assembled a small corpus of 379 artifacts relevant to our research. We processed this corpus with the same algorithm as the original privately owned corpus and released the data together with the source code that applies the algorithm[2].

3.1. Classifier description

The main classifier for our experiments is a *Convolutional Neural Network* (CNN). It classifies an input image as a whole. Although it may be configured to be a *Multi-Label* classifier and thus classifying different objects in an image, there is naturally no link between regions in an image and the classifier's output.

In a first step, we applied a standard implementation of a CNN (namely *VGG19* (Russakovsky et al. (2015))). The results were not satisfactory for our needs. It classified the type of the object depicted in the image with an accuracy of 0.37 for the initial corpus.

In a second step, we decided to employ pre-training, a common technique for improving accuracy in neural networks. We experienced that available pre-trained classifiers for generic image classification proved ineffective in our case. Most of them are trained on a subset of *ImageNet* (J. Deng et al. (2009)), containing 1000 classes. These classes are broadly spread around everyday objects like "dogs", "cats" and "planes" but also "chairs" and "tables". This led us to assume that the amount of very different classes (e.g. "bottle" and "car") that occur nowhere in our corpora interfere with the recognition. Following that hypothesis, we decided to train the algorithm on a more specific subset compiled from *ImageNet* mainly containing different furniture objects like tables, chairs, sofas and cabinets. They sum up to 35,000 images, more than ten times the amount of images in our corpus. The first training step with these images resulted in an accuracy of 0.54 of classifying the object correctly.

3.2. First layout

The first corpus contained 2129 images representing 300 European period artifacts mostly in a colored format of highly diverging quality and resolution. They depict the objects fully or partially or are close-up shots of specific forms. We coarsely annotated these images by manually labeling them on the level of folders. The concepts applied during this labeling process are directly taken from the *Neoclassica* ontology and describe concepts for types of artifacts. These concepts were derived from period sources.

The depth of the class hierarchy was partly reflected by the folder structure. To avoid diminishing the total number of instances available for the training by nesting too many folders we stopped when the attained level of specification was considered to be sufficient to adequately

1 http://www.metmuseum.org/press/news/2017/open-access
2 http://www.neoclassica.network/resources

represent the concept by the domain expert. For instance, the folder labeled *Chest of drawers* contains all instances of this class. Their labels in turn reflect the names of all the sub-classes in the most extensive specification (e.g. *Semainier*, *Wellington Chest*, *Commode scriban*).

After pre-training, the next step was fine-tuning with this corpus. The accuracy after these two steps was 0.438. The F1 measure was 0.442. The improvement was 19% in accuracy and 32% in F1 measure.

3.3. Second layout

The second corpus was assembled from the open data released by the Metropolitan Museum of Art. It contains 1246 images representing 379 European and American period artifacts ranging roughly from 1780–1840 including some transition pieces, drawings, and prints. They also depict the objects fully or partially or are close-up shots of specific forms. We used the titles provided by the Museum aligning them with the Neoclassica ontology and annotated the images in a similar way as with the initial corpus.

The overall mean accuracy over all classes was 0.36, the F1 measure 0.21. For the computation of these numbers, all results that are non-computable[3] were removed. These low numbers result from the existence of two many artifacts represented by only one image, thus making a split in training and testing data meaningless. However, applying pre-training using same ImageNet corpus as in the first layout yielded a mean accuracy over all classes of 0.59 and a F1 measure of 0.58.

In order to achieve better results and since the classifier classifies the image as whole and not features, we excluded all images that did not depict the whole artifact. However, we kept multiple images of the same object. We also kept multiple copies of the same image if they were used to describe different but similar object (e.g. the same image representing multiple chairs in a set). We also split the images depicting multiple objects so that the resulting images represent only one artifact. In that case, we also processed the images so that neighboring objects were covered with solid colors. The images that could not be split (e.g. room interiors) were excluded from the corpus. With the goal of having as few variants of single image as possible, in the case of images of black-and-white drawings and prints recorded with different contrast settings, we kept the ones with better visual quality. However, with colored images of drawings and prints we kept images with different contrast settings since we assumed them to contain different information of relevance to the classifier.

Using the same settings with the curated corpus and with pre-training we achieved an overall mean accuracy over all classes of 0.77 and an F1 measure of 0.76.

3.4. Challenges

While pre-training and improving the curation process helped us to raise the accuracy, we assume that there is room for improvement. But to reach a better result, we yet have to overcome some challenges:

3 This happens when there are either no images for this class in the training set or in the testing set.

Parameters to be taken into consideration include the small size of the corpora in humanities and how to overcome this limitation since this limits the effectiveness of a neural net. Additionally, since pre-training has been proven to optimize the results, it is rational to assume that a pre-training corpus better suited to period artifacts would improve the results further. For instance, a modern bed may have features not found in the Neoclassic style, so the connection between these objects is not trivially transferred and must be retrained. Third, our experiment was probably negatively affected by the limitation of the standard implementation of the CNN which classifies the image as a whole and not parts of it. Therefore, images depicting multiple objects had to be split during the curation process. However, we could spare this effort and even get more information about the nature of an artifact by for example including interior scenes.

Outlining parts inside an image and classifying them is a difficult task for any machine learning method. Recently, a new type of neural net emerged that tackles this challenge: *Regional CNN*. It is implemented most prominently in an algorithm called *MultipathNet* (Zagoruyko et al. (2016)).

4. Current improvements: Using a Regional CNN

At the time of writing, we are manually annotating the images with its classes from the ontology and improving the accuracy of the RCNN of locating and classifying the annotated parts of the images.

The implementation of the RCNN is divided into two steps. First, it detects objects in the image and draws their outline freely and not with bounding boxes other methods do. The second step is classifying the outlined objects using a standard, advanced CNN (Zagoruyko et al. (2016)).

Preliminary results with the standard implementation of *MultiPath* with already pre-trained settings show that the first step of the RCNN already achieves good results in locating main objects in our corpora within reasonable limits. The base corpus for pre-training is COCO (T. Y. Lin et al. (2014)). It is similar to *ImageNet*, hence it's based on current-century everyday objects. Naturally, specific domain objects are not found and classes are rather generic. For further classification, fine-tuning on a fully annotated corpus is essential.

An RCNN requires a more detailed corpus for training. All images must have class annotations with bounding boxes. This exhaustive task promises higher quality in locating objects (first step) and is necessary to classify the objects and parts of it according to the ontology (second step).

5. Conclusion & Future work

We decided thus to take two steps in the near future for improving our results.

First, we are compiling a new corpus to train the RCNN with, avoiding pitfalls like inconsistent quality, heterogeneous image rights and an inadequate distribution of image per class. Here we would like to go a dual approach. Together with domain experts, we intend to collate a corpus from the large repositories of a major auction house, providing us not only with a selection of artifacts but also with texts that could be used in a future multimodal analysis.

On the other hand, this kind of artifacts may exhibit provenance issues (e.g. heterogeneity or lack

of provenance). We will thus compensate for such issues by digitizing a major corpus of Neoclassical artifacts forming an ensemble and comprising artifacts in multiple modes having evolved in close reference to each other. Therefore, we have entered a partnership with the Dessau-Wörlitz UNESCO World-heritage site, an almost untouched complex of manor houses and their furnishings in early Neoclassical style.

Regarding the annotations, we started the development of our own semantic annotation and ontology population tool since January 2017. The tool will create a fully annotated corpus suitable for the RCNN. The actual annotation process will be conducted in cooperation with emerging domain experts from the chair of Visual Culture and Art History at the University Passau.

References

Deng, J, W Dong, R Socher, L.-J. Li, K Li, and L Fei-Fei. 2009. “ImageNet: A Large-Scale Hierarchical Image Database.” In *CVPR09*.

Donig, Simon, Maria Christoforaki, and Siegfried Handschuh. 2016. “Neoclassica - A Multilingual Domain Ontology.” In *2nd Ifip International Workshop on Computational History and Data-Driven Humanities*, edited by Bozic, Mendel-Gleason, Debruyne, and O’Sullivan.

Lin, Tsung Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. “Microsoft COCO: Common objects in context.” In *Lecture Notes in Computer Science*, 8693 LNCS:740–55. PART 5. doi:10.1007/978-3-319-10602-1_48.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. 2015. “ImageNet Large Scale Visual Recognition Challenge.” *International Journal of Computer Vision* 115 (3): 211–52. doi:10.1007/s11263-015-0816-y.

Zagoruyko, Sergey, Adam Lerer, Tsung-Yi Lin, Pedro O. Pinheiro, Sam Gross, Soumith Chintala, and Piotr Dollár. 2016. “A MultiPath Network for Object Detection.” In *BMVC*. 1. http://arxiv.org/abs/1604.02135.