

Visual artefacts through the Black Box: Analysing Deep Learning classification of Neoclassical furniture images

Simon Donig, Maria Christoforaki, Bernhard Bermeitinger and Siegfried Handschuh
University of Passau, simon.donig@uni-passau.de

The recent trend towards introducing digital instruments in the history of art is glaring at best. Like so many disruptive technologies new techniques in computer vision and natural language processing are about to fundamentally change the way we conduct and even envision our very disciplines. Image classification experiments for instance have become a trending topic in visual culture research and more traditional art history alike. Prominent among them are Artificial Intelligence Deep Learning techniques such as Convolutional Neural Networks, (Krizhevsky, Sutskever, and Hinton 2012). New instruments however harbor new methodological and heuristic challenges since these techniques are a black box through which images are passed to create what is perceived as hermeneutic knowledge (Castelvecchi 2016).

There are essentially two ways to approach the methodological dilemma of the black box - either by dissecting it in order to assess best how it operates (Mahendran and Vedaldi 2015; Dosovitskiy and Brox 2016; Zeiler and Fergus 2013), or by trying to improve the procedures surrounding the process in order to make them more accurate and meaningful for an application in the field of the Digital Humanities. In this paper we present a contemplation of the application results of Deep Learning in image classification of Neoclassical artefacts, specifically a standard implementation of a CNN (Bermeitinger, Donig, Christoforaki, et al., 2017), and the impact of data curation and augmentation on them.

In the Neoclassica Framework (The Neoclassica Project, 2017), which aims to provide art historians, historians and social anthropologists with new instruments and methods for analysing and classifying material culture and aesthetic forms from the era Classicism, we constantly face the challenge of grasping and enhancing the operation of the black box. Initially we are focussing on works of applied art (for instance furniture and furnishings) as well as architecture and their visual representation, following the preliminary hypotheses that these *modi* show aesthetic forms in closest communication with each other due to constructional commonalities and their shared reference of the Classical.

While Neoclassical artefacts show a trove of similarities in their visual features across a huge geographic range they also diverge due to local traditions, needs and limitations or political conditions supporting the use of particular vocabularies of aesthetic forms, such as the love for Egyptian-Revival forms in the English aristocracy or for Neo-Grecian furnishings among the German elites expressing opposition to the Napoleonic French Empire Style (Harrison-Moore 2007).

We consider it most promising to follow these trajectories in a fresh way, applying the instruments provided by the Neoclassica system, joining a top down approach harnessing the domain expert knowledge by the means of a newly developed ontology (Donig, Christoforaki & Handschuh, 2016), and a data-driven bottom up approach exploiting the powers of computationally processing large amounts of multimodal (text and image) data, where we incorporate

the Deep Learning algorithm for the classification of period artefacts from images (photos, prints and drawings).

We collated a fairly simple yet open corpus derived from the data recently released in the public domain by the Metropolitan Museum of Art (The Metropolitan Museum of Art, 2017). Our corpus comprises of 1246 images depicting fully, partially, or are close-up shots of specific forms of 379 artefacts displaying Neoclassical features, chiefly furniture, small works of art (such as bronzes, silverware etc.) and their depiction in graphics and print (e.g. pattern books). The objects depicted belong to a variety of unequally distributed kinds, as to be expected, since among the furniture preserved chairs, for example, are more common than architect's tables. The geographic range of these artefacts is mostly American East Coast, with a still considerable number of French and some German pieces.

The corpus was assembled by scraping all the images of the designated artefacts supplied by the MET webpages. We annotated the corpus, manually mapping the MET provided titles to the Neoclassica ontology. We then went through an iterative process of fine tuning. After each step we both recorded accuracy and F1 score and conducted visual assessments of the training and test set, hypothesizing about the impact of specific kinds of data on the classification process. We started by improving the input by applying standard techniques such as pre-training with a subset of furniture images from ImageNet (Deng et al., 2009) as well as cleaning and augmenting of the data. Since a CNN classifies images as a whole, we hypothesized that an input of images depicting interiors, furniture sets or suits, and close-ups of specific furniture features, would compromise the results, since the classifier will not be able to correctly identify the features associated with each class. Thus, we excluded all images depicting furniture features like close-ups. Additionally we split images depicting multiple objects so that each image would contain only one artefact and processed them so that neighbouring objects were covered with solid colours. The images that could not be split (e.g. room interiors) were in the end excluded from the corpus. This procedure improved the overall mean accuracy from 0.36 to 0.77 and the F1 measure from 0.21 to 0.76. We ran the algorithm 20 times with different partitions of the corpus in different training and test sets producing highly similar results with each run.

In a final step we summarized the qualitative assessments with regard to whether specific artefacts were successfully, or rather in the most interesting case, not successfully classified.

As may have been expected the most populous classes yielded greater accuracy. For instance the chair and arm-chair classes are the most populous, having the additional benefit of displaying rather clean shapes and of being relatively consistent over historical epochs (Boyce and Butler 2013: 54), giving additional weight to pre training process.

Another interesting observation was that, usually, artefacts in drawings were classified correctly together with their counterparts depicted in photographs. This made us consider augmenting our corpus with drawings from period pattern books, thus enhancing our repository with other public domain images. Additionally, this will enable the domain expert to trace spe-

cific patterns found on physical pieces and relate them to design drawings such as by Sheraton, Percier, Dannhauser, etc.

We also realized that depictions of different artefacts are sometimes indistinguishable for the algorithm, either because artefacts are made to mimic each other (e.g. a *Sécretaire à abat-tant* and a German *Blenderschrank*) or due to the perspective and medium applied in depictions, as in the case of a drawing of a canopied bed which was incorrectly identified as a fire-screen.

Finally, another rather severe limitation of the algorithm is that on encounter of artefacts that cannot be classified, it assigns them to the most populous class (in our case a chair). This poses the problem of having a corpus with unequal distribution of images and how this should be handled.

Regarding these limitations, we hypothesise that we can obtain better results by multimodal analysis, which is shown to produce promising results in distributional semantics (Bruni, Tran, and Baroni 2014). Text mining period pattern books, encyclopedias, and architectural treatises, or historical catalogues raisonnés can provide information that goes way beyond anything that can be conveyed in a visual representation of the artefact, pertaining to its description (e.g. parts of the object that are not shown in the image) as well as its provenance, manufacturer, dating, etc.

Regarding the image classification in particular, we are currently implementing feature classification using a Regional CNN algorithm (Zagoruyko et al. 2016). Additionally, we are contemplating testing other methods such as SIFT (Lowe 1999) in order to assess their performance in the given task and the possibility of combining a variety of techniques to optimise results.

Information both from text and image could be structured and semantically augmented by domain knowledge according to the Neoclassica ontology and used for reasoning by a query answering system, comprising an effective research tool for the domain expert.

References

- Bermeitinger, B., Donig, S., Christoforaki, M., Freitas, A. and Handschuh, S. (2017): "Object Classification in Images of Neoclassical Artifacts Using Deep Learning.", DH2017, Montréal, Canada, August 8-11 2017 (peer-reviewed, forthcoming)
- Boyce, C. and Butler, J. T. (2013): *Dictionary of Furniture: Third Edition*. New York, NY Skyhorse Publishing.
- Bruni, E., Nam-Khanh T. and Baroni, M. (2014): "Multimodal Distributional Semantics." *J. Artif. Intell. Res. (JAIR)* 49: 1–47.
- Castelvecchi, D. (2016): "Can We Open the Black Box of AI?" *Nature News* 538 (7623): 20. doi:10.1038/538020a.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009). *ImageNet: A Large-Scale Hierarchical Image Database*. *CVPR 09*: 248-255.
- Donig, S., Christoforaki, M. and Handschuh, S. (2016): "Neoclassica - A Multilingual Domain Ontology. Representing Material Culture from the Era of Classicism in the Semantic Web." In: B. Bozic, G. Mendel-Gleason, C. Debruyne, and O'Sullivan, D. (eds.). *Computational History and Data-Driven Humanities. CHDDH 2016. IFIP Advances in Information and Communication Technology*, vol 482. Cham, Springer: 41–53. doi: 10.1007/978-3-319-46224-0_5

- Dosovitskiy, A. and Brox, T. (2016): “Inverting Visual Representations with Convolutional Networks.” Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 4829–4837.
- Girshick, R., Donahue, J., Darrell, T. and Malik, J.: “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.” Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 580–587. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014.
- Harrison-Moore, A. (2007): “Aristocratic Identity: Regency Furniture and the Egyptian Revival Style”. Sofaer, J. (ed.): Material Identities. Malden, Oxford, Calden: Blackwell Publ. 2007: 101–126.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” In Pereira, F., Burges, C. J. C., Bottou, F. and Weinberger, K. Q. (eds.): Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012: 1097–1105 <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- Lowe, D. G. (1999): “Object Recognition from Local Scale-Invariant Features.” Proceedings of the Seventh IEEE International Conference on Computer Vision 2 (8): 1150–1157.
- Mahendran, A. and Vedaldi, A. (2015): “Understanding Deep Image Representations by Inverting Them.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: 5188–5196
- The Metropolitan Museum of Art (2017). The Met Makes Its Images of Public-Domain Artworks Freely Available through New Open Access Policy. Available from: <http://www.metmuseum.org/press/news/2017/open-access> [Accessed: 28 April 2017]
- The Neoclassica Project (2017) Neoclassica – A Framework for Research in Neoclassicism. Available from: <http://neoclassica.network> [Accessed: 3 April 2017].
- Zagoruyko, S., Lerer, A., Lin, T.-Y., Pinheiro, P. O., Gross, S., Chintala, S. and Dollár, P. (2016): “A MultiPath Network for Object Detection.” BMVC. <http://arxiv.org/abs/1604.02135>.
- Zeiler, M. D. and Fergus, R. (2013): “Visualizing and Understanding Convolutional Networks.” <http://arxiv.org/abs/1311.2901>.