
Script Identification in Medieval Latin Manuscripts Using Convolutional Neural Networks

Mike Kestemont

mike.kestemont@gmail.com

University of Antwerp, Belgium

Dominique Stutzmann

dominique.stutzmann@irht.cnrs.fr

Centre National de la Recherche Scientifique, France

Introduction

In paleography, scholars study the history of handwriting, a crucial aspect of book history and manuscript studies. Paleography has traditionally been dominated by expert-based approaches, driven by the opinions of a small group of highly trained individuals. These have acquired a set of expert skills through year-long training, e.g. the ability to date a handwriting or attribute it to specific individuals. This knowledge remains very hard to render explicit, in order to share it with others. Therefore, paleographers are increasingly interested in digital modelling techniques to enhance the creation and dissemination of paleographic knowledge (Stutzmann, 2015). An important task in paleography is the classification of script types, especially now that digital libraries (BVMM, Gallica, e-Codices, Manuscripta Mediaevalia, etc.) are amassing reproductions of medieval manuscripts, often with scarce metadata. Being able to recognize the script type of such historic artefacts is crucial to date, localize or (semi-)automatically transcribe them. This paper focuses on script identification for medieval Latin manuscripts (ca. 500 AD to 1600 AD) and demonstrates the feasibility of a fairly accurate, meaningful automated classification.

Medieval script classification was the focus of the recent CLaMM (Classification of Latin Medieval Manuscripts) competition. For this shared task, the organizers released a training data set of 2,000 photographic (greyscale, 300 dpi) reproductions of pages extracted from Latin manuscripts, which were classified into a

12 script type classes, including uncial, caroline, textualis and humanistic script, but also more difficult to delineate classes such as the cursiva or (semi)hybrida. The participating teams had to submit a standalone application which was able to classify unseen images and estimate the distance between them. The organizers would then apply the submissions to 1,000 resp. 2,000 test images (Stutzmann, 2016) and evaluate their performance using various evaluation schemes. Here, we discuss the *DeepScript* submission to the CLaMM competition. The competition's results have been officially been released on 26 Oct. 2016. *DeepScript* was ranked first on task 2, i.e. the 'crisp' classification of mixed script images (Cloppet et al., 2016). As the ground truth and results were released too recently, we limit this abstract to a general discussion of the approach; the final version and presentation of this paper will be supplemented with additional information and test results.

The *DeepScript* submission builds upon recent advances in Computer Vision, where the use of so-called 'deep' neural networks has recently led to dramatic breakthroughs in the state of the art of image classification (LeCun et al., 2015). The kind of neural networks applied in Computer Vision are typically convolutional: they slide small 'filters' (feature detectors) across images to make the network robust to small translations of objects. The networks make use of many 'layers' of such feature detectors, where the output of one feature detector always feeds into the next one. The use of such a stack of layers is beneficial, because this 'deep architecture' allows algorithms to model features of an increasing complexity (Bengio et al., 2013): in the first layers of the network, very raw and primitive shapes are detected ('edges'); it is only at the higher layers in the networks that these primitive features are combined into more complex, abstract visual patterns (e.g. entire faces). These neural networks lie at the basis of e.g. modern face verification algorithms on social media websites such as *Facebook*.

Neural networks are composed of millions of parameters which have to be optimized. For this, the available training data is split out in a set of training images and a smaller set of development images (respectively ca. 1,800 and 200 images): the former is used to optimize the parameters of the network during training, the latter is used to monitor the performance of the network. The use of development data is necessary to avoid 'overfitting': it is possible for a network to start 'memorizing' the training images, so that it produces perfect predictions for the training data, but is not able

any more to generalize properly to new, unseen images. By using a development set, we can stop optimizing the network, if its predictions for the development data do not increase in quality anymore. Only at this stage, the algorithm is evaluated on the actual test images.

Modern neural networks are typically trained on hundred thousands of training images. In the field of Cultural Heritage data, a common challenge is that most data sets are much smaller, and CLaMM is no exception, so that the danger of overfitting is much larger. We therefore proceeded as follows: the generous resolution for each training image was downsized by one half. Next, we would select random square crops or patches from the image (150x150 pixels) and train the algorithms on batches of these crops. This approach is blunt, yet innovative, since we make no effort to extract more specific regions of interest from the images, such as individual lines, words or characters. To avoid overfitting, we also applied augmentation: each training crop would be ‘distorted’ through randomly varying the zoom level, rotation and translation. Introducing such noise in the input is a common strategy to combat overfitting. Below goes an example of such a set of augmented patches for a single manuscript page (Fig. 1).



Fig. 1: Example of augmented crops for a single manuscript page.

After each epoch, we evaluated the performance of the current state of the network through inspecting the classification accuracy on the development images: we randomly selected 30 crops from each image (without augmentation), and calculated the average probability for each output class. The full image was assigned to the class with the highest average probability. The best validation accuracy which we achieved was 91.17%, using a network architecture of 14 layers,

inspired by the famous Oxford VGG net (Simonyan et al., 2015). The manual classification of CLaMM images was based on morphological differences and allo-graphs, as defined in standard works on Latin scripts such as (Bischoff, 1986; Derolez, 2003). The confusion matrices in Fig. 2-4 for the actual and the predicted classes in the development and test data illustrate that the confusions generally make sense from a paleo-graphic point of view (the normal textualis letter is for instance often mistaken for the Southern or semi-textualis variant).

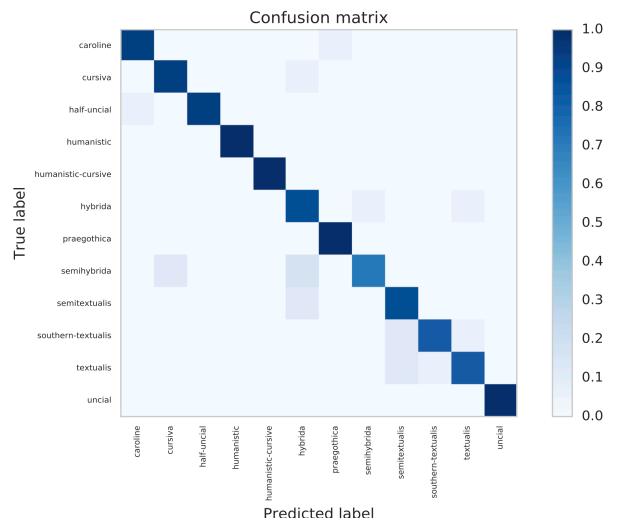


Fig. 2: Confusion matrix for the development data.

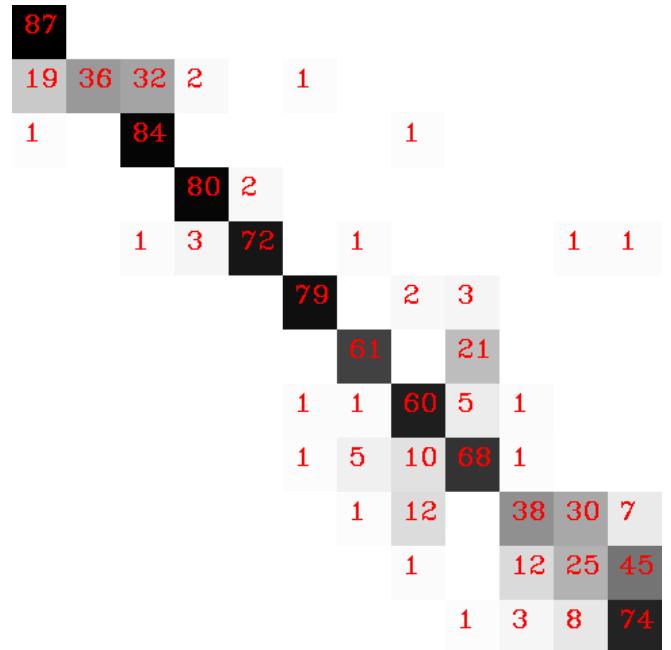


Fig. 3: Classifications for the test data as a confusion matrix (task 1). Horizontal lines: ground-truth; Vertical columns: predictions. Order: 1-Uncial; 2-Half-uncial; 3-Caroline; 4-Humanistic; 5-Humanistic; Cursive; 6-Praegothica; 7-Southern Textualis; 8-Semitemporealis; 9-Textualis; 10-Hybrida; 11-Semihybrida; 12-Cursiva.

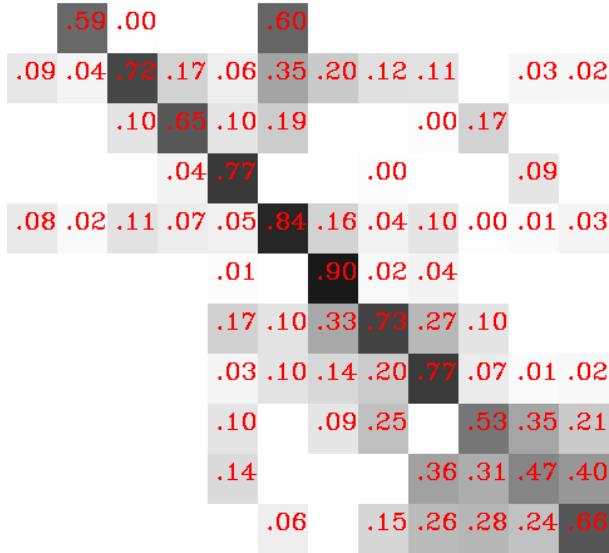


Fig. 4: Membership Degree matrices for task 2, on the 999 test images where only one label is attributed to the image.

There exist interesting methods to visualize which patterns the trained network is sensitive to. Using the principle of gradient ascent, we start from a random noise image and feed it to one of the filters on the last convolutional layer: during 3,000 iterations we change the image so that it maximizes the activation of this particular filter. In Fig. 3, we show the 25 artificially generated images which yielded the strongest results; clearly, the network picks up relevant patterns. The third image from the left in the first line, for instance, clearly captures the presence of loops in the ascenders of individual characters (e.g. in the 'b' or 'h', which is crucial to differentiate between e.g. a textualis and a cursive letter). These visualizations directly tackle the issue of the computational 'black box' in the Digital Humanities, and especially Digital Palaeography (Hassner et al., 2013; Stutzmann et al., 2014). In our paper, we will offer further interpretations and visualizations of our model and confront these with the results from other participants in the CLaMM competition to offer new perspectives on the graphic definition of script classes in traditional paleography.

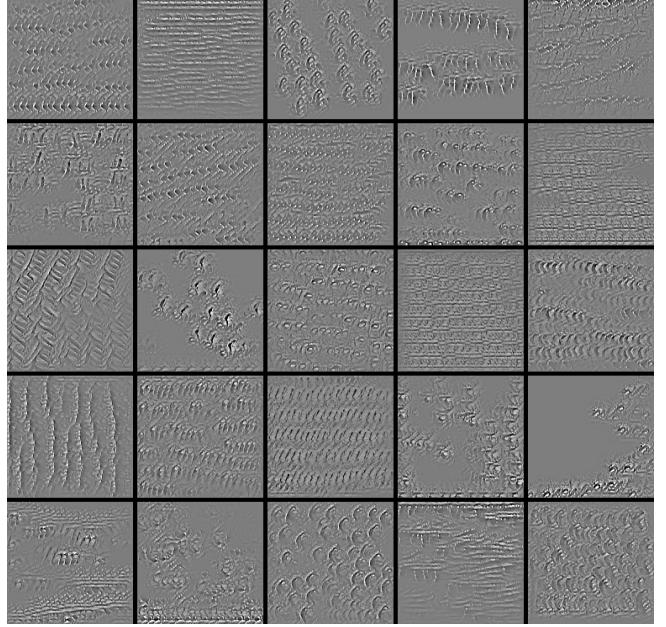


Fig. 5: Artificially generated images that maximally activate filters in the final convolutional layer.

Bibliography

- Bengio, Y., Courville, A. and Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *TPAMI* 35:8, 1798–1828.
- Bischoff, B., *Paläographie des römischen Altertums und des abendländischen Mittelalters*, Berlin, 1986.
- Cloppet, F., Eglin, V., Kieu, V. C., Stutzmann, D. and Vincent, N. (2016), ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script, *Proceedings of the ICFHR 2016*, 590-595.
- Derolez, A., *The Palaeography of Gothic Manuscript Books from the Twelfth to the Early Sixteenth Century*, Cambridge, 2003.
- Hassner, T., Rehbein, M., Stokes, P. and Wolf, L. 2013. Computation and palaeography: Potentials and limits. *Dagstuhl Manifestos* 2: 14–35.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature* 521(7553): 436–44.
- Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Proceedings of ICLR 2015*. <https://arxiv.org/abs/1409.1556/>.
- Stutzmann, D. and Tarte, S. (2014). Digital palaeogra-

phy: New machines and old texts: Executive summary. *Dagstuhl Reports* 4.7: 112–34 (112–14, 132).

Stutzmann, D. (2015). Clustering of medieval scripts through computer image analysis: Towards an evaluation protocol. *Digital Medievalist* 10, <http://digitalmedievalist.org/journal/10/stutzmann/>.

Stutzmann, D. (2016). *ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script*. <http://icfhr2016-clamm.irht.cnrs.fr/>.