From Modern to Medieval: Detecting and Visualizing Entities in Manuscripts of Marco Polo's Devisement du Monde

XML

1. INTRODUCTION

Digital humanities have had a strong focus on algorithmic reading of textual data, which has been brought together with visualization for comparing and analyzing results (Jänicke et al., 2017). In recent years, this textual focus has grown to include other modalities such as audio-visual data (Arnold and Tiltion, 2019; Wevers and Smits, 2020). In turn, methods in computer vision (Arnold and Tiltion, 2019) have been proposed for the specificities of audio-visual corpora. As a starting point for distant viewing of medieval illumination, we applied computer vision methods to a dataset of images from manuscripts of the French Marco Polo textual tradition, images that demonstrate a strong visual coherency. Extant in some 15 manuscripts, the Devisement du monde is famous for descriptions of extra-European travel and the depiction of the exotic wonders of Asian cities (Cruse, 2019). We set out to see if repeated visual features across this image corpus are detectable using object detection, what visualization would allow us both to understand better Polo's depiction of the exotic and how modern image hierarchies might be adapted to the specificities of medieval manuscripts.

For image classification and object detection of images, large datasets with class hierarchies exist like ImageNet (Deng et al., 2009) and Open Images (Kuznetsova et al., 2020). These datasets and their underlying hierarchies are neither particularly effective at identifying the wide variety of entities depicted in medieval manuscripts, nor do they detect entities well given the representational density of medieval illumination. Furthermore, there is insufficient data for training these neural networks. In our work, we argue that networks trained on natural image datasets can provide both a first impression (Crowley and Zisserman, 2014), and a convenient starting point for building new classes and hierarchies and they can be even used to extract some initial training samples from small- to medium-sized image corpora.

We applied computer vision methods on a dataset of some 700 medieval illuminations from seven manuscripts and built a visual interface to explore and annotate the results. We were interested in the possibility of editing the classes of contemporary hierarchies, replacing them with categories more appropriate for the period and the corpus.

2. DATA & IMAGE PROCESSING

Each image shows a page with a visual scene depicting different aspects of Polo's description. We applied object detection by using the Faster R-CNN (Ren et al., 2015) trained on Open Images. The label hierarchy of Open Images consists of 600 different classes including parent and child relations. The object detection extracts 100 bounding boxes for each image with a confidence score and a label for the detected entity. The result was 71,400 bounding boxes. Furthermore, we extracted image embeddings for each bounding box detected with an EfficientNet (Tan and Le, 2019) trained on ImageNet. For the image embeddings, we use faiss (Johnson et al., 2019) to query the most similar bounding boxes for each example based on the Euclidean distance between the embeddings. This allows us to see the most similar parts of another image to an image of interest.

3. VISUAL INTERFACE

The design of the visual interface facilitates exploration of the image dataset and comparison with different depictions of specific entities. For this, the object classes can be accessed through a Tag Cloud where frequency is encoded by font size, or through a tree that visualizes the Open Images hierarchy with all classes found in the Marco Polo dataset. Such interfaces for visual exploration and annotation allow the professional viewer/reader to focus on a given interest to annotate new areas or investigate the objects found inside the image, delete them or even edit their labels (Siemens et al., 2009). To prevent visual clutter, they can filter by confidence value and select or deselect specific classes. When focusing on one specific bounding box, it is also possible to display the bounding boxes that intersect, that are inside or outside the box of interest. Figure 1 shows a page of the dataset with the entities found by the neural network. For a given object class, all depictions are displayed in a 2D grid ordered by the confidence score assigned by the neural network. Examples can be seen in Figures 2 and 3. The interface is designed for both discovery and revision by clicking on a bounding box of interest, which leads us to see the most similar bounding boxes. It is also possible to select multiple bounding boxes and delete or re-label them in case of an imprecise classification. Furthermore, the expert viewer is able to annotate areas in the image with new classes, thereby contributing to a new category in the Tag Cloud (Annotated) and transforming it into a TagPie (Jänicke et al., 2018), seen in Figure 4.

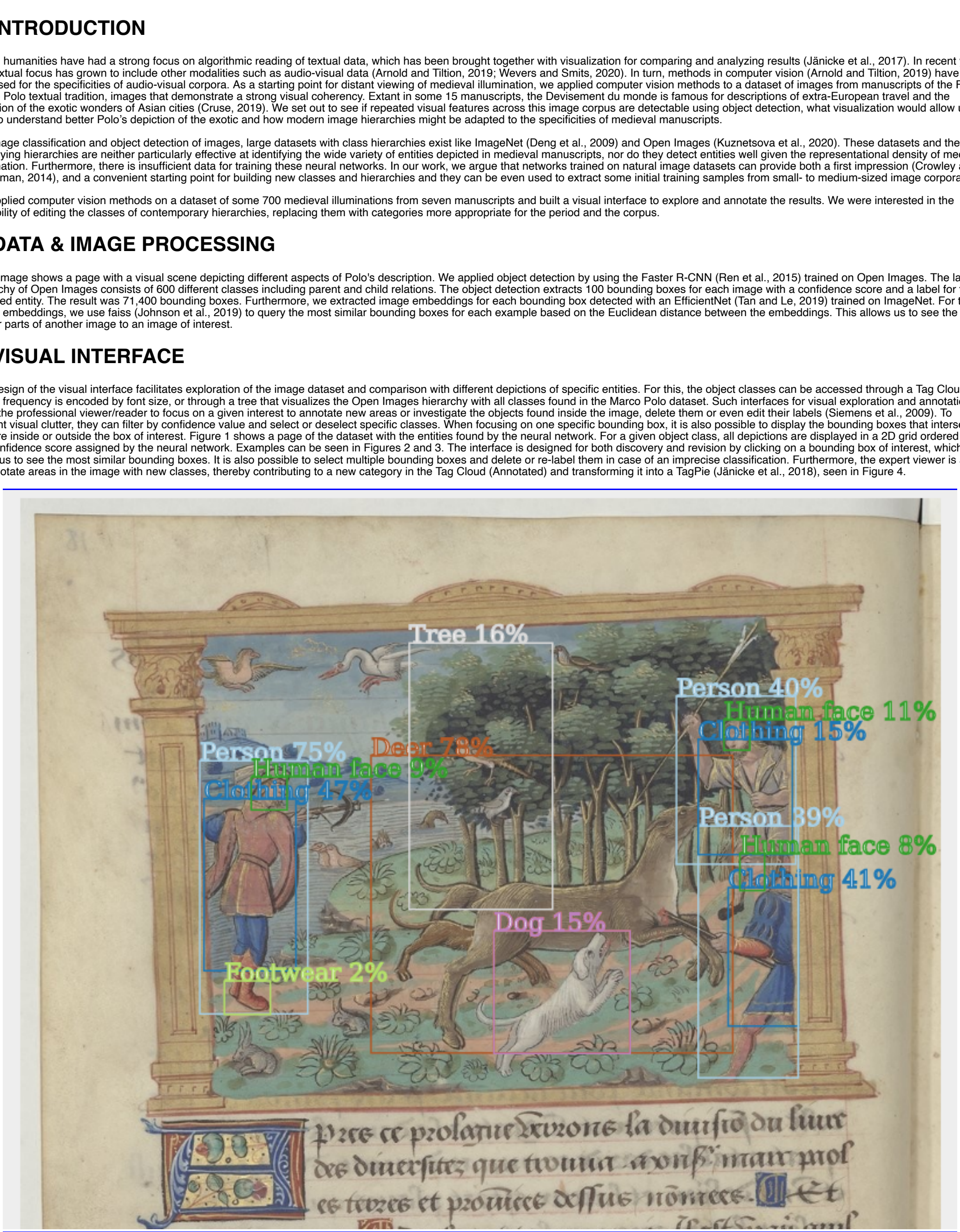

Figure 1: A page of the dataset with entities found by a neural network in an illumination from BnF Arsenal ms 5219

Figure 2: An overview of samples of faces marked by the highest confidence scores

Figure 3: An overview of samples of human figures found in the corpus with the highest confidence scores

Figure 4: The TagPie gives an overview of the classes found by the neural network (green) and those from human annotations (purple)

4. DISCUSSION

Whereas some anachronistic categories were persistent throughout the output of the initial system, other objects such as those mentioned in Figures 1 - 3 led to rather convincing recognition. Furthermore, the summary views of the visual interface proved particularly effective at demonstrating the tensions found between the codified visual languages of medieval French manuscripts and the diachronic innovative attempts at representing the "unprecedented images of the world beyond Europe's borders" as well as domains in which patterns in those tensions were particularly pronounced (Keene, 2019: 196).

On the other hand, the interface created to explore, revise and manipulate features in the Marco Polo visual corpus provides us with a stepping stone for working with larger visual corpora built from across the global middle ages. As our inquiry evolves, finding ways to guide the viewer from the extracted objects and their computed confidence levels back to full images and relevant metadata will be crucial for allowing for sufficient contextualization to facilitate interpretation. Furthermore, our current method for revision and addition of labels is open-ended, but in future work, we intend to lead the annotation toward established art historical vocabularies to ensure future discoverability. Future work will also focus on ways to achieve the "best of both worlds," allowing research to move from the modern to the medieval, that is, for current day hierarchies to be adjusted and augmented by domain- and period-specific terminology with the support of expert knowledge.

Creating this visual pathway for visual exploration and hypothesis generation using computer vision techniques is not a trivial task, since the metadata of legacy databases of manuscript illumination (Mandragore, Initiales, Digital Scriptorium, etc.) also vary in both size and granularity. Furthermore, there is a need for methodologies to combine or unify the vocabulary of different datasets, bridge the gap between general and domain-specific vocabularies, as well as to create expert hierarchies of entities found in manuscript illumination in order to create appropriate training datasets to deal with issues of cross-depiction (Hall et al., 2015).

Appendix A

Bibliography

1. **Jänicke, S. Franzini, G. Cheema, M.F. and Scheuermann, G .** (2017). TagPies: Visual text analysis in digital humanities. Computer Graphics Forum 36, 6 2017, 226--250.
2. **Arnold, T. and Tilton, L.** (2019). Distant viewing: analyzing large visual corpora. Digital Scholarship in the Humanities 34, Supplement_1, i3–i16.
3. **Wevers, M. and Smits, T .** (2020). The visual digital turn: Using neural networks to study historical images. Digital Scholarship in the Humanities 35, 1 (2020), 194–207
4. Cruse, M . (2019) . Novelty and Diversity in Illustrations of Marco Polo's Description of the World.
5. Keene, B.C . (2019) . Toward a Global Middle Ages: Encountering the World Through Illuminated Manuscripts. Los Angeles : J. Paul Getty Museum.
6. **Deng, J. Dong, W. Socher, R. Li, L.J. Li, K. and Fei-Fei, L .** (2009). ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09.
7. **Kuznetsova, A. Rom, H. Alldrin, N. Uijlings, J. Krasin, I. Pont-Tuset, J. Kamali, S. Popov, S. Malloci, M. Kolesnikov, A. Duerig, T. and Ferrari, V.** (2020). The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. International Journal of Computer Vision 128, 7 (2020), 1956—1981.
8. **Crowley, E.J. and Zisserman, A .** (2014). In search of art. In European Conference on Computer Vision. Springer, 54–70.
9. **Ren, S. He, K. Girshick, R. and Sun, J .** (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems. 91–99.
10. **Tan, M. and Le, Q .** (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning 6105--6114
11. **J ohnson, J. Douze, M. and Jégou, H .** (2019). Billion-scale similarity search with gpus. IEEE Transactions on Big Data
12. **Siemens, R. Leitch, C. Blake, A. Armstrong, K. and Willinsky, J.** (2009). “It May Change My Understanding of the Field”: Understanding Reading Tools for Scholars and Professional Readers. Digital Humanities Quarterly, 3, 4.
13. **Jänicke, S. Blumenstein, J. Rücker, M. Zeckzer, D. and Scheuermann, G .** (2018). TagPies: Comparative Visualization of Textual Data. VISIGRAPP (3: IVAPP) 40--51.
14. **Hall, P. Cai, H. Wu, Q. and Corradi, T .** (2015). Cross-depiction problem: Recognition and synthesis of photographs and artwork. Computational Visual Media 1, 2, 91–103.

Christofer Meinecke (cmeinecke_at_informatik.uni-leipzig.de), Leipzig University, Germany und David Joseph Wrisley (djw12_at_nyu.edu), New York University Abu Dhabi und Stefan Jänicke (stjaenicke_at_imada.sdu.dk), University of Southern Denmark