

Reflecting on How Artworks Are Processed and Analyzed by Computer Vision

Sabine Lang^[0000–0003–2543–0085] and Björn Ommer

Heidelberg Collaboratory for Image Processing,
IWR, Heidelberg University,
`firstname.lastname@iwr.uni-heidelberg.de`

Abstract. The intersection between computer vision and art history has resulted in new ways of seeing, engaging and analyzing digital images. Innovative methods and tools have assisted with the evaluation of large datasets, performing tasks such as classification, object detection, image description and style transfer or assisting with a form and content analysis. At this point, in order to progress, past works and established practices must be revisited and evaluated on the ground of their usability for art history. This paper provides a reflection from an art historical perspective to point to erroneous assumptions and where improvements are still needed.

Keywords: computer vision, art history, critical reflection, distant viewing, close reading, object detection, image description, style transfer

1 Introduction

For some time, computer vision and art history are in close collaboration: scholars from both fields work together to find innovative ways to process large digital image sets. These new approaches are beneficial for research, because they offer new modes of how digital images can be seen or analyzed. Computational technologies enable a large scale evaluation and a close-up study, including classification, object retrieval, or a form and content analysis. For computer vision a collaboration is beneficial, since existing algorithms are tested and modified due to new requirements imposed by artistic data. In parallel, art history is compelled to question established methods and terms: how do we describe images and what do we mean by 'style'? At this point, in order to progress, we must revisit past works: how are images produced, processed and understood? Which problematic assumptions have been held? The objective of this paper is to provide a critical reflection, point to problems and research gaps. The paper especially focuses on aspects of distant viewing versus close reading, object detection, image description and style transfer.

2 Image Analysis in Computer Vision and the Arts

Digital art history, which refers to the "use of analytical techniques enabled by computational technology" [6], is the result of the meeting between computer

vision and art history. The presence of large art datasets eventually required efficient computational methods and tools to process and evaluate them. Works included diverse tasks, such as classification, object detection, image description or style transfer. Karayev et al. [15] classified artworks according to style; [27] used a deep convolutional model to categorize images according to genre, style and artist. Other works performed object detection in paintings: classifiers were trained on natural images [4] and paintings or on both to measure the domain shift problem [5]. Karpathy et al. [16] addressed the task of an automatic image description; [21] simultaneously annotated, classified and segmented objects in natural images. Recently, scholars focused on transferring artistic styles to natural images, utilizing deep neural networks [10] or generative adversarial networks [32] – most relied on a single input image. In art history, scholars have been concerned with similar topics for a long time: Aby Warburg (1866-1929) used reproductions of artworks to map ‘the afterlife of antiquity’ [29], resembling current distant viewing efforts [13]. Art historians also discussed topics of image analysis or style: contributions have been made by Alois Riegl (1858-1905) [25] or Heinrich Wölfflin (1864-1945), who used a comparative method to study artworks and formulated his five principles of art history [31]. With his iconographical-iconological method, Erwin Panofsky (1892-1968) established a framework for image understanding and description [12]. Digital humanities scholars have (critically) reflected on the impact of technologies on these traditional practices [17]; [1], for example, pointed to the loose usage of terms and uninterrogative nature of many works. On the basis of current works in computer vision, the paper engages in a critical discussion.

3 Reflecting on a Computer-Based Image Analysis

Distant Viewing AND Close Reading. Art historians aim to understand works of art: why did artists depict a certain subject matter or use a specific color? To find answers, they study images in detail and within a wider context. In the past, scholars in digital art history have commented on the fact that computer-based works focused on a quantitative analysis of data, thus only identifying patterns without providing an interpretation [1], [17]. While more recently, a qualitative analysis has been added, scholars either facilitate a distant viewing approach or answer more pointed questions on the basis of individual artworks. Works, such as the analysis of strokes in a limited collection of drawings by Picasso, Matisse, Egon Schiele and Modigliani to identify forgeries [8] or an ‘Automatic Thread-Level Canvas Analysis’ to conclude whether or not two paintings were made from the same canvas [22], evaluate artworks in detail and show impressive results, but are not applicable on a large scale, because they require specific, costly data, and lack contextualization. Similar, projects utilizing distant viewing mostly remain pure visualizations [23], produce little new knowledge and rarely add a top-down approach to explain origins of patterns [7]. However for art history, an either-or-stance is insufficient [3]; in order to be relevant, an analysis must be quantitative and qualitative.

Finding Objects in Paintings. Object detection in paintings has been based on a quantitative analysis [4], [5], where retrieval systems are mostly conditioned on ImageNet. The visual database contains over fourteen millions of well-aligned natural images gathered alongside pre-defined contemporary categories. While systems confidently detect objects, such as dogs, persons or other modern categories, in naturalistic images, they fail, when confronted with objects belonging to pre-modern times. Failure cases occur for medieval objects or clothing and pre-modern architecture, because systems are simply unfamiliar with these categories. Algorithms are further challenged by less standardized and complex compositions, which are manifold in art. Further complications arise, when the content of an artwork is distorted due to perspective or abstraction. In its current state, many models for object detection are not feasible for art history; to train models directly on art data would be one solution to overcome some limitations [30].

Describing Artworks. Art history and computer vision are both concerned with image description and work has been done to automatize this step [16], [14]. While results on natural images may be convincing, the question remains, if the variety of subject matters, objects or styles in art can be correctly grasped by models and descriptions resemble those of art historians. A full image description of Gustave Courbet’s *La Rencontre* (1854), using Panofsky’s iconographical-iconological method, can be found in the supplementary material and establishes requirements of an art historical description: the method includes a pre-iconographical description, which identifies the manner in which objects are expressed, an iconographical analysis of symbols and motifs, and a placement within a wider historic and biographical context – the iconological analysis [12]. A model for an automatic image description must be able to perform a formal and semantic analysis of the artwork, preferably considering foreground, middle-, and background, understand its composition and relations between objects. Also, it must recognize symbols and cultural conventions and place the image in a wider context. What is possible so far? Works [16] have proven that models can generate descriptions of regions, thereby providing formal descriptors of, for example, color or material and identifying objects correctly. Thus, approaches mostly provide a formal description, but are unable to produce an iconographical or iconological analysis. Although the linkage to other historical digital sources might give further information about artworks, networks do not possess knowledge about symbols and pictorial or cultural conventions. A closer evaluation of works from an art historical perspective reveals further issues: most examples lack to provide an account of the image’s composition or relations between objects; also computer vision mainly performs a single image description and misses a comparison or broader contextualization. However, some works have addressed these issues and studied how objects in images are related by utilizing relative attributes, thereby capturing semantic relationships [24], and identified salient regions [19]. Also, instead of images with simple compositions, more challenging datasets [18] were used, where the complexity is representative of those in artworks. While these approaches are first steps, they are still not suf-

ficient. Automatic models might create a descriptive list of image components, however, it remains the task of art historians to create the story: to interpret artworks and position them within a wider context.

Style Refers to Formal Qualities. Style transfer is a current task in computer vision, where a natural image is being rendered in, for example, the style of Picasso or van Gogh [10], [32]. For art history, these works are relevant, because they lead to a reassessment of the term style; however, works are based on some problematic assumptions. The often used expression ‘in the style of...’ implies that an artist is bound to a single style. However, if we look at Picasso, we find works in many different styles: in an academic, Cubist or Surrealist manner. In the context of style transfer, style mainly refers to color, shape or brush stroke; other formal features, such as composition or modeling of figures, and content are neglected. This is again highlighted, when we look at the referenced styles: most common are Impressionism, Post-Impressionism, Expressionism, Cubism or Abstract Art; less visually distinct and content-based styles, such as Gothic Art, Renaissance, Baroque or Surrealism, are absent. Results then illustrate that style transfer works best with heavy visual styles and when naturalistic images display structure – on planar regions the network produces random artifacts. In computer vision artistic style is assumed to be static, but not – as it is its nature dynamic and evolving. A last point refers to the fact that style transfer is mostly based on one image [10], [11]. However, a single artwork might not display all aspects of a style; a portrait in an Impressionistic style accentuates different style constituents, which a landscape painting in the same style does not. Just as one has to look at the whole image to make a style judgment, because shape or light contrasts vary in different regions, it is necessary to utilize a collection of images in the same style. The work by [26] shows that using multiple instead of single images produces stylistically more convincing results.

4 Conclusion

The paper reflected on the topics of distant viewing versus close reading, object detection, an automatic image description and style transfer. It aimed to highlight problematic assumptions and where work has yet to be done. Computer vision has provided powerful tools to analyze artworks quantitatively and qualitatively, thereby creating verification and new knowledge for art history. In turn, the discipline contributes from how art history approaches, describes and interprets images. An evaluation of previous work is valuable in that it forces both disciplines to reflect on existing terms and practices. Eventually, there is great potential, when scholars from both fields work together, and there are still topics, which require our attention: the study of sculptures [9] or architecture, preservation and documentation of cultural heritage through digital reconstruction and 3D modeling [2], [28], detection of forgeries [20] or provenance research being some examples.

References

1. Bishop, C.: Against digital art history. *International Journal for Digital Art History* (3)
2. Boeykens, S., Maekelberg, S., De Jonge, K.: (re-) creating the past: 10 years of digital historical reconstructions using bim. *International Journal for Digital Art History* (3)
3. Bonfiglioli, R., Nanni, F.: From close to distant and back: How to read with the help of machines. In: *International Conference on History and Philosophy of Computing*. pp. 87–100. Springer (2015)
4. Crowley, E.J., Zisserman, A.: In search of art. In: *Workshop at the European Conference on Computer Vision*. pp. 54–70. Springer (2014)
5. Crowley, E.J., Zisserman, A.: The art of detection. In: *European Conference on Computer Vision*. pp. 721–737. Springer (2016)
6. Drucker, J.: Is there a “digital” art history? *Visual Resources* **29**(1-2), 5–13 (2013)
7. Drucker, J., Helmreich, A., Lincoln, M., Rose, F.: Digital art history: la scène américaine. une discussion entre johanna drucker, anne helmreich et matthew lincoln, introduite et modérée par francesca rose. *Perspective. Actualité en histoire de l’art* (2), 27–42 (2015)
8. Elgammal, A., Kang, Y., Leeuw, M.D.: Picasso, matisse, or a fake? automated analysis of drawings at the stroke level for attribution and authentication. *arXiv preprint arXiv:1711.03536* (2017)
9. Fouhey, D.F., Gupta, A., Zisserman, A.: 3d shape attributes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1516–1524 (2016)
10. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015)
11. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017)
12. Hatt, M., Klonk, C.: *Art history: a critical introduction to its methods*. Manchester University Press (2006)
13. Hristova, S.: Images as data: Cultural analytics and aby warburg’s mnemosyne. *International Journal for Digital Art History* (2)
14. Johnson, J., Karpathy, A., Fei-Fei, L.: Densecap: Fully convolutional localization networks for dense captioning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4565–4574 (2016)
15. Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H.: Recognizing image style. *arXiv preprint arXiv:1311.3715* (2013)
16. Karpathy, A., Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3128–3137 (2015)
17. Kienle, M.: Digital art history “beyond the digitized slide library”: An interview with johanna drucker and miriam posner. *Artl@s Bulletin* **6**(3), 9 (2017)
18. Kinghorn, P., Zhang, L., Shao, L.: A region-based image caption generator with refined descriptions. *Neurocomputing* **272**, 416–424 (2018)
19. Krause, J., Johnson, J., Krishna, R., Fei-Fei, L.: A hierarchical approach for generating descriptive image paragraphs. In: *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. pp. 3337–3345. IEEE (2017)

20. Li, J., Yao, L., Hendriks, E., Wang, J.Z.: Rhythmic brushstrokes distinguish van gogh from his contemporaries: findings via automated brushstroke extraction. *IEEE transactions on pattern analysis and machine intelligence* **34**(6), 1159–1176 (2012)
21. Li, L.J., Socher, R., Fei-Fei, L.: Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. pp. 2036–2043. IEEE (2009)
22. van der Maaten, L., Erdmann, R.G.: Automatic thread-level canvas analysis: A machine-learning approach to analyzing the canvas of paintings. *IEEE Signal Processing Magazine* **32**(4), 38–45 (2015)
23. Manovich, L.: How to compare one million images? In: *Understanding digital humanities*, pp. 249–278. Springer (2012)
24. Parikh, D., Grauman, K.: Relative attributes. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. pp. 503–510. IEEE (2011)
25. Riegl, A., Castriota, D., Zerner, H.: *Problems of style: Foundations for a history of ornament*. Princeton University Press Princeton, NJ (1992)
26. Sanakoyeu, A., Kotovenko, D., Lang, S., Ommer, B.: A style-aware content loss for real-time hd style transfer. *arXiv preprint arXiv:1807.10201* (2018)
27. Tan, W.R., Chan, C.S., Aguirre, H.E., Tanaka, K.: Ceci n’est pas une pipe: A deep convolutional network for fine-art paintings classification. In: *Image Processing (ICIP), 2016 IEEE International Conference on*. pp. 3703–3707. IEEE (2016)
28. Underhill, J.: In conversation with cyark: Digital heritage in the 21st century. *International Journal for Digital Art History* (3)
29. Warburg, A.: *Der Bilderatlas Mnemosyne*, vol. 2. Akademie Verlag (2008)
30. Wilber, M.J., Fang, C., Jin, H., Hertzmann, A., Collomosse, J., Belongie, S.: Bam! the behance artistic media dataset for recognition beyond photography. In: *Proc. ICCV*. vol. 1, p. 4 (2017)
31. Wölfflin, H.: *Principles of art history*. Courier Corporation (2012)
32. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593* (2017)