Reflections on connoisseurship and computer vision

Peter Bell and Fabian Offert

This special issue shows the diversity of approaches to connoisseurship throughout history. One recent area of research where questions of connoisseurship have become particularly relevant is digital art history, specifically where it intersects with computer vision and machine learning. Here, connoisseurship is not re-invented but modelled in a way that seems to stay close to the human connoisseur: as learning from examples. While clearly there is no guarantee that a computer will develop strategies of attribution akin to those of the human connoisseur, both tasks and methods seem to stay essentially the same if connoisseurship is operationalized as machine learning. On the following pages we will demonstrate how this similarity, but also the significant differences between human and machine approaches can be understood as productive interventions in the discourse around connoisseurship. Central to this investigation is the question: How do we teach connoisseurship to a new kind of observer — the computer — and what challenges result from this process?

A connoisseur always contextualises an observed picture with a large number of other pictures, for instance preparatory drawings and later works.[1] He or she always understands the picture as one node in a network of similar pictures, not necessarily by the same artist. He or she also does not only observe the picture as a whole. Or, put differently, the whole is only one field of view that is addressed. The impression of the whole ('Totaleindruck'), as Giovanni Morelli warns, cannot be the end of the analysis.[2] Instead, different elements and aspects of a work need to be considered. One example is the signature, a marker of authenticity par excellence. Others could be the individual characteristic style inherent in the brushstrokes or the hatching, the treatment of objects and figures and, in particular, anatomical details like ears or fingernails.[3] These details are often considered expressions of the subliminal repertoire of form at an artist's disposal, and were particularly relevant to Giovanni Morelli in the context of attribution. After all, connoisseurship is not merely the ability to see 'properly'. Historical knowledge, for example of places,

[1] Morelli emphasises the study of drawings in particular, see Giovanni Morelli, Jean Paul Richter, *Italienische Malerei der Renaissance im Briefwechsel von Giovanni Morelli, Jean Paul Richter, 1876-1891*, edited by Irma Richter, Baden-Baden, 1960, 54.
[2] Giovanni Morelli, 'Die Galerien Borghese und Doria Panfili in Rom', *Kunstkritische Studien über italienische Malerei*, 1, 1890, 23, 26f.
[3] See Bode's description of Ludwig Scheibler's approach to picture analysis: Wilhelm von Bode, *Mein Leben 1845-1929*, Berlin, 1930, 9.

lives or biographies, travel routes, provenances, and of documents in general, is just as much a part of connoisseurship as are art-technological investigations of colour and medium. Wide contextual knowledge is required to attribute a picture not only to specific hands but specific minds.[4]

The inherent limitations of this approach have often been the subject of journalistic debates, and connoisseurship has been repeatedly associated with human 'weaknesses' such as subjectivity, vanity, and ignorance.[5] Moreover, since the 1980s, the computer in art history has been generally associated with a new kind of formalism, with a fallback to the time before Schlosser and Riegl.[6] These accusations aside, it is an undeniable fact that the human pictorial memory is limited and is usually not a photographic memory. Rather, humans memorise what they see in an idiosyncratic way. Based on these limitations, it seems plausible to 'delegate' questions of attribution to a machine — not to replace human connoisseurship but to separate it somewhat from intuition.[7] Delegating attribution to a computer would mean to create an artificial observer whose view would not override but complement the human perspective. Such an artificial observer would not be 'objective' or 'neutral' by any means — after all it would need to be conditioned on human-selected data — but it would provide a deliberately 'alien' point of view.

Generally, this opportunity only presents itself because connoisseurship can be trained.[8] This is impressively demonstrated by Watanabe's experiment on pigeons, which were conditioned to distinguish between Impressionists and Cubists.[9] It is self-evident that pigeons are equipped with a different visual apparatus than humans. Hence, their attribution decisions, although accurate, are made on the basis of entirely different visual and cognitive processes. Here lies the opportunity of the computer as an alternative beholder: in the creation of modes of perception that are entirely different from their human complements.

The computer also allows us to work under laboratory conditions that would be impossible with human observers: it is possible to determine exactly

[4] See Felix Thürlemann, 'Händescheidung ohne Köpfe? Dreizehn Thesen zur Praxis der Kennerschaft am Beispiel der Meister von Flémalle/Rogier van der Weyden-Debatte', *Zeitschrift für Schweizerische Archäologie und Kunstgeschichte*, 62, 2005, 225–232.
[5] See Frank Zöllner, 'Salvator Mundi: Der teuerste Flop der Welt?', *Die Zeit*, 6 January 2019, Section: Kultur, https://www.zeit.de/2019/02/salvator-mundi-leonardo-da-vinci-gemaelde-verkauf.
[6] Karl Clausberg, '1984 wieder hinter Schloss(er) und Riegl? - Ein Kongreß-Ausblick', *kritische berichte - Zeitschrift für Kunst- und Kulturwissenschaften*, 11: 3, 1983, 71–74.
[7] Max J. Friedländer places connoisseurship beyond consciousness ('jenseits der Bewußtseinsschwelle'), quoted after Claudia-Alexandra Schwaighofer, *Von der Kennerschaft zur Wissenschaft*, Munich, 2009, 116.
[8] Every 'rational man' can learn to be a connoisseur, as Jonathan Richardson claims. See Schwaighofer 2009, 38.
[9] Shigeru Watanabe, Junko Sakamoto, Masumi Wakita, 'Pigeons' discrimination of paintings by Monet and Picasso', *Journal of the Experimental Analysis of Behavior*, 63, 1995, 165–174.

which images the machine will learn from, what metadata is added, and which images are used for testing. In that sense, computer vision is 'pure' vision, without any synesthetic interference. It is not supplemented, for instance, by tactile information, which Alois Riegl considers an essential aspect of perception or embodiment in general.[10]

Since William Vaughan's first experiments with image processing in the 1980s, many different applications have been developed to attribute works of art to an author or school. Most of them pick a specific domain of connoisseurship like the analysis of brushstrokes or canvas, others neglect the details and compare works as a whole. Vaughan developed a software to compare Rembrandt's oeuvre and its reproductions.[11] However, the name of the system — 'Morelli' — was more of a homage and had no methodical grounding. Moreover, connoisseurship was only one of many topics among the early attempts to combine art history and computer vision. This may be related to the fact that the question of attribution plays only a minor role in computer vision; innovation happened, and continues to happen, particularly in the areas of image content analysis and image understanding.[12]

One example of a detail-oriented approach is a study by Johnson et al. that compares original Van Gogh paintings with paintings of uncertain or provably different provenance. The study focuses on examining the form and orientation of Van Gogh's brushstrokes[13] to distinguish between the two classes. The wavelet analysis approach used in the study was also applied to the oeuvre of Bruegel the Elder and Perugino.[14] However, it is obvious that this approach is bound to fail in the case of Leiden Fijnschilders like Gerrit Dou and other artists who avoid visible brush strokes.

Another often-invoked distinctive property is colour. Instead of brushstrokes and shape, arranging images by colour is computationally simple and produces impressive visualisations. This is why Lev Manovich's experiments are quite well-

[10] Alois Riegl, *Historische Grammatik der bildenden Künste*, Graz, 1966, 129.
[11] William Vaughan, 'Computergestützte Bildrecherche und Bildanalyse', Hubertus Kohle (ed.), *Kunstgeschichte digital. Eine Einführung für Praktiker und Studierende*, Berlin, 1997, 97–105.
[12] See David G. Stork, 'Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature', *Computer Analysis of Images and Patterns: Proceedings of the 13th International Conference, CAIP 2009, Münster, Germany, September 2-4, 2009*, 9 –24; see also: Peter Bell, Leonardo Impett, 'Ikonographie und Interaktion. Computergestützte Analyse von Szenen der Evangelien', *Das Mittelalter. Perspektiven mediävistischer Forschung. Themenheft Digitale Mediävistik*, 24: 1, 2019, 31–53.
[13] C. Richard Johnson, Ella Hendriks, Igor Berezhnoy, Eugene Brevdo, Shannon Hughes, Ingrid Daubechies, Jia Li, Eric Postma, and James Z. Wang, 'Image processing for artist identification', *IEEE Signal Processing Magazine*, 25: 4, 2008, 37–48.
[14] Siwei Lyu, Daniel Rockmore, Heny Farid, 'A digital technique for art authentication', *Proceedings of the National Academy of Sciences of the United States of America*, 101: 49, 2004, 17006–10.

known in the digital humanities and digital art history communities.[15] Here, Van Gogh's popular oeuvre is sorted by brightness and saturation. Importantly, this sorting was sufficient for Lev Manovich to reconstruct Van Gogh's movement from Paris to Arles, relying solely on the change in colour.

Several approaches to the operationalisation of attribution and connoisseurship have been developed by Ahmed Elgammal's research group at Rutgers. Here, too, classically trained art historians were consulted, and their methods emulated. One project[16] examines brushstrokes, similar to the approach mentioned above but uses CNNs instead of wavelet analysis. Elsewhere[17], in an attempt to 'teach' Heinrich Wölfflin's concepts to the machine, Elgammal's group shows how a neural network can clearly differentiate between styles. In this particular study, however, 'style' is understood not only as a set of formal attributes, but also of motifs, genres, and techniques that were particularly common at the time. The study thus proposes a concept of 'style' that goes far beyond the narrow art-historical sense of the term and is closer to the notion of *zeitgeist*, or period eye. In other words, the study builds on pictures which somehow represent their historical moment but without relying on either a specific notion of style or the picture as a whole. Other, more elaborate approaches look for style indicators within the image to distinguish between different hands or copied parts.[18] Finally, it is important to mention that, beyond the painting as a semantic surface, material aspects like the texture of wood and paper, or the weaving pattern of the canvas can indicate provenance.[19]

In our own experiments below, we choose canonical examples of attribution. This basic approach is just meant to visualise the performance and features of a current computer vision model. We ask if the machine recognises the presence of the artist in the work. We present an attempt to learn to distinguish Braque from Picasso on the one hand, and Filippo Lippi from his son Filippino Lippi on the other. These are connoisseurly tasks that were challenging to earlier art history, but which are regarded as 'solved' today. Our aim is to use them to evaluate the capabilities of the machine in a transparent way. In 2015, we stated that 'the

[15] Lev Manovich, 'Museum without walls, art history without names: Visualization methods for humanities and media studies', Carol Vernallis, Amy Herzog, and John Richardson (eds.), *Oxford Handbook of Sound and Image in Digital Media*, Oxford, 2013.
[16] Elgammal, Ahmed, Yan Kang, and Milko Den Leeuw, 'Picasso, Matisse, or a fake? Automated analysis of drawings at the stroke level for attribution and authentication', arXiv preprint 1711.03536, 2017.
[17] Elgammal, Ahmed, Bingchen Liu, Diana Kim, Mohamed Elhoseiny, and Marian Mazzone, 'The shape of art history in the eyes of the machine', arXiv preprint 1801.07729, 2018.
[18] Nanne Van Noord, Ella Hendriks, Eric Postma, 'Toward discovery of the artist's style: Learning to recognize artists by their artworks', *IEEE Signal Processing Magazine*, 32: 4, 2015, 46–54.
[19] Margaret Holben Ellis, C. Richard Johnson Jr, 'Computational connoisseurship: Enhanced examination using automated image analysis', *Visual Resources*, 35:1-2, 2019, 125–140.

computer lacks intuition, its advantage is the processing time and the capacity to retrieve thousands of images and bring them into visual correspondence.'[20] In 2020, this is still the case, but the paradigm change to deep learning via convolutional neural networks brings new potential operationalisations to the table.[21] Some of these we discuss below.

In our experiments, we work with standardised machine learning approaches. As our primary architecture, we utilise a convolutional neural network architecture called VGG19[22]. Instead of training this architecture from scratch, for both tasks we use established transfer learning techniques to leverage low-level feature detectors already present in models that have been pre-trained on large-scale datasets like ILSVRC2012[23]. In the first experiment, we fine-tune such an ImageNet pre-trained VGG19 architecture on a dataset of Picasso's and Braque's paintings and drawings from the years 1907-1925, downloaded via script from the Prometheus image archive.

As is well known, it is not easy to distinguish between Picasso's and Braque's cubist paintings and to justify such distinctions.[24] Lyon states: 'The discoveries Picasso and Braque had made together during 1911-12 began to lead them in somewhat divergent directions by the end of 1913. When the war broke out in 1914, it spelled an end to their collaboration.'[25] Braque went on with Cubism after 1917 whereas Picasso changed his style. For the experiment, we define the historical period of interest as the years 1907-1925, to potentially also learn something about the individual characteristics of transformation.

The dataset consists of 400 Braque and Picasso paintings each from this period. For the purpose of fine-tuning, the dataset is split into 300 images for training, 60 images for validation, and 40 images for testing. In machine learning, this split is necessary to avoid overfitting, the simple memorisation of data, and facilitate generalisation, the learning of a classification function that generalises to unseen data. The fully trained model reaches 96% validation accuracy, i.e. it is able

[20] Peter Bell, Björn Ommer, 'Digital connoisseur? How computer vision supports art history', Stefan Albl and Alina Aggujaro (eds.), *Il metodo del conoscitore - approcci, limiti, prospettive Connoisseurship nel XXI secolo*, Rome, 2016, 187–200.
[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, 'ImageNet classification with deep convolutional neural networks', *Communications of the ACM 60*: 6, 2017, 84–90.
[22] Karen Simonyan, Andrew Zisserman, 'Very deep convolutional networks for large-scale image recognition', arXiv preprint 1409.1556, 2014.
[23] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei, 'ImageNet large scale visual recognition challenge', *International Journal of Computer Vision*, 115, 2015, 211–252.
[24] Max Imdahl, 'Cézanne - Braque - Picasso. Zum Verhältnis zwischen Bildautonomie und Gegenstandssehen', *Wallraf-Richartz-Jahrbuch*, 36, 1974, 325–365; William Rubin, *Picasso and Braque: Pioneering Cubism*, New York, Boston, 1989.
[25] Christopher Lyon: 'A Shared Vision', introduction to *Picasso and Braque: Pioneering Cubism*, MoMA, 2: 2, Autumn, 1989, 7–13, 8.

to successfully distinguish Picasso from Braque in 96% of cases. Importantly, misclassifications happen primarily on ambiguous images, i.e. those images that a human would also potentially misclassify. To test how the model's learned approach to classification mirrors established historical principles of connoisseurship, we visualise the 'attention' of the model with respect to its internal layers, i.e. with respect to different levels of the hierarchy of its learned features with the help of the Grad/CAM method.[26]

From the visualisations we can infer two things: The fine-tuned classification layer itself does not seem to correspond to any meaningful distinctions between the works of the two painters. The Grad/CAM images show that the attention of the model, with respect to this layer, routinely lies on parts of the image that are obviously insignificant for attribution, with few exceptions (fig. 1).[27]

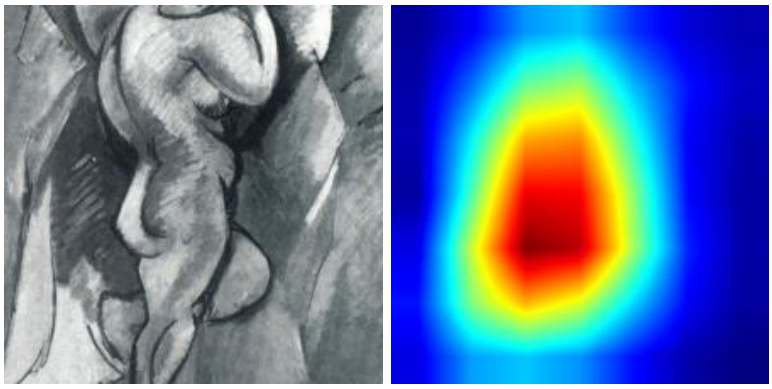

Figure 1 Correctly classified sample ('Picasso') from the Picasso/Braque corpus test set and Grad/CAM visualisation w.r.t. Layer 4 of the VGG19 network. Later layers focus on 'objects' in the image. In this specific case, the existence of 'round' objects (suggesting human figures more prevalent in Picasso) seems to be an important feature for the model © Authors.

Grad/CAM visualisations linked to lower-level layers in the model (fig. 2), however, seem to correspond better to meaningful details in the paintings. Intuitively, this corresponds to the precedence of formal aspects over representational aspects in the work of Picasso and Braque: Low-level features like edges, patterns, etc. are indeed more meaningful than potential 'objects' for attribution. It should be noted that the dataset is too small to empirically validate

[26] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', *Proceedings of the IEEE International Conference on Computer Vision*, 2017, 618–626.

[27] The general question of detecting representation in abstract paintings has also been approached with machine learning, see Shiry Ginosar, Daniel Haas, Timothy Brown, and Jitendra Malik, 'Detecting people in cubist art', *European Conference on Computer Vision*, 2014, 101–116.

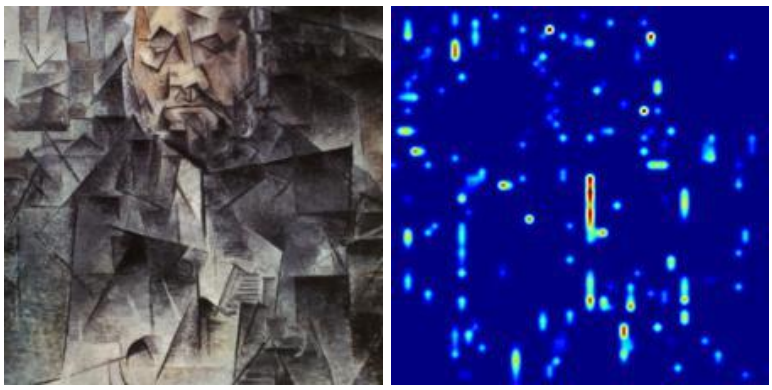

Figure 2 Ambiguous, but correctly classified sample ('Picasso') from the Picasso/Braque corpus test set, and Grad/CAM visualisation w.r.t. Layer 1 of the VGG19 network. As to be expected of an early layer, the visualisation shows an increased attention on lines vs. more representational image features © Authors.

this hypothesis. It is nevertheless relevant to the question of automated connoisseurship in so far as it shows that, in delegating questions of attribution to a machine learning system, we might have to rethink the usual hierarchical approach, depending on the kind of art objects under investigation. Indeed, state-of-the-art deep learning approaches are designed for 'object detection' and thus might need to be revisited when it comes to works of art that have historically abandoned representation.

For the second experiment, which deals with clearly representational works of art, this problem is less relevant. In fact, the representational quality of the dataset allows the introduction of additional approaches based on 'historical' art-historical hypotheses. The corpus on which the second experiment is based has been scraped from the Web Gallery of Art website. It contains 100 images depicting works by Fillipo Lippi (ca. 1406-1469), and 100 images depicting works of Filippino Lippi (ca. 1457-1504). In the late Florentine *quattrocento*, father and son formed a triad and workshop context with Botticelli who, like Filippino, was Filippo's student — a challenge to the connoisseur.

The parameters for the training of the VGG19 architecture stay the same for the second experiment: A VGG19 network, pre-trained on ImageNet, is fine-tuned on the Lippi/Lippi corpus. The resulting classifier is then tested on the holdout images from the test set, and the attention of the model is visualised with the Grad/CAM method. While the fully trained classifier also reaches a reasonable accuracy of 86% for this dataset, the Grad/CAM visualisations do not seem to indicate that any meaningful representation of connoisseurship principles has been learned, except for a slight focus on hands for some test cases.

This is why one of the additional approaches we introduce as part of the second experiment is the separate analysis of human hands in the Lippi/Lippi corpus. Giovanni Morelli is known for introducing the idea of a detailed analysis of anatomical details as a means to solve attribution questions. Instead of relying on

the instant, 'total' impression of an image, he proposed to focus on details like fingers, hands, and ears.

In the second experiment, we take up this proposal by extracting (almost) all hands from the images in the corpus. This is achieved by running a second pre-trained model, a keypoint RCNN based with ResNet50 backend, on the corpus. This model returns a set of keypoints for each human figure identified in an image. As the model has been trained on photographic representations, this identification of human figures does not achieve the best possible results for painted or drawn figures, such as those in the corpus, but still identifies most figures. For the identified figures, we predict the position of the hand (which does not have its own keypoint) by moving the wrist keypoint into the direction of the wrist-elbow vector, and then drawing a bounding box. The dimensions of the bounding box are calculated in relation to the size of the identified figure (fig. 3).

Figure 3 Detected 'human' figure (green), keypoints and computed wrist-elbow vectors (white) and resulting hand bounding boxes (red) for a sample from the Lippi/Lippi corpus © Authors.

The resulting corpus of hands contains about 130 hand images for each artist in the original corpus. This additional corpus is then analysed in the same way as the original corpus, by fine-tuning a VGG19 network pre-trained on ImageNet. Surprisingly, the resulting classifier still reaches 75% accuracy, implying that there are, at least, some operationalisable differences between the way both artists depicted hands. Of course, pinpointing these differences becomes more difficult

with decreasing accuracy, as the classifier becomes less 'trustworthy'. Nevertheless, the Grad/CAM visualizations for the classification layer show that the model's attention lies indeed on the hands in most cases, again making the case for hands as a salient feature for automated attribution (fig. 4).

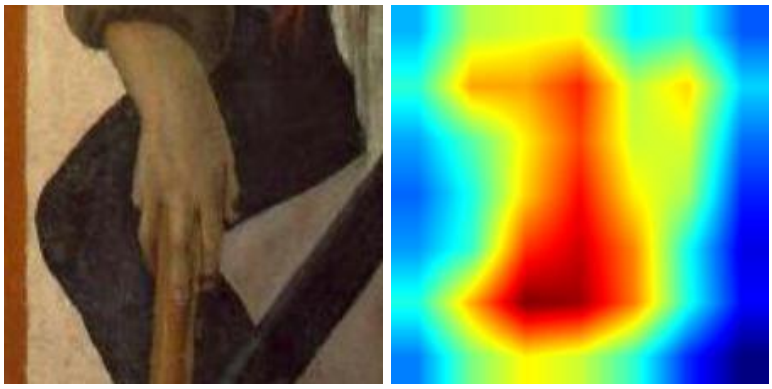

Figure 4 Extracted hand region from the test set and Grad/CAM visualisation that shows that the hand-classifier's attention lies indeed on the actual hand, rather than on auxiliary image details © Authors.

It should be noted that, due to the relatively small sizes of the datasets in our experiments, the described results are not necessarily a confirmation of any potential mapping of human approaches to connoisseurship to neural network features. In fact, it is well known that neural networks often find 'shortcuts'[28] and infer classification-relevant information from semantically irrelevant aspects of the image, like high-frequency textures[29]. With decreasing dataset size, it becomes more likely that any classification task, including attribution, can be solved by simply identifying such 'adversarial'[30] features for the relevant classes. Nevertheless, in the case of connoisseurship, the point of these experiments is less to prove that connoisseurship can be automated — that has already been shown for multiple aspects of connoisseurship in the literature — but rather that such 'alien' modes of perception have to be taken into account. In other words: A working classifier, as those described for all three corpora above, does not necessarily imply that connoisseurship has been learned, much less that specific strategies of connoisseurship are represented as specific learned features in the neural network.

[28] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann, 'Shortcut learning in deep neural networks', arXiv preprint 2004.07780, 2020.
[29] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel, 'ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness', arXiv preprint 1811.12231, 2019.
[30] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, 'Explaining and harnessing adversarial examples', arXiv preprint 1412.6572, 2014; Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, and Rob Fergus, 'Intriguing properties of neural networks', arXiv preprint 1312.6199, 2013.

For the development of a true automated connoisseur, we might abandon the historical context embedding entirely, and further explore the strange but salient strategies of operationalisation that the computer proposes. This experiment will make the digital connoisseur a complementary observer to assist the human connoisseur, who will always – even as Morelli or Longhi – be able to connect the individual hand, the biography, history, and space to his or her perception.

Peter Bell is a Professor for Digital Humanities with a focus on Art History at Friedrich-Alexander-Universität Erlangen-Nürnberg. He held a Postdoc position at Ruprecht-Karls-Universität Heidelberg in the Interdisciplinary Center for Scientific Computing (IWR), and he was research group leader Heidelberg Akademie der Wissenschaften.

peter.bell@fau.de

Fabian Offert is Assistant Professor in History and Theory of Digital Humanities at the University of California, Santa Barbara. Previously, he was a Postdoctoral Researcher in the DFG priority program "The Digital Image" at Friedrich-Alexander-Universität Erlangen-Nürnberg and Affiliated Researcher in the Artificial Intelligence and Media Philosophy Research Group at Karlsruhe University of Arts and Design.

offert@ucsb.edu