

Generative Digital Humanities

Fabian Offert^{a,b}, Peter Bell^b

^aUniversity of California, Santa Barbara, U.S.A.

^bFriedrich Alexander University Erlangen-Nuremberg, Germany

Abstract

While generative machine learning has recently attracted a significant amount of attention in the computer science community, its potential for the digital humanities has so far not been fully evaluated. In this paper, we examine generative adversarial networks, a state-of-the-art generative machine learning technique. We argue that GANs can be particularly useful in digital art history, where they can be employed to facilitate the exploration of the semantic structure of large image corpora. Moreover, we posit that the foundational statistical distinction between discriminative and generative approaches offers an alternative critical perspective on machine learning in the digital humanities context. If “all models are wrong, some are useful”, as the often-cited passage reads, we argue that, in case of the digital humanities, the most useful-wrong models are generative.

Keywords

machine learning, generative models, data augmentation, digital art history

1. Introduction

The computational humanities, recently, have established themselves as a subfield of the digital humanities, or in some readings [25] even as a potentially separate field that pursues a re-alignment with computer science. This re-alignment implies both the identification of humanities questions requiring computational means that go beyond those of widely available tools¹, and the identification of areas of research in computer science that have significant historical or philosophical implications.

Other than, for instance, [18] and [25], however, we do not understand this re-alignment as the beginning of a new split between technical and non-technical digital humanities work. Instead, we understand it as an opportunity to extend the critical focus of the digital humanities to the technical systems themselves. We see the computational humanities not only as an opportunity to open up the digital humanities towards computer science but also as an opportunity to investigate the epistemological implications of the technical systems that drive this development. To put it differently: “computational humanities” can and should not only imply the application of state-of-the-art computational methods to humanities questions, but also the questioning of computational methods from “state of the art” critical perspectives².

CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands


✉ offert@ucsb.edu (F. Offert); peter.bell@fau.de (P. Bell)

🆔 0000-0003-4415-7408 (P. Bell)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

¹By “widely available tools” we mean specifically widely used stand-alone software packages like Gephi or Transcribus, but also “off-the-shelf” libraries like Gensim.

²It should be noted that several recent initiatives within the digital humanities, like the “tool criticism” workshop at DH2020 (<https://dls.hypotheses.org/1110>), propose a similar research direction. Along with an increasing amount of collaboration between the digital humanities and media studies, for instance within the

As a potential demonstration of this approach, this paper attempts to make both a theoretical and preliminary practical contribution to the integration of machine learning into digital humanities research. Concretely, we explore a subfield of machine learning which we believe to have great practical and critical potential, but which so far has not been studied in the digital humanities context. Recent research [1, 17, 34, 24] has shown the huge potential of adopting experimental machine learning methods for digital humanities research in particular. What is missing from these and similar investigations, however, is the subfield of generative machine learning.

On the theoretical level, we investigate the epistemological implications of generative machine learning techniques in the context of digital humanities research. Despite a seemingly infinite variety of highly specific machine learning applications, all machine learning is automated statistical modeling. The rules of statistics, even if they are applied to web-scale corpora of complex, high-dimensional data, stand as unifying principles behind all machine learning approaches. While this is obvious to computer science practitioners, it is often disregarded when machine learning is discussed in the humanities context, mirroring an often-diagnosed epistemological split [22] between computer scientists and engineers and digital humanities scholars. We argue that, somewhat counter-intuitively, statistical notions offer an alternative *critical* perspective on machine learning in the humanities context, and we suggest that the foundational statistical distinction between discriminative and generative approaches [23] can be utilized to guide the further development of the computational humanities. In other words, if “all models are wrong, some are useful”, as the often-cited passage by George Box reads [5], we argue that, in case of the computational humanities, the most useful-wrong models are generative models.

On the practical level, we explore the potential of generative machine learning techniques in the visual domain, targeting applications in digital art history. Concretely, we examine generative adversarial networks [13], a state-of-the-art image-based generative machine learning technique, in regard to their potential for data augmentation, i.e. the production of “realistic” additional data from an image corpus. As visual data augmentation methods using GANs have recently been applied successfully to a number of complex tasks in the sciences (e.g. Ravanbakhsh et al. [26]) we propose that they can also be of practical use in the digital humanities context, particularly in digital art history, where data is often particularly scarce.

We conclude by arguing that much more research is needed regarding both the practical potential and theoretical implications of generative machine learning in the digital and computational humanities.

2. Generative machine learning

For the purpose of this paper, we formally define generative machine learning as one of two possible statistical approaches – discriminative and generative – to modeling real-world data [23]. A toy example best demonstrates the difference between these two approaches in statistical terms. To distinguish between two kinds of objects, say apples and oranges, based on a dataset of labeled images of both, we can imagine two possible strategies of classification. We can design a model that learns the *most salient difference between apples and oranges* from the dataset. A good candidate for a distinctive visual feature would be color: apples are

recently established DFG research cluster “The Digital Image”, one could argue that this suggests the beginning of a “critical turn” in the field.

usually red; oranges are usually orange. The model then uses these most salient features to classify new, unseen samples. This is the discriminative approach. The generative approach, however, learns the *complete distribution of visual features for both apples and oranges*. Apples come in different shades of red, yellow, and green, oranges come in different shades of orange. The model then classifies new, unseen samples by comparing their visual feature distribution to the visual feature distribution for apples and the visual feature distribution for oranges. In other words: the discriminative approach attempts to model a decision boundary *between classes* (it literally learns “where to draw the line” between apples and oranges), while the generative approach attempts to model the actual distribution *of each class*. The generative approach essentially asks: what is the most likely *source* of the *signal* we are seeing, while the discriminative approach simply looks for a way to distinguish one signal from the other and does not take the source into account.

Formally, whereas the discriminative approach attempts to learn the *conditional* probability distribution $p(y|x)$, i.e. the probability of y given x , the generative approach attempts to learn the *joint* probability distribution $p(x, y)$, i.e. the probability of x, y to appear together. While the discriminative approach models $p(y|x)$ directly and thus allows us to find the most likely class y for any given set of features x , the generative approach models $p(y)$ (the so called class priors, i.e. the distribution of labels) and *both* $p(x|y = \textit{apple})$ and $p(x|y = \textit{orange})$. By applying Bayes’ rule, we can then derive the posterior distribution on y given x :

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Note that the generative approach is *unsupervised* (as there is one “model” for each class and the label of each class thus becomes irrelevant), even though it can be transformed into a supervised approach with the steps described above. In other words: the discriminative approach is a *sparse* approach to modeling the world – only what is needed from the world is taken into account – and the generative approach is a *dense* approach – an attempt to model the world as it is. More importantly, while a sparse approach can tell us something about the *abstract* properties of our data, a dense approach can tell us something about the *concrete* properties of our data. While both approaches can thus solve classification problems, the discriminative approach is obviously simpler. As resources are limited, we usually would like to avoid learning things that are not pertinent to the problem. To distinguish apples from oranges, for instance, learning the shape of both seems irrelevant. As Vapnik [33] puts it: “one should solve the [classification] problem directly and never solve a more general problem as an intermediate step [such as modeling $p(y|x)$].”

Taking the generative route, however, comes with one specific benefit: it enables us to synthesize data, to produce *new likely* data points for each of the classes we are modeling. For instance, by sampling the distribution $p(x|y = \textit{apple})$ we would get a set of features x_0, \dots, x_n describing one possible manifestation of “apple”. This is impossible with the discriminative approach, because a discriminative model has only learned the *difference* between apples and oranges, and not what apples and oranges *are* (in terms of their features).

3. Generative digital humanities

The digital humanities have often been broadly criticized for the mere use of quantitative methods, as eloquently summarized in Ted Underwood’s blog post “It looks like you’re writing

an argument against data in literary study...” [32]. While some of these critiques from the early days of the field still resonate, and general critiques of quantitative methods occasionally reappear with force (as recently in Claire Bishop’s critique of digital art history, Bishop [4]), generally, a consensus has grown that such a blanket rejection has no grounding in the reality of digital humanities work. The focus of critique, thus, has shifted to more elaborate discussions of theory building with quantitative methods [30], and the shortcomings of specific statistical tools in the analysis of cultural data.

One of the most powerful recent critiques is Nan Z. Da’s article “The Computational Case against Computational Literary Studies” [8]. Da writes: “all the things that appear in [Computational Literary Studies]—network analysis, digital mapping, linear and nonlinear regressions, topic modeling, topology, entropy—are just fancier ways of talking about word frequency changes.” Based on a formal distinction between discriminative and generative methods, as outlined above, we can see, however, how some of these methods are not the same. In fact, linear regression and topic modeling fall on opposite ends of the spectrum between discriminative and generative approaches, and the prevalence of latent Dirichlet allocation (LDA), also known as (one variation of) topic modeling, in computational literary studies and in the digital humanities in general points to an even stronger claim: the digital humanities “intuitively” choose generative over discriminative approaches because they are more aligned to humanities data.

Why do the digital humanities gravitate towards generative approaches? Because generative approaches mitigate, at least in part the alienation, the general inadequacy of quantitative methods vis-a-vis cultural artifacts. Quantitative methods, obviously, can never fully represent cultural artifacts. Precisely, both the sampling of cultural artifacts into data, and the modeling of this data are reductive. In the domain of modeling, however, generative approaches stay as close to the material as possible, while discriminative approaches essentially “ignore” the material for the sake of classification. In other words, generative approaches, while not being able to mitigate the problems introduced by sampling, can mitigate the problems introduced by modeling within the realm of what can be modeled.

Regardless, many of the problems discussed in Da’s article stand with or without generative machine learning. Shoddy hypothesis building or the lack thereof, intentional or unintentional over- or misinterpretation of the empirical evidence quantitative methods can offer, or too-broad applications of narrow technical concepts are problematic irrespective of the kind of model involved. Hence, a focus on generative approaches does not “solve” or even explicitly address these issues. Generative approaches do not magically produce a “self-reflexive account of what the model has sought to measure and the limitations of its ability to produce such a measurement”, as Richard Jean So writes [29]. On the contrary, generative methods tend to actually, implicitly encourage the problematic “exploratory” approach that became a central argument in the discussion following Da’s article³ [31, 7].

Moreover: famously, the existence of “raw data” is an illusion [12, 21, 10], and the existence of “neutral algorithms” even more so [3, 6, 9]. Thus, if we propose that generative methods “stay as close to the material as possible”, we do not imply the absence of subjective guidance through the design or selection of algorithms and datasets. Indeed, it is not only dataset bias that shapes machine learning models, but inductive biases induced by pragmatic architectural decisions often further entangle subjective and machinic perspectives [24, 11].

What a critical distinction between generative and discriminative approaches offers, regard-

³One could argue that this is an effect of the reduced interpretability of generative methods.

less, is a prospective path through current and future experimental work in computer science, where the digital humanities, we argue, need to critically consider the distinction between generative and discriminative methods in the evaluation of new, experimental tools and methods – all while keeping in mind that the maximum benefit of all machine learning models is a “useful-wrong” model in the sense of Box [5], i.e. a model that stays “reasonably” close to the material.

4. Generative digital art history

In the following, we sketch such a path for digital art history. As Leonardo Impett has pointed out [16] the computer vision problems that art history relates to are almost exclusively search-related, i.e. they are classification problems. What if digital art history would start focusing on generative methods instead? Would a closer relation to the material also establish itself in the domain of images? Generative methods are already implicitly employed in the neural network based clustering of images, which has become increasingly more popular in digital art history in recent years [34]. When embeddings are utilized, a learned sub-system of a classifier is repurposed exactly for its generative properties, which is also why recent research [14] suggests building explicitly generative systems for the specific purpose of clustering. In the realm of explicitly generative systems, then, we argue that *generative adversarial networks* can become for the visual domain what LDA became for the text domain: an instrument of unsupervised exploration for large-scale corpora.

Generative adversarial networks, first introduced by Goodfellow et al. [13] leverage game theory to model the probability distribution of a corpus by means of a minimax game between two deep convolutional neural networks. Effectively, generative adversarial networks define a noise distribution p_z which is mapped to data space via $G(z; \theta_g)$ where G is a “generator”, an “inverted” convolutional neural network with parameters θ_g that “expands” an input variable into an image, rather than “compressing” an image into a classification probability. G is trained in conjunction with a “discriminator”, a second deep convolutional neural network $D(z; \theta_d)$ that outputs a single scalar. $D(x)$ represents the probability that x came from the data rather than G . Note that the whole system, not just the “generator”, realizes the generative approach, as the whole system is needed to model $p(x|y=0)$. Also note that the system effectively learns a compression: a high-dimensional data space with dimensions $> z$ is compressed to be reproducible from a data space with dimensions z .

The original paper by Goodfellow demonstrates the potential of generative adversarial networks to synthesize images in particular by synthesizing new handwritten digits from the MNIST dataset. The MNIST dataset, however, has a resolution of 28x28 pixels, i.e. several orders of magnitude below standard photo resolutions. Scaling up the approach proved difficult, and while a lot of effort was made to go beyond marginal resolutions, progress was slow (for machine learning) until very recently, when StyleGAN [19], a generative adversarial network that implemented several significant optimization tricks to mitigate some of the limitations of generative adversarial networks, was introduced. Current-generation models like StyleGAN2 [20], which presents another improvement over the original StyleGAN, are able to produce extremely realistic samples from large image corpora, samples that – the reader should keep this in mind – are not part of any training corpus, but share the decisive features of the images in a corpus.

Such samples, to the humanist, feel uncanny. GANs obviously learn “something” (maybe

everything?) about a corpus, and GAN samples “tick all the boxes” at the first glance. At the same time, GANs seem almost useless. What knowledge is there to gain from a model that essentially learns to recreate approximations to what exists, and nothing *about* what exists? Interestingly, for a long time and despite impressive early results, the utility of generative adversarial networks was not entirely clear in the computer science community either. And while today there are obvious applications in digital image processing (inpainting, superresolution, image-to-image translation, style transfer etc.) and manipulation (deep fakes etc.), the epistemological qualities of GANs, i.e. their role in scientific (and, we argue, humanist) processes of discovery, are still not fully explored.

5. Generative data augmentation

In the meantime, however, the targeted generation of images with GANs has been improved significantly. Bau et al. [2] had already shown in 2018 that GANs learn less entangled representations than comparable CNNs. More recently, approaches have been found that allow the unsupervised identification of meaningful hyper-directions in GAN latent spaces, with an approach called GANspace [15] currently providing the most efficient method⁴. In other words, if recent experimental approaches are taken into account, latent spaces become accessible to exploration.

Why is this relevant to the potential exploitation of generative methods in the visual domain? An important epistemological aspect of GANs is the continuity of the latent spaces that can be produced. GAN latent spaces are, for better or worse, “filled to the brink” with images. This also means, for each sample there exist millions of other samples that look almost like this sample, except for tiny details. It also means that, between each two samples we can find theoretically infinite “intermediate” samples, hybrid images that combine aspects of both of the samples between which they are positioned. Digital humanities corpora, on the other hand, visual or otherwise, always exist as discrete collections of samples. Concretely, in an art historical corpus, there is no “intermediate” image between, for instance, a *Titian Marriage at Cana* and a *Last Supper*, or between the latter and a *Last Supper of Titian*. However, if we train a GAN on this corpus, such an “intermediate” image suddenly comes into existence. Simply put we argue that GANs can reintroduce a certain continuity to a corpus that allows to study the discreteness of the corpus itself.

In the following, we present a proof of concept for this approach. In a first step, we train a generative adversarial network, following the StyleGAN architecture, on two art historical corpora. First, an iconographic corpus of 20,000 “adoration” scenes. This corpus contains images referencing the adoration of the Christ Child; in particular the *Adoration of the Child (by Mary and Joseph)*, the *Adoration of the Shepherds*, and the *Adoration of the Magi*. The second corpus contains 50,000 images from the Museum of Modern Art, New York, online collection. The hypothesis, here, is that a GAN, by means of compression, would learn the most salient differences between images in a corpus. In the case of the adoration corpus in particular, which is drawn from several centuries of visual culture, these differences would likely not only relate to the number and arrangement of people and objects, but also the style and medium of the works. In a second step, we then analyze the most salient hyper-directions in the learned latent space with the help of the GANspace method. Importantly, GANspace is

⁴In July 2020, another, conceptually different, approach was published [28] that shows even more promising results. Unfortunately, we were unable to test it on our data before the submission deadline of this work.

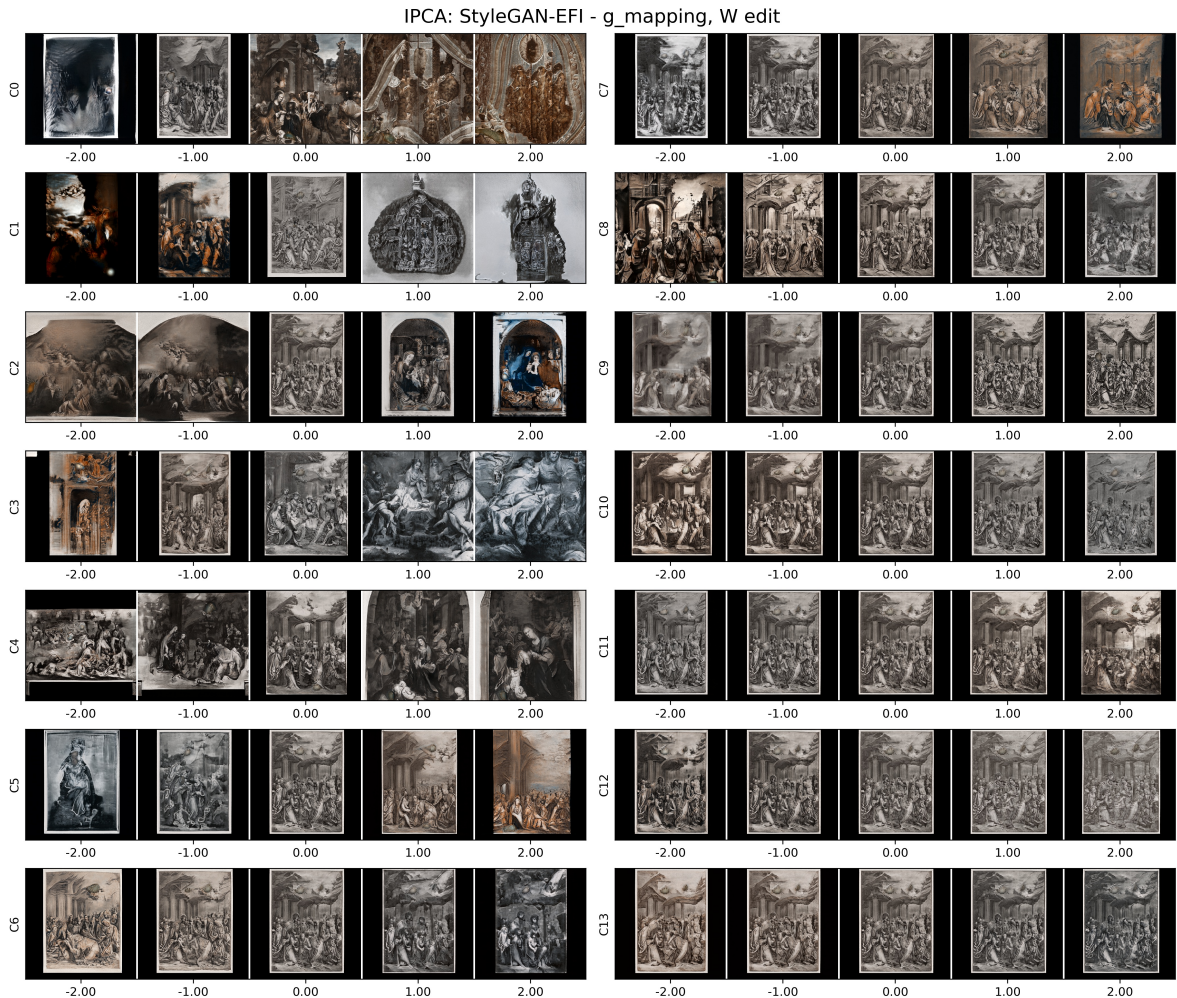


Figure 1: Meaningful directions in the latent space of a StyleGAN model trained on our corpus of 20,000 adoration scenes, as predicted by GANspace. All directions are shown starting from the same random z , going two steps in the discovered direction, and its inverse.

an unsupervised method, i.e. no labeling is involved.

What we find is that, indeed both semantic and syntactic hyper-directions in latent space emerge. For the adoration corpus (fig. 2), an interpretation of these directions suggests that, among other things, they represent syntactic concepts such as “painting or object” (C1), “precious carving or non finito sculpture” (C4), “pencil or colored drawing” (C7), “sharp or blurry outlines” (C9) and semantic concepts such as “zooming into a scene” (C3) and “number of people in a scene” (C5, C6). For the MoMA corpus (fig. 3), unsurprisingly for a corpus composed of mainly abstract art, we find mostly syntactic concepts such as “figurative or abstract” (C0, C13), “organic or technical” (C3), “drawing or painting” (C1), “textural or graphical” (C8).

It is important to point out that, due to the proof-of-concept nature of the above experiment, significant caveats apply. Obviously, our results are exploratory and trivial in the exact sense criticized by Da. At this point, they do not expand our knowledge about either the



Figure 2: Meaningful directions in the latent space of a StyleGAN model trained on our corpus of 50,000 works of modern art from the MoMA collection, as predicted by GANspace.

concrete corpus or about a potential iconography represented by it but simply confirm our preconceived ideas about both. Moreover, comparable results could have likely been achieved by more established methods, like clustering based on CNN features, or a principal component analysis in pixel space. Finally, GAN latent spaces, of course, are imaginary spaces. They are reconstructions of the defining features of a corpus, and exploring such spaces is not the same thing as exploring the corpus itself.

This imaginary quality, however, that is deeply problematic in any other application of generative methods (for instance, in the sciences), can precisely be of use in the digital humanities context. Here, GAN samples are not mistaken for valid information generated from nothing, as in so many recent examples, but can be understood as an additional means to ask questions about the information we do have, about the corpus at hand.

While our concrete results thus remain preliminary, we argue that they point towards a significant potential of generative methods in the visual domain. By reintroducing continuity to a corpus of discrete images, we are forced to precisely quantify the semantic thresholds that

support its discreteness. Discrete concepts are transformed into continuous variables. What, exactly, defines a certain iconography? How far in any direction (in the literal sense of latent space hyper-directions) can we veer off until an image that is clearly recognizable as belonging to a certain iconographic tradition stops being recognizable as such? A synthetic grammar of art emerges that is not historical like Riegl's [27] but diachronic and multimodal. In a sense, if generative approaches automatically stay close to the material, using GANs means staying closer to the material than actually possible by augmenting it.

6. Conclusion

In this paper, we have demonstrated how the statistical distinction between generative and discriminative approaches can inform the methodological discourse in the digital humanities and can be understood as a starting point for a deep technical exploration in the computational humanities. We have argued that generative methods, while they are not immune against many of the general problems pointed out in recent discussions on the methodological grounding of digital humanities work, can mitigate some of the problems introduced on the level of data modeling. Moreover, we have proposed that, while computational literary studies and related sub-disciplines of the digital humanities have already implicitly embraced generative methods, the visual digital humanities lack equivalent tools.

We have also suggested to explore generative adversarial networks as a potential generative approach in digital art history and have documented a proof-of-concept approach utilizing StyleGAN and the GANspace algorithm to identify meaningful directions in the latent spaces of two GANs trained on art historical corpora. Based on the results from this experiment we have argued that, other than in scientific uses of GANs, the imaginary nature of GAN images allows for the emergence of a synthetic, continuous "replacement" corpus that, exactly by means of its continuity, can serve to delineate the semantic thresholds that define a collection of images.

Pragmatically, future research will have to show if GANs are the right tool for this purpose, or if other networks like variational autoencoders need to be considered. More importantly, future research will have to empirically verify the hypothesis that the synthetic corpora produced by GANs and related methods are interpretable enough to serve as a means to evaluate semantic (e.g. iconographic) concepts in specific art-historical corpora, or if more traditional methods remain the more viable approach for the time being.

References

- [1] T. Arnold and L. Tilton. "Distant Viewing: Analyzing Large Visual Corpora". In: *Digital Scholarship in the Humanities* (2019).
- [2] D. Bau et al. "GAN Dissection: Visualizing and Understanding Generative Adversarial Networks". In: *arXiv preprint arXiv:1811.10597* (2018).
- [3] R. Benjamin. *Race after Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons, 2019.
- [4] C. Bishop. "Against Digital Art History". In: *International Journal for Digital Art History* 3 (2018).

- [5] G. E. P. Box. “Robustness in the Strategy of Scientific Model Building”. In: *Robustness in Statistics*. Academic Press, 1979, pp. 201–236.
- [6] J. Buolamwini and T. Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. In: *Conference on Fairness, Accountability and Transparency*. 2018.
- [7] N. Z. Da. “Critical Response III. On EDA, Complexity, and Redundancy: A Response to Underwood and Weatherby”. In: *Critical Inquiry* 46.4 (2020), pp. 913–924.
- [8] N. Z. Da. “The Computational Case against Computational Literary Studies”. In: *Critical Inquiry* 45.3 (2019), pp. 601–639.
- [9] T. de Vries et al. “Does Object Recognition Work for Everyone?” In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2019, pp. 52–59.
- [10] J. E. Dobson. *Critical Digital Humanities: The Search for a Methodology*. University of Illinois Press, 2019.
- [11] R. Geirhos et al. “Shortcut Learning in Deep Neural Networks”. In: *arXiv preprint arXiv:2004.07780* (2020).
- [12] L. Gitelman. *Raw Data Is An Oxymoron*. Cambridge, MA: MIT Press, 2013.
- [13] I. Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems*. 2014, pp. 2672–2680.
- [14] J. M. Graving and I. D. Couzing. “VAE-SNE: A Deep Generative Model for Simultaneous Dimensionality Reduction and Clustering”. In: *bioRxiv preprint* (2020).
- [15] E. Härkönen et al. “GANSpace: Discovering Interpretable GAN Controls”. In: *arXiv preprint arXiv:2004.02546* (2020).
- [16] L. Impett. *Open Problems in Computer Vision*. Friedrich Alexander University Erlangen-Nuremberg, Mar. 2020.
- [17] L. Impett and F. Moretti. “Totentanz. Operationalizing Aby Warburg’s Pathosformeln”. In: *New Left Review* 107 (2017).
- [18] F. Jannidis. *No Results or Wrong? Methodological Challenges in Computational Literary Studies*. University of Leipzig, 2019.
- [19] T. Karras, S. Laine, and T. Aila. “A Style-Based Generator Architecture for Generative Adversarial Networks”. In: *arXiv preprint arXiv:1812.04948* (2018).
- [20] T. Karras et al. “Analyzing and Improving the Image Quality of StyleGAN”. In: *arXiv preprint arXiv:1912.04958* (2019).
- [21] B. Latour. “Circulating Reference: Sampling the Soil in the Amazon Forest”. In: *Pandora’s Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press, 1999, pp. 24–79.
- [22] G. Mercuriali. “Digital Art History and the Computational Imagination”. In: *International Journal for Digital Art History* 3 (2019), p. 141.
- [23] A. Y. Ng and M. I. Jordan. “On Discriminative vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes”. In: *Advances in Neural Information Processing Systems*. 2002, pp. 841–848.

- [24] F. Offert and P. Bell. “Perceptual Bias and Technical Meta-Images. Critical Machine Vision as a Humanities Challenge”. In: *AI & Society* (2020).
- [25] M. Piotrowski. “Ain’t No Way Around It: Why We Need to Be Clear About What We Mean by “Digital Humanities””. In: (2020).
- [26] S. Ravanbakhsh et al. “Enabling Dark Energy Science with Deep Generative Models of Galaxy Images”. In: *Thirty-First AAAI Conference on Artificial Intelligence*. 2017, pp. 1488–1494.
- [27] A. Riegl. *Historical Grammar of the Visual Arts*. New York, NY: Zone Books, 2004.
- [28] Y. Shen and B. Zhou. “Closed-Form Factorization of Latent Semantics in GANs”. In: *arXiv preprint arXiv:2007.06600* (2020).
- [29] R. J. So. “All Models Are Wrong”. In: *PMLA* 132.3 (2017), pp. 668–673.
- [30] N. Tahmasebi and S. Hengchen. “The Strengths and Pitfalls of Large-Scale Text Mining for Literary Studies”. In: *Sammlaren: Tidskrift för Svensk Litteraturvetenskaplig Forskning* 140 (2019), pp. 198–227.
- [31] T. Underwood. “Critical Response II. The Theoretical Divide Driving Debates about Computation”. In: *Critical Inquiry* 46.4 (2020), pp. 900–912.
- [32] T. Underwood. *It Looks like You’re Writing an Argument against Data in Literary Study ...* 2017.
- [33] V. Vapnik. *Statistical Learning Theory*. New York, NY: Wiley, 1998.
- [34] M. Wevers and T. Smits. “The Visual Digital Turn: Using Neural Networks to Study Historical Images”. In: *Digital Scholarship in the Humanities* 35.1 (2020), pp. 194–207.