Reproduced bits

Reproduced Bits

Maarten Delbeke, Benoit Seguin

Until not long ago artistic production was defined in relation to models. Creation was imitation: selecting a model appropriate for the task at hand, which guided the creative process, and helped the recipient understand the intention of the work. In this process, the model remained present in the new work which, in turn, was always a comment on its model.

Given this mode of artistic production, tracing models is important for historians of art and architecture. Identifying models of texts, buildings or works of art reveals artistic intentions, communities of ideas, transfers of knowledge, and critical positions. Paradoxically, the search of models is driven by the same 20th-century obsession with originality that terminated the practice of imitation; innovation can only be measured against models. Shared motifs, ideas or types delineate originality.

Imitation is closely related to reproduction. The invention of moveable type and the development of accessible techniques to reproduce images, enabled the reproduction of models. This allowed for the reliable transmission of identical copies of a single work, and availed ever widening circles of actors of shared models. Exactly the same source could be altered, reinterpreted or transferred between media regardless of whether a single original was available. When it comes to architecture, printed editions of Vitruvius' treatise permitted an army of interpreters to use versions of the same text.

Exploring the relationships between models and their variations is therefore important in the history of art and architecture. It is also difficult, entailing the exploration and comparison of potentially numerous references. Traditionally, this work is guided by bibliographical instruments and studies that define corpora of sources and outline possible connections. While extremely valuable, these instruments tend to limit findings to an established canon. Moreover, exploration of sources can take years to validate a research hypothesis, and will always remain provisionary. After all, the corpus will never be exhaustive. This limitation is compounded by the research method: manual exploration. Researchers study, compare and analyse a body of material, guided by what they have read or experienced before. This work is limited by the human capacity to master large bodies of information.

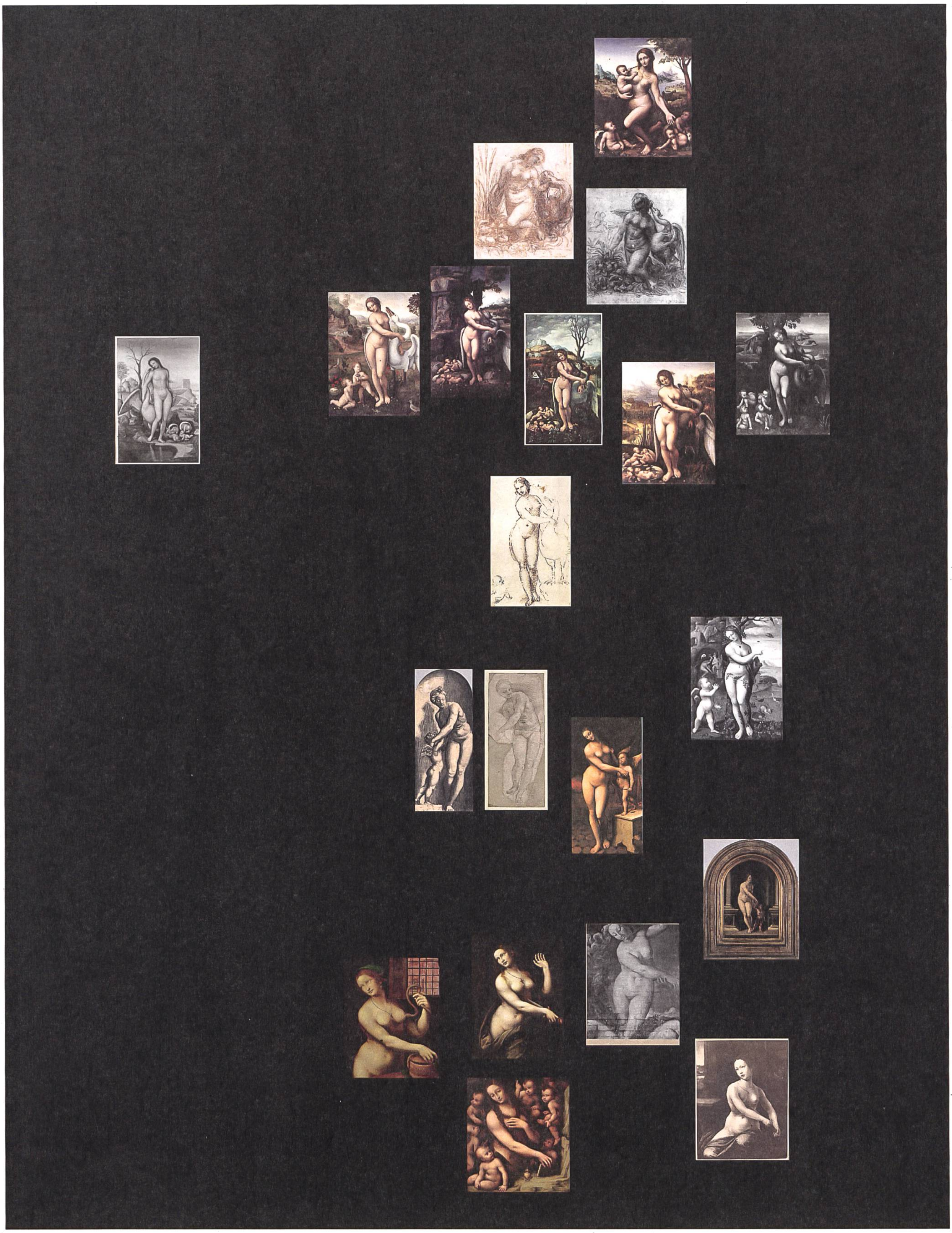
Over the last years, two emerging phenomena have challenged these limitations: the extensive digitization of sources and the development of computational methods. Computation and the rise of artificial intelligence have opened new avenues in the processing and analysis of documents. Contents of digitized materials can be extracted more faithfully than ever, and text and images are understood at an unprecedented level. Thanks to the exponentially increasing availability of digitized sources in a standardized format, new opportunities arise in automated interaction with very large corpora.

Detecting reproductions is possible with computational methods because it is an «operationalizable» concept: it can be encoded in a mathematical formulation that does not require a prior knowledge or an understanding of the material at hand. For instance, the use of the same words in the same order in totally different works can be detected automatically, even if the algorithm does not grasp the meaning of the sentence and only sees an undecipherable sequence of characters. If detecting textual correlation – a pure operationalizable concept – is only the first step in understanding complex model-reproduction processes, it is necessary nonetheless.

Reproduction in text takes multiple forms: from plagiarism, quoting or paraphrasing to representing similar ideas. This scale implies a gradient of semanticity, and requires different degrees of understanding. As such, they are more or less operationalizable. When semantics are involved, algorithms become more complex and success-rates drop. While detecting the same sequence of characters across two documents is relatively easy, being able to handle OCR errors, detect paraphrasing, or process large corpora is possible but complicated. Handling different languages might only be feasible in the near future.

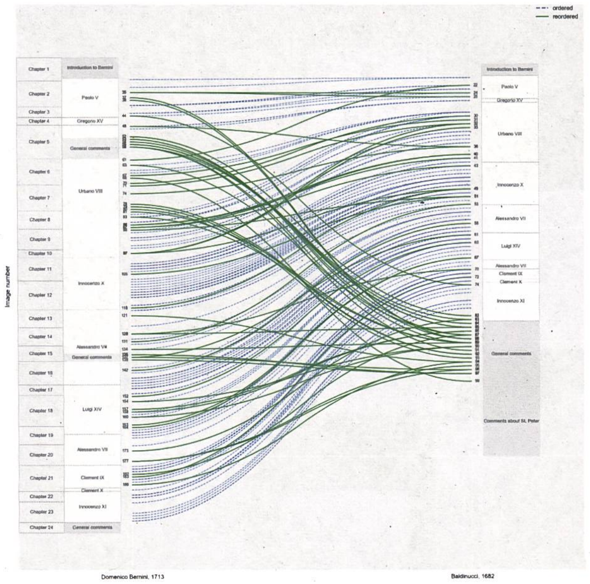
Tracing reproductions of images follows a similar scheme and involves an even more complex typology of relationships: are two digital images artefacts of the same physical object? Different scans of the same physical photograph? Did two prints come out of the same woodblock? Or was one based on the other? Is an engraving or drawing visually similar to a painting? If so, because of its composition or of a near-identical element?... Each question corresponds to different connections between two images, and engages a different

A



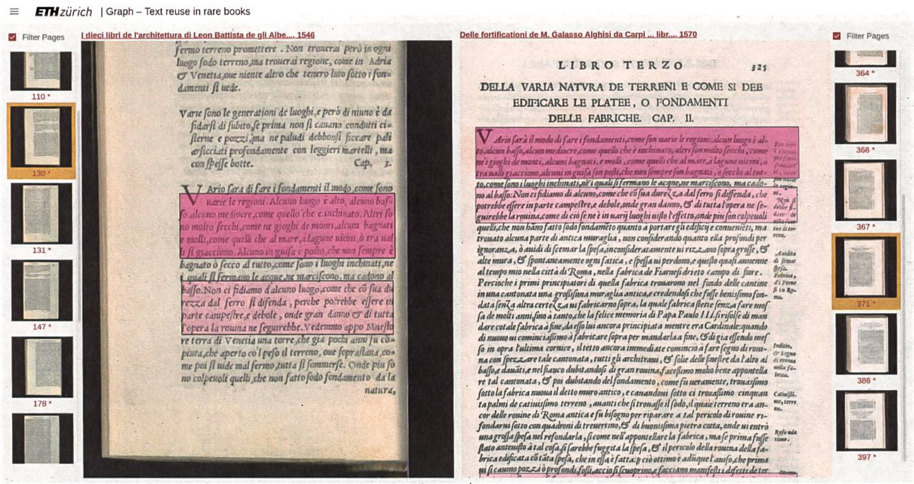
A Complex model propagations identified with the help of computer vision across a large visual corpus. The central female figure evolves to represent six different characters – Leda, Venus, Mary-Magdalen, Lucretia, Cleopatra, and Eve – across the production of nine different artists, 2018. [1] Image: Benoit Seguin

B



C

D



B Automatically identified textual matches between two biographies of Gianlorenzo Bernini, highlighting the difference in narrative ordering of events by the two biographers, 2021. Image: Roberta Berardo and the authors

C Graph Project: Global view of reuses where each dot is a book, and connections are a meaningful amount of text-reuse detected, 2021. Image: authors

D Graph Project: Close reading of the detected matching passages by comparing the two books next to each other directly, 2021. Image: authors type of algorithm. As with textual reproduction, the accuracy of the methods is related to the «level» of the task. Detecting matches at the pixel-level is much easier than determining if a drawing was a preliminary study for a painting. The first task has been possible for 20 years, the second only since recently, thanks to the latest developments in deep learning and significant effort. [2]

To test this approach in architectural history, and explore cross-overs between architectural history and computation, we initiated two complementary pilot projects in the detection of text reuse. Reuse is the occurrence of highly similar passages in two distinct textual artefacts. The first project, Graph, detects and visualizes reuse in a large body of printed books about architecture. The second explores and analyses reuse in two closely related texts, the biographies of the artist Gianlorenzo Bernini published shortly after his death in 1680.

The Graph Project [3] is developed in collaboration with the e-rara collection of the ETH Library. We selected 1019 books pertaining to architecture published before 1850, a corpus where architecture is a specific domain organized around a defined set of topics, often articulated in reference to a single text: the «Ten Books of Architecture». Determining how much each book is reused in another is therefore possible and relevant. By expressing reuse graphically and chronologically, Graph offers an entry point into the reproduction of text within a large body of publications. The relevance of such reuse is up to the user to determine and explore.

The Bernini Project takes the opposite approach. Bernini received two early biographies, the first by Filippo Baldinucci in 1682, [4] the second by Domenico Bernini in 1713. [5] The texts resemble each other to the point that Domenico was long considered a plagiarist. Now we know that both authors based their books on the same sources (now mostly lost), and that Domenico wanted to improve on Baldinucci. The close resemblance between the two texts is well studied, providing a benchmark for the quality and the validity of the computational analysis. Testing against the benchmark improved the computational tool. It also allowed to explore what computational analysis could add to manual exploration. After all, the corpus is within the remit of a single researcher. What are the advantages of computation?

Visual representation of reuse provides a synthetic and accessible view of the relationship between the two texts. This allows for quantification: determining how much the two texts overlap. By breaking the overlap down according to various criteria we can hypothesize about which material the authors shared and what they added individually. Moreover, computational detection catches all instances of resemblance above a certain threshold, not just those a researcher is interested in. This has yielded hitherto unnoticed echoes between the two texts. Finally, resemblances are important moments of divergence or convergence between the two texts. Tagging these passages enhances them semantically, and renders them available for further analysis and connection with data.

These projects suggest that applying computational methods on architectural historical sources offers fresh perspectives on the matter of reproduction. It allows to trace how ideas travelled and transformed across hitherto inaccessible amounts of material. The new scale of this enquiry allows to transcend the limits of canonical collections, and raises questions about the role of meaning of reproduction itself, challenging notions of originality, authorship, and uniqueness.
In this sense, computation contributes to historical and critical enquiry not just as a tool, but as a method.

Maarten Delbeke, born 1970, studied architecture at Ghent University. After completing his PhD in 2001 at Ghent University he was a fellow in Oxford and taught at the universities of Ghent and Leiden. Since 2017 he has held a professorship in the History and Theory of Architecture at ETH Zurich.

Benoit Seguin, born 1988, studied Computer Science at Ecole Polytechnique in Paris. After completing his PhD at EPFL in 2018, where he applied Machine Learning to track visual transmission in large iconographical collections, he accepted a position as a lecturer at ETH Zurich and works as an independent expert on applying Artificial Intelligence to Cultural Heritage data.