Visual Patterns Discovery in Large Databases of Paintings

Isabella di Lenardo, Benoit Seguin, Frédéric Kaplan

Isabella.dilenardo@epfl.ch, benoit.seguin@epfl.ch, frederic.kaplan@epfl.ch

DHLAB – Ecole Polytechnique Fédérale de Lausanne (EPFL)

CDH, INN 116, Station 14, CH-1015 Lausanne, Switzerland

The digitization of large databases of works of arts photographs opens new avenue for research in art history. For instance, collecting and analyzing painting representations beyond the relatively small number of commonly accessible works was previously extremely challenging. In the coming years, researchers are likely to have an easier access not only to representations of paintings from museums archives but also from private collections, fine arts auction houses, art historian photo collections, etc. However, the access to large online database is in itself not sufficient. There is a need for efficient search engines, capable of searching painting representations not only on the basis of textual metadata but also directly through visual queries. In this paper we explore how convolutional neural network descriptors can be used in combination with algebraic queries to express powerful search queries in the context of art history research.

Context and Method

This project is part of project called *Replica*, conducted in collaboration with the Cini Foundation in Venice. This project is based on two parallel developments, the digitization of the Cini Foundation's photo library, a collection of about a million photographs of paintings, engravings, plastic arts and architecture (1300 - 1900) and the creation of dedicated search engine allowing for searches for visual patterns in this database. As the digitization is currently ongoing, the results reported in this paper are performed on a subset of only 39 000 paintings. However, the progressive densification of the database, including a large set of so-called minor painters, should, in the coming months, unfold the full discovery potential of this search engine.

The field of visual pattern recognition has been recently transformed by the surprising performances of so-called deep learning approaches using Convolutional Neural Networks (CNN). CNN are multi-layers architectures used for supervised learning, especially for object classification. Each layer is representing an operation on the previous layer: convolution layer, fully-connected layer, pooling layer, regularization layer, etc. These networks have many parameters (filter parameters, fully-connected weights) that can be learned via backpropagation in a supervised manner. Traditionally, the input (first) layer is a full raster image and the output (last) layer is a vector representing the score of the input image for each class. By showing the network some labeled images and comparing the network's output to the desired label, one can update the parameters of the model.

The theory of deep neural methods have been known for decades, and were already successfully applied with the first convolutional neural networks in the 90s to digit recognition (LeCun et al 1989). However, their computational complexity and the necessity of important amount of training data have seen them being ignored for a long time. With very large datasets available like ImageNet, and GPU computation being more accessible, there has been a sudden surge of interest in deep methods (Deng et al 2009).

In 2012, a convolutional neural network shattered the competition in a difficult 1000 class object recognition challenge, attaining the impressive result of a top–5 error of 15.3% compared to 26.3% for the runner-up (Krizhevsky et al 2012). Ultimately, this work had an important impact on the machine vision community starting the so-called deep learning revolution. A clear manifestation of this trend was that just a year later at the next iteration of this object recognition challenge, almost every entry was based on CNNs as well.

Despite being trained to recognize a precise set of classes. It has been observed that some of the learned parameters of the CNNs will most likely be similar across different datasets. For instance, the first convolutional layer usually learns various edge detectors and basic filters. Some researchers have evaluated the representative power of CNN trained for a specific task to other problems. The model used is one that outperformed the others in 2012 on the *ImageNet* data. Results seemed extremely promising, suggesting that task transfer is possible (Donahe et al 2014)

To calculate the descriptors of our search engine use a pre-trained Convolutional Neural Networks similar to the one described in (Donahe et al 2014). Each painting of our database is associated with 1000 features, corresponding to the last convolutional layer of a trained Convolutional Neural Network. These features are thought to represent high-level characteristics directly usable for the classification tasks. Through this process each painting is associated with a single point in a high dimensional space. When a single image query is sent to the search engine, the results are simply shown ranked by their distance to the query.

However, similarity between paintings could not be the results of single homogenous distance. To enable the users to specify the kind of similarity they want to explore, a more refined language has been introduced. Searches take the form of *algebraic formulas* in which the user can add or subtract examples. For performing such search, we use a binary support vector machine (kernel Radial Basis Function). In the cases were no negative examples are provided a one-class-support vector machine is used (in the case of a query with a single image, this corresponds to the nearest neighbor algorithm we discussed in the previous section). The rest of the paper shows examples of such algebraic queries and the corresponding results.

Examples of queries and results

The classic principles for classifying visual similarities in art history include various dimensions like recurrence of particular pictorial patterns or common compositional structures. As a query illustration, the first criteria of classification chosen is the search for common 'dominant and multiple pictorial motif' in the composition. One classic example in this typology is the Still life, featuring for instance only a large bouquet of flowers. The development of this subject has a long history, from the late sixteenth century, before arriving at its codification during the seventeenth-century Spanish by painter Juan de Arellano. The results of a query with his *Still life of flowers* (Fig. 1) include other famous interpreters of the genre, almost identical in composition and also very close together in chronological and pictorial influences: de la Corte, Snyders, Casteels. However, there are also seventeenth-century painters, Gentileschi, Régnier, Bonito, Vouet, who, while not painting Still lifes, are characterized by the same tonality of *chiaroscuro* typical of a precise moment in history of art. Without further information the similarities found by a single image query include various families of resemblances combining pictorial patterns and color tones.

To focus only on the pictorial motif of flowers excluding any paintings with figures, we subtract one of the paintings by Gentileschi to the initial query (Fig. 2). We obtain all the 'key painters' in this genre including for instance Daniel Seghers. He does not paint a real Still life but flowers around a sacred figure, the Virgin, one of the first subjects, probably invented by Jan Brueghel the Elder. It is probably from this initial subject that evolves the *Still Lifes with flowers*. So in this case, the algebraic query recovered the evolution of a specific pictorial motif with its significant variations during the Seventeenth century.

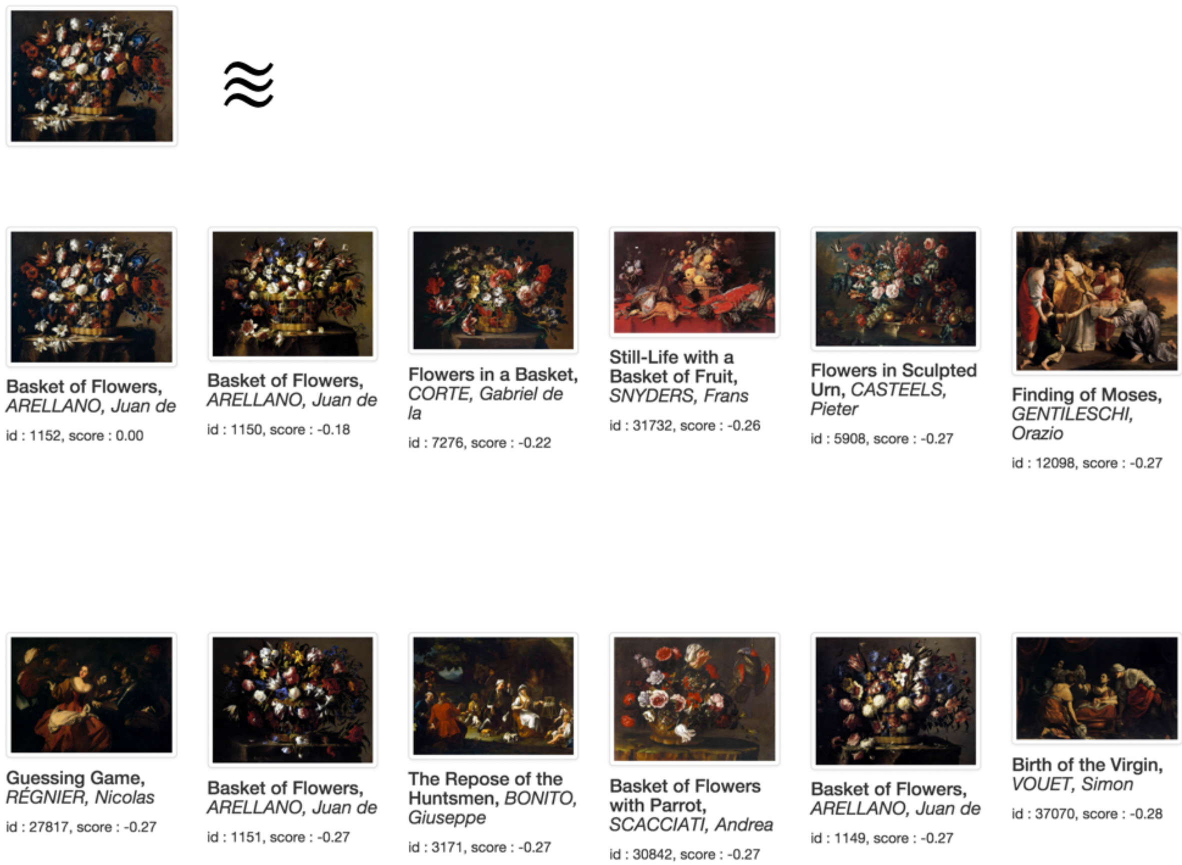

Figure 1 : A query of a Still Life of Flowers by Juan de Arellano returns several paintings with flowers but also other subjects.

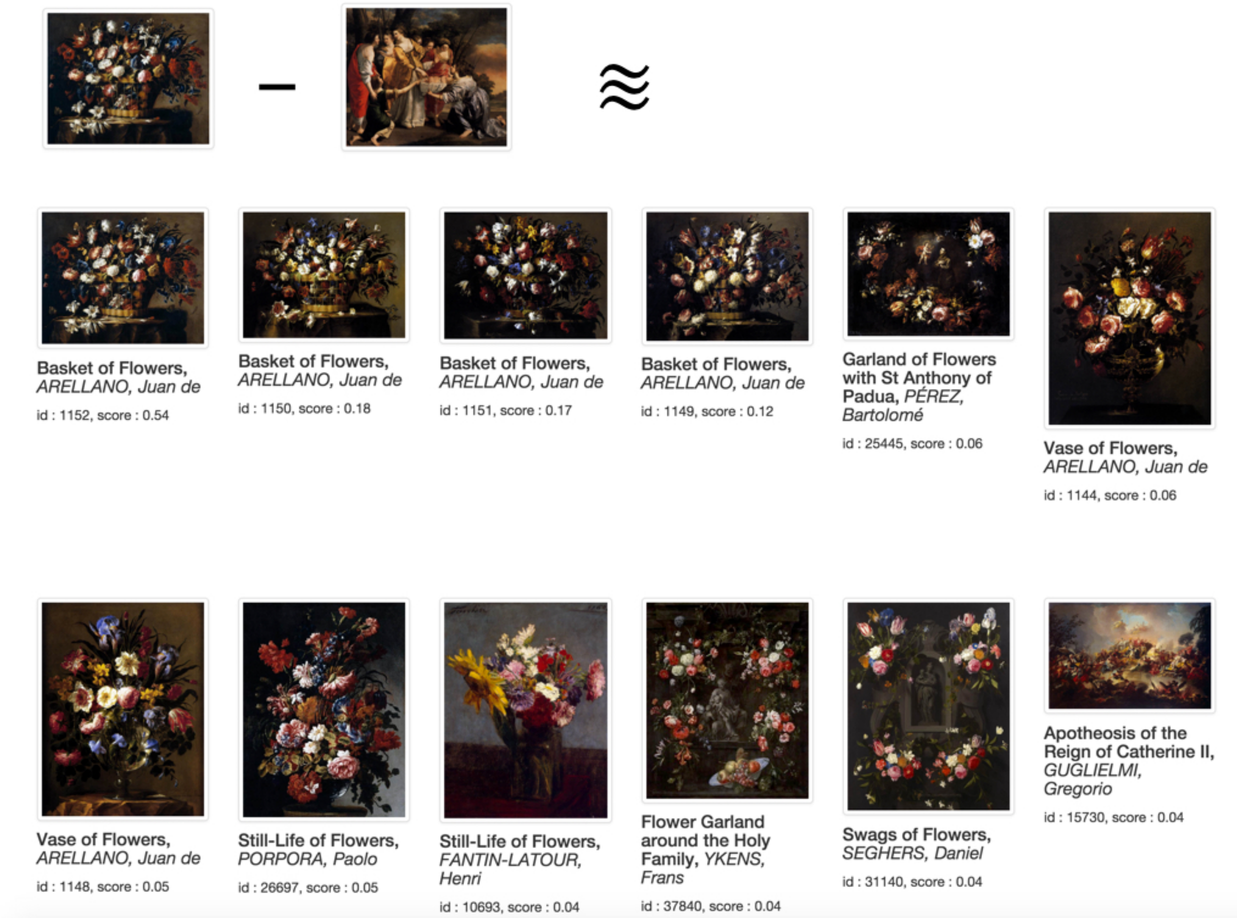

Figure 2 : By subtracting to the flower painting the Finding of Moses by Orazio Gentileschi, only painting featuring flower are returned.

Another criteria of classification in Art History are structural analogies between compositions (Gombrich 1960). Structure of the composition is understood here in a geometric sense, with the reoccurrence of similar geometrical patterns in various paintings. Such kind of structural similarity is one of the classic dimension in the formal analysis of paintings (Focillon 1964; Didi-Huberman 1996)..

So we choose as a second example, *The Gallery of Archduke Leopold* painted by David Teniers the younger, in 1639. This painting is also known to have four different variants. The painting is the reference of a long tradition of paintings subjects featuring cabinet of art lovers and collectors. The contextual meanings of showing famous collections, galleries of great connoisseurs of art have been well studied (Findlen 1996; van der Veen 1993), and are beyond the scope of this paper. A single query returns variations of the same painting (Staatsgalerie Schleissheim, Münich; Prado, Madrid; Kunsthistorisches Museum, Wien), but also painting featuring squares within squares (Fig. 3). For instance on *The Ambassadors depart by Vittore Carpaccio and Baptism of St Libertus by Colijn de Coter,* squares are on the floor or on the wall. To refine further our search and find the “good neighborliness” (Warnke 2000; Freedberg 1989) of David Teniers structure , we can try to exclude these two examples by subtracting them to the initial query. Results of such a query exclude now interior scenes feature geometrical square but now include various scenes of the passion organized as sequences of “squares” (Fig 4). In a third attempt, we can now exclude those by substracting them to the initial query and reinforce the focus on search by adding variant of the first painting by Teniers . Indeed, all the first results feature now paintings with galleries of collectors with examples of the most important authors of the genre, thus facilitating the study of their mutual influence (Fig 5).

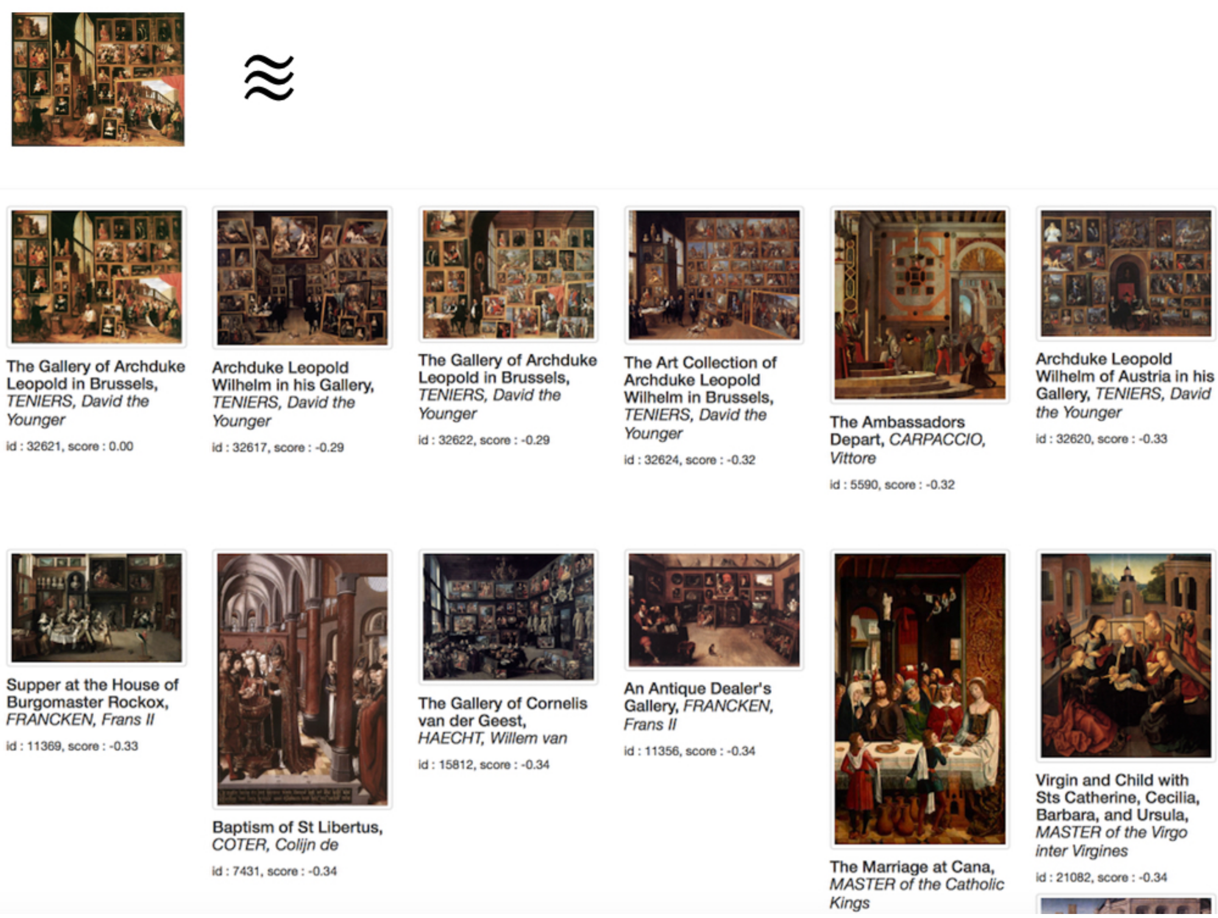

Figure 3 : *A query with the The Gallery of Archduke Leopold by David Teniers the younger (1639) gives a first results four variants of the same painting by the same author. The following results include various kind of painting which same some similarities with the initial query but are not representing the same subject.*

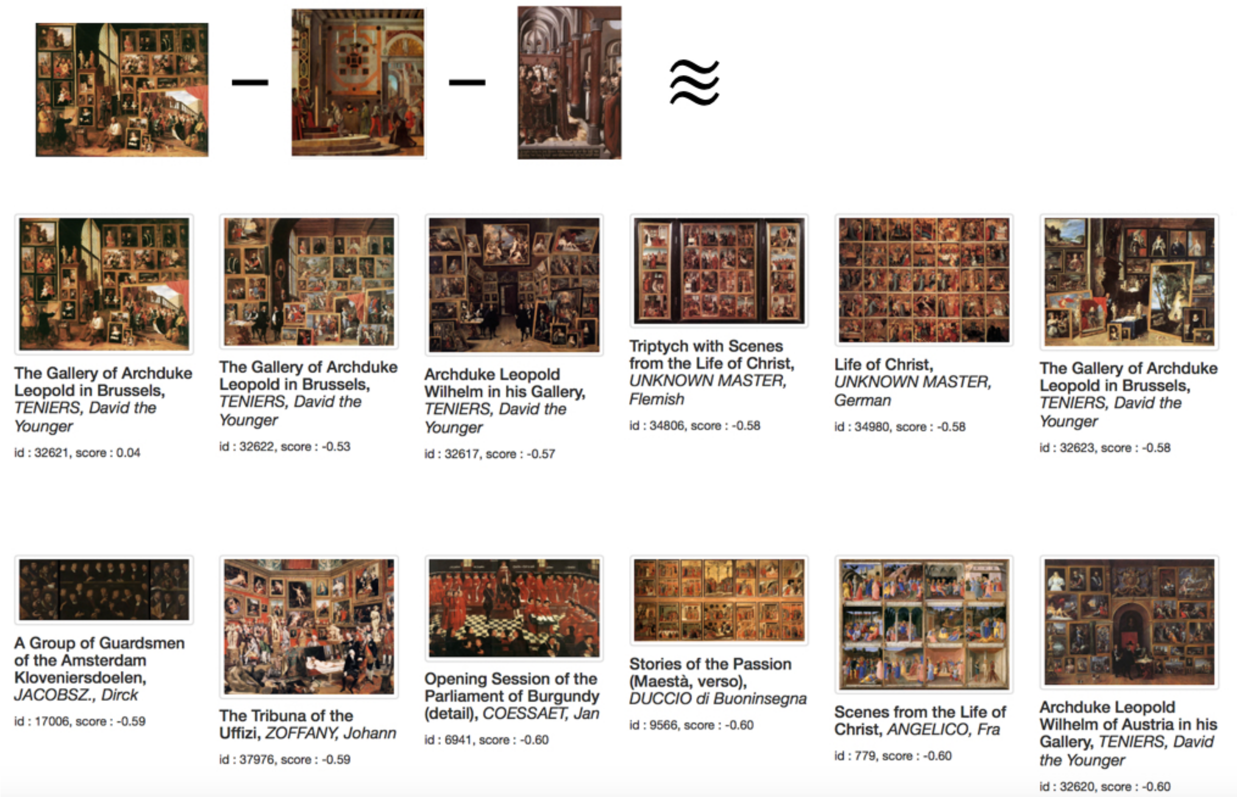

Figure 4 : *When The Gallery of Archduke Leopold is subtracted with The Ambassadors depart by Vittore Carpaccio and Baptism of St Libertus by Colijn de Coter a series of paintings only containing hierarchy of embedded squares are returned. The formula has isolated a specific characteristic in the feature space when the presence of a multiple squares is the most specific trait.*

Figure 5 : In order to search specifically for the paintings containing paintings, the two painting representing the stories of the passion can be subtracted. Most results now feature paintings in which paintings are present.

Perspectives

Pattern recognition methods have been extremely impressive progresses in the recent years, thanks to the progress of convolutional neural network and advent of very large databases of images. As Art History deals with the study of the migration of patterns, there is surely a great opportunity of designing new tools to search through large databases of paintings photographs. This paper is a first examination of the typologies of search use cases that could be envisioned combining convolutional neural network features and simple algebraic formulas. This initial study illustrates that this new way of expressing queries allows for the incremental definition of various kinds of "similarities" between paintings. Convolutional neural networks features manage to capture many dimensions of similarity between paintings, including composition, colors and also common iconographic elements. Combined with a simple language for expressing the specificity of the traits that the user looks for, it could enable new powerful search tools that may in turn have important impact on history of art studies.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A large-scale hierarchical image database, 2009 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2?9, 2009.

G. Didi-Huberman, *Pour une antropologie des singularités formelles. Remarque sur l'invention warburgienne*, in "Genèses", 24, 1996, pp. 145-163.

J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, Int. Conf. Mach. Learn., pp. 647–655, 2014.

P. Findlen,. *Possessing Nature - Museums, Collecting & Scientific Culture in Early Modern Italy*. Reprint. Berkeley: University of California Press, 1996

H. Focillon, *Vie des formes: suivie de l'"Eloge de la main",* Paris 1964

D. Freedberg, *The Power of Images: Studies in the History and Theory of Response*, Chicago-London 1989

E. H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation*, London 1960

A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst., pp. 1097–1105, 2012.

Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, pp. 541–551, 1989.

J. van der Veen, *Galerij en Kabinet, vorst en burger. Schilderijen Collectie in de Nederlanden, in Verzamelen van Rariteitenkabinet tot Kunstmuseum,* E. Bergvelt, D. J. Meijers, M. Rijnders eds., Heerlen 1993, pp. 145-164

Der Bilderatlas. Mnemosyne, M. Warnke eds., Akademie, Berlin 2000