Image and Vision Computing 77 (2018) 21–32

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Can we teach computers to understand art? Domain adaptation for enhancing deep networks capacity to de-abstract art☆

Mihai Badea, Corneliu Florea*, Laura Florea, Constantin Vertan

Image Processing and Analysis Laboratory (LAPI), University Politehnica of Bucharest, Romania

A R T I C L E   I N F O

Article history:
Received 11 December 2017
Accepted 21 June 2018
Available online 4 July 2018

Keywords:
Convolutional Neural Networks
Domain adaptation
Genre recognition
Painting analysis
Style transfer

A B S T R A C T

Humans comprehend a natural scene at a single glance; painters and other visual artists, through their abstract representations, stressed this capacity to the limit. The performance of computer vision solutions matched that of humans in many problems of visual recognition. In this paper we address the problem of recognizing the genre (subject) in digitized paintings using Convolutional Neural Networks (CNN) as part of the more general dealing with abstract and/or artistic representation of scenes. Initially we establish the state of the art performance by training a CNN from scratch. In the next level of evaluation, we identify aspects that hinder the CNNs' recognition, such as artistic abstraction. Further, we test various domain adaptation methods that could enhance the subject recognition capabilities of the CNNs. The evaluation is performed on a database of 80,000 annotated digitized paintings, which is tentatively extended with artistic photographs, either original or stylized, in order to emulate artistic representations. Surprisingly, the most efficient domain adaptation is not the neural style transfer. Finally, the paper provides an experiment-based assessment of the abstraction level that CNNs are able to achieve.

1. Introduction

This paper aims to investigate the differences between the level of abstraction achieved by deep convolutional neural networks as compared to the human performance in the context of painting analysis. To synthesize the motivation, let us recall Pablo Picasso's words: "There is no abstract art. You must always start with something. Afterward you can remove all traces of reality". Art historians and enthusiasts are able to note, while recalling major artistic works through the history, that the level of abstraction steadily increased.

In parallel, in the last period, works that use computer vision techniques to analyze visual art increased with respect to both the quantity and the quality of reported results. Two trends favored these developments. First, there were consistent efforts to digitize more and more paintings, such that modern systems may learn from large databases. Two of such popular efforts are Your Paintings (now Art UK[1]) which contains more than 200,000 paintings tightly connected with historical British culture and WikiArt[2] which contains around 100,000 paintings gathered from multiple national cultures. The databases come with multiple annotations. For this work, we are particulary interested in annotations dealing with the painting's subject or scene type. From this point of view, a more complete database is the WikiArt collection, where the labelling category is named *genre*. The second trend is purely technical and it deals with the development of the Deep Neural Networks, that allowed classification performances that were not imagined before. In this work, we will use the more popular Convolutional Neural Networks (CNN) to recognize the painting genre.

Let us now establish the meaning of "genre", its relation with the scene and with the image subject. A list of definitions for various paintings genres is presented in Table 1. To label a painting into a specific genre, in most of the cases, a user has to identify the subject of that painting. The exceptions are "Abstract Art", "Design", "Illustration" and "Sketch and Study", where the main characteristic is related to the depiction mode. In this majority of cases, the subject is related to the scene represented in the work of art. The term "genre" is typical for art domain, and is a more general, including, concept than mere "subject" or "scene type". In this work, while referring to paintings, we will use all three with the same meaning of "genre". In comparison, for a non-artistic photograph, as there is no artistic intervention in the depiction mode, the subject is more related to the scene, while the genre is hard to be defined. For artistic photos, the "genre" gets meaning again.

☆ This paper has been recommended for acceptance by Yannis Panagakis.

* Corresponding author.
E-mail addresses: mbadea@imag.pub.ro (M. Badea), corneliu.florea@upb.ro (C. Florea), laura.florea@upb.ro (L. Florea), costantin.vertan@upb.ro (C. Vertan).

[1] http://artuk.org/.
[2] http://www.wikiart.org/.

https://doi.org/10.1016/j.imavis.2018.06.009
0262-8856/

Table 1
Overview of the genre organization of the WikiArt database and the explanation of the label meaning. We have marked only genres with more than 200 images.

	Genre	No. imgs.	Description
1	Abstract Art	7201	Uses shapes, forms etc. to replace accurate representations
2	Allegorical Painting	809	Expression of complex ideas using another subject
3	Animal Painting	1233	Paintings which depict animals
4	Battle Painting	273	The main subjects are battles and wars
5	Cityscape	4089	Works which contain cities or other large communities
6	Design	1577	Conceptual schemes of objects and structures
7	Figurative	1782	Forms inspired by objective sources, but altered
8	Flower Painting	1270	Paintings of flowers
9	Genre Painting	10,984	Scenes of everyday life
10	History Painting	656	Depictions of historical events
11	Illustration	2493	Visual representations usually meant for books, magazines, etc.
12	Interior	511	Paintings depicting interiors of structures
13	Landscape	11,548	Contains representations of land or other natural scenes
14	Literary Painting	418	Subject taken from literary work
15	Marina	1385	These paintings show scenes from docks or ports
16	Mythological Painting	1493	Inspired by mythology
17	Nude Painting	1758	Paintings which contain nudes
18	Portrait	12,926	Images of real individuals
19	Poster	229	Works which are usually intended for advertising
20	Religious Painting	5703	Inspiration is drawn from religious scenes
21	Self-Portrait	1199	The subject of the painting is the artist
22	Sketch and Study	2778	Drawings done for personal study or practice
23	Still Life	2464	Images which depict inanimate objects
24	Symbolic Painting	1959	Content suggested by symbols in the forms, lines, shapes and colors
25	Wildlife Painting	259	Paintings of natural scenes, including animals in their habitats

Starting from the idea that Deep Neural Networks share similarities with the human vision [1] and the fact that such networks are already proven to be efficient in other perception-inspired areas, like object recognition or even in creating artistic images, we ask ourselves if they can pass the abstraction limit of artistic paintings and correctly recognize the scene type of such a work.

In this paper, we will first work with Residual Network (ResNet) on the standard WikiArt database so to obtain state of the art results. Afterwards, we will test different domain transfer augmentations to see if they can increase the recognition rate; also we will study if the network is capable to pass the abstraction limit and learn from different types of images that contain the same type of scenes. Furthermore, we introduce several alternatives for domain transfer to achieve a dual-task: improve the scene recognition performance and understand the abstraction capabilities of machine learning systems.

Regarding deep networks, multiple improvements have been proposed. In many situations, if the given task database is small, better performance is reachable if the network parameters are previously trained for a different task on a large database, such as ImageNet. Next, these values are updated to the given task. This is called fine-tuning and it is a case of transfer learning. As our investigation is related to a different domain transfer, we will avoid to use both of them simultaneously, in order to establish clearer conclusions. To compensate, we are relying on the recent architecture of Residual Networks (Resnet [2]) that was shown to be able to overcome the problem of vanishing gradients, reaching better accuracy for the same number of parameters, when compared to previous architectures.

1.1. Contribution and paper organization

This paper extends our previous works [3, 4], being mostly developed from Ref. [3], where we had initiated the discussion about the efficiency of various methods to transfer information from the photographic domain to the paintings domain, such that the recognition by CNNs of paintings genre is improved. In this paper we significantly extend the discussion, by including other transfer methods and by adding more significant results that allow crisper conclusions. In the second work ([4]), we showed that the artistic style transfer remains as efficient even if a reduced number of iterations are performed while over–imposing the style of an artistic painting and the content from a photograph onto a new image, according to the neural style transfer introduced by Gatys et al. [5].

Overall, this paper claims several contributions along a number of directions. On one direction, we investigate which aspects, comprehensible by humans, hinder the CNNs while understanding a painting genre; subsequently by means of domain transfer, we retrieve information about the internal description and the organization of the painting clusters. In order to accomplish such a task, we annotate artistic photographic images with respect to the scene type related to genres and we stylize a large corpus of photographs using different style transfer methods. All this data will be made publicly available to be used in other research works.

On a second direction, this paper is the first to objectively evaluate the efficiency of the currently popular neural style transfer methods. Currently existing solutions [5-8] compare themselves by speed, stability within video sequences or number of transferable styles. By quantifying the improvement while adapting photographs to the painting domain, we reach a surprising conclusion, namely that they are less or at most as efficient as non-neural style transfers solutions. Evermore, a CNN finds as informative the original photographs without any style transfer applied.

The remainder of the paper is organized as follows: Section 2 presents previous relevant works, Section 3 summarizes the CNN choices made and Section 4 will discuss different aspects of painting understanding. Section 5 presents the used databases, while implementation details and results are presented in Section 6. The paper ends with discussions about the impact of the results.

2. Related work

This work investigates the capabilities of CNNs to recognize the subject of paintings as compared with the performance of humans. Thus, relevant prior work refers to solutions for object and scene recognition in paintings. As paintings are an abstraction of real images, scene recognition in photographs is also relevant. At last, we aim to adapt information from photographs to paintings by means of style transfer.