AI Authentication Report by Art Recognition

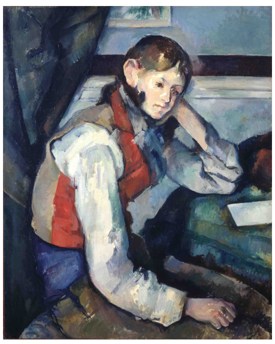
AI authenticity evaluation of the artwork

«Boy in a Red Vest»

assumed to be by

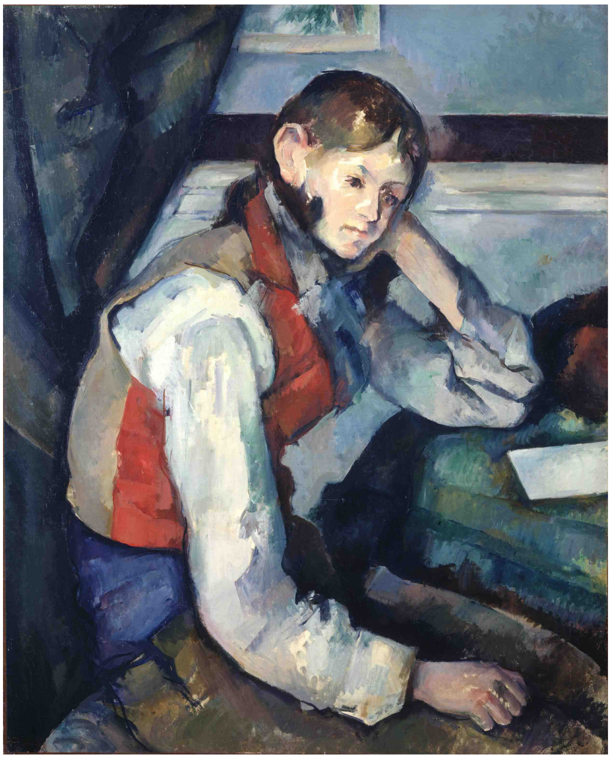
Paul Cézanne

Issued for the website on July 21, 2023



Digitally signed by Art Recognition AG
Date: 2023.07.25 09:41:05 +02'00'

ART RECOGNITION



Artwork information

Attributed to	Paul Cézanne
Title	Boy in a Red Vest
Medium	Oil on Canvas
Physical size	80cm x 64.5cm
Image size	2500 x 3097 px
Training dataset	235 images of authentic artworks, 258 images of non-authentic artworks

Analysis outcome

The Art Recognition AI System assesses «Boy in a Red Vest» to be an **authentic artwork** by Paul Cézanne with a probability of **89.58%**.

1 Method description

This report presents the results of an Artificial Intelligence analysis of the painting «Boy in a Red Vest» attributed to French artist Paul Cézanne (1839 - 1906).
Paul Cézanne is an exceptionally complex artist, whose style roughly spans over four periods: Dark period (c. 1861-1870), Impressionist period (c. 1870 - 1880), Mature or Constructive period (1887-1890), and Final period (1890-1906). The painting «Boy in a Red Vest» has been attributed to the mature period, which distinguishes itself through several characteristics: the importance of form and surface, as well as composition; creating forms from color and brushstroke, not so much from outline; and looking for basic geometric patterns.
The purpose of this analysis is to determine whether the artist's main characteristics, learned by the AI algorithm from a dataset of authentic paintings by Paul Cézanne, match the characteristics identified on the painting in question. Below we describe the main steps of the process:

1. **Data collection.** We compiled a high quality image dataset containing photographic reproductions of verified authentic artworks created by Paul Cézanne. To improve the algorithm's discrimination capabilities, we also fed into the system a set of paintings which are not by Paul Cézanne.

2. **Training the algorithm.** From the training images, the AI is able to capture fine details such as the artist's brushstroke, edges, shapes, variations in colour; high-level composition elements (repetition of certain motifs, prevalent locations of certain figures and objects, overall proportions); as well as other features which are distinctive to the analysed artist.

3. **Testing the algorithm's performance.** The success of the training is measured by two performance metrics: the confusion matrix and the accuracy. These metrics are evaluated on a set of images that were not used during training ("test set").

4. **Assessing the authenticity of the artwork in question.** Following the successful training, the AI compares the features learned from the training images with those identified on the painting «Boy in a Red Vest». Based on this comparison, the AI calculates the probability for the authenticity of the analysed artwork.

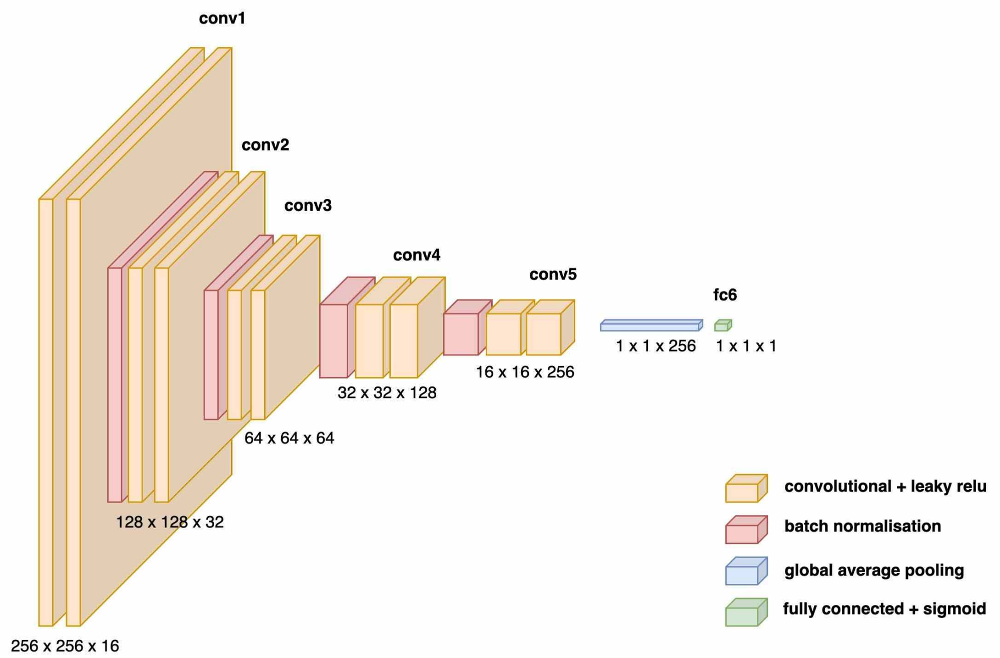
5. **Visualising the results.** On the analysed painting, we superpose a heat map which depicts the areas which most strongly support the algorithm's decision. Furthermore, we generate a brushstroke boundary extraction map showing the brushstrokes learned by the algorithm.

2 Artificial neural network architectures

We use two types of artificial neural networks to learn an artist's main characteristics, namely a Convolutional Neural Network (CNN) which incorporates a module for brushstroke extraction, and Vision Transformer using Shifted Windows (SWIN). Each of these state-of-the-art architectures has been adapted to our art-classification task, where artworks are classified into two categories: "authentic" and "non-authentic".

The networks employed in this analysis are described in detail in Ref.[1]. Our CNN architectures are based on previous studies [2, 3] which demonstrated the remarkable performance of Convolutional Neural Networks (CNNs) in art-classification tasks when trained on extensive digitized art collections. Transformers, initially developed for natural language processing and demonstrated through the remarkable achievements of ChatGPT [4], represent a new class of deep learning architectures that have gained widespread popularity in the computer vision field [5]. Notably, Vision Transformers have showcased exceptional capabilities in image classification. Art Recognition has been at the forefront of pioneering the adaptation of Vision transformers to the realm of art authentication tasks.

It is also worth mentioning that the literature has addressed the machine's role as a novel form of art expert for attributions and authentications [6, 7]. These discussions have underscored the resemblances between human and machine connoisseur approaches, particularly in terms of their knowledge regarding numerous works by the same artist and related pieces.



The figure above illustrates the structure of a convolutional neural network architecture, showing the convolutional layers (*conv*), their sizes (*numbers underneath*), and the last, fully connected layer which calculates the probability of an artwork to be authentic.

Each artwork from the training set is passed through the layers of the artificial neural network from left to right, whereby with each layer the identified features are being consolidated. At the very last layer, the class "authentic" or "non-authentic" is assigned to the artwork. If the classification is correct, the features are remembered, else the artwork is being processed again and different features are being learned. This is done multiple times for all the artworks from the training set, until the neural network has learned the artist's characteristic features.

3 Data collection and pre-processing

We gathered a total of 235 images of authentic paintings and 258 images of non-authentic paintings to train, optimise and validate our model. The non-authentic dataset includes images of known forgeries, artworks by followers or artists of similar style(s) and chronology, as well as synthetic images created by a separate algorithm. In detail, our training dataset contains:

- 235 images of original artworks by Paul Cézanne;
- 29 images of forgeries and copies;
- 216 images of imitations by followers and students, or of artworks in the style of the author;
- 13 synthetic images in the syle of Paul Cézanne.

Since the paintings exhibit high variability in terms of aspect ratios, to increase precision we employ a particular pre-processing strategy which allows us to capture both fine details and coarse structures. To capture finer details, we split the images into patches which are further cropped into squares and shuffled before being fed into the training. The number of created patches depends on the image resolution: we generate 20 patches from images with a resolution greater than 1024×1024 pixels; 4 patches from images with a resolution smaller than 1024×1024 pixels, but greater than 512×512 pixels; whereas images with a resolution smaller than 512×512 pixels are not patched. We also keep the full (undivided) images to capture the high-level structural elements.

2 x 2 grid (left side) and 4 × 4 grid (right side) of patches created from the painting "Boy in a Red Vest". The center cropped squared patches are highlighted in bright.

The strategy described above allowed us to augment the dataset, such that our final, balanced dataset contains a total of 2858 patches stemming from authentic images and 3104 patches from images which are not authentic. All images have been converted to the lossless Portable Network Graphics (PNG) format. The dataset generated as described above forms an excellent basis for training the AI.

4 Training and validating the algorithm

A neural network has to first learn the artist's discriminative features from the training data in order to be able to subsequently identify them on a new, previously unseen artwork. During the learning phase, the network extracts from every training image characteristic features such as brushstrokes, shapes, colours, boundaries of objects, figures, and compositional elements.
The learning process is completely independent, in the sense that we do not introduce any hand-crafted features. Aside from dividing the images into patches, we optimised a set of so-called hyperparameters to reach the most accurate training. The hyperparameters specify different configurations of the neural network, also known as "models".
After completing the training, we test the robustness of the trained neural network. Following a standard protocol, we randomly split the images into training, validation and testing data, preserving 80% of images for training, 10% for validation, and 10% for testing. No painting in the training set appears in the validation/testing sets and vice-versa. Furthermore, all patches of each painting are always assigned to the same partition, which implies that the test set always consists of patches that were not part of the training or validation.

The performance of a trained neural network is measured on a carefully crafted testing set which contains artworks not used during the training, including high quality forgeries. On the testing set we evaluate two metrics, the confusion matrix and accuracy. Based on their values, we select the best model, i.e., which is able to classify correctly the highest number of images in the testing set; this model will be used to calculate the class probability for the artwork in question.

Confusion matrix on the testing set. Since our neural networks are trained to recognise two classes (authentic/non-authentic), we use a binary confusion matrix with the following structure:

True Positives	False Negatives
False Positives	True Negatives

For our application, the four squares can be interpreted as follows:

- True Positives: authentic paintings correctly predicted to be authentic;
- False Positives: authentic paintings incorrectly predicted to be non-authentic;
- False Negatives: non-authentic paintings incorrectly predicted to be authentic;
- True Negatives: non-authentic paintings correctly predicted to be non-authentic.

The table below shows the confusion matrix on the testing set, for the best model. We note that the reported values also take into account the correctly and incorrectly classified patches, not just the entire artworks. It is in principle possible to wrongly classify one patch belonging to an artwork from the testing set, but correctly classify the rest of the patches as well as the entire artwork.

True Positives: 281	False Negatives: 14
False Positives: 17	True Negatives: 254

Accuracy on the testing set. The accuracy is defined as the fraction of correct predictions (of authentic and non-authentic images) over the total predictions. The best model's accuracy on the test set is 96.8 %, indicating an excellent model performance.

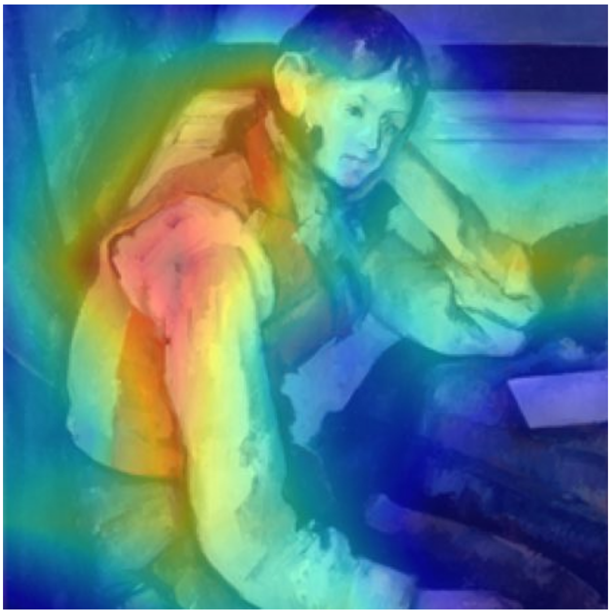
5 Analysing the submitted image

Following the successful training, we passed the image of the painting titled «Boy in a Red Vest» through the best performing model. The algorithm compared the artist's features learned from the training images with those on the image in question. Based on this comparison, the algorithm returned a class probability of 89.58% for a positive response ("authentic"), whereby this probability has been calculated as the median of the probablities of all individual patches.

This result is a strong indicator of a match between the features learned from the training images, and those on the evaluated artwork.

6 Visualising the results

In this section we present two visualisation charts which are meant to give more insight into the analysis: a heat map and a brushstroke identification map.
The interpretation of the heatmap presented below is as follows: the most important areas for the algorithm's decision are those highlighted in red colors, and the importance gradually decreases approaching yellow colors; at the other end of the spectrum, the regions which are colored in blue have the smallest influence on the decision. Therefore, the hotspot areas (colored most intensely in red) have the highest importance for the final decision.



Heatmap providing visual evidence on the algorithm's evaluation. Red areas indicate regions of highest importance, while blue areas are of least importance.

The hotspots tend to appear in the regions which comprise more structure (corners, edges, shapes etc) or are important for the overall composition. But, they can also appear on flat surfaces where the brushstroke changes direction. An important observation is that these regions stem from the analysis of the readable brushstroke and other structural characteristics, and are not related to the artistic representation. Also note that the size and position of the regions marked in red are not directly linked to the final outcome ("authentic"/"not authentic").
On the painting in question, the areas which most strongly support the algorithm's decision ("authentic") are the elongated arm, the right shoulder, and the chair behind the table.
Moving on to the brushstroke representation, we note that brushstrokes are particularly important features as they are a unique signature of the artists' style. Thus, they need to be isolated very carefully. To achieve this, we have implemented a module which enables our neural net-

work to extract the boundaries of brushstrokes from all training images, and explicitly use these features during the training. We stress that brushstrokes are learned alongside other features (colors, shapes, object locations, compositional elements etc.).

The map below depicts the extracted brushstroke contours on the analysed painting:

Sample brushstroke extraction on the artwork in question. We see the lines contouring the brushstrokes as learned by our model. The painting exhibits numerous visible, broad and long strokes.

7 Conclusion

We compiled a solid training dataset not only quantity-wise, but also in terms of its quality (235 original artwork images, 258 contrast images). The accuracy of the best trained model on the testing set has reached 96.8 %, implying an excellent performance. The AI returned a class probability of 89.58% for an "authentic" response. The brushstroke visualization map exhibits broad strokes of planes of colors which create characteristically geometric shapes. In particular, the heatmap visualization shows that the elongated arm captures the attention of the Artificial Intelligence, alongside that of the viewer.

References

[1] Ludovica Schaerf, Carina Popovici, and Eric Postma. "Art Authentication with Vision Transformers". In: (2023). arXiv: 2307.03039 [cs.CV]. URL: https://arxiv.org/abs/2307.03039.

[2] Nanne van Noord, Emile Hendriks, and Eric Postma. "Toward discovery of the artist's style: Learning to recognize artists by their artworks". In: *IEEE Signal Processing Magazine* 32.4 (2015), pp. 46–54.

[3] Thomas Dobbs and Zbigniew Ras. “On art authentication and the Rijksmuseum challenge: A residual neural network approach”. In: *Expert Systems with Applications* (2022), p. 116933.

[4] OpenAI. *GPT-4 Technical Report*. Tech. rep. OpenAI, 2023. URL: https://openai.com/research/technical-reports/gpt-4.

[5] Alexey Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: (May 2021). URL: https://openreview.net/forum?id=YicbFdNTTy.

[6] Petra Bell and Frédéric Offert. “Reflections on connoisseurship and computer vision”. In: *Journal of Art Historiography* 24 (2021).

[7] Yijun Zhu et al. “Machine: The new art connoisseur”. In: *arXiv preprint arXiv:1911.10091* (2019).