How to use artificial intelligence to analyse paintings?

Thibaud PICCINALI, 2316 23 2101

August 21, 2023

Abstract

This report presents a study on using artificial intelligence (AI) to analyze paintings, with a focus on identifying the focal points in art pieces. The goal was to develop a deep-learning model that can predict focal points, aiming to understand what makes a painting popular. Two models were created: one from scratch using Convolutional Neural Networks (CNN), and the other using the *TensorFlow Object Detection API*. While the latter model showed promise in recognizing focal points in specific genres like portraits and landscapes, its performance declined when applied to more diverse and general images. The limitations of the model highlight the importance of defining clear research objectives and dataset curation when employing AI in art analysis.

Keywords: Artificial Intelligence, Painting Analysis, Focal Point, Deep Learning, Convolutional Neural Networks, *TensorFlow Object Detection*.

Introduction

Today, artificial intelligence (AI) is everywhere. With the rise of deep-learning, these machines are now capable of performing tasks that typically require human intelligence, such as reasoning, problem-solving, and decision-making. This technology is transforming our industries and changing the way we live and work.

However, if we often see AI being used in technical fields, it is important to say that the AI impact is not limited to these domains. One of the most important is the art world. Indeed, nowadays, some artists use AI algorithms to generate new and unique art pieces, while others use machine learning techniques to analyze and classify existing art.

This desire to understand how art works and what makes a popular painting popular is what is motivating Professor HAGA Hirohide to work on this domain. I, Thibaud PICCINALI, as an international student in Doshisha University, have worked on this notion during six months and under the supervision of Professor HAGA.

This document presents to you the results for these six months of work at Doshisha.

You can find on this GitHub all the code that I developed for this project.

1 My research plan

1.1 Goals

During these six months, my goal was to help Professor HAGA by developing tools which can be used to analyze paintings. I especially studied the notion of "Focal Point", indeed my goal was to develop a program that can predict where the focal point(s) is/are. For this reason, I built some deep-learning models that can be able to recognize it. I believe that these kind of algorithms could help us to understand what makes a popular painting so popular.

1.2 Focal point: concept

The focal point is defined as an area in a picture that attracts the eye of the viewer, it acts as a kind of "eye-magnet" and its power can be influenced by many factors:

The first one is the use of contrast: any type of difference in imagery will result in that element becoming a focal point.

Figure 1: Contrast concept illustration

Placement is one of the most important. Indeed, if an object is placed in the center (or near the center) it will naturally become a focal point.

Figure 2: Centering concept illustration

A third way to create a focal point is to use implied lines to direct the viewer's eye to an object or element. This technique is known as convergence.

Figure 3: Convergence concept illustration

Another way to create a focal point in an artwork is through isolation. Whenever one element is separated from a group it becomes a focal point.

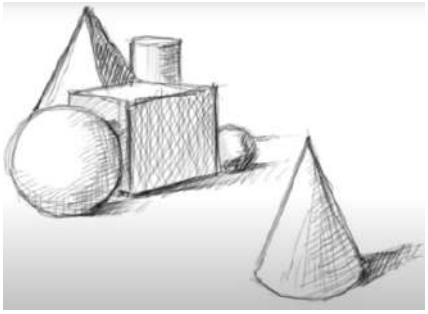

Figure 4: Isolation concept illustration

At last, another way to create a focal point is to introduce an object (or element) that is unusual to the scene.

Figure 5: Unusual concept illustration

2 Presentation of my work

2.1 My first model

2.1.1 Explanation of operations

My idea was to built an object detection model, and train it to recognize focal point(s). At first, I tried to build a model from scratch. To do that, I chose to built my model by implementing an Convolutional Neural Network architecture (CNN).

CNN architecture is a type of artificial intelligence that mimics the way the human brain processes visual information. It is particularly well-suited for tasks like image recognition and computer vision. At its core, a CNN is designed to automatically learn and identify patterns in images. It does this by using multiple layers of specialized filters (convolutional layers) to detect various features like edges, colors, and shapes. These filters learn to recognize simple patterns in the early layers and gradually combine them to identify more complex patterns in deeper layers.

The code of this kind of architecture looks like this (you can see in the code the different layers of filters):

```
model = Sequential([
# First convolutional layer
Conv2D(64, (3, 3), activation='relu', input_shape=(img_height, img_width, 3)),
MaxPooling2D(2, 2),

# Second convolutional layer
Conv2D(128, (3, 3), activation='relu'),
MaxPooling2D(2, 2),

# Flatten layer
Flatten(),

# Fully connected layer
Dense(128, activation='relu'),

# Output layer
Dense(4, activation='linear')
])
```

Furthermore, the network need to be feed with labeled examples of images in order to learn what is a focal point, this is why I created an important dataset (with thousands of different paintings). Each element of the dataset has been labeled by myself: I was in charge of indicating to the model where the focal points of the paintings are situated. I chose to create a diversified dataset (with different painters, styles, eras...) in order to build a strong model.

To help me in this task I used a commonly used software in machine learning named *LabelImg*. *LabelImg* is an open-source graphical image annotation tool used for labeling objects or regions of interest in images. The main purpose of *LabelImg* is to help users annotate images by drawing bounding boxes around objects they want the model to recognize. Thanks to this software, I was able to obtain a *.xml* file, for every painting of my dataset, indicating the coordinates of its focal point.

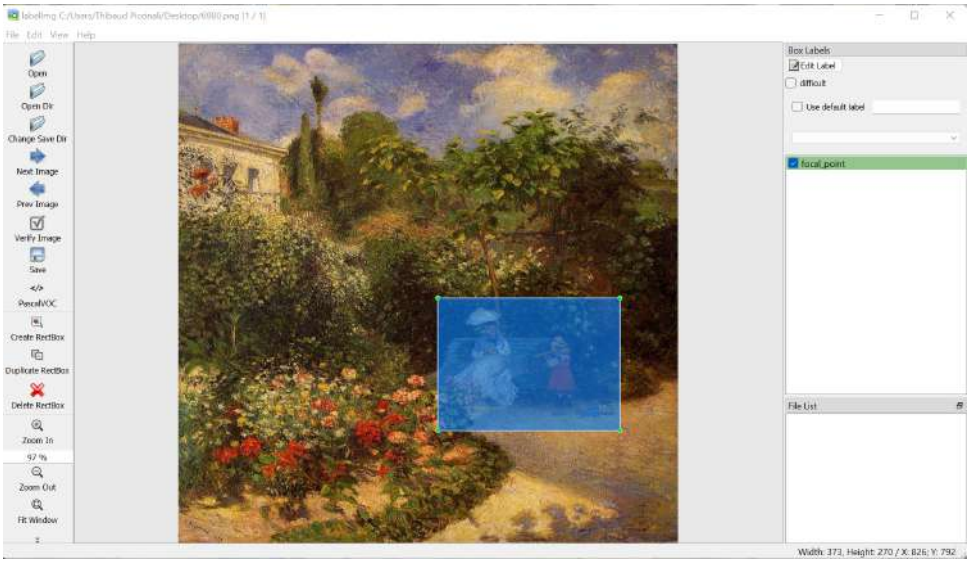

Figure 6: Example of use of the software *LanelImg*

Once all the code has been implemented, the last step was to adapt my network to my dataset in order to have better result. Indeed there are a lot of different parameters (number of layers, activation functions, weights...) that can be changed to optimized the accuracy of the model. For this matter, I developed some programs which automate the prediction process. Thank to them, I managed to find the best parameters for my model to have the best results.

Once all of this parameters has been optimized in order to get the best result possible, I was ready to use the model.

2.1.2 Results

Here is a graph showing the best results of my model that I was able to obtain:

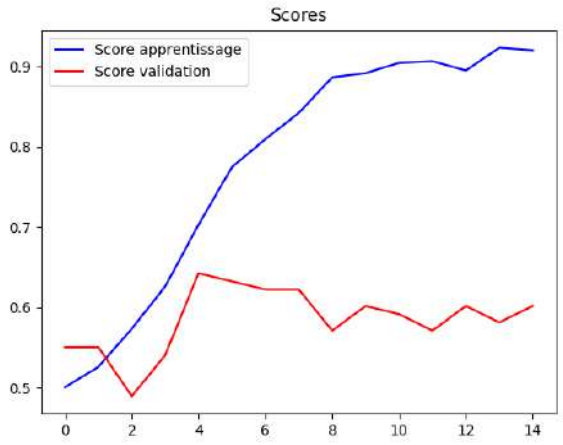

Figure 7: Accuracy of the model

On this diagram, the blue curve shows the accuracy of my model on paintings of my dataset, and the red curve shows the accuracy for new (unknown) paintings. As we can understand with these figures, the prediction of my model for unknown painting seems to be pretty weak. Indeed, if we try to make some tests, we will notice that we have some good but also some bad predictions. Here are some examples (the green box represents the focal point):

Figure 8: Examples of good predictions

Figure 9: Examples of bad predictions

After some studies of what could explain these bad predictions, I thought that my problem could be related to my dataset. Indeed my dataset is composed of many different paintings: there are paintings of landscape, objects, portraits (and many more) in a surrealism, futurism, cubism (and many more) style. The problem is that this important diversity could represents a difficulty for my model to

understand where the focal point is in every situation. To answer this problem, I tried to restrict my model (and its dataset) to a particular kind of paintings. Indeed, I tried to limit my dataset to image like portrait, landscape or buildings, but unfortunately I was not able to obtain better result that the ones that I previously obtained. Faced to this problem, I decided to entirely rework my model and develop a second version.

2.2 My second model

2.2.1 Explanation of operations

For this second version of my work, instead of starting from scratch, I decided to used the *Tensorflow Object Detection API*.

TensorFlow Object Detection is a powerful framework for object detection tasks in the field of deep learning. Developed by Google, it provides a flexible and efficient solution for detecting and localizing objects within images or videos. It includes a collection of pre-trained models, such as Single Shot Multibox Detector (SSD) and Faster R-CNN, which can be fine-tuned on custom datasets to achieve higher accuracy and better generalization.

Work with pre-trained models could be interesting in our problem. Indeed, because this kind of models has already been trained to recognize and identify general features from various data patterns, they could provide us better results. As you may understood, this is the reason why I choose to use the *Tensorflow* framework.

2.2.2 Results

After finishing to setup my computer, download every package and library (I detailed every steps that need to be follow if you want to reproduce this kind of models on your computer on my GitHub) I was able to train a new model on my dataset by using pre-trained models.

I choose the *EfficientDet* model for its high efficiency, accuracy, flexibility and adaptability. At first, I decided to train two models on two datasets: one with only portrait paintings and another one with paintings of landscape with buildings. Here is an example of predictions of my models:

Figure 10: Examples of predictions of the "human" model

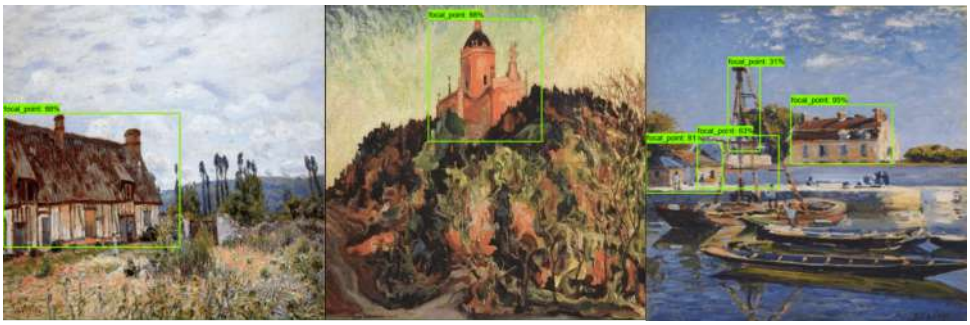

Figure 11: Examples of predictions of the "building" model

As you can see the results are better than previously. Indeed, the program is now able to identify the focal point(s) with an higher efficiency and accuracy. You can also notice that, unlike our previous model, this second version is able to recognize multiple focal point in the same image. This new approach, seems to solve and propose a sustainable solution to our problem.

Having witnessed the impressive performance of our model on theses specific datasets, I was optimistic about its capabilities. However, when I tried to extended the testing to more general images (encompassing diverse styles, types, painters...) I encountered disappointing results.

Figure 12: Examples of predictions of the model

As you can see, when I extended the testing to more general images, the model's performance has declined, revealing its limitations in handling a broader range of artworks. I think that the discrepancy between its success on focused datasets and its struggles with more diverse and general images indicates that the problem may lie in the lack of generality in our training dataset. While the model effectively learned to detect focal points in portraits and landscapes, it seems to lack the adaptability and flexibility to generalize to unfamiliar art genres and unconventional compositions.

This finding emphasizes the importance of implementing a good dataset with clear definitions of what is researched. Hence, in our case where the focal point is not a very well defined concept (it could be subjective: two viewers could identify the focal point on two different spots) it seems to be impossible to define a model that can perfectly find the focal point on a painting.

However, the model that we developed can still be used to find the focal point on specific paintings (likes portraits) or can be used for a different problem that this one (e.g. car detection).

Conclusion

The results demonstrate the potential of AI in art analysis, but also the need for tailored approaches when addressing intricate and subjective concepts like focal points in paintings. The developed models may have valuable applications in specific domains, but further research and dataset refinement are essential to achieving robust AI solutions for painting analysis.

If someone want to pursue this research, here are some recommendations. The main thing that must be improve is the dataset. Indeed, the models (but especially the second one) that I implemented show that they are capable to provide good results if they are feed with good images, hence if we want a model that is able to recognize the focal point(s) all we need to do is build a better dataset. There are different things that we can do to improve this dataset:

1. First, select only paintings where the focal point has been created intentionally by the painters. Indeed, sometimes the artist do not want to grab the attention to the audience to a special point on his painting, and we do not want our model to be able to find a focal point on a painting where there is none.
2. Then, try to restrict the dataset to paintings that only follow one or two concepts of focal point that I presented you in the section 1.2 (contrast, centering...).
3. Finally, increase the size of the dataset, by adding hundreds (or thousands) of new pictures.

I believe that with these advises the model should be able to provide better results for every kind of painting.

Bibliography

- Web Article, *Focal Point Vs. Center of Interest*, Greg Albert
- Web Article, *5 Ways to Create a Focal Point in Artworks*, Matt Fussell
- Web Article, *9 Different Types of Focal Points in Art*, Marla Baggetta
- Web Article, *How to Use Leading Lines to Create More Interesting Compositions*, Dan Scott
- Web Article, *Leading Lines*, Mr. Warrens
- Tutorial, *TensorFlow 2 Object Detection API tutorial*, Mr. Lyudmil Vladimirov