BroadBrush: Final Report

Stephen Brade, Sandra Petkovic

December 3, 2019

Word count: 1975

Introduction

When considering how much of the world is being digitized today, there is a use case for a tool that can classify art so that it may be properly stored and easily-accessible for the masses to enjoy. It is out of this need that we wanted to create BroadBrush: a classifier that, given an artwork, can determine the era and artist from which it came.

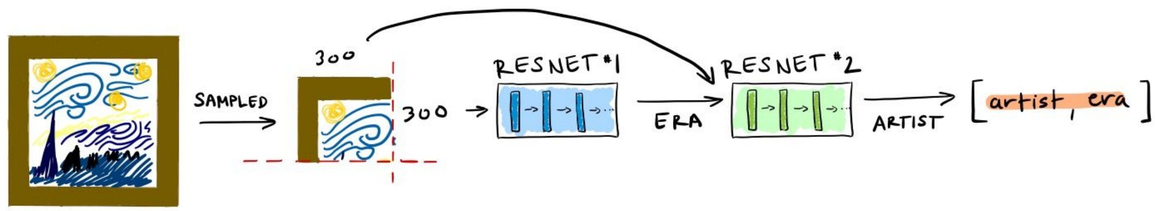

Figure 1: An overview of BroadBrush's architecture.

Figure 1 illustrates a high-level view of how BroadBrush works. We used two separate models, each predicting an era or artist, and placed them in series to improve the accuracy. In our case, we placed the era model before the artist model so the era prediction narrowed the scope of the artists.

Background & Related Work

Similar work has been done by researchers at Stanford on how well a machine learning model can predict a painting's artist from a pool of 57 artists [1]. Using the WikiArt dataset and ResNet models, the researchers were able to achieve over 90% accuracy with their model.

Data and Data Processing

For our dataset, we used the WikiArt dataset, a 27 GB set of all photos on the WikiArt website. To scale down, we chose only 6 eras and 4 artists from each. These are summarized below:

- **Baroque:** Annibale Carracci, Caravaggio, Diego Velasquez, Rembrandt
- **Cubism:** Fernand Leger, Juan Gris, Louis Marcoussis, Marc Chagall
- **Impressionism:** Mary Cassat, Joaquin Sorolla, Edgar Degas, Edouard Manet
- **Pop Art:** Andy Warhol, Hiro Yamagata, Patrick Caulfield, Roy Lichtenstein
- **Realism:** Anders Zorn, Boris Kustodiev, Henri Fantin, Camille Carot

- **Renaissance:** Albrecht Altdorfer, Hans Baldung, Lorenzo Lotto, Michelangelo

To prune the data, we removed all sketches and studies since we were only interested in complete paintings. For augmentation, we randomly sampled multiple 300x300 px squares from the images. To ensure the results were accurate, samples from the same image were not mixed between the training, validation, or test sets; each image and all its sampling were confined to one set.

Before sending the images into the models, we normalized them, applied a Gaussian blur, and integrated a random horizontal flip. The choice to normalize and add a blur were to mimic the ResNet we were using, which had been trained on normalized, blurred images. The random horizontal flip was an empirical decision, as we observed it improved the tool's accuracy.

Below are examples of pre-processed data:

Figure 2: A removed image example. Fernand Leger, Study of Sask*.*

Figure 3: Andy Warhol Marilyn *(left), Andy Warhol* Marilyn *cropped (middle), Andy Warhol* Marilyn *cropped, blurred, normalized, flipped (right).*

Architecture

We implemented our artist and era models with two different architectures: two standard CNNs and two 101-layer ResNets. For the CNNs, we noticed that the more layers we had, the more the model overfit. As a result, we chose to limit the CNN to a single convolutional layer and two hidden layers, all with ReLu activation, batch normalization, and dropout. We assumed that the brush strokes would be the most telling of the artist and era, so it made sense for us to use an architecture of fewer layers with more kernels.

For the convolutional layer, we used 50 kernels of size 10. For the fully connected layers, we used hidden sizes of 100 and 32. Finally, the model was trained with Cross Entropy loss and SGD. The only difference in architecture between the artist- and era-classifying CNNs was the size of the output layer.

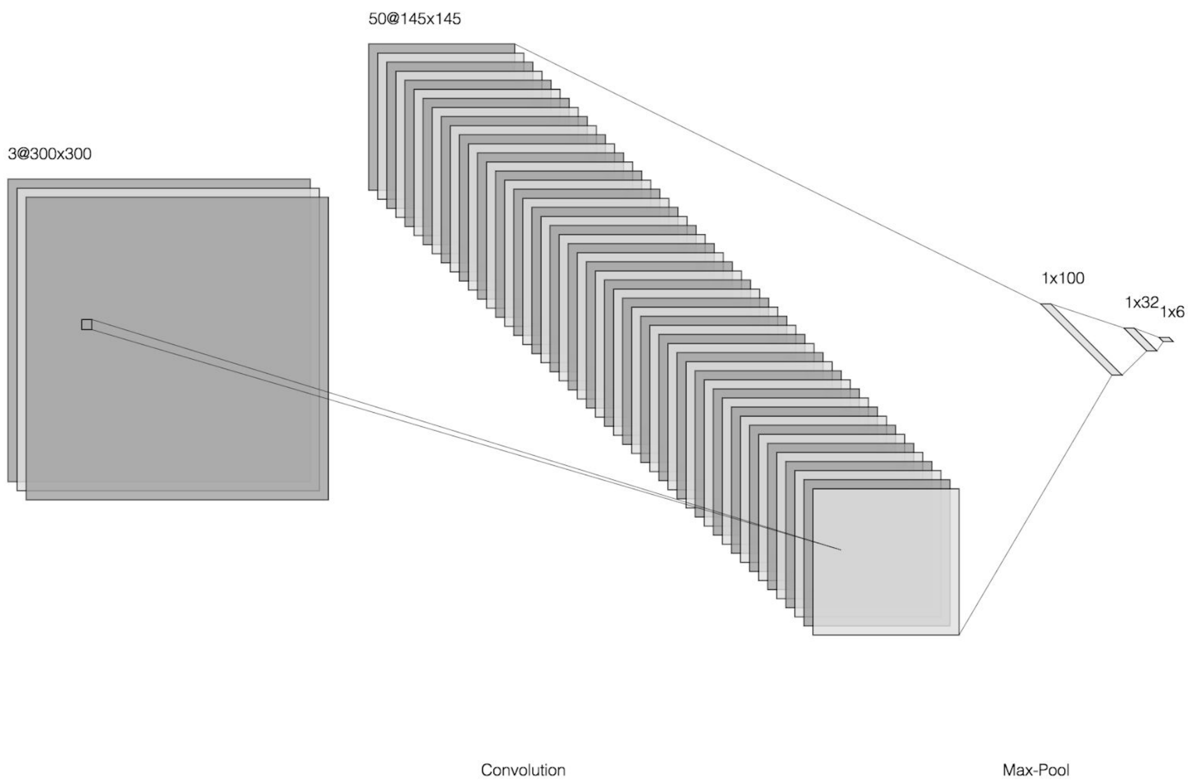

Figure 4: Architecture of CNNs under the context of era classification. The first layer is convolutional, and the small rectangles (right) are fully connected.

For the ResNets, both the artist and era models started as pretrained networks with 101 layers. When training this ResNet, the output layer was adapted to have 6 outputs for era classification and 24 outputs for artist classification. They were trained with Cross Entropy loss and SGD and the pretrained layer was adjusted with a learning-rate of 0.01.

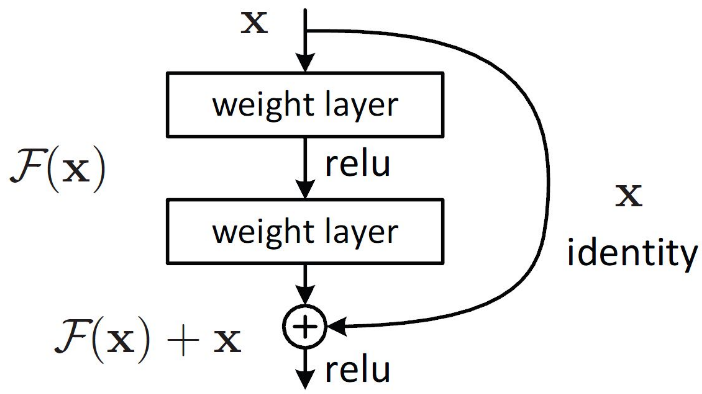

Figure 5: An example of how the residual is calculated in the ResNet-101 that we used. For more detailed architecture, see [2].

Baseline Model

The baseline model used only the colours in the image to make a prediction. The model determined the top colours in the image and compared it to the colour averages of each era in our dataset.

For an accurate comparison, the image vectors were converted from RGB to the CIELAB colorspace, a numerical system designed specifically to compare colours [3]. To find the difference in colours, the “Delta E” function was used in place of Euclidean distance, and the minimum difference was taken as the prediction. From this, the baseline was able to achieve 30% accuracy. All conversions and calculations were done with the help of Python libraries.

Quantitative Results

Table 1: Results Table

	Baseline	Era CNN	Artist CNN	Era ResNet	Artist ResNet
Training Accuracy	**30%**	**100%**	**100%**	**100%**	**100%**
Validation Accuracy	**n/a**	**45.6%**	**20.1%**	**86.0%**	**68.1%**
Testing Accuracy	**n/a**	**40.2%**	**16.8%**	**84.9%**	**63.4%**
Specs	**Based on colours**	**Few layers, many kernels**	**Few layers, many kernels**	**Transfer learning**	**Transfer learning**

Above is a summary from all the models, including their accuracies and general structure. As seen, we were able to overfit on all our intelligent models but had overall better results with the ResNet than with the CNNs. All of our intelligent models also performed better than the baseline 30%.

CNN

Though it didn't take long to implement the CNNs, the results weren't promising. They overfit quickly (figures 6 and 7) while the validation accuracy never improved to above 50%. We attempted to change the number of convolutional layers, apply dropout and batch-normalization, and chang other parameters – none of which significantly improved accuracy. We would have liked to optimize these models more, however the prudent approach was to move onto a ResNet which ended up giving better results in less time.

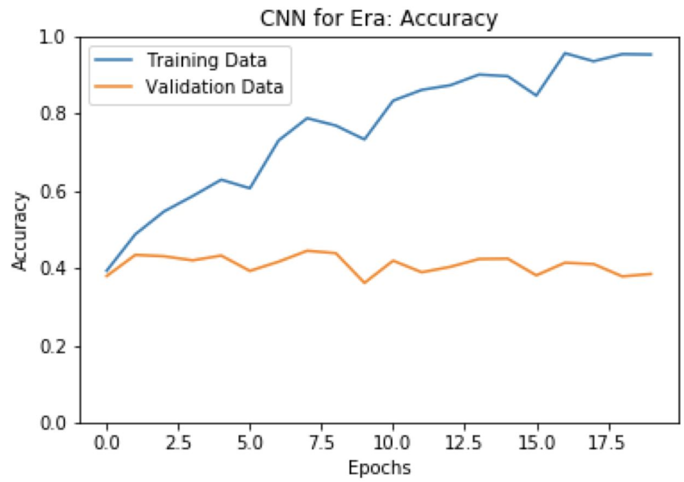

Figure 6: Example of how the CNN overfit in the context of era classification.

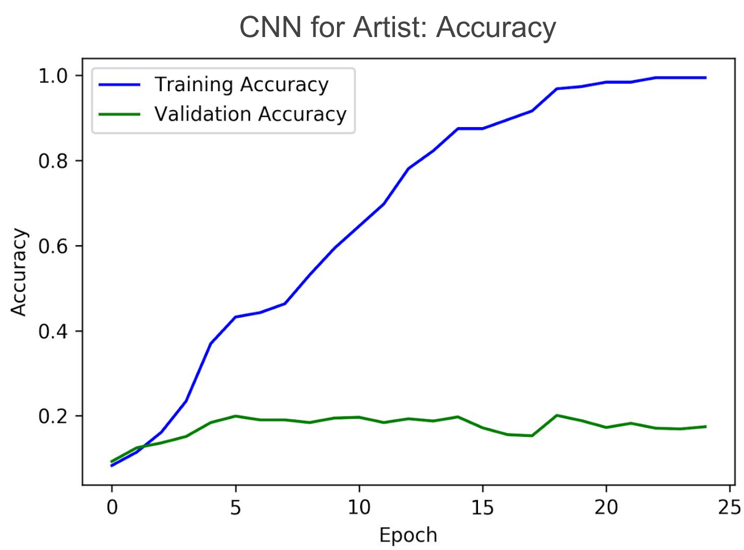

Figure 7: Example of how the CNN overfit in the context of artist classification.

ResNet

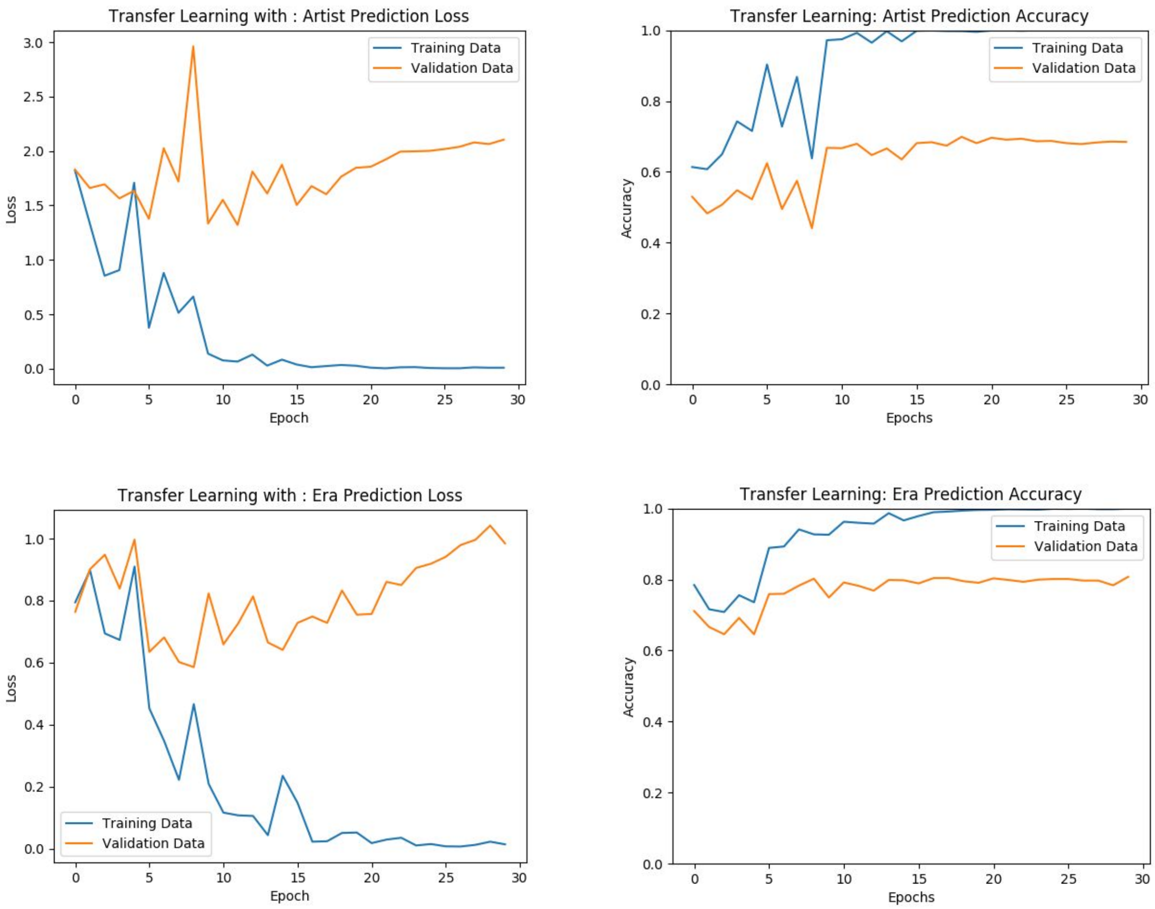

Figure 8: Training and validation accuracy and loss plots for ResNet-101 networks. Note for artist classification that the graphs above are not informed by the era classification. This strategy will be implemented in testing.

As seen in the plots in figure 8, the ResNets gave better results: they didn't overfit as much and we were able to get the test accuracy to 84.9% for era prediction in addition to 80.4% test accuracy for artist prediction once the two networks were linked together.

Qualitative Results

Below are the confusion matrices from our models. In figure 9, we see the confusion matrix for the era ResNet. With a test accuracy of nearly 85%, there's no surprise in seeing a strong diagonal in the matrix. However, it is interesting to note that the matrix is not symmetrical. It's greatest point of confusion is between Baroque and Renaissance. These eras are very similar and are areas of confusion even for humans, but the confusion only went one way: the model confused Baroque for Renaissance, but rarely confused Renaissance for Baroque.

Reasons for this may be a biased classifier, maybe as a result of Baroque croppings that weren't diverse enough. There is the chance that many croppings of the Baroque period were the dark, untelling portions and the model never learned enough. A remedy to this would be to create a larger dataset or perform the croppings manually.

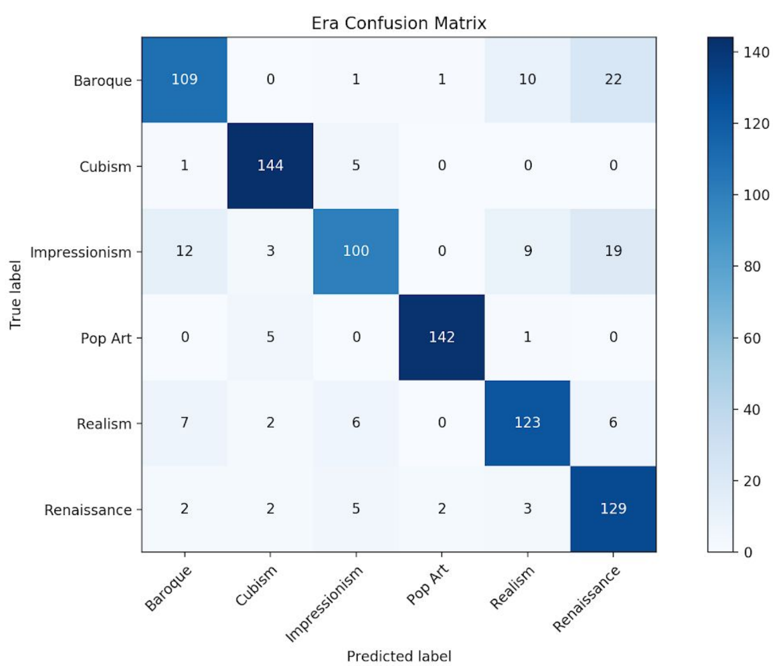

Figure 9: Confusion matrix for era ResNet. The greatest point of confusion occurs in the top right corner between Baroque and Renaissance.

In figures 10 and 11, which display the confusion matrices for the artist ResNet, we can see the difference that adding the era prediction made. In figure 10, without the era prediction, the diagonal is sparse and has distinct weaknesses such as correctly identifying a work by Anders Zorn or Annibale Carracci.

However, once the model had been fed the era prediction, we can see how the results improved in figure 11. The diagonal is much stronger and the previous weaknesses were overcome almost entirely. In the case of Anders Zorn, the model went from 4 correct predictions up to 29. This is reflected in the testing accuracy, which went up to 80.4% with the addition of the era prediction.

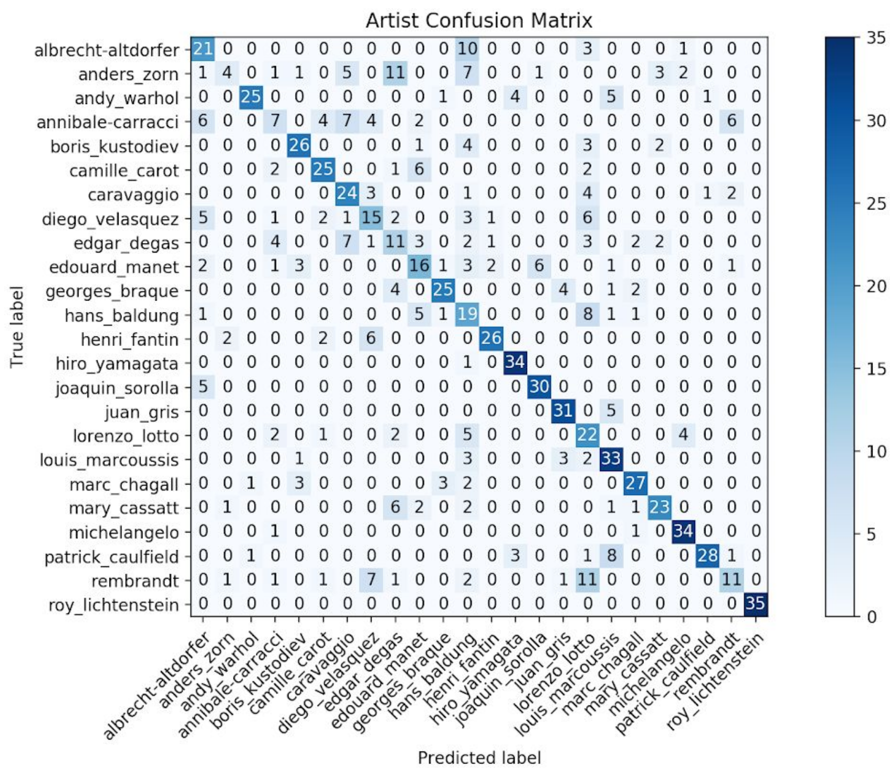

Figure 10: Confusion matrix for artist ResNet without the era prediction.

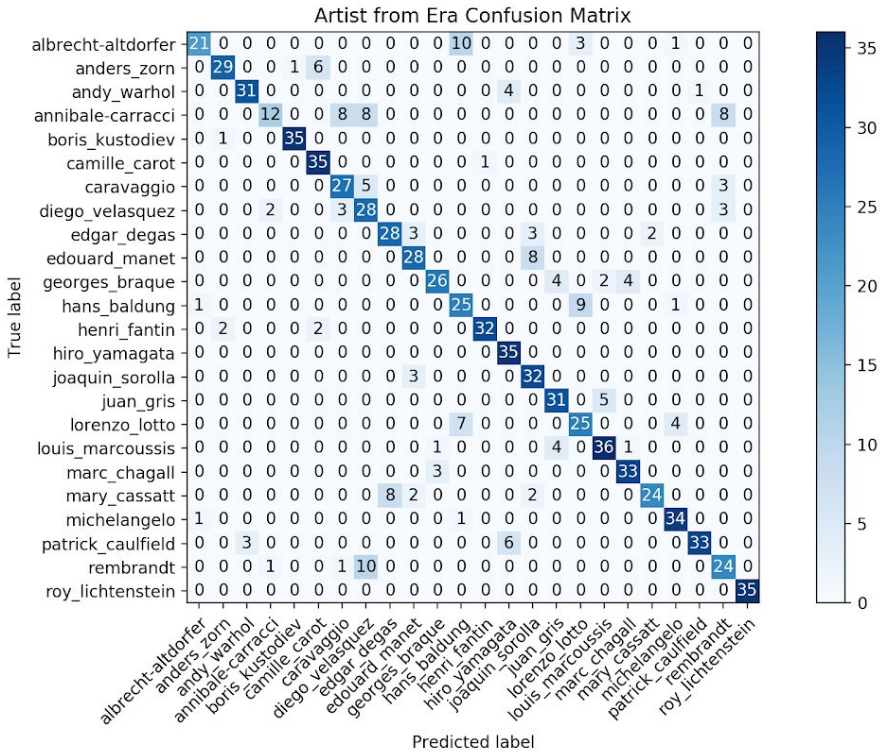

Figure 11: Confusion matrix for artist ResNet with the era prediction.

The points of confusion that remain in the artist confusion matrix, such as in the top middle between Albrecht Altdorfer and Hans Baldung, occur between artists of the same era. This shows how feeding the era prediction helped improve the artist ResNet's areas of confusion between artists of different eras, but couldn't remedy the confusion between artists of the same era.

Discussion and Learnings

The results of this project were successful! We nearly replicated the testing accuracy of other previous art classification projects and gained valuable experience with machine learning.

Data collection and processing was the major bottleneck in our project. To train our model with enough information, we were constantly switching out eras and artists for those with more artwork. This was before we considered the multiple cropping idea, which would have allowed us to keep smaller sets, such as Fauvism. Future advice would be to plan how the images will be preprocessed *before* building the dataset.

From building the models, we found that the code implementation was easier than anticipated, and the real difficulty was the architecture. Creating the best model required time and a deep understanding of both the problem and machine learning techniques. Although we felt that the fine tuning done with the CNNs was enough to show that art classification could be done with machine learning, we accept that a lot of the problem is still not entirely understood. We aren't aware of what the models look for to make their predictions, but with more time and computing power, that is something we'd like to investigate.

Ethical Framework

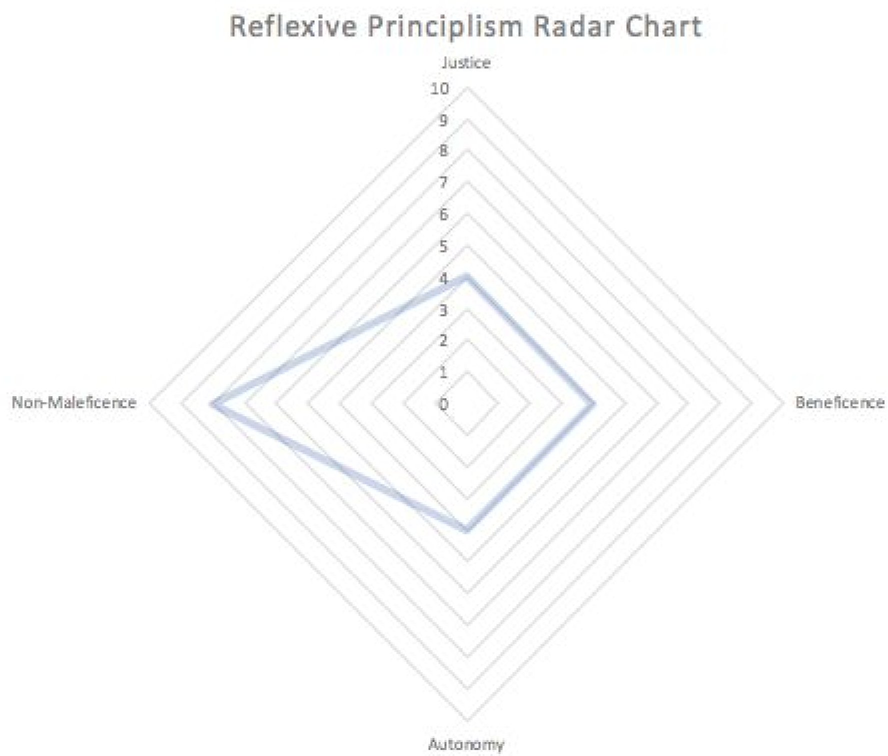

Figure 12: Reflexive Principlism plot.

Figure 12 describes where we believe our project sits on the axis of ethical reasoning, described in reflexive principlism. The numbers in the figure act only as a way to compare each axis' relevance with another, not as actual values.

We believe that our project is non-maleficent and beneficent. This is due to the fact that the classifier's prediction of the artist or era is of little consequence to anyone or anything. In the modern day, there isn't very much that depends on an artwork being classified correctly.

One may mention that the classifier could replace the job of an art curator, however, that is likely not the case. The job of an art curator is to choose and arrange pieces of art to make an appealing and stimulating exhibit. Although we admit that this may one day be achievable by artificial intelligence, it would require much more than an art classifier. That being said, a future version of BroadBrush could possibly aid in discovering when a painting was made and identify the artist.

In terms of justice, we believe that the art classifier is a step towards justice. The knowledge and experience in making a tool like BroadBrush could aid in learning how to identify fraudulent art.

In terms of autonomy, the classifier can be adapted to identify paintings in real-time through a smartphone app and provide information about the artist and era. This can be easily integrated into museums or galleries to provide visitors a more autonomous and informed experience.

References

[1] N. Viswanathan, “Artist Identification with Convolutional Neural Networks.” 2017. http://cs231n.stanford.edu/reports/2017/pdfs/406.pdf.

[2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Cornell University, Dec. 2015.

[3] T. Mouw, “LAB Color Values,” *X-Rite Pantone*, 08-Oct-2018. [Online]. Available: https://www.xrite.com/blog/lab-color-space.

Permissions

Stephen Brade and Sandra Petkovic both consent to the following:

- Posting our video on the course website
- Posting our final report
- Posting our source code