**Classification of Modern Sculpture Using Convolutional Neural Networks**

Anouk Flinkert, s2313820

Leiden University Centre for Digital Humanities

26-6-2023

## 1 Introduction

During my internship at the modern and contemporary sculpture museum Beelden aan Zee (BaZ), I encountered the challenges inherent in the art profession and found inspiration for the practical application of Digital Humanities in the cultural sector. I witnessed the meticulous efforts invested by employers in digitally registering their collection. My responsibilities included manual registration, research, and transcription of pertinent handwritten archival materials for each object. Private museums with dedicated volunteers can afford such labor-intensive tasks, but larger museums like the Rijksmuseum face numerous registration requests, including weekly incoming and outgoing loans, making a lengthy process unmanageable for them. This experience sparked my curiosity and helped me wonder if the efficiency of digital collection registration could be enhanced by leveraging artificial intelligence. In this project, I attempted to employ a Deep Learning algorithm known as a Convolutional Neural Network (CNN), to classify the types of sculpture (e.g., 'head', 'figure', or 'geometric form' etc.) in BaZ's collection using the Transfer Learning technique on a pre-trained model. This research demonstrates the precarious nature of classifying art objects, given the exciting contradiction between the unconventional nature of modern and contemporary art and CNN's reliance on formalities to learn and recognize patterns.

## 2 Related work

In 2002, Daniel Keren made one of the earliest attempts at algorithmic art classification by utilizing local features and naïve Bayes to identify painters.[1] Afterwards, the use of Convolutional Neural Networks (CNN) for fine art classification has been rapidly advancing. In 2018, Eva Cetinic et al. conducted fine-tuning research of pre-trained models that achieved significant test accuracies for the classification of artists (79.8%) and genres (76.9%) in fine art paintings.[2] Even more captivating is the growing emphasis on thoughtful analysis of the data at hand when applying Computer Vision to the arts.[3] Mohammad Reza Mohammadi and Fatemeh Rustaee point out the challenges faced by machines in understanding the visualizations present in abstract art, as well as the confusing categorical distinctions in paintings, such as style, in their experiments on classifying fine-art objects using deep learning neural networks.[4] Furthermore, Cetinic et al.'s groundbreaking research in 2019 tested the historical transformation of stylistic properties described in Heinrich Wöfflin's book

[1] Keren, D. 'Painter identification using local features and naive Bayes.' *In Pattern recognition*, 2002. Proceedings. 16th international conference on: 2.

[2] Cetinic, E., Lipic T., Grgic S., 'Fine-tuning Convolutional Neural Networks for fine art classification', *Expert Systems with Applications*, Volume 114, 2018, pp. 107-118.

[3] 'Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos.' Read 'Computer vision' on Wikipedia: https://en.wikipedia.org/wiki/Computer_vision. Accessed on 26-06-2023.

[4] Mohammadi, M.R., Rustaee, F. 'Hierarchical classification of fine-art paintings using deep neural networks.' *Iran J Comput Sci* 4, 59–66 (2021).

*Principles of Art History* (1915) using CNNs.[5] It appears that recently, models are being adapted to accommodate the data, rather than the other way around.

However, most of these experiments focus on the discipline of painting, particularly pre-modern works. Only in 2017, Dajeong Hong and Jongweon Kim worked on CNNs for the detection of sculptures, rather than classification, using a dataset of sculpture images captured from various angles.[6] Yet, most research is practiced on traditional art movements and periods characterized by figurative styles. Thus, the possibilities of applying Computer Vision to modern and contemporary art remain largely unexplored.

## 3 Research question

Considering modern and contemporary art as a prominent component of today's art industry, I developed a keen interest in examining the obstacles and potentials of integrating this with the field of Computer Vision. This project entails training a pre-trained CNN model, 'EfficientNet_B0_Weights.IMAGENET1K_V1,' using Transfer Learning (TL) on a custom dataset created from BaZ's digital collection.[7] A CNN can be seen as an artificial replica of the neurological processes in our brains. Through TL, it is possible to use a pre-trained model for a related task from what it originally is trained and constructed for.[8] Only the last layer of the model is trained to recognize patterns (or referred to in Deep Learning as calculating the 'weights') specific to the new dataset it is forwarded. TL and Fine-tuning have proven to be highly effective and least time-consuming methods for training CNNs on new or custom datasets.[9]

Following the training process, the results are evaluated using various metrics such as accuracy, precision, recall and the f1-score, along with visualizations of predictions like the confusion matrix and plots. These evaluations are crucial in gaining insights into what the model has or has not learned. Unfortunately, the model exhibited poor performance in predicting the correct types of sculpture based on our dataset. Different computational features could potentially improve the prediction accuracy and minimize the loss, such as hyperparameter tuning. However, the visualizations provided interesting observations, indicating the complex nature of the data that poses significant issues for CNN classification.

## 4 Data

The data for the custom dataset 'baz_dataset_V1' is sourced from the museum BaZ and consists of 630 images in different formats and varying quality.[10] It is important to acknowledge the selective process that has shaped the representation and balance (or imbalance) of the dataset. A macro factor influencing this is the scope of the museum's

---

[5] Cetinic, E., Lipic T., Grgic S., 'Learning the Principles of Art History with convolutional neural networks', *Pattern Recognition Letters*, Volume 129, 2020, pp. 56-62.

[6] Hong, D., and Jongweon K. "Sculpture Detection Method using the Convolution Neural Network." *Proceedings of the 2017 International Conference on Information Technology*. 2017.

[7] The documentation and source for the pre-trained model EfficientNet_B0_Weights.IMAGENET1K_V1 is available at: https://pytorch.org/vision/stable/models/generated/torchvision.models.efficientnet_b0.html#torchvision.models.EfficientNet_B0_Weights

[8] Hussain, M., Bird, J.J., and Faria, D.R. "A study on cnn transfer learning for image classification." *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence*, September 5-7, 2018, Nottingham, UK. Springer International Publishing, 2019.

[9] Cetinic et al., 'Fine-tuning Convolutional Neural Networks for fine art classification'.

[10] The data is obtained with the approval of conservator Dick van Broekhuizen and through correspondence with the host of the digital photobank, Ton Horsten. Recent initiatives have supplemented the database with professional imagery of the entire collection. As a result, the quality and size of some images are more advanced than older versions captured before the start of this process. The images are available in the formats '.jpeg', '.tiff', or '.TIFF'.

collection. The imbalanced distribution of sculpture types within the dataset is inherent to the composition of the museum's collection, which includes a substantial proportion of figures and heads.[11]

Several micro factors played a decisive role in determining the available data range. These factors are limited access to photographs, time constraints, and resource limitations for creating a comprehensive manually supervised dataset. The CNN is trained to classify sculptures into eight predefined classes: 'bust', 'figure', 'fragment', 'geometric form', 'head', 'installation', 'organic form', and 'relievo'.[12] Even for human experts, identifying the distinguishable features of these types is complicated as it necessitates art historical expertise and critical analysis. Modern and contemporary sculpture often contradict traditional forms of representation, leading to ambiguity in type attribution. For instance, Sorel Etrog's (1933-2014) *Dream Chamber* (1976) (Fig. 1) initially appears as a geometric form, but the artist intended to depict a skull, as implied by the title. However, the model encountered most difficulties in differentiating between the categories 'bust', 'figure' and 'head'. Given the complexity of these types, the given results seem intuitive.

Finally, as Hong and Kim shared in their research, the three-dimensionality of the artworks and varying conditions surrounding the object which create extra noise, further complicate the model's learning process.[13] With only a few exceptions, sculptures incorporated in the dataset are represented from a single viewpoint. Considering that the perspective of a sculpture is a pivotal element in the interpretation of its type, this decision contributes to the selective character of the dataset.

## 5 Methods

Given the nature of the data source described above, we shift into a paradox of representation. A decision needed to be made regarding whether to a) maintain the inherent imbalance of the data, thereby focusing on the CNN's accurate classification of the composition of BaZ's sculpture collection; or b) settle a relative imbalance across all classes which would prompt a broader examination of the CNN's performance on images of modern sculpture provided by the museum BaZ. This project opted for the latter approach and calculated a balanced distribution for each class (Table. 1). To evaluate the performance

**Fig. 1** Sorel Etrog, *Dream Chamber*, 1976, bronze casting, 150 x 60 x 60 cm. Collection Beelden aan Zee.

[11] In 1994, the museum was opened to house the private collection of modern and contemporary sculpture belonging to art collectors Theo and Lida Scholten. The Scholtens had a particular interest in the human figure as depicted by Dutch artists, which explains the dominant presence of this type of artwork in the current collection.

[12] Originally, the types 'portrait' and 'monument are included in BaZ's categorization. However, without access to metadata these attributions are impossible to determine for either human or machine and are therefore excluded from this project.

[13] Hong and Jongweon, 'Sculpture Detection Method using the Convolution Neural Network', p. 148.

of the BaZ dataset, the well-defined Fashion-MNIST dataset from the pytorch.datasets library was also run through the same model.

| Classes | Distribution of images | Percentage of total |
|---|---|---|
| bust | 85 | 13,49% |
| figure | 95 | 15,08% |
| fragment | 67 | 10,63% |
| geometric form | 90 | 14,29% |
| head | 96 | 15,24% |
| installation | 64 | 10,16% |
| organic form | 63 | 10% |
| relievo | 70 | 11,11% |
| **Total** | 630 | |

**Table 1**. Distrubition of images across all classes in the BaZ dataset.

I made use of the open-source PyTorch Machine Learning framework in Python, with frequent use of the Torchvision package for specific modules, architectures, and image-related functionalities. The steps undertaken in image classification are well-documented and commented on in the attached notebook 'baz_TL_effnet'. In this section, I will discuss and justify deliberate choices made during these steps.

The transforms must be defined for image processing before downloading the data into the Pytorch environment. The transform is manually defined and incorporates the Grayscale function to reduce computational requirements. Also, the augmentation TrivialAugmentWide is activated to slightly transform the images thereby generalizing the data and improving the accuracy.[14] Once the data is downloaded, it is prepared further in the dataloader functions. To prevent overfitting of the mildly imbalanced dataset, the shuffle parameter is set to 'True', ensuring that the images are shuffled before being divided into batches of size 32. Next, the CNN model EfficientNet_B0 is loaded, and the TL method is applied. All layers and pre-trained weights are frozen and used for making predictions on the new data, except for the classifier, which will be trained separately on the sculpture images. With only 5.3 million parameters, the EfficientNet model reached an accuracy of 93.532% after the fifth epoch of its original training. Respective of this project's scope, available resources, and timeframe, a model with a small number of parameters is a suitable choice due to its low computation time.

For the most optimal loss calculation and minimization CNN multiclass classification, the 'CrossEntropyLoss' loss function and Adam optimizer are selected. These choices are made to effectively measure the loss between predicted and target class probabilities and to efficiently update the model's parameters during the training process. After defining the necessary functions and variables, a train and test loop is initiated including calculations of accuracy, test loss, and train loss. These values are automatically logged in the Wandb application by configuring its API.[15]

Finally, to evaluate the model's performance, its predictions are visualized by plotting images using the matplotlib library. The Torchmetrics mlxtend extension provides a function for plotting the Confusion Matrix, which illustrates the frequency of (in)correct predictions across all classes. This helps to determine the instances where the model confuses one type for another. The sklearn.metrics module enables the generation of a classification report projecting precision, recall, f1-score, and micro accuracy measures. Precision calculates the probability of the model correctly predicting a type as positive, while recall measures the total number of correct predictions of a type (either as negative or positive). The F1 score represents the

[14] Pytorch's documentation on the Grayscale function and TrivialAugmentWide augmentation: https://pytorch.org/vision/stable/generated/torchvision.transforms.Grayscale.html; https://pytorch.org/vision/main/generated/torchvision.transforms.TrivialAugmentWide.html.

[15] For more information: https://wandb.ai.com/

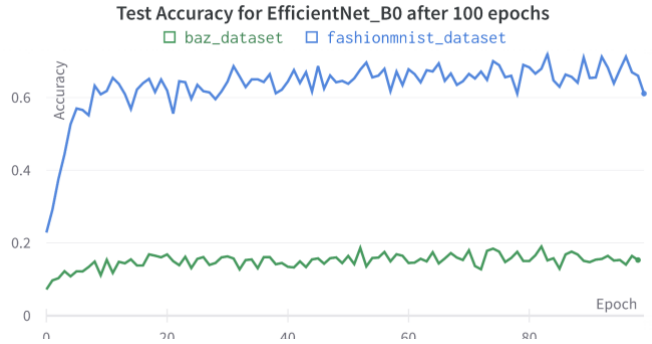


**Fig. 2** Test accuracy curves for the BaZ and FashionMNIST classification

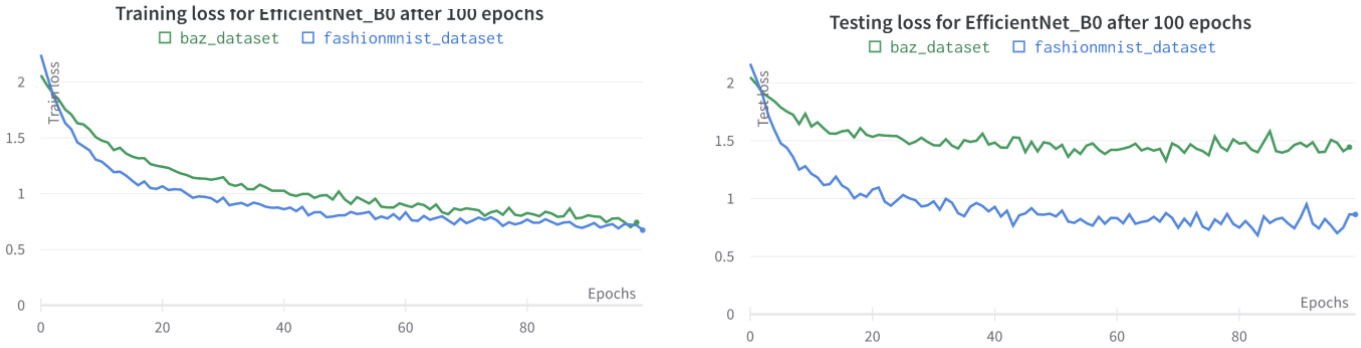


**Fig. 3** Training and test loss curves for the BaZ and FashionMNIST classification

harmonic mean of precision and recall and assesses the model's predictive performance between a score of 0 and 1. The micro accuracy metric calculates the average by aggregating the contributions of all classes,

making it particularly useful for multiclass problems. These measures help to signal possible associations the model is making between certain types.

## Results and evaluation

### *6.1 Test accuracy and train/test loss*

Figure 3 displays the train and test loss EfficientNet during training on both the BaZ and MNIST datasets for 100 epochs. The graph indicates that the model is indeed learning from the BaZ dataset, as the test loss curve consistently decreases over time. This suggests that the model neither overfits nor underfits the training data. However, the low accuracy (18,65%) implies that the model is likely making small errors in most of its predictions. Furthermore, the model's ability to achieve at least 71.95% accuracy on the MNIST dataset demonstrates that the accuracy results are dependent on the input data. The comparison between BaZ's and the MNIST's loss curve indicates that BaZ's data may be overly complex for the model to effectively learn from or handle.

### *6.2 Predicted images*

In Figure 4, a visualization of 9 samples offers a detailed look at the model's predictions. From a human perspective, certain incorrect predictions seem understandable, such as the

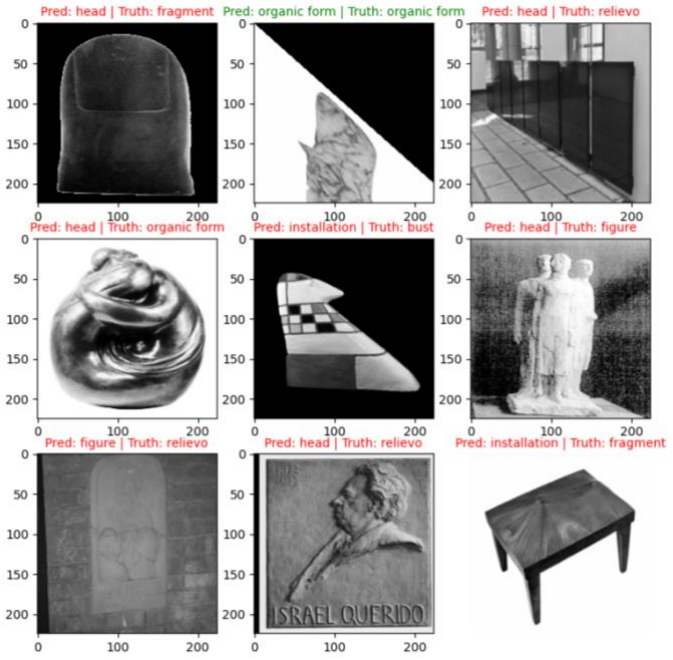


**Fig. 4** A plot of 9 sample images and the model's predictions

confusion between the image of a relievo as a head located in the middle of the lower row. By visualizing the errors, we can estimate patterns that the model is learning for each class.

### *6.3 Confusion Matrix*

The confusion matrix in Figure 5 shows the relations the model detects amongst all types. As previously anticipated, the type ‘bust’ appears to be frequently mistaken for the types ‘figure’ and ‘head’. However, the model excels in making accurate predictions for the 'head' class. On the other hand, the model struggles with accurately classifying sculptures labeled as ‘fragment’ and ‘organic form’, exhibiting a high level of confusion in these categories.

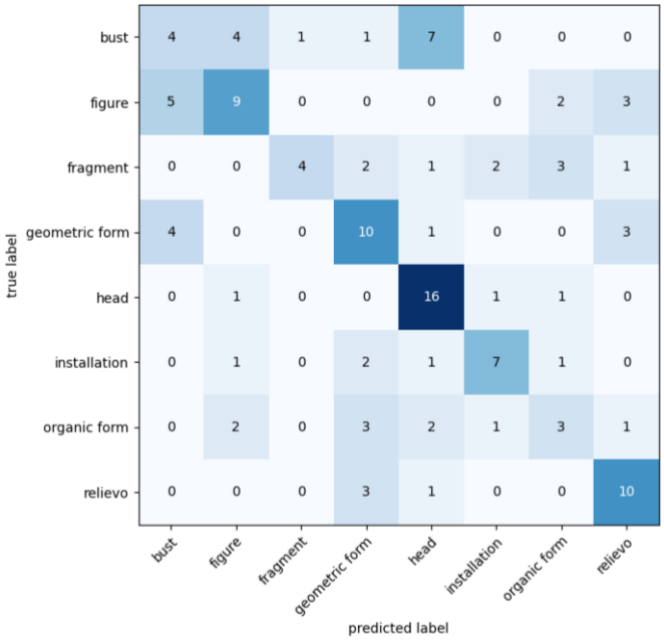


**Fig. 5** Confusion matrix for the BaZ type classification

### *6.4 Precision, recall, F1-score, and micro accuracy*

Though the ‘bust’s’ general low precision and recall scores (0,29) would suggest an overall lowest f1-score, ‘organic form’ holds the minimum value of 0,22 (Fig. 7). This implies that in the model’s correct predictions of ‘organic form’, most instances were False/negative. Interestingly, both the ‘head’ and ‘relievo’ classes achieve similar f1-scores, despite the former being represented by 96 images in the dataset and the latter by only 70. The micro accuracy of the BaZ dataset is 50,8%, which is higher than the macro average (44%) (Fig. 8).

| | precision | recall | f1-score |
|---|---|---|---|
| bust | 0.29 | 0.29 | 0.29 |
| figure | 0.43 | 0.47 | 0.45 |
| fragment | 0.33 | 0.23 | 0.27 |
| geometric form | 0.53 | 0.50 | 0.51 |
| head | 0.55 | 0.84 | 0.67 |
| installation | 0.31 | 0.33 | 0.32 |
| organic form | 0.33 | 0.17 | 0.22 |
| relievo | 0.75 | 0.64 | 0.69 |
| accuracy | | | 0.46 |
| macro avg | 0.44 | 0.44 | 0.43 |

**Fig. 6** Output for precision, recall and f1-score calculations per each class and the macro average

```
Precision: 0.5080645161290323
Recall: 0.5080645161290323
Accuracy: 0.5080645161290323
```

**Fig. 7** Output for total micro precision, recall, f1-score and accuracy calculations

## 7 Synthesis

Given the results of the calculated metrics, the following hypotheses can be proposed to explain the model’s poor performance in classifying the types of sculpture:

- The small size of the BaZ dataset (only 630 images) may hinder the model's ability to accurately learn from the data.
- The images of sculptures contain much information about their surroundings, due to this extra noise the model struggles to distinguish relevant features.
- The complex nature of the class distinctions in sculpture types poses difficulties for both humans and machines.

- The model's training duration of 100 epochs may be insufficient, as many models require longer training sessions to learn intricate patterns.
- The hyperparameters (e.g. batch size, learning rate, and dropout) have not been fine-tuned.

## 8 Conclusion and further debate

As is typical in Digital Humanities research, the project's main challenges existed of practical restraints that affected an attainable outcome. However, apart from working toward a desirable result, the research encompassed several other objectives. In this paper, I have reflected on the process of gaining familiarity with the science of Deep Learning as an art historian and the application of Computer Vision to humanities data. In this final part, I wish to emphasize the relevance of the interdisciplinary study of Digital Humanities drawing from this research.

Within Humanities curricula, theory and conceptual frameworks are often prominent in academic education. In my view, Digital Humanities connects the excessively debated theory to the practical dimensions of everyday life. This interdisciplinary field not only enhances the efficiency of the workflow of professionals in their respective fields, but also comprises the enumeration, measurement, and visualization of complex information. By presenting humanities concepts in a manner that diverges from their inherent nature, Digital Humanities cultivates novel insights and stimulates broader discussions, propelling further exploration and analysis.

When humanities scholars delve into Computer Vision, their aims diverge from those of computer scientists, who often prioritize achieving maximum results. Modern and contemporary sculpture seemed not the most suitable subject of study for the particular task of image classification. However, further exploration of this experiment may validate or challenge the neurological patterns we believe exist.

## References


1. Cetinic, E., Lipic T., Grgic S., 'Fine-tuning Convolutional Neural Networks for fine art classification', *Expert Systems with Applications*, Volume 114, 2018, pp. 107-118.
2. Cetinic, E., Lipic T., Grgic S., 'Learning the Principles of Art History with convolutional neural networks', *Pattern Recognition Letters*, Volume 129, 2020, pp. 56-62.
3. Hong, D., and Jongweon K. "Sculpture Detection Method using the Convolution Neural Network." *Proceedings of the 2017 International Conference on Information Technology*. 2017.
4. Hussain, M., Bird, J.J., and Faria, D.R. "A study on cnn transfer learning for image classification." *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence*, September 5-7, 2018, Nottingham, UK. Springer International Publishing, 2019.
5. Keren, D. 'Painter identification using local features and naive Bayes.' *In Pattern recognition*, 2002. Proceedings. 16th international conference on: 2.
6. Mohammadi, M.R., Rustaee, F. 'Hierarchical classification of fine-art paintings using deep neural networks.' *Iran J Comput Sci* 4, 59–66 (2021).
7. Wentao Zhao, Wei Jiang, Xinguo Qiu, "Big Transfer Learning for Fine Art Classification", *Computational Intelligence and Neuroscience*, vol. 2022.