# Neural Recognition of Media of Art Pieces

Jane M Boettcher
Stanford University
jmsb@stanford.edu

Shridhar Athinarayanan
Stanford University
shriathi@stanford.edu

## Abstract

*Historical art classification serves as an engaging topic within the digital humanities given the vast, rich visual and textual data involved. Most research has classified art pieces by artist and style, but there is little to no prior work on classification by medium. We utilize various Convolutional Neural Network (CNN) architectures (mostly variations of ResNet) as platforms for experimentation to ultimately develop and fine-tune a high performing medium classification paradigm. Using data from Tate collection involving British artwork from 1500 to the present and models ranging from a rudimentary CNN to ResNet18 paired with data augmentation and frozen gradients, our strongest model obtained an overall test accuracy of 74.9%. We also showcase interesting discussion on binary classification through medium tagging. Ultimately, these results showcase the power of CNNs in medium classification, signifying their abilities of learning representational characteristics of various media.*

## 1. Introduction

With the rise of the digital world, there have been numerous efforts to document art collections online. Unfortunately, many public art institutions find themselves struggling to digitize their collections due to the prohibitive costs of photographing and labeling art and images by features such as style and medium [2]. One way to make this process easier is the automatic classification of such meta features. While there has been significant work on the classification of styles and genre, classification of the medium of art pieces — i.e. the materials which are used to compose this art — has not been significantly explored, perhaps in-part due to the unavailability of a large, downloadable dataset which annotates these features. The WikiArt dataset infrastructure, for example, includes genre, style, and artist, but it does not include medium, although this is often easily accessible on art websites and can be manually downloaded from them.

The idea of classifying art pieces under their medium has multiple implications. For instance, such a classifier can ascertain which elements of certain media are the most representational of that media, allowing us to further understand the qualities of, for instance, oil paints or watercolor. Further, understanding this problem has the potential for transference into other key areas that work with material identification outside of art, such as fabric classification. Thus, the problem of classification of art by medium is one that is both important and relatively unexplored.

### 1.1. Problem Statement

For our project, we were guided by the following research inquiry - is it possible to create a classifier which can categorize artwork into its respective media? Our final model is trained on images from the Tate collection, one of the United Kingdom's largest national collections of British art. The input to our algorithm is any 224 pixel by 224 pixel image of artwork. We then use a Deep Residual Neural Network, or ResNet model, paired with a binary classification system to output a predicted medium for the piece from 10 choices of media.

## 2. Related Work

Though there has not been much work on classification of paintings by medium, the deep-learning-for-art space has seen some work on the classification of art by other metadata. For instance, Banerji and Sinha (2016) [3] looks at multiple classifiers beyond CNNs to categorize artwork by their respective authors and also styles. They started with the Painting-91 dataset which consists of over 4,000 paintings in multiple styles. Further, they used CNNs to generate feature representations and then experimented with EFM-KNN and Linear SVM classifiers to classify pieces with these CNN features — their most accurate model generated a 45% test accuracy. This paper, however, was not as robust as the experimentation done by Hong and Kim (2017) [8] who investigated the effect of different CNN architectures on the classification of art pieces by author. They included distortions of artworks to simulate them on a TV screen for detection of artwork in copyright proceed-

ings. Distortions such as scaling, color variation, rotation, and lens distortion were applied to the researcher's set of images to expand their dataset. The researchers tested three main architectures, one based on AlexNet and two based on VGGNet, each with convolutional layers of different filter sizes, padding, and strides. They found that their second VGGNet implementation works better than the standard SIFT model for image classification. As for other CNN-based papers, Cetinic et al. (2018) does well in focusing on discussion regarding fine-tuning hyperparameters, finding high accuracy is achieved with a constant learning rate of $10^{-4}$ with a stochastic gradient descent optimizer of momentum 0.9 and weight decay 0.0005 [4]. Additionally, to inform data selection, Zhao et al. (2021) focuses more on data selection than the other papers — like Banerji and Sinha, they use Painting-91 with 1,250 samples for training and the other 1,088 for evaluation. For the classification task itself, they only selected artists with more than 500 paintings [14].

Zhao et al. utilized a Deep Residual Neural Network, or ResNet, as his classification model. According to He et al. (2015), the developer of ResNets, deep convolutional neural networks can oversaturate and ultimately degrade the accuracy. ResNets involve residual block layers with skip connections, or connections which skip training layers [7]. The researchers found residual nets achieve only a 3.57% error on the ImageNet test set. ResNet is robust enough to even do mood detection of paintings as suggested by Huo et. al. (2020). Using categories such as "horrifying" and "peaceful," ResNet was able to delineate paintings by such categories with an accuracy of 93.7% using an Adam optimizer [9].

One potential issue with classification of media is that pieces often could be developed with multiple types of media. Additionally, media can show up differently on different types of material (i.e. charcoal marks differently on drawing paper verses on canvas). This can expand the feature space for a certain type of media (i.e. charcoal now would have various attributes it can be distinguished by if we lump together its presence on multiple types of material, causing confusion during classification and a reduced accuracy). A method to fix this explored in literature is to break down multi-class classification into binary classification. One standard method for this presented by Allwein et al. (2000) [1]. They showcase that, in a multi-class classifier where each class has multiple attributes, the classes can be decomposed into multiple binary classifiers on whether a class contains one of these specific attributes or not. This can be done using a recoding matrix $\mathbf{M}$ with dimensions $k \times l$ where $k$ is the number of classes from the original multi-class problem and $l$ is the number of attributes which have been decomposed from these classes to formulate binary classifiers. 1s and 0s represent for each row, or class, what attributes that class possesses. Further decomposition strategies are presented by Lorena et al. (2008) [11]. These researchers discuss more robust methods of codifying multi-class problems into binary problems using matrix decomposition methods for $\mathbf{M}$ and using Directed Acyclic Graphs. Further investigated in our own study, we decompose classes using the more simple and straightforward method presented by Allwein et al.

An application of this type of classification is style transfer. A CNN will possess learned features of the particular style of a piece; thus, this allows for the style to be transferred onto any inputted image. In 2015, Gatys et al. [6] investigated this approach by creating a feature space which captures information about the texture of the inputted pieces. Their model is a version of the VGG-Network with 16 convolutional layers and 5 pooling layers. Interestingly, for image creation, the researchers found that replacing max pooling with average pooling improves the optimization of the loss function. Ultimately, the CNN that they developed was able to transfer the style of any painting onto any inputted image, including other paintings. Building on this paper was work done by Johnson et. al. (2016) which conducted real-time style transfer using feed-forward CNNs paired with a per-pixel loss function [10]. Style transfer or even saliency maps (images which also utilize per-pixel loss gradients to showcase what areas of the painting influence its classification decision) can highlight the defining features of different media.

## 3. Methods

We use 5 different CNN architectures to tackle the problem of identifying painting medium categories. 3 of these architectures are structured to solve a multi-class classification problem. 1 architecture is structured to solve 10 binary classification problems and use these results to construct final class predictions. The final architecture was proposed to be a mixed model that first performed multi-class classification and then used the binary classifier on any categories over which it performed better. For all of these architectures we built upon training functions and accuracy function provided in CS231N assignments. Further, we highly modified the CIFAR Pytorch Vision Dataset codebase to create our own VisionDataset class for our TATE multiclass dataset and our TATE binary datasets.

### 3.1. Multi-class classification

The inputs to each multi-class model are 3x224x224 RGB images and their labels are one of 10 medium description categories ('Chalk and graphite on paper', 'Engraving on paper', 'Etching on paper', 'Graphite and watercolour on paper', 'Graphite on paper', 'Ink on paper', 'Line engraving on paper', 'Oil paint on canvas', 'Pen and ink on paper', and 'Watercolour on paper'). The outputs for each

multi-class model are scores for each of the 10 medium description categories. For each multi-class model, we use a softmax classifier with cross-entropy loss (defined below):

$$L_i = -log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$$

$L_i$ gives the loss for each example $i$, $f_{y_i}$ is a the score for that example's class, and $f_j$ is the score of the class $j$.

#### 3.1.1 Baseline

Our baseline model for training is a simple CNN with the following structure:

1. a convolutional 2D net with 32 filters of size 5x5, stride 2, and padding 2
2. a ReLU layer
3. a MaxPool 2D layer with kernel size 2
4. a convolutional 2D net with 32 filters of size 3x3, stride 1, and padding 1
5. a ReLU layer
6. a MaxPool 2D layer with kernel size 2
7. a convolutional 2D net with 32 filters of size 3x3, stride 1, and padding 1
8. a ReLU layer
9. a MaxPool 2D layer with kernel size 2
10. a linear layer that outputs 10 scores one for each medium

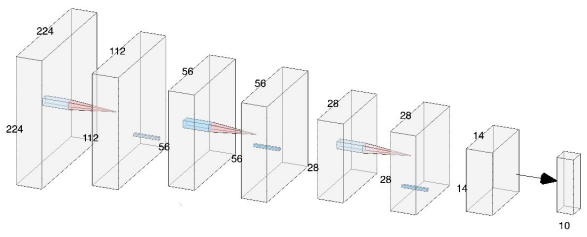


Figure 1. Baseline Model

We used an SGD optimizer with a learning rate of 2e-3, momentum set to 0.9 with Nesterov momentum enabled.

#### 3.1.2 ResNet18

For our next model, we use the ResNet18 architecture. This model is a deep residual neural network with 18 deep layers (72 individual layers total). With its presence of multiple deep layers, the issue of vanishing gradients arises, causing the degradation of accuracy as He et al. described. ResNet uses residual blocks which crucially include skip connections which connect past certain training layers and thereby solve the problems of vanishing gradients. In each residual block, the difference (or residual) between the output from the skipped layers and the input to the skipped layers needs to be learned. In other words, if $H(\mathbf{x})$ is the mapping of the output from the skipped stacked layers with $\mathbf{x}$ being their input, the residual function across the skipped layers can be approximated by $F(\mathbf{x}) := H(\mathbf{x}) - \mathbf{x}$. Then, the output function becomes $F(\mathbf{x}) + \mathbf{x}$ which abstracts away the original complexity of the skipped layers [7].

For our own use of ResNet18, we utilized pretrained weights from ImageNet and fine-tuned the network's learning rate and optimizer. Our final model uses a learning rate of 4e-3 and the SGD optimizer.

#### 3.1.3 ResNet18 with Data Augmentation and Frozen Gradients (Augmented ResNet18 or ResNet18 (Aug))

As we will see later, the ResNet18 architecture seemed to be overfitting. As a result, we first used data augmentation which was not in the initial data preparation described in the next section, including random horizontal and vertical flips with probability .5 and auto augmentation [5]. Further, we froze the gradients of every other layer in ResNet18 in order to attempt to reduce some of the overfitting. Once again, we use pretrained weights from ImageNet and fine-tune the network's learning rate and optimizer. Our final model uses a learning rate of 7e-3 and the SGD optimizer.

### 3.2. Binary Classification

While paintings fall under common combinations of media or medium description categories like 'Chalk and graphite on paper', 'Engraving on paper', 'Etching on paper', 'Graphite and watercolour on paper', 'Graphite on paper', 'Ink on paper', 'Line engraving on paper', 'Oil paint on canvas', 'Pen and ink on paper', and 'Watercolour on paper', these categories share overlapping media that give different categories common features. For example, 'Chalk and graphite on paper', 'Graphite and watercolour on paper', and 'Graphite on paper' all contain graphite.

Since medium description categories are combinations of different media, another way to approach the problem of categorizing art by medium description categories is by using multiple binary tags (chalk, graphite, paper, engraving (non-line), etching, watercolour, ink, line engraving, oil

paint, pen). That is, we run 10 binary classifiers "tagging" if each image includes a medium or not and then convert the combination of these tags assigned to each image into a medium description category. This was done using the recoding matrix suggested by Allwein et al. (2000) [1]. The intention for this was to reduce confusion for the model when it attempts to learn features for pieces engaging multiple types of media — breaking down classification into separate media should increase accuracy.

For each medium (chalk, graphite, paper, engraving (non-line), etching, watercolour, ink, line engraving, oil paint, pen), we take 3x224x224 RGB images as inputs and recode their labels to indicate the presence or absence of the medium. We use ResNet18 with pretrained weights on these inputs. Our outputs are scores for the presence and absence of the given tag. We again use a softmax classifier with cross-entropy loss with the same formulation defined previously, but this time where $L_i$ gives the loss for each example $i$, $f_{y_i}$ is a the score for that example's class (presence or absence), and $f_j$ is the score of the class $j$ (presence or absence). Running 10 of these classifiers, we get predictions for the presence/absence of all 10 media tags. We then convert these tags to back to medium description categories and which we use to calculate overall classification accuracy of medium description classes.

### 3.3. Mixed model

Finally, we propose using a multi-class classification model to first categorize works, followed by a binary classifier for categories the binary classifier performs better in. As we will see in our results though, due to lower performance in the binary classifier, the result of this mixed model ended up just being the ResNet18 with Data Augmentation and Frozen Gradients.

## 4. Dataset

The basis for our constructed dataset is artwork from the Tate collection, a national collection of British art from 1500 to the present as well as international modern and contemporary art. The Tate Collection releases almost 70,000 images of its art on its various web pages under the Creative Commons Public Domain CC0 licence [12].

While the collection has almost 70,000 released images, only 10 medium categories have over 500 images available. Thus, our team constructed a dataset of 5,000 images, 500 randomly sampled per each of the following medium categories: 'Chalk and graphite on paper', 'Engraving on paper', 'Etching on paper', 'Graphite and watercolour on paper', 'Graphite on paper', 'Ink on paper', 'Line engraving on paper', 'Oil paint on canvas', 'Pen and ink on paper', and 'Watercolour on paper'.

Each image was collected from its individual artwork web page on the Tate website at an image size that was

Figure 2. Giorgio de Chirico's *The Uncertainty of the Poet* — example oil painting on canvas from Tate Collection dataset [13]

larger than 224x224 pixels. Fortunately, the Tate website has images of multiple scales. Since the images are of varying sizes, they were then center cropped to 224x224 pixels and converted to tensors with RGB channels. We cropped these images because the medium constitutes the material used to generate the image and thus should be able to be detected from a subsection of said image. We chose a center crop rather than a random crop in order to allow for results that could be replicated for our test set In order to stay consistent with the format of our test data, we also chose to center crop our training data. We made this decision also because photographs of the art center the image and this crop ensures that the art itself is in the crop. Images were separated into training, validation, and test sets in an 70/10/20 split. We then normalized the images based on channel means and standard deviations. Cross-validation was used during training, with each validation set having 500 images. Medium classifications for each were obtained via metadata provided by Tate.

## 5. Experiments

### 5.1. Hyperparameters

For each of our models, except the baseline, we fine-tuned learning rates and optimizers using our validation set. We use learning rate 2e-3 for the baseline, 4e-3 for ResNet18, 7e-3 for ResNet18 with data augmentation and frozen gradients, and 2e-3 for binary classification, trying

learning rates between $10^{-5}$ and $10^{-1}$. We also fine-tuned with both Adam and the SGD optimizer, ultimately choosing the SGD optimizer because of higher performance. We used mini-batch size 64 as we did in class assignments.

## 5.2. Evaluation

We used overall medium description classification accuracy to evaluate all of our models:

$$Accuracy_{overall} = \frac{\sum_i \mathbb{1}(y_{pred_i} == y_{true_i})}{n}$$

where $y_{pred_i}$ is the medium description class prediction from the model and $y_{true_i}$ is the true medium description class for example $i$ and $n$ is the number of examples.

We additionally used by medium description class accuracy to compare our best multi-class model to our binary classification model in an effort to develop a mixed model:

$$Accuracy_j = \frac{\sum_i \mathbb{1}(y_{true_i} == j) * \mathbb{1}(y_{pred_i} == j)}{\sum_i \mathbb{1}(y_{true_i} == j)}$$

where $j$ is a medium description class, $y_{pred_i}$ is the medium description class prediction from the model and $y_{true_i}$ is the true medium description class for example $i$.

Finally, we use tag accuracy to explore how accurately our models can detect the presence of a particular medium in a work:

$$Accuracy_k = \frac{\sum_i \mathbb{1}(y^k_{pred_i} == y^k_{true_i})}{n}$$

where $y^k_{pred_i}$ is the tag presence/absence prediction from the model for tag $k$ and $y_{true_i}$ is the true tag presence/absence for example $i$ and $n$ is the number of examples.

## 5.3. Results

### 5.3.1 Overall performance analysis

We see that by overall metrics shown in Table 1, ResNet18 models (ResNet18 and ResNet18 with data augmentation and frozen layers) does significantly better than the baseline model and the binary tag model. Our baseline model was able to achieve $74.54\%$ on the training set, $57\%$ on the validation set, and $56.5\%$ on the test set. However, even with parameter fine-tuning and experimenting with small deviations in numbers of layers and filter sizes, this model was not able to break $57\%$ accuracy.

In contrast, our ResNet18 models achieved perfect accuracy (indicating some overfitting) on the training set and accuracies around $75\%$ for the validation and test sets – almost $20$ percentage points higher than our baseline model. Our regular ResNet18 model showed steep learning in training sets and validation results in the first few epochs which then seemed to plateau. In order to deal with perhaps some overfitting, we augmented data with random flips and auto augmentation and froze the gradients in half of ResNet18's layers. Unfortunately, while this model resulted in marginally better results, we can see in the training/validation accuracy graphs that even our updated ResNet18 quickly learned and then plateaued in an almost identical way to ResNet18.

Finally, we see that binary tag accuracy is extremely low even though it uses the ResNet18 model. This is because tagging for each medium independently allows for $2^{10}$ media combinations. That is, it allows for medium attribute predictions that are not in the medium's original descriptive class to be classified as present in the piece. Even one inaccuracy in the tagging will cause the whole classification to be marked as incorrect, making it more difficult to attain accurate medium description predictions.

| | Baseline | ResNet18 | ResNet18 (Aug) | Binary |
|---|---|---|---|---|
| Train | .7454 | 1 | 1 | 1 |
| Validation | .5700 | .7660 | .7760 | .4760 |
| Test | .5650 | .7430 | .7490 | .4770 |

Table 1. Final Accuracies for Tested Models

### 5.3.2 By medium description class performance analysis

While overall accuracy is a helpful metric for understanding the efficacy of our model, since some of our medium description classes are more similar to each other (and thus may be confused for each other), looking at by class accuracies and confusion matrices is also important to examining the performance of our models.

In Fig 3, we see that our best model – the ResNet18 with data augmentation and frozen gradients – performs best on classes where there is a distinctive medium such as oil paint on canvas and chalk and graphite on paper. On the other hand, we see that 'graphite and watercolour on paper' works were not classified as well, perhaps because they shared 'graphite', 'watercolour', and 'paper' with other classes.

Finally, we see that performance for 'line engraving on paper' and 'engraving on paper' is very poor. Since line engraving is a specific type of engraving, in light of these results, we might reevaluate that the initial assumption of our model that the class a work is assigned in the dataset is the only appropriate label for the work.

Interestingly, we would expect subtle differences between different medium description classes as a result of similarities between the media themselves or these combinations to make these classification tasks more difficult for humans as well. Thus, by examining our by class accuracies, we're able to see that our model actually performs

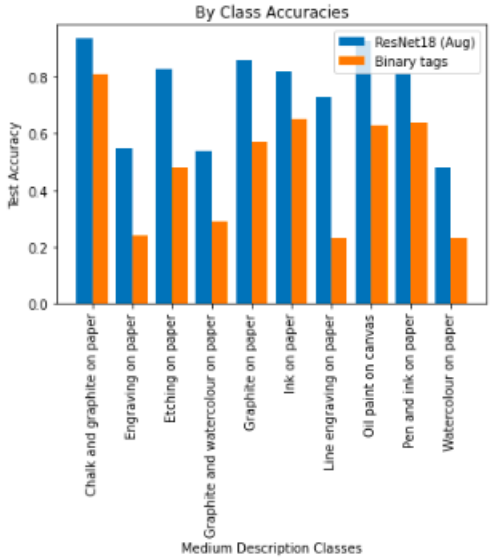


Figure 3. By medium description class accuracies for ResNet18 augmented model and binary classification model

quite well on tasks where there's more of a perceivable difference between classes.

The confusion matrix for our ResNet18 with data augmentation and frozen gradients model as seen in Figure 7 aligns with these explanations. Most notably, we see that Engraving on paper is very often assigned the incorrect label of line engraving on paper and similarly line engraving on paper is often (although less often) assigned the label of engraving on paper. This may indicate either that the perceivable difference between these two classes is actually quite difficult or that there might be more overlap between class categorizations than indicated by the dataset. In Figure 4, we see how similar the media in line engraving on paper and engraving on paper works look to the human eye. Future work may include manually reviewing labeled classes with alternate class assignments.

In addition to highlighting similarities between different media included in medium description categories, the confusion matrix also reflects the problem that categorizing works into mixed media categories can pose. As we predicted, 'graphite and watercolour on paper' works most often get confused with 'graphite on paper' and 'watercolour on paper' works, most likely picking up some but not all of the features of the image. Further, because we crop the image, our model might not be getting all of the media present in an image which would affect its accuracy. For example, in Figure 5, we see an example of a work that is labeled as 'graphite and watercolour on paper' misclassified as 'graphite'. Visually, this image appears to be only graphite to the human eye. Thus, this model might actually have some efficacy at not only classifying entire works with medium descriptions but also identifying what media are present in different regions.

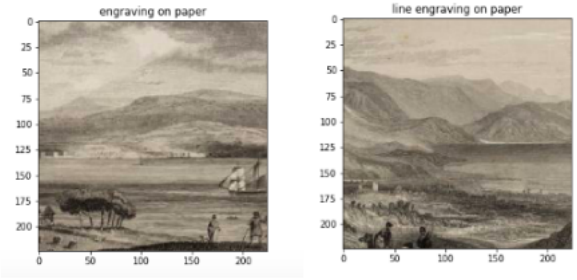


Figure 4. Engraving and line engraving images that don't seem discernibly different media to the non-domain expert human eye

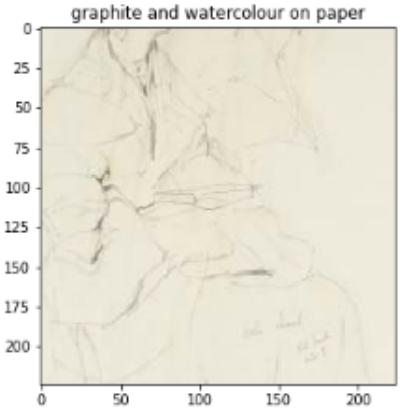


Figure 5. Graphite and watercolour on paper work misclassified by augmented ResNet18 as graphite that appears to be only graphite to the human eye

Finally, with the category 'watercolour on paper' we see a mix of similarity between media and similarity between combinations of media affecting our classifications. Most often, 'watercolour on paper' is misclassified as 'graphite and watercolour on paper' pointing to shared similarity of 'watercolor' between the combinations of media in these medium description classes. Next often, 'watercolour on paper' is misclassified as 'oil painting on canvas', indicating that the similarity in the two painting media (watercolour and oil) might make it more difficult for the model to classify the artworks.

Because of some of these difficulties were posed by the similarities between combinations of different media, we also proposed using multiple binary classification models for the medium description classification task. As we saw

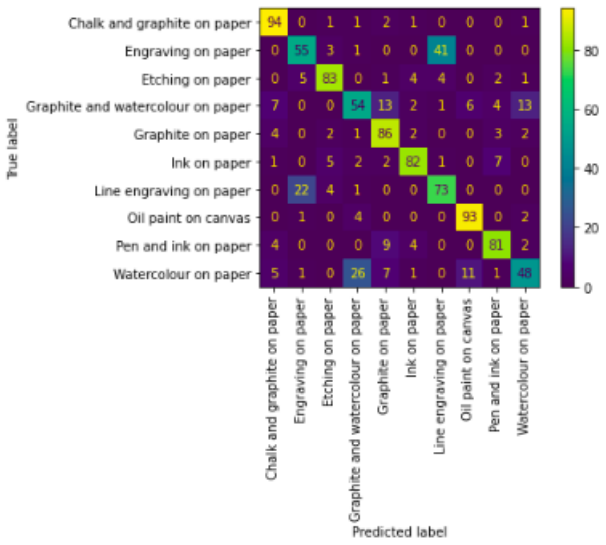


Figure 6. Confusion Matrix for 10 different medium classifications

in Table 1 and in the by class accuracies in Figure 3, this binary classification model did not do as well as our multi-classification model. As a result, our mixed model, we just our best multi-classification model. While the results of our binary classification model were disappointing in relation to our multi-classification model, their results reveal more about the classification problem itself.

For example, the binary class model's by medium class classification accuracies mirror the performance of the multi-class models in the relative performance for each class, with the exception of the 'line engraving on paper' class which performed much more poorly in the binary class model. This is because many of the works rated as 'engraving on paper' marked positive both for 'line engraving' and 'engraving (non-line)' as they were run as binary tags rather than mutually exclusive categories as they are treated in the dataset. The works in Figure 4 were both classified as positive for 'line engraving' and 'engraving (non-line)', aligning with our personal experience as human evaluators of media. Thus we can once again see the utility of using domain experts in the future to reexamine the data labels in this dataset.

### 5.3.3 Salency map analysis

Confusion in classification can be extended though an analysis of salency maps. Using a per-pixel loss function, salency maps will show which areas of an image had the most say in dictating the chosen class for that image. The following figure shows saliency maps for 4 example paintings, each with different media ('pen and ink on paper,' 'etching on paper,' 'watercolor on paper,' and 'oil paint on canvas').

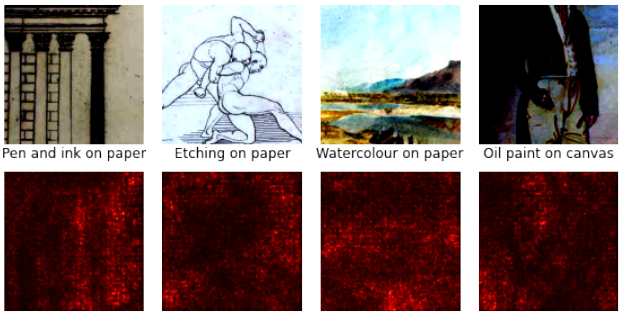


Figure 7. Saliency Maps for 4 different media

Running multiple saliency maps, we found many of the same trends shown in each of these images for different classes. Noticeably, for classes like pen and ink on paper and etching on paper, saliency (red markings) is found in the blank or white areas of the artwork. This is most likely due to these classifications being differentiated by their lack of color and sparsity of marking on paper. Particularly in the pen and ink example, one can clearly map columns of red in the salency map with the columns of exposed paper in the artwork. Contrastingly, red in the salency map is most present where there is filled in color in the watercolor and oil painting images.

### 5.3.4 By medium tag performance analysis

While the multiple binary models didn't perform as well on the medium description classification task, individual tagging is actually quite accurate since the binary classifier for each tag is only trained on whether a piece will contain that tag or not (i.e. if a piece contains graphite or not vs. if it is a piece on paper containing graphite and watercolor). This abstraction of classification yields strong validation and test accuracies, as shown below in Table 2 and in Figure 8.

| | Train | Validation | Test |
|---|---|---|---|
| chalk | 1 | .9760 | .9500 |
| graphite | 1 | .8900 | .8670 |
| paper | 1 | .9620 | .9770 |
| engraving (non-line) | 1 | .8820 | .8790 |
| etching | 1 | .9080 | .9380 |
| watercolour | 1 | .8740 | .8860 |
| ink | 1 | .9320 | .9410 |
| line engraving | 1 | .9100 | .9230 |
| oil paint | 1 | .9620 | .9550 |
| pen | 1 | .9440 | .9530 |

Table 2. Binary Tag Classification Accuracies

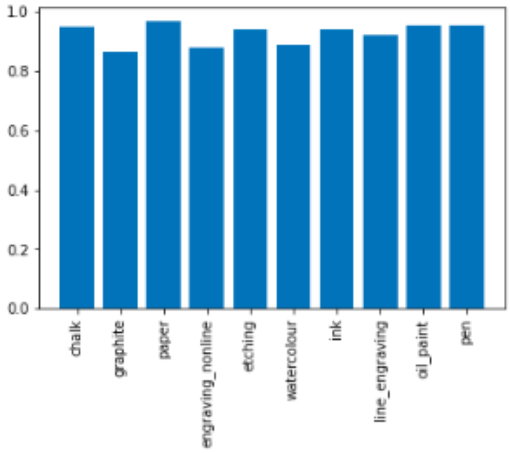


Figure 8. By tag binary classifier test accuracies

That being said the accuracy of these tags for different medium description classes once again parallels some of the performance we saw with the ResNet18 augmented model. In Figure 9, we see how accurately each of our binary models was able to tag works in different medium description tasks. Immediately it stands out that for 'line engraving on paper' and 'engraving on paper', the 'engraving (non-line)' and 'line engraving' tags respectively perform very low (indicating they show false positives at a high rate).

Further, we see that our medium tags seem to be picking up other media tags that are commonly associated with the tag. For example, we see that 'graphite' performs poorly for 'watercolour on paper', indicating that the graphite tag is most likely learning some features from the 'graphite and watercolour on paper' works that are associated with 'watercolour' rather than only 'graphite'.

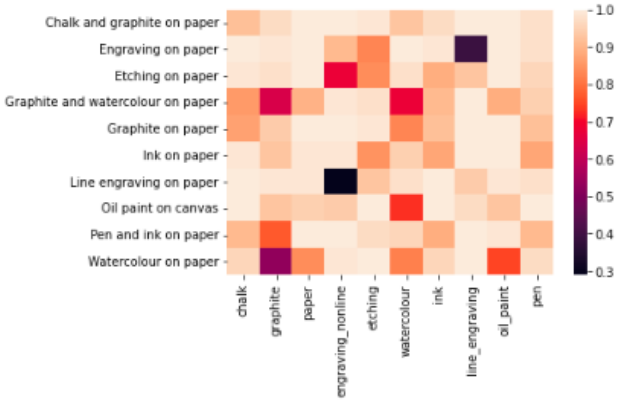


Figure 9. Binary tag model accuracies by medium description class

## 6. Conclusion

In this paper, we investigated different CNN architectures to aid in the classification of art pieces from the Tate collection by medium. We ultimately found the ResNet18 with data augmentation and frozen gradients to yield the highest test accuracy (while binary classification yielded the highest accuracy for classifying by individual tags). The final ResNet model most likely performed best due to its robust layer architecture with skip connections, and binary classification performed well for individual tag classification due to the reduced complexity from the original classes.

Additionally, we provided ample analysis which further highlighted the intuitions behind the model's classifications. We were able to pinpoint where misclassification occurred and why through confusion matrices, as well as understand the reasonings for valid classifications through saliency maps. Ultimately, this project has great implications for the field of art curation. With machine classification, museum curators can automate the tagging of pieces with their associated metadata. Additionally, machine classification provides insight into what are the most representational qualities of certain media, or what is each medium most marked by.

For future work, we would want to engage more robust models with more advanced hyperparameter tuning to increase classification accuracy or experiment with different croppings/resizing of our data (or even expanding our dataset itself). Additionally, we would want to extend the capabilities of the per-pixel loss function from the saliency maps to try and tackle style transfer. Ultimately, however, our work provides a rich analysis to just slightly demystify how CNNs operate in classification tasks.

## 7. Contributions & Acknowledgements

Jane handled dataset construction/processing, development of the multiclass models, and confusion matrix visualizations. Shridhar handled the binary classification modeling and the saliency map visualizations. Both team members trained and ran models and wrote the report.

Special thanks to Sanchit Jain who was involved in the initial stages of determining our problem statement and explored ways in which to extract features from images in a domain-informed way. While we did not use this feature extraction in this current work, this could be the basis for future extended work.

Finally, thank you to the CS231N class for providing code for training and checking accuracies and PyTorch CIFAR-10 dataset source code (PyTorch reference), all of which we adapted significantly for our own models.

## References


[1] Erin L Allwein, Robert E Schapire, and Yoram Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of machine learning research*, 1(Dec):113–141, 2000. 2, 4

[2] Sam Baker. Finding a way to make digitizing art collections profitable. 1

[3] Sugata Banerji and Atreyee Sinha. Painting classification using a pre-trained convolutional neural network. In *International Conference on Computer Vision, Graphics, and Image processing*, pages 168–179, 2016. 1

[4] Eva Cetinic, Tomislav Lipic, and Sonja Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, 114:107–118, 2018. 2

[5] Dandelion Mane Vijay Vasudevan Quoc V. Le Ekin D. Cubuk , Barret Zoph. Autoaugment: Learning augmentation strategies from data. *https://arxiv.org/pdf/1805.09501.pdf*, 2019. 3

[6] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv*, 2015. 2

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 3

[8] Yiyu Hong and Jongweon Kim. Art painting identification using convolutional neural network. *International Journal of Applied Engineering Research*, 12(4):532. 1

[9] Shuqiao Huo. Moods in classical paintings: An ai based classification approach. In *Journal of Physics: Conference Series*, volume 1650, page 032102. IOP Publishing, 2020. 2

[10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 2

[11] Ana Carolina Lorena, André CPLF De Carvalho, and João MP Gama. A review on the combination of binary classifiers in multiclass problems. *Artificial Intelligence Review*, 30(1):19–37, 2008. 2

[12] Tate. Tate images. 4

[13] Tate. 'the uncertainty of the poet', giorgio de chirico, 1913. Jan 1970. 4

[14] Wentao Zhao, Dalin Zhou, Xinguo Qiu, and Wei Jiang. Compare the performance of the models in art classification. *Plos one*, 16(3):e0248414, 2021. 2