Using CNN to Classify and Understand Artists from the Rijksmuseum

Tara Balakrishnan, Sarah Rosston, Emily Tang
{taragb, srosston, emjtang}@stanford.edu

Motivation

Art museums are digitizing their collections so there is a real need for tools that can improve this process, for both museum curators and visitors [1]. To further this goal, the Rijksmuseum, located in Amsterdam, has provided a public dataset consisting of 112,039 art photographs as part of the Rijksmuseum Challenge [1]. Previous work on this dataset includes PigeoNET which is a CNN based on AlexNET which achieves 78% accuracy for the artist classification problem [2].

Problem

We attempt to learn how well our network understands artist styles and visualize what the network learns about the images. We evaluate the performance of our model both quantitatively and by using network visualization tools such as saliency maps, which we compare to known facts about the artists.

Dataset

The Rijkmuseum dataset contains paintings by 12,641 unique artists, 3,949 of whom only have one piece of artwork in the dataset and 23 of whom have over 100 pieces in the collection. For our problem, we took the top 10 artists who had the most images in the dataset, which was over 3,000 images per artist.

Bernard Picart George Breitner Isaac Israels Jan Luyken J. Tavenraat

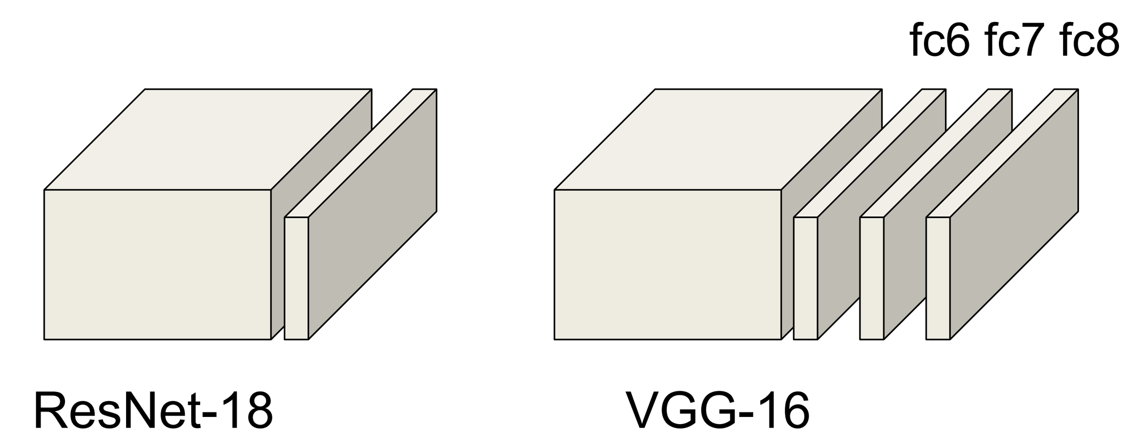
Marius Bauer Reinier Vinkeles Rembrandt Willem Witsen Simon Fokke

References

1. T. Mensink and J. van Gemert. The rijksmuseum challenge: Museum-centered visual recognition. In ACM International Conference on Multimedia Retrieval (ICMR), 2014.
2. K. Simonyan, A. Vedaldi, and A. Zisserman. "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014.
3. N. van Noord, E. Hendriks, and E.O. Postma. Toward discovery of the artist's style: Learning to recognize artists by their artworks. IEEE Signal Process. Mag., 32(4):46–54, 2015.
4. Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European conference on computer vision. Springer International Publishing, 2014.

Approach

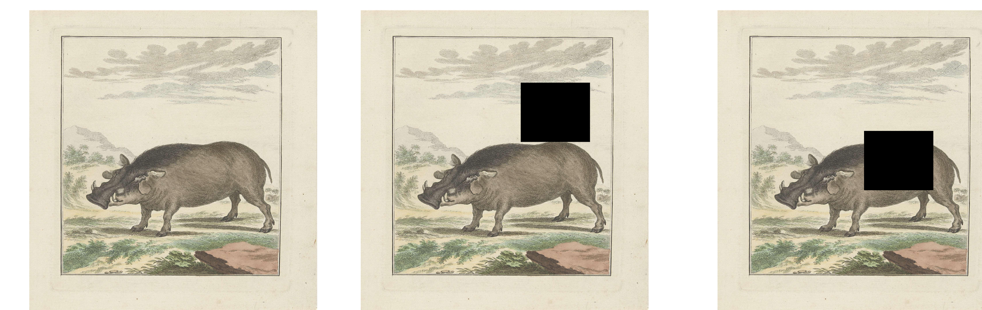
Transfer Learning



To build our CNN model, we apply transfer learning. We use pretrained ResNet-18 and VGG-16 models on ImageNet. We finetune the last layer of ResNet-18, and the last 3 layers of VGG-16.

Visualizing and Understanding Our Network

Occlusion



We apply the sliding window technique with a black box that occludes parts of the test image, and visualize the importance of each pixel in predicting the correct class with heatmaps [3].

Saliency Maps

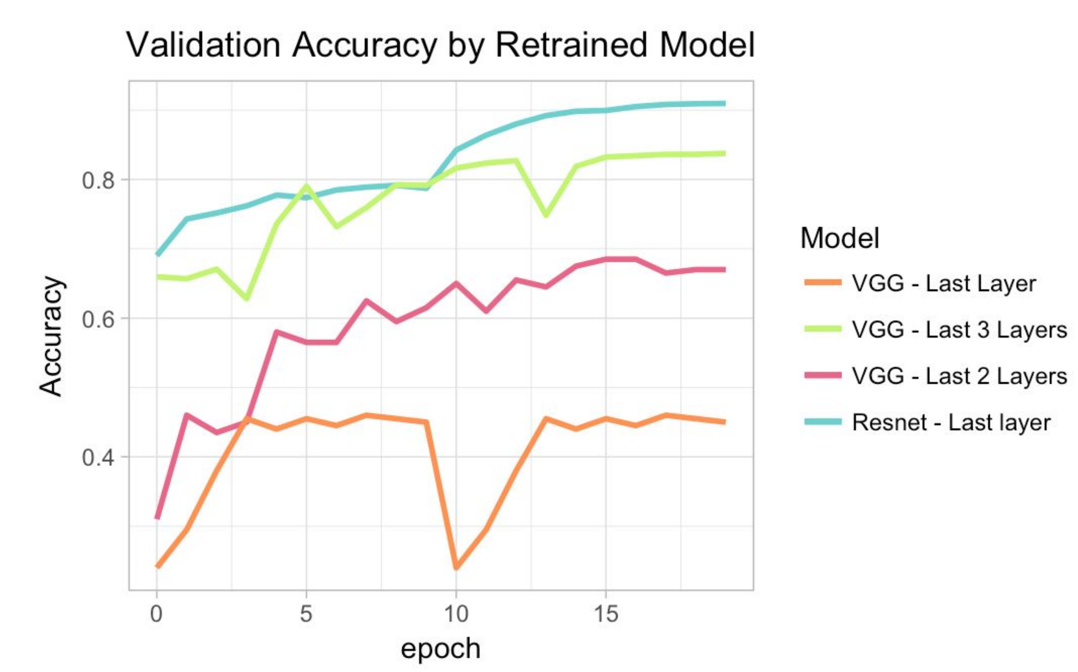
To compute saliency maps, we take the maximum value over the RGB channels of the gradient of the correct scores of an input image [4].

Results

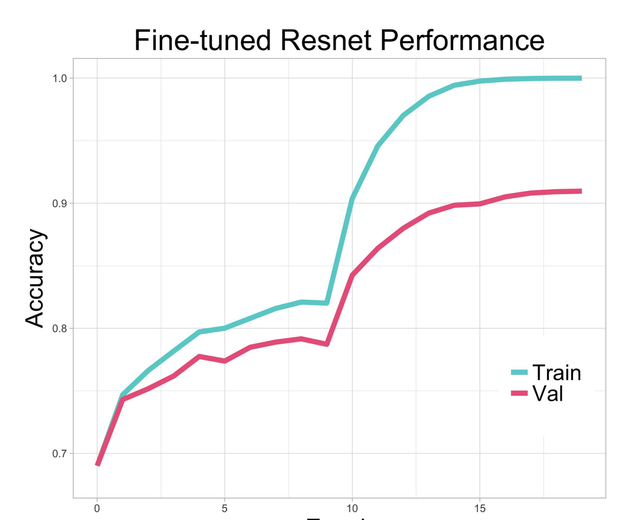
Fine-tuned CNN Performance

Model	Train Acc.	Validation Acc.
VGG-16	0.918636	0.8377031
ResNet-18	0.9999207	0.9096425

Final ResNet-18 Test Accuracy: 0.907523



Resnet Model



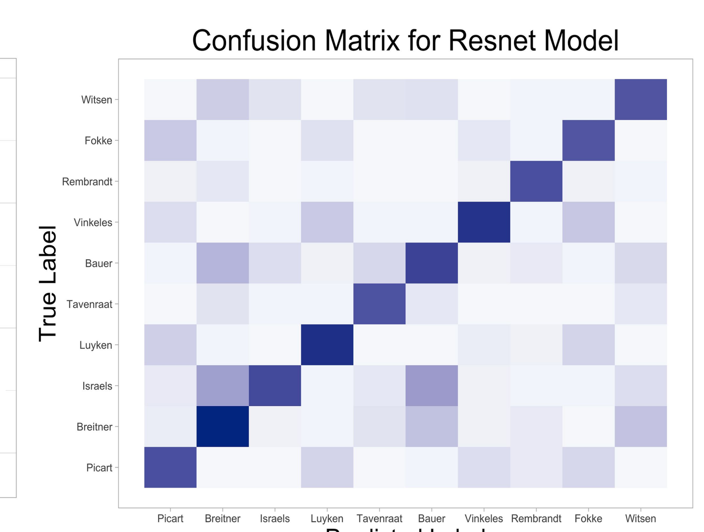
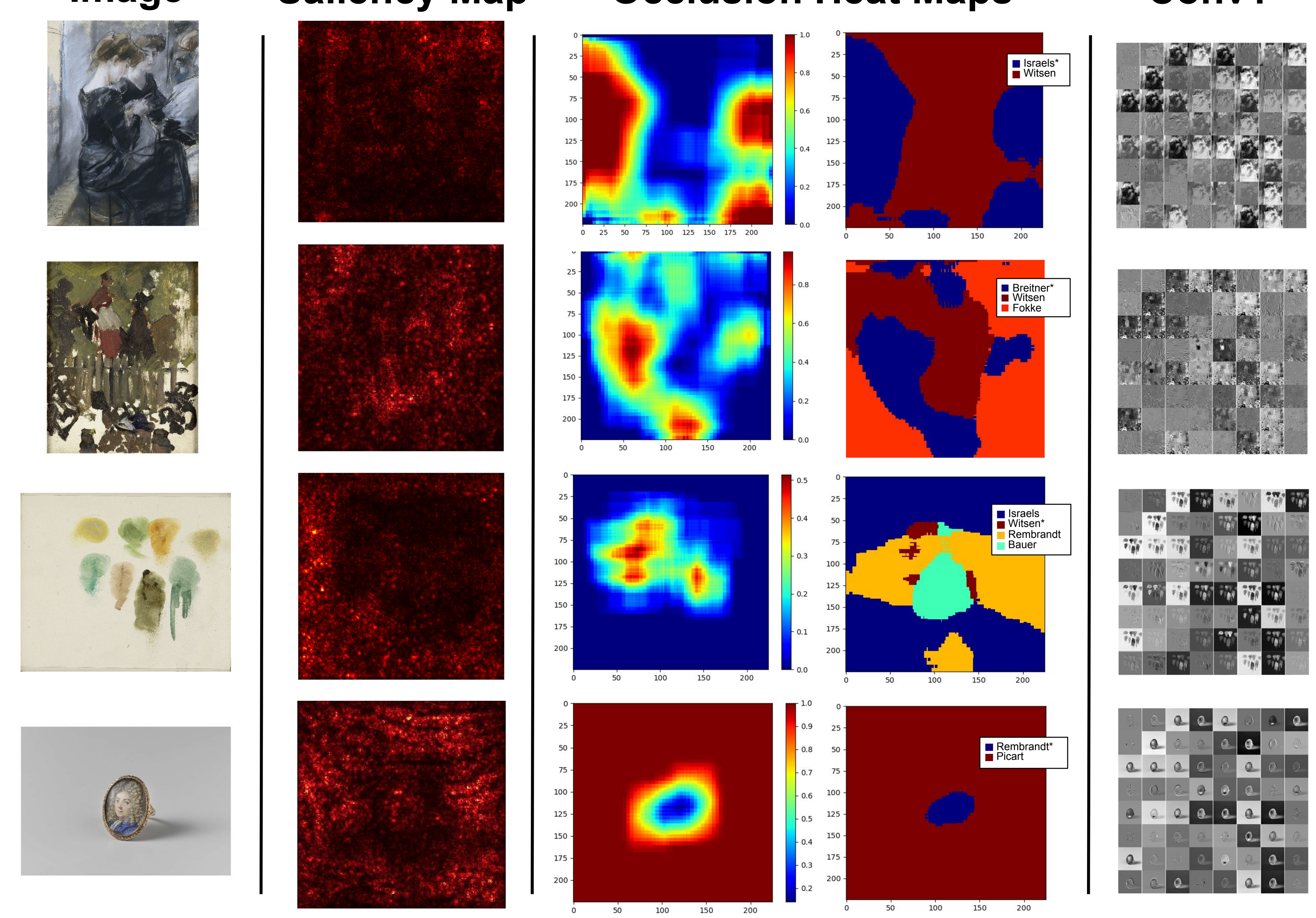


Image | Saliency Map | Occlusion Heat Maps | Conv1



Analyzing Rembrandt. Rembrandt was famous for his self-portraits and portraits; his use of light and atmosphere draws attention to the subjects. Our network's understanding of this concept is evident in our saliency maps and occlusion heat map (see right), where higher importance is given to the subjects' faces.