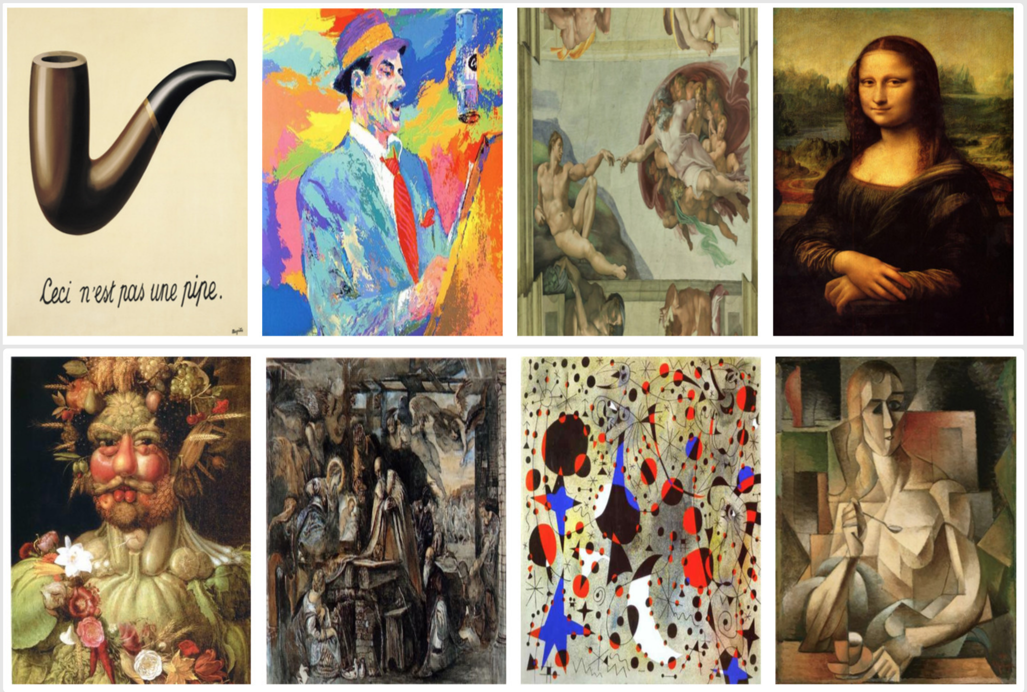
Ceci n'est pas une pipe: A Deep Convolutional Network for Fine-art Paintings Classification

Wei Ren Tan[1], Chee Seng Chan[2], Hernán E. Aguirre[1], and Kiyoshi Tanaka[1]

[1]Shinshu University, Japan. [2]University of Malaya, Malaysia.



Motivations

- Recent years, vast digital paintings have been made available across the Internet and museum
- Paintings analysis through machine learning became an important task to aid curators in their daily work routine
- We want to learn meaningful features from paintings

Challenges

- Small training data
- Many paintings are non-representative nor figurative
- Paintings analysis requires other background knowledge, e.g. history

Goals

- Train an end-to-end Convolutional Neural Network (CNN) for large-scale *style*, *genre*, and *artist* classification
- Investigate the capability of CNN in learning features of fine-art paintings
- Visualize the learned features

Wikiart Paintings Dataset[2]

- Collection of $> 80,000$ fine-art paintings ranging from 15th century to modern times.
- 27 *styles* from **all** paintings.
- 10 *genres* with $> 1,500$ paintings ($\sim 65,000$ samples).
- 23 *artists* with > 500 paintings ($\sim 20,000$ samples).

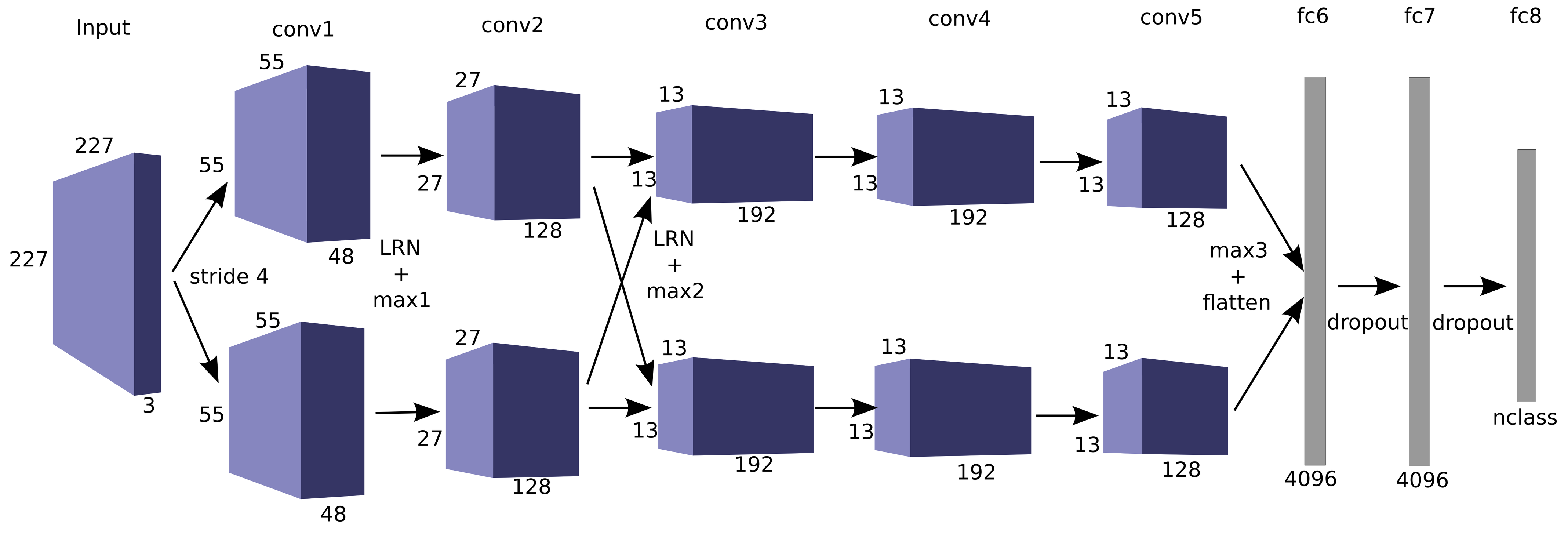
References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, pages 1097–1105, 2012.

[2] B. Saleh and A. Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint arXiv:1505.00855*, 2015.

Convolutional Neural Network

Overview of Architecture:



AlexNet[1]

Training details:

- Learning Scheme: SGD

$$v_{i+1} = 0.9 \cdot v_i - \epsilon \cdot \left(\left\langle \frac{\partial L}{\partial w} \right\rangle_{B_i} + 0.0005 \cdot w_i \right)$$

$$w_{i+1} = w_i + v_{i+1}$$

- Init. ϵ: 0.01 (non-fine-tuning) and 0.001 (fine-tuning)
- Mini-batch size: 128
- ϵ reduction: factor of 10 / 5,000 iterations
- Max. Iter.: 20,000 iterations.

Data Augmentations:

- Centered raw RGB values
- Image translation
- Image size: 227×227
  - Random cropped during training
  - Centered cropped during testing
- Horizontal reflection

Fine-tuning:

- Pre-trained on ImageNet dataset
- Last layer (fc8) is replaced with new SoftMax or SVM layer

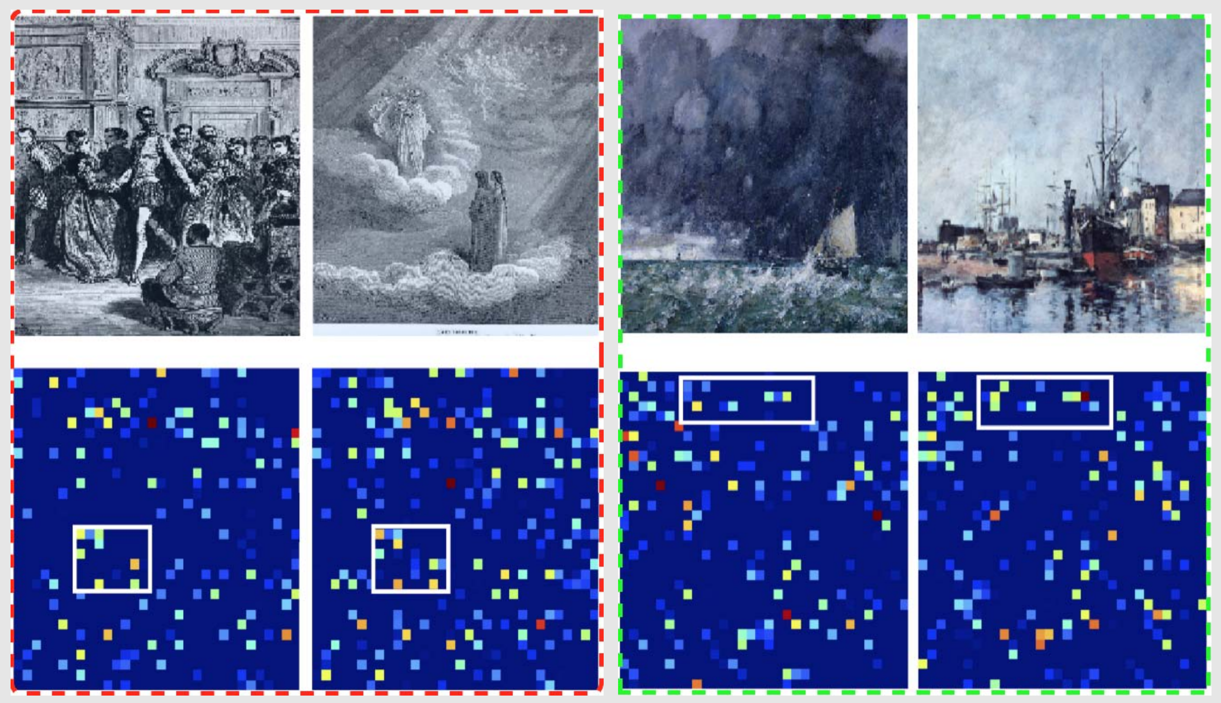
Experimental Results

Model	Accuracy (%)				Size
	Style	Genre	Artist	Overall	
CNN	42.96	65.45	54.39	54.27	61M
CNN-nofine	45.95	69.24	67.02	60.74	61M
CNN-SVM	44.17	69.18	67.17	60.17	61M
CNN-1000	43.56	68.38	64.55	58.83	61M
CNN-finetune	**54.50**	**74.14**	**76.11**	**68.25**	61M
CNN-fc6	51.51	72.11	74.26	65.96	44M
CNN-1024	53.38	73.75	76.02	67.72	48M
CNN-PCA-SVM [2]	21.99	49.98	33.62	35.20	-
Saleh and Elgammal [2]	45.97	60.28	63.06	56.44	-

- Fine-tuning from ImageNet pre-trained network (CNN-finetune) yields best performance.
- SoftMax vs SVM (CNN-nofine vs CNN-SVM) have similar performance
- Preserving fc8 (CNN-1000) does not help.
- Reducing parameters (CNN-fc6 and CNN-1024) only deteriorate $\sim 2\%$ accuracy.
  - **Insight: Better pruning strategy may compress the network without affecting accuracy**

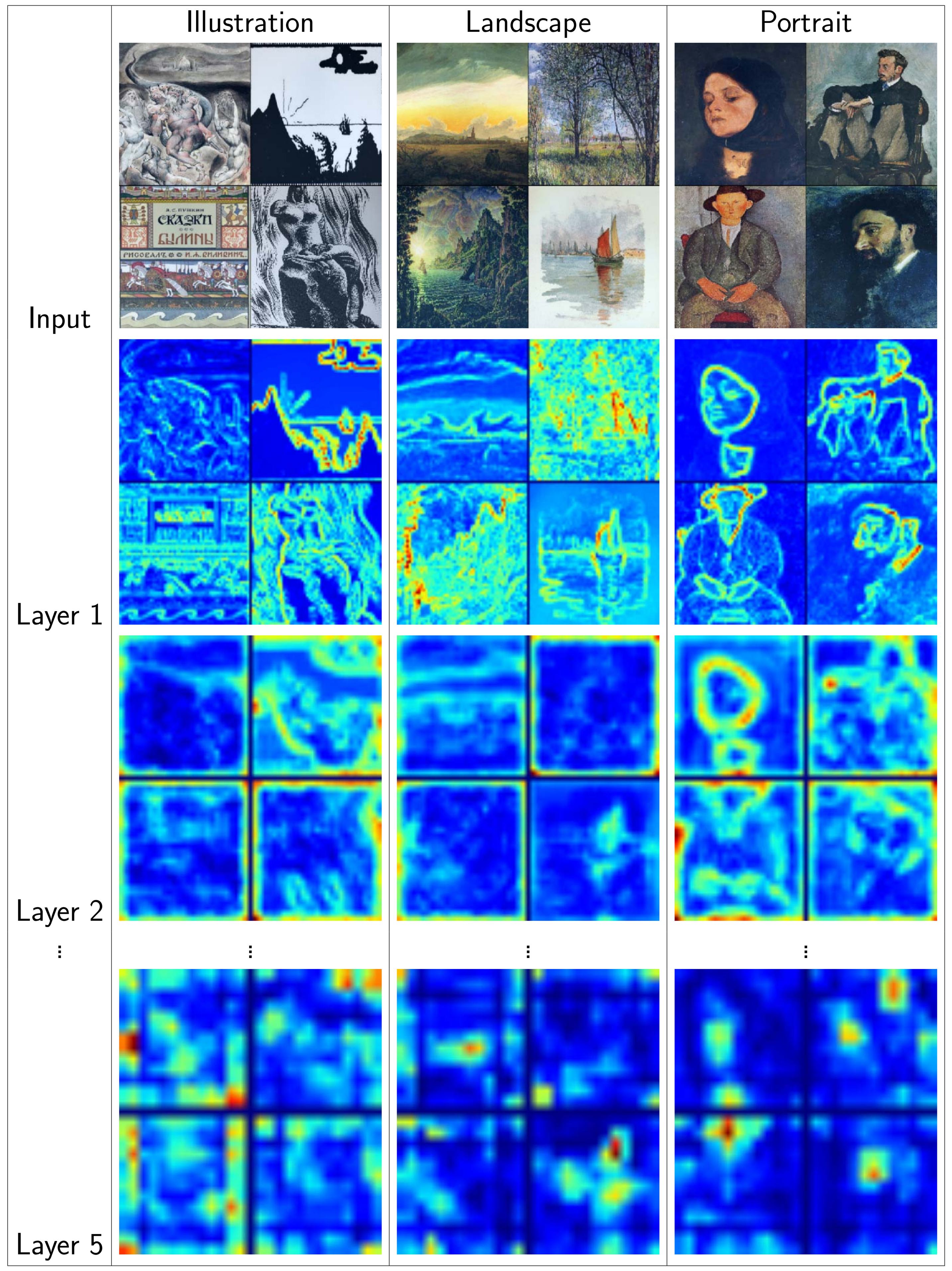
Future works

- To design a better model to learn features from fine-art paintings, and possibly semantically relate them together
- To investigate different visualization techniques for better understanding of how CNN extracts features from paintings.
- To design a generative model that is able to reconstruct and draw the paintings



(a) Gustave Dore (b) Eugene Boudin

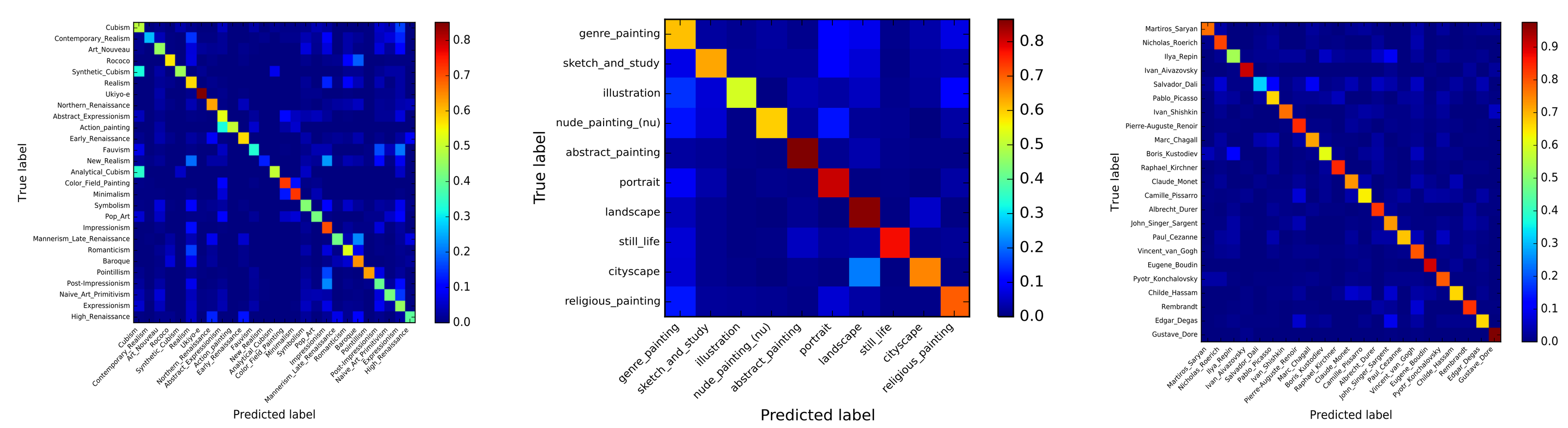
Neurons' Responses Visualization



Observations from the visualizations:

- Same group does not necessarily have similar features activations
- For more structured paintings, CNN tends to find key objects or shapes for cues

Confusion Matrix



(a) Style (b) Genre (c) Artist

- In *styles* classification, poor performance is caused by relationship between styles:
  - Synthetic cubism vs analytical cubism (same root)
  - Rococo vs Baroque (historically related)
- In *genres* classification, top performers are related to other classification problem:
  - Portrait → face/human detection
  - Landscape → scene recognition
- In *artists* classification, artists that are recognized with high precision prefer certain **techniques** or **objects**:
  - Gustave Dore → engraving, etching, and lithography
  - Eugene Boudin → marine and seashore