MATH 6380p Project 1: Feature Extraction and Transfer Learning

M Usman Maqbool Bhutta[1], Candi Zheng[2] and Lan Yuan[2] {20419087, 20489185, 20485165} {mumbhutta, czhengac, ylanaa}@connect.ust.hk
[1]: Department of Electronic and Computer Engineering, HKUST [2]: Department of Mechanical and Aerospace Engineering [3]: Department of Mathematics, HKUST

1. Introduction

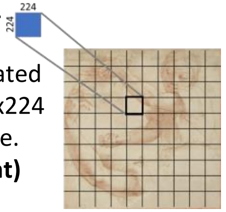
For this project, the Raphael dataset is examed [1]. We compared the performance between transfer learning and ScatNet and discuss the factors affecting in the final performance.

2. Raphael dataset

This dataset contains a 28 digital painting of Raphael or forgeries. There are 13 genuine, 9 not sure and 6 disputed images.

Methodology

We processed the image by cropping the edges and removed stamps manually. We resized all images with avg. dimension of 3000 as scale 1. And further resized to scales 0.5,0.25 and 0.125 respectively. Two different types of methods are used for chopping the image.

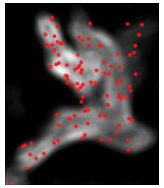


- **Sequentially(Seq)**

We sequentially generated the patches of size 224x224 as shown in above figure.

- **Entropy Random(Ent)**

We calculated the local entropy, and the cropping center (red point) is selected with probability proportional to it.



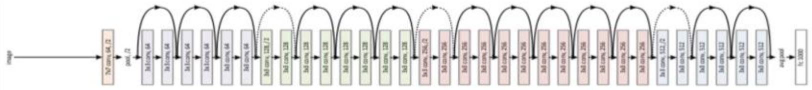
After chopping, patches with top 50 % variance are selected.

3. Methods

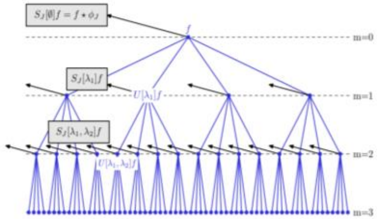
Feature Extractor:

VGG16: This model is a classic convolution neural network, pre-trained on ImageNet with 16 convolution layers and small respective field. We froze the weights for network and deleted the final fully connected layer to make it an feature extractor.

ResNet18: It's a residual learning framework with 18 layers. We do the same operation as that in VGG16 and the weights of this pre-trained model also trained on ImageNet.



Resnet-18 network structure



ScatNet structure

Scat Net: Wavelet scattering network (Scat Net) computes a translation invariant image representation. With an image X, we use 2-layer Scat Net to output a 241 dim feature vector.

Classification: We use three methods (SVM, Logistic Regression, Random Forest) to train a classifier on our features generated from the above feature extraction methods.

Cross Validation: Finally, We use image-wise Leave-One-Out cross validation to calculate accuracy patch-wise, and calculate AUC(area under curve) based on the accuracy.

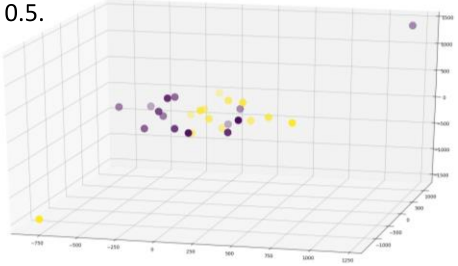
4. Results

The best results for ResNet, VGG and ScatNet are shown in Tabel. ResNet actually outperform VGG and ScatNet in both entropy random cropping (best at scale 0.125) and sequential cropping (best at no resize), with AUC 0.73, scale and 0.72 respectively using logistic regression.

For entropy random cropping, Resnet + LR, AUC decrease from 0.73 to 0.66 with respect to scale from 0.125 to 1. As for sequential cropping, an AUC extreme value 0.72 appears at scale 0.5.

Feature	Cropping	Estimator	AUC
ResNet	Ent	LR	**0.73**
VGG	Seq	LR	0.71
ScatNet	Seq	LR	0.68

Best AUC for each feature extractor



Visualization of ResNet 18 feature based on TSNE

Source Code: https://github.com/UsmanMaqbool/MATH-6380p

5. Analysis

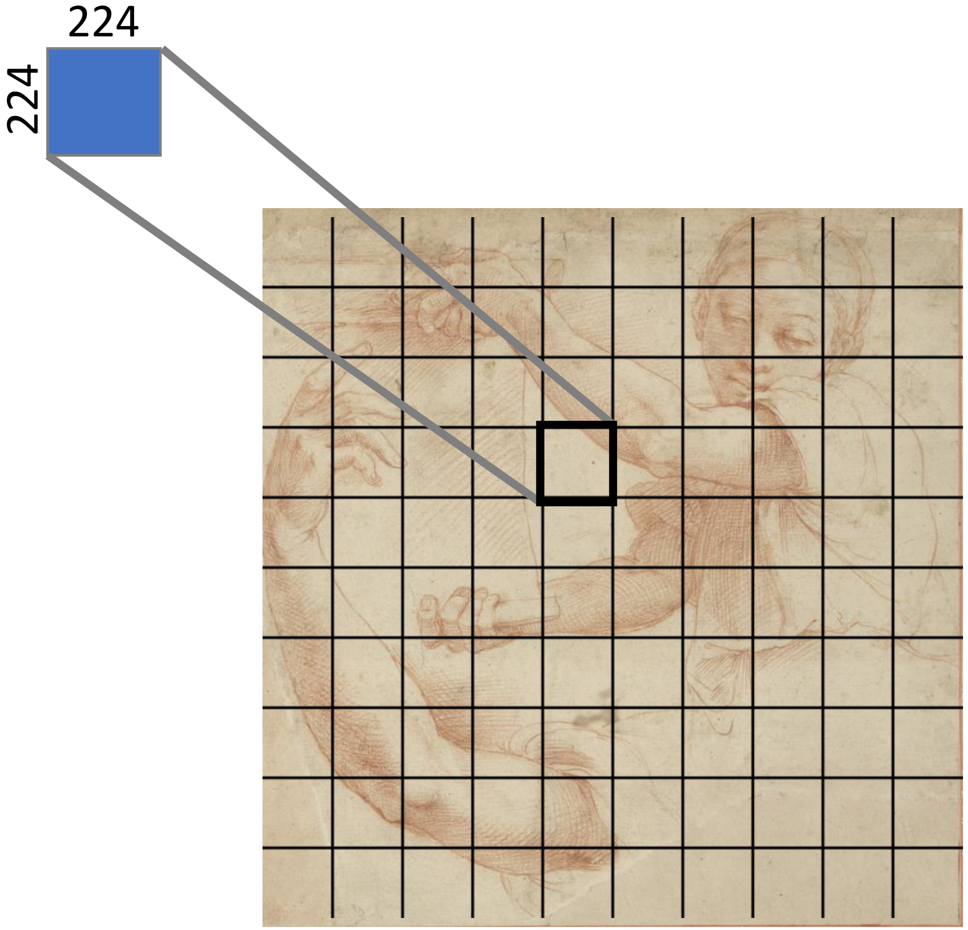
Two main factors, chunks number, reception filed (affected by scale), determines the final outcome. Entropy random cropping (Fixing the chunks number) achieve the best performance at smallest scale (largest reception field) due to the full coverage of picture. For same reception field (scale), large chunks number reduce the variance and increase performance. Keeping the coverage (reception field x number), Sequential cropping result varies, the outcome gets and extreme value at a balanced reception field and chunk numbers.

6. Conclusion

Using Transfer learning and Scat Net, we successfully captured information to distinguish between fake/genuine Raphael's painting. Transfer learning is slightly better than ScatNet, but ScatNet is more invariant to scale change due to its multiscale nature. Cropping the picture while cleaning it using variance/entropy is crucial. A large number of crops (~ 800 each), or a large reception field (about half the picture with randomness crop ~100), helps to get a meaningful result, and they interact with each other to create an balance point. It is hard to achieving large reception field and large number of data without affecting relative independence. Focusing on one approach is more promising to achieve a better performance.

7. References

[1] Dataset provided by Prof. Yang Wang, HKUST
[2] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[3] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
[4] Bruna, Joan, and Stéphane Mallat. "Invariant scattering convolution networks." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1872-1886.
[5] Yosinski, Jason, et al. "How transferable are features in deep neural networks?." Advances in neural information processing systems. 2014.


224
224