https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-317

Image Quality Assessment Using Computer Vision

Zhi Li**[a]**, Palghat Ramesh**[b]**, Chu-Heng Liu**[c]**;
[a] ***School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907;***
[b] ***Palo Alto Research Center, Palo Alto, CA 94304;***
[c] ***Xerox Corp., Webster, NY 14580***

Abstract

For decades, image quality analysis pipeline has been using filters that are derived from human vision system. Although this paradigm is able to capture the basic aspects of human vision, it falls short of characterizing the complex human perception of different visual appearance and image quality. In this work, we propose a new framework that leverages the image recognition capabilities of convolution neural networks to distinguish the visual differences between uniform halftone target samples that are printed on different media using the same printing technology. First, for each scanned target sample, a pre-trained Residual Neural Network is used to generate 2,048-dimension vision feature vector. Then, Principal Component Analysis is used to reduce the dimension to 48 components, which is then used to train a Support Vector Machine to classify the target images. Our model has been tested on various classification and regression tasks and shows very good performance. Further analysis shows that our neural-network-based image quality model learns to makes decisions based on the frequencies of color variations within the target image, and it is capable of characterizing the visual differences under different printer settings.

Introduction

Image quality assessment is very important for electronic imaging systems. It allows us to predict the quality of the image, so that it can be maintained, controlled and possibly enhanced before production or further processing[1]. Hence, a reliable image quality paradigm is crucial in the development of image processing systems.

There have been many efforts developing image quality assessment methods and tools in the imaging community. In general, the methodology of image quality assessment can be divided into two schools: subjective methods [2], where human viewers are involved to evaluate the quality of images, and objective methods [3], where the numerical metrics are calculated from the image. Theoretically, subjective methods tend to be more accurate and reliable, as images are ultimately viewed by human viewers. However, subjective methods are not practical due to time and labor cost. Thus, objective image quality metrics are highly desired in the imaging industry to predict the quality of an image as close as possible to the subjective assessment.

Traditional objective image processing-based image quality assessment pipeline [4] for image noise use filter-based techniques to focus on the structure within the certain (pre-determined) frequency range, and analyze the image linearly channel by channel. However, such approaches based on linear combination of filters may not be best suited for the task of modeling human perception of printed image noise.

Neural network based computer vision model has shown great potential to mimic human vision in many fields, and increasing research efforts are being put into using computer vision to do image quality assessment [5]. In this paper, we use computer vision (CV) models to build appearance-based metrics to evaluate the visual image quality of printed images, in particular, the micro uniformity [6] in a print. We aim to characterize the visual noise that is perceived by human viewers from printed images. Our model leverages recent developments in deep neural networks and machine learning to mimic noise perception of the human vision system. Our experiment shows that the proposed model achieves high accuracy for both subjective and objective micro uniformity assessment tasks. Furthermore, to demonstrate the viability of the proposed model, we show the effectiveness of the off-the-shelf neural network features for decision making in various image quality assessment tasks.

Research supported by PARC and Xerox Corp.

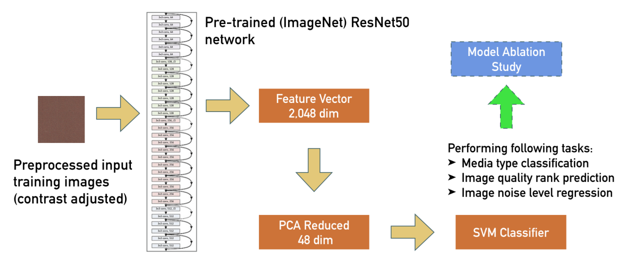

Figure 1: Overall framework of computer vision based image quality analysis system.

Data Acquisition

Our dataset consists of 75,000 RGB image patches (224x224x3) from 600 dpi scans of full page CMYK halftone inkjet prints on 6 different papers. These 6 media types and their respective printing conditions were chosen to produce nearly identical colors but with various levels of graininess noise according to human perception and analytical image analysis. In the context of this paper, paper type is just a convenient label for various levels (classes) of image noise.

The digital image patch used in this experiment is shown in Fig. 1. Note that the dimension of the image is chosen to fit in the input layer of Residual Neural Network. Also, the contrast of scanned images are adjusted and normalized to control the environment and reinforce our pipeline to learn from the noise and nothing else.

Machine Learning Based Image Quality Assessment

The modeling framework can be described as follow: We use a pre-trained Residual Neural Network ResNet50 [7], to retrieve a 2,048 dimension feature vector from an input RGB image. We then use the Principal Component Analysis (PCA) to further reduce the feature dimensions to 48. Finally, the reduced feature is input to a support vector machine that has been trained for various objective and subjective image quality assess-

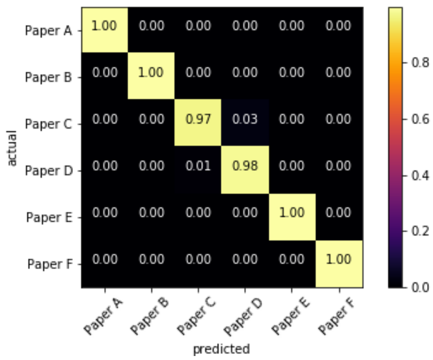

Figure 2: Confusion matrix on test data for media classification.

ment tasks such as media type classification, visual noise metric, and IQ noise ranking.

Media Type Classification

As noted previously, the media type is a surrogate for different levels of image noise. Due to the different physical and chemical attributes of the media and the physics of ink-media interaction, the printed images have different IQ noise appearance on different media. Therefore, we aim to use discriminant features extracted from a Convolution Neural Network (CNN) to pinpoint the source media of image.

We train a multi-class Support Vector Machine (SVM) classifier to predict the correct media types for given images. We split the dataset into 80% and 20%, and train the classifier on the 80 %. We show the results in Fig. 2. It shows that our model achieves higher than 95% accuracy across all paper types. This suggests that the noise patterns for each paper type are distinct and the CV model successfully learns to recognize them. Compared with traditional IQ metrics, where only real values are produced as the representation of the quality, our model can further approach human vision system by reflecting the type of source based on distinct image visual appearance.

Image Quality Ranking

In addition to the media classification capability, the proposed model also shows ability to perform pair wise image quality comparison, and it can produce an overall subjective IQ rankings through out all the studied images.

First, we had a human expert rank the images in order of preference from 0, the best image quality, to 5, the worst image quality. We verified that the ranking is generally consistent for images of the same media type, so we assigned the same rank to all images of a media type. Then, we trained a set of five binary SVM classifiers, and combined them to produce discrete number from 0 to 5 as image quality ranking. Fig. 3 shows the final prediction accuracy. Overall ranking prediction is higher than 95%, except for paper C which appears have some confusion with paper D. Coincidentally, papers C and D are from the same manufacturer.

Note that although the rankings are given by experts by paper types, hence the rankings are directly associated with the paper types, there are differences between media classification and image quality ranking prediction. For media type classification, we consider every class independent, and there is no correlation assumed between classes. However, the pair wise ranking prediction is ordinal regression, and the algorithm needs to find the correct order based on image quality and find the right boundaries between two classes that have similar image quality performance. Fig. 3 also shows the difference between media type classification and

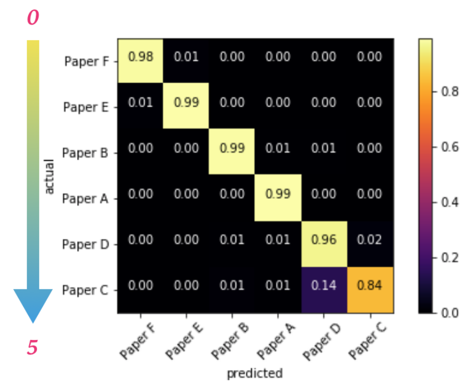

Figure 3: Confusion matrix on test data for image quality ranking. Here we keep all paper types in order of image quality. 0 means the best image quality, and 5 means the worst image quality.

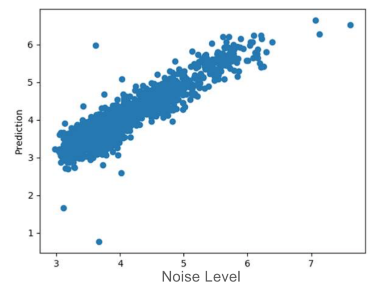

Figure 4: Plot of ground truth noise measure vs prediction from noise level regressor. The x-axis is the ground truth, and y-axis is the prediction from our regression model.

ranking prediction. Comparing Fig. 2 and Fig.3, we can see that ranking prediction has more confusion between papers C and D, but little confusion can be observed in the media type classification.

Image Noise Level Prediction

Finally, in the context of objective image quality analyses, we consider the ability of our image quality assessment model to quantitatively predict the noise level. We use an industry standard image quality tool to measure the visual high frequency noise level from patches, and trained a SVM regression model to predict it.

The regression results on test data are shown in Fig. 4. A linear correlation between ground truth and prediction can be observed. The R-Squared score of our regression model is about 0.86, suggesting a good agreement between objective metric and metric inferred from the neural network model.

Model Ablation Study

To validate and demonstrate the viability of model, we aim to further examine the proposed model. Although, explaining and examining the machine learning classifier, including traditional machine learning methods and deep learning methods, has been drawing more and more attention recently, most of the classifier explainers found in recent literature [8, 9] are focused on explaining spatially localized features for generic image recognition or text classification tasks. However, this is not applicable in our study. All the images in our experiment look roughly the same;

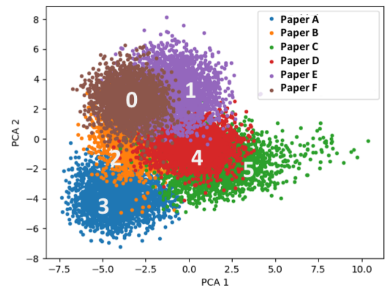

Figure 5: Plot of data points by 1st and 2nd PCA components. Data points are color coded by media types, and image quality rankings from human expert are marked on the different point clouds.

the distinct features of interest is the noise (i.e. texture), which is not local. Therefore, we propose two approaches to explain our model: (a) by checking the separability of features generated by the pre-trained neural network and (b) by determining the confidence level change of the SVM classifier as the noise structure is altered.

Effectiveness of Neural Network Vision Features

Although the IQ assessment pipeline achieves high accuracy using the ResNet50, further test on the effectiveness of the neural network features is still desired for following reasons:

- **Networking repurposing:** It has been shown that the hierarchical combination of convolution layers in the neural networks is capable of extracting localized features. However, for image quality assessment of noise, no localized semantic features are included in the images.
- **Dataset difference:** Our pre-trained network was trained on ImageNet [10] dataset. ImageNet contains millions of images labeled with semantic labels, and it is designed to train and benchmark classifiers to do image and object recognition. Therefore, no information regarding image quality is given during the network training.

Therefore, the effectiveness of neural network features needs to be validated for IQ assessment. Here, we propose that neural network features are effective if the extracted neural network visual feature can separate the data points by their visual appearance differences.

To demonstrate the separability of the neural network features, we plot data points by first and second PCA components, and the visualization is shown in Fig. 5. We can see that datapoints of the same media type are well clustered, and some separability between different clusters can be observed. In fact, one can argue as more latent features (e.g. PCA components) are included, the point clouds will become completely separable from each other. Hence, near perfect classification results can be expected.

To further enhance the between class separability of the neural network features, we use Linear Discriminant Analysis (LDA) to process the neural network features and visualize the first two LDA components in Fig. 6. Unlike the PCA, the LDA uses the class labels to separate the data points in orthogonal spaces. We can see that a good separability can be achieved using just 2 LDA components. The only overlap observed occurs between papers C and D, which correlates with our previous observation in image quality rankings.

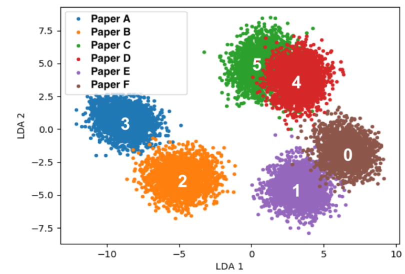

Figure 6: Plot of data points by 1st and 2nd LDA components. Data points are color coded by media types, and image quality rankings from human expert are marked on the different point clouds.

Therefore, we can conclude that neural network features from pre-trained ResNet are effective in discriminating between visual appearances of images required for image quality assessment tasks.

Classification Explaining

We now consider the decision making of the SVM classifier. Unlike most of the classifier explainers, which find localized stimulating regions within an image corresponding to a class label, we attempt to find the relevant frequency ranges in the image that allows the classifier to make its prediction.

Our approach is described as follow: First, we filter original images with a low pass filter. Here we chose filters with cut-off frequency ranging from 0 to 10 cycle/mm. Then, the filtered images are fed into our pipeline to predict the different media types. We monitor the confidence of the true class from the SVM classifier as the cut-off frequency is changed.

Fig. 7 shows the confidence level versus cut-off frequency. The following observations can be made:

1. The confidence of classifier is very low when only near-DC component is included in the image. This is not surprising because that the near-DC component should look very similar on all different media types, which is a constant tone patch of contrast adjusted color.
2. The confidence of classifier is almost at 1 when the cut-off frequency is greater that 10 cycles/mm suggesting there no discriminating features of frequency greater than 10 cycles/mm .
3. The confidence level jump occurs when the cut-off frequency is between 4-6 cycles/mm. Each media type has a distinct response corresponding to the noise frequencies that the classifier has learned to discriminate the media type.

Note that, in Fig. 7, the confidence curve representing paper type B does not go down to near 0 when only near DC component is included. This is caused by the nature of the multi-class classifier. Given an arbitrary image, the classifier is designed to assign one of six labels. Therefore, one of the six labels is selected as default classifier when the classifier is not able to produce any confident decision. In our case, paper B is the default classifier label.

To summarize, we demonstrated that the neural network feature utilized by our image quality assessment model is able to recognize the image quality differences within the test images, and our pipeline makes the classification decision based on the high frequency noise between 4 and 6 cycles/mm. We note that these frequencies are in the visual noise high frequency (VNHF) range typically associated with graininess noise.

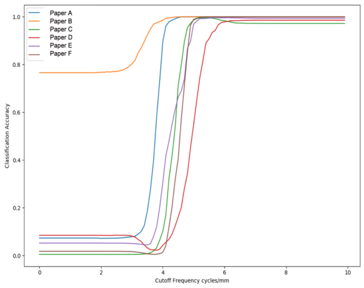

Figure 7: The plot of media type classifier's confidence response corresponding to frequency.

Conclusion

We propose a neural network based image quality assessment model to characterize graininess noise in the printed images. Our pipeline utilizes a pre-trained residual neural network to process the images with various levels of image noise, and extracts features for both subjective and objective image quality assessment tasks, including media type classification, IQ ranking prediction, and IQ noise level prediction. The proposed model achieves high accuracy on the aforementioned IQ assessment tasks without additional neural network training and other processing. Compared with traditional filter based IQ metrics, our model is able to better approximate the human vision system and successfully assess images based on visual appearance.

Furthermore, we propose a new methodology for explaining and examining the IQ assessment classifier. First, We show that visual features extracted from neural networks are effective and highly discriminating for visual appearance and image quality. Then, we show that our trained classifiers are responding to image noise frequencies between 4 6 cycles/mm, which one often attributes to graininess.

Acknowledgments

We thank PARC and Xerox Corp. for supporting this research, and following people for their insights and inspiring discussions on machine learning, image quality and human vision system: Wencheng Wu and Beilei Xu of the University of Rochester, and Mojgan Rabbani of Xerox Corp..

References

[1] K. Thung and P. Raveendran, "A survey of image quality measures," in *2009 International Conference for Technical Postgraduates (TECHPOS)*, Dec 2009, pp. 1–4.

[2] "Recommendation 500-10: Methodology for the subjective assessment of the quality of television pictures," ITU-R Rec. BT.500, 2000.

[3] R. Zeman, W. C. Kress, D. R. Rasmussen, E. K. Zeise, G. Chiu, K. Donohue, and D. Hertel, "Update on incits w1.1 standard for perceptual evaluation of micro-uniformity," 12 2003.

[4] E. N. Dalal, D. R. Rasmussen, F. Nakaya, P. A. Crean, and M. Sato, "Evaluating the overall image quality of hardcopy output," in *Image: Processing, Quality, Capture. Systems (PICS) Conference*, Portland, OR, USA, 1998.

[5] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[6] B. Mishra and R. Rasmussen, "Microuniformity: An image quality metric for measuring noise," in *PICS 2000: Image Processing, Image Quality, Image Capture, Systems Conference, Portland, OR, USA, March 2000*, 2000, pp. 75–78. [Online]. Available: http://www.imaging.org/IST/store/epub.cfm?abstrid=1643

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 770–778. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 1135–1144. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939778

[9] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," in *Deep Learning Workshop, International Conference on Machine Learning (ICML)*, 2015.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *Proceedings of the IEEE Computer Vision and Pattern Recognition, CVPR 2009*, 2009.

Author Biography

Zhi Li is a fifth year doctoral student and teaching/research assistant in the School of Electrical and Computer Engineering at Purdue University, West Lafayette. He works with Prof. Jan Allebach and Poshmark primarily on fashion photography analysis, including fashion photography aesthetics and the autonomous garment color extraction system. He has also been involved multiple research projects such as fashion textural and imagery analysis. The work described in this paper was carried out during his summer internship at Xerox in Webster NY. Beyond academics, Zhi is an active member of the Purdue University Choir since 2014, and the Eta Kappa Nu (HKN) Beta Chapter since 2016. He served as the HKN volunteer director in 2018 Spring.