International Conference on Image Analysis and Recognition, vol. I, pp. 61-65, Jun. 2020.

4K or Not? - Automatic Image Resolution Assessment

Vyas Anirudh Akundy and Zhou Wang

Department of Electrical & Computer Engineering, University of Waterloo
200 University Ave W, Waterloo, ON, N2L 3G1, Canada
Emails: vaakundy@uwaterloo.ca, zhou.wang@uwaterloo.ca

Abstract. Recent years have witnessed a growing popularity of 4K or ultra high definition (UHD) content. However, the acquisition, production, post-production, and distribution pipelines of such content often go through stages where the actual video resolution goes below 4K/UHD level and is then upscaled to 4K/UHD resolution at later stages. As a result, the claimed 4K content in the real world often drops below the intended 4K quality, while final consumers are not well informed about such quality degradation. Here, we present our recent research progress on automatic image resolution assessment methods that determine whether a given image has true 4K resolution or not. Specifically, we developed a largest of its kind database of more than 10,000 true and fake 4K/UHD images with ground-truth labels. We have also made some initial attempts on constructing edge feature, Fourier transform feature, and deep learning based methods for the classification task. We believe that the built database and the attempted methods will help accelerate the research progress on automatic image resolution assessment.

Keywords: image quality assessment · 4K · ultra high definition (UHD) · video resolution · perceptual image quality

1 Introduction

There has been a significant trend in recent years in the media and entertainment industry of producing and delivering 4K or ultra high definition (UHD) content to consumers. Strictly speaking, 4K and UHD represent two different spatial resolutions of 4096×2160 and 3840×2160, respectively, but the UHD resolution is most commonly used in consumer electronics, and thus 4K and UHD are often used interchangeably in practice for 3840×2160 resolution. 4K/UHD videos offer the potential to present significantly increased sharpness and fine details for better quality-of-experience (QoE) of end viewers. However, the acquisition, production, post-production, and distribution pipelines often go through stages where the actual video resolution goes below 4K/UHD level and is then upscaled to 4K/UHD resolution at later stages. Consequently, the claimed 4K content in the real world often drops below the intended 4K quality in terms of their sharpness and fine details, but consumers are often not well informed about such quality degradation. The objective of this research is to develop image resolution assessment methods without access to the pristine-quality reference image that can automatically determine whether a given image has true 4K resolution or not.

Traditional full-reference (FR) image quality assessment (IQA) algorithms such as PSNR and SSIM [1] do not apply because they require access to the reference image. No-reference (NR) methods are desirable but not much NR-IQA work has been dedicated to detecting images whose resolutions have been increased by upscaling from lower resolutions. In [2], a Discrete Fourier Transform (DFT) based technique is proposed with focuses on observing the difference in Fourier power spectra between natural and upscaled images. Natural scene statistics (NSS) based approaches [3], [4] have been developed based on examining the statistical dependencies in natural against artificially generated images. Deep learning approaches have also been explored [6], [5] by detecting manipulation in images during forgeries. Nevertheless, there are still significant gaps in achieving reliable image resolution assessment. In particular, large-scale high-quality databases that cover a wide variety of image content are missing, making it difficult to perform sufficient and convincing validation of objective models. Furthermore, the lack of big data also impedes the development of machine-learning, specifically deep learning, based approaches because of the risk of over-fitting. Therefore, our first focus of this work is on database construction, which is followed by a few first attempts on objective NR resolution assessment algorithms developed upon the database.

2 Database Construction

We develop a large-scale database containing images of real and fake 4K/UHD resolutions together with ground-truth labels. To the best of our knowledge, it is the first and largest database of its kind. The database consists of two classes of images: "True 4K" and "Fake 4K" images. A visual example of a pair of true and fake 4K images are shown in Fig. 1. Both images are cropped and enlarged for visualization purpose. In this particular example, it is apparent that the true 4K image presents more crisp texture details and sharper edges than the fake 4K image. Depending on the native resolution, the up-scaling factor, the up-scaling method, and the image content, the fake 4K images may exhibit reduced perceptual sharpness at different levels.

In constructing the database, the "True 4K" images are acquired by taking videos recorded in 4K and extracting the frames from the videos. The "Fake 4K" images includes two sub-datasets created from two sources. The first dataset is constructed by extracting frames from 1080p resolution (1920×1080) videos, and up-scaling to 4K/UHD resolution. The second set of images is obtained from [9], which is a dataset consisting of 102 classes of flowers with a wide variety of resolutions. This complements the first dataset in that the variation in source image resolution reduces the bias towards the 1080p resolution which has a fixed up-scaling factor of 2 to 4K/UHD images. In both datasets, three up-scaling

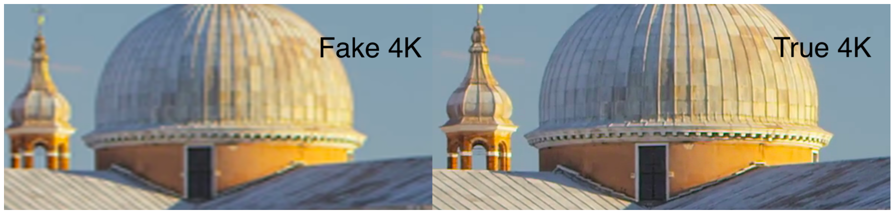

Fig. 1. Sample "true" and "fake" 4K images in the built database. Images are cropped and enlarged for visualization purpose.

filters, bicubic, fast-bilinear and lanczos, have been used to up-scale the images. All operations are performed using the open-source FFMPEG tools.

The full database is divided into two parts with 8,437 and 2,393 images, respectively. The two parts have no overlap in terms of image content, and each contains both "True 4K" images and "Fake 4K" images upscaled from 1080p as well as other arbitrary resolutions (e.g., 667 × 500, 754 × 500, 674 × 500, and 500 × 533) that help improve the robustness of the models being trained and tested with the database. The division is intended for machine learning methods that require independence between a training and a testing sets. The division is flexible in practical use of the database, and does not necessarily follow the suggested division here.

3 "True" and "Fake" 4K Image Classification

Starting from the built database, we make a few first attempts training objective models that classify "True" and "Fake" 4K images. These models are geared towards exploiting potential feature extraction and classification methodologies and are at a premature stage. Diagrams of these methods are illustrated in Fig. 2.

The first method is based on edge features. Specifically, four edge detection filters, Laplacian, Sobel, Prewitt and Scharr, are applied to a test image, and various statistic features are extracted (including variance, mean, median, maximum, among others) from the filtered images. Such features have been successfully used previously in other classification tasks such as shark fish classification [7]. The second method works in the discrete Fourier transform (DFT) domain where a similar statistical feature extraction process is applied. In both cases, support vector machine (SVM) models, namely Model-1 and Model-2 in Fig. 2, are trained to predict the classification labels based on the extracted features. The third method is based on deep convolutional neural networks (CNN), where features learned from other classification tasks (specifically the Inception V3 network [8] features learned for ImageNet large scale visual recognition challenge) are transferred to the current task and the fully-connected layers after the CNN layers are trained for classification.

Using the database division described earlier, we train the three models using the training set (within which 33% of images are used for validation), and test

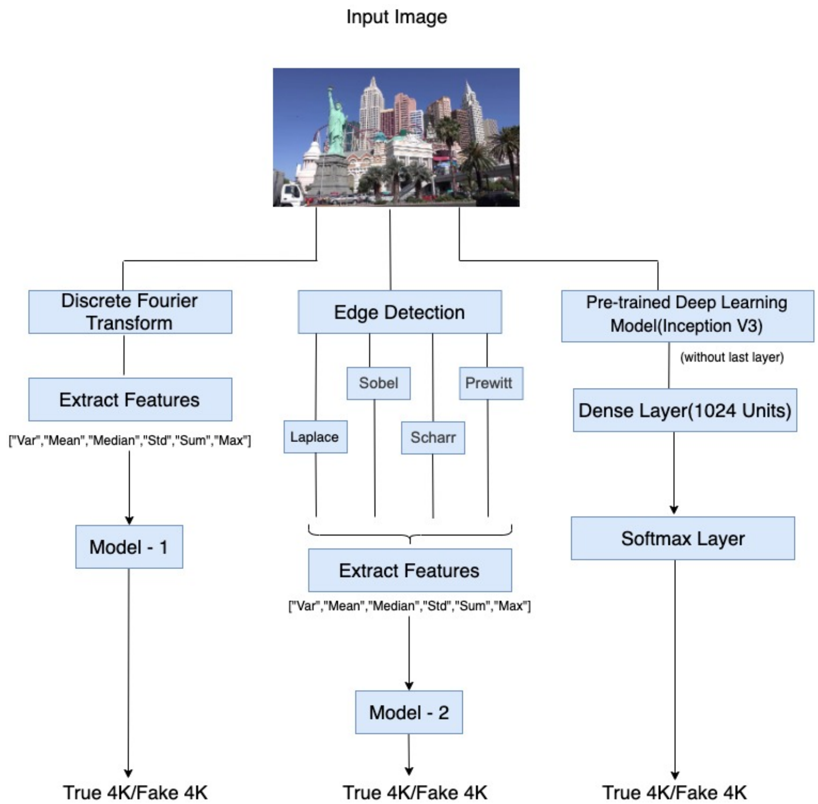

Fig. 2. Diagram of objective image resolution assessment models

the models using the testing set, which is completely independent from the training set. The overall classification accuracy on the testing set is 79.70%, 79.07% and 72.8%, respectively, of the three models. In terms of speed, to test a 4K image, the time required to run the three models are 4.17 second (feature extraction 2.51s and classification 1.66s), 3.96 second (feature extraction 2.19s and classification 1.77s), and 1.66 second, respectively, on a machine with a 1.8 GHz Dual-Core Intel Core i5 Processor, a 8 GB 1600 MHz DDR3, and an Intel HD Graphics 6000 1536 MB. These results are promising as initial attempts, but also leave significant space for improvement, especially the CNN based approach, for which end-to-end training and other network architectures may be investigated. Methods that incorporate both knowledge-driven approaches (such as feature extractions in Model-1 and Model-2) and data-driven approaches (such as the CNN model) may also be combined to improve the classification performance.

4 Conclusion

We present our recent research progress on image resolution assessment, specifically targeting at automated classification of "true" and "fake" 4K image content. We build a first of its kind database that contains over 10,000 "true" and "fake" 4K images with ground truth labels. The database will be made publicly available and is expected to greatly help accelerate the research progress on the topic. We also make several initial attempts in developing image classification methods based on edge features, DFT features and deep CNN predictions. These methods demonstrate promising results but also leave significant space for improvement. Future work includes thorough comparisons with other NR-IQA methods especially those focusing on perceptual sharpness and blur assessment, and further development of advanced methods based on machine learning and perceptual modeling approaches.

References

1. Zhou Wang and A. C. Bovik and H. R. Sheikh and E. P. Simoncelli: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
2. E. Mavridaki and V. Mezaris: 2014 IEEE International Conference on Image Processing (ICIP). No-reference blur assessment in natural images using Fourier transform and spatial pyramids, 566–570 (2014).
3. A. K. Moorthy and A. C. Bovik: Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. IEEE Transactions on Image Processing **20**(12), 3350–3364 (2011)
4. H. R. Sheikh and A. C. Bovik and L. Cormack: No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Transactions on Image Processing **14**(11), 1918–1927 (2005)
5. B. Bayar and M. C. Stamm: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). On the robustness of constrained convolutional neural networks to JPEG post-compression for image resampling detection, 2152–2156 (2017).
6. Bayar, Belhassen and Stamm, Matthew C. : Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security. A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer, 5–10 (2016)
7. Shrivakshan,G. T. and Chandrasekar,C. : A Comparison of various Edge Detection Techniques used in Image Processing. International Journal of Computer Science Issues (IJCSI) **9**(5), 269–276 (2012)
8. Christian Szegedy and Vincent Vanhoucke and Sergey Ioffe and Jonathon Shlens and Zbigniew Wojna : Rethinking the Inception Architecture for Computer Vision. CoRR **abs/1512.00567** (2015).
9. Nilsback, M.-E., Zisserman, A.: Automated flower classification over a large number of classes. In: Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing (2008)