Preprint: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, June 2021

REAL VERSUS FAKE 4K - AUTHENTIC RESOLUTION ASSESSMENT

Rishi Rajesh Shah, Vyas Anirudh Akundy, Zhou Wang

Department of Electrical and Computer Engineering, University of Waterloo
Emails: rr8shah@uwaterloo.ca, vaakundy@uwaterloo.ca, zhou.wang@uwaterloo.ca

ABSTRACT

In recent years, the native 4K/Ultra High Definition (UHD) resolution has been trending towards the new normal of video content creation and distribution, but the practical pipelines of video acquisition, production and delivery often involve downscaling stages where the spatial resolution drops below the 4K level. Even though the video may be upscaled back to 4K/UHD resolution later, the content has lost its authentic resolution. This work aims at authentic resolution assessment (ARA). We first construct a database of over 10,000 real and fake 4K/UHD images. We then develop a two-stage ARA (TSARA) approach that classifies a video frame to have real or fake 4K resolution, where the first stage classifies local patches using a convolutional neural network (CNN), and the second stage aggregates local assessment into a global image level decision using logistical regression. Experimental results show that the proposed approach achieves high accuracy at low computational cost, and outperforms state-of-the-art no-reference (NR) image quality assessment (IQA) and image sharpness assessment (ISA) models. The built database and the proposed method are made publicly available[1].

Index Terms— 4K, Ultra High Definition (UHD), Authentic Resolution Assessment (ARA), Convolutional Neural Network (CNN), Image Quality Assessment (IQA)

1. INTRODUCTION

The technology advancement has enabled the humankind of the 21st century to enjoy image and video content of different resolutions on a variety of devices ranging from mobile phones to Ultra High Definition (UHD) televisions. Although strict 4K (4096×2160) and UHD (3840×2160) represent two different resolutions, in the practice of consumer electronics, 4K and UHD are often used interchangeably for the UHD resolution. In recent times, streaming 4K/UHD or even higher resolution image/video content has been increasing steadily because of the potential to deliver crisp and detail-rich quality-of-experience (QoE) to end-users. In practice, however, the pipeline of video acquisition, production, post-production and delivery often involves stages where video frames are scaled down to lower resolutions, and then upscaled back to 4K/UHD resolution at later stages. As a result, the authentic 4K resolution has been lost in the process while end users are often poorly informed of such quality degradations. For example, it has been shown that a large proportion of the recent UHD Blu-Ray films and TV series do not have true 4K resolution [1]. Here we aim for authentic resolution assessment (ARA), and specifically, the current work focuses on classifying a 4K video frame to have real or fake 4K resolution.

As this research deals with the no-reference image resolution assessment problem, existing conventional full reference (FR) image quality assessment (IQA) techniques such as PSNR and SSIM [2] cannot be applied. On the other hand, no-reference (NR) IQA methods evaluate the quality of a distorted image in the absence of the reference image, and are thus applicable. NR-IQA methods may be categorized into opinion-aware (OR) and opinion-unaware (OU) methods based on whether they are trained on images with human subjective ratings. OA methods may be further classified based on whether handcrafted or learned features are used, where handcrafted features may be derived from natural scene statistics (NSS) models [3], [4], and learned features are obtained from sample images [5], [6]. Recent deep neural network (DNN) based approach also enables joint training of features and quality predictor in an end-to-end fashion [7]. OU methods may employ NSS [8], distortion artifact detection [9] or local binary pattern [10] features, and may also make use of FR-IQA methods to annotate sample images for training [11], [12]. A specific type of NR-IQA methods of high relevance are NR image sharpness assessment (NR-ISA) methods that are designed specifically for evaluating image blur or sharpness. Most of these methods are based on domain knowledge, including human visual system (HVS) [13], [14], Fourier phase [15], complex wavelet domain phase coherence [16], local variation [17], and sparse image representation [18] models.

None of the above mentioned NR-IQA and NR-ISA techniques is dedicated for the image resolution assessment problem. Moreover, large-scale high-quality databases of large spatial images that are essential for learning-based approaches are missing. In this work, we first build a high-quality database of over 10,000 4K/UHD images with ground

[1]Code and dataset access: *https://github.com/rr8shah/TSARA.git*

Fig. 1: Sample "True4K" and "Fake4K" images from the built dataset. Images are cropped and enlarged for visualization.

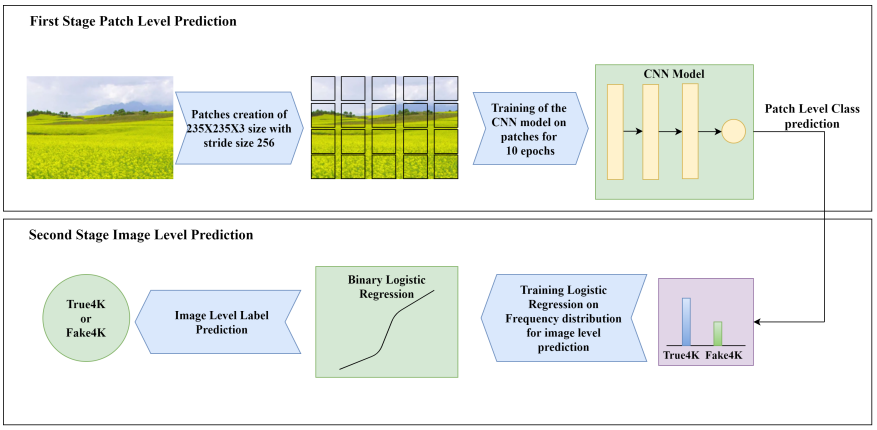

Fig. 2: Flow diagram of the proposed TSARA method for real vs fake 4K classification.

truth labels. We then develop a two-stage ARA (TSARA) method to classify a video frame to have real or fake 4K resolution, where the first stage classifies local patches using a convolutional neural network (CNN), and the second stage aggregates local assessment into a global prediction by leveraging a logistic regression model.

2. REAL VERSUS FAKE 4K CLASSIFICATION

We construct a database consisting of two classes of images labelled as "Fake4K" and "True4K" images respectively. The True 4K images are obtained by extracting frames from videos recorded at 4K/UHD resolutions. The Fake 4K images are produced by upscaling the images extracted from 1080p video and a wide variety of native resolution images of 102 classes of flowers [19]. Three filters, bicubic, faster-bilinear and lanczos from open-source FFMPEG tools, are employed to perform the upscaling operations. A pair of sample "True4K" and "Fake4K" images are shown in Fig. 1. In summary, the full database contains a total of 10,824 images of 4K/UHD resolution, including 5417 "Fake4K" and 5407 "True4K" images. In this work, 60%,20% and 20% of the built database are used for training, validation and testing, respectively. More details of the database are provided in our earlier short progress report in [20].

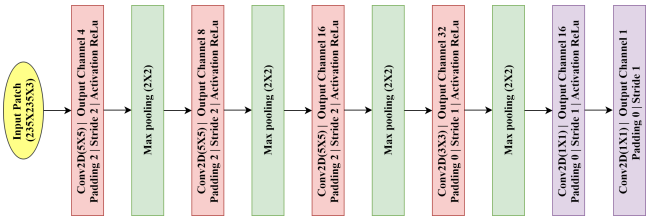

Fig. 3: CNN architecture for patch level label prediction

The diagram of the proposed TSARA algorithm for real vs fake 4K classification is illustrated in Fig. 2. In the first stage, local patches are extracted from the 4K/UHD image or video frame. A CNN model is then used to predict the class labels of the local patches. The use of local patches has several advantages. First, local patches limit content variation and subsequently improve model prediction accuracy. Second, local patches lead to smaller network size and lower computational complexity in both the training and testing phases. Third, the use of local patches makes it easier to increase the amount of data for training. Fourth, local patch assessment offers spatial variations across the image or video frame being tested. In the second stage, patch level label pre-

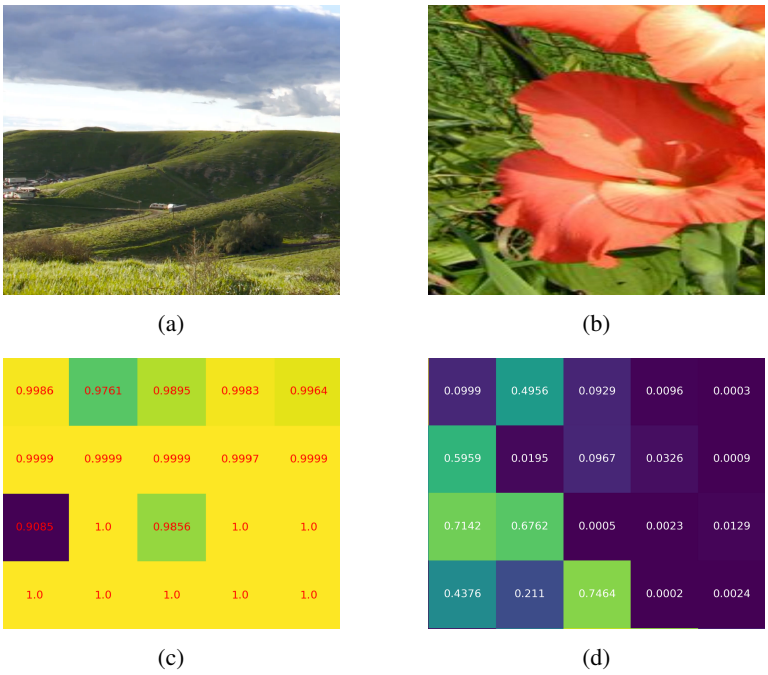

Fig. 4: (a) Cropped and enlarged "True4K" ground truth label image (b) Cropped and enlarged "Fake4K" ground truth label image (c) Confidence map generated using CNN model for (a) (d) Confidence map generated using CNN model for (b)

dictions are aggregated and a logistic regression on detection frequency is used to make an overall assessment of the whole image or video frame.

The CNN architecture of the patch level classifier is shown in Fig. 3. The network takes a color image patch of size 235×235×3 as input. The depth of the feature map gradually increases from 3 to 32 and the spatial size of the feature map is reduced with a total of 4 CNN and pooling layer pairs. ReLu is the activation function used in all the convolutional layers and the first dense layer. The last two layers are dense layers for the classification of the features extracted by the CNN layers. The sigmoid activation function is utilized at the end to predict the probability. A probabilistic threshold is used to determine the label of the patch. The CNN is trained with patches extracted at a stride size of 256 using 60% and 20% of the images in the built database for training and validation, respectively.

The Adam [21] is used as the optimizer keeping learning rate 0.0005. The weight initialization technique for convolutional layer used in this algorithm is adopted from [22]. After training the model on 0.779 million patches of 6494 training images, only non-flat patches are used for prediction in the CNN model. This step largely improves the model accuracy as flat regions do not have sufficient information that allows for accurate classification. The constraint on the statistical variance is applied to detect the flat patch is 5e-5.

When the CNN model is applied to local patches across a test image, a spatial map of label predictions is created. To aggregate local patch labels into a global prediction, we apply a logistic regression model on the label frequency distribution. The logistic regression model is implemented with Ridge regression (L2 regularization) [23] and Limited-memory BFGS optimization algorithm [24]. It is fitted on the label frequency distribution generated from the non-flat patches of the training dataset. The removal of flat patches in the statistics largely reduces the uncertainty in model prediction and appears to be highly effective at improving the global prediction accuracy.

Table 1: Proposed model accuracy

Patch Level Accuracy		
Train	Validation	Test
99.14%	99.20%	98.96%
Image Level Accuracy		
Train	Validation	Test
99.85%	99.88%	99.91%

3. RESULT AND ANALYSIS

As the built dataset is balanced, classification accuracy in terms of the percentage of correct classifications, Area Under the Curve of Receiver Operating Characteristic curve (AUC-ROC), and Area Under the Curve of Precision-Recall curve (AUC-PR), are used to evaluate and compare the performance of the proposed method with NR-IQA and NR-ISA algorithms. Table 1 shows the training, validation and test

Table 2: Accuracy, ROC-AUC, PR- AUC performance of NR-IQA, NR-ISA and TSARA methods

Type	Model	AUC - ROC	AUC - PR	Maximum Testing Accuracy	Execution Time (Seconds/Image)
NR-ISA	SPARISH [18]	0.967	0.968	90.66%	281.430
	HVS-MaxPol-1 [13]	0.934	0.941	85.34%	**0.520**
	MLV [17]	0.900	0.798	81.38%	2.267
	GPC [15]	0.882	0.889	80.83%	77.989
	HVS-MaxPol-2 [13]	0.833	0.824	75.46%	1.018
	LPC [16]	0.829	0.777	74.60%	29.244
	FQPath [14]	0.190	0.346	50.58%	0.897
NR-IQA	dipIQ [11]	0.954	0.943	90.71%	245.261
	CORNIA [5]	0.914	0.919	83.22%	122.905
	EONSS [12]	0.907	0.919	82.55%	2.552
	NIQE [8]	0.866	0.890	77.50%	4.431
	BRISQUE [6]	0.796	0.829	74.76%	2.275
	LPSI [10]	0.789	0.765	74.05%	**0.700**
	MEON [7]	0.788	0.767	72.05%	2.601
	SISBLIM [9]	0.756	0.747	68.78%	39.655
Proposed	TSARA	**0.999**	**0.999**	**99.91%**	0.754

accuracy of the proposed algorithm at both patch and image levels. The results suggest that the proposed method is able to make highly accurate predictions at both patch and image levels. Closer observations on the images that the model makes incorrect predictions suggest that the model is mostly struggling when a majority of the image regions are flat, which significantly limits the information for effective predictions.

A special feature of the proposed TSARA approach is that a spatial probability map may be generated for each test image, where each point in the map represents the raw output (before being thresholded to a binary classification result) of the CNN model applied to the surrounding image patch. Examples of the probability maps of "True4K" and "Fake4K" images are given in Fig. 4, where local probability numbers are given. These probability maps provide useful insights regarding region-specific behaviours of the algorithm and how such local information may be used to infer global decisions.

To the best of our knowledge, there is no other method for the classification of real vs. fake resolutions available in the public domain. Therefore, we compare the proposed TSARA algorithm with existing NR-ISA and NR-IQA techniques on the AUC-ROC, AUC-PR and accuracy evaluation metrics. Since the NR-IQA and NR-ISA methods report scalar values on an image, following which thresholds need to be applied to obtain binary classifications, we perform a line search for each method for the threshold value that results in the best classification accuracy, and these best results are reported in Table 2. Moreover, the computational complexity in terms of execution speed of all the method is also compared. The execution time is reported as the average time of ten times running loop of the two-test input RGB images of resolution 3840×2160 measured on a HP laptop computer with 1.8GHz Intel Core i5-8265U processor, 8 GB of RAM and Windows 10 Home operating system. Table 2 shows that the proposed two-stage algorithm outperforms all the existing NR-ISA and NR-IQA methods by a clear margin. The dipIQ [11] and SPARISH [18] are the best performing algorithms in the NR-IQA and NR-ISA groups, respectively. However, the average time taken by both methods to predict the native resolution of the image is around 245 seconds and 281 seconds respectively, making them difficult to use in practice. By contrast, the proposed method evaluates an image within 1 second, making it an excellent choice in time-critical applications.

4. CONCLUSION

This work targets at assessing the authentic resolution of images, with the current focus on developing automated algorithms that can accurately classify real versus fake 4K images. We construct a database of over 10,000 "True4K" and "Fake4K" images, and develop a two-stage TSARA algorithm for real vs fake 4K classification, where the first stage classifies at local patch level using a CNN, and the second stage applies logistic regression at the full image level. Our experimental results show that TSARA achieves high accuracy at the low computational cost, and significantly outperforms state-of-the-art NR-IQA and NR-ISA methods. Both the database and the proposed algorithm will be made publicly available. Future effort includes improving the generalizibility of the proposed method to handle arbitrary upscaling methods, and automatically identifying the authentic resolution at which the video content is captured or created.

5. REFERENCES

[1] 4K Media, "Real or Fake 4K," https://4kmedia.org/real-or-fake-4k/, Jan 2020.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[3] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[4] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.

[5] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1098–1105.

[6] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[7] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2018.

[8] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[9] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Transactions on Broadcasting*, vol. 60, no. 3, pp. 555–567, 2014.

[10] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 339–343.

[11] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[12] Z. Wang, S. Athar, and Z. Wang, "Blind quality assessment of multiply distorted images using deep neural networks," in *Image Analysis and Recognition*, Fakhri Karray, Aurélio Campilho, and Alfred Yu, Eds., Cham, 2019, pp. 89–101, Springer International Publishing.

[13] M. S. Hosseini, Y. Zhang, and K. N. Plataniotis, "Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4510–4525, 2019.

[14] M. S. Hosseini, J. A. Z. Brawley-Hayes, Y. Zhang, L. Chan, K. N. Plataniotis, and S. Damaskinos, "Focus quality assessment of high-throughput whole slide imaging in digital pathology," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 62–74, 2020.

[15] A. Leclaire and L. Moisan, "No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 145–172, May 2015.

[16] R. Hassen, Z. Wang, and M. M. A. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2798–2810, 2013.

[17] K. Bahrami and A. C. Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 751–755, 2014.

[18] L. Li, D. Wu, J. Wu, H. Li, W. Lin, and A. C. Kot, "Image sharpness assessment by sparse representation," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1085–1097, 2016.

[19] M. E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2008.

[20] V. A. Akundy and Z. Wang, "4k or not? - automatic image resolution assessment," in *Image Analysis and Recognition*, Aurélio Campilho, Fakhri Karray, and Zhou Wang, Eds., Cham, 2020, pp. 61–65, Springer International Publishing.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.

[23] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 regularization for learning kernels," *CoRR*, vol. abs/1205.2653, 2012.

[24] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.