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

CAPTURING BANDING IN IMAGES: DATABASE CONSTRUCTION AND OBJECTIVE ASSESSMENT

Akshay Kapoor, Jatin Sapra and Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Canada
Emails: {a43kapoor, jsapra, zhou.wang}@uwaterloo.ca

ABSTRACT

With the fast technology advancement and the accelerated growth of high-quality image and video production and services, banding or false contour has become a frequently observed artifact in images, creating annoying negative impact on the visual quality-of-experience (QoE) of end users. Nevertheless, thorough investigations on the causes of banding, and effective and efficient methods to detect and reduce banding are largely lacking. This work targets at capturing and quantifying banding artifacts in images. We construct the first of its kind large-scale public database, consisting of 1,250 images with segmented banding regions and 169,501 image patches with class labels. We also develop a deep neural network based no-reference deep banding index (DBI), which not only produces an overall banding assessment of a given image, but also creates a banding map that indicates the variation of banding across the image space. Our experiments show that the proposed DBI method achieves accurate banding prediction with low computational cost. The database and the proposed algorithm are made publicly available[1].

Index Terms— image banding, false contour, no-reference image quality assessment, deep neural network

1. INTRODUCTION

Banding, colour banding, or false contours is a common visual artifact appearing in images and videos, often in large regions of low textures and slow gradients such as sky. When the granularity of bit-depth or display intensity levels mismatches with the visual system's perception of the smooth transition of color and luminance presented in the image content, the discontinuity positions in smooth image gradients are transformed into perceivable, wide, discrete bands. Banding significantly deteriorates the perceptual quality-of-experience (QoE) of end users. An example is shown in Fig. 1 where banding artifacts are clearly visible in the sky. With the recent growing popularity of high dynamic range (HDR), wide color gamut (WCG) production and services, banding has become an even bigger issue because HDR/WCG attempts to cover a wider range of luminance/color levels than that of standard dynamic range (SDR) content, making it more difficult to create smooth visual transitions of slow gradients.

Fig. 1. An example of visual banding artifact.

Commonly used image quality assessment (IQA) methods such as PSNR cannot reliably measure the visibility of banding. The structural similarity index (SSIM) [1] often captures banding in its quality maps but does not give sufficient penalty that accurately reflects the perceptual annoyance of banding. Only a few methods were proposed specifically for detecting and quantifying banding. In [2], directional contrast features are used to differentiate between false contours and object edges in a two-stage banding detection algorithm. In [3], the banding artifacts are quantified based on the size and number of uniform segments. The idea is extended by first identifying uniform segments and then using edge length and contrast to differentiate between banding and other artifacts such as blockiness [4]. While these initial attempts are valuable, there are significant gaps in achieving reliable image banding assessment. First, sufficient and convincing validation has been missing, largely due to the lack of high quality databases that provide wide coverage of the variations in both image content and banding types. Second, existing methods are often of high computational complexity, impeding the wide usage of such methods in practical systems that often require real-time scanning of large volumes of videos.

In this work, we aim for developing automated algo-

[1]Dataset access: https://zenodo.org/record/4512571.YCNhKGhKhPZ, https://zenodo.org/record/4513740.YCNhRmhKhPY; Code access: https://github.com/akshay-kap/Meng-699-Image-Banding-detection

rithms for banding assessment without accessing the original banding-free reference image. To tackle the problem, we first create the first of its kind large-scale public database consisting of images and image patches with and without banding artifacts. As explained in Section 2, we develop a semi-automatic annotation procedure, which segments 1,250 images into banded and non-banded regions, and assign 169,501 image patches with banded/non-banded class labels. The large database allows us to train a deep neural network (DNN) for no-reference (NR) banding classification at patch level. By aggregating DNN predictions of local patches in an image, we create a deep banding index (DBI) for image level banding assessment. Meanwhile, the proposed approach also produces a banding map for each image, which indicates the variations of visual banding over the image space.

2. BANDING DATABASE CONSTRUCTION

We construct one of the first of its kind image database for image banding assessment, which is be made publicly available. To the best of our knowledge, the only previous banding database was briefly mentioned in [4]. It is not publicly available and is limited in size and banding distortion type (VP9 video transcoding of Youtube content only).

We create a database by extracting frames from more than 600 pristine high-definition (HD) videos, resulting in approximately 1,250 images of 1920×1080 resolution. Banding distortion can be introduced by bit-depth reduction (or dynamic range tone mapping) in luma and chroma channels of image/video (where bit-depth reduction can also leads to other perceptually visible artifacts). Therefore, we apply bit-depth reduction followed by bit-depth promotion, where the level of banding varies with the levels of quantization. Six levels of quantization are used to enhance the diversity of the dataset. Close observations suggest that our simulated banding artifacts are perceptually very similar to the strong banding artifacts in images and videos frequently seen on popular video and social media service platforms such as Netflix, Amazon Prime and LinkedIN.

Table 1. Composition of labelled image patch dataset

	Banded	Non-Banded
Training	58917	76507
Validation	7474	9563
Testing	7476	9563
Total	73867(44%)	95634(56%)

We use a semi-automatic approach to create class labels. Firstly, the 1,250 HD images are manually segmented and labelled into banded and non-banded regions. A large number of image patches of size 235×235 are then extracted from these segmented and labelled images to form a large dataset

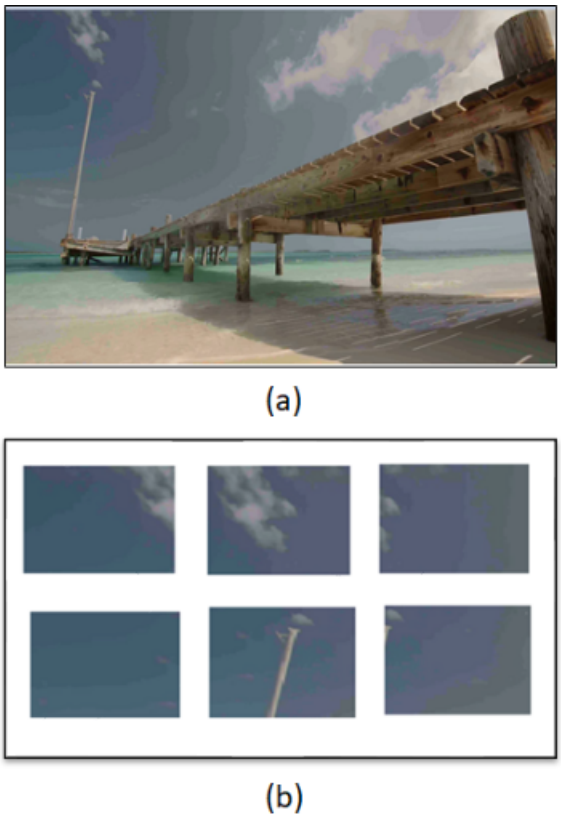

Fig. 2. Sample banded image (a) and patches (b) extracted from its upper-left region.

of image patches where each patch is labelled as banded or non-banded in an automatic manner. Specifically, a patch is labelled as banded if it has more than 30% overlap with banded regions in the image, and as non-banded otherwise. A sample banded image together with extracted patches are shown in Fig. 2.

Eventually, a dataset of 1,250 segmented images of 1920×1080 resolution and 169,501 labeled image patches of size 235×235 are generated. The scale of the data allows for training machine learning based classification methods including DNN based approaches. Table 1 provides more details of the database, together with suggested divisions to the training (∼80%), validation (∼10%), and testing (∼10%) sets. The non-banded dataset is slightly larger due to the inclusion of patches extracted from the original pristine images.

3. OBJECTIVE BANDING ASSESSMENT

The diagram of the proposed banding assessment algorithm is shown in Fig. 3. A DNN model is first applied to classify individual 235×235 image patches into 'banded' and 'non-banded' categories. The local patch classification labels are

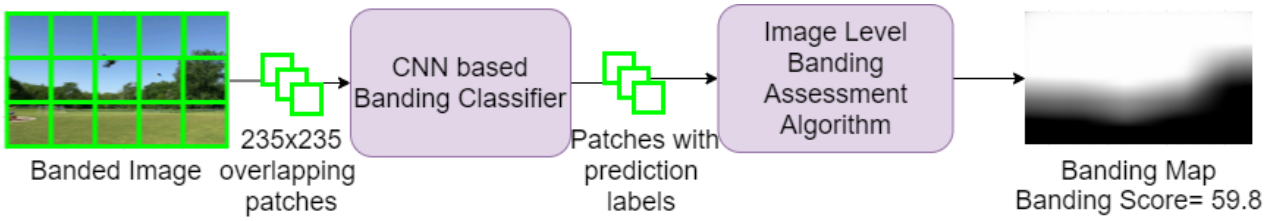

Fig. 3. Diagram of the proposed banding assessment model.

then aggregated aross the image space to yield an overall banding score of the test image. Meanwhile, the local classification labels are smoothed spatially to create a banding map that reflects the variation of perceptual banding across space.

3.1. Patch Level Banding Classification

We train a seven-layer convolutional neural network (CNN) of 6.7K parameters to classify an image patch of size 235×235 to be "banded" or "non-banded". The CNN architecture is shown in Fig. 4. The network uses a sequence of convolutional layers, pooling layers, followed by batch normalization layers and global average pooling, before the final dense layers that produce the final classification result. The use of Batch normalization [7] allows the model to converge faster and produces some regularizing effect. The use of a global average pooling [8] helps flattening the receptive fields into 1D array while utilizing the average spatial information. The best fit for the model is obtained using Adam optimizer with a learning rate of 0.001 and an exponential decay value of 0.9. Rectified linear units are used as activation functions for all the hidden layers in the network architecture and sigmoid activation function is used for the output dense layer. The loss function used is binary cross entropy, and output score for each class is optimized using cross entropy loss.

Since the patch dataset features imbalance of classes (Table 1), precision-recall analysis is carried out on the validation set to select the best threshold to convert the predicted probabilities into binary classifications. The threshold is set as 0.2 which yields the highest F1-score (0.9) and accuracy (0.93) on the validation set.

3.2. Image Level Banding Assessment

To produce an image level banding score, the image is divided into non-overlapping 235×235 patches and CNN based patch level classification is applied, resulting in a raw banding classification label for each patch. The raw label S_i of Patch i is then updated depending on the labels of its neighboring patches based on

$$\tilde{S}_i = W_p \cdot S_i + W_n \cdot \frac{1}{N} \sum_{j=1}^{N} S_j \,, \tag{1}$$

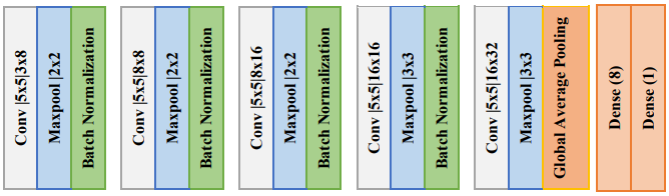

Fig. 4. CNN architecture for patch classification. 'Conv layer | a x a | c x d' denotes a convolutional layer featuring filter size a x a, with c and d filters in the previous and current layers, respectively; 'Maxpool | e x e' denotes the window size of max-pooling layer with valid padding; and Dense(f) denotes a dense layer featuring f neural units.

where N is the number of its neighbors (currently $N = 8$ as its direct neighbors), j is the neighbor index, W_p and W_n are the weights given to the current patch and its neighbors, respectively, and $\tilde{S}_i$ is the updated score. The overall deep banding index (DBI) of the whole image is then calculated as the average of all updated patch scores.

The CNN model for local patch labeling may be applied as a sliding window across the image to create a dense spatial banding map. Specifically, we let the 235×235 window slide through the image with a stride of 5. We then compute the pixel value in the banding map as the average of the classification labels of all windows containing the pixel. This can be expressed as

$$P_i = \frac{1}{K} \cdot \sum_{j=1}^{K} S_j, \tag{2}$$

where K is the total number of neighboring windows that contain the central pixel, j is the window index, and S_j is the classification label of the j-th window.

Fig. 5 shows a set of sample images with their corresponding image level banding scores (DBIs) and banding maps, where brighter pixels in the maps indicate stronger banding. It can be observed that the spatial variation of the banding effects are well predicted by the banding maps, and the overall banding scores provide a good predictions on the spread of banding. Advanced weighting methods may be developed in the future to further improve the overall banding prediction.

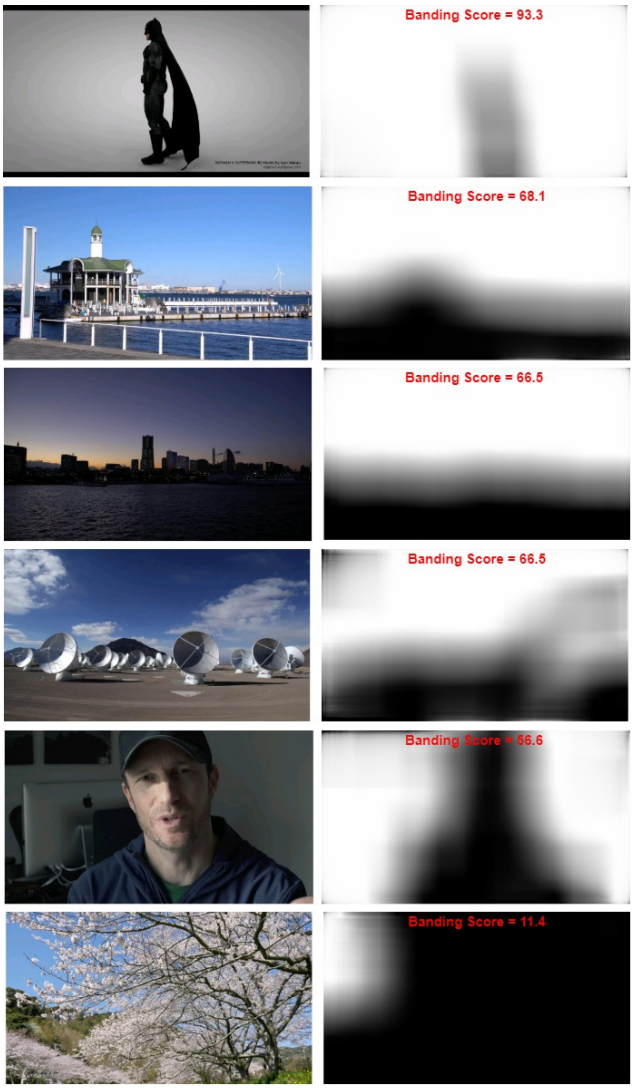

Fig. 5. Sample images with corresponding banding maps (brighter pixel in the map indicates stronger banding) and image level DBI scores.

3.3. Performance Comparison

To the best of our knowledge, no other banding assessment method is available in the public domain. Therefore, we compare the proposed DBI model with existing NR-IQA techniques on the task of patch banding classification. The algorithms under comparison include BRISQUE [9], LPSI [10], SSBLIM [11], MEON [12], dipIQ [13], and NIQE [14]. All these NR-IQA methods produce scalar quality values only, and thus a thresholding step is necessary to convert the scalar values into binary classification decisions. For each method, we conduct a line search to find the optimal threshold value that produces the best classification results, and the corresponding maximum testing accuracy (MTA) is reported. In addition, AUC-ROC (area under the curve - receiver operating characteristics) and AUC-PR (area under the curve - precision recall) are also included as the evaluation criteria. The performance comparison results of existing NR-IQA methods against the proposed DBI method are given in Table 2, where DBI clearly outperforms all NR-IQA methods.

Moreover, the computational complexity in terms of execution time per image patch is also compared, as shown in Table 2. The execution time is reported as the average time for processing an image patch on an Intel(R) Core (TM) machine with i7-8750H CPU at 2.21 GHz. It can be seen that the proposed DBI method is among the fastest, making it a favorable choice in time-critical applications.

Table 2. Performance comparison based on AUC-ROC (area under the curve - receiver operating characteristics), AUC-PR (area under the curve - precision recall), MTA (maximum testing accuracy), and computation speed (execution time in second per image patch).

Model	AUC-ROC	AUC-PR	MTA	Speed
BRISQUE[8]	0.26	0.31	56%	0.6438
LPSI[9]	0.74	0.71	73%	**0.0148**
SSBLIM[10]	0.46	0.46	57%	0.2632
MEON[11]	0.35	0.35	56%	0.6473
dipIQ[12]	0.62	0.58	66%	3.4736
NIQE[13]	0.15	0.20	46%	1.26
DBI	**0.95**	**0.96**	**91%**	0.0487

Despite the superior performance of the proposed DBI method against existing NR-IQA approaches, there are images and patches that DBI produces incorrect predictions. Closer observation on these cases reveal that DBI tends to perform well on large wave like banded patterns, but struggles with image content with macroblocking artifacts or wave-like patterns in the original content or created by artificial design.

4. CONCLUSION

In this work, we target at automatic NR banding assessment of images. We construct so far the largest database dedicated to banding artifacts, containing images with banding region segmented and patches with classification labels. We develop a CNN-based DBI banding classification model for image patch labeling, upon which we compute an overall DBI score for a given test image and create a banding map to predict the spatial variations of visual banding effect. Image samples and performance comparisons show that the proposed method achieves accurate banding prediction with low computational cost. The database and the proposed DBI algorithm build the groundwork to facilitate future reproducible research on banding characterization, detection and reduction methodologies.

5. REFERENCES

[1] 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, 13(4), 600612, (2004).

[2] J. W. Lee, B. R. Lim, R. H. Park, J. S. Kim, and W. Ahn, "Two-stage false contour detection using directional contrast features and its application to adaptive false contour reduction," IEEE Trans. Consum. Electron., 2006, doi: 10.1307/mmj/1144437443.

[3] G. Baugh, A. Kokaram, and F. Pitie, "Advanced video debanding," In European Conference on Visual Media Production (CVMP), 2014.

[4] Y. Wang, S. Kum, C. Chen and A. Kokaram "A perceptual visibility metric for banding artifacts." In: IEEE Int. Conf. on Image Processing, pp. 2067-2071, 2016.I.

[5] J. Kim and S. Lee, "Fully Deep Blind Image Quality Predictor," IEEE J. Sel. Top. Signal Process., 2017, doi: 10.1109/JSTSP.2016.2639328.

[6] J. Kim and S. Lee, "Fully Deep Blind Image Quality Predictor," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 1, pp. 206-220, Feb. 2017, doi: 10.1109/JSTSP.2016.2639328.

[7] Ioffe, Sergey, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In International conference on machine learning, pp. 448-456. PMLR, 2015.

[8] Szegedy et al., "GoogLeNet Going Deeper with Convolutions," arXiv Prepr. arXiv1409.4842, 2014, doi: 10.1109/ICCV.2011.6126456.g. 2015

[9] 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

[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. 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.

[12] 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.

[13] 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.

[14] 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.