# DEEP IMAGE DEBANDING

*Raymond Zhou, Shahrukh Athar, Zhongling Wang, and Zhou Wang*

Department of Electrical & Computer Engineering, University of Waterloo, Canada
Email: {raymond.zhou, shahrukh.athar, zhongling.wang, zhou.wang}@uwaterloo.ca

## ABSTRACT

Banding or false contour is an annoying visual artifact whose impact negatively degrades the perceptual quality of visual content. Since users are increasingly expecting better visual quality from such content and banding leads to deteriorated quality-of-experience, the area of banding removal or debanding has taken paramount importance. Existing debanding approaches are mostly knowledge-driven, while data-driven debanding approaches remain surprisingly missing. In this work, we construct a large-scale dataset of 51,490 pairs of corresponding pristine and banded image patches, which enables us to make one of the first attempts at developing a deep learning based banding artifact removal method for images that we name *deep debanding network* (deepDeband). We also develop a bilateral weighting scheme that fuses patch-level debanding results to full-size images. Extensive performance evaluation shows that deepDeband is successful at greatly reducing banding artifacts in images, outperforming existing methods both quantitatively and visually. The proposed algorithm and dataset are made publicly available. [1]

***Index Terms*—** image banding, false contour, debanding, deep learning, deep convolutional neural network

## 1. INTRODUCTION

Banding artifacts are common annoyances found in visual content and caused by quantization, due to issues like compression, tone mapping, and poor display. They often appear in large regions of smooth content with low textures and slow gradients, such as sky or water. Banding manifests as sharp, discrete colour discontinuities where there otherwise should be smooth transitions, causing notable degradation in quality. Fig. 1 shows an image with severe banding in the sky. Recent technological advances, like increases in display resolution, have led users to expect better visual quality-of-experience, where banding artifacts are particularly annoying. Thus, there is an urgent need to develop accurate banding detection and banding removal (or debanding) methods that are practically applicable.

While some work has been done in banding detection [1, 2, 3, 4] and removal [5, 6, 7], no banding removal efforts have been made with deep learning. Contemporary debanding methods [5, 6, 7] are knowledge-driven, relying on domain knowledge and understanding of the human visual system [8]. A major disadvantage of such methods is that they have multiple parameters that must be carefully calibrated properly for optimal results, which can be a lengthy process. Additionally, many knowledge-driven methods like [7] use dithering, introducing noise, as a means to lessen the visibility of banding. However, this process often reduces the visibility of fine texture details, worsening the user's quality-of-experience [8]. An alternative is to take a data-driven approach where machine learning, especially deep learning, is employed. Indeed, deep learning techniques have been widely used in the image restoration context, such as in the removal of noise [9], blur [10], and blocking artifacts [11]. However, to the best of our knowledge, thus far there have been no efforts to use deep learning targeted specifically at banding removal.

**Fig. 1**: Example of banding artifacts present in the sky region.

Our major contributions are as follows. 1) We make one of the first attempts to develop a deep learning model for removing banding artifacts caused by quantization from images, taking a banded image and returning its debanded version, and call it *deepDeband*. 2) As a major bottleneck in developing robust deep learning models is a lack of annotated training data, we create a large dataset of 51,490 corresponding pairs of image patches with and without banding artifacts, supporting future deep learning debanding work. 3) We present two techniques of applying deepDeband: a direct global method and a patch-level method followed by a bilateral weighting scheme.

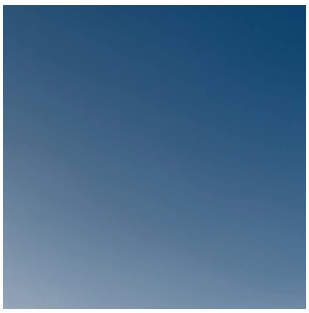
## 2. DATASET CONSTRUCTION

To construct a new dataset that enables training of deep learning based debanding models, we build upon an existing work [2], which contains 1,439 pairs of pristine and their corresponding quantized images of 1920×1080 resolution (FHD), where the quantized images have been segmented and labelled into banded and non-banded regions. To the best of our knowledge, this is the only publicly available dataset containing labelled banded images. From each quantized FHD image, we extract overlapping image patches of size 256×256 with a sliding window of stride 75. Using the provided image labels, we select only those patches that contain banding and also extract their corresponding patches from the pristine FHD images. This process results in 51,490 pairs of image patches, which we partition into training (∼60%), validation (∼20%), and test (∼20%) sets without content-overlapping, such that for any FHD image, all

[1] Access: https://github.com/RaymondLZhou/deepDeband

(a) Banded FHD image

(b) Banded patch



(c) Pristine patch

**Fig. 2**: Banded FHD image (a), extracted banded patch (b), and corresponding pristine patch extracted from the pristine FHD image (c).

**Table 1**: Dataset composition in terms of patches and FHD images.

| Dataset | Patches (256×256) | FHD (1920×1080) |
|---|---|---|
| Training | 30,988 | 872 |
| Validation | 10,203 | 257 |
| Testing | 10,299 | 310 |
| Total | 51,490 | 1439 |

patches extracted from it belong to the same set. Each set contains images of diverse visual content, the scale of which allows for developing deep learning models. This procedure of creating a new patch dataset was necessary as the existing patches from [2] did not have matching pristine and banded pairs needed for deep learning tasks.

Fig. 2 shows an example of a banded FHD image with a zoomed-in banded patch and its corresponding pristine patch that has been extracted from the respective pristine FHD image. Table 1 provides a detailed overview of the dataset. Since each FHD image can result in fairly different numbers of banded patches, the percentage of FHD images corresponding to each of the training, validation, and test sets is not in the same proportion as that of the image patches.

## 3. DEBANDING MODEL DEVELOPMENT

### 3.1. Deep Debanding Network

We opt to adopt the conditional Generative Adversarial Network (cGAN) Pix2Pix [12] as the basis of our deep learning model. Pix2Pix has been successfully used in a wide range of other image-to-image translation and image restoration tasks, such as denoising [13], deblurring [14], and dehazing [15], making it promising for debanding. Compared to other networks with similar architectures, Pix2Pix has lower complexity and its adversarial objective is ideal for capturing the nature of banding artifacts and successfully removing them. Pix2Pix contains a generator and discriminator [12]. During training, the model is given corresponding pairs of banded and pristine patches. The generator takes banded patches and gives debanded patches as output. The discriminator distinguishes pristine patches from generated ones. Both parts are trained together, with the generator and discriminator trying to minimize and maximize the loss function, respectively [12]. During evaluation, the generator is given a banded image and produces its debanded version.

We train our debanding model, called *deep debanding network* (deepDeband), by using the 30,988 pairs of image patches contained in the training set of the dataset constructed in Section 2 (Table 1). Since initial experiments using the Pix2Pix architecture showed promising results, we focus our attention towards different application methods as described in Section 3.2. We also explore different batch sizes, image augmentation, and dataset compositions, covered in Section 3.3. All hyperparameters other than batch size are the same as the default implementation of Pix2Pix [12].

### 3.2. Application to Images

Although deepDeband is trained on 256×256 patches, we need to apply it on images of larger sizes (e.g., 1920×1080 FHD images). We use two techniques to do so. The first method, named deepDeband Full image (deepDeband-F), directly applies the network to the full image. Since the Pix2Pix generator is fully convolutional [12], deepDeband can be applied to images of any size, though here we focus on FHD images. The Pix2Pix generator expects an input whose width and height dimensions are divisible by 256, so we first pad the input FHD image to 2048×1280 through mirroring. We then crop the debanded image returned by the model back to 1920×1080. Padding the input with black or white results in unwanted textures, likely as the solid colour is unnatural visual content.

Our second method operates at the patch-level. First, we pad the input FHD image to 2048×1280 and extract overlapping 256×256 patches with a stride of 128, and deepDeband is applied to all patches. A pixel in the image may be covered by multiple patches, where the $i$-th patch produces a prediction $p_i$ of the subject pixel. These predictions may be different, and merging non-overlapping patches or using a simple average may result in undesirable visible patch boundaries. Presumably the image patches that are closer to the subject pixel should carry more weights, where the closeness may be in terms of the geometric distances or content intensity values. Therefore, we compute both the geometric distance, $d_g$, between the patch center and the subject pixel, and the root mean squared (RMS) difference, $d_c$, between the intensities of the subject pixel and all pixels in the patch. We then define a bilateral weighting function with a two-dimensional Gaussian profile:

$$w = \exp\left(-\frac{d_g^2}{2\sigma_g^2} - \frac{d_c^2}{2\sigma_c^2}\right), \tag{1}$$

where $\sigma_g$ and $\sigma_c$ control the decaying speeds of the Gaussian profile. Finally, we compute the reconstructed subject pixel value $p_r$ using a bilateral weighting scheme given by

$$p_r = \frac{\sum_{i=1}^{N} w_i p_i}{\sum_{i=1}^{N} w_i}, \tag{2}$$

where $N$ is the number of patches that cover the subject pixel. This bilateral weighting approach allows us to make the best use of overlapping patches to produce a smooth spatial transition across the im-

(a) Banded FHD image (b) deepDeband-F FHD image (c) deepDeband-W FHD image

**Fig. 3**: Banding in banded FHD image (a) is significantly reduced in deepDeband-F (b) and deepDeband-W (c) FHD images.

**Table 2**: Mean validation scores of deepDeband-F for different patch sizes. Optimal values are in bold.

| Dataset | DBI ↓ [2] | BBAND ↓ [1] |
|---|---|---|
| 256 × 256 patches | **0.2028** | **0.1803** |
| 572 × 572 patches | 0.3495 | 0.1896 |
| 1920 × 1080 FHD images | 0.2763 | 0.1870 |

**Table 3**: Mean validation scores of deepDeband-F for different dataset compositions. Optimal values are in bold.

| Dataset | DBI ↓ [2] | BBAND ↓ [1] |
|---|---|---|
| Banded patches, no flipping | **0.2028** | **0.1803** |
| Banded patches, with flipping | 0.2037 | 0.1848 |
| All patches, no flipping | 0.3114 | 0.1926 |
| All patches, with flipping | 0.3240 | 0.1924 |

age. Finally, we crop the output to 1920×1080. We name this version of the model deepDeband Weighted merge (deepDeband-W). Fig. 3 shows an image example, where banding in the smooth regions is greatly removed while sharp textures remain well-preserved.

### 3.3. Validation

Since losses reported by GANs are not meaningful due to their adversarial nature [16], we apply DBI [2] and BBAND [1] as evaluation metrics on the validation set of our dataset. Both DBI and BBAND are no-reference banding assessment indices, making them an ideal choice for this task. For both DBI and BBAND, a smaller value indicates less banding. To explore the optimal training conditions for our deep learning models, we create additional datasets and use the deepDeband-F model in the subsequent analysis.

First, to ascertain the optimal patch size for training, we train the deepDeband model on banded (and their corresponding pristine) patches of size 256×256 and 572×572, and on 1920×1080 FHD images at a fixed batch size of 8. Table 2 shows the outcome of this experiment in terms of DBI and BBAND, where it can be seen that using the smaller 256×256 patches leads to superior performance, even when applied directly at the image-level, as in deepDeband-F. This is likely because using smaller patch sizes gives more training examples, an important requirement for deep learning.

Next, we investigate different dataset compositions, using patches of size 256×256 and batch size 8. Beyond the dataset of only banded patches from Section 2, we use datasets containing patches of all types of visual content (banded and non-banded). We also try data augmentation through horizontal flipping. Table 3 summarizes the results in terms of DBI and BBAND, where we see that including all types of patches and incorporating horizontal flipping leads to poor performance, which visibly manifests as undesirable textures introduced in the output images. This finding is somewhat counterintuitive and is worth investigating further. Based on results shown in Tables 2 and 3, we finalize the dataset of only banded 256×256 patches, discussed in Section 2, for model training.

Finally, we vary batch size between 2, 4, 8, and 16, finding smaller values tend to give better results. We train the model for 200 epochs, choosing the epoch with best performance on the validation set, for DBI and BBAND, as our final deepDeband models.

## 4. PERFORMANCE COMPARISON

To the best of our knowledge, deepDeband is so far the only deep learning based debanding method. Thus, we compare its performance against three most recent state-of-the-art knowledge-driven debanding methods, FFmpeg's deband filter [5], AdaDeband [6], and FCDR [7], on the 310 FHD PNG images in the test set of our dataset (Table 1). We use default parameters for FFmpeg. For AdaDeband, originally designed for YUV420p videos, we test the following four versions. 1) AdaDeband 1: The original YUV420p version with default parameter settings; 2) AdaDeband 2: The YUV420p version with parameter settings recommended by the authors for H.264 encoded frames (since images in [2], on which our dataset is built, were extracted from H.264 encoded videos); 3) AdaDeband 3: To prevent any loss in colour due to chroma subsampling to YUV420p, we implement a YUV444p version with default parameter settings; 4) AdaDeband 4: The YUV444p version with parameter settings as in AdaDeband 2. Since the full implementation of FCDR is not publicly available, we implement it ourselves by building upon code provided in [6]. We try nine different parameter settings and choose the top two performers, which we call FCDR 1 and FCDR 2.

For evaluation, we use both DBI [2] (data-driven) and BBAND [1] (knowledge-driven), which to the best of our knowledge are the only publicly available banding assessment indices, making them most appropriate for our analysis. For BBAND, we choose parameter settings based on the authors' recommendations for content in our test set. As additional evaluation metrics, we use three blind image quality assessment (BIQA) methods of varying design philosophies, which are shown to be top performers in the BIQA area in [17]. These include HOSA (opinion-aware) [18], dipIQ (opinion-unaware) [19], and ILNIQE (opinion-unaware) [20].

Table 4 shows the performance of debanding methods in terms of mean evaluation metric scores, including the original banded images as a baseline. Notably, deepDeband-F outperforms all existing methods at all metrics except FCDR 1 at dipIQ, and its DBI score is second only to that of deepDeband-W. Similarly, deepDeband-W surpasses all existing methods except FFmpeg at BBAND, and FCDR 1 and FCDR 2 at dipIQ, making it as competitive as deepDeband-F. Furthermore, deepDeband-F and deepDeband-W perform significantly better than FCDR 1 and FCDR 2 for all met-

**Table 4**: Quantitative performance comparison of debanding methods in terms of mean evaluation metric scores on the entire test set. Optimal values are in bold. The arrow besides each metric name shows whether higher or lower mean values are better.

| Model | DBI ↓ [2] | BBAND ↓ [1] | dipIQ ↑ [19] | HOSA ↓ [18] | ILNIQE ↓ [20] |
|---|---|---|---|---|---|
| Banded Images | 0.4059 | 0.3830 | -6.5261 | 34.8581 | 29.7630 |
| FFmpeg [5] | 0.2240 | 0.1523 | -5.8885 | 34.6408 | 30.0330 |
| AdaDeband 1 [6] | 0.3414 | 0.2085 | -6.3723 | 34.8560 | 28.3333 |
| AdaDeband 2 [6] | 0.3358 | 0.2060 | -6.3735 | 34.9581 | 28.4328 |
| AdaDeband 3 [6] | 0.3374 | 0.2163 | -6.4375 | 35.1767 | 28.1319 |
| AdaDeband 4 [6] | 0.3328 | 0.2135 | -6.4370 | 35.1811 | 28.2034 |
| FCDR 1 [7] | 0.3980 | 0.3468 | **-4.9563** | 35.4358 | 29.0935 |
| FCDR 2 [7] | 0.3813 | 0.3538 | -5.3961 | 35.7727 | 28.8689 |
| deepDeband-F | 0.2026 | **0.1518** | -5.3110 | **32.8358** | **25.4175** |
| deepDeband-W | **0.1774** | 0.1629 | -5.7636 | 34.4330 | 25.8048 |

Banded FHD Image (a) (b) (c) (d) (e) (f)

**Fig. 4**: Visual performance comparison of banded (a) and debanded content using FFmpeg (b), AdaDeband 4 (c), FCDR 2 (d), deepDeband-F (e), and deepDeband-W (f).

**Table 5**: Execution time comparison. Optimal values are in bold.

| Model | Time (seconds) ↓ |
|---|---|
| FFmpeg [5] | **1.2907** |
| AdaDeband 1 [6] | 10.6903 |
| AdaDeband 2 [6] | 12.7411 |
| AdaDeband 3 [6] | 10.9448 |
| AdaDeband 4 [6] | 12.8803 |
| FCDR 1 [7] | 23.4890 |
| FCDR 2 [7] | 38.7096 |
| deepDeband-F | 10.0121 |
| deepDeband-W | 283.0539 |

rics except dipIQ, especially in terms of DBI and BBAND. We also present a visual comparison of the results in Fig. 4, as the ultimate goal is to reduce human perception of banding. It can be seen that although FFmpeg, AdaDeband 4, and FCDR 2 reduce the visibility of banding, the banding contours are still visible, which are hardly discerned in images produced by deepDeband methods.

Finally, we evaluate the execution time of different methods to deband one FHD image on a machine with a 2.70GHz Intel Core i7-7500U processor and 8GB of RAM. Table 5 shows the results for this experiment, where it can be seen that deepDeband-F is faster than all other methods except FFmpeg, while deepDeband-W is the slowest, which can be attributed to the weighted merge approach that it takes. Both of our models can be accelerated by GPUs during evaluation.

## 5. CONCLUSION

We propose the first deep learning based model of its kind for removing banding artifacts from images, deepDeband. It possesses none of the disadvantages of existing knowledge-driven methods, such as the need to carefully fine-tune parameters, and can be applied to images of any size. Extensive performance evaluation shows that deepDeband outperforms all existing debanding methods, both quantitatively and visually. We create a dataset of 51,490 pairs of image patches, comprised of corresponding banded and pristine patches. We also present a novel bilateral weighting method of application at the patch-level. The model builds the foundation to support future work in deep learning based banding removal. While our current approach focuses on FHD images with standard dynamic range, it can be extended to content in ultra high definition, high dynamic range, and wide colour gamut, where banding is even more pronounced.

## 6. REFERENCES


[1] Z. Tu, J. Lin, Y. Wang, B. Adsumilli, and A. C. Bovik, "BBAND Index: A No-Reference Banding Artifact Predictor," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 2712–2716.

[2] A. Kapoor, J. Sapra, and Z. Wang, "Capturing Banding in Images: Database Construction and Objective Assessment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toronto, ON, Canada, June 2021, pp. 2425–2429.

[3] P. Tandon, M. Afonso, J. Sole, and L. Krasula, "CAMBI: Contrast-aware Multiscale Banding Index," in *Proc. Picture Coding Symp. (PCS)*, Bristol, England, June 2021, pp. 1–5.

[4] L.-H. Chen, C. G. Bampis, Z. Li, J. Sole, and A. C. Bovik, "Perceptual video quality prediction emphasizing chroma distortions," *IEEE Trans. Image Process.*, vol. 30, pp. 1408–1422, Dec. 2020.

[5] "FFmpeg Filters - deband," 2021, Accessed: Aug. 31, 2021. [Online]. Available: https://ffmpeg.org/ffmpeg-filters.html#deband.

[6] Z. Tu, J. Lin, Y. Wang, B. Adsumilli, and A. C. Bovik, "Adaptive Debanding Filter," *IEEE Signal Process. Lett.*, vol. 27, pp. 1715–1719, Sept. 2020.

[7] Q. Huang, H. Y. Kim, W.-J. Tsai, S. Y. Jeong, J. S. Choi, and C.-C. J. Kuo, "Understanding and Removal of False Contour in HEVC Compressed Images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 2, pp. 378–391, Feb. 2018.

[8] H. Yeganeh, K. Zeng, and Z. Wang, "Understanding Banding – Perceptual Modeling and Machine Learning Approaches for Banding Detection and Removal," *SMPTE Motion Imaging Journal*, vol. 131, no. 3, pp. 35–41, 2022.

[9] Z. Yue, Q. Zhao, L. Zhang, and D. Meng, "Dual Adversarial Network: Toward Real-World Noise Removal and Noise Generation," in *Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, UK, Aug. 2020, pp. 41–58.

[10] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, June 2015, pp. 769–777.

[11] K. Li, B. Bare, and B. Yan, "An efficient deep convolutional neural networks model for compressed image deblocking," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Hong Kong, China, July 2017, pp. 1320–1325.

[12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, July 2017, pp. 5967–5976.

[13] S. Kaji and S. Kida, "Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging," *Radiol. Phys. Technol.*, vol. 12, no. 3, pp. 235—248, Sept. 2019.

[14] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 8183–8192.

[15] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2pix Dehazing Network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 8152–8160.

[16] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, Dec. 2016, p. 2234–2242.

[17] S. Athar and Z. Wang, "A Comprehensive Performance Evaluation of Image Quality Assessment Algorithms," *IEEE Access*, vol. 7, pp. 140030–140070, Sept. 2019.

[18] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind Image Quality Assessment based on High Order Statistics Aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sept. 2016.

[19] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.

[20] L. Zhang, L. Zhang, and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.