Image Debanding Using Cross-Scale Invertible Networks with Banded Deformable Convolutions

Yuhui Quan[a], Xuyi He[a], Ruotao Xu[b,c,1], Yong Xu[a,d,e], Hui Ji[f]

[a]*School of Computer Science and Engineering, South China University of Technology, Guangzhou, 510006, China*
[b]*Institute for Super Robotics, South China University of Technology, Guangzhou, China*
[c]*Key Laboratory of Large-Model Embodied-Intelligent Humanoid Robot, Guangzhou, China*
[d]*Peng Cheng Laboratory, Shenzhen, China*
[e]*Guangdong Provincial Key Laboratory of Multimodal Big Data Intelligent Analysis, Guangzhou, China*
[f]*Department of Mathematics, National University of Singapore, Singapore*

Abstract

Banding artifacts in images stem from limitations in color bit depth, image compression, or over-editing, significantly degrades image quality, especially in regions with smooth gradients. Image debanding is about eliminating these artifacts while preserving the authenticity of image details. This paper introduces a novel approach to image debanding using a cross-scale invertible neural network (INN). The proposed INN is information-lossless and enhanced by a more effective cross-scale scheme. Additionally, we present a technique called banded deformable convolution, which fully leverages the anisotropic properties of banding artifacts. This technique is more compact, efficient, and exhibits better generalization compared to existing deformable convolution methods. Our proposed INN exhibits superior performance in both quantitative metrics and visual quality, as evidenced by the results of the experiments.

Keywords: image debanding, invertible networks, banded deformable convolution

Email addresses: csyhquan@scut.edu.cn (Yuhui Quan), csxuyihe@mail.scut.edu.cn (Xuyi He), rtxu@superobots.com (Ruotao Xu), yxu@scut.edu.cn (Yong Xu), matjh@nus.edu.sg (Hui Ji)

[1]*Corresponding author

1. Introduction

Banding artifacts, visible as bands or stripes in an image, significantly diminish its visual quality. These artifacts are characterized by their tendency to emerge in areas with smooth transitions between colors or shades, where instead of a continuous gradient, abrupt breaks in smoothness lead to distinct bands of colors. The bands within these artifacts often exhibit a regular and repetitive pattern, a critical feature that sets banding apart from other forms of noise or artifacts. Refer to Figure 1 for an illustration exemplifying these properties.

Image debanding is a critical technology with significant value in engineering and various applications. In video streaming services, the quality of visual content directly impacts user experience. By removing banding artifacts from video streams, debanding technology enhances the visual appeal and overall quality of streamed content, leading to higher viewer satisfaction and retention rates. Additionally, debanding technology can be utilized to correct distortions in user-generated or stored dynamic image formats, as well as to address image distortions caused by irreversible bit depth changes. Furthermore, debanding technology can be integrated into various image and video processing software for user convenience. Through these software tools, users can easily rectify banding artifacts and other distortions in images, thereby enhancing image quality and visibility. This is particularly valuable for many users, especially those who need to process large volumes of images.

Various factors contribute to the emergence of banding artifacts in images. For instance, this issue is particularly common in images with limitations in color depth, such as those compressed for online use or displayed on devices with limited color capabilities. Additionally, image compression, especially through lossy compression, introduces banding artifacts due to the loss of information.

The process of inverse tone mapping, where high dynamic range values are mapped back to a limited range for display, also contributes to banding if quantization levels are too coarse or bit depth is insufficient, resulting in visible errors. In essence, these diverse causes all trace back to variations in quantization levels.

Banding artifacts can significantly detract from the visual quality and fidelity of images. They can create a posterization effect, where continuous tones are replaced by distinct, flat regions of color. This effect is particularly pronounced in images featuring smooth gradients. The outcome is reminiscent of the image being segmented into blocks, each showcasing uniform colors with sharp boundaries demarcating them. The study by Huang *et al.* [1] highlights that changes in intensity may not necessarily contribute to contours, however, the disruption of monotonicity in the distribution emerges as a crucial factor for the human visual system to identify sensitive contours. The presence of banding artifacts is intolerable for viewers, hindering their ability to fully appreciate the image and diminishing its intended allure and realism.

There is a compelling demand to address banding artifacts in order to elevate visual quality and uphold data accuracy. *Image debanding*, referring to the process of reducing or eliminating visible bands in an image, plays a pivotal role in achieving these objectives. This technique is especially crucial in diverse applications like digital media, photography and graph design where visual fidelity is paramount. The fundamental challenge in image debanding lies in the precise differentiation between authentic image features and artifacts. This task is often intricate due to the overlapping characteristics of the two. Aggressively removing banding artifacts poses the risk of sacrificing image detail, particularly in areas with smooth gradients. Striking the right balance between artifact removal and the preservation of image content and quality represents a delicate

Figure 1: Two sample images displaying needle-like banding artifacts. The green squares indicate the corresponding region of the pristine image while the yellow squares indicate the debanding results.

trade-off. Furthermore, the complexity arises from the multitude of sources contributing to banding artifacts, making the design of a universally effective method challenging in eliminating all types of banding across diverse images. Addressing debanding requires a nuanced understanding and tailored approaches to mitigate these diverse sources, involving an exploration of the distinct physical characteristics of various types of banding artifacts for the design of corresponding debanding techniques. This intricate process adds to the overall complexity of finding comprehensive solutions.

Despite the creation of numerous algorithms [2, 3, 4, 5, 6] aimed at minimizing visual distortion caused by variations in quantization, the occurrence of banding artifacts is inevitable. Extensive research has been conducted in the past on image debanding to tackle this challenging task. In the pre-deep learning era, the majority of debanding techniques focused on post-processing methods [7, 8, 9, 1, 10]. These techniques detected banding artifacts by imposing specific priors on them and subsequently removing them. A crucial aspect of these methods is defining accurate priors to distinguish between banding patterns and authentic image details. However, banding patterns can vary considerably in scale, shape and density across different images. Consequently, existing hand-crafted priors are often over-simplified and fail to perform consistently and effectively across a wide variety of images.

Recently, deep learning has emerged as a promising solution for image debanding, as evidenced by studies [11, 12, 13, 14, 15, 16]. Some leveraged end-to-end deep neural networks (DNNs) to transform images with banding effects into their banding-free latent versions. While these existing DNN models outperform traditional post-processing techniques, they lack designs specifically tailored for debanding tasks. The architecture of the general model is designed to be applicable to various tasks, such as image deblurring and deraining, but not explicitly optimized for image debanding. This suggests a lack of specialized design exclusively tailored for the image debanding task. Furthermore, the tasks addressed by these general models typically focus on regions with dense gradients exhibiting large variations, which are distinctly different from image debanding that primarily addresses regions with smooth gradients. Some tailored to a specific cause of banding artifacts, lack universality when it comes to removing banding artifacts arising from different causes. Therefore, deep image debanding is still in its early stages, offering ample opportunities for performance improvement.

In this paper, we present a deep learning approach that recasts image debanding as an image decomposition process. That is, an image with banding effect is perceived as a composite of a banding pattern layer and a latent image layer. Within the context of image decomposition, we propose to employ invertible coupling layers [17] to create an Invertible Neural Network (INN). Our proposed INN includes two pathways: one for the extraction of the banding pattern layer, and the other for latent image prediction. The advantage of using an INN can be interpreted in two directions. Firstly, from the view of the forward pass, the invertibility of INN guarantees that all details are preserved when decomposing an image into two layers, which is essential as any loss of information could lead to the disappearance of vital details, subsequently degrading image quality.

Secondly, the invertibility of the process ensures that the original image can be perfectly reconstructed from the separated layers, which is crucial for verifying the effectiveness of the decomposition process.

In the proposed INN, we introduce two key techniques. The first is a cross-scale architecture. Given that banding patterns can greatly vary in scale, shape and density across different images, it is critical to identify/process these patterns at various scales. As such, we introduce a multi-scale INN sharing the same spirit as existing CNNs and U-Net, utilizing sequential upsampling and downsampling. However, we go further by addressing the primary challenge in multi-scale INN development: efficient multi-scale processing with maintained invertibility. To achieve this, the encoder and decoder blocks are designed for cross-scale representation and connected in a coupling fashion. Each encoder/decoder block, composed of a series of invertible coupling blocks [17], is integrated within a broader coupling structure. This allows partial features from the encoder blocks to be channeled to the corresponding decoder blocks via a short path for cross-scale fusion. This design channels partial features from encoder to corresponding decoder blocks via a short path for cross-scale fusion, which enhances cross-scale processing while preserving invertibility.

The second key and novel technique is termed as *banded deformable convolution*, specifically engineered for detecting and processing banding artifacts. Deformable convolution [18, 19] represents a convolution type that allows the filter to spatially vary, enabling a flexible and data-adaptive receptive field as opposed to a fixed one. However, in the context of image debanding, the standard deformable convolution encounters a significant issue. Adapting the convolution's receptive field to handle the needle-like anisotropic pattern of banding artifacts necessitates a very large filter with substantially more weights. This not only increases computational costs but also renders the model more

susceptible to overfitting. To address these limitations, the proposed banded deformable convolution allows the filter to adapt to the needle-like structures, while maintaining a compact model with less computational cost. To conclude, the major contributions of this paper include:

- We propose a decomposition-based deep invertible model for image debanding and present a cross-scale coupling structure within an invertible cross-scale scheme. This approach effectively distinguishes between authentic image details and banding patterns across multiple scales, enhancing the performance of the debanding process.
- We develop a banded deformable convolution technique to efficiently and effectively detect/process anisotropic patterns of banding artifacts in images, while maintaining a compact model size and improving the computational efficiency.
- Experiments demonstrate that our proposed DNN outperforms all existing debanding filters, achieving state-of-the-art performance both quantitatively and qualitatively.

2. Related Work

2.1. No-reference metrics for banding artifacts

Beyond researching image debanding methods, it is essential to develop dedicated no-reference metrics. While mainstream image evaluation metrics such as PSNR, SSIM, and LPIPS focus on supervised measures of similarity to ground truth, they often fall short in accurately quantifying the extent of contamination by banding artifacts. Given that image debanding is a task primarily concerned with enhancing subjective human perception rather than achieving high similarity with ground truth, the design of dedicated no-reference metrics becomes crucial.

These specialized metrics aim to better capture the perceived quality of images affected by banding artifacts, highlighting the importance of subjective evaluation in driving advancements in image debanding research. Tu *et al.* [20] developed BBAND, a Blind BANding Detector index that generates a banding visibility map and output a severity score for an input image based on the map. However, the algorithm employed to generate the map demonstrates limited effectiveness in detecting banding edges across diverse scenarios, which leads to a wrong severity score for many images. Kapoor *et al.* [21] proposed a no-reference deep banding index (DBI), which was trained using a supervised learning mechanism. However, several ground truth within the training dataset inherently exhibit banding artifacts, lead to inaccuracies in the DBI results.

2.2. Image debanding

Image debanding approaches can be categorized into three main groups: pre-processing methods [22], embedded processing methods [23, 24], and post-processing methods [7, 9, 1, 10, 11, 14, 13, 15, 16, 12]. Pre-processing methods are typically employed before or after image encoding to mitigate or eliminate discontinuities, such as bands, in the original image, which may be exacerbated during the encoding process. Another strategy involves embedded processing, where the quantization process is adjusted within the encoder to reduce banding artifacts. Widely utilized techniques fall into the category of post-processing methods, which target the output of the decoder. Among these, post-processing methods have garnered notable attention for their practical flexibility in real-world applications and will be thoroughly explored as the primary focus of this subsection.

Non-learnable approaches The majority of non-learnable post-processing debanding methods [7, 9, 1, 10] are prior-driven approaches whose performance depends on how accurately the priors fit data. Huang *et al.* [1] proposed an

accurate false contour detection method based on the monotonicity of false contours, and utilized probabilistic dithering to remove banding artifacts. Tu *et al.* [10] treated image debanding as a reconstruction-requantization challenge, utilizing adaptive interpolation for high-bit-depth estimation in banded areas, followed by dithered quantization. These methods heavily rely on pre-defined priors for image details and banding artifacts, limiting their adaptability to real-world images with diverse banding patterns and content variations. Moreover, achieving optimal performance often requires meticulous hyperparameter calibration.

Deep learning-based approaches Similar to other image restoration tasks, deep learning [11, 14, 13, 15, 16, 12] has proven to be an effective tool for image debanding, leveraging its abundance of trainable parameters. Deng *et al.* [16] introduced a single Spatio-Temporal Deformable Convolution tailored for compressed video quality enhancement. However, its design is specifically oriented towards video restoration, and banding artifacts are just one manifestation of the broader category of compression artifacts. Jiang *et al.* [15] proposed a flexible blind convolutional neural network that separates the quality factor from the JPEG image, incorporating it into the subsequent reconstructor module. However, its utility is confined to addressing JPEG artifacts, representing only one contributor to the generation of banding artifacts. Zhou *et al.* [14] created a large-scale open-source dataset for image debanding tasks and trained Pix2Pix [25], a conditional generative adversarial network, on this dataset. However, Pix2Pix is a general image translation model that lacks a specific design tailored to image debanding. In another approach, Zhao *et al.* [13] decomposed the debanding process into a flat region detection module and a debanding module, utilizing a generated mask to guide the debanding task. Nonetheless, the target labels for the flat region detection module were generated using an oversimplified

gradient extraction operator, limiting the model's capacity.

2.3. Related DNN Techniques

INNs In recent years, INNs have achieved significant success in various image restoration tasks such as denoising, super-resolution, and image compression [26, 27, 28, 29]. INN is structured as a shared-weight auto-encoder with invertibility, enabling the recovery of the original input from the output. This unique property allows for artifact removal, including noise and other unwanted elements, while preserving all information from the original image. Despite the study of INNs in many image restoration tasks, there is a notable absence of research on their application to image debanding. In this study, we present an approach that specifically tailors INNs for image debanding, introducing novel designs to address this particular task.

Deformable convolution Deformable convolution [18, 19] is a convolutional operation that enhances matching capabilities by adaptively adjusting the kernel to the local structure of an image. It has demonstrated effectiveness across various tasks [16, 30, 31, 32, 33] including image recovery. In the context of image recovery, deformable convolution with spatially-varying kernels allows enlarging the receptive field, resulting in improved prediction without sacrificing fine-scale image details or introducing artifacts. Although deformable convolution has shown its utility in existing DNNs for image recovery tasks, it is essential to design problem-specific deformable convolution modules that do not excessively increase model complexity and computational costs in order to fully leverage its advantages derived from its adaptability to the local structure. In this paper, we propose a tailored deformable convolution module specifically designed for image debanding, aiming to effectively harness its benefits.

3. Methodology

3.1. Cross-Scale Invertible Network Architecture

The proposed DNN, denoted as $\mathcal{G}$, is illustrated in Figure 2 (a). It takes a degraded image $\boldsymbol{Y} \in \mathbb{R}^{C\times H\times W}$ and its replica as input and decomposes them into the desired latent image layer $\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$ and the banding layer $\boldsymbol{B} \in \mathbb{R}^{C\times H\times W}$ using the following mapping:

$$\mathcal{G} : (\boldsymbol{Y}, \boldsymbol{Y}) \to (\boldsymbol{X}, \boldsymbol{B}). \tag{1}$$

Notably, the sizes of the input $(\boldsymbol{Y}, \boldsymbol{Y})$ and the output $(\boldsymbol{X}, \boldsymbol{B})$ are both $2C \times H \times W$, which is configured to satisfy the dimensional consistency constraint of INN. To process the input features, our DNN employs a U-shaped cross-scale invertible structure comprising three Encoder Blocks (EBs) and three Decoder Blocks (DBs), with two Banded Deformable Attentive Coupling Blocks (BDACBs) in between. Each EB includes a pixel unshuffle layer and m Dense Coupling Blocks (DCBs). Similarly, each DB consists of m DCBs and a pixel shuffle layer. The specific values of m for each EB and DB are detailed in Figure 2(a). The BDACB in the bottleneck enhances adaptive and spatially-varying processing of local structure. Throughout the DNN, information from previous blocks is propagated to subsequent blocks, ensuring comprehensive information flow.

The pixel shuffle and unshuffle layers [34], which do not possess learnable parameters, play a crucial role in forming a cross-scale representation within the network. The unshuffle layer reduces the spatial resolution to one-fourth of the original size while quadrupling the channel number. Conversely, the shuffle layer performs the inverse operation, allowing the two layers to be reversible. Moreover we introduce a cross-scale coupling structure for further improvement. The output of each EB is split into two parts, where one part is passed to the corresponding DB via a path with several cascaded DCBs, and the other part

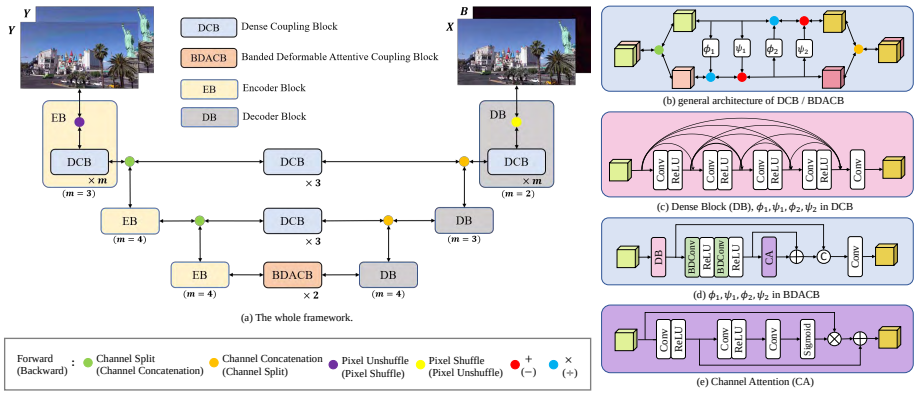


Figure 2: Architecture of proposed DNN for image debanding. All BDConv (Banded Deformable Convolutional) layers use 3×5 kernels.

is passed to the next EB through the skip connection. Accordingly, each DB concatenates the features from the corresponding EB and the features from its previous layer as the input. Indeed, each EB and DB can be viewed as a larger coupling layer nested by some small coupling layers. Such a design empowers the proposed model with better cross-scale analysis capability while keeping invertibility.

3.2. Modules

Dense coupling blocks The DCB depicted in Figure 2(b) is a coupling block [17] specifically designed for invertible feature transformation. It utilizes a double-branch structure that enables straightforward reconstruction of its input from its output using the inverse mode, whose inner learnable functions ϕ_1, ψ_1, ϕ_2, ψ_2 are implemented with the same structure based on dense blocks [35], shown in Figure 2(c). The dense block connects each layer's output to the inputs of all subsequent layers, fostering dense information flow and feature reuse within the network. The invertibility of DCB ensures perfect information fidelity throughout the feature processing stage.

The input features $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are transformed and interact with each other

through the learnable functions ϕ_1, ψ_1, ϕ_2, ψ_2 in a coupling manner. This results in two output features $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$:

$$\boldsymbol{y_2} = \boldsymbol{x_2} \odot e^{\phi_1(\boldsymbol{x_1};\Theta)} + \psi_1(\boldsymbol{x_1};\Theta), \quad \boldsymbol{y_1} = \boldsymbol{x_1} \odot e^{\phi_2(\boldsymbol{x_2};\Theta)} + \psi_2(\boldsymbol{x_2};\Theta), \quad (2)$$

where $\odot$ denotes element-wise multiplication and Θ denotes the whole set of learnable weights. The inversion procedure is defined as follows:

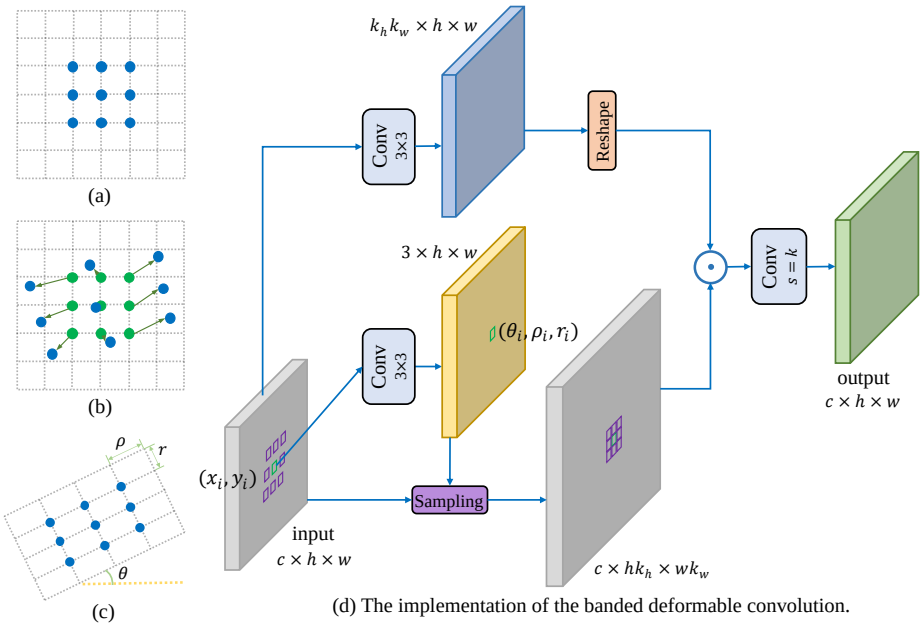
$$\boldsymbol{x_1} = (\boldsymbol{y_1} - \psi_2(\boldsymbol{y_2};\Theta)) \oslash e^{\phi_2(\boldsymbol{y_2};\Theta)}, \quad \boldsymbol{x_2} = (\boldsymbol{y_2} - \psi_1(\boldsymbol{y_1};\Theta)) \oslash e^{\phi_1(\boldsymbol{y_1};\Theta)}, \quad (3)$$

where $\oslash$ denotes element-wise division. Refer to Figure 2(b) for more details.

Banded deformable attentive coupling blocks The Banded deformable attentive coupling block (BDACB) shares the same architecture with DCB, as illustrated in Figure 2(b). It is only inserted at the bottleneck of the neural network for computational efficiency considerations. Its inner functions (i.e., $\phi_1, \psi_1, \phi_2, \psi_2$), as depicted in Figure 2(d), are cascaded with a dense block and a residual channel attention module composed of a series of banded deformable convolutions to effectively distinguish between banding artifacts and image details in the latent feature space.

All banded deformable convolutions here have a kernel size of 3×5, which better exploits the anisotropic structure of edges.[2] The structure depicted in Figure 2(e) illustrates the channel attention module, which utilizes two skip connections to preserve information.

[2]This setting encourages more anisotropic processing. As will be seen later, the banded deformable convolution is rotatable. Therefore, it does not matter whether using a kernel size of 3×5 or 5×3.



(d) The implementation of the banded deformable convolution.

Figure 3: Illustration of sampling principles for three types of convolutions with kernel size 3×3 and dilation 1. Blue points indicate the final sampling locations, while green points represent the reference points for augmented offsets. (a) Regular gridded sampling for standard convolution. (b) Deformed sampling with augmented offsets based on reference points in deformable convolution. (c) Deformed gridding in banded deformable convolution. (d) The implementation of the banded deformable convolution.

3.3. Banded Deformable Convolution

Banded deformable convolution, a specifically tailored version of standard deformable convolution, is extensively utilized in the proposed DNN. To elaborate, in standard deformable convolution, distinct offsets $\{\boldsymbol{\Delta p}_n \in \mathbb{R}^2\}_{n=1}^{k^2}$ are learned for each center location $\boldsymbol{p}_0$ within a convolution sliding window of shape (k, k), based on k^2 pristine sampling points $\{\boldsymbol{p}_n \in \mathbb{Z}^2\}_{n=1}^{k^2}$. The final sampling location set for $\boldsymbol{p}_0$ is then represented as $\boldsymbol{p}_0 + \boldsymbol{p}_n + \boldsymbol{\Delta p}_n$, utilizing bilinear interpolation to calculate final non-integer sampling locations. In this process, $2k^2$ parameters are required to represent the sampling locations for each center pixel. Consequently, when working with an input feature of shape (c, h, w), an intermediary feature of shape $(2k^2, h, w)$ is computed to facilitate the representation of the final sampling locations for the entire feature.

The following mathematical formulas express the sampling principle of banded deformable convolution in polar coordinates. Here, we present its concept using an example of 3×3 kernel with dilation 1. Let the regular point set $\mathbb{P}_{\boldsymbol{p}_0} = \{(p_x, p_y)|(-1,-1),(-1,0),(-1,1),\cdots,(1,1)\}$ define the initial receptive field size and dilation. Then, for each location $\boldsymbol{p}_0$ on the output feature map $\boldsymbol{y}$, we have

$$\boldsymbol{y}\left(\boldsymbol{p}_0\right) = \sum_{\boldsymbol{p}_n \in \mathbb{P}_{\boldsymbol{p}_0}} \omega\left(\boldsymbol{p}_n\right) \cdot \boldsymbol{x}\left(\boldsymbol{p}_0 + \boldsymbol{p}_n + \boldsymbol{\Delta p}_n\right), \tag{4}$$

where $\omega(\boldsymbol{p}_n)$ denotes the kernel weight on $\boldsymbol{p}_n$. In the proposed banded deformable convolution, the regular point set $\mathbb{P}_{\boldsymbol{p}_0}$ is augmented with a rotation angle θ, two scale factors ρ and r:

$$\begin{aligned} \mathbb{P}'_{\boldsymbol{p}_0} = \{(p'_x, p'_y)|p'_x = r\cos\theta \cdot p_x - \rho\sin\theta \cdot p_y, \\ p'_y = r\sin\theta \cdot p_x + \rho\cos\theta \cdot p_y, \forall (p_x, p_y) \in \mathbb{P}_{\boldsymbol{p}_0}\}. \end{aligned} \tag{5}$$

Then, Eq. (4) can be expressed as

$$\boldsymbol{y}\left(\boldsymbol{p}_0\right)=\sum_{\boldsymbol{p}_n^{\prime}\in\mathbb{P}_{\boldsymbol{p}_0}^{\prime}}\omega\left(\boldsymbol{p}_n^{\prime}\right)\cdot\boldsymbol{x}\left(\boldsymbol{p}_0+\boldsymbol{p}_n^{\prime}\right), \tag{6}$$

where non-integer coordinates values are calculated via bi-linear interpolation. The rotation angle θ and two scale factors ρ and r, totally three values for each center pixel, are generated by a convolutional block shown in Figure 3 (d).

See Figure 3 for a comparison between the proposed banded deformable convolution and standard convolution and deformable convolution. Unlike standard deformable convolution, which results in an intermediate feature of shape $(2k^2, h, w)$ for an input with shape (c, h, w), our banded deformable convolution generates an intermediate feature of shape $(3, h, w)$. This significant reduction in parameter number enhances computational efficiency and reduces the risk of over-fitting.

Furthermore, the proposed banded deformable convolution enhances the network's analysis capability. Banding edges have a needle-like structure that can be characterized by two scaling factors (horizontal and vertical) and a rotation factor. Accordingly, the banded deformable convolution utilizes these factors to adaptively parameterize the sampling locations. Specifically, since banding edges as well as image edges can occur in various directions beyond just vertical and horizontal, rotation is employed in our banded deformable convolutions to better capture and discriminate banding edges from image edges. This module allows the convolution to adaptively focus on the most relevant neighboring pixels. The anisotropic property of banded deformable convolution proves beneficial for effectively detecting and processing false contour edges in banding artifacts, as confirmed by our experimental observations.

3.4. Training Loss

Let $\boldsymbol{X}$ and $\boldsymbol{B}$ represent the debanded image layer and banding layer obtained from our proposed DNN for a degraded image $\boldsymbol{Y}$ with its corresponding ground truth image $\boldsymbol{X}'$. Recall that the degradation process caused by banding artifacts is not simply additive, *i.e.*, we cannot assume that $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{B}$. As a result, the supervision data for the banding layer is unavailable as it cannot be simply obtained by subtracting the input with the input image. Therefore, we do not supervise the banding layer $\boldsymbol{B}$ but supervise the latent image layer $\boldsymbol{X}$ only. The total loss $\mathcal{L}$ consists of content loss $\mathcal{L}_{\mathrm{c}}$ and frequency loss $\mathcal{L}_{\mathrm{f}}$, which is given by:

$$\mathcal{L} = \mathcal{L}_{\mathrm{c}} + \lambda_{\mathrm{f}}\mathcal{L}_{\mathrm{f}}, \quad \lambda_{\mathrm{f}} \in \mathbb{R}^{+}$$

The content loss $\mathcal{L}_c$ measuring the reconstruction accuracy of the latent clean image, is defined by

$$\mathcal{L}_{\mathrm{c}} = \|\boldsymbol{X} - \boldsymbol{X}'\|_1. \tag{7}$$

The second term of the total loss is the frequency loss, which is given by:

$$\mathcal{L}_{\mathrm{f}} = \|\mathcal{F}(\boldsymbol{X}) - \mathcal{F}(\boldsymbol{X}')\|_1. \tag{8}$$

Since banding artifacts involve the introduction of erroneous high-frequency information into an image, the utilization of the frequency loss can effectively suppress these artifacts.

4. Experiments

4.1. Experimental Settings

Dataset The dataset is provided by Zhou *et al.* [14], which comprises 1440 pairs of Full High-Definition (FHD) images. Each image is initially divided into

256×256 patches with a step size of 128. After filtering out pairs which degraded image devoid of banding artifacts, the remaining pairs are divided into training (60%), validation (20%), and test (20%) sets while ensuring all patches from the same FHD image belong to the same set. Training involves the use of the segmented patches, totaling 30,829 image pairs, while validation and testing use the complete FHD images.

Implementation details The weights λ_{f} in the loss function is set to 0.1. In the training phase, Xavier [36] initialization is utilized to initial all the model weights. We utilize the Adam optimizer [37] along with a cosine learning rate and a batch size of 4. The implementation of our proposed INN is carried out using the PyTorch framework, and the computations are executed on an NVIDIA Geforce RTX 4090 GPU. The code will be made public via `https://github.com/csxyhe/BDINN`.

Performance metrics We adapt DBI [21] for quantitative comparison, which is a metric specifically designed for assessing the quality of image banding. However, as mentioned above, the capacity of DBI is limited by its supervised training. To comprehensively showcase the debanding capabilities of each method, We also adapt three mainstream metrics including PSNR(Peak Signal-to-Noise Ratio), SSIM (Structural SIMilarity index), LPIPS (Learned Perceptual Image Patch Similarity) [38] for quantitative comparison.

In addition, we implement a subjective quality assessment based on human perception using the double-stimulus impairment scale (DSIS) method. We invited a total of 20 professional observers to rate the debanding results by comparing them to the ground truth images. In this DSIS evaluation, impairment scores range from 1 to 5, with scores of 1, 2, 3, 4, and 5 representing 'very annoying impairment', 'annoying impairment', 'slightly annoying impairment', 'perceptible impairment' and 'imperceptible impairment' respectively. To facilitate the

perception of impairments, observers were allowed to switch back and forth between the reference images and the debanded images processed by the various methods under comparison. Higher average impairment score μ(DSIS) indicates a better visual quality achieved by the method.

Methods for comparison We select four debanding methods for quantitative comparison, including FCDR [1], FFmpeg Filters [39], AdaDeband [10] and DeepDeband [14]. The first three are non-learnable methods, while DeepDeband is the only deep-learning based method for universal image debanding tasks with publicly available code, to the best of our knowledge. As the full implementation of FCDR is not publicly available, we implement it ourselves based on the code provided in [10]. We try six suggested parameter settings and select the best-performing one, i.e., a window size of 9 for both probabilistic dithering and averaging smoothing. We use the default parameters for FFmpeg Filters. For AdaDeband, we compare with its default parameter settings for the YUV420p format. For DeepDeband, we retrain it using the same data configuration as ours and select the best-performing model parameters based on its numerical performance on the validation set.

Since there are few deep learning methods specifically designed for debanding tasks right now, we also consider employing deep learning methods from other related tasks for comparison, including BitNet [40] for bit-depth expansion task, several denoising methods such as ADNet [41] and BRDNet [42], and the general model Uformer-T [43]. For all these deep learning models, we retrain them using the same data configuration and training loss as ours.

4.2. Performance Comparison

Quantitative comparison The quantitative results in terms of four metrics are listed in Table 1. Our proposed method exhibits performance gain over existing methods in terms of all metrics while the second-best performer varies

Table 1: Quantitative performance comparison (3rd-6th columns) in terms of mean evaluation metric scores on the entire test set and comparison of execution time (7th column) on an FHD image with size 1920×1080. Best values are **boldfaced** and second-best values are underlined.

	Method	PSNR(dB)	SSIM	LPIPS	DBI	Time(s)	#Params(M)
N-DL	FCDR	25.73	0.7170	0.3766	0.3657	45.8417	-
	FFmpeg Filter	35.33	0.9352	0.0622	0.1955	-	
	AdaDeband	35.35	0.9392	0.0639	0.2658	4.5162	
Deep Learning	DeepDeband-f	32.95	0.8859	0.0788	0.1735	7.2849	54.414
	DeepDeband-w	32.64	0.9014	0.0717	0.1706	153.6721	54.414
	BitNet	38.24	0.9633	0.0505	0.0939	**0.0346**	0.954
	ADNet	38.29	0.9612	0.0499	0.0689	0.0906	0.521
	BRDNet	38.66	0.9652	0.0460	0.0810	0.3072	1.116
	Uformer-T	39.27	0.9699	0.0463	0.0698	0.3688	5.203
	Ours	**39.37**	**0.9711**	**0.0454**	**0.0668**	0.1728	5.885

regarding different metrics. For instance, it achieves a PSNR improvement of 0.1 dB over Uformer-T and a DBI reduction of 0.0021 compared to ADNet. Additionally, the average impairment scores from the subjective experiment are shown in Table 2, where our proposed method achieves the highest average impairment score among the comparison methods. Such results demonstrate that our proposed model effectively balances between data preservation and visual quality.

Table 2: Average impairment scores on the entire test set. Best values are **boldfaced** and second-best values are underlined.

Method	FCDR	FFmpeg Filter	AdaDeband	DeepDeband-f	DeepDeband-w
μ(DSIS)	1.39	3.16	2.75	3.51	3.43
Method	BitNet	ADNet	BRDNet	Uformer-T	**Ours**
μ(DSIS)	4.24	3.96	4.02	4.42	**4.49**

Complexity comparison We evaluate the execution time as well as the model size of different methods on an 1920×1080 image using the same device. The 'execution time' refers to the time taken for processing the image, excluding the time for loading the image and saving the result. Since obtaining the actual execution time of the FFmpeg toolbox [39] is challenging, its execution time is not provided here. See Table 1 for the results. Our proposed method ranks as the third fastest among all methods, being roughly twice as fast as BRDNet and

Uformer-T. As for model size, the number of parameters of our model is similar to Uformer-T. Such results demonstrate that our proposed model effectively balances between debanding performance and computation complexity.

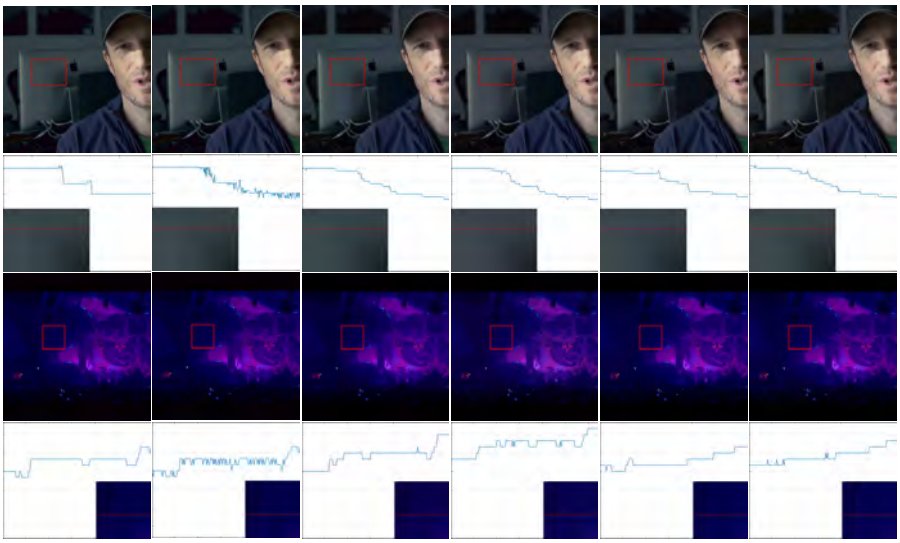


Figure 4: Visual comparison of image debanding results using the zoom-in box with a line profile auxiliary technique. Odd rows display the degradation and corresponding debanding results, while even rows provide zoomed-in views for better inspection. An effective debanding algorithm will result in the line profile of flat areas showing minimal abrupt changes, appearing as horizontal lines or lines with very gentle slopes over relatively long distances.

Qualitative comparison To provide clearer visual comparative analysis of the debanding methods, we attempted two different auxiliary techniques. One is the local zoom-in box with a line profile, and the other is the color segmentation technique described in [9]. While debanding algorithm is effective, the line profile of flat areas will exhibit minimal large jumps and appear as horizontal lines or lines with very small slopes over relatively long distances. The color segmentation technique groups connected pixels with the same RGB color into color blocks and assigns a random color to each block, facilitating intuitive discernment of differences in debanding results. Banding degradation severely affects the flat regions of an image, causing smooth color transitions to be mapped to the same

color. As a result, these flat regions consist of only a few color bands. This is reflected in the color segmentation results, where the corresponding areas display only a limited range of colors. Therefore, the better the debanding result of a method, the smoother the color transitions in the flat regions of the image will be, and the more diverse the colors in the corresponding areas of the color segmentation map will be. We selected four samples of image banding for visual comparative analysis. Two of these samples are presented with zoomed-in boxes showcasing the restoration results. Additionally, the remaining two samples are illustrated using the color segmentation technique. See Figure 4 and Figure 5 for the results under these two scenarios, respectively.

From Figure 4, our proposed method exhibits stronger debanding capability than others. It is apparent that FFmpeg and BRDNet leave many banding effects back. For instance, in the second example of Figure 4, the background of the processed image by FFmpeg, BitNet BRDNet still show multiple significant gradient differences, which are manifested as notable amplitude jumps in the line profile. The debanding visual results of Uformer-T are relatively impressive, but there are still few areas with large pixel value jumps, and its processing time is significantly slower compared to ours. Due to the information fidelity of an INN and the effectiveness of banded deformable convolution, our method preserves image details well and results in natural-looking images.

The results in Figure 5 also show the advantage of our proposed method over other compared ones, which consist with the quantitative results. For instance, in the second example, while methods other than Uformer-T and ours still exhibit noticeable artifacts in the sky area after processing, besides, as indicated by the color segmentation (4-th row of Figure 5), it can be observed that our method achieves a more natural transition compared to Uformer-T. All above results have demonstrated the effectiveness of our proposed method.

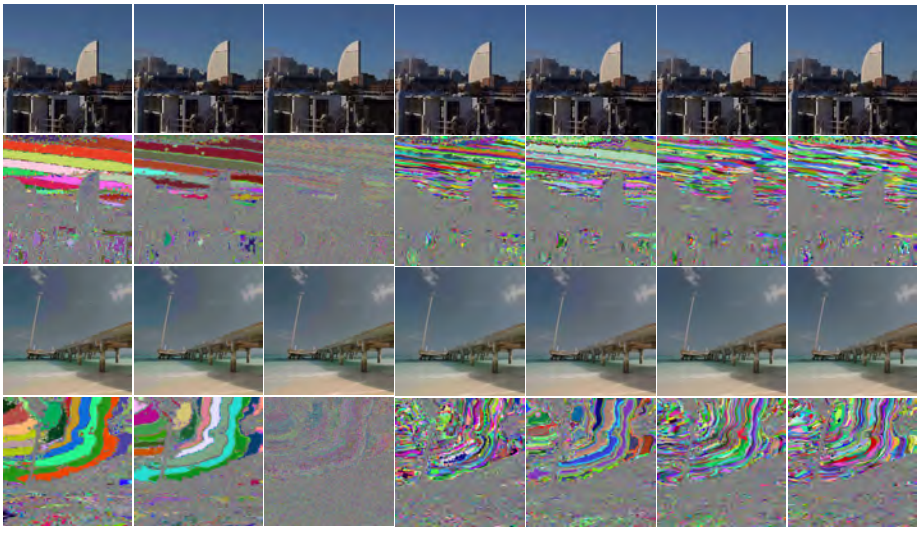


Figure 5: Visual comparison of image debanding results using the color segmentation auxiliary technique. Odd rows display the degradation and corresponding debanding results, while even rows visualize corresponding color segmentation via the method of [9] for better inspection. The better the debanding result of a method, the smoother the color transitions in the flat regions of the image will be, and the more diverse the colors in the corresponding areas of the color segmentation map will be.

Table 3: Ablation study on network's architecture. Best values are **boldfaced**.

Method	PSNR(dB)	SSIM	DBI	#Params(M)	#FLOPs(G)
Non-INN	38.40	0.9634	0.0967	5.943	34.122
Typical INN	38.59	0.9658	0.0843	5.953	97.533
Ours	**39.37**	**0.9711**	**0.0668**	5.885	37.401

4.3. Ablation Study and Analysis

Ablation study on the INN architecture We construct two baseline models for ablation study on our proposed INN architecture for image debanding: (i) Non-INN: Using a standard U-Net with it channel numbers adjusted to have a similar model size to ours; and (ii) Typical INN: the common INN [26] without cross-scale processing, also with it channel numbers adjusted for having a similar model size. The results are listed in Table 3. It can be seen that the performance drops noticeably, *e.g.*, more than 0.5dB PSNR loss, without using the cross-scale invertible structure. Compared to a non-INN structure, employing a standard INN structure yields noticeable improvements, leveraging its information fidelity. Furthermore, utilizing our proposed INN structure results in additional enhancements.

Table 4: Ablation study on key components in terms of PSNR(dB) and DBI.

MSS	DB	BDACB	US/S	PSNR(dB)↑	DBI↓	#Params(M)
	✓			38.57	0.0850	5.942
✓	✓		✓	38.93	0.0792	5.887
✓		✓	✓	39.21	0.0748	5.885
✓	✓	✓		39.06	0.0699	5.985
✓	✓	✓	✓	**39.37**	**0.0668**	5.885

Ablation study on the key components To analyze the contribution of each key component of the model, we form several baseline models. We analyze the effectiveness of the multi-scale structure (MSS), the BDACB module, the dense block (DB) and the unshuffle/shuffle layers (US/S). Specifically, for 'MSS', the multi-scale structure of the coupling pipeline is replaced with several standard coupling blocks. It is worth noting that, the model without 'MSS' does not comprise BDACB module because of the memory constraint. For 'BDACB', we replace the BDACB module with two cascaded DCBs to have a similar model size as original one for a fair comparison. For 'DB', we replace all the dense block with simple convolutions. For fairness, all the above

baseline models are designed to have a similar model size to ours. Besides, the effectiveness of unshuffle/shuffle layers is evaluated by replacing them with convolution/transposed-convolution layers. Different from using unshuffle/shuffle layers, the utilization of convolution/transposed-convolution layers introduce additional learnable parameters. See Table 4 for the results. We can see that each of our utilized components has a performance contribution. For instance, a PSNR gain of 0.44dB and a DBI drop of 0.0124 are achieved by utilizing BDACB module. Since the model replacing unshuffle/shuffle layers with convolution/transposed-convolution layers destroy the network's invertibility, there is a PSNR drop of 0.31dB and a DBI increase of 0.0031.

Ablation study and analysis on banded deformable convolution We test the performance of our proposed DNN using banded deformable convolution with different kernel sizes, including 3×5, 3×7, 3×9, and 5×5. In addition, conventional deformable convolution with kernel sizes 3×5 or 5×5 is also tested. See Table 5 for the results, where we introduce the number of floating-point operations (FLOPs) measured on a 256×256 color image as an additional metric. With less complexity than that of conventional deformable convolution, our proposed banded deformable convolution results in better performance, as it can fit the shape of needle-like structure of banding artifacts more efficiently using only three values to represent the sampling locations for each pixel.

Overall, the 3×5 banded deformable convolution achieved a good balance between performance and computational complexity the optimal choice. In practice, we find that applying a non-square kernel size for banded deformable convolution shows better performance than the one having square kernel size *e.g.*, 5×5. This is probably because the non-square kernel size implicitly includes the prior knowledge of anisotropy banded edges, alleviating possible overfitting.

Analysis on the supervision mechanism To further validate whether

Table 5: Quantitative comparison on conventional/banded deformable convolution with different kernel sizes. Best and second-best values are **boldfaced** and underlined respectively.

Deformable Type	Kernel Size	PSNR(dB)	DBI	#Params(M)	#FLOPs(G)
Conventional	3×5	39.12	0.0764	5.887	37.402
	5×5	39.15	0.0766	6.007	37.525
Banded	3×3	39.14	0.0723	5.861	37.376
	3×5	**39.37**	**0.0668**	5.885	37.401
	3×7	39.23	0.0692	5.909	37.425
	3×9	39.30	**0.0668**	5.933	37.450
	5×5	39.18	0.0720	5.925	37.442

loss on the banding layers helps, we conduct an ablation study that supervised the banding layers $\boldsymbol{B}$ by the residual component $\boldsymbol{R} = \boldsymbol{Y} - \boldsymbol{X}'$. See Table 6 for the results. Notably, the additional supervision on $\boldsymbol{B}$ leads to a PSNR drop of 0.53dB and a DBI increase of 0.0055. This performance drop can be attributed to the fact that imposing $\boldsymbol{B} = \boldsymbol{Y} - \boldsymbol{X}'$ indeed enforces an over-simplified additive modeling of banding, hence leading to negative effects.

Table 6: Quantitative comparison on different supervision mechanisms.

supervise $\boldsymbol{B}$	supervise $\boldsymbol{X}$	PSNR(dB)↑	SSIM↑	LPIPS↓	DBI↓
-	✓	38.84	0.9675	0.0516	0.0723
✓	✓	39.37	0.9711	0.0454	0.0668

5. Conclusion

This paper proposed a novel approach for image debanding by reframing it as an image decomposition problem. Our proposed approach utilizes a cross-scale invertible deep network that is specifically designed for effective image debanding. Additionally, we introduced a module called banded deformable convolution, which is tailored to exploit the anisotropic characteristic of banding artifacts. The introduced banded deformable convolution offers notable advantages in terms of computational efficiency and generalization performance, compared to existing deformable convolutions. In the experiments, our proposed approach

consistently achieved state-of-the-art performance. While our current work focuses on image debanding, it also provides a solid foundation for addressing similar issues in video debanding. Video debanding presents additional challenges, such as maintaining temporal consistency, which may not be optimally addressed by methods designed solely for images. As future work, we plan to extend the model to tackle the complexities of video debanding task, building on the insights gained from our work on image debanding.

Acknowledgments

Yuhui Quan would like to acknowledge the support from National Natural Science Foundation of China (Grant No. 62372186), Natural Science Foundation of Guangdong Province (Grant No. 2023A1515012841), and Fundamental Research Funds for the Central Universities (Grant No. x2jsD2230220). Ruotao Xu would like to acknowledge the partial support from National Natural Science Foundation of China under Grant 62106077 and the partial support Natural Science Foundation of Guangdong Province, China under Grant 2022A1515011087. Yong Xu would like to thank the supports by National Key Research and Development Program of China (Grant No. 2024YFE0105400), National Foreign Expert Project of the Ministry of Science and Technology of China (Grant No. G2023163015L), National Natural Science Foundation of China (Grant Nos. 62472179) and Guangzhou Science and Technology Plan Project - Key R&D Plan(Grant No. 2024B01W0007). Hui Ji would like to thank the support from Singapore MOE AcRF Tier 1 (Grant No. A-8000981-00-00).

References

[1] Q. Huang, H. Y. Kim, W.-J. Tsai, S. Y. Jeong, J. S. Choi, C.-C. J. Kuo, Understanding and removal of false contour in hevc compressed images, IEEE Transactions on Circuits and Systems for Video Technology 28 (2) (2016) 378–391.

[2] X. Yue, D. Miao, L. Cao, Q. Wu, Y. Chen, An efficient color quantization based on generic roughness measure, Pattern Recognition 47 (4) (2014) 1777–1789.

[3] C.-T. Kuo, S.-C. Cheng, Fusion of color edge detection and color quantization for color image watermarking using principal axes analysis, Pattern Recognition 40 (12) (2007) 3691–3704.

[4] B. C. Dhara, B. Chanda, Color image compression based on block truncation coding using pattern fitting principle, Pattern Recognition 40 (9) (2007) 2408–2417.

[5] S. Uchigasaki, T. Miyazaki, S. Omachi, Deep image compression using scene text quality assessment, Pattern Recognition 142 (2023) 109696.

[6] W. K. Han, B. Lee, S. H. Park, K. H. Jin, Abcd: Arbitrary bitwise coefficient for de-quantization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5876–5885.

[7] X. Jin, S. Goto, K. N. Ngan, Composite model-based dc dithering for suppressing contour artifacts in decompressed video, IEEE Transactions on Image Processing 20 (8) (2011) 2110–2121.

[8] H. Noda, M. Niimi, Local map estimation for quality improvement of compressed color images, Pattern Recognition 44 (4) (2011) 788–793.

[9] G. Baugh, A. Kokaram, F. Pitié, Advanced video debanding, in: Proceedings of the European Conference on Visual Media Production, 2014, pp. 1–10.

[10] Z. Tu, J. Lin, Y. Wang, B. Adsumilli, A. C. Bovik, Adaptive debanding filter, IEEE Signal Processing Letters 27 (2020) 1715–1719.

[11] Y. Zhao, R. Wang, W. Jia, W. Zuo, X. Liu, W. Gao, Deep reconstruction of least significant bits for bit-depth expansion, IEEE Transactions on Image Processing 28 (6) (2019) 2847–2859.

[12] Y. Zhao, R. Wang, Y. Chen, W. Jia, X. Liu, W. Gao, Lighter but efficient bit-depth expansion network, IEEE Transactions on Circuits and Systems for Video Technology 31 (5) (2020) 2063–2069.

[13] Y. Zhao, W. Jia, Y. Chen, R. Wang, Fast blind decontouring network, IEEE Transactions on Circuits and Systems for Video Technology (2022).

[14] R. Zhou, S. Athar, Z. Wang, Z. Wang, Deep image debanding, in: Proceedings of the IEEE International Conference on Image Processing, IEEE, 2022, pp. 1951–1955.

[15] J. Jiang, K. Zhang, R. Timofte, Towards flexible blind jpeg artifacts removal, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4997–5006.

[16] J. Deng, L. Wang, S. Pu, C. Zhuo, Spatio-temporal deformable convolution for compressed video quality enhancement, in: Proceedings of the AAAI conference on Artificial Intelligence, Vol. 34, 2020, pp. 10696–10703.

[17] L. Dinh, J. Sohl-Dickstein, S. Bengio, Density estimation using real NVP, in: Proceedings of the International Conference on Learning Representations, 2017.

[18] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 764–773.

[19] X. Zhu, H. Hu, S. Lin, J. Dai, Deformable convnets v2: More deformable, better results, in: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2019, pp. 9308–9316.

[20] Z. Tu, J. Lin, Y. Wang, B. Adsumilli, A. C. Bovik, Bband index: A no-reference banding artifact predictor, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2020, pp. 2712–2716.

[21] A. Kapoor, J. Sapra, Z. Wang, Capturing banding in images: Database construction and objective assessment, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2021, pp. 2425–2429.

[22] S. J. Daly, X. Feng, Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system, in: Proceedings of the Color Imaging VIII: Processing, Hardcopy, and Applications, Vol. 5008, SPIE, 2003, pp. 455–466.

[23] K. Yoo, H. Song, K. Sohn, In-loop selective processing for contour artefact reduction in video coding, Electronics letters 45 (20) (2009) 1020–1022.

[24] N. Casali, M. Naccari, M. Mrak, R. Leonardi, Adaptive quantisation in hevc for contouring artefacts removal in uhd content, in: Proceedings of the IEEE International Conference on Image Processing, IEEE, 2015, pp. 2577–2581.

[25] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[26] Y. Liu, Z. Qin, S. Anwar, P. Ji, D. Kim, S. Caldwell, T. Gedeon, Invertible denoising network: A light solution for real noise removal, in: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2021, pp. 13365–13374.

[27] J. Li, K. Qin, R. Xu, H. Ji, Deep scale-aware image smoothing, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2022, pp. 2105–2109.

[28] M. Xiao, S. Zheng, C. Liu, Y. Wang, D. He, G. Ke, J. Bian, Z. Lin, T.-Y. Liu, Invertible image rescaling, in: Proceedings of the European Conference on Computer Vision, Springer, 2020, pp. 126–144.

[29] J.-J. Huang, P. L. Dragotti, Linn: Lifting inspired invertible neural network for image denoising, in: Proceedings of the European Signal Processing Conference, IEEE, 2021, pp. 636–640.

[30] B. Kim, J. Ponce, B. Ham, Deformable kernel networks for joint image filtering, International Journal of Computer Vision 129 (2) (2021) 579–600.

[31] J. Nie, J. Xie, J. Cao, Y. Pang, Context and detail interaction network for stereo rain streak and raindrop removal, Neural Networks 166 (2023) 215–224.

[32] C. Zhao, W. Zhu, S. Feng, Superpixel guided deformable convolution network for hyperspectral image classification, IEEE Transactions on Image Processing 31 (2022) 3838–3851.

[33] K. Yang, H. Zhang, D. Zhou, L. Liu, Tgan: A simple model update strategy for visual tracking via template-guidance attention network, Neural Networks 144 (2021) 61–74.

[34] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.

[35] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[36] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[37] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[38] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.

[39] F. F. deband, Accessed: Aug. 31, 2021. [online]. available:, `https://ffmpeg.org/ffmpeg-filters.html#deband` (2021).

[40] J. Byun, K. Shim, C. Kim, Bitnet: Learning-based bit-depth expansion, in: Proceedings of the Asian Conference on Computer Vision, Springer, 2019, pp. 67–82.

[41] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, H. Liu, Attention-guided cnn for image denoising, Neural Networks 124 (2020) 117–129.

[42] C. Tian, Y. Xu, W. Zuo, Image denoising using deep cnn with batch renormalization, Neural Networks 121 (2020) 461–473.

[43] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, H. Li, Uformer: A general u-shaped transformer for image restoration, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 17683–17693.