J. Vis. Commun. Image R. 56 (2018) 144–149

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Short Communication

Image up-sampling using deep cascaded neural networks in dual domains for images down-sampled in DCT domain ☆

Kwok-Wai Hung *, Kun Wang, Jianmin Jiang

College of Computer Science and Software Engineering, Shenzhen University, China

ARTICLE INFO

Article history:
Received 10 April 2018
Revised 12 July 2018
Accepted 8 September 2018
Available online 13 September 2018

Keywords:
DCT up-sampling
Interpolation
Super-resolution
Deep neural networks

ABSTRACT

Recent researches show that the high-frequency discrete cosine transform (DCT) coefficients can be estimated from low-frequency DCT coefficients by exploiting the spatial correlations. Hence, images coded by DCT such as JPEG/MJPEG/H.264, etc., can be down-sampled in DCT domain, where the high-frequency information can be accurately restored through image up-sampling. In this letter, we propose a novel deep neural network using the cascaded fully connected layers and convolution layers in dual domains (DCT and spatial domains), in order to restore high-frequency DCT coefficients from observed low-frequency DCT coefficients by exploiting the DCT inter-block and spatial correlations. In the proposed network, many recent techniques are adopted, including residual network in dual domains, batch normalization, denseNet, etc. Experimental results show that the proposed cascaded networks in dual domains significantly outperforms the state-of-the-art DCT up-sampling methods in terms of PSNR (0.63–2.57 dB gain), SSIM values, and subjective evaluations on standard image datasets Set5 and Set14.

1. Introduction

Over the recent years, the discrete cosine transform (DCT) is widely used for image and video coding, such as JPEG, MJPEG, H.264, etc., in order to de-correlate the image signal for higher compression efficiency [1–4]. Hence, there are lots of research efforts devoted to investigate the image re-sizing algorithms in the DCT domain, in order to achieve a higher image quality. DCT re-sizing has become a popular research topic in the last decades [5–18]. Down-sampling in the DCT domain is often done by directly truncating (or discarding) the high-frequency DCT coefficients for simultaneous low-pass filtering and down-sampling process [5–9]. Recent analyses show that down-sampling in the DCT domain can achieve an excellent anti-aliasing and low-pass filtering performance (narrow transition band near Nyquist frequency) and a high level of energy preservation due to characteristics of DCT [7,8]. Moreover, when the image is coded by block-based DCT, the down-sampling process in the DCT domain requires a significantly lower computation compared with spatial down-sampling by converting into the spatial domain. In other words, both the image coding and the image re-sizing processes can be done in the DCT domain only.

The applications of image up-sampling are very wide, such as video coding [1–3], scalable video coding [4,5], image zooming, surveillance, virtual reality, error concealment [10], etc. Hence, many DCT image up-sampling methods using various techniques were proposed to restore truncated high-frequency DCT coefficients [10–18]. However, due to the de-correlation property of DCT, the estimation of high-frequency DCT coefficients from observed low-frequency DCT coefficients is still a non-trivial task [16–18]. Recently, deep learning approaches were proposed for video compression [19,20], enhancement [21] and super-resolution [22–24].

In this letter, we propose a novel deep learning-based approach for restoring the high-frequency DCT coefficients, which were discarded during the DCT down-sampling process. Specifically, our approach uses the deep neural networks in dual domains (DCT and spatial domains) for exploiting the DCT inter-block correlations and spatial correlations respectively, which is the key characteristic of the proposed approach to achieve the state-of-the-art performance. To link the information between dual domains, the proposed cascaded neural network structure have three components, they are the (i) DCT layers (3 fully connected layers) that exploits the inter-block correlations in the DCT domain, (ii) bridge layer (1 fully connected layer) to transform the information from DCT domain to spatial domain, and (iii) spatial layers (33 convolution layers) to exploit the spatial correlations. To achieve the sophisticated performance, the proposed cascaded network structure

☆ This paper has been recommended for acceptance by Dr. Zicheng Liu.

* Corresponding author.
E-mail address: kwhung@szu.edu.cn (K.-W. Hung).

https://doi.org/10.1016/j.jvcir.2018.09.005
1047-3203/

adopts the recent deep learning techniques, including batch normalization [25], residual learning [26], denseNet [27], etc., in dual domains. Experimental results show that the proposed method outperforms deep learning super-resolution methods [22,23] in terms of PSNR, and state-of-the-art DCT up-sampling methods [14,16,17] in terms of objective (PSNR and SSIM) and subjective evaluations. The major contributions of this work are summarized as,

(i) This work proposes a novel deep cascaded structure to exploit the DCT inter-block and spatial correlations in dual domains.
(ii) The proposed deep cascaded neural networks estimate the truncated high-frequency DCT coefficients in dual domains, in order to improve the image up-sampling quality for images down-sampled in the DCT domain.
(iii) The performance of the proposed method significantly outperforms the state-of-the-art DCT up-sampling algorithms in terms of PSNR, SSIM, and subjective evaluations.

The rest of the organization of this paper is as follows. Section 2 gives the literature review of DCT up-sampling methods. The background of DCT down-sampling is described in Section 3. Section 4 describes the details of the proposed deep cascaded networks in dual domains. Section 5 shows the experimental results including objective and subjective evaluations. Section 6 concludes the paper and gives a discussion of future works.

2. Literature review

2.1. Reconstruction-based DCT up-sampling methods

In the literature, up-sampling methods for images down-sampled in the DCT domain can be classified as reconstruction-based and learning-based methods. Reconstruction-based methods include zero padding [7–10], zero padding with overlapping [5], low-pass truncation and subband approximation [11], ringing artifacts reduction [12], trilateral filtering [13], hybrid Wiener filter [14], and frequency sparsity [15], etc. However, the performance of reconstruction-based methods is limited by the information from the observed low-resolution image. Hence, it is essential to exploit the abundant information from external databases.

2.2. Learning-based DCT up-sampling methods

Recent investigations on DCT up-sampling often utilize external information to learn the relationships between the low-resolution and high-resolution images in the spatial domain [16,17] and the DCT domain [18]. However, learning in the DCT domain is a non-trivial task due to the de-correlation properties of DCT coefficients [18]. Hence, recent learning-based approaches often exploit the spatial correlations, for the sake of learning the spatial relationships [16,17] to estimate the truncated high-frequency DCT coefficients in the spatial domain using abundant information from training images.

3. Background of DCT Down-sampling

In this section, let us review the general down-sampling process in the DCT domain for various applications [5–9]. Without loss of generality, let us define the down-sampling factor be dyadic. Let us denote the original high-resolution image in the spatial domain to be $\mathbf{X} \in \Re^{2n \times 2m}$, where n and m are the dimensions of the original image. The high-resolution image is divided into image patches $\mathbf{x}_i \in \mathbf{X}$, such that the block-based DCT down-sampling is formulated as

$$\mathbf{y}_i = DCT^{-1}_{4\times4}(T(DCT_{8\times8}(\mathbf{x}_i))) \tag{1}$$

where $\mathbf{y}_i$ represents the down-sampled low-resolution image blocks, $DCT_{8\times8}(.)$ represents the forward 2D 8×8 DCT transform and $T(.)$ represents the truncation, scaling and down-sampling process. The down-sampled low-resolution blocks forms the low-resolution image $\mathbf{y}_i \in \mathbf{Y}$. Fig. 1 shows the DCT down-sampling process for four concatenated DCT blocks for a clearer illustration. The inverse DCT is not necessary when the down-sampled output is desired to be low-frequency DCT coefficients.

4. Proposed deep neural networks in dual domains

In this section, we will describe the details of the proposed novel DCT up-sampling method, which estimates the high-frequency DCT coefficients in dual domains using the cascaded deep neural networks. Fig. 2 shows the framework of the proposed approach, which has the following steps: (i) the input low-resolution image block is converted into the DCT low-frequency

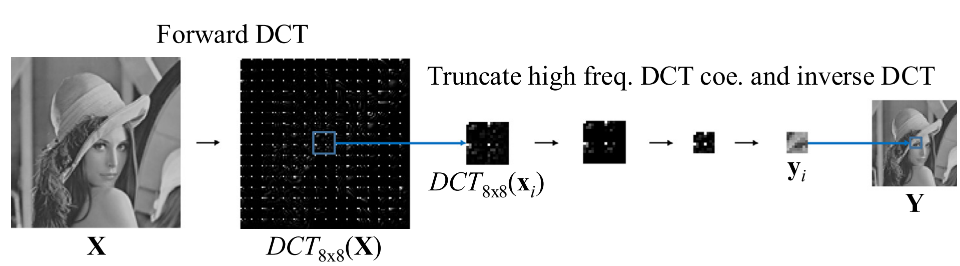

Fig. 1. Dyadic DCT down-sampling process [5–9].

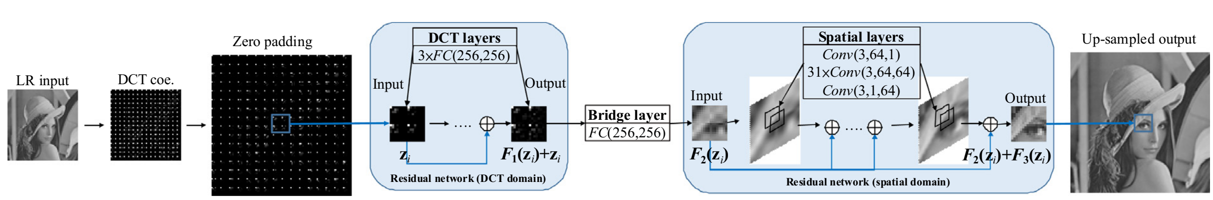

Fig. 2. Detailed framework of proposed deep neural networks in dual domains.