Deep learning approaches for real-time image super-resolution

Pourya Shamsolmoali¹ · M. Emre Celebi² · Ruili Wang³

Published online: 15 July 2020

© Springer-Verlag London Ltd., part of Springer Nature 2020

Generating a high-resolution (HR) image from its corresponding low-resolution (LR) input is referred to image super-resolution (SR). Generally, HR images contain higher pixel densities and more details in comparison with LR images. Image SR has already shown significant performance in many applications, such as video surveillance, remote sensing, face recognition, and medical images. Benefiting from its broad application prospects, SR has attracted enormous interests and is one of the most interesting and active research topics in image processing and computer vision.

Early research in SR mainly focused on the frequency domain. The LR image is transformed into the frequency domain by the Fourier transform or wavelet transform for SR reconstruction. The SR algorithms based on the frequency domain are intuitive and straightforward, but do not consider the degradation process and prior information of an image. Therefore, the reconstruction is not ideal in a complex environment. To address the drawbacks of the frequency domain-based methods, spatial domain-based SR methods have gradually become the focus of mainstream research. Current spatial domain-based methods are mainly divided into two categories: reconstruction-based and learning-based. Reconstruction-based methods are always combined with one or more well-designed priors to estimate the details missed in the reconstruction process.

These methods can obtain good results in preserving edges on the premise that a rational prior has been imposed. Therefore, research has employed a variety of methods to establish reconstruction priors like the sharpening of edge details, regularization, or deconvolution.

Learning-based methods have become a hot spot in SR research in recent years due to their ability to recover high-frequency information of images. Learning-based SR methods establish the mapping of LR image pixels to HR image pixels by learning the spatial structure relationship between HR and LR image and aggregate HR image pixels to reconstruct the HR images. Learning-based methods try to restore missing high-frequency image details by establishing an implicit relationship between LR patches and their corresponding HR patches via machine learning methods. These methods have attracted more and more attention due to their promising and visually desirable reconstruction results. It is a general idea to enhance SR quality by learning relationships from a large quantity of training data. However, applying over data might introduce spurious high frequencies, resulting in noise and blur details. Therefore, it is important to keep a balance between the size of training data and reconstruction visual effects. With the development of machine learning technologies, several learning models have been explored to solve the SR problem. Learning-based methods can be divided into five groups based on differences in their core ideas: neighbor embedding methods, sparse coding methods, self-exemplar methods, locally linear regression methods, and deep learning methods. Recently, due to remarkable advances in deep learning, deep neural networks for SR have shown promising performance in several applications. This special issue collects eight papers reporting the recent developments of deep learning in image SR.

The paper entitled “Perceptual image quality using dual generative adversarial network” develops a variety of generative adversarial networks for image SR that contains two generators and two discriminators. The generators learn from the mixture of real and generated images distributions. This methodology is trained with the feature

✉ Pourya Shamsolmoali
pshams@sjtu.edu.cn

M. Emre Celebi
ecelebi@uca.edu

Ruili Wang
Ruili.wang@massey.ac.nz

¹ Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China

² Department of Computer Science, University of Central Arkansas, Conway, AR, USA

³ School of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand

matching loss to return the detected samples to the corresponding generators to regenerate better real-look samples.

In “CASR: a context-aware residual network for single-image super-resolution,” the authors propose a lightweight context-aware deep residual network, which appropriately encodes channel and spatial attention information to construct a context-aware feature map for single-image SR. An inception network is designed with a novel structure of astrous filters to extract multi-level features from LR images. Then, a Dual-Attention ResNet module is applied to capture context information by dually connecting spatial and channel attention schemes. The proposed inception network only increases a small amount of computation burden and can be easily implemented and adopted by other computer vision tasks.

The paper “Arbitrary-oriented object detection via dense feature fusion and attention model for remote sensing super-resolution image” aims at developing a new arbitrary-oriented object detection method to further push the frontier of object detection for remote sensing images. The proposed method comprehensively takes into account multiple strategies, such as attention mechanism, feature fusion, and SR to boost the performance in terms of localization and classification. Specifically, a dense feature fusion network is designed based on a multi-scale detection framework, which fuses multiple layers of features to improve the sensitivity to small objects. In addition, a rotation anchor strategy is designed to reduce the redundant detection regions.

The paper “A novel super-resolution CT image reconstruction via semi-supervised generative adversarial network” proposes a novel semi-supervised generative adversarial network to accurately recover HR CT images from LR counterparts. A deep unsupervised network of 16 residual blocks is used to design the generator and build a discriminator based on a supervised network in which the bulk specification layer in the commonly used residual network is removed to construct a new type of residual network. Also, for enforcing the mappings between the generator and discriminator, a new variation of the cross-entropy loss function is proposed. The bulk specification layer in the commonly used residual network is removed to construct a new type of residual network.

In “GAN-Poser: an improvised bidirectional GAN model for human motion prediction,” the authors predict human motion using a 3D-based generator. In this model, rather than using the conventional Euclidean loss, a frame-wise geodesic loss is used to have a balanced distribution

of generated data. The discriminator is trained to regress the extrinsic factor, which is used alongside with the intrinsic factor (encoded starting pose sequence) to generate a particular pose sequence. Despite being in a probabilistic framework, the modified discriminator architecture allows predictions of an intermediate part of pose sequence to be used as conditioning for prediction of the latter part of the sequence.

The paper “Spatiotemporal saliency-based multi-stream networks with attention-aware LSTM for action recognition” proposes a novel spatiotemporal saliency-based multi-stream ResNets, which combines three streams (spatial stream, a temporal stream, and a spatiotemporal saliency stream) for human action recognition. The proposed model combines deep convolutional neural network feature extractors with three attention-aware LSTMs to capture the temporal long-term dependency relationships between consecutive video frames, optical flow frames, or spatiotemporal saliency frames.

In “Trainable TV-L1 model as Recurrent Nets for Low-level Vision,” the authors propose a TV-LSTM network to unfold the duality-based iterations of TV-L1 into long short-term memory (LSTM) cells for low-level vision. The proposed end-to-end trainable TV-LSTMs can be naturally connected with various task-specific networks, e.g., optical flow, image decomposition, and event-based optical flow estimation.

The paper “A self-attention-based destruction and construction learning fine-grained image classification method for retail product recognition” proposes an improved fine-grained classification method based on self-attention destruction and construction learning for retail product recognition. Specifically, the proposed method utilizes a self-attention mechanism in the destruction and construction of image information in an end-to-end fashion to calculate a precise fine-grained classification prediction and large information areas in the reasoning process.

Finally, the guest editors wish to thank Professor John MacIntyre (Editor-in-Chief of Neural Computing and Applications), for providing the opportunity to edit this special issue. We would like to thank the authors for submitting contributions and all the reviewers for their most helpful and constructive comments. We hope that the readers benefit from this special issue.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.