Real-Time Deep Video SpaTial Resolution UpConversion SysTem (STRUCT++ Demo)

Wenhan Yang[1*], Shihong Deng[1,2*], Yueyu Hu[1], Junliang Xing[2], Jiaying Liu[1†]
[1]Institute of Computer Science and Technology, Peking University, Beijing, P.R. China
[2]Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China

ABSTRACT

Image and video super-resolution (SR) has been explored for several decades. However, few works are integrated into practical systems for real-time image and video SR. In this work, we present a real-time deep video **S**paTial **R**esolution **U**p**C**onversion Sys**T**em (**STRUCT**++). Our demo system achieves real-time performance (50 fps on CPU for CIF sequences and 45 fps on GPU for HDTV videos) and provides several functions: 1) **batch processing**; 2) **full resolution comparison**; 3) **local region zooming in**. These functions are convenient for super-resolution of a batch of videos (at most 10 videos in parallel), comparisons with other approaches and observations of local details of the SR results. The system is built on a **G**lobal context aggregation and **L**ocal queue jumping **Net**work (**GLNet**). It has a thinner and deeper network structure to *aggregate global context* with an additional *local queue jumping path* to better model local structures of the signal. GLNet achieves state-of-the-art performance for real-time video SR.

KEYWORDS

Real-Time Video Super-Resolution; Batch Processing; Global Context Aggregation; Local Queue Jumping

1 INTRODUCTION

Nowadays, it has gradually become a common demand to embrace high quality video displays. Due to the limitation in current hardware, super-resolution for images and videos by software methods is prevalent and promising. It enlarges a low-resolution (LR) video to a high-resolution (HR) one only employing software techniques. In the past decades, as a scientific research topic, SR methods have been explored widely. Many models are proposed to build the mapping between LR and HR space, *i. e.* Markov random field [4, 7], neighbor embedding [1], sparse coding [6, 11], and anchor regression [10], *etc.* Their results present impressive visual quality. However, most of these works are still far from practical use because of the low time efficiency. Fortunately, the development of deep learning is changing the situation. With deep models proposed in [2, 5, 8], researchers have obtained more promising results with higher time efficiency. To accelerate the SR process, feature extraction and transformation are performed in LR space [9]. Faster super-resolution neural network (FSRCNN) [3] provides the observations that, decreasing the channel number effectively reduces the parameter number of the network, and thus the SR process can be accelerated. Therefore, FSRCNN embeds network shrinking and expanding steps, to save a large part of model parameters and reduce the running time. However, these works have not been integrated into a real-time system for practical image and video SR.

To address this challenge, we construct a practical demo system **STRUCT++** capable of running on both CPU and GPU in real-time manner (50 fps on CPU for CIF sequences and 45 fps on GPU for HDTV videos), supporting three functions including: 1) **batch processing**; 2) **full resolution comparison**; 3) **local region zooming in**. To the best of our knowledge, STRUCT++ is the video SR system that achieves the best evaluation performance comparing with existing approaches and provides very convenient accesses to batch processing and visual comparison.

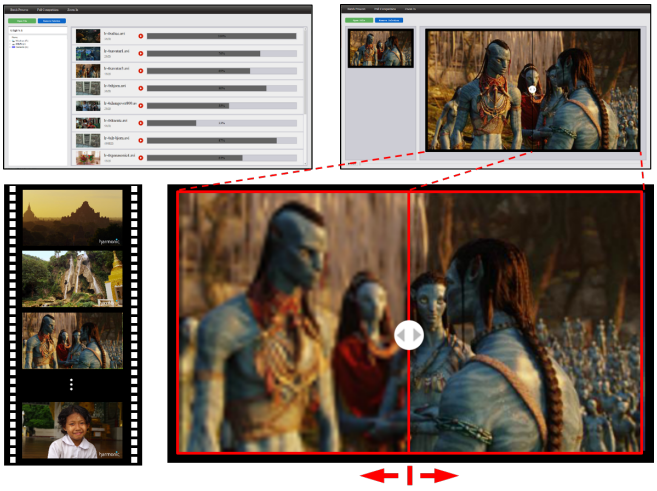

Figure 1: STRUCT++ supports real-time video super-resolution. It provides three functions, including batch processing (the left half), full resolution comparison (the right half) and local region zooming in (Fig. 4).

† Corresponding author. ∗ Wenhan Yang and Shihong Deng contributed equally to this work. This work was supported by National Natural Science Foundation of China under contract No. 61472011 & 61672519 and Microsoft Research Asia (project ID FY17-RES-THEME-013).

MM'17, October 23–27, 2017, Mountain View, CA, USA.
 ISBN 978-1-4503-4906-2/17/10.
DOI: https://doi.org/10.1145/3123266.3127927

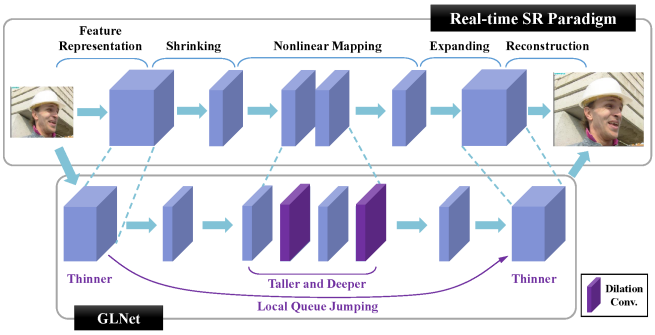

Figure 2: GLNet adopts a thiner and deeper network structure for *global context aggregation*. An additional *local queue jumping* connection helps better model local signals.

2 GLNET FOR REAL-TIME VIDEO SR

STRUCT++ is built on an effective network structure – **G**lobal context aggregation and **L**ocal queue jumping **Net**work (**GLNet**), as shown in Fig. 2. Following real-time SR paradigm, it goes through five steps for image and video SR: feature representation, shrinking, nonlinear mapping, expanding and reconstruction. Comparing with existing real-time SR methods, GLNet has two distinguished characteristics:

- A thiner and deeper network structure for *global context aggregation*. Each layer has fewer channels, leading to a deeper network with the same number of parameters. With dilation convolutions as parts of its units, GLNet has a very large receptive field.
- An additional *local queue jumping* connection between the first and penultimate layers enables the network to better describe the local signal structures.

These two properties jointly help GLNet achieve superior performance to state-of-the-art real-time image and video SR. Fig. 3 demonstrates the relationship between super-resolution quality and running time. As can be observed, GLNet achieves higher response curve, which indicates that GLNet spends less time while generates better results.

3 VIDEO SR RESULT DISPLAYING

In addition to batch processing (as shown in the left half of Fig. 1), and full resolution comparison (illustrated in the right half of Fig. 1), our demo also provides the third function "**Local Region Zooming In**". After clicking the "start" button, the input video will be shown (Fig. 4). User can select a marquee by clicking on the video, and the super-resolved version of the marquee is shown in line on the right side. Multiple marquees are allowed and can be removed by clicking the "selection removal" button.

4 CONCLUSIONS

This paper demonstrates the functions of STRUCT++ and briefly introduces the algorithm in the system. With the efficient GLNet, the system provides convenient operations

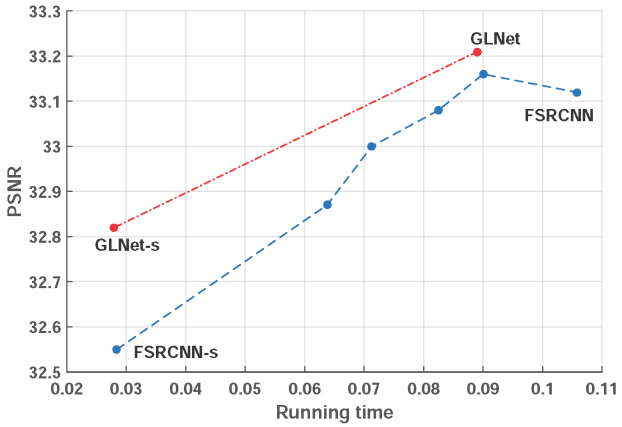

Figure 3: GLNet achieves the best response curve for pairs (*PSNR*, *Running time*) in 3× enlargement on *Set5*.

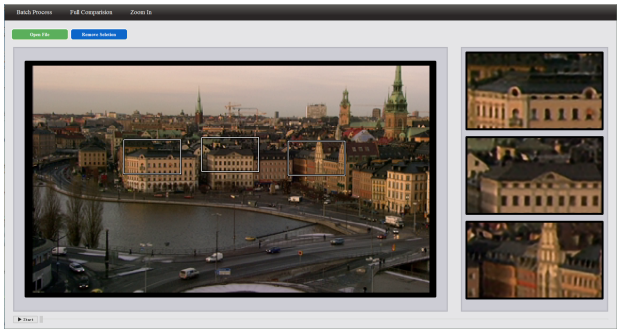

Figure 4: Interface for "Local Region Zooming in".

to super-resolve a batch of videos. The friendly interfaces allow users to compare different methods visually and look into detailed regions of interest in real-time.

REFERENCES

[1] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR*, 2004.
[2] C. Dong, C. Chen, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014.
[3] C. Dong, C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *ECCV*, 2016.
[4] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE CGA*, 2002.
[5] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In *CVPR*, 2016.
[6] J. Liu, W. Yang, X. Zhang, and Z. Guo. Retrieval compensated group structured sparsity for image super-resolution. *TMM*, 2017.
[7] J. Ren, J. Liu, and Z. Guo. Context-aware sparse decomposition for image denoising and super-resolution. *TIP*, 2013.
[8] S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. In *CVPR*, 2015.
[9] W. Shi, J. Caballero, F. Huszr, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*, 2016.
[10] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *ACCV*, 2014.
[11] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *TIP*, 2010.