Temporally adaptive and region-selective signaling of applying multiple neural network models

기세환, 김문철
한국과학기술원 전기 및 전자 공학과
shki@kaist.ac.kr, mkim@ee.kaist.ac.kr

Temporally adaptive and region-selective signaling of applying multiple neural network models

Sehwan Ki, Munchurl Kim
Korea Advanced Institute of Science and Technology Dep. Of Electronic Engineering

요 약

The fine-tuned neural network (NN) model for a whole temporal portion in a video does not always yield the best quality (e.g., PSNR) performance over all regions of each frame in the temporal period. For certain regions (usually homogeneous regions) in a frame for super-resolution (SR), even a simple bicubic interpolation method may yield better PSNR performance than the fine-tuned NN model. When there are multiple NN models available at the receivers where each NN model is trained for a group of images having a specific category of image characteristics, the performance of quality enhancement can be improved by selectively applying an appropriate NN model for each image region according to its image characteristic category to which the NN model was dedicatedly trained. In this case, it is necessary to signal which NN model is applied for each region. This is very advantageous for image restoration and quality enhancement (IRQE) applications at user terminals with limited computing capabilities.

1. Introduction

Convolutional neural networks (CNN) for super-resolution (SR) have shown very promising performance with high fidelity of restoration [2-4]. There have been many studies on CNN-based SR which is still a very hot topic in computer vision. Also, it is possible to deliver encoded video at low spatial resolutions efficiently and to reconstruct them with high restoration fidelity at higher resolution using such CNN-based SR methods at the receiver sides. Moreover, content-adaptive training and transmission of the model parameters of neural networks can boost up the SR performance with higher restoration fidelity. According to the use cases with UC13 Distribution of neural networks for content processing in [1], it was identified that transmission of updated neural network (NN) models is required for different temporal portions of the video. This can be further extended for different spatial regions of the video for better image restoration. However, the size of one single NN model should be large enough to achieve a good performance of quality enhancement and be properly trained with a large amount of training samples. Such large sized NN models necessarily requires a heavy computation and large storage to process input and to produce output, which is prevented from being used in user terminals with a limited computation capability. In practice, it is more practical to use multiple NN models of a (same) manageable size by selectively applying them for appropriate spatial regions of frames and temporal portions in content-adaptive manners. In this case, it is necessary to normatively define the signaling of identifying which NN models are applied for which regions and temporal portions of a video under service. Also, the con-figuration

representation of NN models should be normatively defined in combination of the above mentioned signaling information.

2. Method

The fine-tuned NN model for a whole temporal portion in a video does not always yield the best quality (e.g., PSNR) performance for all regions of each frame in the temporal portion. For certain regions (usually homogeneous regions) in a frame for NSR, a simple bicubic interpolation method may yield better PSNR performance than the fine-tuned NN models. Therefore, when there are multiple NN models available at the receivers where each NN model is trained for a group of images having a specific category of image characteristics, the performance of quality enhancement can be improved by selectively applying an appropriate NN model for each image region according to its image characteristic category to which the NN model was dedicatedly trained. In this case, it is necessary to signal which NN model is applied for each region. This is very advantageous for image restoration and quality enhancement (IRQE) applications at user terminals with limited computation capabilities.

- Case 1: Different NNs are optimized for different categories of content (e.g., image) characteristics. The server signals each NN model for an image region and the region split information of the image before or during the streaming of the image.

- Case 2: Multiple NN models are already available at the client for a target IRQE application. The server needs to deter-mine which NN model is the best applied for each region of the image for IRQE, and to signal the NN model IDs applied for the regions of the image as well as the region split information of the image.

3. Experiments

In experiments, since the AS-model (trained model for all scenes) and the OS-model (trained model for one scene) are already transmitted to client, PSNR performance can be improved by signaling the which NN models applied for spatial regions of a frame with region split information. Fig. 2 shows the signaling process.

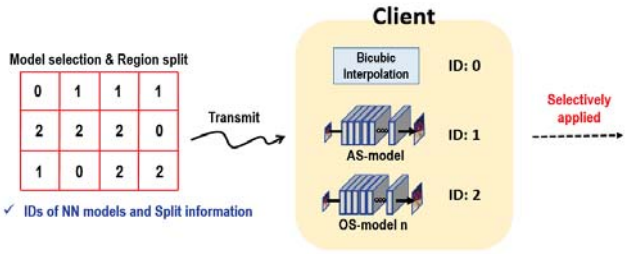

Figure 2. Signaling the which NN models applied for spatial regions of a frame with region split information

Fig. 3 shows the selected model signal maps by each N N block (N is 8, 16, 32, and 64). In this way, when different models are selected for each region, PSNR performance is greatly improved. Table 1 shows the PSNR performance compared to directly applying the same model for all regions. The underlined number means the difference PSNR between the OS-model and the selected model, and the bold number means the difference PSNR between the AS-model and the selected model. In the case of determining the optimal model for each small block, the PSNR performance improvement was higher. However, when deciding the optimal model for

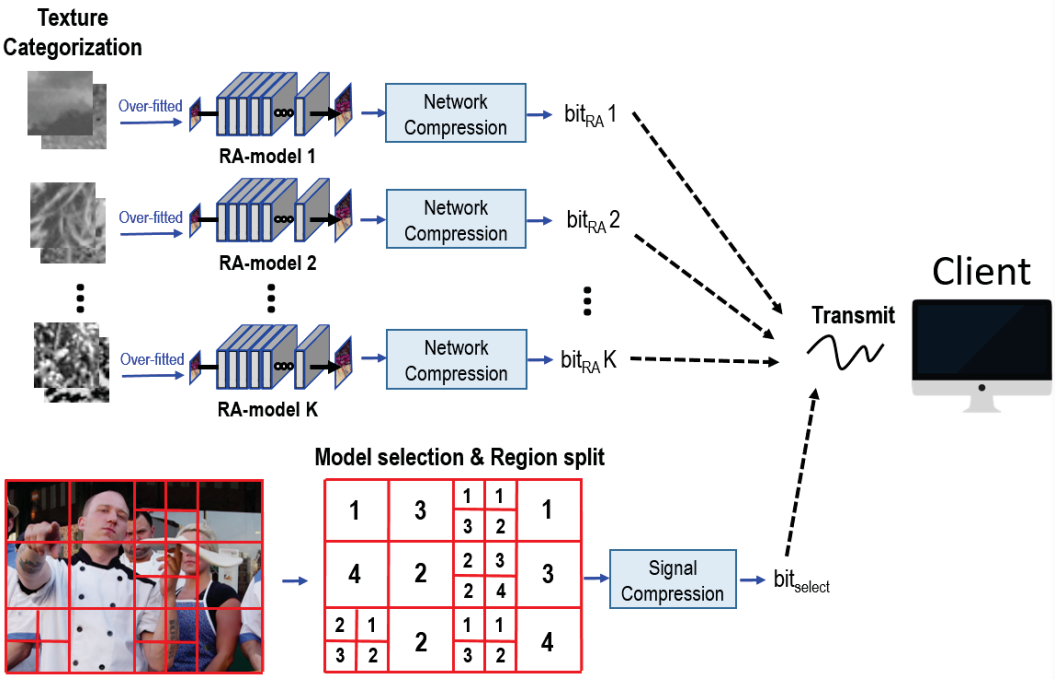

Figure 1. Region-Adaptive NN models and Region-selective signaling

(a) Original frame

: Bicubic Interpolation

: AS-model

: OS-model

(b) Selected model signals by each 8×8 block

(c) Selected model signals by each 16×16 block

(d) Selected model signals by each 32×32 block

(e) Selected model signals by each 64×64 block

Figure 3. The selected model signal maps by each N×N block (N is 8, 16, 32, and 64)

each small block, the amount of signal which is required to be transmitted is very large. Therefore, it is necessary to properly adjust the block size to optimize the tradeoff between the PSNR performance and transmission amount.

Table 1. The PSNR performance of signaling the which NN models applied for spatial regions of a frame

Block Size	PSNR (dB)
8×8	35.9104 (+0.296) (**+0.982**)
16×16	35.7647 (+0.151) (**+0.836**)
32×32	35.7026 (+0.089) (**+0.774**)
64×64	35.6584 (+0.044) (**+0.730**)

4. Conclusions and future works

This contribution presents an updated use cases of the UC13 Distribution of neural networks for content processing in [1]. We propose that WG11 consider the updated use cases of NN model compression for image restoration and quality enhancement such as super-resolution, coding artifact reduction as post processing, contrast enhancement etc. We need to devise an efficient compression method for the signal for model selection in future studies.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00419, Intelligent High Realistic Visual Processing for Smart Broadcasting Media).

References

[1] W. Bailer, et al, “Use cases and requirements for compressed representation of neural networks,” ISO/IEC JTC1/SC29/WG11 N17924, Oct. 2018.

[2] J. Kim, et al., "Accurate image super-resolution using very deep convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016.

[3] Zhang et al, "Image super-resolution using very deep residual channel attention networks." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[4] Zhang, et al. "Residual dense network for image super-resolution." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.

[5] Caglar, et al, “Response to the Call for Proposals on Neural Network Compression: Training Highly Compressible Neural Networks”, ISO/IEC JTC1/SC29/WG11 MPEG2019/m47379, March. 2019.