Multimodal Multi-Head Convolutional Attention with Various Kernel Sizes for Medical Image Super-Resolution – Supplementary

Mariana-Iuliana Georgescu[1], Radu Tudor Ionescu[1], Andreea-Iuliana Miron[2,3], Olivian Savencu[2,3], Nicolae-Cătălin Ristea[1,4], Nicolae Verga[2,3], Fahad Shahbaz Khan[5,6]
[1]University of Bucharest, Romania, [2]"Carol Davila" University of Medicine and Pharmacy, Romania, [3]Colţea Hospital, Romania, [4]University Politehnica of Bucharest, Romania, [5]MBZ University of Artificial Intelligence, UAE, [6]Linköping University, Sweden

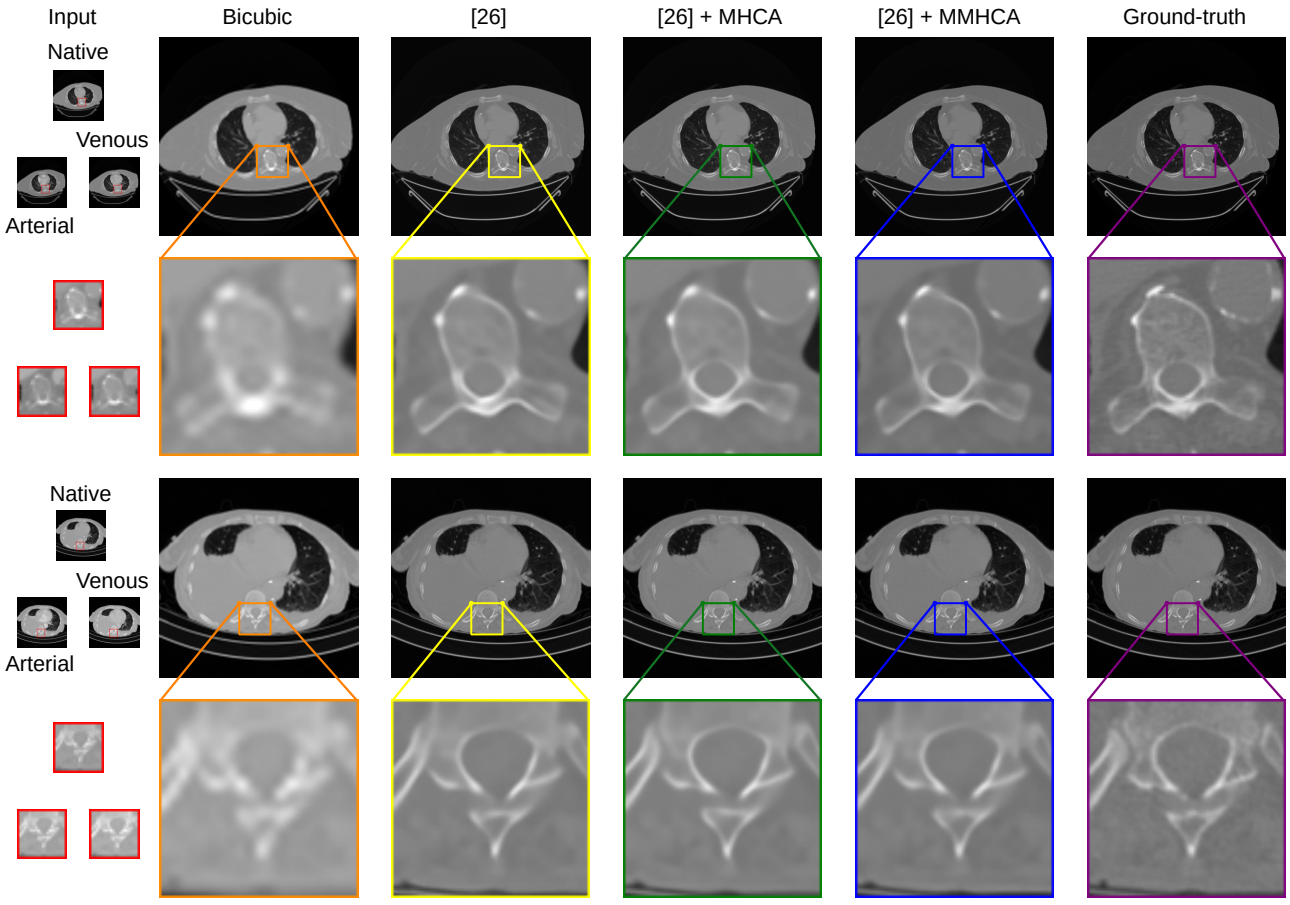

Figure 1: Examples of super-resolved CT images from the Coltea-Lung-CT-100W data set, for an upscaling factor of $4\times$. The HR images produced by two baselines (bicubic interpolation and EDSR [1]) are compared with the images given by two enhanced versions of EDSR [1], one based on our single-contrast attention module (MHCA), and another based on our multimodal attention module (MMHCA).

1. Additional Qualitative Results

In Figure 1, we illustrate qualitative results obtained by two baselines (bicubic and EDSR [1]) versus two enhanced versions of EDSR [1], namely EDSR [1] + MHCA and EDSR [1] + MMHCA, for an upscaling factor of $4\times$. We observe that our EDSR [1] + MMHCA model is able to create sharper reconstructions and to improve the contrast levels.

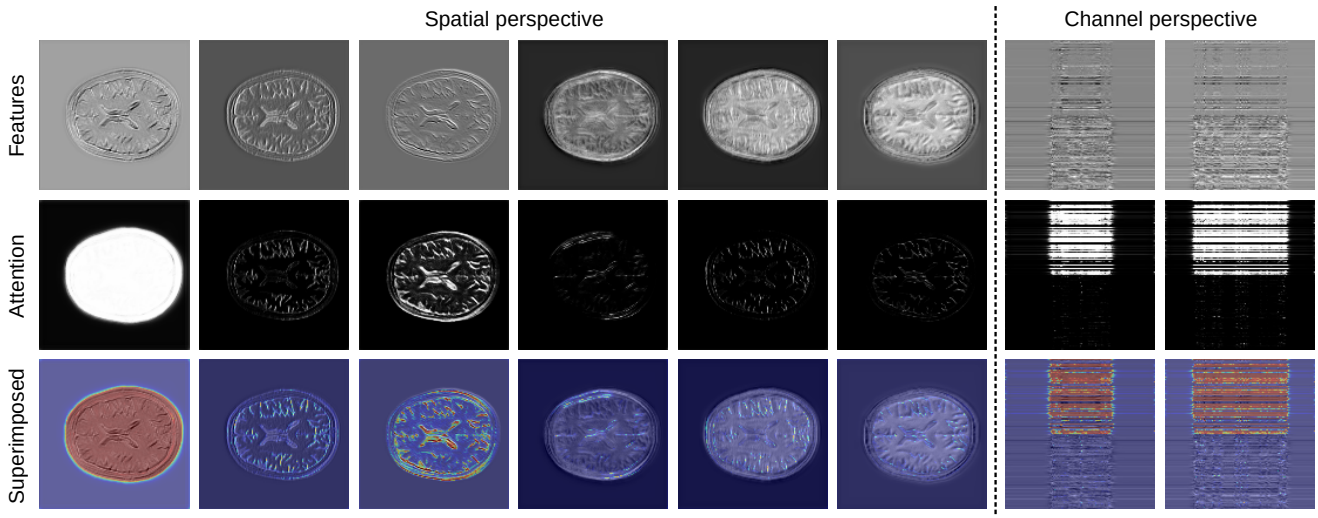

Figure 2: Views (top row) of a tensor computed for an example from NAMIC, which is given as input to MMHCA, and the corresponding attention maps (middle row) along the spatial (first six columns) and channel (last two columns) dimensions, showing that MMHCA performs joint channel and spatial attention. Views with superimposed attention maps are displayed on the bottom row. Best viewed in color.

2. Attention Visualization

In Figure 2, we show various perspectives along the spatial and channel dimensions of a tensor given as input to MMHCA. Looking at the attention corresponding to individual activation maps (first six columns), we observe that our module attends to salient contours and edges, or even full organs. Analyzing the attention along the channel axis (last two columns), we observe that, in this example, our module tends to mainly focus on the first LR input, naturally because the first LR input is the modality (contrast type) that corresponds to the HR output, containing the most relevant information to super-resolve the image. In contrast, the second modality is scarcely attended by our module. Overall, we observe that MMHCA performs both spatial and channel attention, confirming that our attention module works as intended.

References

[1] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of CVPR Workshops*, pages 136–144, 2017.