Image translation using ShuffleUNet

Juhyung (Tony) Ha[1] and JongSeong Park[1]

Indiana University, Bloomington

Abstract. We participated in the "SynthRAD2023" competition, which focuses on medical image translation from MRI/CBCT to CT. To address unsatisfactory image quality coming from artifacts such as blurry and checkboard artifacts, we developed a ShuffleUNet model that uses different down/upsampling modules called "pixel unshuffle" and "pixel shuffle". We also developed 3D pixel unshuffling/shuffling modules to compensate for the 3D dimensionality of medical imaging such as MRI, CT, and CBCT. Our model achieved 63.19/28.67/0.8753 (MAE/PSNR/SSIM) for task 1 validation leaderboard and 58.14/30.36/0.8971 for task 2 validation leaderboard.

Keywords: UNet · Pixel shuffle/unshuffle · Image translation.

1 Introduction

Radiological imaging encompasses a broad spectrum of modalities including MRI, CT, CBCT, X-ray, and more. Each modality offers unique advantages and is optimized for specific clinical scenarios, enabling radiologists to diagnose a plethora of medical conditions with heightened precision. While leveraging multiple modalities concurrently would provide a comprehensive view of patient pathology, it remains infeasible due to practical constraints such as cost, time, and patient comfort. To circumvent these challenges, this study explores the synthetic generation of CT images, termed synthetic CT (sCT), from MRI (task 1) and CBCT (task 2) scans. Generating sCT from MRI eliminates radiation exposure risks, and creating sCT from CBCT reduces radiation doses and scan durations.

The recent proliferation of AI in medicine has spurred the development of myriad CNN-based solutions for medical image translation. This work delves into the potential of deep-learning-driven techniques to facilitate the translation from MRI/CBCT to CT.

2 Methods

2.1 Image processing

The MRI data underwent z-score normalization across the entire 3D image using parameters $\mu = 0$ and $\sigma = 1$. For both CBCT and CT, voxel intensities were scaled according to the formula: $\frac{x+1024}{4024}$, ensuring a standardized intensity range between 0 and 1.

2.2 Model developments

The design and training of our models took into account several considerations, including (1) the choice of up/downsampling modules, (2) the selection of loss functions, and (3) network architecture and data dimensionality specifics. When assessing downsample modules, we evaluated performance metrics across pooling, strided convolution, and pixel unshuffling techniques. Upsampling modules were similarly benchmarked using interpolation, transposed convolution, and pixel shuffling. Our loss function arsenal comprised pixel-wise L1 loss, 2D/3D SSIM loss, and coefficient loss. Furthermore, we conducted experiments comparing the efficacy of 2D versus 3D network structures and input data. After some iterations of empirical experiments, we decided to use pixel shuffling/unshuffling for down/upsampling modules, to solely use pixel-wise L1 loss, and develop 3D-based network.

3 Results

Our model achieved 63.19/28.67/0.8753 (MAE/PSNR/SSIM) for task 1 validation leaderboard and 58.14/30.36/0.8971 for task 2 validation leaderboard.

4 Discussion and Conclusion

With limited available time, we could not perform extensive analysis of various experiments we initially designed. Probably over time, we will develop our model further in this topic.