# Attention Based U-Net for MRI Reconstruction

Qiang Zhao

## Introduction

Introducing deep learning methods has improved the image reconstruction process in fastMRI[1]. Recently, transformer architectures based on attention mechanisms have demonstrated outstanding performance in computer vision by introducing larger receptive fields and long-range dependencies, with numerous applications in medical imaging[2]. In this study, we reference the attention-Unet method to attempt image reconstruction of knee MRI scans[3], testing whether introducing attention mechanisms can enhance performance. By comparing the results of the standard U-Net and the enhanced Attention U-Net on the same dataset and under identical conditions, we aim to determine the effectiveness of this architectural improvement in the context of MRI image reconstruction.

## Methods

### Attention U-NET:

In order to evaluate whether Attention U-Net can enhance image reconstruction performance, we modified partially source code of ***fastmri package*** to incorporate the attention U-Net architecture (Fig.1).

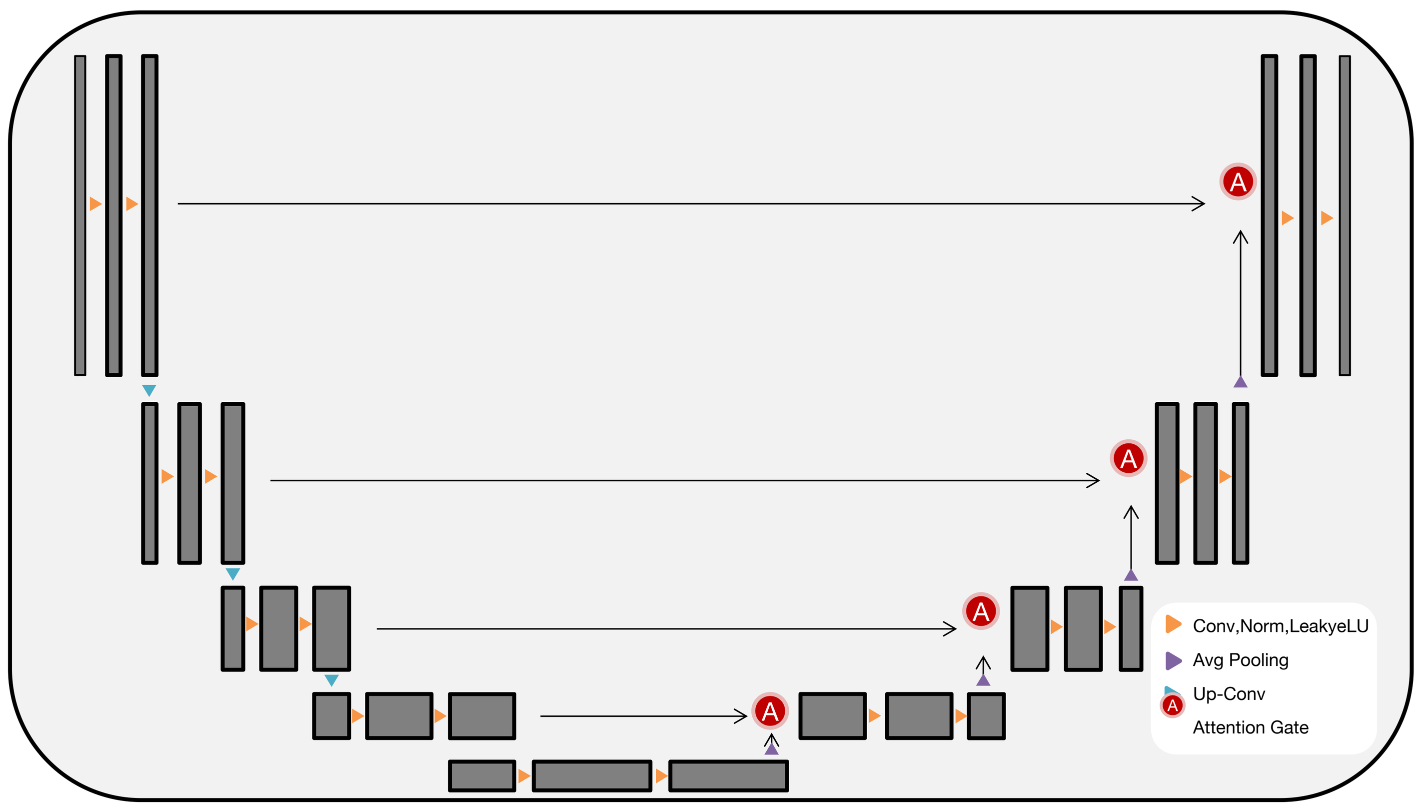



**Figure 1.** Architecture of the Attention U-Net

Attention coefficients (α ∈ [0, 1]) highlight important image regions and prune feature responses to keep only task-relevant activations. The attention gates output is produced by element-wise multiplying input feature maps with attention coefficients. Each pixel vector $x_i^l \in \mathbb{R}^{F_l}$ gets a scalar attention value, where $F$ is the number of feature maps in layer $l$. To handle multiple semantic classes, we use multi-dimensional attention coefficients. A gating vector $g_i \in \mathbb{R}^{F_g}$ identifies focus regions and prunes lower-level feature responses (Fig.2).

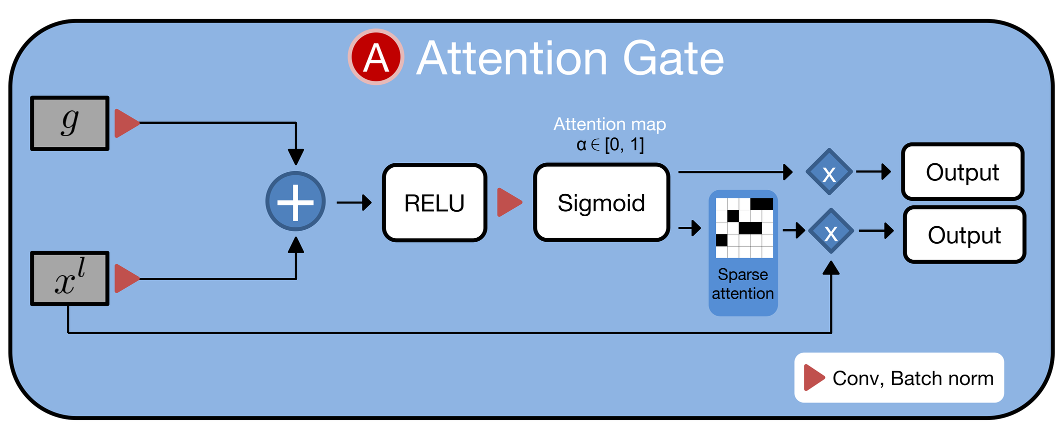



**Figure 2.** Schematic of the attention gate

To evaluate the impact of both conventional attention and sparse attention on model performance, we added sparse attention as a comparison. All image reconstruction results were assessed based on PSNR, SSME and NMSE metrics.

## Results

Figure 3 shows the loss reduction and SSIM comparison on the validation dataset. Attention-based networks initially reduce loss faster, with sparse attention UNet performing similarly to the original UNet but better than non-sparse attention UNet.

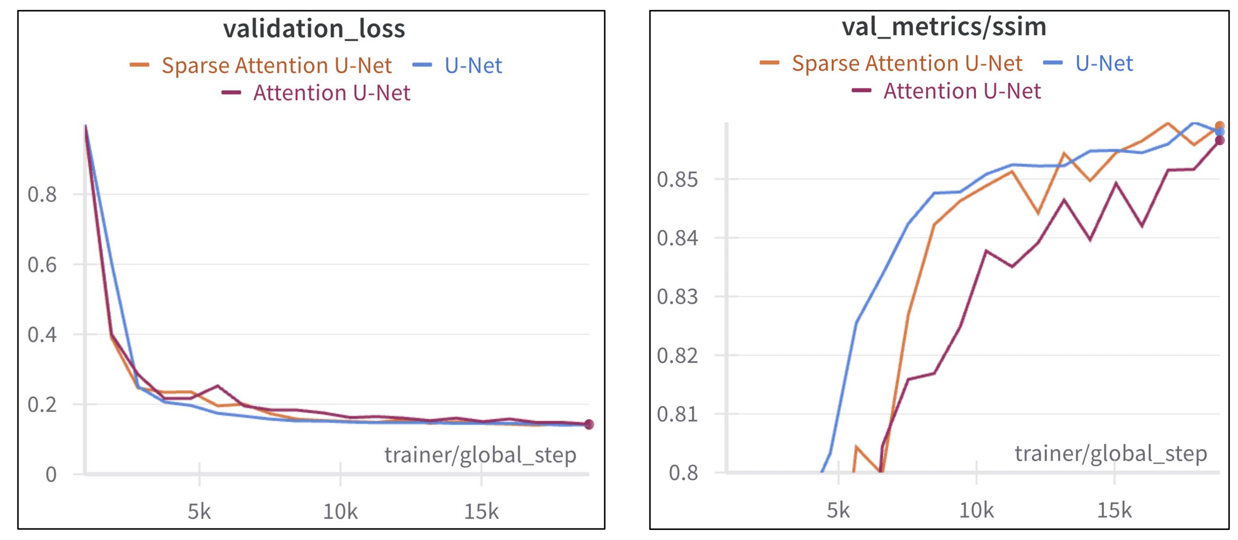



**Figure 3.** The loss and ssim metrics curve on validation data

To further compare model performance, we evaluated SSIM, PSNR, and NMSE on a test set of 41 cases. As shown in Table 1 and Figure 4, the original U-Net outperformed the attention mechanism network in all three metrics.

**Table 1.** Quantitative comparison using different models on test data

| Method | SSIM mean(std) | NMSE mean (std) | PSNR mean(std) |
|---|---|---|---|
| U-Net | **0.843 (0.074)** | **0.023 (0.140)** | **32.798 (2.855)** |
| Attention U-Net | 0.805 (0.097) | 0.049 (0.048) | 30.249 (4.021) |
| Sparse Attention U-Net | 0.837 (0.076) | 0.025 (0.154) | 32.470 (3.052) |

Next, we qualitatively compared the initial and final reconstruction differences with the ground truth. Results show that while all models achieved similar final performance, attention-based models quickly captured image details and contours in the initial stages (Fig.4).

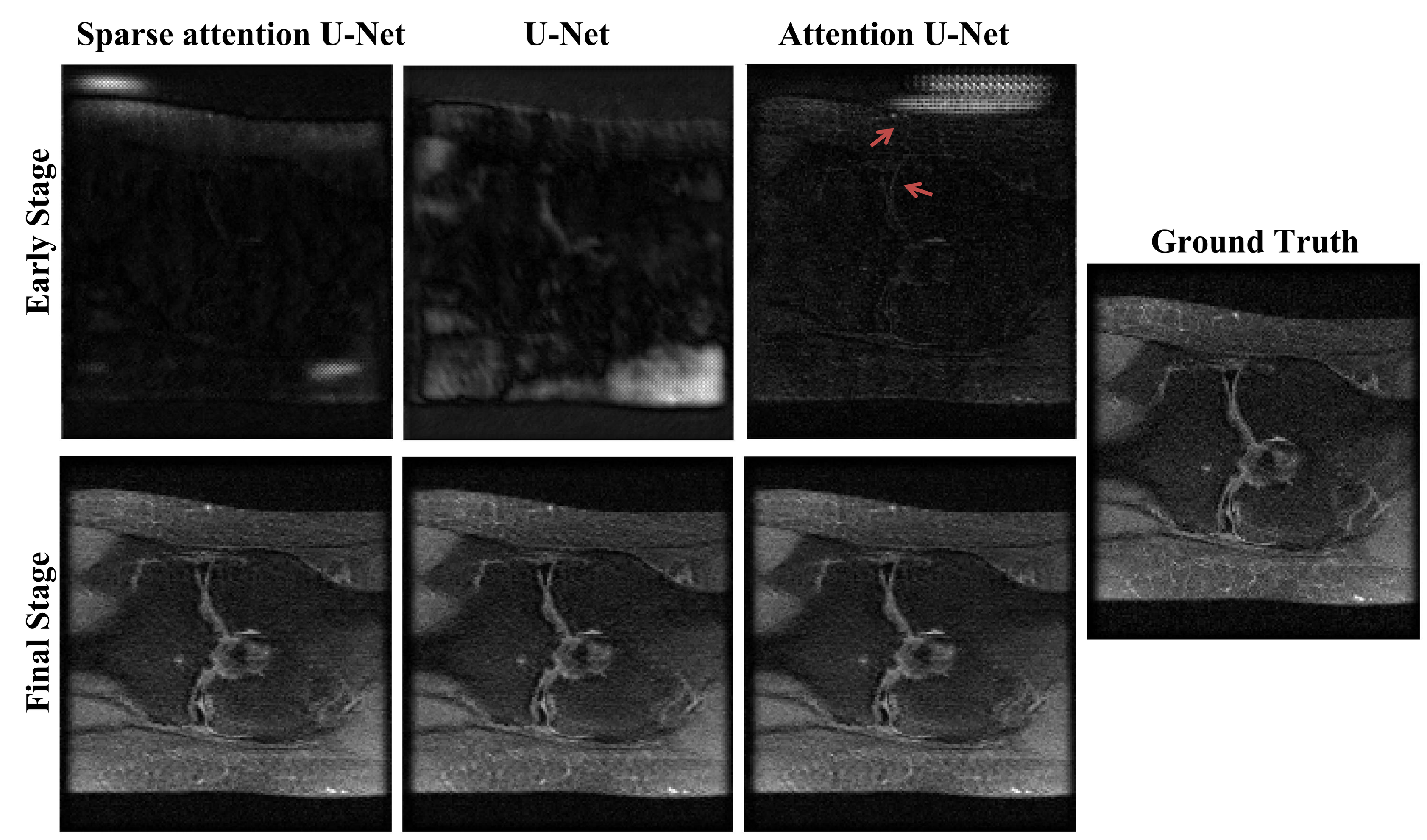



**Figure 4.** The reconstruction comparison using different model

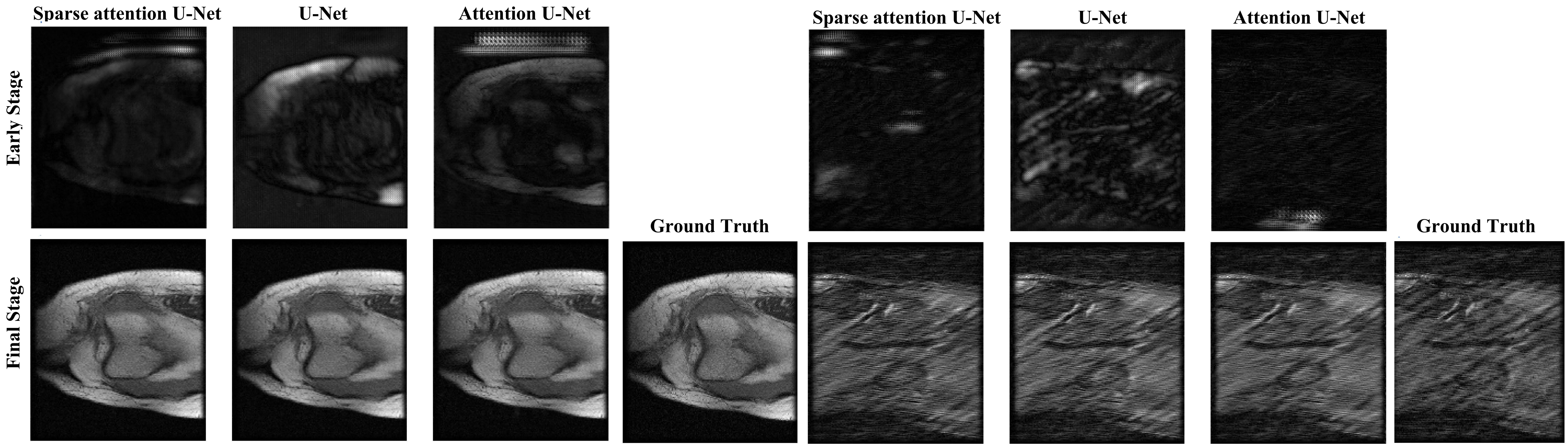
## Conclusion

1. Attention-based models exhibit faster initial loss reduction compared to the default network.

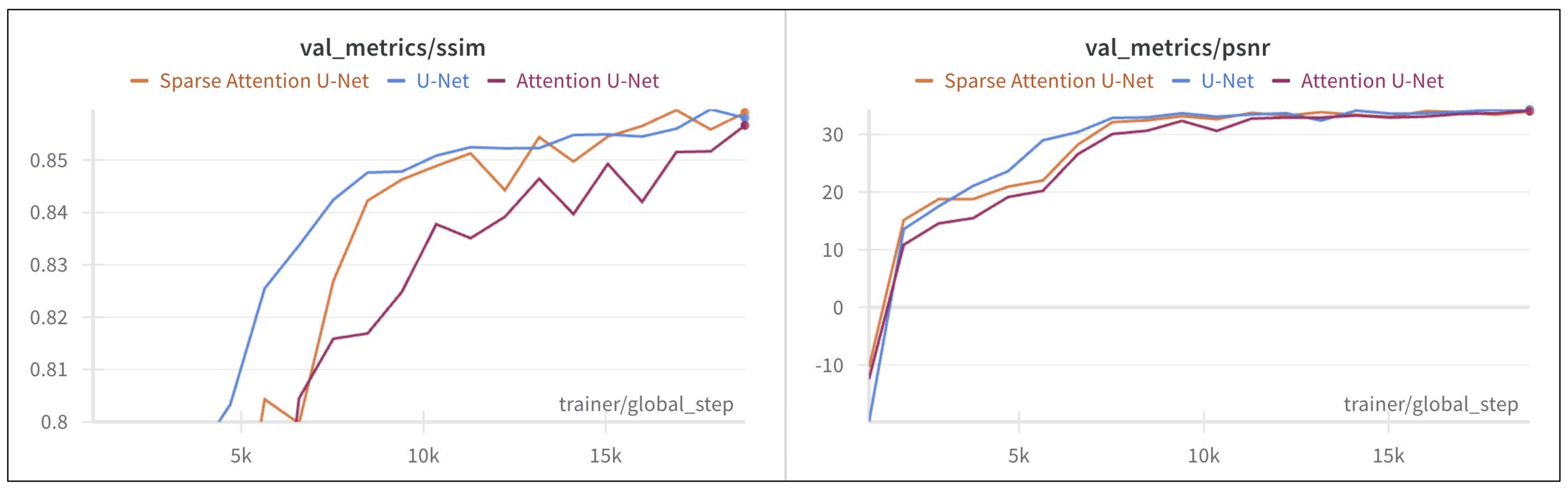
2. Sparse attention UNet performs similarly to the original UNet and better than non-sparse attention UNet in SSIM.

3. Attention-based models quickly capture image details and contours in the initial training stages.

## Supplementy Results



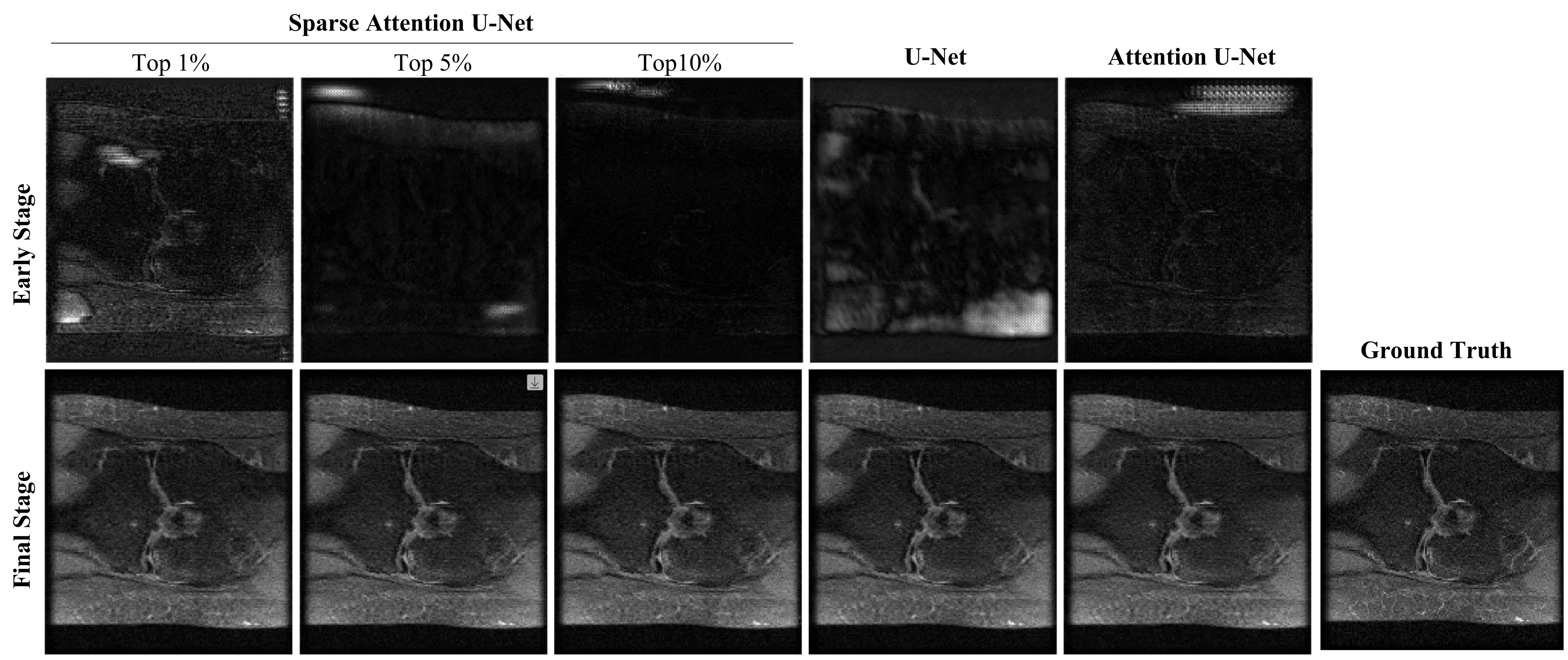

**Supplementary Figure 1.** The reconstruction comparison using different model



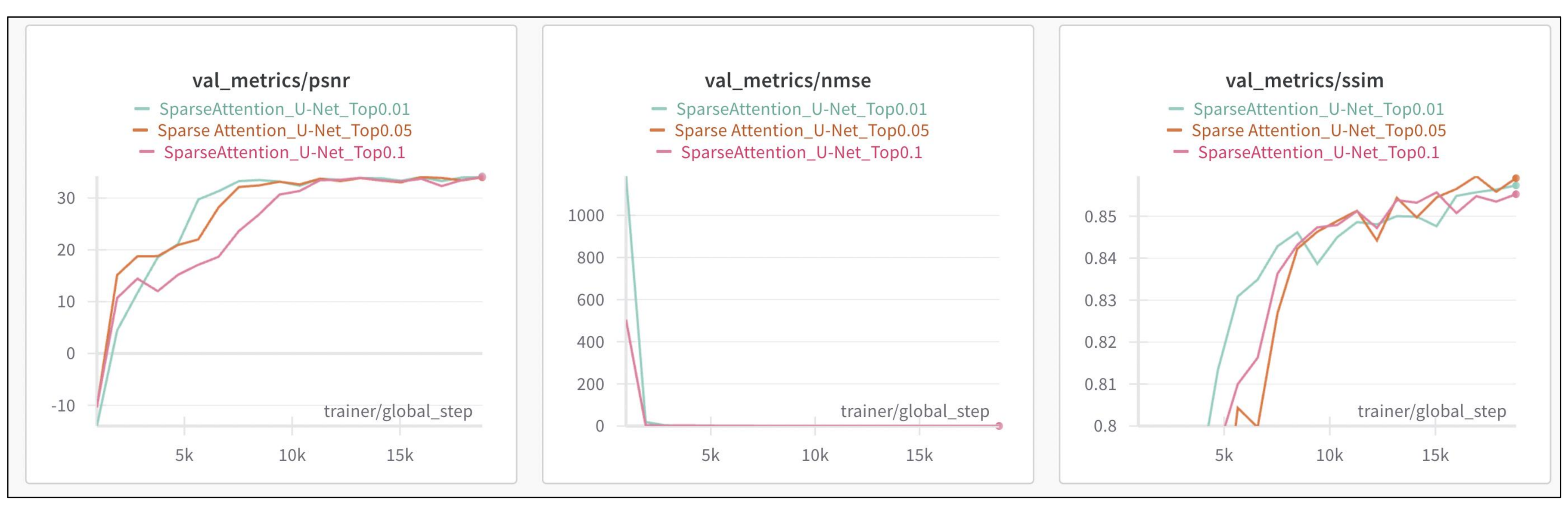

**Supplementary Figure 2.** The SSIM and PSNR metrics curve on validation data

**Supplementary Table 1.** The hyperparameter setting

| Hyperparamter | Value |
|---|---|
| Learning Rate | 0.001 |
| Epoch | 20 |
| Accelerated factor | 2 |
| Mask_type | Random |
| Center_fractions | 0.08 |
| Batch size | 1 |




**Supplementary Figure 3.** The reconstruction comparison, the top k refers to retain top x% of sparse attention




**Supplementary Figure 4.** The PSNR, NMSE and SSIM metrics curve on validation data

## References


[1]Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T., & Knoll, F. (2018). Learning a variational network for reconstruction of accelerated MRI data. Magnetic resonance in medicine, 79(6), 3055-3071.

[2] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M.J., Heinrich, M.P., Misawa, K., Mori, K., McDonagh, S.G., Hammerla, N.Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. ArXiv, abs/1804.03999.

[3] Korkmaz, Y., Yurt, M., Dar, S.U.H., ¨Ozbey, M., Cukur, T.: Deep MRI reconstruction with generative vision transformers. In: Machine Learning for Medical Image Reconstruction. pp. 54–64. Springer International Publishing, Cham (2021)