Intelligent Signal Processing
And MultiMedia lab

SAPIENZA
UNIVERSITÀ DI ROMA

Lightweight Conditional Swin U-Net for Medical Image reconstruction and segmentation

Studente
Luigi Antonelli - 1851425

Relatore
Prof. Danilo Comminiello

Engineering in Computer Science
Facoltà di Ingegneria dell'informazione, informatica e statistica

Anno accademico 2022/2023
Sapienza Università di Roma

1. Medical Image reconstruction and segmentation

Medical Image reconstruction

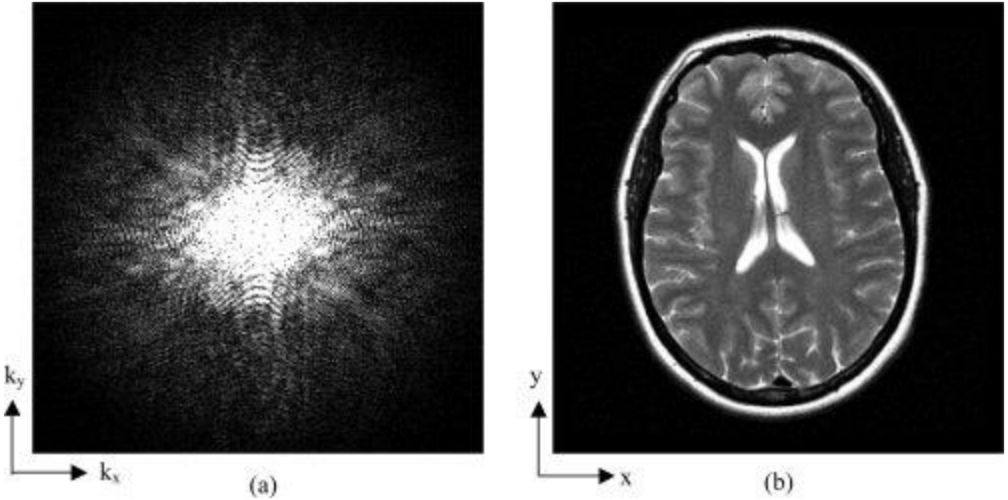
Magnetic Resonance Imaging (MRI) is one of the most used imaging techniques in radiology.

1. MR scanners apply a magnetic field to the human body
2. Hydrogen's protons align with the magnetic field
3. Radio frequency pulses make the protons rotate from this alignment
4. Protons return to the original alignment (relaxation)

With this process the **k-space** data is collected.

... and how Deep Learning can help improve it

The image is obtained from the k-space with the IFFT algorithm.



MRI acquisitions are long and noisy, so they can make some patients uncomfortable.

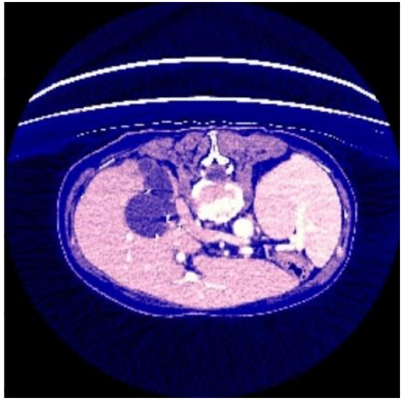
Deep Learning techniques can help reconstruct sharp and detailed images from partially sampled k-spaces.

Deep Learning for Medical Image segmentation

Computed tomography scan (CT scan) uses X-ray to produce detailed images of the human body.

U-shaped neural networks are able to identify the spatial location of tumours from CT scans.

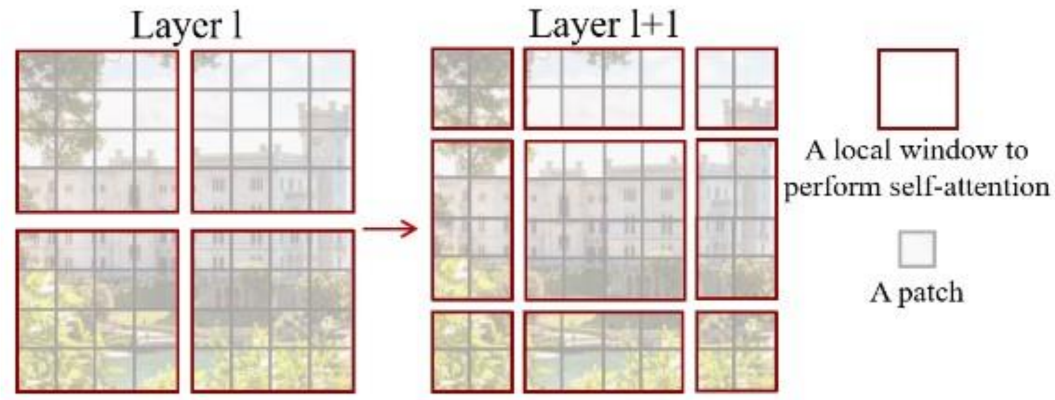
They can be a powerful tool to support Doctors and other healthcare professionals.



2. Why do we need a new approach?

Existing architecture: Swin Transformer (SOTA)

Swin Transformers were able to improve the cost of Multi-Head self-attention by introducing a window partitioning mechanism:



Cost of the standard Multi-Head self-attention layer: $\mathcal{O}(dn^2)$

Cost of the Window Multi-Head self-attention layer:

$$\mathcal{O}(W^4 \cdot d \cdot n_W) = \mathcal{O}(W^4 \cdot d \cdot \frac{hw}{W^2}) = \mathcal{O}(W^2 \cdot d \cdot hw) = \mathcal{O}(W^2 \cdot d \cdot n)$$

Proposed Deep Learning architectures

Two new U-shaped neural networks that separate the computation into **light** and **heavy**:

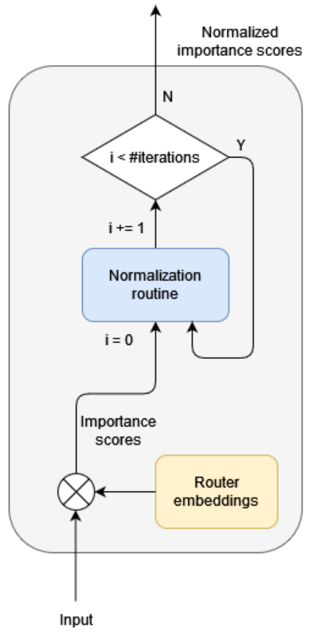
- **Conditional Swin Transformer U-Net with Iterative routers**

- **Conditional Swin Transformer U-Net with Light routers**.

Routers are trainable components that compute importance scores for the patches of the image (and optionally return the most important ones).

Iterative router

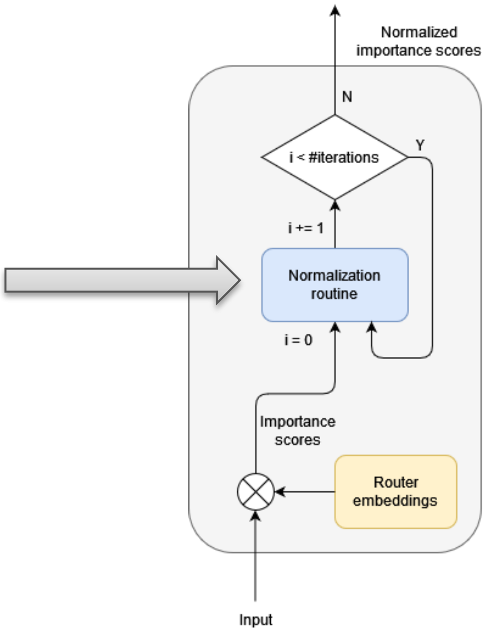
Overview of the Iterative router



Iterative router

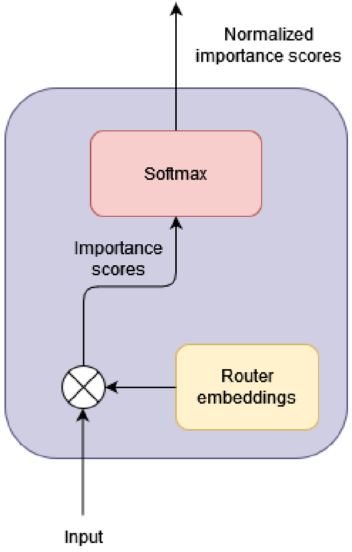
Many hyperparameters:
- Number of iterations
- Epsilon
- Epsilon init
- Epsilon decay

Overview of the Iterative router



Light router

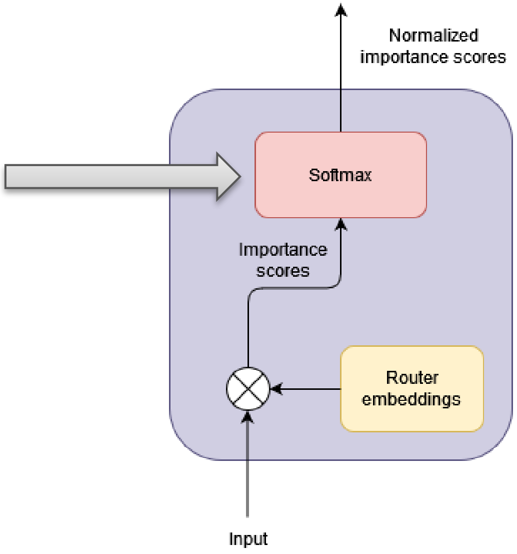
Overview of the Light router



Light router

Overview of the Light router

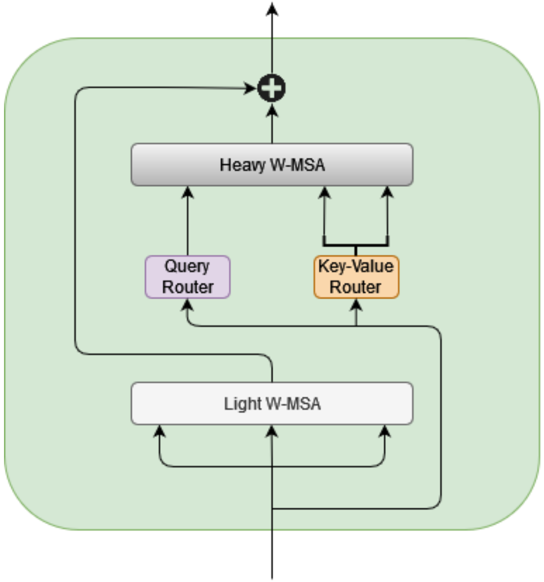
Simple and hyperparameter free mechanism

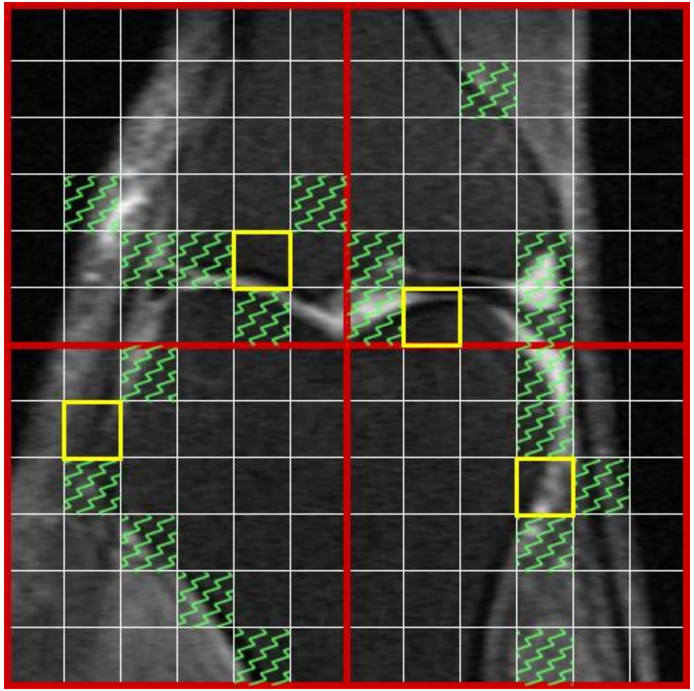


Conditional W-MSA (and SW-MSA)

Overview of the Conditional W-MSA sublayer

$$\text{CondW-MSA}(X_i, s_i^q, \tilde{s}^{kv}) = \text{LightW-MSA}(X_i, X) + s_i^q \,\text{HeavyW-MSA}(X_i, \tilde{s}^{kv}\, X)$$





Conditional W-MSA (and SW-MSA): asymptotic cost

The additional cost introduced by the Conditional Window Multi-Head self-attention sublayer is

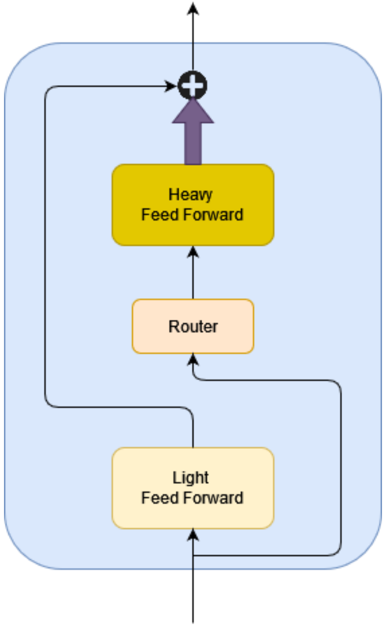
$$\mathcal{O}(W^2 \cdot v \cdot d \cdot n_W) = \mathcal{O}(W^2 \cdot v \cdot d \cdot \frac{hw}{W^2}) = \mathcal{O}(v \cdot d \cdot hw) \in \mathcal{O}(n)$$

$\Longrightarrow$ the asymptotic cost of the layer remains the same.

In practice, the inference time of the two layers is comparable, especially for the Conditional Swin Transformer U-Net with Light routers.

Conditional Feed Forward

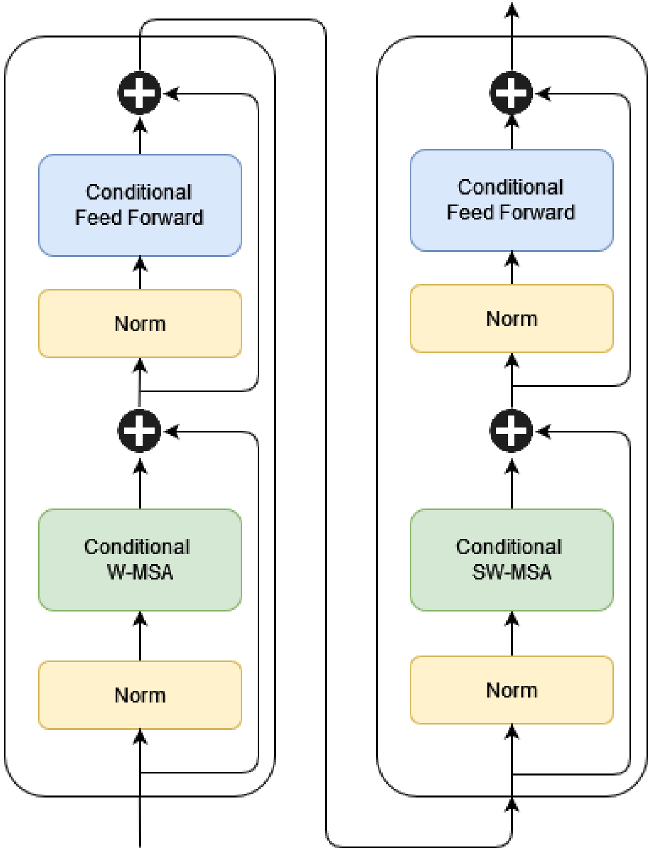
Overview of the Conditional Feed Forward sublayer



$$\mathrm{ConditionalFF}(X_i, \tilde{s}_i) = \mathrm{LightFF}(X_i) + \tilde{s}_i\,\mathrm{HeavyFF}(X_i)$$

Proposed Deep Learning architectures

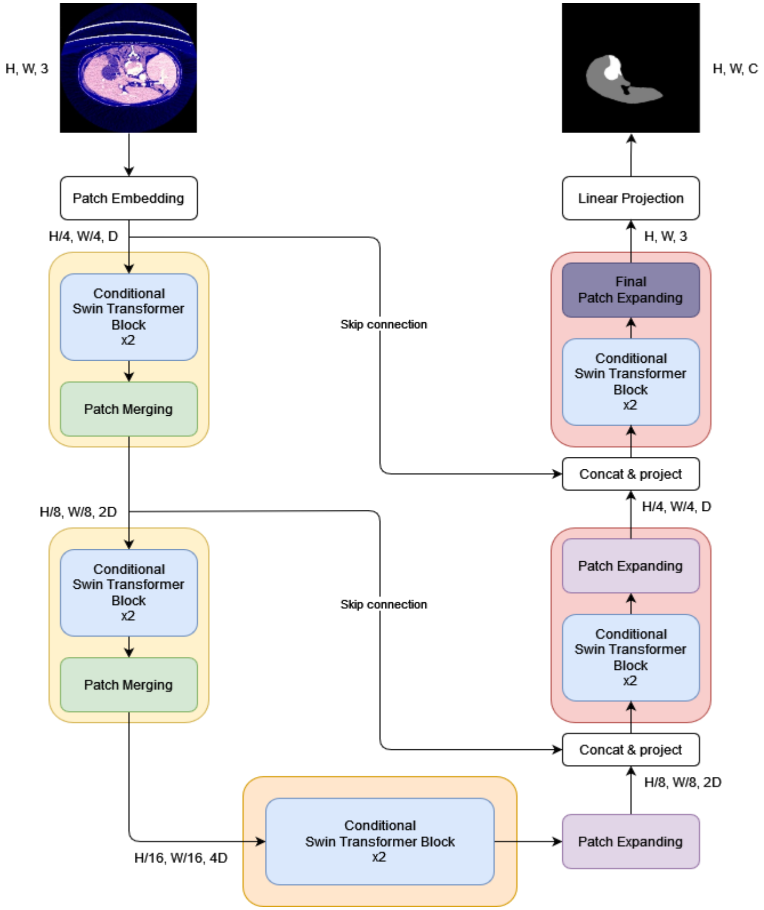
Two consecutive
Conditional Swin Transformer blocks
(both for Iterative and Light routers)



Proposed Deep Learning architectures

Conditional Swin Transformer U-Net
(both for Iterative and Light routers)

with 2 basic layers in the
encoder and in the decoder



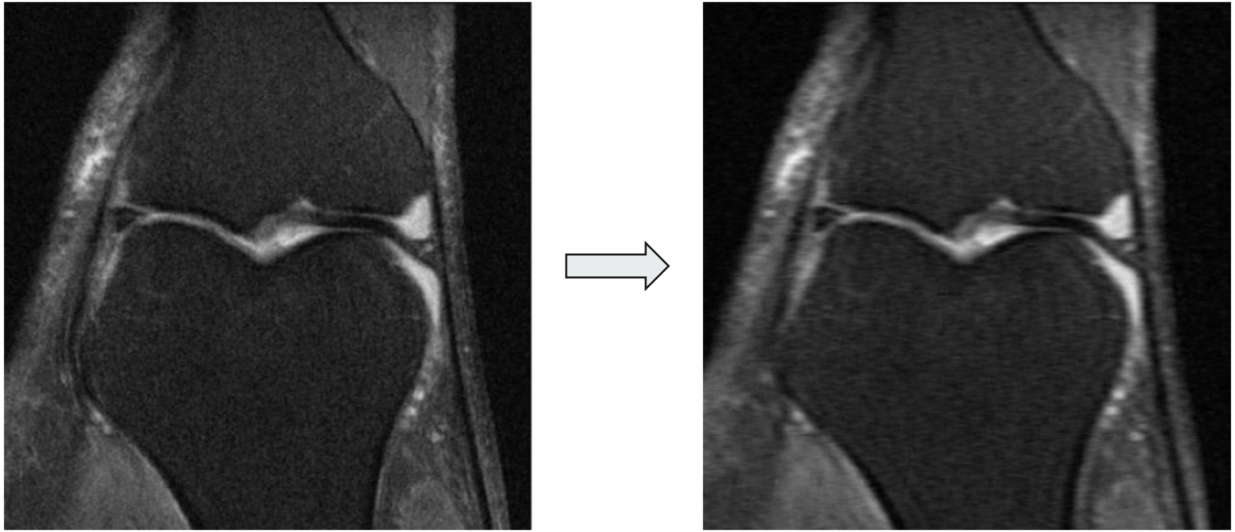
3. Datasets

Datasets

The two Conditional Swin U-Nets and the baseline were trained with the

- **fastMRI dataset** (knee single-coil data) for the image reconstruction task:

Label from the fastMRI dataset (**left**) and the sample for the U-Net (**right**) obtained from the masked k-space of the label

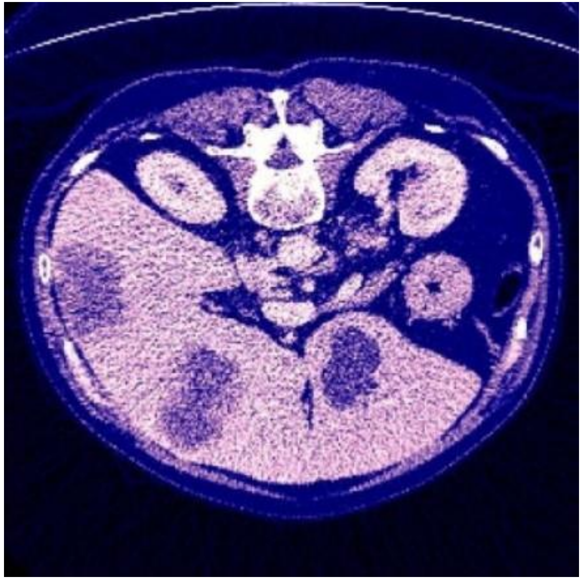


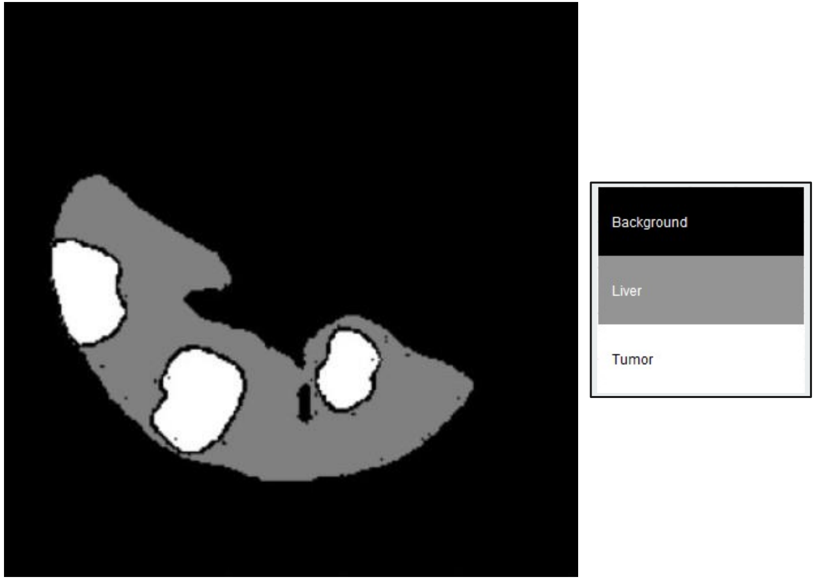
Datasets

The two Conditional Swin U-Nets and the baseline were trained with the

- **Liver Tumor Segmentation dataset** (pt.1 and pt. 2) for the image segmentation task

A <sample, label> pair from the dataset.
The sample is on the **left** and the label on the **right**.





4. Experiments and results

Experiments and results

Training configurations:

Basic layers	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
1	1.9 millions	3.9 millions	3.9 millions
2	6.5 millions	16.3 millions	16.3 millions

- fastMRI: 20/26 epochs, MSE loss
- Liver Tumor Segmentation: 20/30 epochs, Cross Entropy loss

- Adam optimizer with lr = 10^{-4}
- StepLR scheduler with step_size = 5 / 8 epochs and multiplicative factor γ = 0.8

- Dropout for regularization (p = 0.2)

Experiments and results

fastMRI metrics:

- Mean absolute error (MAE)
- Normalized mean squared error (NMSE)
- Peak signal-to-noise ratio (PSNR)
- Structural similarity index (SSIM)

fastMRI results

Basic layers	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
1	1.9 millions	3.9 millions	3.9 millions
2	6.5 millions	16.3 millions	16.3 millions

1 basic layer, smaller dataset

Metric	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
MAE	0.04576	**0.02089**	0.03347
NMSE	0.1328	**0.01228**	0.07208
PSNR	18.7	**29.055**	21.357
SSIM	0.4222	**0.8048**	0.4915

2 basic layers, full dataset

Metric	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
MAE	0.4021	0.5106	**0.02908**
NMSE	0.09043	0.1515	**0.02979**
PSNR	21.84	19.575	**26.798**
SSIM	0.4952	0.4257	**0.7102**

fastMRI results

Performance comparison on a sample

Sample

Target

fastMRI results

Performance comparison on a sample

Swin U-Net

C. Swin U-Net Iterative routers

C. Swin U-Net Light routers

Liver Tumor Segmentation results

Liver Tumor Segmentation metric: Dice Score (with ignore_index = background)

Basic layers	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
1	1.9 millions	3.9 millions	3.9 millions
2	6.5 millions	16.3 millions	16.3 millions

1 basic layer

Metric	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
Dice Score	0.80	**0.81**	0.80

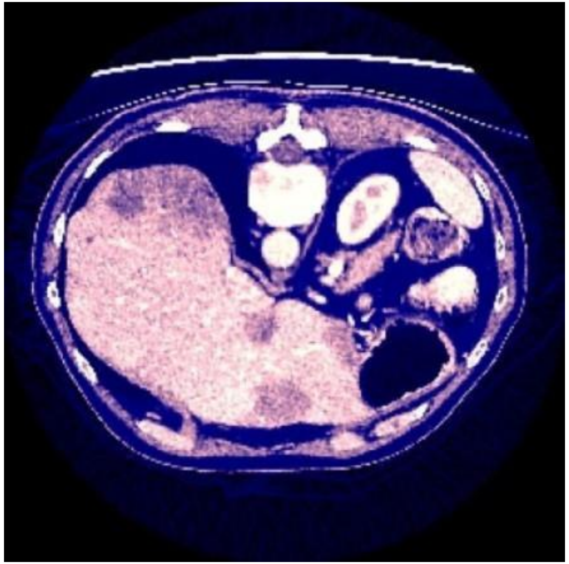
2 basic layers

Metric	Swin U-Net	C. Swin U-Net (I.r.)	C. Swin U-Net (L.r.)
Dice Score	0.92	0.91	**0.93**

Liver Tumor Segmentation results

Performance comparison on a sample

Sample



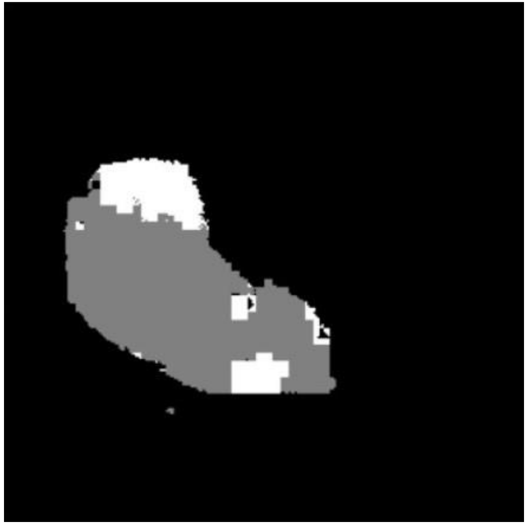
Target

Liver Tumor Segmentation results

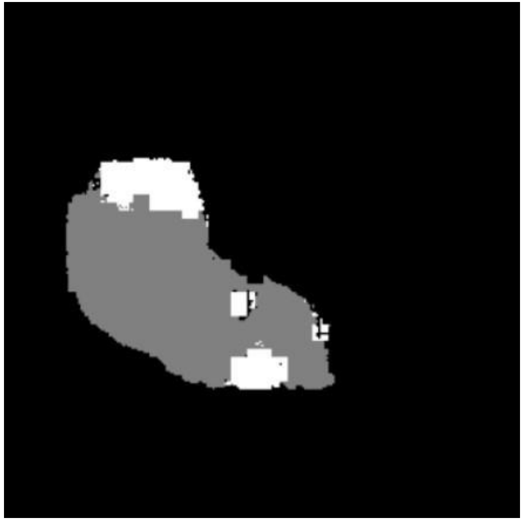
Performance comparison on a sample

Swin U-Net

C. Swin U-Net Iterative routers



C. Swin U-Net Light routers



Conclusions

- Light routers only change how the patches are selected and unlock the potential of the conditional layers

- They provide better performance without introducing any additional hyperparameter and without increasing the inference time.

- Data availability is crucial to exploit the power of Transformer-based U-Nets

- In the future, Light routers can be used in combination with newer sparse attention mechanisms and with self-supervised learning techniques.

Thank you for the attention!

References

1. fastMRI dataset: https://fastmri.med.nyu.edu/
2. Liver Tumor Segmentation dataset: https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation and https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation-part-2
3. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows https://arxiv.org/pdf/2103.14030.pdf
4. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation https://arxiv.org/pdf/2105.05537.pdf
5. Conditional Adapters: Parameter-efficient Transfer Learning with Fast Inference https://arxiv.org/pdf/2304.04947.pdf
6. COLT5: Faster Long-Range Transformers with Conditional Computation https://arxiv.org/pdf/2303.09752.pdf
7. Github repository: https://github.com/luigiantonelli/Lightweight-Conditional-Swin-U-Net-for-Medical-Image-reconstruction-and-segmentation