IEEE ICIP 2025 ANCHORAGE, ALASKA 14-17 SEPTEMBER

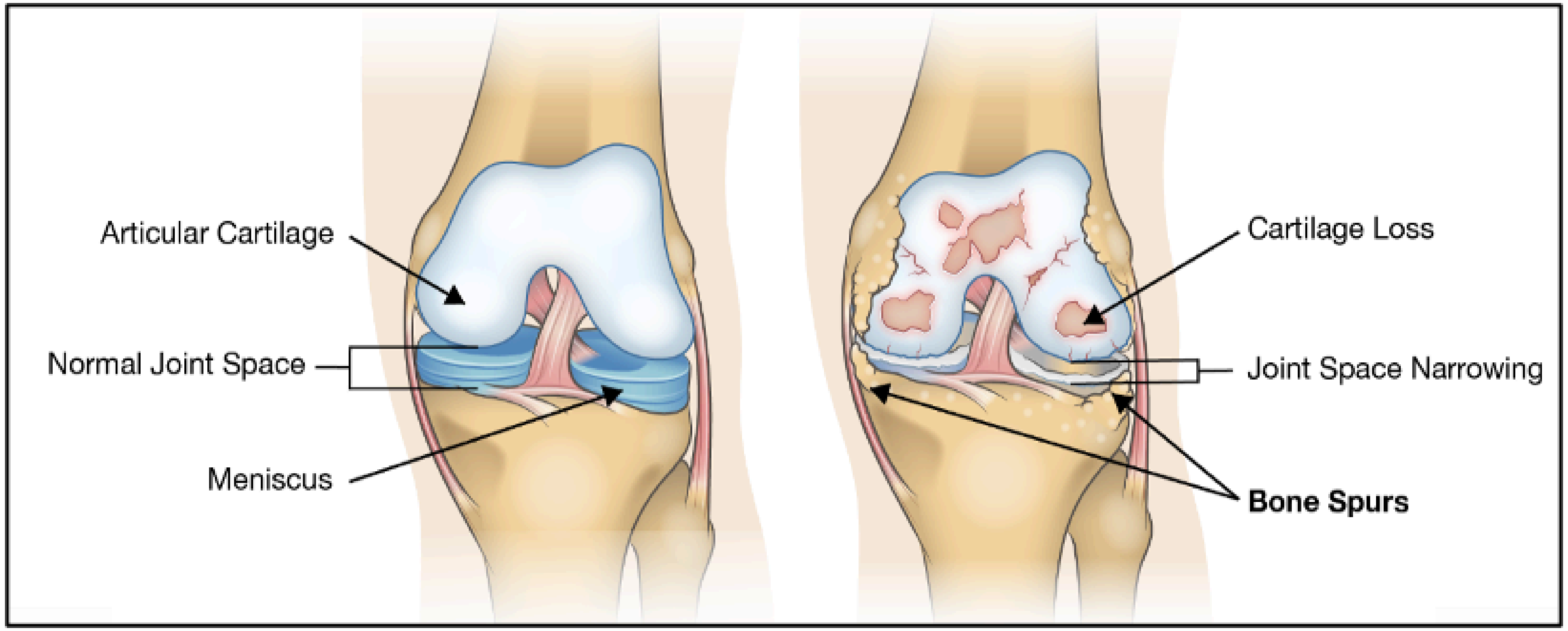
IEEE International Conference on Image Processing

DIFFUSION BASED SHAPE-AWARE LEARNING WITH MULTI-SCALE CONTEXT FOR SEGMENTATION OF TIBIOFEMORAL KNEE JOINT TISSUES: AN END-TO-END APPROACH

PRESENTER NAME: AKSHAY DAYDAR

TABLE OF CONTENTS

INTRODUCTION



WHO DATA

Total Number of people suffering from Musculoskeletal Disorders (MSD): 1.7 B

Total number of people suffering from KOA Worldwide: 343 M

Total Number of people suffering from KOA in India: 47 M

The KOA is preceded by only low back and neck pain amongst MSD category

CHALLENGES

01 Irregularity of pathological structures.

02 Uncertainties in delineating both inter- and intra-cartilage boundaries.

LIMITATIONS OF PREVIOUS WORKS

01 Inconsistencies in the capturing multi-tissue context [4,7]

02 Offline and cumber-some implementations of post-processing stage or segmentation refinement stage [1-3,5]

CONTRIBUTIONS

Multi-Scale Attentive-Unet (MiSA-Unet) model

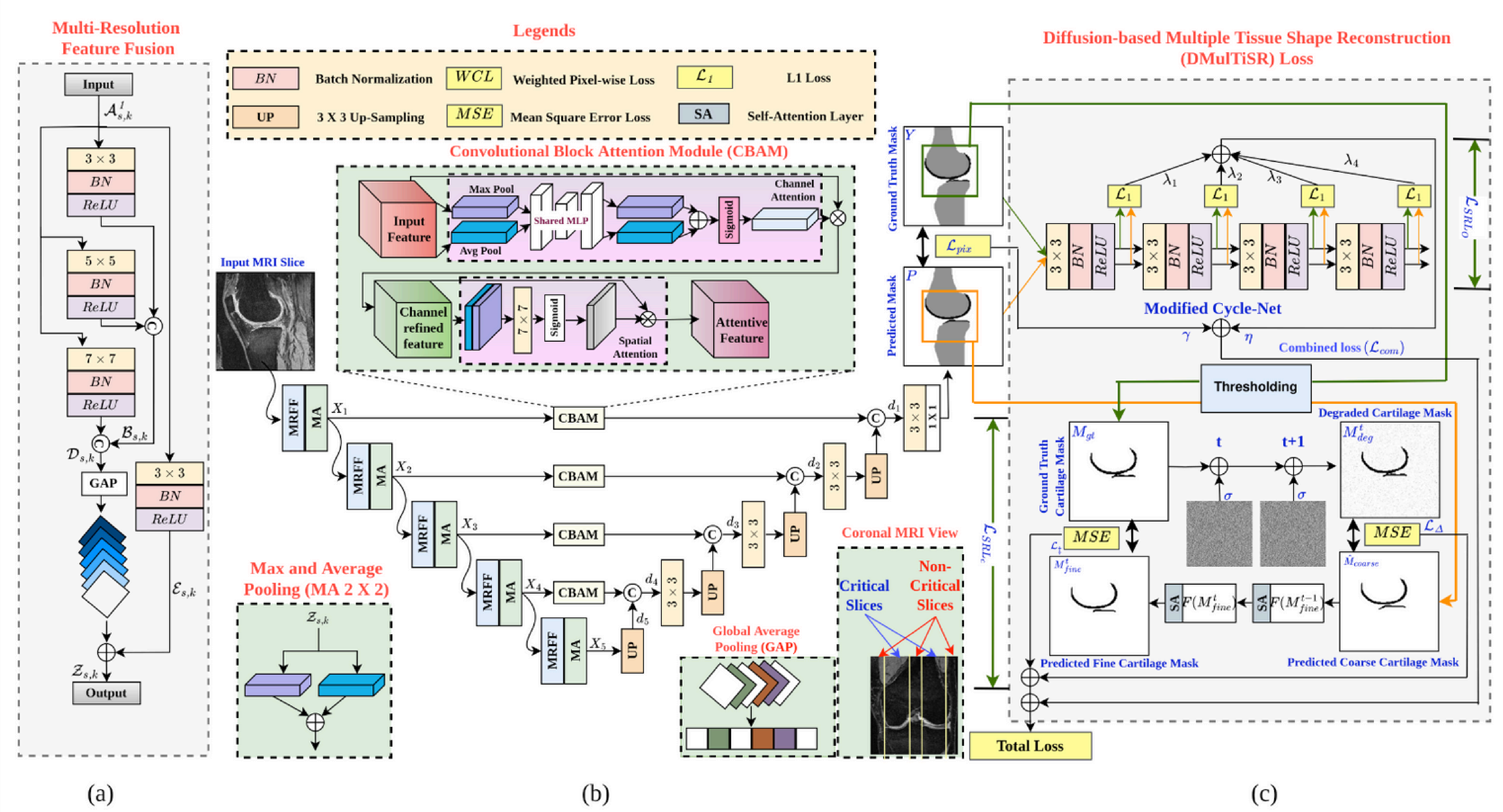


Fig: Schematic of (b) proposed MiSA-Unet model with (a) SAFE module and (c)DMultiSR loss function

Scale-aware Attentive Feature Enhancement module (SAFE)

To focus on multilevel spatial and channel context for accounting relevant local and global

Diffusion-based Multiple Tissue Shape Reconstruction (DMulTiSR) loss

To address structural inaccuracies in the tibiofemoral bones and, more specifically, the cartilages

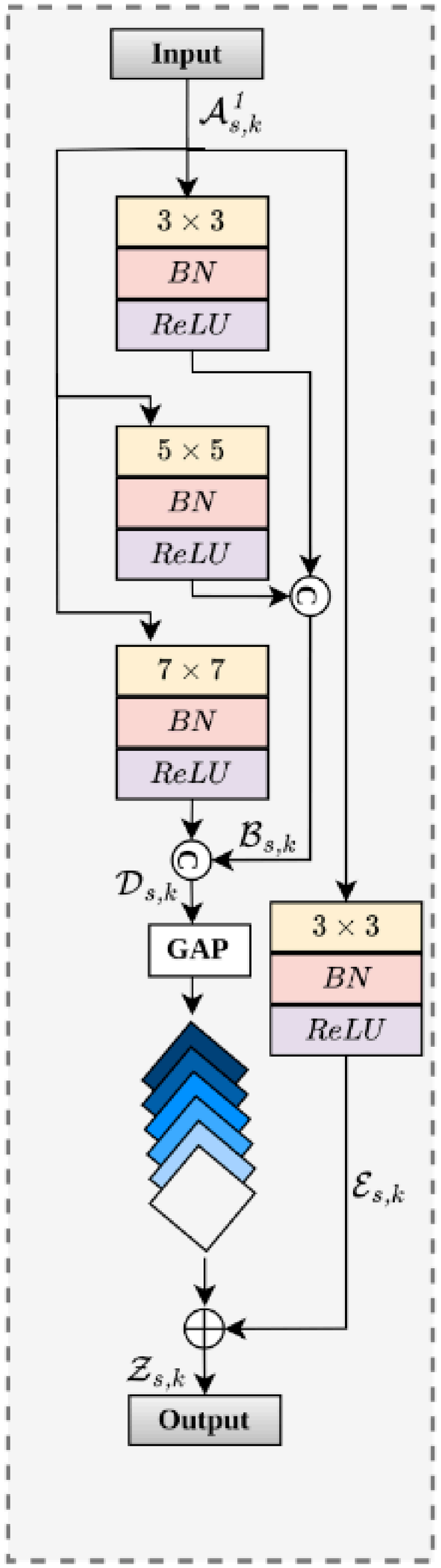
PROPOSED SAFE MODULE

The SAFE module is inspired by inception module, but includes a qualitative improvements to effectively capture task-dependent global and local attention.

$$\mathcal{D}_{s,k} = \mathrm{ReLU}\left(W_p\left[\mathrm{concat}\left(\mathcal{T}_1^{s,k}, \mathcal{T}_2^{s,k}, \mathcal{T}_3^{s,k}\right)\right] + b_p\right) \quad (1)$$

$$\mathcal{Z}_{s,k} = \mathcal{D}_{s,k} + \mathcal{E}_{s,k} \quad (2)$$

Multi-Resolution Feature Fusion

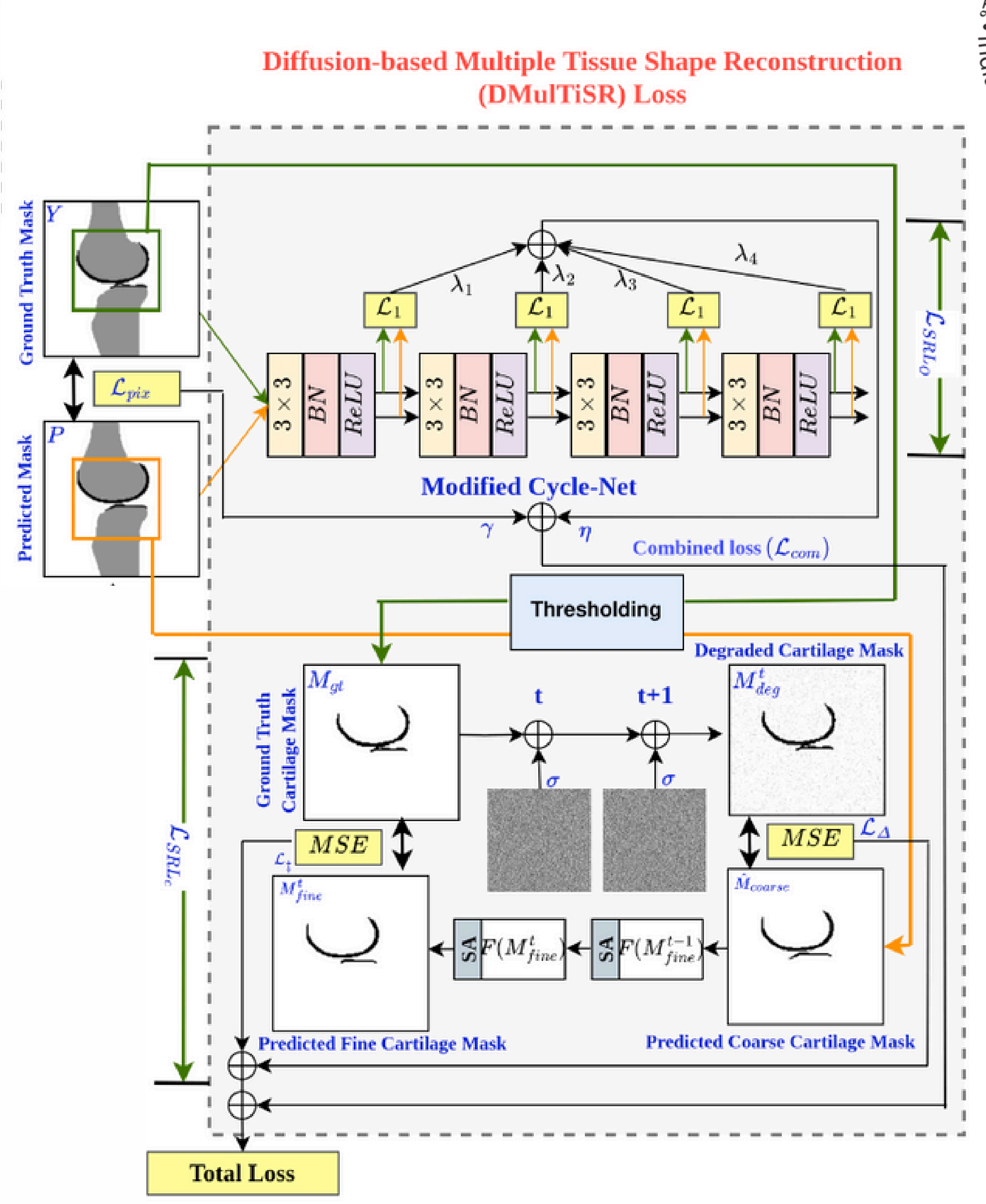


LOSS FUNCTIONS

Pixel-wise loss function

$$\mathcal{L}_{\text{pix},j} = \sum_{j=1}^{m} \beta_j \left[-Y_j \log(P_j) + \frac{2(P_j \cap Y_j)}{|P_j| + |Y_j|} \right], \quad j = 1, 2, \ldots, m \tag{3}$$

$$\mathcal{L}_{\text{pix}} = WDL + WCL \tag{4}$$



Proposed DMultiSR loss function

01 Overall Shape Reconstruction loss

02 Diffusion-based Cartilage Shape Reconstruction loss

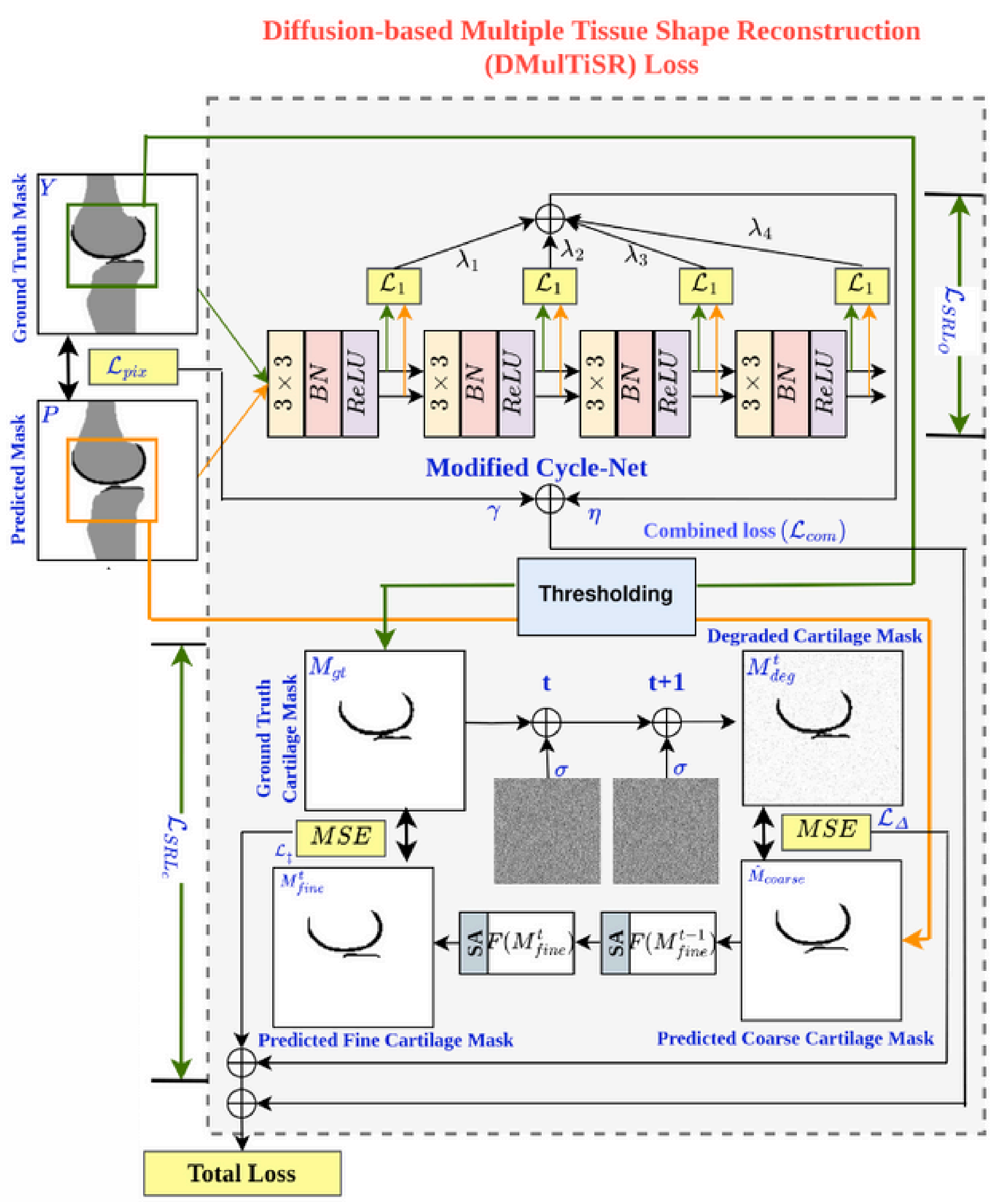
$$\mathcal{L}_{SRL_o} = \lambda \sum_{i=1}^{n} \|P_i - Y_i\|_1$$

$$\mathcal{L}_{SRL_c} = \mathcal{L}_{\ddagger} + \mathcal{L}_{\nabla}$$

LOSS FUNCTIONS

Proposed DMultiSR loss function

Loss inspired by CycleNet [17] and SegRefiner [18] model, but architecturally modified to consider the shape information of multiple tissues and with focus on tibiofemoral cartilage segmentation.



$$\mathcal{L}_{\ddagger} = \frac{1}{N_{\text{steps}}} \sum_{t=1}^{N_{\text{steps}}} \|M_{\text{deg}}^{t} - \hat{M}_{\text{coarse}}\|_2^2 \tag{7}$$

$$\mathcal{L}_{\triangledown} = \frac{1}{N_{\text{steps}}} \sum_{t=1}^{N_{\text{steps}}} \|M_{\text{fine}}^{t} - M_{\text{gt}}\|_2^2 \tag{8}$$

$$M_{\text{deg}}^{t} = \begin{cases} M_{\text{deg}}^{t-1} + \sigma\epsilon, & \text{if}\zeta > \frac{t}{N_{\text{steps}}} \\ \hat{\mathbf{M}}_{\text{coarse}}, & \text{otherwise} \end{cases} \tag{9}$$

$$\mathcal{L}_{\text{total}} = \gamma\mathcal{L}_{\text{pix}} + \eta\mathcal{L}_{SRL_o} + \mathcal{L}_{SRL_c} = \mathcal{L}_{\text{com}} + \mathcal{L}_{SRL_c}$$

EXPERIMENTAL SETUP

Dataset Details

Dataset Size = 512 segmentation maps for each MRI constituting of 160 slices
MRI Sequence = 3D Double Echo Steady-State (DESS)

Experimental Setup

GPU configurations = NVIDIA A100 80 GB GPU

Epochs = 100, MRI slice size = 150*150, batch size = 150, and learning rate = 0.03, Optimizer = Adam

β = [0.01,0.1,0.27,0.12,0.5], λ = [0.1,0.2,0.3,0.4], γ = 0.7, η = 0.3, m = 4 and N_{steps} = 2

EXPERIMENTAL RESULTS

Table: Segmentation SOTA comparison with proposed model

Architecture	Metrics	FC	TC	FB	TB
Knee MRI Segmentation					
2D and 3D CNN + SSM [2]	DSC (%) ↑	89.9	85.6	98.5	98.5
	VOE (%) ↓	18.1	24.9	2.8	2.9
	HD (mm) ↓	5.35	6.35	2.93	3.16
*Modified cGAN [3]	DSC (%)	89.5	83.9	98.5	98.5
	VOE (%)	18.92	27.55	–	–
2D-3D ensemble Unet [4]	DSC (%)	90.3	86.5	98.6	98.8
	VOE (%)	17.5	23.6	2.8	2.4
*Modified Unet++ [1]	DSC (%)	90.9	85.8	99.1	98.2
nnUnet + Entropy Distance Maps [5]	DSC (%)	89.8	86.4	98.6	98.6
	HD (mm)	5.22	4.70	11.82	5.30
Unet-S [7]	DSC (%)	89.7	89.8	98.7	98.7
	HD(mm)	5.58	4.74	4.05	3.82
*Modified Source-free UDA [8]	DSC (%)	74.7	59.4	93.7	94.7
Other Network Architectures					
Unet [13]	DSC (%)	88.6	87.0	98.3	98.3
	VOE (%)	20.06	22.41	3.34	3.29
	HD (mm)	6.69	5.23	6.12	4.05
Attention Unet [20]	DSC (%)	88.7	87.1	98.3	98.2
	VOE (%)	19.62	22.16	3.33	3.24
	HD (mm)	6.88	5.56	6.00	6.46
HRnet [21]	DSC (%)	88.9	86.5	98.2	98.2
	VOE (%)	18.67	22.11	3.19	3.78
	HD (mm)	6.28	5.94	7.10	6.99
SAMed [22]	DSC (%)	89.0	87.1	98.6	98.5
	VOE (%)	17.89	22.89	2.12	2.90
	HD (mm)	5.28	3.94	5.90	3.64
Proposed MiRA-Unet (Critical slices only)	DSC (%)	89.8	88.0	98.5	98.5
	VOE (%)	18.76	20.94	2.76	3.08
	HD (mm)	6.41	4.95	5.47	3.89
Proposed MiRA-Unet$^{\ominus}$ (All slices)	DSC (%)	90.4	90.1	98.7	98.6
	VOE (%)	17.22	18.97	4.09	2.9
	HD (mm)	4.74	3.11	2.54	4.32

• The best and second best results are denoted in red and blue colors, respectively. The * indicates the architectures specifically utilized for the knee MRI segmentation task, and The ⊖ indicates the model's testing on both critical and non-critical slices with the post-processing stage (similar to Deng et al.[1]).

- For Critical MRI slices; excellent results in FC and TC which are nearly 4.5% higher for DSC than modified cGAN [3]
- Slightly lower results for FC (about 1% in DSC than Deng et al. [1]) possibly due to poor delineation of bone cartilage (FB-FC) interface and greater shape variability and discontinuous nature of cartilage in critical MRI slices.
- For all MRI slices, average minimum improvement in DSC, VOE, and HD are 0.24%, 9.85%, and 17.31% respectively.

EXPERIMENTAL RESULTS

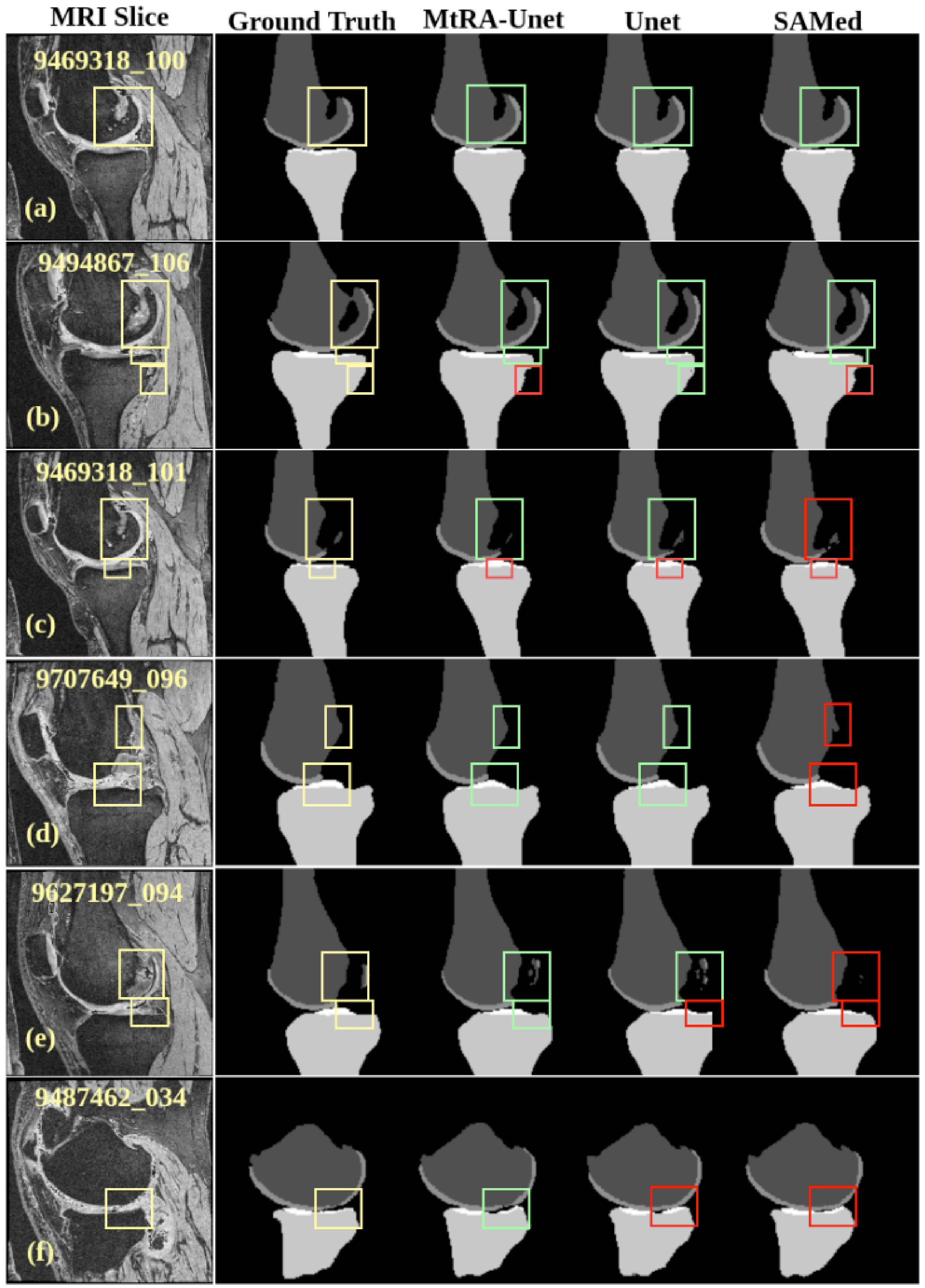


Fig: Segmentation SOTA comparison with proposed model

- Excellent results for femur and tibia in all cases as indicated in Figure 2 (a to f), even in the presence of soft-tissue inflammation (see Figure 2 a, b and c).
- Cartilage performance is improved, specifically at the cartilage-cartilage interface as indicated in Figure 2(d,e,f).
- Failure in some cases in capturing the shape of the tibial bone and cartilage.

ABLATION STUDY

Table: Ablation study with proposed model

Architecture	Loss Function	Metrics	FC	TC	FB	TB
baseline† + SAFE1	WDL	DSC (%)	89.5	87.4	98.4	98.4
		VOE (%)	18.62	21.84	3.07	3.16
		HD (mm)	6.82	5.05	5.48	3.83
	WDL + WCL ($\mathcal{L}_{pix}$)	DSC (%)	89.5	86.4	98.5	98.4
		VOE (%)	18.56	22.64	3.06	3.23
		HD (mm)	6.36	5.19	5.83	4.00
	$\gamma\mathcal{L}_{pix} + \eta\mathcal{L}_{SRL_o}$ ($\mathcal{L}_{com}$)	DSC (%)	89.2	87.2	98.5	98.4
		VOE (%)	19.07	22.18	2.98	3.05
		HD (mm)	6.94	5.11	6.8	3.95
	$\mathcal{L}_{com} + \mathcal{L}_{SRL_c}$ ($\mathcal{L}_{total}$)	DSC (%)	89.7	87.6	98.5	98.5
		VOE (%)	18.23	21.54	2.94	2.97
		HD (mm)	6.87	4.13	6.03	4.45
baseline† + SAFE2	WDL	DSC (%)	89.6	87.5	98.4	98.4
		VOE (%)	18.39	21.73	3.09	3.18
		HD (mm)	6.42	5.25	5.44	4.00
	$\mathcal{L}_{pix}$	DSC (%)	89.2	87.2	98.3	98.4
		VOE (%)	19.01	22.10	3.24	3.07
		HD (mm)	6.75	5.39	5.93	3.54
	$\mathcal{L}_{com}$	DSC (%)	89.7	87.6	98.5	98.5
		VOE (%)	18.20	20.54	3.94	3.17
		HD (mm)	6.53	5.26	5.61	3.82
Proposed MiRA-Unet model	$\mathcal{L}_{total}$	DSC (%)	89.8	88.0	98.5	98.5
		VOE (%)	18.76	20.94	2.96	3.08
		HD (mm)	6.41	4.95	5.47	3.89
baseline† + SAFE3	WDL	DSC (%)	87.1	85.1	96.7	89.5
		VOE (%)	22.41	25.11	3.31	17.82
		HD (mm)	7.03	5.8	7.88	8.22
	$\mathcal{L}_{pix}$	DSC (%)	89.2	87.2	98.4	98.4
		VOE (%)	19.09	22.25	3.12	3.10
		HD (mm)	6.18	5.3	6.13	4.02
	$\mathcal{L}_{com}$	DSC (%)	89.4	86.7	98.5	98.4
		VOE (%)	18.81	22.87	2.84	3.03
		HD (mm)	6.53	5.78	5.72	3.55
	$\mathcal{L}_{total}$	DSC (%)	89.3	87.0	98.5	98.5
		VOE (%)	18.99	22.43	2.97	3.05
		HD (mm)	6.57	4.94	6.11	4.15

- The best and second best results are denoted in red and blue colors, respectively.

- The combined loss function resulted in a minimum improvement of 0.5% in DSC and 4.58% in HD than the pixel-wise loss functions for all SAFE combinations.
- Tibiofemoral cartilages is improved by adding the loss $L_{SRL,c}$ with a combined loss function of an average of 1.68% in VOE and 4.72% HD.
- The TC is observed with a maximum improvement of 0.56% in DSC (for SAFE1).

CONCLUSION

01 Proposed MiSA-Unet is an end-to-end and single-stage segmentation network unlike previous studies.

02 Proposed model improved average DSC by 2.33% (on critical slices) while with post-processing it improved minimum DSC by 0.24%, VOE by 9.85%, and HD by 17.31% (on all slices) over SOTA.

03 In future, an effort will be made to eliminate the postprocessing stage and analyze the segmentation performance for each KOA grade

THANK YOU !!

DIFFUSION BASED SHAPE-AWARE LEARNING WITH MULTI-SCALE CONTEXT FOR SEGMENTATION OF TIBIOFEMORAL KNEE JOINT TISSUES: AN END-TO-END APPROACH

AUTHORS: AKSHAY DAYDAR, ALIK PRAMANICK, ARIJIT SUR AND SUBRAMANI KANAGARAJ

SPECIAL THANKS TO IIT GUWAHATI'S TIDF FOR PROVIDING HIGH-END COMPUTATIONAL FACILITIES

REFERENCES

[1] Yang Deng, Lei You, Yanfei Wang, and Xiaobo Zhou, “A coarse-to-fine framework for automated knee bone and cartilage segmentation data from the osteoarthritis initiative,” Journal of Digital Imaging, vol. 34, no. 4, pp. 833–840, 2021.

[2] Felix Ambellan, Alexander Tack, Moritz Ehlke, and Stefan Zachow, “Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative,” Medical image analysis, vol. 52, pp. 109–118, 2019.

[3] Dimitri A Kessler, James W MacKay, Victoria A Crowe, Frances MD Henson, Martin J Graves, Fiona J Gilbert, and Joshua D Kaggie, “The optimisation of deep neural networks for segmenting multiple knee joint tissues from mris,” Computerized Medical Imaging and Graphics, vol. 86, pp. 101793, 2020.

[4] Muhamad Hafiz Abd Latif and Ibrahima Faye, “Automated tibiofemoral joint segmentation based on deeply supervised 2d-3d ensemble u-net: Data from the osteoarthritis initiative,” Artificial intelligence in medicine, vol. 122, pp. 102213, 2021.

[5] Zezhong Li, Kangming Chen, Peng Liu, Xiaodong Chen, and Guoyan Zheng, “Entropy and distance maps-guided segmentation of articular cartilage: data from the osteoarthritis initiative,” International Journal of Computer Assisted Radiology and Surgery, vol. 17, no. 3, pp. 553–560, 2022.

[7] Boyeong Woo, Craig Engstrom, William Baresic, Jurgen Fripp, Stuart Crozier, and Shekhar S Chandra, “Automated anomaly-aware 3d segmentation of bones and cartilages in knee mr images from the osteoarthritis initiative,” Medical Image Analysis, vol. 93, pp. 103089, 2024.

[8] Siyue Li, Shutian Zhao, Yudong Zhang, Jin Hong, and Weitian Chen, “Source-free unsupervised adaptive segmentation for knee joint mri,” Biomedical Signal Processing and Control, vol. 92, pp. 106028, 2024.