



CAS-Net: Cross-View Aligned Segmentation by Graph Representation of Knees

Zixu Zhuang^{1,3}, Xin Wang¹, Sheng Wang^{1,2,3}, Zhenrong Shen¹,
Xiangyu Zhao¹, Mengjun Liu¹, Zhong Xue³, Dinggang Shen^{2,3},
Lichi Zhang¹(✉), and Qian Wang²(✉)

¹ School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
lichizhang@sjtu.edu.cn

² School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
qianwang@shanghaitech.edu.cn

³ Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. Magnetic Resonance Imaging (MRI) has become an essential tool for clinical knee examinations. In clinical practice, knee scans are acquired from multiple views with stacked 2D slices, ensuring diagnosis accuracy while saving scanning time. However, obtaining fine 3D knee segmentation from multi-view 2D scans is challenging, which is yet necessary for morphological analysis. Moreover, radiologists need to annotate the knee segmentation in multiple 2D scans for medical studies, bringing additional labor. In this paper, we propose the Cross-view Aligned Segmentation Network (CAS-Net) to produce 3D knee segmentation from multi-view 2D MRI scans and annotations of sagittal views only. Specifically, a knee graph representation is firstly built in a 3D isotropic space after the super-resolution of multi-view 2D scans. Then, we utilize a graph-based network to segment individual multi-view patches along the knee surface, and piece together these patch segmentations into a complete knee segmentation with help of the knee graph. Experiments conducted on the Osteoarthritis Initiative (OAI) dataset demonstrate the validity of the CAS-Net to generate accurate 3D segmentation.

Keywords: Knee Segmentation · Multi-view MRI · Super Resolution · Graph Representation

1 Introduction

Magnetic resonance imaging (MRI) plays a pivotal role in knee clinical examinations. It provides a non-invasive and accurate way to visualize various internal knee structures and soft tissues [2]. With its excellent contrast and resolution, MRI can help radiologists detect and diagnose knee injuries or diseases. The choice of MRI sequences and acquisition techniques depends on the specific clinical scenario. While the acquisition of a 3D sequence can render the entire knee

Z. Zhuang, X. Wang and S. Wang—These authors contributed equally to this work.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 110–119, 2023.
https://doi.org/10.1007/978-3-031-43901-8_11

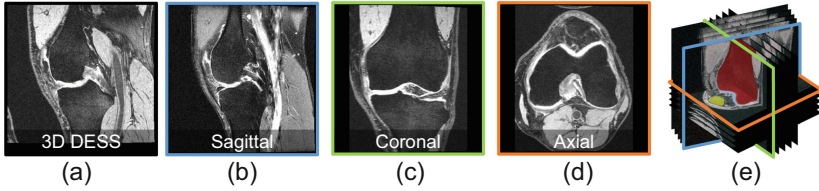


Fig. 1. Example of knee MRI scans in OAI dataset. (a) is a slice of 3D knee volume, which takes much longer scanning time; (b-d) are slices of multi-view 2D knee scans from sagittal, coronal, and axial views; (e) is the multi-view knee MRI scans aligned in the world coordinate system (WCS).

as shown in Fig. 1(a), it is not always feasible in clinical practice where scanner resource is limited [13]. In contrast, multi-view 2D scans can offer higher intra-slice resolution, better tissue contrast, and shorter scanning time [7, 8], as displayed in Fig. 1(b)–(d). Therefore, multi-view 2D scans are widely used for diagnostics in clinical practice.

Although multi-view 2D MRI scanning can provide sufficient image quality for clinical diagnosis in most cases [7, 16], it is challenging for estimating the corresponding 3D knee segmentation, which is essential for the functional and morphological analysis [15, 17]. Previous studies mostly rely on 3D sequences to conduct image segmentation and morphology-based knee analysis [3, 20], which is impractical for clinical data of only 2D scans.

The cross-view consistency of multi-view knee MR scans provides the basis for generating 3D segmentation from 2D scans, as shown in Fig. 1(e). Some studies have already utilized the cross-view consistency for medical image applications. For example, Liu et al. [10] proposed a Transformer-based method to fuse multi-view 2D scans for ventricular segmentation, by learning the cross-view consistency with the self-attention mechanism. Perslev et al. proposed to extract multi-view 2D slices from 3D images for 3D segmentation on knees [12] and brains [11]. Li et al. [9] transferred widely used sagittal segmentation to coronal and axial views based on cross-view consistency. Zhuang et al. [22] introduced a unified knee graph architecture to fuse the multi-view MRIs in a local manner for knee osteoarthritis diagnosis. These studies have demonstrated the value of cross-view consistency in medical analysis.

In this paper, we propose a novel framework, named Cross-view Aligned Segmentation Network (CAS-Net), to generate 3D knee segmentation from clinical multi-view 2D scans. Moreover, following the convention of radiologists, we require the supervising annotation for the sagittal segmentation only. We first align the multi-view knee MRI scans into an isotropic 3D volume by super-resolution. Then we sample multi-view patches and construct a knee graph to cover the whole knee joint. Next, we utilize a graph-based network to derive a fine 3D segmentation. We evaluate our proposed CAS-Net on the Osteoarthritis Initiative (OAI) dataset, demonstrating its effectiveness in cross-view 3D segmentation.

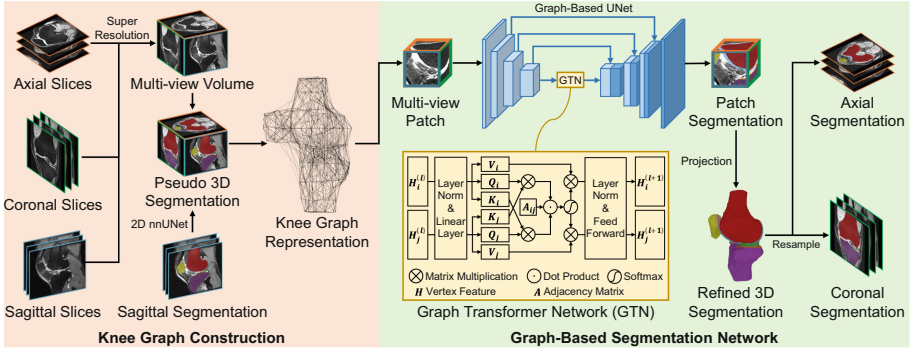


Fig. 2. The overall architecture of the proposed Cross-view Aligned Segmentation Network (CAS-Net). It is composed of two major parts: (1) Knee Graph Construction, which involves super-resolution of the multi-view 2D scans into a 3D volume and representing with a knee graph; (2) Graph-based Segmentation Network, which segments patches on the knee graph, resulting in 3D segmentation.

2 Method

The goal of CAS-Net is to generate 3D knee segmentation with multi-view 2D scans and sagittal segmentation annotation. The overall architecture is shown in Fig. 2, which consists of two major parts.

- 1 **Knee Graph Construction.** We conduct super-resolution on multi-view 2D scans and align them in a 3D volume. Then, we sample multi-view patches along bone-cartilage interfaces and construct a knee graph.
- 2 **Graph-based Segmentation Network.** We deal with the 3D multi-view patches on the knee graph by a 3D UNet [14] combined with Graph Transformer Network (GTN) [4, 21]. We also randomly mask out views of patches to enforce cross-view consistency during network learning.

2.1 Knee Graph Construction

The 2D clinical multi-view scans are acquired from non-orthogonal angles and have large differences in inter- and intra-slice spacing (e.g. 3.5 mm v.s. 0.3 mm in clinical practice). Therefore, we use super-resolution to unify them into one 3D volume with a pseudo segmentation (with tolerable errors), and sample the volume into graph representation.

Isotropic Super-Resolution. We first apply super-resolution [19] to the multi-view 2D scans, which include sagittal, coronal and axial scans, for representing them in the same space and reducing the large inter-slice spacing. The process is shown in the left part of Fig. 2.

Specifically, we follow [19] by projecting multi-view 2D scans into the world coordinate system (WCS) and treating each scan as a continuous function of coordinates. Then, we take advantage of the continuity of coordinates to synthesize an isotropic 3D volume, by querying from the learned function for the coordinates at a equidistantly distributed interval. Since each 2D scan is reconstructed from the same set of WCS coordinates, the multi-view volumes after super-resolution are naturally aligned in the WCS space. Therefore, we concatenate these multi-view volumes as a multi-channel 3D image.

Moreover, we train a 2D nnUNet [6] on the sagittal 2D scan and apply it to the sagittal 3D volume slice-by-slice, to acquire a pseudo-3D segmentation. The pseudo-3D segmentation is the basis of knee graph construction. It also provides supervision for the graph-based segmentation network; yet it contains many errors caused by 2D segmentation, e.g., unsmooth segmentation boundaries of bones and cartilages.

Knee Graph Sampling. Errors in the pseudo-3D segmentation are distributed around bone surfaces. We sample multi-view patches from these areas and refine their segmentations in 3D. Previous studies have demonstrated the usefulness of organizing knee patches with graph-based representation [22]. Therefore, we follow this approach and construct the knee graph \mathbf{G} .

We start extracting points on the bone surface by calculating the bone boundaries on the pseudo-3D segmentation. Then, we uniformly sample from these points until the average distance between them is about 0.8 times the patch size. These sampled points are collected as the vertices \mathbf{V} in the graph. \mathbf{V} stores the center-cropped multi-view patches \mathbf{P} and their WCS positional coordinates \mathbf{C} . The distribution of the sampled points ensures that multi-view patches center-cropped on \mathbf{V} can fully cover the bone surfaces and bone-cartilage interfaces. Finally, we connect the vertices in \mathbf{V} using edges \mathbf{E} , which enables the vertex features to be propagated on the knee graph. The edges are established by connecting each vertex to its k ($k = 10$, following [22]) nearest neighbors in \mathbf{V} .

2.2 Graph-Based Segmentation Network

Graph-Based UNet. The knee graph, which is constructed early, is input to a UNet combined with a 3-layer Graph Transformer Network (GTN) [4, 21] to generate patch segmentation, as shown on the right part of Fig. 2.

To encode the multi-view patches \mathbf{P} as vertex features \mathbf{H} in the bottleneck of the UNet, we first apply average pooling to \mathbf{H} and then add the linearly-projected WCS coordinate \mathbf{C} to enable convolution of \mathbf{H} along the graph. Specifically, we compute $\mathbf{H}^{(0)} = \text{AvgPooling}(\mathbf{H}) + \text{Linear}(\mathbf{C})$. Next, we pass the resulting vertex features $\mathbf{H}^{(0)}$ through a graph convolutional network. Note that the knee graphs have varying numbers of vertices and edges, we utilize the Graph Transformer Networks (GTN) to adapt it.

Similar to Transformers [18], the GTN generates three attention embeddings \mathbf{Q} , \mathbf{K} , and \mathbf{V} from $\mathbf{H}^{(i)}$ by linear projection in the i -th layer. The difference is

that GTN uses the adjacency matrix \mathbf{A} to restrict the self-attention computation to between adjacent vertices instead of all of them: $\alpha = \text{Softmax}(\mathbf{A} \odot \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})$, where α is the self-attention, \odot denotes dot product, and d is the length of attention embeddings. Then, the local feature $\mathbf{H}^{(i)}$ in the i -th layer can be computed as $\mathbf{H}^{(i)} = \sigma(\alpha\mathbf{V})$, where σ is feedforward operation.

After the attention-based graph representation, $\mathbf{H}^{(3)}$ is repeated to match the shape of \mathbf{H} and then summed up, serving as the input to the UNet decoder. And the UNet decoder derives the 3D patch segmentation according to $\mathbf{H}^{(3)}$ and skip connection. We further project the patch segmentations back into the pseudo-3D segmentation, thereby obtaining a rectified 3D knee segmentation.

Random View Masking. To reduce impacts of the errors in the initial pseudo-3D segmentation, we randomly masked out views (channels) in the multi-view (multi-channel) patches during training. The 3D volume, obtained from 2D scans by super-resolution, has a clearer tissue texture in the plane parallel to its original scanning plane. Thus, the random view masking helps to produce more accurate segmentation boundaries by forcing the network to learn from different combinations of patch views (channels). We set the probability of each view to be masked as 0.25. During inference, complete multi-view patches are used, reducing errors caused by the pseudo-3D segmentation.

3 Experimental Results

3.1 Data and Experimental Settings

We have tested the CAS-Net on the Osteoarthritis Initiative (OAI) dataset [5], which contains multi-view 2D scans and 3D volumes aligned in the same WCS space. The 3D DESS modality in the OAI dataset has segmentation annotation [1], which is used as the ground truth to evaluate the 3D segmentation performance in our work. The SAG_IW_TSE_LEFT, COR_MPR_LEFT, and AX_MPR_LEFT modalities in the OAI dataset are selected as multi-view 2D scans. For the 3D DESS is unsmooth when it is resampled to 2D scans, we recruit a radiologist to manually annotate the femur, tibia, their attached cartilages, and meniscus in 2D scans. The sagittal annotation is used in our method, and the coronal or axial annotation is set as the ground truth for evaluation. We extract 350 subjects for our experiments, with 210 for training, 70 for validation, and 70 for testing. We report the results of the testing set in the tables below. The 2D annotations are released in https://github.com/zixuzhuang/OAI_seg.

In the following, we refer to the femur bone and tibia bone as FB and TB for short, their attached cartilages as FC and TC, and meniscus as M. We use GTN and RM to denote the Graph Transformer Network and random masking strategies, respectively. SR and NN stand for super-resolution and nearest neighbor interpolation.

To ensure the fairness of the comparison, we use the same training and testing settings for all experiments: PyTorch1.9.0, RTX 3090, learning rate is $3e-4$, 100 epochs, cosine learning rate decay, Adam optimizer.

3.2 Ablation Study

To find the optimal setting for the CAS-Net, we apply different hyper-parameters on it and test them on cross-view 2D segmentation. We evaluate the performance of the CAS-Net with different patch size, and then remove the GTN, random masking, and super-resolution to see the effect of each component. To ensure the multi-view slices can be aligned in the same space without super-resolution, we use trilinear interpolation as the alternative. The results are shown in Table 1.

Table 1. Results from a controlled ablation study of the proposed CAS-Net, examining the effects of patch size, GTN, randomly masking (RM), and super-resolution (SR).

Patch Size	Method	Coronal Dice Ratio (%)						Axial Dice Ratio (%)					
		FB	FC	TB	TC	M	Mean	FB	FC	TB	TC	M	Mean
64		96.1	79.3	95.3	80.7	79.1	86.1	96.9	79.9	96.9	78.6	76.9	85.8
128		96.5	79.7	96.0	81.9	79.9	86.8	97.1	80.1	97.1	78.8	77.8	86.2
128	w/o GTN	96.1	79.0	95.5	81.0	79.3	86.2	97.1	79.7	97.1	78.4	77.8	86.0
128	w/o RM	95.4	78.4	95.0	80.7	79.0	85.7	97.0	80.1	97.1	78.0	77.0	85.8
128	w/o SR	96.0	74.1	95.2	74.7	72.0	82.4	96.5	74.2	96.3	73.6	70.7	82.3

As shown in Table 1, the best performance appears when patch size is 128 (86.8% mean Dice ratio in coronal view), and the removing of GTN (86.2%), random masking (85.7%) or super-resolution (82.4%) always lead to performance drop. The patch size of 64 is too small to capture local appearance of the knee, and thus causes performance degradation. GTN enables the network to capture knee features globally and therefore brings performance improvement. Random masking encourages CAS-Net to learn knee segmentation with aligned views across different planes, avoiding errors caused by super-resolution and 3D pseudo segmentation, thus resulting in better performance. Super-resolution reduces the aliasing effect during resampling, resulting in better segmentation. Therefore, we select the patch size as 128, and keep GTN, RM and SR enabled in the next experiments.

3.3 Evaluation of 3D Segmentation Performance

We report the performance of the CAS-Net for 3D knee segmentation. To the best of our knowledge, CAS-Net is the first method that can perform knee segmentation from multi-view 2D scans and sagittal segmentation annotation. Therefore, we compare CAS-Net with the following settings: (1) the sagittal annotation that is resampled to 3D DESS space by nearest neighbor interpolation; (2) the pseudo-3D segmentation generated in Sect. 2.1; (3) the 3D prediction from 3D nnUNet based on 3D DESS images and annotations, which is considered as the upper bound for this task.

As can be seen from Table 2, the CAS-Net is capable of producing excellent 3D segmentation results with only sagittal annotation, as demonstrated in

Table 2. Results of the comparison of different methods in 3D segmentation.

Methods	Image View	Annotation View	3D Dice Ratio (%)					
			FB	FC	TB	TC	M	Mean
NN	-	Sagittal	96.1	76.4	95.2	72.5	79.7	83.9
Pseudo Seg	Sagittal	Sagittal	96.3	79.0	95.2	72.6	79.7	84.6
3D nnUNet	3D DESS	3D DESS	92.2	97.8	90.2	90.2	89.8	92.0
Ours	Multi-View	Sagittal	96.7	81.7	95.8	74.9	81.0	86.0

Fig. 3(f). In comparison to 3D nnUNet that utilizes 3D images and annotations, CAS-Net provides better performance on FB (96.7% vs 92.2% mean 3D Dice) and TB (95.8% vs 90.2%), since the 3D DESS sequence has a lower contrast of bone and muscle, resulting in more errors. Additionally, although there is roughly a 16% performance gap of cartilages compared to 3D nnUNet, CAS-Net demonstrates satisfactory outcomes in FC (81.7%) and M (81.0%) due to multi-view image localization errors like slight knee movements during scanning, resulting in misaligned segmentation generation from the multi-view scans and 3D annotation, as shown in Fig. 4. Furthermore, our method significantly improves 3D mean segmentation performance (86.0%) in comparison to the two techniques of using NN interpolation directly (83.9%) and employing super-resolution to generate pseudo sagittal segmentation (84.6%). These experimental findings indicate that our approach effectively leverages multi-view 2D scans and sagittal segment annotations to produce superior 3D segmentation.

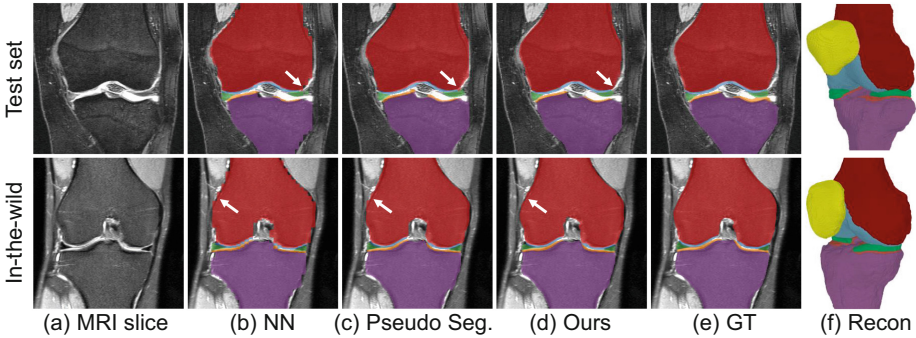


Fig. 3. We test our model on OAI data and in-the-wild clinical data. Here are two example cases with comparison of different methods.

3.4 Evaluation of Cross-View Projection of 3D Segmentation

The 3D segmentation results inferred by CAS-Net can be resampled onto unannotated 2D scans to achieve cross-view segmentation. We evaluate the cross-view

segmentation performance on the coronal view, which is shown in Table 3. We compare with the three alternative methods: (1) the sagittal annotation that is projected to the coronal view by nearest neighbor interpolation, named NN in Table 3, which is considered as the lower bound for the cross-view segmentation task; (2) the pseudo-segmentation generated in Sect. 2.1 projected to the coronal view, named as Pseudo Seg. in Table 3; (3) the 2D prediction from 2D nnUNet based on coronal 2D scans and annotations, which is considered as the upper bound for the cross-view segmentation task. The visualized results are shown in Fig. 3(b)–(d).

Table 3. Results of the comparison of different methods in 2D segmentation.

Methods	Image View	Annotation View	Coronal Dice Ratio (%)					
			FB	FC	TB	TC	M	Mean
NN	-	Sagittal	96.1	76.4	95.2	71.5	79.7	83.8
Pseudo Seg	Sagittal	Sagittal	96.3	79.1	95.3	72.1	79.7	84.5
nnUNet	3D DESS	Coronal	98.9	92.2	98.8	92.8	92.9	95.1
Ours	Sagittal	Sagittal	96.5	79.7	96.0	81.9	79.9	86.8

CAS-Net not only eliminates some segmentation errors but also provides smooth segmentation boundaries in cross-view segmentation tasks. As seen in Table 3 (Row 1/2) and Fig. 3(b) and (c), while the super-resolution method has improved the mean Dice ratio by only 0.7%, it has resulted in noticeably smoother segmentation boundary of the knee tissue. The CAS-Net then further eliminates errors in Pseudo Segmentation, resulting in better segmentation outcomes and leading to additional improvements (86.8% mean coronal Dice).

Compared with nnUNet, there is still improvement room for CAS-Net in the case of small tissues, such as meniscus (79.9% vs. 92.9% Dice ratio). This is likely caused by subtle movements of subjects during the acquisition of multi-view 2D scans. As shown in Fig. 4, compared to sagittal MRI, FC is slightly shifted upwards in coronal and 3D MRIs. Thus the segmentation is aligned in sagittal view but not in coronal and 3D. We have examined subjects in the dataset

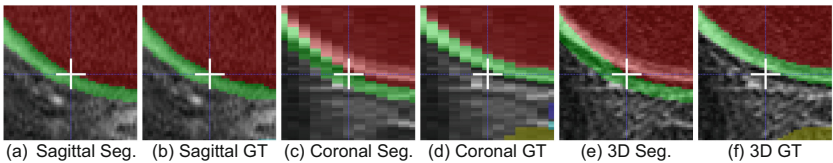


Fig. 4. The multi-view scans and 3D image are misaligned due to knee motion during scanning, though the segmentation have the correct shape as in sagittal view. The white cross indicates the same position in WCS space but cartilage is moved.

that have lower performance (around 50% mean Dice ratio) and confirmed this observation. On the contrary, given correctly aligned multi-view 2D scans, CAS-Net can achieve a mean Dice ratio of 95%.

4 Conclusion

We have proposed a novel framework named CAS-Net for generating cross-view consistent 3D knee segmentation via super-resolution and graph representation with clinical 2D multi-view scans and sagittal annotations. By creating a detailed 3D knee segmentation from clinical 2D multi-view MRI, our framework provides significant benefits to morphology-based knee analysis, with promising applications in knee disease analysis. We believe that this framework can be extended to other 3D segmentation tasks, and we intend to explore those possibilities. However, it should be noted that the performance of the CAS-Net is affected by misaligned multi-view images, as shown in Fig. 4. After removing the misaligned cases, the femur cartilage dice can be increased from 81.7% to 87.3% in 3D segmentation. We plan to address this issue by incorporating multimodal alignment in future work.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (No. 62001292), and partially supported by the National Natural Science Foundation of China (U22A20283), and the Interdisciplinary Program of Shanghai Jiao Tong University (No. YG2023LC07).

References

1. Ambellan, F., Tack, A., Ehlke, M., Zachow, S.: Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the osteoarthritis initiative. *Med. Image Anal.* **52**, 109–118 (2019)
2. Calivà, F., Namiri, N.K., Dubreuil, M., Pedoia, V., Ozhinsky, E., Majumdar, S.: Studying osteoarthritis with artificial intelligence applied to magnetic resonance imaging. *Nat. Rev. Rheumatol.* **18**, 1–10 (2021)
3. Carballido-Gamio, J., et al.: Inter-subject comparison of MRI knee cartilage thickness. *Med. Image Anal.* **12**(2), 120–135 (2008)
4. Dwivedi, V.P., Bresson, X.: A generalization of transformer networks to graphs. arXiv preprint [arXiv:2012.09699](https://arxiv.org/abs/2012.09699) (2020)
5. Eckstein, F., Wirth, W., Nevitt, M.C.: Recent advances in osteoarthritis imaging—the osteoarthritis initiative. *Nat. Rev. Rheumatol.* **8**(10), 622–630 (2012)
6. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021)
7. Kakigi, T., et al.: Diagnostic advantage of thin slice 2D MRI and multiplanar reconstruction of the knee joint using deep learning based denoising approach. *Sci. Rep.* **12**(1), 1–14 (2022)

8. Kijowski, R., Davis, K.W., Blankenbaker, D.G., Woods, M.A., Del Rio, A.M., De Smet, A.A.: Evaluation of the menisci of the knee joint using three-dimensional isotropic resolution fast spin-echo imaging: diagnostic performance in 250 patients with surgical correlation. *Skeletal Radiol.* **41**, 169–178 (2012)
9. Li, T., Xuan, K., Xue, Z., Chen, L., Zhang, L., Qian, D.: Cross-view label transfer in knee MR segmentation using iterative context learning. In: Albarqouni, S., et al. (eds.) DART/DCL -2020. LNCS, vol. 12444, pp. 96–105. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60548-3_10
10. Liu, D., et al.: Transfusion: multi-view divergent fusion for medical image segmentation with transformers. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13435, pp. 485–495. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_47
11. Perslev, M., Dam, E.B., Pai, A., Igel, C.: One network to segment them all: a general, lightweight system for accurate 3D medical image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 30–38. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_4
12. Perslev, M., Pai, A., Runhaar, J., Igel, C., Dam, E.B.: Cross-cohort automatic knee MRI segmentation with multi-planar U-nets. *J. Magn. Reson. Imaging* **55**(6), 1650–1663 (2022)
13. Recht, M.P., et al.: Using deep learning to accelerate knee MRI at 3 T: results of an interchangeability study. *AJR Am. J. Roentgenol.* **215**(6), 1421 (2020)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
15. Schmidt, A.M., et al.: Generalizability of deep learning segmentation algorithms for automated assessment of cartilage morphology and MRI relaxometry. *J. Magn. Reson. Imaging* **57**(4), 1029–1039 (2022)
16. Shakoob, D., et al.: Diagnosis of knee meniscal injuries by using three-dimensional MRI: a systematic review and meta-analysis of diagnostic performance. *Radiology* **290**(2), 435–445 (2019)
17. Vanwanseele, B., et al.: The relationship between knee adduction moment and cartilage and meniscus morphology in women with osteoarthritis. *Osteoarthritis Cartilage* **18**(7), 894–901 (2010)
18. Vaswani, A., et al.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
19. Wang, X., Xuan, K., Wang, S., Xiong, H., Zhang, L., Wang, Q.: Arbitrary reduction of MRI slice spacing based on local-aware implicit representation. *arXiv preprint arXiv:2205.11346* (2022)
20. Wenger, A., et al.: Relationship of 3D meniscal morphology and position with knee pain in subjects with knee osteoarthritis: a pilot study. *Eur. Radiol.* **22**, 211–220 (2012)
21. Zhuang, Z., et al.: Knee cartilage defect assessment by graph representation and surface convolution. *IEEE Trans. Med. Imaging* **42**(2), 368–379 (2022)
22. Zhuang, Z., et al.: Local graph fusion of multi-view MR images for knee osteoarthritis diagnosis. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 554–563. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_53