Letting Osteocytes Teach SR-microCT Bone Lacunae Segmentation: A Feature Variation Distillation Method via Diffusion Denoising

Isabella Poles, Marco D. Santambrogio, and Eleonora D'Arnese*

Politecnico di Milano, Milano, Italy
{isabella.poles, marco.santambrogio, eleonora.darnese}@polimi.it

Abstract. Synchrotron Radiation micro-Computed Tomography (SR-microCT) is a promising imaging technique for osteocyte-lacunar bone pathophysiology study. However, acquiring them costs more than histopathology, thus requiring multi-modal approaches to enrich limited/costly data with complementary information. Nevertheless, paired modalities are rarely available in clinical settings. To overcome these problems, we present a novel histopathology-enhanced disease-aware distillation model for bone microstructure segmentation from SR-microCTs. Our method uses unpaired histopathology images to emphasize lacunae morphology during SR-microCT image training while avoiding the need for histopathologies during testing. Specifically, we leverage denoising diffusion to eliminate the noisy information within the student and distill valuable information effectively. On top of this, a feature variation distillation method pushes the student to learn intra-class semantic variations similar to the teacher, improving label co-occurrence information learning. Experimental results on clinical and public microscopy datasets demonstrate superior performance over single-, multi-modal, and state-of-the-art distillation methods for image segmentation.

Keywords: Lacunae Segmentation · Knowledge Distillation · SR-microCT

1 Introduction

Bone lacunar microstructures are pivotal in controlling bone formation and resorption, reflecting the signals produced by the osteocyte cells they contain in response to mechanical stimuli, diseases, and aging [23]. Indeed, osteocyte-lacunar pathophysiology study is essential for the early detection of bone damage resulting from bone remodeling, osteoporosis, and viral infections [18]. To this extent, histopathology and Synchrotron Radiation micro Computed Tomography (SR-microCT) imaging are among the most employed techniques [3]. While both hold value, SR-microCT stands out for its phase contrast imaging and combination with in-situ mechanical testing [17]. Unfortunately, manual segmentation of bone lacunae is burdensome and costly due to the microstructure complexity. Conversely, automatic segmentation with Convolutional Neural Networks (CNNs) is

* Corresponding author

promising but limited by dataset scarcity and microstructure morphology heterogeneity [20]. Compared to SR-microCTs, histopathologies, thanks to their rapid imaging and low radiation, are more accessible and cost-effective [6], complementing the morphology and densitometric information of SR-microCTs and visualizing osteocytes, which cause lacunae geometry variations [17]. Combining these sources could improve model generalization capabilities, and, guided by this intuition, we propose a novel learning paradigm for bone lacunae segmentation exploiting SR-microCT and unpaired bone histopathology images.

Common strategies to combine multi-modal data in the literature involve independent networks for feature extraction [4, 19, 24, 26–28] or concatenation of multi-modal images at the onset of the framework [14, 16, 25]. In addition, late/early feature fusion with co-/cross-attention modules could facilitate knowledge transfer across modalities. Nevertheless, these approaches present two limitations. Firstly, modality-specific networks may be insufficient for modality interactions, thus negatively affecting the model integration abilities. Although "Chilopod"-shaped multi-modal learning shares all CNN kernels across modalities, it remains constrained when there is a significant information disparity between modalities [5]. Secondly, unpaired data in the fusion/cross-attention interaction or the use of fine-grained image-level labels [8] may lead to acquiring irrelevant information owing to differences in image distributions.

Stemming from these limitations and motivated by bone histopathologies and SR-microCTs interdependence, we propose a method[1] for learning lacunae segmentation from SR-microCTs while effectively integrating histopathological information (Figure 1). Indeed, histopathologies and SR-microCTs have complementary functions but similar semantic characteristics: the first reveals osteocyte variations that provide physiological insights into bone remodeling or disease processes, while the latter shows lacunae, which shape influenced by the osteocytes they contain allows the prediction of microcrack mechanical generation and progression [1]. Therefore, our primary goal is to extract bone status information from a histopathology teacher model and to transfer it to a SR-microCT student in an utterly unpaired setting. To achieve this, we see the student model as a noisy version of the teacher having sub-optimal training ability to learn discriminative features; thus, we leverage **denoising diffusion**, to **denoise the student features and only distill relevant information**. On top of this, a feature variation distillation method encourages the **denoised student to learn intra-class semantic variations similar to the teacher ones**, obtaining additional sharp label co-occurrence information. Unlike current methods, our approach opens to **extract valuable knowledge from many histopathology images** while **relying on the sole SR-microCT during inference**, making it more accessible and cost-effective for microscale bone analysis.

We can summarize our contributions as follows: 1) A novel histopathology-enhanced disease-aware distillation model for bone lacunae segmentation by matching pixel-to-class prototypes from diffusion-denoised features; 2) A diffusion-denoising method that offers flexible intra-class semantic knowledge transfer

[1] `https://github.com/isabellapoles/LOTUS`

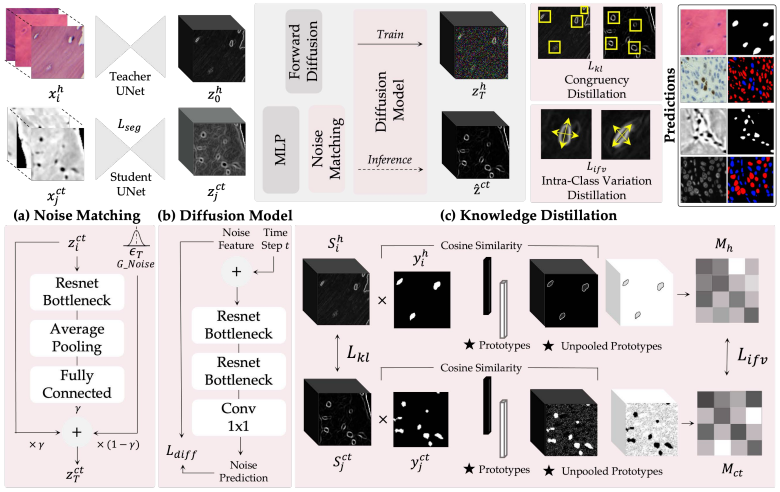

Fig. 1. Overview of our method from the diffusion module, which denoises image features, to their intra-class variation alignment, which distills disease-specific traits.

from any available histopathology image significantly reducing the cost of collecting paired multi-modal data; 3) A validation of our method using both a clinical and a public microscopy dataset demonstrating superior performance in microstructure image segmentation compared to state-of-the-art alternatives.

2 Proposed Methodology

Our methodology is based on two fundamental principles: diffusion-based feature denoising and intra-class semantic knowledge variation distillation. The former denoises the student features, ensuring a consistent distillation, while the latter aims to match pixel-to-class prototypes from the teacher to the denoised student to share disease-specific traits. Details of both components are discussed in the section below, and an overview of the proposed framework is shown in Figure 1.

We define the histopathology dataset as $D_h = \left\{x_i^h, y_i^h\right\}_{i=1}^{K}$ and the SR-microCT dataset as $D_{ct} = \left\{x_j^{ct}, y_j^{ct}\right\}_{j=1}^{L}$. To leverage histopathology knowledge during training, we construct a teacher model trained on D_h. Similarly, a student model learns from SR-microCT images D_{ct} using the same backbone architecture of the teacher. We employ as the segmentation loss L_{seg} a combination of Dice Similarity Coefficient (DSC) and binary cross-entropy. Finally, only SR-microCT data is fed to the SR-microCT model to generate lacunae masks during inference.

2.1 Feature Denoising via Diffusion for Distillation

Student model features exhibit larger values and variances upon an incorrect teacher prediction, being more noisy than the teacher's ones [15]. This noise,

arising from the disparity between the two models' capabilities, limits the student's ability to emulate the teacher's behavior; hence, securing feature similarity during distillation from the teacher to the student model is imperative.

Therefore, to mitigate noise and facilitate the subsequent distillation of disease-specific features, we propose to train a diffusion module with the teacher features and employ it to denoise the student ones (Figure 1(b)) [12]. Formally, leveraging teacher feature z_h and student feature z_{ct}, we iteratively add Gaussian noise $\epsilon_t \in \mathcal{N}(0, I)$ to z_h within the diffusion forward noise process defined as:

$$q(z_t^{(h)}|z^{(h)}) = \mathcal{N}(z_t^{(h)}|\sqrt{\overline{\alpha_t}^{(h)}}z_0^{(h)}, (1-\alpha_t^{(h)})I) \tag{1}$$

where $z_t^{(h)}$ is the transformed noisy data at time step $t \in \{0, 1, ..., T\}$, $\overline{\alpha_t}^h = \prod_{s=0}^{t} \alpha_s^h = \prod_{s=0}^{t}(1-\beta_s)$ is a notation for directly sampling $z_t^{(h)}$ at arbitrary timestamp with noise variance schedule β [11]. To reduce the computation cost of the diffusion mechanism, we fed $z_t^{(h)}$ into a light diffusion model $\Phi_\theta(z_t^{(h)}, t)$, which is a stack of two subsequent ResNet bottleneck blocks, trained to predict the noise ϵ_t in $z_t^{(h)}$ with respect to $z_0^{(h)}$ minimizing the L2 norm between them:

$$L_{diff} = \left\| \epsilon_t - \Phi_\theta(z_t^{(h)}, t) \right\|_2^2 \tag{2}$$

Then, during inference, to generate the denoised student feature $\hat{z}^{(ct)}$, its noisy feature $z_t^{(ct)}$ is fed to the iterative denoising process of our diffusion model:

$$p_\theta(z_{t-1}^{(ct)}|z_t^{(ct)}) = \mathcal{N}(\mathbf{P}(z_{t-1}^{(ct)}); \Phi_\theta(\mathbf{P}(z_t^{(ct)}), t), \sigma_t^2 I) \tag{3}$$

where $\mathbf{P}$ (i.e., multilayer projector) projects the SR-microCT features into the space of the histopathology ones, while σ_t^2 is the denoising diffusion implicit model transition variance, which accelerates the denoising by a small number of score function evaluations sampling [22]. Besides, since the teacher-student noise gap is unknown and may vary depending on the training sample, the initial diffusion process time step remains unknown. Therefore, we adopt a convolutional module to learn a weight γ that fuses student output and Gaussian noise as $z_T^{(ct)} = \gamma z^{(ct)} + (1-\gamma)\epsilon_T$, which allows the matching of the student output to the same noisy level of the noisy feature at the initial time step T (Figure 1(a)).

2.2 Intra-class Semantic Knowledge Variation Distillation

Once the student features have been denoised, we aim to align the *consistency* and *variability* of image features with those of the teacher model (Figure 1(c)). To exemplify, the average count of lacunae is typically constant unless influenced by factors such as bone remodeling under COVID-19 or osteoporosis, yet the shape of the associated osteocytes may undergo alterations. Transferring the consistency and variability of features representing those sample characteristics from teacher to student could enhance the student's fidelity to the teacher's

feature distribution, thus improving its performance. Therefore, we adopt the Kullback-Leibler divergence to add a first constraint on student predictions for features that appear *congruent* with the teacher ones [10]. Formally, we minimize the Kullback-Leibler divergence between the output score maps of the models:

$$L_{kl} = \frac{1}{N} \sum_{p \in \Omega} \sum_{i=1}^{C} S_i^{(ct)}(p) \cdot log \frac{S_i^{(ct)}(p)}{S_i^{(h)}(p)} \tag{4}$$

where Ω is the image domain, C denotes the number of semantic classes, and $S_i^{(h)}(p)$ and $S_i^{(ct)}(p)$ refer to the probability assigned to the i-th class at pixel p by the student model and the teacher model, respectively.

To ensure comprehensive learning of disease-specific feature *variations*, we condense features associated with each semantic class into a class prototype vector. Then, we characterize the intra-class feature variation of each model using a feature similarity map between each image pixel and its respective class-wise prototype. More in detail, the prototype for each semantic class i is computed by averaging the features across all pixels sharing the same class label i. Subsequently, we employ the cosine similarity to determine the feature similarity between each pixel and its corresponding class-wise prototype. The intra-class feature variation map M is formulated as:

$$M(p) = sim(f(p), \frac{1}{|S_p|} \sum_{q \in S_p} f(q)) \tag{5}$$

where $f(p)$ are the feature at pixel p, S_p is the set of pixels with the same label as pixel p, $|S_p|$ indicates the size of set S_p, and sim denotes the similarity function. Subsequently, to transfer intra-class relationship information to the student, we minimize the distance between the teacher and the student intra-class feature variation maps. Specifically, we utilize the Mean Squared (L2) loss defined as:

$$L_{ifv} = \frac{1}{N} \sum_{p \in \Omega} (M_{ct}(p) - M_h(p))^2 \tag{6}$$

where N is the number of pixels, and M_h and M_{ct} denote the corresponding intra-class feature variation maps of the teacher and the student, respectively.

2.3 Overall Framework

In our proposed approach, the whole training aims to minimize the combination of the segmentation, diffusion denoising, and distillation enhancement loss $L_{ct} = L_{seg} + \lambda_1 L_{diff} + \lambda_2 L_{kl} + \lambda_3 L_{ifv}$ where $\lambda_1, \lambda_2, \lambda_3$ are loss weights set to 1 to balance the losses in all experiments. Moreover, we regulate the model optimization process by leveraging the γ weight (Section 2.1). Specifically, we halt the optimization of the diffusion module and pass to distilling knowledge once the noise level in the student features diminishes and γ stabilizes on a plateau (30 epochs patience), aligning with the reduced noise level in the teacher's data.

Table 1. Results on our collected bone histopathology-SR-microCT dataset. "Training" and "Inference" indicate which modalities are required for both phases. "Paired" indicates wheter paired histopathology-SR-microCT images are required in training. All experiments use UNet as the backbone. The best results are in bold, while the (**) and (*) indicate 0.001 and 0.05 p-values significant difference.

Method	Training	Inference	Paired	DSC_{FG} ↑	HD ↓	AJI ↑	PQ ↑
Teacher_UNet	histo	histo	-	65.96±4.73	99.89±22.43	49.78±4.56	40.91±6.09
Student_UNet	μCT	μCT	-	55.69±10.56	148.78±41.88	39.60±9.69	29.36±8.06
Single-Modal Methods							
UNet++ [29]	μCT	μCT	-	56.62±10.11*	138.88±34.41*	40.20±9.53*	29.49±7.28*
U2Net [21]	μCT	μCT	-	57.20±9.66*	137.53±35.04*	40.87±9.35*	30.63±7.93*
Swin-UNet [2]	μCT	μCT	-	59.61±7.79*	122.32±35.23*	43.30±7.60*	32.34±6.88*
nnUNet [13]	μCT	μCT	-	60.10±8.36*	114.04±30.12*	43.95±5.59*	33.06±5.76*
Multi-Modal Methods							
MAML [27]	Both	Both	✓	60.51±3.13*	113.54±28.91*	43.46±3.20*	32.44± 5.88*
MEDIAR [16]	Both	μCT	✓	56.94±6.86**	135.92±33.09*	40.82±10.12*	29.88±9.69*
UMMKD [5]	Both	μCT	×	58.92±6.22*	126.73±32.23*	42.01±8.82*	32.93±7.22*
Ours	**Both**	**μCT**	×	**61.59±7.87**	**109.86±37.35**	**45.12±8.06**	**34.74±6.98**

3 Experiments

3.1 Experimental Setup

Dataset. To assess our approach's effectiveness, we collect a new dataset comprising 404 bone histopathologies and 1077 SR-microCT unpaired image patches from human femoral head samples, covering healthy, osteoporotic, and COVID-19 cases. SR-microCT imaging yielded 2048 slices per volume, with an average size of 3300 pixels per side at 1.6μm pixel resolution, and two operators manually labeled 32.1k lacunar structures. Histopathologies derive from decalcified femoral head bone samples that previously underwent SR-microCT compression cycles where tissue sections of 5μm thickness were stained with H&E and imaged at 10× magnification with an average dimension of 2354×1890 pixels. A single operator segmented 1343 osteocyte cells across all images.

Implementation Details. For bone histopathology images, no processing steps are applied. However, for SR-microCT images, we employ min-max normalization, contrast adjustment to the bottom and top 1% of pixel values, and masking to only preserve the adjusted image content within the bone region. Both modalities undergo data augmentation, including 512×512 patch resizing, random cropping, flipping, rotation and contrast, saturation, and brightness adjustments. The optimization of the ∼3.02M model parameters is carried out using Adam with a learning rate of $1 \cdot 10^{-3}$, a momentum of 0.9, and a weight decay of $1 \cdot 10^{-7}$, with a batch size of 16, employing PyTorch (1.13.0) on a AMD Ryzen 7 5800X 3.8 GHz with a 24 GB NVIDIA RTX A5000 GPU. The dataset is split patient-wise into training, validation, and testing subsets, maintaining an approximate ratio of 8:1:1, while to ensure model robustness, results are reported on the test set after five-fold cross-validation and a paired t-test statistical analysis.

Table 2. Effectiveness of each module in our method. The best results are in bold, while the (**) and (*) indicate 0.001 and 0.05 p-values significant difference.

Method				Metric			
L_{seg}	L_{diff}	L_{kl}	L_{ifv}	DSC_{FG} ↑	HD ↓	AJI↑	PQ↑
✓	×	×	×	55.69±10.56**	148.78±41.88**	39.60±9.69**	29.36±8.06*
✓	×	✓	×	57.58±9.15*	137.95±38.26*	41.09±8.64*	29.24±7.63*
✓	×	×	✓	57.58±9.96*	142.33±36.15*	41.12±9.63*	31.41±8.21*
✓	✓	×	✓	59.66±8.37*	123.91±35.65*	42.91±8.03*	32.04±7.07*
✓	✓	✓	×	59.73±8.39*	113.81±38.58*	42.85±7.88*	30.94±7.27*
✓	✓	✓	✓	**61.59±7.87**	**109.86±37.35**	**45.12±8.06**	**34.74±6.98**

3.2 Comparison with State-of-the-Art Approaches

We conducted a comparative analysis against single-modal, multi-modal, and knowledge distillation techniques to validate our proposed methodology. The results in Table 1 show the superior performance of our histopathology-based teacher model over the student SR-microCT-based one, demonstrating the margin of information that could still be learned. Additionally, our proposed method exhibits substantial improvements over the student single-modal SR-microCT. Specifically, it improves foreground DSC_{FG} from 55.69% to 61.59%, Aggregated Jaccard Index (AJI) from 39.60% to 45.12%, Panoptic Quality (PQ) from 29.36% to 34.74%, while reducing Hausdorff Distance (HD) from 148.78 to 109.86. This underscores our model's capability to leverage valuable insights from the histopathologies to augment the SR-microCT model, which inherently possesses incomplete semantic information. Furthermore, it is noteworthy that current single- [2, 13, 21, 29], multi-modality [16, 27] and knowledge distillation methodologies [5] perform worse on our dataset even upon retraining. Indeed, our approach surpasses the leading multi-modal cell segmentation method [16] by 4.65% in terms of DSC (61.59% *v.s.* 56.94%) and 4.86% in PQ (34.74% *v.s.* 29.88%). Remarkably, while alternative methodologies are constrained by the need for training solely with modality-paired images, our approach operates without such limitations. However, our methodology also outperforms the unpaired knowledge distillation method [21] by 2.67% in DSC (61.59% *v.s.* 58.92%) and 3.19% in AJI (45.12% *v.s.* 42.01%) while being statistically different and limiting WSI inference time at $12.27 \pm 0.02s$.

3.3 Ablation Study

To provide additional insights, Table 2 shows the ablation study of our method. Individually adding L_{kl} and L_{ifv} to the sole L_{seg} improve the overall results. In particular, L_{kl} improves the semantic segmentation performance of 1.89% in DSC by distilling consistent-specific knowledge between the teacher and the student, while L_{ifv} benefits more from the instance counterpart with a 2.05% increase in PQ by attending to intra-class feature variations. Adopting also L_{diff} to denoise student features before knowledge distillation with L_{kl} and L_{ifv} individually improves each score consistently. Nonetheless, the overall framework

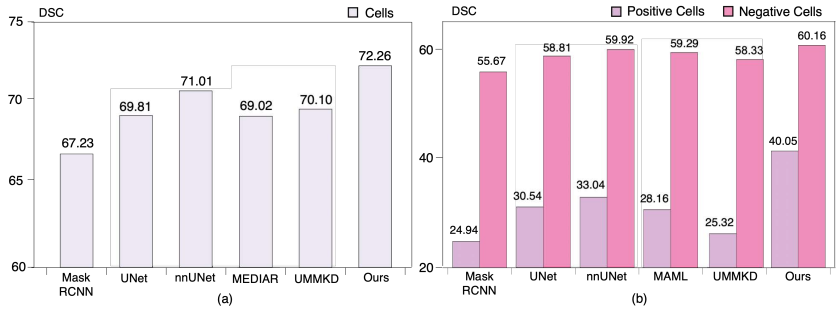

Fig. 2. Results on the publicly available DeepLIIF dataset for cells instance segmentation (a) and multiclass semantic segmentation (b).

reports the best metrics, statistically demonstrating the effectiveness of the combined diffusion feature denoising and semantic knowledge distillation.

3.4 Results with Other Datasets

Since we implement distillation in a disease-aware manner, we do not need paired multi-modal data for training and our method could be applied to any publicly available multi-modal dataset *unless its modalities share complementary semantic information between themselves.* To verify this hypothesis, we reproduce our method on the DeepLIIF dataset [7] during instance (Figure 2(a)) and multiclass semantic cell segmentation (Figure 2(b)). Given the more complex analysis of fluorescence images, we aim at aiding their segmentation with knowledge transfer from its immunochemistry complementary modality. We reproduce MaskR-CNN [9], nnUNet, multi-modal, and distillation methods as baselines. Compared with single-modal fluorescence models, our strategy can increase DSC of 5.03% (72.26% *v.s.* 67.23%) during instance segmentation and 15.11% (40.05% *v.s.* 24.94%) on the minority class during multiclass semantic segmentation. Furthermore, it surpasses the unpaired distillation method up to 2.16% (72.26% *v.s.* 70.10%) and 14.74% (40.05% *v.s.* 25.35%) in DSC. Focusing on clinical application, our method has the efficiency ($0.005 \pm 0.019s$ for each patch prediction) and flexibility to accentuate/disambiguate cellular information from any complex image modality, augmenting their knowledge with lower-cost ones and obtaining a more informative representation.

4 Conclusion

Our work proposes a novel histopathology-enhanced disease-aware distillation model for bone lacunae segmentation. It incorporates a lightweight and adaptive diffusion module to denoise image features during training while distilling only the valuable semantic information to ensure the consistency of disease-related characteristics between both modalities. Our approach deviates from the existing

models that rely on paired multi-modal training and inference, making it possible to extract knowledge from any available image and render predictions with only one modality. As a result, our approach significantly reduces the prerequisites for clinical applications. Our extensive experiments demonstrate that our method outperforms existing baselines by a considerable margin and has the adaptability to emphasize/clarify cellular information from state-of-the-art datasets.

Compliance with Ethical Standards. Femoral heads are collected with prior authorization from the Ethics Committee (approval date: 13/05/2020, ClinicalTrials.gov ID: NCT04787679) of San Raffaele Hospital (Milan, Italy) and signed approval consent of the patients.

Acknowledgments. This work was supported by Polisocial Award 2022 - Politecnico di Milano. The authors acknowledge F. Buccino and M. Vergani for their expertise on bone lacunae and osteocytes mechanics, Elettra Sincrotrone Trieste for providing access to its synchrotron radiation facilities, A. Zeni and D. Conficconi for valuable suggestions and discussions and NVIDIA Corporation for the Academic Hardware Grant Program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Buccino, F., Zagra, L., Longo, E., D'Amico, L., Banfi, G., Berto, F., Tromba, G., Vergani, L.M.: Osteoporosis and covid-19: Detected similarities in bone lacunar-level alterations via combined ai and advanced synchrotron testing. Materials & Design p. 112087 (2023)
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
3. Carter, Y., Thomas, C.D.L., Clement, J.G., Peele, A.G., Hannah, K., Cooper, D.M.: Variation in osteocyte lacunar morphology and density in the human femur—a synchrotron radiation micro-ct study. Bone **52**(1), 126–132 (2013)
4. Chen, X., Zhou, H.Y., Liu, F., Guo, J., Wang, L., Yu, Y.: Mass: Modality-collaborative semi-supervised segmentation by exploiting cross-modal consistency from unpaired ct and mri images. Medical Image Analysis **80**, 102506 (2022)
5. Dou, Q., Liu, Q., Heng, P.A., Glocker, B.: Unpaired multi-modal segmentation via knowledge distillation. IEEE transactions on medical imaging **39**(7), 2415–2425 (2020)
6. Dreyer Vetter, S., Schurman, C.A., Alliston, T., Slabaugh, G., Verbruggen, S.W.: Deep learning models to map osteocyte networks can successfully distinguish between young and aged bone. bioRxiv pp. 2023–12 (2023)
7. Ghahremani, P., Marino, J., Hernandez-Prera, J., V. de la Iglesia, J., JC Slebos, R., H. Chung, C., Nadeem, S.: An ai-ready multiplex staining dataset for reproducible and accurate characterization of tumor immune microenvironment. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2023)

8. Guan, Q., Xie, Y., Yang, B., Zhang, J., Liao, Z., Wu, Q., Xia, Y.: Unpaired cross-modal interaction learning for covid-19 segmentation on limited ct images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 603–613. Springer (2023)
9. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
10. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
12. Huang, T., Zhang, Y., Zheng, M., You, S., Wang, F., Qian, C., Xu, C.: Knowledge diffusion for distillation. Advances in Neural Information Processing Systems **36** (2024)
13. 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. Nature methods **18**(2), 203–211 (2021)
14. Joubbi, S., Ciano, G., Cardamone, D., Maccari, G., Medini, D.: Crossct: Cnn and transformer cross–teaching for multimodal image cell segmentation. In: Competitions in Neural Information Processing Systems. pp. 1–14. PMLR (2023)
15. Kundu, S., Sun, Q., Fu, Y., Pedram, M., Beerel, P.: Analyzing the confidentiality of undistillable teachers in knowledge distillation. Advances in Neural Information Processing Systems **34**, 9181–9192 (2021)
16. Lee, G., Kim, S., Kim, J., Yun, S.Y.: Mediar: Harmony of data-centric and model-centric for multi-modality microscopy. In: Competitions in Neural Information Processing Systems. pp. 1–16. PMLR (2023)
17. Lui, E., Maruyama, M., Guzman, R.A., Moeinzadeh, S., Pan, C.C., Pius, A.K., Quig, M.S., Wong, L.E., Goodman, S.B., Yang, Y.P.: Applying deep learning to quantify empty lacunae in histologic sections of osteonecrosis of the femoral head. Journal of Orthopaedic Research® **40**(8), 1801–1809 (2022)
18. Mastrogiacomo, M., Campi, G., Cancedda, R., Cedola, A.: Synchrotron radiation techniques boost the research in bone tissue engineering. Acta Biomaterialia **89**, 33–46 (2019)
19. Mo, S., Cai, M., Lin, L., Tong, R., Chen, Q., Wang, F., Hu, H., Iwamoto, Y., Han, X.H., Chen, Y.W.: Multimodal priors guided segmentation of liver lesions in mri using mutual information based graph co-attention networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. pp. 429–438. Springer (2020)
20. Poles, I., D'Arnese, E., Buccino, F., Vergani, L., Santambrogio, M.D.: On how to unravel bone microscale phenomena: A mask-guided attention sr-microct image classification approach. In: 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). pp. 1–4. IEEE (2023)
21. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M.: U2-net: Going deeper with nested u-structure for salient object detection. Pattern recognition **106**, 107404 (2020)
22. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
23. Uda, Y., Azab, E., Sun, N., Shi, C., Pajevic, P.D.: Osteocyte mechanobiology. Current osteoporosis reports **15**, 318–325 (2017)

24. Wang, H., Ma, C., Zhang, J., Zhang, Y., Avery, J., Hull, L., Carneiro, G.: Learnable cross-modal knowledge distillation for multi-modal learning with missing modality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2023)
25. Zhang, B., Dong, J., Zhao, Z., Meng, Z., Su, F.: Mt2: Multi-task mean teacher for semi-supervised cell segmentation. In: Competitions in Neural Information Processing Systems. pp. 1–13. PMLR (2023)
26. Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., Zhang, Y., He, Z., Zheng, Y.: mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 107–117. Springer (2022)
27. Zhang, Y., Yang, J., Tian, J., Shi, Z., Zhong, C., Zhang, Y., He, Z.: Modality-aware mutual learning for multi-modal medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 589–599. Springer (2021)
28. Zhao, J., Li, S.: Learning reliability of multi-modality medical images for tumor segmentation via evidence-identified denoising diffusion probabilistic models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 682–691. Springer (2023)
29. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)