Less-than-one shot 3D segmentation hijacking a pre-trained Space-Time Memory network

Cyril Li, Christophe Ducottet, Sylvain Desroziers, Maxime Moreaud

Less-than-one shot 3D segmentation hijacking a pre-trained Space-Time Memory network

Cyril Li[1], Christophe Ducottet[1], Sylvain Deroziers[2], and Maxime Moreaud[3]

[1] Université Jean Monnet Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France
{cyril.li, ducottet}@univ-st-etienne.fr
[2] Manufacture Française des Pneumatiques Michelin, 23 Place des Carmes Déchaux, 63000 Clermont-Ferrand, France
sylvain.desroziers@michelin.com
[3] IFP Energies nouvelles, Rond-point de l'échangeur de Solaize BP 3, 69360 Solaize, France
maxime.moreaud@ifpen.fr

Abstract. In this paper, we propose a semi-supervised setting for semantic segmentation of a whole volume from only a tiny portion of one slice annotated using a memory-aware network pre-trained on video object segmentation without additional fine-tuning. The network is modified to transfer annotations of one partially annotated slice to the whole slice, then to the whole volume. This method discards the need for training the model. Applied to Electron Tomography, where manual annotations are time-consuming, it achieves good segmentation results considering a labeled area of only a few percent of a single slice. The source code is available at https://github.com/licyril1403/hijacked-STM.

Keywords: Deep Neural Network · Electron Tomography · Weakly Annotated Data · Memory Network · Semi-Supervised Segmentation.

1 Introduction

Electron Tomography (ET) [7] is a powerful technique to reconstruct 3D nanoscale material microstructure. A Transmission Electron Microscope (TEM) acquires sets of projections from several angles, allowing the reconstruction of 3D volumes. However, the resulting data contain noisy reconstruction artifacts because the number of projections is limited, and their alignment remains a challenging task [25] (Figure 1). Thus, standard segmentation methods often fail [8], requiring the input of an expert to achieve a good segmentation [9,13,26].

Deep learning (DL) based approaches have achieved excellent results in this area [1,14,11], as advances are made in semantic segmentation of 2D and 3D images [22,5,19,2,24]. Standard approaches rely on training a neural network on fully labeled datasets, which requires many annotated 3D samples. To address the problem of low availability of annotated data, transfer learning [21,28] or semi-supervised learning methods have been proposed with various learning

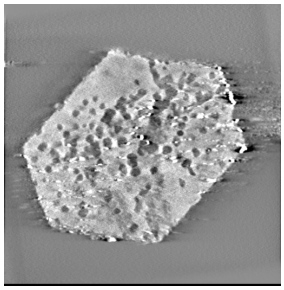

Fig. 1. Reconstruction of a zeolite slice (resolution: 1nm/pixel) showing many artifacts and noise, making automated segmentation difficult.

strategies [12,6,29,15]. These methods still require a specific training step for each new image type or object class to segment. Specifically for ET, a new training step will be necessary for every new material and acquisition condition. Moreover, the training process requires some expertise in machine learning to be done properly.

In video object segmentation, a problem close to volume segmentation, significant progress has been made using memory networks [20,27,4]. The system is fed with an annotated query frame and provides a complete segmentation of the corresponding video. The query frame is usually the first frame of the video, manually segmented by an annotator. This setup, similar to one-shot segmentation, has the benefit of being class independent and does not require any learning for new query images. Memory networks encode the annotated frame in the memory and segment the remaining frames using that memory. However, they have two main drawbacks. First, they need various tedious training steps and large datasets for training. Second, they require a whole segmented frame as input.

In this paper, we propose to hijack a Space-Time Memory (STM) network pre-trained for video object segmentation and use it to segment ET volumes with only a few annotated pixels from one slice. The structure of the hijacked network is slightly modified to take only a few pixels of a slice as a query at the inference step and does not require any training. To the best of our knowledge, this is the first time this type of general-purpose pre-trained video object segmentation network has been used to segment ET images.

Our main contributions are:

- A new semi-supervised volume segmentation method reusing a pre-trained object segmentation STM network without additional training.
- A mask-oriented memory readout module to provide a partially segmented query image at the inference step.
- A detailed experimentation on several actual ET data showing that an accurate segmentation is possible with only one slice and a very small portion of annotated pixels in this slice.

2 Related works

Electron tomography segmentation Segmentation of tomograms remains challenging because of reconstruction artifacts and low signal-to-noise ratio. Manual segmentation is still the preferred method [9] with the support of visualization tools [13] and various image processing methods such as watershed transform [26]. DL-based methods have been applied in electron tomography in more recent work [1,14], with DL models performing better in general semantic segmentation tasks [22,5,19,2,24]. The main bottleneck for ET segmentation tasks is the low availability of labeled training data. Recent works addressed the issue with either a semi-supervised setup with contrastive learning [15] or a scalable DL model, which only requires small- and medium-sized ground-truth datasets [14]. Our method goes further by repurposing a VOS model without any training phase.

Memory network Memory networks have an external module that can access past experiences [23]. Usually, an object in the memory can be referred to as a key feature vector and encoded by a value feature vector. Segmentation memory networks such as STM [20], SwiftNet [27], or STCN [4] encode the first video frame, annotated by the user, into the memory component. The next frame (query) is encoded into key and value feature vectors. The query keys and the memory are then compared, resulting in a query value feature vector used to segment the object on that frame. The memory component is then completed with the new key and value. This technique is often used in video segmentation as the object to segment, whose shapes change as time passes, is constantly added to the memory, providing several examples to help segmentation [20]. Unlike our approach, where only a fraction of the frame is needed, these methods require the annotation of the first entire frame.

Video and volumic interactive segmentation Memory networks are effective but require the segmentation of the entire first frame. An interaction loop can be added where the user is asked to segment the first frame with clicks or scribble to annotate the object of interest to the network [3]. The user corrects the result until they are satisfied. Networks for interactive segmentation are standard semantic segmentation models trained with an image channel, a mask channel, and an interaction channel [17]. By combining the interactive network for the first frame and the memory network to propagate the mask, recent works produce a segmentation mask for the whole video with minimal input from the user [3]. Similar methods have been applied in volumic segmentation [30,31,16]. However, for complex porous networks imaged with ET, standard interactive methods struggle to segment correctly. Adapting an interactive model requires training data composed of many segmented volumes not available for ET. We propose an approach similar to interactive methods using a partially segmented slice. Our method does not require any prior training.

3 Proposed method

Our method uses the same model to reconstruct the partially annotated frame and the entire volume. The images and masks in the memory are stored as key and value feature vectors. The key encodes a visual representation of the object so that objects with similar keys have similar shapes and textures. The value contains information for the decoder on the segmentation. Our intuition is that if we disable areas containing unlabelled data in the memory, we can encode useful information to segment whole slices even with a small amount of labeled data.

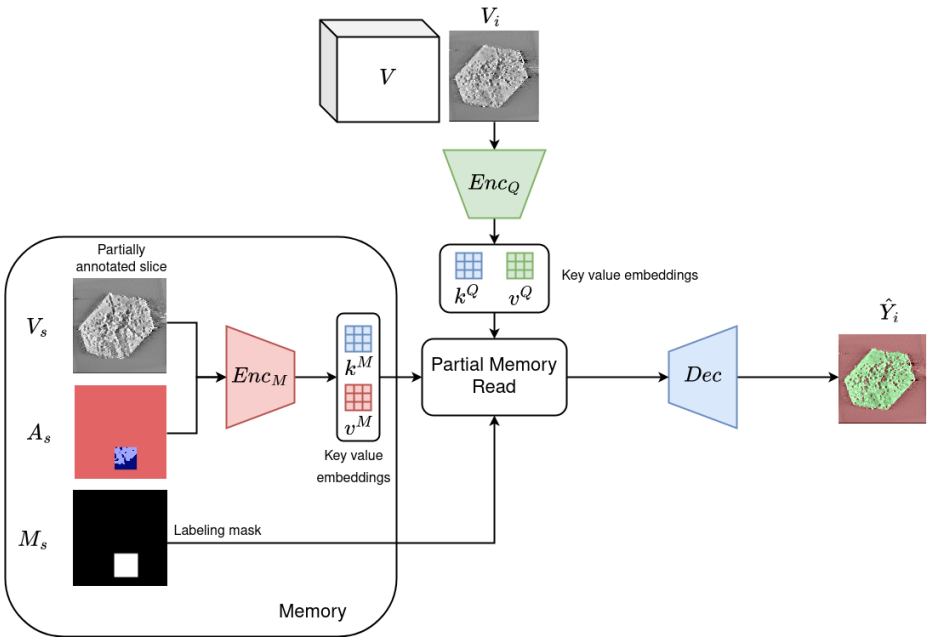

Fig. 2. The partially annotated slice is encoded by the memory encoder into the memory. At inference time, the query slice is encoded by the query encoder into a key and a value and compared with the memory keys and values. During the memory reading, the labeling mask selects only keys from labeled pixels, indicating whether a pixel is annotated. The result is given to the decoder, which reconstructs the whole segmentation.

In our framework, we ask an expert to annotate a small part of a slice A_s of the volume V to get the segmentation $\hat{Y}$ of the entire volume. From the annotations given by the expert, a labeling mask M_s is built where labeled and unlabeled pixels are denoted. The image and the annotations are encoded to one key and one value $\{k^M, v^M\}$ stored in the memory. There are two ways to

propagate the annotation into the entire volume. In the first one, the memory is directly used to segment other slices (Algorithm 1 and Figure 2). On the other hand, we pass the same image V_s as a query into the network to get pseudo-labels of the entire slice $\hat{Y}_s$. The entire slice and the newly acquired pseudo-label mask are then encoded into a key and a value $\{k^S, v^S\}$ to segment other slices $V_i, i \in [1, N]$ with N the number of slices in the volume (Algorithm 2).

During the memory read, the key in memory is modified to mask unknown zones to get the value to segment the whole slice. We use an STM network [20] as the backbone of our method, as it was the first network to use memory networks for 2D semantic segmentation. Moreover, as other methods in the field are based on the STM network, our approach is generalizable to other networks.

Algorithm 1 Procedure for segmenting the whole volume with a portion of a single slice annotated, with only the annotated bit in the memory.

Input: V, s, A_s, M_s, N
Output: $\hat{Y}$
  $\{k^M, v^M\} \leftarrow Enc_M(V_s, A_s)$
  while $i \in \{1, ..., N\}$ **do**
    $\{k^Q, v^Q\} \leftarrow Enc_Q(V_i)$
    $f \leftarrow PartialMemoryRead(M_s, k^M, k^Q, v^M, v^Q)$
    $\hat{Y}_i \leftarrow Decoder(V_i, f)$
  end while

Algorithm 2 Procedure for segmenting the whole volume with a portion of a single slice annotated, with pseudo-labels of the entire slice in the memory.

Input: V, s, A_s, M_s, N
Output: $\hat{Y}$
  $\{k^M, v^M\} \leftarrow Enc_M(V_s, A_s)$ ▷ First slice reconstruction
  $\{k^Q, v^Q\} \leftarrow Enc_Q(V_s)$
  $f \leftarrow PartialMemoryRead(M_s, k^M, k^Q, v^M, v^Q)$
  $\hat{Y}_s \leftarrow Decoder(V_s, f)$
  $\{k^S, v^S\} \leftarrow Enc_M(V_s, \hat{Y}_s)$
  while $i \in \{1, ..., s-1\} \cup \{s+1, ..., N\}$ **do** ▷ Whole volume propation
    $\{k^Q, v^Q\} \leftarrow Enc_Q(V_i)$
    $f \leftarrow MemoryRead(k^S, k^Q, v^S, v^Q)$
    $\hat{Y}_i \leftarrow Decoder(V_i, f)$
  end while

3.1 Key value embedding

The memory and the query encodings are slightly different. We consider, for a set of image $I \in \mathbb{R}^{H \times W}$ and its annotation $A \in \mathbb{R}^{H \times W}$, the memory encoder Enc_M

composed of a backbone network and followed by two parallel convolutional layers, outputting a memory key $k^M \in \mathbb{R}^{H/8\times W/8\times C/8}$ and a memory value $v^M \in \mathbb{R}^{H/8\times W/8\times C/2}$ such as:

$$Enc_M(I, A) = \{k^M, v^M\} \tag{1}$$

where W and H are the image size and C is the number of dimensions of the feature vectors at the output of the backbone network.

The query encoder shares the same architecture with different weights, but since the mask of the image is not available, only the slice passes through the query encoder Enc_Q:

$$Enc_Q(I) = \{k^Q, v^Q\} \tag{2}$$

with $k^Q \in \mathbb{R}^{H/8\times W/8\times C/8}$ the query key and $v^Q \in \mathbb{R}^{H/8\times W/8\times C/2}$ the query value.

3.2 Partial memory read

The standard method for memory reading from the STM network is modified to account for partially annotated slices. Let $M \in \mathbb{R}^{H\times W}$ be the annotation mask:

$$M_{i,j} = \begin{cases} 0 & \text{if } A_{i,j} \text{ is unannotated} \\ 1 & \text{if } A_{i,j} \text{ is annotated} \end{cases} \tag{3}$$

The mask M is then downsampled to be applied directly to the memory key:

$$M^D = Downsample(M, 8) \in \mathbb{R}^{H/8\times W/8} \tag{4}$$

A bilinear interpolation is used, which smoothes the boundaries. Next, the similarity map $S \in \mathbb{R}^{HW/16\times HW/16}$ is computed between a reshaped memory $k_r^M \in \mathbb{R}^{HW/16\times C/8}$ and a query $k_r^Q \in \mathbb{R}^{HW/16\times C/8}$ keys:

$$S = k_r^Q \times {k_r^M}^T \tag{5}$$

where $\times$ is the matrix product. A softmax is then applied to S. However, we mask S with M^D to cancel the unannotated areas' memory key's contribution k^M. Since S is the matrix product of k^Q and k^M, to properly mask k^M, M^D is reshaped into one dimension $M^L \in \mathbb{R}^{HW/16}$. We then multiply each row of S with M^L:

$$R_{i,j} = \frac{\exp(S_{i,j}) M_i^L}{\sum_{k,l} \exp(S_{k,l}) M_k^L} \tag{6}$$

The resulting matrix $R \in \mathbb{R}^{HW/16 \times HW/16}$ is the similarity between each zone of k^M and k^Q without the contribution of unannotated zones. The segmentation key $f \in \mathbb{R}^{H/8 \times W/8 \times C}$ is obtained by concatenating the query value and the memory value, weighted by R.

$$f = [v^Q, R \times v^M] \tag{7}$$

where $\times$ denotes the matrix product.

4 Experiments and results

4.1 Implementation details

We use the STM network architecture proposed in [20] as well as the weights proposed by the authors. The network's backbone is a ResNet, trained for a video segmentation task with Youtube-VOS and DAVIS as training datasets. The decoder outputs a probability map that is 1/4 of the initial input size, which degrades the results for ET where fine details on porous areas are needed. An upsampling operation is performed on the input slice before entering the network. The input slice is upscaled two times as a compromise between memory consumption and finer details.

All the results are computed with Intersection Over Union (IOU) on the entire volume V:

$$IOU(V) = \frac{\sum_{i=1}^{N} \hat{Y}_i \cap Y_i}{\sum_{i=1}^{N} \hat{Y}_i \cup Y_i} \tag{8}$$

with $\hat{Y}$ the segmented volume and Y the ground truth. The closer the IOU is to 1, the better the segmentation is.

4.2 Data

Chemical processes in the energy field often require using zeolites [10]. However, the numerous nanometric scale cavities make zeolites complex to segment. We evaluated our method on three volumes of hierarchical zeolites, NaX Siliporite G5 from Ceca-Arkema [18]. Volumes' sizes are 592×600×623, 512×512×52, and 520×512×24.

The slices are automatically partially annotated to simulate real-world data. A rectangle window of the area A_w is considered labeled. The remaining slice is unlabelled. The center of the window is drawn randomly on the border between the object and the background (Figure 3). The window is adjusted to fit entirely on the screen while maintaining its area. We define the labeling rate as $r = \frac{A_w}{H \times W}$.

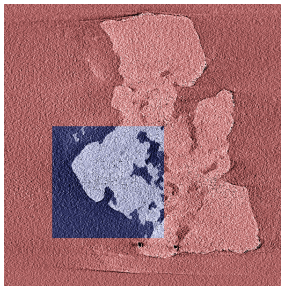

Fig. 3. A partially annotated slice of a zeolite. A window of an area A_w is considered to be annotated, while the pixel label on the outside of the window is unknown. The center of the window is randomly selected near the border between the object and the background to include pixels from both classes.

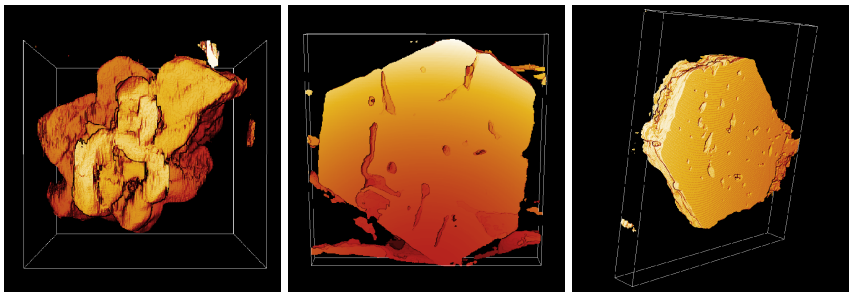

Fig. 4. 3D visualization of segmented volumes of hierarchical zeolites NaX Siliporite G5. A random window of 6% of one slice has been annotated. Segmentations are provided using our approach (Algorithm 1).

4.3 Results

For each volume, we run each experiment on the same five randomly selected slices. The mean IOU of the three volumes is reported.

Comparaison with the STM network We first compare our method with an unmodified STM network for several labeling rates r. We give the same partially annotated slices for the STM network and our method. The results are shown in Table 1. The STM network can not process the partially labeled slice because it was not intended to deal with such data. As a result, there is no way for the STM network to differentiate labeled and unlabeled pixels, which leads to poor segmentation. Our key masking allows the STM network to achieve significantly better results with accurate segmentation (Figure 4).

Table 1. Mean IOU on our volumes for our method and an unmodified STM network. The modification from our approach allows an STM-like model to produce good segmentation.

r	0.02	0.06	0.12	0.18	0.25
Ours	**0.551**	**0.625**	**0.693**	**0.725**	**0.758**
STM	0.013	0.024	0.070	0.150	0.285

First slice propagation We then study the different approaches for the first slice. We tested our method with only the annotated parts in the memory shown (Ours) in the Algorithm 1 against our method with the pseudo-labels of the entire first slice in the memory (Ours+F) described by the Algorithm 2. The results in Table 2 demonstrate that better results are obtained with only the annotated bit in the memory. The STM network performs better with accurate data instead of more variety in memory. Our method's implementation only uses the partially labeled parts in the memory.

Table 2. Mean IOU on our volumes for our method with only the labeled parts in the memory (Ours) and our method with the pseudo-labels of the first slice in the memory (Ours+F). Our approach uses only the annotated zones in the memory.

r	0.02	0.06	0.12	0.18	0.25
Ours	0.550	**0.625**	**0.693**	**0.725**	**0.758**
Ours+F	**0.564**	0.588	0.650	0.671	0.710

Comparaison with other methods Finally, we compare our methods with other approaches that can handle partially labeled slices [15]. We study the performance of our method against a UNET with a weighted cross entropy to train only on labeled zones and a contrastive UNET that uses contrastive learning to exploit unlabeled areas. Both methods require a training phase. The results are shown in Table 3. Our method performs better than a standard UNET but still lags behind the contrastive UNET. Nevertheless, our method shows promising results with scores close to methods with a training procedure. Figure 5 shows all the results from the experiments previously mentioned.

5 Conclusion

In this paper, we illustrate that a slightly modified STM network handles accurate volumetric segmentation of 3D scans from ET with only a tiny portion of one slice labeled needed without any further fine-tuning. This approach achieves results close to methods that require a training procedure. The masking of the

Table 3. Mean IOU on our volumes for our method, a UNET adapted for partially segmented areas, and a UNET using a contrastive loss to exploit both labeled and unlabeled zones. Our proposed method achieves results close to these methods despite no training phase.

r	0.02	0.06	0.12	0.18	0.25
Ours	0.551	0.625	0.693	0.725	0.758
UNET	0.544	0.600	0.671	0.695	0.839
Contrastive UNET	0.737	0.768	0.793	0.813	0.815

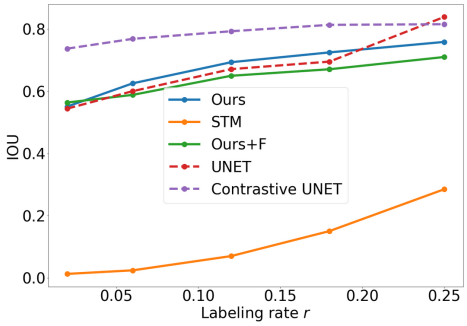

Fig. 5. Mean IOU for several labeling rates r. All methods do not need an additional training procedure except UNET and contrastive UNET [15].

memory shows that semi-labeled slices can be used to propagate accurate segmentation in fields where annotated data are not widely available. A more detailed segmentation mask can be obtained with further investigations, as the original STM network output size is 1/4 of the original size.

Acknowledgements This work was supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX- 0007) operated by the French National Research Agency (ANR).

References

1. Akers, S., Kautz, E., Trevino-Gavito, A., Olszta, M., Matthews, B.E., Wang, L., Du, Y., Spurgeon, S.R.: Rapid and flexible segmentation of electron microscopy data using few-shot machine learning. npj Computational Materials **7**(1), 1–9 (2021)
2. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(4), 834–848 (2018)

3. Cheng, H.K., Tai, Y.W., Tang, C.K.: Modular interactive video object segmentation: Interaction-to-mask, propagation and difference-aware fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5559–5568 (2021)
4. Cheng, H.K., Tai, Y.W., Tang, C.K.: Rethinking space-time networks with improved memory coverage for efficient video object segmentation. Advances in Neural Information Processing Systems **34**, 11781–11794 (2021)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016 pp. 424–432 (2016)
6. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. Advances in neural information processing systems **27** (2014)
7. Ersen, O., Hirlimann, C., Drillon, M., Werckmann, J., Tihay, F., Pham-Huu, C., Crucifix, C., Schultz, P.: 3D-TEM characterization of nanometric objects. Solid State Sciences **9**(12), 1088–1098 (2007)
8. Evin, B., Leroy, É., Baaziz, W., Segard, M., Paul-Boncour, V., Challet, S., Fabre, A., Thiébaut, S., Latroche, M., Ersen, O.: 3D Analysis of Helium-3 Nanobubbles in Palladium Aged under Tritium by Electron Tomography. Journal of Physical Chemistry C **125**(46), 25404–25409 (2021)
9. Fernandez, J.J.: Computational methods for electron tomography. Micron **43**(10), 1010–1030 (2012)
10. Flores, C., Zholobenko, V.L., Gu, B., Batalha, N., Valtchev, V., Baaziz, W., Ersen, O., Marcilio, N.R., Ordomsky, V., Khodakov, A.Y.: Versatile Roles of Metal Species in Carbon Nanotube Templates for the Synthesis of Metal–Zeolite Nanocomposite Catalysts. ACS Applied Nano Materials **2**(7), 4507–4517 (Jun 2019)
11. Genc, A., Kovarik, L., Fraser, H.L.: A deep learning approach for semantic segmentation of unbalanced data in electron tomography of catalytic materials. arXiv preprint arXiv:2201.07342 (2022)
12. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. 2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06) **2**, 1735–1742 (2006)
13. He, W., Ladinsky, M.S., Huey-Tubman, K.E., Jensen, G.J., McIntosh, J.R., Björkman, P.J.: FcRn-mediated antibody transport across epithelial cells revealed by electron tomography. nature **455**(7212), 542–546 (2008)
14. Khadangi, A., Boudier, T., Rajagopal, V.: EM-net: Deep learning for electron microscopy image segmentation. 2020 25th International Conference on Pattern Recognition (ICPR) pp. 31–38 (2021)
15. Li, C., Ducottet, C., Desroziers, S., Moreaud, M.: Toward few pixel annotations for 3D segmentation of material from electron tomography. In: International Conference on Computer Vision Theory and Applications, VISAPP 2023 (2023)
16. Liu, Q., Xu, Z., Jiao, Y., Niethammer, M.: iSegFormer: Interactive segmentation via transformers with application to 3D knee MR images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V. pp. 464–474. Springer (2022)
17. Mahadevan, S., Voigtlaender, P., Leibe, B.: Iteratively trained interactive segmentation. In: British Machine Vision Conference (BMVC) (2018)

18. Medeiros-Costa, I.C., Laroche, C., Pérez-Pellitero, J., Coasne, B.: Characterization of hierarchical zeolites: Combining adsorption/intrusion, electron microscopy, diffraction and spectroscopic techniques. Microporous and Mesoporous Materials **287**, 167–176 (2019)
19. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV) pp. 565–571 (2016)
20. Oh, S.W., Lee, J.Y., Xu, N., Kim, S.J.: Video object segmentation using space-time memory networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9226–9235 (2019)
21. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on knowledge and data engineering **22**(10), 1345–1359 (2009)
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 pp. 234–241 (2015)
23. Sukhbaatar, S., Weston, J., Fergus, R., et al.: End-to-end memory networks. Advances in neural information processing systems **28** (2015)
24. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 5693–5703 (2019)
25. Tran, V.D., Moreaud, M., Thiébaut, É., Denis, L., Becker, J.M.: Inverse problem approach for the alignment of electron tomographic series. Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles **69**(2), 279–291 (2014)
26. Volkmann, N.: Methods for segmentation and interpretation of electron tomographic reconstructions. Methods in enzymology **483**, 31–46 (2010)
27. Wang, H., Jiang, X., Ren, H., Hu, Y., Bai, S.: Swiftnet: Real-time video object segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1296–1305 (2021)
28. Wurm, M., Stark, T., Zhu, X.X., Weigand, M., Taubenböck, H.: Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing **150**, 59–69 (2019)
29. Zhao, X., Vemulapalli, R., Mansfield, P.A., Gong, B., Green, B., Shapira, L., Wu, Y.: Contrastive learning for label efficient semantic segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision pp. 10623–10633 (2021)
30. Zhou, T., Li, L., Bredell, G., Li, J., Konukoglu, E.: Quality-aware memory network for interactive volumetric image segmentation. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 560–570. Springer International Publishing, Cham (2021)
31. Zhou, T., Li, L., Bredell, G., Li, J., Unkelbach, J., Konukoglu, E.: Volumetric memory network for interactive medical image segmentation. Medical Image Analysis **83**, 102599 (2023)