Document Version
Final published version

Citation (APA)
Aswath, A., Alsahaf, A. M. J., Duinkerken, B. H. P., Hoogenboom, J. P., Giepmans, B. N. G., & Azzopardi, G. (2025). ColorEM-Net: Automated Segmentation of Structures in Large-Scale Electron Microscopy Using Element-Derived Ground Truth. In M. Castrillón-Santana, C. M. Travieso-González, D. Freire-Obregón, D. Hernández-Sosa, J. Lorenzo-Navarro, O. J. Santana, & O. Deniz Suarez (Eds.), *Computer Analysis of Images and Patterns: Proceedings of the 21st International Conference, CAIP 2025* (pp. 220-231). (Lecture Notes in Computer Science; Vol. 15621 LNCS). Springer. https://doi.org/10.1007/978-3-032-04968-1_19

ColorEM-Net: Automated Segmentation of Structures in Large-Scale Electron Microscopy Using Element-Derived Ground Truth

Anusha Aswath[1,2](✉), Ahmad M.J. Alsahaf[2], B. H. Peter Duinkerken[2], Jacob P. Hoogenboom[3], Ben N.G. Giepmans[2], and George Azzopardi[1]

[1] Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen,Groningen, The Netherlands

[2] Department of Biomedical Sciences, University Medical Center Groningen, Groningen, The Netherlands

anusha.aswath@gmail.com

[3] Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands

Abstract. Electron microscopy (EM) combined with energy dispersive x-ray (EDX) imaging (or 'ColorEM') of cells and tissues provides ultrastructural insight complemented with elemental context. The resulting hyperspectral datasets can be used to map the relative abundance of specific elements or subjected to more data-driven approaches such as spectral mixture analysis or clustering to highlight the ultrastructural components of interest. Despite the benefits of automatic segmentation over manual annotation, EDX imaging is two orders of magnitude slower than EM imaging precluding its routine use for segmentation. Large-scale ColorEM, however, does generate sufficient annotated labels, which we use as ground truth to train U-Net models, and thus enables the transfer of these labels to conventional EM data. Here, we present ColorEM-Net, a label-free segmentation technique based on features obtained from unsupervised clustering of ColorEM data. ColorEM-Net achieves label-free identification with over 95% accuracy for nuclei, lysosomes and exocrine granules. However, with an accuracy of 79%, the recognition of endocrine granules needs further effort in training for reliable segmentation. By reusing open-access ColorEM datasets, this approach facilitates automated segmentation of EM data, while eliminating the need for manual annotation and achieving scalability for tissue-scale segmentation.

Keywords: Segmentation · electron microscopy · analytical pixel labels

1 Introduction

Understanding how the building blocks of life are arranged is crucial for biomedical research. The study of biomolecules, organelles and cells at the highest

M. Castrillón-Santana et al. (Eds.): CAIP 2025, LNCS 15621, pp. 220–231, 2026.
https://doi.org/10.1007/978-3-032-04968-1_19

resolution can be done by electron microscopy (EM), which enables cellular ultrastructure to be inspected with nanometer resolution. Visualization of the cellular ultrastructure with relevant biological labels is a powerful approach to understand health and disease. Unbiased imaging of large two-dimensional sections at tissue scale (Nano-anatomy or nanotomy) is nowadays routine, providing datasets similar to zoomable Google Earth images [7]. Moreover, with recent technological developments in faster acquisition and higher throughput [10,12,17,26] a data avalanche is foreseen. Hence, automation of data analysis is crucial to maximize the biological insight gathered from such gigapixel datasets.

Advanced features learned from labeled images via deep learning (DL) are being explored and implemented to advance EM analysis [3], building on traditional methods. However, identifying subcellular structures require DL models to capture diverse morphologies and sizes, typically demanding manual labeling of many grayscale images. Therefore, complementary methods such as ColorEM [22,25] and correlated microscopy, i.e. using fluorescent labels projected on EM maps [6] are used to gain information of biostructures in the grayscale data. The former yields hyperspectral images (HSIs) based on the elemental composition, allowing data-driven and more extensive analysis methods which have been widely used in material sciences and remote sensing for classification, segmentation and anomaly detection [2]. ColorEM includes analytical methods in EM, such as energy dispersive X-ray (EDX; [22]) imaging, that bring the benefits of HSI analysis to cellular biology, combining spectral information with spatial context. Recently, large-scale ColorEM was introduced to aid in the unbiased identification of sub-cellular content through spectral unmixing [11] as well as user-guided iterative clustering approach [4]. Like the endmembers in spectral analysis [11], the resulting cluster masks also correspond to specific biological structures.

EDX acquisition is ~100 times slower than conventional EM, which makes it impractical for its routine implementation for the analysis of large-scale EM datasets. Here, we present an automated segmentation method that learns structures in large-scale EM datasets using supervised models trained on labels derived from ColorEM datasets.

2 Methods

Figure 1 shows the ColorEM-Net pipeline, outlining the key stages from hyperspectral data processing to EM segmentation, using examples such as heterochromatin and two types of endocrine granules: insulin and glucagon. In our main experiments, we focus on four representative pancreatic structures: heterochromatin in the nucleus, lysosomes, and both endocrine and exocrine granules. Cluster masks derived from hyperspectral ColorEM data serve as pseudo-labels to enable label-free segmentation of large-scale EM datasets, following the CLEM-Net-inspired methodology [1].

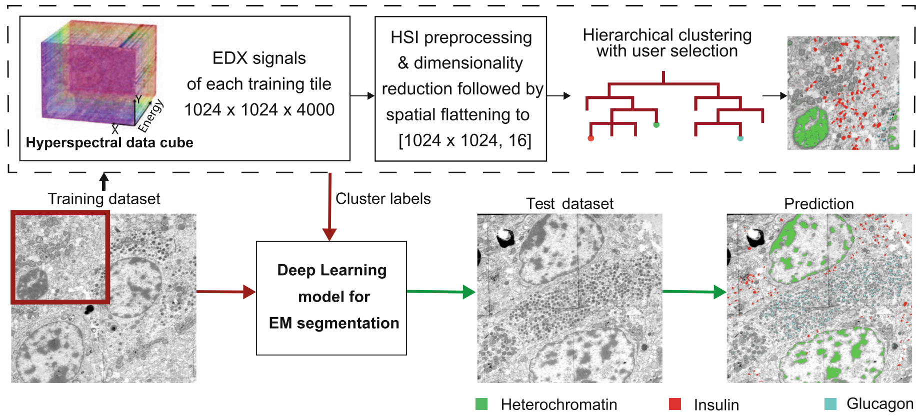


Fig. 1. Overview of the ColorEM-Net pipeline for label-free EM segmentation using EDX-derived pseudo-labels. A hyperspectral data cube ($1024 \times 1024 \times 4000$), shown for a single tile highlighted by a red box in the training dataset, is acquired via energy-dispersive X-ray (EDX) imaging capturing elemental composition at nanometer resolution. After preprocessing and dimensionality reduction, the data is flattened and clustered in a hierarchical way. The resulting cluster masks, each corresponding to a specific biological structure (e.g., insulin, glucagon, heterochromatin), are manually selected and used as pseudo-labels. These masks supervise the training of a deep learning segmentation model using a corresponding grayscale EM image (red path). Once trained, the model is applied to EM-only images without labels (green path), enabling automated segmentation across large-scale datasets. The hyperspectral cube is reproduced from [4]. (Color figure online)

2.1 Segmentation Model

ColorEM-Net uses a U-Net architecture with a ResNet-50 backbone (pre-trained on ImageNet) for enhanced feature extraction [23]. To adapt the pre-trained ResNet model to grayscale images, the first convolutional layer is modified for accepting a single channel by averaging the original multi-channel weights. The input training masks are encoded with zero for the background and one for the foreground in case of binary masks, and values greater than zero (up to one) for the foreground enabling the model to learn confidence-weighted scores for segmenting the target structure from the EM image. Training on large EM images is performed by dividing the image into smaller patches and using data augmentation to increase image diversity for training (Fig. 2).

2.2 Loss Functions

In binary segmentation tasks, addressing class imbalance is a critical challenge, particularly when foreground and background pixels differ significantly in the number of training samples. The Dice loss, which is one minus the Dice similarity coefficient, is widely used to mitigate this issue, especially in EM segmentation tasks [21]. To improve robustness, we employ a combined loss function incorporating both binary cross entropy (BCE) and Dice loss:

$$\text{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i), \tag{1}$$

$$\text{Dice loss} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}, \tag{2}$$

$$\text{Combined loss} = \gamma \cdot \text{BCE} + (1 - \gamma) \cdot \text{Dice loss}, \tag{3}$$

where y_i is the ground truth label, $\hat{y}_i$ is the predicted probability for pixel i, and N is the total number of pixels. A small constant ϵ ensures numerical stability; in our experiments, we set $\gamma = 0.5$ and $\epsilon = 10^{-5}$.

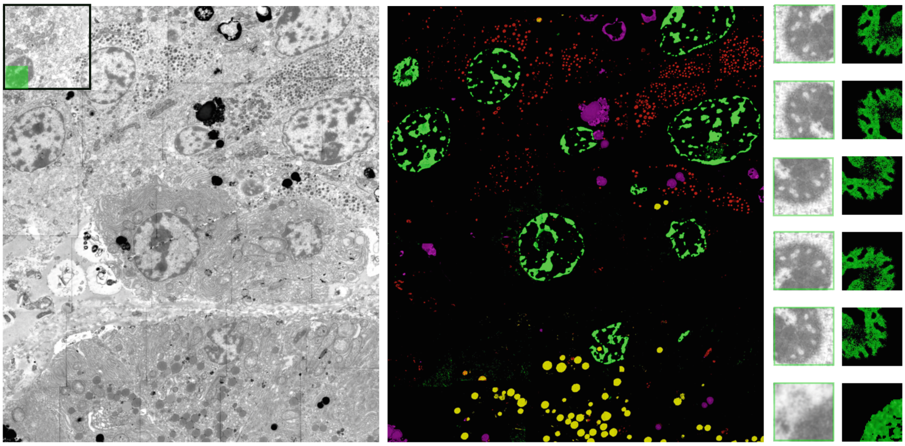

(a) (b) (c)

Fig. 2. ColorEM-Net using EDX-derived pseudo-labels for training binary segmentation models enable prediction on EM-only images. (a) The EM dataset used for training is shown as a stitched grayscale image arranged in a 5-column by 6-row grid following preprocessing and inversion. **(b)** The cluster masks from [4] serve as pseudo-labels for heterochromatin (green), lysosomes (pink), exocrine granules (yellow), and endocrine granules (red), corresponding to the grayscale dataset. The training set is derived from selected grayscale image tiles containing the target structure (e.g., heterochromatin, shown in the black box of the first tile). **(c)** Each tile is divided into smaller patches for efficient processing. An example patch, highlighted in green, undergoes various rigid transformations along with its cluster mask for five augmented variations. (Color figure online)

While effective against class imbalance, this combined loss is susceptible to noisy pseudo-labels: mislabeled samples contribute equally to the loss, potentially degrading performance. To address this, we incorporate pixel-level confidence scores derived from clustering results. Each pixel's confidence C_i is computed as $C_i = 1 - (d_i/mx)$, where d_i is the distance to its cluster centroid and mx is the maximum distance in the cluster. This gives higher weight to samples closer to the centroid. To incorporate these confidence scores, we use them as weights in the weighted BCE loss [9]:

$$\text{weighted BCE} = -\frac{1}{N}\sum_{i=1}^{N} C_i\big[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\big] \tag{4}$$

Finally, this confidence-weighted BCE loss is incorporated into the previously defined combined loss function (Eq. 3). We refer to the combined version using this weighted BCE as the *confidence loss.*

To improve the robustness of pseudo-labels for confidence scoring, we apply hierarchical clustering (Fig. 1). This method captures structural variability by identifying nested substructures, and filtering out high-dimensional noise or outliers at higher cluster levels. Hence, the interactive clustering algorithm is suitable for segmentation tasks with ambiguous boundaries or varying object sizes [27], even when spatial information is flattened.

Since clustering is inherently ill-posed, both model choice and uncertainty characterization are crucial [5]. For instance, [20] shows how confidence calibration via Monte Carlo dropout or deep ensembles can support robust loss weighting or using uncertainty from variance or entropy of predictions [13].

3 Experiments and Results

3.1 Datasets and Image Processing

The pseudo-labels are derived from cluster masks, reusing an open access ColorEM dataset [11] (Fig. 1). The transfer was performed on a separate dataset from the same tissue, but a different section. This dataset consists solely of grayscale EM images and is four times larger than the training dataset, comprising 120 tiles, each of 2000×2000 pixels at a resolution of 4nm. The predictions are included only for the first four rows (12 columns each) and the bottom four rows (also 12 columns each). The remaining 24 tiles were excluded due to section folds and four out of focus tiles.

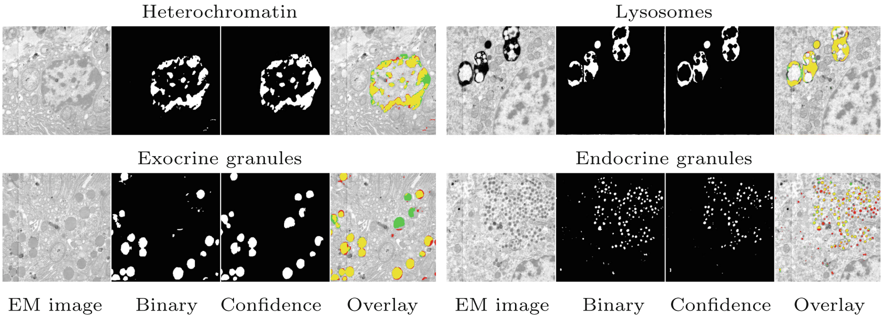
The EM images are acquired as high-angle annular dark-field (HAADF) images that contain overlapping areas for stitching of the images into a single large dataset using TrakEM2 plugin in ImageJ [24]. Although the granule content is relatively uniform and theoretically possible to annotate, manual quantification remains difficult due to discrepancies between expert annotations leading to incomplete data. From the clustering method that was applied on the ColorEM dataset [4], the obtained cluster masks enable the automatic identification of 198 endocrine granules, 54 exocrine granules, and 10 heterochromatin regions (classified as nucleus), which are used as pseudo-labels for learning EM image segmentation models. The spectrally distinct lysosomes were additionally included as training masks.

First, pre-processing is performed by subjecting the datasets to 2×2 binning [4] followed by global normalization within each dataset using min-max scaling. Here the 1st and 99th percentiles are used as the lower and upper bounds, respectively. Values below the 1st percentile and those above the 99th percentile are clipped to remove artifacts, effectively scaling the data in the range [0, 1]. By minimizing the impact of outliers, this step improves the uniformity of intensity distributions across each dataset. Following normalization, all images were inverted to align with the conventional grayscale appearance of EM images, where darker regions correspond to denser materials. Second, image distribution shift is addressed between the datasets (ColorEM and EM-only datasets) by mean-centering to avoid variations in image brightness.

3.2 Training and Inference

To ensure robust model performance, we divided the 30 tiles from the ColorEM dataset into three distinct sets: training (20 tiles), validation (5 tiles), and test (5 tiles). Each set was chosen in such a way that there is sufficient representation of the target structure. Firstly, the labeled tiles for each structure are assigned as 67% for training, 13% for validation, and 13% for testing. Then, the remaining tiles (unlabeled) are allocated randomly to each of the sets to fill the remaining tiles needed for the train-validation-test split.

To facilitate efficient processing of large images, each individual tile was subdivided into patches of 256 × 256 pixels. Constant padding (with a value of zero) was applied to the right and bottom edges to ensure that no partial patches were formed at the image borders. A sliding window with an overlap of 50% was used to extract patches. To enhance image diversity, five types of data augmentations were separately applied; horizontal and vertical flips, random rotations, transpositions, and crops [19].



Heterochromatin

Lysosomes

Exocrine granules

Endocrine granules

Fig. 3. Qualitative results on EM-only images showing the effective transfer of EDX-derived cluster masks as ground truth for training. Each EM image (1000 × 1000 pixels) shows predictions from two methods: (i) binary segmentation masks trained with the combined loss, and (ii) masks obtained using confidence-weighted BCE loss in the combined loss, as an overlay indicating differences between the two methods. Predictions from the confidence-weighted method (green) show improved localization of heterochromatin, lysosomes, and exocrine granules, while binary masks (red) perform better for endocrine granules but exhibit a higher rate of false positives. (Color figure online)

Training was conducted for up to a maximum of 100 epochs, with each epoch iterating through all training patches in batches of size 32. For the encoder, a pre-trained ResNet backbone initialized with ImageNet weights was used, which leveraged transfer learning to provide a strong starting point for feature extraction. The decoder, on the other hand, was trained from scratch. Optimization was performed using the Adam optimizer with an initial learning rate of 10^{-3}. To prevent overfitting and to refine learning, the learning rate was reduced by a factor of 10 if validation performance did not improve for 20 consecutive epochs. Additionally, model weights were saved whenever validation performance improved by more than 10^{-3}. The best model based on validation performance was retained, which could be restored after training for testing or deployment.

For inference, the input image was divided into overlapping patches (50%). Each patch was processed by a neural network that generated a confidence map representing the likelihood of each pixel belonging to the target structure. Overlapping regions were averaged to produce a smooth, artifact-free prediction. The final output was obtained by thresholding pixels with confidence above zero as foreground; others were background.

3.3 Evaluation Protocol

The EM-only images lack labels of ground truth, as cluster masks are only available for the EDX dataset. The evaluation aims to address two key aspects: first, the comparative performance of two training approaches – binary and confidence mask-based on segmentation metrics, and second the transferability of labels from ColorEM to the EM-only dataset.

Models trained with cluster masks are evaluated for each target structure in their respective test sets, which were set aside during training. While the cluster masks were used as-is during training, for evaluation they were manually corrected by removing evident false positives for a fair assessment. The evaluation metrics include precision, recall and region-based scores, namely Dice Similarity Coefficient (DSC) and Intersection over Union (IoU).

To evaluate the model's performance on the EM-only dataset, for which no EDX-based ground-truth is available, a proxy of ground-truth was synthesized by three separate field experts using ImageJ's multi-point tool through which the individual structures were identified. Subsequently, a separate evaluator compared the model's predictions to the identified structures and the number of true positives is reported upon.

3.4 Results

ColorEM-Net achieves image segmentation on the EM-only dataset by training binary segmentation models using cluster masks derived from EDX-based features for four selected target structures: heterochromatin, lysosomes, exocrine granules, and endocrine granules.

Segmentation performance on the EM-only dataset was evaluated using the pseudo-ground truth labels and model predictions based on the binary and

Table 1. Segmentation performance metrics comparing model predictions using binary-trained and confidence-weighted models. Evaluation metrics include Precision (P), Recall (R), Dice Similarity Coefficient (DSC), and Intersection over Union (IoU) across four cellular structures. The results are obtained from the ColorEM dataset where pseudo-ground truth is available.

Structure	Method	P	R	DSC	IoU
Heterochromatin	Binary/Confidence	0.78/0.82	0.80/0.82	0.79/0.81	0.65/0.69
Lysosomes	Binary/Confidence	0.71/0.75	0.86/0.85	0.78/0.80	0.64/0.67
Exocrine granules	Binary/Confidence	0.84/0.90	0.92/0.87	0.86/0.88	0.78/0.79
Endocrine granules	Binary/Confidence	0.73/0.73	0.98/0.94	0.83/0.82	0.71/0.70

confidence-weighted loss (Fig. 3). The application of the confidence-weighted model yielded substantial improvements in precision and recall for most target structures. Specifically, heterochromatin segmentation shows consistent enhancements, with reduced false positives and false negatives, highlighting the model's ability to refine predictions under confidence-based constraints (Table 1).

Predictions on the unlabeled EM-only dataset for the four structures are shown as colored overlays on the 120-tile dataset in Fig. 4. The displayed predictions demonstrate the model's performance across diverse unlabeled EM tiles, highlighting the spatial distribution of all four cellular structures.

Counts of subcellular structures are derived from 10% of the images in the EM-only dataset (12 tiles). Heterochromatin (nucleus), lysosomes, and zymogen were predicted with over 95% accuracy, closely matching annotator counts (Fig. 5). Predictions for endocrine granules, particularly glucagon and polypeptide were highly concentrated in regions corresponding to these granules, consistent with results from the ColorEM dataset. Insulin was excluded due to a severe distribution shift in image intensity or contrast, making it non-transferable. Among the considered granules, severe class imbalance from an uneven distribution of samples or pixels between granules and the background led to a reduced count, accounting for only 79% of the segmentation results. Also, quantitative analysis using biological labels identified 19 nuclei, 49 lysosomes, 169 exocrine granules and 706 endocrine granules. Note that the endocrine counts here only reflect glucagon and polypeptide granules. Overall, the confidence-weighted model excelled in image quality but not in count accuracy for the endocrine granules.

4 Discussion

Label-free segmentation of large-scale EM images was achieved using EDX-derived cluster labels as pseudo-ground truth. Integrating analytical features from additional EM modalities (e.g. EDX) offers a promising approach for achieving precise pixel-level segmentation of critical image features. By adopting cross-

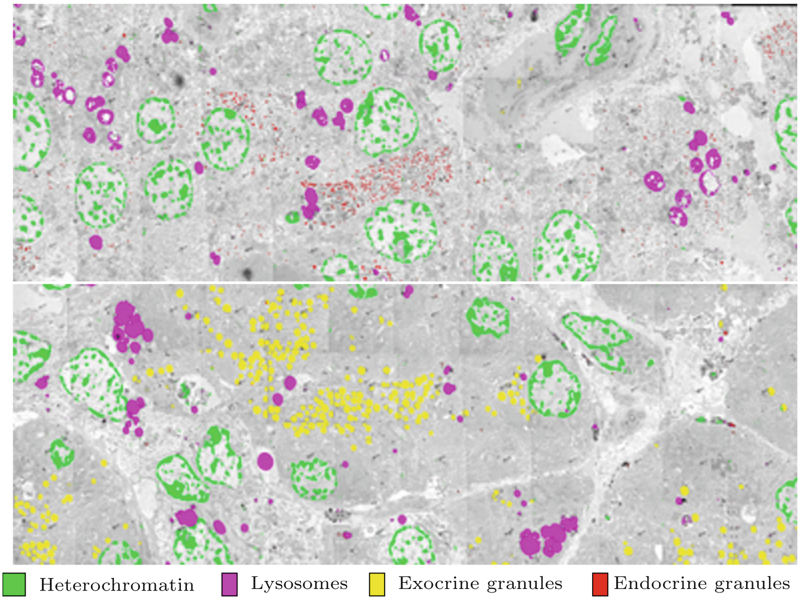


Fig. 4. Predictions of the trained binary models on the EM-only dataset. The EM-only dataset consists of 120 grayscale tiles without EDX signals, in contrast to the 30-tile ColorEM dataset which includes EDX. The white line indicates regions omitted due to artifacts like folds or out-of-focus tiles.

modal insights, similar distributions of image regions were extracted, which are useful to segment grayscale EM in a discriminative manner.

The performance of CNNs depends on the choice of patch size, which must be optimized for each specific structure [18]. For example, the small size of endocrine granules may require smaller patch dimensions to avoid severe class imbalances.

One limitation we observed was the insufficient transferability of insulin-related structures, likely due to strong contrast differences between training and target datasets. Future work could explore contrast enhancement through targeted augmentation. Additionally, while mean subtraction was used to correct brightness variations, more advanced normalization methods, such as histogram matching or class-specific corrections, may further improve robustness.

While our method focuses on spectral features derived from elemental composition, an important direction for future work involves integrating morphological context. The shift from object-based pseudo-labels to biological measurements presents opportunities for improved transferability. However, different ultrastructural components are not always differentiable based on their elemental composition. Integrating morphology- or shape-based encodings particularly for structures like mitochondria, the Golgi apparatus, or the nuclear envelope could improve segmentation, as these organelles exhibit greater variability in appearance and shape, often appearing as larger-scale patterns in images. Using

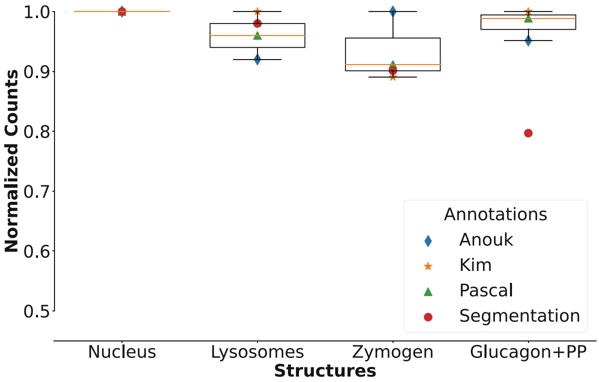


Fig. 5. Quantification of selected biostructures using expert annotations as ground truth and ColorEM-Net outputs as predictions show similar structural counts. The plot shows the variability among expert annotation and normalized counts (scaled by dividing by the maximum of all four counts) for each structure. The accuracy for heterochromatin (nucleus), lysosomes and zymogen (exocrine granules) is very close to all annotators, while that of endocrine granules for glucagon and polypeptide (PP) requires further improvement.

such morphological cues may enable more accurate segmentation of additional sub-cellular structures.

Additional EM and EDX data processed by the methods like those presented in [4, 11] could be used to generate pseudo labels that are refined and corrected by experts to produce proper groundtruth labels. This will allow the application of fully and semi-supervised methods, against which label-free method like ours could be consistently benchmarked.

The current confidence score based on pixel distance to cluster centroids showed important improvements in segmentation quality. To build on these results, future work could explore incorporating cluster-level features such as size or density to further capture structural complexity and biological variability.

Finally, a promising direction for improving unsupervised segmentation is the integration of contrastive and self-supervised learning. Including non-target ultrastructural components, even as unlabeled regions, may support contrastive frameworks that help the model distinguish between subtle classes [15]. Self-supervised methods like SimCLR [8], MoCo [16], and BYOL [14] have shown that robust, discriminative visual representations can be learned from unlabeled data. Applying these techniques to EM images could complement our clustering-based confidence scoring by better capturing fine-grained structural similarities and differences. Moreover, contrastive pretraining may improve generalization in settings with sparse or noisy pseudo-labels.

5 Conclusions

Label-free segmentation in large-scale EM images using EDX-derived cluster labels demonstrates a promising approach for precise pixel-level segmentation without requiring manual annotations. The method allows for the extraction of biologically relevant features (using pixel-level information) that traditional object-based labeling approaches often overlook. The integration of biological measurements and advanced staining techniques offers a promising direction for improving feature discrimination and enabling the segmentation of more complex organelles, leading to improved biological interpretation of EM datasets.

Acknowledgments. We thank the University of Groningen for providing funding through the Center for Data Science and Systems Complexity, and the Center for Information Technology for access to the Hábrók high-performance computing cluster. Part of this work was performed at the UMCG Microscopy and Imaging Center (UMIC), sponsored by the NWO National Roadmap for Large-Scale Research Infrastructure (NEMI; NWO 184.034.014, NL-BioImaging: NWO 184.036.012); NWO-ENPPS.LIFT.019.030; The sample shown involves reused data from tissue provided by the Network for Pancreatic Organ donors with Diabetes (nPOD; RRID:SCR 014641). We also thank Jeroen Kuipers, Kim Kats, Anouk Wolters and Pascal de Boer for their expert assistance and feedback.

References

1. Lane et al., R.: Organelle segmentation facilitated by correlative light microscopy data. Microsc. Microanal. (2022)
2. Amigo, J.M., Babamoradi, H., Elcoroaristizabal, S.: Hyperspectral image analysis. a tutorial. Analytica chimica acta **896**, pp.34–51 (2015)
3. Aswath, A., Alsahaf, A.M.J., Giepmans, B.N.G., Azzopardi, G.: Segmentation in large-scale cellular electron microscopy with deep learning: a literature survey. Med. image anal. p. 102920 (2023)
4. Aswath, A., Duinkerken, B.H.P., Giepmans, B.N.G., Azzopardi, G., Alsahaf, A.M.J.: Interactive segmentation of biostructures through hyperspectral electron microscopy. 14th WHISPERS Workshop, IEEE (2024)
5. Binder, D.A.: Bayesian cluster analysis. Biometrika **65**(1), 31–38 (1978)
6. de Boer, P., Hoogenboom, J.P., Giepmans, B.N.G.: Correlated light and electron microscopy: ultrastructure lights up! Nat. Methods **12**(6), 503–513 (2015)
7. de Boer, P., et al.: Large-scale electron microscopy database for human type 1 diabetes. Nat. Commun. **11**(1), 2475 (2020)
8. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning, pp. 1597–1607. PmLR (2020)
9. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2613–2622 (2021)
10. Collinson, L.M., et al.: Volume em: a quiet revolution takes shape. Nat. Methods **20**(6), 777–782 (2023)

11. Duinkerken, B.H.P., Alsahaf, A.M.J., Hoogenboom, J.P., Giepmans, B.N.G.: Automated analysis of ultrastructure through large-scale hyperspectral electron microscopy. npj Imaging **2**(1),pp. 1–9 (2024)
12. Duinkerken, B.P., et al.: Sample processing and benchmarking for multibeam optical scanning transmission electron microscopy. Microsc. Microanal. **31**(2), ozaf024 (2025)
13. Gomez, C., Drost, A., Roger, J.M.: Analysis of the uncertainties affecting predictions of clay contents from VNIR/SWIR hyperspectral data. Remote Sens. Environ. **156**, 58–70 (2015)
14. Grill, J.B., et al.: Bootstrap your own latent-a new approach to self-supervised learning. Adv. Neural. Inf. Process. Syst. **33**, 21271–21284 (2020)
15. Guo, L.Z., Zhang, Y.G., Wu, Z.F., Shao, J.J., Li, Y.F.: Robust semi-supervised learning when not all classes have labels. Adv. Neural. Inf. Process. Syst. **35**, 3305–3317 (2022)
16. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9729–9738 (2020)
17. Kievits, A.J., Lane, R., Carroll, E.C., Hoogenboom, J.P.: How innovations in methodology offer new prospects for volume electron microscopy. J. Microsc. **287**(3), 114–137 (2022)
18. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
19. Liu, J., Xu, D., Yang, W., Fan, M., Huang, H.: Benchmarking low-light image enhancement and beyond. Int. J. Comput. Vision **129**, 1153–1184 (2021)
20. Mehrtash, A., Wells, W.M., Tempany, C.M., Abolmaesumi, P., Kapur, T.: Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. IEEE Trans. Med. Imaging **39**(12), 3868–3878 (2020)
21. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth international conference on 3D vision (3DV), pp. 565–571 (2016)
22. Pirozzi, N.M., Hoogenboom, J.P., Giepmans, B.N.G.: Colorem: analytical electron microscopy for element-guided identification and imaging of the building blocks of life. Histochem. Cell Biol. **150**(5), 509–520 (2018)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. MICCAI, pp. 234–241 (2015)
24. Schindelin, J., et al.: Fiji: an open-source platform for biological-image analysis. Nat. Methods **9**(7), 676–682 (2012)
25. Scotuzzi, M., et al.: Multi-color electron microscopy by element-guided identification of cells, organelles and molecules. Sci. Rep. **7**(1), 45970 (2017)
26. Shapson-Coe, A., et al.: A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. Science **384** (2024)
27. Yang, S.T., Lu, J.C., Tsao, Y.C.: Clustering and representative selection for high-dimensional data with human-in-the-loop. INFORMS J. Data Sci. (2025)