Microtomography: material 3-d structure recovery having electron-microscopy scans

Vadim Artemov[1], Egor Bulavko[2], Denis Kukushkin[3], Airat Kotlyar-Shapirov[4], Alexandra Razorenova[5]

TA – Egor Zakharov

Project repository: https://github.com/p0l0satik/DL-Microtomography

Author Contributions

1 - was responsible for developing, training and evaluating of hybrid UVNet model. He also assisted in writing the report and was involved in 3d visualization.

2 - participated in data pre-processing pipeline investigation. He trained and evaluated classical UNet architecture, as well as studied the effect of loss weighting on neural network performance. Additionally, was responsible for models' quality metrics calculation and material reconstruction quality estimation.

3 - was responsible for developing, training and evaluating of stacked UNet approach. He also helped in finding an optimal UNet solution. Finally, he came up with a 3D representation and it's converter.

4 - worked on programming minimal 3Conv architecture, optimized 2d to 3d converting for 3d models training, adapted metrics (IOU, accuracy and class recall) for 3d data representation (volumes) and evaluated models on test data. He also contributed to checking dataset validity and visualizing final results.

5 - was responsible for training data generation and evaluation, numerical solutions (provided by prof. Nikolay Koshev's numerical solution), 3D visualization, U-net with 3D convolutions architecture as well as project management and results collection and interpretation.

We thank Egor Zakharov for fruitful discussions and advices regarding models' implementation.

Introduction

Scanning electron microscopy (SEM) is a powerful non-invasive method of examination of microscopic structures, which has spatial resolution 1-100 nm. The principle of SEM operation is based on irradiation of the object under study with a monoenergetic electron beam and subsequent registration of intensity and/or energy distribution of electrons in the beam after reflection by the object (reflected SEM).

The problem of microtomography in reflected electrons microscopy is a classical ill-posed problem of mathematical physics. Standard methods of reflected electron tomography usually only aim at separating superimposed images of layers rather than investigating their characteristics or composition. The relatively low starting energy of the probe's electrons will make it possible to examine non-biological micro-volumes without any risk of damaging them.

The current project involves reconstructing the internal structure of a layered object (cladding) in terms of layer thickness, material density. In 2021 convolutional neural networks were effectively used to solve the inverse problem of fluorescence microtomography. CNN networks (UNet, VNet) presumably can be used as effective solver of reflected electron microtomography inverse problem.

The project is aimed to test and compare different CNN-architectures on volumetric layered structure recovery from SEM scans set with different canning energy. DL approach may be at use as a fast method more stable with respect to low SNR compared to numerical inverse solvers. Within a project we formulate problem of 3d structure recovery in two ways. Since we considered 2 layered structures, we regressed each layer height on a plane. In more complicated setup we considered volumetric segmentation problem.

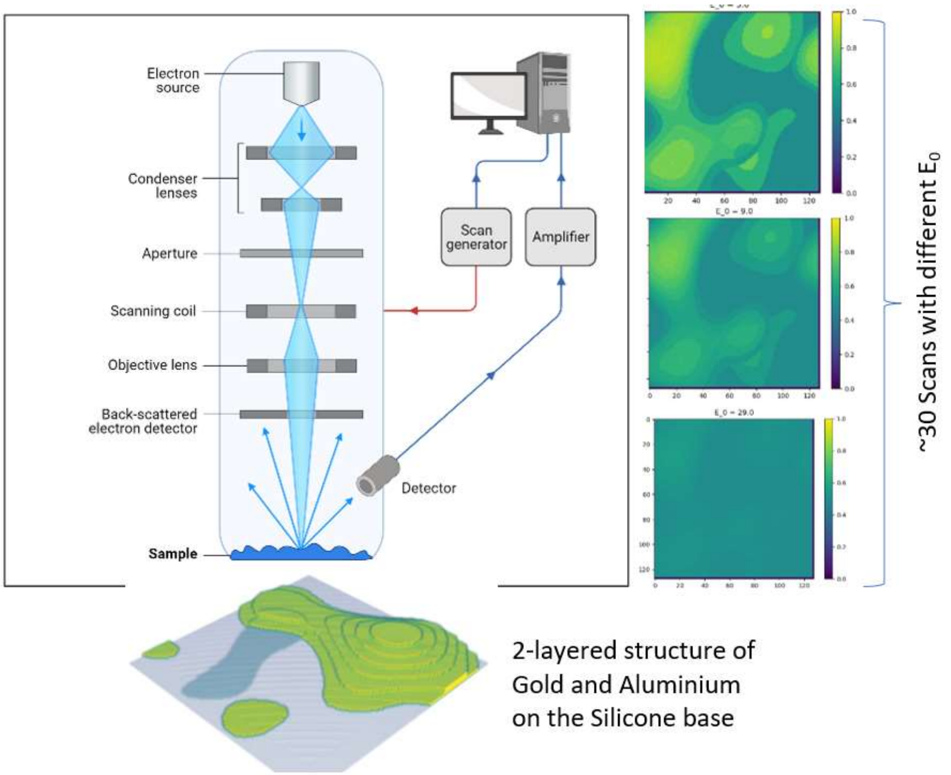

Electron source
Condenser lenses
Aperture
Scanning coil
Objective lens
Back-scattered electron detector
Sample
Scan generator
Amplifier
Detector
E_0 = 9.0
E_0 = 29.0
~30 Scans with different E_0
2-layered structure of Gold and Aluminium on the Silicone base

Figure 1. SEM set schematic and data origin

Methods

Data Representation and Generation

In current project we used a simulation algorithm to generate SEM scans of random 3d composite structures composed of two materials – stacked gold and aluminium layers with arbitrary height. The algorithm considers SEM detector as orthographical camera, thus each pixel on a scan register signal from the same one on a structure plane. Generation algorithm returns signal for each pixel normalized to silicone substrate attenuation (see Figure 1).

Random 3d structures were obtained via Perlin noise generator, each of two layers has an independent generator, thus resulting layers can be highly overlapped which results that gold layer was totally covered with aluminium, or the opposite situation with spatially separated materials. Since gold is a heavy metal as a result it has higher electron attenuation coefficient. If a too thick layer of gold is deposited it becomes opaque to the electron beam and thus the silicon substrate becomes invisible. Therefore, we decided to limit the thickness of gold layers 1/10 of the thickness of aluminium layer resulting to 15 nm and 150 nm limit for height.

2% Gaussian noise was added to registered scans. The level of noise was rather moderate in order to provide compatibility with straight-forward numerical inverse solution.

Evaluation Details

Test and Train Dataset structures were generated with different seeds. Scans set for each structure consisted of 31 set (128, 128) matrices corresponding for a signal registered for one of 31 scanning energies. The structures were represented in two different modalities: simplified and volumetric one. A simplified representation was 2 channel matrices each encodes pixel-wise thickness of aluminium/gold layer. Since our nearest plan is to consider more materials and get rid of sequential layers generation our goal is to recover material at each voxel of a volumetric structure independently of layers representation. Thus, from thickness matrices we generated volumetric segmentation masks with filled with material atomic numbers (or zeros if no material is presented in the voxel). We considered rather coarse volumetric representation with 10 nm resolution of height resulting to (16,128,128) tensors filled with material indices.

10400 pairs of scans and corresponding structures were generated, this set was split on train and validation dataset (9400 samples for train, 1000 for validation). We also generated 1000 test structures with different seeds for further test procedure. We also included real geometry of an experimental structure used in SEM lab.

Considered Models

Numerical solution

This is approach is expected to give poor results, as it is a pixel (voxel) based approach based on least squares method with Tikhonov regularization, that does not take into account the neighbouring pixels. In this regard, DL is expected to give better results, because the convolution operations, at their core, combine the information of different neighbouring elements. This method does not claim to be state-of the-art but can be used as initial benchmark.

Straightforward Convolutional model (3Conv)

We choose to apply feedforward network. That was physically plausible since high energy electrons do not have a tendency to high scattering. To avoid splitting images into pixel-wise data we initially performed 1x1 convolutions + batch normalization + ReLU. After a consultation we choose to change them to 3x3 convolutions for model to take into account neighbouring topography and final network architecture was (see Figure 2)

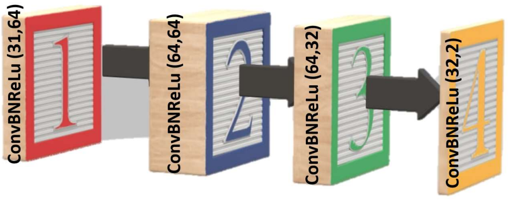


Figure 2. 3Conv model

In order to preserve image size, we did not do any pooling so model could aggregate data without need to compress and decompress it. Since number of layers was small, number of trainable elements did not explode (in total it was below 100k)

Standard UNet

UNet and VNet based models were considered and modified. Following models were trained and tested.

We utilized standard UNet architecture with 4 up-down blocks and skip connections (Figure below) and used the architecture as a basic building block for more complicated models (see below). To adapt it for our task, output two-channel images are being smoothed by sigmoid function and then renormalized: bottom layer (gold thickness matrix) – by 15 – and upper layer (aluminum thickness matrix) – by 150. We performed 150 epoch training using MSE as loss function and Adam optimizer (learning rate 3e-5, weight decay 5e-5) (see corresponding plot below). The same architecture was used to reconstruct volumetric segmentation.

Stacked UNet for layer thickness prediction

Stacked UNet was implemented according to with some minor changes. We use more dropout layers to prevent from overfitting. Like in the original paper the output of the unit model is split one of the branches is mapped assessed and then added to the other branch. We tried different number of stacks and different UNet depth but, eventually, depth 4 stacked 3 times shows the best results. We have also carried out experiments with loss. One option was to just sum all the output loss the other was to calculate its mean. The practice is in favor of the latter.

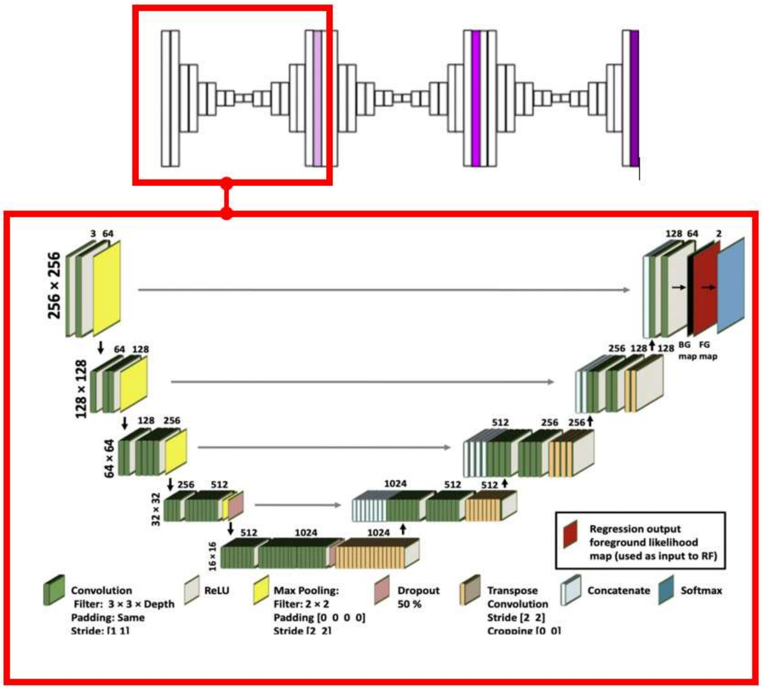

3 64
256 × 256
64 128
128 × 128
128 256
64 × 64
256 512
32 × 32
512 1024
16 × 16
1024
1024
512
512
512
256
256
256
128
128
128 64
2
BG map
FG map
Regression output foreground likelihood map (used as input to RF)
Convolution Filter: 3 × 3 × Depth Padding: Same Stride: [1 1]
ReLU
Max Pooling: Filter: 2 × 2 Padding [0 0 0 0] Stride [2 2]
Dropout 50 %
Transpose Convolution Stride [2 2] Cropping [0 0]
Concatenate
Softmax

Figure 3. Stacked UNet with additional droput layers

Hybrid UVNet

One of the architectures that we used was hybrid UVNet, which was suggested by TA Egor Zakharov. The Down part is using 2D convolution blocks, while the Up part uses 3D analogues for upsampling and deconvolution. The encoder part as usual consists of 3 double convolution blocks and 3 max-pooling layers. After each double convolution, we copy a tensor to a skip connection list.

In the lowest point of the model (bottleneck), we add more channels via the 2D convolution, then flatten the tensor and reshape it to make a 4D structure from it. After this have been done, we use three 3D upsampling and several 3D double convolutions to bring the data to the required shape.

Along the way, we add tensors from the skip connection list by concatenation. In that list we have 3D tensors, so we flatten them and reshape to the shape of a 4D tensor, where depth, width and height is the same, but the number of channels is adjusted to match the shape. The final tensor, after all of the convolutions and transformations, is a 3D structure, where each voxel contains 2 probabilities: one for gold and another one for aluminium, from which we recover the final structure.

We can try to explain how it works by using the explanation from the 2D analogue, we know that convolutions and down-samplings increase the receptive fields of the convolutions, while skip connection tensors provide the unmodified data, so that convolutions could see “broader” picture of the sample. We show by this model that hybrid UVNet is a viable model. It provides good quantitative loss as well as a nice visual final representation of the plates that we have.

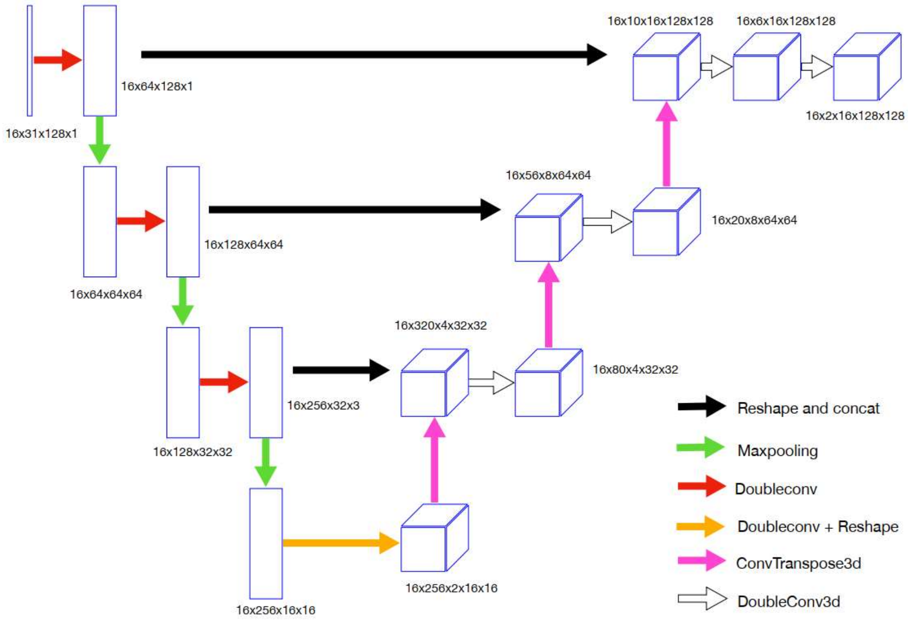


Figure 4. Hybrid UVNet

Experiments

Training statistics

Loss hyperparameters for 2 layer thickness matrices reconstruction.

When comparing UNet generated images with ground truth mask, it may be noticed that while aluminum layer demonstrates almost ideal correspondence with mask, gold layer predictions turns out to be a bit blurry and noisy. Detailed layer-to layer analysis shows that in absolute values average MSE for aluminum is 37 times greater than for gold. Taking into account those fact that normaization constant for aluminum is 10 times larger (which results in 100 times difference for MSE), we may conclude that in fact this ratio is 0.37, i. e. gold is more noisy, which is observed on the picture.

Again, aluminum-directed losses and gradients are heavier than gold-directed, which may be the reason for gold layer predictions being underfitted. We decided to introduce weighted loss function, combining MSE from gold (C_{Au})and aluminum (C_{Al}) layers with different coefficients. We tried C_{Al}/ C_{au} ratio equal to 0.01 (UNet001) and 0.1 (Unet01), Figure 5.

UNet001 demonstrates better performance for gold layer, but aluminum is now noisy too (see below), which means 0.01 is too small coefficient. 0.1 appeared to be a balanced variant, where gold layer noise decreased with no effect on aluminum layer. We managed to reach the situation when the first layer is only 1.7 times noisier than the second.

All considered models showed to be easy-trainable.

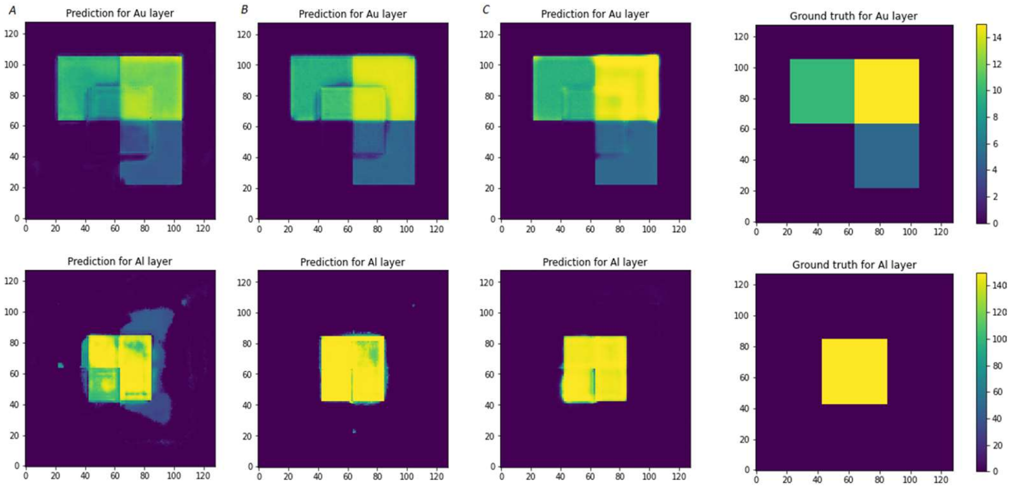
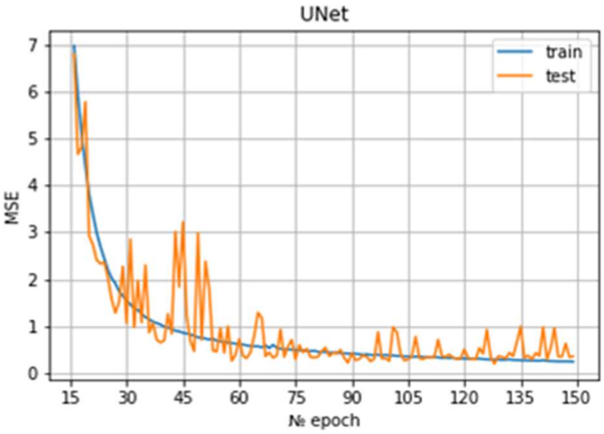


Figure 5. A - UNet. B - UNet001, C - UNet01



Basic UNet-to-2d training

As you can see, considering test, the saturation is reached +- after 100 epochs (MSE value +- 0.3), and after 120 epoch test loss started increasing again (MSE 0.35 on 150 epoch), which means that model became slightly overfitted.

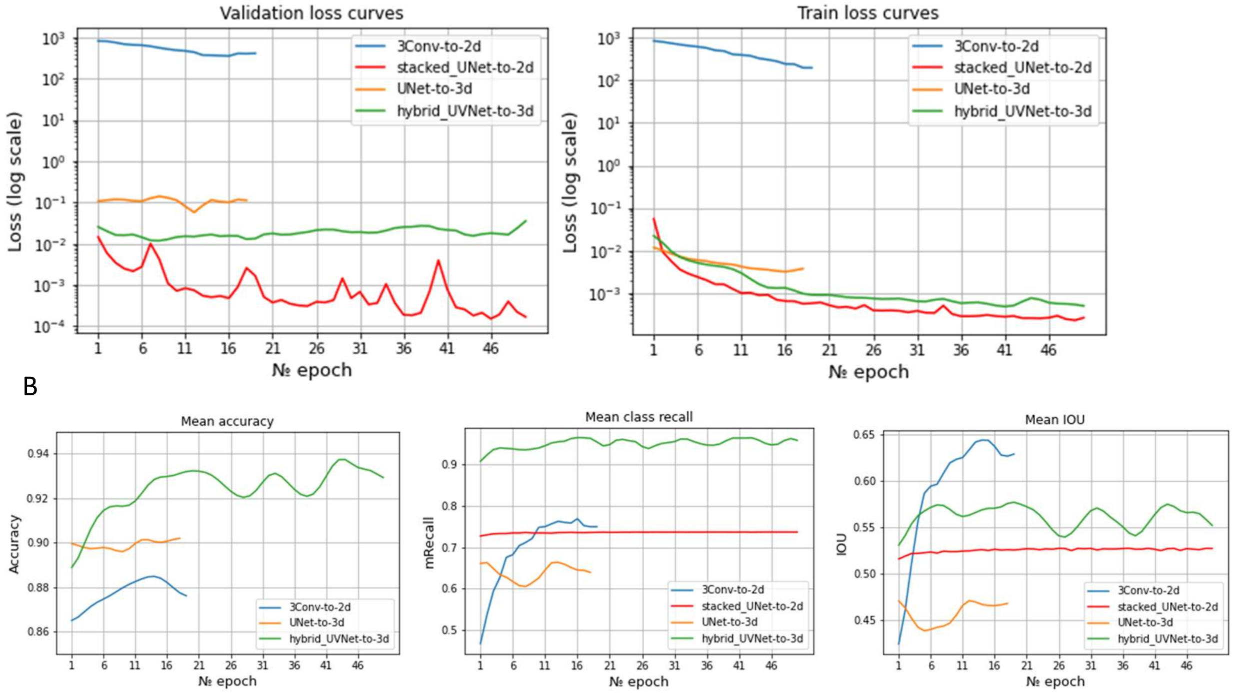


Figure 6. Considered models Training

a) Train and validation loss of considered models; b) Validation metrics

Model-to-2d – 2 layer’s thickness matrices are predicted for gold and aluminium layer.

Model-to-3d – Volumetric classification mask is predicted for 3 classes – gold or aluminium or none.

Results and Discussion

Our preliminary results were obtained for TEST set (Table 1) and experimental structure (Table 2). We estimated models predict on volumes either reconstructed directly from 3d tensor or via converting thickness matrices to 3d structures. Segmentation metrics were adopted for 3d. Mean intersection over union (mIOU), mean class recall (mRecall) and mean accuracy were estimated. We also assessed material reconstruction separately for gold and aluminium (Table 1.2 and Table 2.2). Visual comparison of recovered 3d structure quality with respect to layers can be accessed in Table 3 and 4.

Although applied successfully in microtomography, purely 2D convolutional methods did not yield best results. Here is why: main interests of aforementioned microtomography are segmentation-like tasks where focus is made on distinguishing parts that lie in X-Y plane. in other words, since 2D convolutions uses same weights through Z-axis, the amplitude of the image is integrated along that dimension and serves the purpose to increase the separability of presented parts in the 2D plane, while neglecting the height.

The best example is our baseline network called Conv - it managed to reconstruct ideally the flat shape of each metal while almost neglecting Aluminiums' thickness - thus leading to dramatic drops in mIOU metrics.

Following quite closely comes U-net, which was able to gain a good improvement based on testing results, but when evaluated on close-to-real electron scan it also failed in terms of reconstructing thin gold layer properly. Same goes to stacked U-net - they seem to improve on the test results, but perform still slightly worse than numeric solutions. Our intuition is that being quite deep, it overfitted on non-sharp edge examples provided by random seed generator.

Table 1.1 TEST dataset recovery

	mIOU	**mRecall**	**mAccuracy**
Numerical solution **with Tikhonov regularization**	**.617**	**.642**	**.876**
2d layers thickness mapping			
3Conv-to-2d	.633	.676	.890
UNet-to-2d	**.953**	.957	**.995**
stacked_Unet-to-2d	.909	.942	.978
3d volumetric mask segmentation			
UNet-to-3d	.620	.874	.949
hybrid_UVNet	.938	**.982**	.981
black – numerical solution benchmark; grey* – benchmark result; **green – best result within project			

Table 1.2 TEST dataset recovery material-wise

	GOLD			**ALUMINIUM**	
	mIOU	**mRecall**		**mIOU**	**mRecall**
Numerical (2% noise)	**.871**	**.773**		**.626**	**.756**
2d layers thickness mapping					
3Conv-to-2d	.766	.785		.162	.166
UNet-to-2d	.799	.809		.947	.950
stacked_UNet-to-2d	.934	.939		.846	.911
3d volumetric mask segmentation					
UNet-to-3d	.131	.701		.783	.963
hybrid_UVNet	**.968**	**.974**		**.869**	**.995**
***black** – numerical solution benchmark; grey* – benchmark result; **green** – best result within project; red – result below benchmark (Numerical solution)					

Finally, best results (both qualitatively and quantitatively) are yielded by models leveraging 3D convolutions. Two main ideas why this is happening:

Firstly, the data - electron scans are virtually representing the depth structure of the sample, since each energy bears information from different depths - thus operating on it with 3D convolutions makes more sense in the first place.

Secondly, by switching to 3D we are able to go from unbounded regression task in 2D to quite straightforward voxel classification. In other words, we are using a trick by increasing the dimensionality of the output data by one, we reduce the dimensionality of prediction values from R to C.

Table 2.1 Real experimental structure recovery

noise level [%]		mIOU	mRecall	mAccuracy
Numerical solution with Tikhonov regularization	**no noise**	.893	.900	.980
	.5	.632	.698	.978
	2.0	**0.572**	**0.614**	**0.962**
2d layers thickness mapping				
3Conv-to-2d		.336	.371	.897
UNet-to-2d		.523	.585	.948
stacked_UNet-to-2d		.541	.596	.947
3d mask segmentation				
UNet-to-3d		.684	.714	.964
hybrid_UVNet		**.734**	.815	**.978**
black *– numerical solution benchmark;* **grey** – benchmark result; red – result below benchmark				

Table 2.2 Real experimental structure recovery material-wise

	GOLD			ALUMINIUM	
	Precision	**Recall**		**Precision**	**Recall**
Numerical (2% noise)	**.41**	**.499**		**.53**	**.61**
2d layers thickness mapping					
3Conv-to-2d	.483	.499		.452	.531
UNet-to-2d	.487	.498		.486	.752
stacked_UNet-to-2d	.587	.500		.574	.750
3d mask segmentation					
UNet-to-3d	.199	**.514**		.889	.954
hybrid_UVNet	.33	.5		**.891**	**.959**
black *– numerical solution benchmark;* grey – benchmark result; **green** – best result within project; red – result below benchmark					

Table 2.3 Numerical solution with respect to signal noise level

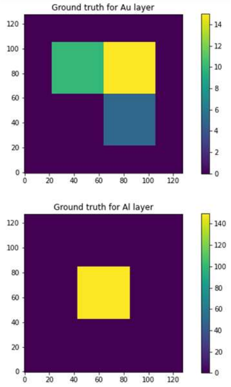
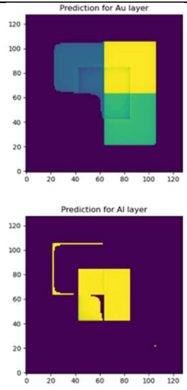
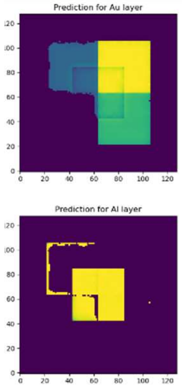
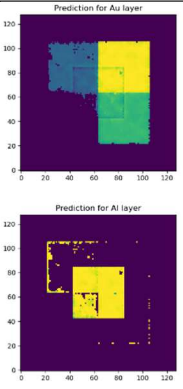
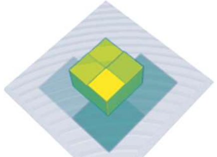
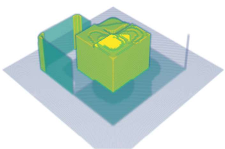
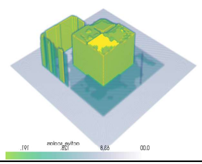
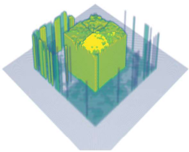
Ground Truth	***Numerical sol no noise***	***Numerical sol 0.5% noise***	***Numerical sol 2% noise***
			
			

Table 3.1 Predicted layer thickness matrices

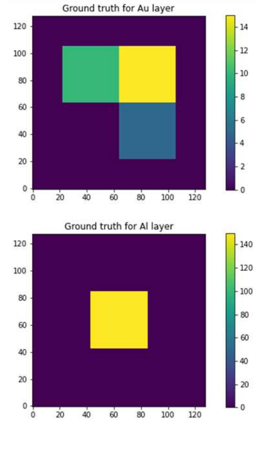
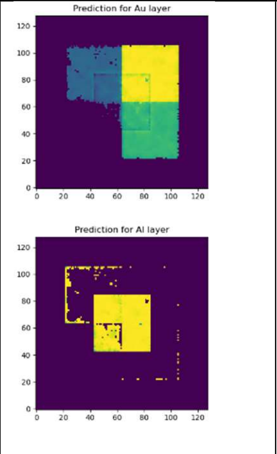
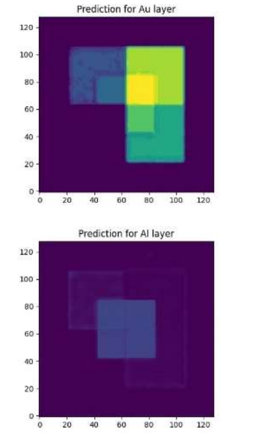
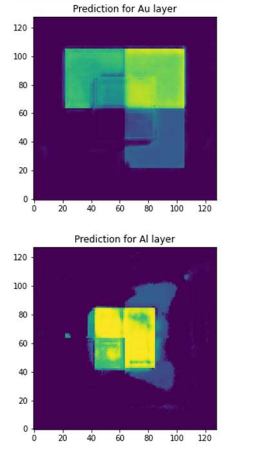
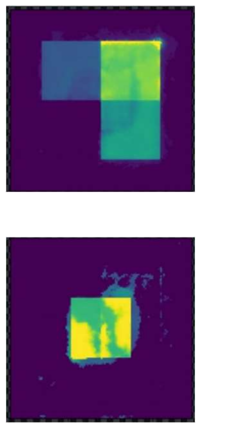
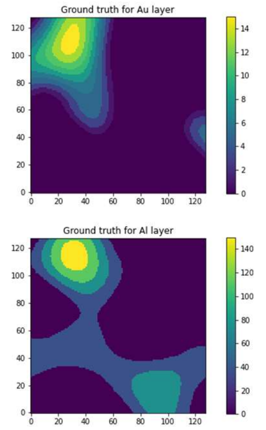
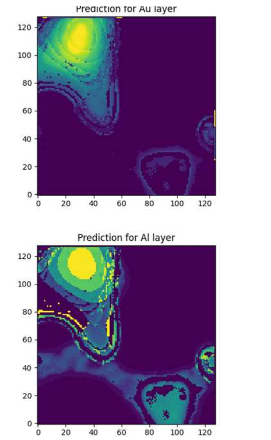
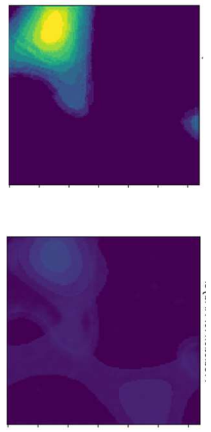
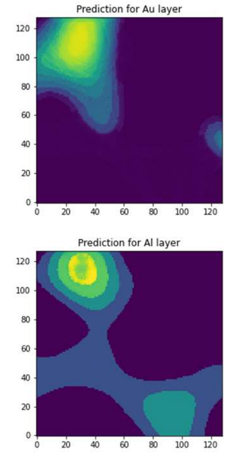
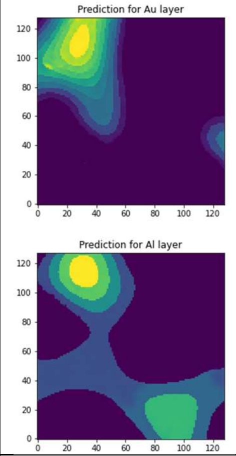
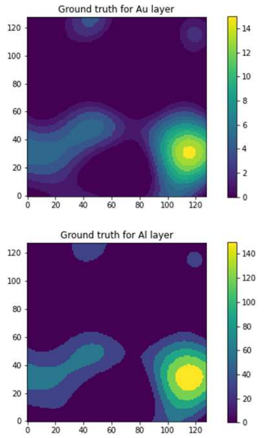
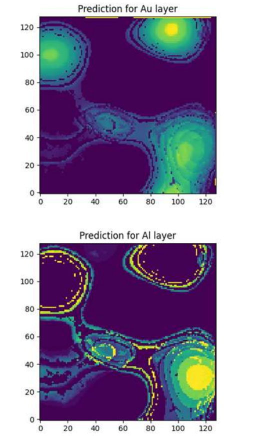
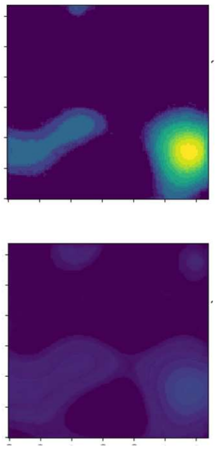
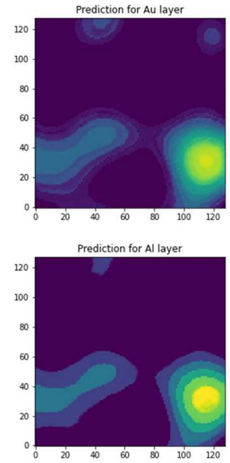
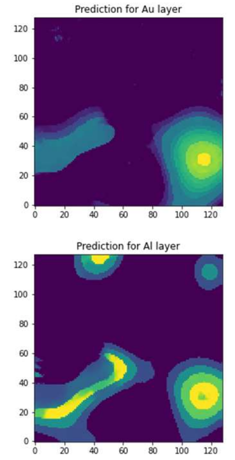
Ground Truth	Numerical solution 2% noise	3Conv	UNet-to-2d	stacked_UNet-to-2d
Real experimental structure				
				
TEST dataset examples				
				
				

Table 3.2 3d structures reconstructed from predicted material thickness matrices

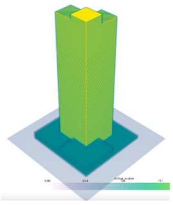
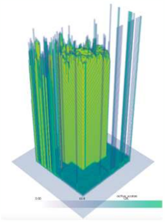
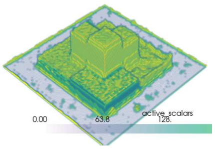
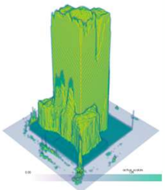
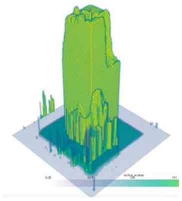
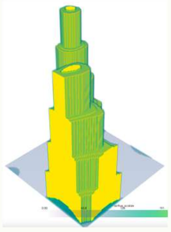
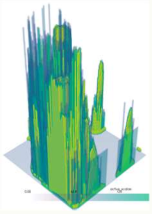
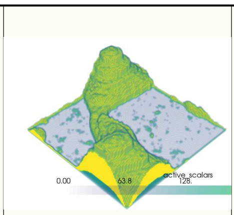
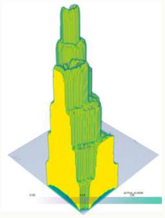
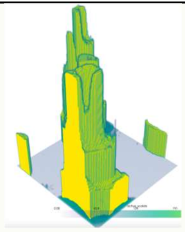
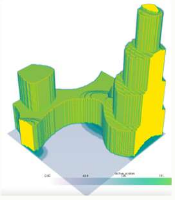
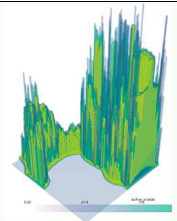
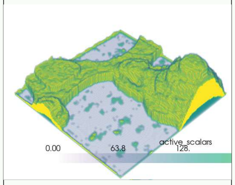
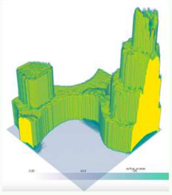
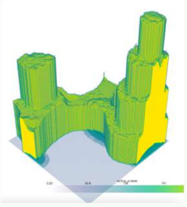
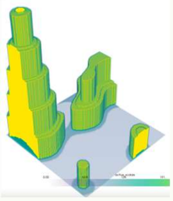
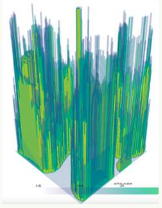
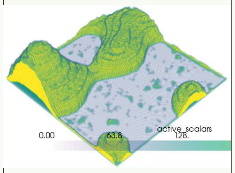
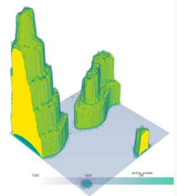
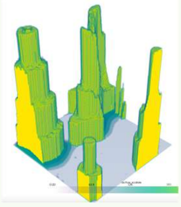
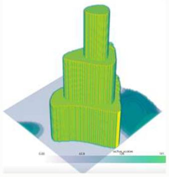
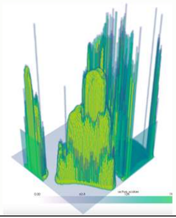
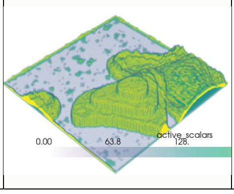
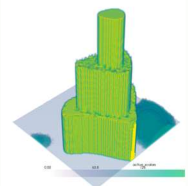
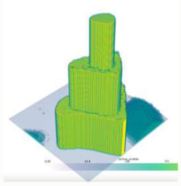
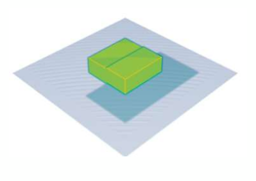
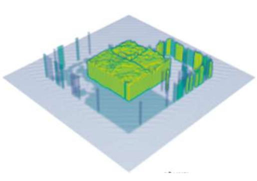
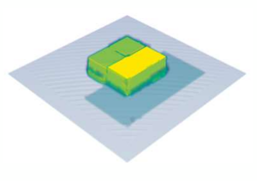
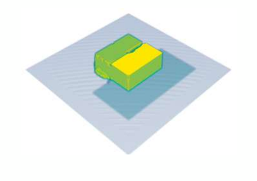
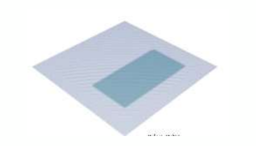
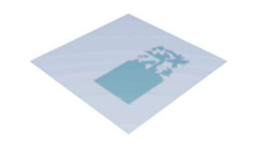
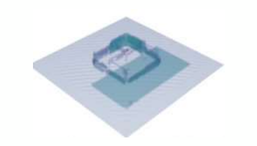
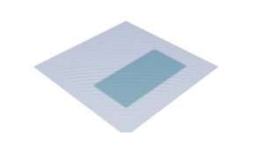
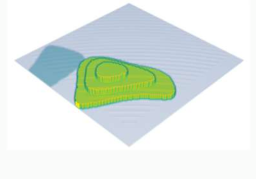
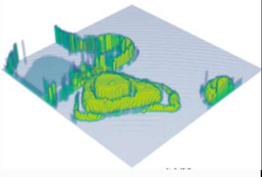
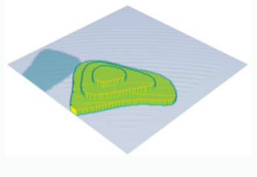
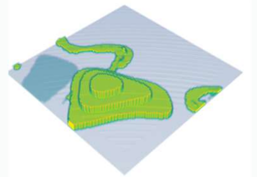
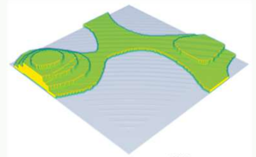
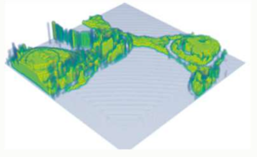
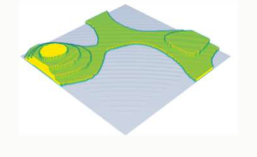
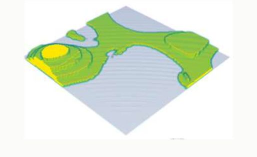
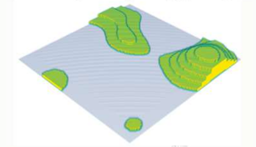
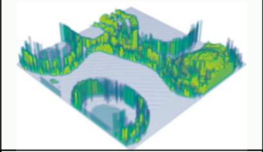
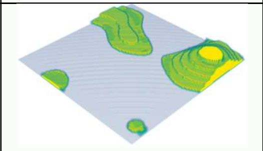
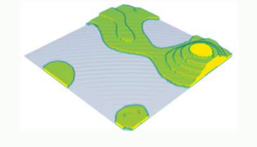
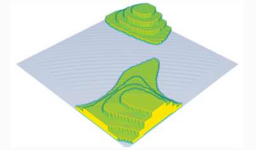
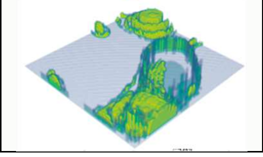
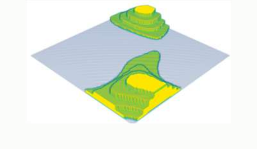
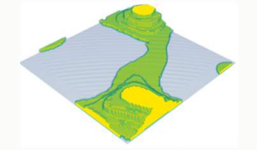
Ground Truth	**Numerical solution**	**3Conv**	**UNet-to-2d**	**stacked_UNet-to-2d**
Real experimental structure				
				
TEST dataset examples				
				
				
				
				

Table 4 Predicted 3d structure as volumetric segmentation mask

Ground Truth	**Numerical solution**	**UNet-to-3d**	**hybrid UVNet**
Real experimental structure			
			
Gold Layer reconstruction quality			
			
TEST dataset examples			
			
			
			
			

Conclusions

Our results revealed that Deep Learning approach can be at use within SEM field. CNNs showed promising preliminary results. The next step is to check the stability of models with respect to low SNR, which appeared to be a problem for classical numerical algorithms (see Table 2.3). Hybrid 3d UVNet should be tested on more fine structures and optimized in terms of loss function cost and input flexibility.

References

1. Kassim, Y. M., Glinskii, O. V., Glinsky, V. V., Huxley, V. H., Guidoboni, G., & Palaniappan, K. (2019, September). Deep U-Net regression and hand-crafted feature fusion for accurate blood vessel segmentation. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 1445-1449). IEEE.
2. Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In European conference on computer vision (pp. 483-499). Springer, Cham.
3. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
4. Abdollahi, A., Pradhan, B., & Alamri, A. (2020). VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. IEEE Access, 8, 179424-179436.
5. Golub, G. H., Hansen, P. C., & O'Leary, D. P. (1999). Tikhonov regularization and total least squares. SIAM journal on matrix analysis and applications, 21(1), 185-194.
6. Belhaj, M., Jbara, O., Filippov, M. N., Rau, E. I., & Andrianov, M. V. (2001). Analysis of two methods of measurement of surface potential of insulators in SEM: electron spectroscopy and X-ray spectroscopy methods. Applied surface science, 177(1-2), 58-65.
7. Rau, E. I., & Reimer, L. (2001). Fundamental problems of imaging subsurface structures in the backscattered electron mode in scanning electron microscopy. Scanning, 23(4), 235-240.
8. Rau, E. I., & Robinson, V. N. E. (1996). An annular toroidal backscattered electron energy analyser for use in scanning electron microscopy. Scanning: The Journal of Scanning Microscopies, 18(8), 556-561.
9. Jo, Y., Cho, H., Park, W. S., Kim, G., Ryu, D., Kim, Y. S., ... & Park, Y. (2021). Label-free multiplexed microtomography of endogenous subcellular dynamics using generalizable deep learning (pp. 1-9). Nature Publishing Group.
10. Parkinson, D. Y., Pelt, D. M., Perciano, T., Ushizima, D., Krishnan, H., Barnard, H. S., ... & Sethian, J. (2017, September). Machine learning for micro-tomography. In Developments in X-Ray Tomography XI (Vol. 10391, p. 103910J). International Society for Optics and Photonics.