Hybrid UNet for segmentation of SEEG electrodes on post-operative CT scans

Anja Pantović[a], Irène Ollivier[b] and Caroline Essert[a]

[a] ICube Laboratory, Université de Strasbourg, CNRS, Strasbourg, France; [b] Department of Neurosurgery, Strasbourg University Hospital, Strasbourg, France

ARTICLE HISTORY
Compiled September 14, 2022

ABSTRACT
A common treatment option for pharmacoresistant epilepsy is to surgically remove epileptogenic zone. Stereoelectroencephalography (SEEG) is a minimally invasive surgical procedure used to identify such zones. After SEEG, precisely determining exact positions of all of the implanted electrodes and their metallic contacts from post-operative CT scans is crucial to design the surgical resection plan. The metallic electrode contacts cause appearance of strong white artefacts in the CT scans which makes the localisation process difficult and imprecise. We propose an automatic approach for accurate localisation of SEEG electrode contacts, using a combination of the 2D and 3D U-Net neural network architecture. The proposed hybrid network makes the best out of the 2D and the 3D model used in our previous work and makes more accurate predictions, resulting in a decrease of false positive and false negative contact segmentations. The network was trained on 36 datasets where 18 belong to pharmacoresistant epilepsy patients from Strasbourg University Hospital, which were augmented using rotation and flipping operations to double the training data. Model was cross validated using the leave-one-out approach and evaluated on different metrics: Intersection over Union, Precision, Recall and F-1 score. The Hybrid model outperformed both the 2D and the 3D U-Net model and the results proved to be statistically significant. To complete the electrode segmentation process, segmented contacts are linked into electrodes using Gaussian Mixture Models.

KEYWORDS
Epilepsy; SEEG; deep learning; segmentation; UNet

1. Introduction

Epilepsy is a chronic neurological disease that affects around 50 million people in the world. Over 30% of the patients are resistant to current antiepileptic drugs. For patients with pharmacoresistant focal epilepsy of structural etiology (Scheffer et al. 2017), surgically removing the epileptogenic zone may be an appropriate treatment option. A successful resection relies on a precise identification of such zones and is usually achieved through Stereoelectroencephalography (SEEG)(Talairach and Bancaud 1966). During a SEEG procedure, between 10 and 18 depth electrodes are placed in targeted brain areas to record neuronal activity through metallic contacts evenly spaced along their body. The recording of SEEG signals and spatial association with the contact location allows for the detection of the spatial and temporal organization

CONTACT Anja Pantović. Email: pantovic@unistra.fr

of the epileptic seizure (Minotti, L., Montavont, A., Scholly, J., Tyvaert, L., Taussig, D. 2018), its starting point, and helps define the zone to resect. In current practice, identifying the exact location of up to 300 metallic contacts is performed manually by neurosurgeons. The presence of metal in electrode contacts produces strong streak artefacts in the CT images, which interfere with the localisation process, making it highly prone to human error and inaccuracy.

In order to save valuable time and increase the effectiveness of the upcoming surgical procedure, several groups have been working to automate this procedure (Meesters et al. 2015; Narizzano et al. 2017; Granados et al. 2018; Benadi et al. 2018; Medina Villalon et al. 2018; Pantovic et al. 2022a). Narizzano et al. suggested an approach based on a geometrical-constrained search. Meesters et al. extracted guiding screws with a multi-scale filter and determined the possible tip locations inside a wedge-shaped region. This method did not take into account electrode bending, assuming rigid electrodes. Granados et al. suggested a method aiming for an automatic segmentation of both bolts and contacts. Their contact search strategy was based on the direction of the bolt, given the distance and angle constraints. The proposed method, however, needs manual adjustments to handle electrodes crossing. All of these methods use co-registered post-implantation CT scans with pre-implantation MRI and hence rely on pre-operative plans. Benadi et al. developed two interactive and an automatic method carried out using the 3D Slicer platform with post-operative CT scans. The automatic method does not take into account electrode bending and inaccurately segments contacts that are surrounded by metallic artefacts. Medina Villalon et al. introduced an automatic approach for electrode segmentation. First extra-cerebral elements from CT scans are removed. The electrodes are segmented using a threshold found from the histogram of grey values, assuming electrode intensity values significantly greater than those of brain structures. They encounter issues with false positive and false negative segmented contacts which are overcome by calculating the median distance between consecutive contacts and using it to add or delete missing or oversegmented contacts.

In a previous work, we proposed to take advantage of deep learning to automatically segment electrode contacts. Two approaches were introduced respectively based on a 2D and a 3D U-Net (Pantovic et al. 2022a), where post-implantation CT scans showing the SEEG electrodes were used for network training. Both networks proved to be highly accurate to segment the most visible electrode contacts. However, the 2D model failed to distinguish contacts from the metallic artefacts when two electrodes were positioned close to one another, while the 3D model failed to detect edge contacts lying close to the skull bone. Yet the two models seemed to complement each other, each one seeming able to deal better with the issues of the other. Inspired by Isensee et al. who used an ensemble of U-Net inspired architectures for segmentation of cardiac structures (Isensee et al. 2018), this paper presents an approach combining a 2D and a 3D U-Net model into a unique hybrid network to make the most of their advantages and overcome their respective limitations. Segmented contacts are grouped together into their corresponding electrodes using Gaussian mixture models. The proposed approach is, up to our knowledge, the first solution to automatically and accurately segment SEEG electrode contacts, robust to electrode bending and electrodes crossing, not requiring prior information about electrode type and linking the segmented instances into corresponding electrodes.

In a previous work, we proposed an approach to generate synthetic training data for SEEG electrode contact segmentation (Pantovic et al. 2022b). Training a 2D U-Net model on an augmented dataset that contains synthetic images generated following the proposed method accounts for better final segmentation results. However, it seems that

such data might not be well suited for 3D models as the 3D U-Net performed poorly when trained on synthetic data and gave less accurate results than when such data were omitted. This can be explained by poor quality of real data along z-axis, whereas synthetic volumes have high resolution along all three axes. As a result, the proposed Hybrid model performed better when trained without any additional synthetic data.

2. Methods

2.1. Dataset

The initial dataset consisting of post-SEEG CT scans from 18 pharmacoresistant epilepsy patients was obtained at Strasbourg University Hospital, each containing between 192 and 259 CT slices of resolution 512×512 in DICOM format. As border pixels of each slice capture only background pixels, all slices were cropped to the dimension of 400×400 to reduce memory usage. Electrode contacts were labeled manually using 3D Slicer (Kikinis et al. 2014) and its segmentation tools. The number of annotated contacts and the total number of electrodes annotated was compared to the planning document to assure a high accuracy of the manual segmentation.

To enlarge the dataset, random rotation and flipping were performed on the original CT images. Each patient's volume was used to create one augmented volume. Annotations were transformed accordingly. The final dataset consisted of 36 volumes, accounting for a total of 8278 images.

2.2. Electrode contact segmentation

To segment SEEG contacts from the post-implantation CT scans, we have implemented and combined a 2D and a 3D version of the commonly used U-Net (Ronneberger et al. 2015; Çiçek et al. 2016) neural network architecture. The two networks are first run separately, then their results are combined to generate the final prediction, as described below. The task at hand was formulated as an instance segmentation problem due to a small size of electrode contacts, inaccurate manual ground truth masks due to the surrounding metal artefacts and the aim to explore which brain zones were reached with each electrode contact, rather than evaluating the prediction for each individual pixel.

2.2.1. U-Net architecture

The U-Net encoder-decoder neural network architecture was chosen for the task due to its outstanding performance in medical image segmentation (Ronneberger et al. 2015). The network consists of a contracting path and an expansive path. The former represents a general convolutional process and consists of the repeated application of two 3×3 convolutions for the 2D model and two $3 \times 3 \times 3$ convolutions for the 3D model, followed by ReLU activation and batch normalization. 2×2 (or $2 \times 2 \times 2$) Max pooling is applied to reduce spatial dimensions and provides high resolution features. The latter represents the expansive path which is built by feature map upsampling, followed by a 2×2 (or $2 \times 2 \times 2$) transpose convolution, halving the number of feature channels and a concatenation with the cropped feature map from the contracting path. The cropping is performed due to the loss of border pixels in each convolution, while the skip connections between layers enable context information and precise localization.

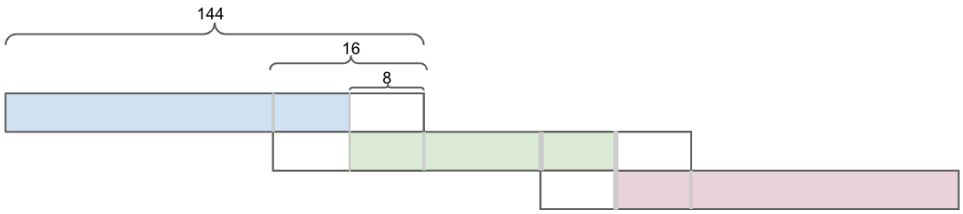


Figure 1.: Simplified illustration of overlapping patches. Numbers represent the number of pixels along the x-axis. The white part of a patch represents output scores which are discarded from the corresponding patch.

Finally, a 1×1 (or $1 \times 1 \times 1$) convolution is applied followed by softmax activation function, after which each pixel (voxel) is assigned a probability of it belonging to an electrode contact.

Both the 2D and 3D models were trained on the same dataset, with the difference being in the input data format.

2.2.2. 2D Model

For the 2D network, the data were organised as 8278 independent images of dimension 400×400 viewed from the axial plane which were fed into the network.

2.2.3. 3D Model

For the 3D U-Net, volumes were formed by grouping CT slices by patient. To fit the data into the GPU RAM memory, the network had to be trained on smaller patches. For this reason, each volume was subdivided into overlapping patches of dimension $144 \times 144 \times 135$. The overlapping step on the x and y axes accounted for 16 pixels, while the overlap on the z axis varied depending on each patient's total number of slices, always fitting 2 patch-depths along the z axis. The patch overlap was introduced to fully exploit the neighboring information of each voxel. The same overlapping technique was used for predictions. As the overlapping pixels were assigned multiple values, for them, prediction for the patch where that pixel belongs to the inner half of the overlap was kept. For better understanding, 2D illustration of this process on 3 consecutive overlapping patches is given in Fig.1, while in practice the same logic was applied along all 3 axes. In Fig.1, the white parts of a patch represent discarded predictions within that patch, while the colored parts of the overlap are the predictions that are kept from the patch of that color. Final output predictions were re-sampled to the original voxel resolution of the initial volume.

2.2.4. Networks configuration

Both models were trained in equivalent hyper parameter setups. *Adam* optimiser was used with the initial learning rate $1 * 10^{-5}$, $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 * 10^{-8}$ and the learning decay of $1.99 * 10^{-7}$, used to help fine-tune the network. The loss function was binary crossentropy. Batch size for the 2D U-Net was 10, while for the 3D U-Net, due to memory constraints, batch size was set to 2. The networks were trained for 150 epochs (determined empirically) and cross-validated using the 'leave-one-out' approach: for each training phase one original volume and its augmented counterpart were left out for testing so that the testing is done on an unseen case. Each training

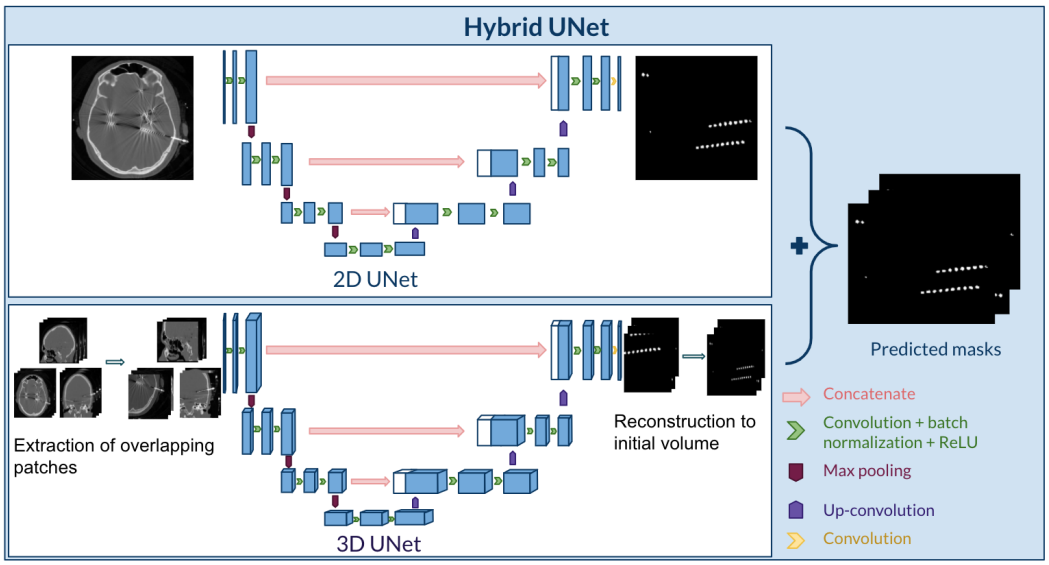


Figure 2.: Hybrid U-Net pipeline

phase was done, for both networks, using the remaining 17 original and 17 augmented patient data, from which 10% served for validation.

2.2.5. Hybrid U-Net

To obtain the final predictions, the output predictions of the 3D network were first rebuilt to the original voxel resolution of the initial volume. Then, they were averaged with the softmax output probabilities of the 2D model. Finally, all of the voxels with a probability lower or equal to 50% of belonging to an electrode contact were rounded to 0 and labeled as background, while the rest were classified as electrode contacts and rounded to 1. An illustration of the overall pipeline is shown in Fig.2.

2.3. Electrode contact clustering

To group segmented electrode contacts into the corresponding electrodes, Gaussian mixture model (GMM) clustering algorithm has been implemented, assuming that all the data points were generated from a mixture of a number of Gaussian distributions, m, equal to the number of implanted electrodes, with unknown parameters. The method used to initialize the weights, the means and the precisions was *k-means*. With the number of initialisations set to 100 and maximum number of iterations set to 1000, the GMM algorithm takes as input number of electrodes, m, and assigns each segmented electrode contact to one of the m clusters.

To ensure that any possible false positive segmented contacts, lying further away from the electrodes, are not interfering with the clustering algorithm, such contacts are filtered out beforehand based on a fixed threshold distance from the nearest connected component.

3. Experimental Setup and Evaluation Metrics

Network training was performed on 4 NVIDIA GeForce RTX2080 Ti GPUs, with 11GB RAM each. Networks were trained for 150 epochs and evaluated on the following metrics: *Dice coefficient (or F1-score), precision*, *recall* and *average precision* (AP). The Dice coefficient was computed using the true positive, false positive and false negative values at two different levels: individual pixels (*pixel-wise*) and full predicted electrode contacts (*contact-wise*). To avoid confusion, the pixel-wise Dice coefficient will be referred to as 'Dice coefficient' while the contact-wise Dice coefficient will be referred to as 'F1-score' in the rest of the paper. As a correct localisation of all the contacts is more relevant for the task than accurately segmenting individual pixels, segmented instances were identified and the other metrics were computed instance-wise. For this reason, an algorithm based on a connected component count was implemented to identify segmented instances and to compute the number of true positive (TP), false positive (FP) and false negative (FN) predicted contacts for each test case, which were then used to calculate the evaluation metrics as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\text{-}score = 2 * \frac{precision * recall}{precision + recall}$$

Average Precision is calculated as the area under the precision-recall curve.

To evaluate the statistical significance of the results, a paired two-sample Student's *t-test* was performed, testing the Hybrid network against the results obtained with the 2D and the 3D models separately.

4. Results and Discussion

Average training times for the 2D and the 3D model were 1h31' and 5h18' respectively. Results obtained with the 2D, 3D and the Hybrid U-Net, calculated after cross-validation and averaged across all folds are given in Table 1. For comparison, results from the previous study (Pantovic et al. 2022a) are also given in Table 1.

We first observe a slight improvement in the results, over all validation metrics, obtained using the 2D and the 3D model when compared to the results previously obtained using the same network models without any data augmentation (Pantovic et al. 2022a). However, when looking at the number of false positive and false negative segmented contacts, we conclude that using simple augmentation techniques alone was not enough to overcome the main issues of these two networks: incorrectly classifying metallic artefacts as electrode contacts with the 2D U-Net (high rate of FP) and failing to segment most of the edge contacts (high rate of FN) with the 3D U-Net. We also observe that the proposed Hybrid network outperformed both the 2D and

the 3D models, taking advantage of each of them, which resulted in a lower number of both FP and FN segmented electrode contacts. This can also be clearly observed when visualizing the resulting predictions.

To test the statistical significance of the results, a paired two-sample Student's t-test, with the significance level $\alpha = 0.05$, was performed over the F1-score, which is the most relevant to our problem, testing the Hybrid network against the 2D and the 3D U-Net. Resulting p-values were 6×10^{-4} and 8×10^{-4}, respectively, proving the improvement to be statistically significant. The 2D and the 3D models were also tested individually against the results obtained in the previous study (Pantovic et al. 2022a) without data augmentation. Both of the resulting p-values were greater than 0.05, indicating no statistically significant difference on this metric. We can conclude that using the Hybrid approach brings more benefit than performing a simple data augmentation.

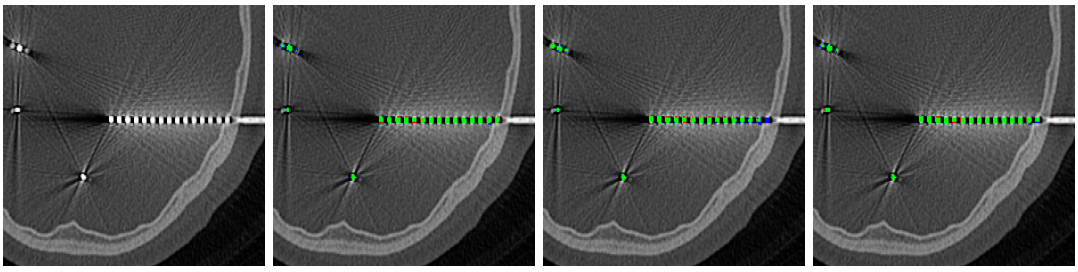
Figure 3 shows 2D visualizations highlighting the advantages/disadvantages of the three models. All models correctly segmented most of the electrode contacts. However, as it can be seen in the first row of Figure 3, the 3D U-Net sometimes failed to segment contacts lying close to the skull (FN in blue). On the other hand, the 2D model sometimes misclassified strong artefacts as electrode contacts, which can be seen in the second row of Figure 3 (FP in red). By averaging the output scores of these 2 models, the Hybrid network successfully overcame those issues, correctly avoiding artefacts and more efficiently segmenting edge contacts (Figure 3d). In extreme cases where an edge contact appeared to be overlapping the bone in the CT image, completely matching bone brightness, Hybrid U-Net failed to segment such contacts. These false negatives could be identified when grouping the contacts into electrodes and correlating with electrode specifications.

3D views of the ground truth and prediction masks obtained with the 2D U-Net, 3D U-Net and the Hybrid U-Net are given in Figure 4. The detected connected components are labelled in different colours. False positive contacts are highlighted with white circles while a red circle is inserted where a contact was missed (false negative). With the 2D model, all false positives are caused by strong artefacts, while with the 3D model both false positive and false negative predictions correspond to the edge contacts - they are either missed or the tip of a screw was misclassified as an electrode contact. As it can be seen in Figure 4c, these flaws are improved when the two models are combined into a Hybrid network.

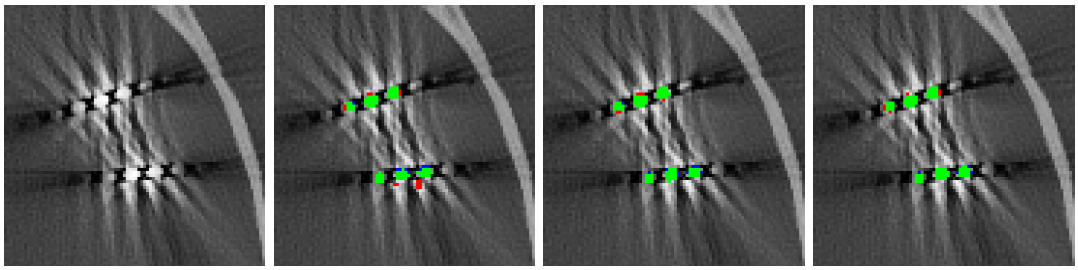
A remaining concern is the segmentation of the vertically implanted electrode. Ver-

Table 1.: Results obtained with the 2D U-Net, 3D U-Net and the Hybrid U-Net model, averaged across all cross-validation folds. The first two rows correspond to results from the previous study (Pantovic et al. 2022a), while the third and the fourth row show results of the 2D and 3D model trained on the augmented dataset.

Network model	**Dice coef.**	**TP**	**FP**	**FN**	**Precision**	**Recall**	**F1 score**	**AP**
2D UNet Pantovic et al.	0.807	180.6	9.6	4.4	0.950	0.976	0.963	0.927
3D U-Net Pantovic et al.	0.806	179.5	5.0	5.5	0.973	0.970	0.971	0.943
2D UNet	0.850	182.2	7.6	2.8	0.960	0.985	0.972	0.945
3D UNet	0.810	180.1	4.0	4.9	0.978	0.973	0.975	0.949
Hybrid UNet	**0.867**	**183.0**	**1.1**	**2.0**	**0.994**	**0.989**	**0.992**	**0.981**



(a) Original image (b) 2D U-Net (c) 3D U-Net (d) Hybrid U-Net



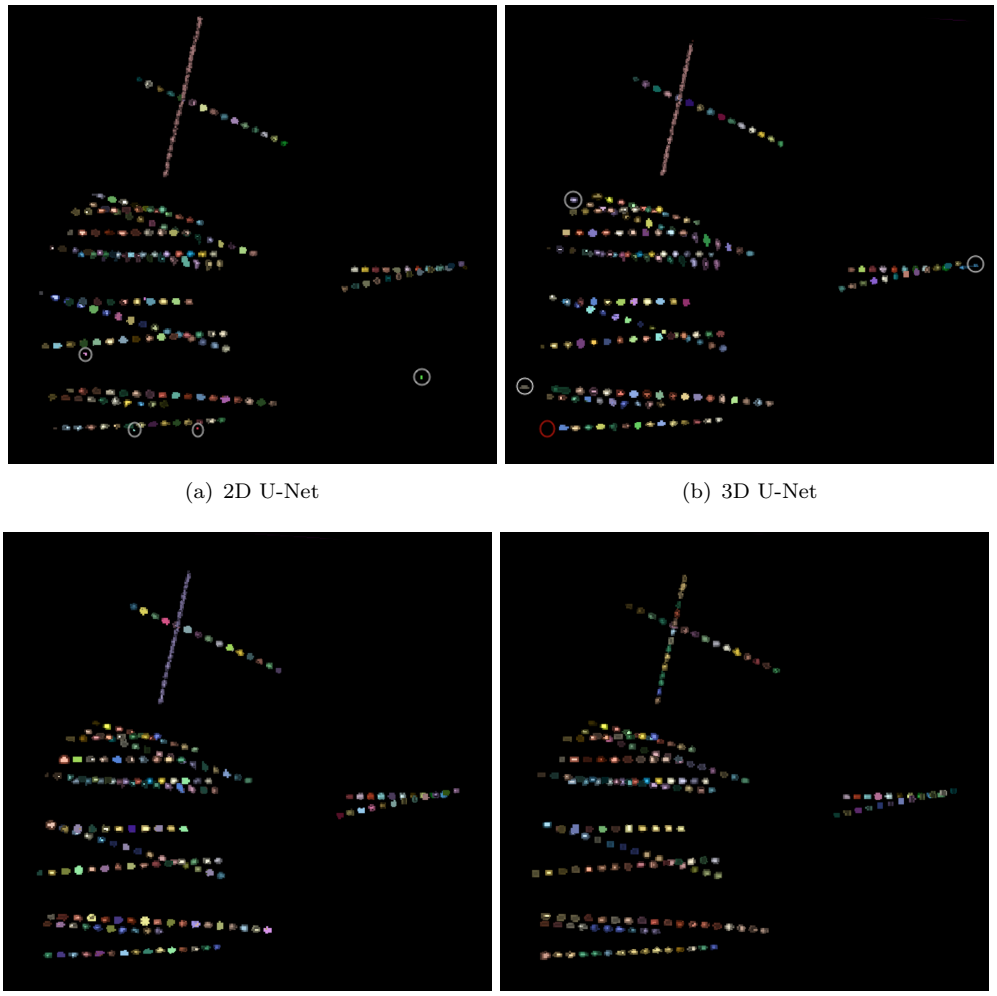
(e) Original image (f) 2D U-Net (g) 3D U-Net (h) Hybrid U-Net

Figure 3.: Original images and predictions obtained with the 2D, 3D and the Hybrid U-Net models. True positive pixels are colored in green, false positives in red and false negatives in blue.

tical electrodes appear on the axial plane as a single contact (see for instance the single white shape at the bottom of Figure 3a) and on most slices as an uninterrupted shape, with only slight brightness variations and no clear space between neighbouring contacts. In 3D, whatever the network used for segmentation, most of their contacts appear connected (see Figure 4c), and would require further postprocessing to obtain a correct segmentation, such as applying erosion and dilation operations.

The proposed segmentation approach is based on the use of both a 2D and 3D U-Net model, strengthening their respective advantages through their combination into a hybrid network. The ensambling method is simple - averaging the softmax activation of the 2 models, yet it drastically improves the individual results of the 2D and the 3D networks alone, successfully overcoming their respective limitations - incorrect classification of metal artefacts as electrode contacts (2D U-Net) and failure to segment most of the edge contacts (3D U-Net). Different ways of combining the networks could be implemented in place of the averaging method, such as using a weighted sum, introducing a different network architecture in the hybrid network and using majority voting. However, the results reach the desired accuracy and we do not think that more complex ways of ensambling the networks would bring any benefits to the task. On the other hand, as the 2D model benefits from additional synthetic training images (Pantovic et al. 2022b), segmentation could possibly be refined by training the 2D network on a larger, combined dataset while the 3D model remains trained on real and augmented volumes only. This remains to be evaluated.

The final stage of the SEEG electrode segmentation is grouping the segmented contacts by their corresponding electrodes, using Gaussian mixture models. Results are represented in Figure 5, where each cluster is labeled with a different color.



(a) 2D U-Net (b) 3D U-Net

(c) Hybrid U-Net (d) Ground truth

Figure 4.: 3D views of prediction masks obtained with the 2D U-Net, 3D U-Net and the Hybrid U-Net and the ground truth. Connected components are labelled in different colours. FP are highlighted with white circles and FN with a red circle.

5. Conclusion and Outlook

In this paper we proposed a method for segmentation of SEEG electrode contacts combining a 2D and a 3D U-Net encoder-decoder architecture into a Hybrid network. The network has been evaluated on the CT images of 18 different patients who underwent SEEG procedure at Strasbourg University Hospital. The proposed solution to provide an accurate segmentation of electrode contacts does not require any prior information about the type nor the number of electrodes. The contacts are then grouped into electrodes using Gaussian mixtures, taking as input number of implanted electrodes.

The proposed method yields the average result of 1.1 false positive contacts and 2.0 false negative contacts per patient, out of 185 contacts (on average) to be identified. Most of the electrodes have all of the contacts correctly segmented and they can be labeled according to the cortical parcellation. The rare cases of false positives or false negatives proved to belong to a contact that lies closest to the screw. As there is a

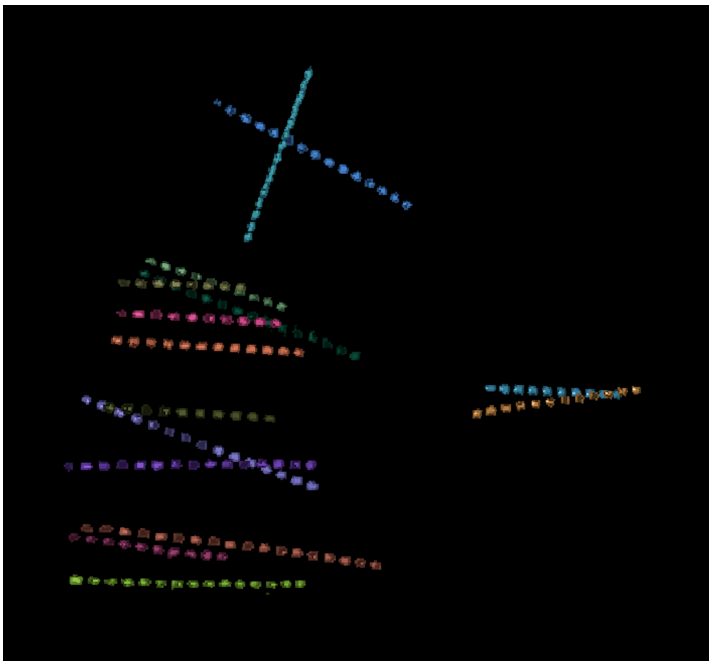


Figure 5.: Results of the contact clustering with Gaussian mixture models. Each cluster is represented with a different color.

known number of electrode contacts that an electrode model can have, this could be used to identify which electrode has a missing (or an additional) contact in order to adjust it. We are confident that this way the Hybrid network could be reliably integrated in a computer-assisted decision-making pipeline, to save neurosurgeons valuable time on locating implanted electrodes and increase the accuracy of the task.

In the future work, our goal is to automate and optimise the trajectory planning process of multiple electrodes that are used during SEEG, which is currently performed manually by neurosurgeons and neurologists. The approach proposed in this paper will allow for an automatic labeling of high number of retrospective cases which would further be used for training the model.

Acknowledgments

This work was funded by ArtIC project "Artificial Intelligence for Care" (grant ANR-20-THIA-0006-01) and co-funded by Région Grand Est, Inria Nancy - Grand Est, IHU of Strasbourg, University of Strasbourg and University of Haute-Alsace, France.

References

Benadi S, Ollivier I, Essert C. 2018. Comparison of interactive and automatic segmentation of stereoelectroencephalography electrodes on computed tomography post-operative images: preliminary results. Healthcare technology letters. 5(5):215–220.

Granados A, Vakharia V, Rodionov R, Schweiger M, Vos S, O'Keeffe A, Li K, Wu C, Miserocchi A, Mcevoy A, et al. 2018. Automatic segmentation of stereoelectroencephalography

(SEEG) electrodes post-implantation considering bending. International Journal of Computer Assisted Radiology and Surgery. 13:935–946.

Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, H MHK. 2018. Statistical atlases and computational models of the heart. acdc and mmwhs challenges. Lecture Notes in Computer Science. Available from: http://dx.doi.org/10.1007/978-3-319-75541-0.

Kikinis R, Pieper SD, Vosburgh KG. 2014. 3D Slicer:A platform for subject-specific image analysis, visualization, and clinical support. In: Intraoperative imaging and image-guided therapy. New York: Springer; p. 277–289.

Medina Villalon S, Paz R, Roehri N, Lagarde S, Pizzo F, Colombet B, Bartolomei F, Carron R, Bénar CG. 2018. EpiTools, A software suite for presurgical brain mapping in epilepsy: Intracerebral EEG. Journal of neuroscience methods. 303:7–15. Available from: https://doi.org/10.1016/j.jneumeth.2018.03.018.

Meesters S, Ossenblok P, Colon A, Schijns O, Florack L, Boon P, Wagner L, Fuster A. 2015. Automated identification of intracranial depth electrodes in computed tomography data. In: IEEE 12th International Symposium on Biomedical Imaging (ISBI). p. 976–979.

Minotti, L, Montavont, A, Scholly, J, Tyvaert, L, Taussig, D. 2018. Indications and limits of stereoelectroencephalography (SEEG). Neurophysiologie Clinique. 48(1):15–24.

Narizzano M, Arnulfo G, Ricci S, Toselli B, Tisdall M, Canessa A, Fato M, Cardinale F. 2017. SEEG assistant: A 3DSlicer extension to support epilepsy surgery. BMC Bioinformatics. 18:124.

Pantovic A, Ollivier I, Essert C. 2022a. 2D and 3D-UNet for segmentation of SEEG electrode contacts on post-operative CT scans. In: Proceedings of Medical Imaging: Visualization and Image-Guided Procedures.

Pantovic A, Ren X, Wemmert C, Ollivier I, Essert C. 2022b. Generation of synthetic training data for SEEG electrodes segmentation. In: Proceedings of Medical Imaging: Visualization and Image-Guided Procedures.

Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Springer. p. 234–241.

Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, Hirsch E, Jain S, Mathern GW, Moshé SL, et al. 2017. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 58(4):512–521.

Talairach J, Bancaud J. 1966. Lesion, "irritative" zone and epileptogenic focus. Stereotactic and Functional Neurosurgery. 27(1-3):91–94.

Çiçek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.