

Investigation on Encoder-Decoder Networks for Segmentation of Very Degraded X-Ray CT Tomograms

Idris Dulau
LaBRI UMR 5800
Bordeaux University
Talence
FRANCE
idris.dulau@labri.fr

Marie
Beurton-Aimar
LaBRI UMR 5800
Bordeaux University
Talence
FRANCE
beurton@labri.fr

Yeykuang Hwu
Institute of Physics
Academia Sinica
Taipei
TAIWAN
phhwu@sinica.edu.tw

Benoit Recur
Institute of Physics
Academia Sinica
Taipei
TAIWAN
benoit.recur@gmail.com

ABSTRACT

Field of View (FOV) Nano-CT X-Ray synchrotron imaging is used for acquiring brain neuronal features from Golgi-stained bio-samples. It theoretically requires a large number of acquired data for compensating CT reconstruction noise and artefacts (both reinforced by the sparsity of brain features). However reducing the number of radiographs is essential in routine applications but it results to degraded tomograms. In such a case, traditional segmentation techniques are no longer able to distinguish neuronal structures from surrounding noise. Thus, we investigate several deep-learning networks to segment brain features from very degraded tomograms. We focus on encoder-decoder networks and define new ones addressing specifically our application. We demonstrate that some networks wildly outperform traditional segmentation and discuss the superiority of the proposed networks.

Keywords

X-Ray nano-tomography, segmentation, deep-learning, brain imaging.

1 INTRODUCTION

Field-Of-View (FOV) Nano-CT X-Ray synchrotron imaging provides 3D images of biological samples at about 300nm by computed tomography (CT). In this study, bio-samples are mouse brains stained with a Golgi solution targeting neuronal connectome (neuron cells, axons, dendrites, ...). Since the whole organ is much larger than the scanner FOV, each brain is cut in several blocks (sized $\approx 3 \times 3 \times 5 \text{ mm}^3$). Each block is introduced in a rod transparent to X-Rays and is positioned on a 3-Axis translational + Z-Axis rotational sample-holder. Numerous FOV CT acquisitions are measured in a sequence (cf. Fig. 1(a-b)) for imaging the whole sample rod. Each FOV acquisition is a set of N_θ radiographs (2560×2160 pixels). According to rod dimensions, scanning resolution and the overlap required between adjacent FOV 3D tiles, the overall rod scan is composed of ≈ 360 CT acquisitions.

Each FOV CT acquisition is processed by a CT algorithm to reconstruct a 3D volume (i.e. tomogram sized

$2560^2 \times 2160$ voxels) imaging the inner features of the sample [Tof96, NW02]. For instance, Fig. 1 (c) shows acquired radiographs (strong absorption reveals Golgi-stained neuronal features), and Fig. 1 (d) is a 3D visualisation of a reconstructed tomogram. Once all tomograms have been reconstructed, the next data processing step consists of segmenting the brain features from the background (empty regions, non-stained parts of the brain) in order to perform further 3D analysis and visualisation of the mouse neuronal connectome. In that context, it is obvious that tomographic reconstruction and segmentation steps are of great significance since both high-quality and accuracy are required in tomograms to segment for achieving such a whole brain analysis.

However overall data acquisition of a whole brain needs to be performed as a routine application [SLH⁺23], thus requiring to drastically reduce both acquisition and CT reconstruction time. This limitation is practically addressed by reducing the number of acquired radiographs in CT acquisition, resulting to a degraded 3D tomogram. Such tomogram can no longer be segmented using traditional methods. Thus in this paper we investigate deep-learning based on encoder-decoder networks since it has been already demonstrated that they are well adapted to segmentation and / or denoising problems. We first introduce our overall data processing sequence and the positioning of our investigation in section 2 and we preface encoder-decoder based seg-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

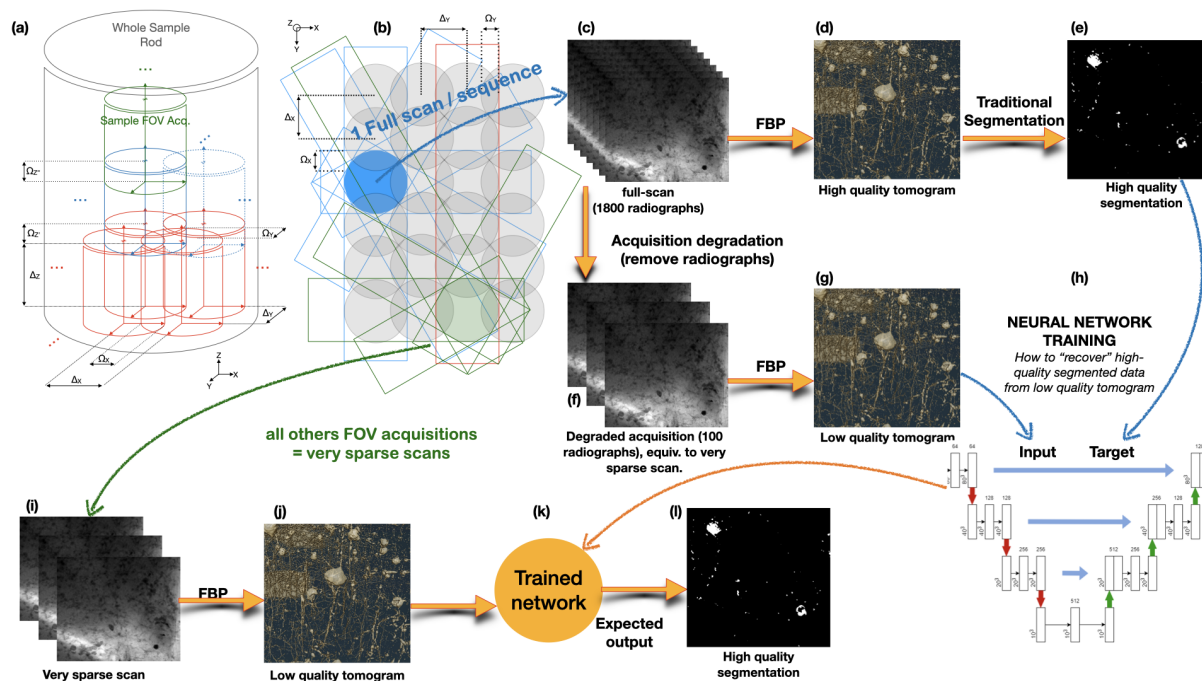


Figure 1: (a) Sample rod with several FOV areas. Translation steps (Δ) calibrated to optimize overlaps (Ω). (b) Acquisitions by X/Y-Axis translations: illustration of two FOV scans in blue and green. A unique full-scan (c) per sequence provides high-quality tomogram (d) and binarisation (e) by traditional segmentation. Degraded full-scan (f) provides degraded tomogram (g). The pair $\{(g), (e)\}$ = (input, target) training dataset (h). All other sequence scans done in very sparse configuration (i), providing 3D volumes (j), that are processed by the trained network (k). Expected output (l) is a high-quality segmentation from degraded tomograms.

mentation methods in section 3. In section 4 we design two networks more suitable to our use-case and explain our design choices according to our data. Before concluding, all networks are experimented in section 5 and we discuss why our networks design specificities lead to better results compared to the state-of-the-art techniques.

2 EXPERIMENTAL DATA PROCESSING

We present in this section the overall data processing sequence developed in this study and the positioning of investigated deep-learning methods (cf. Fig. 1).

Traditional tomographic reconstruction is the Filtered Back Projection (FBP), widely used in both medical and industrial CT scanners and appreciated for its easy implementation and fast computation [Han81]. However, FBP is very sensitive to the noise in the acquisitions and the tomogram quality depends on the number of radiographs (cf. image degradation with the number of radiographs on Fig. 2). Indeed, according to Shannon-Nyquist theory, FBP requires at least $N_{opt} \approx 2000$ radiographs to optimally reconstruct slices sized $N^2 = 2560^2$ pixels [PGF⁺05]. In our case, $N_{\theta} = 2000$ leads to an excessive irradiation damaging bio-sample, but we can safely measure 1800 radiographs (denoted

full-scan in the following) without noticeable accuracy losses on the tomograms. However, due to the huge number of acquired data (≈ 360 FOV acquisition / rod, and ≈ 40 rods / mouse brain), both acquisition and data processing times have to be drastically decreased. The most practical solution consists of reducing the number of radiographs (fast-scan).

Such a fast-scan - i.e when $N_{\theta} \ll N_{opt}$ - leads to angular sampling problem which can be addressed thanks to iterative CT reconstructions (IR-CT). This domain of research has been amazingly fruitful and addressed for decades [HNY⁺13, HL89, GG96, And89, JW03, KS01]. Numerous IR-CT methods have been proposed [DMND⁺00, ZWZ⁺18, RFK⁺14], including radiograph ordering optimizations [Kol05, KB98, EF99], multi-scale / multi-grid methods [MKL⁺16] or GPU-based implementations [SKKH07, RXT07, ZHZ09].

Furthermore, FOV tomography (also denoted interior / region-of-interest / local tomography) can be addressed by dedicated IR-CT methods [ZNG08, HGD⁺10, LHW⁺15, PM17, SKR⁺14]. Despite their efficiency, IR-CT techniques are sparsely deployed in routine applications because of both their complexity and their computation time ($\times 10$ to $\times 100$ FBP requirements). This latter limitation makes them ultimately unrealistic to be used in our project.

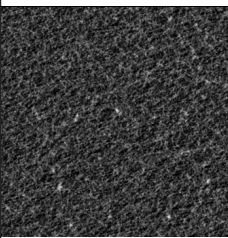
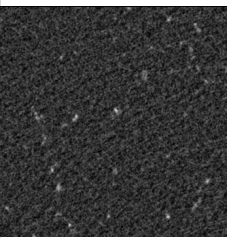
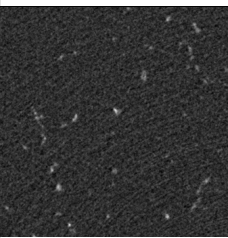
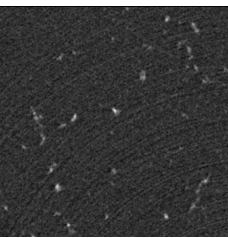
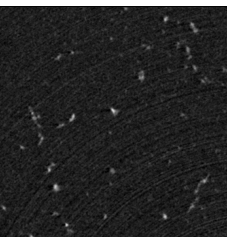
| N_θ | 100 | 300 | 600 | 900 | 1800 |
|------------|---|---|---|--|---|
| |  |  |  |  |  |
| Dice | 0.02 | 0.84 | 0.93 | 0.96 | 1.00 (ref.) |

Figure 2: Reconstructed slices from acquisitions with 900 to 100 radiographs, compared to reference (ref.) full-scan image ($N_\theta = 1800$). Noise and artefacts increase when scanning becomes faster. Dice is obtained by adaptive Renyi entropy based segmentation (compared to segmentation result from full-scan).

Golgi-staining leads to high contrasted features (foreground) on full-scan reconstructed slices (cf. bright features on Fig. 1(d)). Thanks to such a sample preparation, empty regions of the brain are very low intensity and almost confounded with empty space (i.e. both can be considered as background). In that condition, a traditional (binary) segmentation such that an adaptive Renyi entropy method [San04, KR14, SRR08, FZL⁺17] conveniently extracts neuronal features even if they are widely unbalanced (less than 3% of a tomogram are neuronal connectome in our case). Unfortunately, traditional segmentation accuracy directly depends on the tomogram quality, and thus significantly decreases when a fast-scan is processed. As illustrated on Fig. 2, we have for instance experimented that about 30% (*resp.* 50%) of the segmented features are lost when the segmentation is processed from a tomogram reconstructed with 300 (*resp.* 100) radiographs (compared to the segmentation result obtained from full-scan tomogram).

Alternatively to traditional approaches, we investigate deep-learning networks for accurately segmenting degraded tomograms. Our bio-sample acquisition produces massive CT data from a sequence in which all FOV CT scans are exactly measured in the same conditions and on the same sample. This high repeatability is particularly suitable for deep learning. We thus apply the following data processing protocol (cf. Fig. 1): i) only one full-scan FOV CT (c) is acquired per sequence to obtain a high-quality tomogram (d) and its corresponding high-quality segmentation (e) ; ii) the full-scan is degraded (remove radiographs) to reach very sparse scanning configuration (f) ; iii) the resulting degraded tomogram (g) is combined to the segmentation (e) to provide the pair (input, target) feeding the investigated neural network training ; then, iv) all the other FOV regions are acquired in fast-scan conditions (j), providing degraded tomograms (j) which are segmented using the trained network. One may note that the training dataset is automatically obtained thanks to the full-scan combined with a traditional data process-

ing pipeline, leading to an overall processing sequence which is fully automated.

3 RELATED WORKS

Several efficient Convolutional Neural Networks (CNN) have been proposed to address segmentation task. The first network architecture has been proposed by Long *et al.* [LSD15]. It is a Fully Convolutional Network (FCCN) able to achieve a semantic segmentation of natural images. However in these primary works, pixel classification is considered inadequate to achieve accurate segmentation since pixel localisation, which is essential, is ignored. Thus Ronneberg *et al.* have proposed a new architecture called U-Net [RFB15] to address this limitation. U-Net is a CNN based on encoder-decoder operations: i) the encoder compresses the input data into a latent-space representation, and, ii) the decoder aims at predicting the output from the latent-space representation [BKC17].

The U-Net design is one of the most well-known architecture for segmentation. A lot of variants have been proposed to address several applications, such as segmentation of *Drosophilya* cells in microscopy images [RFB15] or road recognition in natural images [ZLW18], for instance. U-Net network also address 3D segmentation thanks to 3D convolutional layers. For example, Cicek *et al.* [ÇAL⁺16] have designed a 3D U-Net with a ReLU activation function to segment kidney in confocal microscopic images.

The reason why U-Net and its variants are suitable for medical image segmentation is that its structure can simultaneously combine low and high level information. The low-level information aims at improving accuracy while the high-level information helps to extract complex features [LSLZ21]. Classical U-Net architecture is based on a binary cross-entropy (BCE) [RFB15] loss function suitable for balanced classes. In our study case, class distribution is very imbalanced so that using such a loss function could lead to roughly classifying background only. A DICE loss function is more dedi-

| | U-Net [RFB15] | V-Net [MNA16] | RED-CNN [HZRS16] | RED-Net [MSY16] |
|--|------------------------|------------------------|------------------------|------------------------|
| Original Application | Segmentation | Segmentation | Denoising | Restoration |
| Original Dimension | 2D | 3D | 2D | 2D |
| 3D Capabilities | Yes | Yes | Yes | Yes |
| Activation function | ReLU | PReLU | ReLU | ReLU |
| Sampling function (encoder path) | Maximum Pooling | Convolution | Convolution | Convolution |
| Sampling function (decoder path) | Transposed Convolution | Transposed Convolution | Transposed Convolution | Transposed Convolution |
| Skip connectors between encoder & decoder | Vector concatenation | Vector concatenation | Scalar sum | Scalar sum |
| Feature map size variations | Yes | Yes | No | No |
| Number of floors | 4 + Bottleneck | 4 + Bottleneck | – | – |
| Total number of Convolutional layers | 28 | 23 | 10 | 10 to 30 |
| Loss function | (C) BCE | (G) DICE | (A) MSE | (B) MSE |
| Other experimented loss | (F) DICE | - | (D) DICE | (E) DICE |

Table 1: Encoder-decoder characteristics. Notations (A) to (G) are refers to discussion Table 5.

cated to our use-case since it only scores estimated foreground pixels. This U-Net variant, denoted V-Net, has been first proposed by Milletari *et al.*[MNA16]. Another major design difference between U-Net and V-Net consists of using convolution layers instead of pooling layers to down/up-sample between each level of the network.

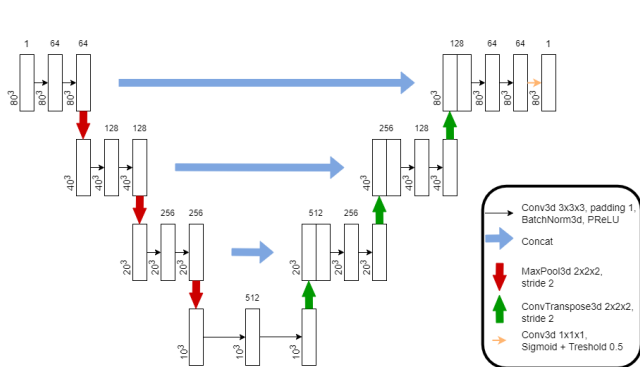
Residual Networks[HZRS16] are deep learning networks mainly made of convolutions and skip connections. Their architecture is based on a deep succession of convolutional layers which have empirically showed an accuracy gain over shallower networks when coupled with skip connections. These skipped connections take places between two not successive layers to address the vanishing/exploding gradients problem by adding a shallow residual mapping to a deeper layer input. Our investigation also focuses on RED-CNN and RED-Net which are residual networks, but also encoder-decoder such as the previously introduced U-Net and V-Net. RED-CNN (Residual Encoder-Decoder Convolutional Neural Network) has been developed for low-dose CT imaging. It has been demonstrated in [HZRS16] that such a network is particularly efficient for preserving structural information while reducing noise. Despite RED-CNN has been proposed for segmenting tomograms reconstructed from a large amount of low-dose X-Ray radiographs, we investigate this network since the noise / artefacts it deals with are quite similar to those observed on our degraded tomograms. We also focus on the capabilities of the RED-Net network which is a very deep fully convolutional encoder-decoder network for image restoration. Since denoising is part of the very de-

graded CT tomogram segmentation problem, RED-Net architecture could provide some ideas to address our task. Usually, RED-Net performs ten to thirty successive convolutions (depending on the version), and is also coupled with skip connections. A summary of the characteristics of all the state-of-the-art networks (U-Net, V-Net, RED-CNN, RED-Net) explored in this study is provided in the Table. 1.

4 OPTIMAL ENCODER-DECODER NETWORKS

In addition to the encoder-decoder networks detailed in section 3, we introduce in this section two new encoder-decoder networks specifically addressing the segmentation of very degraded brain tomograms (Fig. 3). We justify our architectural choices in accordance with our data and the other networks architectures. Each level of the encoder path is composed of a succession of convolutional layers + batch normalization + PReLU activation function ; maximum pooling (network I) or convolutions with strides (network H) are applied between each down-sampling step. Symmetrically, the decoder part is composed of a succession of transposed convolutional layers with concatenations of symmetric encoder feature maps. The final up-sampling step of the decoder is finalized by a sigmoid function coupled with a binary threshold in order to achieve the segmentation.

On top of that, batch normalization is added to speed up the training process enabling a higher learning rate [IS15], while PReLU activation function generalizes the traditional rectified unit and improves network fitting with nearly zero extra computational cost [HZRS15]. We investigated on two variations



| | |
|---|--|
| Our Application | Segmentation of widely unbalanced data in very noisy tomograms |
| Act. function | PReLU |
| Sampling function (encoder path) | Batch normalization + (H) Convolution (I) Maximum Pooling |
| Sampling function (decoder path) | Transposed Convolution |
| Skip connectors | Vector concatenation |
| Number of floors | 4 |
| Total number of Convolutional layers | 18 |
| Loss function | Sigmoid + DICE + Threshold |

Figure 3: New investigated networks architecture and summary of their principal characteristics.

of this architecture assuming: i) network (I) with maximum pooling will maximize Recall (Table. 4); ii) network (H) with convolutions will maximize Precision (Table. 4). Finally, the network minimizes a DICE loss function at the end since it has been already demonstrated it is suitable for accounting on very imbalanced foreground data. However, this minimization is completed with a sigmoid + threshold in order to achieve the binary segmentation. Finally, the overall networks can be designed to work in 2D to perform a segmentation slice by slice, or directly in 3D. Thus both designs have been implemented and results are discussed in the following section.

Second we experiment several level of pre-filtering over the training data to verify the consideration of imbalanced background during the training. All networks investigated in this paper are trained with: i) no filtering (i.e. all patches are considered), ii) > 0 filter (only the patches containing at least one foreground pixel are used), and, iii) $> 5\%$ filter (only the patches containing at least 5% of foreground pixels are used).

Prediction results are given as an illustration on Table 2 for our network (I), but similar behavior is observed for all networks. It leads to a performance superiority for trained networks excluding patches without any data of interest, slightly better than the classic not filtering process, but highly better than the 5% filtering. Considering the reduced amount of training data from a higher filtering rate, we investigated on this potential bias. Prediction results of the filtering process with a same number of training data are given Table 3. The ranking between the processing methods remains the same with still a huge gap between > 0 and $> 5\%$ filtering methods and a superiority of the > 0 filtering over the no filtering method. Training on too much background (no filtering) lowers the features of interest over background ratio while eliminating too much foreground ($> 5\%$ filtering) lowers the absolute features of interest information to train on. The > 0 filtering method take the fullest information while increasing the features of

interest over background ratio leading to higher performances.

5 DISCUSSION

In this section, we first compare our proposed networks to the state-of-the-art encoder-decoders through their segmentation performances from very degraded brain CT tomograms. Notice that all networks are trained using the optimal pre-filtering explained in previous section. Then, since network design can be considered in 2D or in 3D, we compare both design to estimate the quality gain of 3D (which is computationally widely more gloutonous, thus requiring HPC or GPU-box computers) over to 2D (rapidly achievable on a general purpose computer).

Metric comparisons are based on the standard confusion matrix composed of true positive (tp), true negative (tn), false positive (fp) and false negative (fn) ratio compared to ground truth. We also discuss the results using the Precision, Recall, Dice and Jaccard metrics, reminded in the Table 4. Precision (ratio of correctly segmented voxels among all predicted foreground) and Recall (ratio of correctly segmented foreground voxels compared to ground truth) are complementary: a high Precision coupled with a high Recall means a high-quality prediction of the network. Conversely, if one of them is not high enough, the prediction quality downgrades. Dice and Jaccard values are mentioned since they state the similarity and the diversity between predictions and ground truth.

As already mentioned, in all our experiments: i) full-scan are processed with 1800 radiographs while fast-scans only contain 100 radiographs ; ii) FBP is used for all CT reconstructions, and, iii) an adaptive Renyi entropy based segmentation is used from high-quality tomograms to obtain the training target and ground truths to be compared with network predictions.

5.1 Networks comparison

Prediction results are given on Table 5 for networks trained on fast-scans containing only 100 radiographs.

| Filters | nb patches | TP | FP | FN | TN | \mathcal{P} | \mathcal{R} | \mathcal{D} | \mathcal{J} |
|-----------|------------|-------|--------|-------|---------|---------------|---------------|---------------|---------------|
| no filter | 9680 | 18020 | 2280 | 10361 | 3066939 | 0.887 | 0.634 | 0.740 | 0.587 |
| > 0 | 7294 | 21261 | 3703 | 7120 | 3065516 | 0.852 | 0.749 | 0.797 | 0.663 |
| > 5% | 257 | 25118 | 222378 | 3262 | 2846842 | 0.101 | 0.885 | 0.182 | 0.100 |

Table 2: Processing filtered data (maximum data)

| Filters | nb patches | TP | FP | FN | TN | \mathcal{P} | \mathcal{R} | \mathcal{D} | \mathcal{J} |
|-----------|------------|-------|--------|-------|---------|---------------|---------------|---------------|---------------|
| no filter | 257 | 17979 | 14646 | 10402 | 3054573 | 0.551 | 0.633 | 0.589 | 0.418 |
| > 0 | 257 | 17764 | 10697 | 10617 | 3058522 | 0.624 | 0.626 | 0.625 | 0.455 |
| > 5% | 257 | 25118 | 222378 | 3262 | 2846842 | 0.101 | 0.885 | 0.182 | 0.100 |

Table 3: Processing filtered data (reduced data)

| Precision \uparrow | Recall \uparrow | Dice \uparrow | Jaccard \uparrow |
|----------------------------------|----------------------------------|---|-------------------------------------|
| $\mathcal{P} = \frac{tp}{tp+fp}$ | $\mathcal{R} = \frac{tp}{tp+fn}$ | $\mathcal{D} = \frac{2*tp}{2*tp+fp+fn}$ | $\mathcal{J} = \frac{tp}{tp+fp+fn}$ |

Table 4: Definition of Prediction, Recall, Dice and Jaccard metrics.

| Methods | TP | FP | FN | TN | \mathcal{P} | \mathcal{R} | \mathcal{D} | \mathcal{J} |
|-------------------------------|-------|---------|-------|---------|---------------|---------------|---------------|---------------|
| Trad | 28381 | 3067457 | 0 | 1762 | 0.009 | 1.0 | 0.018 | 0.009 |
| (A) RED-CNN (MSE) | 0 | 0 | 28381 | 3069219 | 0 | 0 | 0 | 0 |
| (B) RED-Net (MSE) | 0 | 0 | 28381 | 3069219 | 0 | 0 | 0 | 0 |
| (C) U-Net (BCE) | 0 | 0 | 28381 | 3069219 | 0 | 0 | 0 | 0 |
| (D) RED-CNN (DICE) | 0 | 0 | 28381 | 3069219 | 0 | 0 | 0 | 0 |
| (E) RED-Net (DICE) | 0 | 0 | 28381 | 3069219 | 0 | 0 | 0 | 0 |
| (F) U-Net (DICE) | 19194 | 2451 | 9186 | 3066769 | 0.887 | 0.676 | 0.767 | 0.623 |
| (G) V-Net (DICE) | 20132 | 3380 | 8249 | 3065839 | 0.856 | 0.709 | 0.776 | 0.634 |
| (H) Our Network (conv) | 18715 | 2212 | 9665 | 3067008 | 0.894 | 0.659 | 0.759 | 0.612 |
| (I) Our Network (maxP) | 21261 | 3703 | 7120 | 3065516 | 0.852 | 0.749 | 0.797 | 0.663 |

Table 5: Encoder-decoder performances overview (tomogram reconstructed with 100 radiographs)

In that case, the traditional segmentation performances lead to a nearly white-only pixel prediction. Among the nine encoder-decoder networks, five are not efficient and perform a black pixel only prediction: i) Networks (A) with BCE loss and (B) & (C) using MSE loss which are weak against hugely imbalanced classification tasks; ii) (D) RED-CNN and (E) RED-Net which are customized with a DICE loss but are primarily designed for denoising and restoration thus justifying their inefficiency to our use-case. The four working encoder-decoders are (F) U-Net customized with DICE, (G) V-Net, and our networks, (H) and (I), which lead to a DICE score of 0.759 to 0.797 (*resp.*) compared to the 0.018 for the traditional method. Each of the four networks is built using vector concatenation as skip connectors between encoder and decoder, coupled with feature map size variations and a DICE loss function. Note that it is not the case for any of the networks which does not work (i.e. (A), (B), (C), (D) and (E)).

Investigating on the design specificities of the four networks linked to their performance ranking, we were not able to highlight some evidence. Especially we ex-

pected: i) (F) and (I) networks with maximum pooling to maximize Recall; ii) (G) and (H) networks with convolutions to maximize Precision; iii) (H) and (I) networks with fewer floors to be ranked better as they managed to process more information in the network.

5.2 2D versus 3D DL segmentation

We investigate now our deep learning segmentation model using independent slices (2D segmentation of slices) and 3D (tomogram segmentation). As highlighted in Figure 4 and on the Table 6 metric comparison between 2D and 3D designs, there is a huge performance gap between 2D and 3D processing for our particular task. Processing 3D volumes brings shapes continuity, which should explain the performance superiority. Indeed, considering the tomogram as a whole results in 3D patches decomposition in the network model. Such a volumetric representation propagates the 3D morphological context of the features to segment through the latent space of the network, which would not be the case if the tomogram were considered slice by slice.

| Dimension | TP | FP | FN | TN | \mathcal{P} | \mathcal{R} | \mathcal{D} | \mathcal{J} |
|-----------|-------|-------|-------|---------|---------------|---------------|---------------|---------------|
| 2D | 25577 | 15043 | 31457 | 3025523 | 0.629 | 0.448 | 0.523 | 0.354 |
| 3D | 39630 | 4679 | 17404 | 3035887 | 0.894 | 0.694 | 0.782 | 0.642 |

Table 6: Tomogram processing (3D model) against independent slice processing (2D slice-by-slice segmentation)

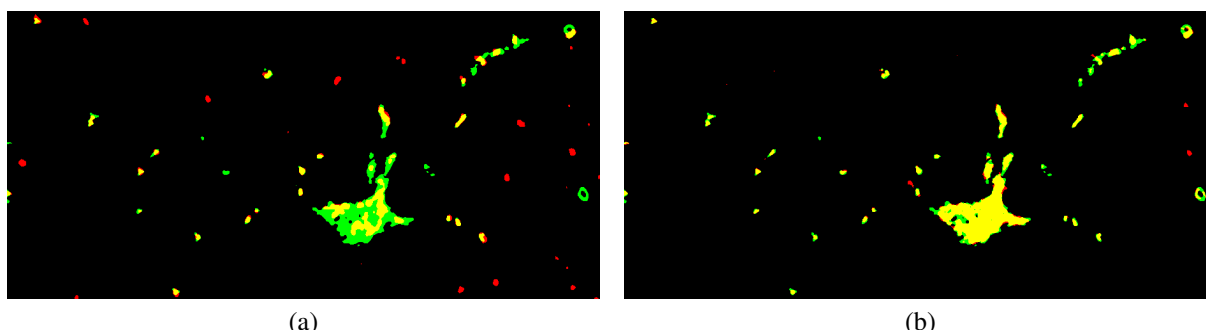


Figure 4: Comparison of 2D model result (a) and 3D model result (b) compared to ground truth (Yellow=TP, Red=FP, Green=FN, Black=TN).

6 CONCLUSION

Being able to reconstruct whole brain network from degraded tomograms remains a challenge as the traditional algorithms are not able to produce useful results. We have investigated several models of deep learning methods and proposed new approaches. From this whole set of experiments we can observe that encoder-decoder performances highly overcome traditional segmentation for 100 radiographs fast-scan processing and such models are good candidates to achieve our addressed task: providing a good quality segmentation from degraded tomograms reconstructed with a limited number of radiographs. More with a small number of layers the best networks are able to manage 3D images and allow to extract information coming from the whole brain conformation. Future works will investigate post processing methods to reconstruct the brain neuronal structures, for example by introducing graph concept into the deep-learning network designs in order to connect positive pixels, i.e. to directly recover neuronal structures instead of the subset of voxels composing it.

7 REFERENCES

- [And89] Anders H Andersen. Algebraic Reconstruction in CT from Limited Views, 1989.
- [BKC17] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [ÇAL⁺16] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- [DMND⁺00] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. Reduction of metal streak artifacts in x-ray computed tomography using a transmission maximum a posteriori algorithm. *IEEE Transactions on Nuclear Science*, 47(3):977–981, 2000.
- [EF99] Hakan Erdogan and Jeffrey A Fessler. Ordered subsets algorithms for transmission tomography. *Physics in Medicine & Biology*, 44(11):2835, 1999.
- [FZL⁺17] Yuncong Feng, Haiying Zhao, Xiongfei Li, Xiaoli Zhang, and Hongpeng Li. A multi-scale 3d otsu thresholding algorithm for medical image segmentation. *Digital Signal Processing*, 60:186–199, 2017.
- [GG96] Huaiqun Guan and Richard Gordon. Computed tomography using algebraic reconstruction techniques (ARTs) with different projection access schemes: a comparison study under practical situations. *Physics in Medicine and Biology*, 41(9):1727–1743, sep 1996.
- [Han81] Kenneth M Hanson. Technical aspects of computed tomography. *Radiology of the Skull and Brain*, 5(1):3941–3955, 1981.
- [HGD⁺10] Benoit Hamelin, Yves Goussard, Jean-Pierre Dussault, Guy Cloutier, Gilles

- Beaudoin, and Gilles Soulez. Design of iterative roi transmission tomography reconstruction procedures and image quality analysis. *Medical physics*, 37(9):4577–4589, 2010.
- [HL89] Tom Hebert and Richard Leahy. A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors. *IEEE transactions on medical imaging*, 8(2):194–202, 1989.
- [HNY⁺13] Jiang Hsieh, Brian Nett, Zhou Yu, Ken Sauer, Jean Baptiste Thibault, and Charles A Bouman. Recent Advances in CT Image Reconstruction. *Current Radiology Reports*, 1(1):39–51, 2013.
- [HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1024–1034, December 2015.
- [HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.
- [IS15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 448–456. PMLR, 2015.
- [JW03] Ming Jiang and Ge Wang. Convergence of the simultaneous algebraic reconstruction technique (sart). *IEEE Transactions on Image Processing*, 12(8):957–961, 2003.
- [KB98] Chris Kamphuis and Freek J. Beekman. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. *IEEE Transactions on Medical Imaging*, 17(6):1101–1105, 1998.
- [Kol05] J. S. Kole. Statistical image reconstruction for transmission tomography using relaxed ordered subset algorithms. *Physics in Medicine and Biology*, 50(7):1533–1545, 2005.
- [KR14] Satish Kumar and Gurdas Ram. A generalization of the Havrda-Charvat and Tsallis entropy and its axiomatic characterization. *Abstract and Applied Analysis*, 2014(5), 2014.
- [KS01] Avinash C. Kak and Malcolm Slaney. 7. Algebraic Reconstruction Algorithms. In *Principles of Computerized Tomographic Imaging*, pages 275–296. SIAM, 2001.
- [LHW⁺15] Minji Lee, Yoseob Han, John Paul Ward, Michael Unser, and Jong Chul Ye. Interior tomography using 1d generalized total variation. part ii: Multiscale implementation. *SIAM Journal on Imaging Sciences*, 8(4):2452–2486, 2015.
- [LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR/Computer Vision and Pattern Recognition)*, pages 3431–3440, June 2015.
- [LSLZ21] Xiangbin Liu, Liping Song, Shuai Liu, and Yudong Zhang. A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3):1224, 2021.
- [MKL⁺16] Glenn R Myers, Andrew M Kingston, Shane J Latham, Benoit Recur, Thomas Li, Michael L Turner, Levi Beeching, and Adrian P Sheppard. Rapidly converging multigrid reconstruction of cone-beam tomographic data. In *Developments in X-Ray tomography X*, volume 9967, page 99671M. International Society for Optics and Photonics, 2016.
- [MNA16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
- [MSY16] Xiaojiao Mao, Chunhua Shen, and Yubin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.
- [NW02] F. Natterer and Ge Wang. The Mathematics of Computerized Tomography. *Medical Physics*, 29(1):107–108, Jan 2002.

- [PGF⁺05] Thammanit Pipatsrisawat, Aca Gačić, Franz Franchetti, Markus Püschel, and José M.F. Moura. Performance analysis of the filtered backprojection image reconstruction algorithms. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume V, 2005.
- [PM17] Pierre Paleo and Alessandro Mirone. Efficient implementation of a local tomography reconstruction algorithm. *Advanced structural and chemical imaging*, 3(1):1–15, 2017.
- [RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [RFK⁺14] Benoit Recur, Mathias Fauconneau, Andrew Kingston, Glenn Myers, and Adrian Sheppard. Iterative reconstruction optimisations for high angle cone-beam micro-ct. In *Developments in X-Ray Tomography IX*, volume 9212, pages 288–299. International Society for Optics and Photonics, 2014.
- [RXT07] Dmitri Riabkov, Xinwei Xue, and Dave Tubbs. Accelerated cone-beam backprojection using GPU-CPU hardware. *Proceedings of the 9th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 68–71, 2007.
- [San04] Bulent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146, 2004.
- [SKKH07] Holger Scherl, Benjamin Keck, Markus Kowarschik, and Joachim Hornegger. Fast GPU-based CT reconstruction using the Common Unified Device Architecture (CUDA). In *IEEE Nuclear Science Symposium Conference Record*, volume 6, pages 4464–4466, 2007.
- [SKR⁺14] Emil Y Sidky, David N Kraemer, Erin G Roth, Christer Ullberg, Ingrid S Reiser, and Xiaochuan Pan. Analysis of iterative region-of-interest image reconstruction for x-ray computed tomography. *Journal of Medical Imaging*, 1(3):031007, 2014.
- [SLH⁺23] Anton PJ Stampfl, Zhongdong Liu, Jun Hu, Kei Sawada, H Takano, Yoshiki Kohmura, Tetsuya Ishikawa, Jae-Hong Lim, Jung-Ho Je, Chian-Ming Low, et al. Synapse: An international roadmap to large brain imaging. *Physics Reports*, 999:1–60, 2023.
- [SRR08] Ahmad Adel Abu Shareha, Mandava Rajeswari, and Dhanesh Ramachandram. Textured renyi entropy for image thresholding. In *Proceedings - Computer Graphics, Imaging and Visualisation, Modern Techniques and Applications, CGIV*, pages 185–192, 2008.
- [Tof96] Peter Toft. The radon transform. *Theory and Implementation (Ph. D. Dissertation)(Copenhagen: Technical University of Denmark)*, 1996.
- [ZHZ09] Xing Zhao, Jing-jing Hu, and Peng Zhang. Gpu-based 3d cone-beam ct image reconstruction for large data volume. *International Journal of Biomedical Imaging*, 2009:1–8, 2009.
- [ZLW18] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.
- [ZNG08] Andy Ziegler, Tim Nielsen, and Michael Grass. Iterative reconstruction of a region of interest for transmission tomography. *Medical physics*, 35(4):1317–1327, 2008.
- [ZWZ⁺18] Hao Zhang, Jing Wang, Dong Zeng, Xi Tao, and Jianhua Ma. Regularization strategies in statistical image reconstruction of low-dose x-ray CT: A review, 2018.