Leveraging 3D Information in Unsupervised Brain MRI Segmentation

Benjamin Lambert[1], Maxime Louis[1], Senan Doyle[1], Florence Forbes[2], Michel Dojat[3], Alan Tucholka[1]

[1] Pixyl, Research and Development Laboratory, 38000 Grenoble, France

[2] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

[3] Univ. Grenoble Alpes, Inserm, U126, Grenoble Institut Neurosciences, GIN, 38000 Grenoble, France

IEEE - ISBI 2021 International Symposium on Biomedical Imaging

PIXYL https://pixyl.ai/

PURPOSE / OBJECTIVES

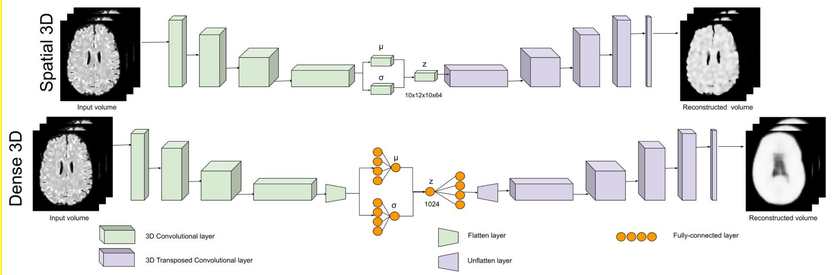
Brain anomalies can manifest in a wide range of shape, location and intensities. To automatically detect them, most state-of-the-art algorithms use a supervised approach, trained on an as large as possible manually annotated dataset, a strenuous task. To alleviate the need for annotated datasets, Unsupervised Anomaly Detection (UAD) methods are proposed, detecting anomalies as outliers of a healthy model learned using a Variational Autoencoder (VAE). Previous work on UAD adopted 2D approaches, meaning that MRIs are processed as a collection of independent slices. Yet, it does not fully exploit the spatial information contained in MRI.

In this study, we propose to perform UAD in a 3D fashion and compare 2D and 3D VAEs. Training is performed on healthy brain scans using a new loss function that we present as a side contribution. Our models are then evaluated on Multiple Sclerosis (MS) and Tumors scans.

METHODS

We use VAEs to perform UAD. These architectures are composed of an encoder followed by a decoder. The encoder compresses the input scan into a normal distribution from which we sample a latent vector z. This vector is then presented to the decoder which outputs a reconstruction of the scan.

We investigate two types of VAE architectures : Spatial and Dense VAEs, differing by the type of encoding. In the spatial configuration, input scans are encoded into a multi-dimensional tensor whereas in the dense configuration, they are encoded into 1-dimensional latent vectors, meaning that spatial information is discarded. To obtain 3D adaptations, we replaced 2D convolutions by their 3D counterparts.



Spatial and Dense 3D VAES. Latent vectors z have a shape of 10x12x10x64 in the spatial configuration, and a shape of 1024 in the dense configuration.

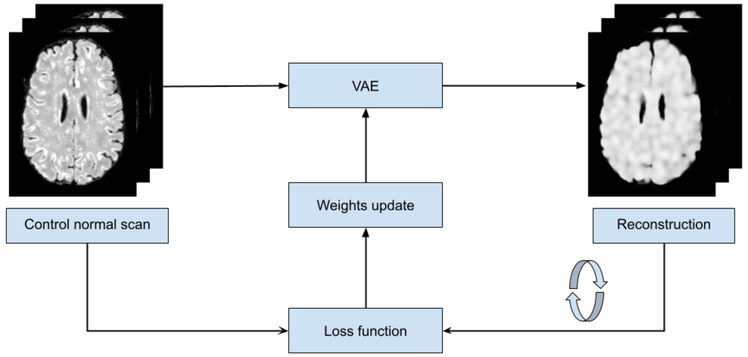
VAEs are trained by reconstructing healthy scans. By doing so, they learn a manifold of healthy scans. To alleviate instabilities in the training process, we propose a variation of the VAE usual loss function, which we named the collapsing robust loss function, defined as :

$$L_T = \frac{\|X - \hat{X}\|_1}{\Sigma} + \beta(t) D_{KL}(q_\phi(z|X)\,|\,P(z)) \text{ with } \beta(t) = \begin{cases} \frac{2t}{T} & \text{for t} \in [0, \frac{T}{2}[\\ 1 & \text{else} \end{cases} \text{ and } \beta(t+T) = \beta(t)$$

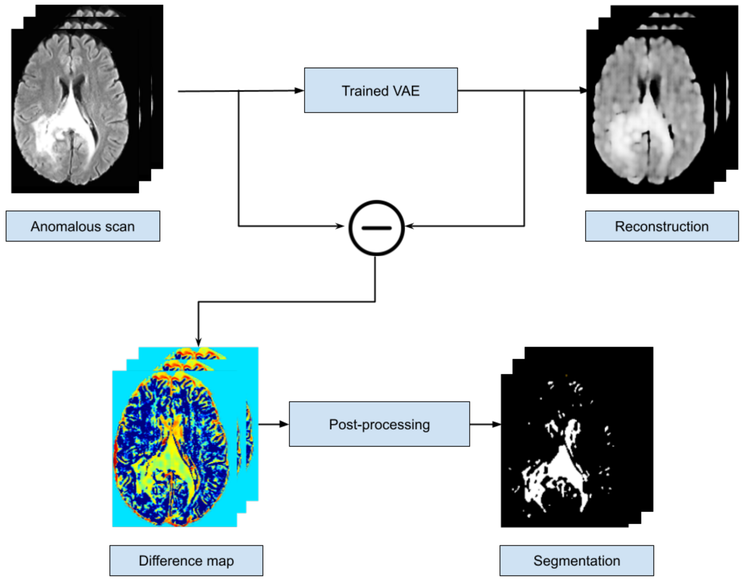
First term of this equation is the reconstruction term, used to estimate the distance between the input scan and its reconstruction. We normalize this term by its moving mean Σ, computed on its L last values. Second term is a regularization term on the latent space, ensuring that scans are encoded into normal distributions. To prevent the collapsing of this term, we add a factor that cyclically anneal it with a period T.

We train our VAEs using this loss on 79 healthy MRI FLAIR. To evaluate our models, we use 196 scans with White-Matter Hyper-intensities (WMH) and 100 scans with tumors.

TRAINING PIPELINE



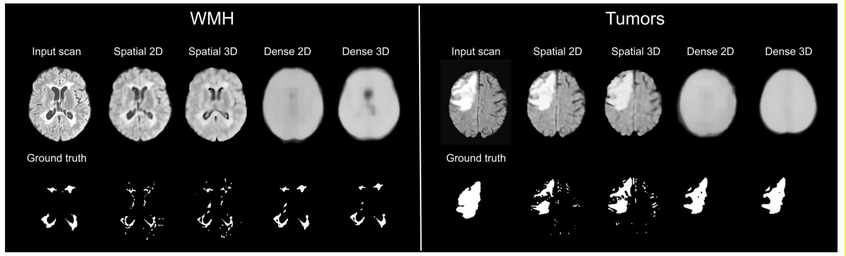
INFERENCE PIPELINE



RESULTS

Segmentation performances are evaluated using Dice, sensitivity, and specificity. Reconstruction performance is estimated using the voxel-wise mean average error (MAE) between the input scan and its reconstruction. To fairly compare 2D and 3D networks, they are all evaluated on the same 75 central slices for each test scan. Finally, to compare our unsupervised method with classic supervised DL approaches, we trained 3D U-Nets in a 4-fold cross-validation setting, on each test dataset.

3D adaptations demonstrate a boost in performance as compared to 2D. On both the WMH and Tumors datasets, the best network is the Dense 3D VAE.



Reconstructions and segmentations produced by our networks.

	WMH				Tumors			
Model	Dice	Spe	Sen	MAE	Dice	Spe	Sen	MAE
Spatial 2D	0.280 ± 0.174	0.986	0.547	0.049	0.260 ± 0.121	0.912	0.376	0.079
Spatial 3D	0.336 ± 0.203	0.989	0.605	0.056	0.386 ± 0.160	0.898	0.655	0.065
Dense 2D	0.460 ± 0.262	0.994	0.604	0.102	0.618 ± 0.167	0.978	0.676	0.133
Dense 3D	$\mathbf{0.463 \pm 0.259}$	0.996	0.511	0.104	$\mathbf{0.650 \pm 0.190}$	0.981	0.711	0.151
Supervised 3D	0.750 ± 0.097	0.992	0.789	–	0.740 ± 0.174	0.999	0.738	–

Performance metrics obtained on the test datasets. Spe = Specificity; Sen = Sensitivity, MAE = Mean Average Error.

CONCLUSION

In this study, we extend UAD to 3D and demonstrate the interest of 3D VAEs over their 2D counterparts. Our framework allow to process whole MRI volumes without the need of manually extracting slices. Our key messages are presented below :

- The gain in performance is the most significant for Spatial networks, which benefit from the introduction of additional spatial context with the use of 3D. For dense networks, which discard spatial information during the encoding-decoding scheme, this gain is less important.
- We observe a tradeoff between reconstruction and segmentation performances. Dense networks that produce rawer reconstructions also demonstrates better segmentation performances. Indeed, anomalies are entirely discarded, which make them more easily detectable.
- Despite alleviating the need for annotated datasets, performance of our UAD approaches still remains lower than standard supervised approaches.