

Unsupervised Anomaly Segmentation in Brain MR Images using Context Variational Autoencoder

Albert Gubaidullin
University of Bonn, Department of Computer Science

Abstract

An MRI is an important part of a medical examination of the brain. It can visualise several brain disorders. Good visualisation makes it easy to comprehend by a doctor, but it is a hard task for a machine. This article proposes an unsupervised algorithm that can outline a brain tumor by reconstructing model-internal representation of the brain without a tumor and comparing it to the actual MRI scan. The main part of the algorithm is a deep learning-based variational autoencoder. It is trained with augmented healthy brain MRI scans from BraTS 2019 dataset for which, the autoencoder learns to reconstruct the healthy images. Given unhealthy brain MRI scan, i.e. image containing tumor, autoencoder returns a healthy version of it. By calculating residual between two images and applying Gaussian Mixture Model we are able to outline the tumor on a scan.

1. Introduction

Brain MRI are one of the most important instruments for detecting and diagnosing brain disorders. It can detect a variety of conditions of the brain such as cysts, tumors, bleeding, swelling, developmental and structural abnormalities, infections, inflammatory conditions, or problems with the blood vessels. It can determine if a shunt is working and detect damage to the brain caused by an injury or a stroke. The biggest problem of this approach for computer science is representation of the results. Despite great visualisation, complexity of a brain internal structure makes it challenging to determine presence or absence of any anomalies on the scan. Development of new approaches in this field can save physicians time and provide an accurate reproducible solution for further analysis and monitoring.

Recently, deep learning based segmentation techniques proved to be at least as useful as traditional computer vision methods. Convolutional neural networks (CNN), for example, are able to learn from examples and demonstrate state-of-the-art segmentation accuracy both in 2D natural images and in 3D medical image modalities. However this is a supervised learning model, it needs tremendous amount of annotated data, which is not usually available. Unsupervised learning-based models, on the other hand, have also shown a great potential in detecting anomalies in brain MRI analysis. The advantage of this approach is to use raw data without additional labeling. There are many articles proving, that this algorithms are working and results are constantly enhancing.

Unsupervised anomaly detection won't outline the anomaly by itself, but it can provide a basis for detecting lesion contours. These unsupervised deep learning methods mostly rely on the use of various types of autoencoders (AEs) and generative adversarial networks (GANs).

Since all healthy adult brains show similar structures, we can teach the an autoencoder to learn it's internal internal structure. In this approach a VAE learns to denoise augmented brain scans and given an unhealthy scan, should return healthy version of the image. By subtracting original image and generated one, we can outline the tumor, if it exists.

2. Related Work

Last years we can see a constant improvements in a field of image analysis, especially in field of medical imaging analysis. Unfortunately, the problem of brain anomalies detection based on MRI scans is still a big challenge. We have seen a lot of approaches and models, trying to solve this problem. In [HDWF*17] was one of the first mentions of using Convolutional Neural Networks for tumor segmentation. Models their proposed in the article used a supervised techniques and requires a huge amount of annotated data to get a decent result.

Different AEs have been applied for anomaly detection in brain images. [BWAN18] applied VAEs and used the reconstruction error for localization of Multiple Sclerosis lesions on an in-house MRI dataset. In their experiments an adversarial reconstruction loss slightly outperformed a vanilla VAE. [CK18] show that a combination of a AE with an adversarial loss on the latent variables can boost performance in detecting brain tumors in the BraTS 2015 MRI dataset using a pixel-wise reconstruction error.

[VTAK16] used a single layer autoencoder with noise reduction to detect and segment glioma in MR images of the brain. Test data reconstruction error patches have been used to detect areas containing brain lesions.

3. Data

The BraTS 2019 dataset were used for training, validation and testing of the network. It consists of 335 patients from multiple institutions and is further divided into 76 low-grade (LGG) and 259 high-grade (HGG) glioma. For each patient the T1, T1CE, T2 and FLAIR sequences as well as a segmentation map are available. The data shows a high variability between patients, as different scanners with varying degrees of quality are used.

The original data has resolution 240 x 240 pixels with 150 slices for each brain scan. After checking that certain amount of pixels is taken by the image of the brain and the slice is around the center, the image is added to the dataset. Approximately half of slices are taken, for each brain scan. Then images get min-max normalized, according to the formula

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

where $x = (x_1, \dots, x_n)$ and z_i is i -th normalized data.

Given a segmentation mask of glioma, dataset is divided into *healthy* and *unhealthy* datasets. Unhealthy images are used in final testing. Healthy ones are divided into training, validation and testing set for autoencoder training. Finally, all images go through augmentation process.

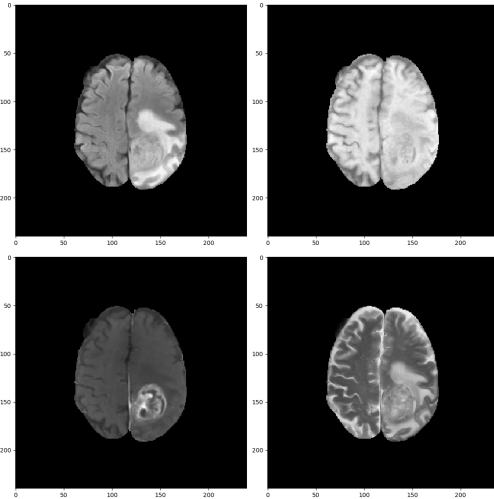


Figure 1: Raw unhealthy brain scans from the (Clockwise from top left: Flair, T1, T1-weighted, T2)

4. Methods

4.1. Autoencoders

An Autoencoder (AE) is a multi-layer neural network structure that is trained to reconstruct its input x from a learned representation z . It consists of two components: encoder and decoder. An encoder $z = h(x)$, encodes the high-dimensional input x to a latent low-dimensional feature representation z . A decoder $r = g(z)$ reconstructs the original input by decoding the representation z .

4.2. Context Autoencoders

A Denoising Autoencoder (DAE) is trained to reconstruct the image from a noisy input. This usually results in more robust and perturbation invariant representations. The most commonly used noise is additive Gaussian noise.

Context Encoders (CEs) are a special class of DAEs where instead of Gaussian noise local patches of the input are masked out. This type of Autoencoder tends to learn semantic aspects of given input. [PKD*16].

4.3. Variational Autoencoders

The mathematical basis of Variational Autoencoders (VAEs) has relatively little to do with classical autoencoders. They are called “autoencoders” only because the final network setup does have an encoder and a decoder, and resembles a traditional autoencoder

Instead of mapping the input into a fixed vector, VAEs maps it into a distribution p_θ , parameterized by θ . [KW19] Assuming that we know the real parameter θ^* for this distribution, in order to generate a sample that looks like a real data point x_i , we follow these steps:

- Sample a z_i from a prior distribution $p_{\theta^*}(z)$
- Then a value x_i is generated from a conditional distribution $p_{\theta^*}(x|z = z_i)$.

Data generation process can be represented by the following formula:

$$p_\theta(x_i) = \int p_\theta(x_i|z)p_\theta(z)dz$$

Unfortunately it is not easy to compute $p_\theta(x_i)$ in this way, as it is very expensive to check all the possible values of z and sum them up. To narrow down the value space to facilitate faster search, we need a new approximation function to output what is a likely code given an input x , $q_\phi(z|x)$, parameterized by ϕ .

4.4. Loss function

The estimated posterior $q_\phi(z|x)$ should be very close to the real one $p_\theta(z|x)$. We can use Kullback-Leibler divergence to quantify the distance between these two distributions. KL divergence $D_{KL}(X||Y)$ measures how much information is lost if the distribution Y is used to represent X .

In our case we want to minimize $D_{KL}(q_\phi(z|x)||p_\theta(z|x))$ with respect to ϕ .

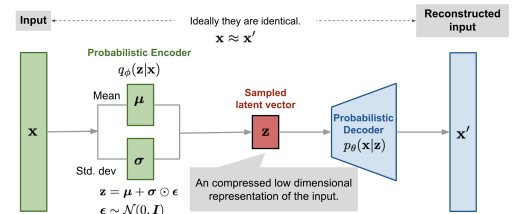


Figure 2: CVAE structure.

5. Experiments

5.1. Preprocessing

To train Autoencoder to denoise unhealthy brain image, it should be trained on noisy augmented images. To imitate glioma, random shapes of different intensity of colour and different sizes were added to the image. Types of shapes used are squares and circles, colour randomly chosen in range from 0 to 1, sizes are random from 15 to 40 pixels, there can be one or multiple shapes on a single image. Position is determined randomly by normal distribution in order to make a cluster of shapes around the center. Figure 1 represents an example of average brain scan from the dataset.

5.2. Network architectures

Different model parameters were tried on a small datasets of 10-20 images. Three models with better results were tested on full BraTS 2019 dataset:

- Denoising Autoencoder (DAE).
- Context Variational Autoencoder (CVAE)
- Context Variational Autoencoder with dilation filters (dCVAE)

Exact model structures can be found in Table 1 and Table 2

DAE
conv3, 1,1-64
conv3, 1,1-64
MaxPool
conv3, 1,1-128
conv3, 1,1-128
MaxPool
conv3, 1,1-256
conv3, 1,1-256
MaxPool
conv3, 1,1-512
MaxPool

Table 1: Encoder architecture. In the encoder, conv designates a convolutional block consisting of one convolutional layer, batch-norm normalization and LeakyReLU activation function. The conv-layer is denoted as "conv [kernel size], [stride], [dilation]-[number of channels]". The decoder is directly transposed to this architecture, using fractionally strided convolutional layers.

The proposed method, utilises Context Variational Autoencoder proposed by [ZKP*18] to extract features from input data. By combining CEs and VAEs, method aims at capturing a better-calibrated reconstruction error and model-internal variations, to yield more complete estimates of anomaly.

Both CVAE uses four 2D-Conv-Layers and 2D-Transposed-Conv-Layers respectively each layer followed by a LeakyReLU function and Batch normalization. The encoder and decoder are symmetric with 16, 32, 64, 64 feature maps for CVAE and 8, 16, 16, 16 for dCVAE This exact parameters were chosen similarly to [WKM19] Encoders, with the last layer having two heads, one predicting the mean, and one predicting the log standard-deviation.

CVAE	dCVAE
conv4, 2,1-16	conv6, 1,1-8
conv4, 2,1-32	conv5, 2,2-16
conv4, 2,1-64	conv6, 1,2-16
conv4, 2,1-64	conv5, 3,2-16
Latent space	
conv3, 2,1-64	conv3, 1,1-16
conv3, 2,1-64	conv3, 1,1-16

Table 2: Encoder and latent space architecture. In the encoder, conv designates a convolutional block consisting of one convolutional layer, batch-norm normalization and LeakyReLU activation function. The conv-layer is denoted as "conv [kernel size], [stride], [dilation]-[number of channels]". The decoder is directly transposed to this architecture, using fractionally strided convolutional layers. The two layers of the latent space describe $\mu(X)$ and $\Sigma^{0.5}(X)$. No normalization or activation functions are used here.

Objective function is represented by sum of KL divergence and MSELoss:

$$L_{ceVAE} = L_{KL}(f_{\mu}(x), f_{\sigma}(x)^2) + L_{recVAE}(x, g(z)) + L_{recCE}(x, g(f_{\mu}(\tilde{x}))),$$

5.3. Postprocessing

After getting the denoised version of the brain scan from the network, it gets min-max normalized to match the input. Then residual between output and input is calculated.

Then Gaussian mixture model (GMM) with a neighborhood-based weight is applied to the image. There are three components are expected from the the residual brain image: healthy brain mass, tumor and background. Between these clusters tumor should be the smallest one.

Since results of the models are noisy, morphological operations were applied to improve results. Firstly binary closure were added to fill in possible holes, then all small pixel groups were removed from GMM output.

6. Results

Evaluating the performance of an anomaly detection algorithm is a challenging problem. Intersection over Union (IoU) method were used in final results evaluation. According to this metric, CVAE proved to be the best model for brain tumor segmentation. Most probably this results could be further improved by additional post processing. Final results are presented in Table 3.

Model	IoU score
DAE	0.326
CVAE	0.417
dCVAE	0.392

Table 3: Final results

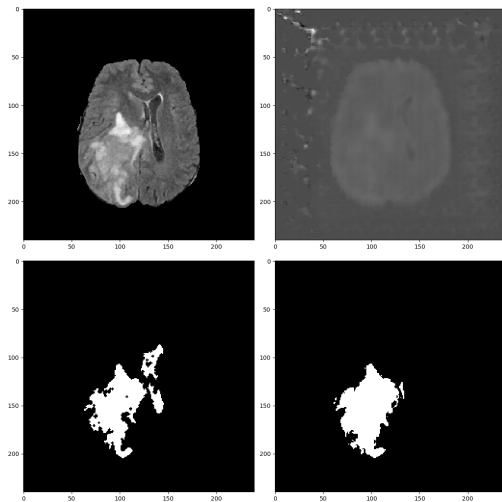


Figure 3: Example results of the CVAE model (Clockwise from top left: Input, Raw output, Final result, Ground truth)

In this Lab was presented CVAE for unsupervised anomaly detection, combining CEs with VAEs for unsupervised training and detection as well as localization of anomalies in medical images. The result shows that presented models were able to outline a tumor with acceptable error.

It might be an interesting future direction to see how different reconstruction losses can boost the performance or interpretability. Unsupervised learning proved to be a reliable solution for brain tumor segmentation problem. The results are promising and have the potential to improve and speed up the future inspection and evaluation of medical images,

References

- [BWAN18] BAUR C., WIESTLER B., ALBARQOUNI S., NAVAB N.: Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. *ArXiv abs/1804.04488* (2018). [1](#)
- [CK18] CHEN X., KONUKOGLU E.: Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. *ArXiv abs/1806.04972* (2018). [1](#)
- [HDWF*17] HAVAEI M., DAVY A., WARDE-FARLEY D., BIARD A., COURVILLE A., BENGIO Y., PAL C., JODOIN P.-M., LAROCHELLE H.: Brain tumor segmentation with deep neural networks. *Medical Image Analysis* 35 (2017), 18 – 31. URL: <http://www.sciencedirect.com/science/article/pii/S1361841516300330>, doi:<https://doi.org/10.1016/j.media.2016.05.004>. [1](#)
- [KW19] KINGMA D. P., WELING M.: An introduction to variational autoencoders. *Found. Trends Mach. Learn.* 12 (2019), 307–392. [2](#)
- [PKD*16] PATHAK D., KRÄHENBÜHL P., DONAHUE J., DARRELL T., EFROS A. A.: Context encoders: Feature learning by inpainting. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 2536–2544. [2](#)
- [VTAK16] VAIDHYA K., THIRUNAVUKKARASU S., ALEX V., KRISHNAMURTHI G.: Multi-modal brain tumor segmentation using stacked denoising autoencoders. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Cham, 2016), Springer International Publishing, pp. 181–194. [1](#)

[WKM19] WENINGER L., KRAUHAUSEN I., MERHOF D.: In *Eurographics Workshop on Visual Computing for Biology and Medicine* (2019), The Eurographics Association, pp. 45–49. doi:[10.2312/vcbm.20191230](https://doi.org/10.2312/vcbm.20191230). [3](#)

[ZKP*18] ZIMMERER D., KOHL S. A. A., PETERSEN J., ISENSEE F., MAIER-HEIN K.: Context-encoding variational autoencoder for unsupervised anomaly detection. *ArXiv abs/1812.05941* (2018). [3](#)