

GLIOBLASTOMA SYNTHESIS AND SEGMENTATION WITH 3D MULTI-MODAL MRI: A STUDY USING GENERATIVE ADVERSARIAL NETWORKS

Edmond Wang

Westminster School, London, United Kingdom

ABSTRACT

The Grade IV cancer Glioblastoma is an extremely common and aggressive brain tumour. It is of significant consequence that histopathologic examinations should be able to identify and capture the tumour's genetic variability for assistance in treatment. The use of Deep Learning - in particular CNNs and GANs - have become prominent in dealing with various image segmentation and detection tasks. The use of GANs have another importance - to expand the available training set by generating realistic pseudo-medical images. Multi-modal MRIs, moreover, are also crucial as they lead to more successful performances. Nonetheless, accurate segmentation and realistic image synthesis remain challenging tasks. In this study, the history and various breakthroughs/challenges of utilising deep learning in glioblastoma detection is outlined and evaluated. To see networks in action, an adjusted and calibrated Vox2Vox network - a 3D implementation of the Pix2Pix translator - is trained on the biggest public brain tumour dataset BraTS 2020. The experimental results demonstrate the versatility and improvability of GAN networks in both fields of augmentation and segmentation. Overall, deep learning in medical imaging remains an extremely intoxicating field full of meticulous and innovative new studies.

KEYWORDS

Medical Imaging Synthesis; Brain Tumour Segmentation; Glioblastoma; Convolutional Neural Networks; Generative Adversarial Networks; 3D MRI.

1. INTRODUCTION

Often referred to as the deadliest human cancer, Glioblastoma (GBM) is the most common brain cancer and an extremely dangerous one. The brain, being the humans' most supreme but mystifying organ, makes the eradication and remedy of this cancer all the more critical. GBM is typically caused by the canceration of glial cells within the brain and the spinal cord. World Health Organisation (WHO) has graded it as a grade IV cancer, now accounting for over 80% of all primary malignant tumours [1]. This grading is done based on the presence/absence of marked mitotic activity, necrosis and florid microvascular proliferation, between grades I, II, III and IV [2]. Gliomas with grade II and I are referred to as low grade gliomas (LGG), and those above are called high grade gliomas (HGG).

Generally, the standard glioma therapy comprises of Stupp protocol with maximal surgical resection followed by chemo radiation (temozolomide and radiation therapy) [3]. Despite many clinical trials and the combined effort of countless medical communities, the median survival age of GBM is still only around 18 to 24 months [4]. Some of the most common obstacles include the aggressive nature of a high-grade neoplasm, perivascular niches of 'persister cancer cells', the inherent genetic variability of glioma, as well as the lack of efficiency for chemotherapeutic agents to cross the blood-brain barrier [5]. As a result, treatment targeted to a particular mutation

may only be partially effective. Currently, before we envision any ‘precision medicine’ approaches to combat glioma, histopathologic examinations must first be able to fully examine and capture the tumour’s genetic variability and mutability [6].

In recent years, the upswing of machine and deep learning in radiology offers an unprecedented opportunity. Now, the millions of voxels of an MR image can become useful in the segmentation/classification of glioma features - paramount in treatment management and prognostication. While Genomic sequencing is still not widely available, a cheaper and much more open technology like the MRI-coupled AI algorithms can significantly benefit the neurobiology and neurological-oncology practices.

Within AI and machine learning, deep learning is an increasingly growing subset training neural networks resembling human brains. In the case of radiology analysis, feature selection and classification can occur concurrently in one algorithm without any human intervention. ImageNet - the world’s biggest image database - hosts annual competitions on the use of AI for classifying animals and objects; in particular, convolutional neural networks (CNNs) have consistently outperformed all traditional machine learning techniques and other deep learning competitors [7~9]. This explains the CNN’s explosive popularity and marks the drift from traditional machine learning techniques like regressions, support vector machines (SVM), random forests and dimensionality reductions for digital imaging analysis. As an example, CNNs can use the quantitative information from millions of voxels from a glioma brain MRI - typically with the modalities of T1WI (T1 weighted image), T1CE (T1 contrast enhanced), T2WI (T2 weighted image) and T2-FLAIR (T2 fluid attenuated inversion recovery) - to solve the complex mathematical analysis tasks. Given enough information - high quantities of high-quality data, CNNs will be able to determine the radiologic features and their relative importance, before making a predictive model to classify images with high accuracy. This is where data becomes exceptionally valuable.

Traditionally, tumour segmentation, being a necessary step before classification or prediction, is performed by mapping (also finding this mapping function) real medical images (e.g. MRI or CT) to semantic label images of the tumour described by radionics and medical professionals. Because of a rapid development in medical imaging equipment, multi-modal data pairs have come to be preferred as they capture more information about the tumour and can lead to better performance results [10]. This is mentioned earlier with T1WI, T2-FLAIR, etc being the 4 primary MRI modalities. Unlike classification, nevertheless, labelling medical images by hand is not only time-consuming but expensive due to the requirement of medical specialists with years of experience. As a result, generative adversarial networks (GANs) have come to be used for image synthesis and data augmentation. In particular, image-to-image translation models like CycleGAN [11] and Pix2Pix [12] have become distinguished very recently due to their accuracy and potential for pixel-wise image augmentation. Now, the use of 3D Multi-modal data has increased the complexity of the model but also the amount of information available - hence improving the models’ performances.

Overall, this paper aims to give a general encapsulation over the field of deep learning on glioma medical imaging; not just the histopathologic aspect on multi-modal MR images, but also the mathematical architectures of various cutting-edge convolutional neural networks and generative adversarial networks. Furthermore, this paper implements an adjusted and calibrated version of a most recent and developed network Vox2Vox [13] by Cirillo et al. built utilising many of the state-of-the-art execution techniques on the task of Glioblastoma image synthesis and segmentation.

2. RELATED WORKS

2.1. Deep Learning in Glioma Medical Imaging

Many recent research studies have achieved remarkable success in the use of deep learning for numerous glioma medical imaging tasks. These include tumour grading, IDH Mutations, 1P19Q Codeletion, MGMT Promoter, Pseudo-progression and overall survival rate predictions, etc [6].

First, tumour grading is a classification task crucial for patient management. The WHO's have categorised different gliomas into the categories of grade I, grade II, III or IV. The use of CNNs, in particular, have been favoured for their great potential in both histology and imaging. Many algorithms, including one using Weighted Neighbour Distance using Compound Hierarchy of Algorithms Representing Morphology (WNCCHRM) and VGG-19 Deep CNN [14], AlexNet and GoogLeNet [15] as well as an ensemble of CNNs [16] have been able to distinguish between high- and low-grade tumours to produce stunning classification results. Isocitrate dehydrogenase (IDH) mutations, on the other hand, are another key area of study as they represent improved chances of survival. Being a prediction task, algorithms including a deep learning-based radionics (DLR) approach [17], a residual CNN [18], 2D/3D hybrid CNNs [19], and a 3D-Dense-UNet CNN [20] have been trained using various datasets from the Tumour Cancer Imaging Archive (TCIA) to The Cancer Genome Atlas (TCGA). Similar to IDH mutations, 1P19Q Codeletion also signifies a lower-grade tumour with prediction techniques ranging from a 2D/3D CNN [19], a novel multi stream deep CNN (7-layer 2D CNN) [21], and other multi scale CNNs [22], etc.

Next, the hypermethylation of O⁶-methylguanine-DNA-methyltransferase (MGMT) promoter is an enzyme indicating the need of different prognosis procedures. Being a more delicate task, success from CNNs have been rather modest including the uses of a bi-directional recurrent CNN [23], 3 different residual CNNs [24], and the same novel 2D-3D hybrid CNN [19] as before. It is perhaps notable that the use of principal component analysis (PCA) for dimensionality reduction have identified key imaging features for predictions of MGMT; including heterogeneous and nodular enhancement, more T2/FLAIR signal with cortical involvement, the presence of eccentric cysts, as well as a slight tendency toward frontal/temporal lobe locations [19]. Likewise, another task is the differentiation of Pseudoprogression (PSP) from true progression (TP) as disease management is radically different in both cases. Some comparatively successful results have been found with a hybrid deep and machine learning CNN-LSTM (long short-term memory) approach, with LSTM algorithms being prominently used in sequence predictions like automatic image caption generation, automatic translation of text, and automatic handwriting generation [25]. Here, the lack of 'ground truth' (histopathologically proven and labelled/annotated MR images) made PSP remain as an unmet and critical area of neuro-oncology study. In fact, the author of this paper originally planned to use pseudo-progression as the target for GAN synthesis, but unfortunately had to forgo the task due to the lack of publicly available datasets. At last, the overall survival rate prediction is another ongoing area of development. Many algorithms, including the combination of a 3D CNN for segmentation and a decision tree regressor as predictor [26], and a 3D CNN to extract features combined with a support vector machine to predict long vs short-term OS [27], have all achieved relatively appreciable successes. Notably, some risk factors indicating poor survival rates after diagnosis with GBM include Caucasians ethnicity, the male sex, 60+ years of age, Karnofsky scores of below 70, as well as advanced tumour and surgery without adjuvant chemoradiation, etc.

Overall, despite hype and certain impressive algorithms, the use of deep learning in neuroimaging still faces significant challenges. The lack of datasets remains the biggest obstacle. Compared to the traditional machine learning approaches, CNNs require much larger and well-annotated

images to improve performance, which is very costly and time-consuming. Radiologists capable of performing such tasks are often busy and difficult to finance, and gliomas, being relatively rare, have made the acquisition of real test cases a difficult problem. Even good CNNs may underperform and overfit due to the independent, external and unregulated datasets. Here, the issue of overfitting occurs as with increased learnable parameters, the model requires more training samples as well. As a result, the need for medical synthesis becomes apparent and a most indispensable task in brain tumour medical imaging [28].

2.2. The Implementation of Generative Adversarial Networks

As a result of insufficient medical imaging datasets, many image augmentation methods have been developed to superficially expand the number of training/validation instances. Earlier methods consist of simple geometric deformations, such as random scaling, rotation and flipping. Nevertheless, for obvious reasons, basic augmentation methods fail to generate and represent the variations in the shape, location, and pathology of brain tumours [29].

Generative adversarial networks (GANs), on the other hand, have provided an exceptional solution for image augmentation. Over the entire field of computer vision, GANs have gained tremendous popularity due to the ability to synthesise virtually any type of image. The standard protocols include style transfer [30], image synthesis from noise [31], image to image translation [32], as well as image segmentation [33]. Compared to other computer vision datasets like the aforementioned ImageNet, GANs have become especially popular in medical imaging due to the discipline's significantly smaller datasets. Furthermore, since the collection of multiple image modalities significantly improves model performances, many GANs - like the Cycle GAN [11] - have been introduced to synthesise the missing modalities as well.

It is in the field of medical imaging segmentation that generative adversarial networks have gained remarkable traction. Han et al. devised a GAN to segment MRI spinal structures [34]; Dong et al. proposed a U-Net style GAN for organs-at-risk (OARs) segmentation [35]; while Skandarani et al. developed a variational auto encoder (VAE) + GAN model for cardiac MRI generation [36]. Specifically on the topic of Glioma and brain tumours, moreover, Nema et al. designed a 2D GAN named RescueNet to segment brain tumours [37]; Armanious et al. created an end-to-end image-to-image translation system utilising the CasNet generator, named MedGAN [38]; and Qingyun et al. proposed a novel generator with two discriminators augmentation network known as TumourGAN [28].

Additionally, in their 2016 paper [12], Isola et al. introduced the Pix2Pix GAN which was particularly impactful in that it was able to generate a type A image from a paired type B. It is an implementation of conditional GANs (cGAN) where the image's generation is conditional on a given image. For Glioblastoma/brain tumour segmentation, both CycleGAN and Pix2Pix GAN are the state-of-the-art industry standards. Since CycleGAN's expertise is more on image-to-image translation, the Pix2Pix model is preferred here due to its closer relevance to the synthesis task. A 3D Implementation of Pix2Pix, Vox2Vox [13], developed by Cirillo et al. in 2020 is utilised here as the principal network of this study. It is chosen as one of the most cutting-edge and systematically robust networks for the concerning task. While a normal convolutional neural network, even the state-of-the-art ones like U-Net [39], segments by examining the image voxel by voxel and maximising a segmentation metric (for example, weighted dice coefficient and intersection over union (IOU)); GANs have the added level of scrutiny by also punishing results that do not look realistic.

This way, it is hypothesised that GANs can be both useful in data augmentation as well as segmentation. To test this, a version of the Vox2Vox adversarial network is trained on the task of segmentation label synthesis.

3. MATERIALS AND METHODS

3.1 will introduce the dataset being used in the paper. 3.2 then describes the mathematical fundamentals behind convolutional neural network and generative adversarial networks.

3.1. Data

For this project the MR images are obtained from the Multimodal Brain Tumour Segmentation Challenge (BraTS) 2020 [40]; being the biggest public dataset on brain tumours. The BraTS training dataset consists of MR volumes of format $240 * 240 * 155$ (155 films each of pixel size $240*240$) from 369 patients. For each patient, 4 types of MR images are provided: T1WI, T1CE, T2, and T2 FLAIR - being the 4 aforementioned MRI modalities. The BraTS validation dataset contains equivalent MR volumes from a further 125 patients.

Here, the images were acquired from 19 different institutions each with their own clinical protocols. The group truth labels were all segmented manually, by 1~4 broad-certified neuro-radiologists following the same annotation protocol. The annotations were then approved by more experienced radiologists for authenticity. No ground truths are provided for the validation set. To ensure consistency, all data were co-registered to the same anatomical template, skull-stripped and interpolated to the same resolution (1 mm^3). The below Figure 1 shows one example of training T1 MR images overlapped with the true segmentation in different planes.

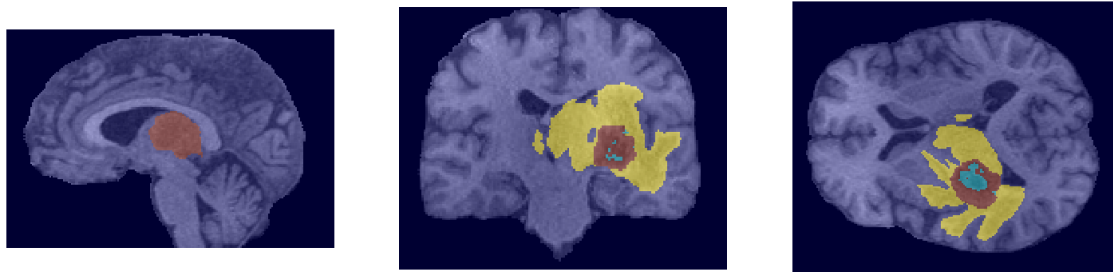


Figure 1. The T1 MR Images in sagittal, coronal and transverse plane overlapped with the true segmentation. Highlighted in yellow, red and cyan are the peritumoural edema (ED), necrotic and non-enhancing core (NCR/NET), and GD-enhancing tumour (ET) [13].



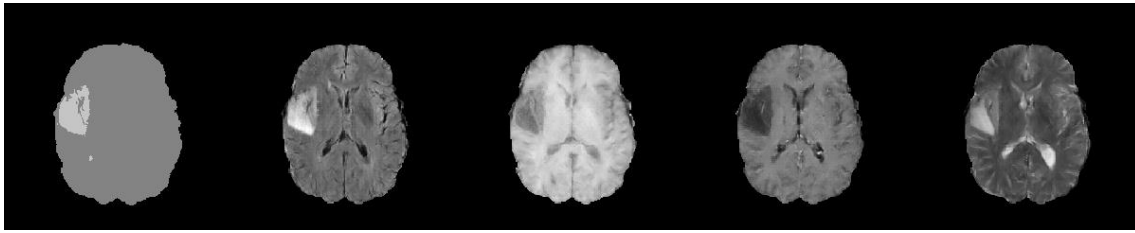


Figure 2. Synthetic image examples from one patient, representing the 58th and 73rd slices. Images from left to right correspond to the semantic labels and the 4 modality images (flair, t1, t1ce and t2), respectively [28].

3.2. Network Architecture

3.2.1. Convolutional Neural Networks

In recent years, CNNs have had significant success being applied to visual recognition problems. In their 1998 paper [41], Yann LeCun et al. developed the famous LeNet-5 architecture to recognise handwritten check numbers. This was widely regarded as the first ever CNN [42].

In detail, CNNs contain convolutional layers where neurons are not connected to all pixels in the input image (the standard practice for MLPs), instead only those in their receptive fields. A second convolutional layer will also only be connected to neurons located within the receptive field of the previous layer. A neuron located in (i, j) is connected to the outputs of the neuron in the previous layer located at $(i + fh - 1, j + fw - 1)$ where fh and fw are the height and width of the receptive field. Filters, then, being sets of weights, can be employed by neuron-full layers to produce feature maps - highlighting the areas in an image which activates the filter the most. By stacking feature maps - where all neurons within a feature map share the same parameters (weights and bias) - the number of parameters is dramatically reduced. Another notable feature of CNNs is the pooling layer responsible for subsampling input images to reduce the computational load, memory usage and the number of parameters (hence reducing the risks of overfitting). Like convolutional layers, pooling layers are connected to a receptive field of neurons from the previous layer, the difference being pooling layers do not have weights, only aggregating the input with aggregation functions like maximum or mean pooling. Typically, CNNs consist of a few convolutional layers (generally followed by RELUs), a pooling layer, a few more convolutional layers, another pooling, and so on. At the end, typically a regular feedforward net is added with a few fully connected layers. The final output layer tends to be a softmax layer outputting class probability predictions.

Some of the most prominent and fundamental CNNs include the LeNet-5 (1998) [41], AlexNet (2012) [43], GoogLeNet (2014) [44], VVGNet (2014) [45], ResNet (2015) [46], Xception (2016) [47], SENet (2017) [48], among others. Here we briefly explicate on the architecture of AlexNet. As winner of the 2012 ImageNet ILSVRC Challenge by a tremendous margin, AlexNet was revolutionary at the time being the first network to stack convolutional layers directly on top of each other. Table 1. shows the network's architecture.

Table 1. Illustration of AlexNet's network architecture. It is similar to LeNet5 but larger and deeper.

Layer	Type	Maps	Size	Kernel Size	Stride	Padding	Activation
In	Input	3 (RGB)	227*227	-	-	-	-
C1	Convolution	96	55*55	11*11	4	VALID	ReLU
S2	Max Pooling	96	27*27	3*3	2	VALID	-
C3	Convolution	256	27*27	5*5	1	SAME	ReLU
S4	Max Pooling	256	13*13	3*3	2	VALID	-
C5	Convolution	384	13*13	3*3	1	SAME	ReLU
C6	Convolution	384	13*13	3*3	1	SAME	ReLU
C7	Convolution	256	13*13	3*3	1	SAME	ReLU
F8	Fully connected	-	4096	-	-	-	ReLU
F9	Fully connected	-	4096	-	-	-	ReLU
Out	Fully connected	-	1000	-	-	-	Softmax

3.2.2. Generative Adversarial Networks

GANs, on the other hand, constitute deep learning frameworks trained in an adversarial fashion. Developed by Ian Goodfellow in 2014 [49], GANs took a few years to overcome initial training difficulties and become mainstream. The underlying principle behind GAN is the use of competition to push neural networks to excel. It is comprised of 2 parts: a generator which tries to map a random latent space to a given data distribution, and a discriminator distinguishing the real distribution from the fake ones. In image recognition, the generator imitates the original distribution to synthesise images, while the discriminator rejects fake images and learns to differentiate between the two. This is done iteratively, separately, and alternatively following the minimax game-playing algorithm.

First, from practically random noise, the generator produces fake samples which can initially fool the discriminator. Then, the discriminator is given both genuine and fake images and improves at differentiating the two. The most basic structure is as followed. During training, the discriminator is trained first with batches of real image and fake images produced by the generator. Here, back propagation only optimises the weights of the discriminator during this phase. Then, the generator is trained with the updated gradients, and real images are never fed into the network. Here weights of the discriminator are frozen and back propagation only affects weights of the generator. The two parts of the network are updated simultaneously, until Nash Equilibrium - the state at which no player is better off changing their strategy, assuming the other does not change theirs - is reached. The following optimisation equation demonstrates the minimax game rules:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

Here $D(x)$ represents the probability that x belongs to the class of real images. The discriminator wants to calculate the probability near 1 when the input data is real. Otherwise, to judge the data and distinguish them when input samples are fake. As $D(G(z))$ draws close to 0, the generator aims to approach 1. Essentially, both networks try to outsmart each other with the goal of reaching the Nash Equilibrium where the probabilities of real vs fake are exactly (50, 50). Idealistically, if trained long enough, this state can be reached and the generator is almost perfect.

Nevertheless, the issue of mode collapse - where the generator's output becomes less diverse, focusing on the production of one convincing image class and forgetting all others - might occur where the generator only cycles across a few classes, never becoming too good at producing any. Some of the current solutions include experience replay, mini-batch discrimination, as well as many newer procedures combatting the difficulties of training GANs. Yann LeCun, the creator of CNN, described generative adversarial training as "the most interesting idea in the last 10 years in Machine Learning", exemplifying how the field of GAN is a most compelling and active one.

For the purpose of this study, the Vox2Vox model architecture will be examined. It is a novel approach building the generator with a U-Net [39] and Res-Net [46] architecture, while the discriminator is built with PatchGAN [33] - the same original style outlined in Pix2Pix.

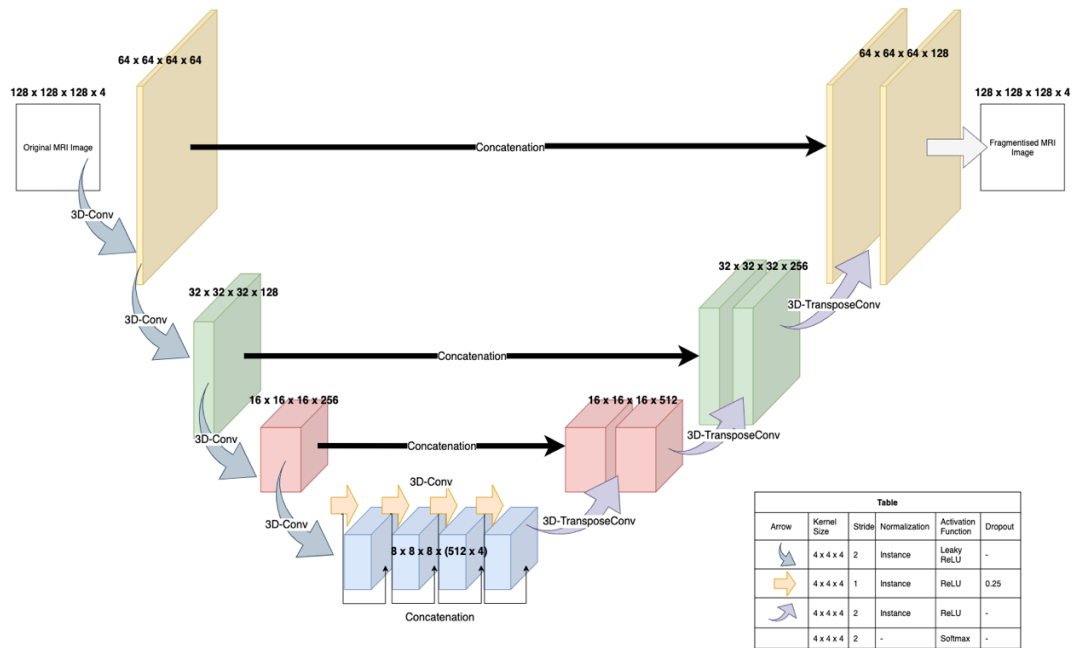


Fig. 3. The Vox2Vox generator model.

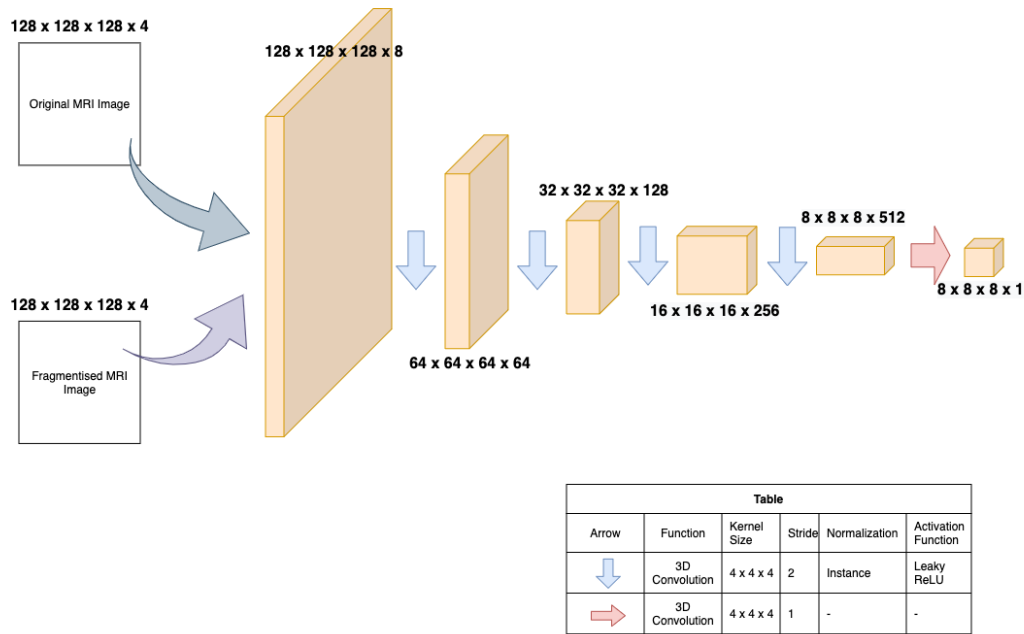


Fig. 4. The Vox2Vox Discriminator model.

4. EXPERIMENT

To demonstrate the use of GAN in synthesis and segmentation, an adjusted and calibrated version of the Vox2Vox network is applied and trained on the BraTS 2020 dataset.

4.1 will detail the data preprocessing/augmentation pipeline; 4.2 will describe more precise implementations; while 4.3 entails the results obtained after training.

4.1. Data Pre-Processing and Augmentation

The pre-processing pipeline stays true to the Vox2Vox proposal. First, the background voxels are fixed to 0 for noise reduction. Intensity normalisation is also executed with NumPy maths modules. The ground truths are then categorised into background, peritumoural edema (ED), necrotic and non-enhancing tumour core (CNR/NET), GD-enhancing tumour (ET) with the labels 0, 1, 2, 4 respectively.

Since training with large images can be over-demanding for such complex models, a patch augmentation technique is also applied to extract a 128*128*128 patch from the original 240*240*155 arrays. This way, only 23.5% of the entire training set is used in the training epoch.

To prevent the networks from overfitting - a prevalent issue for all feature extraction models, where the network memorises the exact training details with high bias and low variance - random data augmentation is also applied. For each sample, the model performs one, or a combination, of the 4 possible augmentation techniques: including *3D flipping* in respect to one of the axes

(chosen randomly), *3D rotation* with random angles, *brightness adjustment* using the power-law gamma transformation, as well as *elastic deformation* with a square deformation grid.

A separate pipeline from a 3D-UNet network [50] is also considered and compared with the Vox2Vox implementation. Although the 3D-UNet includes some additional functions like to randomly scale the image or generate random permutation keys, the Vox2Vox code is far more succinct and covers more advanced and relevant augmentation techniques.

4.2. Implementation

4.2.1. Losses

For the Vox2Vox model, two loss functions are implemented for the generator and discriminator networks. The Discriminator loss is constructed with the sum of the L2 error between the original image and ground truths, and between the original image and the respective segmentation predictions. The generator loss, moreover, is the sum of the L2 error between the original image and segmentation prediction and the generalised dice loss between the ground-truth and generator's output.

Below are the formulae for the discriminator and generator losses:

$$L_D = L_2 [D(x, y), \mathbf{1}] + L_2 [D(x, \hat{y}), \mathbf{0}] ,$$

$$L_G = L_2 [D(x, \hat{y}), \mathbf{1}] + \alpha GDL (y, \hat{y}) .$$

In addition, the traditional Generalised Dice Loss (GDL) [51] used in the generator loss and for performance evaluation, is replaced with the Tversky loss function [52]. This loss function was specifically designed to optimise segmentation on imbalanced medical datasets, being weighted by the constants 'alpha' and 'beta' which penalise false positive and false negatives to a higher degree in the loss function.

$$GDL = 1 - 2 \frac{\sum_{l=1}^2 w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 w_l \sum_n r_{ln} + p_{ln}} ,$$

$$T(\alpha, \beta) = \frac{\sum_{i=1}^N p_{0i} g_{0i}}{\sum_{i=1}^N p_{0i} g_{0i} + \alpha \sum_{i=1}^N p_{0i} g_{1i} + \beta \sum_{i=1}^N p_{1i} g_{0i}}$$

4.2.2. Optimisation and Regularisation

This portion mostly stays true to the Vox2Vox implementation, where both the generator and discriminator are trained using the Adam optimiser [53], with the parameters: $\lambda = 2 \cdot 10^{-4}$, $\beta_1 = 0.5$, and $\beta_2 = 0.999$. In addition, dropout regularisation is a predominant regularisation technique where at every training step, each neuron has a probability of being dropped out (entirely ignored during the step), forcing them to not become overly co-dependent. Here instead of the 0.2 probability in Vox2Vox the dropout rate is increased to 0.25 to further reduce overfitting.

4.3. Results

The calibrated-Vox2Vox model is implemented with Python 3.7, Tensorflow 2.6 and Keras 2.6. As mentioned earlier, it is trained and validated on 128*128*128 images from 369 and 125 subjects using batch size 4. Due to the lack of access to accelerated training hardware, the model is trained on Google Colab Pro. Since the runtime automatically disconnects after 12 hours, only 9 epochs were able to be trained on a randomly allocated TPU with high-RAM.

After each epoch of training, the loss, Tversky loss, discriminator loss on both the training and validation data are displayed for simple comparisons. The training loss decreased to the lowest of 1.55, Tversky loss reached around 0.234 on the 9th epoch, and the validation loss plateaued around 1.74. The validation loss reached 1.74 on the 5th epoch and stopped decreasing for the further 4 epochs of training. To compare, another 6 epochs were trained obtaining very similar results. As illustrated on Figure 3, the label generation is visibly improving after just a few epochs of training.

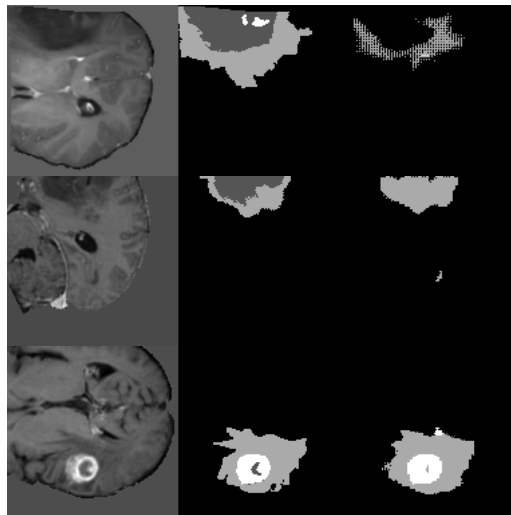


Figure 5. The generator's prediction of ground truth given the image data, on epochs 2, 3 and 8 respectively. Left side is the actual brain image, the middle being ground truth, and right being the predicted segmentation.

In the original Vox2Vox paper, the author undertook over 200 epochs of training on a computer equipped with 128GB RAM and Nvidia GeForce RTX 2080 Ti graphics card, with 11GB of memory. There is also post-processing with 10-fold cross-validation on the training set with an ensemble of 10 Vox2Vox models, achieving the state-of-the-art results of 93.4% on the WT region, 92.6% on the TC region and 87.2% on the ET region (all median dice scores on the training set). The median was proposed since certain bad predictions tended to compromise the mean values, as a generally more reliable deciding metric.

5. CONCLUSIONS

Overall, this paper aims to examine both the general history of deep learning on glioma medical imaging, as well as focusing on the adjustment and implementations of one specific cutting-edge model. In recent years, AI algorithms are becoming increasingly popular dealing with clinical neuroimaging tasks, only restrained by challenges like the lack of large, well-annotated datasets and expert-reviewed 'ground truths'. As a result, generative adversarial networks have been used to solve this issue synthesizing both authentic-looking training data as well as ground truths.

Many aforementioned models have combined and utilised many complex mechanisms such as image-to-image translation, hybrid systems, ensemble learning as well as taking advantage of 3D multi-modal MRI data with 3D convolutional pipelines.

Moreover, the scrutinization and development of the new Vox2Vox training model also yielded unexpectedly positive results. After such few epochs of training, the predictions were already improving visibly and if continued will eventually yield more state-of-the-art results. The complexity of generating Glioblastoma segmentation ground-truths exceeds most of the other image synthesis tasks, consequently demonstrating the immense powers of GANs in both fields of augmentation and segmentation.

6. DISCUSSIONS

Due to the conceptual complexity of the Vox2Vox model and in dealing with multi-modal MR images, and more importantly, the time constraint and the author's limited access to training hardware, the experiment and result sections are unsatisfactory with tremendous areas of originality improvements.

In future studies, the Vox2Vox model can be ameliorated in many ways such as the implementation of more advanced metrics, optimization of the PatchGAN architecture, and optimizations of the training time, etc. The model can also be altered in that instead of synthesising segmentation images, it also synthesises MR images to facilitate another CNN (or SVM and any other deep or machine learning algorithms) on segmentation. The results of Vox2Vox's segmentation synthesiser and the CNN's segmentor can then be compared. In this case, the other 3D U-NET model could have also been trained, and the two segmentation accuracies contrasted. Furthermore, with other image-to-image translation models such as TumourGAN achieving substantial success over even state-of-the-art algorithms like CycleGAN and Pix2Pix, the Vox2Vox model and others similar can be altered to combine the advanced 3D feature extraction techniques with translation networks to achieve even more worthwhile breakthroughs.

Being an active field, recent developments in the use of GAN or CNN/other models on glioma medical imaging, or any medical imaging tasks, is appreciable to say the least. The author is optimistic of a future filled with more innovatory, progressive and responsible establishments in the applications of Artificial Intelligence.

REFERENCES

- [1] Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. *Neuro Oncol.* 2015;17(suppl 4):iv1–iv62.
- [2] F. E. Bleeker, R. J. Molenaar, and L. Sieger, "Recent advances in the molecular understanding of glioblastoma," *Journal of Neuro-Oncology*, vol. 108, pp. 11–27, 2012.
- [3] Stupp R, Mason WP, van der Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med.* 2005;352:987–996.
- [4] Chow DS, Chang P, Weinberg B, et al. Imaging genetic heterogeneity in glioblastoma. *AJR Am J Roentgen.* 2018;210:30–38.
- [5] Sattiraju A, Mintz A. Pericytes in glioblastoma: multifaceted role within tumor microenvironments and potential for therapeutic interventions. *Adv Exp Med Biol.* 2019;1147:65–91.
- [6] Zlochower A, Chow DS, Chang P, Khatri D, Boockvar JA, Filippi CG. Deep Learning AI Applications in the Imaging of Glioma. *Top Magn Reson Imaging.* 2020 Apr;29(2):115–0. doi: 10.1097/RMR.000000000000237. PMID: 32271288.
- [7] Le Cun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–444.

- [8] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
- [9] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. arXiv/1512.03385.
- [10] Tseng, K.L.; Lin, Y.L.; Hsu, W.; Huang, C.Y. Joint sequence learning and cross-modality convolution for 3d biomedical segmentation.
- [11] Zhu J.Y; Park T, Isola P, Efros A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv:1703.10593.
- [12] Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. arXiv:1611.07004
- [13] Cirillo, M.D.; Abramian D.; Eklund A. Vox2Vox: 3D-GAN for Brain Tumour Segmentation. ArXiv:2003.13653 [Cs, Eess], Nov. 2020. arXiv.org, <http://arxiv.org/abs/2003.13653>.
- [14] Ahammed Muneer KV, Rajendran VR, Paul JK. Glioma tumor grade identification using artificial intelligence techniques. J Med Syst. 2019;43:113.
- [15] Yang Y, Yan LF, Zhang X, et al. Glioma grading on conventional MR images: a deep learning study with transfer learning. Front Neurosci. 2018;12:804.
- [16] Ertosun MG, Rubin DL. Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. AMIA Ann Symp Proc. 2015;2015:1899–1908.
- [17] Li Z, Wang Y, Yu J, et al. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep. 2017;7:5467.
- [18] Chang K, Bai HX, Zhou H, et al. Residual convolutional neural networks for determination of IDH status in low- and high grade gliomas from MR imaging. Clin Cancer Res. 2018;24:1073–1081.
- [19] Chang P, Grinband J, Weinberg BD, et al. Deep learning convolutional neural networks accurately classify genetic mutations in glioma. AJNR Am J Neuroradiol. 2018;39:1201–1207.
- [20] Bangalore Yogananda CG, Shah BR, Vejdani-Jahromi M, et al. A novel fully automated MRI-based deep learning method for classification of IDH mutation status in brain gliomas. Neuro Oncol. 2020;22:402–411.
- [21] Ge C, Gu IY, Jakola AS, Yang J. Deep Learning and Multi-Sensor Fusion for Glioma Classification Using Multistream 2D Convolutional Networks. Abstract in Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 58945897.
- [22] Akkus Z, Ali I, Sedlar J, et al. Predicting deletion of chromosomal arms of 1p/ 19q in low-grade glioma from MR images using machine intelligence. J Digit Imaging. 2017;30:469–476.
- [23] Han L, Kamdar MR. MRI to MGMT: predicting methylation status in glioblastoma using convolutional recurrent neural networks. Pac Symp Biocomput. 2018;23:331–342.
- [24] Korfiatis P, Kline TL, Lachance DH, et al. Residual deep convolutional neural network predicts MGMT methylation status. J Digit Imaging. 2017;30:622–628.
- [25] Jang BS, Jeon SH, Kim IH, et al. Predictor of pseudoprogression versus progression using machine learning algorithm in glioblastoma. Sci Rep. 2018;8:12516.
- [26] Li S, Zhang S, Chen H, et al. Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning. Front Neurosci. 2019;13:1–8.
- [27] Nie D, Lu J, Zhang H, et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. Sci Rep. 2019;9:1103.
- [28] Li, Q., Yu, Z., Wang, Y., & Zheng, H. (2020). TumorGAN: A Multi-Modal Data Augmentation Framework for Brain Tumor Segmentation. Sensors (Basel, Switzerland), 20(15), 4203. <https://doi.org/10.3390/s20154203>
- [29] Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best practices for convolutional neural networks applied to visual document analysis. Icdar 2003, 2, p. 958.
- [30] Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
- [31] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. ICLR (2018)
- [32] Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)

- [33] Sato, M., Hotta, K., Imanishi, A., Matsuda, M., Terai, K.: Segmentation of Cell Membrane and Nucleus by Improving Pix2pix. In: BIOSIGNALS. pp. 216–220 (2018)
- [34] Han, Z., Wei, B., Mercado, A., Leung, S., Li, S.: Spine-GAN: Semantic segmentation of multiple spinal structures. *Medical image analysis* 50, 23–35 (2018)
- [35] Dong, X., Lei, Y., Wang, T., Thomas, M., Tang, L., Curran, W.J., Liu, T., Yang, X.: Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Medical physics* 46(5), 2157–2168 (2019)
- [36] Skandarani Y.; Painchaud N.; Jodoin P.M.; Lalonde A.; On the effectiveness of GAN generated cardiac MRIs for segmentation. arXiv:2005.09026
- [37] Nema, S., Dudhane, A., Murala, S., Naidu, S.: RescueNet: An unpaired GAN for brain tumor segmentation. *Biomedical Signal Processing and Control* 55, 101641 (2020)
- [38] Armanious K.; Jiang C.; Fischer M.; Küstner T.; Nikolaou K.; Gatidis S.; Yang B: MedGAN: Medical Image Translation using GANs: arXiv:1806.06397
- [39] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
- [40] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M. A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., Corso, J. J., ... Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10), 1993–2024. <https://doi.org/10.1109/TMI.2014.2377694>
- [41] Lecun, Y., Jackel, L. D., Bottou, L., Brunot, A., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Muller, U. A., Sackinger, E., Simard, P., & Vapnik, V. (1995). Comparison of learning algorithms for handwritten digit recognition. In F. Fogelman, & P. Gallinari (Eds.), *International Conference on Artificial Neural Networks*, Paris (pp. 53-60). EC2 & Cie.
- [42] Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly.
- [43] Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 84 - 90.
- [44] Szegedy C., Liu W., Jia Y. Servant P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A.: Going Deeper with Convolutions: arXiv:1409.4842
- [45] Simonyan K., Zisserman A.; Very Deep Convolutional Networks for Large-Scale Image Recognition; arXiv:1409.1556
- [46] He K., Zhang X., Ran S., Sun J.; Deep Residual Learning for Image Recognition; arXiv:1512.03385
- [47] Chollet F.; Xception: Deep Learning with Depthwise Separable Convolutions; arXiv:1610.02357
- [48] Hu J., Shen L., Albanie S., Sun G., Wu E.; Squeeze-and-Excitation Networks; arXiv:1709.01507
- [49] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y.; Generative Adversarial Networks; arXiv:1406.2661
- [50] Wang F., Jiang R., Zheng L., Meng C., Biswal B.; 3D U-Net Based Brain Tumour Segmentation and Survival Days Prediction; arXiv:1909.12901
- [51] Sudre C.H., Li W., Vercauteren T., Ourselin S., Cardoso M. J.; Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations; arXiv:1707.03237
- [52] Selehi S.S.M., Erdogmus D., Gholipour A.; Tversky loss function for image segmentation using 3D fully convolutional deep networks; arXiv:1706.05721
- [53] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

AUTHORS

Edmond Wang is a year 12 (year 13 in the UK) student currently studying at Westminster School, London. Originally from Beijing, Hong Kong, he came to the UK during year 8 (UK year 9). He is currently studying Mathematics, Further Mathematics, Economics and Philosophy for A-Levels. Academically, his passion lies in Maths/Computer Science and AI and aspires to become an AI-related entrepreneur in the future. Outside of school, he also enjoys basketball, piano, reading, and sneaker-collecting, etc.

