Brain MRI Tumor Segmentation with Adversarial Networks

Edoardo Giacomello
Dipartimento di Elettronica, Informazione e Bioinformatica
Politecnico di Milano
edoardo.giacomello@polimi.it

Daniele Loiacono
Dipartimento di Elettronica, Informazione e Bioinformatica
Politecnico di Milano
daniele.loiacono@polimi.it

Luca Mainardi
Dipartimento di Elettronica, Informazione e Bioinformatica
Politecnico di Milano
luca.mainardi@polimi.it

***Abstract*—Deep Learning is a promising approach to either automate or simplify several tasks in the healthcare domain. In this work, we introduce *SegAN-CAT*, an end-to-end approach to brain tumor segmentation in Magnetic Resonance Images (MRI), based on Adversarial Networks. In particular, we extend *SegAN*, successfully applied to the same task in a previous work, in two respects: (i) we used a different model input and (ii) we employed a modified loss function to train the model. We tested our approach on two large datasets, made available by the *Brain Tumor Image Segmentation Benchmark* (BraTS). First, we trained and tested some segmentation models assuming the availability of all the major MRI contrast modalities, i.e., T1-weighted, T1 weighted contrast enhanced, T2-weighted, and T2-FLAIR. However, as these four modalities are not always all available for each patient, we also trained and tested four segmentation models that take as input MRIs acquired with a single contrast modality. Finally, we proposed to apply transfer learning across different contrast modalities to improve the performance of these single-modality models. Our results are promising and show that not only *SegAN-CAT* is able to outperform *SegAN* when all the four modalities are available, but also that transfer learning can actually lead to better performances when only a single modality is available.**

I. Introduction

In the last decade, machine learning has been applied to a large number of tasks in the healthcare domain and proved to be a promising approach to either automate or simplify clinical processes that would be otherwise performed manually, requiring a great amount of time and increasing the possibility of human error. These applications usually rely on identifying and extracting a set of effective features from data, a time consuming and problem dependent process. In contrast, deep learning promises to solve this issue due to the capability of working directly with raw data by learning automatically an efficient data representation to solve the problems they are applied to [1]. In addition, the data representation learned is often suitable also for different problems, in some cases even across different application domains. These advantages come with a cost: deep learning algorithms generally require a huge amount of data to learn effective data representations and, hence, competent models to solve the problems they are applied to. For this reason, several works in the literature focus on data augmentation – i.e., how to exploit as much as possible the data available – and transfer learning techniques – i.e., how to train models with data from similar problems or domains.

In this work, we focus on the MRI brain tumor segmentation task, a problem that has been widely investigated in the past few years and is the object of an international scientific competition, the *Brain Tumor Image Segmentation Benchmark* (BraTS). In particular, we introduce *SegAN-CAT*, an end-to-end approach to MRI brain tumor segmentation based on generative adversarial networks [2]. We designed our architecture after *SegAN* [3], a top performing approach in the BraTS challenge, extending it in two respects: (i) we re-designed the input of one of the networks of the adversarial architecture and (ii) we introduced an additional term in the loss function.

Our goal in this paper is threefold. First, we aim at comparing the performance of *SegAN* and *SegAN-CAT* on brain tumor segmentation using as input the four MRI modalities most commonly acquired: T1-weighted (T1), T1-weighted contrast-enhanced (T1c), T2-weighted (T2), and T2-FLAIR (FLAIR). Second, as in real-world scenarios these modalities might not all be always available for each patient, we aim at training and comparing the performance of four *SegAN-CAT* models that use as input only a single MRI modality, i.e., one among T1, T1c, T2, and FLAIR. Third, we investigate whether a transfer learning approach can be used to exploit the knowledge extracted from a model, previously trained on a *source* MRI modality, to train a model that works with a different *target* MRI modality.

We tested our approach by training several brain tumor segmentation models on the BraTS 2015 and BraTS 2019 datasets provided for the BraTS challenge. Our results show that *SegAN-CAT* is able to outperform *SegAN*, especially thanks to the modified loss function employed. As expected the results also show that using all the MRI modalities together allows to achieve a much better performance than the one achieved by the models that use a single MRI modality as input. In addition, we show that the performance of single-modality models might be slightly improved by employing a transfer learning approach, that is by transferring the knowledge extracted from other modalities.

The remainder of this paper is organized as follows. In Section II we provide an overview of some of the most relevant previous works. Then, in Section III we describe our approach

978-1-7281-1884-0/19/$31.00 ©2019 IEEE

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

and in Section IV we describe our experimental design. In Section V we report and discuss our experimental results. Finally, we draw some conclusions in Section VI.

II. Related Work

In this section we provide an overview of some of the previous works in the literature that applied adversarial networks to the image segmentation problem as well as the previous works that combined transfer learning with deep neural networks. In addition, we also provide a brief overview of major approaches introduced in the literature to deal with missing MRI modalities.

A. Adversarial Models for Segmentation

Inspired by the recent success of *Generative Adversarial Networks* [2] (GANs), Luc et al. [4] proposed a method in which a segmentation network is trained to perform pixel-wise classifications on images, while an adversarial network (called *discriminator* or *critic*) is trained to discriminate segmentations coming from the segmentation network and the ground truth. This approach has been tested on the *PASCAL VOC 2012* [5] and on the *Stanford Background* [6] datasets and the results show an improvement in segmentation performances when an adversarial loss is used. In the medical imaging domain, multiple works applied adversarial methods to segment MRI, CT, PET and other domain specific image formats. In particular, two previous works applied adversarial learning to brain MRI. Moeskops et al. [7] investigated the effectiveness of adversarial training and dilated convolution. Xue et. al. [3] proposed an Adversarial Network with a Multi-Scale loss, called *SegAN*, achieving better performances compared to the state-of-the-art methods for brain tumor segmentation [8], [9].

B. Transfer Learning

Transfer learning investigates how to exploit a *source model* trained on a *source domain* for a *source task* on either a different *target domain* or a different *target task*. This usually involves either a complete or a partial retraining of the *source model* to adapt it to the *target domain/task*. In particular, transferring deep neural networks across different image analysis tasks has been proved successful [10] and it seems a promising approach to deal with machine learning applications for which the amount of data is limited.

An example of the ability of CNNs to learn features that are useful for different tasks is given in [11], where the authors follow a *multi-task learning* approach to demonstrate that it is possible to develop a single model to perform brain, breast and cardiac segmentation jointly. When working with different MRI modalities, transfer learning has been used in [12] for accelerating MRI acquisition times by applying MRI reconstruction. In [13], the authors uses transfer learning between dataset from different institutions for the prostate gland segmentation task, comparing performances on different dataset sizes. A subclass of transfer learning, called *domain adaptation* studies the setting in which *source* and *target* data have the same feature space but different marginal probability functions, while the task is the same for both the domains. For example, in [14] domain adaptation is applied from CT to MRI for the lung cancer segmentation task. In [15] the authors considered the issue of the different protocols adopted to acquire MRIs in the clinical practice: they used a set of MRIs as Source Domain and their follow-ups as Target Domain, studying the minimal amount of Target Domain data that is necessary to achieve acceptable performance when fine-tuning a segmentation model.

However, transfer learning applied to Adversarial Networks is still an ongoing field of research. In [16], the authors focused on transfer learning using *WGAN-GP* [17] models and investigated how the selection of the network weights to transfer, the size of the target dataset, and the relation between source and target domain affect the final performances. Finally, in a recent work [18], Fregier and Gouray propose a new method to perform transfer learning with WGANs [19].

C. Missing MRI Modalities

A commonly used approach to address the problem of MRI segmentation in the case of missing modalities is to use *image synthesis* techniques to generate artificial data. In [20], Dar et al. used *Cycle GAN* [21] to generate missing modalities in a uni-modal setting, while Sharma and Hamarneh [22] designed a multi-modal architecture where the missing modalities are considered as zero-valued inputs. In [23], 3D FLAIR images are generated from T1 and later used to train a classifier together with T1 images, which led to a performance improvement. Similarly, in [24] the authors addressed the problem of missing FLAIR sequences in *white matter hyper-intensity segmentation* task by generating the FLAIR images from T1 images while performing the segmentation at the same time. Finally, Ge et al. [25], addressed the problem of missing modalities using pairs of GANs: given two modalities for which there are missing samples (eg. T1 and T2), a pair of generators are trained to produce one modality from the other (eg. G_a: T1 $\rightarrow$ T2, G_b: T2 $\rightarrow$ T1) and the corresponding discriminators use the input of the other generator as ground truth (e.g., $D_a(T2, G_b(T1))$ and $D_b(T1, G_a(T2))$ respectively).

Another common approach to deal with missing modalities consists of training a model that is invariant to the input contrast modalities. Working on this idea, Havaci et al. [26] computed a latent vector for each input modality and then combined all the latent vectors into a single representation that accounts for every modality in input. This method was later extended in [27] to allow unlabeled inputs, i.e., to no longer specify which modalities are provided as input. Finally, in [28] the authors exploited Variational AutoEncoders (VAE) [29], to train an encoder for each modality and, at the same time, to generate the missing modalities.

III. Segmentation with Adversarial Networks

Adversarial Networks became popular as *generative models* (Generative Adversarial Networks), in which two networks, called *generator* and *discriminator* are alternately trained to solve a minimax game. In the basic formulation of GANs,

the discriminator network is trained to assign higher scores to samples that comes from the ground-truth and lower scores to sample that are generated by the generator network. The generator is trained to produce samples that can be wrongly classified from the discriminator. As a result, the generator learns to produce samples that follow the distribution of data of the training set from input noise vectors, that are drawn from a so-called *latent space*.

This adversarial mechanism can be used also for solving segmentation tasks: the generator network – dubbed *segmentator* – is trained to produce a *segmentation map* of an input image, while the discriminator takes as input both an image along with its segmentation map and is trained to distinguish generated segmentations from the ground truth ones.

Our approach is based on *SegAN* [3], an adversarial network architecture, that was extended in two respects: (i) a *dice loss* [30] term has been added to the multi-scale loss function employed in [3]; (ii) we provide the discriminator with the input image *concatenated* to its segmentation map, while in *SegAN* the discriminator is provided with the input image *masked* using its segmentation map.

In the remainder of this section we discuss more in details our proposed architecture (dubbed from now on *SegAN-CAT*), the loss function used to train it, and the differences with respect to the original *SegAN*.

A. Network Architecture

The *SegAN-CAT*, shown in Figure 1, involves a *segmentation network* and a *discriminator network*. The segmentation network takes in input an MRI slice and gives in output the segmentation of the slice as a probability label map; the slices have size 160x160xM, where M is the number of MRI modalities used to train the model (in this work M is either 1 or 4); the probability label maps have size 160x160, representing the probability for each pixel of the input slice of being part of the area of interest [1]. Instead, the discriminator network takes in input an MRI slice and its probability label map, which can be either the one computed by the segmentation network or the ground-truth one: the slice and probability label map are either combined or concatenated together as described later in this Section; thus, the network computes a feature vector that represents a multi-scale representation of the input, which is used to compute the loss function during the training of the two networks. Figure 1 shows the structure of the two networks, based on the following types of computational blocks: (i) $S_{in,\,k}$, that is a *segmentation input block* composed of a *2D Convolution* layer with *k* filters of size 4 and *stride=2* [31], [32], followed by a *LeakyReLU* [33] activation layer; (ii) $S_{enc,\,k}$, that is a *segmentation encoder block* has the same structure as the segmentation input block but has a *batch normalization* layer [34], before the activation layer; (iii) $S_{dec,\,k}$, that is a *segmentation decoder block* composed of a *2D Bilinear Upsampling* layer (factor=2) followed by a *2D Convolution* layer with *k* filters of size 3 and *stride=1*, followed by a *batch normalization* layer and a *ReLU* [35] activation layer; (iv) $S_{out,\,k}$, that is a *segmentation output block* composed of a *2D Bilinear Upsampling* layer (factor=2) followed by a *2D Convolution* layer with *k* filters of size 3 and *stride=1*, followed by a *Sigmoid* [36] activation layer; (v) $D_{in,\,k}$, that is a *discriminator input block* composed of a *2D Convolution* layer with *k* filters of size 4 and *stride=2*, followed by a *LeakyReLU* activation layer. (vi) $D_{enc,\,k}$, that is a *discriminator encoder block* has the same structure as the discriminator input block but has a batch normalization layer, before the activation layer. The convolutions weights in all *discriminator blocks* are constrained between *[-0.05;0.05]* for stabilizing the training process [19]. The output of the discriminator, indicated as *Feature Vector* in Figure 1, is computed by concatenating the discriminator input and the flattened output of every discriminator block. The slope for the *LeakyReLU* is 0.3; batch normalization parameters are $\epsilon = 1 * 10^{-}5$ and momentum=0.1 for both networks.

B. Discriminator Network Input

In the *SegAN* architecture the discriminator network is given in input a *masked* slice image, computed by pixel-wise multiplication of the label probability map and each channel of the MRI slice. Instead, in *SegAN-CAT* architecture the label probability map and each channel of the MRI slice are simply concatenated and provided to the discriminator network as input. While the input definition used in the original *SegAN* architecture is more compact, with the input definition we propose it is possible for the discriminator network to extract features that describe also the area of the input MRI that is *not* included into the segmentation. As a result, we expect *SegAN-CAT* to have slightly better generalization capabilities than the architecture introduced in [3].

C. Loss Function

To train our networks, we use the *Multiscale Adversarial Loss*, as in [3]. This particular loss function, applied also in generative problems [37], allows to perform *feature matching* between the ground truth and the output of segmentation network, optimizing also the network weights on features extracted at multiple resolutions rather than focusing just on the pixel level. Thus, the *Multiscale Adversarial Loss* is defined as follows:

$$\ell_{\mathrm{mae}}(f_D(x), f_D(x')) = \frac{1}{L}\sum_{i=1}^{L} ||f_D^i(x) - f_D^i(x')||_1 \qquad (1)$$

where L is the number of layers in the discriminator network; $f_D^i(\cdot)$ is the output of the activation layer after the discriminator block at position $i+1$, e.g., $f_D^1(\cdot)$ is the discriminator input, $f_D^2(\cdot)$ is the output of the first activation layer, etc; x denotes the input of the discriminator when the label probability map is computed by the segmentation

[1]Please notice that in this work we focused on a binary segmentation problem, but our model can be extended to multi-class segmentation problems by computing a probability label map of size 160x160xL, i.e., computing for each pixel a vector that represents the probability distribution among the L labels.

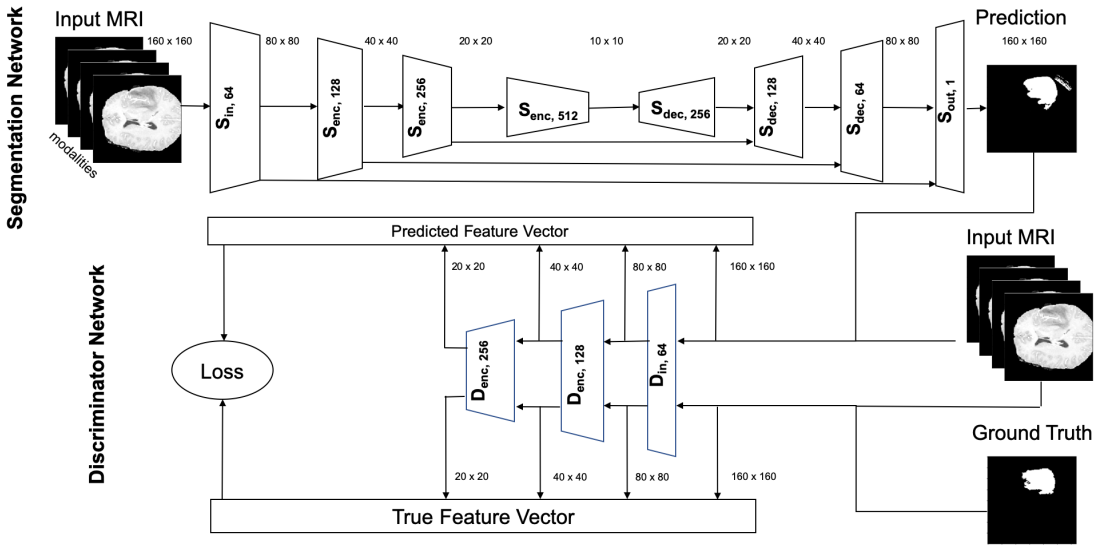

Fig. 1. The *SegAN-CAT* architecture.

network, while x' is the input of the discriminator when the ground-truth is used.

In addition to the *Multiscale Adversarial Loss*, we introduced an additional term to the loss function defined as:

$$\ell_{\text{dice}}(g, p) = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{2}$$

where g is a ground truth image, p denotes the predicted values and the sums run over the pixels of the image; the definition of ℓ_{dice} is based on a differentiable form of the well known *Sørensen-Dice* coefficient [30], or *Dice Score*, it ranges between 0 and 1, and accounts for the overlap between the segmented areas in the ground-truth and in the output of the segmentation network (where 0 means perfect overlap and 1 means no overlap). The use of this term, which allows to assess the quality of the generated segmentation maps, should result in a more stable training process and eventually in a model with better performances.

Overall, the complete objective function used to train the networks is defined as:

$$\min_{\theta_S} \max_{\theta_D} \frac{1}{N} \sum_{n=1}^{N} \ell_{\text{mae}}(f_D(x_n \circ S(x_n)), f_D(x_n \circ y_n)) + \ell_{\text{dice}}(y_n, S(x_n)) \tag{3}$$

where x_n is an input MRI, y_n is the ground truth, and the $\circ$ operator is either a concatenation on the channel axis in *SegAN-CAT* or a pixel-wise multiplication in *SegAN* (as illustrated in Section III-B).

IV. Experimental Design

In this work, we compared the *SegAN-CAT* and the *SegAN* architecture on the brain tumor image segmentation problem based on the *Multi-modal Brain Tumor Image Segmentation Benchmark* (BraTS) [38], a scientific competition organized in conjunction with the international conference on Medical Image Computing and Computer Assisted Interventions (MICCAI) since 2012.

Dataset. We used both the BraTS 2015 [38] and the BraTS 2019 [39] datasets. The Brain tumor segmentation (BraTS) challenge focus on the evaluation of methods for segmenting brain tumors, particularly Gliomas, which are the most common primary brain malignancies. Both the datasets we use are composed of MRI volumes of resolution 240x240x155 and four contrast modalities: T1, T1c, T2, FLAIR. Each voxel is annotated with one among five possible labels representing the tumor sub-regions: *Necrosis* (NCR), *Edema* (ED), *Non-Enhancing Tumor* (NET), *Enhancing/Active Tumor* (AT) and *Everything Else*. The BraTS 2015 dataset consists of 220 high grade gliomas (HGG) and 54 low grade gliomas (LGG) MRIs. The BraTS 2019 is the latest revision of the BraTS dataset and it consists of a total of 335 MRI volumes (259 HGG, 76 LGG); The major similarities and differences between the two datasets are: (i) both datasets contain 30 human-annotated MRIs from a previous dataset, BraTS 2013; (ii) the remaining volumes in BraTS 2015 are a mixture of pre and post-operative scans that have been labeled by an ensemble of algorithms and later evaluated by a team of human experts, while in BraTS 2019 all the post-operative scans have been discarded and all the volumes have been manually re-labeled; (iii) additional data from different institutions have been included in BraTS 2019; (iv) in BraTS 2019, the *Non-Enhancing Tumor* label has been merged with *Necrosis* due to a bias that exists in the evaluation of that area.

Data Preprocessing. Since images in both datasets have an

isotropic resolution of $1mm^3$ per voxels we do not perform any further spatial processing to data. As done in [3], we center-crop each MRI to a 180 x 180 x 128 volume in order to remove black regions while keeping all the relevant data. For each MRI volume, we clip voxel values to the 2^{nd} and 98^{nd} percentile in order to remove outliers, then we apply Feature Scaling [40] to normalize the intensity range between 0 and 1. Finally, we split the data in Training/Validation sets (respectively of size 80% and 20%) using stratified sampling to keep balanced HG and LG subjects within each subset. In Brats2019 dataset we also apply stratified sampling based on the institution that provided the data.

Task Definition. BraTS2015 challenge defines 3 regions of the tumor based on the combination of ground truth labels. Instead, in our experimental analysis, we focused only on one region, that roughly corresponds to the *Whole Tumor* region in BraTS challenge. Accordingly, we re-labeled as 1 all the voxels labeled as NCR, ED, NET, or AT in BraTS datasets, and as 0 the remaining ones.

Evaluation. For assessing the performances of our models, we define 3 metrics derived from the well-known *Confusion Matrix* [41]. To evaluate the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) and thus to compute the metrics, we threshold (using 0.5 as threshold value) the output of the segmentation network in order to obtain a binary classification for each pixel. Despite our models work on a single MRI slice at a time, we compute the metrics by considering the TP, FP, TN, and FN on every slice of the same MRI volume. In particular we considered the following three metrics: (i) the *precision* ($TP/(TP+FP)$), which measures how many voxels, classified as a tumoral lesion, are effectively part of the tumor; (ii) the *sensitivity* ($TP/(TP+FN)$), which measures how many voxels, that are part of the tumor, are correctly identified as such; (iii) the *dice score* ($2TP/(2TP+FP+FN)$), which measures the overlap between the pixels of a binary ground truth and a given prediction.

In our experiment we selected our best model according to the dice score value as it allows to account for both False Positives and False Negatives, leading to a better assessment of the segmentation quality with respect to the precision and sensitivity scores.

Transfer Learning. In our experimental analysis, we used the most straightforward approach to transfer a *SegAN-CAT* model trained from MRIs that include a single modality (*source*) to train a model from MRIs that include a single but *different* modality (*target*): we took the segmentation and discriminator networks trained on the source modality and re-trained (or *fine-tuned*) them on the target modality. In particular, we studied two different fine-tuning process to retrain the networks on the target modality: (i) we adapted all the weights of both the segmentation and the discriminator networks; (ii) we adapted all the weights of the segmentation network and only the weights of the input layer of the discriminator network.

Although keeping the several layers of discriminator fixed during the fine-tuning might prevent it from adapting entirely to the target domain, we believe that this solution could help to retain more knowledge from the source domain while adapting the segmentation network. This choice is motivated by the fact that the discriminator has been found to be the most important part of an adversarial network to transfer to the target domain [16].

Experimental Equipment. We developed and trained all the models described in this paper using the Tensorflow 2.0a Docker environment [42]–[44]. The experiments have been carried out on a server equipped with a 12GB nVidia Titan V, Intel Xeon E5-2609 CPU and 64GB of RAM.

V. Experimental Results

We trained both multi-modal models, i.e., models designed to work with MRIs that include different contrast modalities, as well as uni-modal models, i.e., designed to work with MRIs that include only one specific contrast modality.

In the first experiment we trained three multi-modal models that use all the four contrast modalities, i.e., T1, T1c, T2, FLAIR, available in the datasets: (i) a model that implements the original *SegAN* architecture, (ii) a model that implements the *SegAN* architecture with the additional dice loss term, and (iii) a model that implements completely the *SegAN-CAT* architecture, i.e., with the additional dice loss term and the input discriminator concatenation.

Each model is trained and tested both on BraTS 2015 and BraTS 2019 datasets using a batch size of 64 slices. During the training phase, we perform data augmentation by applying *random cropping* [32] using a window of size 160 x 160, as proposed in [3]. During the validation phase we apply *center cropping* [32] to match the input size of the network, discarding most of the black border of the MRI. All the models are trained using the same initialization seed, RMSprop [45] (*lr: 2*10^{-5}*), and *Early Stopping* [46] (*patience = 300 epochs*) on Dice Score evaluation metric. An epoch consist of a full iteration over the dataset, i.e. approximately 28000 slices for the Brats 2015 dataset and 34000 for the Brats 2019 dataset.

Table I compares the performance of all the multi-modal models trained on two datasets. The results show that both the dice loss term and the discriminator input concatenation, introduced in the *SegAN-CAT*, led to better performances on both datasets. Results also suggest that BraTS 2019 is slightly more challenging than BraTS 2015, leading to a lower performance of all the the three models.

In the second experiment we wanted to investigate how the information content of each contrast modality affects the model performance. To this purpose, we trained four uni-modal models, each one using only a single contrast modality.

Table II compares the performance of these four *SegAN-CAT* models trained using a single contrast modality. The results show that none of the model trained with a single modality is able to achieve the same performance achieved by *SegAN-CAT* which uses all the four contrast modalities together. This

TABLE I

PERFORMANCE OF *SegAN*, *SegAN* WITH DICE LOSS, AND *SegAN-CAT*. THE RESULTS REPORTED ARE THE AVERAGE OF THE DICE SCORE, THE PRECISION, AND THE SENSITIVITY COMPUTED ON EACH ONE OF THE MRI VOLUME INCLUDED IN TEST SET OF BRATS 2015 AND BRATS 2019. WE REPORTED IN BOLD THE BEST PERFORMANCE FOR EACH DATASET.

Model	BraTS 2015			BraTS 2019		
	Dice Score	Precision	Sensitivity	Dice Score	Precision	Sensitivity
SegAN	0.705	0.759	0.694	0.766	0.745	0.834
SegAN w/ Dice Loss	0.825	0.901	0.785	0.814	0.850	0.810
SegAN-CAT	**0.859**	0.882	0.852	**0.825**	0.842	0.835

TABLE II

PERFORMANCE OF *SegAN-CAT* MODELS TRAINED FROM MRIS WITH ONLY ONE CONTRAST MODALITY. THE RESULTS ARE REPORTED FOR EACH MODALITY AS THE AVERAGE OF THE DICE SCORE, THE PRECISION, AND THE SENSITIVITY COMPUTED ON EACH ONE OF THE MRI VOLUME INCLUDED IN TEST SET OF BRATS 2015 AND BRATS 2019. AS A REFERENCE, THE PERFORMANCE OF *SegAN-CAT* TRAINED WITH ALL THE CONTRAST MODALITIES IS REPORTED IN THE LAST ROW OF THE TABLE.

Modality	BraTS 2015			BraTS 2019		
	Dice Score	Precision	Sensitivity	Dice Score	Precision	Sensitivity
T1	0.538	0.570	0.557	0.542	0.586	0.609
T1c	0.578	0.613	0.581	0.587	0.678	0.577
T2	0.721	0.773	0.724	0.675	0.828	0.607
FLAIR	0.810	0.858	0.787	0.763	0.757	0.810
ALL	0.859	0.882	0.852	0.825	0.842	0.835

suggests that none of the four modalities alone contains all the relevant information to identify the tumor. However, the model trained using only the FLAIR modality is able to achieve a much better performance than all the other models on both the datasets. These results are not surprising as FLAIR modality allows to identify more clearly the edema and the lesions in specific areas of the brain, due to the suppression of the cerebrospinal fluid in the images.

In the final experiment we investigated whether it is possible to transfer a model trained on a specific contrast modality to a different contrast modality. Thus, as preliminary analysis, we applied all the uni-modal models previously trained on each of the contrast modality to images taken with different modalities to evaluate the similarities among the models and to have an insight on how easily the knowledge learned from a modality could be transferred to the the others.

Figure 2 shows the performance of all the uni-modal models when applied to each modality alone. As expected, all the uni-modal models reached the best performance on images acquired with the same contrast modality they have been trained for. Therefore, to transfer successfully a model across the contrast modalities, it is necessary to adapt the trained networks with data from the target domain, i.e., the target contrast modality. The data also show how models could be easily transferred across modalities that are known to be similar among them: models trained on T1 and T1c perform poorly on T2 and FLAIR, while they perform much better when applied to the other modality similar to the one they are trained on (i.e., T1 or T1c); the same behavior, the other way around, is found when looking at the performance of the models trained on T2 and on FLAIR.

Based on the results discussed above, we applied the transfer learning to train three uni-modal models respectively on T1, T1c and T2 images by adapting a model trained on FLAIR images. In fact, our results show that the FLAIR images account for a large part of the model performance and our aim is to understand whether the transfer learning process might improve the final performances. In this experiment the fine-tuning of the model on the target modality was limited to 300 epochs.

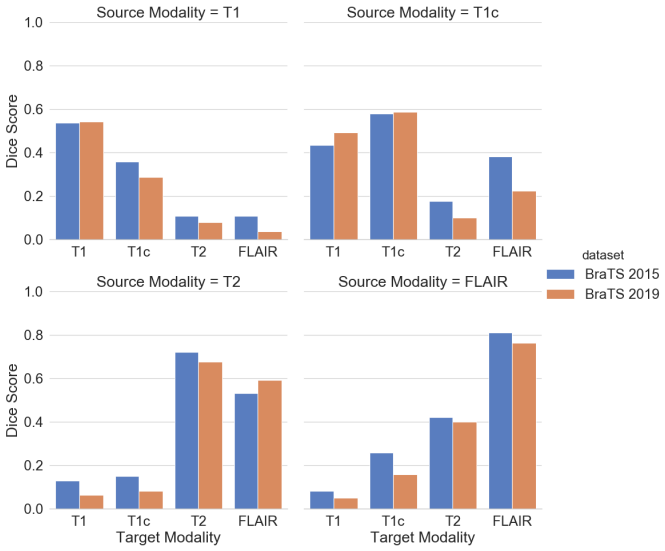

Fig. 2. Performance of each uni-modal model when applied to images acquired using other modalities. The performances on both BraTS 2015 and BraTS 2019 are reported. In addition, we also report as a reference the performance of the uni-modal models on images with the same modality the models are trained for.

Table III shows the performance of the models trained on T1, T1c, and T2 by adapting a model trained on FLAIR. The results suggest that it is more convenient to adapt, i.e.,

TABLE III

PERFORMANCE OF THE MODELS TRAINED BY TRANSFERRING THE MODEL TRAINED ON FLAIR IMAGES. THE COLUMN FINE TUNING REPORTS WHETHER BOTH THE MODEL NETWORKS (S,D) OR ONLY THE SEGMENTATOR AND THE DISCRIMINATOR INPUT LAYER (S,D_{IN}) HAVE BEEN TRAINED ON THE TARGET MODALITY. THE RESULTS ARE REPORTED FOR EACH TARGET MODALITY AS THE AVERAGE OF THE DICE SCORE, THE PRECISION, AND THE SENSITIVITY COMPUTED ON EACH ONE OF THE MRI VOLUME INCLUDED IN TEST SET OF BRATS 2015 AND BRATS 2019. WE REPORTED IN BOLD THE SCORES THAT ARE BETTER THAN THE CORRESPONDING ONES IN TABLE II.

Target	Fine Tuning	BraTS 2015			BraTS 2019		
		Dice Score	Precision	Sensitivity	Dice Score	Precision	Sensitivity
T1	S,D	0.496	0.503	0.553	0.502	0.571	0.582
T1	S,D_{in}	**0.561**	0.614	0.576	0.528	0.558	0.538
T1c	S,D	0.467	0.464	0.538	0.468	0.527	0.494
T1c	S,D_{in}	0.577	0.661	0.541	**0.598**	0.705	0.563
T2	S,D	0.692	0.776	0.668	0.674	0.643	0.775
T2	S,D_{in}	**0.781**	0.818	0.771	**0.741**	0.878	0.681

to train on images with the target contrast modality, *only* the input layer of the discriminator network of the model, keeping the other layers of the discriminator network trained on images with the FLAIR contrast modality. We believe that this result is due to the high instability of the adversarial learning mechanism that makes it very difficult to incrementally adapt a previously trained model [16]. Accordingly, to exploit the knowledge of the model trained on images with FLAIR contrast, we adapted on the target modalities only the segmentator network and the first layer of the discriminator network. Overall, this solution often led to models that perform better or similarly to the models trained from scratch on the target modalities (see Table II), despite being trained for 300 epochs only. As expected, the major benefits of transfer learning mechanism is achieved on the target modality more similar to FLAIR, i.e., T2.

VI. CONCLUSIONS

In this work, we introduced *SegAN-CAT*, an adversarial network architecture based on *SegAN* [3]. Our approach differs from *SegAN* mainly in two respects: (i) the loss function has been extended with a dice loss term and (ii) the input of the discriminator network consists of a concatenation of the MRI images and their segmentation.

We applied *SegAN-CAT* to an MRI brain tumor segmentation problem, the same task the *SegAN* architecture was successfully applied to. In particular, we defined a binary segmentation problem on two datasets, BraTS 2015 and BraTS 2019, used in a scientific competition organized in conjunction with the international conference on Medical Image Computing and Computer Assisted Interventions (MICCAI) since 2012. The aim of this work was (i) to compare the performance of *SegAN-CAT* and *SegAN*, (ii) to assess the performance of uni-modal models for each contrast modality (i.e., T1, T1c, T2, and FLAIR), and (iii) to study the problem of transferring a previously trained model across different modalities. To this purpose, we first trained *SegAN*, *SegAN* with a dice loss term, and *SegAN-CAT* on MRIs acquired with all the four contrast modalities. Our results on both BraTS 2015 and BraTS 2019 datasets showed that both the dice loss term and the discriminator input concatenation proposed in this paper allow to improve the final performance of the model. Then, we trained one uni-modal *SegAN-CAT* model for each one of the four contrast modalities in order to assess the performance that can be achieved using only the information available in each modality alone. As expected, none of these four models reached the performance of the multi-modal one. However, the model trained on FLAIR contrast modality outperformed all the others and was able to reach a performance quite close to the one achieved by the multi-modal model. Thus, we investigated the possibility of transferring a model trained with FLAIR to train better model from images acquired with different contrast modalities. Despite confirming that is rather difficult to use transfer learning with adversarial networks as discussed in [16], [18], our results suggest that it is possible to successfully transfer a model trained on FLAIR contrast modality also to other modalities. Indeed, our results show that transfer learning often allows to train uni-modal models with a better performance than the same models trained entirely on images with their specific contrast modality.

We believe that the transfer learning with adversarial networks is a topic worth to be further investigated in future works. In fact, the limited amount of data is one of the major issue when it comes to the application of deep learning to healthcare domain. So far, transfer learning has been widely used in several domains to deal with the limited availability of data, but it has been not yet fully exploited in the healthcare domain. In particular, applying transfer learning to train adversarial networks is a challenging and yet rather unexplored research area that deserve additional investigation. In future works we plan to investigate additional strategies to adapt a previously trained model to a target domain, such as adapting only some layers of the networks, alternating previously trained and untrained networks during the adapting process, and combining together networks trained using different contrast modalities.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[3] Y. Xue, T. Xu, H. Zhang *et al.*, "Segan: Adversarial network with multi-scale l_1 loss for medical image segmentation," *CoRR*, vol. abs/1706.01805, 2017. [Online]. Available: http://arxiv.org/abs/1706.01805

[4] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *CoRR*, vol. abs/1611.08408, 2016. [Online]. Available: http://arxiv.org/abs/1611.08408

[5] M. Everingham, L. Van Gool, C. K. I. Williams *et al.*, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[6] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 1–8.

[7] P. Moeskops, M. Veta, M. W. Lafarge *et al.*, "Adversarial training and dilated convolutions for brain MRI segmentation," *CoRR*, vol. abs/1707.03195, 2017. [Online]. Available: http://arxiv.org/abs/1707.03195

[8] M. Havaei, A. Davy, D. Warde-Farley *et al.*, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[9] K. Kamnitsas, C. Ledig, V. F. Newcombe *et al.*, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[10] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *CoRR*, vol. abs/1411.1792, 2014. [Online]. Available: http://arxiv.org/abs/1411.1792

[11] P. Moeskops, J. M. Wolterink, B. H. M. van der Velden *et al.*, "Deep learning for multi-task medical image segmentation in multiple modalities," *CoRR*, vol. abs/1704.03379, 2017. [Online]. Available: http://arxiv.org/abs/1704.03379

[12] S. U. H. Dar and T. Çukur, "A transfer-learning approach for accelerated MRI using deep neural networks," *CoRR*, vol. abs/1710.02615, 2017. [Online]. Available: http://arxiv.org/abs/1710.02615

[13] S. Motamed, I. Gujrathi, D. Deniffel *et al.*, "A transfer learning approach for automated segmentation of prostate whole gland and transition zone in diffusion weighted mri," 2019.

[14] J. Jiang, Y.-C. Hu, N. Tyagi *et al.*, "Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, A. F. Frangi, J. A. Schnabel, C. Davatzikos *et al.*, Eds. Cham: Springer International Publishing, 2018, pp. 777–785.

[15] M. Ghafoorian, A. Mehrtash, T. Kapur *et al.*, "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz *et al.*, Eds. Cham: Springer International Publishing, 2017, pp. 516–524.

[16] Y. Wang, C. Wu, L. Herranz *et al.*, "Transferring gans: Generating images from limited data," in *ECCV*, 2018.

[17] I. Gulrajani, F. Ahmed, M. Arjovsky *et al.*, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, 2017, pp. 5767–5777.

[18] Y. Frégier and J. Gouray, "Mind2mind : transfer learning for gans," *CoRR*, vol. abs/1906.11613, 2019. [Online]. Available: http://arxiv.org/abs/1906.11613

[19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv e-prints*, p. arXiv:1701.07875, Jan 2017.

[20] S. U. H. Dar, M. Yurt, L. Karacan *et al.*, "Image synthesis in multi-contrast MRI with conditional generative adversarial networks," *CoRR*, vol. abs/1802.01221, 2018. [Online]. Available: http://arxiv.org/abs/1802.01221

[21] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017. [Online]. Available: http://arxiv.org/abs/1703.10593

[22] A. Sharma and G. Hamarneh, "Missing MRI Pulse Sequence Synthesis using Multi-Modal Generative Adversarial Network," *arXiv e-prints*, p. arXiv:1904.12200, Apr 2019.

[23] B. Yu, L. Zhou, L. Wang *et al.*, "3d cgan based cross-modality mr image synthesis for brain tumor segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, April 2018, pp. 626–630.

[24] M. Orbes-Arteaga, M. J. Cardoso, L. Sørensen *et al.*, "Simultaneous synthesis of FLAIR and segmentation of white matter hypointensities from T1 mris," *CoRR*, vol. abs/1808.06519, 2018. [Online]. Available: http://arxiv.org/abs/1808.06519

[25] C. Ge, I. Y. Gu, A. Store Jakola, and J. Yang, "Cross-modality augmentation of brain mr images using a novel pairwise generative adversarial network for enhanced glioma classification," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 559–563.

[26] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "Hemis: Hetero-modal image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu *et al.*, Eds. Cham: Springer International Publishing, 2016, pp. 469–477.

[27] T. Varsavsky, Z. Eaton-Rosen, C. H. Sudre *et al.*, "PIMMS: permutation invariant multi-modal segmentation," *CoRR*, vol. abs/1807.06537, 2018. [Online]. Available: http://arxiv.org/abs/1807.06537

[28] R. Dorent, S. Joutard, M. Modat *et al.*, "Hetero-Modal Variational Encoder-Decoder for Joint Modality Completion and Segmentation," *arXiv e-prints*, p. arXiv:1907.11150, Jul 2019.

[29] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv e-prints*, p. arXiv:1312.6114, Dec 2013.

[30] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, 2016. [Online]. Available: http://arxiv.org/abs/1606.04797

[31] Y. LeCun, B. E. Boser, J. S. Denker *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396–404.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[33] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1, 2013, p. 3.

[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[35] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[36] T. M. Mitchell, *Machine Learning*, 1st ed. USA: McGraw-Hill, Inc., 1997.

[37] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu *et al.*, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[38] B. H. Menze, A. Jakab, S. Bauer *et al.*, "The multi-modal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct 2015.

[39] S. Bakas, M. Reyes, A. Jakab *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BraTS challenge," *CoRR*, vol. abs/1811.02629, 2018. [Online]. Available: http://arxiv.org/abs/1811.02629

[40] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern recognition letters*, vol. 22, no. 5, pp. 563–582, 2001.

[41] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997.

[42] M. Abadi, A. Agarwal, P. Barham *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[43] D. Merkel, "Docker: lightweight linux containers for consistent development and deployment," *Linux journal*, vol. 2014, no. 239, p. 2, 2014.

[44] C. Boettiger, "An introduction to docker for reproducible research," *ACM SIGOPS Operating Systems Review*, vol. 49, no. 1, pp. 71–79, 2015.

[45] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[46] Y. Yao, L. Rosasco, and A. Caponnetto, "On early stopping in gradient descent learning," *Constructive Approximation*, vol. 26, no. 2, pp. 289–315, 2007.