Data Efficient Segmentation of Various 3D Medical Images Using Guided Generative Adversarial Networks

HOSNA ASMA-ULL, IL DONG YUN, (Member, IEEE), AND DONGJIN HAN

1Department of Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin 17055, South Korea
2Department of Broadcasting and Film, Cheju-Halla University, Jeju City 63092, South Korea

ABSTRACT

The recent significant increase in accuracy of medical image processing is attributed to the use of deep neural networks as manual segmentation generates errors in interpretation besides, is very arduous and inefficient. Generative adversarial networks (GANs) is a particular interest to medical researchers, as it implements adversarial loss without explicit modeling of the probability density function. Medical image segmentation methods face challenges of generalization and over-fitting, as medical data suffers from various shapes and diversity of organs. Furthermore, generating a sufficiently large annotated dataset at a clinical site is costly. To generalize learning with a small amount of training data, we propose guided GANs (GGANs) that can decimate samples from an input image and guide networks to generate images and corresponding segmentation mask. The decimated sampling is the key element of the proposed method employed to reduce network size using only a few parameters. Moreover, this method yields promising results by generating several outputs, such as bagging approach. Furthermore, errors of loss function increase, during the generation of original images and corresponding segmentation mask, in comparison to generating only the segmentation mask. Minimization of increased error leads (GGANs) to enhance the performance of segmentation using smaller datasets and less testing time. This method can be applied to a wide range of segmentation problems for different modalities and various organs (such as aortic root, left atrium, knee cartilage, and brain tumors) during a real-time crisis in hospitals. The proposed network also yields high accuracy compared to state-of-the-art networks.

INDEX TERMS

Medical image segmentation, generative adversarial network, aortic valve, left atrium, knee cartilage, brain tumor, decimated sample.

I. INTRODUCTION

In medical image interpretation, specialists decision is most challenging tasks as this directly influenced by their experience. Segmentation is applied for medical images to identify different anatomical structures through the human body, such as bones, blood vessels, vertebrae, and major organs. However, these images suffer from noise, artifacts, variation of machines, variability of organ shape, size, and orientation [12], [24], [30]. These create challenges for segmentation. Even specialists can easily generate errors in interpretation. To generate and maintain an annotated dataset is very costly, laborious, and inefficient. Under these circumstances, reliable and less complex automatic segmentation technique using deep learning is greatly required to hasten the segmentation process for medical images. However, many processes are developed using deep learning in recent times. But these are designed for a specific organ or with a large dataset. In professional fields, an accurate reliable approach needs to achieve to rely upon. Such motivation inspires us to contribute our idea to enhance medical image analysis quality with a generalized process for various organs. However, this is also efficient in terms of smaller datasets, testing time, and accuracy.

We present an overview of the proposed (medical image) segmentation procedure that is designed based on the latest...
TABLE 1. Challenges of various data-sets.

| Organ          | Challenge                                                                                                                                 |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Aortic Valve   | AV surrounded by other organs with similar three prior intensity regions and an invisible boundary between the aortic valve and left ventricle having similar intensity. |
| Left Atrium    | Thin myocardial wall making it difficult to distinguish boundaries from other surrounding organs such as LV. Surrounding organs, like pulmonary veins, varies cause of topological variations. The shape and size of the left atrial appendage (LAA) varied. |
| Knee           | Every volume has a different size (height - width). The size of the upper and lower cartilage is very small. The true location of the exact boundary of cartilage is considerably far from the measurement. |
| Brain Tumor    | Shape and appearances vary for different tumor regions. Tumors are heterogeneous and limited in type and volume.                                                                                   |

deepllearning approach: generative adversarial networks (GANs), where two networks (generator and discriminator) compete against one another to generate an image. The generator tries to learn how to generate samples resembling real data, and the discriminator tries to learn how to discriminate real and generated data. Thus, generator attempts to minimize loss function while discriminator strives to maximize it, simultaneously. However, traditional GANs use noise to generate resemble data or entire image to generate segmentation mask. This proposed network employs sole labeling of each pixel, similar to other traditional GANs, but this represents the first attempt to guide architecture using decimated samples of images (original, immediately prior, and post slice). The reason behind using a decimated sample instead of the entire image is, samples require fewer parameters and extract relevant information using multiple images and this, in turn, helps to stabilize architecture with smaller datasets. In contrast, complete image increases parameters exponentially and creates an over-fitting problem with small datasets. Therefore, the entire image leads (architecture) toward instability. Furthermore, this decimated sample technique can generate internal several outputs (such as bagging approach [5]) and ultimately yield promising segmentation mask. The proposed method is evaluated on various image modalities and different human organs. Furthermore, this can be considered as efficient for real-time applications like diagnosing various diseases of patients in a short period during an emergency. The key challenges encountered for different datasets summarize in Table 1 and a detailed explanation provides in section IV-(A).

The contribution of the proposed method is subdivided into three points based on the uniqueness of approach, performance of the network, and information utilization:

- To best of knowledge, for the first time, this approach uses decimated samples to guide GANs, including medical image segmentation.
- Segmentation performance is improved by generating original images and corresponding segmentation mask in separate channels. The algorithm performs well despite the variability among human organs (e.g., aortic valve, left atrium, knee, and brain) and various image modalities, with smaller training datasets. Furthermore, this is very efficient for real-time applications like diagnosing disease in a short time as testing takes less than 1 second (for each slice).
- Spatial detail is preserved using a skip mechanism to improve accuracy by obtaining a precise boundary despite using a few parameters with decimated samples of the images.

The rest of the paper is organized as follows. Related work is described in Section II. The decimation of the sample extraction process, objective function, and proposed architecture of networks are presented in Section III. In section IV, results are presented and analyzed. Section V concludes the paper by summarizing it with a recommended idea and notions to implement in the future.

II. RELATED WORK

Deep learning is the most advanced technology employed to date for medical image segmentation. Among deep learning techniques, generative adversarial networks (GANs) [13] are promising addition that capable of implementing a significant number of parameters, which are trainable using deep convolutional neural networks. In network, GANs takes samples from a fixed distribution, such as Gaussian, and transforms samples utilizing a deterministic differentiation deep network to approximate distribution of training samples. GANs are highly effective in generating realistic images as they can learn local and global information of pixels. This capability helps to overcome certain disadvantages such as image blur and sensitivity of outlines, which require to use traditional pixel-wise loss functions, such as SOFTMAX in convolutional neural networks (CNNs) [20]. Furthermore, this helps overcome the lack of training annotations. However, the architecture of GANs suffers from stability, which is addressed by deep convolutional GANs (DCGANs) [25]. Conditional GANs (cGANs) [3] includes additional information to control image generation. The authors used additional information, such as class labels, and showed that it improves training stability while preserving detailed features of the generated image. Another conditional GANs framework referred to as Markovian GANs (MGANs) [23]
exhibits fast and high-quality style transfer while simultaneously preserving image content. In [27], two public datasets, namely, DRIVE and STARE, were employed using U-net for retina segmentation [26]. They likewise incorporated skip-connected [15] idea, which helped preserve low-level features, such as edges and blobs. These features are considered crucial for segmentation accuracy. A segmentation pipeline used for both annotated and unannotated images in [31], where element-wise and adversarial losses were applied for the annotated image. Next, multi-scale L1 loss employed to propose SegANs [29] to enforce multi-scale spatial constraints that helped in achieving state-of-the-art performance in BRATS challenge of 2013 and 2015. However, GANs for medical image segmentation (MI-GANs) [17] generate synthetic medical images from noise to enhance dataset, and then train model using original and synthetic images to generate segmentation mask. Recently, semantic segmentation was applied using GANs for spine structures, where a recurrent network called spine-GANs performed automated segmentation and classification of various organs in magnetic resonance imaging (MRI) image, in [14]. Furthermore, two other recent methods were proposed in [16] and, [19] to segment images of human organs. In conventional approaches, GANs generate a synthetic image or segmentation mask from random noises or complete original image, respectively. The networks, however, suffer from the generalization of architecture for different datasets and over-fitting in case of smaller datasets. Moreover, the size of the annotated dataset is limited, in comparison to the field of computer vision. Because of sole dependency on medical experts for manual segmentation to generate an annotated medical dataset is costly and time-consuming.

Several past key studies on different datasets discuss as follows. There are numerous publications in literature specific to the topic of aortic valve segmentation [6], [18], [21]. Compared to other approaches addressing the aortic root segmentation from computed tomography (CT) scans, a complete automatic segmentation procedure provided in [33], where a marginal space learning technique uses for pre- or post-operative planning. Furthermore, in [34], a technique employs to segment the left atrium to propagate a single atlas to an unseen image, where the propagation performed by local affine and deformable registration techniques. The improved concept of multiple atlases is presented in [35] and [32]. The significance necessity of knee cartilage segmentation described in [10]. The study shows that the primary reason for chronic disability is osteoarthritis (OA) in knee joints. Advances in medical science help observe this effect through MRI [7]. Among numerous research viewpoints, biomarkers segmentation use to identify the stages of osteoarthritis [9]. Knee implantation is another additional key issue, where image segmentation of bones around knee and cartilage is crucial. Studies addressing the issue of predicting joint kinematics of knee [4] and identifying the health of knee joints [11] are likewise worth mention. Lastly, Lee et al. [22] introduced a method to segment different organs of the knee by emphasizing on local shape and appearance.

III. PROPOSED METHOD

Numerous constraints require to consider in the case of biomedical images cause more than one class needs to define in such images. Thus, to improve medical image analysis, we propose a novel method involving advanced GANs architecture for medical image segmentation.

A. DECIMATE SAMPLE

This section describes how this newly designed architecture attempts to encompass traditional GANs for segmentation using decimated samples. We begin by describing the extraction process of decimated samples. Decimation is a term that historically denotes the removal of a tenth. However, in signal processing, decimation by a factor of ten implies retaining only every tenth sample. Here, we define this procedure as keeping the $M$th sample, when this says to decimate image by $M$. For example, if the image is decimated by 4, it means that 25% of pixels retain from images. Similarly, 50%, 75%, and 100% of pixels are retained for image decimation by 2, (4/3), and 1, respectively. Furthermore, ground-truth (gt) also samples following a similar procedure. Subsequently, architecture is guided based on these decimated samples. Notably, this network uses four-channels, where images (original, immediately prior, and post slice) and ground-truth assemble through 2nd, 1st, 3rd, and 4th channels, respectively. Furthermore, before extracting samples, the proposed method, converts images to a 1D array.

Decimated samples do not affect the distinguishing errors of real images and generated ones, as the latter is a similar size to the input image. However, the reason behind reducing the dataset size and parameters is, this architecture designs to handle parameters based on various decimated sample sizes (Section III-(B)). Therefore, the increasing number of parameters requires a larger dataset to overcome the over-fitting problem. Moreover, the reason for choosing 2D GANs over 3D is, to reduce parameters number. 3D GANs increase parameters number exponentially compared to 2D GANs. As a consequence, they require a larger dataset, which is contradictory to the objective of the proposed method.

B. GUIDED GENERATIVE ADVERSARIAL NETWORKS (GGANs)

The architecture is termed as guided generative adversarial networks (GGANs), as it comprises the basic idea of traditional GANs, where two networks compete against one another to generate an image. The generator tries to learn how to generate samples resembling real data, and the discriminator tries to learn how to discriminate real and generated data. Thus, generator attempts to minimize loss function while discriminator strives to maximize it, simultaneously. This simultaneous competition leads architecture toward a Nash equilibrium, where neither can further unilaterally minimize or maximize loss function. Finally, the discriminator...
of GANs provides an abstract unsupervised representation of images. Unlike in traditional GANs, generator of this network applies multi-channel decimated samples extraction from images (original, immediately prior, and post slice) and ground-truth, to use as a guide for generating images and corresponding segmentation mask. However, our proposed method employs U-net to extract spatial information apart from initial features such as blobs and edges. In this architecture (Fig. 1 shows training session), the generator comprises four convolutions (followed by max-pooling with ReLu activation) and deconvolution (followed by ReLu) layers. The convolution-pooling layers organize, from upper (C1, P1) to lower (C4, P4) layer, and deconvolutional layers, in reverse order from lower (D1) to upper (D4) layer. Furthermore, three deconvolution layers (D1, D2, and D3) connect using skip-connection (red arrow) with three convolutional layers (C1, C2, and C3). In contrast, convolutional layers of discriminator are followed by leaky ReLu, as this passes a small negative gradient signal for negative values instead of passing a gradient (slope) of zero for back-propagation to allow the flow of a stronger gradient into the generator. The configuration of the later stage is the same as that of the generator. The sigmoid and tanh activation functions are used for discriminator and generator, respectively.

Architecture processes input images in a specific manner (explained in earlier Section III-(A)) besides handling functionality of generator and discriminator. However, the reason behind including immediate images is to collect relevant information, as medical image encompasses relevant useful information from immediate images compared to the computer vision field. Furthermore, due to compact 2D architecture, the relevant information is considered crucial in improving segmentation performance. On the other network, discriminator identifies the difference between all generated and original images. Thus, errors of real and generated images (four-channel) become non-negligible for the network to minimize easily, compared to the case of generating only segmentation mask. Consequently, generator and discriminator lead architecture to improve the overall performance of segmentation with smaller datasets. However, to calculate final results, Dice similarity coefficient (DSC) scores are considered only for segmentation mask. For testing, the ground-truth channel provided with all zeros and architecture generates (segmentation) mask alongside original images using previously trained parameters.

C. VALUE FUNCTION

Recently, GANs have flourished as one of the most promising deep learning frameworks. They first introduced in [13]. The entire system is a combination of two-distinctive networks, namely, generator, referred to as G, and discriminator, as D. Primary function of the generator is to learn features of similar data to generate a fake image analogous to the real one. Discriminator distinguishes between real data (from true distribution $p_{data}(x)$) and fake data (generated by generator). These two networks operate against each other as long as discriminator is being fooled by trained generator following formulation mentioned in (1):

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

(1)

$E$ is an empirical estimate of expected probability. $G$ transforms a noise variable $p_z(z)$ into $G(z)$, which is basically
TABLE 2. Summary of different dataset.

| Organ               | Vol | Type | Height | Width | Depth | Availability          |
|---------------------|-----|------|--------|-------|-------|-----------------------|
| Aortic Valve        | 100 | CCTA | 512    | 512   | 240   | In – House            |
| Left Atrium         | 10  | MRI  | 320    | 320   | 150   | MICCAI’13             |
| Left Atrium         | 20  | MRI  | 320    | 320   | 150   | Decathlon             |
| Knee Cartilage      | 100 | MRI  | 250    | 250   | 105   | SKT10                 |
| Brain Tumor         | 100 | MRI  | 256    | 256   | 150   | Decathlon             |

A sample from distribution $P(z)$. Ideally, distribution $P(z)$ should converge model. Minimizing $\log(1 - (D(G(z))))$ is equivalent to maximizing $\log(D(G(z)))$. However, $D(x)$ represents the probability that came from the data rather than distribution.

Now, if this system proceeds to generate only segmentation mask in (2) using guidance from decimated samples of the real image, then objective function changes as follows:

$$
\min_G \max_D V(D, G)
= E_{(gt)-P_{data}(gt)}[\log(D(gt))]
+ E_{(\bar{x}, \bar{gt})-P_{x,\bar{gt}}(\bar{x}, \bar{gt})}[\log(1 - D(G(\bar{x}, \bar{gt})))]
$$  (2)

Here, $gt =$ ground-truth, $\bar{x} =$ decimated samples of images (original, immediately prior, and post slice), $gt =$ decimated samples of ground truth of original image. Beginning of this section, extraction process of original images and ground-truth are explained. $D(gt)$ calculates the probability for the generation of segmentation mask that came from the data $x$ rather than distribution $D(G(\bar{x}, \bar{gt}))$. This error consider for only segmentation mask.

Hence, if the system proceeds to generate segmentation mask along with images (original, immediately prior, and post slice) using guidance from decimated samples, then value function changed in (3) as follows:

$$
\min_G \max_D V(D, G)
= E_{(x, gt)-P_{data}(x, gt)}[\log(D(x, gt))]
+ E_{(\bar{x}, \bar{gt})-P_{x,\bar{gt}}(\bar{x}, \bar{gt})}[\log(1 - D(G(\bar{x}, \bar{gt})))]
$$  (3)

Here, $x =$ images (original, prior, and post slice), $gt =$ ground-truth, $\bar{x} =$ decimated sample of images (original, immediately prior, and post slice), $\bar{gt} =$ decimated samples of ground truth of original image. The above equation shows that discriminator attempts to minimize the error of $\log(1 - (D(G(x, gt))))$. Discriminator considers the error of images and ground-truth, $G(x, gt)$ that is considerably higher than the case in which only segmentation mask is generated, where the error is $G(gt)$. Therefore, discriminator attempts to minimize further error, thereby enhancing accuracy.

However, it also has an impact on loss function of the generator, which is defined as follows:

$$
\min_G V(G)
= Loss(\text{Imagechannel}(x, gt)-P_{data}(x, gt))
+ \alpha E_{(\bar{x}, \bar{gt})-P_{x,\bar{gt}}(\bar{x}, \bar{gt})}[\log(1 - D(G(\bar{x}, \bar{gt})))]
$$  (4)

The generator of this architecture attempts to minimize errors in generated images. Unlike the traditional method, this procedure considers two more images (immediately prior and post slices of the original image, which incorporates in additional two-channels as input) to extract relative information. Thus, architecture considers errors of original image and segmentation mask along with two additional images, which reflect first loss in above equation. Furthermore, architecture considers loss parameter $Imagechannel(x, gt)$ along with overall discriminator prediction error of discriminator $\log(1 - (D(G(\bar{x}, \bar{gt}))))$. After considering this channel-wise error, the total error becomes non-negligible for GANs value function, which leads architecture to generate a sharp image, compare to errors during the generation of only segmentation mask. Here, regularize parameter $\alpha$ is used to stabilize the network.

IV. EXPERIMENTAL EVALUATION AND DISCUSSION

To evaluate the proposed method, segmentation performed on five different types of 3D medical datasets, including that of the aortic valve, left atrium (two different datasets), knee cartilage, and brain tumor. Notably, there are variations in the modality of datasets, including those of Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). The challenges of datasets summarize in Table 1. In particular, it is difficult to segment aortic valve (AV) owing to the indistinguishable wall between aortic valve (AV) and left ventricle (LV). They are sharing similar intensity levels. Moreover, left atrium (LA) has a thin myocardial wall, which makes it difficult to distinguish from the boundary with other surrounding organs, such as left ventricle. Furthermore, the shape and size of LA appendage (LAA) differ. Unlike other organs, challenges in the case of knee segmentation are the sharpness, position, and thinness small sizes of upper and lower cartilage. Furthermore, the exact boundary of cartilage may be considerably far from its true location. Moreover, shapes and appearances vary for various regions of brain tumor (edema/tumor infiltration, enhancing tumor core and non-enhancing tumor core). Furthermore, a summary of different datasets is presented in Table 2. This table indicates that left atrium, knee, and brain tumor datasets were acquired from public open sources such as MICCAI’13, medical decathlon Decathlon website, and SKI10. However, aortic valve data collected in private, whose ground-truth also marked by experts. Besides, height, width, and depth of organs vary across different datasets.
TABLE 3. Key characteristics of different methods.

| Methods               | Key characteristics of different methods                                                                 |
|-----------------------|-----------------------------------------------------------------------------------------------------------|
| Traditional GANs or GANs | In this architecture, the generator does not consist of a U-net connection and extraction of decimating samples. It uses the entire image as input and generates a segmentation mask as sole output. |
| Traditional U-net     | The traditional U-net architecture use without GANs. It uses the entire image as input and generates a segmentation mask as sole output. |
| Traditional GANs + U-net | In this architecture, the generator comprises a U-net connection but does not extract decimated samples. It uses the entire image as input and generates a segmentation mask as sole output. |
| Only mask             | The generator comprises U-net connection and decimated samples extraction process. However, it uses the decimated sample as input to generate a segmentation mask as a sole output. |
| Decimated sample + 2-channel | The generator comprises (both) U-net connection and decimated samples extraction process. Four types of decimated samples are used (25%, 50%, 75%, and 100%). Here, decimated samples of images (original, immediately prior, and post slice) are used as input and generate images (along with segmentation mask). |
| Proposed               | The generator comprises (both) U-net connection and decimated samples extraction process of decimated samples, where a 25% decimated sample is used. It also uses the decimated sample of images (original, immediately prior, and post slice) as input and generates images (along with segmentation mask). |

A. ANALYSIS OF DIFFERENT METHODS

This experiment conducted using various structures of architecture. A summary of the key characteristics of different methods is listed in Table 3. Despite the diversity in design, the entire image is used as input and generates a segmentation mask, which is only output produced in the first three methods. However, (both) U-net connection and decimated sample (by 25%, 50%, 75%, and 100%) extraction are used in the last two methods. The samples are used as input to generate images and segmentation mask. Thus, the last two methods distinguish more errors compared to the rest of the methods and generate the most promising result (Equation-3). Between them, last-one extracts 25% decimated samples and considers as “Proposed” method. “Mask only” refers to architecture comprising (both) U-net connection and decimated samples extraction process, which generates only mask (Equation-2).

In this experiment, GPU ‘GeForce GTX TITAN X’ is used with fourfold cross-validation. Training datasets were divided into two subgroups for training and testing (as ground-truth of the testing dataset is unavailable), it is important to effectively tune parameters to obtain the optimal result as this architecture designed for varying numbers of decimated samples and image sizes. In this experiment, Adam optimizer with a learning rate of 0.0002, convolutional kernel size (3 x 3), pooling kernel size (2 x 2) with stride 2, and batch size 32, is used for all datasets. Filters 32, 64, 80, and 80 are used for C1, C2, C3, and C4, convolution layers, respectively. Images are cropped to cover the area of organs, as architecture is capable of treating various image sizes and decimated samples.

To determine parameters, an example is given as follows. Consider, a 50% decimated sample for an (128 x 128) input image. The sizes of decimated input and final output image are (64 x 64) and (128 x 128), respectively. The generator of architecture calculates (32 x 32) output of first convolution-pooling layer (C1, P1). Subsequently, output of (C2, P2), (C3, P3), and (C4, P4) layers is (16 x 16), (8 x 8), and (4 x 4), respectively. After obtaining the output from last convolution-pooling (C4, P4) layer, generator calculates intermediate image size and filter numbers to obtain the output of the first deconvolutional (D1) layer. Skip-connection is then used for the following three deconvolutional layers, as mentioned earlier. Output image (64 x 64) is obtained from D3 layer. Final (D4) deconvolution layer generates an output image (128 x 128), which is similar to the input image. This experiment generates the most promising result compared to other state-of-art methods using only 25% of decimated samples.

The qualitative results of five different organs are illustrated in Fig. 2. Different datasets are used, including AV, knee cartilage, LA, and brain tumor depicted in row-1 to row-4, respectively. Original image, ground truth, segmentation for only mask generation, and segmentation mask for the images and corresponding segmentation mask, are organized in column-1 to column-4, respectively. The figure shows that architecture successfully generates mask, despite the diversity of datasets. Only mask generation detects false regions and generates more noise compared to the generation of images and corresponding segmentation mask.
FIGURE 2. Different datasets showing aortic valve, knee cartilage, left atrium, and brain tumor depicted in row-1 to row-4, respectively. Original image, ground truth, segmentation mask for mask generation only, and segmentation mask for images and segmentation mask, are organized in column-1 to column-4, respectively. Only mask generation detects false regions and generates more noise compared to the generation of images and segmented mask.

TABLE 4. Variation of DSC score depending on percentage of different decimated sample.

| Method                        | AV  | LA(1) | KC  | LA(2) | BT(L1) | BT(L2) | BT(L3) |
|-------------------------------|-----|-------|-----|-------|--------|--------|--------|
| Traditional GANs             | 0.68| 0.60  | 0.50| 0.61  | 0.45   | 0.30   | 0.45   |
| Traditional U-net            | 0.80| 0.78  | 0.63| 0.79  | 0.57   | 0.37   | 0.56   |
| Traditional GANs + U-net     | 0.91| 0.85  | 0.68| 0.85  | 0.60   | 0.44   | 0.60   |
| Decimated sample(25%) + two-channel | 0.96| 0.93  | 0.68| 0.93  | 0.67   | 0.46   | 0.68   |
| Decimated sample(50%) + two-channel | 0.95| 0.90  | 0.73| 0.91  | 0.65   | 0.45   | 0.66   |
| Decimated sample(75%) + two-channel | 0.94| 0.89  | 0.71| 0.88  | 0.63   | 0.43   | 0.65   |
| Decimated sample(100%) + two-channel | 0.92| 0.87  | 0.70| 0.85  | 0.62   | 0.42   | 0.64   |
| State-of-art                 | 0.95(a*)| 0.93(b*)| 0.73(c*)| 0.93(d*)| 0.68(d*)| 0.46(d*)| 0.68(d*)|

Overall results based on various decimated samples percentages of different methods summarize in Table 4, including those of state-of-art results. Complete datasets of all organs are used for this experiment. DSC scores for various datasets using traditional methods are listed in the top three rows. Four rows (last) indicate scores for proposed methods for different decimated samples. Results indicate that architecture produces the most promising result using a decimated sample of 25% (marked bold). Parameter increases exponentially as a consequence of the increased number of decimated samples (of the last four rows), which affects the performance of architecture. This gradual digression of performance with an increasing number of decimated samples is due to overfitting. However, architecture shows a similar pattern across different datasets, which defines the generalization capability of proposed methods. In this table, the best result is marked bold for various parameters of the proposed method. The state-of-art result is not marked in bold, as these are collected from five different sources. The aortic valve, left atrium (MICCAI), knee cartilage, left atrium (medical Decathlon),
brain (Edemas), brain (non-enhancing) and brain (enhancing) are represented as AV, LA(1), KC, LA(2), BT(L1), BT(L2), and BT(L3), respectively. The state-of-the-art sources of AV, LA (MICCAI), knee cartilage (SKI), LA, and brain tumor (medical Decathlon website) refer to (a*), (b*), (c*), and (d*), respectively. These are cited from [8], [28], [1], and [2], respectively. However, all datasets perform optimally with 25% of decimated samples, except for knee cartilage, which performs best with 50% of decimated samples. Regarding AV, a previous study on AV segmentation [8] achieved a DSC score of 0.95. Although AV dataset of the proposed method was acquired in private, it was manually marked by experts, and therefore, this result is considered comparable to that obtained in previous study. The LA for MICCAI'13 achieved promising results with a benchmark [28], and our proposed method was able to achieve this result. Remaining results of medical Decathlon (LA, brain tumor) and SKI’10 (knee cartilage) were collected from the website as reference study is not available for citation.

DSC score based on the variation of dataset size is presented in Table 5. Here, 25% of decimated samples are used. Results indicate that this proposed method achieved the best DSC score (approximately) after using only 50% of datasets (marked bold), where traditional GANs achieve similar performance after using entire datasets. DSC score difference between the proposed method and traditional GANs is significant when 50% of datasets are employed. Furthermore, this table also shows that the proposed method achieves better DSC scores compared to the case when only the segmentation mask is generated. Different datasets exhibit up to 3% to 5% fluctuations in DSC scores between the proposed method and method of generating only mask. In particular, owing to small shape, knee cartilage fluctuated almost 5% in every scenario. Furthermore, the network becomes stabilized after 50% of data is used for training. Notably, stabilizing GANs with small datasets is challenging. Overall, we can conclude that the proposed method performs better compared to both traditional GANs and “Only Mask” methods.

### B. ABLATION STUDY

Besides, the proposed decimated sample method, two more methods are evaluated to evaluate the performance of various extraction methods. Table 6 presents the variation in DSC scores based on different extracting methods. Random, Proposed, and GANs(2D) represent various methods for extracting pixels from images. The random method retrained random pixels such as 25%, 50%, 75%, and 100% from images. However, pixel extraction directly from 2D images named GANs(2D) method. This method leaves (3 × 3), (2 × 2), (1 × 1) and (0 × 0) blocks for 25%, 50%, 75% and 100% pixels, respectively, before performing sample extraction. Results show that the proposed method performs better than the remaining two methods (Random and GANs(2D)). The proposed method extracts more relevant and deterministic pixels compared to the other two methods. However, whereas random may generate more intermediate images, there is no assurance that those are sufficiently relevant and deterministic to segment images for organs. Furthermore, our architecture is designed to handle smaller datasets, and as such, unlimited intermediary images cannot be generated using our method. This table (similar to Table 5) shows, proposed architecture generates promising results with 25% of decimated samples. Due to the increasing number of decimated samples, parameters increase exponentially, which affects the performance of architecture. This gradual digression of performance with an increasing number of decimated samples faces problems caused by over-fitting. As architecture is small, it performs better for fewer parameters. However, this architecture shows a similar pattern across different datasets, which defines the generalization capability of proposed methods.

### C. PERSPECTIVE OF FUTURE WORK

This segmentation method can be applied to identify different anatomical structures through the human body, such as bones, blood vessels, vertebrae, and major organs using...
small architecture and dataset. However, this method could extend dimensionality from 2D to 3D operations in the future. 3D operations will allow implementing diverse and new techniques such as V-net to improve performance, which is contradictory to the objective of the current version of the proposed method. Furthermore, various decimated samples could be used for data argumentation as a single type of pixel extraction is used for 25%, 50%, and 75% decimated samples, in the current experiment. It can extract pixels from various positions to use as a guide for different decimated samples.

V. CONCLUSION

Recently, deep learning has made a colossal impact in various image processing fields. In particular, GANs-based learning has significantly improved medical image segmentation over other deep-learning-based methods and attracted the attention of medical image researchers worldwide. Conventional GANs perform segmentation using the complete image and generate a segmentation mask. The complete image requires a larger dataset and deeper architecture to ensure network stability. In this study, a novel segmentation method using GGANs proposes to enable data-efficient learning by extracting decimated samples. This method evaluated five different segmentation problems of medical datasets, which is comparable to other state-of-art methods. However, using decimated sample traits, this proposed method able to achieve most promising DSC scores after using only 50% of datasets, whereas traditional GANs-based method requires entire dataset to achieve a similar level of performance. The proposed approach also yields a better DSC score compared to the approach in which (only) segmentation mask is generated, as this enhances error between generator and discriminator during the generation of original images and corresponding segmentation mask, thereby rendering errors non-negligible for architecture. Therefore, the proposed method showed better performance than traditional GANs. Notably, the testing procedure takes less than a second to test each slice. Consequently, this can be considered to implement for real-time applications like diagnosing diseases of patients in a short period during an emergency. The proposed approach turns segmentation into an artificial visualization problem, thereby facilitating a better understanding of images.

REFERENCES

[1] Results Knee Segmnetation. Accessed: Dec. 30, 2017. [Online]. Available: http://www.sk10.org/results.php

[2] Results Leftratium-Brain tumour. Accessed: Dec. 30, 2017. [Online]. Available: http://medicaldecathlon.com/results.html

[3] G. Antipov, M. Baccouche, and J.-L. Dugelay, “Face aging with conditional generative adversarial networks,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2017, pp. 2089–2093.

[4] M. A. Baldwin, J. E. Langenderfer, P. J. Rullkoetter, and P. J. Laz, “Development of subject-specific and statistical shape models of the knee using an efficient segmentation and mesh-morphing approach,” Comput. Methods Programs Biomed., vol. 97, no. 3, pp. 232–240, Mar. 2010.

[5] L. Breiman, “Bagging predictors,” Mach. Learn., vol. 24, no. 2, pp. 123–140, Aug. 1996.

[6] A. A. Daquette, P.-M. Jodoin, O. Bouchot, and A. Lalande, “3D segmentation of abdominal aorta from CT-scan and MR images,” Computerized Med. Imag. Graph., vol. 36, no. 4, pp. 294–303, Jun. 2012.

[7] F. Eckstein, F. Ciclitira, J.-P. Raynauld, J. C. Waterton, and C. Peterfy, “Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): Morphological assessment,” Osteoarthritis Cartilage, vol. 14, no. 1, pp. 46–75, 2006.

[8] M. A. Elattar, E. M. Wiegerinck, R. N. Planken, E. Vanvahlen, H. C. van Assen, J. Baan, and H. A. Marquering, “Automatic segmentation of the aortic root in CT angiography of candidate patients for transcatheter aortic valve implantation,” Med. Biol. Eng. Comput., vol. 52, no. 7, pp. 611–618, Jul. 2014.

[9] J. Folkesson, E. B. Dam, O. F. Olsen, M. A. Karsdal, P. C. Pettersen, and C. Christiansen, “Automatic quantification of local and global articular cartilage surface curvature: Biomarkers for osteoarthritis?” Magn. Reson. Med., vol. 59, no. 6, pp. 1340–1346, 2008.

[10] Centers for Disease Control and Prevention et al., “Prevalence of disabilities and associated health conditions among adults—United States, 1999,” MMWR. Morbidity Mortality Weekly Rep., vol. 50, no. 7, p. 120, 2001.

[11] J. Fripp, S. Crozier, S. K. Warfield, and S. Ourselin, “Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee,” IEEE Trans. Med. Imag., vol. 29, no. 1, pp. 55–64, Jan. 2010.

[12] A. Gandhamal, S. Talbar, S. Gajre, A. F. M. Hani, and D. Kumar, “Local guided level S-curve transformation—A generalized contrast enhancement technique for medical images,” Comput. Biol. Med., vol. 83, pp. 120–133, Apr. 2017.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.

[14] Z. Han, B. Wei, A. Mercado, S. Leung, and S. Li, “Spine-GAN: Semantic segmentation of multiple spinal structures,” Med. Image Anal., vol. 50, pp. 23–35, Dec. 2018.

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. pattern Recognit., Jun. 2016, pp. 770–778.

[16] Y. Huo, Z. Xu, S. Bao, C. Bermudez, A. J. Plassard, J. Liu, Y. Yao, A. Assad, R. G. Abramson, and B. A. Landman, “Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks,” Proc. SPIE, vol. 10574, Mar. 2018, Art. no. 1057409.

[17] T. Iqbal and H. Ali, “Generative adversarial network for medical images (MI-GAN),” J. Med. Syst., vol. 42, no. 11, p. 231, Nov. 2018.

[18] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. van Ginneken, “Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans,” IEEE Trans. Med. Imag., vol. 28, no. 7, pp. 1000–1010, Jul. 2009.

[19] S. Izadi, Z. Mirikharaji, J. Kawahara, and G. Hamarneh, “Generative adversarial networks to segment skin lesions,” in Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI), Apr. 2018, pp. 881–884.

[20] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in Proc. 14th Eur. Conf. Comput. Vis. (ECCV), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 694–711.

[21] U. Kurkure, O. C. Avila-Montes, and I. A. Kakadiaris, “Automated segmentation of thoracic aorta in non-contrast CT images,” in Proc. 5th Int. Symp. Biomed. Imaging. From Nano Macro, May 2008, pp. 29–32.

[22] S. Lee, S. H. Park, H. Shin, I. D. Yun, and S. U. Lee, “Optimization of local shape and appearance probabilities for segmentation of knee cartilage in 3-D MR images,” Comput. Vis. Image Understand., vol. 115, no. 12, pp. 1710–1720, Dec. 2011.

[23] C. Li and M. Wand, “Precomputed real-time texture synthesis with Markovian generative adversarial networks,” in Proc. 14th Eur. Conf. Comput. Vis. (ECCV), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 702–716.

[24] Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, and Z. Zeng, “CLUCNNs: Object detection for medical images,” Neurocomputing, vol. 350, pp. 53–59, Jul. 2019.

[25] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2015, arXiv:1511.06434. [Online]. Available: http://arxiv.org/abs/1511.06434

[26] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. 10th Int. Conf. Med. Image Comput. Comput-Assist. Intervent. (MICCAI), N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Munich, Germany: Springer, Oct. 2015, pp. 234–241.
[27] J. Son, S. Jun Park, and K.-H. Jung, “Retinal vessel segmentation in fundoscopic images with generative adversarial networks,” 2017, arXiv:1706.09318. [Online]. Available: http://arxiv.org/abs/1706.09318

[28] C. Tobon-Gomez et al., “Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets,” IEEE Trans. Med. Imag., vol. 34, no. 7, pp. 1460–1473, Jul. 2015.

[29] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, “SegAN: Adversarial network with multi-scale $L_1$ loss for medical image segmentation,” Neuroinformatics, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018.

[30] J. Zhang, Y. Xia, Y. Xie, M. Fulham, and D. D. Feng, “Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features,” IEEE J. Biomed. Health Inform., vol. 22, no. 5, pp. 1521–1530, Sep. 2018.

[31] Y. Zhang, L. Yang, J. Chen, M. Frederiksen, D. P. Hughes, and D. Z. Chen, “Deep adversarial networks for biomedical image segmentation utilizing unannotated images,” in Proc. 20th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds. Quebec City, QC, Canada: Springer, Sep. 2017, pp. 408–416.

[32] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, “Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features,” IEEE Trans. Med. Imag., vol. 27, no. 11, pp. 1668–1681, Nov. 2008.

[33] Y. Zheng, M. John, R. Liao, A. Nottling, J. Boese, J. Kempfert, T. Walther, G. Brockmann, and D. Comaniciu, “Automatic aorta segmentation and valve landmark detection in C-arm CT for transcatheter aortic valve implantation,” IEEE Trans. Med. Imag., vol. 31, no. 12, pp. 2307–2321, Dec. 2012.

[34] X. Zhuang, K. S. Rhode, R. S. Razavi, D. J. Hawkes, and S. Ourselin, “A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI,” IEEE Trans. Med. Imag., vol. 29, no. 9, pp. 1612–1625, Sep. 2010.

[35] M. A. Zuluaga, M. J. Cardoso, M. Modat, and S. Ourselin, “Multi-atlas propagation whole heart segmentation from mri and cta using a local normalised correlation coefficient criterion,” in Proc. 7th Int. Conf. Funct. Imag. Modeling Heart (FIMH), S. Ourselin, D. Rueckert, and N. Smith, Eds. London, U.K.: Springer, Jun. 2013, pp. 174–181.

HOSNA ASMA-ULL received the B.S. degree from Sookmyung Women’s University, Seoul, in 2011, and the M.S. degree from the Hankuk University of Foreign Studies, Yongin, South Korea, in 2017, where she is currently pursuing the Ph.D. degree with the Department of Computer and Electronic Systems Engineering. She was working as a Faculty Member with the Department of Computer Science and Engineering, Stamford University Bangladesh, Dhaka, Bangladesh, from 2012 to 2019, where she is also an Assistant Professor. Her current research interests include deep learning, medical image analysis, artificial intelligence, and computer vision.

IL DONG YUN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1989, 1991, and 1997, respectively. From 1996 to 1997, he was a Senior Engineer with Daewoo Electronics, Seoul. Since 1997, he has been with the Hankuk University of Foreign Studies, Yongin, South Korea. His research interests include medical image analysis and computer vision.

DONGJIN HAN received the B.S. degree from Seoul National University, Seoul, South Korea, in 1990, and the M.Sc. and Ph.D. degrees in E.E. from Brown University, Providence, USA, in 1995 and 2009, respectively. He was a Researcher with the Samsung Advanced Institute of Technology (SAIT), Kihung, South Korea, from 1994 to 1995. From 1996 to 1999, he was a Lecturer with the Korea Military Academy. In 2009, he was with the Intelligent Robot Research Center, Soongsil University, Seoul. From 2012 to 2016, he was a Research Professor with the Asan Medical Center and Severance Hospital, Seoul. He is currently an Assistant Professor with Cheju-Halla University, Jeju City, South Korea. His research interests are in object recognition, multi view geometry, 3D reconstruction, virtual heritage protection, and stochastic processes.