2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Attention-Based Multi-Scale Generative Adversarial Network for synthesizing contrast-enhanced MRI

Meiqing Pan, Hui Zhang*, Zhenchao Tang, Yinghua Zhao, Jie Tian*

***Abstract*—In clinical practice, about 35% of MRI scans are enhanced with Gadolinium - based contrast agents (GBCAs) worldwide currently. Injecting GBCAs can make the lesions much more visible on contrast-enhanced scans. However, the injection of GBCAs is high-risk, time-consuming, and expensive. Utilizing a generative model such as an adversarial network (GAN) to synthesize the contrast-enhanced MRI without injection of GBCAs becomes a very promising alternative method. Due to the different features of the lesions in contrast-enhanced images while the single-scale feature extraction capabilities of the traditional GAN, we propose a new generative model that a multi-scale strategy is used in the GAN to extract different scale features of the lesions. Moreover, an attention mechanism is also added in our model to learn important features automatically from all scales for better feature aggregation. We name our proposed network with an attention-based multi-scale contrasted-enhanced-image generative adversarial network (AMCGAN). We examine our proposed AMCGAN on a private dataset from 382 ankylosing spondylitis subjects. The result shows our proposed network can achieve state-of-the-art in both visual evaluations and quantitative evaluations than traditional adversarial training.**

***Clinical Relevance*— This study provides a safe, convenient, and inexpensive tool for the clinical practices to get contrast-enhanced MRI without injection of GBCAs.**

I. Introduction

About 35% MRI scans are enhanced with Gadolinium - based contrast agents (GBCAs) worldwide currently. However, GBCAs is high-risk, time-consuming, and expensive. Some researchers showed that the injection of GBCAs would make gadolinium deposition within the human bone and brain tissue even patients have normal kidney function. As for patients who have compromised kidney function, GBCAs may result nephrogenic systemic fibrosis [1]. Also, the injection of the contrast agent itself costs additional money and time.

Generative adversarial network (GAN) has been widely used in medical image fields such as reconstruction, segmentation and detection [2]. If we can use GAN as an image generation tool to produce a contrast-enhanced MRI from the non-contrast-enhanced MRI, the potential harm caused by the injection of GBCAs would be avoided while the needs of clinical examination can be satisfied at the same time.

However, there are some challenges for the generation of contrast-enhanced MRI. The size and shape of lesions in contrast-enhanced MRI are different and the anatomical structure around the lesion is complex which makes the generation of meaningful contrast-enhanced MRI more difficult. Large receptive fields with rich semantic information are helpful to locate the lesions and diminish the impact of the cluttered background, but the geometric details are lost due to the reduction of resolution. On the contrary, small fields facilitate the detail generation of lesions such as the enhancement level and boundary information, but the high resolution makes it lacks semantic information. Traditional GAN is comprised of convolutional neural networks (CNN) which extract features layer by layer. The shallow features of CNN have small receptive fields, with the network going deeper, the receptive field becomes larger gradually. For lesions, the fine-scale features are lost heavily as the network deepens. The performance of the traditional GAN may degrade because of the characteristic of contrast-enhanced MRI. Given that the lesions change greatly in size and shape which corresponding to different scale features, we propose a multi-scale-based generator to synthesize contrast-enhanced MRI. The multi-scale idea for feature extraction is widely employed in convolutional neural networks especially for the task of detection and segmentation, but there is no report for utilizing contrast-enhanced MRI generation. After extracting the features from all scales, an attention mechanism is applied to aggregate them effectively by weighting all features according to their weights.

*Corresponding Authors: Hui Zhang, Jie Tian.

Research supported by the National Natural Science Foundation of China under Grant No. 81871511, 62027901, 81930053.

Meiqing Pan is with Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing; Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing (e-mail: pmq1835@buaa.edu.cn).

Hui Zhang is is with Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing; Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing (e-mail: hui.zhang@buaa.edu.cn, phone: 0086-010-82618465).

Zhenchao Tang is with Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing; Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing (e-mail: tangzhenchao@buaa.edu.cn).

Yinghua Zhao is with Department of Radiology, The Third Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong (e-mail: zyh7258957@163.com).

Jie Tian is with Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing; Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing (e-mail: tian@ieee.org, phone: 86-010-82628760).

Taken together, in our current study, we propose a AMCGAN model that can generate contrast-enhanced MRI without the injection of GBCAs, which provides a safe, convenient, and inexpensive tool for the clinical practices. Our proposed AMCGAN is powerful in extracting both coarse features and fine features to synthesize fine-grained lesions. The proposed AMCGAN also add a feature attention mechanism to enhances the feature representation learning of contrast-enhanced-related features and suppresses the expression of irrelevant features for feature aggregation.

II. Related work

A. Identifying lesions on non-contrast enhanced images

Many active lesions show enhancement on contrast-enhanced imaging after the injection of GBCAs. But given the disadvantages of GBCAs mentioned previously, some alternatives have been introduced to recognize enhanced lesions without the injection of GBCAs. For example, Michoux et al used texture parameters from T2-weighted MRI to assess brain inflammatory activity to replace contrast-enhanced T1-weighted images [3]. Shinohara et al adopted logistic regression to model the enhancement probability of each voxel on MRI without the GBCAs injection [4]. Deep learning is also used to learn features of enhancing lesions more conveniently. Researches showed the deep learning can identify lesions on images at reduced GBCAs dose and the prediction accuracy can achieve 75% even on non-GBCAs images [5].

B. Generating non-contrast enhanced images by GAN

There are many applications of GAN in the medical image. For example, Wolterink used a basic pix2pix framework for denoising [6], calimeri et al used a modified GAN adopted to generate MRI slices of the human brain for data augmentation [7]. Besides that, GAN also has good performance in tasks of image segmentation, detection, classification, and so on. For the issue of synthesizing contrast-enhanced images, Zhao et al introduced a Tripartite-GAN to generate liver contrast-enhanced MRI from non-contrast enhanced MRI and then used it for tumor detecting [8]. The lesions in this study have obvious area and regular shape, but the lesions of ankylosing spondylitis have diverse size and shape.

C. Multi-scale feature capture

Multi-scale feature extraction is widely used in CNN especially for the task of detection and segmentation. Models such as HyperNet concatenate low-level and high-level features from different layers to improve the detection effect of the small object [9]. PPM and ASPP introduce different pooling scales or dilated convolutions to extract both the local and global information simultaneously which improves the segmentation output [10]. DeepLabv3+ uses skip connections between the encoding path and the decoding path to enhance the detail boundary information and yield a more precise segmentation effect [11].

D. Attention mechanism

The attention mechanism is a remarkable method to focus on the target region rather than the whole image or sequence. Since Bahdanau et al used a mechanism similar with attention to translating, various attention mechanisms incorporated into deep learning networks have been widely researched. J. Fu et al adopted a dual attention module to stress effective spatial feature representations and reinforce special semantics in channels [12]. F. Wang et al proposed residual attention modules with two branches, in the attention module, each trunk branch has its mask branch to get its specialized features by the mechanism of attention [13]. Inspired by these methods, the attention module is adopted to learn important features automatically from all scales for better feature aggregation.

III. Methods

A. Attention-Based Multi-Scale Generative Adversarial Network

For the effective generation of contrast-enhanced MRI, our contrast-enhanced generative adversarial networks (AMCGAN) executed the competition between two participants: the novel attention-based multi-scale generator and the CNN-based discriminator. Fig.1 displays the structure of our designed AMCGAN. Specifically, the attention-based multi-scale generator is composed of three main parts: multi-scale feature capture modules (MSFC) made up of four parallel convolution layers, feature attention mechanism (FAM) to exploits the relationship among multi-scale features, and an encoder-decoder structure based on Pix2Pix [14]. The process of encoding consists of three convolution blocks which involved the operation of convolution, batch-normalization(BN), and ReLU. After encoding, there are six Resnet blocks constructed by two paths. In the decoding process, the convolution layer is replaced by the deconvolution layer. The dropout layer is used in the Resnet block to reduce overfitting.

Figure 1. The architecture of AMCGAN. The generator consists of three parts: multi-scale feature capture (MSFC), feature attentive module (FAM), an encoder-decoder structure based on pix2pix. The discriminator consists of a series of convolutions, instance-normalization, and Leaky ReLU layers.

The discriminator is built by six convolution layers. First five convolution layers are followed by the process of instance-normalization (IN) and LeakyReLU. The network is trained as a gap measurement of real and fake images, the gap is backward to the generator to help synthesize more realistic images by minimizing the gap.

B. Multi-Scale Feature Capture

The lesions of contrast-enhanced MRI don't have uniform size and shape which corresponding to different scale features. Therefore, a multi-scale feature capture scheme (MSFC) with different convolutional kernels sizes is designed to extract different scale features of lesions. The detail of the scheme is represented in Fig.2. The MSFC consists of four parallel sets of convolution layer to get feature f^1, f^2, f^3, f^4 respectively. Generally, in the CNN structure, the operation of pooling is

used to enlarge the receptive field, but it may cause the reduction of resolution and the image details may be lost. The operation of dilated convolutions can capture multi-scale features without losing resolution. So, a dilated convolution is also applied in the parallel feature extraction structure. All features from the parallel convolutions are concatenated into the feature attention module. The multi-scale module enables the network to extract both rough features such as the location and shape of lesions and fine features such as enhancement level and boundary information of lesions.

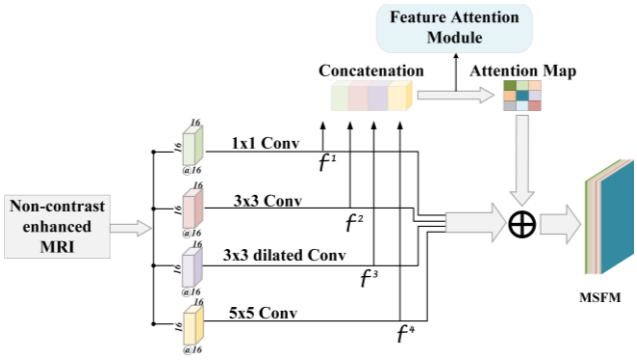

Figure 2. The architecture of MSFC. The MSFC consists of four parallel sets of convolutional kernels with four scales. The features are concatenated into the attention module to get final multi-scale feature maps (MSFM).

C. Attention weighted feature fusion

After getting features from different scales through the multi-scale feature capture module, we further conduct the weight of each feature. More specifically, we introduce a feature attention module into the generator which can calculate the interdependencies between feature maps. The feature attention module will allocate more weight to task-related features and ignore non-related features to make the aggregation of features more effective. The detail of the feature attention mechanism is presented in Fig.3.

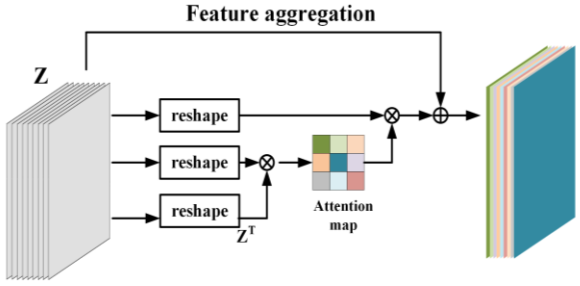

Figure 3. The architecture of feature attention module.

In the beginning, the multi-scale features got by the parallel four convolutions are concatenated as the input of the feature attention module. The concatenated features $Z \in R^{C\times H\times W}$ are reshaped to $Z \in R^{C\times N}$. C is the number of feature maps, H and W are the length and width of the feature map and $N=H \times W$. A matrix multiplication between Z and the transpose of Z is applied. After that, a Softmax layer is taken to obtain the feature attention map $X \in R^{C\times C}$:

$$X_{ij}=\frac{Z_i \cdot Z_j}{\sum_{i=1}^{C} \exp(Z_i \cdot Z_j)} \quad (1)$$

Where X_{ij} measures the ith feature influence on the jth feature. Secondly, matrix multiplication is taken between the transpose of Z and attention map and reshape their result $R^{C\times H\times W}$. Thirdly, a scale parameter β is multiplied to the result and an element-wise sum operation with Z is performed to obtain the output A_j:

$$A_j=\beta \sum_{i=1}^{C}(X_{ij}Z_i)+Z_j \quad (2)$$

D. Implementation Details

The experiment was realized by using python and PyTorch. In the training phase, we set a batch size of 2 and the initial learning rate for Adam optimizer is 0.0002 in the first 300 epochs. The learning rate decays linearly in the last 100 epochs. The peak signal to noise rate (PSNR) value is used as the index to select the best result on training and stored the weights for testing. The objective of AMCGAN is shown in equation 3. G and D are the generator and discriminator of the network.

$$G^{*}=\arg\min\max L_{cGAN}(G,D)+\lambda L_{L1}(G) \quad (3)$$

l_{cGAN} is a conditional GAN loss to map non-contrast enhanced MRI to contrast-enhanced MRI. l_{L1} is an L1 distance to improve the PSNR. Equations 4 and 5 show the loss function of l_{cGAN} and l_{L1}.

$$L_{cGAN}(G,D)=E_{x,y}[\log D(x,y)]+E_{x,z}\left[\log\left(1-D\left(x,G(x,z)\right)\right)\right] \quad (4)$$

$$l_{L1}(G)=E_{x,y,z}[\left|\left|y-G(x,z)\right|\right|_1] \quad (5)$$

IV. Experiments

A. dataset and pre-processing

The dataset was collected from The Third Affiliated Hospital of Southern Medical University including 382 patients diagnosed with ankylosing spondylitis. Active inflammatory lesions of ankylosing spondylitis can show hyperintense signal in contrast-enhanced MRI which is different with chronic inflammatory lesions. The stronger the hyperintense signal the more likely it reflects active inflammation. Each patient had non-contrast enhanced MRIs and contrast-enhanced MRIs. The data are divided into a training set and testing set randomly by patients, there are 5212 slices in the training set and 1628 slices in the testing set.

B. Evaluation and visualization

The performance of our proposed model to synthesis enhanced images is evaluated by the mean absolute error (MAE) and PSNR. Since doctors diagnose the nature of tumors or stage tumors by observing whether the tumor area is enhanced, we only calculate the PSNR value of the slice containing the tumor. Taking Pix2Pix as the baseline, compare it to the network with a multi-scale module and our AMCGAN. The objective image quality evaluation results and the visualization result are shown in Table 1 and Fig.4 respectively.

TABLE I. The Result of Comparison Analysis

Method	*PSNR*	*MSE*
Pix2Pix	25.39±2.31	36.84±9.56
Multi-scale	25.46±2.24	36.97±9.83
AMCGAN	26.29±2.19	34.64±9.20

As shown in Fig.4, the fake MRI generated by our proposed AMCGAN has no obvious visual difference with ground truth which is obtained by GBCAs injection. More importantly, AMCGAN generates a more clearly contrast-enhanced area of lesions than the other two models. The comparison analysis shows the multi-scale module and feature attention mechanism can help the network reach better performance in both visual evaluations and quantitative evaluations. Also, the heatmaps of real contrast-enhanced MRI and generated contrast-enhanced MRI in Fig.5 further demonstrate that the AMCGAN pays attention to an area of the contrast-enhanced lesion in both large and small lesions when generation.

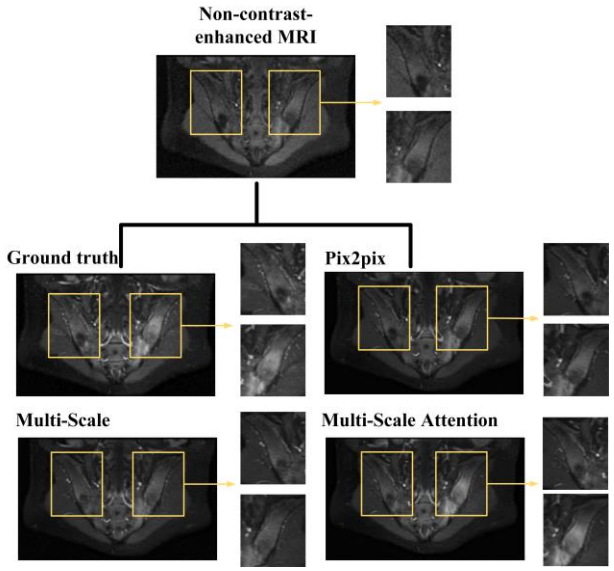

Figure 4. Examples of visual comparison results. The yellow windows of zoomed local patches represent the lesion area. The PSNR value of three generated pictures (Pix2Pix, Multi-Scale, Multi-Scale Attention) are 26.43, 26.91 and 28.45 respectively.

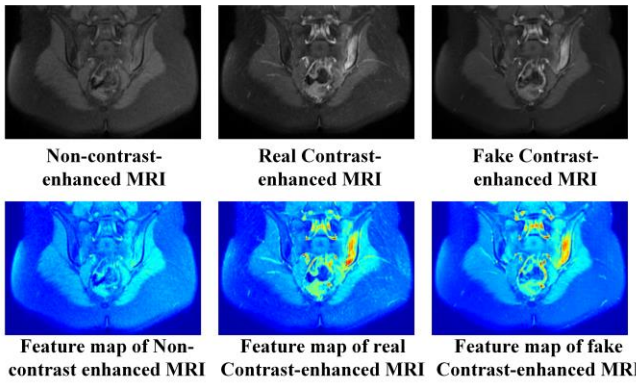

Figure 5. One example of contrast-enhanced MRI generation. The red window in the feature map shows the generation difference of lesion area between non-contrasted enhanced MRI, real contrast-enhanced MRI and fake contrast-enhanced MRI.

V. Conclusion

In this study, we propose an attention-based multi-scale generative adversarial network to synthesize contrast-enhanced MRI without the injection of GBCAs. Based on the structure of Pix2Pix, we add a multi-scale feature module to extract both coarse features and fine features of lesions. Moreover, a feature attention module is used to enhances the expression of important features. Our designed network successfully synthesizes higher quality contrast-enhanced images on one private dataset of 382 subjects than traditional adversarial training. To further illustrate the clinical usefulness of synthesized images, in the future work, we will include the sensitivity and specificity of radiologists in identifying lesions on the generated contrast-enhanced MRI.

References

[1] T. J. Fraum, D. R. Ludwig, M. R. Bashir, and K. J. Fowler, "Gadolinium-based contrast agents: A comprehensive risk assessment," J Magn Reson Imaging, vol. 46, no. 2, pp. 338-353, Aug 2017.

[2] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," Med Image Anal, vol. 58, p. 101552, Dec 2019.

[3] N. Michoux, A. Guillet, D. Rommel, G. Mazzamuto, C. Sindic, and T. Duprez, "Texture analysis of T2-weighted MR images to assess acute inflammation in brain MS lesions," PLoS One, vol. 10, no. 12, p. e0145497, 2015.

[4] R. T. Shinohara, J. Goldsmith, F. J. Mateen, C. Crainiceanu, and D. S. Reich, "Predicting breakdown of the blood-brain barrier in multiple sclerosis without contrast agents," American Journal of Neuroradiology, vol. 33, no. 8, pp. 1586-1590, 2012.

[5] E. Gong, J. M. Pauly, M. Wintermark, and G. Zaharchuk, "Deep learning enables reduced gadolinium dose for contrast-enhanced brain sMRI," J Magn Reson Imaging, vol. 48, no. 2, pp. 330-340, Aug 2018.

[6] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative Adversarial Networks for Noise Reduction in Low-Dose CT," IEEE Transactions on Medical Imaging, vol. 36, no. 12, pp. 2536-2545, 2017.

[7] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in International conference on artificial neural networks, 2017, pp. 626-634: Springer.

[8] J. Zhao et al., "Tripartite-GAN: Synthesizing liver contrast-enhanced MRI to improve tumor detection," Med Image Anal, vol. 63, p. 101667, Jul 2020.

[9] T. Kong, A. Yao, Y. Chen, and F. Sun, "Hypernet: Towards accurate region proposal generation and joint object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 845-853.

[10] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881-2890.

[11] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801-818.

[12] J. Fu et al., "Dual attention network for scene segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146-3154.

[13] F. Wang et al., "Residual attention network for image classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3156-3164.

[14] P. Isola, J. Y. Zhu, T. H. Zhou, A. A. Efros, and Ieee, "Image-to-Image Translation with Conditional Adversarial Networks," in 30th Ieee Conference on Computer Vision and Pattern Recognition(IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5967-5976.