2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

3D CGAN BASED CROSS-MODALITY MR IMAGE SYNTHESIS FOR BRAIN TUMOR SEGMENTATION

Biting Yu[⋆] *Luping Zhou*[⋆,◦] *Lei Wang*[⋆] *Jurgen Fripp*[†] *Pierrick Bourgeat*[†]

[⋆] University of Wollongong, Australia [◦] University of Sydney, Australia [†] CSIRO, Australia

ABSTRACT

Different modalities of magnetic resonance imaging (MRI) can indicate tumor-induced tissue changes from different perspectives, thus benefit brain tumor segmentation when they are considered together. Meanwhile, it is always interesting to examine the diagnosis potential from single modality, considering the cost of acquiring multi-modality images. Clinically, T1-weighted MRI is the most commonly used MR imaging modality, although it may not be the best option for contouring brain tumor. In this paper, we investigate whether synthesizing FLAIR images from T1 could help improve brain tumor segmentation from the single modality of T1. This is achieved by designing a 3D conditional Generative Adversarial Network (cGAN) for FLAIR image synthesis and a local adaptive fusion method to better depict the details of the synthesized FLAIR images. The proposed method can effectively handle the segmentation task of brain tumors that vary in appearance, size and location across samples.

***Index Terms*—** Image synthesis, brain tumor segmentation, Generative Adversarial Network, local adaptive fusion

1. INTRODUCTION

Gliomas as the most frequent brain tumor in adults has threatened large population [1]. To make effective treatment, brain tumor segmentation on medical images is crucial for the quantitative analysis of tumor's appearance, size and location. However, the intensity of tumor parts and surrounding brain tissues change gradually, which could lead to ambiguous tumor outlines in medical images. To mitigate this issue, changing the imaging parameters of Magnetic resonance imaging (MRI) could be an effective option. For example, T1-weighted (T1w) MR images gives sharp contrast between white and grey matter tissues, while Fluid attenuated inversion recovery (FLAIR) highlights lesion tissues [2]. Meanwhile, it is always interesting to examine the diagnosis potential from single modality, considering the cost of acquiring multi-modality images. Since T1 is the most commonly used imaging modality, this paper investigates whether we can synthesize FLAIR images from T1 to improve T1-only based brain tumor segmentation.

Corresponding Author: Dr. Luping Zhou

Image synthesis, which transfers the given images to another specific type, has shown wide applications in medical field [3, 4]. Existing works can be roughly classified into two categories. The first one is registration based methods that conduct the synthesis through the atlas registration of the involved image types [5,6]. However, the unified atlases usually built upon healthy persons may cause significant deformation when abnormal tissues exist. The second category, learning based methods, can mitigate this problem. When the training samples contain pathology, this information can be effectively captured by learning based methods. In this way, abnormal tissue can be synthesized. Some machine learning methods, like dictionary learning [7] and deep convolutional neural network (CNN) [8], have been proposed for this purpose.

Recently, conditional Generative Adversarial Network (cGAN) [9] has demonstrated itself to be a promising method for image synthesis. Different from traditional learning based methods, cGAN consists of two modules, a generator to learn the mapping for realistic images and a discriminator to distinguish the real and the synthesized images. By training the two modules to beat each other, cGAN has achieved excellent performance on medical image synthesis, as shown in [10, 11]. Nevertheless, these methods synthesize each individual slice independently along the axial direction and then concatenate them into a 3D image. This results in discontinuous estimation along the coronal and the sagittal directions. However, image information along these two directions are also crucial for medical image synthesis and analysis. Although some recent work [12] applies 3D cGAN to predict small patches to eliminate the discontinuity caused by 2D estimation, learning on small patches is insufficient to extract both local and global contextual relationship among voxels, and could hurt the synthesis of brain tumor images, as shown in Sec. 3.

In this paper, we propose a FLAIR synthesis method to improve brain tumor segmentation from T1. Our method consists of both a global non-linear mapping and a local linear mapping from T1 to FLAIR. The global non-linear mapping determines the similarity of synthesized images to FLAIR at the whole image level, which is estimated by our proposed 3D cGAN model. The local linear mapping further improves the local details from T1 for tumor segmentation, which is sought locally from the real FLAIR images of the training set and is subject-adaptive.

978-1-5386-3636-7/18/$31.00 ©2018 IEEE

The proposed method has three novel contributions. First, we propose a 3D cGAN model for FLAIR image synthesis. It mitigates th35e problem of discontinuous estimation across slices caused by the 2D cGAN in the current literature. By considering large image patches and hierarchical features from skip connections, our 3D cGAN model could better synthesize FLAIR images by taking contextual information into account. Second, to further improve the synthesized FLAIR images for the segmentation task, a local adaptive synthesis method is proposed, which better depicts the local details of the synthesized FLAIR images. Third, the synthesized FLAIR images are utilized to help brain tumor segmentation from the single T1 modality via training a CNN that considers two imaging modalities jointly. The effectiveness of the proposed method is demonstrated on the public dataset 2015 Brain Tumor Segmentation Challenge (BRATS) [1].

2. PROPOSED METHODS

2.1. Overview

Our proposed cross-modality MR image synthesis framework generates FLAIR-like images from T1 by a 3D cGAN model and a following local adaptive fusion to cater for the similarity at both whole image and local patch levels. The final synthesized images, together with T1 images, are processed by a two-pathway 3D CNN model to segment brain tumor.

2.2. 3D cGAN

2.2.1. Basic ideas

The original GAN [13] consists of two modules, the generator G and the discriminator D, struggling with each other to synthesize images $G(z)$ resembling real images x from the random vector z and meanwhile distinguish x from $G(z)$. cGAN has extended the original GAN to capture auxiliary information y in both the generator G and the discriminator D. For our cross-modality image synthesis, y is the T1 MR images that are the input to synthesize FLAIR-like images $G(y)$. The real pair (y, x) and the estimated pair $(y, G(y))$ are differentiated by the discriminator D. The generator G and the discriminator D are trained simultaneously, as if they are following a two-player min-max game with the following objective:

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{y,x\sim p_{data}(y,x)}[\log D(y, x)] + \mathbb{E}_{y\sim p_{data}(y)}[\log(1 - D(y, G(y)))], \quad (1)$$

where $G(\cdot)$ and $D(\cdot)$ denote the outputs of the generator and the discriminator, respectively.

Furthermore, to ensure the voxel-wise similarity between the synthesized and the real images, an L1-norm penalty [14] is also utilized and formulated as follows:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{y,x\sim p_{data}(y,x)}[\|x - G(y)\|_1]. \quad (2)$$

Combining cGAN objective and L1 loss, the final objective function is formulated as:

$$\mathcal{L}_{total} = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \quad (3)$$

where λ is a hyperparameter to balance the two terms.

2.2.2. Detailed architecture

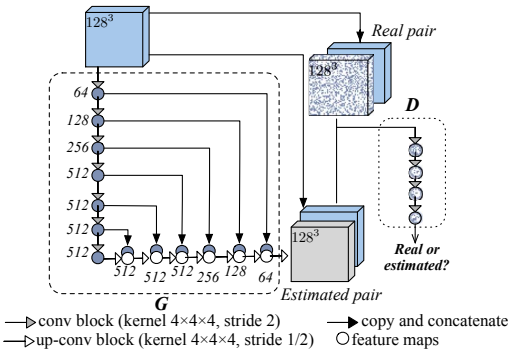

Fig. 1: *Our proposed 3D cGAN consists of a generator G and a discriminator D. All conv and up-conv blocks contain convolutional, batch normalization and ReLU layers. Dropout with 50% is applied in the first three up-conv blocks of G. Batch normalizaiton is not used in the first conv block of G and D. ReLUs in the conv blocks are leaky, with slope 0.2.*

In the literature, U-net has been developed to conduct analysis on whole images or large image patches rather than small ones to guarantee spatial contiguity in the output [15]. It contains contracting and expanding paths with skip connections between them to extract hierarchical features. The skip connections also help reduce the influence from gradient vanishing in deep network, and contracting and expanding paths make it feasible to capture multi-depth information. In this work, to benefit from U-net, we exploit its structure and design a 3D U-net like generator in our 3D cGAN. Specifically, our generator includes seven down-convolutional blocks in the contracting path and also seven up-convolutional blocks in the expanding path. Coupled with a five-convolutional-layer 3D discriminator (including Sigmoid function), it constitutes our proposed 3D cGAN for FLAIR image synthesis from T1. The detailed architecture of the proposed 3D cGAN is illustrated in Fig. 1. Please note that, large patches (128^3), rather than a whole image ($240\times240\times155$) are used as the input of our model to deal with the limited number of training samples, as well as controlling the number of parameters to learn.

2.3. Subject-specific local adaptive fusion

Our ultimate goal is to synthesize the FLAIR images that could help boost the segmentation performance of brain tumor. This puts forward higher requirements on the quality of the output images compared with those synthesis methods merely focusing on improving PSNR (peak signal-to-noise) in the literature. Our task is more challenging due to two factors. First, the pathology involved in T1 MR images significantly increases the difficulty of the synthesis task, since brain tumor varies in appearance, size and location. This is in contrast to the images synthesis for healthy subjects commonly seen in the literature. Second, synthesizing FLAIR images only from T1 has limits, since T1 seems lack of some information observed in FLAIR, such as the diffuse changes

around tumor regions, which may adversely affect the segmentation results. Therefore, in addition to our 3D cGAN, we propose to further polish the local details of our synthesized images through linearly combining the *real* FLAIR images of the training set for approximation, and the combination weights are estimated from the FLAIR-like images output by our 3D cGAN model. This approach is feasible because our FLAIR-like images resemble the real ones so that their combination weights are highly correlated. Our method is both local and adaptive. "Local" means the combination weights vary with different locations in an image. "Adaptive" means the combination weights also change with different subjects.

Specifically, for a test subject that has only T1 MR image, we partition its FLAIR-like image from our 3D c-GAN into non-overlapping small patches (16^3) and approximate each patch $S^{te,gan}$ by the convex combination of the training patches $S_1^{tr,gan}, S_2^{tr,gan} \ldots, S_{N_{tr}}^{tr,gan}$ (N_{tr} denotes the number of training subjects) from the FLAIR-like images at the same location. This is achieved by solving the following optimization problem:

$$\min_{w} \|\sum_{i=1}^{N_{tr}} w_i S_i^{tr,gan} - S^{te,gan}\|_2^2 \quad s.t. \sum w_i = 1, w_i \geq 0. \tag{4}$$

The convex combination in Eq. 4 assigns high weights to only a few very similar training patches, and near-zero weights to those dissimilar ones.

Due to the resemblance of our FLAIR-like images and the real ones, the above learned combination weights are further used to polish $S^{te,cc}$ by linearly combining the real FLAIR image patches $R_1^{tr}, R_2^{tr} \ldots, R_{N_{tr}}^{tr}$ at the same location in the training set:

$$S^{te,cc} = \sum_{i=1}^{N_{tr}} w_i R_i^{tr}. \tag{5}$$

Please note that the linear combination gives even better results than local non-linear mapping as shown in our experimental study. Although this convex combination imputes some artefacts that affect the appearance of the synthesized images, it proves to be an effective strategy to improve the segmentation, which is our ultimate target.

2.4. Brain tumor segmentation model

To effectively segment brain tumor with the synthesized FLAIR-like and T1 MR images, we utilize an 11-layer, two-pathway 3D CNN segmentation model [16], which achieves the state-of-the-art performance on brain tumor segmentation[1]. The first pathway extracts small patches (17^3) with normal resolution. The second one processes context in low-resolution patches (19^3) down-sampled from the actual area of size 51^3. Using this two-pathway structure can investigate different scales of input MR images to cater for both local and contextual image features.

[1]U-net only shows mediocre performance for brain tumor segmentation (as compared in [17]). Therefore, we did not employ it for this task.

In our work, the above segmentation model takes two channels of input: T1 and FLAIR. We train the model in two steps. In the first step, the T1 and real FLAIR images of the training samples are used for training in the usual way. Then, in the second step, this model is further fine-tuned with the T1 and our synthesized FLAIR images of the training samples. The fine-tuning is essential because for a given test sample, it is the synthesized FLAIR image rather than the real unknown FLAIR image that is used for segmentation. Please note that, the FLAIR-like image of a training sample is generated by 3D cGAN and the local patch combination (Sec. 2.3) using all the training samples excluding itself.

3. EXPERIMENTSAL RESULT

3.1. Data and experimental setting

We evaluate our framework on the data set BRATS 2015 [1]. It consists of 274 subjects with the image size $240\times240\times155$ and four modalities: T1, T1C, T2 and FLAIR. Tumors are annotated as: 1) necrotic core, 2) edema, 3) non-enhancing and 4) enhancing core. We randomly select 230 subjects as training samples, and the rest as our test set. For each sample, two modalities, T1 and FLAIR, are used. The whole tumor and the tumor core part (classes 1,3,4) are segmented. We linearly scale the original intensity values in all images to $[-1, 1]$ according to [16], without any additional contrast.

The learning rate of our 3D cGAN is fixed as 0.0002 in the first 100 training epochs and then linearly decays to zero in the next 100 epochs. Adam solver with batch-size 6 is applied, and λ is fixed as 300 when training 3D cGAN. Fine-tuning the segmentation model is executed with 15 epochs.

We evaluate our method from both the synthesis quality and the tumor segmentation performance. Following the literature, to evaluate synthesis quality, we use PSNR and normalized mean squared error (NMSE). To evaluate segmentation performance, we report dice scores (DSC) for both the whole tumor and the tumor core part.

3.2. Results and discussion

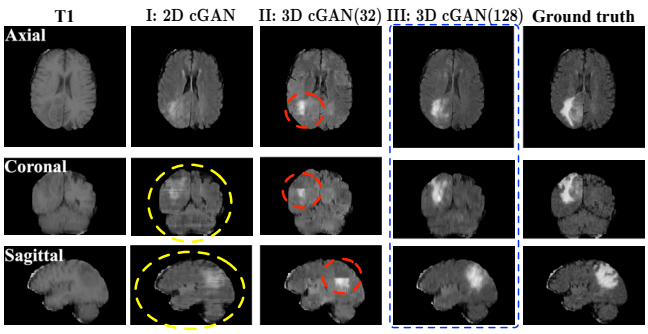

Fig. 2: *Visual comparison of the synthesized FLAIR images by 2D cGAN, 3D cGAN (32) and 3D cGAN (128). Discontinuity in the coronal and sagittal slices (in the yellow circles) is significant when using 2D cGAN. 3D cGAN (32) performs worse in the tumor region (in the red circles) when compared with 3D cGAN (128).*

Table 1: Quantitative evaluation results of the synthesized images and brain tumor segmentation
* Numbers with underline indicate they are statistically significantly different from our proposed method (**III**+local adaptive fusion), according to a two-sided, paired t-test (solid line $p < 0.05$, dotted line $p < 0.1$)

Methods	**Synthesis quality** (PSNR/NMSE%)		**Segmentation performance** (DSC%)	
	whole image	tumor	whole tumor	core
I: 2D cGAN	19.45/25.17	17.70/15.10	59.83	66.59
II: 3D cGAN (32)	19.53/25.12	18.09/13.05	52.06	50.02
III: **3D cGAN (128)**	20.45/25.08	19.13/12.68	66.35	72.09
IV: 3D cGAN (128)+3D cGAN (32)	19.94/24.99	18.73/13.45	66.61	72.14
V: 3D cGAN (128)+3D cGAN (128)	20.23/25.52	19.11/12.85	66.59	67.83
T1	N.A.		67.18	63.00
T1+real FLAIR (ideal scenario)	N.A.		*82.17*	*85.49*
Proposed (III+local adaptive fusion)	**20.68/22.67**	**19.27/11.86**	**68.23**	**72.28**

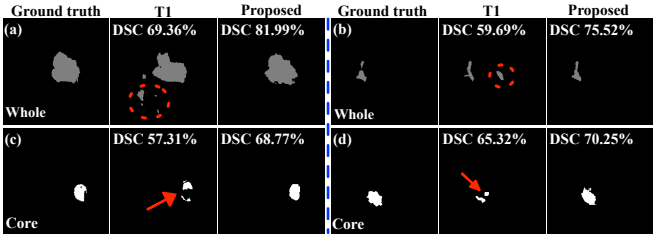

Fig. 3: *Four patients of the segmented whole tumor and core parts by single T1 modality and our proposed method.*

To evaluate the effectiveness of our proposed method, we compare it with **I**: 2D cGAN on the whole zero-padded axial slices (256^2) [14], **II**: 3D cGAN on small patches (32^3), **III**: our proposed 3D cGAN on large patches (128^3), **IV**: local non-linear mapping (3D cGAN on 32^3-size patches) applied after the method **III**, and **V**: whole-image refinement by concatenating 3D cGANs like [12]. Note that the method **III** and **V** are the reduced variants of our proposed method. They are compared in this experiment to better demonstrate the effectiveness and advantage of the proposed method. Performance of the above-mentioned methods on the image synthesis and tumor segmentation is summarized in Table 1.

From Table 1, it can be seen that the synthesized images by our method show the best quality with the highest PSNR and the lowest NMSE. This result is better than that of the 2D cGAN and all other 3D cGAN settings. Also, as shown, all the 3D cGAN based methods outperform the 2D cGAN in terms of synthesis quality, indicating the importance of considering 3D information during image synthesis. In addition, using the concatenated 3D cGANs (**V**) cannot further improve the synthesis results. Fig. 2 shows an example of the visual comparison between the synthesized FLAIR-like images by **I**, **II** and our 3D cGAN (**III**). When looking into the coronal and sagittal slices, it is found that the discontinuity along these two directions is more significant in **I** than in **II** and **III**. Furthermore, it can be seen that using large patches as in our proposed 3D cGAN (**III**) can better synthesize the tumor parts than **II** that uses small patches.

The above observations are further supported by the segmentation performance in Table 1. Again, our method achieves the highest dice score of 68.23%, which is more than 8 percentage points higher than that of the 2D cGAN. Our performance is also better than **III** that merely uses our 3D cGAN, indicating the essence of employing the proposed local adaptive fusion strategy. It is worthy noting that our method also beats **IV** that utilizes local non-linear fusion rather than our local linear one, in both synthesis and segmentation performance. And either non-linear or linear local fusion can further improve the results of 3D cGAN.

Table 1 also gives the dice scores of using the single modality of T1. As shown, jointly considering T1 and our synthesized FLAIR images can improve the tumor core part segmentation from 63% to 72.28% (ours wins in 33 subjects, and loses in 11 subjects), and the whole tumor segmentation from 67.18% to 68.23% (ours wins in 28 subjects, and loses in 16 subjects) compared with using T1 only. This suggests that our synthesized FLAIR-like images may carry complementary information about the soft tissue changes, which is helpful for T1-based brain tumor segmentation. Fig. 3 displays four patients of the segmented whole tumors and core parts by T1-only method and our proposed one. As shown, when using T1 images only, some healthy brain tissues are segmented as parts of tumors (in the red circle) and some tumor core parts (indicated by the red arrow) are missing. The results are improved when our additional synthesized FLAIR images are jointly considered. In addition, in the ideal scenario, when the real FLAIR images are available with T1, the dice score could go up to 82.17%, indicating the large potential of our research direction.

In summary, both the visual and quantitative results demonstrate the capacity of our method in synthesizing FLAIR images from T1, and the benefits of using our synthesized FLAIR images to improve T1-based brain image segmentation, especially for tumor core parts.

4. CONCLUSION

FLAIR images are crucial to achieve accurate brain tumor segmentation. In this paper, we investigate how to synthesize FLAIR images from T1 in order to improve T1-based single modality brain tumor segmentation. Our framework generates the synthesized FLAIR images through the proposed 3D cGAN and the local adaptive fusion scheme. With the two-pathway 3D CNN segmentation model, the synthesized FLAIR images effectively boost the segmentation of whole tumors and tumor core parts from the modality of T1 only.

5. REFERENCES

[1] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al., "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[2] Amod Jog, Aaron Carass, Snehashis Roy, Dzung L Pham, and Jerry L Prince, "Random forest regression for magnetic resonance image synthesis," *Medical image analysis*, vol. 35, pp. 475–488, 2017.

[3] Nicolas Cordier, Hervé Delingette, Matthieu Lê, and Nicholas Ayache, "Extended modality propagation: image synthesis of pathological cases," *IEEE transactions on medical imaging*, vol. 35, no. 12, pp. 2598–2608, 2016.

[4] Yawen Huang, Ling Shao, and Alejandro F Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding," *arXiv preprint arXiv:1705.02596*, 2017.

[5] Snehashis Roy, Aaron Carass, Amod Jog, Jerry L Prince, and Junghoon Lee, "Mr to ct registration of brains using image synthesis," in *Proceedings of SPIE*. NIH Public Access, 2014, vol. 9034.

[6] Johanna Degen and Mattias P Heinrich, "Multi-atlas based pseudo-ct synthesis using multimodal image registration and local atlas fusion strategies," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 160–168.

[7] Yan Wang, Guangkai Ma, Le An, Feng Shi, Pei Zhang, David S Lalush, Xi Wu, Yifei Pu, Jiliu Zhou, and Dinggang Shen, "Semisupervised tripled dictionary learning for standard-dose pet image prediction using low-dose pet and multimodal mri," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 3, pp. 569–579, 2017.

[8] Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 305–312.

[9] Mehdi Mirza and Simon Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[10] Avi Ben-Cohen, Eyal Klang, Stephen P Raskin, Michal Marianne Amitai, and Hayit Greenspan, "Virtual pet images from ct data using deep convolutional networks: Initial results," *arXiv preprint arXiv:1707.09585*, 2017.

[11] Xin Yi and Paul Babyn, "Sharpness-aware low dose ct denoising using conditional generative adversarial network," *arXiv preprint arXiv:1708.06453*, 2017.

[12] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen, "Medical image synthesis with context-aware generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 417–425.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint arXiv:1611.07004*, 2016.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[16] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[17] Kuan-Lun Tseng, Yen-Liang Lin, Winston Hsu, and Chung-Yang Huang, "Joint sequence learning and cross-modality convolution for 3d biomedical segmentation," *arXiv preprint arXiv:1704.07754*, 2017.