

[CVPR](#)**Computer Vision and Pattern Recognition 2018**

Jun 19, 2018 - Jun 21, 2018, Salt Lake City, USA

**Reviews For Paper****Paper ID** 2821**Title** Translating and Segmenting Multimodal Medical Volumes with Cycle- and Shape-Consistency Generative Adversarial Network**Masked Reviewer ID:** Assigned\_Reviewer\_2**Review:**

Question	
<p>[Paper Summary] What is the paper about? Please, be concise (3 to 5 sentences)</p>	<p>This paper presents a very interesting work that translates multimodal medical volume (CT and MRI) and segments anatomical structures (heart) simultaneously. They design two frameworks: generator (based on GAN) and segmentor (based on U-net). The two networks mutually boost each other to achieve better performance.</p>
<p>[Paper Strengths] Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc). For instance, a theoretical paper may need no experiments, while a paper with a new approach may require comparisons to existing methods.</p>	<p>The authors do a good job in explaining the motivation and design of their work. The method was tested on a relatively large data set and was compared with a couple of ad-hoc methods.</p>
<p>[Paper Weaknesses] Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). Note: If you think there is an error in the paper, please explain why it is an error. Also remember that theoretical results/ideas are essential to CVPR (some theoretical papers may not need to have experiments). If the theory is novel and interesting, but the results did not outperform other existing algorithms, it is not necessarily a reason to reject. It is not appropriate to ask for comparisons with unpublished papers and papers published after the CVPR deadline. In all cases, please be polite and constructive. CVPR 2018 policy on dual submission and arxiv appears at:  <a href="http://cvpr2018.thecvf.com/submission/main_conference/author_guidelines">http://cvpr2018.thecvf.com/submission/main_conference/author_guidelines</a>.</p>	<p>Need some details about the network design and explanation of data.</p> <ol style="list-style-type: none"> <li>1. Section 3.6, G A/B, S A/B, D A/B not defined</li> <li>2. Section 4, it will be desirable to have a diagram showing the detailed architecture of the network, such as the modified U-net and CycleGAN.</li> <li>3. Section 5.1, is this a public data set? Need more description about the patient population. Do you have manual segmentation for all 4000 cases?</li> <li>4. Section 5.1, how do you split training and</li> </ol>

	validation sets? 5. How much does the multimodel data improve the segmentation result? If only CT data is available, what would be the segmentation performance? Please conduct experiment or discuss.
[Preliminary Rating] Please rate the paper according to one of the following six choices:	Poster
[Preliminary Evaluation] Please indicate to the AC, your fellow reviewers, and the authors your current opinion on the paper. Please tell the ACs what points you think have the most weight in your reviews and summary, and why.	The idea of translating and segmenting multimodel images is somewhat novel. The authors do a good job in explaining the motivation and design of their work. The experimental result is good.
[Confidence]	Very Confident - to stress that you are pretty sure about your conclusions (e.g., you are an expert who works in the paper's area).

**Masked Reviewer ID:** Assigned\_Reviewer\_3

**Review:**

Question	
[Paper Summary] What is the paper about? Please, be concise (3 to 5 sentences)	The paper concerns cross-modality 3D image synthesis based on a novel cycle- and shape-consistent generative adversarial network (GAN). In contrast to previous similar approaches, training data for the presented network is not required to consist of matching image pairs. As intended application of this method, the authors state radiotherapy. Qualitative and quantitative evaluation has been performed on 4354 computed tomography (CT) scans and 142 magnetic resonance imaging (MRI) scans.
[Paper Strengths] Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty,	Besides some minor typos, the paper is very

theoretical approach and/or technical correctness, adequate evaluation, clarity, etc). For instance, a theoretical paper may need no experiments, while a paper with a new approach may require comparisons to existing methods.

well written and outlines the presented network in sufficient detail and clarity. Related state-of-the-art has been addressed adequately.

The paper addresses a critical challenge that is not only major to radiation therapy but to so many more clinical areas. According to many experts, true image synthesis is the solution to turn away from ionizing and invasive imaging modalities.

[Paper Weaknesses] Please discuss, justifying your comments with the appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). Note: If you think there is an error in the paper, please explain why it is an error. Also remember that theoretical results/ideas are essential to CVPR (some theoretical papers may not need to have experiments). If the theory is novel and interesting, but the results did not outperform other existing algorithms, it is not necessarily a reason to reject. It is not appropriate to ask for comparisons with unpublished papers and papers published after the CVPR deadline. In all cases, please be polite and constructive. CVPR 2018 policy on dual submission and arxiv appears at:  
[http://cvpr2018.thecvf.com/submission/main\\_conference/author\\_guidelines](http://cvpr2018.thecvf.com/submission/main_conference/author_guidelines).

Despite my above appreciation of the relevance of the paper topic, I have some severe doubts that the presented method is mature enough to have an impact on the mentioned development. I will detail my concerns in the following:

1. I am missing a thorough description of the medical background, that is the relevance of image synthesis for radiation therapy. In the current state, it does not become clear that, in order to change the current X-ray-guided workflow to MRI-guidance, positioning planning from a preinterventional CT volume of the patient has to be transferred to an MRI scan of the same patient.

2. Now, it also becomes apparent that it is essentially the second part of the presented network (multi-modal volume segmentation) that is important for radiotherapy application. There are even quite a few experts that question whether perfect volume-to-volume synthesis is necessary for

segmentation transfer... Anyway, the presentation of methodology puts much importance on the image synthesis part, which may be a bit misleading and not representing the actual need of clinical radiotherapy.

3. Having explained the importance of segmentation transfer, I have serious doubts that the shown improvements to 69.2% and 69.6% S-score (is this DICE?) respectively makes any difference to the current state-of-art. Looking at the visual results in Fig. 6, I can make out some promising improvement towards the employed Baseline algorithm. However, knowing that this certainly is not state-of-the-art (atlas-based methods perform superior than that), the relevance is minor.

4. I acknowledge the authors' efforts to come up with a general solution that does not need matching training image pairs. But I also fear that the clinical area the authors have chosen, is highly unsuitable for this. The biggest issue in radiation therapy is that every single outlier, every single pitfall, imposes a direct threat to the patient (under- or over-radiation). I have only very little hope that there is a general solution to this problem NOT using patient-specific data and matching image pairs. And after all, is it really necessary to have a one-for-all solution ready? In today's clinics, everything is patient-specific and personalized.

5. In Fig. 3 it seems that the additional GAN shape-consistency term yields some sort of shrinking of MRI shapes compared to CT shapes. This may explain the very little S-score improvement in Table 1.

6. Although the amount of validation data is sufficient, some very important details are missing. Why did the authors decide to use contrasted CT volumes? Even if they intended to do their shape quality evaluation on the heart, the intensity and contrast situation in a contrasted CT volume compared to an uncontrasted MRI volume is very different. I understand that patients suspected to have a cardiovascular disease are listed for a diagnostic contrasted CT, but this may not be a suitable dataset for the presented network. Besides, there is a variety of easier anatomies available in the human body that may be more suitable.

Minor comments:

\* In 020 and 040: abbreviations CT and MRI are used without prior definition.

\* In 025: ... structures, which could [be] changed by ...

\* In 114: ... due to the [lack] of direct ...

\* In 116: ... as we [explain] in the ...

\* In 235: ..., A and B[,] CycleGAN ...

\* In 259: ... are cycle [consistent] ...

\* In 349: ... force [the] annotation[s to be preserved].

\* In 377: ... will demonstrate [this] in ...

\* In 458: ... maximizes [its] effectiveness.

	<p>* In 628: ... aorta and [spine].</p> <p>* In 651: ... data [and] can help ...</p> <p>* In 666: ... We [join] ...</p>
[Preliminary Rating] Please rate the paper according to one of the following six choices:	Weak Reject
<p>[Preliminary Evaluation] Please indicate to the AC, your fellow reviewers, and the authors your current opinion on the paper. Please tell the ACs what points you think have the most weight in your reviews and summary, and why.</p>	<p>The paper addresses a critical challenge that is not only major to radiation therapy but to so many more clinical areas, but I have severe doubts that the presented method is mature enough to have an impact. In particular the shown improvements to 69.2% and 69.6% S-score (DICE?) is not relevant and is outperformed already by atlas-based methods.</p>
<p>[Rebuttal Requests] Please pose questions you want to be answered in the rebuttal.</p>	<p>Why does the L1 norm in Equ (1) show better visual results than the L2 norm?</p> <p>How and why were the number of training epochs defined?</p> <p>Why did the authors chose the heart as validation anatomy?</p> <p>Why is the amount of available CT volumes so much higher than the MRI volumes? Was this on purpose?</p> <p>Have CT and MRI volumes been acquired at a single site or various?</p> <p>Who performed the data annotations?</p>
[Confidence]	<p>Very Confident - to stress that you are pretty sure about your conclusions (e.g., you are an expert who works in the paper's area).</p>

**Masked Reviewer ID:** Assigned\_Reviewer\_4

**Review:**

Question	
<p>[Paper Summary] What is the paper about? Please, be concise (3 to 5 sentences)</p>	<p>This paper seeks to improve the ability to perform cross modal translation and segmentation of medical volumes, specifically translating between CT and MRI images.</p> <p>This work takes on the more challenging version of this; to learn this mapping from unpaired data. To support this, they use the cycle consistency constraint, and then additionally add shape constraints enforced by semantic segmentation (so the original images can't be warped so much that semantic segmentation fails). They explore reasonable ways to train all the networks and show:</p> <ol style="list-style-type: none"> <li>1. Plausible (and visually better than previous works) cross domain image synth\esis, and</li> <li>2. The cross-modal translation creates synthetic data good enough to quantitatively improve segmentation performance.</li> </ol>
<p>[Paper Strengths] Please discuss, justifying your comments with the appropriate level of details, the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, etc). For instance, a theoretical paper may need no experiments, while a paper with a new approach may require comparisons to existing methods.</p>	<p>This paper offers an approach to a very hard problem. The approach integrates some of the most interesting recent ideas in network architectures and applies them to a relevant problem. The methods could be translated to other domains and the results are promising.</p>
<p>[Paper Weaknesses] Please discuss, justifying your comments with the</p>	<p>The paper is motivated by</p>

<p>appropriate level of details, the weaknesses of the paper (i.e. lack of novelty – given references to prior work-, lack of novelty, technical errors, or/and insufficient evaluation, etc). Note: If you think there is an error in the paper, please explain why it is an error. Also remember that theoretical results/ideas are essential to CVPR (some theoretical papers may not need to have experiments). If the theory is novel and interesting, but the results did not outperform other existing algorithms, it is not necessarily a reason to reject. It is not appropriate to ask for comparisons with unpublished papers and papers published after the CVPR deadline. In all cases, please be polite and constructive. CVPR 2018 policy on dual submission and arxiv appears at:  <a href="http://cvpr2018.thecvf.com/submission/main_conference/author_guidelines">http://cvpr2018.thecvf.com/submission/main_conference/author_guidelines</a>.</p>	<p>medical applications, for "diagnosis and surgical planning". While the results of "images are plausible", and "DICE scores for segmentation improve" are great results, the motivation of the medical domain requires some discussion about how the errors affect that domain. Little discussion of the error modes is offered at all.</p>
<p>[Preliminary Rating] Please rate the paper according to one of the following six choices:</p>	<p>Poster</p>
<p>[Preliminary Evaluation] Please indicate to the AC, your fellow reviewers, and the authors your current opinion on the paper. Please tell the ACs what points you think have the most weight in your reviews and summary, and why.</p>	<p>I like this paper. It uses a combination of several different pieces that are well chosen and the effects of different choices are reasonably well explored. The dataset is rather large within this domain, allowing the authors to evaluate various different data need requirements. The method improves the results over the compared works by more than a trivial amount. Collectively, I believe this merits publication.</p>
<p>[Confidence]</p>	<p>Confident - to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good work from bad work in that area).</p>