Breaking the Dilemma of Medical Image-to-image Translation

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Supervised Pix2Pix and unsupervised Cycle-consistency are two modes that dominate the field of medical image-to-image translation. However, neither modes are ideal. The Pix2Pix mode has excellent performance. But it requires paired and well pixel-wise aligned images, which may not always be achievable due to respiratory motion or anatomy change between times that paired images are acquired. The Cycle-consistency mode is less stringent with training data and works well on unpaired or misaligned images. But its performance may not be optimal. In order to break the dilemma of the existing modes, we propose a new unsupervised mode called RegGAN for medical image-to-image translation. It is based on the theory of "loss-correction". In RegGAN, the misaligned target images are considered as noisy labels and the generator is trained with an additional registration network to fit the misaligned noise distribution adaptively. The goal is to search for the common optimal solution to both image-to-image translation and registration tasks. We incorporated RegGAN into a few state-of-the-art image-to-image translation methods and demonstrated that RegGAN could be easily combined with these methods to improve their performances. Such as a simple CycleGAN in our mode surpasses latest NICEGAN even though using less network parameters. Based on our results, RegGAN outperformed both Pix2Pix on aligned data and Cycle-consistency on misaligned or unpaired data. RegGAN is insensitive to noises which makes it a better choice for a wide range of scenarios, especially for medical image-to-image translation tasks in which well pixel-wise aligned data are not available. Code and data used in this study can be found at `https://github.com/Kid-Liet/Reg-GAN`.

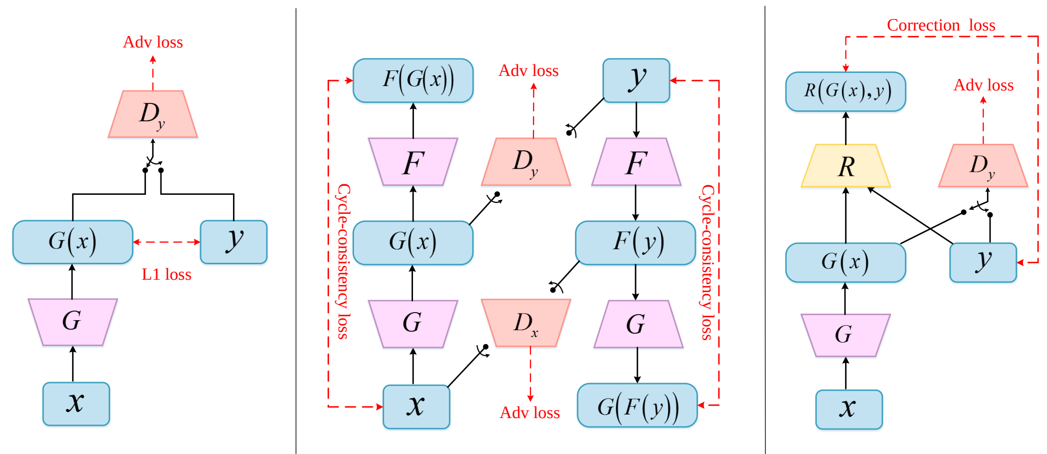
1 Introduction

Generative adversarial networks (GANs)[1] is a framework that simultaneously trains a generator G and a discriminator D through an adversarial process. The generator is used to translate the distribution of source domain images X to the distribution of target domain images Y. The discriminator is used to determine if the target domain images are likely from the generator or from the real data.

$$\min_G \max_D \mathcal{L}_{Adv}(G, D) = \mathbb{E}_y [log(D(y))] + \mathbb{E}_x [log(1 - D(G(x)))] \tag{1}$$

Supervised Pix2Pix[2] and unsupervised Cycle-consistency[3] are the two commonly used modes in GANs. Pix2Pix updates the generator ($G : X \to Y$) by minimizing pixel-level $L1$ loss between the source image x and the target image y. Therefore, it requires well aligned paired images, where each pixel has a corresponding label.

$$\min_G \mathcal{L}_{L1}(G) = \mathbb{E}_{x,y} [\|y - G(x)\|_1] \tag{2}$$



(a) Pix2Pix **(b) Cycle-consistency** **(c) RegGAN**

Figure 1: Comparison among the modes of Pix2Pix, CycleGAN and RegGAN.

Well aligned paired images, however, are not always available in real-world scenarios. To address the challenges caused by misaligned images, Cycle-consistency was developed which was based on the assumption that the generator G from the source domain X to the target domain Y ($G : X \to Y$) was the reverse of the generator F from Y to X ($F : Y \to X$). Compared to the Pix2Pix mode, the Cycle-consistency mode works better on misaligned or unpaired images.

$$\min_{G} \min_{F} \mathcal{L}_{Cyc}(G, F) = \mathbb{E}_x \left[\| F(G(x)) - x \|_1 \right] + \mathbb{E}_y \left[\| G(F(y)) - y \|_1 \right] \tag{3}$$

The Cycle-consistency mode, however, has its limitations. In the field of medical image-to-image translation, it requires not only the style translation between image domains, but also the translation between specific pair of images. The optimal solution should be unique. For example, the translated images should maintain the anatomical features of the original images as much as possible. It is known that the Cycle-consistency mode may produce multiple solutions[4, 5], meaning that the training process may be relatively perturbing and the results may not be accurate. The pix2pix mode is not ideal either. Even though it has a unique solution, it is difficult to satisfy the requirement asking for well aligned paired images. With misaligned images, the errors are propagated through the Pix2Pix mode which may result in unreasonable displacements on the final translated images.

As of today, there is no image-to-image translation mode that can outperform both the Pix2Pix mode on aligned data and the Cycle-consistency mode on misaligned or unpaired data. Inspired by[6–10], we consider the misaligned target images as noisy labels, which means that the existing problem is regarded as supervised learning with noisy labels. So we introduce a new image-to-image translation mode called RegGAN. Figure 1 provides a comparison of the three modes: Pix2Pix, Cycle-consistency and RegGAN. To facilitate reading, we summarize our contributions as follows.

- We demonstrate the feasibility of RegGAN from the theoretical perspective of "loss-correction". Specifically, we train the generator using an additional registration network to fit the misaligned noise distribution adaptively, with the goal to search for the common optimal solution for both image-to-image translation and registration tasks.
- RegGAN eliminates the requirement for well aligned paired images and searches unique solution in training process. Based on our results, RegGAN outperformed both Pix2Pix on aligned data and Cycle-consistency on misaligned or unpaired data.
- RegGAN can be integrated into other methods without changing the original network architecture. Compared to Cycle-consistency with two generators and discriminators, RegGAN can provide better performance using less network parameters.

2 Related Work

Image-to-image Translation: Generative adversarial networks (GANs) have shown great potential in the field of image-to-image translation[11–16]. It has been successfully implemented in medical image analysis like segmentation[17], registration[18, 19] and dose calculation[20]. The existing modes, however, have their limitations. Specifically, the Pix2Pix mode[2] requires well aligned paired images which may not always be available. The Cycle-consistency mode can achieve unsupervised image-to-image translation. With a Cycle-consistency loss, it can be used for misaligned images. Based on Cycle-consistency, many methods[3, 21–30] have been developed including CycleGAN[3] and its variants such as MUNIT[31] and UNIT[32] in which both image content and style information are used to decouple and reconstruct the image-to-image translation task; U-gat-it[33] with a self-attention mechanism added; and NICEGAN[34] proposed to reuse the discriminator for encoding. The main limitation of Cycle-consistency is that it may produce multiple solutions and therefore is sensitive to perturbation, making it difficult to meet the high accuracy requirements of medical image-to-image translation tasks.

Learning from Noisy Labels: Neural network anti-noise training has made great progress. Current research are mainly focused on: estimating the noise transition matrix[7, 35–40], designing a robust loss function[41–44], correcting the noise label[45–50], sampling importance weighting[51–55] and meta-learning[56–59]. Our work is in the category of estimating the noise transition matrix. Compared to conventional noise transition estimation, we mitigate the issue and simplify the task by acquiring prior knowledge of noise distribution.

Closest to our work, Arar.M et al[60] introduced a multi-modal registration method for natural images based on geometry preserving. But their work focused only on registration and did not demonstrate results of image-to-image translation or discuss the relationship between registration and image-to-image translation. The key insight of our work is that we demonstrated the feasibility of using registration to significantly improve the performance of image-to-image translation because the noise could be eliminated adaptively during the joint training process. What we propose in the paper is a completely new mode for medical image-to-image translation.

3 Methodology

3.1 Theoretical Motivation

If we consider misaligned target images as noisy labels, the training for image-to-image translation becomes a supervised learning process with noisy labels. Given a training dataset $\{(x_n, \widetilde{y}_n)\}_{n=1}^{N}$ with N noisy labels in which x_n, $\widetilde{y}_n$ are images from two modalities and assume y_n is the correct label for x_n, but it is unknown in real-world scenarios. Our goal is to train a generator using the dataset $\{(x_n, \widetilde{y}_n)\}_{n=1}^{N}$ with noisy labels and achieve the performance equivalent to trained on clean dataset $\{(x_n, y_n)\}_{n=1}^{N}$ as much as possible. Direct optimization based on Equations 4 usually does not work and can lead to bad results because the generator cannot squeeze out the influence of noise.

$$\hat{G} = \arg\min_{G} -\frac{1}{N}\sum_{n=1}^{N}\mathcal{L}\left(G\left(x_n\right), \widetilde{y}_n\right) \tag{4}$$

To address the noise issue, we propose a solution based on "loss-correction"[7] shown in Equations 5. Our solution corrects the output of the generator $G(x_n)$ by modeling a noise transition ϕ to match the noise distribution. Previously, Patrini et al[7] proved mathematically that the model trained with the noisy labels could be equivalent to the model trained with the clean labels, if the noise transition ϕ matches the noise distribution.

$$\hat{G} = \arg\min_{G} -\frac{1}{N}\sum_{n=1}^{N}\mathcal{L}\left(\phi \circ G\left(x_n\right), \widetilde{y}_n\right) \tag{5}$$

To achieve this, Goldberger et al[36] proposed to view the correct label as a latent random variable and explicitly model the label noise as a part of the network architecture, denoted by R. Then, Equations 5 can be rewritten in the form of log-likelihood, which is used as the loss function for neural network training.

$$\begin{aligned}\mathcal{L}(G,R) &= \sum_{n=1}^{N} log\left(p\left(\widetilde{y}_n|y_n;R\right)p\left(y_n|x_n;G\right)\right)\\ &= \sum_{n=1}^{N} log\left(p\left(\widetilde{y}_n|x_n;G,R\right)\right)\end{aligned} \tag{6}$$

3.2 RegGAN

Compared to existing methods that use expectation-maximum[7, 36], fully connected layers[35], anchor point estimate[37] and Drichlet-distribution[38] to solve Equations 6. In our problem, the type of noise distribution is clearer, it can be expressed as displacement error: $\widetilde{y} = y \circ T$. Here T is expressed as a random deformation field, which produces random displacement for each pixel. So we adopt a registration network R after the generator G as label noise model to correct the results. The Correction loss is shown Equations 7:

$$\min_{G,R} \mathcal{L}_{Corr}(G,R) = \mathbb{E}_{x,\widetilde{y}}\left[\|\widetilde{y} - G(x) \circ R(G(x), \widetilde{y})\|_1\right] \tag{7}$$

where, $R(G(x), \widetilde{y})$ is the deformation field and $\circ$ represents the resamples operation. The registration network is based on U-Net[61]. A smoothness loss[62] is defined in Equations 8 to evaluate the smoothness of the deformation field and minimize the gradient of the deformation field.

$$\min_{R} \mathcal{L}_{Smooth}(R) = \mathbb{E}_{x,\widetilde{y}}\left[\|\nabla R(G(x), \widetilde{y})\|^2\right] \tag{8}$$

Finally, we add the Aversarial loss between the generator and the discriminator (Equations 1), and the total loss is expressed in Equations 9.

$$\min_{G,R}\max_{D} \mathcal{L}_{Total}(G,R,D) = \mathcal{L}_{Corr} + \mathcal{L}_{Smooth} + \mathcal{L}_{Adv} \tag{9}$$

4 Experiments

Performance evaluation of RegGAN was conducted through three investigations to 1) demonstrate the feasibility and superiority of the RegGAN mode in various methods, and 2) assess RegGAN's sensitivity to noise, and 3) explore the availability of the RegGAN on unpaired data.

4.1 Dataset

The open-access dataset (BraTS 2018[63]) was used to evaluate the proposed RegGAN mode. The training dataset and testing dataset contained 8457 and 979 pairs of T1 and T2 MR images, respectively. BraTS 2018 was selected because the original images were paired and well aligned. We created misaligned images by randomly adding different levels of rotation, translation and rescaling to the original images. And we randomly sample one image from T1 and the other one from T2 when training on unpaired images. The availability of well aligned paired images, misaligned paired images, and unpaired images allow us to evaluate the performances of all three modes (Pix2Pix, Cycle-consistency and RegGAN).

4.2 Performances in Different Methods

The primary motivation of introducing RegGAN was to address challenges caused by misaligned data. Therefore, in this section, misaligned data were used in model training to demonstrate the feasibility and superiority of RegGAN. We selected the most popular CycleGAN[3] and its variants MUNIT[31], UNIT[32], and NICEGAN[34] as the methods for evaluation and compared the following four modes for each method.

- **C(Cycle-consistency):** The most primitive mode of all methods, with Cycle-consistency loss (Equations 3) as the main constraint. Two generators and two discriminators are required in this mode.

Table 1: Comparison of CycleGAN, MUNIT, UNIT and NICEGAN using four training modes(C, C+R, NC and NC+R).

Index	Modes	CycleGAN	MUNIT	UNIT	NICEGAN
NMAE ↓	C	0.089	0.11	0.087	0.082
	C+R	(-0.012)0.077	(-0.022)0.088	(-0.013)0.074	(-0.011)**0.071**
	NC	0.11	0.10	0.098	0.089
	NC+R	(-0.038)**0.072**	(-0.021)**0.079**	(-0.027)**0.071**	(-0.019)**0.070**
PSNR ↑	C	23.5	20.6	24.6	25.2
	C+R	(+0.3)23.8	(+2.1)22.7	(+0.7)25.3	(+0.9)26.1
	NC	20.2	21.5	23.7	23.5
	NC+R	(+5.4)**25.6**	(+2.3)**23.8**	(+1.8)**25.5**	(+2.8)**26.3**
SSIM ↑	C	0.83	0.80	0.84	0.83
	C+R	(+0.02)0.85	(+0.03) 0.83	(+0.02)**0.86**	(+0.03)**0.86**
	NC	0.79	0.81	0.83	0.84
	NC+R	(+0.07)**0.86**	(+0.04)**0.85**	(+0.03) **0.86**	(+0.02)**0.86**

- **C+R (Cycle-consistency + Registration):** The RegGAN mode is combined with the mode **C**. Registration network (R) and Correction loss (Equations 7) are added to the constraints.
- **NC(Non Cycle-consistency):** Only Adversarial loss (Equations 1) is used for updating. Compared to the mode **C**, Cycle-consistency loss is removed. Only one generator and one discriminator are required in this mode.
- **NC+R(Non Cycle-consistency + Registration):** A registration network (R) and Correction loss (Equations 7) are added to the mode **NC**. It is the proposed RegGAN mode.

To evaluate the performance of each method on misaligned data, we randomly added [-5, +5] degrees of angle rotation, [-5, +5] percent of translation, and [-5, +5] percent of rescaling to the original T1 and T2 images on the training dataset.

To ensure fair comparison, we used the same training strategy and hyperparameters for all methods and modes (see supplementary materials for details). The Normalized Mean Absolute Error (NMAE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) were used as metrics to evaluate the performances of trained models based on the testing dataset. To avoid false high results of index, we excluded the image background from the calculation. Table 1 summarized the results for all methods and modes under the current investigation.

Based on the results from the Table 1, we can reach several conclusions. First, adding the registration network (**+R**) significantly improves the performances of the methods. This is true for all methods in both **C** and **NC** modes. It clearly demonstrates that RegGAN can be incorporated in various methods or combined with different network architectures to improve the performances. Second, the **C** mode is in general better than the **NC** mode for most of methods. Adding the registration network (**+R**) improves the performance of the **NC** mode more than that of the **C** mode. In fact, our results show that the **NC+R** mode is even better than the **C+R** mode, implying that "Cycle-consistency loss" may play a negative role when it is combined with RegGAN. Compared with the commonly used **C** mode with two generators and two discriminators, RegGAN has fewer parameters but provides better performance. The simple CycleGAN method in the **NC+R** mode outperforms the current state-of-the-art method NICEGAN in the **C** mode by 0.01, 0.4, 0.03 for NMAE, PSNR and SSIM, respectively. The **NC+R** mode can also be used to improve the performance of NICEGAN. In fact, the performance of NICEGAN in the **NC+R** mode is the best among all combinations of the 4 methods and 4 modes.

Figure 2 shows representative results from various combinations of the 4 methods (CycleGAN, MUNIT, UNIT and NICEGAN) and 4 modes (**C**, **C+R**, **NC** and **NC+R**). For all aspects of the image

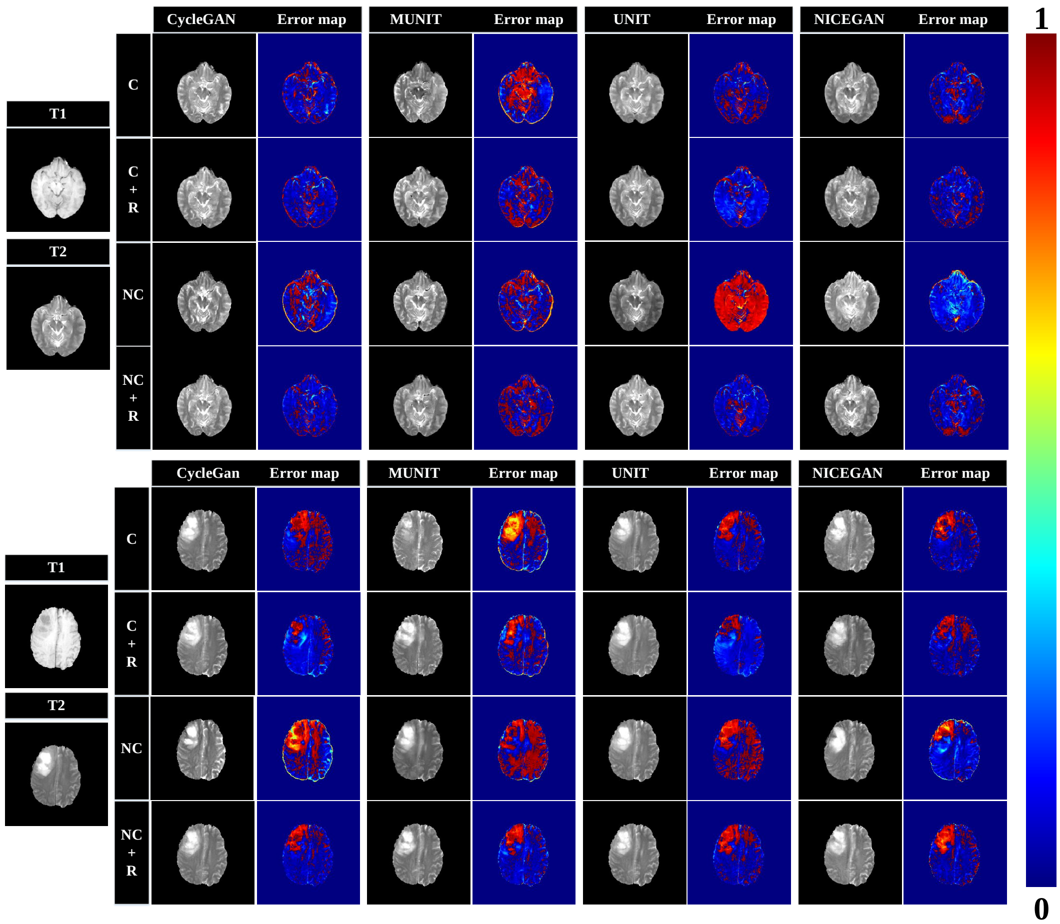


Figure 2: The errors of different modes in different methods.

(from the tumor areas and the details), the combinations that use the registration network (**+R**) always provide more realistic and accurate results than those that do not use the registration network (**+R**).

4.3 Performances in Different Noise Levels

To evaluate the sensitivity of RegGAN to noise, we selected a simple network architecture.(CycleGAN) with the intention to minimize interference from other factors. The same network architecture was used for all three modes: CycleGAN(**C**), Pix2Pix and RegGAN. Six levels of noise were used in the evaluation. Table 2 lists the specific noise setting and range for each noise level. Noise.0 means the original dataset with no added noise. Noise.5 is the highest level of noise. At Noise.5, the data are likely from different patients. Figure 3 shows example images at different levels of introduced noise.

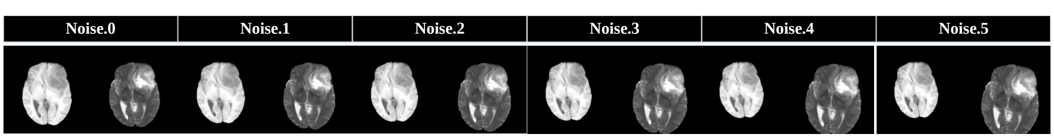


Figure 3: Example images at six different levels of introduced noise.

Table 2 lists the quantitative evaluation metrics from 3 modes at 6 levels of noise. It is clear that RegGAN outperforms CycleGAN(**C**) under all noise levels. Figure 4(**a**) shows the test results from each epoch during the training process for both RegGAN and CycleGAN(**C**). Curves of different

Table 2: Comparison of the NMAE, PSNR and SSIM for CycleGAN(**C**), Pix2Pix and RegGAN under 6 levels of noise.

		Noise.0	Noise.1	Noise.2	Noise.3	Noise.4	Noise.5
Setting	Rotate	$0°$	$\pm1°$	$\pm2°$	$\pm3°$	$\pm4°$	$\pm5°$
	Translation	0%	$\pm2\%$	$\pm4\%$	$\pm6\%$	$\pm8\%$	$\pm10\%$
	Rescaling	0%	$\pm2\%$	$\pm4\%$	$\pm6\%$	$\pm8\%$	$\pm10\%$
CycleGAN(**C**)	NMAE $\downarrow$	0.084	0.095	0.087	0.094	0.087	0.110
	PSNR $\uparrow$	23.9	22.5	23.7	23.3	23.9	23.7
	SSIM $\uparrow$	0.83	0.83	0.82	0.81	0.82	0.79
Pix2Pix	NMAE $\downarrow$	0.075	0.103	0.139	0.161	0.175	0.181
	PSNR $\uparrow$	25.6	22.3	18.9	16.2	15.3	15.0
	SSIM $\uparrow$	0.85	0.82	0.78	0.76	0.74	0.74
RegGAN	NMAE $\downarrow$	**0.071**	**0.073**	**0.071**	**0.072**	**0.072**	**0.072**
	PSNR $\uparrow$	**26**	**25.6**	**25.9**	**25.7**	**25.4**	**25.2**
	SSIM $\uparrow$	**0.86**	**0.86**	**0.86**	**0.86**	**0.86**	**0.85**

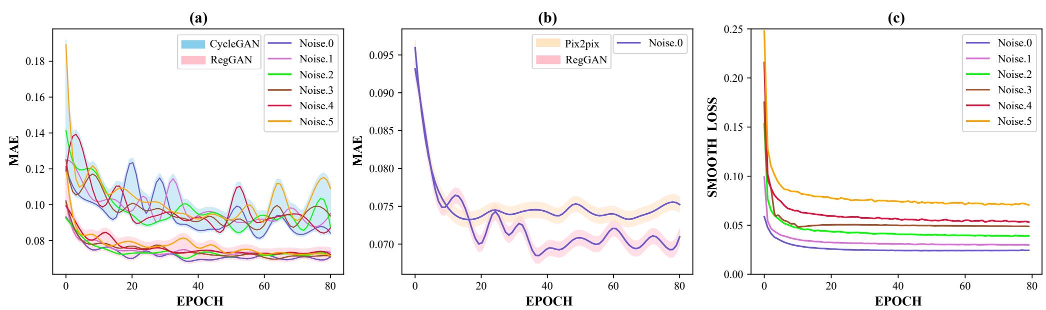


Figure 4: Quantitative evaluation metrics at different epochs in the training process. **(a)** Comparison of CycleGAN and RegGAN at different levels of noise. **(b)** Comparison of Pix2Pix and RegGAN at Noise.0 (i.e., no noise). **(c)** RegGAN's Smoothness loss under different levels of noise.

colors corresponds to different levels of noise. We notice that CycleGAN(**C**) is not very stable during the training process. The test results fluctuate significantly and cannot converge well. This may be caused by the fact that the solution of CycleGAN (**C**) is not unique. As a comparison, RegGAN is quite stable. Although the results from different levels of noise may vary at the beginning of training, all curves converge to a similar result after multiple epoches of training, indicating that RegGAN is more robust to noise compared to CycleGAN (**C**).

Based on Table 2, we notice that the performance of Pix2Pix deteriorates rapidly as the noise increases. This is as expected because Pix2Pix requires well aligned paired images. Surprisingly, the performances of RegGAN at all noise levels exceed those of Pix2Pix with no noise. Figure 4**(b)** shows the test results at each epoch of RegGAN and Pix2Pix under Noise.0 (i.e., no noise). Theoretically, the performances of RegGAN and Pix2Pix should be similar on perfectly aligned paired datasets because the registration network of RegGAN does not help and RegGAN is equivalent to Pix2Pix. A possible explanation to our results is that in the medical field, the perfectly pixel-wise aligned dataset may not practically exist. Even for BraST 2018[62] which is recognized as well aligned, it is still possible that there exists slight misalignment. As a result, adding the registration network is always likely to improve the performances in real-world scenarios. To verify our explanation, we plotted the Smoothness loss of RegGAN under different noise levels as shown in Figure 4**(c)**. Large Smoothness loss corresponds to large deformation field displacement. First, we notice that the Smoothness loss under Noise.0 never completely goes to 0, indicating the existence of misalignment and potential usefulness of the registration network. Second, the noise level and Smoothness loss show a step-like positive correlation, which means that RegGAN can adaptively handle the noise distribution, i.e., the registration network can determine the range of deformation according to the noise level.

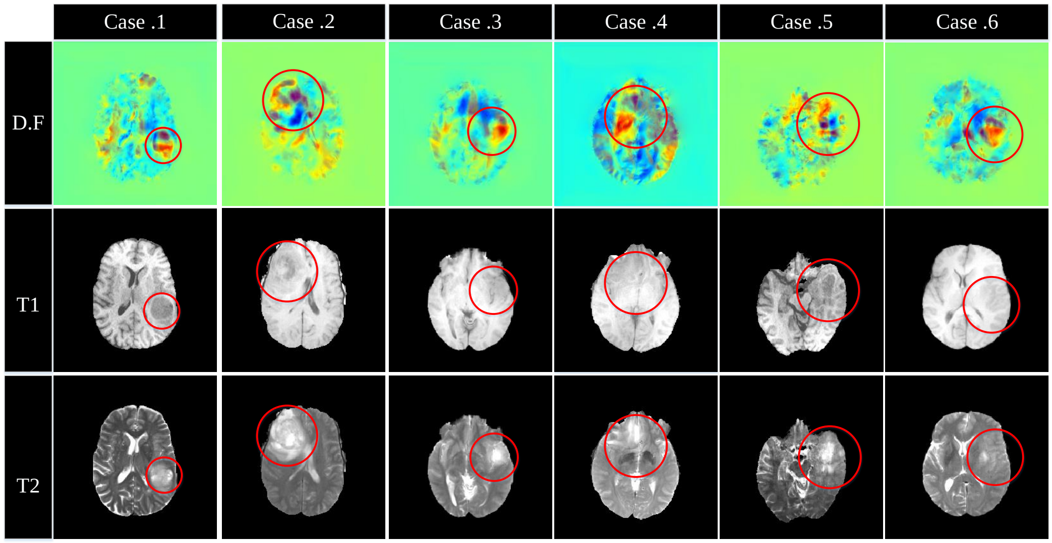
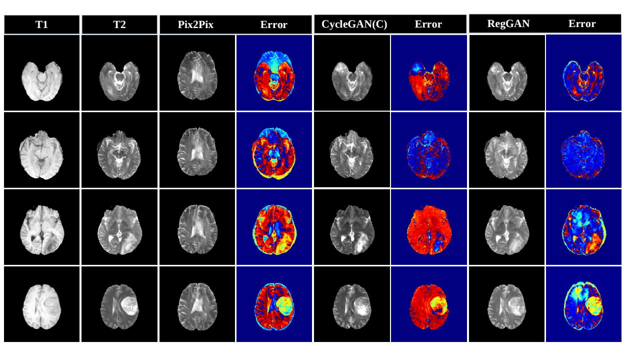


Figure 5: The misalignment of orginal image pairs and corresponding deformation fields.

we also show some original image pairs and visualize the corresponding deformation fields output by registration network in Figure 5. Obviously, there is some misalignment between the original T1 and T2 images, and such misalignment is represented by the deformation fields (highlighted by red circle).

4.4 Performances on Unpaired Dataset

So far, our investigations are based on paired datasets. We also want to explore how RegGAN performs using unpaired datasets. In practice, this is not recommended because even different patients may have similarities in their body tissues of adjacent layers. For unpaired datasets, we can conduct rigid registration first in 3D space and then use RegGAN for training. Unpaired data can be treated as having larger scale noise. If the correction capability is strong enough, RegGAN can still work effectively. The comparison of the performances of three modes on the unpaired dataset is shown in Figure 6.



Mode \ Index	NMAE↓	PSNR↑	SSIM↑
Pix2Pix	0.180	15.5	0.71
CycleGAN(**C**)	0.094	23.6	**0.83**
RegGAN	**0.086**	**24.0**	**0.83**

Figure 6: Performance comparison of the three modes (CycleGAN(**C**), Pix2Pix and RegGAN) on unpaired dataset.

With unpaired datasets, Pix2Pix no longer considers the characteristics of the input T1 images and thus has the worst performance. Due to the challenges in fitting the noise, the performance improvement from replacing CycleGAN(**C**) with RegGAN using unpaired datasets may not be as

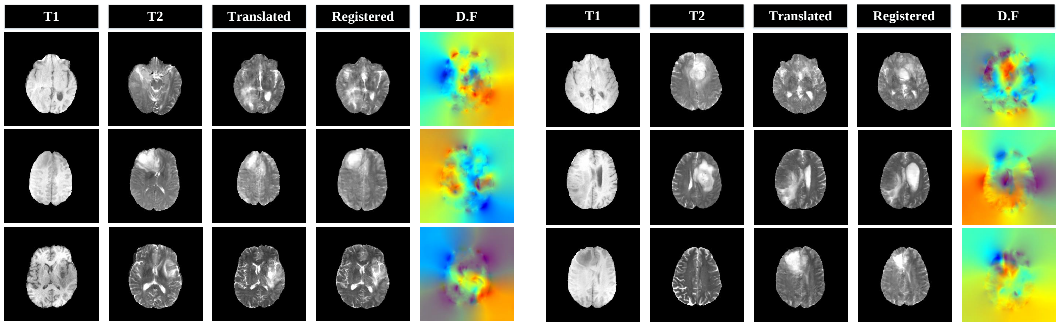


Figure 7: Display of RegGAN's output on unpaired data. **T1** and **T2** are unpaired images. The **Translated** represents the translation result of T1 to T2. **Registered** represents the registration result of the translated images. **D.F** represents deformation fields.

dramatic as that demonstrated using paired datasets, but RegGAN still has the best performance under unpaired conditions. In Figure 7, we show some examples of how RegGAN corrects noise on unpaired dataset. It can be seen that RegGAN will try its best to eliminate the influnce of noise through registration.

Based on our results, it is reasonable to reach the conclusions below. In all circumstances, RegGAN demonstrates better performance compared to Pix2Pix and CycleGAN(**C**).

- For paired and aligned conditions, RegGAN $\geq$ Pix2Pix $>$ CycleGAN(**C**).
- For paired but misaligned conditions, RegGAN $>$ CycleGAN(**C**) $>$Pix2Pix.
- For unpaired conditions, RegGAN $>$ CycleGAN(**C**) $>$Pix2Pix.

Conclusion

In this study, we introduced a new image-to-image translation mode RegGAN to the medical community that can break the dilemma of image-to-image translation task. Using a public BraST 2018 dataset, we demonstrated the feasibility of RegGAN and its superior performance compared to Pix2Pix and Cycle-consistency. We validated that RegGAN can be incorporated into various existing methods to improve their performances. We also evaluated the sensitivity of RegGAN to noise. Our results confirmed that RegGAN could adapt well to various scenarios from no noise to large-scale noise. The superior performance of RegGAN makes it a better choice over Pix2Pix and Cycle-consistency whether datasets are aligned or not. However, this mode may not work well on natural images. The noise may cannot be considered simply as deformation errors due to the differences in natural images are much greater than those in medical images.

Broader Impact

Image-to-image translation has been one of the main focuses in medical image analysis, as it aids in diagnosis and treatment. Previously, physicians had to use different medical imaging equipments if they wanted to get multi-modal images of a patient, which was time-consuming and expensive. Pix2Pix mode is expected to solve this problem by its outstanding performance in image-to-image translation. In most of clinical scenarios, however, it is not practical to create such a large well aligned dataset for Pix2Pix mode. Cycle-consistence mode does not need well aligned dataset but can not meet the high-precision requirements of medical image analysis. Our work aims to provide a general image-to-image translation mode, which not only has no strict requirements on the dataset, but also can meet the clinical requirements in terms of image quality. We foresee positive impacts if the mode is applied to diagnosis in radiology, treatment planning and research.

References

[1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Ghahramani Z, Welling M, Cortes C, et al. Advances in Neural Information Processing Systems. Curran Associates, Inc., 2014.

[2] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[3] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.

[4] Sim B, Oh G, Lim S, et al. Optimal transport, cyclegan, and penalized ls for unsupervised learning in inverse problems[J]. 2019.

[5] Moriakov N, Adler J, Teuwen J. Kernel of cyclegan as a principal homogeneous space[C]// International Conference on Learning Representations. 2020.

[6] Huang L, Zhang C, Zhang H. Self-adaptive training: beyond empirical risk minimization[J]. Advances in Neural Information Processing Systems, 2020, 33.

[7] Patrini G, Rozza A, Krishna Menon A, et al. Making deep neural networks robust to label noise: A loss correction approach[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1944-1952.

[8] Dgani Y, Greenspan H, Goldberger J. Training a neural network based on unreliable human annotation of medical images[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 39-42.

[9] Yao Y, Liu T, Han B, et al. Dual t: Reducing estimation error for transition matrix in label-noise learning[C/OL]//Larochelle H, Ranzato M, Hadsell R, et al. Advances in Neural Information Processing Systems: volume 33. Curran Associates, Inc., 2020: 7260-7271. https://proceedings.neurips.cc/paper/2020/file/512c5cad6c37edb98ae91c8a76c3a291-Paper.pdf.

[10] Zhang L, Tanno R, Xu M C, et al. Disentangling human error from ground truth in segmentation of medical images[C/OL]//Larochelle H, Ranzato M, Hadsell R, et al. Advances in Neural Information Processing Systems: volume 33. Curran Associates, Inc., 2020: 15750-15762. https://proceedings.neurips.cc/paper/2020/file/b5d17ed2b502da15aa727af0d51508d6-Paper.pdf.

[11] wang y, Yu L, van de Weijer J. Deepi2i: Enabling deep hierarchical image-to-image translation by transferring from gans[C/OL]//Larochelle H, Ranzato M, Hadsell R, et al. Advances in Neural Information Processing Systems: volume 33. Curran Associates, Inc., 2020: 11803-11815. https://proceedings.neurips.cc/paper/2020/file/88855547570f7ff053fff7c54e5148cc-Paper.pdf.

[12] Zhang P, Zhang B, Chen D, et al. Cross-domain correspondence learning for exemplar-based image translation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5143-5153.

[13] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//International Conference on Machine Learning. PMLR, 2017: 1857-1865.

[14] Cao Y, Wan X. Divgan: Towards diverse paraphrase generation via diversified generative adversarial network[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings. 2020: 2411-2421.

[15] Choi Y, Choi M, Kim M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8789-8797.

[16] Choi Y, Uh Y, Yoo J, et al. Stargan v2: Diverse image synthesis for multiple domains[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8188-8197.

[17] Chen C, Dou Q, Chen H, et al. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence: volume 33. 2019: 865-872.

[18] Fan J, Cao X, Wang Q, et al. Adversarial learning for mono-or multi-modal registration[J]. Medical image analysis, 2019, 58:101545.

[19] Qin C, Shi B, Liao R, et al. Unsupervised deformable registration for multi-modal images via disentangled representations[C]//International Conference on Information Processing in Medical Imaging. Springer, 2019: 249-261.

[20] Sahiner B, Pezeshk A, Hadjiiski L M, et al. Deep learning in medical imaging and radiation therapy[J]. Medical physics, 2019, 46(1):e1-e36.

[21] Hoffman J, Tzeng E, Park T, et al. Cycada: Cycle-consistent adversarial domain adaptation[C]// International conference on machine learning. PMLR, 2018: 1989-1998.

[22] Li M, Huang H, Ma L, et al. Unsupervised image-to-image translation with stacked cycle-consistent adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 184-199.

[23] Deng W, Zheng L, Ye Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 994-1003.

[24] Li J, Chen E, Ding Z, et al. Cycle-consistent conditional adversarial transfer networks[C]// Proceedings of the 27th ACM International Conference on Multimedia. 2019: 747-755.

[25] Schmidt V, Luccioni A, Mukkavilli S K, et al. Visualizing the consequences of climate change using cycle-consistent adversarial networks[J]. arXiv preprint arXiv:1905.03709, 2019.

[26] Chen Z, Li J, Luo Y, et al. Canzsl: Cycle-consistent adversarial networks for zero-shot learning from natural language[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020: 874-883.

[27] Ren C X, Ziemann A, Theiler J, et al. Cycle-consistent adversarial networks for realistic pervasive change generation in remote sensing imagery[C]//2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI). IEEE, 2020: 42-45.

[28] Liu J, Ding Y, Xiong J, et al. Multi-cycle-consistent adversarial networks for ct image denoising [C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020: 614-618.

[29] Zheng C, Pan L, Wu P. Camu: Cycle-consistent adversarial mapping model for user alignment across social networks[J]. IEEE Transactions on Cybernetics, 2021.

[30] Shah M, Chen X, Rohrbach M, et al. Cycle-consistency for robust visual question answering [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 6649-6658.

[31] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[C]// Proceedings of the European conference on computer vision (ECCV). 2018: 172-189.

[32] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks[C/OL]// Guyon I, Luxburg U V, Bengio S, et al. Advances in Neural Information Processing Systems: volume 30. Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper/2017/file/dc6a6489640ca02b0d42dabeb8e46bb7-Paper.pdf.

[33] Kim J, Kim M, Kang H, et al. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation[C/OL]//International Conference on Learning Representations. 2020. https://openreview.net/forum?id=BJlZ5ySKPH.

[34] Chen R, Huang W, Huang B, et al. Reusing discriminators for encoding: Towards unsupervised image-to-image translation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8168-8177.

[35] S. Sukhbaatar M P L B R F, J. Bruna. Training convolutional networks with noisy labels[C]// International Conference on Learning Representations. 2015.

[36] Goldberger J, Ben-Reuven E. Training deep neural-networks using a noise adaptation layer[J]. 2016.

[37] Xia X, Liu T, Wang N, et al. Are anchor points really indispensable in label-noise learning?[C/OL]//NeurIPS. 2019: 6835-6846. http://papers.nips.cc/paper/8908-are-anchor-points-really-indispensable-in-label-noise-learning.

[38] Yao J, Wu H, Zhang Y, et al. Safeguarded dynamic label regression for noisy supervision[C]// Proceedings of the AAAI Conference on Artificial Intelligence: volume 33. 2019: 9103-9110.

[39] Xiao T, Xia T, Yang Y, et al. Learning from massive noisy labeled data for image classification [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 2691-2699.

[40] Misra I, Lawrence Zitnick C, Mitchell M, et al. Seeing through the human reporting bias: Visual classifiers from noisy human-centric labels[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2930-2939.

[41] Tanno R, Saeedi A, Sankaranarayanan S, et al. Learning from noisy labels by regularized estimation of annotator confusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11244-11253.

[42] Rodrigues F, Pereira F. Deep learning from crowds[C]//Proceedings of the AAAI Conference on Artificial Intelligence: volume 32. 2018.

[43] Branson S, Van Horn G, Perona P. Lean crowdsourcing: Combining humans and machines in an online system[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7474-7483.

[44] Izadinia H, Russell B C, Farhadi A, et al. Deep classifiers from image tags in the wild[C]// Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions. 2015: 13-18.

[45] Jaehwan L, Donggeun Y, Hyo-Eun K. Photometric transformer networks and label adjustment for breast density prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[46] Veit A, Alldrin N, Chechik G, et al. Learning from noisy large-scale datasets with minimal supervision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 839-847.

[47] Tanaka D, Ikami D, Yamasaki T, et al. Joint optimization framework for learning with noisy labels[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5552-5560.

[48] Zheng S, Wu P, Goswami A, et al. Error-bounded correction of noisy labels[C]//International Conference on Machine Learning. PMLR, 2020: 11447-11457.

[49] Han J, Luo P, Wang X. Deep self-learning from noisy labels[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5138-5147.

[50] Durand T, Mehrasa N, Mori G. Learning a deep convnet for multi-label classification with partial labels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 647-657.

[51] Wu X, He R, Sun Z, et al. A light cnn for deep face representation with noisy labels[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11):2884-2896.

[52] Huang J, Qu L, Jia R, et al. O2u-net: A simple noisy label detection approach for deep neural networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3326-3334.

[53] Li J, Socher R, Hoi S C. Dividemix: Learning with noisy labels as semi-supervised learning [C/OL]//International Conference on Learning Representations. 2020. https://openreview.net/forum?id=HJgExaVtwr.

[54] Yan Y, Xu Z, Tsang I, et al. Robust semi-supervised learning through label aggregation[C]//Proceedings of the AAAI Conference on Artificial Intelligence: volume 30. 2016.

[55] Jiang J, Ma J, Wang Z, et al. Hyperspectral image classification in the presence of noisy labels [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(2):851-865.

[56] Li J, Wong Y, Zhao Q, et al. Learning to learn from noisy labeled data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5051-5059.

[57] Li Y, Yang J, Song Y, et al. Learning from noisy labels with distillation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 1910-1918.

[58] Algan G, Ulusoy I. Meta soft label generation for noisy labels[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 7142-7148.

[59] Garcia L P, de Carvalho A C, Lorena A C. Noise detection in the meta-learning level[J]. Neurocomputing, 2016, 176:14-25.

[60] Arar M, Ginger Y, Danon D, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 13410-13419.

[61] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, 2015: 234-241.

[62] Balakrishnan G, Zhao A, Sabuncu M R, et al. Voxelmorph: a learning framework for deformable medical image registration[J]. IEEE transactions on medical imaging, 2019, 38(8):1788-1800.

[63] Menze B H, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (brats)[J]. IEEE transactions on medical imaging, 2014, 34(10):1993-2024.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes]
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] Not applicable
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] Not applicable
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [No] Not applicable
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [No] Not applicable
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No] Not applicable