Joint Registration And Segmentation Of Xray Images Using Generative Adversarial Networks

Dwarikanath Mahapatra[1(✉)], Zongyuan Ge[2], Suman Sedai[1], and Rajib Chakravorty[1]

[1] IBM Research Australia, Melbourne, VIC, Australia
{dwarim,ssedai}@au1.ibm.com, rajib.chakravorty@gmail.com
[2] Monash University, Melbourne, VIC, Australia
zongyuan.ge@monash.edu

Abstract. Medical image registration and segmentation are complementary functions and combining them can improve each other's performance. Conventional deep learning (DL) based approaches tackle the two problems separately without leveraging their mutually beneficial information. We propose a DL based approach for joint registration and segmentation (JRS) of chest Xray images. Generative adversarial networks (GANs) are trained to register a floating image to a reference image by combining their segmentation map similarity with conventional feature maps. Intermediate segmentation maps from the GAN's convolution layers are used in the training stage to generate the final segmentation mask at test time. Experiments on chest Xray images show that JRS gives better registration and segmentation performance than when solving them separately.

1 Introduction

Image registration and segmentation are essential steps of many medical image analysis pipelines. Registration is important for atlas building, correcting deformations and monitoring pathological changes over time. Segmentation is crucial for disease identification, pathology localization and measuring organ function. Accurate segmentation improves registration while accurate registration improves segmentation. Hence a joint registration and segmentation (JRS) framework is expected to improve both over solving them separately. Earlier works combining registration and segmentation have used active contours [17] or Graph cuts [9]. Active contours are iterative, time consuming and may get stuck in local optima, while graph cuts require high computation time. We propose a deep learning (DL) based JRS method that uses generative adversarial networks (GANs) for simultaneous registration and segmentation.

Previous DL based segmentation methods (e.g. brain MRI [13] and lung CT [4]), have used variants of FCN [8] or UNets [12]. DL based approaches for registration have used convolution neural network (CNN) regressors to estimate deformation field [1,10], or combined them with reinforcement learning

Y. Shi et al. (Eds.): MLMI 2018, LNCS 11046, pp. 73–80, 2018.
https://doi.org/10.1007/978-3-030-00919-9_9

[7]. These approaches still use a conventional model to generate the transformed image from the deformation field which increases computation time and does not fully utilize the generative capabilities of DL methods. RegNet [15] and DIR-Net [16] are among the first methods to achieve registration in a single pass but are limited by reliance on spatially corresponding patches to predict transformations. Finding corresponding patches is challenging in low contrast medical images and adversely affects the registration task. Rohe et al. [11] propose SVF-Net trained using reference deformations obtained by registering *previously segmented* regions of interest (ROIs).

Our proposed JRS method is different from existing methods as: (1) we combine registration and segmentation in a single DL framework, which eliminates the need to train a separate segmentation network; (2) registration is driven by segmentation and vice-versa; and '(3) we do not require explicit segmentation of ROIs as in [11], relying instead on segmentation masks generated on the fly from the GAN and use it for registration. We demonstrate its effectiveness for intra-patient lung registration over multiple visits. Our DL approach has the advantage of fast image registration without using conventional time consuming methods, and we outperform DL based registration and segmentation methods, as well as conventional JRS approaches.

2 Methods

In our proposed JRS architecture, the generator network, G, takes three input images: (1) reference image (I^{Ref}), (2) floating image (I^{Flt}) to be registered to I^{Ref}, and (3) I^{Ref}_{Seg}, the segmentation mask of I^{Ref} indicating the organ to be segmented. The outputs of G are: (1) I^{Trans}, the registered image (transformed version of I^{Flt}); (2) I^{Trans}_{Seg}, the segmentation mask of I^{Trans}; and (3) $I^{Def-Recv}$ the recovered deformation field. The discriminator network compares all the three outputs with their corresponding training data to determine if they are real or not. During testing only the generator network is used.

2.1 Joint Registration and Segmentation Using GANs

GANs [3] are generative models trained in an adversarial setting. The generator G outputs a desired image type while a discriminator D outputs a probability of the generated image matching the training data. The training database has chest Xray images and the corresponding masks of the two lungs. To generate training data the images are first translated in the left, right, top or bottom direction with a displacement range of $\pm[25, 40]$ pixels. The translated images are rotated by different angles in the range $\pm[20, 180]^\circ$ at equal steps of 5°. Finally the rotated images are subjected to local elastic deformation using B-splines with the pixel displacements in the range of $\pm[1, 15]$. We denote this deformation field as $I_{Def-App}$, the applied deformation field. The transformations are such that when applied to the corresponding segmentation masks, the Dice Metric (DM) between the original and transformed mask has values less than 0.70. This is done

to ensure that the transformed images are significantly different from the original images and truly test algorithm performance. The original images are I^{Ref} and the transformed images are I^{Flt}. Applying synthetic deformations allows us to: (1) accurately quantify the registration error; and (2) determine the similarity between I^{Trans} and I^{Ref}. G is a feed-forward CNN whose parameters θ_G are,

$$\widehat{\theta} = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{JRS} \left(G_{\theta_G}(I^{Flt}), I^{Ref}, I^{Flt}, I^{Ref}_{Seg} \right), \tag{1}$$

where the loss function l^{JRS} combines content loss (Eq. 2) and adversarial loss (Eq. 3), and $G_{\theta_G}(I^{Flt}) = I^{Trans}$. The content loss is,

$$l_{content}(I^{Trans}, I^{Ref}, I^{Ref}_{Seg}, I^{Trans}_{Seg}) = NMI + [1 - SSIM] + VGG. \tag{2}$$

NMI denotes normalized mutual information between I^{Ref} and $I^{Trans}$and is suitable for multimodal and unimodal deformable registration. $SSIM$ denotes structural similarity index metric (SSIM) based on edge distribution [19] and quantifies landmark correspondence between different images. $SSIM \in [0, 1]$ with higher values indicating greater similarity. VGG is the $L2$ distance between two images using all the multiple feature maps obtained from a pre-trained $VGG16$ network [14]. Note that we extract all the feature maps from all convolution layers of VGG16. This sums up to $64\times2+128\times2+256\times2+512\times3+512\times3 = 3968$ feature maps. The feature maps are of different dimensions due to multiple max pooling steps. Using all feature maps ensures we are comparing information from multiple scales, both coarse and fine, and thus improves robustness. All feature maps are normalized to have values between $[0, 1]$.

2.2 Deformation Field Consistency

CycleGANs [21] learn mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, between image sets $X = I^{Flt}$ and $Y = I^{Ref}$. Adversarial discriminators D_X differentiate between images x and registered images $F(y)$, and D_Y distinguishes between y and $G(x)$. G registers I^{Flt} to I^{Ref} while F registers I^{Ref} to I^{Flt}. Due to space constraints we refer the reader to [21] for details of CyclicGan implementation. In addition to the content loss (Eq. 2) we have: (1) an adversarial loss; and (2) a cycle consistency loss to ensure transformations G, F do not contradict.

The adversarial loss is an important component to ensure that the generated outputs are plausible. In previous works the adversarial loss was based on the similarity of generated image to training data distribution. Since our generator network has three outputs we have additional terms for the adversarial loss. The first term matches the distribution of I^{Trans} to I^{Flt} and is given by:

$$L_{cycGAN}(G, D_Y) = E_{y \in p_{data}(y)} [\log D_Y(y)] + E_{x \in p_{data}(x)} [\log (1 - D_Y(G(x)))], \tag{3}$$

We retain notations X, Y for conciseness. There also exists $L_{cycGAN}(F, D_X)$, the corresponding adversarial loss for F and D_X.

The second component of the adversarial loss incorporates segmentation information by calculating the logarithm of the dice metric (DM) between the generated mask I_{Seg}^{Trans} during each training step, and I_{Seg}^{Ref} the segmentation mask of I^{Ref}. DM is a normalized metric between $[0, 1]$ and acts like a probability measure similar to those in Eq. 3 The third adversarial loss term is the mean square error between $I^{Def-App}$ and $I^{Def-Recv}$, the applied and recovered deformation fields. The final adversarial loss is

$$\begin{aligned} L_{adv} =& L_{cycGAN}(G, D_{I^{Ref}}) + L_{cycGAN}(F, D_{I^{Flt}}) + \log DM(I_{Seg}^{Ref}, I_{Seg}^{Trans}) \\ &+ \log\left(1 - MSE_{Norm}(I^{Def-App}, I^{Def-Recv})\right), \end{aligned} \tag{4}$$

where MSE_{Norm} is the MSE normalized to $[0, 1]$, and $1 - MSE_{Norm}$ ensures that similar deformation fields give a corresponding higher value.

Cycle consistency loss ensures that for each $x \in X$ the reverse deformation should bring x back to the original image, i.e. $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. Similar constraints also apply for mapping F and y. This is achieved using,

$$L_{cyc}(G, F) = E_x \left\| F(G(x)) - x \right\|_1 + E_y \left\| G(F(y)) - y \right\|_1, \tag{5}$$

Thus the full objective function is

$$L(G, F, D_{I^{Flt}}, D_{I^{Ref}}) = L_{adv} + l_{content} + \lambda L_{cyc}(G, F) \tag{6}$$

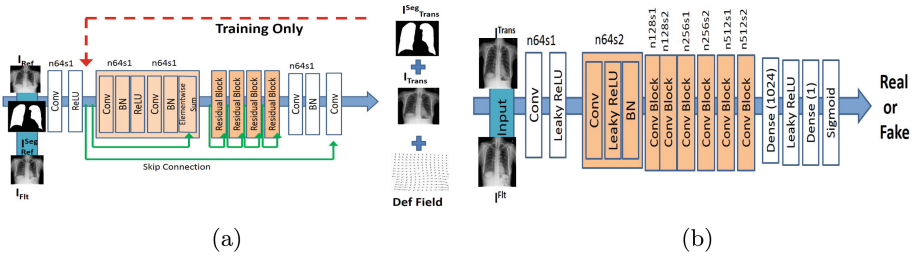
where $\lambda = 10$ controls the contribution of the two objectives. The optimal parameters are given by:

$$G^*, F^* = \arg\min_{F,G} \max_{D_{I^{Flt}}, D_{I^{Ref}}} L(G, F, D_{I^{Flt}}, D_{I^{Ref}}) \tag{7}$$

G (Fig. 1(a)) employs residual blocks having two convolution layers with 3×3 filters and 64 feature maps, followed by batch normalization and ReLU activation. G also outputs the segmentation mask which is fed back for training. F (to ensure cycle consistency) has a similar architecture. The discriminator D (Fig. 1 (b)) determining the similarity between I^{Trans} and I^{Ref} has eight convolutional layers with the kernels increasing by a factor of 2 from 64 to 512 . Leaky ReLU is used and strided convolutions reduce the image dimension when the number of features is doubled. The resulting 512 feature maps are followed by two dense layers and a final sigmoid activation. We do not use max pooling in any layer as we want the input and output images to have the same size.

2.3 Obtaining Segmentation Mask

The segmentation mask is obtained by concatenating the feature maps of different convolution layers which function as activation maps highlighting informative parts of the image [20]. This is similar to the approach by UNet [12] which adds skip connections between corresponding layers of the upsampling and downsampling path to get the final segmentation map. Since our generator network has



(a) (b)

Fig. 1. (a) Generator Network; (b) Discriminator network. $n64s1$ denotes 64 feature maps (n) and stride (s) 1 for each convolutional layer.

no downsampling steps we do not add any skip connections. Instead we take the feature maps of each convolution layer, normalize its values to $[0, 1]$, add them and finally employ Otsu's thresholding to get a segmentation mask. Note that since this mask is generated at each iteration and its similarity with I_{Seg}^{Ref} is being calculated, the feedback is used to update the network weights. Thus, after convergence the segmentation mask thus obtained is an accurate segmentation of the image. We *do not* use a weighted combination similar to [20] because the weights are also being updated.

3 Experiments

Our registration method was tested on the NIH ChestXray14 dataset [18] with $112,120$ frontal-view X-rays from $30K$ patients with 14 disease labels (multiple-labels for each image). Since the original dataset is designed for classification studies, we selected samples and applied the following steps to make it suitable for validating registration experiments.

1. 30 patients each from all the 14 disease classes were selected, giving a total of $14 \times 30 = 420$ different patients. Care was taken to ensure that all the patients had multiple visits (minimum 3 visits and maximum 8 visits.
2. For each set of patient images the left and right lung were manually outlined. We manually annotate corresponding region of disease activity for a particular patient. In some cases there were multiple disease labels for a single patient and each pathology was outlined by the expert. Consequently one image may have multiple labels.
3. In total we had 420 reference images from 420 patients and 1087 floating images (excluding the reference images) across multiple visits of all patients.

Our method was implemented in TensorFlow. We use Adam [5] with $\beta_1 = 0.93$ and batch normalization. The ResNet of G was initialized using mean square error and learning rate of 0.001. Subsequently the final GAN was trained with 10^5 update iterations at learning rate 10^{-3}. Training and test was performed on a NVIDIA Tesla K40 GPU with 12 GB RAM.

We show results for: (1) $JRS-Net$ - our proposed JRS network; (2) JRS_{NoSeg} - registration without using segmentation information; (3) $FlowNet$ - the registration method of [1]; (4) $DIRNet$ the method of [16]; (5) $GC-JRS$ a conventional joint registration and segmentation method using graph cuts ([9]); and (6) Elastix [6]. The following parameter settings were used for Elastix: initial affine transformation and then non rigid registration using normalized mutual information (NMI) as the cost function. Multi grid B-splines were used with spacing of $80, 40, 20, 10, 5$ mm and corresponding downsampling factors being $4, 3, 2, 1, 1$. Average training time for an augmented dataset (rotation and translation) with $98,000$ images is 36 h. Affine registration was applied only for Elastix and not the other methods.

3.1 Results on NIH dataset

The image acquired on the first visit is I^{Ref} and images acquired on subsequent visits are I^{Flt}. I^{Ref}_{Seg} is obtained by manual delineation. For subsequent visits I^{Flt}_{Seg} is obtained by our algorithm. This highlights our JRS algorithm's advantages since the trained model can be applied to different applications using a single manual annotation. The total registration error (TRE) and segmentation overlap measures such as Dice Metric (DM) and 95% Hausdorff Distance (HD_{95}) are calculated before and after registration to quantify each method's efficacy. Intra-patient registration and segmentation results for the lung are summarized in Table 1. We use the UNet trained on the SCR [2] database to segment both lungs from the NIH dataset. The average values for normal images ($DM = 84.9,\ HD = 8.9$) and diseased images ($DM = 84.0,\ HD = 9.3$) is inferior than those reported in Table 1 for $JRS-Net$.

In the example case of patient 5, from day 0 to day 5 had no pathologies in the lung, and hence these 6 images are considered non-diseased. However, infiltration was detected for visits on days $6, 7$ and these images were considered diseased. Figure 2 shows results for non-diseased images where I^{Ref} was day 0 image and I^{Flt} was day 3 image. Figure 3 shows the corresponding results for diseased case where I_{Flt} is from day 6. Superimposed contours (I^{Flt}_{Seg},I^{Ref}_{Seg}) on I^{Ref} (Figs. 2(c), 3(c)) clearly show the difference in lung positions and size on

Table 1. *Intra-patient* image registration results for **left and right lung** using different methods on the NIH-14 database. *Time* indicates computation time in seconds.

	Normal Images							Diseased Images						
	Bef. Reg	After Registration						Bef. Reg	After Registration					
		JRS Net	JRS $_{NoSeg}$	DIR Net	Flow Net	GCJRS	Elastix		JRS Net	JRS $_{NoSeg}$	DIR Net	Flow Net	GCJRS	Elastix
DM(%)	78.9	89.3	85.2	84.8	83.5	85.6	82.1	79.1	88.9	85.0	84.4	83.1	85.2	81.5
HD_{95}(mm)	12.9	6.9	8.4	8.7	9.8	8.0	10.8	11.8	7.3	8.6	8.9	10.1	8.8	11.5
TRE	13.3	7.6	8.9	9.5	10.6	8.9	11.5	12.9	7.9	9.4	9.7	11.0	9.3	12.1
Time(s)		0.5	0.4	0.6	0.5	0.6	21		0.5	0.4	0.6	0.5	53	21

different days due to different acquisition positions. The green and red contours should coincide for ideal registration and results show $JRS-Net$ outperforms all other methods (despite diseased images showing more artificats than normal images) by including segmentation information in the registration task. Segmentation output from UNet is shown in Figs. 2(g), 3(g) using a super imposed yellow contour which demonstrates the superior performance of JRS over conventional segmentation methods.

(a) (b) (c) (d) (e) (f) (g)

Fig. 2. Results for normal lung Xray images from NIH dataset (patient 5). (a) I_{Flt} with I_{Flt}^{Seg} (green); (b) I_{Ref} with I_{Ref}^{Seg} (red) and I_{Flt}^{Seg} before registration; Superimposed registered mask obtained using: (c) $JRS-Net$; (d) $DIR-Net$; (e) $GC-JRS$; (f) Elastix. (g) Segmentation masks of I^{Flt} - manual ground truth (red), $JRS-Net$ (green) and $UNet$ (yellow).

(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Results for diseased lung Xray images from NIH dataset (patient 5). (a) (b) I_{Flt} with I_{Flt}^{Seg} (green); (b) I_{Ref} with I_{Ref}^{Seg} (red) and I_{Flt}^{Seg} before registration; Superimposed registered mask obtained using: (c) $JRS-Net$; (d) $DIR-Net$; (e) $GC-JRS$; (f) $Elastx$; (g) Segmentation masks of I^{Flt} - manual ground truth (red), $JRS-Net$ (green) and $UNet$ (yellow).

4 Conclusion

We have proposed a novel deep learning framework for joint registration and segmentation of lung xray images. Generative adversarial networks are used to register a floating image to a reference image. A simultaneous segmentation of the registered image is achieved by fusing the outputs of the different convolution layers in the GAN. The registration is driven by segmentation information, hence truly integrating registration and segmentation. Experimental results show our joint approach performs better than existing methods that solve registration and segmentation separately. The method's effectiveness is demonstrated on lung xray images of normal and healthy patients with multiple clinical visits.

References

1. Dosovitskiy, A., Fischer, P., et. al.: Flownet: Learning optical flow with convolutional networks. In: Proceedings of IEEE ICCV, pp. 2758–2766 (2015)
2. van Ginneken, B., Stegmann, M., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med. Imag. Anal. **10**(1), 19–40 (2006)
3. Goodfellow, I., et al.: Generative adversarial nets. In: Proceedings of NIPS, pp. 2672–2680 (2014)
4. Harrison, A., Xu, Z., George, K., Lu, L.: Progressive and multi-path holistically nested neural networks for pathological lung segmentation from ct images. In: Proceedings of MICCAI, pp. 621–629 (2017)
5. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: arXiv preprint arXiv:1412.6980 (2014)
6. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: Elastix: a toolbox for intensity based medical image registration. IEEE Trans. Med. Imag. **29**(1), 196–205 (2010)
7. Liao, R., et al.: An artificial agent for robust image registration. In: AAAI, pp. 4168–4175 (2017)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of CVPR, pp. 3431–3440 (2015)
9. Mahapatra, D., Sun, Y.: Joint registration and segmentation of dynamic cardiac perfusion images using mrfs. In: Proceedings of MICCAI, pp. 493–501 (2010)
10. Miao, S., Z.J. Wang, Y.Z., Liao, R.: Real-time 2d/3d registration via cnn regression. In: IEEE ISBI, pp. 1430–1434 (2016)
11. Rohe, M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: Learning deformable image registration using shape matching. In: Proceedings of MICCAI, pp. 266–274 (2017)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of MICCAI, pp. 234–241 (2015)
13. Shen, H., Wang, R., Zhang, J., McKenna, S.: Boundary-aware fully convolutional network for brain tumor segmentation. In: Proceedings of MICCAI, pp. 433–441 (2017)
14. Simonyan, K., Zisserman., A.: Very deep convolutional networks for large-scale image recognition. CoRR absarXiv:1409.1556 (2014)
15. Sokooti, H., de Vos, B., Berendsen, F., Lelieveldt, B., Isgum, I., Staring, M., et al.: Nonrigid image registration using multiscale 3d convolutional neural networks. In: MICCAI, pp. 232–239 (2017)
16. de Vos, B., Berendsen, F., Viergever, M., Staring, M., Isgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: arXiv preprint arXiv:1704.06065 (2017)
17. Wang, F., Vemuri, B., Eisenschenk, S.: Joint registration and segmentation of neuroanatomic structures from brain mri. J Acad. Radiol. **12**(9), 1104–1111 (2006)
18. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R., et al.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of CVPR (2017)
19. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Imag. Proc. **13**(4), 600–612 (2004)
20. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of CVPR, pp. 2921–2929 (2016)
21. Zhu, J., park, T., Isola, P., Efros, A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: arXiv preprint arXiv:1703.10593 (2017)