Registration of Pathological Images

Xiao Yang[1]([✉]), Xu Han[1], Eunbyung Park[1], Stephen Aylward[3],
Roland Kwitt[4], and Marc Niethammer[1,2]

[1] UNC Chapel Hill, Chapel Hill, USA
`xy@cs.unc.edu`
[2] Biomedical Research Imaging Center, Chapel Hill, USA
[3] Kitware, Inc., Carrboro, USA
[4] Department of Computer Science, University of Salzburg, Salzburg, Austria

Abstract. This paper proposes an approach to improve atlas-to-image
registration accuracy with large pathologies. Instead of directly register-
ing an atlas to a pathological image, the method learns a mapping from
the pathological image to a quasi-normal image, for which more accurate
registration is possible. Specifically, the method uses a deep variational
convolutional encoder-decoder network to learn the mapping. Further-
more, the method estimates local mapping uncertainty through network
inference statistics and uses those estimates to down-weight the image
registration similarity measure in areas of high uncertainty. The perfor-
mance of the method is quantified using synthetic brain tumor images
and images from the brain tumor segmentation challenge (BRATS 2015).

1 Introduction

Atlas-to-image registration provides spatial information to map anatomical loca-
tions from an atlas to a patient. This procedure is crucial for atlas-based seg-
mentation which is used in lesion detection and treatment planning for trau-
matic brain injury, tumor and stroke cases [7]. However, large brain pathologies
often produce appearance changes which may result in large misregistrations,
if appearance-mismatch is falsely accounted for by image deformation. This is
especially acute for deformable image registration methods, which are needed to
capture subtle deformations and, for example, mass effects of tumors.

Several approaches have been proposed for atlas-to-image registration[1] with
large pathologies. The most straight-forward method is *cost function masking*,
where the lesion area is not considered during image similarity computation [1].
However, this method could be problematic if the lesion area contains important
brain structure information. Other methods include joint segmentation and regis-
tration that mitigates missing correspondences [3], explicit tumor growth mod-
eling [5], geometric metamorphosis that separates the deformation of healthy

---

[1] Such approaches, as well as our proposed approach, are of course also applicable
to general image-to-image registration. We use atlas-to-image registration as our
motivating application here.

brain areas from lesion changes [14], and registration methods accounting for deformation and intensity changes [21].

While effective, these methods require either explicit lesion segmentation, knowledge of lesion location, or the modeling of tumor growth. Two alternatives exist: (1) using a robust cost-function [17] or a mutual saliency map [15] to mitigate the effect of outliers or, instead, (2) learning desired mappings between image types from large-scale image databases. We follow this second approach. A learned mapping then allows synthesizing one image type from another. Image synthesis has been extensively explored to synthesize MR imaging sequences [8], to facilitate multi-modality registration [2,19] and to segment lesions [18]. Our goal is to synthesize quasi-normal images from images with lesions to simplify atlas-to-lesion-image registration. Using image synthesis rather than a robust cost-function or a mutual saliency map allows reconstructing *structural* information to guide registration even in highly pathological areas.

Liu et al. [10] proposed a low-rank-plus-sparse (LRS) technique to synthesize quasi-normal brain images from pathological images and to simultaneously estimate a quasi-normal atlas. This approach decomposes images into normal (low-rank) and lesion (sparse) parts. The low-rank part then constitutes the synthesized quasi-normal images, effectively removing lesion effects. By *learning* from data, no prior lesion information is required. However, the LRS decomposition itself requires good image alignment, hence decomposition and registration have to be interleaved to obtain good results.

Contributions. Our contributions to improve atlas-to-image registration can be summarized as follows: *First*, similar to [10], we propose a method to directly map a pathology image to a synthesized *quasi-normal* image to simplify the registration problem. No registration is needed in this process. *Second*, we use a deep variational encoder-decoder network to learn this mapping and train it using stochastic gradient variational Bayes [9]. *Third*, since the normal appearance of pathological tissue is unknown per se, we propose loss-function masking and pathology-like "structured noise" to train our model. These strategies ignore mappings between image regions without known correspondence, and artificially create areas with known correspondence which can be used for training, respectively. *Fourth*, based on the variational formulation, we estimate the reconstruction uncertainty of the predicted quasi-normal image and use it to adjust/improve the image similarity measure so that it focuses more on matching areas of low uncertainty. *Fifth*, we validate our approach on synthetic tumor[2] images and data from the BRATS 2015 challenge. Our framework requires no prior knowledge of lesion location (at test time; lesion segmentations are required during training only) and provides comparable or, in many cases, better registration accuracy than the LRS method and cost function masking.

Organization. Section 2 discusses variational Bayes for autoencoders, as well as its denoising criterion. Section 3 introduces our methods to remove brain lesions

---

[2] In this paper we use brain tumors as example pathologies; however, our approach is applicable to other pathologies.

from images and to compute uncertainty estimates for the prediction of quasi-normal images. Section 4 presents experimental results (for 2D synthetic and real data), discusses extensions to 3D, and possible improvements.

2   Denoising Variational Autoencoding

The problem of mapping a pathology image to a quasi-normal image is similar to the objective of a denoising autoencoder, which aims to transform a noisy image into a noise-free image. Next, we introduce variational inference for autoencoders, followed by an explanation of inference for a denoising autoencoder.

Given a clean brain image $\boldsymbol{x}$ and the latent variable $\boldsymbol{z}$, we want to find the posterior distribution $p(\boldsymbol{z}|\boldsymbol{x})$. Since $p(\boldsymbol{z}|\boldsymbol{x})$ is intractable, we approximate it with a tractable distribution $q_\phi(\boldsymbol{z}|\boldsymbol{x})$, where ϕ is the parameter of the variational approximation. For a variational autoencoder, the posterior distribution is $p_\theta(\boldsymbol{z}|\boldsymbol{x}) \propto p_\theta(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})$, where the prior $p(\boldsymbol{z})$ is usually an isotropic Gaussian, and θ are the parameters of the observation model $p_\theta(\boldsymbol{x}|\boldsymbol{z})$. When mapping these parameters to an autoencoder, $\boldsymbol{z}$ corresponds to the hidden layer, $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ refers to the *encoding* operation and $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ refers to *decoding*. Thus, ϕ and θ correspond to the weights in the encoder and decoder.

To approximate the true posterior with the variational posterior, we minimize the Kullback-Leibler (KL) divergence between these two distributions.

$$
\begin{aligned}
D_{\mathrm{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}|\boldsymbol{x})) &= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{q_\phi(\boldsymbol{z}|\boldsymbol{x})}{p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&= \log p_\theta(\boldsymbol{x}) - \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{z},\boldsymbol{x})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right].
\end{aligned}
\tag{1}
$$

Since the data $\boldsymbol{x}$ is independent of the latent variable $\boldsymbol{z}$, $\log p_\theta(\boldsymbol{x})$ in Eq. (1) is constant. Thus, minimizing the KL-divergence is equivalent to maximizing the term $\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}(\log p_\theta(\boldsymbol{z},\boldsymbol{x}) - \log q_\phi(\boldsymbol{z}|\boldsymbol{x}))$. Since the KL-divergence is non-negative, we have $\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{z},\boldsymbol{x}) - \log q_\phi(\boldsymbol{z}|\boldsymbol{x})] \le \log p_\theta(\boldsymbol{x})$, and we call this term the variational lower bound of the data likelihood $\mathcal{L}_{\mathrm{VAE}}$, i.e.,

$$
\begin{aligned}
\mathcal{L}_{\mathrm{VAE}} &= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{z},\boldsymbol{x})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&= -D_{\mathrm{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})) + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})],
\end{aligned}
\tag{2}
$$

where the first term can be regarded as the regularizer, matching the variational posterior to the prior of the latent variable, and the second term is the expected network output likelihood w.r.t. the variational posterior $q_\phi(\boldsymbol{z}|\boldsymbol{x})$. During training, the optimization algorithm maximizes this variational lower bound.

Our goal is a denoising autoencoder for pathology-removal. In other words, we regard lesions as a special structured noise. Removing lesion appearance is then equivalent to removing noise in the denoising autoencoder theory. To do this, we introduce the input noise (lesion) corruption distribution as $p(\widetilde{\boldsymbol{x}}|\boldsymbol{x})$.

The variational posterior distribution is then $\widetilde{q}_\phi(\boldsymbol{z}|\boldsymbol{x}) = \int q_\phi(\boldsymbol{z}|\widetilde{\boldsymbol{x}})p(\widetilde{\boldsymbol{x}}|\boldsymbol{x})d\widetilde{\boldsymbol{x}}$. If the original variational posterior distribution is a Gaussian, this new posterior can be regarded as a mixture of Gaussians, which has better representation power. As shown in [6], the variational lower bound for a denoising autoencoder is

$$\mathcal{L}_{\text{DVAE}} = \mathbb{E}_{\widetilde{q}_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{z}, \boldsymbol{x})}{q_\phi(\boldsymbol{z}|\widetilde{\boldsymbol{x}})}\right] \geq \mathcal{L}_{\text{VAE}} = \mathbb{E}_{\widetilde{q}_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{z}, \boldsymbol{x})}{\widetilde{q}_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]. \quad (3)$$

This means that the denoising variational lower bound is higher than the original one, leading to a smaller KL-divergence between the true and the approximated posterior. In the following section, we discuss our implementation of the encoder-decoder network and how to maximize the denoising variational lower bound.

3   Network Model and Registration with Uncertainty

Figure 1 shows the structure of our denoising variational encoder-decoder network. The input is a brain image $\mathbf{x}$ with intensities normalized to $[0, 1]$. The encoder network consists of convolution followed by max-pooling layers (`ConvPool`), and the decoder has max-unpooling layers followed by convolution (`UnpoolConv`). We choose max-unpooling instead of upsampling as the unpooling operation, because upsampling ignores the pooling location for each pooling patch, which results in severe image degradation. The encoder and decoder are connected by fully connected layers (`FC`) and the re-parameterization layer (`Reparam`) [9]. This layer takes the parameters for the variational posterior as input, which in our case is the mean μ and standard deviation Σ of the Gaussian distribution, and generates a sampled value from the variational posterior. This enables us to compute the gradient of the regularizer $-D_{\text{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}))$ for ϕ using the variational parameters instead of the sampled value, which is not differentiable for ϕ. Below we discuss specific techniques implemented for our task.

Training Normal Brain Appearance Using Pathology Images. A model of normal brain appearance would ideally be learned from a large number of healthy brain images with a consistent imaging protocol. Our goal, instead, is to learn a mapping from a pathological image to a quasi-normal image, i.e., train a denoising autoencoder for the lesion 'noise', and maximize the denoising variational lower bound. This poses two challenges: *first*, in general, we do not know what the normal appearance in a pathological area should be; *second*, pathological images may exhibit spatial deformations not seen in a normal subject population (such as the mass effect for brain tumors). To mitigate these problems, we learn the brain appearance from the *normal areas of the pathological brain images* only. This can be accomplished by (1) introducing lesion-like structured noise (i.e., circles filled with the mean intensity of the normal brain area for brain tumor cases) via the `QuasiLesion` layer in Fig. 1, and (2) *loss function masking*, i.e., ignoring lesion-areas during learning. Suppose we have the lesion segmentation for the training data. For loss-function masking, we first change the
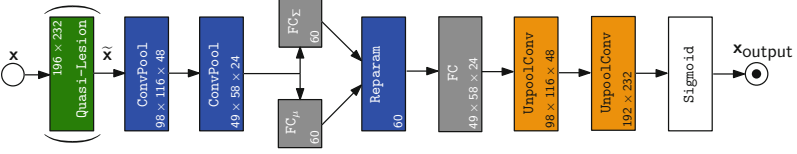
Fig. 1. Network structure (numbers indicate the data size).

input with structured noise $\widetilde{x}$ to $\widetilde{x}_{\text{normal}}$ using the following rule: if $\widetilde{x} \in$ Normal, then $\widetilde{x}_{\text{normal}} = \widetilde{x}$; otherwise, (i.e., $\widetilde{x} \in$ Lesion) $\widetilde{x}_{\text{normal}} = a + \mathcal{N}(0, \sigma)$. This prevents the network from using tumor-appearance. Experiments show only small differences for different settings of a and σ. However, performance suffers when σ is too high, and setting $a = 0$ increases the mean intensity error for the whole image. In our model, we set a to the mean intensity value of the normal area and $\sigma = 0.03$. Second, we set our network output likelihood for x_{output} to

$$\log p_\theta(x_{\text{output}}|z)_{\text{normal}} = \begin{cases} |x_{\text{output}} - x|, & x_{\text{output}} \in \text{Normal} \\ 0, & x_{\text{output}} \in \text{Lesion}. \end{cases} \tag{4}$$

Hence, we disregard any errors in the lesion area during backpropagation. We refer to this two-step strategy as *loss-function masking*.

The overall training procedure for our network is: (1) sample one corrupted input $\widetilde{x}$ from $p(\widetilde{x}|x)$, (2) mask out the lesion area to get $\widetilde{x}_{\text{normal}}$, (3) sample one z from $q_\phi(z|\widetilde{x}_{\text{normal}})$ and obtain a reconstructed image x_{output} from the network, (4) calculate the denoising lower bound $\mathcal{L}_{\text{DVAE}}$ with the change in Eq. (4) and (5) perform stochastic gradient descent backpropagation to update the network.

Reconstruction Uncertainty for Atlas Registration. During testing, due to the small amount of data available and the possibly large appearance differences among training cases, it is useful to utilize the uncertainty of the reconstructed image to guide registration. In our case, we sample z from the approximated posterior $q_\phi(z|x)$ to generate multiple reconstruction images x_{output} with different z. Then, we choose the mean of the sampled images $\mu_{x_{\text{output}}}$ as the reconstruction result, and the (local) standard deviation $\sigma_{x_{\text{output}}}$ as uncertainty measure. We define areas of high uncertainty as those areas with large variance, and, for registration, our method down-weights the contribution of those areas to the image similarity measure. We simply use $w(x_{\text{output}}) = \exp(-\sigma^2_{x_{\text{output}}} \times 2000)$ as a local weight for the image similarity measure in our experiments[3]. This function ensures that the weight drops to near 0 for a large standard deviation. Note that this is different from cost function/pathology masking. Cost function masking uses a simple binary mask, which is equivalent to setting the weight of the lesion area to zero. Our uncertainty-based weighting, on the other hand, *downweights* ambiguous areas in the reconstruction process which may not be

---

[3] Other, potentially better choices are of course possible.

highly reliable for registration. Our uncertainty weight is in $[0, 1]$. Hence, structural information is rarely discarded completely as in cost-function masking. Our experimental results in Sect. 4 show that this is indeed desirable.

4   Experiments and Discussion

We evaluate our model in two experiments: one using 2D synthetic images, and one with real BRATS tumor images. The image intensity range is $[0, 1]$. We implement the network with Torch and use the *rmsprop* [20] optimization algorithm; we set the learning rate to 0.0001, the momentum decay to 0.1 and the update decay to 0.01. Further, we use a batch size of 16, and for a training dataset with 500 images of size 196×232, training 1000 epochs takes about 10 h on a 2012 Nvidia Titan GPU. For data augmentation, we apply random shifting up to 10 pixels in both directions for a training image and add zero-mean Gaussian noise with standard deviation of 0.01. During testing, we sample 100 images for each test case, and calculate their mean and standard deviation. All images for training and testing are extracted from the same slice of their original 3D images, which are pre-aligned to a 3D ICBM T1 atlas [4] using affine registration and judged to be limited to having in-plane deformations. We use `NiftyReg` [13] (with standard settings) together with normalized cross correlation (NCC) to register the 2D ICBM atlas slice to the reconstructed result. Note that we modified `NiftyReg` to integrate image uncertainty into the cost function. We used a large number of B-spline control points (19×23 for a 196×232 image). This ensures that displacements large enough to capture the mass effect observed in the BRATS data can be expressed. B-spline registration approaches similar to `NiftyReg` have successfully been used for registrations of various difficulty [16]; and given sufficient degrees of freedom poor registration performance is likely due to an unsuitable similarity measure, which should be investigated in future work. To capture even larger deformations, `NiftyReg` could easily be replaced by a fluid-based registration approach. The focus here is to *synthesize* quasi-normal images and to exploit them and their associated reconstruction uncertainty for registration. For our images, 1 pixel corresponds to 1mm $\times$ 1mm.

For comparison, we use the LRS method, which is an alternative approach to image synthesis for tumor images. We select the parameters maximizing $2 \times NCC_{\text{tumor}} + NCC_{\text{normal}}$ for the training data. Due to high computational cost of current LRS approaches [10], we use 50 training images for each case. Furthermore, to demonstrate that using synthesized images in fact improves registration accuracy, we also compare our method against using the reconstruction uncertainty map in combination with the original tumor image for registration.

Synthetic Tumor Experiment. We use 436 brain images from the OASIS [11] cross-sectional dataset as base images. This chosen dataset is a mix of 43 % Alzheimer's and 57 % control subjects. We create a synthetic tumor dataset by registering random OASIS images to random BRATS 2015 T1c images (to account for the mass effect of tumors) with tumor area masking, followed by pasting the BRATS' tumor regions into the OASIS images. We generate 500 training
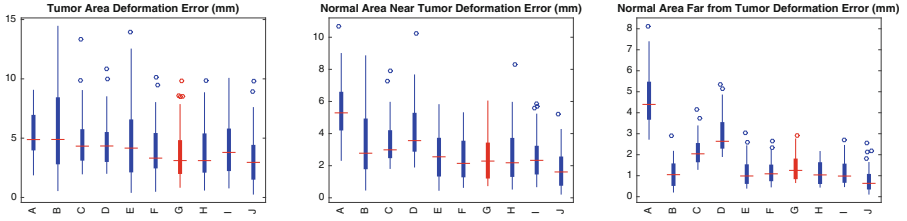
Fig. 2. *Mean deformation error* of all synthetic tumor test cases for various models. Our model is highlighted in red. Masking tumor area = MT. Add structured noise = ASN. Use uncertainty for registration = UR. *(A):* affine registration; *(B):* register to tumor image; *(C):* low-rank-sparse (LRS) with registration; *(D):* LRS w/o registration; *(E):* MT, no ASN, no UR; *(F):* MT, ASN, no UR; *(G):* MT, ASN, UR; *(H):* network trained with clean images, ASN; *(I):* Use uncertainty on tumor image directly; *(J):* cost function masking. (Color figure online)
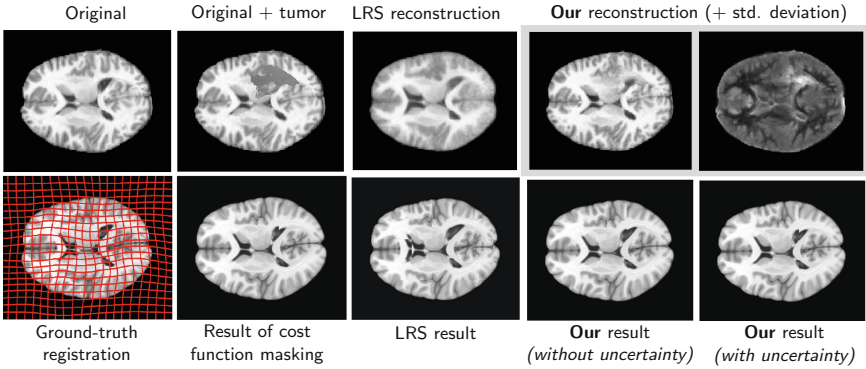
Fig. 3. Exemplary synthetic tumor test case reconstruction and checkerboard comparison with ground truth registration. Best viewed zoomed-in.

and 50 testing images using separate OASIS and BRATS images. Figure 2 shows boxplots of mean deformation errors of different areas per test case, with respect to the ground truth deformation obtained by registering the atlas to the normal image (i.e., without added tumor). The highlighted boxplot is the network model trained with tumor images, added quasi-tumor (i.e., structured noise) and using uncertainty weighting for the registration. We evaluate the deformation error for three areas: (1) the tumor areas, (2) normal areas within 10 mm from the tumor boundary (*near tumor*) and (3) normal areas more than 10 mm away from the boundary (*far from tumor*). By evaluating all three areas we can assess how well the mass effect is captured. This is generally only meaningful for our synthetic experiment. Landmarks (outside the tumor area) are more suitable for real data. For the tumor areas, our method (MT+ASN+UR) outperforms most other methods. For the normal areas, the registration difference between our method and directly registering to the original tumor image is very small, especially

compared with the LRS method which tends to remove fine details. Compared to using the tumor image directly for registration, our model decreases the 99.7 % upper limit of the mean of the tumor area deformation error from 14.43 mm to 7.83 mm, the mean error from 5.62 mm to 3.60 mm, and the standard deviation from 3.49 mm to 2.16 mm. The significantly decreased deformation error only causes a small increase of mean deformation error for the normal area, from 1.36 mm to 1.49 mm. The only method performing better than our model for this *synthetic* test is cost function masking, which *requires* tumor segmentation. Figure 3 shows one example test case. Notice that the LRS method erroneously reconstructs the upper lateral ventricle, resulting in a wrong deformation.

BRATS Experiment. We also evaluate our network using the BRATS 2015 training dataset [12], which contains 274 images. This is a very challenging dataset due to moderate sample size and high variations in image appearance and acquisition. We use cross-validation, and partition the dataset into 4 sets of 244 training images and 30 testing images, resulting in a total of 120 test cases. For preprocessing, we standardize image appearance using adaptive histogram equalization. For evaluation, we manually label, on average, 10 landmarks per case around the tumor area and at major anatomical structures for the test images. We report the target registration error for the landmarks in Table 1. Our method still outperforms most methods, including LRS without registration. Although, the difference of our model and LRS+registration is not statistically significant, the figures in combination with our synthetic results suggest that our method is overall preferable. Note also that LRS requires image registrations for *each* decomposition iteration and introduces blurring to the brain's normal area (see Fig. 4), while our method does not suffer from these problems. Moreover, it is interesting to see that cost function masking performs worse than our method. This could be explained by the observation that in cases where the tumor is very large, cost function masking hides too much of the brain structure, making registration inaccurate. Figure 4 shows one exemplary BRATS test case. Because the tumor covers the majority of the white matter in the left hemisphere, cost function masking removes too much information from the registration. As a result, the left lateral ventricle is misregistered. Combining our network reconstructed image and uncertainty information, our registration result is much better.

Modeling Quasi-tumor Appearance. One interesting problem is the choice of quasi-tumor appearance. In our work we use the mean normal brain area intensity as the appearance, while other choices, such as using simulated[4] tumor appearance or random noise, are also sensible. To show the effect of quasi-tumor appearance choice on the registration result, we conduct additional experiments using 4 textures to create quasi-tumors: (1) real tumors of the BRATS dataset, (2) mean intensity (our approach), (3) random constant intensities and (4) random noise. Registration performance for all 4 methods is similar, with (2) having lower registration error in normal areas (e.g. median of 1.07/2.78 mm compared to 1.28/2.84 mm using (1) for synthetic/BRATS data). A possible reason why

---

[4] Real tumor appearance is not known in such areas.

Table 1. Statistics for landmark errors over the BRATS test cases. The best results in each category are marked in **bold**.

|	Mean error [mm]					Max. error [mm]				
Data Percentile	99.7 %	75 %	50 %	25 %	0.3 %	99.7 %	75 %	50 %	25 %	0.3 %
`Affine (baseline)`	11.29	6.90	5.06	3.72	2.28	19.25	12.61	9.83	7.04	3.52
Use tumor image directly	6.32	4.20	3.29	2.77	**1.54**	20.48	12.47	7.52	5.22	**1.95**
Cost function masking	6.12	4.22	3.12	2.65	1.89	21.00	11.75	6.98	4.97	2.65
LRS+registration	5.26	3.77	3.06	2.74	1.88	12.04	8.20	6.30	5.45	3.54
LRS, no registration	6.15	4.30	3.25	2.79	2.12	14.62	9.61	6.83	5.72	4.05
Tumor image+uncertainty	5.52	3.79	3.08	2.65	1.81	13.91	8.76	6.24	**4.95**	2.63
Our model (no uncertainty)	5.08	3.63	**2.98**	2.64	1.66	12.79	8.12	6.21	4.96	2.82
Our model (with uncertainty)	**4.74**	**3.52**	3.02	**2.61**	1.83	**11.77**	**7.99**	**5.93**	5.08	2.48

using tumor appearance is not superior is the limited training data available (∼200 images). For a larger dataset with more tumor appearance examples to learn from, using tumor appearance could potentially be a better choice.

Discussion. One interesting finding in our work is that while a high-quality lesion area reconstruction is desirable, it is not necessary to improve atlas registration. Lesion reconstruction may be affected by many factors (limited data, large image appearance variance, etc.), but the atlas registration result depends on the quasi-normal reconstruction of the lesion *and* the faithful reconstruction of the normal tissue. For example, in some cases the LRS method achieves visually pleasing results in the lesion area. However, at the same time it smoothes out the normal area losing important details for image registration. Our method on the other hand preserves details in the normal areas more consistently and hence results in overall better registration accuracy. Moreover, for tightly controlled data (e.g., a synthetic dataset) our method generates better reconstructions for the lesion area. Thus, future experiments using more controlled data (e.g. BRATS 2012 synthetic images) would be interesting. Besides, synthesizing a quasi-normal image generates useful structural information that can help guide the registration, and reconstruction uncertainty can be used to focus the registration on regions of high confidence.

Another interesting question is how to extend our approach to 3D images. In initial experiments, we implemented a 2.5D network which reconstructs 14 slices at once. Training the network on 500 2.5D training cases takes 3 days, which, while not fast, is feasible. One possible approach is to learn mappings for 3D patches using patch location as additional feature, which would enable us to train on a much larger dataset (patches) at a reasonable computational cost.

Finally, designing a more "lesion-like" noise model and exploring the impact of training data size on the predictions are interesting directions to explore.
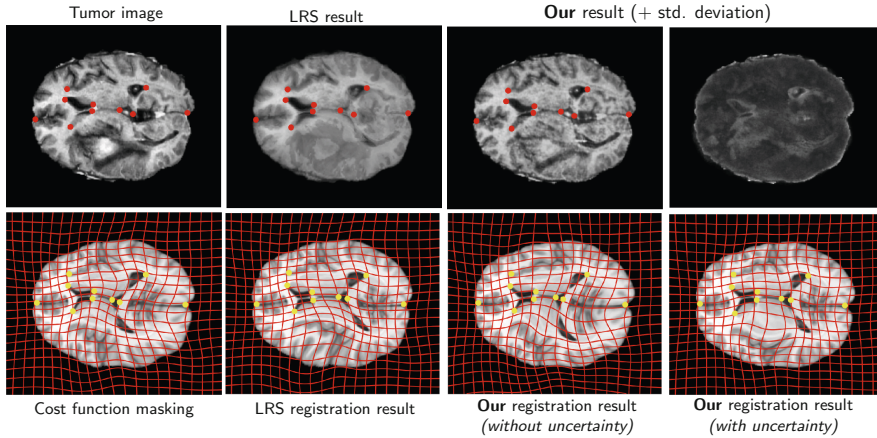
Fig. 4. Exemplary BRATS test case with landmarks for test image (*top row*) and warped atlas (*bottom row*).

References

1. Brett, M., Leff, A.P., Rorden, C., Ashburner, J.: Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage **14**(2), 486–500 (2001)
2. Cao, T., Zach, C., Modla, S., Powell, D., Czymmek, K., Niethammer, M.: Multimodal registration for correlative microscopy using image analogies. MedIA **18**(6), 914–926 (2014)
3. Chitphakdithai, N., Duncan, J.S.: Non-rigid registration with missing correspondences in preoperative and postresection brain images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 367–374. Springer, Heidelberg (2010)
4. Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L.: Unbiased average age-appropriate atlases for pediatric studies. Neuroimage **54**(1), 313–327 (2011)
5. Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: GLISTR: glioma image segmentation and registration. TMI **31**(10), 1941–1954 (2012)
6. Im, D.J., Ahn, S., Memisevic, R., Bengio, Y.: Denoising criterion for variational auto-encoding framework. CoRR abs/1511.06406 (2015)
7. Irimia, A., Wang, B., Aylward, S.R., Prastawa, M.W., Pace, D.F., Gerig, G., Hovda, D.A., Kikinis, R., Vespa, P.M., Horn, J.D.V.: Neuroimaging of structural pathology and connectomics in traumatic brain injury: toward personalized outcome prediction. Neuroimage Clin. **1**(1), 1–17 (2012)
8. Jog, A., Roy, S., Carass, A., Prince, J.L.: Magnetic resonance image synthesis through patch regression. In: ISBI, pp. 350–353 (2013)
9. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. CoRR abs/1312.6114 (2013)
10. Liu, X., Niethammer, M., Kwitt, R., Singh, N., McCormick, M., Aylward, S.: Low-rank atlas image analyses in the presence of pathologies. TMI **34**(12), 2583–2591 (2015)

11. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. **19**(9), 1498–1507 (2007)
12. Menze, B.H., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (BRATS). TMI **34**(10), 1993–2024 (2015)
13. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Comput. Methods Prog. Biomed. **98**(3), 278–284 (2010)
14. Niethammer, M., Hart, G.L., Pace, D.F., Vespa, P.M., Irimia, A., Van Horn, J.D., Aylward, S.R.: Geometric metamorphosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 639–646. Springer, Heidelberg (2011)
15. Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C.: DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. MedIA **15**(4), 622–639 (2011)
16. Ou, Y., Akbari, H., Bilello, M., Da, X., Davatzikos, C.: Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights. TMI **33**(10), 2039–2065 (2014)
17. Reuter, M., Rosas, D.H., Fischl, B.: Highly accurate inverse consistent registration: a robust approach. Neuroimage **53**(4), 1181–1196 (2010)
18. Roy, S., Carass, A., Shiee, N., Pham, D.L., Prince, J.L.: MR contrast synthesis for lesion segmentation. In: ISBI, pp. 932–935 (2010)
19. Roy, S., Carass, A., Jog, A., Prince, J.L., Lee, J.: MR to CT registration of brains using image synthesis. SPIE Med. Imaging **9034**, 903419 (2014)
20. Tieleman, T., Hinton, G.: Lecture 6.5-RMSprop, COURSERA: neural networks for machine learning (2012)
21. Wang, B., Prastawa, M., Awate, S., Irimia, A., Chambers, M., Vespa, P., Horn, J.V., Gerig, G.: Segmentation of serial MRI of TBI patients using personalized atlas construction and topological change estimation. In: ISBI, pp. 1152–1155 (2012)