Joint Synthesis and Registration of MRI and Cone-Beam CT Images Using Deep Evidential Uncertainty Estimation of Deformation Fields

M. Yi,[a] R. Duan,[b] Z. Li,[c] J. H. Siewerdsen,[a,d] A. Uneri,[a] J. Lee,[e] C. K. Jones [c]

[a] Department of Biomedical Engineering, Johns Hopkins University, Baltimore MD
[b] Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore MD
[c] Department of Computer Science, Johns Hopkins University, Baltimore MD
[d] Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston TX
[e] Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University, Baltimore MD

ABSTRACT

Purpose. To accurately register preoperative MRI and intraoperative cone-beam CT (CBCT) in neurosurgery using a joint synthesis and registration network with evidential uncertainty estimation, so surgeons can better interpret the registration quality, and decide whether to use the registered MRI or CBCT in a certain region.

Methods. The prediction of the deformation fields can be formulated as a regression problem, and the output of the network follows a Gaussian distribution, whose mean and variance follow a Normal Inverse Gamma (NIG). According to evidential deep learning, the parameters of the NIG distributions are produced by the network as extra layers of output(denoted as EDL parameters) and the uncertainty can be expressed from them. A loss function based on likelihood forces the network to learn the EDL parameters and an accurate prediction concurrently.

Results. The visualization of the uncertainty map illustrates that the higher magnitude of warping field has a corresponding higher uncertainty in the last layer in the region around the ventricle. We also use numeric results to compare the network performance with and without uncertainty, and proves that the network trained with uncertainty shows better improvement in registration performance than uncertainty-free training.

Conclusions. The preliminary results show that incorporating uncertainty estimation into registration model leads to improved registration performance. Furthermore, the estimated uncertainty map is also useful for surgeons to evaluate the reliability of registered images. By extending the model to predict uncertainty of the entire network, i.e., not freezing any layer, we anticipate that the registration performance can be improved further.

Keywords: Registration, Uncertainty, Evidential Deep Learning

1. DESCRIPTION OF PURPOSE

Magnetic resonance imaging (MRI) serves as the established imaging modality for diagnosing brain-related diseases and planning neurosurgical interventions. Intraoperative MRI imaging, however, is not commonly available in the operating room (OR). Conversely, cone-beam CT (CBCT) is widely used for intraoperative guidance of neurosurgical procedures but lacks the necessary soft tissue image contrast for visualizing target anatomy and surrounding cerebral structures. Registration of MRI to CBCT images can circumvent these limitations by mapping the various soft tissue structures, and the accompanying surgical plan from the preoperative to the up-to-date intraoperative context.

Aside from the rigid cranium, most structures within the brain are subject to deformation from the time of preoperative imaging and during the course of surgery (e.g., due to brain shift or direct surgical manipulation). Deep learning models such as VoxelMorph[1] and JSR[2], are promising for image registration, but for surgical applications requiring high confidence along with high accuracy, an uncertainty map marking the reliability of the output is a good desire. As a result, the estimation of uncertainty in registration is an important topic, as it presents bounds and assurances of registration method, and further indicates a region with high in uncertainty should involve human interpretation with the CBCT. While uncertainty estimation has found success in tasks like medical image segmentation[3], its application in image registration remains largely unexplored. Moreover, the computational demands of real-time uncertainty estimation place constraints on the efficiency of multiple inferences or repeated models, such as Monte Carlo dropout[4] and ensemble[5], thus impacting their usability in fast-paced surgical environments.

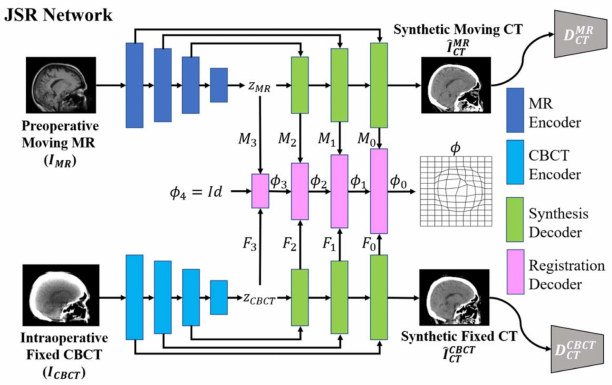

Figure 1. Network architecture of JSR[3] for MRI-CBCT registration.

Our work addresses these challenges by leveraging evidential deep learning within a joint synthesis and registration network (JSR) for MRI-CBCT image registration. Evidential uncertainty estimation employs a two-level distributional assumption to predict uncertainty in one inference. The incorporation of uncertainty analysis alongside image registration improves registration accuracy and provides surgeons with uncertainty feature maps that can aid in assessing the reliability of the registered images.

2. METHODS

2.1 Joint Synthesis and Registration (JSR) Network

Previous work[2] details JSR for registration of preoperative MR images to intraoperative CBCT to provide an up-to-date view of the patient's intraoperative anatomy following deformation during the surgery. As illustrated in Figure 1, the JSR network consists of 3

main parts: MR synthesis, CT synthesis, and deformation field decoding. The two synthesis paths encode the input MR and CBCT images into the same intermediate domain, effectively converting the multi-modal registration problem into a mono-modal one. Both synthesis paths consist of an encoder to map the input images into a latent space, and a decoder to an intermediate domain, i.e., CT. The latent encoder and synthetic decoder follow the structure of a U-Net with skip connections. Simultaneous with the decoding, the feature maps at each layer are input into the multi-resolution deformation fields decoder. The synthesis paths are trained using generative adversarial networks (GAN), where a generator and a discriminator are trained to compete with each other to produce realistic synthetic images. The registration decoder follows an unsupervised training process and uses the similarity metrics between the intermediate modalities, the CT generated from CBCT, and the registered version of synthesized CT from MRI.

2.2 Evidential Deep Learning

Evidential deep learning (EDL)[7] is a technique for direct estimation of predictive uncertainty in deep learning models. EDL learns a higher-order evidential distribution and treats learning as a process of evidence acquisition. Unlike other uncertainty estimation methods, EDL does not require additional computation or modification of the model architecture, thus it is applicable to many complex models. In the regression setting, typically each sample consists of input x and target y. EDL considers the prediction target y as a continuous random variable following a normal distribution, with a mean and variance jointly following a normal inverse gamma (NIG) distribution:

$$y \sim N(\mu, \sigma^2)$$
$$(\mu, \sigma^2) \sim NIG(\gamma, \nu, \alpha, \beta)$$

where $\gamma \in \mathbb{R}, \nu > 0, \alpha > 1, \beta > 0$. To convert the regression neural network to an evidential model, one needs to replace the output layer. For each neuron in the output layer that predicts a y, replace it with 4 outputs to predict the 4 corresponding NIG parameters. The trained evidential model is employed to infer prediction, aleatoric, and epistemic uncertainties associated with any given data x:

$$\mathrm{E}(\mu) = \gamma, \mathrm{E}(\sigma^2) = \underbrace{\frac{\beta}{(\alpha - 1)}}_{aleatoric}, Var(\mu) = \underbrace{\frac{\beta}{\nu(\alpha - 1)}}_{epistemic}$$

2.3 Evidential Deep Learning in JSR

In the proposed work, we adopt EDL into the uncertainty estimation of the deformation fields, by regarding the output value at each pixel as a regression problem. As a result, the extra outputs α, β, ν are all the same size of the final deformation fields (3×128×160×128). Since the uncertainty of the deformation fields in the JSR network is influenced by many factors in the propagation, to better understand the contribution of each layer in the uncertainty, we take the steps of solving simplified problems first. The first step is to freeze all the parameters in the network except the last layer, and then gradually release more parameters. Figure 2 illustrates this process where the light blue blocks indicate the frozen layers, and an extra block in the last layer is for the estimation for the EDL parameters.

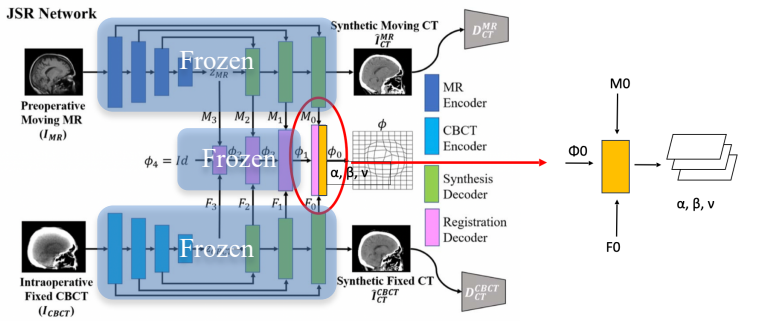

Figure 2. The illustration of the simplified uncertainty estimation problem, where M0 and F0 are feature maps given by synthesis decoders. α, β, ν represent the EDL parameters, and $\phi_1$represents the coarse registration map produced by last layer in registration decoder.

2.4 Loss Functions

The uncertainty loss function in the last layer is composed of the unsupervised registration loss and the EDL uncertainty loss. The unsupervised registration loss is given by:

$$L_{\text{reg}} = L_{\text{intramodal}} + \lambda_{\text{intermodal}} L_{\text{intermodal}} \quad (1)$$

$L_{\text{intramodal}}$ assesses the accuracy of deformation field estimation in the intermediate domain of all four resolution levels. It is determined by calculating the normalized cross-correlation between the CT synthetic images derived from CBCT and the registered CT images synthesized from MR data at each level. The $L_{\text{intermodal}}$ also takes the CT images generated from CBCT, but measures cross-modality similarity with MR, to compensate for the inconsistency caused by synthesis. L_{smooth} forces the output warping fields at each level to be smooth and is defined as the L2 norm of the deformation fields. The expressions of the terms are:

$$L_{\text{intramodal}} = -\sum_{i=0}^{3} \text{NCC}(\text{STN}\left(\emptyset_i, \text{down } 2^i(\hat{I}_{CT}^{MR})\right), \text{down } 2^i(\hat{I}_{CT}^{CBCT})) \quad (2)$$

$$L_{\text{intermodal}} = \sum_{i=0}^{3} |\text{MIND}(\text{STN}\left(\emptyset_i, \text{down } 2^i(I_{MR})\right) - \text{MIND}(\text{down } 2^i(\hat{I}_{CT}^{CBCT}))| \quad (3)$$

$$L_{smooth} = \sum_{i=0}^{3} \| \nabla \emptyset_i \|^2 \quad (4)$$

where STN is the Spatial Transformer Network, and down 2^i denotes the downsampling operation, with $\times 2^i$ smaller in resolution. The uncertainty loss is given by:

$$L_{\text{uncertainty}} = L_{\text{NLL}} + \lambda_{\text{penalty}} L_{\text{penalty}} \quad (5)$$

where L_{NLL} is the negative log likelihood of the network prediction, following the distribution defined by uncertainty parameters α, β, ν. L_{penalty} is introduced to assign high uncertainty to the pixels with great values of absolute error, defined as:

$$L_{NLL} = \frac{1}{2}\log\left(\frac{\pi}{\nu}\right) - \alpha\log\big(2\beta(1+\nu)\big) + \left(\alpha+\frac{1}{2}\right)\log((gt_i-\gamma)^2\,\nu + 2\beta(1+\nu)) + \log\left(\frac{\Gamma(\alpha)}{\Gamma(\alpha+\frac{1}{2})}\right) \tag{6}$$

$$L_{\text{penalty}} = |gt_i - \gamma|(2\nu + \alpha) \tag{7}$$

where gt_i is the ground truth of the warping field, which can be attained from simulated data; γ is the deformation fields prediction, and α, β, ν are outputs of the final layer of the registration network. The final loss function for the training is denoted as:

$$L_{\text{total}} = L_{\text{reg}} + \lambda_{\text{uncertainty}}\, L_{\text{uncertainty}} \tag{8}$$

The hyperparameters $\lambda_{\text{uncertainty}}, \lambda_{\text{penalty}}, \lambda_{\text{intermodel}}, \lambda_{\text{smoothness}}$ are set to balance the assessment of performance based on image similarity and on distance from ground truth; to assign importance to the relation between error and uncertainty; to compensate the information loss caused by domain transformation, and to emphasize the smoothness of the warping field respectively.

2.5 Data simulation

The data used for network training was generated from 50 sets of co-registered CBCT and MRI pairs. A random transform matrix was generated to stimulate a rigid transformation after reshaping images into the size of 128×160×128, followed by a non-rigid transformation simulated according to an inverse power-law with distance:

$$\phi_{sim} = e^{\sum_{i=1}^{N} \frac{\alpha_i}{|s_i - x|^{\beta_i}}(s_i - x)}$$

where $\boldsymbol{s}$ is a set of N random anchor points, $\boldsymbol{x}$ is the location of a pixel, α, β are sets of deformation parameters corresponding to the anchor points and represents the deformation magnitude and decay power respectively. To include randomness, elements from α, β are randomly selected from a range instead of set as a fixed number. The random deformation process is cast 8 times on each of the 50 pairs, resulting in 400 sets of deformed MR and CBCT pairs. Then, each pair was processed by rigid registration, and the 400 rigid-registered MR and CBCT pairs compose of the training set. The same processing was applied to the validation data set (N = 10).

2.6 Experiment Setup

In the experiment, the gradients are back propogated through the weight of the last layer in the registration decoder only, whereas the weights of the JSR network are initialized with a set of pre-trained weights using unsupervised registration loss. The hyperparameters in the loss function are: $\lambda_{\text{uncertainty}} = 0.1, \lambda_{\text{penalty}} = 2, \lambda_{\text{intermodel}} = 0.1, \lambda_{\text{smoothness}} = 1$. The uncertainty layer is trained on NVIDIA RTX 6000 GPU, CUDA 12.2, with a learning rate of 0.0001, and Adam optimizer for 100 epochs.

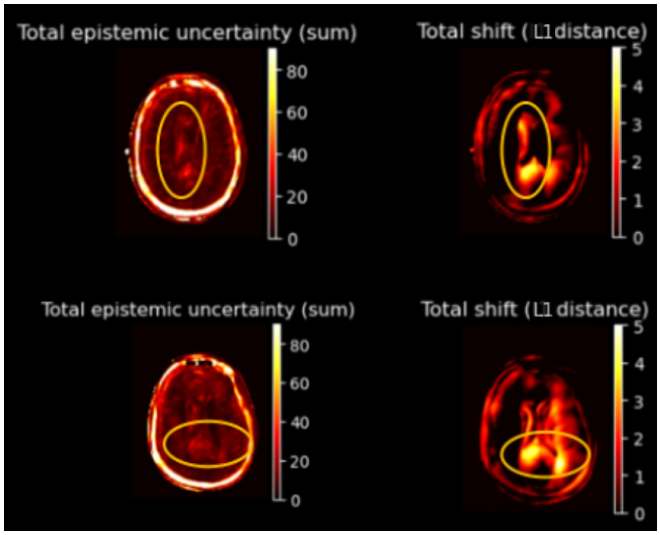

Figure 3. The correspondance between uncertainty map and deformation fields.

3. RESULTS

3.1 Interpretability of Uncertainty Maps

Because the uncertainty propagated by the frozen parameters cannot be reduced, the output uncertainty has relatively high value. However, under current experiment setting we still see the correspondence between the output deformation fields and the uncertainty maps, as demonstrated in Figure 3, where sum denotes the total uncertainty of x, y and z axes. The yellow circle emphasizes the area with corresponding high contrast, mostly around the ventricle, which shows that in these areas the output uncertainty is closely related to the deformation fields.

3.2 Quantitative Analysis of Network performance

In this section, we demonstrate the results supporting the benefit of the uncertainty-aware loss for network training. We computed Dice similarity coefficient (DSC), Hausdorff Distance (HD), and Surface Distance (SD) to quantitatively assess the similarity between the segmentation maps of the registered MR and the ground truth. The ground truth segmentation was obtained from the MR images aligning with input CBCT (deformation-free MRI), as a result the segmentation is well aligned with CBCT. The output registration field is supposed to improve the alignment between CBCT and MR, thus the segmentations. So the segmentation of deformation-free MR is set as the ground truth, to evaluate the metrics on the segmentation of registered MR images. We mainly discuss the metrics on 7 brain regions: 3rd ventricle, 4th ventricle, lateral ventricle, amygdala, hippocampus, thalamus and caudate. The segmented brain anatomy for an instance is presented in Figure 4.

The last layer of the network was trained with only unsupervised registration loss in (1) and the combination total loss in (8) respectively, so that we can compare the results of the training with and without uncertainty estimation. The results are as shown in Table 1. JSR_c is the network initialized with checkpoint parameters, which is a set of pretrained parameters of the whole JSR network. JSR_r is the network with the last layer trained with unsupervised registration loss in (1), while JSR_u is the network with the last layer trained with

total loss in (8). JSR_u produces the best result. Although the magnitude of the increase is small because only the parameters in the last layer were unfrozen, the result is promising. We see that compared with JSR_r, the improvement is much greater in DSC (0.14%/0.04%), HD (0.049/0.025) and in SD (0.004/0.001).

Table 1. Quantitative evaluation of registration (bold indicates the best result)

		3rd Ventricle	4th Ventricle	Lateral Ventricle	Amygdala	Hippocampus	Thalamus	Caudate	Overall
JSR_c	DSC (↑)	0.492	0.580	0.750	0.556	0.590	0.759	0.612	0.620
	HD (↓)	3.326	4.315	4.406	3.565	**3.720**	4.072	**3.892**	3.894
	SD (↓)	0.759	0.666	0.683	0.724	0.737	0.782	0.658	0.716
JSR_r	DSC (↑)	**0.492**	0.582	0.750	0.556	0.590	0.760	0.612	0.620
	HD (↓)	3.358	4.315	**4.337**	3.563	**3.720**	3.879	3.909	3.869
	SD (↓)	0.757	0.665	0.684	0.725	0.736	0.779	0.657	0.715
JSR_u	DSC (↑)	0.491	**0.584**	**0.751**	**0.557**	**0.592**	**0.761**	**0.613**	**0.621**
	HD (↓)	**3.256**	**4.231**	4.356	**3.563**	3.742	**3.855**	3.909	**3.845**
	SD (↓)	**0.755**	**0.659**	**0.681**	**0.723**	**0.733**	**0.774**	**0.655**	**0.711**

4. NEW OR BREAKTHROUGH WORK

The preliminary experiment showed promising results, however, the uncertainty propagated in the synthesis paths is not utilized. Future work will involve further training of the synthesis path for uncertainty-aware optimization. Furthermore, since the JSR network was trained with unsupervised loss function, another focus area will be towards an unsupervised uncertainty prediction method resulting in deformation fields with lower uncertainty, and improved registration accuracy.

5. CONCLUSIONS

In this paper, we present a method of introducing uncertainty into a joint synthesis and registration network. Interpreting the prediction of warping field at each pixel as a regression problem, we applied evidential deep learning for uncertainty estimation. The method is promising, with fine-tuning the network weights resulting in a refined uncertainty prediction and EDL parameters. The results show that the network can produce an uncertainty map, along with an improved registration accuracy evaluated using multiple metrics. The work highlights a promising step towards uncertainty-aware unsupervised MR-CBCT registration networks.

REFERENCES

[1] Balakrishnan, et al., "VoxelMorph: A learning framework for Deformable Medical Image Registration," IEEE Trans. Med. Imag. 38(8), 1788–1800 (2019).

[2] Han, R., et al., "Joint synthesis and registration network for deformable MRI-CBCT image registration for Neurosurgical guidance," Phys. Med. & Biol. 67(12), 125008 (2022).

[3] Jones, C. K., et. al.., "Direct quantification of epistemic and aleatoric uncertainty in 3D U-net segmentation," Jour. of Med. Imag. 9(03) (2022).

[4] Gal, Y., Ghahramani, Z., "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," Int. Conf. on Mach. Lear. (2016).

[5] Lakshminarayanan, B., et. al., "Simple and scalable predictive uncertainty estimation using deep ensembles," Adv. in Neur. Inf. Proc. Sys. 30(2017).

[6] Amini, A., et al., " Deep evidential regression," Adv. in Neur. Inf. Proc. Sys. 33, 14927-14937 (2020)