Dense Deformation Network for High Resolution Tissue Cleared Image Registration

Abdullah Nazib, *Member, IEEE,* Clinton Fookes, *Fellow, IEEE,* and Dimitri Perrin, *Fellow, IEEE*

***Abstract*—The recent application of Deep Learning in various areas of medical image analysis has brought excellent performance gain. The application of deep learning technologies in medical image registration successfully outperformed traditional optimization based registration algorithms both in registration time and accuracy. In this paper, we present a densely connected convolutional architecture for deformable image registration. The training of the network is unsupervised and does not require ground-truth deformation or any synthetic deformation as a label. The proposed architecture is trained and tested on two different version of tissue cleared data, 10% and 25% resolution of high resolution dataset respectively and demonstrated comparable registration performance with the state-of-the-art ANTS registration method. The proposed method is also compared with the deep-learning based Voxelmorph registration method. Due to the memory limitation, original voxelmorph can work at most 15% resolution of Tissue cleared data. For rigorous experimental comparison we developed a patch-based version of Voxelmorph network, and trained it on 10% and 25% resolution. In both resolution, proposed DenseDeformation network outperformed Voxelmorph in registration accuracy.**

***Index Terms*—Registration,High Resolution,Deep Learning, Dense Connection.**

I. Introduction

TRADITIONALLY image registration problem is addressed as optimization problem that maximizes similarity between images until satisfactory parameters are achieved. For a deformable registration problem the number of parameters are high and computationally expensive. The number of parameters and their computation expanse increase with the resolution and dimension of the images. For high resolution images, a deformable image registration algorithm takes very long runtime which makes them inapplicable for quick analysis of biological data in the processing pipeline. Recently, a biochemical process named Tissue clearing is emerged that can remove light obstructing elements from soft-tissues and enable biologist to take images with very high resolution. The pixel spacing of these images are extremely small. Thus, the 3D images obtained from tissue clearing are significantly large and contain cellular level information that are necessary for biological studies. Compared to a MRI human brain volume, a simple tissue cleared mouse(CUBIC dataset in our case) brain is 1000 times larger and takes 6 gigabytes of space. In terms of resolution, tissue cleared volumes are in micrometer scale (6.456.4510m3) whereas MRI volumes are in millimeter scale(0.860.861.5mm3). Therefore, the registration of these images through conventional registration methods is computationally very expensive [1]. Recent application of deep-learning in image registration [2], [3],[4] gives promising way to address this issue. Deep learning based registration methods are computationally very efficient and have satisfactory registration accuracy. Despite their high efficiency, training a deep-learning registration method requires large amount of data. In our case, having large amount of tissue cleared data (similar to [2]) is practically impossible and very expensive. Moreover, training of a deep-learning based registration method on high resolution tissue cleared data requires large memory and other computational resources that is not feasible. In summary, registering high resolution tissue cleared data using deep-learning requires following criterion to be met:

- The learning framework should be trainable with small amount of data.
- The learning framework should have the ability to take high resolution images with limited computational resources.
- It should be scale-able in resolution.
- It should not be dependent on specific reference image, therefore, any arbitrary image pair can be registered without further training.

To address these issues, we propose a patch-based densely connected image registration network. The proposed network generates deformation in two steps between two dense blocks. The dense blocks are densely connected convolutional layers able to capture complex features from the tissue cleared data. Thus the generated deformation fields can capture small deformations in the cellular structures and give better results than auto-encoder based learning frameworks. In an auto-encoder based architecture, the successive down-scaling removes the global information from the training patches. The proposed network generates a deformation field from the dense block without down-scaling and hence be able to capture global deformation more accurately. The deformation in the tissue cleared images are small and smooth. To capture the small deformation, the second dense-block extract convolutional features from the down-sampled feature bank obtained from first dense block. Hence, the a dense deformation field is obtained and the network is named as DenseDeformation network. The proposed network is scale-able in resolution and trainable with limited computational resource.

II. Related Works

Deep learning has been extensively used and investigated in medical image analysis. Despite the success of deep learning in medical image segmentation, its applicability in image registration task is not properly investigated. Only a handful of literature indicates the larger scope of using deep learning

QUT High Performance Computing facility

in registration. Based on the literature found, there are two prevalent approach [5] to use deep learning based image registration:

A. Deep Learning for Similarity Measure

This approach uses deep networks for similarity measures only. In [6] first proposed deep learning based multi-modal registration approach. In [6], two stacked auto-encoder (SAE) networks are proposed to measure similarity between moving and fixed image patches. The SAEs are pre-trained in unsupervised manner and fine tuned by a prediction layer on top of the SAE. The prediction layers determine the similarity between the image patches. In their [6] approach the similarity measure requires a pre-training approach and the authors did not present the effectiveness of their similarity metric in a registration framework. In [7] proposed a similar strategy with CNN. Here, deep CNN network is an integral part of a registration framework. After re-sampling the moving image by registration process, the re-sampled moving image and fixed image patches are fed into the CNN architecture. The CNN then outputs a dissimilarity map, representing receptive field of each input patch. The derivative of dissimilarity map is then directly used for parameter optimization. Despite their [7] success in integrating deep architecture in a continuous registration framework, the CNN architecture does not guarantee the differentiability of the dissimilarity derivative. A more robust similarity based approach is proposed by [8]. In [8] a deep auto-encoder based registration framework scalable to different modality is proposed. The auto-encoder used in their work is a CNN network which has encoder and decoder. The encoder learns the low dimensional representation of high dimensional 3D patches and outputs only 1D representation vectors. The decoder reconstructs 3D patches again from the 1D representation vectors. The reconstructed output from the decoder is used as a feedback to the encoder to optimize encoding for more accurate representation. A simple feature based image registration framework named HAMMER is then used for registration using low dimensional representation vectors from encoder. To reduce performance overhead, an importance sampling is used before the encoding by CNN to select the most discriminative image patches from the training sample. The significant performance improvement makes this method comparable to renowned Demon method [9].

B. Deep Learning for Parameter Estimation

This is a different strategy to use deep learning in image registration. By this approach, deep learning is directly employed to estimate transformation parameters. Miao et al. [10] first proposed a CNN regression parameter estimator for 2D-3D registration. In their method, the rigid transformation parameters are partitioned. The DRR and X-ray image is divided into different region based on transformation parameters and residual features are calculated. A set of CNN networks are used to predict and update of transformation parameters taking DRR and X-ray ROIs as input. This approach shows significant performance improvement in different dataset. The problem of this method is the number of CNN regressors increases with the image dimension and running multiple CNN regressors is extremely resource hungry. Furthermore, this approach is not suitable for deformable registration as parameter space is very large. To capture deformation by deep learning, in another approach [11] considers large deformation diffeomorphic metric mapping as their baseline. In [11], a momentum parameter is initialized for each pixel and evolved in time. A CNN based deep architecture is used to predict the momentum maps. This method gives significant performance boost for both 2D (around 1600x) and for 3D (around 66x) registration. Despite its accuracy, this method requires pre-registration of training dataset to generate deformation fields on which training of the CNN network is dependent. To make registration process independent from other baseline methods Sokooti et al. [12] proposed a to train registration network with randomly generated deformation fields as label. This method provides a clever way to circumvent pre-registration of training dataset with a conventional registration algorithm but require a large set of random deformation for training. Using random deformation as label data does not guarantee the invertibility and smoothness of the deformation fields. Producing large set of ground-truth/label data for a large dataset using an optimization based algorithm gives satisfactory performance but is a cumbersome task while creating large set of synthetic deformation is easy but has performance drawback. In [13] proposed segmentation based ground-truth/label data generation where a segmented region of interest from each training image is registered to a template image giving a set of correspondence points. Deformation fields are then generated by interpolating these correspondence points in a grid. Thus this method can generate deformation field between any image pairs in the training data. Still its dependence on an optimization algorithm makes it vulnerable to the pros and cones of the optimization algorithm used. To make learning based registration algorithm independent, [14] used a fully convolutional structure that directly estimates spatial deformation between fixed and moving image by optimizing similarity and regularization loss. The network gives loss in different spatial resolution and combine them all to optimize with back-propagation. The training strategy is similar to conventional registration approach and independent of ground-truth deformation label. The first learning based registration method that is independent and also achieved state-of-the art performance in different dataset is proposed in [15]. This method combines a U-net FCN architecture with a spatial transformation network for interpolation and is optimized by cross-correlation loss between fixed and interpolated moving image. Rigorous training and testing on eight different MRI dataset with around eight thousand 3D brain images make this method a viable competitor of conventional state-of-the-arts. The high performance and accuracy of this network still does not guarantee the inverse compatibility of the deformation field. Dependency on voxel similarity metric and very slow convergence in training further bolsters the requirement of new method development.

III. Tissue Cleared Data Acquisition

In this work, the proposed network is trained and tested using tissue cleared images of Arc-Devenus mouse brain. The process of clearing tissue of mouse brain is a completely bio-chemical process and out of scope of this work. Following is the brief discussion of the process involved in data acquisition and data pre-processing:

A. Tissue Clearing

There are many tissue clearing methods have been developed. Among them, popular methods include BABB[16], Scale [17], SeeDb [18], CLARITY [19] and iDISCO [20] etc. In this paper we used data from a tissue clearing method named CUBIC protocol. In this method data is acquired in three steps: tissue clearing by chemical process, image acquisition, and image analysis (including registration). The chemical process can be performed in three different ways: i) Simple immersion protocol for whole-organ clearing, ii) CB-perfusion and immersion protocol for faster clearing for whole organs, and iii) CB-perfusion protocol for whole-body clearing [21],[18]. Based on the protocols applied, it takes 11 to 14 days to clear an organ or whole body.

B. Data Acquisition

We briefly describe data acquisition process here, detailed process is explained in [22]. After tissue clearing, a Light-Sheet Fluorescence Microscope (LSFM) with CMOS camera of resolution 2560×2160 pixels and a pair of laser excitation and emission filters of wavelength 488nm for green signal and 588nm for red signal are used altogether for rapid 3D imaging. The acquired images have zoom capacity up to 6.5×6.5 micrometer, which is sufficient to detect signals from a single cell. For whole brain sample with clearer and sharper 3D image volume, images of the mouse brain are taken in two opposite directions, Dorstal-to-Ventral (D-V) and Ventral-to-Dorsal(V-D) respectively from same brain. Thus the LSFM imaging procedure takes four 2D stacks for each brain: two signals (one for cell positioning and one for cell activity), and two directions.

C. Data Preparation

The obtained V-D and D-V stacks, mentioned in previous section, are registered for alignment (using the cell positioning data) and then merged together. Before this alignment, the data has to be down-scaled to 25% of original size due to software limitation *(no existing software tool can register the high-resolution images of 100% size)*. The merging of two stacks is done by calculating edge content, and this edge information is used to determine which slices should be taken from which imaging direction. For each brain sample this procedure generates one structure volume and one signal volume in Nifti format. In our experimentation, we use structural brain files for each brain sample as input to the proposed network. Using same procedure of down sampling and merging we generate a 10% version of the tissue cleared data to have an idea how down sampling affect the registration performance of the proposed deep-net architecture in this two resolution scale.

IV. Method

A. Dense Deformation Network

Considering the problem of large memory requirement for high-resolution (25% as example) image registration by deep learning method, in this paper we propose a patch based deep learning architecture that is trainable in unsupervised manner. The proposed network has dense connection blocks and generate deformation fields in two stage unlike other deep-learning based registration methods most of which employs U-net architectures for deformation generation. Unlike the most deep-learning based image registration architecture, proposed architecture is consists of two dense-block as shown in figure(to be added). This type of architecture is also used in patch based segmentation in [23] Each of these dense block consists of a number of unites. The unites are densely connected with each other and each of these units consists of a LeakyRELU activated batch-normalization layer, a convolution layer. The network takes concatenated source and target image patches as input and pass them through the first dense block without down sampling. Then a transformation block is used to down-sample the features. A deformation field is generated from first dense block and is defined as global flow since it contains more contextual information. After the transformation block, another dense-block is employed from which another deformation field is generated from these more refined feature banks and we termed it local flow. At the end of the network, both of these flows are concatenated to obtain final deformation field in x,y and z direction. The source image then warped with the combined flow by a dense STN.

B. Loss function

Similar to [2] The network is trained with cross-correlation loss and is regularized with diffusion regularization.

$$\mathcal{L}_{Loss}(src, tgt, \phi) = \mathcal{L}_{similarity} + \lambda \mathcal{L}_{smooth} \tag{1}$$

where

$$\mathcal{L}_{similarity} = -CC(src \circ \phi, tgt) \tag{2}$$

and

$$\mathcal{L}_{smooth}(\phi) = \sum_{n} \nabla(\phi(p)) \tag{3}$$

V. Experiments

A. Training Data Generation

To train the network, we use 10% and 25% resolution images of CUBIC brains. The spatial dimension of at this two resolution are $256 \times 216 \times 68$ and $640 \times 540 \times 169$ respectively. For 10% resolution a region of size $184 \times 216 \times 32$ is cropped and for 25% $540 \times 540 \times 169$ is cropped. We have 20 brains for training and 3 brains for testing (in both resolution). To train the network with these 20 brains we have 20×19 pairs of source and target. Before extracting the training patches, we select brain "003" from test dataset as reference to all brains and perform affine registration to make all brains in same rigid space. Intensity normalization (between 0 to 1) is applied to to all data before extracting the patches. To extract training

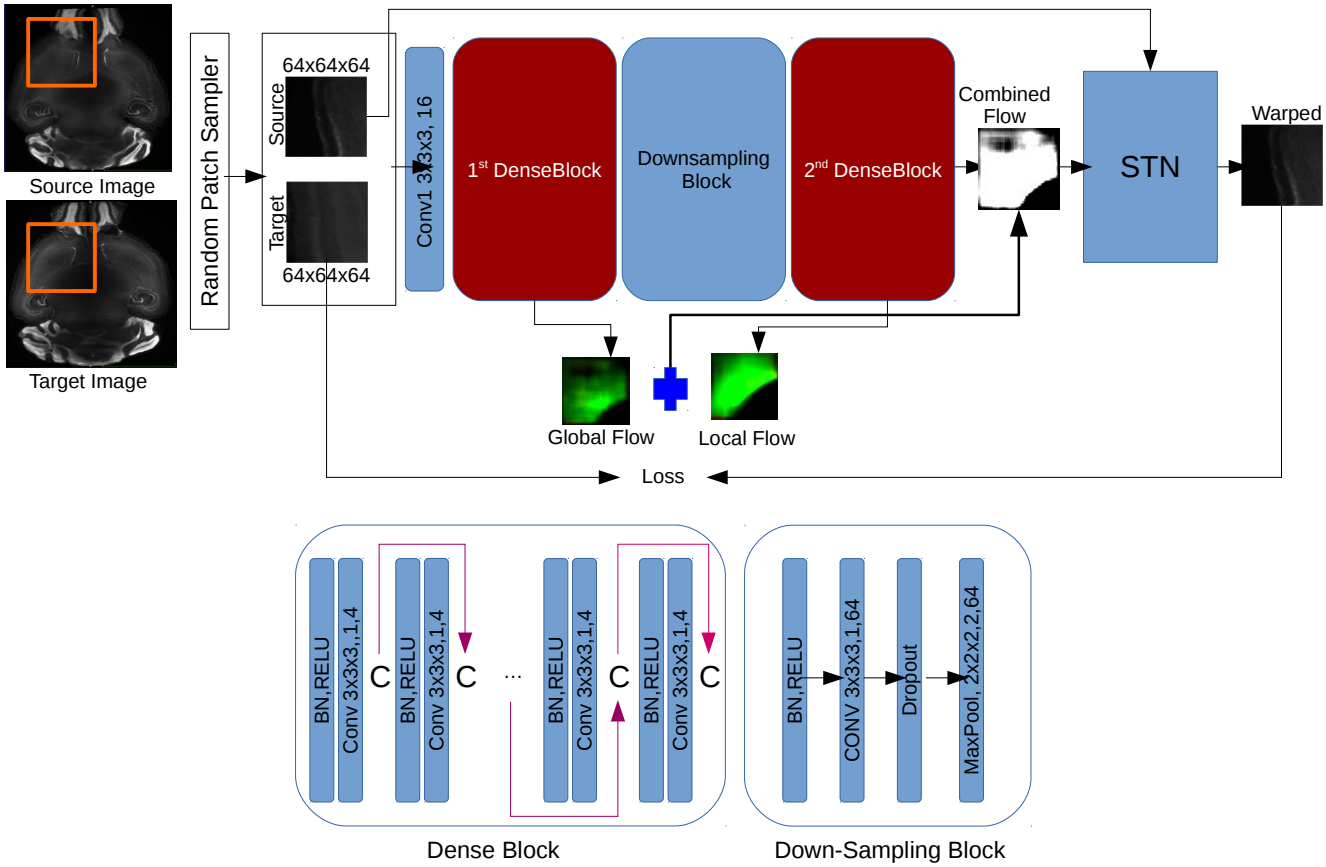


Fig. 1: Architecture of the proposed Dense Deformation Network

patches, each source and target images are selected randomly out of 20 brains. From each pair, 2500 patches are randomly picked. To ensure patches to be informative and discriminative, we apply canny edge detection on both source and target images with upper threshold 0.5 and lower threshold 0.02 and select patches from same location of the edge detected binary version of both source and target. The location where the average intensity of the binary patches are more than 0.1(for 10%) and 0.2(for 25%) are selected as the discriminative patch pairs. Thus 950000 patch-pairs are collected for training the network.

B. Evaluation Metric

The high resolution dataset we are using from CUBIC protocol does not have any ground truth data to validate the registration performance of the proposed network. Generating ground-truth with other registration tools is a time consuming task. Most image registration methods are evaluated based on dice similarity which require hand annotated segmentation data. To avoid these technical difficulties, cross-correlation and mutual information matrices are selected as evaluation metric. To further validate the registration accuracy, a known deformation is applied on test images and each deformed version is then registered against their non-deformed original version. Apart from the quantitative measure we also present a rigorous qualitative evaluation in section VI-A3 and VI-B1.

C. Competing Methods

For the comparison of the proposed DDN method, we select four traditional optimization based method and one learning based method. The details of the settings of the selected tools are given below:

1) ANTS: [24] stands for Advanced normalization tools and use symmetric diffeomorphic normalization for non-rigid registration. For our evaluation, we used cross-correlation as similarity measure in 4 different resolution level and 100 iterations at each level.

2) Elastix: [25] is a large library of different components of image registration. In our experiment, we use parameter settings from [26] for rigid, affine an deformable registration

3) NiftyReg: [27] is also a high performing registration tool. We use exactly same settings as [28] except the number of iterations and intensity threshold. 100 iterations are used for non-rigid registration with free-form deformation model and 500 as intensity threshold.

4) IRTK: [29] is one of the early image registration tool with free form deformation model. we used the same settings as [28] except B-spline control points. Due to the minuscule pixel spacing of CUBIC dataset, the control point spacing is set to 5mm which is the minimum possible value allowed for this method.

5) VoxelMorph: [2] is a recently developed deep learning based image registration tool. This method used U-net architecture for deformation generation. Unlike the proposed DDN

method, original Voxelmorph was developed to take whole resolution instead of patch extraction. For our evaluation we modified the original VoxelMorph to a patch based Voxelmorph to train on extracted patches.

D. Experimental Setup

The proposed network is developed in Keras with tensorflow backend. The network is trained and tested in High Performance Computing environment with 200 hours walltime, 64 GB RAM, 12 GB Video RAM in Tesla K40m GPU and a single core 2.66GHz 64bit Intel Xeon processor.

VI. Results and Evaluation

The performance of the network is evaluated on two different resolution scale, 10% and 25% respectively. We compared the quantitative performance of registration methods by normalized cross-correlation and mutual information for test brains. For testing the methods we have a CUBIC dataset of three brains and for the notational convenience we note them as brain "001", "002" and "003". The performance scores showed in tables are measured by making brain "003" as target and other two brains as source image. For qualitative performance, same brain slice extracted from all test brains and overlaid on the target brain with different color map. In all of our experiments, target brain is mapped with red color and registered brains are mapped with green color. The lower part of the brains, therefore the cerebellum region varies from brain to brain and dissimilarity in this region is expected. For the visual quality assessment we consider (upper part of the brains) regions like hippocampal formation, dentate gyrus to be similar and aligned by the selected methods. Since tissue cleared images contain mixed fore-ground and background and the nature of information (the cells) are discrete, its difficult to evaluate the performance of the registration tools by viewing the whole brain slice. To facilitate the qualitative evaluation process, we crop and zoom the patches from hippocampal formation and dentate gyrus along with the whole brain slice. In most of the image registration research, performance are evaluated by measuring overlap between registered and target image using Dice similarity metric. The overlap measurement require manually segmented ground-truth data. In our case, no ground-truth results are available and manually producing ground-truth for high resolution CUBIC data is costly and time consuming. To validate the registration results due to lack of ground-truth, a validation experiment is designed. The test brains are deformed by applying a random Gaussian deformation. All the registration tools are applied on these deformed brains as source image to register them to their non-deformed version. This experiment ensures the similarity in registration performance as in general case.

A. Performance at 10% Resolution

1) Quantitative: Table I shows the performance of selected tools and proposed method. The performance of the tools varies across the CC and MI scores. In cross-correlation scores and for brain "001", traditional methods obtained best scores with 0.9589 (Elastix) and 0.9494 (ANTS), The deep learning based DDN is in third position with 0.9093 whereas deep-learning based Voxelmorph obtained 0.8107 slightly better than IRTK. In terms of mutual-information score, ANTS is the top scorer with 0.8057 Elastix and NiftyReg obtained second and third respectively with 0.7899 and 0.6806. The proposed DDN obtained 0.6098 in mutual information score. For brain "002" a similar pattern is obtained with slight position changes. ANTS is the top scorer in terms CC. Elastix and proposed DDN obtained very similar scores with tiny difference. In MI score proposed DDN obtained 0.6602. The rank between other tools remain as brain "001".

2) Validation of Performances: The results of the validation experiment is shown in Table II. A different performance pattern is appeared in this experiment. In this experiment, we compared the CC and MI scores of all registered-deformed version of the brain to their original version. The ANTS remains the top in CC scores in brain "001" and "003". Elastix obtained the best CC score in brain "002". The proposed DDN network obtained 0.8664, 0.7826 and 0.9826 respectively. The VM network obtained least performance with 0.8261, 0.6987, 0.9388 among all the registration tools. In terms of MI score, DDN network obtains scores 0.4383, 0.3490 and 0.6005. ANTS scored the best performance while other three tools performed differently in different brains. One of the reason behind this performance difference compared to the Table I for Elastix, NiftyReg and Voxelmorph might be the choice of the tuning parameters. In this experiment we use same parameter sets for all the tools. The performance of the deep-learning based tools in the validation test is quite low compared to their traditional competitors. The reason behind this is the deformation applied is global in nature. The learning based tools are good at capturing local deformation constrained inside the small patches used to train the networks.

3) Qualitative: The qualitative performance of the proposed DDN and four other tools at 10% resolution are presented in figure 4. In figure 4, the registered brain(green color) is overlay-ed on the reference brain(red color). The regions of the brain where colors become yellowish is aligned perfectly. In Figure 4, the first row shows registration results of proposed architecture and Voxelmorph. The second and third rows contains the results from optimization based tools. In alignment results, the visual differences in the cerebellum region (lower part) is common and expected since this region varies brain to brain. We consider the difference in Hippocampal and dentate gyrus region as visual alignment errors. For traditional tools, ANTS and elastix shows best visual alignment in both regions. NiftyReg and IRTK shows erroneous alignment in both regions and brains. Alignment in deep learning based method Voxelmorph showed promising results in both brains with small differences in dentate gyrus region. The proposed DDN method on the other hand, shows clear improvement in alignment in this region. The alignment performance in hippocampal region by both deep learning based algorithm is difficult to judge. In our experiment, we used patch-based training strategy for both network, hence its better to compare their performance on the local patches for

TABLE I: Performance comparison at 10% Resolution

Methods	brain 1(CC)	brain 1(MI)	brain 2(CC)	brain 2(MI)
BEFORE REGISTRATION	0.8056	0.3865	0.7716	0.3619
ANTS	0.9494	0.8057	0.9596	0.8241
Elastix	0.9589	0.7899	0.9388	0.8104
NiftyReg	0.8623	0.6806	0.8489	0.6526
IRTK	0.8013	0.5340	0.6958	0.5050
VoxelMorph(VM)	0.8107	0.4543	0.8405	0.4960
DenseDeformation(DDN)	0.9093	0.6098	0.9258	0.6602

TABLE II: Validation of Registration at 10% Resolution

Methods	brain 1(CC)	brain 1(MI)	brain 2(CC)	brain 2(MI)	brain 3(CC)	brain 3(MI)
ANTS	0.9836	0.6424	0.9716	0.6635	0.9973	0.7538
Elastix	0.9757	0.5880	0.9731	0.6062	0.9956	0.6955
NiftyReg	0.9720	0.6068	0.9641	0.6416	0.9923	0.6556
VoxelMorph	0.8261	0.3706	0.6987	0.2980	0.9388	0.4523
DenseDeformation	0.8664	0.4383	0.7826	0.3490	0.9826	0.6005

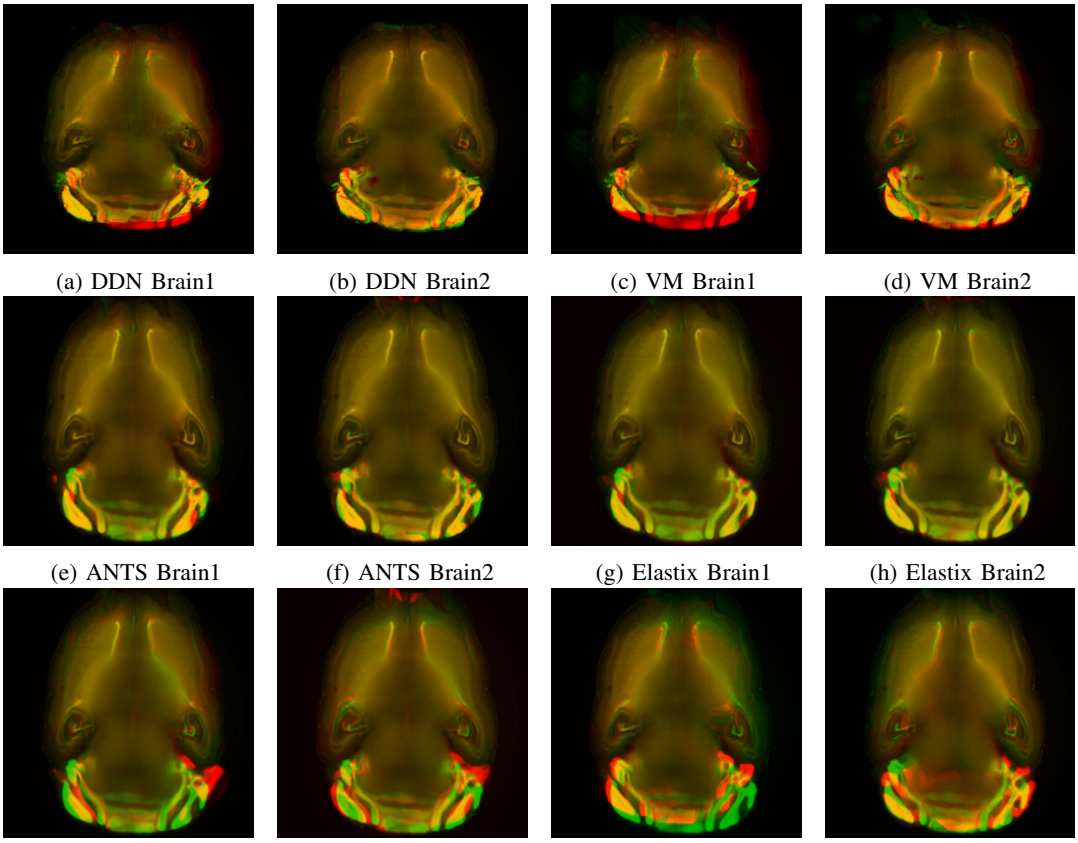
more accurate understanding than whole brain slice. Moreover, visual quality of these patch based algorithms can easily be affected by patch stitching process. Considering these facts and to accurately identify the alignment performance of all the baseline methods and proposed DDN methods, we select two hippocampal region and one dentate gyrus region. For all the methods, we extract patches from selected regions from both registered image and reference image and calculate the difference between the patches. The difference image contains intensity values in the range +1 to -1 (since all brain volumes are intensity normalized from 0 to 1). For accurate alignment the intensity difference should be 0 in the difference image. For proper visualization of the difference image, we transform the difference image intensities into the range 0 to 255 using a linear triangular function centered at zero.The resulting transformed images thus contains intensity values 255 (0 in difference image) or white, in accurately aligned regions and 0 or dark in non-aligned regions. For all the tools, similar process is applied and resulting difference-transformed patches from three selected regions are shown along with the overlayed image slices in Figure 3. From Figure 3, the observation are as follows:

- For patch A (Dentate Gyrus), all the optimization based registration methods have compareable performance with slight differences. Among them IRTK has highest level of dark intensities indicating clearly its low performance as indicated in quantitative performance. Between proposed DDN and Voxelmorph, DDN has much lower darker intensities compared to Voxelmorph which has two clear dark lines clearly indicates the misalignment between brains.
- For patch B (Left Hippocampal),ANTS and NiftyReg both have lowest level of dark intensities compared to other tools. Proposed DDN method is batter than IRTK and Voxelmorph but has slightly more dark regions than the Elastix.
- For patch C (Right Hippocampal), ANTS and Elastix has lowest level of dark intensities among all the tools. The proposed DDN has similar level of dark intensities with slight difference to ANTS and Elastix. Compared to other tools DDN performed well aligning in this region.

B. Performance at 25% Resolution

1) Quantitative: Registration performance of the proposed method and selected baselines at 25% resolution are shown in Table III. At 25% resolution the similarity measurements before the registration is quite low for the brain 001 compared to brain 002. At 25% resolution ANTS obtains top CC and MI score for brain 001 with 0.8712 and 1.3583 respectively. Elastix scored second best while deep learning based proposed DDN network obtain 3rd for brain 001 with 0.8337. In MI score, proposed DDN obtained second with score 1.1765 while Elastix scored 1.1542. NiftyReg IRTK and Voxelmorph obtained the rest of the positions with very small difference in performance metrics. For brain 002, proposed DDN obtained top score in CC with 0.9438 and second best in MI with 1.5003. ANTS, the top scorer in most of the cases obtain 0.9391 in CC and 1.5918 in MI. Elastix and Voxelmorph both highly competed for the third position. Voxelmorph performed slightly better in CC score than Elastix wheres Elastix obtain much better MI score than Voxelmorph. For brain 002 both deep-learning based method performed better than brain 001 and strongly challenged their traditional counterparts. For both test brains, proposed DDN architecture performed much better than its technically closest rival Voxelmorph. Overall, the performance measure at 25% resolution indicates the potential ability of the proposed architecture over the traditional and deep-learning based methods.

2) Validation of Performances: The validation performances at 25% resolution are shown in Table IV. Like 10% resolution, the validation performance at 25% resolution shows similar performance of all the tools. The ANTS remains the top in all the brains in CC score. Elastix and NiftyReg remains 2nd and 3rd position respectively. The Voxelmorph obtained the last position in CC score for all brains. In MI score, the proposed DDN network obtains the top position for Brain ”002” and ”003” with 1.8824 and 1.2238 respectively. In this metric, VM network achieves the 2nd bestfor brain ”002”



(a) DDN Brain1 (b) DDN Brain2 (c) VM Brain1 (d) VM Brain2

(e) ANTS Brain1 (f) ANTS Brain2 (g) Elastix Brain1 (h) Elastix Brain2

(i) NiftiReg Brain1 (j) NiftiReg Brain2 (k) IRTK Brain1 (l) IRTKBrain2

Fig. 2: Visual Comparison of ANTS and Dense Deformation Network at 10% Resolution

TABLE III: Performance comparison at 25% Resolution

Methods	brain 1(CC)	brain 1(MI)	brain 2(CC)	brain 2(MI)
BEFORE REGISTRATION	0.6763	0.9147	0.8016	1.2006
ANTS	0.8712	1.3583	0.9391	1.5918
Elastix	0.8686	1.1542	0.9102	1.2903
NiftyReg	0.7166	1.1319	0.7570	1.1677
IRTK	0.7946	0.9250	0.8091	0.9508
VM(0.50verlap)	0.7877	1.0620	0.9126	1.4009
DDN(0.50verlap)	0.8337	1.1765	0.9438	1.5003

and ”003” suppressing ANTS, Elastix and NiftyReg.In the validation results at 25% resolution, all the optimization based tools performed consistently in all three brains. Both deep learning based methods showed very good performance in brain ”002” and brain ”003” but looses their performance in brain ”001”.

3) Qualitative Performance: The qualitative performance at 25% resolution are shown in Figure VI-B1 for all registration tools. Like, 10% resolution, the reference brain and registered brains with red and green color-map are overlayed and intensities became yellowish in aligned regions. For the evaluation consistency, the same slice is used and represented for qualitative evaluation in both resolution. Like 10% resolution, ANTS gives best alignment interms of color-map mixing and smoothness in both test brains. Elastix performed well for brain ”002” but showed difficulty in aligning at dentate gyrus region in brain ”001”. NiftyReg showd similar pattern of misalignment in brain ”001”. IRTK on the other hand completely failed in aligning both brains. Though quantitative performance of IRTK is at satisfactory, the visual inspection clearly indicates its disability at higher resolution. The performance of proposed DDN and voxelmorph seems very similar. Despite their close proximity in whole slice representation for both test brains, close inspection in both hippocampal regions in both brains indicates dissimilarities. For both test brain, DDN has more color-map mixed yellow intensities in both hippocampal regions compared to Voxelmorp method. The proposed method also performed well than NiftyReg and Elastix but more alignment errors than ANTS. Like 10% resolution, we take same strategy for better visual representation of patches from two hippocampal regions and dentate gyrus region. Alignment results from all the tools with difference-transformed patches at 25% resolution are shown in Figure 5. From Figure 5 we observe following facts:

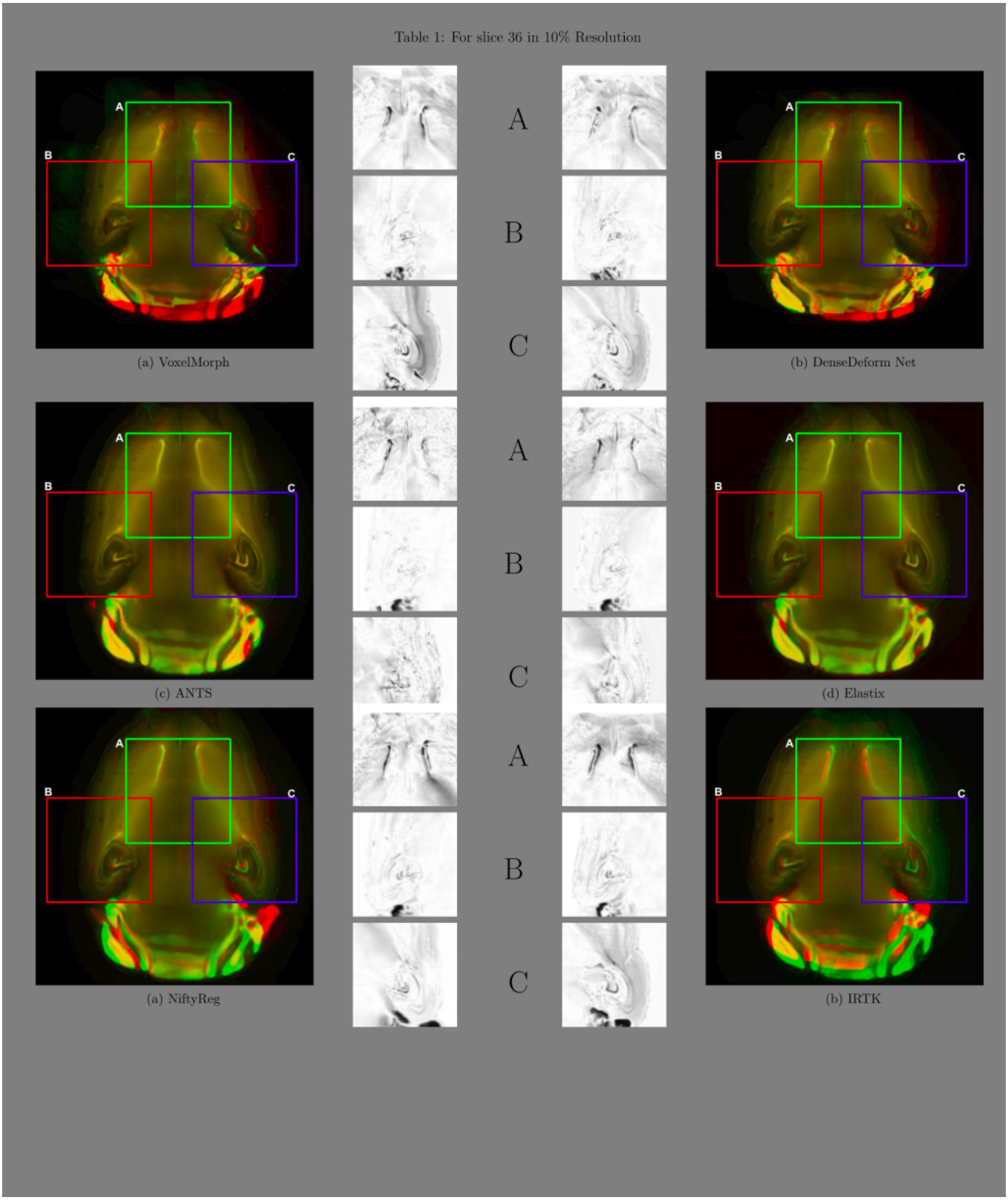


Fig. 3: Visual Comparison of Patches from DDN and VM networks at 10% Resoliution

TABLE IV: Validation of Registration at 25% Resolution

Methods	brain 1(CC)	brain 1(MI)	brain 2(CC)	brain 2(MI)	brain 3(CC)	brain 3(MI)
ANTS	0.9574	1.1299	0.9668	1.2015	0.9698	1.2061
Elastix	0.9562	1.0417	0.9547	1.1095	0.9590	1.0896
NiftyReg	0.9384	1.0192	0.9373	1.0770	0.9234	1.0409
VoxelMorph	0.7858	1.0304	0.9687	1.8053	0.8423	1.1657
DenseDeformation	0.8259	1.1047	0.9791	1.8824	0.8682	1.2238

- For patch A(Dentate Gyrus), ANTS clearly outperformed all others with lowest level of dark regions. Elastix and NiftyReg has different patterns in this region. Elastix has single two single dark curves where NiftyReg has two different dark curves at left side indicating its misalignment. Proposed DDN has more white shades in the right side of the patch and has two thin lines indicating its slightly better performance than Voxelmorph.
- For patch B(Left Hippocampal),its difficult to determine between ANTS and Elastix. Both of them has less dark regions compared to others. Close inspection between proposed DDN and Voxelmorph indicates that DDN has dark regions smaller than Voxelmorph and their their darkness is more blurry compared to Voxelmorph. The blurry darkness indicates less difference between reference and source images.
- For patch C (Right Hippocampal), ANTS remains the superior performer among all. Elastix remains the second best in terms of dark regions. Between proposed DDN and Voxelmorph, both of the patches appears very similar in almost all regions. Close inspection in the middle of the Hippocampal finds that DDN has tiny sharp dark region while Voxelmorph has more thicker dark spots in the same place.

From the visual evaluation at 25% resolution, the proposed DDN method has better performance than its closest rival Voxelmorph. Performance by traditional tools remains are still better than the deep-learning based tools.

VII. Conclusion

In this paper we proposed a patch-based 3D registration framework consists of densely connected convolutional layers. The proposed architecture, utilizes two steps of deformation generation and generate a dense deformation field. The unsupervised patch-based training strategy gives the network ability to learn displacement without need of any ground truth data. The performance of the DDN network at 25% resolution is more convincing than its performance at 10% resolution which indicates its ability to handle higher resolution data than the Voxelmorph network. The qualitative results at two different resolution scale also proved that the proposed DDN network can better handle the complexity of tissue cleared data compared to the Voxelmorph network. Though, the performance of the network still behind the traditional tools, its ability to register images within minutes with satisfactory accuracy, makes it a definite choice for clinical applications. There are still rooms for further investigation and research in deep learning based image registration methods. The DDN network and Voxelmorph both struggles to capture complex and large deformation. An investigation in loss function and regularization parameter is a definite direction for future research. The patch-based training generates discontinuous deformation fields and interpolation using these discontinuity sometime gives unnatural images. Therefore, investigation in patch-by-patch continuous deformation generation can also be a future direction of research for deep learning based registration approaches.

Acknowledgment

Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

References

[1] A. Nazib, J. Galloway, C. Fookes, and D. Perrin, "Performance of Image Registration Tools on High-Resolution 3D Brain Images," jul 2018. [Online]. Available: http://arxiv.org/abs/1807.04917

[2] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An Unsupervised Learning Model for Deformable Medical Image Registration," feb 2018. [Online]. Available: http://arxiv.org/abs/1802.02604

[3] K. A. J. Eppenhof and J. P. W. Pluim, "Pulmonary CT Registration Through Supervised Learning With Convolutional Neural Networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1097–1105, may 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8510836/

[4] X. Cao, J. Yang, J. Zhang, Q. Wang, P.-T. Yap, and D. Shen, "Deformable Image Registration Using a Cue-Aware Deep Regression Network," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1900–1911, sep 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8331111/

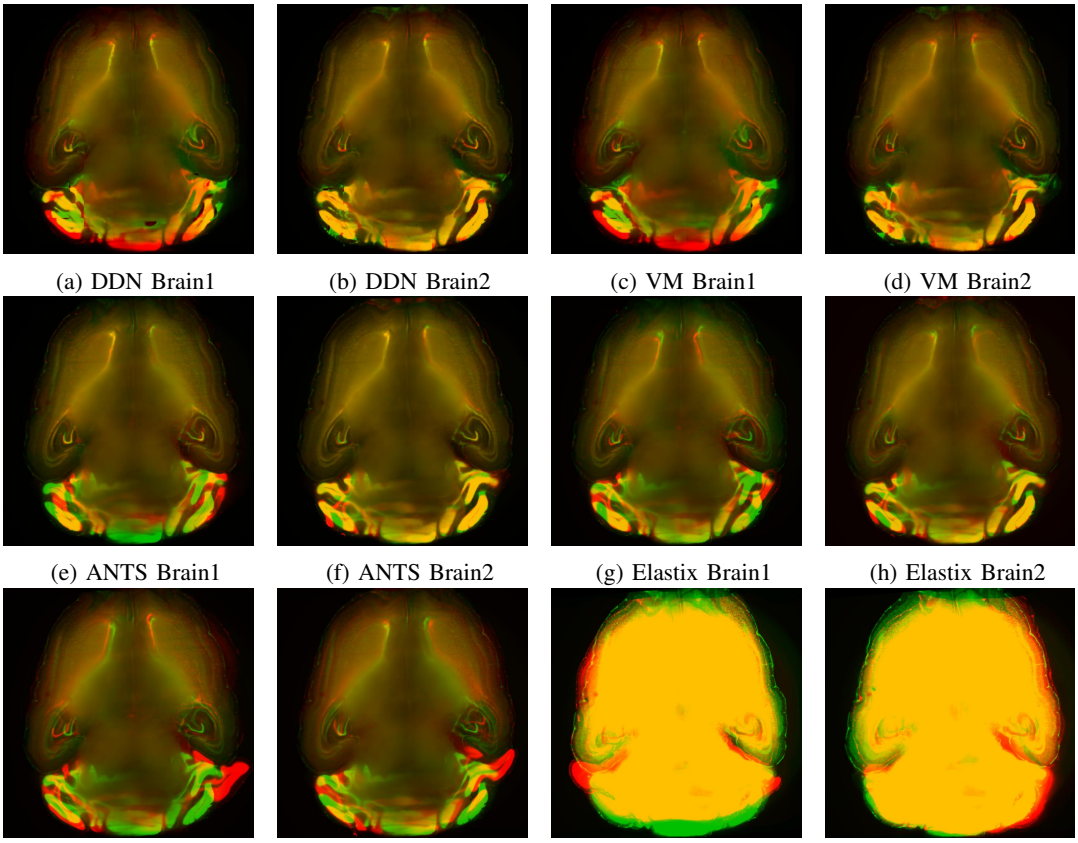
[5] W.-S. Lai, J.-B. Huang, and M.-H. Yang, "Semi-Supervised Learning for Optical Flow with Generative Adversarial Networks," *Nips2017*, vol. 1, no. Nips, pp. 353–363, 2017. [Online]. Available: http://papers.nips.cc/paper/6639-semi-supervised-learning-for-optical-flow-with-generative-adversarial-networks.pdf

[6] X. Cheng, L. Zhang, and Y. Zheng, "Deep similarity learning for multimodal medical images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 6, no. 3, pp. 248–252, apr 2018. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/21681163.2015.1135299

[7] D. Mateus, M. Simonovsky, N. Navab, and N. Komodakis, "A Deep Metric for Multimodal Registration," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 9902, 2016, pp. 10–18. [Online]. Available: http://link.springer.com/10.1007/978-3-319-46726-9

[8] G. Wu, M. Kim, Q. Wang, B. C. Munsell, D. Shen, and f. t. A. D. N. Initiative, "Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning." *IEEE transactions on bio-medical engineering*, vol. 63, no. 7, pp. 1505–16, jul 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/26552069 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4853306

[9] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, sep 1998. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1361841598800224



(a) DDN Brain1 (b) DDN Brain2 (c) VM Brain1 (d) VM Brain2

(e) ANTS Brain1 (f) ANTS Brain2 (g) Elastix Brain1 (h) Elastix Brain2

(i) NiftiReg Brain1 (j) NiftiReg Brain2 (k) IRTK Brain1 (l) IRTK Brain2

Fig. 4: Visual Comparison of ANTS and Dense Deformation Network at 25% Resolution

[10] S. Miao, Z. J. Wang, and R. Liao, "A CNN Regression Approach for Real-Time 2D/3D Registration," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352–1363, may 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7393571/

[11] . Xiao Yang1(B), Roland Kwitt3, and Marc Niethammer1, "Fast Predictive Image Registration," *Springer Verlag*, vol. 1, no. 10, pp. 48–57, 2017. [Online]. Available: http://arxiv.org/abs/1704.06040

[12] H. Sokooti, B. De Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, "Medical Image Computing and Computer Assisted Intervention MICCAI 2017," vol. 10433, 2017. [Online]. Available: http://link.springer.com/10.1007/978-3-319-66182-7

[13] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "SVF-Net: Learning Deformable Image Registration Using Shape Matching." [Online]. Available: https://hal.inria.fr/hal-01557417/document

[14] H. Li and Y. Fan, "Non-Rigid Image Registration Using Self-Supervised Fully Convolutional Networks without Training Data," *arXiv preprint arXiv:1801.04012*, 2018. [Online]. Available: http://arxiv.org/abs/1801.04012

[15] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration," 2018. [Online]. Available: http://arxiv.org/abs/1805.04605

[16] H.-U. Dodt, U. Leischner, A. Schierloh, N. Jährling, C. P. Mauch, K. Deininger, J. M. Deussing, M. Eder, W. Zieglgänsberger, and K. Becker, "Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain," *Nature Methods*, vol. 4, no. 4, pp. 331–336, apr 2007. [Online]. Available: http://www.nature.com/doifinder/10.1038/nmeth1036

[17] H. Hama, H. Kurokawa, H. Kawano, R. Ando, T. Shimogori, H. Noda, K. Fukami, A. Sakaue-Sawano, and A. Miyawaki, "Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain," *Nature Neuroscience*, vol. 14, no. 11, pp. 1481–1488, aug 2011. [Online]. Available: http://www.nature.com/doifinder/10.1038/nn.2928

[18] M.-T. Ke, S. Fujimoto, and T. Imai, "SeeDB: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction," *Nature Neuroscience*, vol. 16, no. 8, pp. 1154–1161, jun 2013. [Online]. Available: http://www.nature.com/doifinder/10.1038/nn.3447

[19] K. Chung and K. Deisseroth, "CLARITY for mapping the nervous system," *Nature Methods*, vol. 10, no. 6, pp. 508–513, may 2013. [Online]. Available: http://www.nature.com/doifinder/10.1038/nmeth.2481

[20] N. Renier, Z. Wu, D. J. Simon, J. Yang, P. Ariel, and M. Tessier-Lavigne, "IDISCO: A simple, rapid method to immunolabel large tissue samples for volume imaging," *Cell*, vol. 159, no. 4, pp. 896–910, 2014.

[21] E. A. Susaki, K. Tainaka, D. Perrin, F. Kishino, T. Tawara, T. M. Watanabe, C. Yokoyama, H. Onoe, M. Eguchi, S. Yamaguchi, T. Abe, H. Kiyonari, Y. Shimizu, A. Miyawaki, H. Yokota, and H. R. Ueda, "Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis," *Cell*, vol. 157, no. 3, pp. 726–739, apr 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24746791 http://linkinghub.elsevier.com/retrieve/pii/S0092867414004188

[22] E. A. Susaki, K. Tainaka, D. Perrin, H. Yukinaga, A. Kuno, and H. R. Ueda, "Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging," *Nature Protocols*, vol. 10, no. 11, pp. 1709–1727, oct 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/nprot.2015.085

[23] L. Yu, J.-Z. Cheng, Q. Dou, X. Yang, H. Chen, J. Qin, and P.-A. Heng, "Automatic 3D Cardiovascular MR Segmentation with Densely-Connected Volumetric ConvNets." Springer, Cham, 2017, pp. 287–295. [Online]. Available: http://link.springer.com/10.1007/978-3-319-66185-8_33

[24] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, feb 2008.

[25] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, jan 2010.

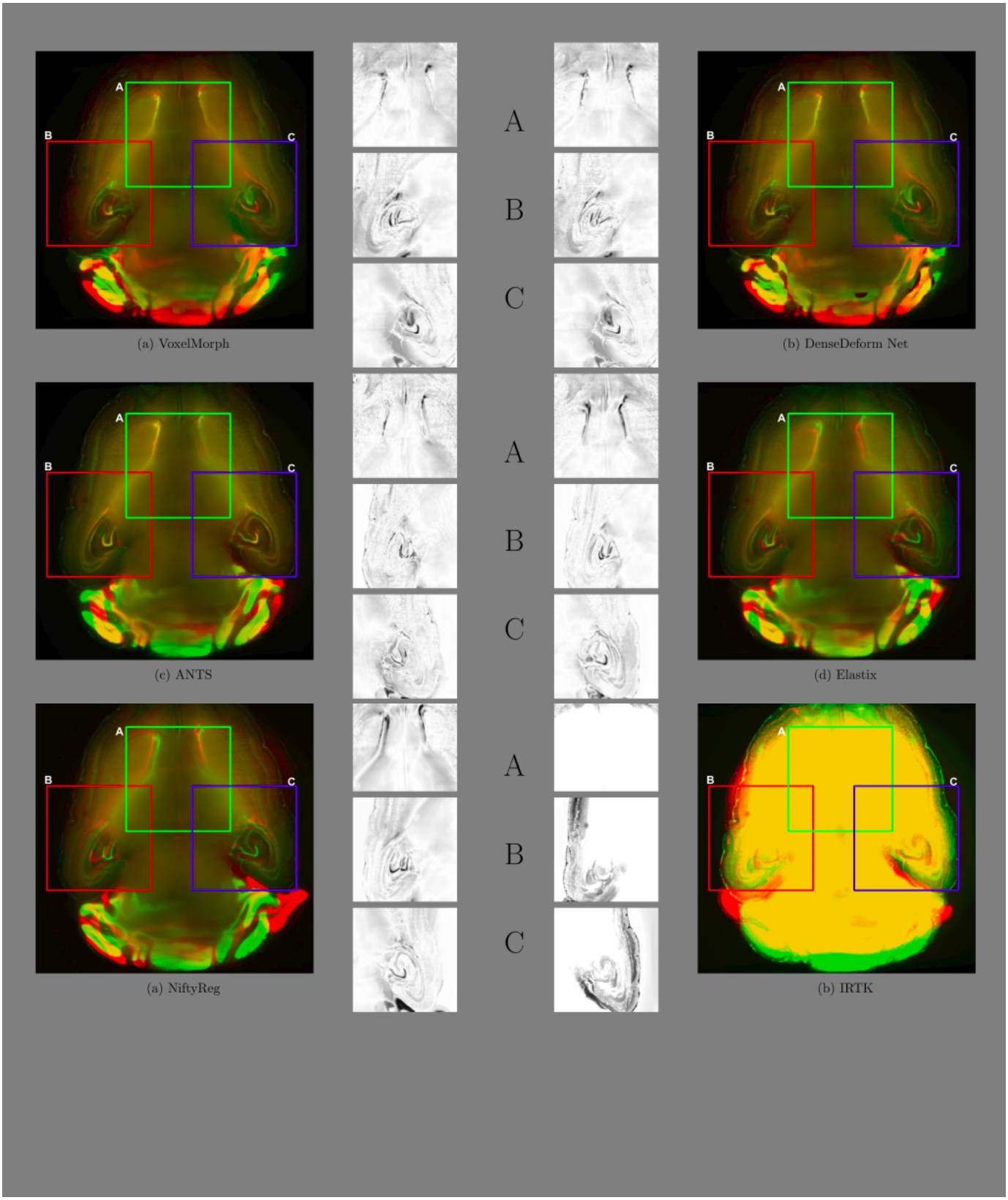


Fig. 5: Visual Comparison of Patches from DDN and VM networks at 25% Resoliution

[Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19923044 http://ieeexplore.ieee.org/document/5338015/

[26] L. Hammelrath, S. Škokić, A. Khmelinskii, A. Hess, N. van der Knaap, M. Staring, B. P. Lelieveldt, D. Wiedermann, and M. Hoehn, "Morphological maturation of the mouse brain: An in vivo MRI and histology investigation," *NeuroImage*, vol. 125, pp. 144–152, 2016.

[27] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *Computer Methods and Programs in Biomedicine*, vol. 98, no. 3, pp. 278–284, 2010.

[28] Z. Xu, C. P. Lee, M. P. Heinrich, M. Modat, D. Rueckert, S. Ourselin, R. G. Abramson, and B. A. Landman, "Evaluation of Six Registration Methods for the Human Abdomen on Clinically Acquired CT," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1563–1572, aug 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7482649/

[29] D. Rueckert, "Nonrigid registration using free-form deformations: Application to breast mr images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999. [Online]. Available: http://ieeexplore.ieee.org/document/796284/