Deformable image registration of worm brains

Bruchez Loic, de Riedmatten Ines, Madrona Antoine
EPF Lausanne, Switzerland

***Abstract*—In this project, VoxelMorph [1] was used to align all the frames of a confocal microscopy video of C. Elegans in order to generalize the segmentation data of seven red fluorescently-labelled neurons to all the frames. The goal was to reach similar or higher accuracy than state-of-the art registration methods but in a shorter time. First, a new pipeline was investigated to roughly pre-align some raw data and test unsupervised VoxelMorph in 2D as a proof-of-concept. In a second time, VoxelMorph was applied with unsupervised and semi-supervised learning on pre-aligned frames to assess the efficiency and the rapidity of the algorithm in 3D. The best results were obtained with the semi-supervised training MSE, a L2-regularizer λ of $5e^{-2}$, 32 epochs and a batch size of 4.**

I. Introduction

Contrarily to the human, C. Elegans does not have a brain per se. Nevertheless, it possesses a nervous system, composed of neurons grouped in ganglia in the head, tail and into a spinal cord-like ventral nerve cord [2]. Therefore, the neuronal activity of C. Elegans can be studied. In this work, seven neurons are experimentally labelled with red fluorescent proteins and the neuronal activity of all neurons is shown with the help of GCaMP green-fluorescent protein. The frames extracted from a confocal microscopy movie must be aligned in order to facilitate the tracking of the neuronal activity of the seven neurons of interest over time.

The actual pipeline of the LPBS Laboratory performs a raw alignment of the frames using the Jian and Vemuri method. Some of the frames are labelled by hand. Our work consists in improving this alignment using non-linear transformations and predicting the labels for the rest of the frames using VoxelMorph. This framework is aimed for deformable, pairwise medical image registration. In this case, the registration consists in a function that maps an input image pair to a deformation field that aligns these images, thanks to a convolutional neural network (CNN) architecture similar to UNet (Figure 1).

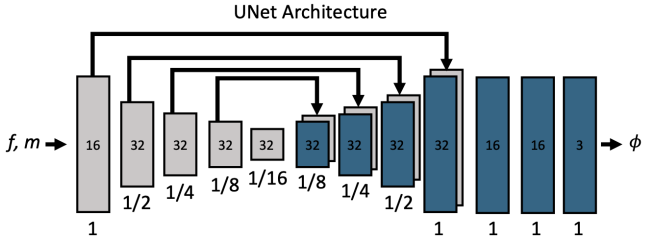

Figure 1: Convolutional UNet architecture, see Figure 3 of [1]

Compared to state-of-the-art methods as for instance ANTs or NiftyReg, VoxelMorph reaches the same accuracy in a significantly smaller running time as it learns a general function instead of pairwise transformations. For this reason, this algorithm appears to be a good choice to improve the current pipeline.

II. Models and Methods

VoxelMorph

The Adam optimization algorithm was used with a learning rate of $1e^{-4}$. The two mostly used unsupervised losses during training are the mean-squared error (MSE) and the normalized cross-correlation (NCC). In this case, the NCC should be preferred, as it is more robust to intensity variations. Those losses penalize the difference between a fixed and a warped image (transformation learnt on a image is applied to another image, for example its labels or another channel).

In the unsupervised and the semi-supervised cases, a L2-regularization parameter λ penalizes the local spatial variation in the registration field. The parameter λ will be optimized during this work.

In the semi-supervised case, an auxiliary data loss function, the Dice score, can be used during the training to quantify the volume overlap between the warped reference label with the hand-made labels of the original frame. This Dice score is regularized with a parameter γ. In this project, the parameter γ will be kept constant at a value of 0.01. Another Dice score and NCC were used to directly quantify the results obtained during testing.

Concerning the data format, the generators based on individual compressed file for each entry were favored as they solve the constraint of live memory limitations.

For both 2D and 3D models, the UNet structure deployed was identical to the one described in the original publication. The encoder and decoder had respectively a 16, 32, 32, 32 and 32, 32, 32, 32, 32, 16, 16 structures (see Figure 1).

Data Acquisition

During the acquisition of the movie, the samples must be moved up and down on a confocal, which only shows fluorescence in one z-slice. Each direction, up or down, causes different distortions that were considered negligible in the context of this algorithm.

Raw data

The raw data consists of 572 unlabelled frames with red and green channels. Each frame has a volume of 512x512x35.

Feature processing: Problematic frames with indistinguishable, fused neurons were identified and removed from the data-set. Frames with artefacts such as considerable smearing or staggering were also removed.
The frames are normalized using the min-max method and median blurred to efficiently reduce the salt-and-pepper noise of the volumes. Then, a 3D mask is generated based on the red channel and applied to the green channel in order to extract the neural activity of the neurons of interest.

In this part of the project, the VoxelMorph application was focused on 2 dimensions data. In order the reduce the dimension of the dataset, a z-axis maximum intensity projection (MIP) of the red channel and the masked green channel was performed. Then, to roughly align the 2D data, a rotation was first applied to horizontally align the two brightest neurons' center of mass in the red MIP. Additionally, the frames were cropped around the center of mass of the red MIP and minimal value padding was used when necessary to finally obtain 256x256 images.

Training: The model was trained on using the red channel MIPs for 90 epochs and batch size of 16 and optimization of the λ regularization was performed by manual grid search. A ratio 0.8-0.1-0.1 for the training-validation-testing was performed. The testing data were kept for visualizing the warped image. The first volume of the data is chosen as the fixed (reference) for the learning, because the two brightest neurons were clearly separated in the MIP.

Once trained, the model can provide a red reference, a red channel MIP and its corresponding green MIP, generate a transformation function based on the red channel images and apply it to the green channel using the warp function.

In order to use all the available data while avoiding overfitting, the training was performed using 32 epochs, 250 steps per epoch, with λ = 0.5, 5e^{-2} and 5e^{-5} with MSE and 0.9, 1, 1.1, 1.25 and 1.5 with NCC.

Pre-aligned data

The pre-aligned data consists of 1715 frames, 118 of which are labelled. The red channel only is available and each frame has a volume of 112x112x32. Those data are aimed at testing the efficacy of VoxelMorph on this alignment task/generalization of segmentation data to all the frames.

Feature processing: The data was translated, rotated and cropped using the Jian and Vemuri algorithm.

A few aberrant values above 255 were observed and clamped with a threshold value of 255. All the frames were min-max normalized in order to improve the training.

Using MIP projection on all 3 axis, the 3D arrangement of several slices and the distribution/dispatching of the neurons were studied. Additionally, a testing dice score was computed for all the labeled frames with respect to each of them in order to identify the best candidates to be used as pre-alignment reference, or atlas. Using this method, the frame 904 was retained and used as an atlas for the rest of this work.

Training: The unsupervised and semi-supervised learning methods implemented by VoxelMorph were tested. Since no public atlas corresponded to the studied data, the frame 904 was used as one, since it was the one offering the best testing dice score on the pre-aligned data when used as reference (see Table II). A ratio 0.8-0.2 for training-testing was performed. To be able to detect and avoid overfitting, a validation data-set consisting of the 30 first frames of the training set was used. It had to contain a restricted number of frames because of memory constraints during the model training (no generator could be used for this purpose)

In order to use all the available data while avoiding overfitting, the training was performed using 32 epochs, 250 steps per epoch, with λ = 0.1, 5e^{-2} or 5e^{-5} for MSE and λ = 6e^{-1} or 9e^{-1} or 1 for NCC.

Depending on the model used, some slight differences occur:

Unsupervised: Learning on the red 3D channel using the homemade atlas as reference.

Semi-supervised: In order to further improve the performance of VoxelMorph, the labels of the training set frames (95 in total) were taken advantage of. In a first time, an optimization of the λ L2-regularization weight was undertaken whereas the γ regularizer of the Dice score and the learning rate were kept constant at 0.01 and $1e^{-4}$, respectively. Three values of λ were bench-marked with 32 epochs, 250 steps and a batch size of 4. The performance was assessed using a testing Dice Score and a testing NCC. For the Dice score, it was either calculated using all labels, or only the label 4, corresponding to the bright neuron used as reference during the pre-alignment.

III. Results

Raw data

With λ of 0.05, both training and validation losses have comparable range of values and converges around epoch 70 (Figure 2). Based on this convergence, the validation losses were average between epoch 70 to epoch 90. The values for MSE are summarized in Table I. Note that all values of λ except λ=1 gives rise to divergence for NCC. For λ=1, the NCC is $-9.07e^{-2}(\pm 7.5e^{-5})$.

After having trained the model, the learned deformation field was applied to new data (Figure 3).

Pre-aligned data

Unsupervised: As no significant results were obtained using the unsupervised, no results are shown.

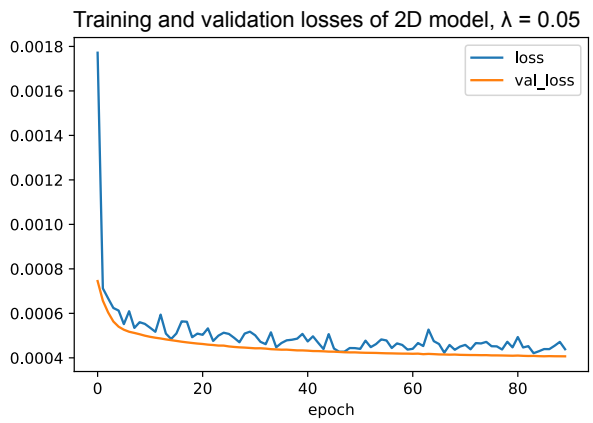

Figure 2: Training and validation losses of the 2D model. Nb_Epochs = 90, batch_size = 16, λ = 0.05. The validation set size is 50.

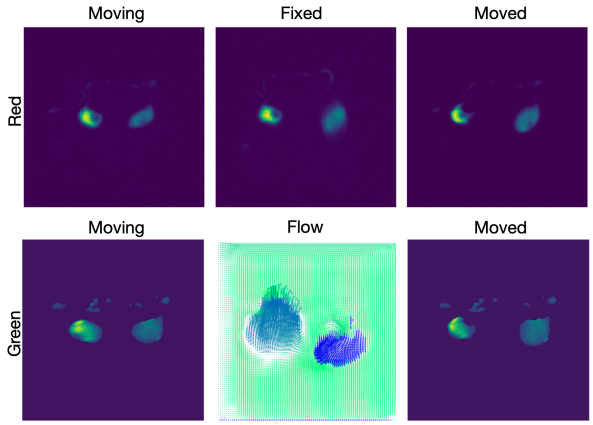

Figure 3: Warping of the entry '50' moving green, based on the model with MSE and λ = 0.05. The warping is generated using the nearest interpolation method

λ	loss with MSE
$5e^{-1}$	$7.04e^{-4}(\pm 4e^{-6})$
$5e^{-2}$	$4.10e^{-4}(\pm 2e^{-6})$
$5e^{-5}$	$1.19e^{-4}(\pm 7e^{-6})$

Table I: Validation losses of a 50 frames validation set of the 2D model, average over epoch 70 to 90, with λ of $5e^{-5}$, $5e^{-2}$ and $5e^{-1}$

Semi-supervised: Based on the average Dice scores and NCC testing metrics for the semi-supervised model, best results for training MSE and training NCC were achieved with λ of $5e^{-5}$ and $6e^{-1}$ respectively (see Table II).

IV. Discussion

Raw Data - 2D

The pre-alignment was fast and pretty robust. Nevertheless, some slices in MIP form exhibited two close bright spots that were difficult to separate. For those cases, the rotation did not work well.

As visible in Figure 3, the VoxelMorph alignment did not bring satisfying results because of information loss due to the MIP. In addition, it was difficult to quantitatively assess the performance of the model because of the lack of a ground truth. During the training, the impact of the reference frame quality was realized.

λ (loss funct.)	Avg DS	Avg DS label 4	NCC
pre − align.	**0.35 (± 0.30)**	**0.55 (± 0.24)**	**0.48 (± 0.27)**
$1e^{-1}(MSE)$	0.30 (± 0.28)	0.52 (± 0.21)	0.64 (± 0.24)
$5e^{-2}(MSE)$	0.32 (± 0.28)	0.53 (± 0.22)	0.64 (± 0.23)
$5e^{-5}(MSE)$	0.26 (± 0.25)	0.35 (± 0.28)	0.47 (± 0.24)
$6e^{-1}(NCC)$	0.31 (± 0.29)	0.50 (± 0.24)	0.58 (± 0.23)
$9e^{-1}(NCC)$	0.30 (± 0.28)	0.50 (± 0.21)	0.54 (± 0.22)
$1(NCC)$	0.29 (± 0.27)	0.50 (± 0.23)	0.57 (± 0.21)

Table II: Average testing dice score (DS) and Normalized Cross-Correlation (NCC), calculated between the atlas labels warped with the transformation learned on the volume red channel and the volume original labels for 22 unseen volumes. The pre-aligned row correspond to the values obtained by using the volume original labels versus the atlas labels. The Dice Score was obtained for all 7 values of labels or just the label 4. NCC was obtained by flattening the volumes and computing the Pearson product-moment correlation. The models were trained using all the labels values, 32 epochs, 250 steps, MSE or NCC, batch size of 4 and frame 904 with its labels as atlas.

Pre-aligned - 3D

1) Unsupervised: The unsupervised learning was slow, with an average of 30 seconds per step, and after some preliminary runs with a reduced number of epochs and steps, it was decided that the computational power at disposition was not sufficient to work with this learning method. However, the preliminary results were obtianed with a training of 32 epochs, 32 steps per epoch and a batch size of 4.

In the light of these results, all the efforts were focused on the semi-supervised learning approach.

2) Semi-supervised: The semi-supervised learning can take advantage of the labels available for 118 frames in the data-set (95 in the training set, 22 in the testing set) and hypothetically improve the prediction by adding a training dice score in the loss formula as a second regularization term, to take into account mismatches between labelled images.

The results are summarized in Table II and $\lambda = 5e^{-2}$ for MSE was identified as the most promising value whereas lower values of λ lead to worse outcomes. The different labels (7 different neurons) are impacted differently by the morphing depending on the λ value. If the testing Dice Score was not necessarily improved by VoxelMorph, depending on the label (neuron), the testing normalized cross-correlation value showed that overall, the similarity between the frames

was improved. This would indicate that a single label volume might be enough to analyze an entire movie.

Based on these results, another training with 60 epochs, 250 steps per epoch for λ=$5e^{-2}$ for MSE was initiated in order to guarantee convergence of the loss and further improve the performance. Training NCC with λ=$6e^{-1}$ was also run with similar number of steps to verify that MSE was indeed the most appropriate loss in the context of this work. Their respective overall dice score are respectively 0.29 ($\pm$ 0.27) and 0.30 ($\pm$ 0.28), as for the normalized cross-correlation; respectively 0.57 ($\pm$ 0.24) and 0.54 ($\pm$ 0.22).

The validation segmentation loss was constant at 0, meaning that the 30 frames used for validation were not labelled. Even though, the total validation loss was still converging.

Using personal laptops, with a batch size of 4, each step took approximately 7 seconds. Since the data-set contained a considerable number of frames, many steps were required at each epoch to get an optimal coverage.

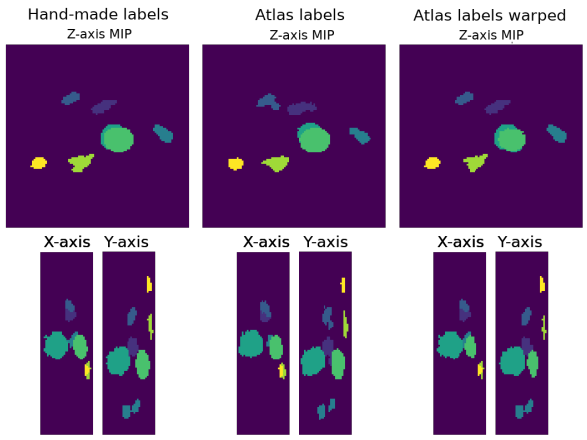

Figure 4: Maximum Intensity Projection on each axis of frame 500 original label, the atlas label and the warped atlas label. The prediction was made by the model MSE, $\lambda = 5e^{-2}$ and the warping generated using the nearest interpolation method.

As observable in Figure 4, the warping transformation managed to get the atlas labels quite close to the volume original labels. In particular, each neurons is clearly distinguishable and no wrong allocation/smearing are visible. The overall structure of the neurons and thus labels is globally maintained.

By using the same reference frame during the entire training, one could say that the model is overfitting the data. This is indeed voluntary as the movies are quite different from each other and would require a full training of the model each to yield accurate warping. By doing so, the model improve its capacity to morph the frames on the reference and extend its labels to the entire data-set.

No cross-validation could be used as the training because despite being faster than other neural networks, it remained too time consuming to be applied with the computational power at disposal.

V. Conclusion

Unfortunately, our supervisor, Professor Rahi wasn't able to provide us with an additional data-set with labelled frames to test this model on unseen data. It would have been useful to know if these learned weights could be applied to another movie or if another model training on the new movie was required.

Future work

Concerning the pre-aligned data, further optimization of the hyperparameter values should be undertaken, especially the learning rate of the model and the γ regularizer of the training Dice score that were not investigated during this project. The UNet shape should also be diagnosed to further optimize neuron detection in worm brains.

Future testing on raw 3D data should be done to expand the proof-of-concept obtained in 2D. Nevertheless, some limitations should still be overcome, as for instance the rotation in 3D during the pre-processing step. As a comparison, the LPBS lab pipeline, based on Jian Vemuri, could also be improved, not only in its accuracy, but also in its running time.

VoxelMorph could also probably replace the Segmentation Convolutional Neuron Network that the LPBS lab is currently using in conjunction of the Jian Vemuri algorithm based pre-alignment. Unfortunately, the output of this work could not be tested with this CNN to eventually assess its performance in the Lab's workflow and should be done later.

With the help of more efficient computational equipment, some longer trainings should be considered to cover the whole data-set and reach better convergence.

References

[1] B. G. et al., "Voxelmorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, feb 2019.

[2] O. Hobert, "Neurogenesis in the nematode Caenorhabditis elegans," *WormBook: The Online Review of C. Elegans Biology*, pp. 1–24, Oct. 2010.