Neural-Bspline for Non-Rigid Image Registration

Shashank Pathak
spathak2@ualberta.ca

Abstract—Non-rigid image registration plays a critical role in medical imaging by facilitating improved analysis of tissue, vessel, and other structural changes arising from variations in image views, acquisition times, or modalities. Traditional registration approaches, such as B-splines, have demonstrated efficiency but often lack the flexibility to capture complex deformations due to their strong local influence. In this work, we introduce a parameterized framework for B-spline-based image registration, termed Neural-Bspline. Rather than directly learning a dense optical flow across the entire image grid, our method focuses on learning a scattered optical flow at randomly sampled control points, which is subsequently interpolated across the image using third-order B-spline interpolation. We train and evaluate our approach with varying numbers of control points and assess its performance using the Dice coefficient on the benchmark Fundus Image Registration Dataset (FIRE) [1]. VoxelMorph [2] was used as a baseline model for comparative analysis. Our results indicate that VoxelMorph outperformed all other models on the FIRE-segmented dataset. However, on the unsegmented FIRE dataset, the performance of our Neural-Bspline method was comparable to that of the VoxelMorph network, demonstrating the potential of our approach for non-rigid registration tasks. The implementation is available at https://github.com/shashankcuber/Neural-B-Spline-Image-Registration/tree/main

1 Introduction

Image registration is a fundamental process in computer vision and medical imaging that involves aligning a moving image (I_m) to a fixed image (I_f). By identifying correspondences between features in the images, the registration process facilitates the integration, comparison, and analysis of datasets captured under varying conditions, such as different times, viewpoints, or imaging modalities [3]. This capability is crucial for numerous applications, including medical diagnostics, where accurate alignment is necessary for reliable interpretation and actionable insights for diagnosis and prognosis.

The registration process typically comprises three core components: a transformation model to describe the spatial relationship between images, an optimization algorithm to iteratively refine the transformation, and a similarity metric to evaluate the quality of alignment. Depending on the nature of the data and the application's requirements, registration methods can be broadly categorized into rigid and non-rigid approaches. While rigid registration assumes a fixed geometric transformation involving translations and rotations, non-rigid registration allows for more complex, non-linear transformations to address local variations. Unlike rigid methods that assume a global transformation, non-rigid techniques address local variations such as tissue deformation, organ motion, or elastic changes. These methods involve creating a dense deformation field that maps corresponding points between the images, allowing for more flexible and precise alignment. This work focuses on non-rigid image registration due to its importance in capturing subtle, localized deformations that rigid methods cannot handle.

In medical imaging, non-rigid registration is indispensable for capturing and aligning fine-grained structural changes. For instance, in ophthalmic imaging, it is crucial to analyze vascular changes between retinal images captured at different times or from different viewpoints. This capability is essential for the diagnosis and treatment of diseases like diabetic retinopathy, where subtle vascular changes can indicate disease progression or treatment efficacy [4]. Traditional methods for non-rigid registration, such as B-splines [5], have been widely used due to their efficiency and ability to model smooth deformations with a high degree of continuity. However, these methods are limited in flexibility, as their local influence is tightly constrained by the choice of control points, making it challenging to capture highly complex deformations. In recent years, neural network-based approaches have emerged as powerful tools for learning dense optical flow or deformation fields directly from image data [6]. These methods use neural networks to model complex spatial relationships. Despite their advantages, existing neural methods typically learn dense optical flow for the entire image grid.

In this work, we propose a parameterized framework for B-spline-based non-rigid image registration, designed for unsupervised learning. Unlike conventional methods that compute dense deformation fields, we focus on learning sparse optical flow at randomly sampled grid points, referred to as control points. These control points serve as the basis for modeling deformations, allowing us to capture complex local variations without the computational overhead of dense grid learning. To warp the I_m, the sparse optical flow vectors at control points are interpolated across the entire image grid using third-order B-spline interpolation. This interpolation method ensures smooth and continuous deformations with C_2 continuity, providing a balance between flexibility and regularization. By integrating the strengths of traditional B-splines with neural network paradigms, our proposed Neural-B-spline framework offers a flexible solution for non-rigid image registration in medical imaging.

2 DATA

The FIRE (Fundus Image Registration Dataset) dataset [1] is used for applying the retinal vessel registration algorithm. It is a publicly available benchmark dataset containing 134 fixed-moving pairs of retinal fundus images with varying degrees of image distortions and intensity variations. The images were captured using a Nidek AFC-210 fundus camera at a resolution of 2912 x 2912 pixels, with a 45-degree field of view in the x and y dimensions. Moreover, the dataset also contains annotated ground truth values for evaluation in the form of point correspondences between the image pairs. Blood vessels being the most critical component of the retinal images. Hence, we used FUNDUS-segmented dataset which was created as a part of CMPUT-617 course project by training a U-net segmentation model on the Digital Retinal Images for Vessel Extraction (DRIVE) dataset [1] [2]. Fig. 1 illustrates the FUNDUS and FUNDUS-segmented dataset.

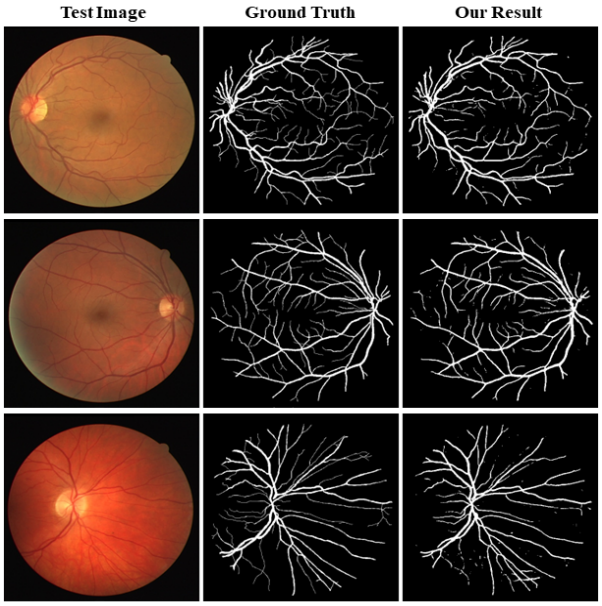

Figure 1: The left most column displays the images from the FUNDUS dataset, the middle column is from the DRIVE dataset and the last column is predicted segmentation mask for the FUNDUS-segmented dataset.

3 METHOD

The proposed method for non-rigid image registration combines sparse control point sampling, neural network-based displacement prediction, and B-spline interpolation to generate a dense deformation field that aligns a moving image (I_m) to a fixed image (I_f). Neural-Bspline algorithm is described in 1.

1. https://drive.google.com/drive/folders/1OJCGpIMxiglWamBSOtOkQfon52WkMPTF?usp=sharing

2. https://drive.google.com/file/d/11LOqYH1tz9XBjASFpVQN-ouYmkzsJpeK/view?usp=sharing

3.1 Control Point Sampling

Control points are uniformly sampled from the moving image grid to reduce the computational complexity of learning displacements for all grid points. The image is divided into $n_p \times n_p$ patches, where $n_p = \sqrt{n_{\text{control}}}$, and one control point is randomly selected from each patch:

$$\mathbf{p}_i = \text{random}(P_{ij}), \tag{1}$$

where P_{ij} denotes the spatial region of the i, j-th patch. The sampled control points for a batch B are represented as $\mathbf{P}_c \in \mathbb{R}^{B \times n_{\text{control}} \times 2}$.

The sampled control points are then normalized to the range $[-1, 1]$ to make it compatible with the neural network training.

$$\mathbf{p}_i^{\text{norm}} = \left(\frac{\mathbf{p}_i[0]}{H} \cdot 2 - 1, \frac{\mathbf{p}_i[1]}{W} \cdot 2 - 1 \right), \tag{2}$$

where H and W are the height and width of the image.

3.2 Feature Extraction

The feature extraction step involved creating embedding for control points and then fusing it with the image pair feature maps.

Coordinate Embedding: The normalized control points are passed through a coordinate embedding network which is a multi layered perceptron (MLP) to generate high-dimensional embeddings:

$$\mathbf{F}_c = \text{CoordEmbed}(\mathbf{P}_c^{\text{norm}}), \tag{3}$$

where $\mathbf{F}_c \in \mathbb{R}^{B \times n_{\text{control}} \times d}$ and d is the embedding dimension taken as 64 in this work.

Image Features: A feature extractor network is a convolution neural network based encoder to encode features from the concatenated image pair:

$$\mathbf{F}_i = \text{FeatureExtractor}([I_m, I_f]), \tag{4}$$

where $\mathbf{F}_i \in \mathbb{R}^{B \times d \times H \times W}$.

Sampled Features: Sparse features are sampled from $\mathbf{F}_i$ at the control points using grid sampling:

$$\mathbf{F}_s = \text{GridSample}(\mathbf{F}_i, \mathbf{P}_c^{\text{norm}}), \tag{5}$$

where $\mathbf{F}_s \in \mathbb{R}^{B \times n_{\text{control}} \times d}$.

3.3 Displacement Prediction

$\mathbf{F}_c$ and $\mathbf{F}_s$ are concatenated and passed through a displacement network which is a MLP to predict sparse displacements for the control points:

$$\Delta \mathbf{u}_i = \text{DisplacementNet}([\mathbf{F}_c, \mathbf{F}_s]), \tag{6}$$

where $\Delta \mathbf{u}_i \in \mathbb{R}^{B \times n_{\text{control}} \times 2}$.

3.4 B-spline Interpolation

The sparse displacements $\Delta \mathbf{u}_i$ are interpolated to a dense deformation field ϕ using cubic B-spline interpolation. For each pixel (x, y) in the image, the deformation is computed as:

$$\phi(x, y) = \frac{\sum_{i=1}^{n_{\text{control}}} B_k(x - \mathbf{p}_{i,x}) \cdot B_k(y - \mathbf{p}_{i,y}) \cdot \Delta \mathbf{u}_i}{\sum_{i=1}^{n_{\text{control}}} B_k(x - \mathbf{p}_{i,x}) \cdot B_k(y - \mathbf{p}_{i,y}) + \epsilon}, \tag{7}$$

where B_k is the $k = 3$ order B-spline basis function, and $\epsilon = 1e - 6$ is a small constant for avoiding division by zero.

3.5 Image Warping

The moving image is finally warped using ϕ to align it with the fixed image using the following equation.

$$I_m^{\phi}(x, y) = I_m(\phi(x, y)). \tag{8}$$

Algorithm 1 Neural B-Spline Algorithm

Require: I_m, I_f, n_{control}, Image size (C, H, W)

1: Divide I_m into $n_p \times n_p$ patches; uniformly sample n_{control} control points $\mathbf{P}_c$.
2: Normalize control points: $\mathbf{P}_c^{\text{norm}} \leftarrow \text{normalize}(\mathbf{P}_c)$.
3: Control points features: $\mathbf{F}_c \leftarrow \text{CoordEmbed}(\mathbf{P}_c^{\text{norm}})$.
4: Extract image features: $\mathbf{F}_i \leftarrow \text{FeatureExtractor}([I_m, I_f])$.
5: Sample image features: $\mathbf{F}_s \leftarrow \text{GridSample}(\mathbf{F}_i, \mathbf{P}_c^{\text{norm}})$.
6: $\Delta\mathbf{u}_i \leftarrow \text{DisplacementNet}([\mathbf{F}_c, \mathbf{F}_s])$.
7: Interpolate: $\phi(x, y) \leftarrow \text{B-Spline}(\mathbf{P}_c, \Delta\mathbf{u}_i)$.
8: Warping: $I_m^{\phi}(x, y) \leftarrow I_m(\phi(x, y))$.
9: **return** ϕ, I_m^{ϕ}

4 Result

4.1 Training And Evaluation Setup

We evaluated our approach Neural-Bspline for 10, 100, and 200 control points (*n_control*) which we refer to as Neural-Bspline-10-num_points, Neural-Bspline-100-num_points, and Neural-Bspline-200-num_points respectively. For the baseline, we trained VoxelMorph architecture [3]. On the FIRE-segmented dataset the models were trained for 156 epochs, whereas on the FIRE dataset they were trained for 10 epochs. The remaining training hyperparameters are provided in Table 1. We performed training on two

Parameters	Value
Batch Size	4
Learning Rate	0.0001
Optimizer	SGD
Evaluation Metric	Dice Score
Loss Function	Mean Squared Error (MSE)

Table 1: Hyper-parameter values used for training

datasets mentioned in section 2. For evaluation, a dice score was used which is calculated using the equation below.

$$DICE = \frac{2 \times |X \cap Y|}{|X| + |Y|}, \tag{9}$$

where X is the predicted set of pixels, Y is the ground truth segmentation mask and $|\,|$ denotes the cardinality of the set.

4.2 Quantitative Result

On the FIRE-segmented dataset, we observe similar loss convergence for all different control points in the Neural-Bspline method illustrated in Fig. 2. For the Neural-Bspline-200-num_points network, the training was found to be unstable as compared to other 10 and 100 control points Neural-Bspline. On the evaluation metric for the training and validation set, all three Neural-Bspline networks converged to the same point with the best dice score of

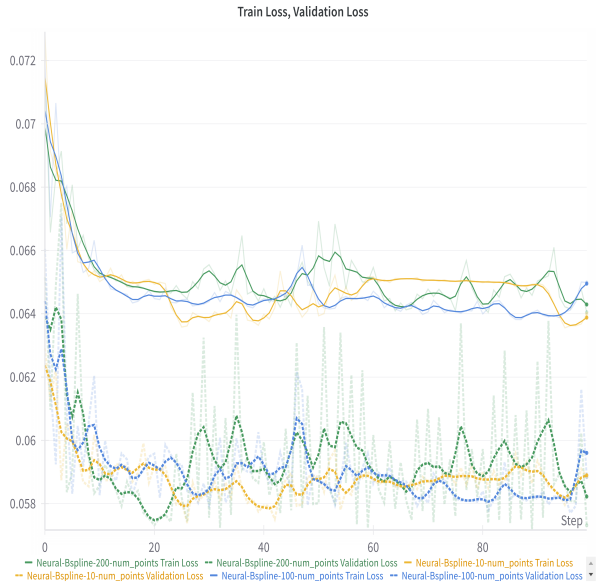

Figure 2: Training and Validation Loss curve of Neural-Bspline on the Fire-segmented dataset

Figure 3: Training and Validation dice score convergence of Neural-Bspline on the Fire-segmented dataset

0.27 achieved by the Nerual-Bspline trained for 200 control points (Fig. 3).

Voxelmoprh which is our baseline performed $2\times$ better with a dice score of 0.54 on the validation set of FIRE-segmented dataset (Fig. 4).

On the FIRE dataset, we observed that our approach and VoxelMorph had a similar dice score on the validation set (Fig. 5).

3. https://github.com/Hsankesara/VoxelMorph-PyTorch

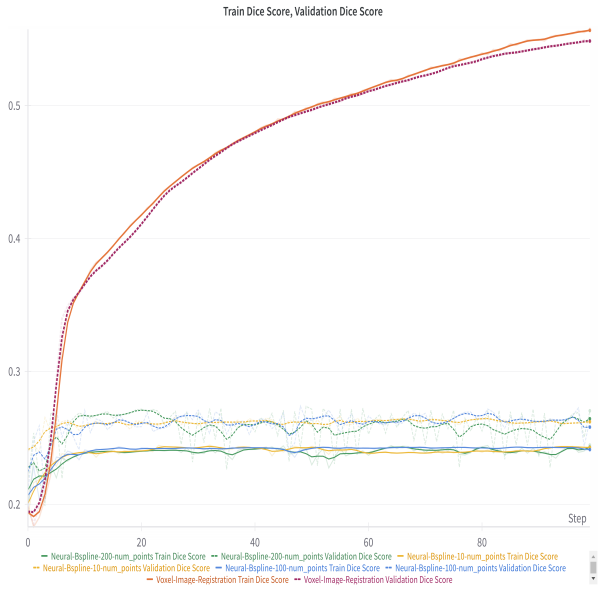

Figure 4: Training and Validation dice score convergence comparison with baseline on the Fire-segmented dataset

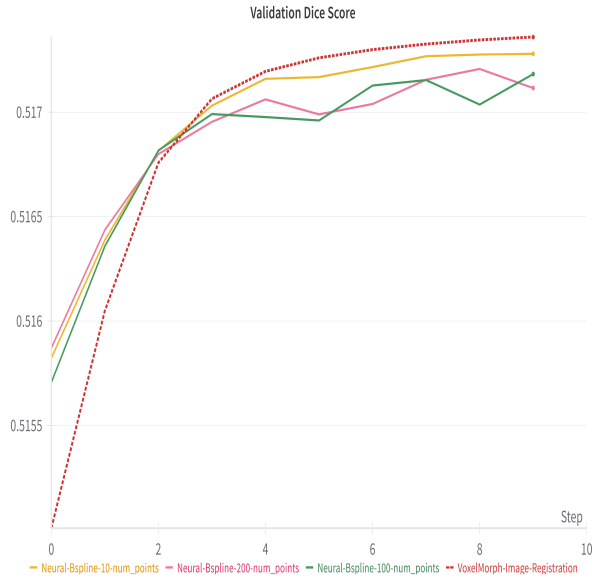

Figure 5: Training and Validation dice score convergence comparison with baseline on the Fire dataset

4.3 Qualitative Result

The qualitative analysis for best dice score sample on the validation set of FIRE-segmented dataset is illustrated in Fig. 6 for Neural-Bspline-200-num_points and in Fig. 7 for the baseline VoxelMorph.

References

[1] Carlos Hernandez-Matas, Xenophon Zabulis, Areti Triantafyllou, Panagiota Anyfanti, Stella Douma, and Antonis A Argyros. Fire: fundus image registration dataset. *Modeling and Artificial Intelligence in Ophthalmology*, 1(4):16–28, 2017.

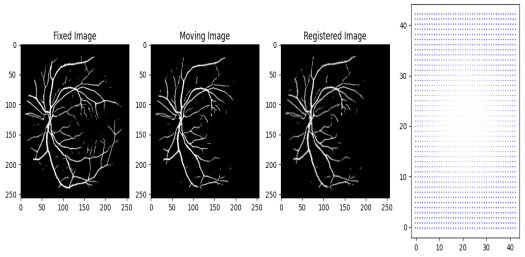

Figure 6: Neural-Bspline for 200 control points visual result on the FIRE-segmented dataset

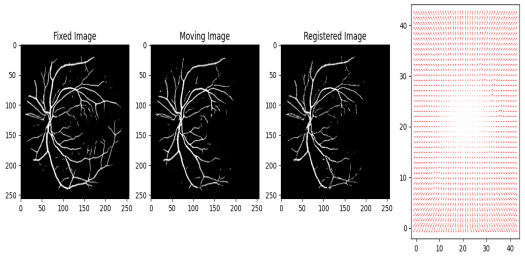

Figure 7: Voxelmorph visual result on the FIRE-segmented dataset

[2] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

[3] Ronald WK So and Albert CS Chung. A novel learning-based dissimilarity metric for rigid and non-rigid medical image registration by using bhattacharyya distances. *Pattern Recognition*, 62:161–174, 2017.

[4] Emily Dawn Cole, Eduardo Amorim Novais, Ricardo Noguera Louzada, and Nadia K Waheed. Contemporary retinal imaging techniques in diabetic retinopathy: a review. *Clinical & experimental ophthalmology*, 44(4):289–299, 2016.

[5] Zufeng Wu, Tian Lan, Jiang Wang, Yi Ding, and Zhiguang Qin. Medical image registration using b-spline transform. *Int. J. Simul. Syst. Sci. Technol*, 17(48):1–1, 2016.

[6] Yaoying She, Mei Zhou, Qingli Li, and Li Sun. Retinal image registration based on features of vessel-segmented image. In *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE, 2021.