A NEW DEEP LEARNING MODEL FOR DIFFEOMORPHIC DEFORMABLE IMAGE

REGISTRATION PROBLEMS

By

MR. Apinan SANONGSIN

A Thesis Submitted in Partial Fulfillment of the Requirements

for Master of Science MATHEMATICS

Department of MATHEMATICS

Silpakorn University

Academic Year 2022

โมเดลการเรียนรู้เชิงลึกชนิดใหม่สำหรับปัญหาการลงทะเบียนภาพอนุพันธสัณฐาน

โดย
นายอภินันท์ สนองสินธุ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาคณิตศาสตร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2565
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

A NEW DEEP LEARNING MODEL FOR DIFFEOMORPHIC DEFORMABLE

IMAGE REGISTRATION PROBLEMS

By

MR. Apinan SANONGSIN

A Thesis Submitted in Partial Fulfillment of the Requirements

for Master of Science MATHEMATICS

Department of MATHEMATICS

Silpakorn University

Academic Year 2022

Title	A New Deep Learning Model for Diffeomorphic Deformable Image Registration Problems
By	MR. Apinan SANONGSIN
Field of Study	MATHEMATICS
Advisor	Assistant Professor Noppadol Chumchob, Ph.D.

Faculty of Science, Silpakorn University in Partial Fulfillment of the Requirements for the Master of Science
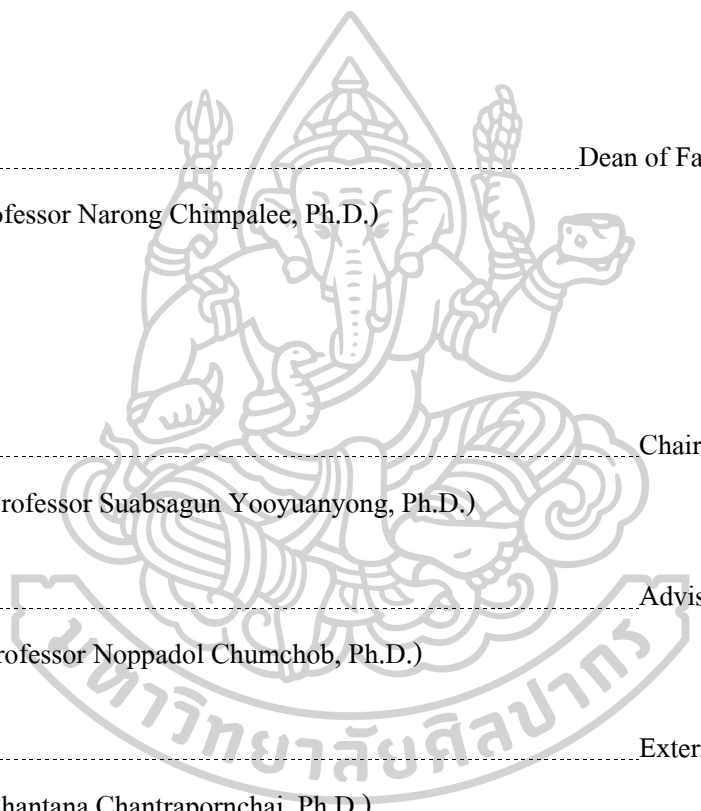
.... Dean of Faculty of Science

(Assistant Professor Narong Chimpalee, Ph.D.)

Approved by

.... Chair person

(Associate Professor Suabsagun Yooyuanyong, Ph.D.)

.... Advisor

(Assistant Professor Noppadol Chumchob, Ph.D.)

.... External Examiner

(Professor Chantana Chantrapornchai, Ph.D.)

630720018 : Major MATHEMATICS

Keyword : Deep Learning, Image Registration, Diffeomorphic Transformation, U-Net, Chest X-rays Image, Unsupervised Learning

MR. Apinan SANONGSIN : A New Deep Learning Model for Diffeomorphic Deformable Image Registration Problems Thesis advisor : Assistant Professor Noppadol Chumchob, Ph.D.

Diffeomorphic deformable image registration is an essential technique in medical image analysis, aiming to find a smooth and invertible mapping between images to align their corresponding anatomical structures. The importance of diffeomorphic transformations lies in their ability to preserve topology, maintain smooth and continuous deformations, and allow for invertibility, ensuring accurate and physiologically plausible results.

Traditional variational methods have been used for diffeomorphic deformable image registrations; however, they can be computationally expensive and require extensive parameter tuning. In contrast, deep learning approaches have shown remarkable success in various image processing tasks due to their ability to learn complex and hierarchical features. These deep learning models can offer improved efficiency, robustness, and generalization in image registration tasks.

In this thesis, we present a novel diffeomorphic deformable image registration model that incorporates a novel diffeomorphic regularization loss with an unsupervised learning strategy. Diffeomorphic regularization enforces smooth and invertible transformations, leading to improved registration results. Our proposed model outperforms other models using diffeomorphic regularization losses in terms of the relative sum of square differences and maintains the topological properties of images, demonstrating its potential in various medical imaging applications.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction to Image Registration

Image registration is a process in computer vision and image processing that involves aligning and overlaying two or more images of the same scene taken at different times, from different viewpoints, or by different sensors. The goal is to bring these images into a common coordinate system so that they can be compared, analyzed, or combined to create a new image with enhanced information content [5].

Image registration has numerous applications across various fields, enhancing our understanding of complex data and improving decision-making processes. In medical imaging, it enables the integration of images from different modalities like MRI, CT, and PET scans, providing a comprehensive view of a patient's anatomy and pathology for more accurate diagnoses and treatment planning [30, 31, 36]. In remote sensing, image registration facilitates the monitoring of land-use changes, urban growth, and natural disasters by merging images taken at different times [12]. It also plays a crucial role in computer vision and robotics, assisting in tasks such as object recognition, tracking, and navigation by combining images from multiple sensors or viewpoints. These diverse applications demonstrate the importance and versatility of image registration in addressing real-world challenges.

In this thesis, we focus on the process of image registration involving a pair of images: a reference image and a template image. Image registration entails aligning the template image with the reference image by applying a geometric transformation, making the template image to become similar to the reference image, as depicted in
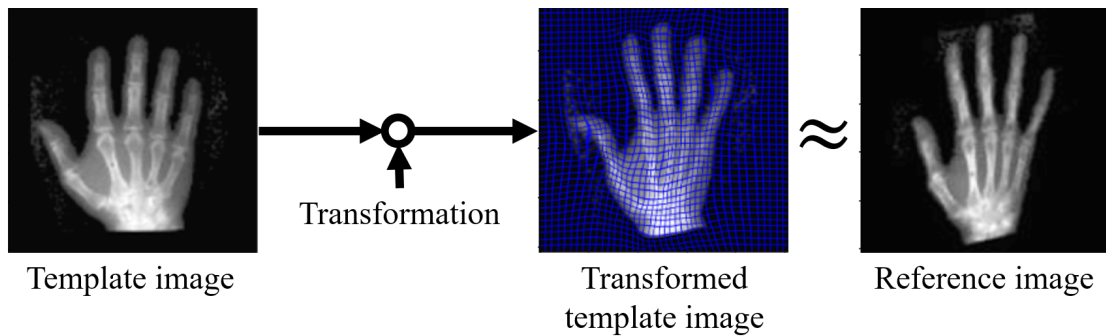
Figure 1.1: The transformed template image and the reference image are spatially matched according to an optimal geometric transformation.

Figure 1.1.

The objective of image registration is to search for a suitable transformation that enables the template image to become similar to the reference image, as previously mentioned. This transformation should be meaningful, such as a deformable transformation, which may not be a linear transformation like translation, rotation, scaling, or shearing. As illustrated in Figure 1.1, the curved line demonstrates the transformation of the grid lines. Additionally, the transformation can involve diffeomorphic transformations that are smooth and invertible, further enhancing the registration process's accuracy and relevance.

Over the past few decades, various traditional methods, such as variational approaches that are non-learning-based, have achieved success in image registration, as demonstrated by examples in [8, 10, 17, 20, 37, 42]. However, since these variational methods rely on mathematical models and optimization techniques to establish the transformation, they often face significant limitations in practical applications due to their intensive and time-consuming computational optimization requirements.

Recently, with the rapid advancements in technology, deep learning (DL) methods that aim to emulate human brain processes have been employed in various fields and problems, including image registration, as reviewed in [19, 22]. Deep learning techniques

can learn to address the problem through advanced training processes and then apply pre-trained models, which are models trained using suitable datasets and algorithms. These models can predict solutions in practical applications within a few seconds or even less, significantly enhancing efficiency.

While deep learning models are efficient, they sometimes face challenges. Some of these models do not account for diffeomorphic transformations, which are essential for specific medical image registration tasks, such as lung image registration between inhalation and exhalation. Consequently, the results may be inaccurate and unsuitable for practical applications.

Considering the previously discussed information, the objective of this thesis is to propose a novel deep-learning model for diffeomorphic deformable image registration problems. This new model specifically incorporates diffeomorphic transformations to enhance accuracy and applicability.

1.2 Thesis Outline

This thesis focuses on the development of a novel diffeomorphic deformable image registration model and its evaluation on synthetic and medical images. The outline of the thesis is as follows:

1. Introduction

   - Briefly introduces the topic of image registration, its importance in medical imaging, and the challenges of deformable image registration.

   - Presents the motivation and objective of the study, which is to develop a deep-learning model for diffeomorphic deformable image registration.

   - Provides an overview of the thesis organization.

2. Mathematical Frameworks and Deep Learning Models for Image Registration

- Mathematics for Image Registration: Explains the fundamental concepts and equations involved in image registration.

- Image Registration Frameworks: Reviews image registration frameworks, focusing on data similarity measures and regularization techniques.

- Deep Learning Models: Presents an overview of deep learning approaches for image registration, focusing on convolutional neural networks (CNNs).

  - Convolution Layer: Describes the function of convolution layers in CNNs.

  - Activation Functions: Explores various activation functions used in CNNs.

  - Pooling Layer: Explains the role of pooling layers in CNNs.

- Loss Function: Discusses the concept of loss functions in the context of image registration, focusing on supervised and unsupervised learning techniques.

3. Proposed Diffeomorphic Deformable Image Registration Model

- Introduction: Introduces the proposed deep learning model for diffeomorphic deformable image registration.

- Neural Network Architecture: Describes the architecture of the proposed deep learning model for diffeomorphic deformable image registration.

  - Architecture Components: Details the components of the neural network architecture.

  - Deformation Prediction: Explains how the deformation is predicted from the network output.

- Loss Function: Presents the components of the proposed loss function, including data similarity, smoothness regularization, and diffeomorphic regu-

larization.

- Discretization: Discusses the discretization of the loss function terms and their computation.

4. Experiments and Results of Synthetic Images

- Synthesizing Image Pairs: Describes the process of generating synthetic image pairs for evaluation.

- Dataset Splitting: Explains how the synthetic image dataset is split into training, validation, and test sets.

- Evaluation Metrics: Introduces the evaluation metrics used for assessing the performance of the proposed deep learning model on synthetic images.

- Experimental Setup for Synthetic Images: Details the implementation and parameter settings for synthetic image experiments.

- Results for Synthetic Images: Registration results of the proposed DL model on the validation set and test set.

5. Experiments and Results of Real Medical Images

- Dataset: Introduces the medical image dataset used for evaluation.

- Pre-processing: Describes the pre-processing steps applied to the medical image dataset.

- Evaluation Metrics: The same evaluation metrics used in synthetic images are also used in medical images.

- Experimental Setup for Medical Images: Details the implementation and parameter settings for medical image experiments.

- Results for Medical Images: Results of registration using the test set and validation set for the proposed deep learning model.

6. Conclusion and Future Work

- Conclusion: Summarizes the registration results of the proposed deep learning model for diffeomorphic deformable image registration.

- Future Work: Outlines potential avenues for future research.

MATHEMATICAL FRAMEWORKS AND DEEP LEARNING MODELS FOR IMAGE REGISTRATION

In this chapter, we will discuss various concepts and methods related to image registration using deep learning techniques. We will begin by exploring the mathematical definitions of images and diffeomorphic transformations, which play a crucial role in image registration. Next, we will delve into the image registration framework, highlighting the key components and steps involved in aligning images.

Subsequently, we will provide an overview of deep learning models, focusing on CNNs and their components, such as convolutional layers, activation functions, and pooling layers. These building blocks allow CNNs to learn complex patterns and features from input images, making them particularly well-suited for image registration tasks. Lastly, we will discuss utilizing deep learning models to solve image registration problems by minimizing a loss function, which quantifies the difference between the predicted outputs and the actual target values.

2.1   Mathematics for Image Registration

We consider an image as a function mapping a two-dimensional spatial coordinate to an intensity value. We have a reference image R and a template image T for image registration. These images can be represented as functions $R, T : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ denote the spatial domain of the images and $\mathbb{R}$ represents the range of intensity values. The goal is to find a transformation $\varphi : \Omega \to \Omega$ such that the transformed template image $T(\varphi(\mathbf{x}))$ aligns with the reference image $R(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2) \in \Omega$ are the spatial coordinates.

In our study, we consider the transformation $\boldsymbol{\varphi}$ to be a diffeomorphic deformable transformation. For deformable transformations, we define the displacement or deformation $\boldsymbol{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))$, which represents the movement of image coordinates. Consequently, $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{x} + \boldsymbol{u}(\mathbf{x})$, and solving for $\boldsymbol{\varphi}(\mathbf{x})$ or $\boldsymbol{u}(\mathbf{x})$ is equivalent. As for diffeomorphic transformations, the Jacobian determinant, denoted as $\det(\nabla\boldsymbol{\varphi}(\mathbf{x}))$, where $\nabla\boldsymbol{\varphi}(\mathbf{x})$ is the Jacobian matrix of the transformation at point $\mathbf{x} \in \Omega$, should be positive [2, 21]. That is for each $\mathbf{x} \in \Omega$,

$$
\begin{aligned}
\det(\nabla\boldsymbol{\varphi}(\mathbf{x})) &= \begin{vmatrix} \frac{\partial\varphi_1(\mathbf{X})}{\partial x_1} & \frac{\partial\varphi_1(\mathbf{X})}{\partial x_2} \\ \frac{\partial\varphi_2(\mathbf{X})}{\partial x_1} & \frac{\partial\varphi_2(\mathbf{X})}{\partial x_2} \end{vmatrix} \\
&= \begin{vmatrix} \frac{\partial(x_1+u_1(\mathbf{X}))}{\partial x_1} & \frac{\partial(x_1+u_1(\mathbf{X}))}{\partial x_2} \\ \frac{\partial(x_2+u_2(\mathbf{X}))}{\partial x_1} & \frac{\partial(x_2+u_2(\mathbf{X}))}{\partial x_2} \end{vmatrix} \\
&= \left(\frac{\partial u_1(\mathbf{x})}{\partial x_1} + 1\right)\left(\frac{\partial u_2(\mathbf{x})}{\partial x_2} + 1\right) - \left(\frac{\partial u_1(\mathbf{x})}{\partial x_2}\right)\left(\frac{\partial u_2(\mathbf{x})}{\partial x_1}\right) \\
&> 0.
\end{aligned} \tag{2.1}
$$

The positive Jacobian determinant ensures the preservation of local areas and prevents folding or tearing in the registered images [25, 36].

2.2 Image Registration Frameworks

From this point forward, we will use the term "deformed template image" instead of "transformed template image" to emphasize that the transformation is deformable. To assess the difference between the reference image R and the deformed template image $T_{\boldsymbol{u}} = T(\mathbf{x} + \boldsymbol{u}(\mathbf{x})) = T(\boldsymbol{\varphi}(\mathbf{x}))$, an appropriate data similarity term, $\mathcal{D}(T_{\boldsymbol{u}}, R)$, should be introduced. Consequently, the image registration problem can be formulated by minimizing the following data similarity term:

$$
\min_{\boldsymbol{u} \in \mathcal{U}} \{\mathcal{D}(T_{\boldsymbol{u}}, R)\}, \tag{2.2}
$$

where $\boldsymbol{u}$ is searched over an appropriate function space $\mathcal{U}$.

There are several data similarity terms used in image registration frameworks. For grayscale images R and T, we will discuss two commonly used similarity measures: the Sum of Squared Differences (SSD) and Normalized Cross-Correlation (NCC).

- **Sum of Squared Differences (SSD):** The Sum of Squared Differences (SSD) is a widely used similarity measure for image registration. It calculates the squared differences between the intensities of corresponding pixels in the deformed template image $T_{\boldsymbol{u}}$ and the reference image R. The SSD is defined as:

$$\mathcal{D}_{\mathrm{SSD}}(T_{\boldsymbol{u}}, R) = \int_{\Omega} \left(T_{\boldsymbol{u}}(\mathbf{x}) - R(\mathbf{x})\right)^2 d\mathbf{x}. \tag{2.3}$$

SSD is computationally simple and relatively fast, which makes it suitable for real-time applications. It performs well when the intensity values of the images are linearly related. However, it is sensitive to changes in intensity due to factors such as lighting conditions or imaging modalities. Thus, it may not work well in situations where the images have significant intensity differences.

- **Normalized Cross-Correlation (NCC):** Normalized Cross-Correlation (NCC) is another popular similarity measure used in image registration. It measures the correlation between the intensities of corresponding pixels in the deformed template image $T_{\boldsymbol{u}}$ and the reference image R. The NCC is defined as:

$$\mathcal{D}_{\mathrm{NCC}}(T_{\boldsymbol{u}}, R) = -\frac{\int_{\Omega} \left(T_{\boldsymbol{u}}(\mathbf{x}) - \bar{T}_{\boldsymbol{u}}\right) \left(R(\mathbf{x}) - \bar{R}\right) d\mathbf{x}}{\sqrt{\int_{\Omega} \left(T_{\boldsymbol{u}}(\mathbf{x}) - \bar{T}_{\boldsymbol{u}}\right)^2 d\mathbf{x} \int_{\Omega} \left(R(\mathbf{x}) - \bar{R}\right)^2 d\mathbf{x}}}, \tag{2.4}$$

where $\bar{T}_{\boldsymbol{u}}$ and $\bar{R}$ are the mean intensity values of $T_{\boldsymbol{u}}$ and R, respectively. NCC is less sensitive to intensity changes compared to SSD, making it more robust to variations in lighting conditions or imaging modalities. It also works well when the images have different intensity distributions. However, NCC is computationally

more complex than SSD, which can result in longer processing times. It may be less accurate when the images have non-linear intensity relationships or when the images are misaligned.

Nonetheless, it is widely recognized that merely minimizing the data similarity term in the previous (2.2) can result in an ill-posed problem, as it does not guarantee the uniqueness and continuity of the solution [32, 38]. To address this challenge, regularization is essential. By combining the data similarity term and a regularization term, the image registration problem can be well-posed as minimizing the following functional:

$$\min_{\boldsymbol{u} \in \mathcal{U}} \{\mathcal{D}(T_{\boldsymbol{u}}, R) + \alpha \mathcal{R}(\boldsymbol{u})\}, \tag{2.5}$$

where $\mathcal{R}(\boldsymbol{u})$ serves as the regularization term that eliminates irregular and undesired solutions, while $\alpha > 0$ is a regularization parameter that balances the two terms.

In image registration frameworks, various regularization terms are often employed to enforce smoothness constraints. In this thesis, we will focus on three widely-used smoothness regularization terms, namely Total Variation (TV), Diffusion, and Curvature regularizations.

- **Total Variation (TV) Regularization [18, 23]:**

$$\mathcal{R}_{\mathrm{TV}}(\boldsymbol{u}) = \int_{\Omega} \sum_{l=1}^{2} |\nabla u_l(\mathbf{x})| d\mathbf{x}. \tag{2.6}$$

TV regularization promotes piecewise smooth solutions. It is particularly effective in handling non-smooth registration problems. However, TV regularization can sometimes introduce staircase artifacts in the estimated deformations, especially when the regularization parameter is not well-tuned. It may also lead to slow convergence rates in the optimization process due to the non-differentiability of $\mathcal{R}_{\mathrm{TV}}(\boldsymbol{u})$.

- **Diffusion Regularization [16]:**

$$\mathcal{R}_{\mathrm{diff}}(\boldsymbol{u}) = \int_{\Omega} \sum_{l=1}^{2} |\nabla u_l(\mathbf{x})|^2 \, d\mathbf{x}. \tag{2.7}$$

Diffusion regularization is a relatively simple and computationally efficient method that promotes smooth solutions. It is easy to implement and is suitable for applications where smooth deformations are desired. While diffusion regularization is computationally efficient, it can over-smooth the deformations, leading to less accurate registration of images with complex deformations [9].

- **Curvature Regularization [17]:**

$$\mathcal{R}_{\mathrm{curv}}(\boldsymbol{u}) = \int_{\Omega} \sum_{l=1}^{2} |\Delta u_l(\mathbf{x})|^2 \, d\mathbf{x} \tag{2.8}$$

where Δ is the Laplacian operator. The advantage of curvature regularization is its capability to accommodate affine transformations - processes encompassing operations such as translation, scaling, rotation, and shearing - without imposing penalties for their utilization [29]. Although curvature regularization can handle affine transformations, it is computationally more expensive due to the need to compute second-order derivatives.

2.3 Deep Learning Models

In recent years, deep learning has emerged as a powerful tool for addressing various challenges in the field of image registration [7, 19, 22]. Leveraging the power of artificial neural networks, deep learning models are capable of learning complex relationships between images, enabling them to discover the optimal transformation parameters for aligning images more accurately and efficiently. By training on large datasets, deep learning models can capture intricate patterns and structures within the images, making them particularly suitable for complex registration tasks. Additionally, the application
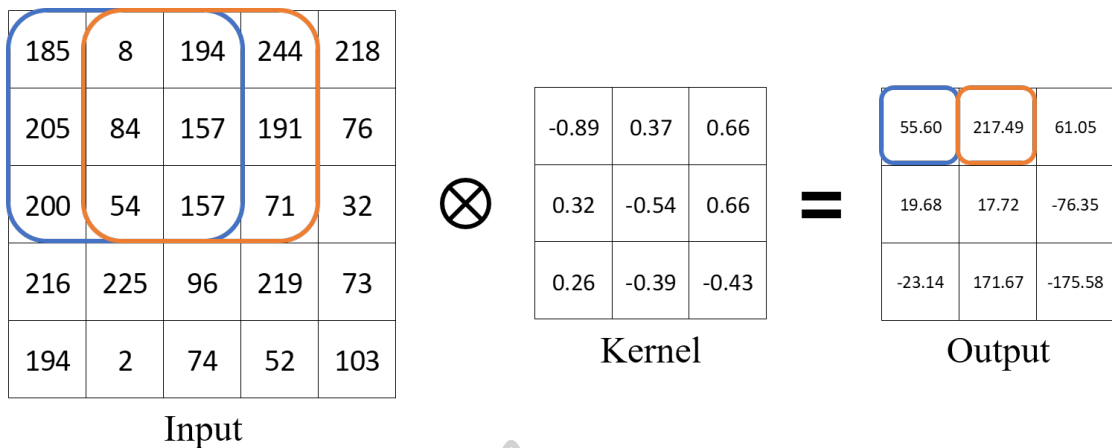
Figure 2.1: Convolution in two-dimension with a kernel size of 3×3 and a stride of (1,1)

of deep learning in image registration has shown promising results in various domains, including medical imaging, remote sensing, and computer vision, thereby revolutionizing the way image registration problems are approached and solved.

2.3.1 Convolutional Neural Networks [1, 41]

Convolutional Neural Network or CNN, a special type of deep learning architecture, is particularly adept at extracting local features and hierarchical representations from images, making them well-suited for image registration tasks. By employing convolutional layers, followed by activation functions and pooling layers, CNNs can learn robust feature representations for both template and reference images. These learned features can then be utilized to compute the optimal transformation parameters for accurate image alignment.

Convolutional Layer

A convolutional layer can be composed of many convolutional kernels. Convolutional kernels, also called filters or masks, are small matrices of weights that slide over the input, which is called stride, performing element-wise multiplication and summing the results.

Figure 2.1 shows an example of convolutions in CNNs. The convolution operation, denoted by $\bigotimes$, is a mapping from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}^2$. In the given example in Figure 2.1, the values 55.60 and 217.49 are derived from the sum of respective input values, i.e.,

$$
\begin{aligned}
55.60 =& 185(-0.89) + 8(0.37) + 194(0.66) \\
& + 205(0.32) + 84(-0.54) + 157(0.66) \\
& + 200(0.26) + 54(-0.39) + 157(-0.43)
\end{aligned}
$$

and

$$
\begin{aligned}
217.49 =& 8(-0.89) + 194(0.37) + 244(0.66) \\
& + 84(0.32) + 157(-0.54) + 191(0.66) \\
& + 54(0.26) + 157(-0.39) + 71(-0.43).
\end{aligned}
$$

A stride of (1,1) is used, meaning that the kernel slides across the input with a step of 1 unit in both vertical and horizontal directions. It is important to note that the output size is smaller than the input size. If we want to maintain the same size of the output as the input, we can apply padding. Padding is the process of extending the input such as adding zeros to all four sides, which effectively preserves the spatial dimensions after the convolution operation.

Activation Functions [34]

Activation functions in neural networks are essential components that introduce non-linearity into the models, enabling them to learn complex relationships and patterns in the data. Here, we present three common activation functions and their formulas:

- **Sigmoid Function σ:**

$$
\sigma(x) = \frac{1}{1 + e^{-x}}
$$

The sigmoid function maps input values to a range between 0 and 1, making it suitable for interpreting output as probabilities. It has a smooth gradient, which can be beneficial during training. The sigmoid function suffers from the vanishing gradient problem, which occurs when the gradients become very small during training, leading to slow learning or convergence.

- **Hyperbolic Tangent Function (tanh):**

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.9}$$

Similar to the sigmoid function, the tanh function has a smooth gradient. However, it maps input values to a range between -1 and 1, making it zero-centered. The tanh function still suffers from the vanishing gradient problem, albeit to a lesser extent than the sigmoid function.

- **Rectified Linear Unit (ReLU):**

$$\text{ReLU}(x) = \max(0, x) \tag{2.10}$$

The ReLU function is computationally efficient and has become the default choice for many deep-learning architectures. It mitigates the vanishing gradient problem and helps to accelerate convergence during training. However, the ReLU function suffers from the dying ReLU problem, where neurons can become inactive and stop learning if their weights are updated such that the input to the ReLU function is always negative.

Figure 2.2 shows plots of three activation functions: sigmoid (σ), hyperbolic tangent (tanh), and rectified linear unit (ReLU) in the same coordinate.
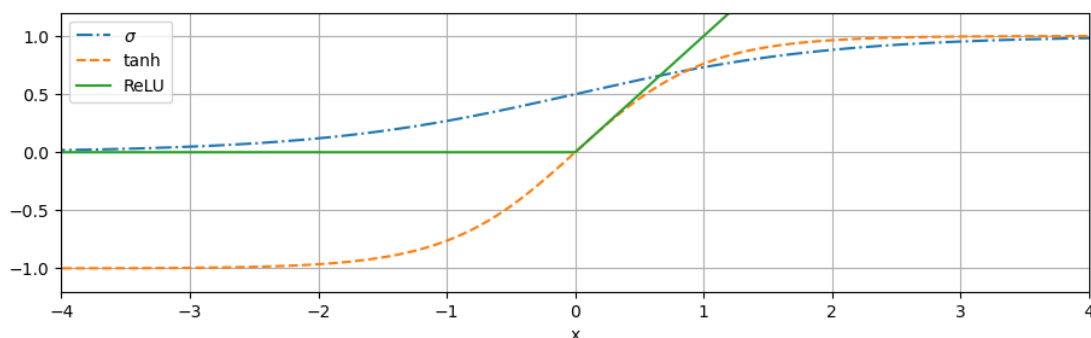
Figure 2.2: Plots of three activation functions: sigmoid (σ), hyperbolic tangent (tanh), and rectified linear unit (ReLU).
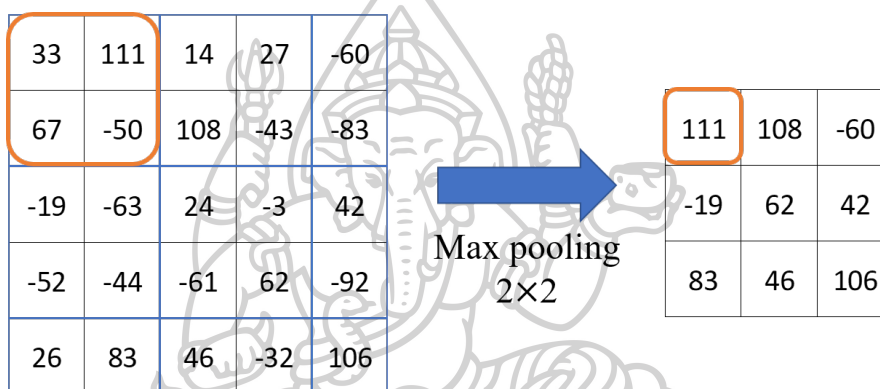
Figure 2.3: Example of max pooling with window size of 2×2

Pooling Layer

Pooling is an operation in CNNs that aims to reduce the spatial dimensions of the feature maps while retaining their most important information. This downsampling step helps to decrease computational complexity, control overfitting, and enhance translation invariance. There are two common types of pooling: max pooling and average pooling.

- **Max Pooling:** Max pooling selects the maximum value within a defined window (usually a 2×2 or 3×3 region) as the representative value for that region. This approach has the advantage of preserving the most prominent features within the pooled regions, which can be beneficial for capturing patterns and textures in images. However, one downside is that max pooling can potentially discard

other valuable information from the input by only considering the maximum value.
Also, it may introduce some loss of spatial resolution, which might affect the final
performance of the network in certain applications. An example of max pooling
is shown in Figure 2.3

- **Average Pooling:** Average pooling, on the other hand, computes the average
  value within a defined window as the representative value for that region. This
  approach provides a smoother representation of the input feature maps and is less
  sensitive to extreme values compared to max pooling. It can help in reducing the
  risk of overfitting and maintaining more contextual information within the pooled
  regions. However, average pooling might result in losing some distinctive features
  that would be otherwise preserved using max pooling, as it considers all values in
  the region equally. This could potentially lead to lower performance in tasks that
  require more precise localization or detection of specific features.

As can be seen, CNNs can be composed of various components, including convolu-
tional layers, activation functions, and pooling layers, depending on the specific archi-
tecture or network design being employed.

2.3.2 Loss Function

In neural networks, the loss function plays a critical role in determining the performance
of the model. The loss function quantifies the difference between the predicted outputs
and the actual target values, providing a metric to evaluate the model's accuracy.

The primary objective of image registration is to determine a meaningful transforma-
tion φ, which results in the accurate alignment of images. In this context, the predicted
outputs are generally associated with the transformation, such as deformation $\boldsymbol{u}$, that

depends on the model weights, as given in the following [26]:

$$\boldsymbol{u_W} = f(R, T | \mathbf{W}), \tag{2.11}$$

where f represents the neural network, and $\mathbf{W}$ denotes its weights. The deformation $\boldsymbol{u_W}$ is obtained by applying the neural network f to the input images, R and T, with the model weights $\mathbf{W}$. The actual target values are determined based on the learning paradigm employed. Neural networks primarily follow two distinct learning paradigms: supervised learning and unsupervised learning [4].

Supervised Learning

Supervised learning involves training a model using labeled data, where the target output values are provided for each input. This allows the model to learn a direct mapping from input features to the desired output, making it suitable for tasks such as classification and regression. In the case of supervised learning for image registration, the loss function $\mathcal{L}$ can be represented as [15]:

$$\mathcal{L}(\boldsymbol{u_W}) = d(\boldsymbol{u_W}, \boldsymbol{u}_{\mathrm{gt}}), \tag{2.12}$$

where $\boldsymbol{u}_{\mathrm{gt}}$ denotes the ground truth solution, and d is a function measuring the difference between the predicted outputs and the actual target values. Supervised learning has been employed in image registration studies, such as those found in [27, 39]. The advantage of supervised learning is that it can achieve high accuracy and generalization performance when provided with sufficient training data. However, the main drawback is the requirement for labeled data, which can be time-consuming and costly to obtain, especially in cases where manual labeling is needed.

Unsupervised Learning

On the other hand, unsupervised learning does not rely on labeled data. Instead, it aims to discover underlying patterns and structures in the input data by leveraging the intrinsic properties of the data itself. Common unsupervised learning tasks include clustering, dimensionality reduction, and density estimation. In the case of unsupervised learning for image registration, the loss function $\mathcal{L}$ can be represented as following [3]:

$$\mathcal{L}(\boldsymbol{u_W}) = \mathcal{D}(T_{\boldsymbol{u_W}}, R) + \alpha \mathcal{R}(\boldsymbol{u_W}), \tag{2.13}$$

where $\mathcal{D}(T_{\boldsymbol{u_W}}, R)$ serves as the data similarity term, which quantifies the dissimilarity between the deformed template image $T_{\boldsymbol{u_W}}$ and the reference image R, $\mathcal{R}(\boldsymbol{u_W})$ is the regularization term that discourages non-smooth or unrealistic deformations, and parameter α is a positive scalar that adjusts the trade-off between these two terms. Unsupervised learning has been employed in image registration studies, such as those found in [11, 13, 40]. The primary advantage of unsupervised learning is that it can be applied to large datasets without the need for labeled examples. However, the lack of explicit target values means that evaluating the performance of unsupervised models can be more challenging, and the resulting representations may not always align with human intuition or the desired task-specific outcome.

During the training process, the goal is to minimize the loss function $\mathcal{L}(\boldsymbol{u_W})$ by adjusting the model weights $\mathbf{W}$, i.e.,

$$\min_{\mathbf{W}}\{\mathcal{L}(u_{\mathbf{W}})\}. \tag{2.14}$$

The choice between supervised and unsupervised learning for image registration tasks depends on factors such as the availability of labeled data, the required level of accuracy, and the specific application domain.

CHAPTER 3

PROPOSED DIFFEOMORPHIC DEFORMABLE IMAGE REGISTRATION MODEL

3.1 Introduction

Over the last decade, there has been a growing interest in diffeomorphic deformable image registration among researchers [31, 36]. Diffeomorphic transformations play a crucial role in image registration due to their unique properties, ensuring that the resulting registration is smooth and invertible. These properties are particularly important in applications where preserving the topological structure of the images is vital, such as in medical imaging.

One notable example of the importance of diffeomorphic transformations can be found in medical image registration, specifically in the registration of lung images during different phases of respiration, such as inhalation and exhalation. By employing diffeomorphic deformable image registration, researchers and clinicians can obtain reliable and meaningful results that aid in the diagnosis, treatment planning, and monitoring of various lung conditions [14].

Han and Wang [21] presented an effective diffeomorphic image registration model that leverages fractional-order regularization and the Cauchy-Riemann constraint to further enhance the smoothness and invertibility of the resulting transformations. The Cauchy-Reimann constraint is transformed into an unconstrained minimization problem. Their work provides a thorough mathematical analysis of the proposed model as well as numerical experiments to demonstrate its effectiveness and compares its performance with other registration methods. Their results indicate that the proposed model

performs well in terms of accuracy and robustness, outperforming other methods in various scenarios.

Kuang and Schmah [28] proposed a promising unsupervised 3D medical image registration method, also known as FAIM, which leverages deep learning to learn diffeomorphic transformations while preserving the Jacobian determinant. To ensure diffeomorphic transformations, the method incorporates a regularization term based on the Jacobian determinant into the loss function. Their work demonstrates the effectiveness of the approach through various experiments, showcasing its potential for a wide range of medical imaging applications.

In this thesis, we aim to propose a novel deep learning model for model diffeomorphic deformable image registration that incorporate a new regularization term, which is based on the Jacobian determinant, into the loss function of CNNs. Furthermore, we perform a comparison with the regularization terms presented in the previous works [21, 28] to evaluate the effectiveness of our proposed model.

Following this, we will delve into the proposed CNN architecture, detailing its individual components. We will also discuss the loss function, highlighting how it integrates the aforementioned regularization terms.

3.2 Proposed Neural Network Architecture

In our research, a Fully Convolutional Neural Network (FCNN) is developed with an architecture similar to U-Net [33] to tackle the diffeomorphic deformable image registration problem. The U-Net architecture has demonstrated remarkable effectiveness in a range of tasks related to image segmentation and registration [35], making it ideal for our study. The FCNN, a type of CNNs where the dimensions of the input and output are identical, accepts a pair of images, the template image T and the reference image
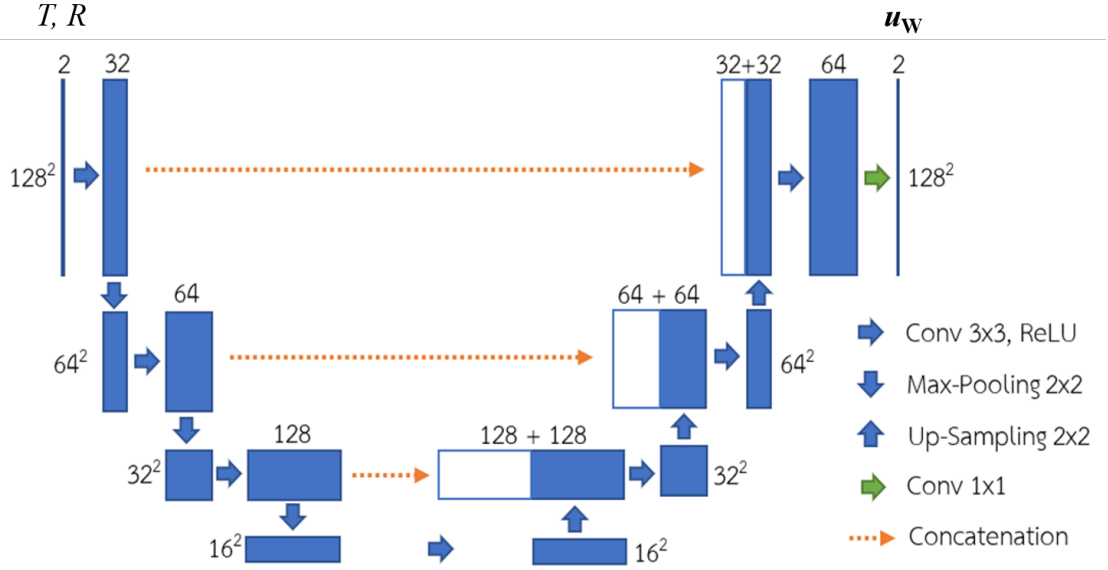
Figure 3.1: FCNN Architecture with a U-Net-like structure for Diffeomorphic Deformable Image Registration.

R, as input and produces the deformation $u_{\mathbf{W}}$ as output, as depicted in Figure 3.1.

3.2.1 Architecture Components

The architecture of the proposed FCNN comprises four levels, with the first level having 32 kernels and each successive layer having twice as many kernels. The network utilizes 2D convolutions with a 3×3 kernel size, and both the encoder and decoder paths are followed by the ReLU activation layer. The decoder path uses 2×2 upsampling layers and half as many kernels as the prior layer, while the encoder path uses 2×2 max pooling and a stride of 2 to downsample the feature maps. The main components of the architecture include:

Encoder path: The encoder path is responsible for extracting features from the input images. It consists of a series of convolutional layers, activation functions (ReLU), and downsampling layers (max pooling).

Decoder path: The decoder path is responsible for reconstructing the deformation from the extracted features. It consists of a series of upsampling layers, convolutional

layers, and activation functions (ReLU).

Skip connections: The proposed FCNN architecture incorporates skip connections or concatenation between the encoder and decoder paths.

Final layer: The final layer of the network, with a 1×1 kernel size convolution kernel, produces the deformation, as represented by $\boldsymbol{u_W}$:

3.2.2 Deformation Prediction

The primary goal of the proposed FCNN is to predict the deformation $\boldsymbol{u_W}$ that the deformed template image $T_{\boldsymbol{u_W}}$ and the reference image R are spatially matched with an optimal or meaningful manner. The final layer of the network produces the deformation $\boldsymbol{u_W}$ by applying a 1x1 convolution kernel. This layer acts as a regression layer, estimating the deformation at each pixel location in the image. The predicted deformation $\boldsymbol{u_W}$ is then used to warp the template image T to align it with the reference image R.

In summary, the proposed FCNN with a U-Net-like architecture efficiently predicts the deformation $\boldsymbol{u_W}$ for diffeomorphic deformable image registration tasks (as can be seen in Chapters 4 and 5). The architecture's components, such as the encoder and decoder paths, convolutional layers, activation functions, and upsampling/downsampling techniques, contribute to the overall performance and accuracy of the image registration process.

3.3 Proposed Loss Function

As mentioned in section 2.3.2, the registration task is to solve the minimization problem:

$$\min_{\mathbf{W}}\{\mathcal{L}\left(\boldsymbol{u_W}\right)\} \tag{3.1}$$

where $\boldsymbol{u_W} = f(R, T|\mathbf{W})$, f is our proposed FCNN, and $\mathbf{W}$ represents its model weights. For the sake of simplicity in the following discussion, we will use $\boldsymbol{u}$ to represent $\boldsymbol{u_W}$.

In this thesis, we design our loss function to train the proposed FCNN for diffeomorphic deformable image registration problems. The proposed loss function consists of three components: data similarity loss, smoothness regularization loss, and diffeomorphic regularization loss. These components work together to minimize the similarity between the reference and deformed template images, enforce smoothness in the deformation, and maintain the properties of diffeomorphic transformations. By using a combination of these loss components, we aim to achieve accurate, smooth, and realistic deformation, enhancing the overall performance of the image registration process. The total loss function can be then expressed as follows:

$$\mathcal{L}(\boldsymbol{u}) = \mathcal{L}_{\text{sim}}(\boldsymbol{u}) + \alpha \mathcal{L}_{\text{smooth}}(\boldsymbol{u}) + \beta \mathcal{L}_{\text{diffeo}}(\boldsymbol{u}). \tag{3.2}$$

where α and β are two parameters that control the trade-off among sub-loss functions. The details of each loss are as follows:

3.3.1 Data Similarity Loss

Assume that the image intensities of the given images R and T are comparable (i.e., in a monomodal registration scenario), the data similarity loss can be given by:

$$\mathcal{L}_{\text{sim}}(\boldsymbol{u}) = \mathcal{D}_{\text{SSD}}(T_{\boldsymbol{u}}, R) = \int_{\Omega} (T_{\boldsymbol{u}}(\mathbf{x}) - R(\mathbf{x}))^2 d\mathbf{x}. \tag{3.3}$$

Lower values of $\mathcal{L}_{\text{sim}}(\boldsymbol{u})$ indicate a better registration result.

3.3.2 Smoothness Regularization Loss

In this work, the diffusion regularization of the form

$$\mathcal{L}_{\text{smooth}}(\boldsymbol{u}) = \mathcal{R}_{\text{diff}}(\boldsymbol{u}) = \int_{\Omega} \sum_{l=1}^{2} |\nabla u_l(\mathbf{x})|^2 \, d\mathbf{x}. \tag{3.4}$$

is used to be smoothness regularization loss in order to ensure that the constructed deformation $\boldsymbol{u}$ is smooth and penalizes unwanted deformation.

3.3.3 Diffeomorphic Regularization Loss

We first review the diffeomorphic regularization techniques: the Cauchy-Riemann-based regularization introduced by Han and Wang [21] and the regularization used in Kuang and Schmah [28] and propose our new diffeomorphic regularization loss. Note that these losses enforce the diffeomorphism constraint, which ensures the constructed deformation is smooth, invertible, and non-self-intersecting.

1. **Cauchy-Riemann-Based Regularization:**

   Han and Wang [21] proposed a diffeomorphic image registration model that incorporates the Cauchy-Riemann constraint. This registration model aims at ensuring that $\det(\nabla\boldsymbol{\varphi}(\mathbf{x})) > 0$ to guarantee $\boldsymbol{\varphi}$ to be diffeomorphic transformation. The regularization in their registration model is given by:

   $$\mathcal{L}_{\text{diffeo,CR}}(\boldsymbol{u}) = \int_{\Omega} \left(\frac{\partial u_1(\mathbf{x})}{\partial x_1} - \frac{\partial u_2(\mathbf{x})}{\partial x_2}\right)^2 + \left(\frac{\partial u_1(\mathbf{x})}{\partial x_2} + \frac{\partial u_2(\mathbf{x})}{\partial x_1}\right)^2 d\mathbf{x}. \quad (3.5)$$

2. **Diffeomorphic Regularization:**

   Kuang and Schmah [28] integrated a Jacobian determinant regularization term into the loss function of their model to ensure diffeomorphic transformations and prevent folding or tearing in the registered images. The regularization in their registration model is defined as follows:

   $$\mathcal{L}_{\text{diffeo},\sigma_1}(\boldsymbol{u}) = \int_{\Omega} \sigma_1(\det(\nabla\boldsymbol{\varphi}(\mathbf{x})))d\mathbf{x}, \quad (3.6)$$

where $\sigma_1(x) = \frac{1}{2}(|x| - x) = \max(0, -x), x \in \mathbb{R}$ which is the ReLU function of $-x$.

Now, we shall propose our new diffeomorphic regularization loss function,

$$\mathcal{L}_{\mathrm{diffeo},\sigma_p}(\boldsymbol{u}) = \int_\Omega \sigma_p \left(\det(\nabla\boldsymbol{\varphi}(\mathbf{x})) \right) d\mathbf{x}, \tag{3.7}$$

which improves the one introduced in [28] where $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{x} + \boldsymbol{u}(\mathbf{x})$ and $\sigma_p(x) = \frac{(x-1)^2}{\max(x,1)^2}$ satisfies the following properties:

- σ_p is a smooth function defined on real numbers. Because the Jacobian determinant is a real number, $\det(\nabla\boldsymbol{\varphi}(\mathbf{x})) \in \mathbb{R}$. Here $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{x} + \boldsymbol{u}(\mathbf{x})$.

- σ_p is a decreasing function for $x < 1$ and an increasing function for $x \geq 1$. This property encourages the Jacobian determinant to be close to 1, which means the local volume change induced by the deformation should be close to the identity transformation. By minimizing the proposed loss function, the algorithm avoids excessive local shrinkage or expansion.

- $\sigma_p(x) = \sigma_p\left(\frac{1}{x}\right)$ if $x > 0$. This symmetry property ensures that local shrinkage and local expansion are penalized symmetrically. This symmetry is crucial for maintaining the balance between local compression and local expansion.

To visualize the behavior of the two diffeomorphic regularization loss functions, $\mathcal{L}_{\mathrm{diffeo},\sigma_1}$ and $\mathcal{L}_{\mathrm{diffeo},\sigma_p}$, we plot σ_1 and σ_1 on a graph as shown in Figure 3.2. The plots can show the different behaviors of the loss functions, illustrating how they penalize local shrinkage and expansion. σ_1 is a constant rate decreasing function for $x < 0$ and a constant function for $x \geq 1$ while σ_p is a decreasing function for $x < 1$ and an increasing function for $x \geq 1$.

3.4 Discretization

In this section, we shall introduce the finite difference discretization for image registration, focusing on the domain setup and discretized losses.
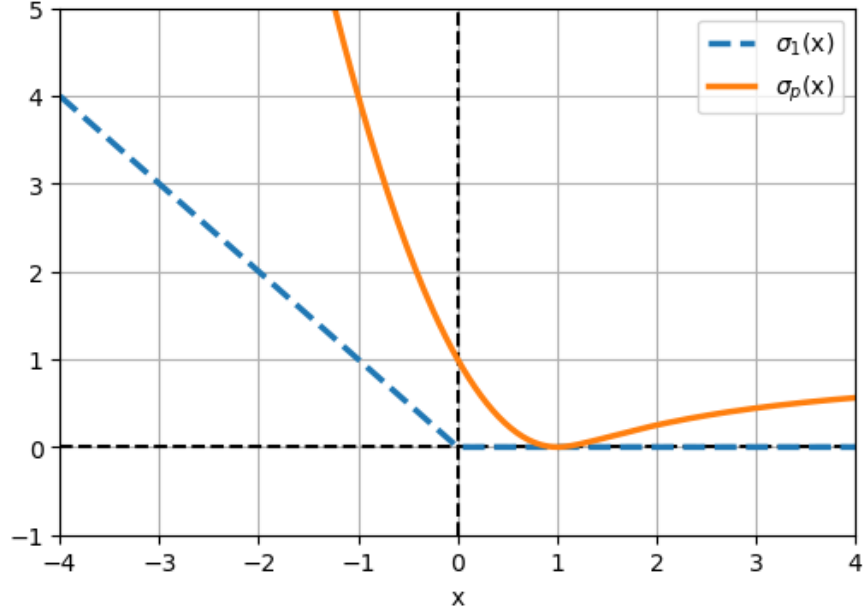
Figure 3.2: Behaviors of σ_1 and σ_p.

Assume that the image domain is represented by a two-dimensional set $\Omega = [0, N - 1]^2 \subset \mathbb{R}$, where the image shape is $N \times N$. To discretize the domain, we define a uniform grid with grid points $(x_i, y_j) = (i\Delta, j\Delta)$, where $i, j \in 0, 1, 2, ..., N - 1$. Here, the grid spacing $\Delta = 1$ is determined by dividing the size of the domain by the number of grid points in each dimension.

The continuous images $T(\mathbf{x})$ and $R(\mathbf{x})$ is then represented by their discretized counterparts $T_{ij} = T(x_i, y_j)$ and $R_{ij} = R(x_i, y_j)$. Similarly, the deformation field $\boldsymbol{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))^\top$ can be discretized as $\boldsymbol{u}_{ij} = \boldsymbol{u}(x_i, y_j) = (u_1(x_i, y_j), u_2(x_i, y_j))^\top$.

3.4.1 Discretization of $\mathcal{L}_{\mathrm{sim}}(\boldsymbol{u})$

To discretize the similarity loss function $\mathcal{L}_{\mathrm{sim}}(\boldsymbol{u})$, we compute the discrete version of the integral over the domain Ω as follows:

$$\mathcal{L}_{\mathrm{sim}}(\boldsymbol{u}) \approx \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (T_{\boldsymbol{u}_{ij}} - R_{ij})^2 \Delta^2. \tag{3.8}$$

We note that the so-called bilinear interpolation is used as a method to compute the

intensity value of a continuous image at non-integer coordinates by interpolating the intensity values of the surrounding grid points.

Given the deformation $\boldsymbol{u}_{ij} = (u_{1_{ij}}, u_{2_{ij}})$ at a grid point (x_i, y_j), we can compute the transformed coordinates $\mathbf{x}' = (x', y')$ as follows:

$$(x', y') = (x_i + u_{1_{ij}}, y_j + u_{2_{ij}}). \tag{3.9}$$

Since (x', y') may not correspond to integer coordinates, we use bilinear interpolation to estimate the intensity value $T_{\boldsymbol{u}_{ij}}$ at this point. Let $i' = \lfloor x' \rfloor$, $j' = \lfloor y' \rfloor$, $a = x' - i'$, and $b = y' - j'$. Then, we can compute the bilinearly interpolated intensity value as follows:

$$T_{\boldsymbol{u}_{ij}} = (1-a)(1-b)T_{i'j'} + a(1-b)T_{(i'+1)j'} + (1-a)bT_{i'(j'+1)} + abT_{(i'+1)(j'+1)}. \tag{3.10}$$

In this equation, $T_{i'j'}, T_{(i'+1)j'}, T_{i'(j'+1)}$, and $T_{(i'+1)(j'+1)}$ are the intensity values of the template image at the surrounding grid points of the transformed coordinate (x', y'). By using bilinear interpolation, we can compute the deformed template image $T_{\boldsymbol{u}_{ij}}$ for any deformation field $\boldsymbol{u}$, allowing us to evaluate the similarity loss function and perform the image registration process.

3.4.2 Discretization of $\mathcal{L}_{\mathrm{smooth}}(\boldsymbol{u})$

We first rewrite the smoothness loss function (3.4) as follows:

$$\mathcal{L}_{\mathrm{smooth}}(\boldsymbol{u}) = \int_\Omega \sum_{l=1}^2 |\nabla u_l(\mathbf{x})|^2 \, d\mathbf{x} = \int_\Omega \left(|\nabla u_1(\mathbf{x})|^2 + |\nabla u_2(\mathbf{x})|^2 \right) d\mathbf{x}. \tag{3.11}$$

Now, we approximate the integrals using the finite sums over the discrete domain and the gradients using the standard finite difference as given by:

$$\mathcal{L}_{\mathrm{smooth}}(\boldsymbol{u}) \approx \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(|\nabla_{ij} u_1|^2 + |\nabla_{ij} u_2|^2 \right) \Delta^2, \tag{3.12}$$

where $\nabla_{ij} u_l$ denotes the gradient of u_l at the grid point (x_i, y_j) for all $l = 1, 2$ and

$|(x, y)| = \sqrt{x^2 + y^2}$. In order to approximate the gradients, we apply the central finite difference as follows:

$$\nabla_{ij} u_l = \left(\frac{u_{l_{(i+1)j}} - u_{l_{(i-1)j}}}{2\Delta}, \frac{u_{l_{i(j+1)}} - u_{l_{i(j-1)}}}{2\Delta} \right), \tag{3.13}$$

with the following boundary conditions:

$$u_{l_{(-1)j}} = u_{l_{(0)j}}, \ u_{l_{(N)j}} = u_{l_{(N-1)j}}, \ u_{l_{i(-1)}} = u_{l_{i(0)}}, \ u_{l_{i(N)}} = u_{l_{i(N-1)}}. \tag{3.14}$$

3.4.3 Discretization of $\mathcal{L}_{\mathbf{diffeo,CR}}(\boldsymbol{u})$

To discretize the Cauchy-Riemann loss term $\mathcal{L}_{\text{diffeo,CR}}(\boldsymbol{u})$ (to be used in our numerical experiment), we first apply the central finite difference approximations to compute the partial derivatives of the deformation on a grid point as follows:

$$\frac{\partial u_{1_{ij}}}{\partial x_1} \approx \frac{u_{1_{(i+1)j}} - u_{1_{(i-1)j}}}{2\Delta},$$
$$\frac{\partial u_{1_{ij}}}{\partial x_2} \approx \frac{u_{1_{i(j+1)}} - u_{1_{i(j-1)}}}{2\Delta},$$
$$\frac{\partial u_{2_{ij}}}{\partial x_1} \approx \frac{u_{2_{(i+1)j}} - u_{2_{(i-1)j}}}{2\Delta},$$
$$\frac{\partial u_{2_{ij}}}{\partial x_2} \approx \frac{u_{2_{i(j+1)}} - u_{2_{i(j-1)}}}{2\Delta}.$$

with the same boundary conditions in section 3.4.2. By substituting the approximated derivatives into the loss function, we obtain the discretized version of the Cauchy-Riemann loss term as given by:

$$\begin{aligned}
\mathcal{L}_{\text{diffeo,CR}}(\boldsymbol{u}) &\approx \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(\left(\frac{\partial u_{1_{ij}}}{\partial x_1} - \frac{\partial u_{2_{ij}}}{\partial x_2} \right)^2 + \left(\frac{\partial u_{1_{ij}}}{\partial x_2} + \frac{\partial u_{2_{ij}}}{\partial x_1} \right)^2 \right) \Delta^2 \\
&\approx \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(\left(\frac{u_{1_{(i+1)j}} - u_{1_{(i-1)j}}}{2\Delta} - \frac{u_{2_{i(j+1)}} - u_{2_{i(j-1)}}}{2\Delta} \right)^2 + \right. \\
&\qquad \left. \left(\frac{u_{1_{i(j+1)}} - u_{1_{i(j-1)}}}{2\Delta} + \frac{u_{2_{(i+1)j}} - u_{2_{(i-1)j}}}{2\Delta} \right)^2 \right) \Delta^2.
\end{aligned}$$

3.4.4 **Discretization of $\mathcal{L}_{\mathbf{diffeo},\sigma_1}(\boldsymbol{u})$ and $\mathcal{L}_{\mathbf{diffeo},\sigma_p}(\boldsymbol{u})$**

To discretize the diffeomorphic regularization loss terms $\mathcal{L}_{\mathrm{diffeo},\sigma_1}(\boldsymbol{u})$ and $\mathcal{L}_{\mathrm{diffeo},\sigma_p}(\boldsymbol{u})$, we need to compute the Jacobian determinant $\det(\nabla\boldsymbol{\varphi}(\mathbf{x}))$, as shown in (2.1), for each grid point in the discrete domain. The Jacobian determinant at grid point (x_i, y_j) can be computed using the partial derivatives of the deformation as given by:

$$\det(\nabla\boldsymbol{\varphi}(\mathbf{x})) = \left(\frac{\partial u_{1_{ij}}}{\partial x_1} + 1\right)\left(\frac{\partial u_{2_{ij}}}{\partial x_2} + 1\right) - \frac{\partial u_{1_{ij}}}{\partial x_2}\frac{\partial u_{2_{ij}}}{\partial x_1}. \tag{3.15}$$

Next, we apply the central finite differences approximation for $\frac{\partial u_{1_{ij}}}{\partial x_1}, \frac{\partial u_{2_{ij}}}{\partial x_2}, \frac{\partial u_{1_{ij}}}{\partial x_2}$, and $\frac{\partial u_{2_{ij}}}{\partial x_1}$ as given in section 3.4.3 to approximate these partial derivatives.

By substituting the above approximations into 3.15, we consider σ as σ_1 or σ_p, the discrete version of diffeomorphic regularization losses are given by:

$$\mathcal{L}_{\mathrm{diffeo},\sigma}(\boldsymbol{u}) \approx \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sigma\left(\left(\frac{\partial u_{1_{ij}}}{\partial x_1}+1\right)\left(\frac{\partial u_{2_{ij}}}{\partial x_2}+1\right) - \frac{\partial u_{1_{ij}}}{\partial x_2}\frac{\partial u_{2_{ij}}}{\partial x_1}\right)\Delta^2$$

$$\approx \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\sigma\left(\left(\frac{u_{1_{(i+1)j}}-u_{1_{(i-1)j}}}{2\Delta}+1\right)\left(\frac{u_{2_{i(j+1)}}-u_{2_{i(j-1)}}}{2\Delta}+1\right)\right.$$

$$\left.- \frac{u_{1_{i(j+1)}}-u_{1_{i(j-1)}}}{2\Delta}\frac{u_{2_{(i+1)j}}-u_{2_{(i-1)j}}}{2\Delta}\right)\Delta^2,$$

CHAPTER 4

EXPERIMENTS AND RESULTS OF SYNTHETIC IMAGES

In this chapter, we present the experiments and results of synthetic images conducted to evaluate the performance of our proposed deep learning (DL) model for the diffeomorphic deformable image registration problems using different loss functions, along with their diffeomorphic regularization terms. We also compare our proposed method with existing diffeomorphic regularization losses [28] and a DL model that incorporates the Cauchy-Riemann constraint from the variational method proposed by Han and Wang [21]. Finally, we examine the effectiveness of diffeomorphic regularization by comparing the registration results obtained using the proposed loss function with and without the diffeomorphic regularization term.

4.1 Synthesizing Image Pairs

This section introduces the process of building image pairs. All steps can be explained as follows:

4.1.1 Generating Deformation

The first step in building image pairs is to generate deformations. These deformations are responsible for determining the spatial transformation between the image pairs. The process of generating deformations can be summarized as follows:

1. A 5×5 uniformly spaced grid of control points is created. This grid spans across the entire image, with points evenly distributed across the width and height of the image.

2. For each control point in the grid, a corresponding target point is randomly gen-

erated. The distance of the target point from its corresponding control point is restricted to less than 20% of the image size.

3. A two-dimensional cubic interpolation is performed using the control points and their corresponding target points. This creates a smooth deformation. The interpolation ensures that the deformation is smooth and continuous, resulting in realistic and natural deformations.

4. After generating the deformation, a check is performed to ensure that the Jacobian determinant of the transformation is positive. This step is crucial to ensure that the deformation does not cause any folding in the image, preserving the topology of the image. If the Jacobian determinant is positive, the deformation is considered valid.

4.1.2 Creating Template images

Each template image is obtained by the following step:

1. An empty or blank image is created with a pre-specified shape. This image serves as the canvas on which different shapes will be drawn. These shapes are assumed to be the 'objects' in the image and are the features that will be subjected to deformations.

2. Three different geometric shapes are selected randomly from the set comprising rectangles, circles, and ellipses. For each chosen shape, its size and position within the image are determined randomly. Each shape is assigned a unique intensity value to ensure they are distinguishable from each other within the image.

3. Once the shapes are placed within the image, a random transformation is applied to the entire image. The transformation includes a combination of rotation and
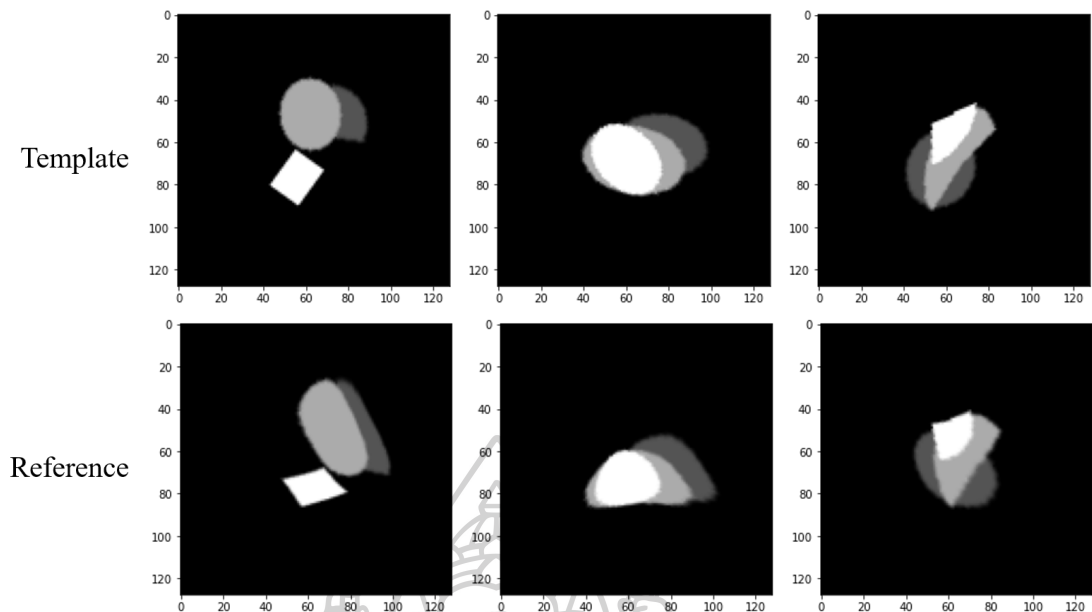
Figure 4.1: Examples of the synthetic image: Top row is template images and the bottom row is reference images.

shearing. The rotation angle and the shear magnitude are both selected randomly. This transformation adds an additional layer of variety among the generated template images, as each image, despite containing similar shapes, appears different due to the transformation applied.

4. After applying a transformation, the intensity of the image is normalized, ensuring that the maximum pixel intensity value in the image is 1. This is a common practice in image processing and deep learning.

5. The generated and transformed image, now termed a 'template image', is added to a collection of such images, building a dataset of template images.

4.1.3 Generating Reference Images

After creating the template images, we generate the corresponding reference images. These images are obtained by transforming the template images using the known deformations generated earlier.

In our approach to generating the dataset of image pairs, we leveraged the capabilities of Python programming to follow the outlined process, resulting in the successful generation of 600 image pairs (template images and corresponding reference images). Representative examples of these synthetic image pairs are depicted in Figure 4.1.

4.2 Dataset Splitting

Our dataset contains 600 image pairs and is split into three parts to ensure proper evaluation and validation of the proposed model as follows:

1. Training: 400 image pairs.

2. Validating: 100 image pairs.

3. Testing: 100 image pairs.

We note that this partitioning strategy allows our registration model to learn from a large set of training data, evaluate its performance on the validation set during training, and assess its generalization capabilities on the unseen test set.

4.3 Evaluation Metrics

Evaluation metrics are crucial for assessing the performance of a learning registration model and play an essential role in the development and evaluation of image registration algorithms. These metrics provide quantitative measures of the algorithm's effectiveness in aligning the synthetic images and are commonly employed to compare various registration methods and assess the accuracy of the results. The following are the evaluation metrics in this thesis.

- **Relative Sum of Squared Differences (Rel. SSD)** is a widely used evaluation metric in registration results. It is computed as the ratio of the sum of squared

differences between the reference image and the transformed template image after registration to the sum of squared differences between the reference image and the template image before registration as given by:

$$\text{Rel. SSD} = \frac{\mathcal{L}_{\text{sim}}(\boldsymbol{u})}{\mathcal{L}_{\text{sim}}(\boldsymbol{0})} = \frac{\int_{\Omega}(T_{\boldsymbol{u}}(\mathbf{x}) - R(\mathbf{x}))^2 d\mathbf{x}}{\int_{\Omega}(T(\mathbf{x}) - R(\mathbf{x}))^2 d\mathbf{x}}. \tag{4.1}$$

The resulting value indicates the degree of changes in the template image after registration, with lower values signifying a better registration outcome. The Rel. SSD is often expressed as a percentage, with values near 0% denoting a highly accurate registration and values closer to 100% indicating a visually unappealing registration result.

- **Number of non-positive Jacobian determinant pixels** addresses cases where the transformation between the given images leads to a fold in the physical space. This occurs when the Jacobian determinant of the transformation function is not positive, implying that the transformation is not invertible at that point or folding. The number of pixels with non-positive values in the determinant of the Jacobian matrix

$$J_{\leq 0}(\boldsymbol{\varphi}) = |\{\mathbf{x} \in \Omega : \det(\nabla\boldsymbol{\varphi}(\mathbf{x})) \leq 0\}| \tag{4.2}$$

is frequently used as an evaluation metric to gauge the quality of a registration result. A larger number of pixels with non-positive values suggests more significant distortion or folding of the transformed synthetic image, resulting in a lower-quality registration outcome.

4.4   Experimental Setup for Synthetic Images

In this section, we detail the experimental setup and parameters used for our proposed DL model for diffeomorphic deformable image registration problems on synthetic images.

4.4.1 Implementation

Our proposed FCNN was designed and implemented using Keras, a powerful open-source neural network library, with TensorFlow serving as the backend. For the synthetic images, we employed a method of 2D linear interpolation to generate the deformed template images. The central finite difference method was used to numerically discretize the partial derivatives in the loss function as mentioned in section 3.4.

The model was trained using the Adaptive Moment Estimation (ADAM) optimizer, a popular and effective gradient descent optimization algorithm. The learning rate was set to 10^{-4}. The training process used mini-batch stochastic gradient descent with a batch size of 20 image pairs. The training was run for a total of 200 epochs.

All experiments were conducted in Python on a computer notebook equipped with an AMD Ryzen 7 4800HS processor and Radeon graphics, operating at a clock speed of 2.90 GHz, and equipped with 16 GB of RAM. All operations were carried out using the default settings.

4.4.2 Loss Function Parameters

We experimented on the total loss function (3.2) with and without diffeomorphic regularization loss to see how well our proposed DL model performs on synthetic images. For the total loss function without diffeomorphic regularization loss ($\beta = 0$), we used $\alpha = 0.001, 0.002, 0.005, 0.01, 0.02, 0.05$, and 0.1. For the total loss function with diffeomorphic regularization loss ($\beta \neq 0$), we used $\alpha = 0.001$, $\beta = 0.01, 0.02, 0.05, 0.1$, and 0.2.

Table 4.1: Average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ from the validation set of the 100 pairs of the synthetic images without diffeomorphic regularization loss in (3.2) (i.e., $\beta = 0$).

α	Rel. SSD (%)	$J_{\leq 0}(\boldsymbol{\varphi})$
0.001	5.25	127
0.002	5.36	72
0.005	5.62	24
0.01	5.52	14
0.02	6.44	5
0.05	8.62	2
0.1	**11.94**	**0**

Table 4.2: Average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ from the validation set 100 pairs the synthetic images with different diffeomorphic regularization losses in (3.2) ($\alpha = 0.001$).

β	$\mathcal{L}_{\text{diffeo,CR}}$ (3.5)		$\mathcal{L}_{\text{diffeo},\sigma_1}$ (3.6)		$\mathcal{L}_{\text{diffeo},\sigma_p}$ (3.7)	
	Rel. SSD	$J_{\leq 0}(\boldsymbol{\varphi})$	Rel. SSD	$J_{\leq 0}(\boldsymbol{\varphi})$	Rel. SSD	$J_{\leq 0}(\boldsymbol{\varphi})$
0.01	5.95%	16	5.01%	92	5.14%	28
0.02	6.45%	5	4.79%	78	5.49%	17
0.05	9.02%	1	5.23%	47	6.45%	4
0.1	**11.48%**	**0**	4.95%	38	7.36%	2
0.2	16.68%	0	5.52%	25	**9.52%**	**0**
100	-	-	7.56%	2	-	-
200	-	-	9.55%	1	-	-
500	-	-	9.67%	1	-	-
1000	-	-	**10.84%**	**0**	-	-

4.5 Results for Synthetic Images

4.5.1 Validation Set Performance

Table 4.1 shows the average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ resulting from the validation set of the 100 pairs of the synthetic images using the proposed total loss functions (3.2) without diffeomorphic regularization loss (i.e., $\beta = 0$). From the registration results, increasing α can reduce $J_{\leq 0}(\boldsymbol{\varphi})$ to 0, achieving diffeomorphic deformation. However, we expect that using the total proposed loss functions (3.2) with diffeomorphic regularization loss can reduce the values of $J_{\leq 0}(\boldsymbol{\varphi})$ more effectively.

Table 4.2 presents the average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ obtained from the validation set of the 100 pairs of the synthetic images using the proposed total loss functions

(3.2) with different diffeomorphic regularization losses. Additionally, we further experimented for the $\mathcal{L}_{\text{diffeo},\sigma_1}$ diffeomorphic regularization with $\beta = 100, 200, 500$ and 1000, since $J_{\leq 0}(\boldsymbol{\varphi})$ had not yet reached 0. The results demonstrate that our proposed DL model using diffeomorphic regularization in (3.7) is more effective for reducing $J_{\leq 0}(\boldsymbol{\varphi})$ in terms of the Rel. SSD when compared to other diffeomorphic regularization losses.

4.5.2 Test Set Performance

We chose the appropriate α and β that yield $J_{\leq 0}(\boldsymbol{\varphi}) = 0$ for each experiment. In our notation, we represent the total loss function as $\mathcal{L}_\lambda(\boldsymbol{u}|\alpha, \beta)$ for convenience. This function is defined as:

$$\mathcal{L}_\lambda(\boldsymbol{u}|\alpha, \beta) = \mathcal{L}_{\text{sim}}(\boldsymbol{u}) + \alpha\mathcal{L}_{\text{smooth}}(\boldsymbol{u}) + \beta\mathcal{L}_{\text{diffeo},\lambda}(\boldsymbol{u}), \qquad (4.3)$$

Recall that $\mathcal{L}_{\text{sim}}(\boldsymbol{u})$ represents the similarity loss term, $\mathcal{L}_{\text{smooth}}(\boldsymbol{u})$ is the smoothness regularization term, and $\mathcal{L}_{\text{diffeo},\lambda}(\boldsymbol{u})$ denotes the diffeomorphic regularization term. The parameters α and β are weights for the smoothness and diffeomorphic regularization terms, respectively. The term λ in the subscript of $\mathcal{L}$ is used to represent the types of diffeomorphic regularization being used.

We compare $\mathcal{L}_-(\boldsymbol{u}|0.1, 0)$, $\mathcal{L}_{\text{CR}}(\boldsymbol{u}|0.001, 0.1)$, and $\mathcal{L}_{\sigma_1}(\boldsymbol{u}|0.001, 1000)$ with the proposed total loss function $\mathcal{L}_{\sigma_p}(\boldsymbol{u}|0.001, 0.2)$. Figure 4.2 shows the loss total function behavior throughout the training process on the synthetic images. Throughout the training process, both the training and validation loss decrease. This observation suggests that our DL model generalizes well to new data and is not overfitting. However, an exception was noted in the case of the model trained with the total loss function $\mathcal{L}_{\sigma_1}(\boldsymbol{u}|0.001, 1000)$, which displayed instability during the training process.

We also used the α and β that yield $J_{\leq 0}(\boldsymbol{\varphi}) = 0$ to compare the average Rel. SSD and $J_{\leq 0}(\boldsymbol{\varphi})$ resulting from the test set, as shown in Table 4.3. These results demonstrate that
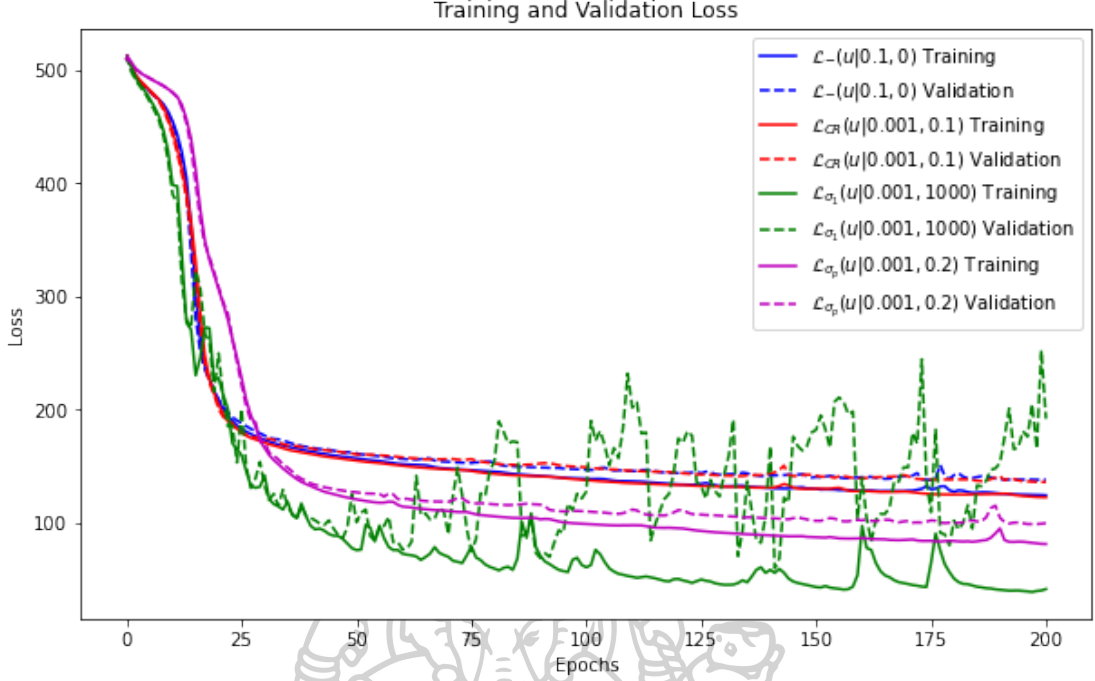
Figure 4.2: The plots of different total loss functions from training and validation sets during the training process on the pairs of the synthetic images for different loss functions

Table 4.3: Comparison of average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ from the test set of the 100 image pairs of the synthetic images with different total loss functions.

Total loss function	Rel. SSD (%)	$J_{\leq 0}(\boldsymbol{\varphi})$
$\mathcal{L}_-(\boldsymbol{u}\|0.1, 0)$	11.69	0
$\mathcal{L}_{\mathrm{CR}}(\boldsymbol{u}\|0.001, 0.1)$	11.13	0
$\mathcal{L}_{\sigma_1}(\boldsymbol{u}\|0.001, 1000)$	10.99	0
$\mathcal{L}_{\sigma_p}(\boldsymbol{u}\|0.001, 0.2)$	**9.74**	**0**

the effectiveness of our proposed diffeomorphic regularization loss in reducing $J_{\leq 0}(\boldsymbol{\varphi})$ more efficiently than other approaches for synthetic images.

Figure 4.3 displays a registration problem from our test set. The first row shows the template T and reference R. The second and third rows show the deformed template images $T_{\boldsymbol{u}}$ and the constructed transformations, from different total loss functions.

As can be seen, the registration results obtained from the synthetic image pairs confirm the effectiveness of our proposed DL model in reducing $J_{\leq 0}(\boldsymbol{\varphi})$ while maintaining a competitive Rel. SSD when compared to other diffeomorphic regularization losses.

(a) T (b) R

(c) $T_{\boldsymbol{u}}^{\mathcal{L}_-}$
Rel.SSD = 5.61%

(d) $T_{\boldsymbol{u}}^{\mathcal{L}_{\mathrm{CR}}}$
Rel.SSD = 4.88%

(e) $T_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_1}}$
Rel.SSD = 4.30%

(f) $T_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_p}}$
Rel.SSD = 2.59%

(g) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_-}) = 0$

(h) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_{\mathrm{CR}}}) = 0$

(i) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_1}}) = 0$

(j) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_p}}) = 0$
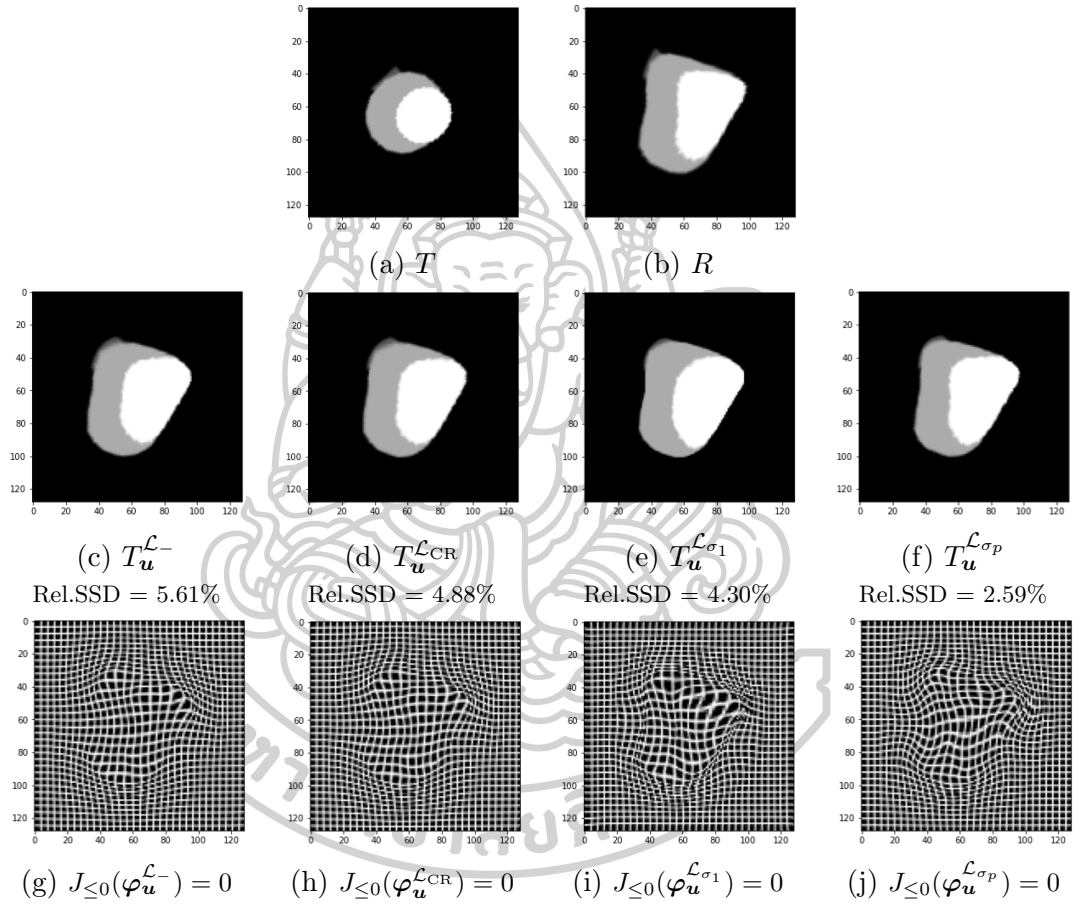
Figure 4.3: Registration results of a registration problem from the test set: (a) template T, (b) reference R, (c)-(f) and (g)-(j) are respectively the deformed template images and constructed transformation from $\mathcal{L}_-(\boldsymbol{u}|0.1, 0)$, $\mathcal{L}_{\mathrm{CR}}(\boldsymbol{u}|0.001, 0.1)$, $\mathcal{L}_{\sigma_1}(\boldsymbol{u}|0.001, 1000)$, and $\mathcal{L}_{\sigma_p}(\boldsymbol{u}|0.001, 0.2)$ (our proposed total loss function).

CHAPTER 5

EXPERIMENTS AND RESULTS OF MEDICAL IMAGES

The previous chapter presents the performance of our proposed DL model on synthetic images. This chapter extends our evaluation to real medical images and demonstrates the robustness and applicability of our DL model with different diffeomorphic regularization losses.

5.1 Dataset

The standard digital image database for tuberculosis[1] dataset includes Chest X-rays obtained during routine clinical procedures using Philips DR Digital Diagnose systems. It comprises 336 cases with tuberculosis manifestations and 326 normal cases. The images in the dataset are in PNG format and vary in size, with each X-ray approximately 3K×3K pixels. For our study, the tuberculosis X-ray dataset was chosen and utilized as the evaluation dataset.

5.2 Pre-processing

The 220 normal male images of 3K×3K pixels from the dataset were undergone pre-processing by resizing them to 128x128 pixels and scaling the intensity values from 0 to 1. Of these 220 images, there are 160 training images, 30 validating images, and 30 testing images. Our image pairing procedure was applied to create image pairs, with one image designated as the template and the other as the reference image. This pairing process resulted in 25,440 image pairs for training and 870 image pairs for validation

---

[1]A publicly accessible dataset, was established through a collaboration between the National Library of Medicine in Maryland, USA, and Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China [6, 24].
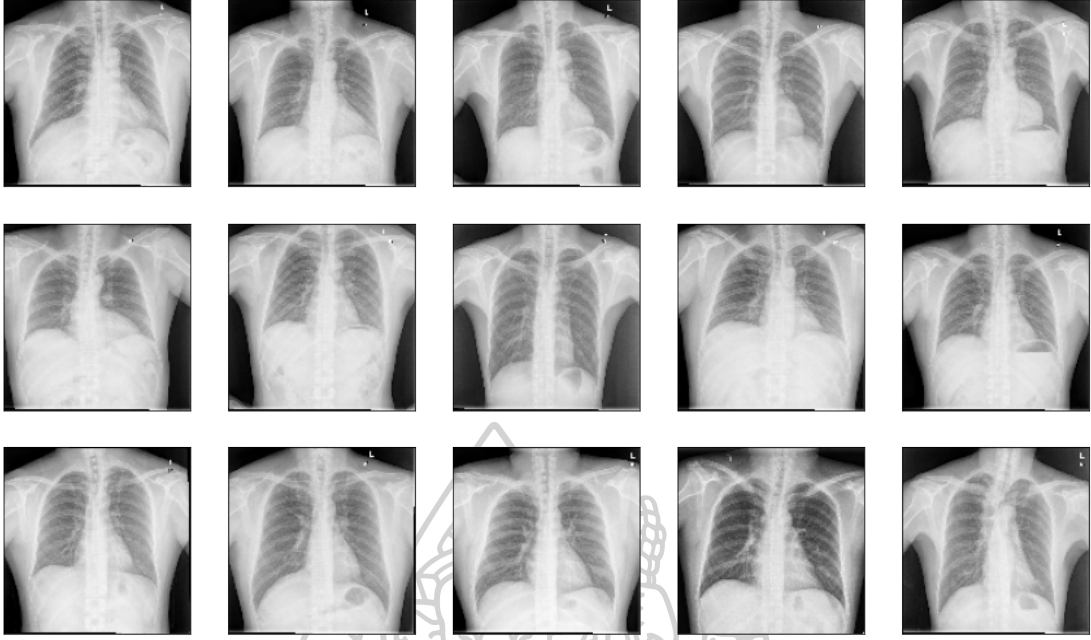
Figure 5.1: Examples of real medical images used in our experiments, showcasing the diversity of images.

and testing.

Figure 5.1 displays some examples of real medical images that were utilized in our experiments. The figure demonstrates the variety of real medical images used.

5.3 Evaluation Metrics

The following two evaluation metrics will be also used in this chapter:

- **Relative Sum of Squared Differences (Rel. SSD):**

$$\text{Rel. SSD} = \frac{\mathcal{L}_{\text{sim}}(\boldsymbol{u})}{\mathcal{L}_{\text{sim}}(\mathbf{0})} = \frac{\int_{\Omega}(T_{\boldsymbol{u}}(\mathbf{x}) - R(\mathbf{x}))^2 d\mathbf{x}}{\int_{\Omega}(T(\mathbf{x}) - R(\mathbf{x}))^2 d\mathbf{x}}.$$

- **Number of non-positive Jacobian determinant pixels:**

$$J_{\leq 0}(\boldsymbol{\varphi}) = |\{\mathbf{x} \in \Omega : \det(\nabla \boldsymbol{\varphi}(\mathbf{x})) \leq 0\}|.$$

5.4 Experimental Setup for Real Medical Images

We outline the experimental setup and parameters employed for our proposed DL model for diffeomorphic deformable image registration on real medical images, which are a bit different from those used for synthetic images.

5.4.1 Implementation

For real medical images, our proposed FCNN was also implemented using Keras with a TensorFlow backend. As with synthetic images, 2D linear interpolation was used to generate the deformed template images, and the central finite difference method was utilized to numerically discretize the partial derivatives in the loss function. The DL model was trained using the ADAM optimizer, maintaining a learning rate of 10^{-4}. Each training batch consisted of pairs of the real medical images, facilitating mini-batch stochastic gradient descent with a batch size of 30 and a total of 12 epochs. All experiments were executed in Python on the same computer notebook as mentioned in the previous chapter in section 4.4.1.

5.4.2 Loss Function Parameters

We experimented on the total loss function (3.2) with and without diffeomorphic regularization losses, to evaluate the performance of our proposed DL model on the real medical images. For the total loss function without diffeomorphic regularization loss ($\beta = 0$), we used $\alpha = 0.01, 0.02, 0.05$, and 0.1. For the total loss function with diffeomorphic regularization loss ($\beta \neq 0$), we used $\alpha = 0.01$, $\beta = 0.01, 0.02, 0.05$.

Table 5.1: Average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ from the validation set of the 870 image pairs of the real medical images without the diffeomorphic regularization loss in (3.2) (i.e., $\beta = 0$).

α	Rel. SSD (%)	$J_{\leq 0}(\boldsymbol{\varphi})$
0.01	20.44	136
0.02	23.34	28
0.05	27.19	3
0.1	**31.87**	**0**

Table 5.2: Average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ from the validation set of the 870 image pairs of the real medical images with different diffeomorphic regularization losses in (3.2) ($\alpha = 0.01$).

β	$\mathcal{L}_{\text{diffeo,CR}}$ (3.5)		$\mathcal{L}_{\text{diffeo},\sigma_1}$ (3.6)		$\mathcal{L}_{\text{diffeo},\sigma_p}$ (3.7)	
	Rel. SSD	$J_{\leq 0}(\boldsymbol{\varphi})$	Rel. SSD	$J_{\leq 0}(\boldsymbol{\varphi})$	Rel. SSD	$J_{\leq 0}(\boldsymbol{\varphi})$
0.01	22.83%	31	20.16%	129	22.14%	39
0.02	24.13%	13	20.23%	105	22.48%	15
0.05	27.57%	2	19.81%	94	24.07%	3
0.1	**31.77%**	**0**	20.48%	53	**26.65%**	**0**
100	-	-	23.94%	1	-	-
200	-	-	23.93%	1	-	-
500	-	-	**26.42%**	**0**	-	-

5.5 Results for Real Medical Images

5.5.1 Validation Set Performance

Table 5.1 presents the average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ obtained from the validation set of the 870 image pairs of the real medical images using the proposed total loss function (3.2) without diffeomorphic regularization ($\beta = 0$). Similar to the registration results obtained from the synthetic images, increasing α reduces $J_{\leq 0}(\boldsymbol{\varphi})$ to 0, delivering diffeomorphic deformation. We also expect that applying the total loss function with diffeomorphic regularization (3.7) can be better in decreasing $J_{\leq 0}(\boldsymbol{\varphi})$.

Table 5.2 displays the average Rel. SSD and average $J_{\leq 0}(\boldsymbol{\varphi})$ resulting from the validation set of the 870 image pairs of the real medical images with different diffeo-morphic regularization losses. Similar to the experiments conducted on the synthetic images, we further experimented $\mathcal{L}_{\text{diffeo},\sigma_1}$ with $\beta = 100$, 200, and 500, as $J_{\leq 0}(\boldsymbol{\varphi})$ had
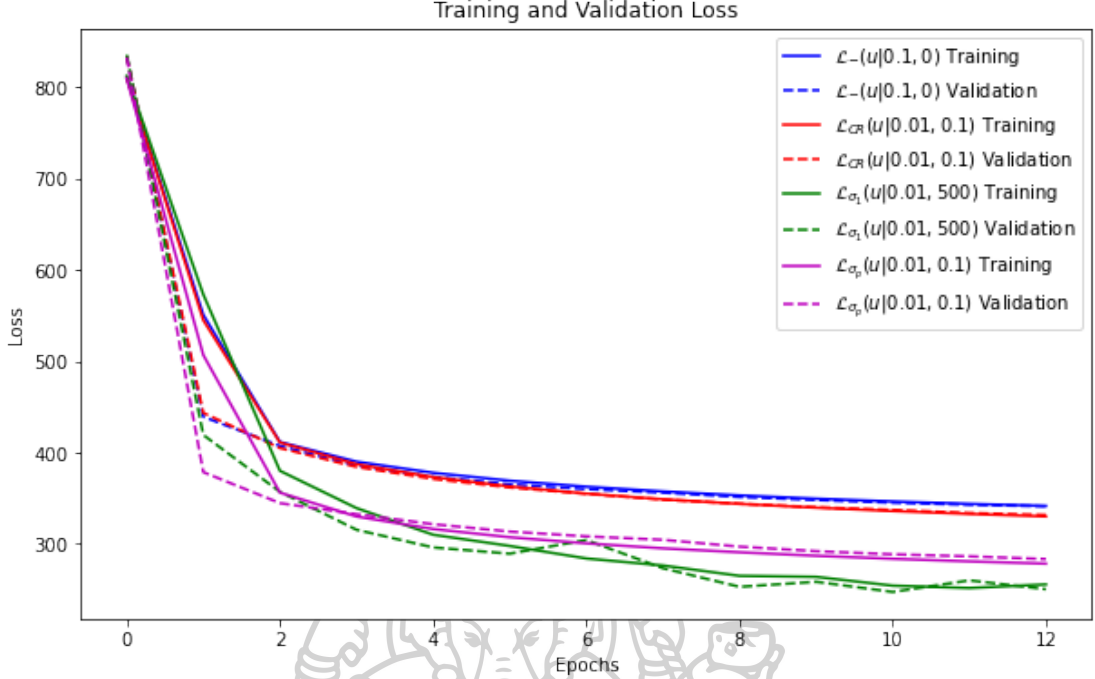
Figure 5.2: Loss value of training and validation sets during the training process on real medical images for different loss functions

not yet reached 0. The results also demonstrate that our proposed DL model is more effective for reducing $J_{\leq 0}(\boldsymbol{\varphi})$ in terms of Rel. SSD compared to other diffeomorphic regularization losses.

5.5.2 Test Set Performance

Similarly, we chose the appropriate α and β that yield $J_{\leq 0}(\boldsymbol{\varphi}) = 0$ for each experiment. Figure 5.2 shows the total loss function behavior throughout the training process on the real medical images. Here, we also compare $\mathcal{L}_{-}(\boldsymbol{u}|0.1,0)$, $\mathcal{L}_{\mathrm{CR}}(\boldsymbol{u}|0.01,0.1)$, and $\mathcal{L}_{\sigma_1}(\boldsymbol{u}|0.01,1000)$ with our proposed total loss function $\mathcal{L}_{\sigma_p}(\boldsymbol{u}|0.01,0.2)$. Similarly, the training and validation loss decrease during the training process so our DL model generalizes well to new data and is not overfitting on real medical images.

We also used the α and β that yield $J_{\leq 0}(\boldsymbol{\varphi}) = 0$ to compare the average Rel. SSD and $J_{\leq 0}(\boldsymbol{\varphi})$ resulting from the test set of the 870 image pairs of the real medical images,

Table 5.3: Comparison of average Rel. SSD and average $J_{\leq 0}(\varphi)$ from the test set of the 870 image pairs of the real medical images with different total loss functions.

Total loss function	Rel. SSD (%)	$J_{\leq 0}(\varphi)$
$\mathcal{L}_-(\boldsymbol{u}\vert 0.1, 0)$	40.87	0
$\mathcal{L}_{\mathrm{CR}}(\boldsymbol{u}\vert 0.01, 0.1)$	32.58	0
$\mathcal{L}_{\sigma_1}(\boldsymbol{u}\vert 0.01, 500)$	27.14	0
$\mathcal{L}_{\sigma_p}(\boldsymbol{u}\vert 0.01, 0.1)$	**26.76**	**0**

as shown in Table 5.3. These results demonstrate the effectiveness of our proposed diffeomorphic regularization loss in reducing $J_{\leq 0}(\varphi)$ more efficiently than other approaches for real medical images.

Figure 5.3 displays a registration problem from our test set of the real medical images. The first row shows the template T and reference R. The second and third rows show the deformed template images $T_{\boldsymbol{u}}$ and the constructed transformations from different total loss functions.

We display the absolute difference between the deformed template images and the reference images, denoted as $|T_{\boldsymbol{u}} - R|$, as shown in Figure 5.4. This difference visualization provides a clear view of the dissimilarities between the registration results obtained from different diffeomorphic regularization losses. We can see that the darker areas represent a smaller difference, indicating better registration. The contrasting light regions, conversely, indicate areas with larger differences or discrepancies between the deformed template and the reference image, signifying less effective registration.

Obviously, the registration results obtained from this experiment confirm the effectiveness of our proposed DL model in reducing $J_{\leq 0}(\varphi)$ while maintaining a competitive Rel. SSD when compared to other techniques. The overall performance improvement is also evident in the real medical images, demonstrating the robustness of the proposed method across various image types.

(a) T

(b) R

(c) $T_{\boldsymbol{u}}^{\mathcal{L}_-}$

Rel.SSD $= 16.44\%$

(d) $T_{\boldsymbol{u}}^{\mathcal{L}_{\mathrm{CR}}}$

Rel.SSD $= 12.44\%$

(e) $T_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_1}}$

Rel.SSD $= 11.33\%$

(f) $T_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_p}}$

Rel.SSD $= 9.82\%$

(g) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_-}) = 0$

(h) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_{\mathrm{CR}}}) = 0$

(i) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_1}}) = 0$

(j) $J_{\leq 0}(\boldsymbol{\varphi}_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_p}}) = 0$
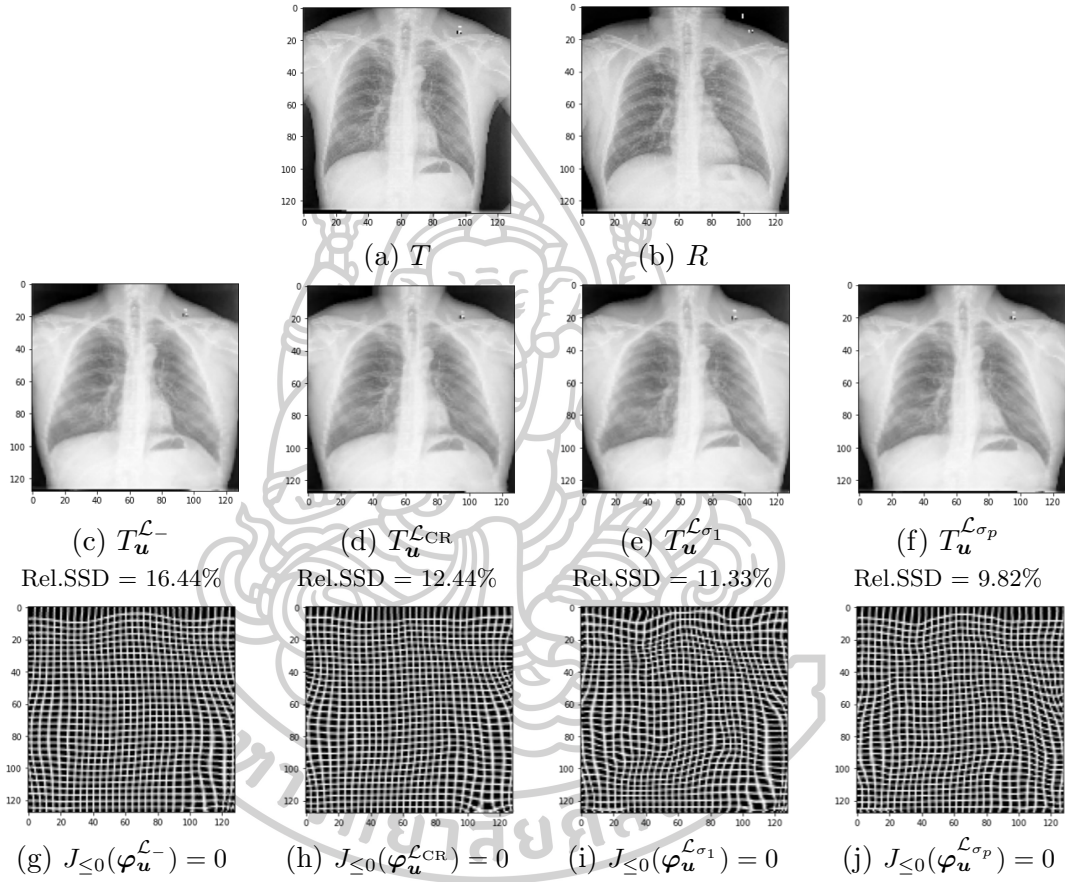
Figure 5.3: Registration results of a registration problem from the test set of the real medical images: (a) template T, (b) reference R, (c)-(f) and (g)-(j) are respectively the deformed template images and constructed transformation from $\mathcal{L}_-(\boldsymbol{u}|0.1, 0)$, $\mathcal{L}_{\mathrm{CR}}(\boldsymbol{u}|0.01, 0.1)$, $\mathcal{L}_{\sigma_1}(\boldsymbol{u}|0.01, 500)$, and $\mathcal{L}_{\sigma_p}(\boldsymbol{u}|0.01, 0.1)$ (our proposed total loss function).

$$|T_{\boldsymbol{u}}^{\mathcal{L}_-} - R| \qquad |T_{\boldsymbol{u}}^{\mathcal{L}_{\mathrm{CR}}} - R| \qquad |T_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_1}} - R| \qquad |T_{\boldsymbol{u}}^{\mathcal{L}_{\sigma_p}} - R|$$
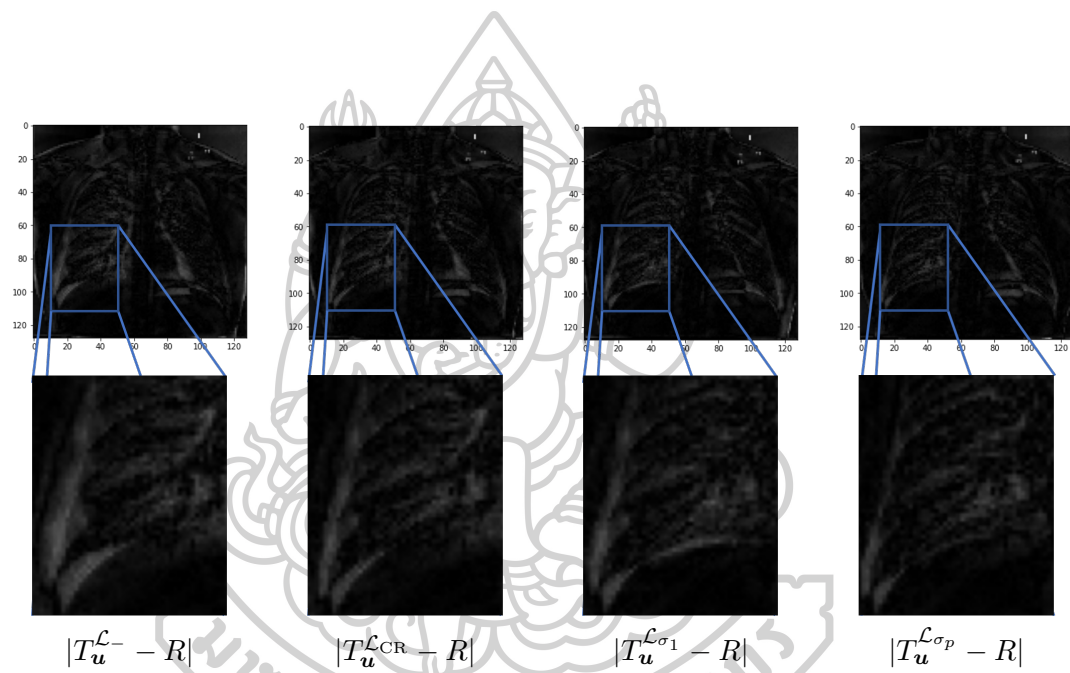
Figure 5.4: Absolute difference between the deformed template image and the reference image from different total loss functions. From left to right: Results from $\mathcal{L}_-$, $\mathcal{L}_{\mathrm{CR}}$, $\mathcal{L}_{\sigma_1}$, and $\mathcal{L}_{\sigma_p}$ (our proposed total loss function).

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

This work contributes to the field of image registration, with a particular focus on medical image analysis. The study aimed at enhancing the registration process by proposing a DL model equipped with diffeomorphic regularization losses, thereby ensuring the topological properties of the transformation.

Our proposed DL model was evaluated against various total loss functions, and the results were thoroughly analyzed. The comparative analysis was made based on the Rel. SSD and $J_{\leq 0}(\varphi)$.

From our results, it is evident that our proposed DL model achieved significant improvements in both synthetic and medical images. Specifically, we demonstrated that by adjusting the parameters α and β, it is possible to achieve diffeomorphic transformations, where $J_{\leq 0}(\varphi)$ is reduced to zero. This result signifies that the transformations constructed by the proposed model are diffeomorphic, thereby preserving the topological properties of the images.

Moreover, the results from the validation and test sets further reinforced the model's effectiveness. Our DL model generalized well to new data. The successful results on the medical images confirm the applicability of the proposed method in real-world scenarios, particularly in medical imaging, which is known for its complexity and high demand for precision.

6.2   Future Work

Although our proposed DL model has shown promising results, there are still areas for future work and improvement:

1. Currently, the model is trained and tested on a limited dataset. To further validate the effectiveness of our approach, we would perform experiments on larger and more diverse datasets, covering different imaging modalities and anatomical structures.

2. The optimization of hyperparameters α and β plays a crucial role in the performance of our model. Therefore, developing an adaptive strategy for selecting the appropriate values for these parameters during the training process could potentially lead to further improvements in registration performance.

3. We plan to extend our model to 3D medical image registration, taking advantage of the additional spatial information available in three-dimensional images. This could potentially provide even better registration results in a broader range of medical imaging applications.

4. Our model can be further improved by incorporating more advanced deep learning architectures, such as the use of attention mechanisms, which could help the model focus on specific regions of interest within the images, thus improving registration accuracy in challenging cases.

By addressing these areas in future work, we believe that our proposed diffeomorphic regularization model can be further refined and applied to a wide range of medical image registration tasks, ultimately contributing to the advancement of the field of medical image analysis.

REFERENCES

[1] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 international conference on engineering and technology (ICET)*, 1–6.

[2] Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*(1), 95–113.

[3] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2018). An unsupervised learning model for deformable medical image registration. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 9252–9260.

[4] Boveiri, H. R., Khayami, R., Javidan, R., & Mehdizadeh, A. (2020). Medical image registration using deep neural networks: A comprehensive review. *Computers & Electrical Engineering*, *87*, 106767.

[5] Brown, L. G. (1992). A survey of image registration techniques. *ACM computing surveys (CSUR)*, *24*(4), 325–376.

[6] Candemir, S., Jaeger, S., Palaniappan, K., Musco, J. P., K, R. K. S., Xue, Z., Karargyris, A., Antani, S., Thoma, G., & McDonald, J. C. (2013). Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE transactions on medical imaging*, *33*(2), 577–590.

[7] Chen, X., Diaz-Pinto, A., Ravikumar, N., & Frangi, A. F. (2021). Deep learning in medical image registration. *Progress in Biomedical Engineering*, *3*(1), 012003.

[8] Chumchob, N. (2013). Vectorial total variation-based regularization for variational image registration. *IEEE transactions on image processing*, *22*(11), 4551–4559.

[9] Chumchob, N., & Chen, K. (2010). A variational approach for discontinuity-preserving image registration. *East-West Journal of Mathematics*, *2010*, 266–282.

[10] Chumchob, N., & Chen, K. (2020). An augmented Lagrangian method for solving total variation (TV)-based image registration model. *Journal of Algorithms & Computational Technology*, *14*, 1748302620973534.

[11] Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2018). Unsupervised learning for fast probabilistic diffeomorphic registration. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, 729–738.

[12] Dawn, S., Saxena, V., & Sharma, B. (2010). Remote sensing image registration techniques: A survey. *Image and Signal Processing: 4th International Conference, ICISP 2010, Trois-Rivières, QC, Canada, June 30-July 2, 2010. Proceedings 4*, 103–112.

[13] De Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., & Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, *52*, 128–143.

[14] Ehrhardt, J., Werner, R., Schmidt-Richberg, A., & Handels, H. (2010). Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE transactions on medical imaging*, *30*(2), 251–265.

[15] Eppenhof, K. A., & Pluim, J. P. (2018). Pulmonary CT registration through supervised learning with convolutional neural networks. *IEEE transactions on medical imaging*, *38*(5), 1097–1105.

[16] Fischer, B., & Modersitzki, J. (2002). Fast diffusion registration. *Contemporary Mathematics*, *313*, 117–128.

[17] Fischer, B., & Modersitzki, J. (2003). Curvature based image registration. *Journal of Mathematical Imaging and Vision*, *18*, 81–85.

[18] Frohn-Schauf, C., Henn, S., & Witsch, K. (2008). Multigrid based total variation image registration. *Computing and Visualization in Science*, *11*(2), 101–113.

[19] Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2020). Deep learning in medical image registration: A review. *Physics in Medicine & Biology*, *65*(20), 20TR01.

[20] Han, H., & Wang, A. (2021). A fast multi grid algorithm for 2D diffeomorphic image registration model. *Journal of Computational and Applied Mathematics*, *394*, 113576.

[21] Han, H., & Wang, Z. (2020). A diffeomorphic image registration model with fractional-order regularization and Cauchy–Riemann constraint. *SIAM Journal on Imaging Sciences*, *13*(3), 1240–1271.

[22] Haskins, G., Kruger, U., & Yan, P. (2020). Deep learning in medical image registration: A survey. *Machine Vision and Applications*, *31*, 1–18.

[23] Hömke, L., Frohn-Schauf, C., Henn, S., & Witsch, K. (2007). Total variation based image registration. *Image Processing Based on Partial Differential Equations: Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems, CMA, Oslo, August 8–12, 2005*, 343–361.

[24] Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R. K., Antani, S., et al. (2013). Automatic tuberculosis screening using chest radiographs. *IEEE transactions on medical imaging*, *33*(2), 233–245.

[25] Janssens, G., Jacques, L., De Xivry, J. O., Geets, X., & Macq, B. (2011). Diffeomorphic registration of images with variable contrast enhancement. *Journal of Biomedical Imaging*, *2011*, 1–12.

[26] Jia, X., Thorley, A., Chen, W., Qiu, H., Shen, L., Styles, I. B., Chang, H. J., Leonardis, A., De Marvao, A., O'Regan, D. P., et al. (2021). Learning a model-driven variational network for deformable image registration. *IEEE Transactions on Medical Imaging*, *41*(1), 199–212.

[27] Krepl, J., Casalegno, F., Delattre, E., Erö, C., Lu, H., Keller, D., Rodarie, D., Markram, H., & Schürmann, F. (2021). Supervised learning with perceptual similarity for multimodal gene expression registration of a mouse brain atlas. *Frontiers in Neuroinformatics*, *15*, 691918.

[28] Kuang, D., & Schmah, T. (2019). Faim–a convnet method for unsupervised 3D medical image registration. *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, 646–654.

[29] MacGillivray, S. A. (2009). Curvature-based image registration: Review and extensions. *An essay presented to the University of Waterloo in fulfillment of the essay requirement for the degree of Master of Mathematics in Combinatorics and Optimization Waterloo, Ontario, Canada*.

[30] Maintz, J. A., & Viergever, M. A. (1998). A survey of medical image registration. *Medical image analysis*, *2*(1), 1–36.

[31] Oliveira, F. P., & Tavares, J. M. R. (2014). Medical image registration: A review. *Computer methods in biomechanics and biomedical engineering*, *17*(2), 73–93.

[32] Pöschl, C., Modersitzki, J., & Scherzer, O. (2010). A variational setting for volume constrained image registration. *Inverse Problems and Imaging*, *4*(3), 505–522.

[33] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241.

[34] Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. *Towards Data Sci*, *6*(12), 310–316.

[35] Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, *9*, 82031–82057.

[36] Sotiras, A., Davatzikos, C., & Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, *32*(7), 1153–1190.

[37] Thompson, T., & Chen, K. (2020). An effective diffeomorphic model and its fast multigrid algorithm for registration of lung CT images. *Computational Methods in Applied Mathematics*, *20*(1), 141–168.

[38] Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, *45*(1), S61–S72.

[39] Wachowiak, M. P., Smolíková, R., Zurada, J. M., & Elmaghraby, A. S. (2002). A supervised learning approach to landmark-based elastic biomedical image registration and interpolation. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, *2*, 1625–1630.

[40] Wodzinski, M., & Müller, H. (2020). Unsupervised learning-based nonrigid registration of high resolution histology images. *Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 11*, 484–493.

[41] Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, *5*(23), 495.

[42]   Zhang, D., & Chen, K. (2018). A novel diffeomorphic model for image registration and its algorithm. *Journal of Mathematical Imaging and Vision, 60*, 1261–1283.

VITA

NAME   Apinan Sanongsin

DATE OF BIRTH

PLACE OF BIRTH   Ratchaburi, Thailand

INSTITUTIONS ATTENDED 2016 – 2019 Bachelor of Science in Mathematics, Silpakorn University

2020 – 2023 Master of Science in Mathematics, Silpakorn University

HOME ADDRESS   52/44 Moo 2 Tambol Pak-rat, Amphoe Banpong, Ratchaburi, 70110, Thailand