Journal of Engineering Research

Volume 5 | Issue 2 Article 2

2021

Nonrigid Medical Image Registration using Adaptive Gradient Optimizer

Mohammed Abdel Rhman Abo Arab, Heba Ali El-Khobby, Amira S. Ashour

Vol. 5, No. 2 – 2021 Journal of Engineering Research (ERJ)

Nonrigid Medical Image Registration using Adaptive Gradient Optimizer

Mohammed A. AboArab, Heba Ali El-Khobby, and Amira S. Ashour

Electronics and Electrical Communication Engineering Department, Faculty of Engineering, Tanta University, 31527, Egypt

Abstract- **Medical image registration has a significant role in several applications. It has sequential processes, including transformation, similarity metric calculation, diffusion regularization, and optimization of the transformation parameters (i.e., rotation, translation, and shear). The optimization process for determining the optimal set of the transformation vectors is considered the main stage affecting the performance of the registration process. Hence, medical image registration can be deliberated as an optimization problem for computing the geometric transformations to realize maximum similarity between the moving image and the fixed one. In this paper a mono-modal nonrigid image registration using B-spline is designed for the alignment of Computed Tomography (CT) images using Adaptive Gradient algorithm (AdaGrad) optimizer. In addition, a comparative study with other first order optimizers, such as Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam) algorithm (AdaMaX), AdamP, and RangerQH were conducted. Also, a comparison with the limited memory Broyden-Fletcher-Goldfarb-Shannon (LBFGS) as a second order optimizer was also carried out. The results showed the superiority of the AdaGrad optimizer by 56.99% and 48.37% improvement in the target registration error (TRE) compared to the SGD, and the LBFGS optimizer, respectively.**

Keywords- **Non-rigid registration; Adaptive Gradient optimizer; Stochastic Gradient Descent; Adaptive Moment Estimation optimizer; limited memory Broyden-Fletcher-Goldfarb-Shannon optimizer.**

I. INTRODUCTION

Image registration has an essential role in registering different images captured from different sensors, different time instances, or different shooting conditions. Medical image registration has three main steps: determining the proper transformation between the fixed and moving images, measuring the similarity degree of the fixed and the moving images, and optimizing the transformation process. Accordingly, the image registration problem can be formulated as an optimization problem [1]. Since in the medical applications, respiration and other physiological processes lead to motion during the acquisition process of the medical images/videos, nonrigid registration is considered using nonlinear transformations.

Such transformation techniques include affine, displacement field and B-spline [2]. Spline-based registration techniques use corresponding control points (features) in the fixed and moving images. Also, a spline function is defined to delineate the correspondences away from these points. B-splines are considered one of the effective transformations used in the image registration method. It is used to optimize the correlation between images by adjusting a series of parameters [3, 4], which have different varieties, including the convex nuclear B-splines, linear interpolation B-splines, and the cubic B-splines [5]. After transformation between corresponding points in the fixed and moving images, the similarity measurement is calculated during the optimization of the transformation process to determine whether the optimal match between the two images is achieved by measuring the similarity between them. Based on the differences in the landmarks of an image, the normalized correlation coefficient (NCC) is considered for evaluating the similarity degree between the two images [6].

Typically, the optimization methods have a significant role in the registration process to obtain the required optimum transformation parameters for accurate registered image [1, 7, 8]. Proper selection of the used optimization technique guarantees the optimal transformation parameters for fast and reliable registration process. In non-rigid registration applications, selecting the used optimizer is complex because more non-rigid the model, the more parameters are mostly required. This leads to long computational time for determining the parameters with the possibility of the local minima problem. The transformation parameters are computed by finding an optimum cost function to define the parameters being optimized.

The cost function measures the similarity between the two images for the used transformation. It is less complex in the mono-modal registrations since there is a linear relationship between the fixed and the moving image and the similarity metric is straightforward. It also contains regularization terms for smoothness and diffeomorphic constraints to maintain the topology.

Accordingly, the optimization methods can be categorized in terms of the gradient information into [9]: i) first-order optimization methods, which uses stochastic gradient methods, ii) high-order optimization methods (Newton’s method), which is LBFGS as a model example, and iii) heuristic derivative-free optimization methods, which is based on the coordinate descent methods. In image registration, the typically used optimizers are based on the gradient descent, adaptive stochastic gradient descent, nonlinear conjugate gradient, quasi-Newton, stochastic gradient descent methods, preconditioned stochastic gradient descent, Robbins-Monro, Kiefer-Wolfowitz, and simultaneous perturbation [1].

Several studies were implemented using different optimization methods during the image registration process. Xu *et al*. [10] developed an iterative consistent-feature-guided registration technique in the forward and inverse matching direction for 4D image registration by using the adaptive stochastic gradient descent optimizer. The initial forward registration took 1 hour, and the inverse registration took 2 hours. In addition, Hong *et al*. [11] developed a nonrigid registration technique for liver CT and Magnetic Resonance (MR) images using L-BFGS optimization method. The affine transformation using B-splines free-form deformation (FFD) was applied, where the normalized mutual information (NMI) was used for similarity measurement. This approach was tested on three groups of CT and liver MR images.

The B-splines FFD transformation showed accurate performance, although it is time-consuming. Before and after registration using affine transformation, B-spline transformation, and the combination between both of them, the achieved the correlation coefficient (CC) were 0.813, 0.930, 0.955, and 0.965, respectively, also, the required registration times were 48.7, 366.4, and 188.3 seconds, respectively. The results showed improved registration accuracy and reduced running time. From the preceding studies, it is obvious that the used optimization method provides optimal parameters for proper match between the fixed and moving image with the fixed one.

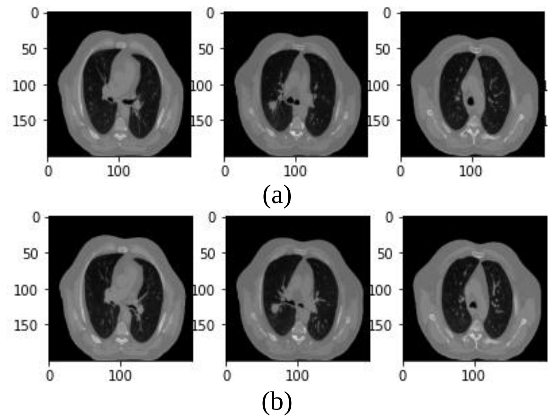
Generally, several registration methods are based on stochastic gradient descent (SGD) [12] for optimization. The SGD uses one sample randomly to update the gradient per iteration instead of directly calculating the exact gradient value. Since it is challenging to choose a suitable learning rate during the optimization of all parameters, the AdaGrad [13] is considered in this paper, which is a straightforward improvement of the SGD [9]. The AdaGrad sets the learning rate dynamically as it relies on the historical gradient in the previous iterations. It is free of parameter adjustment, and fast to converge.

Moreover, a comparative study with the update of the Adaptive Moment Estimation (Adam) algorithm (AdaMaX) [14], the AdamP optimizer [15], RangerQH, and the L-BFGS optimizer was conducted versus the AdaGrad optimizer in this paper. The proposed AdaGrad-based registration to optimize the transformation process in the nonrigid registration taking less computational time compared to the other first order optimizers i.e., SGD optimizer and the LBFGS as a second order optimizer under the same parameter's settings [9].

The following sections include the methodology of the proposed optimization-based nonrigid registration method, including the image pre-processing, the Bspline transformation, the calculations of the NCC as a similarity measurement, and the optimization methods in section II. Afterward, the results are analyzed with comparative studies in section III. Finally, the conclusion of the present work is highlighted in section IV.

II. Materials and Methods

The present work proposed an optimization-based mono-model non-rigid medical image registration using the following stages of the image registration, including transformation, similarity metric computing, regularization, and optimization of the transformation parameters. In addition, we conducted a comparative study with first order optimizers, and a second order optimizer. For evaluating the proposed AdaGrad-based registration method, a CT dataset including four-dimensional records of size $482 \times 360 \times 141$ was used [16]. The dataset includes ten CT files, where 40 feature points are available for each image. Fig. 1 includes a sample of the used images from the dataset.



(a)

(b)

Fig. 1. Sample from the used dataset, where (a) fixed images from the time frame 0, and (b) moving images from the time frame 5.

Initially, linear normalization [17] was applied on the dataset as a pre-processing stage before applying the image registration processes. Then, the Gaussian image pyramid [18, 19] was used for down sampling to reduce processing time and memory size.

A. *Non-rigid medical image registration*

The goal of image registration is to link any point in the fixed image to the moving image for finding the optimal transformation to align the moving image with the fixed one. Recently, the B-spline transformation technique has been proposed for image registration [20, 21], which maps any point in the fixed image $F(x,y,z,t)$ at time t into its corresponding point in the moving image $M(x,y,z,t_0)$, which is acquired at time t_0. Since the dataset is 4D, the B-spline transformation of a volume can be expressed as [20-22]:

$$T_{B-spline}(x,y,z) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} b_{ijk}\ N_i(x)N_j(y)N_k(z) \cdot \qquad (1)$$

where $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, and $k = \lfloor z/n_z \rfloor - 1$, also, N_l represents the basis function of the B-spline model. In addition, b_{ijk} are the control points $n_x \times n_y \times n_z$ of the B-spline.

Then, the degree of similarity/dissimilarity between two the compared images is evaluated by measuring the NCC metric [23]. The NCC has a range between –1 and 1, where NCC=1 refers to perfect positive correlation between the two images, while NCC= -1 refers to perfect negative correlation. Thus, for a fixed and moving images where both images have the same size N, the NCC can be expressed using the following formula [23]:

$$NCC = \frac{\sum_{i=0}^{N-1}(I_f - E(I_f))(T(I_g) - E(T(I_g)))}{\sqrt{\sum_{i=0}^{N-1}(I_f - E(I_f))^2}\sqrt{\sum_{i=0}^{N-1}(T(I_g) - E(T(I_g)))^2}} . \qquad (2)$$

where $E(I_f)$ and $E(T(I_g))$ are the averages of brightness for fixed image and transformation of moving image, respectively.

To penalize the displacement fields, the diffusion regularization, which is an L_2 parameter norm penalty, is added. The diffusive regularization smoothies the transformation by penalizing the gradients in the x, y or z components of the displacement field, where the diffusion regularization is defined by [24]:

$$S_{diff}(u)=\frac{1}{2}\sum_{l=x,y,z}\int_{\Omega_f}\left\|\nabla u_l(X)\right\|^2 dX \cdot \qquad (3)$$

where u is the displacement field, $\nabla u_l(X)$ is the gradient of the L^{th} scalar component of the displacement field evaluated at X, $X\in\Omega_f$ and Ω_f is the fixed image domain.

B. Optimization-based B-spline nonrigid registration

Medical image registration is considered an iterative optimization method using the NCC and regularization to select the optimal search strategy to finally determine the optimal transformation process. The required optimal value occurs once the registration outcome achieves the optimum similarity measurement. This iterative optimization process aims to reduce the search time to increase the time sensitivity of the process. Consequently, some of the first order optimizers can be applied, such as SGD, AdaGrad, AdaMaX, AdamP, RangerQH, and the LBFGS. To obtain the optimal transformation's process, we minimize an objective function T^* using the following formula [25]:

$$T^*=\arg_T \min NCC\left(T\left(I_m\right),I_g\right)+\lambda S_{diff}\left(T,\Omega_f\right). \qquad (4)$$

where the first term represents the objective associated with the NCC in (2), while the second term corresponds to the diffusion regularizer in (3), where λ is the regularization weight, $\lambda\in[0,\infty)$ there is no regularization, if λ equal 0 and the regularization is higher when λ is a higher value [24] , and Ω_f is the domain of the fixed image.

a. Stochastic gradient descent optimizer

Stochastic gradient descent (SGD) updates the weights of the learning process in each training phase image. However, such updates engender huge fluctuation in the loss function due to the increased variance between the two images. The weights are updated using the following expression [26]:

$$w_{t+1}=w_t-\eta.\nabla_w J(w;x^{(i)};y^{(i)}), \qquad (5)$$

where w_t represents the weights at step t and its initially by the 40 landmarks points in the thorax 4D-CT datasets that are used to evaluated the proposed model, η is the learning rate hyperparameter and the default value of it is around 0.001 in all optimizers, $J(.)$ is the cost function, and $\nabla_w J(w;x^{(i)};y^{(i)})$ is the gradient of weight parameters w_t.

b. Adaptive Gradient optimizer

The SGD updates all weights w at once, where the same learning rate η is used with every parameter w_t . Conversely, the AdaGrad is a gradient-based optimizer, which adapts the learning rate of the parameters. It performs smaller updates for frequent features, and larger updates for infrequent features [13]. The AdaGrad uses a different learning rate for every parameter w_t at every time step t. For brevity, set g_t^i to be the gradient of the cost function with respect to the parameter w_t at time step t using the following formula [27]:

$$g_t^i=\nabla_w J(w_t^i) , \qquad (6)$$

Then, the SGD update for every parameter w_t at each time step t becomes in the AdaGrad as follows:

$$w_{t+1}^i=w_t^i-\eta.g_t^i , \qquad (7)$$

In the updating rule, the AdaGrad modifies η at each time step t for every weight w_t based on the past gradients that have been computed for w_t using the following formula [13, 27].

$$w_{t+1}^i=w_t^i-\frac{\eta}{\sqrt{\sum_{\tau=1}^{t}(g_t^i)^2+\varepsilon}}.g_t^i \cdot \qquad (8)$$

The sum of squares $\left(\sum_{\tau=1}^{t}\left(g_t^i\right)^2\right)^{\frac{1}{2}}$ is used to scale the learning rate as it gives a low learning rate for the more frequent gradients, and a high learning rate for the least frequent gradients. The main advantage of the AdaGrad is its ability to eliminate the manual tuning of the learning rate along with its fast procedure.

c. Adaptive Moment Estimation optimizer

AdaMax is the update of the Adaptive Moment Estimation (Adam) algorithm, where the uncentered variance tends to ∞ [14]. The weights of the learning process are updated using the following formula:

$$w_{t+1}=w_t-\frac{\eta}{u_t}\hat{m}_t , \qquad (9)$$

$$u_t=\max(\beta_2.v_{t-1},|g_t|) , \qquad (10$$

$$v_t=\beta_2 v_{t-1}+(1-\beta_2)|g_t|^2 . \qquad (11)$$

where m_t is the first moment, v_t is the second moment and u_t depends on the maximum operation, it is not as suggestible to bias towards zero as v_t and m_t in Adam.

d. AdamP

The AdamP is a modified version of the Adam optimizer to slow down the slowdown for momentum optimizers on scale-invariant weights [15]. The momentum is designed to speed the convergence of gradient-based optimization by letting w get-away high-curvature regions and cope with small and noisy gradients. The momentum is updated as follows:

$$w_{t+1}=w_t-\eta g_t , \qquad (12)$$

$$g_t=\beta g_{t-1}+\nabla_{w_t} J(w_t) . \qquad (13)$$

These optimization methods start with an initial guess for w to minimize a cost function $J(w)$. Then, iteratively, sequences of improving solutions for w are generated until a result criterion is satisfied.

e. Limited memory Broyden-Fletcher- Goldfarb-Shannon (LBFGS)

Newton optimization methods use the second order gradient information to calculate the step direction by calculating the inverse of the Hessian matrix [28]. In a high dimensional setting, the calculation of the inverse of a Hessian matrix can get too expensive because the parameter vector w is very large. To overcome the problem of calculating the inverse of the Hessian matrix in each iteration, the "Quasi-Newton method is used by continuously update the approximation of the inverse of the Hessian matrix in each iteration. The approximation of the inverse of the Hessian matrix is updated in each iteration based on the current s_t and y_t values, where s_t is the position difference and y_t is the gradient difference in the iteration. These vectors are of the same length as the vector W, which are defined as follows:

$$s_t = w_{t+1} - w_t \, , \tag{14}$$

$$y_t = \Delta J_{t+1} - \Delta J_t \, . \tag{15}$$

The BFGS needs to keep an approximation of the inverse of the Hessian matrix in each iteration (an $m \times m$ matrix), where m is the length of the parameter vector w. The L-BFGS (Limited-memory BFGS) algorithm modifies BFGS to obtain Hessian approximations that can be stored in a few vectors of the length m. In L-BFGS, the { s_t , y_t } pairs are stored from the last n iteration which causes the algorithm to need $2 \times n \times m$ storage compared to $m \times m$ storage in the BFGS algorithm.

From the above mentioned optimizers, the main advantage of the AdaGrad optimizer compared to the other gradient descent ones is that the learning rate is no longer fixed during the parameter update process but is computed using all the previous gradients accumulated up to this iteration [9]. Hence, the AdaGrad optimizer eliminates the need to tune the learning rate manually and convergence is faster than simple SGD when the scaling of the weights is unequal, and this leads to an increase to reach an optimal transformation process in medical image registration.

In AdaGrad, each parameter's learning rate is adaptively modified, resulting in a greater learning rate, a lower cumulative gradient, and a quicker learning speed in the early stages of training. This leads to shorter processing time and good accuracy. However, selecting an appropriate learning rate in SGD is complex; also, it is not recommended to use the same learning rate with all parameters. In rare cases, the optimization may become stuck at a local point, causing the accuracy measure to fall short of AdaGrad. The LBFGS method has several disadvantages as it requires huge storage space, costly, and takes long time to calculate the second derivatives, making it unsuitable for handling large-scale problems like medical image registration.

C. Proposed AdaGrad-based B-spline registration

Initially, the two input images (fixed and moving image) are preprocessed by normalization and Gaussian image pyramid to reduce processing time and memory size. Then, the B-spline transformation is used to match any point in the fixed image to its corresponding in the moving image. To find the optimal transformation process, the NCC is calculated for similarity measures between the landmarks and their average in the fixed image and their corresponding in the B-spline transformation of the moving image. Afterward, the diffusion regularization is applied to avoid over-fitting. The block diagram of the proposed AdaGrad-based B-spline registration of the thorax CT images is illustrated in Fig. 2.

Table 1 Comparison between different optimizers.

Optimization Method	Advantages	Disadvantages
AdaGrad	It is a precise and quick method with an adaptable learning rate.	It provides accurate performance with increasing the training time.
SGD	It is a simple approach and takes less time of processing.	It is not accurate to reach the optimal parameters because it stops in a local point.
LBFGS	It can be more accurate using big storage memory.	It is more complicated and time-consuming to process.

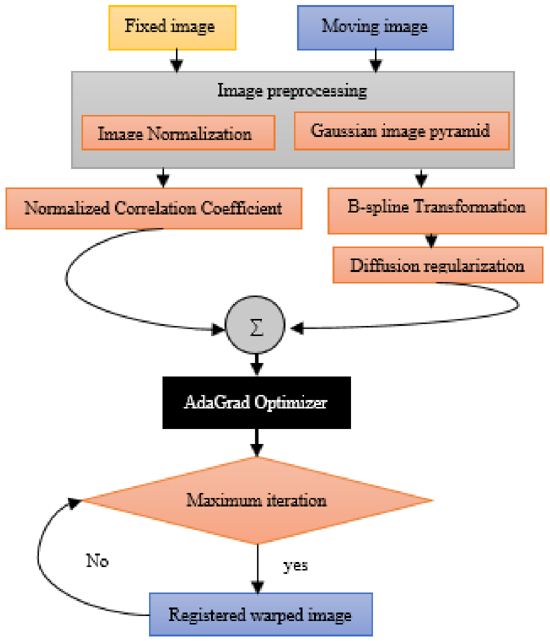


Fig. 2. Framework of our proposed model of medical image registration.

In Fig.2, a summing operator is used to sum the NCC output and the diffusion regularization output before feeding the summed value to the AdaGrad optimizer. The early stopping effect [29] is used to prevent overfitting by forcing the AdaGrad optimizer to stop after reaching a certain number of local points, even if they have not yet reached the maximum number of iterations, thereby reducing processing time. After reaching the maximum number of iterations, to get the optimal transformation and the output is registered warped image. The AdaGrad optimizer method is suitable for dealing with gradient problems where the learning rate of each parameter is adjusted adaptively according to the sum of the squares of all previous gradients using the following algorithm.

Algorithm: AdaGrad optimization

Input the global learning η and the initial parameter w

Set the constant ε to be $10e-7$ for stability

While stopping criterion not met do

Compute gradient: $g_t^i \leftarrow \nabla_w J(w_t^i)$

Compute update: $w_{t+1}^i \leftarrow w_t^i - \frac{\eta}{\sqrt{\sum_{\tau=1}^{t}(g_\tau^i)^2+\varepsilon}} \cdot g_t^i$

Minimize $\left\{NCC\left(T\left(I_m\right), I_g\right) + \lambda S_{diff}\left(T, \Omega_f\right)\right\}$

Apply update.

End while

Output the optimal parameters of the B-spline transform

The proposed model is evaluated by measuring the following metrics:

(i) TRE between the set of landmark points in the fixed and moving images [30]:

$$E_{t_2,t_1} = \frac{1}{\left|P_{t_1}\right|} \sum_{p_{t_1,c} \in P_{t_1}} \left\| T^{t_2,t_1}(p_{t_2,c}) - p_{t_1,c} \right\|. \qquad (16)$$

where E_{t_2,t_1} is the TRE between landmarks of time t_2 and t_1, $p_{t_1,c}$ is landmark c in time t_1, t_2 is the moving time point and T^{t_2,t_1} is the estimated transformation.

(ii) NCC loss as mentioned in (3) to measure the correlation between fixed and moving images.

(iii) The diffusion regularization loss as mentioned in (4) to penalize the gradient in x, y and z components.

III. Results and Discussion

The following section details the results obtained from the proposed technique of medical image registration which is implemented using Google CoLAB with 12GB of graphics memory and 25GB of RAM. In this model of image registration, the CT dataset was used to study the behavior of these six optimizers and determine the best performance optimizer which is evaluated by metrics i.e., TRE, NCC, and $L_2\, loss$. This study showed that the proposed AdaGrad-based registration is better than others and gives low TRE in transformation from moving image to fixed image. The registered warped image takes approximately 50.7 minutes to be registered.

This model is implemented by using a Gaussian image pyramid with four tap low pass filters with $\sigma = [9,9,9]$ [25] was used. After the B-spline transformation is applied on the moving points and the NCC is measured, the L_2 Diffusion regularization term is add to NCC with the optimal values of regularization weight λ are $\left[1e^{-2}, 1e^{-1}, 1e^{-0}, 1e^{+2}\right]$ on each level of Gaussian image pyramid, respectively, to minimize the objective function as mentioned in (5) by the optimizers.

A. AdaGrad results

As shown in Fig. 3, the AdaGrad optimizer reduced TRE by 76.8% than before registered images. The TRE is computed on 40 landmarks among $0^{th},\ldots,9^{th}$ time frames were used, where $\{case0, m\}_{m=0}^{9}$ represents a registered case between a time frame 0 as a fixed image and a time frame m as a moving image.

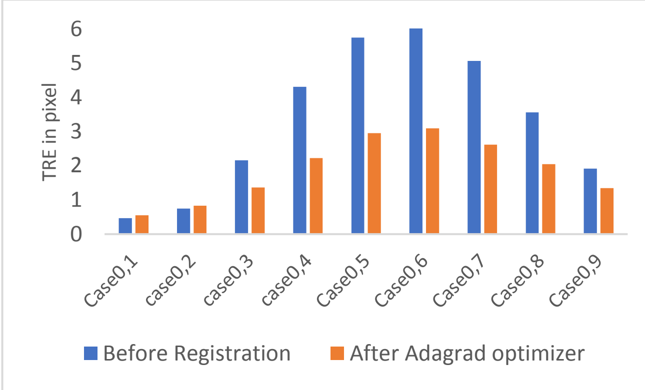


Fig. 3 The TRE before and after registration by using AdaGrad optimizer.

The TRE on $case_{0,1}$ represents a fixed image time frame 0 and moving image time frame 1 is a small value before and after registration, and this is represented in the registered output image as in Fig. 4. In Fig. 4, (a) represents a fixed image time frame 0, and (b) moving image time frame 1 and two-time frames of images are close together and this led to a pure B-spline registered image which is close to the fixed image of it. Against to $case_{0,5}$ which is represented a fixed image time frame 0 and moving image time frame 5, the TRE before registration technique is a large value as shown in Fig. 3. This is represented in Fig. 4. (a) and (d) showing a moving image time frame 5, and the difference between fixed and moving images is clear so, after registered two-time frames of images the output B-spline registered image is not good as registered time frames 0 and 1 as shown in (g) which represents a B-spline registered image of time frame 0 and 5.

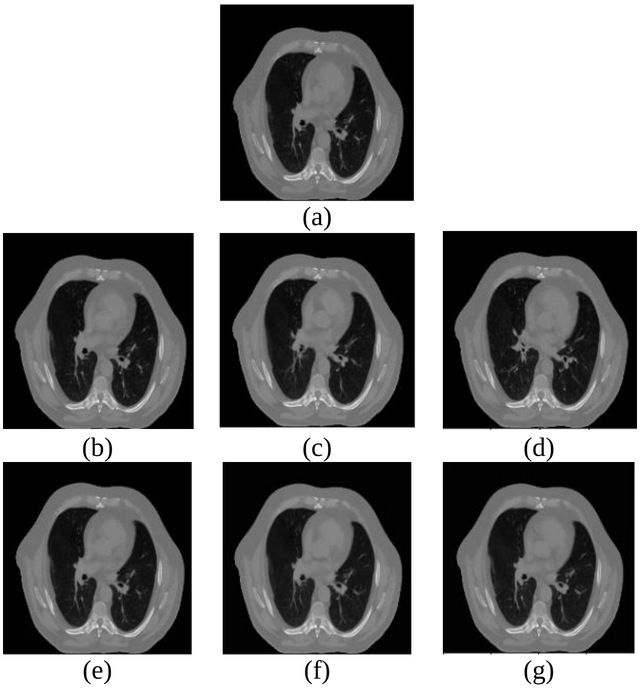


Fig. 4. Sample images, where (a) fixed image of time frame 0, (b) moving image of time frame 1, (c) moving image of time frame 3, (d) moving image of time frame 5, (e) B-spline registered image of time frame 1, (f) B-spline registered image of time frame 3, and (g) B-spline registered image of time frame 5.

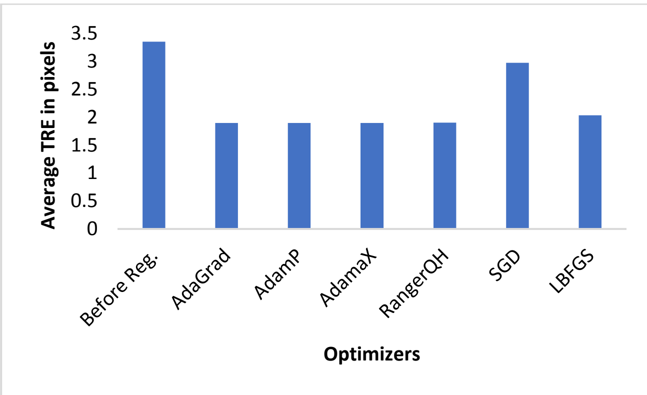
B. Comparative study of the proposed AdaGrad with first- and second-order optimizers

These optimizers are iterated by $[300, 200, 100, 50]$ and the $step\ size = \left[1e^{-2}, 4e^{-3}, 2e^{-3}, 2e^{-3}\right]$ for each level, respectively. From our experiments the increase in the iterated numbers does not affect the evaluation metrics but makes the time increase and this is not desirable. Table 1 shows the value of TRE before registration and the value of it after registration by using AdaGrad, AdamP, AdamaX, RangerQH and SGD which are considered as first order optimizers and LBFGS which is a type of second order optimizer.

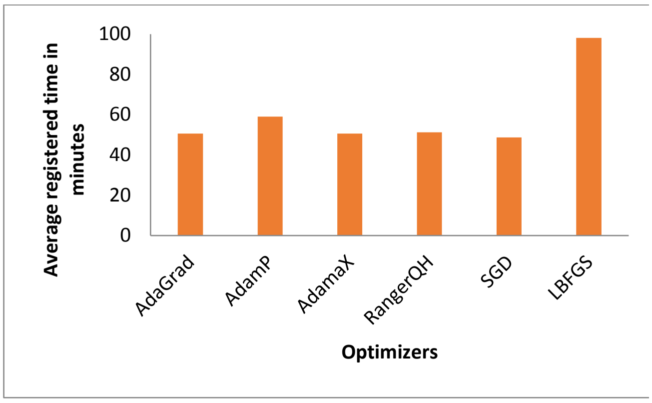
Table 2. The comparative between AdaGrad, AdamP, AdamaX, RangerQH, SGD and LBFGS optimizers in TRE (in pixels).

	Before Reg.	AdaGrad	AdamP	AdamaX	RangerQH	SGD	LBFGS
	$E_{0,n}$	$E_{0,n}$	$E_{0,n}$	$E_{0,n}$	$E_{0,n}$	$E_{0,n}$	$E_{0,n}$
Case0, 1	0.468	0.549	0.553	0.55	0.568	0.53	0.71
Case0, 2	0.745	0.836	0.868	0.862	0.871	0.76	0.87
Case0, 3	2.161	1.368	1.326	1.322	1.325	1.99	1.67
Case0, 4	4.305	2.222	2.219	2.228	2.221	3.82	2.25
Case0, 5	5.747	2.954	2.953	2.992	2.989	5.08	3.27
Case0, 6	6.154	3.096	3.106	3.094	3.093	5.35	3.53
Case0, 7	5.066	2.619	2.623	2.613	2.627	4.38	2.68
Case0, 8	3.563	2.051	2.069	2.067	2.051	3.09	1.96
Case0, 9	1.914	1.347	1.34	1.347	1.35	1.75	1.34

Table 2 demonstrated that the AdaGrad optimizer outperformed to the other optimizers in TRE, where $E_{0,n}$ represent the TRE between fixed image frame 0 and all other frames which considered as a moving image.



(a)



(b)

Fig. 5. The performance of optimizers in (a) average TRE in pixel, and (b) average registered time in minutes for all time frames from 0 to 9.

Figure 5 illustrated that in the first order optimizers, i.e., AdaGrad, AdamP, AdamaX, and RangerQH, outperformed the traditional optimizer SGD by 56.9%, and LBFGS by 7.29% in terms of the average TRE for all time frames. By comparing AdaGrad, AdamP, AdamaX, and RangerQH in terms of the average registered time between all time frames, it was found that the AdaGrad optimizer outperformed the other optimizers with reducing the required registration average time by 16.56% than AdamP, and by 48.37% using LBFGS. Table 3 reported the comparison between the AdaGrad with the other optimizers in terms of the average of NCC loss and regularization loss between a fixed time frame 0 and all other moving time frames.

From the comparisons in Tables 2 and 3, the AdaGrad optimizer gives the best results where the average TRE is 1.893 which is the lowest among the other optimizers and the registered output time frame image is taken around 50.7 minutes.

C. Comparative study of proposed AdaGrad-based B-spline registration with State-of-the-Art

Our experimental study established that the AdaGrad optimizer provided superior results compared to the evaluated first order optimizers and the LBFGS second order optimizer to register thorax images. In addition, a comparison with state-of-the-art studies is reported in Table 4, showing that the proposed AdaGrad-based registration provided the minimum TRE in minimum time with highest correlated values compared to the results in [30], and [10].

Table 4 showed that the AdaGrad optimizer achieved superior performance compared to the adaptive stochastic gradient descent (ASGD) optimizer [10]and [30]. The results showed that our method outperforms [10] and [30] by introducing reduction in the average TRE errors by 5.5%, and 55.5%, respectively.

Table 3. The comparative between AdaGrad, AdamP, AdamaX, RangerQH, SGD, and LBFGS optimizers in evaluation metrics.

	AdaGrad	AdamP	AdamaX	RangerQH	SGD	LBFGS
$\bar{r}_{NCC}$	0.99678	0.9969	0.99687	0.99687	0.9907	0.9936
$\bar{L}_2 loss$	3.71E-05	7.72E-05	5.83E-05	5.57E-05	8.50E-07	2.38E-05

Table 4. Comparison between the proposed method and state-of-the-art studies.

	Before Reg.	3D Reg.[30]	4D Reg.[30]	4^{th}iter.[10]	Ours Proposed Reg.
$\bar{E}(\bar{\sigma})$	3.3(1.8)	2.7(2.1)	2.8(2.2)	1.9(1.6)	1.8(1.0)

IV. Conclusion

Image registration has several applications in medical image analysis. It is a fundamental preprocessing step where two or more images are matched to a common coordinate system. Out of various types of registration methods, we used a mono-model nonrigid registration method, which uses Gaussian image pyramid content to use in the B-spline transformation step for overlaying the input images. The proposed model is tested on the POPI dataset. In this paper we used some of first order optimizers like SGD, AdaGrad, AdaMaX, AdamP and RangerQH. and from the result analysis it is found that the AdaGrad optimizer outperformed the other first order optimizer in overall metric and when compared it with the LBFGS as a second order optimizer, it is found that the AdaGrad optimizer Outperformed to the LBFGS optimizer by 7.29% in TRE and the registered time to produce an output image is taken around 50.7 minutes which is less than the time that the LBFGS is taken by 48.37%. The future scope of work would be, this model will applied in multi-modal images with applying deep learning approaches for feature detection and matching.

References

[1] G. Song, J. Han, Y. Zhao, Z. Wang, and H. J. C. M. I. Du, "A review on medical image registration as an optimization problem," *Current Medical Imaging,* vol. 13, no. 3, pp. 274-283, 2017.

[2] F. L. J. I. T. o. p. a. Bookstein and m. intelligence, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 11, no. 6, pp. 567-585, 1989.

[3] P. Thévenaz, T. Blu, and M. J. I. T. o. m. i. Unser, "Interpolation revisited [medical images application]," *IEEE Transactions on Medical Imaging,* vol. 19, no. 7, pp. 739-758, 2000.

[4] R.-M. Menchón-Lara *et al.*, "Fast 4D elastic group-wise image registration. Convolutional interpolation revisited," *IEEE Transactions on Medical Imaging,* vol. 200, p. 105812, 2021.

[5] S. Klein, M. Staring, and J. P. J. I. t. o. i. p. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *J IEEE transactions on image processing,* vol. 16, no. 12, pp. 2879-2890, 2007.

[6] Y. R. Rao, N. Prathapani, E. J. I. J. o. R. i. E. Nagabhooshanam, and Technology, "Application of normalized cross correlation to image registration," *International Journal of Research in Engineering Technology,* vol. 3, no. 5, pp. 12-16, 2014.

[7] J. A. Maintz and M. A. J. M. i. a. Viergever, "A survey of medical image registration," *Medical image analysis,* vol. 2, no. 1, pp. 1-36, 1998.

[8] V. Mani, S. J. J. o. B. E. Arivazhagan, and Technology, "Survey of medical image registration," *Journal of Biomedical EngineeringTechnology,* vol. 1, no. 2, pp. 8-25, 2013.

[9] S. Sun, Z. Cao, H. Zhu, and J. J. I. t. o. c. Zhao, "A survey of optimization methods from a machine learning perspective," *IEEE transactions on cybernetics,* vol. 50, no. 8, pp. 3668-3681, 2019.

[10] H. Xu and X. Li, "Consistent feature-aligned 4d image registration for respiratory motion modeling," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, 2013, pp. 584-587: IEEE.

[11] H. Song, J.-j. Li, S.-l. Wang, and J.-t. J. J. o. C. S. U. Ma, "Multi-modality liver image registration based on multilevel B-splines free-form deformation and L-BFGS optimal algorithm," *Journal of Central South University,* vol. 21, no. 1, pp. 287-292, 2014.

[12] H. Robbins and S. J. T. a. o. m. s. Monro, "A stochastic approximation method," *The annals of mathematical statistics,* pp. 400-407, 1951.

[13] J. Duchi, E. Hazan, and Y. J. J. o. m. l. r. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of machine learning research,* vol. 12, no. 7, 2011.

[14] D. P. Kingma and J. J. a. p. a. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:,* 2014.

[15] B. Heo, Chun, Sanghyuk,Oh, Seong Joon,Han, Dongyoon,Yun, Sangdoo,Kim, Gyuwan,Uh, Youngjung,and Ha, Jung-Woo, "AdamP: Slowing down the slowdown for momentum optimizers on scale-invariant weights," in *Proceedings of the International Conference on Learning Representations (ICLR), Online*, 2021, pp. 3-7.

[16] J. Vandemeulebroucke, D. Sarrut, and P. Clarysse, "The POPI-model, a point-validated pixel-based breathing thorax model," in *XVth international conference on the use of computers in radiation therapy (ICCR),* 2007, vol. 2, pp. 195-199.

[17] S. Patro and K. K. J. a. p. a. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:.06462,* 2015.

[18] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. J. R. e. Ogden, "Pyramid methods in image processing," *RCA engineer,* vol. 29, no. 6, pp. 33-41, 1984.

[19] A. Shocher, Gandelsman, Yossi,Mosseri, Inbar,Yarom, Michal,Irani, Michal,Freeman, William T, and Dekel, Tali, "Semantic pyramid for image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7457-7466.

[20] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. J. I. t. o. m. i. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE transactions on medical imaging,* vol. 18, no. 8, pp. 712-721, 1999.

[21] Z. Xie, G. E. J. I. T. o. v. Farin, and c. graphics, "Image registration using hierarchical B-splines," *IEEE Transactions on visualization*, vol. 10, no. 1, pp. 85-94, 2004.

[22] W. Zufeng, L. Tian, W. Jiang, D. Yi, Q. J. I. J. o. S.-.-S. Zhiguang, Science, and Technology, "Medical Image Registration Using B-Spline Transform," *International Journal of Simulation--Systems, Science Technology*, vol. 17, no. 48, 2016.

[23] S. i. Kaneko, Y. Satoh, and S. J. P. R. Igarashi, "Using selective correlation coefficient for robust image registration," *Pattern Recognition*, vol. 36, no. 5, pp. 1165-1173, 2003.

[24] D. F. Pace, S. R. Aylward, and M. J. I. t. o. m. i. Niethammer, "A locally adaptive regularization based on anisotropic diffusion for deformable image registration of sliding organs," *IEEE transactions on medical imaging*, vol. 32, no. 11, pp. 2114-2126, 2013.

[25] R. Sandkühler, C. Jud, S. Andermatt, and P. C. J. a. p. a. Cattin, "AirLab: autograd image registration laboratory," *arXiv preprint arXiv:.09907*, 2018.

[26] I. Kandel, M. Castelli, and A. J. J. o. I. Popovič, "Comparative Study of First Order Optimizers for Image Classification Using Convolutional Neural Networks on Histopathology Images," *Journal of Imaging*, vol. 6, no. 9, p. 92, 2020.

[27] S. J. a. p. a. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:.04747*, 2016.

[28] H. O. Aggrawal and J. J. a. p. a. Modersitzki, "Hessian Initialization Strategies for L-BFGS Solving Non-linear Inverse Problems," *arXiv preprint arXiv:.10010*, 2021.

[29] M. Li, M. Soltanolkotabi, and S. Oymak, "Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2020, pp. 4313-4324: PMLR.

[30] C. T. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. J. M. i. a. Niessen, "Nonrigid registration of dynamic medical imaging data using nD+ t B-splines and a groupwise optimization approach," *Medical image analysis*, vol. 15, no. 2, pp. 238-249, 2011.