Unsupervised image registration based on Residual-connected DRMINE for diagnostic metastatic bone tumors

Shogo Baba[1], Tohru Kamiya[1], Takashi Terasawa[2], Takatoshi Aoki[2]
[1] *Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, Fukuoka 804-8550, Japan*
[2] *University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyushu, Fukuoka, 807-8555, Japan*
Email: baba.shogo274@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

Abstract

Metastatic tumors are frequently identified through follow-up surveillance using computed tomography (CT) scans. However, CT scans produce more than 100 images in an examination, which imposes a significant burden on radiologists and entails a potential risk of misdiagnosis. Temporal subtraction is utilized in Computer-Aided Diagnosis (CAD) and proves to be an effective technique in aiding image interpretation process for the radiologists. In this study, we focus on the preliminary stage of CAD development specialized in bone metastasis extraction, with a particular emphasis on rigid registration. We propose a novel rigid registration technique by augmenting DRMINE, which estimates mutual information using neural networks, with skip connections and normalization. From the three datasets, ten images were selected randomly from the cervical, thoracic, and lumbar regions. These images were then augmented with rotation as well as horizontal and vertical translations to create modified versions. The registration accuracy was assessed based on the Full Width at Half Maximum (FWHM) of the difference images. In the proposed method, FWHM values for the thoracic and lumbar regions of the spine exhibited a maximum reduction rate of 2.8% and a minimum reduction rate of 0.533%. However, the cervical spine region exhibited superior FWHM results with DRMINE compared to the proposed methodology. The proposed method was influenced by the capture area, but it indicated the potential to provide stable registration as the standard deviation decreased for all FWHM values.

Keywords: Computed Tomography, Rigid registration, Unsupervised learning, MINE, DRMINE

1. Introduction

As per the estimates of the World Health Organization (WHO) in 2019, cancer is the first or second leading cause of death in 112 of 183 countries [1]. Furthermore, there exists the possibility of cancer metastasizing to distant locations through the vascular and lymphatic systems, with the spine, in particular, being the most susceptible site for skeletal metastasis [2], [3]. Metastatic bone tumors necessitate prompt identification and timely intervention. However, as metastatic bone tumors commonly lack discernible symptoms, they are often detected through Computed Tomography(CT) examinations during follow-up observations. Therefore, the assessment of osseous structures harboring metastases and the early detection thereof predominantly rely upon CT scanning as the most efficacious modality. However, in CT examinations, where 100 to several hundred cross-sectional images can be obtained in an examination, the burden on the diagnosing radiologist increases significantly, and the results vary due to differences in interpretation experience, which poses a significant challenge.

The Temporal Subtraction technique (TS) [4] is a method that involves performing calculations between two images taken at different times to extract lesions that appeared during a specific period. This technique is applied in computer-aided diagnosis (CAD) systems for supporting radiographic interpretation. However, the mere subtraction of a simple current image and a past image results in the occurrence of differential artifacts due to positional discrepancies between the images. Therefore, it becomes crucial to perform registration processing as a preprocessing step. In typical image registration frameworks, prior to performing non-rigid registration, the target image pairs are often pre-aligned based on rigid registration or affine transformations. However, unless there is a significant improvement in the accuracy of affine deformations, the performance of alignment will be greatly diminished [5], [6].

In recent years, numerous studies have been conducted on registration techniques, particularly in the domain of deep learning. However, deep learning requires a substantial amount of annotated data for training, which hampers the practical application of deep learning methods in medical research. Consequently, unsupervised learning, which does not rely on annotated data, has garnered significant attention as a potential solution to this limitation. Differentiable registration with mutual information and matrix exponential (DRMINE) [7] is an unsupervised learning-based rigid registration model that utilizes Mutual Information Neural Estimation (MINE) [8] to estimate mutual information using a neural network, instead of relying on conventional approximations of mutual information.

In this study, we focused on rigid registration and proposed a model incorporated with skip connections and normalization in the MINEnet framework within the DRMINE model.

2. Materials and methods

2.1. Dataset

We conducted experiments employing TOSHIBA Aquilion PRIME and TOSHIBA Aquilion ONE, utilizing data from three datasets encompassing imaging ranges spanning from the cervical to lumbar spine levels. The image size was set at 512×512, with a slice thickness of 1.0 mm and pixel size ranging from 0.625 to 0.820 mm. We conducted experiments by randomly selecting ten images each from the cervical, thoracic, and lumbar regions based on the data of these three datasets cases.

Furthermore, in order to replicate the displacement in positioning, an image was created with horizontal and vertical movements of 15 pixels each along the x-axis and y-axis, along with a 10° rotation (Bicubic interpolation). Fig. 1 illustrates the image transformation, depicting the image before and after the conversion.

Our computing system consisted of an Intel Core i7-7500U CPU and an NVIDIA Tesla V100 GPU with 16GB of memory.

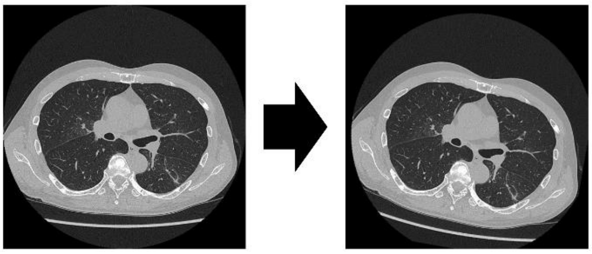


Fig. 1 The image transformation
Left: Original image, Right: transformed image

2.2. DRMINE registration algorithm

MINE utilizes the principle of Donsker-Varadhan (DV) duality to compute mutual information (MI),

$$MI = \sup_{f} J(f) \quad (1)$$

where $J(f)$ is the DV lower bound.

$$J(f) = \int f(x,z)P_{XZ}(x,z)dxdz - \log\left(\int exp\big(f(x,z)\big)P_X(x)P_Z(z)dxdz\right) \quad (2)$$

MINE uses a neural network to compute $f(x,z)$ and uses Monte Carlo technique to approximate the right hand side of (2). MINE claims that computation of (1) scales much better than histogram-based computation of MI.

The optimization of DRMINE is as follows.

$$\max_{\substack{v_1,\cdots,v_6 \\ v_1^1,\cdots v_6^6 \\ \theta}} \{\sum_{l=2}^{L} MINE(T_l, Warp(M_l, M_{exp}(\sum_{i=1}^{6} v_i B_i))) + MINE(T_1, Warp(M_1, M_{exp}(\sum_{i=1}^{6}(v_i + v_i^1)B_i)))\} \quad (3)$$

B_i is an affine transformation on a two-dimensional plane. This group comprises six generators, where $Mexp$ denotes the matrix exponential operation.

$$B_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} B_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} B_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} B_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} B_6 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using the multi-resolution pyramid technique [9], [10], [11], the fix image pyramid :T_l and the moving image pyramid :M_l （$l = 1, \ldots, L$） . L is the maximum level in the image pyramid.

MINE is a method for estimating mutual information utilizing neural network architectures. MINE learns using the error backpropagation method, similar to conventional neural networks, in order to maximize the mutual information between the fixed image and the distorted moving image. θ denotes the parameters of the neural network utilized to accomplish MINE. MINEnet is constructed with two hidden layers composed of 100 neurons each, and utilizes the non-linear activation function ReLu between the hidden layers.

However, note also that image structures are slightly shifted through multi-resolution image pyramids. So, a transformation matrix suitable for a coarse resolution may need a slight correction when used for a finer resolution. To alleviate this issue, in the second item, we employ the parameterization of matrix exponential functions and introduce a dedicated additional parameter vector $v^1 = [v_1^1, \cdots, v_6^1]$ for the finest level of resolution, optimizing the multiresolution approach. Fig. 2 illustrates DRMINE registration algorithm.

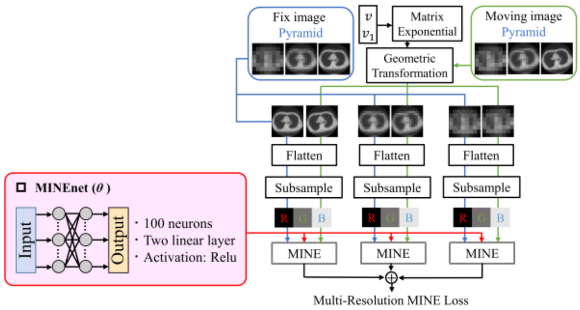


Fig. 2 DRMINE algorithm

2.3. Our method

In order to enhance the registration technique utilizing DRMINE, it is of paramount importance to precisely estimate mutual information. MINEnet is a method that utilizes neural networks for the estimate mutual information, and thus, we postulated that the effective propagation of differentials plays a pivotal role in the estimate mutual information. The conventional approach comprises a simplistic two-layer structure, whereas the proposed methodology constructs a three-layer architecture by incorporating additional layers into MINEnet. Furthermore, we introduced skip connections

[12] between each layer. Additionally, we believed that the normalization of the hidden layers also influences the precision of the neural network. Hence, we incorporated layer normalization [13] into the model, as depicted in Fig. 3. Moreover, we set the number of neurons in each layer to 100, similar to DRMINE.

To demonstrate the enhancement in accuracy attributed to the incorporation of normalization, this study conducted a comparative analysis of precision among three methodologies: the conventional approach, a model enriched solely with skip connections in the DRMINE framework (Method [a]), and the proposed technique (Method [b]).

2.4. Evaluation methods

The evaluation of image registration accuracy involved the computation of the Full Width at Half Maximum (FWHM) from the histogram of the difference image. This difference image was aligned using a reference image and registered with the moving image (Fig. 4). FWHM represents the width at which the distribution encompasses half of its maximum value. A smaller FWHM indicates a reduction in the presence of artifacts attributed to residuals.

3. Experiments and results

3.1. Results

From Table 1, the method[b] yielded the lowest average FWHM for both thoracic and lumbar regions of all data of datasets. The mean reduction rate was found to be 1.389%, with a maximum reduction rate of 2.8% and a minimum reduction rate of 0.294%. However, in the cervical spine region, DRMINE yielded superior precision results across all data of datasets. The model constructed solely with skip connections without normalization did not show a clear improvement in accuracy compared to DRMINE. Therefore, the enhancement of accuracy cannot be achieved solely through the addition of skip connections. By combining Layer Normalization, we have achieved an enhancement in the accuracy of estimations.

Fig. 5, Fig. 6, Fig. 7 depict the results obtained from algorithms applied to the cervical, thoracic, and lumbar regions. The proposed methodology led to the observable reduction of residuals in the contour region.

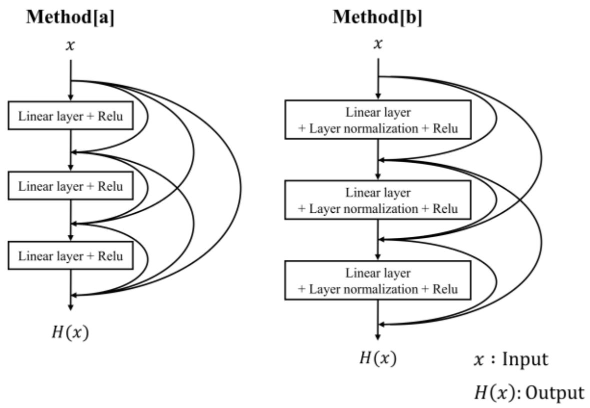


Fig. 3 Proposed network

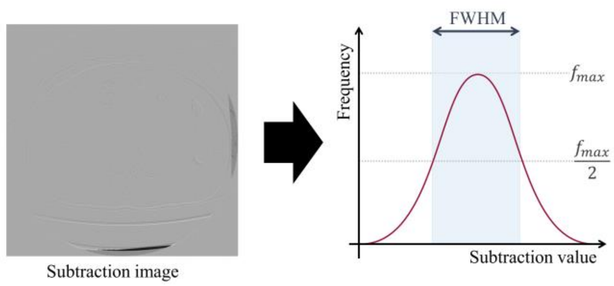


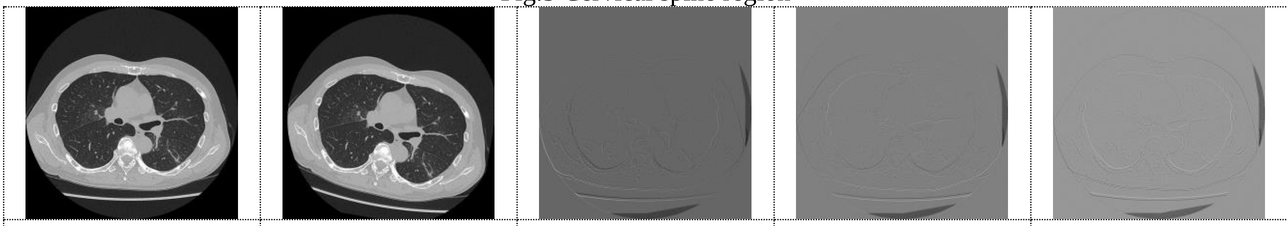
Fig. 4 FWHM (Subtract image histogram)

Table1. Average FWHM and Rate of increase or decrease
Rate[a]: The comparison between DRMINE and Method[a], Rate[b]: The comparison between DRMINE and Method[b]
▲: Rate of increase , ▼: Rate of decrease

Datasets	Level	DRMINE	Method[a]	Method[b]	Rate[a] (%)	Rate[b] (%)
Data[A]	C-spine	1.2686±0.0182	1.2805±0.0177	1.2712±0.0151	0.928▲	0.200▲
	T-spine	1.2251±0.0249	1.2065±0.0196	1.1918±0.0140	1.543▼	2.800▼
	L-spine	1.1812±0.0237	1.1794±0.0194	1.1633±0.0136	0.158▼	1.546▼
Data[B]	C-spine	1.3628±0.0224	1.3724±0.0193	1.3667±0.0116	0.699▲	0.286▲
	T-spine	1.3955±0.0197	1.3984±0.0111	1.3819±0.0127	0.254▼	0.981▼
	L-spine	1.3187±0.0252	1.3153±0.0279	1.2958±0.0115	0.207▲	1.764▼
Data[C]	C-spine	1.2645±0.0180	1.2688±0.0200	1.2693±0.0166	0.344▲	0.381▲
	T-spine	1.3194±0.0147	1.3140±0.0121	1.3101±0.0044	0.026▲	0.712▼
	L-spine	1.2680±0.0106	1.2684±0.0154	1.2613±0.0067	0.417▼	0.533▼

Fix image | Moving image | DRMINE | Method[a] | Method[b]

Fig.5 Cervical spine region



Fix image | Moving image | DRMINE | Method[a] | Method[b]

Fig. 6 Thoracic spine region

Fix image | Moving image | DRMINE | Method[a] | Method[b]

Fig. 7 Lumbar spine region

4. Discussion and conclusion

From Table 1, it can be inferred that the contour's shape exerts an influence on the FWHM. However, it is posited that the stratification of the proposed methodology was rather shallow, resulting in an insufficient learning efficacy. Therefore, we contemplate that further deepening the layers and establishing interconnections between them through skip connections may lead to an improvement in precision. Furthermore, the mere addition of skip connections did not exert any influence on the enhancement of accuracy and stabilization of estimation precision. This demonstrates the contribution of adding layer normalization not only to the improvement of accuracy but also to the stabilization of estimate mutual information.

The proposed methodology has demonstrated further enhancement and stabilization of precision in rigid registration. Additionally, it has suggested its utility in rigid registration of medical images from an unsupervised learning perspective.

References

1. OECD, Health at a Glance 2021: OECD indicators, pp.143, OECD publishing, 2022.
2. David Stegner et al., "Mechanistic Explanation for Platelet Contribution to Cancer Metastasis", Thrombosis Research, Vol.133, No.52, pp.149-157, 2014.
3. R. Guillevin et al., "Spine Metastasis Imaging: Review of the Literature", Journal of Neuroradiology, Vol. 34, No.5, pp.311-321, 2007.
4. Kano A. et al., "Digital image subtraction of temporally sequential chest images for detection of interval change", Medical Physics, Vol.21, No.3, pp.453-461, 1994.
5. Maintz et al., "A survey of medical image registration. Medical image analysis", Vol.2, No.1, pp.1-36, 1998.
6. Pluim et al., "Mutual-information-based registration of medical images: a survey", IEEE Transactions on Medical Imaging, Vol.22, No.8, pp.986-1004. 2003.
7. Nan A. et al., "Drmime: Differentiable mutual information and matrix exponential for multi-resolution image registration", In: Medical Imaging with Deep Learning. PMLR, pp.527-543, 2020.
8. Belghazi et al., "Mutual information neural estimation., In: International conference on machine learning", PMLR, pp.531-540. 2018.
9. P. Thevenaz, et al., "A pyramid approach to subpixel registration based on intensity", IEEE Transactions on Image Processing, Vol.7, No.1, pp.27-41, 1998.
10. S. Kruger et al., "Image registration using multiresolution frequency domain correlation", in BMVC, pp.1-10, 1998.
11. H. S. Alhichri et al., "Multi-resolution image registration using multi-class Hausdorff fraction", Pattern Recognition Letters, Vol.23, No.1-3, pp.279-286, 2002.
12. He K. et al., "Deep residual learning for image recognition", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.
13. Ba Jimmy Lei et al., "Layer normalization", arXiv:1607. 06450, 2016.

Authors Introduction

Mr. Shogo Baba

He received his Master degree in Medical Science from Kagoshima University Japan in 2021. He is a Ph.D. student at Kyushu Institute of Technology Japan.

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Masters and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis.

MD. Takashi Terasawa

He received his MD degree from university of occupational and environmental health japan in 2011 and PhD in medicine in 2020.
He is a radiologist at the university.

MD. Takatoshi Aoki

He is a professor of Radiology at University of Occupational & Environmental Health in Japan, and the vice president of Japanese Musculoskeletal Society. His clinical and research interests include the imaging modalities of musculoskeletal pathology, especially of tumor, arthritis, and bone metabolism.