

Meta-Learning Guided Pelvis MR to CT Translation: Addressing Cross-Modality Misalignments

Daniel Kim¹, Jae-Hun Lee¹, Yoseob Han², Kanghyun Ryu³, and Dong-Hyun Kim¹

¹Yonsei University, Seoul, Korea, Republic of, ²Soongsil University, Seoul, Korea, Republic of, ³Korea Institute of Science and Technology, Seoul, Korea, Republic of

Synopsis

Motivation: In radiation therapy planning, both MR and CT are essential, but there is a potential risk of radiation exposure from CT. To address this problem, MR to CT translation could be an important solution.

Goal(s): In cross-modality translations like MR to CT, misalignment is significant challenge. The goal is to develop a method that can effectively learn to handle this misalignment.

Approach: We propose a method that utilizes meta-learning to focus on reliable regions and employs loss functions and network suited for misalignment.

Results: Our method surpassed existing GAN-based methods in quantitative evaluations, particularly in the reconstruction of bone structures.

Impact: It can be seen that meta-learning can be effectively applied to the problem of misalignment. This can aid in preserving fine details and bone structures in MR to CT translation. It is also broadly applicable to cross-modality translation.

Introduction

MR and CT imaging are widely used in radiation therapy planning¹. Although obtaining both types of data is part of the standard clinical workflow, it can be time-consuming and costly, with the additional concern of potential harm from CT radiation exposure. To reduce these problems, MR to CT translation can be an important solution for MR-only treatment planning². In cross-modality tasks like MR to CT translation, registration plays a critical role, especially in the pelvis where abdominal movement due to breathing can cause serious misalignment³. To address this challenge, two types of registration methods are utilized: linear and nonlinear. Nonlinear methods (e.g., B-spline, SyN, and Deformable) are proactive in registration but often lead to blurring because of interpolation⁴. In contrast, linear methods (e.g., rigid, affine) may have lower registration performance but almost never cause blurring. We present a learning approach that is more robust under the conditions of linear registration, which is typically more prone to misalignment.

Methods

1. Meta-learning-based Re-weighting⁵

Meta-learning reweighting addresses misalignment between source and target by focusing on reliable instances. This process starts by selecting a well-registered dataset as the validation set, then updates the weight map for each training set to enhance translation accuracy on the validation set. We define (x, y) as the source-target pair and $\{(x_i, y_i), 1 \leq i \leq N\}$ as the

training set. A well-registered validation set is assumed to be $\{(x_i^v, y_i^v), 1 \leq i \leq M\}$, and $M \ll N$. And define (n, m) as the count of mini-batches for training and validation, respectively.

$$\hat{\theta}_{t+1} = \theta_t - \alpha \nabla \sum_{i=1}^n \epsilon_i \|y_i - f(x_i; \theta_t)\|_p$$

The auxiliary variable ϵ is utilized to update the weights for the corresponding mini-batch within the training dataset.

$$u_{i,t} = -\eta \frac{\partial}{\partial \epsilon_{i,t}} \frac{1}{m} \sum_{j=1}^m \|y_j^v - f(x_j^v; \theta_{t+1}(\epsilon))\|_p \Big|_{\epsilon_{i,t}=0}$$

The gradient of $\epsilon_{i,t}$ is estimated through the loss of a single validation mini-batch, as computing across the entire validation data would be computationally inefficient. The learning step is indicated by t . The term $u_{i,t}$ reflects the direction of the update for $\epsilon_{i,t}$ in a single gradient descent step.

$$\tilde{w}_{i,t} = \max(u_{i,t}, 0)$$

$$w_{i,t} = \frac{\tilde{w}_{i,t}}{\sum_j \tilde{w}_{j,t}}$$

Weight updates are only performed when, $u_{i,t}$ is positive, and normalization is achieved by the total weight sum of its mini-batch. Ultimately, $w_{i,t}$ represents the spatial weight map for the training data, and by incorporating it into Equation 1 in place of ϵ_i , the network is guided to focus on the more reliable regions.

2. Contextual Loss⁶

The Contextual Loss assesses the similarity of images x and y by computing the average of the closest feature matches across their deep feature maps, which is particularly beneficial in situations of misalignment.

$$\mathcal{L}_{CTX}(x, y) = \frac{1}{N} \sum_j \min_i \mathbb{D}(\Phi(x)_j, \Phi(y)_i)$$

The symbol $\Phi(\cdot)$ denotes the feature maps from a pre-trained VGG-19 network, specifically from specific layers. Here, \mathbb{D} is the cosine similarity measuring the minimum distance between feature j of image x and all features of image y .

3. Implementation detail

We adopted the generator and discriminator architecture from PGAN⁷, training our model on single-slice images cropped to (320,192). Optimization was performed over 100 epochs using a batch size of 6 for training and 32 for meta-validation.

Datasets

We utilized the SynthRAD2023 dataset from the Grand Challenge, comprising pelvis MR and CT pairs. Data was collected from three different centers. The MRI protocol was a T1w imaging. For meta-learning, the top 11% of cases with the highest mutual information between MR and CT images were allocated for validation. All data were provided post-rigid registration, and no further registration was performed.

Results

We compared our approach with existing GAN-based supervised learning techniques (PGAN[7], REG-PGAN[7,8], CTX-PGAN[6,7]), all of which are tasked with translating MR to CT on the Pelvis. Fig 3 contains a quantitative evaluation and a comparison of the result images for each model. The proposed method showed better results for finer details and bones. Fig 4 presents the result of visualizing the bone structure by 3D rendering, which is done by stacking the generated 2D CT data on the slice dimension. The proposed method shows exceptional results in preserving the bone structure.

Discussion and Conclusion

We propose a method to address the issue of misalignment in Cross-Modality Translation. Our method demonstrated superior performance in translating bone structures and clearly defined features compared to existing methods. However, there are limitations for areas where features are not distinct, such as in internal organ regions, and where misalignment is particularly severe. Minimizing blur with non-linear registration and network refinement will be the direction of future research.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1A4A1030579).

References

1. Khoo, V. S., Adams, E. J., Saran, F., Bedford, J. L., Perks, J. R., Warrington, A. P., & Brada, M. (2000). A comparison of clinical target volumes determined by CT and MRI for the radiotherapy planning of base of skull meningiomas. *International Journal of Radiation Oncology* Biology* Physics*, 46(5), 1309-1317.
2. Schmidt, M. A., & Payne, G. S. (2015). Radiotherapy planning using MRI. *Physics in Medicine & Biology*, 60(22), R323.
3. Weckesser, M., Stegger, L., Juergens, K. U., Wormanns, D., Heindel, W., & Schober, O. (2006). Correlation between respiration-induced thoracic expansion and a shift of central structures. *European radiology*, 16, 1614-1620.
4. Su, Y., Dai, X., He, L., & Kong, X. (2022, November). ABN: Anti-Blur Neural Networks for Multi-Stage Deformable Image Registration. In *2022 IEEE International Conference on Data Mining (ICDM)* (pp. 468-477). IEEE.
5. Ren, M., Zeng, W., Yang, B., & Urtaasun, R. (2018, July). Learning to reweight examples for robust deep learning. In *International conference on machine learning* (pp. 4334-4343). PMLR.
6. Mechrez, R., Talmi, I., & Zelnik-Manor, L. (2018). The contextual loss for image transformation with non-aligned data. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 768-783).
7. Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., & Cukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE transactions on medical imaging*, 38(10), 2375-2388.
8. Kong, L., Lian, C., Huang, D., Hu, Y., & Zhou, Q. (2021). Breaking the dilemma of medical image-to-image translation. *Advances in Neural Information Processing Systems*, 34, 1964-1978.

Figures

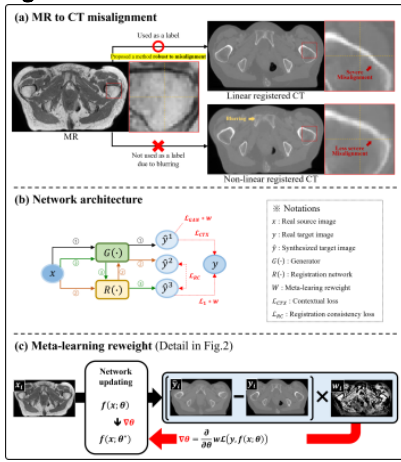


Figure 1. (a) Definition of the problem of MR to CT misalignment and the setting of labeling. (b) An illustrative overview of the network architecture. (c) A simplified visualization of the meta-learning reweighting process.

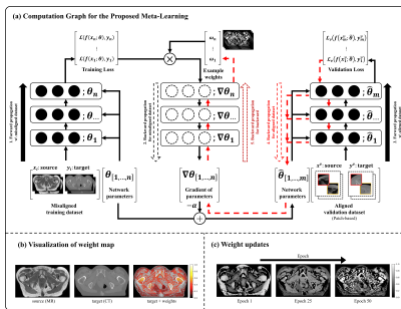


Figure 2. (a) This graph depicts the computation process where weights for each mini-batch are determined and applied to update the network parameters. (b) This shows a color map overlay on the target data (CT) for a specific epoch, where regions with weights closer to 1 are more heavily weighted in the loss calculation. (c) A sequence of images illustrating how the weights change over epochs, with a focus on progressively emphasizing the more reliable regions of the data.

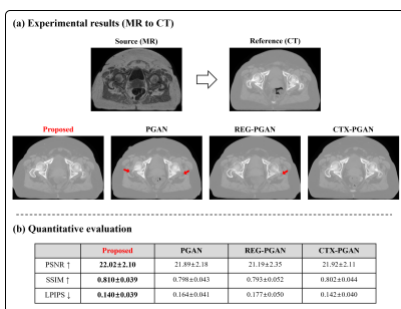


Figure 3. (a) A comparison of the experimental outputs from each model. The red arrows indicate the locations where the bone areas have not been properly translated. (b) The performance of each model is quantitatively compared using the PSNR, SSIM, and LPIPS metrics.

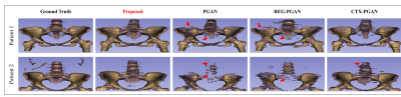


Figure 4. Results of 3D rendering after stacking synthesized 2D CT slices in the slice dimension. A threshold was set during the rendering process to visualize the bone structure. The red arrows indicate areas where the bone synthesis failed during the MR to CT translation.