Image registration of chest CT volumes: 4DCT DIR-Lab Challenge

Medical Image Registration and Applications

1st Xavier Beltran Urbano
EMJMD in Medical Imaging and Applications
University of Girona
Girona, Spain
xavibeltranurbano00@gmail.com

1st Frederik Hartmann
EMJMD in Medical Imaging and Applications
University of Girona
Girona, Spain
frederik.hartmann@gmx.de

Abstract—This paper presents a comprehensive study on image registration of chest CT volumes, focusing on aligning thoracic structures across different respiratory phases. We compare traditional techniques like itk-elastix with modern deep learning approaches such as VoxelMorph, using a dataset of 4 thoracic 4DCT images from various respiratory-binned phases. The study employs multiple preprocessing steps, including segmentation, normalization, and histogram equalization, to enhance registration accuracy. Our analysis uses the target registration error (TRE) as the primary evaluation metric. The results show itk-elastix, especially with custom parameter tuning and preprocessing, achieves a TRE of 2.63 ± 1.43 mm, significantly outperforming VoxelMorph, which yielded a TRE of 28.92 ± 10.53 mm. On the day of the challenge, the achieved TRE was 2.00 ± 0.76 mm using elastix with the mentioned approach. These findings highlight the effectiveness of traditional techniques with optimized parameters and preprocessing for chest CT volume registration.

Keywords—Chest CT Volumes, Image Registration, Affine, TRE, Voxelmorph, Elastix

I. INTRODUCTION

Image registration is an essential task in medical imaging as it enables different images to be aligned into a shared spatial frame of reference. This procedure is crucial for a range of applications, including comparing patient scans over time, combining information from diverse imaging modalities, and mapping individual images to standardized anatomical atlases. Through techniques such as rigid, affine, or non-rigid transformations, image registration adjusts for translational, rotational, and scaling discrepancies between images. In the field of thoracic imaging, the registration of chest CT volumes is an especially difficult yet crucial task. Due to the dynamic nature of thoracic structures, including the diaphragm, heart, and lungs, which can change shape and position between scans as a result of breathing and other bodily movements, this procedure requires the precise alignment of multiple CT scans of the patient's thorax. In order to control these variations and guarantee precise alignment, sophisticated registration methods are utilized. Preciseness of this kind is of the greatest importance in various domains, like tracking the progression of lung diseases, planning and monitoring treatment for thoracic conditions, and conducting detailed lung analyses. In this study, we have presented a comprehensive examination of various chest CT volume registration methods.

II. DATASET

To implement this project, we have utilized a data set consisting of 4 thoracic 4DCT images acquired at the University of Texas M. D. Anderson Cancer Center in Houston, TX [1]. Each CT image in the dataset corresponds to different respiratory-binned phases, ranging from T00 to T90. The T00 phase represented end-inhalation, while the T50 phase represented end-exhalation. Expert manual annotation was conducted to identify 300 landmarks on each patient's CT images of T00 and T50. In the Figure 1 we can observe an example of the inhalation and exhalation phases of patient 1 with their corresponding landmarks.

III. METHOD

In order to align the inhaling lung with the exhaling lung, multiple preprocessing and registration techniques have been explored. Each of the methods used will be explained in the following sections.

A. PREPROCESSING

Three preprocessing techniques (see Figure 2) have been employed and tested:

1) Segmentation
2) Normalization
3) Histogram Equalization

In this project, two different **segmentation** approaches were tested. The first segmentation approach was developed as a part of a previous project. This relies on classical image processing techniques and takes advantage of the intensity range and position of the lung. The segmentation consisted of three steps. First, a preprocessing step is performed to create a coarse segmentation of the lung and remove the table. Second, the coarse segmentation is refined by tracking the contours from
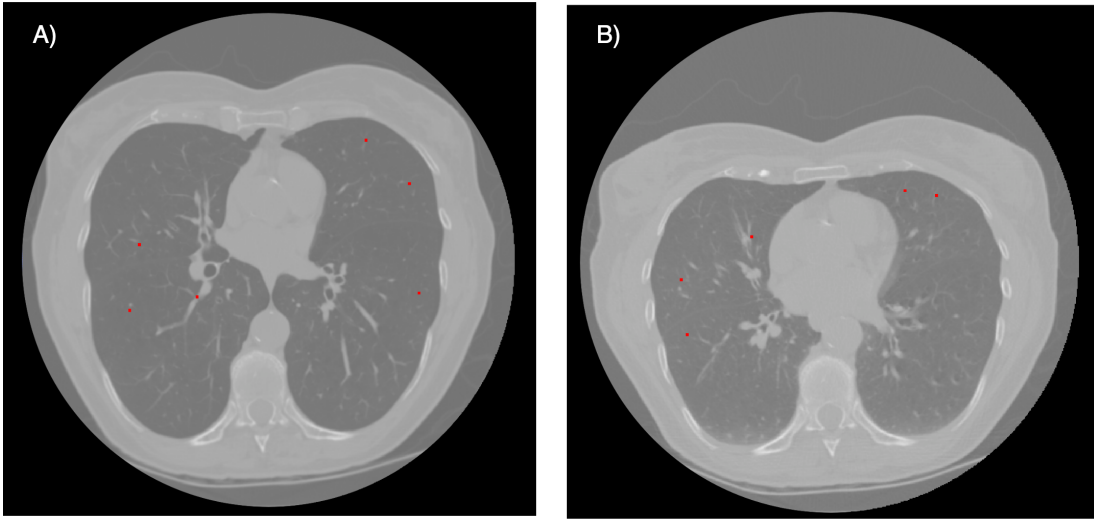
1

Fig. 1. An illustration of the respiratory phases in the COPD1 case. **A)** Inhalation phase with highlighted landmarks (red points). **B)** Exhalation phase, also featuring corresponding landmarks (red points).

the center slice and deciding, based on the overlap, whether they are part of the lung. Third, the remaining holes are closed in the post-processing. This algorithm was tested on the Vessel12 dataset and achieved a dice score (DSC) of 0.98. A detailed explanation can be found here[1].

However, the algorithm had only been tested using unit-spacing, which is not the case for the COPD dataset. There-fore, numerous improvements with regards to the robustness of the algorithm have been made, such as the detection and replacement of wrongly segmented slices in the preprocessing, the improvement of contour detection in the refinement, and the additional detection and removal of artifacts in the post-processing. The second segmentation approach consists of a pretrained model based on UNet architecture [2]. This approach was trained and tested with several datasets, such as VISCERAL Anatomy3 (VISC-36), LTRC (LTRC-36), and LCTSC (LCTSC-36), to perform lung segmentation, achieving a DSC of 0.98 ± 0.03. As the model was trained with CT scans within the Hounsfield Units (HU) range, before using this model, we had to normalize our images to meet this condition. The input image has a shape of $(512, 512, D)$, being D the number of slices of the corresponding image. By leveraging GPU acceleration for the segmentation process, each case was successfully segmented in under 10 seconds.

After a comprehensive qualitative evaluation (see an example of both segmentation approaches in Figure 3), we finally decided to use the approach implemented by ourselves, as it demonstrated better performance than the pretrained UNet model.

Normalization has been performed following the segmen-tation. Here, min-max normalization was chosen. The formula

used can be found in Equation 1.

$$x_{\text{normalized}} = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \quad (1)$$

Finally, **Histogram Equalization** has been used on each axial slice employing contrast-limited adaptive histogram equaliza-tion (CLAHE).

B. REGISTRATION-ELASTIX

One of the methods used for registration was itk-elastix [3]. First, an affine transformation has been carried out to quickly account for scale, shear, rotation, and translation differences. Equation 2 shows the three-dimensional affine transformation.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_x \\ a_{21} & a_{22} & a_{23} & t_y \\ a_{31} & a_{32} & a_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2)$$

In this equation:

- $\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$ is the original point in 3D space.

- $\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix}$ is the transformed point.

- The 4×4 matrix represents the affine transformation, where a_{ij} elements are responsible for rotation, scaling, and shearing, while t_x, t_y, and t_z are the translation components in the x, y, and z directions, respectively.

Second, a bspline registration has been performed. A b-spline registration is able to register non-rigid transformations. This is

---

[1]Accessed on: January 7, 2024. URL: https://frederik-hartmann.github.io/projects/lungSegmentation/
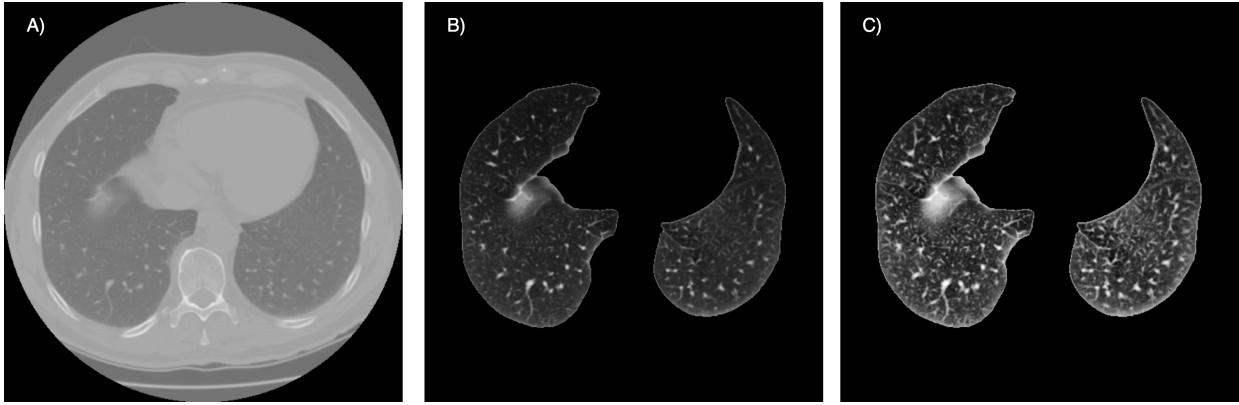
Fig. 2. Sequential overview of preprocessing steps in this approach. **A)** Original image **B)** Segmented Image **C)** Normalized and contrast-enhanced (CLAHE) image.
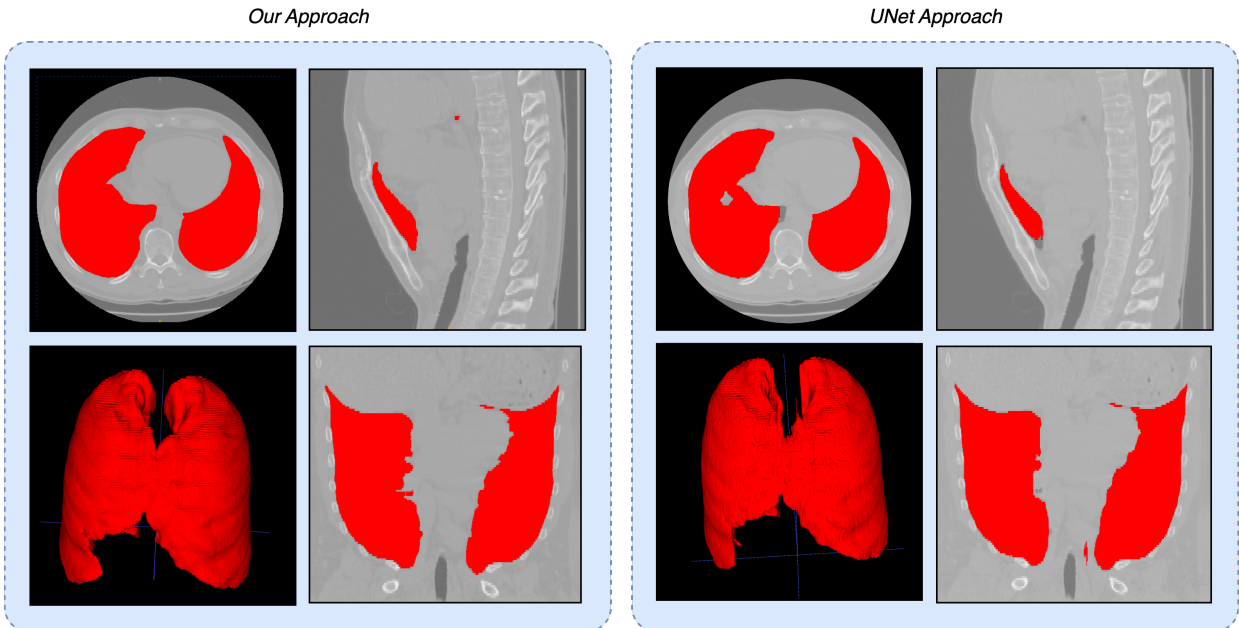

Fig. 3. Comparative visualization of lung segmentation for COPD3 Case. The left panel illustrates the segmentation results using our custom approach, while the right panel showcases the outcomes from the pretrained UNet model [2]. Each panel includes three distinct views - axial, sagittal, and coronal planes - alongside a 3D reconstruction of the segmented lungs.

especially important for this registration case, as the expected transformation is non-rigid. In this challenge, the test sets only consist of inhale landmarks. Due to the fact that itk-elastix uses the forward registration for point sets rather than the normally used inverse registration for images, the fixed image is chosen as the inhale and the moving image as the exhale.

For each of the transformations, multiple parameter sets are available. Here, the publicly available parameter set 11 from modelzoo[2] has been selected. This set of parameters was designed specifically for intra-patient lung CT registration, which is also applicable to this challenge. Additionally, it has been tested for use on slices with spacings of up to

2.5 mm, a characteristic shared by all the COPD images. This parameter set has also been used as a starting point to develop custom parameter files for this specific challenge. More specifically, the number of iterations has been set to a higher value. The most notable changes, however, are in the bspline estimation. The biggest performance improvements were given by changing to a 3D image pyramid, taking into account the differences in voxel spacing. Beyond other changes, the most notable one was the change to randomly sample 25000 voxels to compute the metrics instead of 2000.

C. REGISTRATION-VOXELMORPH

A novel approach based on deep learning, known as VoxelMorph [4], has also been implemented to register chest CT volumes. By employing a U-Net architecture [5], a well-known

---

[2]Accessed on: January 7, 2024. URL: https://github.com/SuperElastix/ElastixModelZoo/tree/master/models/Par0011

CNN model recognized for its expertise in biomedical image segmentation, to discover a correspondence between pairings of input images and a deformation field, this technique permits the unsupervised alignment of medical images. Figure 4 depicts the network used in VoxelMorph, which takes a single input formed by concatenating m (the moving image) and f (the fixed image) into a 2-channel 3D image.
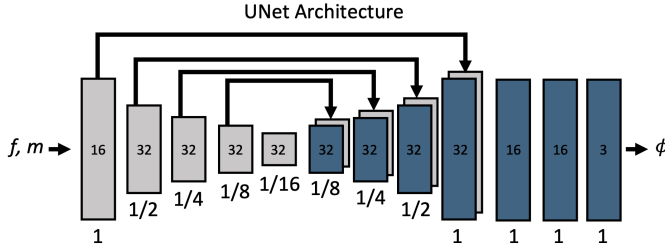

Fig. 4. Convolutional UNet architecture implementing g(f, m). Each rectangle represents a 3D volume, generated from the preceding volume using a 3D convolutional network layer. The spatial resolution of each volume with respect to the input volume is printed underneath. In the decoder, we use several 32-filter convolutions, each followed by an upsampling layer, to bring the volume back to full resolution. Arrows represent skip connections, which concatenate encoder and decoder features. The full-resolution volume is further refined using several convolutions [4].

The core mathematical concept behind this approach lies in the computation of a spatial deformation field ϕ, which is learned to align a moving image m to a fixed image f. The field ϕ is parameterized by a convolutional neural network, and the alignment is formulated as an optimization problem. The objective function is typically a combination of a similarity measure and a regularization term. The similarity measure, $S(f, m \circ \phi)$, ensures the aligned moving image $m \circ \phi$ closely matches the fixed image f, while the regularization term $R(\phi)$ enforces smoothness of the deformation field. The optimization problem is formulated as follows:

$$\phi^* = \operatorname*{argmin}_{\phi} \left[S(f, m \circ \phi) + \lambda R(\phi) \right] \qquad (3)$$

In our implementation, the network architecture is defined by the 'VxmDense' model in TensorFlow, configured with a volume shape of $(256, 256, 128)$ and multiple feature layers. Given the variability in the number of slices per case in the dataset, we added blank slices, with all pixels set to zero, to ensure uniformity across all data inputs for the model. Moreover, all the images have been preprocessed following the approach detailed in subsection III-A. Our loss function combines the Mean Squared Error (MSE) for assessing the similarity and a gradient loss for ensuring the smoothness of the deformation field. We compile the model using the Adam optimizer with a learning rate of $1e - 3$, and it is trained over 100 epochs using a custom data generator and a batch size of 1.

IV. EVALUATION

To evaluate the performance of the described methods, the target registration error (TRE) is used. Equation 4 shows the computation.

$$\text{TRE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \| \mathbf{p}_i' - \mathbf{T}(\mathbf{p}_i) \|^2} \qquad (4)$$

where

- N is the number of target points.
- $\mathbf{p}_i$ is the i-th point in the original space.
- $\mathbf{p}_i'$ is the corresponding point of $\mathbf{p}_i$ in the registered space.
- $\mathbf{T}$ is the transformation function applied for registration.
- The notation $\| \cdot \|$ denotes the Euclidean distance.

For each of the scans (inhale and exhale), 300 points were provided. To explore the effects of each of the preprocessing and registration parameter sets, multiple tests have been carried out. The experiments include:

- no registration
- only the affine transform
- affine + segmentation
- affine + normalization + CLAHE
- affine + segmentation + normalization + CLAHE
- affine + bspline set 1 + segmentation + normalization + CLAHE
- affine + bspline set 1 + bspline set 2 + segmentation + normalization + CLAHE
- custom parameter set affine only
- custom parameter set affine + segmentation
- custom parameter set affine + bspline + segmentation
- voxelmorph

The target registration error of each experiment on the four COPD images is displayed in Table I.

V. DISCUSSION

In the following section, each of the steps undertaken will be discussed critically based on qualitative and quantitative results.

A. COMPARISON OF PREPROCESSING STEPS

Four different preprocessing steps have been tested. As normalization and CLAHE have only been tested in combination, they will be treated as one. Therefore, three different options remain:

1) No preprocessing
2) Segmentation
3) Normalization and CLAHE

All of these options have been tested on the same parameter set using only the affine transform. As all of these options are computed using the same parameter-file, a comparison is valid. Table I shows that using no preprocessing, a target registration error of 20.50 ± 7.55mm is achieved. The use of min-max normalization in combination with CLAHE results in a TRE of 21.38 ± 8.04mm. While the results are worse using

TABLE I
RESULTS OF THE REGISTRATION COMPARING DIFFERENT APPROACHES.

Parameter File	Affine	Bspline	Segmentation	CLAHE	Time (min)	TRE in mm					
						COPD1	COPD2	COPD3	COPD4	Mean	Std
No Registration					-	26.33	21.79	12.64	29.58	22.59	6.38
Par11	✓			✓	0:36 ± 0:04	25.98	26.26	7.50	25.75	21.38	8.01
Par11	✓				0:34 ± 0:04	25.37	26.03	7.56	23.02	20.50	7.55
Par11	✓		✓	✓	0:51 ± 0:11	15.12	14.68	5.63	12.90	12.08	3.82
Par11	✓		✓		0:46 ± 0:15	14.16	12.84	4.87	10.16	10.51	3.56
Par11	✓	✓✓¹	✓	✓	10:07 ± 3:24	6.88	6.73	1.90	10.91	6.60	3.19
Par11	✓	✓²	✓	✓	23:43 ± 5:58	8.11	7.75	2.36	10.07	7.07	2.86
custom	✓		✓		1:25 ± 0:24	14.17	12.88	4.79	10.18	10.50	3.60
custom	✓	✓			4:39 ± 0:33	2.39	5.60	1.95	3.71	3.41	**1.42**
custom	✓	✓	✓		4:57 ± 0:48	**1.39**	**4.86**	**1.34**	**2.63**	**2.55**	1.43
Voxelmorph			✓	✓	46:37 ± 5:12	39.76	11.28	31.50	33.12	28.92	10.53

1: Bspline file 1 and 2; **2:** Bspline file 1

normalization and CLAHE, no statistically evident conclusion can be drawn as the standard deviations of both have a big overlap with respect to the mean. However, as performing the registration using only segmentation performs better than using segmentation, normalization, and CLAHE in combination, a trend can be noted. For this reason, the custom parameter maps have not been evaluated using normalization and CLAHE.

The employment of segmentation, however, improved the results significantly compared to using no preprocessing. This results in TRE as low as 10.56 ± 3.56 mm using only the affine transform and the segmented images. While the difference is less using the custom parameter maps with bsplines with segmentation (2.63 ± 1.43 mm) vs without (3.41 ± 1.42 mm), it is still notable.

B. COMPARISON OF AFFINE AND BSPLINE

Here, two different approaches have been tested. The first one only using the affine transform, and the second employing an affine registration followed by bsplines. From a medical perspective, this comparison raises an interesting question: Do all areas of the lung expand uniformly, or are there specific regions that expand more? If the former is true, using a non-rigid transformation like bsplines might not yield significant improvements. However, if the latter is the case, we can expect better results with bsplines. A comparison of the target registration error reveals that the usage of bsplines (6.60 ± 3.19 mm) performs significantly better than without (10.50 ± 3.60 mm). This assumption not only holds true for the custom parameter set, as shown before, but also for parameter set 11 as visible in Table I. This quantitative result suggests that the lung does not expand similarly in all areas. This assumption can be confirmed qualitatively using, e.g. Figure 1. It can be seen that the lung expands significantly less around the spline than the anterior. Both the quantitative and qualitative results confirm that the usage of a non-rigid transformation such as bsplines is needed to accurately register the inhaling and exhaling images.

C. COMPARISON OF PARAMETER SETS

Using the Elastix library. Two different parameter sets have been tested:

1) Parameter set 11 from modelzoo
2) A custom parameter set based on Parameter set 11

As discussed in the previous sections, both parameter sets benefit from the use of segmentation in combination with an affine and a bspline transform. Here, only the best results from each parameter set are compared. Using the parameter set 11, the best target registration error was (6.60 ± 3.19 mm). The usage of the custom parameter set results in the best overall TRE of all experiments with 2.63 ± 1.43 mm. This improvement has been made possible by accounting for different spacings when building the image pyramid and by using a more robust metrics estimation.

D. COMPARISON OF ELASTIX AND VOXELMORPH

In the comparison between itk-elastix and VoxelMorph for chest CT volume registration, itk-elastix proved to be more effective, achieving a target registration error (TRE) of 2.63 ± 1.43 mm. This result significantly surpassed VoxelMorph, which had a TRE of 28.92 ± 10.53 mm. One possible reason for VoxelMorph's underperformance could be its reliance on a generic deep learning architecture that may not be fully optimized for the specific challenges presented by lung tissue deformation during different respiratory phases. Additionally, the choice of the model's hyperparameters, such as the loss function in VoxelMorph's training process, might not have been ideally suited to capture the complexities of aligning thoracic structures, potentially leading to less accurate results.

VI. CONCLUSION

In conclusion, the great variety of methods tested led to a multitude of results, which will be summarized in the following: First, classical approaches can still outperform deep-learning methods if fine-tuned properly and therefore still have relevance. Second, normalization and contrast enhancement did not lead to improved results, but rather worse ones. Third, segmentation improved results significantly as the registration was targeted only at relevant regions. Fourth, the used parameter set is the biggest influence factor and has to be adapted to the dataset at hand. Once parameter set 11 was changed to use a 3D imaging pyramid accounting for the
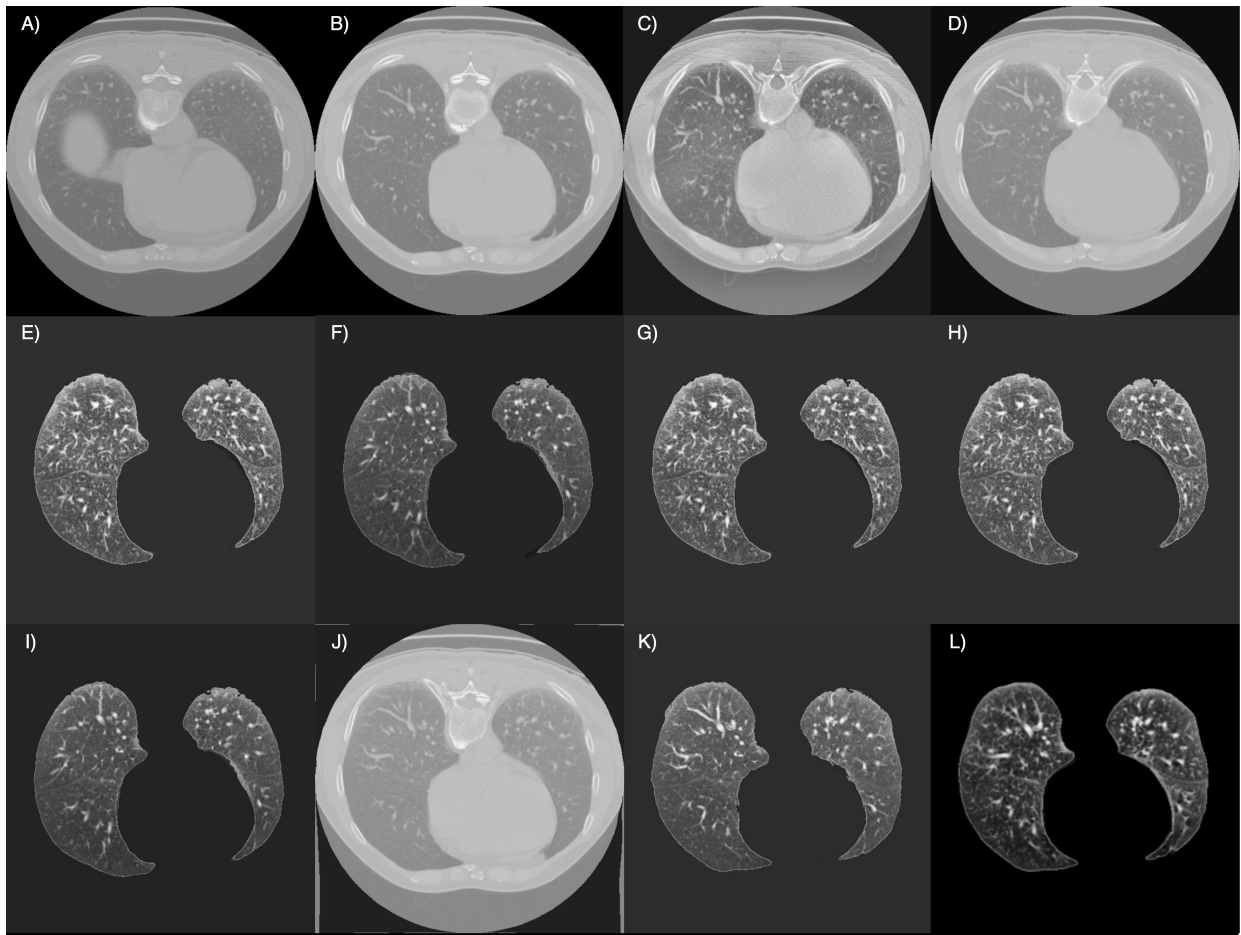
Fig. 5. Example of the registration obtained with the different approaches implemented in this project. **A)** Moving Image **B)** Fixed Image **C)** Affine + Normalization + CLAHE, **D)** Only Affine, **E)** Affine + Segmentation + Normalization + CLAHE, **F)** Affine + Segmentation , **G)** Affine + Bspline set 1 + Bspline set 2 + Segmentation + Normalization + CLAHE, **H)** Affine + Bspline set 1 + Segmentation + Normalization + CLAHE, **I)** Custom parameter set affine only, **J)** Custom parameter set affine + Bspline + Segmentation, **K)** Custom parameter set affine + Segmentation, **L)** Voxelmorph

spacing of the dataset, results improved significantly. Finally, we want to mention that we both enjoyed working on the project, especially the exploration of different preprocessing, segmentation, and registration techniques.

VII. HARDWARE SPECIFICATIONS

In this project, we utilized the NVIDIA RTX A6000 GPU, featuring 48 GB of GDDR6 memory and based on NVIDIA's Ampere architecture, for efficient deep learning computations. Complementing this, our general computational tasks were handled by an Intel(R) Xeon(R) Gold 5315Y CPU @ 3.20GHz, equipped with 8 cores and 96 MiB of L3 cache, ensuring robust overall data processing.

REFERENCES

[1] R. Castillo, E. Castillo, R. Guerra, V. Johnson, T. McPhail, A. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Physics in Medicine and Biology*, vol. 54, no. 7, pp. 1849–1870, 2009, accessed: March 5, 2009.

[2] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, no. 1, Aug 2020. [Online]. Available: https://doi.org/10.1186/s41747-020-00173-2

[3] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[4] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, p. 1788–1800, Aug. 2019. [Online]. Available: http://dx.doi.org/10.1109/TMI.2019.2897538

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.