We would like to thank all reviewers for their time and effort writing these valuable reviews. The evaluation of different image modalities proposed by Reviewer 1 is an interesting idea, we will consider for future work.

Reviewer 3 mentioned that a performance measure with other recent methods would be beneficial. Most of the current methods in literature work on the registration of 3D image data. We will do further performance evaluation when we extend our method for the 3D case. The code for this paper will be released with the camera-ready version.

In the following, we focus on the questions given by Reviewer 2.

1: The network contains fewer parameter than the B-spline method?
The presented network does not contain fewer parameters compared to the classical B-spline method for optimization. However, the number of parameters to describe the final displacement field is smaller than with the B-spline registration, i.e. our method creates a compact representation of the displacement.

2: Why sequence based image registration?
The overall goal of the sequence-based registration approach is that the number of steps that are needed for the registration are not fixed from the beginning. The network should continue the registration until both images are correctly registered. In order to show that the idea of sequence-based registration is able to register two images in the first place, we use a fixed number of steps. For future work, we will investigate how we can integrate the stopping of the registration into the learning process.

3: Why a Gaussian function as local transformation?
We use a Gaussian function as local transformation as it provides an elegant method to modify the shape of the local transformation, i.e anisotropic local deformations. Furthermore, it is straightforward to extend for the 3D case. However, we are not limited to Gaussian functions.

4: Why is the parameter network not explicitly made aware of the result of the position network?
We decided to use a hierarchic approach for the estimation of the position on different spatial resolution in order to get an accurate estimation for the position. This idea is inspired by the hierarchic approaches used in classical image registration methods in order to avoid local minima during the optimization. However, for the parameter estimation, we think it is useful to place the parameter network at the end because with this we can take the global registration status into account. It might also have an additional regularization effect to prevent the parameter network from creating local extreme values, compared to the current registration status. It would be an interesting idea to add the parameter estimation for the different spatial resolutions too.

5: Why is the registration of chest MRI images important and hard to solve?
As mentioned in Section 4 the MR image data we used for the experiments are acquired with a special MR-sequence that is able to detect physiological changes inside the lung. For the analysis of the image data, a registration of corresponding anatomical structures is needed. Based on the registered image data it is possible to detect impaired regions of the lung. Using MRI allows continuous monitoring of the disease state which is critical for the treatment of chronic lung disease like cystic fibrosis, especially for pediatric patients.
The registration of chest images is a major topic in the field of medical image registration, because of the discontinuous displacements at the border of the lung and the thoracic cavity. Here, the major problem is that the thoracic cavity has a different motion direction compared to the lung. On the other hand, the displacement should be smooth inside the lung. Parametric transformation models like B-spline create smooth displacements but have problems at sliding organ boundaries. Non-parametric transformation models are able to create sharp discontinuities but have problems with the smoothness inside the lung. With our method, we can provide smooth displacements inside the lung and we are able to adapt the shape and the position of the local transformations at the lung border to increase the registration accuracy in this regions.

6: Registration of longitudinal image data and other image modalities like MR/CT.
As mentioned in question 5 the registration of thoracic images is challenging. Longitudinal registration or the registration of different modalities for 2D images is difficult as the slice positions may not be the same for two acquisitions and therefore different structures are acquired. Both experiments are very interesting, but we will leave this for future work together with a 3D version of the presented method.

7: How to obtain a corresponding point given the current model?
A point to point transformation is not directly possible with this model as it needs two images as input. After the registration is performed by the network a point in the moving image can be transformed with the final displacement. The final displacement is the sum of all local displacements.

8: Detailed explanation of the B-spline method.
The B-spline method belongs to the class of parametric registration methods. Here, the transformation of interest between the fixed and the moving image is approximated by a number of B-spline kernel functions located on a fixed regular grid over the image domain (Figure 2b). The fixed grid is used because the transformation of interest is not known in advance. The weights of the B-spline kernel functions are optimized using gradient based optimization methods for each image pair. A more detailed description is given in [22].