UCLA
UCLA Previously Published Works

Title

Using neural networks to extend cropped medical images for deformable registration among images with differing scan extents

Permalink

https://escholarship.org/uc/item/3q28d8mh

Journal

Medical Physics, 48(8)

ISSN

0094-2405

Authors

McKenzie, Elizabeth M
Tong, Nuo
Ruan, Dan
et al.

Publication Date

2021-08-01

DOI

10.1002/mp.15039

Peer reviewed

Using Neural Networks to Extend Cropped Medical Images for Deformable Registration Among Images with Differing Scan Extents

Elizabeth M. McKenzie, M.S., **Nuo Tong**, **Dan Ruan, Ph.D.**, **Minsong Cao, Ph.D.**, **Robert K. Chin, M.D. Ph.D.**, **Ke Sheng, Ph.D.**

Department of Radiation Oncology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90024 USA

Abstract

Purpose: Missing or discrepant imaging volumes is a common challenge in deformable image registration (DIR). To minimize the adverse impact, we train a neural network to synthesize cropped portions of head and neck CT's and then test its use in DIR.

Methods: Using a training dataset of 409 head and neck CT's, we trained a generative adversarial network to take in a cropped 3D image and output an image with synthesized anatomy in the cropped region. The network used a 3D U-Net generator along with VGG deep feature losses. To test our technique, for each of the 53 test volumes, we used Elastix to deformably register combinations of a randomly cropped, full, and synthetically full volume to a single cropped, full, and synthetically full target volume. We additionally tested our method's robustness to crop extent by progressively increasing the amount of cropping, synthesizing the missing anatomy using our network, then performing the same registration combinations. Registration performance was measured using 95% Hausdorff distance across 16 contours.

Results: We successfully trained a network to synthesize missing anatomy in superiorly and inferiorly cropped images. The network can estimate large regions in an incomplete image, far from the cropping boundary. Registration using our estimated full images was not significantly different from registration using the original full images. The average contour matching error for full image registration was 9.9mm, while our method was 11.6mm, 12.1mm, and 13.6mm for synthesized-to-full, full-to-synthesized, and synthesized-to-synthesized registrations, respectively. In comparison, registration using the cropped images had errors of 31.7mm and higher. Plotting the registered image contour error as a function of initial pre-registered error shows that our method is robust to registration difficulty. Synthesized-to-full registration was statistically independent of cropping extent up to 18.7cm superiorly cropped. Synthesized-to-synthesized registration was nearly independent, with a –0.04mm change in average contour error for every additional millimeter of cropping.

Corresponding authors: Ke Sheng, Ph.D., Department of Radiation Oncology, University of California, Los Angeles, 200 Medical Plaza, B265, Los Angeles, California, 90024, ksheng@mednet.ucla.edu, Phone: 310-853-1533, Fax: 310-983-3603.

Conflict of Interest: None

Conclusions: Different or inadequate in scan extent is a major cause of DIR inaccuracies. We address this challenge by training a neural network to complete cropped 3D images. We show that with image completion, the source of DIR inaccuracy is eliminated, and the method is robust to varying crop extent.

Keywords

deformable registration; scan extent; deep learning

Introduction

Deformable image registration (DIR) is a topic of intense research and clinical interests in radiation therapy. DIR establishes correspondence between medical images for imaging information synthesis, dose accumulation and adaptive treatment planning. For these applications, DIR is frequently performed on image pairs that exhibit non-rigid motion. On the other hand, the usefulness of DIR can be limited by low accuracy and robustness. The process of matching one image to another can introduce erroneous or unrealistic tissue deformation[1], requiring practice of caution when DIR is involved in clinical decision for interventions. In cases where the DIR accuracy is unsatisfactory or the accuracy cannot be verified, rigid registration is used instead as a compromise[2–4].

Besides differences in multimodal image intensity and large deformation, a common factor contributing to the DIR difficulty is the mismatch in the image scan extents or the field-of-view. Because the boundary conditions are not explicitly available, unrealistic deformation is often introduced in DIR. The unrealistic stretch or compression of tissues is most severe near the edges of an image but can propagate through the entire image volume with smoothness constraints in the DIR DVF. The scan extent mismatch is common in retrospective analysis, where images were acquired with varying scanning protocols, as well as in multimodal registration problems. In image guided radiotherapy, cone beam CT (CBCT) images are used to help with patient set up but CBCTs have a substantially more limited coverage in both the axial and longitudinal dimensions compared with the planning CT. Imaging volume mismatch is also common in MR to CT registration. MR provides superior soft tissue visualization that is helpful for tumor and normal tissue delineation, but the MR imaging volume is often smaller than the planning CT. MR images acquired on oblique orientations further complicate the imaging volume mismatch issues. We previously demonstrated that the challenges in registering MR to CT due to differences in imaging intensities can be mitigated via synthetic image bridge[5] but the issues due to mismatched imaging volumes persist. According to TG-132, differences in scan extent are a major source of deformable registration error[6].

Research to mitigate the adverse impact due to imaging volume mismatch has been reported. A straightforward approach to reduce the registration error due to mismatched imaging extent is to manually crop the larger imaging volume to match the scan length of the shorter image[7,8]. Manually cropping the images not only reduces the workflow efficiency, but also introduces error because the image matching lines are not explicitly available to the operator. The error can be substantial when large patient pitch correction or

deformation is involved. Periaswamy and Farid used an expectation maximization algorithm to simultaneously segment the more complete image volumes and register partial images[9]. The method effectively contained the registration error due to image artifacts but its ability to handle both large deformation and mismatched scan volume was not demonstrated. To address differing scan lengths in CBCT and planning CT registration, researchers and then relied on the DVF smoothness constraint to outside the effective field of view[4,10]. The method was shown to reduce the DIR error, but the registration accuracy was still limited by the lack of contextual information due to missing volumes.

Aside from their specific algorithms, the existing methods share the strategy of using the intersection of the two images as the starting point of DIR. By doing so, the imaging information in the more complete image is discarded despite its potential value for the overall registration accuracy. In this study, we take a fundamentally different approach. Instead of cropping the images, we propose to fill the missing portion of the anatomy using neural networks. The registration can then proceed using the artificially extended image. We design the study to answer two questions: (1) Do registrations with artificially extended images perform as well as registration pairs with equal extent (2) How does the quality of the registration with artificially extended images vary as a function of the initial amount of missing tissue?

Materials and Methods

Dataset

Head and Neck CT images were acquired from The Cancer Imaging Archive (TCIA) dataset[11]. We had a total of 409 training, 53 validation, and 53 testing images. Scan extent went from the top of skull to approximately the carina. Scanning beds and immobilization equipment were masked out of the images. For input into the network, all images were rigidly registered to a template image and downsized to 128×128×128 with 4mm isotropic voxels. Image intensity value were clipped to a range of [−1024, 3000], then normalized to [−1, +1]. For analysis, volumes were automatically segmented using a neural network approach[12,13]. This resulted in 16 contours per patient.

Network

For this work, we used a Generative Adversarial Network (GAN) approach to extend the cropped volume[14]. We term the network **CropGAN**. A GAN consists of a generator to create synthesized data, and a discriminator to judge if data is synthesized or real. The input to our generator was a cropped volume, and the output was a volume with the missing portion replaced with synthesized data. The cropped region was randomly created each iteration of training, where the angle was randomly varied between 0 and 45 degrees in the superior-inferior direction, and between −5 and 5 degrees in the other 2 dimensions; and the amount of cropping was varied on both the superior and inferior edges from 120 to 210mm. We chose to vary the cut angle of the crop to simulate two common scenarios in DIR. First, as a preprocessing step, rigid registration is performed prior to DIR. Correction of the patient pitch and yaw will lead to oblique cutting planes relative to the target image. The second scenario is registration of the MR acquired in oblique orientations. In CropGAN,

the generator was a 3D U-net with skip connections[15]. At the bottom of the U-net, we used 4 dilated convolutions to increase the amount of contextual information for prediction[16]. All of the convolutions in the U-net used instance normalization with elu activation and were gated so the network could adaptively learn feature selection, as was done by Yu et al[17]. The discriminator had 3 inputs: the cropped image, either the original full image (uncropped) or synthesized output from the generator (synthetically uncropped), and the mask used to crop the image. The cropped and uncropped (either full or synthetic) inputs were first concatenated together. The Discriminator was dual branching, with one branch operating on the entire concatenated image, while the other branch applied the mask for cropping to only use the data from within the mask to focus on the fidelity of the synthesized portion. The discriminator used spectral normalization, which has been shown to add stability to discriminator training[18]. The output of the discriminator was a concatenation of the two branches. Figure 1 shows details of the networks.

For the loss function we followed the formulation of Hui et al[19], which uses several deep feature-based losses. We passed the generated and target uncropped image through a previously trained VGG network[20]. This network was trained to classify CT and MR imaging sites from patches and had learned activations pertinent to these modalities' features[21]. An example showing the first 5 activation layers for the generator output and ground truth target is given in Figure 2.

We compared the activations between the generated and target images in two ways. First, we compared the activations from the first 5 convolutional VGG layers [Equation 1].

$$loss_{vgg} = \sum_{l=1}^{5} w^l \frac{\left\| \Psi^l_{I_{gt}} - \Psi^l_{I_{output}} \right\|_1}{N\Psi^l_{I_{gt}}} \qquad \text{Equation 1}$$

Where $\Psi^l_{I_*}$ is the activation map of the lth layer, for image volume I_* (gt = ground truth, output=output from generator). $N\Psi^l_{I_{gt}}$ is the number of elements in the ground truth image's l^{th} layer. w^l weights each addend as a function of the channel size of the lth layer of the ground truth image [Equation 2]. $C^l_{\Psi^l_{I_{gt}}}$ in Equation 2 is the channel size of $\Psi^l_{I_{gt}}$.

$$w_l = \frac{1e6}{C^l_{\Psi^l_{I_{gt}}}} \qquad \text{Equation 2}$$

Second, to focus on more challenging areas of the image, we compared the error map weighted activations from the first two VGG layers. [Equation 3]

$$loss_{vgg\_challenge} = \sum_{l=1}^{2} w^l \frac{\left\| M^l_{guidance} \odot \left(\Psi^l_{I_{gt}} - \Psi^l_{I_{output}} \right) \right\|_1}{N\Psi^l_{I_{gt}}} \qquad \text{Equation 3}$$

Where $M_{guidance}^{l}$ is the error map associated with layer *l* and is used to give more weight to the VGG layer differences which are more challenging to match. For each layer, $M_{guidance}^{l}$ is given by $M_{guidance}^{l+1} = average\ pooling\left(M_{guidance}^{l}\right)$. $M_{guidance}^{1}$ is equal to $M_{guidance,p}$ which is the error map value at position p [Equation 4]. $M_{guidance,p}$ is derived from the generated image and its corresponding ground truth. Average-pooled guidance maps give a spatial correspondence between the differences seen in the images, and the differences seen at deeper layers.

$$M_{guidance,\,p} = \frac{M_{error,\,p} - \min(M_{error})}{\max(M_{error}) - \min(M_{error})} \qquad \text{Equation 4}$$

where, $M_{error} = (I_{out} - I_{gt})^{2}$

Mean absolute error was used to assess the fidelity between the generated and target uncropped images. This is given as $L_{fidelity}$. For the adversarial loss, we used a Wasserstein Hinge loss[22].

In addition to the adversarial loss, the discriminator also used a deep feature-based loss. The activation layers of the cropped-area discriminator branch were used to compare the generator output and ground truth [Equation 5].

$$Loss_{disc\_features} = \sum_{l=1}^{6} w^{l}\frac{\left\|D^{l}(I_{gt}) - D^{l}(I_{output})\right\|_{1}}{N_{D^{l}(I_{gt})}} \qquad \text{Equation 5}$$

The total loss function is thus:

$$Total\ loss = L_{adverarial} + \lambda_{1}L_{fidelity} + \lambda_{2}L_{vgg} + \lambda_{3}L_{vgg\_challenge} + \lambda_{4}L_{disc\_features} \qquad \text{Equation 6}$$

We searched for a stable training result by iteratively varying the weights (λ_{*}) using the validation set. This led to empirically selected weights of 20, 10, 10, and 5, respectively. Further tuning may possibly lead to improved results. Our generator and discriminators used an RMSprop optimizer with a learning rate of 0.00005. We used a batch size of 2 and trained for 2000 epochs.

The output from the network was a 128×128×128 image with 4mm^3 voxels. The synthesized portion was resized to 512×512×512 (1mm^3 voxels) to match the size of the original image. The final synthetically extended image only had synthesized voxels in the cropped region. The non-cropped portion was copied to the final image.

Registration

We tested how well deformable registration with the synthetic cropped images compared to uncropped (ground truth) registration and cropped registration. To do this, we deformably registered the moving images (cropped, uncropped, and synthetic uncropped) to the same target images (cropped, uncropped, and synthetic uncropped) in all unique combinations. It is worth noting that for fair comparison the synthetic image volumes are only used to

assist DIR. When the moving image was synthetically extended, we applied the resulting deformation vector field to the cropped image such that the final result only included actual scanned data.

Without losing generality, we performed registrations using an open source B-spline method (Elastix[23,24]). We used a multi-resolution deformable registration scheme and mutual information as the cost function, as in a previous publication[5]. This method was selected due to its competitive performance in registering head and neck images[25], open-source nature to facilitate comparison, and flexible registration parameter settings; however, the CropGAN images are expected to work with other registration algorithms.

Analysis

We tested our hypothesis that synthetically extending cropped images would lead to the same registration quality as a registration performed with the full, ground truth images by evaluating the similarity between deformed and target contours. To avoid being skewed by the organ size, instead of the Dice index, the similarity was calculated using the 95% Hausdorff distance surface matching metric[26,27]. We analyzed our results using a one-way ANOVA amongst registration pairs, as well as a linear regression between the pre- and post-registration contour similarity. All analyses were performed using GraphPad Prism.

We tested our secondary hypothesis that the registration quality using synthetically extended images would be the same as using full images independent of the initial cropping amount (range of approximately the superior apex of the skull to the inferior nose: 120 to 210 mm superior cropping) by cropping the same image by variable amounts, synthesizing the missing portion using CropGAN, then performing the same registration tests. For this experiment the angle of cropping was kept a constant 23 degrees in the superior-inferior direction. This angle was selected to be the middle of the angle ranges used during training. The other two cropping planes had angles of zero. The results were averaged over 3 patients and analyzed using linear regression.

Results

Execution

Training the CropGAN network took 6 days on one Nvidia Quadro RTX 8000 GPU. Once completely trained, inference took 0.04 seconds to synthesize the missing part of the cropped image.

Registration Comparisons

We performed 364 deformable image registrations to compare all 7 combinations of source and target volumes across the 52 test images (one of the 53 test images was held back as the target image), with random amounts of induced cropping from 120 to 210mm. An example showing one such set of registrations is given in Figure 3 and Figure 4. The columns are for a given target image (synthesized, cropped, or full) and the rows are for a given source image (synthesized, cropped, or full). The intersection of a row and column show the registration result (Figure 4 shows the respective deformed contours overlaid on the target).

From the provided example, it is clear that deformable registration between images with different scan extent leads to unrealistic distortion. Compared with the worst case when the moving image is cropped and the target image is full, registering a full volume to the cropped volume results in less distortion, though it is still worse than registering between two uncropped images. Registering with the synthesized images in all three cases leads to close performance to the registration of uncropped images as shown in the deformed contours of Figure 4.

In the quantitative analysis, the 95% Hausdorff distance averaged across all contours is displayed with 95% confidence intervals in Figure 5. Using a one-way ANOVA with a post-hoc Tukey multiple comparison test, registration using CropGAN synthesized images in all three cases is not statistically different from the best-case full image registration (p>0.9), while it is significantly different from registrations using a cropped image in either the source or target (p<0.0001). While the average contour distance of registrations using synthesized images was approximately half that of a simple rigid alignment, this difference was not statistically significant. We strengthened these conclusions by testing for equivalence between our proposed method and full image registration using a two-one-sided t-test. We chose our equivalency delta to be the average error reported for the automatic contouring algorithm (3.39mm 95% Hausdorff distance[13]). We concluded with 95% confidence that both synthesized-to-full and full-to-synthesized registrations were equivalent to a full-to-full image registration within the error of contouring. Registrations using synthesized images for both the source and target had confidence limits 1.6mm beyond this contouring error threshold. Thus, while it may not be significantly different from full image registration, it is advantageous to have either the source or target image be full. Interestingly, the full-to-cropped registration had confidence intervals 20.7mm beyond the contour error.

Figure 6 shows the 95% Hausdorff distance of individual contours. The trends are consistent with the average distance analysis. The esophagus showed the greatest error across registrations due to inconsistencies in the inferior range of this lower contrast contour.

To analyze the dependence of our results on the degree of registration difficulty, we plotted the average 95% Hausdorff distance of the registrations as a function of the original degree of error (Figure 7). A simple linear regression was used to obtain best fit lines. The slope for the proposed method closely follows that of the best-case full image registration. Specifically, the full-to-full image registration and full-to-synthesized slopes were not significantly different (p=0.29), with a shared slope of 0.27. Synthesized-to-full and synthesized-to-synthesized registrations were also not significantly different (p=0.10) with a shared slope of 0.48; however, they were significantly different from a full-to-full registration. Additionally, the relatively flat nature of these lines indicates that our proposed method performs well across a broad range of registration difficulty. When both the source and target images are cropped, the more challenging registrations can have errors exceeding 10cm. When only the source or target are cropped errors are in the 3–5cm range. This is consistent with the qualitative results shown in Figure 3.

Dependence on Cropping Amount

To analyze the effect the amount of cropping had on the registration result, we progressively increased the amount superiorly cropped in the moving image, while keeping the angle of cropping constant (S-I plane=23°, A-P=L-R=0°). These increasingly cropped images were passed through the trained CropGAN network to synthesize the missing regions. The target images (full, cropped, and synthesized) were kept the same for this experiment. The results averaged across three patients, along with a visualization of the cropping extent, are shown in Figure 8. The lines for registrations where the moving image was full are flat, as these are not changing in this experiment; however, they offer useful reference lines. Our proposed method performs as well as a full image registration until approximately 18.7cm of scan is missing from the superior edge. The synthesized-to-full registration was closest to the full registration result (average difference of 1.5mm). When both the moving and target images were synthesized, the average difference was 4.5mm from the full registration. This is contrasted with the registrations including cropped images, which had an average difference of 22mm and higher. The registrations with cropped moving images show larger error, however there is a noticeable decrease around 170mm. For cropped-to-cropped registration (black line), the error decreases until the cropped moving image extent matches the cropped target image's extent. As the moving image is cropped further, the extents again become mismatched, and the error increases. For the cropped-to-full registration (brown line), the error increases until the regularization of the registration algorithm prevents further stretching of the small moving image to the larger full image. While the average error for cropped-to-full registration appears to decrease slightly in Figure 8, observing the registration results directly reveals extreme distortion for these larger crop amounts.

For our proposed method, the synthesized-to-full and synthesized-to-synthesized registrations are *independent* of crop extent (slope was not significantly different from 0, p=0.2887 and 0.8556, respectively) until this extreme cut point of 18.7cm. This point roughly corresponds to the region of the nose, suggesting this to be an important landmark for synthesis. This is in sharp contrast to the large, varying results with the original cropped registrations. These results show that our method is robust across a wide range of scan extents.

CropGAN Synthesis Visual Performance Variation

While it not the intent to recover the accurate anatomy for individual patients, it is interesting to visually exam the potential for creating missing tissues. Figure 9 shows two representative cases for good (top row) and poor (bottom row) synthesis of missing imaging volumes. In the good case, the network synthesized realistic anatomies including sinuses, sternum, and heart. In the poor case, the network failed to generate the patient nasal and skull base anatomies possibly due to the low number of training images and large variation of metal artifacts. In any case, the anatomies generated using CropGAN in its current form is not actual and cannot be used as such. For registration purposes, however, the quality of image synthesis achieved using CropGAN appears to provide adequate contextual information for DIR.

Discussion

We present here a novel solution to directly address adverse effects due to inadequate or mismatched scan extent in deformable registration. DIR between images of insufficient extents is a major source of registration error. Existing approaches focus on cropping the larger or more complete images to better match the cropped images, which results in loss of information that could have benefited the registration. This is particularly problematic when both the moving and target images have inadequate scan extent for registration. We were able to artificially extend cropped images using a method which is fully 3D. The method is fully automated and able to handle a broad range of scan extent differences. Once trained, our method fills the missing volume in 0.04 seconds, making it an amenable addition to a clinical workflow. To our knowledge, the current study is the first to synthesize the missing or cropped imaging volumes to improve the registration performance. It is worth emphasizing that the synthesized anatomy cannot represent the actual patient anatomy in the missing volumes. It serves the purpose of assisting DIR of the actual imaging volume. Therefore, when the moving image is cropped, we apply the DVF from the synthesized moving image registration to the cropped image, thereby only including the real imaging data in the final result.

The task to synthesize missing image slices itself is also novel. Neural networks have been used to inpaint a missing patch inside a 2D medical image slice[28–30], which is a considerably less challenging problem that is analogous to interpolation with known boundary conditions around the missing patch. In contrast, synthesizing data in a cropped image is analogous to extrapolation with undefined boundary conditions. A study looking at network-based image extension in 2D landscape photos was able to successfully extend natural 2D images, however they caution that their results did not apply well to photos of human faces.[22] Our proposed CropGAN uses generative adversarial networks (GAN) to synthesize missing data, a technique that has been well tested in network-based inpainting tasks[17,31]. Specifically, we based our method on the winner of the AIM 2020 Challenge on Extreme Inpainting[19,32], which used deep features in the generator, discriminator, and a VGG net as terms in the loss function. That study showed impressive results filling in holes of 2D color photos, yet has not yet been pursued for image extension nor for 3D medical images.

Our proposed method of using a neural network to synthetically extend 3D cropped images improves deformable registration between images of differing scan extents. It creates a bespoke synthesis in the cropped region that takes cues from each image's anatomy. In most cases, it synthesizes realistic anatomies even far beyond the line of cropping. CropGAN creates details such as sinuses, lungs, orbits, and heart that continue smoothly from the available anatomical information. These large details help anchor the registration algorithm while it optimizes the correspondence within the real portions of the image. This advantage was seen even with extreme differences in scan extent (e.g., a cranium to carina scan and a scan only including the neck).

It has been observed that deforming a full image to a cropped image is more robust than the reverse. Therefore, implementing inverse consistency[33] in DIR could conceivably improve

the registration of the former case. However, as shown in the study, CropGAN synthesized images still significantly outperforms the case of full-to-crop registration.

We noticed that the synthesized images were poorer when metal imaging artefacts were near the cropping boundary. This may be due to a lack of accurate anatomical cues near the boundary which the network can use to make its prediction. Interestingly, once further away from the artefact, the network can still create realistic anatomies, and the registration result is not significantly different from cases without such artifacts, as well as to the result using full images. Therefore, while the study was not designed to quantify the effects of artefacts, current results suggest our technique is robust to this effect.

We chose to use a b-spline based registration algorithm for our study, but this technique is generalizable for other algorithms. Our preliminary results suggest that CropGAN similarly improves demons based registration. Preprocessing with CropGAN could also aid in other medical imaging neural network tasks, since it can help to standardize the data.

One limitation of this technique is that it requires training data for each region one wishes to extend. We focused this study on CT images in the head and neck since this region can include large non-rigid motion. For other anatomical sites or imaging modalities, one would need a large dataset with similar scan extents to provide supervised training between an induced cropped image and its full image ground truth. For example, cone beam CT deformable registration may benefit from our method due to its limited field of view relative to the simulation CT target. However, the network would need to be trained and verified on this different modality. Improvement may also be seen by better selecting the trained VGG net used for deriving the deep features in the loss function. The network we used was trained for the unrelated task of classifying small 3D patches of CT and MR images by their scan site. The features learned from this network may not be optimal for our task, which used full image volumes. While having a VGG net trained on full images may lead to a better result, recent research has suggested that using deep features to assess image similarity can be surprisingly effective even when the network was trained for an unrelated task[34]. An additional limitation is the inherent uncertainty in the contours used for this study's analyses. We used a previously developed in-house segmentation network to increase reproducibility[13]. While this network demonstrated impressive dice scores, there is still an inherent amount of uncertainty. Despite the abovementioned limitations, we have provided a foundation upon which other studies can extend our work. The code which was used for this manuscript is openly available in Github at https://github.com/emckenzi123/CropGAN/.

Conclusion

Differences and inadequacy in scan extent is a difficult problem in medical image deformable registration. We proposed a solution using a neural network to synthesize the missing portions of the scan. These syntheses were able to successfully create realistic anatomy for the missing volume with details such as sinuses, orbits, skull, lungs, and heart. It was also robust to the amount of cropping in the inferior and superior directions. After filling the cropped volumes using CropGAN synthesis, the two images can then be

deformably registered as though they had the same full scan length. Using 95% Hausdorff distance on a selection of head and neck contours, we found that our registration workflow was able to match contours equally well to a registration with complete scans. CropGAN performance for DIR as a function of cropped tissue is robust up to until 20cm of the superior end of the head was missing. By using CropGAN as a preprocessing step to deformable registration, we have provided an intuitive solution to the challenge of registration with different scan extents.

Funding/Support:

Bibliography

1. Kirby N, Chuang C, Ueda U, Pouliot J. The need for application-based adaptation of deformable image registration. Med Phys. 2012;40(1):011702. doi:10.1118/1.4769114

2. Geets X, Daisne JF, Tomsej M, Duprez T, Lonneux M, Grégoire V. Impact of the type of imaging modality on target volumes delineation and dose distribution in pharyngo-laryngeal squamous cell carcinoma: comparison between pre- and per-treatment studies. Radiother Oncol. 2006;78(3):291–297. doi:10.1016/j.radonc.2006.01.006 [PubMed: 16499982]

3. Geets X, Tomsej M, Lee JA, et al. Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: Impact on target volume delineation and dose distribution using helical tomotherapy. Radiother Oncol. 2007;85(1):105–115. doi:10.1016/j.radonc.2007.05.010 [PubMed: 17562346]

4. Veiga C, McClelland J, Moinuddin S, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. Med Phys. 2014;41(3):031703. doi:10.1118/1.4864240 [PubMed: 24593707]

5. McKenzie EM, Santhanam A, Ruan D, O'Connor D, Cao M, Sheng K. Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge. Med Phys. 2020;47(3):1094–1104. doi:10.1002/mp.13976 [PubMed: 31853975]

6. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132: Report. Med Phys. 2017;44(7):e43–e76. doi:10.1002/mp.12256 [PubMed: 28376237]

7. Zhen X, Yan H, Zhou L, Jia X, Jiang SB. Deformable image registration of CT and truncated cone-beam CT for adaptive radiation therapy. Phys Med Biol. 2013;58(22):7979–7993. doi:10.1088/0031-9155/58/22/7979 [PubMed: 24169817]

8. Ottosson W, Lykkegaard Andersen JA, Borrisova S, Mellemgaard A, Behrens CF. Deformable image registration for geometrical evaluation of DIBH radiotherapy treatment of lung cancer patients. J Phys Conf Ser. 2014;489:012077. doi:10.1088/1742-6596/489/1/012077

9. Periaswamy S, Farid H. Medical image registration with partial data. Med Image Anal. 2006;10(3):452–464. doi:10.1016/j.media.2005.03.006 [PubMed: 15979375]

10. Yang D, Goddu SM, Lu W, et al. Technical Note: Deformable image registration on partially matched images for radiotherapy applications. Med Phys. 2009;37(1):141–145. doi:10.1118/1.3267547

11. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. J Digit Imaging. 2013;26(6):1045–1057. doi:10.1007/s10278-013-9622-7 [PubMed: 23884657]

12. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med Phys. 2018;45(10):4558–4567. doi:10.1002/mp.13147 [PubMed: 30136285]

13. Tong N, Gou S, Yang S, Cao M, Sheng K. Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images. Med Phys. 2019;46(6):2669–2682. doi:10.1002/mp.13553 [PubMed: 31002188]

14. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. 2014:1–9. doi:10.1001/jamainternmed.2016.8245

15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 5 2015. http://arxiv.org/abs/1505.04597.

16. Dinkla AM, Wolterink JM, Maspero M, et al. MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network. Int J Radiat Oncol Biol Phys. 2018. doi:10.1016/j.ijrobp.2018.05.058

17. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang T. Free-Form Image Inpainting With Gated Convolution. 2020:4470–4479. doi:10.1109/iccv.2019.00457

18. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. arXiv. 2018.

19. Hui Z, Li J, Wang X, Gao X. Image fine-grained inpainting. arXiv. 2020;(2):1–11.

20. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc. September 2014:1–14. http://arxiv.org/abs/1409.1556.

21. Avants B, Greenblatt E, Hesterman J, Tustison N. Deep Volumetric Feature Encoding for Biomedical Images. Vol 12120 LNCS. Springer International Publishing; 2020. doi:10.1007/978-3-030-50120-4_9

22. Teterwak P, Sarna A, Krishnan D, et al. Boundless: Generative Adversarial Networks for Image Extension. 2019. http://arxiv.org/abs/1908.07007.

23. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix : A Toolbox for Intensity-Based Medical Image Registration. 2010;29(1):196–205.

24. Shamonin D, Bron E, Lelieveldt B, Smits M, Klein S, Staring M. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. Front Neuroinform. 2014;7(January):1–15. doi:10.3389/fninf.2013.00050

25. Li X, Zhang Y, Shi Y, et al. Comprehensive evaluation of ten deformable image registration algorithms for contour propagation between CT and cone-beam CT images in adaptive head & neck radiotherapy. Zhang Q, ed. PLoS One. 2017;12(4):e0175906. doi:10.1371/journal.pone.0175906 [PubMed: 28414799]

26. Huttenlocher DP, Rucklidge WJ, Klanderman GA. Comparing images using the Hausdorff distance under translation. In: Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol 1992-June. IEEE Comput. Soc. Press; 1992:654–656. doi:10.1109/CVPR.1992.223209

27. Raudaschl PF, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. Med Phys. 2017;44(5):2020–2036. doi:10.1002/mp.12197 [PubMed: 28273355]

28. Armanious K, Mecky Y, Gatidis S, Yang B. Adversarial Inpainting of Medical Image Modalities. ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc. 2019;2019-May:3267–3271. doi:10.1109/ICASSP.2019.8682677

29. Wei D, Ahmad S, Huo J, et al. Synthesis and Inpainting-Based MR-CT Registration for Image-Guided Thermal Ablation of Liver Tumors. In: Shen D, Liu T, Peters TM, et al., eds. Medical Image Computing and Computer Assisted Intervention -- MICCAI 2019. Cham: Springer International Publishing; 2019:512–520.

30. Zhang S, Wang L, Zhang J, et al. Consecutive Context Perceive Generative Adversarial Networks for Serial Sections Inpainting. IEEE Access. 2020;8:190417–190430. doi:10.1109/access.2020.3031973

31. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS. Generative Image Inpainting with Contextual Attention. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2018:5505–5514. doi:10.1109/CVPR.2018.00577

32. Ntavelis E, Romero A, Bigdeli S, Timofte R. AIM 2020 Challenge on Image Extreme Inpainting. 10 2020. http://arxiv.org/abs/2010.01110.

33. Yang D, Li H, Low DA, Deasy JO, Naqa I El. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. Phys Med Biol. 2008;53(21):6143–6165. doi:10.1088/0031-9155/53/21/017 [PubMed: 18854610]

34. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2018;(1):586–595. doi:10.1109/CVPR.2018.00068
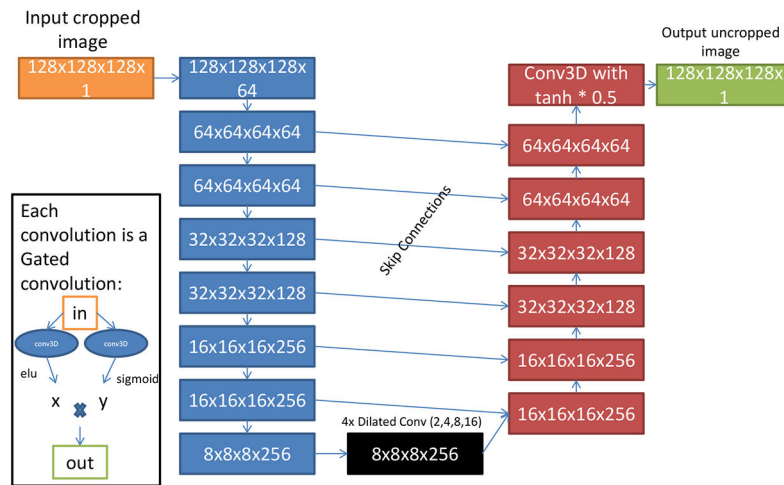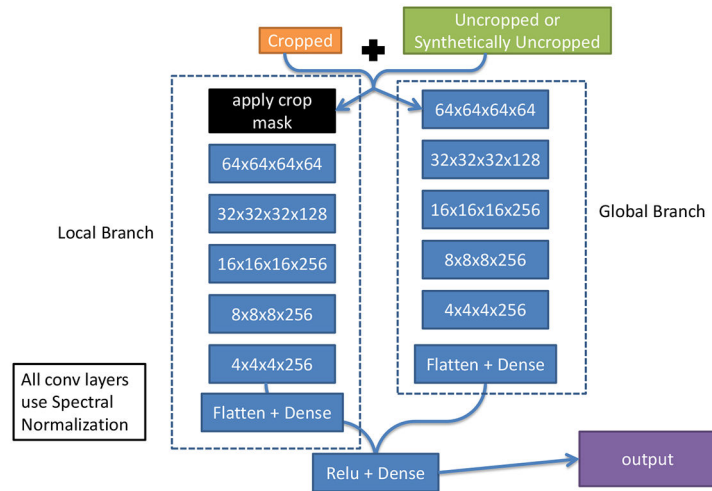
Generator (Unet)

Discriminator

Figure 1.
Architecture of Generator (top) and Discriminator (bottom).

VGG Deep Layer Comparison

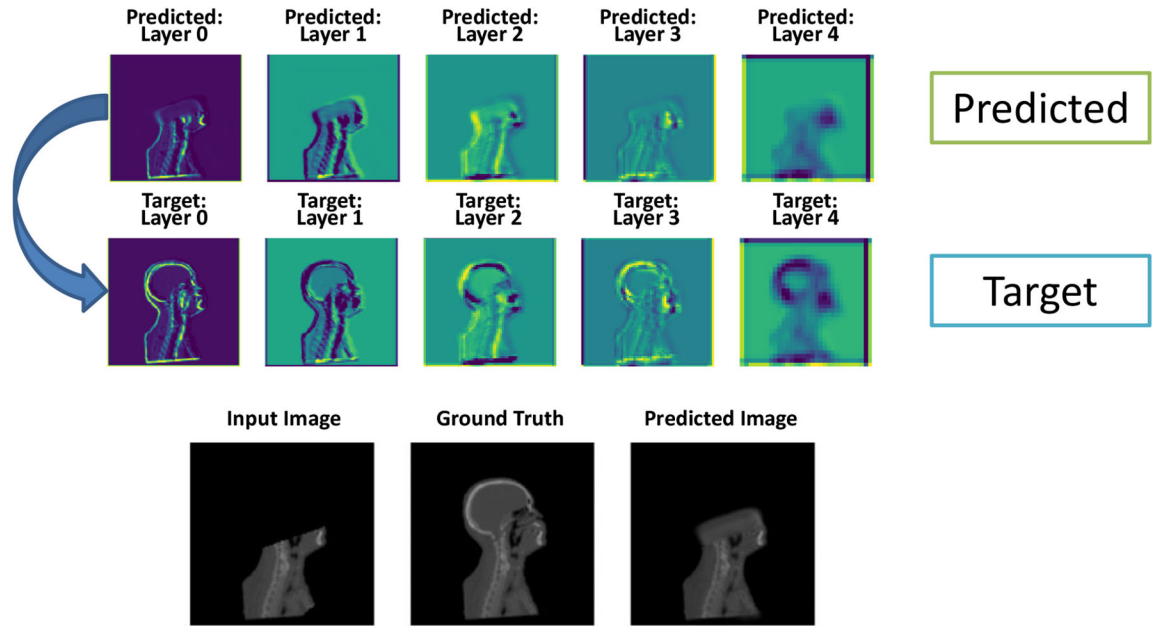**Figure 2.**
visualizes the activations in the first 5 layers of the VGG network for a predicted and uncropped ground truth image. These activations are used as features, which are compared using equations 1, 3, and 5 to produce a similarity metric, driving the predicted image to resemble the target. Best viewed in color (online version).

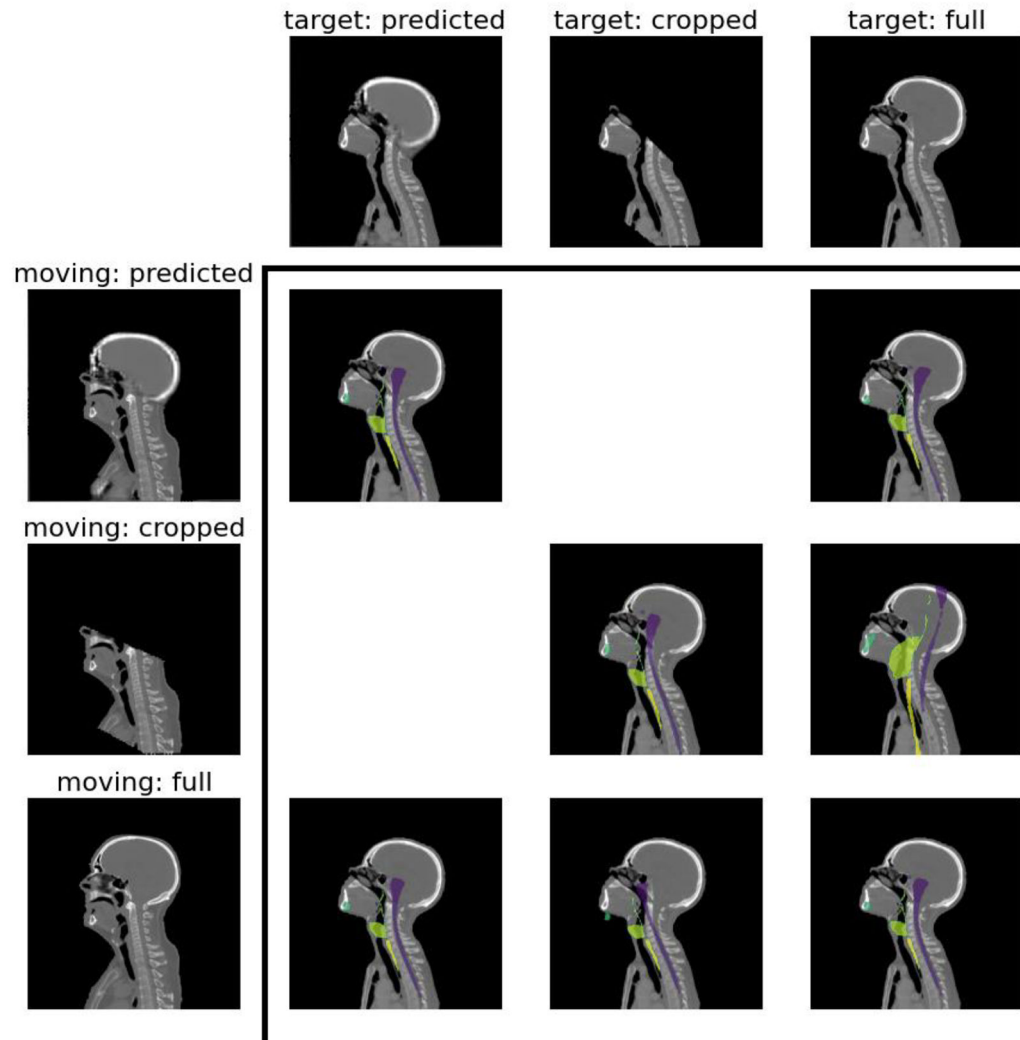# Registration Comparison for Subject: 0000
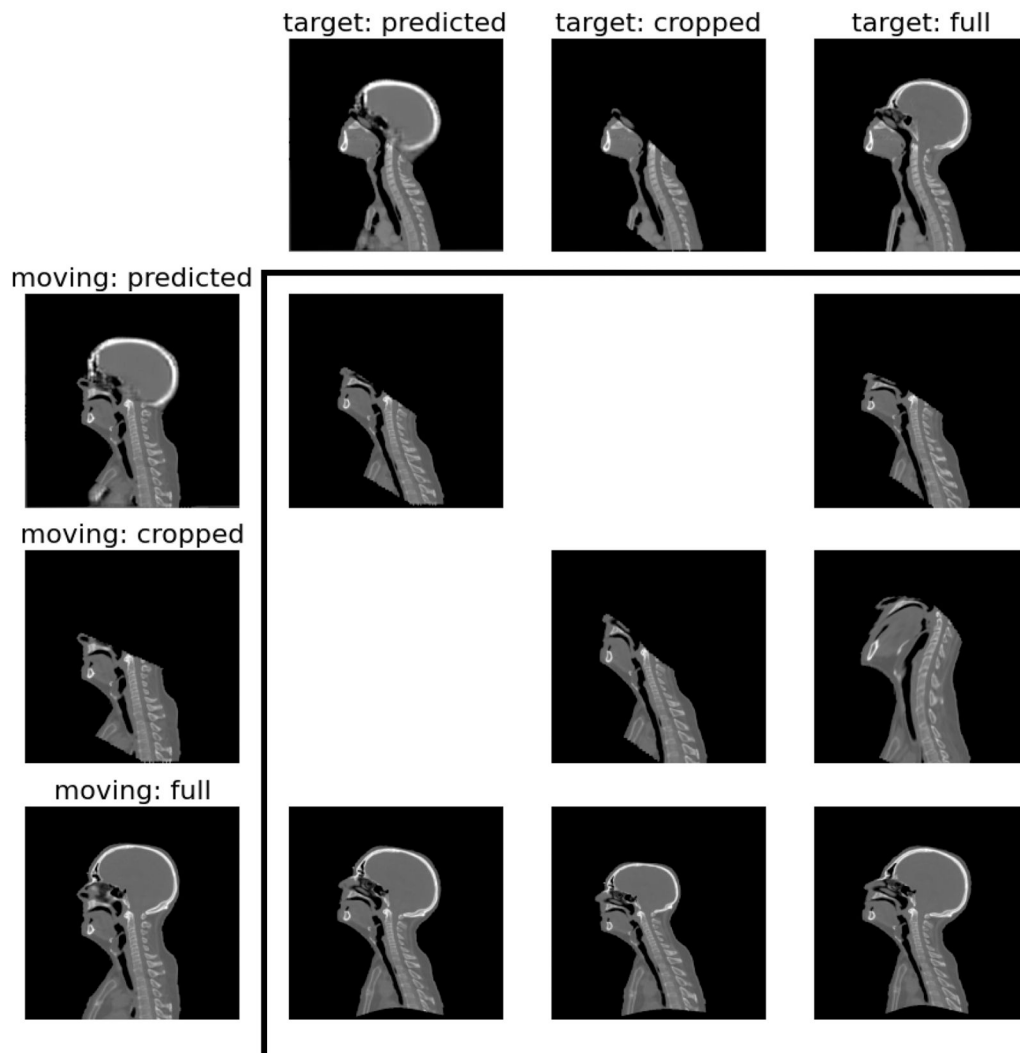
Figure 3.
Example registration of one of the 53 test patients. Columns are for a given target image; Rows are for a given moving image. The intersection shows the deformable registration result. Our method (top row of central grid) has applied the deformation vector field to the original cropped image, so only real data is included in the final registration result. Cropped-to-predicted, and predicted-to-cropped are not shown, since if one cropped image could be predicted, the other could feasibly be predicted as well.

Registration Comparison for Subject: 0000

Figure 4.
Example registration of one of the 53 test patients. Columns are for a given target image; Rows are for a given moving image. The intersection shows the deformable registration result applied to the moving image's contours. Best viewed in color (online version).
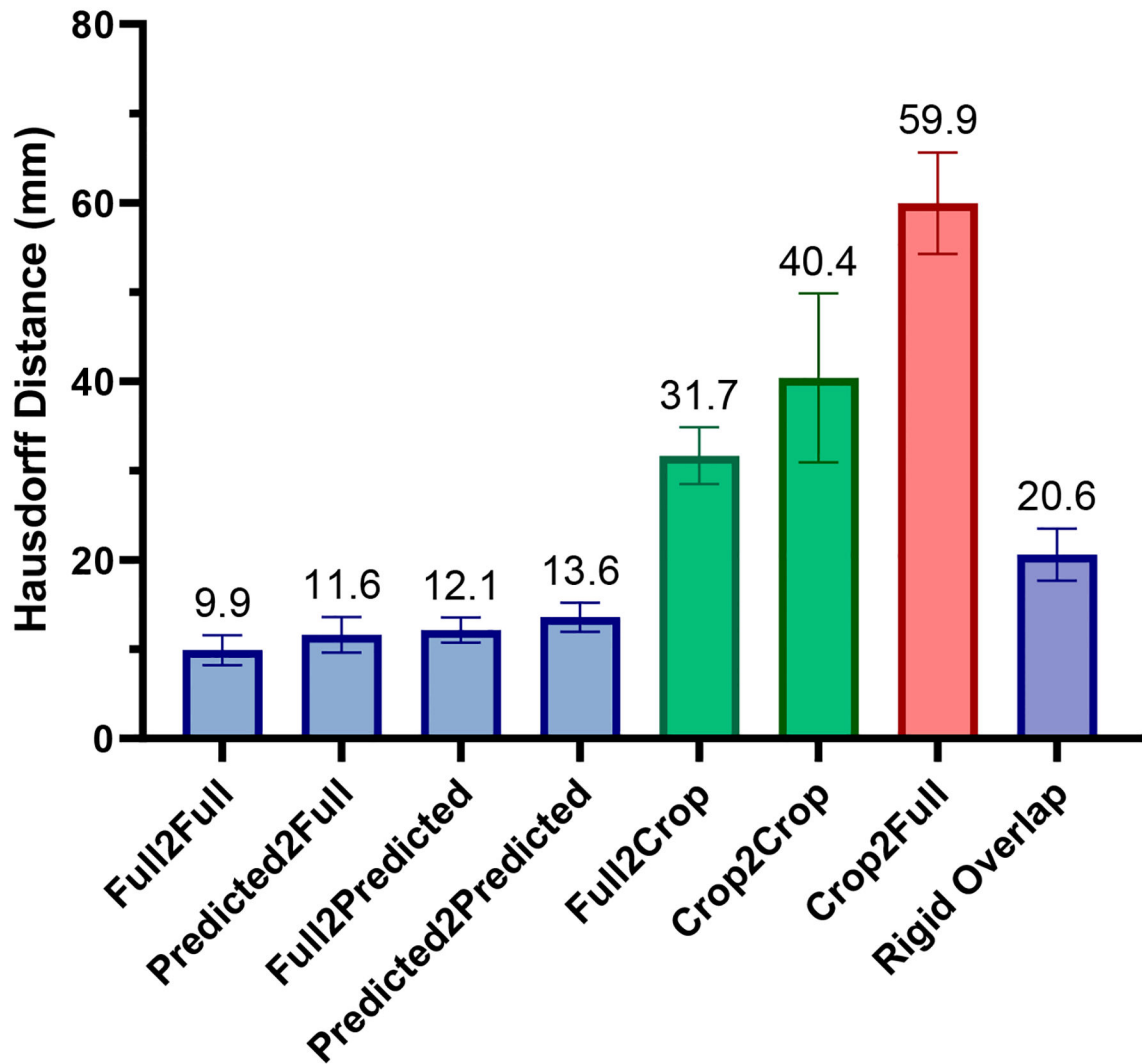
Figure 5.
The average 95% Hausdorff Distance between deformed and target contours for each registration pair, averaged across all 16 contours, and all 53 test patients. The best-case registration is "Full2Full" (leftmost bar), while registrations using our method are shown in the next 3 bars. These 4 leftmost bars are not significantly different (represented by being the same color blue). Full2Crop and Crop2Crop were not significantly different, while Crop2Full was the highest of all. While the difference between our method and rigid overlap did not reach significance, we did see a near halving of the error (20.6mm average error down to ~12mm error). The error bars are the 95% confidence interval.
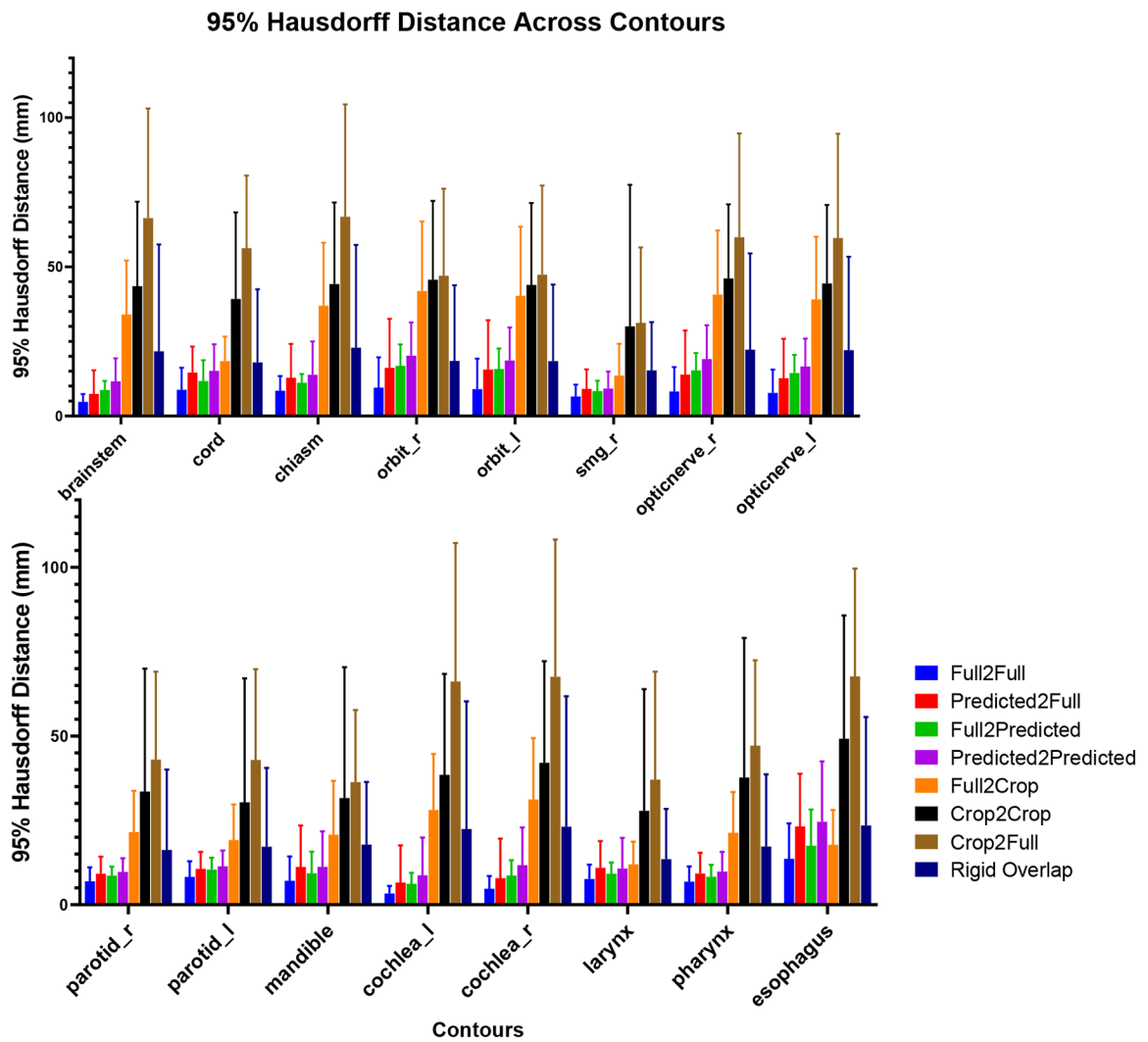
95% Hausdorff Distance Across Contours

Figure 6.

shows the average registration error broken down by each contour used in this study. The error bars are standard deviation. The horizontal axis has been split into two levels to improve readability.
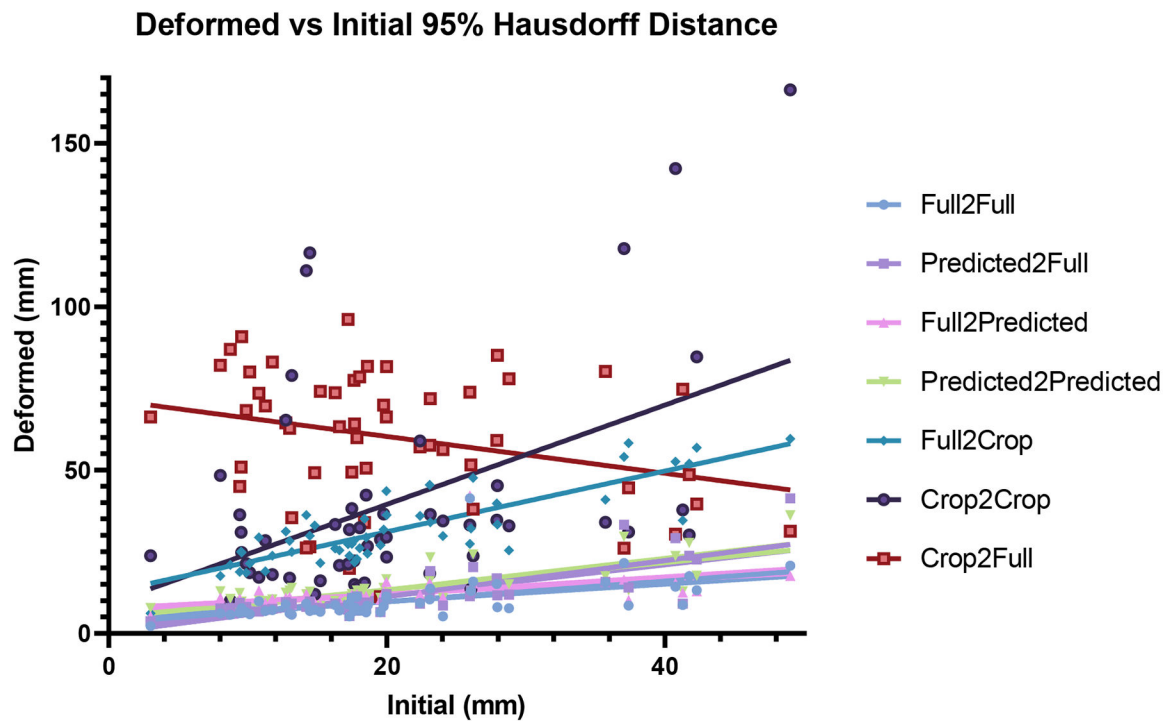
Figure 7.
demonstrates how the registration error varies as a function of initial rigidly aligned 95% Hausdorff distance. The intial overlap value is a measure of the difficulty of the registration. The full-to-full registrations and the ones using our proposed CropGAN technique all have relatively flat lines, showing their robustness to the registration's initial conditions. Best viewed in color (online version).
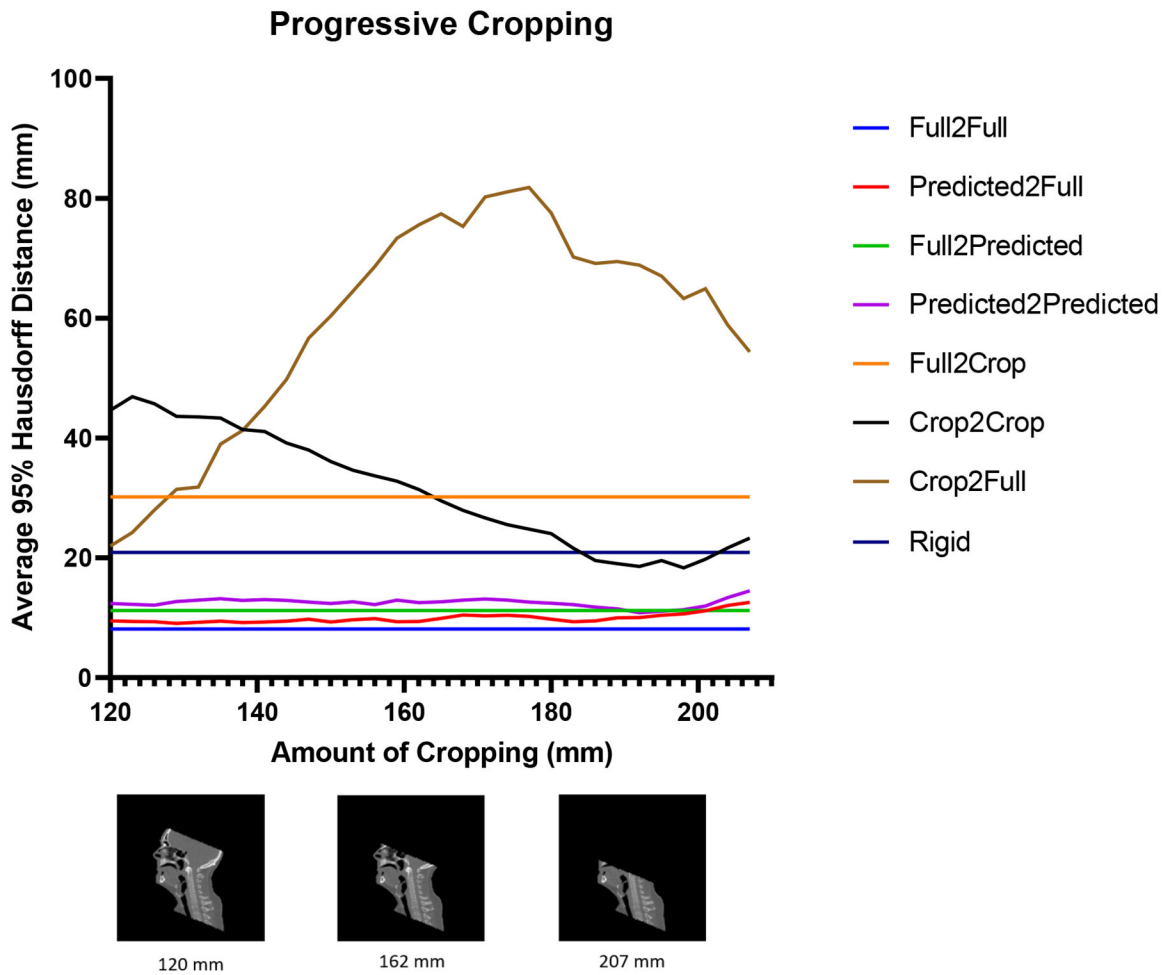
Progressive Cropping

Figure 8.

shows the registration error as a function of missing tissue averaged across three patients. The images below the plot help to visualize a given amount of cropping. We see that for our method (purple and red lines) the amount of superior cropping does not have much of an effect before 18.7cm, where nearly all the superior half of the tissue is missing. This demonstrates the robustness of our technique. Our method is also superior to rigid alignment for all cropping amounts investigated. When the moving image is left cropped (brown and black lines), the registration quality varies wildly.
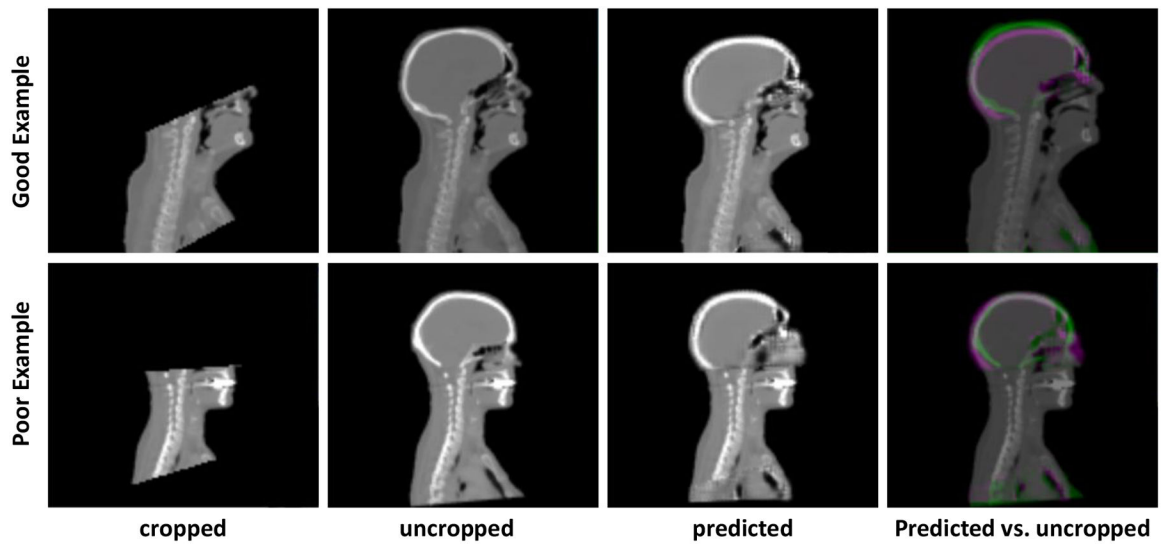
Figure 9.
An example showing a good (top) and poor (bottom) predicted completion of a cropped image. The first column shows the cropped image, the second column shows the uncropped ground truth, and the third column shows the predicted result from our network. A difference image is shown in the right-most column (best seen in color). The poor prediction occurs near an artefact in the mouth.