Whole-Body PET Registration for Multiple Visits using Weakly-Supervised Learning

Author(s) names withheld

Address withheld

Editors: Under Review for MIDL 2020

Abstract

We aim to reduce the tedious nature of developing and evaluating methods for aligning PET-CT scans from multiple patient visits. To reduce the evaluation effort, instead of 3D annotations that are difficult to create, we propose to label multiple key points across several 2D slices, to which we then fit a key curve. We use these key curves to define an error metric for the alignments that can be computed efficiently. To further reduce human effort, we propose a novel deep learning-based method that is trained with weak supervision and validated using our key curves. We show the potential of our method both qualitatively and quantitatively, and that our weak supervision improves over self-supervision.

Keywords: PET-CT Registration, Weakly-Supervised Learning, 3D Annotation

1. Introduction

Aligning PET-CT scans accurately is a crucial step in disease assessment and treatment planning. Physicians need to be able to compare scans from one visit to another to determine how the disease has progressed. However, in practice – even with care – the initial alignments are never perfect. Therefore, semi-automatic registration methods (Nelson, 2018; Avants et al., 2008) have been developed to correct this misalignment and help doctors assess the extent and progression of the disease. More recently, there have been efforts (Balakrishnan et al., 2019) to make use of the recent progress in deep learning for this task. Still, in practice, we rely mostly on physicians to align different scans, as the automated methods are not accurate enough.

A potential reason behind the slow pace of development compared to other applications that use deep learning (He et al., 2015), is that creating labels in 3D is very difficult. For example, creating “ground truths” for PET-CT alignments requires providing 3D correspondences. This could be in the form of an alignment given by an expert through a graphical interface tool that manipulates data in 3D, or perhaps by a few 3D key point correspondences (again, chosen manually). In both cases, a clinical expert needs to interact with the data in 3D, which is cumbersome and error prone as our display devices are still in 2D. Without a substantial amount of these ground truth alignments, it becomes difficult to adopt traditional deep learning-based methods, and – more importantly – evaluate the alignments effectively.

In this short paper, we therefore propose (1) a 2D-based 3D annotation method based on key-curves and show how it could be used, and (2) suggest a novel weakly-supervised deep method that overcome these limitations.
2. Key-curves as geometric representations

We propose to use a 3D curve to represent the geometric configuration of each scan. Given enough corresponding 3D curves between two visits from the same patient, we measure the misalignment between the two scans by measuring the distance between these curves. In more detail, we represent curves in 3D as a second order polynomial that outputs \(x\) and \(y\) coordinates given a \(z\) coordinate. We then compute the distance between the two curves by computing the distance for selected \(z\) slices and averaging them.

To obtain key curves we rely on 2D annotations. As shown in the example in Fig. 1 (left and center), one needs to only label matching points in 2D. Note that the two scans do not have to correspond to the exact same \(z\) slice since with curves – unless the curve is perfectly straight, which in practice does not happen – the average distance will only be zero if and only if the scans are perfectly aligned. To facilitate easy labeling, we have further developed a simple interactive software with a graphical user interface.

**Incorporating uncertainty.** All labeling methods inherently create uncertainty, which we incorporate for visual inspection; see Fig. 2. We estimate the uncertainty in the key point selection by assessing the variance of our annotations for each point. The average selection uncertainty for our dataset is 2.52 mm and 1.96 mm in the \(x\) and \(y\) direction, respectively.

Another source of uncertainty comes from the fact that not all \(z\) slices have key points to select (for example, the bladder does not extend to the back bone). As in (Wolberg, 2006), we model the confidence on the curves to be poorer in these regions where we extrapolate, which can be seen in Fig. 2.

3. Novel method for PET-CT image registration

To further reduce human effort, we propose a novel weakly supervised method that utilizes the proposed metric for validation. In detail, instead of depending on geometric labels, we rely on synthetic transformations of scans (i.e. self-supervision) as in (Rocco et al., 2017), but extend the method to 3D from the original 2D setup. After applying synthetic transformations, we then train a deep network to recover the transformation parameters by just looking at the scans before and after the transformation. We use both PET and CT scans as the input to our framework by concatenating the two as a multi-channel 3D image. The dynamic range of PET scans can differ greatly from one scan to another, and we therefore pre-process our scans by computing the gradient of the log of the PET images.

![Figure 1: An example of the selection of a key point for different visits of the same patient. We show PET scans of one scan overlayed on top of CT scans of another visit.](image1)

![Figure 2: Example key points and the fit curves projected to \(x\) and \(y\) axes. The uncertainty is shown as error bars for two different scans – source and target – and the source aligned to the target.](image2)
Architecture-wise, we first apply 9 consecutive 3D ResNet blocks (He et al., 2016) to the pre-processed scans to extract features from both the “source” and “target” (transformed) scans. We then create a similarity matrix, as in (Rocco et al., 2017), by flattening the feature maps of both our source and target scans, then performing a cross product between the two. This similarity matrix is then provided as an input to a successive 3D CNN that regresses to the transformation parameters. As in (Rocco et al., 2017), we first apply an affine transformation for the initial alignment, then a thin plate spline (TPS) transformation, which accounts for the remaining – more subtle – misalignments. We train these two steps separately, which we empirically found to be important.

The self-supervised approach alone does not provide robustness, as each training sample consists only of a single PET-CT scan along with the transformed version of that image. We therefore enforce consistency among the feature representations (created by the deep network) across different PET-CT scans of the same patient – we do so by maximizing the recently proposed linear centered kernel alignment (LCKA) (Kornblith et al., 2019) between the feature maps. The LCKA encodes the similarity of deep feature responses, while ignoring their geometrical configuration. Lastly, we monitor the convergence of training through our key-curve metric.

We apply our method to the QIN-HEADNECK dataset (Beichel et al., 2015; Fedorov et al., 2016; Clark et al., 2013) of PET-CT images, which consists of 651 studies on 156 patients. We selected 296 patient scans that are of reasonable size (larger than 120×120×240 pixels at a resolution of 3.5mm). We further split this sample into training (267), validation (15), and test (14) sets based on the patient ID. We emphasize here that we only label the validation and test pairs with our key curves, and for training, we strictly rely only on the pairwise relationship and no geometrical supervision. See Fig. 3 for a qualitative example.

In Table 1, we apply the key curve metric to a variety of different visit pairs in our test set, and report the summary of how the proposed registration method performs. Without our weak supervision – denoted as w/o LCKA, which is a 3D extension of (Rocco et al., 2017) – the method is not able to learn to be invariant across visits and provides worse results. This demonstrates the importance and effectiveness of our method.

<table>
<thead>
<tr>
<th>Unaligned</th>
<th>w/o LCKA</th>
<th>Our Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (mm)</td>
<td>29.25</td>
<td>15.89</td>
</tr>
</tbody>
</table>

4. Conclusion
We have proposed (1) an annotation method and (2) a novel weakly-supervised method that improves upon self-supervision to reduce human effort in aligning PET-CT scans of patients from different visits. As future work, we plan to perform comparisons with CT-only intensity based alignments as well as other existing methods for aligning PET-CT scans.
Acknowledgments

Acknowledgments withheld.

References


