Approximate Vertebral Body Instance Segmentation by PET-CT Fusion for Assessment After Hematopoietic Stem Cell Transplantation

Brandon D. Carson[1], Favio Hurtado[2], Joseph P. Havlicek[3], Lucas J. Powers[3], Liza Lindenberg[4], Daniele N. Avila[4], Christopher G. Kanakry[4], Peter Choyke[4], Karen Kurdziel[4], Philip Eclarinal[4], Kirsten M. Williams[5], and Jennifer Holter Chakrabarty[6]

***Abstract*—We introduce a new, fully automatic vertebral instance segmentation method to facilitate the extraction of standard uptake values (SUV) from the medullary cavities of individual vertebral bodies in joint 18 F-fluorothymidine (FLT) PET/CT scans acquired from hematopoietic stem cell transplantation (HSCT) patients at 28 days after transplant. Due to dosing considerations, the CT voxels in these scans are characterized by a large 5 mm axial slice thickness which significantly complicates the vertebral body segmentation problem. The key ideas of our method are to first apply an ensemble of U-Nets to obtain a binary mask for the aggregated collection of vertebral bodies as a single object without estimating the intervertebral boundaries, and then leverage the relatively better 4 mm axial slice spacing in the PET data to estimate a "best fit" axial coordinate to approximate the break between each pair of vertebrae. This PET-CT fusion approach results in an approximate vertebral body segmentation where each estimated intervertebral boundary is, by construction, restricted to lie in a single axial plane. However, because the FLT uptake is well localized within the medullary cavities, our results show that this approximate segmentation is sufficiently accurate to enable FLT SUV data to be isolated for individual vertebral bodies. Compared to traditional methods for assessing engraftment based on single aspirate biopsies, this new technique has potential to facilitate a significantly more comprehensive assessment of the medullary compartment by providing fully automated SUV data for a plurality of individual bones.**

***Index Terms*—vertebra segmentation, PET/CT, HSCT, FLT**

I. Introduction

Hematopoietic stem cell transplantation (HSCT) is used in the treatment of life threatening malignancies of the bone marrow and blood. Post-transplant monitoring is critical for assessing engraftment versus graft failure, predicting relapse, and modulating therapy when needed. Assessment has traditionally been based on single aspirate biopsies. However, this traditional approach fails to provide a comprehensive characterization of the marrow compartment and introduces significant risk of potentially fatal infections [1]. Noninvasive monitoring techniques based on ^{18}F-fluorothymidine (FLT) PET/CT imaging have emerged more recently and can provide a substantially more comprehensive characterization of the full-body marrow compartment [1]–[6]. But this generally requires manual designation of a large number of regions of interest (ROIs) by expert physicians which is both time consuming and labor intensive. Development of an automated procedure for reliably measuring FLT standard uptake values (SUV) in the medullary cavities of individual bones could resolve this issue and facilitate translation of the newer noninvasive monitoring techniques to widespread clinical practice. However, improved fully automatic 3D bone segmentation algorithms are needed as a prerequisite for this to occur.

Here, we focus on fully automated vertebral body instance segmentation as one step towards this goal. The problem of vertebrae segmentation from CT data alone has been widely studied [7]–[12]. State-of-the-art performance has generally been established on good quality CT data with in-plane resolution on the order of 0.35 mm to 1.00 mm and slice thickness also on the order of 1.00 mm [8]. Yet, while Dice scores above 90% have been reported [9], [11], many existing methods still produce relatively less accurate boundary segmentation results [12]. Furthermore, comparatively fewer studies have been reported for the case of low-dose CT. Low dose is an important consideration in post-HSCT assessment due to the inherent vulnerability of the patients after myeloablation [6] and, in general, vertebral segmentation is significantly more difficult in the low-dose case.

For example, Lessmann et al. [8] performed vertebra segmentation on dedicated spine CT scans having in-plane resolution varying from 0.31 mm to 0.36 mm with slice thickness ranging between 0.7 mm and 1.0 mm and obtained a Dice score of 95.8%. By comparison, they obtained a Dice score of 92.1% on low-dose chest CT scans with in-plane resolution between 0.54 mm and 0.82 mm and slice thickness between 1.0 mm and 2.5 mm. Due to dose considerations, the HSCT patient CT data we consider here were acquired at a substantially larger voxel resolution of 1.17 mm $\times$ 1.17 mm $\times$ 5.00 mm, making the vertebral body instance segmentation

This work was supported by the National Cancer Institute, NIH, under contract number HHSN261200800001E and grant number 5R01HL146668.

[1]B.D. Carson is with Camgian, Starkville, MS 37206, USA.

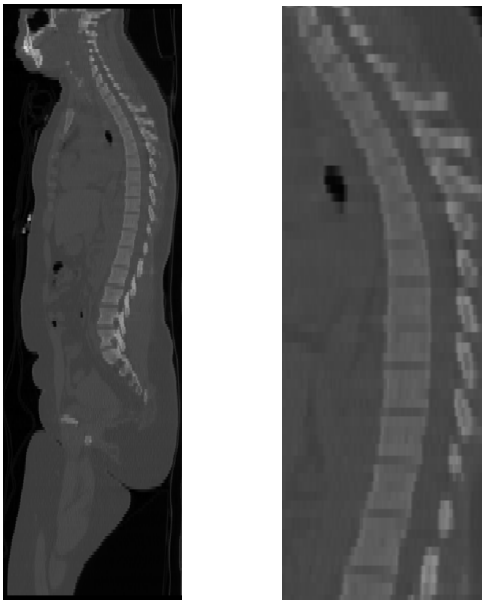
[2]F. Hurtado is with HBE Systems, LLC, Norman, OK 73071, USA.

[3]J.P. Havlicek and L.J. Powers are with the University of Oklahoma, School of Electrical & Computer Engineering, Norman, OK 73019, USA.

[4]L. Lindberg, D.N. Avila, C.G. Kanakry, P. Choyke, K. Kurdziel, and P. Eclarinal are with the National Cancer Institute and National Institutes of Health, Bethesda, MD 20892, USA.

[5]K.M. Williams is with the Dept. of Pediatrics, Emory University, Children's Healthcare of Atlanta, and the AFLAC Cancer and Blood Disorder Center, Atlanta, GA 30322, USA.

[6]J. Holter Chakrabarty is with the Division of Marrow Transplantation and Cell Therapy, Stephenson Cancer Center, Oklahoma City, OK 73104, USA.



(a) (b)

Fig. 1. Typical CT sagittal slice. (a) Patient 1 full sagittal slice; (b) detail view illustrating poorly defined intervertebral boundaries.

problem challenging.

In this paper, we present a new method for fully automatic vertebral body instance segmentation applicable to low-dose FLT PET/CT scans of HSCT patients acquired as described in [6]. The test set consisted of 13 patients out of a total cohort of 23. Each patient was imaged on the day before transplant, between five and 12 days post transplant, and on the 28th day post transplant. As mentioned above, the CT data were acquired with a voxel resolution of 1.17 mm $\times$ 1.17 mm $\times$ 5.00 mm due to dose considerations. A typical example of the CT data is shown in the sagittal slice of Fig. 1, where poorly defined intervertebral boundaries resulting from the 5.00 mm axial resolution can be seen in the detail view in Fig. 1(b). The corresponding PET data were acquired with a resolution of 4 mm $\times$ 4 mm $\times$ 4 mm.

This data set presents several challenges including the small cohort size, which severely limits the quantity of training data available, as well as the coarse CT axial resolution – resulting in the somewhat unusual circumstance that the PET volumes actually provide better axial resolution as compared to the CT volumes. The key idea behind our method is to apply a novel, computationally efficient algorithm for integrating the FLT SUV data within 2D axial regions obtained by intersecting each PET axial plane with a 3D mask of the vertebral column obtained by processing the CT data, thereby leveraging a data fusion approach to exploit both the higher in-plane resolution provided by the CT data and the higher axial resolution provided by the PET data. This results in an *approximate* vertebral body instance segmentation where each intervertebral boundary is approximated by a "best-fit" axial plane as shown below in Fig. 4(b). Consequently, our segmentation results are not directly comparable to those reported in, e.g., [8]–[12] because our method inherently restricts each approximate intervertebral boundary to lie entirely within a single axial plane. However, as shown below in Fig. 4(c)-(f), the FLT SUV PET signature is typically well localized within the marrow cavities; as a result our approximate vertebral body instance segmentation is sufficiently accurate to enable per-vertebra extraction of SUV measurements within the marrow cavities, which is the clinically significant observation that is needed for more comprehensive assessment of these HSCT patients.

II. Methods

In this section, we describe the major stages of our approximate vertebral body instance segmentation method, which include semantic segmentation of the CT volumes, postprocessing of the resulting vertebral column binary mask, resampling and masking of the FLT PET data, and intervertebral boundary approximation.

More specifically, we train an ensemble of convolutional neural networks (CNNs) to perform semantic segmentation on the CT volumes, resulting in a binary mask for the vertebral column as a single monolithic object. This approach avoids the need to estimate the locations of the intervertebral boundaries from the CT data, thereby circumventing the problems associated with CT dosing considerations and the relatively large concomitant 5 mm CT axial voxel size. After 3D postprocessing, we use the segmented binary vertebral column object to mask the FLT PET modality axial slices where the slice spacing is 4 mm (as opposed to 5 mm in the CT data). We then approximate each intervertebral boundary with a "best-fit" axial plane by filtering the PET data using a sliding window filter with an iteratively adjusted window size that is seeded with an automatically determined anatomical measurement prior. Using these detected approximate boundaries, we then obtain the final approximate vertebral body instance segmentation that may be used to extract FLT SUV measurements from the marrow cavity of each vertebral body. As we discuss later in Section IV, this method is effective when applied to patient scans acquired at 28 days after transplant but is not expected to work as well at earlier observation points when the FLT PET signal is at a lower level.

A. Semantic Segmentation

To obtain the initial semantic segmentation of the vertebral column we implement a convolutional neural network model consisting of three independently trained U-Nets, slightly modified from the original U-Net architecture [13] by the addition of a batch normalization layer before each ReLU activation. Following a "pseudo-3D" configuration similar to that implemented in [9], each of the three U-Net models was trained on a unique anatomical view corresponding to one of the axial, sagittal, or coronal anatomical planes. During inference, 3D prediction volumes were obtained by providing each constituent U-Net with a sequence of images from the the anatomical plane on which it was trained. After the inference

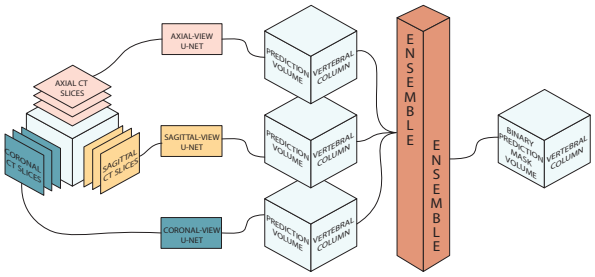


Fig. 2. Block diagram of vertebral body semantic segmentation. Coronal, sagittal, and axial slices from the CT data volume are input, respectively, to an ensemble of three U-Nets trained on coronal, sagittal, and axial slices with manually generated ground truth. Each U-Net produces a 3D prediction volume of confidence scores for the vertebral body class. The 3D binary prediction mask for the vertebral body class may then obtained by averaging the individual U-Net prediction volumes or, alternatively, by considering a single one of them, followed by thresholding at the naïve value of 0.5.

was completed, vertebral class confidence scores for each U-Net output were averaged voxel-wise and the results were thresholded at p = 0.5 to obtain a binary segmentation mask for the vertebral column. A block diagram illustrating this approach is shown in Fig. 2.

Ground truth 3D segmentation masks were manually annotated with four class labels (vertebral body, sternum, pelvis, and background) at the native resolution of the CT input data. The training set was composed of 14 CT image volumes from eight patients and two ground truth volumes were reserved for validation. The CT and ground truth mask volumes were sliced along the axial, sagittal, or coronal plane prior to training. Geometric data augmentations were used to increase the diversity of the training dataset. Input images and their target masks were scaled from -20% to $+20\%$, random rotation was applied from $-45°$ to $+45°$, horizontal and vertical translations were applied up to 20% of the image height and width, and horizontal and vertical flipping were applied. Training images were cropped to a size of 320 $\times$ 320 pixels, with mirror-image padding to account for missing data that might arise from the geometric transformations. All augmentations, excepting the final crop, were performed randomly with a uniform distribution.

The training dataset is unavoidably class-imbalanced, with the background class dominating 99.5% of all voxels. To mitigate the effect of the severely imbalanced dataset on model convergence during training, multi-class weighted cross entropy was used as the loss function, with the background class down-weighted to contribute only 10% of the loss, while the three object classes (vertebral body, sternum, and pelvis) each contributed 30% to the loss. The Adam optimizer [14] (with default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and an initial learning rate of 10^{-4}) was used to update the learned parameters during training. Models were trained over 32 epochs with each model seeing every 2D slice in the training set in each epoch for a total of 2,579 training pairs per epoch in the axial-view model and 7,168 input training pairs per epoch in the sagittal- and coronal-view models. A minibatch size of six was used. Validation was performed at regular intervals to track convergence, and the learning rate was reduced by an order of magnitude when the validation score failed to reach a new maximum within eight epochs.

B. Semantic Segmentation Post Processing

Intervertebral boundaries are typically poorly detected in the binary vertebral column mask obtained by semantic segmentation due to the coarse axial resolution of the CT scans. To overcome this problem, we fill in the gaps between adjacent vertebrae, thus obtaining a contiguous mask to facilitate approximation of the intervertebral boundaries from the FLT PET data. We accomplish this by applying a morphological closing with a small spherical structuring element of radius one followed by connected components labeling to select and extract the largest contiguous object in the morphologically closed CT mask volume. A typical resulting binary mask for the vertebral body class as a single object is shown below in Fig. 4(a).

C. Intervertebral Boundary Detection

We perform intervertebral boundary detection in the FLT PET modality due to the finer axial resolution it provides relative to the CT data in these HSCT patient scans. Furthermore, the FLT uptake is generally well localized in the medullary cavities of the vertebral bodies, resulting in clear gaps between adjacent vertebrae in the FLT PET volumes even when the boundaries are difficult or impossible for a human observer to localize in the CT modality alone, thus motivating our PET-CT fusion approach.

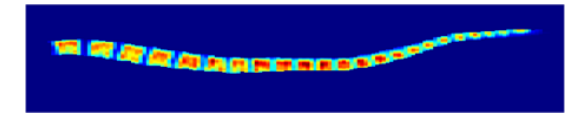
We apply bicubic interpolation to resample each axial PET slice to match the in-plane axial slice resolution of the CT data. We then interpolate the axial resolution of the CT binary mask to match that of the PET data using a global axial coordinate provided by the joint PET/CT scanner for every axial slice in both modalities. Pointwise multiplication then produces a masked FLT PET volume with a voxel resolution of 1.17 mm $\times$ 1.17 mm $\times$ 4.00 mm which we use to segment the individual vertebral bodies from the vertebral column as described below. A typical sagittal slice of the resampled and masked FLT PET volume is shown in Fig. 3(a).

The resampled and masked FLT PET data is then averaged in each axial slice to obtain a 1D signal

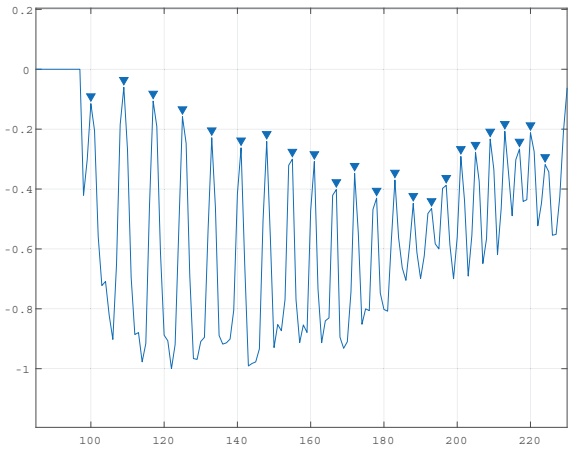
$$I_{\text{avg}}(i) = \frac{1}{N_i} \sum_{j,k} P_{r,m}(i,j,k), \tag{1}$$

where i, j, and k are the axial, coronal, and sagittal coordinates, $P_{r,m}$ is the resampled and masked FLT PET volume, and N_i is the number of vertebral body class pixels in the i'th axial slice of $P_{r,m}$ (which is given by the number of 1's in the i'th slice of the interpolated mask). We then compute a normalized 1D signal

$$\widehat{I}_{\text{avg}}(i) = \frac{-I_{\text{avg}}(i)}{\max_m [I_{\text{avg}}(m)]}, \tag{2}$$



(a)



(b)

Fig. 3. Intervertebral boundary approximation. (a) Sagittal slice of resampled FLT-PET data showing vertebral bodies isolated by applying the binary mask obtained from semantic segmentation and post processing of the CT data; (b) the normalized 1D signal $\widehat{I}_{\text{avg}}(i)$ used in detecting a "best-fit" axial plane approximation for each intervertebral boundary. Detected boundaries for the masked sagittal PET slice in (a) are shown with blue triangles in (b), although it should be noted that parts (a) and (b) are not shown to scale.

where inversion facilitates detection of the intervertebral boundaries as local peaks in $\widehat{I}_{\text{avg}}(i)$ and scaling by $\max_m[I_{\text{avg}}(m)]$ normalizes variability between patients.

As shown in Fig. 3(b), the 1D signal $\widehat{I}_{\text{avg}}(i)$ typically contains spurious local maxima which preclude the use of simple peak detectors for finding the intervertebral boundaries. Instead, we implement a sliding window filter and adjust the window size iteratively as vertebral boundaries are detected along the vertebral column. The window size is initially seeded with the axial span of the L5 vertebral body, denoted d_{L5}, which is computed at run time for each patient without *a priori* knowledge by leveraging the consistency with which the first few peaks appear in the normalized signal $\widehat{I}_{\text{avg}}(i)$. We take i_{L5} to be the index of the first nonzero value in $\widehat{I}_{\text{avg}}(i)$, which corresponds to the axial coordinate of the beginning of the segmentation mask and thus the inferior boundary of the L5 vertebra. A generic peak detector (e.g., Matlab `findpeaks` with minimum peak height of -0.4 and minimum prominence of 0.1) can then reliably detect the boundary i_{L4} between the L5 and L4 vertebrae and we initialize the L5 vertebral span prior according to $d_{\text{L5}} = i_{\text{L4}} - i_{\text{L5}}$.

Based on the detected index i_{L4} of the boundary between the L5 and L4 vertebrae, a window is constructed from which the next boundary will be detected. To avoid multiple detections of the previous boundary, the left edge of the window is set at $i_{\text{last}} + 3$, where i_{last} is the index of the last detected boundary in $\widehat{I}_{\text{avg}}(i)$. The right edge of the window is set at $i_{\text{last}}+\text{round}(1.2d_{\text{last}})$, where d_{last} is the last calculated vertebral span. The index of the next boundary detection is then given by

$$i_{\text{next}} = \arg\max_{i \in W} \widehat{I}_{\text{avg}}(i), \tag{3}$$

where W is the current window with left and right edges set based on the last detected intervertebral boundary as just described. The detection process terminates at the C2 vertebra, excluding C1 due to challenges with the CT-derived segmentation mask in that region for many of the patients in the data set. We use the detected boundary locations, which are illustrated in Fig. 3(b) for a typical case, as "best-fit" axial plane coordinates for approximating the intervertebral boundaries and use them to segment the individual vertebral bodies from the interpolated binary segmentation mask.

III. Results

Testing was performed on 13 scans that were acquired at 28 days post transplant from 13 patients and were not used for training or validation. Semantic segmentation was applied as shown in Fig. 2 and described in Section II-A to label CT voxels belonging to the vertebral body class. Dice coefficient results for this task are reported in Table I. The constituent axial, sagittal, and coronal U-Net models were scored individually and in the ensemble configuration. The scores were obtained by using a classification threshold of $p = 0.5$ on the 13 raw prediction volumes of the test set and then comparing the resulting binary prediction volumes for the vertebral body object class to the manually labeled ground truth. In each row of Table I, the best result is shown in bold and the second best is shown with underline.

Of the three constituent U-Net models, the axial-trained model performed best on the test data with a mean Dice score of 0.9222 on the vertebral body class. This result was unsurprising in view of the fact that the 5 mm axial slice thickness of the CT data was expected to have a significant detrimental effect on performance of the sagittal and coronal models. Nevertheless, the sagittal model still performed reasonably, achieving an average Dice score of 0.9017, while the coronal model performed poorly relative to the other two, achieving an average Dice score of only 0.8489. Also due in part to the poor performance of the coronal model, the ensemble method achieved a reduced performance relative to the axial model alone, scoring an average Dice coefficient of 0.9139. We attribute the fact that our pseudo-3D ensemble method failed to achieve significant performance gains of the type reported in [9] to two main factors: first, the thick 5 mm slices in our low-dose CT data result in relatively poor resolution in the sagittal and coronal views; second, we omitted implementation of the "variational difference image" used as a second input channel in [9]. Although the numerical Dice scores reported here in Table I compare reasonably with those reported for the VerSe challenge in [11], it must be kept in mind that these results are not directly comparable due to the fundamental

TABLE I
DICE SCORE RESULTS FOR VERTEBRAL BODY OBJECT CLASS SEMANTIC SEGMENTATION TASK RELATIVE TO MANUALLY LABELED GROUND TRUTH. RESULTS SHOWN FOR PATIENTS IN THE TEST SET ONLY (PATIENTS 1-6, 15, 20, 22, AND 23 WERE USED FOR TRAINING AND VALIDATION IN THIS TEST).

	U-Net Model			
Patient	Axial	Sagittal	Coronal	Ensemble (Avg)
7	**0.9012**	0.8836	0.7620	0.8967
8	**0.9180**	0.8725	0.7576	0.8924
9	**0.9342**	0.9094	0.9033	0.9303
10	**0.9297**	0.9113	0.8899	0.9255
11	**0.9380**	0.8925	0.8575	0.9207
12	**0.9088**	0.9074	0.7890	0.9047
13	**0.9198**	0.9003	0.8696	0.9097
14	**0.9179**	0.9046	0.8579	0.9075
16	**0.9381**	0.9212	0.8910	0.9311
17	**0.9019**	0.8799	0.8692	0.8952
18	**0.9200**	0.8962	0.8223	0.9082
19	**0.9350**	0.9171	0.8988	0.9323
21	0.9258	0.9261	0.8674	**0.9263**
mean	**0.9222**	0.9017	0.8489	0.9139

differences in the evaluated task. Here, the Dice coefficients reported in Table I are for segmentation of the vertebral body class for the entire spine as a single monolithic object.

Approximate intervertebral boundary detection experiments were conducted using our method described in Section II-C and depicted in Fig. 3. For ground truth, an experienced observer manually selected a single "best-fit" axial plane to approximate the break between each pair of vertebral bodies at the native resolution of the PET scans. In view of the semantic segmentation results presented in Table I, we used the vertebral column binary mask obtained from the axial-trained U-Net model alone rather than the one obtained from the pseudo-3D U-Net ensemble. Mean absolute errors for the detected boundaries are reported in units of voxels and millimeters in Table II for each patient in the test set, averaged over intervertebral boundaries, and in Table III for each intervertebral boundary, averaged over patients in the test set. In Table III, the column labeled "L5 Inf" gives results for the inferior boundary of the L5 vertebra, whereas the rest of the entries are for the superior boundaries of the indicated vertebrae.

While these results show that the algorithm described in Section II-C achieved precise agreement with the manually selected ground truth "best-fit" axial boundary plane for Patients 14, 17, and 18 and for several of the intervertebral boundaries, the L5 inferior boundary error is attributed to inaccuracies in the CNN-derived segmentation mask at the edge of the vertebral column. The C2 boundary error is attributed to the difficulty of filtering the averaged PET intensity signal in that region of the spine; the expected vertebral span is short (making the window harder to define) and the PET response is generally less intense resulting in a lower signal to noise ratio in this region. However, compared to the Kalman filter used on the same dataset in [7], our sliding-window algorithm presented here is able to detect the vertebral boundaries more accurately.

Overall results for the approximate vertebral body instance segmentation method we have proposed in this paper are given in Fig. 4. The vertebral body class binary segmentation mask obtained using the method described in Section II-A with post processing as described in Section II-B is shown in Fig. 4(a). This result segments the vertebral body class as a single object without attempting to estimate the boundaries between individual vertebrae. Figure 4(b) shows the final result for approximate segmentation of the individual vertebral bodies obtained by applying the boundary detection algorithm of Section II-C. This result is *approximate* in the sense that each estimated intervertebral boundary is inherently restricted to lie in a single axial plane. Finally, Figs. 4(c)-(f) show the clinically important FLT SUV measurements overlaid on the approximate individual vertebral body segmentations obtained for four patients. From the results shown in Figs. 4(c)-(f), it may be seen that the approximate segmentation method presented here is sufficiently accurate to enable extraction of SUV measurements from the medullary cavities of the individual vertebral bodies of this data set, thereby achieving our overall signal processing goal.

IV. DISCUSSION AND LIMITATIONS

Hematopoietic stem cell transplantation (HSCT) is an important procedure in the treatment of certain bone marrow and blood malignancies including leukemia and multiple myeloma. Post-transplant monitoring is critical for assessing engraftment, predicting relapse, and planning therapy modulations in these patients and has traditionally been performed via single aspirate biopsies drawn, e.g., from the pelvis. Recent research suggests that identification of transplant complications can be improved through joint FLT PET/CT imaging (see, e.g., [6], [15]). However, this new approach is both labor intensive due to the need to designate and evaluate many regions of interest within the joint volumetric data, making automation highly desirable, and challenging due to the need for low-dose CT imaging to accommodate the inherent vulnerability of HSCT patients. In this paper, we presented an approximate vertebral body instance segmentation method capable of automatically extracting FLT SUV measurements from the individual vertebral bodies by leveraging the PET data to overcome poor axial resolution characteristic of the low-dose CT scans in our HSCT patient data set. The key idea of our method is to detect approximate intervertebral boundaries by integrating the PET data within axial slices. This method is *approximate* in the sense that each detected boundary is restricted by construction to lie in a single axial plane. Nevertheless, as demonstrated in Figs. 4(c)-(f), our method is accurate enough to facilitate extraction of FLT SUV measurements from the individual vertebral bodies in the HSCT patient scans acquired 28 days after transplant that were considered here.

In our ongoing research, we are seeking improved techniques to solve three main limitations of the method proposed in this paper. First, the boundary detection method described in Section II-C relies on the FLT PET data capturing a strong

TABLE II

MEAN ABSOLUTE ERROR IN DETECTED VERTEBRAL BOUNDARIES BY PATIENT. ERROR IS DEFINED AS THE ABSOLUTE DIFFERENCE IN AXIAL COORDINATES BETWEEN THE BEST-FIT AXIAL BOUNDARY PLANES DETECTED BY OUR PROPOSED METHOD AND DESIGNATED BY AN EXPERIENCED OBSERVER. ERRORS ARE AVERAGED OVER VERTEBRAL BOUNDARIES FOR EACH PATIENT IN THE TEST SET. PATIENTS 1-6, 15, 20, 22, AND 23 WERE USED FOR TRAINING AND VALIDATION IN THIS TEST.

Patient	7	8	9	10	11	12	13	14	16	17	18	19	21
MAE, voxels	0.000	0.583	0.042	0.083	0.083	0.167	0.042	0.000	0.125	0.000	0.000	0.125	0.042
MAE, mm	0.000	2.332	0.168	0.332	0.332	0.668	0.168	0.000	0.500	0.000	0.000	0.500	0.168

TABLE III

MEAN ABSOLUTE ERROR IN DETECTED VERTEBRAL BOUNDARIES BY VERTEBRAE. ERROR IS DEFINED AS THE ABSOLUTE DIFFERENCE IN AXIAL COORDINATES BETWEEN THE BEST-FIT AXIAL BOUNDARY PLANES DETECTED BY OUR PROPOSED METHOD AND DESIGNATED BY AN EXPERIENCED OBSERVER. ERRORS ARE AVERAGED OVER 13 PATIENTS IN THE TEST SET FOR EACH VERTEBRAL BOUNDARY.

Boundary	L5 inf	L5 sup	L4 sup	L3 sup	L2 sup	L1 sup	T12 sup	T11 sup	T10 sup	T9 sup	T8 sup	T7 sup
MAE, voxels	0.682	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MAE, mm	2.728	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Boundary	T6 sup	T5 sup	T4 sup	T3 sup	T2 sup	T1 sup	C7 sup	C6 sup	C5 sup	C4 sup	C3 sup	C2 sup
MAE, voxels	0.000	0.000	0.000	0.045	0.045	0.000	0.000	0.000	0.091	0.182	0.273	0.864
MAE, mm	0.000	0.000	0.000	0.180	0.180	0.000	0.000	0.000	0.364	0.728	1.092	3.456

FLT signature. Fig. 5 shows typical PET data acquired at one day before transplant, 5 days after transplant, and 28 days after transplant, demonstrating that the FLT uptake dramatically increases as engraftment proceeds. While our proposed method works well on scans acquired at 28 days post transplant as demonstrated, it is unlikely to be effective at the earlier observation points due to a lack of sufficient FLT uptake. Therefore, the method proposed in this paper may be best for scans obtained on patients to evaluate marrow cancers such as leukemia or metastases or other hyperproliferative processes rather than early post HSCT events when the PET signal is at a lower level. Second, because the boundary detection method works by integrating PET data within axial slices to derive the 1D signal $\widehat{I}_{\text{avg}}(i)$ given in (2) and illustrated in Fig. 3(b), the approximate intervertebral boundaries obtained with this technique are each inherently restricted to lie entirely within a single axial plane. These approximate boundaries may not be accurate enough to enable robust SUV measurements for individual vertebral bodies in scans presenting a high curvature of the spine, particularly in the smaller vertebrae of the upper thoracic and cervical regions. We are currently working on an observer model to resolve this issue by estimating the kyphotic and lordotic curves. Third and finally, training of our three U-Net models is limited by the small cohort size of the HSCT patient data set considered here. While we continue to acquire additional FLT PET/CT patient scans, we are also investigating techniques that may enable us to leverage large, publicly available pre-labeled datasets such as VerSe for improved training by applying signal processing to degrade the axial resolution in these datasets.

REFERENCES

[1] K. M. Williams and J. H. Chakrabarty, "Imaging haemopoietic stem cells and microenvironment dynamics through transplantation," *The Lancet Haematology*, vol. 7, no. 3, pp. e259–e269, Mar. 2020.

[2] A. Agool, B. Schot, P. Jager, and E. Vellenga, "18F-FLT PET in hematologic disorders: A novel technique to analyze the bone marrow compartment," *J. Nuclear Medicine*, vol. 47, no. 10, pp. 1592–1598, 2006.

[3] A.K. Buck, et al., "First demonstration of leukemia imaging with the proliferation marker 18F-Fluorodeoxythymidine," *J. Nuclear Medicine*, vol. 49, no. 11, pp. 1756–1762, 2008.

[4] M. Vanderhoek, M. Juckett, S. Perlman, R. Nickles, and R. Jeraj, "Early assessment of treatment response in patients with AML using [18F]FLT PET imaging," *Leukemia Research*, vol. 35, no. 3, pp. 310 – 316, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0145212610003024

[5] S. Schelhaas, K. Heinzmann, V. R. Bollineni, G. M. Kramer, Y. Liu, J. C. Waterton, E. O. Aboagye, A. F. Shields, D. Soloviev, and A. H. Jacobs, "Preclinical applications of 3'-Deoxy-3'-P[(18)F]Fluorothymidine in oncology - a systematic review," *Theranostics*, vol. 7, no. 1, pp. 40–50, 2017.

[6] K. M. Williams, J. Holter-Chakrabarty, L. Lindenberg, Q. Duong, S. K. Vesely, C. T. Nguyen, J. P. Havlicek, K. Kurdziel, J. Gea-Banacloche, F. I. Lin, D. N. Avila, G. Selby, C. G. Kanakry, S. Li, T. Scordino, S. Adler, C. M. Bollard, P. Choyke, and R. E. Gress, "Imaging of subclinical haemopoiesis after stem-cell transplantation in patients with haematological malignancies: a prospective pilot study," *The Lancet Haematology*, vol. 5, no. 1, pp. e44–e52, Jan. 2018.

[7] C. Nguyen, J. Havlicek, Q. Duong, S. Vesely, R. Gress, L. Lindenberg, P. Choyke, J. H. Chakrabarty, and K. Williams, "An automatic 3D CT/PET segmentation framework for bone marrow proliferation assessment," in *Proc. IEEE Int'l. Conf. Image Process.*, Phoenix, AZ, Sep. 25-28, 2016, pp. 4126–4130.

[8] N. Lessmann, B. van Ginneken, P. de Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation," in *Proc. 1st Conf. Medical Imaging with Deep Learning*, 2018.

[9] N. Shigeta, M. Kamata, and M. Kikuchi, "Effectiveness of pseudo 3D feature learning for spinal segmentation by CNN with U-Net architecture," *J. Image and Graphics*, vol. 7, no. 3, pp. 107–111, Sep. 2019.

[10] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Coarse to fine vertebrae localization and segmentation with SpatialConfiguration-Net

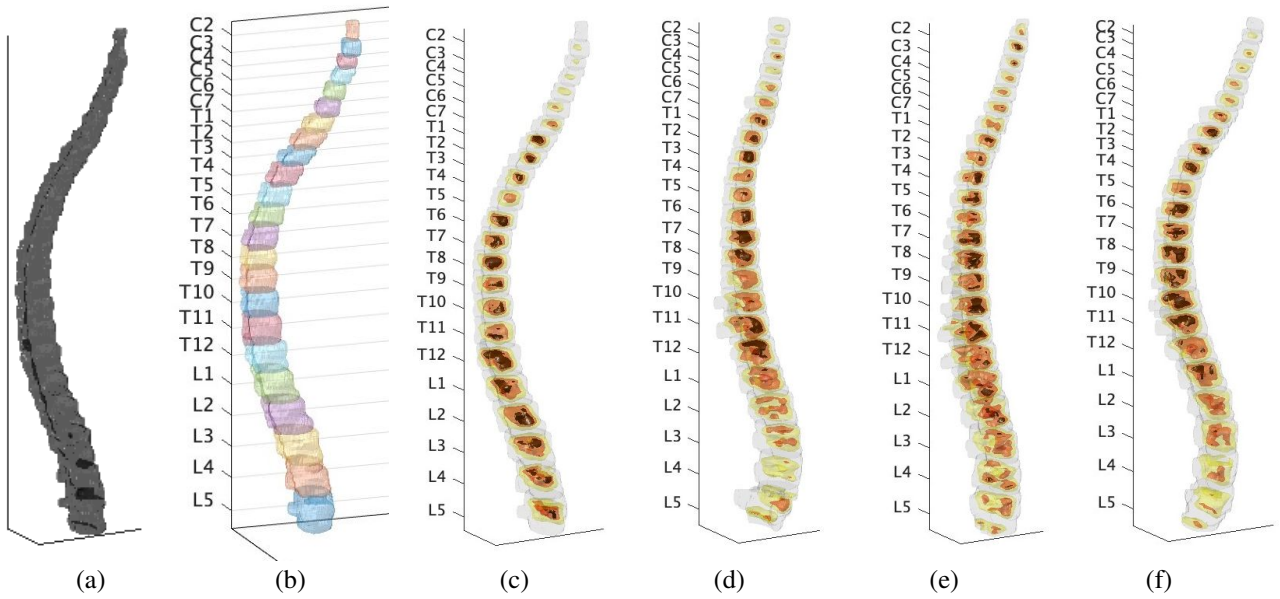


Fig. 4. Typical segmentation results for scans acquired at 28 days after transplant. (a) Binary mask for vertebral body class voxels obtained from CT data after semantic segmentation and post processing; (b) final approximate vertebral body instance segmentation obtained by our proposed PET-CT data fusion method. Note that each approximate intervertebral boundary is restricted to a single axial plane; (c)-(f) overlay of FLT SUV measurements calculated from PET data onto the approximate vertebral body instance segmentations obtained by the proposed method.

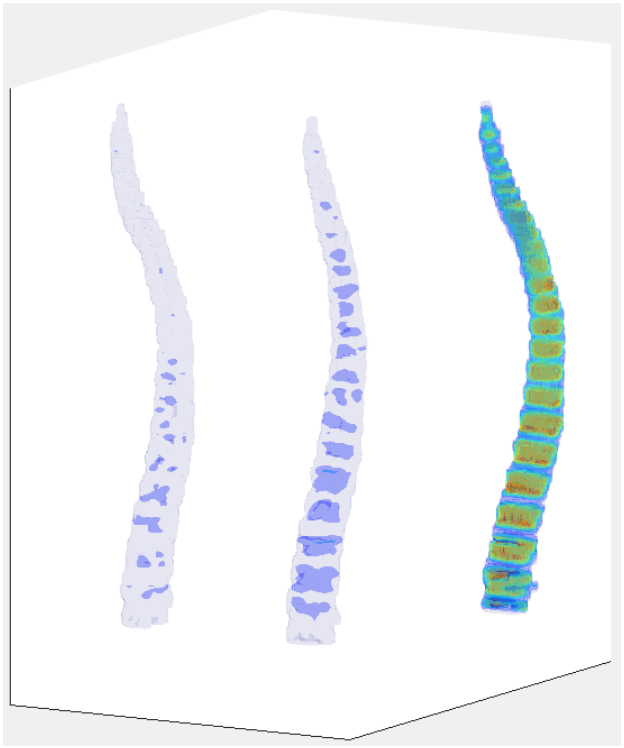


Fig. 5. Typical FLT PET data for the spine acquired at one day before, 5 days after, and 28 days after transplant illustrating dramatically increased FLT uptake as engraftment proceeds.

and U-Net," in *Proc. 15th Joint Int'l. Conf. Comput. Vision, Imaging, Comput. Graphics Theory, Appl.*, Vol. 5, 2020, pp. 124–133.

[11] A. Sekuboyina *et al.*, "VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images," *Medical Image Analysis*, vol. 73, 2021.

[12] X. You, Y. Gu, Y. Liu, S. Lu, X. Tang, and J. Yang, "EG-Trans3DUNet: A single-staged transformer-based model for accurate vertebrae segmentation from spinal CT images," in *Proc. IEEE Int'l. Symp. Biomed. Imaging*, 2022, pp. 1–5.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Comput., Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv preprint 1412.6980.

[15] J. Glover, A. Nadig, S. Vesely, D. Neelakantan, K. M. Williams, and J. Holter-Chakrabarty, "Fluorothymidine PET/CT identifies a case of herpes simplex virus esophagitis," *Radiol Imaging Cancer.*, vol. 5, no. 2, e220141, Mar. 2023.