HAL
open science

Improvement of a skeleton segmentation model of bone scintigrams with a transformer and component tree loss function

Quynh Anh Nguyen, Jean Cousty, Yukiko Kenmochi, Shigeaki Higashiyama, Joji Kawabe, Akinobu Shimizu

To cite this version:

Quynh Anh Nguyen, Jean Cousty, Yukiko Kenmochi, Shigeaki Higashiyama, Joji Kawabe, et al.. Improvement of a skeleton segmentation model of bone scintigrams with a transformer and component tree loss function. International Congress and Exhibition of Computer Assisted Radiology and Surgery CARS 2023, Jun 2023, Munich, Germany. hal-04001790

HAL Id: hal-04001790

https://hal.science/hal-04001790v1

Submitted on 11 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Improvement of a skeleton segmentation model of bone scintigrams with a transformer and component tree loss function

Quynh Anh Nguyen[1], Jean Cousty[2], Yukiko Kenmochi[3],
Shigeaki Higashiyama[4], Joji Kawabe[4], Akinobu Shimizu[1]

[1]TUAT - Tokyo University of Agriculture and Technology
[2]LIGM - Laboratoire d'Informatique Gaspard-Monge
[3]GREYC - Groupe de Recherche en Informatique, Image et Instrumentation de Caen
[4]Osaka Metropolitan University

Abstract

Keyword: Skeleton segmentation; transformer; component tree loss function

1 Purpose

Bone scintigraphy is widely used in clinical radiological examinations and is effective in diagnosing bone metastases from neoplastic diseases. Bone scintigrams of the human body are two-dimensional images obtained from the bone scintigraphy process from the anterior and posterior sides of the patient. In bone scintigraphy diagnostic imaging, the Bone Scan Index (BSI) is useful for quantitatively evaluating the spread of bone metastases. As the calculation of BSI requires the bone metastasis area and the bone area on whole-body anterior-posterior images, an accurate skeleton segmentation of the whole-body anteroposterior images is essential. However, there are cases where the recognition fails owing to high concentrations of positive accumulation and the irregular shape of organs when using the conventional method. In this study, we introduce a novel deep learning-based model with a transformer that combines the component tree loss function for whole-body bone scintigram skeleton segmentation. The proposed method aims to improve the dice similarity coefficient (DSC) of bones for both anterior and posterior images. Additionally, the component tree loss function is expected to reduce the difference in the number of connected components and holes between the predicted results and ground truths.

2 Methods

TransBtrflyNet network. The TransBtrflyNet model is a combination of BtrflyNet and TransUNet [1]. The skeleton segmentation input is a pair of anterior and posterior images. The network consists of an encoder, concatenation part, and a decoder. Following the TransUNet framework, a CNN-Transformer hybrid model was used for the encoder. The size of the input image was 576×256 pixels. First, the anterior and posterior images were fed simultaneously into the CNN blocks to generate feature maps. The transformer was applied to 1×1 patches extracted from the CNN feature maps of anterior and posterior images. The outputs from the transformer block were then reshaped for the next steps. The concatenation part has multiple convolutions, max-pooling layers, deconvolutions, and a skip connection, to enhance features. By concatenating the outputs from the transformer blocks of anterior and posterior features, the model can learn the characteristics of both the anterior and

posterior images. The output from the concatenation part is deconvoluted into anterior and posterior features for the decoder process. Skip connections were used to concatenate the features from the encoder and decoder. Finally, the segmentation heads output the segmentation labels for anterior and posterior images.

Loss function. The loss function consists of cross-entropy loss ($\mathcal{L}_{CE}$), deep supervision (DSV) loss ($\mathcal{L}_{DSV}$) which is calculated based on the dice loss, and component tree (CTr) loss ($\mathcal{L}_{CTr}$) which is designed to reconnect discontinuous components by selecting or discarding image maxima [2]. The loss function is expressed as follows:

$$\mathcal{L}_{CTr} = \sum_{i=1}^{i \leq \ell} \max(m - \mathbf{sm}_{\mathbf{r}_i}, 0)^p + \sum_{i=\ell+1}^{i \leq k} \mathbf{sm}_{\mathbf{r}_i}^p \text{ with } \mathbf{r} = \text{argsort}(\mathbf{im}) \tag{1}$$

where the constant margin $m \in \mathbb{R}$ is set to 1, which is the maximum value of the softmax function. The symbol $\ell \in \mathbb{N}^+$ is a target number of maxima. We select the value of ℓ based on the number of connected components of the ground truth. The values of ℓ differed for each bone. Here, $\mathbf{sm} \in \mathbb{R}^+$ and $\mathbf{im} \in \mathbb{R}^+$ represent a saliency and an important measure of the maxima respectively. We used a combination of dynamic and volume for ($\mathbf{sm}$, $\mathbf{im}$).

$$\mathcal{L}_{TransBtrflyNet} = \frac{(0.5\mathcal{L}_{CE_A} + 0.5\mathcal{L}_{DSV_A} + 0.0001\mathcal{L}_{CTr_A} + 0.5\mathcal{L}_{CE_P} + 0.5\mathcal{L}_{DSV_P} + 0.0001\mathcal{L}_{CTr_P})}{2} \tag{2}$$

where A and P are abbreviations of anterior and posterior.

3 Results

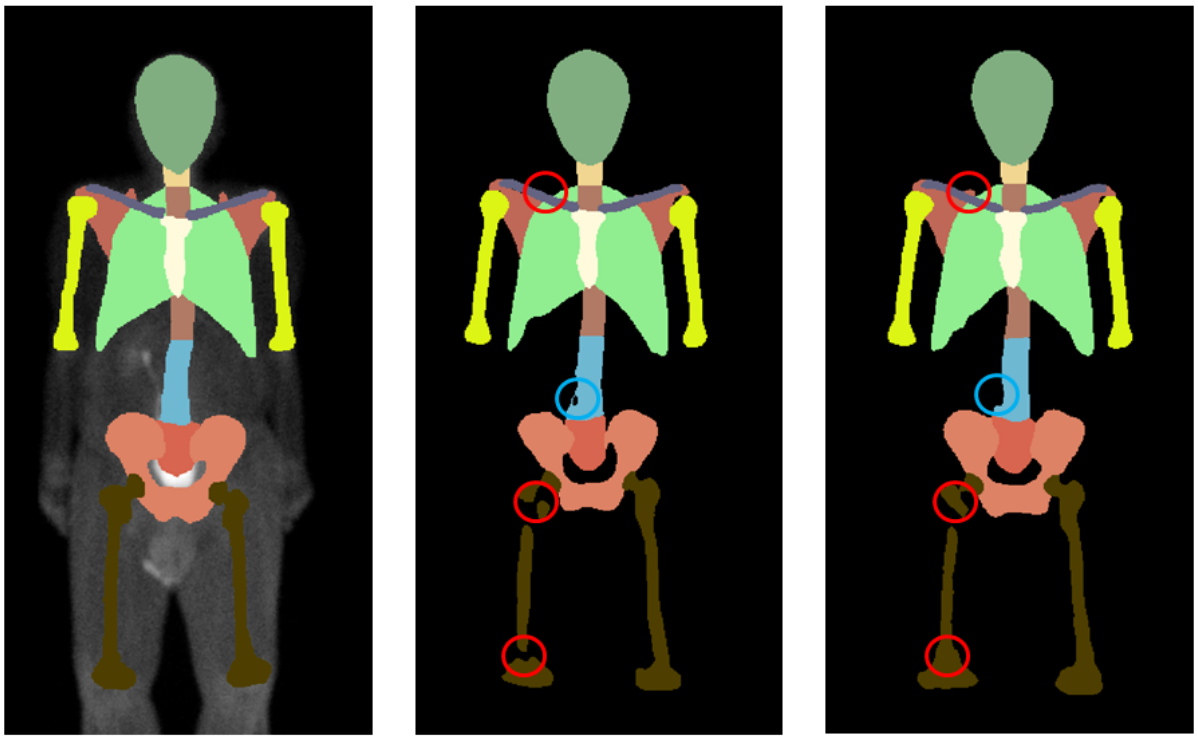
The segmentation targets included 12 anterior bones (skull, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum, pelvis, ribs, scapula, humerus, femur, sternum, and clavicle) and 10 posterior bones (skull, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum, pelvis, ribs, scapula, humerus, and femur). Experiments were conducted on 2,470 bone scintigrams obtained from 1,235 cases, and the respective results were compared and discussed. The 3-fold cross-validation method was used for performance evaluation. We defined the difference between the predictions and ground truths for the number of connected components (delta-CCs) and holes (delta-holes). We compared the results between the TransBtrflyNet and TransBtrflyNet+CTr models in terms of DSC, delta-CCs, and delta-holes. Wilcoxon signed-rank test was used for statistical analysis.

Dice similarity coefficient. We successfully improved the DSC of three anterior bones: skull ($p < 0.01$), pelvis ($p < 0.01$), and sternum ($p < 0.05$). However, the DSC of lumbar ($p < 0.01$), clavicle ($p < 0.01$) of TransBtrflyNet+CTr was inferior to that of TransBtrflyNet.

For posterior results, the DSC of three bones of TransBtrflyNet+CTr: sacrum ($p < 0.05$), whole scapula ($p < 0.05$), and scapula overlapping ribs ($p < 0.01$), significantly outperformed that of TransBtrflyNet. Skull ($p < 0.01$), cervical ($p < 0.01$), and humerus ($p < 0.01$) showed statistically lower results. Besides the DSC, we evaluated the performance of the two models based on the delta-CCs and delta-holes.

Delta-CC and delta-holes. For the results of TransBtrflyNet+CTr, the delta-CCs were reduced in three anterior bones: skull ($p < 0.05$), humerus ($p < 0.05$), clavicle ($p < 0.05$); the delta-holes were reduced in one anterior bone, lumbar ($p < 0.05$) and one posterior bone, scapula ($p < 0.01$). The TransBtrflyNet+CTr model did not produce any inferior results for both delta-CCs and delta-holes.

Figure 1 and Table 1 show an example in which the delta-holes of lumbar of TransBtrflyNet+CTr statistically exceeded that of TransBtrflyNet. The number of connected components of femur also decreased. The average DSC of the model with the CTr loss was better than that of the model without the CTr loss.



(a) Ground truth (b) TransBtrflyNet (c) TransBtrflyNet+CTr

Figure 1: Example of skeleton segmentation by TransBtrflyNet and TransBtrflyNet+CTr. (a) Ground truth, (b) TransBtrflyNet, and (c) TransBtrflyNet+CTr. Red circle: delta-CCs, blue circle: delta holes.

DSC/delta-CCs/delta-holes	TransBtrflyNet	TransBtrflyNet+CTr
All bones	0.8977/0.3333/0.0833	0.9089/0.0833/0
lumbar	0.9165/0/1	0.8916/0/0
femur	0.7867/3/0	0.8125/1/0
scapula	0.8956/1/0	0.8994/0/0

Table 1: DSC, delta-CCs, and delta-holes results of Figure 1.

4 Conclusion

We present a deep learning-based method that integrates a transformer for the skeleton segmentation of bone scintigrams. We also introduced the CTr loss function to reduce the number of connected components and holes. The TransBtrflyNet+CTr model showed promising results compared to the TransBtrflyNet model.

In the future, we will try various combinations of (**sm**, **im**) of CTr loss corresponding to the characteristics of each bone to anatomically improve the segmentation results.

References

[1] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. 2021.

[2] Perret, B., and Cousty, J. Component Tree Loss Function: Definition and Optimization. *Discrete Geometry and Mathematical Morphology. DGMM 2022. Lecture Notes in Computer Science*, 13493:248–260, 2022.