Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Knowledge distillation on individual vertebrae segmentation exploiting 3D U-Net

Luís Serrador [a,b,*], Francesca Pia Villani [c], Sara Moccia [d], Cristina P. Santos [a,b]

[a] Center for MicroElectroMechanical Systems (CMEMS), University of Minho, Guimaraes, Portugal
[b] Clinical Academic Center of Braga (2CA-Braga), Hospital of Braga, Braga, Portugal
[c] Department of Humanities, Università degli Studi di Macerata, Italy
[d] The BioRobotics Institute and Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Italy

ARTICLE INFO

Keywords:
Vertebra segmentation
3D U-net
Knowledge distillation
Computed tomography

ABSTRACT

Recent advances in medical imaging have highlighted the critical development of algorithms for individual vertebral segmentation on computed tomography (CT) scans. Essential for diagnostic accuracy and treatment planning in orthopaedics, neurosurgery and oncology, these algorithms face challenges in clinical implementation, including integration into healthcare systems. Consequently, our focus lies in exploring the application of knowledge distillation (KD) methods to train shallower networks capable of efficiently segmenting vertebrae in CT scans. This approach aims to reduce segmentation time, enhance suitability for emergency cases, and optimize computational and memory resource efficiency. Building upon prior research in the field, a two-step segmentation approach was employed. Firstly, the spine's location was determined by predicting a heatmap, indicating the probability of each voxel belonging to the spine. Subsequently, an iterative segmentation of vertebrae was performed from the top to the bottom of the CT volume over the located spine, using a memory instance to record the already segmented vertebrae. KD methods were implemented by training a teacher network with performance similar to that found in the literature, and this knowledge was distilled to a shallower network (student). Two KD methods were applied: (1) using the soft outputs of both networks and (2) matching logits. Two publicly available datasets, comprising 319 CT scans from 300 patients and a total of 611 cervical, 2387 thoracic, and 1507 lumbar vertebrae, were used. To ensure dataset balance and robustness, effective data augmentation methods were applied, including cleaning the memory instance to replicate the first vertebra segmentation. The teacher network achieved an average Dice similarity coefficient (DSC) of 88.22% and a Hausdorff distance (HD) of 7.71 mm, showcasing performance similar to other approaches in the literature. Through knowledge distillation from the teacher network, the student network's performance improved, with an average DSC increasing from 75.78% to 84.70% and an HD decreasing from 15.17 mm to 8.08 mm. Compared to other methods, our teacher network exhibited up to 99.09% fewer parameters, 90.02% faster inference time, 88.46% shorter total segmentation time, and 89.36% less associated carbon (CO_2) emission rate. Regarding our student network, it featured 75.00% fewer parameters than our teacher, resulting in a 36.15% reduction in inference time, a 33.33% decrease in total segmentation time, and a 42.96% reduction in CO_2 emissions. This study marks the first exploration of applying KD to the problem of individual vertebrae segmentation in CT, demonstrating the feasibility of achieving comparable performance to existing methods using smaller neural networks.

1. Introduction

In recent years, the field of medical imaging has witnessed a surge in technological advancements, particularly in the realm of computer-aided diagnosis and image analysis. Among these innovations, the precise segmentation of individual vertebrae on computed tomography (CT) scans stands out as a pivotal development with far-reaching clinical implications (Altini et al., 2021). Accurate identification and delineation of vertebral structures are fundamental not only for diagnostic purposes but also for planning interventions and monitoring disease progression. This technique finds application in diverse clinical domains, including orthopedics, neurosurgery, and oncology (Ren et al.,

[*] Corresponding author at: Center for MicroElectroMechanical Systems (CMEMS), University of Minho, Guimaraes, Portugal.
E-mail address: id7608@alunos.uminho.pt (L. Serrador).

https://doi.org/10.1016/j.compmedimag.2024.102350
Received 26 September 2023; Received in revised form 1 February 2024; Accepted 1 February 2024
Available online 8 February 2024

Table 1
Overview of the VerSe dataset. Patient/Scan split indicates the split of the data into train/validation/test sets. Cer, Tho and Lumb stands for cervical, thoracic and lumbar, respectively.

VerSe	Patients	Scans	Patient split	Scan split	Vertebrae (Cer/Tho/Lum)
2019	141	160	67/37/37	80/40/40	1725 (220/884/621)
2020	300	319	100/100/100	113/103/103	4141 (581/2255/1305)
Total	355	374	128/117/110	141/120/113	4505 (611/2387/1507)

2022; Yagi et al., 2023; Chen et al., 2023). The ability to automatically segment vertebrae not only expedites the radiological workflow but also enhances the overall accuracy of diagnoses. Despite the promises it holds, deploying algorithms in clinical environments presents a set of challenges. Issues such as algorithm robustness, adaptability to diverse patient populations, and the need for seamless integration into existing healthcare systems are significant concerns that necessitate careful consideration (Saw and Ng, 2022).

In this work, our specific interest lies in determining the feasibility of achieving comparable performance in vertebrae segmentation using shallower neural networks, specifically 3D U-Nets with fewer hidden layers between input and output. Our goal is to reduce segmentation time and computational costs without compromising performance, enabling the deployment of algorithms in clinical devices. KD methods were employed in this study, as their application has proven to enhance accuracy without incurring additional computational overhead in CT segmentation problems, such as thoracic (Noothout et al., 2022) and abdominal organ segmentation (Choi, 2022; Zhang et al., 2022b), pancreas (You et al., 2022), liver and kidney segmentation (Xu et al., 2021; Qin et al., 2021), liver and kidney tumor segmentation (Qin et al., 2021) and COVID-19 lesions segmentation (Xu et al., 2022). In fact, the application of KD is considered a current emerging trend in medical imaging segmentation (Conze et al., 2023).

The contributions of this paper can be summarized as follows:

- This research marks the pioneering application of KD methods to address the challenge of individual vertebrae segmentation on CT scans.
- To enhance the effectiveness of the distillation framework for segmenting the first vertebrae in CT scans, problem-specific data augmentation techniques are proposed.
- To the best of our knowledge, this work stands out as the first to leverage the largest dataset of CT scans, encompassing the highest number of vertebrae during the development phase.
- This work introduces a strategy to reduce segmentation time, energy consumption, and computational requirements for individual vertebrae segmentation in CT scans, achieved by incorporating shallower networks into existing algorithms.

2. Related work

The initial approaches investigated for individual vertebrae segmentation involved a 2D analysis of the CT scan, where each slice was analyzed separately to achieve a 3D segmentation of the vertebrae. However, with the advent of large datasets, the exploration of deep learning (DL) methods, specifically 3D neural networks, emerged as highly advantageous for vertebral segmentation.

Janssens et al. (2018) introduced the first method solely based on 3D neural networks for individual vertebrae segmentation. Their approach uses two cascaded 3D fully convolutional networks (FCNs) - one for localization and one for segmentation. Despite achieving strong performance, it is important to note that the dataset used for development and evaluation was notably uniform, featuring scans with vertebrae from L1 to L5 and a consistent field-of-view (FoV).

To address the challenge of varying FoV in recent datasets, Lessmann et al. (2019) introduced an iterative instance-by-instance segmentation approach based on a 3D FCN. However, the authors noted drawbacks, including the absence of initialization, requiring the network to scan the entire volume to locate the first vertebra, and less effective segmentation of the topmost vertebrae. Additionally, the use of large feature maps along the FCN resulted in extended segmentation times.

In response to the issue of initial vertebrae localization, Payer et al. (2020) proposed a solution involving a coarse-to-fine approach, leveraging their Spatial-Configuration network (Payer et al., 2019) and the 3D U-Net. This three-step method employs a 3D U-Net for spine localization, the Spatial-Configuration Network for vertebrae localization via heatmap regression, and another 3D U-Net for segmenting each identified vertebra. Sekuboyina et al. (2021) secured victory in the VerSe20 challenge with a two-step approach. Similar to Payer et al. (2020), they employed a 3D U-Net for spine localization and another 3D U-Net for iterative vertebrae segmentation akin to Lessmann et al. (2019). They achieved increased accuracy by augmenting the number of model parameters, albeit at the cost of longer segmentation times.

Since the release of the hidden data from the VerSe challenge dataset, three studies have introduced novel implementations for neural networks in vertebral segmentation. Altini et al. (2021) devised a two-step solution involving preprocessing the CT scan to crop the volume around the spine and then using a convolutional network based on the 3D V-Net (Milletari et al., 2016) for vertebrae segmentation. However, this approach is semi-automated, requiring two user inputs: the number of vertebrae and slice selection. Tao et al. (2022) also proposed a two-stage solution for vertebral segmentation, focusing on the detection of vertebral centroids instead of vertebral positions. They achieved improved accuracy by increasing the number of parameters, albeit with longer segmentation times. Meng et al. (2023) introduced a cyclic algorithm for spine and vertebra segmentation, along with vertebra identification, aiming to enhance the segmentation of transitional vertebrae. Despite the improved performance in transitional vertebrae, there is a notable increase in segmentation time due to the cyclic process.

The current state-of-the-art methods primarily involve the implementation of two 3D U-Nets, each dedicated to a specific task within the vertebral segmentation algorithm: i) spine location prediction and ii) iterative individual vertebra segmentation (Lessmann et al., 2019; Payer et al., 2020; Sekuboyina et al., 2021; Tao et al., 2022). These methods have evolved by augmenting the models' structure, introducing more parameters, and subsequently increasing segmentation time. However, this augmentation comes with a notable environmental impact, driven by the computational resources required for training and execution. The size and complexity of these DL models contribute to a heightened environmental footprint (Strubell et al., 2019; Ligozat et al., 2022; Henderson et al., 2020; Lannelongue et al., 2021; Budennyy et al., 2022). Moreover, the computational time needed for these approaches to deliver segmented vertebrae using conventional central processing units (CPU) poses a significant challenge. To mitigate these issues, KD has been explored for U-Net architectures in the segmentation of body structures on CT scans. This technique aims to train a smaller neural network (student network) to emulate the behavior of a larger neural network (teacher network) but with fewer parameters and faster inference times. The objective is to achieve efficient segmentation results, leading to shorter segmentation times and reduced computational requirements.

Noothout et al. (2022) applied KD techniques for thoracic organ segmentation by using an ensemble of neural networks with different structures as teachers. They successfully distilled knowledge from this ensemble to a single neural network, achieving the same performance as the entire ensemble with only one network. Choi (2022) investigated

KD in abdominal organ segmentation, proposing a coarse-to-fine framework that included a teacher and student network pair for each step (coarse and fine segmentation). The student networks demonstrated a performance increase of 7%. Zhang et al. (2022b) employed KD methods to compress the model for abdominal organ segmentation, resulting in an 11% performance boost on the student network. Zhang et al. (2022a) used an ensemble of teacher networks focusing on individual organ segmentation, distilling knowledge to a student neural network capable of performing multi-organ segmentation, improving the student network's performance by 1%.

(Xu et al., 2021) introduced a KD method based on growing teacher assistant networks, which bridge the gap between teacher and student sizes. This approach improved the student network's performance in liver segmentation on CT by 1%–2%. Qin et al. (2021) explored the impact of KD on whole liver and kidney segmentation, as well as liver and kidney tumor segmentation, demonstrating performance improvements for student networks ranging from 1% to 17%. You et al. (2022) enhanced the performance of a student network by 13% through knowledge distillation from a teacher network for pancreas segmentation on CT scans. Xu et al. (2022) investigated a KD method for COVID-19 lesions segmentation, using the encoder segment of an autoencoder for healthy case reconstruction as a teacher and a shallower network with the same architecture as the student. This approach improved the networks' performance by 2% to 7%.

Beyond CT scans, KD has found success in various healthcare applications, including low-dose CT image denoising (Wang et al., 2023), bone suppression on chest X-rays (Liu et al., 2023), brain tumor segmentation on Magnetic Resonance Imaging (MRI) (Qi et al., 2022; Xiong et al., 2023; Rahimpour et al., 2021; Noothout et al., 2022; Lachinov et al., 2020), left atrial segmentation on MRI (You et al., 2022), and heart segmentation in cine-MRI (Noothout et al., 2022).

3. Material and methods

This section introduces the proposed framework, the datasets employed, and the training settings. Fig. 1 provides an overview of the implemented method, showcasing the training data, preprocessing steps, training workflow, and the segmentation algorithm (inference workflow).

3.1. VerSe dataset

In this paper, we leverage the VerSe dataset (Sekuboyina et al., 2021; Liebl et al., 2021), initially introduced in 2019 for the VerSe19 Challenge at the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). The VerSe20 dataset, released in 2020 for the 23rd MICCAI International Conference, includes 105 cases from VerSe19, totaling 319 scans across 300 patients. The dataset encompasses cases with Schmorl nodes, hemangioma, degenerative changes, or the presence of foreign materials for kyphoplasty or spondylodesis. Acquired from various CT scanners at different institutions, the dataset exhibits significant variability in terms of FoV, scan settings, and findings. The images, available in Neuroimaging Informatics Technology Initiative (NIfTI) format, can be accessed on the Open Science Framework (OSF) repository (https://osf.io/nqjyw/, last accessed: 8 April 2023).

Table 1 provides details about the VerSe dataset and its split. For this study, all 374 scans were used, comprising 105 scans common to both VerSe19 and VerSe20 datasets, 55 scans from VerSe19 only, and 214 from VerSe20 only. The data were split on a patient level, as recommended by the dataset authors, resulting in the following distribution: 141 scans for training, 120 scans for validation, and 113 scans for testing.

3.2. Segmentation algorithm

Following a two-step approach inspired by Sekuboyina et al. (2021), depicted in the green dotted box at the bottom of Fig. 1, 3D U-Nets were employed. The first step involves locating the spine in the CT scan using a 3D U-Net (*Spine Location 3D U-Net*). This network outputs a *Spine Location Heatmap*, indicating the probability of each voxel belonging to the spine. In the second step, individual vertebrae within the region specified by the heatmap from the first step are segmented. The segmentation is performed iteratively from the top to the bottom of the heatmap, with a memory instance storing the segmented vertebrae. Fig. 2 illustrates this iterative process of individually segmenting vertebrae on the CT scan. Initially, the topmost vertebra is segmented and recorded in the memory instance. Then, the second vertebra of the CT scan is segmented, which is now the uppermost not yet segmented vertebra. The process is repeated until all vertebrae are segmented. The segmentation is carried out by a 3D U-Net (*Vertebra Segmentation 3D U-Net*), which takes as inputs the CT volume containing the vertebrae to be segmented and the memory instance containing the already segmented vertebrae within that volume. The output is the segmentation of the topmost unsegmented vertebra.

3.3. Spine location with 3D U-net

The Spine Location 3D U-Net is tasked with identifying the region of the CT scan containing the spine, ensuring the iterative vertebrae segmentation occurs within that area. Regardless of the CT scan's FoV, which may encompass the cervical, thoracic, lumbar segments, the entire spine, or even the full body, the Spine Location 3D U-Net predicts the probability of each voxel belonging to the spine.

The input to the Spine Location 3D U-Net is the resized CT scan, set to 64x64x128 voxels (left side of the orange dotted box in Fig. 1). Following the approach of Sekuboyina et al. (2021) and Payer et al. (2020), a 3D U-Net with four depth levels was implemented. Each encoder and decoder level comprises two consecutive blocks involving convolution with a 3x3x3 kernel, followed by batch normalization (BN) and rectified linear unit (ReLU) activation. Max pooling operations reduce spatial dimensions between encoder levels. Nearest neighbor upsampling aligns decoder spatial dimensions. The convolutional layers have 8, 16, 32, and 64 output features in the first, second, third, and fourth levels, respectively.

For improved generalization, training incorporates data augmentation: random translation [−20, +20] in all axes, left–right axis flip, random multiplication in the range [0.75, 1.25] with a shift of [−0.25, +0.25], random zoom [−10, +10]% of the volume's size, random rotation [−10, +10] degrees in all axes, Gaussian noise, and Gaussian blur.

The neural network is trained using binary cross-entropy as the loss function. Training was done using an Nvidia 1660 Dual Super GPU with 6 GB memory, a batch size of 4, and the Adam optimizer with a small learning rate of 0.0001 to ensure slow convergence and good generalization over 300 epochs.

3.4. Vertebra segmentation with 3D U-net

The Vertebra Segmentation 3D U-Net is responsible for segmenting the topmost not yet segmented vertebra within a given volume. A memory instance is employed to store already segmented vertebrae, aiding in identifying which vertebra to segment. The network takes two volumes as input: (i) the CT scan; and (ii) the memory instance (right side of the dotted orange box in Fig. 1).

The input to the Vertebra Segmentation 3D U-Net is the concatenation of the volume and memory instance, resulting in two channels of size 128x128x128, following the recommendation of Lessmann et al. (2019). The network follows the original U-Net architecture with four depth levels, each consisting of two consecutive blocks of one

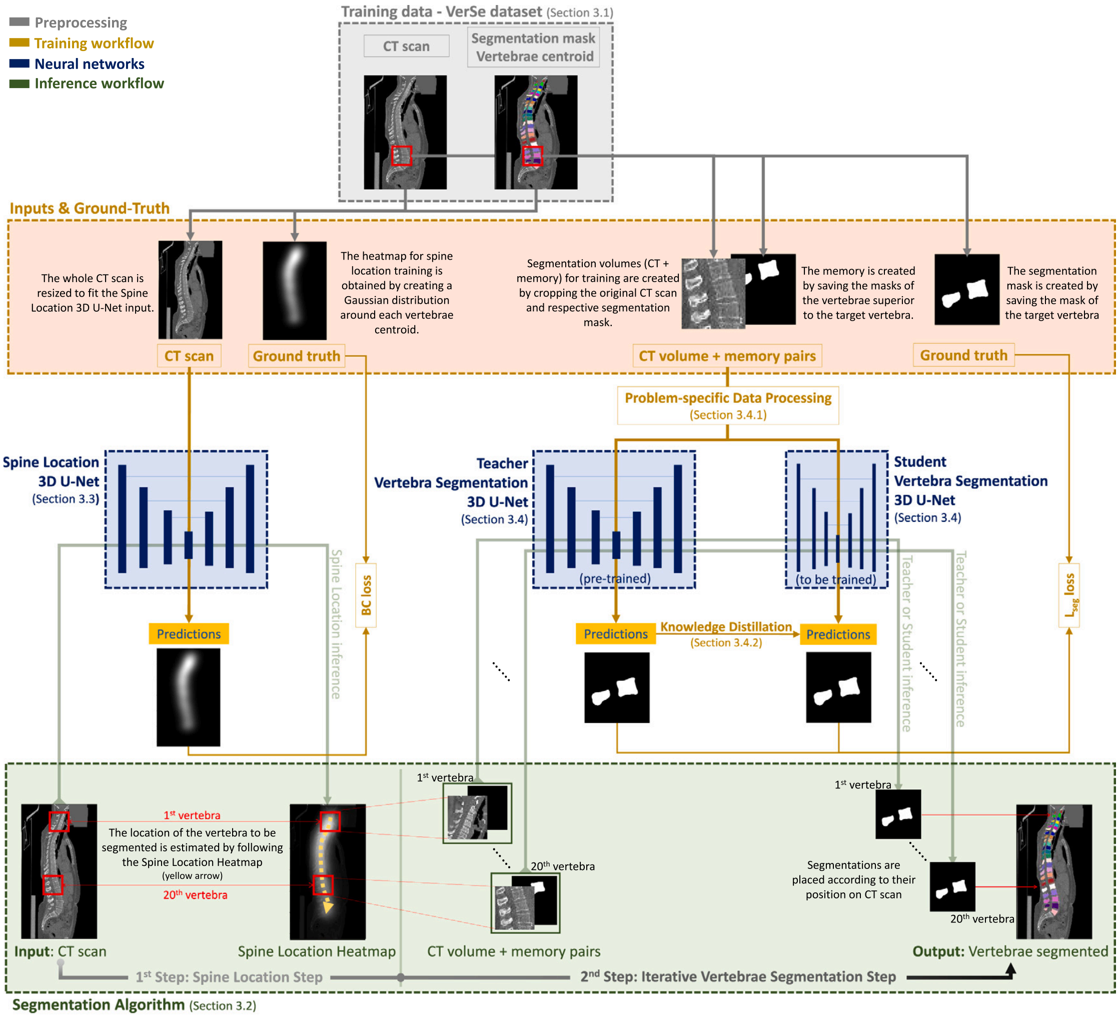
Preprocessing
Training workflow
Neural networks
Inference workflow
Training data - VerSe dataset (Section 3.1)
CT scan
Segmentation mask
Vertebrae centroid
Inputs & Ground-Truth
The whole CT scan is resized to fit the Spine Location 3D U-Net input.
CT scan
The heatmap for spine location training is obtained by creating a Gaussian distribution around each vertebrae centroid.
Ground truth
Segmentation volumes (CT + memory) for training are created by cropping the original CT scan and respective segmentation mask.
The memory is created by saving the masks of the vertebrae superior to the target vertebra.
CT volume + memory pairs
The segmentation mask is created by saving the mask of the target vertebra
Ground truth
Problem-specific Data Processing
(Section 3.4.1)
Spine Location
3D U-Net
(Section 3.3)
Teacher
Vertebra Segmentation
3D U-Net
(Section 3.4)
(pre-trained)
Student
Vertebra Segmentation
3D U-Net
(Section 3.4)
(to be trained)
Spine Location inference
BC loss
Teacher or Student inference
Teacher or Student inference
L_{seg} loss
Predictions
Predictions
Knowledge Distillation
(Section 3.4.2)
Predictions
1st vertebra
The location of the vertebra to be segmented is estimated by following the Spine Location Heatmap
(yellow arrow)
20th vertebra
1st vertebra
20th vertebra
1st vertebra
Segmentations are placed according to their position on CT scan
20th vertebra
Input: CT scan
Spine Location Heatmap
CT volume + memory pairs
Output: Vertebrae segmented
1st Step: Spine Location Step
2nd Step: Iterative Vertebrae Segmentation Step
Segmentation Algorithm (Section 3.2)

Fig. 1. Overview of our method. The VerSe dataset (gray dotted box at top) is used for training purposes. In regards to the Spine Location 3D U-Net the training data is processed in order to get the resized whole CT scan and the heatmap for spine location created from the vertebrae centroid, which are used to train the neural network (left side of orange dotted box). The Spine Location 3D U-Net is trained using the binary-crossentropy loss. For training the Teacher/Student Vertebra Segmentation 3D U-Nets the training data is processed to get the CT, memory and segmentation mask volumes centered in each target vertebra (right side of orange dotted box). The Teacher network is pre-trained using the segmentation loss (L_{seg} loss) exploiting our proposed problem-specific data processing methods. The Student network is trained using the segmentation loss (same process as Teacher) and by distilling knowledge from the Teacher network. The segmentation algorithm (green dotted box at bottom) consist in two main steps: Spine Location Step and Iterative Vertebrae Segmentation. The Spine Location Step consists in locating the spine (Spine Location Heatmap) in the whole CT scan. After this step, an Iterative Vertebrae Segmentation is performed by following the heatmap of the spine location (yellow dotted arrow on Spine Location Heatmap). The volumes from the CT scan and the memory instance (which saves the already segmented vertebrae) are inputs of the Teacher/Student Vertebra Segmentation 3D U-Net that segments the top-most not yet segmented vertebra. When segmenting the 1st vertebra the memory is empty, while when segmenting other than the 1st vertebra, it contains the already segmented vertebrae (memory of the 20th vertebra as example).

convolution, BN layer, and ReLU activation. Max pooling operations between encoder levels halve spatial dimensions, while nearest neighbor upsampling in the decoder aligns dimensions.

KD techniques were applied to the problem of individual vertebrae segmentation on CT by distilling knowledge from a Teacher Vertebra Segmentation 3D U-Net to a smaller Student Vertebra Segmentation 3D U-Net with the same architecture. The convolutional layers of the Teacher network have 16, 32, 64, and 128 output features in the first, second, third, and fourth levels, respectively. The Student network has half the output features: 8, 16, 32, and 64 in the first, second, third, and fourth levels, respectively.

Data augmentation methods were applied to both inputs (volume and memory instance) for better generalization: random translation [−64, +64], left–right axis flip, random multiplication in the range [0.75, 1.25] with a shift of [−0.25, +0.25], random rotation [−20,+20] degrees in all axes, Gaussian noise, Gaussian blur, and random zoom [−30, +30]% in the superior-inferior axis and [−20, +20]% in the other axes. These zoom rates stretch and shrink the vertebrae along the superior-inferior axis.

The cost function for training was based on (Lessmann et al., 2019) segmentation loss, minimizing false positives and false negatives while penalizing incorrect predictions near the vertebra's edge using distance to border penalization weights. These weights are calculated from the distance d_i of voxel i to the closest point on vertebra surface: $\omega_i = \gamma \cdot exp(-d_i^2/\sigma^2) + 1$. We used $\gamma = 8$ and $\sigma = 6$, as suggested by Lessmann et al. (2019). In addition, the penalty for false positives and false negatives is balanced, *i.e.*, false negatives are penalized less than false positives at the beginning of training, and this penalty increases in a sigmoidal fashion until both penalties are equal at the end of training.

The Nvidia RTX 2080 Ti GPU with 12 GB was used for training, allowing only a single batch during Vertebra Segmentation 3D U-Net training. The network trained with Adam optimizer, a fixed learning rate of 0.001, and increased momentum of 0.99 for 30 epochs.

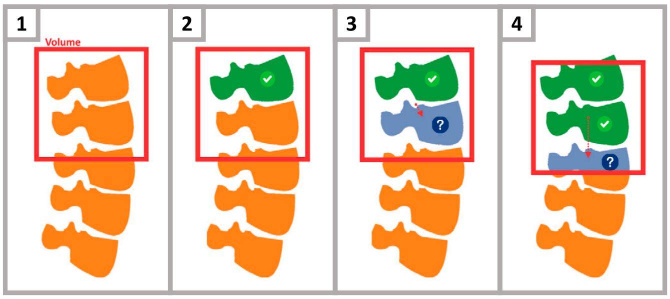

Fig. 2. Illustration of the *Iterative Vertebrae Segmentation* step. The volume is taken to include the top-most vertebra (1). The first vertebra is segmented and added to the memory instance of the Vertebrae Segmentation network (2) and the same volume is analyzed again, yielding now a fragment of the following vertebra ('?' in 3) because the updated memory instance forces the network to ignore the previous vertebra (3). The volume is centered now at the detected fragment of the following vertebra and the process is repeated (4).

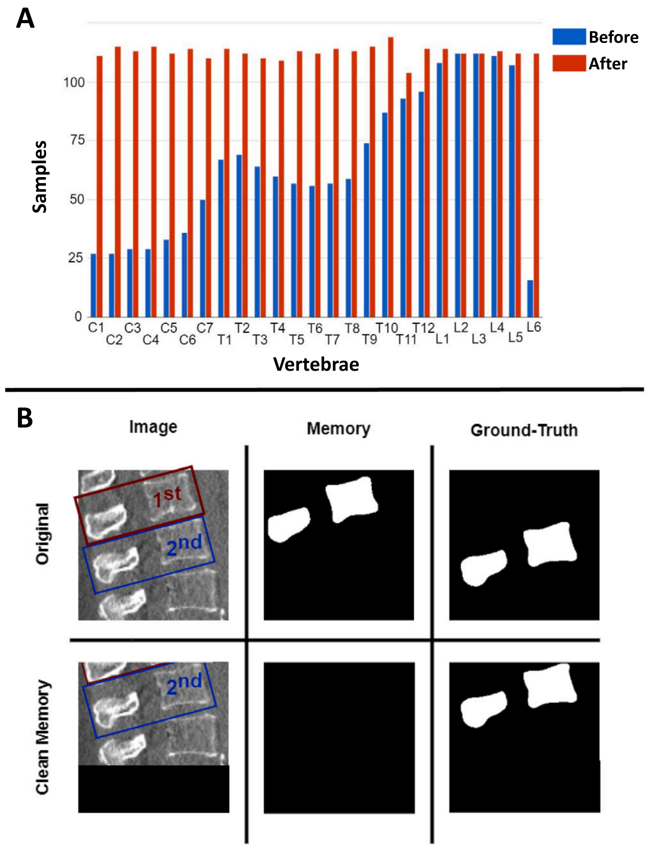

Fig. 3. Problem-specific data processing techniques. (A) Data distribution before (blue) and after (red) repeating aleatory vertebrae during training. (B) Memory cleaning: the first row shows the original image, memory, and ground truth, while the second row shows the case when the memory is emptied and the original image is rolled up to make sure that the first vertebra on the patch is now the second vertebra of the original image - the ground truth is updated accordingly.

3.4.1. *Problem-specific data processing*

In the quest for a more efficient method using shallower neural networks than those in existing literature, the training of Teacher and Student Vertebra Segmentation 3D U-Nets involved leveraging effective data processing techniques. These techniques, validated for enhancing network performance, include data resampling to address dataset imbalances and selectively emptying the memory instance in specific training cases to tackle challenges in first vertebra segmentation, as reported by Lessmann et al. (2019).

Addressing the substantial imbalance in the original training dataset, where there are considerably more thoracic and lumbar vertebrae than cervical vertebrae (as shown in Table 1), a data resampling strategy is employed. The oversampling approach is chosen to maintain a more balanced representation of cervical, thoracic, and lumbar vertebrae per epoch. For each training epoch, random vertebrae are duplicated to ensure similar counts for each class as shown in Fig. 3A, resulting in 2814 training samples after resampling the dataset.

To enhance the segmentation of the first vertebra in the CT scan, a problem-specific data augmentation method is proposed. Since the memory instance is empty when segmenting the first vertebra, and no prior vertebrae have been segmented, additional training volumes with an empty memory are introduced. Approximately one-third of the data is selected for this process, and during training, when the memory is empty, the vertebra is positioned at the top of the patch to simulate the segmentation of the first vertebra. This dynamic process is performed on-the-fly during training, with patches randomly selected in each epoch (as illustrated in Fig. 3B).

3.4.2. *Knowledge distillation for vertebra segmentation*

Our original KD implementation (depicted in Fig. 4A) followed a widely used method in the literature (Wang et al., 2019; Guan et al., 2019; Nekrasov et al., 2019b; Tseng et al., 2020) to distill knowledge from the Teacher to the Student Vertebra Segmentation 3D U-Net. For this purpose, the Student network was designed with two outputs: (i) the predicted segmentation; and (ii) the soft segmentation. The predicted segmentation ($\hat{y}$) was derived by applying the sigmoid activation function to the output of the last convolutional layer (logits). Conversely, the soft segmentation ($\hat{y}_s$) was obtained by dividing the logits by the Temperature hyperparameter (T) before applying the sigmoid activation function. Soft segmentation introduces increased entropy to the predicted segmentation, spreading voxel values in the range [0, 1] more uniformly (He et al., 2019). Also, the output of the Teacher network were soften in order to be used as soft targets during training of the Student Vertebra Segmentation 3D U-Net. Two versions were implemented with$T = 4$ and $T = 5$, respectively.

The total loss of the KD implementation (L_{TOTAL} in Fig. 4A) comprises two terms: i) the loss term between the predicted segmentation of the Student network and the ground truth (GT) ($L_{seg}(y, \hat{y})$); and ii) the loss term between the soft segmentation of the Student network and the soft targets of the Teacher network ($L_{seg}(y_s, \hat{y}_s)$). The total loss function for training was the weighted sum of these two losses ($\lambda\ L_{seg}(y_s, \hat{y}_s) + L_{seg}(y, \hat{y})$), with the hyperparameter λ consistently set to 1.

3.4.3. *Ablation studies*

The Teacher Vertebra Segmentation 3D U-Net underwent training with and without the application of problem-specific data processing methods to the input data (T_S and T_B, respectively). The version of the Teacher network trained without these methods (T_B) served as a baseline to assess the impact of the data processing methods on performance. Also, the initial training iteration of the Student 3D U-Net was performed from scratch (S_S), employing the same training procedure and loss function as T_S and acting as a baseline to evaluate the impact of KD techniques on the Student network's performance. In addition to the original KD implementation with $T = 4$ and $T = 5$ (S_{T4} and S_{T5}), tests were conducted with $T = 3$ (S_{T3}) to compare their performance.

Another KD approach from the literature was explored (Dou et al., 2020; Holliday et al., 2017; Nekrasov et al., 2019a; Park and Heo, 2020; Liu et al., 2019), involving matching the logits of the Student and Teacher networks (Fig. 4B). Two implementations were investigated: i) minimizing the Euclidean distance between Student and

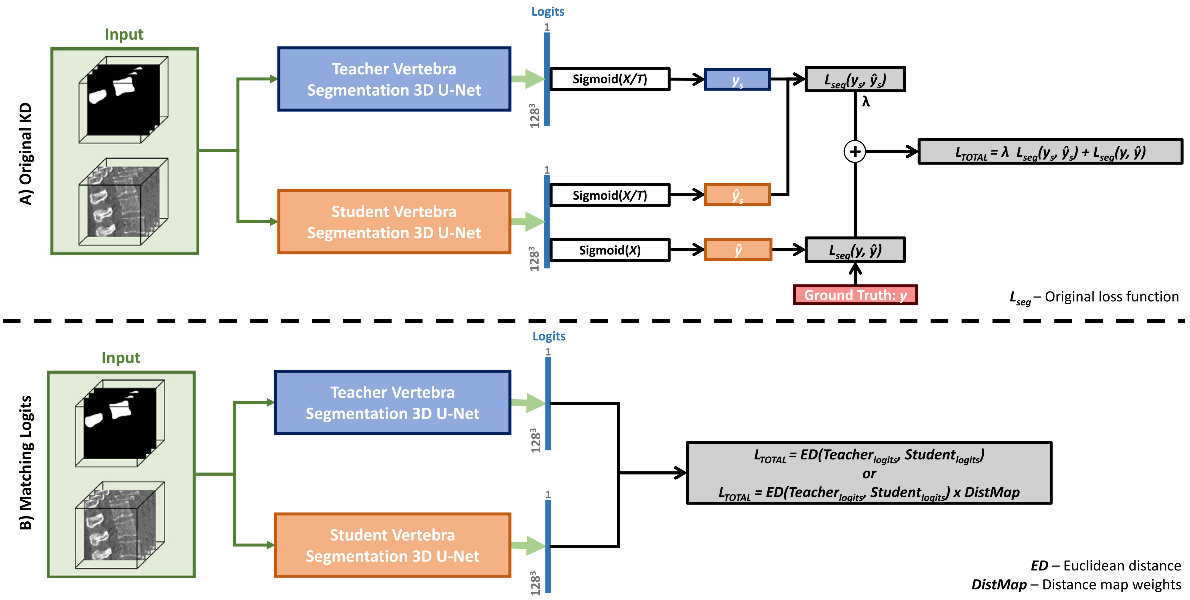
A) Original KD
Input
Teacher Vertebra Segmentation 3D U-Net
Student Vertebra Segmentation 3D U-Net
Logits
1
128³
Sigmoid(X/T)
Sigmoid(X/T)
Sigmoid(X)
y_s
ŷ_s
ŷ
L_seg(y_s, ŷ_s)
λ
L_seg(y, ŷ)
Ground Truth: y
L_TOTAL = λ L_seg(y_s, ŷ_s) + L_seg(y, ŷ)
L_seg – Original loss function
B) Matching Logits
Input
Teacher Vertebra Segmentation 3D U-Net
Student Vertebra Segmentation 3D U-Net
Logits
1
128³
L_TOTAL = ED(Teacher_logits, Student_logits)
or
L_TOTAL = ED(Teacher_logits, Student_logits) x DistMap
ED – Euclidean distance
DistMap – Distance map weights

Fig. 4. Implementation of KD methods from Teacher to Student Vertebra Segmentation 3D U-Net. The top implementation (A) corresponds to the original implementation of KD using the soft targets and the ground truth and using the loss function of the training of the Teacher Vertebra Segmentation 3D U-Net (L_{seg}). At the bottom (B), the implementation that consists in matching the logits of both networks is illustrated, with the two different loss functions that were tested: Euclidean distance (ED) and including the distance map weights ($DistMap$).

Table 2
Identification of Vertebra Segmentation 3D U-Net implementations. Data Pro. - whether problem-specific data processing method were used or not; GT - whether ground truth was used or not; Weights - whether distance map weights were used or not; Soft Outp. - whether soft outputs of teacher and student networks were used or not; Logits - whether the training used the logits; T - temperature value; λ - weight hyperparameter on loss function.

ID	Data Pro.	GT	Weights	Soft Outp.	Logits	T	λ
T_B	No	Yes	Yes	–	–	–	–
T_S[a]	Yes	Yes	Yes	–	–	–	–
S_S[a]	Yes	Yes	Yes	–	–	–	–
S_{T3}	Yes	Yes	Yes	Yes	No	3	1
S_{T4}	Yes	Yes	Yes	Yes	No	4	1
S_{T5}	Yes	Yes	Yes	Yes	No	5	1
S_L	Yes	No	No	No	Yes	–	–
S_{LW}	Yes	No	Yes	No	Yes	–	–

GT-Ground truth; T-Temperature.

[a] Scratch implementation of teacher(T_S)/student(S_S).

Teacher logits (S_L) and an updated version that included distance-to-border penalization weights (S_{LW}). The loss functions for these approaches involved calculating the Euclidean distance between the two predictions for S_L training ($ED(Teacher_{logits}, Student_{logits})$) and, for S_{LW} training, incorporating distance to vertebra's border penalization ($ED(Teacher_{logits}, Student_{logits}) \times DistMap$).

Table 2 provides a summary of the different Vertebra Segmentation 3D U-Net implementations, encompassing scratch implementations of both the Teacher and Student networks trained with the same methodology.

4. Results

The different implementations of the Vertebra Segmentation 3D U-Net were compared through an analysis of four benchmark metrics: (i) DSC to measure the overlap volume; (ii) HD to measure the maximum distance between predicted and ground truth surfaces; (iii) ASSD to gauge the average distance between predicted and ground truth surfaces in both directions (GT-Surf and Surf-GT); and (iv) failure rate (Fail) to specify the percentage of vertebrae segmented with a DSC lower than 50%. Additionally, the analysis included the number of floating-point operations (FLOPs), memory required, inference times and average segmentation times calculated using a standard CPU (AMD Ryzen^{TM} 5 3600). The carbon (CO_2) emissions during the segmentation of all test scans in the VerSe dataset were also tracked using the CodeCarbon package (Courty et al., 2023) to enable a comparison of the environmental footprint with other implementations in the literature.

4.1. VerSe19 and VerSe20 datasets

Table 3 presents the performance metrics of the Student and Teacher networks and Analysis of Variance (ANOVA) between the performance of various implementations of the Student network and the best Teacher network performance. Without the use of problem-specific data processing methods, the Teacher Vertebra Segmentation 3D U-Net (T_B) achieved an average DSC of 80.78%, HD of 16.34 mm, ASSD GT-Surf of 1.02 mm, ASSD Surf-GT of 0.93 mm, and a failure rate of 5.28%. The scratch Teacher network (T_S), trained with problem-specific data processing methods, demonstrated the best results, with an average DSC of 88.22%, HD of 7.71 mm, ASSD GT-Surf of 0.59 mm, ASSD Surf-GT of 0.55 mm, and a failure rate of 2.12%.

The Student scratch implementation (S_S) achieved an average DSC of 75.78%, HD of 15.17 mm, ASSD GT-Surf of 2.41 mm, ASSD Surf-GT of 0.64 mm, and a failure rate of 15.44%. Training the Student Vertebra Segmentation 3D U-Net with knowledge distillation using softened outputs (S_{T4} and S_{T5}) resulted in the best performance, with S_{T4} achieving an average DSC of 84.70%, HD of 9.82 mm, ASSD GT-Surf of 0.67 mm, ASSD Surf-GT of 0.85 mm, and a failure rate of 5.50%. The S_{T5} approach showed similar performance, with a DSC of 84.47%, HD of 8.08 mm, ASSD GT-Surf of 0.73 mm, ASSD Surf-GT of 0.51 mm, and a failure rate of 6.06%. Despite not reaching the same performance as S_{T4} and S_{T5}, the S_{T3} approach improved the performance of the S_S implementation, achieving an average DSC of 82.26%, HD of 11.05 mm, ASSD GT-Surf of 0.53 mm, ASSD Surf-GT of 1.20 mm, and a failure rate of 7.05%.

Matching logits (S_L) resulted in poor performance, with an average DSC of 29.60%, HD of 44.78 mm, ASSD GT-Surf of 1.49 mm, ASSD Surf-GT of 10.90 mm, and a high failure rate of 67.35%. However, including distance map weights (S_{LW}) significantly improved results,

Table 3
Quantitative comparison of the different Vertebra Segmentation 3D U-Net on the VerSe datasets. Table reports average values and STD. Inside the square brackets, is shown the p-value of the ANOVA between each Student network with the best Teacher network (T_S)..

				ASSD (mm)		
	ID	DSC (%)	HD (mm)	GT-Surf	Surf-GT	Fail (%)
Teacher	T_B	80.78 ± 20.43	16.34 ± 12.04	1.02 ± 1.24	0.93 ± 1.13	5.28
	T_S	**88.22** ± **13.07**	**7.71** ± **6.82**	**0.59** ± **0.78**	**0.55** ± **1.05**	**2.12**
Student	S_S	75.78 ± 28.48 [$p < 0.001$]	15.17 ± 16.05 [$p < 0.001$]	2.41 ± 5.22 [$p < 0.001$]	0.64 ± 1.01 [$p = 0.900$]	15.44
	S_{T3}	82.26 ± 21.37 [$p < 0.001$]	11.05 ± 9.42 [$p < 0.001$]	0.63 ± 0.57 [$p = 0.900$]	1.20 ± 2.42 [$p < 0.001$]	7.05
	S_{T4}	**84.70** ± **17.87** [$p = 0.005$]	9.82 ± 8.45 [$p < 0.001$]	**0.67** ± **0.69** [$p = 0.900$]	0.85 ± 1.91 [$p = 0.009$]	**5.50**
	S_{T5}	84.47 ± 21.01 [$p = 0.002$]	**8.08** ± **6.77** [$p = 0.900$]	0.73 ± 1.13 [$p = 0.5974$]	**0.51** ± **0.87** [$p = 0.900$]	6.06
	S_L	29.60 ± 30.20 [$p < 0.001$]	44.78 ± 8.88 [$p < 0.001$]	1.49 ± 1.06 [$p < 0.001$]	10.90 ± 3.96 [$p < 0.001$]	67.35
	S_{LW}	78.94 ± 21.45 [$p < 0.001$]	14.52 ± 10.63 [$p < 0.001$]	1.08 ± 1.54 [$p < 0.001$]	1.57 ± 2.17 [$p < 0.001$]	8.89

GT-Ground truth surface; Surf-Predicted surface.

Table 4
Quantitative comparison of our implementations with the approaches that used the VerSe datasets, including segmentation times in minutes (min.) and carbon emission (CO_2 Emi.) in grams (g).

Author	Seg. Time	CO_2 Emi.	VerSe19		VerSe20	
	(min.)	(g)	DSC (%)	HD (mm)	DSC (%)	HD (mm)
Lessmann et al. (2019)	–	278.80	85.76	8.20	66.96	–
Payer et al. (2020)	–	98.06	89.80	7.08	89.71	6.06
Chen D. et al.	–	421.25	86.44	–	**91.23**	7.15
(Altini et al., 2021)[a]	–	–	–	–	89.17	–
Tao et al. (2022)	–	260.68	89.80	6.35	–	–
Meng et al. (2023)	26[b]	–	**90.84**	–	91.11	6.69
T_S (Ours)	**3**	**44.83**	89.43	9.36	88.23	7.47
S_{T4} (Ours)	**2**	**25.57**	82.79	11.95	84.98	9.47
S_{T5} (Ours)	**2**	**25.57**	84.46	8.60	84.33	7.81

[a] Uses only 50 scans of VerSe20.
[b] On Quadro RTX 5000 GPU (16Gb).

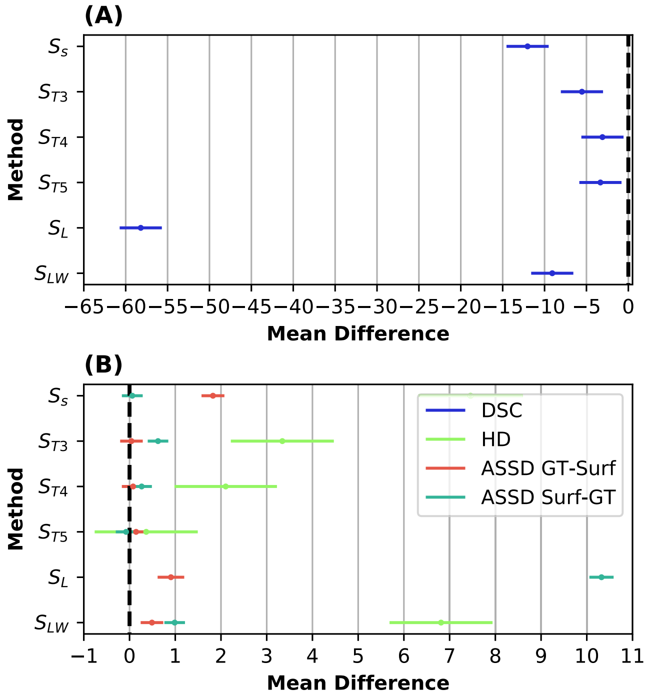

Fig. 5. Results of the Tukey HSD test on DSC, HD and ASSD of the different Student network implementations, when compared with the Teacher network results. A) Tukey 95% confidence interval of DSC; B) Tukey 95% confidence intervals of HD and ASSD. Due to the HD of the S_L implementation is too high (37.07) it has not been included in order not to hinder the visualization of the results.

achieving an average DSC of 78.94%, HD of 14.52 mm, ASSD GT-Surf of 1.08 mm, ASSD Surf-GT of 1.57 mm, and a failure rate of 8.89%. The performance of S_{LW} surpassed that of SS.

The Tukey Honestly Significant Difference (HSD) test allowed to analyze the difference between the means of the Teacher Vertebra Segmentation 3D U-Net (T_S) and each implementation of the Student network, in order to conclude which implementation presents similar performance to the Teacher network. According to DSC intervals (Fig. 5A), none of the implementations obtained similar performance to the Teacher. On the other hand, according to the other metrics (Fig. 5B) the performance of the Student network trained using the soft outputs and the GT with a Temperature of 5 (S_{T5}) achieved a performance that is not significantly different from the Teacher network (HD: $p = 0.90$; ASSD GT-Surf: $p = 0.59$; ASSD Surf-GT: $p = 0.90$) since the 95% confidence intervals on these metrics include the value of 0.

Fig. 6 shows the qualitative segmentation results of two subjects with different implementations of the Vertebra Segmentation 3D U-Net. In the case of subject ‘verse809’, the application of KD techniques enabled the Student network (S_{T4} and S_{T5}) to segment the cervical vertebrae. Regarding subject ‘verse563’, the scratch Teacher and Student networks (T_S and S_S) performed an erroneous segmentation by oversegmenting some vertebrae or segmenting two vertebrae as if they were the same. Through KD, S_{T4} and S_{T5} networks were able to correctly segment the vertebrae, mostly by improving the segmentation of the scratch implementation of the Student (S_S).

4.2. Benchmark

Table 4 provides a comparative analysis of our method with other works that used the VerSe dataset. Using the Teacher Vertebra Segmentation 3D U-Net (T_S) trained with proposed data processing techniques, our algorithm achieved the fourth and fifth positions on the VerSe19 and VerSe20 datasets, respectively, for individual vertebrae segmentation. The Teacher network demonstrated consistent performance on both datasets, with a DSC of 89.43% on VerSe19 and 88.23% on VerSe20, indicating good generalization. These results exhibit a close approximation to other state-of-the-art methods in the field. Specifically, on the VerSe19 dataset, our model’s results are marginally lower, showing only a 1.41% decrease in effectiveness. Similarly, when

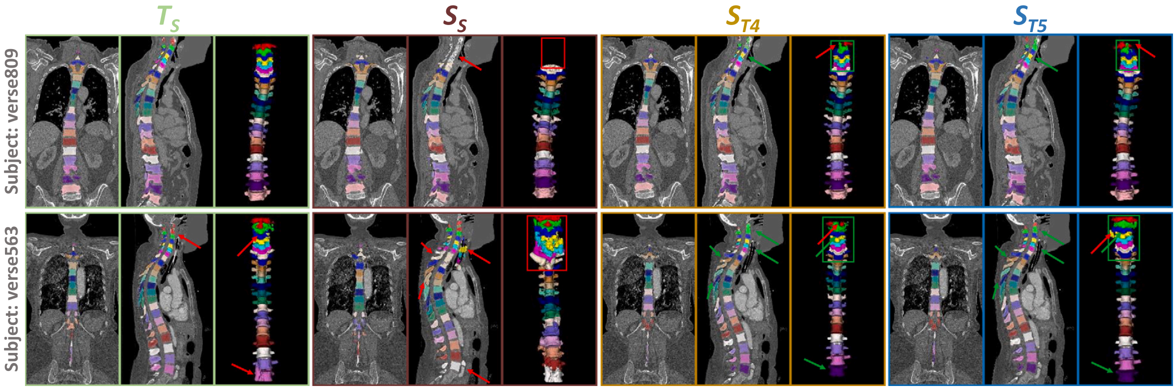

Fig. 6. Qualitative results for patients where KD had the most impact. Red arrows/boxes indicate erroneous segmentation, green arrows/boxes indicate correction of erroneous segmentation.

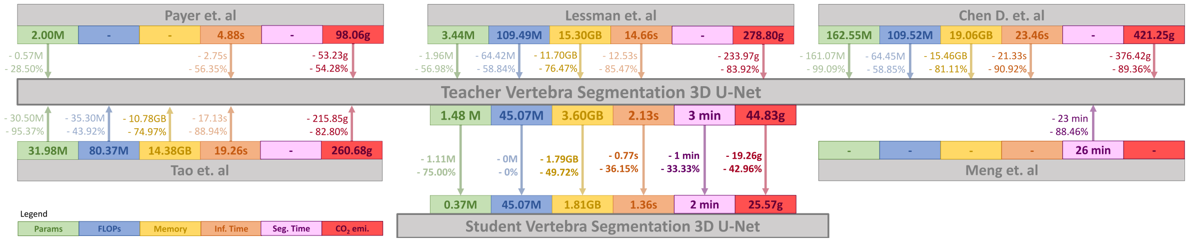

Fig. 7. Comparison of our Teacher and Student Vertebra Segmentation 3D U-Net to other 3D U-Nets found in the literature for individual vertebrae segmentation on CT. The comparison is made considering the number of parameters (green), FLOPs (blue), required memory (yellow), inference time (orange), segmentation time (purple) and CO_2 emissions (red). Arrows point to the shallower network, indicating the difference between the networks in the respective metric (absolute difference and percentage).

Table 5
Comparison of our neural networks architecture with the approaches that used the VerSe datasets, including methods, input size, depth, features per level, number of parameters (Params) in millions (M), floating-point operations (FLOPs) in millions (M), required memory (Memory Req.) for inference in gigabytes (GB) and inference times (Inf. time) in seconds (s).

Author	Methods	Input Size	Depth	Features (per level)	Params (M)	FLOPs (M)	Memory Req. (GB)	Inf. time (s)
Lessmann et al. (2019)	3D U-Net	[128, 128, 128, 2]	4	84-84-84-84	3.44	109.49	15.30	14.66
Payer et al. (2020)	SL: 3D U-Net VL: Spatial-Configuration Net VS: 3D U-Net	[64, 64, 128] – [128, 128, 96, 2]	5 – 5	64-64-64-64-64 – 64-64-64-64-64	2.33 – 2	27.38 – –	2.17 – –	1.67 – 4.88
Chen D. et al.	SL: 3D U-Net VS: 3D U-Net	[64, 64, 128, 1] [128, 128, 128, 2]	5 5	8-16-32-64-128 64-128-256-512-512	1.47 162.55	27.38 109.52	0.36 19.06	0.36 23.46
Altini et al. (2021)	VS: 3D V-Net	[64, 64, 64, 1]	–	–	–	–	–	–
Tao et al. (2022)	VL: Spine-Transformers VS: 3D U-Net	– [144, 144, 96, 2]	– 4	– 64-128-256-512	– 31.98	– 80.37	– 14.38	– 19.26
Meng et al. (2023)	Cyclic SS-VS-VI	–	–	–	–	–	–	–
Our Teacher	SL: 3D U-Net VS: 3D U-Net	**[64, 64, 128, 1]** **[128, 128, 128, 2]**	**4** **4**	**8-16-32-64** **16-32-64-128**	**0.37** **1.48**	**27.37** **45.07**	**0.36** **3.60**	**0.25** **2.13**
Our Student	SL: 3D U-Net VS: 3D U-Net	**[64, 64, 128, 1]** **[128, 128, 128, 2]**	**4** **4**	**8-16-32-64** **8-16-32-64**	**0.37** **0.37**	**27.37** **45.07**	**0.36** **1.81**	**0.25** **1.36**

SL-Spine Location; VL-Vertebrae Location; VS-Vertebra Segmentation; SS-Spine Segmentation; VI-Vertebra Identification.

evaluated against the VerSe20 dataset, the model demonstrates a slight reduction in performance, being only 3% less effective than the current best-performing methods. However, our approach notably reduced segmentation time by 88.46% (3 min vs. 26 min) and CO_2 emissions by 54.28% to 89.36% compared to other methods (44.83 g vs. 98.06 g, 260.68 g, 278.80 g and 421.25 g).

The Student network, trained with $T = 4$, achieved a DSC of 82.79% and HD of 11.95 mm on the VerSe19 dataset, while on the VerSe20 dataset, it achieved a DSC of 84.98% and HD of 9.47 mm. Using $T = 5$ to distill knowledge from the Teacher network, the Student network achieved a DSC of 84.46% and HD of 8.60 mm on VerSe19 and a DSC of 84.33% and HD of 7.81 mm on VerSe20. The Student network demonstrated efficient segmentation time, completing the task in 2 min with an associated CO_2 emission of 25.57 g, indicating a 33.33% and 42.96% improvement compared to the Teacher network.

Table 5 presents a comparison of the architecture of our neural networks with other methods in the literature. Regarding the Spine Location 3D U-Net, our network has one less depth level (4 vs. 5) and fewer features per level, totaling 0.35M parameters. This is 1.98M parameters less than (Payer et al., 2020) and 1.12M parameters less than Sekuboyina et al. (2021), resulting in a shorter inference time (0.25 s vs. 1.67 s and 0.36 s). Our Spine Location 3D U-Net executes 27.37M FLOPs and requires 0.36 GB of memory, comparable to Sekuboyina et al. (2021). Payer et al. (2020) shows a similar number of FLOPs (27.38M) but requires 2.17 GB of memory.

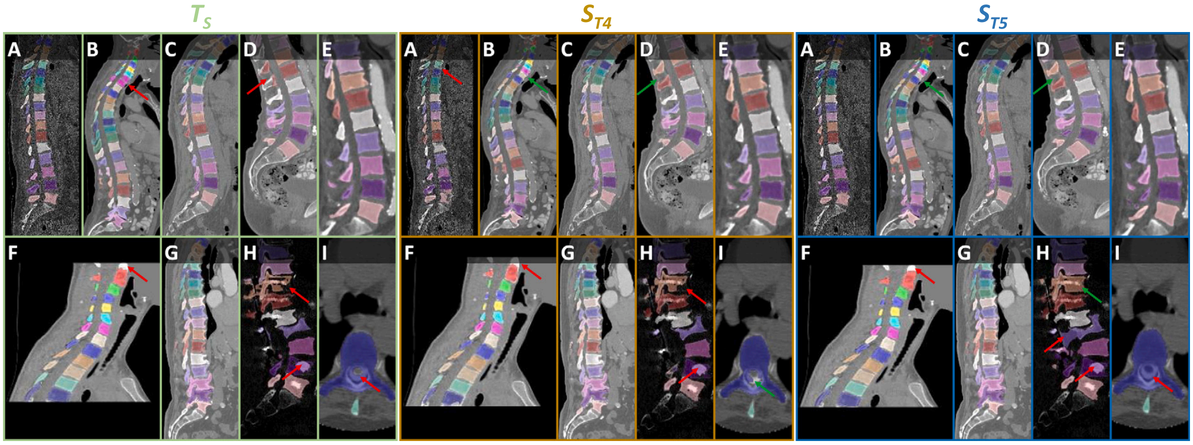

Fig. 8. Qualitative results of the teacher neural network (T_S) and student networks (S_{T4}, S_{T5}) on poor quality (A), thoracic kyphosis (B, C), lumbar lordosis (D), angiography (E), cervical kyphosis (F), osteophytosis (G), instrumentation (H) and myelography (I) scans.

The impact of our methods on the Vertebra Segmentation 3D U-Net is evident in Fig. 7. The Teacher Vertebra Segmentation 3D U-Net, composed of 4 levels and 1.48M parameters, represents a reduction of 0.57M to 161.02M parameters (−28.50% to −99.09%). This reduction leads to a shorter inference time (−56.35% to −90.92%) compared to other methods. Corresponding to 45.07M FLOPs and requiring 3.60 GB of memory, our model achieves a reduction between 35.3M and 64.45M FLOPs (−43.92% to 58.85%) and 10.78 GB to 15.46 GB (−74.97% to −81.11%) of required memory.

The Student network, with fewer features per level than the Teacher, comprises 0.37M parameters, 45.07M FLOPs, requires 1.81 GB of memory, and has an inference time of 1.36 s. This represents an improvement of 75.00%, 0%, 49.72%, and 36.15% compared to the Teacher network in the number of parameters, FLOPs, required memory and inference time, respectively.

5. Discussion

In recent years, efforts have concentrated on enhancing vertebrae segmentation in CT scans, typically by augmenting the 3D U-Net and introducing more layers to its architecture. Our approach, inspired by Sekuboyina et al. (2021), sought to explore whether shallower networks could rival larger ones commonly employed in vertebrae segmentation. We developed an algorithm for segmenting individual vertebrae in CT scans, using 3D U-Net architectures with reduced depth levels and output features per level. This algorithm comprises two steps: (i) spine localization on the CT scan aided by a 3D U-Net (Spine Location 3D U-Net); and (ii) iterative segmentation of the vertebrae from top to bottom using another 3D U-Net (Vertebra Segmentation 3D U-Net). We investigated the application of KD in the context of individual vertebrae segmentation on CT scans for the first time, training a Teacher Vertebra Segmentation 3D U-Net smaller than those in existing literature and then transferring its knowledge to an even smaller Student Vertebra Segmentation 3D U-Net. This objective aligns with the European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence (AI), emphasizing the critical evaluation of resource use and energy consumption throughout the development, deployment, and usage processes, advocating for less harmful choices (European Commission and Directorate-General for Communications Networks, Content and Technology, 2019).

To enhance the performance of smaller neural networks, two strategies were devised to address identified weaknesses in the algorithm reported by other authors. Vertebrae resampling played a crucial role in balancing the dataset, enabling the network to accurately segment the less-represented vertebrae. To tackle the issue of incorrect segmentation of the first vertebrae in CT scans, we proposed cleaning the memory instance during network training to replicate scenarios involving the segmentation of the first vertebra. The results highlight the positive impact of the resampling and memory cleaning techniques on the performance of the Teacher Vertebra Segmentation 3D U-Net (DSC without data processing: 80.78%; DSC with data processing: 88.22%), overcoming challenges associated with the segmentation of cervical vertebrae and the initial vertebrae in CT scans. The Teacher network achieved performance similar to other methods in the literature (DSC: 89.43% vs. 90.43% on VerSe19; 88.23% vs. 91.23%), despite having fewer layers and parameters. Notably, our Teacher network demonstrated faster inference and segmentation times, less FLOPs and required memory and a lower associated CO_2 emission rate than any other reported neural network, corresponding to a reduction of up to 90.92%, 88.46%, 58.85%, 88.94% and 89.36% in these metrics, respectively. These results emphasize that achieving the same performance does not necessarily require larger networks.

To distill knowledge from the Teacher to a Student Vertebra Segmentation 3D U-Net, five approaches were tested, including the use of soft outputs from both networks and the GT, or matching their logits. The most effective method involved using the soft outputs with the GT, particularly employing a Temperature of 4 or 5 to obtain the soft outputs. Results on VerSe19 and VerSe20 datasets suggest that knowledge distillation with$T = 5$ exhibits consistent performance on both datasets (84.46% on VerSe19 and 84.33% on VerSe20), while using$T = 4$ shows a 2% difference in DSC between the datasets (82.79% on VerSe19 and 84.98% on VerSe20). Although not matching the performance of the Teacher network, the results achieved by the Student network through knowledge distillation are promising, surpassing the performance of the scratch implementation of the Student (Scratch: 75.78%; KD: 84.70%). Compared with the Teacher's performance, the Student network shows a 4% reduction in terms of DSC, but opting for the Student network results in a 75%, 36%, 33%, and 42.96% reduction in the number of parameters, inference time, total segmentation time, and CO_2 emissions, respectively. Thus, KD allowed to achieve a balance between better performance and the Student network architecture, contributing to reduced environmental impact as outlined in the European Commission's Ethics Guidelines for Trustworthy AI and suitability for emergency cases involving low computational devices in medical institutions.

5.1. Generalization and limitations

From Fig. 6, it is evident that there are challenges in segmenting cervical vertebrae. This difficulty may arise due to the differences in size and shape compared to thoracic and lumbar vertebrae. Additionally, the junction between the C1 and C2 vertebrae is unique, as the C2 vertebra has the odontoid process that traverses C1 to facilitate the movement of the human head. This specific anatomy can contribute to a lower segmentation performance when dealing with these two vertebrae, as illustrated in the bottom part of Fig. 6.

The Teacher (T_S) and Student (S_{T4}, S_{T5}) Vertebra Segmentation 3D U-Net demonstrated the ability to segment most of the scans, irrespective of the pathologies present. Fig. 8 displays the segmentation results from the Teacher and Student networks on specific test scans from the VerSe dataset, including cases with low image quality, kyphotic/lordotic conditions, angiographs, myelograms, osteophytosis, and instrumentation. Both the Teacher and Student networks successfully segmented the vertebrae without major errors in scenarios involving low-quality scans, angiograms, and the presence of osteophytes (Fig. 8A,E,G). However, it is worth noting that S_{T4} undersegmented some vertebrae in poor-quality scans (as indicated by the red arrow in Fig. 8A for S_{T4}).

In cases of thoracic kyphosis, the Teacher network demonstrates the ability to correctly segment the vertebrae, with some errors occurring when there are more inclined vertebrae (Fig. 8B of T_S). However, the Student network proves effective in correcting these cases (green arrows in Fig. 8B for S_{T4} and S_{T5}). Concerning cases of lumbar lordosis, the Teacher network segmented two spinous processes in the displayed case (Fig. 8D of T_S), a situation that is corrected by the Student network (green arrows in Fig. 8D for S_{T4} and S_{T5}). Another pathology present in the dataset is cervical kyphosis, characterized by a curvature of the cervical spine in the opposite direction to the considered normal cases (lordosis). In these cases, both the Teacher and Student networks correctly segmented the vertebrae, despite the aforementioned difficulty in segmenting the C1-C2 junction (Fig. 8F).

In cases of instrumentation, both the Teacher and S_{T4} networks exhibited undersegmentation at T11 (orange vertebra in Fig. 8H of T_S and S_{T4}) and oversegmentation at L4 (purple in Fig. 8H T_S and S_{T4}), while S_{T5} oversegmented L2 and L4 (red arrows in Fig. 8H of S_{T5}). In the case of myelograms, S_{T4} correctly segmented the vertebrae, while the Teacher network and S_{T5} considered the spinal cord to be part of the vertebra and segmented it as well, which should not have occurred (red arrows in Fig. 8I of T_S and S_{T5}).

Despite its lower performance compared to the Teacher network, the Student network demonstrated the ability to correctly segment some vertebrae that were missegmented by the Teacher, highlighting the capacity of KD methods to empower smaller networks with information from both the Teacher network and the GT.

6. Conclusion

The developed algorithm for individual vertebrae segmentation in CT scans, based on 3D U-Net architecture, has demonstrated high accuracy while maintaining low memory and computational resource requirements. The incorporation of data processing techniques during neural network training proved crucial in enhancing the performance of smaller networks. Additionally, the application of KD techniques in the context of individual vertebrae segmentation on CT scans was explored, resulting in a smaller 3D U-Net with high performance. The algorithm achieved performance comparable to existing methods in the literature, making it suitable for deployment in medical devices with limited computational resources and memory, and potentially aiding in emergency cases where rapid segmentation is crucial.

For future work, it would be valuable to conduct further investigations on different values for Temperature and loss function weight hyperparameters (T and λ) in the KD methods proposed. Exploring the reliability and performance of alternative KD methods, such as transferring knowledge using logits and GT, could provide additional insights. Additionally, repeating the neural network training with input CT scans clipped to bone range values might be explored to determine if it leads to performance improvement.

Funding

This work has been supported by the FCT–Fundação para a Ciência e Tecno-logia – national funds, under the scholarship reference PD/BD/140389/2018.

CRediT authorship contribution statement

Luís Serrador: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Francesca Pia Villani:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing. **Sara Moccia:** Conceptualization, Resources, Supervision, Writing – review & editing. **Cristina P. Santos:** Resources, Supervision, Writing – review & editing,.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Altini, N., Giosa, G.D., Fragasso, N., Coscia, C., Sibilano, E., Prencipe, B., Hussain, S.M., Brunetti, A., Buongiorno, D., Guerriero, A., Tatò, I.S., Brunetti, G., Triggiani, V., Bevilacqua, V., 2021. Segmentation and identification of vertebrae in CT scans using CNN, k-means clustering and k-NN. Informatics 8 (2), 40. http://dx.doi.org/10.3390/informatics8020040.

Budennyy, S.A., Lazarev, V.D., Zakharenko, N.N., Korovin, A.N., Plosskaya, O.A., Dimitrov, D.V., Akhripkin, V.S., Pavlov, I.V., Oseledets, I.V., Barsola, I.S., Egorov, I.V., Kosterina, A.A., Zhukov, L.E., 2022. eco2AI: Carbon emissions tracking of machine learning models as the first step towards sustainable AI. Dokl. Math. 106 (S1), S118–S128. http://dx.doi.org/10.1134/s1064562422060230.

Chen, K., Zhai, X., Wang, S., Li, X., Lu, Z., Xia, D., Li, M., 2023. Emerging trends and research foci of deep learning in spine: bibliometric and visualization study. Neurosurg. Rev. 46 (1), 81. http://dx.doi.org/10.1007/s10143-023-01987-5.

Choi, J.W., 2022. Knowledge distillation from cross teaching teachers for efficient semi-supervised abdominal organ segmentation in CT. In: Fast and Low-Resource Semi-Supervised Abdominal Organ Segmentation. Springer Nature, Switzerland, pp. 101–115. http://dx.doi.org/10.1007/978-3-031-23911-3_10.

Conze, P.-H., Andrade-Miranda, G., Singh, V.K., Jaouen, V., Visvikis, D., 2023. Current and emerging trends in medical image segmentation with deep learning. IEEE Trans. Radiat. Plasma Med. Sci. 7 (6), 545–569. http://dx.doi.org/10.1109/trpms.2023.3265863.

Courty, B., Schmidt, V., Goyal-Kamal, MarionCoutarel, Feld, B., Lecourt, J., SabAmine, kngoyal, Léval, M., Cruveiller, A., ouminasara, Zhao, F., Joshi, A., Bogroff, A., De Lavoreille, H., Laskaris, N., LiamConnell, Saboni, A., Blank, D., Wang, Z., Inimaz, Catovic, A., Michał, S., alencon, JPW, MinervaBooks, SangamSwadiK, Hervé, M., brotherwolf, Pollard, M., 2023. mlco2/Codecarbon: v2.2.7. Zenodo, http://dx.doi.org/10.5281/ZENODO.8181237, URL https://zenodo.org/record/8181237.

Dou, Q., Liu, Q., Heng, P.A., Glocker, B., 2020. Unpaired multi-modal segmentation via knowledge distillation. IEEE Trans. Med. Imaging 39 (7), 2415–2425. http://dx.doi.org/10.1109/tmi.2019.2963882.

European Commission and Directorate-General for Communications Networks, Content and Technology, 2019. Ethics Guidelines for Trustworthy AI. Publications Office, http://dx.doi.org/10.2759/346720.

Guan, C., Wang, S., Liu, G., Liew, A.W.-C., 2019. Lip image segmentation in mobile devices based on alternative knowledge distillation. In: 2019 IEEE International Conference on Image Processing. ICIP, IEEE, http://dx.doi.org/10.1109/icip.2019.8803087.

He, T., Shen, C., Tian, Z., Gong, D., Sun, C., Yan, Y., 2019. Knowledge adaptation for efficient semantic segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, http://dx.doi.org/10.1109/cvpr.2019.00067.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., Pineau, J., 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. J. Mach. Learn. Res. 21 (248), 1–43, URL http://jmlr.org/papers/v21/20-312.html.

Holliday, A., Barekatain, M., Laurmaa, J., Kandaswamy, C., Prendinger, H., 2017. Speedup of deep learning ensembles for semantic segmentation using a model compression technique. Comput. Vis. Image Underst. 164, 16–26. http://dx.doi.org/10.1016/j.cviu.2017.05.004.

Janssens, R., Zeng, G., Zheng, G., 2018. Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging. ISBI 2018, IEEE, http://dx.doi.org/10.1109/isbi.2018.8363715.

Lachinov, D., Shipunova, E., Turlapov, V., 2020. Knowledge distillation for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer International Publishing, pp. 324–332. http://dx.doi.org/10.1007/978-3-030-46643-5_32.

Lannelongue, L., Grealey, J., Inouye, M., 2021. Green algorithms: Quantifying the carbon footprint of computation. Adv. Sci. 8 (12), http://dx.doi.org/10.1002/advs.202100707.

Lessmann, N., van Ginneken, B., de Jong, P.A., Išgum, I., 2019. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. Med. Image Anal. 53, 142–155. http://dx.doi.org/10.1016/j.media.2019.02.005.

Liebl, H., Schinz, D., Sekuboyina, A., Malagutti, L., Löffler, M.T., Bayat, A., Husseini, M.E., Tetteh, G., Grau, K., Niederreiter, E., Baum, T., Wiestler, B., Menze, B., Braren, R., Zimmer, C., Kirschke, J.S., 2021. A computed tomography vertebral segmentation dataset with anatomical variations and multi-vendor scanner data. Sci. Data 8 (1), http://dx.doi.org/10.1038/s41597-021-01060-0.

Ligozat, A.-L., Lefevre, J., Bugeau, A., Combaz, J., 2022. Unraveling the hidden environmental impacts of AI solutions for environment life cycle assessment of AI solutions. Sustainability 14 (9), 5172. http://dx.doi.org/10.3390/su14095172.

Liu, Y., Chen, K., Liu, C., Qin, Z., Luo, Z., Wang, J., 2019. Structured knowledge distillation for semantic segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, http://dx.doi.org/10.1109/cvpr.2019.00271.

Liu, Y., Zeng, F., Ma, M., Zheng, B., Yun, Z., Qin, G., Yang, W., Feng, Q., 2023. Bone suppression of lateral chest x-rays with imperfect and limited dual-energy subtraction images. Comput. Med. Imaging Graph. 105, 102186. http://dx.doi.org/10.1016/j.compmedimag.2023.102186.

Meng, D., Boyer, E., Pujades, S., 2023. Vertebrae localization, segmentation and identification using a graph optimization and an anatomic consistency cycle. Comput. Med. Imaging Graph. 107, 102235. http://dx.doi.org/10.1016/j.compmedimag.2023.102235.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision. (3DV), IEEE, http://dx.doi.org/10.1109/3dv.2016.79.

Nekrasov, V., Chen, H., Shen, C., Reid, I., 2019a. Fast neural architecture search of compact semantic segmentation models via auxiliary cells. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, http://dx.doi.org/10.1109/cvpr.2019.00934.

Nekrasov, V., Dharmasiri, T., Spek, A., Drummond, T., Shen, C., Reid, I., 2019b. Real-time joint semantic segmentation and depth estimation using asymmetric annotations. In: 2019 International Conference on Robotics and Automation. ICRA, IEEE, http://dx.doi.org/10.1109/icra.2019.8794220.

Noothout, J.M.H., Lessmann, N., van Eede, M.C., van Harten, L.D., Sogancioglu, E., Heslinga, F.G., Veta, M., van Ginneken, B., Išgum, I., 2022. Knowledge distillation with ensembles of convolutional neural networks for medical image segmentation. J. Med. Imaging 9 (05), http://dx.doi.org/10.1117/1.jmi.9.5.052407.

Park, S., Heo, Y.S., 2020. Knowledge distillation for semantic segmentation using channel and spatial correlations and adaptive cross entropy. Sensors 20 (16), 4616. http://dx.doi.org/10.3390/s20164616.

Payer, C., Štern, D., Bischof, H., Urschler, M., 2019. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. Med. Image Anal. 54, 207–219. http://dx.doi.org/10.1016/j.media.2019.03.007.

Payer, C., Štern, D., Bischof, H., Urschler, M., 2020. Coarse to fine vertebrae localization and segmentation with SpatialConfiguration-net and U-net. In: Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. SCITEPRESS - Science and Technology Publications, http://dx.doi.org/10.5220/0008975201240133.

Qi, Y., Zhang, W., Wang, X., You, X., Hu, S., Chen, J., 2022. Efficient knowledge distillation for brain tumor segmentation. Appl. Sci. 12 (23), 11980. http://dx.doi.org/10.3390/app122311980.

Qin, D., Bu, J.-J., Liu, Z., Shen, X., Zhou, S., Gu, J.-J., Wang, Z.-H., Wu, L., Dai, H.-F., 2021. Efficient medical image segmentation based on knowledge distillation. IEEE Trans. Med. Imaging 40 (12), 3820–3831. http://dx.doi.org/10.1109/tmi.2021.3098703.

Rahimpour, M., Bertels, J., Vandermeulen, D., Maes, F., Goffin, K., Koole, M., 2021. Improving T1w MRI-based brain tumor segmentation using cross-modal distillation. In: Landman, B.A., Išgum, I. (Eds.), Medical Imaging 2021: Image Processing. SPIE, http://dx.doi.org/10.1117/12.2581067.

Ren, G., Yu, K., Xie, Z., Wang, P., Zhang, W., Huang, Y., Wang, Y., Wu, X., 2022. Current applications of machine learning in spine: From clinical view. Glob. Spine J. 12 (8), 1827–1840. http://dx.doi.org/10.1177/21925682211035363.

Saw, S.N., Ng, K.H., 2022. Current challenges of implementing artificial intelligence in medical imaging. Phys. Med. 100, 12–17. http://dx.doi.org/10.1016/j.ejmp.2022.06.003.

Sekuboyina, A., Husseini, M.E., Bayat, A., Löffler, M., Liebl, H., Li, H., Tetteh, G., Kukačka, J., Payer, C., Štern, D., et al., 2021. VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images. Med. Image Anal. 73, 102166. http://dx.doi.org/10.1016/j.media.2021.102166.

Strubell, E., Ganesh, A., McCallum, A., 2019. Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/p19-1355.

Tao, R., Liu, W., Zheng, G., 2022. Spine-transformers: Vertebra labeling and segmentation in arbitrary field-of-view spine CTs via 3D transformers. Med. Image Anal. 75, 102258. http://dx.doi.org/10.1016/j.media.2021.102258.

Tseng, K.-K., Zhang, R., Chen, C.-M., Hassan, M.M., 2020. DNetUnet: A semi-supervised CNN of medical image segmentation for super-computing AI service. J. Supercomput. 77 (4), 3594–3615. http://dx.doi.org/10.1007/s11227-020-03407-7.

Wang, J., Gou, L., Zhang, W., Yang, H., Shen, H.-W., 2019. Deepvid: Deep visual interpretation and diagnosis for image classifiers via knowledge distillation. IEEE Trans. Vis. Comput. Graph. 25 (6), 2168–2180. http://dx.doi.org/10.1109/tvcg.2019.2903943.

Wang, J., Tang, Y., Wu, Z., Du, Q., Yao, L., Yang, X., Li, M., Zheng, J., 2023. A self-supervised guided knowledge distillation framework for unpaired low-dose CT image denoising. Comput. Med. Imaging Graph. 107, 102237. http://dx.doi.org/10.1016/j.compmedimag.2023.102237.

Xiong, F., Shen, C., Wang, X., 2023. Generalized knowledge distillation for unimodal glioma segmentation from multimodal models. Electronics 12 (7), 1516. http://dx.doi.org/10.3390/electronics12071516.

Xu, P., Kim, K., Koh, J., Wu, D., Lee, Y.R., Park, S.Y., Tak, W.Y., Liu, H., Li, Q., 2021. Efficient knowledge distillation for liver CT segmentation using growing assistant network. Phys. Med. Biol. 66 (23), 235005. http://dx.doi.org/10.1088/1361-6560/ac3935.

Xu, R., Wang, Y., Ye, X., Wu, P., Chen, Y.-W., Xu, F., Zhu, W., Chen, C., Zhou, Y., Hu, H., Qu, X., Kido, S., Tomiyama, N., 2022. Pixel-level and affinity-level knowledge distillation for unsupervised segmentation of Covid-19 lesions. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, http://dx.doi.org/10.1109/icassp43922.2022.9746715.

Yagi, M., Yamanouchi, K., Fujita, N., Funao, H., Ebata, S., 2023. Revolutionizing spinal care: Current applications and future directions of artificial intelligence and machine learning. J. Clin. Med. 12 (13), 4188. http://dx.doi.org/10.3390/jcm12134188.

You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S., 2022. SimCVD: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Trans. Med. Imaging 41 (9), 2228–2237. http://dx.doi.org/10.1109/tmi.2022.3161829.

Zhang, L., Feng, S., Wang, Y., Wang, Y., Zhang, Y., Chen, X., Tian, Q., 2022a. Unsupervised ensemble distillation for multi-organ segmentation. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, IEEE, http://dx.doi.org/10.1109/isbi52829.2022.9761568.

Zhang, F., Wang, M., Yang, H., 2022b. Self-training with selective re-training improves abdominal organ segmentation in CT image. In: Fast and Low-Resource Semi-Supervised Abdominal Organ Segmentation. Springer Nature, Switzerland, pp. 1–10. http://dx.doi.org/10.1007/978-3-031-23911-3_1.