mathematics

MDPI

Article

Enhanced Attention Res-Unet for Segmentation of Knee Bones

Daniel Aibinder [1], Matan Weisberg [1], Anna Ghidotti [2] and Miri Weiss Cohen [1,*]

[1] Department of Software Engineering, Braude College of Engineering, Karmiel 2161002, Israel; daniel.aibinder@e.braude.ac.il (D.A.); matan.weisberg@e.braude.ac.il (M.W.)
[2] Department of Management, Information and Production Engineering (DIGIP), University of Bergamo, Viale G. Marconi, 24044 Dalmine, BG, Italy; anna.ghidotti@unibg.it
* Correspondence: miri@braude.ac.il

Abstract: The objective of this study was to develop a U-net capable of generating highly accurate 3D models of knee bones, in particular the femur. As part of the approach, a U-net was designed, trained, and validated. In order to achieve these goals, a novel architecture was proposed, including an architecture that reduces encoder parameters and incorporates transfer learning, in order to enhance the attention U-net. Additionally, an extra depth layer was added to extract more salient information. Moreover, the model includes a classifier unit to reduce false positives, as well as a Tversky focal loss function, which is an innovative loss function. The proposed architecture achieved a Dice coefficient of 98.05. By using these enhanced tools, clinicians can visualize and analyze knee structures more accurately, improve surgical intervention effectiveness, and improve patient care quality overall.

Keywords: knee bone segmentation; deep learning; attention U-net

MSC: 68T07; 68T45

check for updates

Citation: Aibinder, D.; Weisberg, M.; Ghidotti, A.; Weiss Cohen, M. Enhanced Attention Res-Unet for Segmentation of Knee Bones. *Mathematics* **2024**, *12*, 2284. https://doi.org/10.3390/math12142284

Academic Editors: Monica Bianchini, Maria Lucia Sampoli, Simone Bonechi and Pietro Bongini

Received: 24 June 2024
Revised: 15 July 2024
Accepted: 18 July 2024
Published: 22 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Accurate segmentation of the knee from medical images is a critical component in the diagnosis, treatment, and management of knee-related conditions [1]. This process involves delineating various anatomical structures within the knee joint, such as bones, cartilage, and ligaments, which is essential for understanding joint health and pathology [2].

In clinical applications, precise knee segmentation is indispensable for several reasons [3,4]. Firstly, it enables detailed assessment of knee joint integrity, facilitating the diagnosis of conditions like osteoarthritis, meniscal tears, and ligament injuries. Accurate segmentation provides critical insights into the extent of damage, guiding clinicians in developing appropriate treatment plans [5]. Secondly, precise knee segmentation plays a crucial role in preoperative planning and postoperative evaluation [6]. Surgeons rely on detailed anatomical maps to plan procedures such as knee replacements, ligament reconstructions, and meniscal repairs.

Osteoarthritis is the most common form of arthritis. Symptoms include swelling, pain, and a reduction in range of motion when cartilage in the joints breaks down. The majority of osteoarthritis cases occur in the hands, feet, spine, and large weight-bearing joints, such as the hips and knees [7,8]. Joint cartilage deteriorates, causing the bones to rub against each other, resulting in pain and decreased mobility. The National Institutes of Health (NIH) [9] warn that OA can lead to permanent disability and deformity. Due to these changes, the affected area may lose range of motion and become less stable. The inability to perform everyday tasks, such as walking or carrying objects, can cause difficulty.

Currently, osteoarthritis has no cure, partly due to a lack of comprehensive understanding of the pathological mechanism of the initiation and progression of the disease [10].

Radiologists spend considerable time analyzing MRI images in order to diagnose osteoarthritis accurately [11]. As a result of this significant expenditure of time and energy,

patients with osteoarthritis incur a high cost of care and treatment. The artificial joint replacement procedure involves surgically removing the damaged joint and replacing it with an artificial joint made from metal, plastic, or ceramic [12]. By using an artificial joint, the function of the joint can be restored and pain can be reduced. Traditionally, this has been the only treatment option available for osteoarthritis of the knee, which has a lifetime risk of up to 40 percent [13]. Figure 1 depicts two knee joints; on the left is a normal healthy knee joint, and on the right is a knee joint with osteoarthritis.

Despite its importance, achieving accurate knee segmentation is challenging, due to the complex anatomy of the knee and the variability in imaging quality and protocols. Manual segmentation, although accurate, is time-consuming and prone to observer variability [14]. Automated methods offer a promising solution by providing rapid and consistent segmentation, with minimal manual intervention.

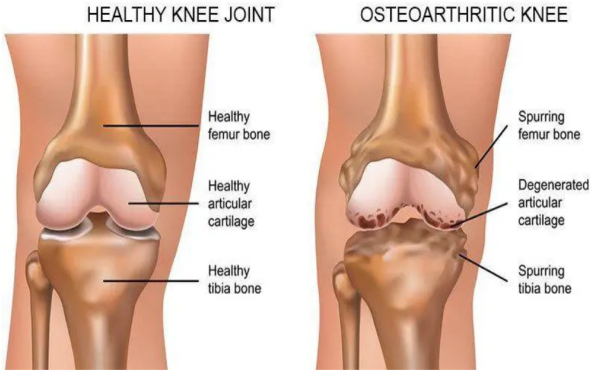


Figure 1. Healthy knee (**left**), osteoarthritis (**right**), [15].

The knee segmentation and marking process is currently performed in three different ways: manually, automatically, and semi-automatically.

- In the manual procedure, a trained doctor manually draws a specific label on an image portion—a lengthy and inefficient process.
- The automatic method involves the use of an algorithm that divides the image into regions with similar characteristics. However, this method does not provide the accuracy required at the present time.
- The semi-automatic method incorporates both techniques as a compromise, but it is limited by multiple 3D modelings [16] and has the disadvantages of both methods.

An accurate segmentation of the patient's knee can enhance the performance of joint replacement procedures. Nowadays, a high percentage of patients are dissatisfied after treatment [17], due to various factors, such as implant malalignment and incorrect implant size. A precise 3D model of the patient's knee can improve alignment and facilitate the design of customized solutions. Therefore, automating the process is essential. Ghidotti et al. [18] discussed the need to automate modeling procedures in order to reduce manual intervention and improve patient satisfaction with the reconstruction of a 3D model of the knee.

Currently, the field of medical engineering has made significant progress due to advances in fields such as three-dimensional imaging [19], neural networks [1], and additive manufacturing [20]. Additionally, research has been conducted on various segmentation techniques [21], as well as on the use of deep learning [22]. It is apparent that further research is required into various segmentation techniques, with deep learning being the most prevalent.

In orthopedics, 3D models provide a detailed representation of complex anatomy, enhancing the precision and personalization of medical interventions. Three-dimensional models can be used by surgeons to visualize the anatomy of each patient, identify pathology, and plan surgeries or implant placements in detail. Mercader et al. [23] used patient-specific 3D models to compare femoral rollback before and after total knee arthroplasty (TKA),

thereby providing knowledge regarding surgical outcomes. Additionally, 3D models can be integrated with human gait analysis to have a complete view of the patient, as demonstrated by Kerdvibulvech et al. [24]. Specifically focused on the femur, this study sought to develop a CNN capable of generating highly accurate 3D models of knee bones. By providing clinicians with better tools to visualize and analyze knee structures, it aims to improve the accuracy of diagnoses, the effectiveness of surgical interventions, and the overall quality of patient care.

2. Related Work

A key challenge in medical segmentation is the high degree of variability in the appearance of bones in medical images as a result of factors such as anatomy, imaging mode, and image quality. Thus, many research studies have focused on developing algorithms for segmentation. A detailed review can be found in [25,26].

One of the first widely used segmentation method was region growing, a type of segmentation that divides an image into regions and segments the image by the similarity of pixels within each region [27]. Deformable-model-based is another common approach to segmentation [28]. The most common models are active shape models (ASMs) [29], statistical shape models (SSMs) [30], and active appearance models (AAMs) [31]. As an example, ASM is a statistical model that combines a deformable shape model with a texture model to represent the shape and appearance of an object. Shape models capture global structure and variability, while texture models capture local appearance and details. A graph-based method for knee bone segmentation was developed in the following years. These models represent knee bones as nodes in a graph connected by edges [32]. Atlas-based models are another approach, Dam et al. [33] proposed a model that used multi-atlas registry before KNN based classification of the cartilage and that was manually trained in different bone combinations.

Machine learning (ML) and deep learning (DL) techniques have been developed in the last decade for segmenting knee bones [34]. With ML, pixel features are extracted from the given data, and segmentation labels are assigned to those pixels. Jamshidi et al. [35] and Klozyk and Matte [36] demonstrated the importance of ML techniques in detecting knee osteoarthritis at an early stage.

Due to advances in hardware, DL methods have become more powerful, and deeper learning techniques have been introduced in recent years. The use of automatic segmentation has been shown to be superior to the use of manual segmentation in some instances [21]. To perform semantic labeling on a two-dimensional knee image, Liu et al. [14] constructed a 10-layer SegNet framework and removed its completely connected layer after the decoder network. Using CNN and statistical shape models, Ambellan et al. [37] reported knee segmentation, but due to the resource-intensive segmentation and the small database it was trained on, the method struggled to scale to larger databases.

In recent years, the U-net architecture has been the most commonly used for deep learning segmentation [38]. Many examples can be found on using U-net for knee segmentation. Norman et al. [39] used U-net to segment the cartilage and meiscus compartments with Dice score results between 0.770 and 0.878 for the cartilage. A three-dimensional U-net with adversarial loss was researched by Chen et al. [3] to automatically segment the knee bone in a re-sampled image volume, which could result in the loss of fine details. A study by Almajalid et al. [40] successfully segmented all knee bones by separating each 3D image into slices for input to a 2D model. However, the dataset used was very small and had many false positives. Deniz et al. [41] performed automatic femur segmentation based on deep convolutional neural networks, achieving a Dice similarity score of 0.95; however, they did not optimize the learning rate and the number of initial feature maps.

Chen et al. [42] performed automatic femur segmentation using a 3D feature-enhanced network using edge detection and multi-scale feature fusion, achieving a 0.96 Dice score. A timeline of DL-based segmentation algorithms for 2D images is found in [43]. In some

instances, automatic segmentation has proven to outperform manual segmentation [21]. A detailed review of segmentation methods utilizing DL can be found in [44].

In this work, three contributions are made. First, a novel attention U-net was developed to solve the problem of segmentation in extreme visual performances of MRI images resulting in the elimination of false positive cases. In addition, the use of transfer learning as part of the training process of the U-net allows for less time and data consumption, thereby improving the constraints of the developed system. Finally, the use of different loss functions contributes to the accuracy.

3. Proposed Approach

An architecture that extends U-net is proposed in this paper, through which new modules are introduced and added to the attention unit [45]. Figure 2 illustrates the multiple operations carried out to construct the proposed U-net.

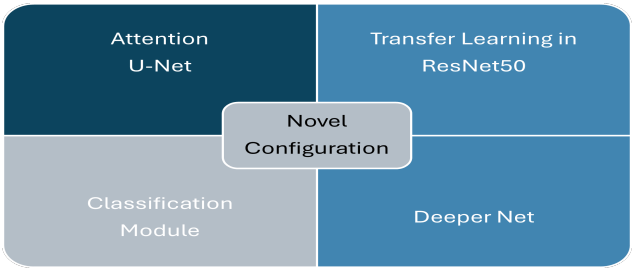


Figure 2. Proposed Approach.

In the following paragraphs, the added modules will be discussed and their contribution to the final extended network will be described.

3.1. Data Pre-Processing

In this study, the dataset comprises MRI scans of patients that showcase various views of the knee, in DICOM file format, accompanied by a corresponding segmentation of the patient's knee bones. The images are represented as gray scale images, with each slice containing 512 × 512 pixels, as seen in Figure 3.

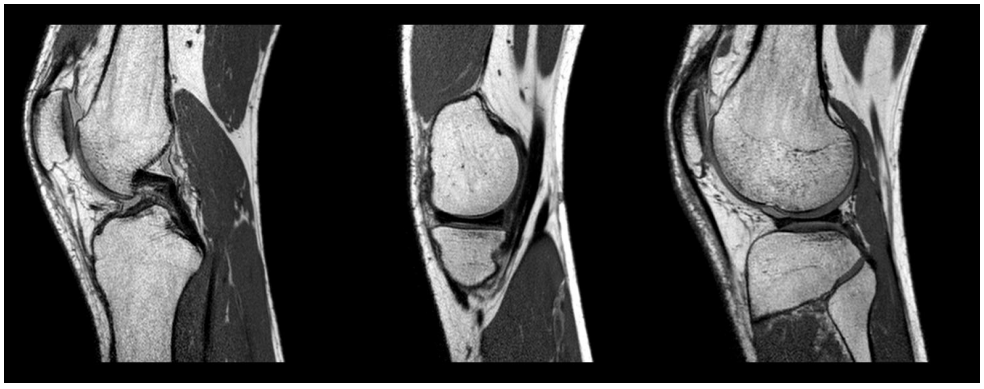
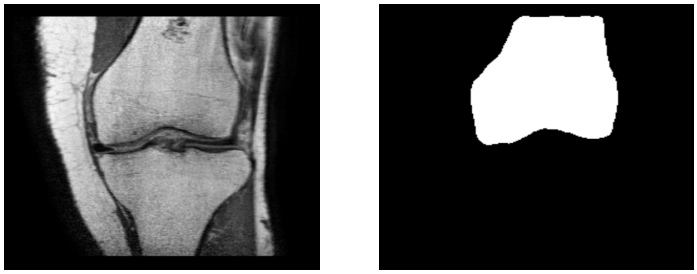


Figure 3. Dataset Examples of the knee bones images.

Based on the data processing methodology used, 12,984 image slices were used during the training process, with 3246 image slices for validation using a 25% split for validation. In order to train, each image was reduced to 256 × 256 pixels, and pixels in the image were represented using a continuous representation, so that each pixel value fell within the range of [0,1]. The image was also normalized using the Housefield unit. It was determined that bone had the highest density, while non-bone had the lowest density [46].

Considering the three-dimensional nature of each patient view, each slice could originate from an x, y, or z axis. As part of the training phase, only images from the x and z

axes were used. With this strategy, a frontal view of the knee was rendered, resulting in a smoother training process and avoiding the possibility of confusion between the tibia and femur. MRI scans of the knee contained 1024 images along the x and z axes; however, less than half of these images indicate the presence of bone. Consequently, our model was primarily trained with black images or images without bones and therefore lacked segmentation data, resulting in underfitting. Due to this predicament, the initial model was prone to producing false segmentation outputs. Therefore, we introduced a supplementary method for segregating our training data. During training, the image was retained if at least five percent of the pixels were white (indicating segmentation). In this procedure, any part of the image that does not contain the necessary data for the U-net to be trained is eliminated, as depicted in Figure 4. Although it is standard to shuffle your training set images, in our scenario, the model benefits from seeing the slices of the MRI in sequence, in order to learn a complete segmentation for each patient. To maintain the sequential order of images within each patient's MRI, we shuffled the order of patients in the training set.



(**a**) Input Image (**b**) True Mask

Figure 4. Example of full segmentation data, MRI scan (**a**) segmentation results (**b**).

3.2. Attention U-Net

Attention U-Net has been successfully employed for pancreas segmentation [47]. The foregoing network was optimized in order to improve the accuracy of segmentation.

The attention U-net combines the strength of U-nets with attention gate mechanisms (AG). It enables the network to concentrate on the most important parts of the input image by reducing noise and ambiguity, thus highlighting the most prominent features. Oktay et al. [47] added an AG before the U-net, joining the encoder and decoder's corresponding features, and readjusted the encoder's output features. As part of an analysis of an image, this refers to how much attention is given to different parts of the image, while indirectly learning to suppress uninteresting areas. The AG mechanism is seen in Figure 5.

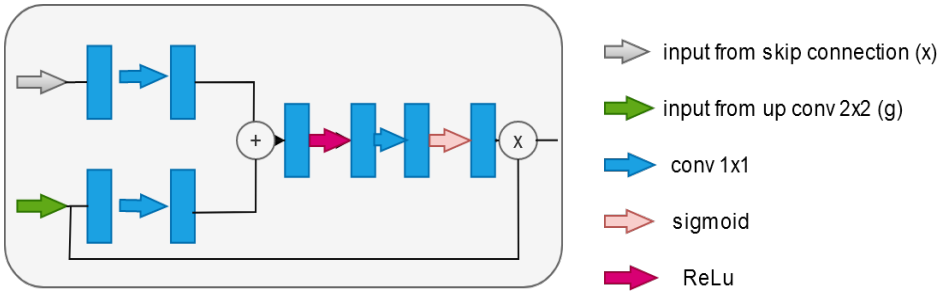


Figure 5. Attention date mechanisms (AG).

The AG is added to each skip connection that receives input from both the encoder feature map and the lower network feature map. The modifications proposed in this work are intended to enhance the existing attention U-net architecture, as can be seen in Figure 6.

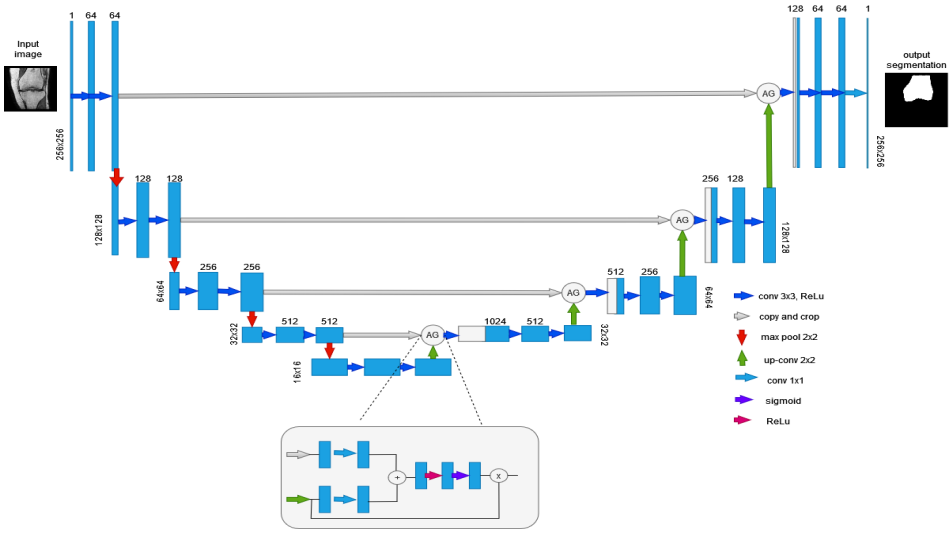


Figure 6. An attention gate is added to each skip connection layer, with input x from the feature map on the left and g from the lower network, the output is then concatenated to the decoder block.

3.3. Transfer Learning

Transfer learning (TL) is used in this work in two different architectures and modules: the first, in the U-net (encoder and decoder) and the second, in the classifier unit, as detailed in Section 3.4.

Bilal et al. [48] introduced TL for U-net networks using the VGG16 architecture, which focuses primarily on transferring encoder functionalities. In order to reduce the number of parameters, transfer learning was applied to both the encoder and decoder of this work using the ResNet50 architecture. ResNet50 has fewer parameters than VGG16, making it more efficient in terms of computation time and complexity. A significant reduction in parameters was achieved using ResNet50. Furthermore, ResNet50 was able to overcome the problem of vanishing gradients.

3.4. Classifier Unit

A substantial portion of volumetric images do not contain any segmentation information, as indicated in Section 3.1. Initially, we trained our network exclusively using segmented data. However, the model produced a significant number of false positive segmentations when it was asked to predict an entire volume consisting of only images without bone structure. Therefore, when we attempted to train the model using all slices of the volume, the segmentation accuracy decreased.

The second enhancement involved adding a new component to the U-net model. This component is called a classification module and functions as a binary classifier, as shown in Figure 7. Using binary cross-entropy, we trained our classifier to identify whether an image contains a femur representation. Once it has been determined that there is no femur in the image, a corrective measure is applied.

As depicted in Figure 8, the segmentation result is multiplied by the classifier result. If the classifier result is 0, indicating the image has no visible femur, the segmentation output is multiplied by zero, resulting in a black image. If the classifier result is 1, indicating the presence of a femur, the segmentation output is multiplied by one, leaving the U-Net model's output unchanged. It is important to note that in the case of a false-positive segmentation, this multiplication will result in a final image that lacks any discernible segmentation, appearing as a black image with no discernible regions of interest.

Segmentation problems inherently involve classification. For example, a model must learn to detect the presence of a femur bone in an image and, if present, identify its location before segmenting the bone. This task of classification and localization can be challenging, often leading to many false positives where the model segments a bone even when none

exists. In the medical field, such errors can be disastrous. To address this, we propose using a classification model to determine whether the segmentation model should process the image. Since binary classification tasks are typically easier to train and achieve high accuracy, we can rely on our classification model to make more accurate decisions about the presence of a bone in the image. This approach reduces the risk of false positives and enhances the overall reliability of the segmentation process. CNL does not add a significant number of parameters to the network, since it is based on a pre-trained VGG16 network that has been fine-tuned. It can be easily integrated into any existing network with little impact on training times, since it is trained concurrently with the segmentation network.

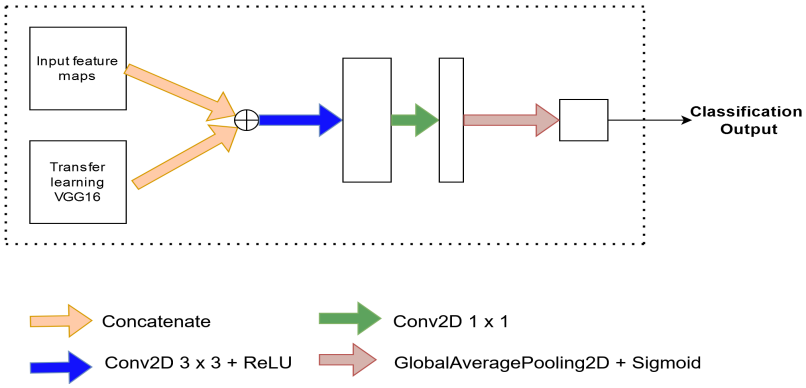


Figure 7. The architecture of the classifier, which takes two inputs, one using transfer learning from VGG16 and one from the bottom layer of the encoder, and produces the classification output.

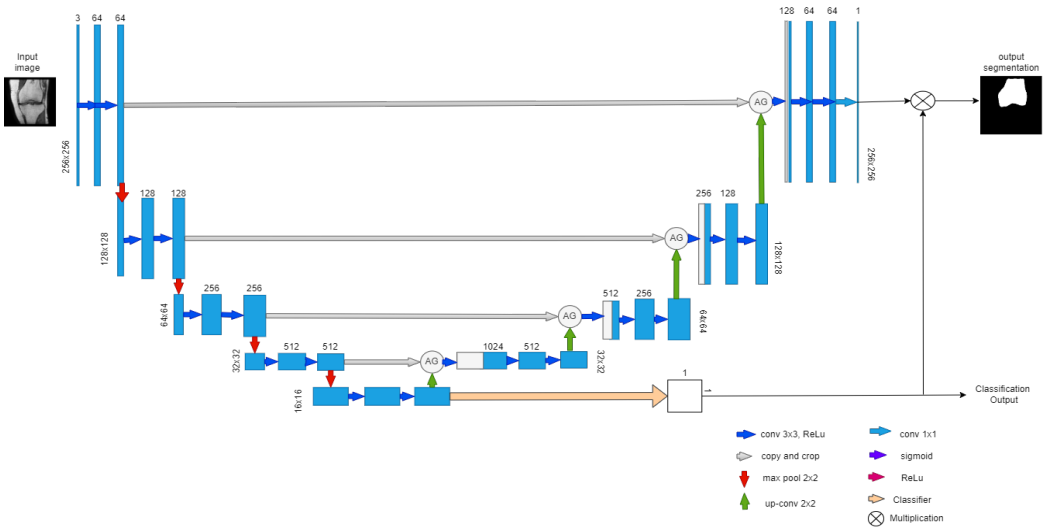


Figure 8. The proposed architecture with the addition of the classifier.

Due to the fact that VGG16 is a network that is trained specifically on the task of classification, our classifier uses it as a base, which should result in a more accurate classification. We further enhance the classification process by concatenating the feature maps from the final block of the VGG16 with the feature maps from the middle block of the U-net decoder. Then, a 3×3 convolution with a Relu activation function is performed, followed by a 1×1 convolution with a global average pooling and a Sigmoid activation function. The classifier output is multiplied by the attention output.

3.5. Deeper U-Net Levels

This work explores the possibility of increasing the network's depth by adding additional layers to the encoder and decoder, as shown in Figure 9. This enables capturing finer details and improves the overall segmentation accuracy. It is important to note that adding

an additional layer resulted in an increase in network parameters, from 35 M to over 90 M . To accommodate an extra layer, the convolution operations use the same number of features as the layer before, rather than increasing the number by two. Adding transfer learning from ResNet50, a lighter network, to the entire decoder further reduces the number of parameters.

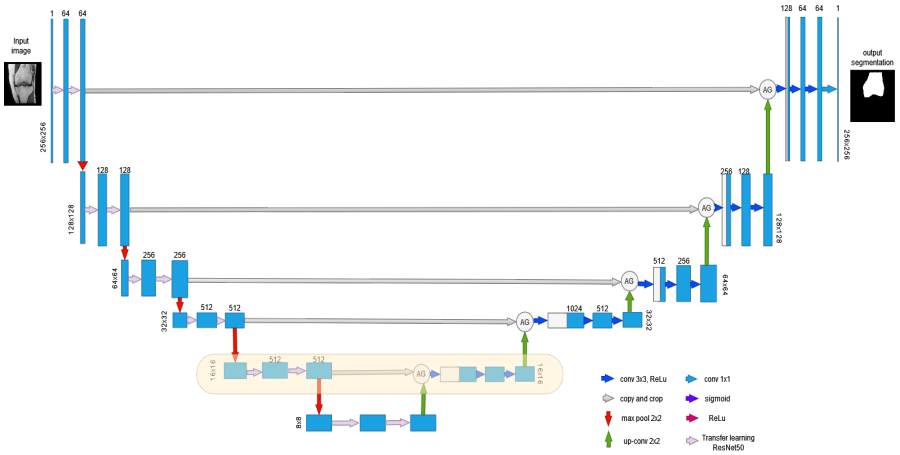


Figure 9. Proposed architecture with an additional layer added.

Defining the extra layer as such increased the network's parameters from 35 M to 49 M, a substantial reduction from the original 90 M parameters required by an extra layer.

3.6. Loss Functions

The primary objective of semantic segmentation is to anticipate the class labels for each pixel within an input image. In this context, we aim to identify a distinct area of interest, the bone, from MRI scans. This simplifies the task into a binary classification problem for individual pixels. However, our area of interest is typically confined to a relatively small portion of the image. In that scenario, the utilization of binary cross-entropy loss or similar metrics can deceive the model into exhibiting a false sense of superiority, for example classifying only foreground pixels and not the actual segmentation. Hence, we require a loss function that precisely represents our priorities. One such solution is the Dice coefficient loss, which quantifies the degree of overlap between the predicted segmentation and the ground truth. It is calculated as the ratio of the intersection of the two sets to the union of the two sets. The Dice loss between the prediction samples p and the ground truth and annotation g is defined by the following Equation (1).

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \quad (1)$$

where

- N is the total number of pixels in the image.
- p_i is the predicted probability for the i-th pixel belonging to the bone (foreground).
- g_i is the ground truth label for the i-th pixel (1 if it belongs to the bone, 0 otherwise).
- ϵ is a small constant added to prevent division by zero.

In spite of the fact that Dice loss is a common function, it has the limitation of weighing false positives and false negatives equally. In segmentation tasks, it struggles when dealing with unbalanced data where the number of bone pixels is much lower than the number of non-bone pixels. Abrham et al. [49] and Salehi at al. [50] suggested using the focal Tversky loss function (FTL), which can help mitigate the imbalance problem and effectively train deep neural networks.

The FTL is similar to the Dice loss in that it measures the overlap between the predicted segmentation and the ground truth. It introduces two additional hyperparameters, α and β,

which allow for more fine-grained control over the trade-off between precision and recall. The FTL is defined as

$$\mathrm{TI}(\alpha, \beta) = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{ic} g_{i\hat{c}} + \beta \sum_{i=1}^{N} p_{\hat{i}c} g_{ic} + \epsilon} \tag{2}$$

where

- N is the total number of pixels in the image.
- p_{ic} is the predicted probability of the i-th pixel belonging to class c, in our case the femur.
- g_{ic} is the ground truth label for the i-th pixel belonging to class c (1 if it belongs to the femur, 0 otherwise).
- $p_{i\hat{c}}$ is the predicted probability of the i-th pixel belonging to the complement of class c (not the femur).
- $g_{i\hat{c}}$ is the ground truth label for the i-th pixel belonging to the complement of class c (not the femur).
- α and β are hyperparameters that control the weighting of false positives and false negatives, respectively.
- ϵ is a small constant added to avoid division by zero.

$\sum_{i=1}^{N} p_{ic} g_{ic}$—represents the sum of the product of the predicted values and the ground truth labels for each pixel belonging to the femur. It effectively counts the number of pixels correctly predicted as the femur (true positives).

$\alpha \sum_{i=1}^{N} p_{ic} g_{i\hat{c}}$—represents the sum of the predicted values for pixels that are incorrectly predicted as the femur when they are not (false positives), weighted by α.

$\beta \sum_{i=1}^{N} p_{\hat{i}c} g_{ic}$—represents the sum of the ground truth labels for pixels that are incorrectly predicted as not the femur when they are actually the femur (false negatives), weighted by β.

To define it as a loss function, we use the focal Tversky loss as mentioned by Abraham et al. [49]:

$$\mathrm{FTL}_c(\alpha, \beta) = \sum_{c} (1 - \mathrm{TI}_c)^{1/\gamma} \tag{3}$$

γ is a hyperparameter that increases the emphasis on harder-to-classify pixels. In our case, we set $\gamma = \frac{4}{3}$ as recommended by Abraham et al. [49], allowing the loss function to focus more on less accurate predictions that have been misclassified.

3.7. Complexity

The U-Net architecture contains 70 million parameters. The model's ability to learn intricate features from medical images is encoded by these parameters. It should be noted, however, that they also impose computational demands throughout both the training and inference processes. As part of the training process, we used a GPU configuration of A100 with 40 GB of RAM on Google Collab. It took approximately thirty minutes to complete each epoch. A basic GPU is sufficient to sample from the model and produce a complete 3D segmentation within a minute following training. The model can be further improved by increasing the input image size from 256 × 256 to 512 × 512. To support the increase in parameters, however, it is necessary to have a stronger GPU and more RAM. Similarly, one can train the model with smaller image sizes with a less powerful GPU, at the cost of some accuracy.

4. Results

A common evaluation metric for neural networks involves comparing the predicted output with a ground truth. In segmentation, one way to achieve this is by using the Dice similarity coefficient (DSC), which is a measure of similarity or overlap between two sets. It is measured with four categories: true positive(TP), false positive(FP), false negative(FN), and true negative(TN). In essence, it gives a quantifiable measurement ranging between

0 and 1 that summarizes the quality of segmentation, with higher values being more accurate. In our case, the ground truth is a segmentation that a doctor has performed and the predicted output is our model's segmentation. In the previous section, we discussed the Dice coefficient loss function described in Equation (1), the Dice coefficient D_{SC} can be described as follows:

$$\mathrm{D_{sc}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

Or by subtracting one from our Dice loss function, as they are a compliment to each other.

$$\mathrm{D_{sc}} = 1 - \mathrm{L_{dice}} \tag{5}$$

The Hausdorff distance is a metric used to assess the similarity or dissimilarity between two sets of points or shapes within a given metric space. It gauges the extent to which one set can approximate the other by considering the maximum distance between points in one set and their closest points in the other set. It is defined as follows:

$$\mathrm{H(A,B)} = \max\left(\sup_{a\in A}\inf_{b\in B} \mathrm{d(a,b)}, \sup_{b\in B}\inf_{a\in A} \mathrm{d(a,b)}\right) \tag{6}$$

where A and B are two surfaces, sup and inf denote the least upper bound and greatest lower bound, respectively.

A 3D slicer [51], known for its ability to visualize and analyze segments, was used to assess the accuracy of a segmentation. For calculating the model's final accuracy, the test stage was carried out on 10 patients. The knee was segmented using the entire MRI scan volumetric image, resulting in a complete 3D model. In order to determine the accuracy of the segmentation, we overlapped the ground truth segmentation with our predicted segmentation and used the SlicerRT add-on inside 3D slicer open source. Segment comparison is a module provided by SlicerRT. Using two 3D knee models, this module calculates the Dice score between them. Accordingly, we computed the Dice scores for each of the 10 patients based on the ground truth segmentation of the entire volumetric image.

Attention U-Net yielded accurate results, owing to the effective functioning of attention gates that focused specifically on the knee bones, while diminishing the surrounding background. This precise attention mechanism led to an enhanced segmentation accuracy. The final output demonstrated an elevated level of smoothness, highlighting the advantage of attention U-Net in producing more refined segmentation outcomes.

Figure 10 illustrates the training and validation process throughout the epochs. Through iteration, the graph facilitates learning and performance. The results clearly indicate that no overfitting occurred during the learning processes. In addition, the data were well-adapted and the system was generalized efficiently.

Upon examining Table 1, we can observe the following results: In some instances, such as the first and fifth rows, the automatic segmentation outperformed the ground truth by producing more homogeneous bone contours. In the second row, the algorithm successfully detected a small black point near the lateral epicondyle. A small portion of the bone in the lateral compartment was incorrectly classified by the algorithm in the third row. In the fourth row, the algorithm accurately identified the absence of femoral bone in the selected MR slice. Comparing the proposed approach with the ground truth in the third and fifth rows revealed that the proposed method achieved a segmentation more consistent with the original image.

The addition of the classifier to the U-net made it more robust to false positives. As a result, the CNL output was prioritized for early classification of segmentation necessity.

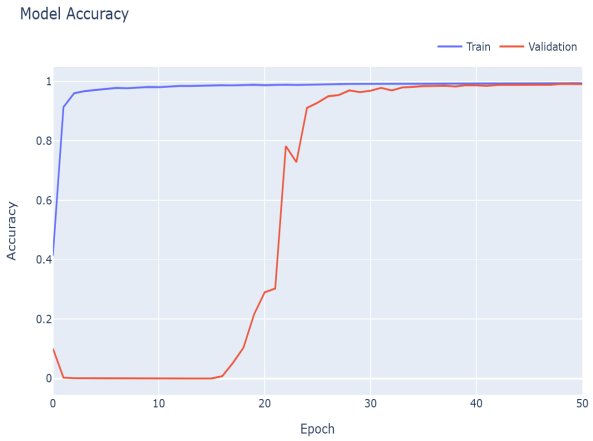
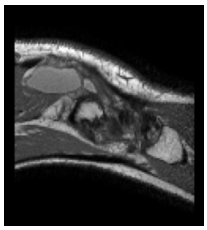
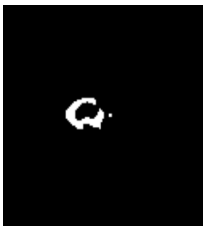
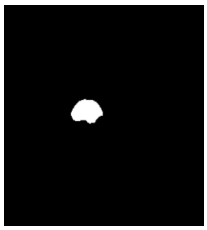
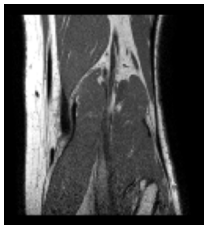
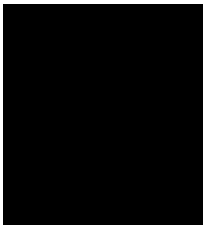
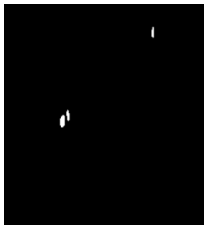
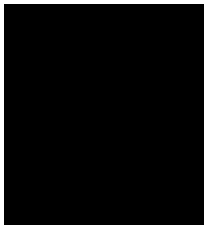
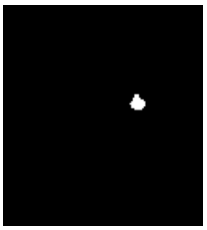
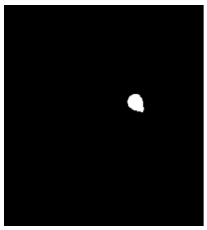
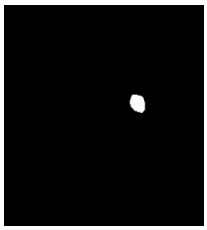


Figure 10. Model accuracy graph.

Table 1. Comparison between tested models.

Original Image	Ground Truth	U-Net	Attention U-Net	Our Approach
				
				
				
				
				

It is also clear that the CNL effectively increased the models accuracy and Hausdorff distance, as seen in Tables 2 and 3. The increase in both metrics confirmed the reduction in false positives, since the CNL unit had no bearing on the quality of the segmentation and only on reducing errors caused by segmentation in the wrong areas. The Dice score results are averages of the Dice scores of ten different patients and their corresponding ground truths. As can be seen in Table 2, some additions contributed to the Dice score, while others, such as focal Tversky (FT) [49], actually decreased it. However, when we combined all the models together, we found that FT added to the overall accuracy of the model. The Hausdorff distance (HD) was calculated by first calculating the average Hausdorff distance for each patient, followed by averaging these distances across all patients. A smaller HD indicates that two sets are similar, while a larger HD indicates that they are dissimilar. Therefore, we aimed to achieve the smallest possible HD.

Table 2. Performance comparison using Dice coefficient score between our proposed enhancements.

Model	Dice Coefficient Scores
AU	96.19%
AU + TL	97.59%
AU + TL + LAYER	95.95%
AU + CLS	97.42%
AU + Focal TV	94.43%
Proposed approach with all modules	**98.05%**

Table 3. Performance comparison using Hausdorff distance between our proposed enhancements.

Model	Hausdorff Distance
AU	0.59 mm
AU + TL	0.47 mm
AU + TL + LAYER	0.6 mm
AU + CLS	0.5 mm
AU + Focal TV	1.3 mm
Proposed approach with all modules	**0.38 mm**

In Table 3, our proposed approach managed to yield an impressive 0.38 mm distance between the ground truth and our prediction.

Three-Dimensional Reconstruction

Figure 11 illustrates how the distance between the ground truth and predicted labels was calculated using a 3D slicer extension. As can be seen from this visualization, the different colors indicate the distance in millimeters between each pixel in the two models. In the case of a small overlap between the models, the distance was closer to zero, and when the absolute value of the distance exceeds 0, this indicates a larger deviation from the ground truth. With the use of this technique, we were able to identify areas of superior and inferior performance within our predictions. Whenever certain regions of a 3D reconstruction lack precision, the overall experience of the patient may be adversely affected. Despite the fact that all other regions overlap perfectly with the ground truth, a deviation can indicate a false result in the extreme case.

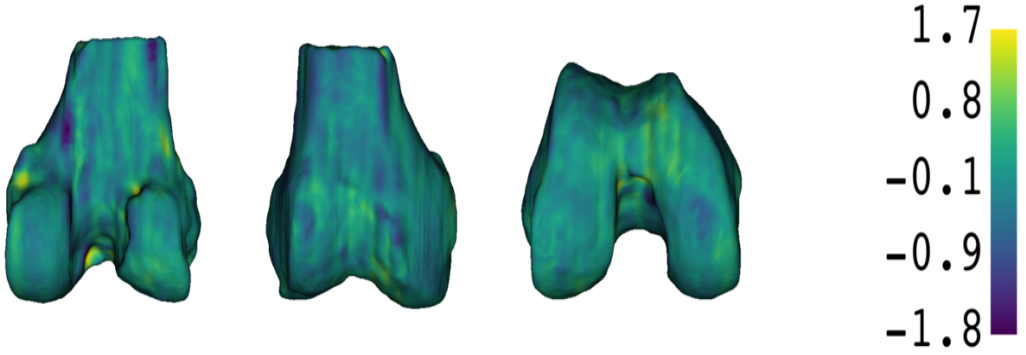


Figure 11. Three-dimensional reconstruction of the segmentation results of the proposed approach in mm.

5. Discussion

The accuracy of automatic segmentation models depends on the implemented algorithm and the characteristics of the input data, in terms of quantity and quality [2,52]. In this study, attention U-net was implemented adding different modules to obtain a novel configuration, including ResNet50 transfer learning , a classification module and one more layer. The input data were a total of 16,230 images, providing a much more detailed anatomical representation than is typically obtained in standard clinical protocols.

The proposed model showed a high performance, achieving a mean Dice score of 98.05% and HD of 0.38 mm.

Although there have been several research studies focusing on the application of DL to MRI images of the knee joint, our quantitative evaluation metrics cannot be directly compared to those from the other studies listed in Table 4. This is primarily due to the variations in the datasets used. Different studies employed different MRI sequences, which affects the image quality, resolution, and the specific anatomical details captured. These differences in the datasets resulted in variability in the performance metrics, making direct comparisons challenging. We have shown the progression of accuracy on the same dataset in Tables 2 and 3, respectively.

Table 4. Results of comparative accuracy in recent literature works.

Model	Neural Networks Architecture	Dice Score
Chen et al., 2022 [1]	3D U-Net	98.0%
Robert et al., 2022 [53]	2.5D U-Net	98.0%
Tang et al., 2022 [54]	3D V-Net	96.4%
Liu et al., 2022 [55]	2D U-Net VGG16	96.0%
Almajalid et al., 2022 [40]	Modified U-Net	96.0%
Kulseng et al., 2023 [52]	Modified U-Net	98.0%
Proposed approach	Extended Atte.Unet	**98.05%**

Improvement in Dice accuracy and Hausdorff distance could enhance clinical applications immensely by creating more precise 3D models for surgical planning and custom implants that are suited to the patient's anatomy.

This study, while achieving an improvement in accuracy, has several limitations. The dataset used is still relatively small and sourced from a single institution. This limits the generalization of the proposed approach's findings to other populations and imaging conditions. Additionally, the exclusion of certain image axes during training might have introduced a bias in the segmentation model, potentially affecting its performance in real-world scenarios.

Future work should test the robustness of the model on different datasets, potentially focusing on specific groups, such as males over the age of 60. Since OA affects older individuals more frequently, fine-tuning the model for these specific scenarios could be highly beneficial. Moreover, the focal Tversky loss function led to a decrease in Dice score.

The focal Tversky loss was intended to address class imbalance problems by focusing on hard-to-segment areas. However, its application might have overly penalized certain predictions, leading to a lower overall Dice score. Future work should explore fine tuning the hyper-parameters of focal Tversky, possibly through a linear interpolation method, to achieve better results.

6. Conclusions

This study introduced several novel modules and integrated them into current segmentation methodologies, resulting in an enhanced attention U-Net. The attention U-Net incorporates innovative components such as transfer learning, increased depth, a focal Tversky technique, and an additional classifier, to overcome the challenges and constraints described in the article. By combining these techniques, commendable segmentation results were obtained, particularly within the relatively unexplored field of femur segmentation. The purpose of this research was to leverage automated 3D modeling to reduce manual intervention by radiologists and provide patients with more precise 3D implants by leveraging femur segmentation. By developing a cutting-edge technique, we aim to develop patient-specific implants that are tailored to the anatomical characteristics of each patient.

Author Contributions: Conceptualization: D.A., M.W. and M.W.C.; methodology: D.A., M.W. and M.W.C.; software: D.A. and M.W.; validation: D.A., M.W. and A.G.; formal analysis: D.A. and M.W.; data curation: A.G.; writing original draft: D.A. and M.W.; writing—review and editing: D.A., M.W., A.G. and M.W.C.; visualization: D.A., M.W. and A.G.; supervision: M.W.C.; project administration: M.W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Braude Research Grant 5210-95624-22-23.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the corresponding author on request

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, X.; Wang, X.; Zhang, K.; Fung, K.M.; Thai, T.C.; Moore, K.; Mannel, R.S.; Liu, H.; Zheng, B.; Qiu, Y. Recent advances and clinical applications of deep learning in medical image analysis. *Med. Image Anal.* **2022**, *79*, 102444. [CrossRef]
2. Mahendrakar, P.; Kumar, D.; Patil, U. A Comprehensive Review on MRI-based knee joint segmentation and analysis techniques. *Curr. Med. Imaging* **2023**, *20*, e150523216894. [CrossRef]
3. Chen, H.; Zhao, N.; Tan, T.; Kang, Y.; Sun, C.; Xie, G.; Verdonschot, N.; Sprengers, A. Knee bone and cartilage segmentation based on a 3D deep neural network using adversarial loss for prior shape constraint. *Front. Med.* **2022**, *9*, 792900. [CrossRef] [PubMed]
4. Zhang, K.; Lu, W.; Marziliano, P. Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies. *Magn. Reson. Imaging* **2013**, *31*, 1731–1743. [CrossRef]
5. Ghidotti, A.; Vitali, A.; Regazzoni, D.; Cohen, M.W.; Rizzi, C. Comparative Analysis of Convolutional Neural Network Architectures for Automated Knee Segmentation in Medical Imaging: A Performance Evaluation. *J. Comput. Inf. Sci. Eng.* **2024**, *24*, 051005. [CrossRef]
6. Marsilio, L.; Faglia, A.; Rossi, M.; Mainardi, L.; Manzotti, A.; Cerveri, P. CEL-Unet: A novel CNN architecture for 3D Segmentation of Knee Bones affected by Severe Osteoarthritis for PSI-Based Surgical Planning. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, 11–15 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 5039–5042.
7. Hawker, G.A. Osteoarthritis is a serious disease. *Clin. Exp. Rheumatol.* **2019**, *37*, 3–6.
8. Conaghan, P.; Abraham, L.; Viktrup, L.; Cappelleri, J.; Beck, C.; Bushmakin, A.; Berry, M.; Jackson, J. Impact of osteoarthritis disease severity on treatment patterns and healthcare resource use: Analysis of real-world data. *Scand. J. Rheumatol.* **2022**, *52*, 353–363. [CrossRef]
9. NIH. NIH, 2023. Available online: https://www.nih.gov (accessed on 3 December 2023).
10. Yao, Q.; Wu, X.; Tao, C.; Gong, W.; Chen, M.; Qu, M.; Zhong, Y.; He, T.; Chen, S.; Xiao, G. Osteoarthritis: Pathogenic signaling pathways and therapeutic targets. *Signal Transduct. Target. Ther.* **2023**, *8*, 56. [CrossRef]
11. Heimann, T.; Morrison, B.J.; Styner, M.A.; Niethammer, M.; Warfield, S.K. Segmentation of knee images: A grand challenge. In Proceedings of the MICCAI Workshop on Medical Image Analysis for the Clinic, Beijing, China, 20–24 September 2010; pp. 207–214.
12. Thirukumaran, C.P.; Cai, X.; Glance, L.G.; Kim, Y.; Ricciardi, B.F.; Fiscella, K.A.; Li, Y. Geographic variation and disparities in total joint replacement use for Medicare beneficiaries: 2009 to 2017. *J. Bone Jt. Surg.* **2020**, *102*, 2120. [CrossRef] [PubMed]

13. Nagano, H. Special Issue on Biomechanical and Biomedical Factors of Knee Osteoarthritis. *Appl. Sci.* **2022**, *12*, 11807. [CrossRef]
14. Liu, F.; Zhou, Z.; Jang, H.; Samsonov, A.; Zhao, G.; Kijowski, R. Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging. *Magn. Reson. Med.* **2018**, *79*, 2379–2391. [CrossRef]
15. Concepts, C. Knee Osteoarthritis Condition. Available online: https://www.coreconcepts.com.sg/knee-osteoarthritis/ (accessed on 10 June 2024).
16. Auricchio, F.; Marconi, S. 3D printing: Clinical applications in orthopaedics and traumatology. *EFORT Open Rev.* **2016**, *1*, 121–127. [CrossRef] [PubMed]
17. Leppänen, S.; Niemeläinen, M.; Huhtala, H.; Eskelinen, A. Mild knee osteoarthritis predicts dissatisfaction after total knee arthroplasty: A prospective study of 186 patients aged 65 years or less with 2-year follow-up. *BMC Musculoskelet. Disord.* **2021**, *22*, 1–8. [CrossRef] [PubMed]
18. Ghidotti, A.; Landi, D.; Regazzoni, D.; Rizzi, C. Evaluation of clinical and technical parameters to customize Total Knee Arthroplasty implants. *J. Comput. Inf. Sci. Eng.* **2022**, *22*, 061006. [CrossRef]
19. Bahadır, B.; Atik, O.Ş.; Kanatlı, U.; Sarıkaya, B. A brief introduction to medical image processing, designing and 3D printing for orthopedic surgeons. *Jt. Dis. Relat. Surg.* **2023**, *34*, 451. [CrossRef] [PubMed]
20. Touri, M.; Kabirian, F.; Saadati, M.; Ramakrishna, S.; Mozafari, M. Additive manufacturing of biomaterials- the evolution of rapid prototyping. *Adv. Eng. Mater.* **2019**, *21*, 1800511. [CrossRef]
21. Ahmed, S.M.; Mstafa, R.J. A Comprehensive Survey on Bone Segmentation Techniques in Knee Osteoarthritis Research: From Conventional Methods to Deep Learning. *Diagnostics* **2022**, *12*, 611. [CrossRef] [PubMed]
22. Hwang, D.; Ahn, S.; Park, Y.B.; Kim, S.H.; Han, H.S.; Lee, M.C.; Ro, D.H. Deep learning-based muscle segmentation and quantification of full-leg plain radiograph for sarcopenia screening in patients undergoing total knee arthroplasty. *J. Clin. Med.* **2022**, *11*, 3612. [CrossRef] [PubMed]
23. Mercader, A.; Röttinger, T.; Bigdeli, A.; Lüth, T.C.; Röttinger, H. A patient-specific 3D model of the knee to compare the femoral rollback before and after total knee arthroplasty (TKA). *J. Exp. Orthop.* **2021**, *8*, 1–10.
24. Kerdvibulvech, C.; Yamauchi, K. Structural human shape analysis for modeling and recognition. In Proceedings of the Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, S+ SSPR 2014, Joensuu, Finland, 20–22 August 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 282–290.
25. Gan, H.S.; Ramlee, M.H.; Wahab, A.A.; Lee, Y.S.; Shimizu, A. From classical to deep learning: Review on cartilage and bone segmentation techniques in knee osteoarthritis research. *Artif. Intell. Rev.* **2021**, *54*, 2445–2494. [CrossRef]
26. Pandey, P.U.; Quader, N.; Guy, P.; Garbi, R.; Hodgson, A.J. Ultrasound bone segmentation: A scoping review of techniques and validation practices. *Ultrasound Med. Biol.* **2020**, *46*, 921–935. [CrossRef]
27. Anshad, P.Y.M.; Kumar, S.S.S. Segmentation of Chondroblastoma from Medical Images Using Modified Region Growing Algorithm. *Clust. Comput.* **2018**, *22*, 13437–13444. [CrossRef]
28. Becker, M.; Magnenat-Thalmann, N. Deformable Models in Medical Image Segmentation. In *3D Multiscale Physiological Human*; Springer: London, UK, 2014; pp. 81–106.
29. Cootes, T.F.; Taylor, C. *Active Shape Models—'Smart Snakes' BT—BMVC92*; Springer: London, UK, 1992; pp. 266–275.
30. Heimann, T.; Meinzer, H. Statistical Shape Models for 3D Medical Image Segmentation: A Review. *Med. Image Anal.* **2009**, *13*, 543–563. [CrossRef]
31. Sarkalkan, N.; Weinans, H.Z.A. Statistical Shape and Appearance Models of Bones. *Bone* **2014**, *2014*, 129–140. [CrossRef]
32. Boykov, Y.Y.; Jolly, M. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 1, pp. 105–112.
33. Dam, E.B.; Lillholm, M.; Marques, J.; Nielsen, M. Automatic segmentation of high-and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. *J. Med. Imaging* **2015**, *2*, 024001. [CrossRef]
34. Kessler, D.A.; MacKay, J.W.; Crowe, V.A.; Henson, F.M.; Graves, M.J.; Gilbert, F.J.; Kaggie, J.D. The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs. *Comput. Med. Imaging Graph.* **2020**, *86*, 101793. [CrossRef]
35. Jamshidi, A.; Pelletier, J.P.; Martel-Pelletier, J. Machine-Learning-Based Patient-Specific Prediction Models for Knee Osteoarthritis. *Nat. Rev. Rheumatol.* **2019**, *15*, 49–60. [CrossRef]
36. Kluzek, S.; Mattei, T. Machine-Learning for Osteoarthritis Research. *Osteoarthr. Cartil.* **2019**, *27*, 977–978. [CrossRef]
37. Ambellan, F.; Tack, A.; Ehlke, M.; Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Med. Image Anal.* **2019**, *52*, 109–118. [CrossRef]
38. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
39. Norman, B.; Pedoia, V.; Majumdar, S. Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology* **2018**, *288*, 177–185. [CrossRef]
40. Almajalid, R.; Zhang, M.; Shan, J. Fully automatic knee bone detection and segmentation on three-dimensional MRI. *Diagnostics* **2022**, *12*, 123. [CrossRef]

41. Deniz, C.M.; Xiang, S.; Hallyburton, R.S.; Welbeck, A.; Babb, J.S.; Honig, S.; Cho, K.; Chang, G. Segmentation of the proximal femur from MR images using deep convolutional neural networks. *Sci. Rep.* **2018**, *8*, 16485. [CrossRef]
42. Chen, F.; Liu, J.; Zhao, Z.; Zhu, M.; Liao, H. Three-Dimensional Feature-Enhanced Network for Automatic Femur Segmentation. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 243–252. [CrossRef]
43. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]
44. Martel-Pelletier, J.; Paiement, P.; Pelletier, J.P. Magnetic resonance imaging assessments for knee segmentation and their use in combination with machine/deep learning as predictors of early osteoarthritis diagnosis and prognosis. *Ther. Adv. Musculoskelet. Dis.* **2023**, *15*, 1759720X231165560. [CrossRef]
45. Maji, D.; Sigedar, P.; Singh, M. Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors. *Biomed. Signal Process. Control* **2022**, *71*, 103077. [CrossRef]
46. DenOtter, T.D.; Schubert, J. Hounsfield Unit 2019. Available online: http://europepmc.org/books/NBK547721 (accessed on 3 December 2023).
47. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net:Learning where to Look for the Pancreas. In Proceedings of the 1st Conference on Medical Imaging with Deep Learning, Amsterdam, The Netherlands, 4–6 July 2018.
48. Bilal, A.; Sun, G.; Mazhar, S.; Imran, A.; Latif, J. A Transfer Learning and U-Net-based automatic detection of diabetic retinopathy from fundus images. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2022**, *10*, 663–674. [CrossRef]
49. Abraham, N.; Khan, N.M. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 683–687.
50. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In Proceedings of the Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, 10 September 2017; Proceedings 8; Springer: Berlin/Heidelberg, Germany, 2017; pp. 379–387.
51. SlicerRt. SlicerRT. 2023. Available online: https://github.com/SlicerRt/ (accessed on 29 November 2023).
52. Kulseng, C.P.S.; Nainamalai, V.; Grøvik, E.; Geitung, J.T.; Årøen, A.; Gjesdal, K.I. Automatic segmentation of human knee anatomy by a convolutional neural network applying a 3D MRI protocol. *BMC Musculoskelet. Disord.* **2023**, *24*, 41. [CrossRef]
53. Robert, B.; Boulanger, P. Automatic Bone Segmentation from MRI for Real-Time Knee Tracking in Fluoroscopic Imaging. *Diagnostics* **2022**, *12*, 2228. [CrossRef]
54. Tang, X.; Guo, D.; Liu, A.; Wu, D.; Liu, J.; Xu, N.; Qin, Y. Fully automatic knee joint segmentation and quantitative analysis for osteoarthritis from magnetic resonance (MR) images using a deep learning model. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2022**, *28*, e936733. [CrossRef]
55. Liu, W.; Ge, T.; Luo, L.; Peng, H.; Xu, X.; Chen, Y.; Zhuang, Z. A novel focal ordinal loss for assessment of knee osteoarthritis severity. *Neural Process. Lett.* **2022**, *54*, 5199–5224. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.