Hip Segmentation for children with cerebral palsy using deep learning

Laya Karimi Chahrogh

November 26, 2022

Abstract

Hip disorders such as hip subluxation and dislocations are common between children with cerebral palsy(CP) because of muscles contracting around the hip. Tight muscles can force the hip to get pulled away from its proper placement in the socket. Pain and a limitation in walking ability are two clinical impacts of hip problems in ambulatory children with cerebral palsy. In non-ambulatory children, the consequences include hindering in sitting balance, skin breakdown, and diapering. In order to shift the painful, arthritic ball away from the socket in older children with painful dislocated hips, the hip joint may need to be removed or a corrective osteotomy performed. Total hip arthroplasty may be necessary for ambulatory patients having painful arthritic hips. [1].

Frequently monitoring and performing physical exams is very important in children with CP, which is called surveillance. Hip surveillance cannot prevent a hip from becoming subluxated or the need for hip surgery. However, it can reduce the probability of hip dislocation occurrence by helping doctors make a proper plan to help children. [2–4]. A patient-specific analysis is crucial in total hip arthroplasty (THA) to assess the surgical result and develop suitable rehabilitation plans. A quantitative study of patient-specific biomechanical simulation is possible thanks to hip segmentation in patients' CT scans. [5]. The placement of joint implants can be simulated using the precise localization and segmentation of the hip joint using CT scans. [6–8]. However, CT scans shouldn't be used frequently on young patients due to the radiation exposure, especially in those who are of childbearing age [9]. Additionally, image degradation, noise, non-homogeneous intensities, and hazy femoral head-acetabulum borders can be seen in CT scans of diseased hips [8, 10, 11]; Therefore, a suitable segmentation scheme is crucial to fully automated segmentation of hip joints in children with CP requires [8].

This work aims to design and evaluate an algorithm capable of segmenting hip structures in CT images of children with CP. We have implemented a modified U-Net architecture along with a transfer learning method to segment the hip structure in the CT data. For segmentation, we have taken into account two separate datasets. The first dataset includes 21 CT scans of hips in children. The second dataset is 140 CT scans of the lung, bones, liver, kidneys, and bladder, which have been used for transfer learning. The desired segmentation was achieved after 100 iterations. An accuracy=0.98, loss=0.01, and IoU of 0.99 were obtained, which are acceptable and illustrate the efficiency of our method.

Contents

1 Introduction

1.1 Motivation

The hip structure (shown in Figure (1)) is among the most important parts of the human body, facilitating movement and stability. The hip is vulnerable and can be easily injured due to its biological function. Pain or injury in the hip can be debilitating and limit motion and the ability to bear weight. Hip deformities are painful disorders that occur in 1 in 500 children [12], and children with hip deformities are prone to develop osteoarthritis and often require hip replacement surgeries as young adults [12].

A group of illnesses known as cerebral palsy (CP) are brought on by harm to the developing brain [13]. Hip disorders such as hip subluxation and dislocations are common between children with CP because the muscles contract around the hip. Tight muscles can pull the hip away from its proper placement in the socket. Pain and a limitation in walking ability are two clinical impacts of hip problems in ambulatory children with cerebral palsy. In non-ambulatory children, the consequences include hindering sitting balance, skin breakdown, and diapering. In order to transfer the painful, arthritic ball away from the socket, older children with painful dislocated hips may require hip removal or a corrective osteotomy. Total hip replacement surgery may be necessary for ambulatory patients with painful arthritic hips. [1].

Monitoring and performing physical exams are very important in children with CP, called "surveillance". Hip surveillance cannot prevent a hip from becoming subluxated or the need for hip surgery. However, it can reduce the probability of hip dislocation by helping doctors make a proper plan to help children [2–4].In addition, patient-specific analysis is crucial in total hip arthroplasty (THA) to assess the success of the treatment and develop suitable rehabilitation plans.

Medical imaging is a technique and a process for visually representing the interior of a patient's body for clinical evaluation and medical intervention. [14]. Healthcare workers can use these images to treat and correctly diagnose patients since they contain crucial information. Clinical diagnosis now depends on medical images such as X-rays, computed tomography (CT), and magnetic resonance imaging (MRI). Medical image segmentation is an important process in the study of medical images. Based on the properties of the image's pixels, it divides an image into numerous meaningfully descriptive pieces or areas. The outcome of the image division can assist in picture analysis and object classification. So it is crucial to produce a properly segmented image, especially for medical photos.

Manual, semi-automatic, and automatic segmentation algorithms are the three categories used to classify traditional medical image segmentation techniques [15].However, these algorithms are based on prior human knowledge. Furthermore, they lack the ability to generalize, which makes it challenging to get the outcomes that are expected. On the other hand, deep learning techniques have advanced in the segmentation of biological images, and these new techniques are assisting us in automating the segmentation problem with high accuracy and quicker processing [15, 16].

Machine learning is a subfield of artificial intelligence (AI) that enables systems to automatically learn from data with minimum human intervention, high accuracy, and faster processing [15, 16]. There are different machine learning methods [17], and deep learning is one of the important ones [18]. The most well-known deep learning algorithm for interpreting medical images is convolutional neural networks (CNN). CNN comprises several layers in which input information, such as an image, produces output, such as the presence or absence of disease [17]. CNN comprises various layers that process input data, like an image, to produce outputs, such as the presence or absence of a disease [17]. Convolutional neural networks (CNN), including fully convolutional networks (FCN) [19], U-Net [18], SegNet [20], PSPNet [21], and deep lab versions [15, 22] were developed to address image recognition problems.

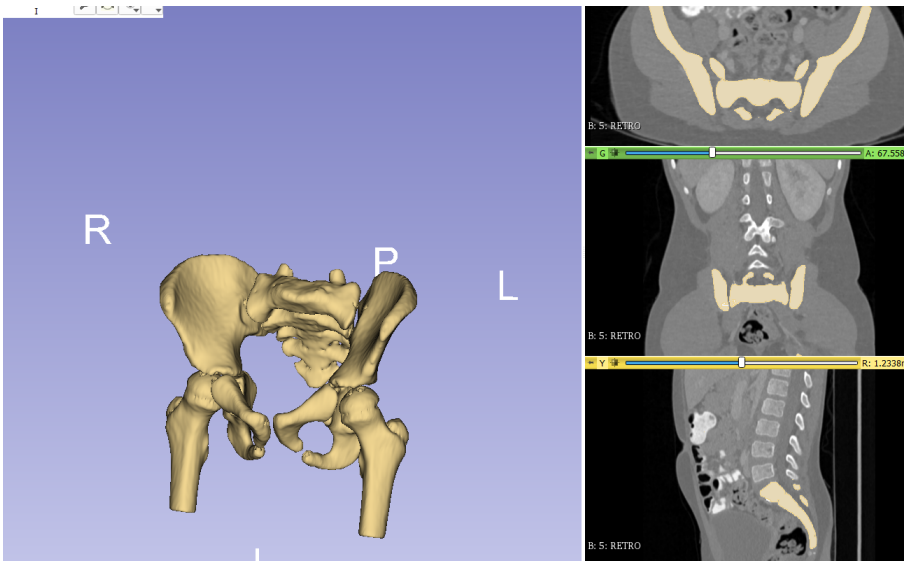


Figure 1: Hip Structure

1.2 Problem Statement

As mentioned in the previous section, in order to monitor and treat hip pain and injuries in children with CP, the segmentation of the hip structure is crucial. However, due to bone deformation and joint space narrowing in diseased hips, particularly in youngsters, precise segmentation of the hip joint is difficult. Hence, the goal of this study is to understand the concept

of deep learning to segment the hip structure in patients with CP with acceptable accuracy, precision, and speed. For a patient-specific and radiation-free strategy for preoperative or treatment planning in CP, we create a modified U-Net network for automated hip segmentation. For segmentation, we have taken into account two distinct datasets. The first dataset consists of 21 hip CT images of children with cerebral palsy. 140 CT scans of the lung, bones, liver, kidneys, and bladder are included in the second dataset, which was utilized for transfer learning.

1.3 Related Works

Deep learning algorithms, in particular CNNs, have quickly taken over as the go-to approach for medical image interpretation in recent years.In earlier investigations [2, 23–37], CNNs were used to automatically segment the hip in medical imaging. In ultrasound pictures, the ilium and acetabulum roof were segmented using CNNs [23–25]. In MRI images, CNNs were also applied to segment the trabecular bone in the proximal femur and detect osteonecrosis of the femoral head [27–31]. The trained CNNs were also used to identify osteoarthritis on hip radiographs [32]. For the purpose of segmenting the proximal femur from 3D MR images, Zeng et al. have suggested a deeply supervised 3D U-net-like fully convolutional network [33]. The usage of very ill hips (i.e., some challenging cases) is not documented, despite the technology being successfully shown on 20 3D MR images of patients with femoroacetabular impingement. A 3D feature-enhanced network for automated femur segmentation from CT images was presented by Chen et al [34]. Their research achieved quick and precise segmentation for a range of therapeutic applications. The performance, however, is strongly dependent on a substantial amount of training data, which is occasionally accessible. The technique may only produce a partial segmentation of the femur in the problematic regions when there is a lack of training data [6]. Wang et al. used 2D and 3D U-Net to segment pelvises [38]. The dice overlap coefficient (DOC) was $> 94\%$. In [36], 31 patients' 3D models based on MRIs were automatically segmented (26 symptomatic patients with hip dysplasia or FAI). They employed the LP-U-net technique, and the results showed that the dice values for the proximal femur and the acetabulum were 0.98 and 0.97, respectively. A Latent 3D U-net technique was used in citezeng2018latent3du [35] to perform an automated segmentation of the proximal femur from a radial MRI of the hip. The obtained dice and IoU were 0.954 and 0.912, respectively. A fully convolutional network that resembles a 3D U-net and has multi-level deep supervision has been used on twenty 3D MR images of patients with femoroacetabular impingement in the study cited in [33]. This experiment's dice and IoU are 0.987 and 0.974, respectively. A CNN architecture using the U-net and its 3D extension has been suggested in [39] for Proximal Femur segmentation from MR Images. Dice and IoU reported are 0.95 0.02 and 0.941.

1.4 Thesis outline

There are six chapters in the remaining section of this thesis. The background of medical image analysis segmentation is covered in Chapter 2, along with the generation of features for the segmentation of CT scans using deep learning. We describe how to segment hip components and evaluate CT images in children with cerebral palsy in the third chapter using a deep learning model. The outcomes are shown in the fourth chapter. In the fifth chapter, a discussion is presented. The conclusion of the thesis is the final section.

2 Literature Review

2.1 Medical Image Diagnosis

Diagnostic imaging, also called medical imaging, refers to several non-invasive methods used to view a patient's body to help find out the causes of an injury or an illness and confirm a diagnosis. It also determines how well the body responds to treating an illness or a fracture. Images of the internal organs and processes inside the body can be produced using a variety of devices and methods. The type of imaging a doctor performs is determined by the body part being checked and your symptoms. They consist of ultrasounds, X-rays, CT scans, nuclear medicine scans, and MRI scans.

A CT scan is ideal for examining bone injuries, identifying lung and chest conditions, and diagnosing cancer. [40]. An MRI is useful for assessing soft tissue in conditions such as brain tumors, spinal cord injuries, and injuries to ligaments and tendons. Because a CT scan may be completed in under 5 minutes, it is frequently used in emergency rooms. On the other hand, an MRI can take up to 30 minutes. Typically, an MRI is more expensive than a CT scan. An MRI has the benefit of not using radiation, whereas CT scans do. Repeated exposure to te radiation is dangerous.

Medical image segmentation is a crucial stage in the study of medical images. Based on the properties of the image's pixels, it divides an image into numerous meaningfully descriptive pieces or areas. Medical image segmentation aims to represent an input image in a meaningful way to study anatomy, find the region of interest (RoI), calculate the volume of tissue, and help determine the dose of medication, plan treatment before applying radiation therapy, or calculate the radiation dose. By highlighting the area of interest, image segmentation facilitates the analysis of medical images. Image segmentation is therefore essential for properly comprehending the content of medical images and making diagnoses since it gives the data a more meaningful representation.

The three types of traditional medical picture segmentation algorithms—manual, semi-automatic, and automatic [15]—include region-based, edge-based, and thresholding procedures. In the thresholding technique, pixels are assigned to various groups based on the values' range within which they fall. In the edge-based method, a filter is used for a picture, classifying the pixels as edged or nonedged based on the filter output. In region-based segmentation techniques, groups of pixels with differing values and adjacent pixels with similar values are separated.

The majority of medical images, however, exhibit noise, severe inhomogeneity, and weak borders necessitating complicated segmentation. As a result, automatic image segmentation methods like thresholding and region growing usually provide segmented images that are less accurate. These algorithms also depend on previously acquired human expertise. Additionally, they have poor generalization skills, which makes it challenging to get the desired outcomes. On the other hand, biomedical image segmentation has advanced thanks to deep learning approaches. They are assisting us in automating the segmentation problems with greater precision and speed [15, 16].

2.2 Deep Learning for Computer Vision

2.2.1 Neural Networks

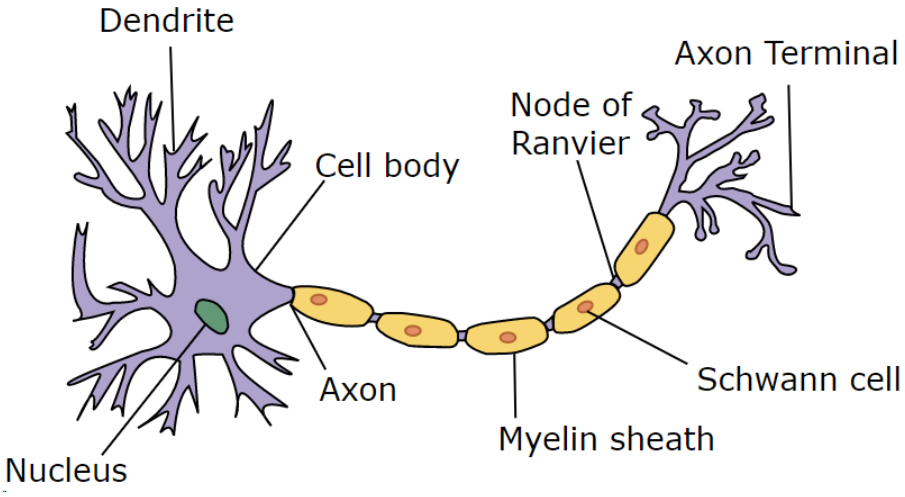


Figure 2: Neuronal Cell Structure [41]

Neural networks (NNs), also known as artificial neural networks (ANNs), are a sort of mathematical model that enables computer systems to detect patterns and relationships in data sets. [42]. They mimic the way biological neurons communicate and take their names and structures from neural networks in living things, illustrated in Figure (2); When a dendrite gets input, it sends it to the nucleus. The axon terminals receive the input from the nucleus. Axon terminals are joined to the dendrite of the following neuron. ANNs repeatedly loop from the output to the input to reinforce the connections, unlike the human neuron structure.

A straightforward structure of NN models is shown in Figure (3). There are three layers in a NN:

- **Input Layer:** This layer takes features as input and feeds the network data from the outside world. Additionally, no calculation is performed at this layer, and nodes pass the data (or features) to the hidden layer.
- **Hidden Layer:** Hidden layers are a sequence of mathematical operations, each of which is designed to yield an output particular to the desired outcome. This layer's nodes are private to the neural network and are not accessible to the outside world. The features entered through the input layer are processed by the hidden layer in a variety of ways, and the output layer receives the results.
- **Output Layer:** The output layer is the neural network's last layer, which produces the program's specified outputs.

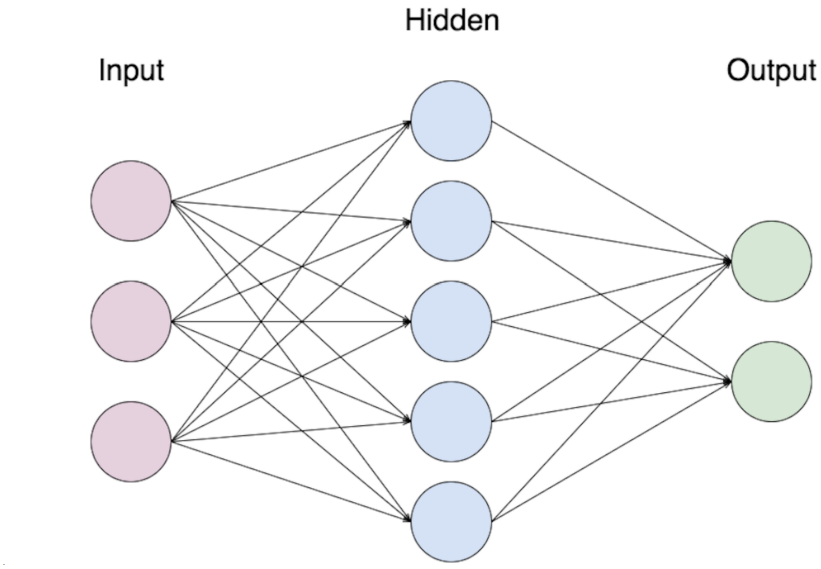


Figure 3: Neuronal Network Scheme [43]

2.2.2 Activation function

The activation function calculates each neuron's output in the network. This function determines whether a neuron should be activated or not. The neuron is activated if the function's output is greater than a certain threshold. Otherwise, the neuron is not activated.

The activation function is crucial to neural networks' precision and efficiency. The type of predictions the model is capable of making is determined by the activation function. The activation function also helps the network understand complicated patterns from the training dataset. A neural network's capacity for learning is limited in the absence of an activation function. The selection of the activation function for the output layers is based on the expected result (the Sigmoid activation function, for instance, is employed for binary classification); see [44]. There are two different categories of activation functions: linear and non-linear.

The linear activation functions are between $(-)$ infinity to $(+)$ infinity. As a result, when given input from multivariable neural networks, these functions perform poorly. These drawbacks make the employment of linear activation functions uncommon. Contrarily, non-linear activation functions are the most widely used in the field of neural networks and perform well in the majority of situations. The most well-known non-linear activation functions include rectified linear unit(ReLU), Sigmoid, and Tanh. Figure (4) illustrates the structure of typical activation functions.

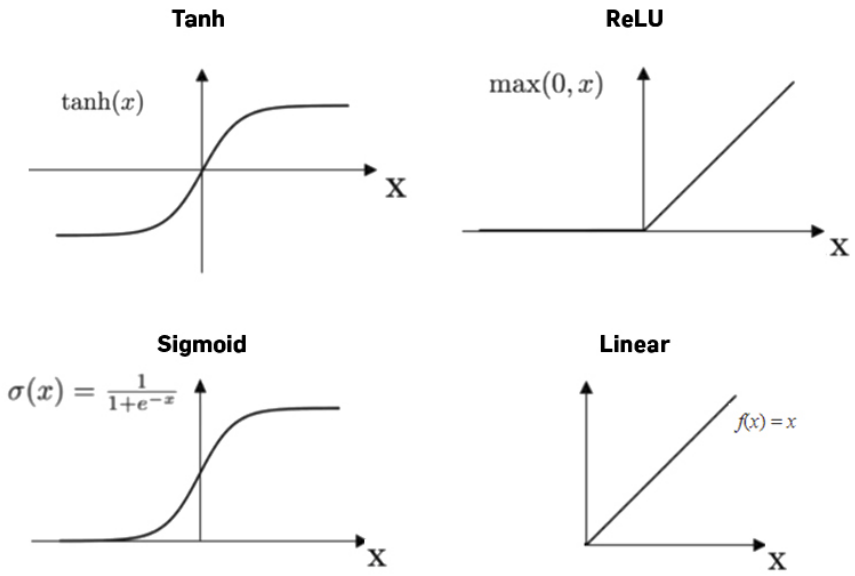


Figure 4: commonly used activation functions [45]

2.2.3 Deep Learning

Deep learning is a branch of machine learning consisting of a neural network with one or more hidden layers [44]. Utilizing strong computational resources and big data, deep learning can be applied to knowledge-based applications such as computer vision, language translation, and voice recognition [44].

Deep learning can learn and identify high-level characteristics from raw data during the learning process, in contrast to standard machine learning techniques like logistic regression and decision trees, which require a collection of features or sample data for training the model. For instance, in a facial recognition system, logistic regression doesn't directly look at a user's face. Instead, it learns how these features correlate by using pre-existing features, such as the edges or shapes of facial parts.

In addition, Traditional machine learning algorithms are unable to affect how features are defined, hence it is dependent on human interaction. Deep learning, in contrast, uses the raw data as input and does not require any pre-feature extraction. A deep-learning example of feature extraction for face recognition in the hidden layer is shown in Figure(5). In this case, the first layer could learn to extract edges, the second could learn to recognize facial components, and the third could learn to connect the facial components to the person.

Deep learning algorithms are divided into two categories supervised learning and unsupervised learning. Supervised learning can learn features from labeled datasets. Unsupervised learning, on the other hand, doesn't need labeled datasets and instead seeks to analyze data patterns. In this project, we will apply a supervised learning method.

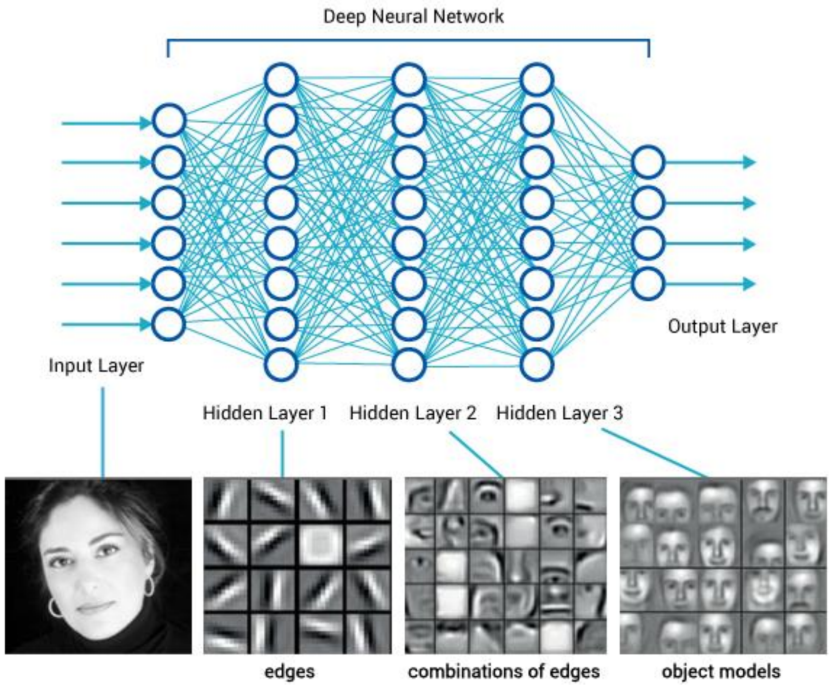


Figure 5: feature extraction from a face image using deep learning [46]

2.2.4 Convolutional Neural Networks (CNNs)

The most prevalent deep learning model used in computer vision applications is CNN. Common neural networks and CNN are quite similar, however, CNN makes the crucial assumption that the input data is organized in a grid-like structure. Images, which have pixels in a 2D grid, are the most simple example of this type of data.

A collection of pixels, commonly in the form of a 3-D pixel array, make up an image. There are various numbers of channels in each pixel. In a grayscale image, there is only one pixel, but in a colored image, there are three channels: red, green, and blue. For instance, the values for red and green would be (255, 0, 0) and (0, 255, 0), respectively. The computer interprets the input (picture) similarly to how people do. In order to learn spatial hierarchies of patterns, it first detects simple edges before recognizing complicated information. Applications for machine learning and computer vision are built on this.

In order to distinguish different elements in a picture, a CNN takes in an input image and assigns priority using weights and biases [47]. CNN has gained popularity due to its shorter processing times and better generalization when there are few training images. According to Chollet's description in his book, CNN is important for two reasons: first, the patterns it learns are "translation invariant," which means that it can detect a pattern anyplace in the image that has been learned from one section of the image [47]. The network's capacity to learn "spatial hierarchies of patterns" is its second attribute. A convolutional layer, a pooling layer, and a fully connected (FC) layer are the three main layers that make up a deep CNN.

The majority of computations take place in the convolutional layer, which is the fundamental component of a CNN. Similar to a standard neural network, the convolution process multiplies an array of input (an image) by a set of weights known as a kernel or a filter. Convolutions work using feature maps, which are produced by repeatedly applying the kernel to the input, creating a map of activations (or feature maps) [47].

The pooling layer is employed to reduce the spatial size of the feature map, just like the convolutional layer. By processing data in fewer dimensions, this aids in reducing computational power. It also aids in the effective training of the model by removing dominating character-

istics that are persistent in their rotation and location. The two different types of pooling strategies are max pooling and average pooling. Average pooling calculates the average value from all the values in the kernel-surrounded region of the image, while max pooling calculates the largest value from that same section of the image. Max pooling helps downsampling by reducing dimensionality, which helps in noise suppression, and functions as a noise suppressant by removing noisy activations. Average pooling, as contrast to max pooling, only aids downsampling as a noise mitigation technique. Therefore, max pooling is a superior pooling method.

The feature map must be flattened after using a pooling layer. The entire pooling matrix is converted into a single feature column by this technique. Fully connected layers are used to produce the final classification. The fully connected layers are connected to every activation in the other layers. Consequently, information gathered from earlier layers is used to construct the final output.

The type of predictions that the model is capable of making is determined by the activation function, which is required to help the network understand complicated patterns from the training dataset. A neural network's capacity for learning is limited in the absence of an activation function. A rectified linear activation function is a typical activation function for hidden layers (ReLU). ReLU solves the vanishing gradient issue, improving model performance. If it receives a negative input, it returns 0, and if it receives a positive input, it returns the same number. Depending on the desired outcome, the activation function for output layers is chosen (for instance, sigmoid is applied for binary classification).

To guarantee that the network trains correctly, it also needs a loss function, an optimizer, and parameters that must be monitored throughout training and testing. On the basis of the training data, a loss function evaluates the network's performance and guides it in the appropriate directions. A network updates itself using an optimizer based on data obtained from the loss function. Additionally, tuning of hyperparameters is recommended for the best outcomes. The network structure and the model's trainability are determined by the hyperparameters.

To sum up, the goal of deep learning or machine learning is generalizing predictions to new data. CNN is one of the best models for computer vision applications since it can learn from little datasets and produce decent results. Figure (6) illustrates the fundamental steps in a CNN. These steps include selecting an input as well as their corresponding labels for supervised learning, operating a convolution operation a long with an activation function, operating a pooling layer, flattening, and using a fully connected layer, compiling CNN utilizing a loss function and optimizer, fitting the CNN, and finally achieving the expected output.

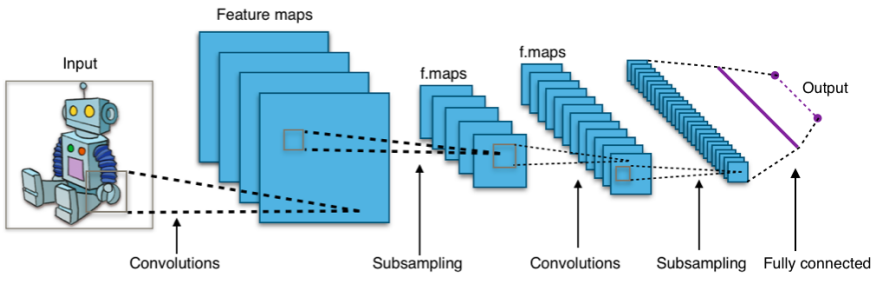


Figure 6: Fully connected convolutional neural network [48]

2.2.5 Transfer Learning for Deep Neural Networks

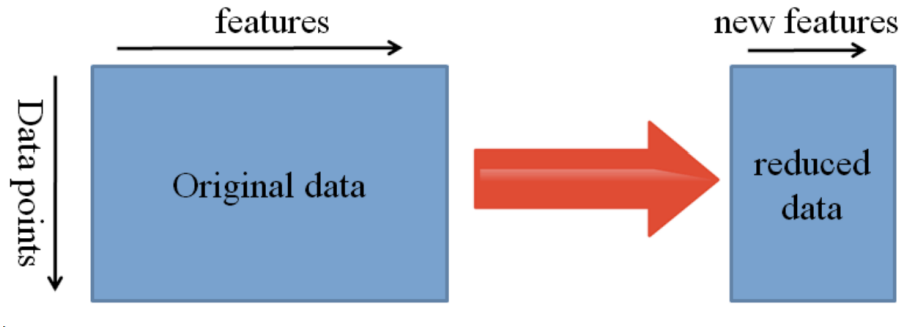
Transfer learning is a technique used in machine learning where the knowledge of a model that has already been trained is applied to a new, related problem. [49]. For instance, the knowledge of a trained classifier for detecting a backpack in an image can be reused to predict other objects

like sunglasses. In other words, we transfer the learned weights of a network at one task to a new task. Therefore, we start learning process with patterns learned from solving a related task instead of starting from scratch. Transfer learning has several advantages such as saving training time, better performance (in most cases), and not requiring a lot of data. Common approaches for Transfer Learning are [49]:

- **Training a model to reuse it** [49]:
  Let's assume you want to solve task A but need more data for training a deep neural network. One way to deal with this issue is to find another similar task B with large data, train the network on task B, and utilize this model as a starting point to solve task A . Whether the whole model is required to use or only a few layers depends on the problem.

- **Using a pre-trained model** [49]**:**
  The second method for transfer learning is applying a pre-trained model, which is most commonly used throughout deep learning. There are a lot of already pre-trained models available on the internet. The number of layers required to reuse and retrain is based on the problem.

- **Feature creation** [49]:
  The feature extraction method breaks down a large input data set into the most important features. Dimensionality reduction is used in this procedure to break enormous input data sets into smaller and more meaningful processing units. Neural networks can learn which features are vital and which aren't. Then, the important learned features can be applied to other tasks as well. This technique is mainly used in computer vision since it can decrease the size of your dataset and reduce computation time.



2.2.6 UNet Model for Medical Image Segmentation

The most used CNN architecture for medical image segmentation is U-Net, and it is the one that my work is based on. A symmetric encoder-decoder structure with skip connections makes up the UNet architecture.

The context of the image is captured by the encoder (the left part in figure(7)). The encoder's blocks are made up of two consecutive 3×3 convolutions, a ReLU activation unit, and a 2×2 max-pooling layer with a stride of two, which would downsample the image by a factor of two.Pooling and convolutional procedures Reduce the image's resolution, or downsample, to create a low-resolution version. By expanding the receptive fiel, the Max Pooling operation aids in understanding "WHAT" is there in the image. As a result, the location of the objects is often forgotten.

Knowing "WHAT" is there in the image is crucial for semantic segmentation, but it's also crucial to know "WHERE" it is present. In order to retrieve the "WHERE" information, we

must find a means to upsample the image from low resolution to high resolution. The best option for upsampling is Transposed Convolution. It uses backpropagation to learn parameters to transform a low-resolution image into a high-resolution image. This upsampling part is done in the decoder(the right part in figure(7)).The feature map is upsampled using 2x×2 up-convolution at each decoding stage.

Low-level information could be lost if the encoder and decoder layers are just stacked.Consequently, the decoder may have created segmentation maps with erroneous boundary lines. Concatenating the intermediate encoder outputs with the intermediate decoder outputs—also known as the skip connections—compensates for the lost data.

In order to decrease the feature map to the necessary number of channels and create the segmented image, an additional 1x1 convolution is done in the last step.

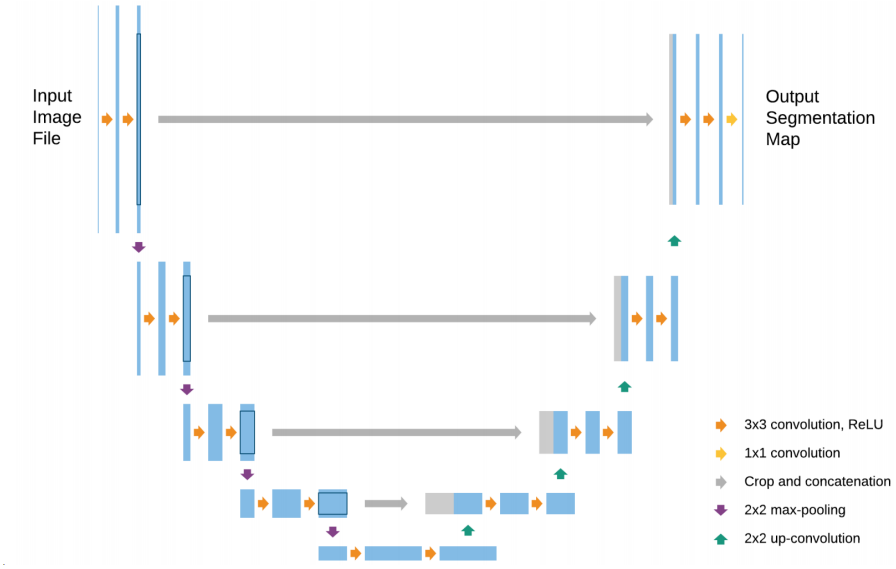


Figure 7: A U-net architecture. Each layer's feature map is displayed in the blue boxes. The cropped feature maps from the contracting path are shown in the gray boxes, and the arrows show the various operations [18, 50]

.

2.2.7 Performance Metrics of Machine Learning Algorithms

The next step after constructing a machine learning algorithm is to evaluate the model's performance using metrics and datasets. For comparing various Machine Learning Algorithms, there are numerous performance metrics available. The following are the typical metrics used to assess the efficiency of any developed segmentation method [51]:

- A true positive (TP) indicates that the class of real data and the class of predicted data are both true.
- A true negative (TN) indicates that the predicted data class and the actual data class are both false.
- A false positive (FP) occurs when the predicted data class is true ,but the actual data class is false.
- False negative (FN) indicates that while the class of predicted data is false, the actual data class is true.

- **Precision:** Precision is an evaluation metric that returns the percentage of input data cases that are reported as true. The formula to calculate the precision is given in equation (1)

$$\text{Precision } = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{1}$$

- **Recal:** The percentage of all relevant results in total that the model properly classified is given by recall, which is illustrated in equation (2).

$$\text{Recall } = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

- **F1 Score:** As shown in the following equation, the F1 score provides information on the model's accuracy. It is characterized as the harmonic average of the recall and precision values.

$$\text{F1 score } = \frac{2 * \text{ precision } * \text{ recall}}{\text{precision } + \text{ recall}} \tag{3}$$

- **Pixel Accuracy:** The percentage of pixels in an input image that the model properly classifies is indicated by the pixel accuracy.

$$\text{Pixel accuracy } = \frac{\text{number of correctly classified pixels}}{\text{overall number of pixels}}. \tag{4}$$

- **Intersection over Union (IoU):** IoU, often known as the Jaccard index, is a statistic frequently used to evaluate how well an image segmentation method is working. It is calculated by dividing the total area of union between the predicted segment mask and the ground truth mask by the amount of intersecting area between the anticipated image segment and the ground truth mask:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|},$$

  where A is the ground truth. Predicted segmentation is represented by B. Modern segmentation algorithms are assessed using Mean IoU. The mean IoU represents the average IoU for every class.

- **Dice Coefficient (DSC):** The DSC is a technique for determining whether two segmentations are similar. It divides the overlap size of the two points by the magnitude of the overlap between the prediction and the ground truth. The DSC value ranges from 0 to 1, with 0 denoting complete spatial overlap and 1 denoting no overlap at all. Using the following formula, the DSC is calculated:

$$\text{DSC} = \frac{2 * |A \cap B|}{|A| + |B|},$$

  in which A and B = predicted and ground truth masks and $\cap$ denotes intersection.

3 Applied Method for 3D Hip Structure Segmentation

3.1 Model Architecture Overview

Using a modified U-Net convolutional neural network architecture, the hip was automatically segmented in CT scans. The following actions were taken with this goal in mind:

- The data was gathered and pre-processed.
- The manually segmented masks were used to train the model.
- Subjects who had not participated in the training procedure were used to test this model.
- The method's block diagram is displayed in figure (8).
- Python, Tensorflow 2.11.0 with an Nvidia GPU platform, and the SIUE Cluster environment were used to implement our model.
-
  - The network takes in images with a resolution of 512×512 pixels and has 16 2D convolutional layers with a 4×4 kernel size. It consists of an image compression path (encoder), where the image size is gradually decreased, then an expanding path (decoder) that allows for a segmentation mask in the output with the same sizes as the input image.
  - There are eight layers in the compression route. In each layer, several 2D convolution filters with the kernel size of 4×4, padding = "same," and stride = 2 were used.A Leaky ReLU activation function with an alpha of 0.02 was also used in each layer. Batch normalization was then performed to each of these layers to prevent overfitting and increase accuracy. Each layer had more convolution filters than the one before it.
  - There are eight layers in the extending path as well. In addition to a ReLU activation function, several 2D transpose convolution filters with the parameters size = 4×4, padding = "same," and stride = 2 were used. There were fewer filters in each layer than in the one before it. Every layer except the first, second, and third layers had batch normalization applied to it.
  - Finally, the sigmoid activation function was applied to the final layer. Additionally, the Adam optimization function was used, with a 0.001 learning rate. Given that there are just two classes (background and hip), binary cross entropy has been employed as the loss function. A value between 0 and 1 has been assigned to each pixel, with 0 denoting the background and 1 denoting the hip. The threshold was considered 0.5 to classify each pixel as 0 or 1.

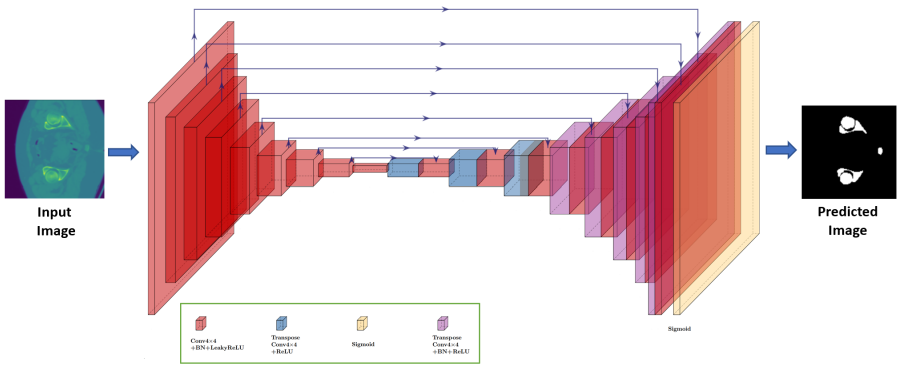


Figure 8: General architecture of network. The arrows are skip connections.

3.2 Transfer Learning

Transfer learning is a technique that makes use of a trained model developed for one task as the basis for a model for another activity. At the beginning of the learning process, machine learning models make a guess for the value of the weights that they want to learn. Transfer learning makes it easier to estimate parameter values roughly and avoid guesswork. This helps the model find the actual weight values more quickly. For this study, we employed a pre-trained model made up of 140 CT scans, each with the five organs (lung, bones, liver, kidneys, and bladder) labeled in three dimensions.

4 Experimental Results

4.1 Data Set Introduction

The hip image dataset include 21 CT scans of children with CP. Each image is 512×512 image patches. 80% of data for the train and 20% of it for the test.

4.2 Hip segmentation

A modified U-Net network introduced in the previous section and shown in Figure (8) was used to segment the hip in the CT image set. The result of the hip segmentation is shown in Figure (9). The segmentation of the hips visually matches the ground truth. The designed U-Net network was trained with 100 iterations. Table (4.2) depicts the quantitative evaluation of the hip segmentation network performance after 5 fold-cross validation.

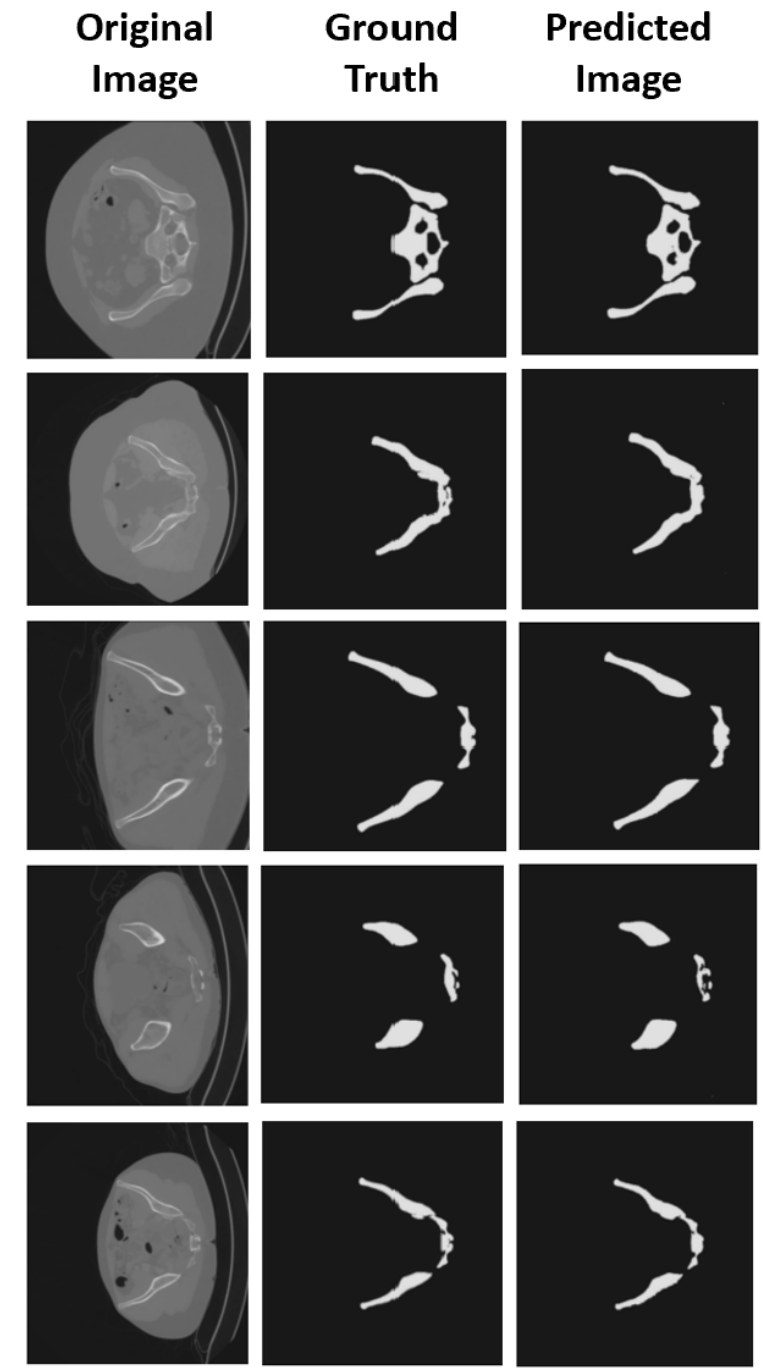


Figure 9: Hip segmentation results from validation data

Metrice	**Performance (%)**
DSC	0.99
Iou	0.99
Accuracy	0.98
Loss	0.01
Precision	0.99
Recall	0.98
F1-sore	0.99

Table 1: Evaluation of the implemented method results

4.3 Manual segmentation

Manual segmentation was done using materialize mimics v24. All segmentations were done by me using a combination of thresholding or automated bone segmentation tools with fine tuning through multislice threshhold techniques. The maps were then converted to parts and smoothed and wrapped. Map and part data were exported as dicom overlays for your use. Occasionally cases were sent to materialize 3 Matic v16 for refinement including local smoothing and removal of false connections before being exported. Cases took between 5-30 minutes depending on how either technique worked. Bone algorithms seemed to create more noise than soft tissue algorithms in a bone window and therefore required more time to exclude the noise from the segmentations.

4.4 Network loss and Accuracy

In this study, binary cross entropy was utilized to measure the model's accuracy during training and validation. The following figure shows the binary cross entropy Iou, loss and accuracy for hip segmentation.

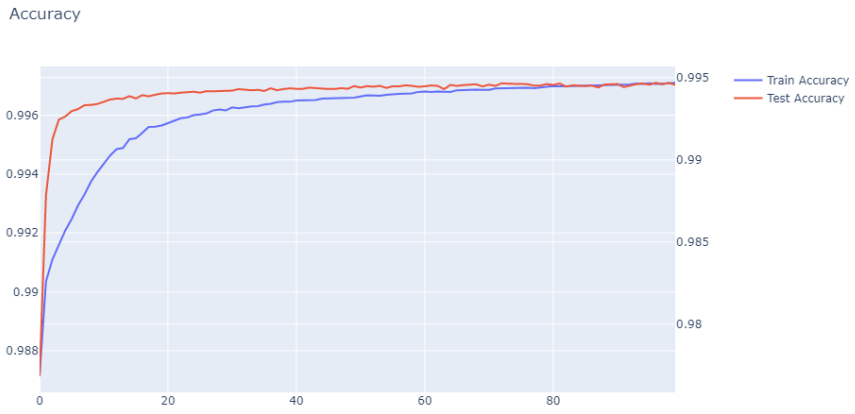


Figure 10: Accuracy

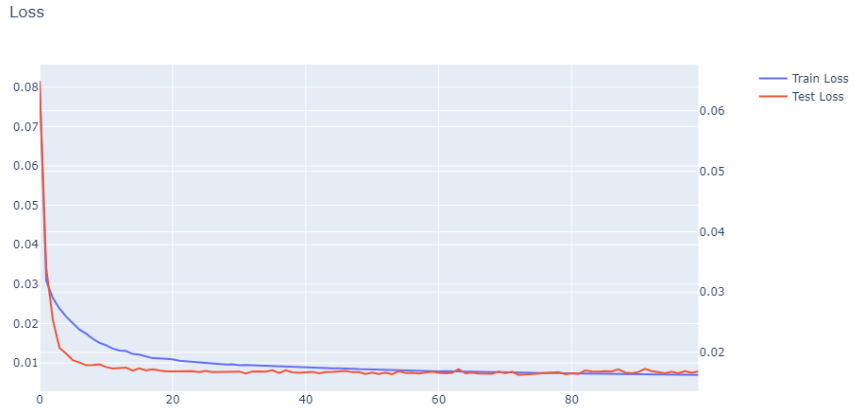


Figure 11: Loss

According to Figure (11), the hip model training loss and validation loss gradually decrease with each epoch, indicating a good learning rate. Furthermore, as can be seen in Figure(10), With each epoch, the hip model's training and validation accuracy increase. Due to the training and validation accuracy graphs being close to one another, the model does not appear to be overfitting.

4.5 Methods Comparison

A comparison of our method with the standard U-Net is provided in the table (4.5). Our applied method showed better performance than standard U-Net which demonstrate the effectiveness of our deep convolutional network. TL in this table is short form for Transfer Learning.

Study	**Method**	**Dice**	**Iou**	**Accuracy**	**Precision**	**Recall**
[18]	U-Net	0.95	0.92	0.99	0.92	0.97
This study (with TL)	Modified U-Net	0.98	0.97	0.98	0.97	0.98
This study (without TL)	Modified U-Net	0.99	0.99	0.98	0.99	0.98

5 Discussion

In this study, we found that the proposed model could effectively and accurately segment diseased hip joints in children with CP from their CT images. With the benefits of convolutional layers, we accelerated the process of learning and improved segmentation outcomes in unhealthy hip joints. The model we used in our experiment differs from the standard U-Net in a few aspects. First, we replaced the pooling layers with standard convolutional layers with stride two based on the research in [52]. The authors in [52] found that a convolutional layer with increased stride can be applied as an alternative to max-pooling without loss in accuracy. Second, we used batch normalization in each layer (except the first, second and third layers of the decoder) which U-Net does not include any batch normalization [18]. Furthermore, the number of layers and filters in each layer in our model is different than U-Net. The results of segmenting the sick hip joints using our suggested method were much better than those obtained using regular U-Net and manual segmentation.

6 Conclusion and Future Work

In this study, we implemented a modified U-net along with transfer learning and successfully used it to automatically segment the hip in CT scans of CP patients. The proposed technique achieved average dice score, recall, and IoU values of 0.99, 0.98, and 0.99, respectively. In summary, the segmentation method that has been suggested is effective for automatically segmenting the hip. However, the current study has some limitations. First, the network's general structure is traditional and lacks innovation in terms of feature extraction. Second, the task set is limited to segmenting the bone sections and there are no further diagnoses or scoring. But in the upcoming study, we intend to look into these issues.

References

[1] C. P. PROGRAM, *Cerebral Palsy Hip Disorders*.

[2] T.-T. Pham, M.-B. Le, L. H. Le, J. Andersen, and E. Lou, "Assessment of hip displacement in children with cerebral palsy using machine learning approach," *Medical & Biological Engineering & Computing*, vol. 59, no. 9, pp. 1877–1887, 2021.

[3] M. Wynter, N. Gibson, K. L. Willoughby, S. Love, M. Kentish, P. Thomason, H. K. Graham, and N. H. S. W. Group, "Australian hip surveillance guidelines for children with cerebral palsy: 5-year review," *Developmental Medicine & Child Neurology*, vol. 57, no. 9, pp. 808–820, 2015.

[4] F. Dobson, R. Boyd, J. Parrott, G. Nattrass, and H. Graham, "Hip surveillance in children with cerebral palsy: impact on the surgical management of spastic hip disease," *The Journal of bone and joint surgery. British volume*, vol. 84, no. 5, pp. 720–726, 2002.

[5] M. Sakamoto, Y. Hiasa, Y. Otake, M. Takao, Y. Suzuki, N. Sugano, and Y. Sato, "Automated segmentation of hip and thigh muscles in metal artifact contaminated ct using cnn," in *International Forum on Medical Imaging in Asia 2019*, vol. 11050, pp. 124–129, SPIE, 2019.

[6] Y. Chang, Y. Yuan, C. Guo, Y. Wang, Y. Cheng, and S. Tamura, "Accurate pelvis and femur segmentation in hip ct with a novel patch-based refinement," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1192–1204, 2018.

[7] T. Ogawa, M. Takao, T. Sakai, and N. Sugano, "Factors related to disagreement in implant size between preoperative ct-based planning and the actual implants used intraoperatively for total hip arthroplasty," *International journal of computer assisted radiology and surgery*, vol. 13, no. 4, pp. 551–562, 2018.

[8] D. Wu, X. Zhi, X. Liu, Y. Zhang, and W. Chai, "Utility of a novel integrated deep convolutional neural network for the segmentation of hip joint from computed tomography images in the preoperative planning of total hip arthroplasty," *Journal of Orthopaedic Surgery and Research*, vol. 17, no. 1, pp. 1–17, 2022.

[9] J. D. Wylie, P. A. Jenkins, J. T. Beckmann, C. L. Peters, S. K. Aoki, and T. G. Maak, "Computed tomography scans in patients with young adult hip pain carry a lifetime risk of malignancy," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 34, no. 1, pp. 155–163, 2018.

[10] J. Wells, J. J. Nepple, K. Crook, J. R. Ross, A. Bedi, P. Schoenecker, and J. C. Clohisy, "Femoral morphology in the dysplastic hip: three-dimensional characterizations with ct," *Clinical Orthopaedics and Related Research®*, vol. 475, no. 4, pp. 1045–1054, 2017.

[11] M. J. Beebe, J. D. Wylie, B. G. Bodine, A. L. Kapron, T. G. Maak, O. Mei-Dan, and S. K. Aoki, "Accuracy and reliability of computed tomography and magnetic resonance imaging compared with true anatomic femoral version," *Journal of Pediatric Orthopaedics*, vol. 37, no. 4, pp. e265–e270, 2017.

[12] A. K. Davison, *OUTCOME RESEARCH IN CHILDREN'S HIP DISEASE*.

[13] P. Rosenbaum, N. Paneth, A. Leviton, M. Goldstein, M. Bax, D. Damiano, B. Dan, B. Jacobsson, *et al.*, "A report: the definition and classification of cerebral palsy april 2006," *Dev Med Child Neurol Suppl*, vol. 109, no. suppl 109, pp. 8–14, 2007.

[14] C. Wang, *Medical Image Segmentation with Deep Learning*. PhD thesis, The University of Wisconsin-Milwaukee, 2020.

[15] H. Lu, Y. She, J. Tie, and S. Xu, "Half-unet: A simplified u-net architecture for medical image segmentation," *Frontiers in Neuroinformatics*, vol. 16, 2022.

[16] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[17] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully con-volutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2014.

[20] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[21] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, 2017.

[22] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[23] D. Golan, Y. Donner, C. Mansi, J. Jaremko, M. Ramachandran, *et al.*, "Fully automating graf's method for ddh diagnosis using deep convolutional neural networks," in *Deep Learning and Data Labeling for Medical Applications*, pp. 130–141, Springer, 2016.

[24] A. R. Hareendranathan, D. Zonoobi, M. Mabee, D. Cobzas, K. Punithakumar, M. Noga, and J. L. Jaremko, "Toward automatic diagnosis of hip dysplasia from 2d ultrasound," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 982–985, IEEE, 2017.

[25] Z. Zhang, M. Tang, D. Cobzas, D. Zonoobi, M. Jagersand, and J. L. Jaremko, "End-to-end detection-segmentation network with roi convolution," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pp. 1509–1512, IEEE, 2018.

[26] O. Paserin, K. Mulpuri, A. Cooper, A. J. Hodgson, and R. Garbi, "Real time rnn based 3d ultrasound scan adequacy for developmental dysplasia of the hip," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 365–373, Springer, 2018.

[27] C. G. Chee, Y. Kim, Y. Kang, K. J. Lee, H.-D. Chae, J. Cho, C.-M. Nam, D. Choi, E. Lee, J. W. Lee, *et al.*, "Performance of a deep learning algorithm in detecting osteonecrosis of the femoral head on digital radiography: a comparison with assessments by radiologists," *American Journal of Roentgenology*, vol. 213, no. 1, pp. 155–162, 2019.

[28] W. Gale, L. Oakden-Rayner, G. Carneiro, A. P. Bradley, and L. J. Palmer, "Detecting hip fractures with radiologist-level performance using deep neural networks," *arXiv preprint arXiv:1711.06504*, 2017.

[29] M. Adams, W. Chen, D. Holcdorf, M. W. McCusker, P. D. Howe, and F. Gaillard, "Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures," *Journal of medical imaging and radiation oncology*, vol. 63, no. 1, pp. 27–32, 2019.

[30] T. Urakawa, Y. Tanaka, S. Goto, H. Matsuzawa, K. Watanabe, and N. Endo, "Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network," *Skeletal radiology*, vol. 48, no. 2, pp. 239–244, 2019.

[31] A. Kazi, S. Albarqouni, A. J. Sanchez, S. Kirchhoff, P. Biberthaler, N. Navab, and D. Mateus, "Automatic classification of proximal femur fractures based on attention models," in *International Workshop on Machine Learning in Medical Imaging*, pp. 70–78, Springer, 2017.

[32] Y. Xue, R. Zhang, Y. Deng, K. Chen, and T. Jiang, "A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis," *PloS one*, vol. 12, no. 6, p. e0178992, 2017.

[33] G. Zeng, X. Yang, J. Li, L. Yu, P.-A. Heng, and G. Zheng, "3d u-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3d mr images," in *International workshop on machine learning in medical imaging*, pp. 274–282, Springer, 2017.

[34] F. Chen, J. Liu, Z. Zhao, M. Zhu, and H. Liao, "Three-dimensional feature-enhanced network for automatic femur segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 243–252, 2017.

[35] G. Zeng, Q. Wang, T. Lerch, F. Schmaranzer, M. Tannast, K. Siebenrock, and G. Zheng, "Latent3du-net: Multi-level latent shape space constrained 3d u-net for automatic segmentation of the proximal femur from radial mri of the hip," in *International Workshop on Machine Learning in Medical Imaging*, pp. 188–196, Springer, 2018.

[36] G. Zeng, F. Schmaranzer, C. Degonda, N. Gerber, K. Gerber, M. Tannast, J. Burger, K. A. Siebenrock, G. Zheng, and T. D. Lerch, "Mri-based 3d models of the hip joint enables radiation-free computer-assisted planning of periacetabular osteotomy for treatment of hip dysplasia using deep learning for automatic segmentation," *European journal of radiology open*, vol. 8, p. 100303, 2021.

[37] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, and Y. Sato, "Automated muscle segmentation from clinical ct using bayesian u-net for personalized musculoskeletal modeling," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1030–1040, 2019.

[38] C. Wang, B. Connolly, P. F. d. Oliveira Lopes, A. F. Frangi, and Ö. Smedby, "Pelvis segmentation using multi-pass u-net and iterative shape estimation," in *International Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging*, pp. 49–57, Springer, 2018.

[39] C. M. Deniz, S. Xiang, R. S. Hallyburton, A. Welbeck, J. S. Babb, S. Honig, K. Cho, and G. Chang, "Segmentation of the proximal femur from mr images using deep convolutional neural networks," *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018.

[40] *CT-Scan-vs-MRI.*

[41] *Contributors to Wikipedia projects,.*

[42] M. Anthony and P. Bartlett, *Neural network learning: Theoretical foundations*. cambridge university press, 1999.

[43] T. Wiederer, *Neural Networks in Javascript - webkid blog*, 2016.

[44] M. Khani, *Medical Image Segmentation Using Machine Learning*. PhD thesis, The University of Wisconsin-Milwaukee, 2021.

[45] A. Wiki, *Activation Function*, 2021.

[46] N. Yali, *A Multi-stage Convolution Machine with Scaling and Dilation for Human Pose Estimation*. PhD thesis, Thesis (Master degree Department of Electronic Engineering), Graduate School . . . , 2018.

[47] V. Cardozo, "A deep learning u-net for detecting and segmenting liver tumors," 2021.

[48] Wikipedia, *Typical CNN Architecture*, 2021.

[49] N. Donges, *What Is Transfer Learning? Exploring the Popular Deep Learning Approach.*

[50] S. Nahian, S. Paheding, E. Colin, and D. Vijay, "U-net and its variants for medical image segmentation: Theory and applications," *Preprint at arxiv*, vol. 1118, p. v1, 2011.

[51] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep neural networks for medical image segmentation," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[52] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.