3D Fusion Hierarchical Net Reconstruction from 2D Transcerebellar Images with Deep Learning

Abarham Martadiansyah, Hadrians Kesuma Putra, Muhammad Rizky Ramadhan, Ermatita, Abdiansah, and Erwin, *Member*, *IAENG*

***Abstract*—Reconstruction of 2-dimensional transcerebellar images into 3-dimensional transcerebellar has an important role in diagnosis and treatment planning in neurology. In this study, we propose the 3D-FHNet method to produce an accurate three-dimensional representation of transcerebellar images. The process begins with the selection of images and the application of image augmentation and enhancement techniques to improve the quality of the initial image. Furthermore, segmentation was carried out using two different architectures, namely U-Net and LinkNet, to compare the performance of the two. After training, both architectures can process object segmentation properly. The best performance was produced by U-Net with a pixel accuracy of 99.83%, Mean IU of 89.71%, FPR of 0.91%, Precision of 85.78%, Recall of 85.31%, and F1 Score of 85.31%. With this accuracy, cross-validation was carried out using a 10-Fold. The experimental results show that U-Net gives better results in terms of transcerebellar image segmentation. After the segmentation process, an initial reconstruction is carried out using the PiFUHD architecture which produces a three-dimensional object. The results of this reconstruction are then taken from four sides to be introduced to the 3D-FHNet architecture, because the main concept of 3D-FHNet is to use multiview input. To measure the accuracy of the model, the IoU (Intersection over Union) metric is used for the ground truth obtained from the PiFUHD method. This metric is used to compare the similarities between the reconstruction results and the ground truth, so that information can be obtained about the extent to which the model has succeeded in reconstructing three-dimensional objects accurately. Model performance can be calculated by using the input as ground truth or the IoU (Intersection over Union) method with an average of 76.76%. By using the 3D-FHNet method, three-dimensional image reconstruction of the ultrasound image of the fetal head can be performed accurately.**

The experimental results show that the 3D-FHNet method produces a more accurate three-dimensional representation of the transcerebellar image compared to the previous method. These results demonstrate great potential in improving diagnosis and treatment planning in neurology, as well as making important contributions to the development of medical image processing technologies.

***Index Terms*—3D- Deep Learning, FHNet, LinkNet, Transcerebellum, U-Net**

Manuscript received June 21, 2023; revised January 18, 2024. This work was supported in part by the Direktorat Riset, Teknologi, dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi under Grant No. 164/E5/PG.02.00.PL/2023.

Abarham Martadiansyah is an Assistant Professor of Medical Faculty from Sriwijaya University, Indralaya, 30662, Indonesia (e_mail: dr_abarham_martadiansyah@yahoo.com).

Hadrians Kesuma Putra is an Assistant Professor of Medical Faculty from Sriwijaya University, Indralaya, 30662, Indonesia (e_mail: hadrianskesuma@fk.unsri.ac.id).

Muhammad Rizky Ramadhan is an undergraduate student of Computer Engineering Department, Computer Science Faculty Universitas Sriwijaya, Indralaya, 30662, Indonesia (e_mail: m.riskyramadhan2001@gmail.com).

Ermatita is an Associate Professor of Information System Department, Computer Science Faculty, Universitas Sriwijaya, Indralaya, 30662, Indonesia (e_mail: ermatita@unsri.ac.id).

Abdiansah is an Assistant Professor of Informatics Engineering Department, Computer Science Faculty, Universitas Sriwijaya, Indralaya, 30662, Indonesia (e_mail: abdiansah@unsri.ac.id).

Erwin is a Professor of Computer Engineering Department, Computer Science Faculty, Universitas Sriwijaya, Indralaya, 30662, Indonesia (Corresponding author, phone: 6281394497590, email: erwin@unsri.ac.id).

I. INTRODUCTION

A FIELD of research that currently continues to exist in the medical sector is computer graphics. Reconstruction of three-dimensional objects is one area of computer graphics research that continues to exist today. This is because more and more software requires the results of the reconstruction process, for example VR (Virtual Reality) technology, automatic 3D modeling, AR (Augmented Reality), and other examples. There are several other methods that continue to be developed in this field of reconstruction. In general, these methods can be grouped into two groups, namely image-based modeling and image-based rendering. Both of these methods have differences that lie in the process used to reconstruct the object. Where in image based modeling many mathematical calculations are carried out to reconstruct objects into 3D. Whereas in image-based rendering, the process is carried out by analyzing a two-dimensional image to determine which part is a three-dimensional object, the result of which is a synthetic image in the form of a reconstructed three-dimensional object. The image based rendering method has the advantage of easily reconstructing photorealistic objects. Meanwhile, image based modeling has a weakness for reconstructing photorealistic objects that must use other methods, such as texture mapping[1].

It is undeniable, this is a challenge in knowing the condition of the fetus whose main goal is to find the parameters contained in the fetal head circumference. Transcerebellar parameters themselves can be characterized by the presence of Cavum Septi Pelucidi (CSP), Frontal Horn, and Choroid Plexus. To observe and find out some of these factors, a tool is needed to be able to detect parameters in the fetal head which is often referred to as Ultrasonography (USG)[2]. Ultrasound or what we generally know as an ultrasound tool is a diagnostic technique for displaying images or images[3]. The reconstruction stages cannot be carried out directly with raw data in the form of ultrasound images only, but rather through several stages starting with the process of image enlightenment, data augmentation, parameter segmentation until finally the data can be trained and tested for 3D reconstruction. In the early stages, image processing is carried out to improve the image starting from reducing noise and increasing image light. The augmentation process is carried out with the aim of increasing the dataset to be used. The dataset which was initially only a single image was expanded with several related methods that might be implemented. The reconstruction activity itself also requires additional architecture to assist the image segmentation process before detection activities. Usually, the manual segmentation process requires high knowledge and qualified experience and also spends a long period of research. On this basis, this study will implement several architectures that will be used for the process of automatic segmentation of the fetal head which will reduce research travel time[4]–[7].

Reconstruction of 3D images or images is a part that needs to be taken into account in the world of medical imaging. The creation of a 3-dimensional model that is done manually usually requires a long time and a large cost. For this reason, many solutions are in the form of techniques that are still being developed that may be used to reconstruct 2D images into 3D images. 2D images generated from image capture devices of human organs can obtain and store information that other images do not have, such as patient-related information and image pixel information [5]. The process of reconstructing a 2D image into a 3D image can help experts and related users to better understand the special information contained in the image, so that in the process of general and medical diagnoses, images can be more precise and accurate[6].

In medical imaging, fetal ultrasound results that can be imaged into several parts of the image can be used as 3D images, to increase clarity and assist medical personnel and patients in obtaining and understanding in detail the body parts of the fetus from ultrasound images using several modalities. There are several techniques for reconstructing a 2D image into a 3D image, for example: Surface Rendering which includes a classic projection method, where the process of rendering the scene is done by projecting polygons onto a platform [7]. This technique is usually used in making games where the weakness is eliminating all existing information and only projecting the outermost part of the captured image. Furthermore, Generalized Voxel Coloring Layer Depth Image is a 3D reconstruction method from a collection of 2D images [8]. Generalized Voxel Coloring produces 3-dimensional image reconstruction results using the theory of visibility and changing camera movements. In addition, Volume Rendering with other methods (Ray Casting) in it that can be expanded and presents results that are close to perfect, close to photorealistic. PIFuHD 3D method which is a multi-level architecture that can be trained from end to end of the image. In observing all images at low resolution and will focus on holistic reasoning. The PIFuHD method projects a high-level context that approximates very detailed geometries by observing high-resolution images. Lastly is the 3DFH-Net method, where this architecture uses the results of the PIFuHD architecture which rotates the corners to obtain images with different views before projecting them into 3-dimensional objects.

II. Theoretical Basis

A. Related Work

Similar research discussing 2D to 3D reconstruction using the texture based method has previously been carried out, but due to limited references and related programs, this topic is currently being discussed a little.

CT Imaging is a non-invasive medical imaging technique that produces detailed cross-sectional images of the body. These images are created by taking multiple X-ray images from various angles and using a computer algorithm to reconstruct a 3D image of the object being imaged. In the context of object reconstruction, CT imaging can be used to scan objects and create high-resolution 3D models. This technique is particularly useful in fields such as archaeology, where it can be used to create digital models of artifacts or structures that are too fragile or valuable to handle directly.

Other techniques used in 3D object reconstruction include laser scanning, structured light scanning and photogrammetry. This method involves taking multiple pictures or measurements of an object from different angles and using a computer algorithm to create a 3D model. Overall, 3D object reconstruction is a rapidly growing field with many applications in medicine, engineering, architecture and other fields. Advances in technology and algorithms continue to improve the accuracy and efficiency of 3D reconstruction, making it an increasingly important tool in many fields of research and industry.

Texture mapping is a method used to add detail, surface texture (a bitmap or raster image), or color to a 3-dimensional model. This method was used in a 3-dimensional graphic model by Edwin Catmull in 1974 [8].

Prior to that, there was also research conducted using the Generalized Voxel Coloring (GVC) method. Generalized Voxel Coloring (GVC) is a 3D computer graphics technique used to render and color complex models with high visual fidelity. It is based on the principle of voxelization, which involves the representation of continuous objects as discrete collections of regularly spaced voxels or small 3D pixels. In GVC, a set of color values are assigned to each voxel based on the material properties of the object being modeled. The color value is determined by considering the interaction between the material and the light source in the scene. For example, the color of a voxel can be affected by factors such as the angle and intensity of light hitting it, as well as material properties, such as reflectivity or transparency. GVC is a flexible technique that can be used to render a wide variety of objects, from simple geometric shapes to complex and highly detailed models.

This is particularly useful for rendering objects with complex surface properties, such as skin, fur, or fabrics, where traditional shading techniques may not be sufficient to capture subtle variations in color and texture. One of the strengths of GVC is its ability to handle dynamic and interactive scenes, where objects or light sources may move. This is achieved by efficiently updating the voxel color values in real-time as the scene changes.

Automatic measurement of fetal head circumference using 2D ultrasound images is a computerized method used to measure the size of the developing fetal head during pregnancy. This is a non-invasive technique that involves analyzing 2D ultrasound images to extract measurements of the fetal head circumference. Measurement of fetal head circumference is an important indicator of fetal growth and development. This can help identify potential problems such as fetal macrosomia (excessive growth of the fetus), microcephaly (abnormally small head size), or other abnormalities of the fetal brain or skull.

Traditionally, measurement of fetal head circumference has been done manually by trained sonographers or physicians using calipers to measure certain markers on the fetal head in ultrasound images. However, this method is time consuming and prone to measurement errors. Automatic measurement of fetal head circumference using 2D ultrasound images offers a faster and more accurate alternative. This technique involves using a computer algorithm to detect the fetal head in ultrasound images and automatically measure the distance between certain markers on the head. Several studies have shown that automatic measurement of fetal head circumference using 2D ultrasound images is a reliable and accurate method of measuring fetal head circumference, with an accuracy equal to or better than manual measurements performed by trained sonographers or physicians. Overall, automatic measurement of fetal head circumference using 2D ultrasound images is a promising technique for increasing the accuracy and efficiency of fetal growth monitoring during pregnancy. This has the potential to improve prenatal care and help identify potential problems early, which can lead to better outcomes for both mother and baby [9].

Initially this method was used to cover and map pixels from what was originally just a texture to a 3-dimensional surface. Finally, now the technique is often referred to as diffuse mapping to distinguish it from other complex mappings. There are many other complex variations on the technique that make it possible to simulate nearphotorealism in real time, by reducing the number of polygons and light counts required to build a more realistic and functional 3D screen[10].

Compression-based 3D texture mapping for real-time rendering is a computer graphics technique used to efficiently render complex, high-resolution 3D textures in real time. This technique involves compressing 3D texture data using compression algorithms that reduce the amount of storage and processing required to render textures.

Texture mapping is a technique used in computer graphics to apply 2D images or patterns to the surface of a 3D model. This gives the model a more realistic look and can be used to add details such as color, texture and shading. However, applying textures to complex 3D models can be computationally expensive, especially when high-resolution textures are used. Compression-based 3D texture mapping addresses this problem by compressing texture data, which reduces the amount of memory and processing power required to render textures. There are several compression algorithms used in compression-based 3D texture mapping, including wavelet-based compression and lossless compression. This algorithm is optimized for real-time rendering, enabling high-quality texture mapping without sacrificing performance[11].

3D PIFuHD is a method built on the pixel aligned implicit function (PIFu) framework recently introduced by metaverse, which takes a 512 x 512 resolution image as input and produces a low resolution (128 x 128) filter embedding[12]. To achieve better resolution output, they stack additional pixel-aligned prediction modules on top of the framework , where the module takes high-resolution images as input (1024 x 1024) and encodes them into high-resolution image features (512 x 512). The second module embeds high-resolution features as well as 3-dimensional embedding from the initial module to predict object probability fields. To improve the quality and accuracy of the reconstruction results, what must be done is to predict the normal map or mapping for the front and back sides in the image object space, and enter it into the network as additional input [13].

B. Image of The Fetal Head

An image is an image that is in a two-dimensional plane obtained from a two-dimensional and continuous analog image into a discrete image [14]. Image is a visual representation of an object, scene or concept. They can be captured through various means such as photography, painting, drawing or digital imaging. Images can be two-dimensional (2D) or three-dimensional (3D), and can be static or dynamic. 2D imagery is a flat representation of an object or scene, and is commonly used in print materials such as books, magazines and newspapers, as well as on digital platforms such as websites, social media and mobile applications. Examples of 2D imagery include photographs, illustrations, diagrams and graphs.

Ultrasonography (USG) is one of the imaging diagnostic fields used to examine the condition of vital organs in humans. Ultrasound itself can study the shape, anatomical size, movement and relationship with the surrounding tissue [15]. Ultrasonography is very helpful in the diagnosis and screening of congenital abnormalities in the prenatal period. Ultrasound examination at 18 and 22 weeks has become the standard in the detection of fetal structural abnormalities for several decades. The results of the report show that the detection of congenital abnormalities using ultrasound ranges from 15 – 85% and is influenced by many factors including the gestational age at the time of detection, the ability of the sonographer, the mother's body mass index and the organ systems examined.

C. Image

An image is an image that is in a two-dimensional plane obtained from a two-dimensional and continuous analog image into a discrete image[16]. Image is a visual representation of an object, scene or concept. They can be captured through various means such as photography, painting, drawing or digital imaging. Images can be two-dimensional (2D) or three-dimensional (3D), and can be static or dynamic. 2D imagery is a flat representation of an object or scene, and is commonly used in print materials such as books, magazines and newspapers, as well as on digital platforms such as websites, social media and mobile applications. Examples of 2D imagery include photographs, illustrations, diagrams and graphs.

D. Transcerebellar

Trancerebellar or transcerebellum refers to the part of the brain that includes the brain region outside the cerebellum. This is the term used to describe functional brain networks that are linked to and involved in various cognitive processes, including working memory, attention, and executive function. The transcerebellar consists of several different parts of the brain, including the dorsal lateral prefrontal cortex, the parietal cortex, and the temporal cortex. These parts work together to support complex cognitive processes and movements.

E. Convolutional Neural Network

In digital image processing, image or image segmentation is the most commonly used technique and analysis to partition or divide an image into parts or regions, often based on the characteristics of the pixels in the image. Image segmentation can also be interpreted by grouping pixel areas based on similarity in color or shape. This activity requires the CNN method, namely U-Net and LinkNet as a comparison.

F. U-Net

U-Net is one of the CNN architectures that is widely used for segmentation in the medical field [17],[18]. Image segmentation is the process of dividing an image into several regions or segments, each of which corresponds to a different object or part of an object[19]. This is an important task in fields such as medical imaging, computer vision and robotics.

The U-Net architecture consists of two main parts: encoder and decoder. Encoder is a series of convolutional layers which downsamples the input image, whereas decoder is a series of up-convolutional layers which upsamples the encoded features and produces the final segmentation output[20]. The encoder and decoder are linked by a bottleneck layer, which stores spatial information and helps the network learn segmentation boundaries more accurately[21]. The U-Net architecture is shown in Figure 1.

Overall, U-Net is a powerful architecture for image segmentation that has been widely used in various fields. Its success can be attributed to its ability to combine deep learning techniques with efficient processing of spatial information, making it especially suitable for tasks requiring accurate segmentation of complex objects in images.

G. LinkNet

LinkNet is a type of Convolutional Neural Network (CNN) architecture that is commonly used in image segmentation processes. This architecture was developed by researchers at Facebook AI Research and published in the paper "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation" in 2017. This architecture is designed to be computationally efficient, and makes this architecture a good choice for applications with related architecture requirements with limited resources, such as on mobile devices or embedded systems. Figure 2. is designed to be computationally efficient, and makes this architecture a good choice for applications related to architectural needs with limited resources, such as mobile devices or embedded systems.

H. PiFUHD : Multi-Level Pixel-Aligned Implicit Function for High Resolution 3D

Specifically, function (*f*) first extracts embedded image features from 2D locations projected at *(X) = x R2*, which we denote by *(x, I)*. Orthogonal projection is used to estimate the query 3D point *X*, and *x = (X) = (Xx, Xy)*. Then, it estimates the occupancy of the query 3D point *X*, and thus it is obtained. The equation can be seen in (1).

$$f(\mathbf{X},\mathbf{I}) = g\left(\Phi\left(\mathbf{x},\mathbf{I}\right), Z\right), \tag{1}$$

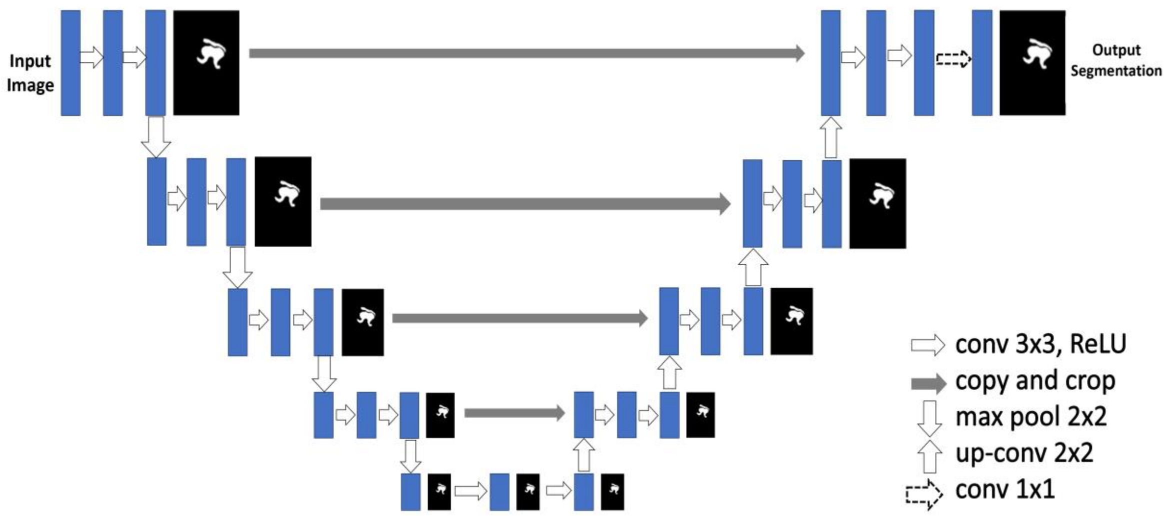


Fig. 1. U-Net architecture with ReLU activation function

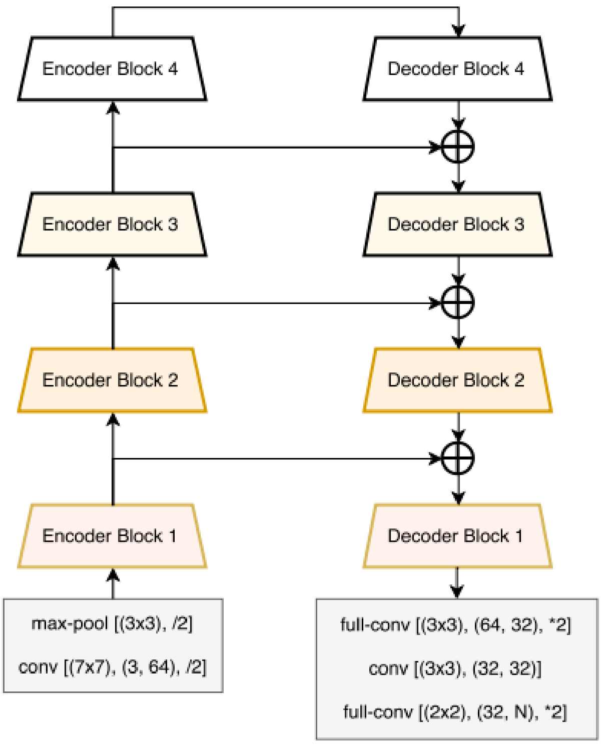


Fig. 2. LinkNet Architecture is used as a comparison in this resear

where $Z = Xz$ is the depth along the ray determined by the 2Dx projection. Note that all 3D points along the same ray have exactly the same image features *(x, I)* from the same projected location *x*, and thus function (*g*) must focus at varying input depth *Z* to clarify the 3D Point occupancy in along the rays. In a Convolutional Neural Network architecture (CNN) is used for the 2D feature embedding function and Multilayer Perceptron (MLP) for function (*g*).

PIFuHD is a method built on top of the pixel aligned implicit function (PIFu) framework recently introduced by metaverse, which takes a 512 x 512 resolution image as input and produces a low resolution (128 x 128) filter embedding. To achieve better resolution output, they stacked additional pixel-aligned prediction modules on top of the framework, where the modules take high-resolution images as input (1024 x 1024) and encode them into high-resolution image features (512 x 512). The second module embeds high-resolution features as well as 3-dimensional embedding from the initial module to predict object probability fields. To improve the quality and accuracy of the reconstruction results, what must be done is to predict the map or normal mapping for the front and back sides in the image object space, and enter it into the network as additional input. In order to further improve the quality and accuracy of the reconstruction, what should be done first is to predict the normal maps for the leading and trailing edges in the image space, and feed them to the network as additional input [13].

Large-scale datasets synthetically generated by rendering hundreds of high-quality scanned 3D human net models were used to train f functions in an end-to-end manner. Unlike voxel-based methods, PIFu does not produce discretized volumes as output, so training can be done by sampling 3D points and calculating occupancy loss at sample locations, without generating a 3D mesh. During inference, the 3D space is uniformly sampled to infer occupancy and the final iso surface is extracted with a threshold of 0.5 using marching cubes[22].

I. 3 Dimensional Reconstruction (3D-FHNet)

Reconstruction is a technique of building and restoring form based on initial events, where the process contains primary values that must still exist in the activity of rebuilding something according to the initial conditions. Previously, 3-dimensional reconstruction was a key issue in many computer vision research areas such as pattern recognition, reverse engineering, and industrial inspection. In recent decades, 3D models have been made manually, which is time-consuming.

The 3D-FHNet method is a hierarchical 3-dimensional reconstruction method that combines a single view and a number of multiple reconstruction views and obtains accurate reconstruction results. This feature proposes a combination of multi-view features that enable this model to improve the reconstruction performance. In figure 3, a two-dimensional encoder will extract image features, map them to three-dimensional features through a feature mapping unit, and decode the predicted voxel probability through a three-dimensional decoder[23].

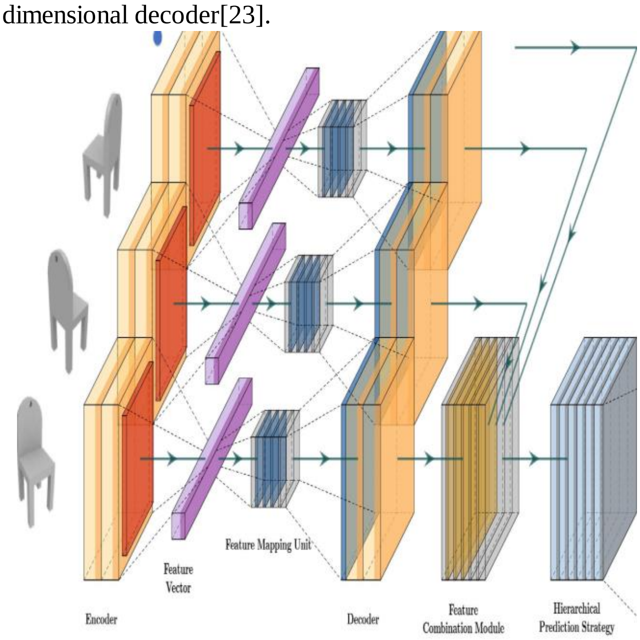


Fig. 3. 3D-FHNet with a two-dimensional encoder to three-dimensional

J. Evaluate Model

Evaluating a model involves measuring its performance and effectiveness in segmentation. Here are the explanations of the evaluation metrics you mentioned:

1. **Pixel Accuracy**: Pixel accuracy measures the percentage of correctly classified pixels in the predicted output compared to the ground truth. It provides an overall assessment of the model's ability to classify pixels correctly.
Pixel Accuracy = (Number of correctly classified pixels) / (Total number of pixels)

2. **Mean IU (Intersection over Union):** Mean IU calculates the intersection over union for each class in the predicted output and then takes the average across all classes. It

measures the overlap between the predicted and ground truth regions and provides a measure of segmentation accuracy.
Mean IU = (1 / Number of classes) * Σ (Intersection of class i) / (Union of class i)

3. **Mean Accuracy**: Mean accuracy calculates the average accuracy across all classes. It measures the overall correctness of the model's predictions, taking into account both true positive and true negative predictions.
Mean Accuracy = (1 / Number of classes) * Σ (Accuracy of class i)

4. **FPR (False Positive Rate)**: The false positive rate represents the proportion of false positive predictions out of all negative samples. It measures the model's tendency to incorrectly classify negative instances as positive.
FPR = (Number of false positive predictions) / (Number of actual negative instances)

5. **Precision**: Precision calculates the proportion of true positive predictions out of all positive predictions. It measures the accuracy of positive predictions, indicating the model's ability to avoid false positive errors.
Precision = (Number of true positive predictions) / (Number of true positive predictions + Number of false positive predictions)

6. **Recall**: Recall, also known as sensitivity or true positive rate, calculates the proportion of true positive predictions out of all actual positive instances. It measures the model's ability to identify positive instances correctly, indicating the ability to avoid false negative errors.
Recall = (Number of true positive predictions) / (Number of true positive predictions + Number of false negative predictions)

7. **F1 Score**: The F1 score is the harmonic mean of precision and recall. It provides a single metric that balances both precision and recall, taking into account false positives and false negatives. It is commonly used as a measure of overall model performance.
F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

These evaluation metrics are commonly used in various machine learning tasks, including image classification, segmentation, and object detection, to assess the performance and quality of the models.

III. Methodology

A. Hardware and Software Environment and Specifications

The Table I below contains specifications for hardware or hardware components that can be used to carry out data processing. The following are related hardware specifications

TABLE I.
HARDWARE SPECIFICATION

OS	Windows 10
Processor	Intel(R) Core(TM) i3-7020U CPU@2.30GHz 2.30 GHz
RAM	4.00 GB
SSD/HDD	500 GB
OS Type	64-Bit
VGA	Intel® HD Graphic 620

In the Table II below, describes the software components that will be used to carry out data processing. The following are the specifications of the software used.

TABLE II.
THE SOFTWARE USED

OS	Windows 10
Text Processing	Google Colab, Jupyter Notebook, Spyder
3D VIsualization	Blender
Language	Python

B. Framework

The framework describes the process of reconstructing a 2D image into a 3D image using the 3D-FHNet method. The following is an explanation of the stages in the framework in figure 4:

Image input: The first stage is to enter a 2-dimensional image that will be reconstructed into a 3-dimensional transcerebellar image[24].

Preprocessing: The subject image goes through a preprocessing stage to improve its quality and clarity. This process may include increasing contrast, removing noise, and adjusting other image parameters. Augmentation: The processed image then undergoes an augmentation process to increase data diversity.

Augmentation can include shifting, rotating, cropping, or other transformations of the image.

Pixel selection: At this stage, pixels that are relevant for 3D image reconstruction are selected. This can involve selecting pixels based on certain criteria, such as pixels in the transcerebellar region.

Segmentation: Images that have gone through the pixel selection stage then undergo a segmentation process, in which the objects in the image are separated and grouped. Segmentation can be performed using methods such as U-Net or LinkNet to produce better segmentation results[25].

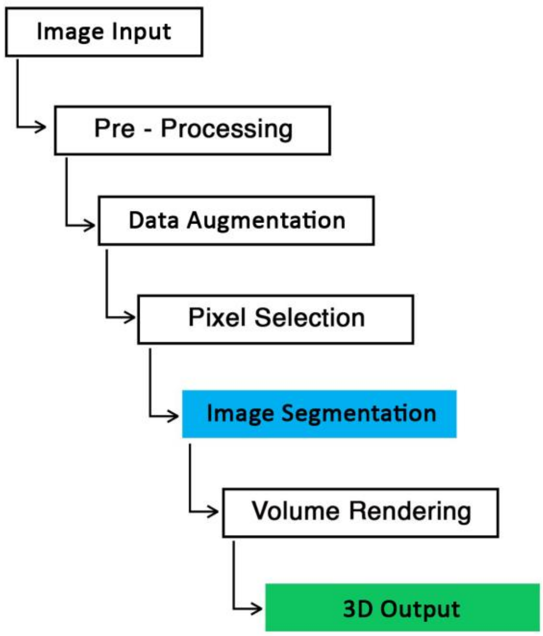


Fig. 4. The framework of reconstructing a 2D image into a 3D image using the 3D-FHNet method

Volume rendering: After segmentation, the reconstructed 3D images are taken from the results of 2D image segmentation. The volume rendering process is used to combine these images into a 3D image that is more comprehensive and shows the object structure in three dimensions[26].

3D Output: The final result of the image reconstruction process is a 3D transcerebellar image that can be used for further analysis, visualization, or other medical applications.

This framework describes the main steps in the reconstruction of a 2D image into a 3D transcerebellar image using the 3D-FHNet method, which includes preprocessing, augmentation, segmentation, and volume rendering[27].

C. Method and Flowchart

Figure 5 is the method and framework for this research. It can be seen that there are 4 major stages. The first stage is the stage where the data acquisition process will be carried out before entering the data pre-processing stage. Data is divided into 2 for split data (Training and Testing) and Unseen Data. After that the train and test data are divided into 80:20 for split. Before pre-processing, a data augmentation process is carried out on the train data to increase the resulting dataset. This stage includes the flip and rotation processes. The second stage, data pre-processing begins with the image input process, improving image quality with a denoising filter, and ending with image quality sharpening (CLAHE). The next stage is object segmentation. This research chooses U-Net as the segmentation architecture and LinkNet as the comparison architecture.

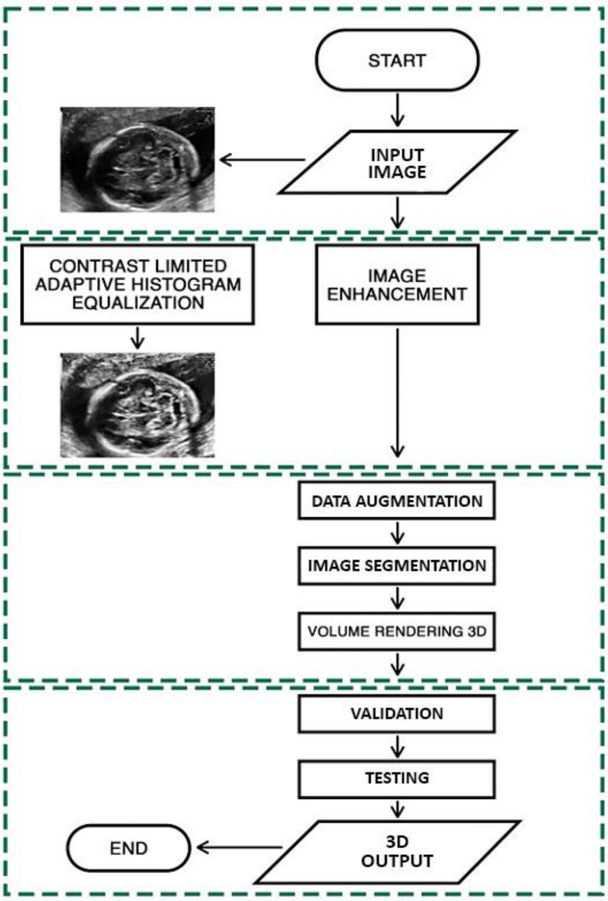


Fig. 5. The method and framework for this research

From the segmentation results, cross validation will be carried out using K-Fold to validate the best model produced. The final stage is 3-dimensional reconstruction. Starting from a single view reconstruction using the PiFuHD method, the results of the PiFuHD method are processed to produce views that will be used for the next reconstruction method, namely 3D-FHNet. In this method, the input used is multi view. This method will produce the same 3D object with the extension .obj which can be rotated using a 3D viewer application. The final results of the reconstruction model will be evaluated against the model used.

IV. Results And Analysis

A. Object Segmentation

A total of 1,960 images were used in the segmentation stage which were divided into training data and testing data. The training data consists of 1,568 images and the train data consists of 392 images. This data is used for the segmentation process using the U-Net architecture. Hyperparameter tuning in this research uses batch size and epoch. This segmentation stage uses a model which can be seen in Table III.

After the U-Net architecture, training is carried out using the LinkNet architecture as a comparison architecture. A total of 1,960 images were used in the segmentation stage which were divided into training data and testing data. The training data consists of 1,568 images and the train data consists of 392 images. Hyperparameter tuning in this model also uses batch size and epoch. Segmentation results using this model can be seen in Table IV.

B. 3 Dimensional Reconstruction

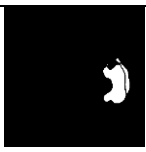
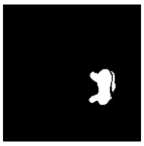
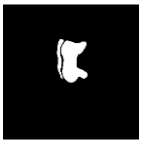
This process starts from the PIFuHD 3D method to obtain an initial 3D image before using the 3D-FHNet method.

TABLE III.
Transcerebellar segmentation results with U-Net architecture

File Name	Before Segmentation	After Segmentation
Patient01565_Plane3_2_of_4		
Patient01565_Plane3_3_of_4		
Patient01599_Plane3_1_of_2		

The PIFuHD 3D method starts by entering the raw data stack resulting from pre-processing, the step is to determine x, y and z, where x is the width value of the image, y is the height value of the image, and z is the number of frames in the raw image stack. Next, create a new frame buffer, create a new texture that is used for the color buffer and create a depthbuffer.

TABLE IV.
TRANSCEREBELLAR SEGMENTATION RESULTS WITH LINKNET ARCHITECTURE

File Name	Before Segmentation	After Segmentation
Patient01565_Plane3_2_of_4		
Patient01565_Plane3_3_of_4		
Patient01599_Plane3_1_of_2		

When all buffers have been created then they can be moved to a new buffer. In this section, the texture and vertex coordinates that correspond to the quad are determined. Using orthogonal projection, set texture coordinates *(0,0)* mapped to vertex *(-1,-1)* and texture coordinates *(1,1)* mapped to vertex (*1,1)*. Texture coordinates *(0,1)* are mapped to *(-1,1)* and texture coordinates *(1,0)* are mapped to *(1,-1)*. This process continues for each quad depicted on the *z* axis. The *z* position of the vertex requires some changes to make it in the range *(-1)* to *(+1)*.

The next step is to delete the black color in each frame. This process in OpenGL is also called an alpha test. The alpha value and alpha criteria value can be set. Check the pixel value which has been set to an alpha value of 0 (fully transparent) and for other pixel values it can be set at 255. This temporary buffer is prepared for the texture. If there are different alpha levels, then the alpha data is the same as the luminance data, which means a pixel with a luminance of 0 will have the same alpha value of 0, a pixel with the number 255 will have the same value of 255 and a pixel with a value between the two ranges it will get the same alpha value too.

The next step in the volume rendering process with this method is applying blending. When different images are mixed together, for various reasons (changing light conditions, schematic effects) the intensities of adjacent pixels are different enough to produce an image that does not match the desired one. Therefore, blending will be implemented to avoid problems caused by different pixel intensities. This process takes place for each frame to be rendered, for example the data used has 100 frames, then the steps in the volume rendering process will be repeated 100 times. Likewise, if the data has more frames, the volume rendering process steps will be repeated as many times as there are data frames.

PIFuHD is a method built on top of the metaverse's newly introduced Pixel-aligned Implicit Function (PIFu) framework. This technique takes images with a resolution of 512x512 as input and obtains feature embeddings with a low resolution (128x128). Table V shows the reconstruction results using the PiFuHD method.

The 3D-FHNet method is a hierarchical 3-dimensional reconstruction method that combines a single view and a number of multiple reconstruction views and obtains accurate reconstruction results. This feature proposes a combination of multi-view features that allows this model to improve reconstruction performance. Table VI shows the reconstruction results using the 3D-FHNet method.

C. Evaluation Score

The U-Net model used in this study shows good accuracy with various evaluation metrics. Pixel Accuracy of 97.39% indicates that the majority of pixels in the reconstructed image have been classified correctly. The IU average of 81.93% indicates a good level of comfort between the prediction results and the ground truth. The Mean Accuracy was 87.27%, indicating the overall proportion of correctly classified pixels. An FPR of 0.91% indicates a low positive misclassification rate. The precision of 85.61% describes the level of accuracy in classifying positive objects. Recall reached 75.44%, an indication of the model's overall ability to find positive objects.

The F1 score of 79.25% reflects the balance between prediction precision and completeness. These results indicate that the U-Net model is successful in reconstructing 2D images into 3D transcerebellar images with a high degree of accuracy and can be an effective alternative in the development of medical image processing technology.

TABLE V.
RESULTS OF 3D TRANSCEREBELLAR RECONSTRUCTION WITH PIFUHD

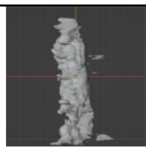
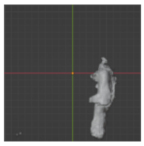
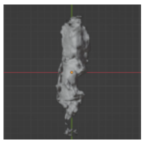
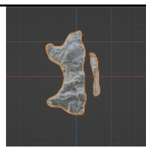
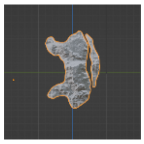
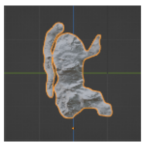
File Name	Raw Images	After Reconstruction
Patient01565_Plane3_2_of_4		
Patient01565_Plane3_3_of_4		
Patient01599_Plane3_1_of_2		

TABLE VI
Transcerebellar 3D reconstruction results with 3D-FHNet

File Name	Raw Images	After Reconstruction
Patient01565_Plane3_2_of_4		
Patient01565_Plane3_3_of_4		
Patient01599_Plane3_1_of_2		

In model testing, segmentation using the U-Net architecture uses epoch 600, batch size 64 and learning rate 0.0001. The results of testing this model can be seen in figures 6 and 7 that the results obtained are not overfitting, because in this model the graph between training and testing can be said to be balanced and does not have a large distance. For model accuracy, see Table VII.

TABLE VII.
U-NET SEGMENTATION EVALUATION SCORE

Parameter	Score
Pixel Accuracy	97.39
Mean IU	81.93
Mean Accuracy	87.27
FPR	0.91
Precision	85.61
Recall	75.44
F1 Score	79.25

The LinkNet model used in this study shows good accuracy, although not as good as the U-Net model. In the LinkNet model, Pixel Accuracy reaches 96.21%, which indicates that most of the pixels in the reconstructed image are classified correctly.

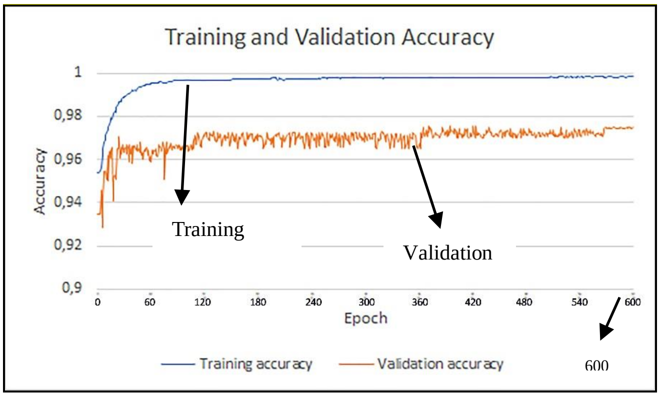


Fig. 6. U-Net Model Training and Validation Accuracy Graph

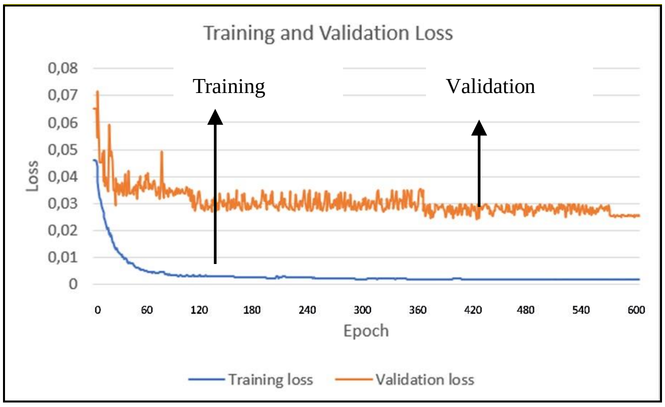


Fig. 7. U-Net Model Training and Validation Loss Graph

Means IU of 78.98% indicates a lower level of comfort compared to the U-Net model. Mean Accuracy reaches 86.90%, indicating the overall proportion of correctly classified pixels. However, the FPR of 1.83% indicates a slightly higher positive misclassification rate compared to the U-Net model. The precision of 77.23% describes the level of accuracy in classifying positive objects, and the recall of 76.01% indicates the ability of the model to find positive objects as a whole. The F1 score of 75.12% reflects the balance between prediction precision and completeness. Although the model with the results of LinkNet's accuracy is not as good as the U-Net model, this model can still be used as an effective alternative in reconstructing 2D images into 3D transcerebellar images.

Model testing for segmentation with the LinkNet architecture uses epoch 600, batch size 64 and learning rate 0.0001. The results of testing this first model can be seen in Figures 8 and 9, that the results obtained are not overfitting but there is accuracy that continues to fall repeatedly which may be caused by an ambiguous dataset. This model may not be good because accuracy errors keep recurring. For model accuracy, see Table VIII.

The Table IX is the result of the accuracy of the 3D-FHNet method on various images with different perspectives. The average accuracy of these images is 76.76%. The results of the accuracy in each image vary, with a range between 57.14% to 88.51%. Higher results can be seen in several images, such as Result_Patient01599_Plane3_1_of_2 with an accuracy of 83.63% and Result_Patient01737_Plane3_1_of_1 with an accuracy of 82.31%. However, there are also images with lower accuracy such as Result_Patient01790_Plane3_1_of_2 with an accuracy of 76.44%. The average accuracy of 76.76% shows a fairly good performance in reconstructing 2D images into 3D transcerebellar using the 3D-FHNet method.

TABLE VIII.
LINKNET SEGMENTATION EVALUATION SCORE

Parameter	Score
Pixel Accuracy	96.21
Mean IU	78.98
Mean Accuracy	86.90
FPR	1.83
Precision	77.23
Recall	76.01
F1 Score	75.12

V. DISCUSSION

This research was conducted using 2D ultrasonography (USG) images on the Transcerebellar section. This research went through several stages to get the final result in the form of a 3-dimensional image. The initial stage is to select the image that will be used for the dataset. After the dataset is ready, the next step is to improve the image by implementing a denoising filter and CLAHE to improve image quality.

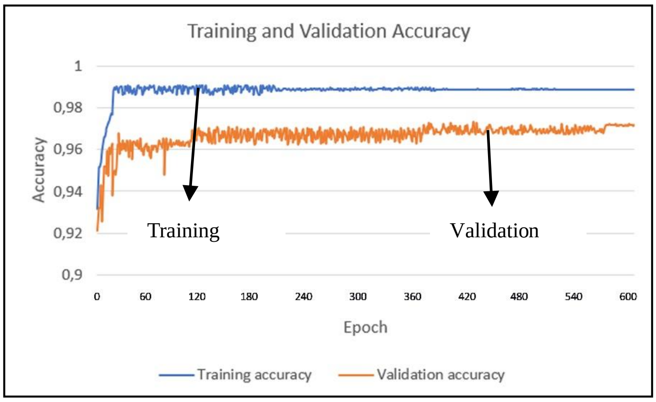


Fig. 8. Graph of Training and Validation Accuracy of the LinkNet Model

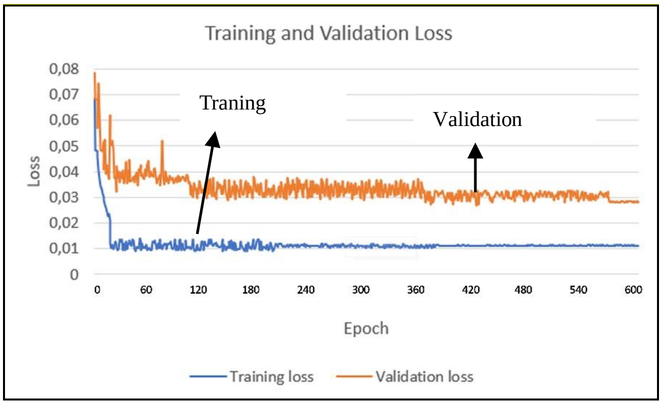


Fig. 9. LinkNet Model Training and Validation Loss Graph

After that is the labeling process on the original image of the object to be segmented, namely transcerebellar. The next process that is carried out is to multiply the data or data augmentation that will be used to increase the segmentation results so that they are more optimal.

TABLE IX.
ACCURATION SCORE (3D-FHNET)

File Name	90	180	270	360	Mean
Result_Patient01565_Plane3_2_of_4	81.58	58.52	86.74	57.14	70.99
Result_Patient01565_Plane3_3_of_4	80.47	61.19	88.18	61.03	72.71
Result_Patient01599_Plane3_1_of_2	86.1	83.25	85.45	79.75	83.63
Result_Patient01601_Plane3_1_of_1	82.36	62.62	88.51	63.45	74.23
Result_Patient01635_Plane3_4_of_4	80.19	66.29	82.86	73.42	75.69
Result_Patient01648_Plane3_1_of_6	84.69	70.68	85.59	76.71	79.41
Result_Patient01653_Plane3_4_of_11	84.58	78.13	84.38	64.75	77.96
Result_Patient01653_Plane3_6_of_11	83.73	83.6	81.36	69.66	79.58
Result_Patient01671_Plane3_3_of_3	85.2	69.54	87.82	75.61	79.54
Result_Patient01672_Plane3_1_of_2	84.03	69.79	79.49	62.24	73.88
Result_Patient01678_Plane3_4_of_8	82.92	63.68	86	67.09	74.92
Result_Patient01692_Plane3_2_of_4	85.35	68.08	81.47	77.84	78.18
Result_Patient01702_Plane3_6_of_8	80.73	64.1	85.28	69.72	74.95
Result_Patient01708_Plane3_1_of_2	83.14	64.92	86.81	66.98	75.46
Result_Patient01717_Plane3_2_of_3	80.96	72.73	80.79	67.27	75.43
Result_Patient01737_Plane3_1_of_1	84.22	73.84	87.17	84.03	82.31
Result_Patient01747_Plane3_1_of_1	80.49	74.14	81.07	68.14	75.96
Result_Patient01779_Plane3_4_of_4	84.36	72.27	81.63	70.11	77.09
Result_Patient01790_Plane3_1_of_2	74.51	77.21	84.44	69.6	76.44
				Mean	76.76

The next step is the segmentation process which is carried out with the image that has been reproduced earlier. This test was carried out with an epoch of 300 and a batch size of 64 with the results shown in table 4.7. The time required to carry out the U-Net model training process is 300 minutes or 3 hours. The results obtained are in the form of pixel accuracy and mean accuracy which are processed based on a comparison of GT (Ground Truth) values or data labels with U-Net Prediction results. The results are quite good because the position of the object is in accordance with the data label. In numbers, accuracy that has exceeded 90% is good accuracy. Mean UI, F1 score and precision are good enough at 80%. However, this result may still be increased to 90%. The FPR value is still quite far from the target which should be close to 0.

After assessing the accuracy of each architectural model previously explained, the best model for segmentation was selected, namely U-Net. A cross validation technique, namely K-Fold, was carried out to ensure that the U-Net model was truly fit, not overfitting or underfitting. The training and testing graphs can be seen in figures 10 and 11. To assess the performance of K-Fold, an accuracy method was used, the results of which can be seen in Table X.

After obtaining the 10-Fold results, it can be said that the model used is fit or not overfitting or underfitting. Next, the model can be tested on unseen data. For the reconstruction method, the model used (3D-FHNet) produces an average of 76.76 for the 19 images tested.

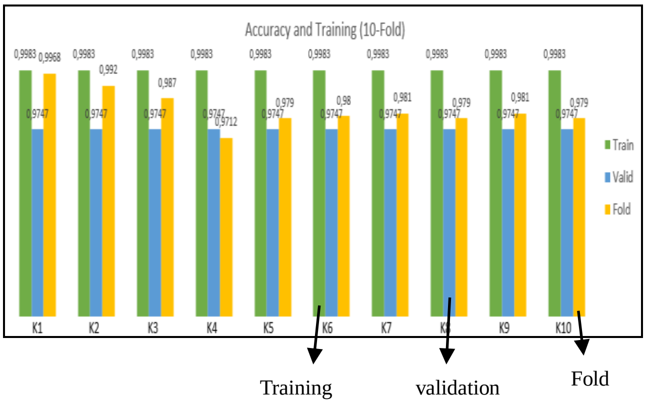


Fig. 10. Training and Validation Accuracy Graph for U-Net Model (10-Fold)

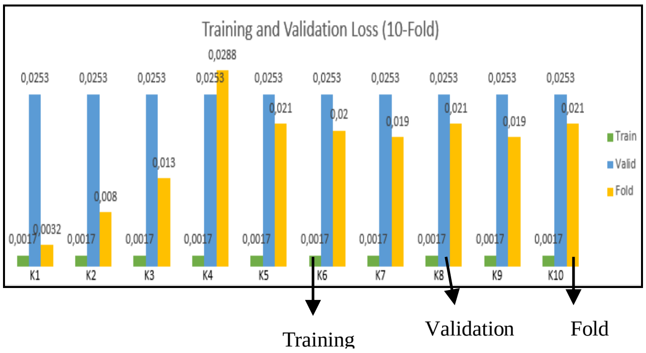


Fig. 11. U-Net Model Training and Validation Loss Graph (10-Fold)

TABLE X.
ACCURACY OF U-NET MODEL (10-FOLD)

K-Fold (Cross Validation)	Pixel Accuracy (%)	Mean IoU (%)
K-1	99.68	89.68
K-2	99.20	88.65
K-3	98.70	86.9
K-4	97.12	89.1
K-5	97.90	87.9
K-6	98.00	86.8
K-7	98.10	87.90
K-8	97.90	85.92
K-9	98.10	87.52
K-10	97.90	86.12

This figure is quite good considering that the method used previously (PIFuHD) did not have an evaluation value which was used as a reference for improving the reconstruction model for this dataset.

VI. CONCLUSIONS

` The dataset used for training consisted of 159 transcerebellar fetal head image data. Model testing is carried out using two architectures, namely U-Net and LinkNet with an epoch of 600 (U-Net), 500 (LinkNet) and a batch size of

64 each. The model training process takes approximately 10 hours with the epoch value and batch size at point 2 for the U-Net architecture and 8 hours for the LinkNet architecture. The accuracy obtained for the U-Net architecture is 97.39% with an F1 score of 79.25% and an accuracy of 90.83% for the LinkNet architecture. The 3D-FHNet method can be used as an alternative to PiFUHD because it succeeds in reconstructing 2-dimensional images into 3 dimensions.

Future work for the next research can be done by increasing datasets for data training activities because this research is still lacking datasets so the results obtained are not optimal. Involve a team of experts in related fields, namely ultrasound of the baby's head so that it can assist in carrying out the labeling process. Subsequent research can use the existing architecture in this research to simplify the labeling process automatically.

REFERENCES

[1] A. Desiani, Erwin, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, "Bi - path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap - smear Images," IAENG International Journal of Computer Science, vol. 48, no. 3, pp. 782-791, 2021.

[2] C. F. Baumgartner *et al.*, "SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound," *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2204–2215, 2017.

[3] X. Deng, S. He, Q. Wu, Z. Weng, M. Yang, and M. Liu, "Prenatal Diagnosis of Fetal Cleft Lip and Palate with Three-Dimensional Ultrasound Information Technology," *Sci. Program.*, vol. 2021, 2021.

[4] M. Han *et al.*, "Automatic Segmentation of Human Placenta Images with U-Net," *IEEE Access*, vol. 7, pp. 180083–180092, 2019.

[5] T. F. Gonzalez, "Handbook of approximation algorithms and metaheuristics," *Handb. Approx. Algorithms Metaheuristics*, pp. 1–1432, 2007.

[6] J. Xiao, J. Wang, S. Cao, and B. Li, "Application of a Novel and Improved VGG-19 Network in the Detection of Workers Wearing Masks," *J. Phys. Conf. Ser.*, vol. 1518, no. 1, 2020.

[7] Erwin, M. N. Rachmatullah, and W. Saputri, "Automatic Detection of Background Diabetic Retinopathy Disease Using Hybrid Multilevel Thresholding and Convolutional Neural Network," *J. Eng. Sci. Technol.*, vol. 14, no. 5, pp. 2522–2539, 2019.

[8] Z. Chen, J. Zhou, Y. Chen, and G. Wang, "3D Texture Ma apping in Multi-view Reconstruction," pp. 359–360, 2012.

[9] T. L. A. van den Heuvel, D. de Bruijn, C. L. de Korte, and B. van Ginneken, "Automated measurement of fetal head circumference using 2D ultrasound images," *PLoS One*, vol. 13, no. 8, pp. 1–20, 2018.

[10] C. Bajaj, I. Ihm, and S. Park, "Compression-based 3D texture mapping for real-time rendering," *Graph. Models*, vol. 62, no. 6, pp. 391–410, 2000.

[11] E. Skelton *et al.*, "Towards automated extraction of 2D standard fetal head planes from 3D ultrasound acquisitions: A clinical evaluation and quality assessment comparison," *Radiography*, vol. 27, no. 2, pp. 519–526, 2021.

[12] A. M. Saleh, F. Al-Saygh, M. Abushama, and B. Ahmed, "The Role of Three-Dimensional Ultrasound in Gynecology," *Res. Women's Heal.*, vol. 1, no. 1, pp. 1–9, 2019.

[13] S. Saito, T. Simon, J. Saragih, and H. Joo, "PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 81–90, 2020.

[14] L. Wu, Y. Xin, S. Li, T. Wang, P. A. Heng, and D. Ni, "Cascaded Fully Convolutional Networks for automatic prenatal ultrasound image segmentation," *Proc. - Int. Symp. Biomed. Imaging*, vol. 0, no. April, pp. 663–666, 2017.

[15] J.-Y. K. and J. K. S. J. Jang, Y. Park, B. Kim, S. M. Lee, "Automatic Estimation of Fetal Abdominal Circumference From Ultrasound Images," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 5, pp. 1512–1520.

[16] Y. Li *et al.*, "Standard plane detection in 3D fetal ultrasound using an iterative transformation network," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11070 LNCS, no. September, pp. 392–400, 2018.

[17] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, pp. 1–28, 2021.

[18] A. Desiani, Erwin, B. Suprihatin, Ermatita, F. R. Husein, and Y. Wahyudi, "A Novelty Patching of Circular Random and Ordered Techniques on Retinal Image to Improve CNN U-Net Performance," Engineering Letters, vol. 30, no. 4, pp. 1217–1229, 2022.

[19] Erwin, A. Noorfizir, M. N. Rachmatullah, Saparudin, and G. Sulong, "Hybrid Multilevel Thresholding-Otsu and Morphology Operation for Retinal Blood Vessel Segmentation," Engineering Letters, vol. 28, no. 1, pp. 180–191, 2020.

[20] Erwin, H. K. Putra, B. Suprihatin, and F. Ramadhini, "A Hybrid CLAHE-GAMMA Adjustment and Densely Connected U-NET for Retinal Blood Vessel Segmentation using Augmentation Data," Engineering Letters, vol. 30, no. 2, pp. 485–493, 2022.

[21] V. Popat, M. Mahdinejad, O. S. Dalmau Cedeño, E. Naredo, and C. Ryan, "GA-based U-Net architecture optimization applied to retina blood vessel segmentation," *IJCCI 2020 - Proc. 12th Int. Jt. Conf. Comput. Intell.*, no. Ijcci, pp. 192–199, 2020.

[22] S. Saito, Z. Huang, R. Natsume, S. Morishima, H. Li, and A. Kanazawa, "PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 2304–2314, 2019.

[23] Q. Lu, Y. Lu, M. Xiao, X. Yuan, and W. Jia, "3d-fhnet: Three-dimensional fusion hierarchical reconstruction method for any number of views," *IEEE Access*, vol. 7, pp. 172902–172912, 2019.

[24] N. Baños *et al.*, "Quantitative analysis of cervical texture by ultrasound in mid-pregnancy and association with spontaneous preterm birth," *Ultrasound Obstet. Gynecol.*, vol. 51, no. 5, pp. 637–643, 2018.

[25] P. Meenakshi, S., Suganthi, M. & Sureshkumar, "Segmentation and Boundary Detection of Fetal Kidney Images in Second and Third Trimesters Using Kernel-Based Fuzzy Clustering," *J Med Syst*, vol. 43, no. 203, 2019.

[26] M. Ebner *et al.*, "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI," *Neuroimage*, vol. 206, no. October 2019, 2020.

[27] S. Tourbier *et al.*, "Automated template-based brain localization and extraction for fetal brain MRI reconstruction," *Neuroimage*, vol. 155, no. April, pp. 460–472, 2017.

Abarham Martadiansyah was born in Baturaja, Indonesia, in 1982. He received his medic of Medical Faculty from Sriwijaya University, Indonesia, in 2005, and an obgin specialist degree in Medical Faculty from Sriwijaya University, Indonesia, in 2012. In 2015, he joined Sriwijaya University, as a Lecturer. His current research interests is Urogynecology.

Hadrians Kesuma Putra was born in Baturaja, Indonesia, in 1977. He received his medic of Medical Faculty from Sriwijaya University, Indonesia, in 2002, and an obgin specialist degree in Medical Faculty from Sriwijaya University, Indonesia, in 2009. In 2005, he joined Sriwijaya University, as a Lecturer. His current research interests is Urogynecology.

Muhammad Rizky Ramadhan, was born in Palembang, Indonesia, in 2021. Since 2019 he was accepted as a student college and became a part of the Department of Computer Engineering Sriwijaya University, Indonesia. His currenct research interest is image processing.

Ermatita was born in Lahat, September 1967. She received mathematics bachelor from Universitas Lampung in 1991, and magister degree in Computer Science from Universitas Indonesia in 1999 and Doktor Degree in Computer Science from Universitas Gadjah Mada in 2013.She currently working on Universitas Sriwijaya start since 2006. She is an Associate Professor in 2013. His research are artificial intelligent, data mining. Machine learning and Information System.

Abdiansah was born in Pagar Alam, Indonesia, in 1984. He received his Bachelor of Informatic Engineering from IST Akprind, Yogyakarta, Indonesia, in 2006. He also received his Master and Doctoral degree in Computer Science at Universitas Gadjah Mada in 2008 and 2019 respectively. In 2008 he joined as lecturer at Department of Informatics, Faculty of Computer Science, Universitas Sriwijaya. His current research interests include Natural Language Processing, Quesstion Answering System, Chatbot System, Text Mining, Big Data, and Computer Vision. He has authored or co-authored some publicaions. He can be contacted at email: abdiansah@unsri.ac.id

.

Erwin was born in Palembang, Indonesia, in 1971. He received his Bachelor of Mathematics from Universitas Sriwijaya, Indonesia, in 1994, and an M.Sc. degree in Actuarial from the Bandung Institute of Technology (ITB), Bandung, Indonesia, in 2002. In 1994, he joined Universitas Sriwijaya, as a Lecturer. Since December 2006, he has been with the Department of Informatics, Universitas Sriwijaya, where he is an Professor in 2023. Since 2012, he has been with the Department of Computer Engineering, Universitas Sriwijaya. Then, in 2019, he received his Doctorate in Engineering, Faculty of Engineering, Universitas Sriwijaya. His current research interests include image processing, and computer vision. Prof. Dr. Erwin, S.Si., M.Si is a member of IAENG and IEEE.