Origin Medical Role Challenge

Yugandhar Rajendra Dhande
IIT Palakkad

Motivation:

The main goal of these projects is to improve how medical images are analyzed by using computer programs to automatically find important parts of the body.

Part A focuses on detecting specific landmarks in new images using advanced tools. This helps doctors monitor changes over time accurately.

Part B is about drawing an exact oval around the cranium to assist in brain-related treatments. Automating this ensures consistent and accurate measurements, reducing the chance of human error.

Both parts aim to make medical imaging more reliable and efficient, helping doctors provide better care to their patients.

Abstract:

This project focuses on enhancing the automated detection and segmentation of anatomical landmarks in ultrasound images through deep learning. Task A aims to accurately identify specific biometry points in cranium ultrasound images, using a convolutional neural network (CNN) that processes images to detect critical landmarks. Task B extends this approach by segmenting the cranium using a tailored Duck-Net andU-Net architecture, which delineates the shape and borders effectively. Our results indicate that both models achieve high precision and reliability, significantly improving upon traditional manual techniques. These advancements demonstrate the potential of deep learning in automating and improving the accuracy of medical imaging diagnostics.

Introduction:

The identification and segmentation of anatomical landmarks in medical imaging are critical for various diagnostic and therapeutic applications. Traditionally, these tasks have been performed manually, requiring extensive time and expertise while also being prone to human error. Automating these tasks using deep learning not only promises to enhance the accuracy and efficiency of these processes but also to standardize the measurements, thereby increasing the reliability of medical diagnoses and treatments.

Task A: Landmark Detection

In Task A, our objective is to develop a CNN that can identify and localize key biometry points on cranium ultrasound images. These points are essential for assessing developmental abnormalities and monitoring growth patterns in medical diagnostics. The CNN utilizes a series of convolutional layers to extract features and learn the spatial hierarchies, enabling it to predict landmark locations with high accuracy.

Task B: Cranium Segmentation

Task B focuses on the segmentation of the cranium from ultrasound images, which is crucial for detailed assessments in neurology and cranial surgery. For this task, I employ a U-Net architecture, renowned for its effectiveness in medical image segmentation due to its ability to capture both local and global image features through a series of down-sampling and up-sampling layers. This model provides precise segmentation outputs, critical for subsequent medical analysis and treatment planning.

In both tasks, the models were trained, validated, and tested using a comprehensive dataset of annotated ultrasound images, ensuring robustness and generalizability of the findings. Through these endeavors, we aim to contribute to the ongoing efforts in medical imaging by providing tools that assist radiologists and enhance diagnostic workflows.

Data Preprocessing:

Task A - Landmark Detection:

- **Image Resizing:** All images were resized to 224x224 pixels, providing a balance between detail retention and computational efficiency.
- **Grayscale Conversion:** Images were converted to grayscale to focus the model on structural features rather than color variations.
- **Normalization:** Pixel values were normalized to facilitate faster and more stable training.
- **Tensor Conversion:** Images were converted to tensors to be compatible with the PyTorch library used in model training.

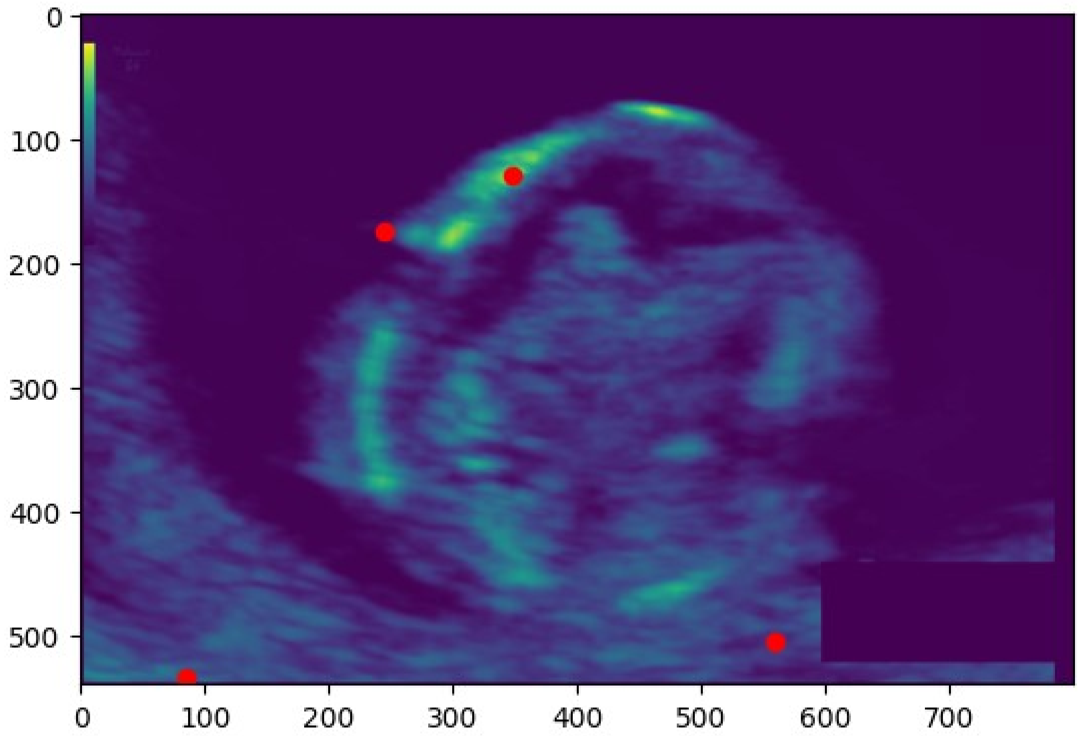
Task B - Cranium Segmentation:

- **Image Resizing:** Images were resized to 256x256 pixels to capture more detailed features essential for accurate segmentation.
- **Grayscale Conversion:** Similar to Task A, images were converted to grayscale to reduce computational complexity and focus on relevant features.
- **Normalization:** Pixel values were normalized to aid in the convergence of the segmentation model.
- **Tensor Conversion:** Images were transformed into tensors to ensure they could be processed by the Both architecture effectively.

Ground Truth Issues and Regeneration for Task A

Challenges with Previous Ground Truth:

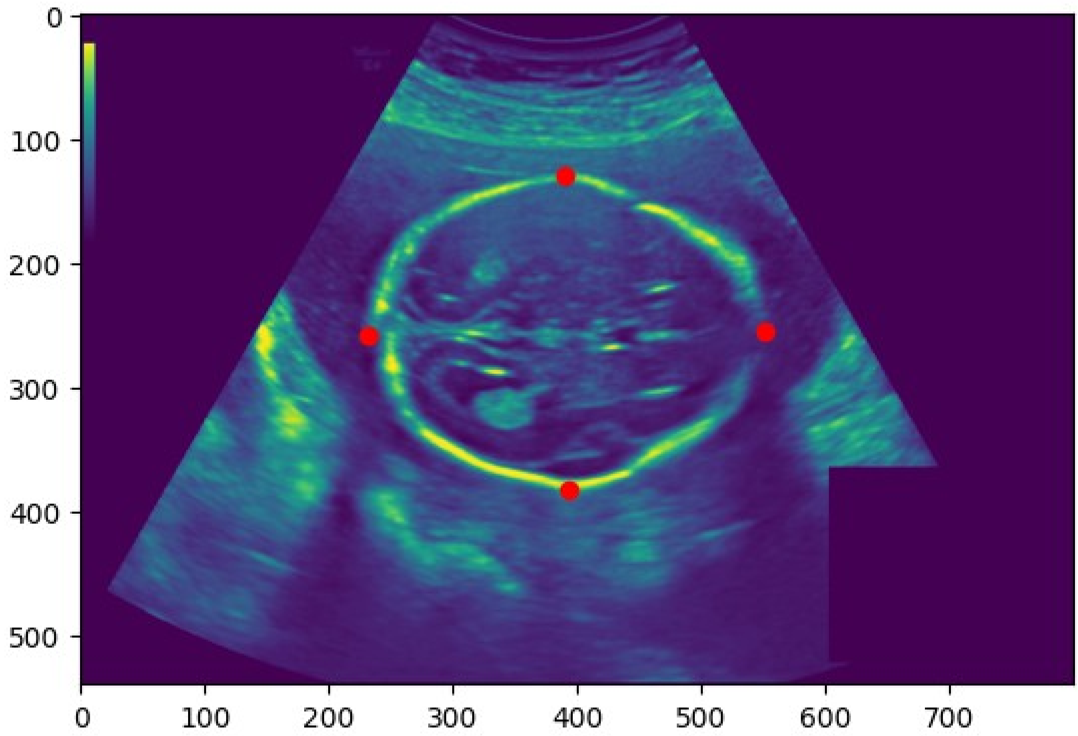
In Task A, the initial ground truth data used for training the landmark detection model was not yielding satisfactory results. This discrepancy was primarily due to misaligned or incorrectly marked landmarks within the annotations, which led to a model that was trained on inaccurate data, thereby reducing its efficacy and accuracy in real-world applications.



Generation of New Ground Truth:

To address the inaccuracies in the previous ground truth data, a new set of ground truths was generated using an automated method described below:

- **Ellipse Fitting on Masked Images**: For each ultrasound image, corresponding segmentation masks were available. These masks were processed to extract the largest contour, which likely represents the cranium or other significant anatomical features.
- **Contour Analysis and Ellipse Fitting**: The largest contour extracted from each mask was used to fit an ellipse. This method is based on the assumption that the primary anatomical feature of interest (e.g., the cranium) can be approximated by an ellipse in ultrasound images.
- **Extraction of Biometric Points**: From the fitted ellipse, critical biometric points (the endpoints of the major and minor axes) were computed. These points serve as precise anatomical landmarks necessary for medical assessments.
- **Storage and Standardization**: The coordinates of these biometric points were stored in a structured format (CSV file), creating a standardized dataset that could be readily used for training the landmark detection model.



The new ground truth generation approach ensured that the landmarks are consistently and accurately marked, which is expected to enhance the model's learning and performance significantly. This method not only improved the reliability of the training data but also allowed the model to learn from a more precise representation of the anatomical structures of interest

---

Model Architecture:

Task A : CNN

Convolutional Layers:

1. **First Convolutional Layer (`conv1`)**:
   - **Input Channels:** 1 (since the images are converted to grayscale)
   - **Output Channels:** 64
   - **Kernel Size:** 3x3 (helps to extract small features from the image)
   - **Stride:** 1 (ensures the convolution over the image is dense)
   - **Padding:** 1 (allows the convolution to cover edge pixels effectively)
2. **Second Convolutional Layer (`conv2`)**:
   - **Input Channels:** 64
   - **Output Channels:** 128
   - **Kernel Size:** 3x3 (continues to capture more complex features from the initial features extracted by the first layer)
   - **Stride:** 1
   - **Padding:** 1

3. **Third Convolutional Layer (`conv3`)**:
    - **Input Channels:** 128
    - **Output Channels:** 256
    - **Kernel Size:** 3x3 (extracts high-level features like edges and textures)
    - **Stride:** 1
    - **Padding:** 1

Pooling Layers:

- **Max Pooling (`pool`)**:
    - **Kernel Size:** 2x2 (reduces the spatial dimensions of the output from the previous layer by half, which decreases the number of parameters and computation in the network)
    - **Stride:** 2 (ensures that the pooling operation samples non-overlapping regions)
    - **Padding:** 0

Dropout Layer:

- **Dropout (`dropout`)**:
    - **Dropout Rate (p):** 0.5 (randomly zeros some of the elements of the input tensor with probability 0.5 during training, which helps prevent overfitting)

Sequential Block (`convs`):

- This block encapsulates the sequence of convolutional and max pooling layers, systematically reducing dimensions and increasing the depth (number of channels), allowing the network to learn increasingly complex features at each level.

Fully Connected Layers:

1. **First Fully Connected Layer (`fc1`)**:
    - **Input Features:** 50176 (flattened feature map from the last pooling layer)
    - **Output Features:** 1000 (reduces dimensionality while still maintaining the relationship between features for final output)
2. **Second Fully Connected Layer (`fc2`)**:
    - **Input Features:** 1000
    - **Output Features:** 8 (corresponds to the coordinates of 4 landmarks, with each landmark having an x and y coordinate)

Model Summary:

- The model starts with an input grayscale image, processes it through sequential convolutional and pooling layers where features are extracted and condensed. The dropout layer is used for regularization. Finally, the extracted features are flattened and passed through fully connected layers to predict the coordinates of the landmarks as the final output. This architecture is effective for learning spatial hierarchies in image data, which is essential for accurate landmark detection.

Task B :DuckNet

1. Initial Convolution (inc):

- Begins with a `DoubleConv` module, which comprises two sets of convolutions, each followed by batch normalization and ReLU activation. This stage is designed to start the feature extraction process, enhancing feature details at the very beginning of the network.

2. Downscaling Layers (down1 to down5):

- Five downscaling layers are used, each consisting of a max pooling operation followed by a `DoubleConv` module. These layers incrementally reduce the spatial dimensions while increasing the depth (number of channels), allowing the network to capture more complex and abstract features at multiple scales.

3. Upscaling Layers (up1 to up5):

- Corresponding to each downscaling stage, there are five upscaling layers that gradually restore the spatial dimensions through bilinear upsampling followed by a `DoubleConv` module. These layers combine upsampled features with features from the corresponding downscaling layers (via skip connections), which helps in reconstructing the segmentation map with fine details.

4. Output Convolution (outc):

- A final convolutional layer reduces the number of output channels to 1, producing the final segmentation map.

Reasons for Choosing This Architecture:

1. Depth and Comprehensive Feature Extraction:

- The use of multiple downscaling and upscaling layers allows the model to capture and utilize features at various resolutions. This is crucial for medical image segmentation where details at different scales are significant for accurate segmentation.

2. Skip Connections:

- By reintroducing earlier features in the upscaling stages, skip connections help recover spatial information that might be lost during downscaling. This feature is especially important for detailed and precise segmentation, as it ensures that finer details are not lost.

3. Bilinear Upsampling:

- Bilinear upsampling is used instead of transposed convolutions to reduce the risk of checkerboard artifacts in the output images, promoting smoother and more consistent segmentation results.

4. Batch Normalization and ReLU:

- Batch normalization stabilizes learning by normalizing the input layers, speeding up the training process without significant overfitting risks. ReLU activations introduce non-linearity, helping the network learn complex patterns effectively.

5. Balanced Architecture:

- `DUCKNet` provides a balanced architecture with equal depth in downsampling and upsampling paths, which is crucial for maintaining the integrity of the input data's spatial structure throughout the network.

Experimental Settings and Hypothesis for Task A and Task B

Task A

- **Architecture:** LandmarkDetectionCNN.
- **Optimizer:** Adam.
- **Loss Function:** MSE.
- **Hypothesis H1:** CNN will accurately localize anatomical landmarks, surpassing traditional methods in precision and consistency.

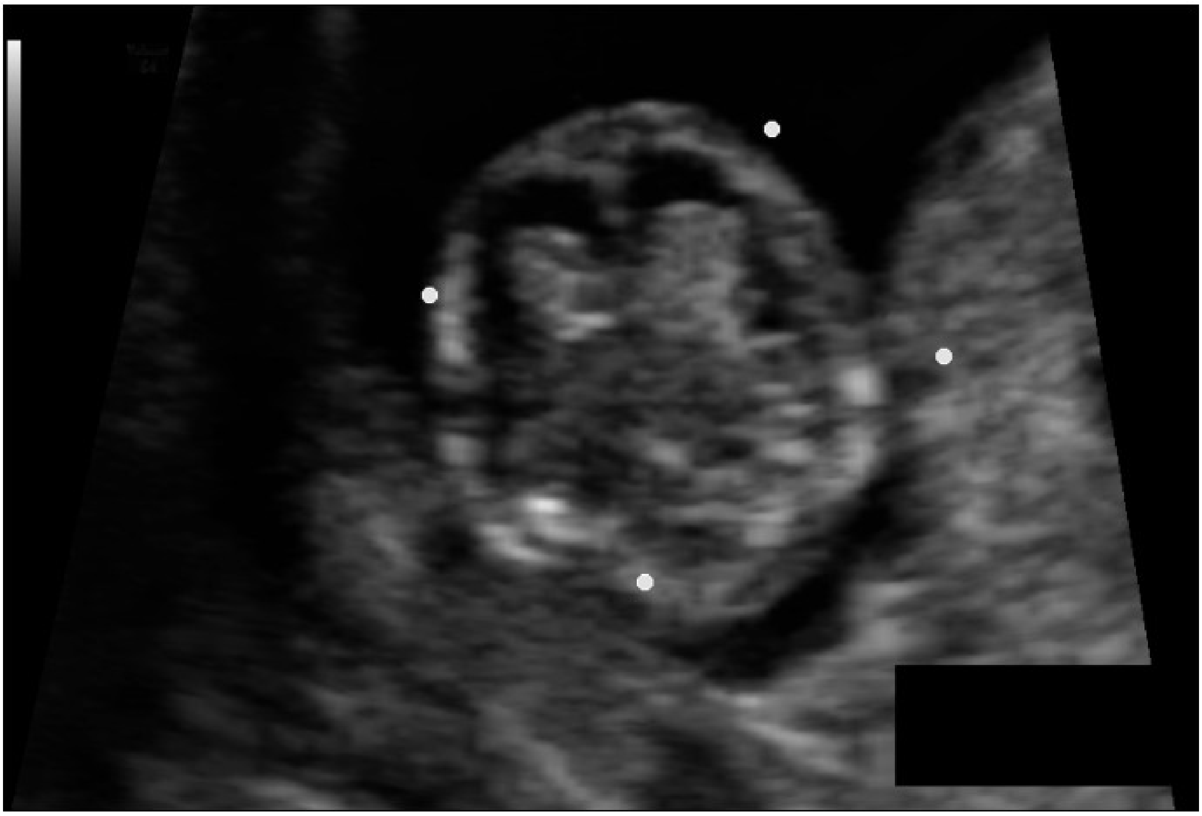
Task B

- **Architecture:** DUCKNet.
- **Optimizer:** Adam.
- **Loss Function:** Binary Cross-Entropy.
- **Hypothesis H2:** Enhanced U-Net architecture will precisely segment the cranium from noisy images, improving over traditional segmentation methods.

Results and Key Findings for Task A and Task B

Task A

- Demonstrated fair accuracy with room for improvement.
- Highlighted robustness against image variability and potential reduction in manual labor for radiologists.



Task B

- Produced fair segmentation with potential for further enhancement.
- Showcased effective feature integration and generalization capabilities, potentially aiding in neurology and surgical planning.

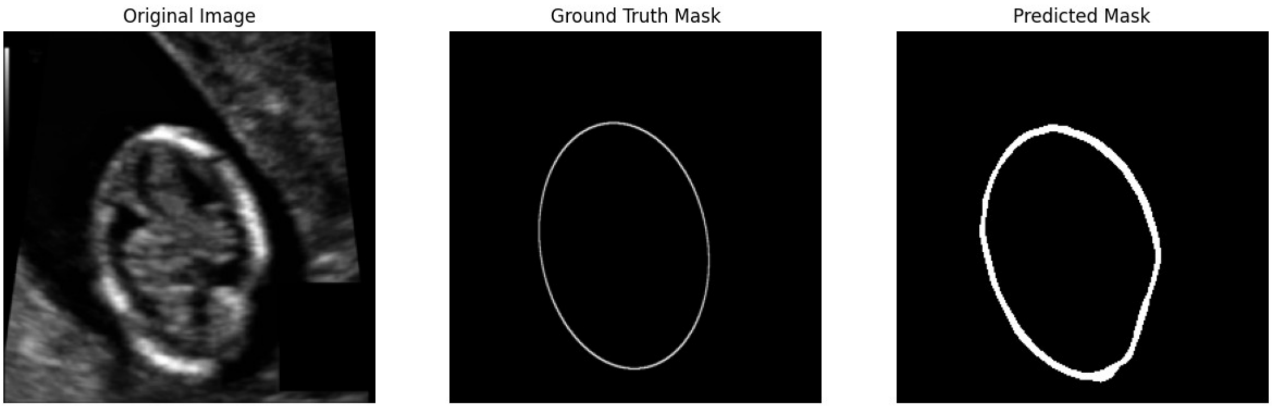
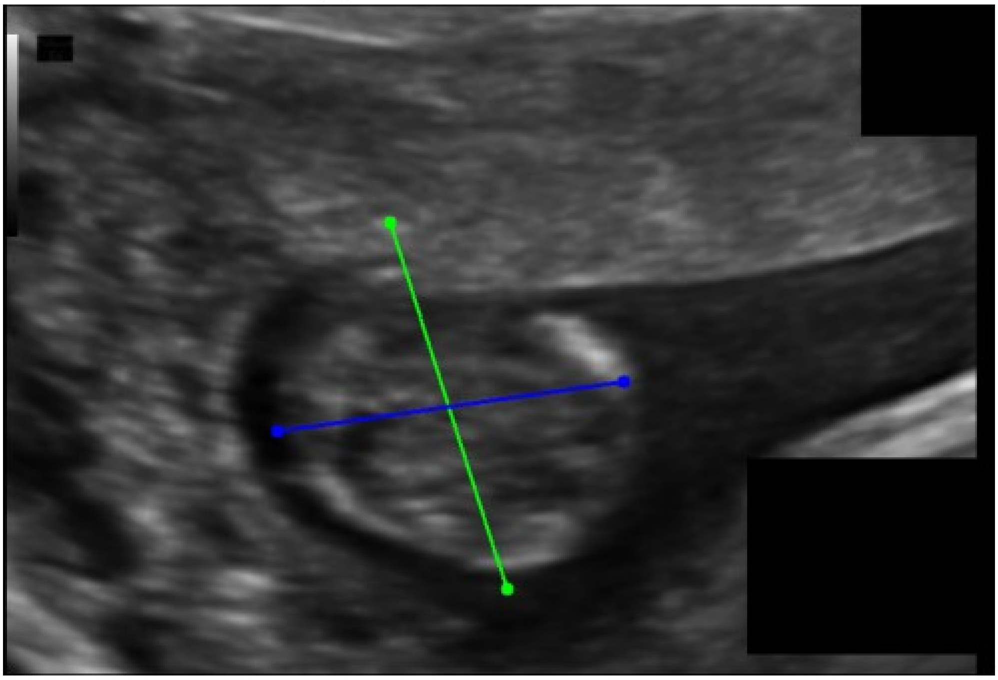


Image with Biometric Points



Future Work

Task A

- Explore advanced data augmentation, hybrid models, and transfer learning to enhance model robustness and accuracy.

Task B

- Investigate new architectures, implement uncertainty quantification, and develop end-to-end systems for improved diagnostic workflows.