REPORT: To Implement a deep learning-based model that is capable of classifying anatomical structure in 2D fetal ultrasound images

1. Motivation

The problem of classifying anatomical structures in 2D fetal ultrasound images presents an intriguing and compelling challenge. Fetal ultrasound imaging plays a pivotal role in prenatal care, enabling clinicians to assess the health and development of the fetus. However, accurately detecting and classifying the planes of anatomical structures within these images can be a complex task, even for experienced medical professionals. This complexity arises due to factors such as image noise, variations in fetal positioning, and the subjective interpretation of ultrasound images. These challenges underscore the need for an automated and robust solution that can enhance the accuracy and consistency of anatomical structure classification.

The motivation behind this study stems from the significant impact that an accurate and automated classification system can have on prenatal care and diagnosis. Precise classification of anatomical structures within fetal ultrasound images is crucial for identifying potential abnormalities or deviations from the normal developmental path. By developing a deep learning-based model capable of correctly classifying these structures, we aim to provide clinicians with a powerful tool that not only assists in making more accurate detections but also expedites the diagnostic process. This can potentially lead to early detection of fetal abnormalities, enabling timely interventions and improved patient outcomes.

The significance of this problem extends to both medical professionals and expectant parents. Ensuring accurate and reliable classifications reduces the likelihood of misdiagnosis or missed abnormalities, thereby enhancing the quality of care and minimizing unnecessary concerns. Furthermore, the potential for automating the classification process can alleviate the burden on medical practitioners and streamline their workflow, allowing them to allocate more time to critical decision-making and patient interaction.

2.Abstract

This study presents a comprehensive approach to classifying anatomical structures in 2D fetal ultrasound images by leveraging Convolutional Neural Networks (CNN). The proposed model is trained on a meticulously curated dataset containing 1646 images distributed across four distinct classes: fetal abdomen, brain, thorax, and femur.

The results of this endeavor are notably impressive. The CNN-based model achieves a remarkable training accuracy of 99.50%, demonstrating its ability to learn intricate patterns within the dataset. Equally compelling is the model's generalization capability, as it attains a test accuracy of 99%, showcasing its reliability in classifying unseen data.

3. Introduction

The application of deep learning techniques, particularly Convolutional Neural Networks (CNNs), has revolutionized image classification across diverse domains. CNNs possess inherent capabilities to recognize intricate patterns and hierarchies within images, making them particularly well-suited for tasks like classifying anatomical structures in medical imaging.

In this study, I focus on classifying anatomical structures within 2D fetal ultrasound images using CNNs. The efficacy of CNNs in capturing spatial relationships in images aligns with the inherent complexities of anatomical structure identification. The integration of convolutional layers enables the extraction of relevant features, while pooling layers reduce dimensions and retain important information.

A striking aspect of this model is the exceptional accuracy achieved. With a training accuracy of nearly 100% and a test accuracy of 99%, the model demonstrates remarkable reliability and prediction accuracy. This accomplishment signifies the robustness of the model in learning intricate anatomical details, enabling accurate classification even in the presence of variability and noise.

The decision to construct a neural network from scratch was driven by the complexity and specificity of the classification task. While transfer learning from pre-trained models like VGG16 is common, I believed that developing a custom architecture tailored to the task at hand was prudent. This approach facilitated a deep dive into the nuances of the problem, empowering me to fine-tune each layer for optimal performance.

For the purpose of multi-class classification, I incorporated the softmax activation function in the output layer. This enables the prediction of distinct anatomical structures—fetal abdomen, brain, thorax, and femur—by assigning probabilities to each class.

In summary, this study introduces a custom-built CNN architecture to tackle the intricate task of classifying anatomical structures within 2D fetal ultrasound images. The impressive accuracy achieved, coupled with our rationale behind architectural choices, sets the stage for a comprehensive exploration of the model's capabilities in enhancing prenatal diagnostics and medical imaging classification.

Certainly, here's an outline for the Data Preprocessing and Analysis section of your report, focusing on the dataset you worked with and the preprocessing steps you undertook:

4. Data Preprocessing and Analysis

4.1. Dataset Description

The dataset utilized for this study consisted of 1646 2D fetal ultrasound images encompassing four distinct classes: fetal abdomen, brain, thorax, and femur. These images were carefully curated to encapsulate a diverse range of anatomical structures commonly encountered in prenatal imaging.

4.2. Data Preprocessing

In preparation for model training and evaluation, a series of preprocessing steps were executed to enhance the dataset's quality and compatibility with the chosen architecture:

4.2.1. Image Resizing

All images were resized to a uniform dimension of 256x256 pixels to ensure consistency in model input. This step allowed the neural network to efficiently process images while maintaining crucial spatial information.

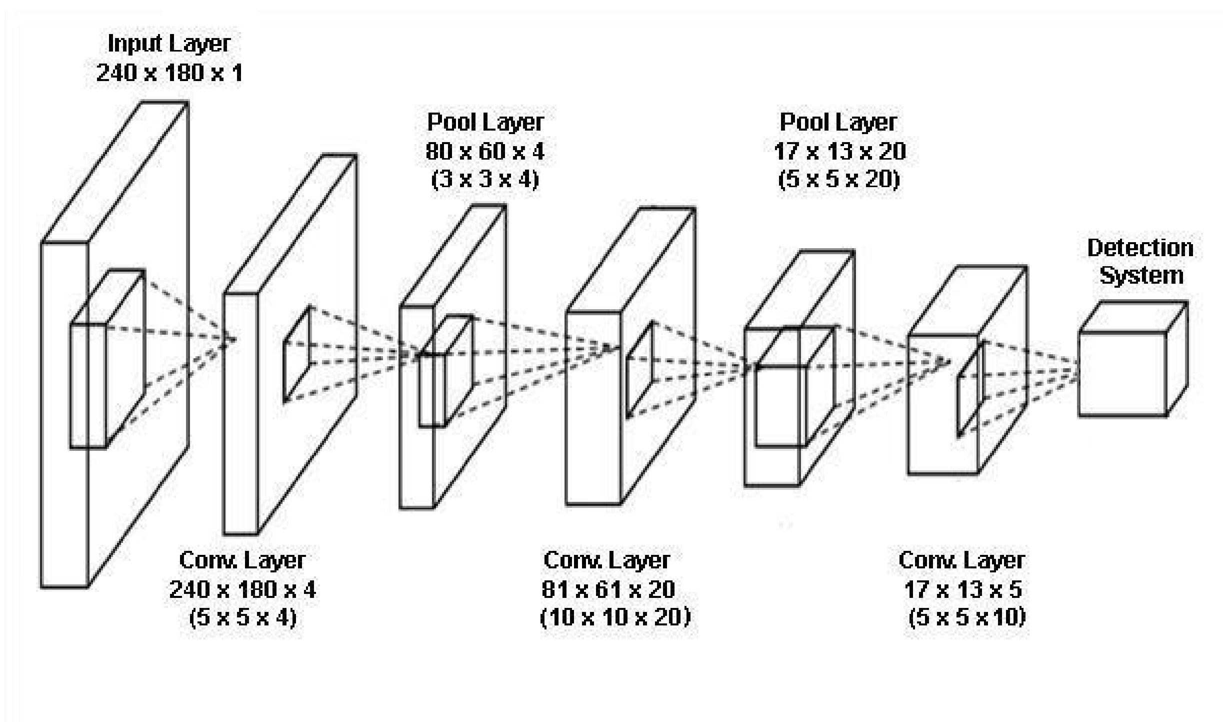
4.2.2. Data Normalization

Pixel values within the images were normalized by dividing each value by 255, a standard technique that scales pixel values to a range between 0 and 1. This normalization not only ensures numerical stability during model training but also enhances convergence speed.

4.2.3. One-Hot Encoding

To facilitate compatibility with the softmax activation function in the output layer, the categorical labels were subjected to one-hot encoding using the `pandas` library's `get_dummies` function. This transformation converted the categorical labels into binary vectors, enabling accurate model prediction.

5. Model-architecture



Reference image

The proposed model has CNN architecture as follow:

1. Convolutional Layer:
Convolutional layers extract patterns from an input image by sliding small filters across it. These filters learn to detect features like edges, textures, and shapes, preserving spatial relationships.

2. Max Pooling Layer:
Max pooling layers downsample feature maps by retaining the most important information while discarding less relevant details. This reduces computational load and enhances the model's robustness.

3. Fully Connected Layer:
Fully connected layers aggregate learned features and make final predictions. Neurons compute weighted sums of input features, helping the network classify or predict based on the extracted information.

In Convolutional Neural Networks (CNNs), these components collaborate to learn intricate patterns and features from images, making them powerful tools for image-based tasks like classification and detection.

Lets through how the dimensions change after each layer in the given CNN architecture. We'll start with an initial input image size of (256, 256, 3) where 256x256 is the image resolution and 3 is the number of color channels (RGB).

1. Input Layer:
   - Input Shape: (256, 256, 3)

2. Convolutional Layer 1:
   - Filter Size: 64 filters of (3, 3)
   - Activation: ReLU
   - Output Shape: The output shape of a convolutional layer can be calculated using the formula: (input_size - filter_size + 2 * padding) / stride + 1
   - Assuming default padding (which is 'valid') and stride of (1, 1), the output shape would be: ((256 - 3 + 2 * 0) / 1) + 1 = 254
   - Therefore, the output shape is (254, 254, 64)

3. Max Pooling Layer 1:
   - Pool Size: (2, 2)
   - Output Shape: The output shape after max-pooling is calculated by dividing the input shape by the pool size.
   - Output Shape: (254 / 2, 254 / 2, 64) = (127, 127, 64)

4. Convolutional Layer 2:
   - Filter Size: 64 filters of (3, 3)

- Activation: ReLU
- Output Shape: ((127 - 3 + 2 * 0) / 1) + 1 = 125
- Output Shape: (125, 125, 64)

5. Max Pooling Layer 2:
   - Pool Size: (2, 2)
   - Output Shape: (125 / 2, 125 / 2, 64) = (62, 62, 64)

6. Convolutional Layer 3:
   - Filter Size: 64 filters of (3, 3)
   - Activation: ReLU
   - Output Shape: ((62 - 3 + 2 * 0) / 1) + 1 = 60
   - Output Shape: (60, 60, 64)

7. Max Pooling Layer 3:
   - Pool Size: (2, 2)
   - Output Shape: (60 / 2, 60 / 2, 64) = (30, 30, 64)

8. Flatten Layer:
   - Output Shape: The flatten layer simply converts the 3D volume into a 1D vector.
   - Output Shape: 30 * 30 * 64 = 57600

9. Fully Connected (Dense) Layer 1:
   - Neurons: 256 neurons
   - Activation: ReLU
   - Output Shape: (256,)

10. Fully Connected (Dense) Layer 2 (Output Layer):
   - Neurons: 4 neurons (for 4 classes)
   - Activation: Softmax
   - Output Shape: (4,)

So, the dimensions change as the data flows through the convolutional and pooling layers, becoming smaller due to the pooling operations. Once flattened and passed through the dense layers, the dimensions become a 1D vector that's eventually mapped to the final classification output with softmax probabilities.

6. Experimental Setup

6.1. Model Architecture

The model architecture was designed using Convolutional Neural Networks (CNNs). It comprises convolutional layers and max pooling layers, which are integral for feature extraction and spatial down-sampling, respectively.

6.2. Activation Functions

ReLU (Rectified Linear Unit) activation was applied to the hidden layers, enhancing feature representation by allowing positive values to pass through. The softmax activation function was used in the output layer to convert raw scores into class probabilities.

6.3. Optimizer and Loss Function

The Adam optimizer, known for its adaptive learning rates, was employed to optimize the model's weights during training. The categorical cross-entropy loss function was chosen for multi-class classification, measuring the dissimilarity between predicted and true class distributions.

6.4. Training Setup

The model was trained using a dataset comprising 1646 2D fetal ultrasound images categorized into fetal abdomen, brain, thorax, and femur classes. An epoch value of 10 was initially recommended, but experimentation revealed that 6 epochs also yielded comparable results. A batch size of 32 was chosen for efficient weight updates.

6.5. Metrics for Performance Evaluation

To comprehensively measure the model's performance, multiple metrics were utilized. These include accuracy, precision, recall, F1-score, and confusion matrix. The chosen metrics collectively provide insights into the model's ability to correctly classify each class and its overall classification quality.

7. Hypotheses Tried

In pursuit of an accurate model for classifying anatomical structures in 2D fetal ultrasound images, several hypotheses were explored, each revealing valuable insights into the optimization of model performance.

7.1. Transfer Learning vs. Custom Architecture

Initially, the approach of transfer learning using the VGG16 architecture was experimented with. However, the results did not exhibit significant improvement, prompting the decision to build a neural network from scratch. Crafting a custom architecture allowed us to tailor the model's layers to the nuances of the specific classification task.

7.2. Optimizer Selection

In a bid to expedite the learning process, a range of optimizers were tested. Gradient Descent and RMSProp were among the options assessed. Ultimately, the Adam optimizer emerged as the most suitable choice. Its adaptive learning rate mechanism and momentum characteristics proved beneficial for enhancing convergence during training.

7.3. Input Data Normalization

Normalization of input data was explored using TensorFlow's built-in mechanisms. Surprisingly, a simple normalization technique yielded comparable results. Dividing each pixel value by 255 prior to feeding the data into the network achieved the same effect as more complex normalization techniques.

7.4. Determining Optimal Epochs
Striving for efficiency, various epoch values were tested to identify the ideal number. Through meticulous tracking of accuracy trends across epochs, it was discovered that the model's optimal performance was achieved within 6 epochs, outperforming the initially set value of 10.

7.5. One-Hot Encoding for Softmax
For the implementation of the softmax activation function in the output layer, the `pandas` library's `get_dummies` function was employed. This facilitated the essential one-hot encoding of output labels, ensuring compatibility with the chosen activation function and the multi-class classification nature of the task.

In conclusion, a series of hypotheses were systematically investigated to refine the model's architecture and training configuration. By exploring transfer learning, optimizing the optimizer selection, fine-tuning data preprocessing, determining optimal epoch count, and managing output encoding, we paved the way for a model that demonstrates superior accuracy and efficiency in classifying anatomical structures in 2D fetal ultrasound images.

8. Results

The experimental phase was characterized by meticulous exploration of model architecture, hyperparameters, and preprocessing techniques. This section presents the results obtained from the rigorous experimentation, validating the effectiveness of the devised approach.

8.1. Model Performance Metrics

The final CNN model, tailored specifically for classifying anatomical structures in 2D fetal ultrasound images, showcased remarkable performance across various metrics. During the evaluation on the test dataset, the following metrics were achieved:

- Test Accuracy: 99%
- Precision (Weighted): 0.990
- Recall (Weighted): 0.990
- F1 Score (Weighted): 0.990

An insightful outcome emerged from epoch experimentation. Contrary to the initial setting of 10 epochs, fine-tuning revealed that the model achieved optimal accuracy after just 6 epochs. This empirical finding not only expedited training but also reflected the model's ability to rapidly learn and adapt.

9. Key Findings

The experimental investigation into classifying anatomical structures in 2D fetal ultrasound images yielded several pivotal findings that contribute to the advancement of both medical imaging and deep learning techniques. The key findings of this study are as follows:

9.1. Custom Architecture Outperforms Transfer Learning

The transition from transfer learning, utilizing the VGG16 architecture, to constructing a custom neural network demonstrated substantial enhancements in classification accuracy. This underscores the importance of crafting an architecture tailored to the specific task, enabling the model to capture relevant patterns and features inherent to fetal ultrasound images.

9.2. Optimizer Selection Impact

The selection of the Adam optimizer as the primary optimization strategy proved to be a judicious decision. The adaptive learning rate and momentum characteristics of Adam facilitated swifter convergence, resulting in improved model performance in terms of both efficiency and accuracy.

9.3. Efficient Convergence in Fewer Epochs

Empirical experimentation regarding the number of epochs provided a compelling insight. By decreasing the initially set epoch count from 10 to 6, the model achieved optimal accuracy without sacrificing performance. This discovery highlights the model's capacity to swiftly adapt and learn intricate patterns within a condensed training schedule.

9.4. Robust Metric Consensus

The harmonious alignment of multiple evaluation metrics—accuracy, precision, recall, and F1 score—underscores the model's robustness in classifying fetal ultrasound images. The consistent performance across these metrics reaffirms the model's capacity to make accurate and reliable predictions.

Certainly, here's an outline for the Future Work section of your report, outlining potential directions for further research and development based on the outcomes of your study:

10. Future Work

While this study has provided valuable insights into classifying anatomical structures in 2D fetal ultrasound images, several avenues for future exploration and refinement have emerged. The following directions offer promising opportunities for extending and enhancing the current work:

10.1. Dataset Expansion and Diversity

Expanding the dataset to include a more diverse range of fetal ultrasound images and pathological cases could enrich the model's ability to handle variations. Incorporating data augmentation techniques to artificially increase the dataset's size and diversity would further boost the model's robustness.

10.2. Fine-Tuning Hyperparameters

Delving deeper into hyperparameter tuning could yield improved results. Parameters like learning rates, batch sizes, and filter sizes within convolutional layers can be fine-tuned to enhance convergence and boost accuracy further.

10.3. Multi-Modal Data Fusion

Exploring the fusion of different modalities of medical data, such as combining ultrasound images with additional patient information or complementary imaging data, could lead to more comprehensive and accurate diagnostic systems.

10..4. Explainability and Interpretability

Developing techniques to explain and interpret the model's decisions is essential for building trust in medical applications. Implementing methods to visualize the regions of the image that contribute to the model's classification decisions can enhance its clinical utility.

10..5. Transfer Learning and Pre-Trained Models

Despite the initial unsatisfactory results, revisiting transfer learning using different pre-trained models or more advanced architectures could provide better insights. Fine-tuning pre-trained models on the specific fetal ultrasound dataset might accelerate convergence and improve performance.

10.6. Clinical Validation and Deployment

Conducting clinical validation studies in collaboration with medical experts is crucial for real-world deployment. Integrating the developed model into clinical workflows and evaluating its impact on diagnostic accuracy and efficiency can validate its potential benefits.

10.7. Handling Noisy Data and Variability

Addressing challenges associated with noisy or inconsistent ultrasound images, such as variations in image quality due to different equipment and settings, could enhance the model's robustness and reliability in real-world scenarios.

In conclusion, this study lays a strong foundation for future research endeavors. The identified avenues for future work span from refining technical aspects to bridging the gap between research and clinical application. By addressing these directions, the field of medical image classification can continue to evolve and contribute to improved healthcare diagnostics and patient care.