Automated Maternal Fetal Ultrasound Image Identification Using a Hybrid Vision Transformer Model

Thunakala Bala Krishna, Ajay Kumar Reddy Poreddy, Kolla Gnapika Sindhu, and Priyanka Kokil

Advanced Signal and Image Processing (ASIP) Lab, Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram, Chennai - 600127, India

Problem Definition and Contribution

Goal: Develop an automated fetal anatomical plane classification approach using a hybrid vision transformer.

Key Contributions:

- A deep learning model called H-ViT, which improves the classification of commonly used fetal anatomical structures by enhancing the interclass variance among fetal planes is proposed.
- The H-ViT model combines DenseNet-121 and Vision Transformer (ViT). Spatial feature maps extracted from the DenseNet-121 backbone are fed into the ViT, which which further refines the fetal US attributes using the attention based capabilities of transformers.
- To ensure the robustness and reliability of the proposed model, we further assessed its performance using a speckle-introduced fetal US image dataset, notably under noisy conditions.

Dataset Description

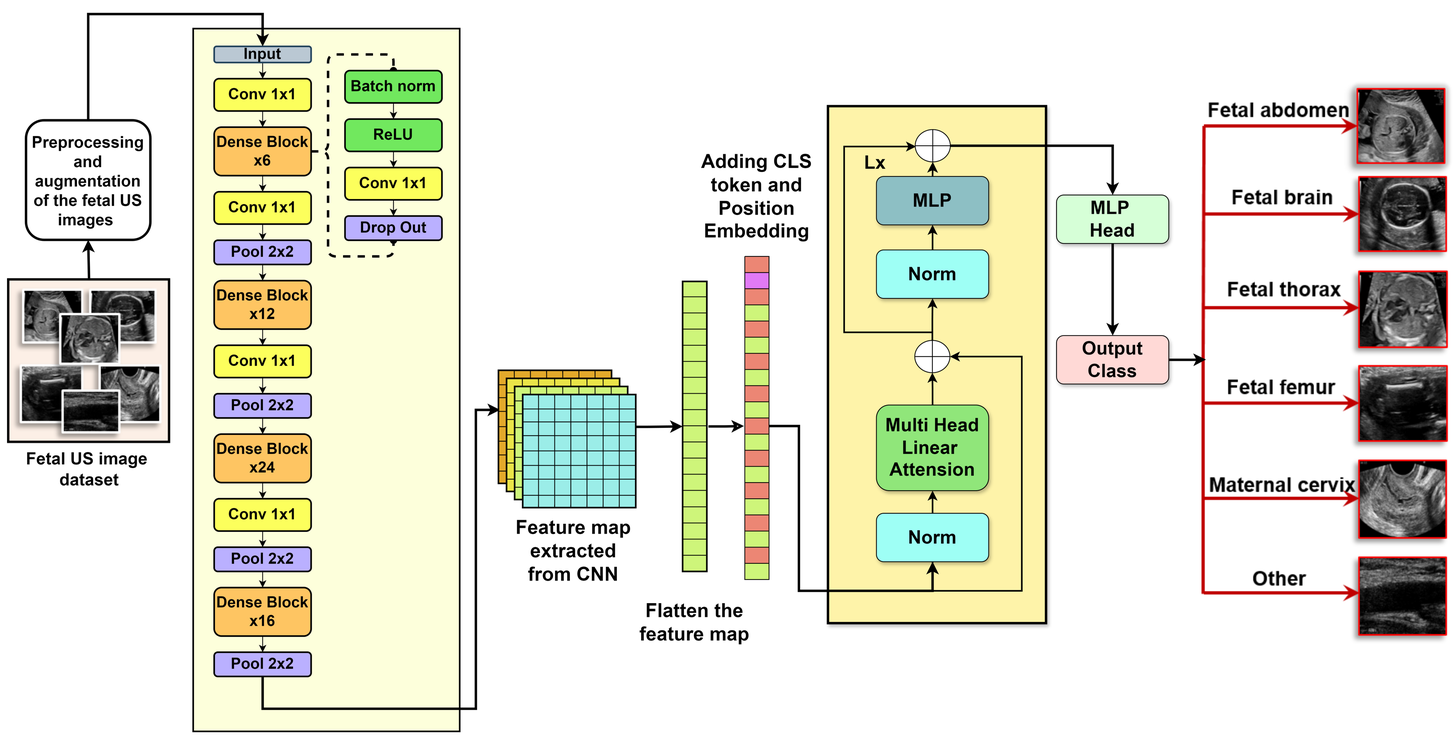
FETAL_PLANES_DB: Common maternal-fetal ultrasound images:

- The dataset consists of 12,400 images, categorized into six classes [1,2].

Fetal Plane	No of Samples
Fetal Abdomen	711
Fetal Femur	1040
Fetal Brain	3092
Fetal Thorax	1718
Maternal Cervix	1626
Other	4216
Total	12400

Method

Overview of the proposed hybrid vision transformer (H-ViT) model:



- The proposed method utilizes hierarchical features extracted from DenseNet-121, which are then inputted into the vision transformer to analyze complex spatial relationships and patterns within fetal US images.
- By incorporating both global and local features, the proposed method enhances feature discriminability, thus alleviating low inter-class variance of fetal US images.

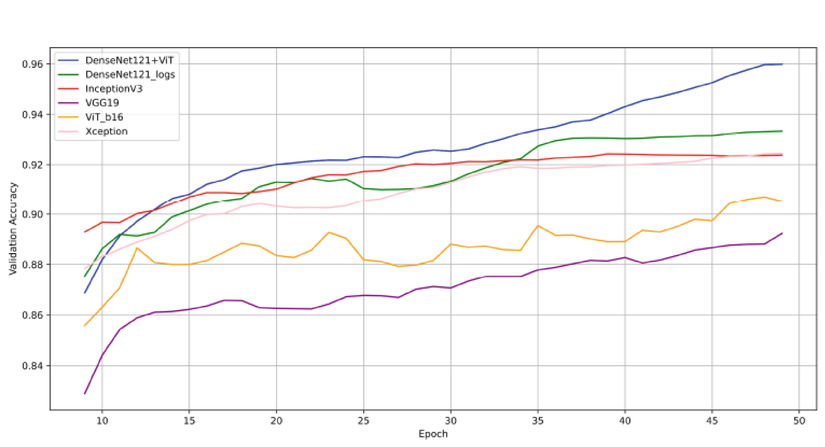
Proposed H-ViT methodology: The following the major sections of the proposed classification model

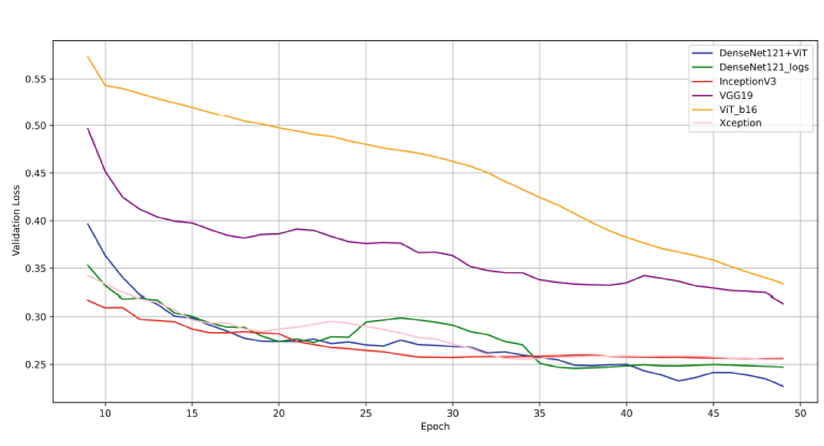
Data augmentation: To enhance the robustness and generalization of the proposed classification method, data augmentation strategy is employed on the maternal-fetal US dataset. The data augmentation techniques such as cropping, rotating, translating, and flipping images are employed.

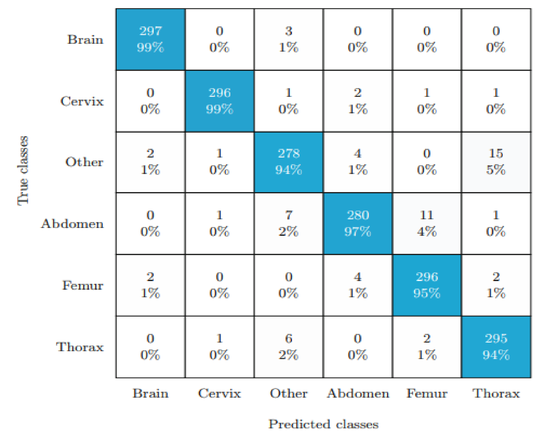
Feature extraction: The pre-trained backbone CNN is fine-tuned on the target fetal US image dataset, enabling the extraction of hierarchical features with varying levels of abstraction.

Vision transformer: Building upon the strengths of ViT, the proposed approach combines ViT with the computational capabilities of the DenseNet-121 backbone feature extractor. The resulting feature maps are systematically divided into smaller, distinct patches for further processing.

Experiments







Accuracy and validation loss values of the state-of-the-art CNN models for different number of epochs, and confusion plot of the proposed method, respectively.

Results

Model	Accuracy	Precision	Recall	F1-score
DenseNet-201	93.00	93.07	93.00	93.01
ViT-base16	89.30	90.81	88.20	89.23
H-ViT	**96.33**	**96.77**	**96.36**	**96.56**

Method	Accuracy	Precision	Recall	F1-score
Xavier et al.	93.73	91.95	93.08	92.5
Zhen Yu et al.	94.12	92.87	94.25	93.48
Baumgartner et al.	94.48	93.56	93.06	93.28
Krishna et al.	95.10	93.83	95.00	94.38
Krishna et al.	95.33	93.58	95.84	94.64
HaifaGhabri et al.	93.63	91.50	93.59	92.48
Sendra et al.	94.94	93.17	95.36	94.19
Proposed method	**96.33**	**96.77**	**96.36**	**96.56**

Model	**Test Acc**	**Precision**	**Recall**	**F1-score**
VGG-19	0.889	0.892	0.889	0.889
Inception-V3	0.931	0.931	0.931	0.931
Xception	0.927	0.927	0.927	0.927
DenseNet-121	**0.940**	0.930	0.93	0.930
ViT	0.893	0.908	0.882	0.892
Proposed (H-ViT)	**0.963**	0.967	0.963	0.965

Model	Acc	Precision	Recall	F1-score
Fold 1	96.33	96.77	96.36	96.56
Fold 2	92.66	93.27	92.41	92.84
Fold 3	95.58	95.97	95.41	95.69
Fold 4	95.16	95.24	95.08	95.16
Fold 5	94.66	94.97	94.41	94.69

Noise level	Test Accuracy	Test Loss	Precision	Recall	F1-score
DenseNet121+Noise					
σ=0.1	0.878	0.366	0.887	0.875	0.876
σ=0.15	0.806	0.639	0.848	0.808	0.814
σ=0.2	0.725	0.978	0.814	0.727	0.735
DenseNet121+VIT+Noise					
σ=0.1	0.928	0.210	0.929	0.928	0.929
σ=0.15	0.898	0.3000	0.900	0.893	0.894
σ=0.2	0.873	0.368	0.881	0.871	0.873

Split ratio	Acc	Precision	Recall	F1-score
60:20:20	85.20	86.79	84.08	85.41
70:15:15	96.33	96.77	96.36	96.56
80:10:10	95.49	95.57	95.49	95.60

Quantitative results such as comparison with state-of-the-art CNN models, existing works, and ablation studies of the proposed work.

Conclusions and Future scope

Conclusions:

- This study introduces an H-ViT that integrates a DenseNet-121 backbone with transformer architecture to enhance the classification accuracy of fetal US planes during prenatal screening.
- To assess the effectiveness of the proposed approach, we utilized a publicly available fetal US image dataset obtained from high-resource settings.

Future scope:

- In the future, it would be interesting to incorporate more advanced preprocessing techniques, such as noise reduction and artifact removal, which can further enhance the quality of input images. Consequently, the performance of the model might be improved.

References

1. Burgos-Artizzu, X.P., Coronado-Guti´errez, D., Valenzuela-Alcaraz, B., BonetCarne, E., Eixarch, E., Crispi, F., Gratac´os, E.: Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. Sci. Rep. 10(1), 10200 (2020).
2. Burgos-Artizzu, X.P., Coronado-Gutierrez, D., Valenzuela-Alcaraz, B., BonetCarne, E., Eixarch, E., Crispi, F., Gratac´os, E.: FETAL PLANES DB: Common maternal-fetal ultrasound images (Jun 2020).
3. Krishna, T.B., Kokil, P.: Automated classification of common maternal fetal ultrasound planes using multi-layer perceptron with deep feature integration. Biomedical Signal Processing and Control 86, 105283 (2023).
4. Krishna, T.B., Kokil, P.: Standard fetal ultrasound plane classification based on stacked ensemble of deep learning models. Expert Systems With Applications. 238, 122153 (2024)