Learnable 2D Gaussian Filters for Computationally Efficient Abdominal Organ Classification

Sifat Zina Karim, Sabyasachi Biswas, and Dr. John Ball

Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762, USA

ABSTRACT

This paper presents a novel approach for real-time classification of abdominal organs in ultrasound images by integrating learnable 2D Gaussian filters into the neural network architecture, superscribing the computational complexities of state-of-the-art models like VGG16, ResNet, DenseNet, etc., with a custom CNN which employs a 2D parameterized Gaussian filter as its initial layer. Unlike conventional convolutional layers, the Gaussian filter layer uses only mean and standard deviation as the learnable parameters. This reduction in complexity enables real-time processing without sacrificing classification accuracy. To test the robustness of our proposed approach, we conducted comparative experiments using VGG-16, ResNet-50, DenseNet-121, ALexNet, custom-based CNN and the proposed 2D-Gaussian model. The proposed model achieved a classification accuracy of 87%, outperforming other methods by around 4% while maintaining better computational efficiency as it reduces the overall parameter count of the network lowering the memory with it. Our findings highlight the potential of parameterized Gaussian filters for real-time medical image processing, offering an efficient alternative to widely used models in the domain.

Keywords: VGG-16, ResNet-50, DenseNet-121, AlexNet Convolutional Neural Network (CNN), deep learning, Ultrasound, abdomen, abdominal organs, Gaussian filter.

1. INTRODUCTION

Medical image classification has been one of the lucrative goals of deep learning. It can revolutionize this field because it can provide huge support to medical associates where they diagnose diseases or abnormalities more quickly, which will require less time. Ultrasonic imaging enables internal structural views without causing any harm. Also, making it adaptive to real-time applications has been challenging as it requires large computational expenses with a proportionately high number of parameters.

Litjens et al, showed deep learning models, especially convolutional neural networks have been fairly successful in medical image analysis.[1] Xie et al, also show deep learning has facilitated disease diagnosis, lesion and abnormality detection, lesion, and organ segmentation, etc.[2] some of the works that focus only on the classification for medical are namely,[3],[4] and.[5] In terms of computational efficiency, the number of parameters as well as real-time application has been difficult to combine.

There have been attempts to use a combination of multiple deep learning models to do a classification task for medical images. An ensemble of deep neural networks for kidney ultrasound image classification has been used in[6] where pre-trained ResNet-101,[7] MobileNet-V2,[8] and ShuffleNet[9] have been used, followed by a support vector machine to classify the images into 4 categories, namely normal, cyst, stone, and tumor. The majority voting technique is used to make the final prediction. The ultrasound images were collected from datasets[10] and[11] and radiologists. Although better performance were acheived combining multiple models together, using multiple models including ResNet-101, MobileNet-V2, and ShuffleNet increased computational complexity and resource requirements. As a result, this deos not suit real-time predictions. Furthermore, transfer learning helps in reducing the training time but the ensemble still requires substantial time for inference.

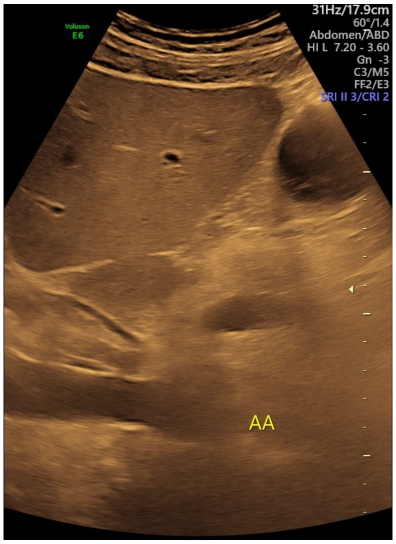
VGG-16 has been popular for classification tasks for its simple architecture. This paper[12] utilizes the Visual Geometry Group (VGG)-16T model, which is based on the VGG-16 architecture with additional batch normalization (BN) and dropout layers in addition to the fully connected layers. This proposed network has provided satisfactory specificity, sensitivity, and accuracy in classifying malignant and benign Thyroid nodules based on the remarks of experienced radiologists. However, this study used a small number of samples limiting the generalizability of the proposed deep-learning algorithm to larger populations or different clinical settings. Moreoever, the study shows overreliance on a single model solely that is VGG-16T. Furthermore, this research did not take into account the correlation between the algorithm's accuracy and factors such as TN (Thyroid Nodules) size or cancer subtypes. which puts restrictions on the clinical applicability and diagnostic accuracy of the proposed model. Additionally, the medical images include substantial information beyond the images themselves—such as labels indicating lesion conditions, clinical diagnosis reports reflecting physicians' conclusions, and descriptive details—that can be challenging to fully utilize.

Another widely utilized architecture in medical image processing is DenseNet due to its feature extraction capabilities. Zhong et al 2020, proposed a hybrid DenseNet model for metastatic cancer image classification.[13] The authors in this paper utilized the PatchCamelyon (PCam) dataset[14] and showed better performance compared to ResNet34 and Vgg-19 models. However, the dense structure of the model increases computational complexity. These computational demands make it difficult to adapt to real-time applications where quick decision-making is required. Zhou et al had found some other obstacles for using dense architecture in their paper[15] where it was mentioned that a large amount of images for creating a medical dataset is often hard to come by.

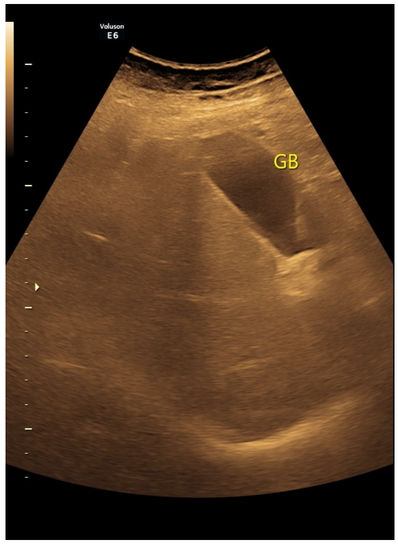
So, the challenge has been to reduce the number of parameters without sacrificing the accuracy of classification. To address this problem, we develop a novel CNN architecture that utilizes learnable 2D Gaussian filters as the initial layer of CNN. Learnable filters were first utilized in[16] for audio signal classification. Where One dimentional learnable sinc filter banks were utilized. Furthermore, Furthermore,[17–21] introduced complex-valued 1D-CNNs and 2D-CNN architectures that uses learnable filters for RF sensing based applications. Taking inspiration from these researches, we developed a 2D learnable gaussian filter based CNN model. This architecture reduces the number of learned parameters from each Gaussian filter applied for convolution. Where M number of 2D convolutional filters with size N learns $M \times N \times\ N$ number of parameters, the 2D Gaussian filters learns only mean and standard deviations in both dimensions. This makes the number of learnable parameters as $M \times 4 \times\ N$. This makes the total number of parameters lower, thus requiring less computational power.

2. DATASET

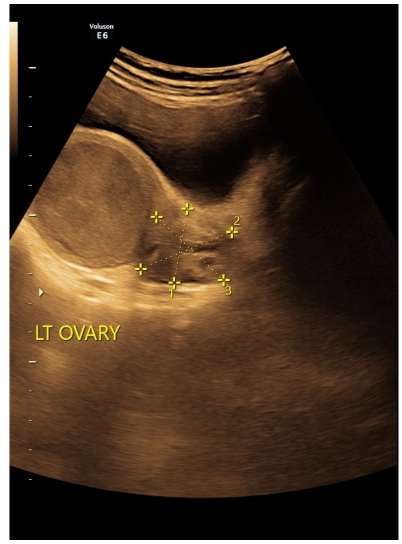
The dataset, comprising 2,786 images of different abdominal organs, namely Aortic Aorta, Gall bladder, Urinary bladder, Ovaries, Kidneys, Prostrate, Uterous, Hepatic vein, Portal vein, Liver, Pancreas and Spleen, was collected from 242 patients at MH Samorita Hospital and Medical College in Dhaka, Bangladesh. The images were collected from August 2023 to December 2024. The ultrasound machine used for acquiring the images is Voluson™ E6 BT16 which is an Active Diagnostic Medical Product pertaining to Class IIa according to the MDD 93/42/EEC regulation for use on human patients. The diversity in anatomy and acquisition in our dataset presents a special challenge that is well-suited for investigating various model designs. The images were preprocessed to remove the patient's information from the images and split into halves to make the classification easier. We used a train-test split meaning 80% for training purposes and the other 20% for testing purposes. Figure 1 shows some of the sample images from the dataset and Table 1 shows the dataset distribution for training and testing purposes. We have utilized NVIDIA GeForce RTX 4070 for both data processing and network training. Pytorch was utilized for the implementation of the models.



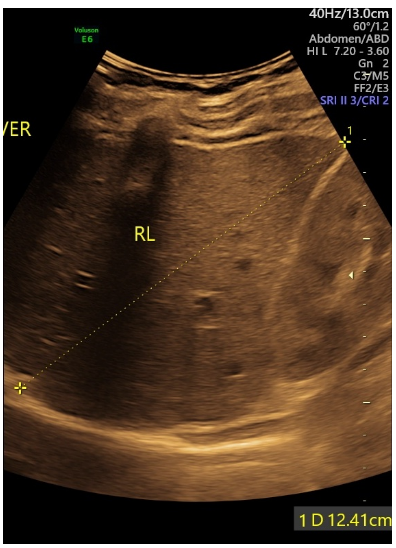
(a) Aortic Aorta



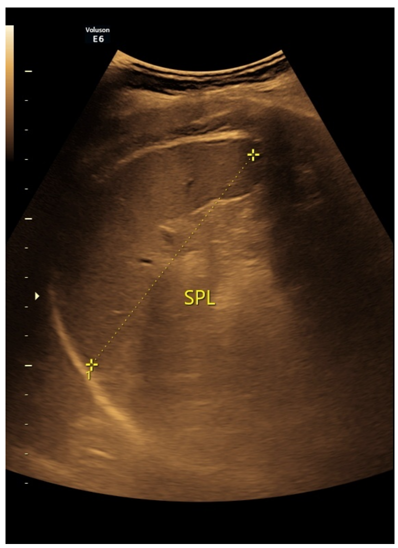
(b) Gall bladder



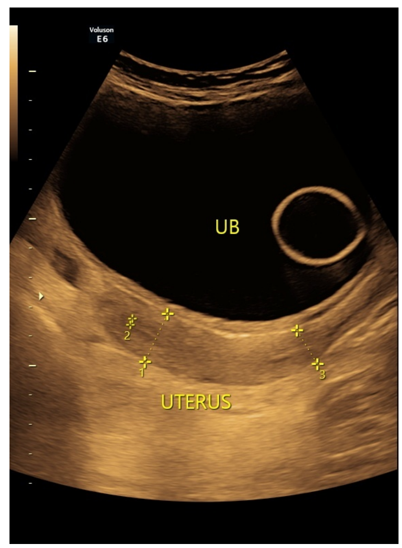
(c) Ovary



(d) Kidney



(e) Spleen



(f) Urinary bladder+ Uterous

Figure 1: Sample images from the dataset

Table 1: Training and Validation Dataset.

	Dataset Samples	**Training**	**Testing**
Track	2,786	2,228	558
Split	100%	80%	20%

3. METHODOLOGY

In this study, we have presented a novel classification approach utilizing a parameterized learnable filter to provide easy interpretability without sacrificing the accuracy of deep neural networks (DNNs) in processing raw ultrasound images. Traditional deep learning models depend on static convolutional filters that are learned from the data in an unrestricted manner. That is why it sometimes fails to extract crucial features from the ultrasound images. We propose a structured approach where the mean (μ) and standard deviation (σ) of a Gaussian filter are taken to be the learnable parameters that let the model flexibly distill the extraction of features in a more physically consequential way. We have utilized the formula for calculating the 2D Gaussian function containing means μ_x andμ_y, σ_x and σ_y representing the standard deviation of the Gaussian in X, and Y direction, X and Y representing the coordinates in X, and Y direction. It creates a Gaussian function centered at μ_x andμ_y.

$$\text{Gaussian} = \exp\left(-\left(\frac{(X-\mu_x)^2}{2\sigma_x^2} + \frac{(Y-\mu_y)^2}{2\sigma_y^2}\right)\right)$$

Here, μ_x and μ_y determine the center of the filter, σ_x, and σ_y sway the spread of the Gaussian which measures up to regulation of the smoothness and edge enhancement. This formulation is differentiable which makes it possible to regulate the mean and variation through backpropagation. It helps the model to learn dynamically from the images. The methodology is structured as follows:

3.1 Deep Learning Architecture

Our architecture is based upon standard deep learning components with a custom Gaussian 2D layer for improved feature extraction. The architecture comprises of: Convolutional Layers that are used for spatial feature extraction with complex filters. Max-Pooling Layers for reducing spatial dimensions while keeping important features. Batch Normalization layers to stabilize and accelerate training. ReLU activation is used to introduce non-linearity and dropout layers to prevent overfitting during training. The flow diagram of our model is provided in figure 2 and figure 3.

3.2 Model Comparison

To evaluate the performance of the Gaussian 2D layer, we have compared its classification accuracy with other popular deep learning models, including: VGG-16,, ResNet-50, DenseNet-121, AlexNet, and Custom CNN (baseline without Gaussian 2D layer). The custom CNN architecture was identical across experiments except for the integration of the Gaussian 2D layer.

4. RESULTS AND OBSERVATIONS

We have recorded the classification accuracy, precision, recall, and sensitivity of each model on the ultrasound images. As shown in Table 2 the Gaussian 2D model achieved the highest accuracy (87%), outperforming standard architectures such as DenseNet-121, VGG-16, ResNet-50, and the baseline custom CNN. The results demonstrate that the Gaussian 2D layer enhances feature extraction and overall model performance for ultrasound image classification tasks.

Table 2 also shows the number of parameters that have been used by each model. VGG-16 which has a very high number of parameters did well whereas ResNet-50 with the smallest number of parameters did the poorest. DenseNet-121 had a high number of parameters and did somewhat fair but AlexNet with a smaller number of parameters did not do well. Custom-based CNN had the second lowest number of parameters and did reasonably well. However, our proposed model had the highest accuracy and smallest number of parameters. A higher F1 score means a better balance between precision and recall. Our proposed model has the highest F1 score which can be seen from the table 2.

Table 2: Evaluation Matrices

	VGG-16	**ResNet-50**	**DenseNet-121**	**AlexNet**	**Custom CNN**	**Gaussian 2D Model**
Accuracy	84.59%	75.27%	84%	82.26%	83.33%	87%
Parameters	728930	1484538	6957834	2121293	176938	176746
Precision	0.83	0.75	0.83	0.81	0.87	0.88
Recall	0.82	0.71	0.83	0.78	0.86	0.85
F1 Score	0.81	0.71	0.82	0.79	0.86	0.86

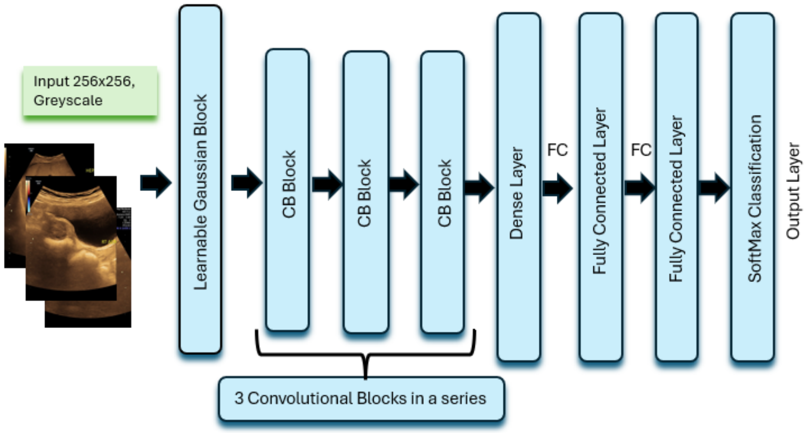


Figure 2: Flow diagram of the proposed architecture

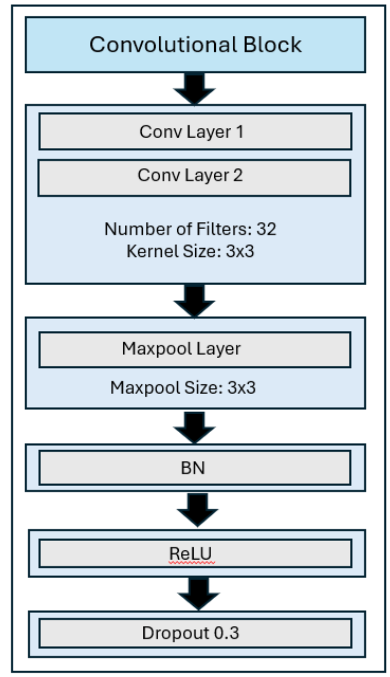


Figure 3: Inside the Convolutional Block

5. ACKNOWLEDGEMENT

We would like to express our sincerest gratitude and heartfelt appreciation to the medical professionals, staff, and individuals of the MH Samorita Hospital and Medical College in Dhaka, Bangladesh who dedicated their time, effort, and their invaluable support in collecting the ultrasound images of 242 patients. It has been instrumental in this research.

6. CONCLUSION AND FUTURE WORK

We have proposed a novel approach for ultrasound image classification using a parameterized learnable filter with only two learnable parameters namely, mean and standard variance in a semblance of Gaussian 2D layer within deep learning architecture. We have compared the performance of our model concerning other famous deep learning models such as VGG-16, ResNet-50, DenseNet-121, AlexNet, and a baseline custom CNN. It has provided the highest accuracy of 87% compared to these state-of-the-art models. The results validate the effectiveness of this proposed model in better feature extraction and classification of the ultrasound images. This proposed model also offers improved latency which makes it a promising candidate for real-time medical image analysis. Our further research goal is to work more on the architecture to optimize it for edge computing platforms and explore its applicability to other medical imaging domains to further enhance its utility in clinical diagnostics.

REFERENCES

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

[2] Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., and Yu, S., “A survey on incorporating domain knowledge into deep learning for medical image analysis,” *Medical Image Analysis* **69**, 101985 (2021).

[3] Kareem, F. Q., Abdulazeez, A. M., et al., “Ultrasound medical images classification based on deep learning algorithms: a review,” *Fusion: Practice and Applications* **3**(1), 29–42 (2021).

[4] Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., and Yao, Y., "Deep learning in medical ultrasound image analysis: a review," *IEEE Access* **9**, 54310–54324 (2021).

[5] Yang, J., Shi, X., Wang, B., Qiu, W., Tian, G., Wang, X., Wang, P., and Yang, J., "Ultrasound image classification of thyroid nodules based on deep learning," *Frontiers in Oncology* **12**, 905955 (2022).

[6] Sudharson, S. and Kokil, P., "An ensemble of deep neural networks for kidney ultrasound image classification," *Computer Methods and Programs in Biomedicine* **197**, 105709 (2020).

[7] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., "Mobilenetv2: Inverted residuals and linear bottlenecks," in [*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 4510–4520 (2018).

[9] Zhang, X., Zhou, X., Lin, M., and Sun, J., "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in [*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 6848–6856 (2018).

[10] Antony, J., "Ultrasound-images.com." https://www.ultrasound-images.com/ (2015). Accessed: March, 2020.

[11] Geertsma, T., "Ultrasoundcases.info." http://www.ultrasoundcases.info/ (2011). Accessed: March, 2020.

[12] Zhu, Y.-C., Jin, P.-F., Bao, J., Jiang, Q., and Wang, X., "Thyroid ultrasound image classification using a convolutional neural network," *Annals of translational medicine* **9**(20) (2021).

[13] Zhong, Z., Zheng, M., Mai, H., Zhao, J., and Liu, X., "Cancer image classification based on densenet model," in [*Journal of physics: conference series*], **1651**(1), 012143, IOP Publishing (2020).

[14] Kaggle, "Metastatic tissue classification - patchcamelyon," (March 2019). Accessed: Feb. 18, 2025.

[15] Zhou, T., Ye, X., Lu, H., Zheng, X., Qiu, S., and Liu, Y., "Dense convolutional network and its application in medical image analysis," *BioMed Research International* **2022**(1), 2384830 (2022).

[16] Ravanelli, M. and Bengio, Y., "Speaker recognition from raw waveform with sincnet," in [*2018 IEEE Spoken Language Technology Workshop (SLT)*], 1021–1028 (2018).

[17] Biswas, S., Ayna, C. O., and Gurbuz, A. C., "Plfnets: Interpretable complex valued parameterized learnable filters for computationally efficient rf classification," *IEEE Transactions on Radar Systems* (2024).

[18] Stadelmayer, T. and Santra, A., "Parametric convolutional neural network for radar-based human activity classification using raw adc data," (09 2020).

[19] Biswas, S., Ayna, C. O., Gurbuz, S. Z., and Gurbuz, A. C., "Complex sincnet for more interpretable radar based activity recognition," in [*2023 IEEE Radar Conference (RadarConf23)*], 1–6 (2023).

[20] Biswas, S., Ayna, C. O., Gurbuz, S. Z., and Gurbuz, A. C., "Cv-sincnet: Learning complex sinc filters from raw radar data for computationally efficient human motion recognition," *IEEE Transactions on Radar Systems* **1**, 493–504 (2023).

[21] Hicks, B., Biswas, S., Dahal, A., Gonzalez, V., and Gurbuz, A. C., "Automatic waveform recognition from complex rf data with filter-based deep learning," in [*NAECON 2024 - IEEE National Aerospace and Electronics Conference*], 168–172 (2024).