# Fetal Ultrasound Image Denoising Autoencoder Network

Chawla, Enakshi
University of North Carolina at Chapel Hill
Joint Department of Biomedical Engineering
Chapel Hill, NC, USA
echawla@unc.edu

Venkatapathy, Varsha
University of North Carolina at Chapel Hill
Joint Department of Biomedical Engineering
Chapel Hill, NC, USA
vvenkatapathy@unc.edu

Wu, Zehao (Matthew)
University of North Carolina at Chapel Hill
Joint Department of Biomedical Engineering
Chapel Hill, NC, USA
wuzehao@unc.edu

Giraud, Nalaya
University of North Carolina at Chapel Hill
Joint Department of Biomedical Engineering
Chapel Hill, NC, USA
ngiraud@unc.ed

***Abstract*— This paper presents a replication study of the LAD-CNN (Lightweight Attention Denoise-Convolutional Neural Network) model designed for ultrasound image denoising, originally proposed by Shi et al. [1] The LAD-CNN model was trained using just the fetal head circumference ultrasound public dataset [2] and its effectiveness in reducing speckle noise while preserving image details was validated. BM3D, a state-of-the-art conventional denoising algorithm, was implemented to compare objective and subjective results. The implementation below was conducted in Jupyter Notebook utilizing Keras/Tensorflow, and the models were assessed using Peak Signal-to-Noise Ratio (PSNR), Equivalent Number of Looks (ENL), and Structural Similarity Index Measure (SSIM) across multiple noise levels. Experimental results in the article demonstrated that the LAD-CNN model outperformed BM3D in all objective metrics and subjective clarity, however the implementation utilized below encountered issues in achieving that standard. Specifically, in the article, LAD-CNN achieved PSNR values about 25 and SSIM values about 0.85 for all noise levels.**

***Keywords—component, formatting, style, styling, insert (****key words****)***

## *I.* INTRODUCTION

Ultrasound imaging is a widely used diagnostic tool in clinical settings due to its non-invasiveness, cost-effectiveness, and real-time imaging capabilities. However, one of the major challenges in ultrasound imaging is the presence of speckle noise, an inherent granular interference caused by the coherent nature of the ultrasound wave propagation. It occurs because sound waves scatter and interfere with each other creating random dark/light spots on the ultrasound image. Speckle noise severely degrades image quality, obscuring important structural and textural details, which in turn affects the accuracy of medical diagnoses. Consequently, developing efficient methods to suppress speckle noise while preserving essential anatomical features is of significant interest in medical image processing.

Conventional approaches to ultrasound image denoising typically involve filtering techniques, such as Block-Matching and 3D Filtering (BM3D) and wavelet-based methods. While these techniques are effective to some extent, they often result in oversmoothing and loss of critical image details, limiting their utility in clinical applications. To address these limitations, deep learning methods using machine learning have gained traction due to their ability to learn complex features and achieve superior denoising performance.

In the original study by Shi et al. [1], a Lightweight Attention Denoise-Convolutional Neural Network (LAD-CNN) was proposed, introducing a novel approach that combines two lightweight attention blocks: the Lightweight Channel Attention (LCA) block and the Lightweight Large-Kernel Attention (LLA) block. The model architecture has an encoder stage that implements the LCA block after pooling layers and a decoding stage that implements the LLA block after upsampling layers. During encoding, downsampling/pooling is used for feature extraction and edge detection. The LCA block is added after downsampling to avoid detail loss and emphasizing important feature channels while suppressing noisy ones. During decoding, subpixel convolution upsampling was used for recovering image size, converting depth to space for feature extraction and rearranging pixels from multi-channel to single channel high resolution. The LLA block was added after upsampling. The LLA block consisted of a very large kernel size to capture relations between far apart pixels. However, this increased the number of parameters and computational time. Two skip connections were used for feature integration between encoding and decoding stages and reduction of gradient vanishing. By incorporating these blocks, the LAD-CNN model enhanced feature extraction, preserved image texture, and mitigated the vanishing gradient problem through the use of skip connections. The authors demonstrated that LAD-CNN achieved superior performance over existing denoising methods, both quantitatively and qualitatively. Standard convolution (Conv2D) performed the channelwise and spatial-wise computation in one step and was used during many layers of implementation. In the LLA block, depthwise convolution and a dilation factor in convolution were used. This broke the computation into two steps: Each filter was applied to a single input channel (no cross-channel interaction). Feature maps for individual channels were created. Pointwise Convolution was used: 1x1 filters to combine the results of the depthwise convolution across all channels. Then, these two steps were added through channel-wise interaction.

In this study, two replications of the LAD-CNN model were created and those were compared against BM3D, a well-established state-of-the-art denoising technique. The images consisted of varying levels of speckle noise, and their effectiveness was evaluated using metrics such as Peak Signal-to-Noise Ratio (PSNR), Equivalent Number of Looks (ENL), and Structural Similarity Index Measure (SSIM). By conducting this comparative analysis, insights into the robustness and efficacy of the LAD-CNN model were gained for practical ultrasound image denoising applications.

## II. METHODS

### *A. Speckle Noise*

In ultrasound images, speckle noise is modeled by:

$$z = \mu * n + \mu \tag{1}$$

where gamma is the parameter affected by the actual ultrasound equipment which is set to 0.5 for the equipment used in the paper as that is consistent with its ultrasound image.

### *B. Preprocessing / Experimental Environment*

To evaluate the effectiveness of the proposed denoising model, experiments were carried out on real fetal ultrasound images with simulated speckle noise. All experiments were implemented on the Keras/Tensorflow Jupyter Notebook framework and accelerated with NVIDIA GeForce 4060 Laptop GPU and Intel(R) Iris(R) Xe Graphics. We used the fetal head 2D ultrasound public data set which contains 1,334 images with a size of 800 px × 540 px where 999 images were in the training set and 335 images in the test set. The pixel size for each image ranged from 0.052 to 0.326 nm. A trained sonographer manually annotated every image's head circumference. [2] The images were resized to 128 x 128 px. The training resized images were augmented five times each through random rotation,

translation, and flipping. This resulted in about 5000 images. These augmented images were clean images, and their corresponding noisy images were created by adding speckle noise. The ground truth clean images and noisy images were matched and separated by noise level. Finally, 25% of the training data was moved to validation, so about 1250 images.

### C. Quantitative Evaluations

The following three commonly used evaluation metrics were used to quantitatively assess the effectiveness of the ultrasound image denoising: the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and equivalent number of looks (ENL). The PSNR measures the pixel-by-pixel difference between the denoised image and the original image, as shown in Equation 2. The SSIM evaluates the similarity in terms of brightness, contrast, and structure as shown in Equation 3. The ENL serves as an indicator for evaluating the smoothness of the images in homogeneous areas, providing a metric for assessing denoising efficiency in the absence of original images as shown in Equation 4. Higher PSNR and SSIM values indicate a greater similarity between the denoised image and the original image, indicating better denoising performance, while a higher ENL value indicates greater accuracy in the denoised image. During training, Adam optimizer was used with an initial learning rate of 0.001 multiplied by a decay factor of 0.1 if the PSNR failed to improve after 5 epochs.

The batch size was 10 images, and they were trained for 120 epochs per noise level. The implementation below incorporated early stopping at 20 epochs when the PSNR and SSIM stopped improving. The weight parameters for the highest PSNR and SSIM in training were reported. To assess the ability of the proposed model to reduce speckle noise, comparative experiments were performed on the same fetal ultrasound dataset using two BM3D models.

The comparison metrics used to assess the effectiveness of the models were PSNR, SSIM, and ENL, which are defined here:

PSNR: $$10 \cdot \log_{10} \frac{(u_{\max})^2}{\mathrm{MSE}(u,x)} \quad (2)$$

SSIM: $$\frac{(2\mu_x\mu_u + c_1)(2\sigma_{xu} + c_2)}{(\mu_x^2 + \mu_u^2 + c_1)(\sigma_x^2 + \sigma_u^2 + c_2)} \quad (3)$$

ENL: $$\mathrm{ENL} = \frac{\mu_x^2}{\sigma_x^2} \quad (4)$$

### D. Loss Function

In this paper, the loss function integrates the Mean Squared Error (MSE) function with the Total Variation (TV) regularization function to enhance image denoising performance. The MSE function shown here:

$$L_{MSE} = \frac{\sum_{i=1}^{n}(u-x)^2}{n} \quad (5)$$

minimizes the difference between the original and denoised images, guiding the model toward accurate reconstruction. However, relying solely on the MSE function may not fully eliminate artifacts that can emerge during the denoising process.

To address these residual artifacts, the TV regularization function described here:

$$L_{TV} = \sum_{w=1}^{W}\sum_{h=1}^{H}\sqrt{\left(x^{w+1,h} - x^{w,h}\right)^2 + \left(x^{w,h+1} - x^{w,h}\right)^2} \quad (6)$$

is incorporated. The TV function smooths the images by suppressing abrupt intensity variations between neighboring pixels, thereby reducing artifacts without excessively blurring important structural details.

Building upon these functions, the overall loss function:

$$L_{LOSS} = L_{\mathrm{MSE}} + \gamma_{TV} L_{TV} \quad (7)$$

is defined as a weighted combination of the MSE and TV losses, where the regularization weight, denoted as $\gamma$TV,

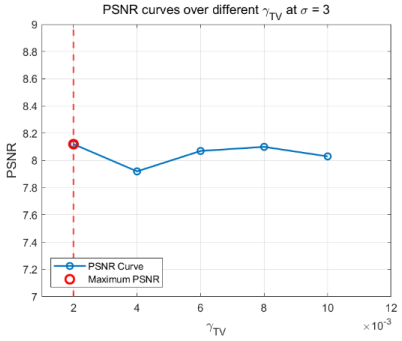


Fig. 1 : PSNR curves over different regularization parameters $\gamma$TV at σ=3. PSNR, peak signal-to-noise ratio; LAD-CNN, lightweight attention-denoise-convolutional neural network; $\gamma$TV, predefined weight for the TV loss function; TV, total variation; σ, noise level.

is carefully selected to balance the trade-off between smoothing and detail preservation by setting the value significantly smaller than 1.

Based on experimental results using the same training setup as the target paper, a regularization weight of $\gamma$TV = 0.002 was determined to yield the optimal denoising performance at noise level σ = 3. This value is lower than that reported in the original paper. Fig. 1 presents the experimental results, showing PSNR values corresponding to various $\gamma$TV values.

### E. Activation Function

The model employs three activation functions across various convolutional layers: Gaussian Error Linear Unit (GELU), Rectified Linear Unit (ReLU), and the Sigmoid function. These functions introduce non-linearity, enabling the network to capture complex patterns essential for effective image denoising.

GELU is an activation function that smoothly adjusts outputs based on input values. It considers both the value and its likelihood under a normal distribution, allowing the model to make more nuanced decisions about which features to activate. ReLU is simple and efficient–it outputs zero for any negative input and passes through positive inputs unchanged. This helps the network train faster and reduces computational complexity. The Sigmoid function squeezes input values into a range between zero and one, effectively normalizing the outputs. It's useful in layers where outputs need to represent probabilities or be within a specific range.

By incorporating these activation functions, the model effectively captures complex data patterns, reduces noise, and preserves important image details.

TABLE I. MODEL NAMES AND DESCRIPTIONS USED IN PAPER

| Model | Model Description |
|---|---|
| Model A | Exact replication of LAD-CNN in target paper |
| Model B | Alternative Variation of LAD-CNN model in target paper |

TABLE II. OBJECTIVE METRICS FOR BM3D

| Metrics | Noise Level (σ) | | | | | |
|---|---|---|---|---|---|---|
| | $\sigma = 2$ | $\sigma = 3$ | $\sigma = 4$ | $\sigma = 5$ | $\sigma = 6$ | $\sigma = 7$ |
| PSNR | 24.50 | 22.35 | 20.73 | 19.43 | 18.6 | 17.86 |
| SSIM | 0.8011 | 0.7453 | 0.6951 | 0.6462 | 0.6098 | 0.5797 |
| ENL | 1.04 | 1.25 | 1.39 | 1.53 | 1.64 | 1.73 |

TABLE III. OBJECTIVE METRICS FOR MODEL A

| Metrics | Noise Level (σ) | | | | | |
|---|---|---|---|---|---|---|
| | $\sigma = 2$ | $\sigma = 3$ | $\sigma = 4$ | $\sigma = 5$ | $\sigma = 6$ | $\sigma = 7$ |
| PSNR | 15.63 | 15.77 | 15.36 | 15.77 | 14.97 | 12.3 |
| SSIM | 0.27 | 0.26 | 0.26 | 0.27 | 0.24 | 0.19 |

## III. RESULTS

### A. Results Intro

In the target paper, six noise models with different noise levels (σ = 2.0, 3.0, 4.0, 5.0, 6.0, and 7.0) were trained and compared to the LAD-CNN model. The effectiveness of each model was evaluated based on both subjective visual effects and objective evaluation metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Equivalent Number of Looks (ENL), and denoising time. In the replication study, the focus was on comparing implemented versions of LAD-CNN to a replicated BM3D model, a well-established traditional denoising method.

### B. BM3D Results

Table 2 presents the average PSNR, SSIM, and ENL values for BM3D across different noise levels. While the BM3D implementation performs slightly worse than the BM3D results reported in the original paper, it still follows a similar trend. Specifically, BM3D performs well at lower noise levels, but its performance deteriorates at higher noise levels, highlighting a limitation in traditional denoising methods.

The SSIM values for BM3D were closer to those reported in the original paper and exhibited a similar trend. It is worth noting that ENL values were calculated on a private dataset in the original paper and may not be directly comparable to the dataset used in this study; however, ENL values are generally expected to increase as the noise level increases.

### C. Implementation of LAD-CNN

Two versions of the LAD-CNN model were implemented, referred to as Model A and Model B, as summarized in Table 1. Model A is an exact replication of the LAD-CNN architecture presented in the original paper, while Model B is an alternative variation that utilizes a CNN-DAE backbone,

TABLE IV. OBJECTIVE METRICS IN MODEL B

| Metrics | Noise Level (σ) | | | | | |
|---|---|---|---|---|---|---|
| | $\sigma = 2$ | $\sigma = 3$ | $\sigma = 4$ | $\sigma = 5$ | $\sigma = 6$ | $\sigma = 7$ |
| PSNR | 15.56 | 15.84 | 15.79 | 15.75 | 15.60 | 15.85 |
| SSIM | 0.263 | 0.2681 | 0.269 | 0.2786 | 0.2622 | 0.2686 |
| ENL | 2132394.36 | 1894886.08 | 1926696.01 | 1894414.04 | 1959841.71 | 1891849.66 |
| Loss | 0.1579 | 0.1567 | 0.1564 | 0.1570 | 0.1572 | 0.1574 |

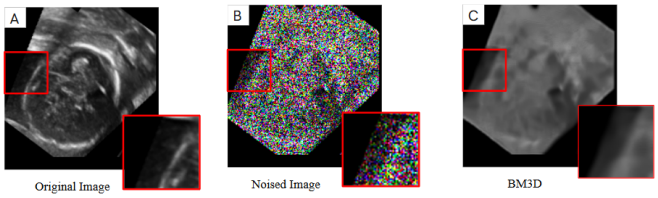


Fig. 1. Subjective visual comparision of the denoised fetal head iamges at σ =5. (A) Original fetal ultrasound, (B) noisy fetal ultrasound, (C) BM3D denoising ultrasound; BM3D, block-matching and three-dimensional filtering; σ, noise level.

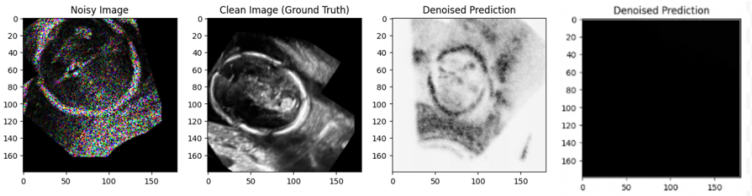


Fig. 2. Subjective Results LAD-CNN as follows, left to right: Noisy, Clean, Model B, Model A

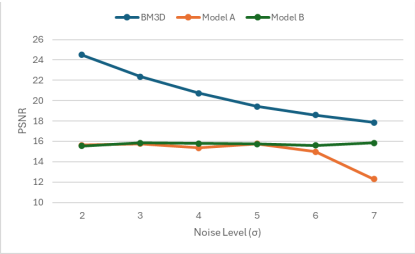


Fig. 3. Variations in the denoising evaluation metrics for each model at different noise levels. The average PSNR for the fetal head data set. PSNR, peak signal-to-noise ratio; σ, noise level; BM3D, block-matching and three dimensional filtering.

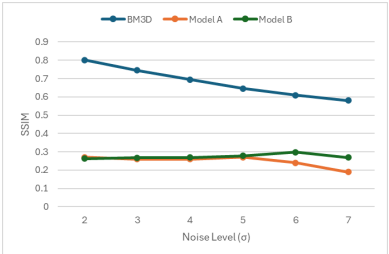


Fig 4. Variations in the denoising evaluation metrics for each model at different noise levels. The average SSIM for the fetal head data set. SSIM, structural similarity; σ, noise level; BM3D, block-matching and three dimensional filtering

TABLE V. COMPARISON OF DENOSING TIMES

| Model | Time(s) |
|---|---|
| BM3D | 1.45 |
| LAD-CNN | 0.08 |

incorporating some features proposed in the original model. Due to challenges in exactly replicating LAD-CNN, Model B was developed as an alternative approach.

Table 4 presents the average PSNR, SSIM, ENL, and Loss values for Model B. These values are significantly lower than those reported in the original paper, and the ENL values deviate substantially. Table 3 illustrates the average PSNR and SSIM values for Model A, the exact replica of the LAD-CNN network architecture. While the performance metrics for Model A are lower than those achieved in the original paper, Model A performs similar to Model B. In fact, Model B performs the most consistent across all noise levels.

### *D. Performance of Model A and B against BM3D*

When comparing these models against the BM3D replication, BM3D outperforms both Model A and Model B in terms of PSNR and SSIM. This contrasts with the original paper, where their implementation of LAD-CNN outperformed all pre-existing traditional and deep-learning methods tested.

One notable observation is that the range of PSNR and SSIM values for Model A and Model B is much smaller than that for BM3D. Specifically, based on Table 4, the PSNR for Model B ranges from approximately 15.56 to 15.85, while Table 3 shows that Model A's PSNR ranges from about 12.30 to 15.80. For SSIM, Model B's values range from 0.26 to 0.28, and Model A's values range from 0.19 to 0.27. In contrast, BM3D's PSNR ranges from f17.90 to 24.50, and its SSIM ranges from 0.58 to 0.80.

This pattern is also evident in Fig. 3 and 4, which show the variation of PSNR and SSIM across the three models. BM3D exhibits a steady decline in performance as the noise level increases, whereas Models A and B maintain relatively consistent performance, with their metrics remaining relatively close to each other across noise levels.

### *E. Denoising Time Comparisons*

As shown in Table 5, BM3D took longer to denoise a single image compared to Models A and B, aligning with findings in the original paper. The increased processing time for BM3D is characteristic of traditional denoising methods, which often require more computational resources.

### *F. Subjective Visual Results*

Regarding subjective visual results, Fig. 2 displays the original image, the noisy images at various noise levels, and the denoised images produced by BM3D. While BM3D effectively reduces noise, it tends to over-smooth the images, leading to a noticeable loss of texture and detail.

Fig. 2 showcases the visual results from Model A and B. The denoised image for Model B has a strong similarity to the clean image but with smoothening effects and an inversion of color. Model A, on the other hand, did not produce meaningful outputs, resulting in images that are effectively black boxes as shown in the rightmost graph in Fig. 2. Note that the visual images vary significantly, from model to model and inside a single model from image to image at different noise levels. For example, in Model A the images varied from black boxes, dark gray, light gray, and white boxes randomly throughout the dataset. In Model B, the images varied from noisy, rotated, over smoothened, etc. throughout each image and noise level.

### *G. Summary of Results*

BM3D outperforms both Model A and Model B in this replication study, contrary to the results reported in the original paper. The challenges encountered in replicating LAD-CNN may have contributed to the lower performance of Models A and B. The reduced performance and narrower range of PSNR and SSIM values for Models A and B suggest that these models are less sensitive to changes in noise levels compared to BM3D but also do not achieve the same level of denoising effectiveness.

## IV. DISCUSSION

The first attempt to replicate the LAD-CNN model exactly as per the information given in the article revealed significant challenges that underscored the difficulties of reproducing deep learning architectures without detailed documentation. To address these issues, an alternative CNN-DAE model inspired by the LAD-CNN design was implemented but encountered similar limitations. The fetal head ultrasound dataset was processed through the state-of-the-art BM3D model for comparative analysis.

### *A. Challenges with the Exact LAD-CNN Replication*

The original LAD-CNN paper exhibited several critical shortcomings that hindered its implementation and replication. Notable inconsistencies between textual descriptions and accompanying figures created ambiguity, particularly in detailing the integration of the channel-attention mechanism within the down-sampling and up-sampling stages. Key implementation steps were mentioned in the text but lacked corresponding visual support, leading to sizing errors during pooling and upsampling, and mismatches during replication efforts. The paper introduced lightweight attention mechanisms, such as the channel-attention (LCA) and large-kernel-attention (LLA) blocks but failed to adequately justify their selection. It did not compare these mechanisms with alternatives like spatial attention or transformer-based approaches, nor did it provide theoretical grounding for embedding these mechanisms in the down-sampling and up-sampling stages. The lack of clarity on how these additions improve denoising performance compared to benchmarks like CNN-DAE further weakened the justification for their use. Implementation challenges were exacerbated by insufficient details on critical parameters, such as training strategies, hyperparameter tuning, filter number, and loss function weighting. By assuming familiarity with the underlying methods and failing to provide open-source code or implementation-ready resources, the paper significantly complicated efforts to accurately reconstruct the architecture.

The subjective visual images from the exact replication of code were black, gray, or white boxes suggesting persistent issues in reconstructing meaningful visual data. This could be due to resizing layers added due to the inconsistencies in the article mentioned above.

### *B. Alternative Approach: LAD-CNN Variant Implementation*

Given the challenges faced with replicating the LAD-CNN model and guidance from the supervising professor, an alternative approach was implemented using the baseline code from the CNN-DAE model referenced in the consulted article, along with several key modifications. These included substituting traditional upsampling with subpixel convolution to expand the receptive field during decoding, integrating the GELU activation function to improve gradient flow in the LLA block and enable smoother nonlinear transformations, and adjusting weighting factors in the loss function to optimize the balance between noise reduction and detail preservation. Subpixel convolution enabled Model B to outperform Model A by efficiently upsampling features while preserving spatial details, unlike traditional upsampling methods in Model A, which relied on fixed interpolation and often

produced blurred results. Subpixel convolution operates in the low-resolution space and rearranges learned features into high-resolution outputs, making it computationally efficient and capable of generating sharper images. Additionally, Model B's skip connections avoided sizing errors due to the precise dimension matching facilitated by subpixel convolution, ensuring seamless feature map integration. This combination of high-quality upsampling and well-aligned skip connections allowed Model B to retain fine details and improve overall performance compared to Model A. After these efforts, the output images from the LAD-CNN variant model had a similar resemblance to an ultrasound, suggesting that critical visual information was being preserved better. However, the black/white colors seem to be inverted (Fig 2), the ultrasound seems to be rotated, and there is a substantial amount of over smoothing. This outcome highlights potential fundamental misalignments in the architecture or unresolved limitations in the original methodology.

### C. *Comparison between the alternative and exact LAD_CNN implementations( Model B vs. Model A)*

Subjective images and objective metrics (PSNR/SSIM) were of better quality in Model B. However, the number of parameters for Model B (795,393) is much higher than Model A (504,793) and higher than the parameters in the article (533,155). This is due to how the number of parameters is counted. Equation 8 shows traditional parameter counting, however, for subpixel convolution $C_{out}$ is multiplied by an upscale factor. to the addition of subpixel convolution which Refer to Figure 2 to see the difference in denoised predictions between the two models.

$$\text{Parameters} = (K_H \times K_W \times C_{\text{in}} \times C_{\text{out}}) + C_{\text{out}} \quad (8)$$

### D. *BM3D Denoising Performance*

The fetal head ultrasound dataset was processed using the BM3D model, a conventional denoising algorithm. BM3D showed measurable improvements in image quality, achieving higher PSNR and SSIM values compared to noisy input images. However, its results demonstrated significant over smoothing, leading to the loss of fine structural details. This reaffirms BM3D's limitations in preserving texture and clarity, further highlighting the need for improved deep-learning-based denoising solutions. Both Figures 1 and 2 display an exponential decay in PSNR and SSIM as the noise level increases.

### E. *Critique LAD-CNN*

The study highlights several challenges presented by the original LAD-CNN paper. The incomplete documentation of layer configurations, mismatched descriptions of activation functions used, pooling operations, and skip connection integration made faithful replication difficult. Ambiguities in textual descriptions, coupled with a lack of visual clarity in figures and the absence of open-source code, further limited reproducibility. Additionally, the insufficient discussion of key design decisions and the lack of comparisons with alternative approaches restricted the ability to validate LAD-CNN's claimed advantages.

### F. *Initial Steps Taken*

Following discussions with the supervising professor, the approach was refined to focus on several key strategies. First, the baseline CNN-DAE architecture was leveraged as a starting point. Next, the impact of specific features, such as subpixel convolution and GELU activation, on the model's denoising performance was evaluated. Additionally, experiments were conducted with individual components, including the Gamma weighting factor, to assess their contributions to the overall effectiveness of the model. Analyzing Figure 1, the peak of the weighting factor versus the PSNR graph is at 0.002, indicating that this value would be ideal for future implementation of the LAD-CNN model.

By analyzing these challenges and limitations, the findings highlight the critical importance of robust documentation, detailed architectural descriptions, and accessible implementation resources for advancing reproducibility and practical feasibility in deep-learning research.

## V. Conclusion

To address the issues of incomplete architectural details and practical implementation challenges, this study attempted to replicate and validate the LAD-CNN model using the fetal head circumference ultrasound dataset. The incomplete documentation of LAD-CNN's architecture led to sizing mismatches and ambiguities, preventing successful replication. An alternative implementation of a CNN-DAE model, which included subpixel convolution and GELU activation, also failed to produce meaningful outputs, resulting in black or gray images.

Comparative analysis using the BM3D model demonstrated its effectiveness in noise reduction but highlighted its limitations in preserving image details, leading to over smoothing. While BM3D remains a reliable traditional method, its performance reinforces the need for advanced deep-learning-based solutions like LAD-CNN to achieve superior denoising performance and enhanced image clarity.

The LAD-CNN architecture possesses theoretical advantages, including faster processing with fewer parameters, the utilization of attention mechanisms, and a lightweight design. However, the findings emphasize the importance of robust documentation—including detailed layer descriptions, hyperparameter specifications, and open-source code—to ensure reproducibility in deep learning research. Additionally, the trade-off between model complexity and real-time usability remains a critical consideration for practical applications.

Future work will focus on addressing these challenges by exploring simplified, well-documented architectures and integrating deep learning models into practical medical applications. The objectives include optimizing the model design to balance performance and usability for real-world applications, investigating alternative attention mechanisms and denoising strategies to enhance image quality while preserving structural details, developing user-friendly implementations to facilitate adoption in clinical workflows, and utilizing model data from LAD-CNN by incorporating parameters such as the weighting factor in subsequent implementations.

By tackling these challenges, the aim is to make preprocessing methods like deep-learning-based denoising accessible to physicians, thereby enhancing diagnostic accuracy. This work underscores the critical importance of transparency, reproducibility, and usability in advancing medical image denoising techniques.

## Acknowledgment

Acknowledgment and thanks to Professor Azarang in the Joint Department of Biomedical Engineering at UNC-Chapel Hill & NC State

## References

[1] Shi L, Di W, Liu J. Ultrasound image denoising autoencoder model based on lightweight attention mechanism. Quant Imaging Med Surg 2024;14(5):3557-3571. doi: 10.21037/qims-23-1654

[2] Van den Heuvel TLA, de Bruijn D, de Korte CL, Ginneken BV. Automated measurement of fetal head circumference using 2D ultrasound images. PLoS One 2018;13:e0200412