Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated Removal of Caliper Annotations from Thyroid Ultrasound Images

Yasaman Sharifi

sharifiy@mums.ac.ir

Mashhad University of Medical Sciences

Morteza Danay Ashgzari

Islamic Azad University, Mashhad

Research Article

Keywords: Ultrasound images, Caliper removal, Artifact removal, Caliper Annotations, Deep learning

Posted Date: April 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8754736/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background:

Ultrasound imaging is widely used owing to its safety, accessibility, and real-time capabilities. However, thyroid ultrasound images frequently contain calipers and measurement lines that obscure anatomy and degrade visual fidelity. Although these overlays are clinically useful during acquisition, their presence in stored images can hinder quantitative analysis and bias the performance of downstream artificial intelligence (AI) models. Compared with artifact reduction in CT and MRI, automated caliper removal in ultrasound remains underexplored.

Methods:

We designed an attention-enhanced Residual U-Net for automated caliper removal, formulated as an image-to-image translation task using paired thyroid ultrasound images acquired before and after caliper placement. The network reconstructs caliper-free images while preserving diagnostically relevant anatomy by combining residual learning with attention-gated skip connections that selectively suppress caliper-related features. A composite loss based on mean squared error and the structural similarity index (SSIM) balances intensity fidelity and structural preservation. The proposed method was compared with multiple convolutional and transformer-based architectures using standard image-quality metrics and qualitative expert review.

Results:

The proposed model consistently outperformed competing architectures, achieving higher SSIM, PSNR, Pearson correlation coefficient, and edge-based F1 scores, together with reduced MSE, MAE, and RMSE. Visual inspection confirmed effective removal of caliper annotations while preserving lesion boundaries, tissue textures, and anatomical continuity, with no observable loss of diagnostically relevant information.

Conclusions:

The proposed approach provides a reliable and computationally efficient solution for automated caliper removal in thyroid ultrasound images. By improving dataset quality and image realism, it enhances image usability for clinical review and strengthens the robustness of AI-driven ultrasound analysis.

1. Introduction

Medical imaging plays a fundamental role in modern clinical practice, supporting disease diagnosis, treatment planning, image-guided interventions, and longitudinal patient monitoring. With the increasing adoption of artificial intelligence (AI) in medical image analysis, the quality and realism of input images

have become critically important. Even minor non-anatomical artifacts can bias feature extraction, impair model generalization, and reduce the reliability and safety of AI-assisted diagnostic systems(1, 2).

Ultrasound imaging is widely used in clinical settings due to its real-time acquisition, low cost, portability, and absence of ionizing radiation. It is particularly valuable for the assessment of soft tissues, including thyroid nodules. However, ultrasound images present unique challenges for automated analysis because of speckle noise, low contrast, operator dependency, and inter-device variability. These challenges have motivated extensive research into deep learning-based methods for ultrasound image enhancement and analysis (3–6).

In addition to inherent image degradation, ultrasound images frequently contain non-anatomical overlays such as calipers, measurement lines, arrows, and textual annotations. While these annotations are clinically useful during image acquisition for lesion measurement and documentation, their presence in stored images can obscure anatomical structures, disrupt local texture patterns, and introduce artificial edges that do not correspond to tissue boundaries(7). For AI-based systems, such overlays violate image realism and may lead to spurious correlations, misinterpretation of annotation edges as anatomical structures, and reduced robustness across datasets(1, 2, 7, 8).

This issue is particularly relevant in retrospective studies and real-world clinical datasets, where access to original, unannotated ultrasound images is often unavailable. Unlike acquisition-related artifacts such as noise or motion blur, caliper annotations exhibit substantial variability in shape, size, orientation, color, and placement, and frequently overlap clinically relevant regions. These characteristics make automated caliper removal a challenging and distinct problem, especially in ultrasound imaging, where preservation of fine structural details is essential for both clinical interpretation and quantitative analysis(4).

Reliable removal of caliper annotations has important clinical and translational implications. For human observers, overlays may obscure lesion margins or internal textures, while for AI algorithms they can distort quantitative features, degrade segmentation performance, and reduce diagnostic confidence. Effective caliper removal, without introducing hallucinated structures or compromising anatomical fidelity, is therefore critical for dataset harmonization, robust AI model training, and safe deployment of automated ultrasound analysis pipelines(1, 7, 9, 10).

The objective of caliper removal is to reconstruct clean, artifact-free images that faithfully preserve underlying anatomy and diagnostically relevant structures. Traditional image processing and physics-based approaches are generally ill-suited to this task, as artificial overlays lack a consistent generative model and exhibit heterogeneous visual patterns. Consequently, data-driven deep learning approaches have emerged as an effective solution for medical image restoration and artifact suppression tasks(11–13).

Although convolutional neural networks and generative models have demonstrated strong performance in artifact removal for modalities such as CT and MRI (12, 13), relatively few studies have directly addressed the problem of annotation or caliper removal in ultrasound imaging. Existing ultrasound-

focused work remains limited and often relies on self-supervised or weakly supervised strategies, highlighting the need for reliable, supervised deep learning frameworks tailored to this task (7, 14).

In this study, we propose an automated deep learning-based approach for removing caliper annotations and measurement lines from thyroid ultrasound images while preserving diagnostically relevant anatomical structures. By formulating the task as an image-to-image translation problem, the proposed method reconstructs clean ultrasound images that maintain structural fidelity, edge integrity, and clinical realism. The approach emphasizes computational efficiency, training stability, and suitability for limited datasets, with the ultimate goal of improving dataset quality, enhancing image usability for clinical review, and enabling more accurate and robust AI-driven ultrasound analysis.

2. Methods

2.1. Data Preprocessing

The input data consisted of diagnostic regions of interest (ROIs) extracted from a subset of images from the publicly available Thyroid Ultrasound-Image Dataset (5). We included only cases for which both the original clean images (prior to caliper placement) and the corresponding physician-annotated images were available. This allowed us to obtain paired images with and without annotations, providing approximate pixel-level correspondence between input and ground truth while preserving the underlying anatomy without post-hoc manual editing.

To increase data diversity and improve model generalization, we combined real clinical annotations with synthetically generated caliper overlays. Synthetic calipers were added offline exclusively to non-annotated images and were designed to closely mimic real clinical annotations, similar in spirit to recent work on ultrasound annotation removal using self-supervised learning (7). Specifically, dashed line patterns with variable spacing were used, with plus (+) or cross (×) markers placed at both ends. Calipers were rendered in white, yellow, or green and randomly positioned within the diagnostic region. The appearance and spatial distribution of synthetic calipers were visually compared with real annotations to ensure realistic representation. This procedure produced up to nine caliper-augmented variants per original clean image, which were stored as additional training samples.

All images were resized to 256 × 256 pixels using bilinear (INTER_LINEAR) interpolation. This interpolation strategy was selected to preserve intensity consistency in speckle-dominated ultrasound images and to avoid introducing artificial high-frequency artifacts during resampling. Pixel intensities were normalized to the range [0,1], and all images were converted to single-channel grayscale before being fed into the network.

2.2. Network Architecture

The proposed method is based on a Residual U-Net architecture augmented with attention gates applied exclusively to the skip connections. The network follows a symmetric encoder–decoder structure similar

to U-Net (15), designed to remove caliper annotations from thyroid ultrasound images while preserving fine anatomical details.

The encoder comprises four resolution levels, each consisting of an initial 3 × 3 convolution followed by a residual block. Each residual block includes two 3 × 3 convolutional layers with batch normalization and an identity shortcut, inspired by residual learning in ResNet (16). When the number of channels changes, a 1 × 1 convolution is applied to the shortcut to match dimensions. Residual connections facilitate stable gradient propagation and effective feature reuse, which is particularly important for preserving subtle structural information in ultrasound images. Spatial downsampling between encoder stages is performed using 2 × 2 max-pooling.

At the deepest level of the network, the bottleneck captures high-level semantic representations of the input image through a 3 × 3 convolution and a residual block, followed by a dropout layer (rate = 0.3) to mitigate overfitting and improve generalization.

The decoder mirrors the encoder structure using 2 × 2 transposed convolutions for upsampling. At each decoder stage, feature maps from the corresponding encoder level are integrated via skip connections to recover spatial detail lost during downsampling, as in standard U-Net (15). To selectively filter encoder features before fusion, attention gates are applied to all skip connections, following the attention U-Net paradigm (17). Each attention gate receives the encoder feature map and the corresponding decoder feature map as inputs and computes an attention mask through a sequence of 1 × 1 convolutions, element-wise addition, a nonlinear activation, and a final 1 × 1 convolution followed by a sigmoid. This mechanism suppresses irrelevant background features and emphasizes spatial regions associated with caliper annotations. The refined encoder features are then concatenated with the decoder features and further processed using residual blocks.

The final output layer consists of a 1 × 1 convolution followed by a sigmoid activation, producing a single-channel grayscale image representing the caliper-free ultrasound output.

Attention mechanisms are applied only to skip connections rather than encoder or bottleneck layers. This design enables effective filtering of transferred spatial features without substantially increasing model complexity or computational cost. In contrast to transformer-based architectures, such as Swin-UNet (18), which typically require larger datasets and higher memory consumption, the proposed attention-enhanced Residual U-Net offers an efficient and stable solution suitable for limited medical imaging datasets. A detailed schematic of the architecture is provided in Additional file 1, and the specific layer configurations and filter sizes are summarized in Table 1.

2.3. Loss Function

The network was trained using a combined loss function that balances pixel-wise accuracy and structural similarity:

$$L_{\text{combined}} = \alpha \cdot \text{MSE} + (1 - \alpha) \cdot (1 - \text{SSIM}),$$

where MSE denotes the mean squared error between the predicted and ground truth images, and SSIM denotes the structural similarity index (19). The weighting factor α was empirically set to 0.5 based on preliminary experiments, providing a balanced trade-off between intensity fidelity and structural preservation

2.4. Training Setup

The network was trained on paired thyroid ultrasound images with and without caliper annotations. The dataset was split at the patient level into training (80%), validation (10%), and test (10%) sets to avoid information leakage between patients. Each input image was resized to 256 × 256 pixels while preserving the full diagnostic region.

To enhance generalization, on-the-fly data augmentation was applied to the input images containing calipers. Augmentations included horizontal and vertical flips, small in-plane rotations (± 12°), random translations (up to 8 pixels), mild isotropic scaling (zoom factors between 0.95 and 1.05), brightness adjustments (multiplicative factors between 0.85 and 1.15), and additive Gaussian noise with a small standard deviation. Geometric transformations were applied identically to both input and ground truth images to preserve perfect spatial alignment, whereas photometric and noise perturbations were applied only to the input images with calipers. This strategy is consistent with best practices in medical image augmentation for deep learning (20).

Training was performed using the Adam optimizer (21) with default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$), a batch size of 8, and a maximum of 50 epochs. Early stopping with a patience of 10 epochs based on validation loss was employed, and the model achieving the lowest validation loss was selected for final evaluation to prevent overfitting.

2.5. Evaluation Metrics

Model performance was evaluated using multiple quantitative metrics assessing both pixel-level accuracy and structural fidelity. These included the structural similarity index (SSIM) (19), peak signal-to-noise ratio (PSNR), mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE).

Edge preservation was assessed using the Edge-F1 score, which evaluates boundary fidelity after caliper removal. Edge-F1 scores were computed by extracting edges from both predicted and ground truth images using the Canny edge detector (22) with fixed thresholds (100 and 200), binarizing the resulting edge maps, and computing the F1-score between them. In addition, the Pearson correlation coefficient (PCC) was computed to measure linear correlation between predicted and ground truth intensities, reflecting overall intensity consistency.

All metrics were computed on a per-image basis and aggregated over the test set to provide robust evaluation across different patients and imaging conditions

2.6. Ablation experiments

To systematically investigate the impact of architectural components and training choices, a series of Ablation experiments were conducted. Baseline models included a simple convolutional autoencoder, standard U-Net (15), U-Net++ (23), Attention U-Net (17), Swin-UNet (18), and a Residual U-Net variant combining U-Net and residual blocks (16). For comparison with transformer-based designs, a lightweight Swin-UNet was employed, with input images resized to 128 × 128 pixels due to GPU memory constraints and attention applied only at the bottleneck layers. This resolution mismatch represents a known limitation and is further discussed in Section X.

After identifying Residual U-Net as the strongest baseline, additional experiments examined the effects of enabling deep supervision, introducing attention gates on skip connections (17), and modifying the loss function (e.g., MSE-only vs. combined MSE–SSIM). The influence of different interpolation strategies during image resizing (bilinear vs. nearest-neighbor and bicubic) was also explored. These controlled experiments enabled isolation of the contribution of each design choice to overall image fidelity and edge preservation.

All experiments were repeated using three different random seeds, and results were reported as mean ± standard deviation. The network was implemented in TensorFlow (Google, Mountain View, CA, USA) and trained on an NVIDIA GeForce RTX 3060 GPU. All experiments were conducted using Python 3.10 in the Spyder environment.

Table 1
Layer and filter configuration of the proposed network

Stage	Operation	Filters	Output Size	Notes
Input	—	—	256×256×1	Grayscale ultrasound image
Encoder L1	Conv 3×3 + ResBlock 3×3	32	256×256×32	MaxPooling 2×2 ↓
Encoder L2	Conv 3×3 + ResBlock 3×3	64	128×128×64	MaxPooling 2×2↓
Encoder L3	Conv 3×3 + ResBlock 3×3	128	64×64×128	MaxPooling 2×2↓
Encoder L4	Conv 3×3 + ResBlock 3×3	256	32×32×256	MaxPooling 2×2↓
Bottleneck	Conv 3×3 + ResBlock 3×3 + Dropout	512	16×16×512	High-level semantic features
Decoder L3	Transposed Conv 2×2 + Attention (skip) + ResBlock 3×3	256	32×32×256	Skip connection from Encoder L4 (attention gate applied)
Decoder L2	Transposed Conv 2×2 + Attention (skip) + ResBlock 3×3	128	64×64×128	Skip connection from Encoder L3 (attention gate applied)
Decoder L1	Transposed Conv 2×2 + Attention (skip) + ResBlock 3×3	64	128×128×64	Skip connection from Encoder L2 (attention gate applied)
Decoder L0	Transposed Conv 2×2 + Attention (skip) + ResBlock 3×3	32	256×256×32	Skip connection from Encoder L1 (attention gate applied)
Output	1×1 Conv + Sigmoid	1	256×256×1	Caliper-free ultrasound

3. Results

3.1 Quantitative evaluation

The performance of the proposed method was quantitatively evaluated on the held-out test set and compared against several baseline and advanced architectures using multiple complementary metrics, including SSIM, PSNR, MSE, MAE, RMSE, PCC, and Edge-F1. These metrics collectively assess global image fidelity, pixel-wise reconstruction accuracy, structural similarity, and edge preservation. All values were computed on a per-image basis and then averaged over the test set and three random initializations. Quantitative results for all architectures are summarized in Table 2.

Among all tested architectures, the proposed Residual U-Net with attention-gated skip connections achieved the best overall performance across nearly all evaluation metrics. In particular, it obtained the

highest SSIM and PSNR values, indicating superior structural similarity and reconstruction fidelity, while simultaneously yielding the lowest error-based metrics (MSE, MAE, and RMSE). High PCC values further demonstrate strong consistency between predicted images and the corresponding ground truth in terms of intensity distribution. In addition, the proposed model maintained strong edge preservation as measured by the Edge-F1 score, confirming its ability to retain fine anatomical boundaries after caliper removal.

Table 2. Quantitative comparison of different architectures for automated caliper removal in thyroid ultrasound images

	SSIM	PSNR	MSE	MAE	RMSE	PCC	EdgeF1
SC-AE*	0.9107 ± 0.0060	34.5480 ± 0.2959	0.0004 ± 0.0000	0.0131 ± 0.0004	0.0192 ± 0.0007	0.9927 ± 0.0005	0.7012 ± 0.0132
U-Net	0.9955 ± 0.0009	39.3775 ± 1.6537	0.0002 ± 0.0001	0.0067 ± 0.0016	0.0123 ± 0.0020	0.9963 ± 0.0014	0.9338 ± 0.0036
U-Net++	0.9937 ± 0.0007	38.5787 ± 0.8477	0.0002 ± 0.0001	0.0071 ± 0.0008	0.0131 ± 0.0017	0.9956 ± 0.0021	0.9293 ± 0.0022
U-Net++ (D-S)	0.9951 ± 0.0002	38.4787 ± 1.5231	0.0002 ± 0.0001	0.0087 ± 0.0017	0.0125 ± 0.0021	0.9985 ± 0.0002	0.9292 ± 0.0054
Attention U-Net	0.9936 ± 0.0005	37.3434 ± 1.6710	0.0003 ± 0.0001	0.0084 ± 0.0028	0.0154 ± 0.0026	0.9949 ± 0.0001	0.9224 ± 0.0041
Residual U-Net	0.9951 ± 0.0008	42.0946 ± 2.4032	0.0001 ± 0.0001	0.0054 ± 0.0021	0.0091 ± 0.0022	0.9979 ± 0.0010	0.9413 ± 0.0022
Swin-UNet	0.9889 ± 0.0055	33.6205 ± 1.7756	0.0007 ± 0.0003	0.0097 ± 0.0035	0.0232 ± 0.0047	0.9870 ± 0.0047	0.9226 ± 0.0202

SC-AE: Simple Convolutional-Autoencoder; D-S: deep supervision

3.2 Qualitative assessment

To complement the quantitative evaluation, a qualitative visual assessment was performed by comparing input ultrasound images containing caliper annotations, the corresponding ground truth images without calipers, and the model predictions. Representative examples are shown in Fig. 1. As illustrated, the proposed method effectively removes caliper annotations and measurement lines while preserving underlying diagnostic tissue structures, without introducing visible artifacts, false structures, or noticeable blurring.

Further qualitative analyses were conducted using absolute difference maps and pixel-wise error histograms, presented in Fig. 2. The absolute difference maps demonstrate that reconstruction errors are largely confined to regions originally occupied by caliper annotations, while surrounding anatomical

structures remain well preserved. Error histograms further confirm a narrow error distribution, indicating high pixel-wise accuracy across the test set.

Overall, the quantitative and qualitative results consistently demonstrate that the proposed attention-enhanced Residual U-Net provides robust and accurate caliper removal in thyroid ultrasound images, achieving a favorable balance between global image quality, structural fidelity, and edge preservation.

3.3 Ablation results

To analyze the contribution of individual architectural components and training choices, a series of Ablation experiments was conducted. The impact of residual connections, deep supervision, attention gates, loss function, and interpolation strategy is analyzed in the ablation study in Table 3. Starting from baseline architectures, progressively enhanced models were evaluated by introducing residual connections, deep supervision, attention mechanisms on skip connections, alternative loss formulations, and different interpolation strategies.

The Residual U-Net baseline achieved strong performance across all evaluation metrics compared with conventional U-Net variants and transformer-based models. Incorporating deep supervision did not yield consistent improvements across quantitative metrics. In contrast, adding attention gates to the skip connections led to improved performance in most global image quality metrics, including SSIM, PSNR, MAE, RMSE, and PCC, while also enhancing edge preservation as measured by the Edge-F1 score. The use of an edge-aware loss formulation primarily improved edge-related performance, whereas variations in interpolation strategies influenced the overall stability of reconstruction quality across metrics.

Table 3. Ablation study of Residual U-Net variants for caliper removal

	SSIM	PSNR	MSE	MAE	RMSE	PCC	EdgeF1
Res U-Net (baseline)	0.9951 ± 0.0008	42.0946 ± 2.4032	0.0001 ± 0.0001	0.0054 ± 0.0021	0.0091 ± 0.0022	0.9979 ± 0.0010	0.9413 ± 0.0022
Res U-Net + D-S	0.9952 ± 0.0006	39.1331 ± 0.5943	0.0002 ± 0.0000	0.0080 ± 0.0006	0.0119 ± 0.0009	0.9977 ± 0.0002	0.9357 ± 0.0078
Res U-Net + S-A (proposed)	0.9963 ± 0.0008	42.7052 ± 2.2088	0.0001 ± 0.0001	0.0045 ± 0.0015	0.0085 ± 0.0026	0.9978 ± 0.0015	0.9444 ± 0.0096
Res U-Net + S-A + edge-aware loss	0.9956 ± 0.0006	41.2880 ± 2.6371	0.0002 ± 0.0000	0.0058 ± 0.0025	0.0108 ± 0.0019	0.9964 ± 0.0006	0.9486 ± 0.0062
Res U-Net + S-A + bicubic-I	0.9924 ± 0.0003	38.8006 ± 1.5842	0.0003 ± 0.0000	0.0074 ± 0.0025	0.0134 ± 0.0022	0.9959 ± 0.0005	0.8747 ± 0.0046

The proposed model corresponds to the “Residual U-Net + skip attention (bilinear, MSE+SSIM)” configuration. Res U-Net: Residual U-Net S-A: skip attention; D-S: deep supervision; bicubic-I: bicubic interpolation

4. Discussion

4.1. Main findings

The proposed Residual U-Net with attention-gated skip connections consistently outperformed alternative architectures, including standard U-Net, U-Net++, Attention U-Net, the Residual U-Net baseline, and the transformer-based Swin-UNet, across a comprehensive set of quantitative metrics. As shown in Table 3, our model achieved the highest SSIM, PSNR, and PCC values, indicating superior global image fidelity, while also maintaining strong edge preservation as measured by the Edge-F1 score. Error-based metrics (MSE, MAE, and RMSE) were lower than those of all other models, reflecting more accurate pixel-wise reconstruction. Qualitative evaluations further support these findings, absolute difference maps and pixel-wise error histograms (Fig. 2) show that reconstruction errors are largely confined to regions originally occupied by caliper annotations, and representative examples (Fig. 1) confirm effective removal of measurement lines without compromising diagnostically relevant tissue structures.

Residual connections within the encoder–decoder architecture facilitated effective feature reuse and stable optimization, enabling the preservation of subtle diagnostic details. To selectively filter encoder features before fusion with the decoder, attention gates were applied exclusively to the skip connections rather than employing global self-attention. This design preserves high-frequency anatomical edges

while suppressing caliper-related artifacts, thereby avoiding the over-smoothing effects sometimes observed with more diffuse transformer-based attention mechanisms. By emphasizing diagnostically relevant features and attenuating irrelevant background information, skip attention enhances both local and global reconstruction quality, which is reflected in the improved SSIM, PSNR, and Edge-F1 scores.

Deep supervision was introduced to provide additional guidance to the decoder via intermediate outputs. However, the ablation study (Table 3) indicates that, for this caliper removal task, deep supervision does not consistently improve performance and can induce mild over-smoothing. Because intermediate decoder representations lack fully refined edge information, enforcing supervision at these scales may encourage suppression of high-frequency details, leading to slightly blurred outputs and modest degradation in edge-related metrics.

An edge-aware loss function was also investigated to further improve boundary preservation. Although this loss improved the Edge-F1 score relative to the baseline combined loss, it did not consistently enhance global image quality metrics such as SSIM and PSNR and, in some cases, slightly reduced intensity fidelity. These findings suggest that over-emphasizing edge information may bias optimization toward local gradients at the expense of overall structural and intensity consistency. Consequently, the combined MSE–SSIM loss was adopted as the primary objective, as it provides a more balanced trade-off between global fidelity and edge preservation.

Despite the strong capability of transformer-based architectures in modeling long-range dependencies, the lightweight Swin-UNet used in this work underperformed compared with convolutional Residual U-Net variants. This outcome can be attributed to three main factors: (i) the need to downsample input images to 128 × 128 pixels due to GPU memory constraints, (ii) restricting attention mechanisms to low-resolution bottleneck layers, and (iii) the intrinsic nature of the task, which relies heavily on precise reconstruction of fine local structures and speckle patterns, for which the inductive bias of convolutional networks is particularly effective. Overall, these results suggest that, in the context of caliper removal from ultrasound images and under limited data conditions, convolutional architectures such as Residual U-Net with skip attention may be better suited than lightweight transformer-based models.

In addition, the ablation study revealed that higher-order interpolation methods, such as bicubic interpolation (INTER_CUBIC), tend to amplify speckle and noise patterns in ultrasound images, negatively affecting reconstruction quality and edge metrics. This observation underscores the importance of selecting preprocessing operations that are tailored to the physical characteristics of ultrasound data, rather than relying on generic image-processing defaults.

4.2. Comparison with previous work

Manual overlays and annotations in ultrasound—such as calipers, arrows, and text—have only recently begun to receive focused attention. Early work on thyroid ultrasound by Narayan et al. and Sun et al. relied on rule-based artifact detection followed by classical inpainting to remove manually introduced measurement tools while preserving speckle texture and anatomy (24, 25). For example, Sun et al.

proposed an edge-connection algorithm to segment artifacts and a Criminisi image restoration step to fill the masked regions using surrounding texture (25). These approaches demonstrated that explicit artifact masks combined with texture synthesis can effectively restore images, but they depend strongly on hand-crafted features and are sensitive to variability in overlay style, scanner parameters, and acquisition protocols.

More recent studies have adopted learning-based strategies. Zhang et al. proposed a self-supervised Noise2Noise framework for ultrasonic annotation removal, in which common annotations are extracted from archived images and randomly superimposed onto a small set of clean ultrasound images to synthesize corrupted pairs; the network is then trained to map one corrupted image to another while preserving the same underlying clean content (7). Li et al. introduced a marker-mask inpainting (MMI) pipeline within a curated thyroid ultrasound database, combining an automatic marker detection module that generates binary masks of artificial markers with an inpainting module that restores marker-free images used to train more robust CAD systems (26). Complementarily, Mikolaj et al. focused on removing confounding information from fetal ultrasound by masking annotated regions and applying model-level regularization so that deep networks cannot rely on non-diagnostic cues (e.g., embedded text or calipers), thereby improving generalization without explicitly reconstructing clean images (27).

Beyond ultrasound, several works have addressed artificial markers and acquisition-related artifacts in other modalities using inpainting and deep restoration networks. These include classical artifact removal in dermoscopy images using edge detection and polygon-based inpainting (28), blind inpainting of artificial markers without explicit masks (29), conditional inpainting for motion and interpolation artifacts in 4D CT (30), transformer-based MRI artifact removal (31), and GAN-based inpainting of metal artifact regions in brain MRI (32). Related studies on low-dose CT denoising and artifact suppression have also shown the effectiveness of deep residual and wavelet-domain CNNs for restoring noisy or artifact-corrupted images (33, 34). From a more general perspective, many of these methods can be viewed as image-to-image translation or inpainting problems, conceptually related to conditional GAN frameworks such as Pix2Pix (9) and hybrid CNN–Transformer architectures such as TransUNet (35).

In summary, prior work on ultrasound annotation and caliper removal includes rule-based detection plus inpainting (24, 25), self-supervised Noise2Noise annotation suppression (7), marker-mask inpainting and dataset construction (26), and confounder masking at the model level (27). More general frameworks for artificial marker removal and artifact inpainting across modalities rely on explicit masks, artifact simulators, or complex adversarial and transformer-based training (28–34). In contrast, the present study focuses specifically on caliper and measurement line removal in thyroid ultrasound using a supervised, paired image-to-image formulation based on an attention-enhanced Residual U-Net. The proposed model simultaneously removes calipers and reconstructs anatomically faithful, caliper-free images without hand-crafted rules or explicit artifact masks, with a particular emphasis on preserving lesion boundaries, speckle patterns, and overall anatomical fidelity.

Compared to prior work, our study is, to our knowledge, the first to address supervised caliper and measurement-line removal in thyroid ultrasound using paired pre- and post-caliper acquisitions and an attention-guided Residual U-Net, with a systematic comparison against multiple encoder–decoder and transformer-based baselines and explicit analysis of edge preservation and image quality.

4.3. Strengths and limitations

The main strengths of the proposed approach include its ability to remove caliper annotations while preserving fine anatomical details and lesion boundaries, resulting in high-fidelity reconstructions. The combined MSE–SSIM loss provides a balanced trade-off between global image quality and structural preservation. Attention-gated skip connections selectively filter irrelevant features and caliper-related artifacts, enhancing both local and global reconstruction performance without incurring the complexity of full transformer-based models. The architecture is computationally efficient and demonstrates robust performance even with a limited number of training subjects. Qualitative analyses corroborate the quantitative results, showing that diagnostically relevant tissue structures remain accurately preserved.

Nonetheless, several limitations should be acknowledged. First, the model was primarily evaluated on thyroid ultrasound images from a specific acquisition protocol; its generalizability to other organs, ultrasound vendors, or imaging modalities remains to be established. Second, deep supervision provided only marginal benefits and, in some configurations, led to over-smoothing, suggesting that its utility may be limited in small datasets focused on fine structural detail. Third, while the edge-aware loss improved edge-related metrics, it slightly reduced global intensity fidelity, reinforcing the need for carefully balanced loss formulations. Fourth, the transformer-based Swin-UNet underperformed in this setting, partly due to limited data and reduced input resolution, so these results should not be taken as definitive evidence against transformer architectures in general. Finally, the proposed approach relies on paired training data (images with and without calipers); fully unpaired, weakly supervised, or self-supervised scenarios were not examined in this study.

4.4. Future work

Future research may extend the proposed methodology to other organ-specific ultrasound datasets (e.g., breast, liver, obstetric imaging) to better assess its generalizability and clinical utility. Self-supervised or semi-supervised pretraining on large collections of unlabeled ultrasound images could further improve feature representations in low-data scenarios. Domain adaptation or harmonization techniques may help adapt the model across different ultrasound devices, imaging presets, and clinical centers. In addition, exploring multi-task learning frameworks—for example, combining caliper removal with lesion segmentation or classification—could strengthen anatomical constraints and improve robustness. Finally, investigating advanced or hybrid loss formulations that jointly optimize global fidelity, edge preservation, and perceptual quality may yield even more accurate and clinically trustworthy artifact removal systems.

5. Conclusion

In this study, we proposed a Residual U-Net architecture enhanced with attention-gated skip connections for removing caliper annotations and measurement lines from thyroid ultrasound images. Using paired images acquired before and after caliper placement, the model was trained in a supervised image-to-image setting to reconstruct caliper-free images while preserving diagnostically relevant tissue structures. Comprehensive quantitative evaluation (SSIM, PSNR, MSE, MAE, RMSE, PCC, Edge-F1) and qualitative visual inspection demonstrated that the proposed method achieves high global image fidelity and strong edge preservation, with reconstruction errors largely confined to regions originally occupied by calipers.

Ablation experiments confirmed the critical contribution of attention-gated skip connections to maintaining structural detail and boundary sharpness, and the overall architecture consistently outperformed alternative models, including standard U-Net, U-Net++, Attention U-Net, and a lightweight Swin-UNet. The proposed approach is computationally efficient and well suited to limited datasets, providing a practical solution for generating clean, caliper-free ultrasound images. By improving dataset quality and anatomical realism, this method has the potential to enhance the performance, robustness, and generalizability of AI-driven diagnostic algorithms for thyroid abnormality detection and related ultrasound-based clinical applications.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of Mashhad University of Medical Sciences (approval number: IR.MUMS.MEDICAL.REC.1403.014) and was conducted in accordance with the Declaration of Helsinki. The requirement for informed consent was waived by the Ethics Committee due to the retrospective use of anonymized publicly available data.

Consent for publication

Not applicable. The manuscript does not contain any individual person’s data. All identifying information in ultrasound images has been removed.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

YS designed the study, developed the methodology, conducted the experiments, analyzed the results, and drafted the manuscript. MDA contributed to data acquisition, interpretation, and manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

The ultrasound images used in this study were collected from the Persian Cohort Center at Mashhad University of Medical Sciences. The authors thank the center for providing access to the data.

References

1. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annual review of biomedical engineering. 2017;19(1):221-48.
2. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, Van Ginneken B, Madabhushi A, et al. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE. 2021;109(5):820-38.
3. Luijten B, Chennakeshava N, Eldar YC, Mischi M, van Sloun RJ. Ultrasound signal processing: From models to deep learning. Ultrasound in medicine & biology. 2023;49(3):677-98.
4. Song K, Feng J, Chen D. A survey on deep learning in medical ultrasound imaging. Frontiers in Physics. 2024;12:1398393.
5. Sharifi Y, Bakhshali MA, Dehghani T, DanaiAshgzari M, Sargolzaei M, Eslami S. Deep learning on ultrasound images of thyroid nodules. Biocybernetics and Biomedical Engineering. 2021;41(2):636-55.
6. Sharifi Y, Danay Ashgzari M, Naseri Z. Amiri Tehranizadeh. TIRADS-Based Artificial Intelligence Systems for Ultrasound Imaging of Thyroid Nodules: A Systematic Review. Austin J Radiol. 2025; 12 (2): 1254. Austin Journal of Radiology.12(2-2025).
7. Zhang Y, Jiang N, Xie Z, Cao J, Teng Y. Ultrasonic image's annotation removal: A self-supervised Noise2Noise approach. arXiv preprint arXiv:230704133. 2023.
8. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60-88.
9. Isola P, Zhu J-Y, Zhou T, Efros AA, editors. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
10. Smistad E, Falch TL, Bozorgi M, Elster AC, Lindseth F. Medical image segmentation on GPUs–A comprehensive review. Medical image analysis. 2015;20(1):1-18.

11. Chobola T, Müller G, Dausmann V, Theileis A, Taucher J, Huisken J, et al., editors. Leveraging classic deconvolution and feature extraction in zero-shot image restoration. Proceedings of the IEEE/CVF International Conference on Computer Vision; 2023.
12. Liao H, Lin W-A, Zhou SK, Luo J. ADN: artifact disentanglement network for unsupervised metal artifact reduction. IEEE Transactions on Medical Imaging. 2019;39(3):634-43.
13. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in x-ray computed tomography. IEEE transactions on medical imaging. 2018;37(6):1370-81.
14. Ye JC, Eldar YC, Unser M. Deep learning for biomedical image reconstruction: Cambridge University Press; 2023.
15. Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention; 2015: Springer.
16. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
17. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.
18. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al., editors. Swin-unet: Unet-like pure transformer for medical image segmentation. European conference on computer vision; 2022: Springer.
19. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing. 2004;13(4):600-12.
20. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. Journal of big data. 2019;6(1):1-48.
21. Kingma DP. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
22. Canny J. A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence. 2009(6):679-98.
23. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J, editors. Unet++: A nested u-net architecture for medical image segmentation. International workshop on deep learning in medical image analysis; 2018: Springer.
24. Narayan NS, Marziliano P, Hobbs CG, editors. Automatic removal of manually induced artefacts in ultrasound images of thyroid gland. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013: IEEE.
25. Sun M, Meng Q, Wang T, Liu T, Zhu Y, Qiu J, et al. Removal of manually induced artifacts in ultrasound images of thyroid nodules based on edge-connection and Criminisi image restoration algorithm. Computer Methods and Programs in Biomedicine. 2021;200:105868.
26. Li X, Fu C, Xu S, Sham C-W. Thyroid ultrasound image database and marker mask inpainting method for research and development. Ultrasound in Medicine & Biology. 2024;50(4):509-19.

27. Mikolaj K, Lin M, Bashir Z, Svendsen MBS, Tolsgaard M, Nymark A, et al. Removing confounding information from fetal ultrasound images. arXiv preprint arXiv:230313918. 2023.
28. Mahmood A, Mahmood H. Artifact Removal from Skin Dermoscopy Imagesto Support Automated Melanoma Diagnosis. Al-Rafidain Engineering Journal. 2015;23(5):22-30.
29. Guo X, Hu W, Ni C, Chai W, Li S, Wang G, editors. Blind Inpainting with Object-Aware Discrimination for Artificial Marker Removal. ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2024: IEEE.
30. Madesta F, Sentker T, Gauer T, Werner R. Deep learning-based conditional inpainting for restoration of artifact-affected 4D CT images. Medical Physics. 2024;51(5):3437-54.
31. der Goten V, Alexander L. Wide Range MRI Artifact Removal with Transformers. arXiv preprint arXiv:221007976. 2022.
32. Xie K, Gao L, Zhang H, Zhang S, Xi Q, Zhang F, et al. GAN-based metal artifacts region inpainting in brain MRI imaging with reflective registration. Medical Physics. 2024;51(3):2066-80.
33. Huang X, Wang J, Tang F, Zhong T, Zhang Y. Metal artifact reduction on cervical CT images by deep residual learning. Biomedical engineering online. 2018;17(1):175.
34. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Medical physics. 2017;44(10):e360-e75.
35. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:210204306. 2021.

Figures

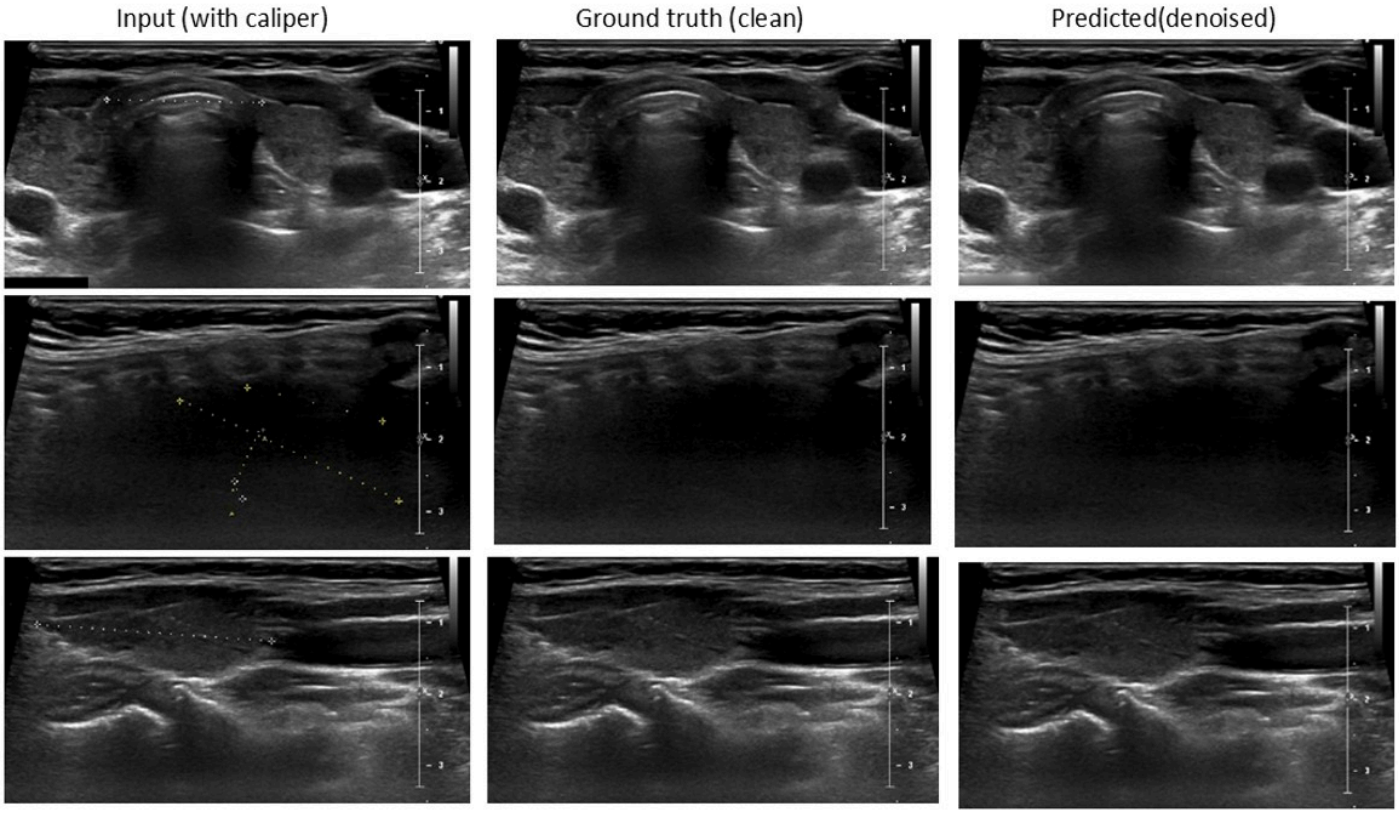


Figure 1

Visual examples of caliper removal in thyroid ultrasound images using the proposed method.

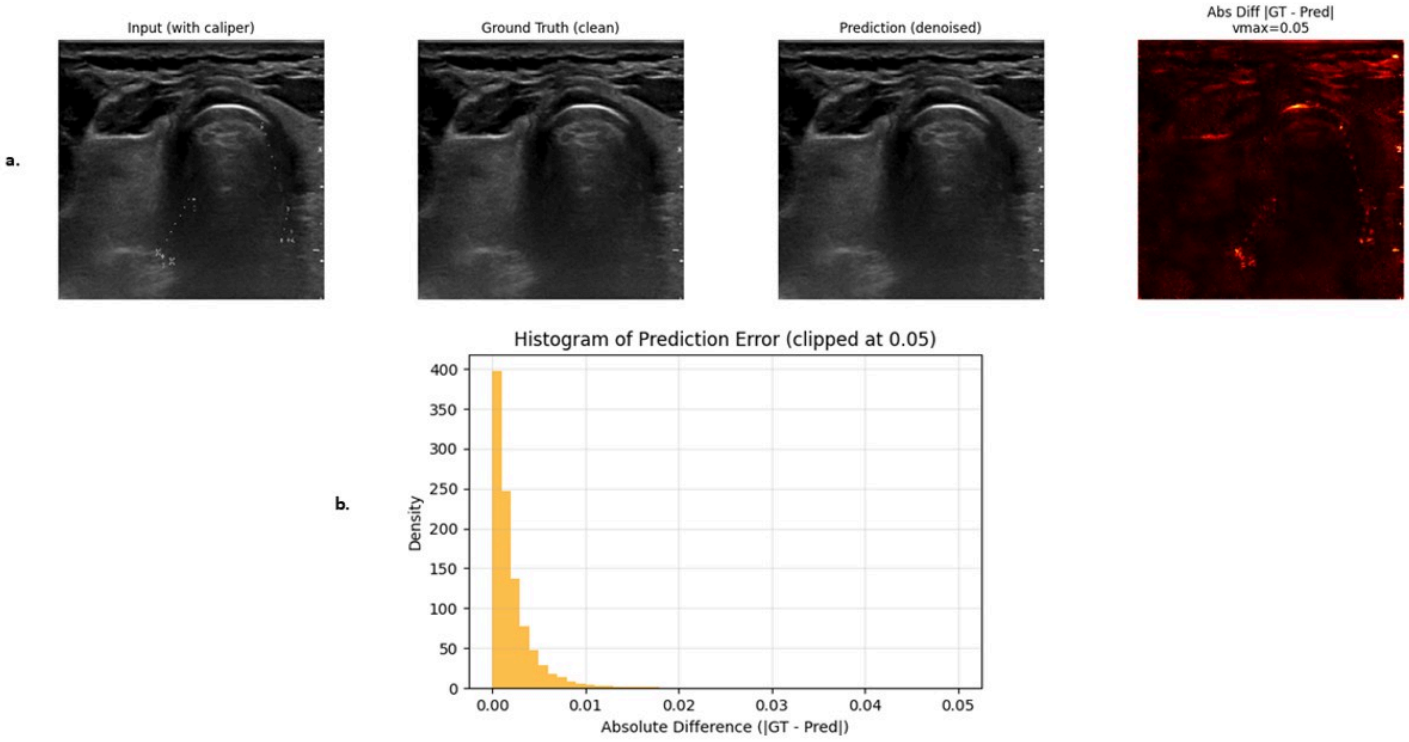


Figure 2

Error analysis of the proposed method**.** (a) Absolute difference maps between predictions and ground truth images. (b) Histogram of pixel-wise prediction errors across the test set

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.OverviewoftheproposedattentionenhancedResidualUNet.png