UltraAugment: Fan-shape and Artifact-based Data Augmentation for 2D Ultrasound Images

Florian Ramakers
KU Leuven
Leuven, Belgium
florian.ramakers@kuleuven.be

Tom Vercauteren
King's College London
London, United Kingdom

Jan Deprest
KU Leuven
Leuven, Belgium

Helena Williams
KU Leuven
Leuven, Belgium

Abstract

Deep learning systems for medical image analysis have shown remarkable performance. However, performance is heavily dependent on the size and diversity of the training data as small datasets might lead to overfitting. Unfortunately, labeled data is often hard to acquire because of the high cost and required medical expertise. Data augmentation is an effective strategy to combat this and has proven to significantly improve model generalisability as it increases the size and diversity of the dataset. However for ultrasound images classic data transformations may not always be appropriate. In this paper we focus on developing data augmentations specifically designed for fan-shaped ultrasound images by simulating artifacts, altering speckle patterns, and adapting conventional techniques to make them fan-shape preserving. We apply the suggested augmentations to two segmentation tasks and demonstrate that the proposed augmentation techniques can improve performance and can remedy the harm caused by there conventional alternatives.

1. Introduction

It is widely known that deep learning system require a vast amount of data to reach acceptable performance. However, medical images are often hard to acquire due to the high cost and the required medical expertise to create ground truth data such as segmentations or classification labels. Since training on small datasets often leads to overfitting, data augmentation is crucial for achieving acceptable generalisability [37]. Conventional data augmentations (e.g. affine transformations, flipping, cropping, zooming, blurring, adding image noise, etc.) have shown to reduce overfitting by exposing the network each time to a slightly different version of the original image [4, 7]. However, for ultrasound (US) images these augmentations may be suboptimal as inappropriate augmentation could result in images that deviate from clinical reality [10, 28]. Application of conventional data techniques may therefore be limited. Previous research involving US images limits data augmentation to rotations [3, 9, 11, 14, 15, 25, 26, 29–32], horizontal flipping [1, 3, 9, 11, 15, 26, 29–32], contrast adjustment [3, 9, 14, 30, 31], cropping [1, 9, 15], elastic deformation [3, 23, 26, 31], grid distortion [3, 9], Gaussian smoothing [9, 14, 31] and additive noise [9, 30, 31]. Unfortunately, there is disagreement in whether these actually simulate anatomical, operator and machine variability. Some argue that rotation should be avoided, since this changes the origin of the US beams [40], while others state that it simulates a difference in pitch of the probe [39]. Furthermore, horizontal flipping, which could be seen as a 180° rotation of the probe [3], may violate probe orientation conventions [40]. Vertical flipping is often avoided as it changes the direction of the beams [40] and might result in unrealistic depiction of acoustic shadows [28]. Contrast and brightness adjustment, which correspond to altering machine settings related to contrast, brightness and time gain control [7, 39], or probe pressure variability [3], are heavily used and show to be highly useful for US [10]. Elastic deformation and grid distortion, while not extensively used, are argued to simulate anatomical variability [3], or patient motion (e.g. due to breathing) [10]. Gaussian or bilateral smoothing is also often used for simulating image degrading due to noise [17] or for simulating bone response to different orientations of the probe [39]. Lastly, additive noise is often deemed less appropriate as noise in US images has a multiplicative nature [13]. Given these connections to anatomical, operator and machine variability, applying conventional

data transformations should be handled with care and with the anatomical area in mind.

Another way to artificially increase dataset size is to generate synthetic data. A popular approach is to use a generative adversarial network (GAN) [27, 35, 39]. However, training a GAN typically requires large amounts of data, which might not be available. Furthermore, the training process of GANs is often difficult and unstable. In addition, the performance boost obtained from synthetic data is often relatively small given the time and resources it consumes [27, 35].

Further, popular techniques such as MixUp [41] or CutMix [38] aim to create new data by combining features from two images. Even though Mixup has been successfully applied to US images [24, 33], it has been shown that combining it with conventional data techniques might negatively impact performance [24]. CutMix creates a new image by copying a small region from one image onto another and adapting the labels appropriately. Given that this may create highly unlikely images CutMix should be applied while keeping the anatomy in mind [12].

Because of the limited applicability of conventional data augmentation methods and complexities of synthetic data generation, one might benefit from developing data augmentations that are designed with the image modality in mind. This allows to make certain assumptions about the data, which enables the use of more powerful and complex transformations. For example in magnetic resonance imaging (MRI) several data augmentation techniques operate in k-space allowing the simulation of motion and noise artifacts [21] or to perform under-sampling to simulate faster MRI sequences [8].

Equivalently to MRI augmentations, augmenting US data could also be performed in another space such as the radio-frequency (RF) space [34]. However, RF data is rarely accessible on commercial machines, therefore, augmentation of the RF space is not possible in most scenarios. A possible solution might be to estimate the RF data [19], but converting it back to high quality B-mode images is a non-trivial task.

Data augmentations specifically for B-mode US images are far less widespread. Tirindelli et al. introduced data augmentations specifically for US based on the simulation of reverberations, deformation, and signal-to-noise ratio alteration [28]. However, all techniques require a bone segmentation, which might not be available in all US images, and is time-consuming to curate. Furthermore, using the data augmentations resulted in a negligible segmentation improvement and a degrading classification accuracy. Singla et al. proposed data augmentations based on simulating acoustic shadows and different machine settings related to zoom and time gain control [22]. Unfortunately, most augmentations relied on statistical assumptions and did not result in realistic images, nor did they add a significant performance boost. Another approach, called *Gaussian Shadow* [23] simulates shadows by darkening a Gaussian region that was randomly sampled. While this significantly improved performance, advertising this as a shadow seems rather farfetched as it's closer to a form of masking instead of simulating a shadow given that the location and shape are not based on the depicted US image features. Xu et al. proposed to simulate acoustic shadowing by copying the shadow of one image to another one [36]. Even though this resulted in a significant performance improvement, the application of this technique seems only valid if the anatomy of the US images is somewhat aligned and depicts the same region. Otherwise, this technique could introduce shadows in unrealistic places within the US image.

More generative approaches such as GANs or variational autoencoders (VAE) have simulated US specific features such as speckle [2] or tissue deformation [6]. However, these typically require large amounts of data to produce acceptable results.

Instead of simulating altered machine settings or artifacts, simulating different probes also seems an effective augmentation technique. In [40] it is shown that by increasing probe diversity, model generalisability can be improved.

In this paper, we propose the *UltraAugment* framework. More specifically, UltraAugment consist of six new data augmentation techniques for 2D fan-shaped US images. We focus on achieving anatomically and physically plausible results by simulating US artifacts, distorting speckle and creating fan-shape preserving versions of conventional cropping, elastic deformation and noise data augmentation techniques. In addition, the proposed augmentations are easily employable for any fan-shaped US images as they do not rely on additional external segmentations or generative models. The impact of the augmentations is analysed by utilising them on a private dataset for 2D external anal sphincter (EAS) segmentation and on the public JNU-IFM [16] dataset for 2D fetal head (FH) and symphysis pubis (SP) segmentation. All proposed data augmentation techniques are shown in Figure 1.

2. Methods

2.1. Eliminating Fan Shapes

Since the fan shape of the US images complicates most of the image transformations involved in the data augmentation techniques, it is removed by warping the images onto a rectangular grid. This is achieved by sampling all scanlines through a polar sweep across the fan angle and arranging them next to each other. After the image is transformed it is warped back such that the original fan shape is restored. For details on the warping operations, we refer to section 5 in the supplementary material. In what follows we denote

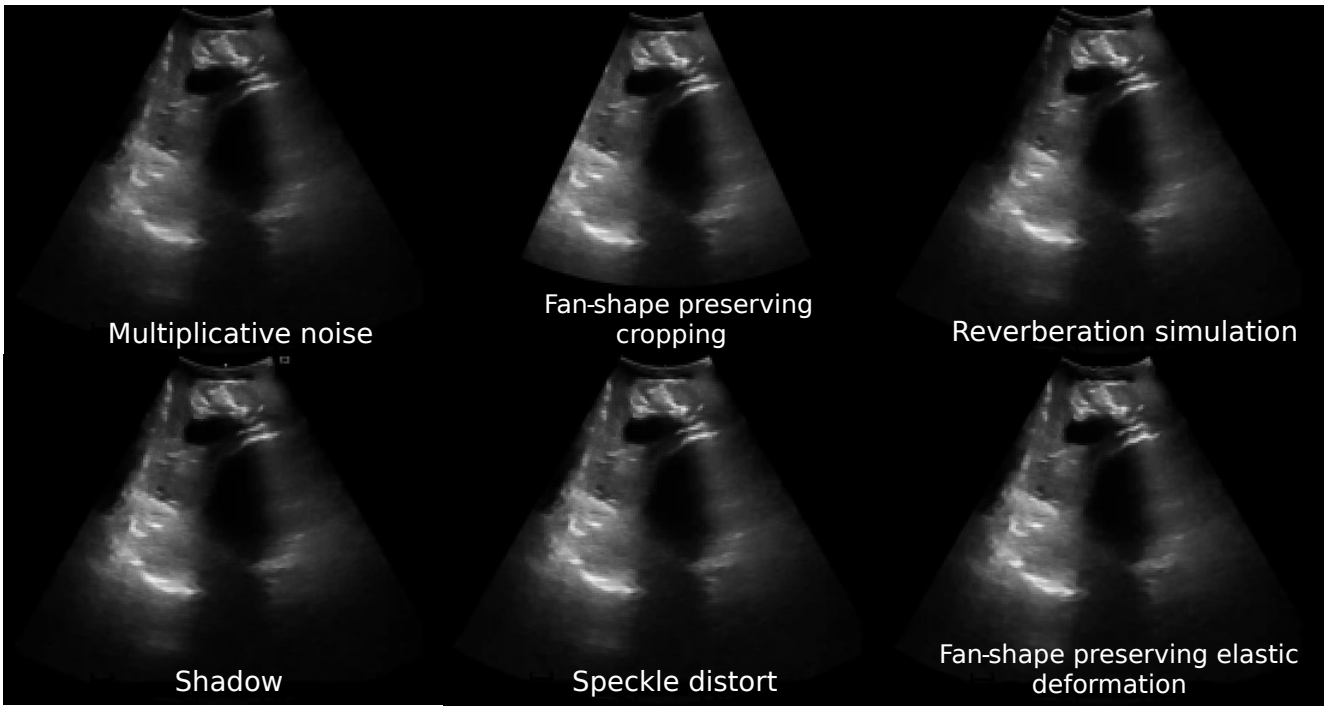

Figure 1. The proposed data augmentations shown on an image of the JNU-IFM dataset.

the warping operation of an image X and its inverse with $\Phi(X)$ and $\Phi(X)^{-1}$ respectively.

2.2. Speckle Distortion

Speckle can be defined as an interference pattern resulting from interfering echoes of a transmitted waveform. While speckle has a direct relationship to tissue composition, there is no consensus regarding a standardized approach for using this information as some consider it to be noise and others as part of the signal [20]. Furthermore, since speckle is partly random due to the unpredictability of scattered waves, it is justifiable to alter the speckle composition of an US image as a form of data augmentation.

To extract the speckle pattern from US images, it is assumed that speckle corresponds to the high frequency components of the US images and can consequently be altered by changing these components. In what follows, $\mathcal{F}_{high}\{X\}$ denotes the components of the Fourier transform of an image X of frequencies higher than frequencies (k_{tx}, k_{ty}) associated to the lowest spatial frequencies of speckle in the x and y directions respectively. For simplicity, k_{tx} and k_{ty} are in reality provided as fraction m_{tx} and m_{ty} respectively and represent a normalized index of the image in Fourier space that corresponds to the lowest speckle frequency. Note that the Fourier transform is also applied to the warped image $\Phi(X)$ as this allows to capture speckle patterns at different depths using the same spatial frequency.

To distort the speckle that is present in an image, the high frequency components are altered in two ways. The magnitude of the components is changed by randomly permuting $\mathcal{F}_{high}\{\Phi(X)\}$ such that the presence of certain frequencies is reduced and others is increased. Second, the phase of each component is altered by adding a randomly sampled phase offset. This is achieved by expressing the Fourier transform in polar form and adding a random phase component:

$$F(k_x, k_y) = \begin{cases} |f(k_x, k_y)| e^{i(\phi + \psi(k_X, k_y))} & \text{if } k_x \geq k_{tx} \text{ and } k_y \geq k_{ty} \\ f(k_x, k_y) & \text{otherwise,} \end{cases} \tag{1}$$

where f is the Fourier transform of the warped image, namely $f = \mathcal{F}_{high}\{\Phi(X)\}$, $|\cdot|$ denotes the magnitude, ϕ is the angle of the Fourier transform and $\psi(k_x, k_y) \sim \text{Uniform}(\psi_{min}, \psi_{max})$ denotes the phase offset.

2.3. Reverberation Simulation

Reverberations arise when an US beam bounces back and forth between two parallel reflectors. The transducer mistakenly interprets this as waves coming from deeper structures, which results in visual repetitions of the same anatomical region. Previous work already tried to simulate reverberations based on replicating an image patch containing bone [28]. However, the need for a bone segmentation makes this approach less widely applicable. As a result, we propose to sample a patch close to the US probe and repli-

cate this patch n times in opposite direction from the probe. The patch location ($p_x \in [x_{min}, x_{max}], p_y \in [y_{min}, y_{max}]$) and size ($p_w \in [w_{min}, w_{max}], p_h \in [h_{min}, h_{max}]$), specified in normalized coordinates, are randomly sampled according to the provided ranges. To mask the obvious transitions between the replicated patches and the image, the image is linearly mixed in using a sinusoidal pattern λ with a period equal to the width of the replicated patch and that decays to zero near the border. The augmented image is obtained by

$$X_{patch}(u,v) = \lambda(u,v)R(u,v) + (1-\lambda(u,v))\,\Phi(X)(u,v), \tag{2}$$

where (u,v) is a coordinate in the image space, R the reverberation patch and $\lambda(u,v)$ the smoothed sinusoidal pattern. The pattern $\lambda(u,v)$ consists of two parts:

$$\lambda(u,v) = \left(\frac{1}{2}\sin\left(\frac{2\pi}{H}\right) + \frac{1}{2}\right)(G_{K,\sigma} * Z_m)(u,v). \tag{3}$$

In this equation $G_{K,\sigma}$ is a $K \times K$ Gaussian kernel with standard deviation σ and Z_m is defined as

$$Z_m(u,v) = \begin{cases} 1 & \text{if } x+m < u \leq x+W-m \\ & \text{and } y+m < v \leq y+nH-m \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $m \in \mathbb{N}$, and (x, y, W, H) defines the bounding box of the replicated patch. We take $K = 11$, $\sigma = 2.0$ and $m = 2$ as this visually results in the most realistic reverberations.

Since reverberations typically cause shadowing behind them, the intensities of the pixels located behind the patch are reduced according to:

$$X_{new}(u,v) = X_{patch}(u,v) - \beta\kappa(u,v)X_{patch}(u,v), \tag{5}$$

where $\beta < 1$ is the reduction factor and $\kappa(u,v)$ a mask that defines the region affected by the shadow. This affected region has the same width as the replicated patches and has a height expressed as a fraction d of the scan line length. To make sure the shadow blends well with the image, $\kappa(u,v)$ decays to zero near the border of the shadow. Similarly for the smoothing of the shadow caused by the reverberations, we define $\kappa(u,v)$ as

$$\kappa(u,v) = (G_{K',\sigma'} * Z_{m'})(u,v), \tag{6}$$

with $K' = 51$, $\sigma' = 25$ and $m' = 20$, as this results in shadows that look similar to real examples.

The entire algorithm is visualized in Figure 2. Note that due to differences in fan shapes across training images, all operations are carried out on the warped version of the image. This has the additional benefit that introduced image features follow the polar directions of the fan shape which results in more realistic image alterations.

2.4. Shadows

Shadows are often the result of US waves not reaching the affected regions. This can be due to bony structure attenuating some of the signal, resulting in acoustic shadows, or due to smooth curved objects reflecting beams in a non-intended direction, creating refraction artifacts. It is hypothesized that further reducing the pixel intensities of these regions, is an effective data augmentation strategy. To retain the same amount of energy in the US signal, the pixel intensities of other regions are increased. To achieve this Otsu's method is used to determine a threshold τ_o that indicates the maximum intensity of a pixel belonging to a shadow. Since binary masking would introduce noticeable transitions at the edges of the segmentation, a soft thresholding function is used based on [36] to obtain a smoothed mask S:

$$S(u,v) = \begin{cases} \frac{1}{2} + \frac{1}{2}\cos\left(\pi\frac{X(u,v)}{\tau_o}\right) & \text{if } X(u,v) \leq \tau_o \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

In this equation τ_o is the threshold obtained by applying Otsu's method. Next, the new image is obtained by scaling the regions according to S:

$$X_{new}(u,v) = \gamma S(u,v)X(u,v) + v(1-S(u,v))X(u,v), \tag{8}$$

where $\gamma \in [0,1)$ is the energy reduction factor and v a normalization factor such that the total energy in X_{new} is the same as the original image X. To obtain the energy of a region of an image X, all intensities in that region are summed.

2.5. Fan-Preserving Cropping

Although cropping is a widely used data transformation, extensive use may result in illogical US images due to corruption of the fan shape. To combat this, we propose to reduce the fan angle and acquisition depth by a factor f_a and f_d respectively such that the result still has a well defined fan shape. This is achieved by using warping operations $\Phi(X)$ and an adapted version of its inverse $\Phi(X)^{-1}$. During unwarping, outer left, right and bottom sides are simply ignored and are replaced by zero-valued pixels. This corresponds to making all scan lines shorter and deleting the outer ones, which results in the desired cropping effect. Note that this transform never crops the side of the image close to the US probe since this would result in anatomically implausible results.

2.6. Fan-Preserving Elastic Deformation

Elastic deformation is a popular data augmentation technique in medical imaging as it can simulate anatomical variability [7]. It works by constructing a rectangular grid consisting of grid points spaced Δx and Δy pixels apart in the x and y direction respectively. By randomly perturbing these

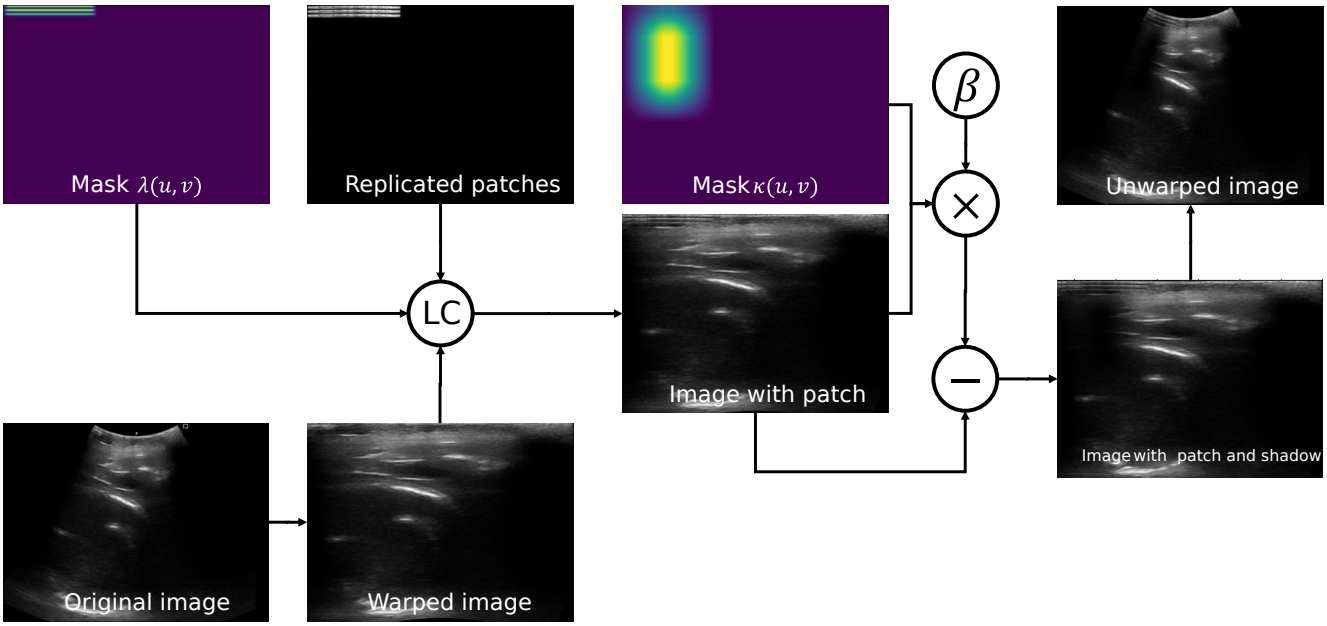

Figure 2. Schematic overview of the reverberation simulation. The reverberation patch is linearly blended with the warped image according to smoothing mask λ. Next, the result is masked according to smoothing mask κ and multiplied with factor β to obtain a shadow. This shadow is subtracted from the image. To obtain the final result the image is unwarped.

points and resampling according to the perturbed grid, the image is elastically deformed. Typically, a grid point p is perturbed by adding a random offset $\delta_p \in \mathbb{R}^2$ with

$$\delta_p = (s_{p,x}, s_{p,y}),\ s_{p,x}, s_{p,y} \sim \mathcal{N}(0, C^2) \tag{9}$$

where $C \sim \text{Uniform}(C_{min}, C_{max})$ is equal for every point p. However, in its typical implementation the applicability of elastic deformation in US augmentation is rather limited since it also deforms the shape of the fan. To alleviate this problem, we propose to perform elastic deformation in the warped image space instead. By doing this the fan-shape of the image is not affected since after unwarping the original fan shape is restored. In a typical elastic deformation transform, the grid can be further altered by an affine transformation. This is not incorporated as affine transformations typically use padding (using for example zero-valued pixels) and would therefore introduce padding pixels inside the fan shape of the unwarped image.

2.7. Fan-Preserving Multiplicative Noise

Additive Gaussian noise is another popular data augmentation technique. However, its application on US images should be avoided as noise in US images has a multiplicative nature. As a result, US images augmented with additive noise might therefore not depict a realistic signal. To combat this, we propose a fan-shape preserving multiplicative noise augmentation that more closely adheres to the nature of US signals. This is achieved in the following way:

$$X_{new}(u,v) = \Phi(X)(u,v)N(u,v), \tag{10}$$

Where $N(u,v)$ a normalized noise mask sampled from the gamma distribution parameterized by shape k and scaling θ. More specifically,

$$N(u,v) = \frac{g}{\mu_N}, \tag{11}$$

where $g \sim \text{Gamma}(k,\theta)$ and μ_N is the mean of $N(u,v)$. Note that we first warp the image X, such that after unwarping the fan-mask is unaltered.

3. Experiments & Results

3.1. Data

To analyze the impact of the proposed data augmentation techniques, each technique is utilized in two segmentation tasks on transperineal ultrasound images. The first task is 2D external anal sphincter (EAS) segmentation on a private dataset referred to as *UZ-EAS*. The second task is 2D fetal head (FH) and symphysis pubis (SP) segmentation on the public *JNU-IFM* dataset [16]. For more robust evaluation, all models are trained using 5-fold cross-validation.

The UZ-EAS dataset consists of 115 transperineal ultrasound (TPUS) volumes acquired in the UZ Leuven pelvic floor dysfunction clinic between October 2019 and March 2021. All volumes were acquired with a Voluson E10 BT16

ultrasound system (GE Healthcare; Zipf, Austria) equipped with a 3D 4-8 MHz convex probe placed transperineally with an average voxel resolution of 0.3mm by 0.3mm by 0.3mm. Ground truth labels for the EAS were created in 4D View VOCAL (GE Healthcare; Zipf, Austria) by a clinical expert with over five years of experience in anal sphincter US imaging and organ segmentation. This dataset was split on patient level into five parts of similar size. For each fold one part was taken as test set, another part as validation set and the other parts were combined to form the training set, resulting in a train-test-validation split of 0.6-0.2-0.2. Since the data augmentation techniques were implemented in 2D, 2D axial planes at the centre of the EAS were extracted.

The JNU-IFM dataset consists of 2D US clips from which certain frames are extracted. However, not all frames have a FH and SP segmentation, so for this experiment only the images which have both were selected. A similar technique as with the UZ-EAS dataset was followed to obtain the folds. However, to simulate a scenario where there is limited data, only two images per patient were retained for training and validation. For the test set all the images of each test patient were used.

3.2. Data Augmentation Configurations

To compare the usability and preference of our techniques over conventional methods, several baseline models with different augmentation configurations are used for comparison. All baselines with corresponding names are listed below.

- **None** – This configuration uses no data augmentation.
- **Conventional** – This configuration uses only conventional data techniques that are frequently used for US images. Used augmentations include horizontal flipping, contrast adjustments with a factor between 0.8 and 1.5, random Gaussian smoothing with a uniform standard deviation between 0.5 and 1.0 and a random rotation between -30° and 30° and a translate range between -40 and 40. All augmentations are applied with a probability of 0.3.
- **Conventional + Crop** – This configuration includes the conventional data techniques but also includes random cropping for comparison with our fan-shape aware cropping technique. Cropping is performed by selecting a centered ROI that has a shape equal to fraction f of the original image size. During training f is sampled from $\mathrm{Uniform}(0.8; 0.99)$ and cropping is applied with a probability of 0.3.
- **Conventional + Elastic** – This configuration includes the conventional data techniques but also includes random elastic deformation with a uniform grid point spacing of 60px and magnitude range in $[1, 3]$ for comparison with our fan-shape preserving elastic deformation technique. Elastic deformation is applied with a probability of 0.3.

parameters for US-EAS	parameters for JNU-IFM
Speckle distort	
$m_{tx} = m_{ty} = 0.5$ $\psi_{min} = -\pi$ $\psi_{max} = \pi$	$m_{tx} = m_{ty} = 0.6$ $\psi_{min} = -\pi$ $\psi_{max} = \pi$
Fan-preserving cropping	
$f_a \sim \mathrm{Uniform}(0.8, 0.99)$ $f_d \sim \mathrm{Uniform}(0.8, 0.99)$	$f_a \sim \mathrm{Uniform}(0.8, 0.99)$ $f_d \sim \mathrm{Uniform}(0.8, 0.99)$
Fan-preserving elastic deformation	
$\Delta x = \Delta y = 60$ $C_{min} = 1$ $C_{max} = 3$	$\Delta x = \Delta y = 60$ $C_{min} = 1$ $C_{max} = 3$
Fan-preserving multiplicative noise	
$k \sim \mathrm{Uniform}(40, 100)$ $\theta = 1$	$k \sim \mathrm{Uniform}(40, 100)$ $\theta = 1$
Shadow	
$\gamma \sim \mathrm{Uniform}(0.5, 0.8)$	$\gamma \sim \mathrm{Uniform}(0.5, 0.8)$
Reverberation Simulation	
$x_{min} = 0$ $x_{max} = 0.85$ $y_{min} = 0$ $y_{max} = 0$ $w_{min} = 0.15$ $w_{max} = 0.75$ $h_{min} = 0.02$ $h_{max} = 0.04$ $\beta = 0.8$ $d = 0.5$	$x_{min} = 0$ $x_{max} = 0.85$ $y_{min} = 0$ $y_{max} = 0$ $w_{min} = 0.15$ $w_{max} = 0.75$ $h_{min} = 0.02$ $h_{max} = 0.04$ $\beta = 0.95$ $d = 0.5$

Table 1. Parameter values used in experiments for the US-EAS and JNU-IFM datasets. For the meaning of the symbols we refer to section 2.

- **Conventional + Noise** – This configuration includes the conventional data techniques but also includes random zero-mean Gaussian noise with a standard deviation of 0.4 for comparison with our fan-shape aware multiplicative noise technique. The noise is added with a probability of 0.3 to the normalized image.

To test the impact of the proposed data augmentations every technique is trained in a configuration where they are added to the conventional data augmentations. Parameters for each augmentation are specified in table 1. All values were decided based on visual coherency with images of the training set. All transformations are applied with a probability of 0.3. The fan-aware cropping and elastic deformation augmentations are also applied to the segmentation masks.

3.3. Model Architecture & Loss Function

The dynamic U-Net [18] implementation of MONAI [5] was chosen as a network architecture. The network contains five levels from which the upper three ones are trained

using deep supervision. As a loss function a combination of dice loss and cross entropy loss was used.

3.4. Implementation Details

The implementation of the data augmentations was done using PyTorch and MONAI packages. Training of all models was performed on a Linux system using a 80GB Nvidia A100 GPU. The Adam optimizer was used with a learning rate of 0.001 and weight decay of 0.001. Each model was trained for a maximum of 100 epochs. Model selection was performed based on the average Dice score of the validation set.

3.5. Results

Table 2 shows the results for both the UZ-EAS and JNU-IFM datasets for all trained configurations. Performance is analyzed using Dice score and the 95^{th} percentile Hausdorff distance, averaged over all five folds. Since some patients have more images than others, metrics were first averaged per patient. Arrows indicate whether a higher (↑) or lower (↓) value is desired. The best-performing values are marked in bold.

4. Discussion & Conclusion

In this paper we proposed six data augmentation techniques to simulate US specific artifacts and solve potential issues with conventional data augmentations. All techniques are easily applicable for any fan-shaped US images as they do not rely on any additional external segmentations or separately trained generative models. Results in table 2 for different augmentation configurations on two datasets showed the possible harm that conventional data augmentations may cause. Even though the conventional data augmentations improve drastically on the "None" configuration, it shows that standard cropping, elastic deformation and additive noise can degrade performance. Preserving a plausible fan-shape is therefore an important factor to keep in mind when augmenting US data.

For segmentation of the EAS and SP, fan-aware cropping seems to remedy the performance drop caused by conventional cropping as it reaches similar performance as the "Conventional" baseline. However, given the similarity to this baseline, one might not benefit from cropping at all. For the segmentation of the FH, conventional cropping seems less harmful as it reaches similar performance as the conventional methods. This might be explained by the fact that the FH is relatively centered on each image such that conventional cropping does not affect the FH in a significant way. Fan-aware cropping might crop the FH more harsher as it's often more located to the sides of the fan.

Elastic deformation of the fan-shape seems less problematic. For the JNU-IFM dataset it often degraded performance but the fan-aware variant was not able to remedy it, implying that elastic deformation might not be appropriate for that dataset. This might be because the SP and FH are rather rigid structures which have a more consistent shape. Elastic deformation might warp these structures to a shape that is anatomically less plausible, which might not be useful since the model will not see such cases during testing. For the UZ-EAS this is different as the visualized area consists mostly of soft tissue and circular striated muscle which are more likely to be deformed by for example probe pressure. However, warping of the fan-mask does not seem to harm performance as the conventional elastic deformation reaches similar performance in both Dice score and Hausdorff distance.

Further, we compare whether it is beneficial to use gamma distributed multiplicative noise instead of additive Gaussian noise. As stated before, additive Gaussian noise was not deemed appropriate since noise in US images typically has a more multiplicative nature. As seen in table 2, this hypothesis is partly true. For the EAS, a slight drop in performance is seen compared to the "Conventional" configuration. There is however a slight decrease in Hausdorff distance but this is only marginal and does not outweigh the Dice score drop. Using multiplicative noise seems to be a better solution as it improves Hausdorff distance and achieves better Dice score. However, a trade-off between Dice score and Hausdorff distance has to be made to decide whether adding noise is beneficial. The same holds for the SP, where the multiplicative noise outperforms the additive noise but comes at a small dice score drop compared to the "Conventional" baseline. For the FH, however, additive noise seems beneficial as both Dice score and Hausdorff distance both improve and even beat the performance of its fan-shape aware variant.

The shadow alteration augmentation does not seem appropriate for both datasets. Performance based on both Dice score and Hausdorff distance is always lower compared to the "Conventional" baseline. The assumption that the shadow augmentation makes, namely that shadows can be relatively easily segmented based on Otsu's method, might therefore not hold in every image.

The same holds for the speckle distort augmentation. The speckle distort augmentation assumes that speckle can be captured by the high frequency components. While this visually holds up, setting the frequency threshold might require more extensive tuning or should be image-specific. Alternatively, there might be information inside the speckle that the model uses. Distorting it could then potentially hurt performance.

The reverberation simulation augmentation proved to be more successful. For the JNU-IFM dataset, a performance improvement is observed in all metrics, except for the Dice score of the SP, where there is a slight drop. For the EAS, however, it does not seem to improve the quality of the seg-

	UZ-EAS		JNU-IFM			
Configuration	**DSC EAS** (↑)	**HD95 EAS** (↓)	**DSC SP** (↑)	**DSC FH** (↑)	**HD95 SP** (↓)	**HD95 FH** (↓)
None	0.8109 ± 0.10	15.80 ± 12.85	0.7231 ± 0.10	0.8274 ± 0.08	21.98 ± 14.02	26.32 ± 16.83
Conventional	0.8531 ± 0.08	12.67 ± 10.84	**0.7890 ± 0.10**	0.8615 ± 0.05	12.53 ± 09.71	22.69 ± 10.25
Conventional + Crop	0.8375 ± 0.09	12.83 ± 09.27	0.7650 ± 0.12	0.8655 ± 0.05	12.52 ± 08.88	22.41 ± 10.46
Conventional + Elastic	0.8530 ± 0.07	11.95 ± 09.50	0.7676 ± 0.11	0.8474 ± 0.06	13.83 ± 12.33	23.41 ± 09.56
Conventional + Noise	0.8400 ± 0.08	12.55 ± 09.79	0.7759 ± 0.11	0.8658 ± 0.06	14.04 ± 10.28	**20.58 ± 09.76**
Speckle Distort	0.8348 ± 0.09	13.53 ± 11.55	0.7677 ± 0.11	0.8446 ± 0.07	15.12 ± 13.17	25.42 ± 09.33
Shadow	0.8486 ± 0.08	12.08 ± 09.38	0.7765 ± 0.10	0.8532 ± 0.06	12.67 ± 09.87	23.57 ± 10.48
Reverberation	0.8418 ± 0.10	13.58 ± 14.70	0.7722 ± 0.10	**0.8691 ± 0.06**	**10.82 ± 07.32**	21.26 ± 09.63
Fan Crop	**0.8576 ± 0.06**	12.95 ± 10.63	0.7851 ± 0.09	0.8618 ± 0.06	14.79 ± 13.41	21.72 ± 09.97
Fan Elastic	0.8529 ± 0.07	11.80 ± 08.11	0.7647 ± 0.12	0.8484 ± 0.07	16.17 ± 12.84	26.96 ± 12.24
Multiplicative Noise	0.8486 ± 0.08	**11.55 ± 09.47**	0.7827 ± 0.11	0.8427 ± 0.08	11.63 ± 14.56	24.40 ± 14.54
All	0.8255 ± 0.09	15.51 ± 12.40	0.7856 ± 0.09	0.8541 ± 0.05	12.13 ± 09.67	24.05 ± 09.94

Table 2. Dice scores (DSC) and 95^{th} percentile Hausdorff distancse (HD95) for 5-fold cross-validation for the segmentation of the external anal sphincter (EAS) on the UZ-EAS dataset and the segmentation of the fetal head (FH) and symphysis pubis (SP) on the JNU-IFM dataset. Best values are indicated in bold.

mentations. This might be because the EAS is often located near the probe. When a reverberation is added above the EAS its shadow might obscure important features which might negatively impact the learning process.

Lastly, combining all augmentations does not seem to elevate performance. This might be because of less performing augmentations that have a destructive impact, or by the incompatibility of certain techniques. In future work, incompatibility should be further investigated to identify if this behaviour could be eliminated.

In conclusion, US augmentation is a complex process, where it seems that more does not always mean better. We proved that certain conventional techniques might corrupt important features but can sometimes be remedied by their fan-aware variants. Further, augmenting by simulation of artifacts could be beneficial but should be cautiously approached and evaluated for each dataset and each structure individually, while keeping anatomical properties in mind.

References

[1] Ajay Anand and Nageswara R. Gurram. Automated deep learning-based single-step diameter estimation of carotid arteries in b-mode ultrasound. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 434–437, 2022. 1

[2] Lennart Bargsten and Alexander Schlaefer. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. *Int. J. Comput. Assist. Radiol. Surg.*, 15(9): 1427–1436, 2020. 2

[3] Sachintha R. Brandigampala, Abdullah F. Al-Battal, and Truong Q. Nguyen. Data augmentation methods for object detection and segmentation in ultrasound scans: An empirical comparative study. In *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 288–291, 2022. 1

[4] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2):125, 2020. 1

[5] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer, Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward, Prerna Dogra, Sebastien Ourselin, and Andrew Feng. Monai: An open-source framework for deep learning in healthcare, 2022. 6

[6] Edward Chen, Howie Choset, and John Galeotti. Uncertainty-based adaptive data augmentation for ultrasound imaging anatomical variations. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 438–442, 2021. 2

[7] Manuel Cossio. Augmenting medical imaging: A comprehensive catalogue of 65 techniques for enhanced data analysis, 2023. 1, 4

[8] Zalan Fabian, Reinhard Heckel, and Mahdi Soltanolkotabi. Data augmentation for deep learning based accelerated mri reconstruction with limited data, 2021. 2

[9] Margarida R. Ferreira, Helena R. Torres, Bruno Oliveira, João Gomes-Fonseca, Pedro Morais, Paulo Novais, and João L. Vilaça. Comparative analysis of current deep learning networks for breast lesion segmentation in ultrasound images. In *2022 44th Annual International Conference of the*

IEEE Engineering in Medicine & Biology Society (EMBC), pages 3878–3881, 2022. 1

[10] Fabio Garcea, Alessio Serra, Fabrizio Lamberti, and Lia Morra. Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*, 152:106391, 2023. 1

[11] Sobhan Goudarzi, Jesse Whyte, Mathieu Boily, Anna Towers, Robert D. Kilgour, and Hassan Rivaz. Segmentation of arm ultrasound images in breast cancer-related lymphedema: A database and deep learning algorithm. *IEEE Transactions on Biomedical Engineering*, 70(9):2552–2563, 2023. 1

[12] Sobhan Goudarzi, Jesse Whyte, Mathieu Boily, Anna Towers, Robert D. Kilgour, and Hassan Rivaz. Segmentation of arm ultrasound images in breast cancer-related lymphedema: A database and deep learning algorithm. *IEEE Transactions on Biomedical Engineering*, 70(9):2552–2563, 2023. 2

[13] Gunti Gunarathne. *Advancements and Breakthroughs in Ultrasound Imaging*. IntechOpen, Rijeka, 2013. 1

[14] Zihao Guo, Jianqiao Zhou, and Di Zhao. Thyroid nodule ultrasonic imaging segmentation based on a deep learning model and data augmentation. In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pages 549–554, 2020. 1

[15] Wenxin Jiang, Xiaotong Chen, Ning Lv, Miao Rao, Yanvan Yu, Weibao Qiu, and Jianming Li. The application of data imputation and deep learning network in the papillary thyroid carcinoma classification. In *2021 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2021. 1

[16] Yaosheng Lu, Mengqiang Zhou, Dengjiang Zhi, Minghong Zhou, Xiaosong Jiang, Ruiyu Qiu, Zhanhong Ou, Huijin Wang, Di Qiu, Mei Zhong, Xiaoxing Lu, Gaowen Chen, and Jieyun Bai. The jnu-ifm dataset for segmenting pubic symphysis-fetal head. *Data in Brief*, 41:107904, 2022. 2, 5

[17] O.V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 53(1):64–78, 2006. 1

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 6

[19] José Seabra and João Miguel Sanches. *RF Ultrasound Estimation from B-Mode Images*, pages 3–24. Springer US, Boston, MA, 2012. 2

[20] C. M. Sehgal. Quantitative relationship between tissue composition and scattering of ultrasound. *The Journal of the Acoustical Society of America*, 94(4):1944–1952, 1993. 3

[21] Richard Shaw, Carole Sudre, Sebastien Ourselin, and M. Jorge Cardoso. Mri k-space motion artefact augmentation: Model robustness and task-specific uncertainty. In *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, pages 427–436. PMLR, 2019. 2

[22] Rohit Singla, Cailin Ringstrom, Ricky Hu, Victoria Lessoway, Janice Reid, Robert Rohling, and Christophe Nguan. Speckle and shadows: Ultrasound-specific physics-based data augmentation for kidney segmentation. In *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*, pages 1139–1148. PMLR, 2022. 2

[23] Erik Smistad, Kaj Fredrik Johansen, Daniel Høyer Iversen, and Ingerid Reinertsen. Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. *Journal of Medical Imaging*, 5(4): 044004, 2018. 1, 2

[24] Eric J. Snider, Sofia I. Hernandez-Torres, and Ryan Hennessey. Using ultrasound image augmentation and ensemble predictions to prevent machine-learning model overfitting. *Diagnostics*, 13(3), 2023. 2

[25] Eric J Snider, Sofia I Hernandez-Torres, and Ryan Hennessey. Using ultrasound image augmentation and ensemble predictions to prevent machine-learning model overfitting. *Diagnostics (Basel)*, 13(3), 2023. 1

[26] Noam Suissa, Sean D Jeffries, Kevin Song, Samer Botros, and Thomas M Hemmerling. Utilizing deep learning to identify an ultrasound-guided nerve block target zone. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4, 2023. 1

[27] Hao Tan, Xun Lang, Bingbing He, Yu Lu, and Yufeng Zhang. Gan-based medical image augmentation for improving cnn performance in myositis ultrasound image classification. In *2023 6th International Conference on Electronics Technology (ICET)*, pages 1329–1333, 2023. 2

[28] Maria Tirindelli, Christine Eilers, Walter Simson, Magdalini Paschali, Mohammad Farid Azampour, and Nassir Navab. Rethinking ultrasound augmentation: A physics-inspired approach, 2021. 1, 2, 3

[29] Yamuna V, Monika M, Amrisha R R, Arun Kumar S, and Sasikala S. Towards improving breast cancer detection in ultrasound images using deep transfer learning. In *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pages 1–6, 2023. 1

[30] Simão Valente, Pedro Morais, Helena R. Torres, Bruno Oliveira, L.R. Buschle, A. Fritz, Jorge Correia-Pinto, Estevão Lima, and João L. Vilaça. A comparative study of deep learning methods for multi-class semantic segmentation of 2d kidney ultrasound images. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4, 2023. 1

[31] R.J.G. van Sloun, R.R. Wildeboer, A.W. Postema, C.K. Mannaerts, M. Gayer, H. Wijkstra, and M. Mischi. Zonal segmentation in transrectal ultrasound images of the prostate through deep learning. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2018. 1

[32] Jianxiong Wang, Shuai Li, Wenfeng Song, Hong Qin, Bo Zhang, and Aimin Hao. Learning from weakly-labeled clinical data for automatic thyroid nodule classification in ultrasound images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3114–3118, 2018. 1

[33] Helena Williams, Laura Cattani, Tom Vercauteren, Jan Deprest, and Jan D'hooge. Automatic tomographic ultrasound imaging sequence extraction of the anal sphincter. In *Simplifying Medical Ultrasound*, pages 35–44, Cham, 2021. Springer International Publishing. 2

[34] Paul F. R. Wilson, Mahdi Gilany, Amoon Jamzad, Fahimeh Fooladgar, Minh Nguyen Nhat To, Brian Wodlinger, Purang Abolmaesumi, and Parvin Mousavi. Self-supervised learning with limited labeled data for prostate cancer detection in high-frequency ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 70(9):1073–1083, 2023. 2

[35] Hongbei Xiang, Yue Zhao, Hao Huang, Kuo Miao, and Xiaoqiu Dong. An intelligent ovarian ultrasound image generation algorithm based on generative adversarial networks. In *2023 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2023. 2

[36] Xuanang Xu, Thomas Sanford, Baris Turkbey, Sheng Xu, Bradford J. Wood, and Pingkun Yan. Shadow-consistent semi-supervised learning for prostate ultrasound segmentation. *IEEE Transactions on Medical Imaging*, 41(6):1331–1345, 2022. 2, 4

[37] Suorong Yang, Weikang Xiao, Mengchen Zhang, Suhan Guo, Jian Zhao, and Furao Shen. Image data augmentation for deep learning: A survey, 2023. 1

[38] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features, 2019. 2

[39] Asaduz Zaman, Sang Hyun Park, Hyunhee Bang, Chul-Woo Park, Ilhyung Park, and Sanghyun Joung. Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images. *Int. J. Comput. Assist. Radiol. Surg.*, 15(6):931–941, 2020. 1, 2

[40] E. Zhixuan Zeng, Ashkan Ebadi, Adrian Florea, and Alexander Wong. Covid-net l2c-ultra: An explainable linear-convex ultrasound augmentation learning framework to improve covid-19 assessment and monitoring. *Sensors*, 24(5), 2024. 1, 2

[41] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization, 2018. 2