Data Augmentation Transformations for Self-Supervised Learning with Ultrasound

Anonymous Author(s)
Affiliation
Address
email

Abstract

Central to joint embedding self-supervised learning is the choice of data augmentation pipeline used to produce positive pairs. This study developed and investigated data augmentation strategies for medical ultrasound. Three pipelines were studied: BYOL augmentations (as a baseline), AugUS-v1 – a pipeline designed to retain semantic content, and AugUS-v2 – a pipeline designed from baseline and AugUS-v1 transformations. Evaluation of SimCLR-pretrained models on diagnostic downstream tasks in lung ultrasound yielded mixed results. The use of AugUS-v1 led to the best performance on COVID-19 classification on a public dataset. However, BYOL and AugUS-v2 outperformed AugUS-v1 on A-line versus B-line classification. AugUS-v2 decidedly obtained the greatest performance on pleural effusion detection. The salient findings were that ultrasound-specific transformations may be suitable for some tasks more than others, and that the random crop and resize transformation was instrumental for all tasks.

1 Introduction

Automated interpretation of medical ultrasound (US) images is increasingly implemented using deep learning strategies [1]. Despite early successes, investigators are limited by the lack of publicly available datasets [2, 3]. It is common for researchers to leverage private repositories of US examinations when accessible, as they may contain far more examinations. Given the expense required to manually annotate US examinations, researchers are turning to self-supervised learning (SSL) methods to pretrain deep neural networks using large unlabeled collections of US data [4]. However, studies investigating joint embedding SSL for US report using standard transformations popularized by SSL publications that did not evaluate on medical imaging datasets [5–11].

In this study, we developed *AugUS-v1* – sets of data augmentation transformations tailored to medical US images. The pipeline was developed from a series of pre-existing and novel stochastic transformations, informed by guidelines in the SSL literature and rooted in an understanding of invariance relationships in ultrasound. It contains geometric transformations (e.g., probe type change simulation), intensity transformations (e.g., contrast change), and texture transformations (e.g., speckle noise simulation). We also developed *AugUS-v2* – a data augmentation pipeline that combines the most impactful transformations from AugUS-v1 and the BYOL pipeline [12], which is widely used in computer vision applications. After evaluating on multiple diagnostic downstream tasks, we found that US-specific transformations alone resulted in the greatest improvement in performance for COVID-19 classification on a public dataset. However, experiments revealed that the baseline and AugUS-v2 pipelines – which may produce semantically inconsistent pairs – outperformed AugUS-v1 on the tasks of A-line versus B-line classification (`AB`) and pleural effusion classification `PE` in lung US. In summary, our contributions are as follows:

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

- Novel transformations for stochastic data augmentation with US images, such as changes to the beam shape and simulation of noise found in US images.
- The development and evaluation of data augmentation pipelines for SSL in US

To our knowledge, this study is the first to explore data augmentation methods for SSL with ultrasound. We are hopeful that the transformations proposed in this study may contribute to the development of foundation models for medical US.

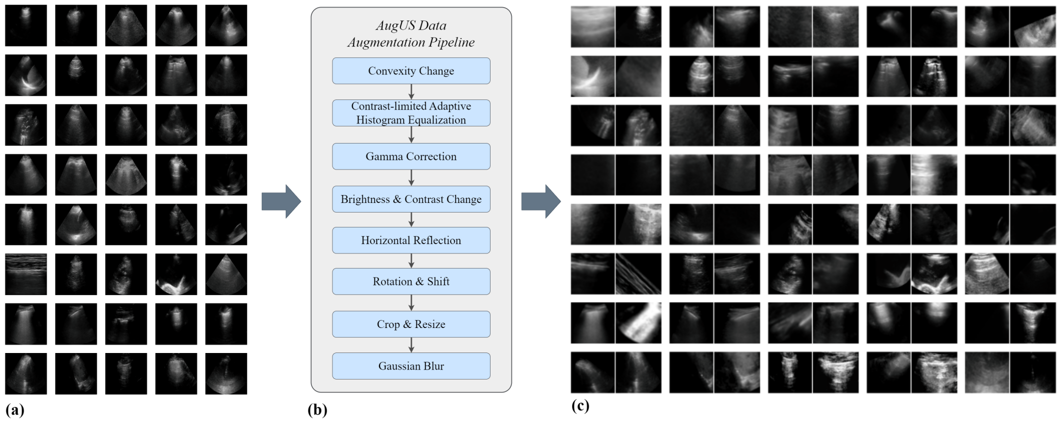


Figure 1: Overview of our methods. A batch of ultrasound images **(a)** is subjected to ultrasound-specific stochastic transformations **(b)** twice to create pairs of images for self-supervised learning **(c)**.

2 Methods

Datasets: We assess our methods using public and private lung US data. COVIDx-US is a public COVID-19 lung US dataset consisting of 242 publicly sourced videos, acquired from a variety of manufacturers and sites [13]. Each example is annotated with one of the following classes: normal, COVID-19 pneumonia, non-COVID-19 pneumonia, and other lung pathology. Referred to as `COVID` hereafter, the task is a four-class classification problem. Since there is no standard test partition, we split the data by patient identifier into training (70%), validation (15%), and test (15%) splits. Pretraining is conducted using the training split.

The second data source is a private collection of lung ultrasound examinations originating from one local and one external institution. Access to this data was granted by [redacted organization] on [redacted date].[1] The dataset contains both parenchymal and pleural views of the lung. A subset of the parenchymal views have binary labels for the presence of A-lines or clinically significant B-lines (i.e., the `AB` task). A-lines are horizontal artifacts signifying normal lung, while B-lines are ray-like artifacts that are generally indicative of abnormal lung tissue [14]. Some of the pleural views have labels for the presence or absence of pleural effusion (i.e., the `PE` task), which refers to a typically pathological collection of fluid between the pleura of the lung. Pretraining is conducted using all unlabelled examples and all labelled examples in the training split. Table 1 provides the video and class counts of the private dataset.

Novel Transformations: We developed novel transformations that could serve as components in a novel data augmentation pipeline for SSL with US. Transformations included probe type change, beam convexity change, wavelet transform denoising, depth change simulation, speckle noise application, and salt & pepper noise application. Refer to Fig. 2 for a visual example of each transformation. Algorithmic details for each transformation are provided in Appendix A.

AugUS-v1: Several SSL studies applied to photographic or medical imaging datasets adapt the BYOL data augmentation pipeline used by Grill *et al.* [12]. With invariance relationships for the US modality in mind, we designed *AugUS-v1* – a data augmentation pipeline specific to ultrasound images. AugUS-v1 includes both the novel US transformations and a selection of transformations from BYOL that were selected to preserve information while imposing nontrivial differences across

[1] Omitted to protect anonymity during the review process.

	Local			External
	Train	Validation	Test	Test
Videos	5679	1184	1249	925
AB labels	2067/999	459/178	458/221	286/327
PE labels	789/762	176/142	162/158	68/110
Patients	1702	364	364	168

Table 1: Breakdown of the private dataset. x/y indicates the number of labeled videos in the negative and positive class for each binary classification task.

invocations. Lists of all transformations constituting the BYOL and AugUS-v1 pipelines may be found in Appendix B.

AugUS-v2: To construct AugUS-v2, we conducted an ablation study that revealed which transformations were contributory to downstream performance across tasks. We pretrained several models for each dataset using SimCLR [15], for each of the BYOL and AugUS-v1 pipelines. A model was pretrained using each complete augmentation pipeline to establish baseline performance. Pretraining was repeated for each transformation in the pipeline, using an incomplete version of the pipeline in which that transformation was omitted. Linear classifiers were trained using each feature extractor. The maximum validation area under the receiver operating curve (AUC) across epochs was recorded (values are reported in Appendix C). Of note was the decrease in performance without the random crop and resize (RCR) transform, which was unexpected because it frequently omits large portions of the image containing the artifacts to be detected. Transformations that, when excluded, did not improve validation AUC compared to baselines were included in AugUS-v2.

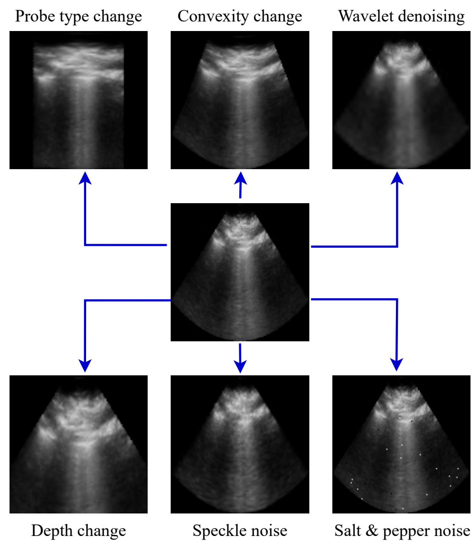


Figure 2: An example of applying each novel transformation to the same US image.

3 Experiments & Evaluation

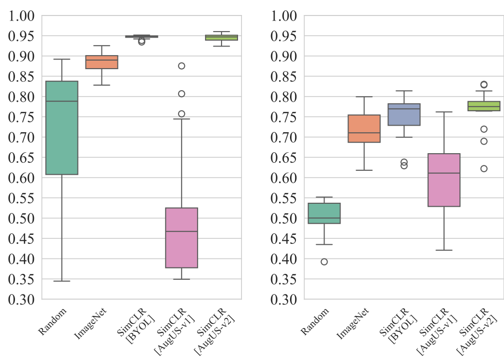
Linear Classification: Table 2 reports the test AUC achieved by linear classifiers trained using pretrained models' feature representations. AugUS-v1 achieved the greatest performance on `COVID`, BYOL and AugUS-v2 performed comparably on `AB`, and AugUS-v2 obtained the greatest test AUC on `PE`. Figure 4 offers visual insight into the separability of pretrained representations for each task.

Fine-Tuning: As shown in Table 2, the AugUS-v2 and BYOL pipelines performed comparably well on the `AB` task on both the local and external test sets. AugUS-v2 attained greater local and external test AUC than the BYOL pipeline on `PE`. On the `COVID` task, AugUS-v1 was observed to have the greatest test performance, with AugUS-v2 achieving the lowest metrics. The test split of the comparatively smaller COVIDx-US may have been particularly difficult, as validation set performance was markedly greater across all pipelines. Training details are in Appendix D.

Label Efficiency: To assess performance in low-label settings, pretrained models were fine-tuned on subsets of approximately 5% of the private dataset's training set and evaluated on the test set. The training set was split randomly by patient identifier and stratified by label. Note that splitting by patient creates a difficult learning problem, since all images belonging to the same video or patient are related. Figure 3 displays the test AUC distribution across trials. One-way repeated ANOVA revealed that the means of the test AUC across trials were significantly different for both `AB` and `PE`. Post-hoc paired t-tests were performed between each condition, using the Bonferroni correction with a family-wise error rate of $\alpha = 0.05$. All means were significantly different for `AB`, except for those between the BYOL and AugUS-v2 pipelines. For `PE`, all means were significantly different.

Weights	Pipeline	Linear Classification			Fine-Tuning				
		COVID	AB Local	PE Local	COVID	AB Local	AB External	PE Local	PE External
Random	-	0.500	0.500	0.500	0.528	0.949	0.816	0.811	0.770
ImageNet	-	0.627	0.950	0.884	0.690	0.947	0.799	0.851	0.801
SimCLR	BYOL	0.779	**0.966**	0.870	0.449	**0.971**	**0.876**	0.870	0.825
SimCLR	AugUS-v1	**0.820**	0.947	0.852	0.787	0.746	0.634	0.746	0.611
SimCLR	AugUS-v2	0.684	0.962	**0.903**	**0.696**	0.970	**0.876**	**0.895**	**0.889**

Table 2: Test set AUC obtained by models fine-tuned for each of the COVID, AB, and PE tasks. For COVID, the average across the classes is reported.



(a) AB (b) PE

Figure 3: Distribution of test AUC for classifiers trained on disjoint subsets of 5% of the patients in the training partition of the private dataset

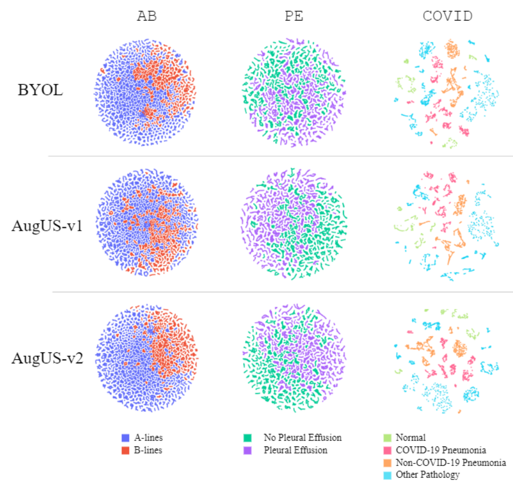


Figure 4: t-SNE projections for test set embeddings produced by pretrained models, for all tasks and data augmentation pipelines.

4 Discussion

The results indicated that the choice of data augmentation pipeline can greatly affect the performance of deep classifiers pretrained using SSL. The handcrafted US-specific pipeline, AugUS-v1, led to the strongest performance for the COVID task, which is a difficult multi-class problem with little data available. However, the AugUS-v1 pipeline resulted in worse performance on AB and PE than not pretraining at all. The baseline BYOL pipeline performed comparably to the AugUS-v2 pipeline, which was empirically designed to combine the best of the BYOL and AugUS-v1 pipelines. However, use of the AugUS-v2 pipeline led to the greatest performance on the PE task, for which less labels were available compared to AB.

A salient result was the impact of the RCR transformation, which is in the BYOL pipeline. In the context of US, this transformation can obliterate artifacts and structures necessary for interpretation. Despite this, ablating RCR had the greatest impact on validation set AUC for each task.

The main conclusion was that the use of data augmentation designed to preserved semantic content did not assuredly lead to improved downstream performance for US diagnostic tasks. A limitation of this study is that SimCLR was the only SSL method investigated. Additionally, some of the novel transformations are computationally expensive – with an efficient choice of feature extractor, the augmentation pipeline becomes a bottleneck during training. Future work can apply the methods of this study to assess the impact of data augmentation pipelines for US diagnostic tasks outside of the lung and for other SSL objectives.

References

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

[2] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: a review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.

[3] M. Y. Ansari, I. A. C. Mangalote, P. K. Meher, O. Aboumarzouk, A. Al-Ansari, O. Halabi, and S. P. Dakua, "Advancements in deep learning for b-mode ultrasound segmentation: A comprehensive review," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024.

[4] B. VanBerlo, J. Hoey, and A. Wong, "A survey of the impact of self-supervised pretraining for diagnostic tasks in medical x-ray, ct, mri, and ultrasound," *BMC Medical Imaging*, vol. 24, no. 1, p. 79, 2024.

[5] Y. Chen, C. Zhang, L. Liu, C. Feng, C. Dong, Y. Luo, and X. Wan, "USCL: Pretraining Deep Ultrasound Image Diagnosis Model Through Video Contrastive Representation Learning," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pp. 627–637, Springer, 2021.

[6] Y. Chen, C. Zhang, C. H. Ding, and L. Liu, "Generating and weighting semantically consistent sample pairs for ultrasound contrastive learning," *IEEE Transactions on Medical Imaging*, 2022.

[7] C. Zhang, Y. Chen, L. Liu, Q. Liu, and X. Zhou, "Hico: Hierarchical contrastive learning for ultrasound video model pretraining," in *Proceedings of the Asian Conference on Computer Vision*, pp. 229–246, 2022.

[8] D. Anand, P. Annangi, and P. Sudhakar, "Benchmarking Self-Supervised Representation Learning from a million Cardiac Ultrasound images," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2022-July, pp. 529–532, Institute of Electrical and Electronics Engineers Inc., 2022. ISSN: 1557170X.

[9] M. Saeed, R. Muhtaseb, and M. Yaqub, "Contrastive Pretraining for Echocardiography Segmentation with Limited Data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13413 LNCS, pp. 680–691, 2022. ISBN: 9783031120527 Publisher: Springer Science and Business Media Deutschland GmbH.

[10] S. Basu, S. Singla, M. Gupta, P. Rana, P. Gupta, and C. Arora, "Unsupervised contrastive learning of image representations from ultrasound videos with hard negative mining," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 423–433, Springer, 2022.

[11] B. VanBerlo, A. Wong, J. Hoey, and R. Arntfield, "Intra-video positive pairs in self-supervised learning for ultrasound," *Frontiers in Imaging*, vol. 3, 2024.

[12] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21271–21284, 2020.

[13] A. Ebadi, P. Xi, A. MacLean, A. Florea, S. Tremblay, S. Kohli, and A. Wong, "COVIDx-US: An open-access benchmark dataset of ultrasound imaging data for AI-driven COVID-19 analytics," *Front. Biosci. (Landmark Ed.)*, vol. 27, p. 198, June 2022.

[14] N. J. Soni, R. Arntfield, and P. Kory, *Point-of-Care Ultrasound*. Philadelphia: Elsevier, second ed., 2020.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.

[16] K. Kim, F. Macruz, D. Wu, C. Bridge, S. McKinney, A. A. Al Saud, E. Sharaf, A. Pely, P. Danset, T. Duffy, *et al.*, "Point-of-care ai-assisted stepwise ultrasound pneumothorax diagnosis," *Physics in Medicine & Biology*, vol. 68, no. 20, p. 205013, 2023.

[17] E. Z. Zeng, A. Ebadi, A. Florea, and A. Wong, "Covid-net l2c-ultra: An explainable linear-convex ultrasound augmentation learning framework to improve covid-19 assessment and monitoring," *Sensors*, vol. 24, no. 5, p. 1664, 2024.

[18] D. Vilimek, J. Kubicek, M. Golian, R. Jaros, R. Kahankova, P. Hanzlikova, D. Barvik, A. Krestanova, M. Penhaker, M. Cerny, *et al.*, "Comparative analysis of wavelet transform filtering systems for noise reduction in ultrasound images," *Plos one*, vol. 17, no. 7, p. e0270745, 2022.

[19] L. Birgé and P. Massart, *From model selection to adaptive estimation*. Springer, 1997.

[20] P. Singh, R. Mukundan, and R. de Ryke, "Synthetic models of ultrasound image formation for speckle noise simulation and analysis," in *2017 International Conference on Signals and Systems (ICSigSys)*, pp. 278–284, IEEE, 2017.

[21] A. Bardes, J. Ponce, and Y. LeCun, "VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning," in *International Conference on Learning Representations*, 2022.

[22] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, *et al.*, "Searching for MobileNetV3," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314–1324, 2019.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[24] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, and C.-J. Hsieh, "Large batch optimization for deep learning: Training bert in 76 minutes," *arXiv preprint arXiv:1904.00962*, 2019.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

A Novel Ultrasound Transformations

To compose the novel transformations, we required information regarding the shape of the US beam. We adopt the naming convention for the vertices of the US beam [16]. Let p_1, p_2, p_3, and p_4 represent the respective locations of the top left, top right, bottom left, and bottom right vertices, and let $\langle x_i, y_i \rangle$ be the x- and y-coordinates of p_i in image space. For convex beam shapes, we denote the intersection of lines $\overleftrightarrow{p_1 p_3}$ and $\overleftrightarrow{p_2 p_4}$ as p_0. The angle at between the vertical and line $\overleftrightarrow{p_0 p_j}$ for some point p_j in image space is denoted as θ_j Fig. 5 depicts the arrangement of these points for each of the three main US beam shapes: linear, curvilinear, and phased array. A software tool was used to estimate the beam vertices and probe type for all videos in each dataset (UltraMask, Deep Breathe Inc., London, Canada).

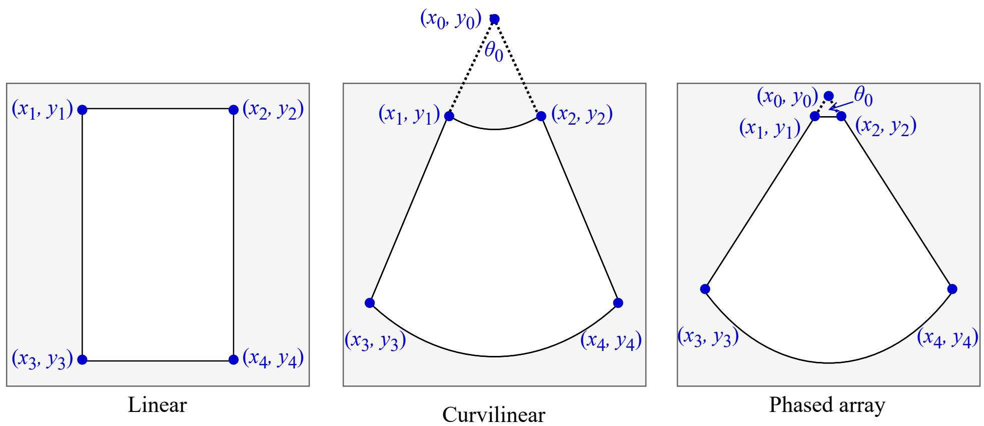


Figure 5: Locations of the named beam vertices for each of the three main beam shapes in US imaging

Probe Type Change: Since US dataset may be acquired with different probes, there often exist examples with different beam shapes. Inspired by Zeng *et al.*'s work [17] that proposed projective and piecewise affine transforms for simulating a conversion between linear and convex beam shapes, we developed a transformation that resamples an image according to a different beam shape using a mapping $\mathbf{f} : \mathbb{R}^2 \to [-1, 1]^2$ that determines pixel locations to sample in the original image, given an input coordinate in the new image. For point (x, y) in the sampled image, Linear beam shapes are converted to curvilinear shapes, and curvilinear and phased array shapes are converted to linear. Algorithm 1 details the calculation of $\mathbf{f}_{\ell \to c}$ for converting linear beams to convex beams with a random radius factor $\rho \sim \mathcal{U}(1, 2)$, along with new beam vertices. Similarly, curvilinear and phased array beam shapes are converted to linear beam shapes, as detailed in Algorithm 2.[2]

Convexity Change: The shape of convex beams can vary, depending on the manufacturer, depth, and field of view of the probe. We produced a transformation that resamples an image to modify the beam shape such that the distance between x_1 and x_2 is different, mimicking a change in θ_0. Depending on the change in θ_0, p_0 is translated vertically, and a new beam shape is computed accordingly. A pixel map $\mathbf{f}_{c \to c'}$ is computed according to Algorithm 3.

Wavelet Denoising: The quality of ultrasound images is often marred by speckle noise, Gaussian noise, and salt & pepper (S&P) noise [18]. We implemented an alternative to the common blur transformation used in data augmentation for SSL. Following the soft thresholding method by Birgé and Massart [19], we apply a wavelet transform, conduct thresholding, then apply the inverse wavelet transform. The mother wavelet is randomly chosen from a set, and the sparsity parameter α is sampled from a uniform distribution. We designated $\{db2, db5\}$ as the set of mother wavelets, which is a subset of those identified by Vilimek *et al.*'s assessment [18] as most suitable for denoising US images.

[2]Since the private dataset was resized to square images that exactly encapsulated the beam, some steps in Algorithms 1, 2, and 3 were modified to account for the resulting non-circular bottom bound of the beam.

Algorithm 1 Compute a point mapping for linear to curvilinear beam shape, along with new beam vertices

Require: Beam vertices p_1, p_2, p_3, p_4; radius factor ρ; coordinate maps $\mathbf{x} = \mathbf{1}_{h\times 1}[0, 1, \ldots, w-1]$ and $\mathbf{y} = [0, 1, \ldots, h-1]\mathbf{1}_{1\times w}$

$r_b \leftarrow \rho(y_3 - y_1)$ ▷ *Bottom sector radius*

$x'_0 \leftarrow (x_4 - x_3)/2$ ▷ *Intersection of lateral bounds*

$y'_0 \leftarrow y_3 - r_b$

$y'_1 = y'_2 \leftarrow y_1$

$y'_3 = y'_4 \leftarrow y'_0 + \sqrt{r_b^2 - (x'_0 - x_1)^2}$

$x'_1 \leftarrow x'_0 - (y_1 - y'_0)(x'_0 - x_3)/(y'_3 - y'_0)$

$x'_2 \leftarrow 2x'_0 - x'_1$

$r_t \leftarrow \sqrt{(x'_0 - x'_1)^2 + (y_1 - y'_0)^2}$ ▷ *Top sector radius*

$\mathbf{f}_{\ell\to c} \leftarrow \begin{pmatrix} \frac{\operatorname{atan2}(\mathbf{x}-x_0, \mathbf{y}-y_0)}{|\operatorname{atan2}(\mathbf{x}-x'_0, y_3-y'_0)|} \\ \frac{2\sqrt{(x'_0-\mathbf{x})^2+(y'_0-\mathbf{y})^2}-r_t}{r_b-r_t} - 1 \end{pmatrix}$

return $\mathbf{f}_{\ell\to c}, p'_1, p'_2, p'_3, p'_4$

Algorithm 2 Compute a point mapping for convex to linear beam shape, along with new beam vertices

Require: Beam vertices p_1, p_2, p_3, p_4; point of intersection p_0; angle of intersection θ_0; width fraction $\omega \in [0, 1]$; coordinate maps $\mathbf{x} = \mathbf{1}_{h\times 1}[0, 1, \ldots, w-1]$ and $\mathbf{y} = [0, 1, \ldots, h-1]\mathbf{1}_{1\times w}$

$r_b \leftarrow \sqrt{(x'_0 - x_3)^2 + (y'_0 - y_3)^2}$ ▷ *Bottom sector radius*

$x'_1 = x'_3 \leftarrow x_0 - \omega w/2$

$x'_2 = x'_4 \leftarrow x_0 + \omega w/2$

$y'_1 = y'_2 \leftarrow y_1$

$y'_3 = y'_4 \leftarrow y_0 + r_b$

$\boldsymbol{\phi} \leftarrow \theta_0((\mathbf{x} - x_3)/(x_4 - x_3) - \frac{1}{2})$ ▷ *Angle with vertical*

$\mathbf{y_n} \leftarrow (\mathbf{y} - y_1)/(y_3 - y_1)$ ▷ *Normalized y-coordinates*

if Probe type is curvilinear **then**

  $r_t \leftarrow \sqrt{(x'_0 - x_1)^2 + (y'_0 - y_1)^2}$ ▷ *Top sector radius*

else

  $r_t \leftarrow (y_1 - y_0)/\cos(\boldsymbol{\phi}/w)$ ▷ *Top line segment*

end if

$\mathbf{f}_{c\to\ell} \leftarrow \begin{pmatrix} x_0 + \sin(\boldsymbol{\phi}/w)(r_t + \mathbf{y_n}(r_b - r_t)) \\ y_0 + \cos(\boldsymbol{\phi}/w)(r_t + \mathbf{y_n}(r_b - r_t)) \end{pmatrix}$

return $\mathbf{f}_{c\to\ell}, p'_1, p'_2, p'_3, p'_4$

Depth Change Simulation: Changing the depth controls on an ultrasound probe impacts how far the range of visibility is from the probe. We simulate a change in depth by implementing a controlled zoom that preserves the centre for linear beam shapes and preserves p_0 for convex beam shapes. The magnitude of the zoom transformation, d, is randomly sampled from a uniform distribution. Increasing the depth corresponds to zooming out ($d > 1$), while decreasing the depth corresponds to zooming in ($d < 1$).

Speckle Noise Simulation: To simulate the addition of speckle noise, we implemented the synthetic speckle noise algorithm by Singh *et al.* [20]. The lateral resolution (δ_ℓ) axial resolution (δ_a) for interpolation and the number of synthetic phasors (M) are randomly drawn from uniform distributions. Sample points on the image are evenly spaced in Cartesian coordinates and polar coordinates for linear beams and convex beams, respectively.

Salt & Pepper Noise Simulation: Imitation of S&P noise was implemented by sampling a random assortment of points in the image and setting their intensities to 255 (salt) or 0 (pepper). The fractions of pixels set to salt and pepper values – f_S and f_P, respectively – are drawn from a uniform distribution.

Algorithm 3 Compute a point mapping from an original to a modified convex beam shape.

Require: Beam vertices p_1, p_2, p_3, p_4; point of intersection p_0; angle of intersection θ_0; new top width w'; coordinate maps $\mathbf{x} = \mathbf{1}_{h\times 1}[0, 1, \ldots, w-1]$ and $\mathbf{y} = [0, 1, \ldots, h-1]\mathbf{1}_{1\times w}$

$s \leftarrow w'(x_4 - x_3)/(x_2 - x_1)$ ▷ *Scale change for top*

$x'_1 \leftarrow x_0 - s(x_0 - x_1)$

$x'_2 \leftarrow x_0 + s(x_2 - x_0)$

$y'_1 = y'_2 \leftarrow y_1$

$y'_3 = y'_4 \leftarrow y_3$

$\theta'_0, p'_0 \leftarrow$ angle, point of intersection of $\overleftrightarrow{p'_1 p'_3}$ and $\overleftrightarrow{p'_2 p'_4}$

$r_b \leftarrow \sqrt{(x_0 - x_3)^2 + (y_0 - y_3)^2}$ ▷ *Current bottom radius*

$r'_b \leftarrow \sqrt{(x'_0 - x'_3)^2 + (y'_0 - y'_3)^2}$ ▷ *New bottom radius*

$r_t \leftarrow \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}$ ▷ *Current top radius*

$r'_t \leftarrow \sqrt{(x'_0 - x'_1)^2 + (y'_0 - y'_1)^2}$ ▷ *New top radius*

$\boldsymbol{\phi}' \leftarrow \mathrm{atan2}(\mathbf{x} - x'_0, \mathbf{y} - y'_0)$

$\mathbf{r}' \leftarrow \sqrt{(x'_0 - \mathbf{x})^2 + (y'_0 - \mathbf{y})^2}$

$\mathbf{r}' \leftarrow (\mathbf{r}' - r'_t)(r_b - r_t)/(r'_b - r'_t) + r_t$

$\mathbf{f}_{c\to c'}(x, y) \leftarrow \begin{pmatrix} x_0 + \mathbf{r}' \sin(\boldsymbol{\phi}'\theta_0/\theta'_0) \\ y_0 + \mathbf{r}' \cos(\boldsymbol{\phi}'\theta_0/\theta'_0) \end{pmatrix}$

return $\mathbf{f}_{c\to c'}, p'_1, p'_2, p'_3, p'_4$

B Data Augmentation Pipeline Details

Tables 3 and 4 detail the transformations in the BYOL and AugUS-v1 pipelines, respectively, along with estimates of the time to transform a single image, averaged over 1000 trials using the same image. For clarity, we assign each transformation an alphanumeric identifier and express a data augmentation pipeline as an ordered sequence of identifiers. Note that we use the symmetrized version of the BYOL pipeline, as in the VICReg paper [21].

Table 3: The sequence of transformations in the BYOL data augmentation pipeline [12]

Identifier	Probability	Transformation	Time [ms]
B00	1.0	Random cropping of $c \sim \mathcal{U}(0.08, 1)$ of the image's area, then resizing to original dimensions	0.29
B01	0.5	Horizontal reflection	0.08
B02	0.8	Color jitter, with maximum brightness, contrast, saturation, and hue changes by up to 0.4, 0.4, 0.2, and 0.1, respectively.	2.40
B03	0.2	Conversion to grayscale	0.19
B04	0.5	Gaussian Blur with a kernel size of $23/224$ of the image's height and standard deviation $\sigma \sim \mathcal{U}(0.1, 2.0)$	0.74
B05	0.1	Solarization, with intensity threshold 128	0.15

C Transformation Ablation Study for AugUS-v2 Design

Table 5 details the results of the transformation ablations. Removal of random crop and resize (RCR) from the BYOL pipeline resulted in the most dramatic performance reduction. Omission of the horizontal flip and rotation & shift transformations also resulted in notable performance drop. In terms of the transformation identifiers from Tables 3 and 4, the AugUS-v2 pipeline can be expressed as the following sequence: [*U01, U03, U04, U05, U10, U11, B00, B04*].

Table 4: The sequence of transformations in the US-specific augmentation pipeline

Identifier	Probability	Transformation	Time [ms]
U00	0.3	Probe type change with $\rho \sim \mathcal{U}(1, 2)$ and $\omega \sim \mathcal{U}(0.5, 1)^\dagger$	2.25
U01	0.75	Convexity change with $w' \sim \mathcal{U}(0, 0.75)^\dagger$	1.92
U02	0.5	Wavelet denoising with $\alpha \sim \mathcal{U}(2.5, 3.5)$, $J_0 = 2$, $J = 3$, and the mother wavelet set $\{db2, db5\}^\dagger$ 5.00	
U03	0.2	Contrast-limited adaptive histogram equalization with a clip limit of $c \sim \mathcal{U}(30, 50)$ and 8-pixel square tiles	4.64
U04	0.5	Gamma correction with $\gamma \sim \mathcal{U}(0.5, 1.75)$	0.52
U05	0.5	Brightness and contrast changes by up to 0.4 each.	0.49
U06	0.5	Depth change with $d \sim \mathcal{U}(0.8, 2)^\dagger$	1.76
U07	0.333	Speckle noise with $\delta_\ell \sim \mathcal{U}(35, 45)$, $\delta_a \sim \mathcal{U}(75, 85)$, and $\delta_\ell \sim \mathcal{U}(5, 10)^\dagger$	3.69
U08	0.333	Gaussian noise with standard deviation $\sigma \sim \mathcal{U}(0.5, 2.5)$	0.28
U09	0.1	Salt & Pepper noise with salt and pepper fractions $f_S, f_P \sim \mathcal{U}(0.001, 0.005)^\dagger$	0.18
U10	0.5	Horizontal reflection	0.19
U11	0.5	Rotation & shift	1.42

† Described in Appendix A

Pipeline	Left-out transform	COVID	AB	PE
BYOL	None	0.891	0.967	0.875
	B00	0.711	0.887	0.741
	B01	0.925	0.969	0.866
	B02	0.868	0.967	0.858
	B03	0.902	0.967	0.868
	B04	0.873	0.966	0.866
	B05	0.921	0.971	0.870
AugUS-v1	None	0.902	0.953	0.856
	U00	0.897	0.965	0.868
	U01	0.891	0.949	0.851
	U02	0.887	0.958	0.866
	U03	0.856	0.950	0.841
	U04	0.870	0.952	0.841
	U05	0.846	0.961	0.854
	U06	0.818	0.962	0.857
	U07	0.906	0.960	0.858
	U08	0.912	0.846	0.722
	U19	0.887	0.957	0.856
	U10	0.847	0.938	0.826
	U11	0.838	0.943	0.818

Table 5: Comparison of the BYOL and AugUS-v1 pipelines with one transformation left out. Performance is reported as the maximum validation set AUC achieved by a linear classifier trained on the output of a frozen feature extractor. `AB` and `PE` refer to binary classification tasks from the private dataset. `COVID` refers to the COVID-19 classification task in the public COVIDxUS dataset.

D Training Details

We adopted the MobileNetV3 architecture [22] for all experiments in this study and pretrained using the SimCLR method [15]. Feature extractors were initialized using ImageNet-pretrained weights [23] and pretrained using the LARS optimizer [24] with a batch size of 2048, a base learning rate of 0.2 and a schedule consisting of warmup with cosine decay. Pretraining was conducted for 15 epochs with 2 warmup epochs for the private dataset, and 100 epochs with 10 warmup epochs for COVIDx-US.

To conduct supervised evaluation, a model head consisting of a fully connected layer was appended to the final pooling layer of the feature extractor. Cross-entropy loss was minimized by training with the Adam optimizer [25] for 10 epochs with a batch size of 256. The weights corresponding to the epoch with the lowest validation loss was retained for test set evaluation. The learning rates for the feature extractor and head were 0.002 and 0.02, respectively. As is customary in SSL evaluation, separate experiments were conducted with the feature extractor's weights held constant (i.e., linear classification) or included as trainable parameters (i.e., fine-tuning). Code for all experiments and transformations will be shared in a public GitHub repository upon publication.