

# Principled Ultrasound Data Augmentation for Classification of Standard Planes

Lok Hin Lee, Yuan Gao and J. Alison Noble

Department of Engineering Science, University of Oxford, UK

**Abstract.** Deep learning models with large learning capacities often overfit to medical imaging datasets. This is because training sets are often relatively small due to the significant time and financial costs incurred in medical data acquisition and labelling. Data augmentation is therefore often used to expand the availability of training data and to increase generalization. However, augmentation strategies are often chosen on an ad-hoc basis without justification. In this paper, we present an augmentation policy search method with the goal of improving model classification performance. We include in the augmentation policy search additional transformations that are often used in medical image analysis and evaluate their performance. In addition, we extend the augmentation policy search to include non-linear mixed-example data augmentation strategies. Using these learned policies, we show that principled data augmentation for medical image model training can lead to significant improvements in ultrasound standard plane detection, with an average F1-score improvement of 7.0% overall over naive data augmentation strategies in ultrasound fetal standard plane classification. We find that the learned representations of ultrasound images are better clustered and defined with optimized data augmentation.

**Keywords:** Data augmentation · Fetal Ultrasound.

## 1 Introduction

The benefits of data augmentation for training deep learning models are well documented in a variety of tasks, including image recognition [20,5,23] and regression problems [12,22]. Data augmentation acts to artificially increase the size and variance of a given training dataset by adding transformed copies of the training examples. This is particularly evident in the context of medical imaging, where data augmentation is used to combat class imbalance [10], increase model generalization [11,18], and expand training data [28,9]. This is usually done with transformations to the input image that are determined based on expert knowledge and cannot be easily transferred to other problems and domains. In ultrasound, this usually manifests as data augmentation strategies consisting of small rotations, translations and scalings [1]. However, whilst it is appealing to base augmentation strategies on “expected” variations in input image presentation, recent work has found that other augmentation strategies that

generate “unrealistic looking” training images [19,24] have led to improvements in generalization capability. There has therefore been great interest in developing data augmentation strategies to automatically generate transformations to images and labels that would lead to the greatest performance increase in a neural network model. In this paper, inspired by the RandAugment [7] augmentation search policy, we automatically look for augmentation policies that outperform standard augmentation strategies in ultrasound imaging based on prior knowledge and extend our algorithm to include mixed-example data augmentation [24] in the base policy search. We evaluate the proposed method on second-trimester fetal ultrasound plane detection, and find that a randomly initialized network with our augmentation policy achieves performance competitive with methods that require external labelled data for network pre-training and self-supervised methods. We also evaluate our method on a fine-tuning a pre-trained model, and find that using an optimized augmentation policy during training improves final performance.

**Contributions:** Our contributions are three fold: 1) We investigate the use of an augmentation search policy with hyperparameters that does not need expensive reinforcement learning policies and can be tuned with simple grid search; 2) We extend this augmentation search policy to combinations that include mixed-example based data augmentation and include common medical imaging transformations; 3) We explain the performance of optimal augmentation strategies by looking at their affinity, diversities and effect on final model performance.

**Related Work** Medical image datasets are difficult and expensive to acquire. There has therefore been previous work that seeks to artificially expand the breadth of training data available in medical image classification [11,16], segmentation [21,10] and regression [9].

**Original Data Manipulation:** Zeshan et al. [16] evaluate the performance of eight different affine and pixel level transformations by training eight different CNNs for predicting mammography masses and find that ensembling the trained models improves the classification performance significantly. Nalepa et al. [21] elastically deform brain MRI scans using diffeomorphic mappings and find that tumour segmentation is improved. However, in the above works, specific augmentations and parameters are selected arbitrarily and are task and modality dependent. In contrast, we propose an automated augmentation policy search method that can out perform conventional medical imaging augmentation baselines.

**Artificial Data Generation:** Florian et al. [9] generates new training samples in by linearly combining existing training examples in regression. Models trained to estimate the volume of white matter hyperintensities had performance comparable to networks trained with larger datasets. Zach et al.[10] also linearly combine training examples and target labels linearly inspired by mix-up [27] but focus on pairing classes with high and low incidence together, which was

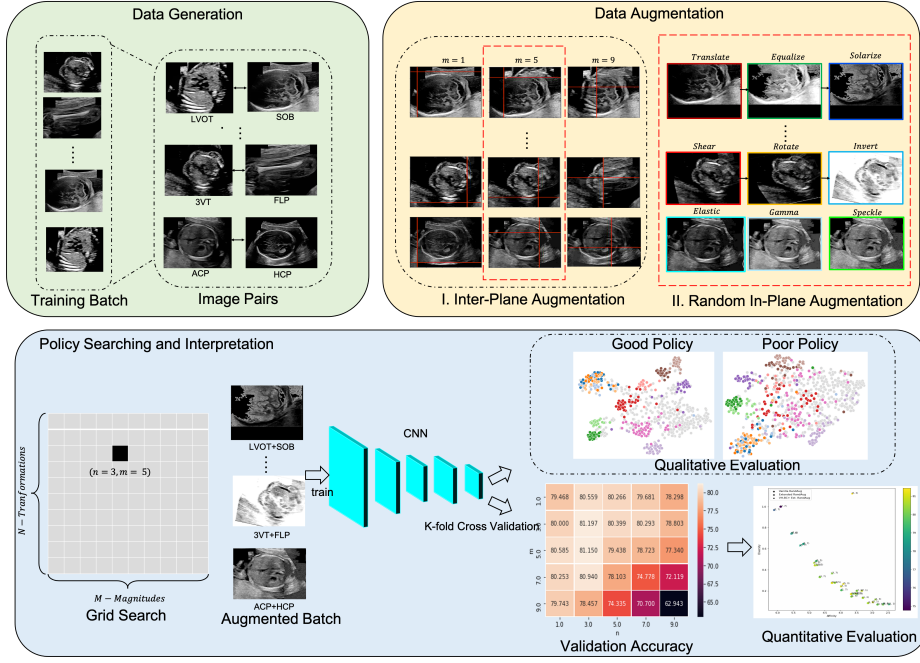
found to be beneficial for tasks with high class imbalance such as in brain tumor segmentation. Maayan et al. [11] train a conditional generative adversarial network (cGAN) to generate different types of liver lesions and use the synthesized samples to train a classification network. Dakai et al. [18] use a cGAN to synthesize 3D lung nodules of different sizes and appearances at multiple locations of a lung CT scan. These generated samples were then used to finetune a pretrained lung nodule segmentation network that improved segmentation of small peripheral nodules. However, cGAN-based methods are difficult to train and have significant computational costs during augmentation.

**Automated Augmentation Policy Search:** There are augmentation policy search methods in the natural image analysis [6,14,19] that learn a series of transformations which are parameterized by their magnitudes and probability. However, these searches are expensive, and cannot be run on the full training dataset as the hyperparameter search for each transformation require significant computational resources. RandAugment (RA) [7] finds that transformations can have a shared magnitude and probability of application and achieve similar performance, without expensive reinforcement learning. However, RA is limited to single-image transformations. We therefore explore the use of an extended RA policy search with additional transformations that are more specific to medical imaging, and expand its capabilities to include mixed-example image examples to include artificial data in model training.

## 2 Methods

In this section we describe our proposed framework for augmentation policy search, depicted in Figure 1 which consists of three key procedures i) data generation, ii) data augmentation, iii) policy searching and interpretation.

**Mixed-Example Data Augmentation** The original dataset  $D = \{(X_i, Y_i)\}$  consists of a series of  $i$  ultrasound frames  $X$  and their associated classes  $Y$ . We first generate a paired dataset  $D_{paired} = \{(x_1, x_2)_{\frac{1}{2}}, (y_1, y_2)_{\frac{1}{2}}\}$  by pairing examples from different classes. Examples of artificial data are then generated using non-linear methods [25], which are found to be more effective than linear intensity averaging (mix-up)[27]. As illustrated in Figure 2, instead of pixel-wise averaging, the bottom  $\lambda_1$  fraction of image  $x_1$  is vertically concatenated with the top  $1 - \lambda_1$  fraction of image  $x_2$ . Similarly, the right  $\lambda_2$  fraction of image  $x_1$  is horizontally concatenated with the left  $1 - \lambda_2$  fraction of image  $x_2$ . After the concatenations, the resulted images are combined to produce an image  $\tilde{x}$  in which the top-left is from  $x_1$ , the bottom right is from  $x_2$ , and the top-right and bottom-left are mixed between the two. Moreover, instead of linear pixel averaging, we treat each image as a zero-mean waveform and normalize mixing coefficients with image intensity energies [26]. Formally, given initial images  $x_{1,2}$  with image intensity means and standard deviations of  $\mu_{1,2}$  and  $\sigma_{1,2}$ , the generated artificial



**Fig. 1.** Overview of our proposed learning framework.

mixed-example image  $\tilde{x}$  is:

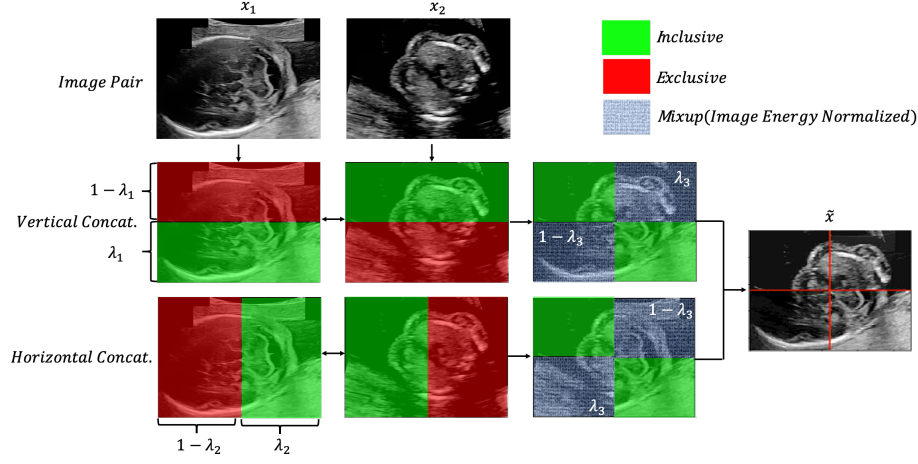
$$\tilde{x} = \begin{cases} x_1(i, j) - \mu_1 & \text{if } i \leq \lambda_1 H \text{ and } j \leq \lambda_2 W \\ \frac{c}{\phi} [x_1(i, j) - \mu_1] + \frac{1-c}{\phi} [x_2(i, j) - \mu_2] & \text{if } i \leq \lambda_1 H \text{ and } j > \lambda_2 W \\ \frac{1-c}{\phi} [x_1(i, j) - \mu_1] + \frac{c}{\phi} [x_2(i, j) - \mu_2] & \text{if } i > \lambda_1 H \text{ and } j \leq \lambda_2 W \\ x_2(i, j) - \mu_2 & \text{if } i > \lambda_1 H \text{ and } j > \lambda_2 W \end{cases}$$

where  $c$  is the mixing coefficient  $(1 + \frac{\sigma_1}{\sigma_2} \cdot \frac{1-\lambda_3}{\lambda_3})^{-1}$  and  $\phi$  is the normalization term defined as  $\sqrt{c^2 + (1-c)^2}$ . The row index and column index is represented by  $i, j$  and the height and width of the images are represented by  $H, W$ .

We sample  $\lambda_{1,2,3} \sim \text{Beta}(m/10, m/10)$  where  $m$  is a learnt hyperparameter varied from 0-10. As  $m$  approaches 10,  $\lambda$  values are more uniformly distributed across 0-1 and artificial images are more interpolated. The ground truth label after interpolation is determined by the mixing coefficients and can be calculated with:

$$\tilde{y} = (\lambda_3 \lambda_1 + (1 - \lambda_3) \lambda_2) y_1 + (\lambda_3 (1 - \lambda_1) + (1 - \lambda_3) (1 - \lambda_2)) y_2$$

**Original Data Augmentation** Augmentation transformations are then applied to the mixed images. Inspired by [7], we do not learn specific magnitudes and probabilities of applying each transformation in a given transformation list.



**Fig. 2.** The procedure for non-linear mixed-example data augmentation using an image pair and the final artificial mixed-example image.

Each augmentation policy is instead defined only by  $n$ , which is the number of transformations from the list an image undergoes, and  $m$ , which is the magnitude distortion of each transformation. Note that  $m$  is a shared hyperparameter with the mixed-example augmentation process. We investigate the inclusion in the transformation list transformations commonly used in ultrasound image analysis augmentation: i) grid distortions and elastic transformation [4] and ii) speckle noise [2]. The transformation list then totals 18 transformations, examples of which can be seen in Figure 3.

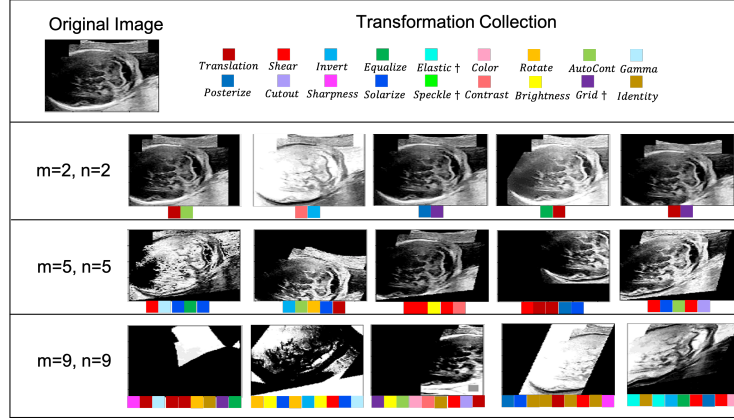
**Optimization** We define  $f$  and  $\theta$  as a convolutional neural network (CNN) and its parameters. As depicted in Figure 1, we train a CNN with the augmented mini-batch data  $\tilde{x}^i$  and obtain the predicted output class scores  $f_\theta(\tilde{x}^i)$ . These are converted into class probabilities  $p(\tilde{x}^i)$  with the softmax function. The KL-divergence between  $f_\theta(\tilde{x}^i)$  and  $\tilde{y}^i$  is then minimized with back-propagation and stochastic gradient descent

$$L = \frac{1}{B} D_{KL}(\tilde{y}^i \parallel p(\tilde{x}^i)) = \frac{1}{B} \sum_{i=1}^B \sum_{j=1}^C \tilde{y}_j^i \log \frac{\tilde{y}_j^i}{\{p(\tilde{x}^i)\}_j}$$

where  $B$  is the batch size,  $C$  is the number of classes and  $L$  is the loss.

Due to the limited search space, the hyperparameters  $n$  and  $m$  that produce the optimum classification performance can then be found using grid search as seen in Figure 1. The best performing  $m, n$  tuple is then used during final model evaluation.

**Quantifying Augmentation Effects** Next, we investigate how augmentation improves model generalization and quantify how different augmentation policies



**Fig. 3.** Examples of how each transformation and augmentation policy affect input images. Each color represents a transformation, and augmented image is transformed using a number of transformations ( $n$ ) at a magnitude of ( $m$ ). †: our additional transformations.

affect augmented data distributions and model performance. We adopt a two dimensional metric - affinity and diversity [13] to do this. Affinity quantifies the distribution shift of augmented data with respect to the unaugmented distribution captured by a baseline model; the diversity quantifies complexity of the augmented data. Given training and validation datasets,  $D_t$  and  $D_v$ , drawn from the original dataset  $D$ , we can generate an augmented validation dataset  $D(m, n)_v'$  derived from  $D_v$  using  $m, n$  as hyperparameters for the augmentation policy. The affinity  $A$  for this augmentation policy is then:

$$A = \mathbb{E}[L(D_v')] - \mathbb{E}[L(D_v)]$$

where  $\mathbb{E}[L(D)]$  represents the expected value of the loss computed on the dataset  $D$  loss of a model trained on  $D_t$ .

The diversity,  $D$ , of the augmentation policy  $a$  is computed on the augmented training dataset  $D_t'$  with respect to the expected final training loss,  $L_t$ , as:

$$D = \mathbb{E}[L(D_t')]$$

Intuitively, the greater the difference in loss between an augmented validation dataset and the original dataset on a model trained with unaugmented data, the greater the distribution shift of the augmented validation dataset. Similarly, the greater the final training loss of a model on augmented data, the more complexity and variation there is in the final augmented dataset.

### 3 Experiments and Results

We use a clinically acquired dataset consisting of ultrasound second-trimester fetal examinations. A GE Voluson E8 scanner was used for ultrasound image

acquisition. For comparison with previous work [8,17], fetal ultrasound images were labelled into 14 categories. Four cardiac view classes (4CH, 3VV, LVOT, RVOT) corresponding to the four chamber view, three vessel view, left and right ventricle tracts respectively; the brain transcerebellar and transventricular views (TC, TV); two fetal spine sagittal and coronal views (SpineSag, SpineCor); the kidney, femur, abdominal circumference standard planes, profile view planes and background images. The standard planes from 135 routine ultrasound clinical scans were labelled, and 1129 standard plane frames were extracted. A further 1127 background images were also extracted and three-fold cross validation was used to verify the performance of our network.

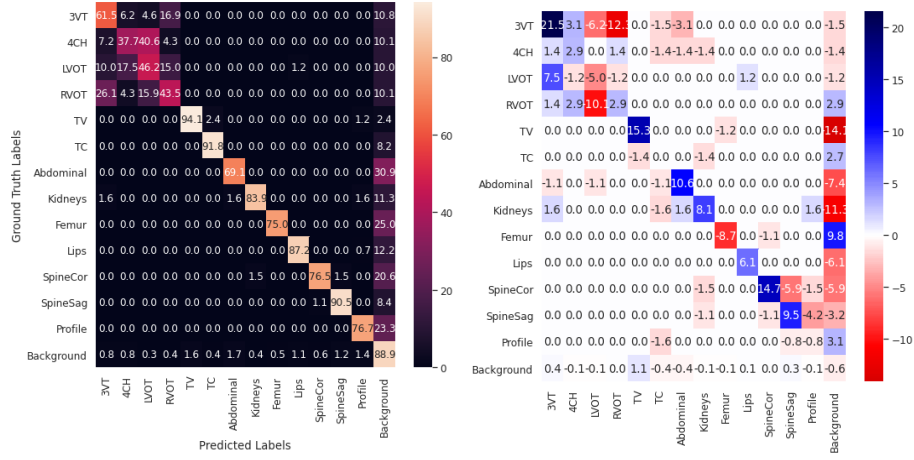
**Network Implementation** The performance of the SE-ResNeXt-50 [15] backbone is well validated on natural images and therefore used. Networks were trained with an SGD optimizer with learning rate of  $10^{-3}$ , a momentum of 0.9 and a weight decay of  $10^{-4}$ . Networks were trained for a minimum of 150 epochs, and training was halted if there was 20 continuous epochs without improvement in validation accuracy. Models were implemented with PyTorch and trained on a NVIDIA GTX 1080 Ti. Random horizontal and vertical flipping were used in all RA policies as a baseline augmentation. Models were trained with a batch size of 50. We evaluated the performance of networks trained with augmentation policies with values of  $m, n$  where  $m, n = \{1, 3, 5, 7, 9\}$  and used a simple grid search for augmentation strategies to find optimal  $m, n$  values.

**Table 1.** Results for standard plane detection (mean  $\pm$  std %). The best performing augmentation strategies are marked in **bold** for each metric.

	Random Initialization				Initialized with external data			
	No Aug.	SN Pol.	RA [7]	Mix. RA (ours)	Siam. Init. [17]	Saliency [8]	SonoNet [8,3]	SonoNet + Mix.RA (ours)
Precision	56.5 $\pm$ 1.2	70.4 $\pm$ 2.3	74.7 $\pm$ 1.8	<b>75.1<math>\pm</math>1.8</b>	75.8 $\pm$ 1.9	79.5 $\pm$ 1.7	82.3 $\pm$ 1.3	<b>86.3<math>\pm</math>1.3</b>
Recall	55.1 $\pm$ 1.2	64.9 $\pm$ 1.6	72.2 $\pm$ 2.3	<b>73.4<math>\pm</math>1.9</b>	76.4 $\pm$ 2.7	75.1 $\pm$ 3.4	<b>87.3<math>\pm</math>1.1</b>	85.1 $\pm$ 1.3
F1-Score	55.4 $\pm$ 1.2	67.0 $\pm$ 1.3	72.8 $\pm$ 1.8	<b>74.0<math>\pm</math>1.8</b>	75.7 $\pm$ 2.0	76.6 $\pm$ 2.6	84.5 $\pm$ 0.9	<b>85.4<math>\pm</math>1.5</b>

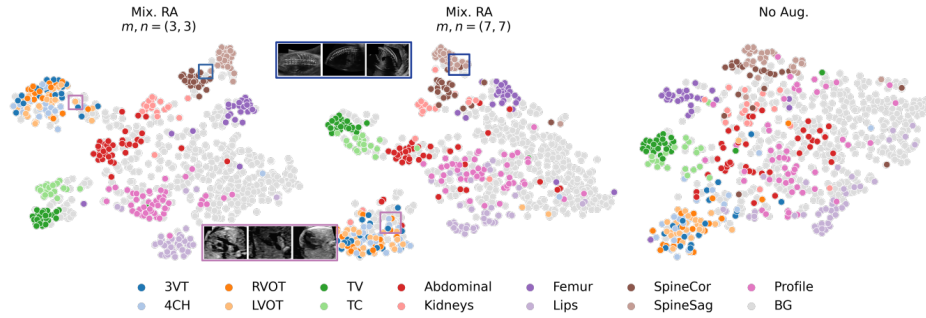
**Results on CNNs with Random Initialization** The effectiveness of our mixed-example augmentation search policy algorithm (*Mix. RA*) on SE-ResNeXt-50 models that are randomly initialized is compared with models trained with the baseline RandAugment (*RA*) augmentation search policy; a commonly used augmentation strategy (*SN Pol.*) in line with that found in [8], where images are augmented with random horizontal flipping, rotation  $\pm 10^\circ$ , aspect ratio changes  $\pm 10\%$ , cropping and changing brightness  $\pm 25\%$  and image cropping 95 – 100%; and no augmentation (*No. Aug.*).

From Table 1 we can see that the proposed method *Mix. RA* outperforms all other augmentation methods on every metric with random network initialization, including the baseline *RA* augmentation policy search.



**Fig. 4.** Confusion matrix for *Mix. RA* (left) and the difference in precision between *Mix. RA* and *SN Pol.*

To better understand how *Mix. RA* outperforms naive augmentation, we show the confusion matrix for the best performing model *Mix. Aug* and the difference in confusion matrix between it and naive augmentation *SN Pol.* We find that in general, heart plane classification is improved with a mean increase in macro F1-Score of 4.0%. Other anatomical planes with the exception of the femur plane also show consistent increases in performance with movement of probability mass away from erroneously classified background images to the correct classes suggesting the model is able to recognize greater variation in each anatomical class.



**Fig. 5.** t-SNE embeddings for different augmentation policies. The boxes represent the regions from which example images are taken from. The blue and purple boxes contain examples of the (SpineCor, SpineSag, and BG) and (3VT, RVOT, BG) classes respectively taken from the highlighted positions in the t-SNE embedding. Best viewed in color.



The t-SNE embeddings of the penultimate layer seen in Figure 5 can also be used to visualize the differences in feature spaces in trained networks with different augmentation policies. Compared to the model trained with no augmentation, our best performing policy leads to a better separation of the abdominal and profile standard planes from the background class as well as clearer decision boundaries between anatomical classes. The two brain views (TC, TV) and the demarcation of the boundary between the kidney view and abdominal plane view is also better defined.

Between the best performing policy  $m, n = (5, 3)$  and an underperforming policy  $m, n = (7, 7)$ , we find that profile planes are better separated from the background class and the abdominal planes better separated from the kidney views, which suggests that the optimum  $m, n$  value increases network recognition of salient anatomical structures. However, in all three cases, the cardiac views remain entangled. This can be attributed to the difficulty of the problem, as even human expert sonographers cannot consistently differentiate between different cardiac standard plane images. We also find that the background class also contains examples of the anatomies in each class, but in sub-optimal plane views, which leads to confusion during classification. This difficulty is illustrated in example background images in Figure 5 where the heart and spine are visible in the BG class.

**Pre-trained Networks** We also compare our work to methods where networks were initialized with external data as seen in the right of Table 1. Baseline methods of self-supervised pre-training using video data [17] (*Siam. Init.*), multi-modal saliency prediction [8] (*Saliency*) and Sononet (*Sonet*) [3] were used to initialize the models and the models fine-tuned on our dataset. Using our augmentation policy during fine-tuning of a pre-trained SonoNet network further increased the performance of standard plane classification with an increase in final F1-score of 0.9% when  $m, n = (5, 1)$ . This reduction in optimum transformation magnitude may be due to the change in network architecture from SE-ResNeXt-50 to a Sononet, as the smaller Sononet network may not be able to capture representations the former is able to. Furthermore, we find that augmentation policy with a randomly initialized network *Mix. RA* approaches the performance of the *Siam. Init.* and *Saliency* pre-trained networks. This is despite the fact that the *Siam. Init.* requires additional 135 US videos for network self-supervised initialization, and *Saliency* required external multi-modal data in the form of sonographer gaze.

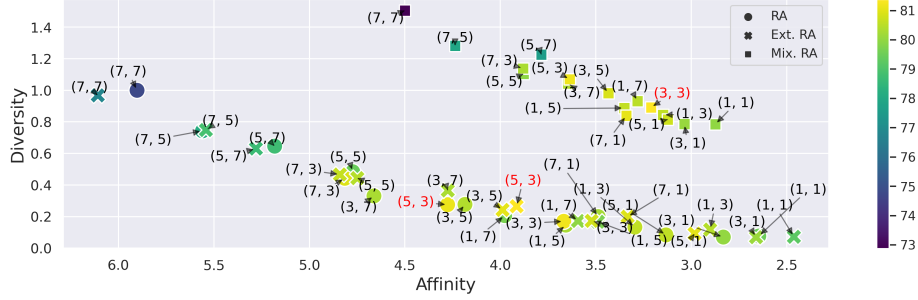
**Ablation Study** To better understand the individual contributions to the *Mix. RA* augmentation search policy, we show the results of an ablation study on the components of *Mix. RA* in Table 2.

It can be seen that both including Speckle noise transformations and deformation (Grid, Elastic) transformations lead to increased classifier performance for standard plane classification of +0.1% and 0.3% respectively with further improvement when both are combined together with *Ext. RA*. We find that both

**Table 2.** Ablation study on the individual components of our *Mix. RA* policy search algorithm on training of a randomly initialized CNN for ultrasound standard plane detection. All metrics are macro-averages due to the class imbalance. The *Linear Mix. RA* is included as a baseline mixed-example augmentation strategy.

	No Aug.	SN Pol.	RA	RA + Speckle ①	RA + Deform. ②	Ext. RA ① + ②	Linear Mix. RA ① + ②	Non-Linear Mix. RA ① + ②
Precision	56.5±1.2	70.4±2.3	73.9±1.7	74.6±1.7	74.6±1.7	74.0±2.4	74.6±1.6	<b>75.1±1.8</b>
Recall	55.1±1.2	64.9±1.6	72.9±2.0	72.1±1.8	72.9±1.8	<b>74.5±1.3</b>	73.2±1.5	73.4±1.9
F1-Score	55.4±1.2	67.0±1.3	72.8±1.8	72.9±1.7	73.2±1.2	73.6±1.3	73.7±1.6	<b>74.0±1.8</b>

*RA* and *Ext. RA* had an optimal  $m, n = (5, 3)$ , suggesting that the magnitude ranges for our additional transformations are well matched to the original transformation list. This performance increase is further boosted when mixed-example augmentation is introduced on top of *Ext. RA*, with non-linear mixed-example augmentations outperforming a linear mix-up based method.



**Fig. 6.** Affinity and diversity metrics for *RA*, *Ext. RA* and *Mix. RA* augmentation policy search algorithms. Each point represents a  $(m, n)$  value in hyperparameter space. Best performing  $m, n$  values are highlighted in red for each policy search method. Color represents final F1-Score on a randomly initialized CNN.

**Affinity and Diversity** The affinity and diversity of the augmentation policies is shown in Fig. 6. We find that there exists a “sweet spot” of affinity and diversity using non-mixed class augmentation strategies at an affinity distance of  $\sim 3.8$  and diversity of  $\sim 0.25$  which maximized model performance, corresponding to  $m, n = (5, 3)$ . At high  $m, n$  values, affinity distance is too high and the distribution of the augmented data is too far away from the validation data, decreasing model performance. However, at low  $m, n$  values, the diversity of the augmented data decreases and the model sees too little variation in input data.

It can also be seen that the *Mix. RA* augmented dataset showed a reduced affinity distance to the original dataset than the *Ext. RA* dataset at the same  $m, n = (5, 3)$  value, implying that our proposed transforms shifts augmented im-

ages to be more similar to the original images. Moreover, using a mixed-example data augmentation strategy drastically increased diversity for any given value of data distribution affinity, which improved final model performance. The best performing mixed-example augmentation policy  $m, n = (3, 3)$  reduced the magnitude of each transformation compared to the optimal non-linear augmentation policy. This suggests that mixed-example augmentation acts to increase the diversity of training images which reduces the magnitude required during further processing.

## 4 Conclusion

The results have shown that we can use a simple hyper-parameter grid search method to find an augmentation strategy that significantly outperforms conventional augmentation methods. For standard plane classification, the best performing augmentation policy had an average increase in F1-Score of 7.0% over that of a standard ultrasound model augmentation strategy. Our augmentation policy method is competitive with the *Siam. Init.* [17] despite the latter needing additional external data for pre-training. Our method also improves the performance of a Sononet pre-trained model when fine-tuned using our augmentation policy search method. From t-SNE plots and confusion matrix differences, we can see that the performance increase is from better classification of background-labelled planes, despite the difficulty in heart plane classification. It should be noted that a large degree of misclassification was due to standard planes being mis-classified into background images or vice-versa, and qualitative evaluation of t-SNE clusters show that this was due to background images being examples of class anatomies. The ablation study also shows that our additional transformations improve model performance, and non-linear mixed-example augmentation further improves classification. The evaluation using affinity and diversity indicate that the hyperparameter search involves a trade-off between diversity and affinity. We find that using non-linear mixed-class data augmentation before transformations drastically increases diversity without further increasing affinity distance between the training data and augmented data, which helps explain the increase in model performance. In conclusion, we have shown that our augmentation policy search method outperforms standard manual choice of augmentation policy. The augmentation policy search method presented does not have any inference-time computational cost, and has the potential to be applied in other medical image settings where training data is insufficient and costly to acquire.

## References

1. Asaduz, Z., Sang, H.P., et al.: Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images. ICARS (2020)
2. Bargsten, L., Schlaefel, A.: Specklegan: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. ICARS (2020)

3. Baumgartner, C.F., et al.: SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound. *IEEE TMI* (2017)
4. Buslaev, A., Iglovikov, V.I., et al.: Albumentations: Fast and flexible image augmentations. *MDPI* (2020)
5. Cecilia, S., Michael, J.D.: Improved mixed-example data augmentation. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2019)
6. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: *CVPR* (2019)
7. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: *CVPR* (2020)
8. Droste, R., et al.: Ultrasound Image Representation Learning by Modeling Sonographer Visual Attention. *IPMI* (2019)
9. Dubost, F., Bortsova, G., et al.: Hydranet: Data Augmentation for Regression Neural Networks. *MICCAI* (2019)
10. Eaton-Rosen, Z., Bragman, F., et al.: Improving data augmentation for medical image segmentation. In: *MIDL* (2018)
11. Frid-Adar, M., Diamant, I., et al.: Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing* (2018)
12. Gan, Z., Henao, R., Carlson, D., Carin, L.: Learning deep sigmoid belief networks with data augmentation. *Proceedings of Machine Learning Research (PMLR)* (2015)
13. Gontijo-Lopes, R., Smullin, S.J., Cubuk, E.D., Dyer, E.: Affinity and diversity: Quantifying mechanisms of data augmentation (2020)
14. Ho, D., Liang, E., et al.: Population based augmentation: Efficient learning of augmentation policy schedules. In: *ICML* (2019)
15. Hu, J., et al.: Squeeze-and-Excitation Networks. *IPAMI* (2020)
16. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. *AMIA* (2017)
17. Jiao, J., et al.: Self-supervised representation learning for ultrasound video. In: *IEEE 17th International Symposium on Biomedical Imaging* (2020)
18. Jin, D., Xu, Z., et al.: CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. in *MICCAI* (2018)
19. Lim, S., Kim, I., Kim, T., Kim, C., Kim, S.: Fast autoaugment. in *Advances in Neural Information Processing Systems (NeurIPs)* (2019)
20. Luke, T., Geoff, N.: Improving deep learning using generic data augmentation. *IEEE Symposium Series on Computational Intelligence* (2018)
21. Nalepa, J., et al.: Data augmentation via image registration. In: *ICIP* (2019)
22. Ohno, H.: Auto-encoder-based generative models for data augmentation on regression problems (2019)
23. Ryo, T., Takashi, M.: Data augmentation using random image cropping and patches for deep CNNs. *IEEE TCSVT* (2020)
24. Summers, C., Dinneen, M.J.: Improved mixed-example data augmentation. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2019)
25. Summers, C., Dinneen, M.J.: Improved mixed example data augmentation. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2019)
26. Tokozume, Y., Ushiku, Y., Harada, T.: Between-class learning for image classification. In: *CVPR* (2018)
27. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: *ICLR* (2018)
28. Zhao, A., Balakrishnan, G., et al.: Data augmentation using learned transformations for one-shot medical image segmentation. *CVPR* (2019)