Daniel Wulff*, Mohamad Mehdi, Floris Ernst, Jannis Hagenah

Cross Data Set Generalization of Ultrasound Image Augmentation using Representation Learning: A Case Study

Abstract: Data augmentation is a common method to make deep learning assessible on limited data sets. However, classical image augmentation methods result in highly unrealistic images on ultrasound data. Another approach is to utilize learning-based augmentation methods, e.g. based on variational autoencoders or generative adversarial networks. However, a large amount of data is necessary to train these models, which is typically not available in scenarios where data augmentation is needed. One solution for this problem could be a transfer of augmentation models between different medical imaging data sets. In this work, we present a qualitative study of the cross data set generalization performance of different learning-based augmentation methods for ultrasound image data. We could show that knowledge transfer is possible in ultrasound image augmentation and that the augmentation partially results in semantically meaningful transfers of structures, e.g. vessels, across domains.

Keywords: 2D Ultrasound, Variational Autoencoder, Generative Adversarial Network, Latent Space

https://doi.org/10.1515/cdbme-2021-2193

1 Introduction

In recent years, deep learning showed outstanding performance in medical image analysis [1]. However, the convergence of a deep neural network to an adequate and generalized state relies on the availability of large amounts of data covering the full variance of possible samples. The collection of such a vast data set is challenging, especially in the medical domain. Therefore, a common technique to artificially enlarge a data set is data augmentation [2], i.e. the perturbation of existing data points to synthesize new samples. Even though being very common in image analysis, typical simplistic small, random image transformations used for augmentation might lead to highly unrealistic synthetic samples in the medical imaging domain. This specifically holds for ultrasound images where rotations, sheering operations or contrast adjustments will lead to unrealistic images due to the physics of ultrasound imaging. To achieve more realistic synthetic samples, several approaches were proposed to utilize deep learning for data augmentation [3, 4]. Typically, these approaches rely on learning a latent representation of the data set from which new, synthetic images can be sampled. The representation can be learned e.g. using a variational autoencoder (VAE) [5] or a generative adversarial network (GAN) [6]. It could be shown that learning-based augmentation models are capable of providing highly realistic samples [7]. However, the augmentation model has to be trained on a sufficiently large data set. Since an adequate amount of data often is rare in the medical domain there is a need of an augmentation model that can be transferred between application domains within the field of medical image analysis, e.g. different scenarios of ultrasound imaging. Such a model should be usable for augmenting a small, unseen data set with good results. However, there is a lack of studies on cross data set generalization performance of learning-based data augmentation methods in medical imaging. In this work, we present a qualitative analysis of the cross data set generalization performance in the scope of ultrasound image augmentation. Therefore, we train and evaluate two different learning-based augmentation models on three different ultrasound image data sets in different transfer scenarios.

Contribution of this Work

We see two main contributions of this study. First, to the best of our knowledge, this study presents the first evaluation of the

***Corresponding author: Daniel Wulff, Mohamad Mehdi, Floris Ernst, Jannis Hagenah:** Institute for Robotics and Cognitive Systems, University of Luebeck, Ratzeburger Allee 160, 23562 Luebeck, Germany, e-mail: wulff@rob.uni-luebeck.de
Jannis Hagenah: Computational Health Informatics Lab, Institute of Biomedical Engineering, University of Oxford, Old Road Campus Research Building, Oxford OX3 7DQ, UK, e-mail: jannis.hagenah@eng.ox.ac.uk

Open Access. © 2021 The Author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

cross data set generalization performance of learning-based augmentation models in the scope of medical image analysis. Since data augmentation is a very common technique and most of the data sets are limited in their size, the qualitative results are of high interest to the medical image analysis community. Second, the models developed and trained in the scope of this study partially provided promising results. Therefore, our pretrained models might be relevant for applied ultrasound image analysis research.

2 Methods

In this section, the two methods for generating image augmentations as well as the ultrasound data and the experiments to evaluate the proposed approaches are presented in detail

2.1 Augmentation Methodology

Image augmentation refers to an extansion of an existing data set by generating synthetic data on the basis of this data set. The main idea of the proposed method is to identify abstract representations of the data, encode a sample and decode another sample that is similar to the real one in terms of this representation. It is assumed that similarity between two samples corresponds to a small distance between them in the abstract representation space.

Two representation learning approaches are used to learn a latent space of the given data set: A variational autoencoder (VAE) [6] and a generative adversarial network (GAN) specifically designed for augmentations [7]. After training the models an image sample is encoded into the latent space. Then, a gaussian distribution is centered around this point via setting its mean to the position and an augmented data point is generated by sampling from this distribution. The augmented image can be synthesized by propagating this point through the decoder. The variance σ of the gaussian distribution describes the degree of similarity between the real and the augmented sample and is a hyperparameter of the method.

2.2 Ultrasound Data

As presented in Table 1, in this study three different ultrasound data sets are used containing femoral artery images of a phantom leg [8], female breast images [9] and images of fetus head circumferences [10]. Note that the fetus images show full cone-shaped ultrasound images while breast and phantom images show region-of-interest sections. For the experiments the data sets are divided into one source data set D^{source} containing the phantom data as well as two target data sets D_1^{target} and D_2^{target} containing the breast and fetus data, respectively. For one experiment, auxiliary target data sets $\widetilde{D}_1^{target}$ and $\widetilde{D}_2^{target}$ are generated containing 100 randomly selected images from the corresponding target data sets. In a preprocessing step all ultrasound images are resized into a unitary shape of 128×128 and the gray scale values are normalized to the range [0,1] for VAE and [-1,1] for GAN, respectively.

Table 1: Information about the data sets used for training and augmentation

Notation	Content	Amount	Usage	Reference
D^{source}	vessel phantom	1950	source	[8]
D_1^{target}	female breast	780	target	[9]
D_2^{target}	fetus head	1334	target	[10]

2.3 Experiments

Three different experiments are performed whose general process is summarized in Figure 1. First, the VAE and GAN models are trained using three different approaches. Then, the trained models are used in combination with a target data set to generate augmentations of this target data set. For this, the sample representation is manipulated by sampling from a gaussian distribution with variance σ around a real sample as described in section 2.1. In all experiments σ is chosen as 0.6, 1.0, 1.6 and 2.0.

First, a baseline is generated to investigate whether the models are able to generate meaningful images performing an *intra domain* experiment. Both, training and augmentation is performed using the same domain which is D_1^{target} or D_2^{target}.

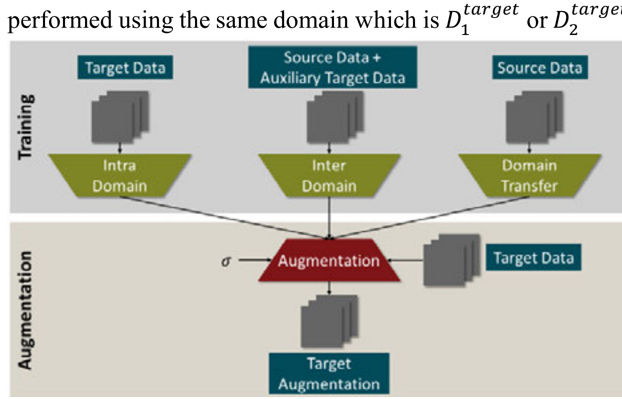

Figure 1: Visualization of training and augmentation process of the *intra domain, inter domain* and *domain transfer* experiments.

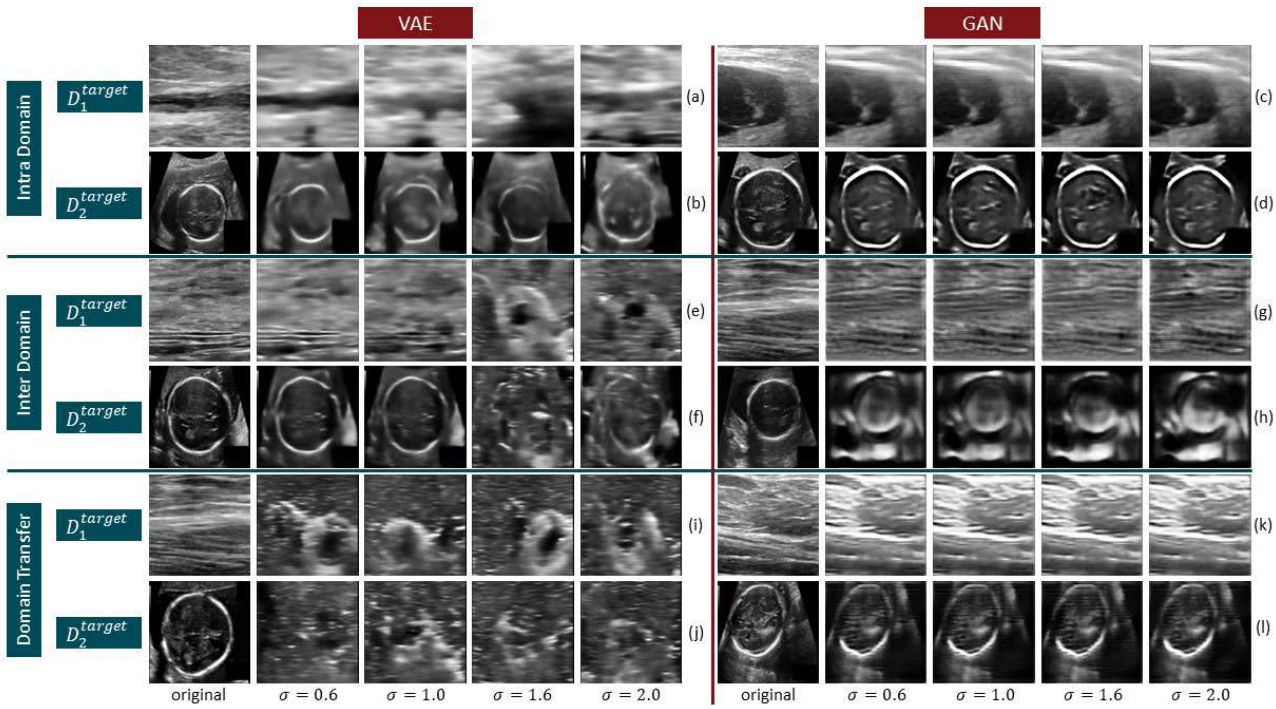

Figure 2: Examples for the augmentations generated in the *intra domain* (top), *inter domain* (middle) and *domain transfer* (bottom) experiments for the target data sets D_1^{target} (breast) and D_2^{target} (fetus) using VAE (left) and GAN (right) at different σ values.

Second, an *inter domain* experiment is performed. It is tested whether a model can augment images of a target data set, though the model is trained mainly on the source data set. For model training D^{source} plus an auxiliary target data set $\widetilde{D}^{target}$ are used. Augmentation is performed using the remainder of the corresponding target data set. The third experiment is a *domain transfer* experiment where training is performed only using D^{source} and augmentation is done using the target data sets. This experiment investigates whether the models are capable to augment images of a target data set though this model was trained on another domain.

3 Results and Discussion

The image augmentations generated in the experiments are evaluated qualitatively, examples for each experiment and model are visualized in Figure 2. In the *intra domain* experiment both models are capable to generate variations of the origin image. In general, the VAE creates images that are strongly blurred. With high σ values of 1.6 and 2.0 the VAE is capable to generate breast augmentations that look highly different than the origin, however, these augmentations seem to be realistic (Figure 2a). Though, fetus augmentations with high σ values are not realistic due to the fact that the fetus head is distorted (Figure 2b). In contrast to VAE, GAN augmentations are less blurry but the augmentations look more similar to the origin even with high σ values. Though, in GAN augmentations several structures look clearer and are distinguished compared to the origin. In addition, GAN fetus augmentations look realistic as there are no unrealistic distortions. The *intra domain* experiment serve as a baseline and indicate that this augmentation approach can lead to meaningful augmentations. The VAE generates blurry images but it can create augmentations that are transformed in a larger scale compared to the GAN. However, this characteristic does also lead to unrealistic augmentations which is not observed in GAN augmentations.

In the *inter domain* experiment VAE augmentations generated with σ values of 0.6 and 1.0 look similar to the ones in the *intra domain* experiment. However, with higher σ values the VAE generates augmentations showing a source data image instead of a target data image in multiple cases in both target data sets (Figure 2e,f). As the VAE was mainly trained with the source data set, this was expected. When the distance to a target sample in latent space gets too high, the augmentation becomes a source data sample. Thus, in this approach $\sigma = 1.0$ should be chosen at maximum to get meaningful target data augmentations. Considering the breast augmentations, changes like structure appearance and deformation of existing structures can be identified (Figure

2e). In the GAN augmentations, the image contrast clearly changed compared to the origin. Though, the breast augmentations look realistic and even fine new structures can be identified in some augmentations (Figure 2g). However, the fetus augmentations do not appear to be realistic, even with low σ values as the fetus head is distorted or the structure is completely lost (Figure 2h). This could be explained with the different field of views in the ultrasound images and the different image contrast. Fetus images show a whole ultrasound image whereas phantom and breast images only show a section of an ultrasound image. Thus, phantom and breast images are more similar. However, even with high σ values the GAN generates target data augmentations. Other than the VAE, the GAN tries to generate an augmentation that is similar to the origin so that it is not possible to get source data augmentations as observed when using the VAE.

In the *domain transfer* experiment, the VAE is not capable to generate augmentations of any target data set. Instead, at all σ values the model generates augmentations of the source data set (Figure 2i,j). As the VAE learned a latent space representing variations of the source data set, using target data for augmentation equates to sampling randomly in latent space. Thus, the VAE is not suitable for *domain transfer* if training data only contains one domain. Training data should rather contain various domains to make the model capable to generalize over several domains. However, the GAN is capable to generate realistic augmentations of the origin for both target data sets. In breast augmentations the image contrast change on a small scale. In addition, in some breast augmentations new structures appear (Figure 2k) which is promising in terms of augmenting meaningful synthetic images. The fetus augmentations generated by the GAN look similar to the origin, but the images are strongly blurred (Figure 2l). The transformations in breast augmentations are clearly higher than in fetus augmentations which again could be explained by the higher similarity between phantom and breast data compared to phantom and fetus data.

4 Conclusion

In this study two methods for generating meaningful ultrasound image augmentations are proposed and investigated whether these methods are capable to generalize for different domains. The results show that both methods hold the potential but for generating a highly generalized model training should be performed on a larger and more variable data set. Though, promising behaviour as new structure appearance and structure deformation can be observed in this study. The results in the *inter domain* and *domain transfer* experiments indicate that it might be possible to use the proposed methods for augmenting an unknown data set from another domain only by performing fine tuning or even without any further model training. These are promising results as ultrasound data sets often are small and common augmentation techniques lead to unrealistic ultrasound images. In further studies we will create more generalized models by using various data sets for training and investigate its performance.

Author Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] Puttagunta M, Ravi S. Medical image analysis based on deep learning approach. Multimed Tools Appl, vol. 80, pp. 24365-24398, 2021

[2] Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications, Journal of Medical Imaging and Radiation Oncology, 2021

[3] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing, vol. 321, pp. 321-331, 2018

[4] Qin Z, Liu Z, Zhu P, Xue Y. A GAN-based image synthesis method for skin lesion classification, Computer Methods and Programs in Biomedicine, vol. 195, 2020

[5] Kingma DP, Welling M. Auto-Encoding Variational Bayes, CoRR, abs/1312.6114, 2014

[6] Antoniou A, Storkey A, Edwards H. Data Augmentation Generative Adversarial Networks, arXiv, 2018

[7] Bart E, Hegdé J. Deep Synthesis of Realistic Medical Images: A Novel Tool in Clinical Research and Training, Front. Neuroinform. Vol. 12, no. 82, 2018

[8] Haxthausen Fv, Hagenah J, Kaschwich M, Kleemann M, Garcia-Vázquez V, Ernst F. Robotized Ultrasound Imaging of the Peripheral Arteries – a Phantom Study. Current Directions in Biomedical Engineering, vol. 6, no. 1, 2020

[9] Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief, vol. 28, 2020

[10] Heuvel TLAvd, Bruijn Dd, Korte CLd, Ginneken Bv. Automated measurement of fetal head circumference using 2D ultrasound images, PLOS ONE, vol. 13, no. 8, 2018