Christian Kaiser*, Matthias Schaufelberger, Reinald Peter Kühle, Andreas Wachter, Frederic Weichel, Niclas Hagen, Friedemann Ringwald, Urs Eisenmann, Michael Engel, Christian Freudlsperger, and Werner Nahm

Generative-Adversarial-Network-Based Data Augmentation for the Classification of Craniosynostosis

https://doi.org/10.1515/cdbme-2022-1005

Abstract: Craniosynostosis is a congenital disease characterized by the premature closure of one or multiple sutures of the infant's skull. For diagnosis, 3D photogrammetric scans are a radiation-free alternative to computed tomography. However, data is only sparsely available and the role of data augmentation for the classification of craniosynostosis has not yet been analyzed.

In this work, we use a 2D distance map representation of the infants' heads with a convolutional-neural-network-based classifier and employ a generative adversarial network (GAN) for data augmentation. We simulate two data scarcity scenarios with 15 % and 10 % training data and test the influence of different degrees of added synthetic data and balancing underrepresented classes. We used total accuracy and F1-score as a metric to evaluate the final classifiers.

For 15 % training data, the GAN-augmented dataset showed an increased F1-score up to 0.1 and classification accuracy up to 3 %. For 10 % training data, both metrics decreased.

We present a deep convolutional GAN capable of creating synthetic data for the classification of craniosynostosis. Using a moderate amount of synthetic data using a GAN showed slightly better performance, but had little effect overall. The simulated scarcity scenario of 10 % training data may have limited the model's ability to learn the underlying data distribution.

Keywords: Generative adversarial network, classification, craniosynostosis, data augmentation

***Corresponding author: Christian Kaiser, Matthias Schaufelberger, Andreas Wachter, Werner Nahm,** Institute of Biomedical Engineering (IBT), Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, Karlsruhe, Germany, publications@ibt.kit.edu
Reinald Peter Kühle, Frederic Weichel, Michael Engel, Christian Freudlsperger, Department of Oral and Maxillofacial Surgery, Heidelberg University Hospital, Im Neuenheimer Feld 400, Heidelberg, Germany
Niclas Hagen, Friedemann Ringwald, Urs Eisenmann, Institute of Medical Informatics, Heidelberg University Hospital, Im Neuenheimer Feld 130.3, Heidelberg, Germany

1 Introduction

Craniosynostosis is a congenital condition characterized by premature ossification of skull sutures and has been linked to increased intracranial pressure which can lead to reduced neuropsychological development. The reported prevalence is four cases per 10,000 live births [1]. As the growth perpendicular to the closed suture is significantly decreased, craniosynostosis is accompanied by characteristic head shapes. Early diagnosis is crucial for surgical treatment with computed tomography being the gold standard for diagnosis exposing the child to harmful ionizing radiation. 3D stereophotogrammetry is a radiation-free alternative to quantify the head shape and can be used as the basis to distinguish craniosynostosis patients from healthy subjects or mild positional head deformities.

A common problem when working with clinical data is the imbalance and the limited amount of available data. This might hamper the performance of machine learning approaches which often need large datasets for robust classification. Generative models such as generative adversarial networks (GANs) have shown to improve the classification performance in medical applications and have been suggested (but not implemented) for craniosynostosis as well [2].

By expanding on a convolutional neural network (CNN)-based classifier for craniosynostosis, we construct a deep convolutional GAN to create synthetic samples and test the data-augmented classifier in a scarce training scenario. To the best of our knowledge, this is the first GAN employed specifically for craniosynostosis. We perform several test cases addressing dataset imbalance and limited data availability and evaluate the data-augmented classifier using accuracy and F1-score.

2 Methods

2.1 Dataset and distance map creation

The dataset was provided by the Department of Oral and Maxillofacial Surgery from the Heidelberg University Hospital. It consists of 367 photogrammetric scans of subjects with craniosynostosis and a control group. The age distribution and the number of cases for each class is shown in Figure 1.

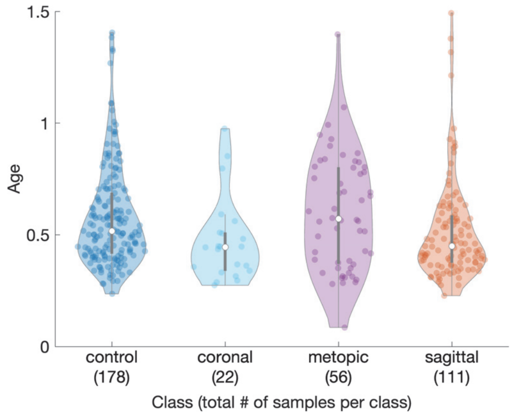


Fig. 1: Age distributions of the clinical dataset. Parentheses indicate the number of samples per class.

To enable a CNN-based classification, we used a ray casting approach [2] on the 3D scans and converted the distances to a 2D image [5]. We defined a center point using anatomical landmarks annotated by clinical experts and defined two angle directions similar to the horizontal coordinate system. We defined the azimuth angle ϕ in the interval $[0, 360°]$ and the altitude angle θ in the interval $[0, 90°]$, describing a semisphere. The angles formed the axes of the 2D image, while the extracted distances were extracted as an intensity value as represented in Figure2. We refer to [7] for an in-depth description of this method.

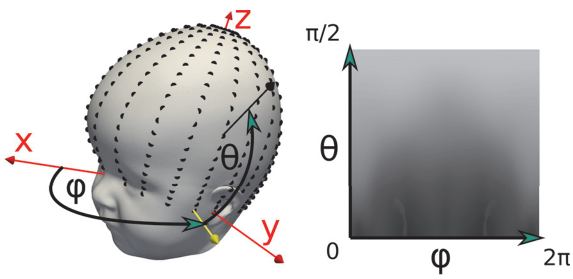


Fig. 2: Representation of the creation of a 2D distance map.

2.2 Generative adversarial network

GANs [3] are based on a zero-sum game. The game has two players, the generator **G**, producing synthetic samples, and the discriminator **D**, distinguishing between synthetic samples and real samples. **G** creates new data by sampling from a noise vector z that follows a uniform or Gaussian distribution. As **G** improves, the generator distribution P_g should match the distribution of the real data P_r and produce more and more realistic images.

We used a deep convolutional generative adversarial network (DCGAN), displayed in Figure3. It is an adaptation of the work of [6], only changing the number of layers to match the input of an image with the size of 224×224, which is the input size of the ResNet18. For **D**, this was achieved by adding an entire layer. It can be hard to find the Nash-equilibrium that solves the minimax game as proposed by [3] as GANs are known to be rather sensitive with respect to the hyperparameters. To partially avoid the problem, the Wasserstein distance was used in combination with a gradient penalty that enforced convergence as proposed by [4] which punished high gradients and forced **D** to a part of a set of K-Lipschitz functions. This resulted in the following minimax game:

$$\min_{G} \max_{D\, \in \mathbf{W}} V(D,G) = \mathbb{E}_{x\sim P_r}[D(x)] - \mathbb{E}_{x\sim P_g}[D(x)] - \lambda_{GP}\mathbb{E}_{\dot{x}\sim P_{\dot{x}}}[||\nabla D(\dot{x})||_2 - 1)^2] \quad (1)$$

λ_{GP} is the hyperparameter that punishes high gradients and $\dot{x}$ is a value interpolated from generated and real samples. This interpolation could be described with the following equation:

$$\begin{aligned} \dot{x} &= t \cdot x_r + (1-t) \cdot x_g \\ \text{with } t &\sim U(0,1), \\ x_r &\sim P_r \\ \text{and} \quad x_g &\sim P_g \end{aligned} \quad (2)$$

Instead of training a GAN for each group, we used a conditional GAN as the overall shape of the head is similar in the sense that it is close to a semi-ellipsoid. The conditional part was implemented by using an embedding marked as red in Figure3. In the case of **G**, the additional information was added by appending the embedding to the noise vector z. For **D**, this was achieved by adding an entire layer. The conditional GAN was trained with the hyperparameters in Table 1. New 2D images could then be created by sampling a noise vector based on the Gaussian distribution and by providing the label of the desired image.

2.3 Test scenarios

To ensure that we were in a case of data scarcity and to test what could be achievable with more data, the ratio between training and test data was gradually increased in favor of the test data. We defined data scarcity as the point when the overall

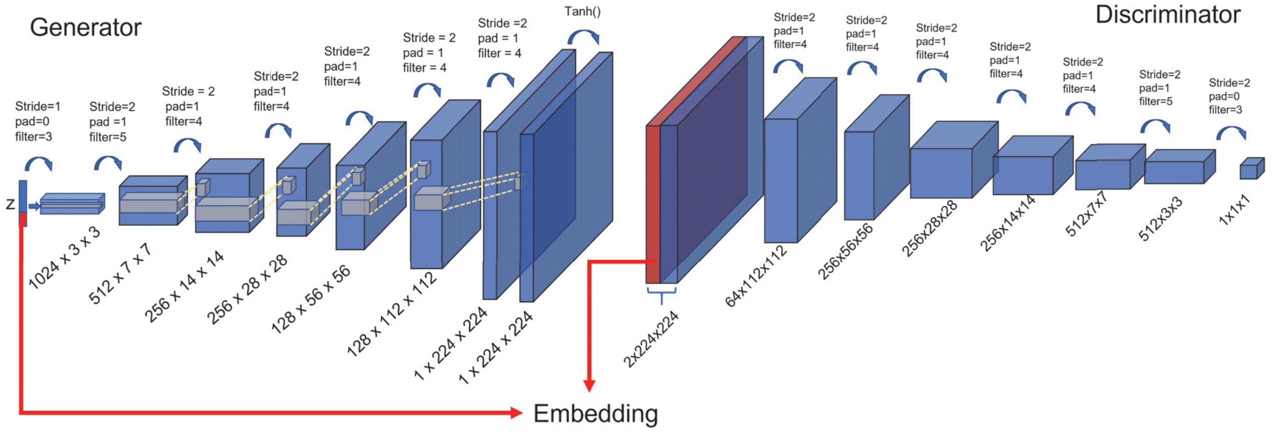


Fig. 3: Conditional DCGAN used for the sampling of synthetic data trained with the original maps.

Tab. 1: Hyperparameters for the training of the DCGAN

Gradient penalty	$\lambda_{GP} = 10$
Iteration used to train the discriminator	10
Batch-size	10
Number of epochs	50
Size of the noise vector z	100
Optimizer discriminator	Adam(β_1 = 0, β_2 = 0.9)
Learning rate	10^{-4}

performance decreased substantially compared to the starting ratio of 80 % training and 20 % test data. For the classification, we chose a ResNet18 model pretrained on ImageNet with the hyperparameters listed in Table 2.

Tab. 2: CNN hyperparameters for the classification of 2D maps.

Training approach	Fine-tuning of pre-trained *Resnet18*
Optimizer	Adam
Learning rate	$\alpha = 2 \cdot 10^{-3}$
Weight decay	$\gamma = 10^{-4}$
Number of epochs	$n = 150$

We defined four experiments to evaluate how synthetic images in the scenario of data scarcity for data augmentation affected the CNN-classifier. The tests are summarized in Table 3. The first two experiments added data proportional to existing data, assuming that the dataset did not need to be balanced in order to achieve a better classification performance. The last two scenarios focused on balancing the dataset with different amounts of synthetic data. As evaluation criteria we used overall accuracy and F1-score to explicitly take into account the imbalance of the dataset.

Tab. 3: Summary of the four test scenarios during data scarcity.

Test scenario	Description
Moderately augmented	Add 100 % synthetic images to each class for a similar number of synthetic and real data.
Highly augmented	Add 1000 % synthetic images to each class to have mostly synthetic data.
Balanced	Balance the dataset, without adding synthetic images to the most frequent class.
Balanced and augmented	Increase the maximum number of images to ten times the number of images in the most represented class and balance the dataset.

3 Results

The final accuracy for each split and the F1-score is displayed in Figure 4. The strongest drop in the two metrics could be observed for 85 %. Thus we assumed the case of data scarcity for a ratio of 85 % test data and considered the case of 90 % test data as a severe case of data scarcity. It should be noted that in all scenarios the accuracy was above 80 %.

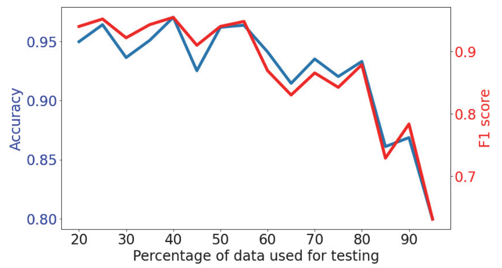
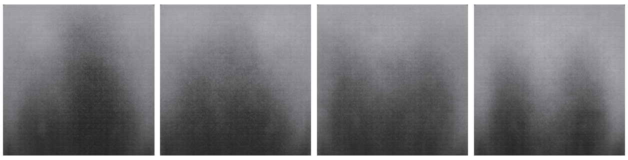


Fig. 4: Mean accuracies and F1-scores for different ratios of test data. Displayed is the mean for three trials.

We show exemplary synthetic images from the DCGAN in Figure 5. The images resembled images from clinical data with additional high-frequency noise.



control coronal metopic sagittal

Fig. 5: Synthetic distance maps of the four classes.

The data augmentation results for the four tests are shown in Table 4. For 85 % test data, F1-score increased when adding synthetic data. Accuracy also increased with the only exception for the *moderately augmented* case. For 90 % test data the addition of synthetic data decreased the F1-score for all scenarios. Overall, accuracy changes were smaller than 5 % and F1-scores 0.1 or below.

Tab. 4: Mean test scenario results after 150 epochs and three trials. Green indicates an increased, red a decreased performance.

		F1	Acc
Reference	85	0.71	0.861
	90	0.79	0.869
Moderately augmented	85	0.81	0.854
(Additional 100% images per class)	90	0.77	0.872
Highly augmented	85	0.77	0.888
(Additional 1000% images per class)	90	0.75	0.812
Balanced	85	0.82	0.873
Images per class: max(largest class)	90	0.77	0.848
Balanced and augmented	85	0.73	0.879
Images per class: max(largest class)·10	90	0.75	0.854

4 Discussion

We presented a DCGAN trained for data augmentation with respect to the classification of craniosynostosis which is capable of synthesizing images with predefined class labels. The images resemble different classes of clinical data with the addition of high-frequency noise which is typical for GANs.

We defined a sparse test scenario and defined four test cases to evaluate the effect of the GAN-based data augmentation with respect to the classifier. The test cases revealed that overall the data augmentation had little effect on the classifier. For the 85 % test case, data augmentation had a slightly positive effect on accuracy and F1-score. Specifically balancing the dataset did not show an improvement compared with class-proportional data augmentation. For 90 %, data augmentation decreased classification performance. We assume that the small amount of data to the train the generative models which might not have been enough to be able to correctly represent the key components important for the classifier.

While the synthesized images visually meet the expectations, the main limitation of this study is the small dataset. More experimental validation is necessary.

Future work might be related to include a statistical shape model based for data augmentation. With a combined data augmentation strategy, it might be possible to also include rare pathologies such as lambdoid synostosis for the classification.

Author Statement

Conflict of interest: Authors state no conflict of interest. Ethics approval: Data acquisition and usage was approved by Ethics Committee Medical Faculty of the University of Heidelberg (Ethics number S-237/2009).

References

[1] Sheree L. Boulet, Sonja A. Rasmussen, and Margaret A. Honein. A population-based study of craniosynostosis in metropolitan atlanta, 1989-2003. *American journal of medical genetics. Part A*, 146A(8):984–991, 2008.

[2] Guido de Jong, Elmar Bijlsma, Jene Meulstee, Myrte Wennen, Erik van Lindert, Thomas Maal, René Aquarius, and Hans Delye. Combining deep learning with 3d stereophotogrammetry for craniosynostosis diagnosis. *Scientific reports*, 10(1):15346, 2020.

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks.

[4] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans.

[5] Stéphanie Lanche, Tron A. Darvann, Hildur Ólafsdóttir, Nuno V. Hermann, Andrea E. van Pelt, Daniel Govier, Marissa J. Tenenbaum, Sybill Naidoo, Per Larsen, Sven Kreiborg, Rasmus Larsen, and Alex A. Kane. A statistical model of head asymmetry in infants with deformational plagiocephaly. In *Image analysis*, volume 4522 of *Lecture notes in computer science*, pages 898–907. Springer, 2007.

[6] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks.

[7] M. Schaufelberger, R. P. Kühle, C. Kaiser, A. Wachter, F. Weichel, N. Hagen, F. Ringwald, U. Eisenmann, C. Freudlsperger, and W. Nahm. Cnn-based classification of craniosynostosis using 2d distance maps. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2022.