# Applying Longitudinal Augmentation and Data Generation (LAUGEN) in Medical Imaging

Nico Albert Disch [1,2,3] * Balint Kovacs[1,5] * Yannick Kirchhoff [1,2,3]
Robin Peretzke [1,5] Maximilian Rokuss [1,3] Saikat Roy [1,3] Constantin Ulrich [1,5]
David Zimmerer [1,2]
Klaus Maier-Hein [1,2,4,6]

[1] Division of Medical Image Computing, German Cancer Research Center, Heidelberg, Germany
[2] HIDSS4Health - Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany
[3] Faculty of Mathematics and Computer Science, University of Heidelberg Heidelberg, Germany
[4] Pattern Analysis and Learning Group, Department of Radiation Oncology Heidelberg University Hospital Heidelberg, Germany
[5] Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany
[6] Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital

nico.disch@dkfz-heidelberg.de, balint.kovacs@dkfz-heidelberg.de

## Abstract

*Deep learning had transformative impact in medical imaging, in areas such as classification, segmentation, and report generation. Another area holding promises, is the area of personalized medicine, especially disease progression modeling. However, longitudinal imaging is even more data-constrained than single time-point imaging, as it requires repeated acquisitions over extended periods, often spanning months or even years. To address this challenge, we introduce Longitudinal Augmentation and Data Generation (LAUGEN) a lightweight, semi-synthetic image generation framework which can be applied in the domain of medical image time series. LAUGEN is efficient, requiring only a single image and its segmentation to produce diverse pseudo-temporal sequences, and is capable of handling typical 3D medical data. We demonstrate its use as a data augmentation strategy for improving model performance and propose its role as a tool for unit testing longitudinal models, where pre-defined latent progressions enable controlled and arbitrarily many evaluations. Our qualitative results on the Brain Tumor Segmentation (BraTS) dataset and quantitative experiments on Automated Cardiac Diagnosis Challenge (ACDC) dataset highlights LAUGEN's potential to enrich datasets and enhance result diversity.*

## 1. Introduction

Medical imaging has become a cornerstone of modern healthcare, enabling non-invasive diagnosis, treatment planning, and disease monitoring. Over the decades, modalities such as X-ray, CT, PET, ultrasound and MRI have changed clinical workflows. The rise of deep learning in medicine has further advanced medical imaging, with models achieving expert-level performance in tasks such as tumor detection, segmentation, and anomaly detection. However, despite the successes, medical imaging remains a data-sparse domain. Unlike expansive natural image datasets such as ImageNet [5], which contains millions of labeled examples, medical imaging datasets are limited by privacy regulations, ethical considerations, and the high costs of expert annotation. This is even more pressing in the domain of image time series, where most modalities are not temporal in nature; In these cases, time series are acquired by imaging patients over a long time period, sometimes over multiple years, such as in Alzheimer's Disease. This makes data collection and curation even more difficult. Consequently,

*These authors contributed equally to this work.

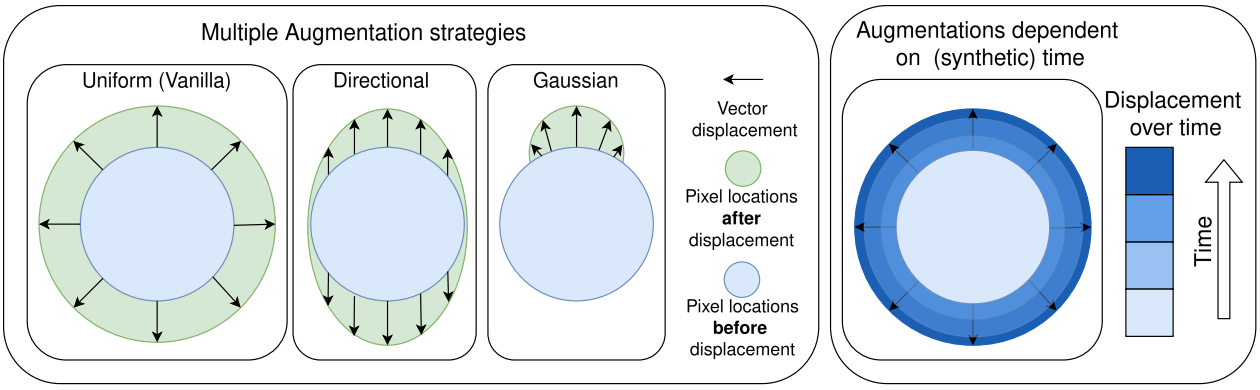



Figure 1. Left: Additional augmentation strategies. From the vanilla uniform, to directional and Gaussian. Right: The second part shows how conceptually the augmentations are turned into time series.

there is a need for data generation methods capable of producing longitudinal series from single images. Most existing approaches require extensive pre-training and are not easily applicable when no temporal data is available. Classical models for disease progression, such as those used in cancer prediction, also exist; however, they often suffer from a modality gap, as the generated outputs are fully synthetic. We therefore introduce Longitudinal Augmentation and Data Generation (LAUGEN), which

1. Is training-free, and both model- and data agnostic;
2. Requires only a single image and segmentation to produce a semi-synthetic longitudinal series, which can leverage ;
3. Serves as augmentation even if there are spatio-temporal data are available.

## 2. Method

### 2.1. Related Work

Traditional augmentations like flipping, rotations, and intensity shifts are common, but do not alleviate the severe data limitations. More advances data augmentations, such as using Diffusion models or GANs [3, 4, 9] aim to diversify the data, typically targeting downstream classification tasks. However, these models often require a lot of computational resources, and can only reproduce the input training data. These methods are generally not suited for image time series data when only single time-point images are available. Synthetic data generation has also been used for pre-training [6, 8, 15], as well as for model validation [1, 10, 11, 14].

### 2.2. Datasets

We focus on two datasets. The first the Brain Tumor Segmentation (BraTS) dataset. This dataset only has single time-point acquisitions. For our purposes we transform this dataset to simulate longitudinal studies. The challenge is to show that there is some merit in using that dataset, so we conceptually show how it can be used. The second dataset is the Automated Cardiac Diagnosis Challenge (ACDC) [2]. This dataset is commonly used as a proxy for longitudinal learning, as it is acquired regularly, in $3D$, and there are no confounding factors in between scans, which are common in cancer imaging datasets. On this dataset, we demonstrate that our approach serves as a data augmentation strategy, slightly enriching the data available. We note here that we train on 90 samples, and validation is performed on 10 samples. In this section we describe the method. We build on the work of [13], which introduces a biological data augmentation. This augmentation was used in [16], in order to improve on locating lesions. We extend the uniform data augmentation with additional steps, namely unidirectional and a single growth in a gaussian distribution. Which will be shown in Figure 1.

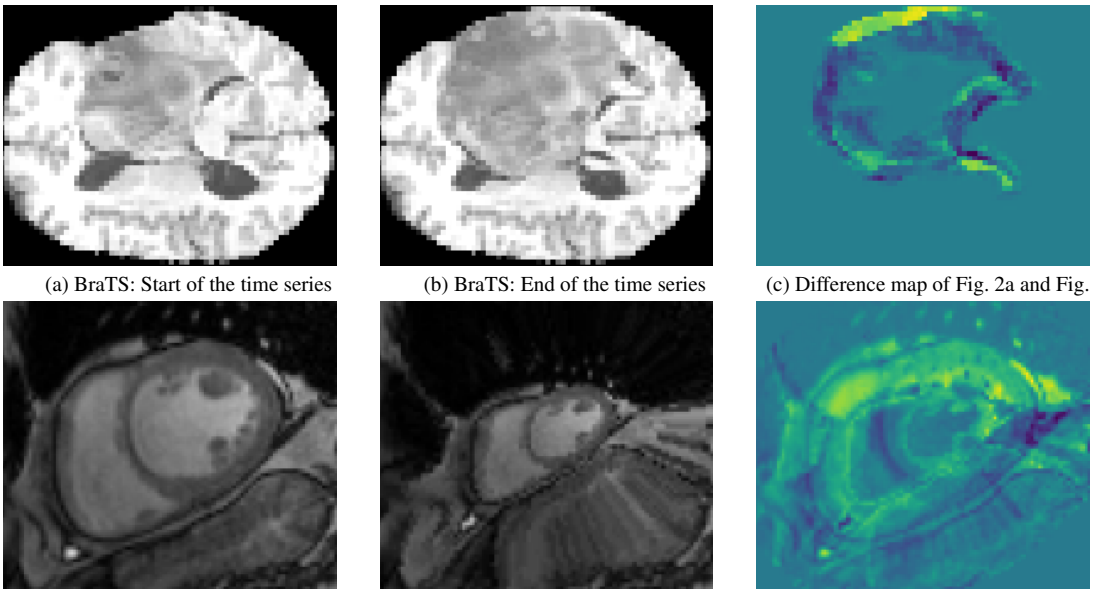
### 2.3. Data Augmentation

**Algorithm 1** Semi-Synthetic Data Augmentation.

**Require:** Image set $\mathcal{I}$ with $|\mathcal{I}| =: N$, Segmentation $S$, Biological Augmentation Function $\mathcal{A}$, Latent Trajectory $\mathcal{T}$, Number of Time Points $T$

Define a synthetic trajectory $\mathcal{T} = \{z_1, z_2, ..., z_T\}$
**for** $I, S \in \mathcal{I}$ **do**
  Extract single image $I$ and segmentation $S$
  **for** $t = 1$ to $T$ **do**
    Generate point from synthetic latent $z_t \in \mathcal{T}$
    Apply augmentation: $(I_t) \leftarrow \mathcal{A}(I, S, z_t)$
  **end for**
  $R_i \leftarrow \{I_t\}_{t=1}^{T}$
**end for**
Return $\{R_i\}_{i=1}^{N}$

In Algorithm 1 we see the general augmentation strategy. In Figure 1 we show the various augmentations, including



(a) BraTS: Start of the time series (b) BraTS: End of the time series (c) Difference map of Fig. 2a and Fig. 2b

(d) ACDC: ED heart phase (real). (e) ACDC: Synthetically augmented end. (f) Difference map of Fig. 2d and Fig. 2e

Figure 2. Figures 2a and 2b: two qualitative temporal samples from BRaTS which were longitudinally augmented with LAUGEN, and c) the difference. d) Shows the ED phase, zoomed in, of the heart, with e) being the synthetic change, and e) the difference.

two additional variants compared to the baseline. We define a latent trajectory $\mathcal{T}$, where we sample from a distribution $\mathcal{D}$. For simplicity and to reduce hyper-parameter tuning, we assume a linear latent trajectory:

$$z_t = m \times t, \tag{1}$$

where $m$ is sampled uniformly in a range, depending on the sample.

### 2.4. Experimental Settings

We train with the SimVP model [7] and their OpenSTL framework. The task is image generation, i.e. predicting the next frame. We choose 4 context images, and 4 target images. For the mixing experiments, **a portion of the training data (defined by the mixing ratio)**, consists of synthetically augmented series, see 1, combined with the true time series, i.e. true medical time series. Additionally, we define a "last image" baseline, which uses the final available context image as the prediction. We find that for ACDC, SimVP outperforms this baseline; however, *in other experiments*, this simple heuristic remains competitive, and **not all models are able to outperform it**.

## 3. Qualitative Results

In Figure 2 we can see an example of the BraTS dataset, that has been longitudinally augmented. In Fig. 2a and Fig. 2b we can see the augmented images, and in Fig. 2c the difference map between the two augmented images. Note that that away from the boundary of the segmentation, the changes are essentially zero. We also notice that despite the simplicity of the change, which is uniform here, the resulting change is still not too simple.

### 3.1. Results Augmentation

To show that method has some merit, we apply it as an actual data augmentation method for the image generation task. For this, we consider the ACDC [2] dataset. We augment the number of real time series, with semi-synthetic augmentations of a single image from the time series. The image is the first image with segmentation, which is the end-diastolic (ED) phase.

In Fig. 3 the effects of the size of the synthetic data addition are shown. In order to give the realistic data a similar chance, we do not let the model train longer, but we repeat the original data. We see that a ratio of $0.1$ gives the best results for this dataset. Given the homogeneity of the changes in this particular dataset, the fact that we are only augmenting in-distribution, and the simplicity of the augmentations, *it is surprising that we observe even a slight improvement*.

In Fig. 4 we can see the difference the augmentation makes on different dataset sizes. We note here that in this dataset, there are 5 different disease characteristics. So depending on the size different morphologies are possible. Depending on the metric, the augmentation helps with the smaller dataset sizes. However, in this experiment the same images were selected. I.e. the augmentations **did not add samples that were *not* seen before**. Considering that this dataset is quite saturated, the improved results are surprising. In future work we could test whether we can further

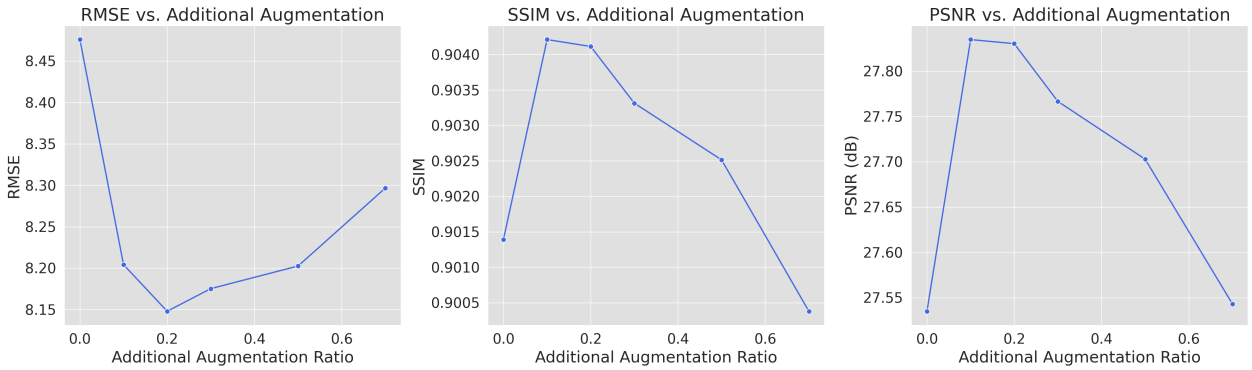



Figure 3. Performance in terms of mixing ratio of synthetic data. The last image baseline is $RMSE$ : 12.144, $SSIM$ : 0.8408, $PSNR$ : 24.00.


Figure 4. Effects of augmentation on time series prediction on the ACDC dataset for the SimVP model. The amount of augmentation is 0.1 chosen from Fig. 3. The x-axis is to the number of datapoints in the dataset. The y-axis is the corresponding metric. As comparison, the last available image has the metrics $RMSE$ : 123.11, $SSIM$ : 0.890, $PSNR$ : 24.80.

add parts that the model has not seen before, thus increasing the diversity of the dataset.

## 3.2. Discussion

We have shown the qualitative and quantitative results of our approach. Despite the small changes in ACDC, even mixing a small amount of synthetic data makes the prediction slightly better. This serves as a first proof of concept for our method.

### 3.2.1. Use Case Scenarios

Our method has several potential applications beyond standard data augmentation. One possible direction is the use of LAUGEN to **unit-test** longitudinal models. The use of semisynthetic models for benchmarking models in [10, 17]. Since we explicitly define the latent trajectories of the generated image time series, we have access to the ground truth progression, unlike in some real longitudinal datasets. This allows us to assess whether trained models behave reliably under known and controlled progression patterns, and to identify potential failure modes in disease modeling.

Another use case is pre-training on semi-synthetic data. Generalization remains a major challenge in medical imaging due to domain shifts across scanners, protocols, and institutions. Training models on diverse, synthetic image time series could act as a form of spatio-temporal pre-training, helping models to better adapt to unseen real-world settings.

### 3.2.2. Limitations

Our approach currently has several limitations. While it is computationally efficient, it currently supports three augmentations, limiting the complexity of the augmentations. Yet, with multiple iterations, more complex patterns would be possible. In addition, it depends on the availability of segmentation masks, which may not be present in all datasets. However, this dependency is mitigated by the increasing availability of high quality, promptable segmentation models such as SAM [12], which can generate masks even with minimal supervision. Since LAUGEN only uses segmentations to extract object boundaries, coarse segmentations may be sufficient. Finally, the current latent trajectory model is too simplistic - limited to a linear progression.

This was a deliberate choice for initial proof-of-concept experiments. Future extensions could include more realistic, disease-specific trajectories, such as sigmoidal curves (e.g. Alzheimer's), or even more complex trajectories for cancer.

## 4. Conclusion

In this work, we propose LAUGEN, a longitudinal augmentation and data generation framework that we have applied to medical imaging. We propose that LAUGEN can be used as a kind of unit-test for models in medical imaging. We also showed that LAUGEN can be used as in-distribution data augmentation on ACDC.

## Acknowledgements

The present contribution is supported by the Helmholtz Association under the joint research school HIDSS4Health – Helmholtz Information and Data Science School for Health.

## References

[1] Brett K. Beaulieu-Jones, Zhiwei Steven Wu, Christopher J. Williams, Ran Lee, S. Bhavnani, James Brian Byrd, and C. Greene. Privacy-preserving generative deep neural networks support clinical data sharing. *bioRxiv*, 2017.

[2] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, Gerard Sanroma, Sandy Napel, Steffen Petersen, Georgios Tziritas, Elias Grinias, Mahendra Khened, Varghese Alex Kollerathu, Ganapathy Krishnamurthi, Marc-Michel Rohé, Xavier Pennec, Maxime Sermesant, Fabian Isensee, Paul Jäger, Klaus H. Maier-Hein, Peter M. Full, Ivo Wolf, Sandy Engelhardt, Christian F. Baumgartner, Lisa M. Koch, Jelmer M. Wolterink, Ivana Išgum, Yeonggul Jang, Yoonmi Hong, Jay Patravali, Shubham Jain, Olivier Humbert, and Pierre-Marc Jodoin. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, 2018.

[3] L. Bonaldi, Andrea Pretto, C. Pirri, F. Uccheddu, C. Fontanella, and C. Stecco. Deep learning-based medical images segmentation of musculoskeletal anatomical structures: A survey of bottlenecks and strategies. *Bioengineering*, 2023.

[4] T. Ceritli, Ghadeer O. Ghosheh, V. Chauhan, T. Zhu, Andrew P. Creagh, and D. Clifton. Synthesizing mixed-type electronic health records using diffusion models. *arXiv.org*, 2023.

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[6] Alvaro Fernandez-Quilez. Deep learning in radiology: ethics of data and on the value of algorithm transparency, interpretability and explainability. *AI and Ethics*, 2022.

[7] Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. SimVP: Simpler Yet Better Video Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3170–3180, 2022.

[8] Gonzalo Garde, Andoni Larumbe-Bergera, Benoît Bossavit, R. Cabeza, Sonia Porta, and A. Villanueva. Gaze estimation problem tackled through synthetic images. *ETRA Short Papers*, 2020.

[9] Fatih Gurcan and Ahmet Soylu. Synthetic boosted resampling using deep generative adversarial networks: A novel approach to improve cancer prediction from imbalanced datasets. *Cancers*, 2024.

[10] Qixing Hu, Alan L. Yuille, and Zongwei Zhou. Synthetic data as validation. *arXiv.org*, 2023.

[11] James Jordon, L. Szpruch, F. Houssiau, M. Bottarelli, Giovanni Cherubin, C. Maple, Samuel N. Cohen, and Adrian Weller. Synthetic data - what, why and how? *arXiv.org*, 2022.

[12] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[13] Balint Kovacs, Nils Netzer, Michael Baumgartner, Carolin Eith, Dimitrios Bounias, Clara Meinzer, Paul F. Jäger, Kevin S. Zhang, Ralf Floca, Adrian Schrader, Fabian Isensee, Regula Gnirs, Magdalena Görtz, Viktoria Schütz, Albrecht Stenzinger, Markus Hohenfellner, Heinz-Peter Schlemmer, Ivo Wolf, David Bonekamp, and Klaus H. Maier-Hein. Anatomy-Informed Data Augmentation for Enhanced Prostate Cancer Detection. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 531–540, Cham, 2023. Springer Nature Switzerland.

[14] Ruochen Pi and Lianlei Shan. Synthetic lung x-ray generation through cross-attention and affinity transformation. 2025.

[15] Moritz Platscher, Jonathan Zopes, and C. Federau. Image translation for medical image generation - ischemic stroke lesions. *arXiv.org*, 2020.

[16] Maximilian Rokuss, Yannick Kirchhoff, Seval Akbal, Balint Kovacs, Saikat Roy, Constantin Ulrich, Tassilo Wald, Lukas T Rotkopf, Heinz-Peter Schlemmer, and Klaus Maier-Hein. LesionLocator: Zero-Shot Universal Tumor Segmentation and Tracking in 3D Whole-Body Imaging. *arXiv.org*, 2025.

[17] Julia Rosenzweig, E. Brito, H. Kobialka, M. Akila, Nico M. Schmidt, Peter Schlicht, Jan David Schneider, Fabian Hüger, M. Rottmann, Sebastian Houben, and Tim Wirtz. Validation of simulation-based testing: Bypassing domain shift with label-to-image synthesis. *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, 2021.