Malte Seemann, Lennart Bargsten* and Alexander Schlaefer

# Data augmentation for computed tomography angiography via synthetic image generation and neural domain adaptation

https://doi.org/10.1515/cdbme-2020-0015

**Abstract:** Deep learning methods produce promising results when applied to a wide range of medical imaging tasks, including segmentation of artery lumen in computed tomography angiography (CTA) data. However, to perform sufficiently, neural networks have to be trained on large amounts of high quality annotated data. In the realm of medical imaging, annotations are not only quite scarce but also often not entirely reliable. To tackle both challenges, we developed a two-step approach for generating realistic synthetic CTA data for the purpose of data augmentation. In the first step moderately realistic images are generated in a purely numerical fashion. In the second step these images are improved by applying neural domain adaptation. We evaluated the impact of synthetic data on lumen segmentation via convolutional neural networks (CNNs) by comparing resulting performances. Improvements of up to 5% in terms of Dice coefficient and 20% for Hausdorff distance represent a proof of concept that the proposed augmentation procedure can be used to enhance deep learning-based segmentation for artery lumen in CTA images.

**Keywords:** computed tomography angiography; data augmentation; deep learning; neural style transfer; segmentation; synthetic image data.

## Introduction

Computed tomography angiography (CTA) is a minimally invasive imaging modality that can help physicians to achieve an accurate diagnosis regarding cardiovascular diseases. Deep learning methods can further improve the quality and quantity of information drawn from such images, but neural networks need to be trained on large amounts of annotated data to perform sufficiently. Especially in medical research, where task-specific annotations for different kinds of image data are required, we see a high demand for domain expert knowledge. But the process of labeling medical image data is costly and time-consuming [1].

One way of dealing with this problem is to implement models that learn efficiently from limited data, e.g., training a convolutional neural network (CNN) for full 3D segmentation by feeding it with only a few manually annotated slices [2]. Nevertheless, performing visual segmentation of vascular structures by hand remains a challenging task. Significant deviations between manual segmentations might occur limiting the development of algorithms [1]. For synthetically created images, the ground truth is known, i.e., reliable labels exist [1]. So additionally to the approach above, simulations can be used for data augmentation, i.e., artificially enlarging the data set.

In this work, we developed a simple algorithm to generate moderately realistic CTA image data of cross-sectional artery representations incorporating pathological changes. We applied neural style transfer (NST) methods in order to improve realism. Finally, we evaluated the impact of this kind of data augmentation on lumen segmentation performance.

## Data

### Data set and preprocessing

We used data available through the Rotterdam Coronary Artery Algorithm Evaluation Framework (RCAAEF) [3]. It comprises 78 artery segments from 18 distinct patients, acquired using three different scanners and annotated by three different observers.

For preprocessing, we followed [4], who converted annotated contour lines into voxel masks and subsequently transformed them into a cross-sectional view by using the

***Corresponding author: Lennart Bargsten**, Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Hamburg, Germany, E-mail: lennart.bargsten@tuhh.de
**Malte Seemann and Alexander Schlaefer,** Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Hamburg, Germany

provided artery centerlines. This representation was found to be preferable for U-Net segmentation [5] of tubular structures. Additionally, the volumes were normalized such that the grayvalues ranged between 0 and 255.

### Generating moderately realistic synthetic data

For simulating 3D cross-sectional CTA artery representations, we needed to achieve high variability but still maintaining a realistic structure, i.e., a circular organic shape in the middle of each slice and variable background due to surrounding tissue. Both were modeled by parameterizing the radius by a sum of sine waves of different period length and random amplitude (e.g., Figure 2). These parameters were slightly changed for sequential slices to maintain inter-slice correlation. The resulting slices were stacked to form 3D data.

To further increase complexity and improve realism we simulated stenoses and plaques with varying degree of calcification. For these pathological changes, three shapes (D-shaped, centric and eccentric) were introduced and added randomly to the synthetic data. Finally, Gaussian noise was added and the resulting volume was filtered with a Gaussian filter.

## Methods

### Synthetic images and stylization

NST is an optimization method that uses a pre-trained CNN to extract statistics describing style and content of images [6]. Those are used to optimize an output image to match the content statistics of a content and the style statistics of a style reference image. For 3D data this optimization problem had to be solved for each slice individually, so we used only a single iteration to reduce the optimization time. For every synthetic image, a real artery segment was chosen randomly from the training set. This resulted in pairs of style and content images for every slice to which the style transfer was applied.

For the sake of completeness, we also implemented real-rime style transfer [7], that significantly speeds up the NST but only allows a single style image (FST).

FastPhotoStyle (FPS) [8] is another domain adaptation approach that is based on whitening and coloring transform (WCT) [9] and does not require further training. It could therefore be efficiently applied slice-wise with different style images using the previously described NST-algorithm.

For evaluation of image synthesis no widely accepted measure exists, so we decided on using multiple metrics to justify the synthetic components and to set the needed parameters for stylization. We tuned the parameters for the shape by measuring shape similarity of synthetic slices and histology images of coronary arteries based on Hu-Moments. Even though constructed for natural RGB images rather than medical images, Fréchet inception distance (FID) [10], which measures the distance between two data distributions via activations of pre-trained image classifiers, indicates at least a correlation regarding image realism. We extended the similarity analysis by multiscale structural similarity (MS-SSIM) [11], which compares contrast, luminance and structure (i.e., correlation) of two corresponding images on multiple scales. Lastly, the Histogram Intersection Similarity Method (HISM) was used to compare the color distributions of the 3D images without taking spatial attributes into account [12].

The same metrics were used to create the set up and evaluate the different style transfer algorithms.

### Segmentation

For segmentation, a state-of-the-art 3D U-Net architecture proposed by [2] was used. A typical U-Net architecture consists of an equal amount of upsampling and downsampling layers [5]. Due to computational resource limitations, the number of feature maps was reduced to 4, 8, 16 and 32 in the four layers. We used the Generalized Dice Loss [13] function as it works well for imbalanced data. During training, random crops and flips were applied. Furthermore, only the labels of the second annotator were used to define a ground-truth and thus neglecting inter-observer variability. The training was performed on Google Colab with a batch size of six and was stopped after 3,000 epochs. We used the model with the best validation performance for further testing.

For evaluation of the segmentation results, we used the Generalized Dice Score [13] and Hausdorff distance. To analyze the effect of data augmentation on training sets of different sizes, we subsequently added artery segments of three patients, one from each scanner (see Figure 1). We performed three-fold cross-validation on the resulting subsets. Fully

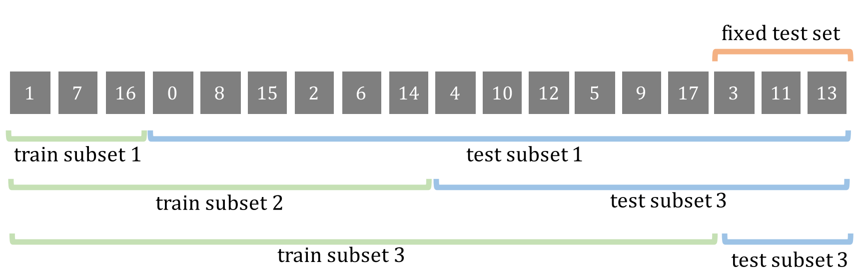


**Figure 1:** Splitting of RCAAEF patient data (gray box) into training (green) and testing (blue) subsets of different sizes.

exploiting the limited total amount of data, we used dynamic sized test sets to compare different approaches trained on the same training subset. In order to compare the results of different training subsets, i.e., sets of different sizes, we introduced a fixed test set.

Two different analyses were carried out: the first one used training subset 1 (see Figure 1) containing data of three patients. It was then enlarged with different amounts of stylized and raw synthetic data. In a second study, 120 synthetic volumes were used to pre-train the segmentation network without stylization (noST) and stylized using FastPhotoStyle (3D-FPS) or neural style transfer (3D-NST). The resulting networks were evaluated with and without fine-tuning. To be able to fairly compare results of different models, we used a majority vote of the two remaining annotators to generate ground-truth for testing.

# Results and discussion

## Synthetic images and stylization

For similarity analysis we successively applied deterioration techniques introduced in section "Generating moderately realistic synthetic data" to 30 synthetically created image volumes (Figure 2) and compared the results to 30 real volumes. Intervals for random value sampling were adjusted such that the similarity increased. Table 1 shows the corresponding results. The improving similarity metrics for each of the intermediate results of our algorithm justify the performed processing steps. Further we see that by applying style transfers to the 'random noise'-images realism measured by all selected metrics improved significantly.

## Segmentation results

Table 2 depicts the results of the pre-training study. The performance of 3D-NST on subset 1 (one network pretrained) shows that networks only trained on the synthetic data already show promising results. Tuning the contracting part of the U-Net with real data (training subset 1 and 2) further improved the segmentation performance in terms of overlap and alignment. It is worth mentioning that on subset 2, this approach outperforms the annotators' segmentation performance measured by GDS, whereas Hausdorff distances are inferior. When increasing the amount of original data further (training subset 3), we found that the positive effect of pre-training and fine-tuning diminishes.

In our second study, we found that also adding synthetic data to the training set improves the segmentation performance. The results can be found in Table 3. Over all three sets of synthetic data, the best performance was reached when having the same amount of original and synthetic data (12). We got the best performance for the network trained on not stylized synthetic data. With a GDS of 72.77 ± 0.48% and a Hausdorff distance of 0.91 ± 0.01mm, we even improved the results of pre-training.

**Table 1:** Result of realism analysis of synthetic data showing mean and standard deviation of MS-SSIM, HISM and FID. Lower FID values are better.

| | MS-SSIM | HISM | FID |
|---|---|---|---|
| Artery only | 0.538 ± 0.110 | 0.028 ± 0.017 | 321.59 |
| With surround | 0.557 ± 0.096 | 0.068 ± 0.042 | 272.7 |
| Filtered randomly | 0.566 ± 0.097 | 0.156 ± 0.070 | 177.3 |
| Random noise | 0.583 ± 0.89 | 0.386 ± 0.164 | 115.1 |
| 3D-NST | *0.604 ± 0.085* | 0.541 ± 0.165 | *60.3* |
| FST | 0.578 ± 0.087 | *0.612 ± 0.144* | 72.0 |
| 3D-FPS | 0.588 ± 0.095 | 0.554 ± 0.140 | 90.4 |

**Table 2:** Test results of training and fine-tuning on training subset 1 and 2 in terms of the mean and standard deviation of cross-validated GDS and Hausdorff distances.

| | | GDS, % | Hausdorff dist., mm |
|---|---|---|---|
| Train subset 1 | Annotator 1 | 74.79 | 0.39 |
| | Original | 68.91 ± 1.23 | 1.14 ± 0.13 |
| | no ST | 68.93 | 1.11 |
| | no ST finetuned | 69.41 ± 0.67 | 1.07 ± 0.04 |
| | 3D-NST | 69.89 | 0.97 |
| | 3D-NST finetuned | *70.97 ± 0.49* | *0.96 ± 0.05* |
| | 3D-FPS | 69.61 | 0.99 |
| | 3D-FPS finetuned | 69.07 ± 0.77 | 0.98 ± 0.03 |
| Train subset 2 | Annotator 1 | 72.37 | 0.38 |
| | Original | 72.08 ± 0.79 | 0.97 ± 0.12 |
| | no ST finetuned | 73.28 ± 0.89 | 0.88 ± 0.02 |
| | 3D-NST finetuned | *73.34 ± 2.62* | 0.89 ± 0.07 |
| | 3D-FPS finetuned | 72.93 ± 0.01 | *0.85 ± 0.07* |

**Table 3:** Test results of training on training subset 1 enlarged with 6, 12 and 18 synthetic images in terms of the mean and standard deviation of cross-validated GDS and Hausdorff distances.

| | | GDS, % | Hausdorff dist., mm |
|---|---|---|---|
| 6 | no ST | 70.62 ± 0.84 | 1.03 ± 0.07 |
| | 3D-NST | 68.65 ± 0.08 | 1.13 ± 0.05 |
| | 3D-FPS | 70.18 ± 1.02 | 1.04 ± 0.06 |
| 12 | no ST | *72.77 ± 0.48* | *0.91 ± 0.01* |
| | 3D-NST | 70.86 ± 1.36 | 0.95 ± 0.04 |
| | 3D-FPS | 71.72 ± 0.6 | 0.93 ± 0.03 |
| 18 | no ST | 70.55 ± 1.07 | 0.93 ± 0.06 |
| | 3D-NST | 69.75 ± 1.28 | 1.08 ± 0.04 |
| | 3D-FPS | 1.49 ± 0.97 | 0.94 ± 0.41 |

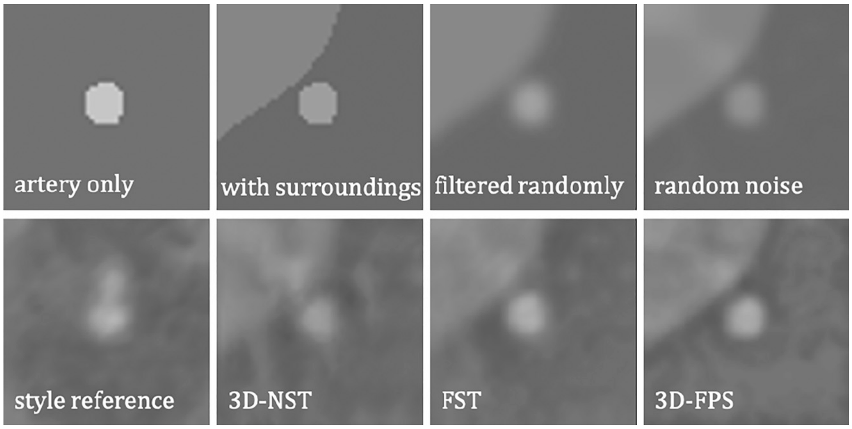


**Figure 2:** Synthetic CTA slices. The first row shows four different processing steps of the developed algorithm. Second row shows an example original CTA slice from RCAAEF data set and three results of stylizing 'random noise'-image

## Conclusion

From the results presented above, we can conclude that using synthetic data has a significant impact on lumen segmentation performance of CTA data in terms of Dice coefficient and Hausdorff distance, especially in cases where annotated data is scarce.

We showed that the domain adaptation methods examined significantly increased visual realism measured by multiple metrics. Using this stylized synthetic data further improved the performance when used for pre-training but not when mixing with real data.

It is noteworthy that training the segmentation network only on synthetic data performed similarly to the ones trained only on real data. Since this could decrease the problem of subjectivity in the process of producing manual segmentations without a loss of performance, this might be an interesting direction for future research.

**Research funding:** This work was partially funded by the European Regional Development Fund (ERDF), by the Hamburgische Investitions – und Förderbank (IFB) and by the Free and Hanseatic City of Hamburg.
**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
**Competing interests:** Authors state no conflict of interest.
**Informed consent:** Informed consent was obtained from all individuals included in this study.

## References

1. Litjens GJS, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
2. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. "3d u-net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016; 2016:424–32 p.
3. Kirişli H, Schaap M, Metz C, Dharampal AS, Meijboom WB, Papadopoulou SL, et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. Med Image Anal Dec 2013;17:859–76 [Online]. Available from: https://doi.org/10.1016/j.media.2013.05.007.
4. Bargsten L, Wendebourg M, Schlaefer A. Data representations for segmentation of vascular structures using convolutional neural networks with u-net architecture. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Berlin, Germany: IEEE; 2019. https://doi.org/10.1109/embc.2019.8857630.
5. Ronneberger PFO, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015:234–41 pp.
6. Gatys LA, Ecker AS, Bethge M. A neural algorithm of artistic style. ArXiv 2016;abs/1508:06576. https://doi.org/10.1167/16.12.326.
7. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision; 2016.
8. Li Y, Liu M, Li X, Yang M, Kautz J. A closed-form solution to photorealistic image stylization. ArXiv 2018;abs/1802:06474.
9. Li Y, Fang C, Yang J, Wang Z, Lu X, Yang M. Universal style transfer via feature transforms. ArXiv 2017;abs/1705:08086.
10. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017:6629–40 p.
11. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: The thrity-seventh asilomar conference on signals, systems computers 2003; 2003, vol 2: 1398–402 pp.
12. Swain MJ, Ballard DH. Color indexing. Int J Comput Vis 1991;7:11–32.
13. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. ArXiv 2017;abs/1707:03237. https://doi.org/10.1007/978-3-319-67558-9_28.