CNN BASED VEHICLE TRACK DETECTION IN COHERENT SAR IMAGERY: AN ANALYSIS OF DATA AUGMENTATION

S. Kuny[1]*, H. Hammer[1], A. Thiele[1,2]

[1] Fraunhofer IOSB, Institute of Optronics, System Technologies, and Image Exploitation, Ettlingen, Germany - (silvia.kuny, horst.hammer, antje.thiele)@iosb.fraunhofer.de
[2] Institute of Photogrammetry and Remote Sensing IPF, Karlsruhe Institute of Technology KIT, Germany

Commission I, WG I/3

KEY WORDS: coherent change detection, vehicle tracks, CNN, data augmentation

ABSTRACT:

The coherence image as a product of a coherent SAR image pair can expose even subtle changes in the surface of a scene, such as vehicle tracks. For machine learning models, the large amount of required training data often is a crucial issue. A general solution for this is data augmentation. Standard techniques, however, were predominantly developed for optical imagery, thus do not account for SAR specific characteristics and thus are only partially applicable to SAR imagery. In this paper several data augmentation techniques are investigated for their performance impact regarding a CNN based vehicle track detection with the aim of generating an optimized data set. Quantitative results are shown on the performance comparison. Furthermore, the performance of the fully-augmented data set is put into relation to the training with a large non-augmented data set.

1. INTRODUCTION

Synthetic aperture radar (SAR) imagery allows for two different approaches to change detection: amplitude change detection or coherent change detection if provided with a coherent image pair. The coherence image as a product of a coherent image pair can expose even subtle changes in the surface of a scene, such as the tracks made by vehicles. Several approaches to vehicle track detection exist in the literature, including e.g. the use of convolutional networks (Quach, 2017) or conditional random fields (Malinas et al., 2015). Others seek to enhance the coherence image with the aim of boosting a threshold-based change detection method (Hammer et al., 2021).

As is common for all image classification via machine learning models, the large amount of required training data often is a crucial issue. A general solution for data scarcity is data augmentation, where different techniques are used to expand the existing data set in size and quality (Shorten and Khoshgoftaar, 2019). Current techniques for coherent track detection seek to make synthetic data look more like measured data using machine learning algorithms (Lewis et al., 2019) or insert simulated tire tracks into non simulated images to obtain a larger variety of images (Turner et al., 2012). Most fundamental are techniques using geometric and color space modifications, however, these standard techniques were predominantly developed for optical imagery, thus do not account for SAR specific characteristics and thus are only partially applicable to SAR imagery. Several of these techniques can be ruled out merely by considering the specific properties of SAR images, however, for some the question arises, how well they are suited for the task of coherent change detection and what impact they have on the actual track detection performance.

In this paper several data augmentation techniques (geometric and color space transformations) are investigated for their performance impact regarding a convolutional neural network (CNN) based vehicle track detection. With the aim of generating an optimized training data set, they are compared among one another and subsequently put into relation to the training with a larger non-augmented data set. It is of interest if the augmentation of samples originating from a single image can compare with the un-augmented large data set extracted from multiple diverse images.

The paper is structured as follows: Section 2 contains a description of the data set used in this study. In Section 3 the process of data augmentation is specified. Section 4 describes the CNN architecture and training process. The results are presented in Section 5, while Section 6 contains the conclusions and an outlook to future work.

2. DATA

The experiment is conducted on an airborne interferometric SAR data set of POLYGONE area, located in southern Rhineland-Palatinate, Germany, where between the two overflights three distinguishable vehicle tracks were generated per vehicle movement. The tracks overlap to an extent and feature an axle width of 2.0 m $\pm$0.2 m and a wheel width between 0.37 m and 0.4 m. Otherwise this area was not affected by human action in-between the times of the overflights. Figure 1 shows an optical image of the scene, where the area of vehicle movement is marked in orange. The recorded data set consists of multiple coherent image pairs, showing the same scene under different aspect angles. A manually performed vector-based extraction of the three vehicle tracks yields the corresponding reference data in the form of a binary image distinguishing track from background.

* Corresponding author

Figure 1. Optical image of the POLYGONE scene.

2.1 SAR imagery

The SAR data set was part of a measurement campaign in 2015, recorded by the SmartRadar experimental sensor of Hensoldt Sensors GmbH, mounted on a Learjet. This is an X-band sensor with resolution in the decimeter range. The six image pairs used in this study were recorded during two overflights over Bann B of POLYGONE Range, approximately 4 hours apart. In-between this time the vehicle movement took place, whereas at the time of the overflights the scene was static. In Table 1 basic properties of the POLYGONE acquisition are summed up. For this investigation six image pairs have been selected featuring only very small acquisition angle differences, so that high coherence values can be achieved. All image pairs were co-registered and subsequently the coherence was computed using the classical formula and a 7 x 7 pixel window. Coherence takes values in the interval [0,1] where zero reflects total incoherence (black) and 1 implies a fully coherent signal (white). In Figure 2, the resulting coherence images C_1-C_6 are depicted, showing the whole scene of the POLYGONE area. Note that all SAR images in this paper are visualized with range direction on the x-axis and azimuth on the y-axis. In all six images wide horizontal stripes are visible which are caused by a flawed motion compensation during SAR data processing. For the task at hand, however, this is not considered to be a problem. As a matter of principle a high coherence level surrounding the changed regions is essential for a successful coherent change detection. The grassland cropped shortly before the measurement campaign features such an area of high coherence, thus making the vehicle track detection in this area possible. For the images C_1-C_6 the coherence levels vary somewhat, which was to be expected, since the acquisition angles and the angle difference between each image pair differ. For this, see the azimuth angle α_M of each master image, the azimuth difference $\Delta\alpha$ regarding each image pair, and the mean local coherence level $\bar{\gamma}_C$ of said grass-land area (measured in a 500 x 500 pixel window) in Table 2.

Sensor	SmartRadar
Band	X-band
Resolution	Decimeter range
Slant range distance	approx. 10,260 m
Flight height above ground	approx. 4270 m
Depression angle first bin	approx. 25°
Time between acquisitions	approx. 4 h

Table 1. Acquisition properties of the POLYGONE measurement campaign.

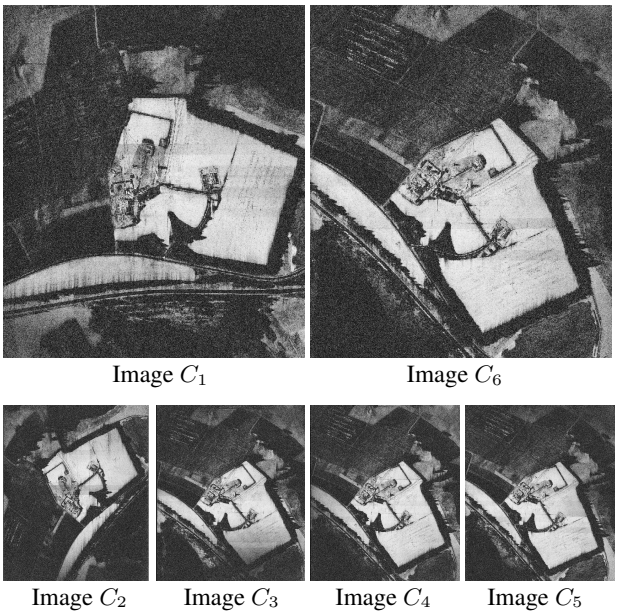


Figure 2. Coherence images: Image C_1 for the basis of an augmented data set, Images C_2-C_5 for the generation of a large un-augmented data set, and Image C_6 for the testing.

Since the images are not calibrated and recorded under different aspect angles, they show varying track orientation and backscattering and they also result in a different coherence level, thus qualifying for independent training and test imagery. In the main part of this investigation, Image C_1 was used as source material for the training data set (un-augmented and augmented), whereas Image C_6 functioned as test environment. Note the profound orientation difference of the two images. Images C_2 - C_5 subsequently were used in combination with Image C_1 to provide for a large, diverse un-augmented data set.

2.2 Reference data

Reference imagery was generated in three steps: Firstly, a manual vector-based extraction of the three vehicle tracks was conducted, where both lanes were accounted for; secondly, the tracks were rasterized to generate a mask; and lastly, the mask was subjected to a morphological dilation operation to enforce a standard wheel width. The result is a binary image distinguishing track from background. Figure 3 shows the process of reference generation for the vehicle tracks in Image C_1. Images C_2 - C_6 were processed accordingly.

3. DATA AUGMENTATION

Using but one image for the extraction of a training data set, as in this case, inevitably leads to some deficits regarding object

	α_M	$\Delta\alpha$	$\bar{\gamma}_C$
C_1	161.56°	0.000045°	0.874
C_2	115.49°	0.005707°	0.875
C_3	203.57°	0.005745°	0.861
C_4	206.58°	0.005237°	0.865
C_5	209.81°	0.011034°	0.836
C_6	208.70°	0.001642°	0.856

Table 2. List of coherence images: Master azimuth α_M; azimuth difference $\Delta\alpha$; local mean coherence level $\bar{\gamma}_C$.

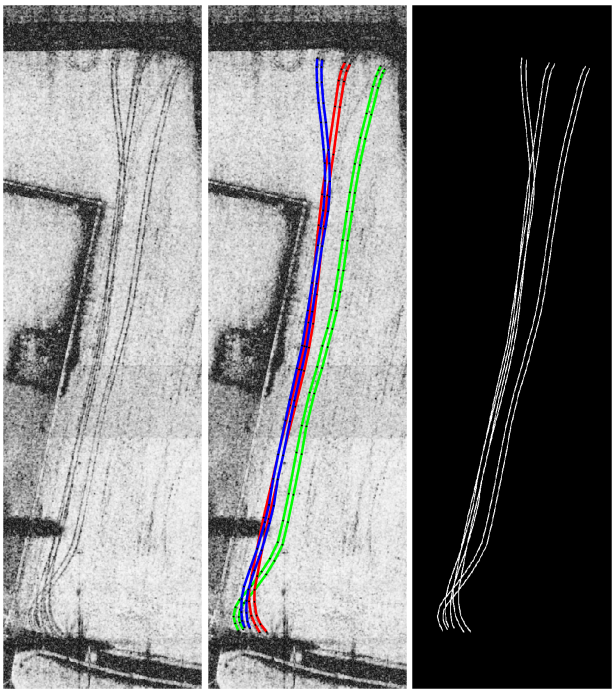


Figure 3. Manual track extraction for reference purposes (here Image C_1): a) coherence image; b) vector-based manual extraction; c) binary mask with dilated tracks.

orientation, and backscattering variety. Data augmentation methods can be used to widen that variety, so that the trained model generalizes better. This raises the question whether standard data augmentation techniques on a small data set can equal the use of a a larger data set with a wider variety. In that regard the interferometric data set of POLYGONE is very suitable for an investigation.

3.1 Standard methods

When it comes to data augmentation a wide range of methods have been introduced, ranging from standard basic image manipulation to more sophisticated methods. GAN-based augmentation or other deep learning based augmentation methods are amongst the most advanced methods and are capable of very complex image generation. For the task at hand, however, these methods are somewhat disproportionate and hence are not taken into account.

The focus instead is on the standard methods of image manipulation such as geometric and color space transformations. However, since these standard augmentation techniques were developed primarily for optical use, not all are reasonable for the SAR specific case. One such case is the widely used method of scaling. While for the optical use the simple resizing of an image to imitate another resolution or object size is valid, in the SAR specific case this disregards the more complex workings of image processing. Not only the radiometry would be tampered with, but for more complex objects with specular reflections also the characteristic features of the signatures. Further, the resolution of the SAR image depends only on the sensor and the acquisition mode and is constant over the entire image. There are further techniques that are obsolete due to the missing radiometric background, such as e.g. shearing, noise injection. Since in SAR images small changes in aspect angle can result in profound signature changes, in particular due to multi-bounce reflections, rotation augmentation usually is an unsuitable method as well. However, for flat objects at ground level, such as in this case it is a different matter. The absence of orientation dependent specular reflections and the object flatness eliminate the main objections. Whether a difference in range and azimuth pixel spacing may cause unrealistic distortions when rotated is deemed insignificant when compared with the high potential of rotation augmentation. Many color space transformations rely on the presence of multiple channels, and for this reason cannot be transferred to SAR imagery. However, simple modifications can also be applied for grayscale images.

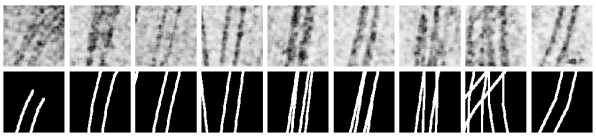


Figure 4. Exemplary un-augmented training samples.

In the following, five augmentation techniques are applied to the training samples from Image C_1, including translation (A), flipping (B) and rotation (C), as well as changes to contrast (D) and brightness (E).

3.2 Applying augmentation methods

Original samples Based on a single coherence image C_1 a base training data set of 2000 samples of the size 128 x 128 pixels was extracted, with the samples being perfectly centered on the tracks. The reference map provides the corresponding label data, accordingly. Figure 4 shows an exemplary set of these training data. Based on these samples, multiple training data sets were generated via the data augmentation techniques A-E.

Translation To avoid positional bias in the data, translation augmentation was used, where the sample centers are moved with a random displacement offset. Taking into account the sample size of 128 x 128 pixels and the track width a maximal translation offset of 45 pixels was chosen so that the maximal cut-off was no more than 35% of the sample. Figure 5 a) shows an exemplary set of these training data.

Flipping Random horizontal and vertical flipping of the original samples was conducted. The resulting training data are depicted in Figure 5 b).

Rotation Based on the original sample centers and Image C_1 a rotation was executed for each sample using nearest-neighbor interpolation. For an optimal coverage of track orientations the rotation was executed randomly in the interval between 0° and 359°. Note that there is little variation regarding the orientation of the vehicle tracks in the original samples, suggesting the rotation augmentation to be able to improve the model training considerably. Figure 5 c) shows the exemplary rotated samples.

Contrast and brightness The challenge of a track detection ultimately is to work with image pairs of poor coherence, so the contrast is bound to be smaller than that of the given training data. To take this into consideration both contrast and brightness were modified to a small extent. For both, there was made a point of making sure the resulting values were in the range a typical coherence image takes, between 0 and 1. Figures 5 d) and e) depict the corresponding augmentation samples with a distinct change in gray values.

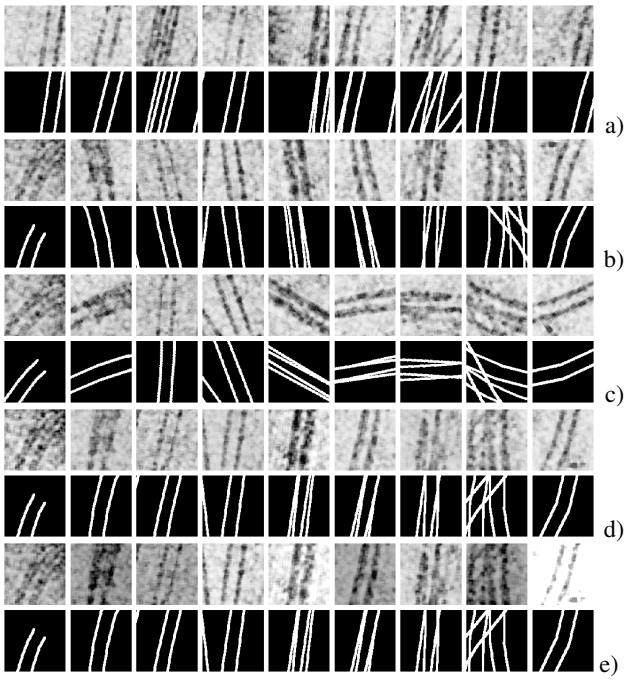


Figure 5. Exemplary training samples after data augmentation: a) translation, b) flipping, c) rotation, d) contrast, e) brightness.

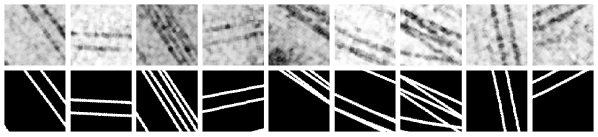


Figure 6. Exemplary training samples with all 5 data augmentation methods.

3.3 Training data sets

In this paper two aspects are targeted: Firstly, the investigation of the five data augmentation techniques and their performance impact in a vehicle track detection; and secondly, the aim of generating an optimized training data set that can match the performance of a larger non-augmented data set. For the first aspect, six training data sets were generated (one for the original un-augmented data and one for each augmentation technique, respectively, each consisting of 2000 samples. They are denoted as listed in Table 3. Regarding the optimized training data set, a further set was generated combining all the augmentation techniques A-E. Some exemplary training samples are depicted in Figure 6. Lastly, a large un-augmented data set was created by extracting samples not only from Image C_1, but also from Images C_2 - C_5, resulting in a data set of 10.000 images. In total, this leads to the generation of 8 training data sets listed in Table 3.

4. CNN TRAINING

The U-Net architecture has been shown to be very effective for fast semantic segmentation of images, and hence was considered a suitable choice for the task at hand. In our experiment a standard 4-layer U-Net architecture was used, consisting of a contracting path, a bridge segment and an expansive path. The encoder subnetwork consists of two sets of convolutional and ReLU layers at a time followed by a max pooling layer. In return, the decoder subnetwork involves a transposed convolutional layer and two sets of convolutional and ReLU layers at a time. The 4-layer structure represents a good compromise between the position independence of the features and the fact that too much information is lost when the images in the lowest U-Net layer become too small. Note that with an input image size of 128 x 128 pixels, the lowest layer image has but a size of 16 x 16 pixels. The network was then trained with an Adam optimizer and fed with the different training data sets respectively, whereas 200 samples of each training data set were used as validation data. Table 4 shows the final validation accuracies and losses for different training data sets. The trained models were then applied to the test image C_1.

data set	augmentation	source image
DS_{orig}	un-augmented	C_1
DS_A	translation	C_1
DS_B	flipping	C_1
DS_C	rotation	C_1
DS_D	contrast	C_1
DS_E	brightness	C_1
DS_{A-E}	methods A-E	C_1
$DS_{BigData}$	un-augmented	C_1 - C_5

Table 3. List of training data sets.

data set	validation accuracy	validation loss
DS_{orig}	99.1571	0.020999
DS_A	98.9971	0.025031
DS_B	97.2638	0.062695
DS_C	96.4036	0.088731
DS_D	98.6041	0.033565
DS_E	98.8415	0.027946
DS_{A-E}	95.9861	0.094365
$DS_{BigData}$	98.9135	0.025894

Table 4. Training information for the different data sets.

5. RESULTS

In the following the predictions for test image C_6 regarding the individual trained networks are described. Aside from a visual inspection, two quality measures are used to assess the track detection performance. Firstly, a detection performance ratio (DPR) is introduced, which describes the ratio of detected pixels in the local track area and calls upon the reference mask to be able to do so. To capture the line continuity of the track detection, the segmentation result is converted into connected components with an 8-pixel connectivity. As a second criterion the maximal length L_{max} of the major ellipse axis of the components is explored, where the full length of the vehicle tracks would equal an L_{max} of 3745.6 pixels.

5.1 Effect of data augmentation

In Figure 7, two details of the prediction results for Image C_6 are visualized, regarding a training with the un-augmented data set DS_{orig} and the augmented data sets DS_A-DS_E. For a better visual impression, the segments (red) are superimposed on the corresponding coherence image. Quantitative results are provided in Table 5. The performance regarding the un-augmented data set DS_{orig} serves as a basis for the assessment of the augmentation impact. So it is of relevance how well this simple training data set can perform. Figure 7 a) demonstrates quite well that the un-augmented samples lack the variety to generalize the network. In particular, the lack of track orientation in the training samples becomes apparent,

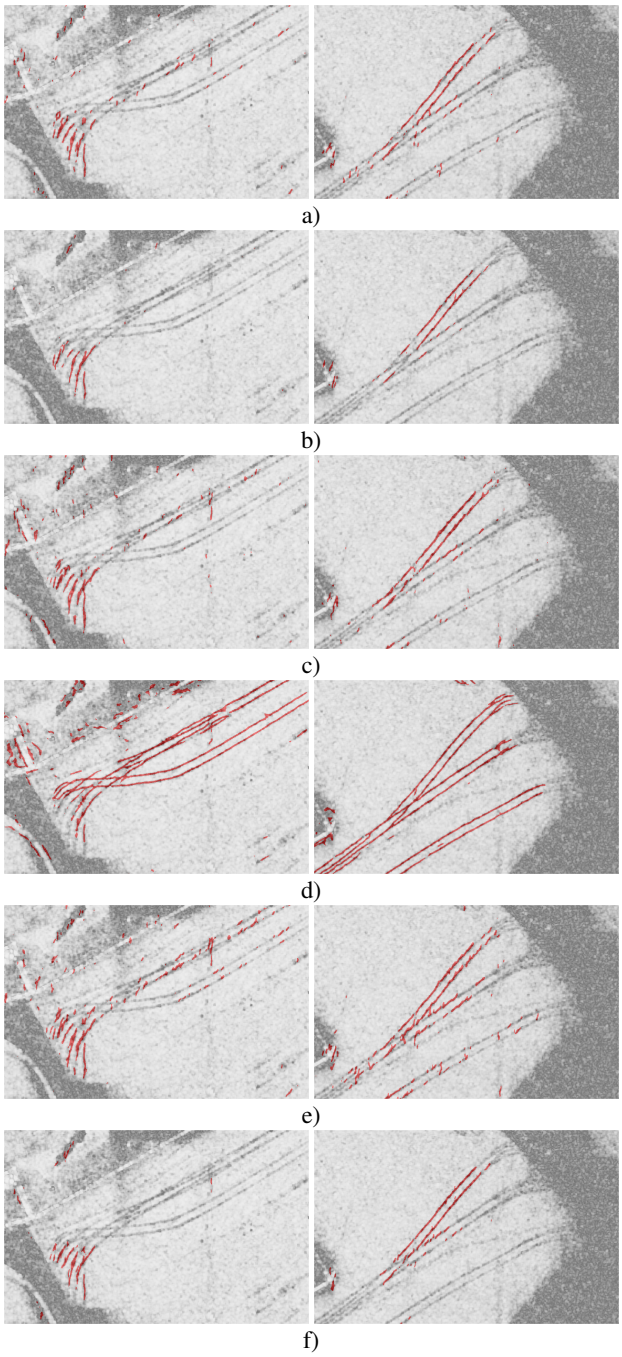


Figure 7. Segmentation results (red) superimposed on the coherence image C_6, regarding the network training with data sets: a) DS_{orig}, b) DS_A, c) DS_B, d) DS_C, e) DS_D, f) DS_E.

	DPR [%]	L_{max}
DS_{orig}	11.8009	313.8592
DS_A	5.6812	329.7666
DS_B	13.2699	374.6575
DS_C	67.5337	2192.9018
DS_D	20.3743	376.3904
DS_E	7.5411	413.1504
DS_{A-E}	72.5374	2278.1166

Table 5. Performance measures for the track detection (detection performance ratio DPR, and the maximal length L_{max} of the connected components) regarding the individual data sets.

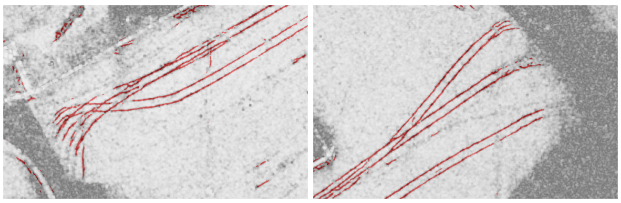


Figure 8. Segmentation results (red) superimposed on the coherence image C_6, regarding the training with set DS_{A-E}.

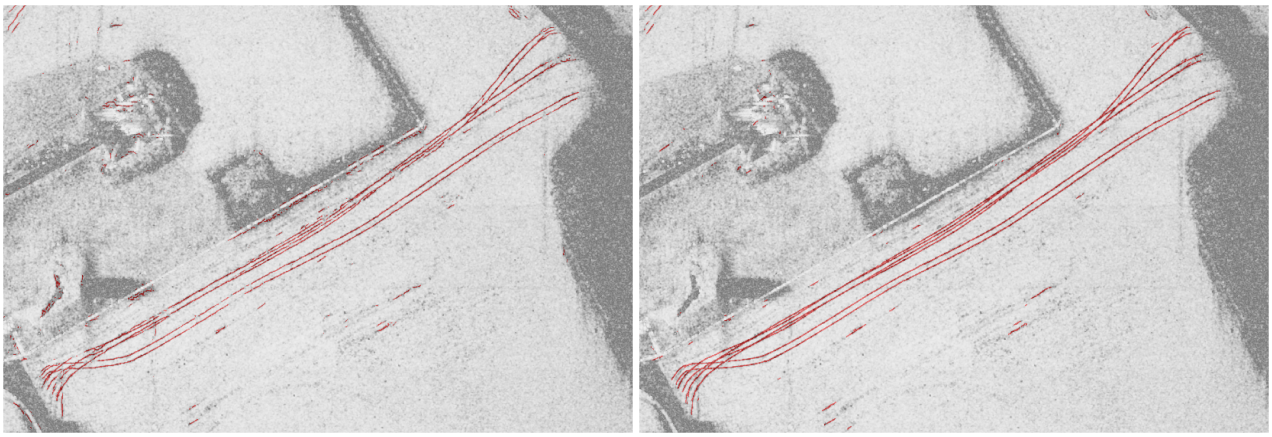
	DPR [%]	L_{max}
DS_{A-E}	72.5374	2278.1166
$DS_{BigData}$	79.4442	3663.1824

Table 6. Performance measures for the track detection (detection performance ratio DPR, and the maximal length L_{max} of the connected components) regarding the individual data sets.

since the performance varies profoundly with the track orientation. Several track segments aligned more in azimuth direction even show an acceptable result. With a DPR of 11.8% and an L_{max} of 313.9 pixels this is used as a basis of comparison. Figures 7 b)-f) show the effect of the individual data augmentation techniques. As was expected, the rotation augmentation has the most profound impact on the track detection performance (see Figure 7 d)), with the observed orientation dependent performance differences seemingly eliminated completely. Overall, a good performance can already be achieved with but this augmentation technique, also showing in the high values of the chosen performance measures, a DPR of 67.5% and an L_{max} of 2192.9 pixels. In comparison, all other techniques have a much smaller effect on the performance. The flipping augmentation, even though by far not as powerful as the rotation technique, has some effect in the same direction. Most track orientations still cannot be detected, however, the performance measures (DPR of 13.3% and an L_{max} of 374.7 pixels) show a certain improvement to the un-augmented data set. Employing contrast augmentation leads again to a small performance increase (DPR of 20.4% and an L_{max} of 376.4 pixels). This improvement is probably due to the somewhat lower coherence level in test image C_6 compared to the training image C_1. Augmentation by translation and brightness modifications show no clear improvement over the un-augmented data set. However, for the optimized data set DS_{A-E} they seem to improve the robustness of the model, so that the optimized data set deliberately includes all five augmentation techniques. The segmentation result of the network trained with data set DS_{A-E} can be observed in Figure 8. The data augmentation with a combination of all five augmentation techniques could again considerably improve the track detection results and achieve a DPR of 72.5% and an L_{max} of 2278.1 pixels.

5.2 Data augmentation vs larger un-augmented data set

To put the performance of the fully augmented data set into relation, a performance comparison to the network trained on the larger un-augmented data set $DS_{BigData}$ is provided in the following. A visual impression is given in Figure 9, showing the segmentation result for both the training with the fully augmented data set and the training with the large un-augmented data set. Although both show a good line continuity, the results of the large un-augmented data set surpass those of the fully augmented data set. This also is reflected in the performance measures, listed in Table 6. The un-augmented data set pro-



a) b)

Figure 9. Segmentation of test area (Image C_6); a) augmented data set DS_{A-E}, b) large un-augmented data set $DS_{BigData}$.

duces a DPR of 79.4% and an L_{max} of 3663.2 pixels. Even though standard methods of image manipulation such as geometric transformations and color space transformations were not able to fully replace the use of a more diverse large data set, the performance increase is profound.

6. CONCLUSION AND OUTLOOK

The experiment was conducted on an airborne dual-pass SAR data set of POLYGONE area, located in southern Rhineland-Palatinate, Germany, where between the two overflights three distinguishable vehicle tracks were generated per vehicle movement. The data set consists of multiple coherent image pairs, showing the same scene under different aspect angles. A manually performed vector-based extraction of the three vehicle tracks provided the corresponding reference data in the form of a binary image distinguishing track from background.

It was discussed which standard augmentation techniques are not reasonable for the SAR specific case and which are to be investigated. Based on a single coherence image a base training data set of 2000 samples was extracted and subsequently multiple training data sets were generated via different data augmentation techniques. These include geometric transformations such as translation, flipping and rotation, as well as color space transformations such as changes to contrast and brightness. A second coherence image of the overflight functioned as test data, showing the same three vehicle tracks in a different orientation. A CNN with a 4-layer U-Net architecture was introduced and trained with an Adam optimizer for the different training data sets respectively. The performance of the trained models was then assessed on the test image, whereas e.g. line continuity served as a quality criterion. As a result the impact each augmentation technique has on the track detection performance can be rated. As a last step the training with augmented data was put into relation to the training with non-augmented data. For this, four additional coherence images of the scene were exploited to receive a large data set matching the augmented data sets in size and featuring the three vehicle tracks for multiple orientations and coherence levels. Finally, a performance comparison was conducted between the best results regarding the augmented data versus the results of training with non-augmented data. Concluding, this brings into light how well the data augmentation techniques can imitate an actual larger data set.

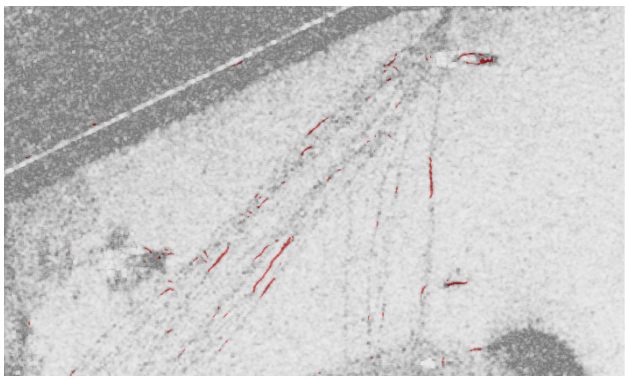


Figure 10. Foot tracks in the POLYGONE data set.

Future work could include, how well this approach can be expanded to the prospect of foot track detection. The POLYGONE data set provides the means for such an investigation, with Figure 10 showing the area in question and the segmentation result regarding the network trained on vehicle tracks.

REFERENCES

Hammer, H., Kuny, S., Thiele, A., 2021. Enhancing coherence images for coherent change detection: An example on vehicle tracks in airborne SAR images. *Remote Sensing*, 13, 5010.

Lewis, B., DeGuchy, O., Sebastian, J., Kaminski, J., 2019. Realistic SAR data augmentation using machine learning techniques. *Algorithms for Synthetic Aperture Radar Imagery XXVI*, 10987, 12-28.

Malinas, R., Quach, T.-T., Koch, M., 2015. Vehicle track detection in CCD imagery via conditional random field. *49th Asilomar Conference on Signals, Systems and Computers*, 1571-1575.

Quach, T.-T., 2017. Convolutional networks for vehicle track segmentation. *Journal of Applied Remote Sensing*, 11(4), 1-10.

Shorten, C., Khoshgoftaar, T., 2019. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60.

Turner, E., Phillips, R., Chiang, C., Cha, M., 2012. Inserting simulated tracks into SAR CCD imagery. *Autumn Simulation Multi-Conference*, 44, 17-24.