CV for anomaly detection in industrial applications

Viacheslav Kozitsin, Iurii Katser, Arman Alahyari, Anton Hinneck, Rahim Tariverdi

Skolkovo Institute of Science and Technology

Abstract

Keywords: deep learning, computer vision, anomaly detection, fault detection, technical systems

1. INTRODUCTION

Anomaly detection problems have a great importance in industrial applications, because anomalies usually represent faults, failures or the emergence of such. To detect these automatically, we propose deep learning algorithms for anomaly / fault detection and their classification. We develop an algorithm achieving Segmentation Defects classification Its performance will be examined on two different problems: anomaly detection in oil pipelines and fault detection on transmission system components. Both these systems can span over thousands of kilometers, which makes manual inspection very costly. How such computer vision techniques can be applied on an industrial scale is discussed in the following.

Transmission infrastructure, on the other hand, physically connects power sources and consumers extending over thousands of kilometers. Many different components exist, which is why maintaining power grids is a serious cost factor for transmission system operators (TSOs). Automated fault detection could potentially help to decrease costs. To use an automated visual system however, specific infrastructure first needs to be identified and segmented, to then perform any type of fault detection. Therefore, the proposed deep learning-based approach, to segment and identify faulty components (i.e. insulators) in handheld footage, will be tested regarding its viability. If a deep-learning based approach would prove reliable, drones could monitor equipment automatically.

The damage of pipelines that transport petroleum and gas products lead to severe environmental problems. Eliminating breakthroughs and their consequences is expensive. To avoid accidents, it is recommended to improve diagnostics quality and to increase the frequency of in-line-inspection (ILI) tools deployment. ILI tools, also referred to as pipeline inspection gauges (Fig. 1), use Hall effect for measuring localized magnetic flux leakage intensity along the pipe wall. While moving along the pipe gauge inspects the wall and detects the magnetic field leaks. The data collected during the inspection can be further analyzed for main diagnostics problems solving: damage and defects detection, their localization, diagnosis or defects classification. Analysis results are useful for assets managing and repair priorities determination.

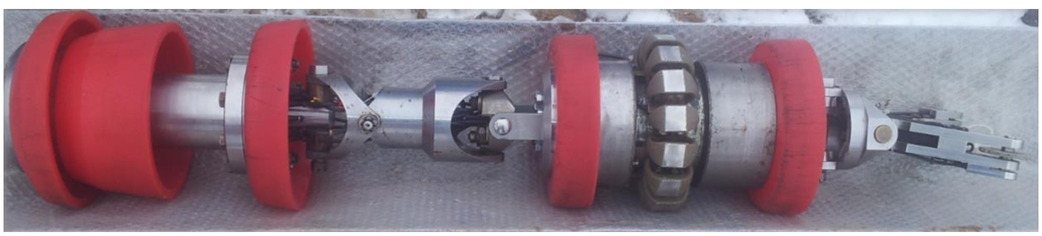

Figure 1: In-line-inspection tool

2. LITERATURE REVIEW

2.1. Pipelines defects diagnostics

Magnetic Flux Leakage (MFL) technique is the most common approach for oil and gas pipelines nondestructive testing. The data obtained during the pipeline inspection process is primarily analyzed by traditional machine learning (ML) methods. A comparison of performance among different ML methods for defects identification problem is presented in [1]. The main challenge for this approach is creating informative and important features that will be used as an input for ML methods. Usually, these diagnostics features are generated using expert knowledge and manually-created heuristics. It imposes the limitation on defects detection problem solving quality. A variety of most successful features is presented and analyzed in details in [2].

Deep Learning showed significant progress and achieved incredible results in numerous applications, just in the past few years. The image classification problem is one of the most successful applications of DL and Convolutional Neural Networks (CNNs) in particular. To automate the process of feature generation in MFL data analysis, CNNs can be used either. As an advantage, they can solve the defects detection and segmentation tasks at the same time. In literature there are examples of applying CNNs for defects detection [3], welds defect detection [4], welds and defects classification [5], defect size estimation [6]. For all mentioned applications, CNNs outperformed existing traditional approaches. Nevertheless, still, there are just a few works dedicated to MFL data analysis using DL. A number of particular problems that can be solved using a novel approach are not covered yet. For instance, we could not find any works on applying CNNs to defects segmentation task, despite the importance of this problem solving according to [3].

In this work, we want to research two different problems:

1. Defects detection (Picture classification task).
2. Defects segmentation (Semantic segmentation task).

For their solving, we propose CNNs of different architectures and compare their results with existing state-of-the-art approaches. Moreover, we research different preprocessing techniques for dealing with typical issues in the MFL data.

2.2. Unet

asd [7]

3. CONTRIBUTION

Viacheslav Kozitsin.

Iurii Katser.

Arman Alahyari.

Anton Hinneck.

Rahim Tariverdi.

4. 3RD PARTY CODE LIST

5. METHODS PIPELINES

We reimplemented CNN from [3] with one difference: we used squared pictures as an input, so we didn't implement Normalization layer (first layer in the Fig. 2).

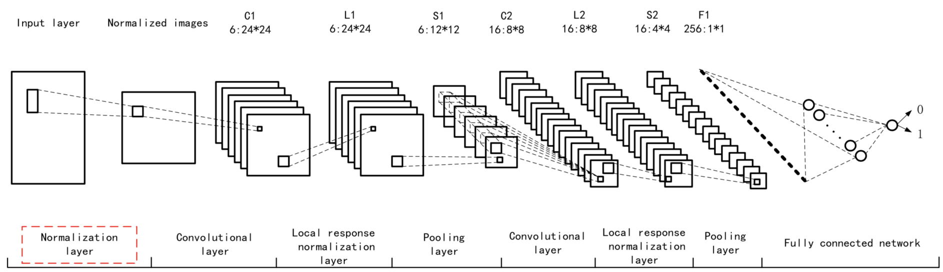

Figure 2: Architecture of CNN from [3]

Our model contains Batch Normalization layers instead of Local Response Normalization layers and Dropout. It consists of 5 Convolutional layers overall.

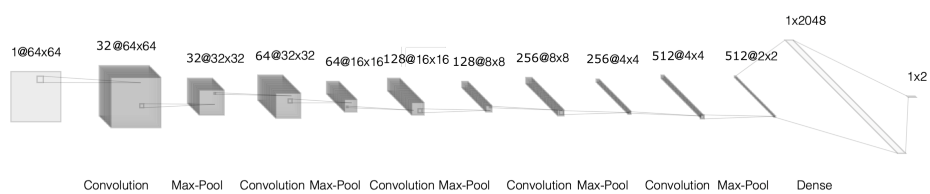

Figure 3: Architecture of our CNN

Batch size is equal to 64, so the input to the network has shape (64, 1, 64, 64). For all experiments we use Adam optimizer with initial learning rate 0.001 and Learning Rate scheduler with parameters: threshold=0.0001, factor=0.5, min lr=0.0001, patience=484. Also for all experiment the number of the epochs is equal to 12.

5.1. Data and preprocessing description

Although MFL data looks quite similar for different pipes and ILI tool types, it can differ a lot. The data mainly depends on pipe size, wall width, sensors geometry and other geometric characteristics. Moreover, the ILI tools differ a lot for different pipe sizes. We have a data from the 219 mm in diameter pipe. MFL dataset provides information about single inspection tool run. Dataset has 64 features collected as an array with a constant step along the ILI tool movement inside the pipe. Dataset has 4470704 samples that represent 15162.85 meters long pipeline part. It has 745 defects of different types and 1462 welds. Fig. 4 shows examples of normal data, data with a weld and with a defect. Attached to the dataset technical report contains information about welds and defects location, defects types, sizes and other related characteristics.

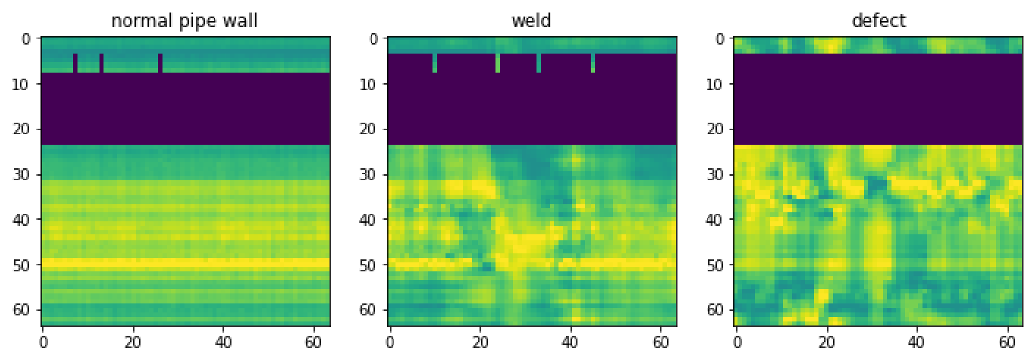

Figure 4: Example of the MFL data

Raw data has several issues that don't allow to solve CV problems without proper preprocessing. They are:

1. Sensors malfunctions (zeroed values cause bold horizontal line in Fig. 4);
2. Displaced origins between data and report files coordinates;
3. Inaccurate annotations, e.g. missed defect, wrong defect location, etc.

In addition to preprocessing, data were annotated for the segmentation task.

5.1.1. Sensors malfunctions problem

To deal with sensors malfunctions we suppose to fill the gaps (zeroed values) with values calculated by different methods:

1. Scaling of picture values to [0.5 : 1] range. Abnormal values (< 2000) are equal to 0.
2. Abnormal values are equal to the mean of normal values from one picture.

3. Abnormal values are equal to the mean of normal values over the column.
4. Abnormal values are equal to the mean of neighboring sensors over the column.
5. Abnormal values are equal to the interpolation results over the column.
6. Rerange initial set of values to 0...255 uint8 range.
7. Rerange initial set of values to 0...1 float range.

The results of all applied methods are presented in Fig. 5.

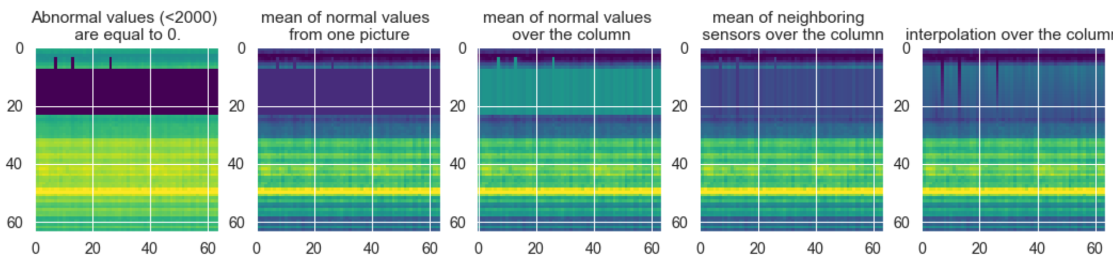

Figure 5: The results of missing values filling methods

5.1.2. Displaced origins problem

Since the data on the location of the robot did not match the defect location data from the report, it was necessary to merge the data. The key factor here turned out to be that the values of signal from magnetic flux sensors grow at the weld site, see picture 6.

The locations of maxima of these values have been founded and then combined with welds locations.

5.1.3. Inaccurate annotations problem

This problem is a common one for oil and gas pipeline nondestructive testing [1]. It appears to be a lot of missing defects that affects the quality of the problem. Besides there are wrong defect type and locations. For wrong location issue elimination we additionaly searched extremums around the provided location and chose the defects or welds taking into account new coordinates.

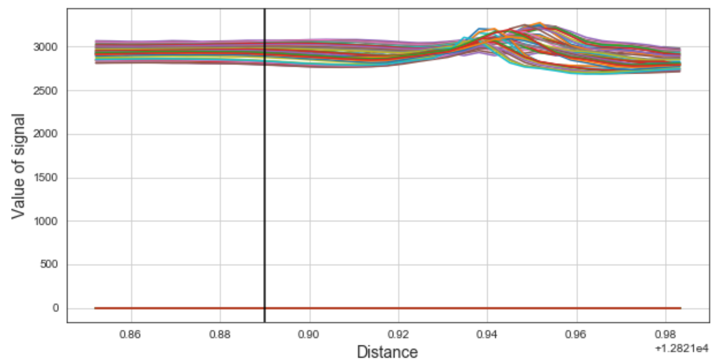

Figure 6: Location of weld. Black vertical line in accordance with defect location data from the report. Others lines are values from sensors.

5.1.4. Augmentation

Although we have a lot of data, we don't have a lot of defects and welds in comparison with normal pipe wall instances. We use augmentation procedure to balance classes of pictures and increase model's quality by increasing number of instances in small classes (defects, welds). As an augmentation tool we use Albumentations library [8]. For welds pictures we apply following augmentations:

1. Rotate (limit=180, p=1),
2. VerticalFlip (p=1),
3. HorizontalFlip (p=1),
4. ElasticTransform (p=1, alpha=20, sigma=120 * 0.05, alpha affine=120 * 0.03),
5. GridDistortion (p=1),
6. OpticalDistortion (p=1, distort limit=2, shift limit=0.5).

And for defects we apply following ones:

1. Transpose (p=1),
2. Rotate (limit=90, p=1),
3. Rotate (limit=180, p=1),
4. Rotate (limit=270, p=1,),
5. VerticalFlip (p=1),
6. HorizontalFlip (p=1),

7. ElasticTransform (p=1.0, alpha=20, sigma=120 * 0.05, alpha affine=120 * 0.03),
8. RandomRotate90 (p=1.0),
9. GridDistortion (p=1.0),
10. OpticalDistortion (p=1, distort limit=1, shift limit=0.3).

Examples of augmentations are shown in Fig. 7.

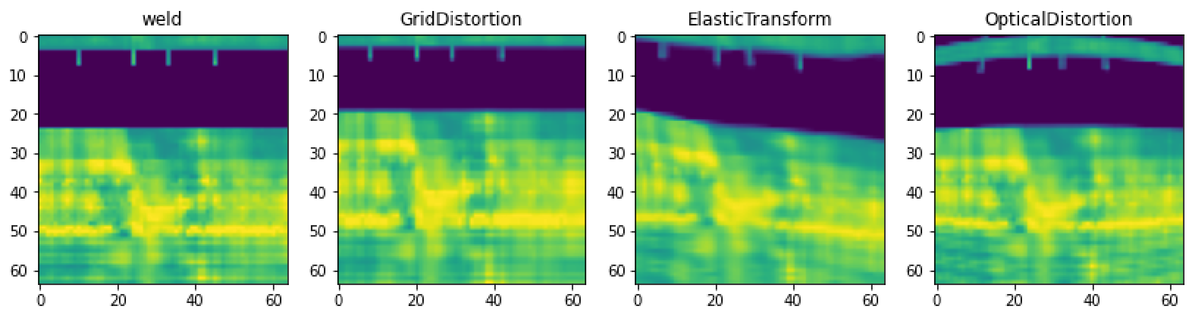

Figure 7: Examples of the augmentation results

Table 1: Dataset size for pipeline defects detection and segmentation problems

Classification			
Data	Normal pipe wall	Defect	Weld
Before augmentation			
Train	11106	1130	569
Validation	584	282	142
After augmentation			
Train	11106	11300	8535
Validation	584	282	142

5.2. Performance metrics

For the classification problem we use Accuracy. Accuracy is defined by the formula:

$$Acc = \frac{\sum_{i=0}^{N} 1_{\{\hat{y}_i = y_i\}}}{N},$$

where N - number of samples, $\hat{y}$ - predicted label, y - true label.

6. RESULTS PIPELINES

Table 2: Comparison of performance among different classification methods for binary classification problem

Method	$\hat{y} = y = 0$ (normal)	$\hat{y} = y = 1$ (defect/weld)	Average
CNN-2	95.55	82.08	89.88
CNN-5	97.95	**91.51**	**95.24**
CNN-5+LRN	**98.29**	89.86	94.74
Filling techniques comparison			
CNN-5 (filling 1)	97.95	**91.51**	**95.24**
CNN-5 (filling 2)	97.95	84.20	92.16
CNN-5 (filling 3)	97.26	83.02	91.27
CNN-5 (filling 4)	**98.63**	81.13	91.27
CNN-5 (filling 5)	98.12	81.84	91.27

Table 3: Comparison of performance among different classification methods for multiclass classification problem

Method	$\hat{y} = y = 0$ (normal)	$\hat{y} = y = 1$ (defect)	$\hat{y} = y = 2$ (weld)	Average
CNN-2	95.55	82.08	89.88	
CNN-5	**97.95**	**91.51**	**95.24**	
Filling techniques comparison				
CNN-5 (filling 1)	97.95	**91.51**	**95.24**	
CNN-5 (filling 2)	97.95	84.20	92.16	
CNN-5 (filling 3)	97.26	83.02	91.27	
CNN-5 (filling 4)	**98.63**	81.13	91.27	
CNN-5 (filling 5)	98.12	81.84	91.27	

7. RESULTS PIPELINES

8. METHODS TRANSMISSION

9. RESULTS TRANSMISSION

10. CONCLUSION

Today, manual analysis of a magnetographic image is a bottleneck for the diagnosis of pipeline transport, since it costs a lot of money and is limited by human resources. This study allows us to hope that this process can be fully automated, which is likely to make the analysis more reliable, faster and cheaper.

There can be defined several project development options:

1. better preprocessing, including manual pictures selection;
2. multiclass classification problem solving for defects;
3. defected welds detection;
4. applying some common archictures like VGG or ResNet;
5. adding layers to the last Convolutional layer for defect depth evaluation.

[1] A. Khodayari-Rostamabad, J. Reilly, N. Nikolova, J. Hare, S. Pasha, Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection, IEEE Transactions on Magnetics 45 (8) (2009) 3073–3084. doi:10.1109/tmag.2009.2020160.

[2] S. D.A., Development of the magnetic method of non-destructive testing due to automation of data processing and optimization of defect detection algorithms, Ph.D. thesis, National Research University "Moscow Power Engineering Institute" (2017).

[3] J. Feng, F. Li, S. Lu, J. Liu, D. Ma, Injurious or noninjurious defect identification from MFL images in pipeline inspection using convolutional neural network, IEEE Transactions on Instrumentation and Measurement 66 (7) (2017) 1883–1892. doi:10.1109/tim.2017.2673024.

[4] Oil pipeline weld defect identification system based on convolutional neural network, KSII Transactions on Internet and Information Systems 14 (3) (mar 2020). doi:10.3837/tiis.2020.03.010.

[5] L. Yang, Z. Wang, S. Gao, Pipeline magnetic flux leakage image detection algorithm based on multiscale SSD network, IEEE Transactions on Industrial Informatics 16 (1) (2020) 501–509. doi:10.1109/tii.2019.2926283.

[6] S. Lu, J. Feng, H. Zhang, J. Liu, Z. Wu, An estimation method of defect size from MFL image using visual transformation convolutional neural network, IEEE Transactions on Industrial Informatics 15 (1) (2019) 213–224. doi:10.1109/tii.2018.2828811.

[7] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 234–241. doi:10.1007/$978-3-319-24574-4_28$.

[8] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, Albumentations: fast and flexible image augmentations, Information 11 (2) (2020) 125.