Sewer-AI: Sustainable Automated Analysis of Real-World Sewer Videos Using DNNs

Rajarshi Biswas, Marcel Mutz, Piyush Pimplikar, Noor Ahmed and Dirk Werth
August-Wilhelm Scheer Institut, Uni-Campus D 5 1, 66123 Saarbrücken, Germany
{firstname.lastname}@aws-institut.de

Keywords: Computer Vision, Deep Learning, Industrial Application.

Abstract: Automated maintenance of sewer networks using computer vision techniques has gained prominence in the vision-research community. In this work, we handle sewer inspection videos with severe challenges. These obstacles hinder direct application of state-of-the-art neural networks in finding a solution. Thus, we perform an exhaustive study on the performance of highly successful neural architectures on our challenging sewer-video-dataset. For complete understanding we analyze their performance in different modes. We propose training strategies for effectively handling the different challenges and obtain balanced accuracy, F1 and F2 scores of more than 90% for 17 out of the 25 defect categories. Furthermore, for developing resource efficient, sustainable versions of the models we study the trade-off between performance and parameter pruning. We show that the drop in average performance of the networks is within 1% with more than 90% weight pruning. We test our models on the state-of-the-art Sewer-ML-dataset and obtained 100% true positive rate for 8 out of 18 defect categories in the Sewer-ML-dataset.

1 INTRODUCTION

The sewage system is an indispensable part of civic life all around the world. In Germany the sewer network is approximately 594,321 kilometers long as of 2018 while the US has almost 2.08 million kilometers in total network length (ASCE, 2017). Various factors cause progressive aging of the sewer networks. Thus, systematic inspection is required for maintaining the sewers. Sewer networks maintenance in most countries is carried out by the network operators who perform sampling inspections with the aid of TV-camera mounted robotic probes. These probes record the sewer conditions using a rotating video camera while being remotely driven by a professional inspector. This is the state-of-the-art in sewer inspection, however, the cost incurred in terms of man hours and finance is exorbitant. The inspectors need to examine the video for identifying the defects for extended time lengths which is exhausting, time-consuming and error prone. This results in inconsistent labels, i.e., the same image is assigned multiple labels by different inspectors. It is depicted in figure 4 where each image is assigned to multiple different categories.

It costs approximately 2500€ (Euros) to inspect one kilometer of the sewer network in Germany. This is compounded by the acute shortage of labor and domain experts. These factors render the current approach for monitoring sewers both unscalable and unsustainable. In this paper, we perform an in-depth analysis of the suitability of deep learning based models for automating the network inspection process and thereby reducing the cost burden on the maintenance industry. However, the prospect of direct application of deep learning models from the computer vision community is greatly limited due to several difficult challenges. The primary obstacle is the inconsistency in data collection, that is, there is no uniform standard with regards to equipment, format and resolution in which data is stored. Moreover, the dataset is

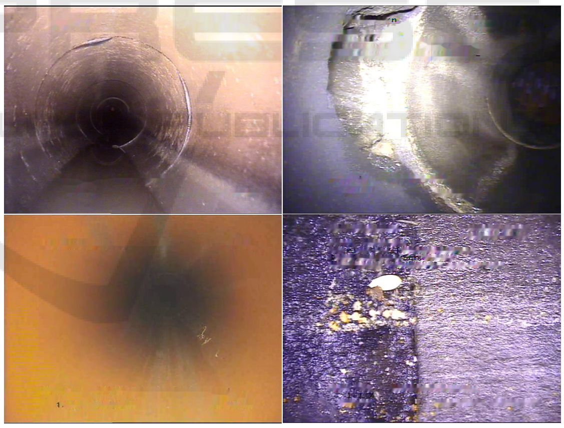

Figure 1: Challenging artefacts in real world sewer data.

Biswas, R., Mutz, M., Pimplikar, P., Ahmed, N. and Werth, D.
Sewer-AI: Sustainable Automated Analysis of Real-World Sewer Videos Using DNNs.
DOI: 10.5220/0011709400003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 897-904
ISBN: 978-989-758-626-2; ISSN: 2184-4313

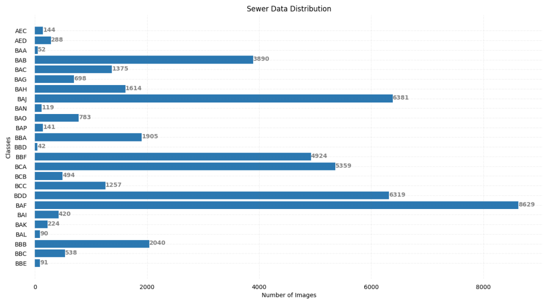

Figure 2: Highly imbalanced class distribution in real world sewer data.

characterized by poor quality as most of the recorded videos contain blur, grain and distortion artefacts (refer to figure 1). There is also huge variation in terms of lighting distribution, camera orientation, material characteristics and inconsistent labeling, i.e., there is high inter-annotator disagreement (refer to figure 4). In addition to this, the dataset suffers from highly imbalanced distribution among the labels (as can be seen in figure 2). In order to systematically handle all these obstacles we perform three experiments with increasing degree of complexity for the classification task. We study the performance of the chosen neural architectures, HRNet (Wang et al., 2019), ResNet-152 (He et al., 2016), MobileNet_V3_Large (Sandler et al., 2018), DenseNet-264 (Huang et al., 2017), Inception_V4 (Szegedy et al., 2016) and Efficient-Net (Tan and Le, 2019) respectively through these experiments. Our motivation behind these experiments is to determine the comparative difficulty-level posed by the different defect classes to these networks under the non-uniform data collection and labeling standards of the sewer maintenance industry. In the first experiment, *Single-defect classification (E1)*, we test the networks for individual defect category identification using dataset containing only positive instances from specific defect category and all the negative instances present in our dataset. Then in the second experiment, *One-vs-all classification (E2)*, we observe the performance of the networks in distinguishing a specific defect category against all the categories, including the non-defect instances, together. Through this experiment, we test the influence of the low inter-class distinction among some of the labels and other inconsistencies in our dataset on the classification performance of the networks for each defect category. Finally, as the last experiment *Classification under Weight-pruning (E3)* we test the drop in performance of the trained models when performing inference with heavily pruned models. Our goal here is to study the suitability of the chosen networks under resource constrained environments. To summarize, our contributions in this paper are the following:

- Comprehensive performance analysis of the most effective neural network architectures for sewer video classification using three different experiments with gradual increase in complexity.
- Performance estimation of the neural networks under heavy weight pruning for developing resource efficient and sustainable solutions.
- Training strategies for handling acute data imbalance in the sewer data.

2 RELATED WORK

The success of deep learning models in solving computer vision problems has prompted their widespread adoption in various industrial applications. The civil engineering discipline too is witnessing a steady adoption of different techniques for conducting visual inspection of large civic infrastructures. There are approaches that borrow different classical vision methods for analysing the condition of pavements, bridges, roads tunnels etc. Additionally, there is emphasis too on using 3D modeling for digital reconstruction and visualization of these large infrastructures for improving the inspection quality. The area of automatic sewer network analysis also generates attention in the vision research community. This has lead to the development of crack detection in the sewers using image processing (Halfawy and Hengmeechai, 2014) and segmentation (Iyer and Sinha, 2006) methods. Mathematical morphology is used by the authors in (Sinha and Fieguth, 2006) for classifying cracks, holes and joints in the sewer pipes whereas the work (Dang et al., 2018) uses morphological operations along with edge-detection, binarization for identifying defects by recognizing the text on the sewer videos. Employment of models with task-specific features or heuristic decision rules are reported in (Myrans et al., 2018). However, classical approaches such as the ones mentioned require a lot of feature engineering, extensive pre-processing routines and suffer more from noisy, low-quality data. Deep learning based approaches resolve this problem. As for example, convolutional neural networks or similar deep neural network derived applications learn directly from the data obviating feature engineering with little to almost non-existent pre-processing efforts. The authors in (Cha et al., 2017; Montero et al., 2015) demonstrate the efficiency of convolutional neural networks over conventional techniques for tunnel inspection. The introduction of deep learning has led to improvements in image & video analysis (Fang et al., 2020; Moradi et al., 2020; Wang et al., 2021), estimation of water level (Haurum et al., 2020; Ji et al., 2020), defect identification (Cheng and Wang, 2018; Kumar et al., 2020; Yin et al., 2020), segmentation (Pan et al., 2020; Piciarelli et al., 2019; Wang and Cheng, 2020) along with the handling of the classification problem in the multi-class setting (Hassan et al., 2019; Kumar et al., 2018; Li et al., 2019; Meijer et al., 2019; Xie et al., 2019). The steady industrial and academic interest in automating the sewer maintenance process has led to utilization of different technologies. For instance, the works (Duran et al., 2002; Liu and Kleiner, 2013) report the use of various sensors in this area. More specifically, there is use of acoustic sensors (Iyer et al., 2012) , depth sensors (Alejo et al., 2017; Haurum et al., 2021; Henriksen et al., 2020) and laser scanners (Lepot et al., 2017; Dehghan et al., 2015) for defect detection and reconstruction of the sewer pipes. However, none of the works mentioned above study the performance of different highly effective neural networks for classifying sewer defects in three different modes as described earlier. Moreover, through the pruning experiment we derive performance bounds for developing lightweight versions of the networks suitable for operation under resource constrained environments. To the best of our knowledge our work is the first in the area of automated sewer video analysis to perform this study which can help in developing sustainable and energy efficient solutions. Furthermore, we benchmark our models against the Sewer-ml dataset (Haurum and Moeslund, 2021) for understanding their generalization capabilities and obtain better results compared to them.

3 DATASET

The data in our work is collected by registered sewer network operators from 221 municipalities in Germany. The operators employ human experts to identify the defects in the videos and use two software, IBAK and PIPEX respectively, for annotating the video data with the identified defects. The defects are categorized into standard defect inspection codes used all across Germany for sewer maintenance. A total of 10,205 video files are collected from the 221 municipal locations amounting to 2.37 Terra-Bytes. For every video, we extract defect frames from the time interval marked as featuring a set of defects and extract the negative frames using an offset of 10 seconds outside this interval in either direction of the time axis. Following this strategy we collect 156,028 frames, having resolutions of 576x480 and 576x768 respectively, in total amounting to 87.10 GB of frame data. The extracted frames contain different information in the form of text overlay. However, such textual information can wrongly influence the decision making process of the neural classifiers. So, we remove the text from the images using optical character recognition, for identifying the text, followed by inpainting (Bertalmio et al., 2001). The effect of this frame processing is seen in figure 3. There are 25 different defect categories, however, the distribution of instances pertaining to each of these categories is highly imbalanced as shown in figure 2.

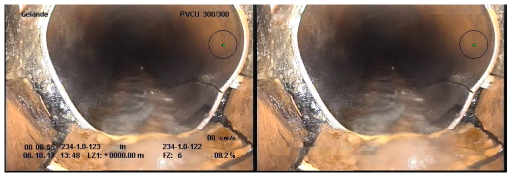

Figure 3: Inpainting based text removal from frames.

(a) BAG,BAH,BCA,BAJ

(b) BAJ,BAB

(c) BBB,BBA

(d) BBB,BAF

Figure 4: Ambiguous instances with multiple labels.

4 METHOD

We perform three different experiments with the chosen neural networks to handle the challenges in our real world dataset. Our goal is to systematically estimate the performance of the different chosen networks. So, as our first experiment (*E1*), we train the networks to perform single defect category identification. This is followed by the second experiment (*E2*), where we test the networks ability to perform one-vs-all classification and finally in our last experiment (*E3*), we prune systematically the parameters of the network up to 98 percent of it's original value and study the drop in performance compared to the results obtained from experiments *E1* and *E2* respectively. The motivation behind our experiments is to handle challenges in our dataset in an increasing degree of complexity. Since our dataset suffers from

Table 1: Average performance of the networks for experiments *E1, E2, E3*.

Neural Net	E1			E3(E1 + Pruning)			E2			E3(E2 + Pruning)		
	BA	F1	F2	BA	F1	F2	BA	F1	F2	BA	F1	F2
HRNet	**85.27**	0.81	**0.83**	**84.46**	**0.82**	0.80	77.67	0.71	0.65	79.12	0.74	0.69
ResNet	82.77	**0.82**	0.82	81.54	0.82	**0.82**	78.65	0.73	0.67	78.75	0.73	0.67
EfficientNet	75.39	0.73	0.73	74.88	0.70	0.70	73.34	0.63	0.59	72.75	0.67	0.65
MobileNet	82.83	0.80	0.78	78.76	0.72	0.70	79.34	0.74	0.68	74.90	0.62	0.58
Inception	82.05	0.77	0.76	82.76	0.79	0.78	76.55	0.70	0.66	76.41	0.69	0.66
DenseNet	82.25	0.81	0.75	81.25	0.78	0.76	**80.55**	**0.76**	**0.72**	**80.94**	**0.76**	**0.70**
Best Avg. Perf.	89.38	0.88	0.88	90.13	0.89	0.90	84.59	0.82	0.79	84.08	0.83	0.82

various limitations, discussed previously, we first consider the simple setting of single defect category identification (*E1*). Subsequently, we consider the comparatively complex task of one-vs-all classification with our dataset (*E2*) followed by the parameter pruning experiment (*E3*).

We divide our dataset into train, validation and test sets using a 80-10-10 split for each category. However, it is important to note that our dataset is highly skewed towards negative images, that is, the number of frames without any defects is much larger compared to the number of frames in any individual defect category. For handling this imbalance while training our models we follow a specific training strategy. In this approach, we first consider all the positive training instances of a particular class and an equal number of random negative instances in our first epoch of training. In the next epoch, we keep the same positive instances and take another set of random negative images of same size and continue training. In this way, in every epoch we keep the same positive images but keep considering set of random negative instances of same size from our negative frames. We repeat this strategy for 100 epochs for training the chosen neural networks. We have intentionally extracted negative frames from the videos in much larger proportion compared to the frames containing the defects as we wanted to understand the performance of the models under real conditions where on average the defects occur in a highly infrequent manner.

For training the models in the one-vs-all classification mode, experiment *E2*, we use the pre-trained single defect classifiers and fine-tune the models on the dataset reorganized in a one-vs-all way. We take the specific defect category as the positive class and all the other defect categories including the non-defective instances as the negative class. For ease of reference we assign the non-defective instances to the "No Defect" category. In each epoch, we take the same number of random "No Defect" instances as the number of positive instances, and in the negative dataset we further add the same number of instances from the other defect categories with equal distribution from each category. Now, our negative dataset size is twice the size of our positive dataset. This setup ensures that the model will encounter a higher proportion of "No Defect" instances compared to specific defect category instances during training. At the same time, in addition to the "No Defect" instances, it will be able to distinguish the specific defect category from the other defect categories. For the experiments *E1 and E2*, we optimize binary cross entropy loss using a batch size of 16 and perform hyperparamter search over the optimizers Adam, SGD and over learning rates of 1e-3, 1e-4 respectively. We compute balanced accuracy, F1 and F2 scores in all our experiments for measuring the performance of the models on the test set.

With the view of developing energy efficient solutions and reducing the computational burden we study the influence of parameter pruning on the performance of our models, obtained from *E1 and E2* respectively, under experiment *E3*. We use global pruning for this purpose. In this approach, different layers of the model are considered as a single global module, and the weights closer to zero are pruned depending upon the amount of pruning percentage specified. This type of pruning converts dense matrix representation to a sparse matrix, due to which the model size gets reduced substantially. We tried pruning percentages of 30, 50, 70, 90, 98 respectively. We keep on increasing the pruning percentage as specified, in succession, if the balanced accuracy drop on test dataset is less than 1 percent. If it drops more than one percent, we stop pruning that network. This way we are able to do dynamic pruning and come up with the best pruning threshold or percentage for all the models for each defect category. Following this strategy, we are able to prune the models to a tenth of it's original size which significantly helps us in reducing the storage and computational budget. Through this pruning exercise we try to find the most accountable weights of our models responsible for making correct predictions and discard those that are either irrelevant or degrade the model's performance. With each pruning threshold, we check performance of each pruned model on

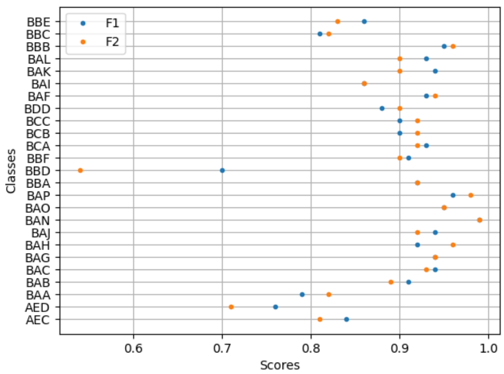

Figure 5: F1 and F2 scores obtained for each defect category.

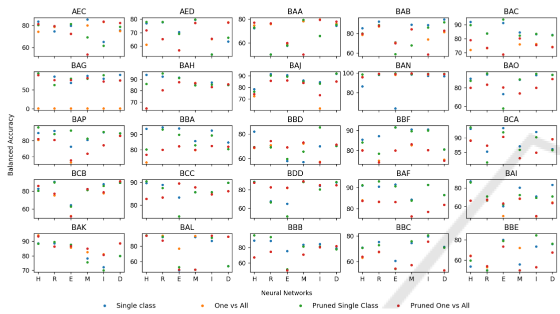

Figure 6: Balanced Accuracy for each defect category under experiments *E1, E2 and E3*.

the test set that is used in our previous experiments and report the results obtained with the pruned models.

5 RESULTS AND DISCUSSION

We trained six different classifiers, HRNet, ResNet-152, EfficientNetV2-L, MobileNet_V3_Large, Inception_V4 and DenseNet-264 with significantly different number of trainable parameters, for identifying each defect category. We follow the training strategy described in section 4 for handling the critical imbalance in our dataset. The results obtained with the chosen networks for classifying each defect category under the experiments *E1, E2 and E3* are reported in detail in figures 5 and 6. The balanced accuracy scores obtained from the chosen neural networks under all three experiments for every defect category in our sewer dataset can be seen in figure 6. We report the best F1 and F2 scores obtained for each defect category from our experiments in figure 5.

Under experiment *E1* we trained the six different neural networks on each of the 25 defect categories. Out of the 25 defect categories, we obtained more than 90 percent balanced accuracy on 14 categories, more than 85 percent balanced accuracy on 5 categories while the rest gave around 80 percent on the same metric. The defect category BAF is the most frequently occurring defect in our dataset with a total number of 8629 instances. On this category we obtained a balanced accuracy score of 91.5 percent. On all the categories containing more than 1000 instances we achieved more than 90 percent balanced accuracy. Interestingly, on the remaining 10 categories with less than 1000 instances we achieved balanced accuracy score between 85 percent and 90 percent. In some of the rare categories, despite having very less instances but due to its visually distinct nature, our models were able to achieve very high accuracy. For instance, we achieved 98.70 percent accuracy on the BAN class which has only 119 instances and 91.91 percent accuracy for the BAP class which has 138 instances respectively. This shows that if the class is visually distinct, then even with as less as 100 images, our models are capable of achieving good results. HRNet (19.25 million parameters) was the best performing model with an average balanced accuracy of 85.27 percent and a F2 score of 0.83 across all the defect categories (refer to Table 1). The second best performing model was MobileNet_V3_Large, which is also the smallest in size with 5.48 million parameters compared to the other networks, with an average balanced accuracy of 82.83 percent. While DenseNet-264 (72.08 Million parameters) and ResNet-152 (60.19 Million parameters) achieved average balanced accuracy scores of 82.25 percent and 82.77 percent respectively. These results demonstrated models with smaller parameter

Table 2: Performance of our single defect models on the Sewer-ML test set.

SewerML-Code	Our-Code	TPR on SewerML
VA	BDD	2.20%
RB	BAB	100.00%
OB	BAF	2.49%
PF	—	—
DE	BAA	100.00%
FS	BAG	98.83%
IS	BAI	2.15%
RO	BBA	100.00%
IN	BBF	100.00%
AF	BBC	62.09%
BE	BBC	68.26%
FO	BBE	0.0%
GR	BCA	0.0%
PH	BAH	100.00%
PB	BAG	96.47%
OS	BAH	100.00%
OP	BAH	100.00%
OK	BAH	100.00%

count, i.e., MobileNet_V3_Large achieved accuracy scores close to or even better than much larger models. Such smaller models might be easier to integrate on the inspection robots and can be used for quick reliable inference during sewer inspection. This fact further motivated us in the direction of pruning (*E3*) in order to explore the feasibility of creating smaller models that require less computational power but can still achieve high, reliable performance.

In our next experiment *E2* we tested the chosen neural networks ability to distinguish each defect category against all the other defect categories including the non-defective instances, i.e., extracted images with no defects. The results obtained in this experiment establish DenseNet-264 as the best performing neural network with an average balanced accuracy of 80.55 percent, F1 score of 0.76 and F2 score of 0.72 respectively across all the defect categories (refer to Table 1). This is different from the results obtained from experiment *E1* where HRNet was the best performing among all the chosen networks. In the results from *E2* we witness a drop of 4.72 percent in average balanced accuracy, however, we expected this since there is very low inter-class variation among many defect categories. For example, the defect categories BAG, BAH, BAJ and BDE are all related to errors in pipe connections and are very similar in their visual appearance. Therefore, having similar images in both the positive and negative classes under the one-vs-all setting decrease the performance of the networks. As further examples, the defect categories BAB, BAC, BAF and BAO are related pipe damage and contain instances which are very similar with some being visually indistinguishable. In spite of this, we were able to obtain more than 90 percent balanced accuracy for 5 defect categories. Visually distinct categories like BAN yielded very high balanced accuracy of 99.57 percent while performance on the shifted connection category BAJ, similar to other connection related defect categories like BAG, BAH, BDE, dropped to 86.01 percent from 91.95 percent obtained under experiment *E1*. Thus, the models under one-vs-all strategy (*E2*) certainly helped us in differentiating one defect category from all the other categories including the non-defective category, but due to low inter-class variability in few of the categories the performance of the models dropped.

Under experiment *E3*, we performed pruning on all the single-defect and one-vs-all classifiers developed in the experiments *E1 and E2* respectively. With the dynamic pruning strategy, described earlier, we were able to find models with least size but still capable of achieving performance similar to the original models. The average pruning on single-defect classifiers from *E1* achieved in our experiment across all the classes was around 50 percent for all the neural networks apart from DenseNet-264 where we achieved 75 percent average pruning. For one-vs-all classifiers from *E2*, the average pruning ranged from 60 to 85 percent. We were able to prune more than 90 percent for 44 different single-defect models and 73 different one-vs-all models. Pruning significantly reduced model size, as with 90 percent or more pruning the model size went down to 1/10th of the original model. As a result, a significant number of models were reduced to a much smaller size and it helped us to reduce the storage budget significantly. Remarkably, 55 single-defect classifiers under pruning gave better results than their corresponding non-pruned versions. Similarly, 62 of our one-vs-all pruned models performed better compared to their corresponding non-pruned versions. Over here 6 models gave significant results, with an increase in balanced accuracy by 6 to 11 percentage points. This shows that pruning resulted in finding the most important weights of the model and it discarded the weights which were relatively less important or which reduced the model's performance. Thus pruning gave us the best compact models which resulted in achieving higher performances. In majority of the pruned models, the drop in balanced accuracy scores stayed within the 1 percent mark of the scores obtained with the corresponding non-pruned versions. Finally, after conducting experiments *E1, E2 and E3* we found that the average best performance over the 25 different defect categories considering only the best performing networks is 90.13 percent (refer to Table 1). This shows that if we use the best performing model for each defect category, we are able to achieve 90% average balanced accuracy across all the 25 classes.

Lastly, we tested the performance of our models on the largest publicly available standardized dataset for sewer inspection (Haurum and Moeslund, 2021). Our goal was to determine whether our models are able to generalize to other data sources and to observe the levels of performance they can offer on these data. We identified corresponding defect categories in our respective datasets and measure the performance of our models on the Sewer-ML test set. To be precise, we computed the True Positive Rate (TPR) or Recall and found that on 8 out of the 18 defect categories in the Sewer-ML dataset our models achieve a TPR of 100 percent and more than 95 percent on the defect catagories FS and PB in the Sewer-ML dataset (refer to Table 2). On two defect categories, our corresponding models achieve TPR of more than 60 percent while on the remaining six very low performance is observed. We believe this can be due to improper mapping between the defect categories in our dataset compared to their dataset.

6 CONCLUSION

To summarize, in this paper we thoroughly analyzed the performance of six different neural networks for identifying the defects present in real world sewer inspection videos. We outlined the critical challenges present in our dataset, proposed strategies for training neural networks for handling these challenges and performed three different experiments, *E1, E2 and E3* respectively, to study the effectiveness of the chosen networks in handling such challenging data. We found that for experiment *E1* HRNet was the best performing neural architecture with an average balanced accuracy of 85.27% across all the defect categories. For the more complex task in experiment *E2* we found that DenseNet-264 performed the best with an average balanced accuracy of 80.55%. In our results obtained for experiment *E2* we witnessed a drop of 4.72% on average in balanced accuracy. We believe this is due to the very low inter-class variation among the defect categories present in our dataset. Under the pruning experiment in *E3* we were able to significantly reduce the model size. With pruning at 90 percent or more, the models got reduced to one-tenth of it's original size. In total, we were able to prune 117 different class specific models, from experiments *E1 and E2* combined, by more than 90 percent. However, the drop in balanced accuracy scores stayed within the 1 percent mark of the scores obtained with the corresponding non-pruned versions. Finally, we tested our models on the Sewer-ML dataset and obtained very high recall of 100% on 8 out of the 18 defect categories presented in the Sewer-ML dataset.

ACKNOWLEDGEMENTS

This research was funded in part by the German Federal Ministry of Education and Research (BMBF) under the project KIKI (grant number 02WDG1594A). It is a joint project of the August-Wilhelm-Scheer Institut, TU Clausthal, AHT AquaGemini GmbH, EURAWASSER Betriebsführungsgesellschaft mbh, IBAK Helmut Hunger GmbH & Co. KG and Entsorgungsverband Saar KdöR. August-Wilhelm-Scheer Institut is mainly entrusted with conducting research in artificial intelligence for automatically diagnosing the condition of the sewers.

REFERENCES

Alejo, D., Caballero, F., and Merino, L. (2017). Rgbd-based robot localization in sewer networks. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4070–4076.

ASCE (2017). American society of civil engineers 2017 infrastructure report card.

Bertalmio, M., Bertozzi, A. L., and Sapiro, G. (2001). Navier-stokes, fluid dynamics, and image and video inpainting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE.

Cha, Y.-J., Choi, W., and Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.*, 32(5):361–378.

Cheng, J. C. and Wang, M. (2018). Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Automation in Construction*, 95:155–171.

Dang, L. M., Hassan, S., Im, S., Mehmood, I., and Moon, H. (2018). Utilizing text recognition for the defects extraction in sewers cctv inspection videos. *Computers in Industry*, 99.

Dehghan, A., Mehrandezh, M., and Paranjape, R. (2015). Defect detection in pipes using a mobile laser-optics technology and digital geometry. *MATEC Web of Conferences*, 32:06006.

Duran, O., Althoefer, K., and Seneviratne, L. (2002). State of the art in sensor technologies for sewer inspection. *IEEE Sensors Journal*, 2(2):73–81.

Fang, X., Guo, W., Li, Q., Zhu, J., Chen, Z., Yu, J., Zhou, B., and Yang, H. (2020). Sewer pipeline fault identification using anomaly detection algorithms on video sequences. *IEEE Access*, 8:39574–39586.

Halfawy, M. and Hengmeechai, J. (2014). Efficient algorithm for crack detection in sewer images from closed-circuit television inspections. *Journal of Infrastructure Systems*, 20:04013014.

Hassan, S., Dang, L. M., Mehmood, I., Im, S., Choi, C., Kang, J., Park, Y.-S., and Moon, H. (2019). Underground sewer pipe condition assessment based on convolutional neural networks. *Automation in Construction*.

Haurum, J. B., Allahham, M. M. J., Lynge, M. S., Henriksen, K. S., Nikolov, I. A., and Moeslund, T. B. (2021). Sewer defect classification using synthetic point clouds. In *VISIGRAPP*.

Haurum, J. B., Bahnsen, C. H., Pedersen, M., and Moeslund, T. B. (2020). Water level estimation in sewer pipes using deep convolutional neural networks. *Water*.

Haurum, J. B. and Moeslund, T. B. (2021). Sewer-ml: A multi-label sewer defect classification dataset and benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13456–13467.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Henriksen, K. S., Lynge, M. S., Jeppesen, M. D. B., Allahham, M. M. J., Nikolov, I. A., Haurum, J. B., and Moeslund, T. B. (2020). Generating synthetic point clouds of sewer networks: An initial investigation. In *Augmented Reality, Virtual Reality, and Computer Graphics: 7th International Conference, AVR 2020, Lecce, Italy, September 7–10, 2020, Proceedings, Part II*, page 364–373, Berlin, Heidelberg. Springer-Verlag.

Huang, G., Liu, Z., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269.

Iyer, S. and Sinha, S. (2006). Segmentation of pipe images for crack detection in buried sewers. *Comp.-Aided Civil and Infrastruct. Engineering*, 21:395–410.

Iyer, S., Sinha, S. K., Pedrick, M., and Tittmann, B. R. (2012). Evaluation of ultrasonic inspection and imaging systems for concrete pipes. *Automation in Construction*, 22:149–164.

Ji, H. W., Yoo, S. S., Lee, B.-J., Koo, D. D., and Kang, J.-H. (2020). Measurement of wastewater discharge in sewer pipes using image analysis. *Water*, 12(6).

Kumar, S., Wang, M., Abraham, D., Jahanshahi, M., Iseley, T., and Cheng, J. (2020). Deep learning–based automated detection of sewer defects in cctv videos. *Journal of Computing in Civil Engineering*, 34:04019047.

Kumar, S. S., Abraham, D. M., Jahanshahi, M. R., Iseley, T., and Starr, J. (2018). Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks. *Automation in Construction*, 91:273–283.

Lepot, M., Stanić, N., and Clemens, F. H. (2017). A technology for sewer pipe inspection (part 2): Experimental assessment of a new laser profiler for sewer defect detection and quantification. *Automation in Construction*, 73:1–11.

Li, D., Cong, A., and Guo, S. (2019). Sewer damage detection from imbalanced cctv inspection data using deep convolutional neural networks with hierarchical classification. *Automation in Construction*, 101:199–208.

Liu, Z. and Kleiner, Y. (2013). State of the art review of inspection technologies for condition assessment of water pipes. *Measurement*, 46:1–15.

Meijer, D., Scholten, L., Clemens, F., and Knobbe, A. (2019). A defect classification methodology for sewer image sets with convolutional neural networks. *Automation in Construction*, 104:281–298.

Montero, R., Victores, J., Martínez, S., Jardón, A., and Balaguer, C. (2015). Past, present and future of robotic tunnel inspection. *Automation in Construction*, 59:99–112.

Moradi, S., Zayed, T., Nasiri, F., and Golkhoo, F. (2020). Automated anomaly detection and localization in sewer inspection videos using proportional data modeling and deep learning–based text recognition. *Journal of Infrastructure Systems*, 26:04020018.

Myrans, J., Everson, R. M., and Kapelan, Z. (2018). Automated detection of faults in sewers using cctv image sequences. *Automation in Construction*.

Pan, G., Zheng, Y., Guo, S., and Lv, Y. (2020). Automatic sewer pipe defect semantic segmentation based on improved u-net. *Automation in Construction*, 119:103383.

Piciarelli, C., Avola, D., Pannone, D., and Foresti, G. L. (2019). A vision-based system for internal pipeline inspection. *IEEE Transactions on Industrial Informatics*, 15(6):3289–3299.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520.

Sinha, K. and Fieguth, W. (2006). Morphological segmentation and classification of underground pipe images. *Mach. Vision Appl.*, 17(1):21–31.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.

Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., and Xiao, B. (2019). Deep high-resolution representation learning for visual recognition. *TPAMI*.

Wang, M. and Cheng, J. C. P. (2020). A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. *Comput.-Aided Civ. Infrastruct. Eng.*, 35(2):162–177.

Wang, M., Kumar, S. S., and Cheng, J. C. (2021). Automated sewer pipe defect tracking in cctv videos based on defect detection and metric learning. *Automation in Construction*, 121:103438.

Xie, Q., Li, D., Xu, J., Yu, Z., and Wang, J. (2019). Automatic detection and classification of sewer defects via hierarchical deep learning. *IEEE Transactions on Automation Science and Engineering*, 16(4):1836–1847.

Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., and Kurach, L. (2020). A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in Construction*, 109:102967.