Sewer-ML: A Multi-Label Sewer Defect Classification Dataset and Benchmark Supplementary Materials

Joakim Bruslund Haurum Thomas B. Moeslund
Visual Analysis and Perception (VAP) Laboratory, Aalborg University, Denmark
joha@create.aau.dk, tbm@create.aau.dk

A. Content

In these supplementary materials we describe in further detail aspects of the dataset and the training process and performance of the tested methods. Specifically, the following will be described:

- Additional examples from the Sewer-ML dataset (Section B).
- Further insights into the Sewer-ML dataset (Section C).
- Full training details and metric performance for the Faster-RCNN text detector (Section D).
- Full details on the Extra Trees hyperparameter grid search (Section E).
- The loss curves of the trained multi-label classification methods (Section F).
- Ablation study of the two-stage methods (Section G).
- Results when evaluating using the common multi-label performance metrics (Section H).

B. Sewer-ML Dataset Examples

In this section we present more examples of the images in the Sewer-ML dataset. All images are annotated using the Danish inspection standard containing 18 classes [3], listed in Table 1. In Figure 1 we present examples of different cases with several co-occurring classes. In Figure 12 we present five examples of each class, where only the mentioned class is present.

C. Sewer-ML Dataset Insights

In this section, we describe the available information in the Sewer-ML dataset in more detail. First, we report the number of occurrences for each class in the dataset splits, see Table 2, where it is observed that the distribution of the classes is similar across the different splits.

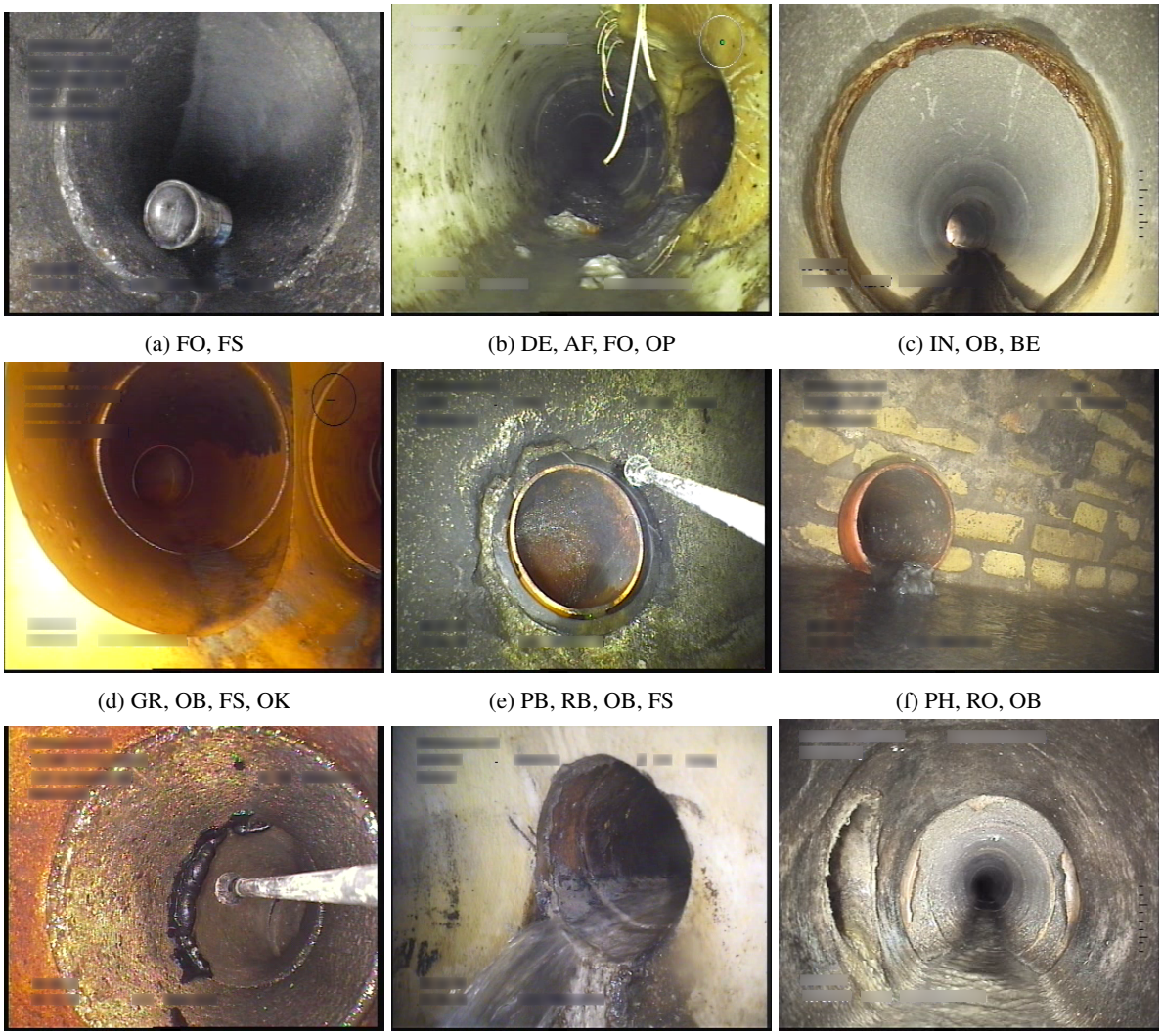
Table 1: **Sewer inspection classes.** Overview and short description of each annotation class [3].

Code	Description
VA	Water Level (in percentages)
RB	Cracks, breaks, and collapses
OB	Surface damage
PF	Production error
DE	Deformation
FS	Displaced joint
IS	Intruding sealing material
RO	Roots
IN	Infiltration
AF	Settled deposits
BE	Attached deposits
FO	Obstacle
GR	Branch pipe
PH	Chiseled connection
PB	Drilled connection
OS	Lateral reinstatement cuts
OP	Connection with transition profile
OK	Connection with construction changes

Moreover, we look into the pipe properties associated with each image. Each image contains information on the pipe shape, material, dimension, and water level.

In Figure 2 we plot the distribution of the eight different pipe material types for the images in each split. We find that the concrete, vitrified clay, plastic, and lining materials are the most common materials in the Sewer-ML dataset. We also observe that all material types are equally represented across the splits, except for the "Brickwork" and "Unknown" material types. The reason these material types are skewed for the validation and test sets, is due to these materials being rarely used anymore, and therefore rarely occur in the sewer inspection videos. Therefore, the images containing these material types are from a small subset of pipes, which were not evenly spread out across the splits.

In Figure 3 we plot the distribution of the six different



(a) FO, FS (b) DE, AF, FO, OP (c) IN, OB, BE

(d) GR, OB, FS, OK (e) PB, RB, OB, FS (f) PH, RO, OB

(g) IS, OK (h) OS, PF (i) OP, OK, FS, OB

Figure 1: **Sewer-ML data examples with co-occurring classes.** A subset of the images in the Sewer-ML showcasing images with multiple classes co-occurring and all annotated classes represented. The class codes are described in Table 1.

Table 2: **Class occurrences per split.** The number of occurrences for each class per dataset split.

Split	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Training	45,821	184,379	16,254	19,084	283,983	6,271	22,637	23,782	74,856	66,499	5,010	53,986	23,685	6,746	4,625	5,325	154,624	552,820
Validation	5,538	23,624	2,021	2,038	36,218	881	2,917	2,812	9,059	7,929	597	6,889	3,432	765	457	612	19,655	68,681
Test	5,501	23,264	1,949	2,307	35,781	924	2,684	3,235	9,182	8,720	649	6,726	2,962	833	530	533	19,420	69,221
Total	56,860	231,267	20,224	23,429	355,982	8,076	28,238	29,829	93,097	83,148	6,256	67,601	30,079	8,344	5,612	6,470	193,699	690,722

pipe shapes for the images in each of the dataset splits. We find that the circular type is by far the most common pipe shape, followed secondly by conical pipes, whereas the remaining pipe shapes only appear a few thousand times each. As with the pipe material, we see that distribution of pipe shapes are similar between dataset splits, except for the "Eye shaped", "Rectangular", and "Other" pipe shapes. This is again due to these pipe shapes occurring in a limited set of sewer inspections, and have therefore not been evenly divided across the splits.

In Figure 4 we plot the occurrences of the pipe dimensions associated with each image. The dimension is denoted in millimeters, as per the industry standard. We see that the majority of images are from pipes with a diameter of 100–1,000 millimeters, with a skew towards 100 millimeters. We

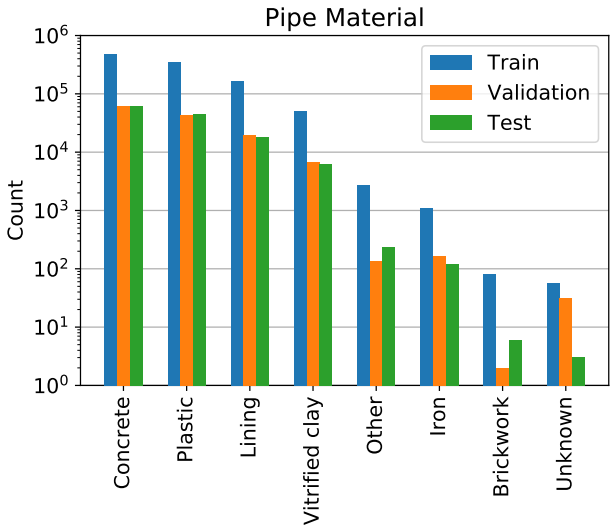


Figure 2: **Distribution of the pipe materials**. We plot the occurrence frequencies for each of the eight pipe materials in the dataset, for each dataset split. Note that the y-axis is log-scaled.

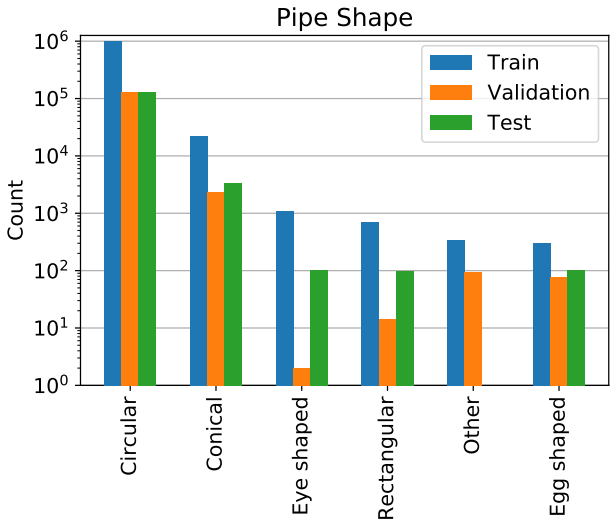


Figure 3: **Distribution of the pipe shapes.** We plot the occurrence frequencies for each of the six pipe shapes in the dataset, for each dataset split. Note that the y-axis is log-scaled.

observe that the distribution of the pipe dimension for the training, validation, and test splits appears to be similar in shape, as expected.

In Figure 5 we plot the distribution of the different water level classes for each data split. We find that the distribution of the water level classes is similar across the three dataset splits We also observe that the majority of the images have

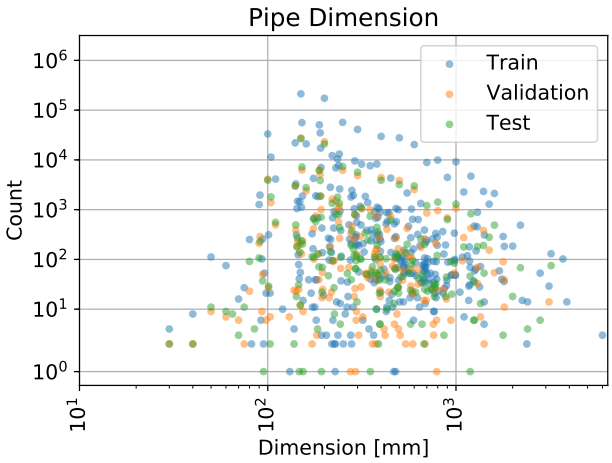


Figure 4: **Distribution of the pipe dimensions.** Plots of the occurrence frequencies of each pipe dimension, for each dataset split. Note that both axes are log-scaled.

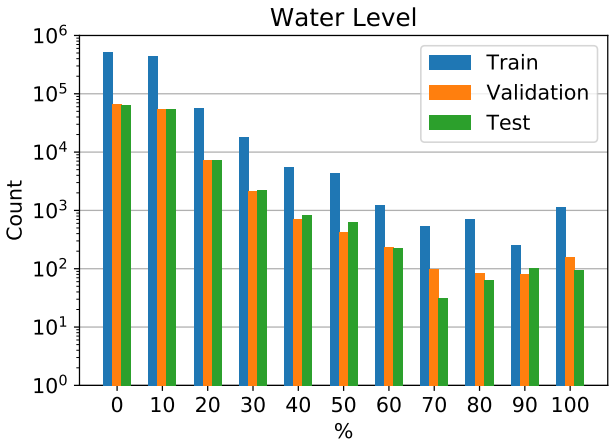


Figure 5: **Distribution of the water level.** We plot the occurrence frequencies for each of the water level classes, for each dataset split. Note that the y-axis is log-scaled.

an associated water level in the range 0–30 %, while the remaining classes occur less often, and not as evenly split between the classes. This can be explained by the fact that when the majority of a pipe is filled with water, the inspections may at times be postponed for a later time and it becomes difficult to accurately access how much water it actually contains. Furthermore, the inspection vehicle will at times be partially or fully submerged in the water, resulting in the inspector losing key reference points used for estimating the water level, such as the pipe wall.

Lastly, in Figure 6 we plot the resolution of the sewer inspection videos in each split. The resolution is denoted as width by height. It should be noted that the video resolutions reported are not the resolutions observed by the

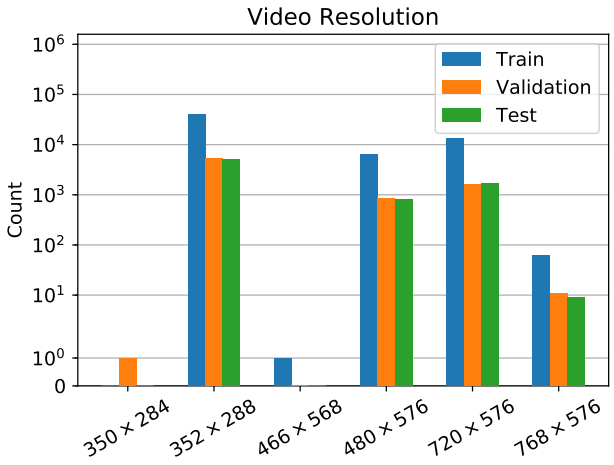


Figure 6: **Distribution of video resolution.** We present the distribution of the different resolutions for the videos in each dataset split. Note the y-axis is log-scaled.

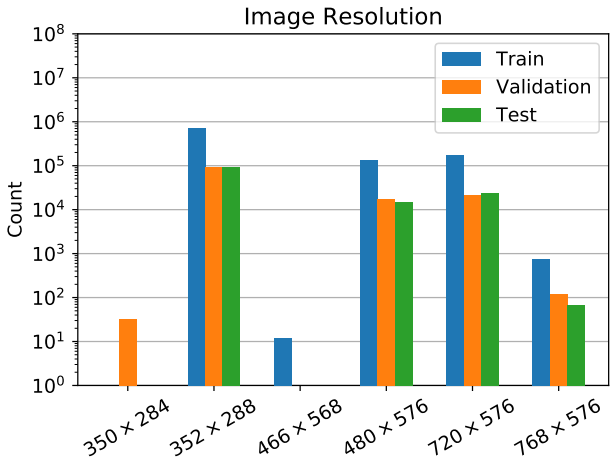


Figure 7: **Distribution of image resolution.** We present the distribution of the different resolutions for the images in each dataset split. Note the y-axis is log-scaled.

inspector. The videos are encoded in such a way that the video data is stored in the resolution reported in this work, but when presented using a media player the width is multiplied by a "sample aspect ratio". We decide not to apply this resizing, in order to not introduce artifacts in the image data. We find that across the videos in each dataset split, the resolutions are evenly distributed. This is also true when looking at the resolution for all the images in the dataset splits, see Figure 7.

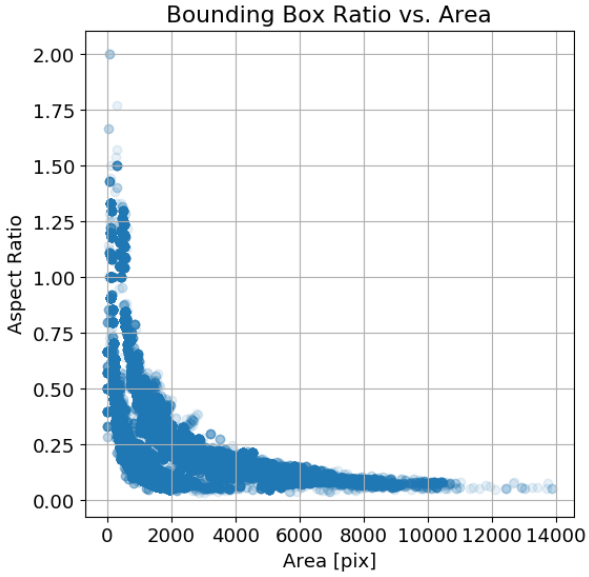


Figure 8: **Training split bounding box information.** The training split bounding box annotations are plotted with the bounding box area against the bounding box ratio.

D. Faster-RCNN Training and Metric Details

In this section we detail the hyperparameters and training settings for the Faster-RCNN [15] model we use to redact overlaid text information on the images. We also present the full COCO [11] metric suite performance, to show how well the network performs. A training split of 20,739 images and a validation split of 2,305 images are used, wherein all text information is manually annotated with bounding boxes.

Hyperarameters. The Faster-RCNN model is trained for 26 epochs with a batch size of 16 batches. An SGD optimizer with momentum is used, with a learning rate of 0.02, momentum of 0.9 and weight decay of 0.0001. The learning rate is multiplied by 0.1 at epoch 16 and 22, respectively. We employ linear warm up of the learning rate during the first 1,000 mini batches of the first epoch, increasing the learning rate from 10^{-3} to 0.02. The backbone is a ResNet-50 FPN [7, 10] pre-trained on ImageNet [17], of which we fine-tune the last three residual blocks. Custom anchor boxes are used, with a bounding box ratios (height over width) of 1:8, 1:4 and 1:2, and bounding box scales with areas of 32^2, 64^2, 128^2, 256^2, and 512^2. These values are determined based on the bounding box information in the training split, see Figure 8. All images are normalized using the ImageNet per channel mean and standard deviation, and horizontal flipping with a 50% chance is used during training. The images are rescaled such that the shortest side is 800 pixels, while enforcing that the largest side is no larger than 1,333 pixels. The training loss and mAP[0.5:0.95] on the validation set are plotted in Figure 9.

Table 3: **Full COCO metric suite.** The performance of the trained Faster-RCNN model on the validation set, for different Average Precision (AP) and Average Recall (AR) settings.

AP, IoU:			**AP@[0.5:0.95], Area:**			**AR@[0.5:0.95], #Dets:**			**AR@[0.5:0.95], Area:**		
0.5:0.95	0.5	0.75	S	M	L	1	10	100	S	M	L
89.10	98.89	96.39	88.08	89.96	95.63	10.06	88.31	92.25	91.72	92.71	96.28

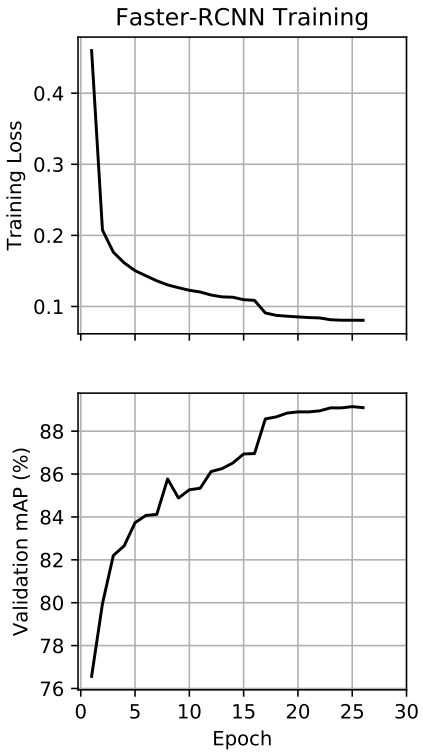


Figure 9: **Faster-RCNN loss and metric curves.** The training loss and validation metrics for the trained Faster-RCNN model. mAP@[0.5:0.95] is denoted as mAP.

Metrics. In order to determine the effect of the Faster-RCNN model, we compute the full COCO metrics suite on the validation set, as shown in Table 3. As shown in the metrics, we have a high precision and recall, though the recall indicates that not all of the text objects have been detected. This is partially due to some text information being annotated with a single bounding box but detected as several boxes, and vice versa. To verify the annotations we manually inspect a set of randomly selected samples.

E. Extra Trees Hyperparameter Grid Search

For the system proposed by Myrans *et al*. [13, 14], two Extra Trees classifiers are used in sequence. However, the hyperparameters of the trees are not specified. Therefore, we conduct a small grid search across three hyperparameters: The amount of trees in the ensemble, the maximum depth of the trees, and the maximum amount of features used when splitting an internal node. The investigated parameters are reported in Table 4. We train the Extra Trees classifier in three settings. First, we train under a binary setting determining whether there is *any* class in the image. Thereafter, we train a multi-label setting, first on a subset of the dataset only containing images with annotated classes, and secondly on the full dataset. The resulting validation losses of the hyperparameter search is shown in Figure 10. From this we conclude that for the binary Extra Trees classifier 100 trees, with a maximum depth of 10 and using $\log_2(d)$ features when splitting, should be used. Similarly, we find that for the multi-label Extra Trees classifiers 250 trees, with a maximum depth of 10 and using $\log_2(d)$ features when splitting, should be used.

Table 4: **Extra Trees grid search intervals.** Hyperparameter search intervals for the Extra Trees classifiers. d denotes the dimensionality of the GIST descriptor.

Parameter	**Values**
Number of Trees	[10, 100, 250]
Max Depth	[10, 20, 30]
Max Features	$[\sqrt{d}, \log_2(d), d/3]$

F. CNN Loss Curves

We present the loss curves for all the tested convolutional neural networks (CNNs) tested, see Figure 11. All networks are trained using the weighted binary cross-entropy loss, and using hyperparameters set based on the guidelines from Goyal *et al*. [5]. Further training details are presented in the main manuscript.

From the loss plots we observe that the validation loss of the majority of the tested networks start diverging after approximately 30-40 epochs, a clear sign of overfitting. The method by Xie *et al*. [19] is an exception, with the first and second stage methods stagnating after 60–70 epochs. We also observe that the first stage of Chen *et al*. [1], the SqueezeNet [8], has a constant loss value for both the training and validation loss. Similarly, the second stage of Xie *et al*. settles on a constant loss after the initial 10 epochs when trained on the full dataset.

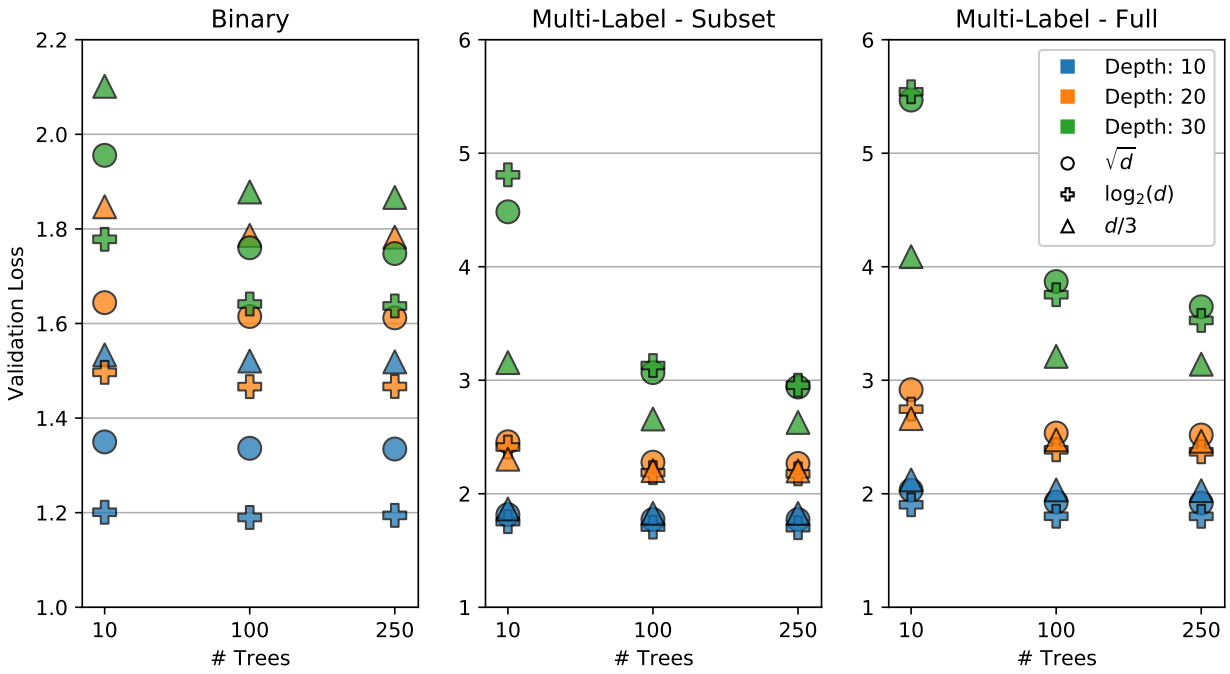


Figure 10: **Extra Trees grid search results.** Results of the grid search of the Extra Trees classifiers for: Binary classifier trained on full dataset, multi-label classifier trained on a subset of the dataset, and multi-label classifier trained on the full dataset.

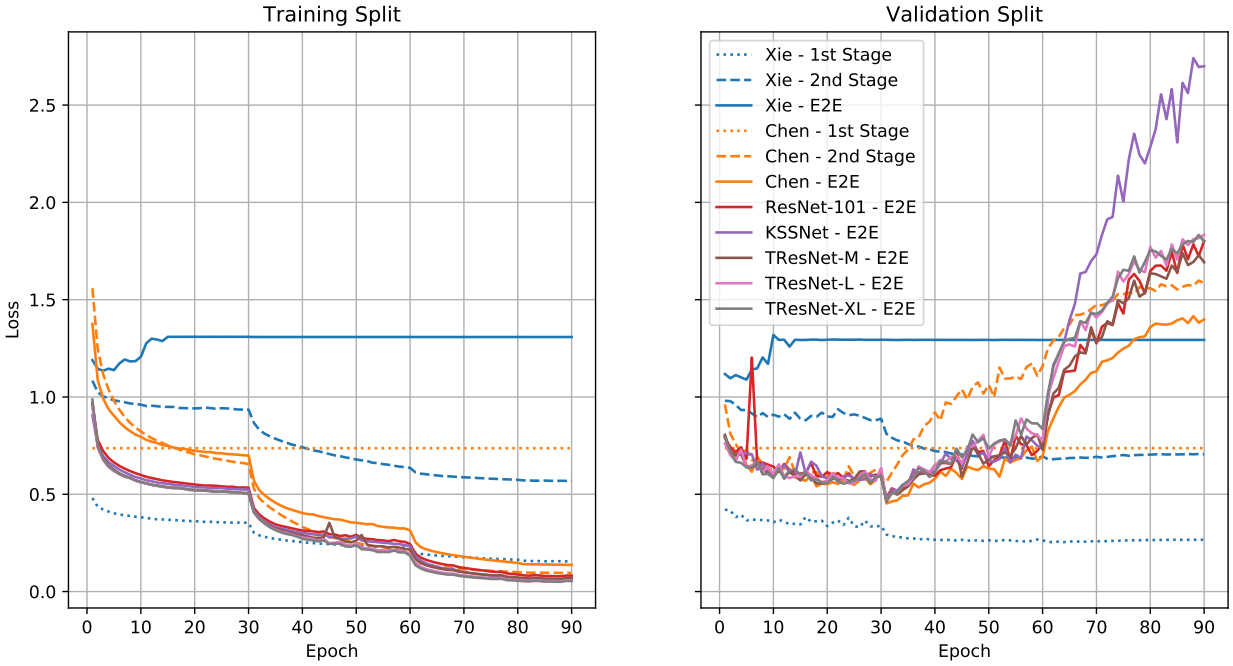


Figure 11: **Multi-label CNN loss curves.** The training and validation loss curves for all tested networks. "1st Stage" indicates a binary classifier, "2nd Stage" indicates a multi-label classifier trained on a subset of the dataset, and "E2E" indicates a multi-label classifier trained in an end-to-end manner with the full dataset.

Table 5: **Effect of binary stage in two-stage classifiers.** We present the metric performance for the two-stage methods, comparing the effect of the full pipeline and using only the multi-label classifier. TS denotes that both stages are used, otherwise only the second stage is used.

Model	TS	Validation		Test	
		$F2_{CIW}$ ↑	$F1_{Normal}$ ↑	$F2_{CIW}$ ↑	$F1_{Normal}$ ↑
Xie [19]	✓	**48.57**	**91.08**	**48.34**	**90.62**
		37.65	0.52	37.83	0.68
Chen [1]	✓	42.03	3.96	41.74	3.59
		42.03	3.96	41.74	3.59
Myrans [13]	✓	4.01	26.03	4.11	27.48
		19.25	0.00	19.19	0.00

Table 6: **Effect of training second stage on full dataset.** The metric performance for the two-stage methods, when training both stages on the full dataset. TS denotes that both stages are used, otherwise only the second stage is used.

Model	TS	Validation		Test	
		$F2_{CIW}$ ↑	$F1_{Normal}$ ↑	$F2_{CIW}$ ↑	$F1_{Normal}$ ↑
Xie [19]	✓	31.98	**88.23**	31.82	**87.95**
		28.12	59.98	27.96	59.99
Chen [1]	✓	**43.45**	76.73	**43.14**	75.68
		43.45	76.73	**43.14**	75.68
Myrans [13]	✓	2.58	25.98	2.61	27.48
		7.48	0.00	7.37	0.00

G. Two-Stage Ablation Study

We conduct two ablation studies on the two-stage classifiers, to determine the effect of the different stages and training methodology.

What is the effect of the binary classifier? We compare the effect on performance of using both stages or only the second stage. These results are presented in Table 5, and indicate that the first stage is crucial. Performance for Xie *et al*. [19] degrades for both metrics when the first stage is missing, whereas for Chen *et al*. [1] there is no difference as the first stage never predicts a normal pipe. For Myrans *et al*. [13] the first stage inaccurately classifies images with classes as normal pipes, causing a lower $F2_{CIW}$ score. This is improved when using only the second stage, but at the cost of an inability to recognize any normal pipes.

Training the second stage on the full dataset. Classically within the sewer classification domain, the second stage is only trained on data which contains some kind of class. We investigate whether performance improves by training on the full dataset, such that the second stage also sees normal pipes. The results are shown in Table 6. For Myrans *et al*. the performance is reduced substantially in both tested settings, and the second stage is still unable to classify normal pipes. For Xie *et al*. both metrics are lower when comparing to Table 5, except for the large increase in F_{Normal} score when only using the second stage. The only performance improvement is achieved by Chen *et al*. through the use of the deeper InceptionV3 network.

H. Multi-Label Metrics and Results

When evaluating multi-label tasks, a large suite of metrics are commonly used, in order to uncover different aspects of the tested methods. Commonly, the F1-score is used in different variations, depending on how the F1-score is calculated or averaged. An overview of the different metrics is provided in Table 8. Each of the metrics are in the range $[0, 1]$, and for all a high score is better. As a reference on how to compute the metrics, we refer to the supplementary materials of the work by Durand *et al*. [4]. We present the classic performance metrics for each of the tested methods on both the validation and test splits, as well as the per-class F1, F2, Recall, Precision, and Average Precision (AP). It should be noted, that AP cannot be calculated for the normal class. This is due to the normal class being an implicit class, and therefore not possible to rank as it does not have a single associated probability. The Kumar *et al*. [9], Meijer *et al*. [12] and ML-GCN [2] methods are not shown in the metric tables as the models diverged during training. The benchmark algorithm consisting of the first stage from Xie *et al*. [19] and the TResNet-L multi-label classifier [16] is reported as "Benchmark". The metrics for the validation split are presented in Table 9-14 and the metrics for the test split are presented in Table 15-20.

Figure 12: **Sewer-ML data examples.** A subset of the images in the Sewer-ML showcasing five images from each of the annotated classes as well as normal pipes in each row. The class codes are described in Table 1.

Continued on next page

Figure 12: **Continued from previous page**

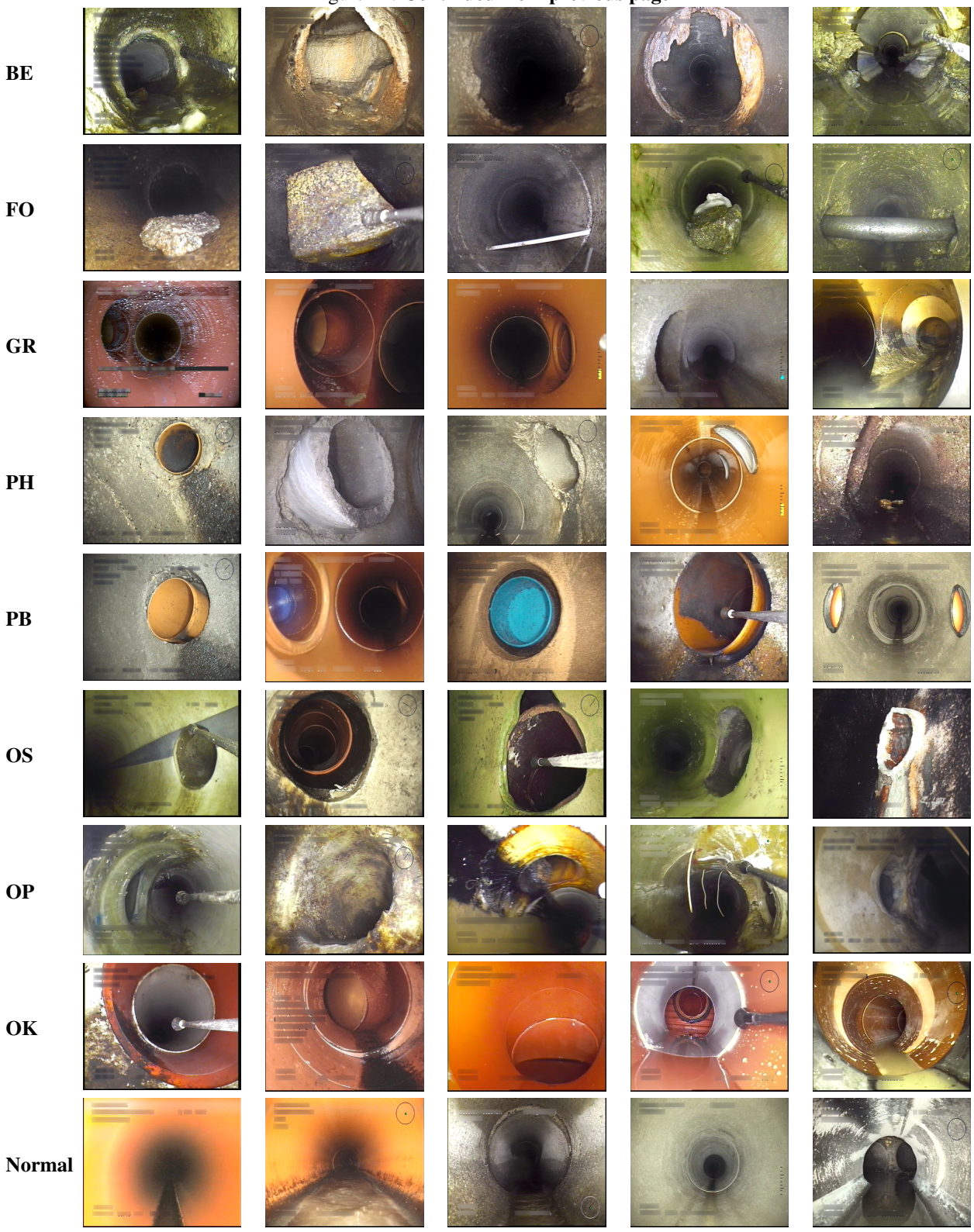


Table 8: **Multi-label classification metrics.** A short description of the commonly used multi-label classification metrics. For details on how the metrics are computed, we refer to Durand *et al*. [4].

Metric	Description
Macro-F1 (M-F1)	Average F1-score across all classes.
Micro-F1 (m-F1)	F1 score calculated over all samples.
Overall Precision (OV-P)	Precision metric calculated over all samples, regardless of class.
Overall Recall (OV-R)	Recall metric calculated over all samples, regardless of class.
Overall F1 (OV-F1)	F1 score calculated using OV-P and OV-R.
Per-class Precision (PC-P)	Average precision metric across all classes.
Per-class Recall (PC-R)	Average recall metric across all classes.
Per-class F1 (PC-F1)	F1-score calculated using PC-P and PC-R.
Zero-one Exact Match Accuracy (0-1)	Ratio of samples with all labels correctly predicted.
mean Average Precision (mAP)	Average of the Average Precision of all annotated classes

Table 9: **Performance metrics for each method - Validation Split.** The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	m-F1	M-F1	OV-F1	OV-P	OV-R	PC-F1	PC-P	PC-R	0-1	mAP
Sewer	Xie *et al*. [19]	59.33	38.10	59.33	46.31	82.52	42.43	29.61	74.79	51.64	66.40
	Chen *et al*. [1]	33.94	26.62	33.94	26.38	47.60	35.09	23.97	65.40	7.96	62.06
	Hassan *et al*. [6]	12.76	6.36	12.76	7.44	44.86	6.92	3.72	50.00	0.00	8.89
	Myrans *et al*. [13]	5.39	3.69	5.39	3.19	17.27	4.80	2.90	14.06	13.66	0.54
General	ResNet-101 [7]	54.47	38.08	54.47	40.63	82.62	43.58	28.98	87.83	39.96	76.27
	KSSNet [18]	56.18	39.37	56.18	42.52	82.77	44.82	30.12	87.56	41.28	77.63
	TResNet-M [16]	55.27	38.69	55.27	41.22	83.88	44.14	29.35	**88.93**	41.07	78.29
	TResNet-L [16]	56.01	39.63	56.01	42.09	83.69	44.90	30.10	88.32	41.22	78.75
	TResNet-XL [16]	55.83	39.30	55.83	41.82	83.98	44.66	29.85	88.67	41.68	78.32
	Benchmark	**61.45**	**42.39**	**61.45**	**47.02**	**88.67**	**46.38**	**32.25**	82.55	**51.65**	**79.79**

Table 10: **Per-class F1 score - Validation Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al*. [19]	22.97	71.40	35.83	22.84	77.12	9.77	17.94	28.39	40.27	38.09	**5.84**	49.03	37.86	24.21	13.80	29.03	70.29	91.08
	Chen *et al*. [1]	24.60	57.19	17.17	10.08	68.37	6.03	**31.78**	21.05	26.71	39.01	5.33	25.26	34.27	8.79	10.10	32.34	57.18	3.96
	Hassan *et al*. [6]	0.00	30.75	3.06	3.09	43.57	1.35	4.39	0.00	13.02	0.00	0.00	10.06	5.14	0.00	0.00	0.00	0.00	0.00
	Myrans *et al*. [13]	1.61	5.70	1.79	1.07	8.07	0.28	0.53	0.84	3.29	2.56	0.20	3.07	0.94	0.94	0.28	0.14	9.16	26.03
General	ResNet-101 [7]	24.08	73.10	30.46	18.47	79.44	9.48	20.43	27.80	39.92	41.07	4.50	47.41	40.62	24.66	18.08	32.83	73.52	79.55
	KSSNet [18]	**25.40**	73.64	31.52	18.84	80.59	10.56	21.06	28.88	41.11	42.15	4.82	51.24	43.69	25.23	19.28	35.55	74.58	80.60
	TResNet-M [16]	24.87	72.90	30.24	19.72	80.13	10.47	20.16	28.31	40.61	40.32	4.54	48.04	44.08	24.17	17.26	35.64	73.80	81.23
	TResNet-L [16]	24.51	73.14	30.87	19.74	79.94	11.57	19.76	29.49	41.24	41.49	4.74	50.31	47.15	28.00	17.88	38.95	73.34	81.22
	TResNet-XL [16]	24.75	73.15	32.20	20.58	79.89	10.21	19.76	29.09	40.30	41.15	4.69	48.35	45.22	26.45	18.23	37.08	74.57	81.81
	Benchmark	24.81	**74.50**	**36.39**	**23.91**	**80.69**	**11.87**	20.05	**31.19**	**42.39**	**42.63**	4.94	**51.40**	**48.53**	**33.09**	**21.58**	**48.75**	**75.00**	**91.32**

Table 11: **Per-class F2 score - Validation Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al*. [19]	38.87	77.86	52.92	37.48	79.78	19.96	32.90	46.59	53.52	53.48	**12.66**	58.74	56.49	40.44	26.37	47.51	76.98	90.32
	Chen *et al*. [1]	38.82	66.38	31.00	20.81	73.90	12.77	**43.33**	33.70	42.02	45.98	11.64	43.15	52.73	19.18	20.63	51.75	66.21	2.52
	Hassan *et al*. [6]	0.00	52.60	7.32	7.37	65.87	3.30	10.29	0.00	27.24	0.00	0.00	21.86	11.94	0.00	0.00	0.00	0.00	0.0
	Myrans *et al*. [13]	3.23	7.81	4.06	1.84	9.59	0.66	1.16	1.85	5.95	4.78	0.49	5.91	2.03	2.26	0.69	0.33	13.32	25.93
General	ResNet-101 [7]	42.45	84.34	51.08	35.34	87.49	19.98	37.81	47.47	59.18	59.87	10.39	64.78	61.24	44.03	34.81	54.23	82.99	71.60
	KSSNet [18]	**43.74**	**84.64**	52.24	35.92	87.45	21.76	38.67	48.54	59.89	**60.81**	11.08	**67.40**	63.94	44.73	36.60	57.05	83.30	72.95
	TResNet-M [16]	43.55	84.49	51.02	37.23	87.79	21.75	37.57	47.93	60.01	59.99	10.51	65.61	64.43	43.62	33.70	57.04	**83.71**	73.71
	TResNet-L [16]	43.08	84.39	51.75	37.39	**87.81**	23.50	37.03	49.20	**60.10**	60.60	10.91	67.05	**66.64**	48.24	34.53	60.12	83.59	73.69
	TResNet-XL [16]	43.34	84.44	52.99	38.50	87.69	21.37	37.01	48.93	59.79	60.42	10.79	65.77	65.07	46.46	35.12	58.61	83.63	74.49
	Benchmark	42.92	83.56	**54.06**	**39.16**	86.99	**23.89**	37.17	**50.41**	59.64	60.12	11.30	66.39	66.40	**50.16**	**37.32**	**64.86**	82.88	**90.79**

Table 12: **Per-class Precision score - Validation Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al.* [19]	13.66	62.73	23.30	13.83	**73.05**	5.28	10.21	17.20	28.51	25.75	**3.08**	**38.45**	24.43	14.51	7.69	17.61	61.39	92.36
	Chen *et al.* [1]	**15.27**	46.48	9.85	5.42	60.78	3.20	**22.00**	12.94	16.62	**31.15**	2.80	14.93	21.64	4.62	5.46	19.90	46.59	91.88
	Hassan *et al.* [6]	0.00	18.17	1.55	1.57	27.85	0.68	2.24	0.00	6.97	0.00	0.00	5.30	2.64	0.00	0.00	0.00	0.00	0.00
	Myrans *et al.* [13]	0.88	3.93	0.93	0.63	6.38	0.14	0.28	0.44	1.88	1.44	0.10	1.71	0.49	0.47	0.14	0.07	6.03	26.21
General	ResNet-101 [7]	13.99	59.82	18.21	10.29	68.89	5.05	11.56	16.44	25.88	26.96	2.32	32.76	26.02	14.23	10.04	19.80	61.78	97.61
	KSSNet [18]	14.95	60.52	18.98	10.51	71.26	5.69	11.97	17.24	26.99	27.89	2.48	36.61	28.59	14.61	10.78	21.84	63.50	97.68
	TResNet-M [16]	14.50	59.34	18.01	11.06	69.95	5.62	11.37	16.83	26.39	26.07	2.33	33.21	28.88	13.86	9.52	21.93	61.64	97.87
	TResNet-L [16]	14.26	59.84	18.46	11.05	69.55	6.27	11.12	17.69	27.08	27.19	2.44	35.53	31.70	16.48	9.91	24.54	60.88	**97.89**
	TResNet-XL [16]	14.43	59.81	19.47	11.59	69.58	5.46	11.12	17.36	26.12	26.87	2.41	33.55	29.98	15.40	10.12	23.00	63.16	97.86
	Benchmark	14.56	**63.11**	**23.56**	**14.50**	72.00	**6.46**	11.35	**19.07**	**28.60**	28.71	2.55	37.34	**33.50**	**21.11**	**12.67**	**34.49**	**64.74**	92.21

Table 13: **Per-class Recall score - Validation Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al.* [19]	72.16	82.86	77.59	65.46	81.67	65.49	74.08	81.33	68.54	73.19	57.29	67.67	84.06	73.07	67.18	82.52	82.20	89.83
	Chen *et al.* [1]	63.16	74.34	66.95	71.59	78.12	50.28	57.18	56.26	68.01	52.19	54.94	81.80	82.28	90.20	67.61	86.27	74.01	2.03
	Hassan *et al.* [6]	0.00	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.00	**100.00**	0.00	0.00	**100.00**	**100.00**	0.00	0.00	0.00	0.00	0.00
	Myrans *et al.* [13]	9.84	10.37	26.37	3.58	10.98	9.88	5.90	9.67	12.94	11.33	9.05	15.42	8.97	38.56	19.47	5.72	19.09	25.86
General	ResNet-101 [7]	86.38	93.96	93.07	90.28	93.82	76.50	87.42	**89.83**	87.25	86.16	80.90	85.73	92.57	92.42	90.81	**95.92**	90.78	67.13
	KSSNet [18]	84.36	94.01	92.97	90.73	92.72	74.23	87.38	88.90	86.12	86.25	82.41	85.34	92.54	92.29	91.25	95.59	90.34	68.61
	TResNet-M [16]	**87.25**	94.51	94.16	91.22	93.77	77.07	88.58	89.12	88.07	**88.93**	**85.26**	86.76	93.07	**94.12**	**92.34**	95.10	91.95	69.42
	TResNet-L [16]	87.04	94.04	94.26	92.59	93.98	75.14	88.79	88.73	86.46	87.45	83.08	86.17	91.99	93.07	91.03	94.28	**92.19**	69.39
	TResNet-XL [16]	86.85	94.13	93.02	91.81	93.80	78.89	88.58	89.69	88.22	87.84	82.24	86.56	91.99	93.73	91.90	95.59	91.00	70.29
	Benchmark	83.66	90.93	79.91	68.11	91.76	73.55	86.25	85.56	81.84	82.75	79.23	82.42	88.02	76.47	72.65	83.17	89.12	**90.44**

Table 14: **Per-class AP score - Validation Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK
Sewer	Xie *et al.* [19]	29.82	83.81	83.82	63.25	87.48	31.80	62.03	54.60	60.44	66.43	48.39	84.66	78.56	68.85	61.24	77.41	86.25
	Chen *et al.* [1]	48.30	73.98	49.98	45.03	83.03	56.31	78.12	45.85	61.99	71.01	42.94	74.62	78.27	56.95	51.00	56.89	80.80
	Hassan *et al.* [6]	7.20	32.00	1.09	0.26	37.38	1.26	5.90	6.25	7.71	11.05	1.44	10.27	4.34	0.00	0.00	5.85	19.14
	Myrans *et al.* [13]	0.05	0.62	0.10	0.64	2.57	0.13	0.12	0.19	1.02	0.28	0.00	0.35	0.00	1.17	0.00	0.00	1.96
General	ResNet-101 [7]	54.54	90.21	84.15	76.73	93.56	49.33	81.88	67.13	74.85	80.20	64.11	90.83	87.63	66.49	59.77	81.58	93.57
	KSSNet [18]	56.86	90.74	85.42	76.50	94.05	56.75	**83.43**	68.86	75.14	81.40	**65.51**	91.27	89.20	66.49	64.58	79.66	93.87
	TResNet-M [16]	**57.22**	90.90	87.74	77.69	93.98	58.68	80.56	**69.94**	**76.17**	**82.39**	60.67	91.55	**89.90**	69.27	65.58	84.39	94.32
	TResNet-L [16]	56.76	90.75	88.32	78.36	93.95	60.42	80.88	69.21	75.64	81.99	64.62	91.37	89.38	69.75	69.39	83.57	94.44
	TResNet-XL [16]	57.15	90.81	87.36	78.34	94.04	56.91	80.92	69.84	76.01	82.00	63.28	91.69	88.97	69.66	68.16	82.09	94.23
	Benchmark	56.68	**90.93**	**90.12**	**80.30**	**94.06**	**60.55**	80.79	69.45	75.99	82.27	65.32	**92.06**	89.89	**75.70**	**72.97**	**84.81**	**94.57**

Table 15: **Performance metrics for each method - Test Split.** The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	m-F1	M-F1	OV-F1	OV-P	OV-R	PC-F1	PC-P	PC-R	0-1	mAP
Sewer	Xie *et al*. [19]	59.05	37.94	59.05	46.06	82.24	42.16	29.48	73.95	51.55	65.32
	Chen *et al*. [1]	33.49	26.23	33.49	26.03	46.94	34.55	23.60	64.48	7.63	59.89
	Hassan *et al*. [6]	12.57	6.27	12.57	7.33	44.12	6.83	3.67	50.00	0.00	7.35
	Myrans *et al*. [13]	5.66	3.88	5.66	3.36	18.07	5.02	3.04	14.43	14.51	0.59
General	ResNet-101 [7]	53.91	37.94	53.91	40.19	81.85	43.46	28.89	87.70	39.38	74.99
	KSSNet [18]	55.64	39.22	55.64	42.12	81.96	44.68	30.02	87.32	40.46	75.70
	TResNet-M [16]	54.62	38.53	54.62	40.72	82.94	43.96	29.24	**88.56**	40.23	76.55
	TResNet-L [16]	55.34	39.45	55.34	41.56	82.79	44.72	29.97	88.05	40.42	76.82
	TResNet-XL [16]	55.08	38.98	55.08	41.21	83.01	44.34	29.61	88.23	40.74	76.61
	Benchmark	**61.26**	**42.35**	**61.26**	**46.90**	**88.30**	**46.22**	**32.24**	81.61	**51.59**	**77.79**

Table 16: **Per-class F1 score - Test Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al*. [19]	23.47	71.12	34.63	23.51	77.25	9.99	16.14	31.33	39.59	39.93	**6.29**	49.22	34.61	24.09	14.27	27.05	69.85	90.62
	Chen *et al*. [1]	24.54	57.24	16.85	11.65	67.47	6.09	**29.99**	23.03	26.47	37.67	5.62	24.84	31.53	8.58	11.01	29.58	56.46	3.59
	Hassan *et al*. [6]	0.00	30.35	2.95	3.49	43.16	1.41	4.04	0.00	13.19	0.00	0.00	9.84	4.45	0.00	0.00	0.00	0.00	0.00
	Myrans *et al*. [13]	1.63	5.60	1.71	1.46	8.18	0.36	0.60	1.02	4.03	2.93	0.21	3.16	0.82	0.89	0.31	0.16	9.28	27.48
General	ResNet-101 [7]	24.42	72.52	28.62	19.75	79.22	9.98	19.07	30.13	38.93	42.15	4.87	47.69	37.80	26.86	18.56	30.54	73.28	78.57
	KSSNet [18]	**26.06**	73.34	29.89	19.64	**80.56**	10.83	19.84	31.47	40.59	43.50	5.24	50.88	40.74	26.39	20.55	32.84	74.38	79.29
	TResNet-M [16]	24.78	72.54	28.94	20.96	79.87	10.89	18.52	31.14	39.64	41.39	4.90	47.97	40.11	24.99	18.34	34.68	73.90	79.91
	TResNet-L [16]	24.78	72.93	28.68	20.62	79.60	12.01	18.20	32.29	40.43	42.56	5.11	49.97	43.33	28.13	19.43	38.68	73.41	79.88
	TResNet-XL [16]	24.76	72.66	30.24	21.49	79.71	10.46	18.32	31.51	39.59	41.94	5.13	48.32	41.21	27.12	19.25	35.10	74.41	80.42
	Benchmark	25.11	**74.40**	**35.58**	**25.01**	80.50	**12.26**	18.59	**34.26**	**41.93**	**44.16**	5.26	**51.28**	**45.09**	**31.60**	**22.20**	**49.17**	**75.04**	**90.94**

Table 17: **Per-class F2 score - Test Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al*. [19]	39.77	77.88	51.49	37.53	79.90	20.42	30.19	49.28	53.16	54.67	**13.50**	59.40	53.74	39.57	26.30	45.24	76.63	89.77
	Chen *et al*. [1]	38.91	66.52	29.90	23.33	73.08	12.94	**42.11**	35.50	41.72	44.01	12.11	42.63	50.53	18.73	21.87	48.37	65.92	2.28
	Hassan *et al*. [6]	0.00	52.14	7.07	8.28	65.50	3.45	9.53	0.00	27.53	0.00	0.00	21.43	10.44	0.00	0.00	0.00	0.00	0.0
	Myrans *et al*. [13]	3.27	7.70	3.89	2.47	9.75	0.87	1.33	2.22	7.27	5.35	0.51	6.11	1.80	2.15	0.77	0.38	13.51	27.35
General	ResNet-101 [7]	42.95	83.92	48.25	37.06	87.22	21.11	35.87	50.06	58.23	60.37	11.18	64.81	58.83	47.12	35.24	51.76	82.63	70.41
	KSSNet [18]	**44.79**	**84.52**	49.68	36.93	87.36	22.42	36.96	51.45	59.54	**61.55**	11.96	**66.99**	61.57	46.35	**37.92**	54.59	83.00	71.32
	TResNet-M [16]	43.39	84.36	48.99	38.84	**87.48**	22.70	35.12	51.09	59.13	60.27	11.27	65.41	60.99	44.71	35.04	56.55	**83.72**	72.08
	TResNet-L [16]	43.50	84.42	48.55	38.39	87.45	24.45	34.67	52.27	**59.55**	61.13	11.70	66.37	63.54	**48.58**	36.69	60.42	83.50	72.03
	TResNet-XL [16]	43.39	84.22	50.14	39.42	87.43	22.00	34.89	51.39	59.15	60.64	11.74	65.47	61.83	47.45	36.31	56.76	83.52	72.74
	Benchmark	43.35	83.82	**52.94**	**39.69**	86.76	**24.70**	34.96	**53.41**	59.45	61.05	11.94	66.05	**64.00**	47.39	36.79	**65.41**	82.72	**90.35**

Table 18: **Per-class Precision score - Test Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al*. [19]	13.95	62.14	22.40	14.49	**73.20**	5.40	9.09	19.50	27.77	27.54	**3.33**	**38.28**	21.72	14.58	8.10	16.20	60.87	92.09
	Chen *et al*. [1]	15.19	46.43	9.75	6.35	59.82	3.24	**20.27**	14.52	16.45	**30.38**	2.97	14.65	19.38	4.51	6.02	17.96	45.57	91.35
	Hassan *et al*. [6]	0.00	17.89	1.50	1.77	27.52	0.71	2.06	0.00	7.06	0.00	0.00	5.17	2.28	0.00	0.00	0.00	0.00	0.00
	Myrans *et al*. [13]	0.89	3.86	0.88	0.87	6.46	0.18	0.31	0.54	2.32	1.67	0.11	1.75	0.43	0.45	0.16	0.08	6.09	27.70
General	ResNet-101 [7]	14.20	59.14	17.06	11.10	68.72	5.31	10.71	18.11	25.08	28.04	2.51	33.12	23.69	15.65	10.37	18.14	61.65	97.37
	KSSNet [18]	**15.36**	60.10	17.96	11.03	71.31	5.82	11.20	19.11	26.52	29.22	2.71	36.33	26.05	15.36	11.66	19.73	63.39	97.46
	TResNet-M [16]	14.45	58.81	17.21	11.86	69.76	5.83	10.36	18.86	25.58	27.19	2.52	33.21	25.54	14.40	10.23	21.09	61.81	97.59
	TResNet-L [16]	14.43	59.44	17.05	11.64	69.23	6.50	10.15	19.73	26.33	28.26	2.64	35.40	28.32	16.53	10.89	24.18	61.10	**97.62**
	TResNet-XL [16]	14.43	59.12	18.20	12.22	69.48	5.58	10.23	19.16	25.52	27.70	2.65	33.64	26.48	15.83	10.80	21.45	62.96	97.58
	Benchmark	14.76	**62.66**	**23.01**	**15.47**	71.87	**6.66**	10.44	**21.44**	**28.11**	30.22	2.72	37.36	**30.22**	**20.32**	**13.37**	**34.79**	**64.99**	91.94

Table 19: **Per-class Recall score - Test Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK	Normal
Sewer	Xie *et al*. [19]	74.02	83.15	76.24	62.29	81.77	66.99	71.98	79.72	68.91	72.53	57.16	68.90	85.11	69.27	60.00	81.99	81.93	89.21
	Chen *et al*. [1]	63.81	74.59	61.83	70.35	77.37	51.41	57.64	55.55	67.74	49.56	52.39	81.56	84.47	88.48	63.96	83.86	74.21	1.83
	Hassan *et al*. [6]	0.00	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.00	**100.00**	0.00	0.00	**100.00**	**100.00**	0.00	0.00	0.00	0.00	0.00
	Myrans *et al*. [13]	9.96	10.26	26.37	4.55	11.17	12.34	7.19	10.26	15.62	11.87	8.63	16.13	8.98	33.61	18.49	7.69	19.42	27.26
General	ResNet-101 [7]	86.91	93.73	88.92	89.16	93.52	82.47	86.85	**89.55**	86.96	84.82	81.66	85.19	93.52	94.72	87.92	96.44	90.31	65.85
	KSSNet [18]	85.98	94.07	88.92	89.34	92.58	78.25	87.07	89.21	86.46	85.09	81.66	84.91	93.42	93.52	86.79	**97.75**	89.96	66.83
	TResNet-M [16]	86.95	94.64	91.02	90.03	93.42	82.03	87.22	89.21	87.95	**86.62**	**84.75**	86.34	93.38	94.36	89.06	97.56	91.85	67.66
	TResNet-L [16]	**87.64**	94.34	90.20	90.16	93.61	79.00	87.48	88.93	86.99	86.19	82.74	84.94	92.20	94.24	**90.00**	96.62	**91.93**	67.60
	TResNet-XL [16]	87.04	94.22	89.33	88.86	93.47	83.23	87.89	88.72	88.22	86.27	83.05	85.74	92.81	**94.84**	88.68	96.44	90.94	68.39
	Benchmark	84.04	91.54	78.45	65.19	91.50	76.52	84.72	85.16	82.42	81.95	77.81	81.74	88.83	71.07	65.47	83.86	88.78	**89.96**

Table 20: **Per-class AP score - Test Split**. The metrics are presented as percentages, and the highest score in each column is denoted in bold.

	Model	RB	OB	PF	DE	FS	IS	RO	IN	AF	BE	FO	GR	PH	PB	OS	OP	OK
Sewer	Xie *et al*. [19]	35.07	83.48	85.86	62.97	87.30	36.64	59.04	58.52	58.15	62.44	36.26	82.11	80.02	65.49	53.70	77.47	85.85
	Chen *et al*. [1]	48.49	74.08	48.09	47.43	81.76	57.95	74.70	48.77	60.86	65.30	31.56	74.91	80.58	43.37	49.87	49.86	80.61
	Hassan *et al*. [6]	6.16	26.79	0.33	3.60	35.10	1.17	1.25	3.16	5.62	5.14	0.44	9.23	4.17	0.00	1.57	2.52	18.69
	Myrans *et al*. [13]	0.11	0.69	0.00	0.63	2.91	0.00	0.09	0.16	1.80	0.43	0.18	0.69	0.00	0.00	0.00	0.24	2.09
General	ResNet-101 [7]	55.25	90.20	88.04	71.96	93.32	65.63	78.98	65.69	71.40	77.78	47.33	90.72	88.05	65.34	52.55	79.33	93.32
	KSSNet [18]	**58.43**	**90.59**	86.80	71.99	93.68	69.97	**80.75**	68.28	72.57	78.92	44.03	91.11	88.46	62.31	55.86	79.26	93.83
	TResNet-M [16]	55.61	90.16	89.82	**76.00**	93.65	65.85	78.58	69.71	73.45	79.82	50.44	91.03	**89.41**	67.57	57.79	78.11	94.32
	TResNet-L [16]	56.95	90.38	89.28	75.03	93.64	68.61	80.04	70.09	73.59	79.43	**48.74**	91.36	88.77	67.29	59.38	79.00	94.40
	TResNet-XL [16]	56.64	90.00	89.17	75.66	93.68	63.87	79.70	68.90	73.62	79.80	48.14	91.33	88.83	69.51	59.38	79.96	94.17
	Benchmark	56.99	90.46	**89.89**	75.09	**93.70**	**70.74**	80.20	**70.81**	**73.99**	**79.96**	48.21	**92.21**	89.08	**72.70**	**62.57**	**81.33**	**94.55**

References

[1] Kefan Chen, Hong Hu, Chaozhan Chen, Long Chen, and Caiying He. An intelligent sewer defect detection method based on convolutional neural network. In *2018 IEEE International Conference on Information and Automation (ICIA)*, pages 1301–1306, Aug 2018. 5, 7, 10, 11, 12, 13

[2] Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo. Multi-label image recognition with graph convolutional networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5172–5181, 2019. 7

[3] Dansk Vand og Spildevandsforening (DANVA). *Fotomanualen: TV-inspektion af afløbsledninger*. Dansk Vand og Spildevandsforening (DANVA), 6 edition, 2010. 1

[4] Thibaut Durand, Nazanin Mehrasa, and Greg Mori. Learning a deep convnet for multi-label classification with partial labels. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 647–657, 2019. 7, 10

[5] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv:1706.02677*, 2017. 5

[6] Syed Ibrahim Hassan, L. Minh Dang, Irfan Mehmood, Suhyeon Im, Changho Choi, Jaemo Kang, Young-Soo Park, and Hyeonjoon Moon. Underground sewer pipe condition assessment based on convolutional neural networks. *Automation in Construction*, 106:102849, 2019. 10, 11, 12, 13

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Jun 2016. 4, 10, 11, 12, 13

[8] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size. *arXiv:1602.07360*, 2016. 5

[9] Srinath S. Kumar, Dulcy M. Abraham, Mohammad R. Jahanshahi, Tom Iseley, and Justin Starr. Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks. *Automation in Construction*, 91:273 – 283, 2018. 7

[10] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 936–944, 2017. 4

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing. 4

[12] Dirk Meijer, Lisa Scholten, Francois Clemens, and Arno Knobbe. A defect classification methodology for sewer image sets with convolutional neural networks. *Automation in Construction*, 104:281 – 298, 2019. 7

[13] Joshua Myrans, Richard Everson, and Zoran Kapelan. Automated detection of fault types in cctv sewer surveys. *Journal of Hydroinformatics*, 21(1):153–163, Oct 2018. 5, 7, 10, 11, 12, 13

[14] Joshua Myrans, Richard Everson, and Zoran Kapelan. Automated detection of faults in sewers using CCTV image sequences. *Automation in Construction*, 95:64–71, Nov. 2018. 5

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc., 2015. 4

[16] Tal Ridnik, Hussam Lawen, Asaf Noy, and Itamar Friedman. Tresnet: High performance gpu-dedicated architecture. *arXiv:2003.13630*, 2020. 7, 10, 11, 12, 13

[17] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 4

[18] Ya Wang, Dongliang He, Fu Li, Xiang Long, Zhichao Zhou, Jinwen Ma, and Shilei Wen. Multi-label classification with label graph superimposing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):12265–12272, Apr. 2020. 10, 11, 12, 13

[19] Qian Xie, Dawei Li, Jinxuan Xu, Zhenghao Yu, and Jun Wang. Automatic detection and classification of sewer defects via hierarchical deep learning. *IEEE Transactions on Automation Science and Engineering*, pages 1–12, 2019. 5, 7, 10, 11, 12, 13