Vision-Based Corrosion Identification Using Data-Driven Semantic Segmentation Techniques

Rui Pimentel de Figueiredo
Department of materials and production
Aalborg University
Aalborg, Denmark
rmhpdf@mp.aau.dk
https://orcid.org/0000-0003-2443-5669

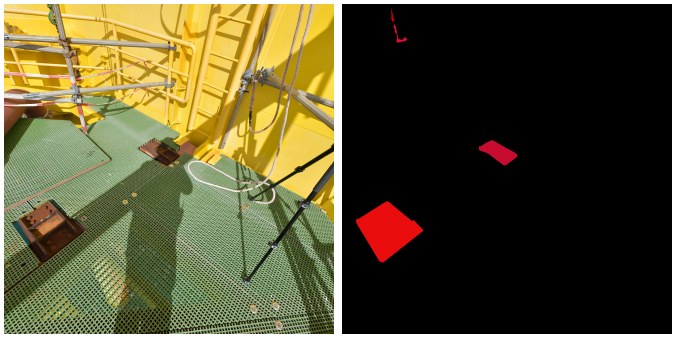
Simon Bøgh
Department of materials and production
Aalborg University
Aalborg, Denmark
sb@mp.aau.dk
https://orcid.org/0000-0002-5960-4365

***Abstract*—Corrosion is a natural process that degrades metal-made materials. Its detection is of primordial importance for quality control and for ensuring longevity of metal-made objects in various contexts, in particular in industrial environments. Different techniques for corrosion identification including ultrasonic testing, radio-graphic testing, and magnetic flux leakage have been proposed in the past. However, these require the use of costly and heavy equipment onsite for successful data acquisition. An under-explored alternative is to deploy conventional lightweight and inexpensive camera systems and computer vision based methods to tackle the former problem. In this work we present a detailed benchmark of four state-of-the-art supervised semantic segmentation techniques, for vision-based pixel-level corrosion identification. We focus our study on four, recently proposed deep learning architectures which have surpassed human-level accuracy on various visual tasks. The results demonstrate that the former approaches may be used for the problem of segmenting highly irregular patterns in industrial settings, such as corrosion, with high accuracy rates.**

***Keywords*—Machine Vision; Semantic Segmentation; Corrosion Identification**

I. Introduction

Corrosion is a natural process that degrades metal-made materials. Its detection is of the utmost importance for ensuring and control the longevity and quality of numerous metal-made infrastructures existing in various contexts, namely in industrial, urban and transportation, such as gas and oil pipelines, buildings, and vehicles [BHH+22], [DSDPM23]. Many different sensing technologies [RPS+21] for corrosion identification including ultrasonic testing [OKS21], radio-graphic testing [VEEA06], magnetic flux leakage [PALT20], and acoustic-based signals [JF21], have been successfully applied in different quality control chains, e.g. in pipeline inspection. However, this processes rely on costly and heavy machinery equipment that has to be deployed and operated by highly trained human operators. Recently, the use of less expensive and lightweight conventional cameras has been investigated to tackle the former problem [YJC+21], [20119], [NZB22]. However the literature falls short on works that attempt to employ artificial intelligence methodologies to automate this task. In this work we assess the feasibility of employing supervised deep learning approaches to solve the former problem. Our experiments demonstrate that current state-of-the-art machine learning methods for semantic segmentation can be used to recognize highly irregular corrosion patterns.



(a) (b)

Fig. 1: Corrosion semantic segmentation in RGB monocular images.

The contributions of our work are the following: The rest of this article is organized as follows: first we overview the background in corrosion detection, and existing technologies and methodologies that attempt to solve this problem. Second, we review vision-based semantic segmentation techniques that have been successfully applied across different domains, namely in biomedical and robotics applications. Then we benchmark the state-of-the-art machine learning techniques for semantic segmentation in a corrosion labeled data-set, comprising images gathered in challenging offshore industrial environments. Finally, we present conclusions and promising research directions for future work.

II. Background

Vision based scene understanding plays a role of major importance in many domains, namely in robotics [GCE21], [dFMB13], manufacturing, medical imaging [SBKV+20], inspection [TSSPdF22], [lFSPDFK21], [dFlFSH+21], [DLHB22] and surveillance [GCE21], to name a few. Scene understanding tasks may belong to one of the following categories: image classification, object detection and semantic segmentation. In this work we focus on the latter, which deals

with the problem of assigning a label, representing an object class, to each pixel. In the rest of this section we revise and explain the main techniques that attempt to solve the later problem. While image classification and object detection deal with the problem of classifying images and localizing regions of interest (i.e. bounding boxes), respectively, semantic segmentation tackles the more complex problem of assigning a class label to every image pixel. Methods existing in the literature may belong to two different paradigms: model-based or data-driven. Classical computer vision model-based approaches are based on theoretically principled methods that attempt to analytically solve geometric and physical image formation aspects, data-driven ones attempt to learn statistical properties of the image, directly from visual data, through machine learning techniques. With the rise of deep learning and the availability of large publicly annotated datasets (e.g. [LMB+14b], [COR+16]), the latter have outperformed the former in increasingly complex tasks. With the invention of AlexNet [KSH17], the DCNN architecture that won the 2012 ImageNet challenge [DDS+09a], the use of DCNNs became ubiquitous among the computer vision literature. Since then, convolutional neural networks architectures have become more accurate and applicable to increasingly complex datasets and visual tasks. In the particular case of semantic segmentation tasks, multiple approaches have been proposed in the last decade. One of the first successful deep learning semantic segmentation approaches was Mask R-CNN [HGDG17]. Its architecture is conceptually simple, and consists of a CNN backbone for feature extraction followed by a region proposal network (RPN) optimized to output candidate regions of interest. Then, three parallel branches perform classification, bounding box regression, and pixel level mask predictions. Mask R-CNN and latter proposed similar architectures achieved state-of-the-art performance in multiple semantic segmentation datasets, namely on Microsoft COCO [LMB+14a]. U-net is a fully convolutional neural network [RFB15] introduced in 2015, that is based on the idea of replacing fully connected layers with upsampling layers enabling pixel-level predictions. More specifically, U-net consists of an encoder-decoder architecture without fully connected layers. The CNN encoder (contracting path) downsamples the input image to a low dimensional feature space, while the decoder (expansive path) up-samples the feature space through deconvolutional layers. U-net has been very successful in biomedical imaging applications, in particular on segmentation of tumour cells. In the remainder of this article we overview and test U-net and three other similar approaches for semantic segmentation of corrosion patterns.

III. Methodologies

In this section we overview in detail the state-of-the-art approaches used for employing image-based semantic segmentation of corrosion in metallic surfaces. In particular, we study the feasibility of employing the state-of-the-art semantic segmentation networks for corrosion detection tasks. In the remainder of this section we describe in detail the chosen semantic segmentation network architectures. Figure 2 depicts the various tested neural networks for semantic segmentation purposes.

1) DeepLab: The first version (V1) of the DeepLab semantic segmentation network [CPK+14] proposed the use of dilated (or atrous) convolutions. The main idea is to control the field-of-view of receptive fields by manipulating the sampling rate in the convolution operation. To segment objects and surrounding context at multiple scales, the authors propose a pyramidal approach that employs dilated convolutions with different sampling rates (i.e. the larger the rate, the larger the field-of-view). Then the output from the network is bi-linearly interpolated to ensure the feature maps are enlarged to the original resolution (see Fig. 2(d)). DeepLabV3 [CPK+18] proposed the use of up-sampled filters (i.e. atrous convolutions), to control the resolution at which convolutional filters operate, i.e., to balance the trade-off between localization accuracy and context awareness. Furthermore, the authors propose the use of fully connected conditional random fields (CRFs) to refine segmentation and improve localization performance. In [HJS+22] the authors propose a multi-scale aware-relation network (MANet), optimized to deal with object scale and scene variability in remote sensing applications. The network learns multi-scale feature representations via multi-scale collaborative learning (MCL) and inter-class and intra-class region refinement (IIRR) to exploit correlations between features among different scales.

2) MANet: The multi-scale aware-relation network (MANet) [HJS+22] was originally proposed to deal with high variability of scene and object scales. The authors introduce an inter-class and intra-class region refinement scheme (IIRR) to enhance the discriminability of multi-scale representations, i.e., to reduce redundancy of fused features. The scheme utilizes a refinement strategy that separately considers the inter-class and intra-class scale variation and utilizes regional high-level semantic representations to refine multi-scale predictions. In addition, they design discrepancy classifiers to augment dissimilarity of features at different scales (see Fig. 2(c)). Then, instead of learning separate classifiers over each scale feature set and combining the predictions to decrease the error at a global level, the authors enforce pixel-level collaborative learning through co-training, by encouraging the model to focus on areas misclassified with large uncertainty, thus exploiting the correlation among different scales.

3) PSPnet: PSPnet [ZSQ+17a] or pyramid parsing network is a semantic segmentation network suitable for tasks where context may improve the performance of scene parsing. Its architecture differs from the previous, by introducing a pyramid pooling module preceding the CNN contracting path.

PSPnet exploits global scene context using a spatial pyramid pooling layer, that allows CNNs to deal with variable size inputs. Traditional CNNs attempt to avoid fixed input lengths by cropping and warping the original image, introducing context loss and distortion, and hence decreased performance. Finally, ta spatial pooling layer on top of the network, converts

the feature map to a fixed size (see Fig. 2(a)). More specificaly, the pyramid parser first extracts features representing different sub-regions at different scales to capture both local and global context. Then, these are up-sampled and concatenated to form the final feature representation, which is fed into a convolution layer to get the final segmentation masks. To deal with the challenging task of detecting objects at multiple scales, the authors of [SIBS18] proposed a Feature Pyramid Network (FPN), which comprises a bottom-up and a top-down pathway. The bottom-up pathway consists of a conventional CNN encoder and the top-down one utilizes nearest neighbour up-sampling succeeded by multi-channel concatenation, spatial dropout and bi-linear interpolation to ensure the output matches the input image size.

The Pyramid Scene Parsing Network (PSPNet) [ZSQ+17b] architecture differs from the previous, by introducing a pyramid pooling module preceding the CNN contracting path.

4) U-Net: U-net [RFB15] is based on the fully convolutional network [LSD15], but altered to allow training with small image samples, while being more accurate then the former. U-net consists of a traditional convolutional network followed by two paths: The contracting path starts by applying two 3x3 unpadded convolutions, followed by rectified linear units (ReLU) and down-sampling via 2x2 max pooling operations. The expansive path up-samples the feature map with 2x2 convolutions, succeeded by cropping to deal with border pixels loss in convolutions, and concatenation with the corresponding cropped feature map from the contracting path, and 3×3 convolutions followed by ReLU, rectifying linear units (see Fig. 2(c)).

IV. Results

In this section we perform a comparative study on the state-of-the-art semantic segmentation networks described in the previous section, namely on segmenting corrosion in metallic structures. In all our experiments, the networks were trained and tested on a 12th Gen Intel® Core™ i9-12900KF x 24, with a GeForce RTX 3090ti graphics card.

The dataset used for training and testing comprises 14265 labeled images, which have been gathered with a high definition DSLR camera in a offshore environment, and manually annotated using an online labeling tool [Seg22].

In our experiments we train the models with random crops of size 1024×1024, of the original input images, and we set the training batch size to 8. The dataset is partitioned into 60% training, 20% validation and 20% testing sample sizes. Finally, all models are pre-trained on the imageNet [DDS+09b] dataset.

A. Evaluation Metrics

Evaluating the performance of semantic segmentation tasks requires simultaneously evaluating pixel-level classification and localization accuracy. Let us consider the number of true positive TP, false positive FP, true negative TN, and false negative FN predictions. In our experiments we use the following metrics to evaluate our model:

a) Intersection over Union (IoU): Similarly to the F-score, the intersection over union measures the amount of overlap between ground truth and prediction masks. However, the penalties are higher:

$$IoU = \frac{TP}{TP + FN + FP} \tag{1}$$

b) Precision: Precision measures how accurate the model is at finding true positives, i.e., from all pixels that the model estimates as belonging to the segmentation mask, how many are correctly estimated according to the ground truth. Precision is computed according to the following expression:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

c) Recall: Recall measures true positive rate i.e. from the pixels belonging to ground truth masks, how many are predicted. Recall is computed according to the following expression:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Considering that the ground truth has missing annotations, predictions may be incorrectly classified as being false positive. Therefore, recall is more relevant than precision when ground truth masks are incomplete.

d) Dice Similarity Coefficient (F-score): A popular computer vision metric used to measure similarity between two images, which penalizes under and over segmentation, according to the following formula:

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{4}$$

B. Quantitative and qualitative analysis

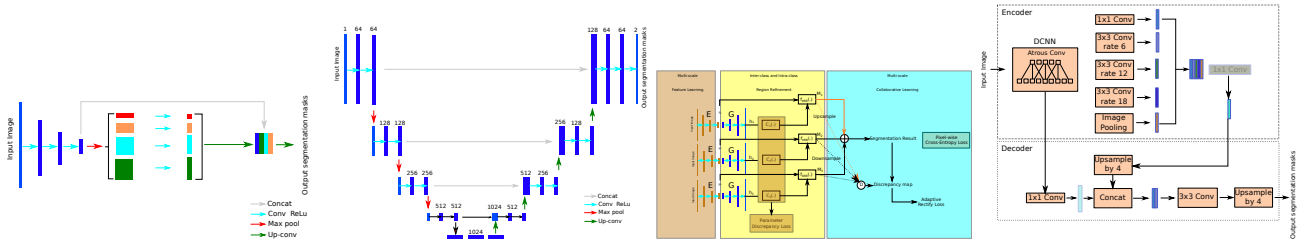
In our experiments we use the dice loss metric for training and validation, and precision, recall and accuracy for testing our models.

Table II presents the obtained results for the architectures described in the methodologies section, on the test set. While PSPnet architecture is the fastest with the lowest average inference time (i.e. $0.0306s$), DeepLab is the slowest one with $0.0637s$ inference time. The most precise one is DeepLab, with a precision score of 0.7485%, and the one with the highest recall is U-net (0.5508).

Figure 3 shows the performance of U-net on different scenes. U-net exhibits high recall, being capable of identifying most ground truth spots. Again, U-net correctly identifies corrosion spots missed by the labeler. Considering that corrosion is hard to label, since small spots are easily missed by human labelers, precision is not the best metric to assess the performance of the former methods.

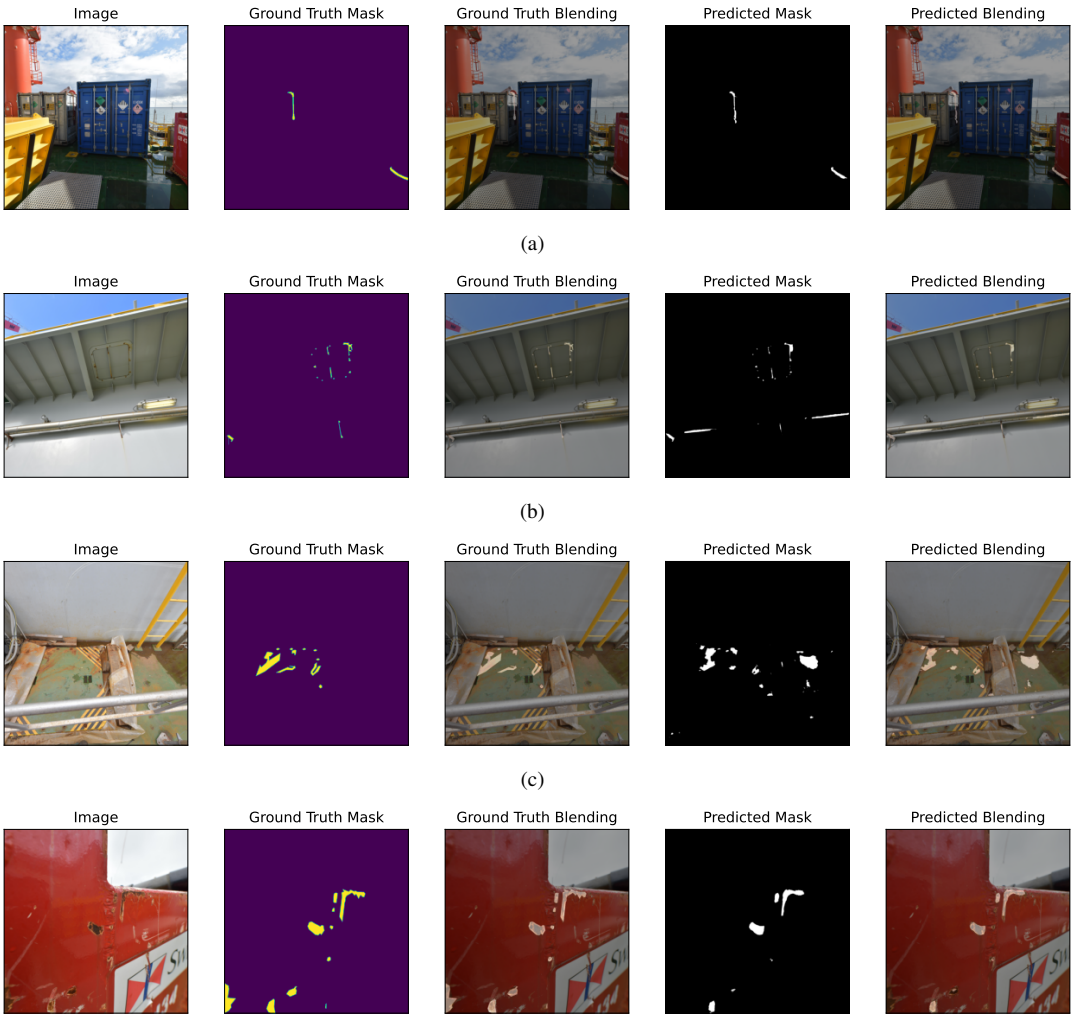
V. Conclusions and Future Work

In this article we assessed the performance of current state-of-the-art neural networks for semantic segmentation, on corrosion segmentation tasks. The results demonstrate that the



(a) PSPNET (b) U-net (c) MANet (d) DeepLabV3

Fig. 2: Architectures of the bench-marked deep neural networks.



(a)

(b)

(c)

(d)

Fig. 3: Semantic segmentation masks obtained for various example scenes using the U-net architecture with ResNet34 backbone.

current state-of-the-art supervised approaches are able to segment corrosion with good performance as long as enough data is provided. One major limitation of supervised deep learning based approaches is that they require the availability of large datasets, manually annotated in a laborious, time-consuming and in a human error prone manner. Furthermore, visual scene analysis tasks are challenging due to high variability in pose, occlusions, clutter and irregular illumination.

In the future we intend to improve our current models using different data augmentation techniques existing in the literature to increase the variability of the data, including extending the dataset via cropping, flipping, rotation, scaling, translation, brightness, and contrast variations. Also, we intend to use sim-to-real transfer [DBF+19] approaches by training the models

TABLE I: Dataset used for training and validating the semantic segmentation networks.

	total images in dataset
train	2853 (20%)
val	2853 (20%)
test	8559 (60%)

TABLE II: Performance results for different network architectures and training parameters.

model	IoU score	Precision	Recall	F-score	Avg inference time
DEEPLABV3-resnet34	0.3736	0.7485	0.4368	0.4688	0.0637
MANET-resnet34	0.4395	0.7151	0.5385	0.5442	0.0471
PSPNET-resnet34	0.3024	0.7049	0.3759	0.3933	0.0306
UNET-resnet34	0.4519	0.7112	0.5508	0.5574	0.0445

with synthetically generated datasets using realistic simulators and rendering engines such as Nvidia Isaac [MWG+21], AirSim [SDLK18] and Gazebo [KH04]. In particular, we intend to evaluate the viability of using physically plausible models of corrosion from the computer graphics literature [MDG01] to synthetically generate large annotated image datasets, to be used for training corrosion segmentation algorithms.

ACKNOWLEDGEMENTS

This work was supported by Predictive Automatic Corrosion Management (EUDP 2021-II PACMAN), project no.: 64021-2072. The authors would further like to thank Semco Maritime for bringing up use-case challenges.

REFERENCES

[20119] *Deep Learning AI for Corrosion Detection*, volume All Days of *NACE CORROSION*, 03 2019. NACE-2019-13267.

[BHH+22] Simon Bøgh, Daniel S Hain, Emil Blixt Hansen, Simon Buus Jensen, Torben Tvedebrink, and Roman Jurowetzki. Predictive analytics applications for small and medium-sized enterprises (smes)—a mini survey and real-world use cases. In *The Future of Smart Production for SMEs: A Methodological and Practical Approach Towards Digitalization in SMEs*, pages 263–279. Springer, 2022.

[COR+16] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[CPK+14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[CPK+18] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.

[DBF+19] Atabak Dehban, João Borrego, Rui Figueiredo, Plinio Moreno, Alexandre Bernardino, and Josá Santos-Victor. The impact of domain randomization on object detection: A case study on parametric shapes and synthetic textures. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2593–2600. IEEE, 2019.

[DDS+09a] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[DDS+09b] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[dFlFSH+21] Rui Pimentel de Figueiredo, Jonas le Fevre Sejersen, Jakob Grimm Hansen, Martim Brandão, and Erdal Kayacan. Real-time volumetric-semantic exploration and mapping: An uncertainty-aware approach. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9064–9070, 2021.

[dFMB13] Rui Pimentel de Figueiredo, Plinio Moreno, and Alexandre Bernardino. Fast 3d object recognition of rotationally symmetric objects. In João M. Sanches, Luisa Micó, and Jaime S. Cardoso, editors, *Pattern Recognition and Image Analysis*, pages 125–132, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[DLHB22] Rui Pimentel De Figueiredo, Jonas Le Fevre Sejersen, Jakob Grimm Hansen, and Martim Brandão. Integrated design-sense-plan architecture for autonomous geometric-semantic mapping with uavs. *Front. Robot. AI*, 9, September 2022. Funding Information: This work was supported by UKRI/EPSRC THuMP (EP/R033722/1) and the Smart Industry Program (European Regional Development Fund and Region Midtjylland, grant no.: RFM-17-0020). Publisher Copyright: Copyright © 2022 Pimentel de Figueiredo, Le Fevre Sejersen, Grimm Hansen and Brandão.

[DSDPM23] Valentina De Simone, Valentina Di Pasquale, and Salvatore Miranda. An overview on the use of ai/ml in manufacturing msmes: solved issues, limits, and challenges. *Procedia Computer Science*, 217:1820–1829, 2023.

[GCE21] Monica Gruosso, Nicola Capece, and Ugo Erra. Human segmentation in surveillance video with deep learning. *Multimedia Tools and Applications*, 80(1):1175–1199, 2021.

[HGDG17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[HJS+22] Pei He, Licheng Jiao, Ronghua Shang, Shuang Wang, Xu Liu, Dou Quan, Kun Yang, and Dong Zhao. Manet: Multi-scale aware-relation network for semantic segmentation in aerial scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

[JF21] Taeho Ju and Alp T. Findikoglu. Monitoring of corrosion effects in pipes with multi-mode acoustic signals. *Applied Acoustics*, 178:107948, 2021.

[KH04] N. Koenig and A. Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, volume 3, pages 2149–2154 vol.3, 2004.

[KSH17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[lFSPDFK21] Jonas le Fevre Sejersen, Rui Pimentel De Figueiredo, and Erdal Kayacan. Safe vessel navigation visually aided by autonomous unmanned aerial vehicles in congested harbors and waterways. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pages 1901–1907, 2021.

[LMB+14a] 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.

[LMB+14b] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. *ArXiv*, abs/1405.0312, 2014.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[MDG01] Stephane Merillou, Jean-michel Dischler, and Djamchid Ghazanfarpour. Corrosion: Simulating and rendering. *Proceedings - Graphics Interface*, 07 2001.

[MWG+21] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.

[NZB22] Will Nash, Liang Zheng, and Nick Birbilis. Deep learning corrosion detection with confidence. *npj Materials Degradation*, 6(1):1–13, 2022.

[OKS21] Samuel Chukwuemeka Olisa, Muhammad A Khan, and Andrew Starr. Review of current guided wave ultrasonic testing (gwut) limitations and future directions. *Sensors*, 21(3):811, 2021.

[PALT20] Xiang Peng, Uchenna Anyaoha, Zheng Liu, and Kazuhiko Tsukada. Analysis of magnetic-flux leakage (mfl) data for pipeline corrosion assessment. *IEEE Transactions on Magnetics*, 56(6):1–15, 2020.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[RPS+21] M. Sai Bhargava Reddy, Deepalekshmi Ponnamma, Kishor Kumar Sadasivuni, Shampa Aich, Saraswathi Kailasa, Hemalatha Parangusan, Muna Ibrahim, Shady Eldeib, Omar Shehata, Mohammad Ismail, and Ranin Zarandah. Sensors in advancing the capabilities of corrosion detection: A review. *Sensors and Actuators A: Physical*, 332:113086, 2021.

[SBKV+20] Hyunseok Seo, Masoud Badiei Khuzani, Varun Vasudevan, Charles Huang, Hongyi Ren, Ruoxiu Xiao, Xiao Jia, and Lei Xing. Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Medical Physics*, 47:e148–e167, 06 2020.

[SDLK18] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and service robotics*, pages 621–635. Springer, 2018.

[Seg22] Segments.ai. 2d 3d data labeling, 2022. https://segments.ai/, Last accessed on 2022-11-30.

[SIBS18] Selim S. Seferbekov, Vladimir I. Iglovikov, Alexander V. Buslaev, and Alexey A. Shvets. Feature pyramid network for multi-class land segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 272–2723, 2018.

[TSSPdF22] Daniel Tøttrup, Stinus Lykke Skovgaard, Jonas le Fevre Sejersen, and Rui Pimentel de Figueiredo. A real-time method for time-to-collision estimation from aerial images. *Journal of Imaging*, 8(3), 2022.

[VEEA06] PR Vaidya, Isaac EINAV, Austria Sinasi EKINCI, and Turkish Atomic Energy Authority. Radiographic evaluation of corrosion and deposits in pipelines: Results of an iaea co-ordinated research programme. *Atomic Energy*, pages 1–14, 2006.

[YJC+21] Biao Yin, Nicholas Josselyn, Thomas Considine, John Kelley, Berend Rinderspacher, Robert Jensen, James Synder, Ziming Zhang, Elke Rundensteiner, and ARL Northeast Regional Extended Site. Corrosion image data set for automating scientific assessment of materials. 2021.

[ZSQ+17a] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *CVPR*, 2017.

[ZSQ+17b] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.