IEEE Copyright notice

This is the author preprint version. ©2023 IEEE, Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A UNIFIED DNN-BASED SYSTEM FOR INDUSTRIAL PIPELINE SEGMENTATION

Dimitrios Psarras *Christos Papaioannidis* *Vasileios Mygdalis* *Ioannis Pitas*

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece

ABSTRACT

This paper presents a unified system tailored for autonomous pipe segmentation within an industrial setting. To this end, it is designed to analyze RGB images captured by Unmanned Aerial Vehicle (UAV)-mounted cameras to predict binary pipe segmentation maps. The overall proposed system consists of three main components: a) a Convolutional Neural Network (CNN) that is used to obtain initial estimates of the pipe segmentation maps, b) a point extraction module that acts on the outputs of the CNN to propose strong pipe class representatives in the input image space, and c) a foundation segmentation model, utilized to refine the initial estimations based on the proposed pipe class representatives. The architecture of the proposed system was specifically designed to ensure increased generalization ability in different, unknown environments, offering an effective solution to a well-known limitation of typical segmentation CNNs, at least in the pipe segmentation task. The effectiveness of the proposed system in this particular setting is evaluated by utilizing two pipe segmentation datasets, originating from two different industrial sites, which were manually annotated with the corresponding pipe segmentation maps. Experimental results demonstrate that the proposed system outperforms the baseline segmentation CNNs, demonstrating its remarkable generalization capabilities.

Index Terms— Industrial Pipeline Segmentation, Convolutional Neural Networks, Foundation Models, Autonomous Inspection

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, are experiencing a growing utilization in various industrial sectors [1, 2] and applications [3, 4]. Among these applications, the inspection of pipes for potential damages stands out as a significant use case [5, 6]. An integral component of the inspection procedure involves accurately determining the precise location of the pipe within the two-dimensional (2D) plane of the RGB image.

This work has received funding from the European Union's Horizon research and innovation programme under grant agreement number 101070604 (SIMAR).

Currently, there exists a limited body of research dedicated to addressing the specific task of pipe segmentation. Earlier approaches have applied classical segmentation techniques on infrared images [6], while more recently pipe segmentation has been also achieved through supervised models that are trained on annotated datasets. The most typical approach in the latter case is to employ a Convolutional Neural Network (CNN) for the pipe segmentation task [1, 5]. However, it is well-known that supervised models trained on small annotated datasets tend to suffer from overfitting [7], which significantly limits their generalization capabilities.

Recently, foundation models for image segmentation have emerged, with the Segment Anything Model (SAM) [8] standing out as a noteworthy example. These foundations models are pre-trained using a vast amount of data, rendering them capable of performing object segmentation with remarkable accuracy and zero-shot generalization to unfamiliar objects and images without requiring additional training. However, these models are not inherently capable of generating inferences without specific prompts, thus rendering them unsuitable for automatic pipe segmentation.

Motivated by the limitations of the aforementioned approaches, the goal of this work is to introduce a deep learning-based system for automatic pipe segmentation, which demonstrates enhanced generalization capabilities, thus rendering it suitable for real-world applications. To this end, the proposed method combines the effectiveness of typical segmentation CNNs and the zero-shot generalization capabilities of foundation models into a unified system for pipe segmentation. More specifically, the proposed system utilizes a pre-trained image segmentation CNN to predict initial estimates of the pipe segmentation maps, which are subsequently used by a point extraction module that proposes strong pipe class representatives in the input image space. The latter are used as prompt inputs to the employed foundation segmentation model, which refines the initial estimates obtained by the CNN. The final pipe segmentation maps are then obtained by matching the refined segmentation maps with the zero-shot object proposals, also produced by the employed foundation model. The generalization ability of the proposed method was evaluated by utilizing two manually annotated pipe segmentation datasets that correspond to two different industrial environments. Experimental results show that the proposed unified system outperforms both typical segmentation CNNs

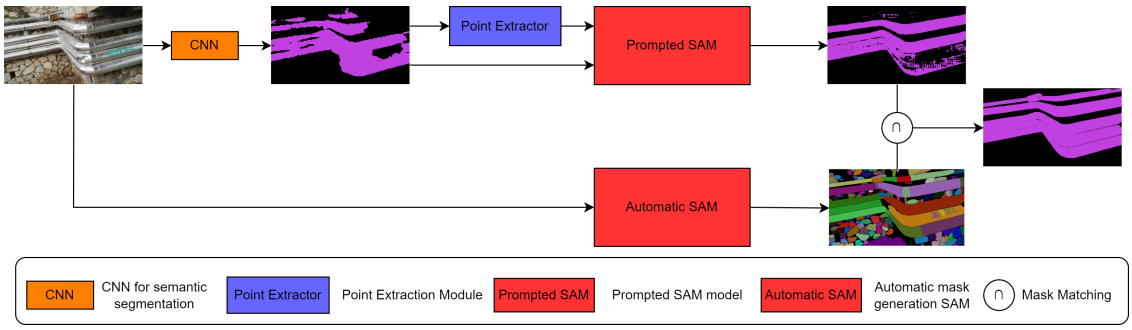

Fig. 1. Overall architecture of the proposed segmentation system. It comprises a pre-trained image segmentation CNN, a point extraction module and the foundation segmenta- tion model both promptable and automatic.

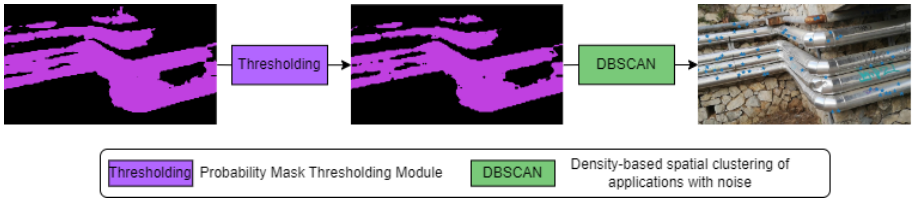

Fig. 2. Point extractor Module extracts from a thresholded version of the CNN segmentation mask point that lay on the pipes. These points are used to prompt the the SAM in the later stages of the proposed system.

and foundation models in this task.

2. RELATED WORK

2.1. Supervised Convolutional models

Pipe segmentation is essentially a semantic image segmentation problem [1, 2], wherein every pixel in an input image is allocated a per-class probability for one of two object categories: *pipe* or *non-pipe*. Besides traditional approaches for pipe segmentation [6], typical segmentation CNNs [1, 9, 10, 11, 12, 13] can also be employed to this end, by simply tasking them to predict binary pipe segmentation masks. A popular example is the U-Net architecture [9] which has emerged as a pivotal deep learning framework in the field of computer vision. The U-Net architecture consists of a contracting path, the encoder, followed by an expansive path, decoder, which outputs the final predictions [9]. Several extensions of U-Net were later proposed, such as U-Net++ [10], that demonstrated increased performance. Having in mind execution speed, the BiSeNet (Bilateral Segmentation Network) model was proposed in [11], which represents a significant advancement in the field of real-time semantic segmentation. BiSeNet's architecture is characterized by its two-branch design, which efficiently combines global and local context information to predict accurate segmentation maps [11].

Based on the BiSeNet, I2I-CNN [12] merges a semantic image segmentation network with an Image-to-Image (I2I) [14] network to precisely predict segmentation maps. The I2I neural branch enriches the segmentation neural branch by providing additional semantic information through skip connections that interconnect the two branches. This integration enhances the accuracy of the segmentation task. Notably, both networks share a single backbone and feature extraction CNN, undergoing joint training through a multi-task objective function.

2.2. Foundation Models for Image Segmentation

Foundation models [8, 15] are large models, typically pre-trained on massive datasets. Due to their large capacity and advanced training procedure on huge amounts of data, they typically offer improved flexibility and adaptability, leading to accurate predictions in diverse scenarios, often surpassing the performance of specialized models. Some typical examples that have recently emerged in the computer vision field are SEEM [15], SAM [8] and SegGPT [16]. Despite their remarkable capabilities, these models are not inherently capable of making accurate predictions without specific prompts.

3. AUTOMATIC PIPE SEGMENTATION

The primary goal of this work is to introduce a method that offers an effective solution for real-world pipe segmentation from RGB images. In this direction, the proposed system combines the state-of-the-art image segmentation I2I-CNN network with the SAM foundation model in a novel configuration, which allows it to predict accurate pipe segmentation maps in unknown environments.

3.1. Unified Pipe Segmentation System

The unified pipe segmentation system consists of three main components, a) the pre-trained image segmentation CNN, b) the point extraction module, and c) the foundation segmentation model, as illustrated in Fig. 1.

Given an RGB input image $\mathbf{X} \in \mathbb{R}^{M \times N \times 3}$ of height M and width N, the employed segmentation I2I-CNN model first calculates an initial estimate of the corresponding pipe segmentation map in the form of a pipe class probability tensor $\mathbf{P} \in \mathbb{R}^{M \times N \times 2}$, where each channel is a probability map for the *pipe* and *non-pipe* classes, respectively. With the probability tensor $\mathbf{P}$ available, the corresponding binary pipe segmentation map $\tilde{\mathbf{S}} \in \mathbb{R}^{M \times N}$ can be calculated by simply choosing the channel index with the maximum probability value for each pixel. Note that $\tilde{\mathbf{S}}$ in this stage is only an initial estimate of the pipe segmentation map.

Subsequently, the initial estimate of the pipe segmentation map $\tilde{\mathbf{S}}$ and the probability tensor $\mathbf{P}$ are exploited as follows. First, $\tilde{\mathbf{S}}$ serves as an input mask prompt to the employed SAM model. Second, the probability tensor $\mathbf{P}$ is utilized in the Point Extraction module to obtain strong pipe class representatives, which act as additional prompts for the SAM model. In order to accomplish this, $\mathbf{P}$ is given to a Probability Mask Thesholding Module to produce a new mask $\mathbf{S}' \in \mathbb{R}^{M \times N}$, where only *pipe* class labels that have a *pipe* class probability over 0.9 are present, while all other entries are labeled as *non-pipe*. Therefore, $\mathbf{S}'$ is calculated according to the following equation:

$$\mathbf{S}'(i,j) = \begin{cases} c_{pipe} & \mathbf{P}_{pipe}(i,j) \geq 0.9 \\ c_{non-pipe} & \mathbf{P}_{pipe}(i,j) < 0.9 \end{cases} \quad (1)$$

where $i = 1, \ldots, M$, $j = 1, \ldots N$, $\mathbf{P}_{pipe}$ denotes the probability tensor $\mathbf{P}$ channel that corresponds to the *pipe* class and c_{pipe}, $c_{non-pipe}$ are the class labels that correspond to the *pipe* and *non-pipe* classes. The calculated $\mathbf{S}'$ is then fed to the second stage of the Point Extraction module, that consists of a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [17] algorithm. More specifically, the DBSCAN algorithm is applied to $\mathbf{S}'$ for cluster identification and centroid extraction. Essentially, the DBSCAN enables us to choose strong *pipe* class representatives from the input image space by extracting the pixel coordinates that correspond to the centroids calculated by the DBSCAN algorithm. The Point Extraction module is presented in Fig. 2. The extracted *pipe* class representatives are then used as additional input prompts, along with $\tilde{\mathbf{S}}$, for prompting the employed SAM model and producing the refined binary segmentation mask $\mathbf{S}_{prompt} \in \mathbb{R}^{M \times N}$.

However, the input prompts of the foundation model may be noisy due to errors in $\tilde{\mathbf{S}}$, originating from the CNN model predictions, and/or errors introduced by imperfect clustering results obtained by DBSCAN. Consequently, the refined segmentation mask $\mathbf{S}_{prompt}$ may still be not accurate enough. To address this challenge and further increase the accuracy of the predicted segmentation masks, $\mathbf{X}$ is processed by the employed SAM in an automatic segmentation mode, producing a new mask tensor $\mathbf{S}_{auto} = \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_K\}, \in \mathbb{R}^{M \times N \times K}$, which consists of K object proposal binary masks $\mathbf{S}_k \in \mathbb{R}^{M \times N}$, $k = 1, \ldots, K$. Given $\mathbf{S}_{prompt}$ and $\mathbf{S}_{auto}$, the final predicted binary pipe segmentation mask $\mathbf{S} \in \mathbb{R}^{M \times N}$ is obtained by applying the mask matching operation between $\mathbf{S}_{prompt}$ and $\mathbf{S}_{auto}$:

$$\mathbf{S} = \cup_{k=1}^{K} \mathbf{S}_k, \text{ if } \mathbf{S}_k \cap \mathbf{S}_{prompt} \neq \emptyset \quad (2)$$

The overall multi-step process of the proposed system ensures the accurate pipe segmentation in unknown environments, as presented in the following Section.

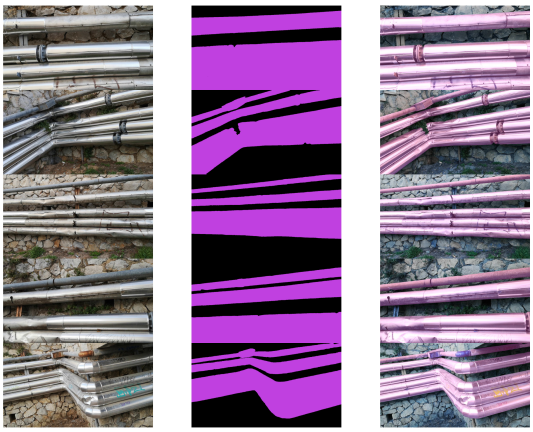

Fig. 3. Example samples of the manually annotated AUTH pipes dataset. Left: RGB images, middle: RGB segmentation masks, right: visualization of annotation on the RGB image.

4. EXPERIMENTS AND RESULTS

4.1. Datasets

While there are publicly accessible datasets pertaining to pipes [18, 19, 20], they are unsuitable for pixel-level segmentation, as they typically offer bounding box annotations instead of segmentation maps [20]. Therefore, in order to evaluate the proposed system for real-world pixel-level pipe segmentation, two RGB image datasets were recorded and manually annotated.

The Refinery pipe dataset includes diverse RGB images of operational insulated pipes, captured at varying angles and featuring different pipe quantities within each frame. These images originate from industrial settings and span resolutions from 1920×1080 to 4032×3024. The dataset comprises 1k RGB images that are all used for training purposes. Please note that, due to confidentiality constraints, this dataset cannot be publicly shared. The AUTH pipe dataset consists of real RGB images depicting operational pipes captured at AUTH facilities, all maintaining a consistent resolution of

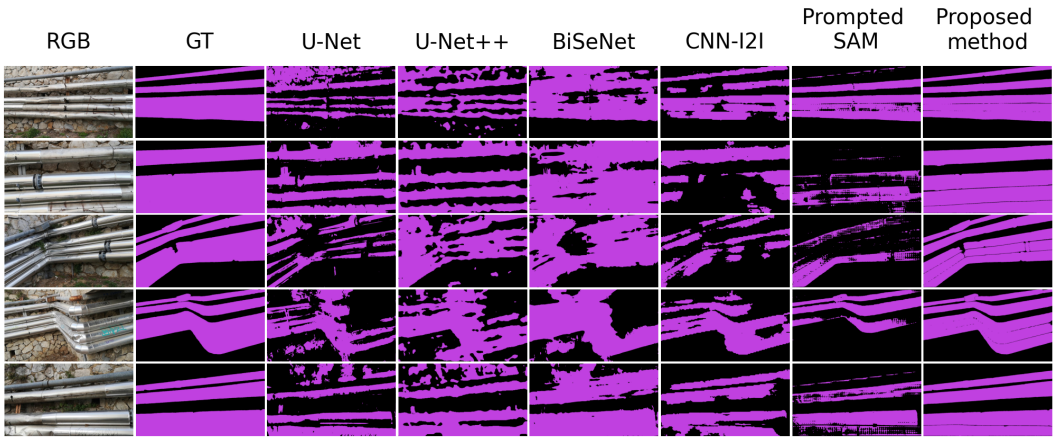

Fig. 4. Pipe segmentation results for test pipe images from AUTH pipe dataset. Each row depicts a random test image along with the corresponding ground-truth and the predicted segmentation masks obtained by all competing methods.

Table 1. Pipe segmentation performance on 1280×720 resolution of both *pipe* and *non-pipe* classes on the manually annotated AUTH dataset.

	IoU (%)			
	non-pipe	**pipe**	**mIoU**	**mPA(%)**
U-Net [9]	52.0	46.1	49.1	66.0
U-Net++ [10]	51.4	58.3	54.8	71.1
BiSeNet [11]	54.2	65.4	59.8	75.4
I2I-CNN [12]	68.5	63.7	66.1	79.7
prompted SAM	78.9	79.3	79.1	88.3
Proposed System	**89.0**	**90.9**	**89.9**	**94.8**

1920×1080. This dataset includes a total of 77 images exclusively designated for evaluation to assess the system's generalization capability. It is publicly accessible and available for use. Example samples from the AUTH pipe dataset are shown in Fig. 3.

4.2. Evaluation Procedure

In all experimental sessions, the CNN models were trained using the Refinery pipe dataset. The proposed method is compared to four competitors:U-Net [9], U-Net++ [10], BiSeNet [11], I2I-CNN [12] and a modified version of the system that excludes the automatic SAM segmentation step. U-Net++ was trained for 100 epochs with batch size 8 and a learning rate 0.0001, while BiSeNet was trained for the pipe segmentation task using up to 120 epochs and the Adam optimizer with batch size 8 and an initial rate 0.001. Note that, the backbone network (ResNet-18) of BiSeNet is pretrained on ImageNet. Similarly, the I2I-CNN was trained using the same parameters as BiSeNet model however for both pipe segmentation and image-to-image translation tasks [12]. In order to evaluate the pipe segmentation performance of the proposed system and all competing methods in a real-world pipe segmentation scenario, Refinery pipe dataset was used for training and the AUTH pipe dataset was used only for testing. All methods were evaluated in the pipe segmentation task using the common Intersection over Union (IoU) and mean Pixel Accuracy (mPA) metrics. The IoU for both *pipe* and *non-pipe* classes and mPA of all models are reported in Table 1. Results are reported at 1280×720 input resolution (by training and testing all models accordingly). As demonstrated in Table 1, the proposed method outperforms all competing methods when applied to the new domain data. This is also evident in the qualitative evaluation presented in Fig. 4, where ground-truth and predicted segmentation masks obtained by each method are depicted. As it can be seen, the CNN models (columns 3-5) predict noisy segmentation masks, misclassifying *non-pipe* pixels as *pipe* ones. These errors affect the Point Extractor Module, leading to noisy prompts that hinder "prompted SAM" performance (forth column). In contrast, the proposed method effectively combines the I2I-CNN model and SAM operating in both prompted and automatic modes in a unified pipe segmentation system that accurately predicts pipe segmentation masks (final column).

5. CONCLUSIONS

In this paper a system designed to generalize in new domain images for the pipe segmentation task was presented. The system is based on combining a segmentation CNN model with a foundation model, in order to automate the prompting of the latter and utilize its zero-shot generalization capabilities without requiring additional training. To train and evaluate the effectiveness of the proposed system, two new datasets originating from different sites were introduced, the Refinery pipe dataset for training and the AUTH pipe dataset for testing. The proposed system significantly outperforms all the other competing models in segmenting new domain images for the pipe segmentation task.

6. REFERENCES

[1] E. Guerra, J. Palacin, Z. Wang, and A. Grau, "Deep learning-based detection of pipes in industrial environments," in *Industrial Robotics-New Paradigms*. IntechOpen, 2020.

[2] P. Ravishankar, S. Hwang, J. Zhang, IX Khalilullah, and B. Eren-Tokgoz, "Darts—drone and artificial intelligence reconsolidated technological solution for increasing the oil and gas pipeline resilience," *International Journal of Disaster Risk Science*, vol. 13, no. 5, pp. 810–821, 2022.

[3] I. Mademlis, A. Torres-Gonzalez, J. Capitan, M. Montagnuolo, A. Messina, F. Negro, C. Le Barz, T. Goncalves, R. Cunha, B. Guerreiro, F. Zhang, S. Boyle, G. Guerout, A.Tefas, N.Nikolaidis, D. Bull, and I.Pitas, "A multiple-uav architecture for autonomous media production," *Springer Multimedia Tools and Applications*, pp. 1–30, 2022.

[4] G. Kalitsios, V. Mygdalis, and I.Pitas, "Enhancing power line segmentation for uav inspection utilizing synthetic data," *IEEE International Conference on Robotics and Automation (ICRA), Workshop on The Role of Robotics Simulators for Unmanned Aerial Vehicles*, 2023.

[5] YMR da Silva, FAA Andrade, L. Sousa, GGR de Castro, JT Dias, G. Berger, J. Lima, and MF Pinto, "Computer vision based path following for autonomous unmanned aerial systems in unburied pipeline onshore inspection," *Drones*, vol. 6, no. 12, pp. 410, 2022.

[6] H. Guan, T. Xiao, W. Luo, J. Gu, R. He, and P. Xu, "Automatic fault diagnosis algorithm for hot water pipes based on infrared thermal images," *Building and Environment*, vol. 218, pp. 109111, 2022.

[7] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[8] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, AC Berg, WY Lo, et al., "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[10] Z. Zhou, MM Rahman Siddiquee, N Tajbakhsh, and J Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *4th International Workshop in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA and ML-CDS)*, 2018.

[11] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018.

[12] C. Papaioannidis, I. Mademlis, and I. Pitas, "Autonomous uav safety by visual human crowd detection using multi-task deep neural networks," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[13] C. Papaioannidis, I. Mademlis, and I. Pitas, "Fast semantic image segmentation for autonomous systems," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022.

[14] P. Isola, JY Zhu, T. Zhou, and AA Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[15] X. Zou, J. Yang, H. Zhang, F. Li, L. Li, J. Gao, and YJ Lee, "Segment everything everywhere all at once," *arXiv preprint arXiv:2304.06718*, 2023.

[16] X. Wang, X. Zhang, Y. Cao, W. Wang, C. Shen, and T. Huang, "Seggpt: Segmenting everything in context," *arXiv preprint arXiv:2304.03284*, 2023.

[17] M. Ester, HP Kriegel, J. Sander, X. Xu, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," *kdd*, vol. 96, pp. 226–231, 1996.

[18] CONVR2022, "Construction item dataset," https://universe.roboflow.com/convr2022/construction-item, 2022.

[19] Synthetic Corrosion, "Synthetic corrosion dataset dataset," https://universe.roboflow.com/synthetic-corrosion/synthetic-corrosion-dataset, 2022.

[20] CEFETRJ, "Pipe quali dataset," https://universe.roboflow.com/cefetrj/pipe_quali, 2023.