A UNIFIED DNN-BASED SYSTEM FOR INDUSTRIAL PIPELINE SEGMENTATION

Dimitrios Psarras, Christos Papaioannidis, Vasileios Mygdalis and Ioannis Pitas
Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, 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.

Introduction

Traditional approaches for inspecting pipes for damages have used classical segmentation techniques and Convolutional Neural Networks (CNNs), both of which have limitations such as overfitting when trained on small datasets. Recently, foundation models like the Segment Anything Model (SAM) have shown promise in object segmentation with zero-shot generalization capabilities but require specific prompts for operation, making them unsuitable for automatic segmentation tasks. This work introduces:

- A unified system for autonomous pipe segmentation in industrial environments using UAV-captured RGB images.
- Implementation of a three-component system involving a Convolutional Neural Network (CNN), a point extraction module and a foundation segmentation model.
- Design focused on enhanced generalization capability to operate effectively in varied and unknown environments, addressing a common limitation of standard segmentation CNNs in pipe segmentation.
- Validation of the system's effectiveness through application on two manually annotated pipe segmentation datasets from distinct industrial sites.

Pipeline Datasets for Segmentation

- Existing public datasets for pipes primarily provide bounding box annotations, not suitable for pixel-level segmentation. Two new RGB image datasets were recorded and manually annotated.
- **Refinery Pipe Dataset**: Contains 1000 diverse RGB images of operational insulated pipes from industrial settings, annotated with segmentation masks. Resolutions range from 1920x1080 to 4032x3024. However, it cannot be shared publicly due to confidentiality constraints.
- **AUTH Pipe Dataset**: Consists of 77 real RGB images of operational pipes from AUTH facilities, annotated with segmentation masks, with a consistent resolution of 1920x1080. The dataset is publicly accessible and available for use.

Unified Pipe Segmentation System

- The unified pipe segmentation system is composed of three integral components, which are: a pre-trained convolutional neural network (CNN) for image segmentation, a point extraction module and a foundational segmentation model that refines and validates the output, as illustrated in Fig. 1.

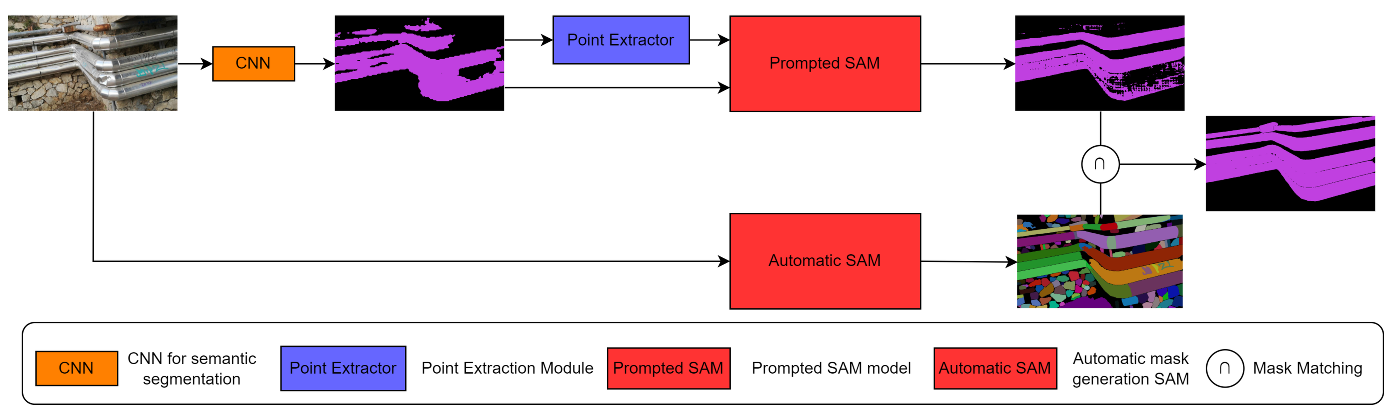

Figure 1: Overall architecture of the proposed segmentation system. It comprises a pre-trained image segmentation CNN, a point extraction module and the foundation segmentation model both promptable and automatic.

- 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 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. The corresponding binary pipe segmentation map $\hat{\mathbf{S}} \in \mathbb{R}^{M \times N}$ can be calculated by choosing the channel index with the maximum probability value for each pixel.
- Following the initial segmentation, the system extracts strong pipe class representatives from the probability tensor $\mathbf{P}$ through the Point Extraction module. $\mathbf{P}$ is given to a Probability Mask Thresholding 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*.

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

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. $\mathbf{S}'$ is then fed to the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [1] algorithm for cluster identification and centroid extraction. The strong *pipe* class representatives are the extracted pixel coordinates that correspond to the centroids. The Point Extraction module is presented in Fig. 2.

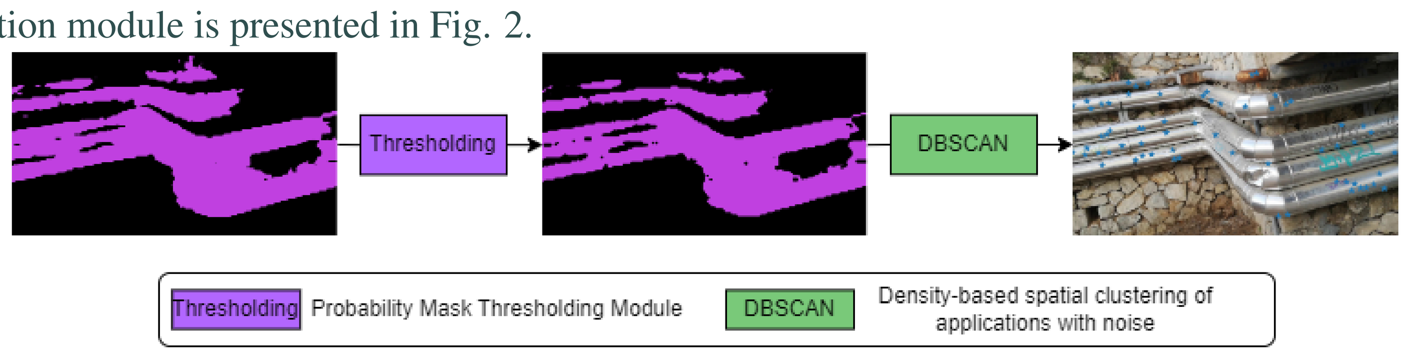

Figure 2: Point extractor Module extracts from a thresholded version of the CNN segmentation mask point that lay on the pipes. These points are used to prompt the the SAM in the later stages of the proposed system.

- The extracted pipe points, indicative of high-confidence pipe areas, are then combined with the initial segmentation map to serve as input for the Segment Anything Model (SAM), producing the refined binary segmentation mask $\mathbf{S}_{prompt} \in \mathbb{R}^{M \times N}$.
- However, this system acknowledges the potential for inaccuracies in the initial predictions made by the CNN and the clustering results. To mitigate this, $\mathbf{X}$ is also 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 $\mathbf{S}_{auto}$:

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

Experiments

In all experimental sessions, the CNN models were trained using the Refinery pipe dataset and tested on the AUTH pipe dataset. The proposed method is compared to four competitors:U-Net [2], U-Net++ [3], BiSeNet [4], I2I-CNN [5] and a modified version of the system that excludes the automatic SAM segmentation step. All methods were evaluated in the pipe segmentation task using the common Intersection over Union (IoU) and mean Pixel Accuracy (mPA) metrics and results are reported at 1280×720 input resolution. 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. 3. As it can be seen, the CNN models (columns 3-5) predict noisy segmentation masks, misclassifying *non-pipe* pixels as *pipe* ones. 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).

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 [2]	52.0	46.1	49.1	66.0
U-Net++ [3]	51.4	58.3	54.8	71.1
BiSeNet [4]	54.2	65.4	59.8	75.4
I2I-CNN [5]	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**

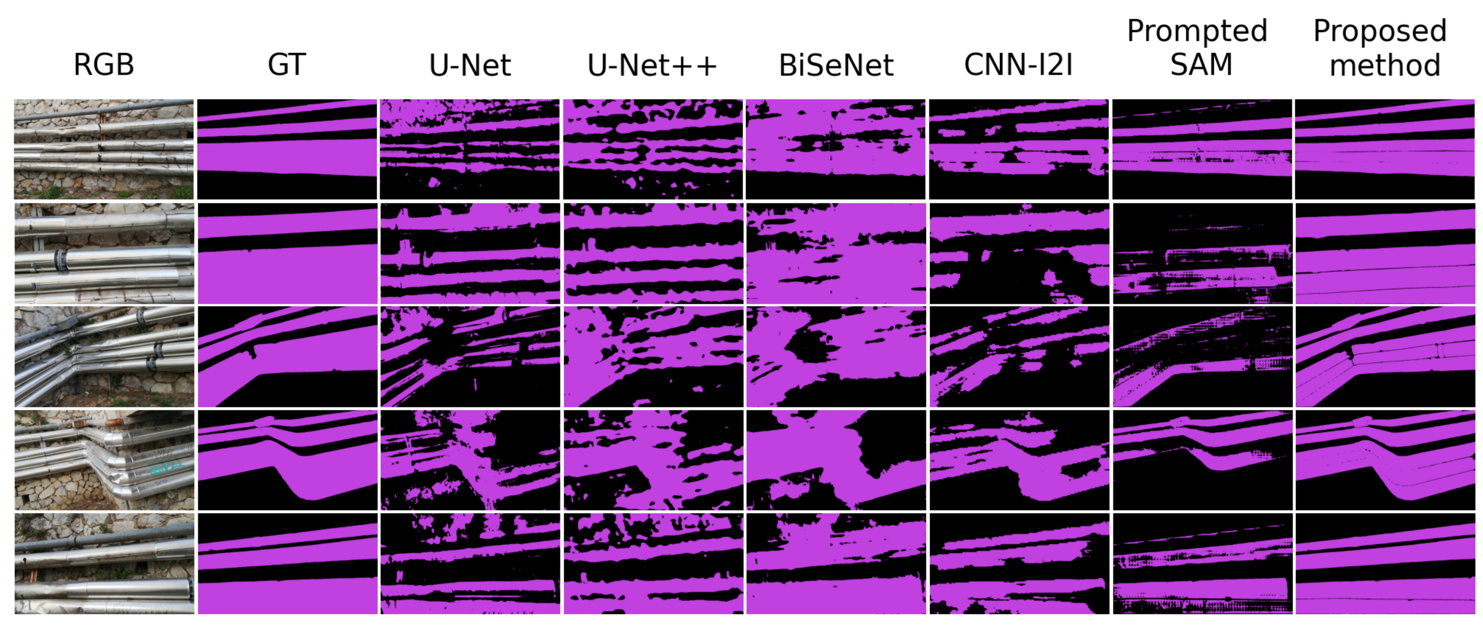

Figure 3: Pipe segmentation results for test pipe images from AUTH pipe dataset. Each row depicts a random test image along with the corresponding ground-truth and the predicted segmentation masks obtained by all competing methods.

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.

References

[1] M. Ester, H. 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.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[3] Z. Zhou, M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *DLMIA and ML-CDS*, 2018.

[4] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *ECCV*, 2018.

[5] C. Papaioannidis, I. Mademlis, and I. Pitas, "Autonomous uav safety by visual human crowd detection using multi-task deep neural networks," in *ICRA*, 2021.

Acknowledgements

Supported by EU Horizon research and innovation programme under g.a.n 101070604 (SIMAR).