Viability of pothole measurement using single image depth estimation using Depth-Anything 3

Alvandy Maulana Yusuf, Shin Soo Young*

Kumoh National Institute of Technology

alvandyyusuf2025@kumoh.ac.kr, *wdragon@kumoh.ac.kr

Abstract

Road distress algorithms are useful in reducing the time-consuming process of manually finding and measuring the potholes in the roads. However, most image-based algorithms like YOLO often only tackle in the detection of the potholes. More detailed measurements, like the depth, would require more advanced methods, such as photogrammetry. Such methods require more numerous images. However, advanced algorithm, such as Depth-Anything, can be used for depth estimation on only a single image, after which masking and relative-depth normalization are applied. The current approach does not recover metric depth, but is able to determine relative depth. This study highlights the potential and drawbacks of single image-based depth estimation for pothole measurement.

Keywords: Depth-Anything, Depth-Estimation, Pothole Measurement

Ⅰ. Introduction

Road damage is a common issue that is usually quite inevitable, given the frequency of the use of roads in our modern world. As such, being able to keep track of the current condition of the road would be immensely beneficial, as it would allow for the maintenance of the roads prior to being too heavily worn out for standard use. Road conditions usually assessed in four main aspects, including roughness, distress, bearing capacity, and skid resistance [2]. One way of measuring the current state of the road would be the use of laser-scanning, stereo camera setups, or photogrammetry, which would require multiple images. For instance, Tan and Li [4] propose an effective method for using UAV for Photogrammetry-based Road damage detection. However, these approaches come with drawbacks, namely in their complexity given their resource-constrained environments.

A potential alternative is in the use of monocular depth estimation from a single RGB image. More recent advancement in this field has shown great results in inferring depth maps from just a single image. However, this usually comes at a cost of accuracy, and the sensitivity of the model.

This paper aims to investigate the feasibility of using Depth-Anything V3 for road pothole measurement. In particular, its strong general performance in a wide variety of situations [1]. While the current scenario is far from an ideal use of the model, it is an interesting alternative to other previously mentioned methods, especially when considering its requirement of only one single RGB image.

Ⅱ. Method

The proposed method consists of four main stages:

A. Pothole Region Extraction

With an RGB image as the input, a binary pothole mask is generated. This mask segments the image into pothole areas and restricts it to these specific areas. Image is cropped to minimize to meet the input constraints.

Fig. 1. Original Pothole Image

Fig. 2. Binary Mask

B. Monocular Depth Estimation

Depth Anything V3 is a transformer based monocular depth model that is used in this section to estimate the depth of the pothole [1].

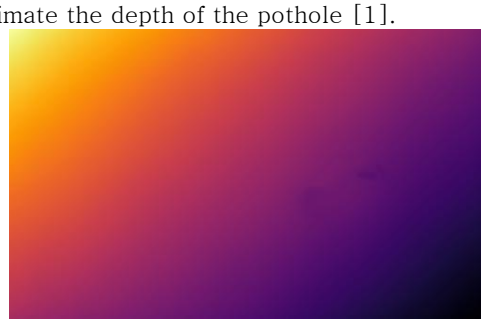

Fig. 3. Raw Depth

C. Relative Depth Normalization

The relative depth generated in the previous section is then normalized relative to the surface of the road, specifically the surfaces near to the boundary of the pothole mask.

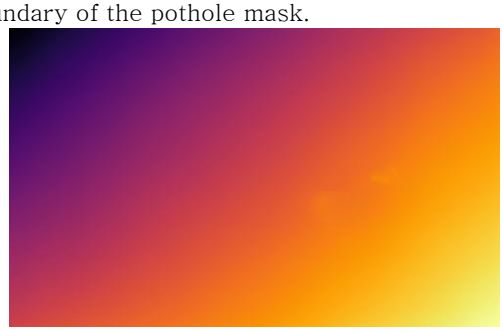

Fig. 4. Relative Depth

D. Local surface reconstruction

Depth values are converted to 3D Point Clouds. While the generated reconstruction is not metrically scaled, the relative depths are maintained.

The current pipeline uses a publicly available dataset containing different examples of potholes in roads [3]. One single image was used from the dataset in order to test the proposed pipeline.

Ⅲ. Conclusion

The proposed method uses Depth Anything V3, alongside binary masking and relative depth normalization in order to obtain the measurement of the Pothole. The current state of the method shows potential, however requires more development, especially in regards to the depth estimation. Due to the limitation of using only one single RGB image, any form of obstruction or distortion will result in unexpected results. Other methods, such as photogrammetry, are much more reliable in this regard.

However, the potential of using only a single image comes with the benefits. It allows for easier data collection, especially in dynamic situations where capturing multiple images may not be viable. Requiring only one single image means instantaneous capture of the scene, whereas methods like Photogrammetry would require multiple images, which means it takes much longer, and as such much less instantaneous.

The current methodology would need to be improved further, as the current results shows only relative depth. In addition to this, monocular depth models usually focus on the whole image, as opposed to viewing finer details, such as small potholes. This results in the pothole being underrepresented.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-00553810, 50%) This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2018R1A6A1A03024003, 50%)

REFERENCES

[1] H. Lin et al., “Depth Anything 3: Recovering the Visual Space from Any Views,” arXiv (Cornell University), Nov. 2025, [Online]. Available: http://arxiv.org/abs/2511.10647

[2] “JTG/T 5142-01-2021 English version, JTG/T 5142-01-2021 Technical Specifications for Preventive Maintenance of Highway Asphalt Pavement (English version) - Code of China,” Jan. 20, 5142. https://codeofchina.com/standard/JTGT5142-01-2021.html

[3] ruirangerfan, “GitHub - ruirangerfan/stereo_pothole_datasets: Pothole Detection Based on Disparity Transformation and Road Surface Modeling (T-IP),” GitHub, 2025. https://github.com/ruirangerfan/stereo_pothole_datasets (accessed Jan. 04, 2026).

[4] Y. Tan and Y. Li, “UAV Photogrammetry-Based 3D Road Distress Detection,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 9, p. 409, Sep. 2019, doi: 10.3390/ijgi8090409.