Two-Stage Pavement Distress Detection and Classification Using UAV Imagery in Unstructured Road Environments

Earl Ryan Aleluya[1,2(✉)], Melody Mae Maluya[2], Francis Jann Alagon[1,2], Edward Rowe Aleluya[3], and Carl John Salaan[2,4]

[1] Department of Computer Engineering and Mechatronics, Mindanao State University - Iligan Institute of Technology, Iligan 9200, Philippines
{earlryan.aleluya,francisjann.alagon}@g.msuiit.edu.ph
[2] Center for Mechatronics and Robotics, Mindanao State University - Iligan Institute of Technology, Iligan 9200, Philippines
{melodymae.maluya0101,carljohn.salaan}@g.msuiit.edu.ph
[3] Department of Physical Sciences and Mathematics, Mindanao State University at Naawan, Naawan 9023, Philippines
edwardrowe.aleluya@msunaawan.edu.ph
[4] Department of Mechanical Engineering and Technology, Mindanao State University - Iligan Institute of Technology, Iligan 9200, Philippines

Abstract. Although tracking road degradation for timely maintenance extends infrastructure lifespan, UAV-based automation struggles on developing regions' unstructured roads due to missing lane markings, irregular surfaces, inconsistent signage, and frequent visual occlusions. This study introduces a two-stage predictive system in which the first stage concurrently detects road lanes and pavement panels, and the second stage uses these outputs to classify each pavement panel into one of four categories: non-distressed, sealed patch, sealed crack, or thin crack. Experimental evaluation on the pavement distress classification task shows that the proposed method achieves a mean Average Precision (mAP) of 0.9775±0.0015 across all classes. Future work will extend the classification framework to incorporate additional distress types observed in concrete, asphalt, and composite road surfaces.

Keywords: Aerial vehicles · Computer vision · Deep learning · Pavement management system · Road transportation · Vision transformers

1 Introduction

Routine road inspection helps modern pavement management systems (PMS) to identify structural degradation early, mitigate severe accident risks, and facilitate proactive maintenance planning for long-term infrastructure sustainability [11,26]. The primary objective of these assessments is the detection of pavement distress, which is defined as any observable surface damage or structural

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16761, pp. 479–490, 2027.
https://doi.org/10.1007/978-3-032-30527-5_31

deterioration—such as cracking and rutting—that negatively impacts a roadway's functional performance and overall safety [9]. To localize this degradation, condition evaluations are spatially organized along a road lane, defined as the continuous longitudinal surface corridor designated to guide directional vehicular transit. For high-resolution, localized assessment, these lanes are frequently partitioned into a pavement panel, which refers to a specific, bounded structural slab or segmented physical unit of the road surface utilized by automated systems and surveyors to precisely isolate damage [12]. Identifying distress features at the granular level of individual panels and lanes is imperative for tailoring material-specific repair strategies. As a result, it can reduce the hazards associated with manual field labor, and improve the transportation infrastructure management [6,14].

Replacing conventional manual surveys, unmanned aerial vehicles (UAVs) recently serve as a transformative technology for the automated identification of pavement degradation. To optimize the precision and speed of these inspections, numerous investigations have merged drone technology with advanced deep learning frameworks. The literature covers various applications of diverse neural network architectures, including Vision Transformers (ViT) [24], Faster R-CNN [26], and multiple YOLO iterations—ranging from YOLOv3 and YOLOv4 [17,26] to YOLOv8 [12,13,16] and YOLOv5 enhanced with an Efficient Channel Attention module [23]. By training on aerial image datasets, these models map distresses such as potholes [16,17,21], patching [21], delamination [16], structural repairs [16], and specific crack morphologies (longitudinal, transverse, alligator, and oblique) [16,21,23].

Vision-based pavement distress analysis is typically addressed by single-stage architectures to integrate detection and classification for real-time applications. While earlier approaches often treated classification using standalone Convolutional Neural Networks (like ResNet and VGG) or Vision Transformers as distinct from spatial detection, modern systems demand rapid, unified processing. Consequently, single-stage detectors—particularly YOLO-based models—emerge as off-the-shelf solution to simultaneously localizing and categorizing distresses like cracks and potholes across vehicle-mounted and UAV platforms [5,16]. Coupling these streamlined single-stage models with advanced UAV sensors, such as multispectral and binocular imaging, broadens distress monitoring frameworks [15,23].

In pavement distress monitoring, the multi-stage architectures divide the complex inspection tasks of localization and pixel-level extraction into distinct, manageable phases. A dominant two-stage implementation utilizes bounding-box object detection models to first isolate distress regions like cracks and potholes, followed immediately by semantic segmentation networks to extract the exact damage pixels [8,10,20]. Other two-stage predictors prioritize broad identification first, employing convolutional neural networks to select crack-containing images before applying transformers for detailed pixel detection [7], using Mask R-CNN for initial detection prior to severity classification via image processing [19], or extracting potential anomalies from smartphone geospatial series before

categorizing them [25]. Expanding into three-stage architectures introduces dedicated preprocessing or refinement steps, such as the sequence of image enhancement, YOLO-SAMT detection, and K-means clustering in [22], or the pipeline of saliency-based localization, CNN-based pixel recognition, and region growing in [18].

While existing multi-stage frameworks isolate distress localization from pixel-level segmentation [8,10,20] or severity classification [19], they predominantly target the damage features directly rather than establishing the broader structural context of the road. These conventional methods typically depend on unified distress datasets and are largely optimized for standard, well-defined driving conditions [7,25], which may become vulnerable to the high variability and noise of unstructured roads. To address these limitations, our work's main contributions are as follows.

1. We introduce a two-stage framework that first performs concurrent binary detection of road lanes and pavement panels to anchor the subsequent distress classification.
2. We design the first stage to suppress false positives by constraining panels to detected lanes and recover false negatives via lane boundary estimation.
3. We implement disjoint multi-task learning approach across three specific datasets—single-class road lane data, single-class pavement panel data, and 4-class labeled pavement panel data.
4. We design our two-stage framework for the unstructured road environments.

The rest of the paper are organized as follows. Section 2 presents the two-stage framework, while the experimental validations are discussed in Sect. 3. Lastly, Sect. 4 concludes the paper.

2 Materials and Methods

2.1 Problem Formulation

The pavement distress identification task is formulated as a supervised object detection and segmentation problem, where a given input image $I \in \mathbb{R}^{(H \times W \times 3)}$ represents an RGB scene of road surface conditions. The objective is to learn a mapping function f_θ parameterized by network weights θ, such that $f_\theta(I) = \{(B_i, M_i, c)\}_{i=1}^{N}$, where $B_i \in \mathbb{R}^4$ denotes the bounding box coordinates, $M_i \in \{0,1\}^{H \times W}$ is the binary segmentation mask, and $c_i \in \mathcal{C}$ is the class label for the i-th detected distress instance among N objects. The learning objective is to minimize a composite loss function $\mathcal{L}(\theta)$, which jointly optimizes classification accuracy, localization precision, and mask quality.

This study introduces a two-stage framework that extends the base detection formulation. As shown in Fig. 1, parallel segmentation predictors in the first stage are defined as $f_{\theta_1}^{\text{lane}}(I) \rightarrow \mathcal{B}_{\text{lane}} = \left\{ b_i^{(l)} \right\}_{i=1}^{m}$ and $f_{\theta_2}^{\text{panel}}(I) \rightarrow \mathcal{B}_{\text{panel}} = \left\{ b_j^{(p)} \right\}_{j=1}^{n}$, where each $b^{(\cdot)} \in \{0,1\}^{H \times W}$ represents a binary mask for road lanes and pavement panels, respectively. A post-processing operator $\mathcal{R}(\cdot)$ is then applied to

enforce spatial consistency, which yields refined panel masks $\tilde{\mathcal{B}}_{\text{panel}} = \left\{\tilde{b}_j^{(p)}\right\}$ such that $\tilde{b}_j^{(p)} = b_j^{(p)} \odot \left(\bigvee_{i=1}^{m} b_i^{(l)}\right)$, where $\odot$ denotes element-wise multiplication and $\bigvee$ is the pixel-wise logical OR to merge lane masks. In the second stage, a classification network g_{θ_3} maps each refined mask region to a class label, i.e., $g_{\theta_3}(\tilde{b}_j^{(p)}) = c_j$, where c_j is either non-distressed, sealed patch, sealed crack, or thin crack. This decomposition adopts structured learning by decoupling geometric segmentation from semantic distress classification.

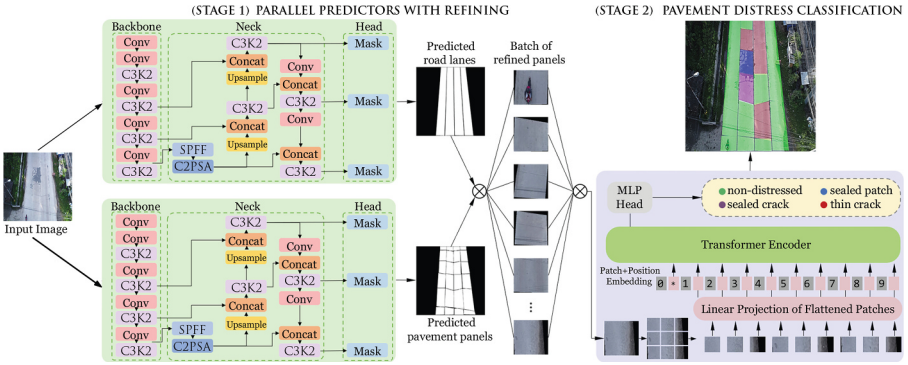

Fig. 1. The overview of the two-stage prediction pipeline, which contains parallel predictors for road lanes and pavement panels, and pavement distress classification using Vision Transformer.

2.2 Parallel Predictors with Refining

The first stage has two disjoint segmentation models that adopt the YOLO11-*seg* architecture. The architecture builds upon its predecessors by integrating advanced feature extraction and segmentation mechanisms, which resulted in better accuracy and computational efficiency. One of its core innovations is the C3K2 block, which employs smaller kernel sizes to refine feature extraction while reducing computational cost. The architecture also incorporates the Spatial Pyramid Pooling Fast (SPPF) module, which facilitates multi-scale feature aggregation. Given an F-sized input sequence of aerial images $I = \{I_i\}_{i=1}^{F}$ such that each input image $I_i \in \mathbb{R}^{H \times W \times C}$, the YOLO11-*seg* models for road lanes and pavement panels produce the corresponding lane segmentation masks $\mathcal{S}_{\text{lane}} = \{S_1^{(l)}, \cdots, S_m^{(l)}\}$ and pavement panel segmentation masks $\mathcal{S}_{\text{panel}} = \{S_1^{(p)}, \cdots, S_n^{(p)}\}$.

To train the lane and pavement panel models, we use a loss function combines four components: bounding box regression loss $\mathcal{L}_{box}$, segmentation loss $\mathcal{L}_{seg}$, classification loss $\mathcal{L}_{cls}$, and distribution focal loss $\mathcal{L}_{dfl}$.

The overall loss function can be elaborated as,

$$\mathcal{L} = \lambda_{box} \sum_{i \in F_G} \mathbb{1}_{\{i \in F_G\}} \mathcal{L}_{\text{CIoU}} \left(B_i, \hat{B}_i \right) + \frac{\lambda_{dfl}}{N} \sum_{i=1}^{N} \sum_{j=1}^{4} \left(-w_{ij} \log \hat{w}_{ij} \right) + \frac{\lambda_{seg}}{N} \sum_{i=1}^{N} \left[\mathcal{L}_{BCE}(M_i, \hat{M}_i) + \mathcal{L}_{Dice}(M_i, \hat{M}_i) \right] + \frac{\lambda_{cls}}{N} \mathcal{L}_{BCE}(y_i, \hat{y}_i) \quad (1)$$

where,

- B_i is the ground truth bounding box,
- $\hat{B}_i$ is the predicted bounding box,
- $\mathbb{1}_{\{i \in F_G\}}$ is an indicator function that is 1 if object i is a foreground sample and 0 otherwise,
- F_G refers to the foreground matched objects,
- N is the number of cells containing an object,
- M_i corresponds to the ground truth segmentation mask for an instance i,
- $\hat{M}_i$ is the predicted segmentation mask,
- y_i is the ground truth class probability,
- $\hat{y}_i$ is the predicted class probability,
- w_{ij} represents the ground truth distribution of bounding box coordinates, and
- $\hat{w}_{ij}$ is the predicted probability distribution.

In this work, the values for the four weight component hyperparameters were set to $\lambda_{box} = 7.5$, $\lambda_{seg} = 7.5$, $\lambda_{cls} = 0.5$, and $\lambda_{dfl} = 1.5$, respectively. Additionally, the training process is influenced by several optimization hyperparameters. The initial learning rate was set to $\eta_0 = 0.01$, with a final learning rate of $\eta_f = 0.01\eta_0$. The momentum coefficient was $\beta = 0.937$, while the weight decay was applied with a factor of $\phi = 5 \times 10^{-4}$.

After 100 epochs, the validation losses for the five YOLO11 variants in the road lane segmentation task using the NoMURLane dataset [2] were $\mathcal{L}_{box} = 0.139 \pm 0.002$, $\mathcal{L}_{seg} = 0.589 \pm 0.017$, $\mathcal{L}_{cls} = 0.075 \pm 0.001$, and $\mathcal{L}_{dfl} = 0.72 \pm 0.02$. In comparison, the training losses using NoMURPanel dataset [3] were $\mathcal{L}_{box} = 0.317 \pm 0.003$, $\mathcal{L}_{seg} = 0.744 \pm 0.06$, $\mathcal{L}_{cls} = 0.191 \pm 0.011$, and $\mathcal{L}_{dfl} = 0.805 \pm 0.014$.

Given an input image I_i, lane segmentation masks $\mathcal{S}_{\text{lane}}$, and pavement segmentation masks $\mathcal{S}_{\text{panel}}$, the post-processing involves bitwise operation between two masks to construct refined pavement panel polygons and extract corresponding image patches for classification.

As presented in Algorithm 1, the process first converts segmentation masks into quadrilateral polygons using extremal coordinates. Each pavement panel is assigned to a lane by minimizing the perpendicular distance to lane boundary curves. The resulting panel groups are vertically ordered to form a structured grid. Missing panels are reconstructed iteratively using neighboring polygon vertices until no further updates are possible. Each polygon is then refined by intersecting its edges with the corresponding lane boundaries, which obtains geometrically consistent panel regions. Finally, the refined polygons are rasterized to extract cropped image patches $\{x_k\}$ for downstream classification.

Algorithm 1 Refine Pavement Panel Masks (Filter Non-Lane Panels and Find Missing Panels)

Require: I_i, $\mathcal{S}_{\text{lane}}$, $\mathcal{S}_{\text{panel}}$
Ensure: Cropped panel images $\{x_k\}$

$\mathcal{B}_{\text{lane}} \leftarrow$ polygons from points $(x_{tl}, y_{tl}), (x_{bl}, y_{bl}), (x_{br}, y_{br}), (x_{tr}, y_{tr})$ of each mask in $\mathcal{S}_{\text{lane}}$
$\mathcal{B}_{\text{panel}} \leftarrow$ polygons formed analogously from $\mathcal{S}_{\text{panel}}$
$\mathcal{C}_{\text{lane}} \leftarrow \{(a, b, c) \mid a = y_{bl} - y_{tl},\ b = x_{tl} - x_{bl},\ c = x_{bl}y_{tl} - x_{tl}y_{bl},\ b_i \in \mathcal{B}_{\text{lane}}\}$
Assign each $b_i \in \mathcal{B}_{\text{panel}}$ to lane index

$$j = \arg\min_{\ell} \frac{|a_\ell x_{tl} + b_\ell y_{tl} + c_\ell|}{\sqrt{a_\ell^2 + b_\ell^2}}$$

$\mathcal{P}_{\text{panel}} \leftarrow \{\text{sort}_y(\{b_i \mid \text{lane}(b_i) = j\})\}$
Iteratively fill missing panels:
while $\exists\, \mathcal{P}_{i,j} = \mathbf{0}$ **do**
    Collect neighboring vertices T_L, B_L, B_R, T_R from valid adjacent panels
    if $|T_L| = |B_L| = |B_R| = |T_R| = 1$ **then**
        $\mathcal{P}_{i,j} \leftarrow [T_L, B_L, B_R, T_R]$
    end if
end while
Refine polygons using lane boundaries:
for each $p \in \mathcal{P}_{\text{panel}}$ **do**
    Form line sets $\mathcal{C}_L^{(l)}, \mathcal{C}_R^{(l)}$ (lane boundaries) and $\mathcal{C}_T^{(p)}, \mathcal{C}_B^{(p)}$ (panel edges)
    Compute vertices via pairwise intersections:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}^{-1} \begin{bmatrix} -c_1 \\ -c_2 \end{bmatrix}$$

    Construct refined polygon p^*
end for
Rasterize polygons and extract patches:
for each $p_k^* \in \mathcal{P}_{panel}^*$ **do**
    Initialize mask $M \leftarrow \mathbf{0}$ with size of I_i
    for each scanline y between $\min_y(p_k^*)$ and $\max_y(p_k^*)$ **do**
        Compute intersection points $\{x_m\}$ of scanline y with polygon edges
        Sort $\{x_m\}$ in ascending order
        for $m = 1$ to $|\{x_m\}|$ step 2 **do**
            Fill pixels $M(x, y) = 1$ for $x \in [x_m, x_{m+1}]$
        end for
    end for
    $x_k \leftarrow I_i \odot M$
end for
return $\{x_k\}$

2.3 Pavement Distress Classification

In our study on pavement distress classification, we modified the Vision Transformer (ViT-Base) model in [4] to categorize pavement conditions into four

classes: *non-distressed*, *sealed patch*, *sealed crack*, and *thin crack*. The ViT-Base model consists of 12 Transformer layers, each with 12 self-attention heads, a hidden size D of 768, and an MLP size of 3072. The input image x_k obtained from the refined pavement panel polygons is resized into a 224×224 RGB image. It is then divided into 16×16 patches, forming a sequence of $N = \frac{HW}{P^2} = \frac{224 \cdot 224}{16^2} = 196$ patches. Each patch is flattened into a vector of size $P^2 \cdot C$ and projected into a 768-dimensional latent space through a trainable linear projection $\mathbf{E} \in \mathbb{R}^{(P^2 \cdot C) \times D}$. We added a single linear layer to the MLP in the transformer's head during fine-tuning to learn our four pavement distress types.

We used the cross-entropy loss for multiclass classification to quantify the discrepancy between the predicted logits and the true class label. The formulation combines softmax and negative log-likelihood. Given logits $\mathbf{x} = [x_1, x_2, \cdots, x_C]$ and the correct class index y, the loss is defined as,

$$\mathcal{L}(\mathbf{x}, y) = -x_y + \log \sum_{c=1}^{C} e^{x_c} \tag{2}$$

Network training on the NoMURDistress dataset [1] terminated after 50 epochs, as the patience-based stopping criterion was met. As a result, our trained ViT model achieved a best validation loss of 0.0575.

3 Results and Discussion

3.1 Precision of Parallel Predictors

Stage 1 of the proposed two-stage network consists of two parallel segmentation predictors. The performance of these trained models affects the overall robustness of the pipeline. To assess their capability in generating segmentation masks for road lanes and pavement panels, the first ablation study focused on evaluating model performance using mean Average Precision (mAP). This was computed by averaging the Intersection over Union (IoU) thresholds ranging from 0.50 to 0.95 in increments of 0.05. Its goal is to provide a comprehensive measure of detection and localization consistency under varying levels of strictness.

As shown in Fig. 2, the road lane segmentation model achieved an mAP of approximately 0.95 among five YOLO11 variants (YOLO11n, YOLO11s, YOLO11m, YOLO11l, and YOLO11x), while the pavement panel segmentation model attained a slightly lower mAP of around 0.88. Model performance remained consistent across different export formats—ONNX and OpenVINO—indicating minimal impact from model conversion. The latency of originally trained PyTorch models and the exported ONNX models were evaluated using CUDA on an NVIDIA GeForce GTX 1050 Ti GPU, whereas the OpenVINO model was tested on an Intel Core i7-7700HQ CPU.

The pavement panel segmentation predictor yielded a noticeably lower mAP than the road lane predictor. It indicates the performance limitations that warrant further analysis. To explore this, the second experiment analyzed missed detections. As shown in the left plot of Fig. 3, the road lane segmentation model

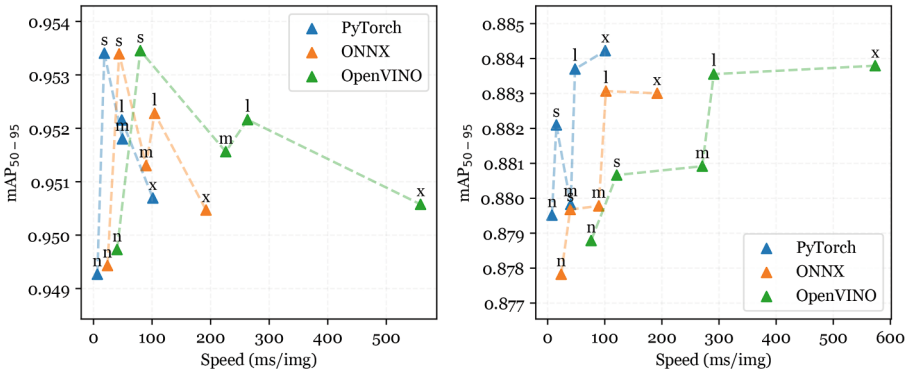

Fig. 2. Mean Average Precisions (@50-95 IoUs) of various Model Exports for (*left*) road lane segmentation and (*right*) pavement panel segmentation.

maintained consistently low false negative rates across various IoU thresholds. It demonstrates its strong ability to identify road lanes in the test images. In contrast, the pavement panel segmentation model exhibited significantly higher false negative rates, particularly at IoU thresholds above 0.7.

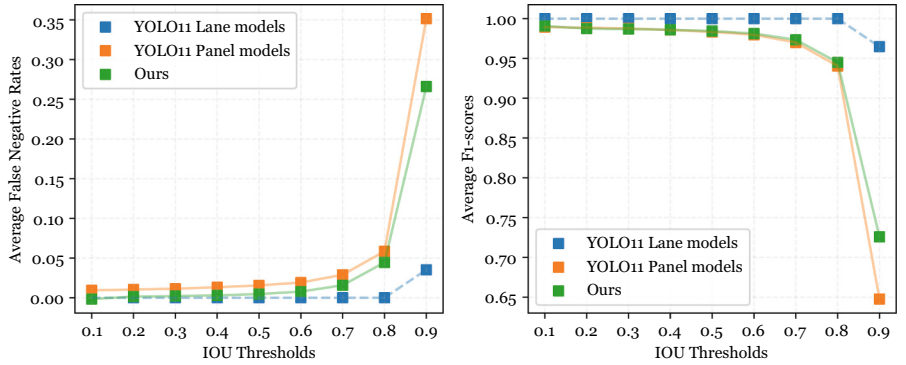

Fig. 3. Recall-sensitive metrics for pavement panel detection—(*left*) average false negative rates and (*right*) F1 scores.

To mitigate this deficiency, the proposed refining technique was introduced. The results show that it substantially reduced the false negative rate compared to the baseline pavement panel model. Additionally, the right plot in Fig. 3 presents the corresponding F1 scores—the harmonic mean of precision and recall—which further confirm the module's effectiveness. The improvement in F1 scores indicates that the refining technique improved both precision and recall.

3.2 Pavement Distress Classification Report

The final ablation study investigates whether the two-stage architecture performs better than directly using the trained YOLO11 pavement panel segmentation models. It also assesses the consistency of classification results between unrefined and refined cropped pavement panel images in the set $\{x_k\}$. In the conventional method, the YOLO11 model segments the panels, and a Vision Transformer classifies the resulting patches, forming the `YOLO11Panel+ViT approach`. We use the `YOLO11n` pre-trained variant for pavement panel segmentation in both methods and for road lane segmentation in the two-stage framework.

The one-vs-rest precision-recall curve analysis compares the performance of the `YOLO11Panel+ViT` approach with the proposed pipeline. As shown in Fig. 4, the pipeline achieves higher average precision (AP) scores for both the *sealed-crack* and *thincrack* classes. The mean Average Precision among the four classes is 0.9775±0.0015. This signifies that the proposed two-stage pipeline improved the classification performance.

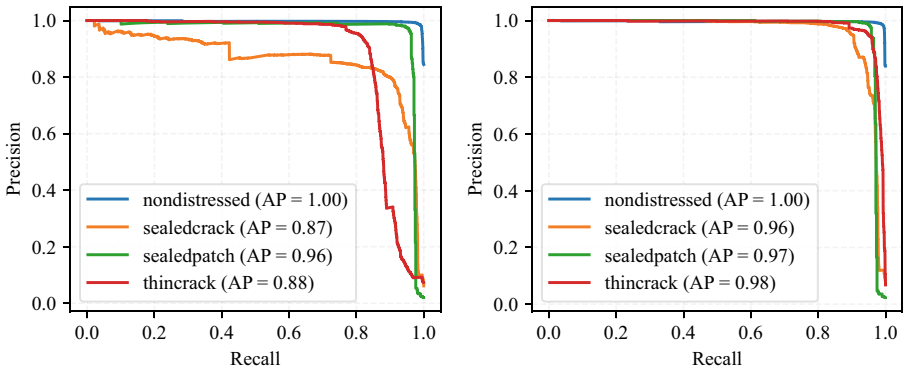

Fig. 4. One-vs.-rest precisionâĂŞrecall curve comparison between (*left*) the `YOLO11Panel+ViT` approach and (*right*) our two-stage architecture.

Figure 5 presents a qualitative demonstration of correct classification results obtained from the proposed pipeline. It can be observed that the extracted pavement panels exhibit irregular geometries, where several instances deviate from ideal rectangular structures with balanced sides. Despite these unstructured shapes, the proposed architecture can delineate pavement panels and can classify the corresponding distress types.

However, the presence of occlusions, particularly from vehicles obstructing the pavement panels, introduces noticeable degradation in classification performance, as evidenced by several prediction samples. This suggests the need for an additional annotation category to explicitly model such cases; specifically, panels occluded by vehicles should be labeled as a distinct vehicle presence class. A further limitation is its inability to adopt to curved road segments. This is

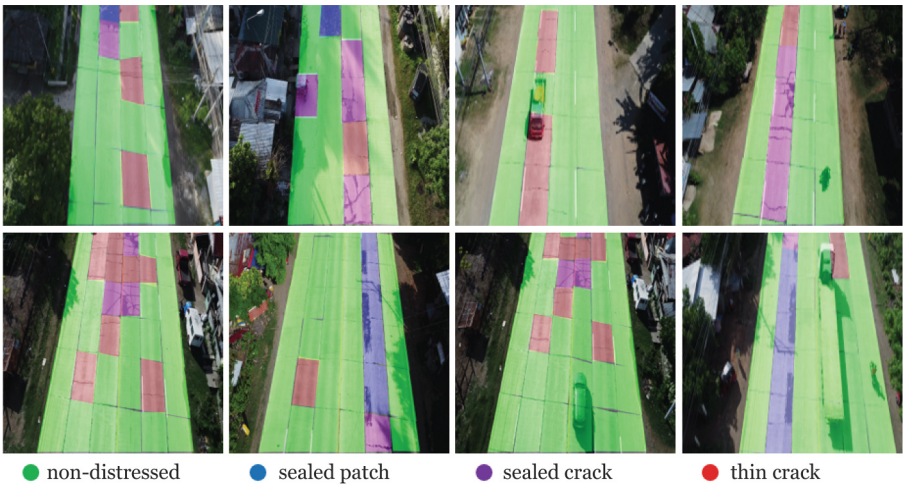

Fig. 5. Qualitative demonstration of correct classification results from our two-stage architecture.

primarily due to the constraints of the refinement procedure, which relies on linear interpolation.

4 Conclusion

This study presents a two-stage architecture designed to detect pavement distresses in aerial images: parallel YOLO-based predictors with refining technique, and a pavement distress classifier. Experimental results show that producing refined pavement panel mask reduces false positive/negatives and improves overall classification performance. The future direction will involve exploring additional distress types found in concrete, asphalt, or hybrid road surfaces. Conducting extensive experiments on these new distress categories will be necessary in the next phase of research.

References

1. Aleluya, E.R., et al.: NoMURDistress (2025). https://doi.org/10.7910/DVN/HUEFOS
2. Aleluya, E.R., et al.: NoMURLane (2025). https://doi.org/10.7910/DVN/BAMYR1
3. Aleluya, E.R., et al.: NoMURPanel (2025). https://doi.org/10.7910/DVN/PSFTEH
4. Dosovitskiy, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (2021). https://doi.org/10.48550/arXiv.2010.11929
5. Du, Y., Pan, N., Xu, Z., Deng, F., Shen, Y., Kang, H.: Pavement distress detection and classification based on YOLO network. Int. J. Pavement Eng. **22**, 1–14 (2020). https://doi.org/10.1080/10298436.2020.1714047

6. Guan, S., Liu, H., Pourreza, H.R., Mahyar, H.: Deep learning approaches in pavement distress identification: A review. arXiv preprint arXiv:2308.00828 (2023). https://doi.org/10.48550/arXiv.2308.00828
7. Guo, F., Liu, J., Xie, Q., Yu, H.: A two-stage framework for pixel-level pavement surface crack detection. Eng. Appl. Artif. Intell. **133**, 108312 (2024). https://doi.org/10.1016/j.engappai.2024.108312
8. Jiang, Y., Pang, D., Li, C., Yu, Y., Cao, Y.: Two-step deep learning approach for pavement crack damage detection and segmentation. Int. J. Pavement Eng. **24**(2), 2065488 (2023). https://doi.org/10.1080/10298436.2022.2065488
9. Kothai, R., Prabakaran, N., Srinivasa Murthy, Y.V., Reddy Cenkeramaddi, L., Kakani, V.: Pavement distress detection, classification, and analysis using machine learning algorithms: a survey. IEEE Access **12**, 126943–126960 (2024). https://doi.org/10.1109/ACCESS.2024.3455093
10. Liu, J., et al.: Automated pavement crack detection and segmentation based on two-step convolutional neural network. Comput.-Aided Civil Infrastruct. Eng. **35**(11), 1291–1305 (2020). https://doi.org/10.1111/mice.12622
11. Lv, Z., Hao, Z., Zhu, Y., Lu, C.: A review on automated detection and identification algorithms for highway pavement distress. Appl. Sci. **15**(11) (2025). https://doi.org/10.3390/app15116112
12. Maluya, M.M.O., et al.: Rural Road Pavement Panel Detection with Unmanned Aerial Vehicles. In: 2023 IEEE 15th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), pp. 1–6 (2023). https://doi.org/10.1109/HNICEM60674.2023.10589124
13. Maluya, M.M.O., Aleluya, E.R.M., Opon, J.G., Salaan, C.J.O.: Lane tracking for autonomous road pavement inspection with unmanned aerial vehicles. In: 2023 6th International Conference on Applied Computational Intelligence in Information Systems (ACIIS), pp. 1–6 (2023). https://doi.org/10.1109/ACIIS59385.2023.10367228
14. Maluya, M.M.O., Aleluya, E.R.M., Salaan, C.J.O.: Flexible and rigid pavement binary classifier using vision transformer. In: 2024 IEEE 16th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), pp. 1–6 (2024). https://doi.org/10.1109/HNICEM64917.2024.11258762
15. Pan, Y., Zhang, X., Cervone, G., Yang, L.: Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery. IEEE J. Selected Top. Appl. Earth Observations Rem. Sens. **11**(10), 3701–3712 (2018). https://doi.org/10.1109/JSTARS.2018.2865528
16. Samadzadegan, F., Dadrass Javan, F., Ashtari Mahini, F., Gholamshahi, M., Nex, F.: Automatic road pavement distress recognition using deep learning networks from unmanned aerial imagery. Drones **8**(6) (2024). https://doi.org/10.3390/drones8060244
17. Silva, L.A., Sanchez San Blas, H., Peral García, D., Sales Mendes, A., Villarubia González, G.: An architectural multi-agent system for a pavement monitoring system with pothole recognition in UAV images. Sensors **20**(21) (2020). https://doi.org/10.3390/s20216205
18. Song, Q., Liu, L., Lu, N., Zhang, Y., Muniyandi, R.C., An, Y.: A three-stage pavement image crack detection framework with positive sample augmentation. Eng. Appl. Artif. Intell. **129**, 107624 (2024). https://doi.org/10.1016/j.engappai.2023.107624

19. Tran, T.S., Tran, V.P., Lee, H.J., Flores, J.M., Le, V.P.: A two-step sequential automated crack detection and severity classification process for asphalt pavements. Int. J. Pavement Eng. **23**(6), 2019–2033 (2022). https://doi.org/10.1080/10298436.2020.1836561
20. Wang, A., Lang, H., Chen, Z., Peng, Y., Ding, S., Lu, J.J.: The two-step method of pavement pothole and raveling detection and segmentation based on deep learning. IEEE Trans. Intell. Transp. Syst. **25**(6), 5402–5417 (2024). https://doi.org/10.1109/TITS.2023.3340340
21. Yan, H., Zhang, J.: UAV-PDD2023: a benchmark dataset for pavement distress detection based on UAV images. Data Brief **51**, 109692 (2023). https://doi.org/10.1016/j.dib.2023.109692
22. Yang, Z., Ni, C., Li, L., Luo, W., Qin, Y.: Three-stage pavement crack localization and segmentation algorithm based on digital image processing and deep learning techniques. Sensors **22**(21) (2022). https://doi.org/10.3390/s22218459
23. Zhang, J., Xia, H., Li, P., Zhang, K., Hong, W., Guo, R.: A pavement crack detection method via deep learning and a binocular-vision-based unmanned aerial vehicle. Appl. Sci. **14**(5) (2024). https://doi.org/10.3390/app14051778
24. Zhang, Y., Chen, J., Wu, Z., Guo, X., Jia, S.: Optimizing pavement distress detection with UAV: a comparative study of vision transformer and convolutional neural networks. KSCE J. Civil Eng. 100095 (2024). https://doi.org/10.1016/j.kscej.2024.100095
25. Zhao, K., Xu, S., Loney, J., Visentin, A., Li, Z.: Road pavement health monitoring system using smartphone sensing with a two-stage machine learning model. Autom. Constr. **167**, 105664 (2024). https://doi.org/10.1016/j.autcon.2024.105664
26. Zhu, J., Zhong, J., Ma, T., Huang, X., Zhang, W., Zhou, Y.: Pavement distress detection using convolutional neural networks with images captured via UAV. Autom. Constr. **133**, 103991 (2022). https://doi.org/10.1016/j.autcon.2021.103991