Multi-Sensor Fusion for Proactive Anomaly Detection in Robot Navigation

[1] DIKSHA M , [2] CHAITHRASHREE A

[1]Assistant Professor, [2]Assiatant Professor
[1] Computer Science & Engineering, [2] Computer Science & Engineering
[1]Brindavan College of Engineering, Bangalore, India
[1]diksha.m364@gmail.com, [2]chaithrashree.a@gmail.com

Abstract : Mobile robots frequently exhibit abnormal behaviors that can impair navigation despite the rapid progress of navigation algorithms. Modern robots need to be able to recognize these unusual behaviors in order to reach high levels of autonomy. Methods for reactive anomaly detection Detect anomalies poor task executions based on the current state of the robot and thus lack the capability to warn the robot before a malfunction actually happens. Due to the possibility of harm to the robot and the environment, such a warning delay is undesirable. For robot navigation in unstructured and uncertain situations, we suggest a proactive anomaly detection network (PAAD). Based on the anticipated movements from the predictive controller and the current observations from the perception module, PAAD forecasts the likelihood of future failure. Effective fusion of multi-sensor signals to provide reliable anomaly detection as seen in the field when there is sensor occlusion environments. Our tests on data from field robots show that our model can catch abnormal actions in real-time while retaining a low false detection rate in congested areas, outperforming earlier methods in failure identification

***IndexTerms* - Face recognition, bias, fairness, soft-biometrics, analysis, privacy, biometrics**

I. INTRODUCTION

Our tests on data from field robots demonstrate that our model outperforms earlier methods in failure identification by detecting aberrant activities in real-time while maintaining a low false detection rate in crowded regions. Robots may still have trouble in real-world situations due to the complexity of the surroundings, the variation of the terrain, and the unreliability of their sensors, despite recent research that has significantly improved trustworthy autonomy for robot navigation [2]–[5]. A lack of a detection system for Anomaly behaviors prior to failures could result in collisions that harm robots and plants. The robot can be prevented from entering failure modes by the detection of such anomalous actions, opening up opportunities for recovery maneuvers to be used and the mission to be completed.

Deep learning-based anomaly detection (AD) techniques have been widely adopted in robotic applications [6]. Many early studies approached the AD problem in a reactive manner. These reactive anomaly detectors can only draw conclusions based on sensory data that is currently available (such as velocity, torque, and LiDAR measurements), hence they are unable to anticipate potential faults in the future.

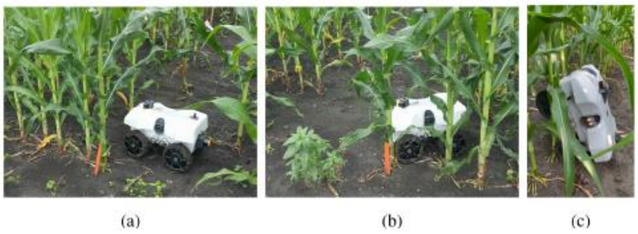


Fig. 1. Field robot platform. (a) The field robot, TerraSentia, navigates between rows of crops under cluttered canopies to collect data on plant traits. (b) The robot fails the navigation task due to anomalous behaviors. (c) Rare but catastrophic failures may occur in extreme cases.

Due to the alert delay, the robot may as a result continue to sustain damage from collisions (Fig. 1(b)) or enter critical states (Fig. 1(c)), the recovery from which is beyond the autonomy of the robot. Proactive anomaly detection is an alternate approach that forecasts the likelihood of future failure based on the planned actions and the current sensory observation. LaND [11] and BADGR [12] have both investigated this type of prediction model to determine the best course of action for outdoor navigation. However, when implementing autonomous systems in typical outdoor contexts, the AD problem for robot navigation across natural field environments creates difficulties that are typically not taken into account. First, during operation, both the perception system and the control system display high levels of uncertainty. The sensory signals produced by weeds, lodged plants, and low-hanging leaves are typically loud and obscure useful information for AD tasks (such as the robot's relative position in relation to the crop rows) (Fig. 2(a)).

While this is going on, the robot's movements (such as its linear and angular velocities) are constantly distorted by the changing wheel-terrain, interactions [2], which can be problematic for pattern recognition by introducing significant volatility in control signals. Second, the regular sensor occlusion makes it difficult for the robot to perceive its surroundings (Fig. 2(b)). Due to the lack of a robust perception system, anomaly detectors that rely on a single sensor modality [7], [11], or [12] are susceptible to being easily tricked. We take a proactive approach to the problem of anomaly detection by identifying aberrant behaviors based on recent observations. Formally, we define an anomalous navigational behavior for a robot as a sequence of future motions that includes at least one time step with failure within the prediction horizon. A series of control actions or a predetermined path can be used to represent this future motion.

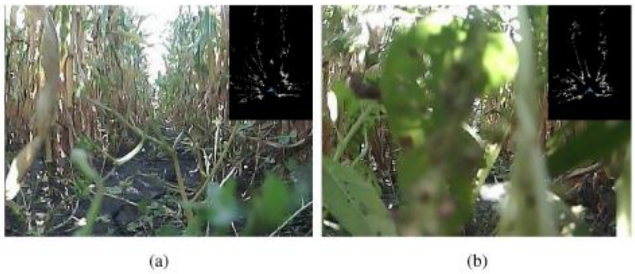

(a) (b)

Fig. 2. Field environment. The robot perceives the environment through a forward-facing camera and a 2D LiDAR. The blue triangle in the 2D point cloud denotes the robot. Weeds and low-hanging leaves introduce high uncertainty in sensory signals and can block the sensor view as the robot navigates under canopy.

We develop a Proactive Anomaly Detection network, or PAAD, that uses the predicted motions from the predictive controller and the most recent observations from the perception system to reason about the likelihood of failure at each time step within the future time horizon. The final likelihood of failure is produced by independently extracting and then fusing features from several modalities twice. To enhance generalization capacity and increase robustness against noisy sensory signals, we train PAAD with a mixed cost function that consists of a prediction job and a reconstruction task.

The following is a summary of our contributions:

1) For robust perception in unstructured and uncertain situations, we present a unique deep neural network architecture termed PAAD that successfully fuses multi-sensor signals.
2) To perform proactive anomaly detection and to enable effective feature extraction from noisy signals, we use a low-variance picture representation of intended motions rather than raw control actions
3) In a real-time test, our suggested detector captures abnormal behavior while maintaining a low false detection rate, outperforming existing approaches in failure identification performance on an offline real-world navigation dataset.

II. RELATED WORK:

A significant issue that has been researched across a variety of academic fields and application domains is anomaly detection, commonly referred to as outlier detection or novelty discovery [6], [13]. In robotics, AD has been utilized to identify manipulation task and navigation task failures [14, 15]. Learning-based AD algorithms have come a long way thanks to recent research efforts. The encoder-decoder technique for multi-sensor anomaly detection (EncDec-AD) introduced by Maalhotra et al. employs reconstruction error to identify abnormalities [10]. A multimodal LSTM-based variational autoencoder (LSTM-VAE) that combines sensory signals and reconstructs their predicted distribution is the idea put forth by Park et al. Then, abnormalities are found using a reconstruction-based anomaly score [8]. In our earlier research, the AD problem is framed as a multi-class classification problem, and the supervised variational autoencoder model (SVAE) is suggested [7]. However, these reactive strategies are unable to identify unusual behaviors in advance of failures, which means that safety is not necessarily increased [17].

The predictive model for upcoming navigational events (like collisions) proposed in LaND [11] and BADGR [12] is the piece of work that most closely resembles PAAD in the proactive anomaly detection/predictive collision avoidance field. The neural network forecasts the likelihood of a collision for each time step inside the prediction horizon using an image and a series of future control actions as input. It has been demonstrated that the model performs consistently well at detecting anomalies in sidewalk and off-road areas with lots of open space. The unimodal input, however, causes sensor occlusion, and the input uncertainty makes it difficult for the network to acquire valuable features. In this study, we combine camera and LiDAR data to enhance the robot's perception abilities, and we replace noisy control actions with an image representation of the planned path to speed up model training. Traversability analysis in unstructured environments is another actively investigated area of research that is pertinent to our work.

The topic of evaluating how difficult it will be for a ground vehicle to navigate a terrain is known as terrain traversability analysis [18]. In order to predict vibrations using only picture texture data, Bekhti et al. train a Gaussian process regressor using terrain photos and acceleration signals [19]. With a 2.5D grid map centered on the vehicle frame and containing both geometry and semantic information about the surroundings, Mat- urana et al.'s real-time mapping technique is proposed [20]. Traversing a traversable terrain does not necessarily mean that the robot behavior is not anomalous, despite the methodologies of traversability estimation and anomaly detection being identical. However, such behavior should be classified as an anomaly as the robot is deviating from the specified navigation task. In field environments, for example, a trajectory that drives off the trail from one to another due to large gaps between crops can be collision free and incur no additional traversal cost.

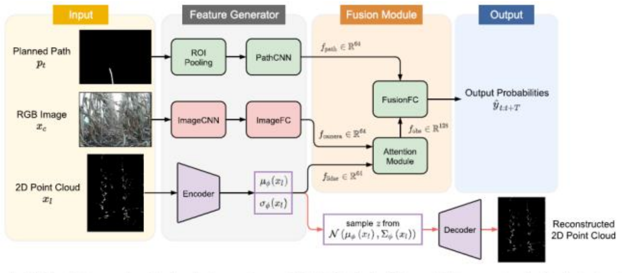


Fig. 3. Model architecture overview. The planned path, camera image and LiDAR point cloud are fed to parallel feature generators, then fused in two stages to learn the probability of failure within the prediction horizon. The processing pipelines for the planned path, camera image, and LiDAR point cloud are in green, red, and purple, respectively. PAAD in test time is highlighted in background colors while red arrows are used for training only.

The camera-lidar fusion is a developing research theme that has been used in numerous robotics and autonomous driving projects [21]. Depth completion [22], object detection [23], object tracking [24], and simultaneous localization and mapping (SLAM) [25] are examples of common uses. However, The camera-lidar fusion is a developing research theme that has been used in numerous robotics and autonomous driving projects [21]. Depth completion [22], object detection [23], object tracking [24], and simultaneous localization and mapping (SLAM) [25] are examples of common uses. However, one typical assumption that these application domains mak is that the camera and LiDAR data are consistent (i.e., the perceived worlds from the two modalities can be matched with each other). Such an assumption is false in a field setting since one of the sensors can frequently become obscured, which makes it more difficult to use earlier methods. The perception error caused by sensor occlusion is frequently treated as noise in agricultural settings and is handled by the Kalman filter [4], [5]. Although the issue of occlusion can be somewhat overcome by such filtering techniques, the necessary premise—for example, that the center line is unobstructed—does not always hold true in practice. To tackle sensor occlusion in congested surroundings, we create a novel sensor fusion approach in this work.

III. METHOD

Our objective is to create an AD module that will allow a mobile robot to recognize unusual behaviors while navigating in the field. We suppose that the sensor observations at time t, ot, are multi-modal and comprise range data from a 2D LiDAR and an RGB picture from a camera, represented by xc RH W C. A predictive controller is used by the robot to prepare a series of actions for the following T time steps in a receding horizon fashion.The current intended path is further retrieved from future actions, and the resulting path is represented as a distinct picture (pt, RH, W1) with the path projected onto a clear front-view image plane.The job for the AD module is to map, for each time step along the path as indicated in Fig. 3, from a collection of current sensor observations and a planned path (ot, pt) RHW C RL RHW 1 to a sequence of navigation failure prob- abilities.

PAAD may leverage the modality from the planning module to detect abnormal behaviors, which is superior to the current reactive anomaly detection method. Such a proactive nature of PAAD alerts the robot before entering critical states from which human interventions are required to recover the robot. Furthermore, resistance against uncertainty and sensor occlusion in complicated field situations is provided by the efficient merging of multi-modal perception signals. Contrarily, in anomaly detectors that utilise unimodal sensory signals, incorrect detection of an anomaly can be commonly generated by camera occlusion [11], [12]. Last but not least, the adopted visual representation of the planned path has less variance than the raw control actions, making the training process more effective.

3.1 Data Collection

The TerraSentia robot is a remarkably small, four-wheeled, skid-steering mobile robot that can be used for automated phenotyping of crops [2]. A forward-facing monocular camera sensor (OV2710) and a 2D horizontal-scanning LiDAR (Hokuyu UST-10LX) with a 270-degree field of view and 0.25-degree angular resolution are both installed on the robot. An RGB picture measuring 240 by 320 pixels and a vector of 1081-dimensional LiDAR ranges make up the observation. Perspective projection is used to create an image representation of the intended path pt from the onboard model predictive controller's output. The robot is either in a normal state or fails the navigation task, according to the ground truth probability of failure yt.

The robot uses an autonomous control policy during data gathering; in this case, it is the LiDAR-based navigation algorithm for agricultural mobile robots [26]. The human deactivates the autonomy once the robot enters a failure mode, moves it back to the center line, and then activates it again. We define the failure mode as any condition that the robot enters and cannot leave without human assistance in order to complete the specified navigation job (such as following a crop row). At each time step t, the robot gathers the observations, planned trajectories, and drive modes (ot, pt, yt). We point out that PAAD does not demand for any information beyond what is normally gathered to assess robot autonomy. In actuality, the data gathering procedure outlined above is intended to demonstrate LiDAR-based autonomy for agricultural robots rather than PAAD [4], [26].

3.2 Model Architecture

We refer to PAAD as a function g: (ot, pt) yt:t+T that receives as input a collection of current observations and a planned path (ot, pt) and produces a series of failure probability. within the predicted horizon, yt:t+T. In Fig. 3, the network structure is displayed. For each modality, separate feature generators (FGs) are created to separate out reliable features from various inputs. We use feature-level camera-lidar fusion rather than signal-level fusion, which can struggle with inconsistent perception signals caused by frequent blockage of one of the sensors, to improve the perception capabilities in harsh and congested agricultural settings. The final result is an assessment of the likelihoods of navigation failure along the intended course over the following T time steps based on the most recent observations. The planned path and RGB image are processed by two distinct convolutional pipelines to provide, respectively, camera features fcamera and path features fpath.
Following each CNN module comes a flattening process. We crop the path image based on the area of interest (ROI), preventing the model from receiving unnecessary information that does not pertain to the path. We adapt the concept from SVAEs [7] to extract features from the LiDAR point cloud. As a regularization [27]–[29], the reconstruction task in the LiDAR pipeline drives the encoder to learn representative characteristics of high-dimensional LiDAR data that are essential to both the downstream inference and generative model. The model tends to perform better at generalization on the inference task when the reconstruction task is given more focus [7]. We approximate the latent variable z's posterior distribution as a Gaussian with variational parameters:

1) qφ(z|xl) = N (z | μφ(xl), diag(σφ(xl))), where μφ(xl) is a mean vector, σφ(xl) is a variance vector, and the nonlinear transformations μ ϕ : RL ——— → Rd and σ ϕ : RL ——— → Rd are parameterized by multilayer perceptrons (MLPs) in the encoder. For the downstream prediction task, we choose LiDAR features as: flidar = [μφ(xl), σφ(xl)]. For the reconstruction task,1 the decoder uses a generative model of the form: where MLP(z; θ) is a mean vector formed by

a nonlinear transformation of the latent variable z, and σ is a hyperparameter.Here, we choose the nonlinear transformation to be an MLP parameterized by θ. Note that the reconstruction branch in LiDAR pipeline follows the structure of a vanilla variational autoencoder (VAE).pθ(xl|z) = N (xl | MLP(z; θ), σ2 · I).

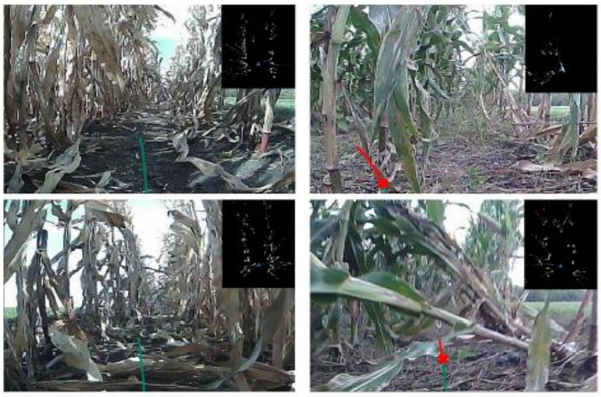
2) Fusion Module: To form observation features from sensors, we employ a feature-level camera-lidar fusion by using a multi-head attention (MHA) with a residual connection [30]: which corresponds to the attention module in Fig. 3. fobs = [fcamera, flidar] + MHA(Q, K, V = [fcamera, flidar]. The concatenation of the terms fcamera and flidar, which can be thought of as a sequence of length 2, is chosen as the question, key, and value in the same way.

For camera-lidar fusion, we prefer an MHA to an MLP because we anticipate the model to produce observation features dependent on the signal quality of each sensor. The point cloud should contribute more to observation features than the image, for instance, when the LiDAR view is clear but the camera is obscured by leaves. The estimated likelihood of failure in the following T time steps is produced by the final fusion of observation features and path features at time t: yt:t+T = Sigmoid (FusionFC ([fobs, fpath])). A sigmoid function is used to ensure that the final output probabilities are scaled into the valid range.

3.3 Training

Using a ResNet-18 backbone that has been trained on a visual navigation problem, the ImageCNN in camera pipeline can anticipate a robot's heading and placement in a crop row. Utilizing an RGB front-view picture [5]. We build the ImageCNN. Module by cutting off the visual navigation model just before completely interconnected layers. The ImageCNN's weights are after pretraining fixed Denoting the dataset collected in Section III-A by D, we specify the overall loss function for PAAD as:

L = Σ(ot,pt,yt:t+T)∈D
α · LBCE(g(ot, pt), yt:t+T)
− Eqφ(z|xl)[log pθ (xl|z)] + DKL[qφ(z|xl) ∥ pθ (z)], (6)



(a) Normal behavior (b) Anomalous behavior

Fig. 4. Samples from the collected dataset. The planned path is overlayed onto the image for visualization purposes. The ground truth probability of failure along the path is indicated by the color (green for normal and red for failure), and the red cross marks the start of a sequence of failures.

where LBCE is the binary cross-entropy loss, α is a hyper parameter controlling the relative weight between the discriminative and generative learning, and pθ (z) is a prior distribution over the latent variable z. As in SVAEs, we choose pθ (z) to be a standardGaussian distribution z ∼ N (0, I) There are two tasks in the training objective: a prediction task and a task of reconstruction. The prediction error is penalized by the first term in (6). We established = 0.1 N, where N represents the total number of data points in [7] format. the loss function's final two terms, which the lower bound on the evidence (ELBO) is likewise the negative Plain VAEs punish the LiDAR data reconstruction mistake.

The last KL divergence term can be viewed as a regularization. The inference model and the generative model can be optimized jointly by stochastic gradient descent of the overall objective function (6). To enable the backpropagation through the sampling layer within the network, a common reparameterization trick is used to move the sampling process to a stochastic input layer [31].

IV. EXPERIMENTAL ANALYSIS

In our tests, we assess the performance of PAAD's anomaly identification on 4.1 km of real-world navigation data that the TerraSentia robot collected in corn fields between September 2018 and October 2020. Under a congested canopy, the robot maneuvers between rows of crops without harming the vegetation. Depending on the surrounding circumstances, the robot may or may not fail during a run. The robot's reference speed is set to 0.6 m/s, and the distance between any two points on its intended path from the onboard MPC is 0.2 meters. We subsample the data to 3 Hz after data collection to align the ground truth failure probability with the expected one along the intended path. We employ a prediction horizon of T = 10 time steps (i.e., a lookahead distance of 1.8 meters) for all proactive anomaly detectors. Figure 4 shows a portion of our dataset 2 in visual form. We create the training set and test set from trials on independent days to reduce the unfavorable impact on the evaluation of various models introduced by the covariance of datapoints closely connected in time. The training set has 29292 datapoints and 2258 aberrant behaviors that were gathered over five days, whereas the test set has 6869 datapoints. datapoints and 689 unusual actions from the data collected over two more days. The information was gathered in part. At the Autonomous Farm of Illinois. We perform under-sampling of normal cases and over-sampling of anomalous cases on the training set to balance the learning of both types of behaviors while keeping the test set unchanged.

In tests utilizing PAAD, we build the PathCNN with three convolutional layers and the following filter parameters: stride 2, filter number 8, and filter size 16 32. A max pooling layer comes after each convolutional layer. One hidden layer with 64 hidden units is how the ImageFC is implemented. Similar to SVAEs, the LiDAR pipeline's encoder is built from one hidden layer and 128 hidden units, while the decoder has a similar construction to the encoder. We select a latent space with 32 dimensions (z R32). The MHA has eight attention heads and the FusionFC has two hidden layers with 128 T hidden units each in the fusion module. The network is trained using ReLU activation functions and an Adam optimizer with a constant learning rate of 0.0005.

4.1 Baselines and Numerical Evaluation

We evaluate the performance of the proposed method on the test set, along with the following baseline methods:

- CNN-LSTM: A convolutional recurrent deep neural network with an image-based, action-conditioned architecture that was first described in LaND [11] and BADGR [12]. Each of the future T control actions is progressively processed by an LSTM unit, which is initialized with image features produced by a backbone convolutional network, and the corresponding projected failure probability is output.
- A feedforward convolutional neural network processing an image and a robot's actions for behavior prediction, according to Cui et al. [32]. A multimodal fusion network for robot navigation in challenging situations is NMFNet [33].
- We take the two branches that process sensor observations using LiDAR data and 2D pictures to calculate future failure probabilities, and we swap out the 3D point cloud branch for an MLP that analyzes robot actions. The approaches listed above are cutting-edge, and to our knowledge, our study is the first to explore with sensor fusion of unprocessed camera and LiDAR data for proactive anomaly identification. For either unimodal perception tasks for anomaly detection tasks involving multimodal sensory signals or similar signals. For to make an accurate comparison, we use all of the backbone convolutional neural networks used in a variety of camera techniques.

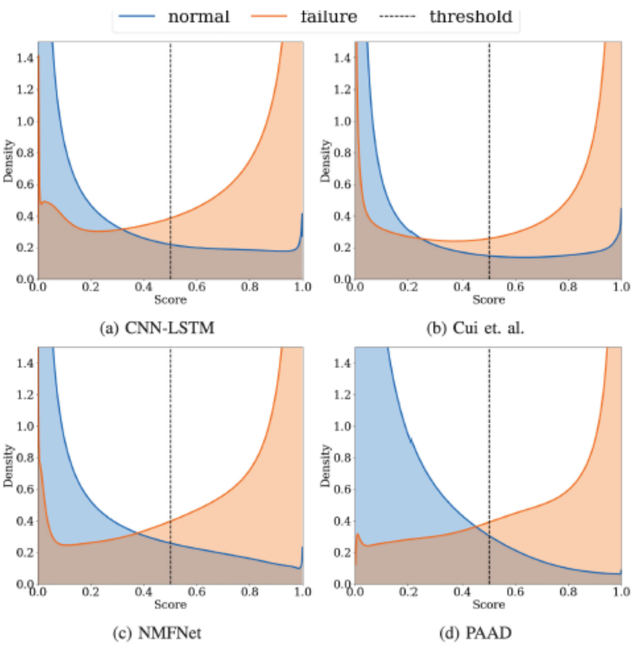
To our knowledge, our work is the first to experiment sensor fusion of raw camera and LiDAR data for proactive anomaly detection, and the above baselines are state-of-the-art methods for either anomaly detection tasks using unimodal perception signals or related tasks using multimodal perception signals. For a fair comparison, we implement all the backbone convolutional neural networks used across different methods for the camera image as the ResNet-18 pretrained on visual navigation task, as described in Section III-C. All methods are trained on the same.

Quantitatively, we compare different methods using the following two metrics:

- *F1-score:* A comprehensive threshold-dependent index considering precision P and recall R, which can be expressed as $2PR/(P + R)$. We set the threshold to be 0.5, i.e., we declare a navigation failure if the predicted probability of failure is greater than that of being "normal" at a point in time. Dataset.
- *PR-AUC:* A threshold-independent index indicating the area under the Precision-Recall Curve. PR-AUC describes the ability to distinguish between positive and negative samples for anomaly detection models.

TABLE I
ANOMALY DETECTION PERFORMANCE WITH DIFFERENT METHODS

Model	F1-score	PR-AUC
CNN-LSTM [11], [12]	0.5352	0.6988
Cui et. al. [32]	0.5748	0.7468
NMFNet [33]	0.5651	0.7554
PAAD (ours)	**0.6453**	**0.8281**



(a) CNN-LSTM

(b) Cui et. al.

(c) NMFNet

(d) PAAD

Fig. 5. Distribution of predicted probability of failure for normal and failure points. Upper parts of the pdfs are omitted. A clear separation of the two distributions is desired. An ideal anomaly detector should render the two pdfs as delta functions with impulses at 0 and 1 for normal and failure points, respectively.

fit probability density functions (pdfs) for normal and failure samples on the test set, respectively. We use a Gaussian kernel and apply the transformation trick [35] to make sure that the estimated pdfs have support on [0,1]. The results are presented in Table I and Fig. 5. As shown, PAAD achieves the best F1-score and highest PR-AUC with a large margin over other baselines. Although the CNN-LSTM model has been shown to have reliable anomaly detection performance for navigation tasks on sidewalks and off-road environments with large free space [11], [12], the method has not been shown to generalize well to harsh and cluttered field environments with limited open space.

We argue that this is due to the fact that the control actions in such uncertain environments are high variance, making the network struggle with identifying true anomalous actions from noises. In fact, all the three baselines, which take the future control actions as input, make overconfident predictions for false positives and false negatives as shown in Fig. 5. As a result, these three models in general show inferior F1-score and PR-AUC compared to PAAD, which makes use of the image representation of the planned path. Despite an additional sensor modality from LiDAR, NMFNet fails to provide a solid improvement over unimodal approaches, which highlights the importance of robust feature generator and fusion mechanism in highly uncertain environments.

Fig. 6 shows the anomaly detection results of different methods in several challenging scenarios. In the first row, the LiDARbased navigation algorithm falsely predict the orientation of the crop rows, making the robot take a left turn. As is further illustrated in Fig. 7, CNN-LSTM and NMFNet make the prediction of navigation failure merely based on the image without considering the future behavior, thus refusing to declare failures in such a clear image near the center line. Cui *et al.* [32] successfully detects a failure at the end of the path; however, the failure alert is too late to prevent the catastrophic collision. By contrast, the start time of the collision is more accurately

predicted by PAAD. The second row shows a near-miss case where the robot manages to recover to the center line from the edge. Although PAAD falsely predicts a failure at the last point with a score of 0.52, most part of the path is classified as normal correctly. However, all the other three methods generate overconfident scores for the entire path. The last row shows a normal case where the robot is tracking the center line while the camera is occluded by low-hanging leaves. The three baselines all failed while PAAD successfully distinguishes such normal behavior from an anomalous one.
To further verify our hypothesis that noisy actions, as opposed to planned paths, hinder the network from learning useful features of robot's behavior, we feed an image and several sequences of actions/paths sampled from the test set through different models to predict probability of failure within the horizon. As shown in Fig. 7(a), the three networks based on control actions always predict normal behaviors no matter how the future motion looks like, which indicates that the models are only making use of the image for anomaly detection. By contrast, PAAD can predict navigation failures based on the planned path, thus producing more promising results as shown in Fig. 7(b).

We further conduct an ablation study to reveal the benefit of different components in PAAD. The ablated versions of PAAD that we consider include: 1) *LiDAR only:* only the LiDAR pipeline is used to generate the observation features; 2) *camera only:* only the camera pipeline is used to generate the observation features; 3) *w/o MHA:* the residual MHA module is replaced with a simple MLP; 4) *w/o reconstruction:* the reconstruction branch in LiDAR pipeline is removed during training; 5) BEV: the planned path is projected to bird's eye view (BEV) instead of front view. The results are summarized in Table II. With an extra sensor modality, PAAD is able to correctly identify normal cases where either camera or LiDAR is occluded, which can otherwise be classified as anomalies by LiDAR-only or camera-only. Such strengthened perception capability results in a higher F1-score and higher PR-AUC. The ablation study on other key components indicates the importance of each design choice to the overall performance of PAAD.

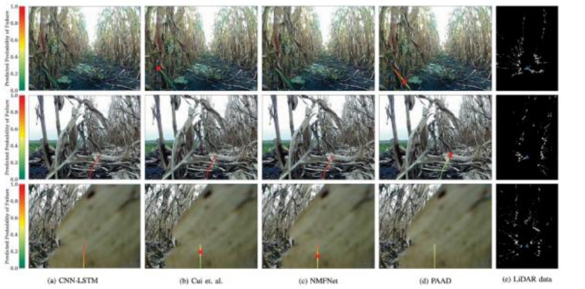


Fig. 6. Qualitative comparisons of different methods in challenging environments. CNN-LSTM, Cui *et al.*, and NMFNet are fed with planned control *actions*, and the path shown above is only for visualization purpose. The LiDAR data is only used by NMFNet and PAAD for prediction. The red cross on the path denotes the first point at which the predicted probability of failure is over 0.5.

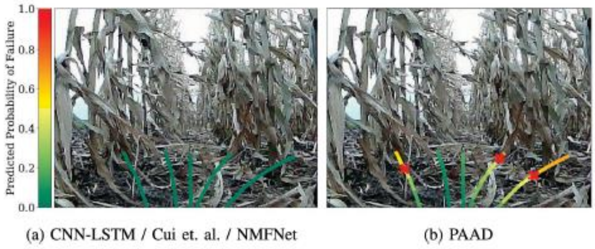

1.0
0.8
0.6
0.4
0.2
0.0
Predicted Probability of Failure
(a) CNN-LSTM / Cui et. al. / NMFNet
(b) PAAD

Fig. 7. Comparative study on AD performance with different sampled actions/paths. All the three baselines generate similar probabilities of failure along the sampled paths and thus are condensed in one figure.

TABLE II
ABLATION STUDY ON PAAD

Model	F1-score	PR-AUC
LiDAR only	0.5237	0.6738
camera only	0.5700	0.7784
w/o MHA	0.5941	0.7902
w/o reconstruction	0.6218	0.7997
BEV	0.6384	0.7965
PAAD	**0.6453**	**0.8281**

4.2 Real time Test

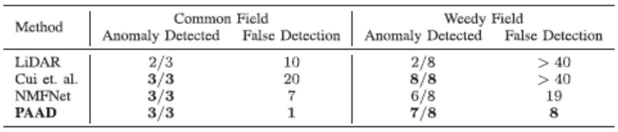
To test the ability of PAAD to alert the robot before executing an anomalous behavior, we further perform a real-time anomaly detection task on additional data.[3] In this experiment, the robot was driven by the vision-based navigation algorithm [5] on 1.3 km of field trails, consisting of 750 m of common field environment and 550 m of densely weedy environment. Three and eight human interventions were required to reset the robot after an anomaly occurred in common and weedy environment, respectively. We define the current *anomaly score* as a linear combination of probabilities of failure within the prediction horizon:

$$s_t = \beta \sum_{k=0}^{T-1} \gamma^k y_{t+k},$$

where γ is a discount factor compensating the uncertainty in the future, and β is a scaling factor ensuring that the summation $\sum_{k=0}^{T-1} y^k$ equals 1. At each time step t, we declare an anomaly if s_t is greater than 0.5. To calibrate the difficulty of the task, we implement a LiDAR baseline for the real-time test. Given range measurements within the forward-facing 90° field of view, we declare an anomaly if 85% of the viewis blocked by objectswithin 0.3meters.We also compare

PAAD against a unimodal approach, Cui *et al.* [32], and a multimodal approach, NMFNet [33], from Section IVA. To increase the robustness against frequent occlusions of camera and LiDARsensors in cluttered field environment, all the anomaly detectors declare an anomaly only when 3 consecutive anomaly scores are over 0.5. We implement all the methods at a frequency of 10 Hz. Table III summarizes the results. As shown, PAAD is able to detect anomalies reliably in both environments while maintaining a low false detection rate. On the contrary, the three baselines struggle with sensor occlusions and noisy actions in such cluttered and uncertain environments, thus frequently intervene the navigation system during the normal operation of the robot. Furthermore, we observe that PAAD is able to capture some rare failure modes, such as driving off the trail due to large gaps between crops. Scenarios in which PAAD failed usually contain dense weeds on the path and/or the robot executing near-miss maneuvers (see video). The detection of these anomalies could be potentially improved with additional data. Lastly, the reliable anomaly detection performance of the PAAD shown in the LiDAR-based navigation system (Section IV-A) and the vision-based navigation system (Section IV-B) indicate that our method is agnostic to the underlying controller and can be applied to general systems that employ predictive control.

TABLE III
REAL-TIME TEST RESULTS



Method	Common Field		Weedy Field	
	Anomaly Detected	False Detection	Anomaly Detected	False Detection
LiDAR	2/3	10	2/8	> 40
Cui et. al.	**3/3**	20	**8/8**	> 40
NMFNet	**3/3**	7	6/8	19
PAAD	**3/3**	**1**	**7/8**	**8**

V. CONCLUSION

In this work, we presented a proactive anomaly detection method for robot navigation in challenging field environment usingmulti-sensor signals. Our approach predicts the probability of future failure based on the planned path and the current sensor observation. By introducing a feature-level camera-lidar fusion, the detector successfully detected navigation failures in agricultural environment with higher F1-score and PR-AUC than other previous state-of-the-art methods. We also demonstrated the reliable anomaly detection performance of the PAAD with low false alarms in the real-time test. Although our method showed robustness in uncertain environments, false detection is unavoidable when both camera and LiDAR are blocked. Active perception, which encourages the robot to collect richer sensory signals through additional interaction with the environment, could decrease perception uncertainty in such cases of full sensor occlusion and would be a future work direction.

REFERENCE

[1] R. Xu, C. Li, and J. M. Velni, "Development of an autonomous ground robot for field high throughput phenotyping," IFAC-PapersOnLine, vol. 51, no. 17, pp. 70–74, 2018. [2] G. Balakrishnan, Y. Xiong, W. Xia, and P. Perona, "Towards causal benchmarking of bias in face analysis algorithms," in Proc. 16th Eur. Conf. Comput. Vis.,Glasgow, U.K., Aug. 2020, pp. 547–563

[2] Z. Zhang, E. Kayacan, B. Thompson, and G. Chowdhary, "High precision control and deep learning-based corn stand counting algorithms for agricultural robot," Auton. Robots, vol. 44, no. 7, pp. 1289–1302, 2020. [4] N. Almudhahka, M. S. Nixon, and J. S. Hare, "Human face identification via comparative soft biometrics," in Proc. IEEE Int. Conf. Identity Security Behav. Anal. 2016, pp. 1–6.

[3] E. Kayacan, Z.-Z. Zhang, and G. Chowdhary, "Embedded high precision control and corn stand counting algorithms for an ultra-compact 3D printed field robot," in Proc. Robotics: Sci. Syst., 2018. [6] F. Boutros, N. Damer, P. Terhörst, F. Kirchbuchner, and A. Kuijper, "Exploring the channels of multiple color spaces for age and gender estimation from face images," in Proc. 22th Int. Conf. Inf. Fusion (FUSION), Ottawa, ON, Canada, Jul. 2019, pp.1–8.

[4] A. E. B. Velasquez, V. A. H. Higuti, M. V. Gasparino, A. N. Sivakumar, M. Becker, and G. Chowdhary, "Multi-sensor fusion based robust row following for compact agricultural robots," 2021, arXiv:2106.15029.

[5] A. N. Sivakumar et al., "Learned visual navigation for under-canopy agricultural robots," in Proc. Robotics: Sci. Syst., Jul. 2021.

[6] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, arXiv:1901.03407.

[7] T. Ji, S. T. Vuppala, G. Chowdhary, and K. Driggs-Campbell, "Multimodal anomaly detection for unstructured and uncertain environments," in Proc. Conf. Robot Learn., 2020, pp. 1443–1455.

[8] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder," IEEE Robot. Automat. Lett., vol. 3, no. 3, pp. 1544–1551, Jul. 2018.

[9] D. Park, H. Kim, and C. C. Kemp, "Multimodal anomaly detection for assistive robots," Auton. Robots, vol. 43, no. 3, pp. 611–629, 2019.

[10] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," 2016, arXiv:1607.00148.

[11] G. Kahn, P. Abbeel, and S. Levine, "Land: Learning to navigate from disengagements," IEEE Robot. Automat. Lett., vol. 6, no. 2, pp. 1872–1879, Apr. 2021.

[12] G. Kahn, P. Abbeel, and S.Levine, "Badgr:An autonomous self-supervised learning-based navigation system," IEEE Robot. Automat. Lett., vol. 6, no. 2, pp. 1312–1319, Apr. 2021.

[13] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM comput. Surv., vol. 41, no. 3, pp. 1–58, 2009.

[14] D. Park, Z. Erickson, T. Bhattacharjee, and C. C. Kemp, "Multimodal execution monitoring for anomaly detection during robot manipulation," in Proc. IEEE Int. Conf. Robot. Automat., 2016, pp. 407–414.

[15] D. Kappler, P. Pastor, M. Kalakrishnan, M. Wüthrich, and S. Schaal, "Datadriven online decision making for autonomous manipulation," in Proc. Robotics: Sci. Syst., 2015.

[16] L. Wellhausen, R. Ranftl, and M. Hutter, “Safe robot navigation via multi-modal anomaly detection,” IEEE Robot. Automat. Lett., vol. 5, no. 2, pp. 1326–1333, Apr. 2020.

[17] R. Hornung, H. Urbanek, J. Klodmann, C. Osendorfer, and P. Van Der Smagt, “Model-free robot anomaly detection,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2014, pp. 3676–3683.

[18] D. C. Guastella and G. Muscato, “Learning-based methods of perception and navigation for ground vehicles in unstructured environments: A review,” Sensors, vol. 21, no. 1, p. 73, 2021.

[19] M. A. Bekhti and Y. Kobayashi, “Regressed terrain traversability cost for autonomous navigation based on image textures,” Appl. Sci., vol. 10, no. 4, 2020, Art. no. 1195.

[20] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, “Real-time semantic mapping for autonomous off-road navigation,” in Proc. Int. Conf.Field Service Robot., Sep. 2018, pp. 335–350.

.