AeroNex: Aerial Intelligence for Precision Operations

Ankit Shaw*
University of Washington
ankit25@uw.edu

Xinyi Zhou*
University of Washington
xinyigix@uw.edu

Ravinder Gulla*
University of Washington
rgulla@uw.edu

Figure1: AeroNex website

Abstract
In the era of advancing technology, the AeroNex project pioneers a novel approach to environmental hazard detection and inventory management through the integration of cutting-edge drone technology and image processing techniques. Employing a drone as the main data acquisition tool, AeroNex captures image and video data in hazardous environments, particularly focusing on construction sites. These images undergo a comprehensive image processing for distinct feature extraction such as edge detection. Subsequently, we use Machine Learning algorithms to processed the images for object classification, enabling precise inventory management through object detection and counting. AeroNex represents a holistic solution for monitoring and managing hazardous environments (fig.3), pushing the boundaries of technology for enhanced safety and efficiency.

Keywords
Drone, Image Processing, Sobel Filter, Laplacian Filter, Black and White Filter, Colorization Filter, Morphological Filters, YOLOv3, Object Classification, Inventory Management, Hazardous Environments, Edge Detection.

Figure3: Hazardous environments

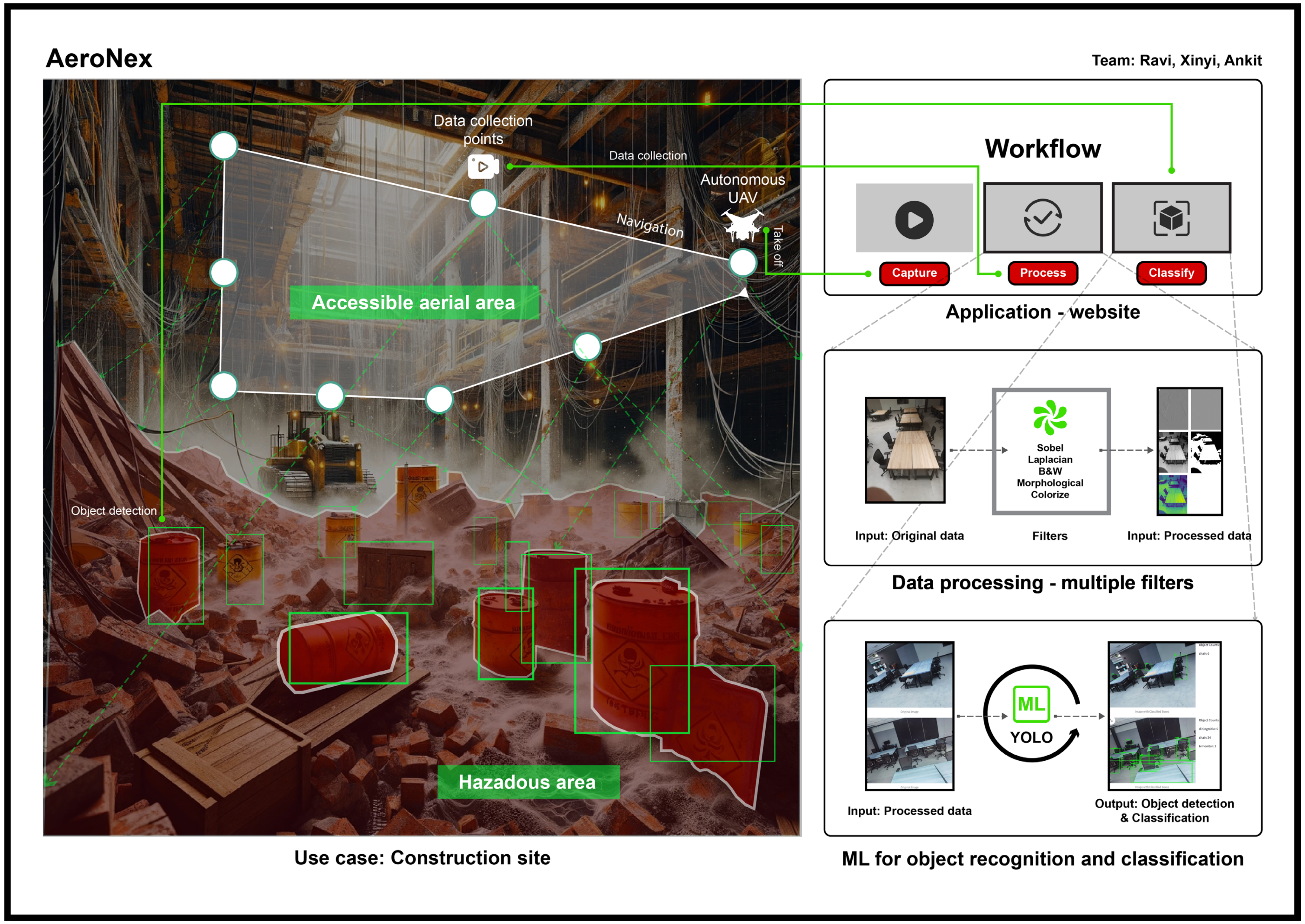

Figure3: Use case

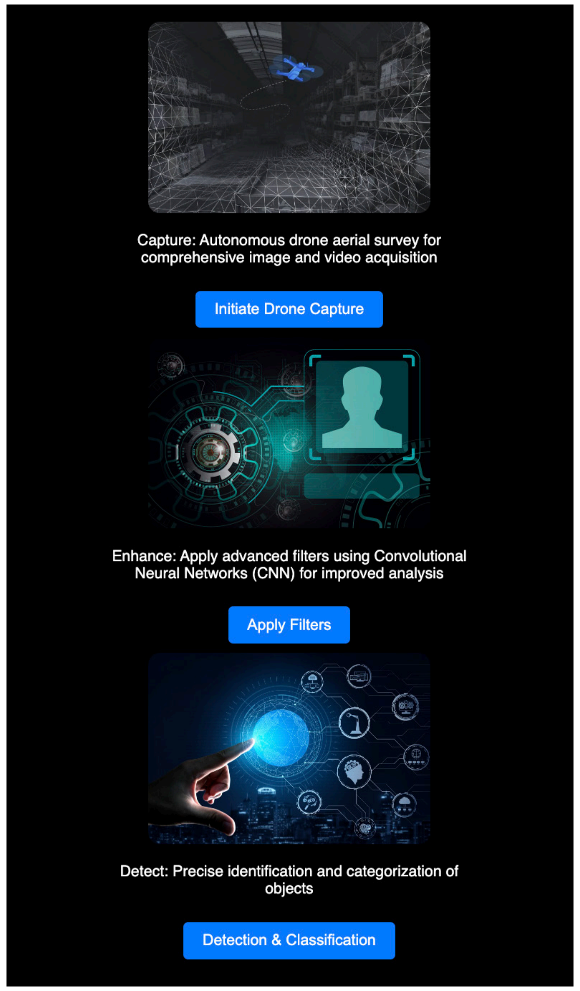

Figure4: Technology integration

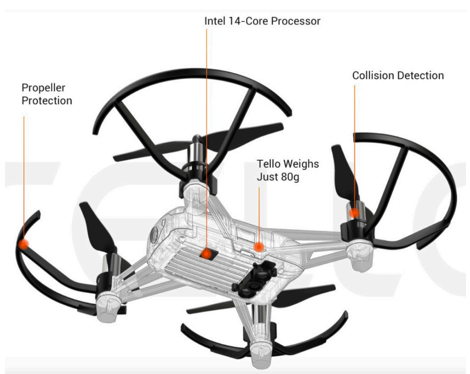

Figure5: Autonomous Drone Navigation

Technology Integration (fig.4):

a. Drone Technology: We used a drone as a central component of the project for data collection in hazardous environments, due to their agility, accessibility, and versatility.
b. Image Processing Techniques: We employed advanced image processing techniques, such as Sobel and Laplacian filters, colorization, and morphological filters for extracting meaningful features from the captured images.
c.YOLOv3 Model: We use YOLOv3 model for object classification and inventory management for its real-time capabilities and accuracy in detecting and counting objects in the images. (fig.3)

Navigation

A drone is employed for seamless navigation through hazardous environments, capturing images from vantage points inaccessible to humans. The autonomous flight capabilities of the drone ensure efficient coverage of the designated areas, facilitating comprehensive data collection. The navigation aspect of the AeroNex project is a crucial element that involves guiding the drone through inaccessible environments to capture images. The autonomous flight capabilities of the drone play a pivotal role in ensuring the efficient coverage of designated areas. AeroNex involves meticulous planning, autonomous flight control, obstacle avoidance, and integration with image capture mechanisms. This comprehensive approach ensures that the drone effectively navigates hazardous environments, captures relevant data, and contributes to the success of the project's objectives.
Autonomous Drone Navigation using Tello Boost (fig.5):
a. Waypoint Planning: Before deployment, the AeroNex system plans a flight path using predefined waypoints. These waypoints are strategically chosen to cover the entire hazardous environment while ensuring optimal image capture.
b. Collision Avoidance: Sensors, such as obstacle detection camera, is employed to ensure collision avoidance. The drone dynamically adjusts its flight path based instructions provide in the code and navigates safely.
c. Path Following Algorithm: A path-following algorithm guides the drone along the predefined waypoints. This algorithm takes into account the drone's current position, desired trajectory, and obstacle avoidance measures.
d. Altitude Control: We maintain a 10 ft altitude to ensure that the drone maintains a consistent height above ground level. This is crucial for capturing images at the desired resolution and perspective.
e. Energy Management: The navigation system also considers energy management to optimize the drone's flight time. Efficient route planning and dynamic adjustments contribute to extending the drone's operational duration. The drone's flying time is 12 minutes with 3 batteries. The navigation time for the drone to complete one flight is programmed to be under 2 minutes for all our simulations.

Integration with Image Capture

The navigation process is tightly integrated with the image capture mechanism. As the drone navigates through the hazardous environment, it captures images at predefined intervals and locations. The coordinated efforts of navigation and image capture contribute to the comprehensive data acquisition needed for subsequent analyses.
Safety Protocols: To ensure the safety of both the drone and the environment, the navigation system adheres to strict safety protocols. The program restricts the drone's flight within predefined boundaries, emergency landing procedures.

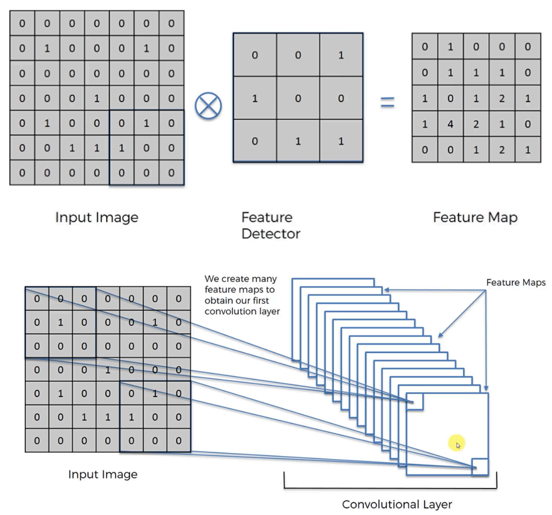

Figure6: Convolution process

Filters

The heart of AeroNex's image processing lies in the application of advanced filters to the captured images. Tello drone has a basic camera and we use Image filters to the drone-captured images for diverse purposes, including enhancing visual appeal and ensuring specific feature extraction. They help correct distortions, enhance color, adapt to environmental conditions, and contribute to a consistent and standardized dataset for analysis. Filters are employed for feature extraction of the images.

Convolution Process (fig.6)

a. Place the Filter matrix (Feature Detector) over a pixel in the image: The central pixel of the matrix aligns with the pixel in the image to which the convolution is being applied.
b. Multiply each element of the matrix by the corresponding pixel value in the image: Multiply each value in the matrix by the corresponding pixel value in the image covered by that element of the matrix.
c. Sum the results: Sum up all the products obtained in step 2.
d. Move the matrix to the next pixel and repeat: Slide the matrix to the next pixel in the image and repeat the multiplication and summation process.
e. Repeat for the entire image: Continue this process for every pixel in the image, performing the convolution operation at each location.
f. Generate a new image (convolved image): The result of the convolution operation is a new image, often called the "convolved image" or "gradient image." This image highlights areas where there are significant intensity changes along the vertical axis, indicating potential vertical edges.

Sobel Filter (fig.7): Applied for edge detection, the Sobel filter enhances the visibility of gradient changes in the images, highlighting significant features. The Sobel filter is a commonly used operator for edge detection in image processing. It operates by convolving the image with a convolution matrix, typically either Sobel-X or Sobel-Y, to highlight edges along the horizontal or vertical axis. Sobel filter convolution happens step by step

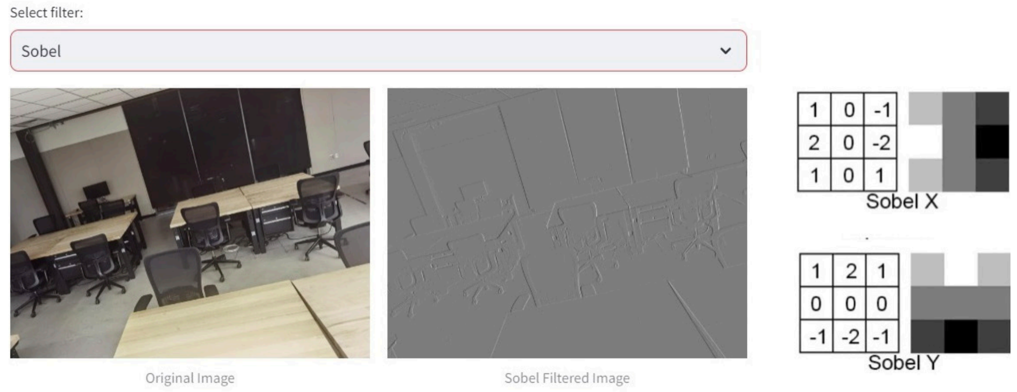

Figure7: Sobel filter

Sobel-X Convolution: The Sobel-X convolution is designed to emphasize vertical edges in an image. The convolution matrix for Sobel-X is:

1	0	-1
2	0	-2
1	0	1

Sobel-Y Convolution: The Sobel-Y convolution is similar but emphasizes horizontal edges. The Sobel-Y matrix is:

1	2	1
0	0	0
-1	-2	-1

The steps for Sobel-Y convolution are the same as for Sobel-X, but the emphasis is on capturing intensity changes along the horizontal axis.
In summary, Sobel filter convolution involves moving a small matrix (Sobel-X or Sobel-Y) over each pixel in an image, multiplying the matrix values by the corresponding pixel values, summing the results, and producing a new image that highlights edges along the specified axis. This process enhances the visibility of edges and is particularly useful for feature extraction in image analysis tasks. We use Sobel-X filter.

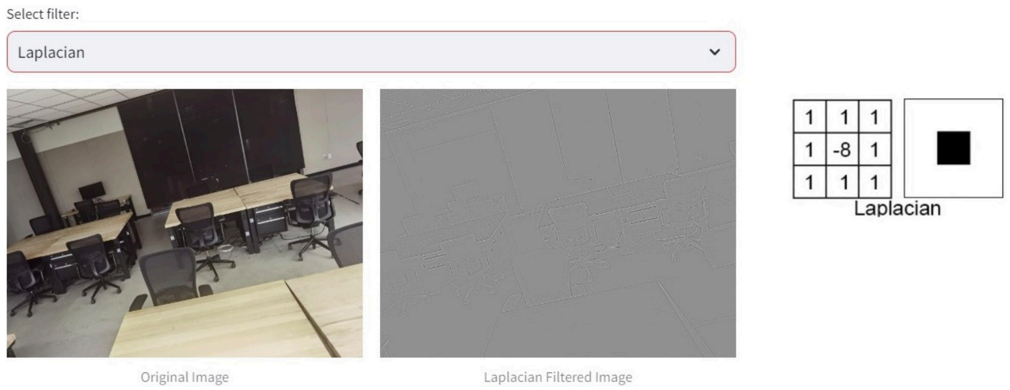

Figure8: Laplacian filter

Laplacian Filter (fig.8): Designed for detecting edges and fine details, the Laplacian filter contributes to a more refined analysis of the images, capturing intricate structures. The Laplacian filter is an image processing operator used for edge detection and the enhancement of fine details. It highlights regions of rapid intensity change in an image, making it particularly useful for identifying edges. The convolution matrix for the Laplacian filter is:

1	1	1
1	-8	1
1	1	1

In summary, the Laplacian convolution matrix focuses on detecting changes in intensity in the immediate neighborhood of each pixel, making it a powerful tool for edge detection and feature enhancement in image processing.

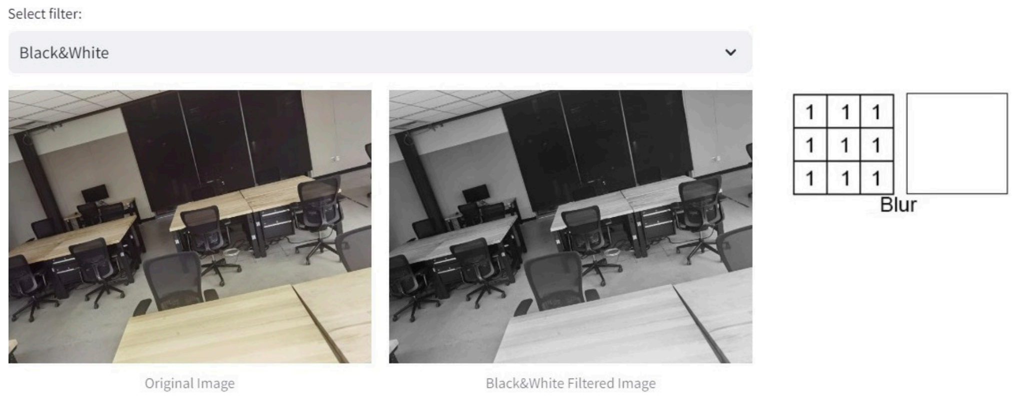

Figure9: Black and white filter

Black and White Filter (fig.9): Simplifying the images to grayscale aids in focusing on structural elements, providing a foundation for subsequent analyses.
Colorization Filter (fig.10): This filter enhances visual information by introducing color, providing a clearer representation of the environment and aiding in object recognition.
Morphological Filters (fig.11): Employed for shape analysis and manipulation, morphological filters play a crucial role in refining the extracted features and preparing the images for further processing.Morphological operations, specifically dilation and erosion, are fundamental techniques in image processing that modify the shape and structure of objects within an image. These operations involve the use of a structuring element (also known as a kernel) to process the

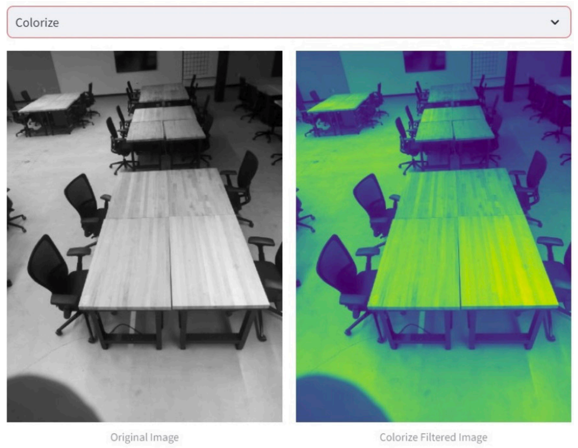

Figure10: Colorization filter

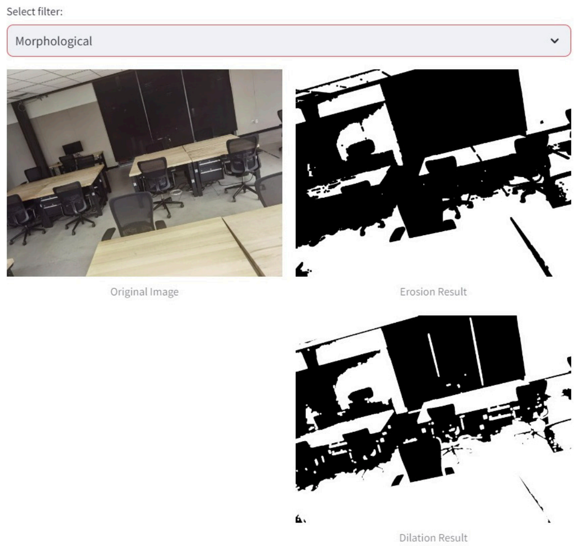

Figure11: Morphological filter

pixel values in an image. Let's delve into the detailed explanations of the convolution process for both dilation and erosion.

Dilation Convolution: Dilation is a morphological operation that enhances the brighter regions in an image, making objects more prominent. The convolution process involves the following steps:

Structuring Element (Kernel):

0	1	0
1	1	1
0	1	0

The result is a new image, often called the "dilated image," where brighter regions have expanded, and objects are made more prominent.

Erosion Convolution: Erosion is another morphological operation that shrinks the brighter regions in an image. The convolution process involves the following steps:

Structuring Element (Kernel):

1	1	1
1	1	1
1	1	1

The result is a new image, often called the "eroded image," where brighter regions have contracted, and finer details are retained.

In summary, the convolution of morphological filters, such as dilation and erosion, involves sliding a structuring element over each pixel in the image, modifying pixel values based on the specific rules outlined for each operation. Dilation expands brighter regions, while erosion contracts them. These morphological operations are essential for tasks such as image segmentation, noise reduction, and feature extraction in image processing applications.

Detailed explanations of the convolution matrices associated with each filter contribute to a comprehensive understanding of the image processing techniques employed in AeroNex.

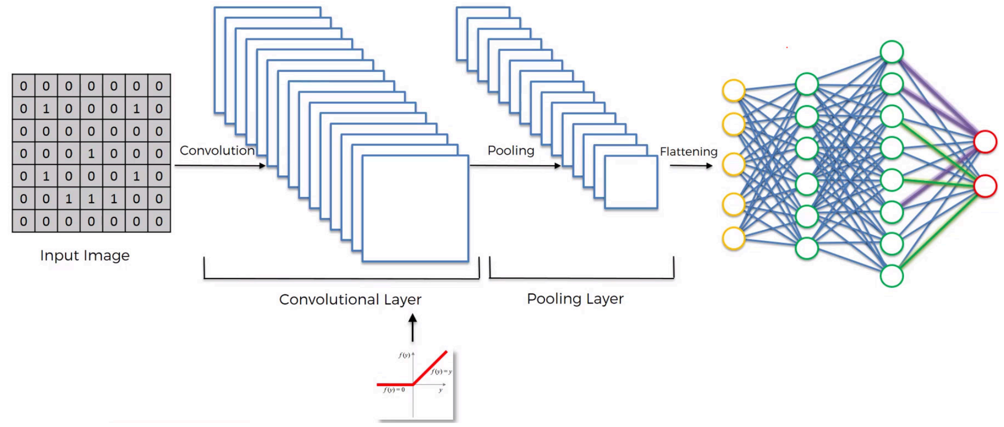

Figure12: Process of image classification

Classification (fig.12)

YOLOv3 (You Only Look Once version 3): Image Classification for AeroNex

YOLOv3 is a state-of-the-art real-time object detection model that excels in identifying and classifying multiple objects within an image. Its architecture allows for efficient and accurate object recognition, making it a suitable choice for the AeroNex project's inventory management component.

Architecture: YOLOv3 adopts a deep neural network architecture based on convolutional neural networks (CNNs). It divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell.

Bounding Box Prediction: YOLOv3 predicts bounding boxes by regressing the coordinates (x, y, width, height) for

Bounding Box Prediction: YOLOv3 predicts bounding boxes by regressing the coordinates (x, y, width, height) for each object within a grid cell. Multiple bounding boxes are predicted for each grid cell, each associated with a confidence score representing the model's confidence in the box containing an object.
Class Prediction: YOLOv3 predicts class probabilities for each bounding box. The model performs multi-label classification, allowing for the identification of multiple objects within a single bounding box. The class predictions are conditioned on the presence of an object within a bounding box, contributing to the overall detection and classification accuracy.
Feature Pyramid Network (FPN): YOLOv3 incorporates FPN, which enhances the network's ability to detect objects at different scales. It combines features from different levels of the network, allowing the model to capture both fine and coarse details.
Anchor Boxes: YOLOv3 employs anchor boxes to improve bounding box predictions. These anchor boxes are predefined shapes that the model uses to refine its predictions, leading to more accurate localization.
Non-Maximum Suppression (NMS):*
To eliminate duplicate detections, YOLOv3 employs NMS. This post-processing step selects the most confident bounding box among those with overlapping predictions and suppresses redundant boxes.

Classification for AeroNex Images (fig.13)

Input Processing: Images captured by the drone that are sent to the filters are passed into the YOLOv3 model as input for object detection and classification. YOLOv3 operates on the images offline, and helps in the crucial and prompt decision-making and inventory management.

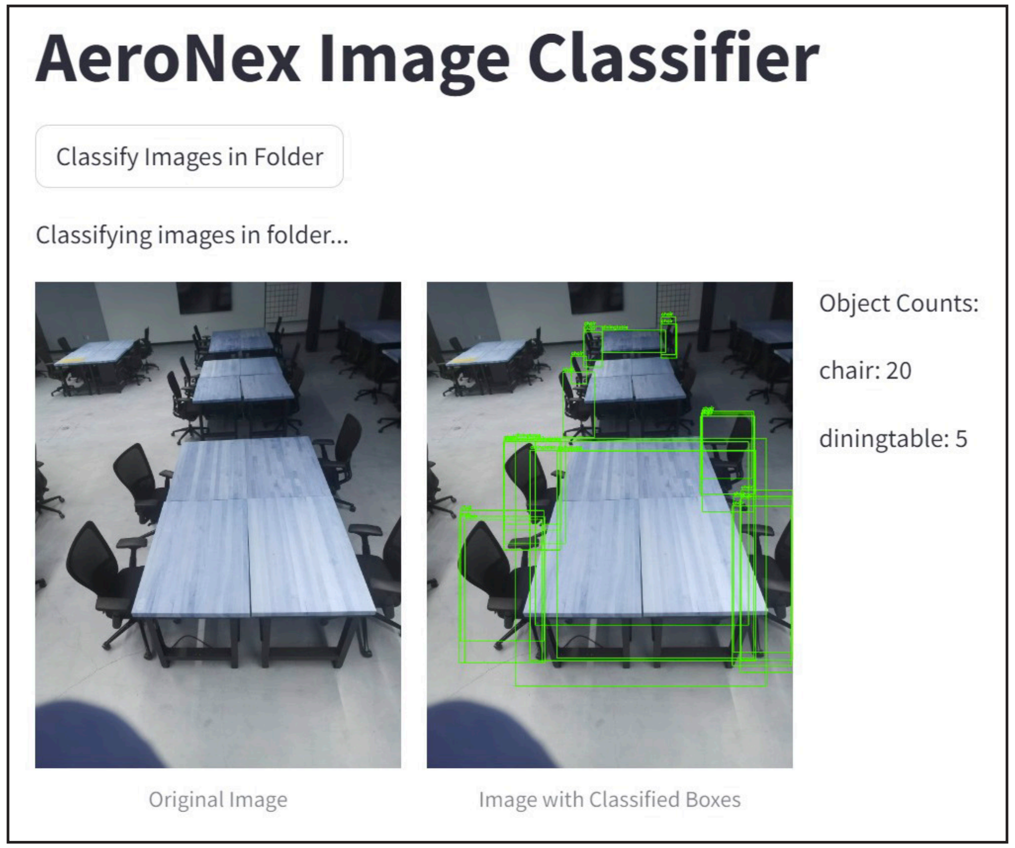

Figure13: Aero image classifier

Inventory Management:
YOLOv3 plays a pivotal role in inventory management by accurately identifying and classifying objects in the images. The model's ability to handle multiple objects in a single image contributes to the efficiency of counting and managing inventory in hazardous environments.
Integration with Navigation and Image Processing:
YOLOv3 seamlessly integrates with the overall AeroNex system, complementing the navigation system's autono-

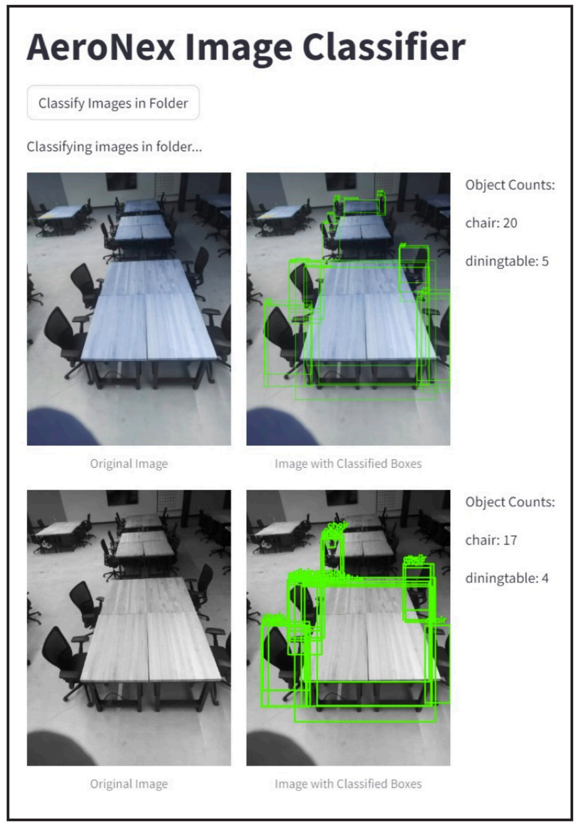

Figure14: 13: Aero image classifier

mous flight capabilities and the image processing pipeline's feature extraction. This integration creates a comprehensive solution for hazard detection and inventory management.

Performance Evaluation (fig.14)

Metrics: The performance of YOLOv3 in the AeroNex project is evaluated using metrics such as precision, recall, and F1 score. These metrics provide insights into the model's accuracy in detecting and classifying objects in hazardous environments. It can be seen from the image that the original image and the image after applying the b&W filter are being classified differently by the YOLO Classifier.

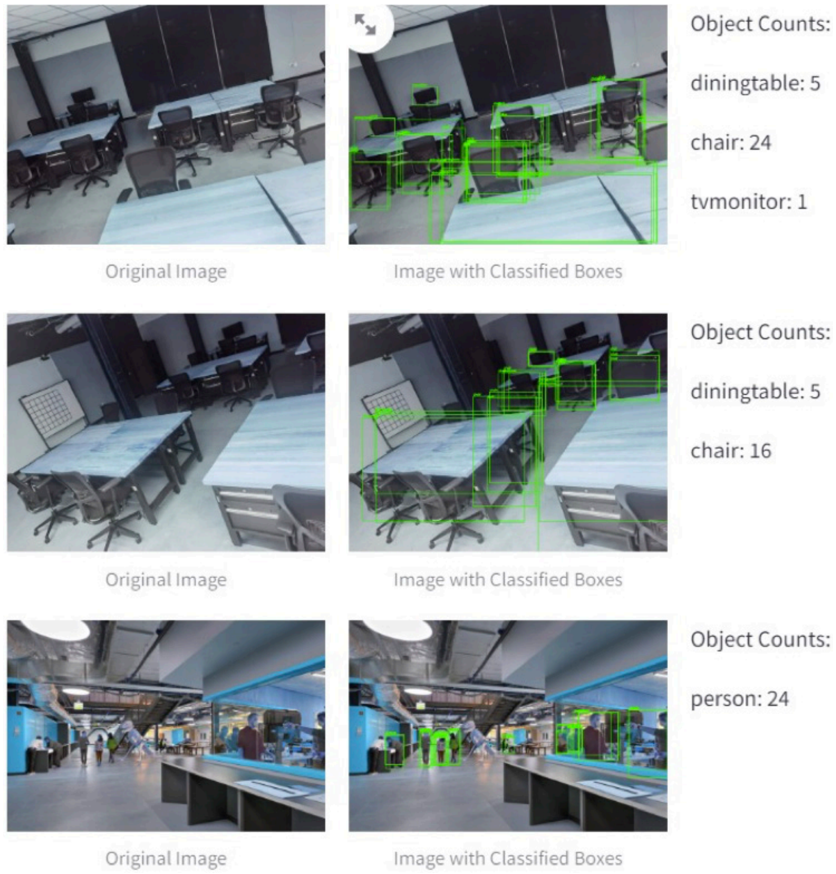

Figure14: Performance evaluation

Optimization: YOLOv3 may undergo optimization techniques, such as model quantization or pruning, to reduce its computational requirements, enabling efficient deployment on edge devices like the drone.
Conclusion: YOLOv3's robust object detection capabilities significantly contribute to the success of the AeroNex project. Its real-time processing, accuracy in object classification, and seamless integration make it a powerful tool for hazard detection and inventory management in challenging environments.

Future Research

Multi threaded process: Currently the drone runs on a single process. For this reason, the drone operations happen synchronously i.e., the drone flies along the path, stops, takes images/video for 2 seconds, then moves the next waypoint and then takes images/video for 2 seconds. We want to increase
SLAM: This effective technology will improve the localization and mapping criteria for the drone. As the drone moves through the environment, it updates its estimate of its own location while also updating the map based on sensor data. SLAM algorithms handle the uncertainty associated with both localization and mapping, providing a more robust solution.
Real-time Adjustments: The navigation system continuously monitors the drone's surroundings and makes real-time adjustments to the flight path using on Board Collision Detection. This adaptability ensures that unexpected obstacles or environmental changes do not disrupt the mission.
Return-to-Base Functionality: In the case of low battery levels, completion of the mission, or emergency situations, the drone will be programmed with a return-to-base functionality.

References

Introduction to Convolutional Neural Netrwork: by Jianxin Xu 2017. http://cs.nju.edu.cn/wujx/paper/CNN.pdf

Gradient-Based Learning applied to Document Recognition: Yan LeCun, Leon bottou, Yoshua bengio and Patrick Haffner. https://www.iro.umontreal.ca/~lisa/pointeurs/lecun-01a.pdf

https://medium.com/@er_95882/convolution-image-filters-cnns-and-examples-in-python-pytorch-bd3f3ac5df9c

https://www.v7labs.com/blog/neural-networks-activation-functions#:~:text=ReLU%20activation%20function%20should%20only,depth%20greater%20than%2040%20layers

https://www.v7labs.com/blog/neural-networks-activation-functions#:~:text=ReLU%20activation%20function%20should%20only,depth%20greater%20than%2040%20layers

https://dl.acm.org/doi/abs/10.1145/3459104.3459114?casa_token=28r7Fb7V6QcAAAAA:pbVCAopHHhafvktoMgolNX4SePRTZC217IGbkL9JXVbrasQWaFCnqstFCoFTowzV5QjAU-k_nrTd

https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21495?casa_token=vW8OGavbA4MAAAAA:b4lcd7w4PoHVqPvPpfH_Srlt9Z_m4hiTyndmx6sYKw-c1f_PtUhO_xw_UVhvacqxdTk9tpoagMfpAvk

https://ieeexplore.ieee.org/abstract/document/5513152?casa_token=5GgRXV_49UMAAAAA:zo4QrVJld3KJx1a-ggHvpV5Gd3xi9NrPh4_2ogmKHfuymss8TUsPQQ1tZiQOzlNcFO_yu6-B

Code and Readme file : AeroNex_Code.pdf