Dissertation Submitted for the partial fulfillment of the M.Sc. (Integrated) Data Science degree to the Department of AI&ML and Data Science

M.Sc. Project Dissertation

Semester- 10

Multimodal AI for Industry Safety and Security

submitted to

By

Agarwal Shruti Hemant

Under the guidance of

Mr. Harshal Trivedi

M.Sc. (Integrated) Data Science

Department of AIML & Data Science

School of Emerging Science and Technology

Gujarat University

April, 2024

DECLARATION

I hereby declare that the study entitled **"Multimodal Al for Industry Safety and Security"** submitted to the Gujarat University, Navarangpura, Ahmedabad (Gujarat) in partial fulfillment of M.Sc. (Int) Data Science degree is the result of investigation done by myself. The material that has been obtained (and used) from other sources has been duly acknowledged in this study. It has not been previously submitted either in part or whole to this or any other university or institution for the award of any degree or diploma.

Place: Ahmedabad

Date: 29th April 2024

Signature

Shruti Hemant Agarwal

ACKNOWLEDGMENT

I am writing this acknowledgment to express my sincere gratitude to everyone who has supported me throughout my project.

Firstly, I would like to thank my academic supervisor, Prof. Dr. Nita H. Shah, for their continuous support and insightful feedback during this project. Their advice was instrumental in shaping the direction of my project and ensuring its academic rigor.

I extend my heartfelt appreciation to the team at TuskerAI, where I completed my internship, for providing me with the opportunity to apply theoretical knowledge into practical experience. Special thanks to Mr. Harshal Trivedi, Mr. Rushik Vora and Mr. Soham Thakkar whose expertise, insights, and assistance have significantly enriched my understanding of the subject matter and facilitated the smooth progress of my project.

I am also grateful to my professors and academic staff at Gujarat University, whose teachings and mentorship have laid the foundation for my academic and professional growth. Their continuous support and encouragement have inspired me to strive for excellence and pursue my passion for research projects.

I would like to acknowledge the invaluable support of my family and friends, whose unwavering encouragement, understanding, and love have sustained me through the challenges and triumphs of this endeavor. Their belief in my abilities has been a constant source of motivation, and I am profoundly grateful for their presence in my life.

In conclusion, I recognize and appreciate the collective efforts of everyone who has played a part in shaping this dissertation. Your contributions have been invaluable, and I am honored to have had the opportunity to work with such talented individuals. Thank you all for your support and encouragement.

~ Agarwal Shruti Hemant

Index

Chapter 1:
Abstract & Key Words

Abstract

Industrial accidents cast a long shadow, jeopardizing worker well-being, operational efficiency, and incurring substantial financial burdens. Traditional safety methods, heavily reliant on human vigilance, are susceptible to fatigue and distraction, leaving room for unforeseen hazards. This project proposes a paradigm shift in industrial safety with the development of a powerful Multimodal AI system. The core of the system consists of multiple Vision – Based AI models that analyze real-time video feeds from strategically placed cameras within industrial environments. These models are trained to detect the absence or presence of critical personal protective equipment (PPE), such as hardhats, safety vests, gloves, and shoes. Additionally, the system can identify potential fire and smoke incidents and vehicles moving in the wrong direction, which are common precursors to more significant accidents. By automating the detection of these elements, the system aims to significantly reduce the reliance on human monitoring, thereby decreasing the probability of human error. The ultimate goal is to significantly reduce accidents, human suffering, operational disruptions, and financial losses. This project not only focuses on the effectiveness of Vision-Based AI but also paves the way for a future where Multimodal AI becomes an indispensable tool in creating a safer and more secure industrial environment.

Key Words

Vision – Based AI, Multimodal AI, Industry, Accidents, Detection

Chapter 2:
Introduction

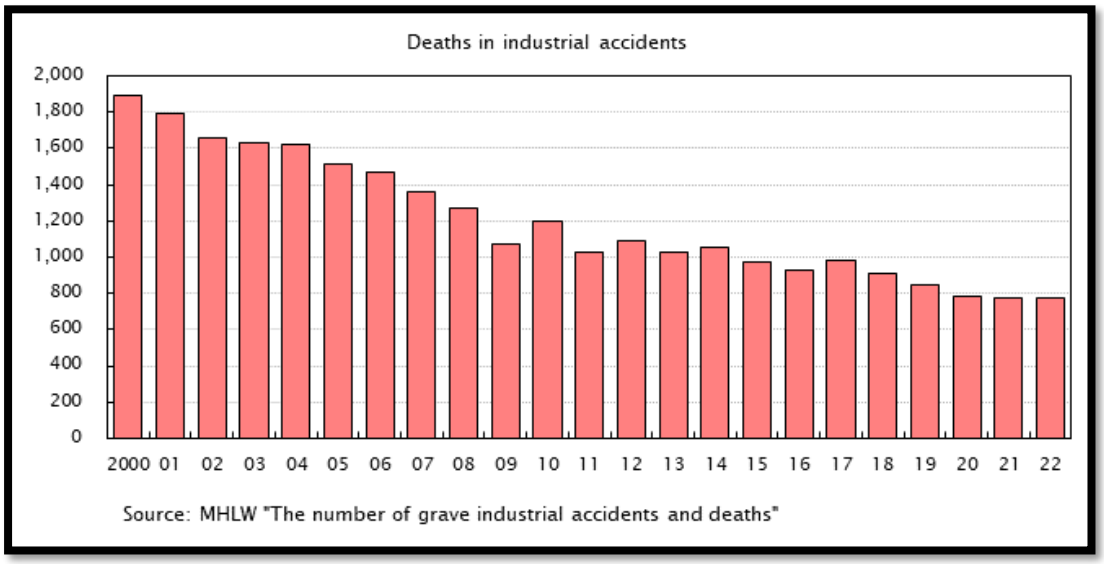
2.1 Background

The industrial sector is a vital contributor to global economic growth. However, it also comes with inherent risks, with industrial accidents posing a significant threat to worker safety, operational efficiency, and financial stability.

Industrial environments pose significant risks to worker safety and security due to the presence of various hazards, ranging from heavy machinery and equipment to hazardous materials and environmental conditions. Despite stringent safety regulations and protocols, industrial accidents continue to occur, resulting in injuries, fatalities, operational disruptions, and financial losses.

Traditional safety measures in industrial settings heavily rely on human vigilance and compliance with safety protocols. However, human factors such as fatigue, distraction, and complacency can compromise the effectiveness of these measures, leaving workers vulnerable to unforeseen hazards. Furthermore, the sheer scale and complexity of industrial operations make it challenging for human operators to monitor every aspect of safety consistently.

2.2 Problem Statement



Industrial accidents are a persistent nightmare. Every year, countless workers are injured due to unforeseen hazards. According to the International Labor Organization (ILO), around 2.3 million people die from work-related accidents or diseases every year, which is over 6,000 deaths per day. The Asia and Pacific region have the highest work-related mortality, accounting for 63% of the global total. The most hazardous sectors are agriculture, construction, forestry and fishing, and manufacturing, which account for 200,000 fatal injuries per year. Traditional safety methods rely on human vigilance, but fatigue and distraction leave dangerous gaps. These accidents are more than just statistics. They cause immense human suffering, disrupt operations, and incur significant financial burdens. We need a robust and proactive approach to industrial safety.

2.3 Objective

The mission is to develop a robust Vision Based AI System tailored for industrial environments, aimed at enhancing safety and security through proactive detection and analyze the potential impact of Multimodal AI on improving overall industry safety and security using the existing infrastructure.

2.4 Challenges of Traditional Safety Methods:

- **Human Reliance:** Traditional safety approaches heavily rely on human vigilance for hazard detection. Unfortunately, human attention is susceptible to fatigue, distraction, and lapses in judgment, leading to potential oversights and accidents.

- **Reactive Approach:** Traditional methods often focus on responding to accidents after they occur, rather than proactively preventing them. This reactive approach can be costly in terms of human suffering, lost productivity, and financial repercussions.

- **Limited Scope:** Conventional methods might struggle to address all safety concerns, especially in large or complex industrial environments.

2.5 Impact of Industrial Accidents:

- **Human Suffering:** Industrial accidents can cause severe injuries or even fatalities. Beyond the physical toll, these accidents can lead to psychological trauma and long-term disabilities for workers.

- **Operational Disruptions:** Accidents can cause significant disruptions to production processes, leading to delays, lost output, and financial losses.

- **Financial Burden:** The costs associated with industrial accidents are substantial, encompassing medical expenses, compensation claims, and potential regulatory fines.

In recent years, there has been a growing recognition of the limitations of traditional safety approaches and the need for more proactive and predictive safety measures in industrial environments. Advances in artificial intelligence (AI) and computer vision technologies have opened up new possibilities for enhancing industrial safety and security through automation and real-time monitoring.

AI-powered vision-based systems have emerged as a promising solution for improving safety in industrial environments. These systems leverage computer vision algorithms to analyze real-time video feeds from strategically placed cameras within industrial facilities. By continuously monitoring the environment, these systems can detect potential safety hazards, identify safety violations, and alert operators in real-time, thereby reducing the reliance on manual monitoring and intervention.

Vision-based AI models can be trained to recognize various safety-critical elements, such as personal protective equipment (PPE) worn by workers, including hardhats, safety vests, gloves, and shoes. Additionally, these models can detect anomalies and potential safety threats, such as fire and smoke incidents, unauthorized personnel in restricted areas, and vehicles moving in the wrong direction.

2.6 Need for a Proactive Approach:

The limitations of traditional safety methods necessitate a paradigm shift towards a more proactive approach. This is where Multimodal AI presents a compelling solution. The development of a robust vision-based AI system tailored for industrial safety and security represents a critical step towards creating safer and more secure work environments. By leveraging the power of AI and computer vision, this project aims to significantly reduce the incidence of industrial accidents, minimize human suffering, and enhance operational efficiency in industrial settings.

2.7 What is Industrial Safety?

Industrial safety is a multi-disciplinary approach to ensuring compliance

with regulations, safe working practices, and the health of employees in a workplace. It involves managing all operations and procedures within an industry to minimize hazards, risks, accidents, and near misses, while protecting industrial workers, machinery, facilities, structures, and the environment.

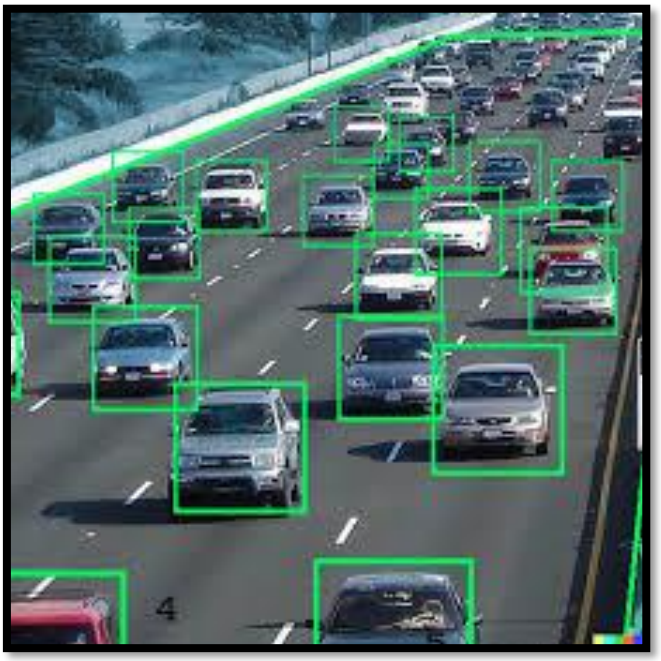
2.7 Why is safety important in industry?

Industrial safety is important as it safeguards human life, especially in high-risk areas such as nuclear, aircraft, chemical, oil and gases, and mining industries, where a fatal mistake can be catastrophic. Industrial Safety reduces risks to people, and processes.

2.8 What is Wrong way Vehicle Detection?

Wrong way vehicle detection count is a system or technology designed to identify vehicles traveling in the wrong direction on roads, highways, or within industrial premises. This technology typically utilizes camera monitoring devices to detect vehicles moving against the flow of traffic. The detection process involves analyzing the movement patterns of vehicles and comparing them to the expected direction of travel. When a vehicle is detected traveling in the wrong direction, the system generates an alert or notification to alert authorities or relevant personnel. The "count" aspect of wrong way vehicle detection refers to the ability of the system to accurately tally the number of instances where vehicles are detected traveling in the wrong direction. This count is essential for monitoring and managing traffic safety, as it provides insights into the frequency and location of wrong-way incidents. Wrong way vehicle

detection count systems are crucial for enhancing road safety and preventing accidents caused by vehicles traveling in the wrong direction. By promptly identifying and alerting authorities to these incidents, these systems help mitigate the risk of head-on collisions and other serious accidents, ultimately improving overall traffic safety.



Chapter 3:
Basic Terminology

- **Vision - Based AI:** Vision-based AI, also known as computer vision, is a subfield of artificial intelligence (AI) that empowers computers to interpret and understand visual information from the real world. This information can be in the form of digital images, videos, or camera feeds.

- **Multimodal AI:** Multimodal AI refers to the use of an artificial intelligence system that integrates and processes data from multiple types of inputs. This integration allows the AI to perform more comprehensively across various scenarios. It allows to integrate all the models.

- **Industrial Safety:** Industrial safety is a set of safety rules, policies, and regulations that protect workers, the workplace, equipment, and the environment from hazards.

- **Object Detection:** Object detection is a computer vision solution that identifies objects, and their locations, in an image. An object detection system will return the coordinates of the objects in an image that it has been trained to recognize. The system will also return a confidence level, which shows how confident the system is that a prediction is accurate.

- **Bounding Box:** A bounding box is a rectangular or square-shaped region that is drawn around an object in an image or a video frame to indicate its location and extent within the visual data. Bounding boxes are commonly used in computer vision tasks, particularly in object detection, where they serve as a way to precisely localize objects of interest.

- **Confidence Level:** refers to the threshold at which the model considers an object detection prediction to be confident enough to be included in the final output. In object detection models, confidence level is typically set as a parameter during inference.

Chapter 4:

Literature review

4.1 Impact of Multimodal Artificial Intelligence (AI) Across Various Industries

This literature review explores the transformative impact of multimodal artificial intelligence (AI) across several key sectors, including healthcare, retail, education, and security. Multimodal AI, which integrates multiple forms of data such as text, image, and sound, is enhancing precision in diagnostics, personalizing customer service, revolutionizing educational approaches, and improving security surveillance systems. Multimodal AI is significantly transforming industries such as healthcare, retail, education, and security by integrating diverse data types to enhance diagnostic precision, personalize customer interactions, tailor educational content, and improve surveillance effectiveness. This technology not only boosts operational efficiencies but also elevates user experiences, promising further advancements and broader applications in various sectors. The ongoing evolution of multimodal AI continues to demonstrate its pivotal role in shaping technological progress across these key industries.

4.2 Advancements in AI for Industrial Safety and Security

The field of industrial safety and security has witnessed significant advancements with the integration of artificial intelligence (AI) technologies.
AI-driven systems offer real-time monitoring capabilities and have shown promise in detecting safety hazards and preventing accidents in industrial environments.

4.3 Object Detection Models in Industrial Applications

YOLOv5 and YOLOv8 architectures have emerged as prominent choices for object detection tasks in industrial settings (Wang et al., 2021).

Previous research has demonstrated the effectiveness of YOLO-based models in detecting safety hazards such as fire and smoke in industrial environments (Zhang et al., 2020).

4.4 Conclusion

Recent research papers have highlighted a shift in the application of AI within industrial safety and security contexts. Traditionally, AI has been utilized post-incident, primarily for analyzing data retrospectively. However, a growing body of studies underscores the transformative potential of AI-based systems in bolstering safety measures within industrial environments. These systems offer real-time monitoring capabilities, enabling proactive risk management by promptly identifying safety hazards as they emerge. Furthermore, AI algorithms exhibit remarkable efficiency in detecting and analyzing safety risks, thanks to their ability to process vast amounts of data with speed and accuracy. Vision-based AI systems, in particular, have emerged as a cornerstone in this domain, leveraging computer vision techniques to analyze visual data from cameras in real-time. By continuously monitoring activities and detecting potential risks, these systems pave the way for early intervention and prevention of incidents. Overall, the integration of AI-driven solutions represents a paradigm shift towards proactive safety management in industrial settings, promising to create safer working environments and mitigate risks effectively.

Chapter 5:
Methodology

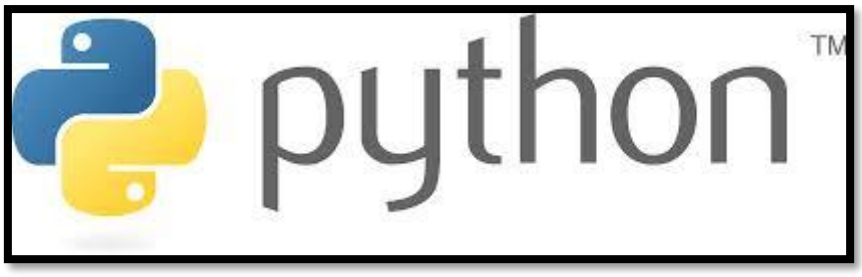
5.1 Hardware & System Requirements

- A combination of Microsoft Windows and Ubuntu was used for this project.
- The code works on all OS.
- Minimum 16 GB RAM is required.
- A GPU is required.

5.2 Tools, Technologies & Software Used

- **Python**

  Python is a high-level, general-purpose programming language that is widely used for web development, data analysis, scientific computing, and many other purposes. It is known for its simplicity, readability, and flexibility, as well as its large and active developer community. Some of the key features of Python include A large standard library that supports many common programming tasks, such as connecting to web servers, reading and writing files, and working with data, An interactive interpreter, which allows you to try out code snippets and experiment with the language in an interactive environment, Support for object-oriented, imperative, and functional programming styles, Dynamically-typed, which means that you don't have to specify the data type of a variable when you declare it, Cross-platform compatibility, which means that Python programs can run on multiple operating systems.



- **LabelImg**

  LabelImg is an open-source graphical image annotation tool. It's primarily used for manually labeling objects in images for tasks such as object detection or image classification. LabelImg allows users to draw bounding boxes around objects of interest within an image and assign labels to those boxes. It supports various annotation formats, including Pascal VOC and YOLO, making it compatible with many machine learning frameworks and libraries. LabelImg is commonly used in computer vision projects where labeled training data is needed to train models for tasks like object detection or semantic segmentation. Its user-friendly interface and support for multiple annotation formats make it a popular choice among developers and researchers working on computer vision applications.

- **Roboflow**

  Roboflow is a platform that provides tools and services for managing, annotating, and preparing image datasets for machine learning applications, particularly in the field of computer vision. It offers a range of features to streamline the process of creating, organizing, and augmenting image datasets, making it easier for developers and researchers to train machine learning models.

- **Amazon Web Services**

  Amazon Web Services (AWS) is a comprehensive cloud computing platform offered by Amazon.com. It provides a wide range of cloud services, including computing power, storage, networking, databases, machine learning, analytics, security, and more. AWS offers these services on a pay-as-you-go basis, allowing businesses to scale and innovate without the upfront costs and complexities of managing physical infrastructure.

- **PyCharm**

  PyCharm is a popular integrated development environment (IDE) specifically designed for Python programming. It's developed by JetBrains, known for their suite of powerful IDEs for various programming languages. PyCharm provides a comprehensive set of tools for Python development, including code analysis, debugging, version control integration, and support for web development frameworks such as Django and Flask. It offers features like syntax highlighting, code completion, refactoring tools, and a customizable user interface to enhance productivity for Python developers.

5.3 Libraries Used

- **ultralytics**

  A library offering deep learning utilities with a focus on computer vision tasks, providing tools for training, inference, and evaluation of models, particularly optimized for object detection and image classification.

- **matplotlib**

  A comprehensive plotting library for Python, widely used for creating visualizations such as histograms, scatter plots, and heatmaps, which can be helpful for analyzing image data and model performance in computer vision projects.

- **numpy**

  Essential for numerical computing in Python, NumPy provides powerful array manipulation capabilities, facilitating efficient handling and processing of image data in computer vision applications.

- **opencv-python**

  A popular computer vision library providing a wide range of functionalities for tasks like image and video manipulation, object detection, and feature extraction, serving as a cornerstone for many computer vision projects.

- **pandas**

  Pandas is a library for the Python programming language that is used for data manipulation and analysis. It provides functions and utilities for working with tabular data, such as data stored in a spreadsheet or a database table. One of the main advantages of Pandas is that it provides a high-level interface for working with data, making it easy to perform tasks such as filtering, aggregating, and transforming data. It also integrates well with other libraries for data analysis, such as NumPy and Matplotlib.

- **seaborn**

  Seaborn is a Python data visualization library built on top of Matplotlib, designed to make creating informative and attractive statistical graphics more straightforward. Seaborn is known for its ability to generate complex visualizations with concise code, making it a valuable tool for data analysts and researchers.

- **scikit-learn**

  sklearn is a popular open-source machine learning library for Python, offering a versatile set of tools and algorithms for a wide array of machine learning tasks. It is renowned for its user-friendly interface, making it accessible to both novice and experienced data scientists. Scikit-learn features an extensive collection of machine learning algorithms, tools for data preprocessing, model evaluation, and seamless integration with other Python libraries like NumPy and SciPy. It is actively maintained and has a vibrant community, ensuring it stays up-to-date and well-supported for diverse machine learning and data science projects.

- **tensorflow**

  TensorFlow is a popular open-source machine learning framework developed by Google. TensorFlow is used for developing and deploying machine learning models, with a focus on deep learning. It's employed in various applications, including image and speech recognition, natural language processing, recommendation systems, and more. TensorFlow is used in research, industry, and academia to build and train advanced artificial intelligence models for a wide range of tasks.

- **Pillow**

  A user-friendly library for image processing tasks, offering functionalities for image opening, manipulation, and saving in various formats, supporting tasks like resizing, cropping, and filtering.

- **psutil**

  python system and process utilities: A utility library providing system monitoring functionalities, which can be useful for tracking resource usage during the execution of computer vision algorithms, ensuring efficient utilization of hardware resources.

- **requests**

  A versatile HTTP library for Python, enabling easy integration with web services and APIs, which can be beneficial for tasks like fetching image data from online sources or interacting with cloud-based services.

- **scipy**

  Offering a diverse set of scientific computing tools, SciPy includes modules for optimization, integration, interpolation, and statistical functions, which can complement computer vision algorithms for tasks like image enhancement or feature extraction.

- **thop**

  An efficient tool for estimating the FLOPs (floating-point operations) and model size of neural networks, aiding in model optimization and resource allocation for computer vision tasks on various hardware platforms.

- **torch & torchvision**

  The PyTorch framework along with its vision library, Torchvision, provide a flexible and powerful platform for deep learning in computer vision, offering high-level abstractions for building and training neural networks.

- **tensorboard**

  A visualization tool integrated with TensorFlow, Tensorboard enables tracking and visualizing various metrics such as loss curves, model architectures, and embedding visualizations during the training process,

aiding in model debugging and optimization.

- **clearml & comet**

  Experiment management platforms designed to streamline the development and deployment of machine learning models, providing features like experiment tracking, version control, and collaboration, which can enhance productivity in computer vision projects.

- **setuptools**

  A package management library for Python, Setuptools facilitates the distribution and installation of Python packages, ensuring smooth integration of dependencies required for computer vision projects.

- **ipython**

  An interactive computing environment for Python, IPython provides an enhanced REPL (Read-Eval-Print Loop) experience with features like syntax highlighting, tab completion, and rich media display, which can be beneficial for interactive exploration and experimentation in computer vision workflows.

5.4 Data Analysis

5.4.1 About Data

The data was obtained from different sources and it is a private data of the company. Image & Video Data was used.

5.4.2 Data Preprocessing

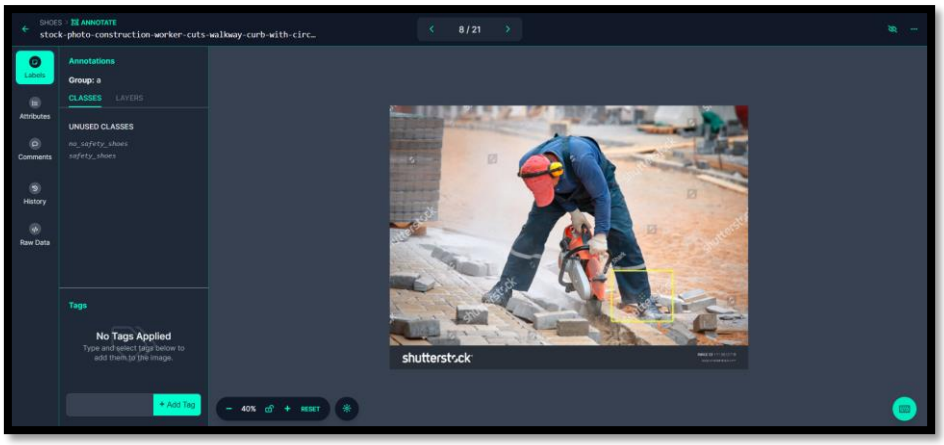
The data obtained was preprocessed before being trained. The preprocessing steps included:

- **Data Cleaning:**

  Data cleaning typically involves reviewing the dataset to identify and remove any incorrect or unusual images that may negatively impact model training and performance. The process of data cleaning begins by inspecting the dataset to identify anomalies such as corrupted images, images with low resolution, or images containing irrelevant or misleading content. These incorrect or unusual images can adversely affect model training by introducing noise or bias into the training data. Once identified, these problematic images are removed from the dataset to ensure that the training data is of high quality and accurately represents the target objects and scenarios. By eliminating irrelevant or misleading data, data cleaning helps improve the overall reliability and effectiveness of the trained model.

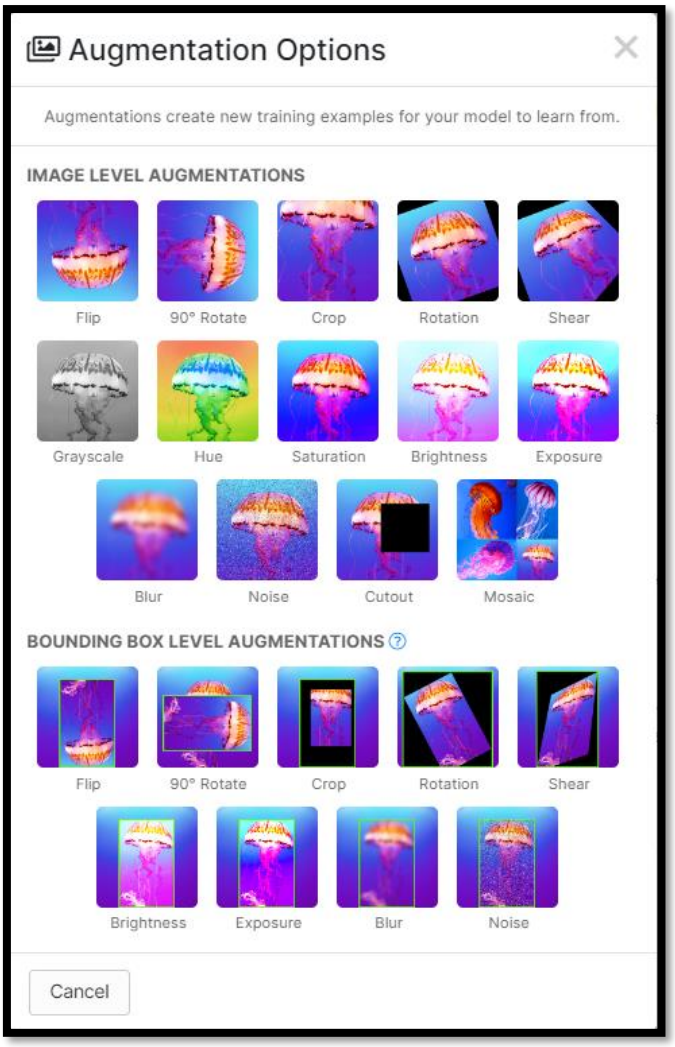
- **Data Annotations:**

  Data annotation refers to the process of labeling objects of interest within images or videos to create training data for object detection models. This annotation process involves marking bounding boxes around objects and assigning corresponding class labels to indicate the type of object present. For example, in a dataset for detecting vehicles, each image or frame in the dataset would be annotated with bounding boxes around individual vehicles (such as cars, trucks, or buses) and labeled with their respective class names. Similarly, in a dataset for safety hazard detection, annotations would mark bounding boxes around hazards like fire, smoke, or safety equipment violations. Data annotation is a critical step in training object detection models as it provides the necessary ground truth information for the model to learn and recognize objects during the training process. Proper and accurate annotation ensures that the model can effectively generalize and make accurate predictions on new, unseen data.



- **Data Augmentations:**

  Data augmentation in YOLOv5 involves artificially increasing the size and diversity of the training dataset by applying various transformations to the original images. This technique helps improve the generalization and robustness of the model by exposing it to a wider range of variations and scenarios that it may encounter during inference on real-world data. Some common data augmentation techniques used in object detection models are Random Cropping, Random Rotation, Random Flipping, Color Jittering, Random Scaling and Random Noise Addition. By applying these augmentation techniques, the training dataset becomes more diverse, allowing the model to learn from a wider range of scenarios. This ultimately leads to improved performance and generalization when the model is deployed in real-world settings.

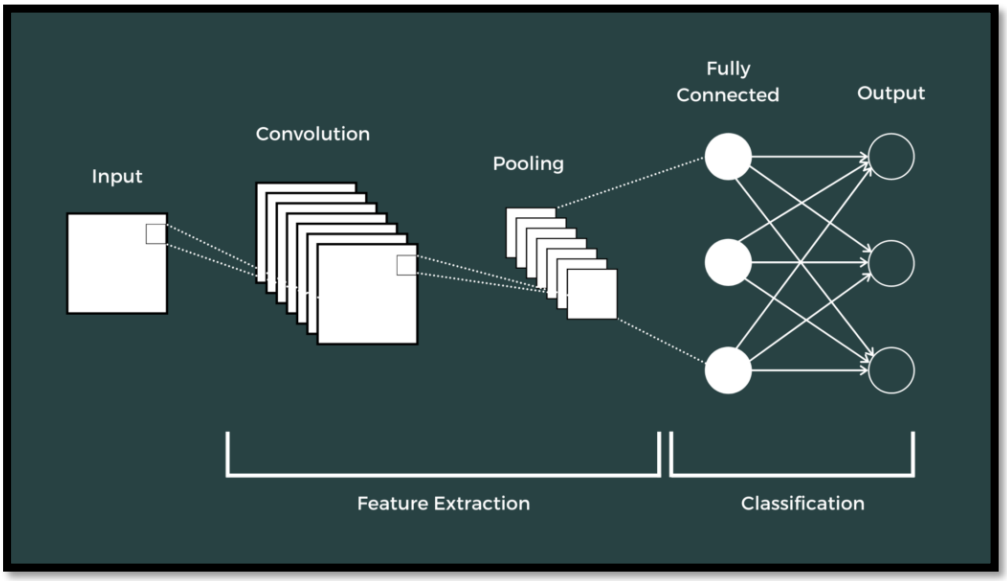


- **Data Resizing:**

  Data resizing is a common preprocessing step in training object detection models. It involves resizing all images in the dataset to a uniform size before feeding them into the model for training. The purpose of data resizing is to ensure that all images have consistent dimensions, which is necessary for the model to process them efficiently during training. Additionally, resizing helps optimize memory usage and computational resources, especially when working with large datasets.

5.5 Concepts & Algorithms Used

5.5.1 Convolutional Neural Network



Convolutional Neural Networks (CNNs) are a specific type of artificial neural network, particularly well-suited for image recognition and processing tasks. They are especially adept at finding patterns in data that has a grid-like structure, such as images.

Key characteristics of CNNs include:

- **Convolutional Layers:** These layers apply convolution operations to input images, extracting local patterns and features through filters or kernels. Convolution helps capture spatial hierarchies of features in an image.
- **Pooling Layers:** Pooling layers downsample the feature maps generated by convolutional layers, reducing their spatial dimensions. Max pooling, for example, retains the maximum value within each region, effectively reducing computational complexity while preserving important features.
- **Activation Functions:** Non-linear activation functions like ReLU (Rectified Linear Unit) introduce non-linearity into the network, allowing CNNs to learn complex relationships in data.
- **Fully Connected Layers:** Following convolutional and pooling layers, fully connected layers aggregate extracted features and

perform high-level reasoning tasks. They connect every neuron in one layer to every neuron in the next layer, enabling the network to learn intricate patterns and make predictions.

5.5.2 YOLOv5

YOLOv5, short for "You Only Look Once version 5," is an advanced object detection model renowned for its speed, accuracy, and simplicity. Developed by Ultralytics, YOLOv5 represents a significant evolution in the YOLO series of models.

Key features of YOLOv5 include:

- **Efficiency:** YOLOv5 achieves remarkable speed and efficiency, making it suitable for real-time applications.
- **Accuracy:** Despite its speed, YOLOv5 maintains high accuracy in detecting objects within images and videos.
- **Simplicity:** The model architecture is straightforward, making it easy to understand, implement, and customize for various tasks.
- **Scalability:** YOLOv5 is highly scalable and can be trained on different datasets with varying sizes and complexities.
- **Versatility:** It supports a wide range of object detection tasks, including detecting common objects, counting objects, and detecting specific attributes within objects.

Overall, YOLOv5 represents a state-of-the-art solution for object detection tasks, offering a balance between speed, accuracy, and simplicity that makes it ideal for a variety of applications in computer vision and beyond.

5.6 Model Selection

Use cases:

- Fire & Smoke Detection
- Personal Protective Equipment (PPE) Vest Detection
- Hardhat Detection
- Safety Gloves Detection
- Safety Shoes Detection

- Wrong Direction Vehicle Detection

Selecting the most appropriate deep learning models is crucial for the success of our Multimodal AI safety system. This section delves into the key considerations and candidate models for this critical stage.

Essential Requirements:

- **Object Detection Accuracy:** The core function of our system is to accurately detect specific objects like PPE equipment, fire/smoke, and vehicles traveling in the wrong direction. High precision is vital to minimize false positives that could trigger unnecessary alarms, and high recall is essential to ensure all relevant hazards are identified.

- **Real-Time Performance:** Industrial environments demand real-time processing for immediate response to potential safety threats. Models with fast inference speeds are necessary for timely detection and intervention.

- **Computational Efficiency:** Training and deploying these models might have resource constraints. Models that offer a balance between accuracy and efficiency are desirable.

Candidate Model Evaluation:

- **YOLOv5:** The pre-trained model achieves a strong balance between speed and accuracy, making it well-suited for real-time object detection tasks.

5.7 Model Training

Model Training includes the following:

- **Learning Rate:** The learning rate is a hyperparameter that determines the step size at which the model parameters are updated during training.

- **Batch Size:** The batch size refers to the number of training examples used in each iteration of the optimization algorithm during training. Choosing an appropriate batch size is essential for balancing computational efficiency and model performance. Larger batch sizes generally result in faster training times but may require more memory and computational resources. Smaller batch sizes, on the other hand, may lead to more stable

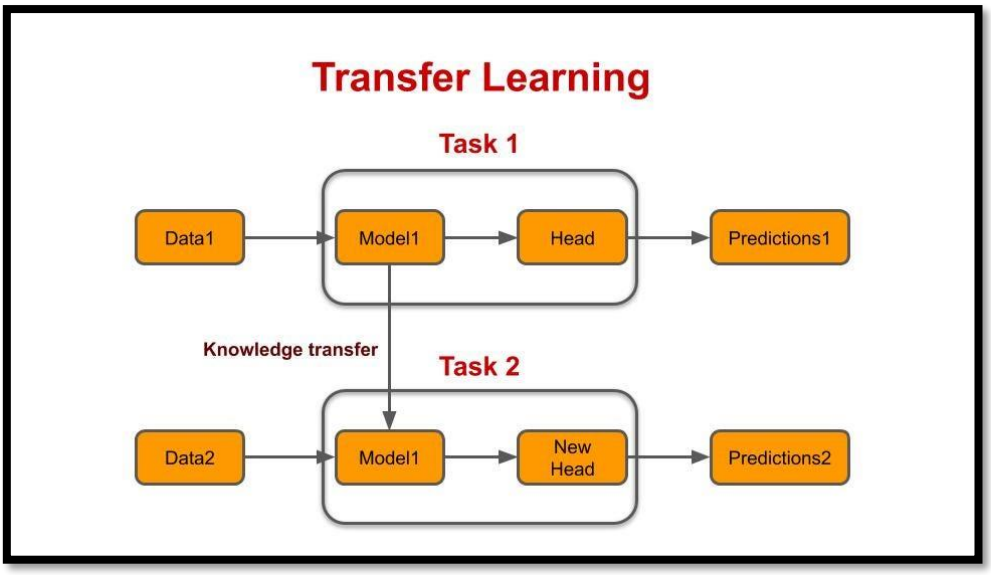
training dynamics and better generalization but require more iterations to converge.

- **Optimizer:** The optimizer is an algorithm that updates the model parameters based on the gradients of the loss function with respect to those parameters. Common optimizers include stochastic gradient descent (SGD) and Adam, each with its own advantages and disadvantages. Adam is a popular choice due to its adaptive learning rate capabilities and efficient handling of sparse gradients, making it well-suited for a wide range of tasks and architectures. However, the choice of optimizer may depend on factors such as the nature of the dataset, the model architecture, and the training objectives, and it is often beneficial to experiment with different optimizers to find the most effective one for a given task.

- **Regularization:** Regularization techniques are used to prevent overfitting and improve the generalization ability of machine learning models. Common regularization methods include L1 and L2 regularization, dropout, and batch normalization.

- **Hyperparameter Tuning:** Experiment with different hyperparameters such as learning rate, batch size, optimizer, and regularization techniques to optimize model performance. Hyperparameter tuning involves training multiple model configurations and selecting the one that yields the best results.

- **Transfer Learning:** Utilize transfer learning by fine-tuning pre-trained models on the target detection tasks. Transfer learning allows the model to leverage knowledge learned from a large, generic dataset to improve performance on a specific task with a smaller dataset.



5.8 Model Testing

Following the model training phase, a comprehensive testing process is crucial to ensure the robustness and generalizability of Multimodal AI safety system. This involves evaluating the model's performance on unseen data to assess its ability to function effectively in real-world industrial environments.

Testing Stages:

- **Test Set Evaluation:**
  This is the initial testing stage after training. A separate test set, meticulously carved out from the original dataset but unseen by the model during training, is used for evaluation. The model's performance on this test set provides insights into how well it has learned to generalize and detect objects in unseen scenarios. Metrics like mean average precision (mAP), false positive rate (FPR), and inference speed are measured on the test set. If the model performs well on the test set, achieving the desired accuracy and efficiency levels, it indicates promising generalizability.

- **Unseen Data Testing:**
  This stage goes beyond the test set and involves evaluating the model on entirely new, unseen data that was not part of the original dataset creation process. This data can be collected from different industrial sites or

represent scenarios not explicitly included in the training data. The goal is to assess how well the model adapts to real-world variations in lighting, object appearances, and environmental conditions that might not have been perfectly captured in the training dataset.

5.9 Model Deployment

After passing all the testing stages the models were finally deployed in the real world and continuously monitored for improvement.

Chapter 6:
Result & Discussion

Model: Fire & Smoke Detection

(The model can detect fire and smoke)

In the above image fire is detected in an industrial setup at 65% and 76% confidence & Smoke at 65% confidence.

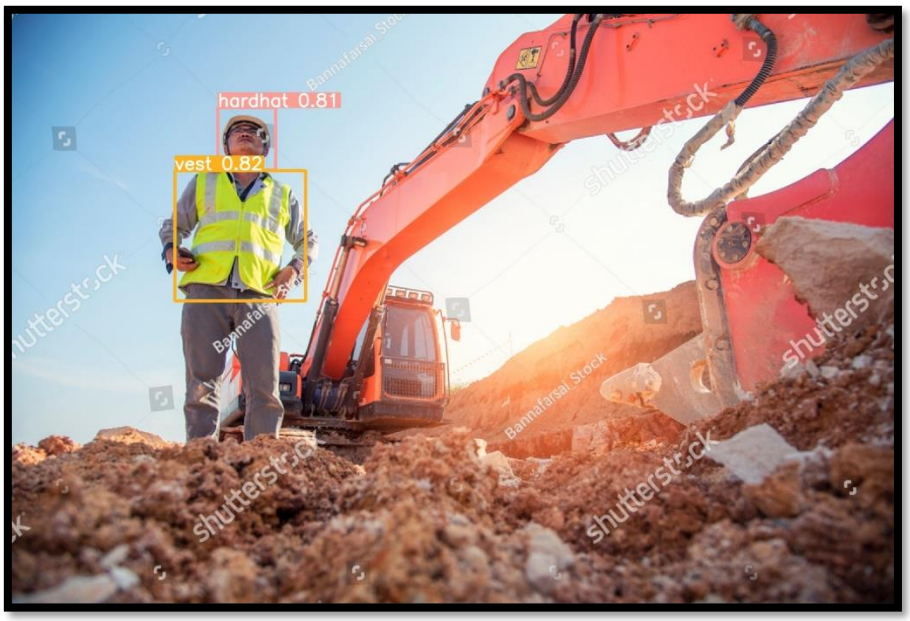
In the above image fire is detected at petrol pump at 70% confidence.

In the above image fire is detected in an industrial setup at 56%, 72% and 76% confidence & Smoke at 51% confidence.

In the above image fire is detected in an industrial setup at 64% confidence and Smoke at 64% confidence.

Model: Personal Protective Equipment (PPE) Vest Detection

(The model can detect if a person is wearing a PPE vest or not)



In the above image vest is detected at 82% confidence at a construction site.

In the above image no vest is detected at 80% confidence in an industrial office.

In the above image vest is detected at 86% and 92% at a construction site.

Model: Hardhat Detection Model

(The model can detect if a person is wearing hardhat or not)

In the above image the no hardhat is detected at 83% confidence in an industrial setup.

In the above image hardhat is detected at 80% and 85% confidence at a construction site.

Model: Safety Gloves Detection

(The model can detect if a person is wearing Safety Shoes or not)

In the above image safety gloves are detected at 82% and 84% in an industrial setup.

In the above image no gloves are detected at 65% and 66% at construction site.

Model: Safety Shoes Detection

(The model can detect if a person is wearing Safety Shoes or not)

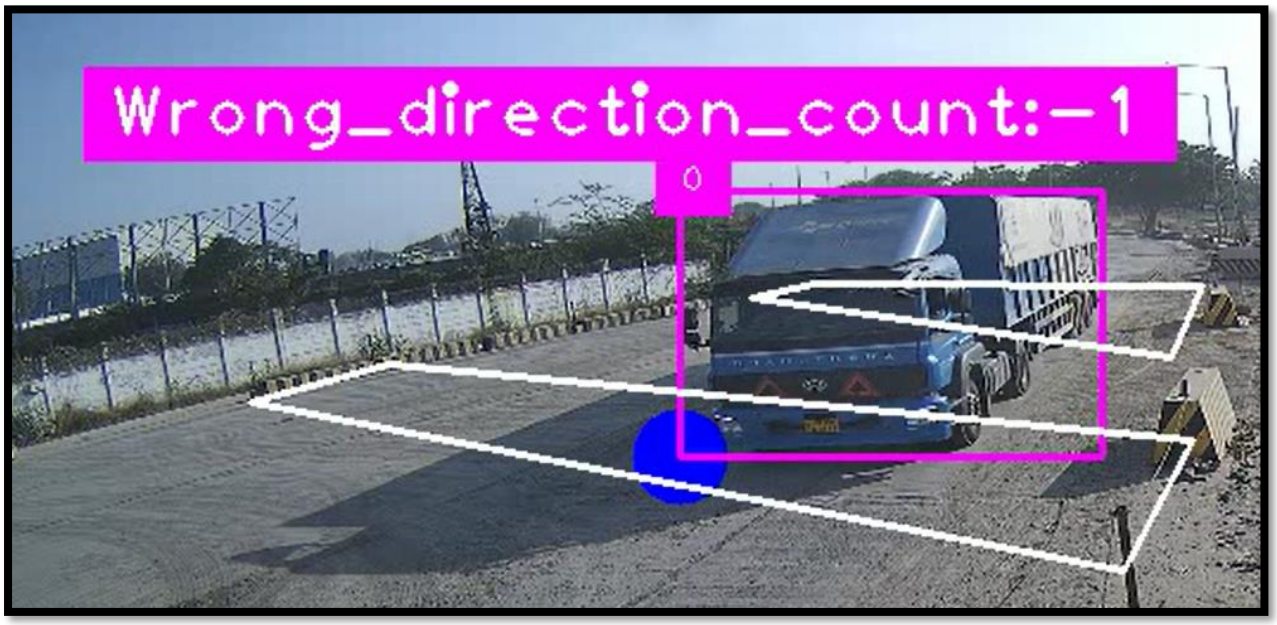
In the above image no safety shoes are detected at 70% and 78% confidence at a construction site.

In the above image safety shoes are detected at 80% confidence in an industrial setup.

Model: Wrong Direction Vehicle Detection

(The model can detect if vehicles are moving in correct direction or not)

As the vehicle will pass from box 1 (small box) its tracking will start.



As soon as the vehicle will pass from box 2 (large box) If its in wrong direction the count will get updated.

Chapter 7:
Conclusion

This project serves as the inception of a Vision-Powered Industrial Safety AI framework, leveraging existing infrastructure to revolutionize safety measures in industrial environments. By harnessing real-time object detection capabilities, the system proactively identifies safety hazards before accidents occur, rather than reacting after incidents have occurred. This proactive approach marks a significant departure from traditional safety methods, which often prioritize post-incident analysis.

The implementation of real-time object detection enables the system to continuously monitor industrial environments and swiftly detect potential safety threats. By integrating Vision-Based AI models, the system can analyze video feeds in real-time, identifying critical safety elements such as fire, smoke, improper use of personal protective equipment (PPE), and wrong-direction vehicle movement. This proactive detection mechanism empowers operators to take timely preventive measures, mitigating risks and preventing accidents before they escalate.

Moreover, the utilization of existing infrastructure ensures a cost-effective and scalable solution, enabling seamless integration into industrial operations without requiring extensive hardware investments. This approach not only maximizes the utility of existing resources but also facilitates widespread adoption across diverse industrial sectors.

In essence, this project lays the groundwork for a paradigm shift in industrial safety, where AI-driven technologies proactively safeguard personnel and assets by detecting and addressing safety hazards in real-time. By embracing this Vision-Powered Industrial Safety AI framework, industries can enhance operational resilience, minimize downtime, and prioritize the well-being of their workforce, ushering in a new era of proactive safety management.

Chapter 8:
Future Work

- **Advanced Data Privacy Measures:** Implementing advanced encryption techniques, access controls, and anonymization methods to ensure data privacy and security within industrial AI systems.

- **Robust Algorithm Development:** Enhancing the robustness and adaptability of AI algorithms to effectively handle complex and dynamic industrial environments, including variations in lighting, occlusions, and environmental factors.

- **Integration of Reinforcement Learning:** Exploring the integration of reinforcement learning techniques to optimize safety monitoring systems, enabling adaptive learning and continuous improvement of safety policies over time.

- **Multimodal Data Fusion:** Investigating the integration of multimodal data sources, such as sensor data and textual information, to enhance context-awareness and decision-making capabilities in industrial safety AI systems.

- **Human - Machine Collaboration:** Facilitating seamless collaboration between AI systems and human operators by designing user-friendly interfaces, incorporating explainable AI techniques, and optimizing decision support systems for informed safety decisions.

Chapter 9:
Bibliography

Bibliography

Cheng, S. e. (2020). AI-enabled industrial safety management: challenges, solutions, and future directions. *Industrial Informatics, IEEE Transactions on*, 16(8), 5065-5075.

Guo, S. e. (2020). Real-Time Monitoring and Early Warning System for Occupational Safety and Health Based on Multi-Source Data Fusion. *IEEE Access*, 8, 103169-103177.

Li, H. e. (2019). Deep learning-based safety protection system in industrial internet of things. *IEEE Transactions on Industrial Informatics*, 15(7), 4079-4087.

Liu, Q. e. (2022). Privacy-Preserving Machine Learning in Industrial IoT Environments: Challenges, Solutions, and Future Directions. *IEEE Internet of Things Journal*, 9(3), 2157-2169.

Wang, J. e. (2021). YOLOv5: A Benchmark for Real-Time Object Detection in Traffic Scenes. *arXiv preprint arXiv:2105.13208*.

What is Multimodal AI + Use cases for Multimodal AI. (2023, 12 1). Retrieved from Divi: https://skimai.com/what-is-multimodal-ai-use-cases-for-multimodal-ai/

Xu, Y. e. (2021). A Survey of Reinforcement Learning for Industrial Internet of Things: Foundations, Methods, and Applications. *IEEE Transactions on Industrial Informatics*, 17(8), 5595-5606.

Yang, Z. e. (2021). Real-Time Monitoring System Based on Computer Vision and Deep Learning for Industry 4.0. *IEEE Transactions on Industrial Informatics*, 17(7), 4600-4609.

YOLOv5. (2023). Retrieved from Ultralytics: https://docs.ultralytics.com/yolov5/

YOLOv8. (2023). Retrieved from Ultralytics: https://docs.ultralytics.com/

Zhang, L. e. (2020). Fire Detection in Industrial Environments using YOLO-Based Deep Learning. *IEEE Transactions on Industrial Informatics*, 17(2), 1050-1058.