Enhancing Autonomous Truck Navigation in Underground Mines. A Review of 3D Object Detection Systems, Challenges, and Future Trends

Ellen Essien [*] and Samuel Frimpong

Review

Enhancing Autonomous Truck Navigation in Underground Mines: A Review of 3D Object Detection Systems, Challenges, and Future Trends

Ellen Essien * and Samuel Frimpong

Missouri University of Science and Technology

* Correspondence: ee2df@mst.edu; Tel.: +1-573-612-5552

Abstract: Integrating autonomous haulage systems in underground mining has revolutionized safety and operational efficiency. However, deploying 3D detection systems for autonomous truck navigation in such an environment faces persistent challenges due to dust, occlusion, complex terrains, and low visibility. This affects their reliability and real-time processing. While existing reviews have discussed object detection techniques and sensor-based systems, providing valuable insights into their applications, only a few have addressed the unique underground challenges that affect 3D detection models. This review synthesizes the current advancements in 3D object detection models for underground autonomous truck navigation. It assesses deep learning algorithms, fusion techniques, multi-modal sensor suites, and limited datasets in an underground detection system. The study uses systematic database searches with selection criteria for relevance to underground perception. The findings of this work show that the mid-level fusion method to combine different sensor suites enhances robust detection. Though YOLO (You Only Look Once)-based detection models provide superior real-time performance, challenges persist in small object detection, computational trade-offs, and data scarcity. This paper concludes by identifying research gaps and proposing future directions for a more scalable and resilient underground perception system. The main novelty is its review of underground 3D detection systems in autonomous trucks.

Keywords: 3D object detection; autonomous trucks; deep learning; sensor fusion; underground mines; YOLO algorithms

1. Introduction

The evolution of autonomous driving systems in the mining sector has brought significant interest in enhancing computer vision for accurate and real-time 3D object detection. Unlike urban [1–3], or surface mining driving [4–6]. Underground mining environments present unique constraints such as limited visibility, dust, confined spaces, and uneven terrain that pose significant challenges for 3D object detection. These object detection systems are the perceptual backbone of autonomous truck haulage for obstacle recognition, object classification, and navigation in real-time.

Recent advancements in computer vision/image processing powered by deep learning (DL) and artificial intelligence (AI) have significantly improved situational awareness in autonomous driving systems [7]. AL/ML-based object detection and tracking models have seen applications in robotics [8–10], urban autonomous driving [4,11,12], collision avoidance systems [13,15], and security systems for monitoring and surveillance [15]. AI-ML techniques, particularly those involving deep learning, play a significant role in these systems [7,16]. They have addressed the challenges of machine-human interactions, particularly for collision prevention, injuries, and fatalities. In the underground environment, research increasingly implements AI/ML architectures with (LiDAR Light Detection and Ranging), thermal infrared (IR), and RGB (Red, Green, Blue) cameras to detect pedestrians, machinery, and hazards [5,6,17–20]. 3D object detection models, which give richer spatial information than 2D models, have become a critical area of research for safe autonomous truck navigation [21–

24]. The DL algorithms include Convolutional Neural Networks (CNNs) and YOLO (You Only Look Once), which have become efficient in identifying and classifying objects in real time [6,25,26].

These current models have demonstrated effectiveness in improving detection capabilities under diverse and harsh conditions [15,27,28]. However, they are limited due to small and occluded object detection inaccuracies amidst noise and obstructions. Also, they lack robustness when applied to specific environments such as underground mines, necessitating more innovative solutions [15,29].

As the mining industry transitions to Industry 4.0 and 5.0, which prioritize human-machine interactions, it is crucial to continually develop and innovate new systems, methods, and solutions for object detection and anti-collision systems in underground mining environments. Deploying autonomous vehicles into the underground mining industry is a paradigm shift toward operational efficiency and safety enhancement. These autonomous haulage trucks are designed to navigate dynamic and intricate mining environments, enhancing rapid decision-making accuracy and situational awareness. To ensure this, they heavily depend on sophisticated 3D object detection systems. However, the capabilities of these models are currently limited, which hinders the exploitation of their full potential in autonomous mining operations. Current detection systems frequently encounter limitations like trade-offs between speed and accuracy, especially under adverse conditions such as variable terrains, fluctuating lighting, and diverse objects [6,12,30]. Sensor modalities such as LiDAR, IR, and RGB cameras often result in suboptimal performance, particularly in dynamic obstacles and rapidly changing settings. Also, the real-time application capabilities of these detection systems remain insufficient, resulting in latency issues that could compromise operational efficiency and safety.

Many review articles such as [2,16,18,31–37] have explored the advancements of object detection models in autonomous vehicles and mining environments. Immam et al. [18] reviewed anti-collision systems based on computer vision in underground mines. The study examined machine learning algorithms, which included CNNs, Fast R-CNN, and YOLO series, for real-time object detection and sensors employed in autonomous trucks to reduce accidents. Tang et al. [38] studied multi-sensor fusion detection methods for 3D object detection in urban autonomous driving. They developed a taxonomy that categorized fusion approaches and assessed their efficacy in improving detection accuracy and safe driving in autonomous vehicle driving scenarios. Cui et al. [32] reviewed navigation and positioning technologies in underground coal mines. The study investigated multiple techniques, like visual image feature-based systems, inertial navigation, visible light communication (VLC), and sensor fusion methods, to improve the accuracy and robustness of the detection systems.

Patrucco et al. [33] underscored the importance of multi-sensor systems by integrating different sensors to address the limitations of individual sensors. The study also investigated anti-collision technologies like cameras, LiDAR, and radar, discussing their principles, advantages, limitations, and costs. Wang et al. [34] also reviewed the detection systems advancements in unmanned driving technology for coal mine transportation systems. They discussed multi-sensor fusion strategies to address the limitations of single sensors. Nevertheless, there is a limited examination of underground settings' complex and dynamic conditions and the scarcity of standardized datasets in detection models for benchmarking.

The underground mining environment requires robust sensor fusion techniques, low-latency processing, and models resilient to noise and occlusion, which conventional surveys overlook. A standardized dataset with underground conditions for practical model evaluation and performance also exists. This review paper addresses the lacuna by synthesizing the most recent developments in 3D object detection, sensor fusion strategies, and dataset challenges within underground autonomous haulage systems. It highlights the strengths, limitations, and suitability of various 3D detection systems for autonomous truck navigation in an underground environment. The objectives of this review are:

- To categorize and evaluate the various sensor modalities employed in underground autonomous haulage 3D detection systems
- To explore multi-sensor fusion approaches, their performance, and trade-offs

- To analyze and synthesize the deep learning architectures used in object detection models, particularly YOLO variants and CNNs.
- Identify key underground dataset limitations for object detection models.
- To identify significant challenges in underground autonomous truck object detection deployment and propose future directions for developing scalable and reliable detection systems.

This paper presents a comprehensive review to evaluate the current literature on 3D object detection systems for underground autonomous trucks. The process commenced with clearly defining the research questions, followed by a comprehensive search of pertinent studies in reputable academic databases, such as SpringerLink, Multidisciplinary Digital Publishing Institute (MDPI), ResearchGate, Institute of Electrical and Electronics (IEEE) Xplore, and Google Scholar. The search was conducted using keywords such as "3D object detection," "autonomous trucks," and " mining," in conjunction with relevant terms such as "pedestrian detection," "underground," and "object detection." There were defined criteria for inclusion and exclusion that guided the selection process. The studies considered for article inclusion were restricted to journals published in reputable academic sources, conference papers, and peer-reviewed articles. The only research included was recent, within the last 5-10 years, and specifically addressed 3D object detection in autonomous vehicles in underground mining environments. Data extracted from the selected studies were categorized and analyzed in key areas, including sensor modalities, 3D detection systems, multi-sensor fusion, fusion strategies, underground datasets, and specific mining challenges. In real-world scenarios, categorization enabled a comparative analysis of various approaches and their performance. The review followed established frameworks to guarantee consistency, transparency, data synthesis, and reporting comprehensiveness.

The review aims to enhance the reliability and robustness of autonomous systems in the underground mining industry by comprehensively analyzing 3D object detection methodologies in autonomous vehicles. The study employs a pragmatic approach, acknowledging that underground mining environments present distinct requirements that substantially differ from those in more controlled industrial environments. It is, therefore, essential to comprehensively evaluate and compare current 3D object detection techniques in such environments to ascertain their strengths, limitations, and application-specific challenges. This investigation will help discover current literature and technology limitations for developing more efficient detection systems that deliver real-time, precise, and resilient performance in underground mining environments. The study will emphasize recent advances and establish a basis for recommending future directions, particularly optimizing algorithms for rugged, resource-limited, and unpredictable mining environments. The study contributes to the advancement of autonomous haulage truck technology consistent with the mining industry's safety protocols for zero fatality and operational requirements, thereby fostering a more efficient and secure future for the sector.

The structure of the paper is organized as follows: Section 2 provides a comprehensive review of the key sensor-based 3D detection systems for autonomous truck navigation, multi-sensor detection systems, and their strengths and limitations. Section 3 delves into multi-sensor fusion strategies, grouping them into early, mid, and decision-level fusion, and assesses their impact on detection accuracy. Section 4 discusses current deep learning architectures, their deployments, and state-of-the-art underground 3D detection model comparisons. Section 5 synthesizes underground dataset challenges related to autonomous object detection. Section 6 identifies the key challenges detection systems face in underground settings and proposes future directions. Finally, Section 7 synthesizes the findings of this review. Figure 1 shows the structure of the entire review report.
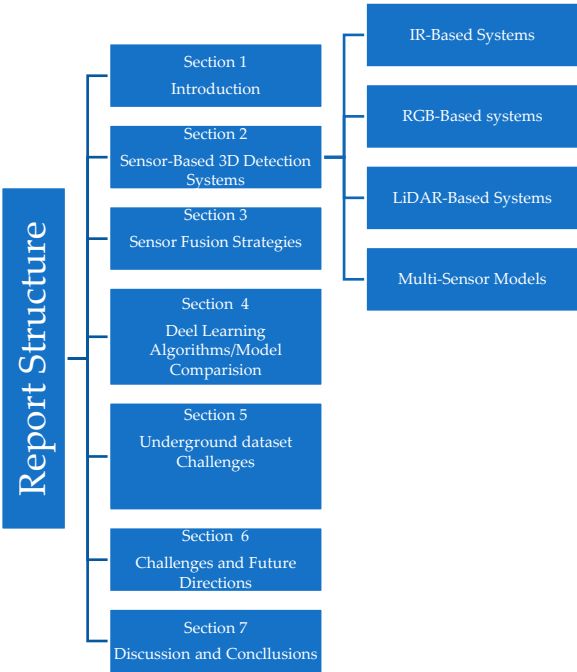
Figure 1. Structure of the Review Report.

2. Overview of Sensor Modalities for 3D Object Detection in Underground Autonomous Haulage Trucks

Three-dimensional (3D) Object Detection is pivotal to autonomous haulage truck operation in underground mining environments where safety, situational awareness, and navigation are critical. These detection systems rely on different sensor modalities to provide spatial awareness, track moving and static objects such as workers, equipment, and structural features, and detect hazards. Unlike controlled urban environments, underground settings have variable lighting, dust, occlusions, and uneven terrains, which necessitate the integration of complementary and robust sensors for detection.

This section provides a detailed analysis of key sensor-based detection perception systems used in underground autonomous trucks, which include IR, RGB cameras, and LiDAR systems. Each system is evaluated for its operational principles, integration with deep learning algorithms, underground applicability, and performance trade-offs.

2.1. Infrared Thermal (IR) Systems

Infrared (IR) thermal sensors (Figure 2) detect heat signals from objects and convert them into thermal images or temperature maps. IR's ability to perceive heat from objects or obstacles makes it significantly valuable for underground mining environments, where smoke, low light, and dust often compromise traditional optical systems. Unlike RGB cameras that depend on ambient lighting, thermal sensors can function in complete darkness and thermally unstable conditions.

These sensor modalities are effective in:

1. Pedestrian or worker detection by identifying humans' heat signatures.
2. Collision avoidance in situations where RGB cameras and LiDAR sensors struggle.

Figure 2. Thermal Infrared camera [39].

IR imagery can be segmented and classified in real-time when integrated with YOLO-based or CNN architecture. Many studies have demonstrated the successful application of different YOLO and two-stage CNN algorithms on thermal imagery for object detection in underground mining environments [18,40–44]. Figures 3 and 4 illustrate the application of IR with DL architectures in underground environments. Keza et al. [43] presented a pedestrian detection system that enhances safety in underground mines by integrating thermal imaging with 3D sensors. Using a FLIR thermal camera and depth sensors (TOF and Kinect), the system classifies regions using four methods and segments thermal images based on temperature thresholds. However, the model demonstrated susceptibility to motion distortion and mist and lacked enough underground datasets for model development.

Key Features and Advantages of IR Systems

- **Effective in Low or No Light Conditions:** IR systems rely on detecting heat signatures rather than visible light, making them highly effective in poorly lit or completely dark environments in underground mines.
- **Thermal Object Detection:** These sensors distinguish objects based on their heat signatures to identify equipment, vehicles, and workers, even in smoke, fog, or dust.
- **Long-Range Detection Capabilities:** Certain IR sensors, like long-wave infrared (LWIR), can detect objects over significant distances, enabling early identification of hazards or obstacles.
- **Insensitive to Ambient Light Variations:** Unlike RGB cameras, IR systems are unaffected by ambient light changes, ensuring consistent performance in dynamic lighting conditions.
- **Compact and Durable Designs:** IR sensors are lightweight, built to endure the harsh mining environment, and can withstand hot temperatures, vibrations, and dust.
- **Resistance to Ambient Light Variations:** Unlike RGB cameras, IR sensors are not affected by changes in ambient light, offering consistent performance both day and night.

Despite these strengths of IR sensor modalities, there are notable limitations:

- Limited spatial resolution. This often affects fine-grained object classification
- High purchase cost. IR sensors such as long-wave IR (LWIR) are often costly
- Inference from reflective equipment or heat surfaces with similar thermal profiles degrades their performance

Despite these limitations, their fusion into multisensory systems is invaluable, which enhances system robustness and reliability. When integrated with LiDAR (for depth information) and RGB (for texture), IR sensors contribute uniquely to real-time perception and classification.
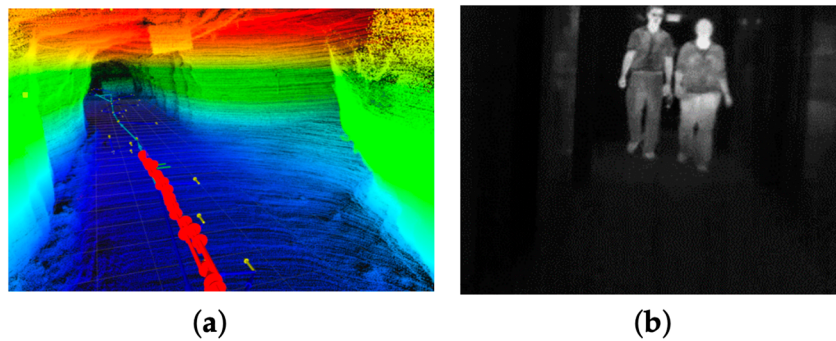
(a)　　　　　　　　　　　　　**(b)**

Figure 3. Thermal Sensor for Pedestrian Detection in Underground Operations (a) Navigating System [45], (b) Pedestrian Detection [46].

Figure 4. Thermal Imagery using YOLO Algorithm for Underground Pedestrian Detection [41].

2.2. RGB-Based 3D Detection Systems

RGB Cameras in Figure 5 remain one of the foundational sensor types utilized in 3D object detection systems because of their ability to capture high-resolution visual data with texture and color details [45,47–49]. They are deployed in underground mining autonomous trucks for object classification, scene understanding, and equipment recognition. RGB camera sensors illustrated in Figure 5 capture 2D images, which comprise red, green, and blue color channels combined to form a full-color image and a per-pixel depth report. In 3D object detection systems, RCB cameras are often used alone or with active stereo or time-of-flight sensing technology [18]. They are integrated with other sensors, such as LiDAR or IR sensors, for multi-sensor systems to enhance the perception model's scene understanding and robustness. Current detection algorithms, such as YOLO and CNNs, have significantly improved the accuracy and speed of RGB camera models [19,26,49,50]. These algorithms use features such as pattern recognition and color differentiation for precise detections and classifications.

RGB sensor detection systems are vital in autonomous mining trucks because they leverage visual data from cameras to identify, classify, and track objects in the environment. RGB-Camera has various configurations:

1. The monocular cameras, which are lightweight and cost-effective, do not have depth perception
2. Stereo cameras estimate depth by triangulating differences between two lenses
3. Depth cameras incorporate active sensors for near-field 3D imaging.

Figure 5. RGB Camera [18].

These sensor types are very compatible with DL models like YOLO and CNN. Zhang et al. [47] proposed an LDSI-YOLOv8 framework to enhance missed detection and low recognition for multiple targets in underground coal mines, leveraging an RGB camera. The work reported an accuracy of 91.4%, demonstrating a 4.3% increase compared to the original YOLOv8 algorithm. Imam et al. [51] also developed an anti-collision system for underground mines, focusing on pedestrian detection via RGB cameras with a YOLOv5 DL-based algorithm to enhance pedestrian detection accuracy in low-visibility conditions.

Philion and Fidler [52] also presented the Lift-Splat-Shoot model, which encodes multi-camera images into bird's-eye-view (BEV) representations for autonomous vehicle driving applications. The key objective was to create an end-to-end architecture that directly transforms multi-camera data into a unified BEV frame for semantic perception, understanding, and motion planning. The research demonstrated that the model could segment vehicles, drivable areas, and lanes by combining frustum-based depth inference and pooling techniques. However, there was reduced performance in low-light conditions and reliance on simple-frame data, which affects depth estimation accuracy compared to LiDAR-based models.

Versatility and affordability make camera-based models attractive for underground applications. They are less expensive than LiDAR and can be applied in areas such as identifying potential hazards and miner helmets for safety [17,53,54]. The key features and advantages of RGB Cameras include:

1. High-Resolution Imaging. RGB cameras can capture rich and detailed visual information, providing the necessary resolution to identify and classify objects based on color and texture.
2. Cost-Effectiveness. They are relatively inexpensive compared to other sensors, such as LiDAR sensors, making them a known choice for cost-effective object detection models. Their affordability enables widespread applicability in autonomous vehicles.
3. Color-Based Object Recognition. An additional layer of information is provided by the ability of RGB cameras to perceive colors, to distinguish between similar-shaped objects, and to identify warning signals.
4. Lightweight Design. RBG cameras are compact and easy to integrate into autonomous vehicle platforms, enabling more sensor placement, system design, and integration flexibility.

    However, there are significant challenges encountered:

1. High reliance on lighting makes them unsuitable for low-light environments
2. Affected by visual occlusions from debris and dust, which reduce detection accuracy
3. Lack of depth perception unless integrated with a stereo or LiDAR sensor.
4. High computational load for processing high-resolution image frames in real time.

    Multi-sensor fusion systems integrating RGB cameras with LiDAR or thermal sensor data address these limitations [25,28,29,54,55]. Fusion provides enhanced object classification and detection by combining color from RGB cameras with depth and range information from other sensors such as LiDAR. Autonomous systems achieve a more comprehensive and accurate perception of their environments by fusing this data from complementary sensor sources. Additionally, advancements in image technologies and algorithms have been introduced to address some of these challenges. These include high-dynamic range (HDR) imaging to improve the camera's ability to handle extreme variations in brightness and enhance visibility in poor and variable lighting

conditions [56–58]. ML algorithms are increasingly trained on augmented datasets to handle noisy and distorted visual data inputs better, improving model robustness in adverse conditions [26,29].

2.3. LiDAR (Light Detection and Ranging) System

The capability of LIDAR technology to generate precise depth measurements of the encompassing environment has made it a pivotal sensor type for detecting objects in autonomous vehicles. As shown in Figure 6, LIDAR sensors can accurately detect the distance between objects in a scanning area of up to about 50 meters [33]. The technology of LiDAR sensors has made significant strides since their inception, making it a critical component in detecting 3D objects for autonomous vehicle navigation. LiDAR technology was initially developed for military and atmospheric applications [59]. It has since improved to accommodate the needs of commercial and industrial sectors, such as the underground mine. LiDAR is the primary solution when incorporated with deep learning algorithms in mining, where the transition to autonomous operations requires real-time and precise detection. Solid-state designs, multi-beam configurations, and scanning mechanisms have significantly enhanced detection range, resolution, and robustness. This has addressed the limitations of early systems in challenging environments that depend on only single-point laser systems. LiDAR sensors emit laser pulses that travel through the atmosphere, bounce off objects, and measure their return time to the sensor, generating high-density point clouds representing 3D information about the object. This information includes object size, position, and orientation. This ability allows autonomous trucks to detect nearby equipment, obstacles, workers, and infrastructure and navigate in GPS-denied underground tunnels.
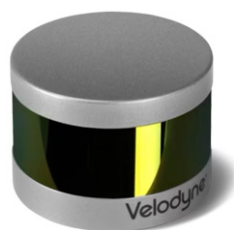

Figure 6. 3D LiDAR Sensor [46].

Modern LiDAR systems employ solid-state designs or rotating mirrors to perform rapid scans across 360 degrees of targeted areas. Data from these scans is processed in real-time to create dense point clouds, which serve as the foundation for object detection, tracking, and classification algorithms.

Key advantages of LiDAR systems in underground mining environments include:

1. High-Precision Depth Measurement. LiDAR sensors' time of flight range **ensures the precise localization of objects:** This makes it particularly useful in underground mining, where exact spatial awareness is critical for safety and overall operational efficiency.
2. **Resilience to Environmental Interference**: LiDAR is highly resistant to environmental interferences, such as vibration and glare. Advanced multi-echo and solid-state have built-in glare filters, enabling reliable performance in some of the underground mining environments.
3. **Wide Field of View and Real-Time Mapping**: Several LiDAR sensors offer a full 360° coverage view. This ensures comprehensive detection of objects, obstacles, and workers around mining vehicles, enhancing situational awareness.
4. **Efficient and Reliable Performance:** LiDAR sensors are unaffected by an object's color or texture, allowing consistent detection regardless of the visual appearance of objects. Unlike RGB cameras, LiDAR systems are effective in low-light conditions as they solely rely on active laser emissions rather than ambient light.

5. **Compatibility with SLAM systems**: LiDAR systems support simultaneous localization and mapping (SLAM), enabling autonomous haulage trucks to track their positions while dynamically building maps of the environment.

LiDAR produces precise depth measurements in the harsh and complex underground mining environment, making it significant. Advanced processing algorithms applied to LiDAR data can classify, detect, and predict objects in the environment for autonomous truck navigation. Integrating these algorithms with LiDAR can distinguish between stationary walls, moving equipment, and other objects, which is crucial for dynamic decision-making in real-time detection operations.

LiDAR sensors are essential for recent 3D object detection systems in autonomous vehicles in underground mining environments. LiDAR point cloud data comprises millions of points (Figure 7) that map the objects' surroundings in 3D (positions, size, and shape), providing a precise view beyond 2D images or videos, which is critical for accurately identifying objects in a scene. LiDAR-based systems' resilience in harsh environments is a significant advantage for their application. Underground mines are characterized by conditions impairing the performance of other sensor-based systems, like cameras or radar sensors. However, LiDAR systems, such as moisture, are less affected by these conditions. Combined with DL algorithms, LiDAR data can improve object detection accuracy and provide a robust obstacle avoidance mechanism for safe autonomous truck navigation. Compared to RGB data, LiDAR 3D point clouds are critical in providing structural and spatial information of precise depth. However, the 3D point clouds are unordered, sparse, and sensitive to local variations, which makes raw LiDAR data processing challenging [60].
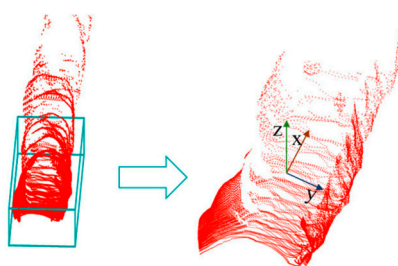

Figure 7. LiDAR Point Clouds for 3D Detection.

Generally, LiDAR object detection systems can be categorized into traditional and DL systems.

2.3.1. Traditional Methods.

Traditional methods often depend on geometric and clustering techniques to detect objects. Algorithms such as DBSCAN [61–63] assume that objects have a higher point density than their surroundings. This method groups nearby points into clusters, then applies shape fit to detect objects based on geometric assumptions. Though this approach is practical in structured environments and computationally efficient, it often struggles in noisy, complex, and unstructured environments. Usually, it fails to detect dynamic elements such as pedestrians or moving vehicles. They are more rule-based and best suited for controlled and simpler environments.

2.3.2. Deep-Learning-Based Methods

Recent technological advancements have enabled more accurate detection of complex real-world environments. These models automatically learn features from large, annotated datasets and are robust to occlusions and noisy inputs, providing accurate and more flexible detection in dynamic environments. LiDAR-based DL Systems are generally classified into view-based, voxel-based, point-based, and hybrid point-voxel-based methods [21,64–66].

1. **Voxel-Based Methods:** Voxel-based models transform raw and irregular 3D point clouds into structured grids known as voxels for processing using 3D convolutional neural [65–68]. This

generates simplified data, allowing the use of mature CNN architectures for object detection. Popular techniques include:

o VoxelNet [67]. This end-to-end framework partitions LiDAR point clouds into voxel grids and encodes object features using stacked voxel feature encoding (VFE) layers, as illustrated in Figure 8. This process is followed by 3D convolution to extract geometric and spatial features. The method streamlines the detection pipeline by eliminating separate feature extraction stages.

o SECOND (Sparsely Embedded Convolutions Detection) [66]:   This is an optimized version of VoxelNet that employs sparse 3D convolutions, which improves computational efficiency without sacrificing accuracy [69].

These models simplify point clouds' irregular and sparse nature and have high localization accuracy due to precise spatial encoding in the voxel grid. Additionally, their ability to process volumetric data directly makes them suitable for underground mine environments. However, the limitations of voxel-based methods include a lack of semantic richness, which limits their classification accuracy when implemented alone. Fin-grained voxelization can lead to high computational demand. Lastly, they struggle with small object detection due to limited resolution in sparse voxel grids.
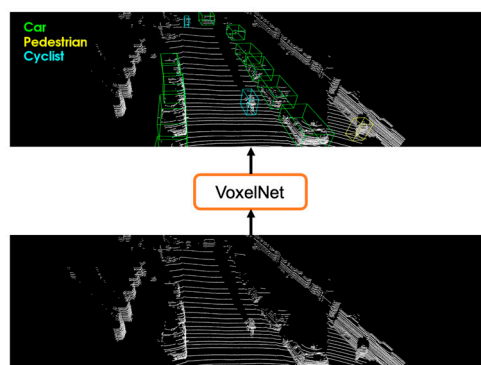

Figure 8. VoxelNet Processing of Raw Point Cloud Providing 3D Detection Results in Underground Environment [67].

2. **Point-Based Methods.** Point-based methods directly process raw LiDAR point clouds without converting them into voxel grids (voxelization) to 2D projections. This method preserves fine-grained geometric details of the environment by operating on ordered 3D points. This makes them highly effective in detecting partially occluded or irregularly shaped objects. The key pioneering family of point-based methods is PointNet [70]. This uses symmetric functions to learn object global shape features from unordered point sets. PointNett++ [71] extends this by using hierarchical feature learning to capture local context in clustered regions. PointPillars [72] (Figure 9) are a widely adopted method in autonomous driving. The method partitions point clouds into vertical columns known as "pillars" to offer a balance between detail preservation and computational efficiency. It converts spatial features into a pseudo-image format, which enables fast detection through 2D convolution while retaining meaningful 3D structures.

Despite their advantages in handling detailed geometry, point-based methods face limitations in computational cost, especially in cluttered environments, limited scalability in real-time, and the need for substantial GPU memory for training and inference
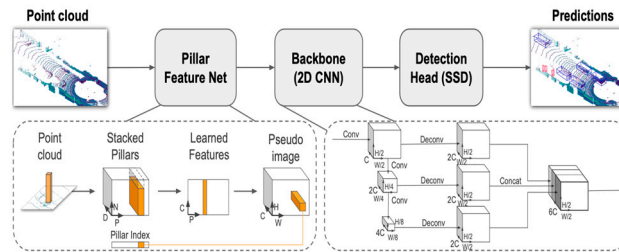
Figure 9. PointPillar Network Framework [72].

3. **Hybrid-Based Methods:** Hybrid approaches integrate the strengths of voxel-based and point-based techniques to enhance the accuracy and efficiency of 3D object detection. These methods utilize voxelization for structured data representation while leveraging fine local feature extraction of point-based models. A hybrid model typically starts by extracting local geometric object features from raw point clouds using point-based encoders. The features are then embedded into voxel grids, where the voxel-based backbone utilizes convolutions to learn global context and make predictions. This design helps the hybrid model to balance robustness, precision, and processing speed, making it suitable for cluttered and complex underground environments.

Despite the fast adaptation of LiDAR-based systems, several limitations impact their deployment in underground mining environments:

1. **Lack of Object Identification:** LiDAR systems can detect the presence and position of objects, but do not provide detailed information about object types or characteristics, which is essential for some mining applications.
2. **Computational Demand:** Processing of dense and sparse point clouds
3. **Environmental Sensitivity:** Accuracy is impacted by environmental conditions, such as dust, fog, water vapor, and snow. These factors degrade signal quality and require additional hardware and software processing to mitigate detection failures.
4. **Sensor Surface Contamination Challenges:** Dust and debris accumulating on the sensor surface impair detection functionality, necessitating regular cleaning or protective mechanisms to maintain reliability.
5. **High Cost. High-resolution** LiDAR units are expensive, which often hinders their large-scale deployment in budget-sensitive underground operations
6. **Range and Field of View Limitations:** The typical range of LiDAR sensors is limited to around 50 meters [29], restricting their effectiveness in larger underground mining operations. Furthermore, planar scanning systems may miss obstacles above or below the scanning plane, posing safety risks.
7. **Energy Consumption and Infrastructure Requirements:** LiDAR systems consume more energy than other sensors and demand robust infrastructure for effective operation, complicating deployment in confined underground mining spaces.

While models such as 3DSG [5,68] have demonstrated enhanced 3D object detection for surface mining trucks using LiDAR sensors, their direct applicability to underground environments remains limited due to differences in spatial constraints, lighting conditions, and operational challenges.

2.4. Multi-sensor Fusion in Underground Mining: Perception Enhancement Through Integration

Multi-sensor fusion models [73–75] are at the forefront of recent 3D detection systems for autonomous vehicle navigation in underground mining environments. These systems integrate data from multiple sensors such as cameras, LiDAR, radar, or thermal to develop a comprehensive perception of the environment. Each sensor type has its strengths and limitations, and by combining them, multisensory fusion systems capitalize on their complementary capabilities to enhance

detection accuracy and reliability for the safe navigation of autonomous trucks. LiDAR provides precise 3D spatial information, excelling in in-depth measurement, but often struggles in particulate-heavy environments. Cameras offer high visual data, capturing color and texture information for more detailed feature extraction, but their performance degrades in low-light or obscured conditions. Radar is reliable in adverse conditions and excels at long-range detection, but is less precise in resolution. Thermal sensors capture heat signatures from objects and surroundings, making them capable of detecting objects in completely dark and challenging conditions, such as fog, smoke, and dust. By fusing these data streams, multisensory models address the individual limitations of each sensor to create a robust perception framework suitable for challenging conditions in underground mines. As sensor technologies and fusion algorithms continue to evolve, they promise a future where autonomous trucks achieve even greater safety, efficiency, and adaptability in mining operations.

Szrek et al.[41] evaluated a UGV-based human detection system in underground mines using RGB and IR cameras alongside YOLOv3 and HOG algorithms. While RGB imagery provided visual context, it struggled in low-light and cluttered environments, especially for non-standing individuals. The study showed that RGB detection alone was often insufficient, but combining it with IR data improved reliability. A key limitation was using pre-trained models that were not optimized for underground settings, which affected accuracy. Xu et al. [76] proposed an autonomous vehicle localization method for underground coal mine tunnels based on fusing vision and ultrasonic sensors. They use infrared cameras to detect wall-mounted barcodes and ultrasonic sensors to measure distances, enabling geometric calculation of vehicle position without relying on complex SLAM. The method achieves sub-meter accuracy but is limited by dependence on manual barcode deployment and potential occlusion in dynamic tunnel environments.

Zhang et al. [77] also developed a real-time underground mining vehicle localization method by fusing YOLOv5-based object detection with high-precision laser distance measurement. The system identifies mining trucks visually using YOLOv5s and calculates exact positioning via laser sensors. However, the system had limitations, including its sensitivity to environmental conditions such as humidity and dust, which reduced detection robustness for small and fast-moving objects, with future improvements suggested through upgrading to YOLOv7/YOLOv8 and enhancing multi-object tracking capability. While recent advances have explored multi-sensor fusion using RGB, LiDAR, and thermal imagery with CNN or YOLO-based algorithms in environments such as search-and-rescue tunnels, urban settings, and surface mine environments, applications specifically targeting underground autonomous haulage trucks remain extremely limited. Most existing work focuses on drones, indicating a critical research gap for truck-scale 3D object detection and navigation in confined mining conditions.

Key advantages of multi-sensor fusion include:

1. **Enhanced Detection Accuracy:** Combining the complementary strengths of different sensors, like LiDAR's depth information with RGB camera's texture data and thermal imaging's heat signatures, will improve overall object detection performance.
2. **Robustness in Harsh Conditions:** Multi-sensor models maintain improved detection capabilities in challenging underground conditions where individual sensors may fail
3. **Redundancy for Safety:** These models provide multiple sources of information, allowing the system to continue functioning safely if one sensor fails or becomes unreliable.
4. **Improved Localization and Mapping:** Fusing vision (RGB Cameras) with depth measurements (LiDAR) strengthens localization precision and supports robust mapping in GPS-denied underground environments.
5. **Adaptability to Dynamic Conditions:** Provides dynamic sensor prioritization, where the system can rely more heavily on the most reliable sensors depending on environmental changes, like thermal sensors in low-visibility conditions and LiDAR for depth information.

With these capabilities of multi-sensor systems, notable challenges associated with them include

1. **High Computational Demand:** Processing large volumes of synchronized data from different sensors in real time requires powerful and often costly computational hardware.

2. **Complex Synchronization and Calibration:** Multi-sensor fusion requires precise calibration and alignment across different sensors, which is technically challenging, especially in underground settings with vibrations and environmental noise.

3. **Scarcity of Datasets for Training:** There is a lack of large, annotated, multi-sensor datasets that specifically capture underground environment conditions, limiting the ability to train robust deep learning models

4. **Increased System Weight:** Adding multiple high-end sensors such as LiDAR, thermal, and RGB cameras raises the overall cost and maintenance complexity and may impact autonomous truck payload or energy efficiency.

3. Sensor Data Fusion Methods

Fusing data from multi-modal sensors is vital in 3D object detection for underground autonomous trucks. The harsh underground conditions require robust perception systems capable of overcoming occlusions, dust, and low-visibility environments. Fusion methods are generally classified into early-stage (raw data), mid-stage (feature-level), and late-stage(decision-level) fusions depending on when the fusion occurs in the data preprocessing pipeline, as shown in Figures 10 and 11. Each strategy has distinct strengths and significant trade-offs, as described below. Figure 12 illustrates the accuracy and complexity levels of different fusion strategies of multi-sensor models.
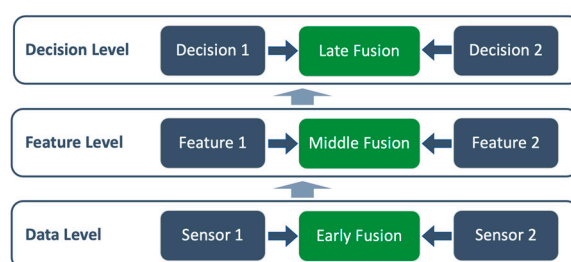

Figure 10. Sensor Level Fusion Stages in Multisensory Fusion [78].

3.1. Multi-Fusion Level Methods

3.1.1. Early-Stage Fusion

Early-stage fusion, or sensor-level fusion, integrates raw, unprocessed data streams from multiple sensors into a single dataset before performing feature extraction [79]. This method is valuable for creating robust detection systems in environments requiring fine-grained detection details, as it captures the full signal fidelity of individual sensor modalities.

For instance, LiDAR point clouds can be registered onto the RGB camera image pixel grid to generate depth-colored visual maps. Thermal sensor outputs can also be overlaid on visual data to identify heat-based anomalies. This strategy significantly detects small, obscured objects that a single sensor type may not fully recognize. This approach maximizes the amount of extracted information, which enables detailed feature extraction. The richness of the raw data captured ensures that critical details are not lost during the fusion process

The primary advantage of early-stage fusion is its ability to retain rich and detailed information across multiple sensor modalities. This gives an expressive input for DL models, enhancing detection precision and accuracy. However, this imposes significant computational demands, as raw sensor data requires huge memory and processing power. This computational burden causes a bottleneck to systems requiring real-time processing, such as the underground mine.

Moreover, data temporal misalignment from different sensors, resolution disparities, and distinct sensor frame rates introduce another challenge. This can be addressed by applying advanced calibration techniques and synchronization protocols for adequate alignment. Misaligned data can

lead to inaccuracies in the final fused dataset, undermining the system's reliability, accuracy, and effectiveness. Despite the earlier limitations, early fusion is essential in environments requiring precise spatial understanding, and computational overhead is not the primary constraint.

3.1.2. Mid-Level Fusion

Mid-level fusion involves the integration of independent extraction features from individual sensor modalities to create a more unified and comprehensive representation of the environment. Rather than performing a fusion of raw sensor data, each sensor undergoes preliminary processing via a convolutional neural network (CNN) or other feature encoders to extract high-level object representations such as contours, depth features, temperature gradients, and motion cues. These extracted features are consequently aligned and integrated into a shared feature space.

This fusion strategy provides an effective balance between perception richness and computational efficiency. By processing feature representations with low dimensionality, memory and bandwidth are reduced instead of the significant volume of raw data, while capturing rich information from each sensor modality. Mid-level fusion provides several advantages, including data efficiency. This reduction in computational requirements offers a more efficient system for real-time operations. Mid-level fusion is particularly relevant for real-time applications in underground autonomous trucks where edge deployment and split-second decision-making are critical.

Nevertheless, mid-level fusion introduces its challenges. There is difficulty in effectively aligning extracted features from heterogeneous sensors, which require spatial and temporal synchronization and are often trivial. Misalignments can degrade fusion quality and present detection errors. Additionally, fusion performance solely depends on the design and quality of the feature extraction modules as well as fusion architecture, such as cross-modal attention networks, which require careful design and tuning

3.1.3. Late-Stage Fusion

Late-level (decision-level) fusion integrates the final outputs of independently processed sensor data, such as bounding boxes, to make high-level decisions. Each sensor operates autonomously in this approach, and the outputs are generated using techniques like weighted averages or voting schemes to obtain final decisions. In [80], the study employed a decision fusion method utilizing camera and LiDAR sensors for mine track object detection. This strategy is favored for its simplicity and modularity, enabling sensors to be trained and maintained separately, simplifying the model design. The decision fusion approach is computationally lightweight, making it suitable for embedded systems or as a fallback in redundant safety layers. As individual sensors process their data autonomously, the fusion process occurs at the decision level, which does not require handling extensive raw data. This makes late-stage fusion suitable for resource-constrained environments and real-time processing applications.

Despite these advantages, there are limitations, such as overreliance on high-level decisions. Additionally, it lacks access to raw or intermediate data. It cannot refine ambiguities that are introduced earlier in the training pipeline. If one sensor misclassifies an object, the fusion process cannot correct it. This challenge often reduces overall detection accuracy and system performance, particularly in complex scenarios of late fusion models, especially in occluded and multiple overlapping scenes. Table 1 comprises the different sensor fusion strategies, strengths, and limitations in multi-sensor underground autonomous truck haulage models.
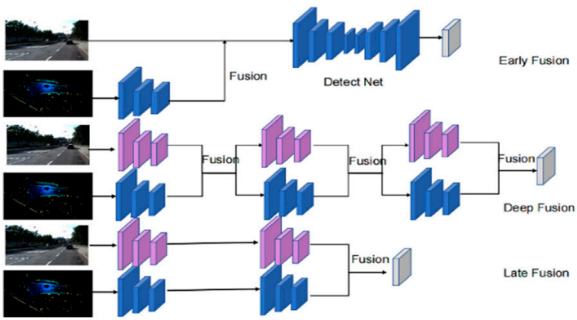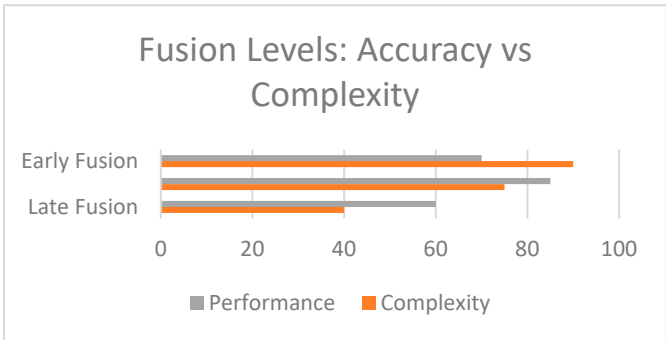
Figure 11. Multi-Fusion Level Methods [3].

Figure 12. Fusion Strategies Performance Comparison.

Table 1. Sensor Fusion Strategies in 3D Object Detection.

Fusion Approach	Level	Advantages	Limitations
Early-Level Fusion	Raw data	-Has rich, joint data representation -Fine-grained detection of small and obscured objects -Supports detailed spatial modeling -High-resolution outputs	-High computational cost, complex calibration -Raw data is sensitive to noise, distortions, and misalignments -High Memory -Demands precise spatial and temporal data alignment and calibration
Mid-Level Fusion	Feature Level	-Balances accuracy and efficiency -Reduction in data volume -Adaptable to real-time models -Can exploit diverse features across sensor modalities	-Requires accurate feature alignment and resolution compatibility -Often lose granularity from raw data - Has complex architecture tuning
Late-Stage Fusion	Decision Level	-No raw data alignment and its complexities - Simple to implement -Computationally efficient -Effective in redundancy layers -Suitable for real-time detection applications	-Cannot resolve early-stage mistakes - Heavy reliance on accurate individual sensor performance - Not suitable for applications requiring detailed integration of raw or feature-level data detection accuracy in complex scenes

This section and the comparison below demonstrate how each fusion strategy proves its unique strengths and limitations based on the deployment context. In complex underground mining environments, where safety-critical decisions are required in real-time, mid-level fusion often provides the best balance between detection accuracy and real-time decision-making. However, hybrid strategies that combine early and late fusion may also provide good results for fail-safe designs.

3.2. Sensor Fusion Advantages and Limitations

3.2.1. Advantages of Sensor Fusion in the Underground Mine Environment

Sensor fusion significantly enhances perception performance in autonomous haulage systems, especially within complex and dynamic underground mine environments. By integrating different sensor modalities such as LiDAR, camera, and IR, sensor fusion improves decision-making accuracy by overcoming individual sensor limitations. The key advantages include the following:

1. **Resilience in Harsh Conditions:** Sensor fusion compensates for individual sensor weaknesses. For example, LiDAR often degrades in dense particulate environments, cameras struggle in low light, and thermal sensors may lose range in open spaces. The fusion of these sensors provides a more consistent and robust perception across such extreme conditions.

2. **Enhanced Object Detection and Classification:** Multi-modal sensor integration improves object detection precision. Combining LiDAR and thermal data enhances the recognition of equipment, workers, and structural features, especially in low-visibility or occluded scenes.

3. **Improved Operational Safety:** Integrating multiple sensors increases system reliability and robustness, providing better object identification and more reliable obstacle avoidance. This ensures workers' and equipment safety in confined underground settings.

4. **Real-Time Decision Making**: Fusion systems enable faster and more informed decision-making, allowing real-time analysis of extreme environmental conditions. This enables autonomous trucks to respond promptly to dynamic changes in the environment.

5. **Model reliability and Redundancy:** Fusion systems enhance fault tolerance and robustness by ensuring that the failure of one sensor modality does not compromise the entire perception system

3.2.2. Challenges in Sensor Data Fusion

Despite the transformative benefits of fusion modalities, sensor fusion in underground autonomous haulage systems faces several technical and environmental challenges that complicate its implementation. These challenges must be addressed to ensure accurate, scalable, reliable, and real-time detection systems.

1. **Sensor Signal Reliability and Noise:** Underground environmental conditions, such as noise, vibrations, dust, and electromagnetic interference, can significantly introduce noise into sensor outputs, degrading performance. LiDAR sensors may produce inaccurate point clouds due to reflective surfaces, while RGB cameras may struggle in low-light conditions. Filtering this noise without losing essential features requires advanced adaptive filtering and denoising techniques.

2. **Heterogeneous Sensor Integration:** Each sensor type, such as LiDAR, cameras, radar, and IR, produces data in distinct resolutions, formats, and operational ranges. Integrating this heterogeneous input data requires overcoming challenges associated with data pre-processing, feature extraction, and data representation. For instance, fusing LiDAR point cloud data with pixel-based data from cameras involves significant computational effort and advanced algorithms to ensure meaningful integration and representation.

3. **Data Synchronization and Latency Management:** Precise synchronization of data streams from various sensors is crucial for real-time sensor fusion performance. Differences in signal delays, sampling rates, and processing times can lead to temporal data mismatches, resulting in inaccurate fusion outputs. For example, data from a camera capturing images at 30 frames per second (fps) must be aligned with LiDAR data that operates at a different frequency (e.g., 15 Hz). Sophisticated temporal alignment algorithms or interpolation methods are required to ensure all sensor inputs contribute meaningful data to the fused output.

4. **Computational Complexity.** The large volume of high-dimensional data that multiple sensors generate presents significant computational challenges. Real-time processing, critical for applications such as autonomous truck navigation, demands high-performance hardware and

optimized algorithms. Resource constraints, including limited processing power and energy availability in mining vehicles, further complicate this task. Efficient techniques that balance computational load with fusion accuracy are crucial for deployment in such environments.

5. **Environmental Factors:** Underground conditions pose severe challenges to sensor performance. Poor lighting, dust, smoke, noise, and fog can obscure camera, radar, and LiDAR data, while extreme temperatures can impact sensor calibration and accuracy. Additionally, confined spaces and irregular terrains can cause occlusions or reflections that lead to distorted sensor readings. Designing fusion algorithms that can compensate for these environmental factors is a critical area of research. Addressing these challenges requires interdisciplinary solutions drawn from signal processing and Machine Learning.

4. Algorithms for 3D Object Detection in Autonomous Trucks

The rapid evolution of ML algorithms has revolutionized numerous sectors, including autonomous vehicles and industrial robotics. It enables these systems to learn from experience, process complex sensory data, adapt, and make real-time decisions. Over the past decade, ML techniques have transitioned from basic pattern recognition to sophisticated algorithms that address complex, real-world challenges. AI/ML technologies provide automation, safety, and productivity advancements in the mining industry, particularly underground operations.

Historically, traditional object detection techniques such as Histogram of Oriented Gradients (HOG) and Support Vector Machines (SVM) struggled to perform under harsh underground conditions because of occluded or noisy environments. Developing more robust and adaptable DL algorithms, particularly Convolutional Neural Networks (CNNs) and YOLO, has emerged to address many of these challenges [81]. These algorithms allow for more precise, reliable, and adaptive perception in complex underground environments.

This section synthesizes major ML algorithms and their integration for 3D object detection in underground autonomous haulage navigation. The section emphasizes their architectural principles, recent applications in underground operations, and performance trade-offs.

4.1. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are the foundational technologies to modern computer vision because they can learn spatial hierarchies of image data features[82]. CNNs have proven resilient in underground environments due to poor lighting, dust, and occluded conditions.

CNN-based object detection models follow two-stage processes [83]. Girshick et al. [84] introduced the original R-CNN (Region-based), which utilizes selective search to propose regions of interest, followed by CNN-based classification. Though this method is accurate, it is computationally expensive and unsuitable for real-time applications. Subsequent models such as Fast R-CNN [85] and Faster R-CNN [86] directly used the region proposal stage in the CNN, minimizing latency and enabling near real-time applications. Mask R-CNN [87] further enhanced this by adding instance segmentation alongside detecting and classifying objects valuable for cluttered environments like underground mines.

Several studies have shown the use of CNNs for vehicle and personnel detection in underground tunnels, which validates the robustness of hierarchical feature extraction [83,86]. However, CNN algorithms remain challenged by their high computational load, local receptive fields that may miss the global context, and limitations in the 3D point cloud [24,88]. Lightweight CNN variants that employ quantization, pruning, and multispectral imaging have been explored to overcome these challenges. These techniques reduce model size and energy consumption capabilities, making CNNs suitable for edge deployment in underground autonomous haulage. However, challenges remain in meeting real-time performance and adaptability to complex underground scenarios in underground autonomous haulage. However, challenges remain in meeting real-time performance and adaptability to complex underground scenarios.

4.2. YOLO (You Only Look Once) Series Algorithm

YOLO (You Only Look Once) algorithms have gained prominence and recognition for their single-stage framework approach. This feature enables simultaneous object localization and classification in one network pass, delivering real-time detection inference critical for dynamic underground environments [89–92]. Redmon et al. [93] pioneered the introduction of YOLOv1 to reframe the detection task as a regression problem, significantly enhancing inference speed. Later versions, such as YOLOv3, YOLOv4, and YOLOv5, presented features including multi-scale detection (Darknet-53), cross-stage partial networks (CSPDarknet), and enhanced model flexibility [94]. YOLOv8, the latest version, includes lightweight designs, transformer-based enhancements, and an improved feature pyramid for better accuracy-speed balance. Figure 13 shows the traditional architecture of the YOLO framework.

Underground-specific YOLO model adaptations include:

Research has leveraged the YOLO algorithms to develop robust object detection models.[21].
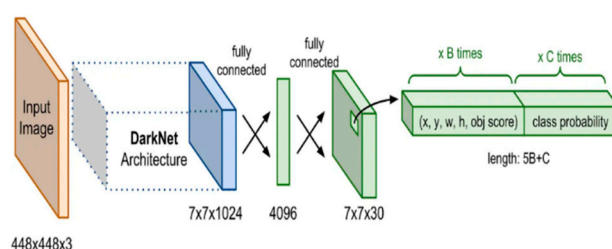

Figure 13. YOLO Architecture.

Zhang et al. [19] proposed LDSI YOLOv8 to address issues of multi-target detection in the coal mine excavation environment. The work leveraged the YOLOv8 architecture and had an mAP improvement of 4.3% compared to the original YOLOv8 model shown in Figure 14. The study achieved 91.4% mAP and 82.2 FPS, demonstrating strong adaptability to underground dusty, low-light, and occluded underground conditions. Despite its potential detection performance, the model's training on a specialized excavation dataset may limit its generalization to different mining sites without retraining or domain adaptation. Ní et al. [89] also developed a YOLOv8-based pedestrian and hazard detection model for underground mining environments, which achieves real-time capability and improved accuracy. Zhang et al. [95] proposed YOLO-UCM, an improved YOLOv5 model, to enhance pedestrian detection in underground coal. It integrated Vision Transformers (ViT) and Meta-AconC for enhanced feature extraction and detection accuracy.

DDEB-YOLOv5s model, which incorporated a C3-Dense feature extraction module, weighted BiFPN, and a decoupled detection head, improved feature extraction, achieving multi-target tracking accuracy of 93.1% and stability in mining environments [96]. Li et al. [97] presented an improved YOLOv11-based miner detection model for underground coal mines, enhancing feature extraction with Efficient Channel Attention (ECA) and refining localization with a weighted CIoU loss. The model achieves 95.8% at mAP@50 and 59.6 FPS on a custom underground dataset, outperforming existing detectors. However, it mainly focuses on personnel detection, with future work needed for broader underground object recognition.
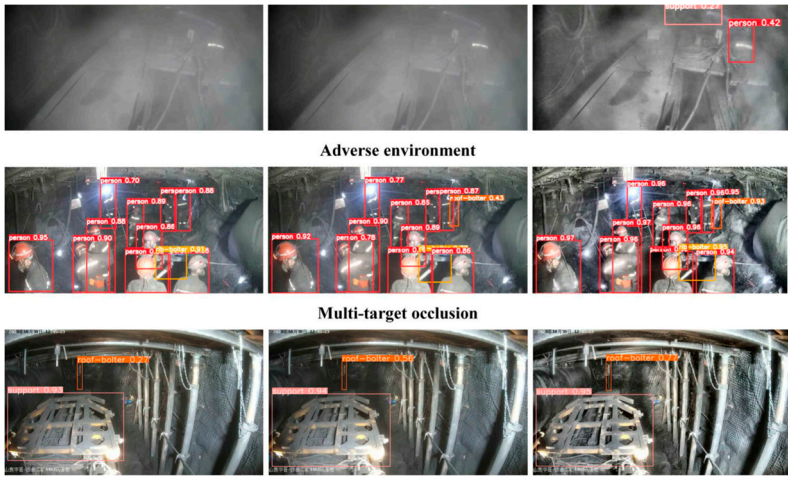
Figure 14. YOLOv8 detection model in an underground mining environment [19].

While YOLO models have achieved considerable success, they are challenged by difficulty detecting objects in highly occluded or cluttered conditions. This often requires additional computational resources as the scale increases. The trade-off between detection speed, accuracy, and model size remains a key challenge when deployed on embedded GPU systems in underground haulage trucks. Therefore, ongoing research aims to optimize YOLO's performance in such settings, ensuring that real-time, accurate detection remains feasible even with limited computational power. A performance comparison of YOLO and CNN algorithms is presented in Table 2. The table highlights key advantages and limitations of CNN-based and YOLO-based object detection frameworks concerning application constraints in underground settings characterized by low lighting, occlusion, computational overhead, and real-time processing inference.

Table 2. Comparison of YOLO and CNN DL Models for Object Detection in Underground Mining Environments.

DL-Approach	Advantages	Limitation
CNN	-Extracts hierarchical features effectively from image data - Effective in noisy and low-light underground conditions - Enhanced detection accuracy with multispectral data - Effective for segmentation and complex object shapes	-Computationally expensive, limiting real-time deployment - Need large, labeled datasets and have long training hours - Difficulty in deployment of embedded edge computers - Require large, labeled datasets for training - Prone to overfitting in complex and dynamic mine layouts
YOLO Series	-Single-stage architecture enables fast inference speed for real-time autonomous navigation - Well-structured Tracing of overlapping and moving objects - Lightweight versions are suitable for deployment on GPU edge platforms - Easy to fine-tune across variable mining datasets - Latest YOLOv8 variants support transformer-based attention	- Trade-offs in accuracy for small or occluded objects. - Reduced robustness without data augmentation or custom tuning practices - Improved performance requires extensive hyperparameter tuning - Limited generalization to unseen mining data - Persistent challenge by sensor misalignment and fusion latency

In summary, two-stage detection frameworks such as CNNs provide high detection accuracy but introduce computational overhead, making them unsuitable for real-time underground applications. On the other hand, YOLO-based models provide a viable direction for fast and efficient object detection, with recent versions demonstrating adaptability to underground conditions.

However, they often suffer in small object detection and tradeoffs between accuracy and inference speed. Continued efforts to enhance these models for edge computing and improve their robustness and real-time performance under occlusions and dynamic environments will be crucial to advancing autonomous haulage safety and efficiency in underground mining operations.

4.3. Detection Model Comparison for Underground Autonomous Haulage Systems

Current 3D object detection models designed for underground autonomous haulage systems navigation differ in structure, sensor dependency, and performance. The performance of 3D object detection systems is crucial for ensuring the safety and efficiency of autonomous trucks, particularly in complex and dynamic environments such as underground mines. Table 3 below evaluates and summarizes the performance of the leading 3D object detection models used in underground autonomous truck applications. It highlights their capabilities to handle the unique challenges posed by underground environments, detection approaches, advantages, limitations, and key performance metrics.

Table 3. Comparison of Recent 3D Detection Models for Underground Applications, where x means no information.

Model/Framework	Detection Algorithm	Sensor Modalities	mAP (%)	FPS	Limitations
LDSI-YOLOv8 [19]	YOLOv8n	RGB Camera	91.4	82.2	It has limited scalability in other mining environments.
YOLOv8 for Hazard Detection [89]	YOLOv8-based	RGB Camera	99.5	45	Limited robustness and generalization due to reliance on a small, self-constructed dataset
YOLO-UCM [95]	YOLOv5	RGB Camera	93.5	15	Model trained on a simulated dataset
DDEB-YOLOv5s + StrongSORT [96]	YOLOv5s with StrongSORT	RGB Camera	91.7	98	High model complexity - Requires significant computational resources - Reduced speed (98 FPS compared to lighter models)
YOLOv11-based Model [97]	YOLOv11	RGB Camera	95.8	59.6	Focuses mainly on personnel detection
Pedestrian Detection Model [98]	YOLOv5 (Deep Learning)	RGB Camera	71.6	x	Challenges with occlusion and detection in crowded scenes
Slim-YOLO-PR_KD [99]	YOLOv8s	RGB Camera	92.4	67	Scope limited to pedestrian detection

5. Dataset Analysis, Challenges, and Proposed Strategies

Annotated datasets tailored for underground environments are limited. Existing datasets are heavily focused on surface mine and urban driving conditions and are limited in diverse lighting, dust interference, and confined space conditions. While synthetic datasets for underground applications have been used as a cost-effective alternative, they often lack the complexities of real-world situations. Dynamic underground conditions create challenges that synthetic data fail to

replicate accurately, leading to poor model generalization and reduced performance when applied in actual mining environments. Real-world datasets are crucial for models to handle the varied and unpredictable nature of underground mining. The better the dataset reflects real underground scenarios, the more effectively the model detects objects accurately and reliably. Table 4 outlines the key underground-specific datasets and their characteristics. This section presents a comprehensive breakdown of challenges relating to datasets in underground 3D object detection and maps each challenge to implementable solutions. Addressing real-world underground constraints like dust, occlusion, sensor misalignments, and data imbalance will advance the frontier of autonomous truck navigation in an underground mine environment. This fills a significant gap in the literature and provides actionable insights for future system model designs.

Table 4. Underground Specific Dataset and Its Characteristics.

Dataset Name	Sensor Type(s)	Objects Annotated	Environment	Limitations
LDSI-YOLOv8 Excavation Scenes [19]	RGB Camera	Pedestrian	Underground coal mine	Limited scalability across diverse mining environments
Thermal image set [54]	Thermal IR	Workers, conveyor loads	Real coal mine	Scene specific Lacks scalability
YOLO-UCM [95]	RGB Camera	Pedestrians	Underground mines	Model trained on a simulated dataset; real underground variability may affect model performance
Real-time perception excavation dataset [96]	RGB Camera	Miners, Equipment	Excavation working faces in coal mines	Generalization to highly dynamic or new tunnel layouts is untested
MANAGEM Pedestrian Detection Model [98]	RGB Camera	Pedestrians	Underground coal mines	Sensitive to occlusion and crowded scenes

Recent detection frameworks in underground mining environments have achieved impressive results using RGB imagery, LiDAR, and thermal sensors. These rely solely on single data points and pose limitations under dust, fog, smoke, or complete dark conditions. Multi-sensor fusion offers a promising direction, integrating LiDAR for depth perception, thermal infrared for heat-based detection, and a camera for rich texture. Future research should prioritize developing fusion-based models that combine complementary sensor data, enhancing situational awareness and resilience for autonomous truck operations in the highly variable and constrained underground mining environment

5.1. Dataset Challenges in the Underground Environment

The development of scalable and effective 3D object detection systems for autonomous underground haulage trucks is hindered by significant dataset-related challenges:

1. **Environmental Complexity**: Underground environmental conditions pose significant challenges, which hinder effective dataset collection for detection models. Such challenges include poor lighting conditions, as these settings lack natural light. This leads to images with low contrast and clarity, making it difficult for optical sensors such as cameras to capture objects accurately. Additionally, dust and smoke from drilling, blasting, and transportation activities often scatter light and obscure sensors, limiting data quality. Furthermore, uneven terrain, cluttered backgrounds, and waste materials make distinguishing between detection objects and irrelevant objects challenging for safe detection.

2. **Data Annotation Challenges:** Annotating 3D data like LiDAR point clouds for detection models is particularly challenging as it's labor-intensive and needs expert knowledge of mining scenarios. Unlike 2D image annotation, identifying objects in 3D, especially in scenarios involving overlapping objects or partial occlusions, demands significant effort. The annotating process is prone to human errors and can significantly impact model performance.

3. **Suboptimal Dataset Representation:** Mining datasets are often characterized by issues related to overrepresenting common objects, such as trucks, and underrepresenting critical but rare classes, such as pedestrians. Frequently encountered objects dominate the dataset, while less common but critical elements, such as workers, are underrepresented in the data. This imbalance causes biases that prioritize common classes at the expense of rarer but safety-critical objects to be detected. This degrades model performance and limits generalizability.

4. **Inefficient Model Generalization:** A significant challenge in underground mining is the variability in mine layout and equipment types across different mines. This makes models trained in one mine less effective in another. The dynamic nature of mining operations necessitates continuous model updates that can adapt to variable scenarios.

5. **High Computational Complexity and Real-Time Constraints:** High-resolution sensors generate terabytes of data daily, which requires massive infrastructure for storage, robust computational efficiency, and real-time processing. Managing such large datasets while maintaining efficiency is a persistent issue. Latency in data processing can potentially compromise the system's effectiveness and safety-critical decisions.

6. **Temporal Synchronization in Multi-sensor Models:** Temporal synchronization is a significant challenge in multi-sensor 3D object detection models. Variability in sampling rates, operating modes, and data speed from different sensors causes a misaligned data stream, impairs fusion quality, and detection accuracy. Delays in transmission and hardware limitations worsen synchronization challenges.

5.2. Proposed Strategies for Dataset Optimization and Model Robustness in Underground

The following strategies are proposed to address these challenges:

1. **Enhance Sensor Capabilities:** Deploy robust sensors that withstand uneven terrain, dust, extreme vibrations, and temperature fluctuations to ensure consistent detection system performance. Integrate multi-sensor configurations by combining complementary sensing modalities. This will provide a robust and reliable detection system to enhance safety in underground mining operations.

2. **Advanced Data Preprocessing and Augmentation Techniques**: Effective preprocessing and data augmentation techniques should be employed to significantly simulate dust, noise, occlusion, and variabilities in lightning [7,9,79]. Synthetic dust clouds, altering textures, and adjusting brightness can help the system adapt to variable surface types or equipment, ensuring robust object detection models for challenging underground conditions.

3. **Improved Data Synchronization:** Solving temporal synchronization issues in multi-modal 3D object detection demands a combination of advanced computational methods and real-time data management strategies. Use software-based interpolation, a Kalman filter, or a deep learning-driven alignment framework to offer more flexible solutions for proper data synchronization. Catching strategies can also compensate for data transmission delays.

4. **Effective Data Annotation:** Leverage data annotation tools such as semi-automated tools, active learning, and pre-trained models [87]. This will reduce manual annotation effort and improve labeling accuracy by minimizing human error and maximizing efficiency. Additionally, domain-specific annotation guidelines can ensure consistency.

5. **Optimize Dataset Representation**: Accurate object detection requires balanced datasets. Applying class weighting, oversampling, and Generative Adversarial Networks (GAN)- based synthetic data generation will balance rare and common object instances in datasets. This approach ensures that objects like pedestrians and/or rare but critical events receive more

attention during training. This improves model robustness in accurately detecting safety-critical objects in real-world scenarios.

6. **Improvement in Model Generalization:** Employ domain adaptation techniques such as transfer learning and adversarial training to improve models' generalizability in cross-site applications. Consistent and continuous fine-tuning based on environmental feedback and characteristics enhances model adaptation.

7. **Edge Processing** and **Optimized Data Handling:** Efficient data handling is crucial in managing large data volumes. Use efficient compression techniques to reduce the size of datasets without sacrificing critical features and integrity. Employing edge computing to optimize data will enable real-time and on-site data preprocessing. This reduces latency and bandwidth usage, enhancing the system's ability to operate in real-time. It will reduce data transmission time and minimize the load on central systems, allowing for quicker decision-making.

6. Key Challenges and Future Directions In 3D Detection Systems for Underground Mines

6.1. Challenges in the Underground Environment.

The application of 3D object detection systems in underground mining environments is complex due to the unique and extreme conditions. These challenges can be categorized into environmental constraints, real-time operational intricacies, and computational limitations. Overcoming these barriers, especially with sensor fusion, real-time processing, detection accuracy, and pedestrian safety, will require further advancements in AI/ML, sensor technology, and edge computing. Future 3D detection systems in autonomous trucks must be intelligent, robust, and adaptive enough to operate safely and efficiently in unpredictable, diverse underground mining environments.

1. **Environmental Challenges.** Underground environments present near-zero natural light, which requires reliance on artificial illumination. This condition introduces variable lighting, low-light noise, shadows, and glare, adversely impacting vision-based systems. The prevalence of low-light imaging noise and incomplete data under artificial lighting conditions affects the performance of object detectors. IR modalities improve visibility but are limited in range and resolution. Dust, airborne particulates, and smoke generated by drilling, blasting, and other unit mining operational activities interfere with sensor signals[4]. These conditions degrade the quality of LiDAR point cloud data, leading to degraded object detection. Although radar sensors are less affected by these environmental factors, they lack resolution, necessitating multi-sensor fusion techniques to ensure robust detection. Creating fail-safes, redundancy in sensor data, and methods for ensuring detection capabilities and reliability are essential for reliable performance.

2. **Dynamic and Irregular Obstacles**. Continuous machinery and other object movements of other objects create a dynamic and unpredictable detection environment. Debris, uneven terrain, and narrow layouts can lead to misclassifications due to their resemblance to natural geological features in sensor data. Advanced semantic segmentation and computationally efficient ML-based classifiers are needed to address these challenges.

3. **Equipment Blind Spots and Sensor Occlusion.** Large mining trucks have extensive blind spots, particularly around corners and confined spaces, and occlusions from structures or materials often obscure key objects. This increases the risk of undetected obstacles, raising the likelihood of collisions causing fatalities. Multi-view and re-identification methods for occluded objects have shown improved continuity in detection. However, these solutions often introduce significant computational overheads, limiting real-time performance implementation in underground environments.

4. **Sensor Data Integration and Overload**. High-resolution LiDAR generates millions of data points per second, requiring advanced data fusion algorithms to integrate and process information from multiple sensors efficiently. Real-time data fusion across LiDAR, IR, and

camera presents a substantial computational load. Managing data overload is critical to ensuring timely and accurate object detection.

5. **Real-Time Latency Challenges.** The maximum need for split-second decision-making in autonomous trucks operating in high-risk environments means delays are unacceptable. Edge computing systems, which process data locally to reduce latency, have demonstrated effectiveness in mitigating this issue. However, challenges remain in the trade-offs between model complexity and detection accuracy. While YOLO algorithms have shown potential enhancement in 3D object detection, they are often affected by low-light, cluttered, small objects, and dynamic underground mine settings.

6. **Efficiency vs Accuracy Trade-Offs**. Large models give high accuracy but are computationally intensive and unsuitable for resource-constrained environments. Lightweight models such as YOLO-Nano have high speed but may lack robustness for detecting small or partially occluded objects [100].

7. **Generalization Across Mining Sites**. Every underground mining environment has unique variability in tunnel geometry, layout, infrastructure, machinery, and operational processes. Detection models trained in one site may not be scalable in another, which poses a significant obstacle to generalizing 3D object detection systems. Due to these variations, domain shifts cause models to fail to perform effectively on another site.

6.2. Future Research Directions

3D object detection remains crucial for both safety and efficiency. While significant progress has been made, challenges remain in optimizing algorithms, improving sensor fusion, and enabling real-time decision-making in dynamic underground environments. The following research directions are proposed to push the frontier of autonomous haulage detection systems:

1. **Multi-Sensor Fusion and Edge Computing**. Integrating data from sensors like LiDAR, IR, and cameras with edge computing is needed to reduce latency and improve real-time processing. Enhanced fusion techniques that combine high-resolution LiDAR data with camera visual information could provide more detailed and accurate object detection. Additionally, by processing data locally, autonomous trucks can make real-time decisions without relying on external servers, improving response times and operational efficiency.

2. **Collaborative Detection Systems**: Future autonomous trucks may not operate in isolation but as part of a larger network of autonomous systems, including other trucks, drones, and support equipment. Collaboration between these systems to share detection data and build shared scene understanding can improve object detection. Multi-agent communication and coordination could improve detection capabilities, object tracking, awareness, and cooperative navigation.

3. **Design Lightweight and Real-time DL Models**: Research into the development of optimized lightweight models for embedded GPUs in autonomous trucks. Advancements in pruning and quantization can minimize model size without significantly compromising detection accuracy.

4. **Real-Time Object Classification and Localization:** Improvements should be made in detection, real-time object classification, and 3D localization. Improved situational awareness enhances safe navigation and operational efficiency in underground mining operations.

5. **Development of Standardized, open-access datasets:** More effort is needed to develop large-scale, labeled datasets that are open-access and reflect underground mining conditions, which will improve research into more robust models for autonomous truck safe navigation. Simulation data combined with real underground mining footage will provide efficient training benchmarks.

6. **Advanced AI and ML Models**: Leveraging cutting-edge AI techniques, such as transfer learning, reinforcement learning, or deep reinforcement learning, could significantly enhance the adaptability and robustness of 3D object detection systems. These models could enable them to learn from diverse datasets and improve their detection capabilities, even in challenging environments like the underground mine. More sophisticated and newer versions of YOLO, like

YOLOv8, can be integrated to improve detection speed and accuracy for maximum system performance. Developing domain-adaptive and site-specific fine-tuning models capable of learning and generalizing across different environments needs much research focus.

7. **Field Validation of Object Detection Systems**: Validation of detection systems in operational underground sites to evaluate the model's real-world performance. Additionally, safety-critical metrics should be incorporated into the evaluation process to assess the developed models' real-world capabilities and safety impact.

8. **Regulatory and Ethical Compliance**: As autonomous trucks become more recognized in the mining industry, ensuring that object detection systems meet safety standards and regulatory requirements will be essential. Developing frameworks that align detection systems with safety regulations and ethical standards is crucial. This will ensure transparency in model design and performance benchmarking, which is key to operational and public trust.

9. **Long-Term Robustness and Reliability:** Long-term deployment of autonomous trucks in underground mines will require systems that can withstand harsh conditions, such as exposure to dust, vibration, and moisture. Ensure long-term stability of systems through real-time health monitoring, regular sensor updates, and robustness against sensor drift and wear. This will be essential to ensure continued safety and model performance.

Addressing these challenges and leveraging emerging research directions will enable the mining industry to accelerate the safe and efficient implementation of autonomous truck haulage in underground operations. This will mark a transformative growth in automation in industry.

7. Discussions

The application of 3D detection systems in underground autonomous trucks presents critical challenges and opportunities different from other domains like urban driving and indoor robotic machines. The underground mining environment is known for its complex terrain, occlusion, high particulate matter, limited visibility, and dynamic equipment-worker interactions. Therefore, it necessitates more specialized sensor configurations and robust perception models capable of operating in these extreme conditions. This paper reveals that multi-modal sensor fusion, especially combining LiDAR and RGB cameras, provides the most robust perception capabilities. While the camera contributes contextual texture data, LiDAR offers high-resolution depth information. When fused effectively, they compensate for each other's limitations, especially under low visibility or high dust conditions. Mid-level fusion is demonstrated to be the most balanced fusion approach, providing efficient feature integration while maintaining sufficient data richness for real-time object detection. Deep learning algorithms, particularly the YOLO-based framework, have shown strong performance in object detection tasks. Architectures such as YOLOv5 and YOLOv8 demonstrate real-time detection capabilities with significant mean Average Precision (mAP). However, small object detection and a lack of large-scale annotated datasets often hinder their performance in the underground environment. The following key trades were:

1. **Accuracy vs. Speed:** It was noticed that high-accuracy models like YOLOv8 often demonstrate slower inference times. This can be problematic for real-time underground navigation

2. **Sensor Cost vs. Redundancy:** Multi-sensor models improve robustness. However, they also increase hardware costs and integration complexity.

3. **Fusion Complexity vs. Benefit:** The early fusion approach provides detailed insights and features but is highly computationally expensive. Late fusion is computationally efficient but is less accurate.

Benchmarking for diverse datasets shows limited consistency in model evaluation, making it difficult to compare models directly. This shows the need for standardized mining-specific data sets and benchmarking protocols for efficient performance evaluation.

8. Conclusion

This review has comprehensively surveyed 3D object detection systems tailored explicitly for underground mining autonomous truck navigation. The review uniquely bridges the research gap between urban autonomous systems and the specific challenges of underground truck navigation. The study evaluates the current state of sensor modalities, detection algorithms, and fusion techniques. Additionally, it highlighted the unique constraints posed by underground complex conditions. Though technologies such as LiDAR, RGB cameras, and thermal sensors have proven individual strengths, integrating these sensors provides a promising path forward. Deep learning architecture, particularly YOLO-based object detectors, has demonstrated strong potential in real-time detection. However, challenges persist, including occlusion handling, limited underground datasets, computational overhead, and performance trade-offs. While significant strides have been made in 3D object detection for autonomous trucks in mining, ongoing innovation, and research are essential to overcoming the persisting challenges. These advancements will enhance operational efficiency and play a crucial role in safeguarding the well-being of workers.

References

1. J. Mao, S. Shi, X. Wang, and H. Li, "3D Object Detection for Autonomous Driving: A Comprehensive Survey," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.09474

2. L. Wang *et al.*, "Multi-Modal 3D Object Detection in Autonomous Driving: A Survey and Taxonomy," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 7, pp. 3781–3798, Jul. 2023, doi: 10.1109/TIV.2023.3264658.

3. H. Wang, X. Chen, Q. Yuan, and P. Liu, "A review of 3D object detection based on autonomous driving," 2024, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s00371-024-03480-6.

4. G. Wang, J. Wu, T. Xu, and B. Tian, "3D Vehicle Detection with RSU LiDAR for Autonomous Mine," *IEEE Trans Veh Technol*, vol. 70, no. 1, pp. 344–355, Jan. 2021, doi: 10.1109/TVT.2020.3048985.

5. H. Li *et al.*, "3DSG: A 3D LiDAR-Based Object Detection Method for Autonomous Mining Trucks Fusing Semantic and Geometric Features," *Applied Sciences (Switzerland)*, vol. 12, no. 23, Dec. 2022, doi: 10.3390/app122312444.

6. P. Peng, J. Pan, Z. Zhao, M. Xi, and L. Chen, "A Novel Obstacle Detection Method in Underground Mines Based on 3D LiDAR," *IEEE Access*, vol. 12, pp. 106685–106694, 2024, doi: 10.1109/ACCESS.2024.3437784.

7. P. Dakshinamoorthy, G. Rajaram, S. Garg, P. Murugan, A. Manimaran, and R. Sundar, "Artiicial Intelligence Algorithms For Object Detection and Recognition In video and Images," 2024, doi: 10.21203/rs.3.rs-3849848/v1.

8. Y. Dai, D. Kim, and K. Lee, "An Advanced Approach to Object Detection and Tracking in Robotics and Autonomous Vehicles Using YOLOv8 and LiDAR Data Fusion," *Electronics (Switzerland)*, vol. 13, no. 12, Jun. 2024, doi: 10.3390/electronics13122250.

9. J. Shahmoradi, E. Talebi, P. Roghanchi, and M. Hassanalian, "A comprehensive review of applications of drone technology in the mining industry," Sep. 01, 2020, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/drones4030034.

10. C. Addy, V. Sriram, S. Nadendla, and · Kwame Awuah-Offei, "YOLO-Based Miner Detection Using Thermal Images in Underground Mines," *Min Metall Explor*, 123AD, doi: 10.1007/s42461-025-01249-6.

11. G. Somua-Gyimah, G. Somua-Gyimah, S. Frimpong, and E. Gbadam, "A computer vision system for terrain recognition and object detection tasks in mining and construction environments," 2018. [Online]. Available: https://www.researchgate.net/publication/330130008

12. T. Rao, H. Xu, and T. Pan, "Pedestrian Detection Model in Underground Coal Mine Based on Active and Semi-supervised Learning," in *2023 8th International Conference on Signal and Image Processing, ICSIP 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 104–108. doi: 10.1109/ICSIP57908.2023.10270885.

13. M. W. Hanif, Z. Yu, R. Bashir, Z. Li, S. A. Farooq, and M. U. Sana, "A new network model for multiple object detection for autonomous vehicle detection in mining environment," *IET Image Process*, 2024, doi: 10.1049/ipr2.13173.

14. N. Gageik, P. Benz, and S. Montenegro, "Obstacle detection and collision avoidance for a UAV with complementary low-cost sensors," *IEEE Access*, vol. 3, pp. 599–609, 2015, doi: 10.1109/ACCESS.2015.2432455.

15. M. Kristo, M. Ivasic-Kos, and M. Pobar, "Thermal Object Detection in Difficult Weather Conditions Using YOLO," *IEEE Access*, vol. 8, pp. 125459–125476, 2020, doi: 10.1109/ACCESS.2020.3007481.

16. K. H. D. Tang, "Artificial Intelligence in Occupational Health and Safety Risk Management of Construction, Mining, and Oil and Gas Sectors: Advances and Prospects," *Journal of Engineering Research and Reports*, vol. 26, no. 6, pp. 241–253, May 2024, doi: 10.9734/jerr/2024/v26i61177.

17. D. P. Tripathy and C. K. Ala, "Identification of safety hazards in Indian underground coal mines," *Journal of Sustainable Mining*, vol. 17, no. 4, pp. 175–183, Jan. 2018, doi: 10.1016/j.jsm.2018.07.005.

18. M. Imam *et al.*, "The Future of Mine Safety: A Comprehensive Review of Anti-Collision Systems Based on Computer Vision in Underground Mines," May 01, 2023, *MDPI*. doi: 10.3390/s23094294.

19. Z. Zhang, L. Tao, L. Yao, J. Li, C. Li, and H. Wang, "LDSI-YOLOv8: Real-time detection method for multiple targets in coal mine excavation scenes," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3450582.

20. C. Li, G. Yao, T. Long, X. Yuan, and P. Li, "A Novel Method for 3D Object Detection in Open-Pit Mine Based on Hybrid Solid-State LiDAR Point Cloud," *J Sens*, vol. 2024, 2024, doi: 10.1155/2024/5854745.

21. P. H. Salmane *et al.*, "3D Object Detection for Self-Driving Cars Using Video and LiDAR: An Ablation Study," *Sensors*, vol. 23, no. 6, Mar. 2023, doi: 10.3390/s23063223.

22. J. Mao, S. Shi, X. Wang, and H. Li, "3D Object Detection for Autonomous Driving: A Comprehensive Survey," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.09474

23. P. Zhang, X. Li, X. Lin, and L. He, "A New Literature Review of 3D Object Detection on Autonomous Driving," 2025.

24. Y. Wang, S. Wang, Y. Li, and M. Liu, "A Comprehensive Review of 3D Object Detection in Autonomous Driving: Technological Advances and Future Directions," Aug. 2024, [Online]. Available: http://arxiv.org/abs/2408.16530

25. X. Xu *et al.*, "FusionRCNN: LiDAR-Camera Fusion for Two-Stage 3D Object Detection," *Remote Sens (Basel)*, vol. 15, no. 7, Apr. 2023, doi: 10.3390/rs15071839.

26. Z. Fu, J. Ling, X. Yuan, H. Li, H. Li, and Y. Li, "Yolov8n-FADS: A Study for Enhancing Miners' Helmet Detection Accuracy in Complex Underground Environments," *Sensors*, vol. 24, no. 12, Jun. 2024, doi: 10.3390/s24123767.

27. B. Das and P. Agrawal, "Object Detection for Self-Driving Car in Complex Traffic Scenarios," *MATEC Web of Conferences*, vol. 393, p. 04002, 2024, doi: 10.1051/matecconf/202439304002.

28. I. Ogunrinde and S. Bernadin, "Deep Camera–Radar Fusion with an Attention Framework for Autonomous Vehicle Vision in Foggy Weather Conditions," *Sensors*, vol. 23, no. 14, Jul. 2023, doi: 10.3390/s23146255.

29. Z. Ren, "Enhanced YOLOv8 Infrared Image Object Detection Method with SPD Module," 2024. [Online]. Available: https://www.woodyinternational.com/https://woodyinternational.com/index.php/jtpet/article/view/21.

30. F. Inostroza, I. Parra-Tsunekawa, and J. Ruiz-del-Solar, "Robust Localization for Underground Mining Vehicles: An Application in a Room and Pillar Mine," *Sensors*, vol. 23, no. 19, Oct. 2023, doi: 10.3390/s23198059.

31. D. Parekh *et al.*, "A Review on Autonomous Vehicles: Progress, Methods and Challenges," *Electronics (Switzerland)*, vol. 11, no. 14, Jul. 2022, doi: 10.3390/electronics11142162.

32. Y. Cui, S. Liu, and Q. Liu, "Navigation and positioning technology in underground coal mines and tunnels: A review," *The Journal of the Southern African Institute of Mining and Metallurgy*, vol. 121, 2021, doi: 10.17159/2411.

33. M. Patrucco, E. Pira, S. Pentimalli, R. Nebbia, and A. Sorlini, "Anti-collision systems in tunneling to improve effectiveness and safety in a system-quality approach: A review of the state of the art," Mar. 01, 2021, *MDPI AG*. doi: 10.3390/infrastructures6030042.

34. M. Wang, J. Bao, X. Yuan, Y. Yin, and S. Khalid, "Research Status and Development Trend of Unmanned Driving Technology in Coal Mine Transportation," Dec. 01, 2022, *MDPI*. doi: 10.3390/en15239133.

35. Y. Du, H. Zhang, L. Liang, J. Zhang, and B. Song, "Applications of Machine Vision in Coal Mine Fully Mechanized Tunneling Faces: A Review," 2023, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2023.3317048.

36. K. Wang, T. Zhou, X. Li, and F. Ren, "Performance and Challenges of 3D Object Detection Methods in Complex Scenes for Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1699–1716, Feb. 2023, doi: 10.1109/TIV.2022.3213796.

37. M. Contreras, A. Jain, N. P. Bhatt, A. Banerjee, and E. Hashemi, "A survey on 3D object detection in real time for autonomous driving," 2024, *Frontiers Media SA*. doi: 10.3389/frobt.2024.1212070.

38. Y. Tang, H. He, Y. Wang, Z. Mao, and H. Wang, "Multi-modality 3D object detection in autonomous driving: A review," *Neurocomputing*, vol. 553, Oct. 2023, doi: 10.1016/j.neucom.2023.126587.

39. "Tau 2 | Teledyne FLIR." Accessed: Apr. 26, 2025. [Online]. Available: https://www.flir.fr/products/tau-2/

40. "IEEE Xplore Full-Text PDF:" Accessed: May 04, 2025. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6072167

41. J. Szrek, R. Zimroz, J. Wodecki, A. Michalak, M. Góralczyk, and M. Worsa-Kozak, "Application of the infrared thermography and unmanned ground vehicle for rescue action support in underground mine— the amicos project," *Remote Sens (Basel)*, vol. 13, no. 1, pp. 1–20, Jan. 2021, doi: 10.3390/rs13010069.

42. X. Dai, X. Yuan, and X. Wei, "TIRNet: Object detection in thermal infrared images for autonomous driving," *Applied Intelligence*, vol. 51, no. 3, pp. 1244–1261, Mar. 2021, doi: 10.1007/s10489-020-01882-2.

43. N. Kazi and D. Parasar, "Human identification using thermal sensing inside mines," in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, Institute of Electrical and Electronics Engineers Inc., May 2021, pp. 608–615. doi: 10.1109/ICICCS51141.2021.9432093.

44. J. S. Dickens, M. A. Van Wyk, and J. J. Green, "Pedestrian detection for underground mine vehicles using thermal images," in *IEEE AFRICON Conference*, 2011. doi: 10.1109/AFRCON.2011.6072167.

45. *2019 IEEE Aerospace Conference*. IEEE, 2019.

46. M. Imam *et al.*, "The Future of Mine Safety: A Comprehensive Review of Anti-Collision Systems Based on Computer Vision in Underground Mines," May 01, 2023, *MDPI*. doi: 10.3390/s23094294.

47. H. Lou *et al.*, "DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor," *Electronics (Switzerland)*, vol. 12, no. 10, May 2023, doi: 10.3390/electronics12102323.

48. Q. Liu, H. Ye, S. Wang, and Z. Xu, "YOLOv8-CB: Dense Pedestrian Detection Algorithm Based on In-Vehicle Camera," *Electronics (Switzerland)*, vol. 13, no. 1, Jan. 2024, doi: 10.3390/electronics13010236.

49. M. Apoorva, N. M. Shanbhogue, S. S. Hegde, Y. P. Rao, and L. Chaitanya, "RGB Camera Based Object Detection and Object Co-ordinate Extraction," in *2022 IEEE 7th International conference for Convergence in Technology, I2CT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/I2CT54291.2022.9825065.

50. Rahul and B. B. Nair, "Camera-based object detection, identification and distance estimation," in *Proceedings - 2nd International Conference on Micro-Electronics and Telecommunication Engineering, ICMETE 2018*, Institute of Electrical and Electronics Engineers Inc., Sep. 2018, pp. 203–205. doi: 10.1109/ICMETE.2018.00052.

51. M. Imam *et al.*, "Anti-Collision System for Accident Prevention in Underground Mines using Computer Vision," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Oct. 2022, pp. 94–101. doi: 10.1145/3571560.3571574.

52. J. Philion and S. Fidler, "Lift, Splat, Shoot: Encoding Images From Arbitrary Camera Rigs by Implicitly Unprojecting to 3D," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05711

53. K. Huang, S. Li, F. Cai, and R. Zhou, "Detection of Large Foreign Objects on Coal Mine Belt Conveyor Based on Improved," *Processes*, vol. 11, no. 8, Aug. 2023, doi: 10.3390/pr11082469.

54. F. Uth, B. Polnik, W. Kurpiel, P. Kriegsch, R. Baltes, and E. Clausen, "AN INNOVATIVE PERSON DETECTION SYSTEM BASED ON THERMAL IMAGING CAMERAS DEDICATE FOR UNDERGROUND BELT CONVEYORS," *Mining Science*, vol. 26, pp. 263–276, 2019, doi: 10.5277/msc192618.

55. S. Li, K. Geng, G. Yin, Z. Wang, and M. Qian, "MVMM: Multiview Multimodal 3-D Object Detection for Autonomous Driving," *IEEE Trans Industr Inform*, vol. 20, no. 1, pp. 845–853, Jan. 2024, doi: 10.1109/TII.2023.3263274.

56. S. Seo, Y. Ko, and M. Chung, "Evaluation of Field Applicability of High-Speed 3D Digital Image Correlation for Shock Vibration Measurement in Underground Mining," Jul. 01, 2022, *MDPI*. doi: 10.3390/rs14133133.

57. P. Xu, Z. Zhou, and Z. Geng, "Safety monitoring method of moving target in underground coal mine based on computer vision processing," 2022.

58. S. K. Nayar and T. Mitsunaga, "High Dynamic Range Imaging: Spatially Varying Pixel Exposures £."

59. J. Aldrighettoni and M. Grazia D'urso, "Number 2, 1-10 ACTA IMEKO | www," 2023. [Online]. Available: www.zrc-sazu.si/en/rvt

60. Z. Yang, Y. Sun, S. Liu, X. Shen, J. Jia, and Y. Lab, "STD: Sparse-to-Dense 3D Object Detector for Point Cloud," 2019.

61. H. T. Nguyen, E. H. Lee, C. H. Bae, and S. Lee, "Multiple object detection based on clustering and deep learning methods," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–14, Aug. 2020, doi: 10.3390/s20164424.

62. S. Lee, S. An, R. Kim, J. Oh, and S. E. Lee, "Point Cloud Clustering System with DBSCAN Algorithm for Low-Resolution LiDAR," in *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICCE59016.2024.10444271.

63. M. El Yabroudi, K. Awedat, R. C. Chabaan, O. Abudayyeh, and I. Abdel-Qader, "Adaptive DBSCAN LiDAR Point Cloud Clustering For Autonomous Driving Applications," in *IEEE International Conference on Electro Information Technology*, IEEE Computer Society, 2022, pp. 221–224. doi: 10.1109/eIT53891.2022.9814025.

64. M. Zhu, Y. Gong, C. Tian, and Z. Zhu, "A Systematic Survey of Transformer-Based 3D Object Detection for Autonomous Driving: Methods, Challenges and Trends," Aug. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/drones8080412.

65. B. Yang, W. Luo, and R. Urtasun, "PIXOR: Real-time 3D Object Detection from Point Clouds," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2018, pp. 7652–7660. doi: 10.1109/CVPR.2018.00798.

66. Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors (Switzerland)*, vol. 18, no. 10, Oct. 2018, doi: 10.3390/s18103337.

67. Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection."

68. G. Park, J. Koh, J. Kim, J. Moon, and J. W. Choi, "LiDAR-Based 3D Temporal Object Detection via Motion-Aware LiDAR Feature Fusion," *Sensors*, vol. 24, no. 14, Jul. 2024, doi: 10.3390/s24144667.

69. Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors (Switzerland)*, vol. 18, no. 10, Oct. 2018, doi: 10.3390/s18103337.

70. C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation."

71. C. R. Q. Li, Y. Hao, S. Leonidas, and J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space."

72. A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection from Point Clouds," 2019. [Online]. Available: https://github.com/nutonomy/second.pytorch

73. H. Liu, W. Pan, Y. Hu, C. Li, X. Yuan, and T. Long, "A Detection and Tracking Method Based on Heterogeneous Multi-Sensor Fusion for Unmanned Mining Trucks," *Sensors*, vol. 22, no. 16, Aug. 2022, doi: 10.3390/s22165989.

74. P. Wei, L. Cagle, T. Reza, J. Ball, and J. Gafford, "LiDAR and camera detection fusion in a real-time industrial multi-sensor collision avoidance system," *Electronics (Switzerland)*, vol. 7, no. 6, Jun. 2018, doi: 10.3390/electronics7060084.

75. P. Trybała *et al.*, "MIN3D Dataset: MultI-seNsor 3D Mapping with an Unmanned Ground Vehicle," *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 91, no. 6, pp. 425–442, Dec. 2023, doi: 10.1007/s41064-023-00260-0.

76. Z. Xu, W. Yang, K. You, W. Li, and Y. Il Kim, "Vehicle autonomous localization in local area of coal mine tunnel based on vision sensors and ultrasonic sensors," *PLoS One*, vol. 12, no. 1, Jan. 2017, doi: 10.1371/journal.pone.0171012.

77. L. Zhang, X. Li, Y. Sun, J. Liu, and Y. Xu, "Research on Positioning and Tracking Method of Intelligent Mine Car in Underground Mine Based on YOLOv5 Algorithm and Laser Sensor Fusion," *Sustainability (Switzerland)*, vol. 17, no. 2, Jan. 2025, doi: 10.3390/su17020542.

78. M. R. Nabati, "Sensor Fusion for Object Detection and Tracking in Autonomous Sensor Fusion for Object Detection and Tracking in Autonomous Vehicles Vehicles." [Online]. Available: https://trace.tennessee.edu/utk_graddiss

79. M. Haris and A. Glowacz, "Navigating an Automated Driving Vehicle via the Early Fusion of Multi-Modality," *Sensors*, vol. 22, no. 4, Feb. 2022, doi: 10.3390/s22041425.

80. B. Liu, B. Tian, and J. Qiao, "Mine track obstacle detection method based on information fusion," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2022. doi: 10.1088/1742-6596/2229/1/012023.

81. W. Zimmer, E. Ercelik, X. Zhou, X. J. D. Ortiz, and A. Knoll, "A Survey of Robust 3D Object Detection Methods in Point Clouds," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2204.00106

82. M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Jul. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/machines11070677.

83. D. G. Schneider and M. R. Stemmer, "CNN-Based Multi-Object Detection and Segmentation in 3D LiDAR Data for Dynamic Industrial Environments," *Robotics*, vol. 13, no. 12, Dec. 2024, doi: 10.3390/robotics13120174.

84. R. Girshick, J. Donahue, T. Darrell, J. Malik, and E. Mercan, "R-CNN for Object Detection Outline 1. Problem Statement: Object Detection (and Segmentation) 2. Background: DPM, Selective Search, Regionlets 3. Method overview 4. Evaluation 5. Extensions to DPM and RGB-D 6. Discussion," 20214.

85. R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Dec. 2015, pp. 1440–1448. doi: 10.1109/ICCV.2015.169.

86. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

87. K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, Institute of Electrical and Electronics Engineers Inc., Dec. 2017, pp. 2980–2988. doi: 10.1109/ICCV.2017.322.

88. D. Lu, Q. Xie, M. Wei, K. Gao, L. Xu, and J. Li, "Transformers in 3D Point Clouds: A Survey," May 2022, [Online]. Available: http://arxiv.org/abs/2205.07417

89. Y. Ni, J. Huo, Y. Hou, J. Wang, and P. Guo, "Detection of Underground Dangerous Area Based on Improving YOLOV8," *Electronics (Switzerland)*, vol. 13, no. 3, Feb. 2024, doi: 10.3390/electronics13030623.

90. T. Ye *et al.*, "An adaptive focused target feature fusion network for detection of foreign bodies in coal flow," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 8, pp. 2777–2791, Aug. 2023, doi: 10.1007/s13042-023-01798-6.

91. M. W. Hanif, Z. Yu, R. Bashir, Z. Li, S. A. Farooq, and M. U. Sana, "A new network model for multiple object detection for autonomous vehicle detection in mining environment," *IET Image Process*, Oct. 2024, doi: 10.1049/ipr2.13173.

92. Z. Song, X. Qing, M. Zhou, and Y. Men, "Mine underground object detection algorithm based on TTFNet and anchor-free," *Open Computer Science*, vol. 14, no. 1, Jan. 2024, doi: 10.1515/comp-2024-0015.

93. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

94. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10934

95. Y. Zhang and Y. Zhou, "YOLOv5 Based Pedestrian Safety Detection in Underground Coal Mines," in *2021 IEEE International Conference on Robotics and Biomimetics, ROBIO 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1700–1705. doi: 10.1109/ROBIO54168.2021.9739594.

96. D. Zhao, G. Su, G. Cheng, P. Wang, W. Chen, and Y. Yang, "Research on real-time perception method of key targets in the comprehensive excavation working face of coal mine," *Meas Sci Technol*, vol. 35, no. 1, Jan. 2024, doi: 10.1088/1361-6501/ad060e.

97. Y. Li, H. Yan, D. Li, and H. Wang, "Robust Miner Detection in Challenging Underground Environments: An Improved YOLOv11 Approach," *Applied Sciences (Switzerland)*, vol. 14, no. 24, Dec. 2024, doi: 10.3390/app142411700.

98. M. Imam *et al.*, "Anti-Collision System for Accident Prevention in Underground Mines using Computer Vision," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Oct. 2022, pp. 94–101. doi: 10.1145/3571560.3571574.

99. H. Mu, J. Liu, Y. Guan, W. Chen, T. Xu, and Z. Wang, "Slim-YOLO-PR_KD: an efficient pose-varied object detection method for underground coal mine," *J Real Time Image Process*, vol. 21, no. 5, Oct. 2024, doi: 10.1007/s11554-024-01539-0.

100. N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2411.00201