Fraunhofer IDMT

Photo: istockphoto.com/NDinfinity

White Paper Construction-s**AI**t

Multi-modal AI-driven technologies for automatic construction site monitoring

In the construction industry, unexpected accidents, thefts, and vandalism can cause high costs due to construction delays. Also, logistic monitoring is of central importance to ensure a smooth construction process, especially for large-scale projects. Manual attempts to ensure the safety of construction workers and to monitor traffic flows and deliveries of construction materials are often costly and inefficient. As a promising solution, audio-visual analysis algorithms combined with distributed sensor networks across construction sites can help to automate these processes.

In recent years with the help of deep multi-layer neural networks, a methodology called deep learning has revolutionized the field of machine learning and artificial intelligence (AI) in general. Nowadays in the fields of computer vision and machine listening, algorithms driven by artificial intelligence are constantly improved and can already compete with the human's ability to see and listen. Inspired by human perception, the combination of multiple sensor modalities in a multi-modal approach is most promising since it allows to balance the individual weaknesses of a pure visual or acoustic analysis in particular application scenarios.

In this white paper, ten use-cases for construction site monitoring using multi-modal AI-driven algorithms are presented and challenges as well as possible solutions are discussed.

Proposed Use Cases

Use Case 1: Safety of Construction Workers

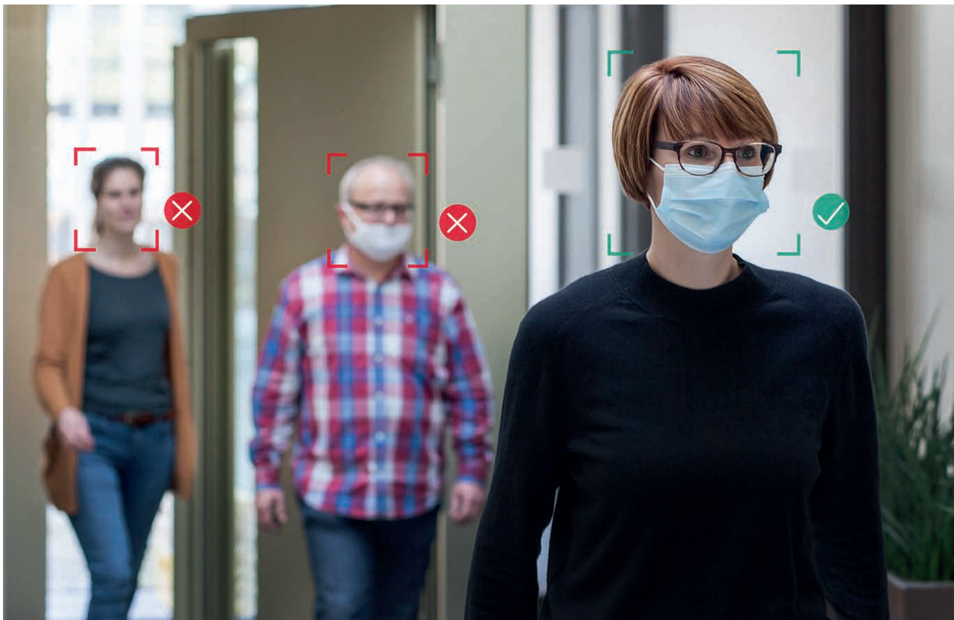
According to [1], the frequency of accidents at construction sites is about twice as high as at all other workplaces. What's more, the consequences of accidents are usually much more serious. The topic of occupational safety is therefore even more important on construction sites. Thus, extensive legal regulations have been established to ensure safety on construction sites. One the one hand, the architect, the site manager, the builder, the coordinator, as well as the contractors and their safety specialists are responsible for occupational health and safety on the construction site itself. On the other hand, company owners are primarily in charge of the occupational safety of their employees and the compliance with existing rules and regulations. Recent advances in the field of computer vision and machine listening on embedded computing devices, i.e., low-power mini computers designed for accelerating machine learning applications, can help to meet these guidelines.

UC1.1: Construction Hard Hat Verification
In 2019 alone, the German Social Accident Insurance recorded more than 44,200 reportable occupational accidents involving head injuries (excluding eye injuries) and 91 fatal accidents [2]. In several cases, a hard hat could have prevented worse. In some accident locations, hard hats were even mandatory. Modern deep learning based visual object detectors are capable of robustly detecting different object types, including hard hats, in real time. Thus, violations of the obligation to wear a hard hat can be quickly detected in order to initiate countermeasures. Furthermore, small autonomous verification devices could be positioned next to entrances of the construction site or at scaffolds for instance to ensure that hard hats are worn correctly. However, insufficient lighting or challenging backlight as well as non-frontal faces and occlusion might pose challenges to the system and therefore must be considered.

UC1.2: Mouthguard Verification
No matter what kind of work is carried out in the construction industry, dusts and other minute particles are generated by a wide variety of work processes and can affect the health of the workers as well as other people in the vicinity. Not only since the emergence of the Corona virus has the wearing of face masks and mouthguards been recommended on many construction sites [3]. The AI-based MaskCognizer, which was developed by Fraunhofer IDMT, is a technology that detects and verifies in real-time whether people are wearing their face masks correctly [4]. It classifies up to ten faces simultaneously and works even when people are in motion. Although the MaskCognizer technology was specifically developed for monitoring compliance with public health and safety regulations for the ongoing Corona crisis, adaption to the task of mouthguard verification as well as to the detection of different types of face masks is possible. The same challenges as outlined in UC1.1 do apply for this use case.

UC1.3: Protective Vest Verification
High-visibility vests and clothing should be standard equipment on construction sites since they protect the wearer in hectic road and rail traffic. Even on construction sites that are located away from public traffic, one should wear conspicuous high-visibility vests, because any excavator driver and the like can recognize the endangered persons more easily. The legislator prescribes the wearing of high-visibility clothing for activities in road traffic. Safety vests must have a CE marking and the current DIN EN ISO 20471 [5] test number which, among other things, regulates that only yellow,

The AI-based MaskCognizer detects in real-time whether people are wearing their face masks correctly



red, and red-orange fluorophores may be used. Therefore, by utilizing these properties, automatically and accurately detecting high-visibility vests using state-of-the-art computer vision methods is undoubtedly possible and could assist in complying with regulations for safety at construction sites. As already stated for use cases 1.1 and 1.2., occlusions, difficult body poses, and insufficient lighting at night or twilight pose difficulties to the technology and have thus to be considered.

UC 1.4: Detection of Cries for Help
Despite protective measures for working safety, working incidents can occur. Based on audio streams that are recorded at various locations across the construction site acoustic events such as human screaming sounds and cries for help can be detected. This is especially relevant when injured workers are out of earshot of their colleagues. As a result, rescue measures such as notifying co-workers and calling an ambulance can be rapidly initiated.

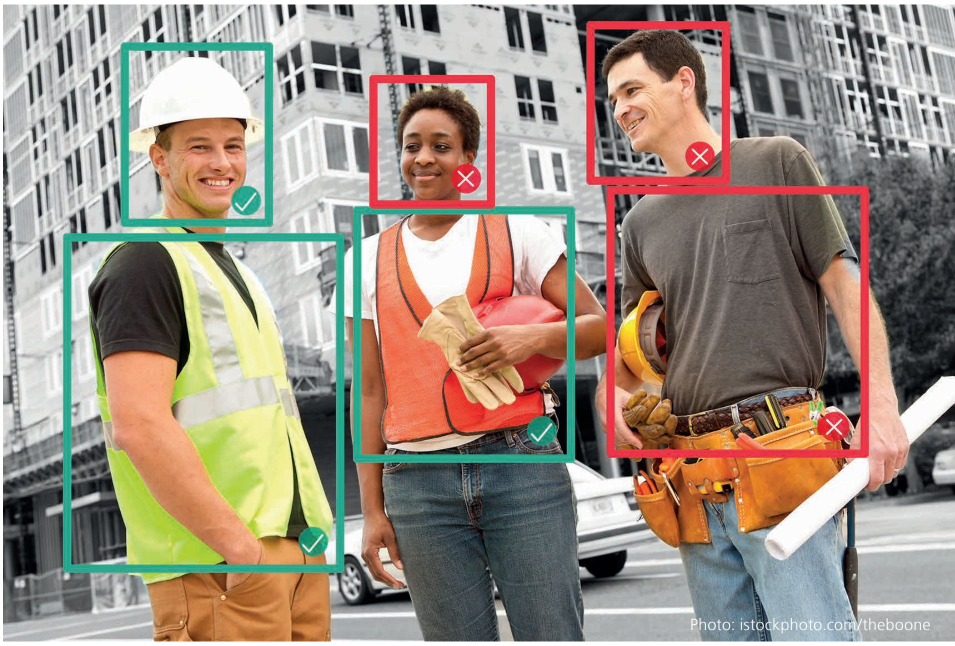
Use Case 2: Construction Site Surveillance

Crime, theft, and sabotage are among the top causes for damages on construction sites. AI-driven monitoring approaches can help to provide a 24 h surveillance of all relevant sections of a construction site area in order to prevent illegal entry vandalism and to detect events such as collapse or explosions.

UC 2.1: Illegal Entry Detection
Construction sites are usually secured by fences and barriers. Nevertheless, illegal access to the construction site can occur, especially at night. To avoid false alarms, it is first important to distinguish between animals and humans when movement is detected on the construction site. While most animal intrusions can be harmless, illegal human intrusion need to be recognized and countermeasures need to be taken either by triggering an alarm or informing security forces.

Visual object detection and tracking algorithms are the core methods to analyze video footage taken close to the boundary areas of the construction site. A particular challenge is the change in lighting conditions between day and night recordings, which can be approached using alternative sensor modalities such as infrared light. Acoustic analysis works regardless of the time of day and allows to recognize suspicious sounds such as footsteps or rattling noises that indicate stepping on scaffolds. A major challenge in deploying audio analysis algorithms at construction sites lies in the unique acoustic conditions at the audio sensor locations. Particularly for construction sites, manifold sound sources create complex soundscapes with a high degree of sound overlap. Furthermore, sound reflections on walls and ceilings as well as the high dynamic range of typical sound events can complicate a robust sound event detection [6].

UC 2.2: Vandalism and Theft Detection
Vandalism and theft can cause great damage and high costs. Various types of physical damages such as broken glass or damaged components made of metal or stone can be recognized using both the visual and acoustic sensory domain. Sound event detection algorithms can recognize sound events such as glass shattering, which can indicate vandalism. As a special case, graffiti can be recognized both visually by detecting characteristic images and patterns on walls, which often combine images and text-like structures, and acoustically by detecting characteristic spraying sounds.



The automatic detection of hard hats and safety vests can help improve safety on construction sites.

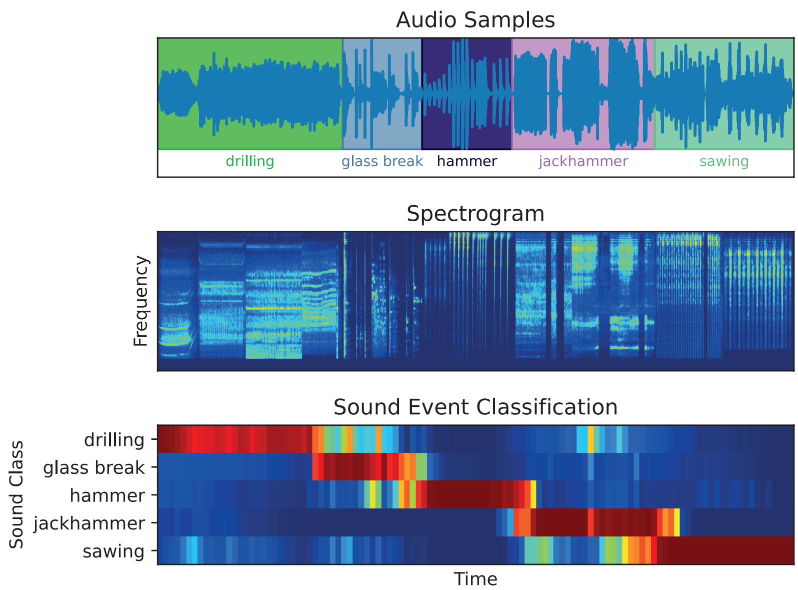
UC 2.3: Collapse, Explosion, and Fire Detection
Serious accidents can happen on construction sites and lead to the collapse of building parts, explosions, and fire, which can also endanger neighboring building complexes and properties. While a visual recognition is often hampered by swirling dust and smoke, an acoustic detection of explosion events is promising due to their short and loud nature. Given a network of distributed audio sensors, sound source localization techniques can be applied to also report the exact location of an explosion in order to better coordinate rescue efforts. Another promising approach is to integrate additional sensor devices such as smoke detectors as well as CO2 and NOx sensors.

Use Case 3: Logistics Monitoring

Logistical planning of goods deliveries is an essential part of any construction project. Each delivery must be timed correctly and deliver the exact quantity required. Especially if construction sites are located in densely populated urban areas, traffic and material flows need to be optimized using real-time traffic monitoring techniques.

UC 3.1: Audio-Visual Traffic Monitoring
With the progress of deep learning-based real-time object detectors, automatic long-term visual traffic monitoring becomes available also for monitoring vehicles on construction sites. Automatic detection of arriving and departing vehicles, automatically classifying different vehicle types such as cars, trucks, cargo vehicles, excavators, cranes, or vans as well as estimating speed and direction of movement of vehicles using object tracking algorithms might be useful tools for monitoring and controlling the logistics on construction sites. Existing technologies for object detection and tracking as well as object type classification at Fraunhofer IDMT can be modified and extended for the task at hand. Furthermore, previous studies already revealed the potential of state-of-the-art deep learning technologies for vehicle type classification [7]. However, the study also revealed that training multi-property neural networks, i.e., deep neural networks with more than one output layer, might be hard to train. Moreover, for special types of vehicles realistic multilabel datasets must be assembled, which can be a tedious and time-consuming task. Again, insufficient lighting and special visual modalities such as night-vision or infrared imagery might be problematic because such datasets are hardly available.

Complementary to the visual domain, acoustic traffic monitoring algorithms can detect vehicles based on two-channel or multi-channel audio recordings and estimate the vehicle type, the direction of movement, as well as the vehicle speed [8]. The combination of both sensory modalities allows for a multi-modal traffic monitoring approach, which can be applied in different logistics monitoring scenarios.



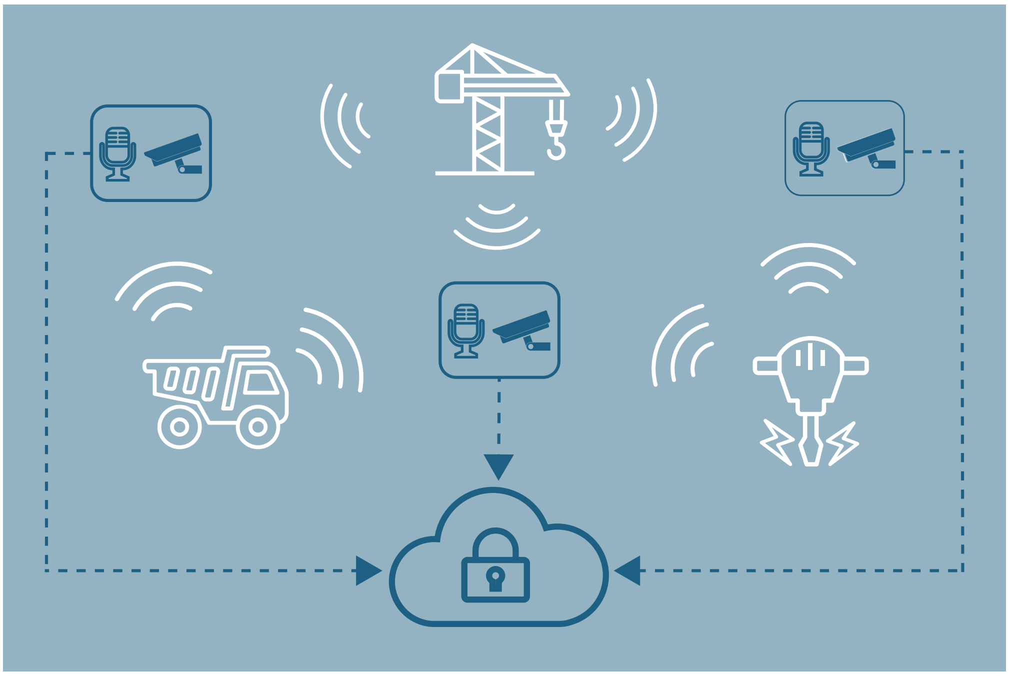
Acoustic Recognition and Classification of Sound Events

UC 3.2: Quantities of Goods

Another important clue for monitoring the logistics on construction sites is to automatically estimate the quantities of goods at each delivery. Object detection and tracking might enable counting of containers, cargo boxes, or pallets and thus give valuable insights into the delivered quantities of goods. However, insufficient lighting conditions as well as massively occluded object might lead to inaccuracies in detection and consequently miscounting.

UC 3.3. Construction progress

The adherence to construction schedules is of importance in large-scale construction projects since delays of only a few days can already result in large delay costs. Especially the building construction is characterized by temporal repetition patterns in the construction steps, which are repeated on every floor. A key component in order to identify the current phase in a construction schedule is the acoustic recognition of sound events, which are specific to work phases such as sawing, drilling, hammering, or unloading of goods. Such fine-grained segmentation of long-term audio recordings allows for a temporal synchronization with the construction project schedule and a quick identification of possible delays and initiation of countermeasures [9]



Secure, privacy-preserving data transmission and processing

Data Transmission and Exchange, Data Security and Privacy Protection

While some of the above-mentioned use cases can solely be processed on a mini-computer or other edge devices, others might require a more powerful computing device. While audio-visual footage can be recorded at one location, it is often necessary to first send it to a workstation where it can be processed / analyzed. Thus, robust and secure data and metadata transmission via a wireless local area network is key for a portable, easy-to-use system and should be taken into account when planning the necessary infrastructure. Systems for wireless data transmission are used to transmit sensor data and control information across network infrastructures. In recent years secure, cost- and energy-efficient systems for both industrial and home automation applications have been developed either as open frameworks or business solutions:

- Long Range Wide Area Network (LoRaWAN®) [10] provides a set of open tools and a global, open network to build an IoT application at low cost, featuring maximum security and ready to scale. Through robust end-to-end encryption, a secure and collaborative Internet of Things (IoT) network is built that spans across many countries around the globe, now operating thousands of gateways providing coverage to millions of people. The LoRa Alliance®, a non-profit association and the fastest growing technology alliance, drives the standardization and global harmonization of the LoRaWAN protocol [11].
- Fraunhofer IIS developed a miniaturized IoT system called mioty® [12], which sets new standards in the field of wireless data transmission when it comes to cost efficiency, transmission range, transmission security, and battery life. This solution relies on an asymmetrical transmission method that uses scores of simple sensor nodes and a complex receiver. A single receiver can handle a large number of simultaneous and robust transmissions. mioty® boasts a transmission range of up to 15 kilometers and the energy-efficient operation results in a battery life of up to 20 years. Fraunhofer IIS is a founding member of the mioty® Alliance where interested companies are given access to relevant component and software suppliers and end customers

Beyond transmission and data exchange, many use cases also require addressing of data security requirements: A/V data, annotations and models need to be stored and transferred e.g. to other systems in a manner that prevents unauthorized access (confidentiality), and ensures that information remains unaltered, including information about e.g. recording time and location (authentication). Both can be achieved by combining and implementing security standards related to key management, symmetric, and asymmetric encryption, hashing and perceptual hashing, and means to securely capture GPS and time information.

In addition to that, data protection can be an important issue, e.g. when it comes to storing and processing audio data: In some use cases, unintended recording of speech is possible, representing a problem with respect to capturing both private conversations, but also speaker characteristics of the persons involved. In order to avoid that, it is possible to apply specific speech filtering / scrambling approaches which render the speech »unusable«, while preserving high levels of classification performance.

Finally, concerns related to security and privacy regarding the training process can be addressed with »secure federated learning« approaches: By keeping training data »on-prem« and only exchanging weights to build a common AI model, and ensuring that the common model is calculated using homomorphic encryption (i.e. the aggregator can perform the necessary calculations on the encrypted data but cannot decrypt it) and so-called »differential privacy«, security and privacy concerns can be kept to a minimum.

Conclusion and Outlook

Artificial Intelligence in combination with state-of-the-art computer vision and audio processing techniques on edge devices has the potential to revolutionize the construction industry. In this paper we exemplarily showed on ten use-cases how modern multi-modal AI-driven algorithms can be exploited for automatic construction site monitoring, ranging from compliance with existing rules and regulations for occupational health and safety to construction site surveillance and even logistic monitoring.

The Fraunhofer Institute for Digital Media Technology IDMT has a long-lasting experience in the fields of audio analysis, computer vision, multi-modal processing, machine learning, and artificial intelligence. Together with our partners from the industry we develop unique solutions for specific individual problems.

References

[1] D. H. Arbeitsschutz, „https://www.deutsche-handwerks-zeitung.de/arbeitsschutz-auf-der-baustelle-vorschriften-im-ueberblick-113308/," [Online].

[2] D. H. Helmpflicht, „https://www.deutsche-handwerks-zeitung.de/helmpflicht-auf-dem-bau-wann-schutzhelme-getragen-werden-muessen-148050/," [Online].

[3] S. w. g. Mundschutz, „https://www.staub-war-gestern.de/," [Online].

[4] F. I. MaskCognizer, „https://www.idmt.fraunhofer.de/en/institute/projects-products/maskcognizer-face-mask-detection.html," [Online].

[5] D. I. Warnkleidung, „https://www.din.de/de/mitwirken/normenausschuesse/nps/veroeffentlichungen/wdc-beuth:din21:254660763," [Online].

[6] J. Abeßer, M. Götze, S. Kühnlenz, R. Gräfe, C. Kühn, T. Clauß und H. Lukashevich, „A Distributed Sensor Network for Monitoring Noise Level and Noise Sources in Urban Environments," in Proceedings of the IEEE 6th International Conference on Future Internet of Things and Cloud, Barcelona, Spain, 2018.

[7] M. Hayat und N. Khalid, Investigation on Deep Learning Approaches for Multi-Property Car Classification, Fraunhofer IDMT/TU Ilmenau, December 2019.

[8] J. Abeßer, S. Gourishetti, A. Kátai, T. Clauß, P. Sharma und J. Liebetrau, „IDMT-Traffic: An Open Benchmark Dataset for Acoustic Traffic Monitoring Research," in Proceedings of the European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021.

[9] S. Holly, Y. Körber, R. Ruch, H. Lukashevich und J. Abeßer, „Apparatus, method, and computer program for acoustically monitoring a monitoring region". Patent EP3791370A1, 2018.

[10] LoraWAN, „https://www.thethingsnetwork.org/," [Online].

[11] A. LORA, „https://lora-alliance.org/," [Online].

[12] M. Fraunhofer IIS, „https://www.iis.fraunhofer.de/en/ff/lv/net/telemetrie.html," [Online].

Contact

Dr.-Ing. Uwe Kühhirt
Media Management and Delivery
Phone +49 3677 467-205
uwe.kuehhirt@idmt.fraunhofer.de

Fraunhofer IDMT
Ehrenbergstr. 31
98693 Ilmenau
Germany
www.idmt.fraunhofer.de/m2d

Version 01/2022