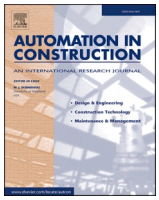
Automation in Construction 176 (2025) 106297

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon



Weighted adaptive active transfer learning for imbalanced multi-object classification in construction site imagery

Karunakar Reddy Mannem [a,b,*], Samuel A. Prieto [c], Borja García de Soto [c], Fernando Bacao [b]

[a] Center for Research Computing, New York University Abu Dhabi (NYUAD), A2 Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
[b] NOVA Information Management School (NOVA IMS), Campus de Campolide, Universidade Nova de Lisboa, 1070-312 Lisboa, Portugal
[c] S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
InceptionV3
BiLSTM
WATLAS
Active learning
Transfer learning
Adaptive sampling

ABSTRACT

Construction site monitoring relies on robust image classification to enhance safety, track progress, and optimize resource management. However, the amount of clutter and the high cost of manual labeling pose significant challenges. This paper presents an approach to multi-object classification in construction sites using Adaptive Active Transfer Learning. The Weighted Active Transfer Learning with Adaptive Sampling (WATLAS) framework is introduced, where Transfer Learning is combined with weighted Active Learning to efficiently classify diverse objects. A pre-trained InceptionV3 architecture integrated with bidirectional long short-term memory (BiLSTM) layers is utilized, and superior performance is achieved through adaptive sampling techniques compared to traditional methods. WATLAS achieves 97 % accuracy on a comprehensive dataset of 9344 construction site images spanning 15 object categories and maintaining 90 % accuracy with only 5 % labeled data. By optimizing performance metrics, the framework demonstrates significant improvements over traditional methods, making it a scalable solution for construction site monitoring.

1. Introduction

The construction industry is quickly embracing digital technologies to improve efficiency, safety, and project management [1]. Computer vision and machine learning play a crucial role in this transformation by enabling automated analysis of construction site imagery [2]. Yet, the complex and dynamic nature of construction sites poses significant challenges for object classification tasks. Image classification offers nonintrusive means to assess construction sites [3,4]. It provides a valuable tool for evaluating the construction environments, aiding in proactive risk mitigation. Nevertheless, labeling extensive datasets for model training poses challenges.

Recent advancements in artificial intelligence and machine learning have led to the development of innovative construction monitoring and management approaches. For instance, researchers have developed frameworks integrating Building Information Modeling (BIM), digital twins, augmented reality (AR), virtual reality (VR), mixed reality (MR), and sensing technologies for automated construction progress monitoring [5]. These technologies are being applied across various aspects of construction, including progress tracking, quality control, and safety management [6].

Integrating computer vision with other technologies like BIM and digital twins enhances project visualization and management [7,8]. Furthermore, computer vision is employed in building automation and robotics, increasing efficiency and precision in construction tasks [9].

Despite the advancements in digital technologies for construction, implementing these innovations still presents significant challenges. These include issues related to data acquisition, model learning, and model validation [10]. Construction site images are diverse and unique, making it difficult to obtain large, representative datasets. Even when substantial data is available, the labeling process is tedious and manual, often prone to errors due to the lack of expert knowledge in accurately identifying construction-specific objects. To address these challenges, researchers are exploring innovative approaches, such as creating synthetic datasets based on virtual environments and using Transfer Learning techniques to develop efficient object recognition models [11].

This paper introduces an Active Learning-based approach to tackle the persistent challenges of limited labeled data and the need for expert

* Corresponding author at: Center for Research Computing, New York University Abu Dhabi (NYUAD), A2 Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates.
E-mail addresses: karunakar.mannem@nyu.edu, 20230691@novaims.unl.pt (K.R. Mannem).

https://doi.org/10.1016/j.autcon.2025.106297
Received 13 January 2025; Received in revised form 19 May 2025; Accepted 19 May 2025
Available online 26 May 2025

knowledge in training robust machine learning models for the construction domain. By employing a Convolutional Neural Network (CNN) architecture, the model learns intricate features from construction site images, enabling accurate classification. The Active Learning framework facilitates model improvement through iterative data selection, allowing the algorithm to focus on the most informative instances. The dynamic nature of construction sites poses a special challenge if the labels need to be defined in advance, which is why the adaptive aspect of this framework allows for adaptability to the evolving characteristics of the dataset, thus reducing the reliance on large amounts of manually labeled data.

Building on these concepts, our research develops an adaptive active Transfer Learning approach for multi-object classification in construction site images. Our method, WATLAS, combines the strengths of Transfer Learning and Active Learning to achieve high accuracy with minimal labeled data. This approach directly addresses the challenges mentioned earlier by leveraging existing knowledge and efficiently selecting the most valuable data for model training, thereby minimizing the need for extensive manual labeling and expert intervention. The primary objectives of this paper are.

- To develop a robust multi-object classification model for construction site imagery.
- To implement and evaluate various Active Learning strategies for efficient data labeling.
- To assess the effectiveness of active Transfer Learning in improving model performance.
- To compare the proposed WATLAS method with other sampling approaches.

In the following sections, we will discuss the background of the different methods involved in our approach and the methodology of WATLAS in detail, including its unique adaptive sampling strategy and how it integrates Transfer Learning principles. We will also present experimental results that demonstrate its effectiveness in improving object classification accuracy with minimal labeled data, potentially revolutionizing how we approach data-driven decision-making in construction site management. Finally, we will present the limitations of our framework and future work to overcome these limitations.

2. Background

Civil infrastructure and construction monitoring plays a pivotal role in tracking progress, ensuring safety, and maintaining the resilience of built environments. This process encompasses the detection and classification of various objects within construction sites. Traditional approaches to object classification often rely on supervised learning, requiring large, labeled datasets for training. However, obtaining extensive labeled data can be challenging in the civil infrastructure domain due to the diversity and dynamic nature of construction environments.

2.1. Computer vision in construction

Computer vision applications in construction have gained significant attention in recent years. Researchers have explored various object detection, activity recognition, and progress monitoring techniques using image data from construction sites. However, the diversity of objects and environmental conditions in construction settings presents unique challenges for these applications.

CNN has proven highly successful in image classification tasks, capturing hierarchical features from input images. These architectures excel in learning complex patterns, making them well-suited for object classification in civil infrastructure monitoring. Krizhevsky et al. [12] demonstrated the effectiveness of deep CNN in the ImageNet Large Scale Visual Recognition Challenge, marking a significant milestone in image classification. Recent research in Construction task monitoring has leveraged CNN with LSTM for feature extraction and classification. For example, Mannem et al. [13] applied CNN-LSTM to track construction activities, showcasing the potential for automated construction project management.

Despite their success, CNN often requires substantial labeled datasets, motivating the exploration of other learning techniques to enhance their efficiency. To overcome this limitation, other deep learning architectures have been explored for construction applications. For instance, Region-based Convolutional Neural Networks (R-CNN) and its variants like Fast R-CNN, Faster R-CNN [14] have been employed for object detection tasks in construction sites [15]. These models have significantly improved the detection and localizing of objects in complex environments.

Moreover, the integration of 3D computer vision techniques has been investigated to address the challenges posed by the dynamic and cluttered nature of construction sites. Techniques such as Structure from Motion (SfM) and Multi-View Stereo (MVS) have been utilized to create 3D models of construction sites, aiding in progress monitoring and quality control [16].

Another promising approach is using Generative Adversarial Networks (GANs) for data augmentation and synthetic data generation. GANs can generate realistic images of construction sites, which can be used to augment training datasets and improve the performance of computer vision models [17].

Furthermore, the application of Semantic Segmentation techniques, such as Fully Convolutional Networks (FCNs) and U-Net, has been explored for pixel-level classification of construction site images [18]. These techniques enable detailed analysis of construction activities and materials, providing valuable insights for project management [19].

Building on these foundations, recent work looks beyond CNN-centric pipelines by combining fast single-stage detectors such as YOLO for real-time PPE and equipment tracking [20], transformer-based backbones (e.g., ViT/DETR) that capture long-range spatial context [21], spatio-temporal models for action and behavior recognition [22], and synthetic data and digital-twin environments to mitigate labeling scarcity [23]. The fusion of these techniques, often alongside IoT sensor feeds, has produced robust multi-algorithm site-monitoring systems capable of simultaneous detection, tracking, and safety assessment under challenging on-site conditions [24].

2.2. Challenges and limitations

Despite these advancements, several challenges remain in the application of computer vision in construction.

- Variability in environmental conditions: The variability in lighting conditions, occlusions, and the presence of dynamic objects necessitate robust and adaptive models.
- Data scarcity: The need for large labeled datasets for training deep learning models is a significant hurdle, especially in construction where site-specific data is often limited.
- Class imbalance: Safety-critical categories like rebar or drywall panels are often underrepresented in datasets, leading to biased models that struggle to detect these minority objects accurately.

Active Learning and Transfer Learning approaches have been proposed as potential solutions to address the need for large labeled datasets and improve model generalization [25].

2.3. Active learning

Settles ([26,27]) defines Active Learning as a process where the algorithm interactively queries the user to obtain labels for the most informative instances, thereby optimizing the learning process as shown in Fig. 1. This iterative labeling strategy enables the model to focus on

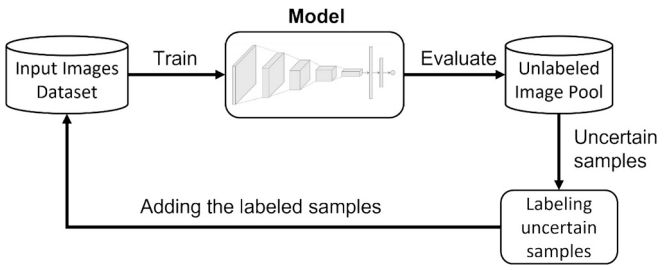


Fig. 1. Iterative active learning framework for image classification with human-in-the-loop labeling.

challenging instances, improving its generalization performance. Active Learning strategies aim to minimize the amount of labeled data required for training by selectively querying the most informative samples. These approaches have shown promise in various domains, including computer vision tasks in construction. Active Learning has emerged as a powerful paradigm in computer vision to address the limitations of supervised learning, particularly in scenarios with a scarcity of labeled data.

In the context of image classification, various studies have demonstrated the effectiveness of Active Learning. For instance, Gal et al. ([28,29]) proposed using deep learning uncertainty estimates to guide Active Learning, showing improvements in model performance. These methods leverage Bayesian neural networks to estimate uncertainty, which helps in selecting the most informative samples for labeling. Similarly, Sener et al. [30] introduced a core-set approach for Active Learning in deep learning, which selects a diverse subset of data points that best represent the entire dataset, further enhancing model performance.

Active Learning has also been applied to medical image analysis, where limited labeled data pose challenges similar to those in civil infrastructure monitoring. Smailagic et al. [31] demonstrated the application of Active Learning in medical imaging, showing that it can significantly reduce the amount of labeled data required while maintaining high diagnostic accuracy. This is particularly relevant in medical domains where expert annotations are costly and time-consuming.

In addressing the challenges of vision-based monitoring in construction, Kim et al. [32] proposed an innovative database-free approach using deep Active Learning. Their work demonstrates how Active Learning can reduce reliance on extensive pre-labeled databases while maintaining monitoring effectiveness. The authors developed a system that actively selects the most informative frames for labeling, thereby minimizing the manual annotation burden. This approach is particularly significant for construction monitoring where site conditions are dynamic and pre-existing databases may not adequately represent site-specific characteristics.

Recent advancements have further expanded the scope of Active Learning. For example, Sinha et al. [33] introduced a variational adversarial Active Learning framework that combines adversarial training with Active Learning to improve sample efficiency. This method has shown promising results in various computer vision tasks, including object detection and semantic segmentation.

In summary, Active Learning continues to evolve, with new methodologies and applications emerging across different domains. Its ability to reduce the need for extensive labeled datasets while maintaining or improving model performance makes it a valuable tool in scenarios where labeled data is scarce or expensive.

2.4. Sampling strategies

Effective sampling strategies are crucial for optimizing model performance in an Active Learning context, especially when dealing with large, unlabeled datasets. We can categorize the sampling methods in mainly three groups: uncertainty sampling, diversity sampling, and a combined strategy. These sampling methods enhance the learning process by selecting the most informative samples for labeling [34].

Uncertainty sampling identifies instances where the model is least confident in its predictions. This method selects samples with the smallest margin between the top two predicted probabilities, indicating ambiguity in classification. By prioritizing these uncertain samples, the model can improve its understanding of challenging cases. Machine learning research supports this approach and highlights its effectiveness in Active Learning scenarios [35]. One of the most notable strategies in the uncertainty sampling group is the Bayesian Active Learning By Disagreement (BALD). BALD is a strategy that uses Bayesian principles to select the most informative samples for labeling. It seeks to identify data points that yield the most significant expected decrease in model uncertainty, making it useful for cases with few labeled instances. This approach has shown robust performance across various data scenarios, balancing exploration and exploitation while being less sensitive to noisy labels than other uncertainty-based methods.

Diversity sampling aims to capture a wide range of data points by selecting samples that are distinct from each other. This method uses clustering techniques, such as k-means, to ensure that the selected instances represent different regions of the input space. This strategy effectively ensures that the training dataset remains representative of the entire input space [36].

The combined strategy leverages uncertainty and diversity sampling to select uncertain and diverse samples. This approach balances exploration and exploitation by refining model predictions while maintaining a comprehensive view of the data landscape.

2.5. Transfer learning

Transfer Learning is a paradigm where a model trained on a particular task (Source Data) is reused for another similar task (Target Data) as shown in Fig. 2, allowing the model to reuse the previous knowledge gained from the original task to improve its performance on a similar new task.

A model is first pre-trained on source data to learn features and then retrained on target data with knowledge transfer to adapt to new labels. Transfer Learning has emerged as a powerful technique for leveraging knowledge from pre-trained models to improve performance on new tasks. In the context of construction site imagery, Transfer Learning can help overcome the limitations of small datasets and reduce training time. Kim et al. [37] demonstrated the effectiveness of Transfer Learning in detecting construction equipment using region-based fully convolutional networks. Mengiste et al. [38] applied Transfer learning on data-scarce construction image datasets to track construction progress successfully. Bharathi et al. [39] used Unsupervised Deep Domain Adaptation to effectively identify safety hazards in construction sites, using

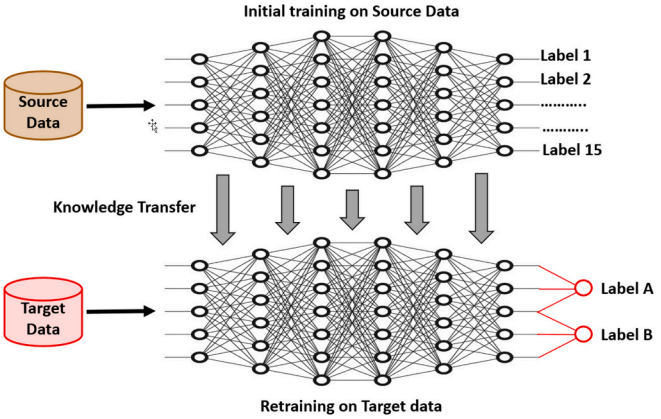


Fig. 2. Illustration of a transfer learning approach.

labeled data from a controlled environment and transferring the training to the real construction site.

In summary, transfer learning can substantially reduce the training data and time, thereby enhancing the model performance, with fewer data points and more predictability can be enhanced.

2.6. Integration of active learning and transfer learning

Active Learning (AL) and Transfer Learning (TL) are powerful methods that by themselves can significantly enhance model performance in data-scarce situations, common in construction and civil engineering domains. However, when integrated, these approaches can offer even more robust solutions to the challenges of limited labeled data.

TL allows models to leverage knowledge gained from pre-trained networks on large datasets, which is particularly useful in domains where obtaining large amounts of labeled data is challenging. For instance, Chen et al. [40] applied TL to develop a new semantic region detection approach on construction sites to avoid the need for labeling.

On the other hand, AL focuses on selecting the most informative samples for labeling, thereby reducing the overall labeling effort while maximizing model performance. Frick et al. [41] applied an Active Learning approach to monitoring the aging civil infrastructure damages and defects, showing that carefully selected training samples can lead to models that perform well with less labeled data.

While these studies demonstrate the potential of AL and TL in construction and civil engineering applications, research integrating both approaches remain limited in this domain. The scarcity of such studies highlights a significant gap in addressing the unique challenges of data scarcity and diversity in construction site monitoring and safety management. The integration of AL and TL can create a synergistic effect. TL provides a strong initial model, while AL helps fine-tune this model efficiently for the specific task at hand. This combination is particularly powerful in the construction domain, where data is often scarce and expensive to label. For example, Zheng et al. [42] proposed an Active Transfer Learning framework for crack detection, outperforming traditional machine learning methods while requiring significantly less labeled data.

Construction sites present unique challenges for computer vision, including dynamic lighting, frequent occlusions, and the movement of workers and machinery. These factors complicate the collection of large, balanced, and fully annotated datasets. Active Learning (AL) addresses this by strategically selecting the most informative samples, thus reducing labeling costs while maximizing model performance [43]. Transfer Learning (TL) leverages knowledge from large, pre-annotated datasets, often from related domains, to jumpstart model training in the target domain, which is critical when labeled construction site data is scarce [44].

Compared to classical machine learning methods, which typically require extensive hand-crafted features and large labeled datasets, AL and TL offer more scalable and efficient solutions for real-world deployment. Data augmentation can help, but may not fully capture the complexity and variability of real construction environments. The combined use of AL and TL thus represents a novel and practical approach for addressing these domain-specific constraints.

While previous standards AL addresses the labeling issue and TL helps mitigate data scarcity, they do not directly address the severe class imbalance in construction imagery. For instance, Duan et al. [4] note in their work on the SODA dataset that safety-critical categories (e.g., rebar, drywall panels) appear in very low proportions (less than 5 % of the total annotations). This underrepresentation of minority categories contributes to elevated misclassification rates for minority objects, with some studies reporting errors, such as the one of He et al. [45], which highlights that learners trained on imbalanced data tend to be biased toward the majority class.

This imbalanced learning creates systemic detection gaps in models. Though adaptive weighting strategies have shown promise in other domains, for instance, Gao et al. [46] demonstrate that dynamically reweighting samples can substantially improve the recognition of underrepresented classes. However, the current construction vision systems lack mechanisms to prioritize underrepresented elements dynamically.

To bridge this gap in construction image analysis, we have developed an extension of the Active Transfer Learning for Adaptive Sampling (ATLAS), developed by Monarch et al. [47], introducing a weighted mechanism that we term WATLAS This innovative method combines the strengths of TL and AL, tailored specifically for the construction industry, by leveraging pre-existing knowledge from TL and intelligently selecting the most informative samples through AL, while strategically weighting samples to mitigate class imbalance and prioritize underrepresented data points. WATLAS aims to create more efficient and effective object classification models for construction site monitoring and safety management.

In summary, the WATLAS approach offers a promising solution to the persistent challenges of limited labeled data and the need for expert knowledge in the construction domain. It addresses the following:

- What previous methods did: Previous methods focused on supervised learning, requiring large labeled datasets, and used techniques like CNNs, R-CNN, and GANs for object detection and classification.
- What their limitations are: These methods struggle with data scarcity, class imbalance, and the dynamic nature of construction sites, leading to biased models and poor performance on minority classes.
- How WATLAS addresses those gaps: WATLAS integrates Active Learning for efficient labeling, Transfer Learning for leveraging pre-existing knowledge, and adaptive sampling to address class imbalance, thereby improving model performance and reducing the need for extensive labeled data.

3. Methodology

This section explains the methodological framework for developing and evaluating the WATLAS approach. The following subsections outline the Active Transfer Learning, data preprocessing steps, model components, training procedures, and the sampling mechanisms that underpin the WATLAS framework.

3.1. Active transfer learning framework

The Active Transfer Learning (ATL) Framework is designed to handle multi-object classification in construction site imagery efficiently. It integrates Transfer Learning and Active Learning techniques to optimize the learning process. Fig. 3 illustrates the proposed framework, which consists of several key components. A pre-trained model is fine-tuned on a construction dataset, predicts on an unlabeled dataset, and undergoes performance validation. If unacceptable, active learning refines the model until it is finalized.

3.1.1. Data preprocessing

The construction site images need a series of preprocessing before being fed to the deep learning model:

1. Resizing: Standardizing image dimensions is essential for maintaining consistency in the input tensor shape for the model architecture.
2. Normalization: Pixel values normalized to a range between [−1,1]. Normalization stabilizes the learning process by preventing large pixel values from dominating the loss function and the gradients. We used the formula:

$$\text{Normalized\_pixel} = \frac{\text{pixel} - 127.5}{127.5}$$

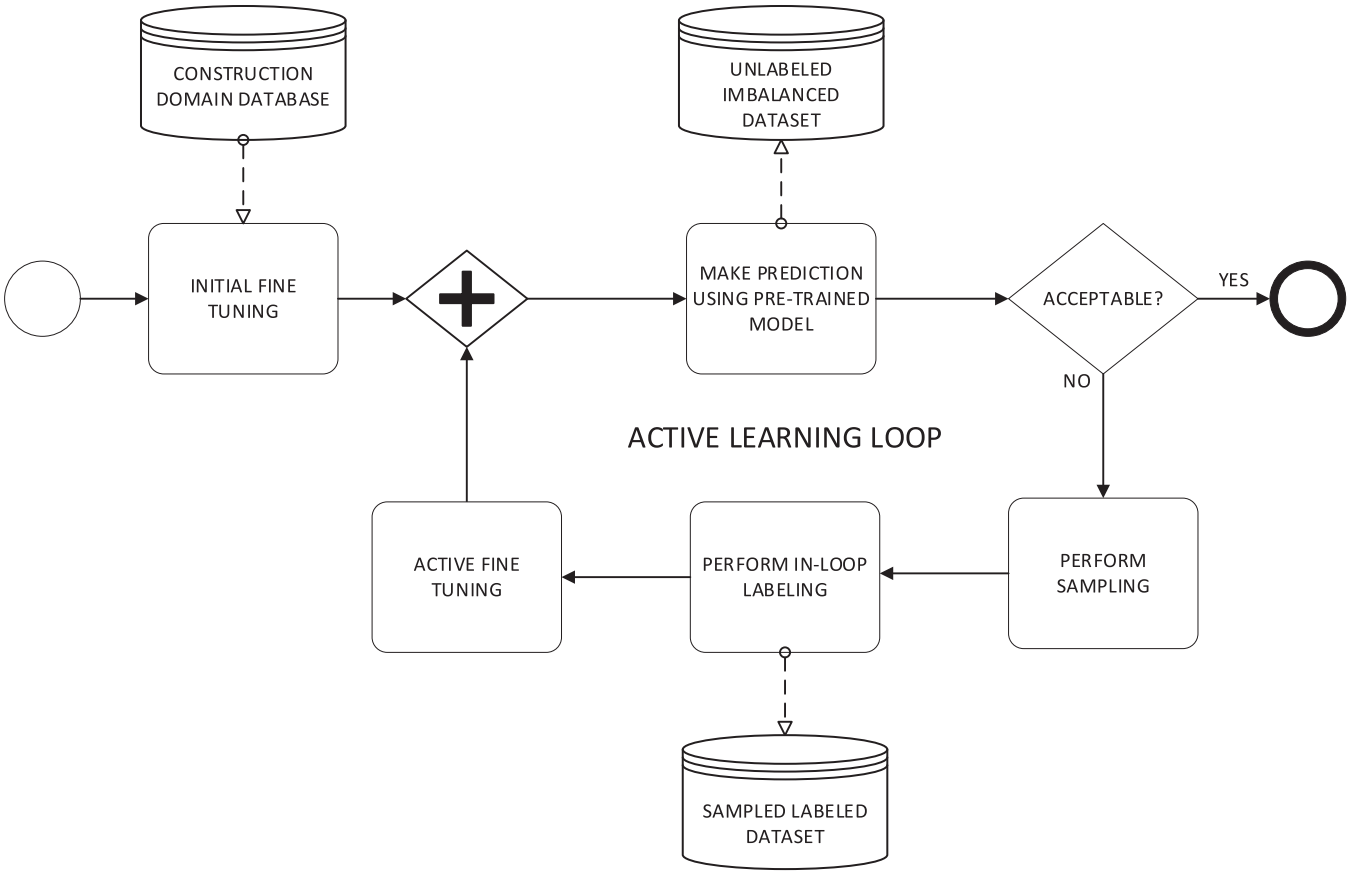


Fig. 3. Overview of the Proposed ATL (Active Transfer Learning) Framework.

Where 'pixel' is the original pixel value (0–255).

3. Data Augmentation: To improve the model's generalization and to reduce overfitting, a series of data augmentation techniques were applied to artificially increase the size of the training set in the dataset. Our augmentation pipeline included operations such as: rotating, flipping, cropping, and brightness adjusting the images.
4. Handling missing/corrupted data: In cases where images were missing or corrupted, we excluded those samples from the dataset

3.1.2. Framework description

The ATL Framework leverages a pre-trained model, such as InceptionV3, fine-tuned on a similar domain dataset. This model is the foundation for feature extraction. It is initially trained on a large-scale dataset and fine-tuned on a domain-specific dataset from the construction domain, leveraging its robust capability to capture complex image patterns [48]. TL is employed to adapt the pre-trained model to the specific characteristics of construction site data. By fine-tuning the model on a smaller, labeled dataset from the construction domain, we can enhance its ability to recognize relevant features and improve classification accuracy.

In addition to the pre-trained model, the ATL Framework incorporates an Active Learning loop to address the challenge of limited labeled data. The loop involves selecting the most informative samples from an unlabeled, imbalanced dataset. These samples are chosen based on uncertainty and diversity sampling methods, prioritizing instances likely to improve model performance when labeled [35]. AL reduces the manual effort required for labeling while maximizing the impact of each labeled sample.

Various sampling methods, discussed in Section 2.3, will be evaluated to compare them with the proposed method. For example, the adaptive sampling strategy that dynamically updates sample weights during training. This approach ensures that the most valuable data points are prioritized for labeling, reducing manual effort and enhancing model efficiency.

Once samples are selected through Active Learning, they undergo in-loop labeling, where human annotators provide labels for these high-priority instances. This iterative process continues until the model achieves satisfactory performance.

Finally, the newly labeled samples are used to fine-tune the model further, enhancing its predictive accuracy on construction site imagery. The framework iteratively refines the model through a cycle of prediction, sampling, labeling, and fine-tuning.

By combining the strengths of Transfer Learning and Active Learning, the ATL Framework effectively addresses the challenges of data scarcity and diversity in construction site monitoring. This integrated approach enables more efficient utilization of available data and improves object classification models' performance.

3.2. Proposed model architecture

We propose an architecture that combines the Transfer Learning model (a pre-trained model in Fig. 3) and BiLSTM for effective multi-object classification, as depicted in Fig. 4. This model is designed to efficiently handle complex data by combining advanced feature extraction and temporal modeling techniques. The network's input is a batch of RGB images resized to 299 × 299 × 3 pixels, normalized to standardized intensity values.

3.2.1. Feature extraction

The InceptionV3 model, pre-trained on the ImageNet dataset, serves as the base feature extractor in our architecture. Due to its deep

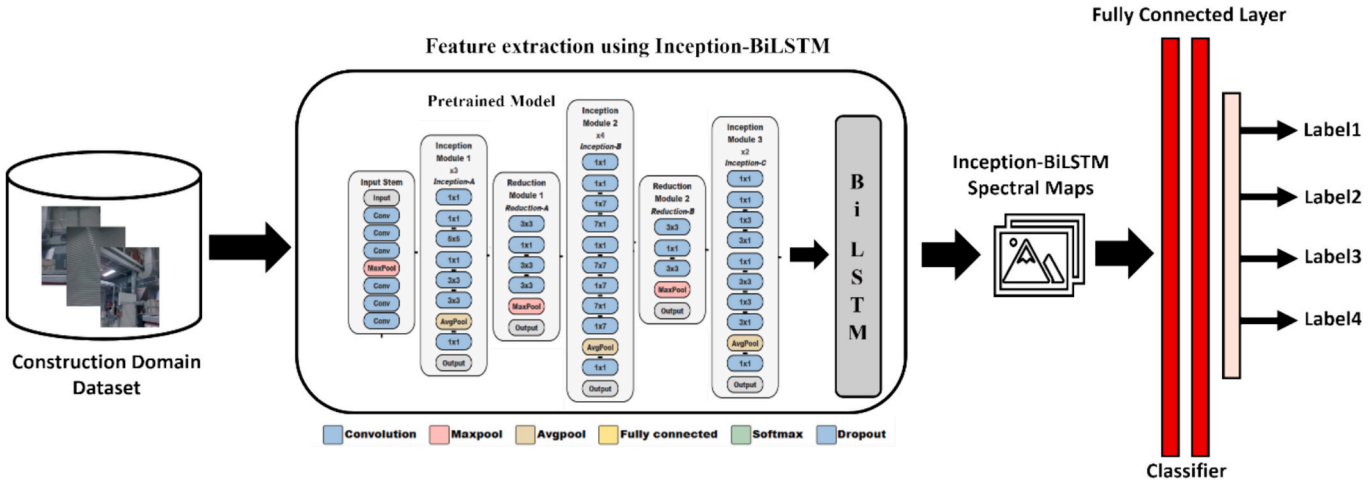


Fig. 4. Proposed model architecture.

convolutional layers and unique inception modules, InceptionV3 is renowned for capturing rich and diverse features from images [48]. The last 50 layers of the pre-trained model are fine-tuned on the construction domain dataset, enabling it to learn domain-specific features and improve classification accuracy, producing an output feature map of size **8 × 8 × 2048**; the model configuration is shown in Table 1.

3.2.2. Temporal modeling

The extracted feature map undergoes additional transformations to prepare it for temporal modeling. The steps include flattening, dense layer projection, and dropout regularization. To capture spatial dependencies and temporal sequences in the feature maps, we incorporate BiLSTM layers. LSTMs are a recurrent neural network that excels at learning long-range dependencies in sequential data [49]. Using bidirectional LSTMs, our model can process the feature maps in both forward and backward directions, enhancing its ability to understand the context and improve classification accuracy. Although the input images from construction sites are typically static, the inclusion of BiLSTM layers in our architecture is intended to capture richer, bidirectional feature relationships across spatial dimensions rather than explicit temporal sequences. This approach has improved the model's ability to integrate contextual information from multiple directions, leading to more robust feature extraction, particularly in cluttered or occluded scenes. The configuration setup of the BiLSTM layers is detailed in Table 2.

3.2.3. Classification

The final component of our architecture is a fully connected layer with **15 units**, corresponding to the number of object classes with sigmoid activation functions, designed for multi-label classification tasks. This layer takes the output from the BiLSTM and maps it to multiple labels, allowing the model to classify multiple objects within an image simultaneously. This integrated approach leverages the strengths of both Transfer Learning and recurrent networks, making it particularly effective for complex multi-object classification tasks in diverse environments such as construction sites.

Table 1
Feature extraction module configuration.

Component	Details
Base Network	InceptionV3
Pre-training Dataset	ImageNet
Fine-tuning Strategy	Last 50 layers
Output Dimensions	8 × 8 × 2048 feature map

Table 2
Temporal Processing Module Configuration.

Layer	Units	Activation	Additional Settings
BiLSTM Layer 1	128	ReLU	return_sequences = True
BiLSTM Layer 2	64	ReLU	–

3.2.4. Training proposed model

A two-phase training approach using the transfer learning strategy is adopted to optimize the model for construction-specific features.

- **Phase 1**: Training the top layers of InceptionV3 for 10 epochs.
- **Phase 2**: Fine-tuning the last 50 layers over 25 epochs, gradually unfreezing to preserve low-level feature representations.

The binary cross-entropy loss function is utilized, specifically chosen for multi-label classification. To address class imbalance, weights are assigned inversely proportional to class frequencies. The following hyperparameters are tuned for optimal performance, as shown in Table 3.

3.2.5. Regularization

Regularization techniques are employed to reduce overfitting:

- **L2 Regularization**: Weight decay with a coefficient of **1.00E-04**.
- **Dropout**: Applied with a probability of **0.5** in dense layers.
- **Batch Normalization**: Applied after convolutional layers to stabilize training dynamics.

Finally, the model is compiled with the Adam optimizer and binary cross-entropy loss and evaluated using accuracy, precision, and recall metrics. It demonstrates state-of-the-art performance in multi-object classification tasks, showcasing robust feature extraction, effective

Table 3
Hyperparameter settings.

Parameter	Value
Batch Size	512
Optimizer	Adam
Learning Rate	0.001
Dropout Rate	0.5
Training Epochs	50
Early Stopping Patience	5
Learning Rate Reduction	0.1
Minimum Learning Rate	1.00E-06

temporal modeling, and resilience to class imbalance. The model generalizes well across diverse construction site scenarios by fine-tuning the architecture and adopting appropriate regularization strategies.

3.3. Proposed sampling strategy WATLAS

WATLAS integrates AL with TL by predicting sample correctness and selecting those most likely to be misclassified. This method refines the model's ability to handle complex cases by focusing on improving prediction accuracy.

The proposed WATLAS is designed to improve the efficiency and precision of adaptive sampling strategies within the ATL paradigm. Building upon the existing Active Transfer Learning for Adaptive Sampling (ATLAS) strategy [47], WATLAS incorporates a novel class-weighted sampling approach that prioritizes underrepresented classes during the learning process. This is especially critical in highly imbalanced datasets, such as those encountered in construction site monitoring, where the framework aims to detect and classify various structural elements and activities across diverse scenes. Below, we detail the WATLAS framework, its architecture, and the innovations introduced for superior handling of imbalanced data. We also explain the visualized process and the architecture used.

3.3.1. ATLAS model architecture

The ATLAS model consists of the below steps:

1. Initial Model Training: An initial neural network is trained with image inputs with a set of basic classes (Label 1, Label 2, etc.) as per Source Dataset, and the model is used to predict the labels on the validation data. With the help of actual labels, the predicted labels are bucketed as "Correct Label" or "Incorrect Label" accordingly.
2. The same model is repurposed by changing the last layer and then retraining on the two bucketed data to predict the "Correct Label" or "Incorrect Label." Then, this new binary output model is applied to the unlabeled dataset to predict the "Correct Label" or "Incorrect Label."
3. Then, the images with the "Incorrect Label" with the highest confidence will be sampled, assuming that these images will be labeled in the future. These images will become part of the active ATL sampling for training so that the model will later predict those images correctly and change the image label from "Incorrect Label" to "Correct Label" to repeat Step 2.
4. Adaptive Sampling: As new " Incorrect Label " data are sampled, the model dynamically adds them to the training pool. This sampling is guided by predictions on validation data, which categorizes samples as " Correct Label " or "Incorrect Label "based on how they would improve model accuracy.

3.3.2. WATLAS model architecture

The WATLAS framework as shown in Fig. 5 represents a substantial advancement in active Transfer Learning, tailored explicitly for imbalanced datasets. By introducing weighted class adjustments, the WATLAS architecture follows the same steps as ATLAS: adding sampling categories/labels with computed class weights according to their distribution. This ensures that the model does not overly prioritize classes with abundant examples, thereby improving the representation of minority classes in the adaptive sampling process. This weighted approach adapts dynamically to the inherent class distributions in each sampled batch, promoting more accurate and balanced model learning.

A pseudocode of the entire strategy is provided in Fig. 6. The explanation of the different steps is as follows:

1. Initialization (Step 1 to 4 in the pseudocode):
   - Class Weight Calculation: Compute the initial class weights W inversely proportional to the frequency of each class in the labeled dataset L. In mathematical terms, for each class c,

$$W[c] = \frac{1}{n_c + \varepsilon}$$

where n_c is the number of samples in class c and ε is a small constant added to prevent division by zero.

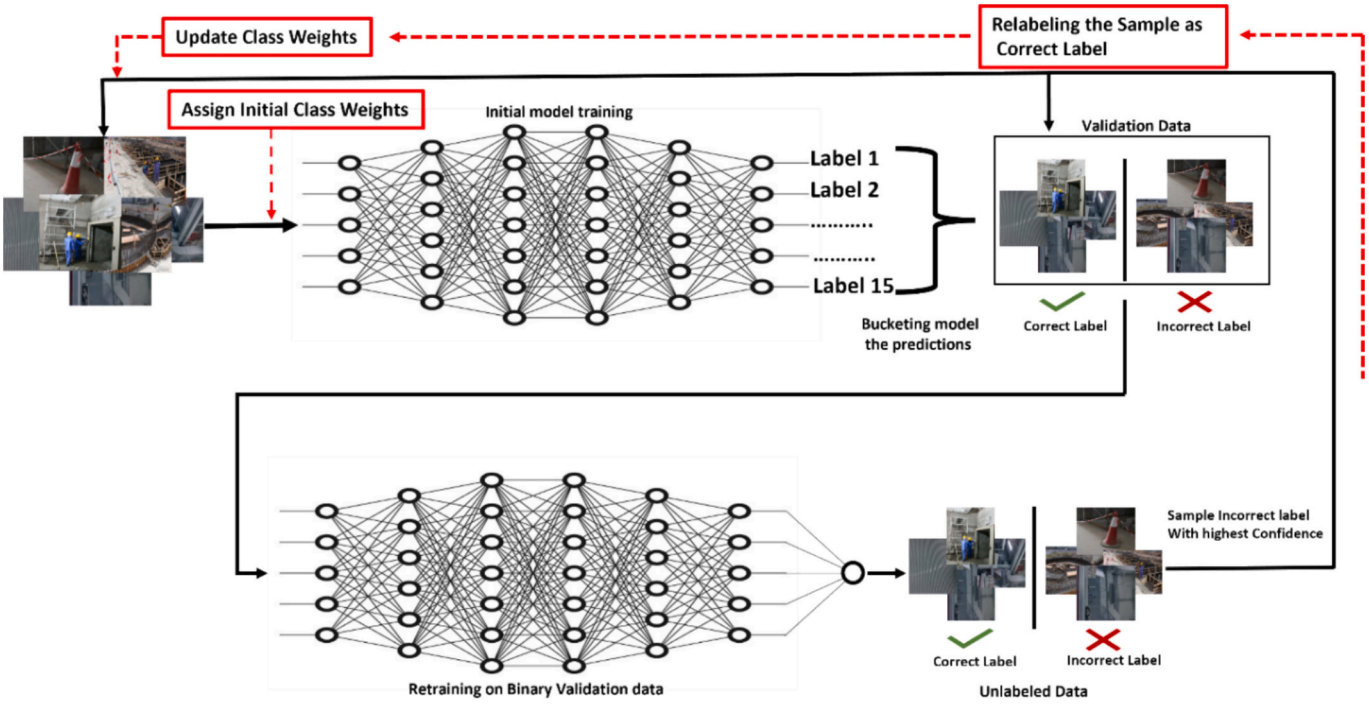


Fig. 5. Illustrates the proposed WATLAS framework, with the new class weights assignment shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

```
Input:
    L: Labeled dataset
    U: Unlabeled dataset
    M: Pre-trained model
    k: Number of samples to label per iteration
    α, β: Hyperparameters controlling the trade-off between uncertainty and diversity
    ε: A small constant to avoid division by zero
    Convergence criterion (e.g., model performance or a maximum number of iterations)

Output:
    M': Updated (final) trained model
    L': Final labeled dataset

1. // INITIALIZATION
2.    For each class c in L, compute initial class weight:
         W[c] ← 1 / (Frequency(c in L) + ε)
3.    Normalize class weights:
         W[c] ← W[c] / (Σ_{c=1}^{C} W[c])
4.    Train initial model M on the labeled dataset L.

5. // ACTIVE LEARNING LOOP
6.    while (Convergence criterion is not met) do
7.         // Step A: Prediction and Feature Extraction on U
8.         For each sample u ∈ U:
9.              P[u] ← M(u)                    // Class probabilities
10.             U_s[u] ← 1 - max(P[u])         // Uncertainty score for sample u
11.             E[u] ← ExtractFeatures(M, u)   // Embedding/feature vector from M
12.
13.         // Step B: Compute Diversity Scores
14.         Cluster all embeddings {E[u] : u ∈ U} into k clusters.
15.         For each sample u ∈ U:
16.             D[u] ← Distance( E[u], Centroid(u's cluster) )
17.
18.         // Step C: Combine Scores for Sampling Priority
19.         For each sample u ∈ U:
20.             S[u] ← α * U_s[u] + β * D[u]
21.
22.         // Step D: Select Samples for Labeling
23.         S_k ← Top-k samples from U with highest S[u]
24.
25.         // Step E: Update Labeled and Unlabeled Datasets
26.         Obtain true labels for samples in S_k.
27.         Update L: L ← L ∪ S_k
28.         Remove S_k from U: U ← U \ S_k
29.
30.         // Step F: Update Class Weights
31.         For each class c in L:
32.             W[c] ← 1 / (Frequency(c in L) + ε)
33.         Normalize W: W[c] ← W[c] / (Σ_{c=1}^{C} W[c])
34.
35.         // Step G: Retrain the Model using Weighted Loss
36.         Retrain model M on updated L using the loss:
             Loss = Σ_{u ∈ L} Σ_{c=1}^{C} W[c] * BinaryCrossEntropy( y[u][c], M(u)[c] )
37.    end while

38. Return final model M' and final labeled dataset L'
```

Fig. 6. WATLAS pseudocode.

- Normalization: Normalize the weights so that they sum to one:

$$W'[c] = \frac{W[c]}{\sum_{i=1}^{C} W[i]}$$

- Baseline Training: Train the initial pre-trained model M on the labeled dataset L.

2. Uncertainty and Diversity Calculation (Step 5 to 17 in the pseudocode):
   - Uncertainty Scores (U): For each sample $s \in U$ (the unlabeled dataset), use M to predict class probabilities $P(s)$ and compute the uncertainty score as:

$$U(s) = 1 - \max(P(s)),$$

where $\max(P(s))$ is the highest predicted probability for sample s.

- Diversity Scores (D): Extract feature embeddings $E(s)$ for each sample s using M (e.g., from an intermediate layer). Cluster these embeddings into k clusters (where k is the number of samples to be queried) and compute the diversity score for each sample as the distance between its embedding and the centroid of its assigned cluster.

  D(s) = Distance(E(s), Centroid of the cluster assigned to s)

3. Adaptive Sampling (Step 18 to 24 in the pseudocode):
   - Combined Sampling Score (S): Combine the uncertainty and diversity scores into a single sampling priority score for each sample:

$$S(s) = \alpha \cdot U(s) + \beta \cdot D(s),$$

     Where α and β are hyperparameters that control the balance between uncertainty and diversity. For instance, we used $\alpha = 0.7$ and $\beta = 0.3$ could be used to prioritize uncertainty while still considering diversity.
   - Sample Selection: Select the top-k samples from U based on the highest values of $S(s)$. These samples are then sent for labeling. We used top 10 samples.
4. Dynamic Weight Updates (Step 25 to 34 in the pseudocode):
   - Recompute Class Weights: After updating the labeled dataset L with the newly labeled samples, recompute the class weights using the updated class frequencies.
   - Normalization: Normalize the updated weights as before to ensure balanced contributions during model retraining.
5. Model Retraining (Step 35 to 38 in the pseudocode):
   - Weighted Loss Function: Retrain the model M on the updated dataset L using a weighted binary cross-entropy loss that incorporates the updated class weights. The loss function for a single sample is given by:

$$\mathscr{L} = -\sum_{u \in L} \sum_{c=1}^{C} W[c] \cdot y[u][c] \cdot \log(M(u)[c])$$

     where y[u][c] is the ground-truth label and $M(u)[c]$ is the predicted probability for sample u and class c.W[c] is the normalized class weight for class c.
   - Iteration: Repeat the sampling and retraining process iteratively until the convergence criterion is met.

4. Case study

The dataset for this study was meticulously collected from an ongoing construction site in a university campus. This unique dataset comprises images capturing assorted objects typically encountered in active construction environments. These images provide a comprehensive view of the dynamic and complex conditions present on-site, which are often overlooked in conventional datasets.

The images reflect authentic construction scenarios, including diverse object types and varying environmental conditions, as depicted in Fig. 7. However, the dataset is inherently imbalanced, with certain object categories underrepresented. This imbalance poses challenges for model performance, as it can lead to biased predictions favoring more prevalent categories.

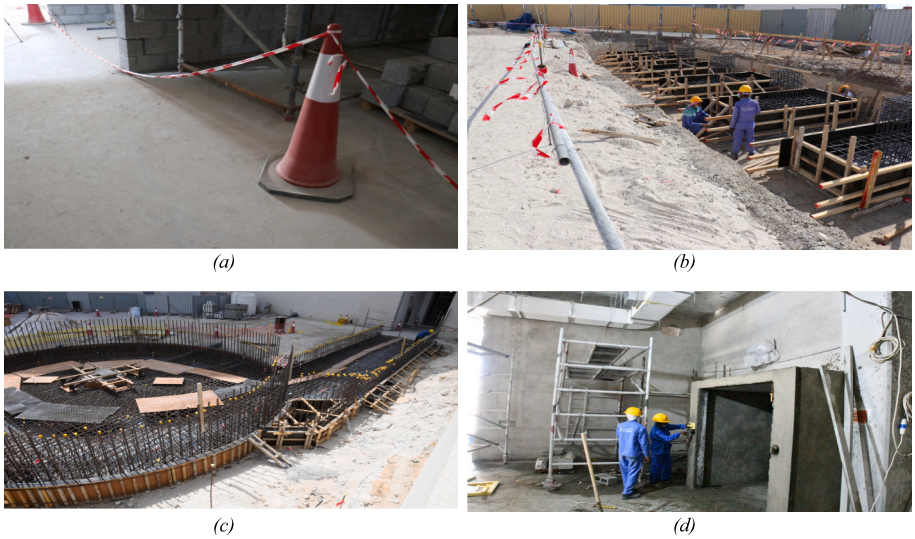
Our dataset consists of 9344 images collected from various areas of the construction site, encompassing 15 object categories that are crucial to construction activities. This diverse dataset is designed to reflect the real-world complexity and variability found in construction environments, providing a robust foundation for training and evaluating our model.

The distribution of object categories is detailed in Table 4, illustrating the varying frequency of each category within the dataset:

Sole labeling of each image takes approximately 1.5 min, and cross-validation takes about 0.5 min, totaling 2 min per image. For 9344

Table 4
Data distribution of image labels.

Image Labels	Instances
Walls	7560
HVAC Rectangular ducts	861
Scaffolding	2559
Ceiling support systems	784
Labels/signs/barricade cones/tapes	2679
Circular ducts	179
Pipes and conduits	179
Formwork	397
Rebar (steel)	480
Workers/people	2029
Electrical items	810
Wires	810
Studs (metal/wood)	325
Drywall insulation filler	290
Drywall panel	54



(a) *(b)* *(c)* *(d)*

Fig. 7. Selected images from the collected dataset, showing (a) corridor safety barriers, (b) foundation formwork and rebar installation, (c) curved wall construction preparation, and (d) HVAC and electrical system installation.

images, at a rate of 2 min per image, the total time equals 18,688 min, which translates to roughly 311.5 h. Based on the $12 per hour estimate from NYUAD crowdsourcing, the total spending on labor for cross-validation and labeling amounts to $3738. The manual annotations comprised a significant portion of the budget, which poses the business case for more active learning initiatives that aim to achieve speedy resources in terms of finances and time while optimizing annotation quality.

Three domain professionals conducted the annotation of the dataset. For discussions on ambiguous cases, the annotators held discussions to reach a conclusion. In case no conclusions were reached, a senior specialist would make the decision. This technique was effective in minimizing label discrepancies and ensuring consistency, which is vital for obtaining reliable results when evaluating the models.

This dataset provides a comprehensive representation of typical objects encountered on construction sites (Fig. 7). The imbalance in category distribution reflects real-world scenarios where certain objects are more prevalent than others. This diversity and imbalance are critical considerations for developing robust machine learning models capable of accurate object classification in complex environments. The dataset supports the evaluation of our proposed ATL Framework by challenging it with both common and rare object categories, ensuring its adaptability and effectiveness across a wide range of construction site conditions.

To address these challenges, our approach integrates advanced machine learning techniques that specifically cater to imbalanced datasets. Sample images with labels are depicted below, illustrating the range of objects and conditions captured in this dataset. This case study not only highlights the practical applications of our model but also underscores the importance of addressing data imbalance in real-world construction monitoring.

4.1. Experimental process

As depicted in Fig. 3, the experimental process involves a detailed configuration of the model to meet the specific requirements of both AL and TL frameworks.

Initially, a pre-trained InceptionV3 model is fine-tuned using a labeled dataset from the construction domain as mentioned above. This step leverages TL, allowing the model to adapt its pre-existing knowledge to recognize patterns and features specific to construction site imagery. By doing so, the model benefits from prior learning, reducing the need for extensive labeled data in this new domain.

Following the initial training phase, the model is evaluated on a pool of unlabeled images. During this evaluation, uncertainty metrics are calculated based on the model's predictions to identify instances where ambiguity exists. These uncertain instances are crucial as they highlight areas where the model's understanding is limited. AL comes into play by selecting these ambiguous samples for labeling, effectively expanding and refining the training dataset.

This iterative process, known as the Active Learning loop, focuses on continuously improving the model's understanding by concentrating on these uncertain areas. The model is retrained with this augmented dataset, enhancing its ability to classify previously uncertain instances accurately. This approach ensures that the model remains adaptable and improves over time, addressing challenges related to data scarcity and imbalance effectively.

4.2. Model comparison for active transfer learning process

This paper systematically evaluates several pre-trained models to identify the optimal architecture for implementing ATL in monitoring construction sites. It uses the full dataset, with 80 % of the data for training and 20 % for testing. The models considered include InceptionV3 + BiLSTM, InceptionV3, VGG16, ResNet152V2, EfficientNetV2B0, and Vision Transformer. Each model's performance is assessed using various metrics, as shown in Table 5 below. These metrics include loss, accuracy, precision, recall, F1-score, Top-1 Accuracy, and Top-2 Accuracy, providing a comprehensive overview of their suitability for the ATL framework.

Among the models:

- The InceptionV3 + BiLSTM model demonstrates the highest overall performance, achieving an accuracy of 0.90 and an F1-score of 0.81. Adding BiLSTM layers is instrumental in capturing temporal patterns, which are crucial for analyzing dynamic construction site imagery. This model benefits from an adaptive sampling strategy, which enables it to selectively focus on challenging samples, enhancing its predictive accuracy over time and aligning well with the ATL approach.
- InceptionV3 achieves an accuracy of 0.88 and demonstrates strong performance while minimizing computational complexity. The architecture leverages factorized convolutions, making it computationally efficient without sacrificing accuracy [48]. This efficiency makes InceptionV3 suitable for scenarios requiring high performance with moderate resource demands, though it lacks the temporal depth offered by the BiLSTM integration.
- VGG16 shows reliable performance with an accuracy of 0.85 and a higher loss of 3.0, indicating slower convergence compared to other models. While its straightforward architecture and depth support robust feature extraction, its large parameter count increases the computational burden [50]. VGG16 is therefore effective in environments where computational resources are not a limiting factor, though it lacks the adaptability for nuanced temporal analysis.
- The ResNet152V2 model provides balanced performance with an accuracy of 0.87 and an F1-score of 0.78. The residual connections within its architecture help alleviate vanishing gradient issues, particularly in deep networks, which enhances training stability and efficiency [51]. ResNet152V2's depth and resilience to training degradation make it a reliable choice, although it does not capture temporal dependencies as effectively as InceptionV3 + BiLSTM.
- With an accuracy of 0.86, EfficientNetV2B0 provides competitive performance alongside computational efficiency, achieved through its compound scaling approach, which optimizes network depth, width, and resolution for balanced performance [52]. This scaling methodology allows EfficientNetV2B0 to perform well in resource-constrained settings, though its accuracy falls slightly behind the more complex InceptionV3 + BiLSTM model.
- The Vision Transformer (ViT) attains an accuracy of 0.87, similar to ResNet152V2, and demonstrates robust capability in handling sequential dependencies, evidenced by its Top-2 Accuracy of 0.90.

Table 5
Illustrates the metrics of the pre-trained model comparison.

Model	Data %	Loss	Accuracy	Precision	Recall	F1-Score	Top-1 Accuracy	Top-2 Accuracy
InceptionV3 + BiLSTM	100 %	2.1	0.9	0.88	0.75	**0.81**	**0.89**	**0.92**
InceptionV3	100 %	2.5	0.88	0.85	0.7	0.77	0.87	0.9
VGG16	100 %	3	0.85	0.82	0.68	0.74	0.84	0.88
ResNet152V2	100 %	2.7	0.87	0.84	0.72	0.78	0.85	0.89
EfficientNetV2B0	100 %	2.8	0.86	0.83	0.7	0.76	0.85	0.88
Vision Transformer	100 %	2.6	0.87	0.84	0.73	0.78	0.86	0.9

Self-attention mechanisms within the ViT architecture enable it to capture intricate patterns within image data, particularly valuable for sequential analysis in construction monitoring [53]. However, while it performs well in feature extraction, its integration with adaptive sampling is less effective than InceptionV3 + BiLSTM.

The comparative analysis indicates that InceptionV3 + BiLSTM is the most suitable model for the ATL process due to its superior handling of spatial and temporal features. While other models, like ResNet152V2 and Vision Transformer, also perform well, they do not integrate Active Learning as effectively as InceptionV3 + BiLSTM.

We chose InceptionV3 because of its robust image feature extraction, which has proven effective in large-scale tasks. Adding BiLSTM layers handles temporal patterns in construction site images, like easily tracking object movements over time and remembering the most common objects. While GRUs or Transformers are alternatives, BiLSTMs balance computational efficiency and capture long-range dependencies better for our needs. Combining TL (using pre-trained models) with AL optimizes limited labeled data, outperforming purely supervised methods (which demand excessive labeling) and self-supervised approaches (which underuse labeled data). This hybrid strategy achieves substantial accuracy with manageable computational costs, which makes it ideal for real-world monitoring.

5. Results and discussion

This section presents the experimental results and analysis of the WATLAS framework. Model performance is assessed using various metrics, different architectures are compared, and the implications of the findings for object classification on construction sites are explored.

5.1. Evaluation metrics

To evaluate our model's performance and the impact of various sampling strategies comprehensively, we employ a range of metrics that collectively assess classification accuracy, reliability, and effectiveness. These metrics provide a holistic understanding of the model's strengths and weaknesses, particularly in handling imbalanced datasets common in construction site imagery.

We include accuracy, precision, recall, and F1-score as key metrics for evaluating classification performance. These metrics are well-known and widely used in the field; thus, detailed explanations and equations are omitted here for brevity and can be referred to in standard references.

Additionally, we use Top-1 Accuracy to assess the proportion of predictions where the model's top-ranked output matches the true label, reflecting the ability to make correct predictions on the first attempt. Top-2 Accuracy extends this by evaluating whether the true label is among the top two predicted labels, a particularly valuable metric in scenarios with closely related classes.

$$Top1\ Accuracy = \frac{\text{Number of correct top predictions}}{Total\ number\ of\ predictions}$$

$$Top2\ Accuracy = \frac{Number\ of\ times\ true\ label\ is\ in\ top\ 2\ predictions}{Total\ number\ of\ predictions}$$

In addition to these metrics, we analyze category-wise accuracy to understand the model's performance across different object classes, identifying areas for further refinement. This multifaceted evaluation approach ensures robust insights into the model's capabilities, enabling high overall performance and consistency across diverse datasets.

5.2. Model performance

The performance of various Active Learning strategies, WATLAS, ATLAS, Combined Sampling, Diversity Sampling, BALD, and Uncertainty Sampling, was evaluated based on Accuracy and F1 Score metrics, as shown in Fig. 8. To evaluate these strategies, an 80/20 split was used. 80 % of the data was allocated for training the models, while the remaining 20 % was reserved for testing. The left plot illustrates the Accuracy progression as a function of data percentage, while the right plot provides a comparative distribution of F1 Scores across the strategies. Table 9 at the end of the paper provides a summary of all the results and metrics for each one of the approaches.

In terms of accuracy, WATLAS demonstrates the highest performance, reaching approximately 0.97 at full data usage, highlighting its efficiency in utilizing labeled data. ATLAS and BALD follow with a steady increase, achieving up to 0.90, though consistently performing below WATLAS. Combined Sampling, and Diversity Sampling show moderate accuracy, both improving gradually to around 0.85, with Diversity Sampling consistently performing slightly lower. Uncertainty Sampling shows the lowest accuracy overall, peaking at about 0.83, indicating limited effectiveness in identifying informative samples.

When analyzing F1 Score distributions, WATLAS again leads with the highest median and the largest range, reflecting its adaptability across diverse datasets. ATLAS and BALD, while slightly lower in median F1 Score, maintain a consistent performance with a narrower distribution. Combined Sampling, Diversity Sampling, and Uncertainty Sampling exhibit lower medians and smaller ranges, suggesting reduced effectiveness and limited adaptability to complex datasets.

Overall, WATLAS, ATLAS, and BALD stand out as the most effective strategies, with WATLAS leading in both metrics. This highlights WATLAS's weighted Active Transfer Learning approach as the optimal method for balancing accuracy and robustness in model performance.

5.3. Category-wise performance analysis

The category-wise performance analysis, as illustrated in Fig. 9, highlights the effectiveness of the WATLAS strategy in managing the imbalanced nature of construction site data. WATLAS excels particularly in underrepresented classes by applying adaptive weighting to improve detection accuracy. A summary of all the results can be found at the end of the paper in Table 10.

For well-represented categories such as Walls (7560 samples) and Scaffolding (2559 samples), most strategies, including WATLAS, demonstrate strong performance. However, WATLAS maintains an edge through its refined detection capabilities enabled by weighted sampling, ensuring consistent accuracy across these prevalent categories.

In moderately represented categories like HVAC Rectangular Ducts (861), Labels/Signs/Barricade Cones/Tapes (2679), and Workers/People (2029), WATLAS's adaptive weighting proves beneficial. It balances performance across diverse items critical for site safety and monitoring, ensuring that these essential elements are accurately detected.

For rare categories such as Circular Ducts (179), Pipes and Conduits (179), Formwork (397), Rebar (Steel) (480), Studs (Metal/Wood) (325), and Drywall-related elements (54 for panels and 290 for insulation filler), WATLAS achieves noticeably higher accuracy compared to other strategies. By applying higher weights to these underrepresented classes, WATLAS ensures the accurate detection of these crucial but less frequent elements. This approach addresses the specific needs of construction environments where detecting every object, regardless of frequency, is vital.

The weighted sampling approach of WATLAS contrasts with standard strategies that often struggle to generalize well in rare categories. This underscores WATLAS's adaptability to imbalanced datasets and its practical utility in real-world, high-risk scenarios. The innovative application of weighted adaptive sampling allows WATLAS to outperform traditional methods by focusing on the most informative samples and ensuring comprehensive coverage across all object categories.

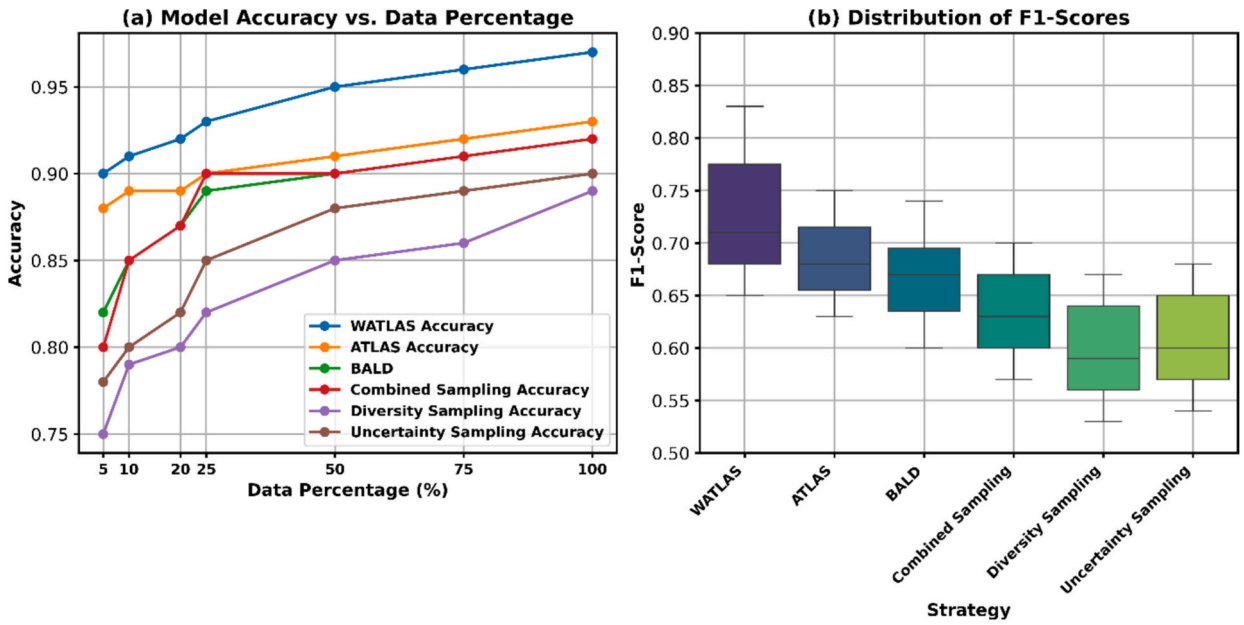


Fig. 8. Model performance comparison showing (a) accuracy trends across different data percentages for various sampling strategies, and (b) F1-score distribution boxplots demonstrating the relative effectiveness of each sampling approach.

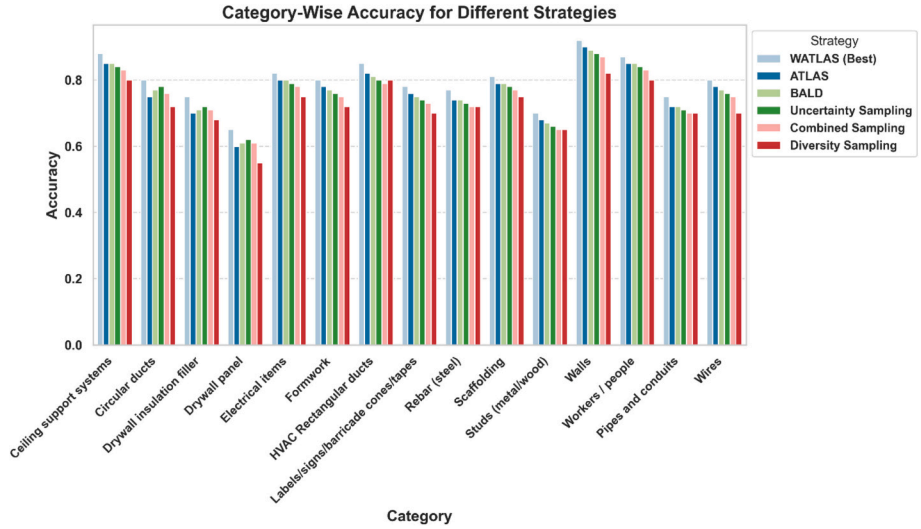


Fig. 9. Category Wise Performance on Various Sampling Strategies.

5.4. Confusion matrix and error analysis

We investigated the WATLAS model performance with labeled data proportions of 5 %, 20 %, 50 %, and 100 %. This examination included analyzing general accuracy patterns and confusion matrices for all object classes. The WATLAS Confusion Matrices at various labeling proportions are presented in Fig. 10, while Fig. 11 displays Error Rates by Category with minimal (5 %) and complete (100 %) labeling for comparison.

5.4.1. Common misclassifications

- Underrepresented categories: due to their small sample sizes, drywall panels and circular ducts (54 and 179 instances, respectively) exhibited lower accuracy, particularly at lower labeling percentages. This limited representation often led the model to confuse drywall panels with drywall insulation fillers and circular ducts with HVAC rectangular ducts.
- Lighting and occlusion: objects such as rebar and wires were frequently misclassified in cluttered scenes or those with poor lighting. Shadows or partial occlusions obscured key features, reducing the model's ability to distinguish subtle differences.
- Class overlap: certain classes occasionally appeared together in the same region of the image (e.g., labels/signs/barricades near workers/people). This overlap sometimes led to false positives for both categories, as shown in the confusion matrices (Fig. 10).

Across all labeled data proportions, the diagonal cells in each confusion matrix (Fig. 10) became darker (indicating higher true positive rates) as more labeled data became available. However, off-diagonal cells in categories such as drywall panels and circular ducts remained relatively pronounced, reaffirming the difficulty in learning rare classes. By contrast, classes with abundant data (e.g., walls and labels/signs/

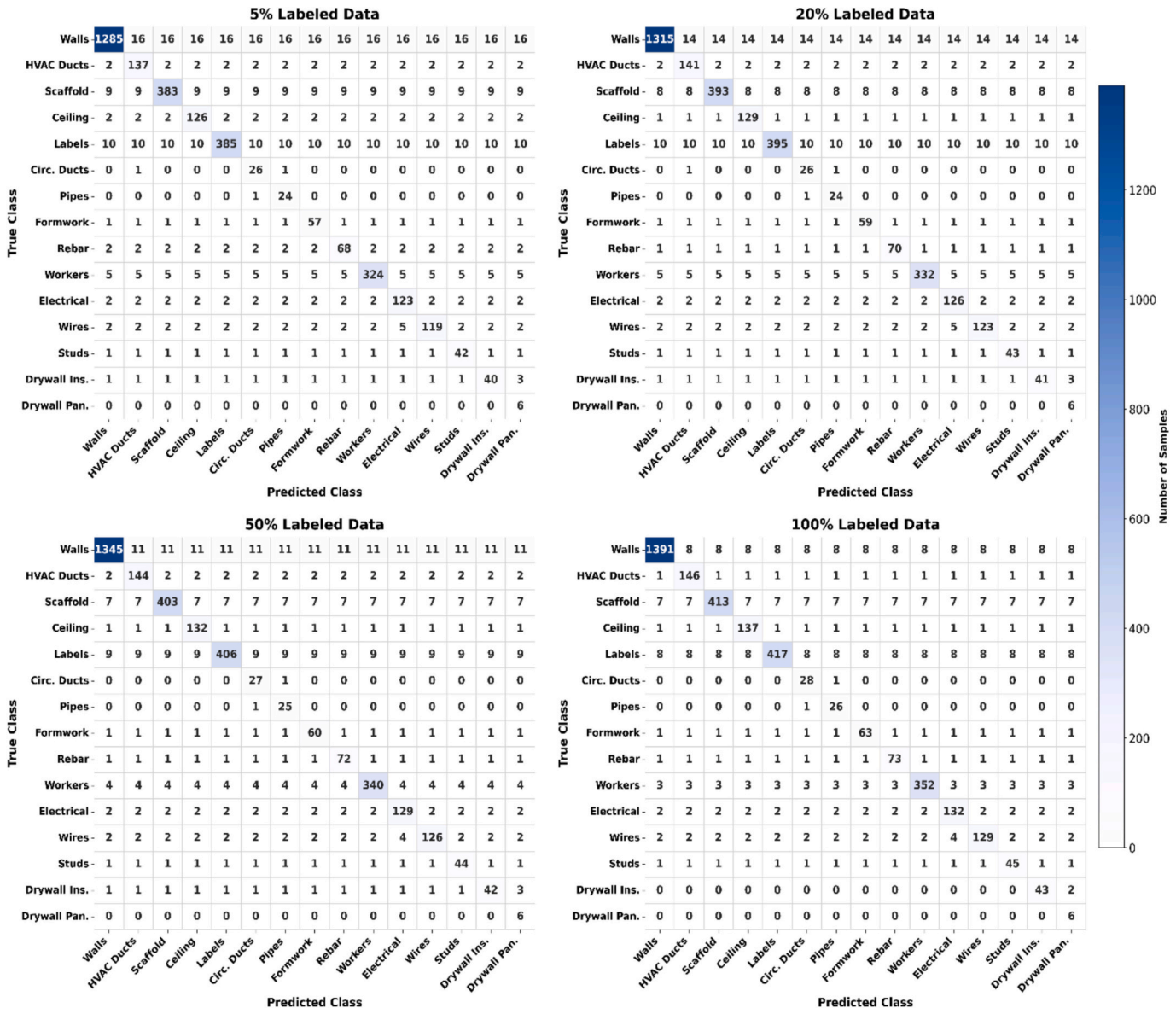


Fig. 10. WATLAS confusion matrices across different percentages of labeled data.

barricades) showed significant improvements, reflecting WATLAS's effectiveness when sufficient training examples are provided.

5.4.2. *Impact of increasing labeled data*

Fig. 11 illustrates the impact of increasing the amount of labeled data across different categories. The effect is particularly pronounced in well-represented categories, where walls exhibit a significant reduction in error rate, decreasing from 15 % to 8 %, highlighting the model's ability to refine its predictions with additional data. Similarly, scaffolding shows substantial improvement, with error rates dropping from 25 % to 19 %.

In contrast, categories with fewer labeled samples demonstrate limited improvement. Drywall panels, for example, see only a marginal reduction from 36 % to 35 %, suggesting that increasing data quantity alone may not be sufficient for underrepresented classes. A slightly better trend is observed in circular ducts, where the error rate improves from 25 % to 20 %.

For mid-representation categories, a more moderate yet consistent enhancement is observed. Both formwork and electrical items exhibit steady reductions in error rates when additional labeled data is introduced, reinforcing the model's capacity to benefit from increased supervision in these cases.

5.5. *Computational environment and performance*

We detail the specifications of our experimental environment in Table 6 to ensure reproducibility and provide transparency regarding the computational resources utilized in this paper. The case study was carried out on a High-Performance Computing (HPC) cluster, which facilitated the efficient training of our deep learning models.

5.5.1. *Training time and computational cost*

The training time for each experiment varied depending on the percentage of labeled data used and the number of active learning iterations. On average, the active learning loop required 8–10 iterations to reach the best performance, depending on the model. Each iteration, including model retraining and sample selection, took approximately 45–55 min based on the number of samples. This level of computational demand is feasible for environments with moderate hardware resources, though further optimization may be necessary for real-time or edge

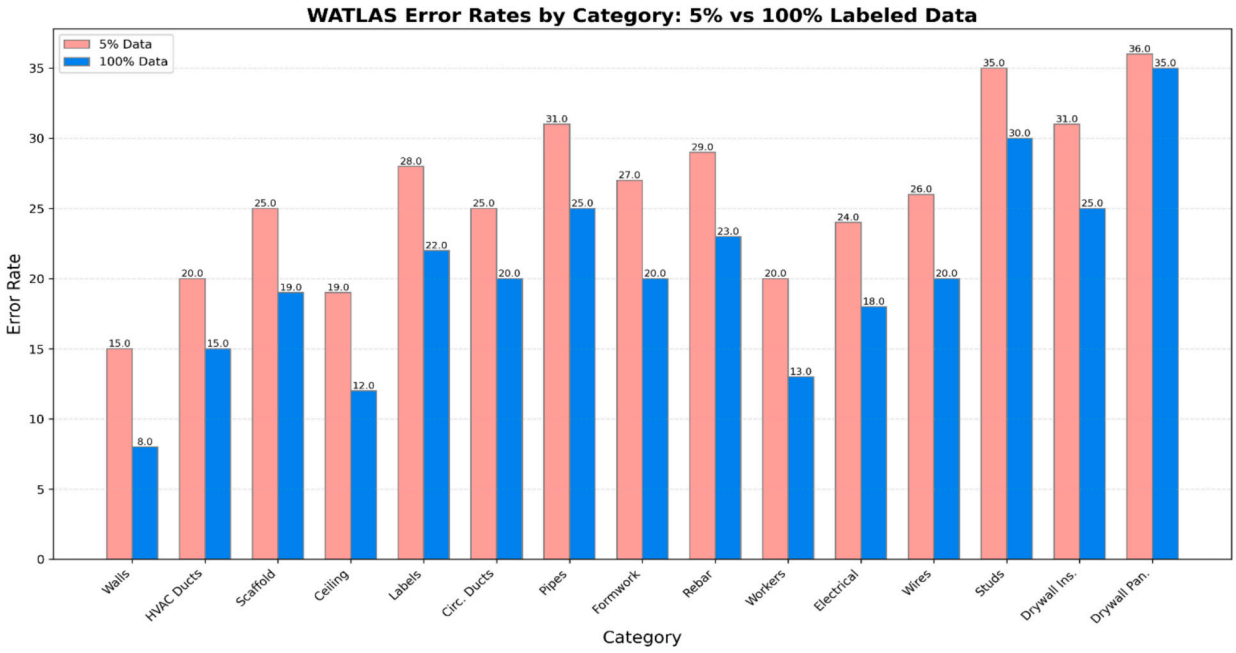


Fig. 11. WATLAS error rates by category: 5 % vs. 100 % labeled data.

Table 6
Hardware and software specifications.

Component	Specification
GPU	NVIDIA Tesla V100-SXM2-32GB
CPU	Intel Xeon Gold 6148R (40 cores)
RAM	1 TB DDR4
Storage	2 TB SSD
OS	Rocky Linux 8.9 (Green Obsidian)
Python Version	Python 3.9
Frameworks	TensorFlow 2.9.0
Libraries	NumPy 1.23, OpenCV 4.6, Scikit-learn 1.0.2, Keras 2.9.0

scenarios. On average, training a single model instance with the full dataset required approximately hours, as shown in Table 7. We conducted a detailed cost estimation for the Google Cloud Platform (GCP) virtual machine configured to match our hardware specifications. Our VM configuration included the following components, as in Table 6. Based on current on-demand pricing from GCP, the estimated hourly cost was derived as shown in Supplementary Table 11, and the total estimated cost amounts to approximately $11.30 per hour.

- The InceptionV3 + BiLSTM model demonstrates superior accuracy (90 %) with a reasonable training time (~8 h), making it the most effective model for classifying construction site imagery.
- The Vision Transformer has the highest number of trainable parameters (88 M), leading to longer training times (~12 h) and higher costs ($135.6), but its performance is comparable to ResNet152V2 (87 % accuracy).
- VGG16 has the smallest number of trainable parameters (2.56 M) and is computationally efficient (~3 h), but its accuracy (85 %) is lower than other models.

Table 7
Model training on full dataset.

Model	Trainable Parameters	Training Time (hours)	Inference Time (ms)	Cost ($)	Accuracy (%)
InceptionV3 + BiLSTM	41,190,351	~8	23	90.4	90
InceptionV3	40,731,855	~6	19	67.8	88
VGG16	2,561,436	~3	11	33.9	85
ResNet152V2	2,126,876	~4	18	45.2	87
EfficientNetV2B0	33,958,684	~5	15	56.5	86
Vision Transformer	88,245,020	~12	38	135.6	87

EfficientNetV2B0 balances accuracy (86 %) and computational cost ($56.5), making it suitable for resource-constrained environments. Unlike the Inception V3 + BiLSTM or Vision Transformer models, VGG16 and ResNet152V2 have a substantially smaller number of trainable parameters. This indicates their lower computational cost, making them a better fit for edge devices with a lower processing capacity.

As expected, more complex models outperform them in accuracy, but the efficiency losses from using these models can be compensated for by the gains made in resource-constrained real-time applications. As noted by Hao et al. [54], InceptionV3 is well known for its trade-off between accuracy and computation. Its optimized structure makes it suitable for edged devices with low to moderate computational power. Regarding edge deployment, the selected model varies with the application's needs because high precision needs to be balanced with the deployment's rigidity. For constrained real-time resource applications, the efficiency of using InceptionV3, if resources permit, brings forth a positive cost-benefit impact. On the other hand, extremely low-level edge devices may need the much lighter VGG16 or even ResNet152V2.

5.5.2. *Final model computational performance across data percentages*

The proposed InceptionV3 + BiLSTM model's performance was evaluated using different percentages of labeled data to demonstrate its

Table 8
Final model computational performance.

Strategy	Data Percentage (%)	Training Time (hours)	Cost ($)	Accuracy (%)
WATLAS	5	~0.5	5.65	90
	10	~1	11.3	91
	20	~2	22.6	92
	25	~2.5	28.25	93
	50	~4	45.2	95
	75	~6	67.8	96
	100	~8	90.4	97

efficiency with minimal labeled data as shown in Table 8.

With only 5 % labeled data, the model achieves impressive accuracy of 90 % and only $5.65 cost for training, showcasing its ability to significantly reduce manual labeling efforts as more labeled data is added (e.g., at 10 %, accuracy reaches 91 %, and at 20 %, accuracy improves to 92 %), WATLAS effectively utilizes additional data to enhance performance. The model achieves near-optimal accuracy (95 %) with just half the dataset labeled (50 %), highlighting its efficiency in handling imbalanced datasets through adaptive sampling.

6. Limitations and future outlook

Despite its strengths, the WATLAS framework has some limitations. The reliance on pre-trained models means that the initial feature extraction might not fully capture domain-specific nuances of construction imagery. Additionally, the computational complexity of adaptive sampling and model retraining can be resource-intensive, potentially limiting its scalability in environments with constrained computational resources.

Another limitation is the fact that images are currently sourced from a single geographic region. Future work will explore expanding our dataset to include diverse environmental conditions, such as snowy or rainy scenarios, to improve model generalization.

WATLAS's performance depends on the similarity between source (ImageNet) and target (construction) domains. While our fine-tuning mitigates domain shifts, future work should quantify this overlap. Additionally, initial labeled data quality impacts adaptive sampling; noisy labels may propagate errors.

Complex architectures like InceptionV3 + BiLSTM increase overfitting risk, especially with smaller labeled datasets. Various regularization techniques need further validation for active learning scenarios.

Future research should focus on refining the WATLAS framework to address its limitations. This includes developing more efficient algorithms for adaptive sampling that reduce computational overhead while maintaining accuracy. Exploring domain-specific pre-training techniques could also enhance feature extraction tailored to construction environments. Furthermore, expanding the application of WATLAS to other domains with similar data challenges could validate its versatility and effectiveness.

We will examine implementing lightweight versions of WATLAS using techniques like knowledge distillation or model compression using quantization, pruning to optimize model size and minimize computational demand. Early-exit networks are an alternative that improves performance since they enable predictions to be made at intermediate stages for simpler inputs, lessening the need for extensive computation at later stages. These techniques will enable WATLAS to be deployed on edge devices for real-time monitoring, which will also be explored for more specific applications. This would allow to combine WATLAS with Building Information Modeling (BIM) or digital twin technologies for comprehensive construction site analysis, improving project management and safety monitoring.

In addition, we plan to explore additional loss functions such as the Focal Loss, which has shown promise in handling class imbalance in object detection tasks, and the Dice Loss, which is adequate for segmentation tasks and could improve our model's performance on overlapping objects. We will also investigate Wing Loss, which has demonstrated effectiveness in face alignment tasks and may enhance our model's robustness to challenging samples.

Continued advancements in this area will contribute to more reliable and efficient monitoring systems in construction and beyond, ultimately enhancing safety and operational efficiency.

7. Conclusions

The paper introduced WATLAS, an advanced framework that addresses critical challenges in construction site imagery classification, particularly using imbalanced datasets. Through its innovative framework, WATLAS significantly improves detection accuracy, particularly for underrepresented classes like Circular Ducts and Drywall Panels, which are typically overlooked in conventional approaches. Integrating Active Learning, Transfer Learning, and Weighted Adaptive Sampling techniques makes WATLAS a robust solution tailored to real-world scenarios where imbalanced data often undermines model performance.

One of the standout results from WATLAS is its ability to achieve 90 % accuracy with only 5 % of the labeled data. This capability demonstrates the power of the framework to minimize the data required for effective learning, thereby reducing the burden of manual labeling in scenarios where data is scarce.

WATLAS's ability to adaptively focus on high-impact samples while maintaining class balance demonstrates its potential to set a new benchmark in construction site analysis. This framework highlights how advanced sampling and weighting strategies can transform traditional machine learning pipelines by reducing biases and accelerating model convergence.

The main contributions of this paper are:

1. First research application of ATLAS: To the authors' knowledge, this paper represents the first formal research implementation and publication of the ATLAS framework, a cutting-edge extension of active transfer learning to adaptive active learning, and sets a new standard for research in this Construction domain.
2. Adaptive class weights for improved representation: Unlike traditional Active Learning frameworks, the WATLAS framework computes class weights on each sampled batch based on their distributions within the training set. This weighted scheme gives underrepresented classes a proportional influence on the loss function, reducing bias in the model.
3. Improved sample selection with adaptiveness: The framework can anticipate its future state by integrating Weighted Active Adaptive Sampling and Transfer Learning. While it does not initially know the labels of the sampled items, it accounts for the fact that they will be labeled, enabling it to make more informed and strategic sampling decisions based on the expected outcomes.
4. Class-aware model refinement: WATLAS enhances model performance by training a secondary classifier to assess the correctness of predictions, enabling intelligent sample selection. The framework streamlines labeling efforts, optimizes resource utilization, and accelerates convergence by focusing on potentially misclassified instances.

In the future, the outcomes from this study will offer multiple new pathways for exploration and practical application. With little labeled data, WATLAS's demonstrated high accuracy suggests that it could be quickly operated in new construction projects, where annotated datasets are limited or changing. Future research could explore how WATLAS adapts to continuous data streams or changing site conditions, enabling real-time model updates as construction progresses.

The framework's class-aware sampling and weighting strategies may inspire new approaches in other domains where rare event detection is critical, such as industrial inspection, disaster response, or medical imaging. Investigating how WATLAS can be integrated with human-in-the-loop systems, where expert feedback further guides sample selection, could enhance its efficiency and reliability.

Finally, as construction sites become increasingly digitized, there is a significant opportunity to apply WATLAS in combination with automated robotics, sensor networks, and digital twin platforms, supporting a new generation of intelligent, adaptive monitoring systems that improve safety, productivity, and decision-making on-site.

CRediT authorship contribution statement

Karunakar Reddy Mannem: Writing – original draft, Visualization,

Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Samuel A. Prieto:** Writing – review & editing, Visualization, Data curation. **Borja García de Soto:** Writing – review & editing, Validation, Supervision, Resources, Funding acquisition. **Fernando Bacao:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research benefited from utilizing resources available at the Core Technology Platforms (CTP) of New York University Abu Dhabi (NYUAD). In particular, the algorithms developed in this study used the research computing services at NYUAD's Center for Research Computing and High-Performance Computing (HPC). This research was partially supported by different Centers at NYUAD. In particular, the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES) funded by Tamkeen under the NYUAD Research Institute Award CG013, the Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001, and the Center for Artificial Intelligence and Robotics (CAIR), funded by Tamkeen under the NYUAD Research Institute Award CG010. This work was also supported by national funds through FCT (Foundation for Science and Technology), under the project - UIDB/04152 - Information Management Research Centre (MagIC)/ NOVA IMS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.autcon.2025.106297.

Data availability

Data will be made available on request.

References

[1] S.D. Datta, M. Islam, M.H.R. Sobuz, S. Ahmed, M. Kar, Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: a comprehensive review, Heliyon 10 (5) (Feb. 2024) e26888, https://doi.org/10.1016/j.heliyon.2024.e26888.

[2] L. Liu, Z. Song, P. Zhou, X. He, L. Zhao, AI-based rock strength assessment from tunnel face images using hybrid neural networks, Sci. Rep. 14 (1) (2024) 17512, https://doi.org/10.1038/s41598-024-68704-0.

[3] M. Arfan, S. Sumardi, H. Huboyo, Advancing workplace safety: a deep learning approach for personal protective equipment detection using single shot detector, in: IWAIIP 2023 - Conference Proceeding: International Workshop on Artificial Intelligence and Image Processing, 2023, pp. 127–132, https://doi.org/10.1109/IWAIIP58158.2023.10462804.

[4] R. Duan, H. Deng, M. Tian, Y. Deng, J. Lin, SODA: a large-scale open site object detection dataset for deep learning in construction, Autom. Constr. 142 (2022) 104499, https://doi.org/10.1016/j.autcon.2022.104499.

[5] X. Chen, A. Chang-Richards, F.Y.Y. Ling, T.W. Yiu, A. Pelosi, N. Yang, Digital technologies in the AEC sector: a comparative study of digital competence among industry practitioners, Int. J. Constr. Manag. (2025) 1–14, https://doi.org/10.1080/15623599.2024.2304453.

[6] S.I. Hassan, S.A. Syed, S.W. Ali, H. Zahid, S. Tariq, M. Mohd Suud, M.M. Alam, Systematic literature review on the application of machine learning for the prediction of properties of different types of concrete, PeerJ Comput. Sci. 10 (2024) e1853, https://doi.org/10.7717/PEERJ-CS.1853.

[7] Qiuchen Lu, Ajith Kumar Parlikad, Philip Woodall, Xiang Xie, Zhenglin Liang, Eirini Konstantinou, James Heaton, Jennifer Schooling, Developing a digital twin at building and city levels: case study of West Cambridge campus, J. Manag. Eng. 36 (3) (2020) 05020004, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000763.

[8] R. Sacks, I. Brilakis, E. Pikas, H.S. Xie, M. Girolami, Construction with digital twin information systems, Data-Centric Eng. 1 (2020) e14, https://doi.org/10.1017/dce.2020.16.

[9] T. Kikuta, P.-J. Chun, Development of an action classification method for construction sites combining pose assessment and object proximity evaluation, J. Ambient. Intell. Humaniz. Comput. 15 (4) (2024) 2255–2267, https://doi.org/10.1007/s12652-024-04753-7.

[10] J. Liu, H. Luo, H. Liu, Deep learning-based data analytics for safety in construction, Autom. Constr. 140 (2022) 104302, https://doi.org/10.1016/j.autcon.2022.104302.

[11] H. Kim, J.-S. Yi, Image generation of hazardous situations in construction sites using text-to-image generative model for training deep neural networks, Autom. Constr. 166 (2024) 105615, https://doi.org/10.1016/j.autcon.2024.105615.

[12] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, Curran Associates, Inc, 2012. Accessed: Nov. 30, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

[13] K.R. Mannem, E. Mengiste, S. Hasan, B. García de Soto, R. Sacks, Smart audio signal classification for tracking of construction tasks, Autom. Constr. 165 (2024) 105485, https://doi.org/10.1016/j.autcon.2024.105485.

[14] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (Jun. 2017) 1137–1149, https://doi.org/10.1109/TPAMI.2016.2577031.

[15] J. Li, G. Zhou, D. Li, M. Zhang, X. Zhao, Recognizing workers' construction activities on a reinforcement processing area through the position relationship of objects detected by faster R-CNN, Eng. Constr. Archit. Manag. 30 (4) (Jan. 2022) 1657–1678, https://doi.org/10.1108/ECAM-04-2021-0312.

[16] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Autom. Constr. 20 (8) (Dec. 2011) 1143–1155, https://doi.org/10.1016/j.autcon.2011.04.016.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Xu Bing, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial networks, Commun. ACM 63 (11) (Oct. 2020) 139–144, https://doi.org/10.1145/3422622.

[18] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

[19] Z. Liu, Y. Cao, Y. Wang, W. Wang, Computer vision-based concrete crack detection using U-net fully convolutional networks, Autom. Constr. 104 (Aug. 2019) 129–139, https://doi.org/10.1016/j.autcon.2019.04.005.

[20] S. Zhang, L. Zhang, Construction site safety monitoring and excavator activity analysis system, arXiv (2021), https://doi.org/10.48550/ARXIV.2110.03083.

[21] Z. Jiang, J.I. Messner, Computer vision applications in construction and asset management phases: a literature review, J. Inf. Technol. Constr. 28 (Apr. 2023) 176–199, https://doi.org/10.36680/j.itcon.2023.009.

[22] Behnam Sherafat, Changbum Ahn, Reza Akhavian, Amir Behzadan, Mani Golparvar-Fard, Hyunsoo Kim, Yongcheol Lee, Abbas Rashidi, Ehsan Azar, Automated methods for activity recognition of construction workers and equipment: state-of-the-art review, J. Constr. Eng. Manag. 146 (6) (Jun. 2020) 03120002, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001843.

[23] Z. Wu, Y. Feng, Y. Demiris, P. Angeloudis, Development of a digital twin-based simulation system and a novel synthetic video dataset for enhancing computer vision in construction site safety, in: Presented at the 2024 European Conference on Computing in Construction, Jul. 2024, https://doi.org/10.35490/EC3.2024.195.

[24] Jiaqi Li, Qi Miao, Zheng Zou, Huaguo Gao, Lixiao Zhang, Zhaobo Li, Nan Wang, A review of computer vision-based monitoring approaches for construction workers' work-related behaviors, IEEE Access 12 (2024) 7134–7155, https://doi.org/10.1109/ACCESS.2024.3350773.

[25] H. Kath, P.P. Serafini, I.B. Campos, T.S. Gouvêa, D. Sonntag, Leveraging transfer learning and active learning for data annotation in passive acoustic monitoring of wildlife, Ecol. Inform. 82 (Sep. 2024) 102710, https://doi.org/10.1016/j.ecoinf.2024.102710.

[26] B. Settles, Active learning literature survey, in: University of Wisconsin-Madison Department of Computer Sciences, Technical Report, 2009. Accessed: Nov. 30, 2023. [Online]. Available: https://minds.wisconsin.edu/handle/1793/60660.

[27] B. Settles, From theories to queries: active learning in practice, in: Active Learning and Experimental Design workshop in conjunction with AISTATS 2010, JMLR Workshop and Conference Proceedings, Apr. 2011, pp. 1–18. Accessed: Nov. 30, 2023. [Online]. Available: https://proceedings.mlr.press/v16/settles11a.html.

[28] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: representing model uncertainty in deep learning, in: Proceedings of The 33rd International Conference on Machine Learning, PMLR, Jun. 2016, pp. 1050–1059. Accessed: Nov. 30, 2023. [Online]. Available: https://proceedings.mlr.press/v48/gal16.html.

[29] Y. Gal, R. Islam, Z. Ghahramani, Deep bayesian active learning with image data, arXiv (Mar. 08, 2017), https://doi.org/10.48550/arXiv.1703.02910 arXiv: 1703.02910.

[30] O. Sener, S. Savarese, Active learning for convolutional neural networks: a core-set approach, arXiv (Jun. 01, 2018), https://doi.org/10.48550/arXiv.1708.00489 arXiv:1708.00489.

[31] Asim Costa Smailagic, Pedro Young Noh, Hae Walawalkar, Devesh Khandelwal, Kartik Galdran, Adrian Mirshekari, Mostafa Fagert, Jonathon Xu, Susu Zhang, Pei Campilho Aurélio, MedAL: accurate and robust deep active learning for medical image analysis, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2018, pp. 481–488, https://doi.org/10.1109/ICMLA.2018.00078.

[32] J. Kim, J. Hwang, S. Chi, J. Seo, Towards database-free vision-based monitoring on construction sites: a deep active learning approach, Autom. Constr. 120 (2020) 103376, https://doi.org/10.1016/j.autcon.2020.103376.
[33] S. Sinha, S. Ebrahimi, T. Darrell, Variational adversarial active learning, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2019, pp. 5971–5980, https://doi.org/10.1109/ICCV.2019.00607.
[34] D.D. Lewis, W.A. Gale, A sequential algorithm for training text classifiers, in: Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1994, https://doi.org/10.1007/978-1-4471-2099-5_1.
[35] B. Settles, Active Learning Literature Survey, University of Wisconsin-Madison Department of Computer Sciences, 2009 [Online]. Available: https://api.semanticscholar.org/CorpusID:324600.
[36] D.A. Cohn, L.E. Atlas, R.E. Ladner, Improving generalization with active learning, Mach. Learn. 15 (1994) 201–221, https://doi.org/10.1007/BF00993277.
[37] H. Kim, H. Kim, Y.W. Hong, H. Byun, Detecting construction equipment using a region-based fully convolutional network and transfer learning, J. Comput. Civ. Eng. 32 (2) (2018) 04017082, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731.
[38] E. Mengiste, K.R. Mannem, S.A. Prieto, B. Garcia de Soto, Transfer-learning and texture features for recognition of the conditions of construction materials with small data sets, J. Comput. Civ. Eng. 38 (1) (2024) 04023036, https://doi.org/10.1061/JCCEE5.CPENG-5478.
[39] V.V. Bharathi, S. Prieto, B.G. de Soto, J. Teizer, Automating construction safety inspections using robots and unsupervised deep domain adaptation by backpropagation, in: Int. Symp. Autom. Robot. Constr. ISARC Proc., vol. 2024 Proceedings of the 41st ISARC, Lille, France, Jun. 2024, pp. 855–862, https://doi.org/10.22260/ISARC2024/0111.
[40] L. Chen, Y. Wang, M.-F.F. Siu, Detecting semantic regions of construction site images by transfer learning and saliency computation, Autom. Constr. 114 (Jun. 2020) 103185, https://doi.org/10.1016/j.autcon.2020.103185.
[41] T. Frick, D. Antognini, M. Rigotti, I. Giurgiu, B. Grewe, C. Malossi, Active learning for imbalanced civil infrastructure data, in: L. Karlinsky, T. Michaeli, K. Nishino (Eds.), Computer Vision – ECCV 2022 Workshops, Springer Nature, Switzerland, 2023, pp. 283–298, https://doi.org/10.1007/978-3-031-25082-8_19.
[42] Y. Zheng, Y. Gao, S. Lu, K.M. Mosalam, Multistage semisupervised active learning framework for crack identification, segmentation, and measurement of bridges, Comput. Aid. Civ. Inf. Eng. 37 (9) (2022) 1089–1108, https://doi.org/10.1111/mice.12851.
[43] X. Peng, X. Jin, S. Duan, C. Sankavaram, Active learning-assisted semi-supervised learning for fault detection and diagnostics with imbalanced dataset, IISE Trans. 55 (7) (Jul. 2023) 672–686, https://doi.org/10.1080/24725854.2022.2074579.
[44] A. Singh, S. Chakraborty, Deep active transfer learning for image recognition, in: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, Glasgow, United Kingdom, Jul. 2020, pp. 1–9, https://doi.org/10.1109/IJCNN48605.2020.9207391.
[45] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (Sep. 2009) 1263–1284, https://doi.org/10.1109/TKDE.2008.239.
[46] X. Gao, Z. Chen, S. Tang, Y. Zhang, J. Li, Adaptive weighted imbalance learning with application to abnormal activity recognition, Neurocomput 173 (P3) (Jan. 2016) 1927–1935, https://doi.org/10.1016/j.neucom.2015.09.064.
[47] R. Monarch, R. Munro, C.D. Manning, Human-in-the-loop machine learning: active learning and annotation for human-centered AI, Manning (2021) 1–272 [Online]. Available: https://books.google.ae/books?id=LCh0zQEACAAJ.
[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826, https://doi.org/10.1109/CVPR.2016.308.
[49] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[50] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition vol. abs/1409.1556, CoRR, 2014, pp. 1–14.
[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[52] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks. Proc. 36th Int. Conf. Mach. Learn. (ICML) 97, Long Beach, CA, USA, 2019, pp. 6105–6114. ArXiv (2019) vol. abs/1905.11946.
[53] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: transformers for image recognition at scale, in: Proc. Int. Conf. Learn. Represent. (ICLR), 2021 [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy.
[54] J. Hao, P. Subedi, L. Ramaswamy, I.K. Kim, Reaching for the sky: maximizing deep learning inference throughput on edge devices with AI multi-tenancy, ACM Trans. Internet Technol. 23 (1) (Feb. 2023) 1–33, https://doi.org/10.1145/3546192.