The use of artificial intelligence to identify objects in a construction site

Alexandre Almeida Del Savio[1], Ana Luna[1], Daniel Cárdenas-Salas[2], Mónica Vergara Olivera[1], Gianella Urday Ibarra[2]

[1] Civil Engineering, Faculty of Engineering and Architecture, University of Lima, Lima, Peru
[2] Systems Engineering, Faculty of Engineering and Architecture, University of Lima, Lima, Peru
delsavio@gmail.com, aluna@ulima.edu.pe, decarden@ulima.edu.pe, mvergara@ulima.edu.pe, g.urday97@gmail.com

Abstract. The construction industry invests a large amount of effort and resources in construction processes such as the follow-up, control, and monitoring of construction works, which, compared to other areas, present a low level of automation. Thus, increasing automation would reduce the times and costs of such activities. This research aims to evaluate a computer vision technique to identify objects of interest in construction sites, from videos and images of drones and static surveillance cameras. The "You Look Only Once" (YOLO) object detection neural network was used to identify eight classes of objects in 1000 drone images and 1046 static camera images of a construction site, achieving an accuracy varying between 78.8% to 82.8% and 73.56% to 93.76%, respectively. The feasibility of using classification algorithms to identify complex objects such as trucks and cranes was verified. Its application can be extended to various other forms to have an intelligent and automated process of monitoring and control project construction activities.

Abstract: artificial intelligence, machine learning, computer vision techniques, neural network models, construction monitoring, YOLO.

1. Introduction

New trends in the construction industry are oriented towards the automation of follow-up, control, and monitoring phases, which are critical stages within the construction process [1]. These stages have direct control over the eventualities that may occur during the execution of the process and improve preventive and corrective decision-making processes, almost in real-time [2]. One of the most significant worldwide challenges that this sector is facing consists of improving the monitoring of construction projects to meet adequate deadlines and execution times [3]. Research in this area ranges from using computer vision techniques [4-7] to some more complex techniques and algorithms such as those of computational intelligence [8-10]. Other works monitor construction sites using the BIM (Building Information Modeling) methodology as a baseline and criteria for comparing the achieved progress [11-12], working with static images and videos, and adapting the obtained vanishing lines to the equivalent perspective in the BIM model.

On the other hand, the productivity of the construction sector in Peru has been affected, throughout the years, by the lack of automated and industrialized construction processes, low-quality control, and, consequently, delays in the schedule and cost overruns. However, both public and private institutions are aware of these setbacks and are interested in finding solutions through investments in new

technologies [13-14].

To offer an improvement to the construction sector in the monitoring of construction processes, a supervised machine learning model is proposed as a monitoring tool that could identify, on an ongoing basis, situations that may affect the construction works.

To this end, a methodology is proposed to monitor construction progress development using an object detector trained for objects commonly found in construction sites. A construction project under execution, located in Lima, Peru, is used as a case study. Drones and four high-resolution video cameras (installed at strategic points of the construction site) were used to capture videos and images from different angles and heights. Then, the obtained images were used to train the neural network model for object detection. Finally, this model automatically detects objects that constitute a visual obstruction to the built structure.

2. Methodology

2.1. Construction site and data generation devices

The building under construction is a four-story building with a total area of approximately 11,696 m^2 located at the University of Lima, Lima, Peru. The work began in February 2020. Initially, a drone was used to obtain photographs from various angles around the area, and then, later, four cameras were installed at strategic points around the perimeter of the construction site at heights that vary between 12 and 35 meters, as shown in Figure 1, all pointing towards the construction site. These cameras constantly capture high-resolution videos stored in the University of Lima's servers for later processing.

Figure 1. General plan of the construction site showing the location of the four cameras (left), and an example of the images obtained from each camera (right)

2.2. Proposed methodology to monitor construction progress development

Videos are recorded from cameras located at strategic points around the area under construction (Figure 1), as well as from drones, which are revised by a semi-automated selection process (Section 2.4) to obtain "clean" images. Nevertheless, these images still contain elements that obstruct the visualization of the constructed building, such as machinery (trucks, excavators, and cranes), people, and similar objects.

The next step consists of preselecting or filtering the images using an artificial neural network, whose training allows the artificial intelligence algorithm to recognize the most common elements found in the

construction area. These images are then filtered out (using Algorithm 2 described in section 3) for the next stage of the process, but they still have elements of the environment, such as nearby buildings that have already been constructed; therefore, it is necessary to apply a non-relevant background removal technique. Once a set of filtered images is available, they are sent to the processing stage responsible for identifying the main structural elements, such as tiles, walls, or columns.

Figure 2 shows the proposed methodology to monitor construction progress development. The processes grouped under "Identify Backgrounds" and "Training System" are carried out only once per construction project, while the remaining stages can be repeated according to the desired frequency for the progress evaluation. The activities marked with "M" indicate a predominantly manual activity, while the activities in light blue correspond to those executed in the present investigation.

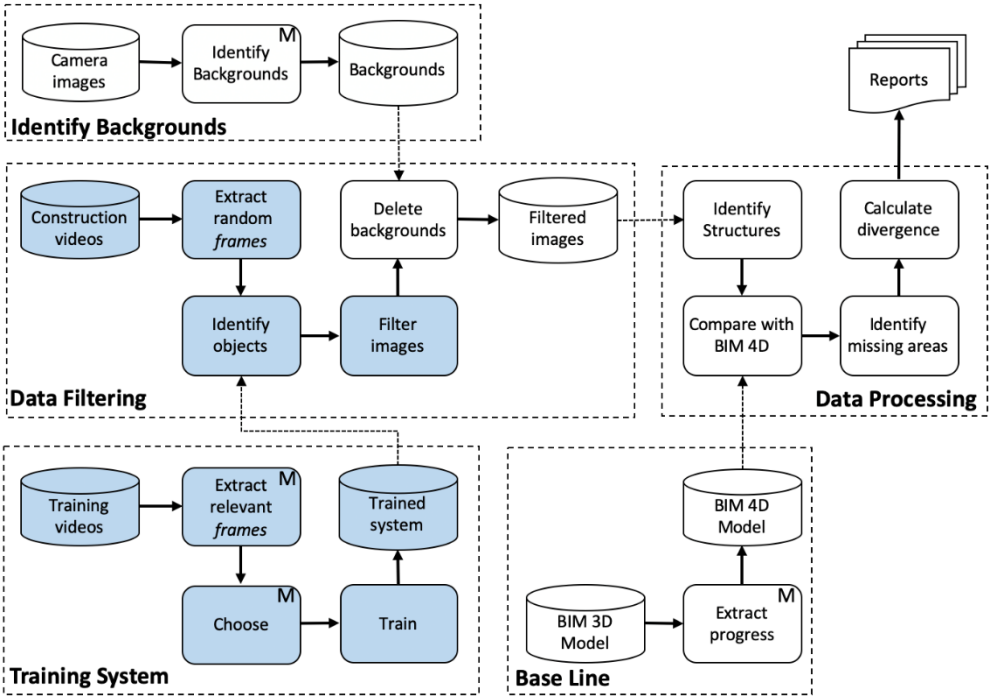

Figure 2. Proposed methodology to monitor construction progress development.

2.3. Training System

This stage is executed once per project before the filtering stage (section 2.4) since its purpose is to train the neural network with the objects, also referred to as the classes, that need to be identified (see the bottom left section in Figure 2). Around 1000 frames were selected from the drone videos, and the target objects in each frame were manually classified and labeled to train the neural network. These 1000 frames were manually selected based on their point of view variability to avoid having several images from the same point in space if selected sequentially. It should be noted that the following eight relevant classes for the construction site were used in this training step: "Dump_truck", "Excavator", "Concrete_mixer", "Skid_steer", "Tower_crane", "Truck_crane", "Truck", and "Person" (Figure 3).

a) Dump Truck b) Excavator c) Concrete mixer d) Truck crane

e) Skid steer f) Tower crane g) Truck h) Person

Figure 3. Examples of the eight classes used for classification.

2.4. Data Filtering

The purpose of filtering is to select "clean" images with the least possible amount of distracting or obstructing elements, such as machinery and people, both of which are very common in construction areas. To achieve this, videos are obtained from fixed cameras located in the construction areas. Frames are obtained from these videos in a pseudo-random manner to ensure an adequate variability among selected frames. The previously trained neural network (section 2.3) is then applied to identify these elements, and both the number of identified objects and their probability of success are obtained. Finally, only the images that contain detected objects whose quantity was less than the maximum allowed (k2), as well as an average detection probability value greater than the confidence level required (k3), are selected (Algorithm 2) to be processed on the next stage. Both k2 and k3 parameters are initially set to 10 and 0.9, respectively, but can be adjusted to an environment's particular characteristics.

3. Implementation and Results

Figure 4 shows the components of the filtering process. This process performs the filtering of images after carrying out the identification of objects in the frames extracted from the videos, as previously introduced in section 2.4.

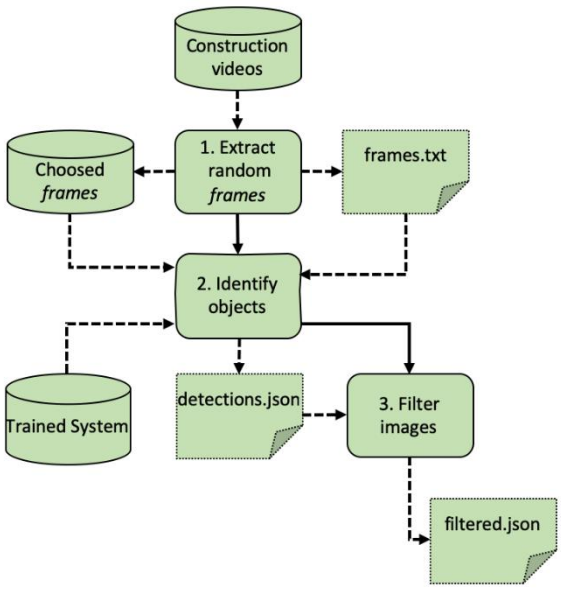

Figure 4. Components of the filtering process.

To ensure a certain degree of variability for the selected frames, a dedicated procedure was created (see Algorithm 1 below), which took as inputs the path of the videos, the type of files to use (mp4 files), the path where the resulting images are to be stored, and an arbitrary interval between frames (200 frames), which can be adjusted based on the frames per second setting used to record the videos. This process's results are stored in a text file (frames.txt) which contains the extracted frames.

Algorithm 1: Frame extraction

Inputs: Video file path, result path, image filter, interval between frames

1. Get the list of files to process
2. For each file:
3. Create a directory with the name of the video file
4. Get the total list of frames
5. Capture frames to process according to a predefined interval
6. For each frame:
7. Generate image in jpeg format
8. Update the results file "frames.txt"
9. Update the image counter

Output: frames.txt file (including selected images/frames)

The YOLO v4 ("You Look Only Once" version four) object detector [15] was used through its implementation in the Darknet library [16]. The default pre-trained weights were used as a starting point to train the neural network further to detect the eight classes mentioned before in section 2.3, using transfer learning. Details of the number of drone images used for training and validation and the obtained results are shown in Table 1, set to approximately 70% of the total number of images for training, and 30% of the total for validation. While the total number of images in datasets 2 to 4 are the same, the distribution of such images between training and validation subsets was done randomly.

Table 1. Neural network training and results using drone videos.

N° of Dataset	Training images (70%)	Validation images (30%)	Total images (100%)	Highest MAP*
1	351	111	462	82.76 %
2	700	300	1000	79.00 %
3	741	259	1000	78.77 %
4	741	259	1000	79.38 %

*MAP = Mean Average Precision.

Using a 50% IoU (Intersection over Union) shows that the first dataset (DS) has the greatest MAP (Mean Average Precision) of 82.76 %.

After this stage, another training step was carried out using images obtained with the fixed cameras (Figure 1). Details of the number of training and validation images and their corresponding results are shown in Table 2. Like the drone images, the 1046 images for each dataset in this scenario were randomly distributed between training and validation subsets.

Table 2. Neural network training and results using the fixed cameras.

DS	Training images (70%)	Validation images (30%)	Total images (100%)	Highest MAP*
1	351	111	462	73.56%
2	733	313	1046	80.30%
3	729	317	1046	84.20%
4	733	313	1046	93.76%

*MAP = Mean Average Precision.

The results using a 50% IoU show that database four has the best MAP. Datasets three and four more than double the number of images of dataset number 1, and they were generated using more randomness.

Thus, dataset four was selected due to its greatest MAP (93.76%). Comparing datasets four in Tables 1 and 2, the results obtained with images from the four fixed cameras (Table 2) exceeds the accuracy of the results obtained with images coming from the drone (Table 1).

The frames selected using Algorithm 1, included in the file frames.txt, were the input to the object detector, which was configured to generate a text file in JSON format containing the identification of the detected class, its location in the image (left, top, height, and width), and the probability of correct identification. To filter individual images, an algorithm was developed (Algorithm 2) that considered the maximum number of objects allowed in the image (k2=10 images) and a confidence threshold (k3=0.9) that configured how high the probability must be that the detected object was, in fact, the real object. If a particular detected object, like a crane, must be considered in further stages of the process, it can be included in a list (k4) to prevent its exclusion.

Algorithm 2: Image filtering

Input: Files (k1), maximum allowed objects (k2), confidence threshold (k3), classes to exclude from filtering (k4)

```
1. Get a list of files to filter (k1)
2. For each image:
3.      Initialize counters
4.      For each detected class:
5.          If the class is not excluded (k4)
6.              Increase item counter
7.              Increase total confidence level
8.      Obtain the average confidence level of the image
9.      If total objects < (k1) and average confidence level > (k3)
10.         Add the image name to a list of filtered items
```

Output: filtered.json file (including filtered images)

4. Conclusions and recommendations

The construction project selected for this research is part of a broader project that seeks to automate the follow-up and monitoring of construction works. As part of the proposed methodology, this research concentrated on evaluating the level of precision of the YOLO v4 algorithm for identifying eight different types of objects commonly present in these environments.

Using the images generated from drones the model precision ranged from 78.77 % to 82.76 %, while the precision using the ones from the cameras ranged from 73.56 % to 93.76 %.

To continue improving the performance of object identification, it is recommended to evaluate the performance of other neural networks models such as EfficientDet [17] or the new version (v5) of YOLO [18] that was released after completing this research, focusing on models that put more emphasis on having higher average precision values rather than real-time speed detection.

References

[1] Yang J, Park M W, Vela P, and Golparvar-Fard M 2015 Construction performance monitoring via still images, time-lapse photos and video streams: Now, tomorrow, and the future *Advanced Engineering Informatics* pp 1-14

[2] Kim C, Son H and Kim, C 2013 Fully automated registration of 3D data to a 3D CAD model for project progress monitoring *Automation in Construction* pp 1-8

[3] Alizadehslehi S and Yitmen I 2018 A Concept for Automated Construction Progress Monitoring: Technologies Adoption for Benchmarking Project Performance Control *Arabian Journal for Science and Engineering* 44 pp 4993-5008

[4] Golparvar-Fard M, Bohn J, Teizer J, Savarese S and Peña-Mora F 2011 Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques *Automation in Construction* pp 1143-1155
[5] Kim C, Son H and Kim C 2013 Automated construction progress measurement using a 4D building information model and 3D data *Automation in Construction* pp 75-82
[6] Martinez P, Al-Hussein M and Ahmad R 2019 A scientrometic analysis and critical review of computer vision applications for construction Automation in Construction pp 1-17
[7] Zhang L, Cao Y, McCabe B and Shahi A 2019 The adoption of Building Information Modelling in Canada
[8] Bosché F, Ahmed M, Turkan Y and Hass R 2015 The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components *Automation in Construction* pp 201-213
[9] Son H, Bosché F and Kim C 2015 As-built data acquisition and its use in production monitoring and automated layout of civil infrastructure: A survey *Advanced Engineering Informatics* pp1-12
[10] Lei L, Zhou Y, Luo H and Love P 2019 A CNN-based 3D patch registration approach for integrating sequential models in support of progress monitoring *Advanced Engineering Informatics* pp 1-11
[11] Han K K and Golparvar-Fard M 2015 Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs *Automation in Construction* pp 44-57
[12] Asadi K, Ramshankar H, Noghabaei M and Han K 2019 Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction *Journal of Computing in Civil Engineering* 33 p 04019031
[13] Salinas J, Prado G 2019 Building information modeling (BIM) to manage design and construction phases of Peruvian public projects *Building & Management* pp 48-59
[14] Palomino J, Hennings J, Echevarría V 2017 *Quipukamayoc* Vol 25 N° 47 pp 95-101
[15] Bochkovskiy A, Wang C Y and Liao H Y M 2020 Yolov4: Optimal speed and accuracy of object detection *arXiv preprint arXiv:2004.10934*
[16] Redmon J 2020 Darknet: Open Source Neural Networks in C https://pjreddie.com/darknet/. Accessed January 15, 2021
[17] Tan M, Pang R and Le Q V 2020 Efficientdet: Scalable and efficient object detection In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp 108781-10790
[18] Jocher G, Nishimura K, Minerva T and Vilariño R 2020 YOLOv5 https://github.com/ultralytics/yolov5. Accessed March 7, 2021