Tracking Productivity in Real-time Using Computer Vision

A Kadoura and E P Small
Department of Civil and Environmental Engineering, University of Delaware, Newark, Delaware, USA

akadoura@udel.edu

Abstract. The construction industry is lagging behind other industries in terms of productivity gains with stagnant growth over several decades. The reasons for the lack of growth are complex and multifaceted, yet all causes and the resulting effects are realized on the output of the workers and machinery on-site. Often, typical site management practices do not identify productivity issues in a timely fashion when corrective actions may impact construction activity progress. Better management tools are required to optimize the productivity of on-site crews. This promises to positively impact the completion of individual tasks, work packages, projects, and ultimately the industry with respect to productivity, which is the driving force behind this research effort. The approach pursued focuses on utilizing commercially available cameras together with computer vision and artificial intelligence. With these tools, the main objective is to develop an algorithm capable of providing real-time feedback on the productivity of workers. Ideally, the algorithm will further identify each type of activity being conducted, thereby capturing useful productivity data. The initial model is a simple classifier that checks whether a worker performs work or stands idle. This is performed through algorithms that identify and track a person's pose and joints and translate them to data points that can be evaluated and translated into helpful productivity measures. Finally, successfully developing a model capable of providing real-time productivity data will allow project managers and planners to better manage and utilize on-site resources. Additionally, since a large amount of data is collected and saved, trends in productivity levels can be tracked and studied further to optimize and improve them.

1. Introduction
The construction industry is regarded as one of the most important economic sectors globally, with an estimated market spending of 10.4 trillion USD. According to McKinsey research, the construction industry accounts for "13% of [global] GDP... [and]... employs 7% of world's working population." [1]. It is expected that with this amount of spending, productivity and total output will improve significantly. This is not the case, as McKinsey estimated that the average growth in construction productivity over the last two decades was 1 percent, compared to a global average of 2.8 percent [1]. This demonstrates that the construction industry lags behind other industries, and steps must be taken to realize the potential for increased productivity growth.

The study will look into the possibilities of leveraging existing tools to track labor productivity in real-time. It is exploratory in nature; thus, it is not meant to be comprehensive or all-encompassing for all construction industries or sorts of construction operations. Computer vision models will be investigated and used on sample imagery/videography rather than a comprehensive, representative set of construction activities. The ultimate, long-term goal is to get productivity data to management.
decision-makers so they can intervene in a timely fashion and make required adjustments to resources and activities. Once proof of concept has been achieved, more studies can be done to expand on this effort.

There are several advantages to exploring the topic of real-time productivity tracking utilizing existing tools. The first benefit is that identifying and tracking productivity in real-time provides quick feedback on the construction operations and crews' performance. If productivity is reduced unexpectedly owing to a site fault, it could be quickly discovered and fixed. On the other hand, current ways to assess productivity rely on limited site observations, making it difficult to promptly address issues if there are delays or productivity reductions.

2. Literature Review

Several factors contributing to the low productivity were identified. Regulatory/contractual issues, design issues, procurement and supply-chain, and on-site execution have been identified as areas where significant differences can be made [1]. The main concern is with the approach taken to improve and correct these productivity problems. Typical improvement approaches rely on forensic analysis, where the productivity is calculated at a later stage of construction and then compared with the planned actions. With this approach, results are obtained too late to implement any corrective actions. Therefore, a significant improvement can be made by reducing the time between observing and calculating the productivity and comparing it to the planned value enabling the opportunity for corrective actions.

2.1. Productivity

The first step required to improve productivity is to be able to define it and measure it. This is not very easy because measuring productivity involves and depends on many factors that vary from project to project. In general, productivity can be quantified as a ratio of input to output costs or daily work output. The exact metrics and units used to calculate productivity typically depend on the type of data being analyzed by the project management team. According to Dozzi and AbouRizk, there is no universal industry standard available to measure or compare the productivity values [2]. To complicate matters further, productivity values are often measured on different scales. The most common levels used in the industry are macro-level and micro-level productivity analysis [2]. Macro-level productivity analysis compares productivity levels across countries or industries, while micro-level productivity analysis is more detailed and can be used to compare the productivity levels between different projects, work packages, or activities. Calvetti et al. introduced a third more detailed level of analysis: nano-level analysis. The nano-level analysis is based on the basic motion and work elements of laborers [3]. Basic motion elements build into work elements which then collectively complete the tasks. This identifies the productivity of each worker and allows for analysis between workers or crews.

2.1.1. Measuring and Collecting Productivity Data. Other industries have made strides toward automating the process of measuring productivity with new sensor technologies. Fortunately, these new technologies are gradually making their way into the construction industry. However, the current widely used methods rely on manual observations of workers by assigning a person to monitor workers on-site and producing a productivity measurement using hand calculations. Dozzi and AbouRizk presented three methods for measuring productivity on the job site: field rating, work sampling, and the five-minute rating method [2]. Each of these methods calculates productivity based on human-based observation and can be burdensome. Therefore, a better approach is to introduce new technologies to automate the process.

The observational limits can be overcome by employing approaches that rely on computer vision and data processing. Calculations can be completed in minutes, and results can be delivered quickly, which is required for using the data to make beneficial changes to site operations. With this methodology, any of Dozzi and AbouRizk's three approaches for calculating worker efficiency in performing tasks can be easily modified to provide an immediate alarm in the event of a work halt or other critical data.
Because the study aims to measure and track productivity in real-time, a significant portion of the literature review will be devoted to examining the various approaches that can be used to establish a measurement. Numerous approaches and equations have been identified to collect productivity-related field data. The majority of the approaches are conceptually similar but appear to differ in the procedure for collecting data or calculating productivity; thus, only a few approaches will be examined to better understand estimating field productivity.

**Factors Impacting Productivity.** While researching the factors that influence productivity in the construction industry, it was discovered that the factors are not constant but rather constantly change, even within the same project. However, these factors can be classified into major categories that have an impact on productivity levels. Although hundreds of factors have been identified in the literature, some of these factors include, but are not limited to, construction management and planning issues, improper or insufficient material and tools, incorrect construction methods, a lack of supervision, and reworks [4].

### 2.2. Construction 4.0

Construction 4.0 is the fourth construction revolution, sparked by Industry 4.0, the fourth industrial revolution. Industry 4.0 was launched in Germany in 2011 to create a fully integrated industry based on newly accessible technologies [5]. Based on the literature, Construction 4.0 can be defined as the process of digitizing and automating the construction industry by utilizing technologies and techniques to enhance performance. This, in turn, improves collaboration and information flow, enhances triple bottom line attributes, preserves safety, and increases productivity.

Construction 4.0 adopted new technologies, which are introduced as pillars. Examples of these pillars include artificial intelligence, big data analysis, the internet of things (IoT), field data-capturing technologies, and so on. The two most appropriate pillars used in this research are artificial intelligence (AI) and field data-capturing technologies (FDCT). These pillars are most related because computer vision is part of AI, and the cameras used to collect the data are part of FDCT. The research will use FDCT to capture and digitize the information from the construction site, while the AI model will predict whether the worker is working or not.

#### 2.2.1. Computer Vision

Computer vision is an interdisciplinary field in computer data science that deals with how computers perceive data from images and videos. Computer vision attempts to imitate how humans use their visual system to perceive and understand their surroundings by using cameras and convolution neural network (CNN) algorithms [6]. The CNN model comprises of four main layers, the input layer, the convolution layer, the pooling layer, and the output layer. Depending on the complexity of the model, more convolution and pooling layers can be supplemented. It should be noted that the performance of CNN or any other AI model is primarily determined by the data used to train the model. Since a construction project is considered unique, it is usually challenging to achieve high accuracy due to the constantly changing nature of the project. Furthermore, there is limited available video of construction operations to use in training the models. Ideally, videos could be captured of actual site operations; however, this proved difficult as construction companies are often reluctant to share data due to confidentiality agreements [7], legal concerns, or site worker disinclination.

#### 2.2.2. Field Data-Capturing Technologies

Field Data-Capturing Technologies (FDCT) are a group of technologies that capture and record information from the field and digitize it to raw data that can be used for further analysis, as the name implies. These technologies are intended to supplement or replace traditional manual methods, increasing efficiency and providing more reliable data. Productivity studies with surveillance cameras enable data collection with fewer direct observers to measure productivity. This method removes the observers from the work site, which has no effect on the crew's work performance. Workers, in effect, do not behave as if they are being observed. Because the process of gathering information still requires a significant amount of manual labor, using FDCT
has significant advantages. Using FDCT to automate data collection and measurement of productivity is thus very appealing for this process.

3. Research Methodology
The goal of this research is to determine whether labor productivity can be assessed in real-time using computer vision, with the results being used to influence positive change in the construction process. This expands on the cutting-edge techniques summarized in the literature review and contributes to the body of knowledge in Construction 4.0 applications and labor productivity optimization.

3.1. Computer Vision Models
Several available computer vision models were tested for the applicability of determining productivity. Some of these models include YOLO and DEEPSort. The main issue with these models is that the model identifies and tracks the worker, but with no further details to the type of work being conducted. A better identified model is TF-Pose-Estimation, which uses a previously developed library (OpenPose) and builds upon it. The primary purpose of OpenPose, and TF-Pose-Estimation, is to provide real-time multi-person key-point detection [8]. It should be noted that there are several other models that are also capable of identifying joints and body parts with no significant difference between them and this model.

The benefits provided in TF-Pose-Estimation and similar models compared to other models is the ability to identify joints and body parts. Since the body movements and joint locations can be identified, the model can be capable of determining whether a worker is working or standing idle. Therefore, TF-Pose-Estimation was chosen to move forward with the research. The algorithms rely on supervised training, and a pretrained dataset was made available for use and application in this study. The model will not need to be re-trained because the developers made available a pre-trained algorithm, saving time and effort in the research process. The models' training recognizes individuals but does not distinguish between working and non-working individuals. This is accomplished through the use of a classifier variable that has been trained to indicate work positions and differentiate whether the labor is productive or idle based on the personnel's joint positions and poses. If necessary, the pre-trained classifier can be retrained using a dataset of construction workers in the future.

The model must be capable of translating and storing the FDCT input (i.e. imagery and videography) into useful data that can be analyzed for additional insightful and useful data. Ideally, the approach will be tested on-site with photography and videography. This study will provide proof of concept, and subsequent studies will use the technique in "live" on-site conditions.

3.2. Data Processing
Without going into specifics, the model predicts the coordinates of body parts and joints using pre-trained models. Then, using OpenCV, these body parts are redrawn on the original photo. This does not predict the worker's productivity but rather provides visual feedback to ensure that the predicted body parts correspond to the person and that the prediction is correct. Finally, the produced stance is compared to a pre-trained collection of poses in order to determine if the person is working or idling. Retraining MobileNet's last layer accomplished this on a dataset of working and idle workers. Figure 1 shows how the model generates outcomes. These are illustrated using royalty-free photos for demonstration reasons. These figures include subjects that are in the photograph's foreground and take up a large portion of the frame, resulting in better results. It is worth noting that in these illustrations, not all of the body parts need to be seen in order to make a prediction, which is what is likely to happen in real-life settings.
3.3. Experiment Design – Development and Testing of the Model

Model testing in this effort requires video footage. The video is less clear and has a lower resolution than the photographs shown in Figure 1. This is expected to affect prediction accuracy. Furthermore, in their positions, specific productive job tasks resemble standing/idling activities. Perfection is neither required nor expected. The goal is to see if the strategy can help and if some of the accuracy constraints can be addressed when deciding whether to notify decision-makers about non-productive periods.

The model's results must also be statistically significant. This is accomplished by creating an experiment and testing the results within a certain level of confidence. The following methodology was used for this study: Video footage of workers engaged in various activities will be used. A table with the manual observation and model categorization as columns and a 15 to 20 second time interval for the rows, indicated as observation number will be generated. A confusion matrix will be generated for each activity and subsequently for the total observations when the model has been run, and the actual vs. anticipated data have been collected. The confusion matrix is useful because it calculates Type I and Type II errors, which refer to false positives and negatives, respectively. In addition, the model's performance and other relevant features can be estimated. The accuracy, precision, sensitivity, and specificity of the model are typically examined when determining the performance of a classifier. These measures show whether the model is appropriately classifying observations and, if not, what type of error is there.

3.4. Preliminary Model Results and Future Refinements

As discussed, computer vision models have been discussed and evaluated for the purposes of this research effort. The model has been evaluated with a small set of videos culled from instructional videos of construction techniques. These were chosen as they were readily available for use and were royalty-free sources. The videos were evaluated manually to identify observations and compared to the predictive models. Results are shown in Table 1.

| Prediction | Observation |
|------------|------------|
| Positive   | Negative   |
| 106        | 78         |
| 3          | 55         |

Table 1: Correlation of Predicted Results to Observations of Work versus Non-Work States

Evaluation of the table indicates a small proportion of Type 1 errors with 1.2% false positives. False negatives, reflective of type 2 errors, were quite high at 32%. There are several reasons for false negatives, including similarities of body poses between work and non-work states and a lack of a sufficient population of different work activities for training. This shows the necessity of additional research in this area before reliable results can be implemented on-site.
4. Conclusion

In conclusion, the poor productivity growth recorded by the construction industry over the last several decades was the driving force behind our research. The construction sector is in the midst of a productivity crisis, and this study aims to leverage Construction 4.0 technology to build tools to help enhance labor productivity on job sites. The goal of the study was to combine AI and computer vision models with commonly used field data gathering devices, such as cameras and other data collection methods. The goal was to create a real-time system that would monitor ongoing activities and notify project management people when workers became idle, allowing for possible productivity consequences to be mitigated and work to continue.

The study began with a literature review to better understand current productivity issues and available construction 4.0 technologies for productivity enhancement. Completed research on AI, computer vision, and field data collecting approaches was also highlighted in the literature study. The literature review revealed that there is a wide range of metrics used for assessing productivity. There is no single, unambiguous definition of productivity. Instead, different definitions and equations exist depending on the level of analysis and the purpose of the analysis. In general, productivity is defined as the ratio of input to output. This is measured differently depending on the perspective being evaluated. Productivity evaluation at the macro-level differs from that at the micro- or even nano-level. After further investigation, it was determined that a nano-level analysis is required for the purposes of this research, in which individual productivity is measured rather than evaluating productivity for large components or project phases. Alternative methods of assessing an individual's productivity were investigated. The factors influencing productivity were investigated and documented. The following section of the literature review focused on Construction 4.0 and determining which technologies are required for this research.

The research methodology was developed after gaining a better understanding of the problem through research and literature review. Several neural network-based AI models, such as YOLO, DeepSearch, and TF-Pose-Estimation, were investigated. Each of these was tested, and TF-Pose-Estimation was found to be the most applicable model for the research objectives. The model provided a representation of a worker's pose and joint positions, which is helpful in developing a real-time prediction of whether the person is working or idling. Furthermore, the model provides pose (skeletal) coordinates that can be used in more in-depth analysis. As a result, while the model provides some raw data, the amount of data captured can be valuable if used for the proper analysis.

Preliminary results of the model show that Type 1 errors are minimal, but there are significant numbers of false negatives. This is problematic and may result in alerts being issued when there are no issues with productivity. Further research is required, and there are numerous additional opportunities to expand the research effort. This includes additional refinement and testing, as well as the use of different data sources and site-specific implementations. Continuing efforts show significant promise in developing tools capable of having a significant impact on the industry and bringing about a small amount of positive change. The models can be improved to provide more specific information about the type of activity being carried out. Furthermore, because there is little to no data on nano-level productivity analysis, this method can be used to create a database and baseline productivity for various activities and variables.

Ideally, the research will provide input to a real-time alert system that can distribute information to management staff where extended times of non-productive work periods are observed. This permits active intervention and correction of the causes of the non-productive states at times where the intervention will be effective. It is the hope and goal that such a system, if implemented on a small/nanoscale in the field, would ultimately have a broader impact on increasing the productivity on-site and the profitability of the individual projects.
References

[1] McKinsey&Company 2017 *Reinventing construction: A Route To Higher Productivity*. https://www.mckinsey.com/~/media/McKinsey/Industries/Capital%20Projects%20and%20Infrastructure/Our%20Insights/Reinventing%20construction%20through%20productivity%20revolution/MGI-Reinventing-Construction-Executive-summary.pdf

[2] Dozzi S P and AbouRizk S M 1993 *Productivity in Construction*. Ottawa, Ontario, Canada: National Research Council Canada.

[3] Calvetti D, Mêda P, Gonçalves M C and Sousa H 2020 Worker 4.0: The Future of Sensored Construction Sites. *Buildings*, 10(169), 1-22.

[4] Ghoddousi P and Hosseini M R 2012 A survey of the factors affecting the productivity of construction projects in Iran. *Technological and Economic Development of Economy*, 18(1), 99-116.

[5] Hofmann E and Rüsch M 2017 Industry 4.0 and the current status as well as future prospects on logistics. *Computers in Industry*, 89, 23-34.

[6] Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., & Wang, X. (2020, October 19). Computer Vision Techniques in Construction: A Critical Review. *Archives of Computational Methods in Engineering*, 1-15.

[7] Darko A, Chan A P, Adabre M A, Edwards D J, Hosseini M R and Ameyaw E E 2020 Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, 112, 1-19.

[8] Cao Z, Hidalgo G, Simon T, Wei S E and Sheikh Y 2018 OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *Computer Research Repository (CoRR)*, 80(8), 1-14.