DL-Based Solution Proposal for Monitoring Concreting Operations

Presented by Group 8
Oct. 7th, 2020

Executive Summary

The Status Quo

- The construction industry suffers a **slow productivity growth** due to the lag in infusing new technology (e.g. IoT, computer vision, etc.)
- **Field labor cost** accounts for a large proportion of the total construction costs, therefore requiring special attention
- **Chrono-analysis** of the on-site operations may be a key breakthrough in improving labor productivity
- **Approaches based on Deep Learning and Computer Vision** have been developed to extract the information from monitory images and/or videos

Improvement Points

- Improve construction productivity and reduce cost by:
  - Smoothing the transition between different tasks
  - Keeping balanced combination of labor and machines to achieve the optimal marginal productivity

  with the help of new technologies (object detection, automation, etc.)

Executable Solutions

- Enhance the monitoring quality by:
  - Easing the detection and related computations
  - Ensuring the continuity of the monitoring images/videos
- Optimize the operations by:
  - Leveraging on productivity KPIs to reorganize the tasks timetable
  - Assigning the most appropriate number of staff to tasks

AGENDA

1. Issue Analysis
2. Methodology
    a. DL-Based Object Detection
    b. The Analysis Model
3. The Experimentation Process
4. Results & Solutions
5. Discussion

AGENDA

1. Issue Analysis
2. Methodology
    a. DL-Based Object Detection
    b. The Analysis Model
3. The Experimentation Process
4. Results & Solutions
5. Discussion

Improvement of productivity in construction is on the way, but inertia is strong

Construction dilemma:
a strong presence in the world economy yet with weak productivity growth

Construction-related spending accounts for

13% of the world's GDP

...but the sector's annual productivity growth has only increased

1% over the past 20 years

$1.6 trillion
of additional value added could be created through enhanced productivity

Applicable measures:
improving onsite execution and infusing technology

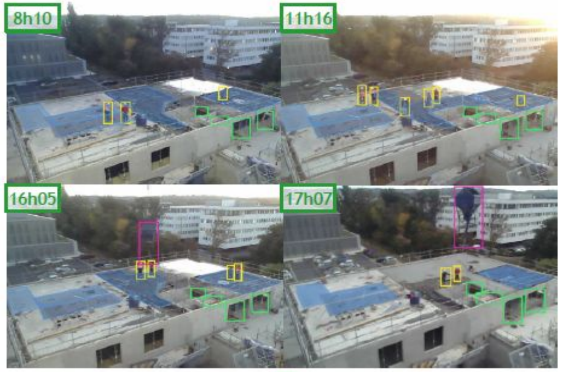
- **rigorous planning process** to ensure that key activities are achieved on time and on budget
- application of **lean construction** principles to reduce waste and variability
- adoption of **production control** and **quality inspection** processes

Main challenges:
production and quality control

- difficulties in obtaining **real-time and accurate data** from a construction site
- struggle to keep workers engaged in the **right tasks, at the right time**
- workers can rush into taking on **unplanned tasks**, which often results in **incomplete work** that has to be finalized later in haste

Source: McKinsey (*Reinventing Construction*, 2017), Aalto University

Chrono-analysis is an efficient monitoring tool yet not generalized in construction



Date	Time	Task	[position]	ID	Nb of people
14/10/2019	8:10:53	Rebars		3	4
			...		
14/10/2019	16:05:23	Concreting		12	2

An efficient monitoring tool ...

- based on the fact that **process inefficiencies** become visible when images are used to track progress

... to boost productivity ...

- shows significant results, **increasing up to 25% in productivity**

... however not yet generalized

- requires the mobilization of specialized engineers
- is time-consuming

need to automate visual construction site monitoring

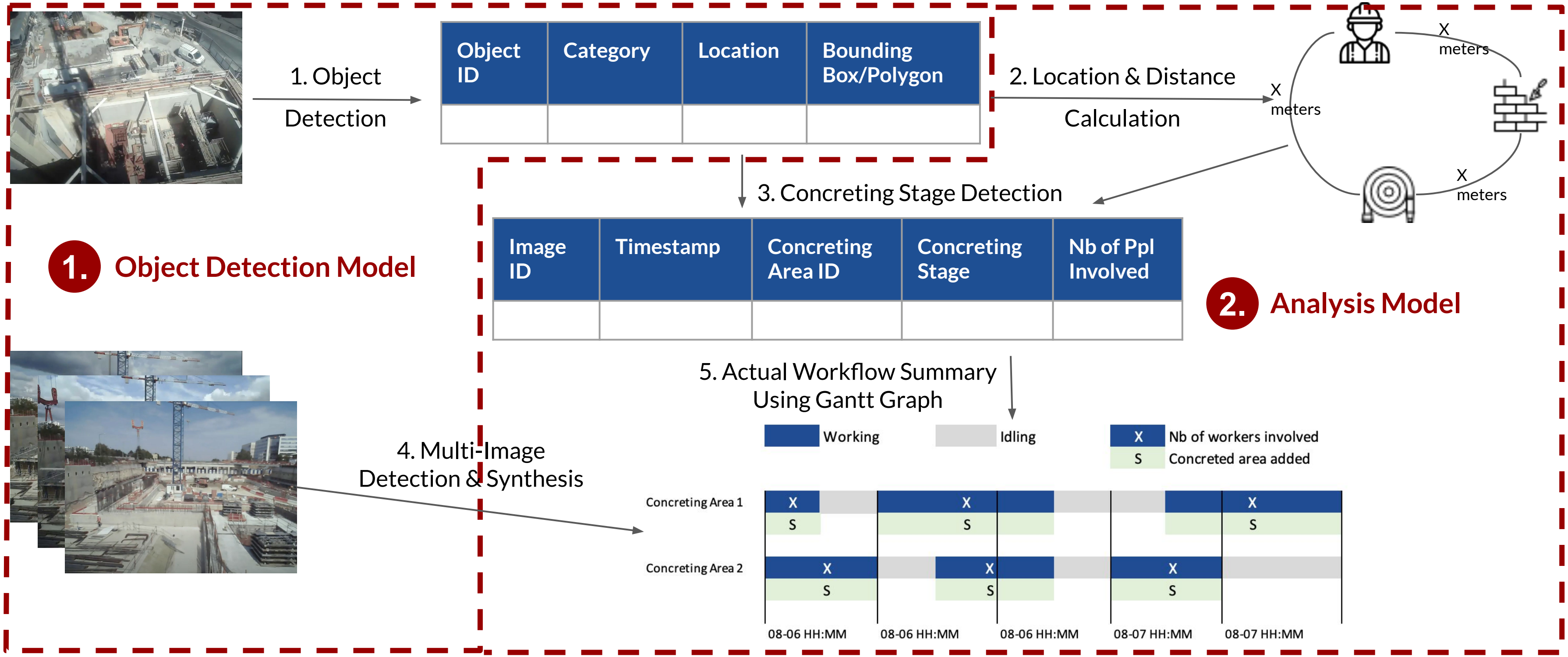
Source: Aalto University

AGENDA

1. Issue Analysis
2. **Methodology**
    a. DL-Based Object Detection
    b. The Analysis Model
3. The Experimentation Process
4. Results & Solutions
5. Discussion

An analysis framework with 2 parts and 5 steps is used to tackle the issue

Our analysis framework consists of **two parts**: an **object detection model** where computer vision methods are applied to detecting the key objects and an **analysis model** to analyze the operation efficiency.



Mask R-CNN with pretrained weights serves as our object detection model

Detection Model: Mask R-CNN

- **High Performance**:
  The state-of-art R-CNN architecture with impressive performance on the COCO dataset.

- **Versatile Output**:
  Capable of producing both bounding boxes and masks for the detected objects.

- **Easy Deployment**:
  Readily open-source and easy to use with full OOP programming style.

- **Efficient Computing Backend:**
  Keras as the backend with optimization for CPU computation, bring high computing efficiency.

Dataset

- Images collected from 6 construction sites with manual annotations of workers, vertical formworks and concrete pumps.

- Train/Valid sets splitted for each construction site to enforce a consistent sampling across different construction sites.

Transfer Learning

- Started from the model weights pre-trained on COCO* dataset.

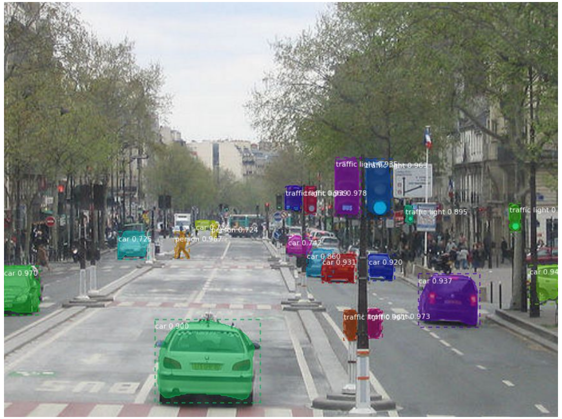
- An open-source external TinyBenchmark** dataset was included to ameliorate the detection rate of workers. (final training data = client's dataset + TinyBenchmark)

*: https://cocodataset.org/#home
**: https://github.com/ucas-vg/TinyBenchmark

Transfer learning and external data facilitate training our deep neural network

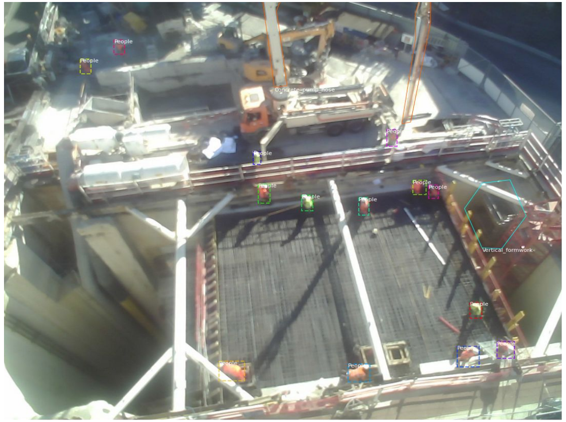
We used pre-trained model weights to facilitate model training and external dataset to feed more useful features to the model.

COCO



- Widely-used object detection benchmark dataset
- **Content**: (>200K) cityscapes
- **Usage**: provided pre-trained weights to start with for **faster convergence**.

Client's Dataset

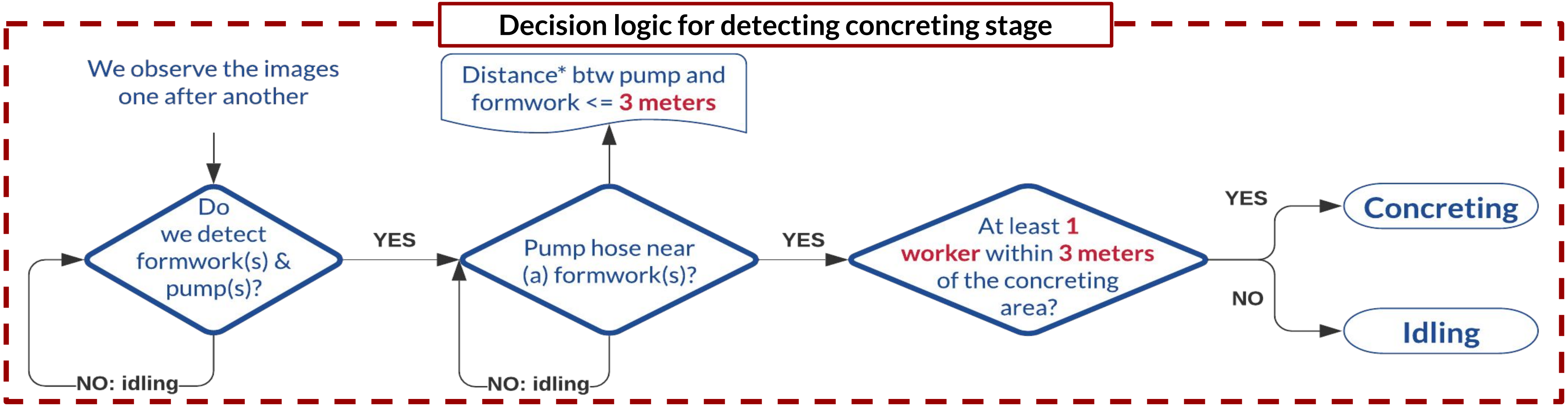


- Images collected by client
- **Content**: (~2K) construction site
- **Usage**: served as part of the training set

TinyBenchmark

- Benchmark dataset for models detecting **small people**
- **Content**: (~1K) **crowded people**
- **Usage**: a subset served as part of the training set

The detected objects are translated into concreting stages using a decision logic



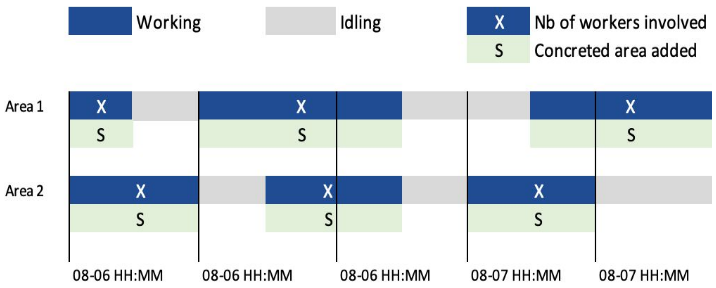
Object detection →

Object
Class (pump, worker, formwork)
Location
Bounding box / polygon

Decision logic →

Concreting Area
Concreting Area ID
Timestamp
Stage & area
Nb of workers involved

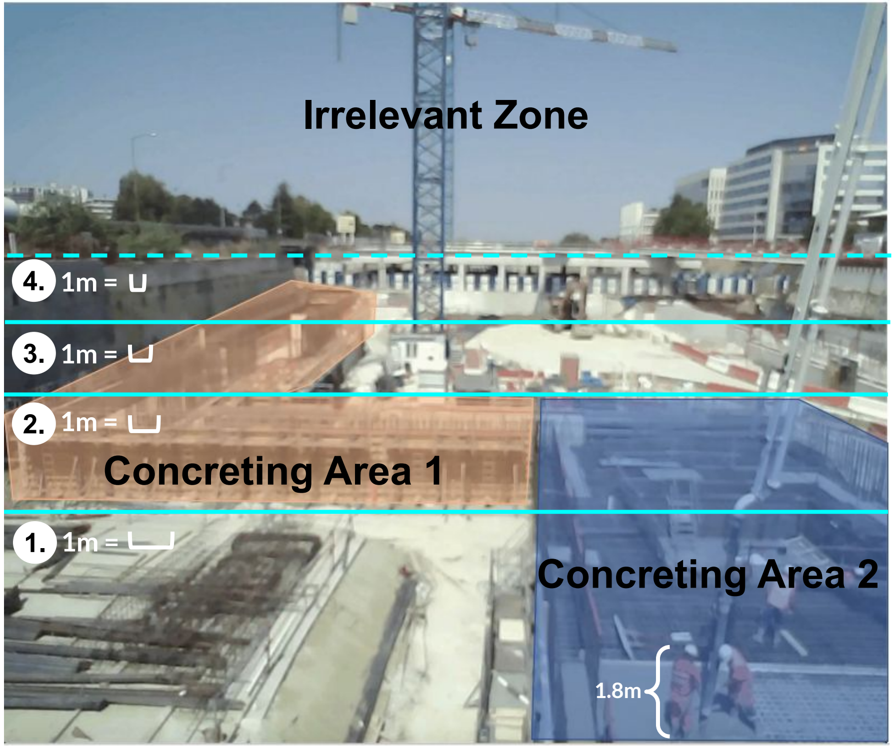
Synthesis of images →



*: The distances between objects were calculated using PyImageSearch (https://www.pyimagesearch.com/2016/04/04/measuring-distance-between-objects-in-an-image-with-opencv/); for the threshold values, please refer to the next slide

Several threshold values were chosen to complement the decision logic

We've defined the maximal distance between pump hose/worker and the concreting area for identifying the concreting stage. In consideration of the perspective, we also designed a table for scaling the sizes and distances depending on objects' locations.



Definitive criteria for concreting stage	Threshold
Maximal distance between pump hose(s) and the concreting area(s)	3 meters
Maximal distance between worker(s) and the concreting area(s)	3 meters

Perspective discounted scale*		
Zone	Discounting factor	Plotting scale
1	1	1 meter = 60 pixels
2	0.67	1 meter = 40 pixels
3	0.5	1 meter = 30 pixels
4	0.33	1 meter = 20 pixels

*: Perspective theory told us that an object's dimensions along the line of sight appear shorter than its dimensions across the line of sight, and that objects appear smaller as their distance from the observer increases (Wikipedia). Therefore, we have made a scaling table to scale the sizes and distances of/between the objects depending on their locations.

AGENDA

1. Issue Analysis
2. Methodology
   a. DL-Based Object Detection
   b. The Analysis Model
3. **The Experimentation Process**
4. Results & Solutions
5. Discussion

Mask R-CNN produced promising results whereas YOLOv3 didn't

Our experimentation with the Mask R-CNN model mainly focused on adjusting the learning rate and wrangling with the data. Both training loss and validation loss decreased as we continued to train the model on newly-added data. However, the YOLO model failed to produce satisfying results due to low computing efficiency.

Day 0 – **Mask R-CNN deployment and testing**

Day 1 – **Dataset download and visualization**

Day 2 – **Pipeline setup for Mask R-CNN and YOLO**

Day 3 – **Training started**

Day 5 – **Additional datasets added**

Day 7 – **Stopped YOLO training**

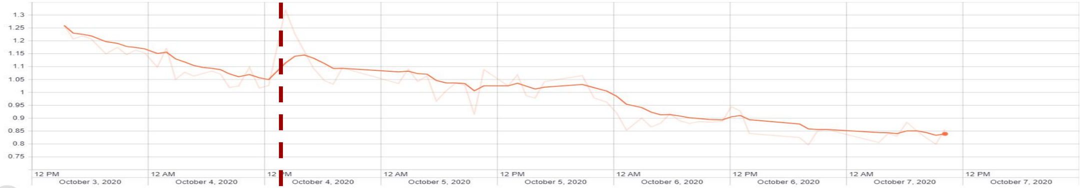
...

Day 9 – **Model to be evaluated**

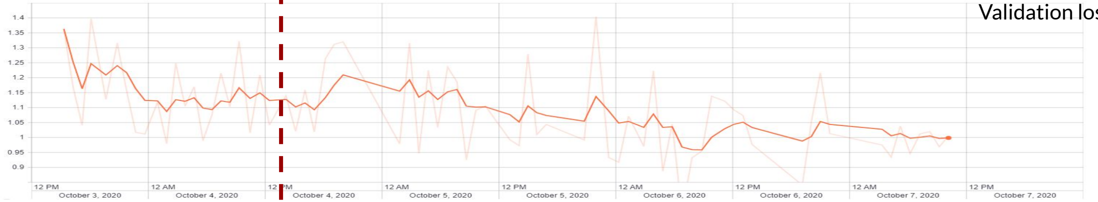
Dataset preparing (Day 1 – Day 3) | **Model training** (Day 3 – Day 9)

- ~ 2000 images in the training set.
- Mask R-CNN is **not overfitted** yet so far.
- Due to the **limitation of available memory**, the batch size is 1 for Mask R-CNN model.
- **~6 hrs** to go through one epoch
- We abandoned YOLO model due to unacceptable **low efficiency** in training with CPU.

Train loss



Validation loss



AGENDA

1. Issue Analysis
2. Methodology
    a. DL-Based Object Detection
    b. The Analysis Model
3. The Experimentation Process
4. **Results & Solutions**
5. Discussion

Detection quality: good at large objects but not so much for small ones

By comparing the ground truth annotations and our predictions through sample visualization, we can see that our model is generally good at detecting larger objects (mixer trucks & vertical formworks) but not so much for people which are small.

Ground truth annotations

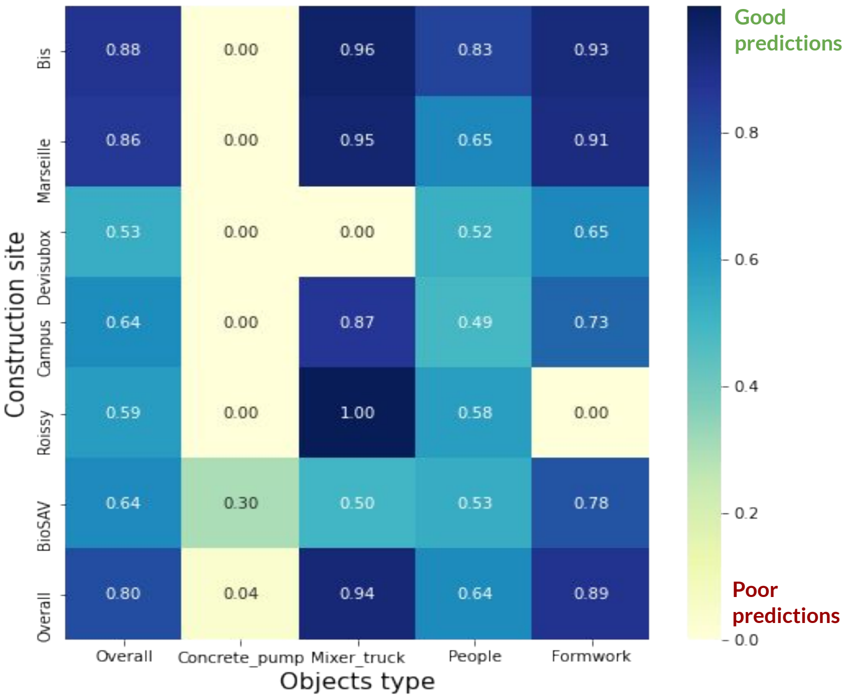
Predictions made by the model

\+ Good predictions: Mixer_truck, vertical formwork

\- Poor predictions: People

Detection quality*: satisfying but limited by insufficient data and poor image quality

Evaluated on the validation set, the key performance indicators of our model are:
mAP@0.5IoU = 0.80 and **mIoU@0.9 =** 0.64



mAP@IoU0.5 for each class and construction site

- Breakdown by construction site
  - Good prediction on
    - Bis, Marseille
  - Poor prediction on
    - Devisubox, Campus, Roissy, BioSAV
- Breakdown by object type
  - Good prediction on
    - Mixer truck, Formwork
  - Poor prediction on
    - People, Concrete pump

- Poor predictions due to
  - **Insufficient data**
    - ~1000 images in Bis/Marseille
      While only ~200 images in others
    - Only ~200 annotated concrete pumps
  - **Poor image quality**
    - People with size of only a few pixels

*: Model evaluated on valid set (20% of the dataset provided by the client)

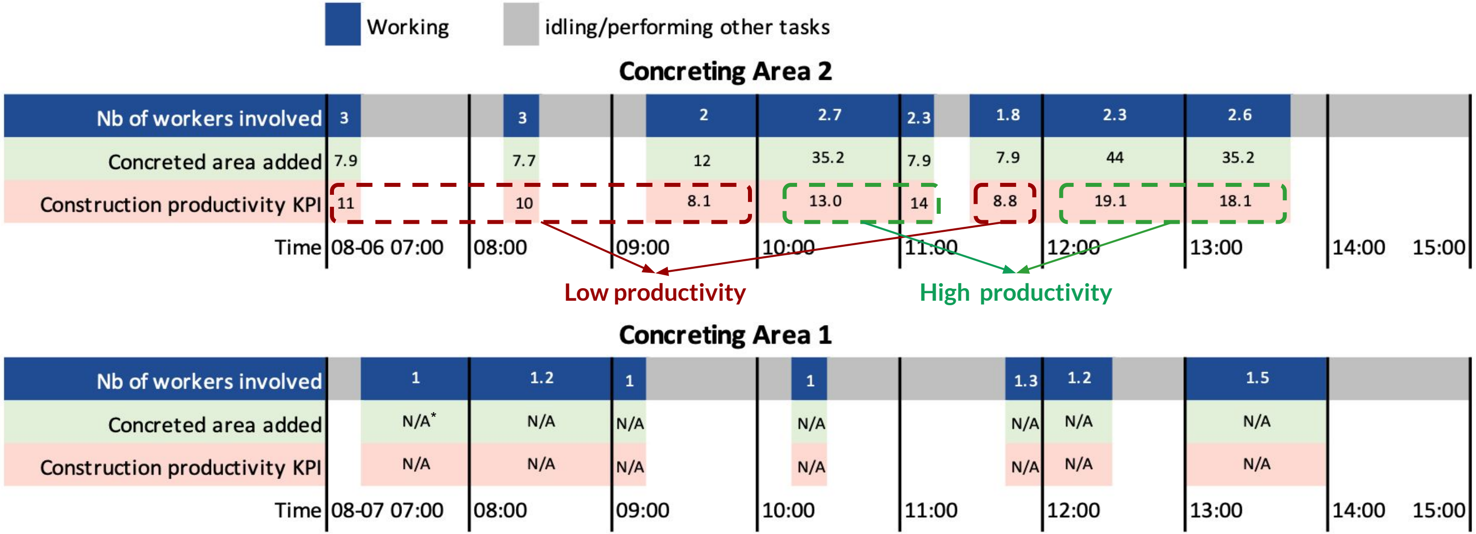
There exists large variability in the productivity across the day

Based on the concreting productivity KPI we've defined, the labor productivity has been quite **unstable** throughout the day with workers being less productive in the early morning compared to late morning and noon.

- Concreting productivity KPI:

$$\textit{concreting productivity KPI} = \frac{\textit{Concreted area added}}{\textit{avg.\# of workers involved} \times \textit{\# of working hours}}$$

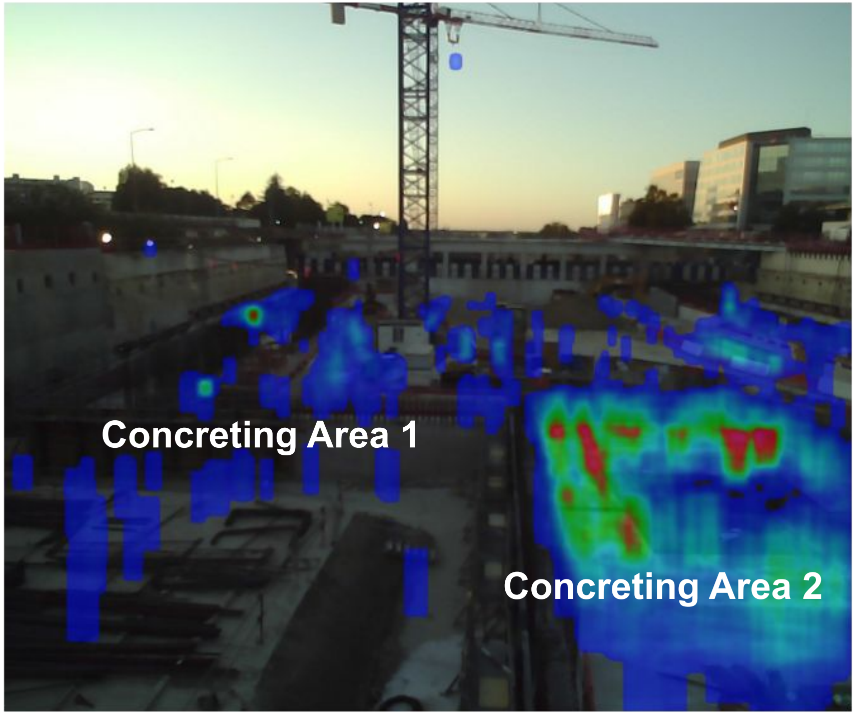
- Current status:



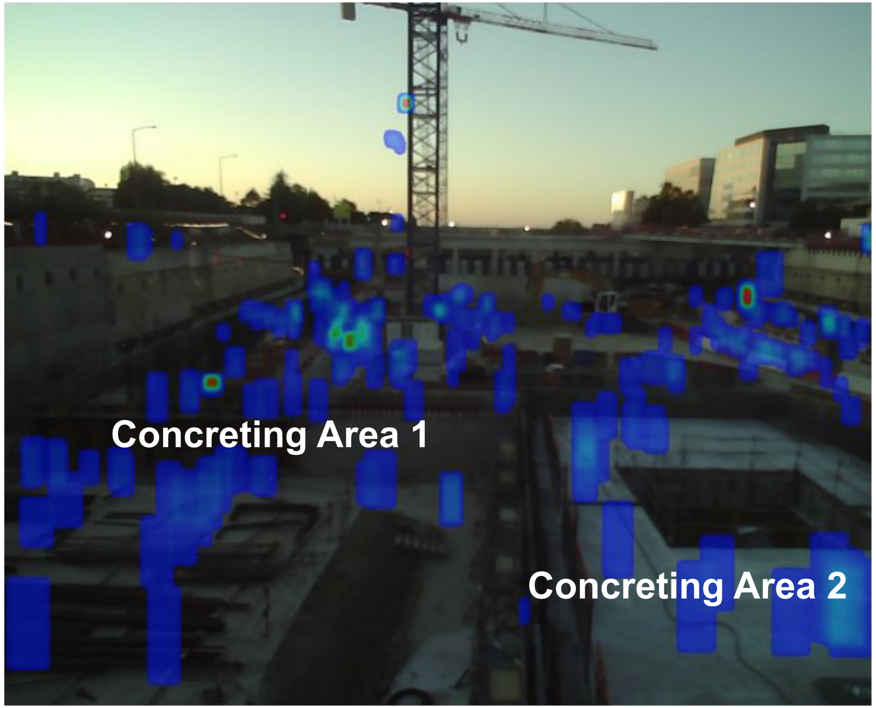
*: we didn't calculate the values for Concreting Area 1 due to the flaws in our detections (caused by the incomplete annotations).

Heatmap of workers' presence verified the credibility of our productivity analysis

The workers have been concentrated in different concreting areas in the two given dates as shown in the heatmaps, which is in accordance with our detection results.



2020-08-06



2020-08-07

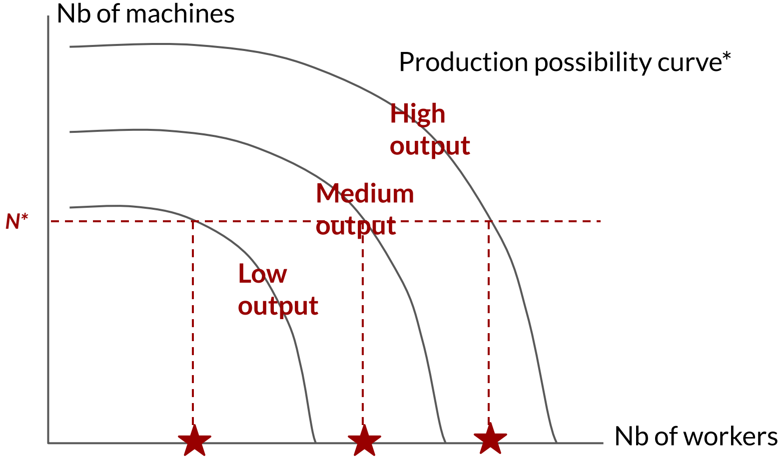
It is advised to enhance the monitoring quality and optimize the operations

Monitoring Quality Enhancement

- Set up high-resolution monitoring cameras for each concreting area from an overlooking angle (ease the detection and related computations)
- Set up a reliable internet network to ensure the continuity of the monitoring images/videos

Operation Optimization

- Rearrange the task timetable based on their productivity KPI across time (concreting tasks are better to be performed in the noon/afternoon)
- Assign an appropriate number of workers to the tasks based on the optimal ratio of labor to machines



*: A curve which shows various combinations of the amounts of two goods which produces the same amount of output. The red stars mark the optimal number of workers given the available number of machines.

AGENDA

1. Issue Analysis
2. Methodology
    a. DL-Based Object Detection
    b. The Analysis Model
3. The Experimentation Process
4. Results & Solutions
5. **Discussion**

Our analysis quality is limited by insufficient computing resources, data and time.

Object Detection Model

- No GPU available on HEC data factory.
    - Didn't use 3rd-party free GPU resources due to data security and confidentiality.
    - Tight project time didn't allow us to make mistakes or try different approaches
- Pump_hoses are difficult to detect due to insufficient ground truth annotations in the training set.
- Workers that takes only a few pixels are difficult to detect.

Analysis Model

- Analytic_train_set is difficult to train due to less satisfying annotations
    - Incorrect annotations (e.g. pump_hose, people).
    - Irregular bounding boxes.
- Analysis based on our hypothesis due to lack of field experience (details of concreting process).

Our proposed steps to improve the results considering return on investment...

Object Detection Model (ranked over feasibility)

- Access more computing resource (e.g. GPU, extra memory, etc.) for more efficient and deeper training.
- Give a try to other deep learning architectures of interest: SSD, CenterNet, etc.
- Look for more data to train our model:
    - Better balance between classes (e.g. more images with pump_hose).
    - Better quality of images (e.g. higher resolution cameras).

Analysis Model

- Generate predictions with our object detection model, train our analysis model for better accuracy.

THANK YOU!