TOWARDS AN AUTOMATIZED AND OBJECTIVE ASSESSMENT OF DATA FROM VISUAL INSPECTIONS OF BUILDING ENVELOPES

JAN MANDINEC*, PÄR JOHANSSON

Chalmers University of Technology, Department of Architecture and Civil engineering, Sven Hultins gata 6, 412 58 Gothenburg, Sweden

* corresponding author: jan.mandinec@chalmers.se

ABSTRACT. The renovation planning process is filled with uncertainties and subjective decisions. These make the decisions upon what and when to renovate a complex and ambiguous problem. Selection of renovation measures related to building envelope are often far from optimal as decisions are usually made based on visual inspections. These are manned and thus prone to subjective assessment and the knowhow of individual inspectors. Furthermore, objective criteria which could indicate non-structural failures are often missing. The objective based planning process allowing the estimation of the current damage status of the building envelope by only using non-destructive measurements is still in its infancy. The first step requires establishing reliable and objective based data collection. These could be efficiently collected by Unmanned Aerial Vehicles (UAV) with subsequent image recognition algorithms allowing the identification of imperfections and store the position and extent of such deviations into the building's digital assessment database. Such tools do not exist. The aim of this study is to investigate the current objectivization possibilities in the domain of building inspections. The first part provides a literature review describing how an autonomous UAV survey of a building envelope may be planned and what computer vision techniques may be used for automatic damage recognition and classification. Subsequently, an objective detection model based on the YOLO-tiny (You Only Look Once) computer vision framework is employed in a case study investigating a building envelope of historical Tjolöholm castle in Sweden. This study contributes to developing a methodology for an objective based visual inspection process.

KEYWORDS: UAV, computer vision, object detection, YOLO, case study, building, renovation, brick, masonry.

1. INTRODUCTION

In terms of environmental aspects, renovation of a building or its adaptive reuse has proven to outperform constructing a new building [1]. The environmental effectiveness of renovations may be further improved by transforming the maintenance paradigm from reactive (emergency renovation after a component breaks down) to proactive (planned renovation before a component breaks down). However, this paradigm shift requires a full digitalization of the maintenance planning allowing to perform data driven prediction of failures (predictive maintenance). Inspiration could be taken from the world of wind turbines, or jet fighters where data can already predict whether any component will fail within a given period allowing not only to reduce down-times but also optimize the usage of resources and thus the environment. Ideal maintenance system in buildings should be capable of doing the same type of assessment to recommend measures which can prevent a failure. However, reaching this level of predictive maintenance requires systematic changes in data collection processes. The performance of machine learning algorithms allowing a computer to find patterns in the data and thus automatically learn and decide over complex systems, are highly dependent on the type and the total amount of the data.

Building inspection provide invaluable information for future failure prediction as it shows how much damaged building is at the time of inspection. However, common manned inspections are insufficient in terms of large data collection. Moreover, data are often not stored in a format suitable for applying machine learning algorithms. Note that insufficient in this context means that building inspectors, however experienced, cannot fully quantify every aspect of how a building envelope is damaged (e.g., count the number of all cracks, measure, and sum up all damaged areas). Contrastingly, Unmanned Aerial Vehicles (UAVs or drones) have such possibilities as drones can gather large sets of data by scanning whole building envelope with various sensors (i.e., RGB or infrared cameras, lidar, radar). Given the premise from above, one need to objectively obtain the current status of the building envelope. The most readily available sensor technique is represented by RGB (red, green and blue) cameras.

The overarching research questions this study answers are:

(1.) Which data can objectively be used to assess what the current and future performance level of a building envelope are?

(2.) How can a damage on a building envelope be quantified based on images from UAVs?

(3.) How can the whole process become as automatized as possible?

This paper is divided into a literature review and the case study. The literature review is further divided into two distinct parts. Based on the conclusions from the literature review, the case study is used to optimize the methodology into the practice. This is done by applying the in-house Build Sense model, developed based on the YOLO-tiny (You Only Look Once, tiny version) object detection framework [2], on a wall of Tjolöholms castle in the west of Sweden. Finally, the study concludes with the discussion about how damaged areas may be quantified giving the outlook for future data-driven UAV based building inspections.

2. Literature review

The first part of the literature review investigates how a flight pattern of UAVs may be automatically generated for building envelope inspection. The second part presents automatic damage recognition from images to investigate how damages on a building envelope may be automatically recognized by the use of computer vision algorithms.

2.1. Planning of UAV based inspections of building envelope

Usage of Unmanned Aerial Vehicles (UAVs or drones) in a building inspection process yields many advantages compare to the classical manned inspections. The two of the most distinct advantages are the time efficiency and the reachability over the whole building envelope. The subsequent analysis of the images captured by drones may provide a comprehensive look on the status of the building envelope.

The building inspection process based on a UAV survey may be either conducted manually i.e., the drone is controlled remotely by an operator on the ground, or it can be conducted semi-automatically. In both cases local laws and restrictions regarding operations of the UAVs need to be followed. Duque [3] identified 4 necessary steps which need to be taken prior to the inspection of any construction:

(1.) UAV fly operation training,

(2.) documentation review of inspected structure,

(3.) observation of surroundings and

(4.) UAV pre-flight check (e.g., battery levels).

In the latter semi-automatic case, drone is often pre-programed to follow specific GPS coordination waypoints and flight path making the flight itself fully automatic. Such an approach was adopted e.g. in a study by Rakha [4] who used images captured by infrared camera to develop heat-leakage segmentation framework based on computer vision algorithms. However, the loss of the GPS signal may lead to navigational hazards. To cope with this problem, a SLAM (Simultaneous Localization And Mapping) technique, was developed, allowing drones to detect objects in the surroundings by employing lidar or similar sensor and thus to avoid collisions.

Current development in the semi-autonomous flight systems adopts 3D models (or BIM models) of an inspected building as a source of prior information for UAV inspection planning. Freimuth [5] utilizes BIM models to automatically generate safe flight paths for UAVs. Similar framework developed for automatized inspections of bridges, allowing to generate safe flight paths can be found in the work by Morgenthal [6]. This framework was practically utilized for inspecting building envelopes by Benz [7] who performed a case study to automatically generate flight paths for obtaining infrared images and subsequent thermal conductance value determination of the building envelope. Note that flight path in this case needed to be manually adjusted to avoid collision with nearby obstacles (e.g. trees).

Regardless the type of flight control, UAVs need to employ specific routing strategies to comply with the needs of different sensors and the intentions of the quest. For this purpose, a different flight pattern is needed for drone scanning to produce high resolution images suitable for subsequent defects analysis compared to a drone taking infrared images suitable for thermography analysis.

2.2. Automatize damage recognition and classification from images

Recent development of computational power made the machine learning and computer vision algorithms accessible to the point where increasing numbers of research papers dealing with their application on building inspections are being published every year. The foundation for the majority of algorithm for image interpretations is a Convolutional Neural Network (CNN) as it mainly specializes on extracting basic features from images like straight lines, round shapes etc. Extracted features are then subsequently fed to a classification algorithm providing the predictions. The aforementioned method was successfully used for crack detection where Support Vector Machine (SVM) algorithm achieved 85.94% of the validation accuracy [8]. Note that substantial number of labelled images i.e., images where the labels are specified manually, are needed for reaching high levels of accuracy. The aforementioned study used $6\,002$ images of "cracks" and "non-cracks". More examples of similar detection strategy using different classification algorithms can be found in a literature review conducted by Sony [9].

An interesting option allowing to overcome the need of the high number of images is represented by the concept of transferred learning which is the

domain of computer vision frameworks e.g., R-CNN (Region-based CNN), Faster R-CNN, AlexNet, Oxfor-Net, YOLO (You Only Look Once), ResNet50, VGG-16. Such frameworks are usually based on layer(s) of CNNs subsequently connected to feed-forward neural networks or similar for classification purposes and in case of object detection problems also to a regression algorithm, which also allows to accurately draw a bounding box around the objects area. Note that frameworks are obtainable as pre-trained i.e., the user doesn't need to train the whole network from scratch but rather tune it for a specific problem. In some specific cases, this may lead to the reduction of the total number of images needed for achieving accurate predictions. Özgenel [10] compared the performance of 7 different computer vision frameworks and concluded that transfer learning applied to crack detection problems is feasible. In terms of transfer learning, the study also advised to use limited number of training images especially when the variance among the images is low as the high number of images increases computational time as well as the risk that the model learn some specific crack patterns too well and fails to comprehend general appearance of cracks (overfitting).

The majority of research papers investigating the usage of presented computer vision techniques comes mainly from three research areas: bridge health monitoring, pavement monitoring and the large-scale monitoring of buildings after earthquakes. The first two areas use the techniques described above for damage detection. The latter mainly uses 3D cloud of points (set of points in virtual 3D space usually forming a base for 3D models) to quickly find collapsed buildings and subsequently applies computer vision techniques to detect damages of smaller scale [11]. Practical example for concrete bridges is provided by Yang [12] who performed a field test with a drone and by the means of VGG-16 framework successfully detecting over 70 % of cracks and spalling areas on a ground level of a bridge. The means of transferred learning was used to detect cracks on pavements in the work of Gopalakrishna [13] who used 1 056 images in total for training purposes. More computer vision based examples of bridge health and pavement monitoring alongside with examples of different applications (building health monitoring or inspection of underground structures) may be found in the work of Sony [9].

Practical applicability of pre-trained frameworks in detecting defects on building envelopes may be well illustrated on the work done by Wang [14] who feed 500 annotated images to Faster R-CNN framework to identify and draw bounding boxes around bricks subjected to spalling and efflorescence with resulting accuracy 95 %. Another good example is provided by the research of Cha [15] where modified Faster R-CNN is used to differentiate concrete cracks, steel corrosion, bolt corrosion and steel delamination. In total 2 366

annotated training images were fed to the framework achieving mean average precision 89.7 % while testing on previously unseen images.

Image segmentation is another computer vision technique allowing to precisely highlight the area of interest i.e., to color out pixels which are associated with a crack or other type of damage. Hoskere [16] used a CNN on a level of individual pixels rather than on whole images eventually adding a label to each pixel and thus highlighting six different types of damage. Already mentioned work by Gallareta [11] combined multiresolution segmentation allowing to differentiate building objects (i.e., façades, windows etc.) and spectral difference segmentation allowing to highlight damaged areas (i.e., cracks, holes etc.). Outputting highlighted areas from image segmentation may be used for measuring distances within an image eventually allowing to measure the scope of a damage. Such approach was illustrated in the work of Lins [17] who used images to measure width and length of cracks.

3. Case study

In the following case study, an in-house computer vision algorithm (Build-Sense model) based on the YOLO-tiny framework [2] capable of detecting damaged areas is applied to images of historical Tjolöholm castle in the west of Sweden. The main goal is to detect missing joints in a stone masonry wall.

3.1. Object detection model

The Build-Sense model was originally built for detecting damage on brick walls specializing on drawing bounding boxes around cracks and bricks subjected to spalling and efflorescence. The model is utilizing the concepts of transfer learning and was trained based on 225 images of damaged brick walls (approximately 75 images for each damage type) manually annotated with more than 2 000 bounding boxes providing the model with the interpretation of how a damage looks like. Figure 1 shows bounding boxes around efflorescence as they were predicted by Build-Sense model.

The current predictive power of the Build-Sense model may be characterized as good enough for initial stage of inspections as it in theory may notify the inspector about possible building envelope failure. However, given its rough estimation of the damaged area it is not yet suitable for accurate damage quantification i.e., measuring the area of damage, width of damage etc.

3.2. Tjolöholms castle and data collection

The construction of Tjolöholms castle was finished in the beginning of 20$^{\text{th}}$ century. Thenceforth, it has been subjected to many refurbishments given its continuous moisture related problems which occurs in the form of mold growth and ongoing disintegration of joints in the stone masonry walls forming the building envelope of the castle. Problems are more obvious on
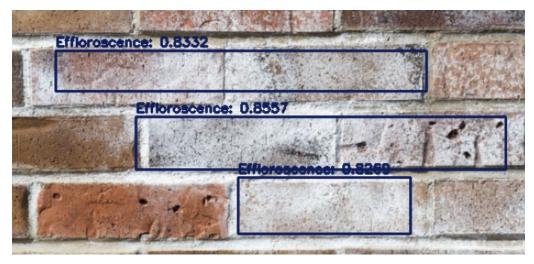
Figure 1. Example of output from the Build-Sense model – model draw bounding boxes around damaged area alongside with its level of confidence [18].

the façade facing the south-east with the open view towards the coast of North Sea.

As foundation for the status assessment a basic ground level visual inspection of the building envelope was performed and complemented by inspecting building insides, including the roof. Following the many years of subsequent repair work in the castle, there is a substantial amount of data. This is stored in the castle management's own databases containing technical drawings, data from hygrothermal wall sensors, 3D model of the building (based on cloud of points), panoramic images of each room from interior and panoramic images taken on 78 different locations around the main building. Note that performing an UAV survey for acquisition of close-up images of building envelope was not possible at the time of the visit. Therefore, obtained outdoor panoramic images are used in subsequent computer vision analysis.

3.3. Scope and the methodology of the case study

Given the fact that joints are continuously disintegrated, mainly on the east side of the south-west façade, the study is limited to the assessment of the Build-Sense model only to one panoramic image which closely depicts the area of interest. The positions of the panoramic image as well as the overall view over the wall of interest is on the Figure 2.

Given the limitations of the position i.e., testing images will not be in direct opposition to the wall surface but rather tilted, and given the complexity of light exposure conditions (uneven surface is under direct sunlight casting a lot of shadows), one area of interest on a building envelope will be analyzed from different angles. The confidence threshold for the model was set to be 50 % i.e., the model draws bounding boxes around objects only where the probability of accurate detection is higher than 50 %.

3.4. Retraining the Build-Sense model

Direct application of the current Build-Sense model is not feasible in this problem as it was not able to detect any missing joints in the castle's stone masonry wall. This is expected as it was originally tuned for detecting cracks, efflorescence and spalling on brick walls i.e., each construction has different structure and different damage patterns which are different from missing joints in stone masonry walls. Therefore, the model needed to be retrained specifically for the purpose of detecting missing joints in stone masonry walls.

Images for re-training purposes were collected in the streets of Gothenburg city (in the west of Sweden) by the in-built high-resolution camera in iPhone SE 2019. Initial 115 images of stone masonry walls were annotated highlighting damaged joints. The total number of images was multiplied in the image augmentation process by performing 90°, 180° and 270° rotation. The images were then split to train/test samples by the ratio of 80/20. Training itself was performed in the Google Collaboration environment allowing to borrow Graphics Processing Units (GPU) for the purpose of calculation. The whole training session took 6.3 hours to complete on a cloud based dedicated GPU Nvidia Tesla K80 controlled by a local computer with 2,.3 GHz Quad-Core Intel Core i5 processor. Note that the time of the calculation could been decreased by limiting the pixel size of images used for the training purposes.

The model iterated over the images for 6 000 times internally using mean average precision (mAP) as a metric for network's weights evaluation. Final output of the training is a set of weights which are the best in predicting missing joints given the training images and the number of iterations. The best weights are then used for subsequent analysis of images from Tjolöholms castle.

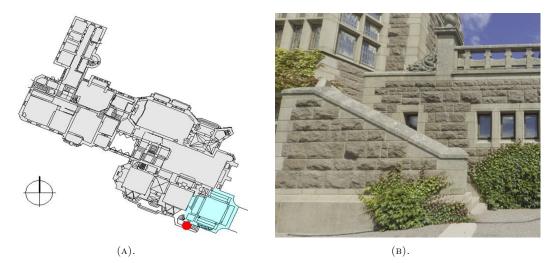(A).   (B).

FIGURE 2. (A) Schematic plan on the building's ground floor to the left shows the location where the panoramic photograph was taken (red dot). Gray background indicates multi-storey parts while cyan colour indicates single storey part of the building. (B) Photograph was taken on the stairs leading to the roof of single storey part [19].

FIGURE 3. Default view from the panoramic image directly facing the area of interest [19].
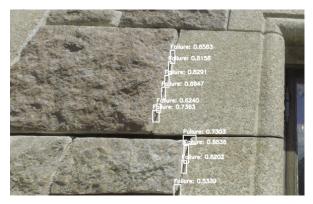
3.5. RESULTS AND DISCUSSION

All images altogether presented in this subchapter were analyzed using retrained Build-Sense model. Figure 3 shows the default (un-zoomed) view from the panoramic image directly facing the building envelope. Note that there are missing joints on the left side of the window. The model there fails to detect any objects of interest.
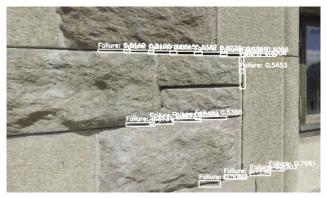
However, this was not a case for zoomed images. Figure 4a shows missing joints in the vertical direction but the algorithm failed to comprehend damages in the horizontal direction. Observe that the confidence of the model in predicting missing veneer is relatively high i.e., always above 70 %. Complete opposite picture gives the Figure 4b where the missing joints were detected mainly in horizontal direction while the model's confidence is in general lower.

The model failed to comprehend a majority of miss-

ing joints in Figure 4c. However, vertical joints are in this point of view badly visible and difficult to correctly interpret even for human eyes. Furthermore, the model again fails to comprehend joints in horizontal direction which makes the confidence of the model over individual bounding boxes rather low. Observe that the model prefers detecting missing joints of relatively thin size but fails to detect thicker joints. This is probably caused by the limited number of training images and by the low variance of images i.e., training set does not include missing joints of this size.

The presumption that the low confidence in the previous case is caused by the low number of relevant training images is further strengthened by Figure 5 where most missing joints were correctly detected with high level of confidence. The model only fails to grasp missing veneers in the opposite wall in the distance and few parts of the joints on the sides.

(A). Detailed view 1.

(B). Detailed view 2.

(C). Detailed view 3.

Figure 4. Detailed views. Credit: Anders Jansson, all rights reserved.

Figure 5. Image facing stairs [19].

4. Conclusion

This study investigated how the inspection process of a building envelope may be automated by the means of UAVs and subsequent computer vision based analysis of images. The first part of the paper is a literature review summarizing how drones may be automatically navigated to capture data from a building envelope. The second part of the literature review shows how computer vision algorithms are used in image analysis. The latter part of the review was exemplified using computer vision based model (Build-Sense model) for detection of missing joints on a wall of Tjolöholms castle, a historical building in Sweden. The model failed to comprehend majority missing joints in vertical direction and joints of larger thicknesses. It was concluded that the model may be improved by adding a variance to the training set (more different examples) and by expanding the total number of the training set. Nevertheless, current predictive power of the model may be typified as good enough for notifying inspectors, who should be responsible for final damage evaluation, about potential building envelope failure.

Despite the performance limits of the presented model, the conjunction of fast and overreaching UAVs with the computer vision promises to bring revolution to the building inspection process. When properly set, UAVs may scan whole building envelope in terms of minutes. Subsequent analysis may be used for accurate damage quantification of building envelope i.e., detecting the number of cracks per square meter, measuring the scope of a damage etc. By separating the envelope into smaller parts, one could differentiate the performance of the envelope over those parts and subsequently use this information for accurate data-based failure predictions, opening the gate of predictive maintenance.

Acknowledgements

References

[1] V. Hasik, E. Escott, R. Bates, et al. Comparative whole-building life cycle assessment of renovation and new construction. *Building and Environment* **161**:106218, 2019. https://doi.org/10.1016/j.buildenv.2019.106218

[2] Z. Jiang, L. Zhao, S. Li, Y. Jia. Real-time object detection method based on improved YOLOv4-tiny, 2020. [2022-02-12]. https://doi.org/10.48550/arXiv.2011.04244

[3] L. Duque, J. Seo, J. Wacker. Bridge deterioration quantification protocol using UAV. *Journal of Bridge Engineering* **23**(10):04018080, 2018. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001289

[4] T. Rakha, A. Liberty, A. Gorodetsky, et al. Heat mapping drones: An autonomous computer-vision-based procedure for building envelope inspection using Unmanned Aerial Systems (UAS). *Technology|Architecture + Design* **2**(1):30–44, 2018. https://doi.org/10.1080/24751448.2018.1420963

[5] H. Freimuth, M. König. Planning and executing construction inspections with unmanned aerial vehicles. *Automation in Construction* **96**:540–553, 2018. https://doi.org/10.1016/j.autcon.2018.10.016

[6] G. Morgenthal, N. Hallermann, J. Kersten, et al. Framework for automated UAS-based structural condition assessment of bridges. *Automation in Construction* **97**:77–95, 2019. https://doi.org/10.1016/j.autcon.2018.10.006

[7] A. Benz, J. Taraben, P. Debus, et al. Framework for a UAS-based assessment of energy performance of buildings. *Energy and Buildings* **250**:111266, 2021. https://doi.org/10.1016/j.enbuild.2021.111266

[8] K. Chaiyasarn, M. Sharma, L. Ali, et al. Crack detection in historical structures based on Convolutional Neural Network. *International Journal of GEOMATE* **15**(51):240–251, 2018. https://doi.org/10.21660/2018.51.35376

[9] S. Sony, K. Dunphy, A. Sadhu, M. Capretz. A systematic review of convolutional neural network-based structural condition assessment techniques. *Engineering Structures* **226**:111347, 2021. https://doi.org/10.1016/j.engstruct.2020.111347

[10] Ç. F. Özgenel, A. G. Sorguç. Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In J. Teizer (ed.), *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC)*, pp. 693–700. 2018. https://doi.org/10.22260/ISARC2018/0094

[11] J. Fernandez Galarreta, N. Kerle, M. Gerke. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Sciences* **15**(6):1087–1101, 2015. https://doi.org/10.5194/nhess-15-1087-2015

[12] L. Yang, B. Li, W. Li, et al. Deep concrete inspection using unmanned aerial vehicle towards CSSC database. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2017. 9 p.

[13] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, A. Agrawal. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials* **157**:322–330, 2017. https://doi.org/10.1016/j.conbuildmat.2017.09.110

[14] N. Wang, X. Zhao, P. Zhao, et al. Automatic damage detection of historic masonry buildings based on mobile deep learning. *Automation in Construction* **103**:53–66, 2019. https://doi.org/10.1016/j.autcon.2019.03.003

[15] Y.-J. Cha, W. Choi, G. Suh, et al. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering* **33**(9):731–747, 2018. https://doi.org/10.1111/mice.12334

[16] V. Hoskere, Y. Narazaki, T. A. Hoang, B. Spencer. Vision-based structural inspection using multiscale deep convolutional neural networks, 2018. [2022-02-14]. https://doi.org/10.48550/arXiv.1805.01055

[17] R. G. Lins, S. N. Givigi. Automatic crack detection and measurement based on image analysis. *IEEE Transactions on Instrumentation and Measurement* **65**(3):583–590, 2016. https://doi.org/10.1109/TIM.2015.2509278

[18] Karagrubis. Brick efflorescence. https://rainguardpro.com/

[19] A. Jansson. Tjolöholms castle.