ELECTRIC POWER
RESEARCH INSTITUTE

Automated Analysis of Remote Visual Inspection of Containment Buildings

Technical Brief — *Nuclear Plant Modernization*

This technical brief summarizes the progress to date on the efforts to develop machine vision models to automatically detect damage in concrete structures. For example, it is envisioned that such tools would enable utilities to maximize the benefits found in deploying unmanned aerial systems for remote visual inspection of containment buildings. Although not ready for deployment, the results of the initial model on a limited dataset show that this approach is feasible and can provide value to the industry. A number of activities for improvement have been identified and are underway, targeting field trials in 2021-2022.

INTRODUCTION

Traditionally, visual inspection of containment buildings and other large concrete structures in the nuclear industry are time-consuming tasks incurring considerable risk to personnel safety. More recently, the industry begun assessing the use of unmanned aerial systems (UAS) to perform such inspections, and noted that besides considerably reducing the risks to personnel safety, such an approach can also decrease inspection time and cost while still satisfying all regulatory inspection requirements [1] [2]. A business case for the use of drones in nuclear power plants is planned for early 2021 that will provide more details on the cost benefit of the approach.

The benefit of this approach is diminished by the increased burden on data analysis that come from the large amounts of high quality image or video data that are typically generated. Looking to enable utilities to maximize the value of remote inspections with UAS, EPRI has started the development of machine vision models to automatically detect damage in concrete to assist in the analysis and review of the data.

WHAT DATA DID WE USE?

Artificial Intelligence (AI) projects rely to a great extent on the available data to support it. To date, EPRI has leveraged the data obtained in a field trial [1] covering a fraction of a containment building, and data from a full containment building inspection performed with UAS provided by one member utility.

Data was recorded either in the form of videos or sequential images with varying resolutions (most commonly 1080x1920; some as high as 4912x7360) and at different distances from the structures. Individual frames were extracted from the videos at a rate of approximately one frame every 2 seconds. Altogether, there were roughly 2,500 images available for the effort, out of which approximately 90% were used for model development; the remaining images were held back for model assessment.

The different types of damage included in the available dataset and their approximate quantity are listed in Figure 1 (see [3] for a description of each damage type). Note that this count is the approximate number of times a defect appears in the image collection; it does not reflect the number of defects in the structures, since the same defect may appear in multiple sequential images. Abrasion, honeycomb and pattern cracking do not appear in sufficient quantities to be part of the scope of the project; the limited quantities of corrosion and spall examples may also limit the model performance on these damage types.

WHAT SOLUTIONS ARE WE DEVELOPING?

Two separate models, image classification and defect localization, were initially developed as described below. For computational efficiency, both models break up the image into smaller tiles of fixed sized, with prescribed overlap, and operate on the tiles.

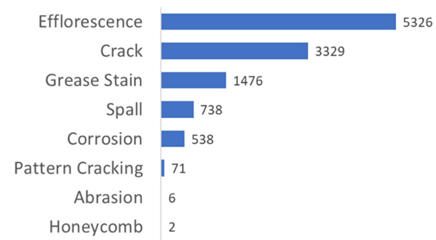

Figure 1 – Feature Count in Available Dataset

Image Classification

The Image Classification model simply indicates the presence of any one of the given types of damage in each image tile, without localizing it. The result is illustrated in Figure 2: each tile that composes the larger input image is tagged with the type of damage (if any) that the model predicts it to contain. Although not localizing the damage, it provides information about what damage type to look for and reduces the search area: the currently chosen tile size of 640x640 pixels corresponds to approximately 20% of the area of the typical high-definition image (1920x1080 pixels). Computationally, the classification model is considerably lighter and faster than the defect localization model, and thus more amenable for real-time implementation. One possible use case is that of informing the pilot in real-time of identified regions that warrant a closer inspection.

Results from an initial classification model are summarized in Table 1. The metrics in this table are:

- ***Positive Tiles (P):*** Number of tiles that contain the damage.
- ***Negative Tiles (N):*** Number of tiles that do not contain the damage.
- ***True Positive (TP):*** Number of tiles correctly predicted to contain the damage.
- ***True Negative (TN):*** Number of tiles correctly predicted not to contain the damage.

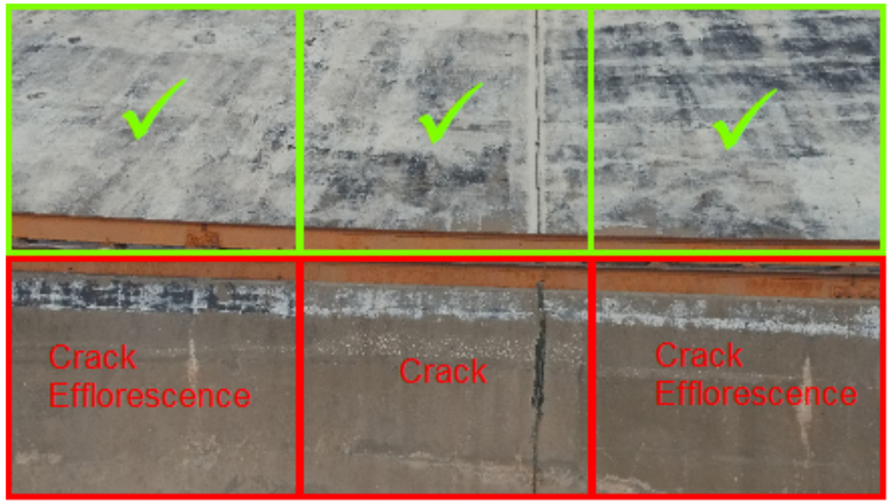

Figure 2 – Illustration of the Results of the Image Classification Model (Tile Overlap not Shown)

Table 1 – Initial Classification Model Results

Damage Type	P	N	TP	TN	FP	FN	TDR	FCR	Defect Count	Defects Detected	DDR
Abrasion	0	4584	–	4584	0	0	–	0%	0	0	–
Corrosion	75	4509	61	3863	646	14	81%	14%	8	7	88%
Crack	383	4201	294	2227	1974	89	77%	47%	101	87	86%
Efflorescence	292	4292	239	4160	132	53	82%	3%	108	106	98%
Grease Stain	282	4302	179	4003	299	103	63%	7%	10	10	100%
Honeycomb	0	4584	0	4575	9	0	–	0%	0	0	–
Pattern Cracking	67	4517	3	4511	6	64	4%	0%	1	1	100%
Spall	124	4460	90	3299	1161	34	73%	26%	21	18	86%

- ***False Positive (FP):*** Number of incorrect positive predictions.
- ***False Negative (FN):*** Number of incorrect negative predictions.
- ***Tile Detection Rate (TDR):*** Fraction of tiles with damage that have been detected (TDR = TP/P).
- ***False Call Rate (FCR):*** Fraction of negative tiles classified as positive (FCR = FP/N).
- ***Defect Count (DC):*** Number of unique defects.
- ***Defects Detected (DD):*** Number of unique defects that were detected.
- ***Defect Detection Rate (DDR):*** Fraction of unique defects that were detected (DDR = DD/DC).

Key observations about the performance of this initial classification model are:

1. ***Reasonable detection:*** Tile Detection Rate (TDR) indicates how many of the tiles with damage have been detected and is, therefore, a measure of the detection capability of the model. For crack, corrosion and efflorescence, TDR is near or above 80%. Grease stain and spall show lower tile recall values (63% and 73%, respectively).

   Because of the sequential nature of the images and tile overlap, the same physical defect will typically appear in multiple tiles. Considering that in practice each defect only needs to be detected in one of those opportunities, the detection rate at the tile level typically underestimates the true detection rate of the model. As seen in Table 2, the detection rate for unique defects for all damage types within the scope of the model is higher than 85%.

2. ***Screening despite high number of false positives:*** The results in Table 1 show that the current model is typically performing with a relatively high number of false positives. However, with the exception of cracks, the number of false positives is small compared to the total number of negative tiles, and the model still provides screening (low false call rate). Additionally, this may be an application where the desired model performance is biased towards better detection at the expense of false calls to a certain extent. Notwithstanding, efforts are underway to improve the precision of the models to decrease the number of false positives.

3. ***Performance & data availability:*** Considering both the detection rate and the number of false positives, the best performance is seen for efflorescence, which is the damage type with most examples in the dataset. Cracks have a considerably poorer performance especially in terms of false calls; data sufficiency may be an issue (comparatively, cracks have 60% of the examples of efflorescence). Grease stain shows reasonable performance (better than cracks if detection and false calls are balanced) despite having fewer examples, showing that this may be an easier damage type to characterize as compared to cracks. This indicates that perhaps different damage types may require different models, which is something that can be explored in continuing efforts. With fewer than 1,000 examples in the dataset, spall and corrosion show the worst overall performance. While these results show that this approach is feasible, more examples and datasets are needed to bring model performance to the desired level. EPRI continues with efforts to obtain more field datasets from utilities as well as generating synthetic data to support further model development.

Damage Detection

The damage detection model localizes the damage in the individual image tiles before aggregating the results back to the original image. Typical results are illustrated in Figure 3 through Figure 5, where the red shaded regions indicate the model predictions with accompanying label. As can be seen, the model provides precise localization of the defect in the image. Being heavier computationally, this model is more suitable for offline analysis, although initial estimates make it possible to run at a sub-sampling of the typical image frame rate, such as 1 frame per second, which is enough to provide coverage. Another possible real-time implementation involves chaining the models, so that localization would only be performed in specific regions of a subset of frames as selected by the classification model.

Performance is summarized in Table 2. The initial model achieves high detection rates (95%) for efflorescence and corrosion (despite the low number of corrosion examples) and around 80% for cracks and grease stain. Spalls are seen to be challenging, and the model was unable to detect any of its instances. Again, it is noted that a defect needs to be detected only once, so the instance detection rate presented in Table 2 typically underestimates the true defect detection rate.

In this case, the number of negative examples or the true negative count is not defined, and one looks at the precision of the model instead: the fraction of the model calls that are actually correct. For instance, the 24% precision for corrosion indicates that only about one quarter of the model predictions for that damage type are actually correct. Similar to the classification model, the defect localization model has a considerable number of false positives, lowering the model precision. Here, it should be noted that the model is identifying a small, very specific region of the image, and false calls would be less burdensome in review than for model classification. Additionally, in some cases the model correctly detected damage that was not identified during the initial labeling process, so that review of the model results is also leading to improvement of the labeling process in an iterative fashion; at this time, such cases are computed as false positives in Table 2. Further efforts are underway to improve model precision.

Again, in terms of both detection rate and precision, the best performance is seen for efflorescence, the most common defect type in the dataset. The model performs similarly on cracks and grease stains despite the larger number of crack examples.

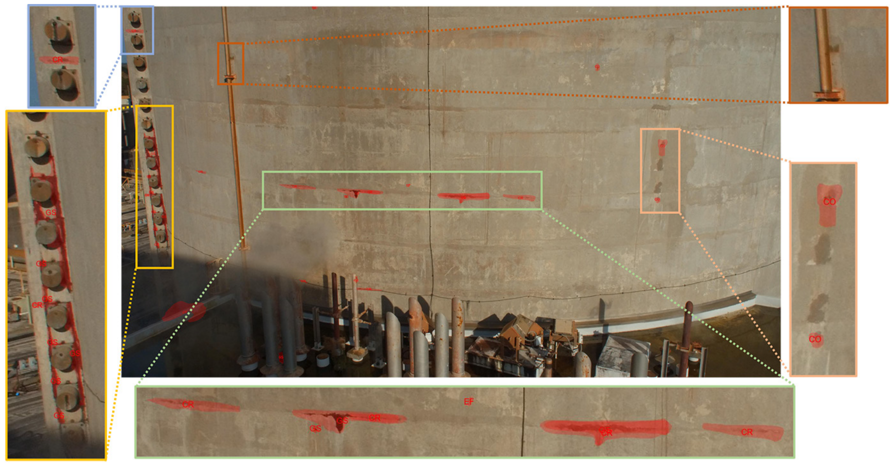

Figure 3 – Example of the Results of the Damage Detection Model. Details show: a crack detected between tendon caps (top left); grease stain detection around tendon caps (bottom left); crack and grease stain detections on the wall, with a miss-called efflorescence (bottom); a missed spall (top right); and corrosion detection (bottom right).

Figure 4 – Damage Detection Model Results Showing Mainly Efflorescence on the Dome

Figure 5 – Damage Detection Model Results Showing Cracks on the Wall

Table 2 – Initial Damage Detection Model Results

Damage Type	P	TP	FP	FN	Detection Rate $\left(\frac{TP}{P}\right)$	Precision $\left(\frac{TP}{TP+FP}\right)$
Abrasion	0	–	–	0	–	–
Corrosion	63	60	195	3	95%	24%
Crack	477	403	883	74	84%	31%
Efflorescence	749	710	646	39	95%	52%
Grease Stain	230	174	398	56	76%	30%
Honeycomb	0	0	44	0	–	0%
Pattern Cracking	12	6	104	6	50%	5%
Spall	85	0	0	85	0%	–

GOING FORWARD

In summary, the results obtained in this first-cut model with a limited dataset show that developing machine vision algorithms to detect damage in concrete and aid in the analysis of remote visual inspection of large structures is feasible. While the performance of the current models may not be sufficient for efficient implementation (especially in terms of false positives), a number of improvement efforts informed by the above assessment have been identified and are underway. Future activities include:

- ***Enrich the dataset:*** EPRI is engaging in the development of synthetic images for this application, focusing on the less common or more challenging defect types: corrosion, cracks, pattern cracking and spalls. It is expected that having more examples of these damage types will help improve model performance. Also, more field datasets are needed for testing and performance assessment; while the model can leverage synthetic data for development and training, it should be assessed in real field data. EPRI continues to engage with member utilities to obtain data for this purpose. Utility members interested in participating should contact EPRI; benefits of providing data include receiving annotations (labels) for their data and ensuring representation of their structures in the model.
- ***Re-train, tune and expand models:*** Once more data is available, the models can first be further assessed on the new field data and then re-trained to incorporate both the new field and synthetic datasets. Observations from this assessment will inform model tuning and selection. For instance, it may be more adequate to have different classification models for different damage types.
- ***Integrate & implement models:*** Once a new generation of models is developed, implementation strategies will be considered. One possibility is to chain classification and damage detection models, for instance. EPRI will look at different implementation approaches to be tested in field trials.
- ***Share datasets & models:*** EPRI aims to make the anonymized datasets and benchmarked models available to member utilities. They can then leverage the models and data to develop their own applications, which may go beyond the types of structures included in the scope of this project.

It is expected that the activities above will improve model performance and allow for initial field implementation of the technology in the near future, enabling utilities to automate the bulk of the analysis of data resulting from remote visual containment inspections or to obtain real-time feedback and assessment to improve inspection efficiency and reliability.

REFERENCES

1. *Remote Visual Inspections with Unmanned Aerial Systems.* EPRI, Palo Alto, CA: 2018. 3002013193.
2. R. Knight, "Nuclear Industry Putting Trust in UAS." Unmanned Systems, https://insideunmannedsystems.com/nuclear-industry-putting-trust-uas/. March 5, 2018
3. *Field Guide: Visual Inspection of Concrete Structures in the Nuclear Fleet.* EPRI, Palo Alto, CA: 2016. 3002007799.

FOR MORE INFORMATION

For more information, contact the EPRI Customer Assistance Center at 800.313.3774 (askepri@epri.com).

Sam Johnson | *Senior Technical Leader*
Program | Nondestructive Evaluation
Phone | 704.595.2596
Email | sjohnson@epri.com

Thiago Seuaciuc-Osorio | *Senior Technical Leader*
Program | Nondestructive Evaluation
Phone | 704.595.2841
Email | tsosorio@epri.com

Note

For further information about EPRI, call the EPRI Customer Assistance Cen- ter at 800.313.3774 or e-mail askepri@epri.com.

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI also provides technology, policy and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI members represent 90% of the electric utility revenue in the United States with international participation in 35 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; and Lenox, Mass.

Together . . . Shaping the Future of Electricity

3002018419 November 2020

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com