# Hierarchical Residual Network with Bayesian Modeling for Robotic Vision-Based Crack Detection

*Qiuchen Zhu and Quang Ha -- University of Technology Sydney*
***Faculty of Engineering & IT***

***E-mail: {Qiuchen.Zhu, Quang.Ha@}uts.edu.au***

## Outline

- Introduction
  - Motivation
  - Infrastructure Monitoring
- Robotic Surface Inspection
  - UAV System
  - SYDCracks
- Methodology
  - Convolutional Neural Network
  - Hierarchical Residual Network
  - Information Loss and Nonlinearity Tradeoff
  - Bayesian Inference
- Quantitative Measures and Results
- Conclusion
- Future work

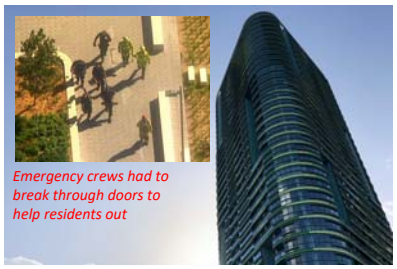
## Motivation

https://www.abc.net.au/news/2019-02-22/opal-tower-report-finds-beams-burst-low-strength-concrete-used/10835498

### Opal Tower in Auburn Australia

Cracks in precast panels at level 10 (Jan 2019)

Homeless at Xmas




Emergency evacuation right before Christmas 2018 (Dec 2018)

"This is in response to an ongoing and persistent cracking and structural deformation observed within the primary support structure and the facade masonry. This deterioration has been rapid, hence expedited propping was deemed a necessary precaution to ensure the safety of the building and its occupants."

## Motivation

https://www.abc.net.au/news/2019-06-15/mascot-towers-sydney-evacuation-what-we-know-about-cracks/11213664

### Surface cracks: structural health indicator of built assets

Identified crack on the wall (Jun 2019)

Alert rises for residents

Emergency evacuation in Mascot, Sydney (Jun 2019)

**If the cracks are detected in the early stage, the potential risk can be avoided in advance.**

## Motivation

### Practical solution: UAV inspection

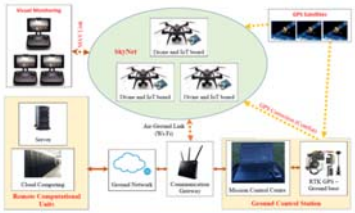
Unclimbable surface

Unreachable height

Endangered areas

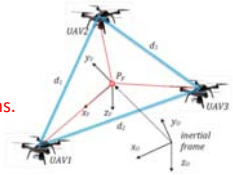
**UAVs can complete several crack detection tasks which are difficult to human.**

## UAV Development @ UTS



Coverage, Communication
Low level control, High-level control
Cooperation Path planning,
Search & Rescue- Inspection
Angle-encoded PSO
Image processing

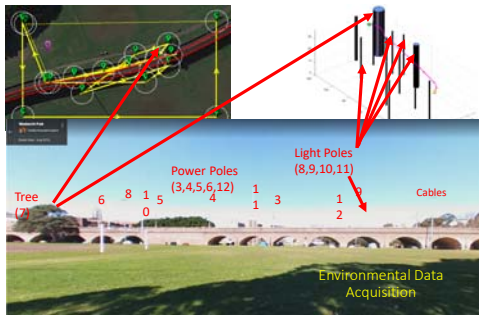
Internet of Things increases processing capabilities of RCU and reduce communication distance burdens.



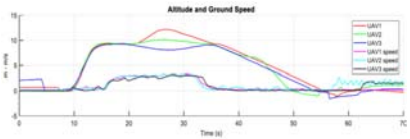


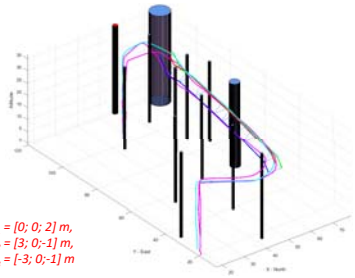
Van T. Hoang ; Manh D. Phung ; Tran H. Dinh ; Quang P. Ha, System Architecture for Real-Time Surface Inspection Using Multiple UAVs, *IEEE Systems Journal*, vol. 14, no. 2, 2020, pp. 2925-2936. DOI: 10.1109/JSYST.2019.2922290.

## UAV Formation Path Planning




V.T. Hoang, M.D. Phung, T.H. Dinh, Q. Zhu and Q.P. Ha, "Reconfigurable Multi-UAV Formation Using Angle-Encoded PSO," *Proc. 2019 IEEE International Conference on Automation Science and Engineering* CASE 2019, Vancouver Canada, 22-26 Aug 2019, pp. 1670-1675. doi: 10.1109/COASE.2019.8843165.



$\Delta T_1$ = [0; 0; 2] m,
$\Delta T_2$ = [3; 0;-1] m,
$\Delta T_3$ = [-3; 0;-1] m

# Image acquisition for crack detection test

### Image acquisition via UAV vision-based Surface Inspection:

- cruising around the identified infrastructure in a queue
- take photos for infrastructure surfaces

Manh Duong Phung, Cong Hoang Quach, Tran Hiep Dinh, and Q. Ha, "Enhanced Discrete Particle Swarm Optimization Path Planning for UAV Vision-based Surface Inspection," *Automation in Construction*, Vol. 81, pp. 25-33, 2017. DOI: 10.1016/j.autcon.2017.04.013.

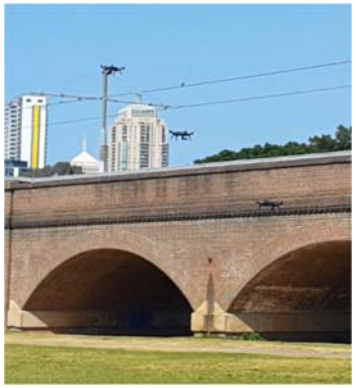


Image Acquisition

# Original UAV collected dataset

SYDCracks:

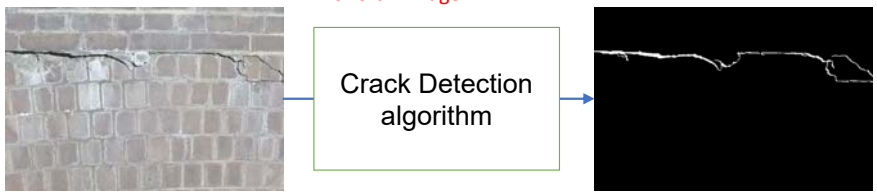
170 comprehensive infrastructure images Collected by our UAV in Sydney

Including typical surface cracks on concrete wall, bridge and brick fence

## SYDCracks

170 infrastructure crack images collected by our UAVs

Monorail Bridge

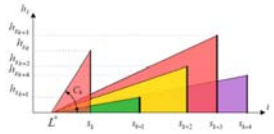



Surface image collected by our own UAVs

Binary Crack map

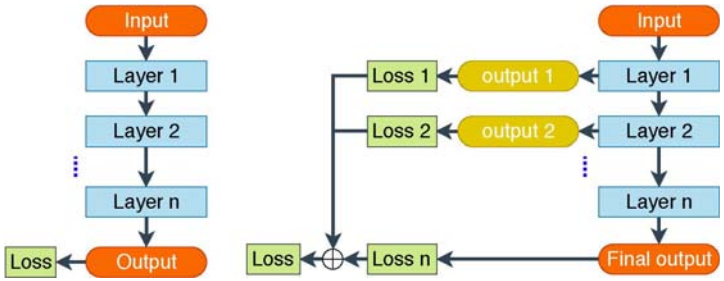
- **Crack detection:** developing an automatic and adaptive algorithm to detect cracks from UAV collected images

Tran H. Dinh ; Manh D. Phung ; Quang P. Ha, "Summit Navigator: A Novel Approach for Local Maxima Extraction," *IEEE Transactions on Image Processing*, Vol. 29, pp. 551-564, 2020. IF 6.79.



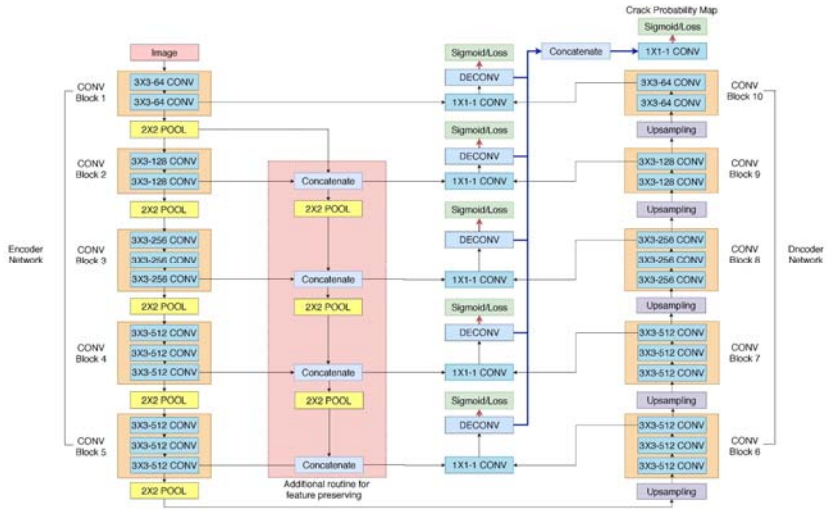
# Sequential vs Hierarchical model

Difference in outputting a feature abstraction:




Flowcharts of forward propagation: (a)Sequential; (b)Hierarchical

Hierarchical model has less bias due to the involvement of multiscale information.

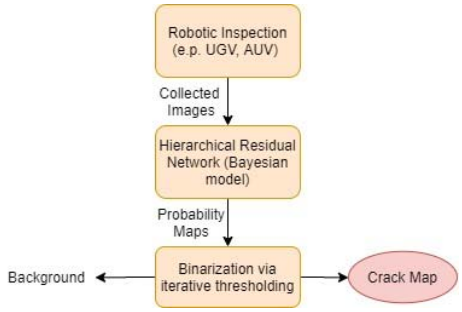



Enhanced Hierarchical Convolutional Neural Networks(EHCNN)

# Proposed Crack Detection Method

**Hierarchical residual network with Bayesian modelling (HRNBM)**

- Bayesian auto-encoder as the main network for feature extraction
- Hierarchical residual module to present detail of crack features
- Crack probability map generated via merging activated features
- Binary crack map obtained using thresholding




Flowchart of crack detection approach

# Trade-off on nonlinearity

Probability model of an identified pixel:

$$P(C_1|x_i) = \frac{P(C_1, x_i)}{P(x_1)} = \frac{P(x_i|C_1)P(C_1)}{P(x_i|C_1)P(C_1) + P(x_i|C_0)P(C_0)} = \frac{1}{1 + \frac{P(x_i|C_0)P(C_0)}{P(x_i|C_1)P(C_1)}} = \frac{1}{1 + e^{-a(x_i)}} \quad (4)$$

where $a(x_i) = \ln \frac{P(x_i|C_1)P(C_1)}{P(x_i|C_0)P(C_0)}$

- $x_i$ -- the intensity value of the $i^{th}$ pixel.
- $C_0, C_1$ -- a pair of random events: being a crack pixel or a non-crack background respectively.

The model is consistent with a sigmoid function :

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

# Trade-off on nonlinearity

Assume that the conditional probabilities follow Gaussian distribution with the same variance [*]

$$P(x_i|C_j) \sim \mathcal{N}(x_i|\mu_j, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{(x_i-\mu_j)^2}{2\sigma^2}]$$

- $j = 0,1$ - the two classes in the binary segmentation.

Then, $a(x_i) = \ln P(x_i|C_0) - \ln P(x_i|C_1) + \ln \frac{P(C_0)}{P(C_1)} = \frac{\mu_0-\mu_1}{\sigma^2} x_i + \frac{\mu_1^2-\mu_0^2}{2\sigma^2} + \ln \frac{P(C_0)}{P(C_1)} = \omega x_i + \omega_0$

- $\omega = \frac{\mu_0-\mu_1}{\sigma^2}$ and $\omega_0 = \frac{\mu_1^2-\mu_0^2}{2\sigma^2} + \ln \frac{P(C_0)}{P(C_1)}$ are two constants.

The right probability is achieved when the output is linear.
The deviation due to nonlinearity in middle layers is therefore inevitable but can be alleviated.

[*] K.-P. Murphy. Machine learning: a probabilistic perspective. *MIT Press*, Cambridge, Massachusetts, 2012.

# Entropy loss

- Entropy loss for a single pixel:

$$L_{lr}(i,j) = -q \log \{P[F_{lr}(i,j)]\} - (1-q) \log \{1 - P[F_{lr}(i,j)]\} \quad (1)$$

- Entropy loss for merged results from blocks on the same level:

$$L_{lr} = \sum_{i=1}^{m} \sum_{j=1}^{n} L_{lr}(i,j) \quad (2)$$

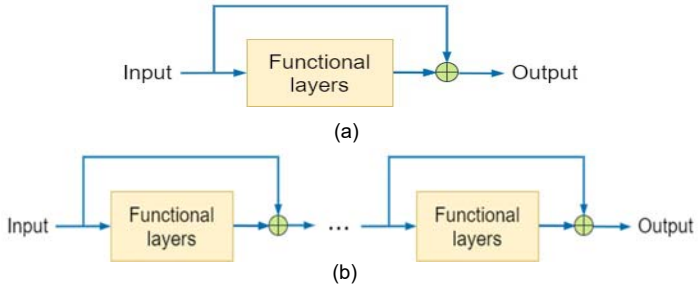
- Entropy loss for a single block:

$$L = \frac{1}{N_r} \sum_{l=0}^{5} \sum_{r=1}^{N_r} L_{lr} \quad (3)$$

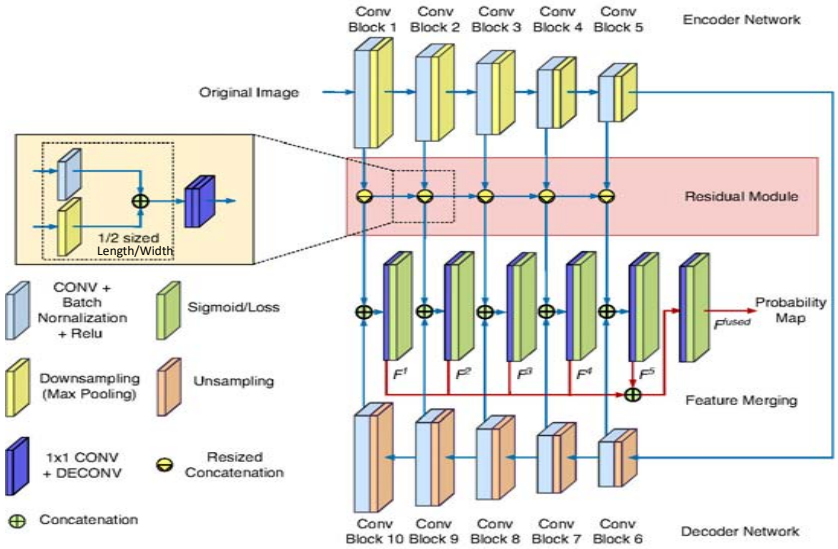
- $q$ -- the label in the ground truth (binary value 0,1).
- $F_{lr}$ -- feature map generated from $l^{th}$ level of blocks for sample $r (l = 0$ repsent for crack map).
- $P$ – probability of $F_{lr}(i,j)$ (calculated via sigmoid function).
- $N_r$– number of training samples.

# Residual Modules

Residual connections in deep neural networks:




Flowcharts of residual connections: (a)Single; (b)Recursive
Functional layers include highly nonlinear operations
Residual Modules can reduce the redundant nonlinearity in deep neural networks.

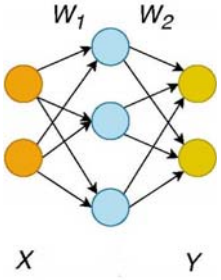



Hierarchical Residual Networks(HRN)

# Frequentist models vs Bayesian models

The difference in the inference for a dataset $X = \{ x_i | i = 1,2 \dots N\}$; pixel values of the original image

$Y = \{ x_i | i = 1,2 \dots N\}$: the ground – truth mask




Frequentist: determine the weights $W$ using maximum likelihood estimation

$$W_{ML} = \text{argmax}(p(X|W)) = \text{argmax} \prod_{i=1}^{n} p(x_1|W)$$

$$= \text{argmax} \sum_{i=1}^{n} \log p(x_1|W)$$

Bayesian: valuing $W$ as a distribution via Bayesian Inference

$$W \sim p(W|X,Y) = \frac{\prod_{i=1}^{n} p(Y|X,W)p(W)}{\int \prod_{i=1}^{n} p(Y|X,W)p(W)dW}$$

Bayesian models are more robust to uncertainties while able to avoid overconfidence wrt. samples.

# Bayesian Inference

Training:

$$p(W|X_{tr},Y_{tr}) = \frac{p(Y_{tr}|X_{tr},W)p(W)}{\int p(Y_{tr}|X_{tr},W)p(W)dW}$$

Testing:

$$p(y|x,X_{tr},Y_{tr}) = \int p(y|x,W)p(W|X_{tr},Y_{tr})dW$$

It is impractical to calculate the conditional probability with tremendous datasets which is common in image processing.

Probability model: $p(x,W) = p(x|W)p(W)$

Find a distribution closest to the posterior, $q(W) \approx p(W|x)$

by minimizing **the Kullback–Leibler divergence (relative entropy)** [**]

$$KL(q(W)||p(W|x)) = \int q(W) \log \frac{q(W)}{p(W|x)} dW$$

[**] Wang, Hao, and Dit-Yan Yeung. "Towards Bayesian deep learning: A framework and some existing methods." *IEEE Transactions on Knowledge and Data Engineering* , vol . 28, no. 12, pp. 3395-3408, 2016.

# Bayesian Inference

Evidence lower bound (ELBO):

$$\log p(X) = \int q(W)\log p(X)dW = \int q(W)\log \frac{p(x,W)}{p(W|x)}dW$$

$$= \int q(W)\log \frac{p(x,W)}{q(W)}dW + \int q(W)\log \frac{q(W)}{p(W|x)}dW$$

$$= \mathcal{L}\big(q(W)\big) + KL\big(q(W)||p(W|x)\big)$$

Minimizing relative entropy, ie. KL-divergence
~Maximizing evidence lower bound

$$\mathcal{L}(q(W)) = \int q(W)\log \frac{p(\mathcal{D},W)}{q(W)}dW$$

ELBO $\mathcal{L}\big(q(W)\big) = \int q(W)\log p(x,W)dW + \int q(W)\log \frac{p(W)}{q(W)}dW$

$$= \mathbb{E}_{q(W)}\log p(x|W) - KL(q(W)||p(W))$$

The inference issue is therefore converted to an optimization problem:

$$KL(q(W)||p(W|X) \to \min_{q_\phi(W)\in\mathbb{Q}} \Leftrightarrow L\big(q(W)\big) \to \max_{q(W)\in\mathbb{Q}}$$

[***] Graves, Alex. "Practical variational inference for neural networks." In *Advances in neural information processing systems*, pp. 2348-2356. 2011.

# Bayesian Inference

Introducing a Gaussian parameter \phi into the ELBO

$\emptyset \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma})$

ELBO for Bayesian neural networks:

$$\mathcal{L}\Big(q_\emptyset(W)\Big) = \sum_{i=1}^{N} \mathbb{E}_{q_\emptyset(W)}\log p(Y|X,W) - KL(q_\emptyset(W)||p(W))$$

Reparameterization, $W \sim q(W|\emptyset) \Leftrightarrow W \sim g(\boldsymbol{\epsilon},\emptyset), \epsilon \sim g(\boldsymbol{\epsilon})$

$$\mathcal{L}\Big(q_\emptyset(W)\Big) = \sum_{i=1}^{N} \mathbb{E}_{q_\emptyset(W)}\log p\big(Y|X,W=g(\boldsymbol{\epsilon},\emptyset)\big) - KL(q_\emptyset(W)||p(W))$$

Estimation using a mini-batch:

$$\mathcal{L}\Big(q_\emptyset(W)\Big) \simeq \frac{N}{M}\sum_{i=1}^{N} \log p\left(Y_{m_i}|X_{m_i},W=g(\boldsymbol{\epsilon},\emptyset)\right) - KL(q_\emptyset(W)||p(W))$$

So the proposed frequentist framework has been converted into a Bayesian one by adjusting the weights *W* subject to noise.

# Bayesian Inference

With a Gaussian implementation, the new weights in the same architecture:

$$W = \boldsymbol{\mu} + \boldsymbol{\sigma}\odot\boldsymbol{\epsilon}, \qquad \epsilon_i \sim \mathcal{N}(0,1)$$

Consequently, the current objective function is in the *max-a-posteriori* (MAP) form with L2 regularization:

$$\mathcal{L}(\phi) = \sum_{i=1}^{N} \mathbb{E}_{q_\phi(W)}\log p(Y|X,W)$$

By maximizing this function using the same gradient approach used for frequentist frameworks, we can obtain a more generalized model for crack detection.

$$\nabla_\emptyset \mathcal{L}(W,\emptyset) = \mathbb{E}_{p(\epsilon)}\left[\frac{\partial}{\partial W}\left[\log p(\mathcal{D},W) - \log q_\emptyset(W)\right]\frac{\partial g(\emptyset,\epsilon)}{\partial \emptyset}\right]$$

# Quantitative evaluation

Mean absolute error (MAE) [6]:

$$MAE = \frac{1}{N_{pixel}}\sum_{k=1}^{N_{pixel}} |s_i - y_i|$$

- $N_{pixel}$ represents the the number of pixels in the image.
- $s_i$ represents the pixel values on the detection result.
- $y_i$ represents the pixel values on the ground truth.

Smaller MAE means more accurate detection in matching the ground truth.

[6] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, P. H. Torr, Deeply supervised salient object detection with short connections, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3203-3212.

# Quantitative evaluation

## Q-measure [2]:

$$Q(I) = \frac{1}{10000(m \times n)} \sqrt{N_c} \times \sum_{n=1}^{N_c} \left\{ \frac{e_n^2}{1 + \log A_n} + \left( \frac{N(A_n)}{A_n} \right)^2 \right\} \quad (7)$$

- $m \times n$ -- the size of processing image.
- $N_c$ -- number of classes( being 2 in crack detection).
- $A_n$ -- number of pixels belonging to class $n^{th}$ class.
- $N(A_n)$ -- number of classes having same number of pixels as $n^{th}$ class.

**Smaller Q means better quality of segmentation.**

[2] M. Borsotti, P. Campadelli, and R. Schettini, "Quantitative evaluation of color image segmentation results," *Pattern recognition letters*, vol. 19, no. 8, pp. 741-747,1998.

# Quantitative evaluation

## F-measure[7]:

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

- *Precision* represents the ratio of correctly-labelled crack pixels among all the predicted crack pixels.
- *Recall* represents the ratio of correctly-labelled crack pixels among all the correctly-labelled pixels.

Larger F means more accurate detection in matching the ground truth.

[7] ] L. Zhang, F. Yang, Y. D. Zhang and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proceedings of the 2016 23th IEEE International Conference on Image Processing (ICIP)*, pp. 3708-3712, Sept 2016.

# Comparison with available DL based crack detection algorithms

## Datasets for test:

- CrackForest [1]: 118 pavement crack images
- SYDCrack [2]: 170 infrastructure crack images collected by our UAVs

## Sequential model compared:

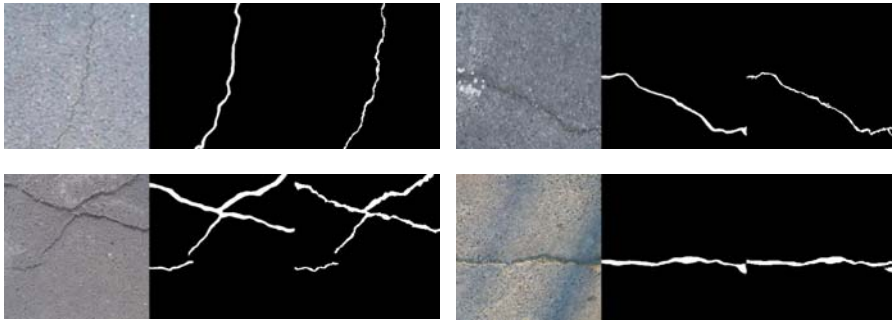
CrackNet-V [3]: a specific sequential convolutional network for crack detection with a customized activation layer.

[1] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 3434-3445, 2016.

[2] Q. Zhu, T. H. Dinh, V. T. Hoang, M. D. Phung, Q. P. Ha, "Crack Detection Using Enhanced Thresholding on UAV based Collected Images," in Proceedings *of the 2018 Australasian Conference of Robotics and Automation*, pp. 1-7, Dec 2018.

[3] Y. Fei, K. C. P. Wang, A. Zhang et al, " Pixel-Level Cracking Detection on 3D Asphalt Pavement Images Through Deep-Learning-Based CrackNet-V," *IEEE Transactions on Intelligent Transportation Systems*, 2019, Early Access.

# Detection results



From left to right: (a) Original image (b) Ground truth (c) Detection result

# Comparison with available DL based crack detection algorithm

| Methods | F-measure | | Q-measure | |
|---|---|---|---|---|
| | CrackForest | SYDCrack | CrackForest | SYDCrack |
| CrackNet-V | 0.6195 | 0.5605 | 2.3679 | 2.5080 |
| Proposed | **0.7807** | **0.7393** | **2.1901** | **2.4588** |

Average F-measure and Q-measure

- The proposed method is expected to provide better quantitative results compared with current sequential deep learning model.

Thank you!

# Q & A