A Versatile Crack Inspection Portable System based on Classifier Ensemble and Controlled Illumination



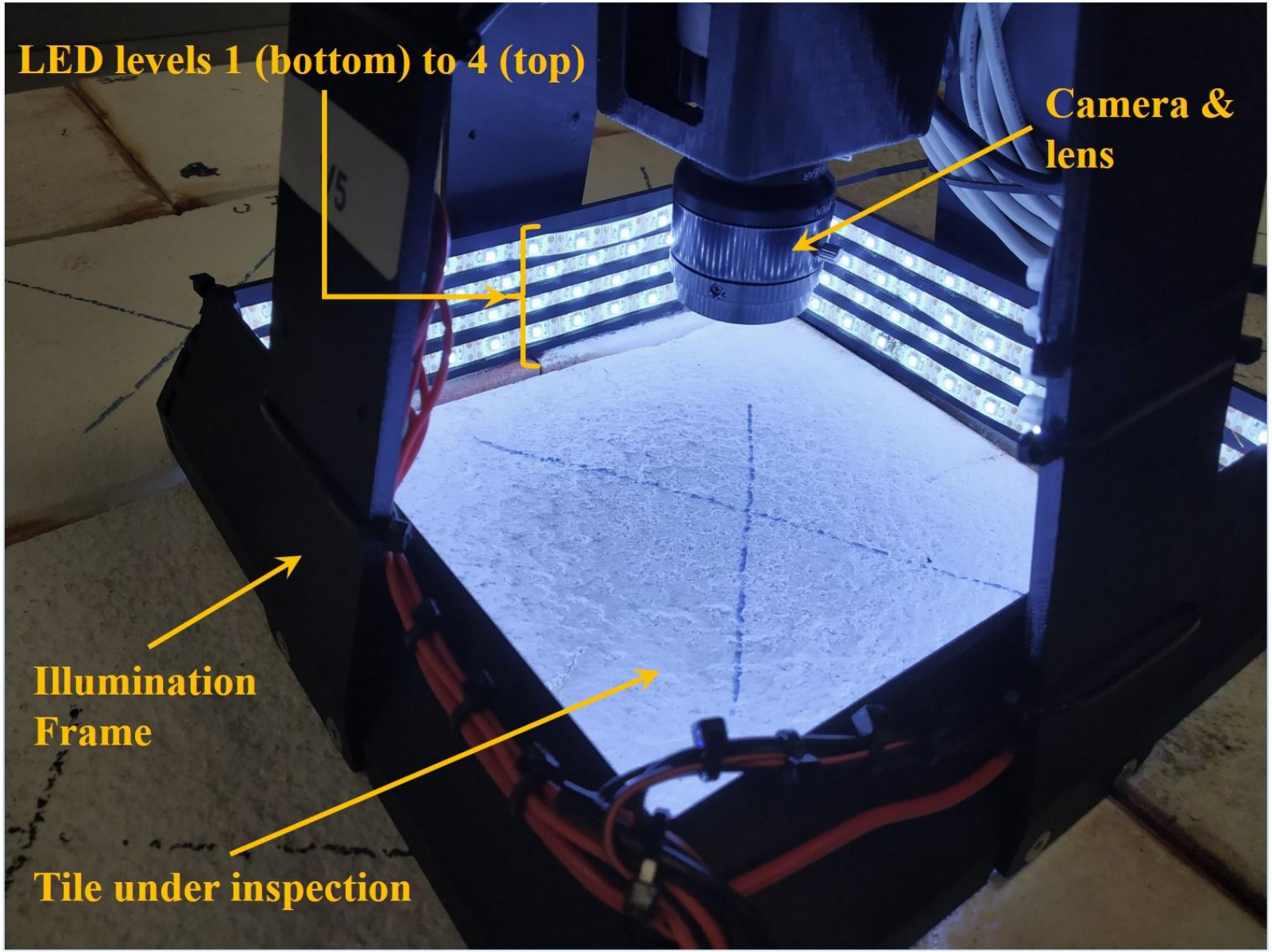


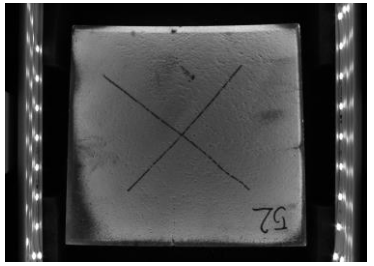
Milind G. Padalkar[1] Carlos Beltrán-González[1] Matteo Bustreo[1,5] Alessio Del Bue[2] Vittorio Murino[1,3,4]

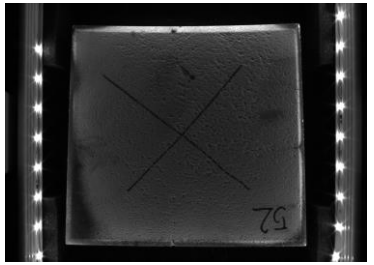
[1]Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Genova, Italy
[2]Visual Geometry and Modelling (VGM), Istituto Italiano di Tecnologia, Genova, Italy
[3]Ireland Research Center, Huawei Technologies Co., Ltd., Dublin, Ireland
[4]Dipartimento di Informatica, University of Verona, Verona, Italy
[5]Dipartimento di Ingegneria Navale, Elettrica, Elettronica e della Telecomunicazioni, University of Genova, Italy

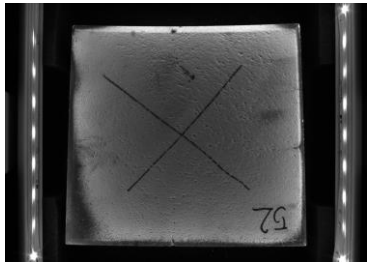
Paper Highlights

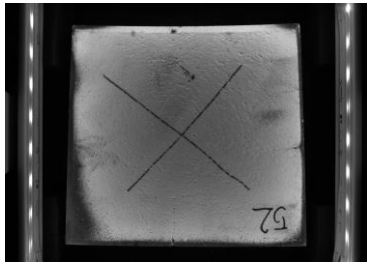
- A novel setup for automatic visual inspection of cracks in ceramic tile
  - Designed for field work with constraints in its maximum dimensions
- Studies the effect of various classifiers and height-varying illumination
  - Classifiers trained on customized as well as state-of-the-art architectures
  - Performance evaluation at patch as well as image-level
- Insights about illumination
  - Which illumination configuration can help in a challenging real-world industrial environment

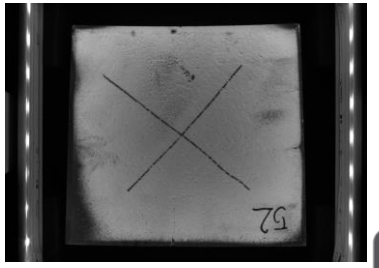


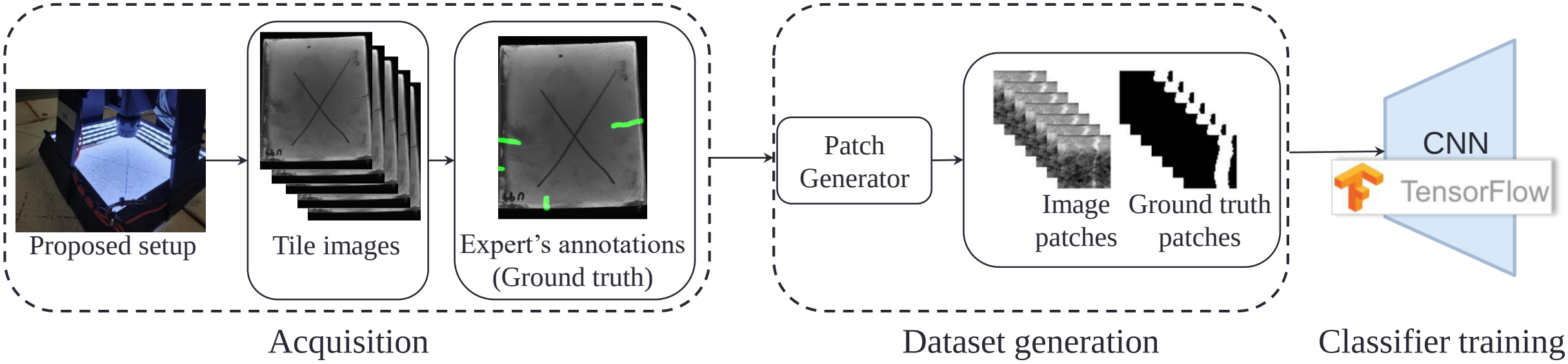

All lights


Lights at level 1

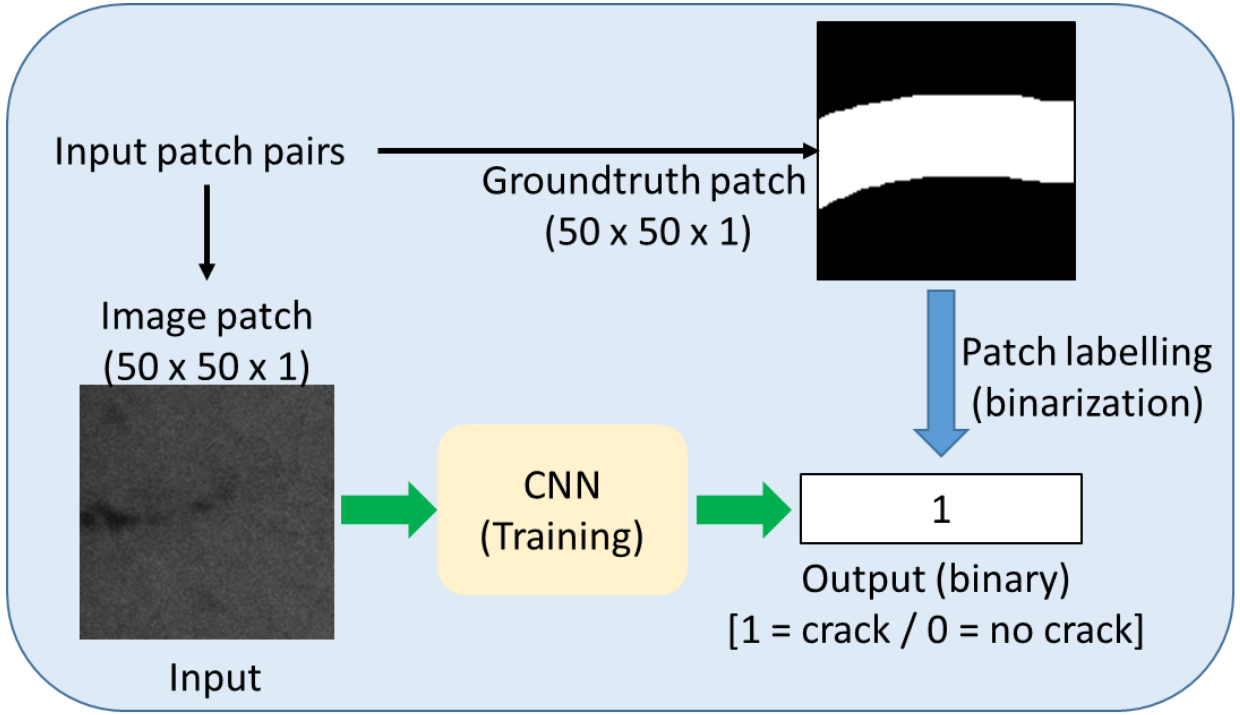

Lights at level 2


Lights at level 3

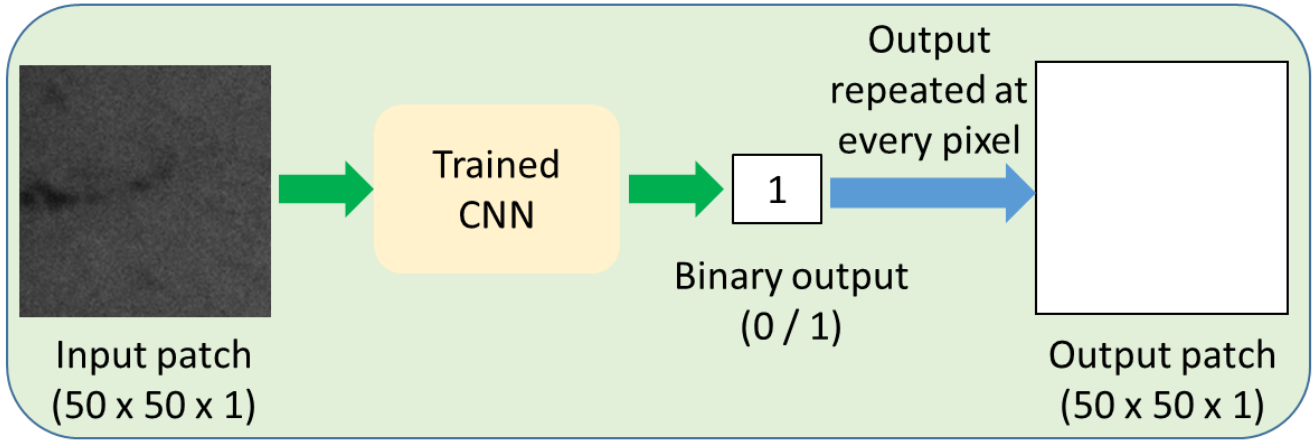

Lights at level 4


Proposed setup
Tile images
Expert's annotations
(Ground truth)
Patch
Generator
Image
patches
Ground truth
patches
CNN
TensorFlow
Acquisition
Dataset generation
Classifier training

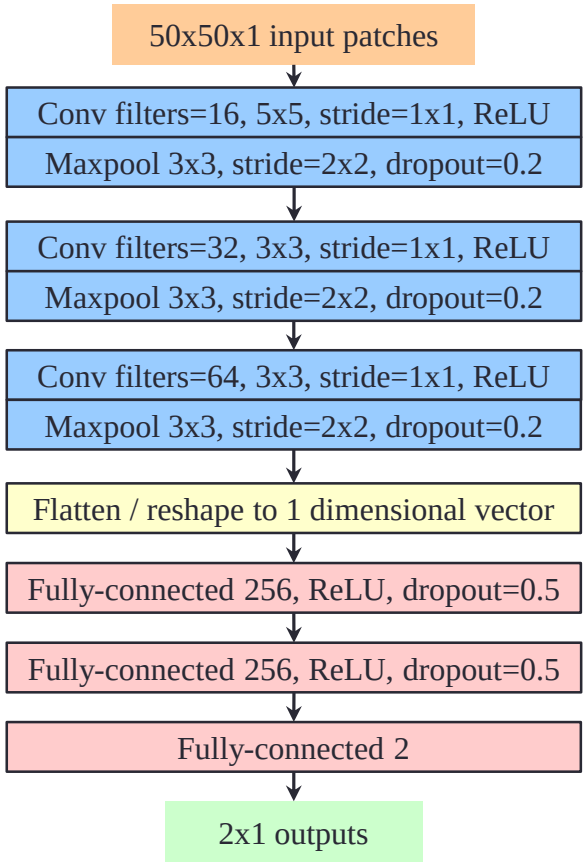
Training



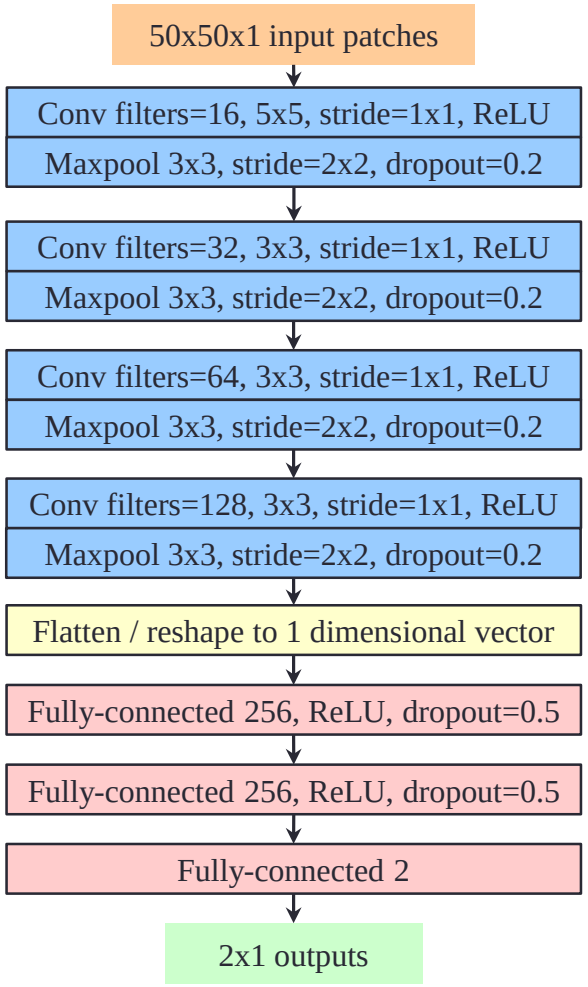
Inference

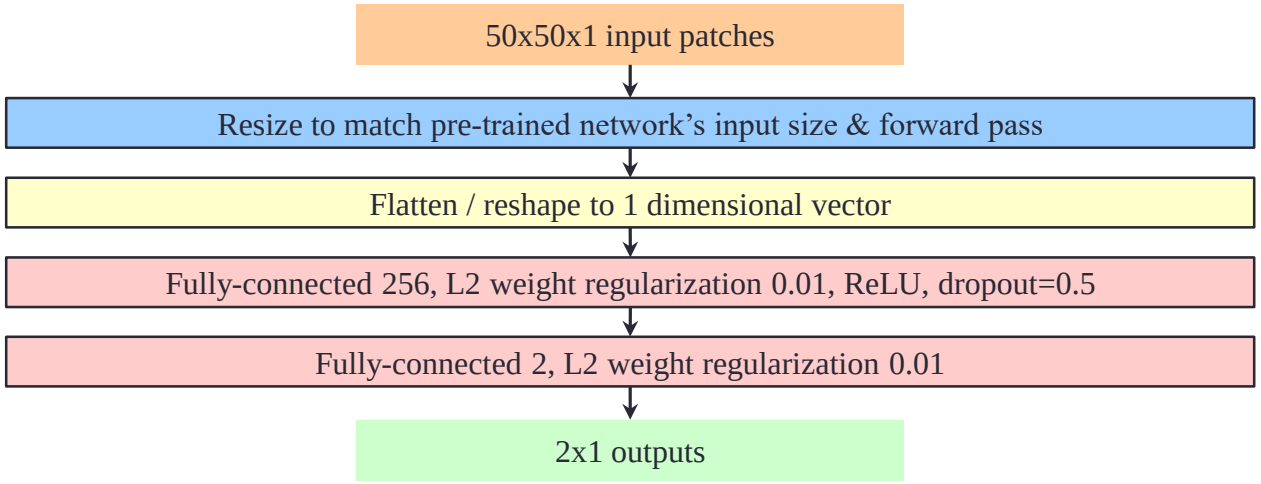


Architectures



TileNet6





Fine-tuning

Architectures used for fine-tuning:

- **VGG16** (Simonyan and A. Zisserman, ICLR 2015)
- **ResNet50** (He et al., CVPR 2016)
- **DenseNet121** (Huang et al., CVPR 2017)
- **InceptionResNet-v2** (Szegedy et al., AAAI 2017)
- **NASNetLarge** (Zoph et al., CVPR 2018 Spotlight)
- **Xception** (Chollet, CVPR 2017)

Patch-level Metrics

- Accuracy

$$\frac{TP + TN}{TP + FP + FN + TN}$$

- Matthew's Correlation Coefficient

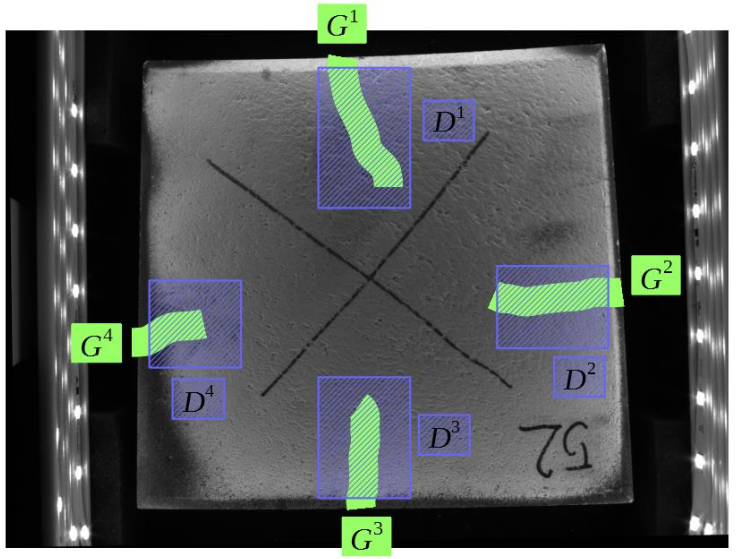
$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

G: Set of cracks (closed contours or connected components) in ground truth

D: Set of detected cracks (closed contours or connected components) in the given image

n(*X*): Number of elements in set *X*, (*X* represents *D*;*G* or any other set of cracks)

I(•) : Indicator function, *I*(*True*) = 1; *I*(*False*) = 0



$$G \cap D = \{(G^1, D^1), (G^2, D^2), (G^3, D^3), (G^4, D^4)\}$$

Image-level Metrics

- Crack Presence Accuracy

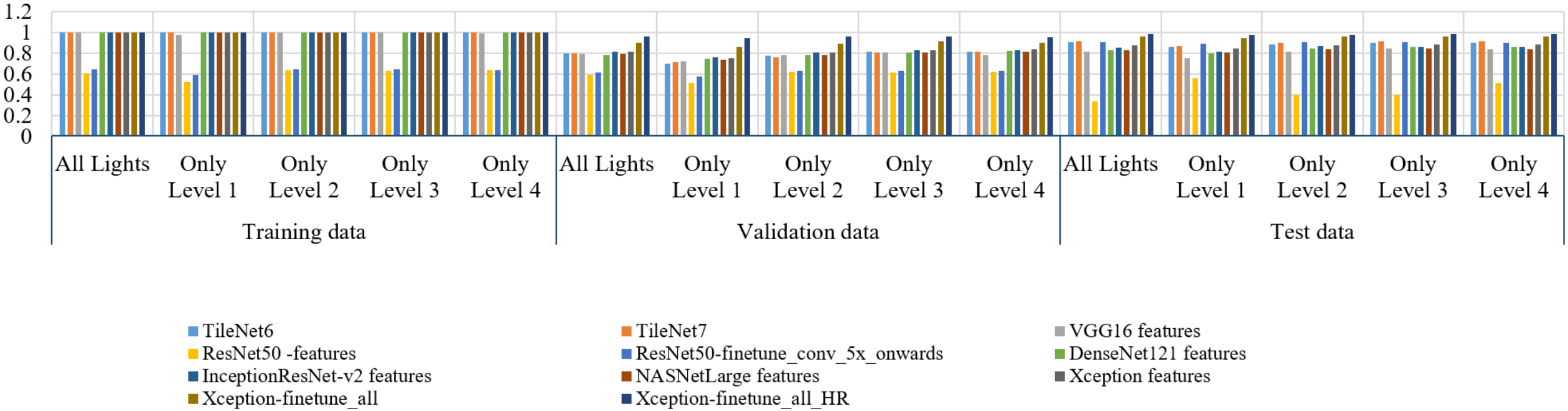
$$CPA = average(PM) = \frac{1}{N}\sum_{t=1}^{N}(PM)_t \quad \text{where} \quad (PM)_t = I(I(n(G_t) > 0) == I(n(D_t) > 0))$$

- Crack Count F1 Score

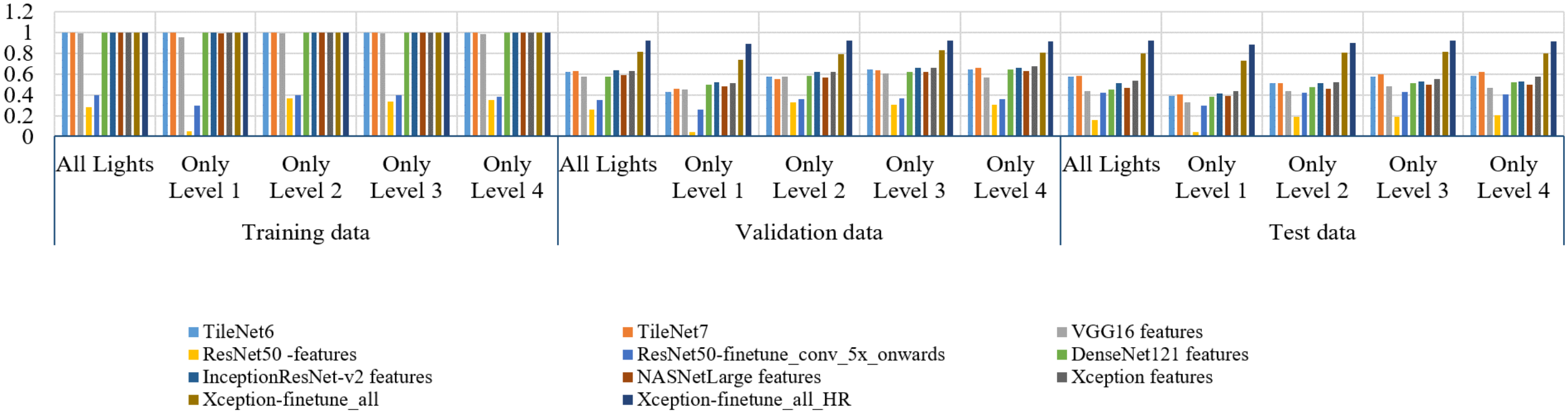
$$CCF1 = \frac{\sum_{t=1}^{N} F1_t * a_t}{\sum_{t=1}^{N} a_t}, \text{ where } a_t = n(G_t) + 1 \qquad F1_t = 2 * \frac{P_t * R_t}{P_t + R_t}$$

$$\text{Precision } (P_t) = \begin{cases} \frac{n(G_t \cap D_t)}{n(D_t)}, & \text{if } n(D_t) > 0, \\ 1, & \text{otherwise,} \end{cases} \qquad \text{Recall } (R_t) = \begin{cases} \frac{n(G_t \cap D_t)}{n(G_t)}, & \text{if } n(G_t) > 0, \\ 1, & \text{otherwise,} \end{cases}$$

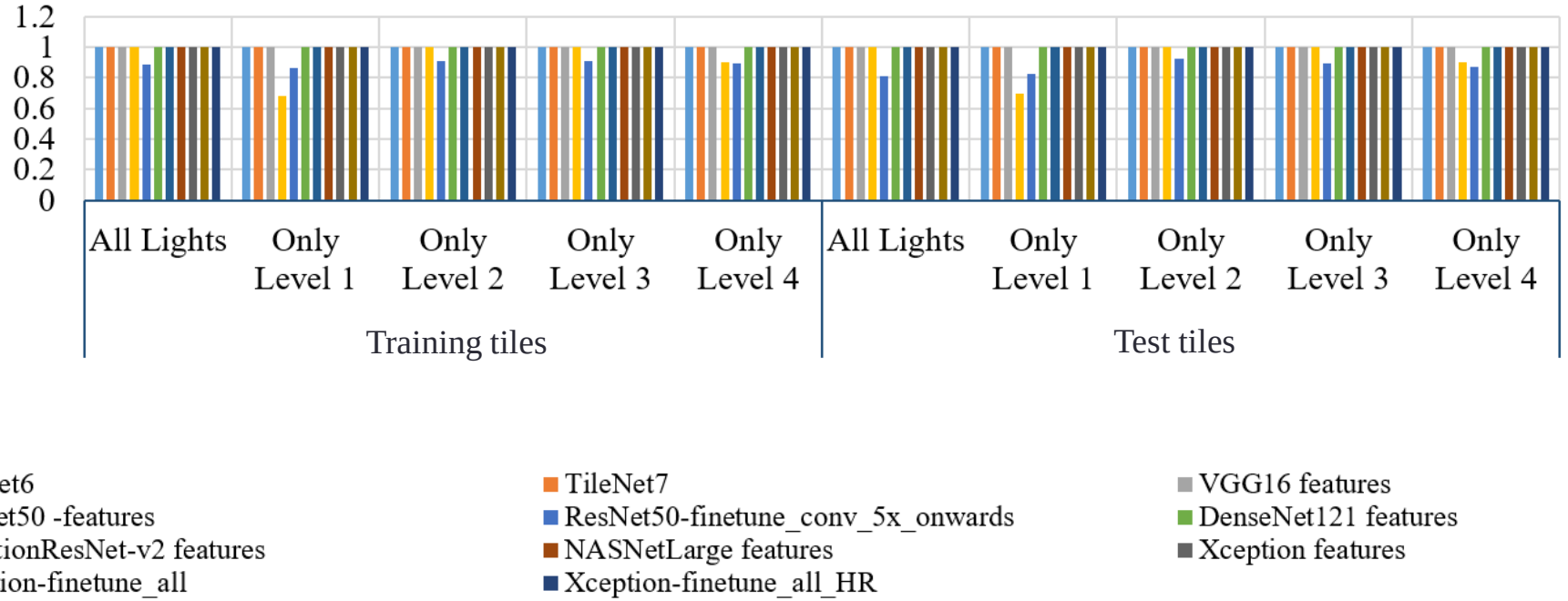
Accuracy


1.2
1
0.8
0.6
0.4
0.2
0
All Lights
Only Level 1
Only Level 2
Only Level 3
Only Level 4
Training data
Validation data
Test data
TileNet6
TileNet7
VGG16 features
ResNet50 -features
ResNet50-finetune_conv_5x_onwards
DenseNet121 features
InceptionResNet-v2 features
NASNetLarge features
Xception features
Xception-finetune_all
Xception-finetune_all_HR

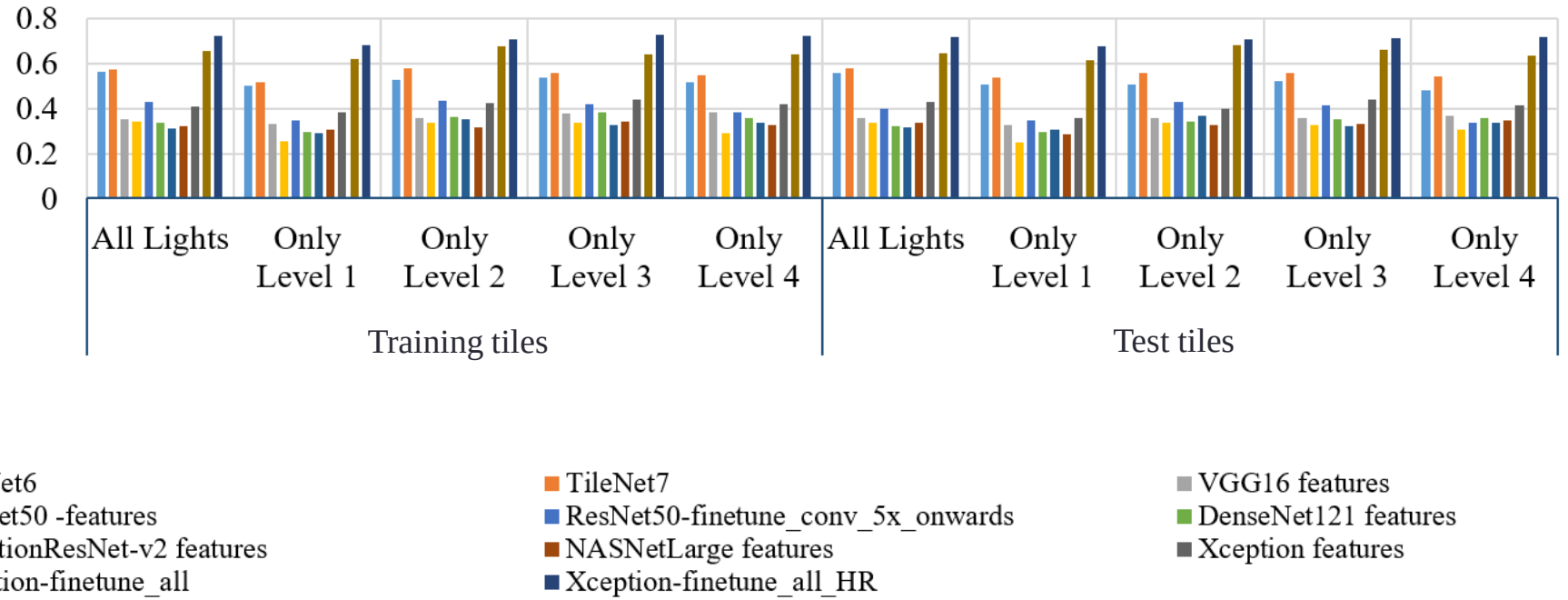
MCC


1.2
1
0.8
0.6
0.4
0.2
0
All Lights
Only Level 1
Only Level 2
Only Level 3
Only Level 4
Training data
Validation data
Test data
TileNet6
TileNet7
VGG16 features
ResNet50 -features
ResNet50-finetune_conv_5x_onwards
DenseNet121 features
InceptionResNet-v2 features
NASNetLarge features
Xception features
Xception-finetune_all
Xception-finetune_all_HR

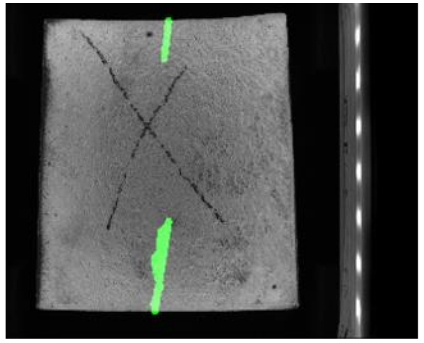
Crack Presence Accuracy



Crack Count F1 Score

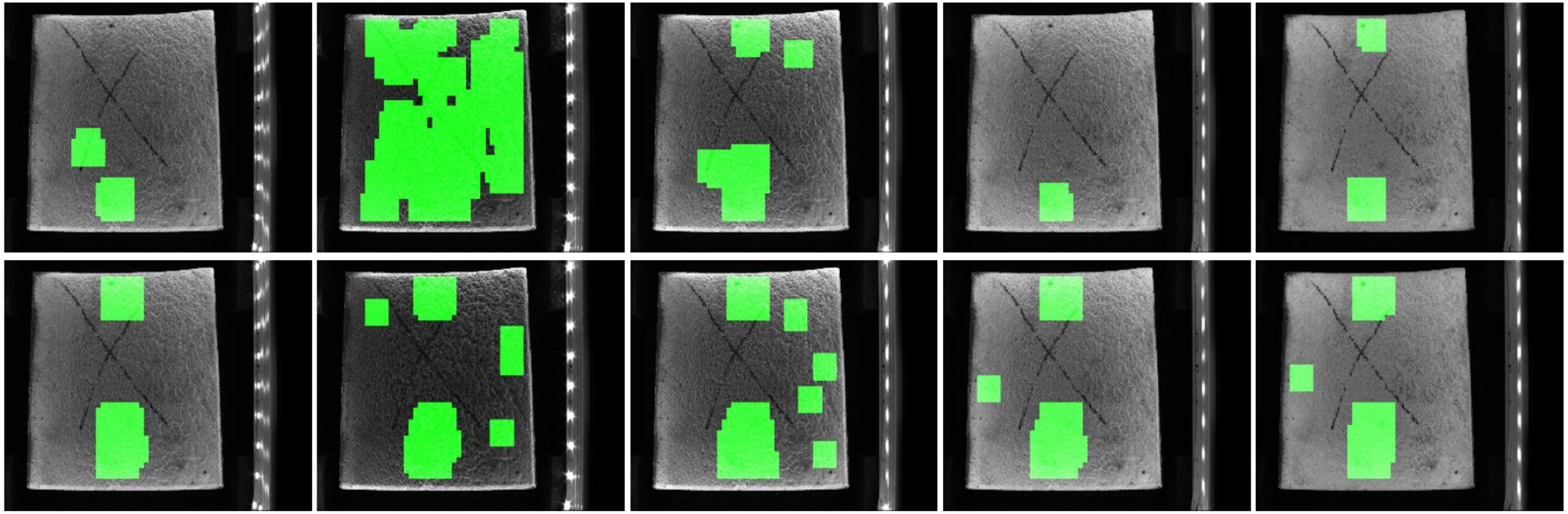


Ground truth



Low-resolution

High-resolution



All lights | Only Level 1 | Only Level 2 | Only Level 3 | Only Level 4

Conclusions

- Lights
  - All lights configuration provides the best results
  - Lights placed at greater heights more effective than those placed near the tile's surface
- Architecture
  - Increasing depth of the network improves the results
  - Fine-tuning pre-trained weights of the Xception architecture provided the best results
- Spatial resolution
  - Use of high-resolution patches improves the results compared to low-resolution
  - Study should help in deciding the resolution versus performance trade-off for field use