

# Virtual CSA-Net : 가상 깊이 생성 및 교차 융합 기반의 콘크리트 균열 검출

박우성, 김남수

서울대학교 전기정보공학부 뉴미디어통신공동연구소 휴먼인터페이스 연구실

wsp156@snu.ac.kr, nkim@snu.ac.kr

## Virtual CSA-Net: Concrete Crack Detection based on Virtual Depth Generation and Cross-Fusion

Woo Seong Park, Nam Soo Kim

Human Interface Laboratory

Department of Electrical and Computer Engineering and INMC,

Seoul National Univ.

### 요 약

콘크리트 구조물의 안전 진단을 위한 딥러닝 기반 균열 검출 기술은 조명 변화나 오염 등 환경적 요인으로 인해 RGB 이미지만으로는 미세 균열 검출에 한계를 보인다. 깊이(Depth)정보를 활용하면 이를 개선할 수 있으나, 고가의 센서가 필요하다는 단점이 있다. 이에 본 연구에서는 학습 단계에서 생성 모델을 통해 RGB로부터 고품질 가상 깊이(Virtual Depth)를 생성하고, 이를 분할 모델과 결합하는 Virtual CSA-Net을 제안한다. 제안 모델은 추론 시 추가 센서 없이도 깊이 정보의 이점을 활용할 수 있다. 실험 결과, mIoU 89.58%를 달성하여 기본 U-Net(88.93%) 대비 우수한 성능을 입증하였다.

### I. 서 론

도로 및 건물과 같은 인프라의 노후화에 따라 균열을 조기에 탐지하는 작업이 필수적이다. 기존의 육안 검사는 비효율적이며, 이를 대체하기 위해 U-Net[1]과 같은 합성곱 신경망(CNN)을 이용한 픽셀 단위의 균열 분할(Segmentation) 기법들이 연구되어 왔다. 그러나 RGB 이미지는 조명 변화, 그림자, 물 자국 등 노이즈에 취약하여, 균열과 배경을 명확히 구분하지 못하는 경우가 빈번하다.

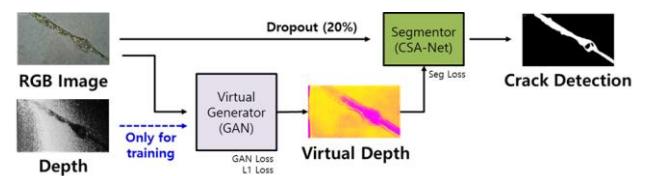
깊이(Depth) 정보는 이러한 광학적 노이즈에 강인한 특성이 있어 균열 검출 성능을 비약적으로 높일 수 있다. 하지만 모든 운영 환경에서 깊이 센서를 운용하는 것은 비용 및 장비 제약상 현실적으로 어렵다.

따라서 본 연구에서는 제한된 깊이 데이터셋으로 학습된 생성기를 통해, 추론 시에는 RGB 이미지만으로 깊이 정보를 복원하고 이를 분할 성능 극대화에 활용하는 Virtual CSA-Net을 제안한다. 본 연구는 깊이 맵을 추정하여 생성하는 적대적 학습(GAN)기반의 가상 깊이 생성기와 교차 융합(Cross-Fusion)을 활용한 CSA-Net[2] 구조를 결합하여 RGB 기반 모델의 한계를 극복하는 것을 목표로 한다.

### II. 본론

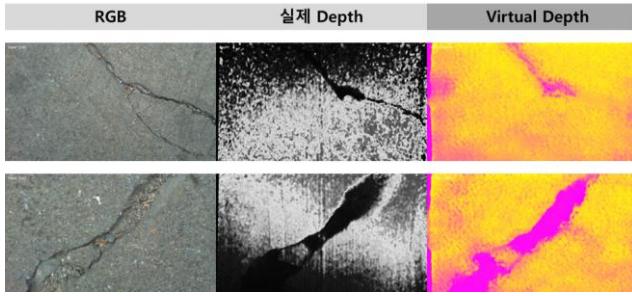
#### i. 제안 방법

본 연구에서 제안하는 Virtual CSA-Net은 크게 가상 생성기(Virtual Generator)와 분할 모델(Segmentor)로 구성되며, 전체 구조는 [그림 1]과 같다.



[그림 1] 전체 모델 구조

먼저, 적대적 학습 기반의 가상 깊이 생성(Virtual Depth Generation) 모듈은 Pix2Pix[3] 구조를 기반으로 설계되었다. 단순한 L1 Loss만 사용할 경우 생성된 깊이 맵이 흐릿해지는 문제가 발생하므로, 본 연구에서는 생성기(G)와 판별자(D) 간의 경쟁적 학습(Adversarial Training)을 도입하였다. 학습 단계에서는 RGB 이미지와 실제 깊이 쌍을 이용해 판별자를 속이도록 생성기를 훈련시키지만, 학습이 완료된 추론 단계에서는 깊이 센서 없이 RGB 이미지만을 입력 받아 가상 깊이를 생성한다. [그림 2]는 생성된 가상 깊이의 예시로, 균열 부위의 깊이 정보가 선명하게 복원됨을 확인할 수 있다.



[그림 2] 생성된 Virtual Depth 시각화

생성된 가상 깊이는 CSA-Net 기반의 분할 모델로 입력되어 RGB 특징과 융합된다. 이 과정에서 모델이 선명한 RGB 정보에만 의존하는 문제를 방지하기 위해, 본 연구에서는 RGB 드롭아웃(RGB Dropout) 훈련 전략을 적용하였다. 학습 중 20%의 확률로 RGB 입력을 차단함으로써, 모델이 RGB 정보가 없는 상황에서도 가상 깊이 정보만을 단서로 균열을 검출하도록 강제하여 깊이 정보의 활용도를 극대화하였다.

## ii. 실험 및 결과

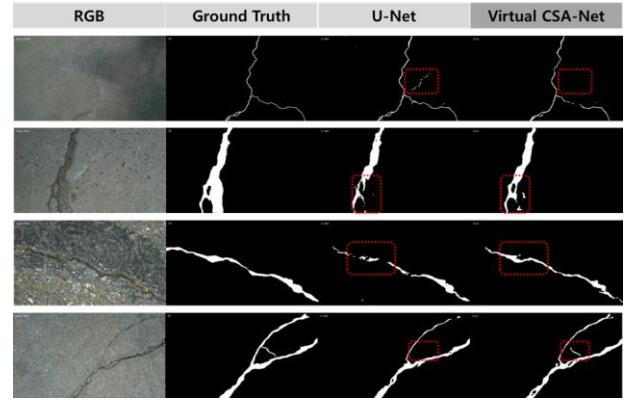
본 실험은 Crack RGBD 데이터셋을 사용하여 진행되었으며, 성능 검증을 위해 RGB 이미지만을 사용하는 U-Net과 FCN[4], 실제 깊이 데이터도 입력으로 사용하는 CSA-Net을 비교군으로 설정하였다. 정량적 성능 지표인 mIoU(mean Intersection over Union) 측정 결과와 정성적 시각화 결과는 [표 1]과 [그림 3]에 제시되어 있다.

| Model                         | Input       | Precision    | Recall       | mIoU (%)     |
|-------------------------------|-------------|--------------|--------------|--------------|
| U-Net                         | RGB         | 91.34        | 84.83        | 88.93        |
| FCN(ResNet-18)                | RGB         | 83.95        | 63.8         | 77.75        |
| <b>Virtual CSA-Net (Ours)</b> | <b>RGB</b>  | <b>91.39</b> | <b>86.27</b> | <b>89.58</b> |
| CSA-Net                       | RGB + Depth | 91.55        | 87.12        | 90.02        |

[표 1] 실험 결과

실험 결과, 제안하는 Virtual CSA-Net은 mIoU 89.58%를 기록하여 U-Net(88.93%) 대비 유의미한 성능 향상을 달성하였다. 제안 모델은 추론 시 RGB만 사용했음에도 불구하고 가상 깊이 정보를 통해 한계를 극복하였으며, 추론시에도 고가의 센서를 사용한 CSA-Net(90.02%)과의 격차를 최소화하였다.

이러한 수치적 향상은 [그림 3]의 시각화 결과에서 더욱 명확히 확인된다. U-Net은 짙은 그림자나 아스팔트의 얼룩을 균열로 오인하거나, 미세하게 끊어진 균열을 감지하지 못하는 한계를 보였다. 그러나 제안 모델은 가상 깊이 맵(Virtual Depth)을 통해 기하학적 깊이 특징을 인지함으로써, 비균열 노이즈를 효과적으로 제거하고 끊어진 균열을 온전하게 연결하여 검출하는 우수한 성능을 입증하였다.



[그림 3] RGB 모델별 예측 결과 비교

## III. 결론

본 논문에서는 추론 시 추가적인 센서 없이 RGB 이미지만으로 고성능 균열 검출이 가능한 Virtual CSA-Net을 제안했다. 학습 단계에서만 깊이 정보를 활용하고 추론 시에는 RGB만을 사용하는 효율적인 구조를 통해, 비용 효율성과 검출 성능을 동시에 확보하였다. 특히 GAN 기반의 깊이 생성과 RGB Dropout 학습 전략을 통해 기존 RGB 기반 모델의 한계를 넘어섰음을 실험적으로 증명하였다. 본 연구는 고가의 장비 없이도 인프라 정밀 안전 진단이 가능함을 시사하며, 향후 모바일 기기 등에서의 실시간 검출 연구로 확장될 수 있다.

## 참 고 문 헌

- [1] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing, 2015.
- [2] Hussain, Tasneem, et al. "Pixel-level crack segmentation and quantification enabled by multi-modality cross-fusion of RGB and depth images." Construction and Building Materials 487 (2025): 141961.
- [3] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [4] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.