

얕은 CNN 층과 Transformer의 결합을 통한 균열 Segmentation 성능 향상에 관한 연구

A Study on Performance Enhancing of Crack Segmentation through the Integration of Shallow CNN Layers and Transformer

SFAI Lab

Dept. of Artificial Intelligence Engineering, Chosun University, Korea 문지환 2023.11.17



Contents

- I. Introduction
- II. Related Work
- III. Proposed PCTC-Net
- IV. Experimental Environment
- V. Experiment and Results
- VI. Conclusion

I. Introduction

1. 균열이란?

- 균열은 일반적으로 구조물에 심각한 손상과 결과를 초래할 수 있
- 작은 균열도 더 크고 더 위협적인 결함으로 변할 수 있음
- 안전과 유지관리를 위해 균열은 초기에 발견하고 조치하는 것이



2. 균열 탐지

- 균열 탐지의 종류에는 크게 고전적인 균열 탐지, 이미지 기반 균열 탐지, 딥러닝 기반 균열 탐지가 있음
- 고전적인 균열 탐지 : 인력을 사용하는 균열 탐지
- 이미지 기반 균열 탐지 : 픽셀 강도의 급격한 변화를 포착하여 균열 탐지
- 딥러닝 기반 균열 탐지: 컴퓨터 비전과 인공지능 기술을 이용한 균열 탐지

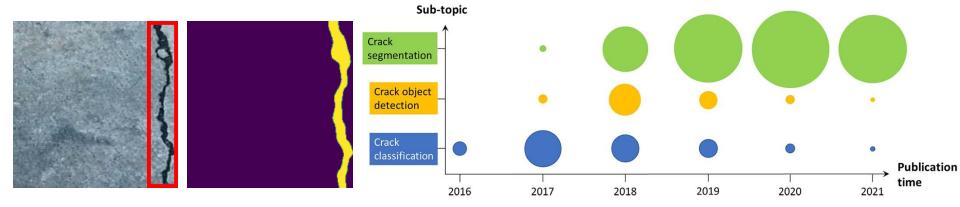


- (a) block cracks
- (b) edge cracks
- (c) reflection cracks
- (d) longitudinal cracks
- (e) fatigue cracks
- (f) transverse cracks

I. Introduction

3. 딥러닝 기반 균열 탐지의 종류

- 딥러닝 기반 균열 탐지는 Classification, Object Detection, Segmentation 3가지로 나눌 수
 있음
- Classification : 균열이 있음, 균열이 없음 두 가지로 분류
- Object Detection : 균열을 Bounding Box로 그려 균열의 위치를 파악
- Segmentation : 균열을 픽셀 단위로 Classification하여 길이와 너비 같은 상세한 특징 파악
- 균열 탐지 분야에서 Segmentation 기법에 대한 연구가 증가하는 추세



(a) Object Detection (b) Segmentation Comparison of different crack recognition tasks in computer vison

Distribution of journal publications in deep learning-based crack detection[1]



1. 균열 Segmentation 모델의 종류

- 1) CNN 기반 Segmentation 모델
 - \rightarrow FCN(2014), U-Net(2015)
- 2) Transformer 기반 Segmentation 모델
 - \rightarrow CrackFormer(2021)
- 3) CNN + Transformer 기반 Segmentation 모델
 - \rightarrow SegCrack(2022), DTrC(2023)

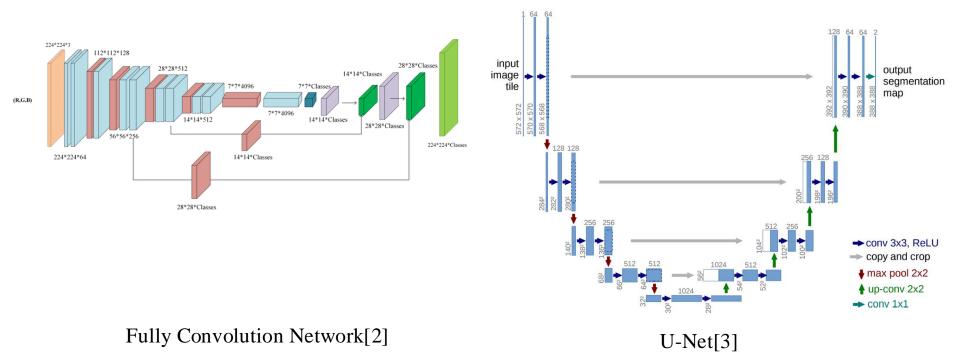
Table I. Crack segmentation model overview

| Year | Model | Architecture | Methodology | Dataset |
|------|-------------|-------------------|-------------------|---------------------|
| 2014 | FCN | CNN | Skip architecture | PASCAL VOC |
| 2015 | U-Net | CNN | Skip connection | ISBI Challenge |
| 2021 | CrackFormer | Transformer | Self Attention | CrackTree260 |
| 2022 | SegCrack | CNN + Transformer | Pyramid | Crack500, DeepCrack |
| 2023 | DTrC | CNN + Transformer | FFM, RPM | Crack709, Crack919 |



2. CNN 기반 Segmentation 모델

- 일반적인 CNN Segmentation모델은 일정한 Scale에서만 특징을 학습해 다양한 크기의 균열을 탐지하지 못함
- 따라서, Multi Scale 특징을 추출하기위해 U-Net, FCN 등의 Encoder-Decoder 구조로 발전함
- 그러나 여전히 Local receptive field에 기반하여 전역적인 특징을 파악하기 어려움

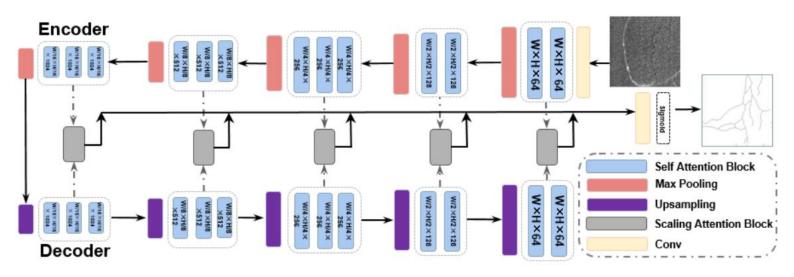


Ref. [2]: Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. Ref. [3]: Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedi- cal image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer 2015



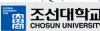
2. Transformer 기반 Segmentation 모델

- 긴 균열과 같이 전역적인 특징을 추출해야 하는 경우 Transformer가 적합
- 따라서, 전역적인 특징을 추출하기위해 CrackFormer[4]에서는 Self Attention Block과
 Scaling Attention Block으로 구성된 Encoder Decoder 구조 사용
- 그러나 미세한 균열과 같이 **지역적인 특징**을 추출하기 복잡함



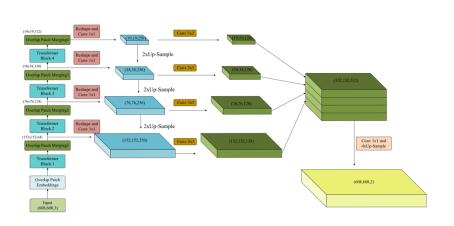
CrackFormer[4]

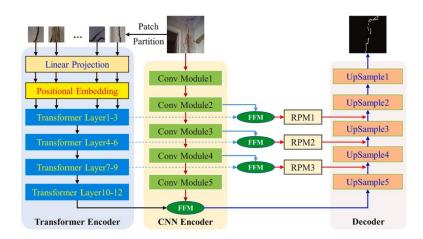
Ref. [4]: H. Liu, X. Miao, C. Mertz, C. Xu and H. Kong, "CrackFormer: Transformer Network for Fine-Grained Crack Detection," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 3763-3772, 2021.



3. CNN + Transformer 기반 Segmentation 모델

- DTrC-Net, SegCrack등의 연구에서 전역적인 특징을 추출하는데 적합한 Transformer를 지역적인 특징을 추출하는 CNN과 결합하는 균열 탐지에 적용함
- 미세하거나 긴 균열 등 복잡한 균열 탐지 성능이 개선됨
- 그러나 Transformer의 낮은 Inductive bias로 인해 방대한 데이터셋이 필요함





SegCrack[5]

DTrC-Net[6]

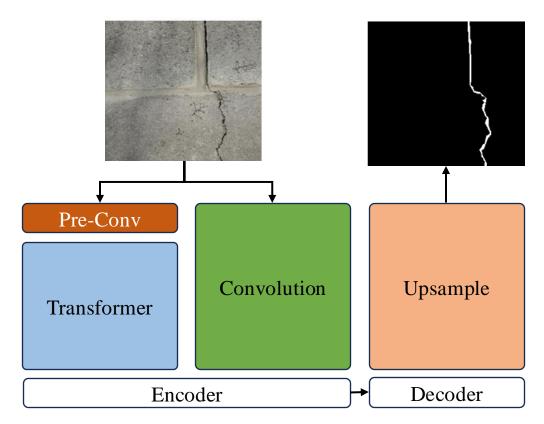
Ref. [5]: Wenjun Wang, Chao Su, "Automatic concrete crack segmentation model based on transformer", Automation in Construction, Volume 139, 2022.

Ref. [6]: Chao Xiang, Jingjing Guo, Ran Cao, Lu Deng, "A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex". Automation in Construction, Volume 152, 2023.



1. Pre Convolution Transformer CNN(PCTC-Net)

- Pre-Conv, Transformer Encoder, CNN Encoder, CNN decoder 로 구성
- Transformer의 낮은 Inductive bias를 개선하기 위한 Pre-Conv를 제안

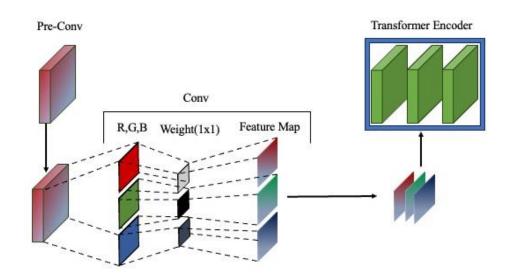


The overall architecture of PCTC-Net



2. Pre-Conv

- 입력 이미지의 R,G,B 총 3개의 채널을 받아 각 채널에 대응하는 1x1 커널과 합성곱 연산 수행
- 추출된 3개의 Feature Map을 하나의 이미지로 변환 후 Transformer Encoder에 입력

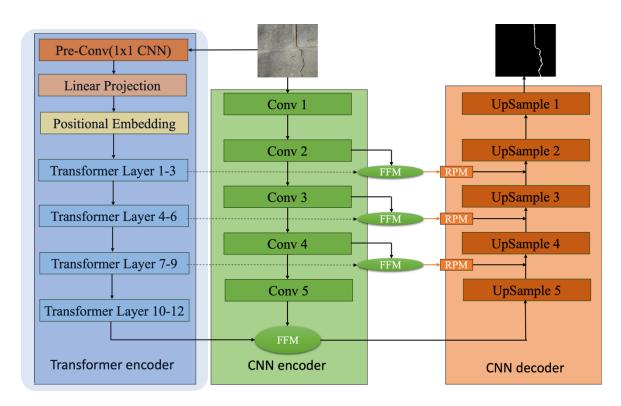


The architecture of Pre-Conv



3. Transformer Encoder

- 균열의 **전역적인 특징**들을 잡기 위한 모듈로 사용함
- 3개의 Transformer layer가 포함된 Transformer Block 각 4개로 구성

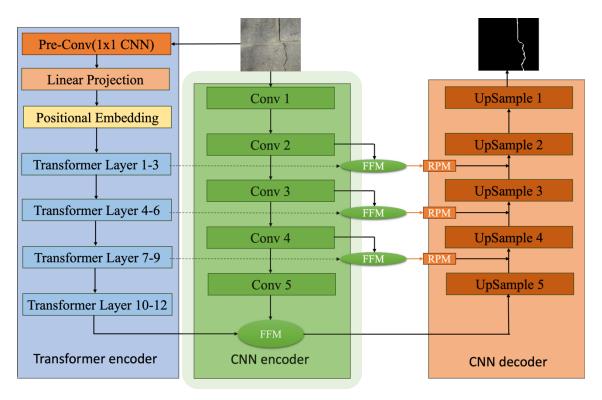


The architecture of PCTC-Net



4. CNN Encoder

- 균열의 지역적인 특징들을 잡기 위한 모듈로 사용함
- 총 5개의 CNN Block으로 구성되며 Backbone은 Resnet-34을 사용하여 배경 정보보다
 균열의 상세한 정보에 집중되도록 함

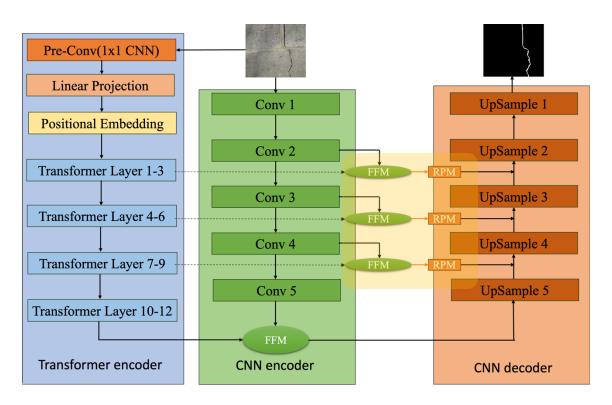


The architecture of PCTC-Net



5. Feature Fusion Module(FFM), Residual Path Module(RPM)

- FFM : Transformer Encoder와 CNN Encoder에서 추출된 정보의 취합을 위한 모듈
- RPM : Transformer Encoder와 CNN Encoder 간의 의미적 차이를 개선하기 위한 모듈

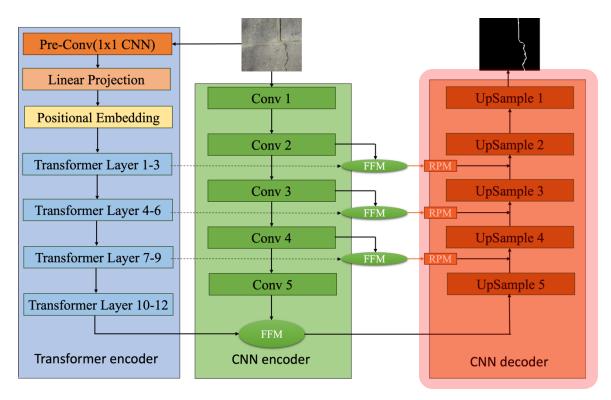


The architecture of PCTC-Net



6. CNN Decoder

- UpSampling을 통해 목표 해상도로 복원하기 위한 모듈
- 5개의 UpSample Block으로 구성되며, 각각의 UpSample Block은 순서대로 Conv + BN + ReLU 및 ConvTranspose2d + BN + ReLU, Conv + BN + ReLU 로 구성



The architecture of PCTC-Net



IV. Experimental Environment

1. DeepCrack

- 300개의 Train data, 239개의 Test data로 구성
- 다양한 유형의 도로 균열 이미지
- 배경이 매끄러워 균열의 길이, 너비와 같은 특징 학습이 단순함

2. Crack 500

- 2,000 x 1,500 픽셀의 도로 균열 이미지 500개
- Crop을 통해 1,896개의 Train data, 1,124개의 Test data로 분리
- 높은 해상도의 이미지로 구성
- 상대적으로 다른 데이터셋에 비해 배경이 복잡

3. Crack9k

- 400 x 400 픽셀의 다양한 균열 이미지 9,255개
- CrackTree, DeepCrack 등 10가지의 데이터 셋이 합쳐진 데이터 셋



IV. Experimental Environment



DeepCrack, Crack500, Crack9k dataset images



IV. Experimental Environment

4. Data Augmentation

- Flip, Horizontal Flip, Rotate90, RandomRotate
- BrightnessContrast, Randomgamma
- 256 x 256 pixel

5. GPU

RTX 4090

6. 하이퍼파라미터 설정

- Batch 3, Epoch 300, Loss Function 0.5 Dice + 0.5 Focal
- Learning Rate : StepLR(0.001, gamma = 0.1)

7. 평가지표

- Precision= $\frac{TP}{TP+FP}$
- Recall = $\frac{TP}{TP+FN}$
- F1-score = $2 \times \frac{Precision \times Recall}{Precision + Recall}$

- 1. Pre-Conv Size에 따른 실험 결과 및 분석
 - 1x1 일 때 F1-score 기준 최대 1.2% 성능 상승함
 - Pre-Conv가 영향을 끼치는 Transformer Encoder는 균열의 전역적인 특성을 추출하기 위해 설계됨
 - 3 x 3, 5 x 5 커널의 Pre-Conv를 사용하면 Inductive bias가 CNN의 지역적인 특성을 추출하는 역할을 수행하기 때문에 Transformer Encoder의 전역적인 특성 추출 성능에 영향을 미칠 수 있음
 - 256 x 256 사이즈의 이미지에서 3 x 3, 5 x 5 커널은 미세한 균열의 특징을 blur 처리하는 경우가 생김

Table I. The results of proposed model according to pre-conv size

| Pre-conv size | Precision (%) | Recall (%) | F1-Score (%) |
|---------------|---------------|------------|---------------------|
| × | 85.46 | 87.16 | 86.11 |
| 1×1 | 86.62 | 88.01 | 87.31 |
| 3×3 | 85.48 | 87.68 | 86.57 |
| 5×5 | 86.53 | 87.59 | 87.06 |

3. Transformer layer 절제 실험 결과 및 분석

- DTrC는 4개의 layer를 줄였을 때 F1-score 기준 약 3.2%의 성능 하락
- PCTC는 8개의 layer를 줄였음에도 F1-score 기준 약 0.9%의 성능 하락
- Precision 기준으로 PCTC가 DTrC보다 <u>약 3% 낮지만</u> Recall 기준으로 <u>약 8% 높음</u>
- 균열 탐지 Task에서는 안전성을 우선순위로 두기 때문에 Recall 수치가 중요
- 트랜스포머의 낮은 Inductive bias를 개선
- 모델의 경량화 가능성을 입증

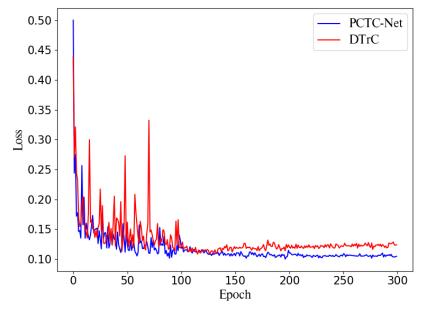
Table II. The results of proposed model according to transformer layer

| Model | Transformer | Precision (%) | Recall (%) | F1-Score (%) |
|-------|-------------|---------------|------------|--------------|
| | 4 layer | 86.83 | 85.99 | 86.41 |
| PCTC | 8 layer | 86.81 | 86.82 | 86.82 |
| | 12 layer | 86.61 | 88.03 | 87.31 |
| | 4 layer | 89.82 | 78.96 | 84.04 |
| DTrC | 8 layer | 89.39 | 78.36 | 83.51 |
| | 12 layer | 87.23 | 86.27 | 86.75 |



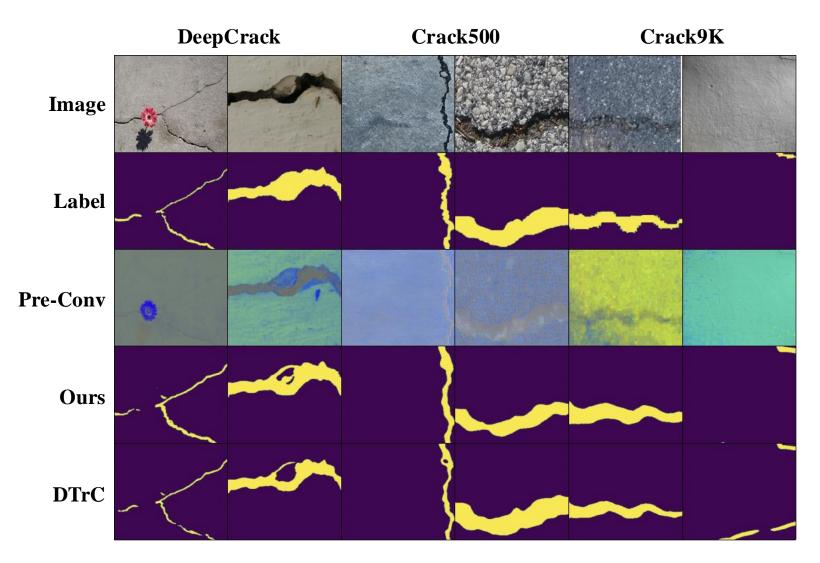
5. Loss 그래프 결과 및 분석

- DTrC 모델 분석: 불안정한 Loss 피크와 대략 130 Epoch을 기준으로 과적합되는 경향을 나타냄
- PCTC 모델 분석 : 불안정한 Loss 피크가 상대적으로 적고 Loss 값이 수렴하는 모습을 나타냄
- Pre-Conv가 모델의 **안정적인 학습**에 기여
- Transformer Encoder의 학습 수렴 속도가 상승



Comparison loss of PCTC-Net and DTrC





Comparison results from DeepCrack, Crack500, Crack9k using PCTC and DTrC



VI. Conclusion

- ▶ 본 논문에서는 Transformer Encoder의 Inductive Bias를 개선하기 위해 Pre-Conv 레이어를 결합한 PCTC-Net을 제안하였음
 - ✓ 3x3 Pre-Conv는 Transformer의 전역적인 특성 추출 성능을 낮추는데 반해 1x1 Pre-Conv는 균열의 미세한 경계 부분의 특징을 잃지 않는 것과 동시에 전역적인 특성 추출성능을 개선함
 - ✓ Pre-Conv를 통해 모델의 경량화 가능성을 입증함
 - ✓ 또한 Transformer의 학습 수렴 속도 개선함

Future Work

- ➤ Transformer Encoder의 전역적인 특성 추출 성능을 위한 Pre-Conv의 확장 및 개선
- ➤ Transformer Encoder와 CNN Encoder에서 추출된 정보를 효과적으로 취합하기 위한 FFM과 RPM의 개선



Thank you for listening

Q&A

