

도로 노면 파손 영상의 다중 분류 심층 신경망 평가를 통한 Backbone Network 선정 기법

A Selection Method of Backbone Network through Multi-Classification Deep Neural Network Evaluation of Road Surface Damage Images

심 승 보* · 송 영 은**

* 주저자 : 한국건설기술연구원 전임연구원

** 교신저자 : 호서대학교 조교수

Seungbo Shim* · Young Eun Song**

* Korea Institute of Civil Engineering and Building Technology

** Hoseo University

† Corresponding author : Young Eun Song, tdsong@hoseo.edu

Vol.18 No.3(2019)

June, 2019

pp.106~118

pISSN 1738-0774

eISSN 2384-1729

<https://doi.org/10.12815/kits.2019.18.3.106>

2019.18.3.106

요 약

최근 들어 인공지능을 이용한 영상 객체 인식에 대한 연구 및 개발이 활발하게 진행되고 있다. 그 연장선상에서 도로 유지 및 관리 분야에도 관련 연구의 활용도가 크게 향상될 것으로 기대된다. 그 중에서도 특히 도로 노면 파손 객체 인식 (Object Detection) 을 위한 인공 지능 모델이 지속적으로 개발되고 있다. 이러한 객체 인식 알고리즘을 개발하려면 우선적으로 특징 지도를 생성하는 Backbone Network가 반드시 필요한데, 본 논문에서는 이를 선정하는 방법을 제안하고자 한다. 이를 위해 6,000여 장의 도로 노면 파손 영상 데이터를 확보하고, 근래에 많이 사용되는 4종류의 심층 신경망을 활용하여 성능을 비교한다. 3가지의 성능 평가 방법을 적용하여 심층 신경망의 특징을 분석하고 최적의 심층 신경망을 결정한다. 또한 하이퍼 파라미터의 최적 조율을 통해 성능을 향상시키고, 최종적으로 도로 노면 파손 영상 분류를 위하여 85.9%의 정확도로 수행이 가능한 경량화된 Backbone Network용 심층 신경망을 제안한다.

핵심어 : 도로 노면 파손, 심층 신경망, 유지보수, 영상 분류, 척추 신경망

ABSTRACT

In recent years, research and development on image object recognition using artificial intelligence have been actively carried out, and it is expected to be used for road maintenance. Among them, artificial intelligence models for object detection of road surface are continuously introduced. In order to develop such object recognition algorithms, a backbone network that extracts feature maps is essential. In this paper, we will discuss how to select the appropriate neural network. To accomplish it, we compared with 4 different deep neural networks using 6,000 road surface damage images. Based on three evaluation methods for analyzing characteristics of neural networks, we propose a method to determine optimal neural networks. In addition, we improved the performance through optimal tuning of hyper-parameters, and finally developed a light backbone network that can achieve 85.9% accuracy of road surface damage classification.

Key words : Road Surface Damage, Deep Learning, Maintenance, Image Classification, Backbone Network

Received 6 May 2019
Revised 31 May 2019
Accepted 25 June 2019

© 2019. The Korea Institute of Intelligent Transport Systems. All rights reserved.

I. 소개

1. 연구의 배경

4차 산업혁명 시대를 맞이하여 국내에서는 IT 기술과의 융합을 통한 새로운 방향을 모색하는 연구들이 여러 분야에서 다양한 형태로 진행되고 있다. 그 중에서도 건설 산업 분야는 정부 차원의 지속적인 관심을 바탕으로 기술 개발에 노력을 기울이고 있으며, 2030년까지 건설자동화 완성이라는 목표 아래 ICT 기술을 설계, 시공, 유지관리 분야에 고르게 접목하기 위한 기술 개발에 집중하고 있다. (MOLIT, 2018)

이러한 흐름 속에서 도로 노면 유지관리 분야에서도 새로운 기술 개발 및 연구가 활발하게 진행되고 있다. 통상적으로 도로 노면 관리는 도로의 수명 확보와 주행 차량의 안전 확보를 위하여 중요하므로 파손 발생 시 도로 관리 주체의 보수 조치가 신속히 이루어져야 한다. 원활한 보수 조치 수행에 있어서 가장 중요한 점은 도로 노면의 파손 위치를 빠르고 정확하게 파악하는 것이다. 그러나 현재까지 이와 같은 정보는 운전자 또는 보행자가 직접 민원 시스템에 접속하여 신고하는 내용을 토대로 파악하는 인력 기반의 신고에 의존할 수밖에 없는 실정이었다. 이처럼 수동적인 정보 수집으로 인하여 도로 노면 파손 상황과 그 실태를 파악하는 데 시간이 지체되고 그만큼 보수도 지연된다는 점을 간과해서는 안 된다.

이러한 보수 과정의 효율성을 제고하기 위해 다수의 기술 개발과 연구들이 수행되고 있는데, 최근에는 이 중에서도 인공 지능을 적용한 연구들이 주목받고 있다. 이는 현재까지의 고전적인 영상 처리 방식인 Feature extraction을 기반으로 해온 연구에 비해 그 성능이 월등히 향상된 것이다. 뿐만 아니라, Tensorflow, PyTorch, Keras와 같은 인공 지능 기술을 쉽게 활용할 수 있도록 하는 프레임워크도 무료로 사용할 수 있어 상대적으로 적용이 수월해졌다. 이런 시류에 따라 정부 주도형 국가 R&D 과제를 통하여 기술 개발에 집중적인 투자가 이루어졌고, 그 성능 및 기술도 지속적으로 발전되고 있다.

2. 도로 노면 파손 탐지를 위한 선행 연구

영상처리를 통한 도로 노면 파손 탐지를 위한 다양한 연구가 진행되어 왔다. 첫 번째로 국내에서는 Jo et al.이 개발한 내용으로 Feature Extraction 기반의 알고리즘 중 하나인 Saliency Map을 이용한 포트홀 탐지 방법이 있다 (Jo et al., 2016). 이 탐지 방식은 크게 후보 추출 단계와 결정 단계로 구성된다. 후보 영역 추출에서는 총 7가지 과정을 거쳐 수행되는데, 이 중에서 Saliency 기반의 이진화 알고리즘을 사용했다는 점이 가장 큰 특징이다. 다음으로 결정 단계에서는 ROI 영역과 소실점을 활용한 영역 필터, 표준 편차와 크기를 고려한 필터 등을 사용하여 최종 포트홀 영역을 찾는다.

두 번째로 심층 신경망을 활용한 연구로는 Chun et al.이 제안한 방법이 있다 (Chun et al., 2018). 이 논문에서 의하면 자율주행에서 많이 사용되는 SegNet과 같이 자기부호화 형태의 심층 신경망을 활용하여 포트홀을 탐지하였다 (Badrinarayanan et al., 2017). 이 논문에서는 포트홀 뿐만 아니라 도로 노면에서 발생할 수 있는 균열까지 탐지할 수 있는데, 이를 위하여 실제 주행을 통해 총 3,000여 장의 영상 데이터를 수집하였다. 그리고 도로 분야의 전문가들이 이 영상에 도로 파손 영역을 표시하였으며, 이를 바탕으로 높은 인식 성능을 가지는 심층 신경망을 개발하였다.

Lee et al.은 Feature extraction 방법과 인공 지능 방법이 융합된 알고리즘을 제안하였다 (Lee et al., 2018). 이는 고전적인 영상 처리 기법을 통해 전처리를 수행하여 포트홀 후보군을 추출하고, 그 후보군을 기계 학습을 이용하여 최종 결정함으로써 탐지 정확도를 향상시켰다. 또한 Kim et al.은 인공 지능을 활용한 방법과 영

상 처리 기법을 융합하고 더 나아가 스마트 폰에 탑재하여 실제 주행 실험을 수행하였다. 그 결과 시멘트 도로 파손 영역과 포트홀을 탐지할 수 있는 가능성을 확인하였다 (Kim et al., 2018)

국외에서도 이와 관련한 연구가 활발히 진행 중인데, Maeda et al.은 딥러닝을 통해 도로 노면에서 발생하는 이상 객체를 탐지하는 연구를 수행했다 (Maeda et al., 2018). 학습 데이터로는 스마트폰에서 촬영된 약 9,000여 장의 영상을 활용하였다. 그리고 이 영상에 포함된 15,000여 개 이상의 객체들을 표시하고 구분하여 학습하기 위한 데이터 세트를 구성하였다. 이를 활용한 딥러닝 기반의 여러 객체 인식 알고리즘을 적용하였고 그 인식 성능을 확인하였으며 스마트 폰에도 적용하여 실험하였다. 다음으로 Yang et al.이 제안한 방식을 살펴보면 FCN 기반의 의미론적 분할(Semantic Segmentation) 방법을 적용한 점이 특징이다 (Yang et al., 2018; Long et al., 2015). 인공 지능 모델을 학습하기 위하여 총 800장 이상의 영상을 확보하였고, 합성곱 연산과 역합성곱 연산의 비대칭형으로 심층 신경망을 구성하였다. 그 결과 pixel-to-pixel 탐지 알고리즘을 제안하였고, 97.96%라는 인식 성능을 확보하였다. 끝으로 Zhang et al.은 아스팔트 표면으로부터 포장 균열을 추출하는 심층 신경망을 제안하였다 (Zhang et al., 2017). 1mm의 정확도를 가지는 카메라를 활용해 시속 100km의 속도로 주행하면서 영상 데이터를 수집하였다. 이 때 균열의 깊이 정보까지 수집하여 3차원 데이터 기반의 심층 신경망을 학습했다는 점이 특징이다. 결론적으로 90.13%의 정밀도와 87.63%의 재현성을 가진 알고리즘을 제안하였다.

3. Backbone Network 관련 선행 연구

앞서 언급한 도로 노면 파손 탐지에서 사용된 객체 인식 알고리즘과는 달리 Backbone Network를 활용한 R-CNN 계열의 알고리즘도 있다. 대표적인 예로는 Fast R-CNN, Faster R-CNN, Mask R-CNN 등이 있는데, 이와 같은 방식 또한 도로 노면 파손을 탐지하는데 충분히 활용이 가능하다 (Girshick et al., 2015; Ren et al., 2015; He et al., 2017). 하지만 이런 방식을 사용하기에 앞서 선행되어야 하는 점은 Backbone Network의 선정이다. Backbone Network는 대부분 합성곱으로 구성되어 있어 R-CNN 방식에서 객체 인식을 하는데 특징 지도를 생성하는 중추적인 역할을 하는 신경망이다. 이런 R-CNN 방식에서 공통적으로 사용하는 심층 신경망으로 이 신경망의 구조에 따라 객체 인식 성능에 차이가 발생하므로 신경망의 특성을 고려하여 선택하는 것이 필수적이다. 이를 위하여 최근 들어 대표적인 Backbone Network로 많이 활용되는 VGGNet, ResNet, DenseNet, DetNet의 특성에 대하여 살펴보고자 한다.

우선 VGGNet의 경우는 모델의 깊이에 대하여 초점을 맞추었다 (Simonyan et al., 2014). GPU의 연산 속도가 빨라지면서 신경망 층이 깊어짐에도 불구하고 빠른 연산이 가능해졌다. 이런 하드웨어 환경에서 3×3의 작은 필터로 구성된 단순한 구조를 갖는 신경망의 깊이에 따른 성능의 변화에 대해 관찰하였으며 신경망의 깊이는 11개에서 19개까지 조절하며 그 성능을 비교하였다. 이렇게 작은 필터를 사용하는 이유는 큰 필터를 사용했을 때보다 더 많은 비선형을 다양하게 표현할 수 있다는 장점 때문이다. 하지만 파라미터가 과도하게 많아져 Vanishing Gradient 문제, Over Fitting과 같은 문제가 발생할 가능성이 높다는 단점도 있다.

ResNet은 2015년 ILSVRC 대회에서 우승한 모델로 Top-5 Error가 3.6%로 나올 만큼 성능이 우수한 모델이다 (He et al., 2016). 이 모델은 VGGNet에서 발생했던 문제와 같이 신경망이 깊어질수록 성능이 나빠지는 점을 개선하기 위한 새로운 방법을 사용하였다. 특히 깊은 신경망이 얇은 신경망보다 성능이 나빠지는 Degradation 문제 해결을 위한 구조를 제시하였는데, Skip Connection이라는 기법을 사용한 잔차 신경망 (Residual Network)이다. 이는 입력으로 들어온 정보를 여러 합성곱을 거친 후의 정보와 연결시킴으로써 정보의 소실을 줄이는 역할을 한다. 따라서 합성곱을 통해 특징 정보를 생성함과 동시에 최초의 정보를 계속해서

유지함으로써 본래의 특징을 보존하도록 학습이 진행된다.

DenseNet은 2016년도에 새롭게 제안된 CNN 모델이다 (Huang et al., 2017). ResNet에서 언급한 Skip Connection과는 달리 Dense Connectivity를 제안하였는데, 이는 입력과 출력을 더하는 것이 아니라 합치는 개념이다. 이는 입력 정보가 출력 정보에서 그대로 보존되게 하는 역할을 수행한다. 일반적으로 GPU의 제한된 메모리로 인해 CNN 연산을 수행하기 위해서는 입력 영상의 크기를 축소하여 정보의 손실이 불가피한데, DenseNet 입력 정보를 다양하게 다음 신경망으로 연결함으로써 소실될 수 있는 정보를 문제없이 보존할 수 있다는 장점이 있다. 또한 이 신경망은 ResNet에 비해 훨씬 깊은 신경망 구조를 가졌지만, Average Pooling을 포함한 병목 신경망을 사용하여 연산량을 줄이기 위한 시도를 하였고, 그 결과 신경망 개수에 비해 상대적으로 파라미터가 작아 빠른 연산 속도를 가지는 특징이 있다.

끝으로 소개하는 DetNet은 우선 ResNet50과 유사하게 구성되어 있다 (Li et al., 2018). DetNet은 총 6단계로 구성되어 있는데, 4번째 단계까지는 ResNet50과 동일한 신경망을 사용한다. 하지만 5번째 단계부터는 확장 합성곱 (Dilated Convolution)을 사용하는 차이점을 보인다 (Yu et al., 2015). 이를 사용하여 효율적으로 수용영역 (Receptive Field)을 확장하는데 주안점을 두었고 연산 속도를 고려하여 4번째 단계 이후로부터 채널수를 늘리지 않고 256개로 유지하였다. 이 같은 전략을 바탕으로 개발한 신경망으로 동일한 데이터를 적용해 실험한 결과 ResNet50 보다 우위에 있는 것이 Backbone Network임이 확인되었다.

본 논문에서는 앞서 언급한 4가지의 심층 신경망을 활용해 도로 노면 파손 영역을 탐지하는데 최적의 Backbone Network를 결정하는 기법을 소개하고자 한다. 이를 위하여 우선 심층 신경망 모델을 개발하는 영상 데이터를 확보하고 분류 기준을 수립한다. 그리고 이를 바탕으로 영상 분류 (Image Classification) 문제를 풀도록 한다. 끝으로 다양한 성능 평가 방법을 바탕으로 심층 신경망 간의 성능 비교를 통해 최적의 신경망을 최종적으로 선정한다.

II. 도로노면 파손 영상 데이터 확보 및 분류기준

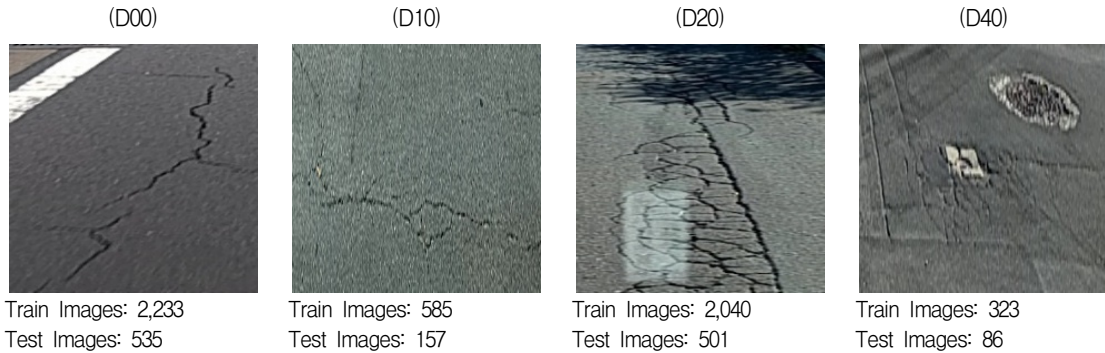
1. 도로 노면 파손 영상 데이터의 확보

도로 노면 파손의 형태는 가로 방향 및 세로 방향 균열, 거북등 균열, 포트홀, 러팅, 라벨링 등과 같이 다양하게 나타난다. 이러한 파손 형태 가운데 어떤 것을 인식할 수 있는 알고리즘을 개발하는지 여부는 개발자의 의도에 따라 다양하게 결정된다. 우선 Jo. et al.의 경우는 포트홀만을 인식할 수 있도록 알고리즘을 설계하였고, Chun et al.의 경우에서 균열과 포트홀을 모두 파손으로 인식할 수 있게 개발하였으나 분류하지는 않았다. (Jo et al., 2015; Chun et al., 2018). 또, Eisenbach et al.의 논문에서는 FGSV 규정에 따라 총 5가지로 분류하였는데 균열, 포트홀, 패치와 같은 형태를 구분할 수 있도록 심층 신경망을 개발하였다 (Eisenbach et al., 2018). 끝으로 Maeda et al.은 도로 노면 파손 영상을 스마트폰으로 촬영하여 수집한 후 총 8종류로 구분하였다. 선형 균열과 피로 균열은 물론이고 도로 마커의 형태까지 인식하기 위해 영상 데이터를 수집하고 분류하였다. 본 논문에서는 이들 가운데 Maeda et al.이 사용한 데이터를 사용하였는데 (Maeda et al., 2018), 그 이유는 본 연구의 최종적인 목표가 모바일 형태의 장치로부터 영상을 취득하고 이를 활용하여 도로 노면 파손을 인식하는 알고리즘을 개발하는 것이기 때문이다. 따라서 최종 적용 분야를 고려했을 때 유사한 방식으로 취득한 영상을 사용하는 것이 적합한 것으로 판단된다.

2. 도로 노면 파손 분류 알고리즘 개발을 위한 기준

본 논문에서 사용하는 도로 노면 파손 분류를 위한 기준은 <Fig. 1>과 같다. Maeda et al.이 제공하는 영상 데이터 중에서 4가지 종류만을 선별하여 사용하였다. 여기서 4가지 종류는 세로 균열, 가로 균열, 거북등 균열, 포트홀로 구성된다. D00은 균열의 방향이 차량의 진행 방향과 평행하고, D10은 진행 방향과 수직이다. 그리고 D20은 거북등 모양으로 생긴 균열을 나타내고, D40은 포장된 표면이 파손되어 구멍이 된 것을 일컫는다.

이렇게 선별한 이유는 도로 교통 안전 측면에서 포트홀과 같이 운전자의 조향에 영향을 주는 대상의 탐지가 가장 중요하다고 판단되기 때문이다. 따라서 자연적으로 발생한 것이 아닌, 포장 공사 과정에서 발생한 Joint Point는 대상에서 제외하였다. 이와 함께 도로 노면 마커에 대한 훼손 여부도 제외한다. 비록 마커의 훼손 또한 운전자에게 도로 상황 정보를 정확하게 제공하지 못하여 피해를 줄 여지가 있지만, 본 연구에서는 도로 포장 상태 유지보수와 관련된 노면 상태 불량에 따른 사고 예방에 초점을 맞추었기 때문이다.



<Fig. 1> Road Damage Images and its Numbers: (D00) Longitudinal Linear Crack and Wheel Mark Part, (D10) Lateral Linear Crack, (D20) Alligator Crack, (D40) Rutting Bump and Pothole and Separation (Maeda et al., 2018)

Ⅲ. 심층 신경망 분석을 위한 평가 및 학습 방법

1. 심층 신경망 분석을 위한 평가 방법

본 논문에서 심층 신경망의 성능 비교를 위하여 3가지 방법을 사용하였다. 그 중 첫 번째는 홀드 아웃 검증 (Holdout method) 방법으로 학습용 데이터 세트와 실험용 데이터 세트를 구분하여 학습된 인공 지능의 모델 성능을 평가하는 것이다. 본 논문에서는 전체 데이터 6,000여 장을 8대 2로 구분하여 각각 학습용과 실험용으로 사용하였고, 4개의 심층 신경망 모델에 동일하게 적용하였다.

다음으로 평가에 활용한 방법은 Top-N 정확도 방식이다. 이 방식은 인공지능을 통한 분류 결과의 성능을 파악하는데 주로 사용하는 방식이다. 대부분의 분류 결과는 확률 값으로 나오게 되는데, 이 확률값 중에서 상위 N개의 정확도를 분석하는 방식이다. 본 논문에서는 총 4개의 도로 노면 파손을 분류하지만, Top-1 정확도는 홀드 아웃 검증과 중복되므로 Top-2 정확도만을 분석하여 인공지능 모델을 평가하였다.

끝으로 K-겹 교차검증 (K-fold Cross Validation) 방법을 활용하였다. He et al.이 제안한 논문에서는 홀드 아

아웃 검증과 Top-N 정확도만을 사용하지만, 본 논문에서는 검증에 있어서 신뢰성을 확보하기 위하여 이 방법을 추가적으로 사용하고자 한다 (He et al., 2016). 데이터를 K개의 그룹으로 분할하고 특정 하나의 그룹만을 제외한 후 나머지 데이터를 학습용 데이터 세트로 활용하는 방식이다. 그리고 사용하지 않은 하나의 그룹 데이터는 실험용 데이터 세트로 사용한다. 이렇게 K개 그룹의 모든 데이터 세트를 같은 방식으로 K번 반복한다. 일반적으로 K의 값은 5~10으로 설정하는데, 본 논문에서 K는 5로 하였다. 따라서 5번의 모델 평가 값을 바탕으로 평균과 편차를 분석하여 모델의 성능을 평가하였다.

2. 학습 파라미터 설정

인공 지능을 학습하기 위해서는 여러 가지 파라미터의 설정이 필수적이다. 우선 영상의 크기를 모두 동일하게 변경하였다. Maeda et al.이 제공하는 영상의 크기는 모두 600×600인데, 도로 노면 파손 영역의 크기는 다양하다. 이럴 경우 특별한 기법을 사용하지 않는 한 CNN 모델을 통하여 학습하기가 쉽지 않아 본 논문에서는 224×224로 통일하였다. 또한 인공지능 모델의 가중치의 초기 값을 모두 Xavier로 설정하였고, 최적화 함수는 ADAM을 적용하였다 (Glorot and Bengio, 2010; Kingma et al., 2015). 이때 사용한 Learning Rate는 0.0001이고 beta-1은 0.9이며 beta-2는 0.999이다. 그 외 파라미터로는 배치 크기와 반복 횟수가 있는데 홀드아웃 검증에서는 50개과 500회 그리고 K겹 교차 검증에서는 30개와 250회로 각각 설정하였다.

이 같은 방법을 활용하여 도로 노면 파손 분류를 위한 가장 성능이 뛰어난 심층 신경망을 선정한 후 최적화 함수 파라미터의 조율을 수행한다. 동일한 심층 신경망이어도 하이퍼 파라미터에 의하여 학습된 결과는 달라질 수 있다. 이 같은 이유로 파라미터 변화에 따른 인식 성능의 차이점을 확인해 볼 필요가 있으므로 본 논문에서는 최적화 함수의 Learning Rate와 종류에 따른 학습 결과에 대하여 분석하였다.

IV. 실험 결과

1. 홀드 아웃 검증 분석

동일한 데이터와 학습 파라미터를 사용한 결과는 <Table 1>과 같다. 4개의 인공지능 모델에 대한 최대 학습 정확도, 최소 학습 손실값, 최대 실험 정확도, 최소 실험 손실값을 기록하였다. 먼저 학습 정확도는 모델에 관계없이 1로 수렴하는 것을 확인할 수 있었다. 하지만 그 외의 결과들은 모델마다 다르게 나타났다. 그 중에서도 최소 학습 손실값은 VGGNet이 가장 낮게 나타났지만, 실험 정확도와 실험 손실값을 모두 고려했을 때 DenseNet이 가장 우수한 것으로 나타났다. 이번 실험을 통해 DenseNet이 도로노면 파손 영역을 탐지하는데 유리할 수 있다는 가능성을 확인하였다.

<Table 1> Holdout Validation Results

Name of Model	VGGNet-19	ResNet-50	DenseNet-121	DetNet-59
Maximum Train Accuracy	1.000	1.000	1.000	1.000
Minimum Train Loss	1.192e-5	1.859e-5	2.230e-5	1.355e-5
Maximum Test Accuracy	0.790	0.826	0.855	0.832
Minimum Test Loss	0.628	0.733	0.573	0.679

2. Top-N 정확도 분석

본 실험에서는 영상 분류 정확도를 확인할 수 있다. 본 논문에서 사용한 4가지의 인공지능 모델에 대한 분류 정확도는 <Table 2>와 같다. Top-2 정확도는 인식 결과 4개 중에서 높은 확률 기준으로 상위 2개가 정답일 확률을 나타낸 것이다. 이는 DetNet이 0.002 차이로 DenseNet보다 정확한 것을 확인할 수 있었다. VGGNet을 제외하고는 모두 0.95를 넘을 정도로 정확도가 높게 나왔고 그 차이는 서로 크지 않음을 알 수 있다.

<Table 2> Top-2 Accuracy Results

Name of Model	VGGNet-19	ResNet-50	DenseNet-121	DetNet-59
Top-2 Accuracy	0.919	0.951	0.956	0.958

3. K-겹 교차 검증 분석

본 실험은 데이터가 학습 데이터와 실험 데이터를 일정한 규칙을 바탕으로 교차함으로써 신경망의 성능을 평가하는 방법이다. 이를 통해 편향된 데이터의 분할로 인하여 심층 신경망의 잘못된 성능 평가를 예방하고자 한다. 다시 말해, 이 분석은 데이터의 분포 및 분할과 무관하게 오로지 심층 신경망의 성능을 평가한 것이다. 신경망마다 각각 5번의 실험 데이터 정확도를 구한 뒤, 이 값들을 바탕으로 평균과 편차를 구하였고 그 결과는 <Table 3>와 같다. 실험 데이터의 평균 정확도는 DenseNet이 0.831로 가장 높았다. 그리고 실험 데이터의 표준 편차는 ResNet이 0.0136으로 가장 낮았지만, DetNet의 표준 편차는 0.0139로 유사한 수치를 보였다.

이 실험을 통해 2가지의 결론을 얻을 수 있다. 첫째는 DenseNet의 성능이 다른 심층 신경망보다 우수하다는 점이다. 표준 편차를 제외한 모든 값들이 다른 신경망에 비해 높게 나왔다. 두 번째는 잔차 신경망 개념을 사용하고 있는 심층 신경망의 정확도가 높게 나타났다. ResNet과 DetNet은 서로 유사하게 Skip Connection을 사용하고 있고, DenseNet 또한 앞단의 신경망 정보가 뒷단까지 소실되지 않게 Dense Connectivity를 적용하고 있다. 이 같은 방식은 균열과 같이 Down Sampling을 하면서 고유한 Feature가 소실될 수 있는 객체가 있을 경우에 큰 효과가 있는 것으로 보인다. 따라서 도로 노면에서 발생할 수 있는 미세한 균열과 같은 객체가 많고 이를 인식해야할 상황이라면 Feature의 소실을 막을 수 있도록 이 같은 기법으로 신경망을 설계하는 것이 유리할 것이다.

<Table 3> K-fold Cross Validation Results

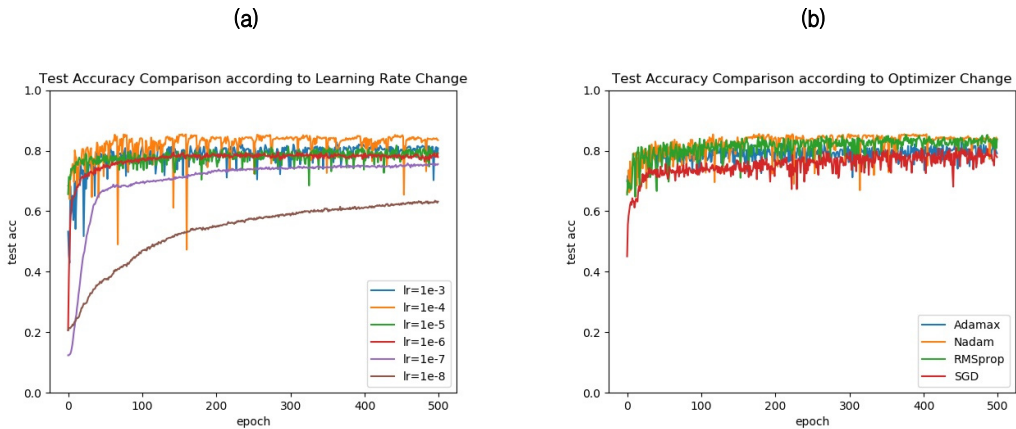
Name of Model	VGGNet-19	ResNet-50	DenseNet-121	DetNet-59
Mean Test Accuracy	0.783	0.807	0.831	0.803
Std. Test Accuracy	0.016	0.014	0.016	0.014
Maximum Test Accuracy	0.807	0.826	0.846	0.820
Minimum Test Accuracy	0.757	0.786	0.799	0.783

4. 최적화 함수의 선정

3가지 성능평가 방법을 통해 선정한 DensNet을 활용하여 인식 성능을 높이기 위한 파라미터 최적화 방법을 제안하였다. 이를 위해 본 논문에서는 Learning Rate를 10^{-3} 에서부터 10^{-8} 까지 10^{-1} 씩 간격으로 조정하였다.

그리고 실험 데이터 세트의 정확도를 살펴보았는데, 그 결과는 <Fig. 2(a)>와 같다. 이 그래프를 보면 Learning Rate가 10^{-3} 에서부터 시작해서 10^{-4} 일까지는 정확도가 높게 나타나고, 다시 10^{-5} 일 때부터는 정확도가 낮은 것으로 드러났다. 그리고 Learning Rate가 계속해서 커질수록 곡선의 변화율이 완만한 양상을 보이는 것에 반해 정확도가 점진적으로 하락하는 것을 확인할 수 있었다. 그리하여 마지막으로 Learning Rate가 10^{-8} 일 때는 정확도가 현저히 하락하는 것을 확인할 수 있다. 결과적으로 Learning Rate가 10^{-4} 일 때 0.8546으로 가장 높게 나타났다.

다음으로는 최적화 함수의 종류에 따른 실험 데이터 세트의 정확도 차이를 살펴보았다. 이를 위하여 총 4개의 최적화 함수를 사용하였고 Learning Rate는 10^{-4} 로 하였다. 그 결과는 <Fig. 2(b)>와 같다. Dozat이 제안한 Nadam을 최적화 함수로 사용했을 때가 0.8554로 가장 높은 정확도를 보였다 (Dozat, 2016). 이 최적화 함수로 학습했을 때, 비록 수렴하는 추이는 다른 최적화 함수를 사용했을 때보다 큰 변화율을 보이지만 결과적으로 인식 성능은 85.54%로 다른 것에 비해 가장 높은 것으로 확인되었다.



<Fig. 2> Test Accuracy Comparison: (a) Test Accuracy according to Learning Rate Change when Adam is used and (b) Test Accuracy according to Optimizer Change when Learning Rate is 10^{-4} .

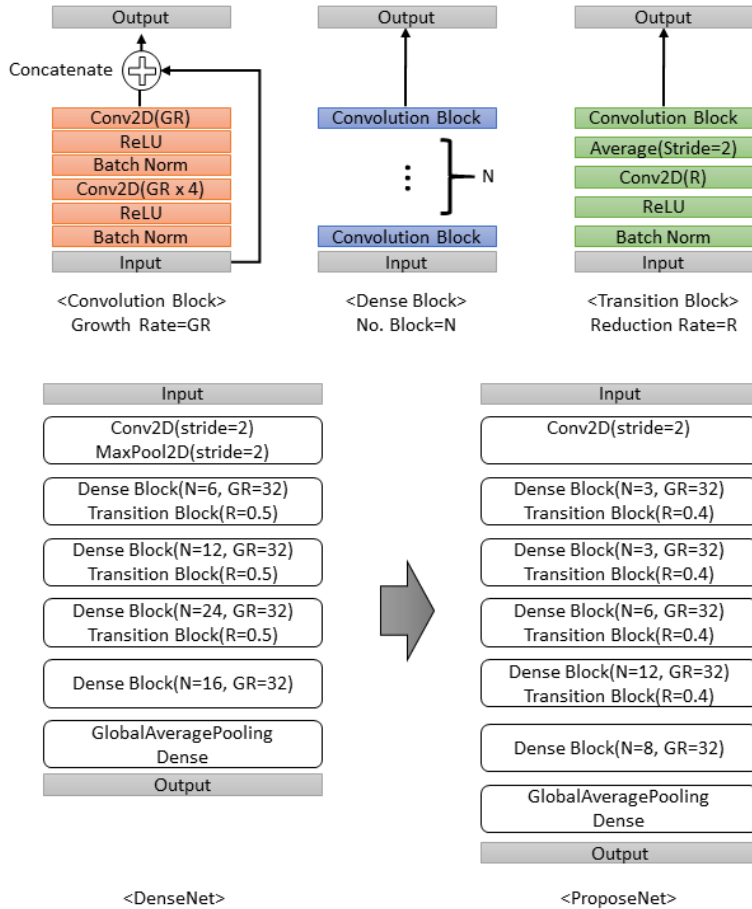
V. 경량 심층 신경망 설계 및 개발

1. 심층 신경망의 경량화를 위한 설계

앞선 실험을 통해 4가지의 각기 다른 심층 신경망 중에서 DenseNet의 인식률이 가장 높은 것으로 파악되었다. 다음 단계는 이 심층 신경망을 활용하여 도로 노면 파손 분야에 적합하게 신경망을 재설계하는 것이다. 재설계의 목적은 신경망의 크기를 축소하고 연산 시간을 단축하는 것이다. 기존의 DenseNet은 CIFAR 데이터 세트와 SVHN 데이터 세트를 대상으로 실험을 수행하였으며, 그에 초점을 맞추어 설계되어 있다 (Krizhevsky et al., 2009; Netzer et al., 2011). 따라서 새로운 데이터 세트에 맞게 신경망을 수정하여 적용 분야에 적합하게 개선할 필요가 있다. 이 같은 이유로 본 논문에서는 <Fig. 3>과 같이 수정된 신경망(ProposeNet)을 제안한다.

ProposeNet의 특징은 첫째로 DenseNet에 있었던 Maxpooling을 사용하지 않는다는 점이다. 모든 신경망을 합성곱의 형태로 구성한 이유는 단순한 Down sampling을 사용하여 영상에서 남아 있는 정보의 손실을 최소화

화하기 위함이다. 그리고 Maxpooling을 대신하여 Dense Block과 Transition Block을 추가하였다. 두 번째 특징은 신경망의 크기를 절반으로 축소한 점이다. DenseNet의 Dense Block에 사용되는 No. Block은 6, 12, 24, 16이다. 하지만 ProposeNet의 Dense Block에서 사용되는 No. Block은 3, 6, 12, 8로 수정하였다. 그리고 Transition Block에 사용되는 Reduction Rate는 0.5에서 0.4로 변경하였다. 그 이유는 본 논문에서 가정한 도로 노면 파손 영상의 4가지 종류가 CIFAR과 SVHN 데이터 세트의 종류에 비하여 적기 때문이다. 따라서 영상 분류를 위한 심층 신경망의 Feature 또한 단순화가 가능하여 신경망의 수도 감소시켜 설계하였다.



<Fig. 3> Comparison between DenseNet and ProposeNet

2. ProposeNet의 성능

새롭게 제안한 심층 신경망의 성능을 파악하기 위하여 DenseNet과 비교하였다. 총 4가지 지표를 사용하였고 그 결과는 <Table 4>와 같다. 첫 번째 지표는 Test Accuracy로 결과적으로 0.4% 향상되었다. 두 번째는 Top-2 Accuracy로 ProposeNet이 DenseNet에 비해 0.6% 낮은 것으로 나타났다. 마지막으로 Total Parameter는 DenseNet에 비하여 32.8%의 수준으로 연산량이 크게 감소되었다. 그 결과 1,279장을 연산하는데 소요되는 시간이 13.98초로 약 2배가량 빨라진 것으로 나타났다.

<Table 4> Comparison of the Evaluation Results between DenseNet and ProposeNet

Evaluation Method	Test Accuracy	Top-N Accuracy (N=2)	Total Parameter	Processing Time (S)
DenseNet-121	0.855	0.956	7,041,604	22.34
ProposeNet	0.859	0.950	2,307,992	13.98

VI. 결 론

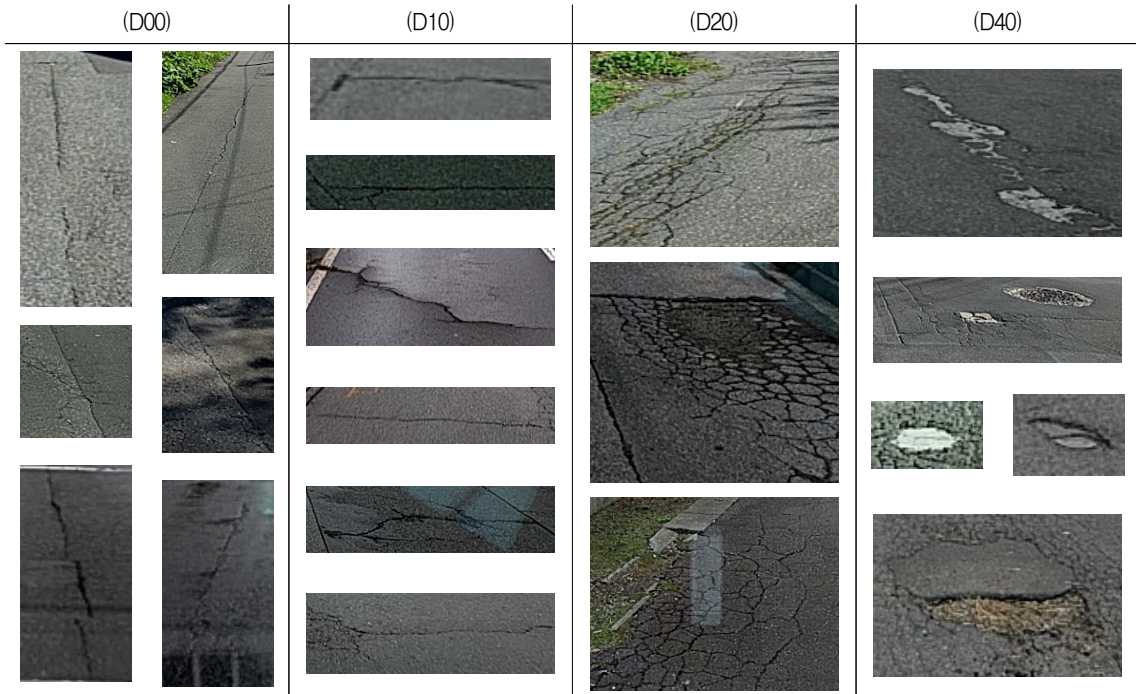
1. Backbone Network의 설계

본 연구에서는 도로 노면 파손 객체 인식을 위하여 필요한 Backbone Network를 선정하는 기법에 대하여 제안하였다. 이는 R-CNN 계열의 객체 인식을 구현하는데 있어 가장 중요한 심층 신경망을 선정하는 단계로 인식 성능과도 직접적인 연관성이 있다. 최적의 Backbone Network를 선정하기 위하여 우선 최근에 많이 활용되고 있는 4개의 심층 신경망을 선별하였고, 이를 활용하여 영상 분류 (Classification) 기능을 가진 인공지능 모델을 개발하였다. 데이터 세트로는 Maeda. et al.이 제공하는 영상을 대상으로 가로 균열, 세로 균열, 거북등 균열, 포트홀이 포함된 영상만 추출하여 사용하였다. 그리하여 학습을 위해 총 5,000여 장의 영상을 확보하였고, 실험을 위해 1,000여 장의 영상도 확보하였다. 이 영상 데이터와 심층 신경망을 이용하여 본 연구에서는 3가지의 평가 기법을 적용하여 최적의 심층 신경망을 선정하는 절차를 제안하였다. 그 결과 홀드 아웃 검증과 K-겹 교차 검증을 사용했을 때는 DenseNet이 우수한 성능으로 나왔으나, Top-2 정확도를 분석했을 때는 DetNet이 우수한 성능으로 나왔다. 그런데 본 연구의 목적은 가장 정확하고 인식 성능이 높은 심층 신경망을 개발하는 것에 있으므로 DenseNet을 선택하는 것이 합리적으로 보인다. 다음으로 최종 결정된 DenseNet을 사용하여 하이퍼 파라미터 조율을 통해 85.5%의 인식 성능을 가지는 Backbone Network를 확보하였다. 끝으로 DenseNet의 구조를 수정 및 경량화 하여 인식 성능을 0.4% 향상시킴과 동시에 속도 역시 약 2배가량 빠른 심층 신경망을 제안하였다. 그리고 이 심층 신경망을 이용해 도로 노면 파손 영상을 분류할 경우 <Fig. 4>와 같은 결과를 얻을 수 있었다. 이는 ProposeNet이 학습된 바와 같이 가로 균열, 세로 균열, 거북등 균열, 포트홀에 대하여 정확한 구분이 가능하다는 것을 의미한다.

2. 향후 연구 계획

본 연구에서는 Backbone Network의 선정을 위한 절차와 기법에 대하여 제안하였다. 향후에는 이를 바탕으로 다양한 객체 인식 알고리즘의 개발이 가능할 것으로 전망된다. 그 가운데에서도 최근에 많이 사용되고 있는 R-CNN 계열의 알고리즘이 대표적이다. 구체적인 예로는 Backbone Network의 종단에 위치 추정 (Localization)과 분류 (Classification)의 기능을 동시에 수행하는 심층 신경망의 개발이 되겠다. 또한 고속 연산 처리를 위한 각종 방법 적용해야 한다는 점을 고려해야 하며, 임베디드 시스템과의 연계를 통한 소형화 및 경량화에 대한 기술 개발도 필요할 것이다.

또한 본 논문에서 사용한 영상 데이터는 일본 도로 상황을 반영한 것으로 우리나라 도로 실정과는 다소 차이가 있다. 이 점을 개선하고자 우리나라 도로에서 발생할 수 있는 도로 노면 파손 정보에 대한 대규모 수집이 요구된다. 또한 이 데이터를 활용하여 우리나라 도로 상황에 맞게 가장 최적화된 형태의 객체 인식 기술을 개발하여 도로 포장 및 유지 관리에 기여할 점을 구체적으로 모색해야 할 것이다.



<Fig. 4> Road Damage Image Classification Result

ACKNOWLEDGEMENTS

본 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2017R1C1B5018218). 그리고 산업통상자원부 ‘산업전문인력역량강화사업’의 재원으로 한국산업기술진흥원(KITA)의 지원을 받아 수행된 연구임. (2019년 산업용 무인비행장치 전문인력양성사업, 과제번호 : N0002431)

REFERENCES

- Badrinarayanan V., Kendall A. and Cipolla R.(2017), “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp.2481-2495.
- Chun C., Shim S., Kang S. and Ryu S.(2018), “Development and evaluation of automatic pothole detection using fully convolutional neural networks,” *Journal of Korea Institute of Intelligent Transport System*, vol. 17, no. 5, pp.55-64.
- Dozat T.(2016), *Incorporating nesterov momentum into adam*.
- Eisenbach M., Stricker R., Seichter D., Amende K., Debes K., Sesselmann M., Ebersbach D., Stoeckert U. and Gross H.(2017), “How to get pavement distress detection ready for deep

- learning? A systematic approach,” in *Proc. International Joint Conference on Neural Networks (IJCNN)*, Anchorage, Alaska, pp.2039-2047.
- Girshick R.(2015), “Fast r-cnn,” in *Proc. The IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp.1440-1448.
- Glorot X. and Bengio Y.(2010), “Understanding the difficulty of training deep feedforward neural networks,” in *Proc. the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy, pp.249-256.
- He K., Gkioxari G., Dollár P. and Girshick R.(2017), “Mask R-CNN,” in *Proc. The IEEE International Conference on Computer Vision (ICCV)*, Venezia, Italy, pp.2980-2988.
- He K., Zhang X., Ren S. and Sun J.(2016), “Deep residual learning for image recognition,” in *Proc. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, pp.770-778.
- Huang G., Liu Z., Van Der Maaten L. and Weinberger K. Q.(2017), “Densely connected convolutional networks,” in *Proc. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, Honolulu, pp.4700-4708.
- Jo Y. and Ryu S.(2015), “Pothole detection system using black-box camera,” *Sensors*, vol. 15, no. 11, pp.29316-29331.
- Jo Y. and Ryu S.(2016), “Pothole detection algorithm based on saliency map for improving detection performance,” *Journal of Korea Institute of Intelligent Transport System*, vol. 15, no. 4, pp.104-114.
- Kim S. J., An K. E., Lee S. W., Ryu S. and Seo D.(2018), “A study of image processing and deep learning algorithm for efficient of mobile-based pothole detection system,” in *Proc. Korea Software Congress 2018 (KSC)*, Pyeongchang, Korea, pp.1595-1597.
- Kingma D. P. and Ba J. L.(2015), “ADAM: a method for stochastic optimization,” in *Proc. third International Conference on Learning Representations (ICLR)*, San Diego, California, pp.1-15.
- Krizhevsky A. and Hinton G.(2009), “Learning multiple layers of features from tiny images,” *Technical report, University of Toronto*, vol. 1, no. 4, pp.7-66.
- Lee S. W., An K. E., Jeong Y. J., Kim S. J., Ryu S. and Seo D.(2018), “A proposal of a real-time pothole detection system based on machine learning and computer vision,” in *Proc. Korea Computer Congress 2018 (KCC)*, Jeju, Korea, pp.7-9.
- Li Z., Peng C., Yu G., Zhang X., Deng Y. and Sun J.(2018), “Detnet: A backbone network for object detection,” *arXiv preprint arXiv:1804.06215*.
- Long J., Shelhamer E. and Darrell T.(2015), “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp.3431-3440.
- Maeda H., Sekimoto Y., Seto T., Kashiyama T. and Omata H.(2018), “Road damage detection and classification using deep neural networks with smartphone images,” *Computer-Aided Civil Infrastructure Engineering*, vol. 33, no. 12, pp.1127-1141.
- MOLIT (Ministry of Land, Infrastructure and Transport). (Oct., 31, 2018). *Smart Construction Technology Road Map*. Available From: http://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?id=95081506 (accessed Feb., 8, 2019).

- Netzer Y., Wang T., Coates A., Bissacco A., Wu B. and Ng A. Y.(2011), “Reading digits in natural images with unsupervised feature learning,” *In NIPS workshop*.
- Ren S., He K., Girshick R. and Sun J.(2015), “Faster R-CNN: towards real-time object detection with region proposal networks,” *in Proc. Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Montreal, Canada, pp.91-99.
- Simonyan K. and Zisserman A.(2014), “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*.
- Yang X., Li H., Yu Y., Luo X., Huang X. and Yang X.(2018), “Automatic pixel-level crack detection and measurement using fully convolutional networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp.1090-1109.
- Yu F. and Koltun V.(2015), “Multi-scale context aggregation by dilated convolution,” *arXiv preprint arXiv:1511.07122*.
- Zhang A., Wang K. C., Li B., Yang E., Dai X., Peng Y., Fei Y., Liu Y., Li J. Q. and Chen C.(2017), “Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 10, pp.805-819.