

# 서울 관악구 도심지역 미세먼지(PM<sub>10</sub>) 관측 값을 활용한 딥러닝 기반의 농도변동 예측

최한수<sup>1</sup> · 강명주<sup>2</sup> · 김용철<sup>3</sup> · 최한나<sup>3,\*</sup>

<sup>1</sup>서울대학교 수학교육연구소

<sup>2</sup>서울대학교 수리과학부

<sup>3</sup>한국지질자원연구원

## Deep Learning-based Prediction of PM<sub>10</sub> Fluctuation from Gwanak-gu Urban Area, Seoul, Korea

Han-Soo Choi<sup>1</sup> · Myungjoo Kang<sup>2</sup> · Yong Cheol Kim<sup>3</sup> · Hanna Choi<sup>3,\*</sup>

<sup>1</sup>Research Institute of Mathematics, Seoul National University, Seoul 08826, Korea

<sup>2</sup>Department of Mathematical Sciences, Seoul National University, Seoul 08826, Korea

<sup>3</sup>Korea Institute of Geoscience and Mineral Resources, Daejeon 34132, Korea

### ABSTRACT

Since fine dust (PM<sub>10</sub>) has a significant influence on soil and groundwater composition during dry and wet deposition processes, it is of a vital importance to understand the fate and transport of aerosol in geological environments. Fine dust is formed after the chemical reaction of several precursors, typically observed in short intervals within a few hours. In this study, deep learning approach was applied to predict the fate of fine dust in an urban area. Deep learning training was performed by combining convolutional neural network (CNN) and recurrent neural network (RNN) techniques. The PM<sub>10</sub> concentration after 1 hour was predicted based on three-hour data by setting SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, and PM<sub>10</sub> as training data. The obtained coefficient of determination value, R<sup>2</sup>, was 0.8973 between predicted and measured values for the entire concentration range of PM<sub>10</sub>, suggesting deep learning method can be developed into a reliable and viable tool for prediction of fine dust concentration.

**Key words :** fine dust (PM<sub>10</sub>), precursor, deep learning, convolutional neural network, recurrent neural network, prediction

### 1. 서 론

국내 대기환경은 계절에 따라 황사 및 미세먼지의 영향이 많이 반영되며, 2010년 중후반부터 고농도 미세먼지의 빈번한 발생으로 대기환경에 대한 인식이 상당히 높아지고 있다. 대기오염은 시정의 불편함을 야기할 뿐 아니라 알레르기성 비염, 천식, 결막염, 아토피 등의 호흡기 및 피부 질환과 연관되어 있음이 전 세계적으로 보고되고 있

다(Ostro et al, 2009; Rückerl et al., 2011). 세계 4대 대도시 중 하나로 꼽히고 있는 서울은(Kim et al., 2018), 좁은 면적에 인구밀도가 높아 자동차 배기가스, 난방가스, 발전소와 같은 에너지 산업연소 등의 다양한 대기오염인자가 존재하고 있다.

미세먼지라고도 불리는 PM<sub>10</sub>은 입자의 크기가 지름 10 μm 이하의 물질을 뜻하며 이들 입자의 농도가 증가하면 가시거리가 감소하여 시각적으로 대기오염을 판별할 수 있다. 이는 초미세먼지 PM<sub>2.5</sub>(직경 2.5 μm 이하의 입자물질)와 함께 현대의 공해병 원인물질로서 많은 주목을 받고 있다. 미세먼지는 생성과정에 따라 액체상/고체상 금속 및 탄소 혼합물이 주를 이루는 1차 미세먼지(Primary aerosol)와, 가스상 유기 및 무기 화합물(질소화합물(NOx), 황화합물(SOx), 휘발성 유기화합물(VOCs) 등의 전구체)

주저자: 최한수, 박사후연구원

공저자: 강명주, 교수; 김용철, 책임연구원

\*교신저자: 최한나, 선임연구원

E-mail: pythagoras84@kigam.re.kr

Received : 2020. 8. 26 Reviewed : 2020. 9. 3 Accepted : 2020. 9. 22

Discussion until : 2020. 12. 31

이 광화학 반응을 통해 형성된 2차 오염물질(Secondary aerosol)로 분류한다(Korea Environment Corporation, 2020). 1차 미세먼지는 연료 연소시설로부터 직접 배출되는 형태로, 대기오염배출량을 산출할 수 있으나 2차 오염물질은 직접적인 배출량 산정이 어렵다.

2010년 서울에서 측정한 미세먼지 분석결과에 의하면, PM<sub>10</sub>의 약 20-40%가 2차 미세먼지로 구성되어 있어 1차 미세먼지 배출자료만으로는 배출원의 미세먼지 형성 기여도평가 및 농도변화 예측이 어려운 상황이다(Seoul Metropolitan Government, 2010). 또한 대기오염물질 배출량 통계에 따르면, 서울의 2014년 PM<sub>10</sub> 미세먼지 발생량은 97,918톤이 배출되었다고 보고되었으나 그 다음해인 2015년은 233,177톤으로 두 배가 넘는 배출량을 보여 미세먼지의 형성 및 배출억제에 관한 사회적 인식과 노력이 필요하다(Korea Environment Corporation, 2020).

PM<sub>10</sub> 이하의 미세분진은 장거리를 이동하기도 하며, 기상현상에 참여한 미세먼지는 건식 및 습식침전 과정을 거쳐 필연적으로 토양지하수 환경에 영향을 미치게 된다(Possanzini et al., 1988). 강수 시기의 미세먼지는 수증기와 결합 혹은 빔물과의 혼합에 의해 토양층 상부에 퇴적되며, 건조한 시기의 미세먼지는 중력낙하에 의한 침전 후 재부유가 일어나기도 한다(Aubrey et al., 2020). 국내의 미세먼지 연구에서도 PM<sub>2.5</sub> 이상의 입도를 가진 부유분진에서 K, Na, Mg, Ca, Fe, Al 등의 금속원소 성분이, 그리고 PM<sub>2.5</sub>에서 K와 Na 성분이 NO<sub>3</sub>와 SO<sub>4</sub> 성분과 높은 상관관계를 가지는 것이 보고된 바 있다(Choung et al., 2016). 특히 가을철 PM<sub>2.5</sub>는 Fe, Cu, Mn, Pb, Zn, Ni, Cr, Co, As, Cd, Sr, Rb, Sn, Sb 등의 금속 함량이 전함량의 50% 이상을 차지하는 것으로

분석되었다. 이와 같은 이유로, 다양한 경로로 토양 및 지하수계로 유입된 미세먼지에 의한 산성비, 금속성분 농축 및 토양 산성화 등의 가능성에 관한 우려가 제기되고 있다(Fig. 1). 토양에 유입된 금속성분은 침출, 침식, 식물 흡수 등의 기작으로 서서히 농도가 감소할 수 있으나, 인간의 시간관점(human time scale)에서는 영구적으로 축적된다고 할 수 있다. 특히 토양 심도에 따라 금속성분의 체류시간은 달라지는데, 표토 2.5 cm 깊이까지는 수십 년에 달하며 상부 20 cm 깊이의 토양은 수세기에 걸쳐 지속될 수 있다고 알려져 있다(Aubrey et al., 2020).

강이나 호수와 같은 수계에 편입된 부유분진 가운데 불용성분은 총용존고형물 형태로 남아있기도 하다. 이와 관련하여 정수시스템을 거쳐 가정으로 공급된 물을 초음파 가습기에 넣었을 때 실내의 미세먼지 수치가 매우 크게 상승하는 연구가 보고된 바 있다(Highsmith et al., 1992). 이는 수증기를 미세입자로 인식한 것이 아니라, 총용존고형물이 미세한 분진으로 실내 공기에 발산되는 문제를 제기한 것으로, 다양한 기원에서 형성된 미세먼지의 수계 유입 및 신체에 미치는 직접적 영향에 관해 활발한 연구가 이어지고 있다(Davis et al., 2016; Sain et al., 2018). 2005년 서울의 강남구에서 도로청소로 제거된 미세먼지의 양은 106톤으로, 차량배출 미세먼지가 전체의 0.56%를 차지한다고 보고되었다(Na et al., 2007). 도로비산 먼지에서 주목할 만한 성분은 납(Pb)과 아연(Zn)의 금속성분이었으며, 질소와 인 성분도 용출되어 이들이 강수와 함께 수계로 흘러들어갈 경우 총용존고형물, 금속먼지, 영양염류 등의 성분이 유입 및 퇴적된다고 할 수 있다. 이러한 이유로, 미세먼지의 농도변동에 대한 예측은 지질환경 분야 가운데 특히 토양 및 지하수에 지속적으로 미칠 수 있는

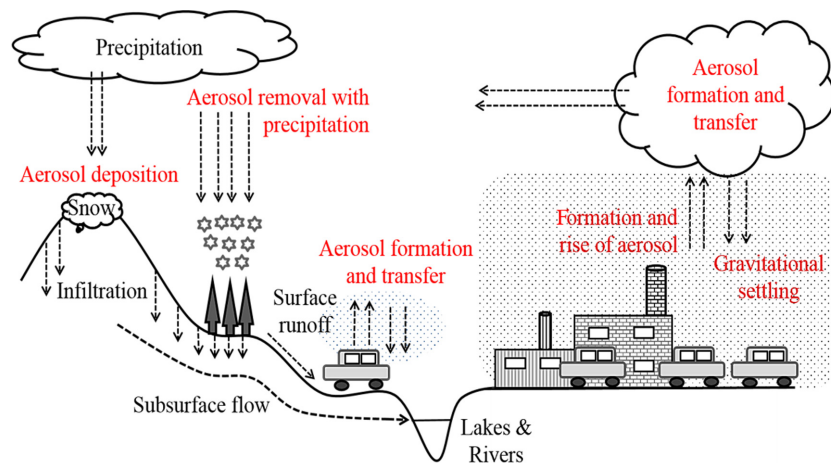


Fig. 1. Schematic diagram of the connection among air, surface water, soil, groundwater due to fine dust.

유해 금속성분의 퇴적 및 산성화 특성을 기능케 하는 기초자료가 될 수 있을 것이다.

도심지역에서 관찰되고 있는 고농도 미세먼지나 황사는 며칠씩 이어지는 경우도 있으나, 많은 경우 몇 시간 단위로 나타났다가 사라지는 경우가 상당수 발생하고 있다 (Park et al, 2012). 따라서 기존의 미세먼지 예보 시스템보다 더욱 시간적 조밀성이 높고, 지역적으로 세분되어 있는 예측모델이 필요하다고 여겨진다. 최근, 딥러닝 기술은 보다 빠르고 정확한 예측이 필요한 분야에서 널리 제시되고 사용되는 추세이다. 딥러닝 기술은 방대하게 축적된 데이터를 통해 스스로 학습을 수행한다는 장점을 갖고 있어, 다량의 데이터가 축적되는 빅 데이터 환경과 지속적으로 결합한다면 더욱 발전 가능성이 높은 기술이다. 이러한 딥러닝 기술을 활용한다면 기존에 비해 더 높은 정확도를 갖는 예측모델 설계가 가능하다. 이 연구에서는 딥러닝 기술을 사용하여 오존( $O_3$ ), 이산화질소( $NO_2$ ), 일산화탄소( $CO$ ), 아황산가스( $SO_2$ ) 등의 미세먼지 전구체를 통해  $PM_{10}$ 의 농도를 예측하는 모델을 제안하였다.

## 2. 연구지역

서울시는 한반도 중부에 위치하고 있으며, 연구지역인 관악구는 서울의 남서부에 위치하고 있으며 남측이 관악

산, 삼성산, 호암산으로 둘러싸여 있다(Fig. 2). 따라서 도심으로부터 형성 및 유입된 미세먼지 농도의 측정 및 변화 예측에 적절한 장소라고 판단하였다.

국내 미세먼지 농도는 통계적으로 겨울철에 높게 나타나는 것으로 보고되고 있으며(Korea Environment Corporation, 2020), 연구기간에 해당하는 2017년 11월부터 2018년 1월의 초미세먼지, 미세먼지, 강수의 관계에서도 12월과 1월에 집중적으로 관악구의 미세먼지 농도가 상승하는 것을 Fig. 3에서 확인할 수 있다(Korea Meteorological Administration, 2020; National Institute of Environmental Research, 2020). 강수사건이 일어나는 시기 혹은 그 이후에는 미세먼지와 초미세먼지의 농도가 감소하는 경향을 보였으며, 일부 시기에는 미세먼지 나쁨 기준인  $80-150 \mu g/m^3$  사이의 농도가 관측되기도 하였다. 그러나 연구지역의 겨울철 미세먼지( $PM_{10}$ )는 국내의 매우 나쁨 기준( $150 \mu g/m^3$  이상)을 넘어서지 않는 것을 관찰할 수 있다.

## 3. 연구방법

### 3.1. Convolutional Neural Network(CNN) 분석법과 활용

딥러닝의 학습모델로는 CNN(Convolutional Neural

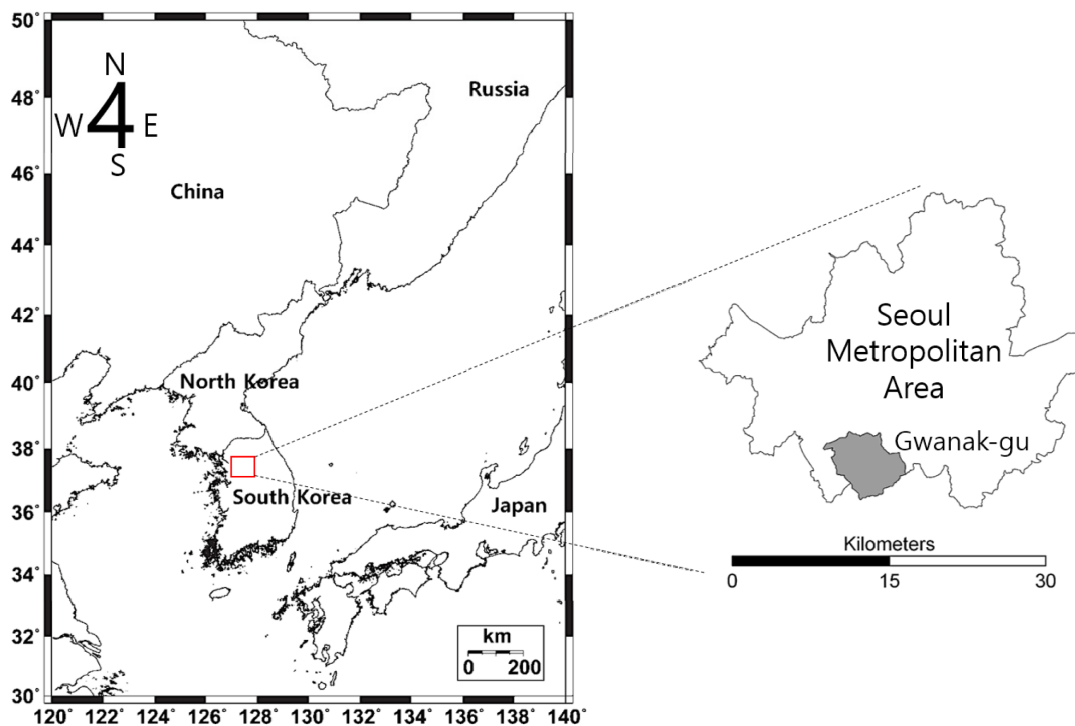


Fig. 2. Location of Gwanak-gu in the Seoul Metropolitan Area, Korea.

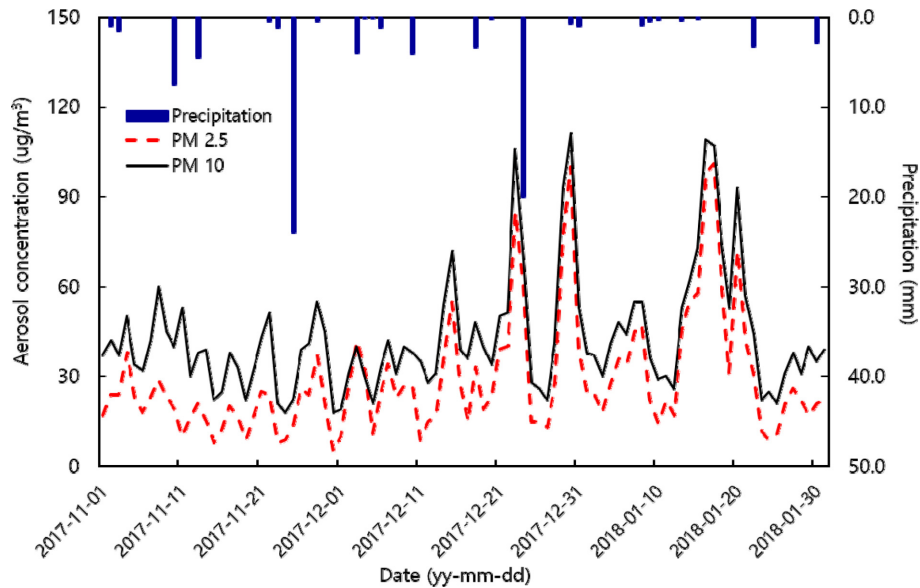


Fig. 3. Fluctuation of fine dust and ultra fine dust concentration according to precipitation.

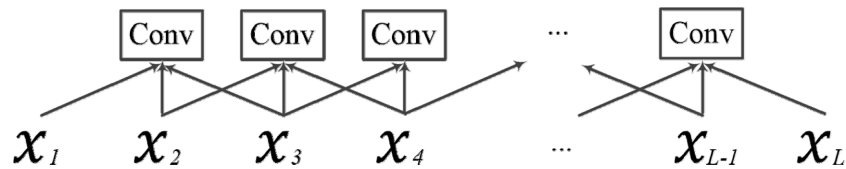


Fig. 4. An example image of a 1D-CNN with kernel size 3 and stride 1.

Network) 및 RNN(Recurrent Neural Network)이 혼합된 모델을 사용하였으며, 분석을 위해 한국환경공단에서 배포한 2014년 1월부터 2017년 12월 사이의 관악구 대기 측정자료를 사용하였다(Korea Environment Corporation, 2020).

CNN을 사용한 딥러닝 네트워크는 LeCun et al. (1998)이 제안한 LeNet-5 기법을 기본으로 하고 있으며, 예측의 오차율을 낮추고 정확도를 높이하고자 AlexNet, ZFNet, GoogLeNet, VGGNet 등과 같은 모델들이 지속적으로 제안되었다(Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Zeiler and Fergus, 2014; Szegedy et al., 2015). 최근에는 CNN 딥러닝 학습이 이루어지는 동안 역전파(backpropagation)가 진행되는 경우, 레이어가 깊어짐에 따른 기울기 소실(vanishing gradient)의 문제를 해결하기 위해 ResNet 모델이 제안되었다(He et al., 2016).

ResNet은 잔차 블록(residual block)을 만들어, 기울기 (gradient)가 건너뛰기 연결(skip connection)을 통해 흐르면서 정보전달이 더 잘 이루어질 수 있도록 한다. 건너뛰기 연결은 모델이 깊어져도 역전파를 통한 학습이 잘 이루어

진다는 것을 보여주었으며, 152개의 레이어로 구성된 모델을 통해 높은 정확도(Top-five error: 4.49%)를 얻는 것으로 보고되었다. 현재 이러한 기법은 이미지 분류기법 (image classification) 이외에도 물체 감지(object detection), 영상 의미론적 영역 분할(image segmentation), 영상 잡음제거(image denoising) 등과 같은 영역에서 널리 사용되고 있다. 하지만, 자연과학 영역에서 많이 사용되는 시계열 데이터의 경우, 이와 같은 2차원 컨볼루션 기법을 사용할 수 없다. 따라서 시계열 특성을 갖는 데이터에 딥러닝을 적용할 수 있도록 1차원 컨볼루션 기법이 제안되었다(Kim, 2014). 이 방법은 입력 자료가 컨볼루션 층을 통과하면서 데이터의 지역적 주요 특징들을 추출할 수 있도록 필터 값을 학습시킨다. 이후 이렇게 추출된 값은 예측모델의 입력 값으로 사용된다. Fig. 4는 kernel size=3, stride=1에 대한 예시이다. 이러한 1차원 컨볼루션은 날씨 예측, 반도체수율 예측 등과 같이 다양한 분야에서 널리 사용되고 있다(Haidar and Verma, 2018; Fu et al 2019).

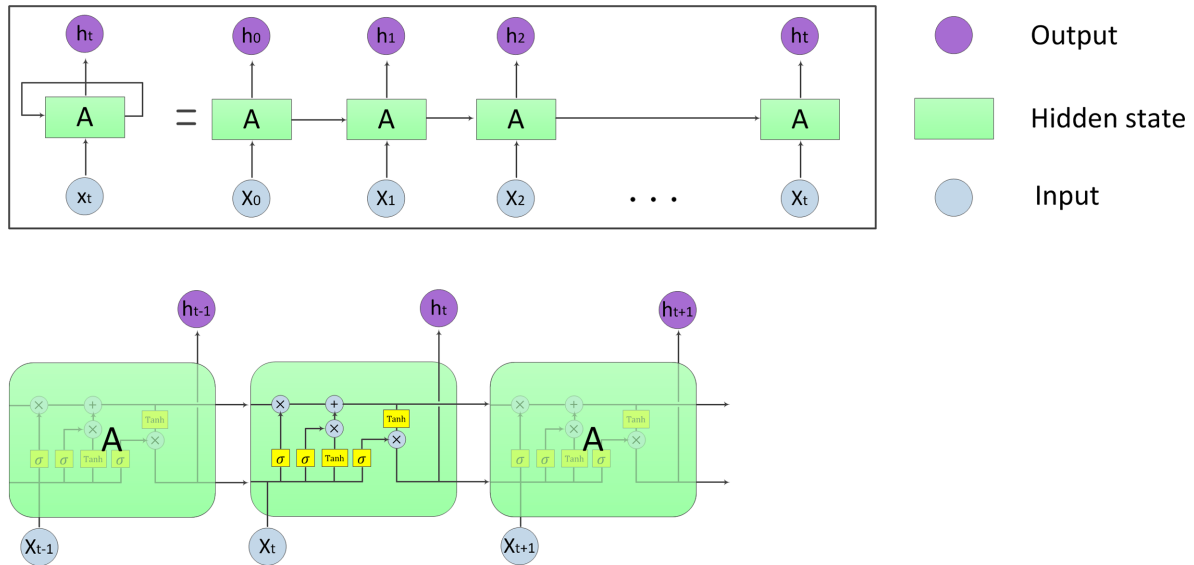


Fig. 5. Recurrent Neural Network (RNN) diagram explaining input, hidden state, and output.

### 3.2. Recurrent Neural Network(RNN) 분석법과 활용

RNN은 기억(hidden state, 기존 입력데이터를 요약한 정보)이 남아있다는 점에서 기존 신경망(neural network)들과 구별되며, 연속 자료(sequence data)를 모델링하는데 유용하게 사용된다. Fig. 5 (A)의 모식도에 나타난 것처럼, 첫 번째 입력이 들어오면 첫 번째 기억이 만들어진다. 두 번째 입력이 들어오면 기존의 기억과 새로운 입력을 참고하여 새 기억을 만든다. 즉, 입력의 길이만큼 이 과정을 계속해서 반복하는 것이 가능하다. 그러나 입력과 출력 사이의 거리(sequence)가 멀어질수록 관계를 학습하기 어려워지며, 출력과 가까운 입력 값의 영향이 가장 강하게 반영된다는 단점을 극복하기 위해 제안된 대표적 변형 모델로 LSTM(Long Short-Term Memory models)이 있다(Hochreiter and Shmidhuber, 1997). LSTM은 특별한 작업이 없어도 오랜 시간동안 정보를 기억하고 있으며, 단일 신경망 층(neural network layer)을 갖는 대신, Fig. 5 (B)와 같이 4개의 상호작용이 가능한 방식의 구조를 가진다. 또한, 입력, 과거 값, 출력에 각각의 가중치를 두고 학습이 이뤄진 더하기 연산으로 구성되어 학습 시 기울기 소실(vanishing gradient)문제를 피해갈 수 있다. 따라서 LSTM모델은 음성인식, 언어 모델링, 번역 및 이미지 캡셔닝 같은 여러 분야에서 사용되고 있다.

### 3.3. 딥러닝을 사용한 미세먼지 농도변화 예측

지금까지 수행된 미세먼지의 농도변화 예측은 주로 알고리즘 기반의 연구가 수행되어 왔으며, Ghassoun et al.

(2017)은 이차원 및 삼차원 파라미터 모델링 기법을 사용하여 도시의 형태학적 미세먼지를 예측하는 방법을 제안한 바 있다. 이차원 파라미터 모델에는 Open Street Map project data(OSM)를 사용하였고, 삼차원 파라미터 모델에는 CityGML-based 3D city model를 사용하였다. 아시아권에서는 Wang et al.(2018)이 NAQPMS(Nested Air Quality Prediction Model System)을 사용하여 동아시아 지역의  $PM_{2.5}$ 와  $PM_{10}$ 에 혼합된  $NO_3$ 의 농도를 예측하는 연구를 진행하였다. 최근, 4차 산업에 대한 국내외의 관심이 높아지며 기계학습과 딥러닝을 사용한 미세먼지의 예측에 대한 연구가 국내에서도 차츰 수행되고 있다. 한국의 대기를 대상으로 수행한 연구로는 Son et al. (2018)이 분산모델링(dispersion modeling)을 사용하여 대전의 미세먼지 농도를 예측하는 연구가 보고된 바 있다. Xayasouk and Lee(2018)는 MLP(Multi-Layer Perceptron) 신경망을 기반으로 풍속, 풍향과 기온, 날씨, 습도를 입력 데이터로 하여 대한민국 영역의  $PM_{2.5}$ 와  $PM_{10}$ 을 예측하는 기법을 제안하였다. Jeon and Son(2018)은 널리 알려진 머신러닝 기법인 SVM(Support Vector Machine), 다항 로지스틱 회귀모형, 랜덤 포레스트(Random Forest) 기법을 사용하여  $PM_{10}$ 의 농도를 예측하는 방법을 제안하였다. 또한, LSTM을 사용하여 서울 지역의 미세먼지의 오염도를 예측하는 기법도 Kim et al.(2019)에 의해 제안되었다.

딥러닝을 사용한 미세먼지의  $PM_{10}$ 에 대한 예측 기법은 현재까지 잘 알려진 수율 예측(yield prediction) 기법과



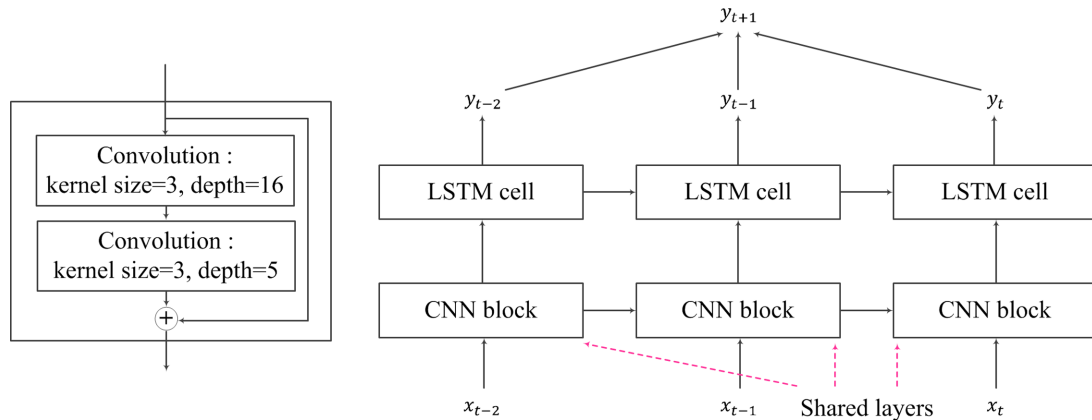


Fig. 6. A process image for feature map extraction using residual block.

유사하며, 정확도를 높이기 위해 데이터에서 전처리의 형태로 CNN기법의 잔차 블록을 적용한 뒤, 결과 값을 RNN의 입력변수로 넣어주었다. 먼저, Fig. 6 (A)에 도시한 것처럼 잔차 블록으로 특징 값(feature value)을 추출하였다. 여기서 자료의 입력 크기(input size)와 특징 값의 출력 크기(output size)를 맞춰주기 위해 자료의 크기(dimension)를 첫 번째 컨볼루션 필터에서 16으로 늘려준 뒤, 두 번째 컨볼루션 필터에서 입력 크기로 다시 줄여주었다. 컨볼루션을 통과한 특징 값은 각각 배치정규화(batch normalization)와 ReLU 활성화함수(Rectified Linear Unit)를 적용하였다. 이후 CNN을 통과한 값과 input data를 더하여 LSTM cell의 input으로 넣어주었다. 이 연구에서 사용한 전체적인 아키텍처(architecture)는 Fig. 6 (B)와 같다. Output의 크기는 PM<sub>10</sub>의 값을 예측하기 위해 1로 설정하였다.

#### 4. 서울시 관악구 미세먼지 농도예측 모델링 결과 및 고찰

관악구의 PM<sub>10</sub> 실측자료를 딥러닝 모델에 적용하여 얻은 예측 값에 대해 실제 값과 비교하고, 해당 예측 값의 타당성을 검증하였다. 국내 초미세먼지의 공식적인 관측은 2015년부터 시작되어 딥러닝을 적용하기에 자료의 수가 충분하지 않다고 여겨져 이번 학습에서는 초미세먼지 자료를 제외하고 사용하였다. 실험기간(2014-2017년)의 대기관측 데이터에서 80%는 학습용으로 설정하였고 20%는 실험용으로 설정하였다. 실험에서 사용한 시간순서의 길이(time sequence length)는 4로 선택하여, 4시간의 자료 가운데 앞의 3시간의 자료를 사용해 뒤의 1시간 값을 예측하도록 하였다. 다시 말해, 이전의  $[x_{t-2}, y_{t-2}]$ ,  $[x_{t-1}, y_{t-1}]$ ,

$[x_t, y_t]$ 의 데이터를 사용하여  $y_{t+1}$  값을 예측하도록 설계하였다. 여기서  $x_t$ 의 값으로 SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, PM<sub>10</sub>을 설정하였고( $i = t-2, t-1, t$ )  $y_t$ 값은 PM<sub>10</sub>으로 설정하였다( $i = t-2, t-1, t, t+1$ ). 또한, time length = 4를 만족하지 못하는 데이터는 제외하였다. 이렇게 선정된 데이터의 수는 총 30,806개이다.

학습 프로그램은 Python 2.7.12, Tensor Flow 1.12.0, Linux Ubuntu 16.04를 사용하였으며, CPU는 Intel Core i-5 6500, GPU는 Tesla K80을 사용하였다. 학습 파라미터 설정은 잔차 블록에서 입력 및 출력 정보의 크기(input/output shape)를 sequence length = 3, input dimension = 5로 하였다. LSTM cell 사이의 은닉층(hidden layer)의 크기는 128로 하였고, 망각게이트(forget gate)의 편향 값(bias)은 0.7로 하였다. 학습 시, 학습률(learning rate, lr)에 의해 성능이 저하되는 것을 방지하기 위해, 학습률 일정에 대한 파라미터 설정은 초기 학습률(initial lr) =  $10^{-2}$ 로 하였고, epoch이 100씩 반복될 때마다 lr 값을 반으로 줄여주었다. 전체 epoch은 1000으로 하였고, 손실함수는 RMSE(root mean square error), 최적화는 Adam을 사용하였다. 또한, 데이터의 패턴양상에 편향되는 것을 방지하기 위해, 학습 시, 1 epoch 이 종료될 때 마다 학습데이터를 혼합(shuffling)해 주었다. 이에 대한 학습 결과는 Fig. 7과 같다.

Epoch 별로 손실(loss) 값을 뽑았을 때, Fig. 7 (A)는 학습 손실(training loss)의 결과이고, Fig. 7 (B)는 실험 손실(test loss) 결과이다. 이들 그래프를 통해, epoch값이 증가함에 따라 손실 값이 감소하면서 미세먼지 농도예측을 위한 학습이 안정적으로 잘 진행되는 것을 알 수 있다. 실제 값과 예측 값의 분포를 보여주는  $y_{target}$ 과  $y_{prediction}$ 의 분포도는 Fig. 8과 같다. 가로축은

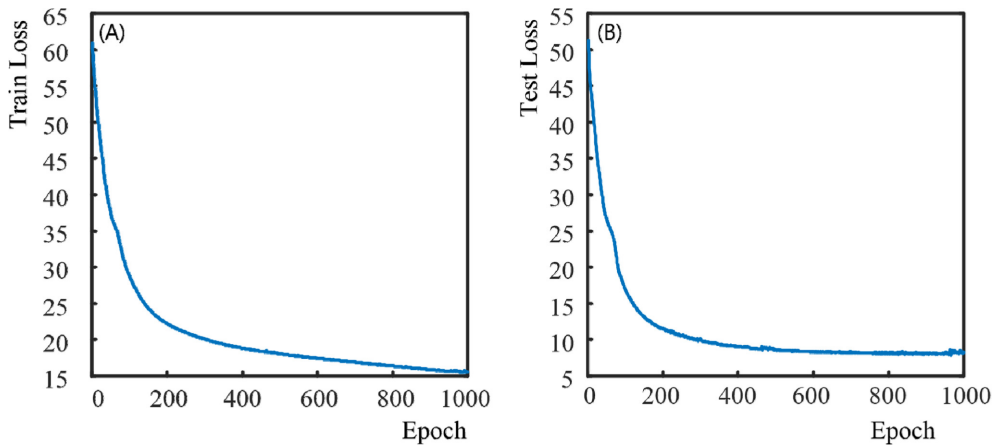


Fig. 7. Variation of train loss and test loss value according to Epoch number.

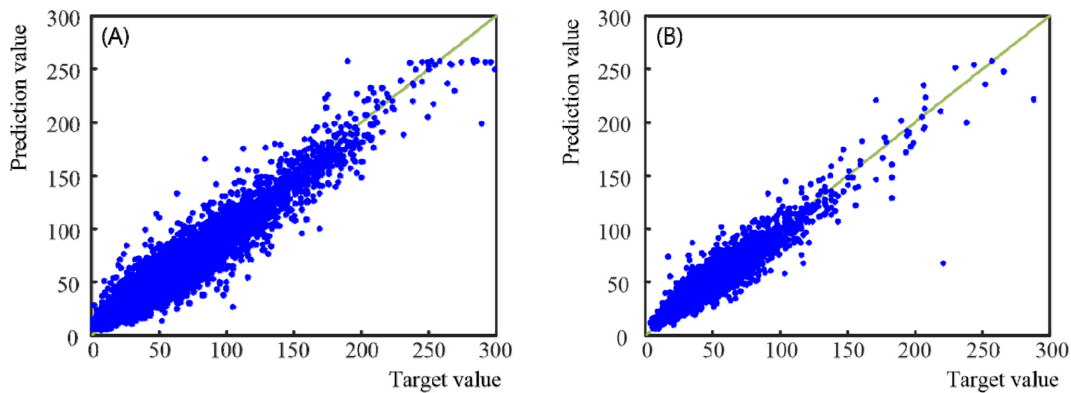


Fig. 8. Distribution diagram between label value and prediction value.

$y_{target}$ 으로 실제 관측한  $PM_{10}$ 의 값이고, 세로축의  $y_{prediction}$ 은 학습을 통해 예측된  $PM_{10}$ 의 값으로, 실제 값을 딥러닝 모델에 넣고 학습된 결과를 실제 자료와 비교하였다.

이 때, Fig. 8 (A)는 학습에 사용된 데이터의 실험결과이고, Fig. 8 (B)는 테스트에 사용된 데이터의 실험결과로서, 추세선  $y=x$  (연녹색 실선)을 표시하여 구분이 용이하게 하였다. 실험에 대한 평가는  $R^2$  (R squared), RMSE, MAPE, MAE 지표를 사용하였으며, 이에 대한 식은 각각 (1)부터 (4)와 같다.

$$R^2 = 1 - \frac{\sum (y_{tar} - y_{pred})^2}{(\sum y_{tar} - \overline{y_{tar}})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_{tar} - y_{pred})^2} \quad (2)$$

$$MAPE = \frac{1}{n} \sum \frac{1}{y_{tar}} |y_{tar} - y_{pred}| \times 100 \quad (3)$$

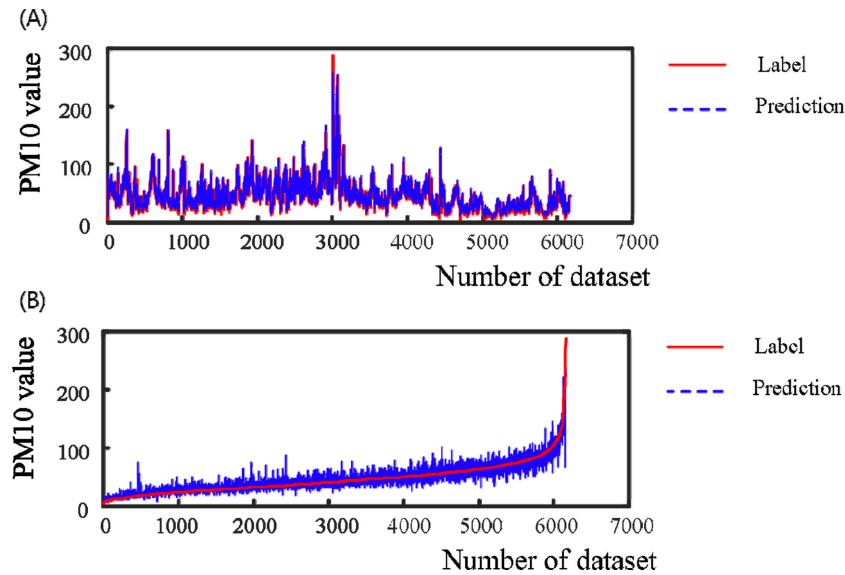
$$MAE = \frac{1}{n} \sum |y_{tar} - y_{pred}| \quad (4)$$

여기서,  $y_{tar}$ 는 실제  $PM_{10}$ 의 값을 의미하고,  $y_{pred}$ 는 예측된  $PM_{10}$ 의 값을 의미한다. 그리고  $n$ 은  $y_{tar}$ 의 데이터의 수,  $\overline{y_{tar}}$ 는  $y_{tar}$ 의 평균값을 의미한다. R squared의 값은 [0, 1]의 범위 내에서 값을 가지며 1에 가까울수록 높은 정확도를 의미한다. RMSE, MAPE, MAE 지표는 0에 가까울수록 높은 정확도를 의미한다. 이러한 지표를 사용하여 얻은 실험 결과는 다음 Table 1과 같다. 실험결과를 통해 얻은 값은 R squared, RMSE, MAPE, MAE의 4개 지표 모두에서 상당한 신뢰도를 가지는 것으로 여겨지며, 앞으로 다양한 기법의 적용을 통해 정확도와 신뢰도가 더욱 개선 가능할 것으로 여겨진다.

Fig. 9 (A)는  $PM_{10}$ 의 실제 값과 예측 값에 대하여 시간별로 도시한 것이며, Fig. 9 (B)는 실제 값을 오름차순으로 정리하여 미세먼지의 농도에 따른 예측 값과 비교한 자료이다. 적색 실선은 실제 값이고, 청색 파선은 예측 값

**Table 1.** Evaluation results of proposed model on test dataset

Metric	R squared	RMSE	MAPE	MAE
Gwanak-Gu, Seoul, Korea	0.8973	8.2450	14.850	5.8111

**Fig. 9.** Label value and prediction value over time (A) and label value with ascending order (B).

이다. 이를 통해, 적용한 딥러닝 학습기법은 예측 값에 비해 다소 편차가 존재하지만, 자료 전체적으로는 실제 값이 증가함에 따라 예측 값도 증가하는 추세를 확인할 수 있다.

딥러닝 기법에 기반을 둔 연구방법을 사용하였을 때, 모든 미세먼지 농도범위에서 높은 정확도를 가진 예측이 가능하였으며, 이에 근거하여 고농도 미세먼지 예측 뿐 아니라 일상적인 생활 예보에도 이 연구에서 제시한 학습모델의 적용이 가능할 것으로 생각된다. 또한, 충분한 자료가 존재하는 경우 다른 도심 지역에도 미세먼지 농도예측 기법의 적용이 가능할 것으로 생각된다.

## 5. 결 론

이 연구는 CNN 기법의 잔차 블록과, RNN 기법의 LSTM을 혼합 적용하여 도심지역 미세먼지(PM<sub>10</sub>)의 농도변동을 예측하는 딥러닝 모델을 제안하였다. 2014년부터 2017년 전 기간에 걸쳐 오존(O<sub>3</sub>), 이산화질소(NO<sub>2</sub>), 일산화탄소(CO), 아황산가스(SO<sub>2</sub>) 등의 미세먼지 전구체를 통해 PM<sub>10</sub>의 농도를 예측하고, 세분화된 시간 간격으로 미세먼지 예측 정확도를 높이하고자 하였다. 이 기법을 사용하여 얻은 결과에 대한 대표적 평가지표는 R squared 값으로 학습자료와 실제자료의 관계에 관하여 0.8973의 값

을 얻어, PM<sub>10</sub>의 예측 값과 실제 값이 잘 일치하는 것으로 나타났다.

미세먼지의 농도를 정확하게 예측할 수 있다면, 단시간 예측의 경우에는 예보를 통해 미세먼지 발생원의 활동을 억제하여 미세먼지의 직접적인 물리적, 화학적 피해를 최소화 할 수 있을 것이다. 일주일 이상의 장기 예측이 가능하게 된다면, 지역 혹은 국가적 차원에서 산업계와 연계하여 미세먼지 배출을 억제하는 등의 대비를 할 수 있다. 이와 함께 기상청의 예보자료를 함께 활용하여 고농도 미세먼지 예측이 있을 때 강수확률이 높은 경우, 산성비 대응, 토양산성화 혹은 금속성 분진의 수계유입 억제와 같은 지하수토양환경 측면의 대비를 할 수 있다.

이 연구는 미세먼지 예측 정확도를 높인 연구라는 점에서 의미가 있으며, 연구에 사용한 기법은 서울과 같은 도심지역 이외에도, 지형 및 배출원의 특성에 따라 지역을 세분화 하여 적용할 수 있다. 앞으로 장기간의 고농도 미세먼지 자료를 확보하게 된다면 미세먼지 매우 나쁨 이상의 농도(150 µg/m<sup>3</sup> 이상)에 관하여도 예측을 시도할 수 있을 것으로 생각된다. 딥러닝 학습의 특성 상, 더 많은 분량의 자료가 축적될수록 예측의 정확도가 높아질 것이며 국내의 미세먼지 및 그와 관련된 토양지하수 오염가능성 연구에 기여할 수 있을 것으로 기대된다.



## 사 사

이 논문은 2019년 교육과학기술부의 재원으로 한국연구재단 개인기초연구지원사업의 지원을 받아 수행한 연구 과제(도심지역 미세먼지의 수용성 금속 성분들에 의한 토양 및 지하수 오염경로 분석 및 특성평가, 과제번호 2018R1D1A1B07044596)입니다. 이 논문을 심사하고 세심한 조언을 주신 심사위원님에게 감사드립니다.

## References

- Aubrey, E.S., Andrew, D.G., Sara, P.E., and Howard, W.M., 2020, Mechanisms of children's soil exposure, *Current Problems in Pediatric and Adolescent Health Care*, **50**(1), 100742.
- Choung, S., Oh, J., Han, W.S., Chon, C.M., Kwon, Y., Kim, D.Y., and Shin, W., 2016, Comparison of physicochemical properties between fine (PM<sub>2.5</sub>) and coarse airborne particles at cold season in Korea, *Science of the Total Environment*, **541**, 1132-1138.
- Davis, M.J., Janke, R., and Taxon, T.N., 2016, Assessing inhalation exposures associated with contamination events in water distribution systems, *PLoS one*, **11**(12), e0168051.
- Fu, Q., Niu, D., Zang, Z., Huang, J., and Diao, L., 2019, Multi-stations' weather prediction based on hybrid model using 1D CNN and Bi-LSTM, Chinese Control Conference, IEEE, p.3771-3775.
- Ghassoun, Y. and Löwner, M.O., 2017, Comparison of 2D & 3D parameter-based models in urban fine dust distribution modelling. *Advances in 3D Geoinformation*, Springer, Cham, p.231-246.
- Haidar, A. and Verma, B., 2018, Monthly rainfall forecasting using one-dimensional deep convolutional neural network, *IEEE Access*, **6**, 69053-69063.
- He, K., Zhang, X., Ren, S., and Sun, J., 2016, Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, p.770-778.
- Highsmith, V.R., Hardy, R.J., Costa, D.L., and Germani, M.S., 1992, Physical and chemical characterization of indoor aerosols resulting from the use of tap water in portable home humidifiers, *Environmental Science & Technology*, **26**(4), 673-680.
- Hochreiter, S. and Schmidhuber, J., 1997, Long short-term memory, *Neural Computation*, **9**(8), 1735-1780.
- Jeon, S. and Son, Y.S., 2018, Prediction of fine dust PM<sub>10</sub> using a deep neural network model, *The Korean Journal of Applied Statistics*, **31**(2), 265-285.
- Kim, H., Zhang, Q., and Heo, J., 2018, Influence of intense secondary aerosol formation and long-range transport on aerosol chemistry and properties in the Seoul Metropolitan Area during spring time: results from KORUS-AQ, *Atmospheric Chemistry and Physics*, **18**(10), 7149-7168.
- Kim, S., Lee, J.M., Lee, J., and Seo, J., 2019, Deep-dust: Predicting concentrations of fine dust in Seoul using LSTM, arXiv preprint arXiv:1901.10106.
- Kim, Y., 2014, Convolutional neural networks for sentence classification, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, p.1746-1751.
- Korea Environment Corporation 2020, Air Korea, Statistics of air quality in 2014-2017, <http://www.keco.or.kr/>
- Korea Meteorological Administration 2020, Weather Information in 2017-2018, <https://www.weather.go.kr/>
- Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, p.1097-1105.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., 1998, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, **86**(11), 2278-2324.
- Na, Y.M., Yoo, K.Y., and Kim, I.S., 2007, The environmental pollutants removal effectiveness of street cleaning in Seoul, *Seoul Studies*, **8**(3), 85-103.
- National Institute of Environmental Research, 2020, National Air Pollutants Emission in 2017-2018, <https://www.nier.go.kr/>
- Ostro, B., Roth, L., Malig, B., and Marty, M., 2009, The effects of fine particle components on respiratory hospital admissions in children, *Environmental Health Perspectives*, **117**(3), 475-480.
- Park, S.M., Moon, K.J., Park, J.S., Kim, H.J., Ahn, J.Y., and Kim, J.S., 2012, Chemical characteristics of ambient aerosol during Asian dusts and high PM episodes at Seoul intensive monitoring site in 2009, *Journal of Korean Society for Atmospheric Environment*, **28**(3), 282-293.
- Possanzini, M., Buttini, P., and Di, P.V., 1988, Characterization of a rural area in terms of dry and wet deposition, *Science of the Total Environment*, **74**, 111-120.
- Rückerl, R., Schneider, A., Breitner, S., Cyrys, J., and Peters, A., 2011, Health effects of particulate air pollution: a review of epidemiological evidence, *Inhalation Toxicology*, **23**(10), 555-592.
- Sain, A.E., Zook, J., Davy, B.M., Marr, L.C., and Dietrich, A.M., 2018, Size and mineral composition of airborne particles generated by an ultrasonic humidifier, *Indoor Air*, **28**(1), 80-88.
- Seoul Metropolitan Government, 2010, *Characteristic study for detailed monitoring of the particulate matters (PM<sub>10</sub>) in the Seoul Metropolitan Area*. 80p.
- Simonyan, K. and Zisserman, A., 2014, Very deep convolu-

tional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Son, S., Kim, D., Kang, Y., Jeon, H., Kim, S., Cho, K., and Yu, J., 2018, Fine-resolution mapping of fine dust concentration in urban areas and population exposure analysis via dispersion modeling, preprint.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., and Rabinovich, A., 2015, Going deeper with convolutions, *IEEE Conference on Computer Vision and Pattern Recognition*, p.1-9.

Wang, Z., Pan, X., Uno, I., Chen, X., Yamamoto, S., Zheng, H., Li, J., and Wang, Z., 2018, Importance of mineral dust and anthropogenic pollutants mixing during a long-lasting high PM event over East Asia, *Environmental Pollution*, **234**, 368-378.

Xayasouk, T. and Lee, H., 2018, Air pollution prediction system using deep learning, *WIT Trans. Ecol. Environ*, **230**, 71-79.

Zeiler, M.D. and Fergus, R., 2014, Visualizing and understanding convolutional networks, (D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars eds.) In *Proc. European conference on computer vision*, p.818-833 (Springer, Cham).