



기술자료

TSN을 이용한 도로 감시 카메라 영상의 강우량 인식 방법

Rainfall Recognition from Road Surveillance Videos Using TSN

Li Zhun, 현중환, 최호진*
 한국과학기술원 전산학부

Zhun Li, Jonghwan Hyeon, Ho-Jin Choi*
 School of Computing, Korea Advanced Institute of Science and Technology (KAIST)

접수일 2018년 9월 19일
 수정일 2018년 10월 5일
 채택일 2018년 10월 22일

Received 19 September 2018
 Revised 5 October 2018
 Accepted 22 October 2018

*Corresponding author
 Tel : +82-(0)42-350-3561
 E-mail : hojinc@kaist.ac.kr

Abstract Rainfall depth is an important meteorological information. Generally, high spatial resolution rainfall data such as road-level rainfall data are more beneficial. However, it is expensive to set up sufficient Automatic Weather Systems to get the road-level rainfall data. In this paper, we propose to use deep learning to recognize rainfall depth from road surveillance videos. To achieve this goal, we collect a new video dataset and propose a procedure to calculate refined rainfall depth from the original meteorological data. We also propose to utilize the differential frame as well as the optical flow image for better recognition of rainfall depth. Under the Temporal Segment Networks framework, the experimental results show that the combination of the video frame and the differential frame is a superior solution for the rainfall depth recognition. The final model is able to achieve high performance in the single-location low sensitivity classification task and reasonable accuracy in the higher sensitivity classification task for both the single-location and the multi-location case.

Key words: Deep learning, Rainfall depth, Surveillance video, Temporal segment networks

1. 서 론

생물 지구 화학적 순환(Biogeochemical Cycle)은 생물권과 비생물권 사이 상호작용을 통해 지구 상의 물질이 이동하는 현상으로, 물의 순환(Oki and Kan-ae, 2006; Ramanathan *et al.*, 2001; Chahine, 1992), 탄소의 순환(Prentice *et al.*, 2001; Falkowski *et al.*, 2000), 질소의 순환(Galloway *et al.*, 2004), 산소의 순환(Walker, 1980) 등이 대표적이다. 그 중, 강우는 물의 순환 과정에서 매우 핵심적인 요소 중 하나이며, 이를 통해 바다의 물이 강, 지하, 호수 등으로 유입될 수 있다. 또한, 강우는 대기 중 미세 먼지 농도(Guo *et al.*, 2016; Kim *et al.*, 2014), 대기의 온도 감소 등 대기

의 상태에도 영향을 미친다고 알려져 있다. 이러한 강우와 관련된 대기 및 기후 연구를 수행하기 위해선 일정 시간 동안 일정한 곳에 내린 비의 분량을 나타내는 강우량을 지속적으로 관측 및 분석하는 과정이 매우 중요하다. 특히, 높은 공간 해상도의 강우량 정보(Berg *et al.*, 2016; Lobligois *et al.*, 2014)는 이러한 효과들에 대한 정량적인 분석에 도움을 줄 수 있으며, 이를 통해 보다 통찰력 있는 설명을 제공할 수 있다.

본 논문은 낮은 장비 비용으로 도로 수준의 강우량 정보를 생성하는 방법에 대해 논한다. 전통적으로, 기상청은 다수의 종관 기상 관측 장비(Automated Surface Observing System, ASOS)와 방재 기상 관측 장

비(Automatic Weather System, AWS)를 설치해 기상 정보를 수집한다. 하지만, 도로 수준의 기상 정보를 수집하기 위해 충분한 수의 기상 관측 장비를 설치하는 것은 큰 비용을 필요로 한다. 한편, 교통 상황을 감시하기 위해 도로에는 다수의 감시 카메라가 설치되어 있다. 만약 추가 비용 없이 도로의 감시 카메라 영상을 획득할 수 있고, 해당 영상으로부터 도로에 내리는 강우량을 예측할 수 있다면, 고가의 기상 관측 장비를 도로에 설치하지 않아도 도로 수준의 강우량 정보를 획득할 수 있을 것이다. 최근 Convolutional Neural Network (CNN) (LeCun *et al.*, 1998)와 같은 딥 러닝(Deep Learning) 모델이 다양한 작업에서 높은 성능을 보여주고 있다. 특히, 이미지와 영상을 분류하는 문제에서 매우 만족할 만한 성능을 보여주고 있기 때문에, 본 논문에서는 딥 러닝 모델을 활용해 도로의 감시 카메라 영상으로부터 강우량 인식을 수행하려 한다.

이러한 연구 목표를 달성하기 위해선, 영상으로부터 비의 특징(Feature)을 학습하고, 이러한 특징을 강우량으로 변환하는 방법의 제시가 필요하였다. 강우량 인식은 컴퓨터 비전(Computer Vision) 분야의 영상 분류/회귀 작업으로 생각할 수 있다. 현재 대부분의 영상 분류 연구는 영상의 이해에 집중되어 있으며 (Abu-El-Haija *et al.*, 2016; Soomro *et al.*, 2012), 해당 연구를 진행하기 위해 Single-Stream Network (Diba *et al.*, 2017; Qiu *et al.*, 2017; Donahue *et al.*, 2015; Tran *et al.*, 2015; Karpathy *et al.*, 2014)와 Two-Stream Network (Carreira and Zisserman, 2017; Wang *et al.*, 2016; Simonyan and Zisserman, 2014), 두 구조의 신경망이 사용되고 있다. 또한, 비와 관련된 기존의 컴퓨터 비전 연구는 강우 여부를 인식하고, 빗방울을 제거하는 것에 국한되어 있었다 (Kang *et al.*, 2012; Xue *et al.*, 2012; Shen and Xue, 2011; Garg and Nayar, 2004). Shi *et al.* (2015)은 레이더의 반향 데이터를 사용해 현재 강우량을 예측하는 Convolutional Long Short Term Memory (LSTM) Network를 제안하였다. 본 논문은 도로의 감시 카메라의 영상을 사용하기 때

문에 해당 연구와의 차별성이 존재한다. Shin *et al.* (2015)은 모서리(Edge) 픽셀의 개수를 활용해 비와 맑음 두 기상 상황을 구분하는 방법을 제안하였다. 해당 연구는 본 논문과 마찬가지로 감시 카메라의 영상을 활용하지만, 강우 여부 인식과 강우량 수치 인식이라는 목표에 차이가 존재한다. Bae *et al.* (2016)은 수동으로 강우의 특징을 정의해 감시 카메라 이미지의 픽셀 정보로부터 강우량을 판별하는 방법을 제안하였다. 해당 연구는 적은 수의 데이터를 사용하며, 강우량 판별 절차를 사람이 쉽게 이해할 수 있었다. 하지만 해당 연구에는 여전히 두 가지 문제점이 존재한다. 첫째, 해당 연구는 강우량 단계를 영상의 레이블로 정의했지만, 강우량 단계와 강우량 사이의 관계를 명확하게 언급하지 않았다. 둘째, 해당 연구는 영상 내 빗방울의 가시성(Visibility)에 의존하기 때문에, 빗방울이 뚜렷하지 않을 경우 판별 성능이 저하되었다. Dong *et al.* (2017) 또한 감시 카메라 영상 내 빗방울을 탐지해 강우량을 측정하는 방법을 제시하였다. 해당 연구는 강우량을 연속적인 수치로 측정할 수 있었지만, 추정 오차가 여전히 크다는 문제가 있었으며, 마찬가지로 빗방울의 가시성이 성능에 많은 영향을 주었다. Lee *et al.* (2018)은 Region Of Interest (ROI) 데이터 패턴이라는 새로운 특징을 정의하고, 해당 패턴의 군집화를 통한 강우량 추정 방법을 제안하였다. 또한, 해당 연구는 감시 카메라 영상 내 최적의 ROI를 탐색하기 위한 알고리즘을 제안하였다. 하지만, 다양한 환경에 대한 해당 알고리즘의 강건성(Robustness)은 여전히 불확실하다. 이러한 연구들과 비교했을 때, 본 논문의 방법은 영상 내 비의 특징을 자동으로 추출할 수 있으며, 단대단(End-to-End) 방식으로 영상의 강우량 인식을 수행할 수 있다는 장점이 있다. 하지만 본 논문의 방법은 학습을 위해 많은 수의 데이터를 필요하며, 자동으로 추출한 특징을 이해하기 어렵다는 단점이 있다. Ko *et al.* (2017)은 CNN-LSTM Network를 사용해 강우량을 판별하는 방법을 제안하였다. 해당 연구는 미리 학습된 CNN 모델을 사용해 영상 내 프레임의 특징을 추출하였고,

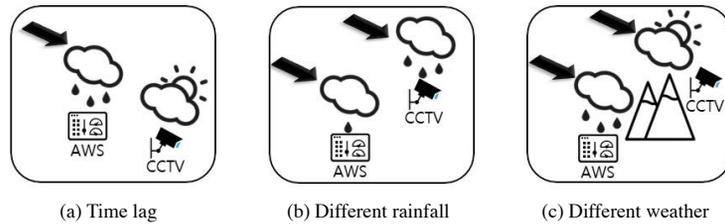


Fig. 1. Distance between surveillance camera and AWS/ASOS station leads to mismatch between video and rainfall depth.

시계열(Time Series) 분석이 가능한 LSTM (Hochreiter and Schmidhuber, 1997) 모델을 통해 프레임과 프레임 사이 관계를 학습하였다. 본 논문의 방법은 미리 학습된 CNN 모델을 사용하면서도, 해당 모델의 특징 추출 성능을 향상시키기 위해 미세 조정(Fine-tune) (Bengio, 2012; Caruana, 1995)을 수행하였다. 게다가, 본 논문의 방법은 영상의 RGB 프레임을 보완하기 위해 RGB 프레임 사이 차이와 광학 흐름(Optical Flow) (Gibson, 1950) 이미지를 활용하였다.

그 다음으로, 딥 러닝 모델을 학습하기 위해선 충분한 양의 검증된 데이터를 수집하는 방법의 제시가 필요하였다. 본 논문은 궁극적으로 고속도로에 설치된 감시 카메라의 영상을 수집하고, 감시 카메라 근처의 ASOS/AWS에서 측정된 강우량을 수집한 영상의 레이블로 취급해 강우량 인식 모델을 학습하려 한다. 하지만, 실제 고속도로 감시 카메라 영상의 해상도가 매우 낮기 때문에, 사람조차도 영상으로부터 날씨를 판별하기가 매우 어려웠으며, 감시 카메라가 설치된 장소에 강우량 측정 장치가 위치한 것이 아니기 때문에, 영상의 강우량 레이블을 정밀하게 생성하는 과정에도 많은 어려움이 있었다. 예를 들어, 그림 1(a)를 참고하면, 비 구름이 감시 카메라와 ASOS/AWS의 사이 직선 방향과 평행하게 움직일 경우, 이러한 거리가 시간 지연을 발생시킬 수 있었다. 그림 1(b)와 (c)의 경우, 비 구름이 감시 카메라와 ASOS/AWS 사이 직선 방향과 수직으로 움직이기 때문에, 두 지역의 강우량 혹은 날씨가 상이할 수 있었다. 또한, ASOS/AWS는 강우량을 0.1 또는 0.5mm 단위로 측정한다. 따라서 임의의 고정된 시간 간격 사이 강우량

을 정확하게 산출할 수 없었다.

이러한 문제를 해결하기 위해, 본 논문에서는 교내에 직접 감시 카메라를 설치하고, 감시 영상을 수집하였다. 그리고, 원시 기상 데이터로부터 더욱 정밀한 강우량을 추정하는 방법과, Temporal Segment Network (TSN) (Wang *et al.*, 2016) 딥 러닝 모델을 활용한 영상 내 강우량 인식 방법을 제안하였다. 본 논문은 세 종류의 입력 데이터를 준비하고, 다양한 데이터 확장(Data Augmentation) 기법을 적용해 영상 내 강우량 인식 실험을 수행하였다. 그 결과, 최종적으로 제안한 모델이 단일 지역 대상 낮은 정밀도 분류 실험에서 가장 높은 성능을 보였으며, 단일 및 다중 지역 대상 높은 정밀도 분류 실험에서도 상당한 정확도를 보였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 앞에서 언급한 문제의 해결 방법을 제시한다. 3장에서는 본 논문에서 수행한 실험의 결과를 소개하며, 4장에서 해당 결과에 대한 논의를 진행한다. 마지막으로, 5장에서는 본 논문의 결론을 논한다.

2. 실험 방법

2.1 비 특징 학습

영상에서 비의 특징을 학습하는 것은 매우 어려운 문제이다. 엄청난 폭우가 내리지 않는 한, 단일 프레임에서 비를 명확하게 관찰하기 어렵기 때문이다. 또한, 강우의 강도에 따라 다양한 밀도의 빗줄기가 한 프레임 내 다양한 구역에 존재할 수 있다. 따라서 본

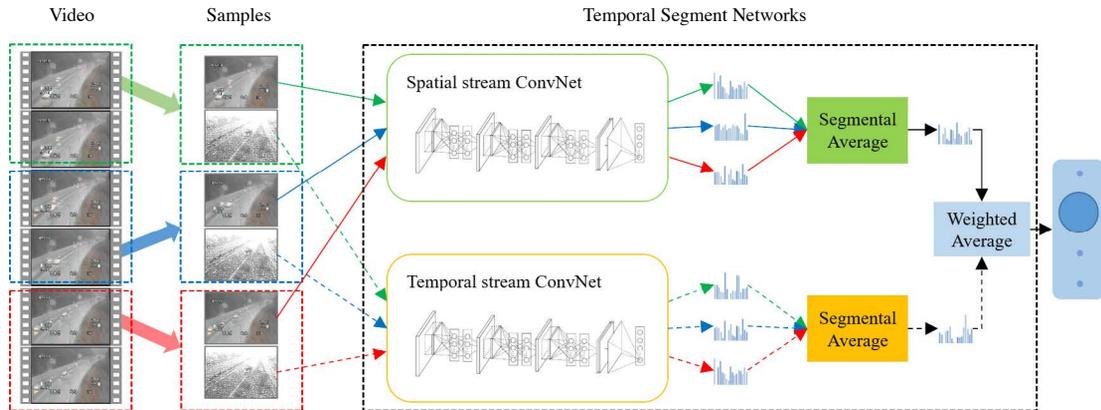


Fig. 2. The architecture of temporal segment networks.

논문에서는 복잡한 이미지에서도 특징을 효과적으로 학습할 수 있는 CNN 모델을 활용해 비의 특징을 학습하였다.

모델이 비를 더 효율적으로 학습할 수 있도록, 본 논문에서는 영상의 RGB 프레임과 함께 RGB 프레임 사이의 차이를 입력으로 사용하였다. RGB 프레임 사이의 차이는 영상 내 연속하는 두 프레임 사이의 차이를 계산하는 방식으로 생성하였다. 이러한 RGB 프레임 사이 차이를 사용해, 카메라의 움직임이 없는 환경에서 비와 배경의 분리를 수행하였다. 또한, 본 논문에서는 영상의 광학 흐름 이미지를 모델의 입력으로 사용하였다. 광학 흐름 이미지는 시간에 따라 변화하는 행동을 인식하기 위해 주로 사용되는 방법이다.

본 논문은 TSN 딥 러닝 모델을 사용해 강우량 인식을 수행하였다. 본 논문에서 TSN을 선택한 이유는 다음과 같다. 첫째, TSN은 3D CNN(Tran *et al.*, 2015) 및 다수의 다른 모델보다 데이터 효율적이다(de Souza, 2018). TSN은 2016년 이후로 영상의 행동 인식 분야에서 가장 높은 성능을 보이고 있는 딥 러닝 모델이다. 둘째, TSN은 Two-Stream 구조를 확장하고 있기 때문에 근본적으로 다양한 종류의 입력을 지원한다. 이를 활용해, 본 논문에서는 비를 효율적으로 탐지할 수 있다고 생각되는 RGB 프레임 사이 차이나

광학 흐름 이미지를 입력으로 사용하였다. 셋째, TSN은 배치 정규화(Batch Normalization)(Ioffe and Szegedy, 2015), Dropout(Srivastava *et al.*, 2014), 사전 훈련(Pre-training)(Erhan *et al.*, 2010; Bengio *et al.*, 2007) 기술을 적극적으로 사용한다. 본 논문에서는 사용 가능한 데이터의 수가 소량이었기 때문에, 과적합(Overfitting)을 방지하기 위해 이와 같은 기술이 매우 중요하였다.

강우량 인식 모델을 학습하기 위해, 본 논문은 그림 2와 같이 입력 영상을 동일한 길이의 세 조각으로 나누었다. 그 다음, 각 조각에서 무작위로 프레임 표본들을 추출하였다. 그 후, 기존의 Two-Stream 구조를 따라 공간 네트워크와 시간 네트워크를 각각 독립적으로 학습하였다. 본 논문에서는 표본들 사이의 평균을 통해 공간 및 시간 점수를 각각 결합하였다. 그 후, 최종적인 공간 및 시간 점수를 가중치 평균을 통해 융합하였고, Softmax를 적용해 강우량 레이블의 분포를 산출하였다.

2.2 영상 수집 및 강우량 레이블 할당

검증된 데이터를 수집하기 위해, 본 논문에서는 감시 카메라를 교내에 설치하고, 감시 영상을 수집하였다. 1장에서 언급한 이유로 인해, 감시 영상이 나타내는 정확한 강우량을 측정하는 것은 불가능하였다. 따

Original data
 Estimated data

Date & Time	1-Minute Rainfall Depth (mm)	Is Raining or Not (Yes: 10, No: 0)	1-Minute Rainfall Depth (mm)	5-Minute Rainfall Depth (mm)
4/18/2017 17:01	0.5	10	0.08	
4/18/2017 17:02	0	10	0.17	
4/18/2017 17:03	0	10	0.17	
4/18/2017 17:04	0.5	10	0.17	
4/18/2017 17:05	0	10	0.13	0.71
4/18/2017 17:06	0	10	0.13	0.75
4/18/2017 17:07	0	10	0.13	0.71
4/18/2017 17:08	0.5	10	0.13	0.67
4/18/2017 17:09	0	10	0.03	0.53
4/18/2017 17:10	0	10	0.03	0.44
4/18/2017 17:11	0	10	0.03	0.34
4/18/2017 17:12	0	10	0.03	0.25
4/18/2017 17:13	0	10	0.03	0.16
4/18/2017 17:14	0	10	0.03	0.16
4/18/2017 17:15	0	10	0.03	0.16
4/18/2017 17:16	0	10	0.03	0.16
4/18/2017 17:17	0	10	0.03	0.16
4/18/2017 17:18	0	0	0	0.13
4/18/2017 17:19	0	0	0	0.09
4/18/2017 17:20	0	0	0	0.06

Fig. 3. Estimate 5-minute rainfall depth.

라서 본 논문에서는 원시 기상 데이터에 통계적인 보정을 적용해 더 정밀한 강우량을 추정하는 방법을 다음과 같이 제시하였다. 본 논문의 실험은 이렇게 추정한 강우량을 영상의 정답(Ground Truth) 레이블로 사용하였다.

단계 1: 교내에 설치한 감시 카메라와 가장 가까운 ASOS/AWS에서 측정한 강우량 데이터를 수집하였다. 본 논문의 경우, 감시 카메라와 650 m 거리에 위치한 대전 지방 기상청의 원시 기상 데이터를 사용하였다. 해당 원시 기상 데이터는 기상청의 기상자료개방포털에서 내려 받을 수 있었다.

단계 2: ASOS/AWS에서 측정한 1분 강우량과 강우 유무를 사용해, 통계적으로 1분 강우량을 추정하였고, 그 후 더욱 정확한 5분 강우량을 계산하였다. 그림 3을 참고하면, ASOS/AWS에서 측정한 1분 강우량은 대부분 0을 보고하고 있었다. 하지만, 강우는 연속적으로 발생하기 때문에 보고된 수치가 실제 상황과

일치하지 않았다. 이는 용기가 가득 찰 경우에만 강우량을 보고하는 우량계의 작동 방식과도 관련이 있었다. 이러한 문제를 해결하기 위해, 본 논문에서는 0이 아닌 1분 강우량이 보고된 시점에서, 이전에 처음으로 강우량이 0으로 보고된 시점 사이의 1분 강우량을 평균해, 실제 1분 강우량을 추정하였다. 또한, AWS의 경우 강우 유무를 보고하기 때문에, 강우가 멈춘 시점을 결정할 때 해당 정보를 활용하였다. 그 후, 추정한 1분 강우량을 5분 단위로 누적해, 최종 5분 강우량을 계산하였다.

단계 3: 추정한 5분 강우량을 수동으로 검증하였다. 감시 카메라와 가장 가까운 ASOS/AWS에서 측정한 강우량을 사용하였지만, 두 지점 사이의 거리로 인해 영상과 강우량 사이의 불일치가 발생할 수 있었다. 게다가, 원시 기상 데이터가 오류를 포함하는 경우도 존재하였다. 따라서 본 논문에서는 모든 영상을 확인하며, 영상과 강우량이 명백히 불일치한 경우를

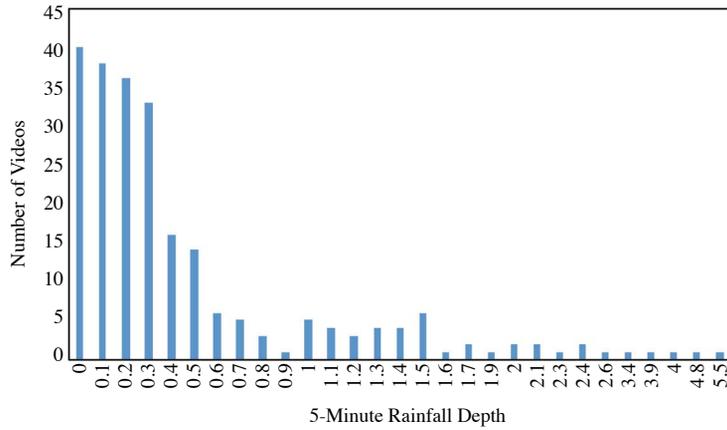


Fig. 4. Rainfall depth distribution.

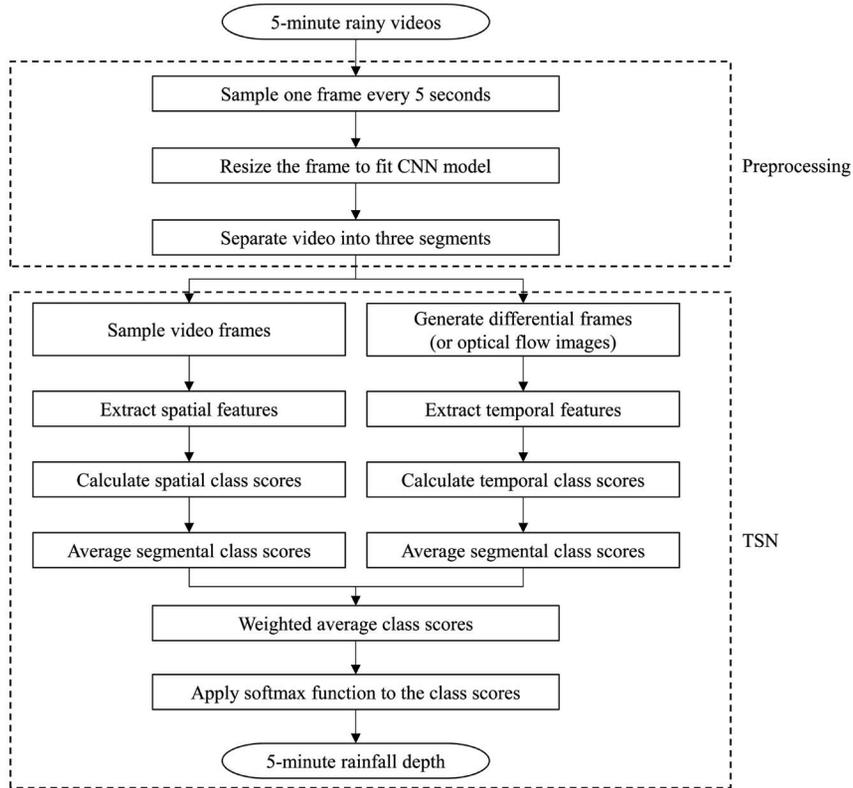


Fig. 5. Rainfall depth recognition flow chart.

제외하였다.

본 논문은 강우량 판별 문제를 분류 방식으로 해결하고자 하였으며, 따라서 입력 영상을 5분 강우량으

로 분류하였다. 이를 위해, 본 논문은 200개 이상의 5분 길이 감시 카메라 영상을 무작위 추출하였다. 그림 4는 추출한 5분 길이 감시 카메라 영상의 강우량

레이블 분포를 보여준다. 참고로, 0 mm는 비가 방금 멈추었음을 의미한다. 그림 4를 보면, 높은 강우량 레이블에 속한 영상보다 낮은 강우량 레이블에 속한 영상의 개수가 더 많음을 확인할 수 있었다. 딥 러닝 모델을 학습하기 위해선 많은 수의 데이터가 필요하기 때문에, 본 논문에서는 0, 0.1, 0.2, 0.3, 0.4, 0.5 mm 5분 강우량 레이블에 속한 영상만을 사용하였다.

2.3 강우량 인식

본 논문의 방법은 영상을 대상으로 하기 때문에, 영상 내 물체, 카메라의 각도, 확대 여부, 주변 밝기 등 영상의 프레임에 영향을 미치는 많은 요인들이 강우량 인식에 영향을 줄 수 있었다. 이러한 요인들의 영향을 줄이며 궁극적으로는 데이터의 규모를 확장하기 위해, 본 논문에서는 올려내는 위치, 확대 비율, 밝기 변화를 통해 새로운 영상의 프레임을 합성하는 데이터 확장을 수행하였다. 참고로, 안개와 같이 본 논문에서 다루지 않은 많은 변수가 여전히 존재함을 유의해야 한다.

본 논문은 TSN의 원 논문에서 언급된 다양한 기술을 사용해 TSN 모델의 학습을 진행하였다(Wang *et al.*, 2016). 준비한 감시 카메라 영상 및 강우량 레이블을 사용해 TSN 모델의 학습을 완료하면, 해당 모델은 그림 5의 절차에 따라 새로운 영상의 강우량을 인식할 수 있었다.

3. 실험 결과

3.1 데이터

본 논문은 0, 0.1, 0.2, 0.3, 0.4, 0.5 mm 5분 강우량 레이블의 영상을 대상으로 단일 지역에서 실험을 진행하였다. 표 1은 강우량 레이블에 따른 단일 지역에서 수집한 감시 영상의 개수를 나타낸다. 해당 표를 확인하면, 수집한 영상의 개수가 딥 러닝 모델을 처음부터 학습하기에는 충분하지 않았다. 따라서, 본 논문에서는 수집한 영상에 데이터 확장을 수행하고, 미세 조정 방식으로 딥 러닝 모델을 학습하였다. 준비한

Table 1. Single-location dataset overview.

Camera location	Label	5-minute rainfall depth (mm)						
		0	0.1	0.2	0.3	0.4	0.5	Total
Three way intersection		40	38	36	33	16	14	177

Table 2. Surveillance video specification.

Item	Value
Frame size	1280 × 720
Video length	5 minutes
Frames per second (fps)	30
Number of frames	9000

데이터 중 20%는 모델의 성능 평가를 위한 시험용, 15%는 모델의 초모수(Hyperparameter) 탐색을 위한 검증용, 나머지 65%는 학습용으로 사용하였다.

표 2는 본 논문에서 수집한 영상의 규격 및 상세 정보를 나타낸다. 일반적으로 영상 분류에 자주 사용되는 UCF101(Soomro *et al.*, 2012) 및 YouTube-8M(Abu-El-Haija *et al.*, 2016) 데이터와 비교했을 때, 본 논문에서 사용하는 데이터가 시간적 규모와 공간적 규모에서 모두 거대했다.

본 논문에서는 수집한 데이터에 다음과 같은 전처리 기술을 적용하였다. 우선, 입력 영상에서 매 5초 간격마다 한 프레임을 무작위로 추출하였다. 이러한 과정을 반복해 각 영상마다 N개의 표본을 추출하였다. N은 각 영상마다 독립적으로 결정되는데, 이는 각 강우량 레이블에 해당하는 프레임의 개수를 균등하게 분포하기 위함이었다. RGB 이미지와 RGB 프레임 사이 차이 입력을 사용할 때, 이미 학습된 CNN 모델의 입력 크기에 맞도록 이미지의 크기를 조절하였다. 광학 흐름 입력을 사용할 때, 영상 내 고정된 작은 영역을 발췌하였고, 해당 영역에 TVL1 알고리즘(Zach *et al.*, 2007)을 사용하였다. 작은 영역을 발췌하는 이유는 계산 복잡도를 줄이기 위함이었다.

또한, 본 논문에서는 소규모의 다중 지역 감시 카메라 영상 데이터에 대해서도 강우량 판별 실험을 수행하였다. 해당 데이터는 한국도로공사에서 제공한 실제 고속도로 감시 카메라 영상으로 구성하였다. 본

Table 3. Multi-location dataset overview.

Camera location	Label	5-minute rainfall depth (mm)				Total
		0	0.1	0.2	0.3	
Gangneung service area (To Incheon)	-	7	3	10	20	
Gyohyang 1st bridge	8	5	17	6	36	
Munmak 1	-	10	6	3	19	
Seongsan 2nd bridge	6	4	-	8	18	
Soksa	2	5	6	11	24	
Wonju	-	8	2	6	16	
Total		16	39	34	44	133

논문에서는 해당 영상의 강우량 레이블을 상기 제안한 방법을 통해 할당하였다. 표 3은 각 지역 및 강우량 레이블에 따른 영상의 개수를 나타낸다. 해당 영상의 규격 및 상세 정보는 단일 지역 영상 데이터와 동일하였다.

3.2 단일 지역 대상 낮은 민감도 분류

이 절에서는, 앞서 언급한 데이터를 사용해 단일 지역을 대상으로 5분 강우량을 0.1, 0.3, 0.5 mm 세 레이블로 분류하는 실험을 수행하였고 성능을 평가하였다. 본 실험에서는 미리 학습된 Inception-v2 (Szegedy *et al.*, 2016)를 영상 내 프레임의 특징 추출을 위한 CNN 모델로 사용하며, 3개의 영상 조각을 사용해 해당 모델을 미세 조정하였다. 그 후, 15개의 영상 조각을 사용해 모델을 시험하였다. 입력 종류와 조합에 따른 강우량 인식 정확도는 표 4에 요약되어 있다.

표 4를 참고하면, 세 가지 입력 종류 중 RGB 이미지를 입력으로 사용했을 때 가장 높은 성능을, 광학 흐름을 사용했을 때 가장 낮은 성능을 관찰할 수 있었다. RGB 이미지와 RGB 프레임 사이 차이를 모두 입력으로 사용한 경우, 인식 성능이 94.4%로 크게 상승하였다. 이는 RGB 이미지와 RGB 프레임 사이 차이가 서로 보완하는 정보를 함유한다는 것을 의미한다. 또한, RGB 이미지와 광학 흐름을 모두 입력으로 사용한 경우에도 성능이 90.9%로 개선되었다. 세 가지 종류의 입력을 모두 사용한 경우, 95.8%의 정확도를 기록하였다. 본 실험에서는 여러 종류의 입력을

Table 4. Comparison of input modality for 0.1, 0.3, 0.5 mm 5-minute rainfall depth classification using single-location dataset.

Input modality	Accuracy
RGB image	82.1%
RGB difference	72.6%
Optical flow	59.1%
RGB image + RGB difference	94.4%
RGB image + Optical flow	90.9%
RGB image + RGB difference + Optical flow	95.8%

Table 5. Comparison of input modality for 0, 0.1, 0.2, 0.3 mm 5-minute rainfall depth classification using single-location dataset.

Input Modality	Accuracy
RGB image	69.6%
RGB difference	64.7%
Optical flow	50.7%
RGB image + RGB difference	70.8%
RGB image + Optical flow	68.5%
RGB image + RGB difference + Optical flow	69.1%

사용할 때, RGB 이미지, RGB 프레임 사이 차이, 광학 흐름에 각각 1, 3, 3의 가중치를 부여하였다.

3.3 단일 지역 대상 높은 민감도 분류

이 절에서는, 단일 지역을 대상으로 5분 강우량을 0, 0.1, 0.2, 0.3 mm 네 레이블로 분류하는 실험을 수행하였다. 해당 실험은 이전 절의 실험보다 더 높은 민감도로 강우량 분류를 시도하였다. 입력 종류와 조합에 따른 강우량 인식 정확도는 표 5에 요약되어 있다.

표 4에 나타난 이전 실험과 유사하게, RGB 이미지는 가장 높은 성능을, 광학 흐름은 가장 낮은 성능을 기록하였다. RGB 이미지와 RGB 프레임 사이 차이를 모두 입력으로 사용할 경우, 강우량 인식 성능이 70.8%로 상승하였다. 하지만 이전 실험과 비교했을 때, 전반적으로 낮은 성능을 기록하였으며, 광학 흐름을 추가 입력으로 사용해도 성능이 개선되지 않았다.

3.4 다중 지역 대상 분류

이 절에서는, 본 논문에서 제안한 방법을 더욱 일

반화된 환경에 적용하였다. 이를 위해, 다중 지역을 대상으로 5분 강우량을 0, 0.1, 0.2, 0.3 mm 네 레이블로 분류하는 실험을 수행하였다. 다중 지역 대상 데이터의 규모가 매우 작기 때문에, 이전 절의 단일 지역 대상 모델을 미세 조정하는 방식으로 다중 지역 대상 모델을 학습하였다. 해당 실험은 광학 흐름 이미지를 사용하지 않았으며, 실험의 결과는 표 6에 요약되어 있다.

단일 지역 대상 실험과 다르게, 해당 실험은 RGB 프레임 사이 차이가 높은 성능을 기록하였다. 매우 다양한 배경의 영상으로부터 강우량을 인식하는 것은 더 어려운 문제이다. 따라서, 배경 정보가 제거된 RGB 프레임 사이 차이가 더 높은 성능을 보였다. 단일 지역 대상 실험과 마찬가지로, RGB 이미지와 RGB 프레임 사이 차이를 모두 입력으로 사용할 경우 78.9%로 가장 높은 강우량 인식 성능을 확인할 수 있

었다.

더욱 자세하게 실험의 결과를 분석하기 위해, 본 논문에서는 각 입력 종류에 따라 오차 행렬(Confusion Matrix)을 계산하였다. 그림 6을 확인하면, RGB 이미지는 두 번째 레이블에서 높은 성능을, RGB 프레임 사이 차이는 첫 번째와 네 번째 레이블에서 높은 성능을 기록하였다. 즉, 두 입력이 서로 상호 보완 관계임을 알 수 있었다. 따라서 RGB 이미지와 RGB 프레임 사이 차이를 모두 사용할 경우 성능 향상을 확인할 수 있었다. 추가로, RGB 이미지와 RGB 프레임 사이 차이 모두 세 번째 레이블에서 낮은 성능을 보였다. 이는 세 번째 레이블(0.2 mm)이 인접한 두 레이블(0.1 mm 및 0.3 mm)과 매우 유사하기 때문에 발생하는 것으로 보인다. 참고로, 첫 번째 레이블(0 mm)은 두 번째 레이블(0.1 mm)과 명백한 구분이 가능하기 때문에 해당 문제가 발생하지 않았다.

Table 6. Comparison of input modality for 0, 0.1, 0.2, 0.3 mm 5-minute rainfall depth classification using multi-location dataset.

Input modality	Accuracy
RGB image	66.2%
RGB difference	73.0%
RGB image + RGB difference	78.9%

4. 고 찰

서론에서 언급한 문제로 돌아와서, 본 논문의 주요 내용은 영상 내 비의 특징을 학습하고, 이 특징을 강우량으로 변환하는 방법이다. 본 논문에서 진행한 실

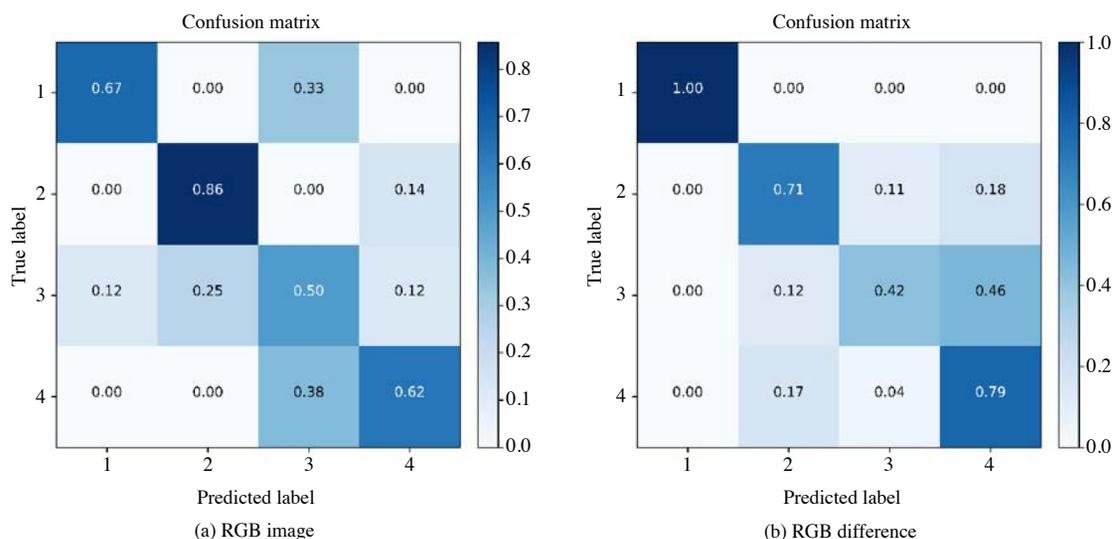


Fig. 6. Confusion matrix.

험은 TSN이 단대단으로 비의 특징을 자동으로 학습하고, 해당 특징을 강우량으로 대응시킬 수 있음을 보여주었다. RGB 프레임 사이 차이와 광학 흐름 이미지는 단일 RGB 이미지를 보완할 수 있는 매우 중요한 입력이다. 그러나 행동 인식 태스크에서 팔목할 만한 성능 향상을 보여주었던 것과 달리, 강우량 인식 문제에서는 높은 성능 향상을 보여주지 못하였다. TVL1 알고리즘 대신, Gunnar Farneback (2003)의 알고리즘으로 광학 흐름을 계산하고 추가 실험을 진행했지만, 유사한 결과를 확인할 수 있었다. 이는 빗방울이 매우 작고 빠르게 움직여, 정확한 추적이 어렵기 때문이라고 생각된다. 성능과 계산량을 고려했을 때, RGB 이미지와 RGB 프레임 사이 차이를 입력으로 사용했을 때 가장 효율적으로 강우량을 인식할 수 있었다.

그 다음으로, 본 논문은 딥 러닝 모델을 학습하기 위한 충분하고 검증된 데이터 수집 방법을 제시하였다. 교내에 감시 카메라를 설치해, 본 논문에서는 감시 영상을 낮은 비용으로 수집할 수 있었다. 현재 사용 가능한 데이터의 규모는 작지만, 장기적인 관점에서 향후 더 많은 데이터를 수집할 수 있을 것이다. 하지만, 수집한 영상의 정답 강우량 레이블이 정확하지 않다는 문제가 있었다. 비록 본 논문의 경우, 기상 관측 장비까지의 거리가 상대적으로 짧은 650m이었지만, 감시 카메라와 ASOS/AWS 사이의 거리는 여전히 영상과 강우량의 불일치를 일으킬 수 있었다. 또한, 영상을 여러 조각 또는 여러 프레임으로 분해하는 과정은 거짓 강우량 레이블 할당을 발생시킬 수 있었다. 본 논문에서 제시한 강우량 레이블 할당 방법은 이러한 불일치를 완화할 수 있었지만, 문제를 근본적으로 해결하지는 못하였다. 이 문제를 매우 확실하게 해결하기 위해서는 감시 카메라와 동일한 위치에 더욱 정밀한 단위로 강우량을 측정할 수 있는 기상 관측 장비를 설치하는 것이 필요하다. 일부 연구는 딥 러닝이 레이블의 노이즈에 매우 강인하다고 보고하고 있다(Rolnick *et al.*, 2017). 따라서, 본 논문은 추정된 강우량 레이블로부터 유용한 강우량 인식 모델을

학습할 수 있었다. 레이블의 노이즈가 얼마나 모델의 성능에 영향을 미치는지는 불확실하지만, 정확하지 않은 레이블을 사용해 학습하는 것이 정확한 레이블을 사용해 학습하는 것보다 어렵다는 것은 매우 확실하다. 다시 말해서, 본 논문에서 제안한 방법은 더욱 정확한 강우량 레이블을 사용한다면 더욱 높은 성능을 달성할 수 있을 것이다. 본 논문에서는 해당 요인이 낮은 민감도의 분류에서 더 높은 성능을 보이는 이유 중 하나라고 생각한다.

5. 결 론

본 논문에서는 딥 러닝을 이용하여 도로의 감시 영상으로부터 강우량을 인식하는 방법을 제안하였다. 목표를 달성하기 위해, 본 논문은 새로운 데이터를 수집하고, 원시 기상 데이터로부터 정제된 강우량 계산 방법을 제안하였다. 또한, 영상의 프레임 사이 차이와 광학 흐름 이미지를 활용한 더 높은 성능의 강우량 인식 방법을 제안하였다. 본 논문은 TSN 모델을 사용해 다수의 비교 실험을 진행하였다. 실험의 결과는 RGB 이미지와 RGB 프레임 사이 차이 조합을 입력으로 사용할 때 가장 효과적으로 강우량 인식 문제를 해결할 수 있음을 보였다. 해당 조합은 낮은 민감도 실험에서 94.4%의 정확도를, 높은 민감도 실험에서 70.8%의 정확도를, 다중 지역 대상 실험에서 78.9%의 정확도를 기록하였다. 이는 본 논문에서 제시한 방법이 도로의 감시 카메라 영상으로부터 70% 이상의 정확도로 강우량을 인식할 수 있음을 의미한다. 또한, 영상 사이의 강우량 레이블 간격이 증가할수록 강우량 인식 정확도가 증가함을 확인할 수 있었다.

본 논문은 작은 범위의 강우량을 대상으로 상대적으로 적은 수의 데이터를 사용해 딥 러닝 모델을 학습하고 평가하였다. 따라서, 제시한 방법이 카메라가 움직이는 환경과 넓은 범위의 강우량에서도 여전히 효과적이지 확인할 필요가 있다. 또한, 본 논문에서는

TSN 이외에 영상 분류를 위해 사용하는 다양한 딥러닝 모델을 적용해 비교 평가를 수행하지 못하였다. 따라서 향후, 본 논문의 저자들은 I3D (Carreira and Zisserman, 2017), P3D (Qiu *et al.*, 2017) 같은 최근에 떠오르고 있는 다양한 모델을 강우량 인식에 적용해 볼 예정이다. 또한, 영상 내 영역에 따른 효과(ROI)를 분석하기 위한 실험을 진행할 예정이다.

감사의 글

이 연구는 기상청 「기상·지진See-At기술개발연구」(KMI2017-00410)의 지원으로 수행되었습니다.

References

- Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., Vijayanarasimhan, S. (2016) Youtube-8m: A large-scale video classification benchmark, arXiv preprint arXiv:1609.08675.
- Bae, G., Kim, J., Kwon, E., Lee, J., Hong, B., Kim, K. (2016) A method for distinguish rainfall by using pixels of CCTV images, Proceedings of Korea Information Science Society Winter Conference 2016, 1698-1700.
- Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. (2007) Greedy layer-wise training of deep networks, Advances in Neural Information Processing Systems, 19, 153-160.
- Bengio, Y. (2012) Deep learning of representations for unsupervised and transfer learning, Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 17-36.
- Berg, P., Norin, L., Olsson, J. (2016) Creation of a high resolution precipitation data set by merging gridded gauge data and radar observations for Sweden, Journal of Hydrology, 541, 6-13.
- Carreira, J., Zisserman, A. (2017) Quo vadis, action recognition? A new model and the kinetics dataset, Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 4724-4733.
- Caruana, R. (1995) Learning many related tasks at the same time with backpropagation, Advances in Neural Information Processing Systems, 7, 657-664.
- Chahine, M.T. (1992) The hydrological cycle and its influence on climate, Nature, 359(6394), 373-380.
- de Souza, C.R. (2018) Action recognition in videos: data-efficient approaches for supervised learning of human action classification models for video (Doctoral dissertation, Universitat Autònoma de Barcelona), Retrieved from https://ddd.uab.cat/pub/tesis/2018/hdl_10803_565827/crds1de1.pdf.
- Diba, A., Fayyaz, M., Sharma, V., Karami, A.H., Arzani, M.M., Yousefzadeh, R., Van Gool, L. (2017) Temporal 3D ConvNets: new architecture and transfer learning for video classification, arXiv preprint arXiv: 1711.08200.
- Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T. (2015) Long-term recurrent convolutional networks for visual recognition and description, Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2625-2634.
- Dong, R., Liao, J., Li, B., Zhou, H., Crookes, D. (2017) Measurements of rainfall rates from videos, Proceedings of the 2017 IEEE International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, 1-9.
- Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P., Bengio, S. (2010) Why does unsupervised pre-training help deep learning?, Journal of Machine Learning Research, 11(Feb), 625-660.
- Falkowski, P., Scholes, R.J., Boyle, E.E.A., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Höglberg, P., Linder, S., Mackenzie, F.T. (2000) The global carbon cycle: a test of our knowledge of earth as a system, Science, 290(5490), 291-296.
- Farneback, G. (2003) Two-frame motion estimation based on polynomial expansion, Proceedings of Scandinavian Conference on Image analysis, Springer, Berlin, Heidelberg, 363-370.
- Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S.P., Asner, G.P., Cleveland, C.C., Green, P.A., Holland, E.A., Karl, D.M. (2004) Nitrogen cycles: past, present, and future, Biogeochemistry, 70(2), 153-226.
- Garg, K., Nayar, S.K. (2004) Detection and removal of rain from videos, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 528-535.
- Gibson, J.J. (1950) The Perception of the Visual World, Houghton Mifflin, Boston, U.S.A., 259pp.
- Guo, L.C., Zhang, Y., Lin, H., Zeng, W., Liu, T., Xiao, J., Rutherford,

- S., You, J., Ma, W. (2016) The washout effects of rainfall on atmospheric particulate pollution in two Chinese cities, *Environmental Pollution*, 215, 195-202.
- Hochreiter, S., Schmidhuber, J. (1997) Long short-term memory, *Neural Computation*, 9(8), 1735-1780.
- Ioffe, S., Szegedy, C. (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift, *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, 37, 448-456.
- Kang, L.W., Lin, C.W., Fu, Y.H. (2012) Automatic single-image-based rain streaks removal via image decomposition, *IEEE Transactions on Image Processing*, 21(4), 1742-1755.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L. (2014) Large-scale video classification with convolutional neural networks, *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1725-1732.
- Kim, S., Hong, K.H., Jun, H., Park, Y.J., Park, M., Sunwoo, Y. (2014) Effect of precipitation on air pollutant concentration in Seoul, Korea, *Asian Journal of Atmospheric Environment*, 8(4), 202-211.
- Ko, B., Li, Z., Choi, H. (2017) Determination of precipitation from road CCTV video by using CNN-LSTM, *Proceedings of Korea Software Congress 2017*, 820-822.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998) Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86(11), 2278-2324.
- Lee, J., Hong, B., Jung, S., Chang, V. (2018) Clustering learning model of CCTV image pattern for producing road hazard meteorological information, 86, *Future Generation Computer Systems*, 1338-1350.
- Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C. (2014) When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, *Hydrology and Earth System Sciences*, 18(2), 575-594.
- Oki, T., Kanae, S. (2006) Global hydrological cycles and world water resources, *Science*, 313(5790), 1068-1072.
- Prentice, I.C., Farquhar, G.D., Fasham, M.J.R., Goulden, M.L., Heimann, M., Jaramillo, V.J., Kleshgi, H.S., LeQuéré, C., Scholes, R.J., Wallace, D.W.R. (2001) The carbon cycle and atmospheric carbon dioxide, *Climate Change 2001*, 183-237.
- Qiu, Z., Yao, T., Mei, T. (2017) Learning spatio-temporal representation with pseudo-3d residual networks, *Proceedings of the 2017 IEEE International Conference on Computer Vision*, 5534-5542.
- Ramanathan, V.C.P.J., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D. (2001) Aerosols, climate, and the hydrological cycle, *Science*, 294(5549), 2119-2124.
- Rolnick, D., Veit, A., Belongie, S., Shavit, N. (2017) Deep learning is robust to massive label noise, *arXiv preprint arXiv:1705.10694*.
- Shen, M., Xue, P. (2011) A fast algorithm for rain detection and removal from videos, *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo*, 1-6.
- Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C. (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting, *Advances in Neural Information Processing Systems*, 28, 802-810.
- Shin, Y., Hong, B., Lee, J. (2015) Weather condition distinguishing method based data analysis using CCTV video, *Proceedings of Korea Information Science Society Winter Conference 2015*, 247-249.
- Simonyan, K., Zisserman, A. (2014) Two-stream convolutional networks for action recognition in videos, *Advances in Neural Information Processing Systems*, 27, 568-576.
- Soomro, K., Zamir, A.R., Shah, M. (2012) UCF101: A dataset of 101 human actions classes from videos in the wild, *arXiv preprint arXiv:1212.0402*.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, 15(1), 1929-1958.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016) Rethinking the inception architecture for computer vision, *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2818-2826.
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M. (2015) Learning spatiotemporal features with 3d convolutional networks, *Proceedings of the 2015 IEEE Conference on Computer Vision*, 4489-4497.
- Walker, J.C.G. (1980) The oxygen cycle in the natural environment and the biogeochemical cycles, *The Handbook of Environmental Chemistry*, Springer, Berlin, Heidelberg, 87-104.
- Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., Van Gool, L. (2016) Temporal segment networks: Towards good practices for deep action recognition, *Proceedings of the 2016 European Conference on Computer Vision*, Springer, Cham, 20-36.

- Xue, X., Jin, X., Zhang, C., Goto, S. (2012) Motion robust rain detection and removal from videos, Proceedings of the 2012 IEEE International Workshop on Multimedia Signal Processing, 170-174.
- Zach, C., Pock, T., Bischof, H. (2007) A duality based approach for realtime TV-L1 optical flow, Proceedings of the 29th DAGM Conference on Pattern Recognition, Springer, Berlin, Heidelberg, 214-223.

Authors Information

Li Zhun (한국과학기술원 전산학부 석사과정)
현종환 (한국과학기술원 전산학부 박사과정)
최호진 (한국과학기술원 전산학부 교수)