

# 구름 운형 분류 시스템

김윤지, 성예은, 김진영, 염다빈, 최승호\*  
이화여자대학교,

\*jcn99250@naver.com

## Cloud Type Classification System

Kim Youn-ji, Seong Ye-eun, Kim Jin Young, Yeom Dabin, Choi Seung-Ho \*  
Ewha W. Univ.

### 요약

기존의 구름 모양을 알려주는 서비스는 시간 정보와 사용자의 위치 정보를 기반으로 실제 구름의 관측 정보를 제공한다. 이는 위치 정보와 시간 정보가 있어야지만 구름의 모양을 알 수 있다는 문제점이 있다. 본 논문에서는 위치 정보와 위치 정보를 사용하지 않고 오직 구름의 모습이 담긴 이미지를 통해 구름의 모양을 분류하고 사용자에게 제공하는 시스템을 제안한다. 3 가지 종류인 층운형, 적운형, 권운형 3 가지 범주의 데이터를 크롤링을 통해 수집한다. 그리고 CNN 알고리즘을 이용하여 3 가지 종류로 분류한다. 실험 결과, 제안한 방법이 구름 분류를 수행할 수 있음을 확인했다. 이를 통해 제안한 구름 분류 시스템이 구름 분류를 할 수 있음을 확인했다.

### I. 서론

구름은 평균적으로 전 세계 표면의 68%를 덮고 있을 만큼, 대기 중으로의 에너지 전달 및 기후에 영향을 주는 중요한 요소이다.[1] 또한, 그 유형에 따라 대기의 온도와 날씨에도 영향을 미치는데, 기상 연구자들은 구름의 종류인 운형을 정확하게 분류하는 것이 다양한 기상 상황을 예측하는 데에 필수적인 요소로 인식하고 있다.[2] 따라서 기존의 구름의 모양을 알려주는 서비스는 시간 정보와 사용자의 위치 정보를 기반으로 실제 구름의 관측 정보를 제공한다. 이는 위치 정보와 시간 정보가 있어야지만 구름의 모양을 알 수 있다는 문제점이 있다.

이를 경감하고자 위치 정보와 위치 정보를 사용하지 않고 오직 구름의 모습이 담긴 이미지를 통해 구름의 모양을 분류하고 사용자에게 제공하는 시스템을 제안한다.

### II. 관련 연구

지상 기반의 컨볼루션 신경망 구름 분류 모델인 CloudNET 은 5 개의 합성곱 층(Convolutional layer)과 2 개의 완전 연결 층(Fully-Connected, FC layer)으로 이루어진 분류기이다.[2] 이는 이전의 모델들과 달리, 더 간단하고 직관적인 모델로 차이점을 보인다.[3] CCSN(Cirrus Cumulus Stratus Nimbus) 데이터셋[4]을 이용하여 11 개의 운형으로 분류한 후, 학습을 하였다. 이 데이터셋은 특히 인간 활동에 의해 생성되는 비행운(Contrail)을 분류에 포함하였는데, 이는 기존 지상 기반 구름 데이터 베이스보다 포괄적인 점에서 차별적이다.

### III. 본론

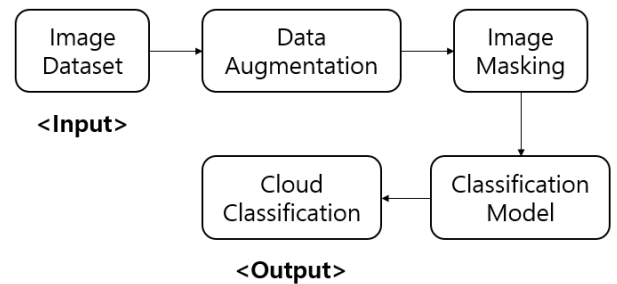


Figure 1. Our Proposal

본 논문에서 제안하는 방법을 그림 1 과 같다. 이미지 데이터셋을 증강하고, 구름만 남기고 나머지 피처를 없애는 마스크 이미지를 제작한 후, 해당 이미지를 이용하여 분류 모델기에 입력하여 운형 분류를 얻는 것이다.

데이터셋은 오픈 데이터셋과 직접 찍은 이미지로 구성된다. 지형 및 지리 조건에 따라 발생하는 구름의 형태가 달라지기 때문에, 두 이미지셋을 모두 선택했다. 오픈 데이터셋은 하버드 대학교에서 구축한 CCSN[4]와 Howard-Cloud-X 데이터셋 두 데이터 베이스를 함께 이용했다. CCSN 의 데이터 중 일부 이미지는 워터마크(watermark)가 포함되어 있어, 해당 이미지는 제외 한 후, 본 연구 목적에 맞게 층운형(Stratiform), 적운형(Cumuliform), 권운형(Cirriform) 구름으로, 총 3 가지 유형으로 재 분류한다. 결과적으로, 적운형 구름은 총 1108 장, 층운형 구름은 1055 장, 권운형 구름은 728 장의 이미지 데이터를 획득했다.

수집할 수 있는 이미지 데이터가 한정적인 상황에서 각 운형 별 이미지 데이터 수를 일정하게 구성하기 위해 GAN(Generative Adversarial Network) 모델 중, DCGAN(Deep Convolutional Generative Adversarial Network) 모델을 이용하여 이미지 데이터를 증강한다.

DCGAN 은 합성곱 신경망(Convolutional Neural Network, CNN)을 사용하여, 이미지 생성에 특화된 구조를 제공한다.[5] 이를 통해, 각 라벨(label)마다 약 1100 장의 이미지를 가지도록 데이터셋을 구성했다.



Figure 2. Data preprocessing, a) Origin image, b) feature elimination image

그림 2 에 보이듯이 이미지에서 구름의 특징을 효과적으로 추출하기 위해, 구름 외의 피쳐들을 OpenCV 라이브러리를 이용하여 제거하였다. 구름의 픽셀 값이 주변 물체들보다 확연히 높기 때문에, 적당한 임계 값을 설정한 후, 구름을 제외한 물체들은 검은색으로 제외시키는 과정을 거쳤다.

CloudNET 을 기반으로 모델을 구성했다. 5 개의 합성곱 층 뒤에 최대 풀링 층(Max-pooling layer)을 추가하여, 특징을 강조하고 연산량을 줄일 수 있도록 수정했다. 또한, 각 합성곱 층 뒤에 배치 정규화(Batch normalization)를 적용하여, 모델의 안정성을 향상시키고 학습 속도를 높였다. 그리고 마지막으로, 선행 연구와 동일하게, 2 개의 완전 연결 층을 사용했다.

본 모델에 적용한 데이터는 표 1 와 같이 학습, 검증, 테스트 데이터로 나누어 적용했다.

Table 1. Dataset Configuration

Type	Train	Validation	Test	Total
Cirriiform	930	89	108	1127
Cumuliform	887	96	125	1108
Stratiform	847	112	96	1055

테스트 데이터셋을 이용하여 분류를 수행한 결과, 총 정확도는 61.09%이며, 유형 별 정확도는 층 유형 60.41%, 적운형 49.6%, 권운형 75%로, 적운형 구름을 제외한다면 좋은 성능을 보였다. 층운형 구름의 경우, 적지 않은 수를 적운형 구름으로 분류하였으며, 그 반대로 적운형 구름을 층운형 구름으로 분류된 이미지가 많았다. 이는 층운형과 적운형 구름이 동시에 나타난 이미지가 많았기 때문으로 보이며, [5]에서도 나타난 오분류이다. 이는 적운형 구름이 시간이 지남에 따라 층운형으로 변형되기 때문에 생기는 문제이다.

Table 2. Model Performance

Type	Precision	Recall	F1 score	Accuracy
Cirriiform	0.68	0.75	0.71	0.75
Cumuliform	0.58	0.50	0.54	0.50
Stratiform	0.56	0.60	0.58	0.60

표 2 에서 정밀도와 재현율, F1 점수의 경우, 다음과 같은 결과가 나타났다. 적운형 구름의 재현율은 0.5 라는

낮은 값을 보였는데, 이는 앞서 언급한 고질적인 문제로 인해 발생한 값이다. 그림 3 에서 모델에 대한 성능 평가를 혼동행렬을 이용하여 시각화했다.

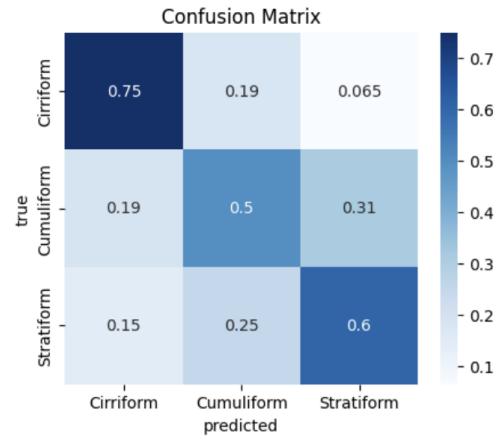


Figure 3. Model Performance

#### IV. 결론

기존의 구름의 모양을 알려주는 서비스는 시간 정보와 사용자의 위치 정보를 기반으로 실제 구름의 관측 정보를 제공한다. 위치 정보와 시간 정보가 있어야지만 구름의 모양을 알 수 있다는 문제점이 있다. 본 논문에서는 위치 정보와 위치 정보를 사용하지 않고 오직 구름의 모습이 담긴 이미지를 통해 구름의 모양을 분류하고 사용자에게 제공하는 시스템을 제안했다.

#### 참고 문헌

- [1] Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, W. P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., & Zhao, G. (2013). Assessment of Global Cloud Datasets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel. Bulletin of the American Meteorological Society, 94(7), 1031-1049. <https://doi.org/10.1175/BAMS-D-12-00117.1>
- [2] Zhang, J. L., Liu, P., Zhang, F., & Song, Q. Q. (2018). CloudNet: Ground-based cloud classification with deep convolutional neural network. Geophysical Research Letters, 45, 8665- 8672. <https://doi.org/10.1029/2018GL077787>
- [3] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
- [4] Liu, Pu, 2019, "Cirrus Cumulus Stratus Nimbus (CCSN) Database", <https://doi.org/10.7910/DVN/CADDPD>, Harvard Dataverse, V2
- [5] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.