JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 8, pp. 51-61, August 2026
https://doi.org/10.9708/jksci.2026.31.08.051

Rethinking Data Augmentation for CCTV-Based Rainfall Intensity Classification

Ju-Hwan Han*, Gye-Young Kim**

*Ph.D. Student, Dept. of AI and Software Convergence, Soongsil University, Seoul, Korea
**Professor, Dept. of AI and Software Convergence, Soongsil University, Seoul, Korea

[Abstract]

As the frequency and intensity of localized heavy rainfall increase due to climate change, research aimed at utilizing the dense network of CCTV cameras installed throughout cities as supplementary rainfall observation data is gaining attention. However, a common constraint in this field is the lack of labeled video data, and the effectiveness of standard techniques—such as transfer learning, data augmentation, and multi-class classification—under small-scale, imbalanced conditions has not been sufficiently validated. This study conducts four series of controlled experiments using a small-scale CCTV video dataset (6,805 clips, 3-class) collected from a single site, and cross-validates the results through augmentation decomposition, baseline, and multi-seed experiments. Under the single-site CCTV rainy-weather video conditions examined in this study, removing the standard augmentation combination (horizontal flip + color jitter) consistently improved Macro-F1 across five random seeds, with a mean improvement of +0.215 (0.428→0.643; paired t-test, $t(4) = 8.90$, $p = 0.0009$, two-sided). These results suggest that augmentation strategies validated in general action recognition need to be re-examined before being applied to the rainfall video domain.

▸ **Key words**: localized heavy rainfall, CCTV, rainfall intensity classification, transfer learning, data augmentation, ablation study, video recognition

[요 약]

기후 변화로 국지성 집중호우의 빈도와 강도가 증가하면서, 도시 전역에 조밀하게 설치된 CCTV를 보조 강우 관측 자료로 활용하려는 연구가 주목받고 있다. 그러나 이 분야의 공통 제약은 라벨된 영상의 부족이며, 소규모 불균형 조건에서 전이 학습·데이터 증강·다중 클래스 분류 같은 표준 기법이 어떻게 작동하는지는 충분히 검증되지 않았다. 본 논문은 단일 지점에서 수집한 소규모 CCTV 영상 데이터셋(6,805클립, 3-클래스)을 대상으로 네 개의 통제 실험 시리즈를 수행하고, 그 결과를 증강 분해, 베이스라인, 그리고 다중 시드 실험으로 교차 검증한다. 본 논문이 다룬 단일 지점 CCTV 강우 영상 조건에서는, 행동 인식에서 표준으로 쓰이는 데이터 증강 조합 즉, 수평 뒤집기와 색상 지터를 제거했을 때 5개 시드 모두에서 Macro-F1이 일관되게 향상되었으며, 평균 향상 폭은 +0.215였다(0.428→0.643; 대응표본 t-검정, $t(4) = 8.90$, $p = 0.0009$, 양측). 이 결과는 일반 행동 인식에서 검증된 증강 전략을 강우 영상 도메인에 적용하기 전에 재검토가 필요함을 시사한다.

▸ **주제어**: 국지성 호우, CCTV, 강우강도 분류, 전이학습, 데이터 증강, 절제 연구, 비디오 인식

• First Author: Ju-Hwan Han, Corresponding Author: Gye-Young Kim
*Ju-Hwan Han (jhhan@bdvr.co.kr), Dept. of AI and Software Convergence, Soongsil University
**Gye-Young Kim (gykim11@ssu.ac.kr), Dept. of AI and Software Convergence, Soongsil University
• Received: 2026. 07. 07, Revised: 2026. 07. 17, Accepted: 2026. 08. 11.

Copyright © 2026 The Korea Society of Computer and Information
http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945
www.kci.go.kr

I. Introduction

기후 변화의 영향으로 짧은 시간에 좁은 지역에 집중되는 국지성 호우가 빈발하고 있다. 도심 저지대 침수, 하천 범람, 산사태 등 재난 피해를 줄이기 위해서는 짧은 선행 시간 안에 정확한 강우 정보를 확보하는 것이 핵심이다. 그러나 지상 우량계는 점 단위 관측이어서 공간 대표성이 낮고, 기상 레이더는 도심의 미세 규모 현상을 포착하는 데 한계가 있다.

도시 전역에 조밀하게 설치된 CCTV는 이러한 공백을 보완할 수 있는 유망한 대안이다. CCTV 영상에는 빗줄기의 굵기와 밀도, 노면 젖음, 시정 저하 등 강우강도와 관련된 시각적 단서가 담겨 있어[1], 별도 장비 없이 고밀도 관측망을 구성할 수 있는 잠재력이 있다. 그러나 영상 기반 강우 관측 논문의 공통적인 제약은 라벨링 된 영상 데이터의 부족이며, 이 조건에서 전이학습·데이터 증강·다중 클래스 분류 같은 표준 기법이 어떻게 상호작용하는지는 충분히 검증되지 않았다.

본 논문의 중심 기여는 새로운 분류 모델의 제안이 아니라, 소규모·불균형 CCTV 강우 영상이라는 제약 조건에서 표준 데이터 증강의 효과를 통제 실험으로 검증하는데 있다. 이를 위해 단일 지점에서 수집한 소규모 CCTV 영상 데이터셋을 대상으로 네 개의 통제 실험 시리즈를 수행하고, 그 결과를 증강 분해·베이스라인·다중 시드 실험으로 교차 검증한다. 특히 행동 인식에서 표준으로 쓰이는 데이터 증강(수평 뒤집기+색상 지터)이 본 데이터셋 및 실험 조건에서는 오히려 성능을 저하시킬 수 있음을 보인다. 이 논문의 기여는 다음 네 가지이다.

(1) 강우강도 분류에서 표준 데이터 증강의 효과를 통제 실험으로 분리하고, 증강 제거가 5개 시드 모두에서 Macro-F1을 일관되게 향상시키며 평균 향상 폭이 +0.215임을 대응표본 t-검정과 함께 제시한다.

(2) 증강을 플립과 색상 지터로 분해하여, 어느 한 증강만 제거해서는 성능이 회복되지 않고 두 증강을 모두 제거할 때에만 향상이 나타나는 비가산적 양상을 단일 시드 탐색 실험으로 관찰하고, 그 해석과 한계를 명시한다.

(3) 사전학습 기여, 클래스 통합(3-클래스 대 5-클래스), 입력 해상도·시간 윈도우의 영향을 동일 파이프라인에서 일관되게 비교한다.

(4) 2D/3D ResNet[2] 베이스라인과 다중 시드 재현성 분석을 통해 결과의 강건성과 한계를 함께 제시한다.

II. Preliminaries

1. Related works

1.1 Video-Based Rainfall Observation

영상으로부터 강우를 정량화하려는 시도는 빗줄기의 광학적 모델링에서 출발한다. Garg와 Nayar[1]는 빗방울이 카메라에 맺히는 동적 외형과 광선 굴절을 물리적으로 모델링하여, 빗줄기가 만드는 밝기 변화가 강우량과 체계적으로 연관됨을 보였다. 이 이론적 토대 위에서 Allamano 등[3]은 교통 카메라 영상의 빗방울 크기 분포로부터 강우강도를 추정하였고, Zen 등[4]은 교통 카메라를 활용한 강우량 추정 프레임워크를 제안하였다. Zheng 등[5]은 영상 감시 카메라에서 실시간 강우강도 추정을 수행하였으며, Haurum 등[6]은 일반 목적 감시 카메라로 강우 탐지(이진 분류)를 수행하였다.

최근 논문은 딥러닝으로 추정 정확도와 신뢰성을 높이는 방향으로 발전하였다. Yin 등[7]은 ResNet 계열 CNN으로 CCTV 단일 프레임에서 강우강도를 회귀 추정하였고, Zheng 등[8]은 CNN-LSTM에 베이지안 불확실성 추정을 결합하여 추정값의 신뢰구간을 함께 제공하였다. 영상 외 보조 신호를 활용하는 흐름도 있는데, Wang 등[9]은 감시 영상 음향으로 강우를 관측하였고, Wang 등[10]은 근적외선 영상 기반 우량계를 제안하였다. Rajabi 등[11]은 지역 네트워크에서 딥러닝 기반 실시간 추정을 수행하였다. 보다 최근에는 Fiallos-Salguero 등[20]이 감시 카메라 데이터와 하이브리드 딥러닝 프레임워크를 결합하여 확장성 있는 강우 추정을 시도하였고, Wang 등[21]은 야간 조명 조건의 도시 감시카메라 영상에서 강우강도를 추정하여 CCTV 기반 환경 인식의 적용 범위를 넓혔다. 이들 선행 논문의 공통 제약은 소규모·불균형 데이터에서 표준 증강 기법의 적절성을 체계적으로 검증하지 않았다는 점이다.

1.2 Video Recognition Models

본 논문이 비교하는 세 모델은 비디오 행동 인식에서 서로 다른 설계 철학을 대표한다. X3D[12]는 2D 영상 분류기를 채널·시간·공간·깊이의 여러 축을 따라 점진적으로 확장하여, 적은 연산량과 파라미터로 높은 정확도를 달성하는 경량 3D CNN 계열이다. SlowFast[13]는 낮은 프레임률로 외형을 포착하는 느린 경로와 높은 프레임률로 빠른 움직임을 포착하는 빠른 경로를 병렬로 운용한다. MViTv2[14]는 계층적 멀티스케일 어텐션으로 지역적 세부와 전역적 맥락을 동시에 모델링하는 비전 트랜스포머 계열이다.

www.kci.go.kr

1.3 Transfer Learning and Domain-Specific Augmentation

소규모 데이터에서 대규모 사전학습 가중치를 재사용하는 전이학습은 표준 해법이며, 전이되는 특징의 일반성과 전이 가능성에 관한 분석도 폭넓게 이루어졌다[15]. 그러나 데이터 증강의 효과는 도메인에 강하게 의존한다는 점이 점차 분명해지고 있다. Shorten과 Khoshgoftaar[16]의 조망은 증강 기법의 적합성이 과제·데이터 특성에 따라 달라짐을 강조하며, Raghu 등[17]은 의료 영상에서 일반 특징 전이의 이점이 제한적임을 보고하였다. 강우 도메인에서는 빗줄기·노면 젖음·시정 저하 등이 강우강도를 가르는 핵심 광측정 단서이므로, 색상 불변성을 학습시키는 색상 지터가 오히려 판별 정보를 훼손할 위험이 있다.

1.4 Robust Learning on Small-Scale, Imbalanced Data

단일 지점 CCTV 데이터는 표본이 적고 강우 등급이 심하게 불균형하다는 두 제약을 동시에 갖는다. 불균형 완화에는 손실 가중치, 라벨 스무딩, 등급 통합 등이 흔히 쓰이나 그 효과는 상황 의존적이다. 또한 소규모 데이터에서는 과적합과 평가의 분산이 커지므로, 단일 실행 결과만으로 결론을 내리기 어렵다. 최근에는 강도 레이블을 갖는 날씨 영상 데이터셋 VARG[22]가 공개되는 등 기상 영상 분류 연구가 활발해지고 있으나, 소규모·불균형 조건에서의 평가 프로토콜은 여전히 표준화되어 있지 않다. 본 논문은 다중 시드 반복과 통계적 유의 검정, 시간적 분할 실험을 통해 이 한계를 보완한다.

III. The Proposed Scheme

1. Dataset

1.1 Collection and Composition

데이터셋은 단일 관측 지점에서 수집하였다. 동일 지점에 설치된 전도형 우량계 1대가 1분 단위 강우강도를 측정하고, 서로 다른 방향·화각을 가진 카메라 3대(다채널)가 같은 지점을 동시에 촬영하여, 우량계 측정값과 각 카메라 영상을 시각적으로 정합하였다. 전체 6,805개 클립은 세 채널과 주간·야간 데이터를 모두 포함하며, 각 클립은 분 단위 시간 윈도우에 해당한다.

수집 기간은 2025년 7월 1일부터 7월 31일까지의 여름 장마철 1개월이며, 이 기간 중 12일에 걸쳐 총 2,285분의 관측 구간이 수집되었다. 총 강우 이벤트 수는 25개(강한 비 포함 이벤트 11개, 약한 비 전용 이벤트 14개)이고, 이벤트 지속 시간은 7분에서 273분(중앙값 34분)이며, 최대 1분 강우강도는 90 mm/hr에 달한다. 주간/야간 클립 비율은 57.3%/42.7%이고, 카메라별 클립 수는 CH1 2,265개, CH3(기준 채널) 2,285개, CH4 2,255개로 총 6,805개이다. 세 카메라는 동일 강우 시점을 서로 다른 방향·화각에서 촬영하므로 카메라 간 시점 중복이 존재하며, 이는 1.4절의 이벤트 단위 분할에서 구조적으로 통제된다. 라벨은 사람이 영상을 판독하여 부여한 것이 아니라, 동일 지점 우량계의 분 단위 측정값으로부터 자동 할당하였으며, 주간·야간 구분은 일출·일몰 기반으로 부여하였다.

1.2 Temporal Alignment between Rain Gauge and Clips

카메라 영상은 1분 단위 파일로 저장되며, 파일명에 기록된 분 단위 타임스탬프를 우량계 기록의 분 단위 절사 시각과 일치시키는 방식으로 정합하였다. 각 분에 대해 해당 분에 기록된 우량계 측정값의 평균을 그 분의 강우강도 r로 정의하고, 이를 동일 분의 모든 채널 클립에 라벨로 부여하였다. 학습 입력의 30초/60초 시간 윈도우는 해당 1분 클립에서 추출한다. 클래스 경계(r=0 및 r=7.5 mm/hr) 부근의 분에 대한 별도 제외 규칙은 두지 않았으며 해당 분의 평균값을 기준으로 그대로 분류하였다. 클립 시각은 카메라 타임스탬프를 기준으로 하며, 카메라-우량계 간 잠재적 시간 정합 오차는 IV장 11.3절의 한계에서 논의한다.

1.3 Class Consolidation

Table 1. Class distribution after class consolidation

Class	WMO class	Rain Intensity (mm/hr)	Clip count	Ratio (%)
no_rain	no_rain	r=0	2,130	31.3
weak	light + moderate	0<r≤7.5	3,469	51.0
strong	heavy + violent	r>7.5	1,206	17.7
total	-	-	6,805	100

세계기상기구(WMO) 5등급(no_rain, light, moderate, heavy, violent) 강우 분류[18]에서 heavy와 violent 등급의 표본이 극히 희소하여 심한 클래스 불균형이 발생하였다. 이를 완화하고 재난 대응이라는 응용 목적에 맞추기 위해, 임계값 7.5mm/hr를 기준으로 Table 1과 같이 무강우(no_rain)·약한 비(weak)·강한 비(strong) 3-클래스로 통합하였다. 임계값 7.5 mm/hr는 본 논문이 사용한 5

www.kci.go.kr

등급 경계 체계에서 moderate와 heavy를 구분하는 경계로, 3-클래스 통합은 이 경계를 그대로 유지하여 light·moderate를 weak로, heavy·violent를 strong으로 묶은 것이다. 이 경계값은 방재 운영 기준과 WMO 등급 경계와의 대응 관계를 고려하여 설정하였으며, 두 체계의 대응 관계는 Table 1에 명시하였다.

1.4 Event-Based Temporal Splitting

연속된 클립은 같은 강우 이벤트에 속할 가능성이 높아 분할 간 정보 누수(data leakage)가 발생할 수 있다. 30분 이상의 건조 구간으로 분리된 연속 강우 구간을 하나의 이벤트로 정의하고, 동일 이벤트의 클립을 반드시 같은 분할에 배정하는 이벤트 단위 분할을 적용하였다. 분할은 분(minute) 단위로 결정한 뒤 동일 분에 촬영된 모든 채널(CH1·CH3·CH4)의 클립에 동일하게 전파되므로, 동일 시점의 다채널 클립이 분할을 가로질러 배정되는 정보 누수는 구조적으로 발생하지 않는다. 강한 비 이벤트는 세 분할 모두에 최소 1개가 보장되도록 강한 비 클립 수 기준의 탐욕(greedy) 방식으로 배분하였고, 약한 비 이벤트는 시간 순서를 유지하며 비율에 따라 배분하였으며, 무강우 분은 이벤트와 독립적으로 시간 순서에 따라 비율 배분하였다. 총 이벤트 수와 분할별 이벤트 수는 Table 2에 클립 수와 함께 제시한다. 최종 분할은 Table 2와 같다.

Table 2. Class distribution by train/validation/test split (number of clips)

Class	Train	Val	Test
no_rain	1,404	420	306
weak	2,563	369	537
strong	648	312	246
total (clips)	4,615	1,101	1,089
rain events	18	4	3

2. Methodology

2.1 Model Architectures

X3D-L(약 5.3M 파라미터), SlowFast R50(약 33.7M), MViTv2-S(약 26M) 세 모델을 비교하였다. 별도의 명시가 없는 한 모든 모델은 Kinetics-400[19] 사전학습 가중치로 초기화하였고, 출력층은 3-클래스에 맞게 교체하였다.

2.2 Training Setup

Table 3. Common training settings

Item	Setting
Optimizer	AdamW (weight decay 1×10^{-5})
Learning Rate Schedule	Cosine decay after 3-epoch linear warmup
Gradient Clipping	Global norm max_norm = 1.0
Precision / Distributed	Mixed precision (fp16), 4-GPU DDP, Effective batch size 16
Early Stopping	Based on validation Macro-F1 with patience 15 (Max 50 epochs)
Class Imbalance Mitigation	Inverse frequency-based loss weighting
Default Input	224^2 resolution, 30-second time window
Default Data Augmentation	Horizontal flip(p = 0.5) + Color jitter(brightness, contrast, saturation, hue)
Pre-training	Initialized with Kinetics-400 weights (except A1)
Frame sampling	Source clip 60 s @ 30 fps (≈1,800 frames); 32 frames pre-extracted uniformly (short side 512); 30 s window uses the first 16 of 32; uniform temporal subsampling to model input – MViTv2-S 16, SlowFast R50 32 (slow 8 / fast 32), X3D-L 16
GPU / training time	NVIDIA RTX 3090 (23.7 GiB) × 4; A0 (224^2, 30s): 18 epochs, ≈ 85 h wall-clock
Balanced sampling in data loader	Not used (imbalance mitigated only by inverse-frequency loss weighting)
Random seed control	torch.manual_seed(seed+rank) and np.random.seed(seed+rank) per DDP rank; cuDNN determinism not enforced

공통 학습 설정은 Table 3과 같다. 별도 명시가 없는 한 모든 실험에 동일하게 적용하였으며, 시리즈별로 달라지는 항목(모델·학습률·해상도·증강 등)은 해당 절에 기술한다. 재현성을 위해 프레임 샘플링, 하드웨어, 시드 고정 방식을 Table 3에 추가로 명시하였다. cuDNN 결정론은 강제하지 않았으므로 동일 seed에서도 실행 간 미세한 변동이 존재할 수 있으며, 이러한 변동은 다중 시드 반복(IV장 9절)으로 통제하였다.

www.kci.go.kr

2.3 Experimental Series Design

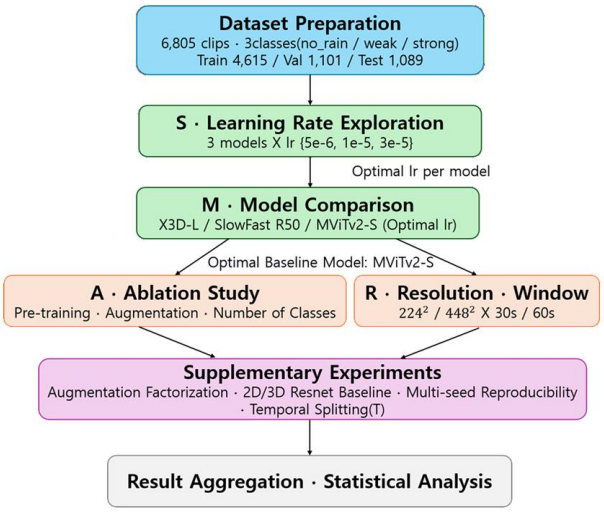

Fig. 1. Experimental pipeline

Table 4. Seed configuration of each experimental series

Series	Runs	Seed(s)
S (learning-rate search)	single run	42
M (model comparison)	single run	42
A (ablation)	single run	42
R (resolution·window)	single run	42
Augmentation decomposition	single run	42
B (baselines)	single run	42
T (temporal split)	single run	42
Multi-seed (A0, A2, M)	5 runs each	{42, 123, 456, 789, 1234}

실험은 네 시리즈로 구성된다. (S) 학습률 탐색: 세 모델 각각에 학습률 ∈ {5e-6, 1e-5, 3e-5}를 적용하여 모델별 최적 학습률을 결정한다. (M) 모델 비교: S에서 얻은 최적 학습률을 적용하여 세 모델을 동일 조건에서 비교한다. (A) 절제 연구: 모델 비교에서 가장 안정적이었던 MViTv 2-S를 기준으로, 사전학습·증강·클래스 수·라벨 스무딩·주야간 분할 항목을 단계적으로 변경한다. (R) 해상도·시간 윈도우: 입력 해상도(224^2, 448^2)와 시간 윈도우(30초, 60초)를 동일 백본(MViTv2-S)으로 비교한다. 이에 더해 증강 분해, 베이스라인(B), 다중 시드 재현성, 시간적 분할(T) 실험을 보조적으로 수행하였다. 단일 실행 실험과 다중 시드 실험을 명확히 구분하기 위해 각 실험의 시드 구성을 Table 4에 정리하였다. 단일 실행 결과는 탐색적 성격을 가지며, 핵심 주장(증강 제거 효과)은 다중 시드 결과(IV장 9절)에 근거한다. 전체 실험 파이프라인은 Fig. 1과 같다.

2.4 Evaluation Metrics

주 지표는 클래스 불균형에 강건한 Macro-F1이며, 클래스별 F1과 정확도(Accuracy)를 함께 제시한다. C개 클래스에 대해 클래스별 F1과 Macro-F1은 정밀도(Pc)와 재현율(Rc)의 조화평균으로 정의된다.

기상 운영 관점의 평가를 위해, 예측을 무강우 대 강우(weak ∪ strong)의 이진 사건으로 환산하여 탐지확률(POD), 오경보율(FAR), 임계성공지수(CSI)를 산출하였다. 이진 혼동행렬에서 TP(강우를 강우로 예측), FN(강우를 무강우로 예측), FP(무강우를 강우로 예측)를 정의하면 각 지표는 다음과 같다.

$$POD = TP / (TP + FN)$$

$$FAR = FP / (TP + FP)$$

$$CSI = TP / (TP + FP + FN)$$

여기서 positive event는 실제 또는 예측 강우강도가 weak 또는 strong인 경우(즉 $r > 0$으로 예측·관측된 사건)이며, FAR은 예측된 강우 사건 중 오경보의 비율(false alarm ratio)이다. POD는 높을수록, FAR은 낮을수록 우수하며, CSI는 POD와 FAR을 동시에 반영하는 종합지표이다.

IV. Experimental Results and Analysis

1. Learning-Rate Search (Series S)

Table 5는 S 시리즈의 결과이다. MViTv2-S는 5e-6에서 Macro-F1 0.483을 기록하며 세 학습률 모두에서 가장 안정적으로 높은 성능을 보였다. SlowFast R50은 5e-6(0.384)만 경쟁력이 있었고 더 큰 학습률에서는 성능이 정체하였으며, 실제로 1e-5와 3e-5 조건은 모두 다수 클래스 예측으로 붕괴해 동일한 테스트 지표로 수렴하였다. X3D-L은 검증 손실 기반 조기 종료로 5~10 에폭 내에 학습이 일찍 종료되었다.

Table 5. S Series: Learning-rate search results

Model	X3D-L		
LR	**5e-6**	1e-5	3e-5
Test F1	**0.375**	0.297	0.352
Acc.	**0.400**	0.318	0.397
POD	**0.688**	0.593	0.658
FAR	**0.268**	0.313	0.279
CSI	**0.550**	0.467	0.524
F1(strong)	**0.317**	0.182	0.241
Model	SlowFast R50		
LR	**5e-6**	1e-5	3e-5
Test F1	**0.384**	0.338	0.338
Acc.	**0.442**	0.477	0.477
POD	**0.983**	0.936	0.936
FAR	**0.256**	0.263	0.263
CSI	**0.735**	0.701	0.701
F1(strong)	**0.384**	0.174	0.174
Model	MViTv2-S		
LR	**5e-6**	1e-5	3e-5

www.kci.go.kr

Test F1	**0.483**	0.477	0.460
Acc.	**0.557**	0.545	0.506
POD	**0.835**	0.826	0.791
FAR	**0.142**	0.143	0.158
CSI	**0.734**	0.726	0.689
F1(strong)	**0.161**	0.166	0.190

2. Model Comparison (Series M)

Table 6은 최적 학습률을 적용한 모델 비교 결과이다. 지표별로 최적 모델이 다르다. 증강을 적용한 이 조건에서 MViTv2-S는 Macro-F1(0.487), Accuracy(0.560), FAR(0.142)에서 가장 우수했으나, 강한 비 클래스 F1은 SlowFast R50(0.383)이 MViTv2-S(0.171)보다 높았고 POD도 SlowFast R50(0.983)이 가장 높았다. 따라서 '가장 우수한 모델'은 운영 목적(균형 식별 대 강우 누락 최소화)에 따라 달라지며, 본 논문은 Macro-F1 기준의 안정성을 근거로 MViTv2-S를 후속 실험의 기준 백본으로 선택하였다. 세 모델 모두 학습 F1이 0.90 이상으로 상승하는 동안 검증 F1은 낮은 수준에 머무는 과적합이 관찰되며, 특히 SlowFast R50의 검증 손실 발산이 두드러졌다. SlowFast R50은 POD가 0.983으로 가장 높지만 FAR도 0.256으로 높아 실제 운영에서 과도한 오경보를 유발할 수 있다.

Table 6. M Series: Model comparison under optimal learning rates

Model	MViTv2-S	SlowFast R50	X3D-L
Test F1	0.487	0.384	0.375
Acc.	0.560	0.442	0.400
POD	0.835	0.983	0.688
FAR	0.142	0.256	0.268
CSI	0.734	0.735	0.550
F1(weak)	0.664	0.540	0.478
F1(strong)	0.171	0.383	0.317

3. Ablation Study (Series A)

Table 7과 Table 8은 MViTv2-S를 기준으로 한 절제 연구 조건과 결과이다. 데이터 증강을 모두 제거한 A2 조건이 단일 실행 기준 Macro-F1 0.713, 정확도 0.731, FAR 0.045, CSI 0.852로 A 시리즈 전체에서 가장 우수하였으며, 증강을 적용한 베이스라인 A0(0.494) 대비 +0.219 향상되었다. 다만 이는 단일 실행(seed=42) 결과로 시드 분산의 영향을 받으며, 5개 시드 평균 기준의 향상폭은 +0.215(0.428±0.058 → 0.643±0.051, IV장 9절)이다. 두 조건의 모델 구조와 학습 설정이 동일(MViTv2-S)하므로 이 향상의 주된 요인은 증강 제거로 해석되나, 학습의 확률적 요인이 완전히 배제되는 것은 아니다. 사전학습 제거(A1)는 Macro-F1을 0.370으로 급락시켜 전이학습의 중요성을 확인하였다. 5-클래스 미통합(A3)은 strong 클래스의 희소성 문제가 심화되어 F1(strong)이 0으로 붕괴하였다.

Table 7. A Series: Ablation study contents

ID	Description
A0	Baseline(Pre-training + Augmentation + Class weighting + 3cls)
A1	Remove pre-training (Random initialization)
A2	**Remove all augmentations (flip + color jitter)**
A3	5-class (WMO unmerged)
A4	Label smoothing + Enhanced class weighting

Table 8. A Series: Ablation study results (MViTv2-S, 224², 30s; single run)

ID	Test F1	Acc.	FAR	CSI	F1 (strong)	F1 (no_rain)
A0	0.494	0.569	0.130	0.732	0.177	0.640
A1	0.370	0.462	0.270	0.705	0.361	0.161
A2	**0.713**	**0.731**	**0.045**	**0.852**	**0.571**	**0.819**
A3	0.323	0.367	0.272	0.620	0.000	0.266
A4	0.457	0.512	0.144	0.713	0.159	0.609

4. Per-Class F1 Analysis

Fig. 2는 전체 실험의 Macro-F1 비교이다. 강한 비 클래스는 대부분의 실험에서 0.15~0.32 수준으로 취약하지만, A2에서만 0.57로 두드러지게 향상된다. 이는 본 실험 조건에서 증강 제거가 단지 평균 지표뿐 아니라 가장 어려운 소수 클래스의 식별력도 함께 개선함을 시사한다.

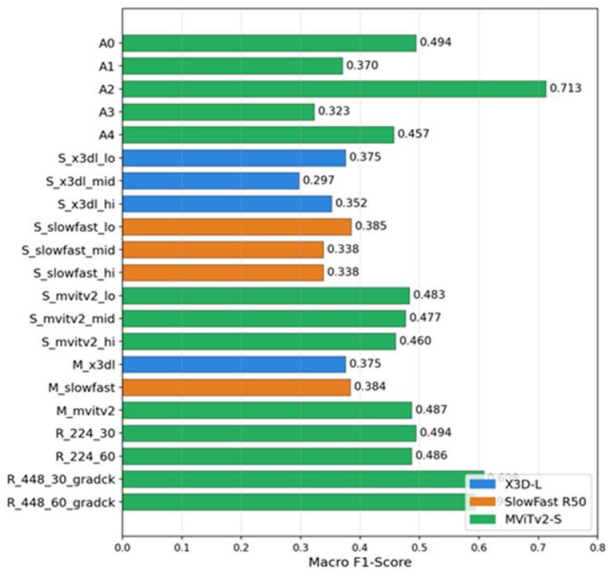

Fig. 2. Macro-F1 comparison across all experiments

www.kci.go.kr

5. Meteorological Operational Metrics Analysis

기상 운영 관점에서 모델·조건별 지표를 비교하면, A2 조건은 단일 실행 기준 FAR이 0.045로 베이스라인(0.130) 대비 약 65% 낮았으나, 다중 시드 기준으로는 FAR 차이가 유의하지 않았다(IV장 9절). 반면 SlowFast R50은 POD가 0.983으로 매우 높지만, FAR도 0.256으로 높아 실제 운영에서는 과도한 오경보를 유발할 수 있다. 강우 누락(FN)이 치명적인 응용에서는 높은 POD가 우선이나, 균형 잡힌 식별이 중요한 응용에서는 A2 구성이 적합하다.

6. Resolution·Window Analysis (Series R)

R 시리즈는 입력 해상도와 시간 윈도우의 영향을 동일 백본(MViTv2-S)으로 통제하여 비교한다. Table 9에서 30초(0.494)와 60초(0.486) 윈도우가 사실상 동일하여, 60초 확대는 에폭당 학습 시간만 약1.9배 늘릴 뿐 이득이 없었다. 표준 448^2 학습은 RTX 3090(23.7 GiB) 환경에서 메모리 부족(OOM)으로 실패하였으나, 그래디언트 체크포인팅 적용 시 448^2에서 30초(0.608), 60초(0.594)로 224^2 대비 +0.114의 향상을 달성하였다.

해상도 상승의 효과는 CSI에서도 나타난다. 224^2 30초 조건의 CSI 0.732에서 448^2 30초 조건의 0.783으로 향상되었으며, 이는 강한 비의 미세한 빗줄기·노면 반사 단서가 고해상도 입력에서 더 잘 보존됨을 시사한다. 다만 FAR는 224^2 30초(0.130)에서 448^2 30초(0.166)로 소폭 증가하여, 고해상도가 탐지 감도를 높이는 동시에 오경보도 일부 늘릴 수 있음을 나타낸다.

Table 9. R Series – Input resolution and temporal window analysis

ID	Resolution	Window	Test F1	F1(strong)	FAR	CSI
R_224_30	224^2	30s	0.494	0.177	0.130	0.732
R_224_60	224^2	60s	0.486	0.171	0.138	0.714
R_448_30	448^2	30s	0.608	0.478	0.166	0.783
R_448_60	448^2	60s	0.594	0.470	0.174	0.778

7. Augmentation Decomposition Analysis

Table 10. Augmentation decomposition results (MViTv2-S, seed = 42)

ID	Flip	Color	Macro-F1	F1(strong)	FAR	CSI
A0	O	O	0.494	0.177	0.130	0.732
A2_1	O	X	0.494	0.177	0.130	0.732
A2_2	X	O	0.484	0.149	0.115	0.753
A2_3	X	weak	0.507	0.150	0.072	0.770
A2_4	**X**	**X**	**0.713**	**0.571**	**0.045**	**0.852**

증강 제거(A2)의 이득이 수평 뒤집기와 색상 지터 중 무엇에서 비롯되는지를 규명하기 위해, 두 요소를 독립적으로 제어한 변형을 비교하였다(Table 10). 이 분해 실험은 seed=42 단일 실행에 기반한 탐색적 분석임을 먼저 밝힌다. 결과에서 주목할 점은 어느 한 증강만 제거해서는 성능이 회복되지 않았다는 것이다. 색상 지터만 제거하고 플립을 유지한 A2_1은 A0와 동일한 결과(0.494)를 보였고, 플립만 제거하고 색상 지터를 유지한 A2_2도 0.484로 개선이 없었다. 두 증강을 모두 제거한 A2_4에서만 0.713으로 향상이 나타났다. 즉 성능 저하는 특정한 증강에 단독으로 귀속되지 않으며, 두 증강이 결합된 저하 효과가 비가산적으로 나타나는 양상이다. 다만 단일 실행 결과이므로 이 상호작용 양상은 가설 수준의 관찰이며, 확정을 위해서는 각 분해 조건의 다중 시드 반복이 필요하다(11.3절).

색상 지터가 강우 판별 단서를 교란할 수 있다는 해석은 다음과 같이 이해할 수 있다. 밝기·대비·채도·색조의 무작위 변동은 빗줄기의 밝은 사선, 노면 젖음에 따른 색조, 시정 저하로 인한 색온도 등 강우강도를 가르는 저수준 광측정 단서를 직접 변형한다. 일반 행동 인식에서 유용한 색상 불변성 학습이, 색 자체가 핵심 신호인 강우 도메인에서는 오히려 판별 정보를 훼손할 가능성이 있다. 다만 Table 10에서 색상 지터 제거만으로는 개선이 나타나지 않았으므로, 이 메커니즘 설명은 두 증강의 결합 효과에 대한 부분적 해석으로 한정된다.

8. Baseline Comparison (Series B)

제안 구성의 상대적 위치를 가늠하기 위해 단순 베이스라인 세 종을 비교하였다(Table 11). B_resnet50_single은 단일 프레임 ResNet-50, B_resnet50_avg는 클립 내 프레임들을 ResNet-50으로 처리한 뒤 평균 풀링, B_resnet18_3d는 소규모 3D ResNet-18이다. 결과는 각각 Macro-F1 0.556, 0.550, 0.669로, 2D 모델들에 비해 3D ResNet-18의 성능이 높았다. 제안 구성(MViTv2-S, A2, 무증강)은 단일 실행 기준 0.713으로 가장 높았다. 비교의 공정성과 관련하여, 세 베이스라인은 제안 구성과 동일한 전처리, 증강 조건, class weighting, early stopping, seed 조건에서 학습되었다.

www.kci.go.kr

Table 11. Baseline comparison (B Series)

ID	Macro-F1	Accuracy	Note
B_resnet50_single	0.556	0.570	Single-frame 2D
B_resnet50_avg	0.550	0.543	Frame average pooling
B_resnet18_3d	**0.669**	**0.672**	**Small-scale 3D CNN**
MViTv2-S (A2, no aug)	0.713	0.731	Proposed (Single run)

9. Multi-Seed Reproducibility and Statistical Significance

단일 실행값의 시드 분산 영향을 통제하기 위해, 핵심 세 구성(A0_baseline, A2_no_aug, M_mvitv2)을 seed ∈ {42, 123, 456, 789, 1234}로 5회 반복하였다(Table 12). A0_baseline과 M_mvitv2는 동일 구성(동일 백본·최적 학습률·기본 증강)이므로 seed 123~1234의 반복 실행을 공유하였으며, seed 42는 각각 독립 실행하였다(두 단일 실행값 0.494와 0.487의 차이는 cuDNN 비결정론에 따른 실행 간 변동을 반영한다). A2_no_aug는 Macro-F1 0.643±0.051로, A0_baseline(0.428±0.058) 대비 평균 +0.215 우수하였다. 동일 시드 쌍을 비교하는 설계이므로 대응표본 t-검정을 주 검정으로 사용하였으며, 시드별 결과와 쌍별 차이(paired difference)는 Table 13에, 검정통계량·자유도·정확 p-value·효과크기·95% 신뢰구간은 Table 14에 제시한다. 쌍별 차이는 5개 시드 모두 양수(+0.138 ~ +0.290)로 개선 방향이 일관되었고, 대응표본 t-검정 결과 $t(4) = 8.90$, $p = 0.0009$(양측), Cohen's $d_z = 3.98$, 평균 차이의 95% 신뢰구간은 [+0.148, +0.282]였다. Wilcoxon 부호순위검정은 n=5에서 정확검정의 최소 양측 p-value가 0.0625로 제한된다는 점을 명시하여 보조 검정으로만 제시한다(관측값 $W^+ = 15$, exact $p = 0.0625$). FAR의 경우 단일 실행에서는 A2(0.045)가 A0(0.130)보다 훨씬 낮았으나, 다중 시드 기준으로는 A2(0.126±0.058)와 A0(0.148±0.042) 간 차이가 통계적으로 유의하지 않아(대응표본 t-검정, $t(4) = -0.69$, $p = 0.527$) 단일 실행의 낮은 FAR은 과대평가된 것으로 해석된다.

Table 12. Multi-seed results (n=5, mean ± standard deviation)

ID	A0_baseline	**A2_no_aug**	MViTv2-S
Macro-F1	0.428 ± 0.058	**0.643 ± 0.051**	0.426 ± 0.056
Accuracy	0.484 ± 0.073	**0.664 ± 0.045**	0.482 ± 0.071
F1 (strong)	0.167 ± 0.018	**0.549 ± 0.060**	0.166 ± 0.017
FAR	0.148 ± 0.042	**0.126 ± 0.058**	0.150 ± 0.041

주: A0_baseline과M_mvitv2(M 시리즈)는 동일 구성이며 seed 123~1234의 실행을 공유한다(seed 42만 독립 실행). 따라서 두 열의 통계는 독립적 반복이 아니다

Table 13. Per-seed Macro-F1 and paired differences (A2_no_aug − A0_baseline)

Seed	A0_baseline	A2_no_aug	Paired diff.
42	0.4943	0.7127	+0.2184
123	0.3818	0.5918	+0.2101
456	0.3545	0.6448	+0.2903
789	0.4492	0.6686	+0.2194
1234	0.4586	0.5964	+0.1378
mean ± sd	0.428 ± 0.058	0.643 ± 0.051	+0.215 ± 0.054

Table 14. Statistical test summary for Macro-F1 (A2_no_aug vs. A0_baseline, n=5)

Item	Paired t-test (primary)	Wilcoxon signed-rank (supplementary)
Statistic	$t(4) = 8.90$	$W^+ = 15.0$
Sidedness	two-sided	two-sided
Exact / asymptotic	exact	exact (min two-sided p = 0.0625 at n = 5)
p-value	0.0009	0.0625
Effect size	Cohen's $d_z = 3.98$	–
95% CI of mean diff.	[+0.148, +0.282]	–

10. Temporal Generalization (Series T)

이벤트 기반 무작위 분할은 정보 누수를 줄이지만, 동일 기간의 분포를 학습·평가가 공유한다는 한계가 있다. 이를 보완하는 보조 실험으로, 과거 기간 데이터로 학습하고 이후 기간 데이터로 평가하는 시간적 분할(temporal split) 실험을 추가하였다. MViTv2-S 기준으로 Macro-F1 0.772, 정확도 0.805, 강한 비 클래스 F1이 0.904로 무작위 분할보다 높게 나타났다(Table 15).

그러나 이 수치는 시간적 일반화 성능이 우수하다는 근거가 아니다. 시간적 분할의 평가 구간은 가장 분류가 어려운 약한 비(weak) 비율이 무작위 분할의 49.3%에서 16.7%로 급감하고, 상대적으로 쉬운 무강우가 28.1%에서

www.kci.go.kr

58.8%로 증가한다. 따라서 두 분할 방식의 Macro-F1을 직접 비교하는 것은 부적절하며, Table 15는 시간적 분할 조건에서도 학습이 붕괴하지 않음을 확인한 보조 결과로만 해석해야 한다. 서로 다른 기간·평가셋 구성에 대한 엄밀한 시간적 일반화 검증은 후속 과제로 남긴다.

Table 15. T Series – Temporal-split evaluation (MViTv2-S, 448², 30s)

Exp.	T_temporal_split
Macro-F1	0.772
Acc.	0.805
POD	0.941
FAR	0.251
CSI	0.715
F1(strong)	0.904

11. Discussion

11.1 Why Standard Augmentation Backfires

증강 제거의 이득을 설명하는 세 가지 근거를 제시한다. 첫째, 효과가 모델 구조가 아닌 학습 조건의 차이에서 비롯된다. A 시리즈와 M 시리즈는 모두 동일한 MViTv2-S를 사용하므로, 증강을 적용한 0.487(M)에서 증강을 제거한 0.713(A2)으로의 향상은 증강 제거가 주된 요인으로 해석된다. 다만 학습 stochasticity, early stopping 시점, 클래스 분포의 영향이 완전히 배제되는 것은 아니며, 5개 시드 평균(0.428 → 0.643)의 일관된 개선 경향이 이 해석을 뒷받침한다.

둘째, 색상 지터는 강우 판별에 중요한 저수준 광측정 단서를 직접 교란한다. 빗줄기의 밝은 사선, 노면 젖음에 따른 색조 변화, 시정 저하로 인한 색온도 등은 강우강도 등급을 구분하는 핵심 단서인데, 색상 지터는 이를 무작위로 변형하여 모델이 이러한 단서를 학습하지 못하게 할 수 있다. 다만 IV장 7절의 분해 실험에서 색상 지터 제거만으로는 개선이 나타나지 않았으므로, 이 메커니즘은 플립과의 결합 효과 안에서 부분적으로 작동하는 것으로 해석해야 한다.

셋째, Grad-CAM 분석에서 증강을 제거한 A2는 빗줄기 등 국소 영역에 집중된 주의를, 증강을 적용한 A0는 강한 초점 없이 상대적으로 분산된 주의를 보이는 경향이 관찰된다(Fig. 3). Fig. 3은 주간·야간× 강우강도별 대표 샘플 6종(1~6행)과 오분류(no_rain→weak) 실패 사례 2종(7~8행)의 비교로, 단일 사례가 아닌 다수 샘플에 대한 관찰이다. 다만 실패 사례에서는 A2 역시 강우와 무관한 국소 영역에 주의가 형성되어 오분류로 이어졌으며, 정량적 localization 분석은 수행하지 않았으므로 이 결과는 증강 제거가 강우 관련 영역에 대한 주의 형성을 도울 수 있다는 해석을 지지하는 보조 근거로 한정된다.

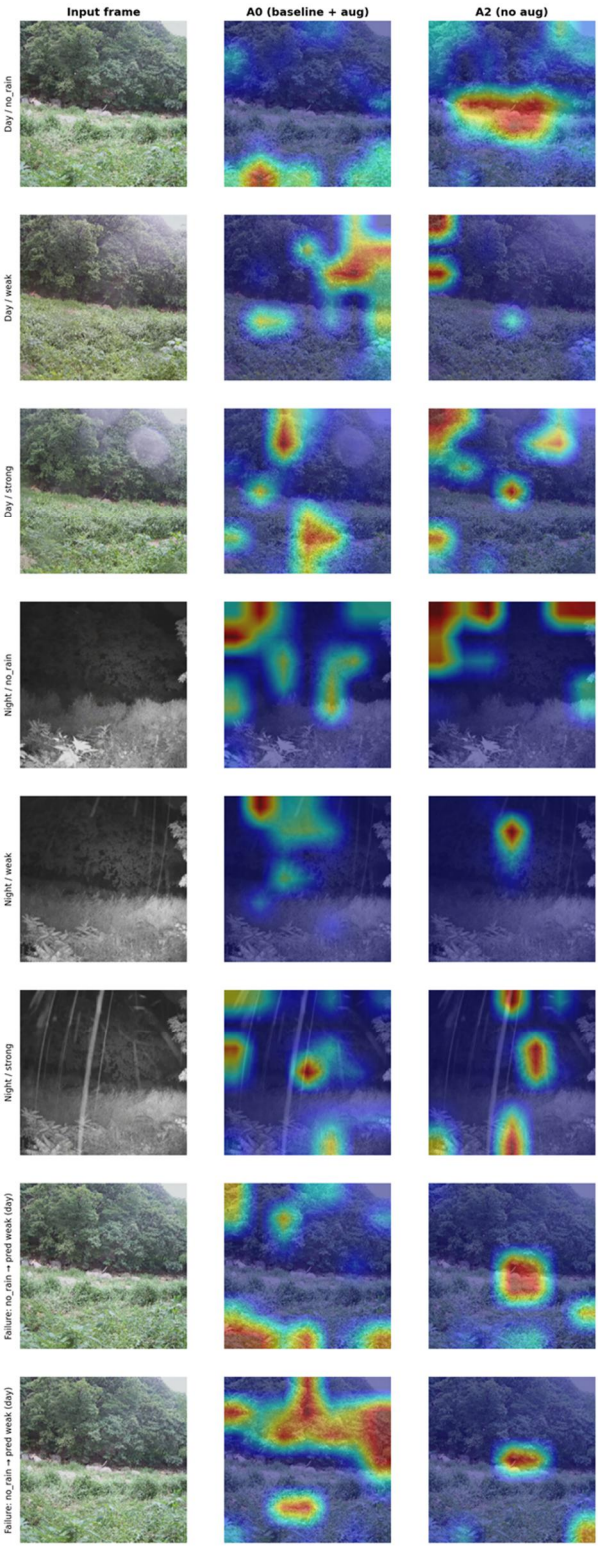

Fig. 3. Grad-CAM comparison (A0 vs A2) for day/night × intensity classes (rows 1–6) and misclassified cases (rows 7–8)

www.kci.go.kr

11.2 Operational Trade-offs

운영 목적에 따라 구성 선택이 달라진다. 강우 누락 최소화가 중요하면 POD가 가장 높은 SlowFast R50이 유리하나 오경보가 많고, 균형 잡힌 식별과 강한 비 탐지가 중요하면 Macro-F1과 강한 비 클래스 F1이 가장 높은 A2(증강 제거) 구성이 적합하다. 다만 A2의 낮은 오경보율은 단일 실행에서 두드러졌을 뿐 다중 시드 기준으로는 통계적으로 유의하지 않으므로, 이 점을 운영 적용 시 고려해야 한다. 448^2 고해상도는 강한 비 탐지력을 추가로 향상시키므로, 자원 여건이 허용된다면 무증강+고해상도 조합의 평가가 유망하다.

11.3 Limitations

1) 데이터는 단일 관측 지점에서 카메라3대(다채널)로 수집하였다. 세 카메라가 화각 다양성을 일부 제공하고 시간적 분할로 시간 일반화를 점검하였으나, 모두 같은 지점·배경·기후에서 얻은 것이므로 다른 지점으로의 공간적 일반화는 검증되지 않았다. 서로 다른 지역의 다지점 교차 검증이 후속 과제로 필요하다.

2) 해상도 분석(R 시리즈)과 절제 연구(A 시리즈)는 모두 동일 백본(MViTv2-S)으로 수행하였으나, R 시리즈는 표준(증강 포함) 파이프라인을 사용하였다. 증강 제거(A2)와 고해상도(448^2)를 결합한 구성은 아직 평가되지 않아 추가 향상의 여지가 있다.

3) 다중 시드 반복은 A0·A2·M 세 구성(각 5 시드)에 한정되며, 증강 분해 및 베이스라인 등은 단일 시드 결과이다. 시드 수(n=5)가 적어 더 많은 반복으로 추정 정밀도를 높일 여지가 있다고, 특히 증강 분해 실험의 비가산적 상호작용 관찰은 다중 시드 반복으로 확정되어야 한다.

4) 라벨은 우량계 측정값 기반이므로 사람의 판독 오류는 없으나, 전도형 우량계 자체의 측정 불확실성과 우량계-영상 간 시간 정합 오차(III장 1.2절)가 라벨 노이즈로 작용할 수 있다. 특히 클래스 경계(r=7.5 mm/hr) 부근 클립의 라벨 불확실성이 크다.

5) 카메라별 성능 차이와 카메라 간 일반화(한 카메라로 학습하고 다른 카메라로 평가)는 분석하지 않았다. 또한 계절 변화, 야간·조명 조건 변화에 따른 성능 변동도 체계적으로 평가하지 않았다.

6) 데이터셋은 사내 보유 자료로 비공개이며, 이는 결과의 외부 재현을 제약한다. 본 논문에서는 이를 보완하기 위해 데이터 구성·분할 절차와 학습 설정을 상세히 기술하였다.

V. Conclusions

본 논문은 소규모 CCTV 영상을 활용한 3-클래스 강우강도 분류에서, 표준 데이터 증강의 제거가 성능을 큰 폭으로 개선함을 통제 실험으로 확인하였다. 동일 백본(MViTv2-S)을 사용할 때 증강을 제거한 구성은 5개 시드 평균 Macro-F1 0.643±0.051(최고 단일 실행 0.713)로 증강 적용 베이스라인(0.428±0.058)을 모든 시드에서 일관되게 상회하였다(대응표본 t-검정, $t(4) = 8.90$, $p = 0.0009$, 양측). 증강 분해 실험(단일 시드)은 어느 한 증강의 제거만으로는 성능이 회복되지 않고 두 증강을 모두 제거할 때만 향상이 나타나는 비가산적 양상을 보였다. 고해상도(448^2) 입력은 강한 비 클래스 식별력을 추가로 향상시키며, 다중 시드 분석은 소규모 데이터 연구에서 단일 실행 결과의 불안정성을 명확히 드러냈다. 이러한 결과는 단일 지점 CCTV 강우 영상이라는 본 논문의 조건에서 도메인 특화적인 증강 설계의 중요성을 보여주는 사례이며, 향후 다지점 공간 일반화 검증 및 무증강·고해상도 결합 실험을 핵심 과제로 남긴다.

ACKNOWLEDGEMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government(MSIT)(IITP-2026-RS-2022-00156360)

REFERENCES

[1] K. Garg, and S. K. Nayar, "Vision and Rain," International Journal of Computer Vision, Vol. 75, No. 1, pp. 3-27, 2007. DOI: 10.1007/s11263-006-0028-6

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the CVPR, pp. 770-778, 2016. DOI: 10.1109/CVPR.2016.90

[3] P. Allamano, A. Croci, and F. Laio, "Toward the camera rain gauge," Water Resources Research, Vol. 51, No. 3, pp. 1744-1757, 2015. DOI: 10.1002/2014WR016298

[4] R. Zen, et al., "Rainfall Estimation from Traffic Cameras," Proceedings of the DEXA, pp. 18-32, 2019. DOI: 10.1007/

www.kci.go.kr

978-3-030-27615-7_2

[5] F. Zheng, et al., "Toward Improved Real-Time Rainfall Intensity Estimation Using Video Surveillance Cameras," Water Resources Research, Vol. 59, No. 8, 2023. DOI: 10.1029/2023WR034831.

[6] J. B. Haurum, et al., "Is it Raining Outside? Detection of Rainfall using General-Purpose Surveillance Cameras," Proceedings of the CVPRW, pp. 55-64, 2019.

[7] H. Yin, et al., "Estimating Rainfall Intensity Using an Image-Based Deep Learning Model," Engineering, Vol. 21, No. 2, pp. 162-174, 2023. DOI: 10.1016/j.eng.2021.11.021

[8] ZF. Zheng, et al., "A Bayesian deep learning approach for video-based estimation and uncertainty quantification of urban rainfall intensity estimation," Journal of Hydrology, Vol. 640, pp. 131701, 2024. DOI: 10.1016/j.jhydrol.2024.131706

[9] X. Wang, et al., "Rainfall observation using surveillance audio," Applied Acoustics, Vol. 186, pp. 108478, 2022. DOI: 10.1016/j.apacoust.2021.108478

[10] X. Wang, et al., "Near-infrared surveillance video-based rain gauge," Journal of Hydrology, Vol. 618, pp. 129173, 2023. DOI: 10.1016/j.jhydrol.2023.129173

[11] R. Rajabi, N. Faraji, and M. Hashemi, "An efficient video-based rainfall intensity estimation employing different recurrent neural network models," Earth Science Informatics, Vol. 17, pp. 2367-2380, 2024. DOI: 10.1007/s12145-024-01290-x

[12] C. Feichtenhofer, "X3D: Expanding Architectures for Efficient Video Recognition," Proceedings of the CVPR, pp. 203-213, 2020. DOI: 10.1109/CVPR42600.2020.00028

[13] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "SlowFast Networks for Video Recognition," Proceedings of the ICCV, pp. 6202-6211, 2019. DOI: 10.1109/ICCV.2019.00630

[14] Y. Li, et al., "MViTv2: Improved Multiscale Vision Transformers for Classification and Detection," Proceedings of the CVPR, pp. 4804-4814, 2022. DOI: 10.1109/CVPR52688.2022.00476

[15] J. Yosinski, et al., "How transferable are features in deep neural networks?" Proceedings of the NeurIPS, pp. 3320-3328, 2014. arXiv DOI: *https://doi.org/10.48550/arXiv.1411.1792*)

[16] C. Shorten, and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, Vol. 6, No. 1, pp. 60, 2019. DOI: 10.1186/s40537-019-0197-0

[17] M. Raghu, et al., "Transfusion: Understanding Transfer Learning for Medical Imaging," Proceedings of the NeurIPS, pp. 3342-3352, 2019. DOI: *https://doi.org/10.48550/arXiv.1902.07208*)

[18] WMO, "Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8)," 2018.

[19] W. Kay, et al., "The Kinetics Human Action Video Dataset," arXiv:1705.06950, 2017. DOI: 10.48550/arXiv.1705.06950

[20] M. Fiallos-Salguero, S.-T. Khu, J. Guan, and M. Wang, "Toward accurate and scalable rainfall estimation using surveillance camera data and a hybrid deep-learning framework," Environmental Science and Ecotechnology, Vol. 25, pp. 100562, 2025. DOI: 10.1016/j.ese.2025.100562

[21] X. Wang, H. Chen, A. Zhou, and Y. Chen, "Rainfall intensity estimation at night using deep learning and urban surveillance cameras in Jiangsu Province, China," Journal of Hydrology: Regional Studies, Vol. 64, pp. 103112, 2026. DOI: 10.1016/j.ejrh.2026.103112

[22] H. Gupta, O. Kotlyar, H. Andreasson, and A. J. Lilienthal, "Video WeAther RecoGnition (VARG): An Intensity-Labeled Video Weather Recognition Dataset," Journal of Imaging, Vol. 10, No. 11, pp. 281, 2024. DOI: 10.3390/jimaging10110281

Authors

Ju-Hwan Han received his M.S. degree in AI and Software Convergence from Soongsil University, Seoul, Korea, in 2023, where he is currently pursuing his Ph.D. degree. He is interested in computer vision and deep learning.

Gye-Young Kim received his B.S., M.S, and Ph. D. degrees in Computer Science from Soongsil University, Seoul, Korea, in 1990, 1992, and 1996, respectively. From March 1996 to February 1997, he served as a Post-Doctoral Fellow at the Electronics and Telecommunication Research Institute (ETRI), Daejon, Korea. From March 1997 to February 2001, he was a Senior Researcher at the Korea Electric Power Research Institute (KEPRI), Korea. Since March 2001, he has been a Professor in the School of Computing at Soongsil University. From March 2008 to June 2008, he was a Visiting Professor at the India Institute of Technology, Bombay, India. His primary research interests include computer vision, multimedia databases, augmented reality, human-computer interfaces, and biometrics.

www.kci.go.kr