

정현구\*, 강현수\*\*\*

충북대학교 정보통신공학부

e-mail : wjdgusrn951@naver.com, hskang@chungbuk.ac.kr

## Deep learning-based video abnormal situation search system

Hyun-Gu Jung\*, Hyun-Soo Kang\*\*\*

Chungbuk National University

### Abstract

In This paper, we introduce a deep learning-based method for the automatic detection and classification of anomalies in video streams. We pre-trained our model on a diverse set of traffic-related images extracted from the ImageNet dataset using the ConvNext technique and then applied it to a video anomaly detection model. Experimental results demonstrated that the proposed method significantly outperformed existing methods, particularly in terms of real-time processing speed and accuracy. This illustrates the effectiveness of the ConvNext architecture in rapid and accurate recognition of traffic incidents and emergencies within video streams.

### I. 서론

최근 디지털 감시 시스템은 공공 안전과 보안을 위한 중요한 도구로 부상하였다. 특히, 도시화와 인구 밀도의 증가로 인해 발생하는 교통사고를 포함한 이상 상황의 자동 감지는 이 분야의 중대한 연구 과제로 자리 잡았다. 그러나 기존의 감시 시스템들은 대량의 비디오 데이터를 처리하는 데 있어 상당한 인적 자원을 필요로 하며, 때때로 사고의 신속한 탐지와 대응에 실패하기도 한다. 이러한 문제를 해결하기 위해, 본 연구는 깊은 신경망을 활용하여 비디오 스트림에서 자동으로 이상 상황을 탐지하고 분류하는 방법을 제안한다.

최근 딥러닝 기술의 급속한 발전은 이미지와 비디오 분석 분야에서 큰 변화를 가져왔다. 여러 모델 중 ConvNext는 최근 아키텍처와 학습 기법의 기술을 통해 주목받고 있다. ConvNext는 높은 수준의 특징 추출 능력과 계산 효율성을 갖추고 있어, 실시간 비디오 처리와 같은 고성능이 요구되는 작업에 적합하다.

본 논문은 ImageNet 데이터셋을 활용하여 ConvNext 모델을 사전 훈련하고, 이를 실시간 교통사고 탐지 시스템에 적용하였다. 이어서, 제안된 방법의 구조와 학습 프로세스를 자세히 설명하며, 실험 설정 및 결과 분석을 통해 모델의 성능을 평가한다. 실험을 통해, 제안된 모델이 기존의 방법들보다 우수한 성능이 보임을 확인하였으며, 이는 교통 관리 시스템과 자율주행 자동차의 발전에 기여할 것으로 기대된다.

본 연구에서는 '사고(accident)' 데이터와 '비사고(no accident)' 데이터를 함께 사용하여 딥러닝 모델을 훈련하고 평가하였다. 구체적으로, 주요 데이터셋으로 활용한 ImageNet의 Dataset 은 'accident'로 레이블 된 13,338장의 이미지와 'no accident'로 분류된 25,663장의 이미지로 구성되었다.

### II. 본론

본 연구에서는 '사고(accident)' 데이터와 '비사고(no accident)' 데이터를 함께 사용하여 딥러닝 모델을 훈련하고 평가하였다. 구체적으로, 주요 데이터셋으로 활용한 ImageNet의 Dataset 은 'accident'로 레이블 된 13,338장의 이미지와 'no accident'로 분류된 25,663장의 이미지로 구성되었다.

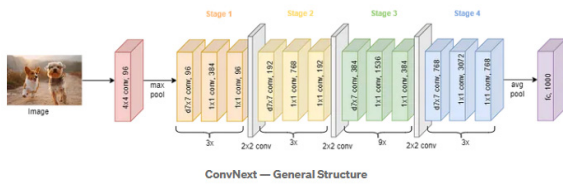


그림1. ConvNeXt - General Structure

ConvNeXt Large는 사전 학습된 데이터를 바탕으로 강력한 특징 추출 능력을 갖추고 있어 복잡한 비주얼 패턴을 잘 파악할 수 있어 본 연구에 매우 적합하다. [그림 1]에서 볼 수 있듯이, ConvNeXt 구조는 여러 단계의 변환을 거치며, 각 단계는 입력 이미지의 정보를 점점 더 추상적인 형태로 변환한다. 초기 단계에서는 주로 이미지의 저수준 특징을 인식하며, 모델이 깊어질수록 더 복잡한 고수준 특징을 학습한다.

이러한 구조적 특성 덕분에, ConvNeXt Large는 사전 학습된 데이터로부터 다양한 특징을 강력하게 추출할 수 있으며, 이를 통해 비디오 내 이상 상황과 같은 복잡한 패턴을 정확히 분류하고 인식하는 데 매우 효과적이다. 모델의 마지막 단계에서는 고수준의 특징이 완전 연결 레이어를 통해 최종 출력으로 변환된다.

참고문헌 [4]에 나타난 바와 같이, ConvNeXt 아키텍처의 여러 변형에 걸쳐 훈련과 검증을 통해 얻어진 성능을 증명한다. 특히 ConvNeXt-L의 경우 APbox 값이  $AP_{50} \text{ box}$ 는 73.8%,  $AP_{75} \text{ box}$ 는 59.5%로 나타났다. 또한 APmask 측면에서도 각각 71.3%, 51.7%로 나타났다. 이는 고해상도의 사진 데이터에 대한 높은 인식 능력과 더불어, 실시간 비디오 스트림에서 이상 상황을 정밀하게 구분해낼 수 있음을 시사한다.

이상상황 검출 시스템에서 모델의 효율성 역시 중요한 측면이다. ConvNeXt Large 모델은 상당히 높은 계산 복잡성을 지니고 있음에도 불구하고 FPS는 10.0을 기록하며, 실시간 처리가 가능한 속도를 유지하는 것으로 나타났다.

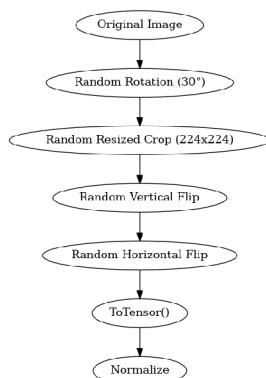


그림2. Flowchart of Image Preprocessing Steps

ConvNeXt Large 모델의 성능을 최적화하기 위해 [그림2]와 같이 다양한 이미지 전처리 기법을 적용하였다. 이미지는 최대 30도의 임의 각도로 회전(RandomRotation(30))되어, 모델이 다양한 방향에서의 객체를 인식할 수 있도록 하였다. 이어서, 이미지는 무작위로 크롭 및 크기 조정(RandomResizedCrop(224))되어, 모델이 핵심적인 특징을 학습하며 과적합을 방지하도록 설계되었다. 또한, 수직 및 수평 뒤집기(RandomVerticalFlip과 RandomHorizontalFlip)를 통해 이미지의 방향성에 대한 모델의 민감도를 줄였다. 마지막으로, 이미지는 텐서 형식으로 변환(ToTensor())되고, 각 채널별로 정규화(Normalize)되어, 모델의 훈련과정에서 안정성과 수렴 속도가 향상되었다.

모델의 평가를 위해 혼동 행렬(confusion matrix)을 사용하였는데, 이를 통해 True Positive, False Positive, True Negative, False Negative 비율을 상세히 분석할 수 있었다. '사고'로 정확히 예측된 경우(True Positive) 1,645건, '비사고'로 정확히 예측된 경우(True Negative) 3,699건을 보여주었다. 오류 분석에서 '사고'로 잘못 분류된 '비사고'(False Positive) 사례는 97건, '비사고'로 잘못 분류된 '사고'(False Negative) 사례는 409건이었다. 이 데이터 분석을 위해 총 5,850개의 테스트 데이터를 사용하였다.

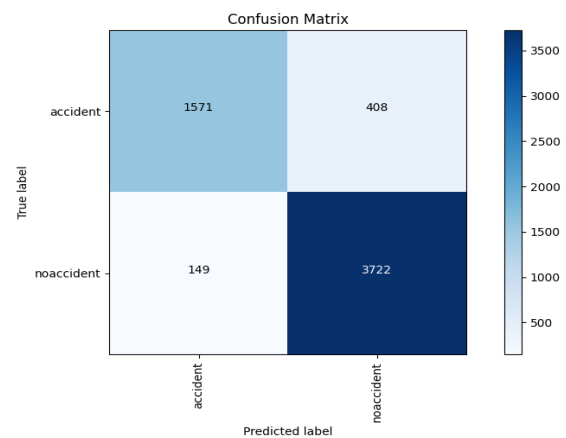


그림 2. Confusion Matrix

혼동 행렬에서는 핵심 성능 지표인 Precision(정밀도), Recall(재현율), 그리고 F1 Score를 계산하였다. 이러한 결과로부터 다음의 성능 지표들이 도출되었다. Precision(정밀도)은 0.91, Recall(재현율)은 0.96, 그리고 F1 Score는 0.93로 계산되었다. 이는 모델이 '사고'로 예측한 사례 중 실제 '사고'인 비율이 높다는 것을 의미하는 정밀도와, 실제 '사고' 사례 중 모델이 '사고'로 올바르게 예측한 비율이 높다는 것을 의미하는 재현율, 두 지표의 균형을 평가하는 F1 Score가 모두 높은 값을 나타내고 있다.

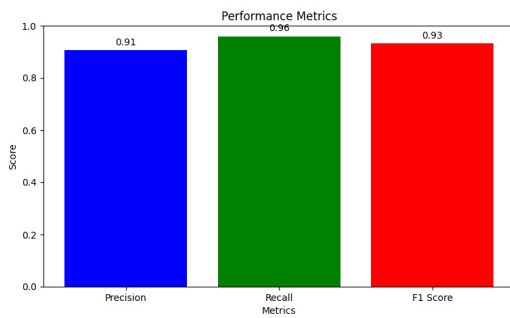


그림 3. Performance Metrics

이러한 높은 성능 지표는 ConvNext Large 모델이 실제 상황에서 교통사고를 탐지하는 데 매우 효과적임을 시사한다. 이는 실시간 교통 감시 시스템에 모델을 통합할 때 중요한 고려 사항으로, 올바른 사고 예측뿐만 아니라 잘못된 알람의 최소화도 중요하다는 점을 강조한다.



그림 4. Traffic Incident Detection: Before and After

결과적으로 앞서 학습된 모델을 활용하여 실시간 비디오 피드에서 자동차 교통사고의 상황을 식별할 수 있다. 비디오 데이터 스트림에서 사고와 비사고 상황을 식별하고, 각 프레임에서의 사고 발생 여부를 'accident' 혹은 'no accident'로 실시간으로 판단하고 화면에 표시하는 기능을 구현하였다. 이 기능은 카메라가 캡처한 영상에서 발생할 수 있는 예기치 않은 상황을 신속하게 감지하고, 적절한 경고를 제공함으로써 교통관제 시스템의 반응 속도와 정확성을 향상시킨다.

### III. 결론 및 향후 연구 방향

본 연구는 비디오 스트림에서의 이상 상황을 자동으로 탐지하고 분류하기 위한 심층 학습 기반 방법을 성공적으로 제안하고 검증하였다. ConvNext Large 모델을 활용하여 ImageNet 데이터셋에서 추출한 교통 관련 이미지들에 대해 사전 학습을 수행하였고, 이를 바탕으로 실제 '사고'와 '비사고' 데이터에 대한 분류 실험을 시행하였다. 이러한 접근 방식은 특히 정밀도와 재현율 면에서 높은 성능을 보였으며, F1 Score가 0.93에 달하는 뛰어난 결과를 얻을 수 있었다. 이는 본 모

델이 현실 세계의 비디오 감시 시스템에 적용될 경우, 실시간으로 사고를 정확하게 탐지하고, 비사고 상황에서의 오경보를 최소화할 수 있는 능력을 갖추고 있음을 시사한다. 그러나, 연구 과정에서 몇 가지 한계점 또한 드러났다. 특히, 특정 유형의 사고나 덜 흔한 상황에서 성능이 충분히 검증되지 않았으며, 다양한 환경과 조건에서의 모델의 범용성을 확보하기 위한 추가적인 연구가 필요하다.

향후 연구에서는 더 다양하고 균형 잡힌 데이터셋을 구축하여 모델의 탐지 능력을 강화하고, 다양한 환경과 조건에서의 범용성을 검증할 수 있도록 할 것이다. 또한, 보다 복잡한 이상 상황에 대한 탐지 능력을 향상시키고, 실시간 데이터 처리의 효율성을 높이는 방향으로 모델을 최적화할 계획이다. 본 연구는 딥러닝 기술을 이용한 비디오 이상 상황 탐지 분야에서 중요한 기여를 하였으며, 실제 교통 감시 시스템 및 자율주행의 발전에 중대한 영향을 미칠 것으로 기대된다.

### Acknowledgment

“이 논문은 과학기술정보통신부 및 정보통신기획평가원의 SW중심대사업(2019-0-01183)의 지원을 받아 작성되었음”

### 참고문헌

- [1] ConvNeXt – “Next Generation of Convolutional Networks, Atakan Erdoğan”, Apr. 2023
- [2] Tushar, Kundan Kumar, Sanjay Kumar – “Object Detection using OpenCV and Deep Learning”, Dec. 2022
- [3] P. Josephinshermila, S. Sharon priya b, K. Malarvizhi c, Ramakrishna hegde d, S. Gokul Pran e, B. Veerasamy f – “Accident detection using Automotive Smart Black-Box based Monitoring system”, June. 2023
- [4] ConvNext: “The Return Of Convolution Networks” Feb. 2022, AUGMENTED A.I.