

# SwinVP: 비디오 예측 모델 기반 교통사고 위험 예측

진현석, MD Ilias Bappi, David J. Richter, 김경백\*

전남대학교(대학원생, \*교수)

ggyo003@jnu.ac.kr, i\_bappi@jnu.ac.kr, david\_richter@jnu.ac.kr, \*kyungbaekkim@jnu.ac.kr

## SwinVP: Traffic Accident Risk Forecasting Based on Video Prediction Model

Jin Hyeonseok, MD Ilias Bappi, David J. Richter, Kim Kyungbaek\*

Chonnam National University(Graduate student, \*Professor)

### 요약

교통사고는 기물 파손 등으로 인한 재산 피해를 유발함과 더불어 많은 인명 피해와 교통 체증 유발 등 많은 불편함을 야기한다. 이러한 문제를 해결하기 위해 인공지능 모델을 사용하여 교통사고 위험을 예측하는 연구가 진행되고 있으나, 기존 연구들은 시공간 상관관계를 학습하고 주말, 날씨 정보 등 추가적인 데이터를 활용하기 위해 복잡한 아키텍처를 활용해야 한다는 한계가 있다. 이러한 한계를 극복하기 위해 본 논문에서는 최근 성공적으로 경량화가 수행된 비디오 예측연구에서 영감을 받아 높은 사고 감지 및 간단한 방법만으로 학습이 가능한 SwinVP를 제안한다. SwinVP는 추가적인 데이터를 사고 데이터와 함께 입력받아 간단한 방법만으로 학습이 가능하며, SwinTransformer 기반의 인코더를 사용하여 CNN 기반의 기존 비디오 모델 대비 공간적 특징을 보다 효과적으로 추출할 수 있음으로 인해 NYC, Chicago 두 가지 데이터를 사용한 실험에서 더 뛰어난 RMSE, Recall, MAP 성능을 보였다. 또한 기존 교통 예측 모델 대비 약 2% 뛰어난 Recall을 보여 교통사고 위험도 감지 능력이 더 뛰어남을 확인하였다.

### I. 서론

교통사고는 기물 파손 등으로 인한 재산 피해를 유발함과 더불어 전 세계적으로 연간 약 119만 명의 사망자를 발생시켜 이를 감소시키기 위한 조치가 필요하다[1]. 이러한 문제를 해결하기 위해 최근 이미지와 유사한 히트맵 형식의 데이터와 인공지능 모델을 사용하여 교통사고 위험을 예측하는 연구[2-3]가 진행되고 있으나, 시공간 상관관계를 학습하고 주말, 날씨, 교통 흐름 등 교통사고에 영향을 미치는 다양한 요인들을 활용하기 위해 복잡한 아키텍처를 활용해야 한다는 한계가 존재한다.

이러한 한계를 극복하기 위해 본 논문에서는 최근 성공적으로 경량화가 수행된 비디오 예측 모델 연구[4-5]에서 영감을 받아 SimVP[5]와 SwinTransformer[6]를 기반으로 한 SwinVP 모델을 제안한다. SwinVP는 기존 CNN 기반의 비디오 예측 모델 대비 히트맵 데이터의 전역적인 정보를 학습할 수 있어 높은 사고 위험도 감지가 가능하며 사고 데이터와 함께 시간, 날씨, 교통량 등 추가적인 데이터를 함께 입력받아 간단한 방법만으로 학습이 가능하다.

### II. 본론

#### 2.1 SwinVP

본 논문에서는 SimVP[5]와 SwinTransformer[6]를 기반으로 한 SwinVP를 제안하였으며 전체적인 구조는 그림 1과 같다. SimVP[5] 모델은 데이터의 공간적인 특징을 학습하기 위한 Encoder, Decoder 사이에 시간적인 특징을 학습할 수 있는 Translator가 배치되어 있는 구조를 지니며, CNN만을 사용하여 계산 비용을 효과적으로 줄일 수 있다.

그러나 일반적인 이미지와 달리 히트맵 데이터를 사용하여 예측을 수행하고자 할 때, 지역적인 공간적 정보를 추출하는 CNN만으로는 전역적인

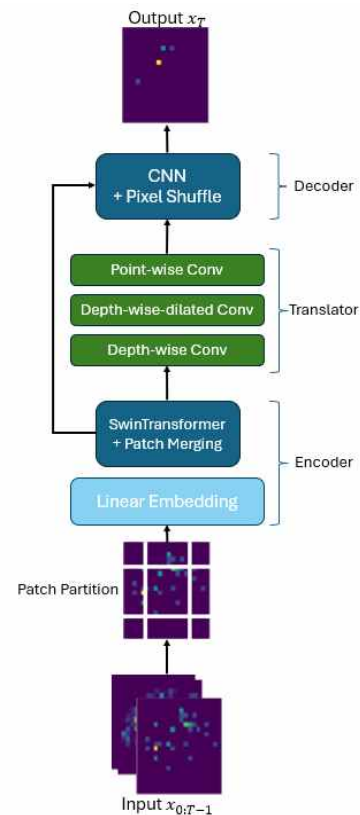


그림 1. SwinVP 구조

공간적 정보를 추출하여 상관관계를 파악하기 어렵다는 한계가 있다. 이러한 한계를 극복하고자 본 논문에서는 공간적 특징을 압축하는 Encoder

구조를 SwinTransformer[6]로 변경한 SwinVP를 제안하였다. SwinVP는 SwinTransformer Encoder를 사용하여 데이터의 공간적 정보를 보다 효과적으로 추출할 수 있다.

## 2.2 데이터셋

본 논문에서 제안한 SwinVP의 성능을 평가하기 위해 선행 연구[2-3]를 참고하여 20x20 사이즈를 갖는 히트맵으로 변환된 NYC[7], Chicago[8] 데이터셋을 사용하였으며 각각 8760, 5832건의 데이터가 포함되어 있다. 각 셀에는 발생한 사고값의 합계가 부여되어 있으며, 사고값은 심각도에 따라 1과 3사이의 값이 부여되어 있다. 그 외 구성 요소는 표 1과 같으며, Min-Max 정규화 적용 및 6:2:2 비율로 분할하여 각각 학습, 검증, 테스트에 사용하였다.

표 1. 데이터셋 구성 요소

데이터셋	구성 요소
NYC	시간, 요일, 휴일
	POI(Point of Interest)
	온도
	날씨
	교통량
Chicago	시간, 요일, 휴일
	온도
	날씨
	교통량

## 2.3 실험

실험은 Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz, RTX 3070, pytorch 2.1.2 환경에서 Adam optimizer, 32 batch size로 설정 후 조기멈춤과 함께 200 epoch 동안 학습을 진행하며 7시간 동안의 과거 데이터를 사용하여 미래 1시간 동안의 사고 예측을 수행하도록 하였으며, 손실함수로 MAE와 MSE의 합계인 L1L2 Loss, 감지 능력 평가는 전체 데이터셋과 더불어 사고 빈도가 높은 시간 구간을 대상으로 RootMeanSquaredError(RMSE), Recall, MeanAveragePrecision(MAP)를 사용하였다.

제안한 SwinVP의 성능을 평가하기 위해 SimVP[5]와 기존 State-of-the-Art(SOTA) 교통 예측 모델인 C-ViT[3]를 대상으로 감지 능력을 비교 실험 하였으며, 결과는 표 2, 표 3과 같다.

표 2. 교통사고 위험도 감지 능력 비교 결과

Dataset	models	RMSE	Recall	MAP
NYC	C-ViT	<b>7.0053</b>	33.86%	0.1875
	SimVP	10.8232	34.00%	0.1918
	SwinVP	10.8507	<b>35.03%</b>	<b>0.1961</b>
Chicago	C-ViT	<b>9.4456</b>	20.93%	0.0980
	SimVP	13.4410	21.94%	0.0933
	SwinVP	13.3699	<b>22.84%</b>	<b>0.1037</b>

표 3. 사고 빈도가 높은 시간 구간의 교통사고 위험도 감지 능력 비교 결과

Dataset	models	RMSE	Recall	MAP
NYC	C-ViT	<b>6.2658</b>	34.46%	0.1802
	SimVP	9.8490	35.23%	0.1852
	SwinVP	9.8122	<b>35.44%</b>	<b>0.1894</b>
Chicago	C-ViT	<b>7.0353</b>	21.95%	0.1247
	SimVP	10.6161	23.46%	0.1175
	SwinVP	10.5491	<b>24.01%</b>	<b>0.1217</b>

## III. 결론

본 논문에서는 교통사고 위험 예측의 복잡도를 감소시키고, 전역적인 공간적 정보를 활용하기 위해 SwinVP를 제안하였다. SwinVP는 사고 데이터와 시간, 날씨, 교통량 등 추가적인 정보를 한번에 입력받아 예측을 수행하여 보다 간단하고, SwinTransformer Encoder를 기반으로 데이터의 전역적인 공간적 정보를 추출하여 보다 효과적으로 교통사고 위험 예측을 수행할 수 있다. 그러나 RMSE가 낮아 사고가 일어나지 않는 지역에 대한 예측력이 부족하다는 한계가 있다.

향후 연구에서는 추가적인 정보를 결합하는 방법 및 모델 구조 고도화를 수행하여 전체 지역에 대한 예측력을 높여 RMSE를 감소시키고 더불어 보다 장기간의 데이터를 예측하는 Multi-out 성능 검토에 관한 연구를 수행할 예정이다.

## ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업임(IITP-2024-00156 287, 50%).

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음(IITP-2023-RS-2023-00256 629, 50%)

## 참고 문헌

- [1] World Health Organization, "Global status report on road safety 2023" 2023, (<https://www.who.int/publications/i/item/9789240086517>).
- [2] Wang, Beibei, et al. "GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting." Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 5. 2021.
- [3] Saleh, Khaled, Artur Grigorev, and Adriana-Simona Mihaita. "Traffic accident risk forecasting using contextual vision transformers." 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2022.
- [4] Gao, Zhangyang, et al. "Simvp: Simpler yet better video prediction." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- [5] Tan, Cheng, et al. "Simvp: Towards simple yet powerful spatiotemporal predictive learning." arXiv preprint arXiv:2211.12509 (2022).
- [6] Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
- [7] NYC Open Data, (<https://opendata.cityofnewyork.us/>).
- [8] Chicago Data Portal, (<https://data.cityofchicago.org/>).