

# Spatial-Temporal Graph Attention Networks: A Deep Learning Approach for Traffic Forecasting

Zhang, C., James, J. Q., & Liu, Y. (2019).

Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access*, 7, 166246-166256.

2025.03.11.

발표자 : 김성희

# 목차

---

1. 연구 소개 및 배경
2. 기존 연구의 한계점
3. ST-GAT 모델
4. 실험 설계 및 결과
5. 결론 및 한계점

# 1. 연구 소개 및 배경

## 연구 배경

- 교통은 생활과 밀접하게 관련된 중요한 주제로, 미국에서만 2017년 기준 2억 명 이상의 운전자 인구가 존재함
- 실시간 교통 상태 예측은 정부와 관련 기관이 지능형 교통 시스템(ITS)을 개발하는 데 매우 중요한 요소
- 교통 속도 예측(TSP)은 경로 안내, 흐름 제어, 내비게이션 등 다양한 교통 응용 프로그램에 유용함이 입증됨

## 연구 동기

- 기존 모델 기반 접근법은 복잡한 시뮬레이션과 비현실적 가정으로 실제 환경에서 성능 저하 문제 발생
- 통계 기반 방법론과 기존 딥러닝 모델들은 비선형 교통 데이터 처리와 공간-시간 특성 학습에 한계 존재
- 그래프 기반 모델도 정적 공간 의존성만 고려하고 시간에 따라 변화하는 교통 네트워크의 동적 특성을 충분히 반영하지 못함

## 연구 목적

- 교통 네트워크의 동적 공간-시간 의존성을 효과적으로 포착하는 그래프 어텐션 네트워크(GAT)와 LSTM을 결합한 ST-GAT 모델 제안
- 시계열 교통 데이터를 특성 벡터로 변환하는 Speed2Vec 임베딩 메커니즘 개발과 공간 의존성에 대한 해석 가능성 향상
- 노이즈와 축소 그래프에 대한 모델의 견고성 검증 및 기존 최신 기법들보다 우수한 교통 속도 예측 성능 달성

1. 연구 소개 및 배경

2. 기존 연구의 한계점

3. ST-GAT 모델

4. 실험 설계 및 결과

5. 결론 및 한계점

- 1. 연구 소개 및 배경
- 2. 기존 연구의 한계점
- 3. ST-GAT 모델
- 4. 실험 설계 및 결과
- 5. 결론 및 한계점

## 2. 기존 연구의 한계점

### 전통적 접근법

- **모델 기반:** 큐잉 이론과 시뮬레이션 활용이 가능하나 비현실적 가정으로 실제 환경에서 성능 저하
- **통계 기반(ARIMA, 칼만 필터):** 시계열 정상성 가정으로 비선형 교통 데이터에 적용 한계
- **비모수 기반(KNN, SVR):** 복잡한 특성 일부 학습이 가능하나 시공간적 의존성 모델링 제한적

### 딥러닝 접근법

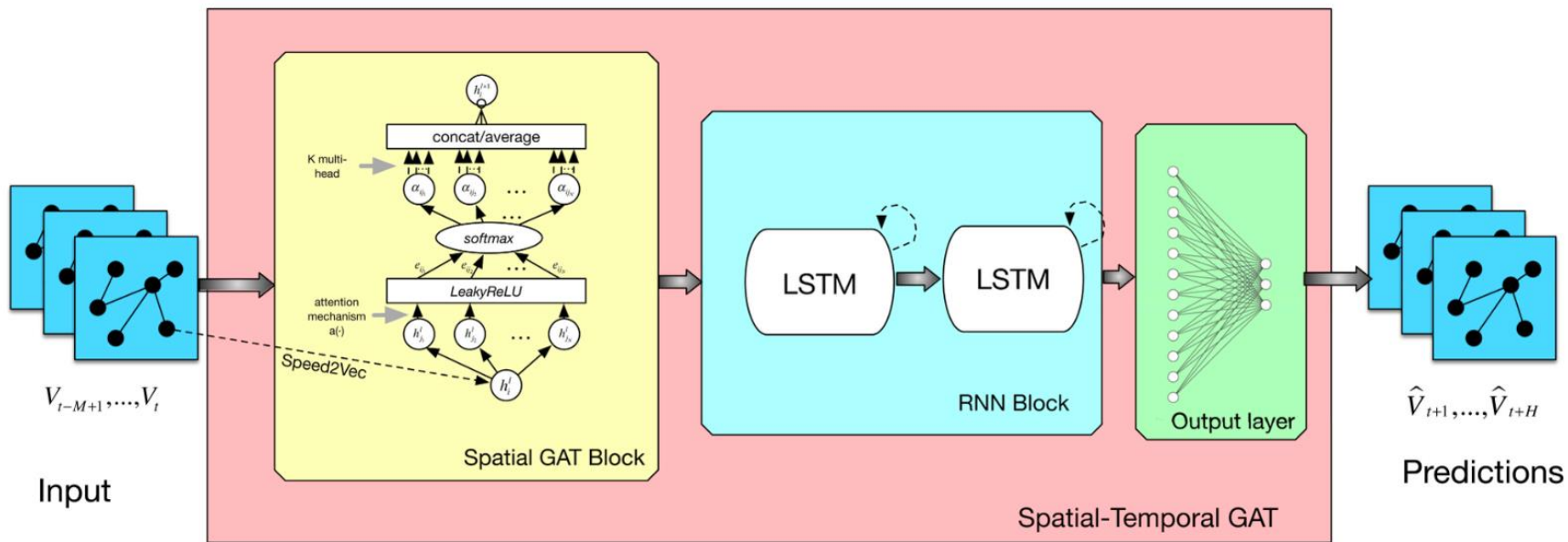
- **일반 딥러닝(DBN, SAE):** 단일 시퀀스 학습만 가능, 공간 정보 활용 부족
- **CNN 기반:** 그리드 구조 데이터에만 적용 가능, 복잡한 도로 네트워크 토폴로지 반영 어려움
- **그래프 기반(GCN, STGCN):** 정적 공간 의존성만 추출, 시간에 따른 동적 변화 포착 실패

### 핵심 한계점

- **동적 공간 의존성 포착 부재:** 시간에 따라 변화하는 도로 간 영향 관계 학습 불가
- **모델 해석 가능성 부족:** 블랙박스 특성으로 예측 근거와 공간적 의존성 이해 어려움
- **노이즈 내성 연구 미흡:** 실제 환경의 측정 노이즈와 결측치에 대한 견고성 검증 부족

### 3. ST-GAT(Spatial-Temporal Graph Attention Networks)

ST-GAT(Spatial-Temporal Graph Attention Networks)는 교통 속도 예측을 위해 제안된 모델로 그래프 어텐션 메커니즘을 활용하여 교통 네트워크의 동적 공간적 의존성을 효과적으로 포착하는 것을 목표로 함



#### ST-GAT 모델 구조

- **Spatial GAT Block(Spatial Dependency Extraction Block)** : 도로 세그먼트 간 공간적 의존성 추출
- **RNN Block(Temporal Feature Extraction Block)** : 시간적 특성 학습
- **Output layer** : LSTM 출력을 변환하여 최종 예측 결과 도출

1. 연구 소개 및 배경
2. 기존 연구의 한계점
3. ST-GAT 모델
4. 실험 설계 및 결과
5. 결론 및 한계점

### 3. ST-GAT(Spatial-Temporal Graph Attention Networks)문제 정의 및 데이터 입/출력 형태

#### 교통 속도 예측 문제의 수학적 정의

$$v_{t-M+1}, \dots, v_t \xrightarrow{f(\cdot)} \hat{v}_{t+1}, \dots, \hat{v}_{t+H},$$

- $v_t \in \mathbb{R}^n$  : 시간 단계  $t$ 에서  $n$  도로 세그먼트의 속도 값
- $f(\cdot)$  : 과거  $M$  시간 단계의 정보가 주어진 다음  $H$ 시간 단계에서 교통 속도를 예측하기 위한 함수

#### 그래프 데이터 표현

$$G_t = (V_t, \mathcal{E}, W)$$

- 노드( $V_t$ ): 시점  $t$ 의 도로 세그먼트의 센서 노드 집합
  - 엣지( $\mathcal{E}$ ) : 엣지 집합
  - 인접행렬( $W$ ): 도로 세그먼트 간 연결
- 초기 인접 행렬 구성: 임계화 된 가우시안 커널 방법으로 도로 네트워크의 기본 구조 정의
  - 노드 간 유클리드 거리를 기반으로 인접 행렬 구축

1. 연구 소개 및 배경

2. 기존 연구의 한계점

3. ST-GAT 모델

4. 실험 설계 및 결과

5. 결론 및 한계점

### 3. ST-GAT(Spatial-Temporal Graph Attention Networks) Spatial GAT Block

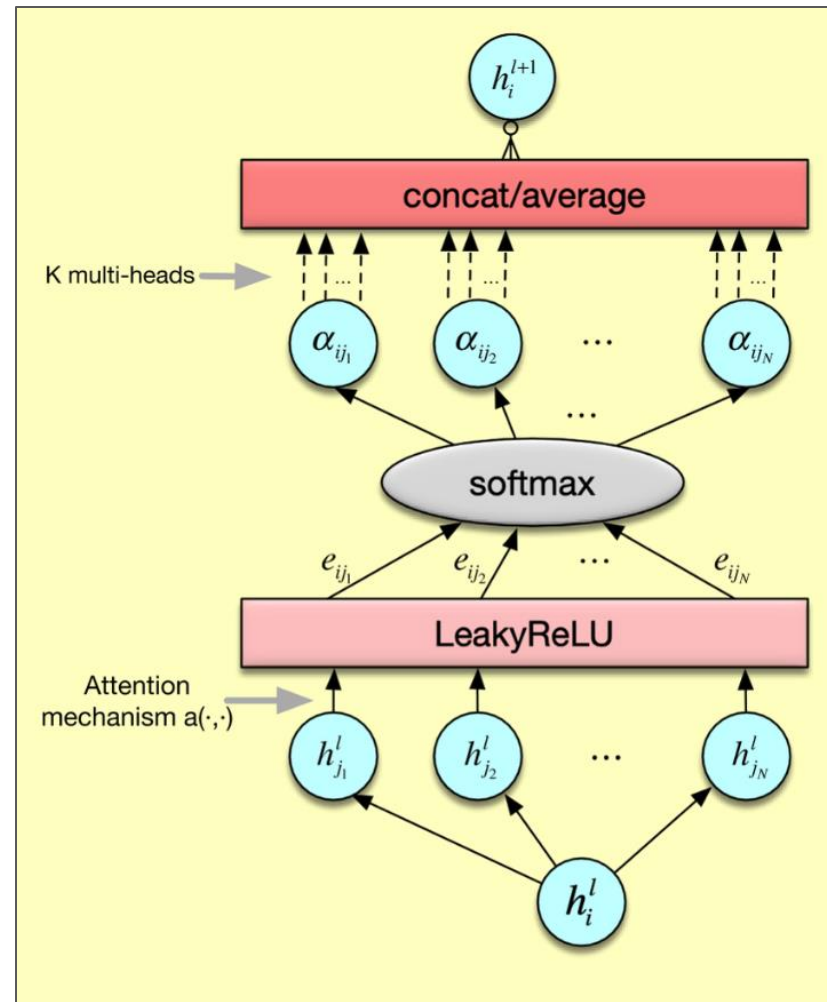
#### 도로 간 공간적 의존성 추출 (GAT)

- Speed2Vec을 통해 생성된 특성 표현을 입력으로 받아 다중 헤드 어텐션 메커니즘을 적용
- 노드 간 어텐션 계수를 계산하고 어텐션 인접 행렬을 생성하여 공간적 의존성을 학습

#### Speed2Vec

- 각 노드의 과거 속도 관측치를 벡터로 변환
- 네트워크 전체 입력 행렬  $H_T^N$  형성

$$H_T^N = \begin{bmatrix} h_1^1 & h_2^1 & \dots & h_T^1 \\ h_1^2 & h_2^2 & \dots & h_T^2 \\ \vdots & \vdots & \ddots & \vdots \\ h_1^N & h_2^N & \dots & h_T^N \end{bmatrix}$$



<K 멀티헤드 어텐션 메커니즘>

1. 연구 소개 및 배경
2. 기존 연구의 한계점
3. ST-GAT 모델
4. 실험 설계 및 결과
5. 결론 및 한계점

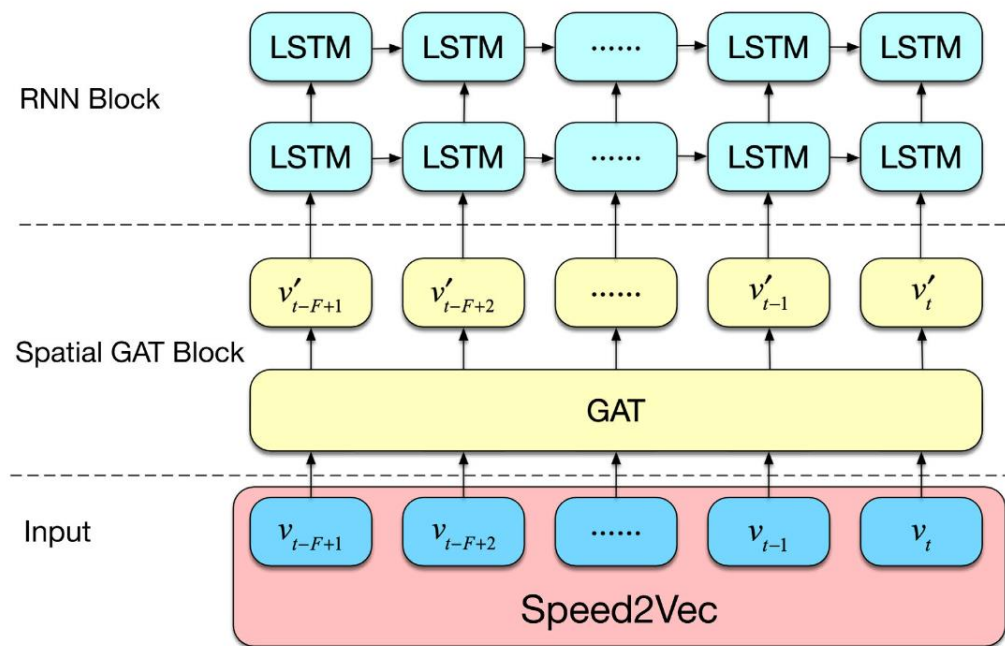
### 3. ST-GAT(Spatial-Temporal Graph Attention Networks) RNN Block & Output Layer

#### 시간 특성 학습 (LSTM)

- 2계층 LSTM 레이어(32, 128 유닛)로 시간적 특성과 장기 의존성 학습
- GAT의 출력(공간적으로 강화된 특성 벡터)을 입력으로 받아 시간적 패턴 학습

#### 최종 아웃풋 (Output Layer)

- LSTM의 출력을 최종 예측 값으로 변환하는 fully-connected 레이어
$$\hat{v}_{t+H} = w \times h_t + b,$$
- $v_{t+H}$ : 미래 시간  $t+H$ 에서의 예측 속도 값
- $w$ : 가중치 행렬로 LSTM의 hidden state를 실제 속도 값으로 변환하는 선형 변환을 수행
- $h_t$ : LSTM의 마지막 단계에서의 hidden state 출력
- $b$ : 편향(bias) 값으로, 선형 변환 후 더해지는 상수



<시공간 특징 업데이트 메커니즘>

1. 연구 소개 및 배경
2. 기존 연구의 한계점
3. ST-GAT 모델
4. 실험 설계 및 결과
5. 결론 및 한계점

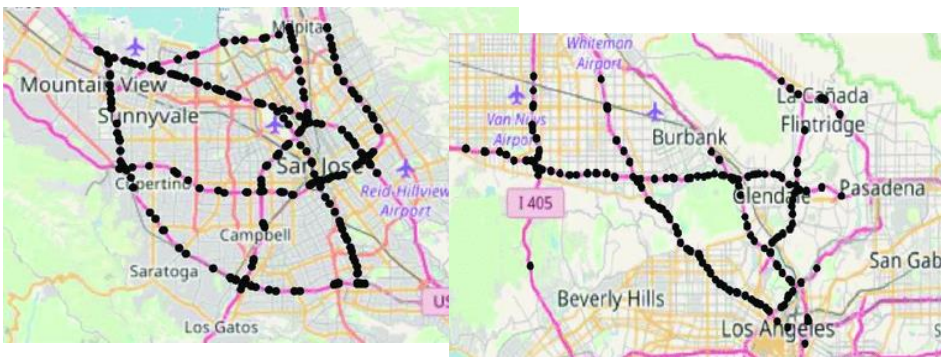


## 4. 실험 설계 및 결과

데이터 및 실험 설계

### PeMSD7 데이터셋

- 출처: 캘리포니아 교통국(Caltrans)
- 기간: 2012년 5월 1일~6월 30일 (평일만)
- 센서: 228개 관측소
- 샘플링: 5분 단위 (하루 288개 관측치)
- 선형 보간법으로 결측치 처리,  
Z-점수 정규화 적용
- 훈련/검증/테스트 분할: 60%/20%/20%



<PeMSD7 데이터셋>

### 그래프 구성

- 임계화 된 가우시안 커널 방법으로 인접 행렬 구축
- 자체 연결 포함, 무방향 그래프 설정
- K-NN 기반 축소 그래프 추가 구성  
( $K=5,10,20,50,100$ )

### 모델 구현 세부사항

- 과거 60분 데이터로 15/30/45분 예측
- 학습률:  $2e-4$ , 배치 크기: 50
- 하이퍼파라미터
  - 어텐션 헤드 수: 8
  - LSTM 레이어: 2층 (32, 128 유닛)
  - Speed2Vec 차원: 12
- 최적화: Adam 옵티마이저
- 손실 함수: MSE (평균 제곱 오차)

## 4. 실험 설계 및 결과

모델 비교 및 평가

### 비교 모델

- 전통적 방법: HA, ARIMA, LSVR
- 딥러닝 방법: FNN, FC-LSTM, DCRNN, STGCN

### 평가 지표

- RMSE (Root Mean Square Error)
- MAE (Mean Absolute Error)
- MAPE (Mean Absolute Percentage Error)

Model	15 min			30 min			45 min		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
HA	7.20	4.01	10.61	7.20	4.01	10.61	7.20	4.01	10.61
ARIMA	9.00	5.55	12.92	9.13	5.86	13.94	9.38	6.27	15.20
LSVR	4.55	2.50	5.81	6.67	3.63	8.88	8.28	4.54	11.50
FNN	4.75	2.74	6.38	6.98	4.02	9.72	8.58	5.04	12.38
FC-LSTM	6.20	3.57	8.60	7.03	3.94	9.55	7.51	4.16	10.10
GCGRU	4.21	2.37	5.54	5.96	3.31	8.06	7.13	4.01	9.99
STGCN	4.04	2.25	5.26	5.70	3.03	7.33	6.77	3.57	8.69
ST-GAT	<b>3.45</b>	<b>2.01</b>	<b>4.76</b>	<b>4.68</b>	<b>2.76</b>	<b>6.57</b>	<b>5.30</b>	<b>3.20</b>	<b>7.86</b>

→ ST-GAT가 모든 시간대, 모든 지표에서 최고 성능 달성

→ MAPE 기준으로 STGCN보다 0.5%(15분), 0.76%(30분), 0.83%(45분) 개선

## 4. 실험 설계 및 결과 하이퍼파라미터 민감도 분석

### 하이퍼파라미터 영향 분석

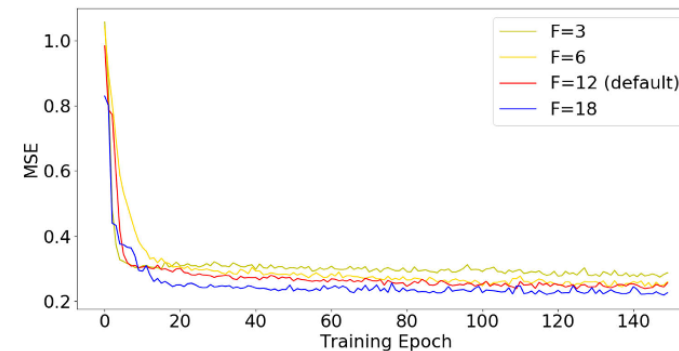
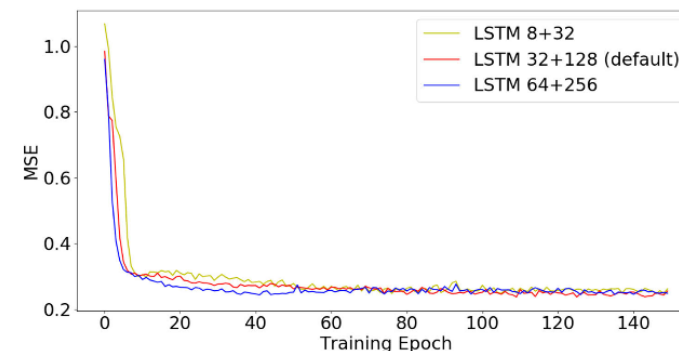
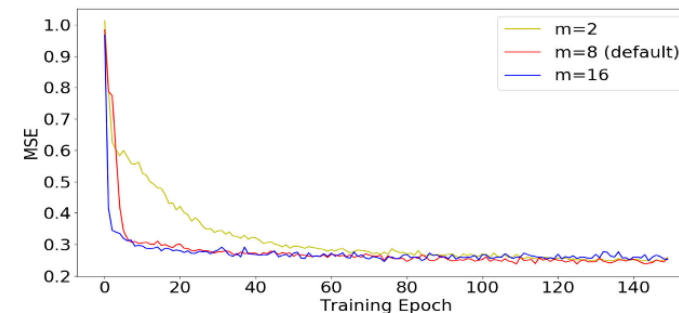
- 어텐션 헤드 수(m): 2/8/16
- LSTM 유닛 수: 8+32/32+128/64+256
- Speed2Vec 차원(F): 3/6/12/18

〈ST-GAT 하이퍼파라미터 분석〉

Hyperparameter	45 min		
	RMSE	MAE	MAPE (%)
m=2	5.55	3.31	8.38
m=8 (default)	5.30	3.20	7.86
m=16	5.22	3.08	7.40
LSTM=8/32	5.38	3.24	7.89
LSTM=32/128 (default)	5.30	3.20	7.86
LSTM=64/256	5.24	3.06	7.38
F=3	5.44	3.28	8.03
F=6	5.37	3.22	7.99
F=12 (default)	5.30	3.20	7.86
F=18	5.20	3.03	7.51

→ 하이퍼파라미터 증가에 따른 성능 향상

→ 8개 어텐션 헤드에서 MSE 수렴 속도 크게 개선



〈어텐션 헤드 수, LSTM 유닛 수, Speed2Vec 차원에 따른 학습 수렴 곡선〉

1. 연구 소개 및 배경

2. 기존 연구의 한계점

3. ST-GAT 모델

4. 실험 설계 및 결과

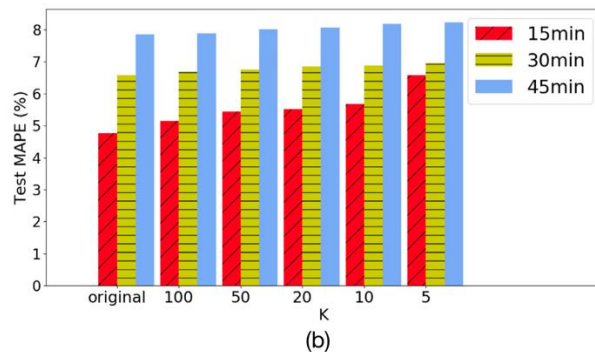
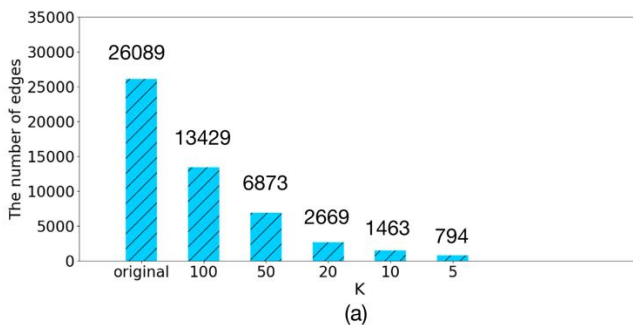
5. 결론 및 한계점

## 4. 실험 설계 및 결과

그래프 크기 효율성 및 노이즈/결측치 견고성 분석

### 그래프 축소에 따른 효율성 검증

- 원래 그래프(26,089개 엣지)에서 K=20일 때 2,669개로 약 90% 감소
- 단기 예측(15분)이 장기 예측(45분)보다 그래프 축소에 더 민감한 경향



### 노이즈와 결측치에 대한 견고성 분석

- 2% 가우시안 노이즈에도 MAPE가 7.86%에서 7.99%로 0.13%만 증가하여 강한 노이즈에 대한 내성 존재
- 0.5% 결측치에서도 MAPE가 8.15%로 0.29% 증가하여 결측치에 상대적으로 취약

<노이즈/결측치 수준별 예측 정확도 비교>

Disturbance		45 min		
		RMSE	MAE	MAPE (%)
Noise	Original	5.30	3.20	7.86
	0.5%	5.31	3.22	7.87
	1%	5.35	3.27	7.94
	2%	5.48	3.35	7.99
Missing	Original	5.30	3.20	7.86
	0.5%	6.83	3.73	8.15
	1%	8.01	3.88	8.49
	2%	9.91	4.91	8.64

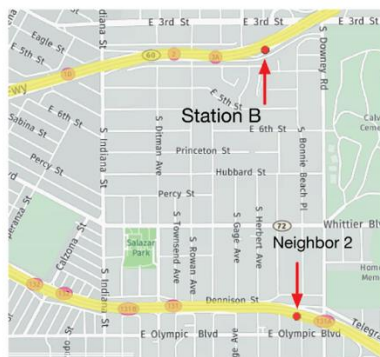
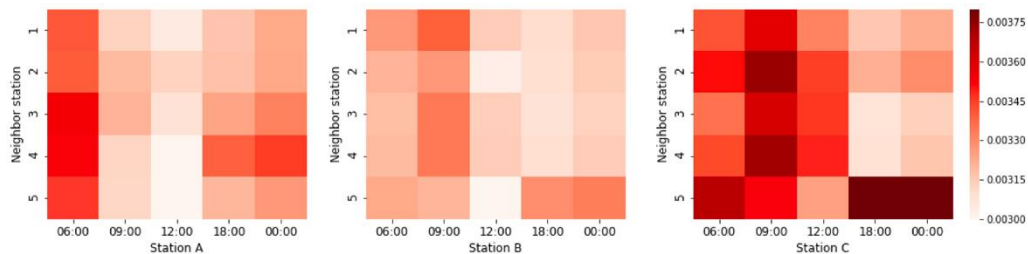
<(a)축소된 그래프 별 총 연결 수 비교, (b)그래프 크기에 따른 예측 성능 비교>

## 4. 실험 설계 및 결과

모델 결과 해석에 대한 시각화

### 어텐션 계수 상관 관계 비교

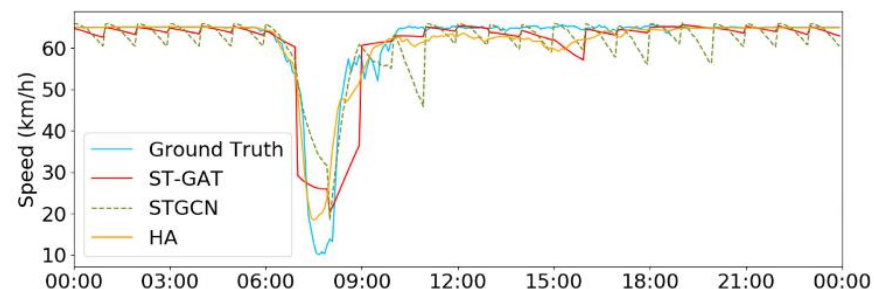
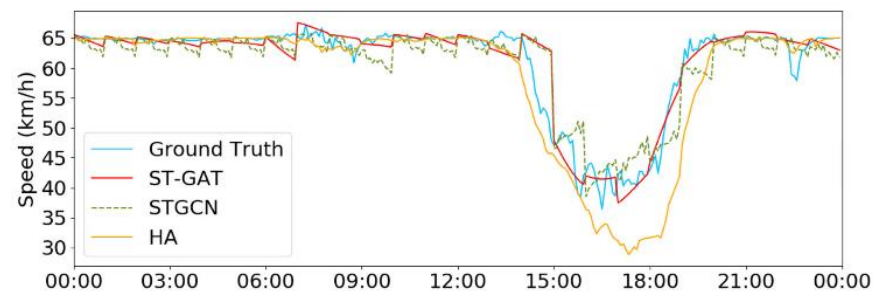
- 각 도로 세그먼트마다 이웃 노드들과의 어텐션 계수 패턴이 서로 다르게 나타나며, 이는 노드 간 공간적 관계가 다양함을 보여줌
- 단순한 지리적 거리보다 도로 네트워크 토폴로지가 어텐션 계수에 더 큰 영향을 미침



<(상)시간별 어텐션 계수 히트맵, (하) 도로 관계에 따른 어텐션 계수 차이>

### ST-GAT와 STGCN의 예측 비교

- ST-GAT는 특히 15:00-21:00 사이의 피크 시간대 시작과 종료 지점을 STGCN보다 더 정확하게 예측
- 더 급격한 속도 변화가 있는 도로 구간에서 ST-GAT는 변화 패턴을 더 민감하게 포착
- LSTM의 효과적인 시간적 특성 학습



<모델별 교통 속도 예측 정확도 비교>

- 1. 연구 소개 및 배경
- 2. 기존 연구의 한계점
- 3. ST-GAT 모델
- 4. 실험 설계 및 결과
- 5. 결론 및 한계점

## 5. 결론 및 한계점

### 주요 연구 성과

- 그래프 어텐션 메커니즘과 LSTM을 결합한 ST-GAT 모델 개발로 교통 네트워크의 동적 공간-시간 의존성 효과적으로 포착
- Speed2Vec 임베딩 방법 제안으로 시계열 교통 데이터를 그래프 어텐션 네트워크에 직접 적용 가능
- 실제 데이터셋에서 기존 최신 모델 대비 15분/30분/45분 예측 성능 각각 0.5%, 0.76%, 0.83% 개선
- 어텐션 계수 시각화를 통한 교통 네트워크 동적 의존성 해석 가능성 제공
- 90% 축소된 그래프에서도 안정적 성능 유지 및 측정 노이즈에 대한 강한 견고성 입증

### 한계 및 향후 연구 방향

- 결측치에 대한 상대적 취약성 및 대규모 그래프에서의 계산 부담 존재
- 결측치 처리 메커니즘 개발 및 외부 요인(날씨, 교통 통제 등) 통합 필요
- 교통 흐름, 사고 예측 등 다양한 응용 분야로 확장 가능성 존재

### 연구 활용성

- 네비게이션 시스템의 도착 시간 예측 및 실시간 최적 경로 제안 정확도 향상
- 미래 교통 상황을 예측하여 선제적으로 대응할 수 있기에 공유 모빌리티 서비스의 수요-공급 밸런싱 및 동적 가격 책정 모델에 적용 가능

**Q&A**