

2025년 자율주행 인공지능 챌린지: 시멘틱 세그멘테이션

발표자 : 이정윤

2025.11.14

팀명: SSU VIP

팀원: 이정윤, 임규일, 길다영

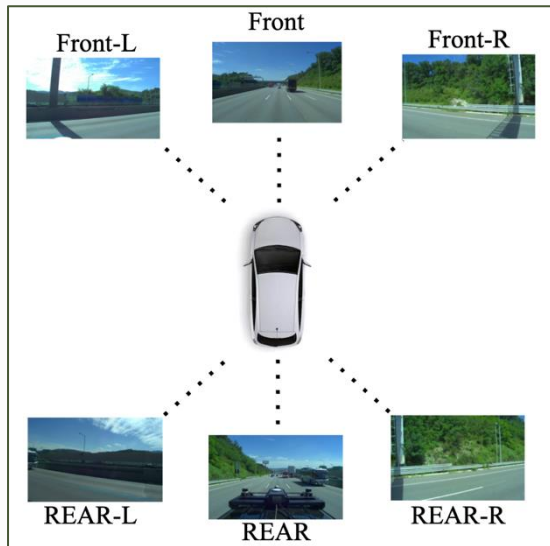
Contents

- 01 과제(Task) 및 제약사항
- 02 Real-time 모델 선정
- 03 전략 : 합성 데이터 및 학습 파이프라인
- 04 실험결과
- 05 결론 및 향후 계획

01. 과제(Task) 및 제약사항

과제 개요

- 카메라 기반 자율주행 환경과 객체의 픽셀별 구분(시맨틱 세그멘테이션)
- 도로, 차선, 노면 표시, 차량, 보행자 등을 화소(Pixel) 단위로 분할



분류	학습 데이터	평가 데이터
총 이미지 수	16,600	800
전방 이미지	3600	-
그 외 이미지	2600 x 5	-
클래스 수	19	19

01. 과제(Task) 및 제약사항

제약사항

- 모델 연산량(FLOPs) 제한

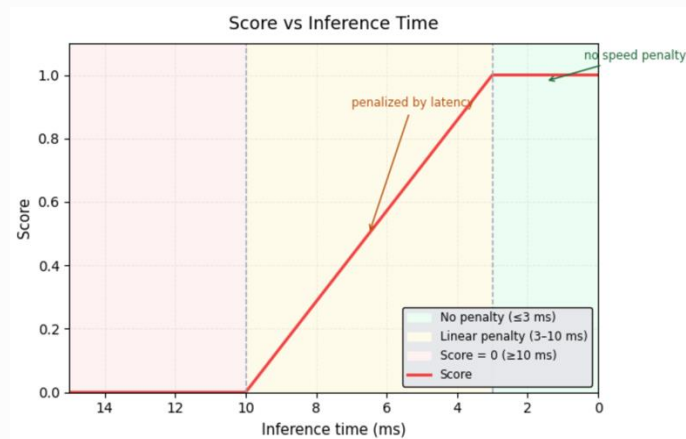
베이스라인 DDRNet-23s (33.85 GFLOPs)의 2배 이내(≤ 67.7 GFLOPs) 초과 금지

- 평가 방식

1차 평가: mIoU 평가, 상위 4팀 본선 진출

2차 평가: 본선 4팀 대상 task-specific score로 최종 순위 결정

$$Score = mIoU \cdot \max\left(0, 1 - \frac{\max(0, T_{infer} - 3ms)}{7ms}\right)$$



01. 과제(Task) 및 제약사항

제약사항

- 모델 연산량(FLOPs) 제한

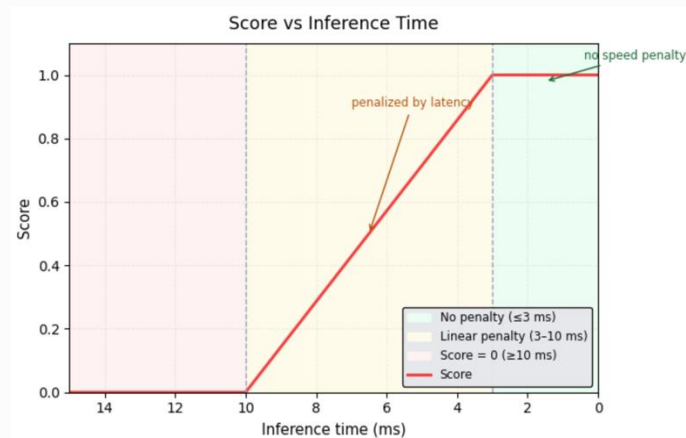
베이스라인 DDRNet-23s (33.85 GFLOPs)의 2배 이내(≤ 67.7 GFLOPs) 초과 금지

- 평가 방식

1차 평가: mIoU 평가, 상위 4팀 본선 진출

2차 평가: 본선 4팀 대상 task-specific score로 최종 순위 결정

$$Score = mIoU \cdot \max\left(0, 1 - \frac{\max(0, T_{infer} - 3ms)}{7ms}\right)$$



02. Real-time 모델 선정

FLOPs 및 FPS 측정

- 해상도 : 1080 * 1920
- 데이터셋 : 대회 제공 데이터셋
- GPU 환경 : RTX 4090
- 비교 모델 : CNN 기반 모델 선택

ViT 기반 모델은 FLOPs 조건을 충족하지만 낮은

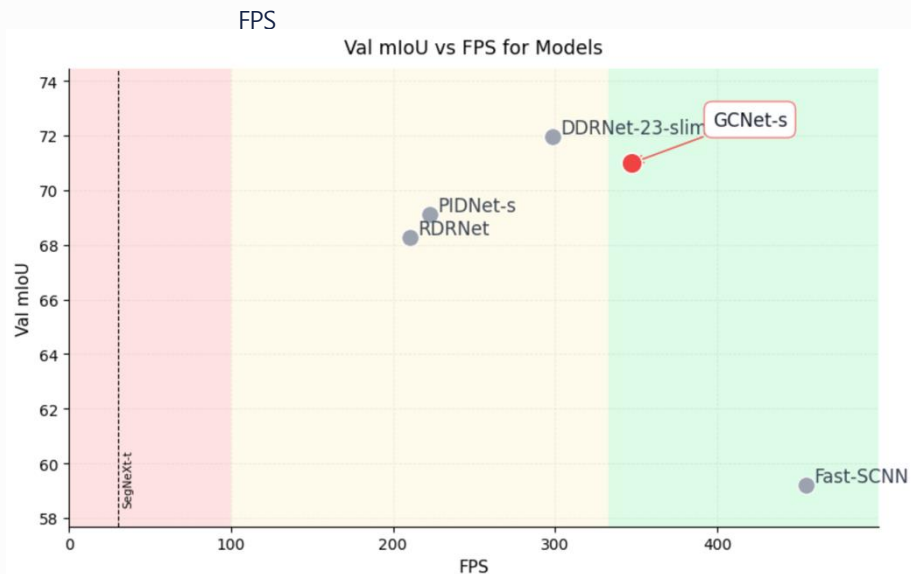
FPS						
Model	Venue	Pretrained	Val mIoU	GFLOPs	FPS	ms
Fast-SCNN	BMVC 2019	Cityscapes	59.19	7.105	454.72	2.19
DDNet-23-slim	T-ITS 2021	-	71.95	36.511	298.54	3.35
SegNeXt	NeurIPS 2022	-	-	47.34	30.68	32.59
PIDNet-S	CVPR 2023	-	69.13	47.517	222.60	4.49
RDRNet	Arxiv 2024	-	68.28	43.4	210.58	4.75
GCNet-S	CVPR 2025	-	70.98	41.76	346.99	2.88

02. Real-time 모델 선정

FLOPs 및 FPS 측정

- 해상도 : 1080 * 1920
- 데이터셋 : 대회 제공 데이터셋
- GPU 환경 : RTX 4090
- 비교 모델 : CNN 기반 모델 선택

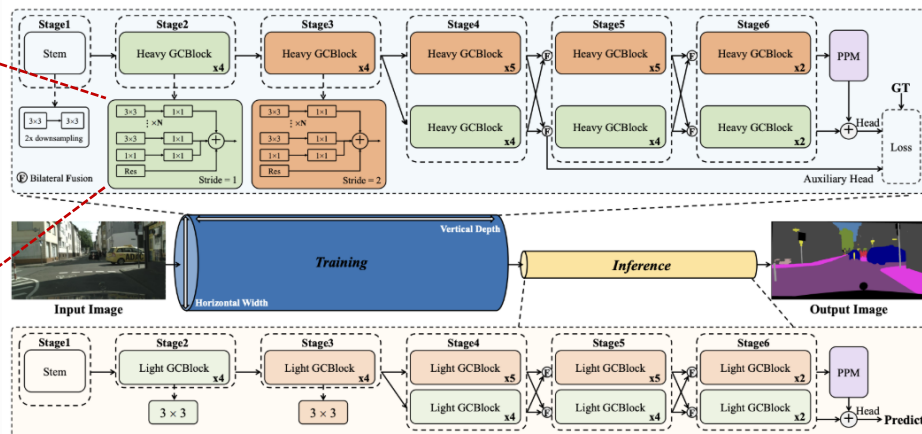
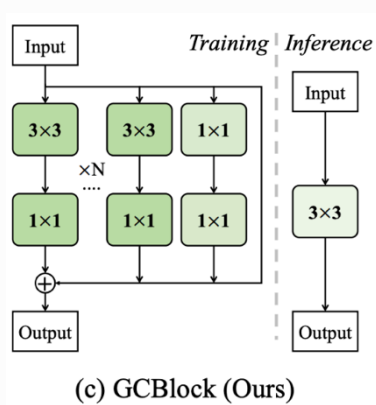
ViT 기반 모델은 FLOPs 조건을 충족하지만 낮은



02. Real-time 모델 선정

GCNet Overview

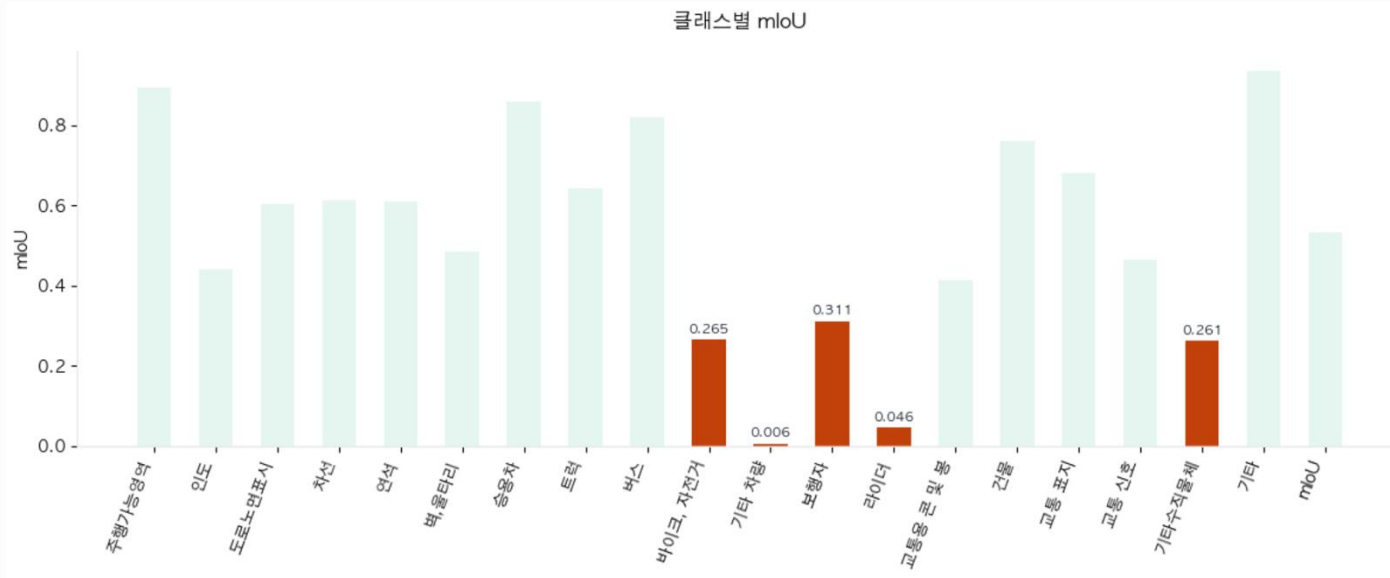
- Yang, Guoyu, et al. "Golden Cudgel Network for Real-Time Semantic Segmentation."
- Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.



03. 주요 전략 – 합성 데이터

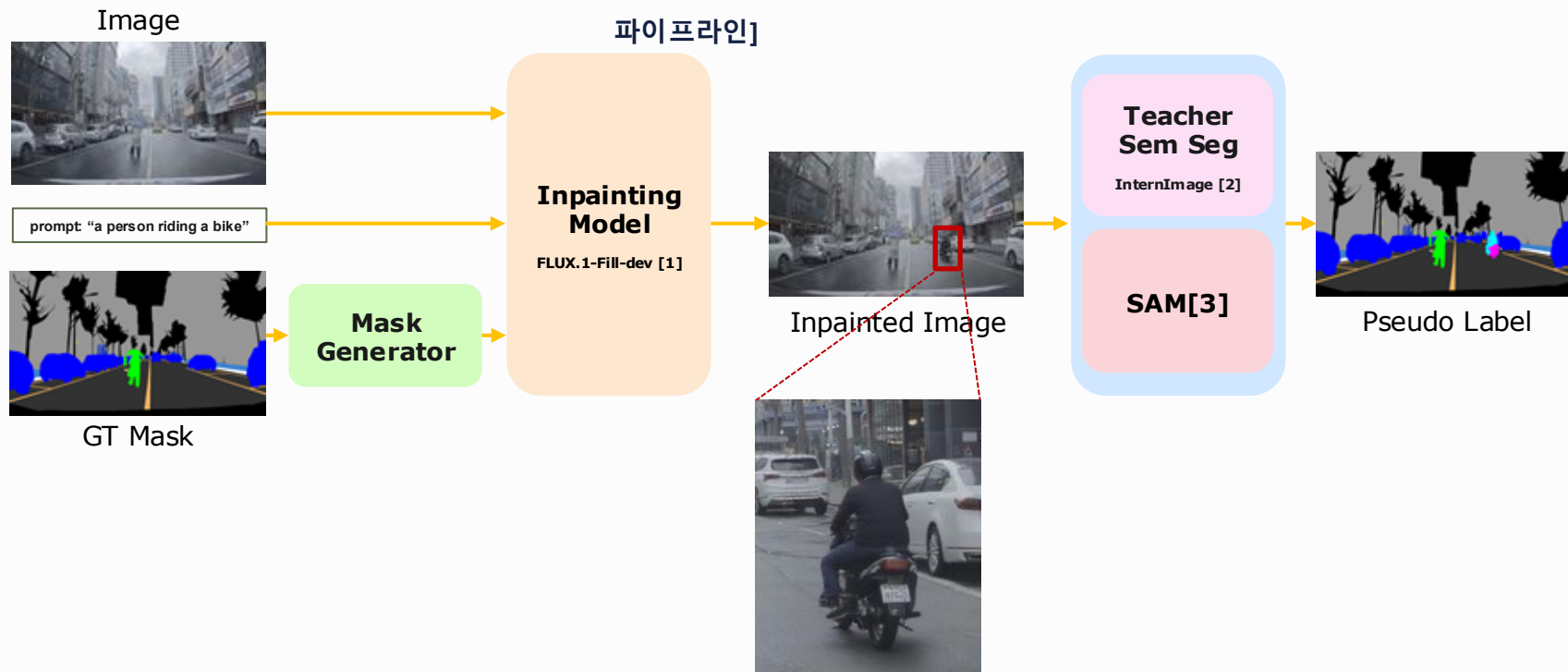
합성 데이터

- FPS 제한으로 신규 모듈 추가 대신 데이터 측면 개선 선택
- 클래스별 mIoU 분석으로 취약 클래스 식별 (모터사이클, 자전거, 보행자, 라이더, 킥보드)
- 취약 클래스 중심 합성 데이터를 추가 생성해 표본 빈도 ↑ · 가시성 ↑



03. 주요 전략 – 합성 데이터

[합성 데이터 생성
파이프라인]



[1] Baffol, Stephen, et al. "FLUX.1 Kortex: Flow Matching for In-Context Image Generation and Editing in Latent Space."

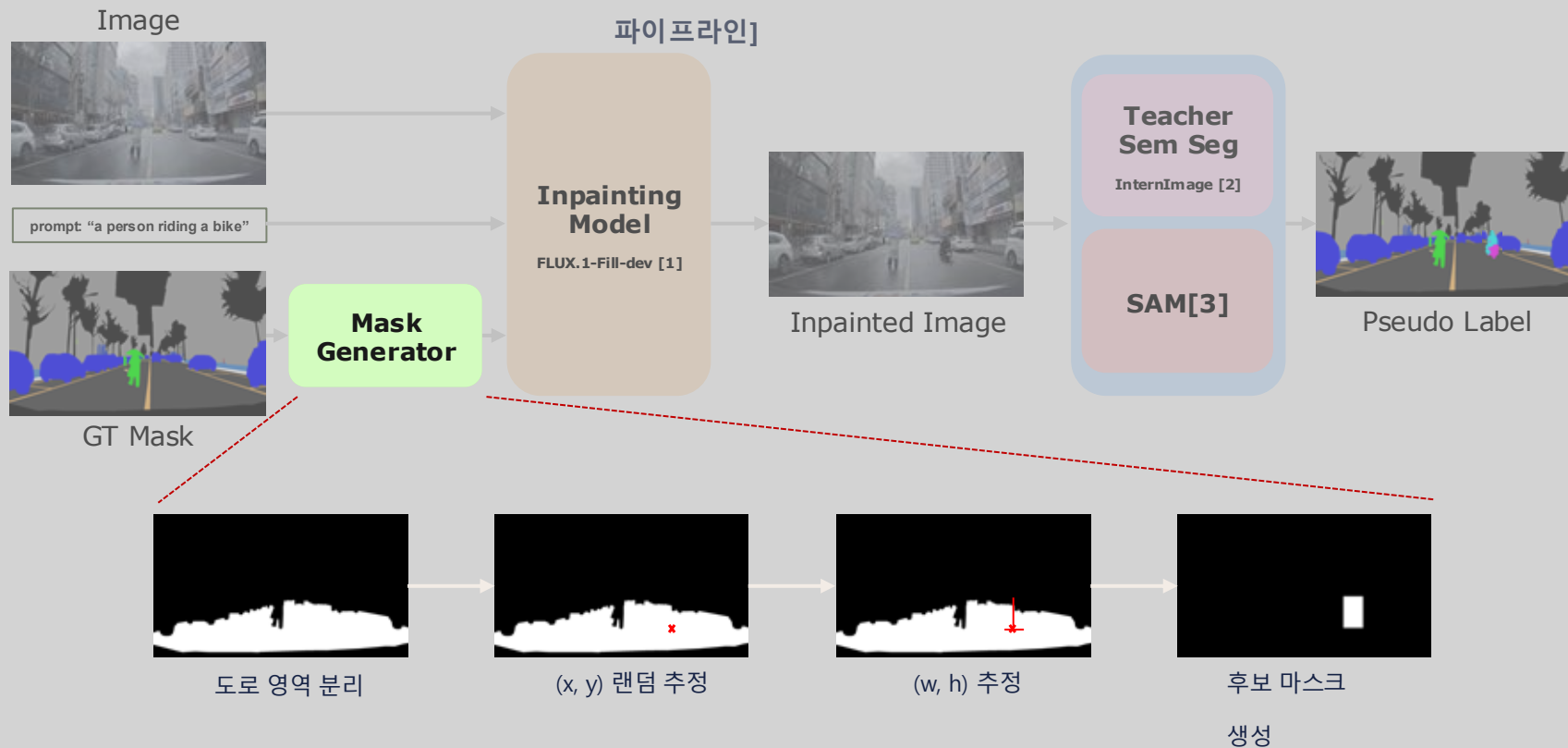
[2] Wang, Wenhai, et al. "InternImage: Exploring large-scale vision foundation models with deformable convolutions."

[3] Kirillov, Alexander, et al. "Segment anything"

03. 주요 전략 – 합성 데이터

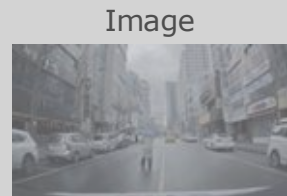
[합성 데이터 생성

파이프라인]

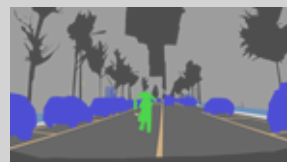


03. 주요 전략 – 합성 데이터

[합성 데이터 생성
파이프라인]



prompt: "a person riding a bike"



GT Mask

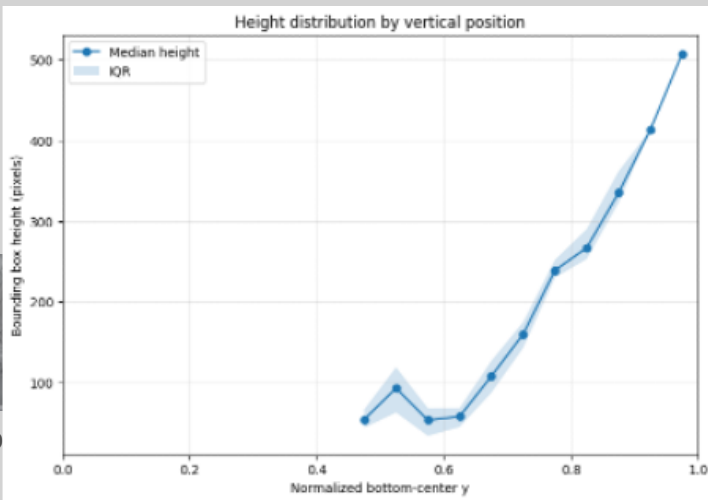
Mask
Generator

Inpainting
Model

FLUX.1-Fill-dev [1]



Input



Label



도로 영역 분리



(x, y) 랜덤 추정



(w, h) 추정



후보 마스크
생성

03. 주요 전략 – 합성 데이터

[예시 이미지]



person



kickboard



motorcycle

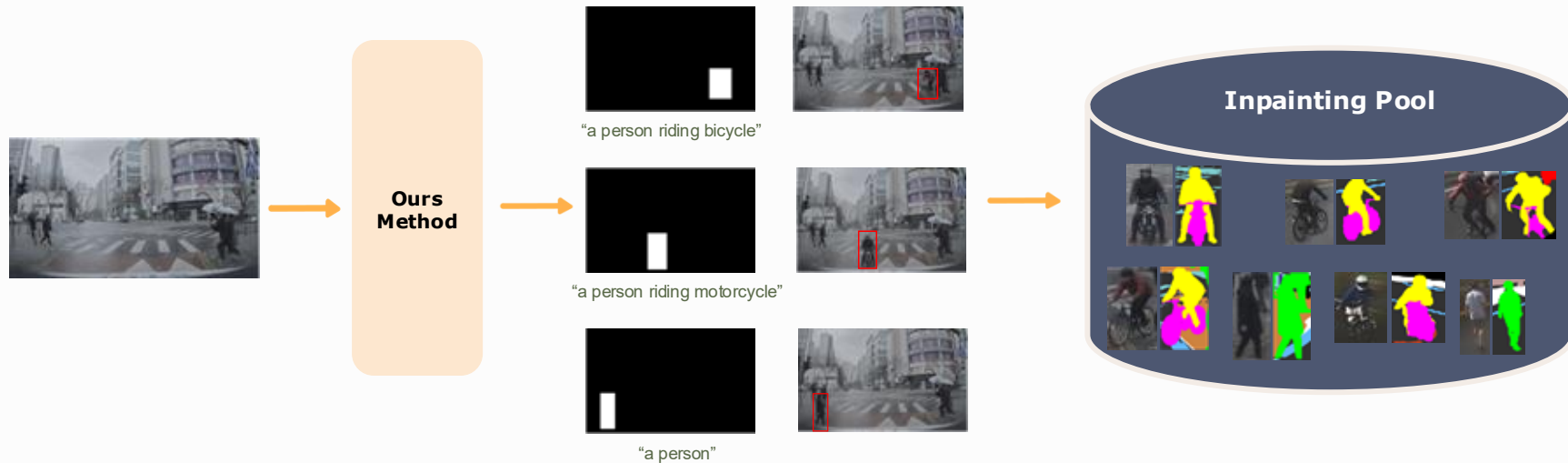


bicycle

03. 주요 전략 – 합성 데이터

Inpainting Pool 구성

- 삽입 후보 영역에 사용할 객체 인스턴스 풀(pool) 사전 구축
- GCNet(Segmentation) 학습 시 Inpainting/Copy&Paste 증강에 활용

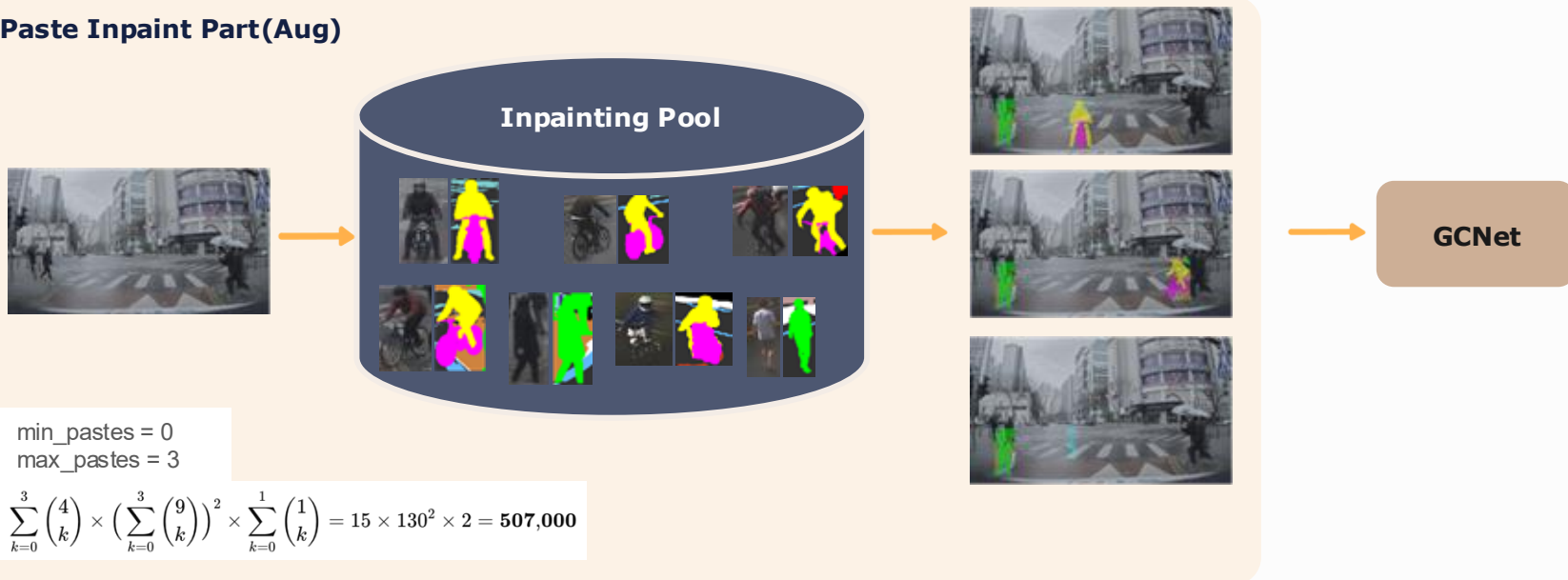


03. 주요 전략 – 합성 데이터

Inpainting Pool 구성

- 삽입 후보 영역에 사용할 객체 인스턴스 풀(pool) 사전 구축
- GCNet(Segmentation) 학습 시 Inpainting/Copy&Paste 증강에 활용

Paste Inpaint Part(Aug)



합성 데이터 생성

FLUX.1-Fill-dev

- **GPU** : 2 x NVIDIA A100
- **Category** : Bicycle(9개), Motorcycle(9개), Person(4개), Kickboard(1개)

InternImage-H(Mask2Former)

- **GPU** : 2 x NVIDIA A100
- **Dataset** : 대회 제공 데이터셋
- **Batch Size** : 2 per GPU (Total 4)
- **Iteration** : 80,000
- **Augmentation** : Random Resize, -Crop, -Flip, ...

세그멘테이션 모델 학습

GCNet-S

- **GPU** : 4 x RTX 4090
- **Dataset** : 대회 제공 데이터셋
- **Validation Set** : Random(7 : 3)
- **Batch Size** : 3 per GPU (Total 12)
- **Iteration** : 320,000
- **Augmentation** : Random Resize, -Crop, -Flip, ...

04. 실험 결과

실제 제출 결과

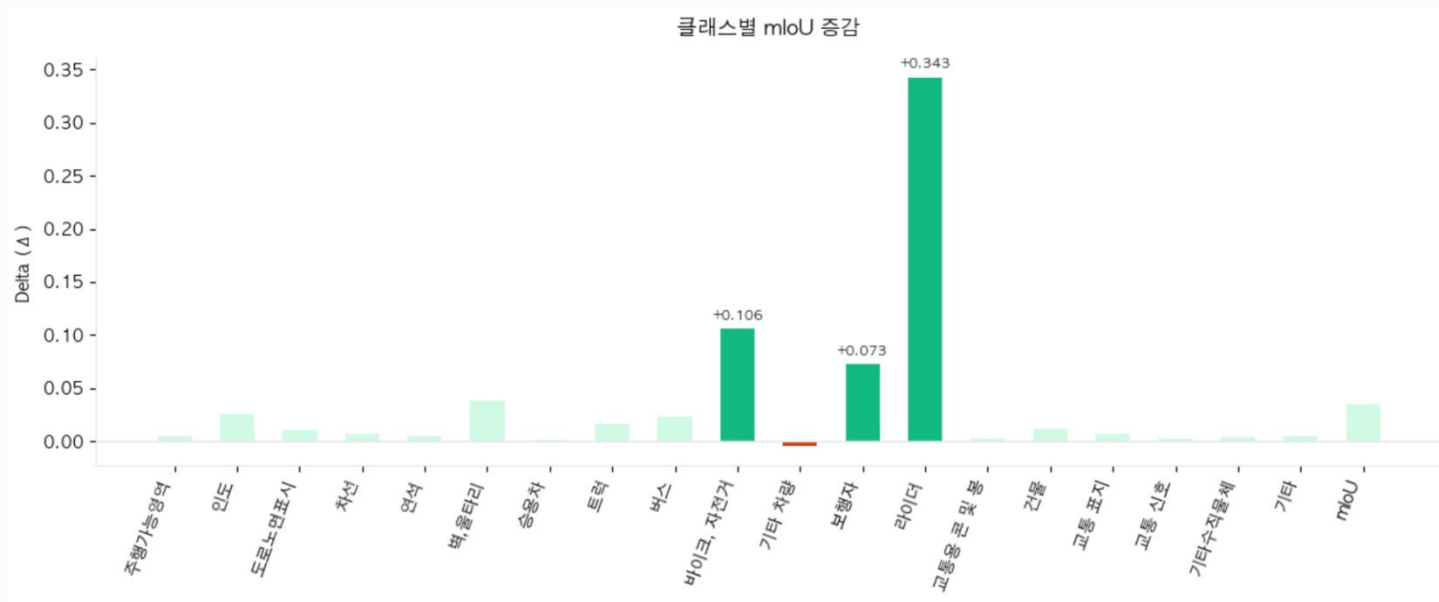
- Test mIoU (리더보드 점수): 57.11%
- 모델 : GCNet-S (*Cityscapes* 사전학습)

Model	Pretrained	Method	Test mIoU	Δ mIoU	Iters	Notes
GCNet-S	Cityscapes	-	53.1622	-	120,000	
		+ synth	56.7387	+ 3.765	120,000	
		+ more train	57.1111	+ 0.3724	318,000	Final Submission
			57.0441		320,000	

04. 실험 결과

Inpainting

- 바이크·자전거, 보행자, 라이더에 대한 인페인팅 효과



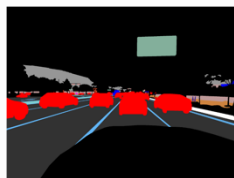
05. 결론 – 한계점 및 향후 개선 방향

ControlNet을 사용한 데이터 합성

- 라벨 맵을 조건으로 전체 이미지를 직접 생성 (단순 인페인팅/붙여넣기 아님)
- 분포 제어 용이: 날씨, 시간대, 조도, 배경 밀도 등 다양한 변형 적용 가능



RGB



GT Mask



Night



Foggy

Removing Inpainting

- 복잡한 동적 장면에서 LaMa(WACV 2022) 기반 제거 인페인팅이 배경을 과도 평탄화 → 학습 품질 저하
- 마스크 크기, 시점 변화에 일관성이 낮음 → 실제 분포와 괴리 확대, 도메인 시프트가 증가 → 검증 성능 악화 → 채택하지 않음



[성공 사례]



[실패 사례]

