

최근 정보통신기술 One-Step Image-to-Image Translation 의 원격주행용 Diffusion Model 기반 Semantic Compression 적용에 관한 연구

윤성빈, 김예준, 강민우, 서효운
성균관대학교

emsekr@g.skku.edu, kyj8130@g.skku.edu, kmw8123@g.skku.edu, hyowoonseo@skku.edu

A Study on the Application of Diffusion Model-Based Semantic Compression For One-Step Image-to-Image Translation in Remote Driving Systems

Yun Sung Bin, Kim Ye Jun, Kang Min Woo, Seo Hyowoon
Sungkyunkwan Univ.

요 약

본 논문은 의미 압축 프레임워크인 GESCO 모델이 야간 주행 환경에서 겪는 성능 저하 문제를 해결하고자 도메인 변환 모듈을 결합한 복원 체계를 제안한다. 제안 방식은 야간 이미지를 주간으로 변환하여 의미 압축을 수행하여 mIoU와 LPIPS 지표를 각각 평균 707.01%와 12.88% 향상시켰다. 이를 통해 저조도 환경에서도 의미 보존과 시각 품질을 개선하고 통신 시스템의 견고성을 확보하였다.

I. 서 론

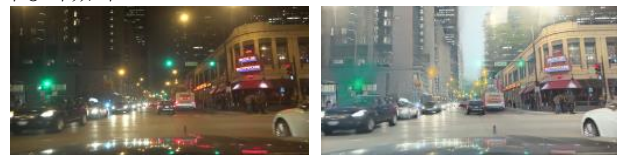
의미 기반 통신은 원본 데이터의 모든 세부 정보를 그대로 전송하는 대신, 수신 단말의 의사결정에 필요한 의미적 요소를 추출 및 압축하여 전송함으로써 대역폭 사용량을 획기적으로 절감할 수 있다.[1] Generative Semantic Communication(GESCO) 모델은 확산 모델 기반 의미 압축 프레임워크이다. 송신단이 One-Hot Segmentation Map 을 압축·전송하고 수신단이 FDS 블록과 SPADE 기반 조건부 확산 모델을 통해 의미를 보존하며 이미지를 재생성하는 방식으로, 낮은 SNR 환경에서도 높은 의미 복원 성능을 유지한다.[2] 그러나 해당 모델은 주간 주행 환경에서 안정적인 성능을 보였으나, 야간 주행과 같이 조명 조건이 극도로 변화하는 상황에서는 복원 성능이 급격히 저하되는 한계가 존재한다.

이러한 한계를 극복하기 위해, 본 연구는 GESCO 모델에 One-Step Image-to-Image Translation 기법을 결합하여 야간 주행 데이터를 주간 데이터로 변환한 후, 이를 GESCO 송신단에 입력해 의미 압축·복원을 수행하는 방식을 제안한다. 이를 위해 CycleGAN-Turbo 구조를 기반으로 한 One-Step 변환 프레임워크를 채택하였다. 이 접근법은 다단계 확산 기반 이미지 변환 모델보다 연산 복잡도가 낮으면서도, 원본 영상의 장면 구조를 유지한 채 조명·색상 특성을 효과적으로 변환할 수 있는 장점이 있다.[3]

결과적으로 제안 시스템은 원격 자율주행 환경에서 다양한 조명 조건에도 안정적인 의미 복원 성능을 보장하며, 특히 야간 주행 상황에서도 주간 대비 유사한 mIoU, LPIPS 성능 확보를 목표로 한다. [4,5]

II. 본론

본 논문에서는 67 epoch 를 학습한 GESCO 모델에 야간 주행 데이터로 의미 압축·복원을 수행했을 때와 동일한 모델에 야간 주행 데이터를 One-Step Image-to-Image 로 주간 도메인 주행 데이터처럼 변환 후 의미 압축·복원했을 때의 결과를 비교한다. 각 데이터에 대한 결과 품질 평가는 생성 이미지에서 주요 객체의 보존 정도를 평가하는 mIoU, 생성 이미지와 원본 이미지 간의 시각적 유사도를 정량화하기 위한 지표인 LPIPS 를 사용하였다.



<그림 1. One-Step Image-to-Image Translation 을 이용해 야간 주행 데이터(좌)를 주간 주행 데이터(우)로 변환한 모습>



<그림 2. 원본 야간 데이터셋 GESCO 출력 결과(좌)와 주간으로 변환한 데이터셋 출력 결과(우)>

<그림 1.>처럼 야간 주행 데이터를 주간 주행 데이터처럼 변환해 색, 명암 대비 측면에서 보다 개선된 데이터를 <그림 2.>와 같이 GESCO 모델의 입력으로 하여 의미 압축·복원 시 차·도로, 건물 등 의미상 주요 객체들의 보존 정도가 향상되었다.

SNR[dB]	mIoU ↑ [%]		mIoU 향상 비율[%]
	야간 이미지	주간 변환 이미지	
1	0.12	0.32	166.71
5	0.13	0.25	92.31
10	0.11	1.67	1418.18
15	0.11	2.28	1972.72
20	0.44	2.36	436.36
30	0.44	2.71	515.91
100	0.64	2.86	346.88

<표 1. 동일 학습 GESCO 모델에서 출력한 야간 이미지와 주간 변환 이미지의 mIoU 평가 결과 비교>

생성 이미지 16 개에 대한 SNR 별 mIoU 값을 비교한 결과, <표 1.>에서 알 수 있듯 야간 주행 데이터의 주간 주행 데이터 형식으로 변환 후 평균 707.01%로 매우 높은 수준의 mIoU 향상을 보였다.

SNR[dB]	LPIPS ↓ 평균값 [%]		LPIPS 향상 비율[%]
	야간 이미지	주간 변환 이미지	
1	0.7892	0.7164	9.22
5	0.7863	0.6878	12.53
10	0.7829	0.6457	17.52
15	0.7463	0.6246	16.31
20	0.6979	0.6097	12.60
30	0.6687	0.6008	10.15
100	0.6758	0.5959	11.82

<표 2. 동일 학습 GESCO 모델에서 출력한 야간 이미지와 주간 변환 이미지의 LPIPS 평가 결과 비교>

생성 이미지 16 개에 대한 SNR 별 LPIPS 평균값을 비교한 결과, <표 2.>에서 알 수 있듯 야간 주행 데이터의 주간 주행 데이터 형식으로 변환 후 평균 12.88%의 준수한 LPIPS 향상 수준을 보였다.

mIoU 및 LPIPS 모두 향상된 것을 보이며, 이는 도메인을 변환함으로써 조명, 색상 개선을 통해 객체 의미 보존 및 시각적 유사도를 향상시킨 결과이다.

III. 결론

본 논문에서는 확산 모델 기반 의미 압축 모델에 CycleGAN-Turbo 기반 도메인 변환 모듈을 결합하여, 원격 자율주행 환경에서 야간 주행 데이터 전송 시 성능 저하 문제의 해결 방향을 제안하였다.

제안 방식은 모든 SNR 구간에서 mIoU, LPIPS 가 유의미하게 향상되었다. mIoU 는 평균 707.01%, LPIPS 는 평균 12.88%의 향상을 보여 야간 도메인에서의 성능 개선을 입증하였다.

그러나 본 연구는 몇 가지 한계점이 존재한다. 첫째, 절대적인 mIoU 값이 낮다는 점이다. GESCO 선행 논문에서 보고된 90% 수준의 mIoU 에 비해 본 연구의 절대 수치는 최대 2~3% 수준에 불과하다. 이는 사용한 Segmentation 모델, 야간 데이터셋 품질, 변환 후 색채/조명 재현과 학습 환경의 한계 등에 기인할 수 있다.

둘째, 도메인 변환 모듈의 일반화 관점에서, 즉 CycleGAN-Turbo 가 기존 학습 데이터 분포와 크게 다른 야간 환경(여름과 겨울의 차이 등)에서는 변환 품질이 저하될 수 있다는 한계점도 존재한다. 셋째, 실시간 처리 성능 부족 또한 한계가 될 수 있다. 변환과 복원 모두 GPU 연산을 요구하며, 고해상도(2048×1024) 환경에서 실시간성 확보가 어렵다는 점은 본 모델의 실시간 원격 주행 시스템에 즉각적인 도입을 불가하게 만든다.

향후 연구에서는 다양한 야간 환경 조건을 포함한 학습 데이터 확충, 도메인 변환과 의미 복원기의 공동 최적화를 통한 품질 개선, 그리고 경량화 및 최적화 기반의 실시간 처리 가능성 검토를 진행할 예정이다. 이를 통해 제안 방법은 원격 자율주행 뿐만 아니라 저조도 영상 통신, 스마트 시티 감시 등 다양한 차세대 통신·비전 융합 응용에 폭넓게 적용 가능할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로

한국연구재단의 지원을 받아 수행된 연구임(RS-2025-

00555907).

참 고 문 헌

- [1] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," IEEE Wireless Communications, vol. 29, no. 1, pp. 210–219, Feb. 2022.
- [2] E. Grassucci, S. Barbarossa, and D. Comminiello, "Generative semantic communication: Diffusion models beyond bit recovery," arXiv:2306.04321, Jun. 2023. (<https://arxiv.org/abs/2306.04321>)
- [3] G. Parmar, T. Park, S. Narasimhan, and J.-Y. Zhu, "One-step image translation with text-to-image models," arXiv:2403.12036, Mar. 2024. (<https://arxiv.org/abs/2403.12036>)
- [4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes dataset for semantic urban scene understanding," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3213–3223, 2016.
- [5] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 586–595, 2018.