

# 객체 중심 이미지 전송을 위한 시맨틱 통신

김시현, 이상훈, 전요셉  
포항공과대학교

{n00ughtism, shlee423, yoseb.jeon}@postech.ac.kr

## Semantic Communication for Object-Centric Image Transmission

Si-Hyun Kim, Sang-Hun Lee, Yo-Seb Jeon  
Pohang Univ. of Science and Technology (POSTECH)

### 요약

본 연구는 이미지 내 관심 객체를 효과적으로 전송하는 시맨틱 통신을 실현하기 위해 객체 중심 이미지 전송 기법을 제안한다. 제안된 기법은 이미지와 함께 입력된 텍스트를 기준으로 CLIP 기반 attention map 과 class activation map 을 활용해 이미지 내 관련 객체의 관심 영역을 자동으로 추출한 뒤, 관심 영역과 비-관심 영역에 차등적인 압축 및 변조 방식을 적용함으로써 시맨틱 정보 전달에 효과적인 전송 방식을 제공한다. 특별히, 관심 영역 전송 시 attention 점수에 따른 적응적 양자화 및 변조를 적용하여 주어진 무선 자원을 효과적으로 분배한다. 모의 실험을 통해 동일한 이미지라도 입력 텍스트에 따라 전송되는 관심 영역이 달라짐을 보이고, 이를 통해 원하는 객체 중심의 시맨틱 정보를 효과적으로 전송할 수 있음을 입증한다.

### I. 서론

기존 이미지 전송 방식은 데이터의 완전한 복원에 중점을 두었으나, 최근에는 의미 기반 핵심 정보만을 선택적으로 전송해 효율을 높이는 시맨틱 통신(Semantic Communication)이 주목받고 있다. 이는 특히 고차원 비정형 데이터인 이미지에서, 전체 전송보다 인간 인지에 중요한 객체 중심 정보만을 전달함으로써 통신 자원을 절감하고 의미 전달을 극대화할 수 있다. 그러나 기존 시맨틱 통신 연구는 이미지 전체를 압축하거나 중단 간 학습 기반 구조에 의존해 실제 디지털 통신 시스템과의 연동 및 유연성에서 한계가 있다. 아직까지 이미지 내 객체 중 관심을 두고 있는 정보 전송에 집중하는 선택적 전송 방식은 고려된 바 없다.

이에 본 연구에서는, 이미지와 함께 입력된 텍스트를 기준으로 관심 객체가 포함된 관심 영역(Region of Interest, ROI)을 자동으로 추출하고, CLIP 기반 attention map 과 class activation map (CAM)을 활용하여 ROI 와 비-관심 영역 (Region of Non-Interest, RONI)를 효과적으로 추출한다. 이후 각 영역별로 상이한 압축 및 변조 전략을 적용하여, 관심 객체 중심으로 정보량을 최적화하면서도 의미 보존이 가능한 이미지 전송 기법을 제시한다. 특히, 객체 단위로 중요도를 반영한 전송 구조를 구현함으로써, 향후 다양한 시맨틱 기반 통신 시나리오에 실용적으로 적용할 수 있는 가능성을 제시한다.

### II. 본론

본 연구에서는 이미지 내 관심 객체를 효과적으로 전송하는 시맨틱 통신을 실현하기 위해 객체 중심 이미지 전송 기법을 제안한다. 먼저, 이미지에 포함된 객체 중 하나를 텍스트 형태로 입력하면, 이를 기준으로 이미지 내 시맨틱 중요도를 분석한다. 구체적으로, CLIP 기반 attention map 과 class activation map (CAM)을 생성하여 해당 객체의 위치를 강조하고, attention 점수 0.5 를

기준으로 ROI 와 RONI 를 분리한다. 이렇게 분리된 ROI 와 RONI 영역에는 각기 다른 양자화 및 변조 방식이 적용된다. ROI 영역에는 attention 점수에 따른 비트를 할당하는 적응적 양자화 방식을 적용한다. 이미지의  $i$  번째 패치의 attention 점수를  $A_i$  라고 한다면, 이때 할당되는 비트 수  $b_i$ 는 다음과 같이 결정된다.

$$b_i = \left\lfloor b_{\min} + (b_{\max} - b_{\min}) \cdot \frac{A_i - 0.5}{0.5} + 0.5 \right\rfloor.$$

여기서,  $b_{\min} = 4$ 는 최소 할당 비트 수,  $b_{\max} = 8$ 는 최대 할당 비트 수를 나타낸다. 또한, RONI 에 해당하는 패치들에는 균일하게 3 비트 양자화를 적용한다. 이러한 방식으로 시맨틱 중요도 및 관심 영역에 따라 전송 자원을 차등 할당함으로써, 통신 효율과 의미 보존 사이의 균형을 도모한다.

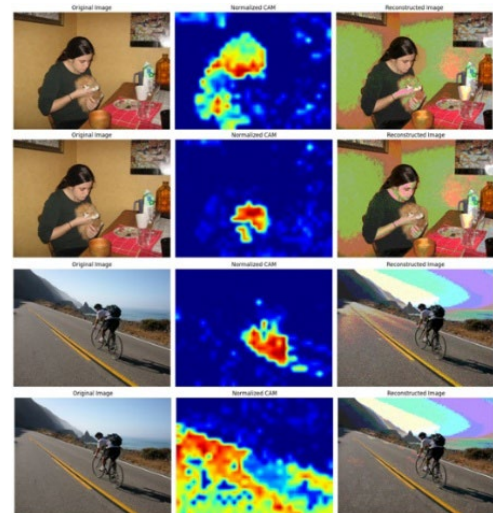


그림 1. 객체 별 텍스트 입력에 따른 ROI 중심 이미지 복원 결과 (위: 여자, 고양이 / 아래: 자전거, 도로)

그림 1 은 앞서 언급한 양자화 방식을 적용하여 SNR=20dB 환경에서 ROI 는 QPSK, RONI 는 64-QAM 변조 방식을 통해 복원한 이미지 결과이다. 같은 이미지에 다른 텍스트를 입력했을 때 해당 ROI 영역 중심으로 이미지가 복원되는 것을 확인할 수 있다.

다음으로, ROI 와 RONI 에 대해 최적의 변조 방식을 결정하기 위해 다양한 변조 기법에 대한 성능 평가를 수행한다.

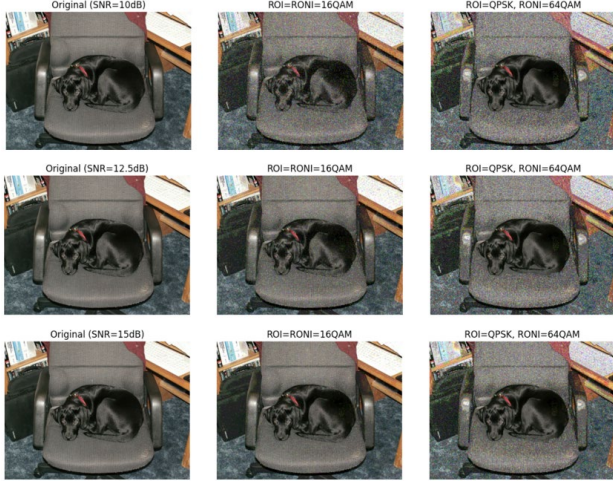


그림 2. 동일 이미지에 두 가지 변조 방식을 적용한 전송 및 복원 결과

그림 2 는 동일한 입력 이미지에 대해 두 가지 변조 설정 (ROI, RONI)=(16QAM, 16QAM), (ROI, RONI)=(QPSK, 64QAM) 을 적용해 SNR 10.0, 12.5, 15.0 dB 조건에서 복원된 결과를 시각적으로 비교한 것이다. 전송은 AWGN 채널을 통해 이루어졌으며, 좌측은 원본, 중앙은 전 영역 16QAM, 우측은 ROI 에 QPSK, RONI 에 64QAM 을 적용한 결과이다. SNR 이 낮을수록 ROI 에 낮은 차수의 변조 방식을 적용한 쪽이 안정적이고, SNR 이 15.0 dB 이상일 경우 고차 변조 방식이 오히려 더 나은 품질을 보였다. 이는 평균적으로 같은 수의 심볼을 전송하더라도 ROI-RONI 를 고려한 선택적 변조 전략이 효과적임을 나타낸다.

본 논문에서는 기존의 PSNR 지표를 확장하여 ROI 와 RONI 의 상대적 중요도를 반영할 수 있는  $\Theta$ PSNR(Theta-weighted Peak Signal-to-Noise Ratio)을 평가 지표로 활용하였다.  $\Theta$ PSNR 은 일반적인 PSNR 과 마찬가지로 MSE(Mean Squared Error)에 기반하지만, ROI 와 RONI 에 다른 가중치를 부여함으로써 ROI 중심의 왜곡을 보다 정밀하게 반영할 수 있다.

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right).$$

$$\Theta PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE_{ROI} \cdot \theta + MSE_{RONI} \cdot (1 - \theta)} \right).$$

여기서 MAX 는 이미지 픽셀의 최대값인 255,  $MSE_{ROI}$  는 ROI 영역의 평균 제곱 오차,  $MSE_{RONI}$  는 RONI 영역의 평균 제곱 오차,  $\theta \in [0, 1]$  는 ROI 영역의 상대적 중요도를 나타내는 가중치 파라미터이다.  $\theta$  가 클수록 ROI 복원 오류가 전체 지표에 더 크게 반영되며,  $\theta=1$  일 경우 PSNR 은 ROI 만을 기준으로 계산된다. 따라서  $\Theta$ PSNR 은 시맨틱

통신에서 정보 중요도 기반 왜곡 평가에 적합하며, ROI 중심 복원 품질을 정량적으로 비교하는 데 유용하다.

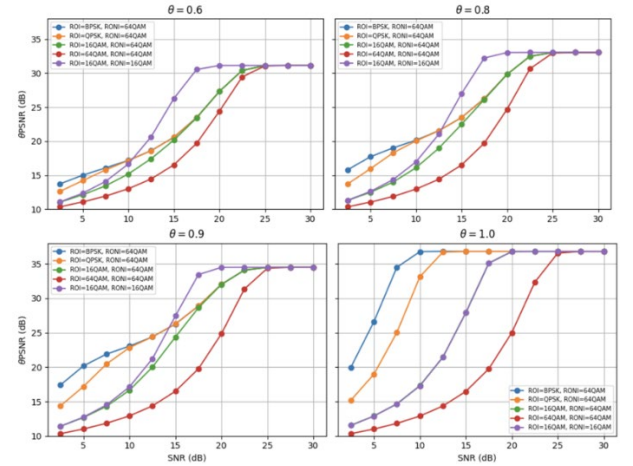


그림 3. 각  $\theta$  에 대해 SNR 에 따른  $\Theta$ PSNR 시각화

그림 3 은  $\theta \in \{0.6, 0.8, 0.9, 1.0\}$  에 따른  $\Theta$ PSNR 성능을 비교한 그래프로, 다양한 ROI/RONI 변조 조합을 SNR 조건별로 시각화한 결과이다. AWGN 채널을 통해 VOC2012 이미지 40 장을 각 SNR 당 5 회 반복 전송한  $\Theta$ PSNR 평균값을 기준으로 하였다.  $\theta$  가 클수록 ROI 를 강하게 보호한 조합(BPSK/QPSK for ROI, 64QAM for RONI)이 우수한 성능을 보이며, 특히  $\theta = 1.0$  에서는 전 구간에서 가장 높은  $\Theta$ PSNR 을 기록해 ROI 중심 지표에서의 적응적 변조 전략 효과를 입증한다. 반면 SNR 이 12.5~15 dB 이상에서는 ROI 와 RONI 모두에 16QAM 을 적용한 방식이 더 높은 성능을 보이며, 이는 채널 환경이 개선되면 일관된 고차 변조가 더 효율적일 수 있음을 시사한다. 즉, ROI 보호 전략은 낮은 SNR 에서 효과적이며, 채널 조건과 응용 목적에 따라 적절한 변조 조합이 필요함을 보여준다.

### III. 결론

본 연구는 텍스트 기반 객체 정보를 활용해 이미지 내 관심 영역(ROI)과 비-관심 영역(RONI)을 구분하고, 각 영역에 상이한 압축 및 변조 방식을 적용하는 객체 중심 시맨틱 이미지 전송 기법을 제안하였다. 실험을 통해, 입력 텍스트에 따라 의미적으로 중요한 객체가 달라지고, 이에 따라 서로 다른 ROI 가 추출되어 해당 영역 중심으로 정보가 선택적으로 전송됨을 확인하였다. 또한, ROI 에 강건한 변조 방식을 적용하면 저잡음(SNR) 환경에서 시각적 품질이 안정적으로 유지되며,  $\Theta$ PSNR 지표를 통해 ROI 중심 복원 성능을 정량적으로 평가할 수 있음을 입증하였다. 본 연구는 시맨틱 중요도와 채널 환경에 따라 자원을 유연하게 분배하는 전송 전략의 가능성을 제시하며, 향후 시맨틱 기반 통신 시스템의 실용화에 기여할 수 있을 것으로 기대된다.

### ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2024-00453301).

### 참고 문헌

- [1] J. Wu et al. "Semantic segmentation-based semantic communication system for image transmission." Digital Communications and Networks, vol. 10, no. 3, pp. 519-527, 2024.