

머신 비전 태스크에서 비트레이트-정확도 최적화를 위한 이미지 전처리

Image Preprocessing for Bitrate-Accuracy Optimization in Machine Vision Tasks

서영홍¹, 김예지¹, 이채은², 이한호¹

인하대학교 정보통신공학과¹, 한양대학교 융합전자공학부²

syh@inha.edu, wl3952@gmail.com, crhee@hanyang.ac.kr, hhlee@inha.ac.kr

최근 객체탐지와 같은 Machine Vision Task를 위해 미분 불가능한 연산이기 때문에 Bitrate-Accuracy 스마트폰과 같은 에지 디바이스에서 촬영한 이미지를 원격 Optimization을 위한 end-to-end training(중단간 서버로 이미지를 전송하는 것이 급격하게 증가하고 있다 학습)이 불가능하기 때문이다. Encoder에서 Discrete [1]. 보통 에지 디바이스에서 서버로 이미지를 그대로 Cosine Transform(DCT)을 수행하고, 계수들을 보내지 않고, Joint Photographic Experts Group(JPEG) Quality Factor(QF)에 맞게 Quantization을 진행하여 표준과 같은 코덱을 이용하여 이미지를 압축 후 서버로 8x8 block으로 변환 후 bpp estimator에 입력으로 전송함으로써 전송 데이터의 크기를 줄인다. 본 논문은 넣어 출력으로 JPEG 압축 시 생성될 bitstream의 이미지 압축 및 전송 이전에 Tone Mapping과 Gaussian bpp를 예측한다. bpp estimator는 ImageNet Blur를 통한 Image Preprocessing 기법을 통해 JPEG 압축 Validation Set에서 Random Quality로 압축된 JPEG 효율성을 증가시켜, 데이터 전송 효율(Bitrate)과 Machine 이미지 5000장에 대해 테스트를 진행하였고, Vision task의 정확도(Accuracy) Optimization 방법 설계와 Symmetric Mean Absolute Percentage Error(SAMPE)는 실험결과를 제시한다. 5.3% ± 1.60를 달성하였다[3]. Machine Vision Task는

그림 1은 본 논문에서 제안하는 아키텍처이다. Tone & Blur 객체탐지를 수행하는 YOLOv5를 사용하였다. 해당 Network는 ResNet18을 Fine-tuning 하였으며, 이미지를 Network의 Total Loss는 (3)의 수식과 같다.

입력으로 받아 이미지 마다 개별적인 α, β (순서대로 Tone Mapping Factor, Blur Factor)를 출력하여 해당 값을 통해 λ 는 Distortion 가중치인 hyperparameter이다. λ 가 커질수록 Bitrate를 줄이기 보다는 Machine Vision Task의 Accuracy 성능을 더 높인다. 이미지를 전처리 한다. Factor 범위는 [0.5, 1.0]로 동일하다. Tone mapping 기법으로는 기존 Reinhard Tone Mapping에서 밝기 압축 강도를 조절하였다. 이미지의 R,G,B 각 채널로부터 L (Luminance)값을 구하고 (1)의 수식을 통해 밝기를 압축한다[2].

$$L_{mapped} = \frac{L}{1+\frac{L}{\alpha}} (\alpha : \text{Tone Mapping Factor}) \quad (1)$$

α 값이 작아질수록 밝기는 더 크게 압축되며, 이때 이미지의 밝기 분포가 균일하게 되어 JPEG 압축 효율성이 더 커지게 된다. Blur 기법으로는 Gaussian Blur를 사용하였으며, k (kernel size)와 σ (sigma)를 parameter로 한다. sigma를 β 값으로 사용하고 kernel size 결정은 (2)의 수식과 같다.

$$k = \begin{cases} 3 & \text{if } \beta \geq 0.8 \\ 5 & \text{if } \beta < 0.8 \end{cases} (\beta : \text{Blur Factor}) \quad (2)$$

kernel size와 sigma가 커지면 이미지의 고주파 성분이 더 크게 제거되며, JPEG 압축 효율성이 더 커지게 된다.

Machine Vision Task를 수행하기전 JPEG 압축 결과 bit-per-pixel(bpp)를 측정하기 위해 bpp estimator를 사용했다. bpp estimator를 사용하는 이유는 JPEG을 수행하는 과정이

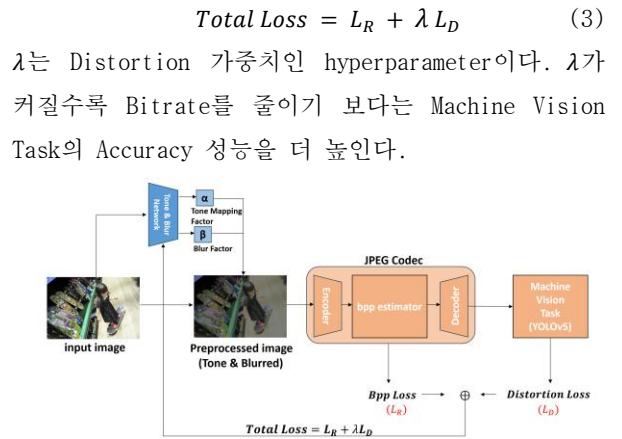


그림 1. Proposed Rate-Accuracy Optimization architecture.

그림 2는 원본 이미지와 제안한 model을 통해 전처리된 이미지를 YOLOv5에 Inference한 결과 사진이다. Object Detection 결과를 비교해보면, 원본 이미지보다 Bitrate를 22.7% 줄였음에도, Machine Vision Task 성능 측면에서 고양이 객체를 모두 탐지하는 것을 볼 수 있다.



그림 2. YOLOv5 Inference 결과(QF = 90)

그림 3은 $\lambda = 8$ 로 고정하고 JPEG Quality Factor를 왼쪽 점부터 60, 70, 80, 90으로 했을 때, Rate Distortion Curve(RD Curve)이다. 실험 결과 BD-Rate(동일 Accuracy 대비 비트레이트 절감량)이 -21.6%로 확인되었다.

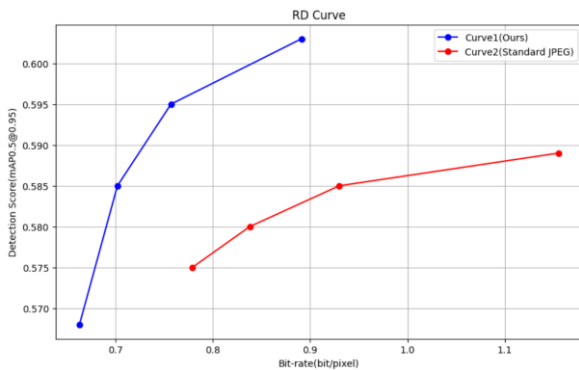


그림 3. RD Curve (Ours vs Standard JPEG)

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 지원을 받아 수행된 연구임 (IITP-2021-0-02052). 본 연구는 IDEC에서 EDA 툴을 지원받아 수행하였음.

참고문헌

- [1] HARELL, Alon; DE ANDRADE, Anderson; BAJIĆ, Ivan V. Rate-distortion in image coding for machines. In: *2022 Picture Coding Symposium (PCS)*. IEEE, 2022. p. 199-203.
- [2] REINHARD, Erik, et al. Photographic tone reproduction for digital images. In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 2023. p. 661-670.
- [3] 정영철, 이채은. "비디오를 위한 딥러닝 네트워크에서 움직임 특성을 활용한 효율적인 Feature Map 압축," 인하대학교 대학원 석사학위논문, 2024년 2월.