

기계 시각을 위한 다중 스케일 특징맵 압축 모델

김영웅 / 경희대학교 영상미디어연구실

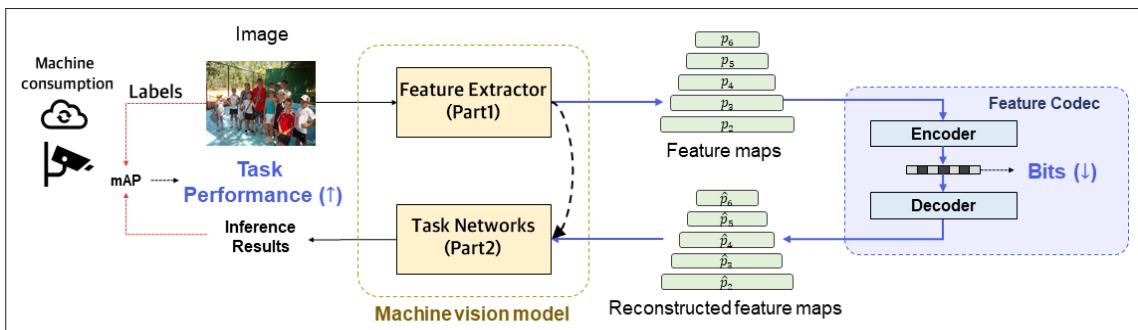
I. 서론

전통적인 비디오 압축은 효율적인 통신 및 저장을 위해 적은 데이터양으로 비디오를 압축하고, 다시 비디오로 복원하였을 때 화질 손실을 최소화하는 것을 목표로 한다. 최근에는 기계 시각(Machine Vision)을 활용한 기술들의 등장으로 〈그림 1〉과 같이 인간이 아닌 기계 시각을 위한 압축에 관한 연구 또한 활발히 진행 중이다. 이러한 기계 시각을 위한 압축 방법은 스마트시티(smart city), 자율 주행 등 기계 시각이 탑재된 기기 간의 통신

을 더욱 효율적으로 개선할 수 있다. 본 연구에서는 기계 시각 모델에서 사용되는 데이터를 효과적으로 압축할 수 있는 모델을 제안한다.

II. 기계 시각 작업 및 다중 스케일 특징맵

〈그림 3〉은 기계 시각의 종류인 객체 검출 및 객체 분할에서 사용되는 다중 스케일 특징맵의 예시를 나

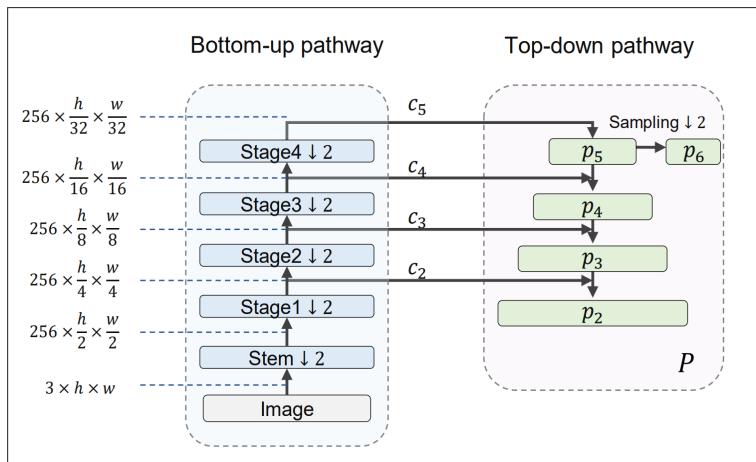


<그림 1> 기계 시각을 위한 다중 스케일 특징맵 압축 과정

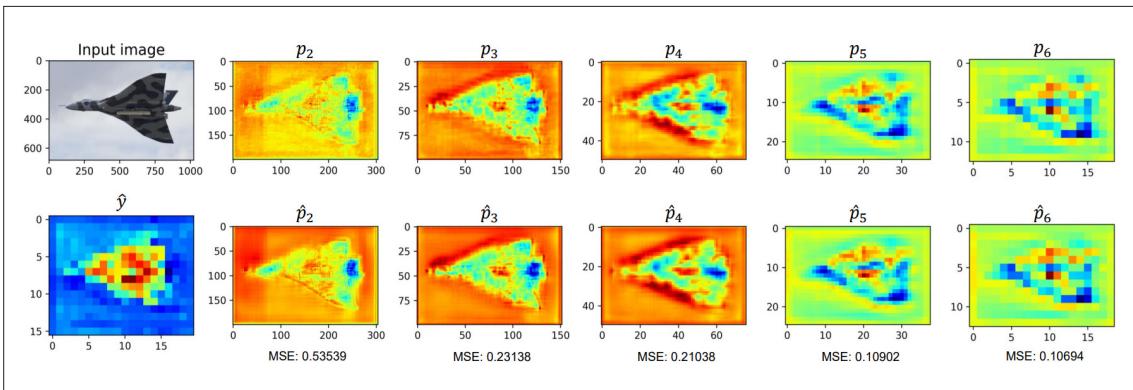
타낸다. 이러한 특징맵은 이미지를 입력으로 하여 〈그림 2〉와 같은 특징 추출기를 통해 추출된다. 다중 스케일 특징맵은 이미지 내에 존재하는 객체 정보를 담고 있으며, 서로 다른 해상도를 가짐으로써 다양한 크기의 객체를 인식하는 데에 효과적으로 사용될 수 있다. 제안 모델은 이러한 다중 스케일 특징맵을 압축하여, 압축된 특징맵을 이용하여 기계 작업을 수행하였을 때 성능 하락을 최소화하는 것을 목표로 한다.

III. 제안된 다중 스케일 특징맵 압축 방법

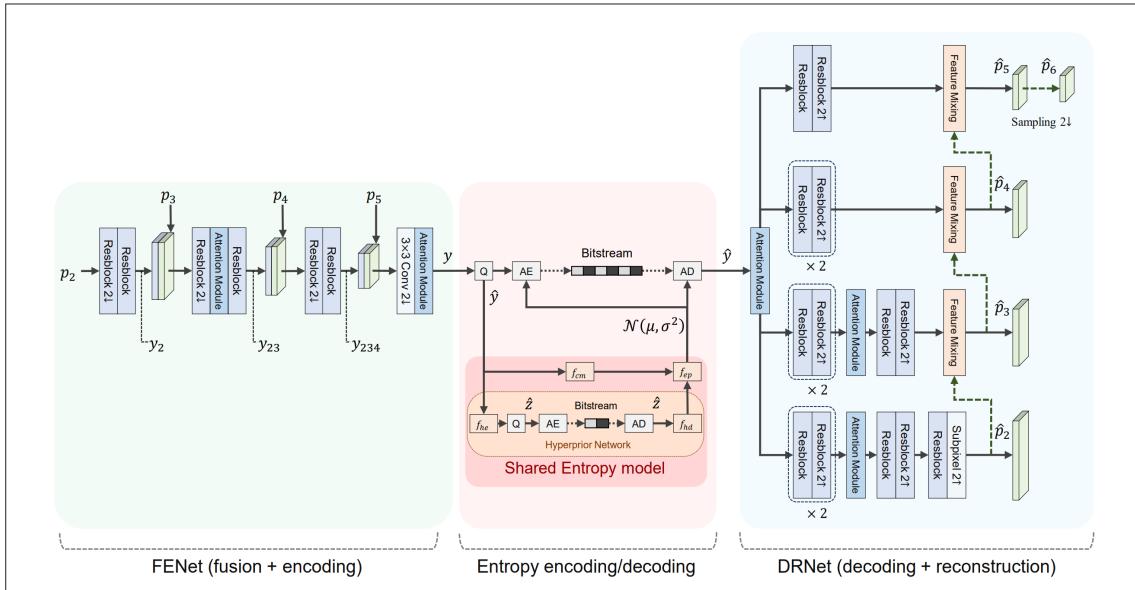
제안 모델의 부호화기는 더 큰 스케일의 특징맵을 더 작은 스케일의 중간 임계 특징맵(latent feature map)으로 변환한 후 반복적으로 작은 스케일의 특징맵과 융합(fusion)하는 과정을 통해 최종적으로 전체 다중 스케일 특징맵의 정보를 내재하는 작은 크기의 임계 특징맵을 생성한다.



〈그림 2〉 다중 스케일 특징맵을 추출하기 위한 FPN 구조



〈그림 3〉 다중 스케일 특징맵 및 압축된 다중 스케일 특징맵의 예시



<그림 4> 제안 모델의 전체 구성도

제안 모델의 복호화기는 부호화기에서 생성된 잠재 특징맵으로부터 다중 스케일 특징맵을 다시 복원하기 위해 대칭적으로 구성되었으며, 부호화기와 복호화기 사이에는 잠재 특징맵을 무손실 부호화하기 위한 엔트로피 모델(Entropy Model)[3]이 사용된다. 제안 모델의 전체 구성은 <그림 4>에서 확인할 수 있다.

제안 모델은 기존 모델 [4], [5]와 달리 종단 간 학습이 가능하기 때문에 다중 스케일 특징맵에 최적화된 잠재 공간(latent space)을 학습하여 높은 부호화 효율을 달성할 수 있다는 이점이 있다. 또한, 부호화기 및 복호화기를 다중 스케일 특징맵의 특성을 고려하여 설계함으로써 구조적 중복성을 최소화했다.

IV. 실험 결과

실험은 객체 탐지 및 객체 분할 작업에 대해 수행되었

으며, 제안 모델은 종래 기술들 대비로 각 작업에 대해 각각 -52.23%, -84.58% BD-rate 감소 효율을 보였다. 또한, 일정 수준 이상의 비트량을 사용하였을 때, 압축되지 않은 특징맵으로 달성할 수 있는 성능과 거의 동일한 성능을 달성할 수 있다. 다시 말해, 제안 모델을 사용하면, 기계 시각 작업 모델이 탑재된 단말 간의 통신 과정에서 무 압축 대비 약 0.002-0.003% 데이터양만으로 동일한 기계 작업 성능을 달성할 수 있게 된다.

V. 결론

본 연구에서는 특징맵 융합 기반의 종단 간 학습 가능한 다중 스케일 특징맵 압축 모델을 제안한다. 제안된 모델은 객체 감지 및 객체 분할 작업에서 종래 기술들과 함께 평가되었다. 결과는 제안 모델이 가장 높은 부호화 효율뿐만 아니라 기존 방법에 비해 훨씬 적은 부호화 시간

으로 최적의 복잡도-성능 균형을 제공한다. 더 나아가, 우리의 모델은 압축되지 않은 것보다 적어도 20,000배 이상 적은 데이터양으로 거의 손실이 없는 기계 작업 성능을 달

성한다. 미래에는 제안된 모델을 시간적 종복성이 중요하게 고려되어야 하는 객체 추적과 같은 비디오 작업을 지원하도록 확장할 수 있다.

참 고 문 헌

-
- [1] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real time object detection with region proposal networks,” in Advances in neural information processing systems, 2015, pp. 91-99.
 - [2] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.
 - [3] Cheng, H. Sun, M. Takeuchi, and J. Katto, “Learned image compression with discretized gaussian mixture likelihoods and attention modules,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7939-7948.
 - [4] H. Han, M. Choi, H. Choi, S.-h. Jung, S. Kwak, H.-G. Choo, W.-S. Cheong, and J. Seo, “[VCM Track 1] Experimental results of multi-scale feature anchor generation on m59576,” MPEG Technical requirements, vol. ISO/IEC JTC 1/SC 29/WG 2, no. m60130, Jul 2022.
 - [5] D.-H. Kim, Y.-U. Yoon, J.-G. Kim, J. Lee, and S.-Y. Jeong, “[VCM Track 1] Performance of the enhanced msfc with bottom-up msff,” MPEG Technical requirements, vol. ISO/IEC JTC 1/SC 29/WG 2, no. m60197, Jul 2022.
 - [6] Y. Zhang, Z. Liu, X. Xu, S. Liu, and Z. Chen, “[VCM Track 1] Learning-based feature compression for instance segmentation,” MPEG Technical requirements, vol. ISO/IEC JTC 1/SC 29/WG 2, no. m60240, Jul 2022.
-

김 영 응



- 2022 : 경희대학교 응용수학과 학사
- 2022 ~ 현재 : 경희대학교 컴퓨터공학과 석사과정
- 주관심분야 : 비디오 압축, 기계 시각을 위한 비디오 압축