VCM 의 바텀-업 MSFF 를 이용한 MSFC 기반 멀티-스케일 특징 압축 네트워크 개선

김동하, 한규웅, 차준석, 김재곤
한국항공대학교
{donghakim, woong2614, wnstjr401}@kau.kr, jgkim@kau.ac.kr

Enhancement of MSFC-Based Multi-Scale Features Compression Network with Bottom-UP MSFF in VCM

Dong-Ha Kim, Gyu-Woong Han, Jun-Seok Cha, and Jae-Gon Kim
Korea Aerospace University

요 약

MPEG-VCM(Video Coding for Machine)은 입력된 이미지/비디오의 특징(feature)를 압축하는 Track 1 과 입력 이미지/비디오를 직접 압축하는 Track 2 로 나뉘어 표준화가 진행 중이다. 본 논문은 Track 1 의 비전임무 네트워크로 사용하는 Detectron2 의 FPN(Feature Pyramid Network)에서 추출한 멀티-스케일 특징을 효율적으로 압축하는 MSFC 기반의 압축 모델의 개선 기법을 제시한다. 제안기법은 해상도를 줄여서 단일-스케일 압축맵을 압축하는 기존의 압축 모델에서 저해상도 특징맵을 고해상도 특징맵에 바텀-업(Bottom-Up) 구조로 합성하여 단일-스케일 특징맵을 구성하는 바텀-업 MSFF 를 가지는 압축 모델을 제시한다. 제안방법은 기존의 모델 보다 BPP-mAP 성능에서 1 ~ 2.7%의 개선된 BD-rate 성능을 보이며 VCM 의 이미지 앵커(image anchor) 대비 최대 -85.94%의 BD-rate 성능향상을 보인다.

1. 서론

최근 딥러닝 기반의 머신비전(machine vision) 응용이 확산되면서 기계를 위한 새로운 비디오 부호화 표준으로 MPEG-VCM(Video Coding for Machines) 표준화가 진행 중이다[1]. VCM 은 사람이 소비하는 기존의 HVS 기반의 부호화 보다 비전 임무(task)를 수행하는 기계의 소비를 주 목적으로 한 것으로 화질 보다는 임무 수행 성능을 고려한 보다 효율적인 압축을 제공하고자 한다. 현재 VCM 은 임무수행 네트워크에서 입력 이미지/비디오로부터 추출한 특징(feature)을 부호화 하는 Track 1 과 입력 이미지/비디오를 직접 부호화하는 Track 2 로 나누어 표준화를 진행 중이며, 올 10 월 회의에서 Track 2 의 기술제안요청서(CfP)와 Track 1 의 기술조사요청서(CfE)[2]에 대한 응답 기고 기술을 논의하고 본격적인 표준화를 진행할 예정이다.

VCM Track 1 기술로 제시된 MSFC(Multi-Scale Feature Compression) 기반의 압축 네트워크 모델은 임무수행 네트워크의 백본(backbone) 네트워크에서 추출한 멀티-스케일(multi-scale) 특징을 뛰어난 성능으로 압축하는 것으로 보고되었다[3]. 본 논문에서는 바텀-업(bottom-up) MSFF(Multi-Scale Feature Fusion)를 가지는 개선된 MSFC 기반의 압축 모델을 제시한다.

2. MSFC 기반의 멀티-스케일 특징 압축

그림 1 은 VCM 후보 기술로 제시된 MSFC 기반의 압축 네트워크 구조이다. 그림 1 과 같이 MSFC 기반의 압축 모델은 MSFF 와 MSFR(Multi-Scale Feature Reconstruction) 모듈로 구성되어, 임무수행 네트워크의 백본 네트워크에서 추출한 특징맵 Px (x = 2, 3, 4, 5)를 압축한다. MSFF 모듈은 멀티-스케일의 특징맵 Px (x = 2, 3, 4)를 가장 상위 계층의 특징맵인 P_5와 동일한 해상도를 가지는 1,024 채널로 정렬된 단일-스케일 특징맵을 64 채널로 축소한 특징맵 F를 출력한다. 이를 위해 MSFF 는 그림 2 와 같이 Align & Concatenation, SE(Squeeze and Excitation) Block 과

컨볼루션(Convolution) 계층으로 구성된다. 축소된 특징맵은 기존 코덱인 VVC 로 압축 복호화하고, 복호화된 특징맵을 MSFR 모듈에서 다시 멀티-스케일의 특징으로 복원한다. 이때, MSFR 에서 복호화된 특징맵 F' 로부터 탑-다운(top-down) 방법으로 멀티-스케일의 특징맵 Px' $(x = 2,3,4,5)$을 재구성한다.

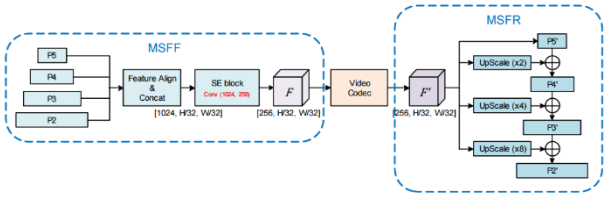
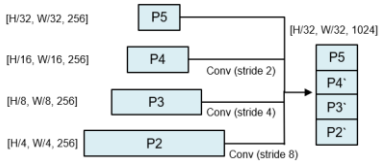
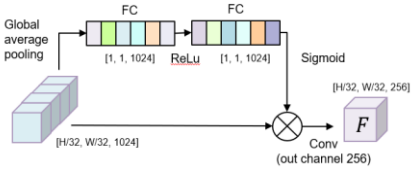


그림 1. MSFC 기반의 압축 모델 구조



(가) Feature Align & Concatenation



(나) SE Block

그림 2. MSFF 구조

3. 바텀-업 멀티-스케일 특징 압축

MSFR에서 탑-다운 방법으로 멀티-스케일 특징을 복원할 가장 상위 계층 특징맵 P_5' 를 활용하여 차례로 상위 계층 특징맵부터 하위 계층 특징맵까지 재구성한다. 이 때, 상위 계층의 특징맵은 잘 유지되지만 하위 계층의 특징맵은 재구성시 정보 손실이 될 수 있다. 이것은 그림 3과 같이 기존 MSFC 기반의 압축 모델에서 특징 간의 합성 없이 단순히 채널 축소의 목적으로 MSFF 가 구성되었기 때문이다. 따라서, 본 논문에서는 압축율 대비 임무수행 정확도 성능을 개선하기 위해서 상대적으로 저해상도 특징을 많이 포함하고 있는 특징 P_3, P_4를 합성을 통해 저해상도 특징 손실 현상을 개선하는 바텀-업 MSFF 구조를 제안한다[4].

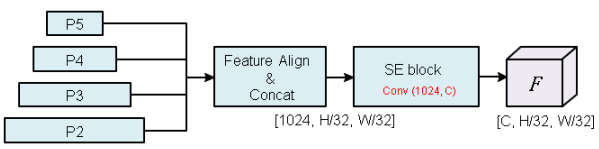


그림 3. MSFC 기반의 압축 모델의 MSFF 구조

그림 4는 기존 MSFF 구조에서 계층간 특징 합성을 적용한 제안된 바텀-업 MSFF 구조이다. 먼저 특징 P_2를 컨볼루션 계층을 통해 다운 스케일링(down scaling)하여 특징 P_3와 더하고, 그 결과를 미세 조정을 위한 컨볼루션 계층을 통해서 특징 P_3'을 생성한다. 같은 방식으로 특징 P_3'과 P_4을 합성해서 P_4'을 생성하고 이들을 기존의 Align & Concatenation과 SE Blcok을 거쳐서 최종 바텀-업 MSFF의 특징 F를 출력한다.

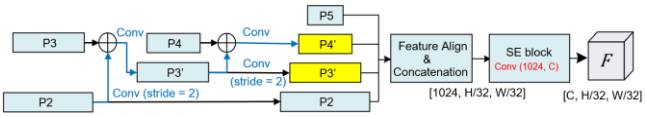


그림 4. 바텀-업 MSFF 를 가지는 개선된 MSFC 모델

4. 실험결과

Detectron2[5]의 Faster R-CNN X101-FPN 의 FPN(Feature Pyramid Network)으로 멀티-스케일 특징을 추출한다. MSFC 모델은 모든 R-CNN 네트워크의 파라미터를 고정하고 바텀-업 MSFF 와 기존 MSFR 모듈만을 학습하다. 학습에는 COCO train2017[6] 데이터셋을 사용하였고, OpenImages V6[7]의 검증 영상 5,000 장을 사용하여 학습된 모델을 평가했다. 배치 크기는 2, 학습률(learning rate)은 0.0005 로 설정하고 300,000 회 반복하였다. 바텀-업 MSFF 모듈의 출력 F 는 특징맵으로 변환하여 VTM 12.0 으로 QP {22, 27, 32, 37, 42, 47}으로 부호화 하였다.

표 1. BPP_mAP 성능비교(KAU-MSFC: 기존 MSFC 모델, Bottom-up MSFF: 제안 MSFC 모델)

	KAU MSFC 192		Bottom up MSFC 192		KAU MSFC 64		Bottom up MSFC 64	
	BPP	mAP	BPP	mAP	BPP	mAP	BPP	mAP
Uncom.	-	78.806	-	79.230	-	78.549	-	78.733
qp22	0.520	78.804	0.519	79.307	0.176	78.711	0.175	78.971
qp27	0.341	78.685	0.340	79.409	0.117	78.554	0.116	78.777
qp32	0.181	78.544	0.182	79.165	0.064	78.811	0.063	78.997
qp37	0.076	75.524	0.077	76.469	0.028	76.675	0.027	76.862
qp42	0.029	67.262	0.030	67.977	0.011	67.665	0.011	68.359
qp47	0.010	51.977	0.011	52.388	0.004	53.202	0.004	53.545
BD rate	-59.20%		-61.92%		-84.98%		-85.94%	
	-2.7%				-0.96%			

그림 5 는 제안한 바텀-업 MSFF 를 적용한 MSFC 와 기존의 KAU-MSFC 의 BPP-mAP 성능을 192, 64 채널에 대해 비교한 것이다. 표 1 에 보인 결과와 같이 제안된 기법 또한 KAU-MSFC 와 같이 VCM 영상 앵커보다 우수한 BPP-mAP 성능을 보였으며, 기존 KAU-MSFC 에 비해 64 채널과 192 채널에서 각각 0.96%, 2.72%의 BD-rate 성능 개선을 보였다. 이것은 VCM 의 이미지 앵커(Image Anchor) 대비 최대 85.94%의 BD-rate 이득을 얻는다.

5. 결 론

본 논문에서는 기존 MSFC 기반의 특징 압축 모델의 성능 개선 기법으로 바텀-업 MSFF 구조를 가지는 MSFC 기반의 압축 모델을 제시하였다. 바텀-업 MSFF 는 백본 네트워크에서 추출된 멀티-스케일 특징들 중 상대적으로 하계계층에 해당하는 특징 P_3, P_4를 합성을 통해 특징 사이의 상관성이 높은 단일-스케일 특징을 도출하는 방법이다. 실험결과를 기존 MSFC 기반의 압축 모델과 비교하였을 때 192 채널, 64 채널 모두에서 1 ~ 2.7%의 추가 BD-rate 성능 향상을 보였으며 이는 VCM 의 이미지 앵커 대비 최대 85.94%의 큰 BD-rate 이득을 제공하는 것으로 향후 표준기술로 고려될 수 있을 것으로 기대된다.

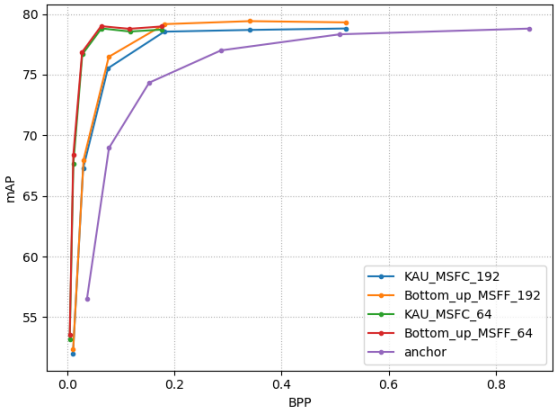


그림 5. 제안모델의 BPP-mAP 성능

Acknowledgement

본 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2021-0-00191, 기계를 위한 영상 부호화 기술).

참 고 문 헌(References)

[1] “Evaluation Framework for Video Coding for Machines”, ISO/IEC JTC1/SC29/WG2/N00162, Apr. 2022.

[2] “Call for Evidence on Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2, N00215, Jul. 2022.

[3] D. Kim, Y. Yoon, J. Kim, J. Lee, Y. Kim, and S. Jeong “[VCM Track1] Compression of FPN Multi-Scale Features for Object Detection Using VVC,” ISO/IEC JTC 1/SC 29/WG2, m59562, Apr. 2022.

[4] D. Kim, Y. Yoon, J. Kim, J. Lee, Y. Kim, and S. Jeong, “[VCM-Track1] Performance of the Enhanced MSFC with Bottom-Up MSFF,” ISO/IEC JTC 1/SC 29/WG2, m60197, Jul. 2022.

[5] “detectron2”,[Online]. Available: https://github.com/facebookresearch/ detectron2

[6] “COCO dataset”, 2017, [Online]. Available: https://cocodataset.org/#download

[7] “OpenImages”, V6, [Online]. Available: https://storage.googleapis.com/openimages/web/index.html