

## VCM 을 위한 FPN 다중 스케일 특징 압축

김동하, 윤용욱, \*이주영, \*정세윤, 김재곤, 정대권

한국항공대학교, \*한국전자통신연구원

{donghakim, yuyoon}@kau.kr, jgkim@kau.ac.kr

### Compression of Multiscale Features of FPN for VCM

Dong-Ha Kim, Yong-Uk Yoon, \*Jooyoung Lee, \*Se-Yoon Jeong, Jae-Gon Kim, Dae-Gwon Jeong

Korea Aerospace University, \*ETRI

#### 요 약

MPEG-VCM(Video Coding for Machine)은 입력된 비디오 특징(feature)를 압축하는 Track1 과 입력 영상을 직접 압축하는 Track2 로 나뉘어 표준화가 진행중이다. 본 논문은 VCM Track 1 에 해당하는 Detectron2 FPN(Feature Pyramid Network)에서 추출한 다중 스케일 특징맵을 VVC 로 압축하는 MSFC(Multi-Scale Feature Compression)을 구조를 제안한다. 본 논문의 MSFC 에서는 다중 스케일 특징을 결합하여 부호화/복호화하는 기존의 구조에서 특징맵의 해상도를 줄여 압축하는 개선된 MSFC 를 제시한다. 제안 방법은 VCM 의 Track2 의 영상 앵커(image anchor) 보다 우수한 BPP-mAP 성능을 보이고 최대 -84.98%의 BD-rate 성능향상을 보인다.

#### 1. 서론

최근 딥러닝 네트워크를 활용한 객체탐지나 객체분할 등의 머신 비전(machine vision)에 관한 연구들이 활발히 진행되고 있다. 이때, 머신 비전에 입력하는 영상이나 딥러닝 네트워크에서 발생하는 정보를 압축 및 전송할 때, 휴먼 비전(human vision)을 목적으로 하는 기존의 비디오 코덱 보다 머신 비전 응용에 적합한 기계를 위한 새로운 표준 코덱이 보다 효율적일 수 있다[1]. 따라서, MPEG 은 머신 비전을 위한 새로운 비디오 부호화 표준인 VCM(Video Coding for Machine)의 표준화를 진행하고 있다. 현재 VCM 은 입력된 비디오로부터 추출한 특징(feature)을 부호화 하는 Track1 과 입력 영상을 직접 부호화 하는 Track2 로 나뉘어 표준화를 진행중이다[2].

본 논문은 VCM Track 1 에 해당하는 Detectron2 FPN(Feature Pyramid Network) [3]에서 추출한 다중 스케일 특징(Multi-Scale feature)을 VVC(Versatile Video Coding)를 이용하여 압축하는 개선된 MSFC(Multi-Scale Feature Compression) [4], [5]를 제안한다.

#### 2. 다중 스케일 특징 압축

그림 1 은 MSFC 의 구조를 나타낸 것이다. 다중 스케일 특징  $P_x$  ( $x = 2, 3, 4, 5$ ) 는 입력된 영상 해상도의 가로 및 세로의 각각  $1/4$ ,  $1/8$ ,  $1/16$ ,  $1/32$  의 크기를 가지고, 각 특징들은 256 채널로 구성된다. MSFC 는  $P_x$  ( $x = 2, 3, 4, 5$ )를 입력 받아 하나의 특징  $F$ 로 결합하는 MSFF(Multi-Scale Feature Fusion) 모듈과, VVC 를 사용해서 부호화/복호화된 특징  $F'$ 로부터 원래의 다중 스케일 특징의 형태인  $P_{x'}$  ( $x = 2, 3, 4, 5$ ) 로 재구성(reconstruction)하는 MSFR(Multi-Scale Feature Reconstruction) 모듈로 구성된다.

그림 2 는 MSFF 의 상세 구조를 나타낸 것이다. MSFF 는 다중 스케일 특징을 하나의 스케일로 크기를 조정된 다음 결합하는 과정을 거친다. 즉, 그림 2(가)에서  $P_{x'}$  ( $x = 2, 3, 4$ ) 들은 컨볼루션 계층(convolutional layer)을 통해  $P_5$  와 동일한 해상도로 스케일링된 후,  $P_5$ ,  $P_4$ ,  $P_3$ ,  $P_2$  의 순서로 채널 방향으로 연결한다. 연결된 특징은 1024 개의 채널을 가지며, 그림 2(나)와 같이 두 개의 FC(Fully Connected) 계층과 활성화 함수로 이루어진 SE block(Squeeze and Excitation

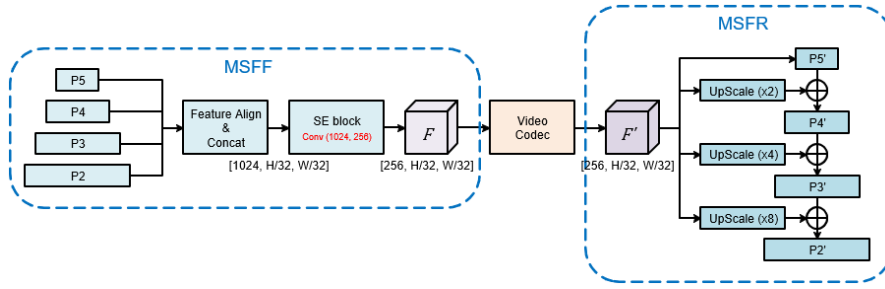
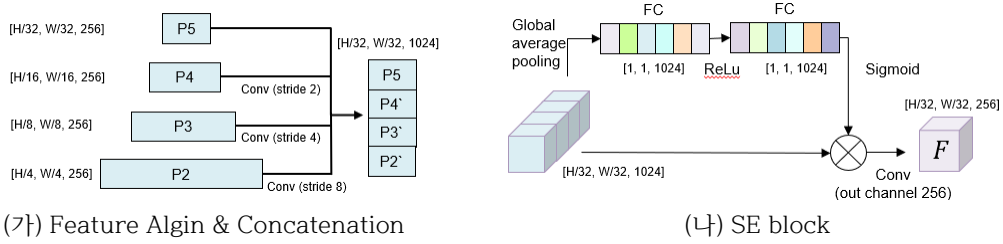


그림 1. MSFC 구조



(가) Feature Algin & Concatenation

(나) SE block

그림 2. MSFF 구조

block)을 통해 채널별로 중요도에 따라 가중화 된다. 이와 같이 중요도가 재설정된 1024 채널로 구성된 특징은 컨볼루션 계층을 통해 최종 256 채널로 축소된 특징  $F$ 로 출력된다. 특징  $F$ 는 다중 스케일 특징 중  $P5$ 와 동일한 크기와 채널 수를 갖는다.

정규화를 거치게 되며, 이때의 최소값과 최대값을 전송하여  $F'$  복호화에 사용한다.

### 3. 채널 축소 다중 스케일 특징 압축

VVC 로 압축되는 특징맵의 해상도는 특징  $F$ 를 구성하는 채널 수에 따라 결정된다. 따라서, 본 논문에서는 압축율(BPP: Bit-Per-Pixel) 대비 임무 수행 정확도(mAP: mean Average Precision) 성능을 개선하기 위해서 특징  $F$ 의 채널 수를 줄임으로써 특징맵의 해상도를 줄여 부호화 하는 개선된 MSFC 구조를 제안한다.

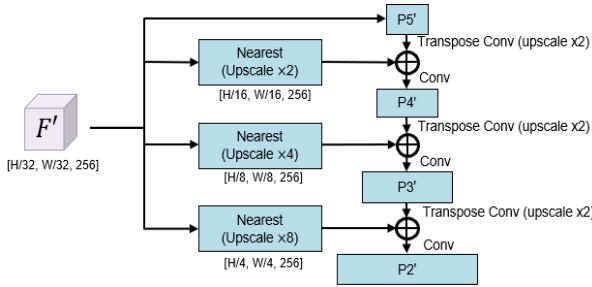


그림 3. MSFR 구조

그림 3은 MSFR의 구조이다. MSFR은 VVC를 이용하여 부호화/복호화된 특징  $F'$ 로부터 원래의 다중 스케일 형태를 가지는 특징  $Px'$  ( $x = 2, 3, 4, 5$ )을 재구성한다.  $F'$ 은  $P5$ 와 동일한 크기와 채널 수를 가지므로 그대로  $P5'$ 로 재구성된다. 나머지 특징들  $Px'$  ( $x = 3, 4, 5$ )은 그림 3과 같이 최접근 보간(nearest interpolation)으로 하위 계층 해상도로 업스케일링(upsampling)된 특징  $F'$ 와 전치(transpose) 컨볼루션 계층을 사용하여 업스케일링된 특징  $Px'$  ( $x = 3, 4, 5$ )을 더하여 재구성된다. 이 재구성 과정은 그림 3과 같이 탑다운(top-down) 방식으로 진행되어  $P5'$ ,  $P4'$ ,  $P3'$ ,  $P2'$ 의 순서로 재구성된다.

그림 4는 제안된 MSFC의 구조로 확장된 MSFF와 확장된 MSFR 모듈로 구성된다. 기존의 MSFF가 채널별로 가중화된 특징을 1024 채널에서 256 채널로 줄이는 것에 비해서, 확장된 MSFF는  $F$ 채널 축소를 기존의 256 채널보다 적은 채널  $C$ 로 축소한다. 이에 따라 부호화 할 특징맵의 해상도가 줄어들게 된다.

제안된 MSFC 구조는 기존의 MFSC의 MSFF와 MSFR 모듈을 사용하여 MSFF의 출력 특징  $F$ 를 VVC의 시험모델인 VTM(VVC Test Model)로 부호화하고 특징  $F'$ 으로 복호화한다. 이때, 특징  $F$ 를 구성하는 모든 채널들을 공간 패킹(packaging)하여 특징맵을 생성하고 최소-최대(Min-Max)

확장된 MSFR은 채널  $C$ 로 축소된 특징  $F'$ 을 다중 스케일 특징  $Px'$  ( $x = 2, 3, 4, 5$ )으로 재구성한다. 이때,  $C$  채널로 축소된 특징  $F'$ 을 기존과 같이 256 채널로 확장한 후 기존의 방법으로 특징  $Px'$  ( $x = 2, 3, 4, 5$ )으로 재구성하도록 기존의 MSFR 구조를 확장한다. 즉,  $F'$ 은 추가된 컨볼루션 계층을 통하여 채널  $C$ 가 256으로 복원된 후  $P5'$ 로 사용한다. 최접근 보간을 적용하여 해상도가 증가된 특징  $F'$ 도 컨볼루션 계층채널  $C$ 를 256으로 복원한다. 256 채널로 복원된 해상도가 증가한 특징  $F'$ 과 256 채널로 복원된 특징  $P5'$ 를 활용하여 기존의 MSFR과 같이 탑다운 방식으로 다중 스케일의 특징을 재구성한다.

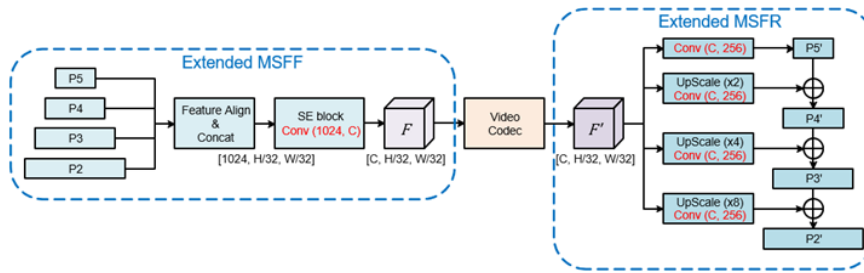


그림 4. 개선된 MSFC 구조

### 4. 실험결과

제안하는 채널 축소 다중 스케일 특징 압축은 VCM 의 객체탐지 임무의 평가 네트워크로 정의된 Detectron2 의 Faster R-CNN X101-FPN 의 FPN 으로부터 추출된 특징  $P_x$  ( $x = 2, 3, 4, 5$ ) 를 사용하여 성능을 평가했다. MSFC 를 학습할 때, 모든 R-CNN 네트워크의 파라미터를 동결한 채, 비디오 코덱은 제외하고 MSFF 와 MSFR 모듈만을 학습했다. 배치(batch) 크기는 2 로 하고 300,000 반복 수(iteration)와 0.0005 학습률(learning rate)로 학습하였다. 학습에는 COCO train2017 [6] 데이터셋을 사용하였고, OpenImages V6 [7]의 검증 영상 5 천 장을 사용하여 학습된 모델을 평가했다. 특징  $F$  는 특징맵으로 변환하여 VTM 12.0 으로 QP {22, 27, 32, 37, 42, 47}에 대해 All Intra 로 부호화 하였다.

그림 5 는 제안한 개선된 MSFC 의 BPP-mAP 성능을 기존 MSFC 와 비교한 것이다. 개선된 MSFC 의 채널  $C$  는 192, 144, 64 채널에 대해서 성능을 확인했다. 제안기법은 모든 채널의 경우에서 VCM 영상 앵커(anchor) 보다 월등히 우수한 BPP-mAP 성능을 보였고, 채널 축소를 통한 64 채널에서 VCM 앵커 대비 최대 -84.98%의 BD-rate 이득을 얻었다.

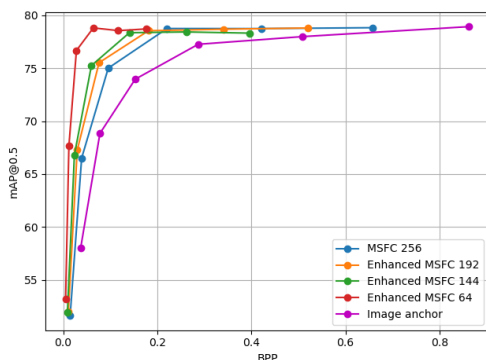


그림 5. 개선된 MSFC 의 BPP-mAP 성능

### 5. 결론

본 논문은 객체탐지 머신 비전 네트워크의 FPN 에서 추출한 다중 스케일 특징을 VVC 를 이용해 압축하는 개선된

MSFC 구조를 제안하였다. 제안기법에서 확장된 MSFF 는 다중 스케일 특징을 단일 스케일의 특징으로 연결되는 특징의 채널 수를 축소하여, VVC 로 압축하는 특징맵의 해상도를 줄인다. 또한, MSFR 을 확장하여 축소된 특징 채널을 다시 복원한 후 기존 방법으로 다중 스케일 특징을 재구성한다. 개선된 MSFC 는 제안한 특징맵 구성 및 단일 특징 채널 축소를 통하여 VCM 의 영상 앵커 대비 최대 -84.58%의 BD-rate 이득을 얻었다. 본 논문의 VVC 를 이용한 다중 스케일 특징 압축 기법은 VCM의 특징 압축의 접근방법으로 고려될 수 있을 것으로 보인다.

### Acknowledgement

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

### 참고 문헌(References)

- [1] "Use cases and Requirements for Video Coding for Machines" ISO/IEC JTC 1/SC 29/WG 2, N00190, Online, Apr. 2022.
- [2] "Evaluation Framework for Video coding for Machines," ISO/IEC JTC 1/SC 29/WG 2 N00162, Online, Jan. 2022.
- [3] "detectron2", [Online]. Available: <https://github.com/facebookresearch/detectron2>
- [4] H. Han, H. Choi, S.-h. Jung, S. Kwak, J. Yun, W.-S. Cheong, and J. Seo, "[VCM] investigation on deep feature compression framework for multi-task," ISO/IEC JTC 1/SC 29/WG 2, m58772, Jan. 2022.
- [5] 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.
- [6] "COCO dataset", 2017, [Online]. Available: <https://cocodataset.org/#download>
- [7] "OpenImages", V6, [Online]. Available: <https://storage.googleapis.com/openimages/web/index.html>