

다중 임무 수행을 위한 관심 객체 기반 전처리 네트워크

이희경 이진영 추현곤 서정일

한국전자통신연구원

lhk95@etri.re.kr jinlee@etri.re.kr hyongonchoo@etri.re.kr seoji@etri.re.kr

Object-of-interest-based preprocessing network for multi-tasking

Lee, HeeKyung Lee, Jin Young, Choo, Hyon-Gon, Seo Jeongil

Electronics and Telecommunications Research Institute

요 약

기계 소비(machine consumption)를 대상으로 하는 이미지 및 비디오의 압축 목표는 사람이 소비하는 것과 다를 수 있다. 예를 들어, 인간의 소비에는 이미지 및 비디오의 전체 캡처 영역이 필요하지만, 기계는 응용 프로그램에서 요구하는 이미지 또는 비디오의 일부만 사용할 수 있다. 이 논문에서는, 효과적인 기계 소비를 위한 영상 압축 방법 개발을 통한 다중 임무(task) 수행 성능 개선 및 압축률(bpp) 향상을 목적으로, 관심 객체(object of interest) 또는 관심 영역(region of interest)을 식별하기 위한 신경망 기반 영상 분석 작업을 수행하는 전처리 네트워크 구성 방법을 제안한다.

1. 서론

최근 MPEG에서는 기존의 인간 시각 시스템을 위한 영상 부-복호화 기술이 아닌, 기계 소비(machine consumption)에 효과적인 압축 기술인 VCM (Video Coding for Machine) 표준화가 진행 중이다. 이는 기존 압축 기술의 목표인 압축 효율 향상뿐 아니라, 기계의 분석 효율 향상, 연산량 감소, 개인 정보 보호를 목표로 한다. 지난 4월 138차 회의에서 이미지/비디오 압축 트랙(Track)에 대한 Call-for-Proposal (CFP)가 발간되었으며, 10월 회의에서 이에 대한 평가를 진행한다. 본 논문에서는 효과적인 기계 소비를 위한 영상 압축 방법 개발을 통한 다중 임무(task) 수행 성능 개선 및 압축률(bpp) 향상을 목적으로, 관심 객체(object of interest) 또는 관심 영역(region of interest)을 식별하기 위한 신경망 기반 영상 분석 작업을 수행하는 전처리 네트워크 구성 방법을 제안한다.

2. 본론

높은 기계 분석 성능 및 압축 효율을 낼 수 있는 가장 효과적인 방법은 기계가 집중해야 하는 관심 객체(Object of Interest)에 대한 정보, 즉 서술자(Descriptor)를 전달하여 다중 임무를 수행하는 것이다^[1,2,3,4,5]. 서술자(descriptor) 기반 방법을 사용하면 기계 응용 프로그램의 활용도와 재사용률을 높일 수 있다. 서술자 정보 기반의 1차적인 기계 분석이 끝나면, 선택적으로 중요한 관심 객체(Object of Interest) 영상을 분석할 필요가 있을 수 있다. 따라서, 전처리 네트워크를 이용하여 관심 객체 영역만을 추출한 후, 이를 압축, 전송하여 개체 감지(object detection), 인스턴스 분할(instance segmentation), 개체 추적(object tracking)과 같은 다중 기계 임무(machine task)를 수행하는 것은 낮은 비트율에

서도 높은 기계 분석 성능을 얻는데 큰 도움이 된다.

그림 1은 전술한 관심 객체 추출 전처리 네트워크의 파이프라인이다. 비디오/영상 입력 단계에서 입력된 영상으로부터, 관심 객체 추출 전처리 네트워크를 수행하여 관심 객체(object of interest) 또는 관심 객체가 포함된 일부 영역, 즉 관심 객체 영역을 분리한다. 이때, 관심 객체 영역은 입력 비디오/영상에 속한 객체를 그 윤곽선을 따라 정확하게 분할한 임의의 형상(arbitrary shape)을 갖는 세그먼트(segment) 영역 또는 해당 객체를 포함하는 사각형(rectangular shape) 영역일 수 있다. 이 관심 객체 영역을 영상 압축 기술(예를 들어, VCM 또는 VVC(Versatile Video Coding))로 부호화하여 전송한다. 수신 및 복호화된 영상에 대해서, 용도에 따라 개체 감지, 인스턴스 분할, 개체 추적, 판옵틱 분할(panoptic segmentation), 키포인트 검출(keypoint detection), 자세 검출(densepose detection) 등의 다양한 다중 임무(task)를 수행할 수 있다.

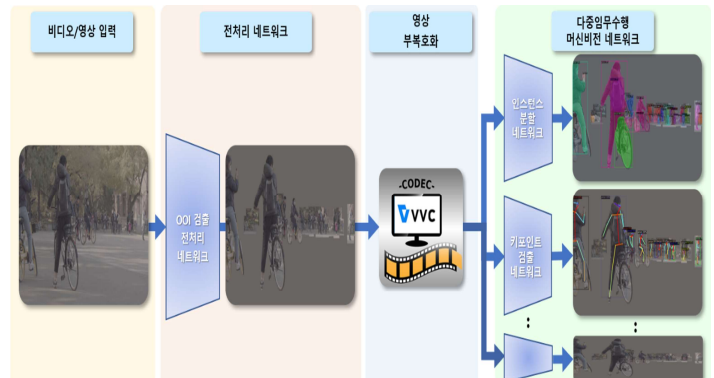


그림 1. 관심 객체 기반 전처리 네트워크 파이프라인

그림 2는 관심 객체 검출 전처리 네트워크의 한 예로 Detectron2 개체 감지 또는 인스턴스 분할 네트워크나 판옵틱 분할 네트워크를 사용

하여 입력 비디오 또는 이미지로부터 하나 또는 그 이상의 관심 객체 영역을 추출한다. 추출된 초기 관심 객체 영역은 후처리 네트워크에서의 다중 임무 수행 성능 향상을 위해 다양하게 조정하여 전송할 수 있다. 예를 들면, 특정 크기 이하인 관심 객체 영역을 저해상도 영상을 고해상도 영상으로 변환하는 super resolution 네트워크를 활용해 확대(upsampling)하거나, 특정 크기 이하인 세그먼트 영역을 바운딩 박스(bounding box) 영역으로 대체할 수 있다. 이때, 크기에 대한 기준 값은 임의의 픽셀 개수가 될 수 있으며, 그림 3은 기준 값을 20000 픽셀로 설정한 예이다.

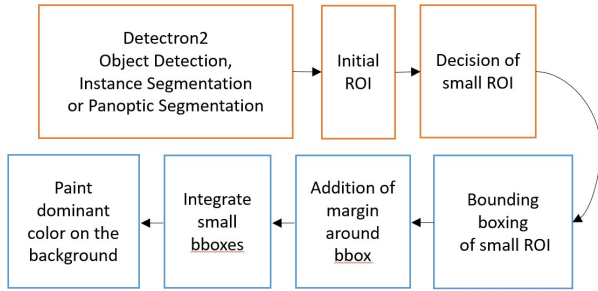


그림 2. 관심 객체 검출 전처리 네트워크의 한 예



그림 3. 그림 2의 전처리 네트워크 적용 결과

뿐만 아니라, 크기가 또 다른 기준 값 이하인 관심 객체 영역에 대해서, 세그먼트 또는 바운딩 박스의 주변 영역에 마진(margin)을 추가할 수 있다. 이때 크기 기준값들은 전처리 네트워크, 수신단, 또는 다중 임무 수행을 위한 머신 비전 네트워크 중 하나에 지정된 값일 수 있으며, 두 값을 동일하게 설정할 수도 있다. 사각형인 바운딩 박스의 주변 영역의 마진은 바운딩 박스의 크기를 상단, 하단, 좌측, 또는 우측 중 적어도 하나의 방향으로 특정 픽셀 단위만큼 확장함으로써 포함시킬 수 있다.

또한, 관심 객체 영역은 인접한 세그먼트 또는 바운딩 박스들과의 결합을 통해 하나의 큰 영역으로 대체될 수 있다. 결합되는 인접한 세그먼트 또는 바운딩 박스 영역들은 동일한 클래스에 해당하는 객체 영역으로 한정할 수 있다. 마지막으로, 관심 객체 영역을 제외한 배경 영역에 우세 칼라(dominant color), 평균 칼라(average color), 사용자 지정 칼라, 열화(degradation)시킨 배경 영상, 임의의 질감(texture)을 갖는 영상을 입혀 전송할 수 있다. 그림 3은 배경 영역에 우세 칼라(dominant color)를 사용하였다.

기계 임무(machine task) 수행 성능 평가를 위한 데이터셋으로는 텐센트(Tencent)가 제공하는 TVD(Tencent video dataset) 이미지 데이터셋을 사용했다. 그림 4는 TVD 이미지에 제안한 전처리 알고리즘을 적용한 후 수신단의 다중 임무 수행 머신 비전 네트워크 중 하나인 인스턴스 분할 네트워크를 적용해 객체를 분할한 결과이다.

그림 5는 관심 객체 추출 전처리 네트워크 적용시 인스턴스 분할 성능(mAP) 및 압축률(bpp) 그래프를 나타낸다.

능(mAP) 및 압축률(bpp) 그래프를 나타낸다. 인스턴스 분할 성능 평가에 사용된 Anchor 네트워크는 COCO2017로 훈련된 Mask R-CNN X101-FPN 네트워크이며, 압축은 VTM(Versatile Video Coding Test Model)-12를 사용하였다.

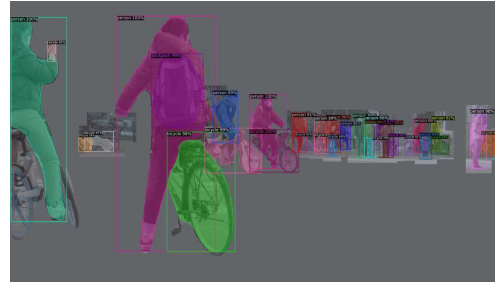


그림 4. 전처리 후 인스턴스 분할 결과

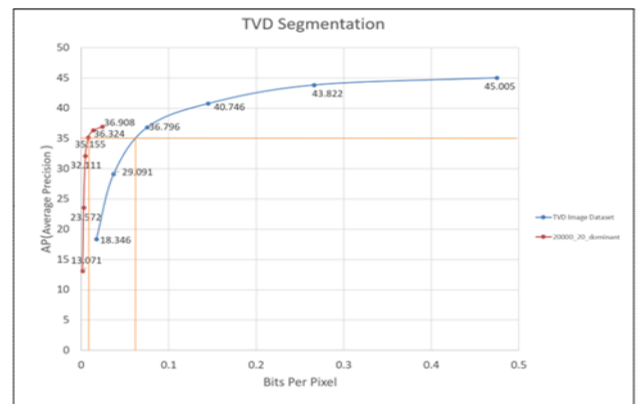


그림 5. 전처리 후 인스턴스 분할 성능(mAP) 및 압축률(bpp) 그래프

제안하는 관심 객체 기반 전처리 네트워크 구성 방법을 적용할 시 전체 영상을 전송할 때에 비해 높은 압축률(BPP, bits per pixel) 뿐 아니라, 정확도면에서도 동일 BPP에서 원본 영상 대비 높은 정확도를 보인다.

3. 결론

본 논문에서 서술한 관심 객체 기반 전처리 네트워크를 사용하면 배경 영역의 부호화를 최소화하여 부호화 효율을 높일 수 있다. 또한, 관심 객체의 정확한 윤곽을 이용하여 낮은 비트율에서도 고정밀 다중 임무 수행이 가능하다. 더불어, 서술자를 사용하면 기계 응용 프로그램에서 활용도와 재사용률을 높일 수 있다.

감사의 글

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, 기계를 위한 영상 부호화 기술 개발).

참고문헌

[1] S.Kim, M.H.Jeong, H.Jin, J.Y.Lee, H.Lee, and W.Cheong, "[VCM] Use case of VCM for Object of

- Interest", ISO/IEC JTC 1/SC 29/WG2 input document, m57487, July 2021, Online.
- [2] S.Kim, M.H.Jeong, J.Y.Lee, H.Lee, and W.Cheong, "[VCM]Preliminary experiment results with objects of interest on pose estimation", ISO/IEC JTC 1/SC 29/WG2 input document, m57998, October 2021, Online.
- [3] J.Y.Lee, H.Lee, W.Cheong, and J.Seo, "[VCM] Descriptor based VCM for multi-task", ISO/IEC JTC 1/SC 29/WG2 input document, m57412, July 2021, Online.
- [4] H.Lee, J.Y.Lee, H.Choo, W.Cheong, and J.Seo, "[VCM] Object of interest based VCM for multi-task", ISO/IEC JTC 1/SC 29/WG2 input document, m58846, January 2022, Online.
- [5] W.Gao, Z.Liu, X.Xu, and S.Liu, "[VCM] Updated annotations and anchor results for TVD object tracking dataset", ISO/IEC JTC 1/SC 29/WG2 input document, m59675, April 2022, Online.