

IVCL

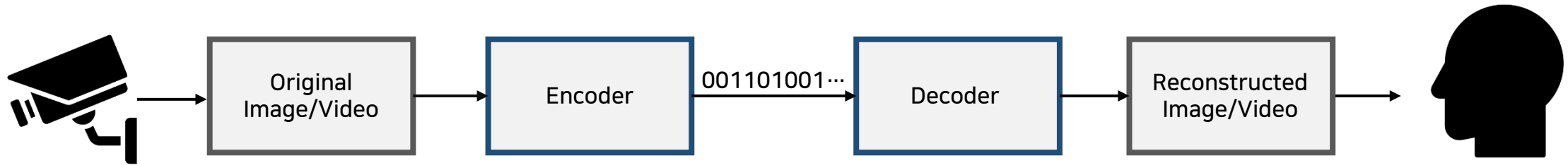
Video Coding for Machines

백전능



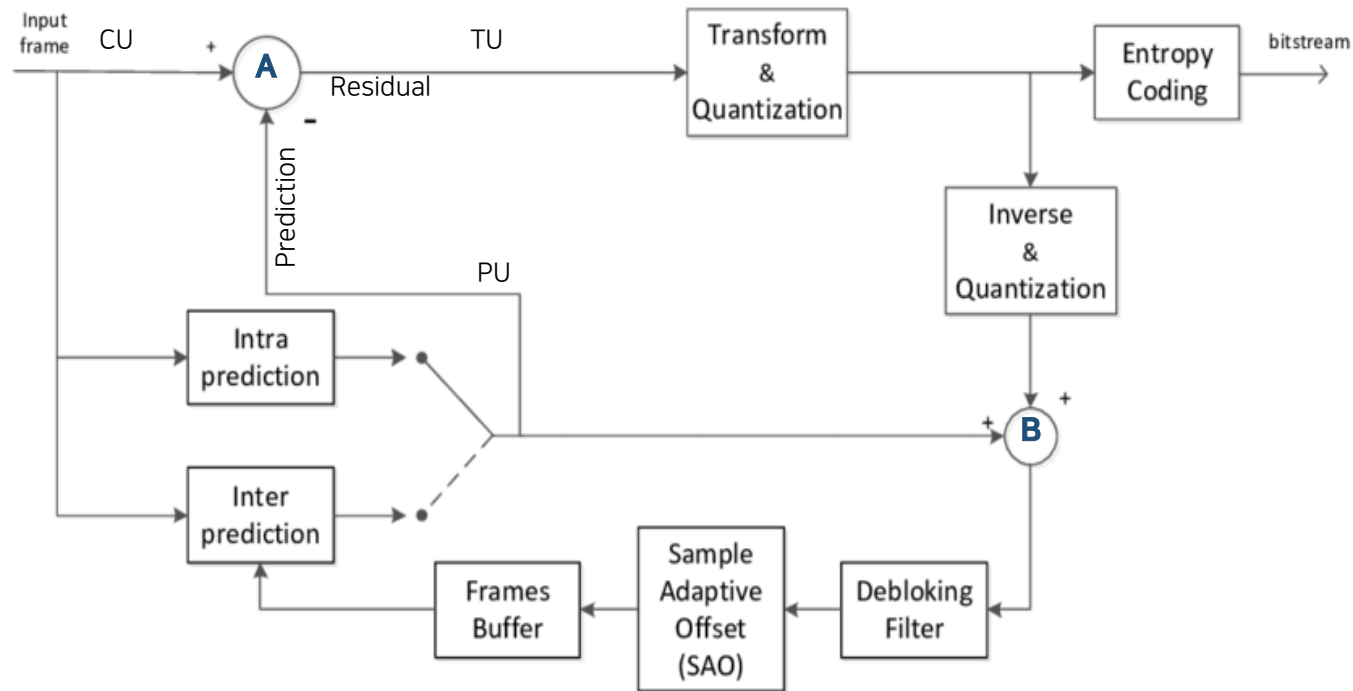
Video Coding for Human

- 기존에 존재하는 전송 방식
- 영상에 대한 압축 진행 후 영상으로 복원
- 화질과 같은 인간 시각적인 관점에서의 성능이 중요



Video Coding for Human

- 기존의 Video Coding 파이프라인



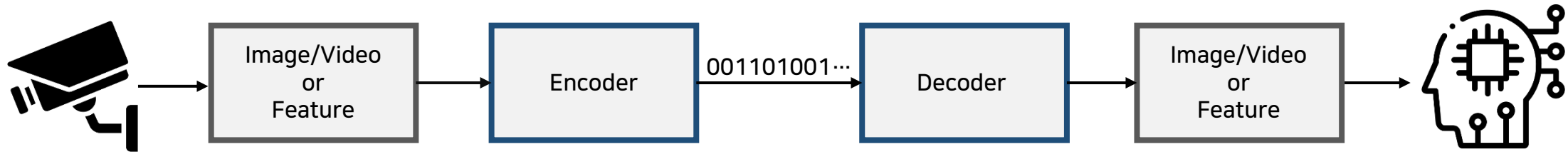
Necessity of VCM

- 영상 데이터의 기하급수적 증가
- 자율주행차량, 스마트팩토리 등의 무인 시스템에 대한 활발한 연구, 개발
- 무인 시스템에서 이미지, 비디오의 중요성
- 사람이 아닌 기계 관점에서의 전송 방식 필요



Video Coding for Machines

- 영상 또는 영상에서 추출된 특징(feature)을 압축 및 복원 후 머신 비전 태스크에 활용
- 화질이 아닌 머신 비전 태스크의 성능이 중요

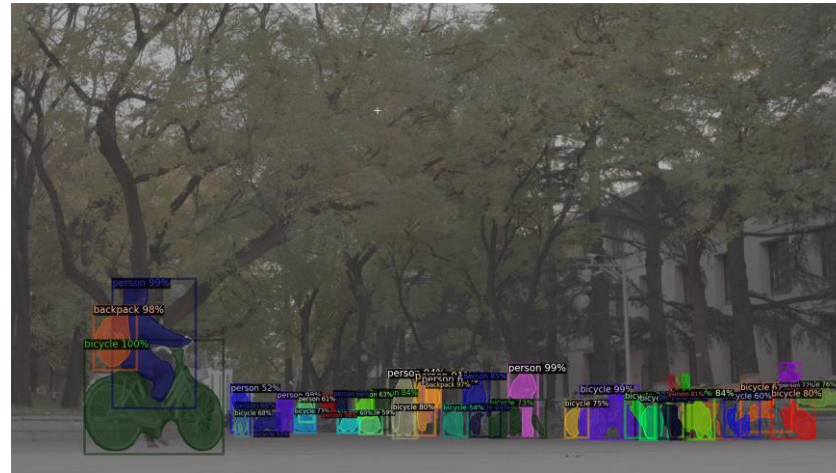


Machine Vision Task

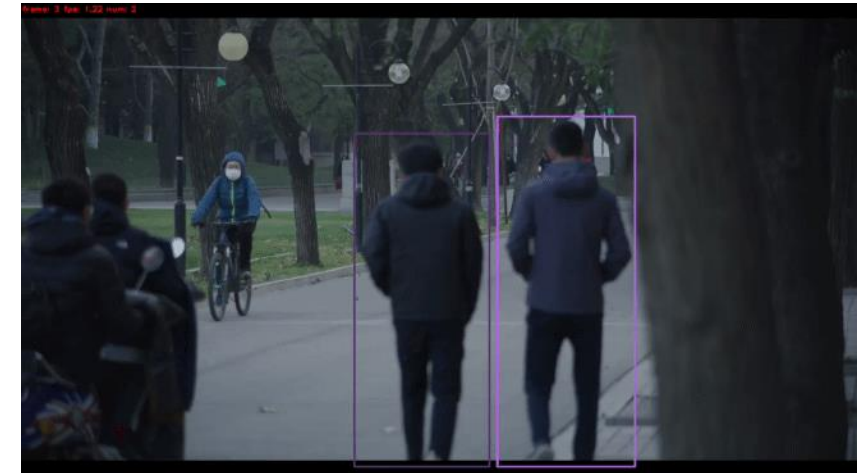
Object Detection



Object Segmentation

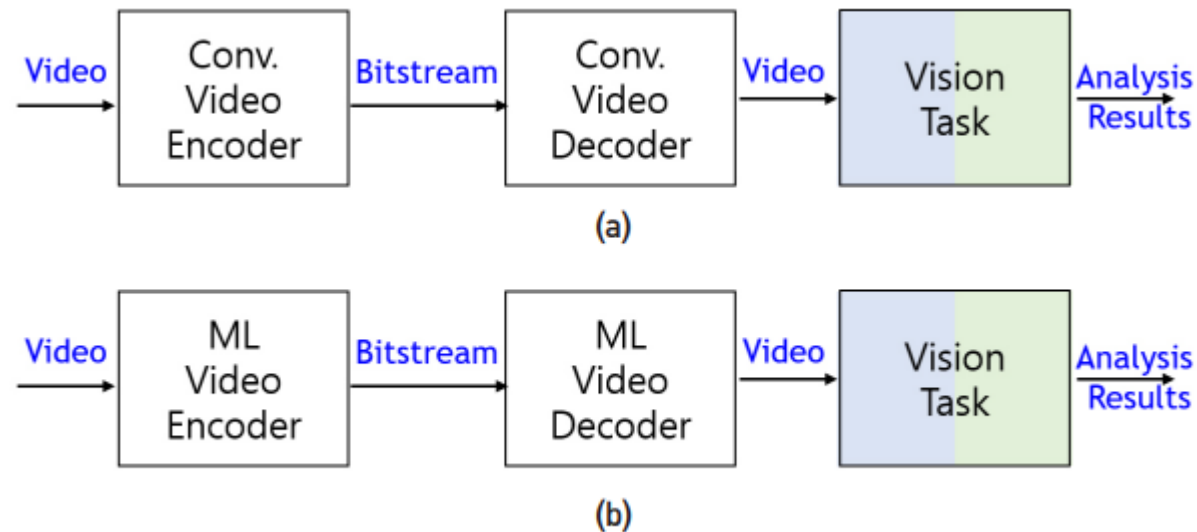


Object Tracking



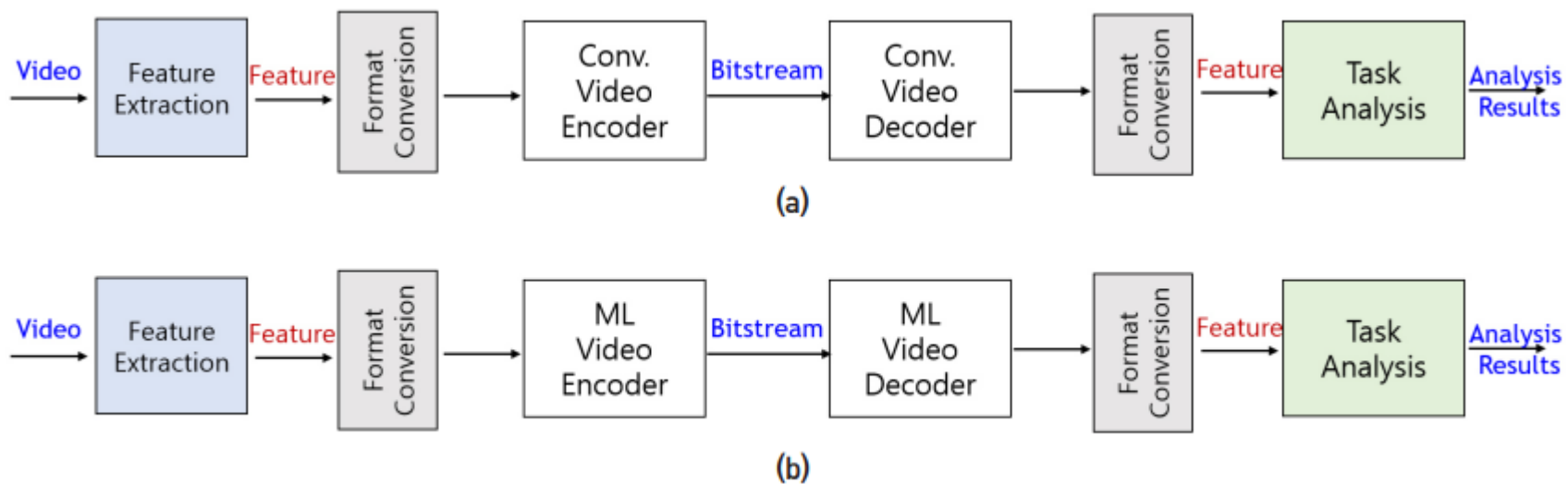
Video Coding for Machines

- 기존의 영상 전송 방식 활용 후 마지막에 머신 비전 태스크 시행



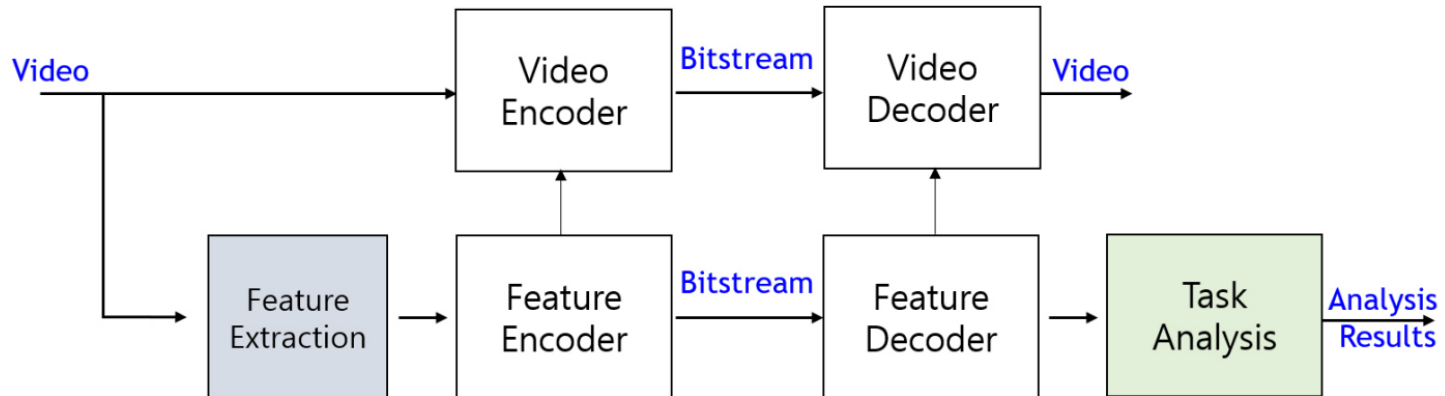
Video Coding for Machines

- 인공지능 네트워크 중간 부분에서 추출된 Feature 압축 및 복원.
- 복원된 Feature를 네트워크 중간 부분에 입력하여 머신 비전 태스크 수행.



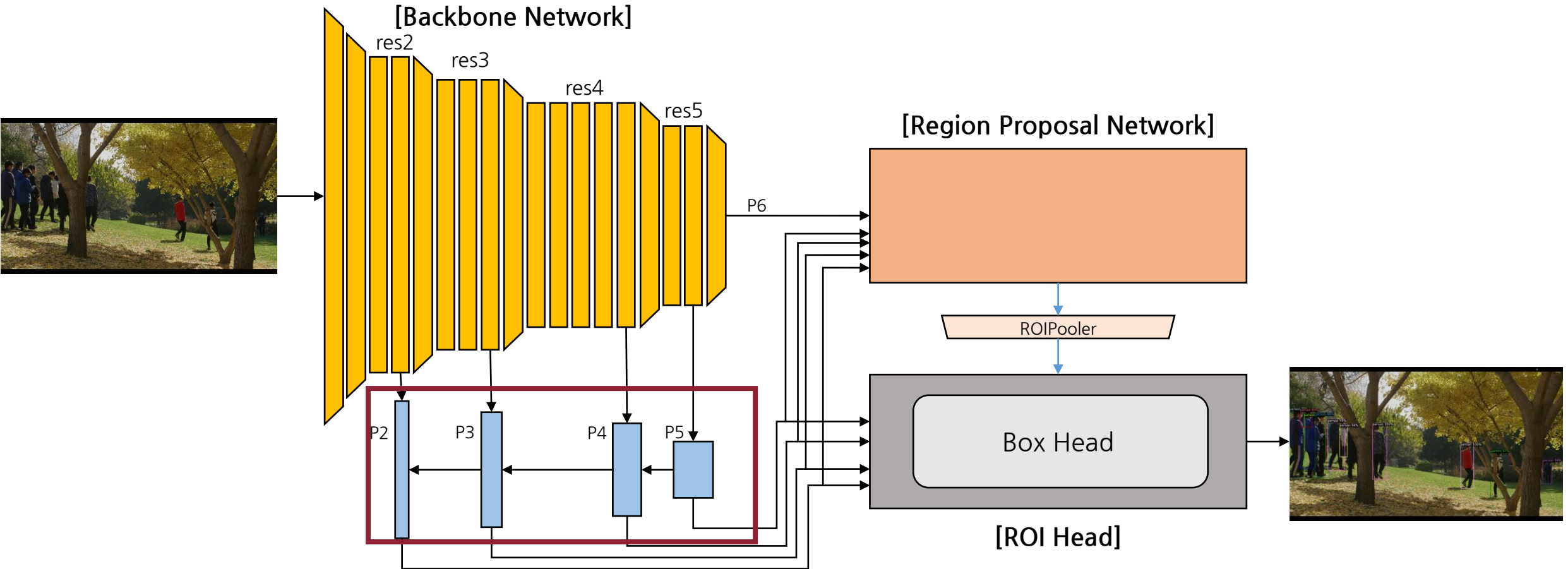
Video Coding for Machines

- 영상과 Feature에 대한 압축 및 복원을 병행하는 하이브리드 방식.
- 머신 비전 태스크에 오류가 생겼을 경우 사람이 감독 및 관리하기 위함.



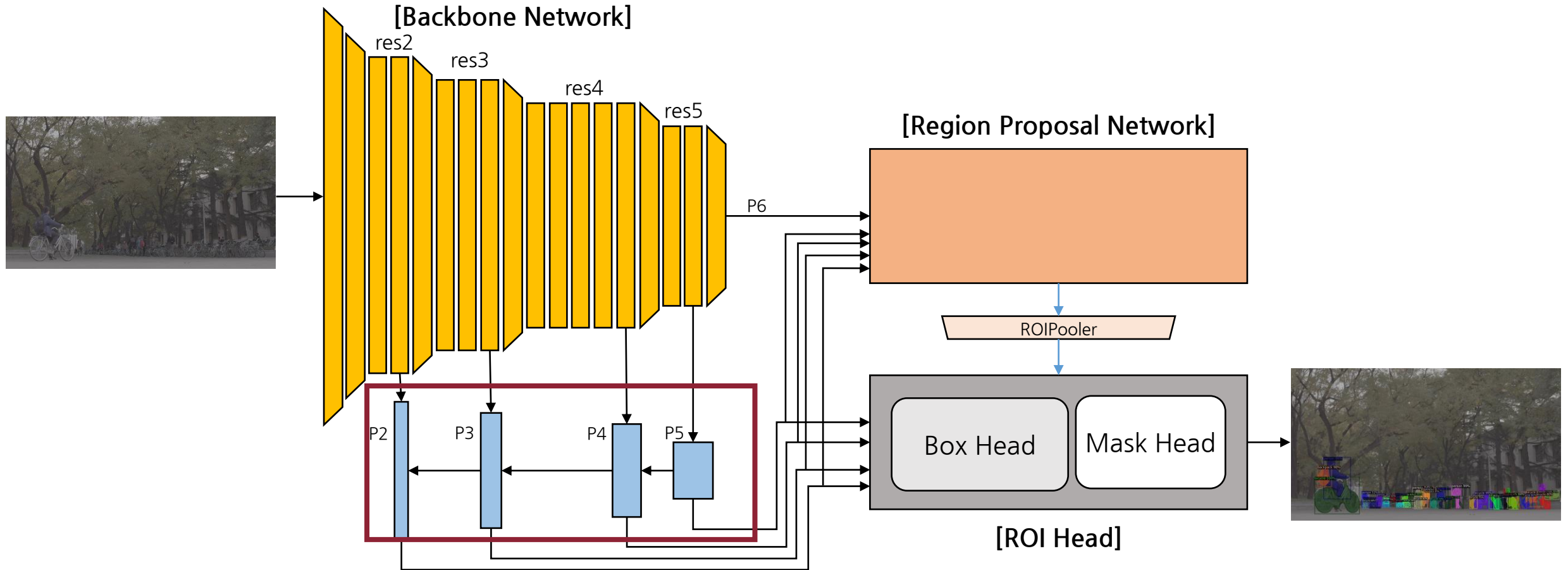
Feature Extraction

- Object Detection : Faster R-CNN



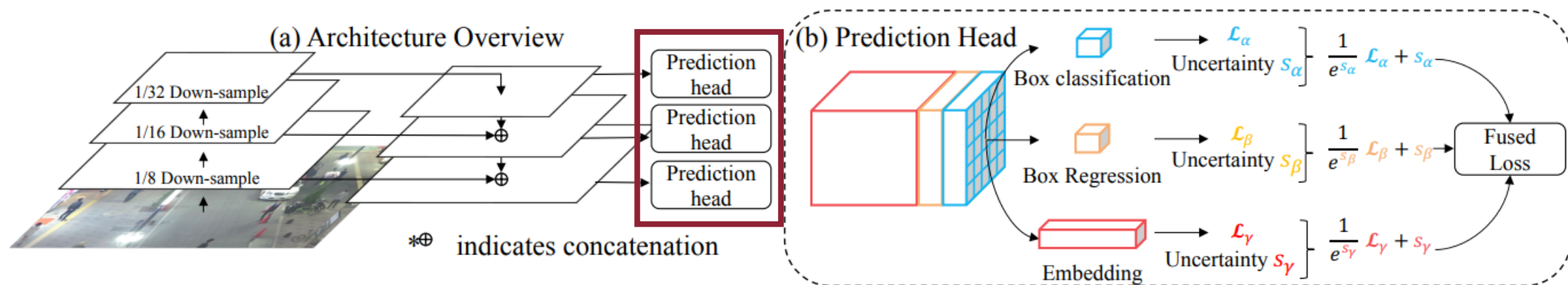
Feature Extraction

- Object Segmentation : Mask R-CNN



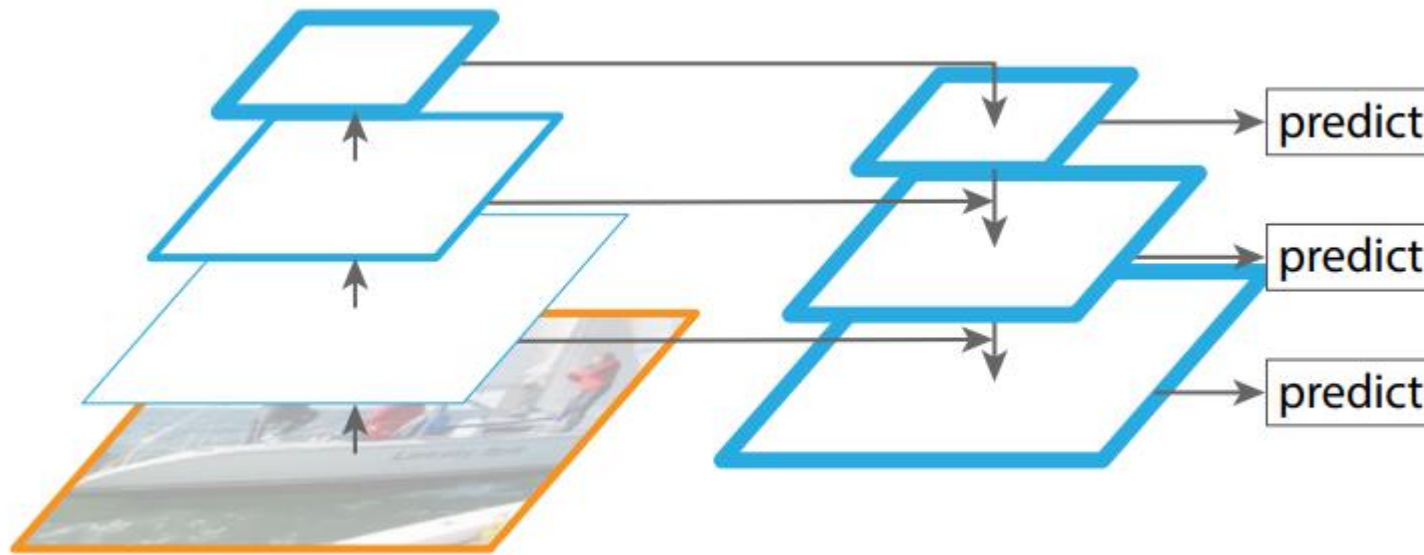
Feature Extraction

- Object Tracking : JDE



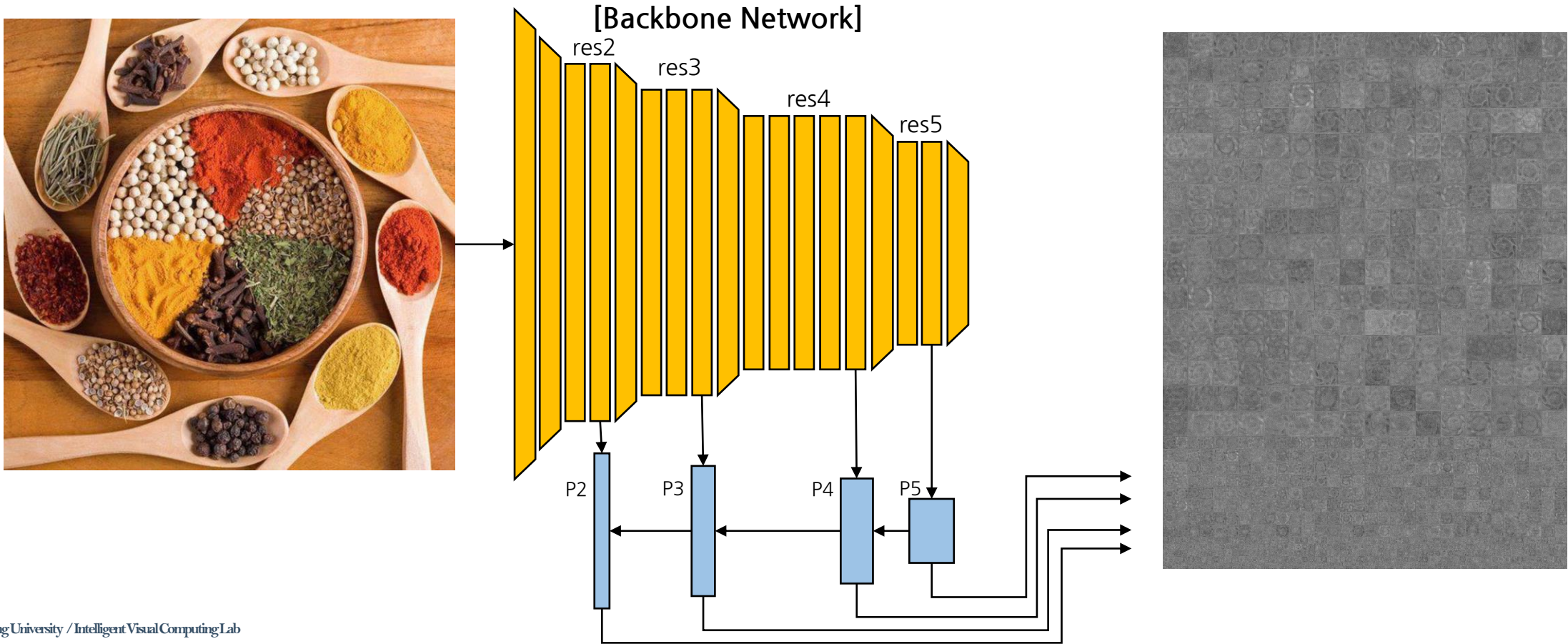
Feature Extraction

- FPN(feature pyramid network)



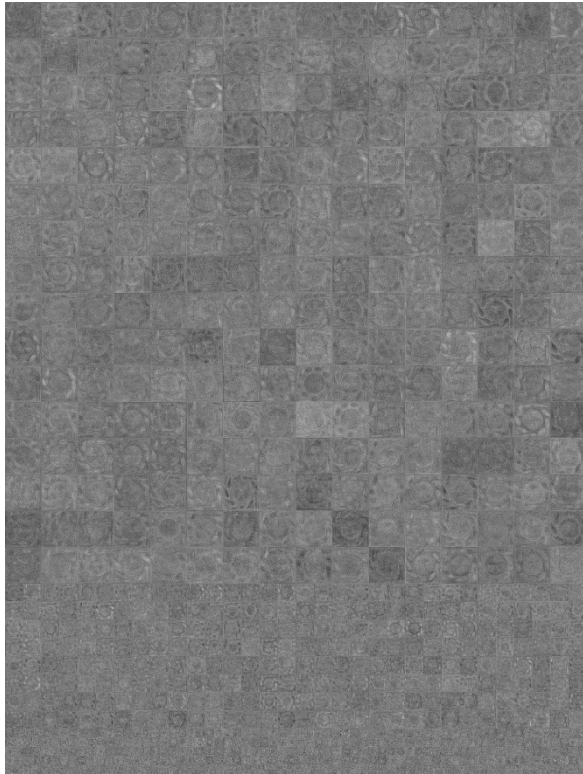
Feature Map

- Backbone Network를 활용해 이미지의 Feature 추출
- 1차원 텐서 형태의 Feature를 2차원 이미지 형태의 Feature Map으로 변환



Feature Map Coding

- Feature Map을 비트스트림으로 압축 후 전송



VVC Encoder



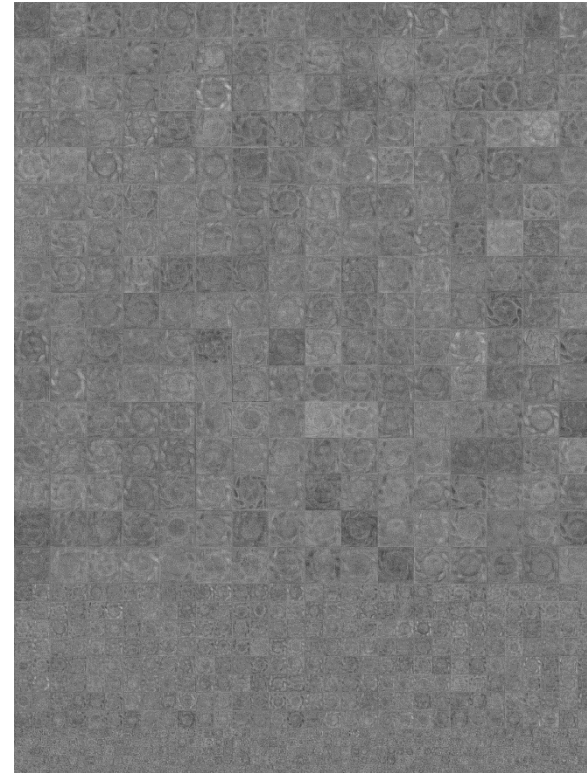
```
0010001111001100  
0000011101010000  
01110001111001...
```

Feature Map Coding

- 압축되어 전송된 비트스트림을 2차원 Feature Map으로 복원

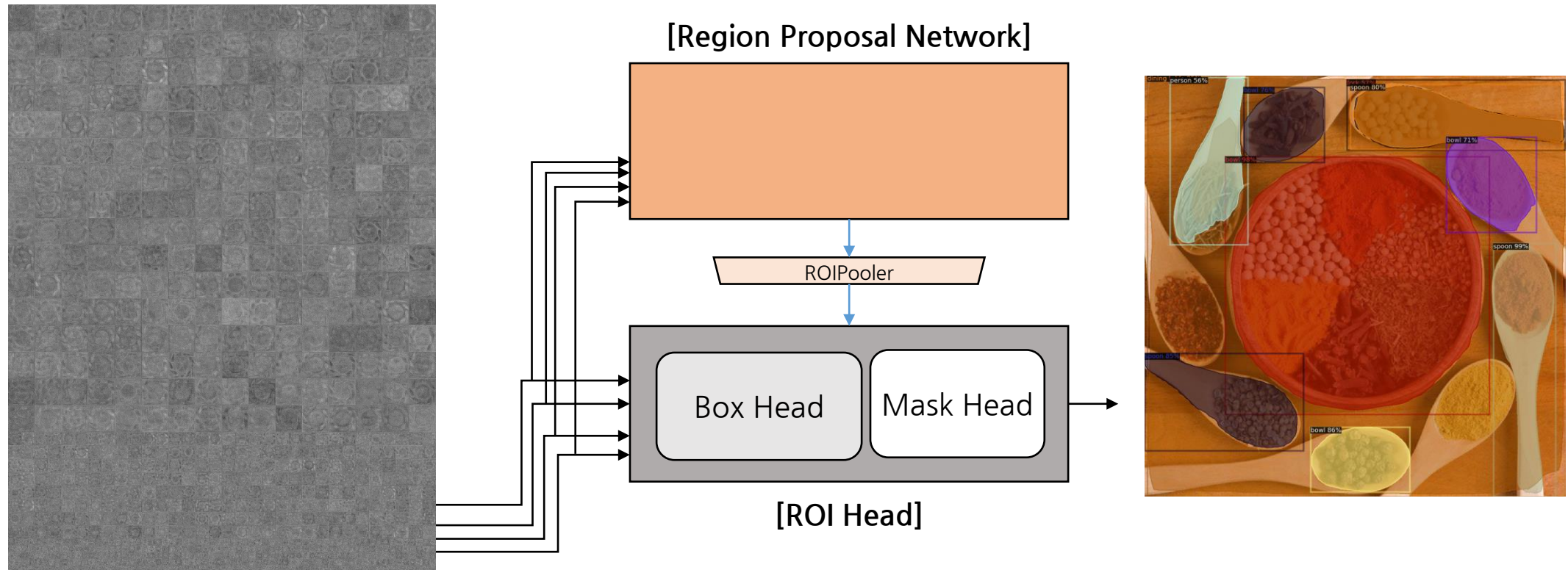
0010001111001100
0000011101010000
01110001111001...

VVC Decoder



Feature Map Coding

- Feature Map을 다시 1차원 텐서 형태의 Feature로 변환 후 네트워크 중간에 입력



Summary

- 자율주행차량, 스마트 팩토리, 스마트 시티 등 기계의 영상 활용 증가
- 기존 인간 시각을 위한 영상 전송이 아닌 머신 비전 태스크를 위한 영상 전송 방식이 대두됨
- 화질이 아닌 비전 태스크의 성능을 우선시
- 영상 압축, Feature 압축 등의 방법론들을 논의 및 연구 중

References

- 권형진, 정세윤, 최진수, 이태진, and 서정일, "Standardization Trends in Video Coding for Machines," *Electronics and Telecommunications Trends*, vol. 35, no. 5, pp. 102–111, Oct. 2020.
- S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.
- K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.
- Wang, Zhongdao, et al. "Towards real-time multi-object tracking." *European Conference on Computer Vision*. Springer, Cham, 2020.
- T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.
- 고효율 영상 부호화 기술 HEVC 표준기술의 이해 (2015)
- Tencent Video Dataset
- Open Images Dataset V6

Thank you for listening.

