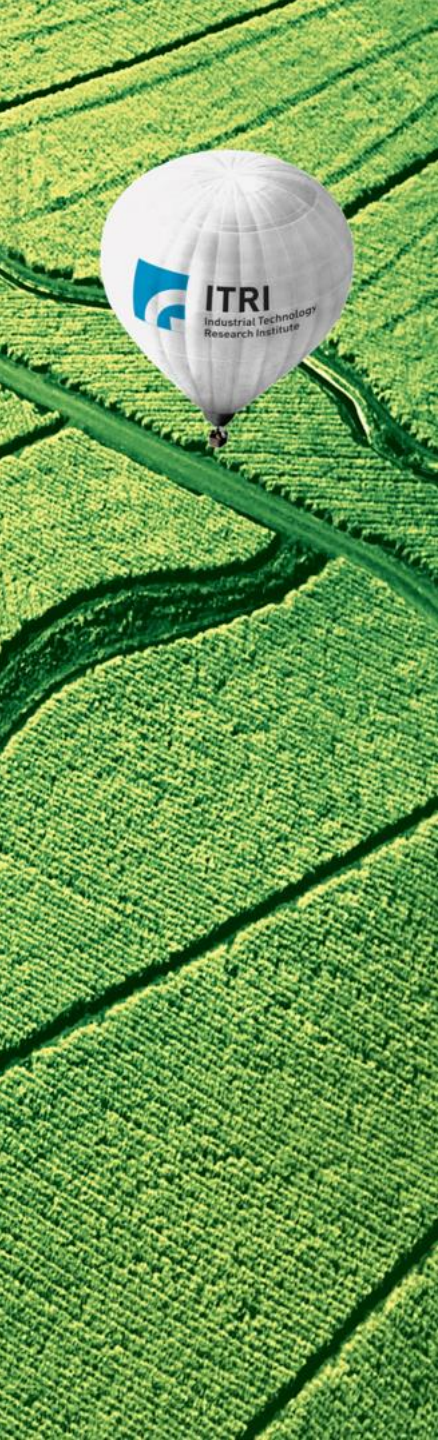
ITRI
Industrial Technology
Research Institute

MPEG國際會議分享



ITRI ICL V200
Ching-Chieh Lin

27th Nov. 2020

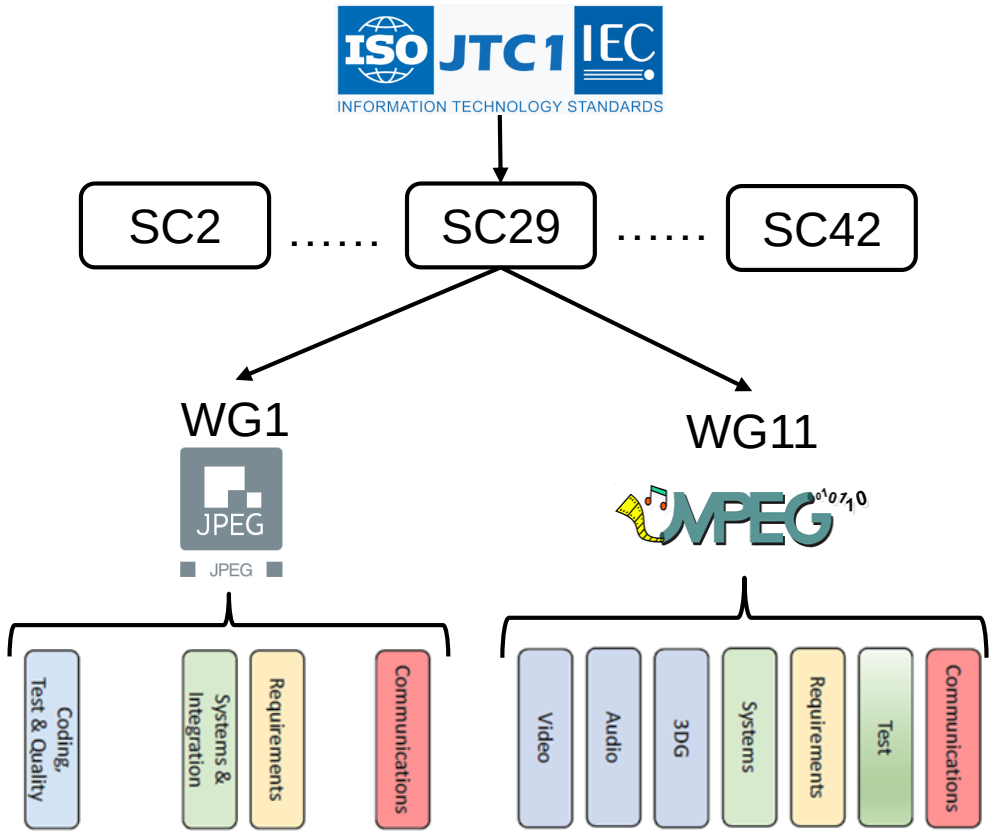
Outline

- News from MPEG
- Joint Video Expert Team(VVC/H.266)
  - Timeline and future works
- Video coding for Machine
  - Introduction and use cases

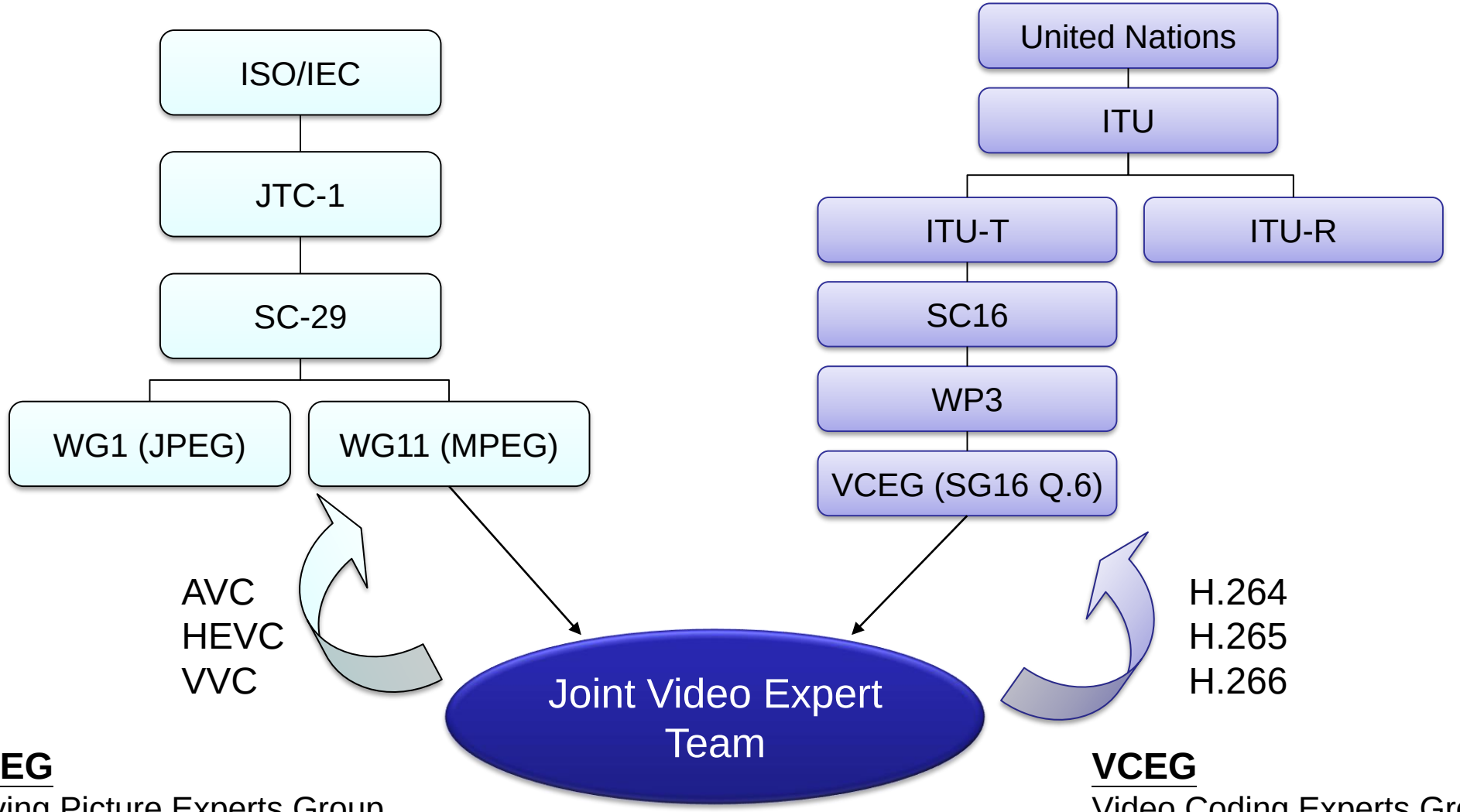
ITRI
Industrial Technology Research Institute

MPEG的組織架構

- MPEG(Moving Picture Experts Group)
  - ISO/IEC Joint Technical Committee 1, Subcommittee 29, Working Group 11– Coding of moving pictures and audio
- MPEG成立於1988年，為一源自ISO與IEC等國際組織的工作小組，用以制定影音壓縮及傳輸的規格標準。
- MPEG 每年舉辦4次例行性會議，平均參與專家人數約400人，來自20個國家與200家公司或研究單位。



MPEG and ITU-T

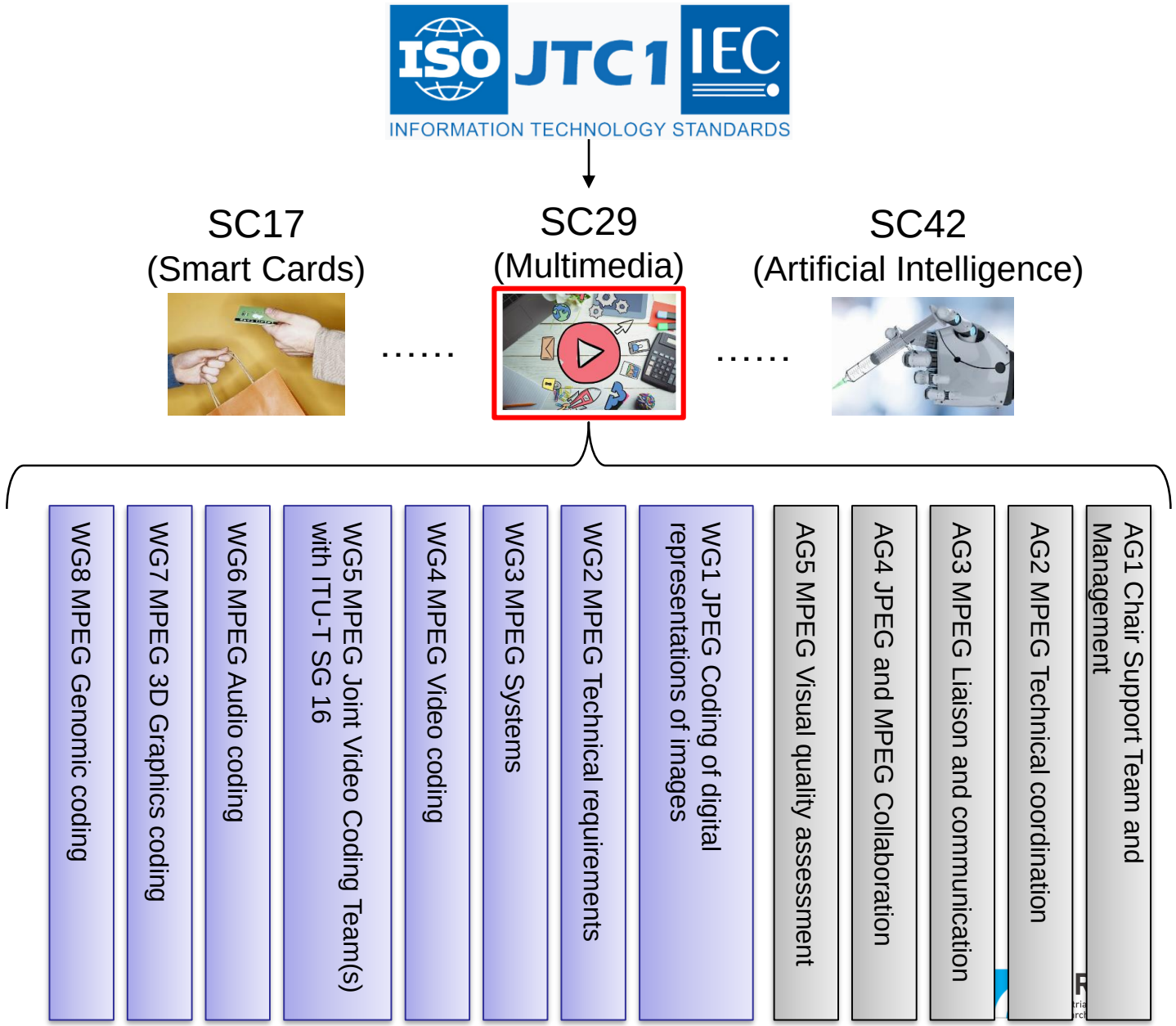


MPEG
Moving Picture Experts Group

VCEG
Video Coding Experts Group

MPEG的組織變革

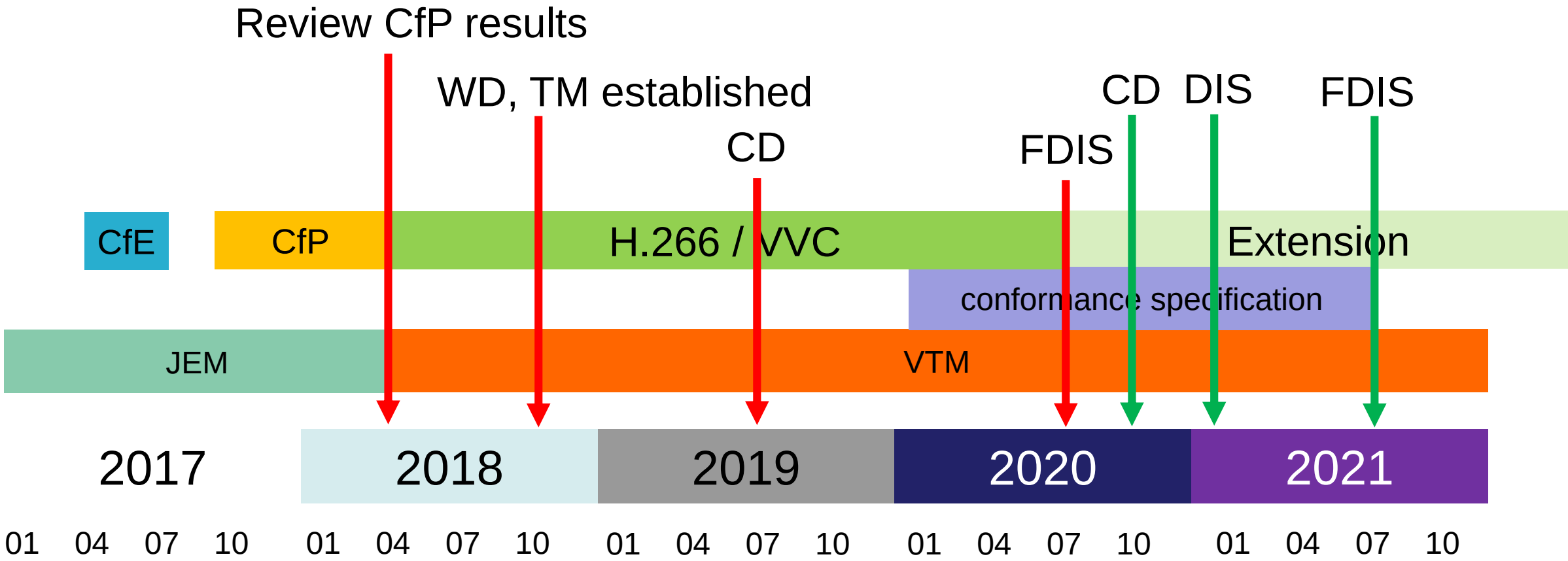
- SC29重整WG1與WG11轄下的11個Sub-group
- 2020年10月會期開始，建立5個Advisory group與8個working group
  - Working group 針對特定主題進行技術討論與標準制定工作
  - Advisory group 負責組織管理、聯繫與共通性議題的處理



Joint Video Expert Team

H.266 Timeline

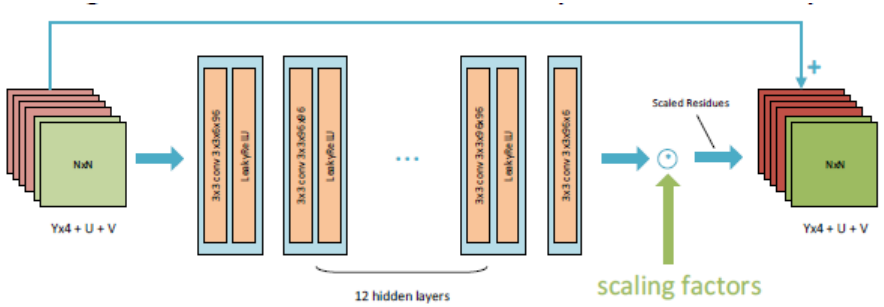
*WD : Working Draft
*TM : Test Model
*CD : Committee Draft
*FDIS: Final Draft International Standard



Future of VVC

- Extend set of SEI messages
  - Digital signature
  - Post-processing control
- High bit depth, high bit rate, high frame rate coding
  - Definition of test sequences and test conditions
  - Possible extensions in transforms and entropy coding
- Neural-network-based video coding
  - Definition of testing conditions and reporting template for unique analysis of complexity and performance, and training conditions
  - Two approaches to be considered
    - Hybrid approach
    - End-to-end approach

Source: Epic Games



Source: JVET-T0079, AHG11: Neural Network-based In-Loop Filter, H. Wang, M. Karczewicz, J. Chen, A.M. Kotra (Qualcomm)

ITRI Industrial Technology Research Institute

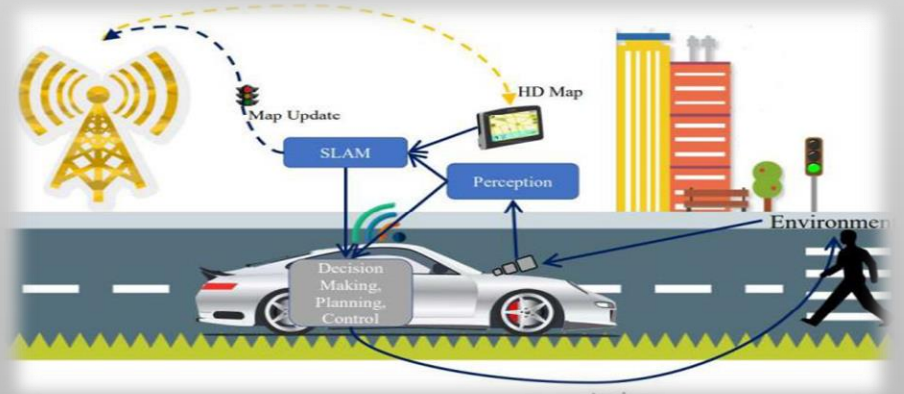
Video Coding for Machine(VCM)

ITRI
Industrial Technology Research Institute

Video Coding for Machine

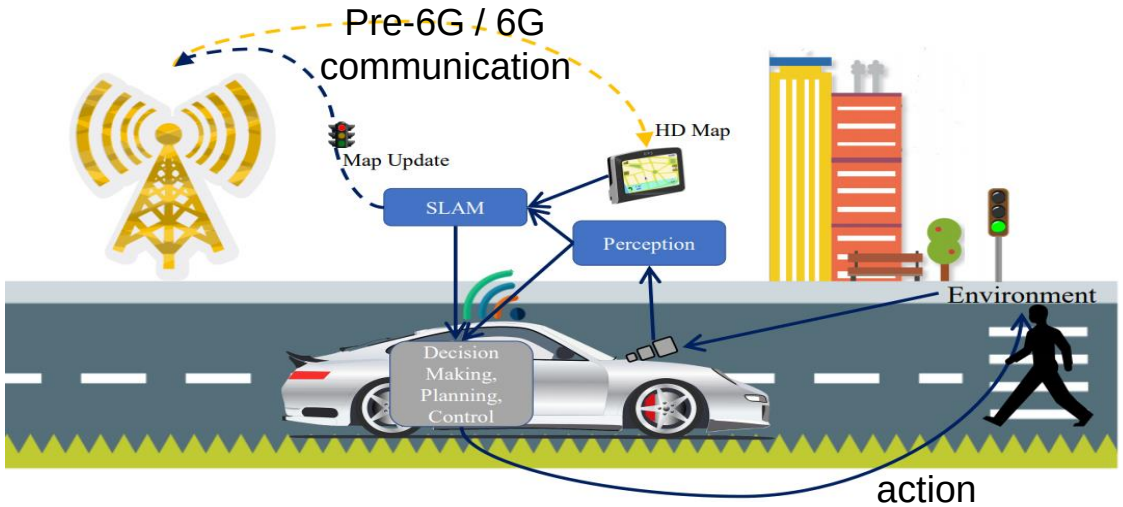
- Traditional coding methods aim for the best video under certain bit-rate constraint for human consumption.

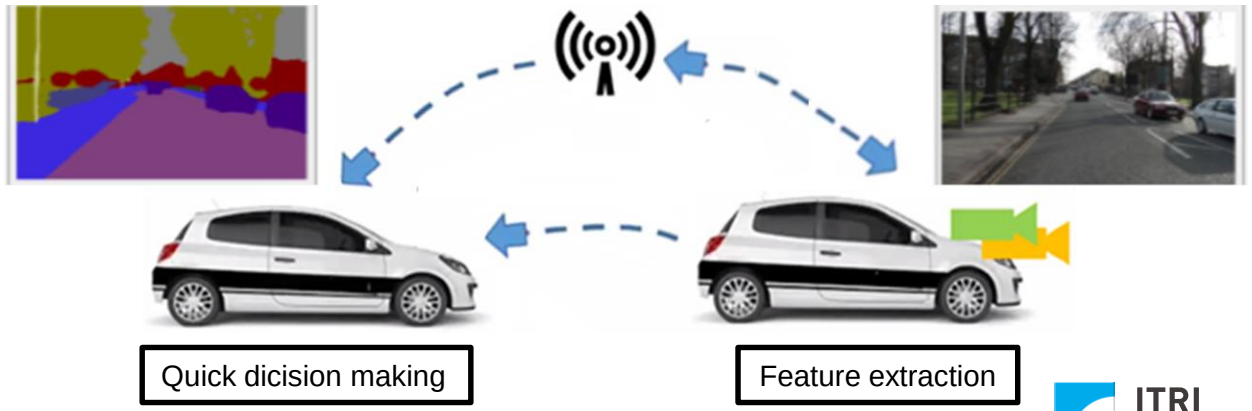
- Machines will communicate amongst themselves to perform tasks without a human.
- Transmission and archive systems require a more compact data representation and low latency solution for machine vision.



VCM for intelligent transportation

- With the rise of IoT, there are large amount of interconnectivities between different node sensors and devices.
- Machine vision tasks could be split into the feature extraction and decision making implemented in the front-end devices and computing servers respectively.
- The separated architecture helps release the workload of computing server and fulfill applications requiring large amount of real-time computing (e.g. V2V/V2X).





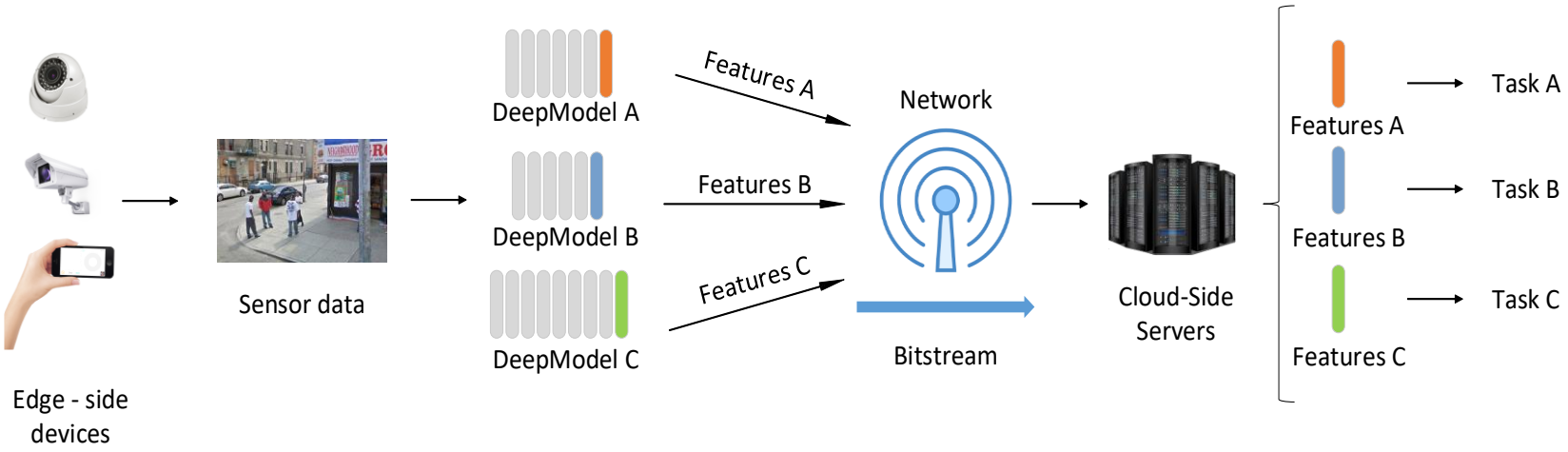
ITRI
Industrial Technology
Research Institute

Source: w19506 Use cases and requirements for Video Coding for Machines, ISO/JCT SC29 WG11 Online July 2020

VCM for intelligent surveillance and smart city

- Current surveillance systems take up large amounts of data due to the number of sensors and length of video.
- Smart City applications encompass use cases such as traffic monitoring, density detection / prediction and traffic flow prediction.

Key tasks:
- Multi-object detection
- Object segmentation
- Object Tracking
- Activity recognition
- Event prediction
- Abnormal recognition
- Super resolution
- Privacy protection

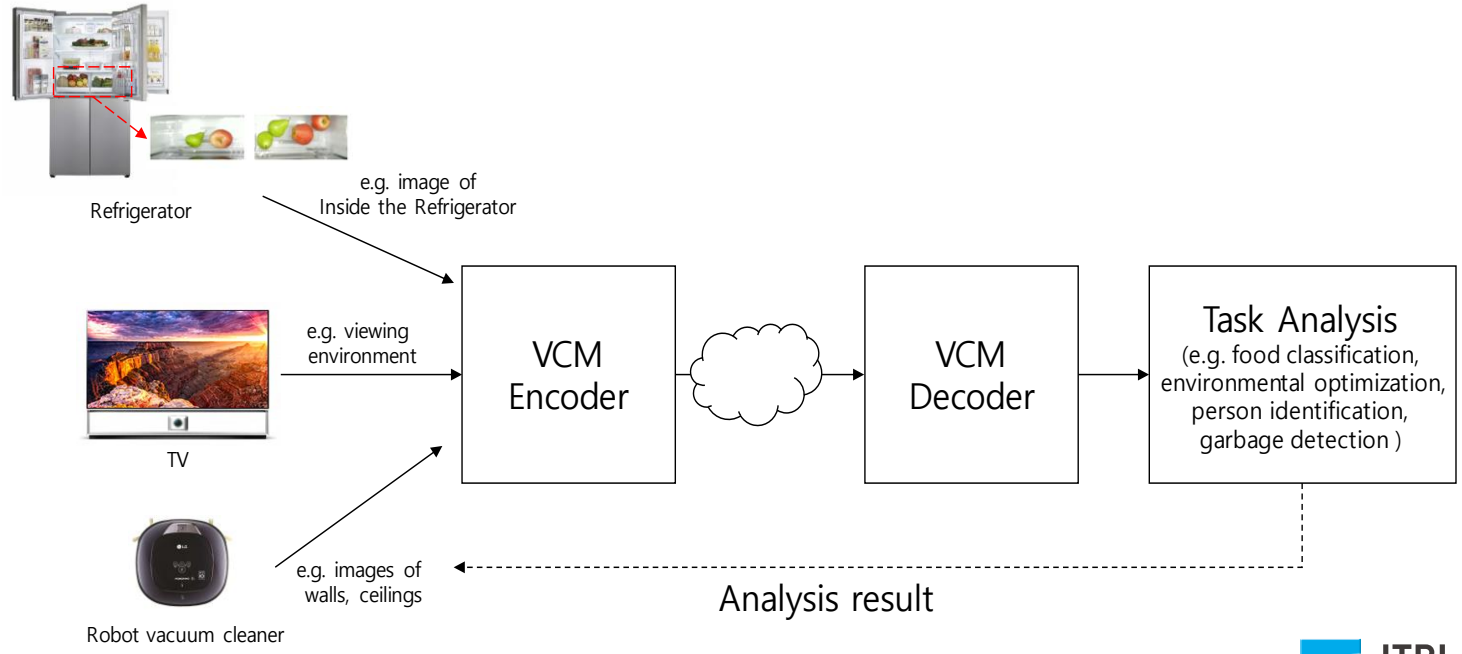


Source: m52162 CNN Feature Map Coding for Cloud-based Visual Analysis
m54260 [VCM] Comments on VCM Use cases and Requirements

VCM for consumer electronics

- VCM enables the consumer electronic devices to provide context-aware services and information to users with neural networks.
- VCM can provide interoperability between devices from different vendors.
- Communication under low bandwidth and personal information protection are desirable.

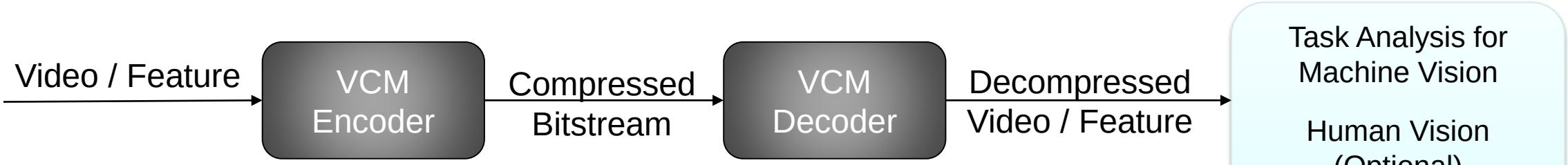
Key tasks:
- Object detection
- Instance segmentation
- Event detection, recognition and prediction
- Image search, reconstruction and enhancement
- Pose estimation and tracking
- Action Recognition
- Anomaly detection



ITRI Industrial Technology Research Institute

Standardization of VCM

- Initiated from July 2020, averaged 100 participants from 30+ Affiliations
- Current status:
  - Issued first Call for Evidence at October 2020
- VCM shall be able to:
  - Efficiently compress the bitstream.
  - Use a common bitstream to support single or multiple tasks.
  - Support varying performance for multiple tasks as measured by the appropriate metrics.

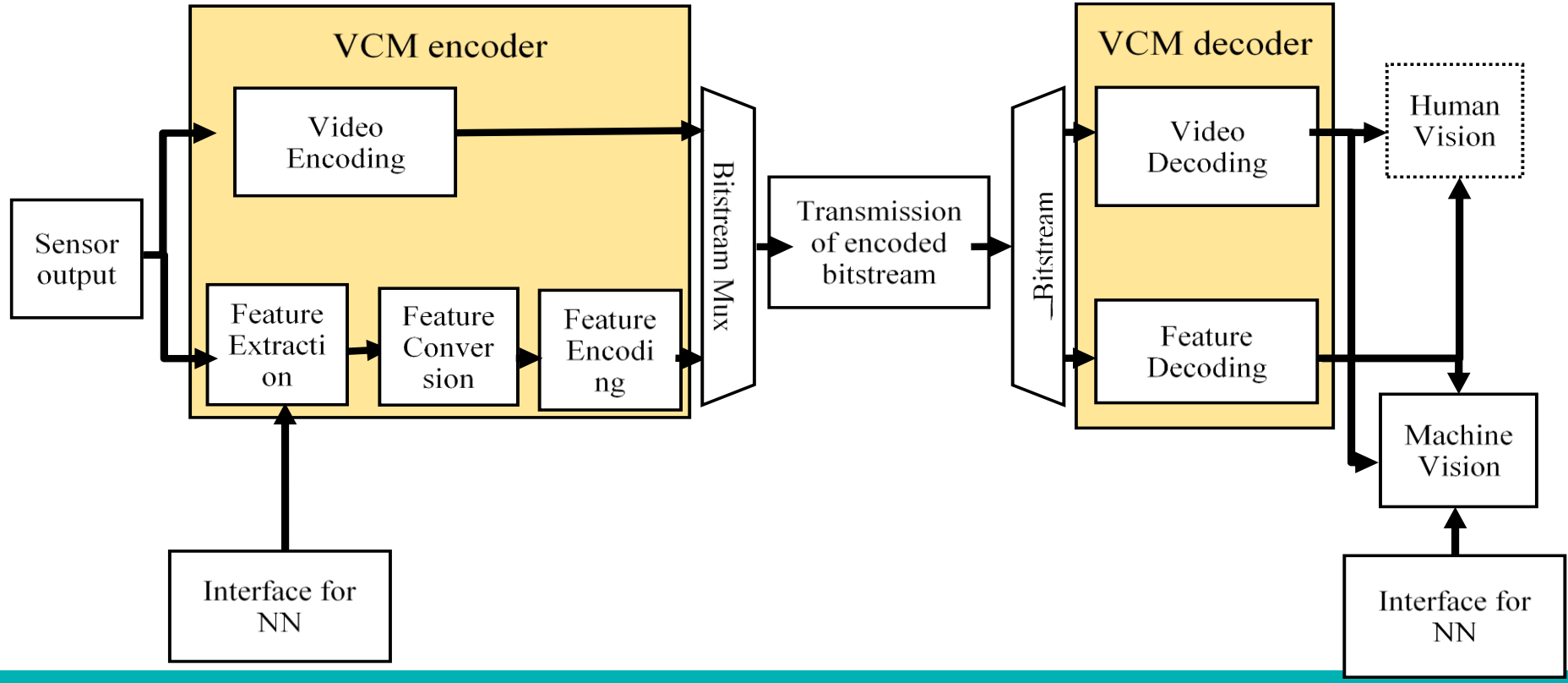


Pipeline for VCM

Source: m55604 Draft of Evaluation Framework for Video Coding for Machines

Potential VCM architecture

- The VCM codec could be video codec, feature codec, or both.
- Machine vision tasks could be split into two stages and being implemented in the encoder side and decoder side respectively.

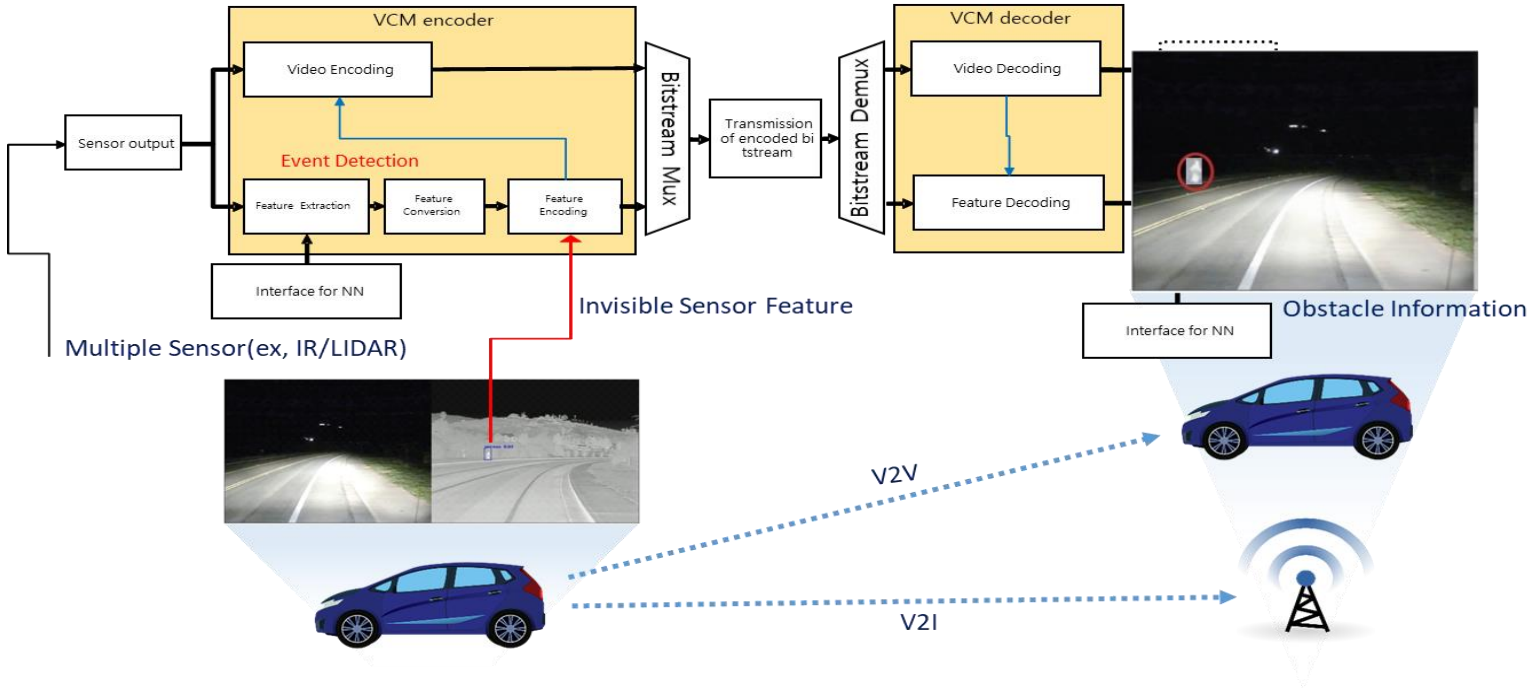


Source: w19506 Use cases and requirements for Video Coding for Machines, ISO/JCT SC29 WG11 Online July 2020

A typical use case for VCM

- VCM is an enabling technology for Autonomous Driving.
- Generative adversarial networks are trained to confront the Turing Test.

Key tasks:
- Multi-object detection
- Object segmentation
- Object (Lane) Tracking
- Activity recognition
- Event prediction
- Optical Flow



Anchor Delimitation and Requirements

- Training and test conditions, key metrics, datasets, benchmarks for various tasks

Task	Metrics	Datasets	Benchmarks	Training/Testing
Object Detection	mAP vs BPP/Rate	COCO [compressed] (image)	http://cocodataset.org/#detection-leaderboard	For COCO, use 2014/2017 Val set for evaluation and 2014/2017 Train in the case of retraining.
		CityScapes [uncompressed](image) CityPersons[uncompressed](image)	https://www.cityscapes-dataset.com/benchmarks/	For CityScapes, use defined train and validation sets
		ImageNet [compressed](image)	https://kobiso.github.io/Computer-Vision-Leaderboard/imagenet.html	For Imagenet, use the training and validation data as published from ILSVRC 2014.
Object Segmentation	mAP vs BPP/Rate	COCO[compressed](image)	http://cocodataset.org/#detection-leaderboard	(see above)
		CityScapes [uncompressed](image)		(see above)
		KITTI(image)	http://www.cvlibs.net/datasets/kitti/eval_object.php	We recommend using the predefined splits.
	mAP vs Rate	DAVIS 2016 / 2017(video)	https://davischallenge.org/	We recommend using the semi-supervised mode for higher accuracy.
Object Tracking	MOTA vs Rate	MOT20[compressed](video)	https://arxiv.org/pdf/1906.04567.pdf	Dataset split is available from the Tracking Challenge, available on their website.

Thank You

INNOVATING A BETTER FUTURE!

ITRI
Industrial Technology Research Institute