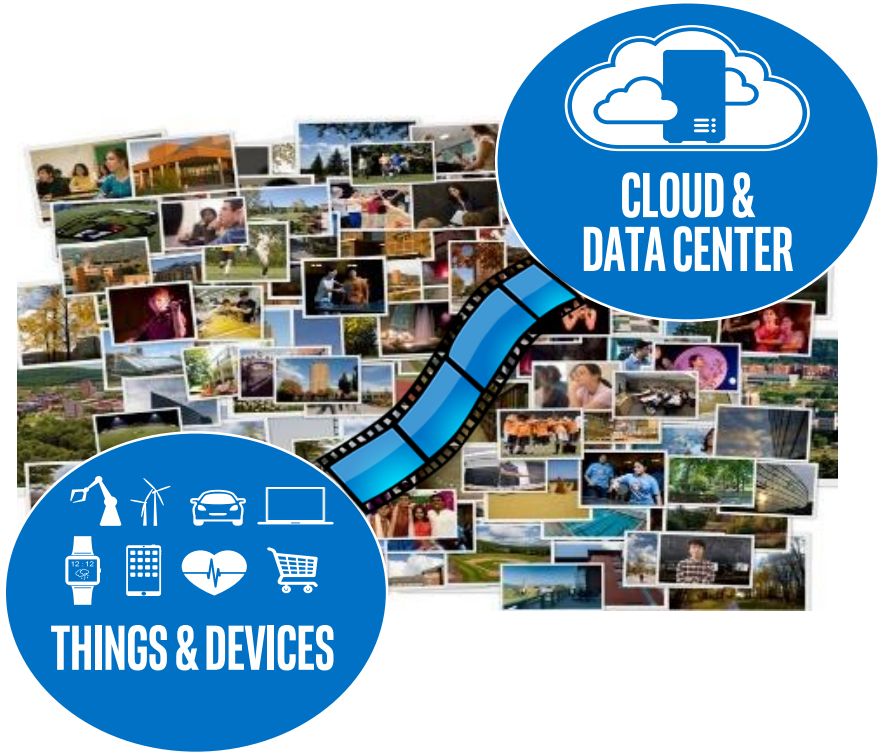
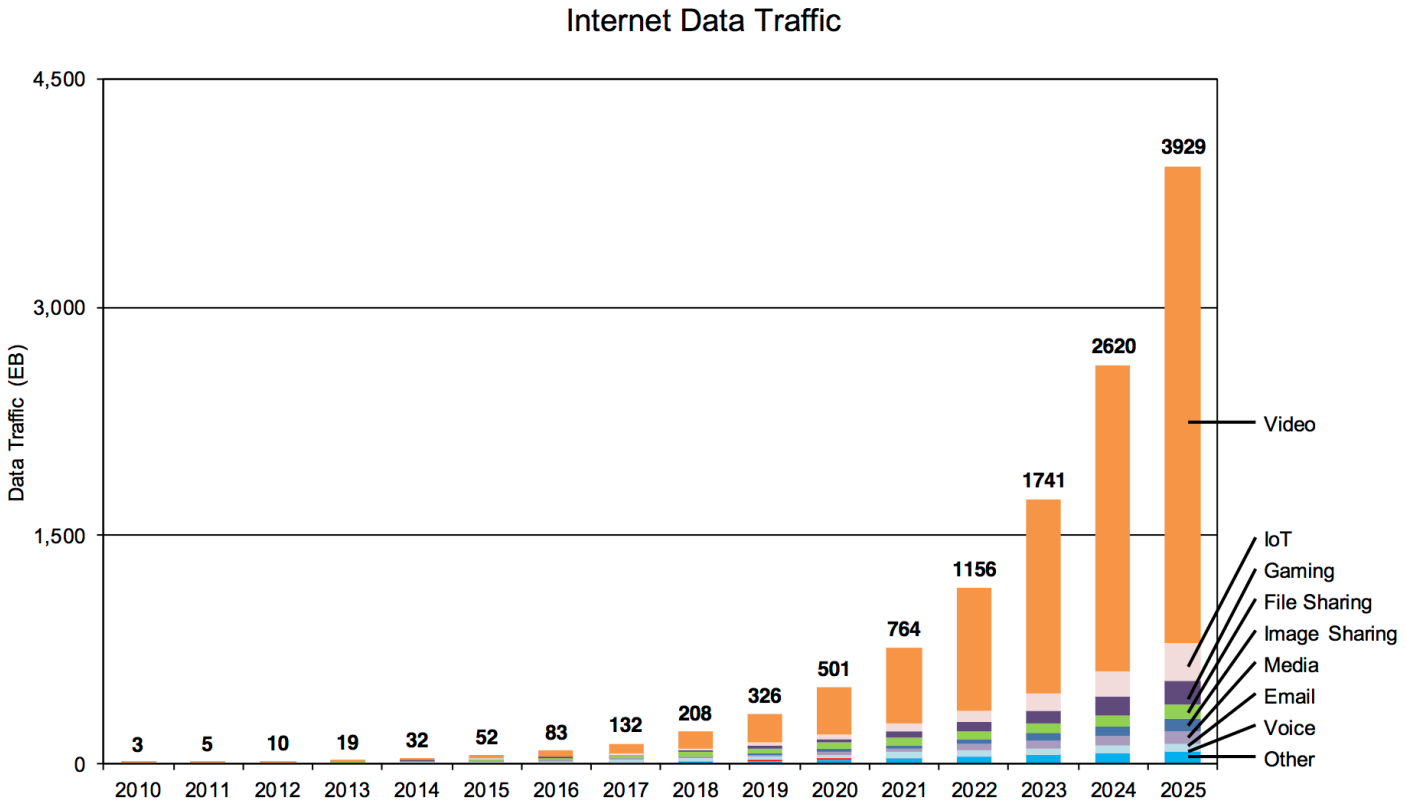
MAKING SENSE OF VISUAL DATA
理解视觉数据

陈玉荣 博士
英特尔首席研究员/高级研究总监
英特尔中国研究院认知计算实验室主任
2018年4月13日

视觉数据爆炸





Source: IBS Global System IC Industry Service Report, IoT Technology and Market Analysis, June 2015

终端和云端上绝大多数数据都是视觉的!!

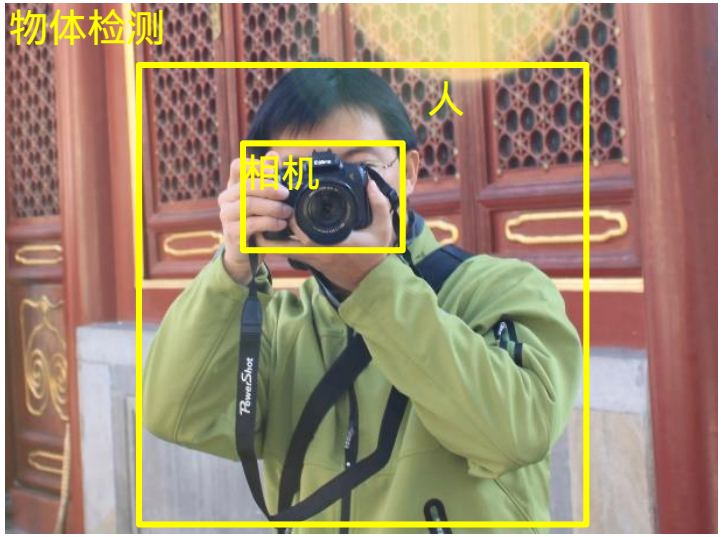
关键挑战：如果处理和理解这些视觉数据？

视觉理解 – 是什么?

计算机视觉领域包括各种用于获取、处理、**分析和理解图像**以及来自现实世界的高维数据，以生成数字或符号信息的方法 Wikipedia.org

功能列举	
人脸检测和识别	情感识别
自然环境下的文本识别	物体分类/检测
行为分类/识别	场景分类/理解
视频分类/摘要	•••

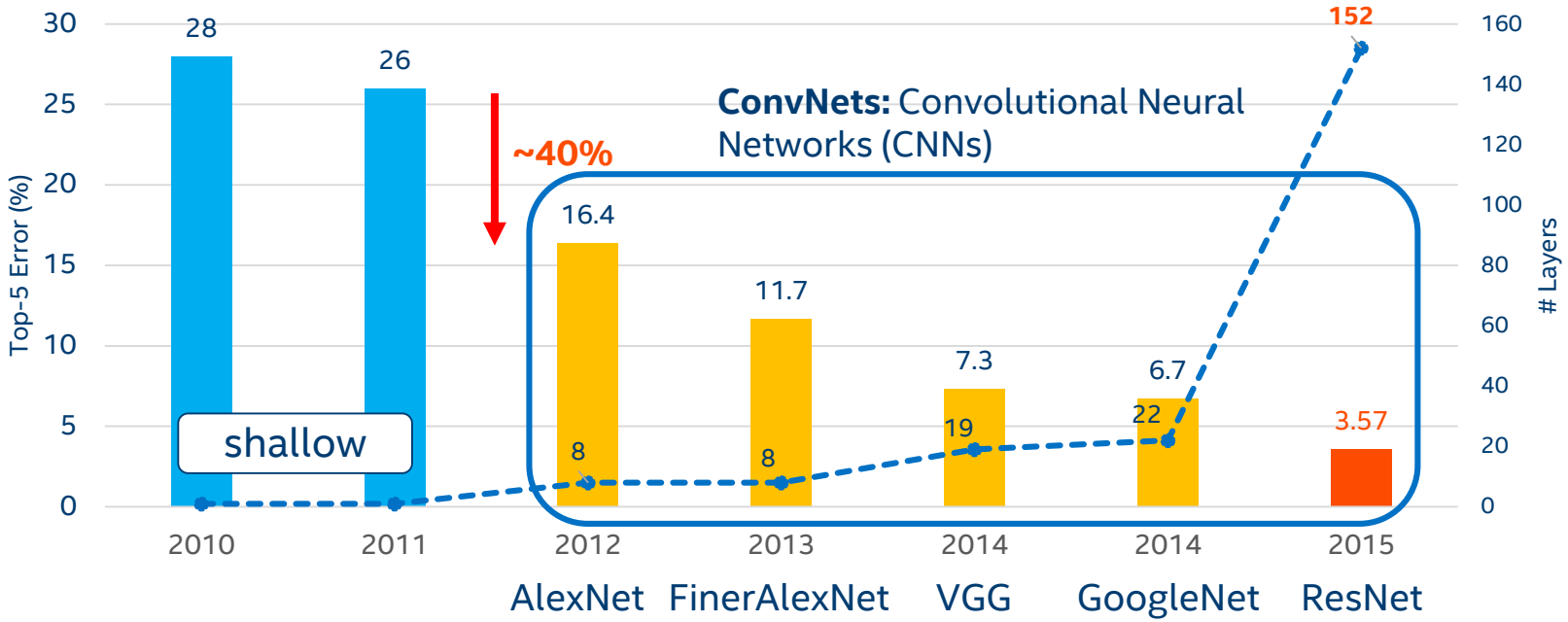
物体分类：人，相机



行为识别：拍照

视觉理解目标：从图像和视频中获取有用信息和知识

视觉识别：深度学习的突破



ImageNet Large Scale Visual Recognition Challenge (ILSVRC): 1000-catg Object Classification

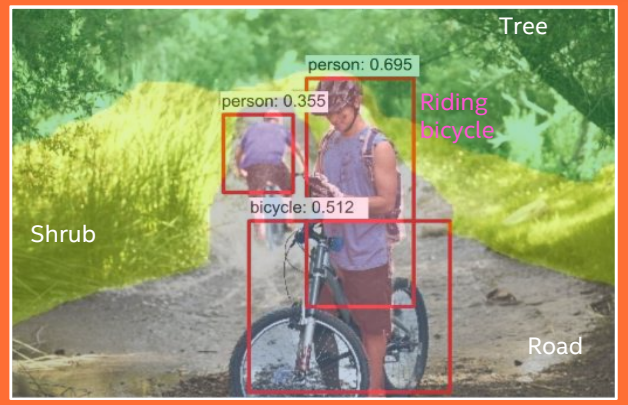
“卷积神经网络(CNNs)目前已几乎成为所有识别和检测任务的**主流方法**，并在一些任务中接近人类水平。”

LeCun, Bengio, Hinton, 深度学习，《自然》，2015年5月

英特尔研究院视觉理解与合成研究

英特尔平台上**智能视觉数据处理技术**研究创新

视觉理解



- 物体识别/检测
- 行为/活动识别
- 图像语义分割
- 几何布局估计
- 高层场景理解
- 视觉为中心的多模态理解，包括情感、视觉内容等

基础组件
- CNN架构、视觉测距、SLAM、视觉索引等

图像/视频合成

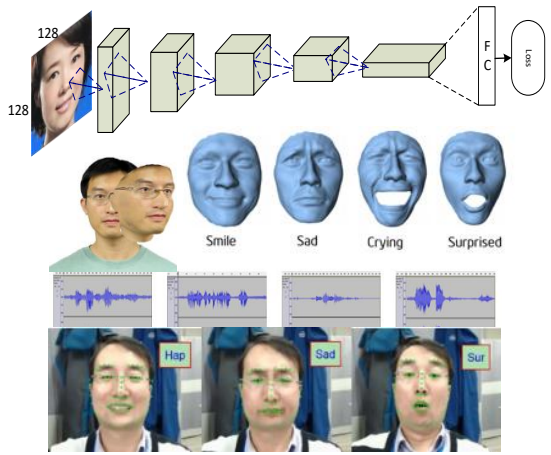
- 三维建模与重构
- 几何处理
- 动画与渲染

软硬件协同设计 **软件接口/工具** **应用/系统原型** **视觉决策/控制**

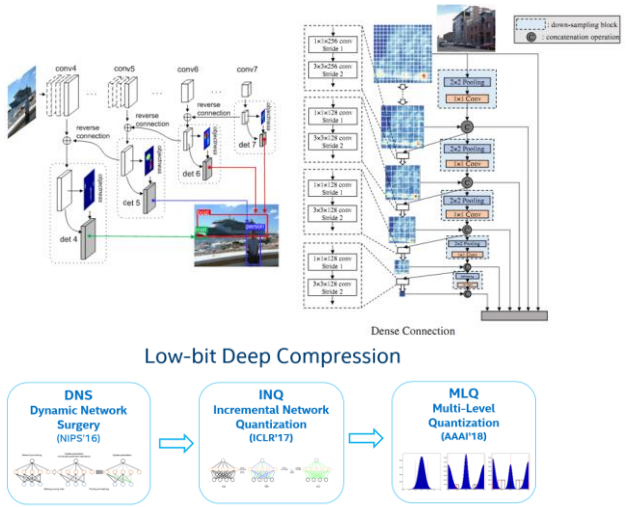
英特尔中国研究院视觉理解研究

前沿视觉认知和机器学习技术研究：感知→认知

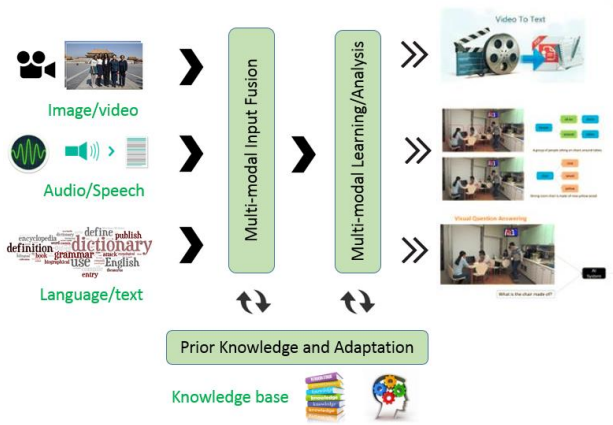
人：人脸分析与情感识别



物：高效CNN设计和DNN压缩



景：视觉解析与多模态分析

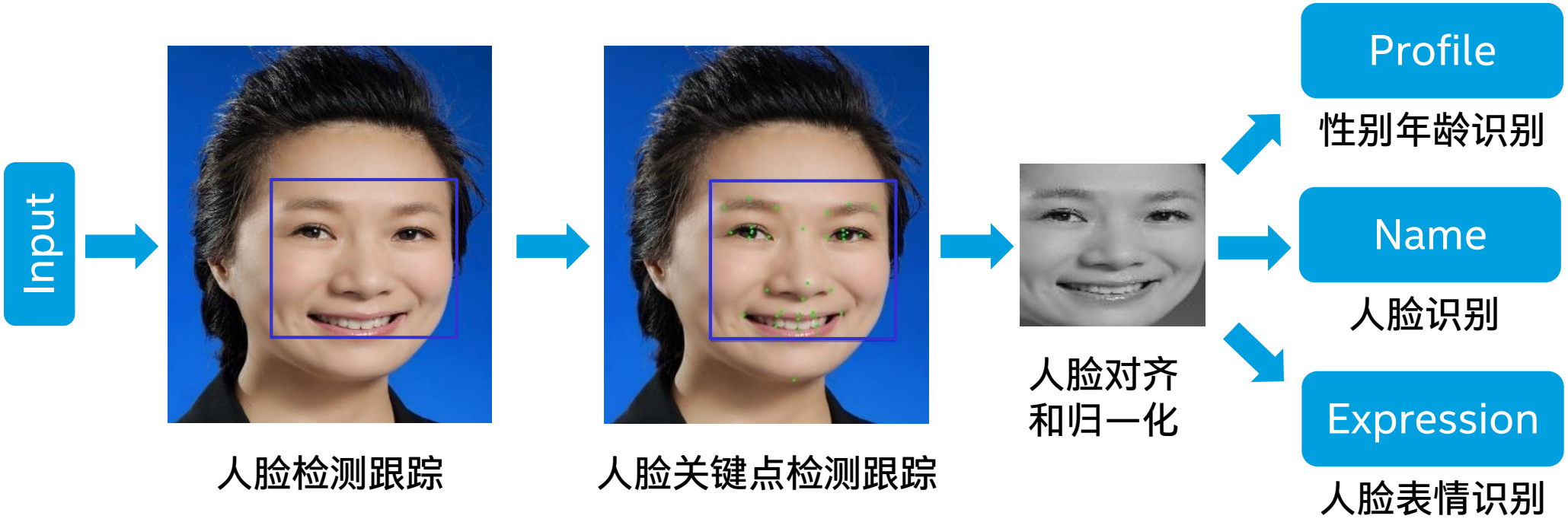


从二维到三维到多模态

人脸分析及情感识别技术

人脸分析技术研究

英特尔中国研究院研发完整人脸分析算法框架

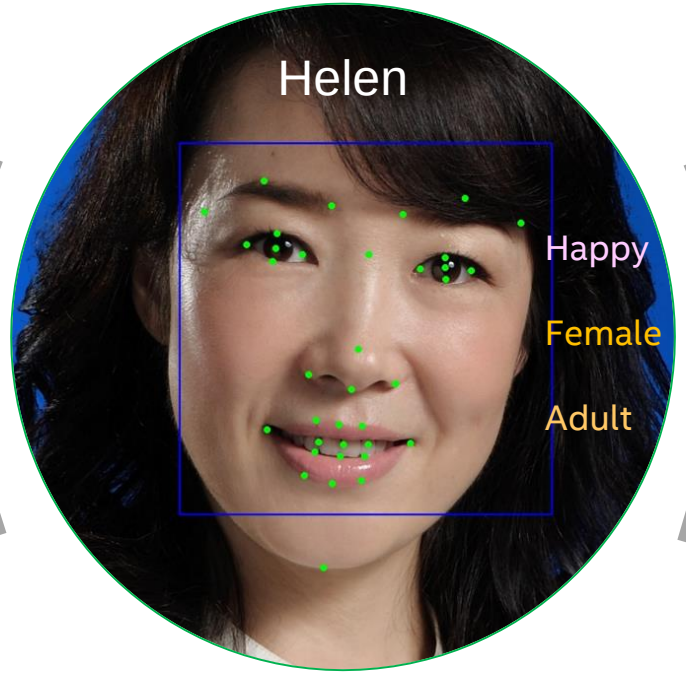


人脸分析技术演进

2011 2012 2013 2014 2015 2016 2017

人脸技术增强用户体验

集显加速



应用原型

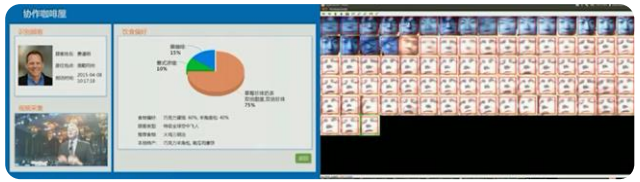
SDK 集成

intel REALSENSE TECHNOLOGY

Intel® Computer Vision SDK

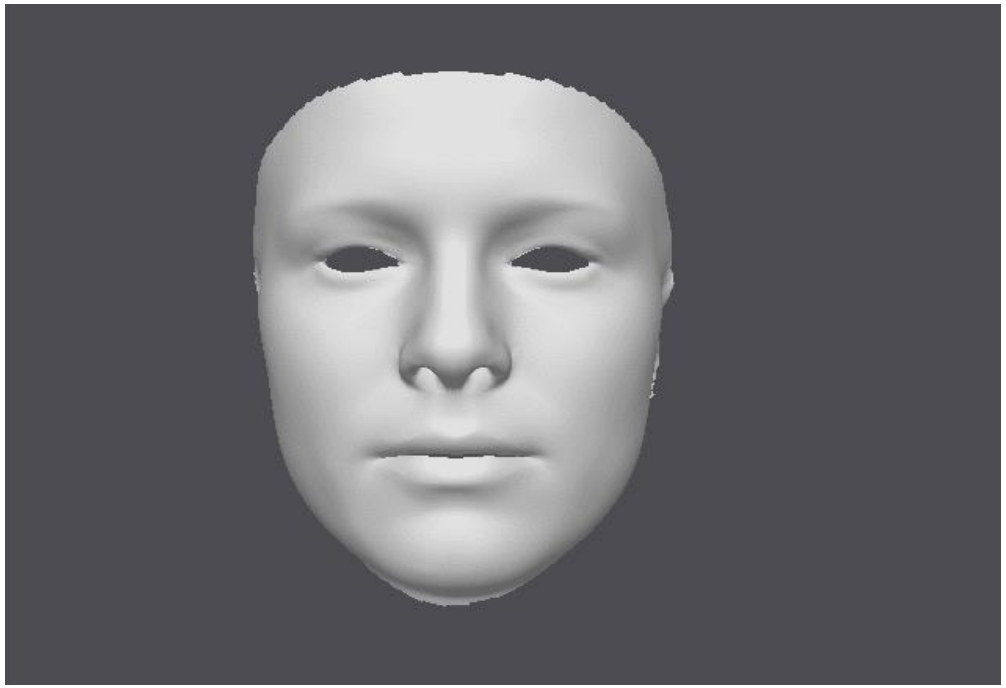
英特尔中国研究院
先进人脸分析技术

视频端到端解决方案



三维人脸技术

领先的 实时三维人脸建模和跟踪增强技术，可用于VR/AR/游戏等多领域



三维人脸技术演示视频

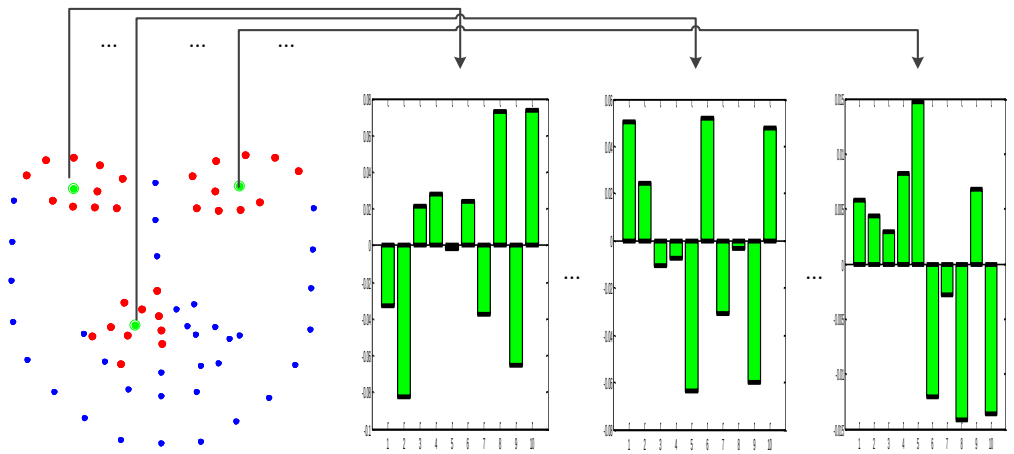
李宇春全球首支基于 Intel AI 技术的MV
——《今天雨，可是我们在一起》

视觉情感识别

智能世界应该是一个充满情感的世界...

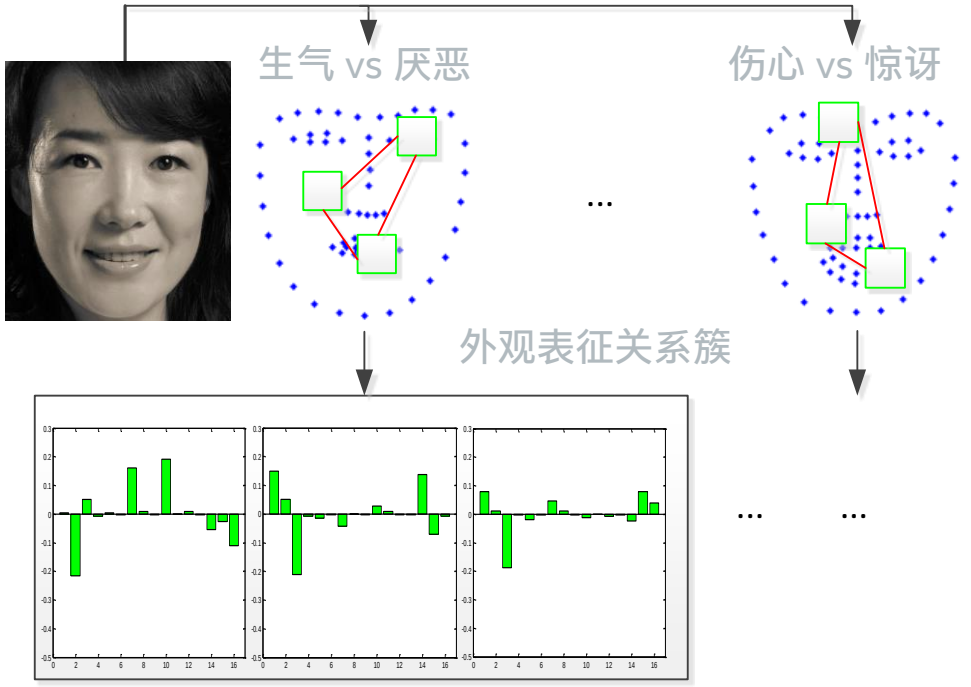
表情识别：显式及稠密地解码人脸肌肉运动及其内在交互关系

几何特征关系簇
（基于线性回归）



从人脸关键点中提取 1 对 10 的几何关系

动作单元感知特征及交互
（基于多任务学习）



自然环境下情感识别挑战赛

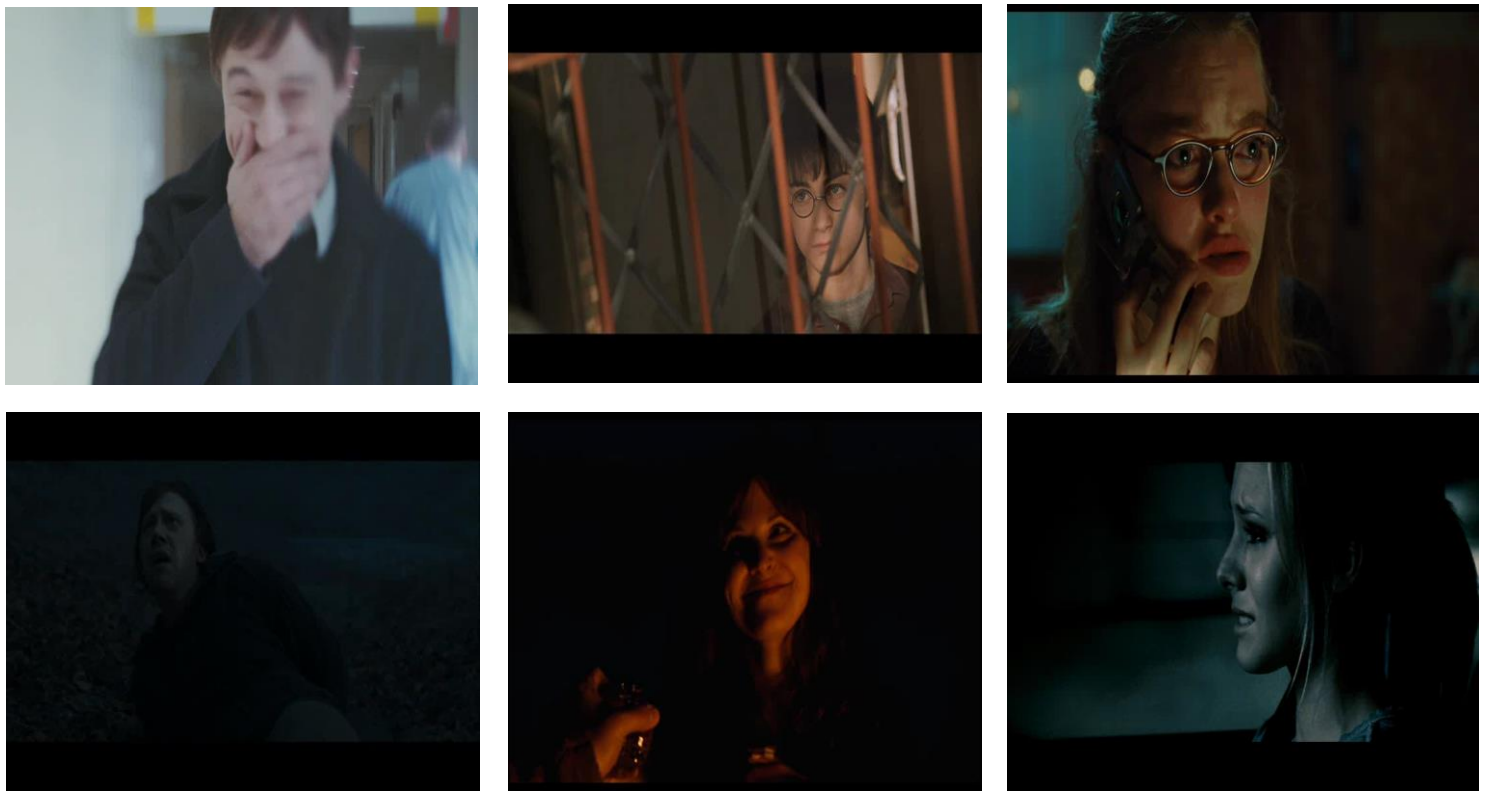
英特尔中国研究院在 EmotiW 2015 (ACM ICMI 2015) 音视频挑战赛中**拔得头筹**，竞争对手来自全球 74 个团队 (CMU、UIUC、MSR 等)

- **任务 1：EmotiW 2015 AFEW 数据集**
  723/383/539 个电影片段（带音频），在完全自然环境下展示出 7 种基本人脸表情
- **任务 2：EmotiW 2015 SFEW 数据集**
  958/436/372 个静态图像，在完全自然环境下展示出 7 种基本人脸表情

EmotiW 2015 测试集上的整体识别率 (%)

方法	AFEW	SFEW
基准	39.33	39.13
2014 年获奖者	50.37	不适用
英特尔解决方案	53.80	55.38

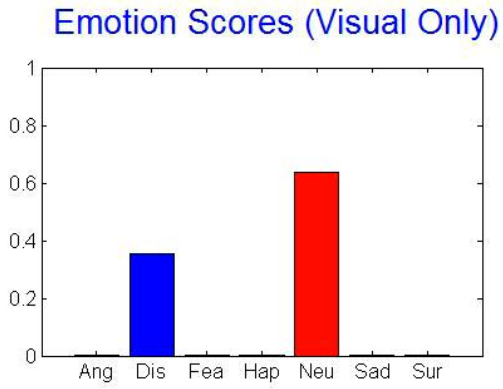
EmotiW 2015 视频片段示例

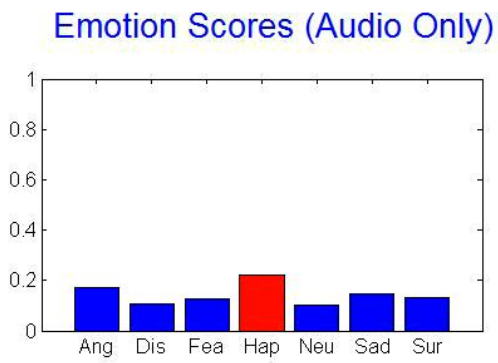


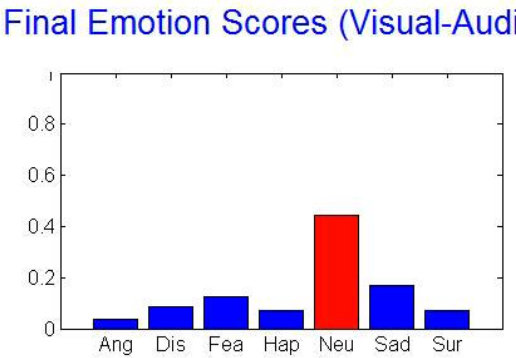
视频来源: EmotiW 2015

实时多模态情感识别系统

视频来源: EmotiW2015





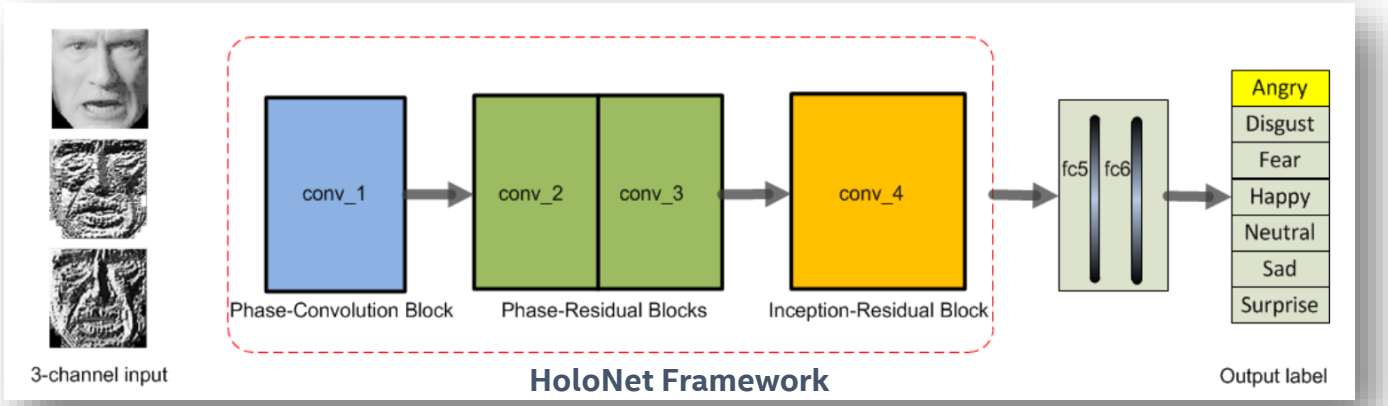


Loading Emotion Data

Run/Pause

HoloNet: 超实时情感识别解决方案

英特尔中国研究院在EmotiW 2016 (ACM ICMI 2016) 音视频比赛中获得亚军 (100个注册团队)，并获得过去四年挑战赛**最具影响力论文奖**



"... You showed me a really novel method, no use of extra data and its speed is hundreds of times faster than the other competitors.

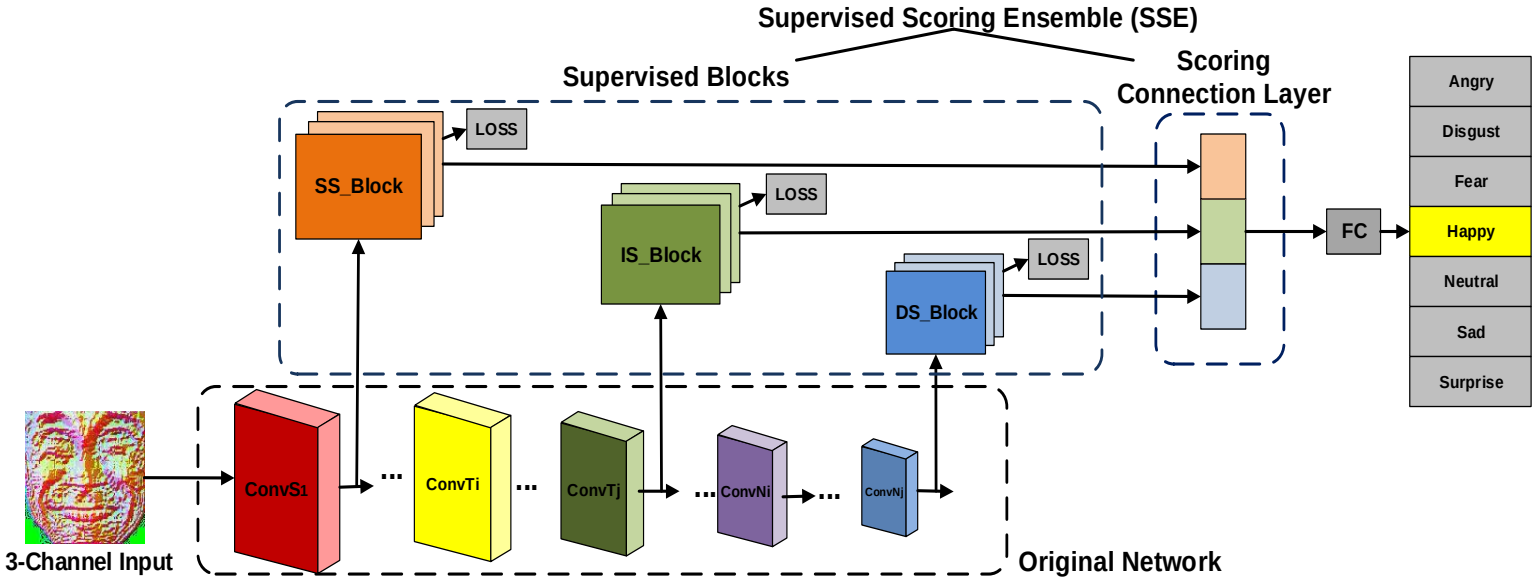
Abhinav, EmotiW 2016 Chair

Submission#	Validation (%)	Test (%)	Method
1	47.78	54.30	Fusion of HoloNet model A + 1 audio model
2	48.83	55.14	Fusion of HoloNet model B + 1 audio model
3	50.13	56.83	1st Fusion of HoloNet model A&B + 1 audio model
4	50.91	55.14	2nd Fusion of HoloNet model A&B + 1 audio model
5	**51.96**	**57.84**	Fusion of HoloNet model A&B + 1 audio model + 1 iDT model

Total recognition accuracy of our 5 submissions to AFEW 6.0, both on the validation and the test sets.

SSE：高精度情感识别解决方案

英特尔中国研究院提出聚合监督情感识别算法 SSE 显著提高识别率，再次问鼎 EmotiW 2017 (ACM ICMI 2017) 音视频比赛 (100多个注册团队)



SSE 学习框架

单模型识别准确率比 HoloNet高 ***5.5%***

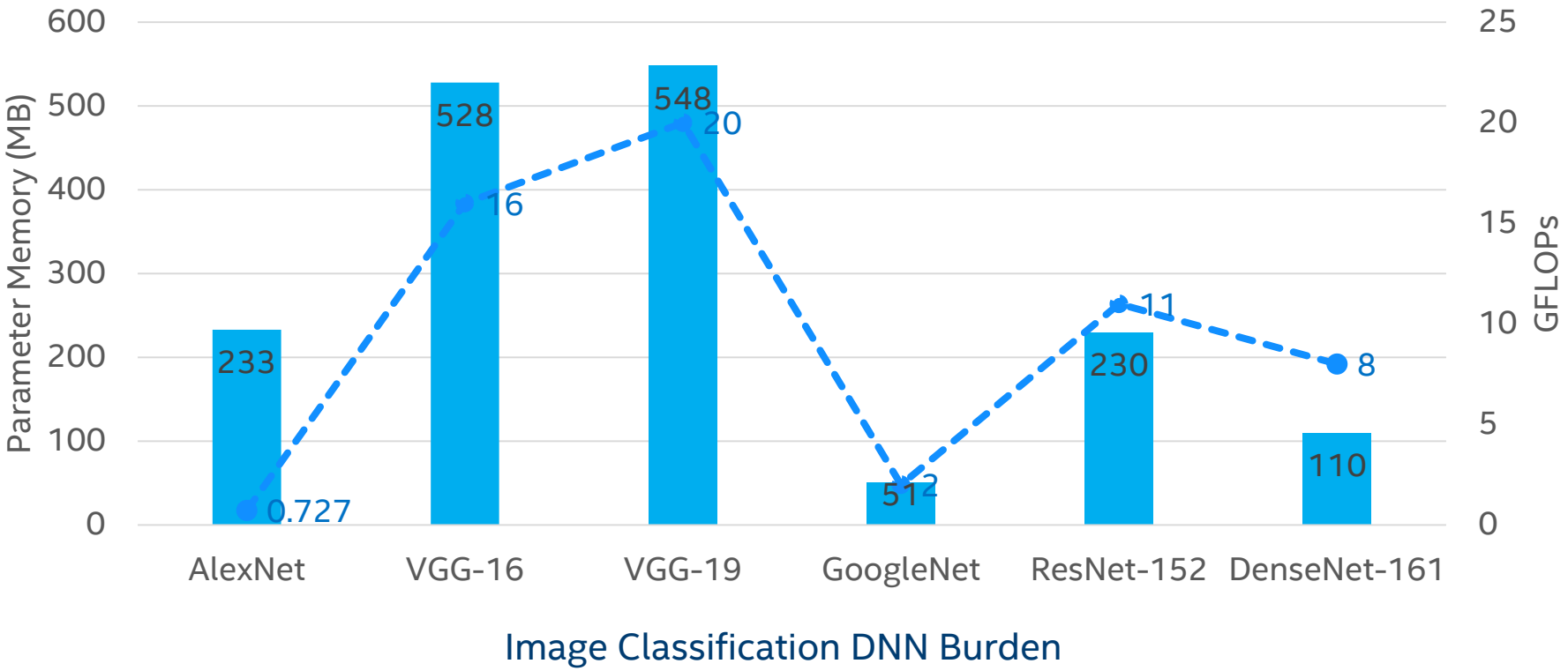
EmotiW 2017中取得 ***60.3%*** 识别准确率

P. Hu, D. Cai, S. Wang, A. Yao, Y. Chen, "Learning Supervised Scoring Ensemble for Emotion Recognition in the Wild", ACM ICMI 2017.

算法创新赋能前端设备

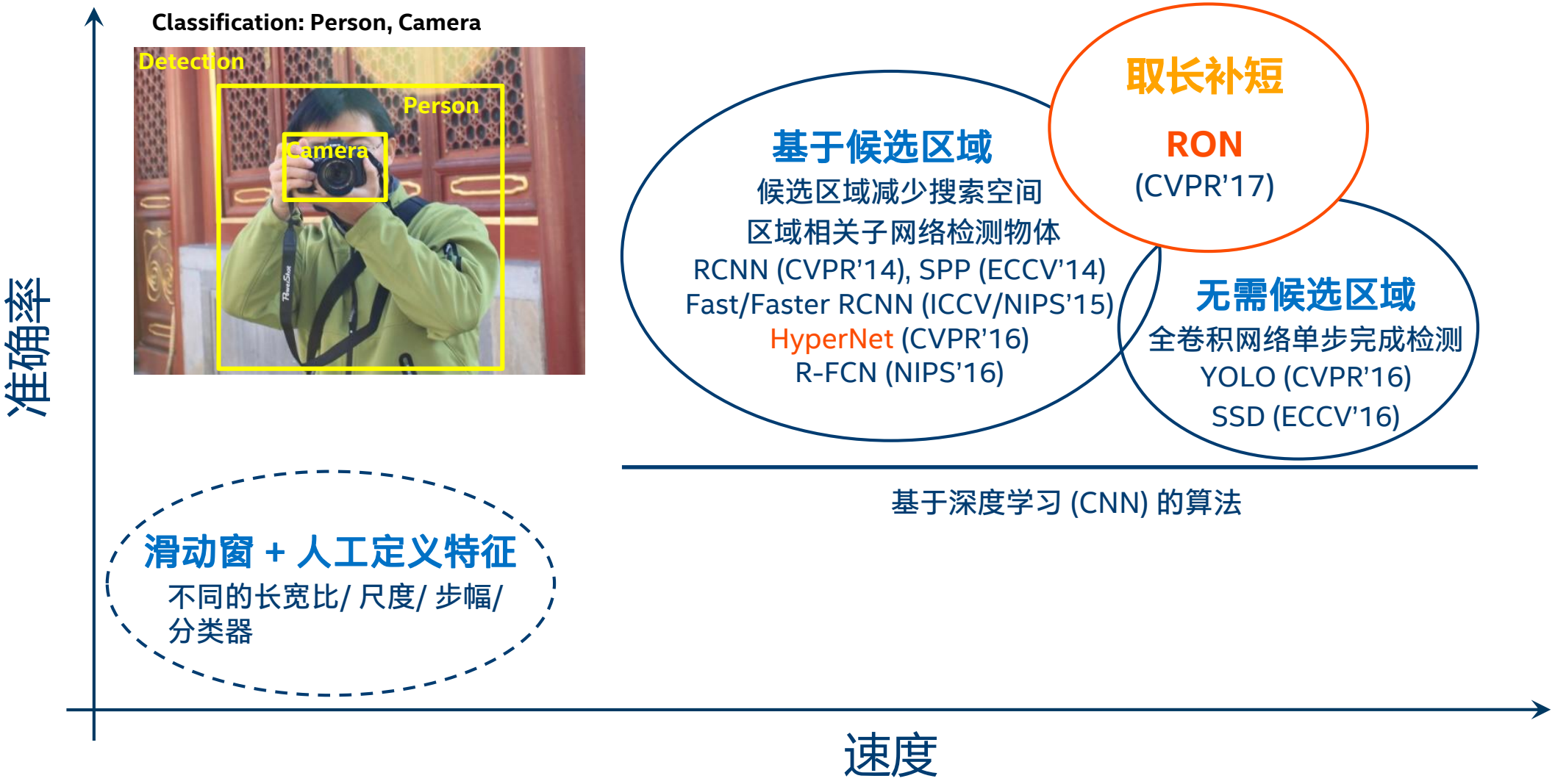
物体识别网络设计与压缩技术

深度学习计算与存储挑战



部署挑战：主流DNNs都是计算和存储密集型的！很难部署到边缘、嵌入式设备上

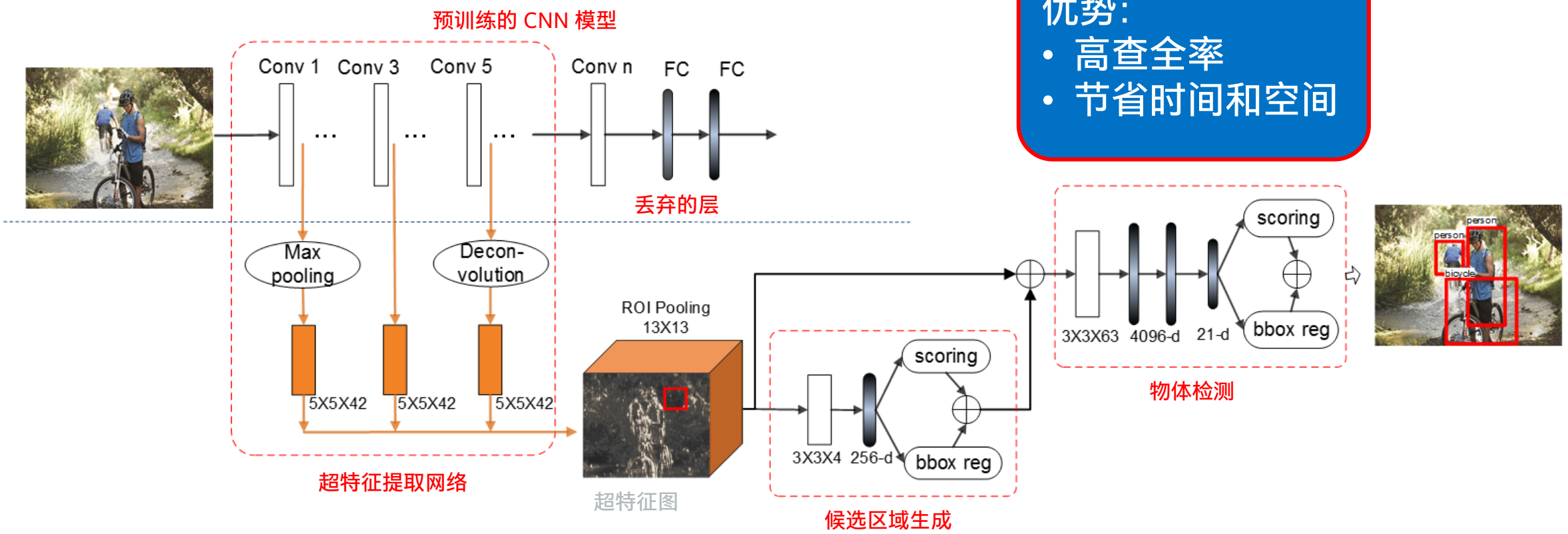
DNN模型设计：物体检测算法的演化



HyperNet: 高效物体检测解决方案

针对候选区域生成和物体检测提供统一框架 (CVPR'16)

- 在各种任务间共享超特征 (Hyper Feature)



Conv: 卷积层 FC: 完全连接层 ROI: 感兴趣区域 bbox reg: 包围盒回归

RON：基于显著性先验网络的反向连接金字塔算法

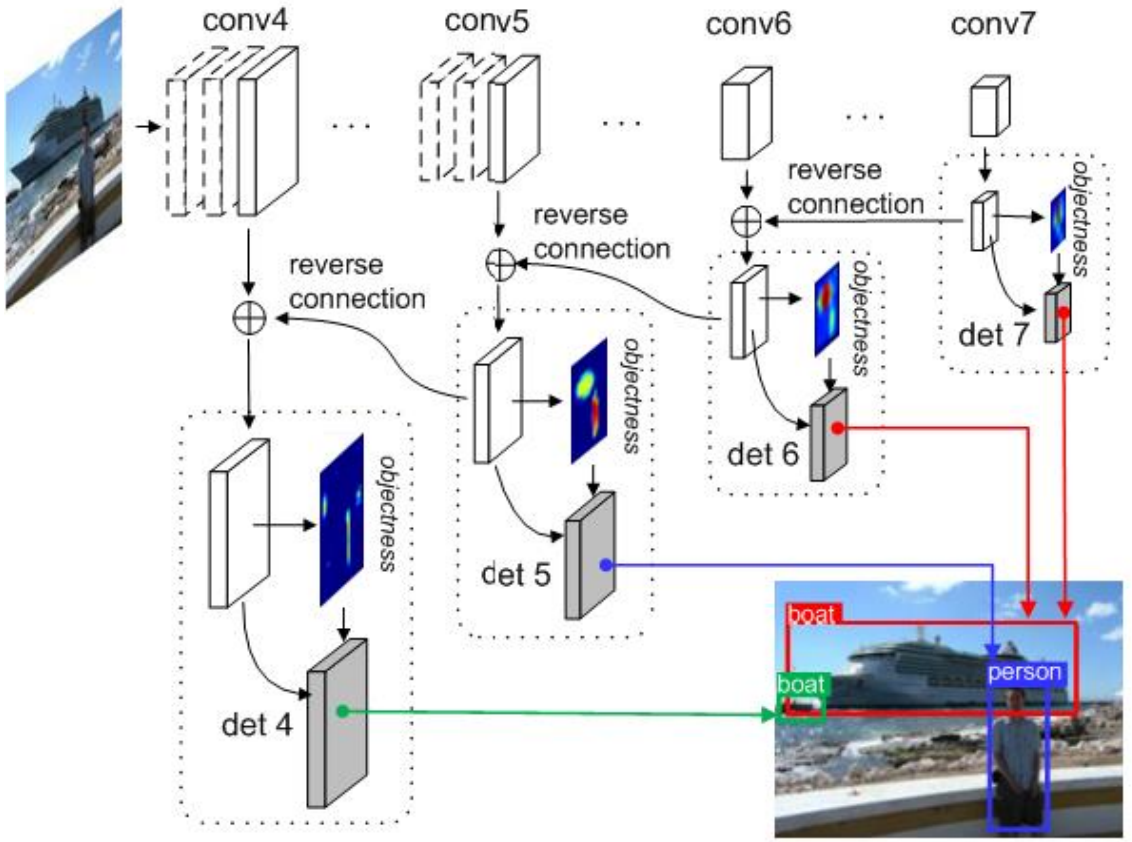
取长补短 的全卷积网络物体检测算法 (CVPR'17)

- 基于和无需候选区域的两类CNN算法基础上

成功解决了两个基本问题

- 多尺度物体定位
- 高效负样本空间挖掘

实现领先的准确率和速度



T. Kong, A. Yao, F. Sun, M. Lu, H. Liu, Y. Chen, "RON: Reverse Connection with Objectness Prior Networks for Object Detection", CVPR 2017.

DSOD：深度监督物体检测器

首项 从头开始训练的物体检测工作 (ICCV'17)

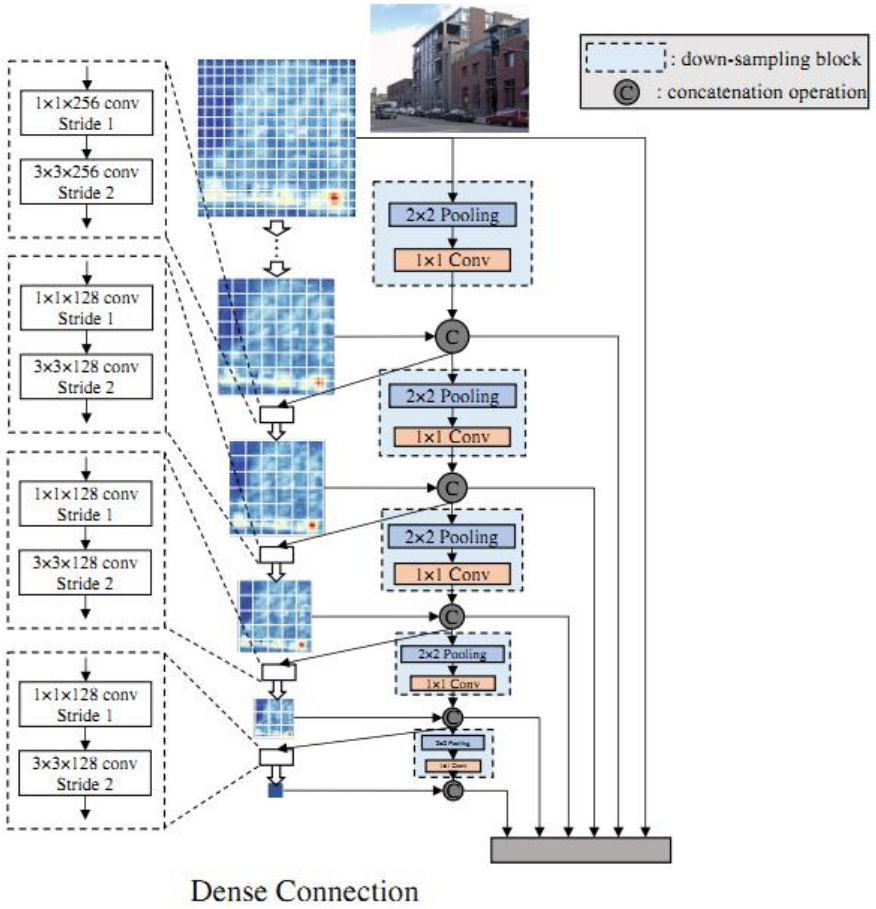
- 无需 ImageNet 上百万张图像的预训练
- 只需数万张注图像

先进的准确率和效率

- 模型参数：½ SSD, ¼ R-FCN, 1/10 Faster-RCNN
- 准确率比SSD/YOLOv2好 (PASCAL VOC/MS COCO)
- 速度：20fps (GPU，无特殊代码优化)

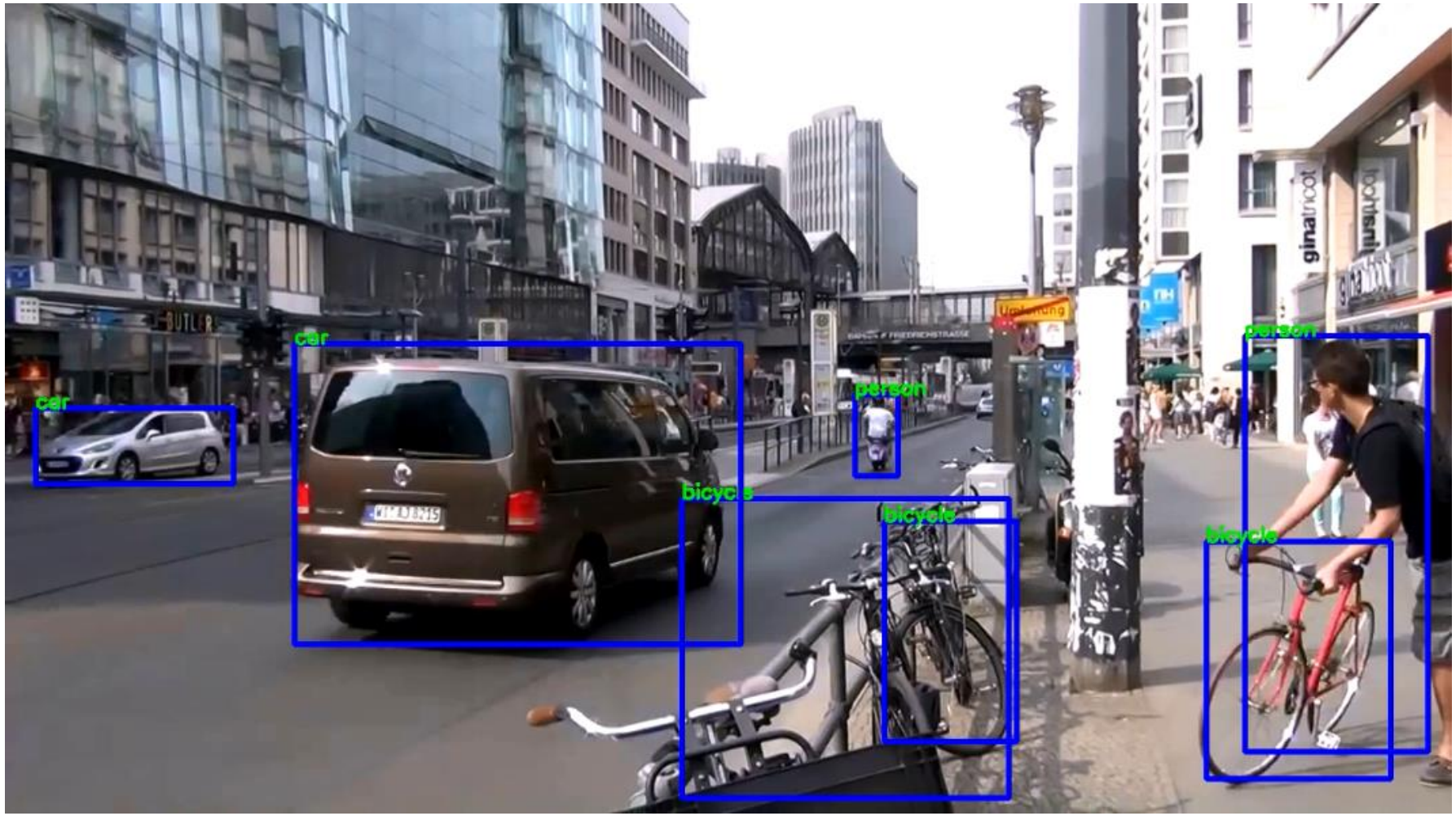
物体检测器设计更多可能性

- 针对特定问题的有限数据集训练
- 其它领域，如深度、医学、多光谱图像网络设计



Z. Shen, Z. Liu, J. Li, Y. Jiang, Y. Chen, X. Xue, "DSOD: Learning Deeply Supervised Object Detectors from Scratch", ICCV 2017.

多类别物体检测实时系统



视频来源: youtube

模型压缩：低精度深度压缩三部曲

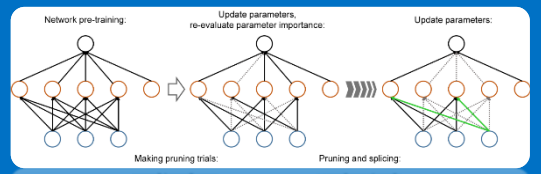
简洁方案取得百倍DNN模型无损压缩性能

低精度权值和激活参数

DNS

动态网络手术

(NIPS'16)

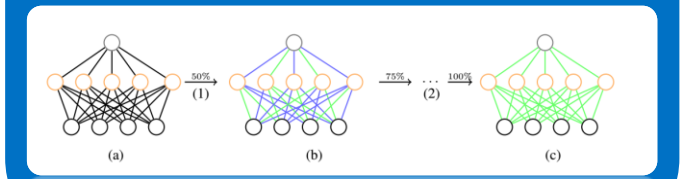


优化DNN结构

INQ

渐进网络量化

(ICLR'17)

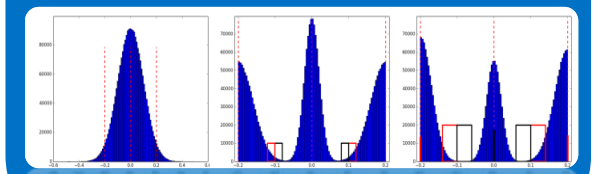


降低DNN权值精度

MLQ

多尺度编码量化

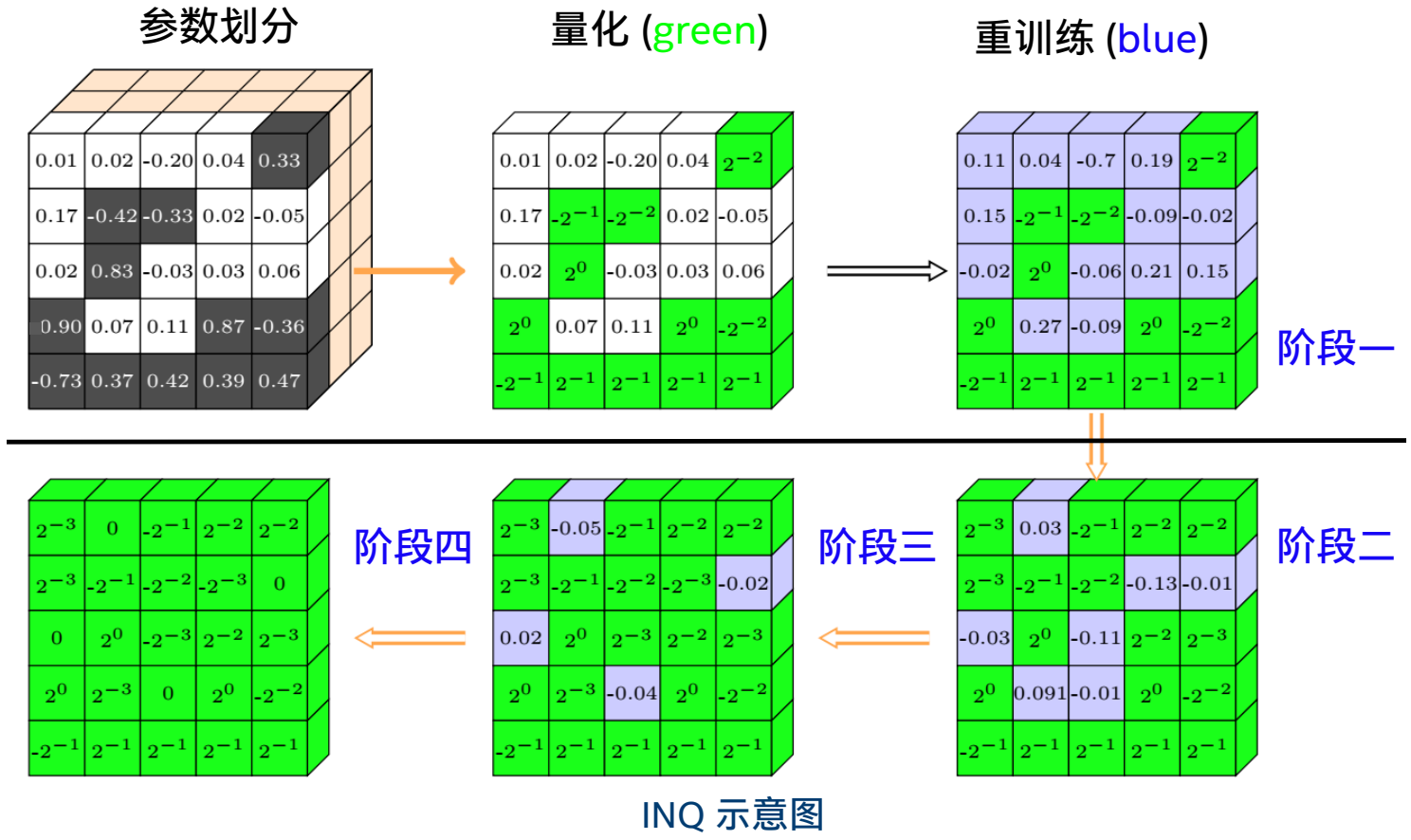
(AAAI'18)



降低DNN激活参数精度

INQ：渐进网络量化算法

创新参数划分，量化和重训练框架实现无损压缩



INQ 示意图

优点:

- 首个无损网络量化方案
- 与网络模型无关
- 移位替换浮点乘法
- 高效训练

A. Zhou, A. Yao, Y. Guo, L Xu and Y. Chen, “Incremental Network Quantization: Towards Lossless CNNs with Low-precession Weights”. ICLR 2017.

INQ效果显著

对于主流DNN网络，5-bit 量化提高识别准确率, 在超低精度（2/3-bit）时仍可取得接近全精度模型识别准确率

5-bit 量化结果

Network	Bit-width	Top-1 error	Top-5 error
AlexNet ref	32	42.76%	19.77%
AlexNet	5	**42.61%**	**19.54%**
VGG-16 ref	32	31.46%	11.35%
VGG-16	5	**29.18%**	**9.70%**
GoogleNet ref	32	31.11%	10.97%
GoogleNet	5	**30.98%**	**10.72%**
ResNet-18 ref	32	31.73%	11.31%
ResNet-18	5	**31.02%**	**10.90%**
ResNet-50 ref	32	26.78%	8.76%
ResNet-50	5	**25.19%**	**7.55%**

多种量化结果对比

Model	Bit-width	Top-1 error	Top-5 error
ResNet-18 ref	32	31.73%	11.31%
INQ	5	31.02%	10.90%
INQ	4	31.11%	10.99%
INQ	3	31.92%	11.64%
INQ	2 (ternary)	33.98%	12.87%

Method	Bit-width	Top-1 error	Top-5 error
BWN	1	39.20%	17.00%
TWN	2 (ternary)	38.20%	15.80%
INQ (ours)	2 (ternary)	**33.98%**	**12.87%**

低精度深度压缩的优异性能

AlexNet测试结果表明：超越主流深度压缩方案至少一倍, 在2/4-bit精度下达到 >100倍网络压缩

Method	Bit-width (Conv/FC)	Bit-width (Act)	Compression ratio	Decrease in top-1 / top-5 error rate
P+Q *	8/5	32	27x	0.00% / 0.03%
P+Q+H *	8/5	32	35x	0.00% / 0.03%
Our method	4/4	4	71x	0.08% / 0.03%
P+Q+H *	4/2	32	-	-1.99% / -2.60%
Our method	3/3	4	89x	-0.52% / -0.20%
Our method	2/2	4	142x	-1.47% / -0.96%

Comparison of our low-bit deep compression and deep compression method (P+Q+H, LCLR'16 and ISCA'16) on AlexNet. Conv: Convolutional layer, FC: Fully connected layer, Act: Activation, P: Pruning, Q: Quantization, H: Huffman coding.

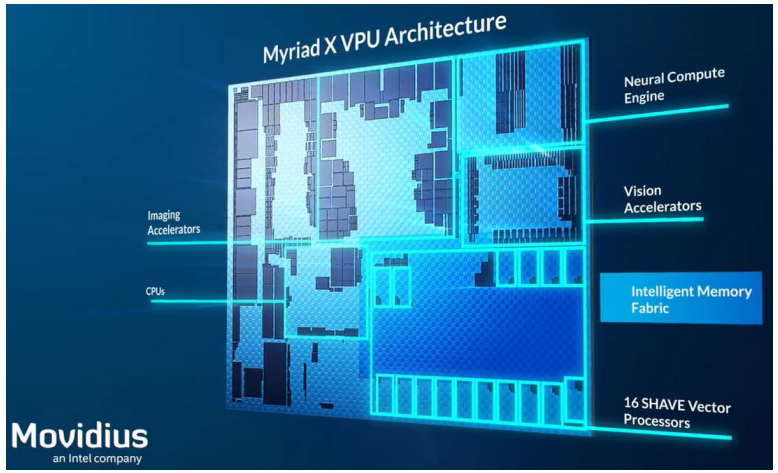
* S. Han, J. Pool, J. Tran, and W. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. Best paper in ICLR 2016.

促进深度学习推断硬件加速

低精度深度压缩技术配合英特尔低功耗硬件，为雾计算和边缘计算提供深度学习推断硬件加速能力

INTEL FPGAS

Movidius



让机器看懂图和视频

场景理解：多模态图像视频解析

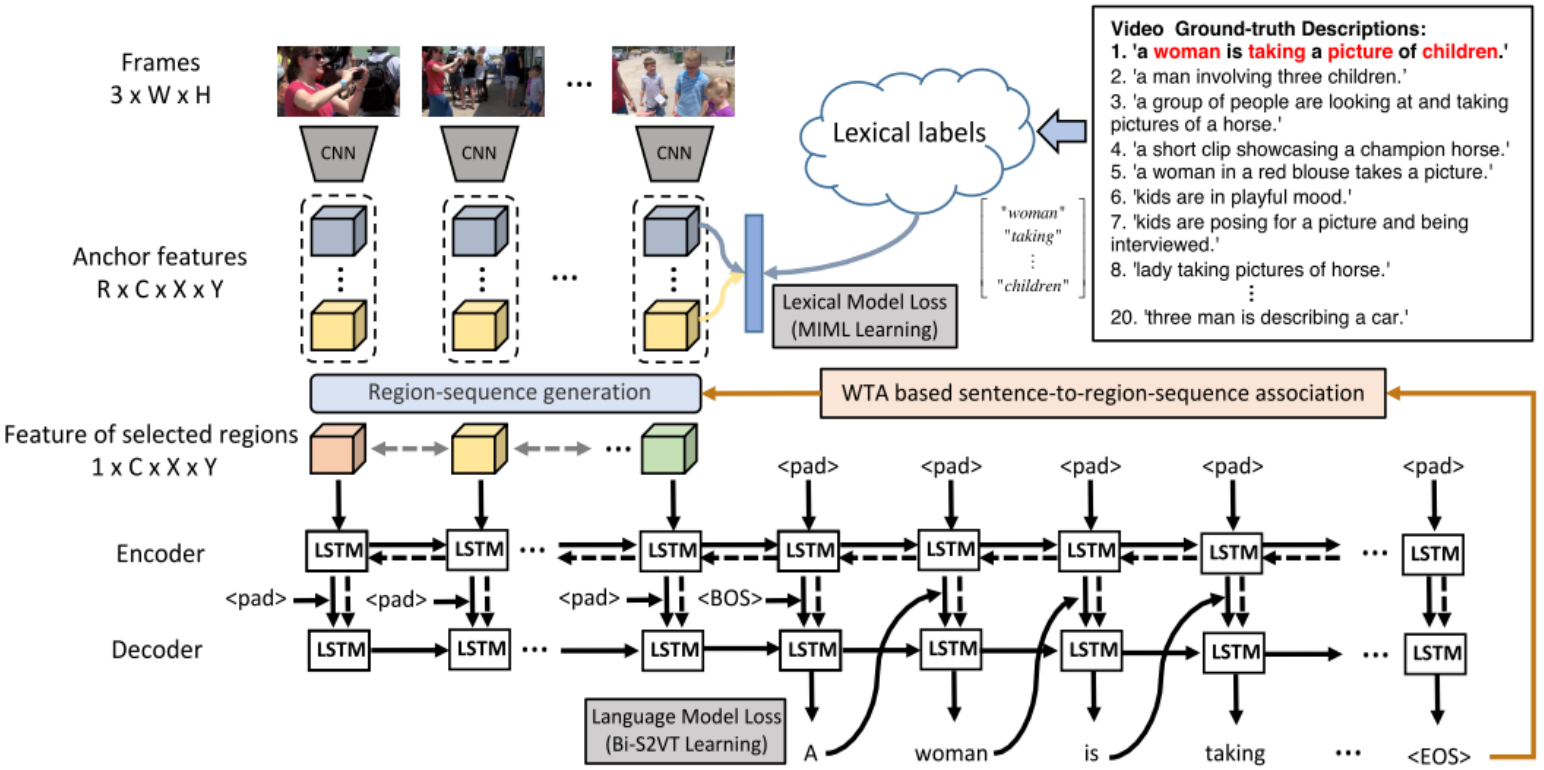
视觉解析和多模态分析

领先多模态融合和学习算法研究将视觉分析能力从识别推进到理解

详细视频描述研究

自动生成信息丰富的多样性详细视频标注，效果超过优秀单视频标注方法

视频来源: MSR-VTT 数据集*



Z. Shen, J. Li, Z. Su, M. Li, Y. Chen, Y. Jiang, X. Xue, "Weakly Supervised Dense Video Captioning ", CVPR 2017.

* J. Xu, T. Mei, T. Yao, and Y. Rui. Msr-vtt: A large video description dataset for bridging video and language. CVPR 2016.

详细视频描述研究

自动生成信息丰富的多样性详细视频标注，效果超过优秀单视频标注方法

视频来源: MSR-VTT 数据集*

Region Sequences & DenseVidCap

A **woman** in **red blouse** is **taking pictures** of **children**

A group of people are **taking pictures** of a **horse**

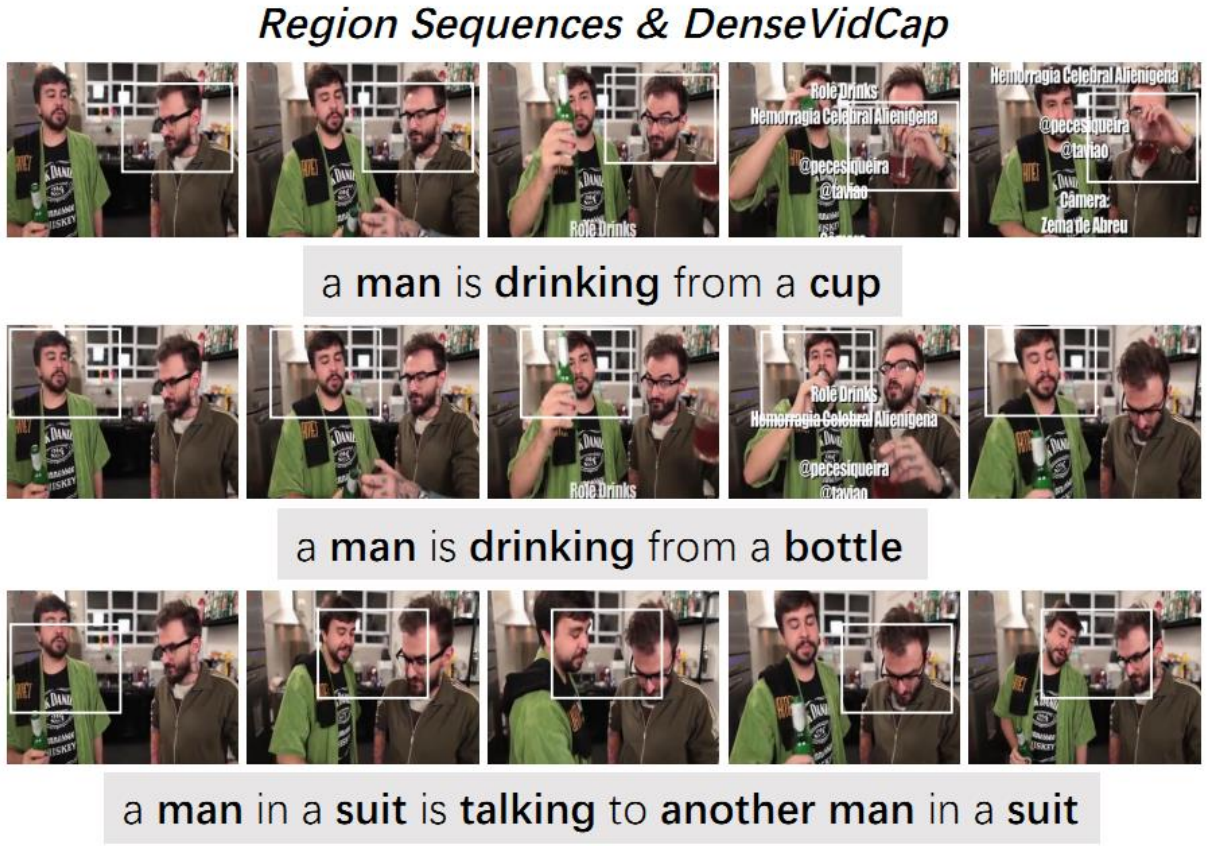
Kids are being **interviewed**

Z. Shen, J. Li, Z. Su, M. Li, Y. Chen, Y. Jiang, X. Xue, "Weakly Supervised Dense Video Captioning ", CVPR 2017.

* J. Xu, T. Mei, T. Yao, and Y. Rui. Msr-vtt: A large video description dataset for bridging video and language. CVPR 2016.

详细视频描述研究

自动生成信息丰富的多样性详细视频标注，效果超过优秀单视频标注方法



Z. Shen, J. Li, Z. Su, M. Li, Y. Chen, Y. Jiang, X. Xue, “Weakly Supervised Dense Video Captioning ”, CVPR 2017.

* J. Xu, T. Mei, T. Yao, and Y. Rui. Msr-vtt: A large video description dataset for bridging video and language. CVPR 2016.

详细视频描述研究

自动生成信息丰富的多样性详细视频标注，效果超过优秀单视频标注方法

Results on MSR-VTT Challenge

Team	Memo	METEOR	BLEU-4	ROUGE-L	CIDEr
Ruc-UVA	RUC + UVA + ZJU	26.9	38.7	58.7	45.9
VideoLab	UCB + Austin +...	27.7	39.1	60.6	44.1
Aalto	Aalto Univ.	26.9	39.8	59.8	45.7
V2t-navigator	RUC + CMU	28.2	40.8	60.9	44.8
Ours	ILC	**28.3**	**41.4**	**61.1**	**48.9**

Z. Shen, J. Li, Z. Su, M. Li, Y. Chen, Y. Jiang, X. Xue, “Weakly Supervised Dense Video Captioning ”, CVPR 2017.

* J. Xu, T. Mei, T. Yao, and Y. Rui. Msr-vtt: A large video description dataset for bridging video and language. CVPR 2016.

视频描述演示


A passenger train is pulling into a station.
A man in orange uniform is walking on a platform.

总结

智能视觉数据处理技术研究解决视觉数据爆炸挑战

前沿深度学习视觉理解研究影响英特尔架构及平台设计，提升英特尔产品用户体验

广泛学术界合作加速研究创新，推动视觉理解技术发展


知未来，创未来
We know the future
because we are building it


intel
experience
what's inside™