



中国科学技术大学
University of Science and Technology of China

Ch11-智能计算系统

中国科学技术大学 王超
2024年春

智能应用已成为与人类生活密切相关的负载

人工智能技术切实改变人类生活

超级计算机



数据中心



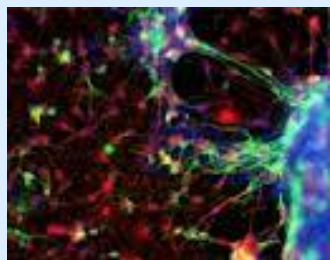
智能手机



嵌入式设备



神经网络
机器学习



药物研制



自动翻译



图像分析



机器人

高效灵活的智能计算需求

数据规模大

能耗开销高

应用种类多

算法迭代快



智慧城市



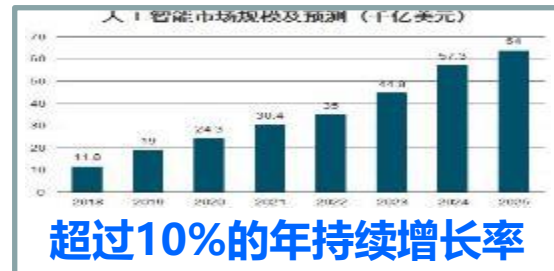
智能医疗



智慧交通



智能制造



维护投入



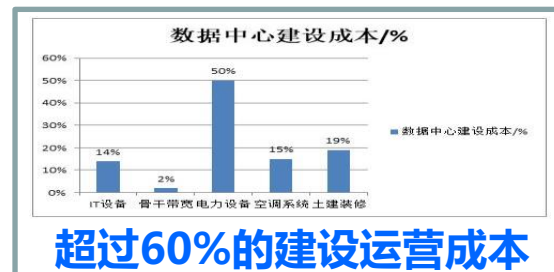
环境影响



电力开销



基建成本



目标检测



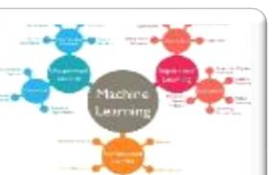
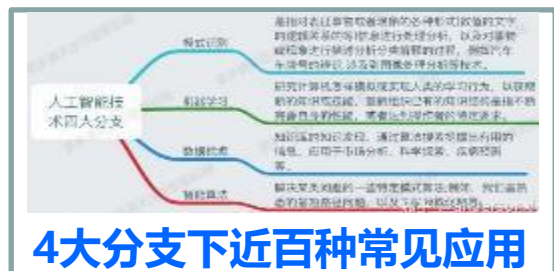
图像分类



语音识别



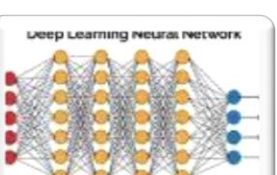
文字处理



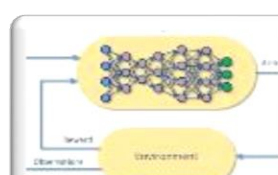
机器学习



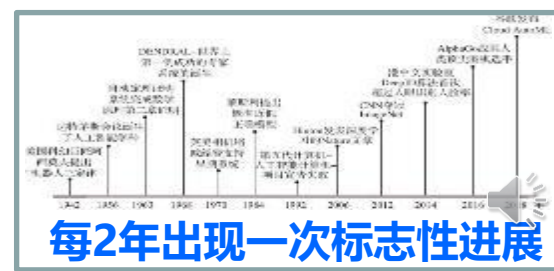
数据挖掘



深度学习



强化学习



万物互联时代，人工智能具有很强的“头雁”效应



人工智能是引领这一轮科技革命和产业变革的战略性技术,具有溢出带动性很强的“**头雁**”效应

-2018年《习近平在中共中央政治局第九次集体学习时强调加强领导 做好规划明确任务夯实基础 推动我国新一代人工智能健康发展》

“以信息技术、**人工智能**为代表的新兴科技快速发展，大大拓展了时间、空间和人们认知范围，人类正在进入一个“人机物”三元融合的**万物智能互联时代**”、

“**人工智能**、数字经济蓬勃发展，图像识别、语音识别走在全球前列”、

“要在事关发展全局和国家安全的基础核心领域，瞄准**人工智能**、量子信息、集成电路、先进制造、生命健康、脑科学、生物育种、空天科技、深地深海等前沿领域，前瞻部署一批战略性、储备性技术研发项目”

-2021年5月28日在中国科学院第二十次院士大会、中国工程院第十五次院士大会、中国科协第十次全国代表大会上的讲话

人工智能已经成为世界各国争夺的技术制高点



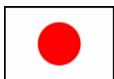
2018年，欧盟发布《欧盟人工智能协调计划》



2018年，德国出台《人工智能战略》



2019年，美国通过《美国人工智能倡议》



2019年，日本出台《人工智能战略2019》



2019年，俄罗斯通过《2030年前俄罗斯国家AI发展战略》



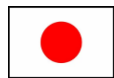
2021年，美国国家AI安全委员会发布《AI国家安全委员会建议报告》



2021年，澳大利亚政府发布“AI行动计划”



2021年，英国制定《国家人工智能战略》战略规划



2021年，日本内阁府提出《AI战略2021》草案



2021年，法国政府出台《AI国家战略》新计划



2022年，以色列国防部公布新的AI战略

人工智能已上升为国家战略（新一代人工智能发展规划-2017）



中华人民共和国中央人民政府

www.gov.cn



国务院

总理

新闻

政策

互动

服务

数据

首页 > 信息公开 > 国务院文件 > 科技、教育 > 科技

国务院关于印发 新一代人工智能发展规划的通知

国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

国务院

2017年7月8日

（此件公开发布）

新一代人工智能发展规划

人工智能的迅速发展将深刻改变人类社会生活、改变世界。为抢抓人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国，按照党中央、国务院部署要求，制定本规划。

战略目标

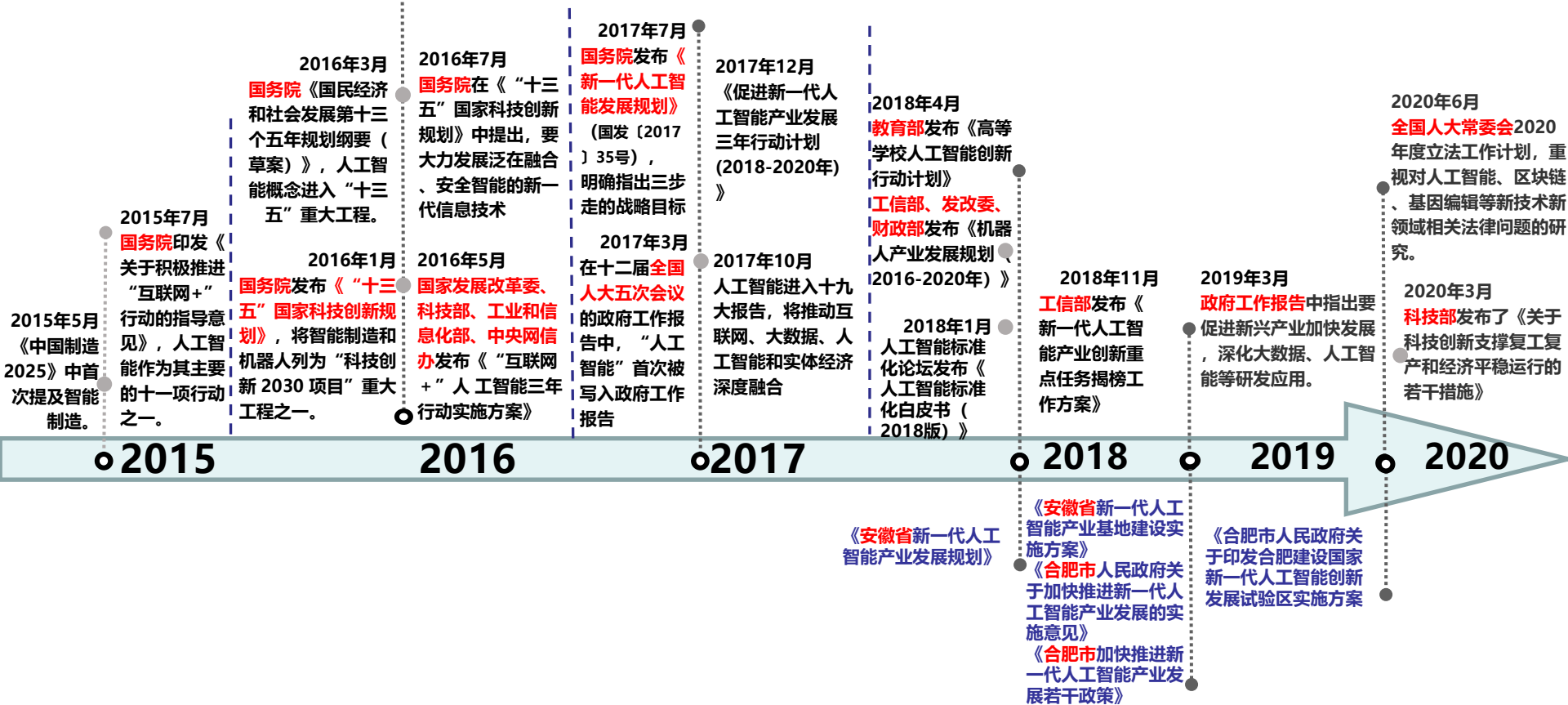
第一步，到2020年人工智能总体技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点，人工智能技术应用成为改善民生的新途径，有力支撑进入创新型国家行列和实现全面建成小康社会的奋斗目标。

第二步，到2025年人工智能基础理论实现重大突破，部分技术与应用达到世界领先水平，人工智能成为带动我国产业升级和经济转型的主要动力，智能社会建设取得积极进展。

第三步，到2030年人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心，智能经济、智能社会取得明显成效，为跻身创新型国家前列和经济强国奠定重要基础。

国家政策支持、区域经济发展动力

- 我国政府高度重视人工智能的技术进步与产业发展，人工智能已上升为国家战略。
- 《新一代人工智能发展规划》：到2030年，使中国成为世界主要人工智能创新中心



国家与地方密集出台人工智能相关政策，大力推动人工智能技术发展

人工智能技术的三个层次



人工智能技术的三个层次



运算智能
能存会算



感知智能
能听会说，能看会认

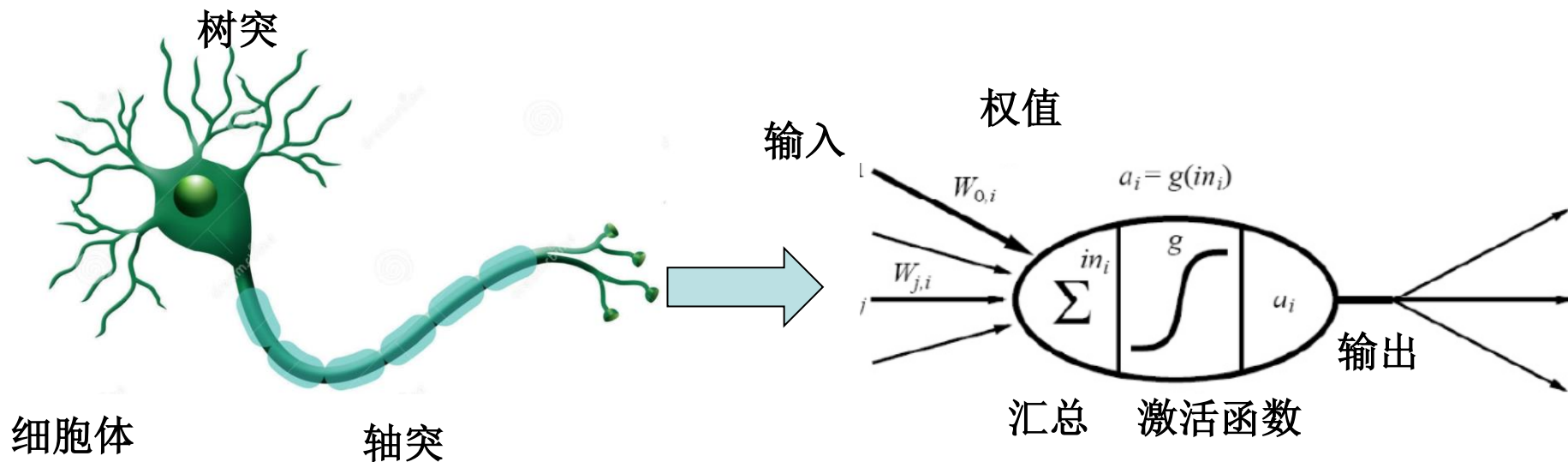


认知智能
能理解会思考

神经网络-人工智能的代表技术与实例

人工神经网络是生物神经元的抽象

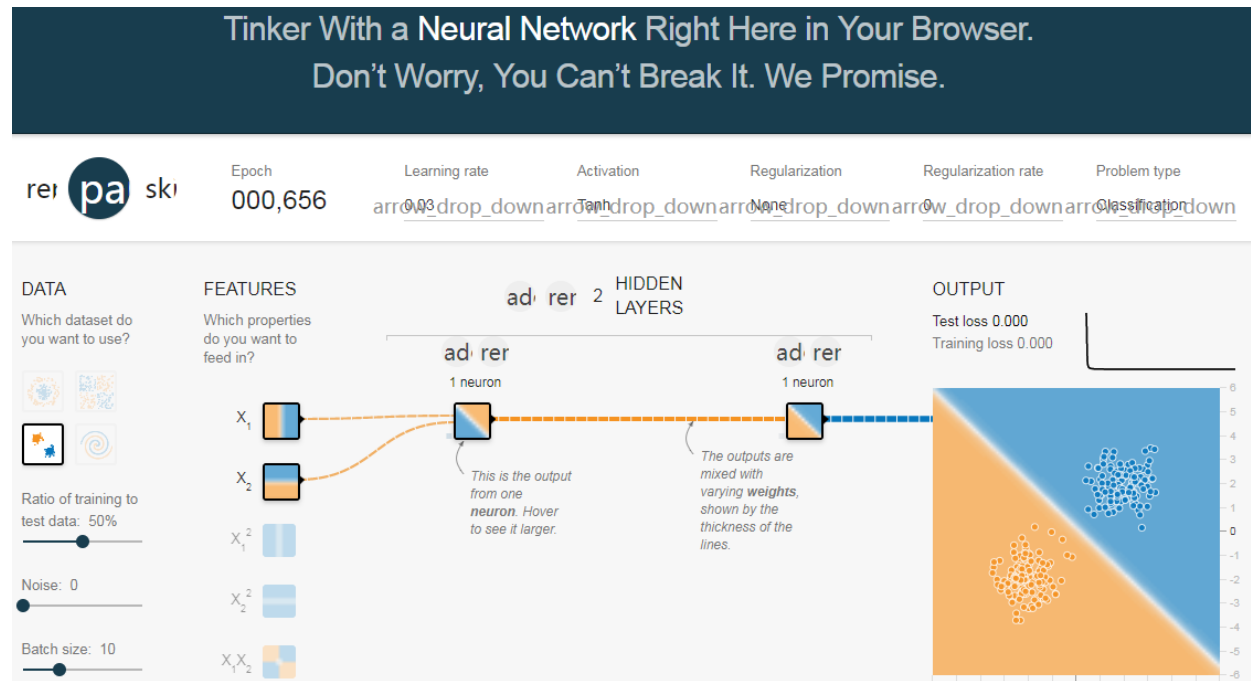
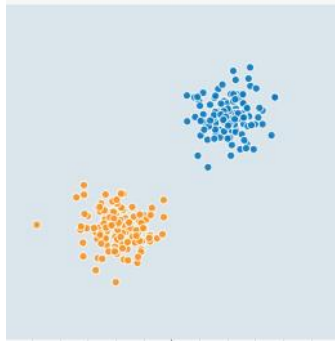
1943年，莫克罗-彼特神经模型 (McCulloch-Pitts neuron model)



人的大脑由千亿神经元
细胞和百万亿突触构成

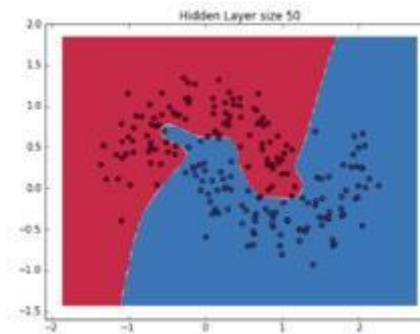
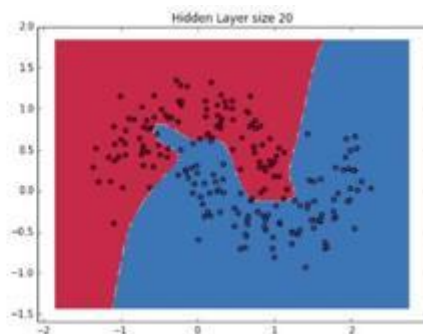
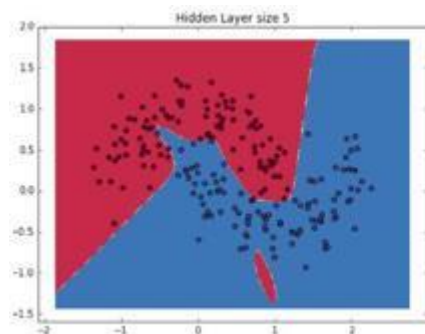
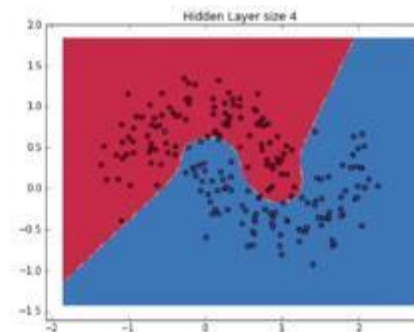
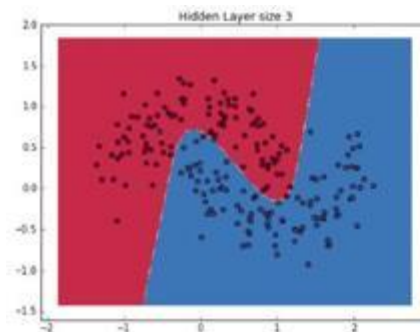
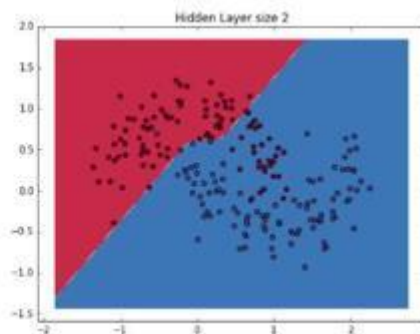
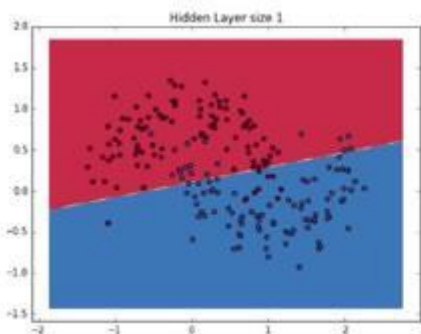
单个神经元能干什么

□ 分类问题



多个神经元能做什么？

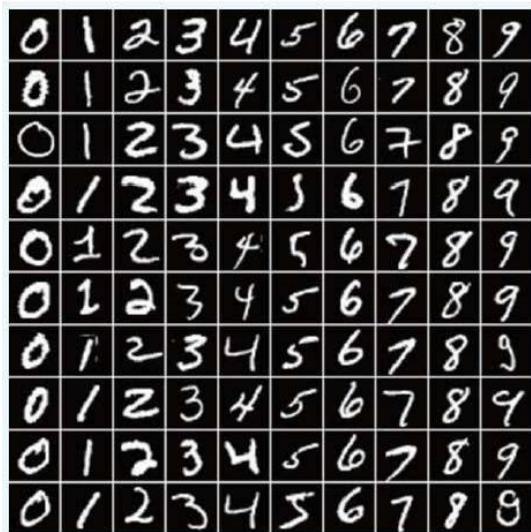
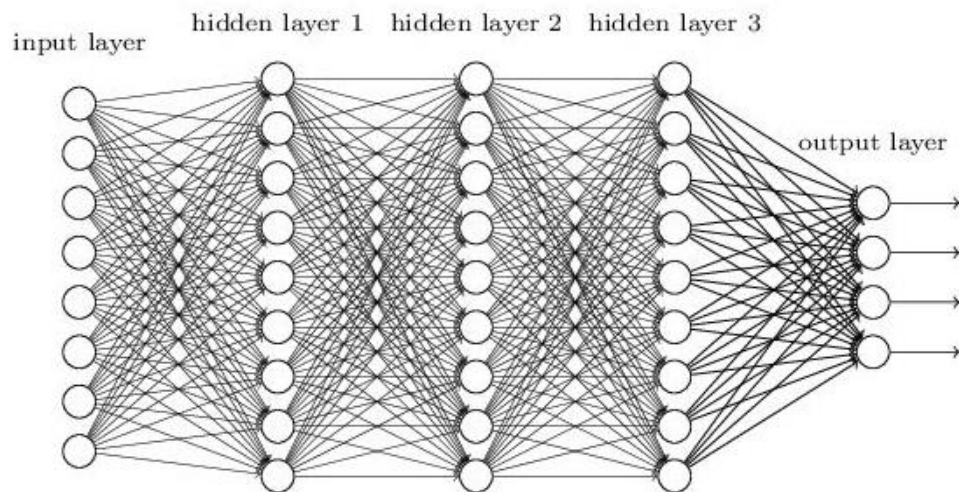
□ 分类问题



多层大规模神经网络的作用

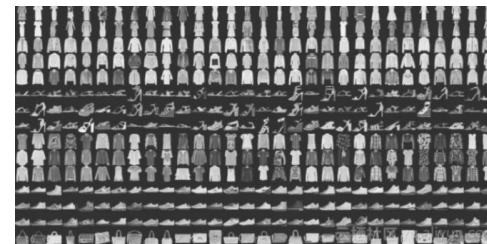
□ 复杂的分类问题

- 垃圾邮件
- 疾病判断
- 猫狗分类

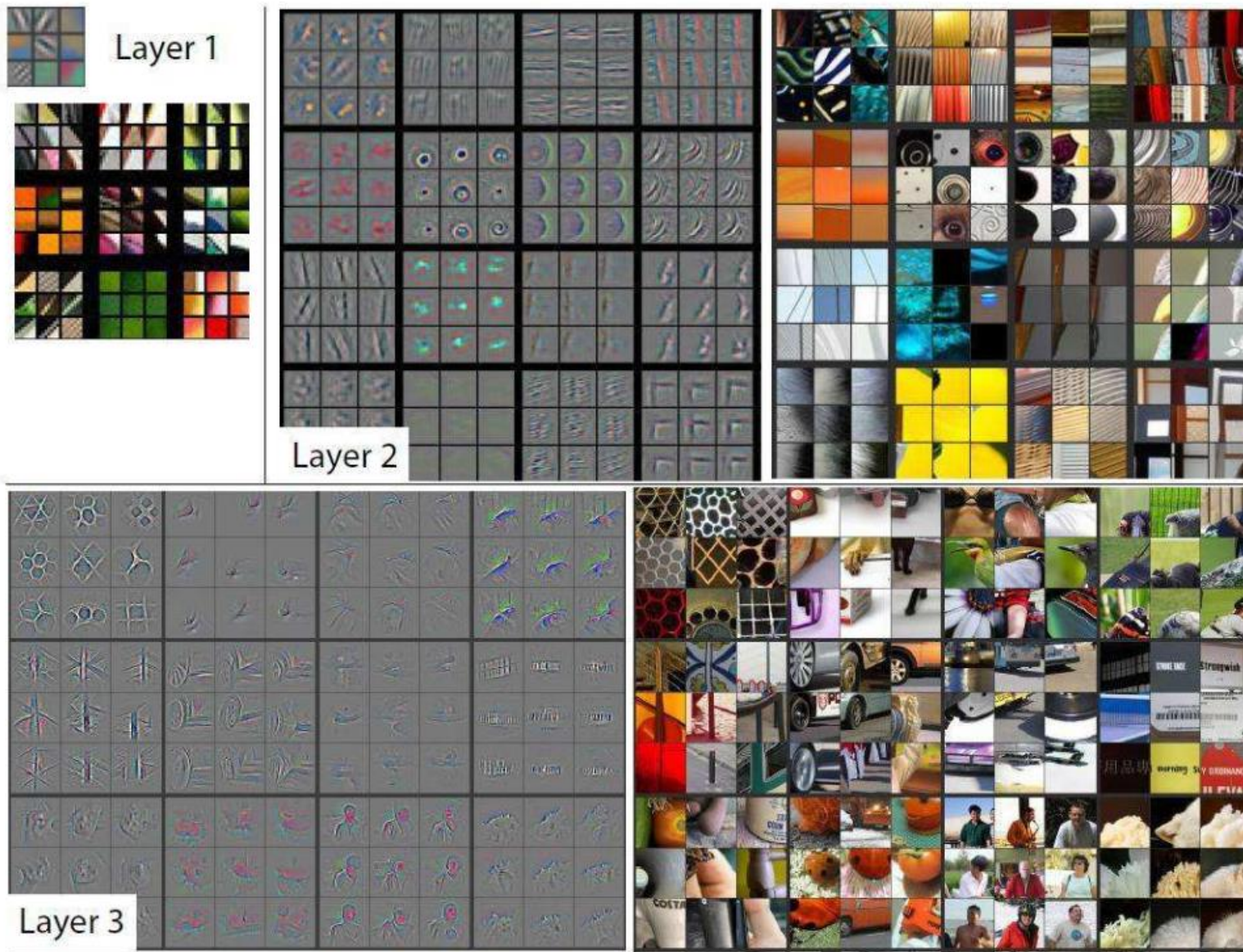


Open Datasets for Deep Learning

MNIST、MS-COCO、ImageNet、OpenImage、VisualQA、SVHN、Cifar10、Fashion—MNIST、IMDB、Twenty Newsgroups、Sentiment140、WordNet、Yelp、WikiPedia、Blogger (图像、视频、文本、社交、推荐)

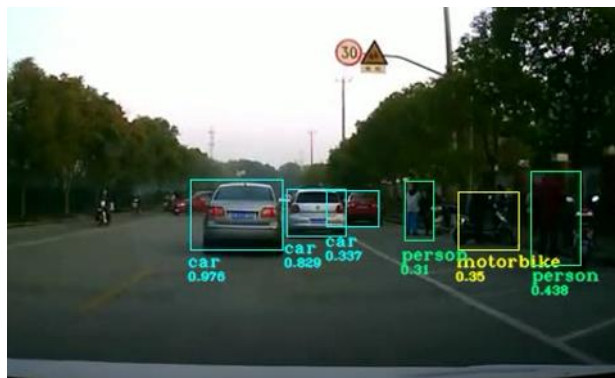


深度学习是如何工作的？



深度学习还能干什么？

- 图像/视频识别
- 自然语言理解
- 语音处理



this small bird has a pink breast and crown, and black primaries and secondaries.

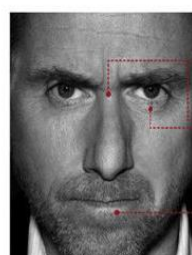


图 1: 内容图像

图 2: 风格图像



Prisma



Lie To Me

深度学习的用途-语音合成

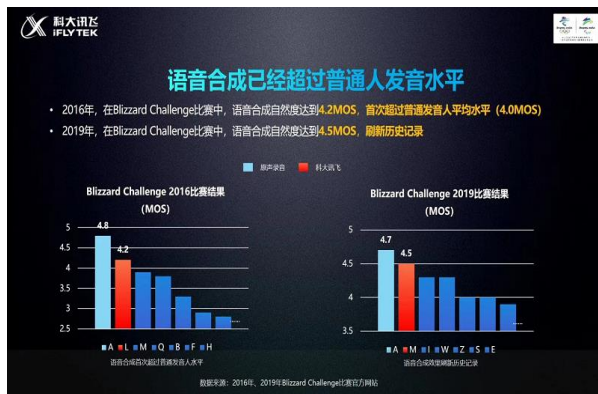


来源：科大讯飞《配音阁》

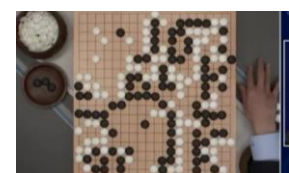
神经网络强大功能



人脸识别准确度超过人类
(ImageNet)



语音合成超过普通人发音水平
语音速记战胜人类专业速记员



AI在多种游戏比赛中夺冠
(桥牌、围棋、及时战略)

LeNet-5

5层CNN模型
用于手写字符识别

AlexNet

ILSVRC 2012冠军
8层CNN

ZFNet

ILSVRC 2013冠军
7层CNN

GoogleNet

ILSVRC 2014冠军
22层CNN

VGGNet

ILSVRC 2014
19层CNN

ResNet

ILSVRC 2015冠军
152层CNN

SqueezeNet

AlexNet简化模型
权值<0.5MB

1998

2012

2013

2014

2015

2017

支撑智能计算的计算机系统

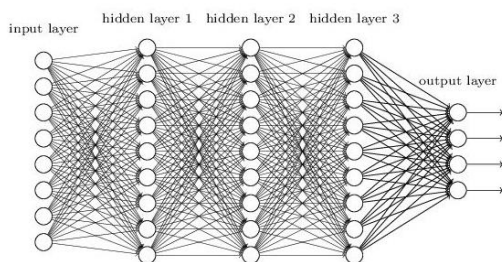
□多个层次

- 1、应用（行业、领域）
- 2、算法（模型、算法）
- 3、编程框架（TensorFlow,Caffe,MXNet, Pytorch）操作系统（OS, 库等）
- 4、硬件（CPU,GPU,ASIC,FPGA,DSP,）

人工智能三大关键要素

大数据时代的 ABC

A算法(Algorithm)+**B数据(B**ig Data)+**C算力(C**omputing Power)
+**D领域(D**omain) +**E环境(E**cosystem)



A算法(Algorithm)



B数据(Big Data)

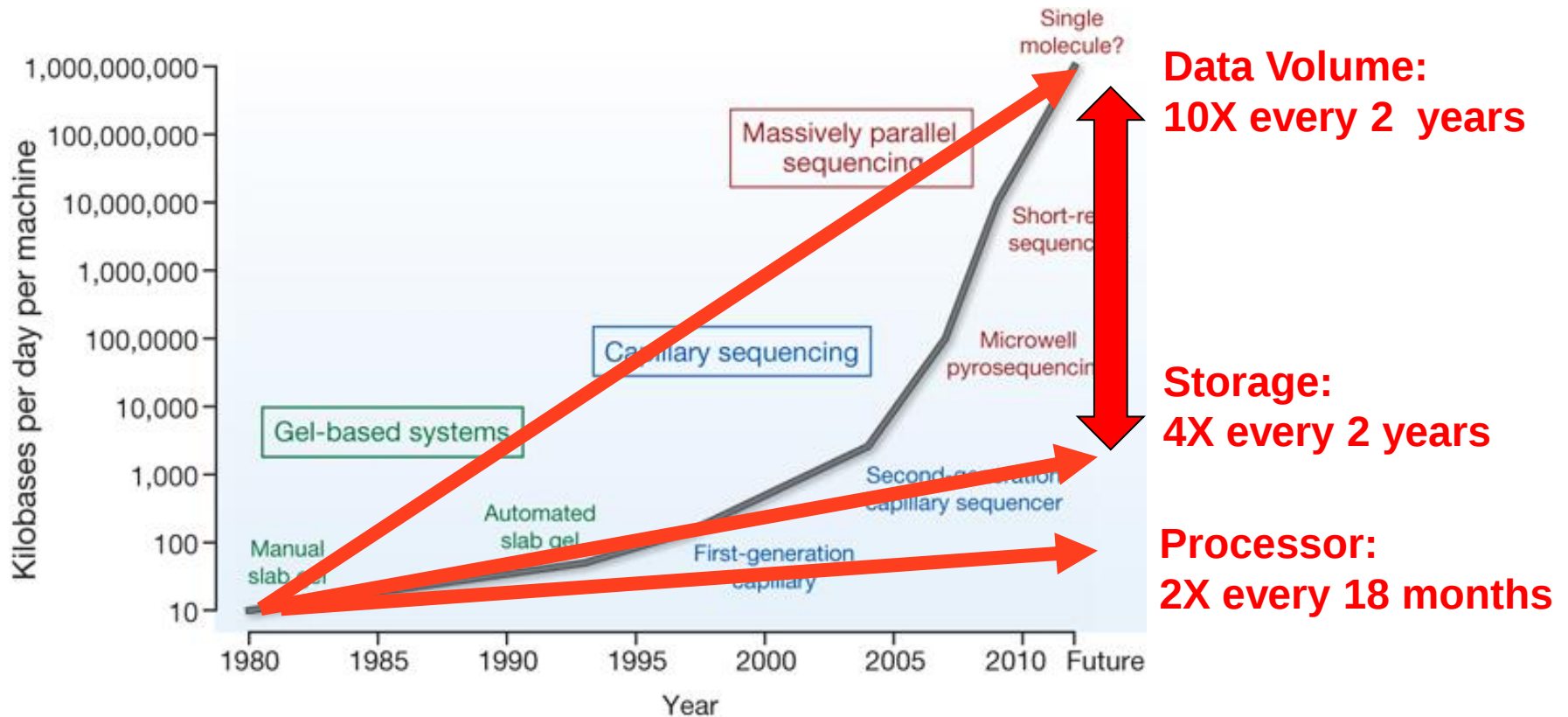


C算力 (Computing Power)

掌握智能计算系统的全栈技术

数据规模驱动系统创新

Explosive Data Scale



The cancer genome, Michael R. Stratton, Peter J. Campbell & P. Andrew Futreal, Nature 458, 719-724(9 April 2009)

nature

1、典型应用

典型应用

网络模型

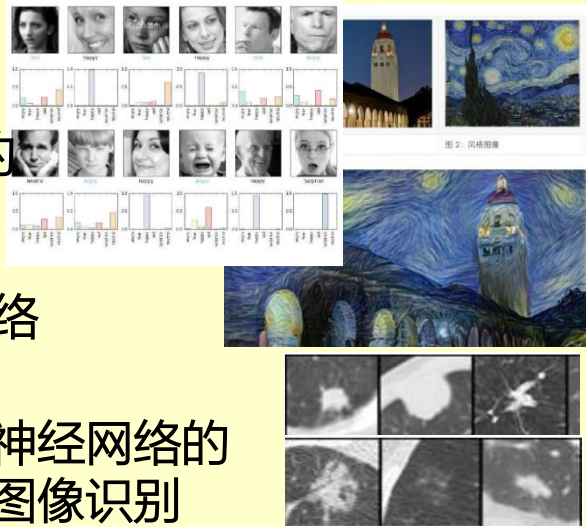
编程框架

硬件平台



基于神经网络的
视频目标检测

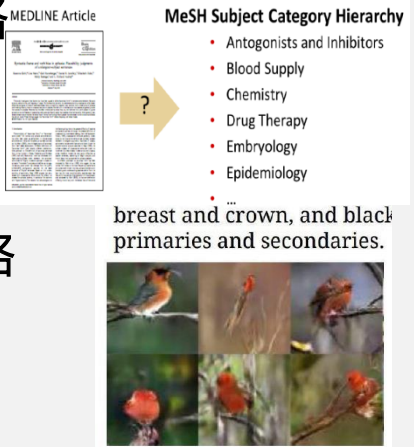
基于神经网络的
模式识别



基于神经网络的面部表情识别

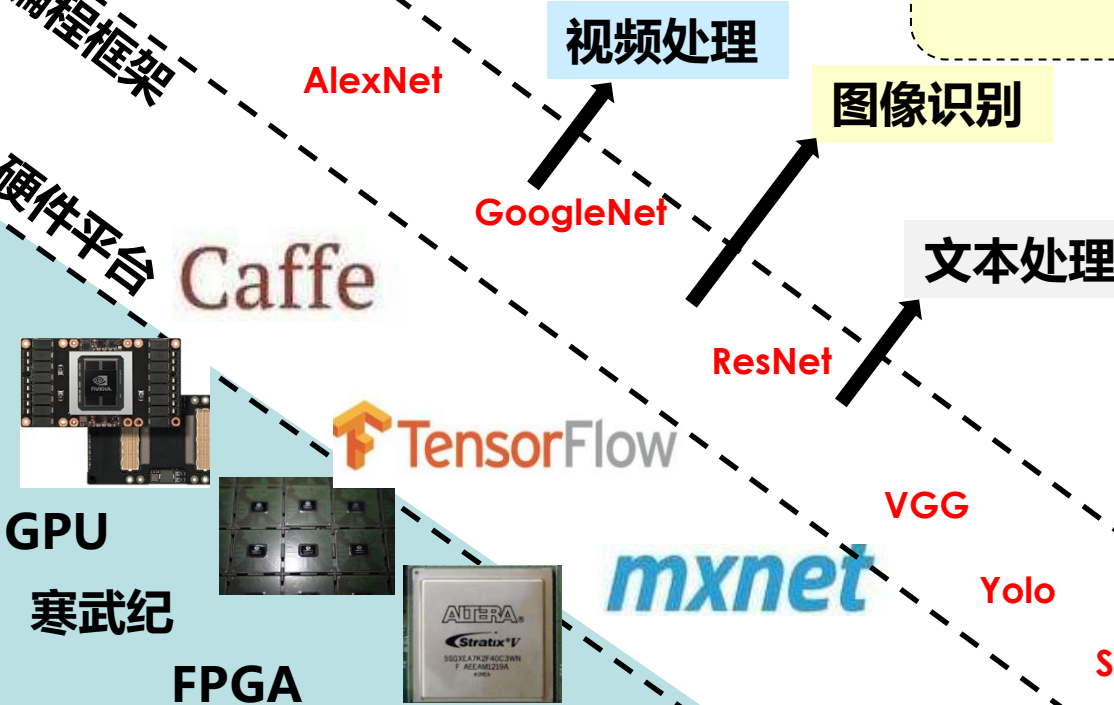
基于神经网络的图像处理

基于神经网络的医学图像识别



基于神经网络的文本分类

基于神经网络的文本摘要



智能计算云平台演示应用

“人脸识别”应用介绍



应用简介
人脸识别应用：基于深度学习神经网络VGG模型，通过人脸图像中提取人脸特征，实现人脸特征提取、人脸特征识别和人脸特征匹配。可用于门禁控制、考勤管理、身份验证等场景。

“物体识别”应用介绍



应用简介
物体识别应用：基于深度学习神经网络SSD模型，通过图像中提取物体特征，实现物体特征提取、物体特征识别和物体特征匹配。可用于垃圾分类、商品识别、图像检索等场景。

“道路检测”应用介绍



应用简介
道路检测应用：基于深度学习神经网络SSD模型，通过图像中提取道路特征，实现道路特征提取、道路特征识别和道路特征匹配。可用于道路检测、交通流量统计、交通事故分析等场景。

“房屋检测”应用介绍



应用简介
房屋检测应用：基于深度学习神经网络VGG模型，通过图像中提取房屋特征，实现房屋特征提取、房屋特征识别和房屋特征匹配。可用于房屋检测、房屋分类、房屋估值等场景。

2、编程模型

□ 编程框架、编程模型、库的发展-动态过程

□ 不同框架之间的区别和联系

Deep learning libraries:
Accumulated GitHub metrics

Aggregate popularity (30•contrib + 10•issues + 5•forks)•1e-3		
#1: 172.29		tensorflow/tensorflow
#2: 89.78		BVLC/caffe
#3: 69.70		fchollet/keras
#4: 53.09		dmlc/mxnet
#5: 38.23		Theano/Theano
#6: 29.86		deeplearning4j/deeplearning4j
#7: 27.99		Microsoft/CNTK
#8: 17.36		torch/torch7
#9: 14.43		baidu/paddle
#10: 13.10		pfnet/chainer
#11: 12.37		NVIDIA/DIGITS
#12: 10.42		tflearn/tflearn
#13: 9.20		pytorch/pytorch

Deep learning libraries: growth over past three months

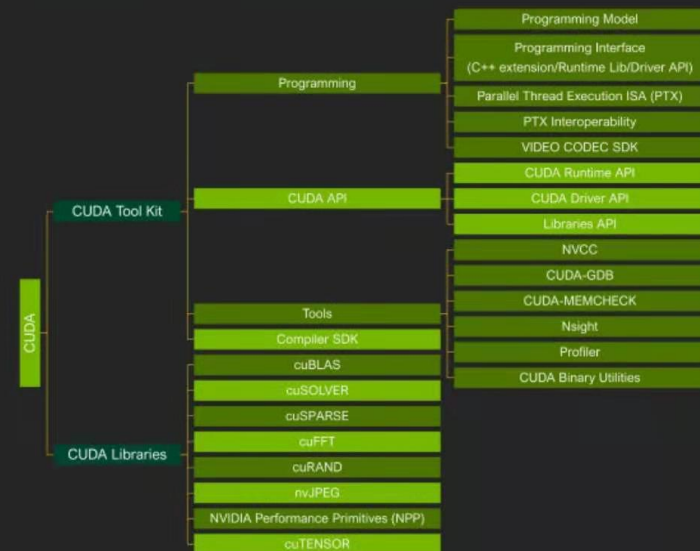
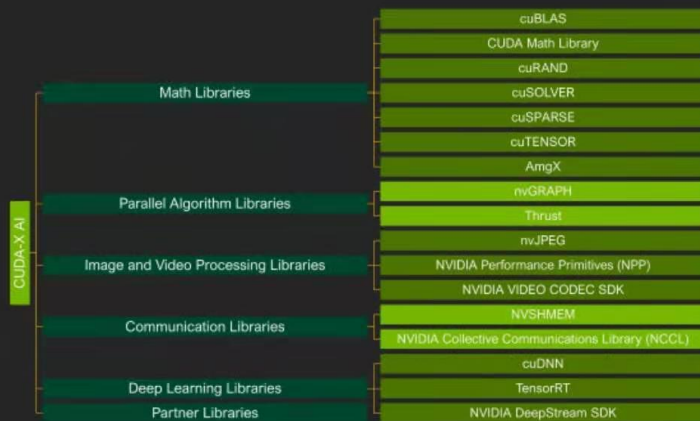
new contributors from 2016-10-09 to 2017-02-10		new forks from 2016-10-09 to 2017-02-10	
#1: 192	tensorflow/tensorflow	#1: 6525	tensorflow/tensorflow
#2: 89	dmlc/mxnet	#2: 1822	BVLC/caffe
#3: 78	fchollet/keras	#3: 1316	fchollet/keras
#4: 42	baidu/paddle	#4: 999	dmlc/mxnet
#5: 29	Microsoft/CNTK	#5: 909	deeplearning4j/deeplearning4j
#6: 23	pfnet/chainer	#6: 887	Microsoft/CNTK
#7: 21	Theano/Theano	#7: 324	tflearn/tflearn
#8: 20	deeplearning4j/deeplearning4j	#8: 321	baidu/paddle
#9: 20	tflearn/tflearn	#9: 287	Theano/Theano
#10: 19	BVLC/caffe	#10: 257	torch/torch7
#11: 9	torch/torch7	#11: 175	NVIDIA/DIGITS
#12: 3	NVIDIA/DIGITS	#12: 142	pfnet/chainer

new issues from 2016-10-09 to 2017-02-10		Aggregate metrics growth from 2016-10-09 to 2017-02-10	
#1: 1563	tensorflow/tensorflow	#1: 54.01	tensorflow/tensorflow
#2: 979	fchollet/keras	#2: 18.71	fchollet/keras
#3: 871	dmlc/mxnet	#3: 16.38	dmlc/mxnet
#4: 646	baidu/paddle	#4: 12.86	BVLC/caffe
#5: 486	Microsoft/CNTK	#5: 10.17	Microsoft/CNTK
#6: 361	deeplearning4j/deeplearning4j	#6: 9.32	baidu/paddle
#7: 318	BVLC/caffe	#7: 8.75	deeplearning4j/deeplearning4j
#8: 217	NVIDIA/DIGITS	#8: 4.23	Theano/Theano
#9: 214	Theano/Theano	#9: 3.89	tflearn/tflearn
#10: 167	tflearn/tflearn	#10: 3.14	NVIDIA/DIGITS
#11: 150	pfnet/chainer	#11: 2.90	pfnet/chainer
#12: 90	torch/torch7	#12: 2.46	torch/torch7

NVIDIA CUDA-X AI ECOSYSTEM



SOURCE: NVIDIA.COM



2、编程语言受欢迎情况（1986-2021）

Very Long Term History

To see the bigger picture, please find below the positions of the top 10 programming languages for a period of 12 months.

Programming Language	2021	2016	2011	20
C	1	2	2	1
Java	2	1	1	2
Python	3	5	6	7
C++	4	3	3	3
C#	5	4	5	6
JavaScript	6	7	10	9
PHP	7	6	4	4
R	8	16	41	-
SQL	9	-	-	-
Swift	10	15	-	-
Perl	14	9	7	5
Lisp	29	26	14	13
Ada	33	24	21	16

Programming Language Hall of Fame

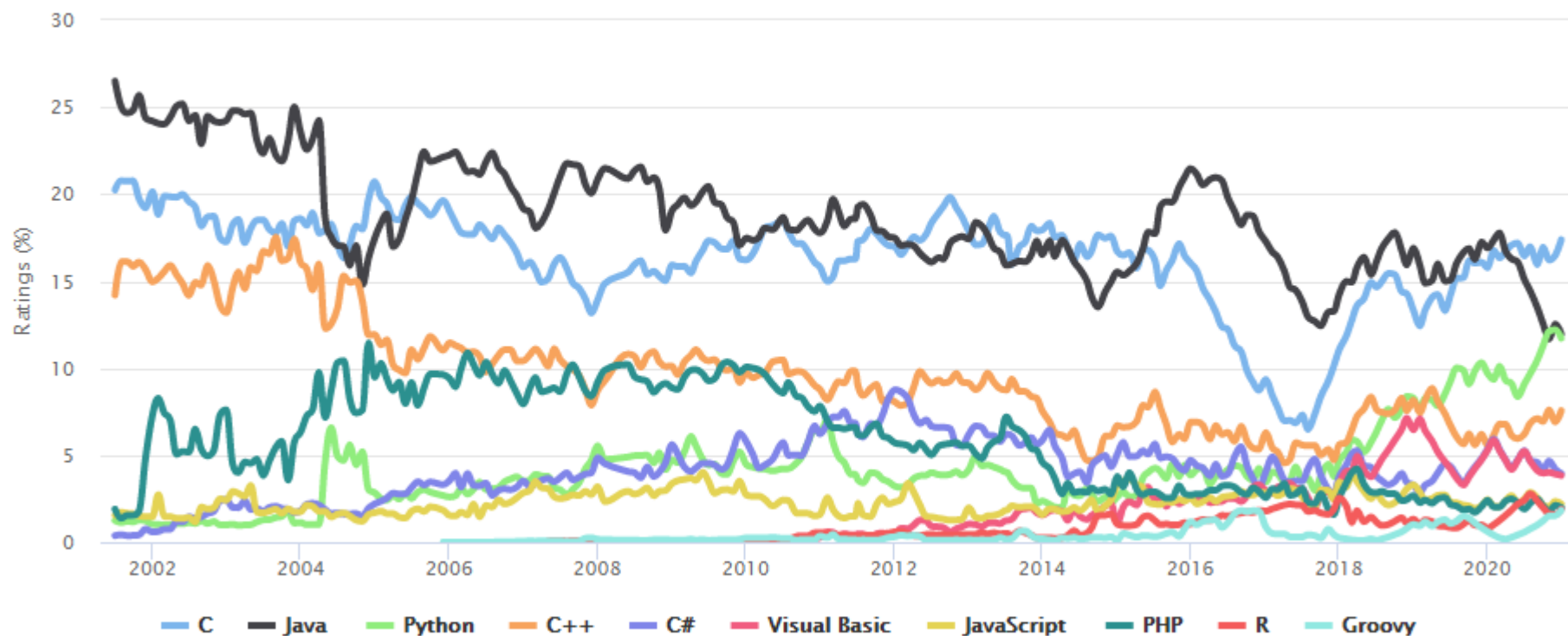
The hall of fame listing all "Programming Language of the Year" award winners is shown below. The language that has the highest rise in ratings in a year.

Year	Winner
2019	🏆 C
2018	🏆 Python
2017	🏆 C
2016	🏆 Go
2015	🏆 Java
2014	🏆 JavaScript
2013	🏆 Transact-SQL
2012	🏆 Objective-C
2011	🏆 Objective-C
2010	🏆 Python
2009	🏆 Go
2008	🏆 C
2007	🏆 Python
2006	🏆 Ruby
2005	🏆 Java
2004	🏆 PHP
2003	🏆 C++

2、编程语言指数走势(2002-2020)

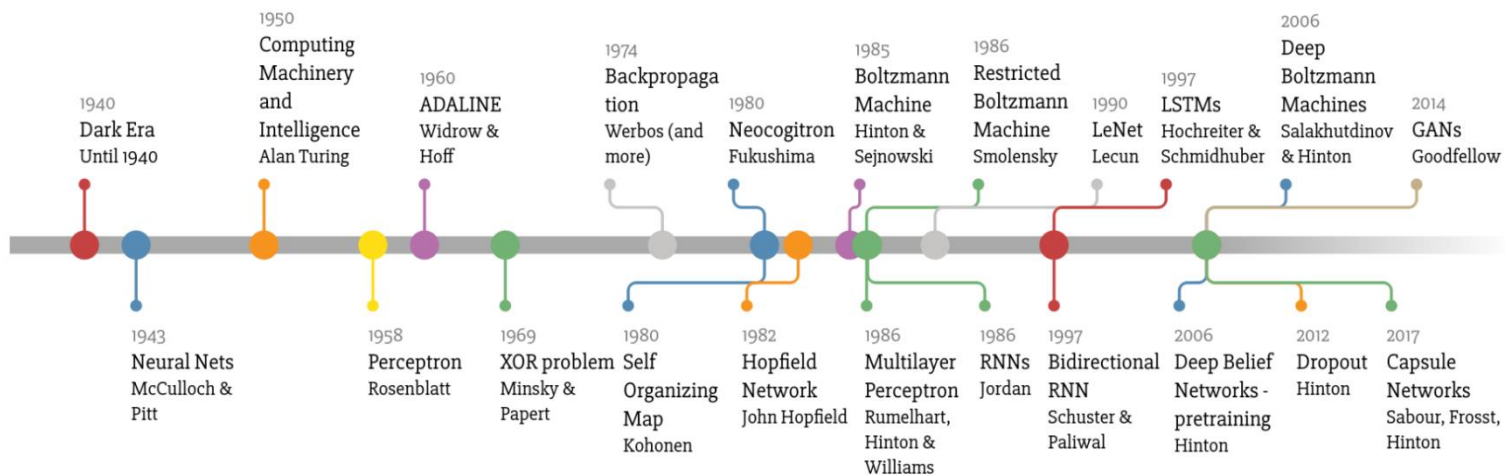
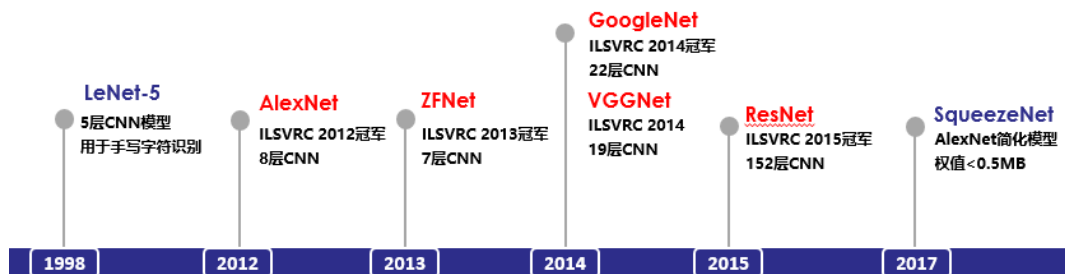
TIOBE Programming Community Index

Source: www.tiobe.com



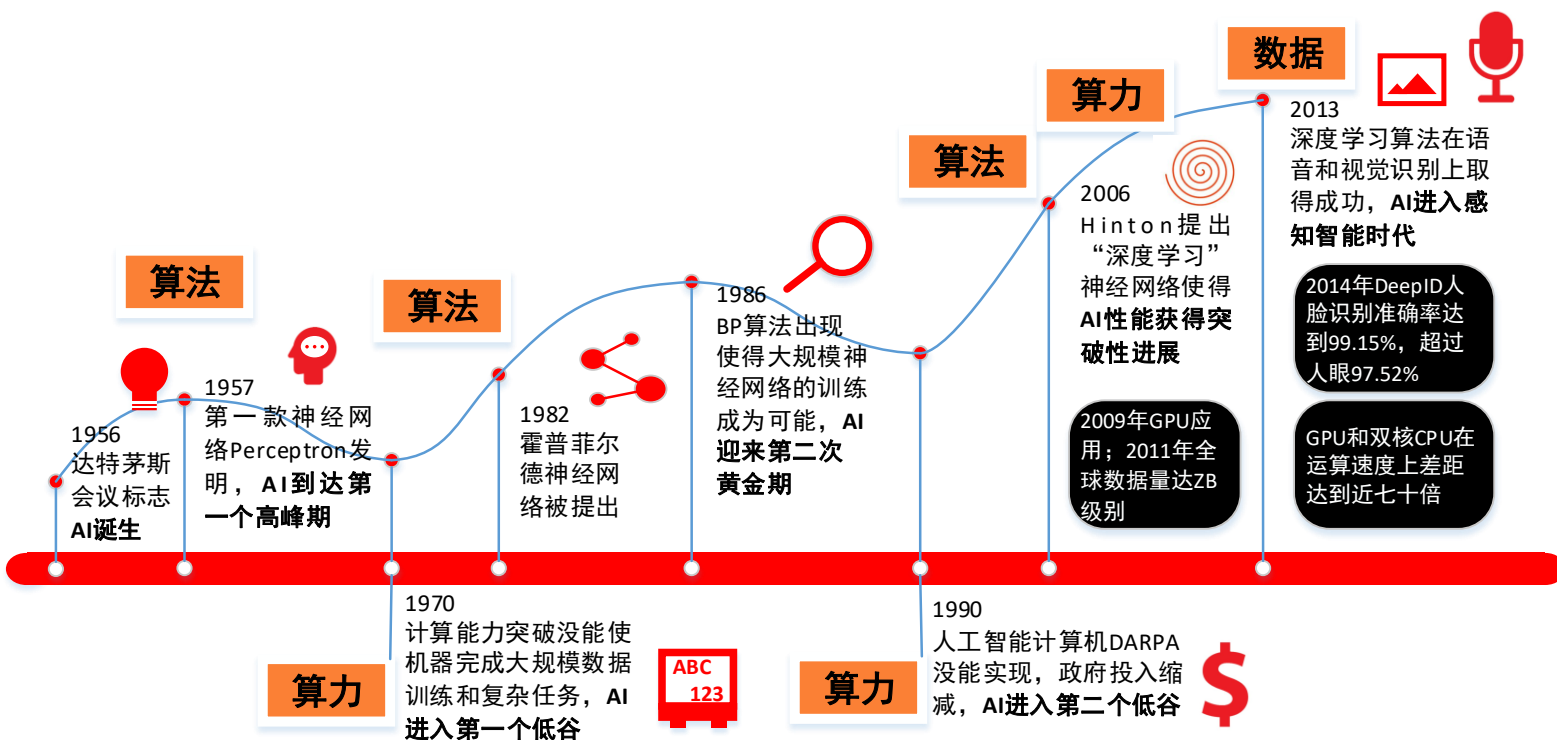
3、神经网络拓扑结构

- 网络面向的应用
- 不同版本
- 在不同平台上如何实现
- 有什么特征，优点缺点
- 是否有更好的解决方案



4、智能硬件平台

□人工智能发展历史：每一次算法突破迎来一波高潮，因算力限制陷入低谷



如何提升算力，降低能效是目前智能计算系统的设计面临的关键问题

4、提升算力的有效途径—智能加速器

基于GPU的加速器

• NVIDIA P100--V100--A100



优势

• 高性能，编程容易

缺点

• 功耗高

基于ASIC的加速器

• DianNao 寒武纪

• Google TPU

• 华为昇腾910



优势

• 高性能低功耗

缺点

• 周期长

基于FPGA的加速器

• Intel+Altera DLIA

• Microsoft Catapult

• Xilinx DeepHi



优势

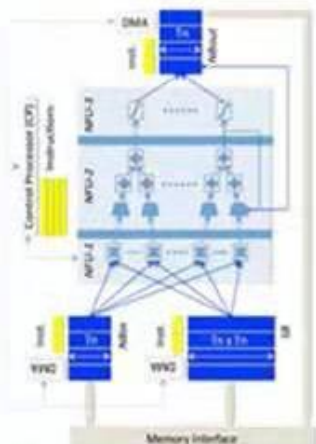
• 高性能、低功耗、高灵活性

缺点

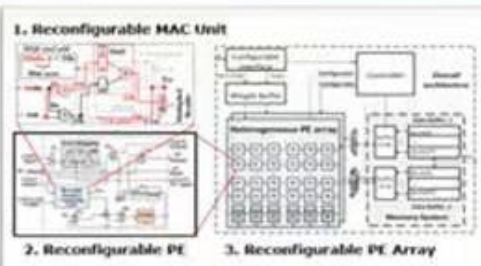
• 编程困难

4、智能芯片简单分类

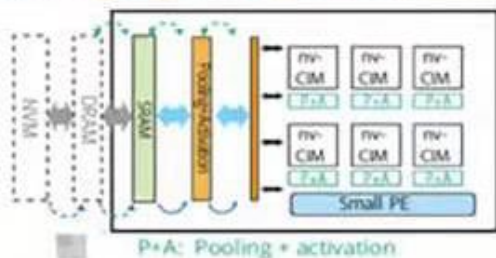
指令集架构处理器 (例如：GPU，MLU)



可重构处理器 (例如：Thinker)

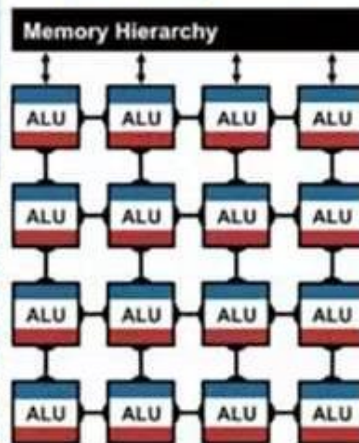


DNN处理器



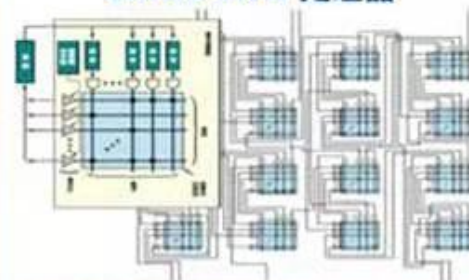
存内计算处理器 (例如：RRAM， SRAM，FLASH)

数据流处理器 (例如：Eyeriss)

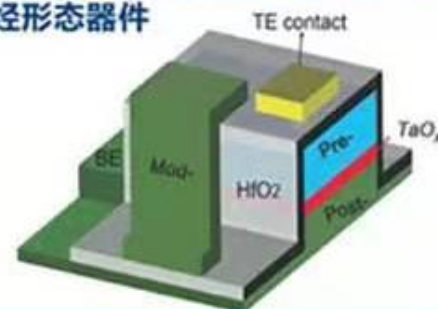


神经形态处理器

CMOS SNN 处理器



神经形态器件



小结：智能计算系统的几个特征

- 1、从通用到专用(智能计算是领域计算的一种形式)
- ISCA 2018, 2017 图灵奖, John L. Hennessy和 David A. Patterson

“A New Golden Age for Computer Architecture:

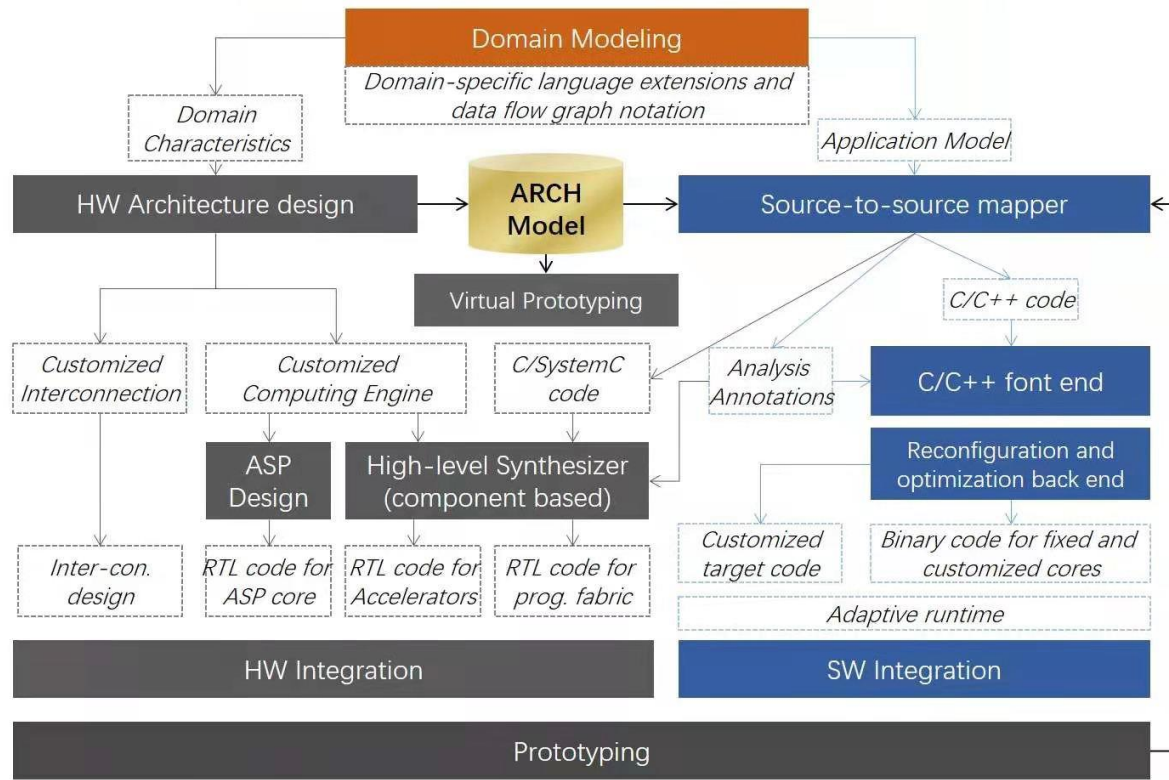
Domain-Specific Hardware/Software Co-Design, Enhanced Security, Open Instruction Sets, and Agile Chip Development,”



小结：智能计算系统的几个特征

□2、多层次全栈知识

领域专用计算是需要软硬件的全栈支持



Source: J. Cong, V. Sarkar, G. Reinman and A. Bui, "Customizable domain-specific computing", *IEEE Des. Test Comput.*, Mar. Apr. 2011.



Jason Cong
UCLA

AutoESL

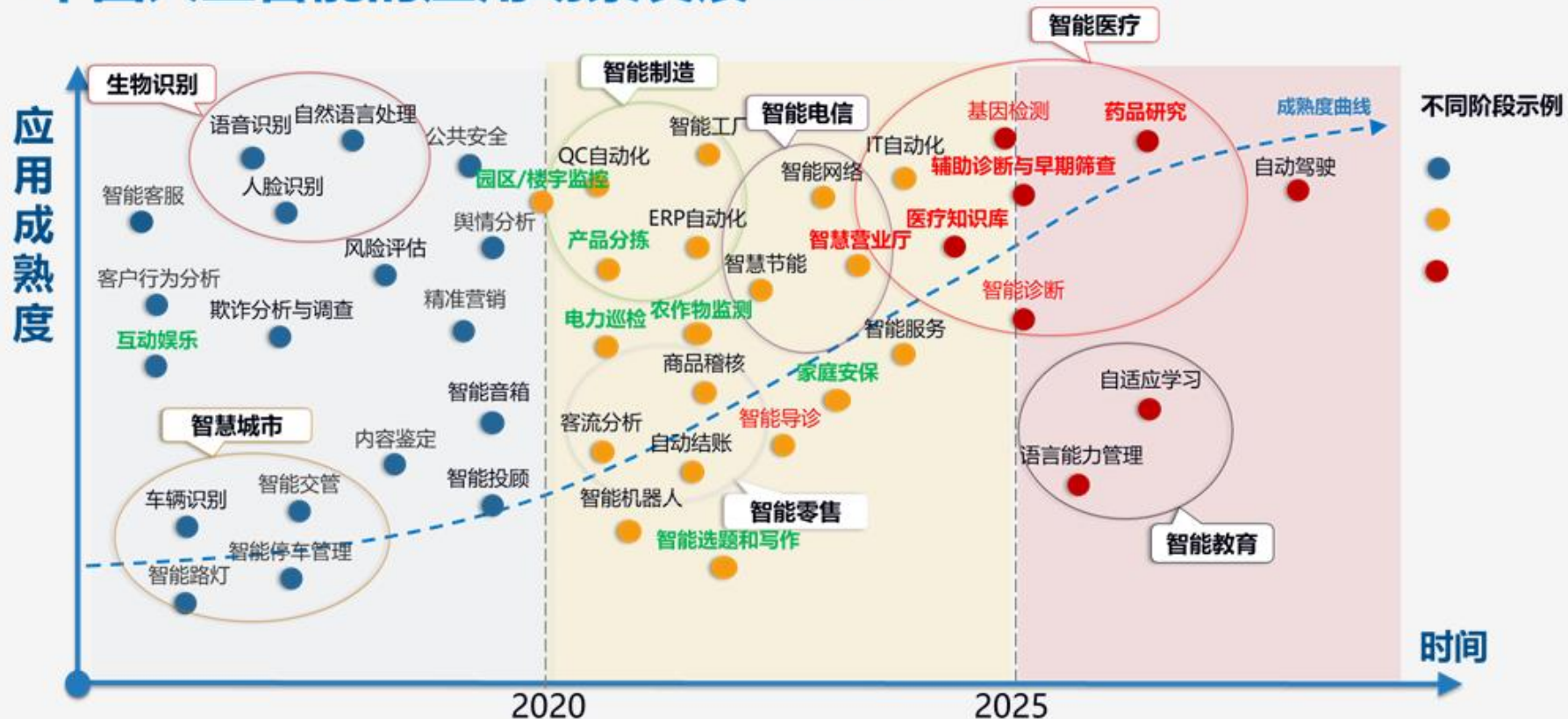
硬

软

小结：智能计算系统的几个特征

□3、交叉学科很关键-“四个面向”

中国人工智能的应用场景发展 2020



来源：IDC中国人工智能基础架构市场跟踪报告，2020H1

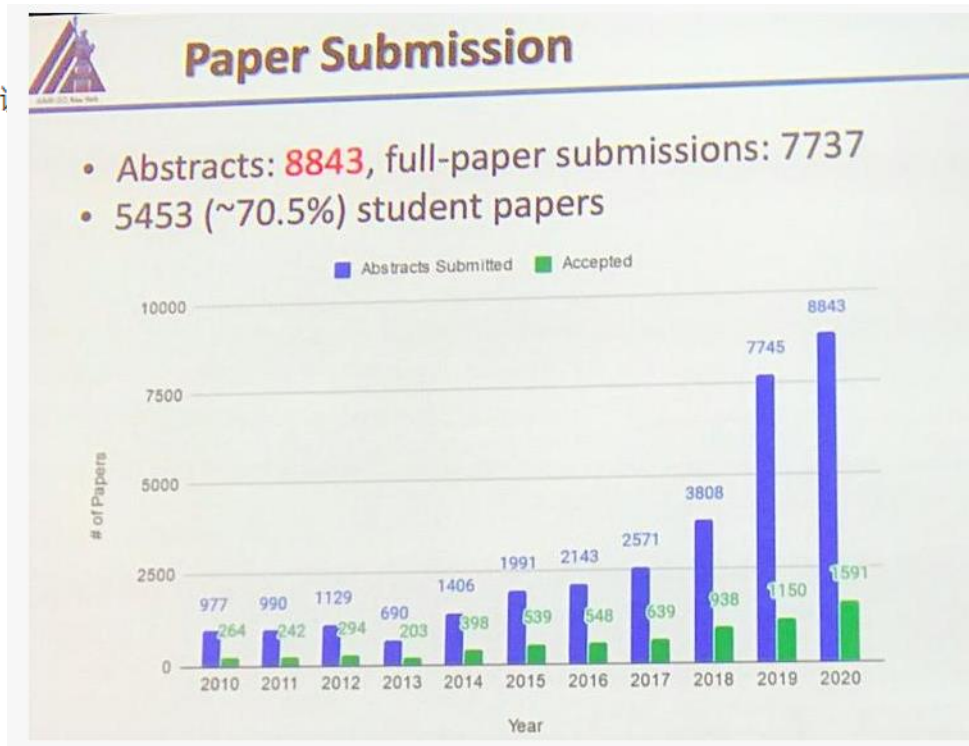
小结：智能计算系统的几个特征

□4、快速发展、快速迭代-软件、算法、模型

中国计算机学会推荐国际学术会议
(• 人工智能)

A类

序号	刊物名称	刊物全称	出版社
1	AAAI	AAAI Conference on Artificial Intelligence	AAAI
2	NeurIPS	Annual Conference on Neural Information Processing Systems	MIT Press
3	ACL	Annual Meeting of the Association for Computational Linguistics	ACL
4	CVPR	IEEE Conference on Computer Vision and Pattern Recognition	IEEE
5	ICCV	International Conference on Computer Vision	IEEE
6	ICML	International Conference on Machine Learning	ACM
7	IJCAI	International Joint Conference on Artificial Intelligence	Morgan Kaufmann



CVPR 2020: 主会总计1467篇录用论文

ICML 2020 收到4990篇投稿，录用1088

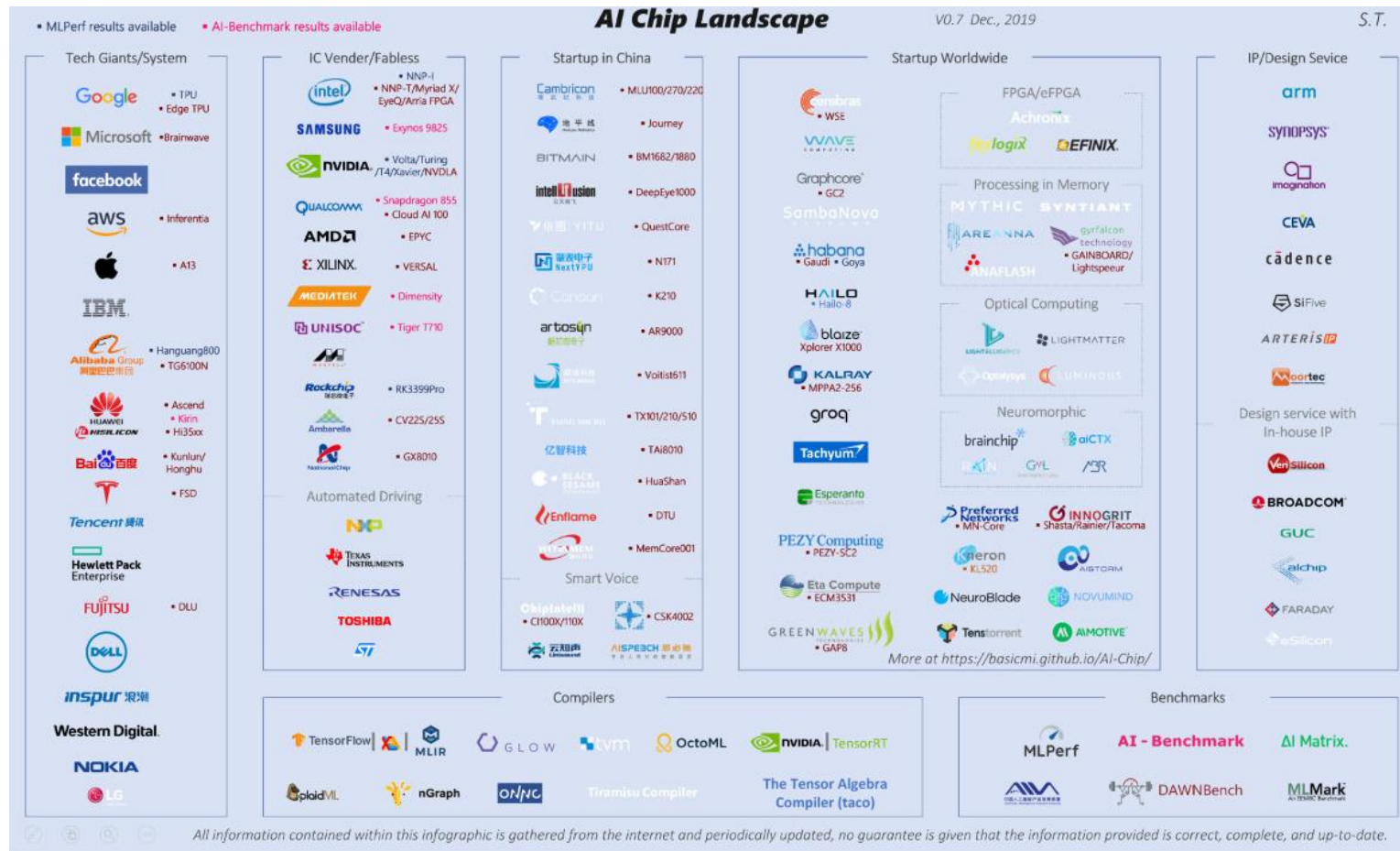
IJCAI 2020共收到5147篇投稿，录用592篇

AAAI 2020 收到7737投稿，录用1591篇

小结：智能计算系统的几个特征

4、快速发展、快速迭代-芯片 大量Startup

非传统芯片厂商



工具链EDA

结论：提高快速学习能力（科研+工程）

新Benchmark

快速自我学习的工具、网站

□ 中国计算机学会



□ IEEE数据库



□ ACM数据库



□ DBLP数据库



□ Google Scholar 数据库



□ 课题组、个人网站等

□ Arxiv等



□ Bilibili



□ 工具： Word Excel PPT LateX Endnote



Thank you

王 超
中国科学技术大学
cswang@ustc.edu.cn