基于云际计算的分布式深度学习及其在声学建模中的应用

国防科技大学

并行与分布计算国防科技重点实验室

许可乐

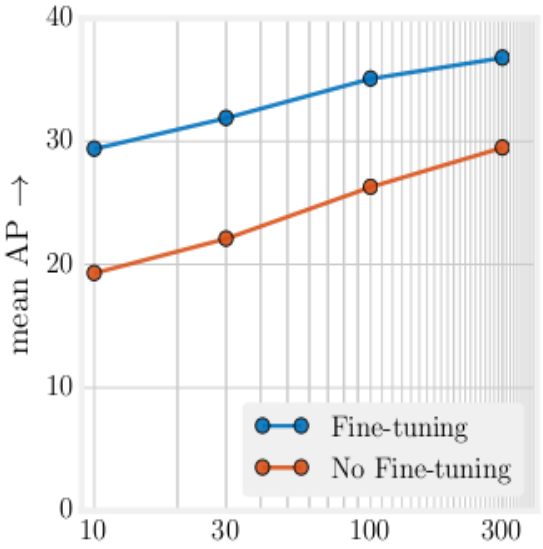
大纲

- 背景
- 基于云际计算的分布式深度学习最新进展
  - 分布式深度学习方法
  - 声学建模中的应用
  - 其它应用场景
- 小结

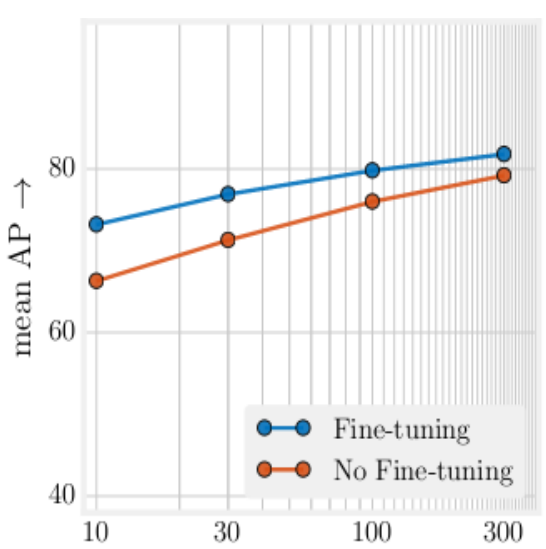
基于云际计算的分布式深度学习背景

- 机器学习模型性能随训练数据量呈线性增长

COCO 目标检测



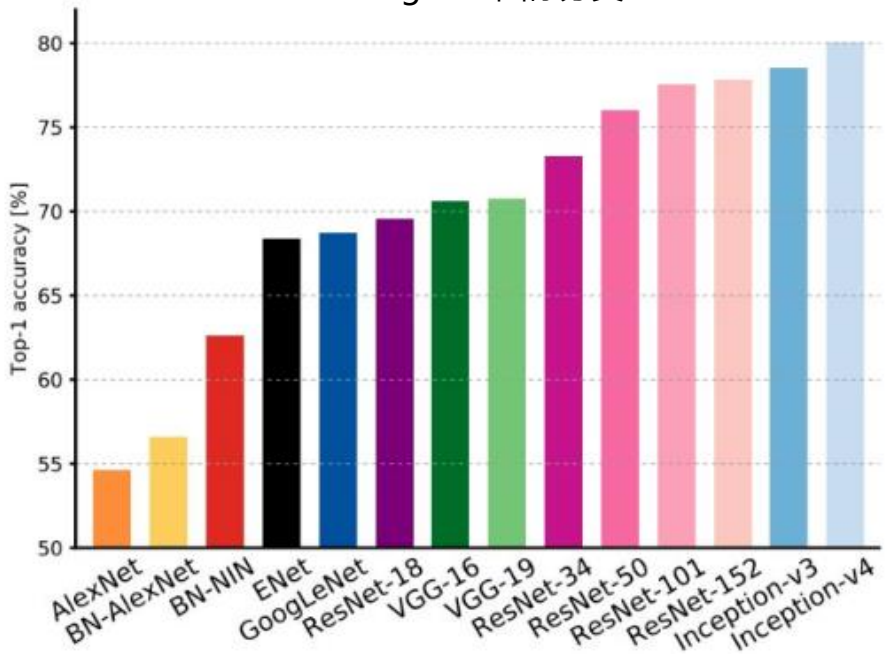
PASCAL VOC2007 目标检测



数据量 VS 目标检测性能

- 机器学习模型性能随模型大小和深度的增长而增长

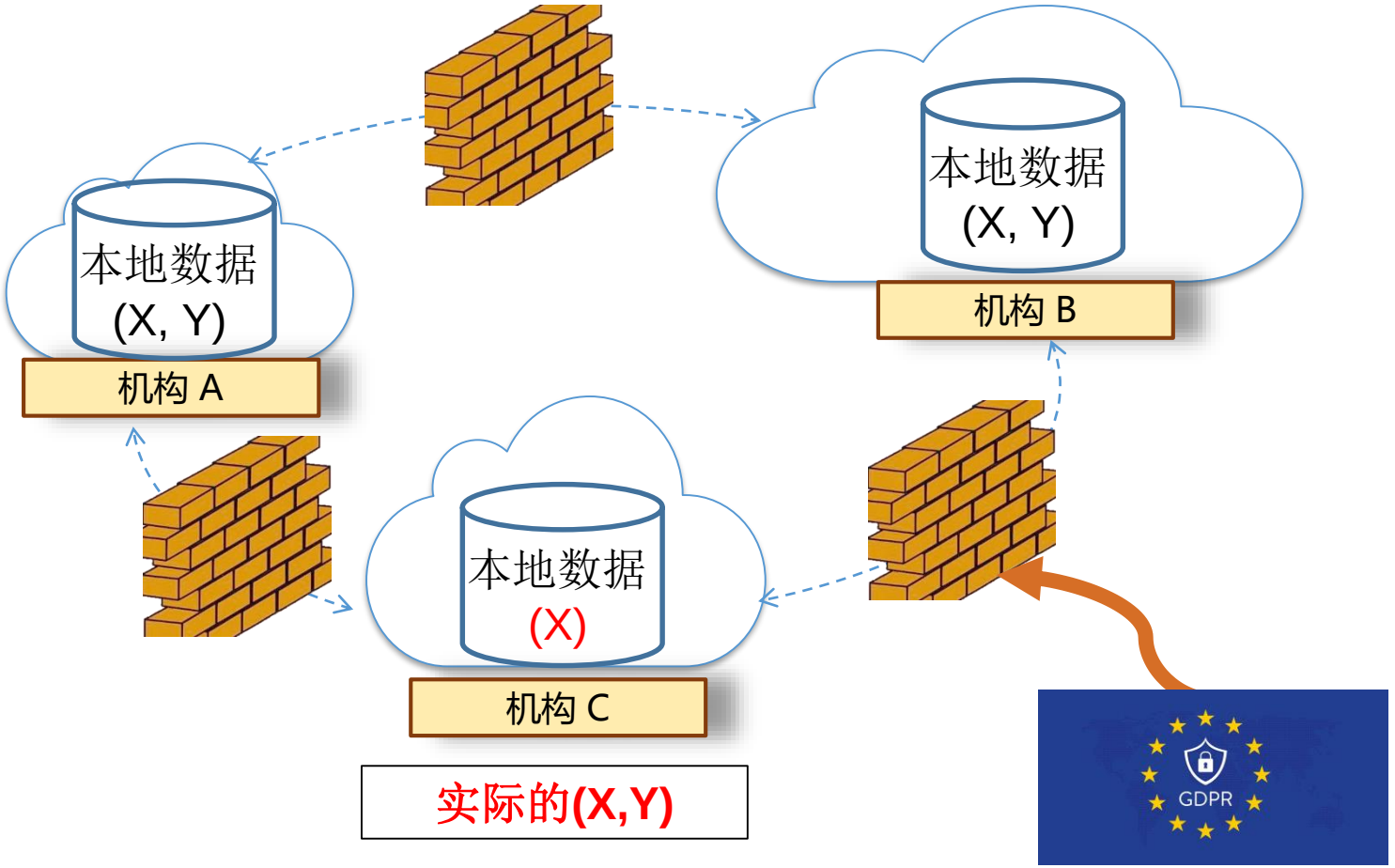
ImageNet图像分类



模型深度 VS 图像分类性能

基于云际计算的分布式深度学习背景

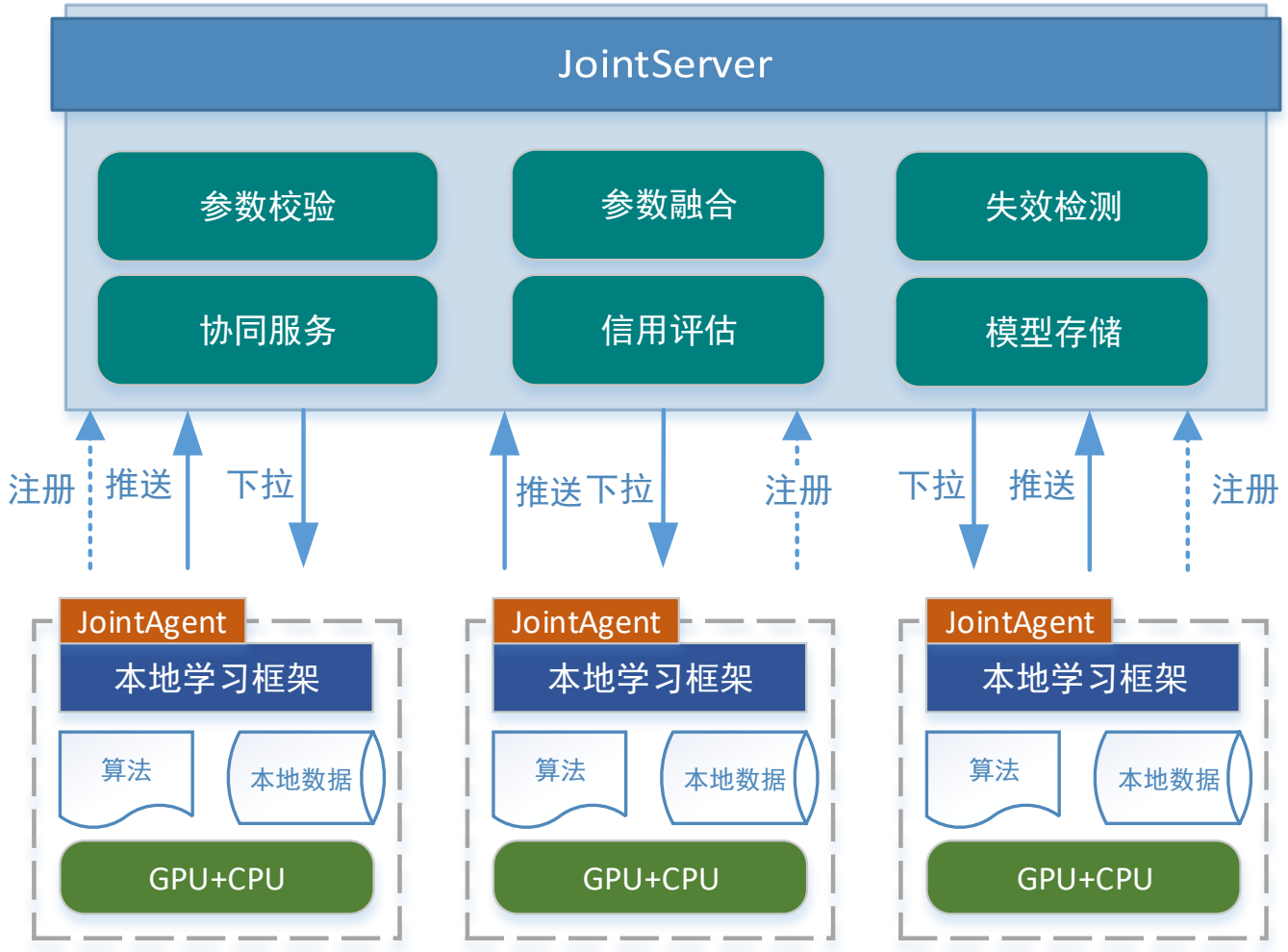
- 机器学习模型的性能依赖于大量高质量的标记数据**(X,Y)**
- 实际的标记数据分布不均，质量差异性大
- 数据以孤岛形式存在
- 数据聚合成本大，风险高



基于云际计算的分布式深度学习背景

- **目的**：数据拥有方在不用给出己方数据的情况下，也可进行模型训练得到模型的计算过程，并能够保证模型的效果与数据集中式的模型效果之间的差距足够小。

- **云际分布学习定义**：在确保各机构数据隐私的前提下，数据所有方利用本地算力进行协作学习，学习过程中彼此之间不交换原始数据，仅需共享训练的中间结果，即可达到将数据集中起来训练后的模型性能。

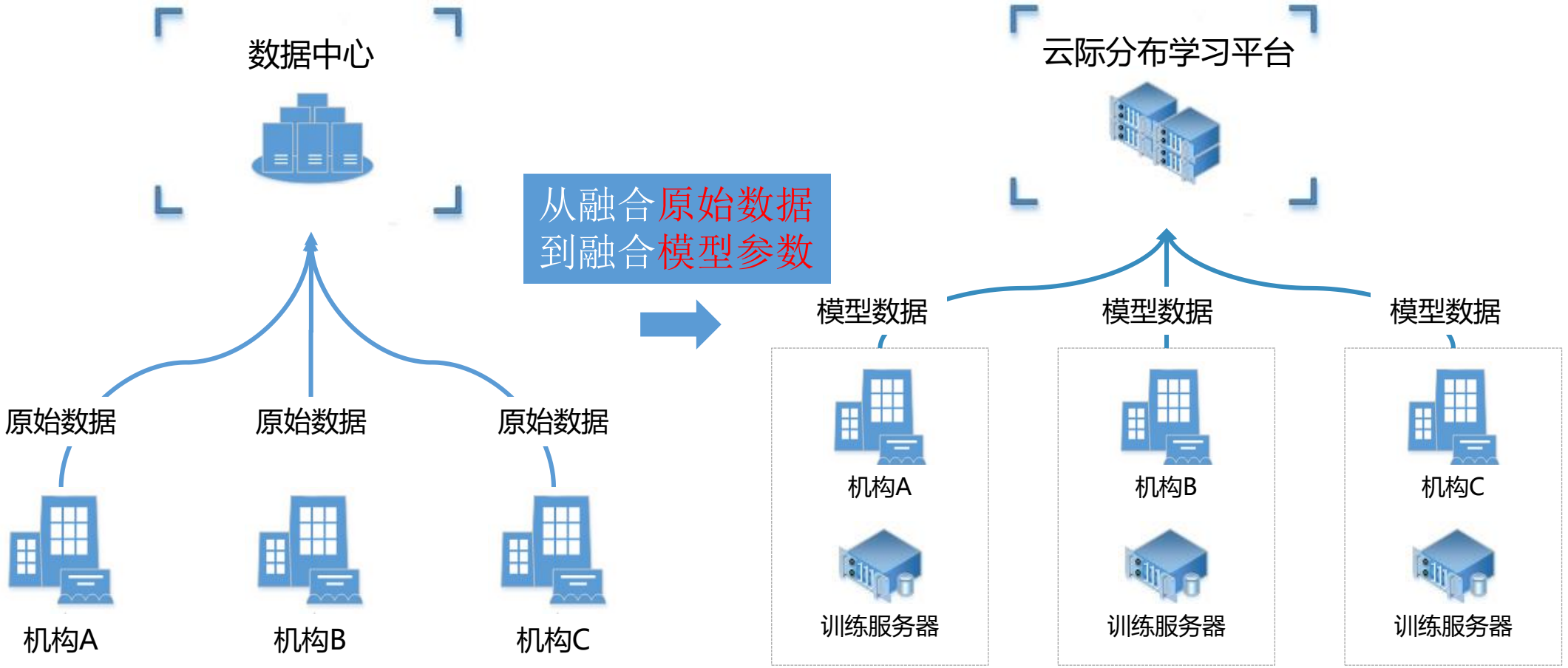


大纲

- 应用场景
- **基于云际计算的分布式深度学习最新进展**
  - 分布式深度学习方法
    - ✓ 模型平均及其改进
    - ✓ 模型蒸馏
  - 声学建模中的应用
  - 其它应用场景
- 小结

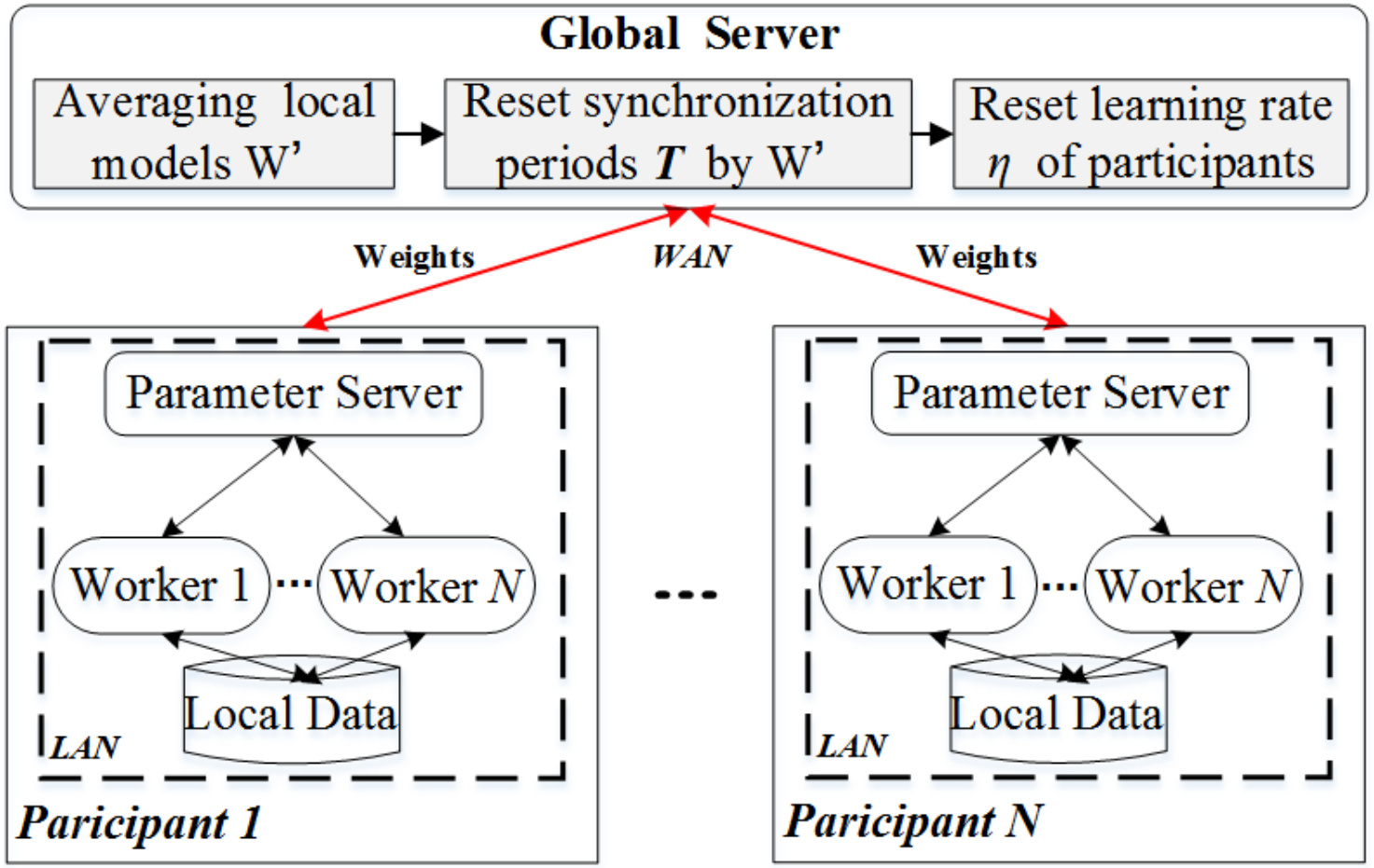
基于云际计算的分布式深度学习最新进展

- 交换模型参数——模型平均



分布式深度学习方法——模型平均

- 传统模型平均的问题
  - **带宽需求大，每1个epoch同步一次**
  - **不同子模型差异性较小，容易陷入局部最优，性能较差**
  - 无法充分利用大量无标签数据

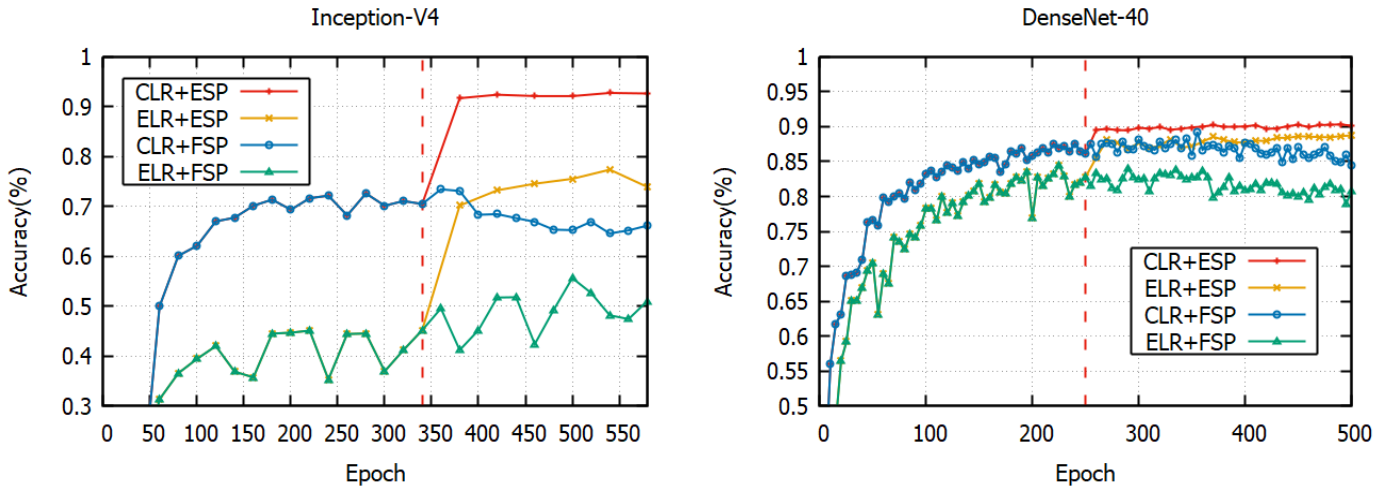


基于模型平均的云际协作学习

[1]15-ICLR-Parallel training of DNNs with natural gradient and parameter averaging

分布式深度学习方法——模型平均的改进

- 方法
  - 扩大局部训练周期(ESP)
  - 使用循环学习率(CLR)
- 目的
  - 有效降低带宽
  - 避免学习过程中的局部最优
  - 提高分布式学习性能
- 结果
  - 带宽降低
  - 性能提升



Models	Comm. interval (min. / T_0)	Comm. volume (MB)
DenseNet-40	4.5 / 5	13
ResNet-152	30 / 5	223
Inception-V4	60 / 20	168
Inception-ResNet-V2	27.5 / 5	218

Xu, Kele, et al. "Collaborative Deep Learning Across Multiple Data Centers." *arXiv preprint arXiv:1810.06877* (2018).

分布式深度学习方法——模型平均的改进

- 在图像、声学和文本三个领域对云际分布学习平台验证
  - 6类开源数据集、11类主流模型

云际分布学习模型与全量数据模型性能一致

图像 (ImageNet)	全量数据模型精度	云际分布学习模型精度
VGG-19	0.8944	0.8964
Resnet-152	0.9264	0.9289
Inception-V4	0.9134	0.9207
Inception-Resnet-V2	0.9286	0.9224
Densenet-40	0.9135	0.9143

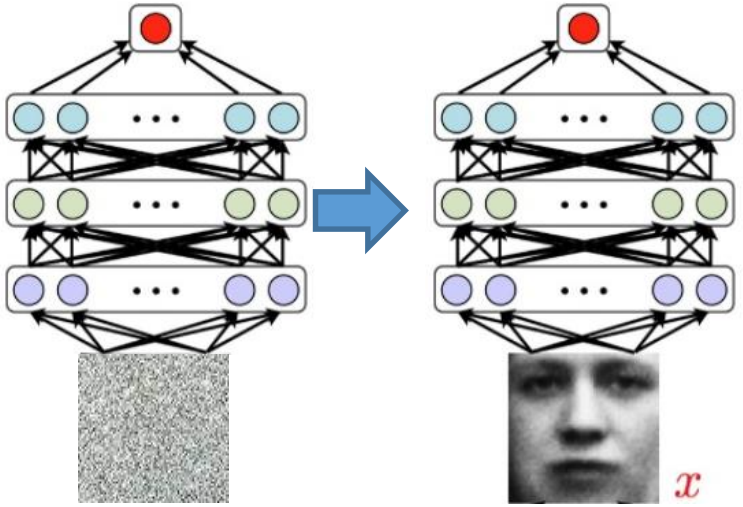
文本 (Toxic Comment)	全量数据模型精度	云际分布学习模型精度
LSTM	98.52	98.57
CapsuleNet	98.32	98.75

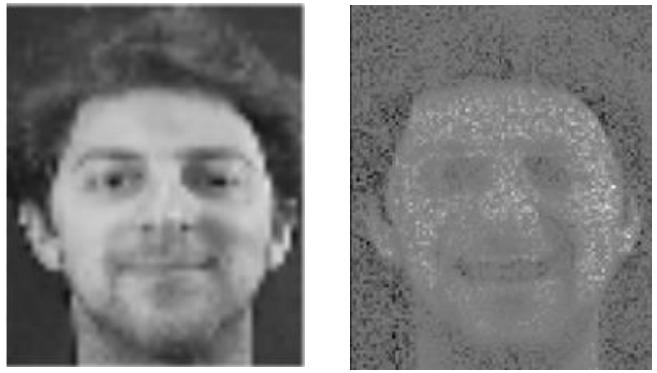
语音 (AudioSet)	全量数据模型AUC	云际分布学习模型AUC
Average Pooling	0.964	0.962
Max Pooling	0.960	0.959
Single Attention	0.968	0.966
Multi Attention	0.968	0.968

Xu, Kele, et al. "Collaborative Deep Learning Across Multiple Data Centers." *arXiv preprint arXiv:1810.06877* (2018).

分布式深度学习方法——模型平均的隐私问题

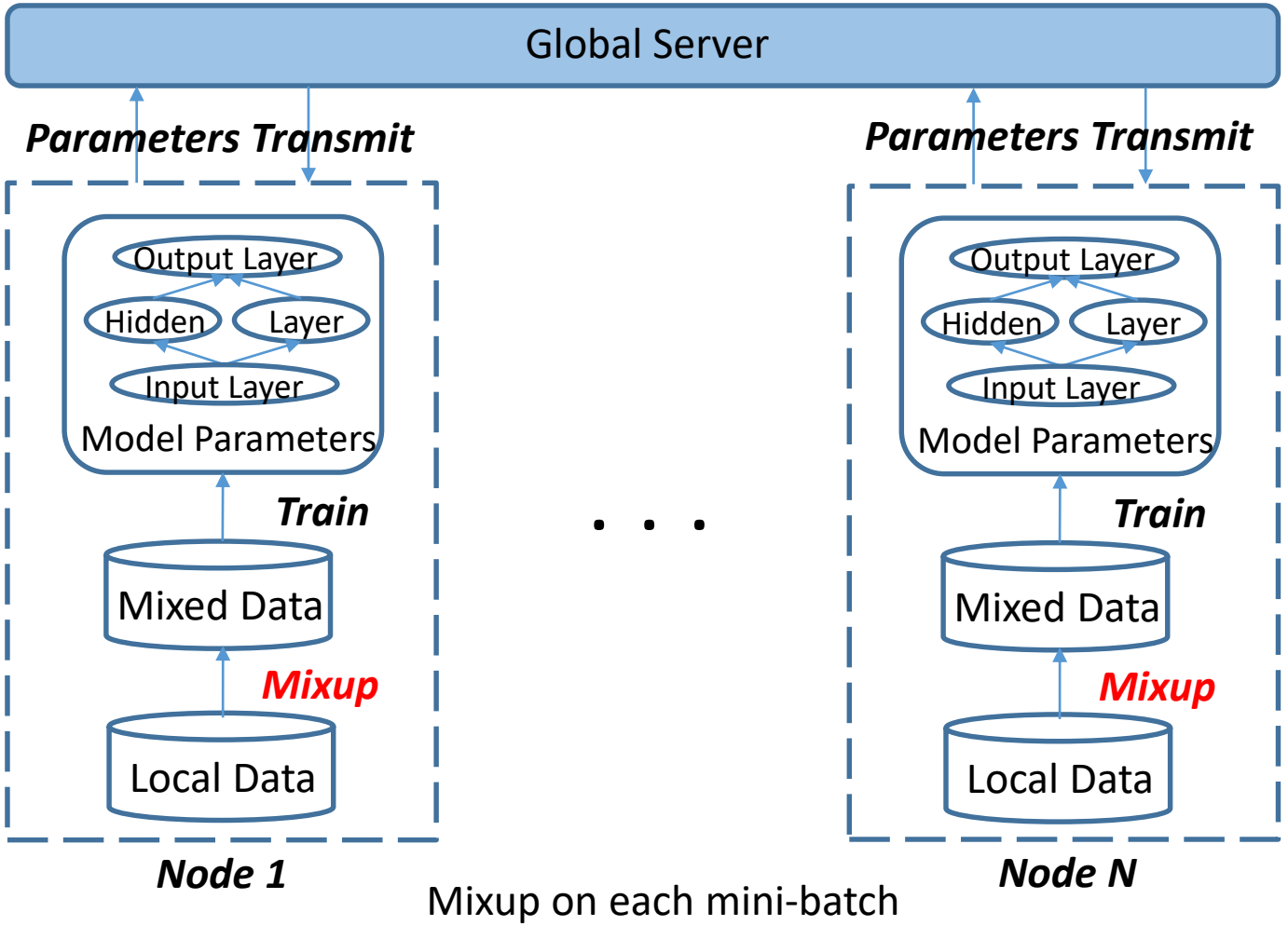
反向攻击模型





基于模型攻击的重构图像

解决思路：样本融合(Sample Mixup)



分布式深度学习方法——样本融合

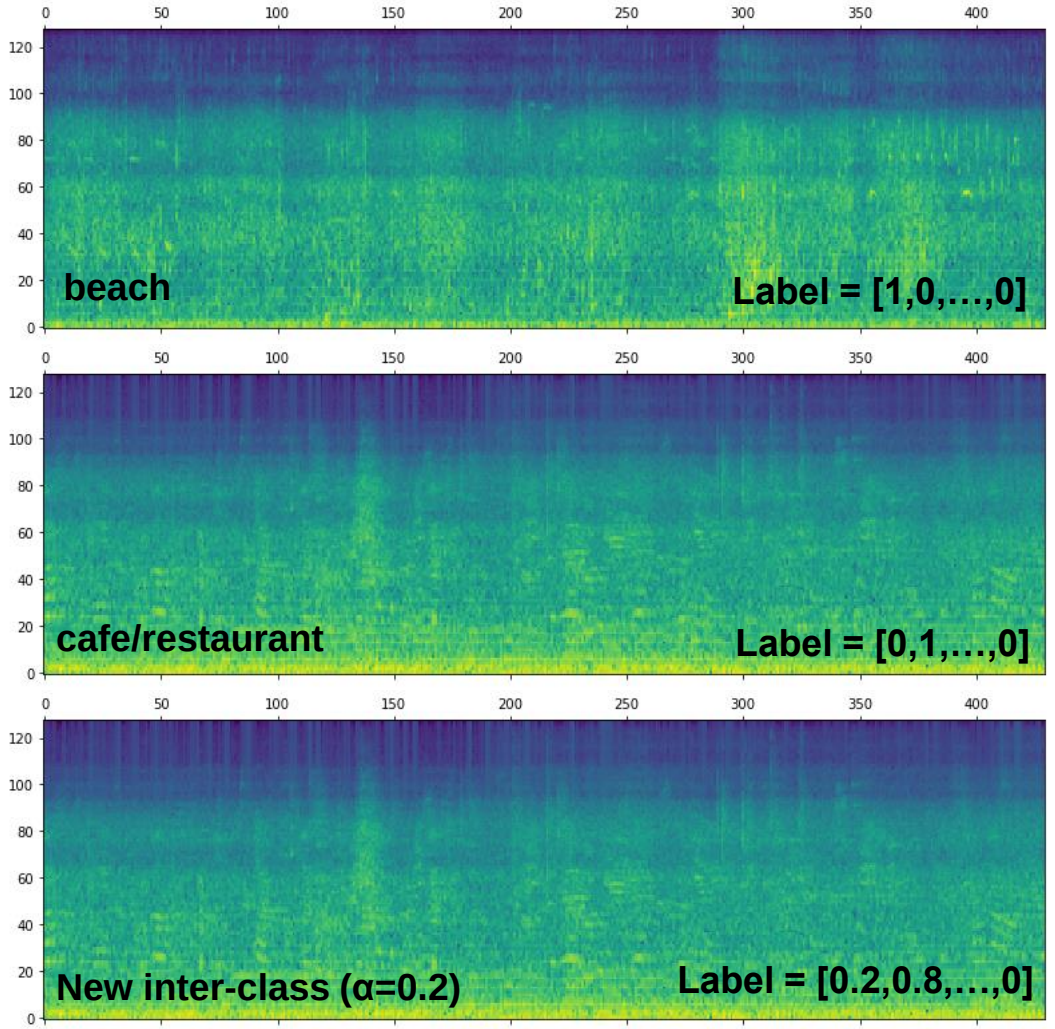
图像样本随机融合：

1. 基于生成中间类别提升模型性能
2. 保护原始数据安全

0.4*  +0.6*  = 

$$\tilde{x}_{k,i} = \lambda x_{k,i} + (1-\lambda) x_{k,j}$$
$$\tilde{y}_{k,i} = \lambda y_{k,i} + (1-\lambda) y_{k,j}$$

音频



基于样本融合的云际协作学习(MCL)

反向攻击模型

实际样本

协作学习(CL)

样本混合协作学习(MCL)

	aHash	pHash	dHash
With MCL	23	10	14
Without MCL	6	4	12

图像		Accuracy(%)	
Model	Method	CIFAR-10	CIFAR-100
VGG-19	CL	93.34	71.95
	MCL	93.67	74.88
ResNet-110	CL	93.89	71.14
	MCL	94.78	76.34
DenseNet-100	CL	95.46	77.12
	MCL	96.01	78.17

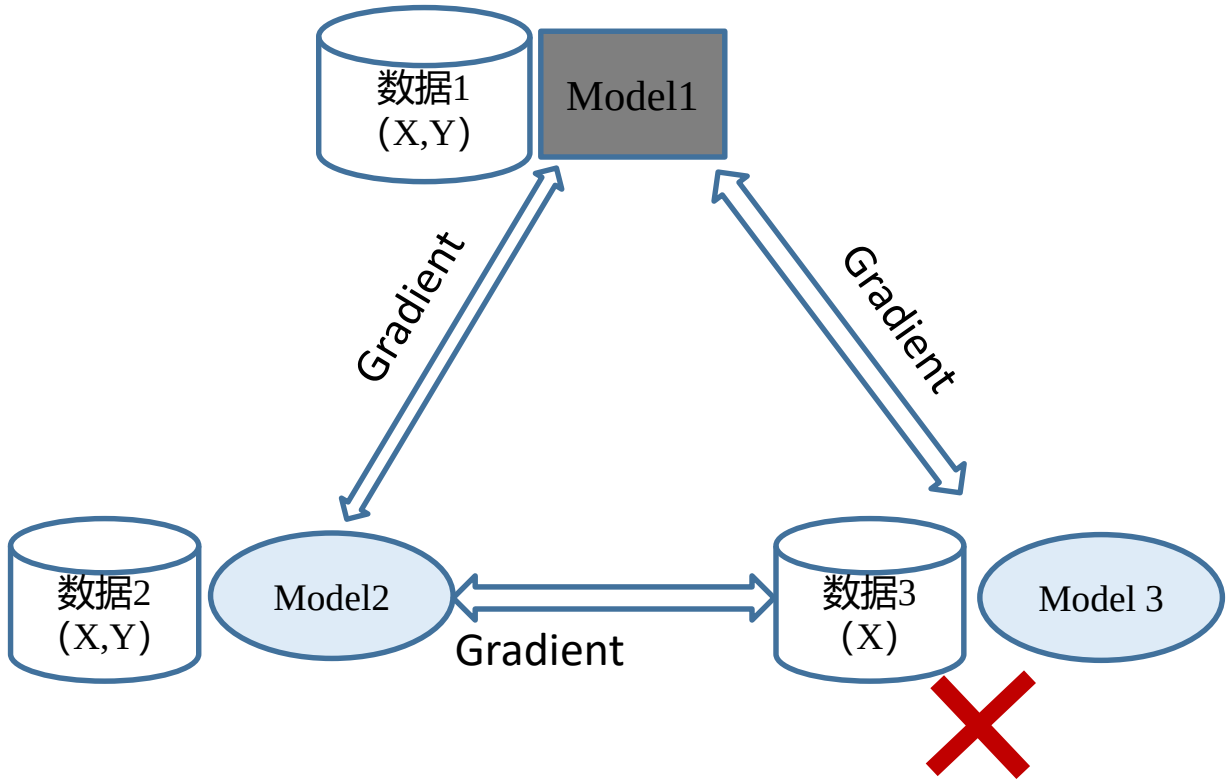
文本（Toxic Comment）	Multi-class AUC(%)	
Model	CL	MCL
LSTM	98.52	98.83
Capsule	98.32	98.71

大纲

- 应用场景
- **基于云际计算的分布式深度学习最新进展**
  - 分布式深度学习方法
    - ✓ 模型平均及其改进
    - ✓ 模型蒸馏
  - 声学建模中的应用
  - 其它应用场景
- 小结

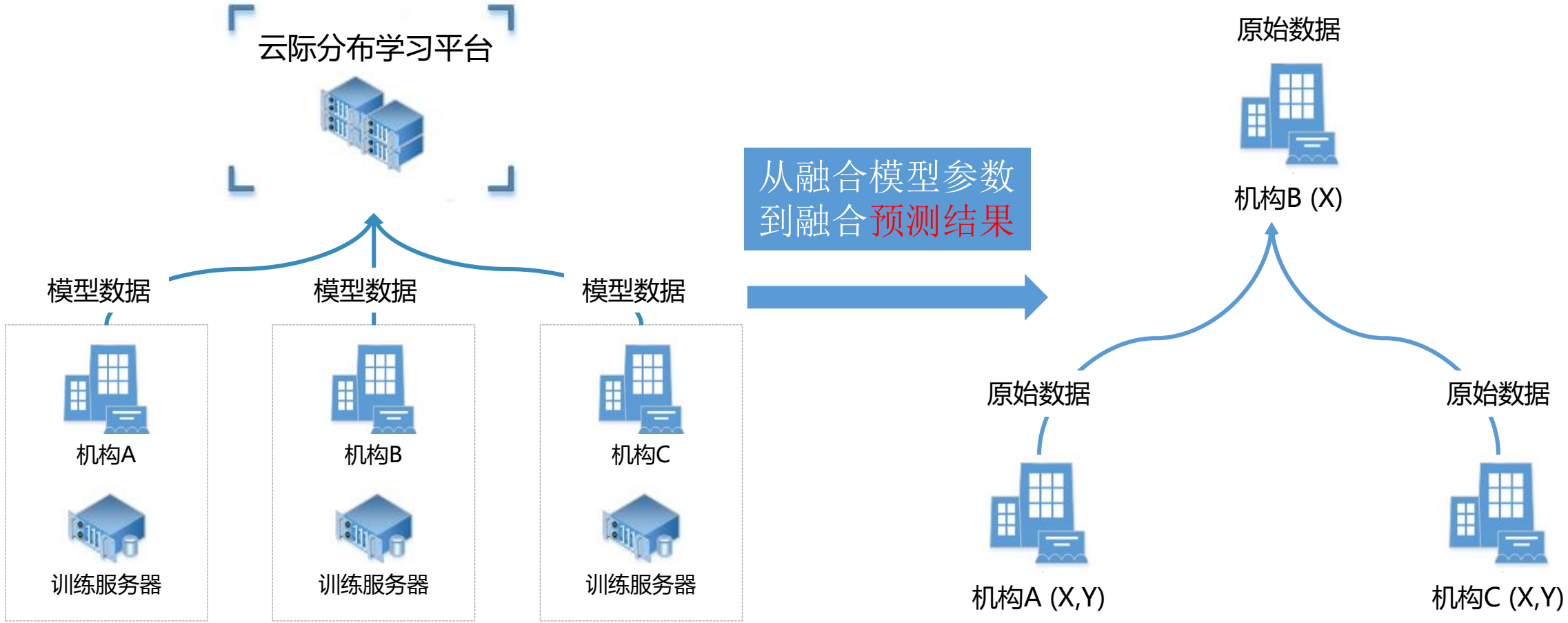
分布式深度学习方法

- 模型平均的问题
  - 带宽需求大，每1个epoch同步一次
  - **子模型结构需要完全相同**
  - **无法充分利用大量无标签数据**



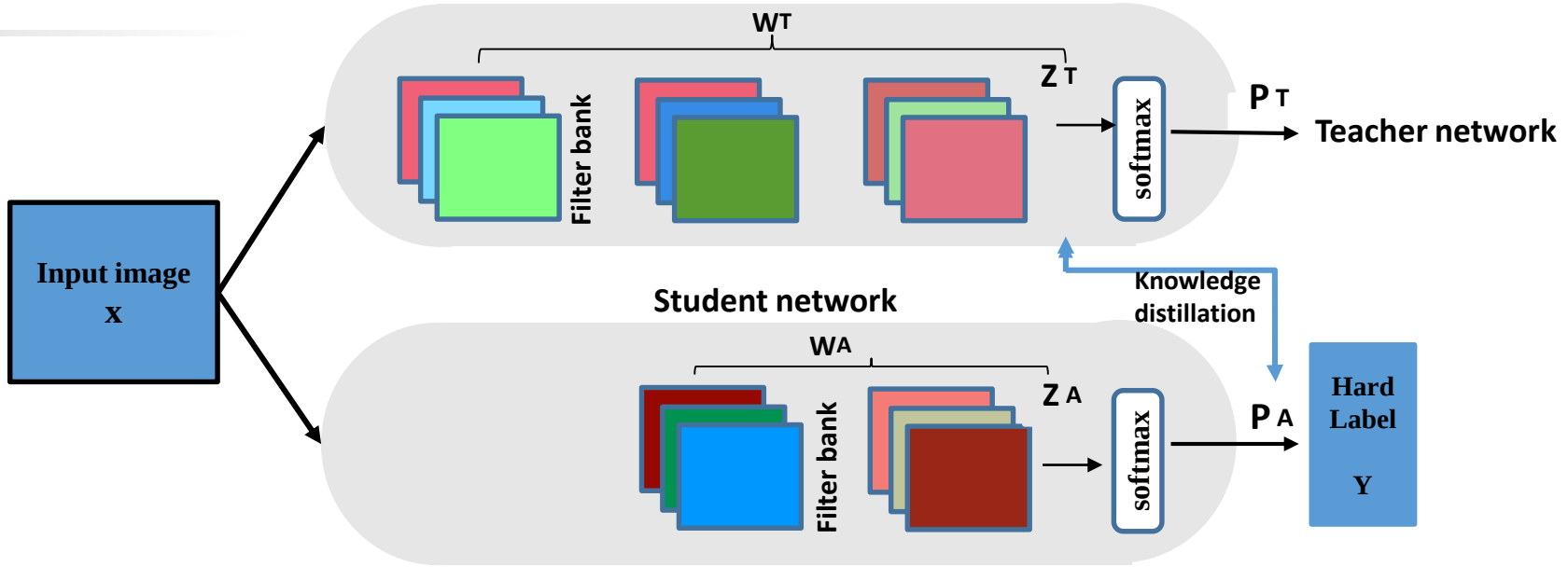
分布式深度学习方法——协作在线蒸馏

- 交换数据——交换模型——交换第三方数据预测结果

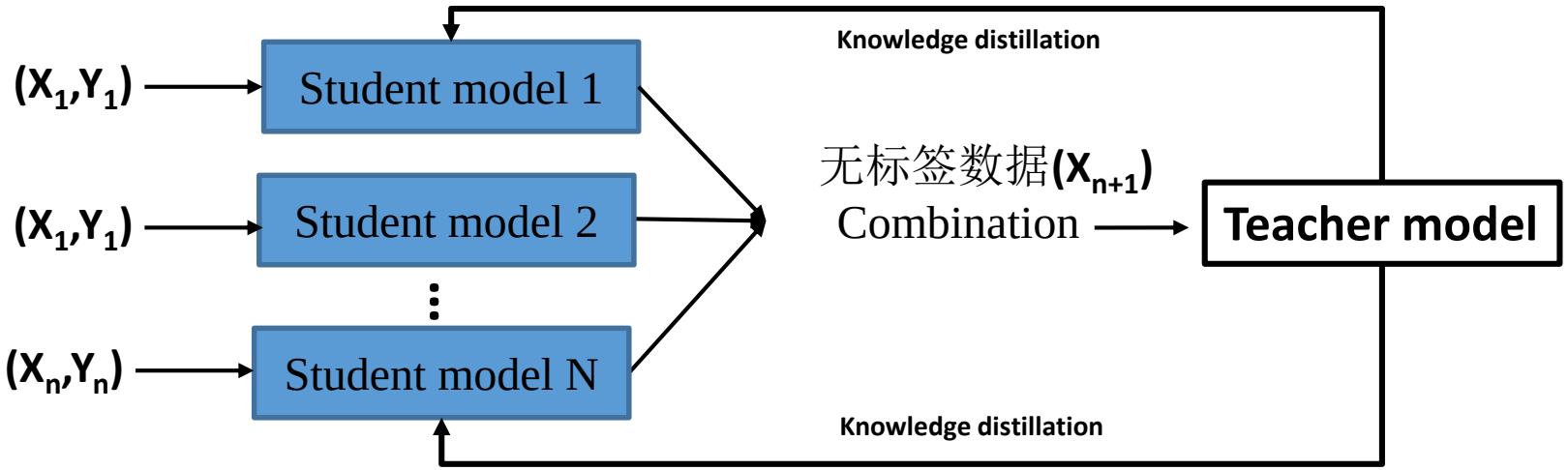


分布式深度学习方法——协作在线蒸馏

蒸馏(Distill)



集成(Ensemble)



Chaojie Zhao, Peng Zhang, Jian Zhu, Chengrui Wu, Huaimin Wang, **Kele Xu**. Predicting tongue motion in unlabeled ultrasound videos with convolutional recurrent neural network. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019).*

分布式深度学习方法——协作在线蒸馏

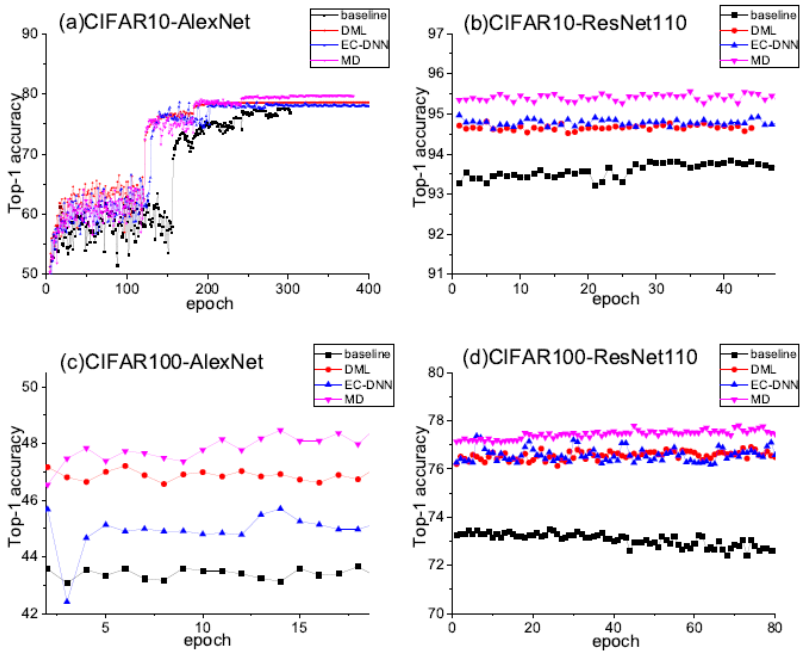


Fig. 1. Classification Accuracy compared to the State of Art Online Distillation. Baseline is the node training alone method, EC-DNN is the Ensemble-Compression method, and DML is the Deep Mutual Learning method, and MD is our method.

model1	model2	I(m1)	I(m2)	MD(m1)	MD(m2)
CIFAR10					
VGG19	ResNet	93.36	93.84	**94.27**	**95.6**
DenseNet	ResNet	95.35	93.84	**95.72**	**95.48**
AlexNet	VGG19	77.58	93.36	**80.23**	**93.81**
ResNet	ResNet	93.84	93.84	**95.72**	**95.75**
DenseNet	DenseNet	95.35	95.35	**95.62**	**95.76**
VGG19	VGG19	93.36	93.36	**94.10**	**94.33**
CIFAR100					
VGG19	ResNet	72.57	74.02	**74.85**	**77.75**
DenseNet	ResNet	77.57	74.02	**78.57**	**79.09**
AlexNet	VGG19	43.85	72.57	**49.79**	**74.08**
ResNet	ResNet	74.02	74.02	**77.81**	**77.47**
DenseNet	DenseNet	77.57	77.57	**78.55**	**78.85**
VGG19	VGG19	72.57	72.57	**75.43**	**75.45**

Liang Gao, Haibo Mi, Dawei Feng, **Kele Xu**, Yuxing Peng. Multi-structures based collaborative online distillation. ***2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)***.

大纲

- 应用场景
- 基于云际计算的分布式深度学习最新进展
  - 分布式深度学习方法
  - **声学建模中的应用**
  - 其它应用场景
- 小结

声学建模中的应用——人类语音

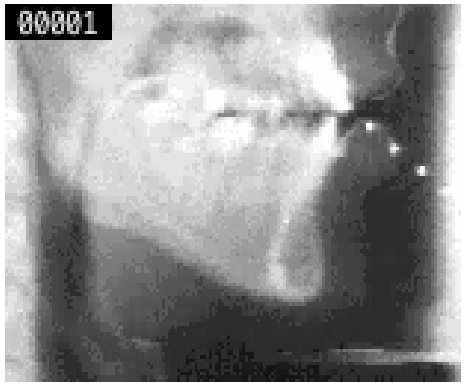
- 发音器官发出来的具有一定意义的声音
- 多发音器官复杂协作
- 语音是人类最自然，最直接的交流方式
- 语音存在若干问题
  - 噪声敏感性
  - 传播介质敏感
  - 交互安全问题
  - 语种问题
  - 若干疾病使人无法发声

- **Xu, Kele**, et al. "A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization." *The Journal of the Acoustical Society of America* (2016)
- **Xu, Kele**, et al. Ultrasound tongue gestural sequence classification using the recurrent neural network. *The Journal of the Acoustic Society of America* (2018).

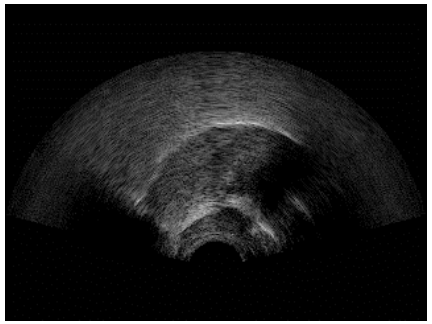
声学建模中的应用

- 问题原因
  - 声学信号的传播特性，依靠传播介质，点对面的传播方式
- 潜在解决方案
  - 利用非声学信号(间接成像方式)进行语音识别
  - 基于图像的语音识别受到越来越多人的重视

MRI
(Courtesy: UCLA Phonetics Lab)



X-ray



Ultrasound

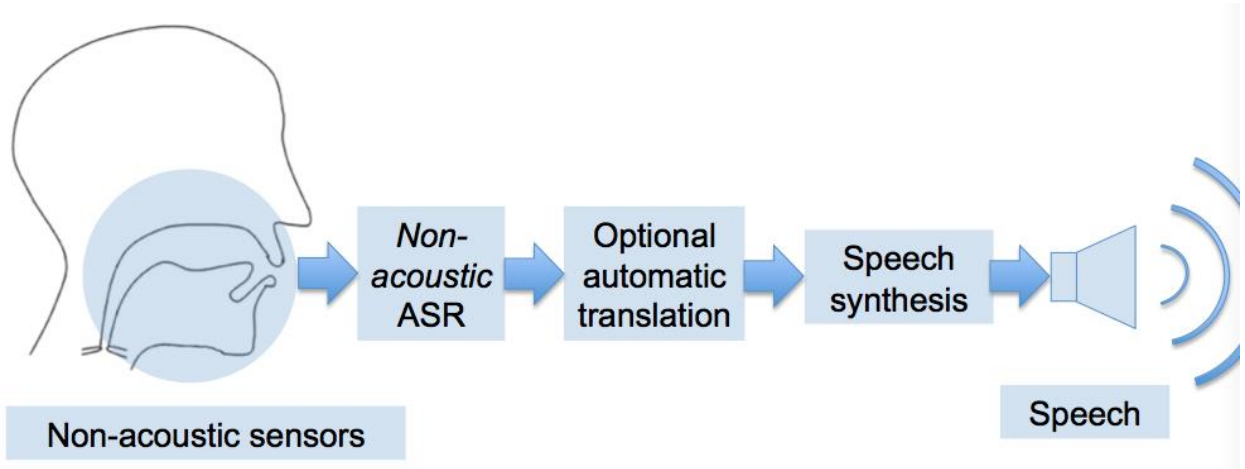
EMA
(Courtesy: 玛丽皇后学院)

人类发声研究的不同成像方式

- Xu, Kele, et al. "Is Speckle Tracking Feasible for Ultrasound Tongue Images?." *Acta Acustica united with Acustica*. 2017
- Xu, Kele, et al. A Comparative Performance Study on Convolutional Neural Network Architecture for Audio Scene Classification. *Acta Acustica united with Acustica. 2018*

基于声道成像的语音识别

- 无声语音接口



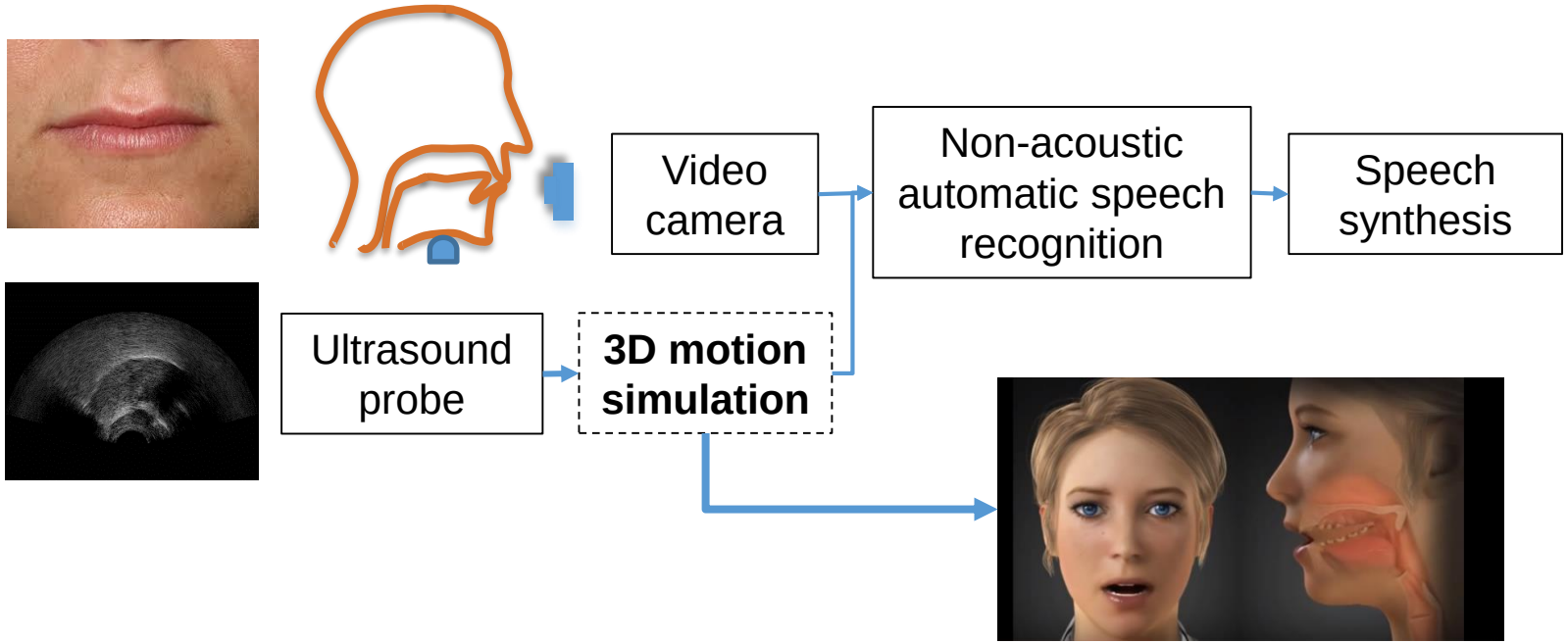
应用场景

- **医学**
  - 帮助患者重新“发声”
- **通信**
  - 点对点加密通信
  - 静态人人/人机交互
  - 高噪声背景下交互

- Xu, Kele, et al. "Convolutional neural network-based automatic classification of midsagittal tongue gestural targets using B-mode ultrasound images." *The Journal of the Acoustical Society of America* (2017).
- Xu, Kele, et al. General audio tagging with ensembling convolutional neural networks and statistical features. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019).*

分布式无声语音接口

改进的基于超声图像三维重构的无声语音接口流程图



方法	精度
CNN+HMM	72%
CNN+RNN	91%
3D CNN+RNN	94%
分布式3D CNN+RNN	95%

- Xu, Kele, et al. "Robust contour tracking in ultrasound tongue image sequences." *Clinical linguistics & phonetics* (2016).
- Bo Li, **Kele Xu***, Jian Zhu, Haibo Mi, Dezhi Wang, Huaiman Wang. Transfer learning-based ultrasound tongue image classification. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019).*
- Bo Li, **Kele Xu***, Dawei Feng, Haibo Mi, Dezhi Wang, Huaiman Wang, Jian Zhu. Denoising convolutional autoencoder based B-mode ultrasound tongue image feature extraction. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019).*

声学建模中的应用

❖研究动机

❑在海量数据、复杂海杂波的情况下，开发海洋大型生物检测算法

❑为潜艇以及大型船只提供航线规划，有效避免碰撞

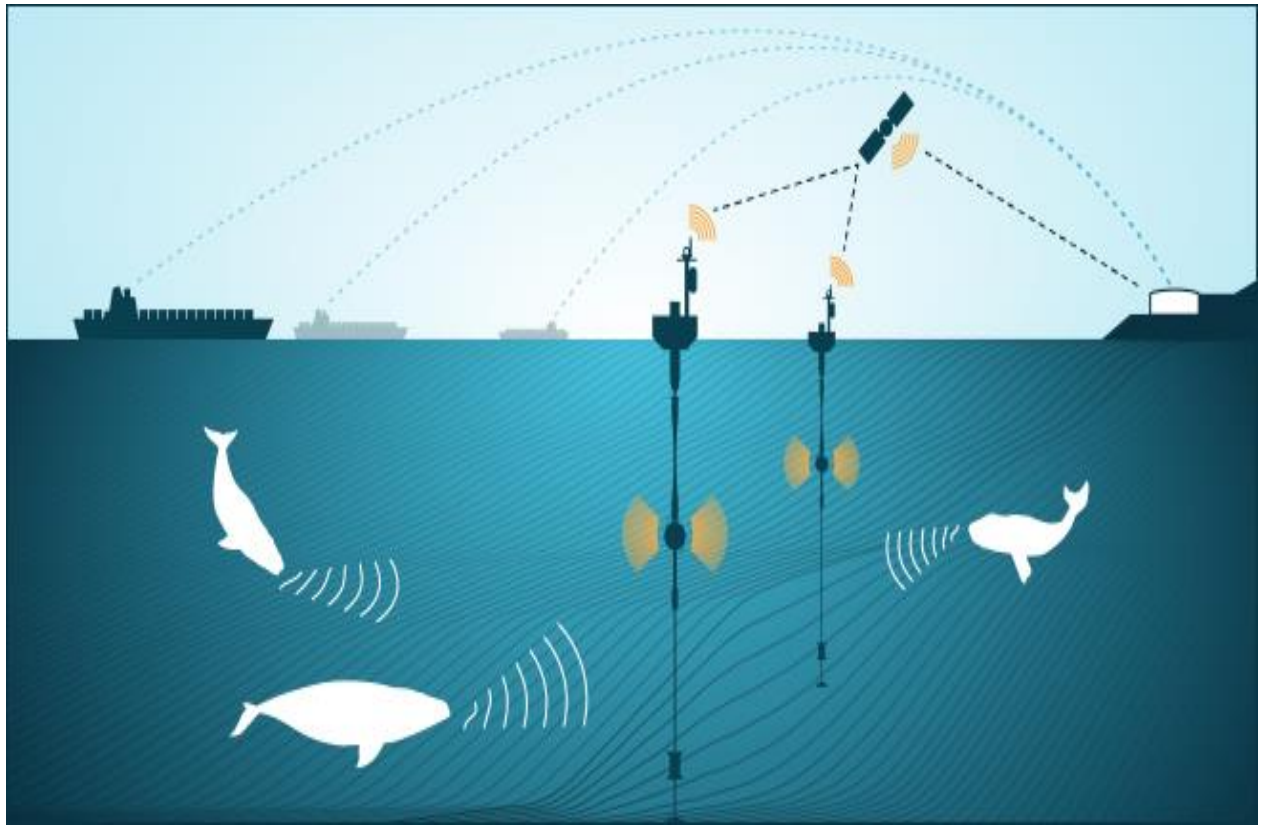
❑为海洋探测及科考提供技术支撑

❖研究难点

❑海杂波噪声复杂

❑传统方法效率低，鲁棒性低

❑数据规模海量，>5TB/天



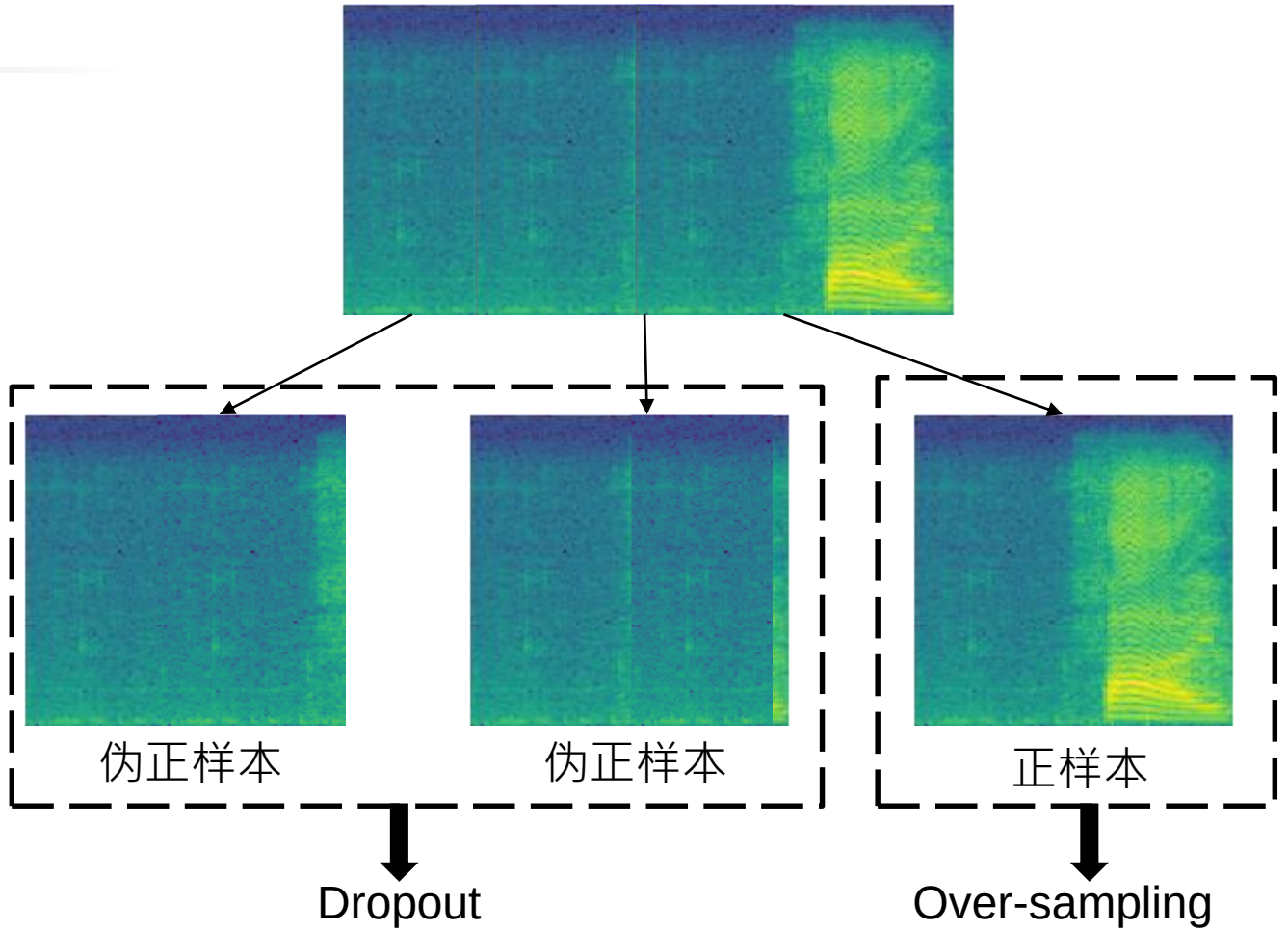
分布式声纹检测

解决办法

伪样本丢弃
(Sample dropout)
正样本多尺度过采样

分布式训练方法

检测错误率:
5.3%降低到1.7%



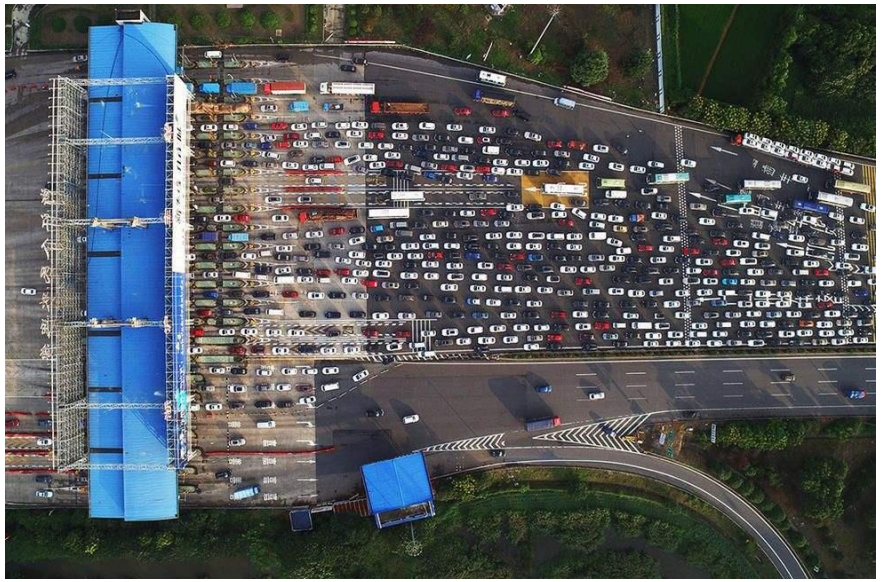
- **Kele Xu** et al. North Atlantic Right Whale Call Detection with Very Deep Convolutional Neural Networks. *The Journal of the Acoustic Society of America*. 2017
- Mingyang Geng, **Kele Xu**, Bo Ding, Huaiman Wang, Lei Zhang. Learning data augmentation policies using augmented random search. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019).*

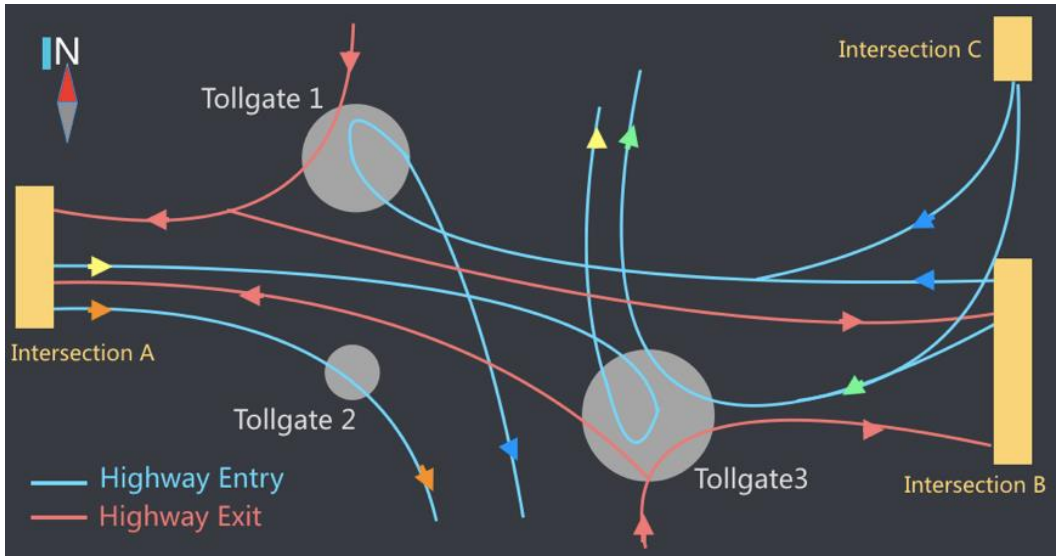
大纲

- 应用场景
- 基于云际计算的分布式深度学习最新进展
  - 分布式深度学习方法
  - 声学建模中的应用
  - **其它应用场景**
- 小结

高速公路行程时间预测

- 问题背景
  - 高速公路收费站是交通网络中的瓶颈
  - 在高峰时段，高速收费站收费缓慢可能导致交通瘫痪
  - 有效预测高速路段行程所需时间，能缓解交通压力，为公路管理决策者提供技术支撑
- 竞赛信息
  - ACM KDD CUP
  - 阿里云主办

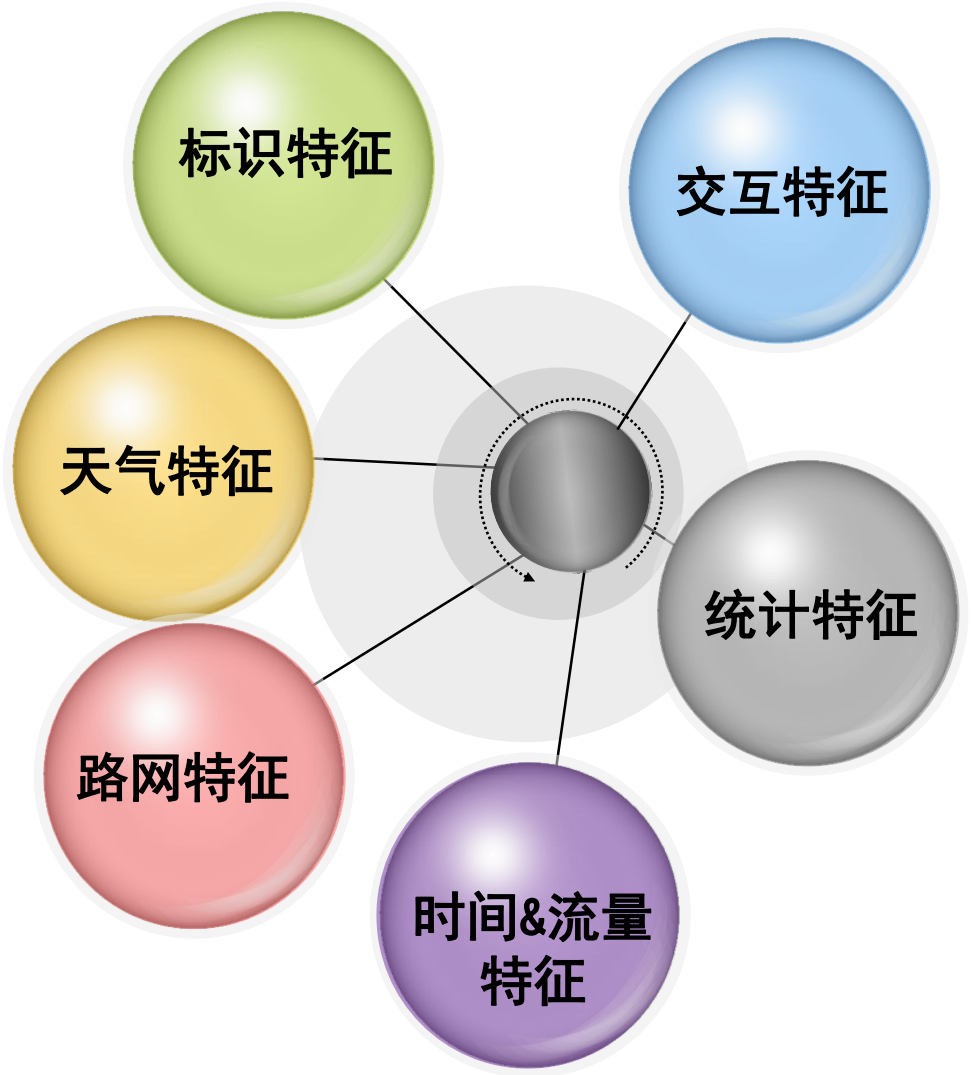




高速公路行程时间预测

❖挑战

- 道路网络
  - 如何建模道路网络内的交通流量
  - 道路的层级信息
- 收费站信息
  - 如何建模收费站的层级信息
  - 收费站的层级信息
- 天气信息
  - 如何历史天气信息及未来天气预报
- 收费站历史流量
- 通行时间受天气、假期、事故等多因素影响
- 数百万条，高维稀疏数据（6万维）



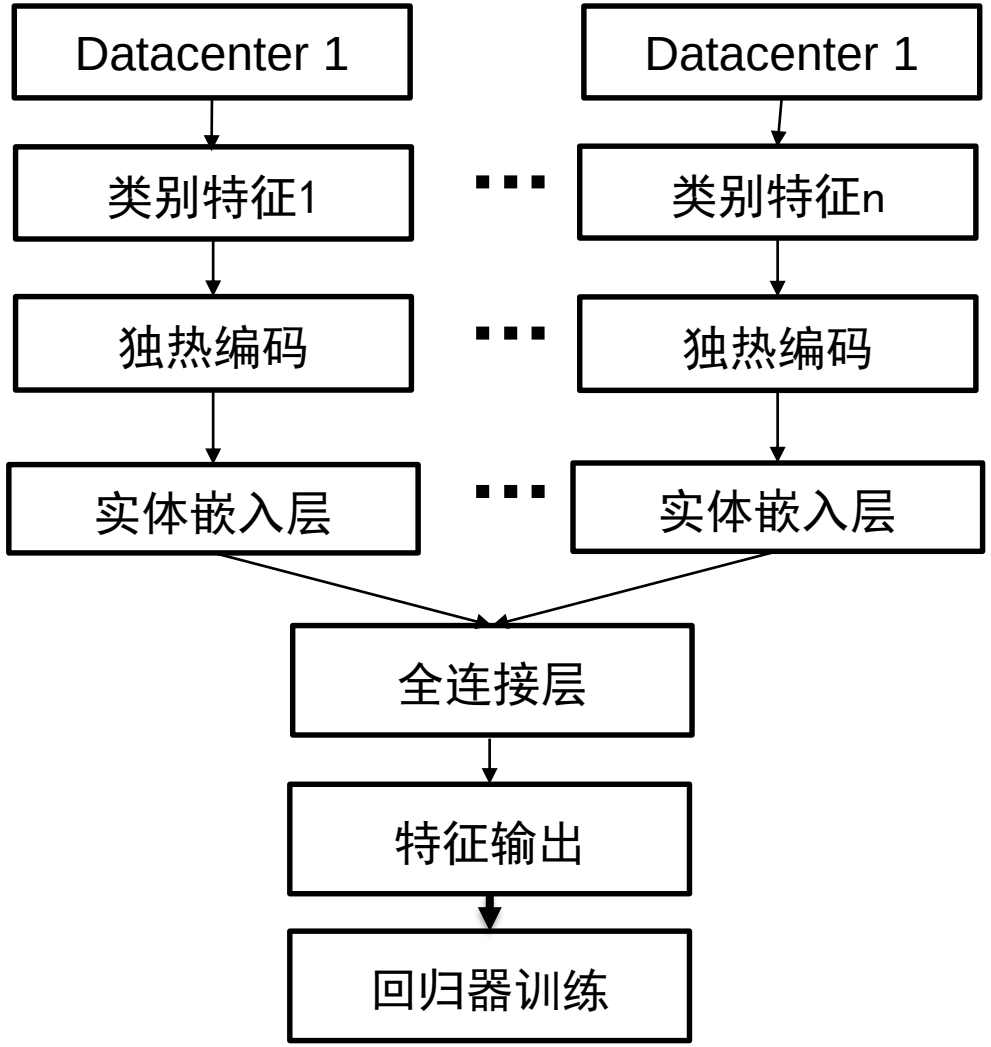
基于分布式深度学习的特征嵌入

解决方法：

- 将不同数据源的类别特征进行独热编码
- 基于类别特征独热编码进行实体嵌入层学习，保持类别层次关系不变
- 原本超过6万维的特征经过特征嵌入学习后，维度降至200维
- 基于200维度特征进行回归器训练

方法	Mean absolute percentage error
无类别特征嵌入	0.2419
集中式类别特征嵌入	0.1774
分布式类别特征嵌入	**0.1678**

在ACM KDD Cup比赛中位列第三位(共计3582组),
受邀在KDD主会上口头报告

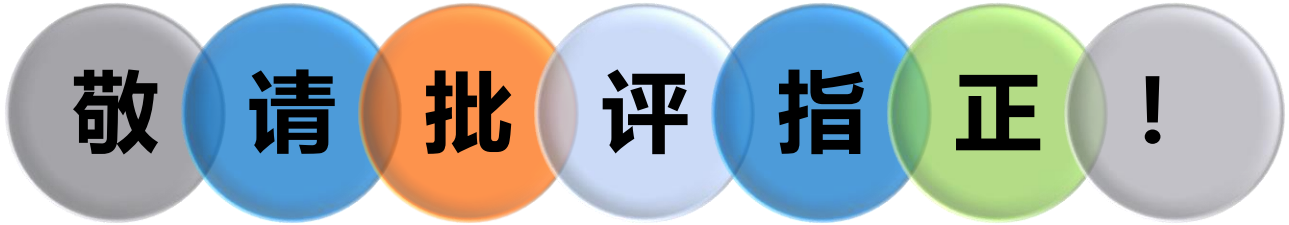


机器学习竞赛获奖情况

1	**2018年：Google General Audio Tagging第1名。 (共557组)**
2	2016年：中国计算机学会CCF大数据与计算智能大赛第1名。 (共985组)
2	2017年：中国计算机学会CCF大数据与计算智能大赛第2名。 (共1145组)
3	2017年：全国大数据创新应用大赛交通赛初赛第1名，金融赛第1名，教育赛第2名。 (共3356组)
4	2016年：携程酒店三个预订渠道的总产量预测第1名。 (共570组)
5	**2018年：第三届阿里云安全算法挑战赛第1名。 (共622组)**
6	**2018年：深圳医疗健康大数据创新应用国际大赛第1名。**
7	2017年：国际知识发现和数据挖掘竞赛(ACM KDD-Cup) 中排名第3位。 (共3582组)
8	2017年：Kaggle Santa背包优化算法竞赛第3名。 (共694组)
9	2016年：融360“天机” 风控大数据竞赛第4名。 (共700组)
10	**2018年：Kaggle地下盐体图像分割第4名。 (共3234组)**
11	2016年：ECML-PKDD (欧洲机器学习会议) Network Classification Challenge Competition第5名
12	2016年：中国人工智能协会CAAI Byte Cup国际机器学习竞赛第9名。 (共1000组)

小结

- 云际协作学习可以基于模型平均实现。
- 传统模型平均抵抗模型攻击能力较差。
- 基于样本融合的云际协作学习可以提高模型抗攻击能力。
- 跨模型蒸馏为云际协作学习提供了有效手段，同时该方法可以有效利用无标签数据。
- 云际协作学习应用场景广泛。


敬请批评指正！