# Industrializing AI with Baidu AutoDL

Haoyi Xiong, Ph.D

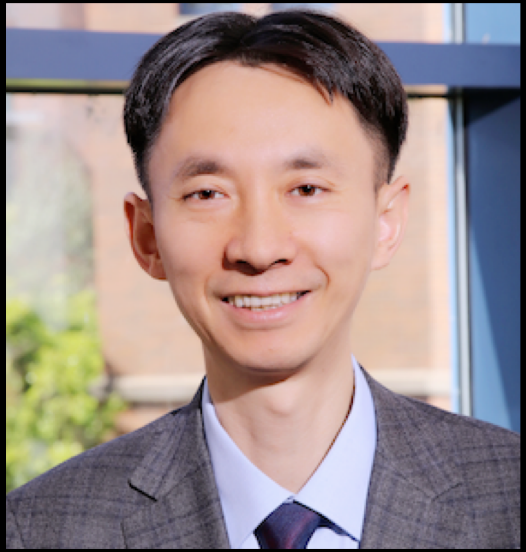
Big Data Lab,  Baidu, Inc.

# About the Talk

- What is Baidu AutoDL?
- What can we do with Baidu AutoDL?
- What technical advantages could Baidu AutoDL provide?
- What broader impacts and intellectual merits could Baidu AutoDL achieve?

# About Baidu Big Data Lab

- One of the six key research labs operated by Baidu, Inc.
- Established under The **United Nations** Development Programme, Aug 2014
- Currently, under the leadership of **Prof. Dejing Dou**
  - from the Department of Computer and Information Sciences, University of Oregon, Eugen, ORE
- In charge of the Baidu Research efforts on
  - Baidu AutoDL
  - Federated Learning
  - Interpolation of Deep Neural Networks
  - Foundations of big data…



Prof. Dejing Dou

Baidu Research

UNDP

# About the Speaker

- Research Interests
  - **Automated Deep Learning**, **Ubiquitous Big Data**
- Professional Experience
  - **Tech Lead & Staff Research Scientist** **2018.05—present**
    - Big Data Lab, Baidu, Inc.
  - **Tenure-Track Assistant Professor/Ph.D Advisor** **2016.08—2010.08**
    - Missouri University of Science and Technology (ex. UM Rolla), Rolla, MO
  - **Postdoctoral Research Associate** **2015.07—2016.08**
    - University of Virginia, Charlottesville, VA
- Education
  - Ph.D, Institut Mines-Télécom and UPMC Paris 6, Evry, France
  - MSc., Hong Kong University of Science and Technology, Hong Kong SAR, China
  - BEng., Huazhong University of Science and Technology, Wuhan, China

Dr. Haoyi Xiong

Personal homepage: https://sites.google.com/site/haoyixiongshomepage/

# Deep Learning as Large-scale Industrial Applications

✓ McKinsey research shows that deep learning can help 2/3 of corporate applications improve business performance, with an average growth of 62% in all industries.

AlphaGo vs. Lee Sedol

2016

Internet giant started adopting large-scale deep learning techniques for their core business: e.g., Google, Baidu search and advertising; Facebook information flow

2015

- Baidu released her first deep learning framework for Machine Translation
- Google Tensorflow became open-sourced.

2013

The first Deep learning framework Theano released

2012

- AlexNet network won the ImageNet competition
- Hinton leveraged deep learning to make large-scale speech recognition progress with Microsoft Research
- Baidu made her first attempt at using deep learning algorithms for speech recognition and OCR

2006

2010

Hinton proposed the world first deep neural network algorithm

Theoretical breakthrough

Technological breakthrough

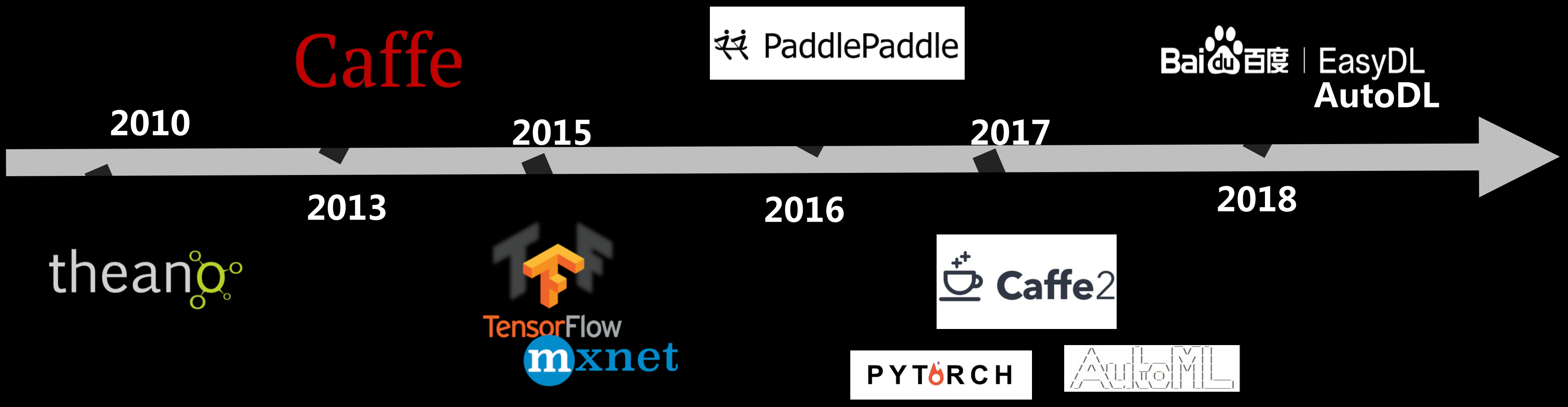
widely used

social influence

# Evolution of the Deep Learning Framework

From the prestigious universities to industrial giants—The battle field moved!

Theano (2010, U Montreal), Caffe (2013, Berkeley), Tensorflow (2015, Google), MXNET (2015, AWS support), PaddlePaddle (2016, Baidu), PyTorch (2017) , Facebook), Caffe2 (2017, Facebook)



# The Development of Baidu's Deep Learning

**2008**
- Built large-scale machine learning infrastructures, models, tools and experimental platforms

**2012**
- Extent deep learning to speech recognition, OCR, etc.

**2013**
- DNN models released for Baidu Search, and Fengchao CTR

**2015**
- Released the world's first NMT online translation engine
- Released Innovative acoustic model based on LSTM-CTC
- Released Innovative general deep learning framework for NLP tasks

**2016**
- PaddlePaddle--The Baidu Flagship Deep Learning Platform

**2017**
- Extent deep learning to recommendation system
- Released PaddlePaddle Fluid

**2018**
- Won multiple international Deep learning contests, including MSMARCO
- Release PaddlePaddle 3.0

# Baidu AutoDL as a part of PaddlePaddle Suite

**Service Platform**

- EasyDL
- AI Studio
- AutoDL

**Core Framework**

- PaddleRec
- PaddleCV
- PaddleNLP
- Training Framework
- PaddlePaddle Serving
- PaddlePaddle Mobile

**Modules**

- Visual DL
- PARL
- EDL

# Some Industry Applications of-course IN CHINA


Example 1

Long Video Summarization

Content–aware Image Extraction and Reorganization

Result

Precision:80%+


Example 2

Biometrics–Identity Co–verification

Recognizing faces with their IDs using mobile phones

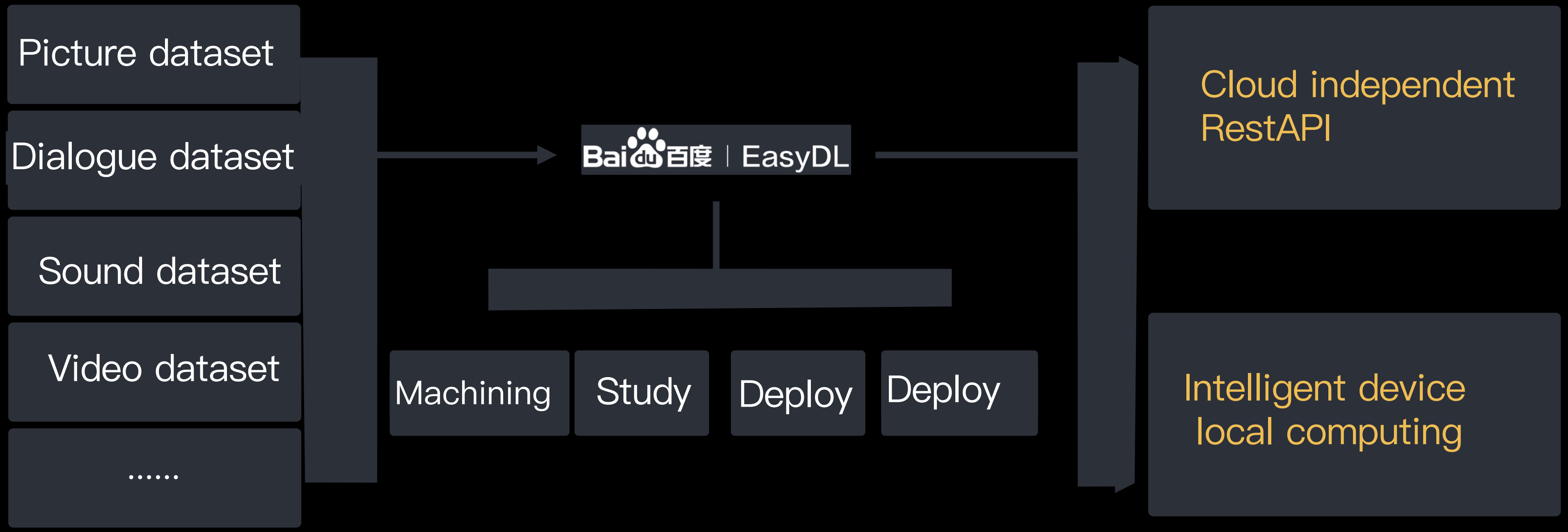
Result

Precision:99%+

Reducing the cost of insurance for customer verification

# Baidu AutoDL Service Provision through EasyDL

Producing high quality models for a wide spectrum of tasks, such as image classification and object detection
Capable of handling more than 6,000 EasyDL developers designing and tuning DNNs in parallel



# About Baidu AutoDL

- Baidu AutoDL Platform
  - AutoDL Design – neural architecture search systems for automated DNN designs
  - AutoDL Transfer – automated pre-trained DNN model sourcing and knowledge transfer
  - AutoDL Finetuner – automated DNN training with hyperparameter optimization
  - AutoDL Federation – automated federated learning with horizontal/vertical separation
  - AutoDL Edge–automated DNN training and inference over mobile devices

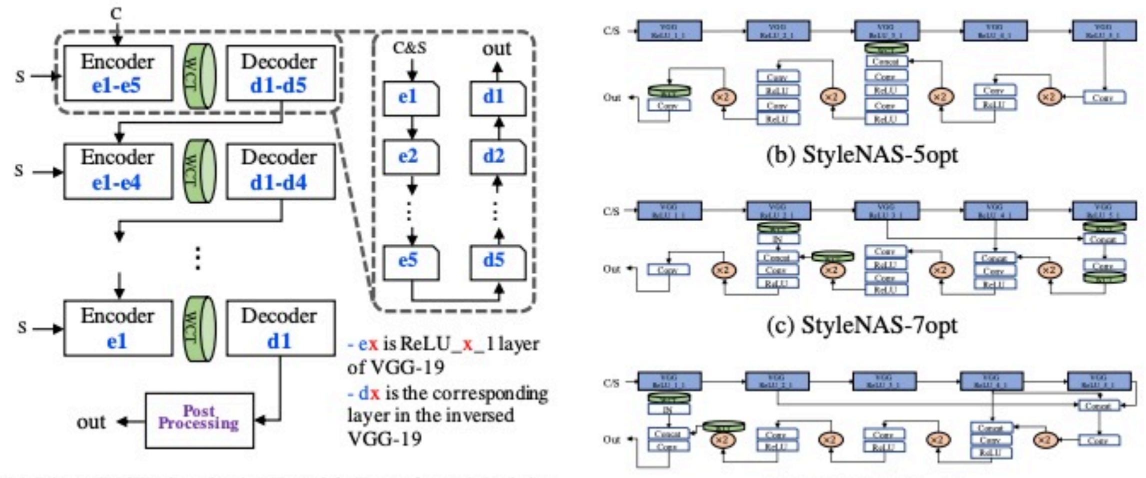
# AutoDL Design - Beyond all Manual Design Networks and (so far all) Architecture Search Algorithms

## AutoDL Design

The accuracy of the network designed by Auto DL technology is over 98%, which is higher than the networks designed by any human experts or other architecture search algorithms.
CIFAR 10 is a major image classification dataset for Benchmark

| Model | Acc. | Model | Acc. |
|---|---|---|---|
| vgg_15_BN_64 | 93 | resnet_v2_bottleneck_164 | 94.2 |
| vgg_16 | 93.6 | wide_resnet | 95 |
| resnet_32 | 92.5 | densenet_BC_100_12 | 95.5 |
| resnet_44 | 93 | resnext_29_8x64d | 96.2 |
| resnet_56 | 93.3 | shake_shake_64d_cutout | 97.1 |
| resnet_110 | 93.5 | AutoDL | 98.03 |

# AutoDL Design StyleNAS – the World First “Design-by-Design” Style Transfer Neural Networks




(a) PhotoWCT: Stacked AEs with Transform Modules.

(b) StyleNAS-5opt

(c) StyleNAS-7opt

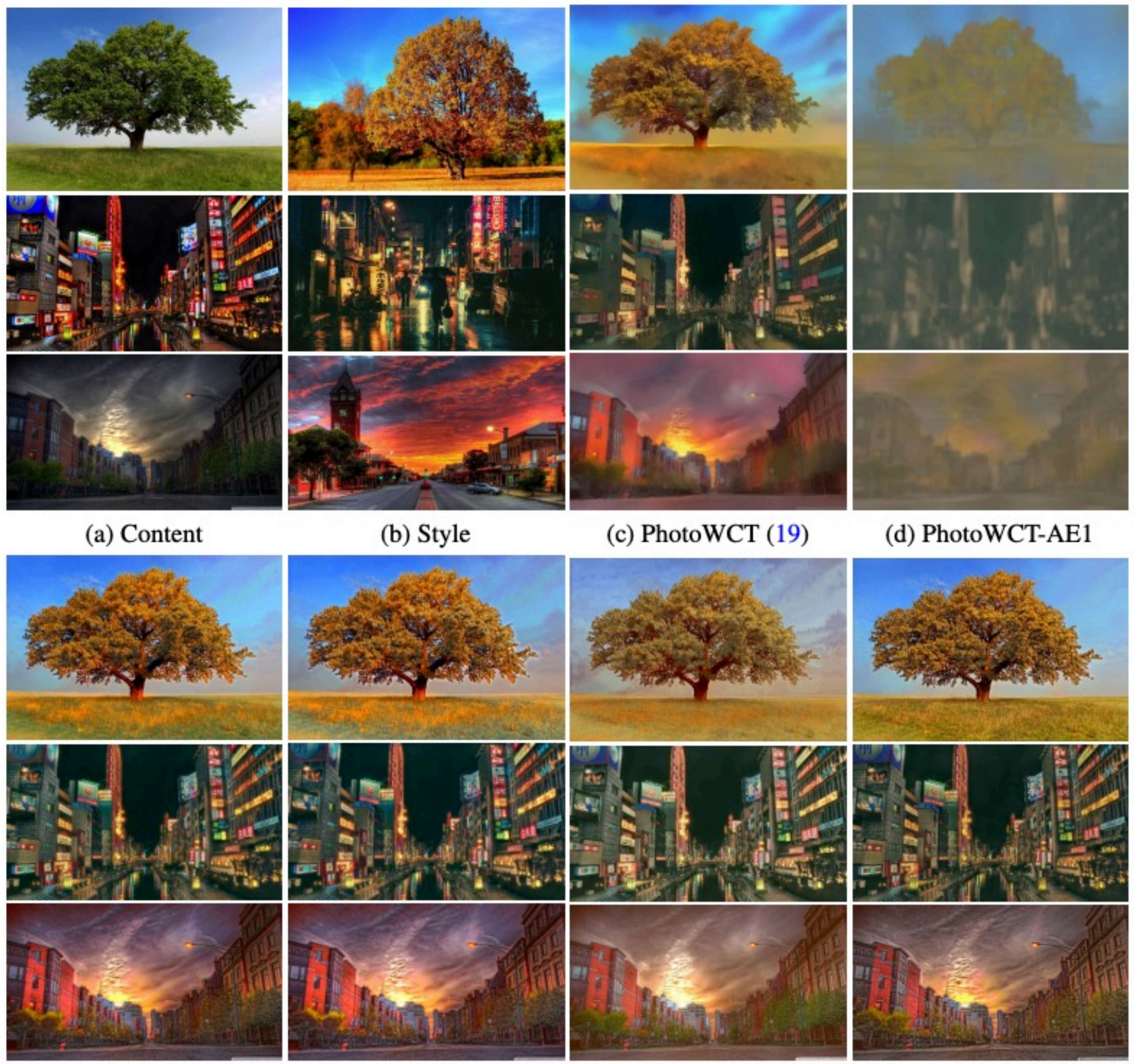
(d) StyleNAS-9opt

Table 1: **Quantitative evaluation results.** A lower FID score means the evaluated method creates images with a more similar style to the reference supervisory oracles/style images. A higher TV score indicates that the measured method preserves more details.

| Method | PhotoWCT | PhotoWCT-AE1 | StyleNAS-RS | StyleNAS-5opt | StyleNAS-7opt | StyleNAS-9opt |
|---|---|---|---|---|---|---|
| FID-Style ↓ | 180.19 | 469.54 | 208.67 | 183.95 | 173.18 | **172.00** |
| FID-Oracle ↓ | - | - | 142.83 | 65.94 | 50.71 | **45.81** |
| TV score ↑ | 5.11 | 0.43 | 4.09 | 5.73 | **6.14** | 5.83 |

Table 2: **Computation time comparison.**

| Method | PhotoWCT | PhotoWCT-AE1 | StyleNAS-RS | StyleNAS-5opt | StyleNAS-7opt | StyleNAS-9opt |
|---|---|---|---|---|---|---|
| 256 × 128 | 4.38 | 0.83 | 0.30 | **0.05** | 0.07 | 0.47 |
| 512 × 256 | 25.37 | 0.99 | 0.35 | **0.09** | 0.10 | 0.67 |
| 768 × 384 | 64.73 | 1.10 | 0.42 | **0.15** | 0.18 | 0.76 |
| 1024 × 512 | 153.25 | - | 0.52 | **0.23** | 0.29 | 0.91 |

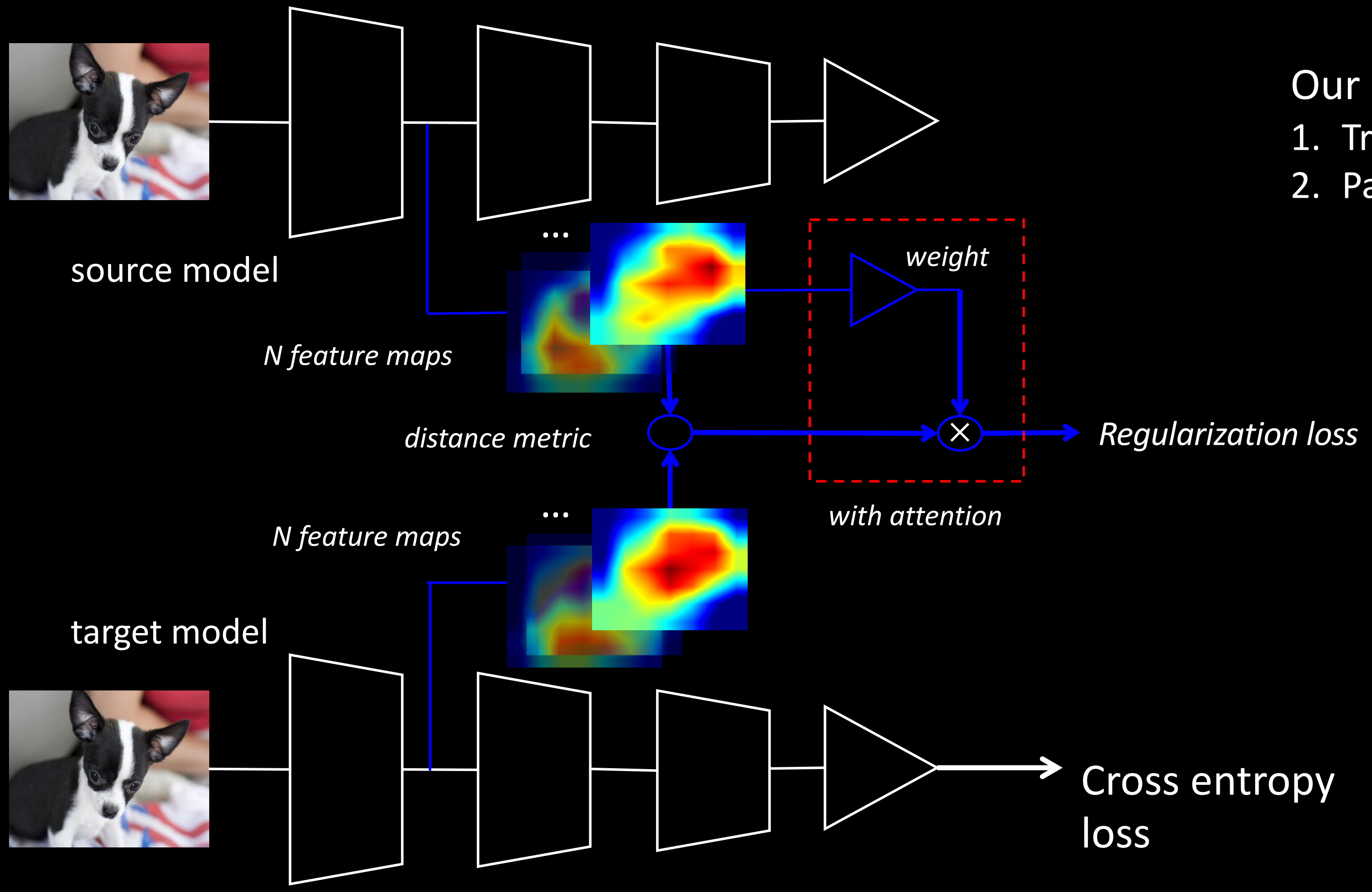


(a) Content (b) Style (c) PhotoWCT (19) (d) PhotoWCT-AE1

(e) StyleNAS-RS (f) StyleNAS-5opt (g) StyleNAS-7opt (h) StyleNAS-9opt

StyleNAS: An Empirical Study of Neural Architecture Search to Uncover Surprisingly Fast End-to-End Universal Style Transfer Networks J An, H Xiong, J Ma, J Luo, J Huan - arXiv preprint arXiv:1906.02470, 2019

# AutoDL Transfer – Pay Attentions to Features, Transfer via Attentive Knowledge Distillation




Our methods

1. Transfer via Knowledge Distillation of Feature Maps
2. Pay attention to “important” Features using weights


[ICLR'19] Xingjian Li, **Haoyi Xiong***, Hanchao Wang, Yuxuan Rao, Liping Liu, and Jun Huan. DELTA: Deep Learning Transfer using Feature Map with Attention for Convolutional Neural Networks. In Proceedings of The Seventh International Conference on Learning Representations, New Orleans, LA, 2019.

# AutoDL Transfer – Pay Attentions to Features, Transfer via Attentive Knowledge Distillation

- Experiments

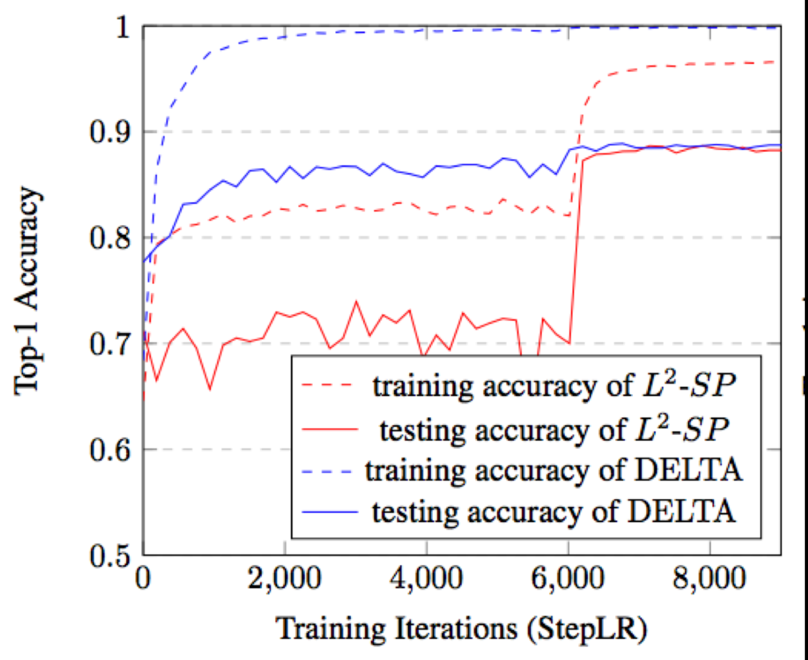
Table 1: Comparison of top-1 accuracy with different implementation approaches. $L^2$-$FE$: Using the pre-trained model as a feature extractor. Baselines: $L^2$-$FE$, $L^2$ and $L^2$-$SP$.

| | $L^2$-$FE$ | $L^2$ | $L^2$-$SP$ | DELTA(w/o ATT) | DELTA |
|---|---|---|---|---|---|
| MIT Indoors 67 | $84.5 \pm 0.2$ | $83.7 \pm 0.3$ | $85.1 \pm 0.1$ | $85.3 \pm 0.2$ | **85.5±0.3** |
| Stanford Dogs 120 | $83.6 \pm 0.1$ | $83.3 \pm 0.2$ | $88.3 \pm 0.2$ | $88.3 \pm 0.2$ | **88.7±0.1** |
| Caltech 256 30 | $84.6 \pm 0.2$ | $84.7 \pm 0.3$ | $85.4 \pm 0.2$ | $85.7 \pm 0.3$ | **86.6±0.1** |
| Caltech 256 60 | $87.2 \pm 0.2$ | $87.2 \pm 0.3$ | $87.2 \pm 0.1$ | $87.6 \pm 0.2$ | **88.7±0.1** |

Table 2: Comparing top-1 accuracy using data augmentation for three regularization methods.

| | $L^2$ | $L^2$-$SP$ | DELTA |
|---|---|---|---|
| MIT Indoors 67 | $84.4 \pm 0.7$ | $85.2 \pm 0.3$ | **85.9±0.3** |
| Stanford Dogs 120 | $85.7 \pm 0.2$ | $90.8 \pm 0.2$ | **91.2±0.2** |
| Caltech 256 30 | $85.1 \pm 0.4$ | $86.4 \pm 0.2$ | **87.1±0.2** |
| Caltech 256 60 | $87.4 \pm 0.2$ | $88.3 \pm 0.1$ | **89.1±0.1** |

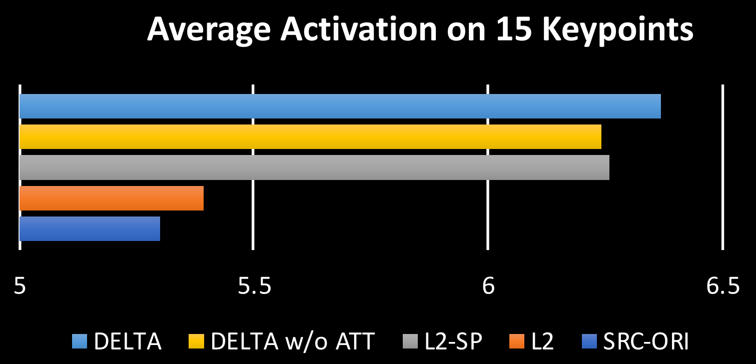
better performance

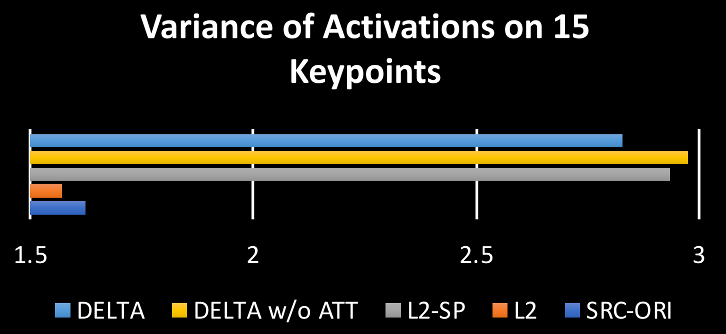


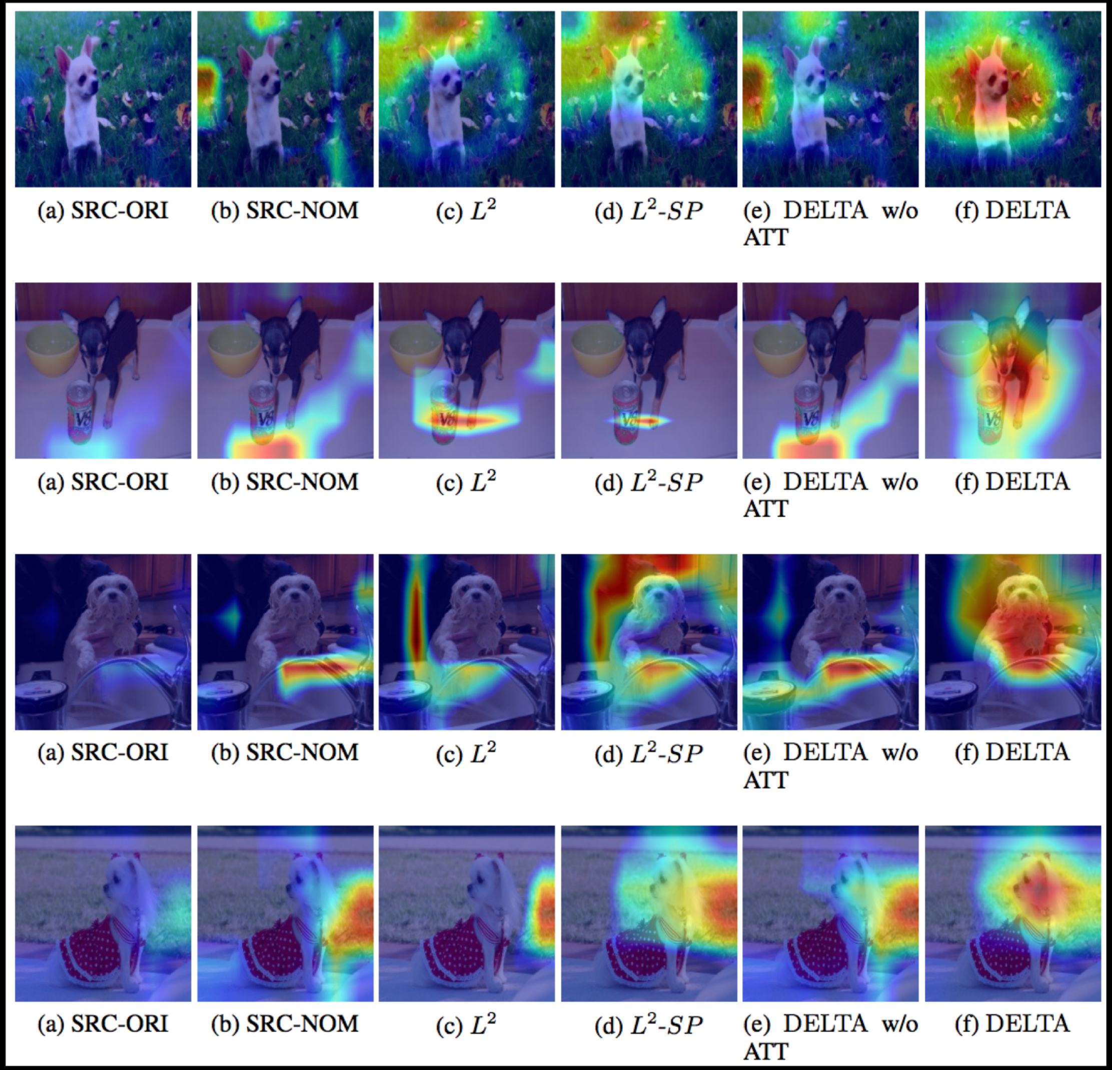

faster and more stable convergence

[ICLR'19] Xingjian Li, **Haoyi Xiong***, Hanchao Wang, Yuxuan Rao, Liping Liu, and Jun Huan. DELTA: Deep Learning Transfer using Feature Map with Attention for Convolutional Neural Networks. In Proceedings of The Seventh International Conference on Learning Representations, New Orleans, LA, 2019.

# AutoDL Transfer – Pay Attentions to Features, Transfer via Attentive Knowledge Distillation




[ICLR'19] Xingjian Li, **Haoyi Xiong***, Hanchao Wang, Yuxuan Rao, Liping Liu, and Jun Huan. DELTA: Deep Learning Transfer using Feature Map with Attention for Convolutional Neural Networks. In Proceedings of The Seventh International Conference on Learning Representations, New Orleans, LA, 2019.

# AutoDL Transfer – Make Deep Transfer Learning Never Hurt

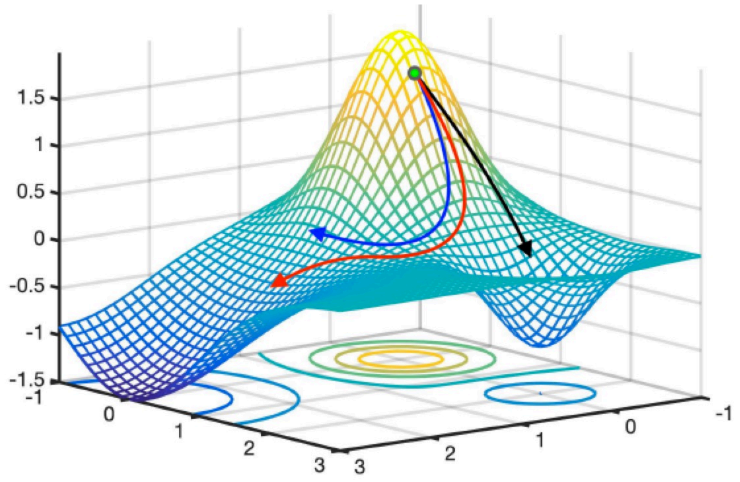


*Figure 1.* Flows of Descent Directions on the Empirical Loss. **Black Line**: the flow via descent direction of empirical loss (i.e., gradient of empirical loss) from a common starting point, where the descent direction quickly leads the learning procedure converged to a local minimum of over-fitting; **Blue Line**: the flow via the descent direction linearly combining gradients of empirical loss and the regularization term, where the regularization term diminishes the minimization of empirical loss; **Red Line**: the flow via the descent direction balancing the gradients of empirical loss and the regularzation term, where the descend direction leads to a flat area with low empirical loss (i.e., potentials of great generalizability).

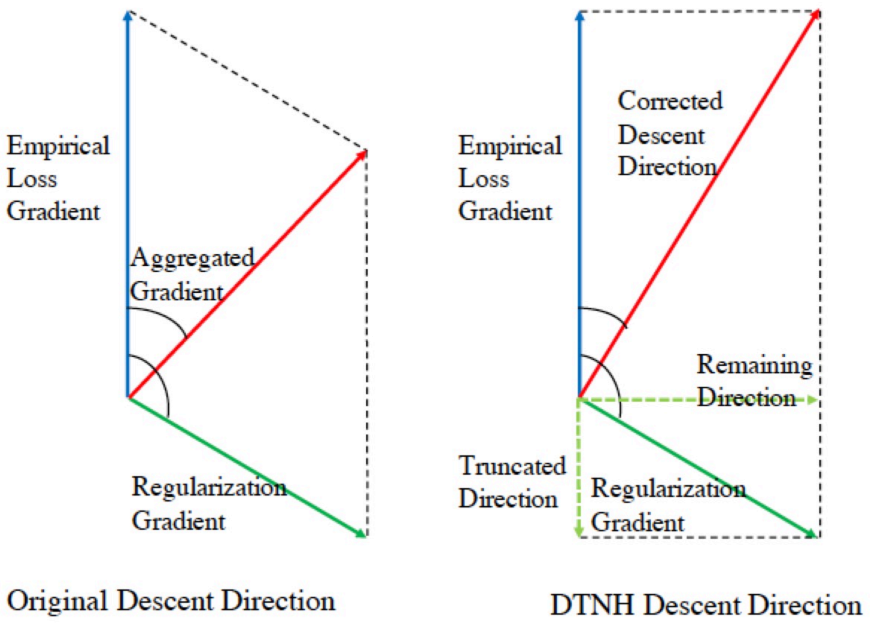



*Figure 2.* Example of **GroD** Descent Direction Estimation

[ICDM'19] Ruosi Wan, Haoyi Xiong*, Xingjian Li, Zhanxing Zhu, and Jun Huan. Towards Making Deep Transfer Learning Never Hurt. Proceedings of the 19th IEEE International Conference on Data Mining, 2019.

# AutoDL Transfer – Make Deep Transfer Learning Never Hurt (DTNH)

Table 1: Statistics on Source/Target Datasets

| Datasets | Domains | # Train/Test |
|---|---|---|
| **Source Tasks** | | |
| ImageNet | visual objects | 1,419K+/100K |
| Place 365 | indoor scenes | 10,000K+ |
| Stanford Dog | dogs | 12K/8.5K |
| **Target Tasks** | | |
| CIFAR 10 | visual objects | 50K/10K |
| Caltech 256 | visual objects | 30K+ |
| MIT Indoors 67 | indoor scenes | 5K+/1K+ |
| Flowers 102 | visual objects | 1K+/6K+ |

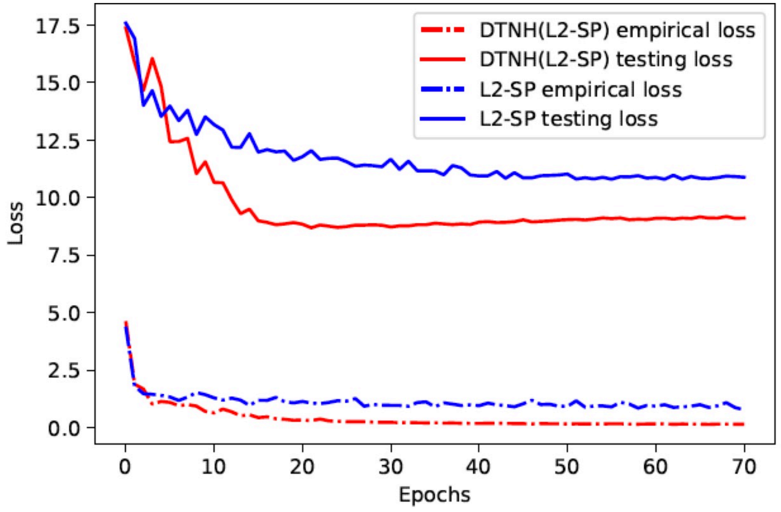



Figure 2: Empirical Loss Minimization

Table 2: Classification Accuracy Comparison from Source **ImageNet** (Numbers in RED refer to the negative transfer)

| Target Datasets | Fine-tuning | $L^2$-SP | **DTNH** ($L^2$-SP) | KnowDist | **DTNH** (KnowDist) |
|---|---|---|---|---|---|
| Caltech 256 | $82.68 \pm 0.2$ | $83.69 \pm 0.09$ | $\mathbf{84.14 \pm 0.08}$ | $82.93 \pm 0.08$ | $\mathbf{83.27 \pm 0.4}$ |
| MIT Indoors 67 | $76.73 \pm 0.77$ | $75.11 \pm 0.43$ | $\mathbf{77.46 \pm 0.29}$ | $78.05 \pm 0.32$ | $\mathbf{78.77 \pm 0.31}$ |
| Flowers 102 | $90.24 \pm 0.31$ | $88.96 \pm 0.21$ | $\mathbf{90.68 \pm 0.31}$ | $90.43 \pm 0.4$ | $\mathbf{90.91 \pm 0.4}$ |
| CIFAR 10 | $96.40 \pm 0.4$ | $93.30 \pm 0.16$ | $\mathbf{96.41 \pm 0.11}$ | $96.43 \pm 0.08$ | $\mathbf{96.57 \pm 0.2}$ |

Table 3: Classification Accuracy Comparison from Source **Places 365** (Numbers in RED refer to the negative transfer)

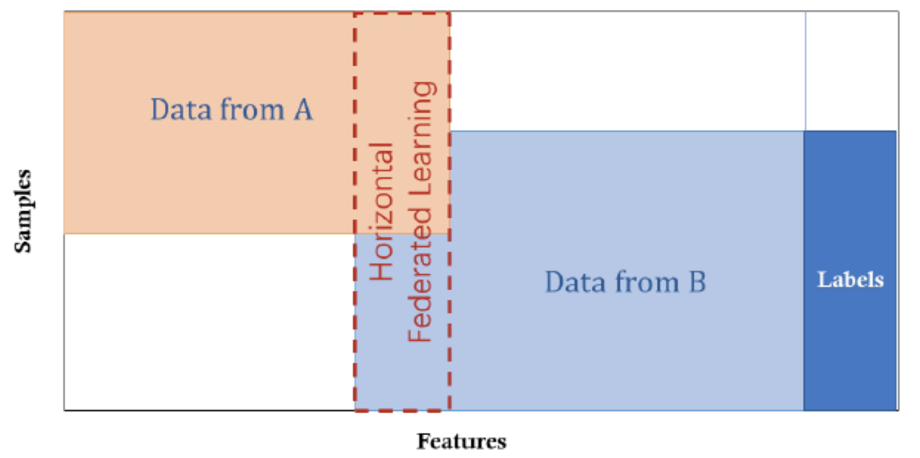
| Target Databsets | Fine-tuning | $L^2$-SP | **DTNH** ($L^2$-SP) | KnowDist | **DTNH** (KnowDist) |
|---|---|---|---|---|---|
| Caltech 256 | $73.13 \pm 0.20$ | $66.99 \pm 0.20$ | $\mathbf{73.32 \pm 0.1}$ | $72.8 \pm 0.22$ | $\mathbf{73.18 \pm 0.24}$ |
| MIT Indoors 67 | $82.64 \pm 0.16$ | $84.09 \pm 0.09$ | $\mathbf{84.19 \pm 0.07}$ | $83.29 \pm 0.42$ | $\mathbf{84.40 \pm 0.41}$ |
| Flowers 102 | $83.77 \pm 0.68$ | $77.66 \pm 0.13$ | $\mathbf{84.11 \pm 0.06}$ | $83.50 \pm 0.26$ | $\mathbf{84.12 \pm 0.56}$ |
| CIFAR 10 | $89.35 \pm 0.59$ | $89.78 \pm 0.05$ | $\mathbf{90.85 \pm 0.11}$ | $94.96 \pm 0.05$ | $\mathbf{95.02 \pm 0.13}$ |

Table 4: Classification Accuracy Comparison from Source **Stanford Dogs 120** (Numbers in RED refer to the negative transfer)

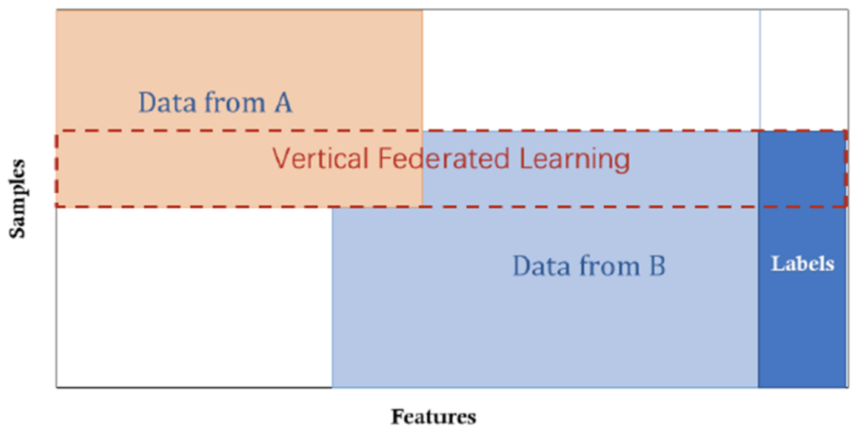
| Target Datasets | Fine-tuning | $L^2$-SP | **DTNH** ($L^2$-SP) | KnowDist | **DTNH** (KnowDist) |
|---|---|---|---|---|---|
| Caltech 256 | $82.29 \pm 0.04$ | $83.44 \pm 0.23$ | $\mathbf{83.84 \pm 0.08}$ | $82.73 \pm 0.26$ | $\mathbf{82.85 \pm 0.27}$ |
| MIT Indoors 67 | $75.69 \pm 0.21$ | $74.64 \pm 0.07$ | $\mathbf{76.46 \pm 0.22}$ | $76.36 \pm 0.19$ | $\mathbf{76.74 \pm 026}$ |
| Flowers 102 | $90.20 \pm 0.39$ | $88.14 \pm 0.06$ | $\mathbf{89.98 \pm 0.04}$ | $89.86 \pm 0.07$ | $\mathbf{90.29 \pm 0.34}$ |
| CIFAR 10 | $96.34 \pm 0.13$ | $94.16 \pm 0.10$ | $\mathbf{96.39 \pm 0.08}$ | $96.11 \pm 0.53$ | $\mathbf{96.41 \pm 0.18}$ |

[ICDM'19] Ruosi Wan, Haoyi Xiong*, Xingjian Li, Zhanxing Zhu, and Jun Huan. Towards Making Deep Transfer Learning Never Hurt. Proceedings of the 19th IEEE International Conference on Data Mining, 2019.

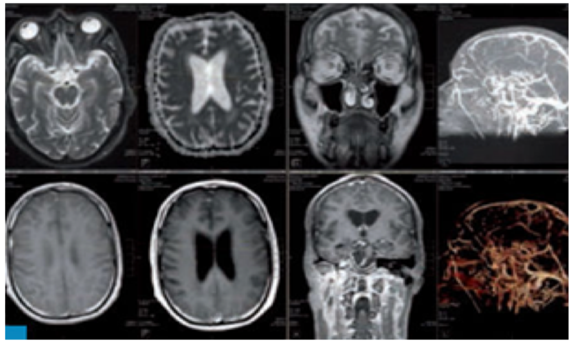
# AutoDL Federation: Federated Learning with Vertical "View Separation"



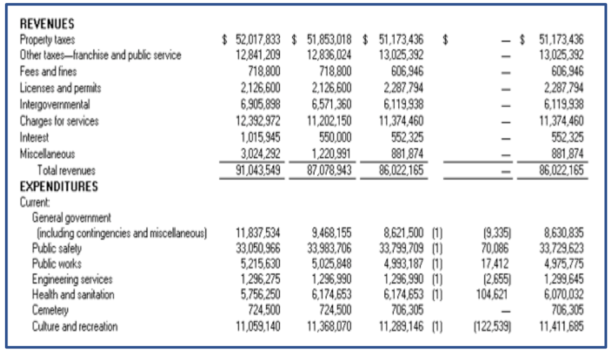

$$\mathcal{X}_i = \mathcal{X}_j,\ \ \mathcal{Y}_i = \mathcal{Y}_j,\ \ I_i \neq I_j,\ \ \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j$$




$$\mathcal{X}_i \neq \mathcal{X}_j,\ \ \mathcal{Y}_i \neq \mathcal{Y}_j,\ \ I_i = I_j\ \ \forall \mathcal{D}_i, \mathcal{D}_j, i \neq j$$



Medical Data



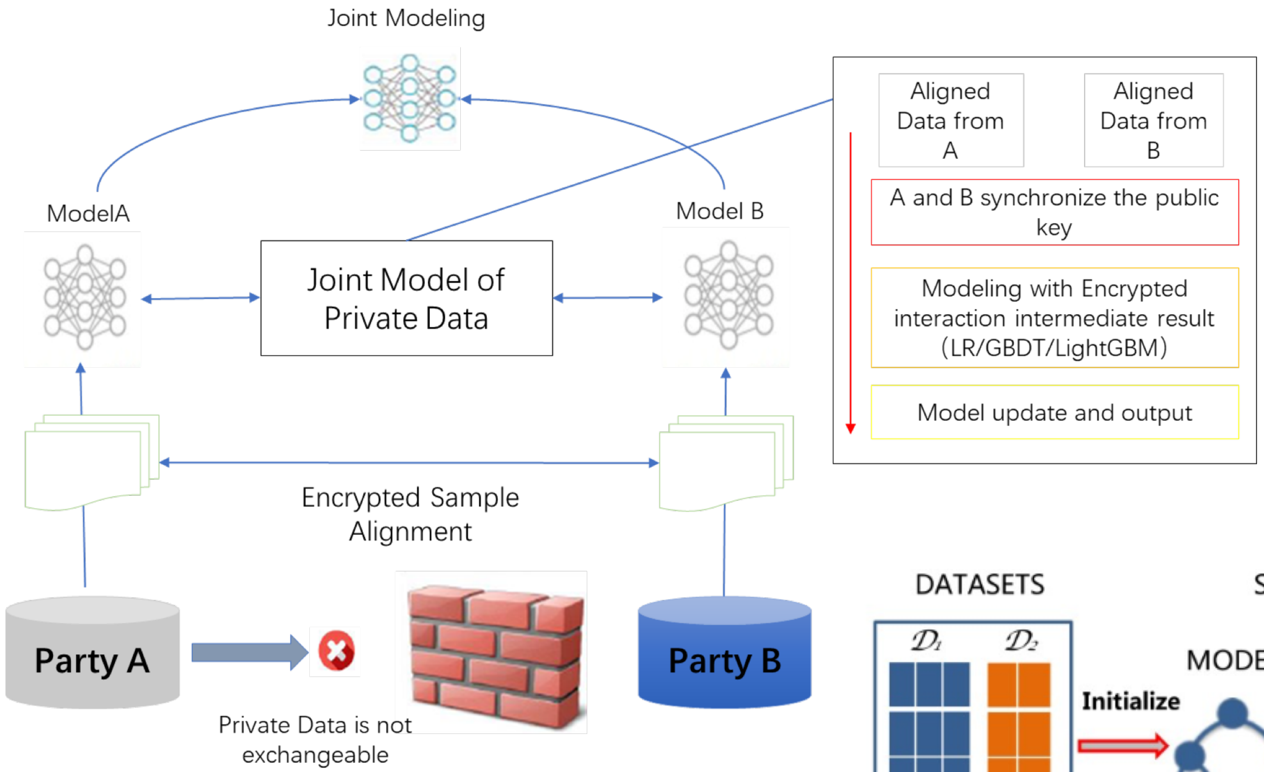
| | | | | | | |
|---|---|---|---|---|---|---|
| **REVENUES** | | | | | | |
| Property taxes | $ 52,017,833 | $ 51,853,018 | $ 51,173,436 | | $ — | $ 51,173,436 |
| Other taxes—franchise and public service | 12,841,209 | 12,836,024 | 13,025,392 | | — | 13,025,392 |
| Fees and fines | 718,800 | 718,800 | 606,946 | | — | 606,946 |
| Licenses and permits | 2,126,600 | 2,126,600 | 2,287,794 | | — | 2,287,794 |
| Intergovernmental | 6,905,898 | 6,571,360 | 6,119,938 | | — | 6,119,938 |
| Charges for services | 12,392,972 | 11,202,150 | 11,374,460 | | — | 11,374,460 |
| Interest | 1,015,945 | 550,000 | 552,325 | | — | 552,325 |
| Miscellaneous | 3,024,292 | 1,220,991 | 881,874 | | — | 881,874 |
| Total revenues | 91,043,549 | 87,078,943 | 86,022,165 | | — | 86,022,165 |
| **EXPENDITURES** | | | | | | |
| Current: | | | | | | |
| General government (including contingencies and miscellaneous) | 11,837,534 | 9,468,155 | 8,621,500 | (1) | (9,335) | 8,630,835 |
| Public safety | 33,050,966 | 33,983,706 | 33,799,709 | (1) | 70,086 | 33,729,623 |
| Public works | 5,215,630 | 5,025,848 | 4,993,187 | (1) | 17,412 | 4,975,775 |
| Engineering services | 1,296,275 | 1,296,990 | 1,296,990 | (1) | (2,655) | 1,299,645 |
| Health and sanitation | 5,756,250 | 6,174,653 | 6,174,653 | (1) | 104,621 | 6,070,032 |
| Cemetery | 724,500 | 724,500 | 706,305 | | — | 706,305 |
| Culture and recreation | 11,059,140 | 11,368,070 | 11,289,146 | (1) | (122,539) | 11,411,685 |

Bank Data

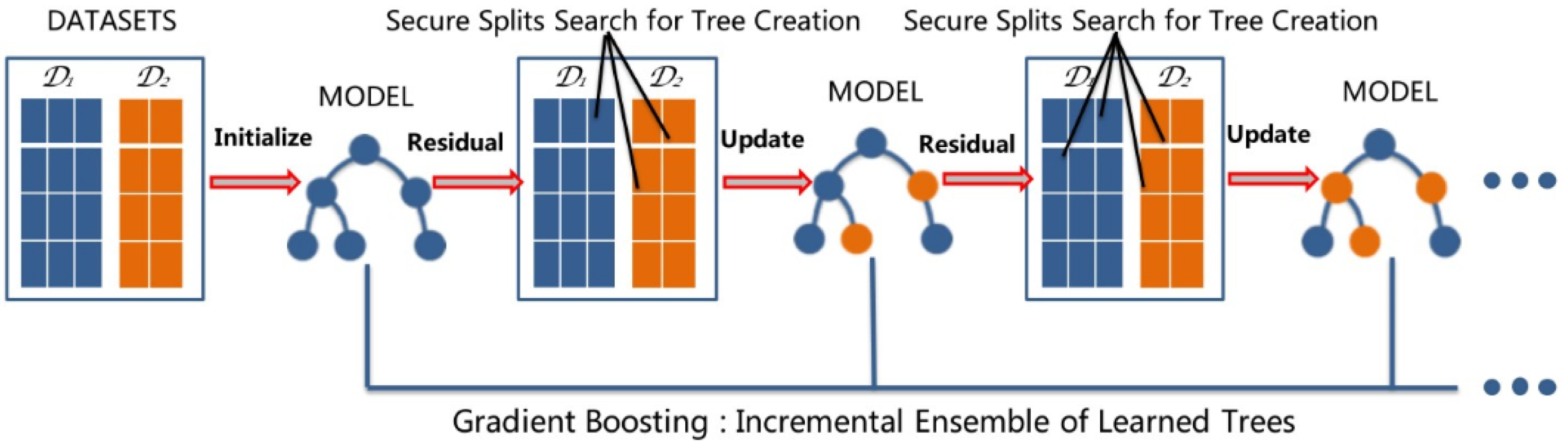

Insurance Data

Learn a model for Insurance Recommendation With "Privacy" Protection

# AutoDL Federation with SecureGBM




Our methods
- Distributed Gradient Boosting Decision Trees (GBDT)
- Semi-Homomorphic Encryption (SHE)
- Multi-Party Computation (MPC)
- GPU + Statistical Acceleration




[BigData'19] Zhi Feng, **Haoyi Xiong***, Chuanyuan Song, Sijia Yang, Baoxin Zhao, Licheng Wang, Zeyu Chen, Liping Liu, and Jun Huan. SecureGBM: Secure Multi-Party Gradient Boosting. The 2019 IEEE International Conference on Big Data, Los Angeles, CA, 2019.

Overall Classification AUC (%) Comparison (N/A: During the experiments, LightGBM reported failure to train the model due as the features of the given datasets are too sparse to learn.)

# AutoDL Federation with SecureGBM

[BigData'19] Zhi Feng, **Haoyi Xiong***, Chuanyuan Song, Sijia Yang, Baoxin Zhao, Licheng Wang, Zeyu Chen, Liping Liu, and Jun Huan. SecureGBM: Secure Multi-Party Gradient Boosting. The 2019 IEEE International Conference on Big Data, Los Angeles, CA, 2019.

| | Sparse | | Adult | | Phishing | |
|---|---|---|---|---|---|---|
| Methods | Training | Testing | Training | Testing | Training | Testing |
| SecureGBM | 93.227 | 66.220 | 92.465 | 90.080 | 62.855 | 61.823 |
| Using the Aggregated Datasets from $\mathbb{A}$ and $\mathbb{B}$ | | | | | | |
| LightGBM-($\mathbb{A}$,$\mathbb{B}$) | 96.102 | 68.528 | 92.199 | 90.145 | 67.994 | 63.430 |
| XGBoost-($\mathbb{A}$,$\mathbb{B}$) | 93.120 | 67.220 | 91.830 | 89.340 | 67.090 | 61.990 |
| LIBSVM-Linear-($\mathbb{A}$,$\mathbb{B}$) | 73.490 | 64.560 | 58.641 | 59.280 | 50.073 | 50.980 |
| LIBSVM-RBF-($\mathbb{A}$,$\mathbb{B}$) | 79.850 | 63.210 | 75.549 | 72.060 | 52.789 | 47.479 |
| Using the Single Dataset at $\mathbb{A}$ | | | | | | |
| LightGBM-$\mathbb{A}$ | N/A* | N/A* | 89.849 | 88.052 | 64.693 | 59.743 |
| XGBoost-$\mathbb{A}$ | 65.170 | 57.370 | 89.490 | 87.620 | 64.070 | 59.740 |
| LIBSVM-Linear-$\mathbb{A}$ | 52.360 | 50.675 | 66.293 | 34.347 | 50.007 | 50.489 |
| LIBSVM-RBF-$\mathbb{A}$ | 56.740 | 52.380 | 72.909 | 55.076 | 50.248 | 50.306 |
| Using the Datasets that aggregate Features from $\mathbb{B}$ and Labels from $\mathbb{A}$ (Not exist in the real case) | | | | | | |
| LightGBM-$\mathbb{B}$* | 96.102 | 68.528 | 85.708 | 84.587 | 62.396 | 58.929 |
| XGBoost-$\mathbb{B}$* | 93.190 | 67.390 | 85.700 | 85.410 | 61.720 | 58.420 |
| LIBSVM-Linear-$\mathbb{B}$* | 67.480 | 60.990 | 46.527 | 46.840 | 50.627 | 48.567 |
| LIBSVM-RBF-$\mathbb{B}$* | 78.230 | 64.880 | 56.927 | 74.987 | 50.336 | 50.415 |

| # Samples | 1,000 | 2,000 | 4,000 | 8,000 | 16,000 |
|---|---|---|---|---|---|
| SecureGBM | 10.20 | 10.50 | 11.4 | 11.75 | 14.30 |
| LightGBM | 0.16 | 0.32 | 0.70 | 1.41 | 2.45 |
| XGBoost | 0.51 | 0.73 | 1.09 | 2.20 | 4.30 |

- Time Consumption per Iteration (seconds) on a Synthesized Dataset over Varying Number of Training Samples

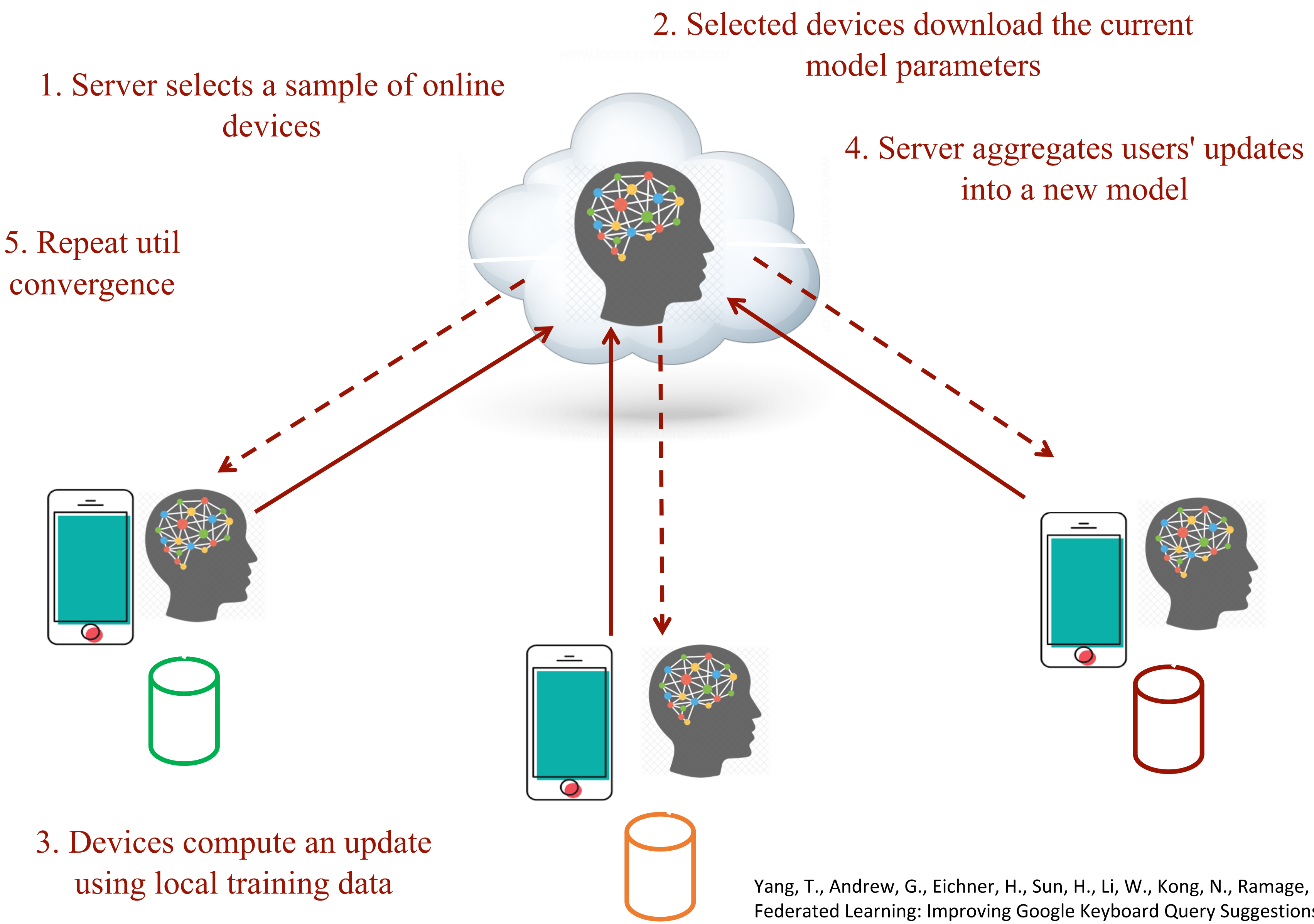
# AutoDL Edge Inference – Accelerated and Energy Efficient Inference over Mobile Devices

## AutoDL Edge

Efficient model compression with performance/accuracy preserved for model deployment on pervasive Edge Computing devices

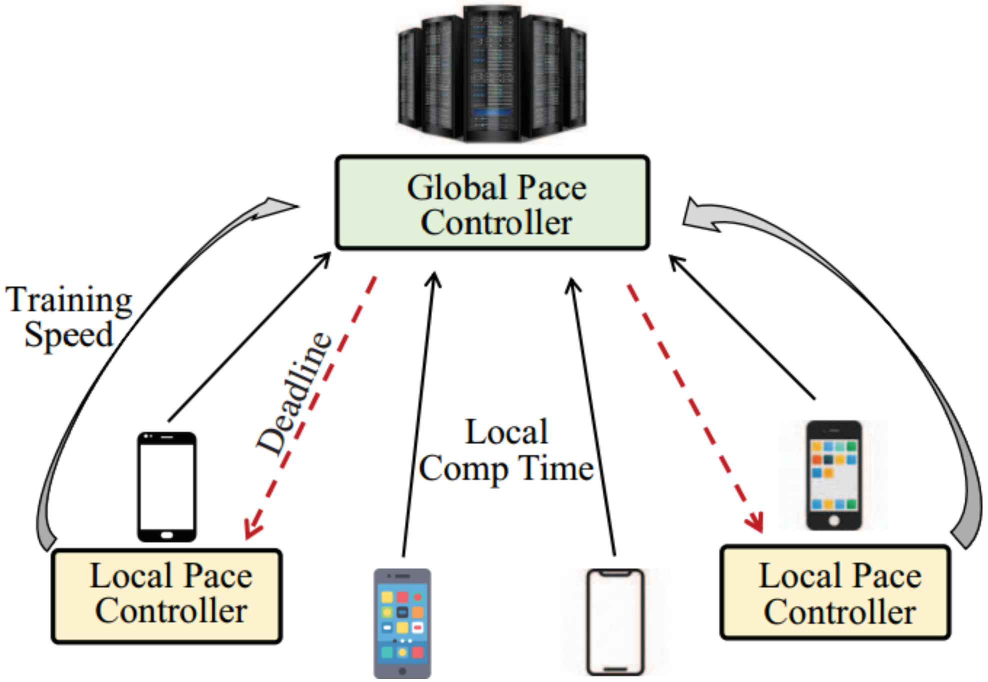
| | | Before Compression | | After Compression | | Compression Ratio |
|---|---|---|---|---|---|---|
| | | Parameters | Accuracy | Parameters | Accuracy | |
| SoundNet on ESC-50 | | 13.00M | 0.66 | 0.07M | 0.656 | 180 |
| ResNet on CIFAR-10 | ResNet-18 | 11.17M | 0.9418 | 0.82M | 0.939 | 13.62 |
| | ResNet-34 | 21.28M | 0.9472 | 1.69M | 0.9429 | 12.59 |
| | ResNet-50 | 23.52M | 0.9516 | 3.97M | 0.9491 | 5.92 |
| DenseNet on CIFAR-10 | DenseNet121 | 6.96M | 0.9513 | 1.75M | 0.9472 | 3.97 |

# AutoDL Edge Training: Efficient Federated Deep Learning over Mobile Devices




Yang, T., Andrew, G., Eichner, H., Sun, H., Li, W., Kong, N., Ramage, D. and Beaufays, F., 2018. Applied Federated Learning: Improving Google Keyboard Query Suggestions.

# AutoDL Edge Training: Efficient Federated Deep Learning over Mobile Devices with SmartPC (Smart Pace Control)



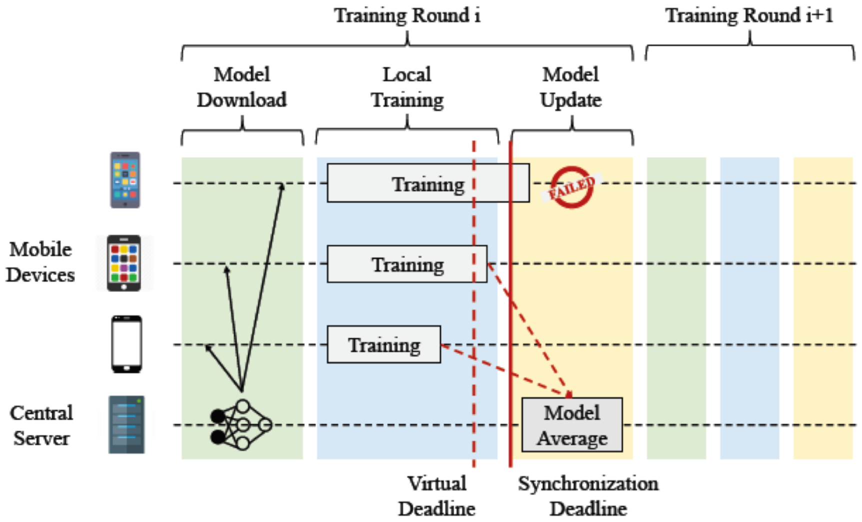




Fig. 6: Deadline assignment and local training.

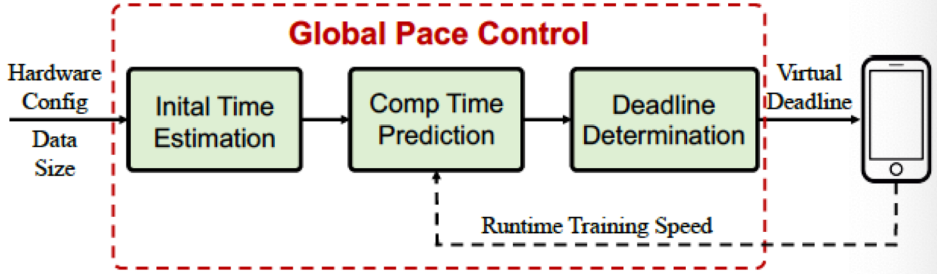



Fig. 7: Workflow of Global Pace Control.

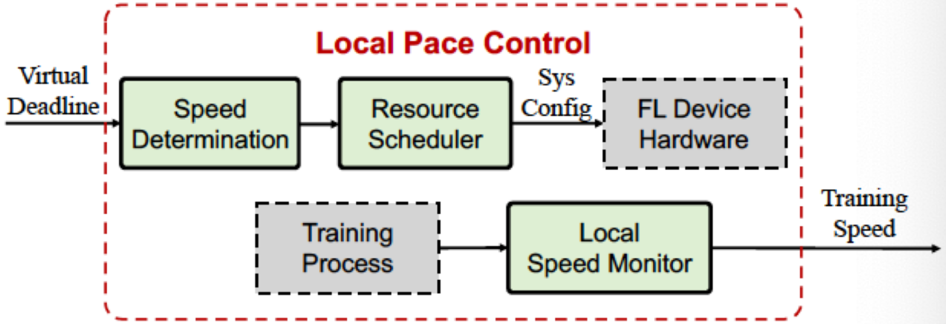
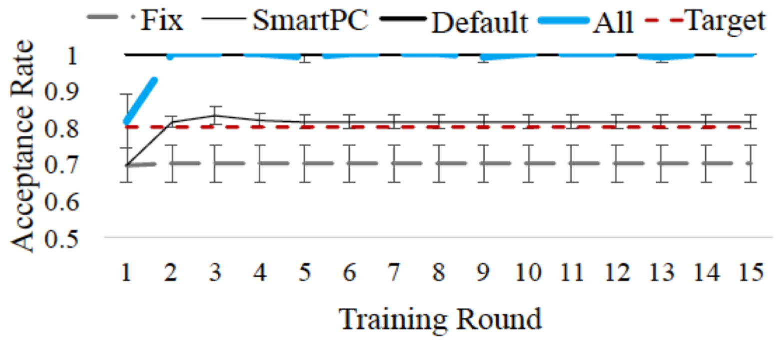
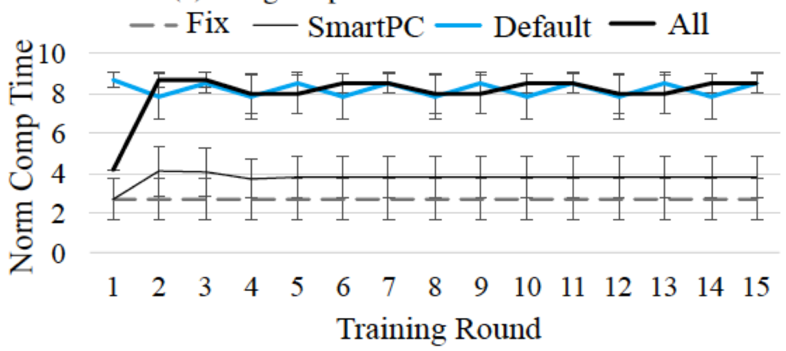



Fig. 8: Local pace control in SmartPC.

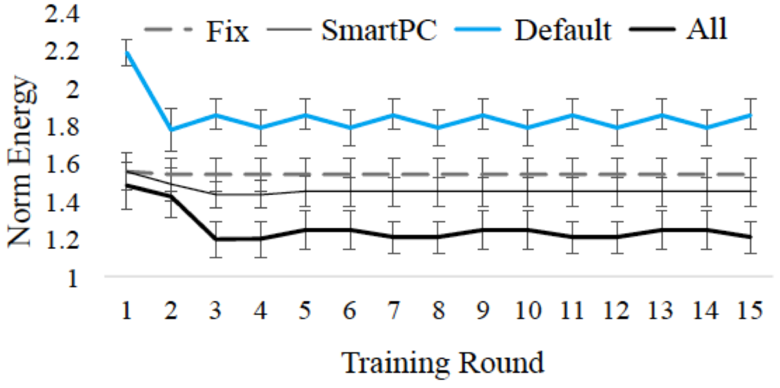
[RTSS'19] Li Li*, Haoyi Xiong*, Zhishan Guo, Jun Wang, and Cheng-Zhong Xu. SmartPC: Hierarchical Pace Control in Real-Time Federated Learning System. In Proceedings of the the 40th IEEE Real-Time Systems Symposium, Hong Kong, China, 2019, IEEE.




(a) Weight updates submission rate.

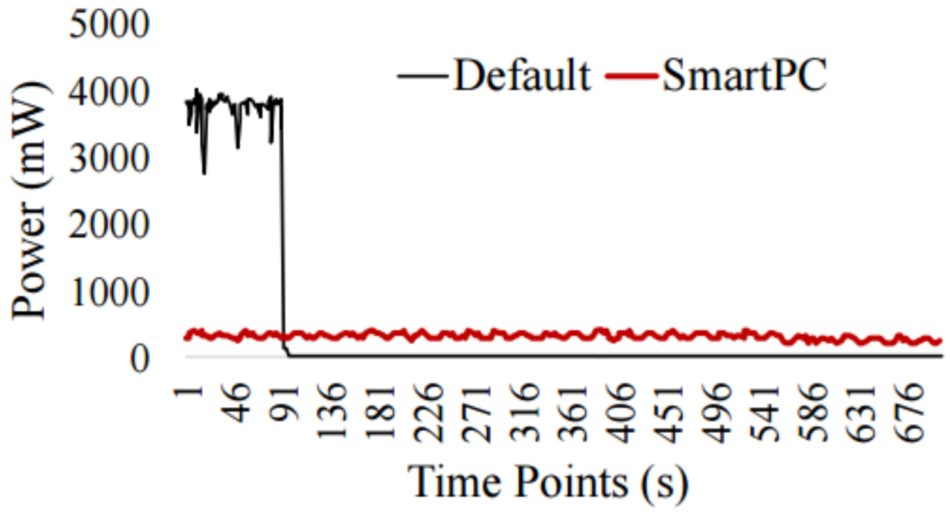



(b) Completion time of each round.

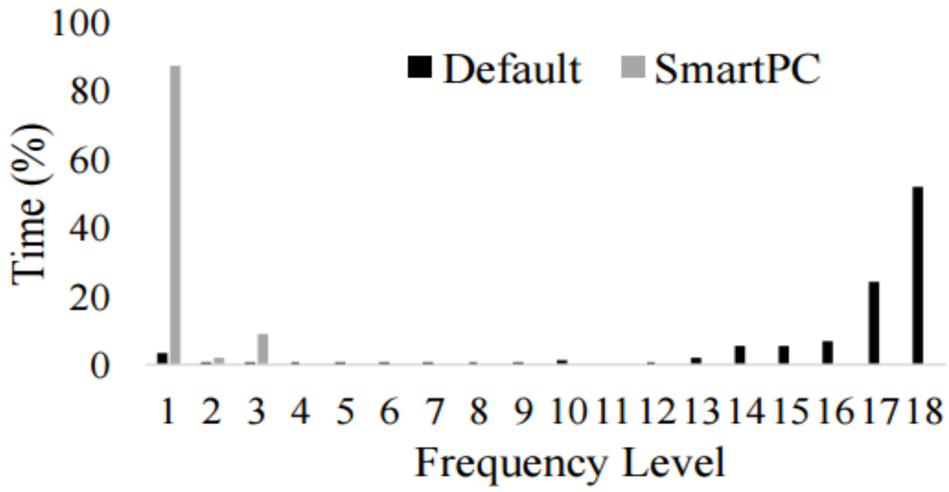



(c) Energy consumption of each round.

Fig. 10: Comparison of different schemes in global pace control.




(a) Power consumption of one training round.




(b) Frequency level selection with different schemes.

[RTSS'19] Li Li*, Haoyi Xiong*, Zhishan Guo, Jun Wang, and Cheng-Zhong Xu. SmartPC: Hierarchical Pace Control in Real-Time Federated Learning System. In Proceedings of the the 40th IEEE Real-Time Systems Symposium, Hong Kong, China, 2019, IEEE.

# Broader Impacts and Intellectual Merits

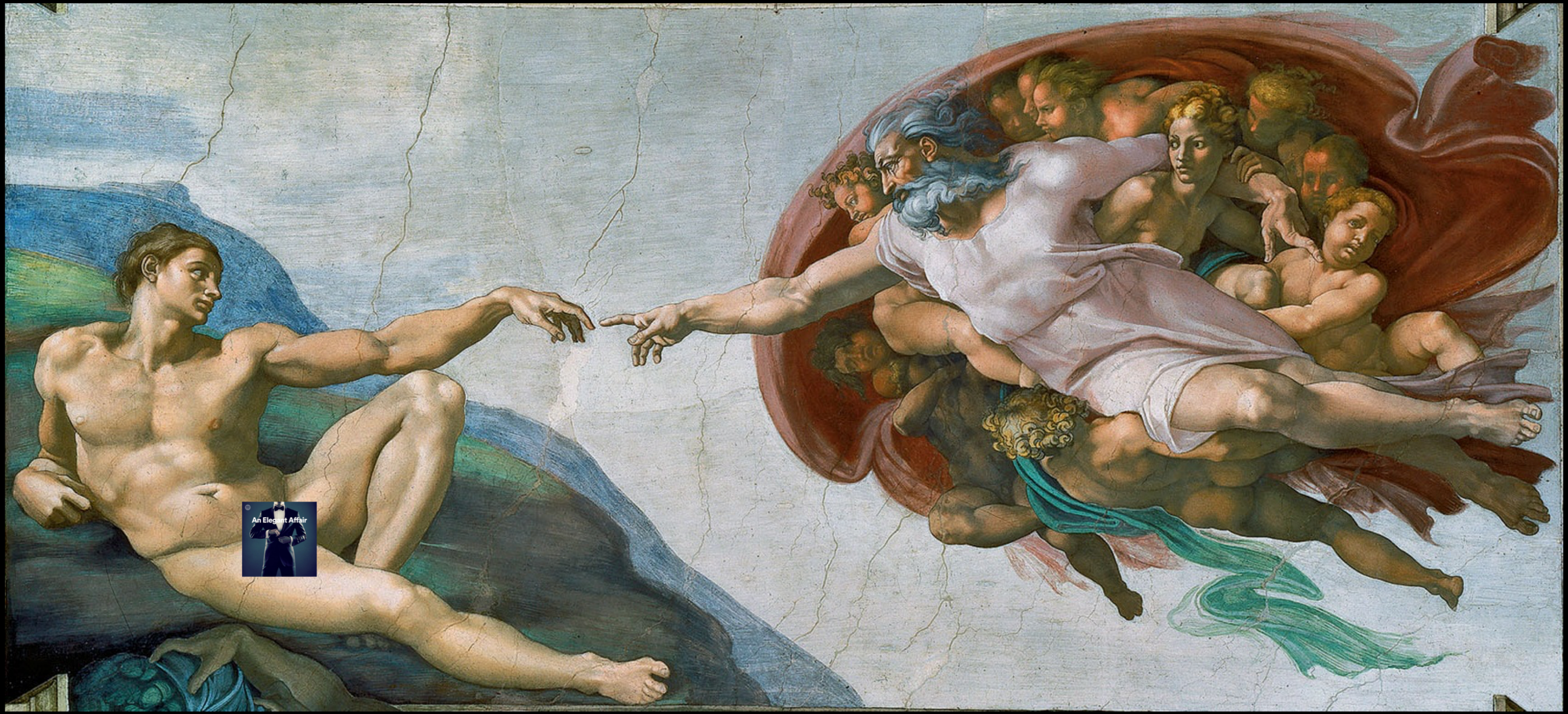
- Intellectual Merits
  - New infrastructure for AI Industrialization
  - New algorithms and research that push the frontiers of automated deep learning
  - Innovative applications that leverage the power of AI to do AI⋯

- Broader Impacts
  - Quest for intelligence

  - New interface to design and implement AI
    - like human-readable programming language (vs. machine code) to programming

  - Data-driven? Human-driven? Or Model-driven? ⋯

# Who is Adam··· Has he come?



“Creazione di Adamo” from Sistine Chapel ceiling
by *Michelangelo di Lodovico Buonarroti Simoni*