# PaddlePaddle

## Baidu Deep Learning Open Source Framework

2019.01

**PA**rallel **D**istributed **D**eep **L**earning
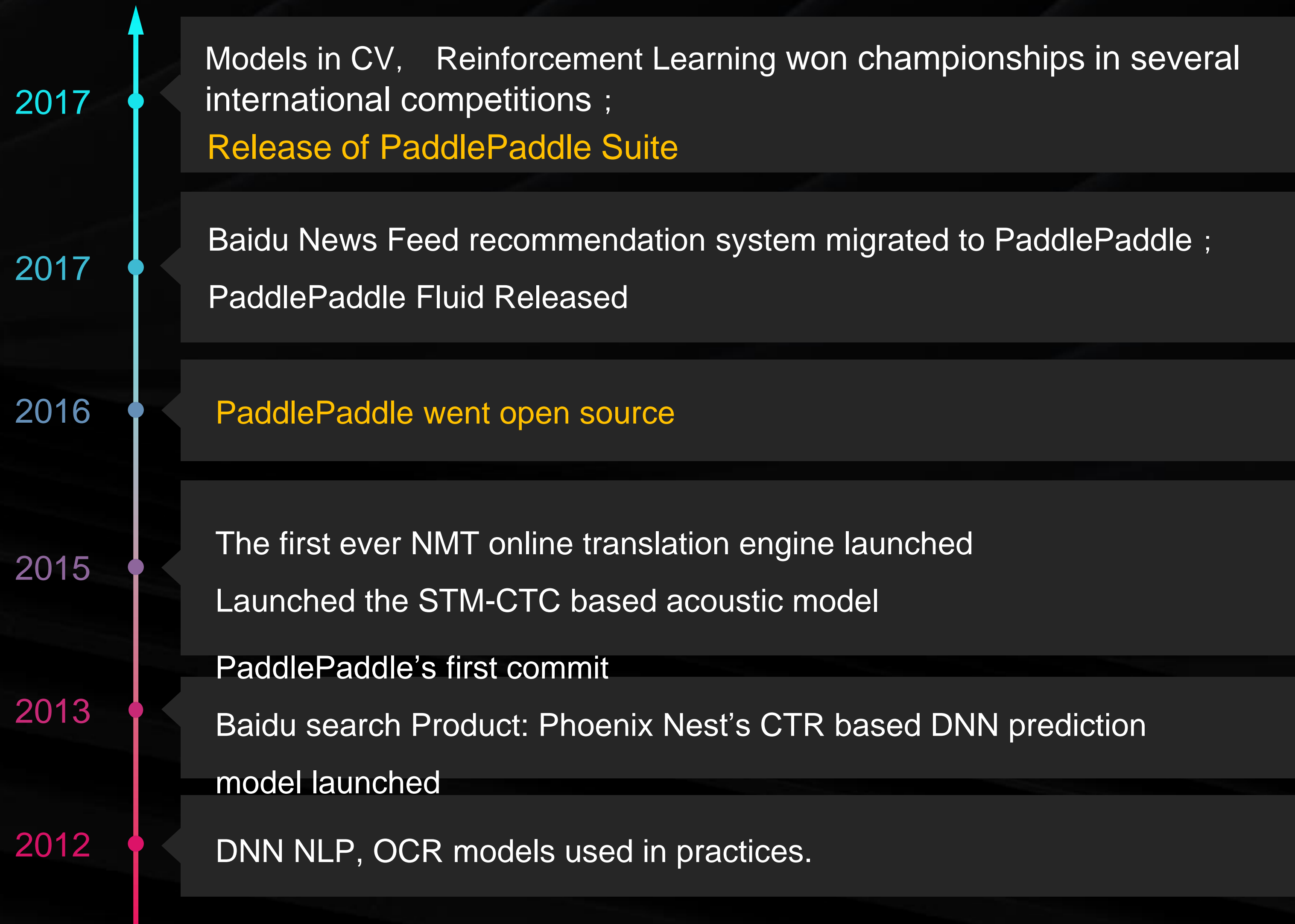
http://www.paddlepaddle.org

# Agenda

A. **Brief Introduction**

B. **Ecosystem**

C. **Real World Use Cases**

D. **Developers Community**

# History of Baidu Deep Learning

**2017** — Models in CV, Reinforcement Learning won championships in several international competitions；
Release of PaddlePaddle Suite

**2017** — Baidu News Feed recommendation system migrated to PaddlePaddle；

PaddlePaddle Fluid Released

**2016** — PaddlePaddle went open source

**2015** — The first ever NMT online translation engine launched

Launched the STM-CTC based acoustic model

PaddlePaddle's first commit

**2013** — Baidu search Product: Phoenix Nest's CTR based DNN prediction

model launched

**2012** — DNN NLP, OCR models used in practices.

# PaddlePaddle 3.0 – Towards Maturity

## PaddlePaddle 1.0

Initial Open Source Edition, command line interaction interface,
Support common DL networks

## PaddlePaddle 2.0

Friendly python API,
Released the core framework , model zoo and Paddle Book,
Improved the ease of use and flexibility

## PaddlePaddle 3.0

July 2018
Released PaddlePaddle 3.0
Including functional components like EasyDL, AI Studio, AutoDL, VisualDL

Nov. 2018
Released PaddlePaddle Fluid 1.0
Release PaddlePaddle Suite
--Full-featured Deep
Learning development kit for businesses and developers

# Widely Recognized by the Government and Industry

**The Only Deep Learning Technology and Application National Engineering Laboratory**

| Engineering platform | Development Tools | Open Data Set | Education and Training |
|---|---|---|---|

PaddlePaddle has established a "Deep Learning National Team" with a number of domestic research institutions and universities.

# Open Sourced Several International Competition Winning Models

The world's top technology level, leading the direction of deep learning technology

| Award-Winning Model | | Awards | |
|---|---|---|---|
| CV | PyramidBo Model | WIDER FACE （3 test subsets） | First place |
| | Attention Clusters Network Model | ActivityNet2017/2018 kinetics | First place |
| | Several Model Based on Faster R-CNN | Google AI Open Images-Object Detection Track | First place |
| PARL(Reinforcement Learning) | | NIPS AI for Prosthetics Challenge | First place |

# Agenda

A. **Brief Introduction**

B. **Ecosystem**

C. **Real World Use Cases**

D. **Developers Community**

# PaddlePaddle Suite

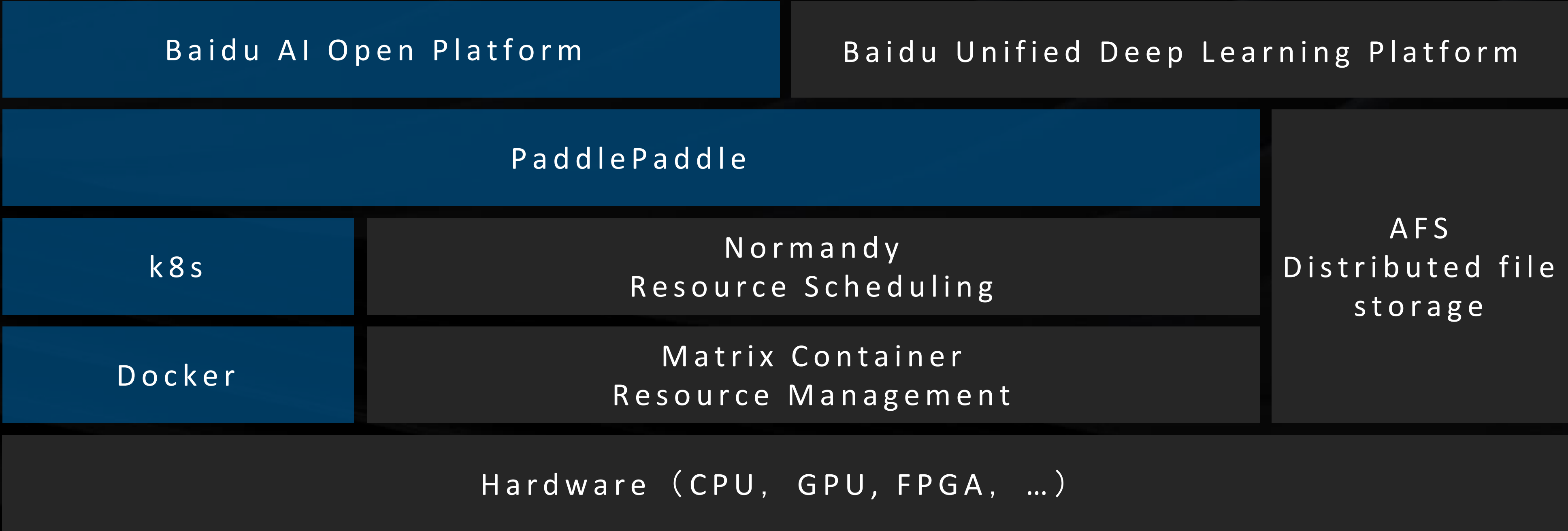## Full-featured Deep Learning Suite with Comprehensive, Leading Technology

### Service Platform

**EasyDL**
Zero-based customized training and service platform

**AI Studio**
One-stop development platform

**AutoDL**
Network structure automation design

### Modules and Components

**Visual DL**
Visualization Tool for Training

### Core Framework

**PaddleRec**
Intelligent Recommendation

**PaddleCV**
Intelligent vision

**PaddleNLP**
Intelligent text processing

**PARL**
Deep Reinforcement Learning

**Paddle Fluid**

**Paddle Serving**

**AutoDL**

**EDL**
Elastic deep learning calculation

# Features of PaddlePaddle Core framework

| Heterogeneous Computing | Parallel Training | Multiple Algorithms | Multi-End Deployment |
| --- | --- | --- | --- |
| Fully supports for large-scale heterogeneous computing clusters CPU、GPU、DSP、FPGA | Supports multi-machine multi-thread asynchronous training and synchronous training mode | Personalized recommendation, image classification, semantic segmentation, face detection, machine translation, reading comprehension, lexical analysis, sentiment analysis | Rapid Deployment Multiple mobile end support |

# Large-Scale Heterogeneous Computing Cluster

| Baidu AI Open Platform | Baidu Unified Deep Learning Platform |

| PaddlePaddle | AFS Distributed file storage |

| k8s | Normandy Resource Scheduling |

| Docker | Matrix Container Resource Management |

| Hardware（CPU, GPU, FPGA, …） |

■ Open source or open module

# Supports Parallel Training of Dense Parameters and Sparse Parameters



**Large-scale dense parameter**

Data0   Data1   Data2   Data3

GPU 0   GPU 1   GPU 2   GPU 3
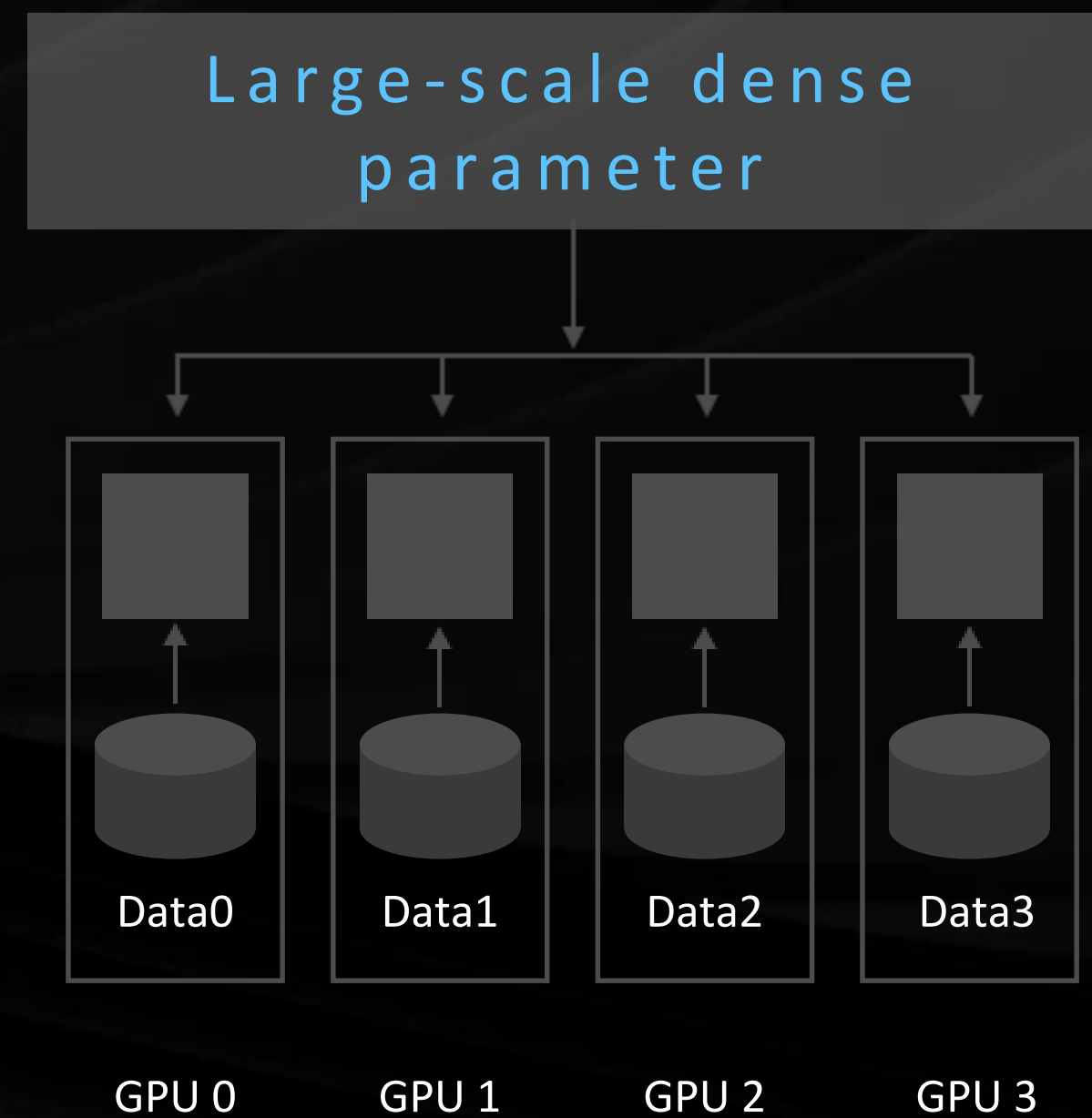
Computationally intensive tasks such as image classification and machine translation

Parameter synchronization mode : Synchronous Collective operation

GPU parallel training speed surpasses similar frameworks in mainstream tasks

**Ultra-Large-Scale sparse parameter**

Data0   Data1   Data2   Data3   Data4
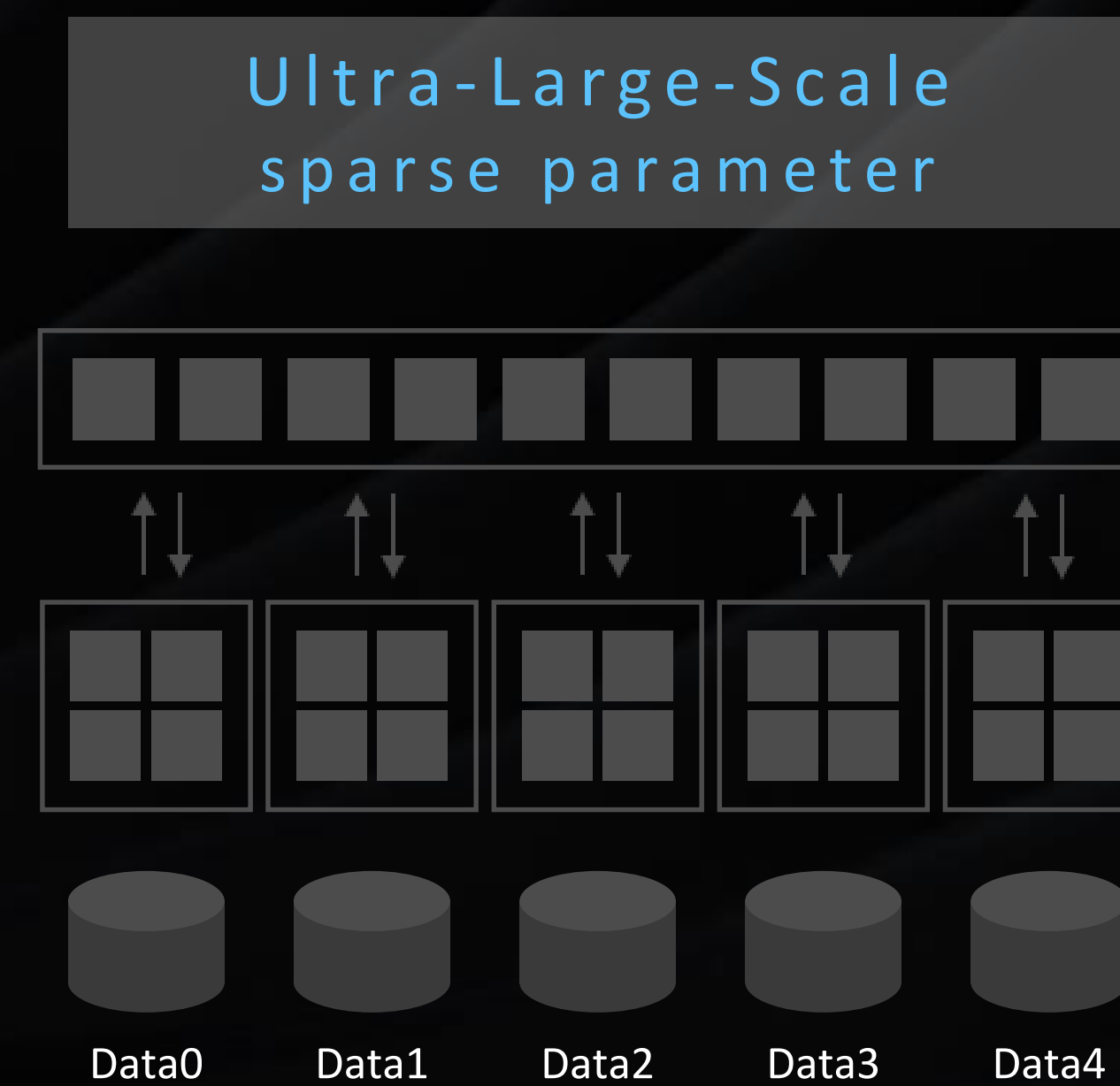
CTR estimation, semantic matching, and other tasks with large data throughput

Parameter synchronization mode:
Asynchronous large-scale sparse parameter server

CPU based ultra-large-scale asynchronous training is unique, supporting 100 billion scale parameters, hundreds of nodes parallel training

# Keep Building the Most Complete Model Collection

| Scenario | PaddleRec – Recommendation | | | PaddleRec - CV intelligence | | | PaddleRec - NLP | | |
|---|---|---|---|---|---|---|---|---|---|
| | Feed | Intelligent marketing | Video analysis | Medical imaging | Autonomous driving | Industry inspection | public sentiment | Search engine | Machine translation | Intelligent dialogue |
| Models set | Provision of many classic recall and ranking algorithms | | | Covers all the cv application scenarios | | | Fulfills mainstream NLP tasks | | |
| | DeepCTR | GRU4Rec | Text label | Image classification | Object detection | Face detection | Chinese | semantic matching | Comprehension |
| | | | | OCR | Semantic Segmentation | GAN | | | |
| | Sequence semantic recall | Multi-view Simnet | | Metric learning | Video classification | | 机器翻译 | Chinese semantic segmentation | |
| Application examples | Baidu feed | Haokan video | Baidu Map | Baidu OCR | Baidu feed | Baidu Nuomi | Baidu | Baidu translation | |

# Multi-platform Service Deployment

## Paddle Serving

- Flexible adaptation to multiple inference engines
- Compatible with mainstream engine TensorRT
- Inference API, lib library
- CPU, GPU performance deep optimization
- Forward pass specific optimization

## Paddle Mobile

- Multiple hardware platform support: ARM CPU, Mali GPU, Qualcomm DSP, FPGA
- Fixed point quantization
- Low precision and efficient quantitative calculation

## Paddle Anywhere
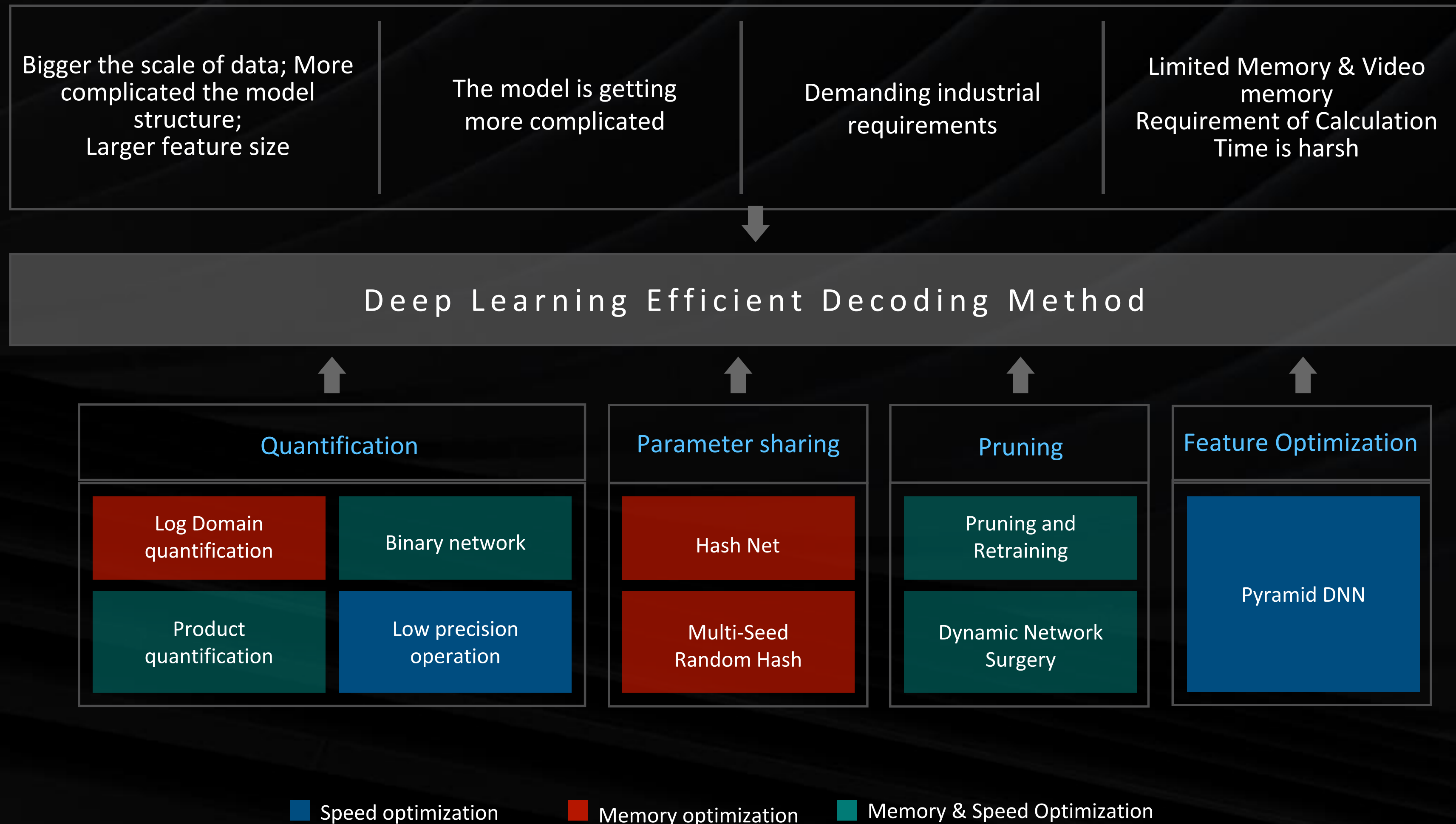
# PaddlePaddle Assistive Tools and Platform

| PARL | Visual DL | AutoDL | EasyDL | AI Studio |
|------|-----------|--------|--------|-----------|
| Deep Reinforcement learning | Visualized Deep Learning Tool | Automatic Network Structure Design | Zero skill required deep learning training and service platform | One-stop development platform |

# PARL
## Tools for Reinforced Learning

Env1 | Env2 | Env3

Agent

CPU1 | CPU2 | CPU3

Agent Wrapper
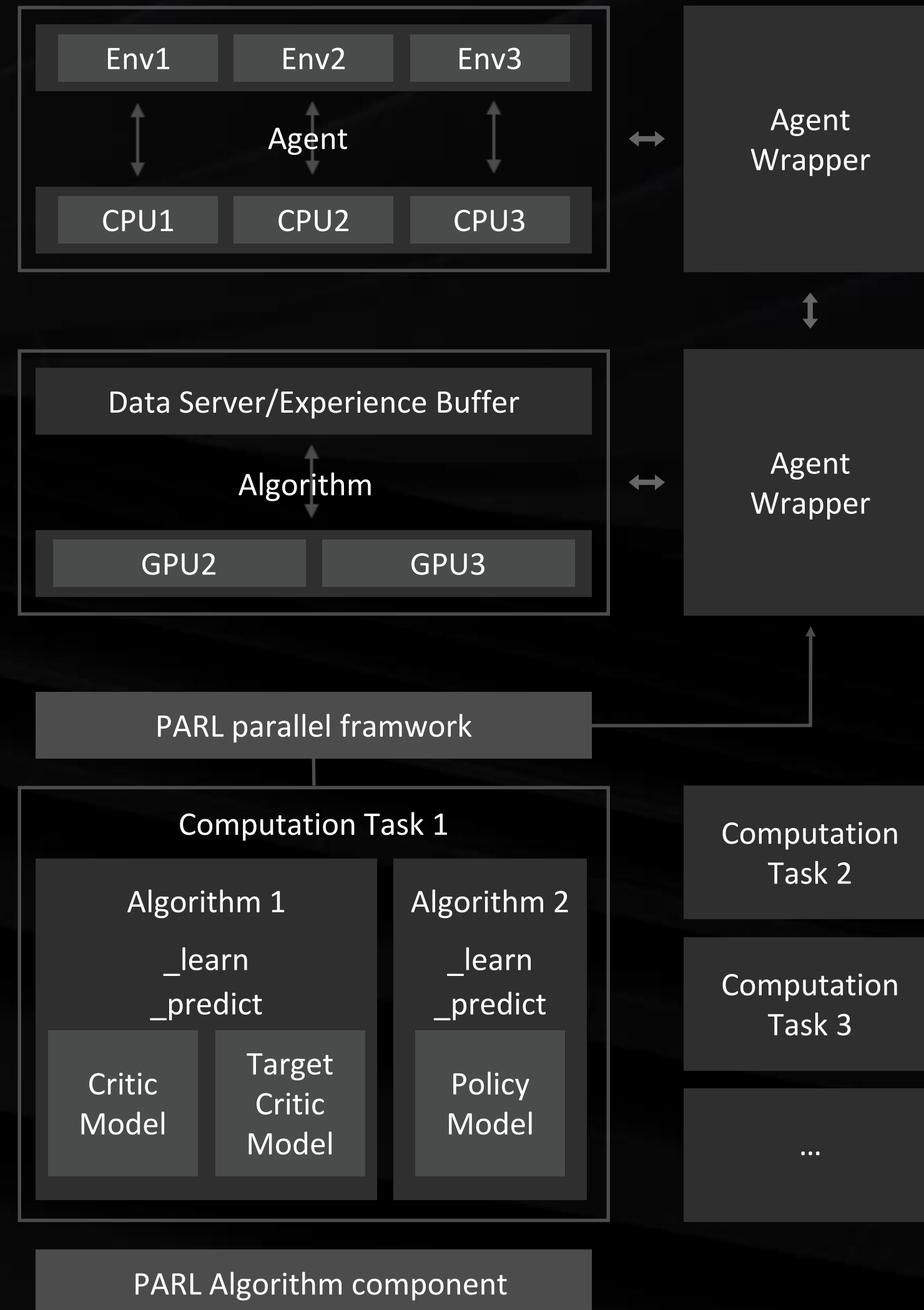
**Won NIPS 2018 AI Prosthetics Challenge**

Data Server/Experience Buffer

Algorithm

GPU2 | GPU3

Agent Wrapper

PARL parallel framwork

Computation Task 1

Algorithm 1
_learn
_predict

Critic Model | Target Critic Model

Algorithm 2
_learn
_predict

Policy Model

Computation Task 2

Computation Task 3

...

PARL Algorithm component
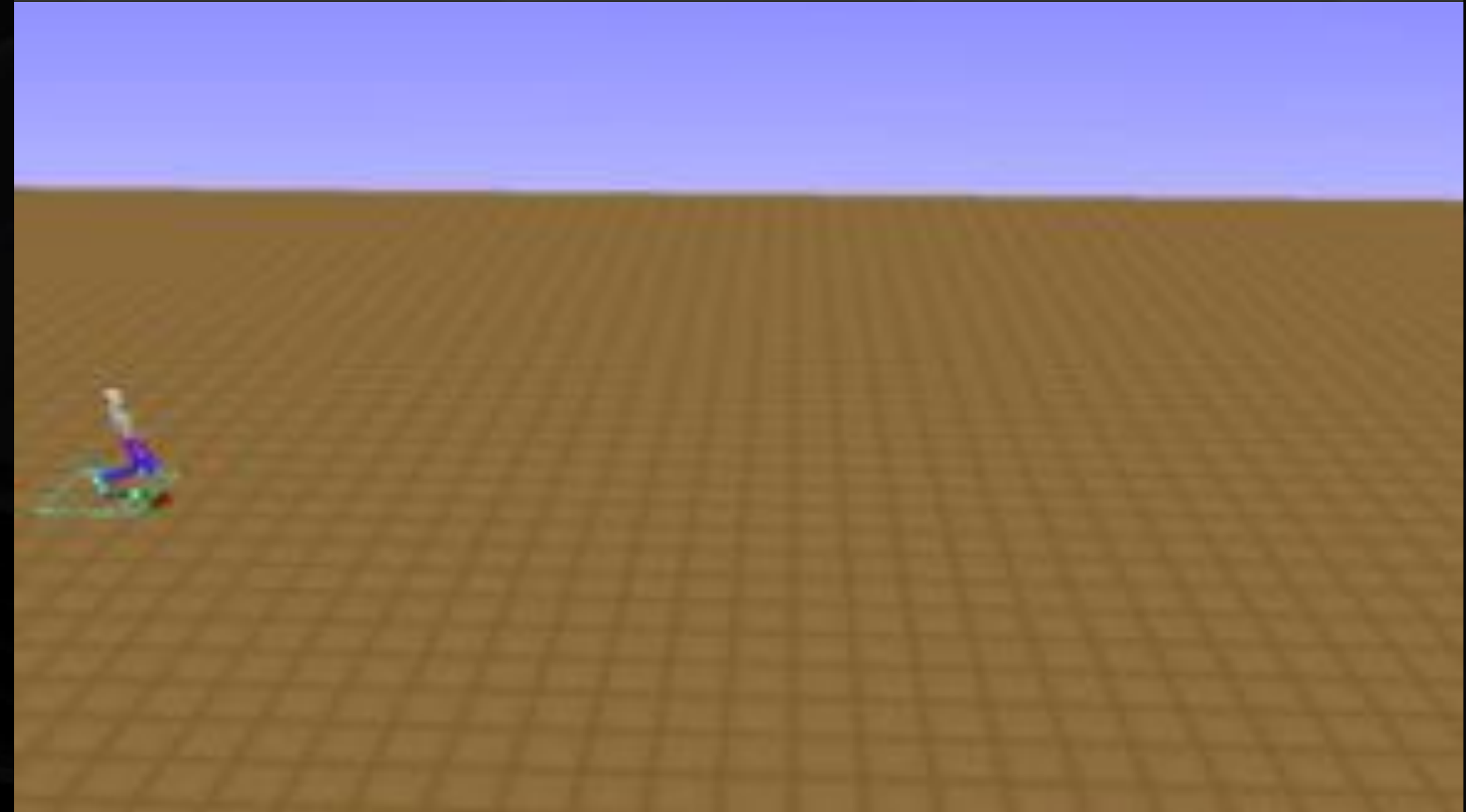
Target Driven DDPG + Bootstrapping
One thousand of CPU + Single GPU

# Visual DL
## Visualize the overall Process of Training and Inferring



Scalar

Six components on visualization

Two SDK : C++, Python

Supports ONNX
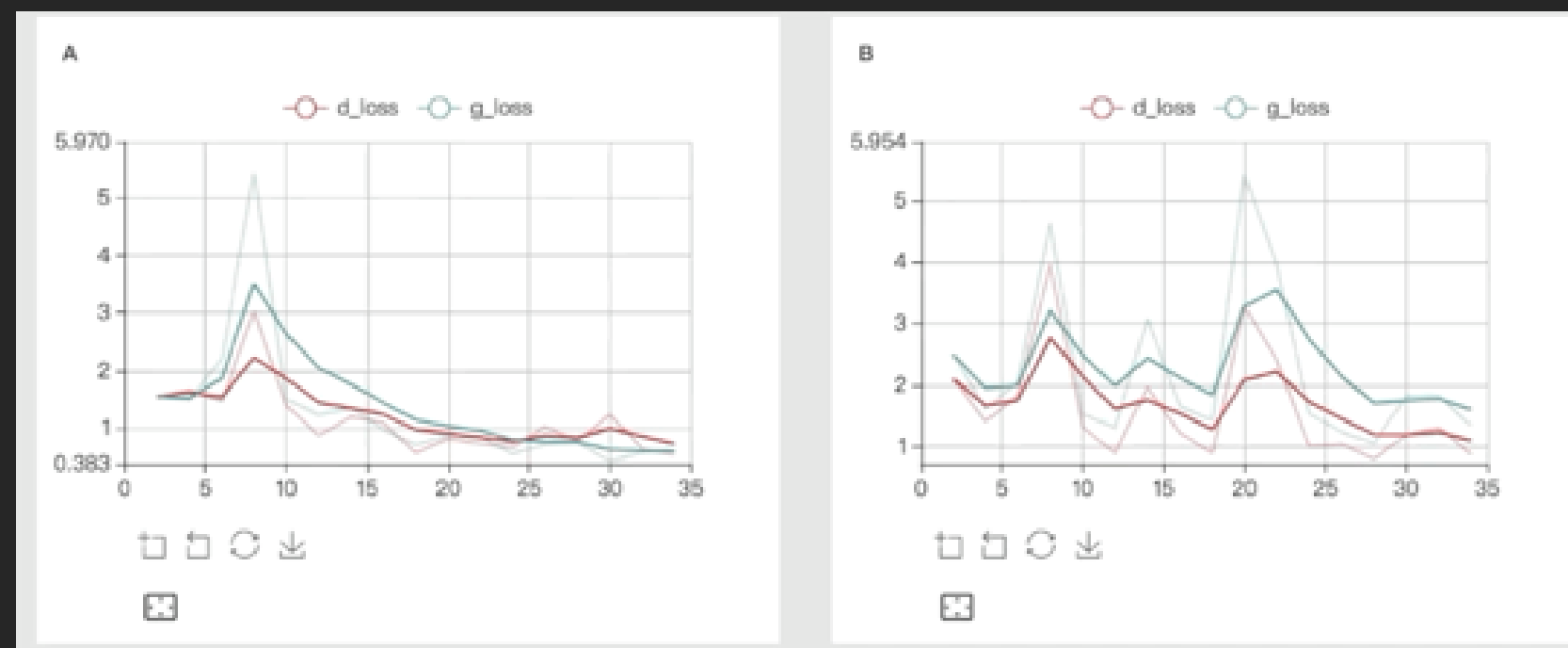
Histogram
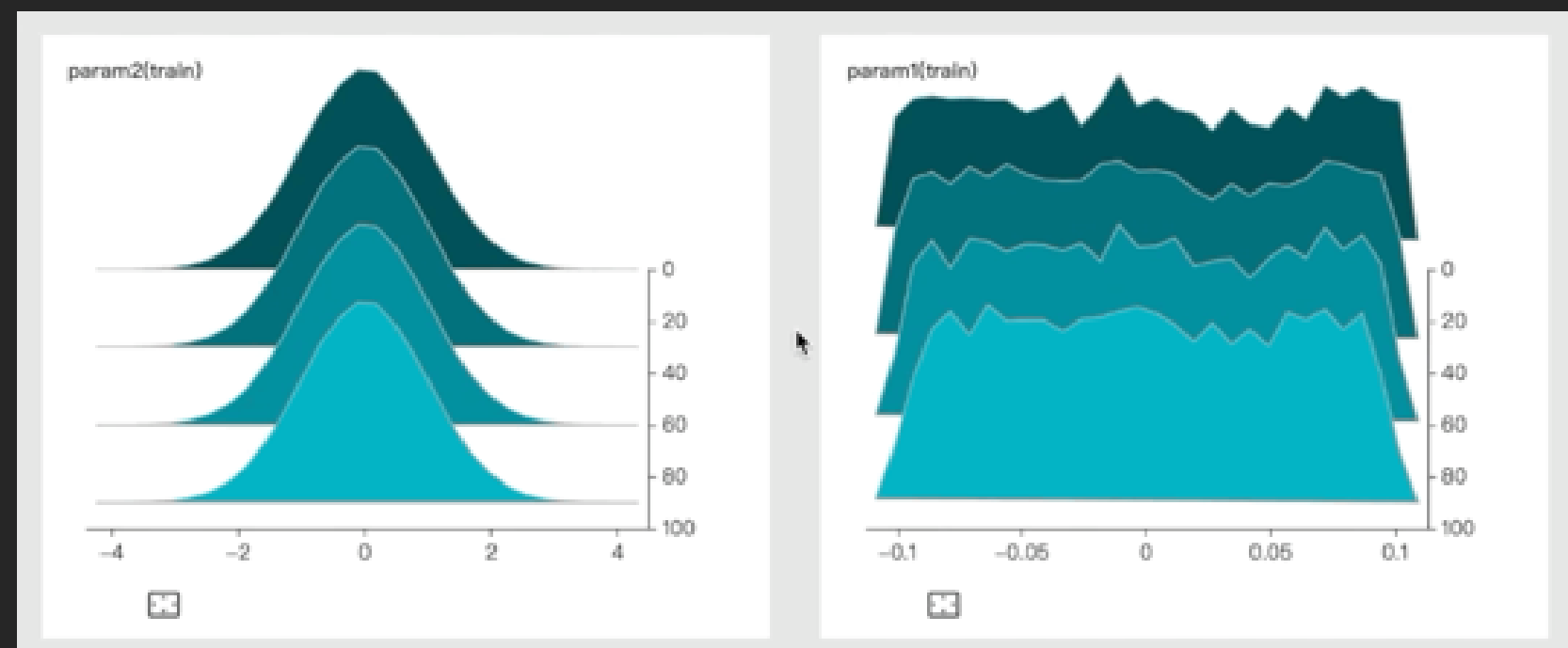
ONNX network graph
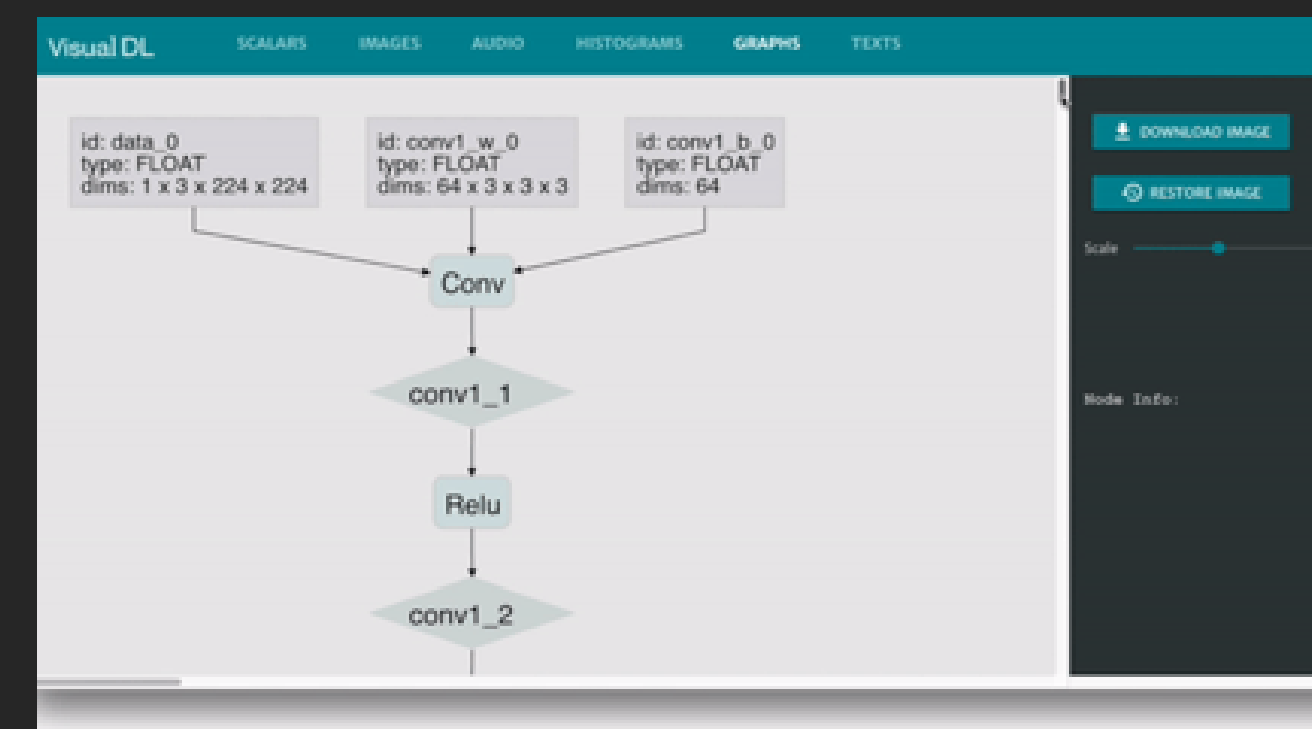
# AutoDL

## Support the Design, Transfer and Adaption of DL

| | |
|---|---|
| **AutoDL Design**<br><br>Network design | Search for several neural networks with excellent performance and different structures |
| **AutoDL Transfer**<br>Create transfer model with small amount of data | Transfer pretrained models to new applications |
| **AutoDL Edge**<br>Adapt to edge computing | Network complexity optimization based on classic model, suits better for mobile deployment |

# AutoDL Design
## Better than manual design

network structure search based on deep reinforcement learning



The network designed by AutoDL has the precision of 98% on CIFAR10
image classification dataset
Surpassing classic network designed manually

# AutoDL Edge
## Adapt for DL Edge Computing

## Network Optimization for DL

- Based on classic DL models
- Optimization on network complexity, suitable for mobile devices

## Optimization Results

- Remain accuracy
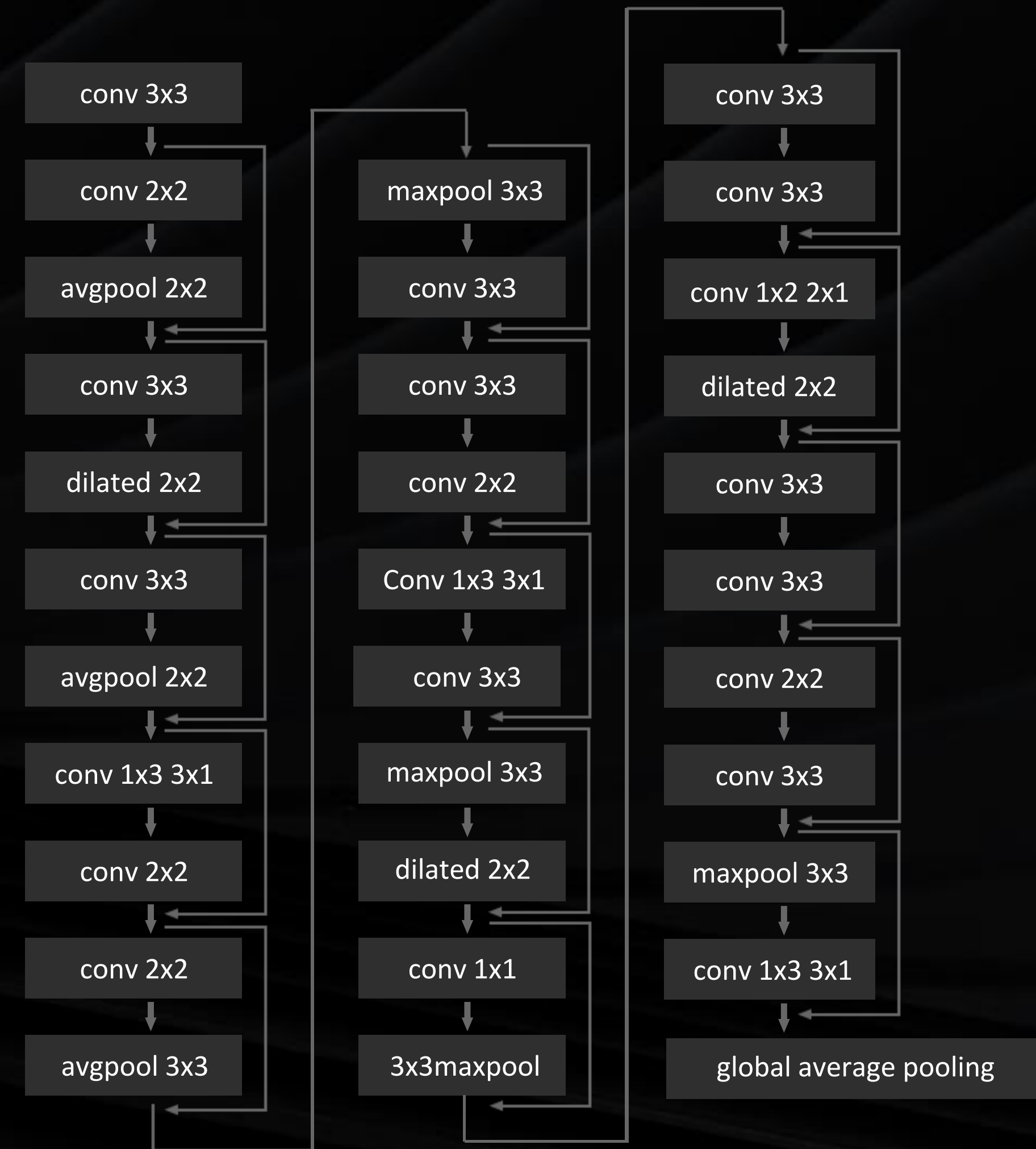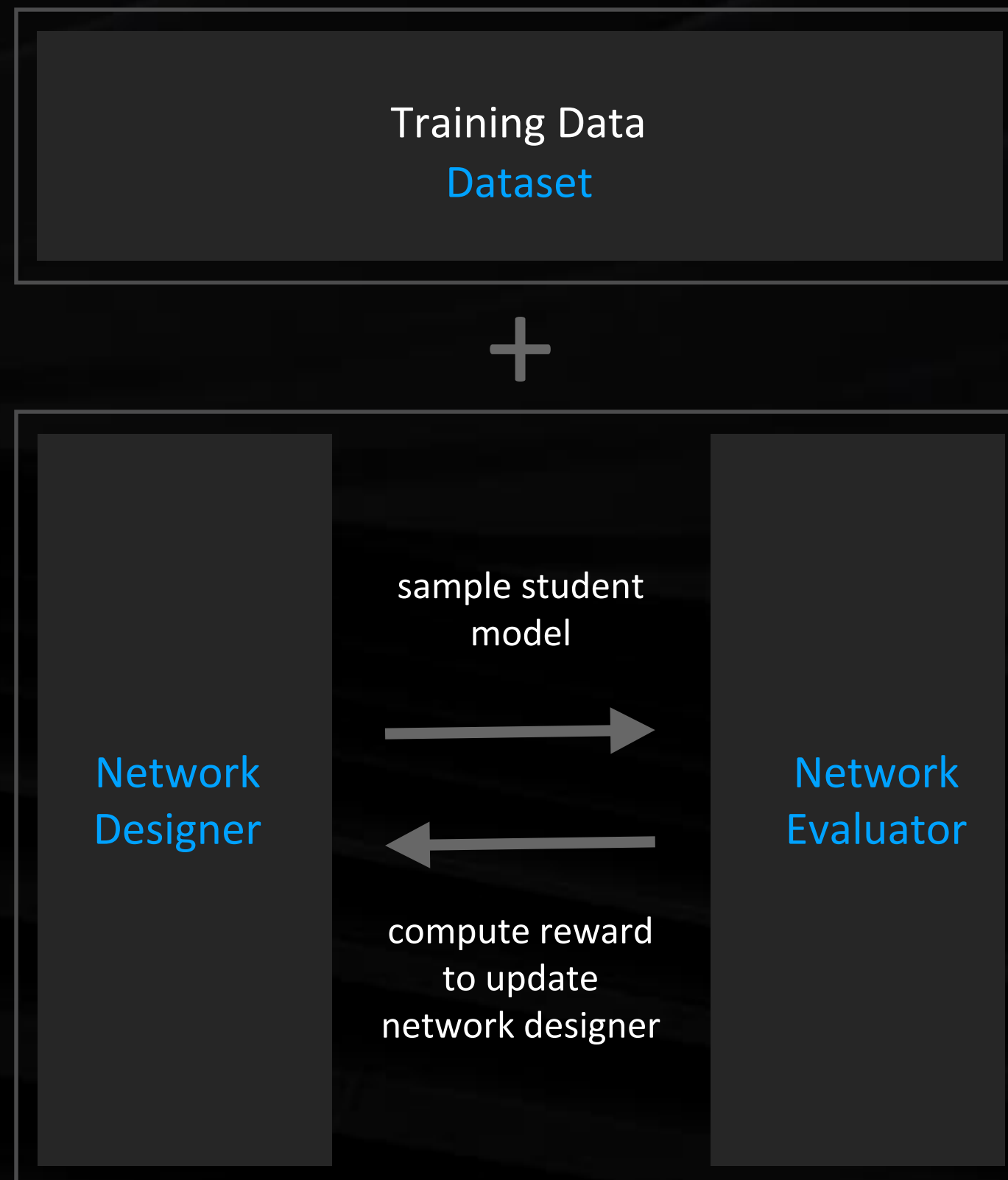- Highly compressed model parameters
- Run more AI tasks within the same computational capibility

| | | Before suppressing | | After suppressing | | Suppression ratio |
|---|---|---|---|---|---|---|
| | | Parameter amount | Precision | Parameter amount | Precision | |
| SoundNet on ESC-50 | | 13.00M | 66.00% | 0.07M | 65.60% | 180 |
| ResNet on CIFAR-10 | ResNet-18 | 11.17M | 94.18% | 0.82M | 93.90% | 13.62 |
| | ResNet-34 | 21.28M | 94.72% | 1.69M | 94.29% | 12.59 |
| | ResNet-50 | 23.52M | 95.16% | 3.97M | 94.91% | 5.92 |
| DenseNet on CIFAR-10 | DenseNet-121 | 6.96M | 95.13% | 1.75M | 94.72% | 3.97 |
| VOC Object detection | | 26.29M | 77.51% | 20.94M | 77.21% | 1.26 |
| Goods identification for retailers | | 31.36M | 84.55% | 22.09M | 84.76% | 1.42 |

# AutoDL Transfer
## Efficiency Modeling with small dataset

### Network transfer

- Transfer pretrained networks to new applications
- Network design automation, less time consuming

### Works better than classic models

- Need less samples
- Improve original model's capibility

**AutoDL Transfer--Comparing with classic models**



Legend: ■ Baseline ■ AutoDLStatic model ■ AutoDL Dynamic model
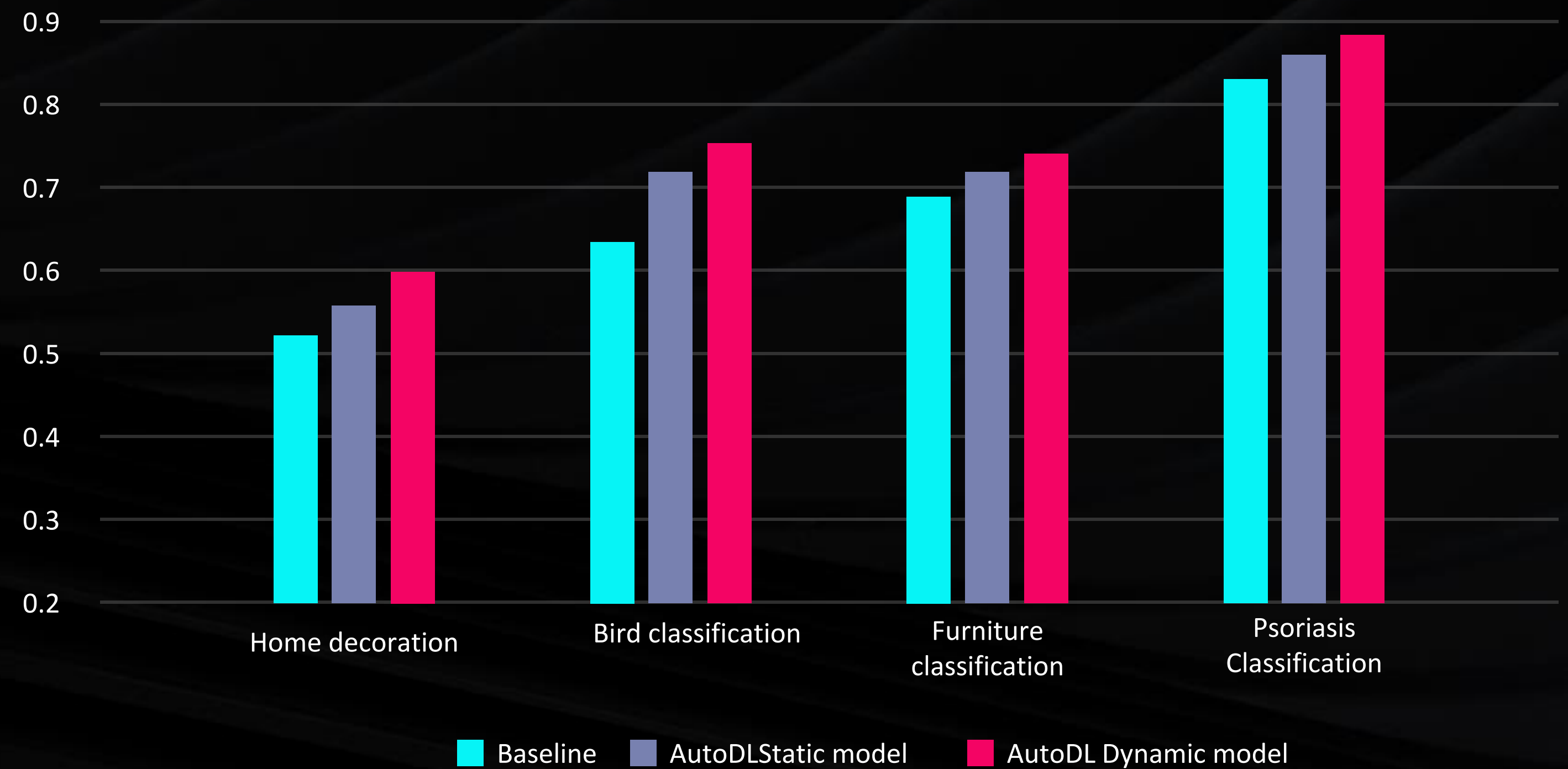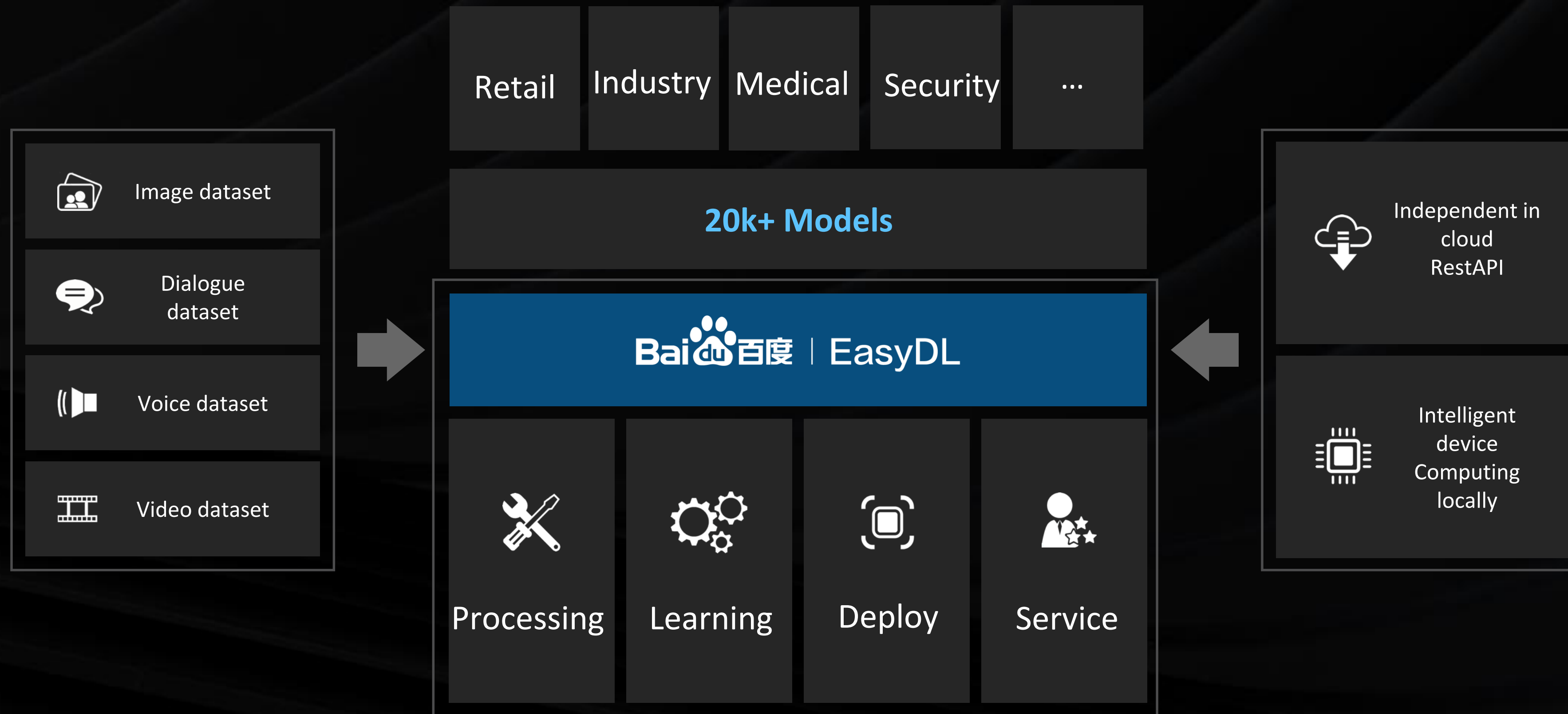
# AI Studio
## One-stop AI Developing Platform

| 20k+ Programs | 46k+ Developers | 3700+ Dataset |
|---|---|---|

## Online Training Platform

**Learning**

Systematic tutorials

Coding examples

Classic datasets

integration in the cloud

Easy to use

Efficient executions

Free resources

**Developing**

Python online coding

Predefined DL framework

Online training

## Large Scaled Open Dataset based on real world industry data

| Video segments | Recognition scenes for autonomous cars | Machine comprehension | Information extraction | Knowledge extraction | Traffic prediction | Object labeling |
|---|---|---|---|---|---|---|

# Agenda

A. **Brief Introduction**

B. **Ecosystem**

C. **Real World Use Cases**

D. **Developers Community**

# PaddlePaddle — Industry Application

## Empower AI ability for industry with our partners 100+ cooperative corporations

**Agriculture**

Intelligent peach sorting
Save 90% manpower

**Forestry**

Worms inspection
Detection accuracy 90%

**Industry**

Smoking
monitoring

**Retail**

Goods sales prediction
Decrease 30%
wastage for fresh items

**Human resource**

Matching system by AI
5 times successful interview
invitation

**Manufacturing**

Machine parts sorting
Double the efficiency

**Petroleum**

Prospection of
petroleum

**Telecommunication**

Base station monitoring

**Real estate**

Building management
Save 20% electricity

**Automobile**

Failure prediction of
charging stall
With accuracy of 90%

# Partners of PaddlePaddle

# Intelligent Sorting of Precision Parts

**PaddlePaddle cooperates with dominant domestic enterprises performing quality assurance for rare-earth permanent magnet, to push the landing of deep learning on manufacturing sector**

PaddlePaddle assists enterprises with the landing of projects in the entire procedure.

Analysis of Needs → Technical Model Selecting → Training (Optimization) → Hardware Preparation → Practical Testing
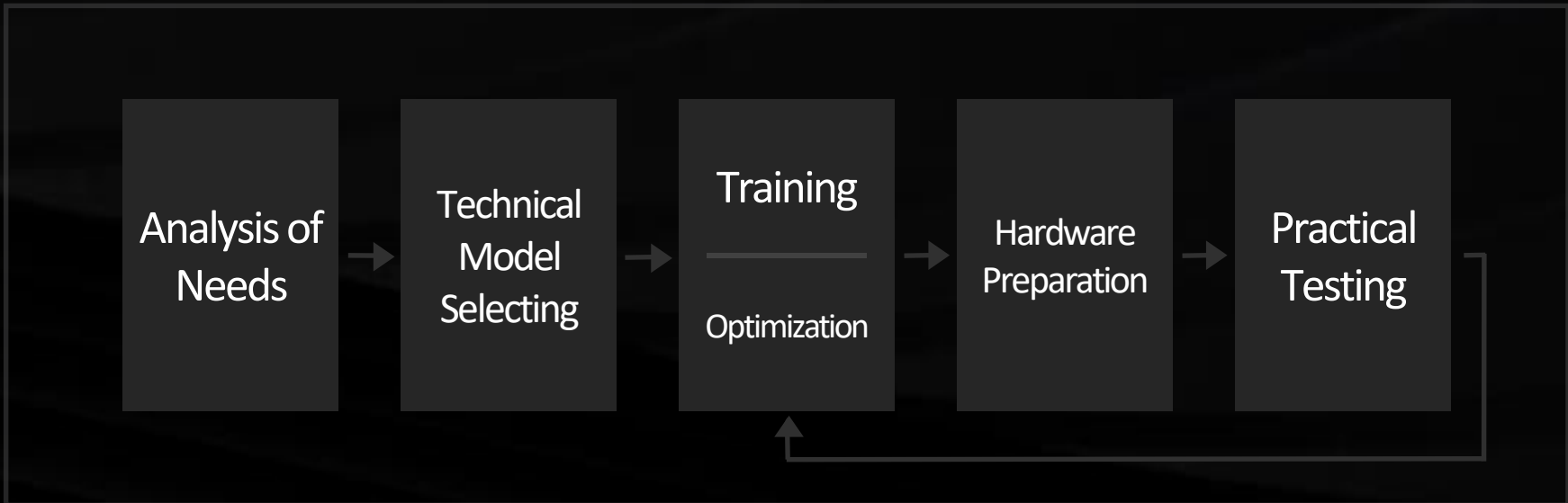
Custom Model Optimization, Predominant Effects in the field

| | |
|---|---|
| ICnet | Models of Semantic Segmentation |
| 0.1% | Rate of mistaken sort （Under a mistake recognition rate at 5%） |
| 25ms | The inference speed of single part exceeds other deep learning frameworks at a rate of 20%. |

# Monitoring System for the Red Turpentine Beetles
# (AI insect Recognition）

PaddlePaddle Cooperates with Beijing Forestry University on "Intelligent Insect Monitoring Project"

PaddlePaddle + Baidu Map+ Experts collaborated on this project

Data Collection → Model Training / Model Optimization → Model Preparation / Capturing devices → Offline Recognition / Baidu Map → Distribution of insect population

Custom Model Optimization,
Predominant Result in the field

| SSD | Models for Semantic Segmentation |
|---|---|
| 90% | The accuracy can reach 90%, which is similar to professionals |
| 1 week VS 1 hour | Enhance the efficiency greatly from the manual assessment time of a week or so |

# Intelligent Candidate Matching System

## CloudBrain adopts PaddlePaddle to invent an "AI HR"

Takes full advantage of PaddlePaddle NLP capibility in Chinese

| | | | |
|---|---|---|---|
| Textual Data | Training | Interview Invitation Open positin Clicks | Result statistics collection |
| Behavior Data | Optimization | | |

Significantly increase the rate of successful interview invitation for enterprises

| | |
|---|---|
| DSSM | Deep Structured Semantic Models |
| 5倍 | The increase in successful interview invitation rate |
| 50% | The increase of click-through rate to the recommended posts |

企业HR通过主动搜索、下载简历和发布JD等待投递后进行人工筛选和面试邀约

企业HR发布JD后，智能招聘服务会给企业HR主动推荐合适的候选人

过

现

传统招聘方法

AI-EDU 智能招聘服务

求职者通过搜索JD和更新简历进行投递简历

求职者更新简历后，智能招聘服务会给求职者更多的企业邀约机会

去

在

时间成本和人工成本高且效率低下

使成本更低，面试邀约率更高

# Monitoring System for Floor Quality



Raw Material

→ DL →

Tagging Platform

→ Tagged Data →

EasyDL platform to perform model training

↓ Model Exporting

↑ SDK integration

Mechanical Control, Channel Transmission

200m²/day  400m²/day

2x single-worker processing amount

← Transmit Signals ←

Business processing & summaries and analyses

合格

Qualified

← Flaw judgment and transmit data ←

Intelligent Camera
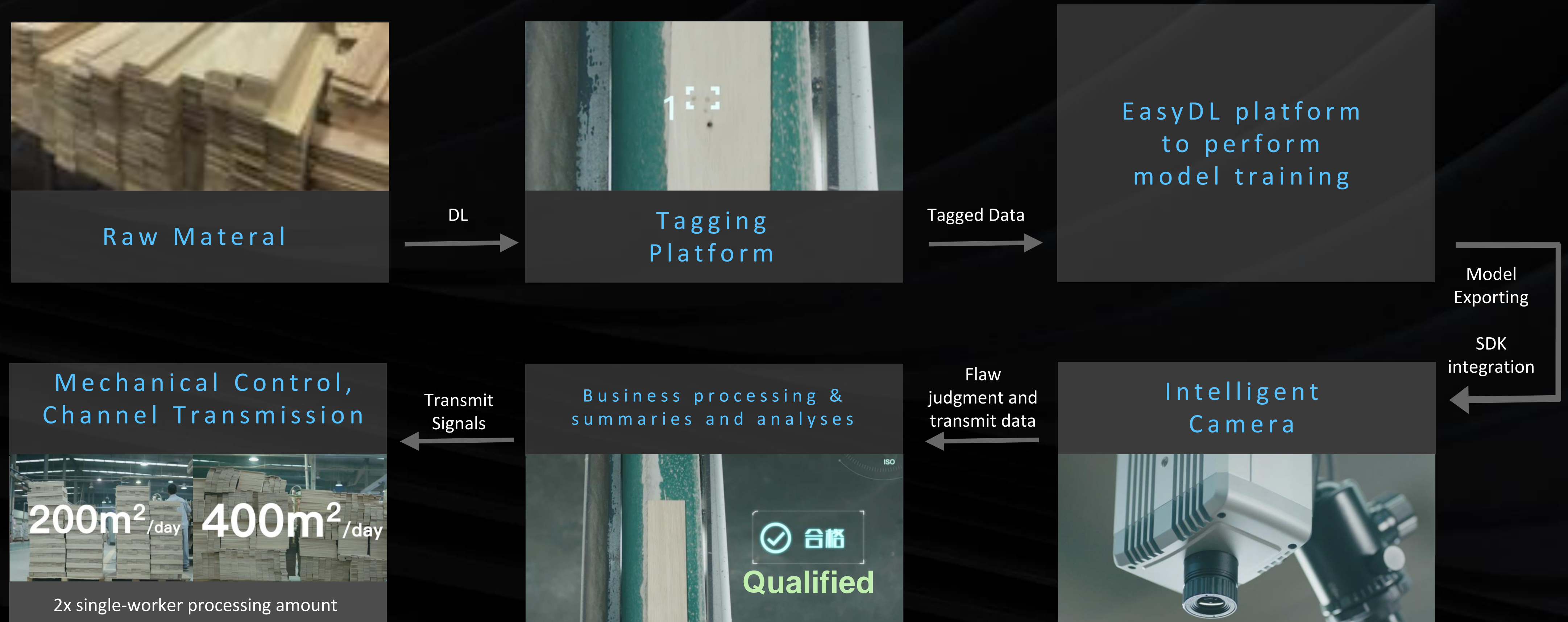
# Advantages of PaddlePaddle
# in Enterprise Empowerment

## Leading technology

Officially supported ICNET model

The precision can reach 99.5%

The inference speed is 20% higher than similar products

## Reliability

The only Chinese Deep Learning Framework

Performances with stability and reliability, thanks to the internal business lines of Baidu

## More understandings for domestic enterprises

Official Technical Support responses within 24h

Official Chinese Community and documentation

Follows AI project all the way through

Published 『HuangPu Plan』 Chinese AI talent training program

Across the globe, there have been many enterprises adopting PaddlePaddle and EZDL
30% of Chinese Enterprises have already remarked PaddlePaddle as one of Top3
deep learning frameworks.

# Agenda

A. **Brief Introduction**

B. **Ecosystem**

C. **Real World Use Cases**
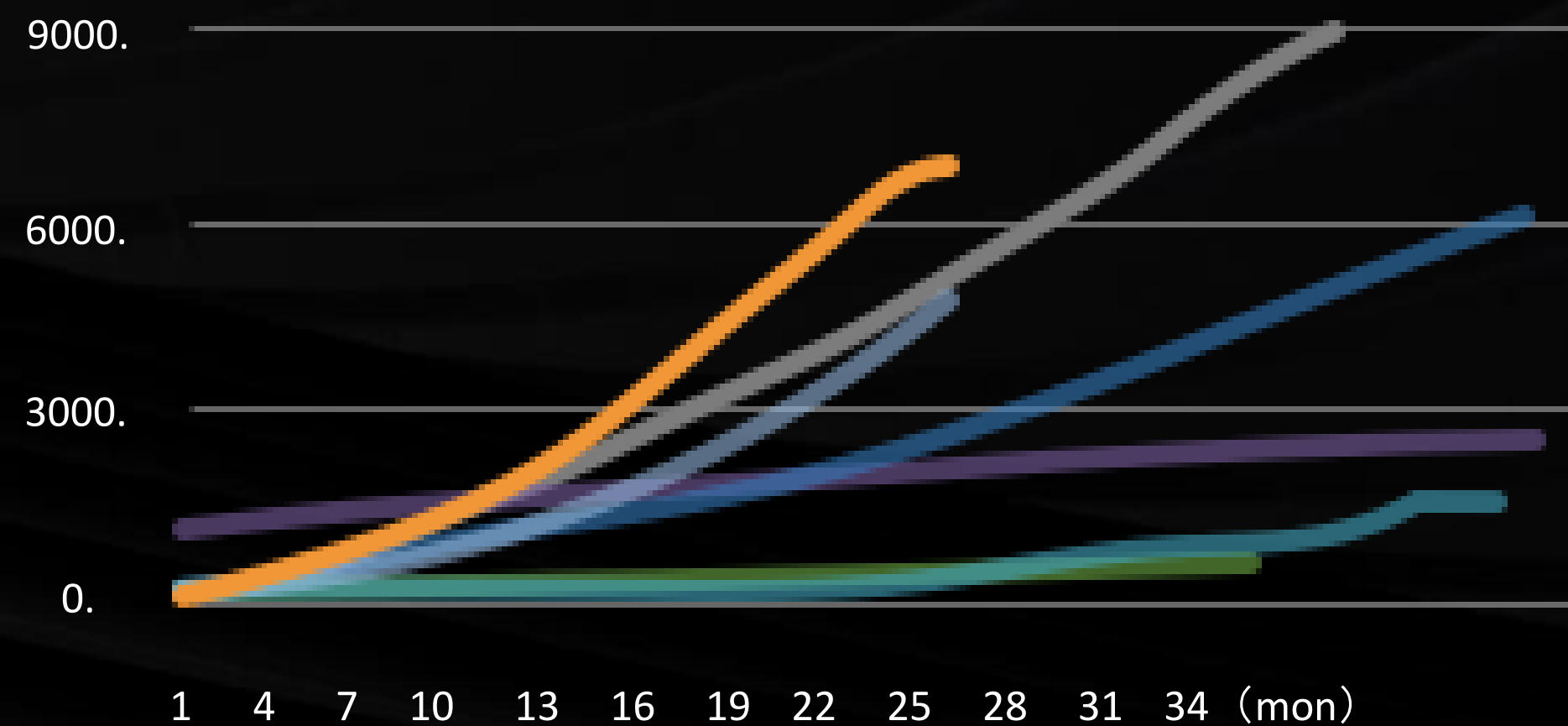
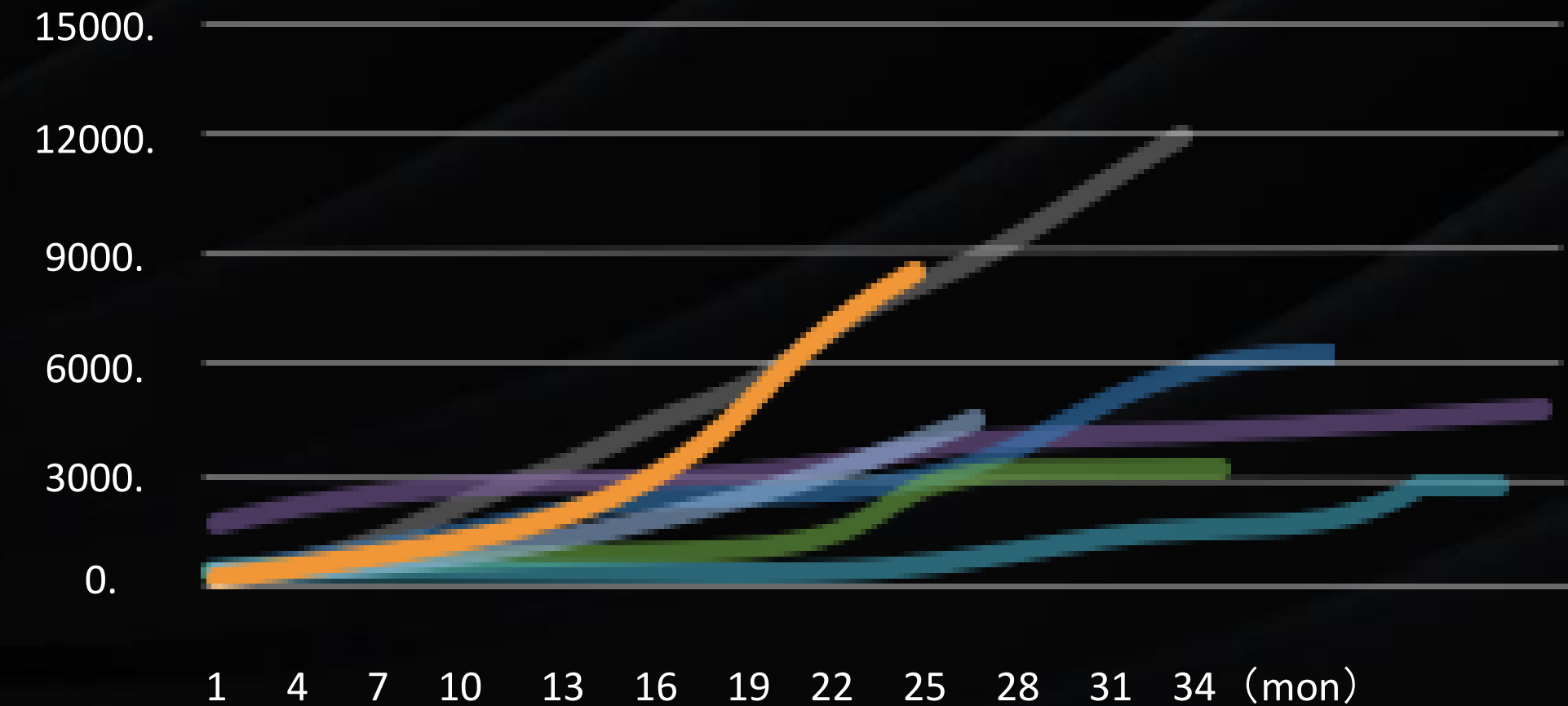D. **Developer Community**

# Active Developer Ecosystem

PaddlePaddle has a relatively high vitality at GitHub open-source community, even 530k+ Downloads and
higher than other frameworks in the same period                                            Counting



# Pull requests

# Issues

● PaddlePaddle    ● Tensorflow    ● MxNet    ● Caffe    ● Caffe2    ● CNTK    ● Pytorch

# PaddlePaddle Education



出版物    公开课    比赛    考试认证    师资培训    布道师沙龙    7X24小时服务

数据开放 BROAD    算力开放 AI Studio    框架开放 算法开源    评价标准开放 认证体系    教研开放 课程研究专家组    师资开放 布道师/讲师认证

PaddlePaddle
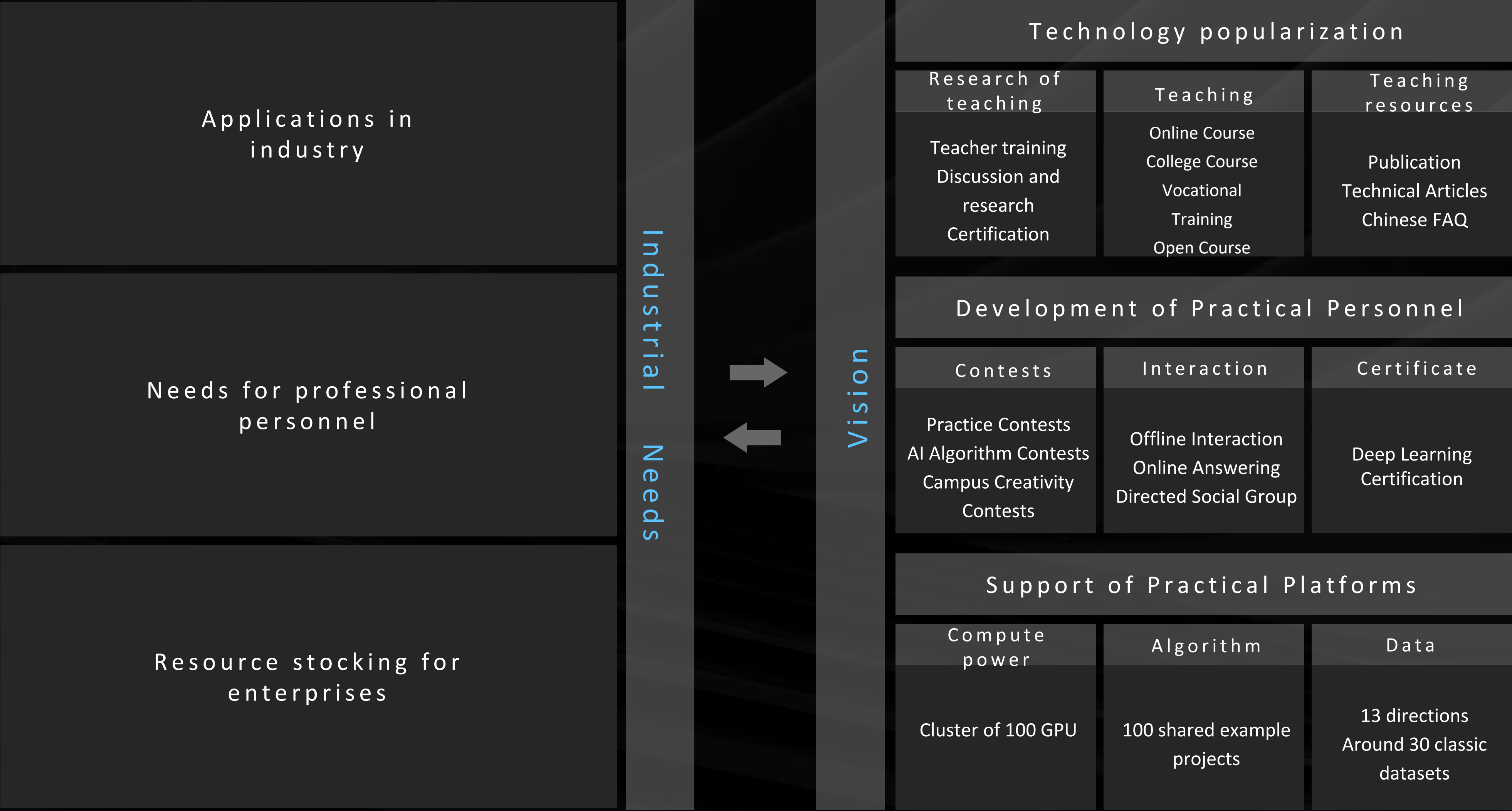
10k+ of active AI studio PaddlePaddle users

Published "Certification Standard for DL Engineering" with China software association

3 training courses for 300 university teachers from 100s of schools

Publications of books and training videos

# PaddlePaddle Education Ecosystem

Applications in industry

Needs for professional personnel

Resource stocking for enterprises

Industrial Needs

Vision

## Technology popularization

| Research of teaching | Teaching | Teaching resources |
| --- | --- | --- |
| Teacher training Discussion and research Certification | Online Course College Course Vocational Training Open Course | Publication Technical Articles Chinese FAQ |

## Development of Practical Personnel

| Contests | Interaction | Certificate |
| --- | --- | --- |
| Practice Contests AI Algorithm Contests Campus Creativity Contests | Offline Interaction Online Answering Directed Social Group | Deep Learning Certification |

## Support of Practical Platforms

| Compute power | Algorithm | Data |
| --- | --- | --- |
| Cluster of 100 GPU | 100 shared example projects | 13 directions Around 30 classic datasets |

# Whampoa College - Training the First Batch of Chief AI Architects for Chinese Industry

Baidu Established the "Whampoa College" with
National Engineering and Applications Laboratory of Deep Learning

| | | | |
|---|---|---|---|
| Face-to-face communication with Baidu Deep Learning T10 Architects | Unlock the key point of implementing DL in Baidu's core business know how | Analysis of the typical case of the combination of business and deep learning in Baidu AI Cooperators in Ecosystem | Help companies use AI thinking, AI tools, and methodologies to solve real business problems |

**20+DL Experts Waiting to Sail Together**

郭国栋
百度深度学习实验室副主任

浣 军
百度大数据实验室主任

马艳军
百度深度学习技术平台部总监

杨睿刚
百度机器人与自动驾驶实验室主任

于佃海
百度PaddlePaddle总架构师

Kenneth Church
Baidu Research Fellow

**[ Launch ]**
**Way Of Deep**
1.19

**[ Second ]**
**CV Fierce**
3.9-3.10

**[Third ]**
**NLP Leap**
4.20-4.21

**[Fourth]**
**Enhance Together**
6.1-6.2
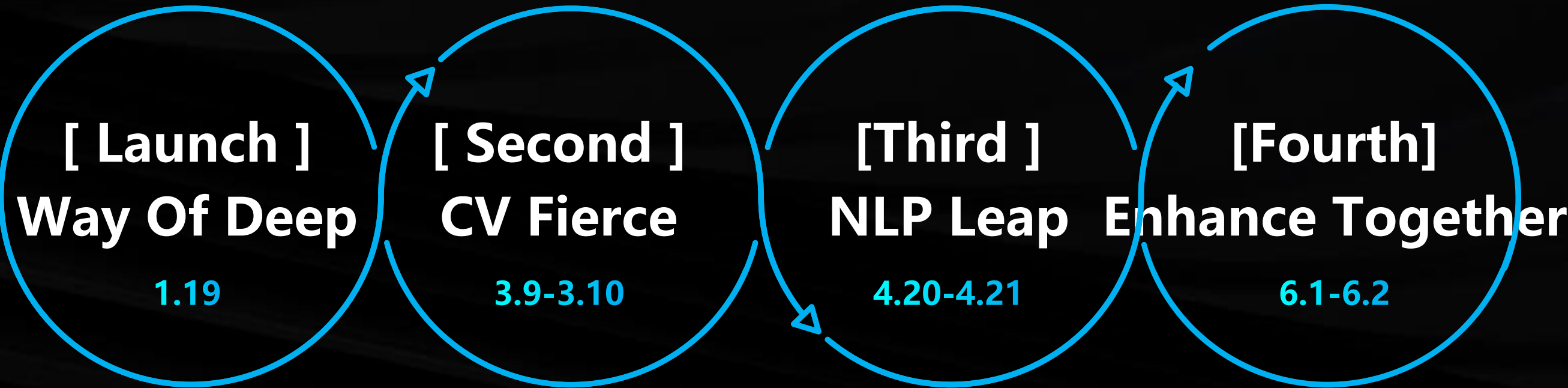
**Hard Core Technology**

Experimental Course throughout the entire process

# Cooperations with Cloud Platform

**Significant AI brand influence power,   and a sharing-oriented attitude towards market resource**

➤ PaddlePaddle is about to bring profit boosts to cloud services through market share expansion

➤ PaddlePaddle is devoted to developing a framework in line with needs of cloud providers

➤ PaddlePaddle is obliged to share cloud-end solutions with partners

➤ PaddlePaddle is willing to share part of promotion resources with all our partners

## Share Customers
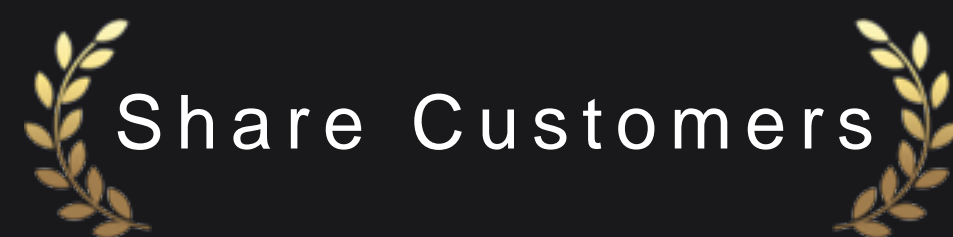
Partners will be listed as important cloud SP forging a great guidance for potential users

## Co-Branding

PaddlePaddle willing to share promotion resources with all our partners

## Technical Support

Expert assistance for deploying PaddlePaddle to cloud

# PaddlePaddle

## The Deep Learning Framework that Truly Stems From Industry Practice

http://paddlepaddle.org
https://github.com/PaddlePaddle