AI Vision (VisionBrain)：
The Deep Learning Development Platform for image/video analysis

Yonghua Lin
Leader of Cognitive System Research
IBM Research - China
Email: linyh@cn.ibm.com
Deep Learning for Computer Vision

- Applied to every corner of computer vision
- Feature extraction and representation through deep neural network
- Learning feature hierarchies
- More effective with deeper net and less parameters (e.g. CNN)
- Multimodal learning with mixed neural networks
2 Stages Deep Learning for Cognitive Solution Build

**Training (development) Stage**
- Train Data Set
- DNN Net File
- Data Preprocessing
- Feature Engineering
- Modeling
- Trained model
- Model pool

**Inference (deployment) Stage**
- Application Data from User
- Recognition, classification
- Application
- Deep Learning platform
- Application servers, DB service, messaging, etc.

**CPU + GPU cluster**

**CPU + GPU/FPGA cluster**

Data Management
Big data platform (Hadoop, Spark)
Deep Learning platform (caffe, Torch, Theano, TensorFlow, etc.)
Steps for Deep Learning Development

• Usually, developers need following steps to develop a DNN model and make it usable for application

1. Define training task
2. Prepare training data
3. Data preprocessing
4. DNN model selection
5. Configure the training hyper-parameter
6. DL training framework preparation
7. DNN model training
8. Package the new DNN model together with preprocessing into inference proc.

Most of enterprises are facing the challenges ...

• No experience on DNN design and develop
• No experience on computer vision
• No experience on how to build a platform to support enterprise scale deep learning, including data preparation, training, and inference
**AI Vision** enables enterprise level DNN easier

- **AI Vision** automates the deep learning development cycles for developers.
- Deep knowledges of ML/DL and computer vision have been embedded into **AI Vision**.

On-line training by looping the inference and training capability together

Steps automatically done by AI Vision

- User defined categories
- Data set management
- Format transformation
  - Support both training and evaluation sets
  - Support different preprocessing plugin
- Provide base models for different scenarios
  - Predict training time
- Training process visualization
  - Training with GPU
- Scalability and HA deployment are supported

User could use the deployed API for visual recognition

Define training task
Prepare training Data
Data Pre-processing
DNN Model selection
DL training framework preparation
Configure the training hyper-parameter
DNN Model Training
Package the new DNN model together with preprocessing into inference proc., and deploy API

Start
Inference
Image Classification example with AI Vision

I’m Aethopyga

I’m Pycnonotus

Result on public cloud API: white, red, yellow and teal bird

Result on public cloud API: white and black short beak bird

We need to get a new model to classify birds with professional knowledge.

User defines categories in AI Vision

Acridotheres

Acrocephalus

Aethopyga

Butorides

Corvus

... >20 categories

Aethopyga: 0.90708

Pycnonotus: 0.99988

I'm Aethopyga

I'm Pycnonotus
Object Detection Example with AI Vision

Want to develop an application to automatically monitor the safety in field

Define the regulation requirement through data labeling

“Click” to finish all the steps, and get the API for detection.
AI Vision: The Deep Learning Development Platform for image/video analysis

Vision Recognition Service Layer

- Image Labeling and Preprocessing
- Video Labeling Service
- Custom Learning for Image Classification
- Custom Learning for Object Detection
- Self-defined Training with graphic/visual monitoring
- Inference API deployment

Service Management Layer

- Image preprocessing management
- Data label management
- Data set management
- Training task management
- Model management
- Inference API management

ML/DL Computation Layer

- Data preprocessing component
- Model Training Component (Caffe, and others)
- Inference Component (Caffe, and others)

Resource management layer (CPU/GPU/FPGA) (Docker, Kubernetes)

- Docker, KVM (POWER)
- Accelerator (GPU/FPGA)
- Data Store (distributed FS and object store)
- Network

DL-Insight: Monitoring and Optimization tool for Deep Learning
AI Vision supports deep learning development

1. User will define the categories and upload data set for new model training

2. Start the model training

3. During the training, AI Vision will show the performance data

4. After training, AI Vision will deploy the model as API. User could also do the test with web page.
High Accuracy and High Efficiency with Fine Tuning and Fine Grained Technologies

- Trained base model on large amount of data
- Selective Scenarios to ensure the similarity of data
- Hyperparameter adjustment based on input data
- Fine Grained Classification

Classify Cadillac Cars on AI Vision

<table>
<thead>
<tr>
<th>w/o BaseModel</th>
<th>BaseModel+FineTuning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.32812</td>
</tr>
<tr>
<td></td>
<td>0.85417</td>
</tr>
</tbody>
</table>

Classification Results for Stanford Cars

<table>
<thead>
<tr>
<th>BaseModel+FineTuning</th>
<th>Fine Grained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.86375</td>
</tr>
<tr>
<td></td>
<td>0.886</td>
</tr>
</tbody>
</table>

Share GPU resource for inference processing

Use GPU to accelerate Image Object Detection (15000 pics/hour ~ 50000 pics/hour):

<table>
<thead>
<tr>
<th>DL training</th>
<th>Pre-process.</th>
<th>DL training</th>
<th>Pre-process.</th>
<th>DL Inference</th>
<th>DL Inference</th>
<th>Pre-process</th>
<th>Pre-process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kubernetes Slave</td>
<td>Compute (docker)</td>
<td>Distributed file system</td>
<td>KVM/Bare Metal CPU + GPUs</td>
<td>Kubernetes Slave</td>
<td>Compute (docker)</td>
<td>Distributed file system</td>
<td>KVM/Bare Metal CPU + GPUs</td>
</tr>
</tbody>
</table>

AI Vision will schedule the inference instances on different computing nodes, and let them share the GPU physical resource.

<table>
<thead>
<tr>
<th>GPU memory used</th>
<th>% of K80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Detection with VGG16</td>
<td>~ 1.7 GB</td>
</tr>
<tr>
<td>Object Detection with ZF</td>
<td>~ 1 GB</td>
</tr>
</tbody>
</table>
Deep Learning Platform: Container Cloud with GPU

- Based on Docker Container Tech.
- Support >10K concurrent instances
- Dynamic resource scheduling
- GPU optimized scheduling
- GPU Resource sharing
- GPU accelerator
- Scale-out with heterogeneous nodes
- Scalability
- HA and load balance

Support Deployment on both Public Cloud and Private Cloud
Developer Workflow with AI Vision

**Self-Training for new model**

**VisionBrain - Training**

- Self-training process (with simple visualization)
- Data preprocess methods
- Base models (image/video)
- NN for classification/object detection
- DL training run-time environment on server cluster with GPUs

**VisionBrain - Inference**

- Inferencing engine with API
- Data preprocess methods
- NN for classification/object detection
- DL Inferencing run-time environment on server cluster with GPUs

- Deploy executable model for production
- Trained model
- On-line retrain for error detection result
- Optimized NN
- Performance data during training
- Optimized NN
- Data collection and Visualization

**Data set upload & labeling**

**Application or solution developer**

**Training data set and label management**

**Self-training process**

**Base models** (image/video)

**Data preprocess methods**

**Trained model**

**Performance data during training**
**AI Vision** provides strong AI capability based on PowerAI

**AI Vision**: Deep learning development tool and run-time environment for image/video recognition

- **ML & DL Libraries & Frameworks**
- **Distributed Computing**
- **Data Lake & Data Stores**
- **Accelerated Servers**
- **Storage**

**IBM PowerAI**: Enterprise Deep Learning Distribution

- Optimized deep learning frameworks
- Enterprise class support
- Optimized for Performance To Take Advantage of NVLink
**AI Vision:** Help Enterprise Enable AI Capability

**Industry app. Developer:**
- No deep learning and vision analysis skill required
- Help them to easily enable cognitive capability in industry application

**Image/video analysis specialist:**
- Help them easily inject the analysis tech. into a platform

**Deep Learning Specialist:**
- Can easily optimize the neural network, and apply different deep learning methodologies onto the platform