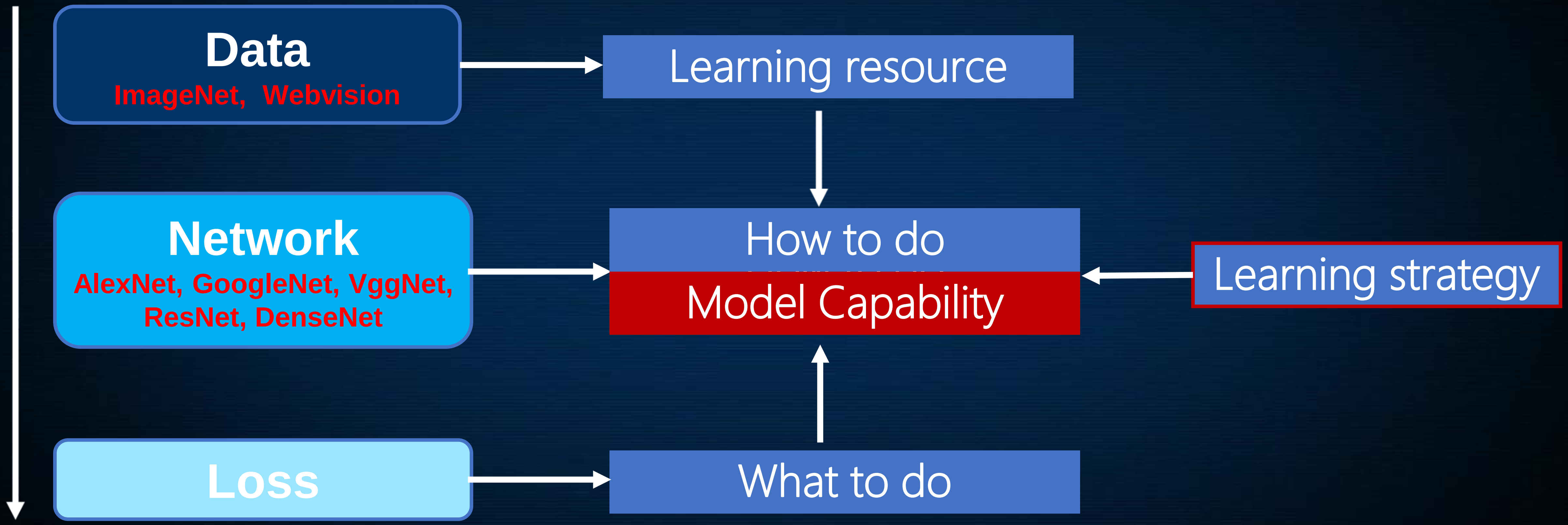
Learn CNNs from Large-scale Web Images without Human Annotations

Weilin Huang

Malong Technologies

How to Train a High-Performance CNN



webvision | MALONG

Motivations

- Performance on ImageNet has been saturation.

  From ~30% (2009) --> ~2.2% (2017)

- Train CNNs without human labelling --> weakly-supervised learning

- Develop new approaches working on large-scale data in real-world scenarios

- Data, model architecture, loss, training strategy are all important

- Train CNNs from web images are most common tasks in industries

Introduction: WebVision Workshop Organizers

General Chairs

J. Berent

A. Gupta

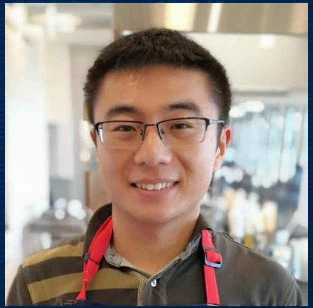
R. Sukthankar

L. Van Gool

Program Chairs

Wen Li

Limin Wang



Wei Li

E. Agustsson

CVL Computer Vision Lab | ETH zürich | Google | Carnegie Mellon University The Robotics Institute

Introduction: Database Construction

1,000 semantic concepts from ILSVRC 2012

Google

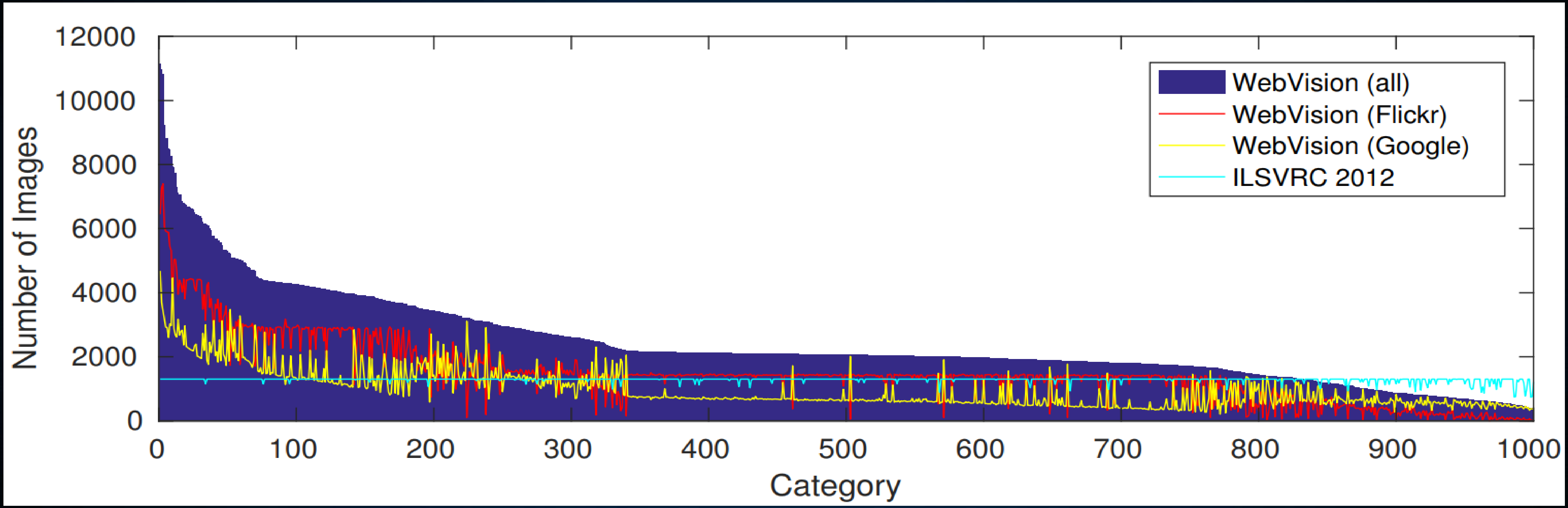
flickr

WebVision Dataset:

- 2 Sources
- 1,000 categories
- 2.4M training images
- 50K validation Images
- 50K test Images

Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, Luc Van Gool, *"WebVision Database: Visual Learning and Understanding from Web Data"*.arXiv: 1708.02862, 2017.

Main Challenge: Data Imbalance

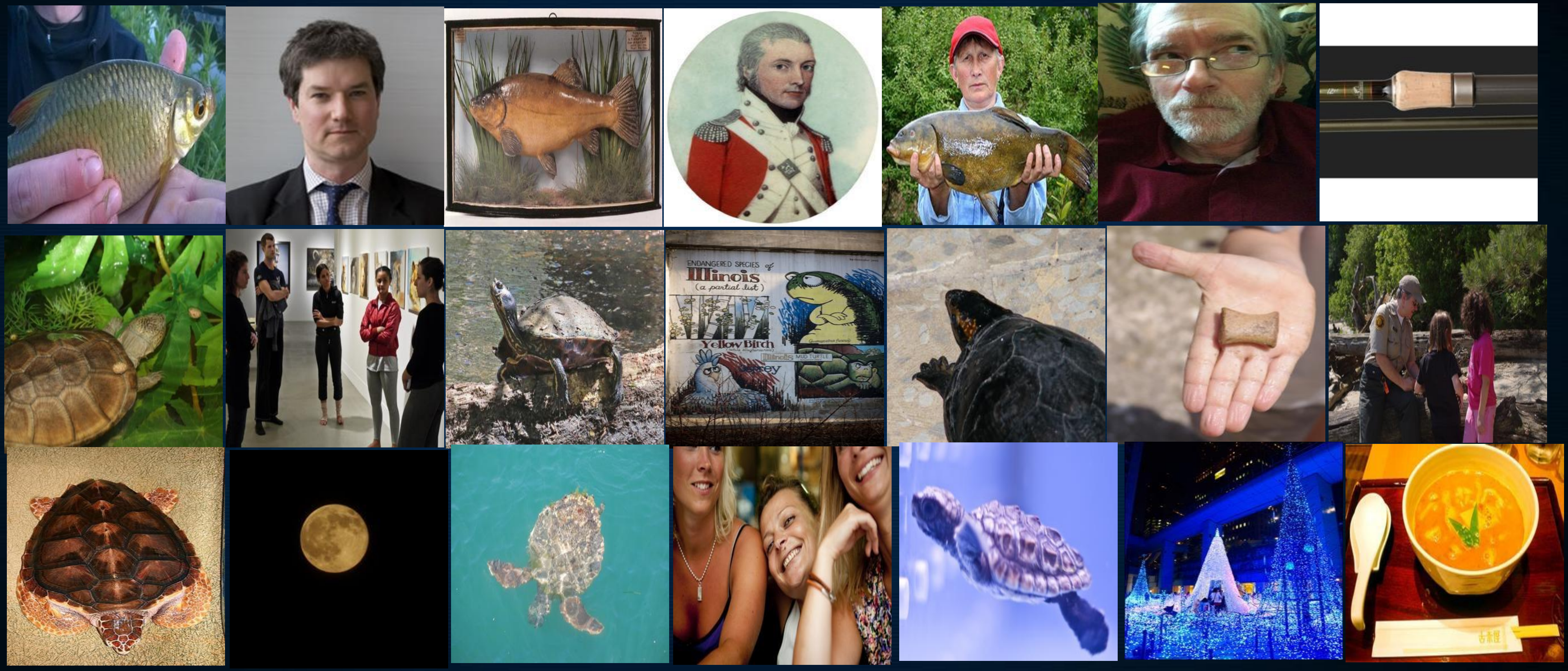


Main Challenge: Label Noise

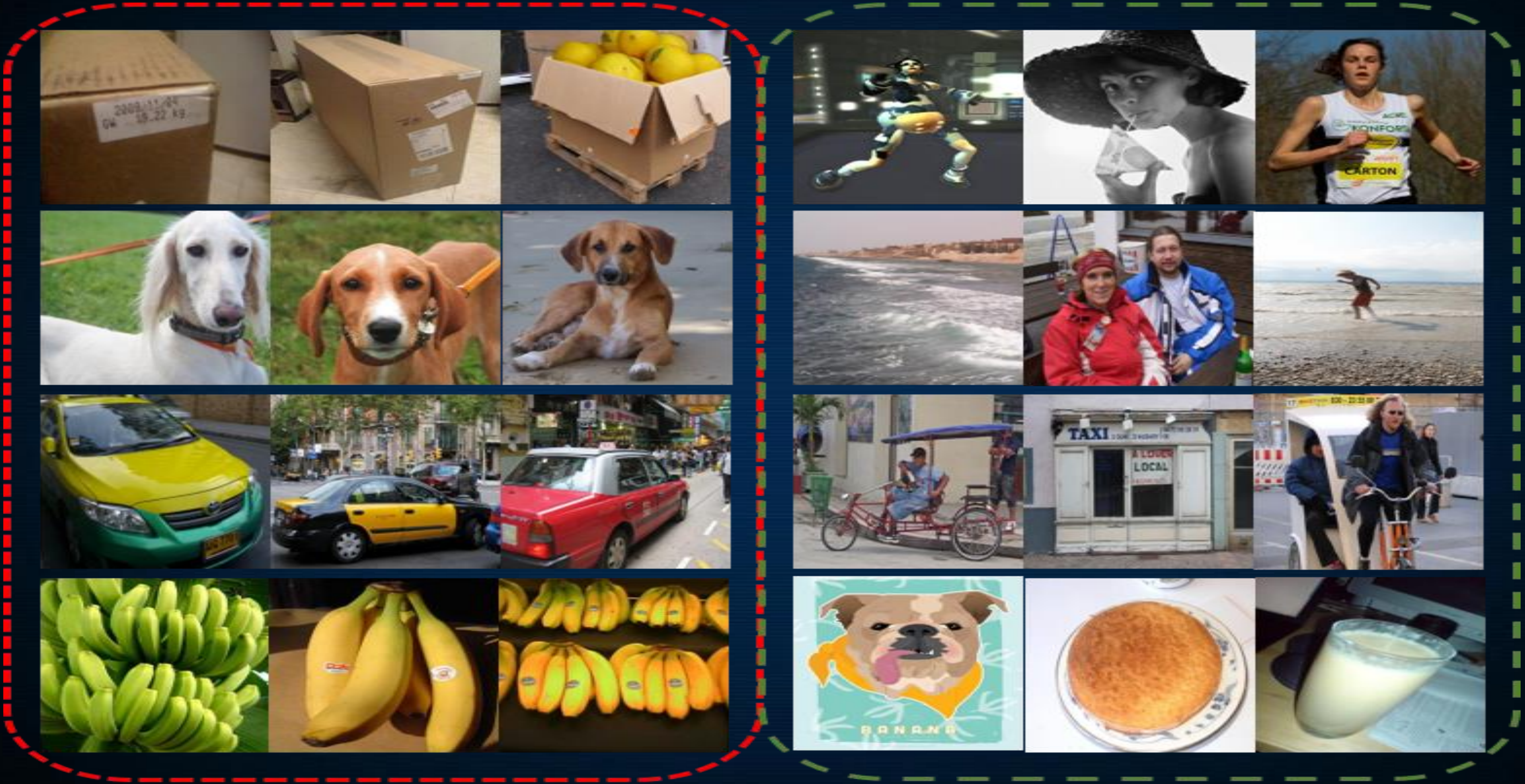
Tench

Terrapin

Caretta



Main Challenge: Label Noise



Clean images

Noise Images

webvision | MALONG

Related Works

- Directly learn from noisy labels

  1. Noise-robust Algorithms

  2. Label-cleansing methods

  --> difficult to identify mislabeled samples from hard training samples

- Semi-Supervised methods

  -->Need a small set of manually-labeled

- Recent deep learning approaches developed for both groups of methods

Improve model capability of standard neural networks by introducing new training strategies.

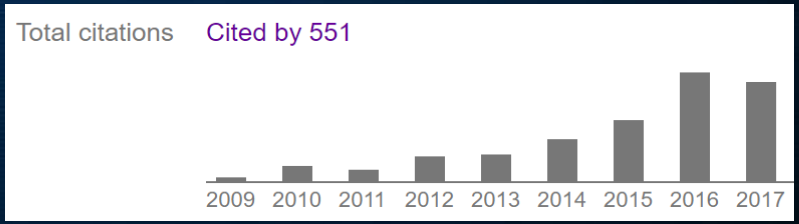
Methodology: Curriculum Learning

Curriculum learning

—Train CNNs on tasks with increasing difficulty

—Train CNNs using samples with increasing complexity

"Humans and animals learn much better when the examples are not randomly presented but organized in a meaningful order which illustrates gradually more concepts, and gradually more complex ones."



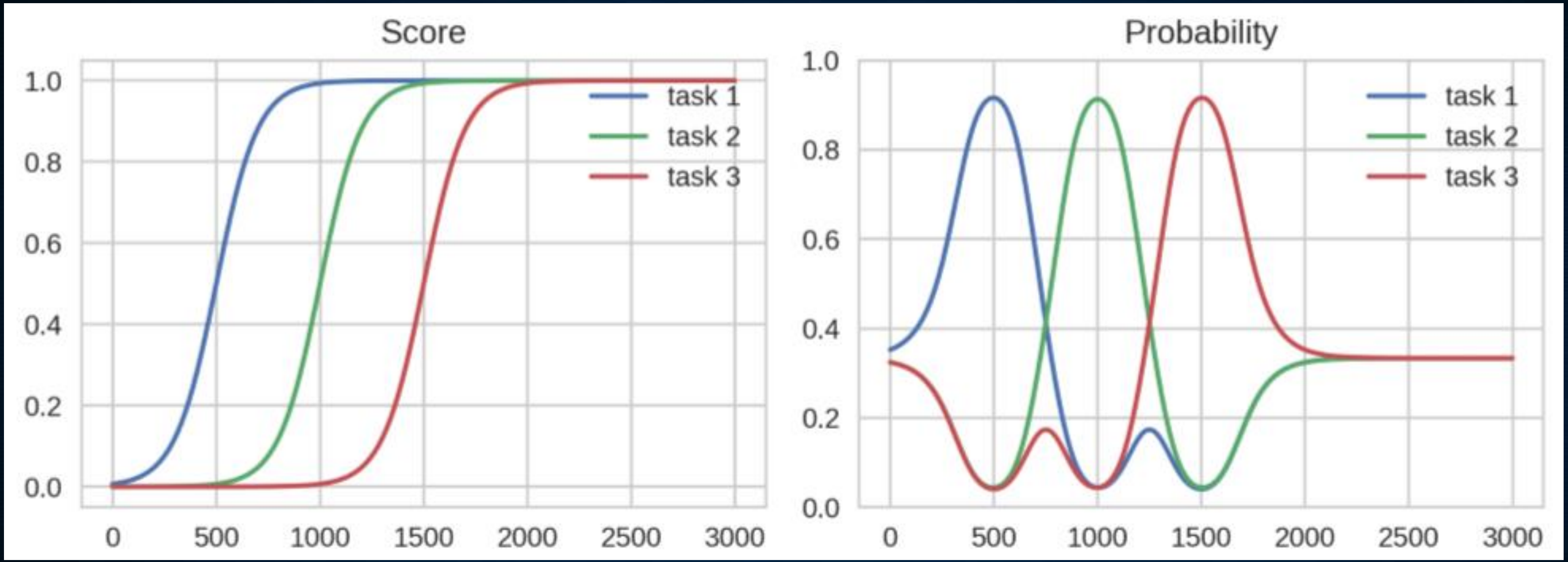
Y. Bengio, J. Louradour, R. Collobert, and J. Weston, Curriculum Learning, ICML, 2009.

Methodology: Curriculum Learning Processing

Steps:

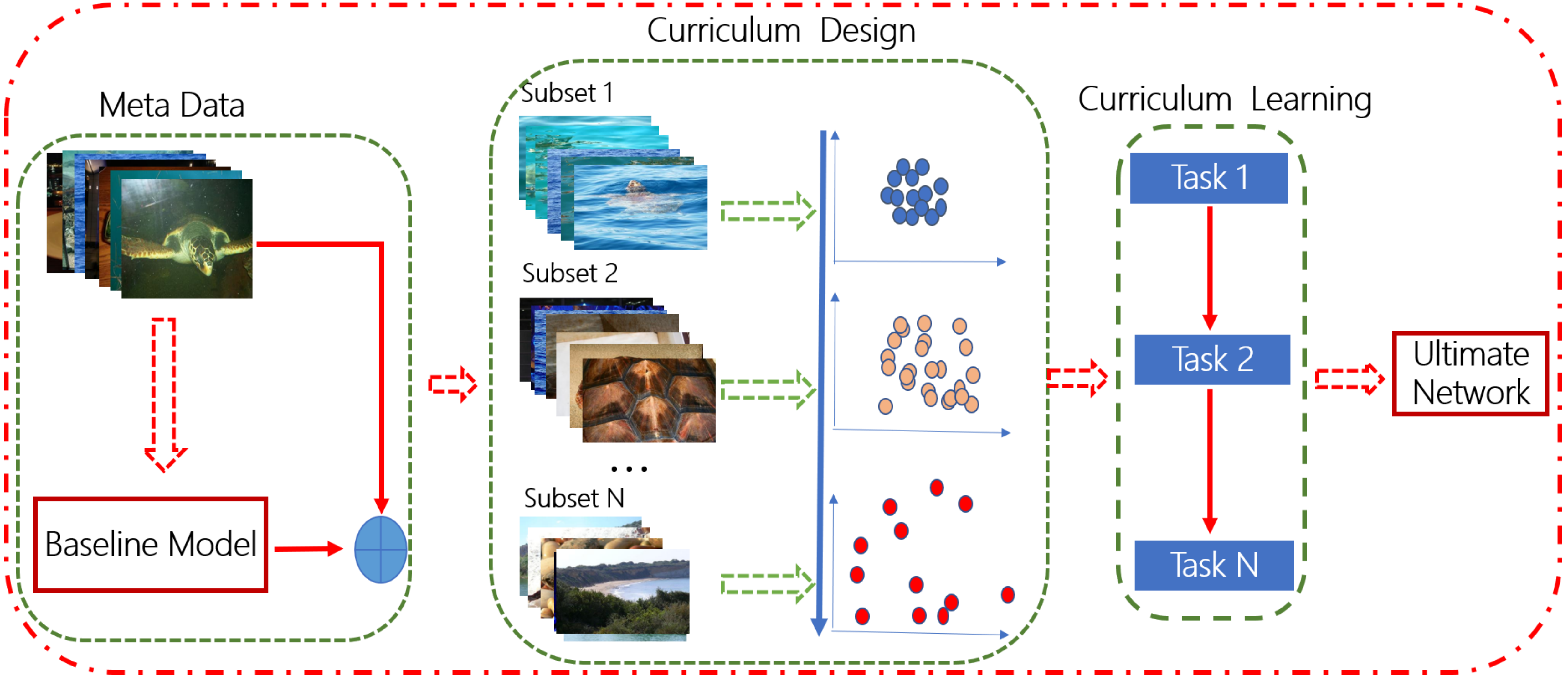
- Split a learning problem into a number of subtasks
- Order subtasks by difficulty
- Decide a task-transform threshold
- Find an optimized path that leads to fast convergence and better generalization
- Simple principle: proceed harder tasks once easier ones are handled

Methodology: Idealistic Curriculum Learning Processing



T. Matiisen, A. Oliver, T. Cohen, and J. Schulman, Teacher-Student Curriculum Learning, arXiv:1707.00183, 2017.

Methodology: Formulate our problem



Methodology: Curriculum Design

- **Split the whole training set into multiple subsets**
- **Rank subsets with increasing complexity**
- **Density-Distance clustering *in each category***

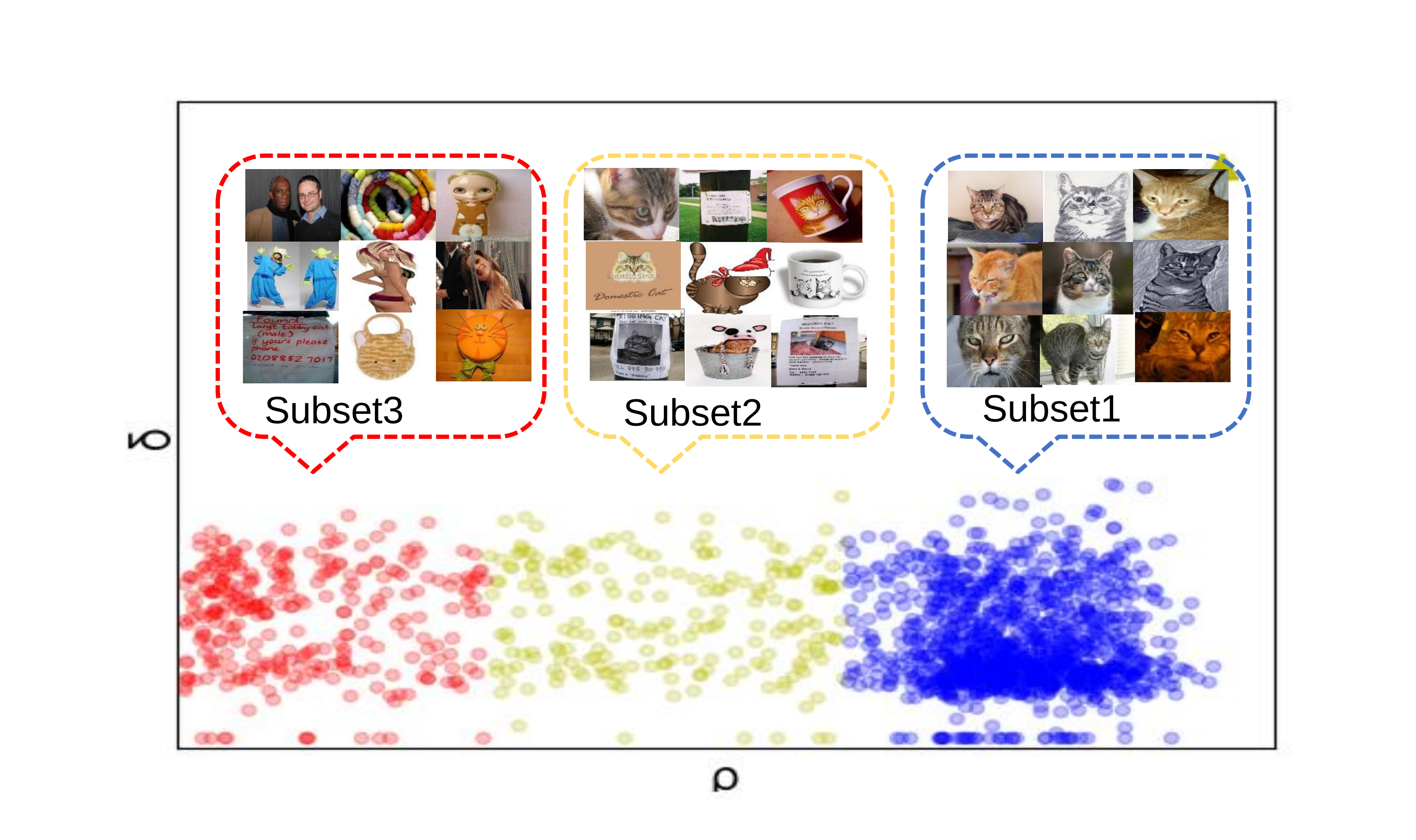
Step One: **Similarity Matrix** : $P_i \rightarrow f(P_i) \quad D_{ij} = \|f(P_i) - f(P_j)\|^2$

Step Two: **Sample Density** : $\rho_i = \sum_j X(D_{ij} - d_c) \quad X(d) = \begin{cases} 1 & d < 0 \\ 0 & \text{other} \end{cases}$

Step Three: **Sample Distance** : $\delta_i = \begin{cases} min_{j:\rho_j > \rho_i}(D_{ij}) & if \exists j s.t. \rho_j > \rho_i \\ max(D_{ij}) & \text{otherwise} \end{cases}$

(Rodriguez and Laio, ***Science***, 2014.)

Methodology: Curriculum Design



webvision | MALONG

Methodology: Curriculum Design

Subset 1

Tench

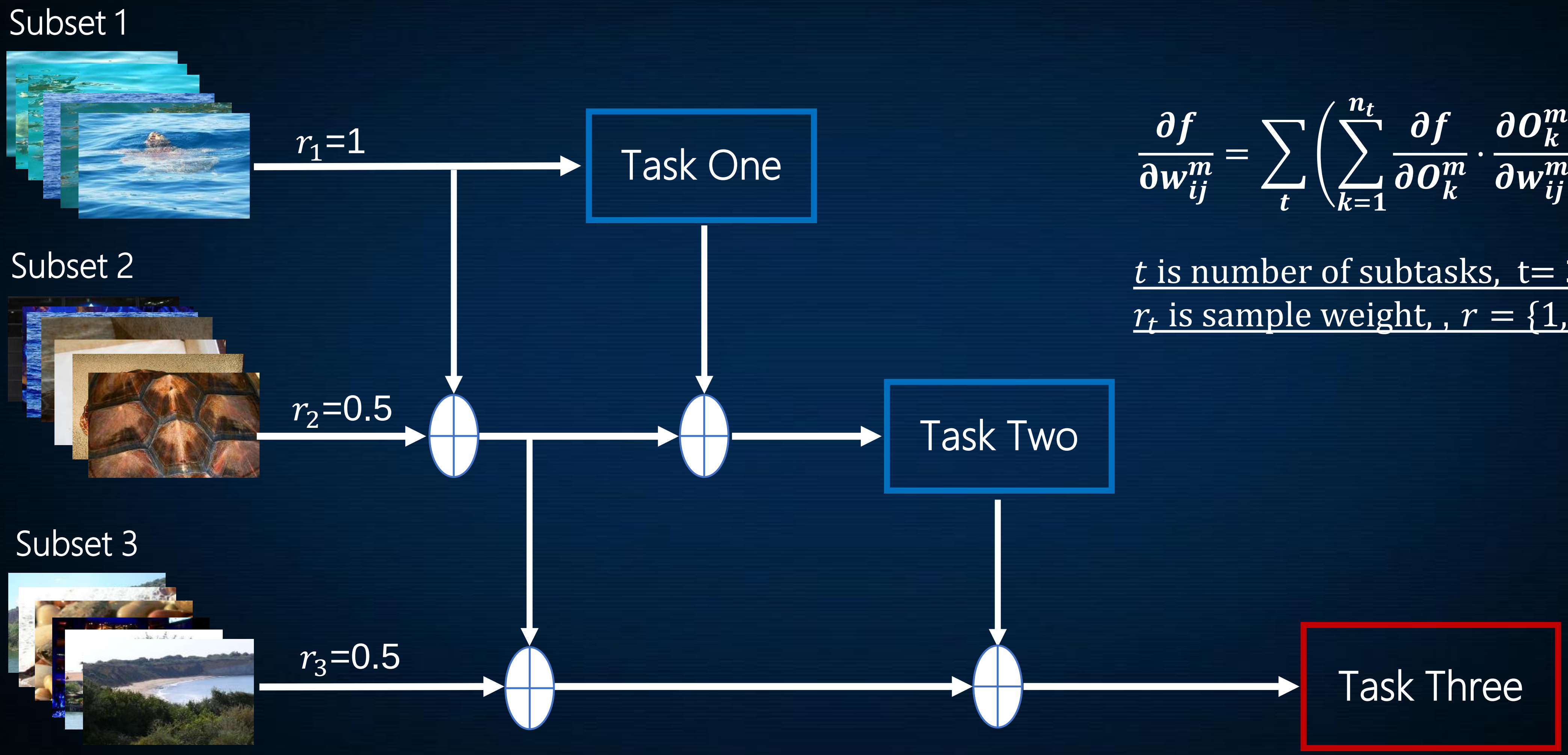
Terrapin

Subset N

Tench

Terrapin

Methodology: Train with Curriculum Learning

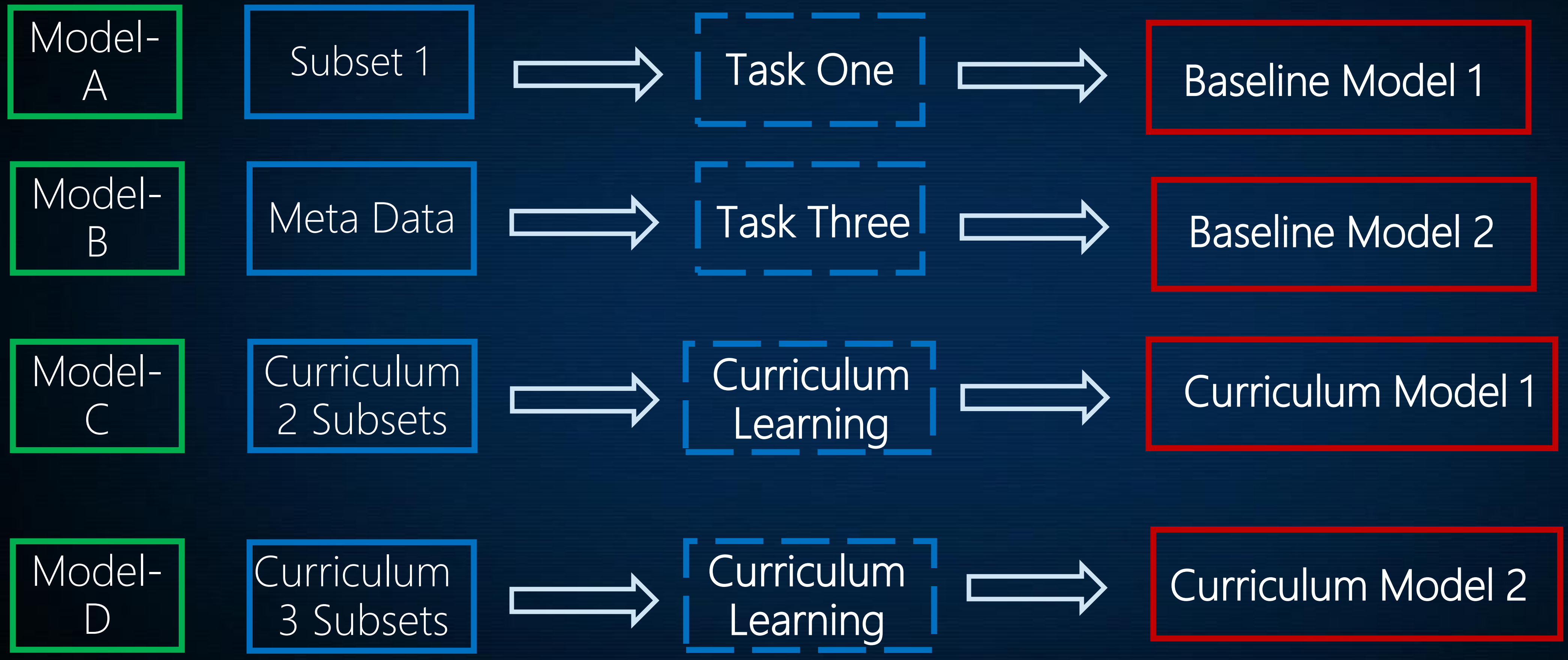


$$\frac{\partial f}{\partial w_{ij}^{m}} = \sum_{t} \left(\sum_{k=1}^{n_t} \frac{\partial f}{\partial O_k^m} \cdot \frac{\partial O_k^m}{\partial w_{ij}^m} \right) \cdot r_t$$

t is number of subtasks, t= 3

r_t is sample weight, , $r = \{1, 0.5, 0.5\}$

Methodology: Models with Different Training Schemes



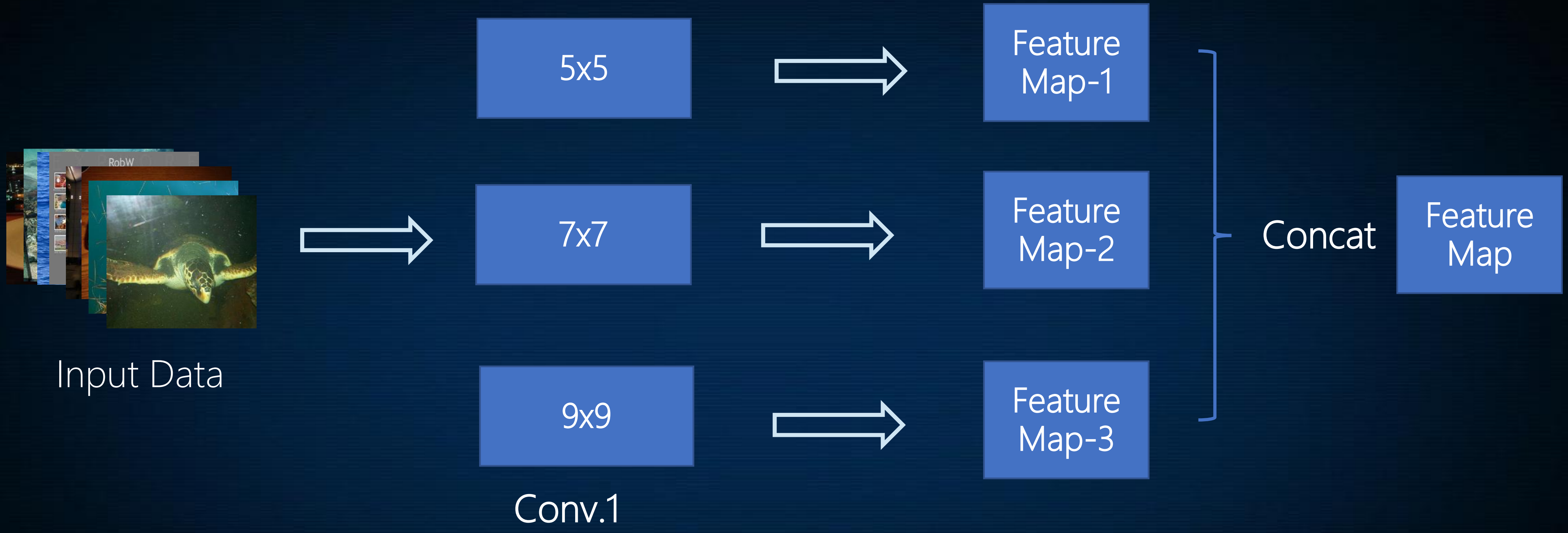
Methodology: Selective Data Balance

Curriculum Design = 3 subsets

Mini-batch = 256

- Samples balance among subsets (three subsets applied)
  [Subset_1 = 128, Subset_2 = 64, Subset_3 = 64]
- Classes balance only on Subset_1
  —> Randomly select 128 classes
  —> Each class only has one sample

Methodology: Multi-Scale Convolutional Kernel



Enhance low-level features which improve the performance (about 0.5%).

Result: Testing Loss

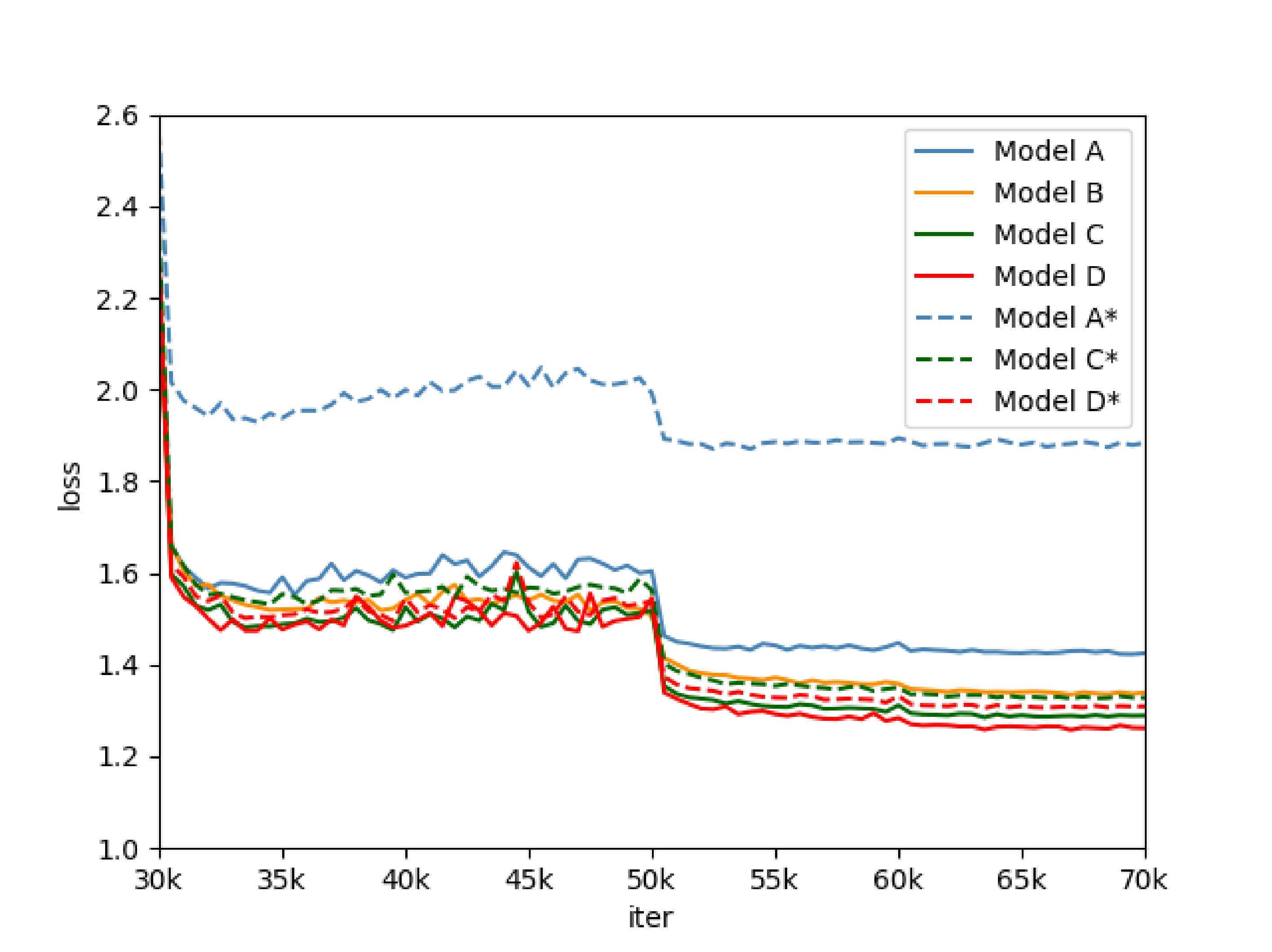


Figure1. Testing loss of four different models with Inception_v2 (also comparing to K-mean clustering in curriculum design)

Results: Single Model, 10 Crops

Model	Top1	Top5
Model-A	30.28%	12.98%
Model-B	30.16%	12.43%
Model-C	28.44%	11.38%
Model-D	**27.91%**	**10.82%**

Table 1. Different models based on Inception_v2 on validation set.

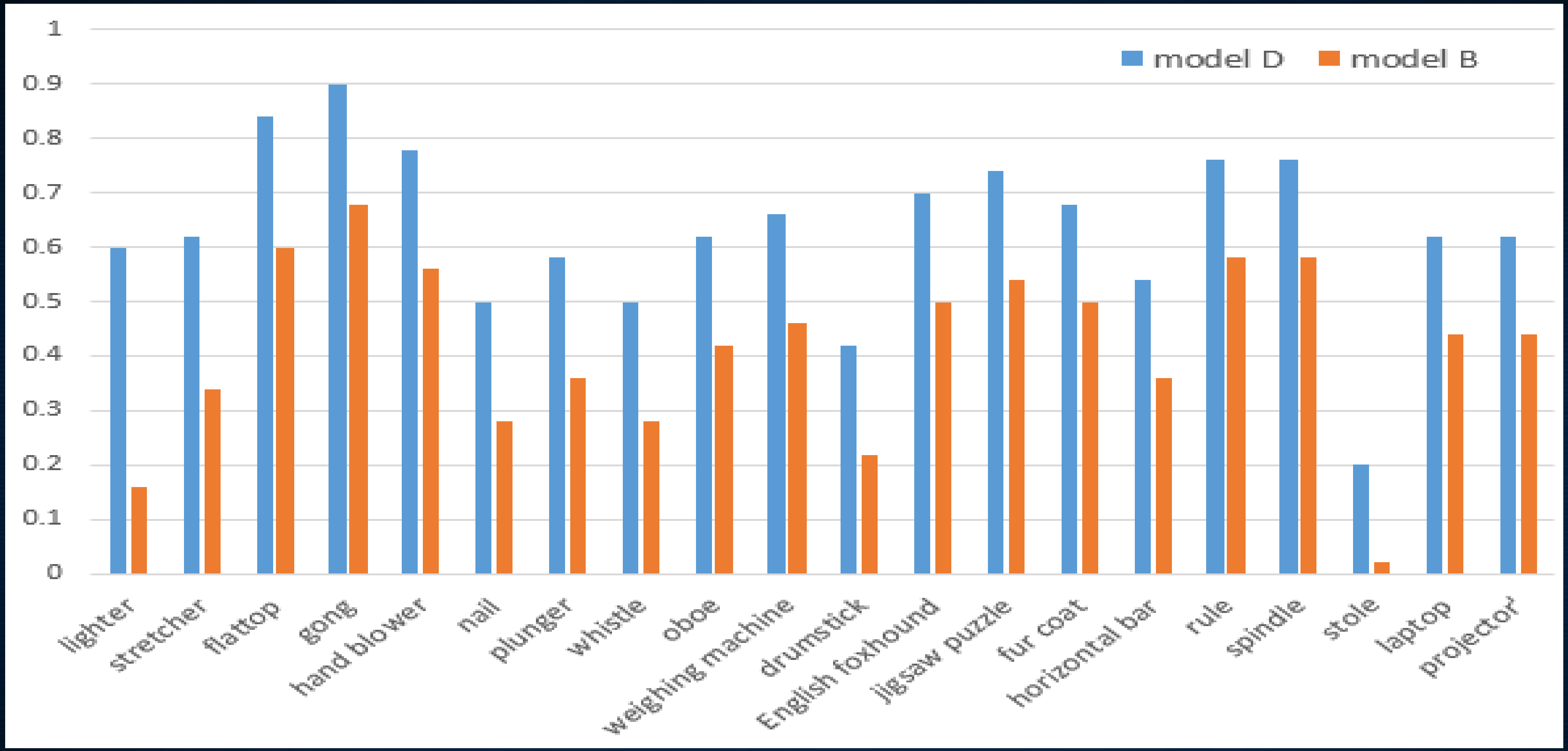
Noise data(%)	Top1	Top5
0	28.44%	11.38%
25%	28.17%	10.93%
50%	**27.91%**	**10.82%**
75%	28.48%	11.07%
100%	28.33%	10.94%

Table 2. Model-D with various amounts of highly noisy data.

Networks	Top1	Top5
Inception_v2	27.91%	10.82%
Inception_v3	22.21%	7.88%
Inception_v4	21.97%	6.64%
Inception_resnet_v2	**20.70%**	**6.38%**

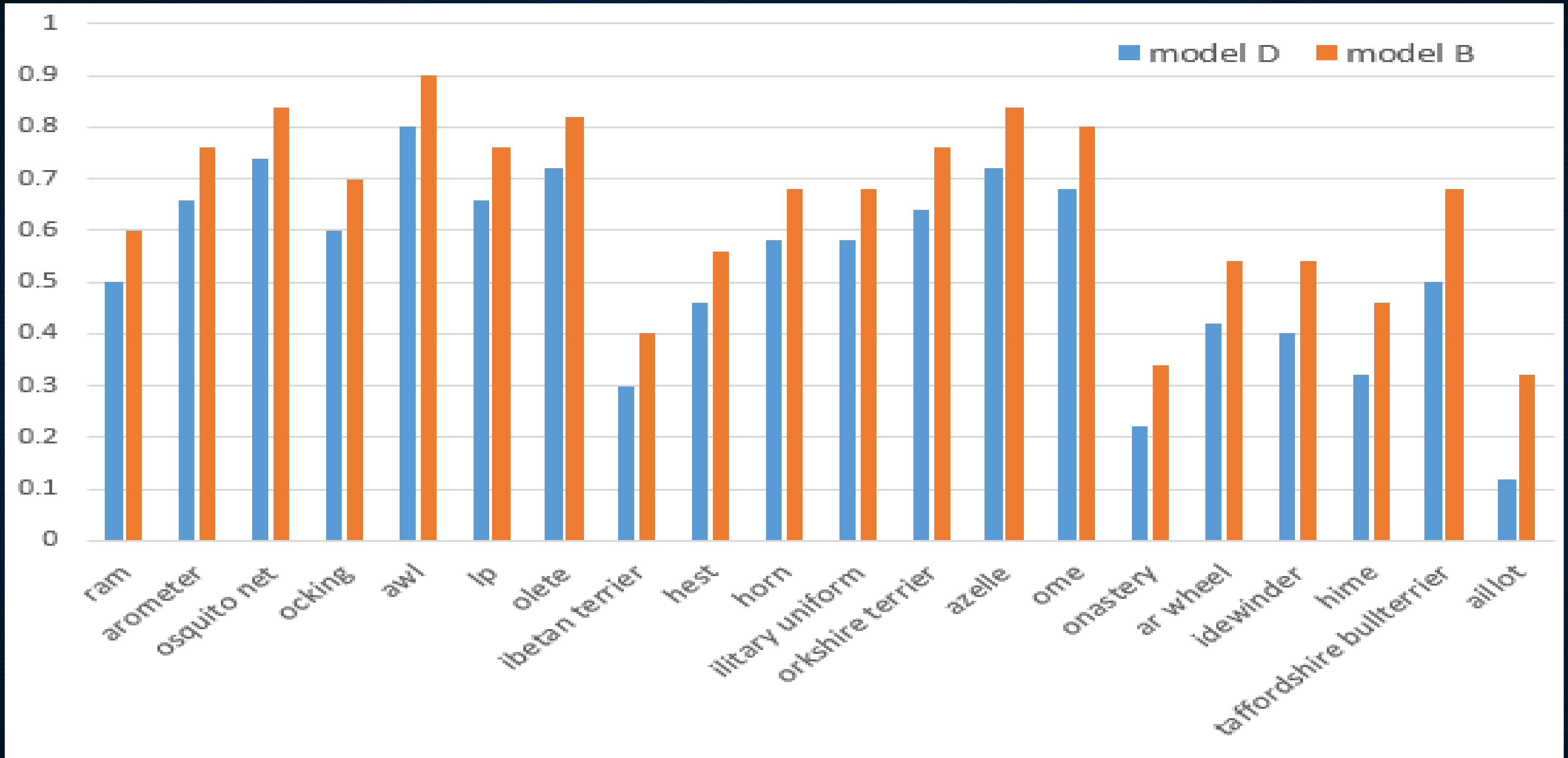
Table 3. Model-D with various networks

Comparisons: Model B & D – Top Positive Categories



Improve 668 categories, reduce 195 categories, and 137 unchanged

Comparisons: Model B & D – Top Negative Categories



Improve 668 categories, reduce 195 categories, and 137 unchanged

webvision | MALONG

Rank	Team name	Run1	Run2	Run3	Run4	Run5
1	Malong AI Research	0.9358	0.9467	**0.9478**	**0.9478**	0.9470
2	SHTU_SIST	0.9223	**0.9225**	0.9218	0.9219	0.9216
3	HG-AI	**0.9189**	0.9152	0.9152	0.9189	0.9189
4	VISTA	0.8979	**0.9005**	0.8980	0.8992	0.8980
5	LZ_NES	**0.8853**	0.8758	0.8723	0.8504	0.8504
6	CRCV	0.8707	0.8717	0.8701	0.8712	**0.8721**
7	Chahrazad	**0.8705**	0.8705	0.8705	0.8705	0.8705
8	Gombru (CVC and Eurecat)	0.8475	0.8374	**0.8586**	0.8586	0.8586

Conclusions

Summary:

—> Train high-performance CNNs from large-scale web images

—> Handle label inconsistence and data unbalance

—> Better generalization capability

—> Improve our products where real-world data was clawed from Internet with less human labelling or labels are inconsistence

—> Will develop semi-supervised and weakly-supervised approaches

Team Members

Our Team:

Sheng Guo, Weilin Huang, Chenfan Zhuang, Dengke Dong, Haozhi Zhang, Matthew R. Scott, Dinglong Huang

Malong Technologies Co., Ltd.

Team history

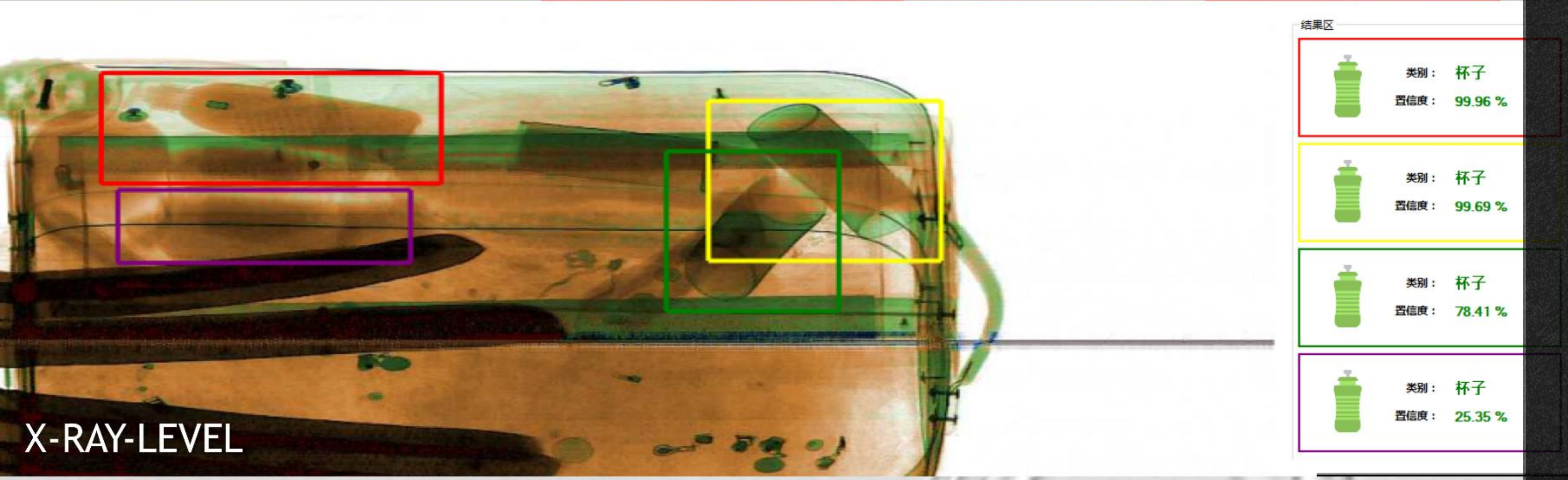
Team members achievements on large-scale challenges:

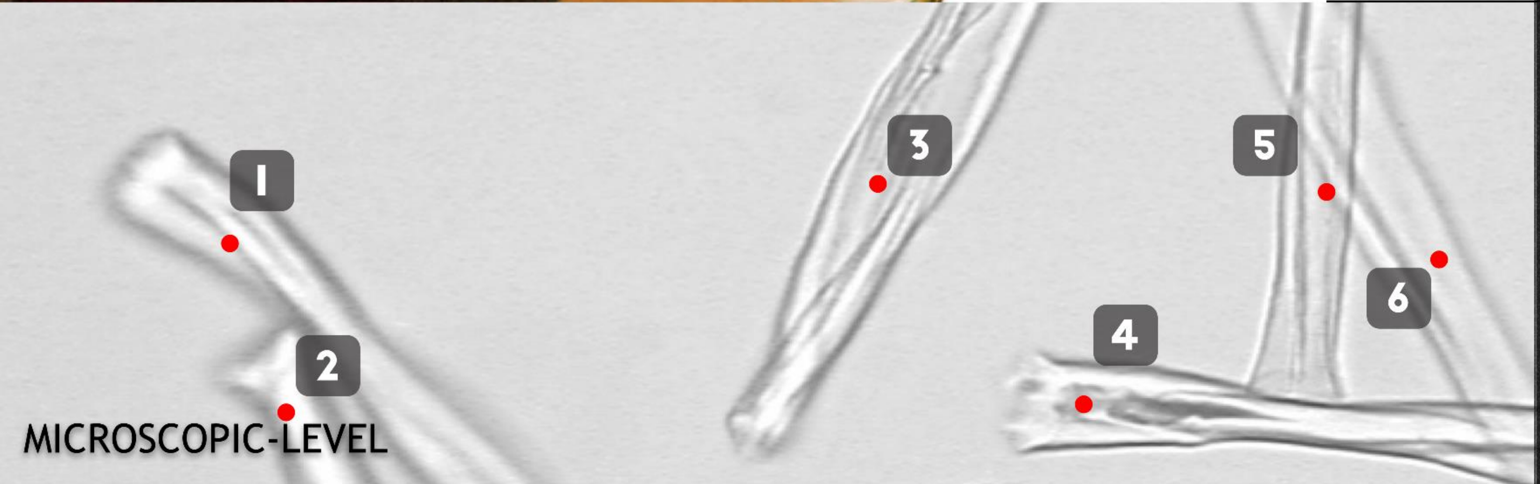
- ICCV - 15: ILSVRC2015 (ImageNet): scene classification - 2nd
- CVPR - 15 :Large-scale Scene Understanding Challenge (LSUN): scene classification - 2nd
- CVPR - 15 : ChaLearn Looking at People Challenge 2015: cultural event recognition - 3rd
- CVPR - 16 : Large-scale Scene Understanding Challenge (LSUN): scene classification - 1st
- ECCV - 16: ILSVRC2016 (ImageNet): scene classification - 4th
- CVPR -17: Webvision Image classification – 1st

About Malong

PRODUCTAI

AI for Product Recognition.





FULL-STACK PRODUCT RECOGNITION

We live in a world of products. In retail, manufacturing, and security scenarios, products need to be routinely recognized at a high-level, microscopic-level, and even the invisible (x-ray) level. If a machine can "see" products as well as people can, higher efficiency can be achieved in retail product checkouts, higher quality in manufacturing product testing, and higher safety via baggage scanning of products – just to name a few. Using breakthrough GPU-powered semi-supervised deep learning algorithms, scientists at Malong invented product recognition technology which operates at human-level performance across the full-stack of visual input levels – the big, the small, and the invisible, to help improve efficiency, quality, and safety, for our world.

PRODUCTAI


Thank you !
We are hiring - Say hello at:
HR@MALONG.COM