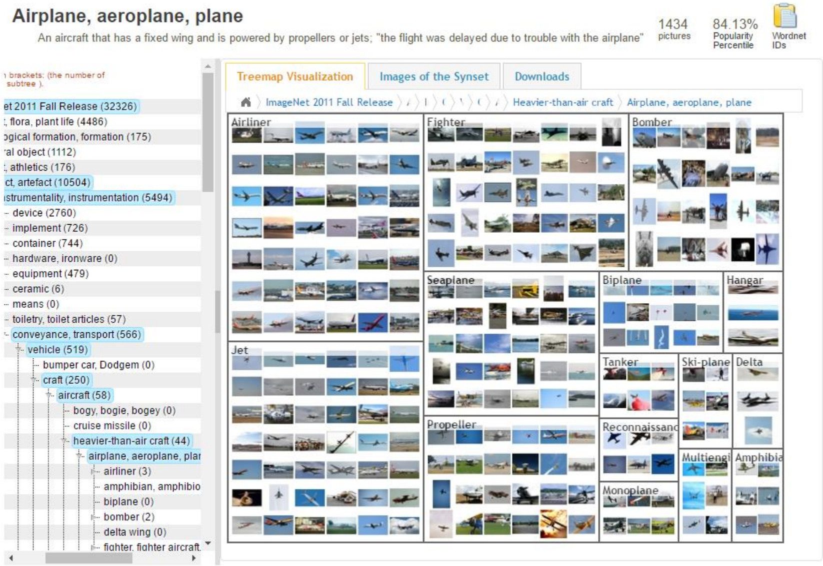
동국대학교
dongguk university

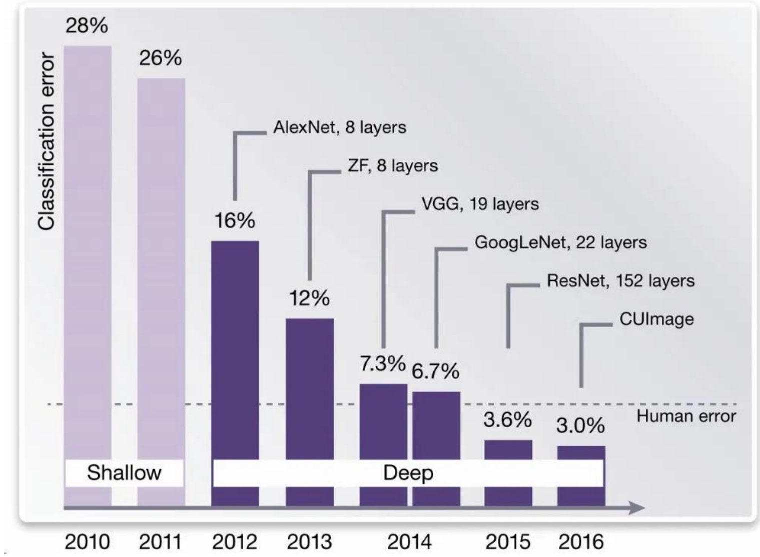
Introduction to Meta-Learning

What is Meta-Learning ?

- Deep Learning
  - Large number of data
  - High computing power
  - Efficient neural network architecture + learning techniques



<example of imagenet dataset>



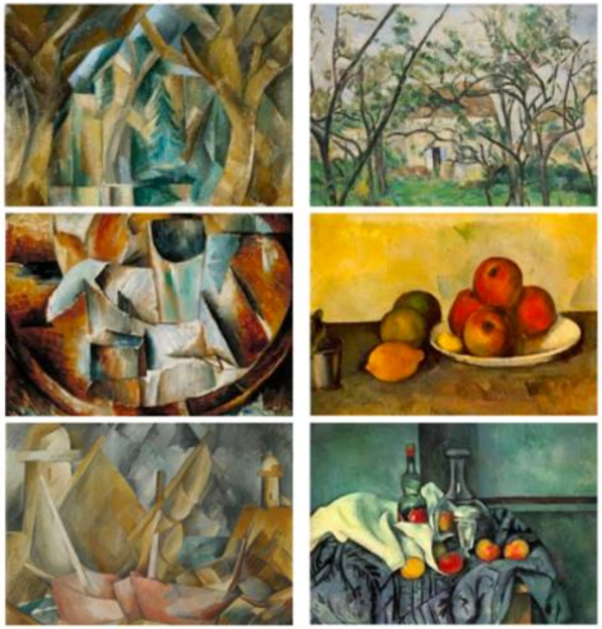
<Imagenet Challenge Resutls>

What is Meta-Learning ?

- What if you don’t have a large dataset?
  - Medical imaging, translation for rare languages, robotics, recommendations, …
  - NOT learning everything from scratch

training data

Braque Cezanne

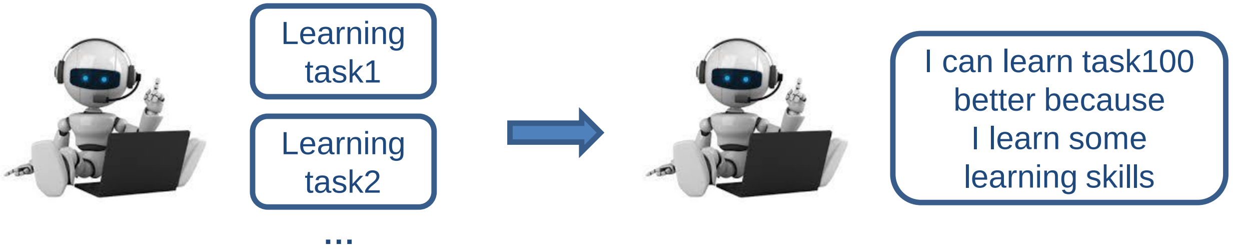


test datapoint

By Braque or Cezanne?

What is Meta-Learning ?

- Meta-Learning: Learning to learn
  - Go beyond train from samples from a single task
  - Learn how to learn efficiently from a sequence of learning experiences


Learning task1
Learning task2
...
I can learn task100 better because I learn some learning skills

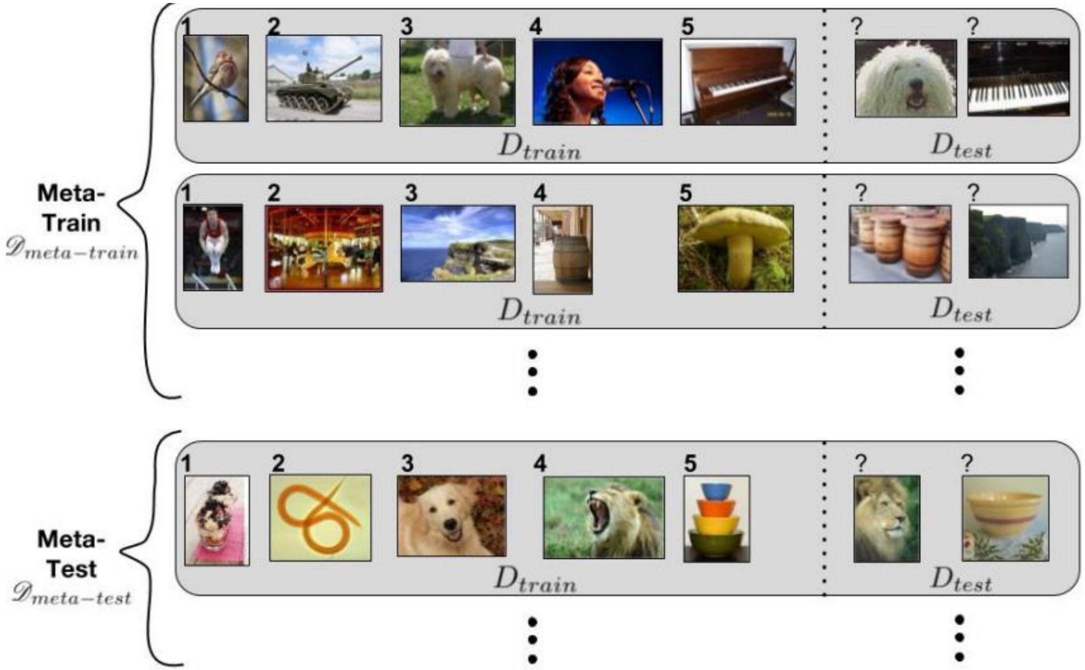
  - Meta-train → Meta-test

$$\left(D_i^{train}, D_i^{test}\right) \rightarrow \left(D^{train}, D^{test}\right)$$

  - 3 common approaches
    - Model-based : using networks with external or internal memory
    - Metrics-based : learning effective distance metrics
    - Optimization-based : explicitly optimizing model parameters for fast learning

Few-shot Learning

- N-way K-shot
  - N classes
  - Only K data instances
- Meta-train → Meta-test



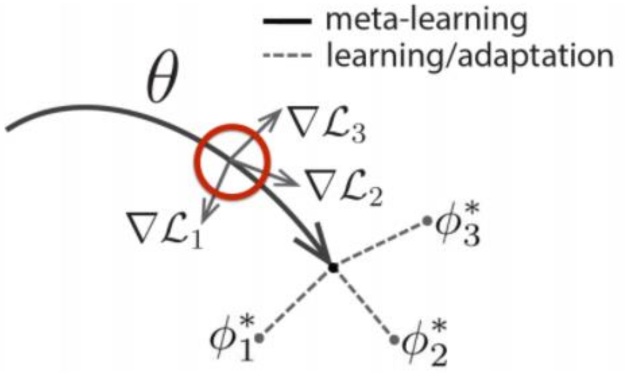
MAML

- Model-Agnostic Meta Learning
  - An algorithm for meta-learning that trains a model's parameters such that a small number of gradient updates will lead to fast learning on a new task

Meta-learning $$\min_{\theta} \sum_{\text{task } i} \mathcal{L}(\theta - \alpha \nabla_{\theta} \mathcal{L}(\theta, \mathcal{D}_i^{\text{tr}}), \mathcal{D}_i^{\text{ts}})$$

θ parameter vector being meta-learned

ϕ_i^* optimal parameter vector for task i



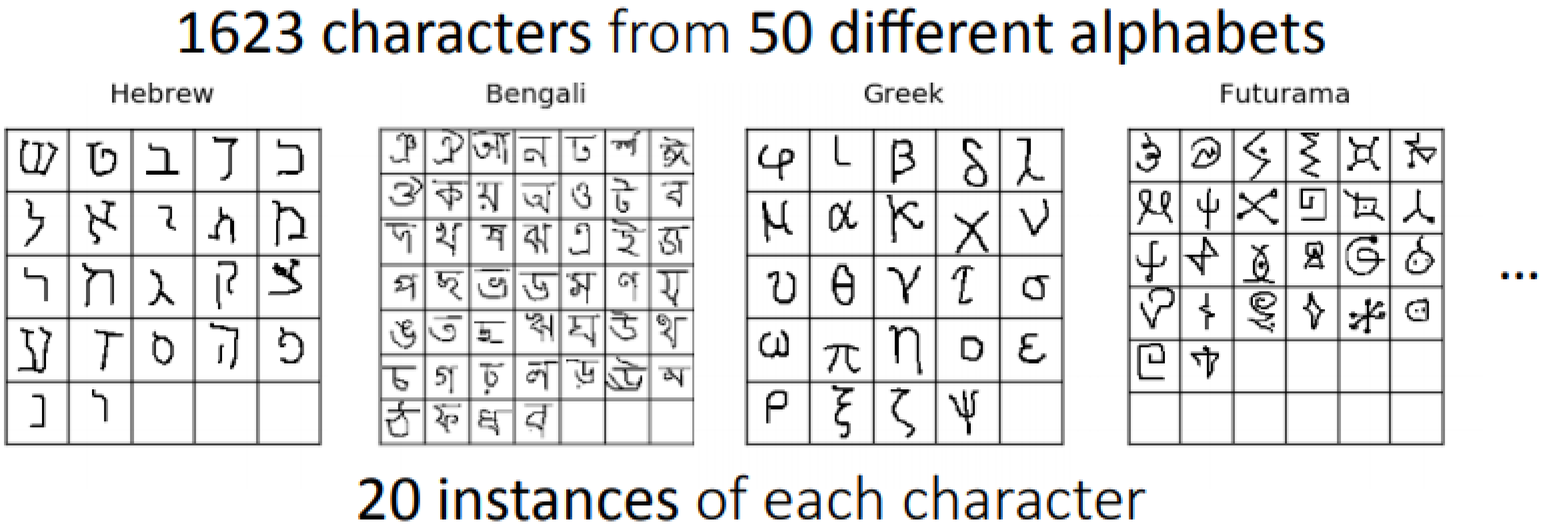
pre-trained parameters

Fine-tuning [test-time] $$\phi \leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}(\theta, \mathcal{D}^{\text{tr}})$$

training data for new task

Omniglot Dataset

- Omniglot
  - A dataset that is used to evaluate a meta-learning algorithm
  - 1623 different handwritten characters from 50 different alphabets written by 20 different people
  - Each image is of size 105x105


1623 characters from 50 different alphabets
Hebrew
Bengali
Greek
Futurama
...
20 instances of each character