Deep meta-learning: learning to learn in the concept space

---Zev Xie, Yuxuan Luo

The outline of the presentation:

- Some definitions
- The starting point of this paper
- The structure of deep meta-learning(DML)
- The result and the contributions.

1.Some definition.

- Meta-Learning: Meta-learning, or learning to learn, is the science of systematically observing how different machine learning approaches perform on a wide range of learning tasks, and then learning from this experience, or meta-data, to learn new tasks much faster than otherwise possible.
- Few-shot Learning: As the name implies, few-shot learning refers to the practice of feeding a learning model with a very small amount of training data, contrary to the normal practice of using a large amount of data.
- Deep Meta-learning: Due to the deep structure of Concept generator, the network for meta-learning call Deep meta-learning.

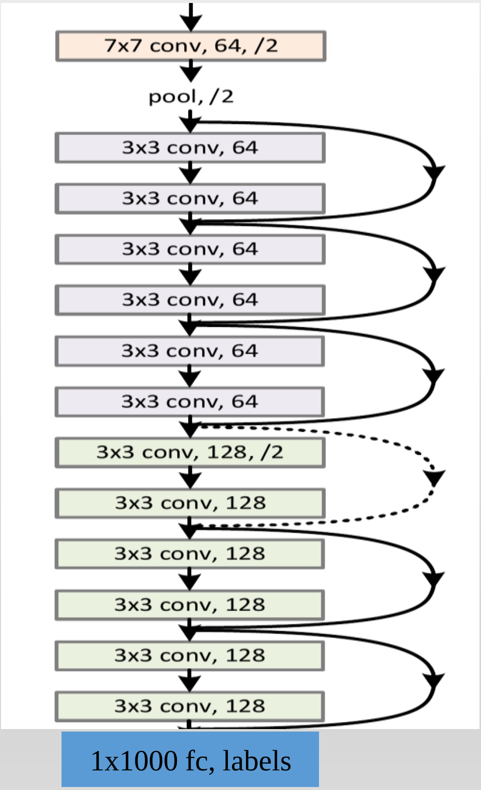
2.The starting point of this paper

【Abstract】 Few-shot learning remains challenging for meta-learning that learns a learning algorithm (meta-learner) from many related tasks. In this work, we argue that this is due to the lack of a good representation for meta-learning, and propose deep meta-learning to integrate the representation power of deep learning into meta-learning. The framework is composed of three modules, a concept generator, a meta-learner, and a concept discriminator, which are learned jointly. The concept generator, e.g. a deep residual net, extracts a representation for each instance that captures its high-level concept, on which the meta-learner performs few-shot learning, and the concept discriminator recognizes the concepts. By learning to learn in the concept space rather than in the complicated instance space, deep meta-learning can substantially improve vanilla meta-learning, which is demonstrated on various few-shot image recognition problems. For example, on 5-way-1-shot image recognition on CIFAR-100 and CUB-200, it improves Matching Nets from 50.53% and 56.53% to 58.18% and 63.47%, improves MAML from 49.28% and 50.45% to 56.65% and 64.63%, and improves Meta-SGD from 53.83% and 53.34% to 61.62% and 66.95%, respectively.

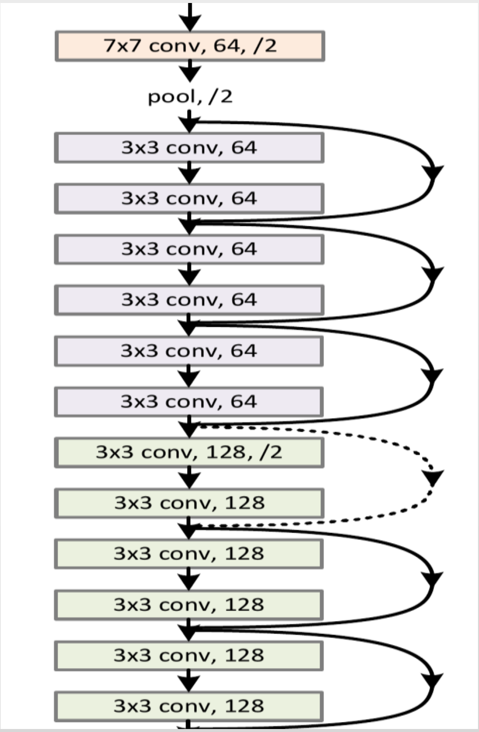
3.1 The meaning of concept.

Concept: The set of features differentiate the subset can be called concept. The concept is a kind of higher level information than the raw data. In this model, every image will be transformed into concept by Resnet-50.

ResNet-50



Concept generator



3.2 The three modules of DML.

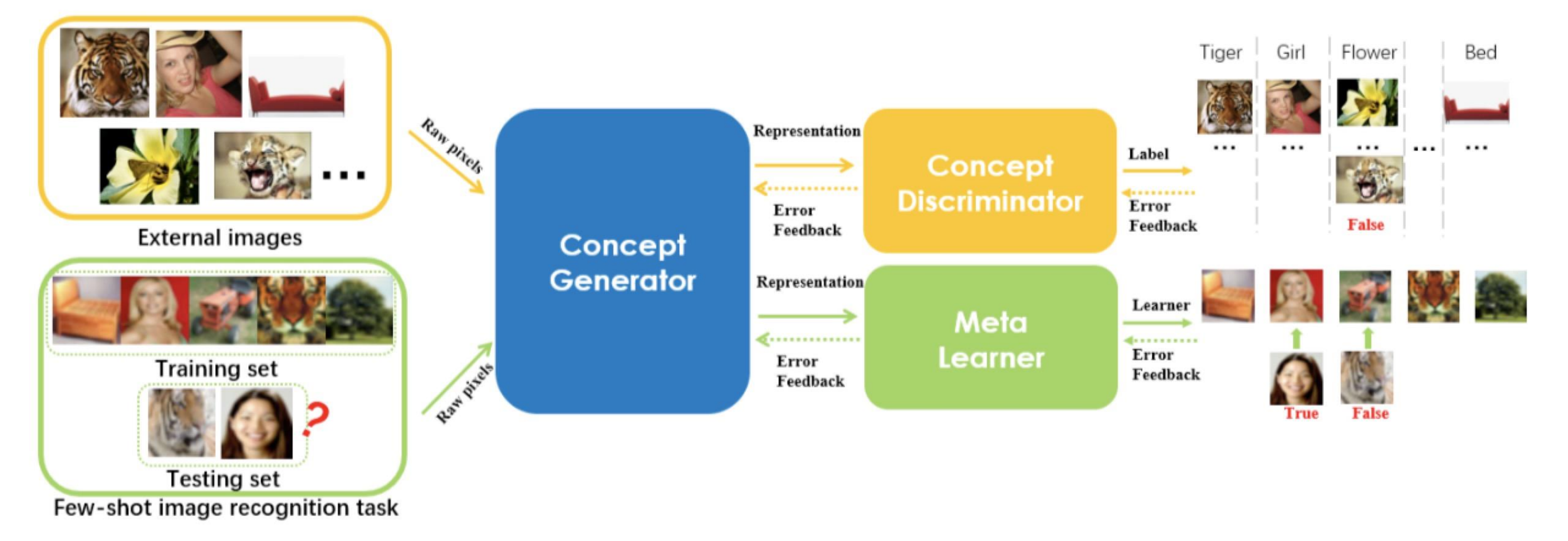
Concept generator : encoder the Image to a vector(concept space)

- Concept generator CNN: AlexNet, Inception, VGG, ResNet

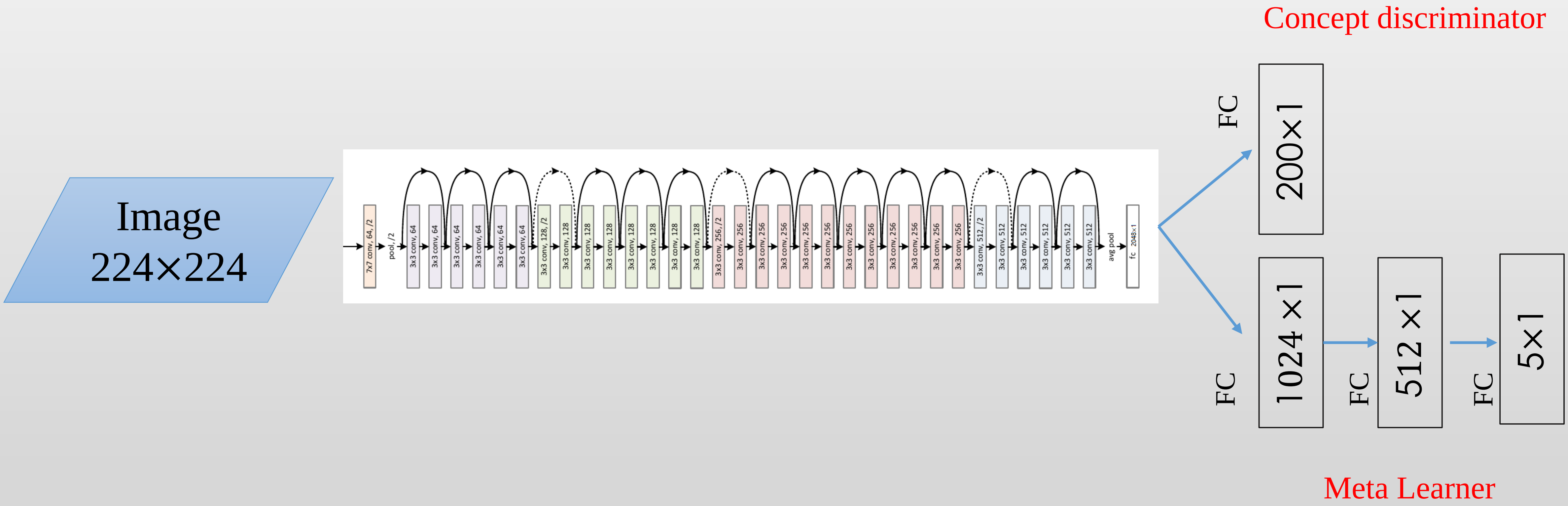
Concept discriminator :enhances the concept generator by handling image recognition tasks from an external large-scale datasets.

- Concept discriminator SVM, nearest network, neural network
- Meta-learner MatchingNets, MAML, Meta-SGD

Meta Learner:Fast adaptive learning with few samples.



3.3 The DML structure:



3.4 The algorithm.

Algorithm 2 Deep Meta-Learning with Meta-SGD

Input: task distribution $p(\mathcal{T})$**, labeled dataset** $\mathbb{D}$**, batch size** n **of tasks, batch size** m **of instances, learning rate** β

Output:$\boldsymbol{\theta}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{D}}, \boldsymbol{\theta}_{\mathcal{M}} = \{\boldsymbol{\phi}, \boldsymbol{\alpha}\}$

Initialize $\boldsymbol{\theta}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{D}}, \boldsymbol{\phi}, \boldsymbol{\alpha}$

while not done **do**

  Sample n tasks $\mathcal{T}_i \sim p(\mathcal{T})$ and m instances $(\mathbf{x}_j, \mathbf{y}_j) \sim \mathbb{D}$

  for each $\mathcal{T}_i$ **do**

    $\mathcal{L}_{\text{train}(\mathcal{T}_i)}(\boldsymbol{\phi}, \boldsymbol{\theta}_{\mathcal{G}}) \leftarrow \frac{1}{|\text{train}(\mathcal{T}_i)|} \sum_{(\mathbf{x},\mathbf{y}) \in \text{train}(\mathcal{T}_i)} \ell(f_{\boldsymbol{\phi}}(\mathcal{G}(\mathbf{x})), \mathbf{y});$

    $\boldsymbol{\phi}'_i \leftarrow \boldsymbol{\phi} - \boldsymbol{\alpha} \circ \nabla_{\boldsymbol{\phi}} \mathcal{L}_{\text{train}(\mathcal{T}_i)}(\boldsymbol{\phi}, \boldsymbol{\theta}_{\mathcal{G}});$

    $\mathcal{L}_{\text{test}(\mathcal{T}_i)}(\boldsymbol{\phi}'_i, \boldsymbol{\theta}_{\mathcal{G}}) \leftarrow \frac{1}{|\text{test}(\mathcal{T}_i)|} \sum_{(\mathbf{x},\mathbf{y}) \in \text{test}(\mathcal{T}_i)} \ell(f_{\boldsymbol{\phi}'_i}(\mathcal{G}(\mathbf{x})), \mathbf{y});$

  end for

  $(\boldsymbol{\theta}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{D}}, \boldsymbol{\phi}, \boldsymbol{\alpha}) \leftarrow (\boldsymbol{\theta}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{D}}, \boldsymbol{\phi}, \boldsymbol{\alpha}) - \beta \nabla \left[\frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_{\text{test}(\mathcal{T}_i)}(\boldsymbol{\phi}'_i, \boldsymbol{\theta}_{\mathcal{G}}) + \lambda \frac{1}{m} \sum_{j=1}^{m} \ell(\mathcal{D}(\mathcal{G}(\mathbf{x}_j)), \mathbf{y}_j) \right];$

end while

4. The result and the contributions.

Method	MiniImagenet		Caltech-256		CIFAR-100		CUB-200	
	5-way-1-shot	5-way-5-shot	5-way-1-shot	5-way-5-shot	5-way-1-shot	5-way-5-shot	5-way-1-shot	5-way-5-shot
Matching Nets	43.56 ± 0.84	55.31 ± 0.73	48.09 ± 0.83	57.45 ± 0.74	50.53 ± 0.87	60.30 ± 0.82	56.53 ± 0.99	63.54 ± 0.85
DEML+Matching Nets	55.84 ± 0.94	59.88 ± 0.73	52.97 ± 0.99	59.42 ± 0.75	58.18 ± 1.09	63.12 ± 0.85	63.47 ± 1.10	64.86 ± 0.87
MAML	48.70 ± 1.84	63.11 ± 0.92	45.59 ± 0.77	54.61 ± 0.73	49.28 ± 0.90	58.30 ± 0.80	50.45 ± 0.97	59.60 ± 0.84
DEML+MAML	53.71 ± 0.89	68.13 ± 0.77	56.81 ± 1.01	70.54 ± 0.73	56.65 ± 1.09	68.66 ± 0.85	64.63 ± 1.08	66.75 ± 0.89
Meta-SGD	50.47 ± 1.87	64.03 ± 0.94	48.65 ± 0.82	64.74 ± 0.75	53.83 ± 0.89	70.40 ± 0.74	53.34 ± 0.97	67.59 ± 0.82
DEML+Meta-SGD	$\mathbf{58.49 \pm 0.91}$	$\mathbf{71.28 \pm 0.69}$	$\mathbf{62.25 \pm 1.00}$	$\mathbf{79.52 \pm 0.63}$	$\mathbf{61.62 \pm 1.01}$	$\mathbf{77.94 \pm 0.74}$	$\mathbf{66.95 \pm 1.06}$	$\mathbf{77.11 \pm 0.78}$

Comparison between DML and vanilla meta-learning.

4. The result and the contributions.

- Propose deep meta-learning to integrate the power of deep learning into meta-learning
- Equip a meta-learner with a concept generator to enable learning in the concept space while employing a concept discriminator to enhance the concept generator, and show that all three modules can be trained jointly in an end-to-end manner.

$$(\boldsymbol{\theta}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{D}}, \boldsymbol{\phi}, \boldsymbol{\alpha}) \leftarrow (\boldsymbol{\theta}_{\mathcal{G}}, \boldsymbol{\theta}_{\mathcal{D}}, \boldsymbol{\phi}, \boldsymbol{\alpha}) - \beta \nabla \left[\frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_{\mathrm{test}(\mathcal{T}_i)}(\boldsymbol{\phi}_i', \boldsymbol{\theta}_{\mathcal{G}}) + \lambda \frac{1}{m} \sum_{j=1}^{m} \ell(\mathcal{D}(\mathcal{G}(\mathbf{x}_j)), \mathbf{y}_j) \right];$$

Thanks for your Listening.