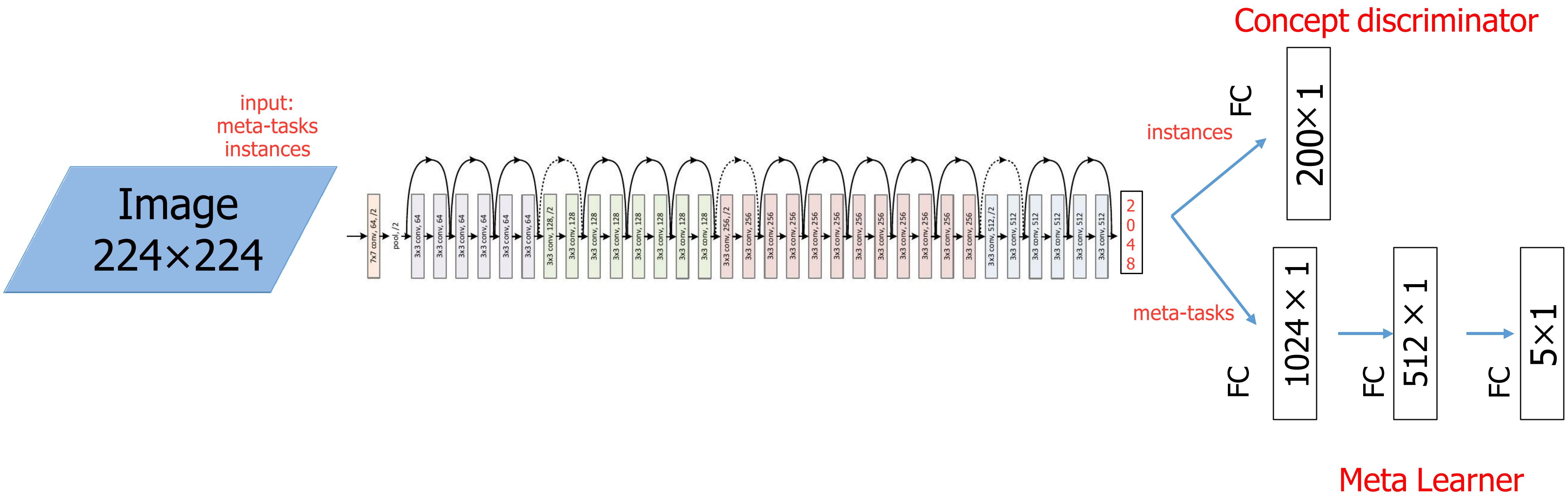
Report about the result of recent experiments

Contents

2.Model structure



2.Starting point of the experiment

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:$\theta_{\mathcal{G}}, \theta_{\mathcal{D}}, \theta_{\mathcal{M}} = \{\phi, \alpha\}$

Initialize $\theta_{\mathcal{G}}, \theta_{\mathcal{D}}, \phi, \alpha$

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**

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

    $\phi_i' \leftarrow \phi - \alpha \circ \nabla_{\phi} \mathcal{L}_{\mathrm{train}(\mathcal{T}_i)}(\phi, \theta_{\mathcal{G}});$

    $\mathcal{L}_{\mathrm{test}(\mathcal{T}_i)}(\phi_i', \theta_{\mathcal{G}}) \leftarrow \frac{1}{|\mathrm{test}(\mathcal{T}_i)|} \sum_{(\mathbf{x},\mathbf{y}) \in \mathrm{test}(\mathcal{T}_i)} \ell(f_{\phi_i'}(\mathcal{G}(\mathbf{x})), \mathbf{y});$

  end for

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

end while

The generator plays a role of prior knowledge , but it learns by using the external dataset and the few-shot image recognition task. The experiment of the paper assume the datasets are similar. However, there is no scrutinizing step to guarantee that. Will the unrelated dataset contaminate the prior knowledge?

2.How we build the model

1.The model structure is totally the same as shown before coded by PyTorch.

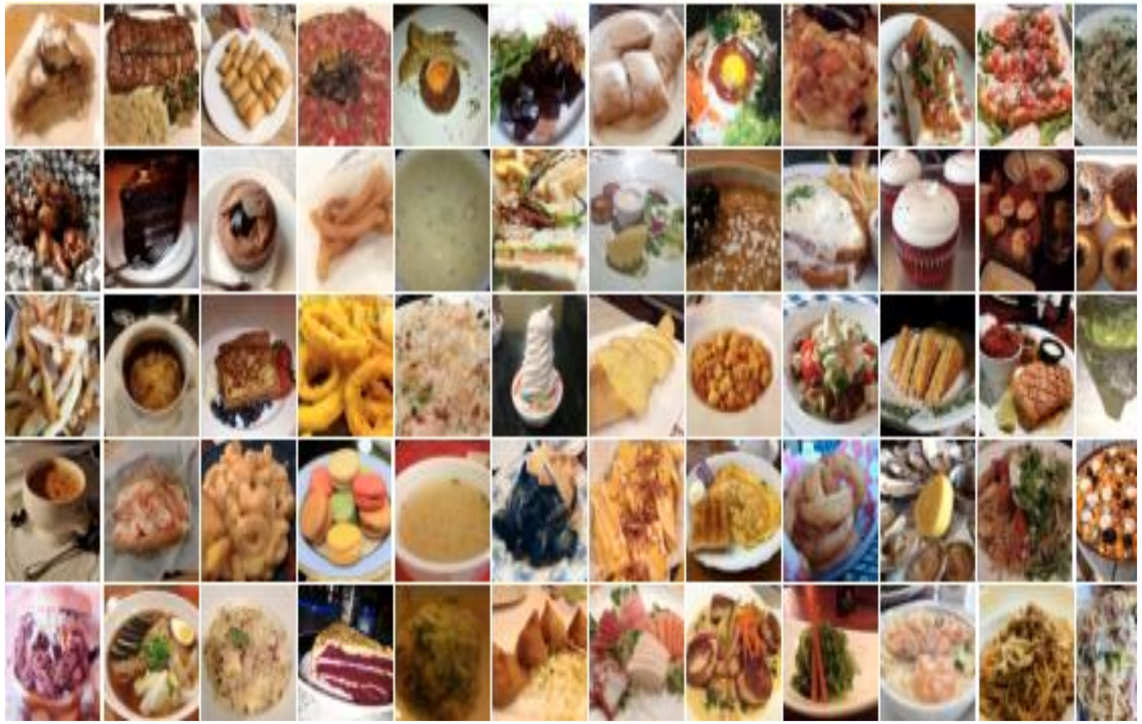
2.In order to update the three part of model simultaneously, we calculate and sum up the loss of Discriminator and meta-learner. Backward the grad of Discriminator and meta-learner.

3. In the output layer of Generator, we sum up the grad from Discriminator and meta-learner using parameter λ.

4.Modify the grad of Generator and then update the model.

3.Datasets

101 classes of foods



CUB_200(200 classes of birds)

3.Datasets

1.Just use CUB_200 meta-learner with the same architecture excluding Discriminator.
2.Selected 200 classes without any animals from ImageNet for Discriminator and CUB_200 for meta-learner.
3.101 classes of foods from Kaggle and CUB_200 for meta-learner.
4.Randomly extracted 200 classes from ImageNet for Discriminator and CUB_200 for meta-learner.

Parameters setting

1.learning-rate:0.001

2.inner-task:4

3.task:5 classes each with 5 samples

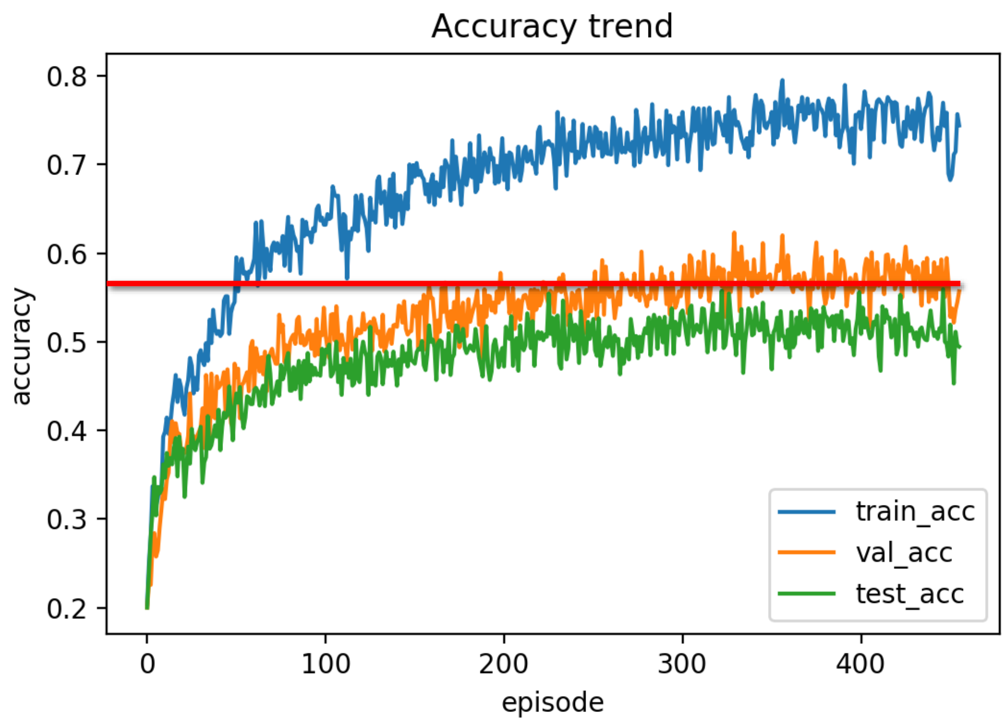
4.batch-size for discriminator:64

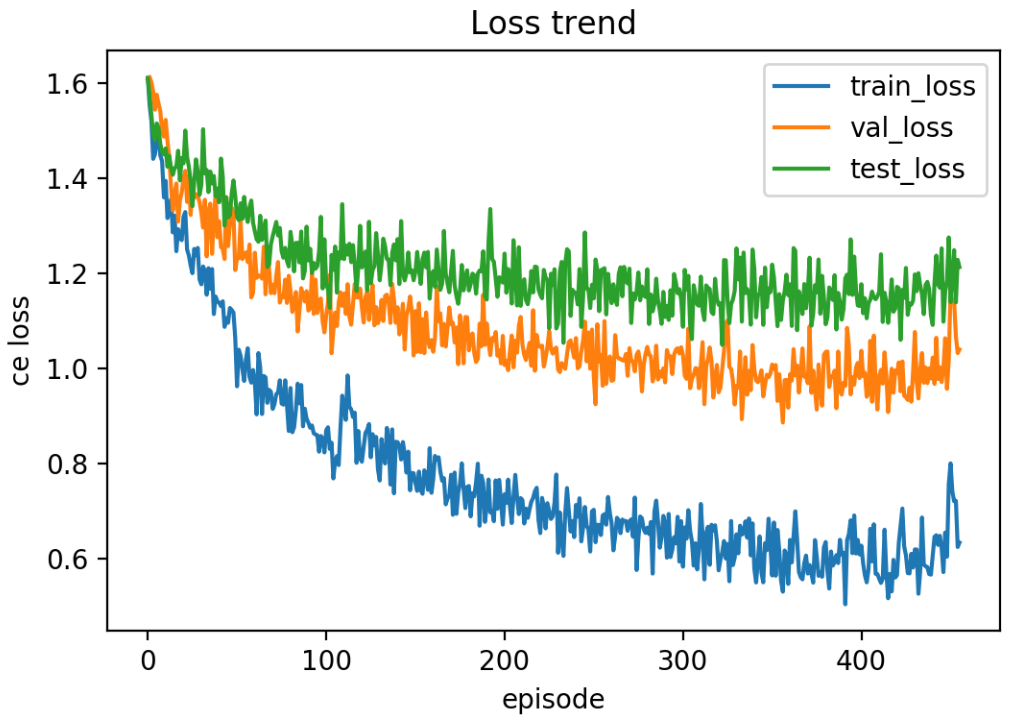
5.learning-rate decay:5000 epochs decay 50%

6.λ :1

7.epochs:30000

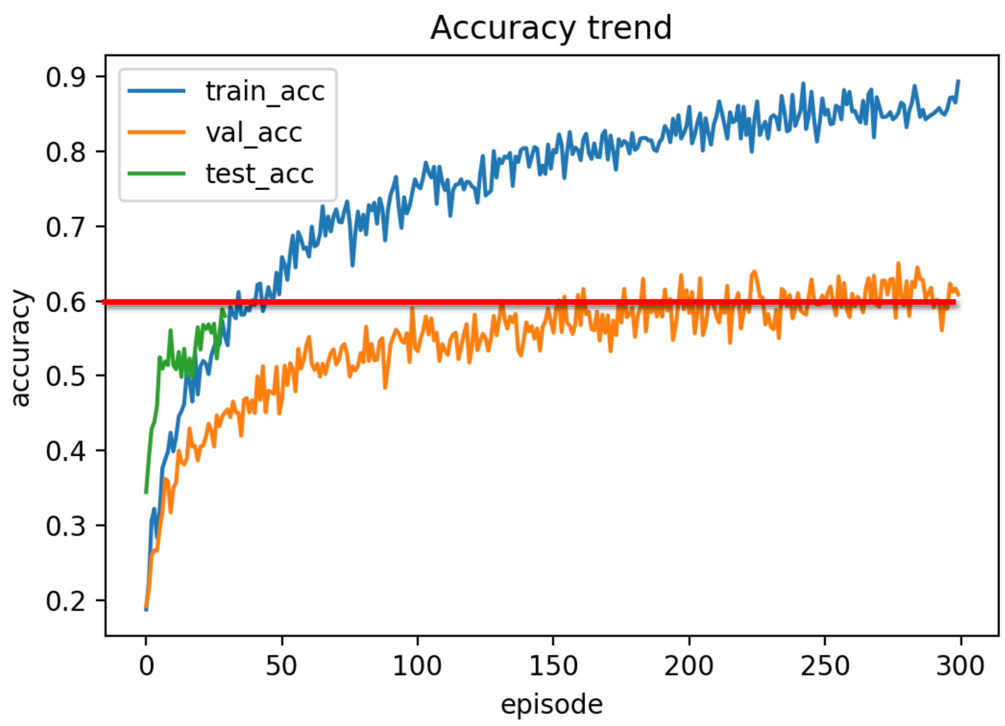
4.Result of original dataset

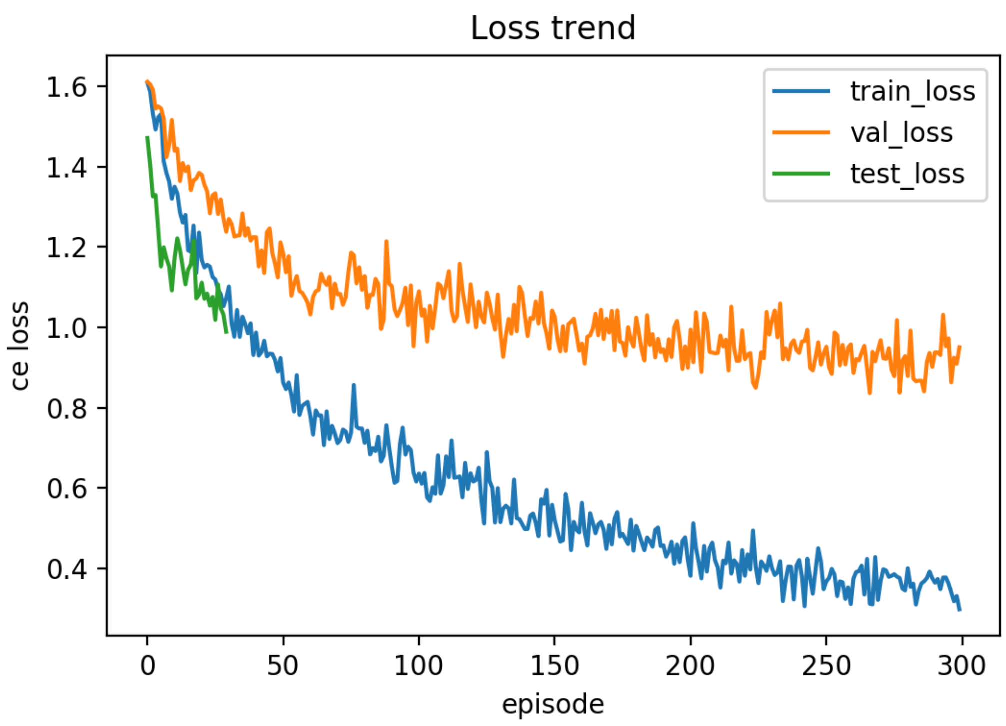




2. CUB_200 for meta-learner.

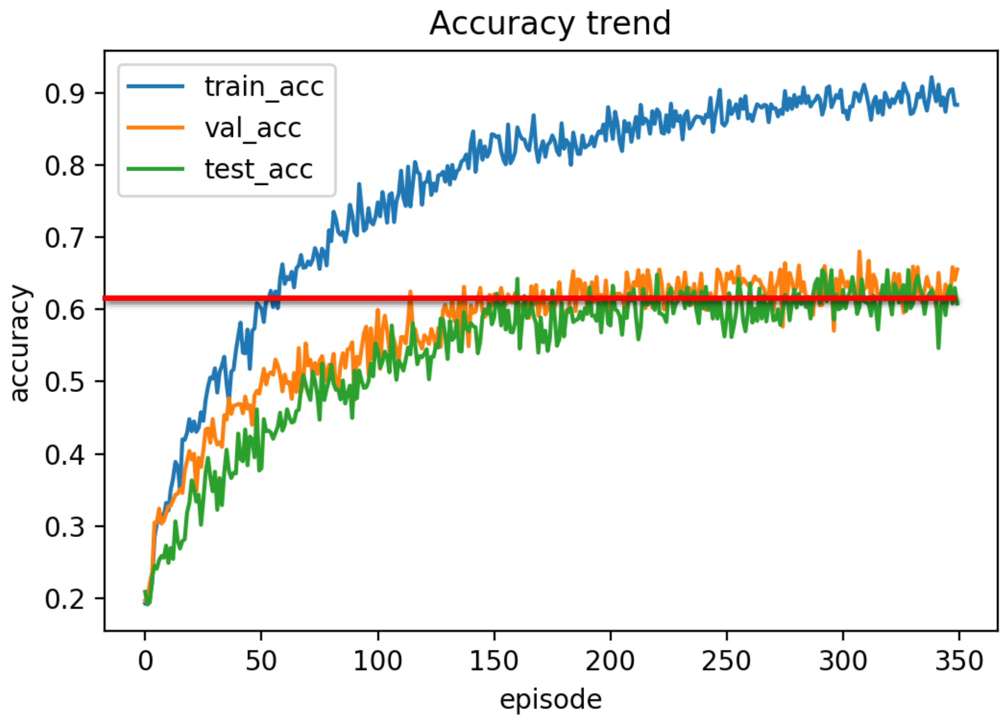
4.Result of irrelative dataset

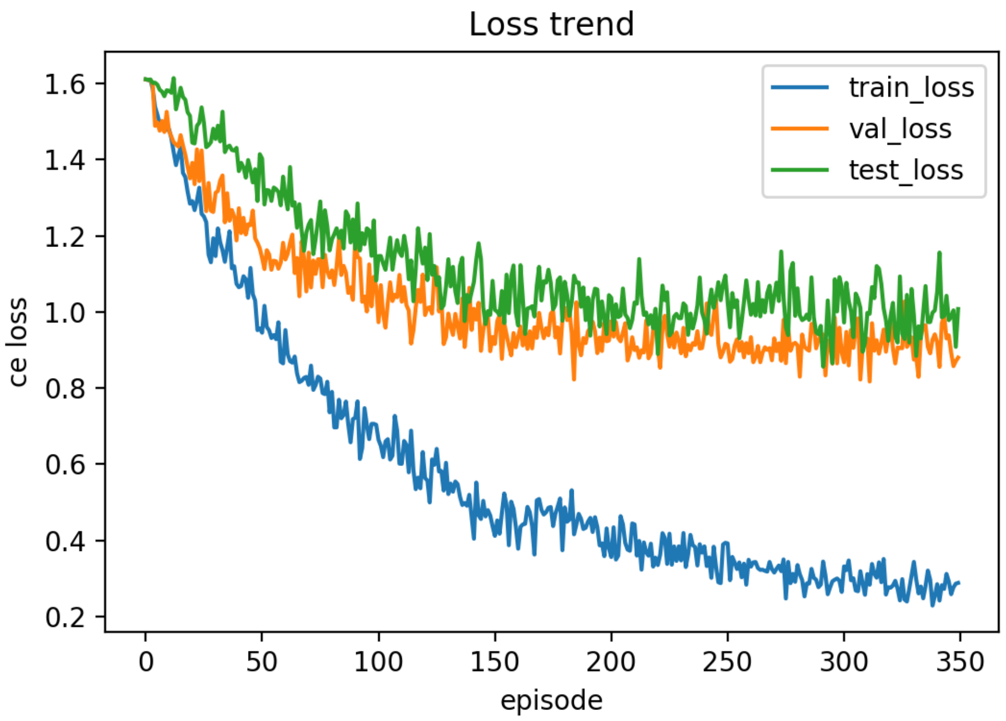




2.Selected 200 classes without any animals from ImageNet for Discriminator and CUB_200 for meta-learner.

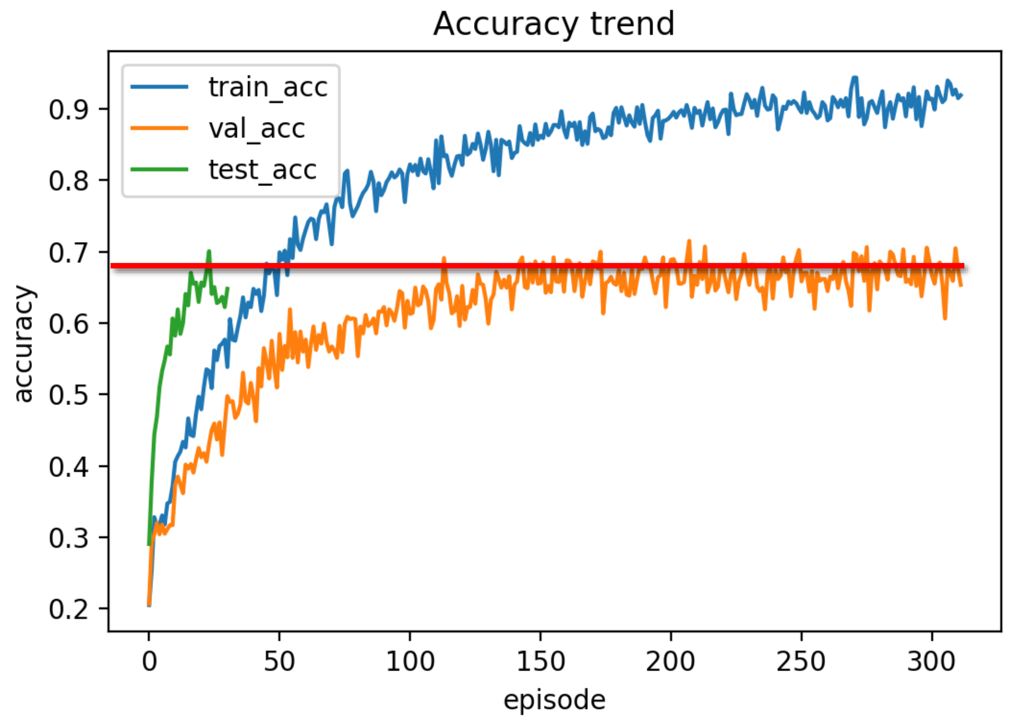
4.Result of irrelative dataset

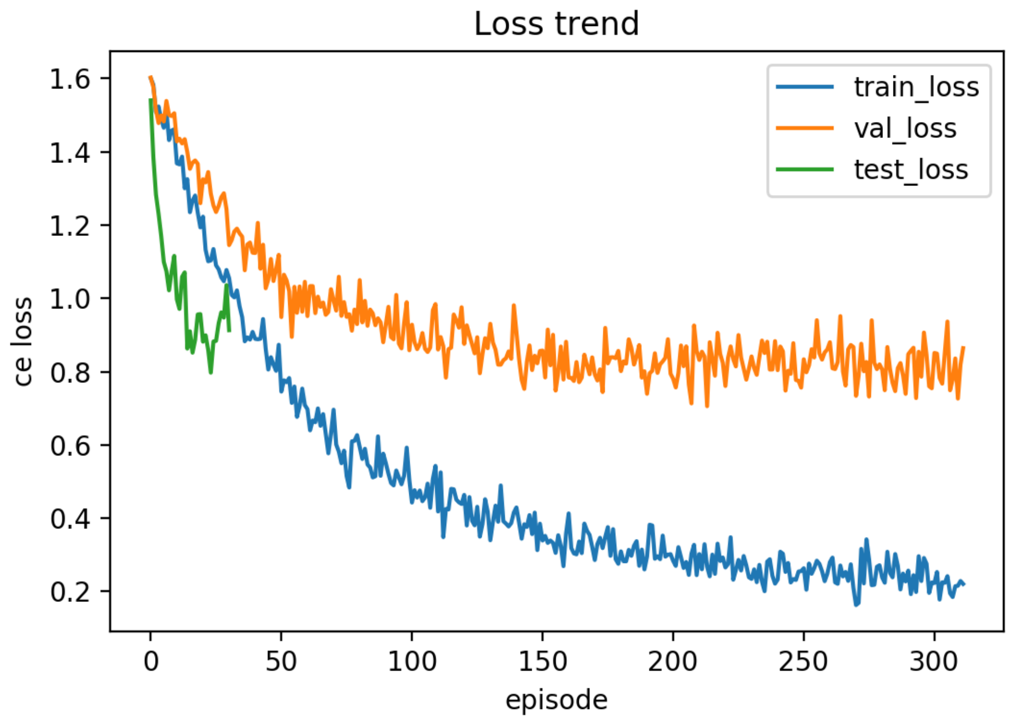




2.Selected 101 kinds of food for Discriminator and CUB_200 for meta-learner.

4.Result of relative dataset





2.Selected 200 classes with many animals from ImageNet for Discriminator and CUB_200 for meta-learner.

Conclusion

1.Relative external datasets make great contribution to model training.

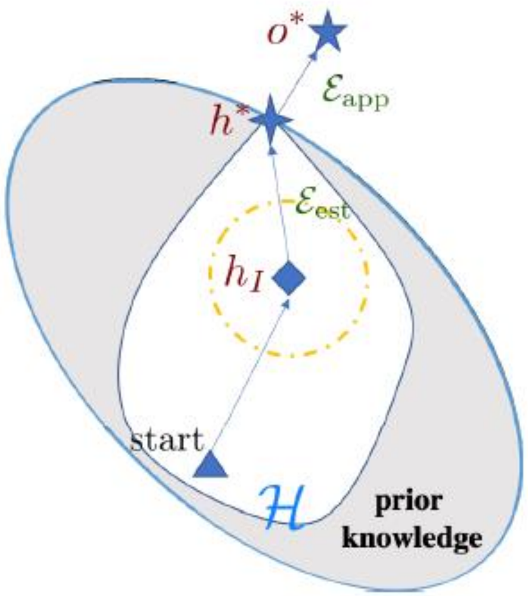
2.Irrelative external datasets make very small contribution to Generator and meta-learner but not affect the performance.

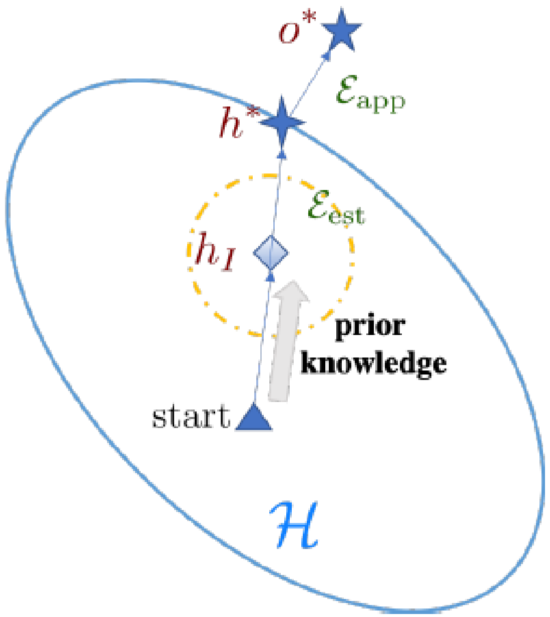
3.Without external datasets, the model performs worse than the model with

irrelative external datasets.

4.The uncertainty

1.the role of the shared parameter θ.(Constrain the complexity of H or help to search for θ?)





2. θ_m aims to learn the sensitive parameters from the changing of tasks. As a result, the small change in grad and parameters will cause larger loss.(Is there any criterion in the small change will cause how big the loss change?)

5.Future work

1.Using the attention mechanism to prove the correlation between two datasets.

2.Try to combining GAN with this model