CLOUD CLASSIFICATION OF CELLPHONE PHOTOS BY MACHINE LEARNING

Kun Yue, Jing Zhuang, Lili Ding
Moji Co.,Ltd Beijing China

2017.12.31

2009
MOJI
Weather

2011
Weather
Alarm

2013
Weather photo
Community

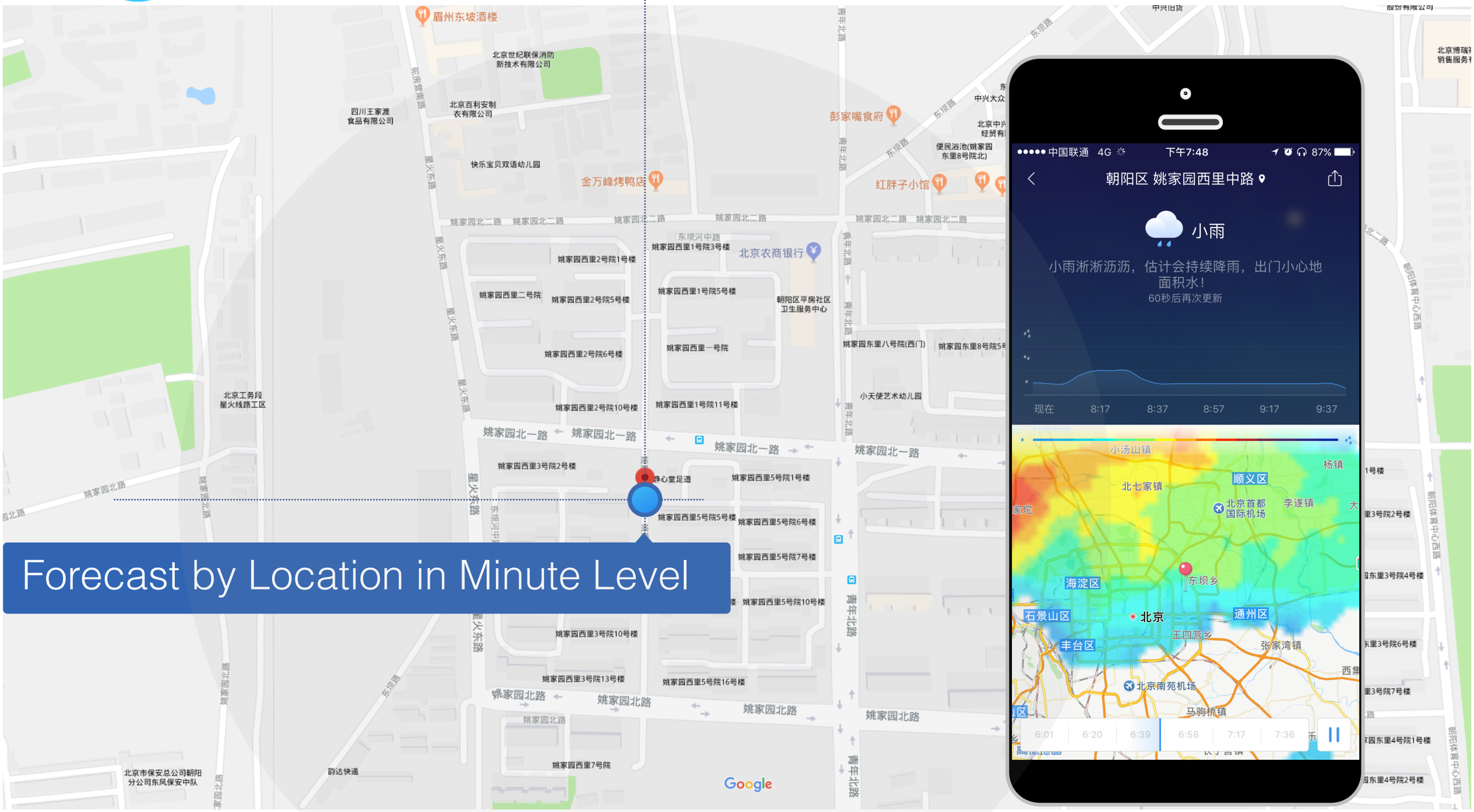
2014
AirNut

2015
Weather
Research

2016
Big Data

2017
MOJI
A Dream

Nowcasting by Machine Learning



Forecast by Location in Minute Level

Nowcasting by Machine Learning

Radar Images

Radar Images Preprocessing

Radar Images Denoising

Machine Learning

TensorFlow

OpenCV

python™

Caffe

scikits learn machine learning in Python

NVIDIA.

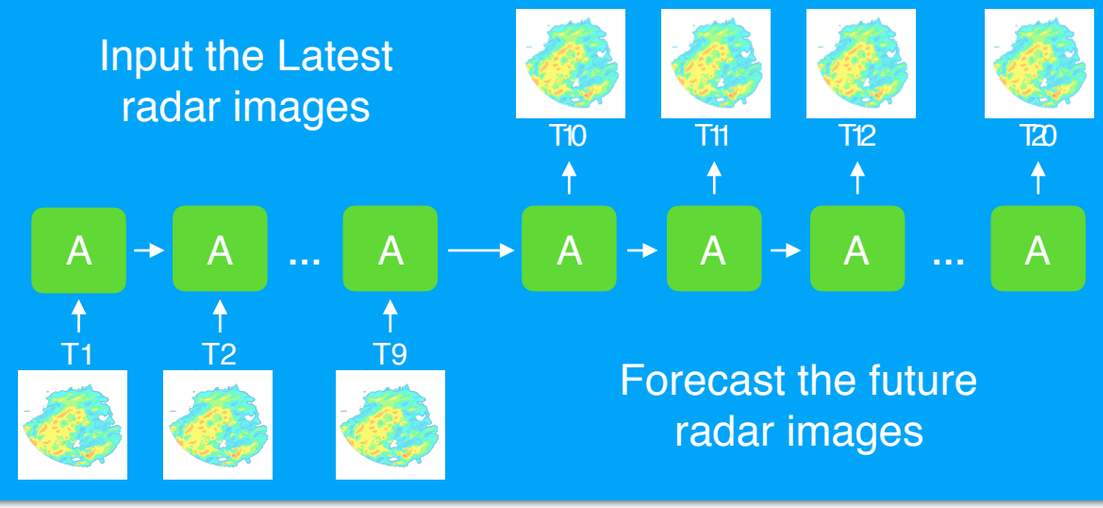
pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

NumPy

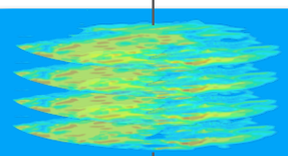
Spatiotemporal Forecasting

Input the Latest radar images



Forecast the future radar images

Display



Results Merging

Nowcasting by Machine Learning

Mainly forecast the precipitation for the next tow hours in minutes level, the forecast data is updated every 6 minutes. Users request data according to the latitude and longitude.

Observed data inputed in the nowcasting system include:

- Radar echo Image (updated per 6 minutes)
- Satellite Image (updated per 10 minutes)
- Numerical Model (updated per hour)

Problems in the nowcasting system:

- Radar echo images is inaccurate sometimes because of noise, topography or hardware problem
- Satellite Image cannot reflect the amount of rainfall accurately
- Numerical model and Observed station is distribute sparsely
- Interpolation reduces accuracy and updates slowly

Nowcasting by Machine Learning

Is there any other data that can be used as a supplement?

Demand：Timeliness、Geographic accuracy

Mobile Phone data could match two points above

What UGC data Moji has?

- Depend on **500 million** users
- Depend on **50 million** Daily active users
- **50,000 ~ 100,000** images/day
- **50,000 ~ 60,000** commons/day
- **60,000,000** cellphone barometers/day

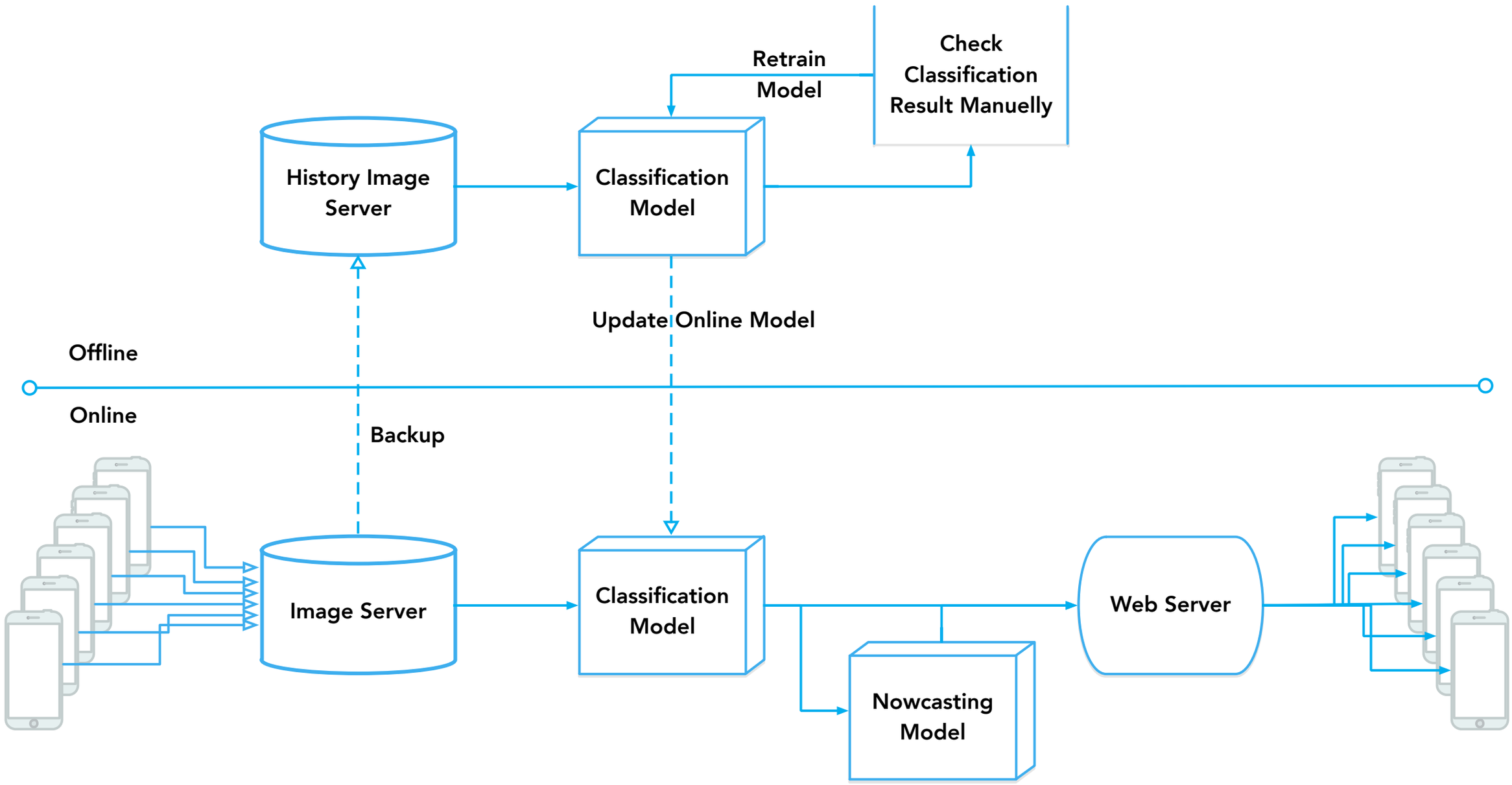
The most intuitive data is images

There are various categories of images

Cloud maps can accurately analyze the weather conditions in the area where the user is located

What we need to do is to identify the information in the cloud through image recognition

Nowcasting by Machine Learning



Nowcasting by Machine Learning

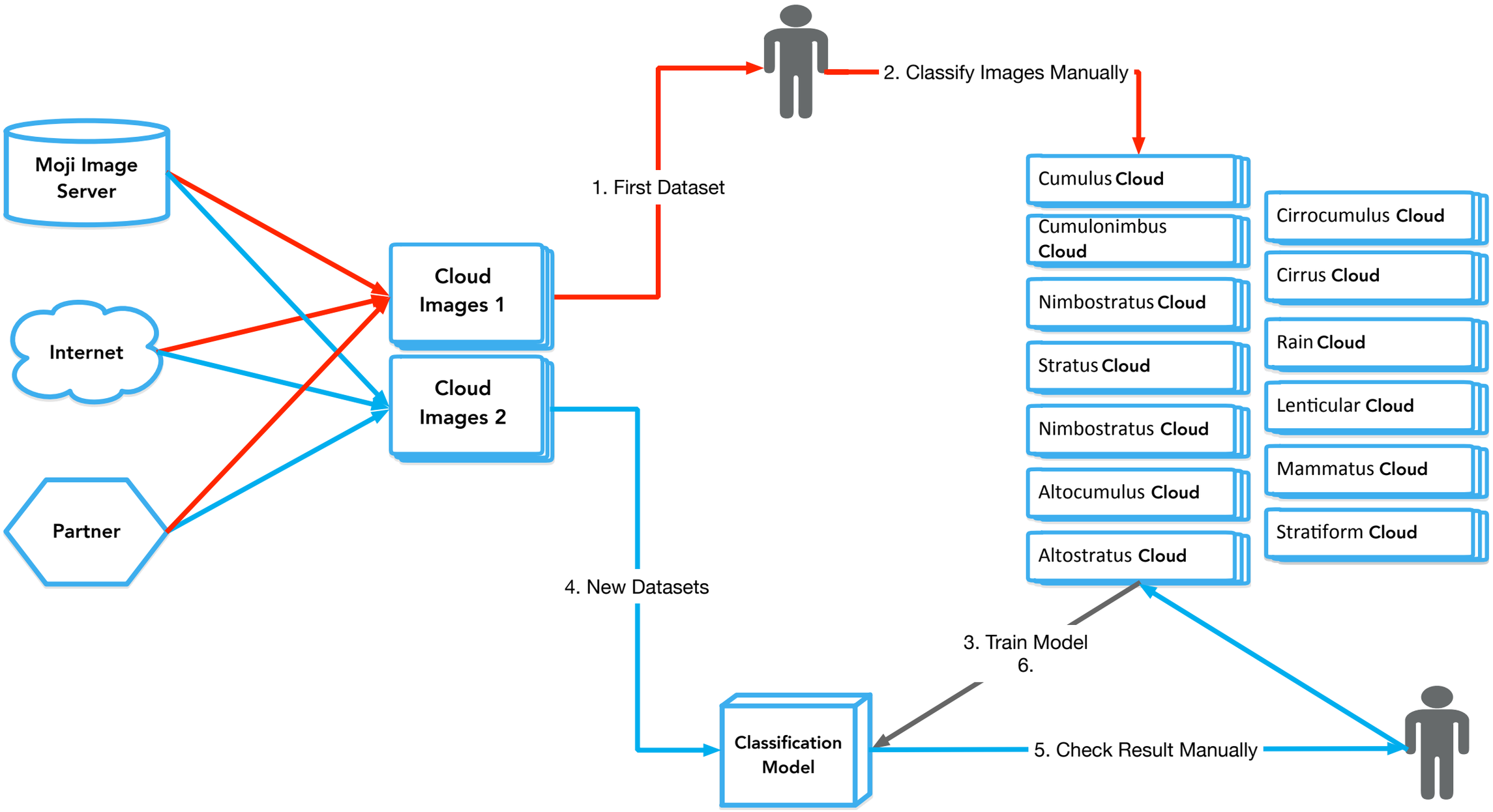
- Infrequent dataset for cloud classification
- Less public hand-classified network dataset
- Hard to classify by non-professional people

- Image data collection by engineers
- Cloud image tagging by meteorologist

In the first phase, our meteorologist spent 1 month tagging nearly 5000 cloud images, including the type and the position.

However, there are 13 categories, with an average of fewer than 500 in each category. For machine learning and training, the number of images is still small, and it costs a lot to continue tagging.

Data Preparation for Cloud Classification Model



Training for Cloud Classification Model

Difficulties in step from 0 to 1：

- Infrequent training data
- Too much cloud categories
- Some categories have very small differences

Simple network for rapid training and testing in early stage
CNN Network (10 layers)
Network parameters Xavier initialization
Training data for each type of 500 images (less than the number of categories on the image were randomly intercepted or repeat complement)

Cyclically repeated training and check the prediction results, the similar types of cloud images were merged, the different types of cloud split, the final establishment of 13 categories of clouds as a training target

However, the training effect is still not satisfactory, Top1 accuracy can only reach 40-50%

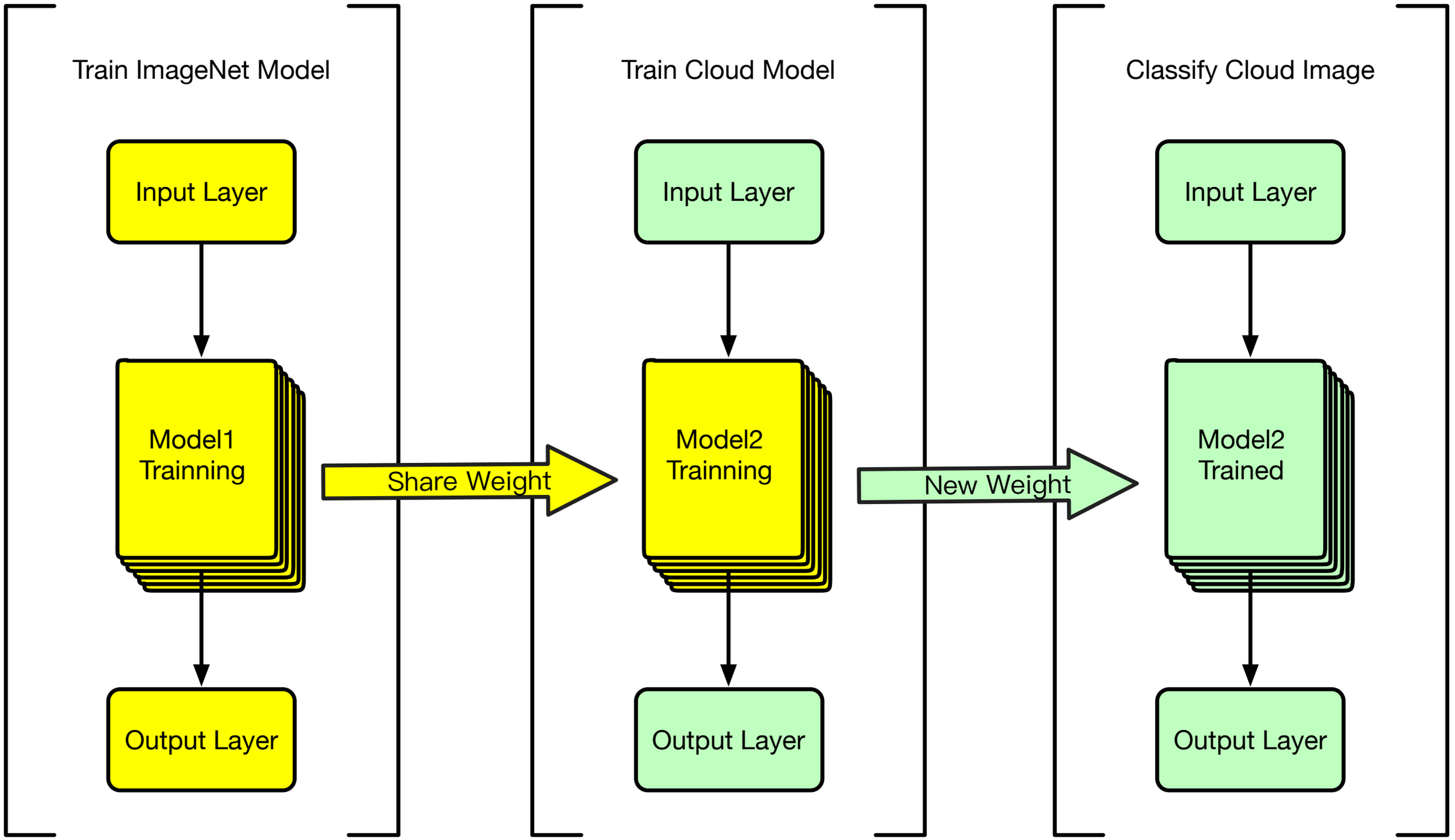
Training for Cloud Classification Model

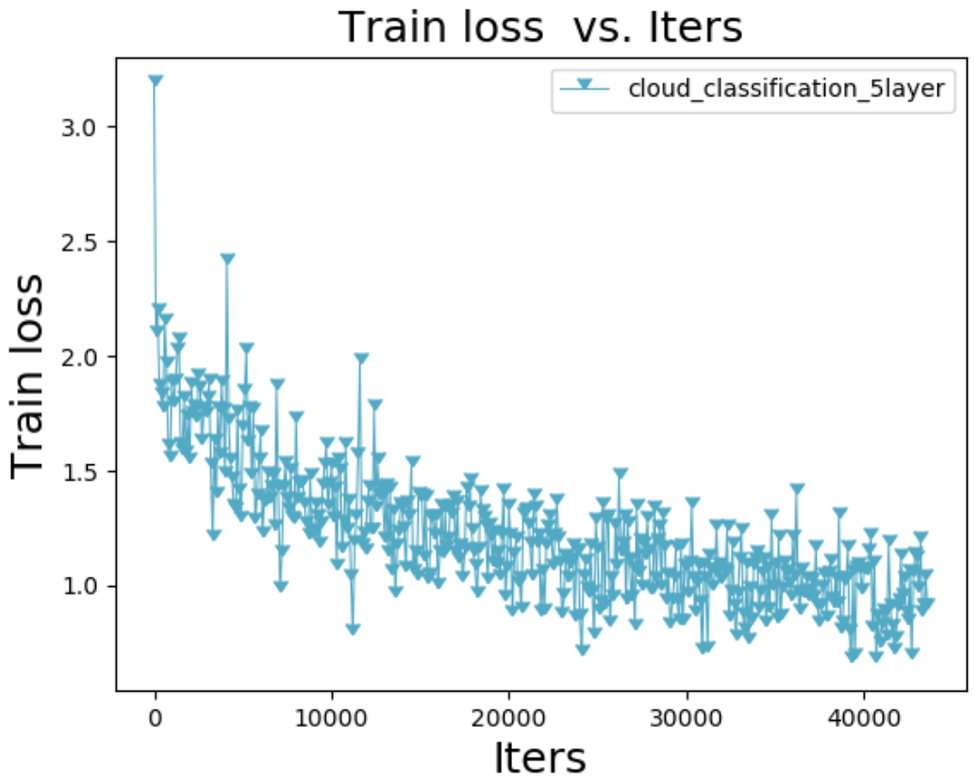
Transfer learning

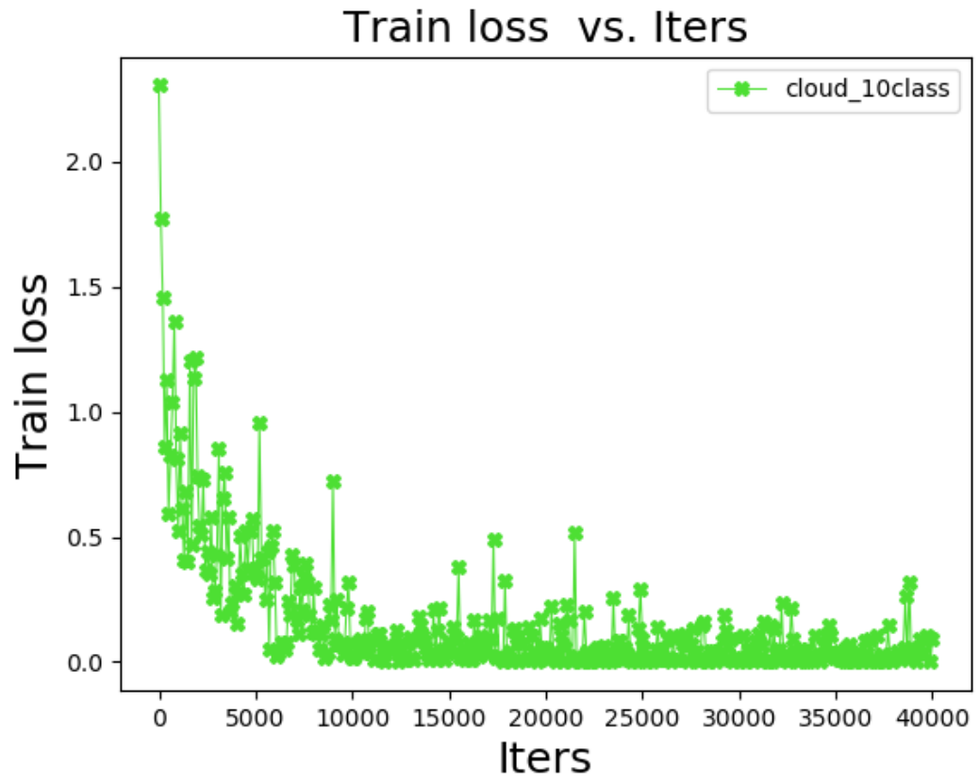
For the short time, we can not mark large quantities of available training samples, we adopt the transfer learning, improve the model accuracy, and achieve a better prediction level

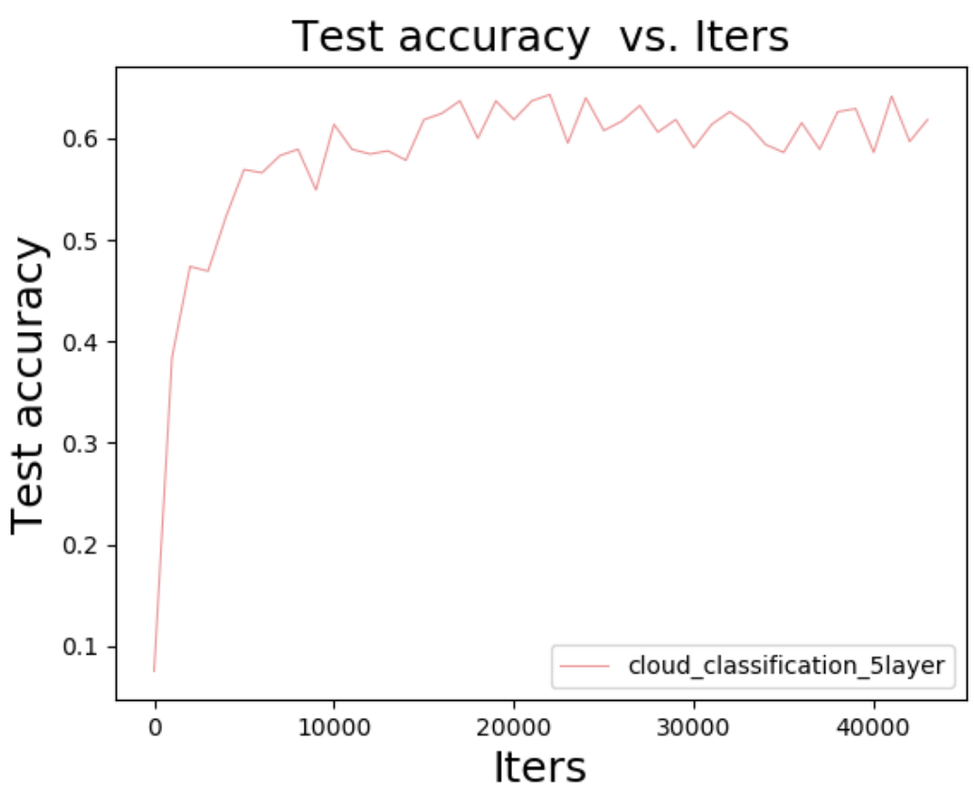
We used the Resnet network trained for the IMAGENET Large Scale Visual Recognition Challenge data set (containing the 1000 categories and 1.2 million images) as a pre-training model

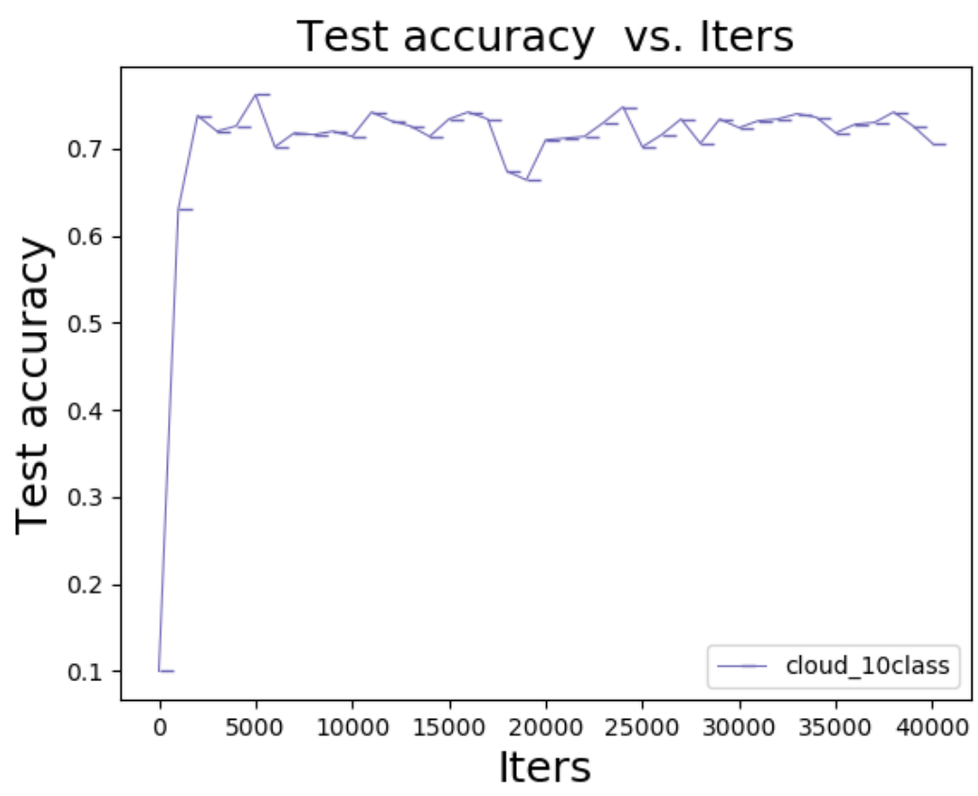
Training for Cloud Classification Model

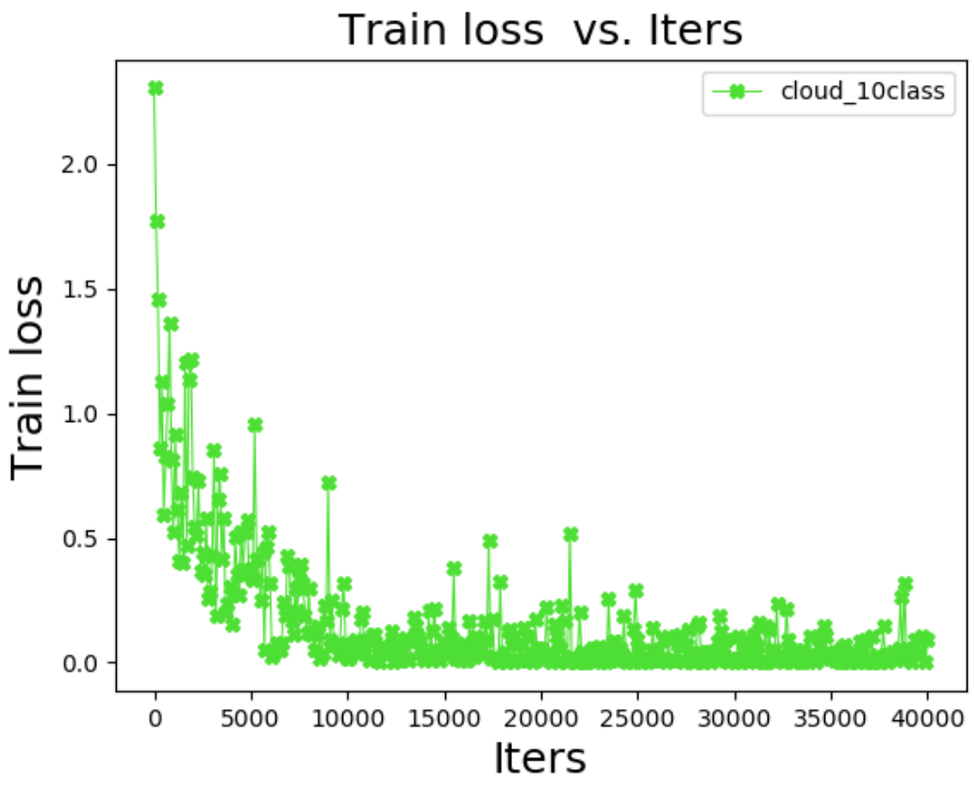


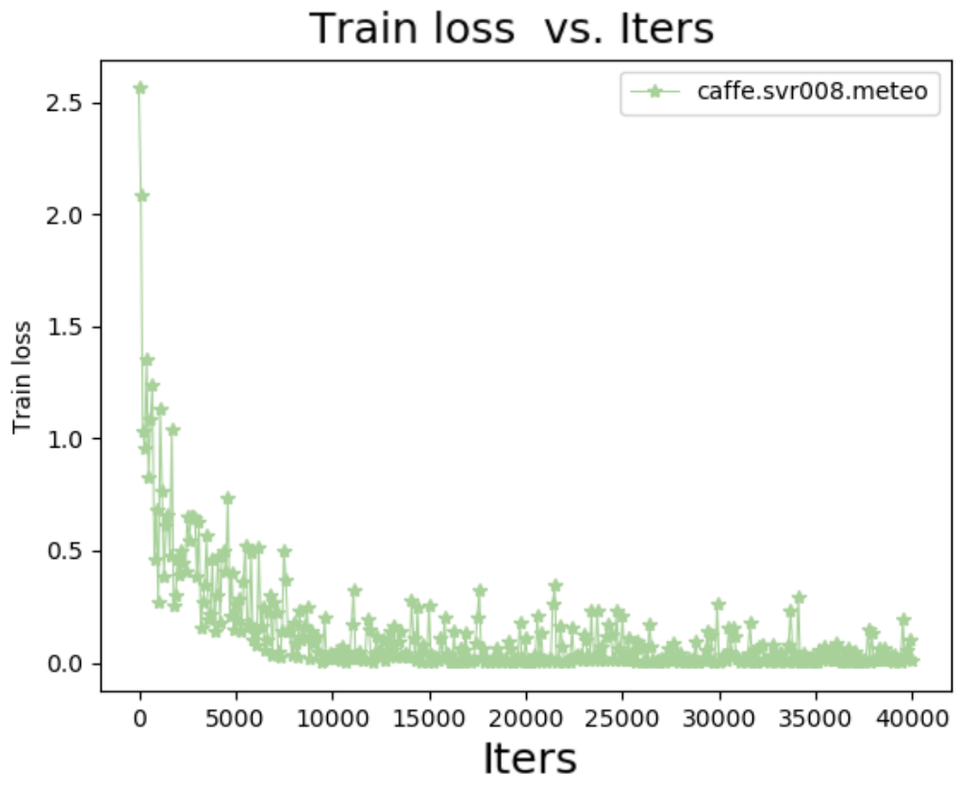

Train loss vs. Iters
cloud_classification_5layer
Train loss
Iters
3.0
2.5
2.0
1.5
1.0
0
10000
20000
30000
40000

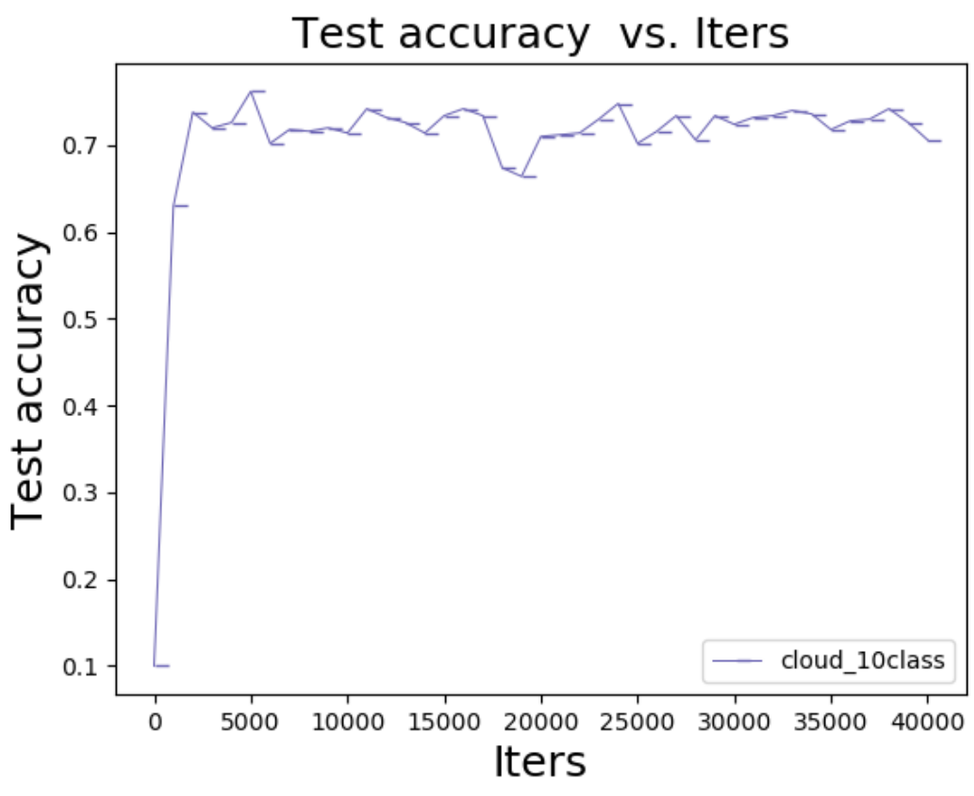

Train loss vs. Iters
cloud_10class
Train loss
Iters
2.0
1.5
1.0
0.5
0.0
0
5000
10000
15000
20000
25000
30000
35000
40000

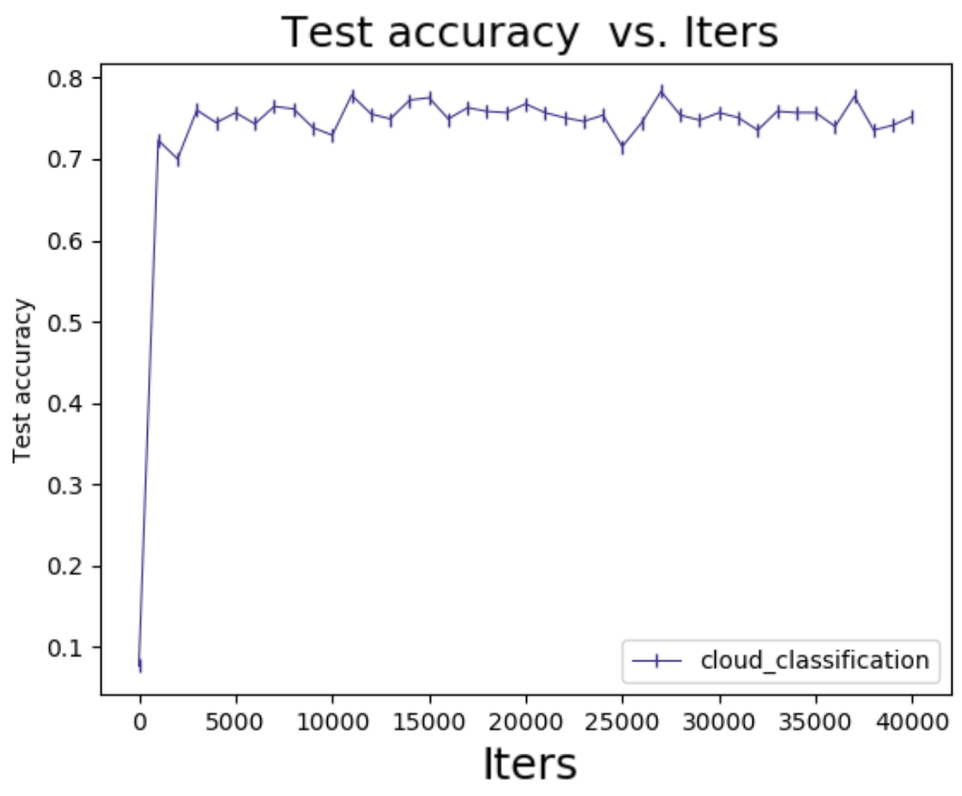

Test accuracy vs. Iters
cloud_classification_5layer
Test accuracy
Iters
0.6
0.5
0.4
0.3
0.2
0.1
0
10000
20000
30000
40000


Test accuracy vs. Iters
cloud_10class
Test accuracy
Iters
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5000
10000
15000
20000
25000
30000
35000
40000

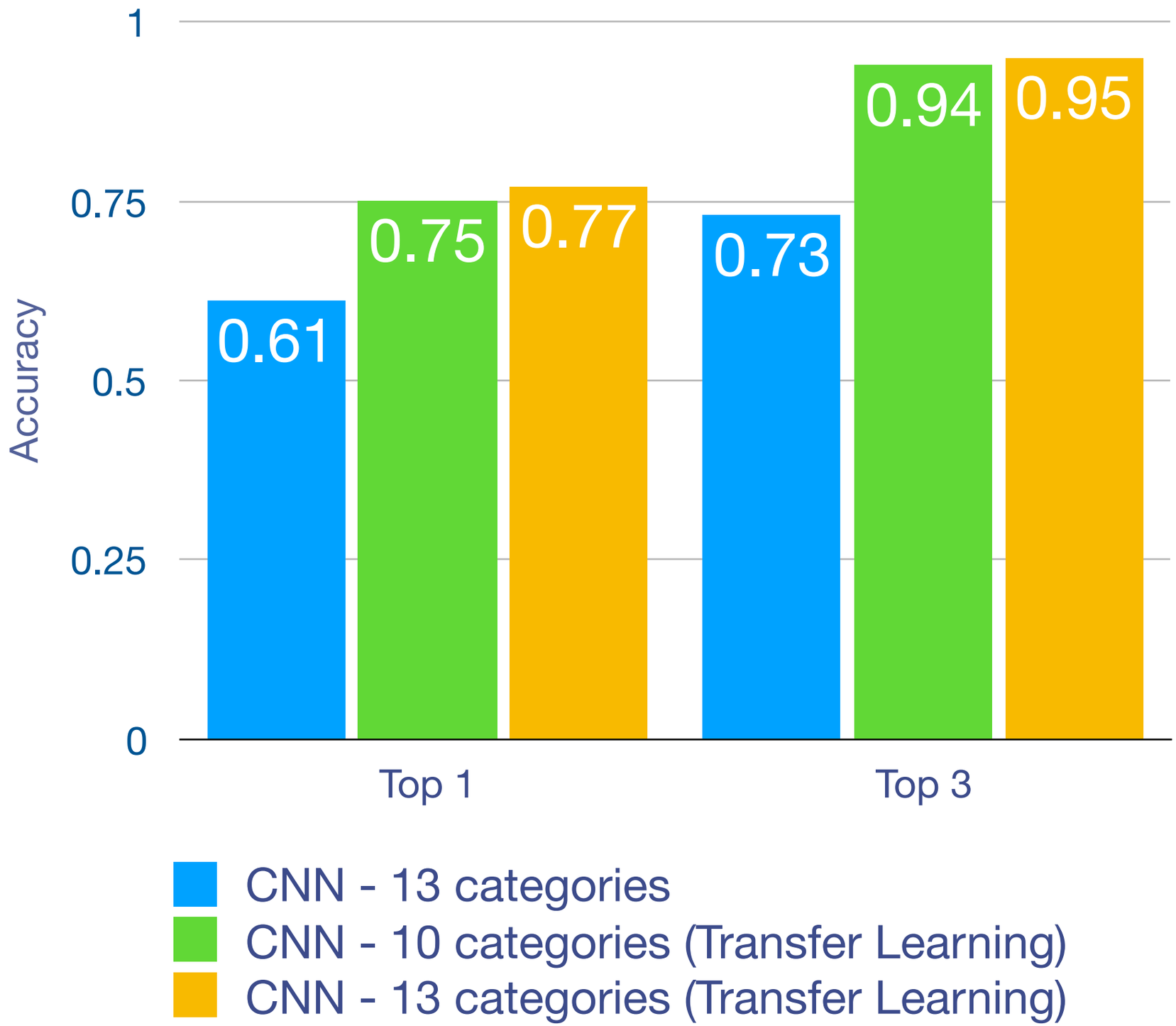

Train loss vs. Iters
cloud_10class
Train loss
Iters
0.0
0.5
1.0
1.5
2.0
0
5000
10000
15000
20000
25000
30000
35000
40000


Train loss vs. Iters
caffe.svr008.meteo
Train loss
Iters
0.0
0.5
1.0
1.5
2.0
2.5
0
5000
10000
15000
20000
25000
30000
35000
40000


Test accuracy vs. Iters
cloud_10class
Test accuracy
Iters
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
5000
10000
15000
20000
25000
30000
35000
40000

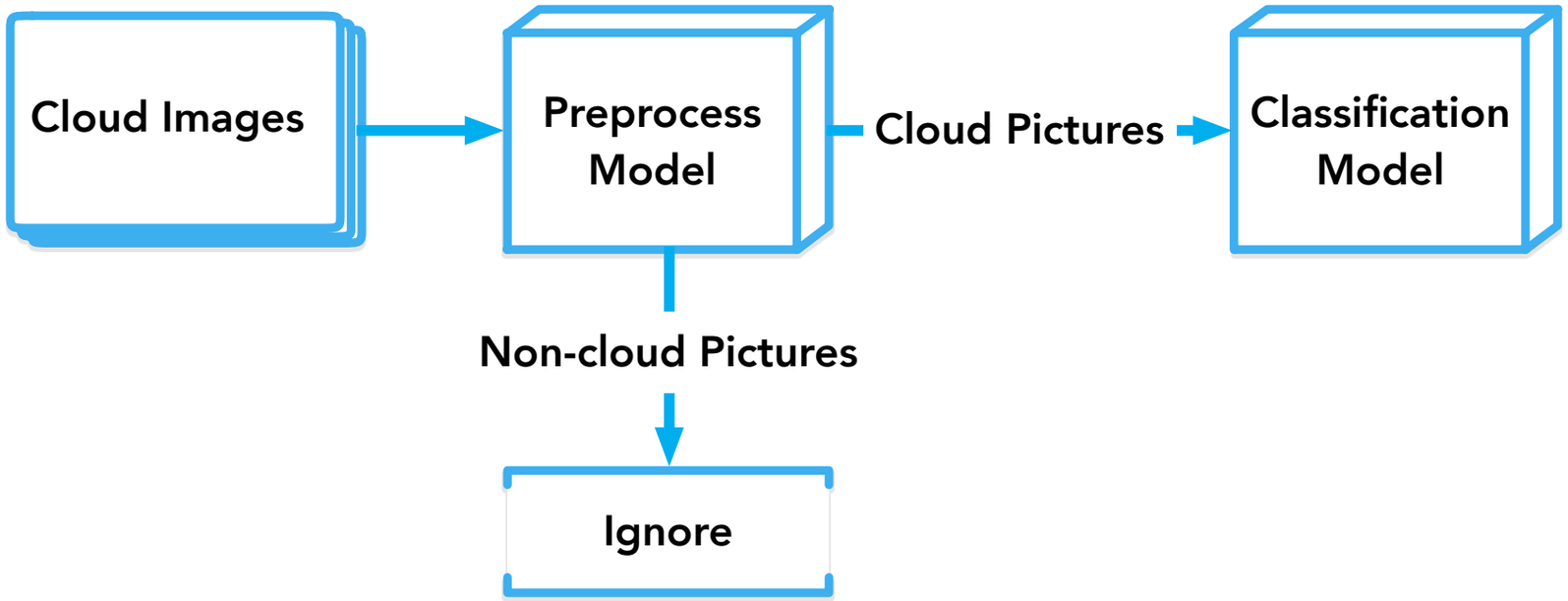

Test accuracy vs. Iters
cloud_classification
Test accuracy
Iters
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
5000
10000
15000
20000
25000
30000
35000
40000

Model Accuracy Comparison



Training for Cloud Classification Model

Only cloud classification model can not be deployed in our online environment:
Users of our application upload all kinds of images they took



We need a preprocess model to distinguish the non-cloud images which will be filtered by the pre-model.
The pre-model is simple, only need to predict two categories. We use the cloud images we prepared for the cloud model and some non-cloud images that our app users upload to train the pre-model and get a accuracy of 95% easily.

Application scenarios and Future Work

- Application on consumer sides and business sides
- Post for product on Exhibit Hall 3(ACC) Tuesday 9th Jan 2018
- Expand the image dataset Jan - June 2018
- Improve accuracy with cellphone barometers，satellite datas Mar - Sep 2018
- Expand recognition type，like：sunny、hail、rain、snow Sep-Dec 2018