Cdiscount's Image Classification

MinWoo Byeon

Cdiscount

- Cdiscount.com is the largest non-food e-commerce company in France.
-

Overview

- https://www.kaggle.com/c/cdiscount-image-classification-challenge
-

Let’s Go

Environments

- Python3 (with PyCharm)
- MXNet 0.12.0
- Pascal based 8-GPUs Machine

Dataset

Dataset Distribution

Dataset Cleaning

-

Models

	Batch Size	GPUs	Memory Usage (per GPU)	Speed (8-GPU, Sync)
ResNext-50	512	8		
ResNext-101-32x4d	512	8	12,461MiB	579.86 / sec
ResNext-101-64x4d	512	8	18,185 MiB	338.29 / sec
SE-ResNext-101-32x4d	512	8	14,057 MiB	411.55 / sec
SE-ResNext-101-64x4d	512	8	19,805 MiB	274.04 / sec
DPNs-92	512	8		
DPNs-98	512	8	16,367 MiB	380.22 / sec
DPNs-98 (224)	384	8	17,487 MiB	269.09 / sec
DPNs-107	256	8	17,707 MiB	246.06 / sec

Weight Initialization

- MSRA PReLu
  - https://arxiv.org/abs/1502.01852
  - Used in Squeeze-and-Excitation Networks
-

Accuracy

		Pre-trained	Training		Testing (with flip)	
	Models	**Valid Acc.**	**Train Acc.**	**Valid Acc.**	**Valid Acc.**	**L.B**
M1	**ResNext-50**					
M2	**ResNext-101-32x4d**					
M3	**ResNext-101-64x4d**				0.758480	
M4	**DPNs-92**					
M5	**DPNs-98**				0.756047	

Optimizer

- SGD
- NAG
- Adam
- NAdam

Augmentation

- ResNet-34, Batch=512, Opt=Adam, Base-LR=0.0005, Epoch=15

Input Size	Option	Local Train (Top-1)	Local Validation (Top-1)
180 x 180	No Aug.	0.862802	0.688090
180 x 180	Flip	**0.787349**	**0.692491**
180 x 180	Flip + HSL	0.774842	0.690422
224 x 224 (NN)	Flip	**0.789405**	**0.695116**

Image Size

- ResNet, stride//2 만큼 pad

Input Size	conv	stage1	stage2	stage3	stage4	Local Train (Top-1)	Local Validation (Top-1)
160 x 160	80 x 80	40 x 40	20 x 20	10 x 10	5 x 5		
180 x 180	90 x 90	45 x 45	23 x 23	12 x 12	6 x 6		
192 x 192	96 x 96	48 x 48	24 x 24	12 x 12	6 x 6		
224 x 224	112 x 112	56 x 56	28 x 28	14 x 14	7 x 7		

Label Smoothing

- ResNet-34, Batch=512, Opt=Adam, Base-LR=0.0005, Epoch=15

Smooth alpha	Local Train (Top-1)	Local Validation (Top-1)
0.0	**0.785019**	0.691917
0.1	0.779468	**0.694035**
0.2	0.770593	0.692766

GAP + Dropout

- ResNet-34, Batch=512, Opt=NAdam, Base-LR=0.0005, Epoch=15

Dropout Ratio	Local Train (Top-1)	Local Validation (Top-1)
0.0	**0.788686**	**0.693604**
0.2	0.743764	0.689954

Multi-View Testing

-

MD5 Dictionary

- ResNext-101-64x4d (epoch 15, original+flip)

Method	Local Test (geo)		Leader Board
None			0.75671 (arith-maen)
Unique	0.763840		0.76025 (arith-mean)
Majority			0.75308 (geo-mean)
L1			0.76205 (geo-mean)
L2			0.76206 (geo-mean)
Softmax			

Ensemble

		Training			Testing (with flip)	
	Models	**Epochs**	**Train Acc.**	**Valid Acc.**	**Valid Acc.**	**L.B**
M2	**ResNext-101-32x4d**	11				0.74718
M3	**ResNext-101-64x4d**	15			0.758480	0.75671
M5	**DPNs-98**	15			0.756047	0.75470
	M3 + M5	-	-	-	0.770675	0.76972

Ensemble Methods

-

		Local Validation		Public Leader Board	
Model	View	Arithmetic Mean	Geometric Mean	Arithmetic Mean	Geometric Mean
3-Ensemble	6			0.77784	0.77918
6-Ensemble	12			0.78268	0.77468

Ensemble Methods

- Ensemble 4 models

Method	Local Validation (Top-1)	Desc
model arith, img arith	0.782746	
model arith, img geo (1 imgs)	0.781875	
model arith, img geo (2 imgs)	0.782525	
model arith, img geo (3 imgs)	0.782848	
model arith, img geo (4 imgs)	**0.783498**	
model geo, img geo	**0.785558**	

Ensemble Models

Model	Model Validation (Product Level)	Ensemble Accuracy
dpn107e19	0.767340	0.767340 (+0.0)
seresnext101-c64-15	0.763164	0.778684 (+0.011344)
dpn131e15	0.762358	0.782644 (+0.015304)
resnext101-c64-e17	0.757778	0.780899
densenet161-e15	0.755631	0.782757

MD5 Dictionary

- ResNext-101-64x4d (epoch 15, original+flip)

Method	Arithmetic Mean	Geometric Mean
None	0.758480 (LB: 0.75671)	0.761119
Unique	0.761724 (LB: 0.76025)	**0.763840**
Majority	0.757597	0.754174 (LB: 0.75308)
L1	0.761747	0.763490 (LB: 0.76205)
L2	0.761574	0.763490 (LB: 0.76206)
Softmax	0.753390	0.756910

Large Size Prediction

ILSVRC 2017 Classification Competition

A. The number of first 1×1 convolutional channels for each bottleneck building block was halved to reduce the computation cost of the network with a minimal decrease in performance
B. The first 7 × 7 convolutional layer was replaced with three consecutive 3×3 convolutional layers.
C. The down-sampling projection 1 × 1 with stride-2 convolution was replaced with a 3 × 3 stride-2 convolution to preserve information

ILSVRC 2017 Classification Competition

D. A dropout layer (with a drop ratio of 0.2) was inserted before the classifier layer to prevent over-fitting
E. Label-smoothing regularisation was used during training.
F. The parameters of all BN layers were frozen for the last few training epochs to ensure consistency between training and testing

https://arxiv.org/pdf/1709.01507.pdf

Sub-category model

- Library Category
  - All Category: 0.511904
  - Library category: 0.530334
-

Batch Prediction

- ZMQ + OpenCV

Mistakes

- 평균 잘못 구함
- RGB 채널 순서
- object reference

```
>>> import numpy as np
>>> a = np.array([1,2,3])
>>> a
array([1, 2, 3])
>>> b = a
>>> b
array([1, 2, 3])
>>> a
array([1, 2, 3])
>>> b += a
>>> a
array([2, 4, 6])
>>> b
array([2, 4, 6])
```

Leaderboard

-

Time

-

Additional Methods

Large Margin Softmax

- https://arxiv.org/pdf/1612.02295.pdf
- https://github.com/luoyetx/mx-lsoftmax
-

Mean-Max Pooling

OCR

- Book, Album, ...
- Tesseract OCR
  - https://github.com/tesseract-ocr/tesseract
- Too Bad...

B-CNN

Multi-input CNN

- 12-channel(4 images) input
-

References

-