iMet Collection 2019 - FGVC6

Kirill Brodt

Task description

- Multi-label classification on images of art (minimum size of short side is 300)
- 1103 classes
  - 398 cultures (french, italian, american, …)
  - 705 tags (men, women, flowers, ...)
- 110K train data
  - stage I - 7.5K
  - stage II - 39K
- Kernel only
  - 9h GPU
  - pretrain is allowed

Evaluation

F2-score averaged by classes.

$$\frac{(1+\beta^2)pr}{\beta^2 p + r} \text{ where } p = \frac{tp}{tp+fp}, \quad r = \frac{tp}{tp+fn}, \quad \beta = 2.$$

F2 score weights recall higher then precision


1034;194;671
79;954
121;433
147;283;647;671;780
1046;335;949
161;487
1064;189;408;616;835;994
1092;147
79;925
147;552

First and last submit

- PNasNetLarge5 + float16
- RandomResizedCrop 331x331 with total square resize between 0.4 and 1.0 + Horizontal flips
- 10 folds (but averaged only over 9)
- 32 batch size
- 10 epochs with scheduler (manual), drop lr by 2 times on 7 and 9 epochs
- 2xTTA (Original + Horizontal flip + RandomResizedCrops)
- Search over threshold
- Divide predictions on maximum value of predictions

CV 0.615 -> LB Stage I 0.64 (65) -> LB Stage II 0.636 (47)

Efforts to improve baseline

- Se-ResNeXt doesn’t work in MXNet!
- ResNet, ResNext - likes
- other image sizes 224, 288, 300
- individual thresholds

No chance

Final (446)

g: [0.672 - 0.654]

s: [0.653 - 0.635]

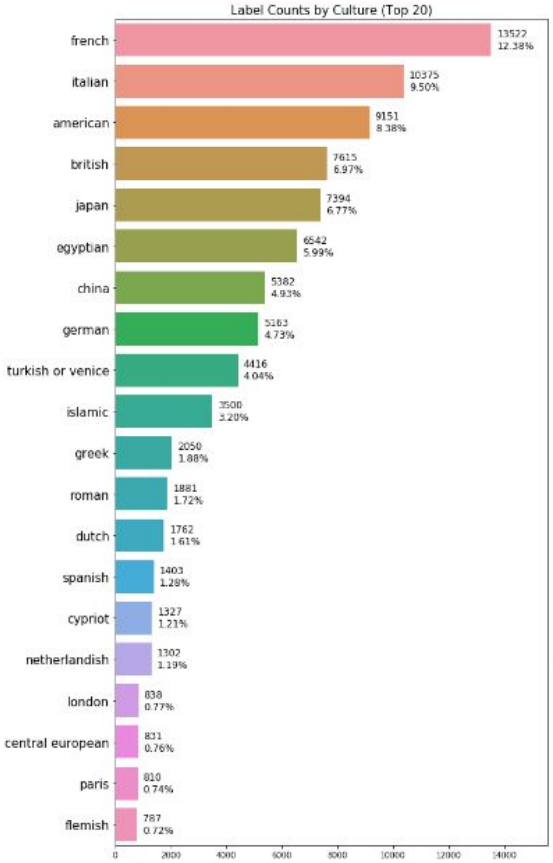
b: [0.634 - 0.610]

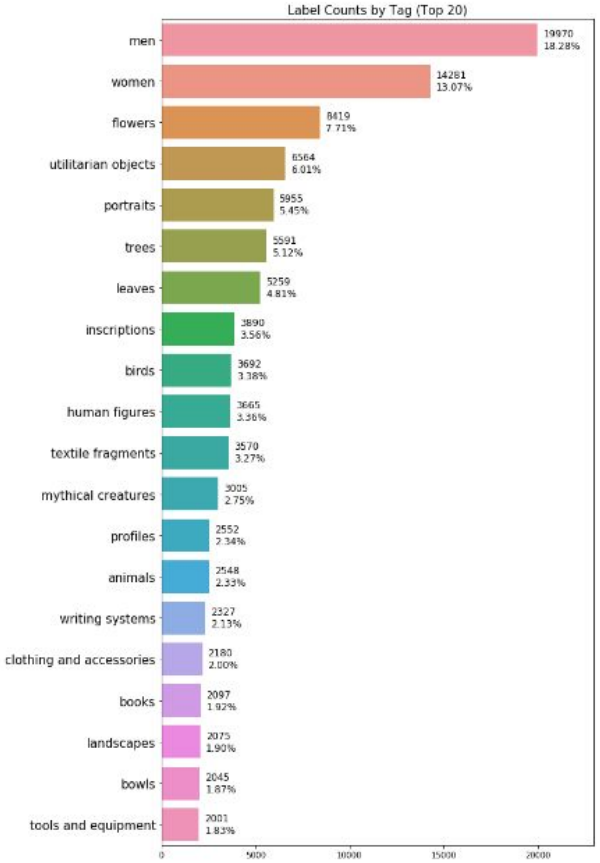
1	▲ 400	[ods.ai] Konstantin Gavrilchik	0.672
2	▲ 425	ʅ(ᴖ◡ᴖ)ㄱ—☆*:・゜	0.667
3	▲ 426	[ods.ai] Ilya Kibardin	0.664
4	▲ 411	pudae	0.663
5	▲ 349	[ods.ai] n01z3	0.662
6	▲ 332	みんなをStarlightしちゃいます	0.660
7	▲ 419	Kaggler-JP&CN	0.659
8	▲ 425	X5, Best Russian Company	0.658
9	▲ 321	Appian	0.658
10	▲ 422	Alchemists' Creed: Obey the ...	0.655
11	▲ 292	頼む！！！！！	0.654

Difficulties: Super long images

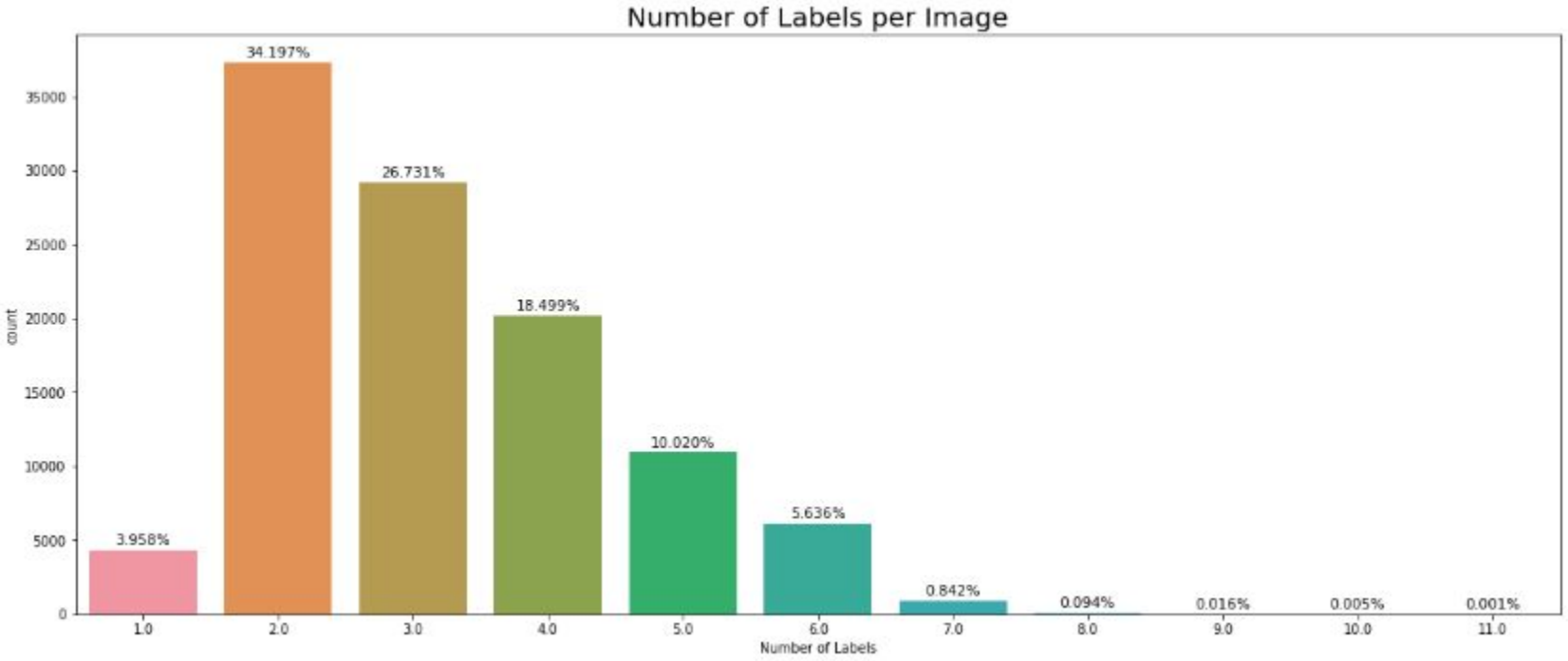
Difficulties: Highly imbalanced

Most images have 2-5 labels (90%)





Difficulties: Highly imbalanced



Mother of images

cultures: coromandel coast, turkish or venice

tags: birds, carriages, dogs, flowers, horses, men, rabbits, textiles, women

1st place solution (LB 0.672)

6x1080Ti with 36 cores and 120 RAM

Stage 0. The same part for all stages:

- SENet154, PNasNet-5, SE-ResNeXt101
- 5 folds
- Horizontal flip, One of Random Brightness or Contrarst, ShiftScaleRotate, Gaussian noise
- Crop if needed (2*SIZE) + resize (SIZE) [SIZE=331 for PNasNet-5 and 320 for others]
    - crop 600x600 from 500x300 -> 500x300, from 500x900 -> 500x600
- 1TTA (Original + Horizontal flip)
- base lr 0.005, 15 epoch with manual scheduler dropping lr by 5 times

1st place solution (LB 0.672)

Stage 1. Training the zoo:

- Focal loss
- batch sampling with logarithmic weights (log of probability of classes in dataset)
- Batch size 1000-1500 (20 accumulations)

Stage 2. Filtering predictions:

- Drop samples with very high error between OOF predictions and labels (noisy)
- Focal loss
- Hard negative mining (sample 5% of hardest samples each epoch)

Retrain from scratch

1st place solution (LB 0.672)

Stage 3. Pseudo labeling:

- Focal loss
- add most confident predictions (highest mean(abs(proba - 0.5)))

Stage 4. Culture and tags separately:

- tags less noisy than cultures
  - train only for tags (Focal to Cross-entropy loss)

1st place solution (LB 0.672)

Stage 5. Second-level model:

- Binary classifier: this class relates to this image (0 or 1)
  - dataset length became 1103*(#imgs)

Hints:

- different thresholds for culture and tags

10-15 days of training

Didn’t submit all these ensembles to the Stage II (faced kernel limit :facepalm:)

2nd place solution (LB 0.667)

- 2x1080Ti
- SE-ResNeXt-50 and SE-DenseNet-161 with partial convs
- 6 folds
- Crop 640, resize 320, horizontal flip, rotate, gamma, brightness, contrast
- 1TTA (original + horizontal flip)
- Train with freezed encoder -> whole network -> tags only
- 512 batch size (batch accumulation)
- AdamW with weight decay 0.01, lr = 1e-4 -> SGD lr = 1e-3 + Cosine scheduler with warmup.
- Label smoothing + MixUp
- Separate thresholds for culture and tags
- Cleaning (high error) + pseudo labeling (high confidence)

6th place solution (LB 0.66)

- RandomCrop 320 + pad if needed
- MixUp + random erasing
- 40 folds
  - SE-ResNeXt-101 (10 folds)
  - InceptionResNet-v2 (5 folds)
  - PNasNet-5 (5 folds)
  - 20 models
- 1TTA (original + hflip)
- Threshold for each image: proba > max_proba / 7