iWildCam 2019 - FGVC6

Categorize animals in the wild

Abuduweili
Xingchen Tao
Xin Wu

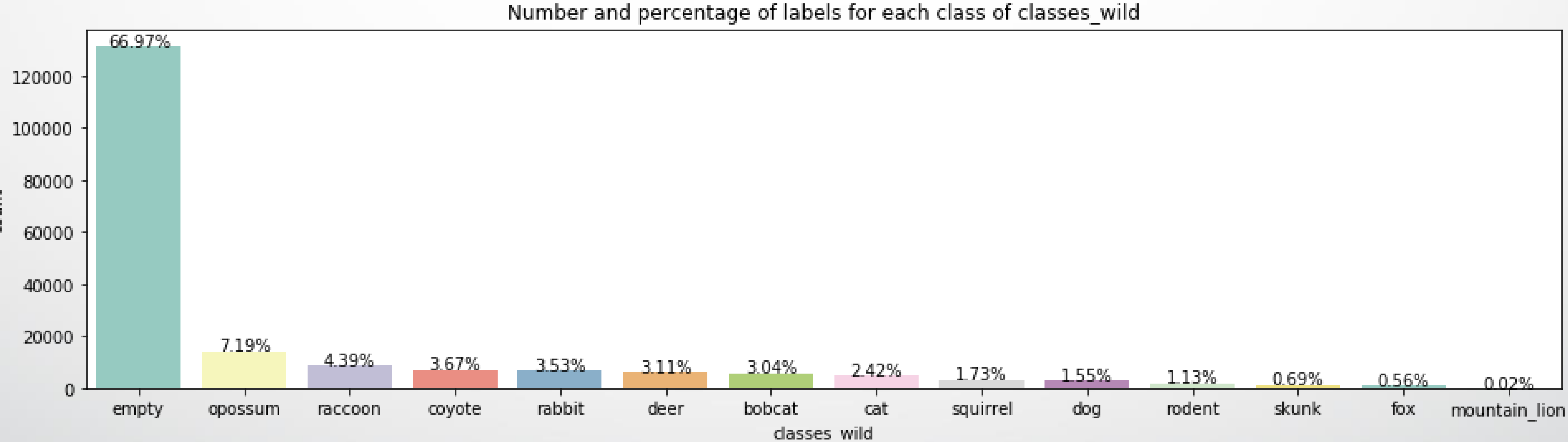
2019-06-10

CONTENTS

Challenges

Challenges on categories

- the training data and test data are from different regions. The American Southwest the American Northwest.
- The species seen in each region overlap, but are not identical. The total number of label is 23， the number of training data label is 14.



Challenges on images

- **Illumination:** Images can be poorly illuminated, especially at night. The example below contains a skunk to the center left of the frame.
- **Motion Blur:** The shutter speed of the camera is not fast enough to eliminate motion blur, so animals are sometimes blurry. The example contains a blurred coyote.
- **Small ROI:** Some animals are small or far from the camera, and can be difficult to spot even for humans. The example image has a mouse on a brance to the center right of the frame.
- **Occlusion:** Animals can be occluded by vegetation or the edge of the frame. This example shows a location where weeds grew in front of the camera, obscuring the view.
- **Perspective:** Sometimes animals come very close to the camera, causing a forced perspective.
- **Weather Conditions:** Poor weather, including rain, snow, or dust, can obstruct the lens and cause false triggers.
- **Camera Malfunctions:** Sometimes the camera malfunctions, causing strange discolorations.
- **Temporal Changes:** At any given location, the background changes over time as the seasons change. Below, you can see a single loction at three different points in time.
- **Non-Animal Variability:** What causes the non-animal images to trigger varies based on location.

Challenges on images

Camera trap data provides several challenges that can make it difficult to achieve accurate results.

Dataset

Smaller CCT images 27GB

iNat Idaho images and annotations 7.2GB zipped （supplementary dataset）

Method

- Animal Detection
- Image Augmentation
- Apply transfer learning on models pretrained
- Advanced Tricks
- Ensemble learning

1. Detection

◆ The Camera Trap Animal Detection Model is a tensorflow Faster-RCNN model with Inception-Resnet-v2 backbone and atrous convolution.

2. Image Preprocessing and Augmentation

◆ Image preprocessing

- White-balance： To solve color reduction and tone processing problems
- Histogram equalization: This method is useful for images that are too bright or too dark in the background and foreground, but it also may increase the contrast of background noise and reduce the contrast of useful signals
- Image denoising: reducing noise in digital image
- Image resizing: reducing images size

◆ Data enhancement

- Traditional methods: random cropping, rotation, translation, increasing brightness, increasing gaussian noise
- Histogram equalization
- Grayscale

◆ Data balancing

2. Image Preprocessing and Augmentation

Original Image

White Balance
+ Histogram equalization

2. Image Preprocessing and Augmentation

Image Augmentaion

Original Image

Blur

Rotate & Shift

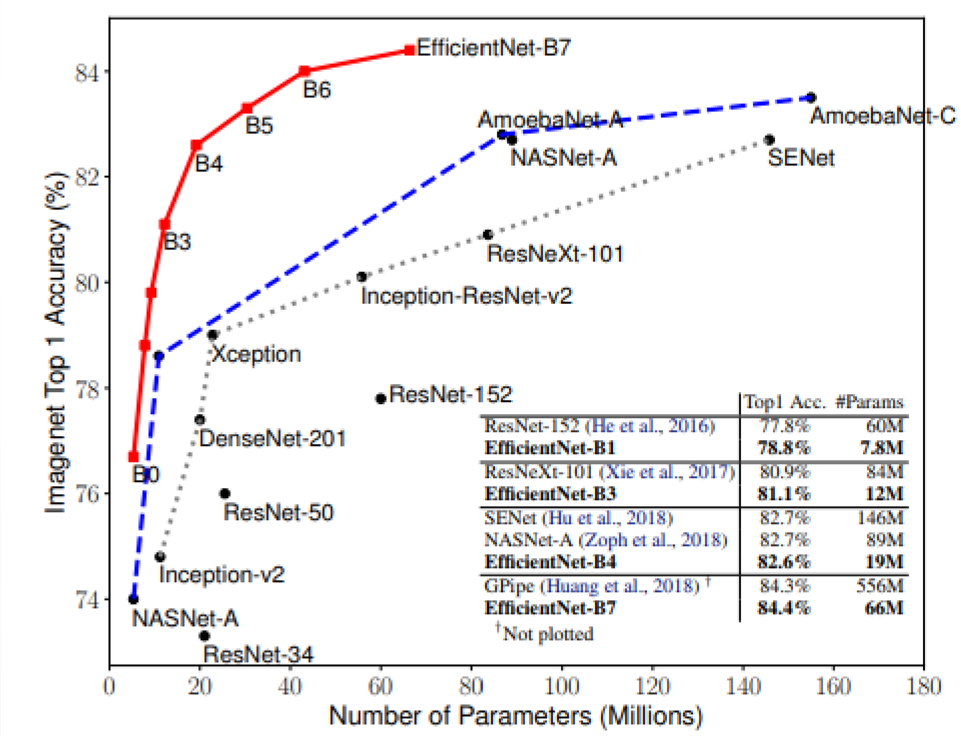
To Gray

3. Pretrained Models for transfer learning

- Resnet50
- Resnet101
- Efficientnet-b0
- Efficientnet-b3
- Vgg16
- Vgg19
- Densenet121
- Xception

3. Pretrained Models for transfer learning

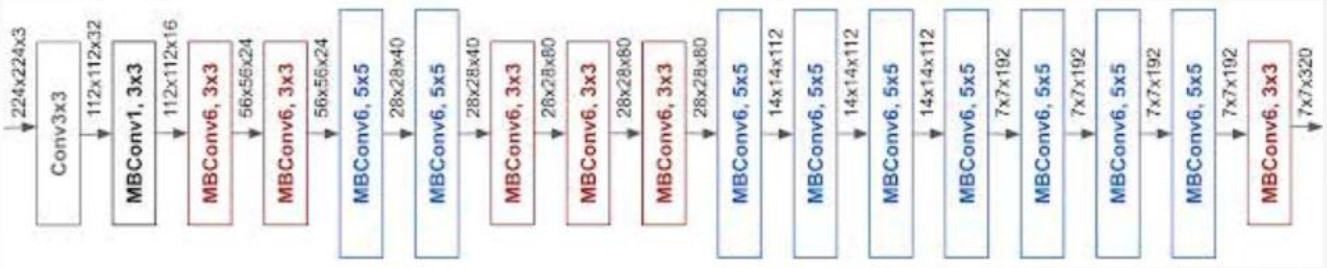
Efficientnet



Tan, Mingxing , and Q. V. Le . "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." (2019).

3. Pretrained Models for transfer learning

Efficientnet-b0



Model Scaling

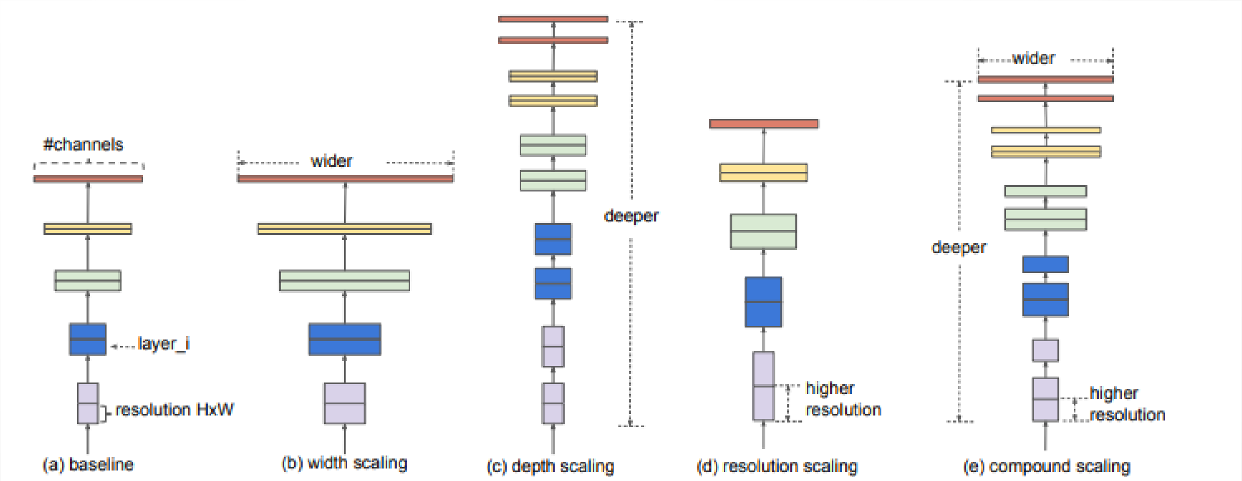


Figure 2. **Model Scaling.** (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

Tan, Mingxing , and Q. V. Le . "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." (2019).

4. Tricks

- **Label smoothing：** it can effectively suppress the over-fitting phenomenon when calculating the loss value by “softening” the traditional one-hot type label
- **Cut out：** a simple regularization technique of randomly masking out square regions of input during training, can be used to improve the robustness and to avoid overfitting
- **Mixup：** a large deep neural network may be sensitivity to adversarial and memorization, mixup is a simple learning principle to alleviate these issues.

4. Tricks

◆ Label smoothing

- Modify your one-hot targets with equation (original notation from paper):

that

$$q'(k) = (1 - \epsilon)\delta_{k,y} + \frac{\epsilon}{K}.$$

We refer to this change in ground-truth label distribution as *label-smoothing regularization*, or LSR.

K - is the number of classes
epsilon is a hyperparameter, usually 0.1

Szegedy, Christian , et al. " [IEEE 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Las Vegas, NV, USA (2016.6.27-2016.6.30)] 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Rethinking the Inception Architecture for Computer Vision." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* IEEE, 2016:2818-2826.

4. Tricks

◆ Cut out

Randomly masking out square regions of input during training

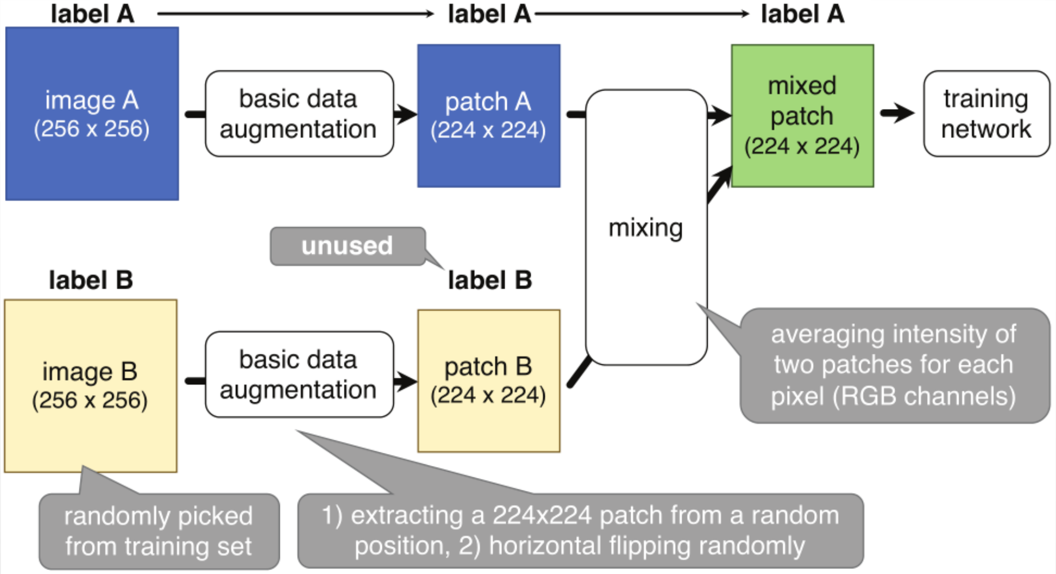
Devries, Terrance , and G. W. Taylor . "Improved Regularization of Convolutional Neural Networks with Cutout." (2017).

4. Tricks

◆ Mixup

Contribution Motivated by these issues, we introduce a simple and data-agnostic data augmentation routine, termed *mixup* (Section 2). In a nutshell, *mixup* constructs virtual training examples

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j, \qquad \text{where } x_i, x_j \text{ are raw input vectors}$$
$$\tilde{y} = \lambda y_i + (1-\lambda)y_j, \qquad \text{where } y_i, y_j \text{ are one-hot label encodings}$$



Inoue, Hiroshi . "Data Augmentation by Pairing Samples for Images Classification." (2018).

5. Ensemble learning

◆ Averaging the probability of following 9 predictions:

3 models * 3 different test image preprocessing = 9 prediction results

3 models :

(1) Resnet 101, (2) Efficientnet-b0, (3) Efficientnet-b3

3 data preprocessing for test image:

(1) original test image, (2) CLAHE, (3) to Gray

5. Ensemble learning

- **Training**
  - Data enhancement
  - 3 pretrained models: resnet101, efficientnet-b0, efficientnet-b3
  - Optimizing strategies
- **Testing**
  - Test data resized to 128*128
  - Test data based on Histogram equalization and resized to 128*128
  - Test data based on grayscale and resized to 128*128
- **Integrate 9 prediction**
  - F1 score finally reached 0.228
  - Rank the tenth

Result

Private leaderboard: **top 3 % (10/344)**

#	∆pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	epsilon			0.399	75	3d
2	▲1	Rb			0.362	2	3d
3	▼1	Hayder Yousif			0.340	29	3d
4	▲3	Local Minima			0.319	18	3d
5	▲4	rewolfiluac			0.303	65	3d
6	▼2	xiang			0.300	30	10d
7	▼1	DeepBlueAI in TH			0.300	30	3d
8	—	Kyrie			0.287	8	19d
9	▼4	ARF MIARFID			0.286	68	22d
10	▲2	JustVeg			0.228	8	3d
11	▼1	Hugo Touvron			0.227	8	4d

1. Best single model:
Efficientnet_b0 with image augmentation and tricks just mentioned
Efficientnet_b0 Private leaderboard score: 0.225

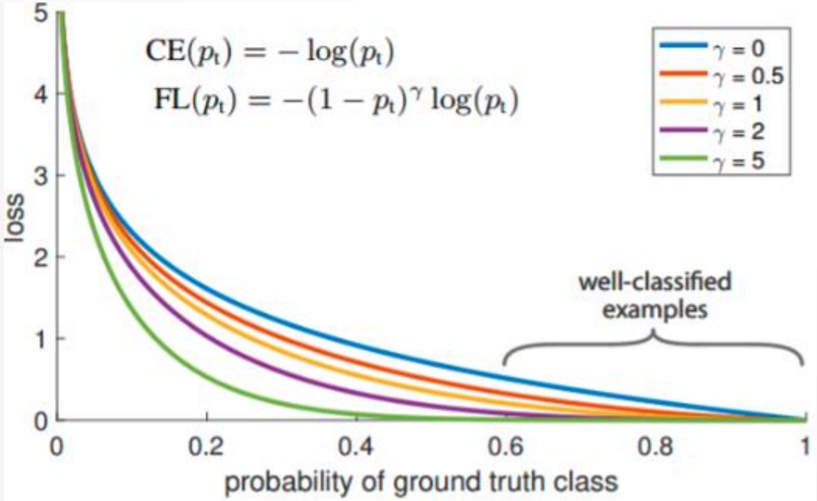
2. Ensemble model:
Private leaderboard score: 0.228
(PS: I've just adopted the ensemble strategy of averaging the results of different models, and if I change to other ensemble methods (e.g. stacking), the performance will be better.)

3. Why are we so much worse than the top 5 teams?
I think, because we cannot consider the data shift (domain shift) between train and test set

Other Attempts

1. Focal Loss

Which may helpful for the class unbalance problem, but which is not improve the results in my solution.



2. Concatenate the original image and detect-cropped image
(1) The original image has a bigger resolution, but which has a small ROI for animals.
(2) Detect-cropped image may have a bigger ROI for animals, but which limited by the performance for the detection model.
Can we use both the original and detect-cropped image for classification?
I tried the classification with the concatenation of original and detect-cropped images, but it makes the model worse. (May be need to some attention mechanisms)

+

Original Image

Detect-cropped Image

Todo …

1. GAN for domain adaptation

the training data and test data are from different regions.
There are domain shift between two dataset.

2. Advanced Transfer learning
Transfer learning with domain shift

3. Advanced Ensemble Strategy
Stacking, Boosting …

THANK YOU

Abuduweili
Xingchen Tao
Xin Wu

https://github.com/Walleclipse/iWildCam_2019_FGVC6