

Object Detection for Shopee Data

Kou Wen

Problem Description

Given a shopee women's apparel image, the designed system will draw bounding box of clothes and give tags of detected clothes.

Original Image



Result



Solution

- Object Detection:

Typically it will be separated into two steps; the first step is to localize the object using bounding box; in the second step, the object in each bounding box will be classified into a category respectively.



feed the window
size patch into
image classification

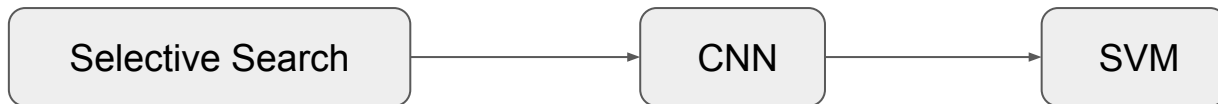
resize image at
multiple scales to
solve object size
variations

Object detection

- **HOG**

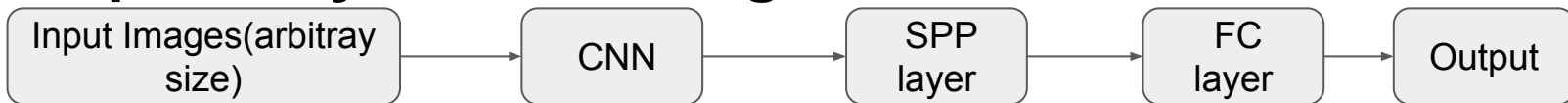


- **Region CNN**

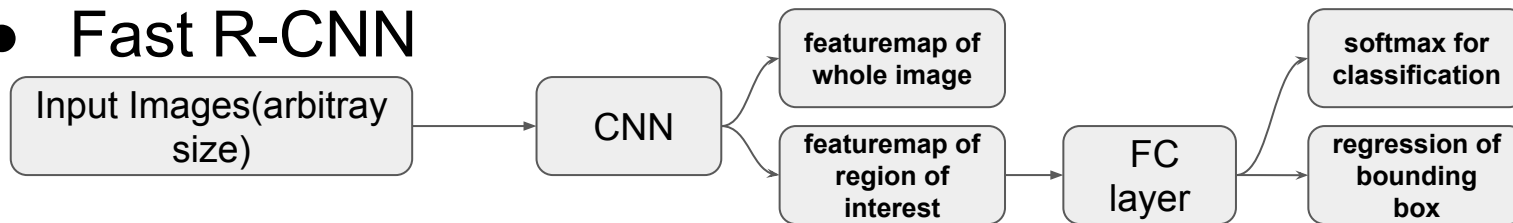


Object detection

• Spatial Pyramid Pooling Net



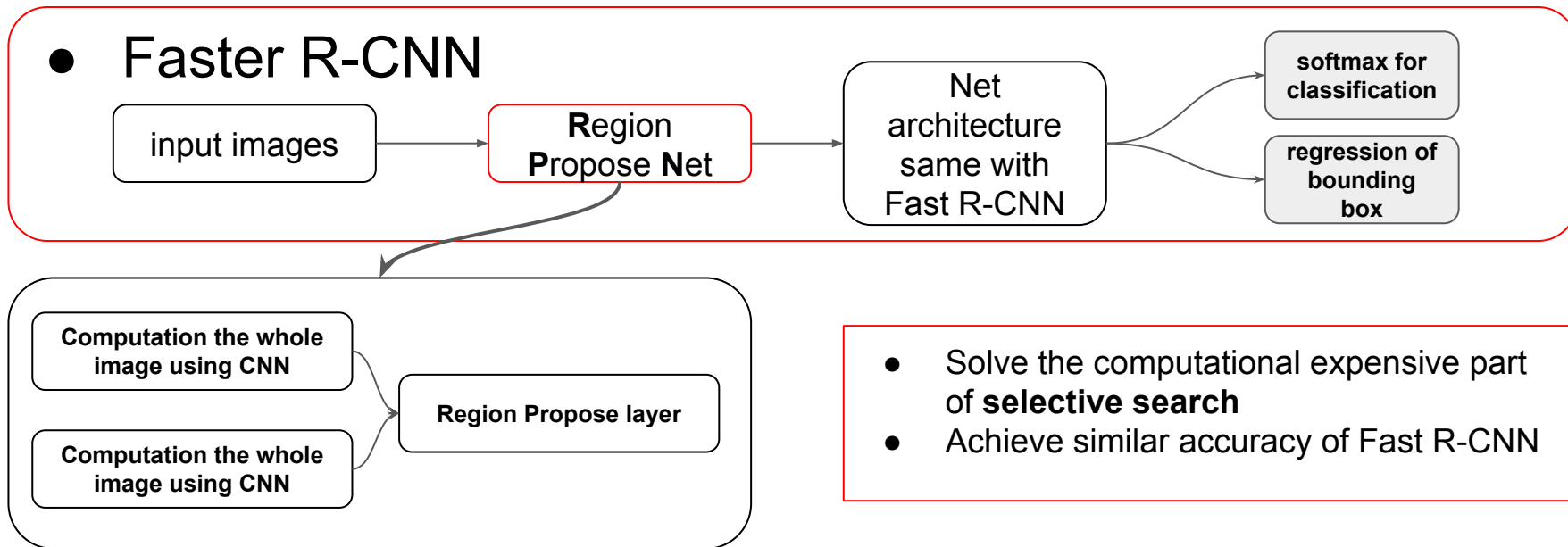
• Fast R-CNN



It is difficult for SPP net for fine-tune using the whole net architecture with SPP layer. Fast R-CNN later conquer the barrier. Both algorithms apply shared computation so they save a lot of time.

Object detection

- Faster R-CNN



- Solve the computational expensive part of **selective search**
- Achieve similar accuracy of Fast R-CNN

The current Faster R-CNN is one of the most advanced algorithm, so I choose this algorithm as my solution.

Implementation: source code

The source code is from <https://github.com/yhenon/keras-frcnn.git> . Now there is better version of <https://github.com/fizyr/keras-retinanet.git> . In this version, the original author improves the code quality for usability, extends the interface to read more different dataset and improves the speed.

My implementation is

https://github.com/kwkwvenusgod/Shopee_ObjectDetection.git

Implementation: dataset

In this task, the dataset is obtained from deep fashion. I applied the [Consumer-to-shop Clothes Retrieval Benchmark](#)

For Consumer-to-shop dataset, it has 239557 data with annotated bounding box and contains 16 categories in the total. Moreover, it covers both photo in shop and also consumers' selfies photo. So the **diversity** and the **volume** is desirable for training and cover the shopee dataset. The drawback is that the dataset does not contain sufficient women's underwear. There is dataset named [Objects segmentation in the fashion field](#) which contains women's underwear. But compare with the size with Deepfashion, the size is too small (400 images) in total; so I did not use it.

Implementation: training facility

1 GeForce GTX 1080 GPU

Result

Training time: the total training number is about 180,000 images. One full epoch will take approximate 30-40 hours.

Test time: for each image the recognition time will take 300ms to 700ms based on the image size.

Results



Conclusion

Issues

- a. The base deep network architecture is [resnet50](#) which is really deep and complicated; It will exhaust a lot of resources, in general it will occupy about 6GB memory only for resnet50.
- b. Since the dataset is too big and the training takes really long time
- c. The implementation will handle image input with different size as Fast R-CNN and SPP-net, it is difficult to train on batch, since each input image size is different. Keras may not be able to handle input tensor with different sizes. (Correct me if it can!!) It is also one reason for the
- d. Since so far the training is not so complete, it is difficult to judge the final model. But some output from shopee data can give some straightforward and intuitive result based on the trained model.
- e. There is still a lot to improve. For example, from the business view the category of women apparel may not be so appropriate, the summer wear can be a little abstract.

Q&A

Thanks