Counting Objects with Faster R-CNN

Krzysztof Grajek

02 Jun 2017. 6 minutes read

Accurately counting objects instances in a given image or video frame is a hard problem to solve in machine learning.
A number of solutions have been developed to count people, cars and other objects and none of them is perfect.
Of course, we are talking about image processing here, so a neural network seems to be a good tool for the job.

Below you can find a description of different approaches, common problems, challenges and latest solutions in the Neural Networks object counting field.
As a proof of concept, existing model for Faster R-CNN network will be used to count objects on the street with video examples given at the end of the post.

Challenges
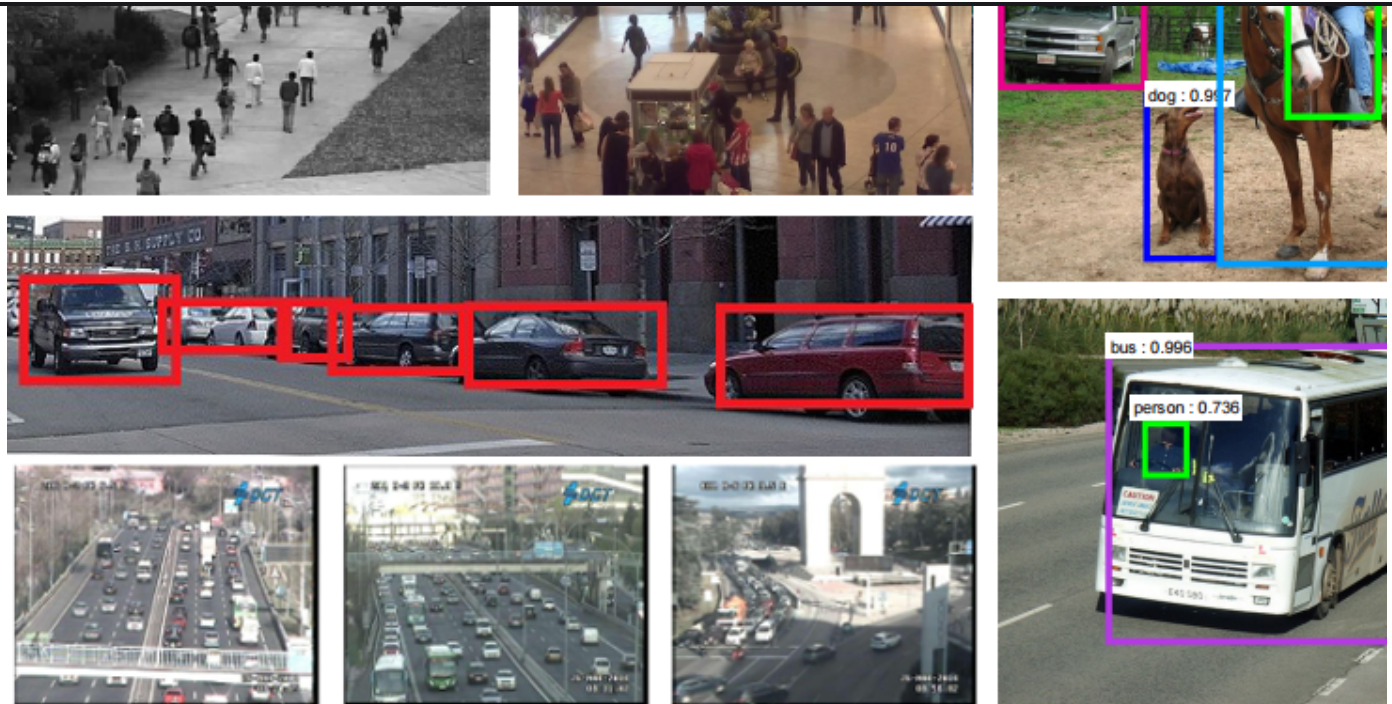
Finding a proper solution to the problem of counting objects depends on many factors. Besides some challanges common to all image processing with Neural Networks - like size of the training data, its quality etc.

Specific challenges to the counting objects problem:

- type of the objects to be counted
- overlapping
- perspective view
- the minimum size of detected objects
- training and testing speed

The approach taken to count cars on a highway or crowds of people at a stadium, where most objects overlap and the perspective view usually allows very small objects in the far distance, will be completely different to counting people on a family photo.
Also, the solution for counting objects on a single photo could be different to a solution suitable for counting objects in a video in real time manner.

Simple Needs, Simple Solutions

In this post I will try to tackle the problem of counting objects on the street, using sample videos with multiple objects visible at the same time, but not too overcrowded.

For processing images of a crowded scene or a traffic jam to count the object instances accurately I recommend diving into the latest research in the field: Towards perspective-free object counting with deep learning. The results from the paper can be reproduced using the code found at GitHub.

Methods like CCNN and Hydra CNN described in the aforementioned paper perform poorly when given an image with just a few objects of different types, therefore a different approach had to be taken.

There is a very interesting method in the field of machine learning (and in Deep Learning with Convolutional Neural Networks in particular), called Region based Convolutional Neural Network (RCNN), where we identify multiple objects and their location on a given image.

For our Proof Of Concept work I will use the Keras implementation of 'Faster R-CNN' modified to process video files and annotate the images with the count of detected objects of a given class.

Fast and Faster

There were number of approaches to combine the tasks of finding the object location and identifying the object to increase speed and accuracy. Over the years, we have moved forward from using standard RCNN networks, through Fast R-CNN and up to Faster R-CNN which we are using to solve our simple counting problem.

R-CNN introduced several innovations to improve training and testing speed, and detection accuracy.

Approaches using RCNN-trained models in multi-stage pipelines (first detecting object boundaries and then performing identification) were rather slow and not suited for real time processing.
The drawback of this approach is mainly its speed, both during the training and during the actual testing while object detection was performed.
Using the famous VGG16, the training process for a standard RCNN takes 2.5 GPU-days for the 5k images and requires hundreds of GB of storage. Detecting objects at test-time takes 47s/image using a GPU. This is mainly caused by performing a forward pass on the convolutional network for each object proposal, without sharing the computation.

Fast R-CNN improved RCNN by introducing a single-stage training algorithm which classifies objects and their spatial locations in a single processing stage. The improvements introduced in Fast R-CNN are:

- Higher detection quality
- Training in a single stage using multi-task loss
- Training can update all network layers
- No disk storage is required for feature caching

Faster R-CNN introduces a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, enabling nearly cost-free region proposals.
The RPN component of this solution tells the unified network where to look. For the same VGG-16 model, Faster R-CNN has a frame rate of 5 fps on a GPU while achieving state-of-the-art object detection accuracy. The RPN is a kind of a fully convolutional network and can be trained

of scales and aspect ratios.

Faster R-CNN was used last year by Pinterest as a solution enabling visual search on their website and it will be our choice to detect and count objects on sample videos in the described PoC below.

Proof Of Concept

To solve our imaginary problem, we are going to use the aforementioned Faster R-CNN model with Keras on a GPU-enabled AWS instance. Living in the era of multiple deep learning frameworks available and ongoing competitions, we are in a comfortable position to download already pretrained models best suited to our needs and the framework of choice.
Of course, you can train the model yourself using the provided training python script, just keep in mind that it can take many days to process.

There exist multiple implementations for Faster R-CNN, including Caffe, TensorFlow and possibly many others. We are going to use Keras (v. 2.0.3) with TensorFlow in the backend.
The code is available as a fork of original Keras F R-CNN implementation on GitHub.

The script for testing the network was modified so that it can process the video files and annotate each frame with appropriate data for detected objects (with probability) as well as a summary of counted objects.
I'm using opencv heavily to process the videos and already trained model (available for download here) while processing the frames.
There are a number of utility methods for processing the video, eg:

```
        flag, frame = cam.read()
        if flag:
            cv2.imwrite(os.path.join(img_path, str(counter) + '.jpg'),frame)
            counter = counter + 1
        else:
            break
        if cv2.waitKey(1) == 27:
            break
            # press esc to quit
    cv2.destroyAllWindows()
```

and saving the video from processed frames:

```
def save_to_video():
    list_files = sorted(get_file_names(output_path), key=lambda var:[int(x) if x.isdigit() else
    img0 = cv2.imread(os.path.join(output_path,'0.jpg'))
    height , width , layers =  img0.shape

    # fourcc = cv2.cv.CV_FOURCC(*'mp4v')
    fourcc = cv2.VideoWriter_fourcc(*'mp4v')
    #fourcc = cv2.cv.CV_FOURCC(*'XVID')
    videowriter = cv2.VideoWriter(output_video_file,fourcc, frame_rate, (width,height))
    for f in list_files:
        print("saving..." + f)
        img = cv2.imread(os.path.join(output_path, f))
        videowriter.write(img)
    videowriter.release()
    cv2.destroyAllWindows()
```

While object detection takes place during testing, we are creating the list of tuples with detected object class and number 1, which later gets reduced to count the number of occurrences for particular object class:

```
for jk in range(new_boxes.shape[0]):
            (x1, y1, x2, y2) = new_boxes[jk,:]

            cv2.rectangle(img_scaled,(x1, y1), (x2, y2), class_to_color[key],2)

            textLabel = '{}: {}'.format(key,int(100*new_probs[jk]))
            all_dets.append((key,100*new_probs[jk]))
            all_objects.append((key, 1))
```

and reducing method:

Script arguments are rather self explanatory:

- "--input_file", Path to input video file.
- "--output_file", Path to output video file.
- "--input_dir", Path to input working directory where the processed frames are stored
- "--output_dir" Path to output working directory where annotated processed frames are stored
- "--frame_rate" Frame rate to use while constructing the video output

Example usage:

```
python test_frcnn_count.py --input_file ~/videos/MVI_6848.mp4 --output_file ~/output4.mp4 --fram
```

A few examples processed by the script:

Keras RCNN Object Counting - Example 2

Keras RCNN Object Counting - Example 3

Summary

Region-based Deep Convolutional Networks are exciting tools, enabling software developers to solve many interesting problems. The presented

learning from other trained models, we can achieve high accuracy and speed while detecting objects.

Links and Downloads

- PoC project github page: GitHub
- Pre-trained model for Keras s3 link
- Fast R-CNN paper: Ross Girshick
- Faster R-CNN paper: Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun
- Video samples processed for this blog post needs were taken from Videezy

Comments **Community** 🔒 **Privacy Policy** ① **Login** ▾

♡ Recommend 5   🐦 Tweet   f Share   Sort by Best ▾

Join the discussion…

LOG IN WITH   OR SIGN UP WITH DISQUS ⑦

Name

Mir Akhtar Ali • 2 years ago

I want to count unique object i.e means it does not count same person more than one time
suppose if same person come in 3 frames there count must be one

1 ∧ | ∨ • Reply • Share ›

nandu yadav regati ➜ Mir Akhtar Ali • 2 years ago

Bro,did u get any solution for ur question???

∧ | ∨ • Reply • Share ›

Mir Akhtar Ali ➜ nandu yadav regati
• 2 years ago

yes with deep sort (open source)

1 ∧ | ∨ • Reply • Share ›

nandu yadav regati ➜ Mir Akhtar Ali
• 2 years ago

please help me to solve this.

^ | ∨ • Reply • Share ›

nandu yadav regati → Mir Akhtar Ali
• 2 years ago • edited

great job bro

^ | ∨ 1 • Reply • Share ›

Rumi • 5 days ago • edited

Traceback (most recent call last):
File "test_frcnn_count.py", line 281, in <module>
main()
File "test_frcnn_count.py", line 259, in main
cv2.rectangle(img_scaled,(x1, y1), (x2, y2),
class_to_color[key],2)
TypeError: only size-1 arrays can be converted to Python
scalars

How can I fix it Sir? please help me

^ | ∨ • Reply • Share ›

think studio • 2 years ago

list_files = sorted(get_file_names(output_path), key=lambda
var:[int(x) if x.isdigit() else x for x in re.findall(r'[^0-9]|[0-9]+',
var)])
TypeError: unorderable types: str() < int()

how to fix this error

^ | ∨ • Reply • Share ›

Kabir • 2 years ago

ModuleNotFoundError: No module named 'data_generators'
can you help me out. data_generators.py is already there in
the keras_frcnn folder Still it is giving me error

^ | ∨ • Reply • Share ›

Rochan Sharma → Kabir • 2 years ago

Add blank __init__.py file

^ | ∨ • Reply • Share ›

Owen Qing • 2 years ago

I am new at this job.how to deal with that...
error: ambiguous option: --input (--input_dir, --input_file?)
yuluojingliudeMacBook-Air:faster-rcnn-keras-master Owen$ --
output ~ /Users/ameochishizuru/Desktop/faster-rcnn-keras-
master/output --frame_rate=25

^ | ∨ • Reply • Share ›

amare mahtsentu • 2 years ago

thank you for the great post!!
can I use SSD multibox detectors for object counting ? how ?

Tam Nguyen • 2 years ago

Hi! I have a proplem:

ImportError: No moduel named 'data_generators'

I clone github: https://github.com/softberr...

Run cmd: python test_frcnn_count.py --input_file/video2.mp4 --output_file/output2.mp4 --frame_rate=10

Can you help me! Tks <3

∧ | ∨ • Reply • Share ›

demcoderseinnachbar • 3 years ago • edited

Where does the "Pre-trained model for Keras s3 link" come from? Which data have been used for training? Are there any additional pretrained models (in particular for traffic light detection)?

∧ | ∨ • Reply • Share ›

> KrisG → demcoderseinnachbar • 3 years ago
>
> You can possibly find some pre-trained models on the yhenon/keras-frcnn github issue pages (like the one copied to s3 bucket), eg:
> https://github.com/yhenon/k.... The training is executed against Pascal VOC data set.
>
> ∧ | ∨ • Reply • Share ›

>> demcoderseinnachbar → KrisG • 3 years ago
>>
>> THX
>>
>> ∧ | ∨ • Reply • Share ›

Nakhonsoft • 3 years ago

Error

anotating...
saving to video..
Traceback (most recent call last):
File "test_frcnn_count.py", line 281, in <module>
main()
File "test_frcnn_count.py", line 278, in main
save_to_video()
File "test_frcnn_count.py", line 66, in save_to_video
height , width , layers = img0.shape
AttributeError: 'NoneType' object has no attribute 'shape'

∧ | ∨ • Reply • Share ›

> Nakhonsoft → Nakhonsoft • 3 years ago
>
> how to fix?
>
> ∧ | ∨ • Reply • Share ›

>> KrisG → Nakhonsoft • 3 years ago
>>
>> check the path to 'img0 = cv2.imread(os.path.join(output_path,'0.jpg'))', it

Nakhonsoft • 3 years ago

ERROR:

File "test_frcnn_count.py", line 281, in <module>
main()
File "test_frcnn_count.py", line 112, in main
with open(config_output_filename, 'r') as f_in:
IOError: [Errno 2] No such file or directory: 'config.pickle'

∧ | ∨ • Reply • Share ›

> **Krzysztof Grajek** ➜ Nakhonsoft • 3 years ago
> pls check your paths, the file config.pickle is available
> at: https://github.com/softberr... folder
>
> ∧ | ∨ • Reply • Share ›
>
> > **Nakhonsoft** ➜ Krzysztof Grajek • 3 years ago
> > thank you, so now i fix it already
> > but i have problem in python3.6
> > ModuleNotFoundError: No module named
> > 'apt_pkg'
> >
> > ∧ | ∨ • Reply • Share ›
> >
> > > **Krzysztof Grajek** ➜ Nakhonsoft
> > > • 3 years ago
> > > This project was built with python 2.7
> > >
> > > ∧ | ∨ • Reply • Share ›
> > >
> > > **Nakhonsoft** ➜ Krzysztof Grajek
> > > • 3 years ago
> > > Thank you
> > >
> > > ∧ | ∨ • Reply • Share ›

polarix • 3 years ago

Hi there, thanks for a great work!
I'm not able to recognize all dogs in this photo. Why?

https://photos.app.goo.gl/r...

Thank you.

∧ | ∨ • Reply • Share ›

> **KrisG** ➜ polarix • 3 years ago
> Difficult to say without any information about the way
> you are doing object detection, network used etc.
> Maybe your network outputs only 5 bounding boxes
> with highest probablities? Maybe you need more
> training data cos not all your dogs are detected? Etc
>
> ∧ | ∨ • Reply • Share ›
>
> > **polarix** ➜ KrisG • 3 years ago

code's row where this limit is.
The network is Pre-trained with "model for
Keras" you linked.
Thank you.

∧ | ∨ • Reply • Share ›

Krzysztof Grajek ➔ polarix
• 3 years ago

You would need to debug it a bit
yourself, I can see that 'new_boxes,
new_probs =
roi_helpers.non_max_suppression_fast(b
np.array(probs[key]),
overlap_thresh=0.5)' should return max
300 boxes, have a look what it actually
returns and then you can step by step
proceed to the place where the boxes
are drawn.

∧ | ∨ • Reply • Share ›

okeke stephen • 3 years ago

Hello
@Krzysztof Grajek
I've been trying to run this code on anaconda Jupyter
notebook but I keep on getting error massages. Please how
can I achieve this? Thanks.

∧ | ∨ • Reply • Share ›

Vladimir Iglovikov • 3 years ago • edited

Did you try it for a dense crowd counting?

In the recent Deep Learning competition
(https://www.kaggle.com/c/no... where one needed to count
sea lions of different types Faster RCNN, SSD and all other
standard object detection tools worked to some extent, but not
as well as one may expect after reading scientific literature.
Direct regression after convolution layers worked much better.

I attached a picture of my predictions using Faster RCNN. It
does not look bad but it allowed me to finish 51 out of 393
which is pretty far from the top.

I know that Alexander Buslaev who finished 15th used SSD +
intelligent prosprocessing, but still 15th is not first.

Basically I am pretty curious is someone had good results for
object counting using Faster RCNN for the case when objects
are pretty dense at the pictures.

see more

∧ | ∨ • Reply • Share ›

overlapping objects the faster rcnn is not the best solution.

^ | ⌄ • Reply • Share ›

Akshay Deep Lamba • 3 years ago
where do I Get the training data .. I Actually need to see the structure of the data,annotations....

^ | ⌄ • Reply • Share ›

Akshay Deep Lamba • 3 years ago • edited
running your code gives the below error

Converting video to images..

(gst-plugin-scanner:29654): GStreamer-WARNING **: Failed to load plugin '/usr/lib/x86_64-linux-gnu/gstreamer-1.0/libgstopencv.so': libopencv_calib3d.so.2.4: cannot open shared object file: No such file or directory

(gst-plugin-scanner:29654): GStreamer-WARNING **: Failed to load plugin '/usr/lib/x86_64-linux-gnu/gstreamer-1.0/libgsteglglessink.so': /usr/lib/x86_64-linux-gnu/libgstegl-1.0.so.0: undefined symbol: eglDestroyImageKHR
GStreamer: Error opening bin: Unrecoverable syntax error while parsing pipeline /home/ubuntu/input/input.mp4
anotating...
saving to video..
Traceback (most recent call last):
File "test_frcnn_count.py", line 283, in <module>
main()
File "test_frcnn_count.py", line 280, in main
save_to_video()
File "test_frcnn_count.py", line 68, in save_to_video
height , width , layers = img0.shape
AttributeError: 'NoneType' object has no attribute 'shape'

^ | ⌄ • Reply • Share ›

KrisG ➜ Akshay Deep Lamba • 3 years ago
try to modify your input/output dir parameters, this error indicates that the program cannot find the image named 0.jpg which should be generated during the conversion process.

^ | ⌄ • Reply • Share ›

Avatar This comment was deleted.

KrisG ➜ Guest • 3 years ago
It probably means that it cannot find generated image named '0.jpg' (which should be generated in the earlier

Frank • 3 years ago • edited

Hi there, thanks for a great article!

Just wondering how long did it take you to train (via
https://github.com/softberr...)?

I'm trying to run this on my own dataset, but it's really slow -
any suggestions for speeding up the training? (I've already
reduces the image size to 224 in config.py)

Cheers

∧ | ∨ • Reply • Share ›

KrisG → Frank • 3 years ago

Not really, just a trivial one, assuming you are using
GPU already you can move onto a machine with more
cores as a huge part of the computations are
happening on CPU anyway.

1 ∧ | ∨ • Reply • Share ›

LernGer k • 2 years ago

Hi,
is it possible to use the pre-trained model with new image
data.. pls let me know the possibilities, as i am looking to test
this with some new data..

∧ | ∨ 1 • Reply • Share ›

Pooja • 3 years ago

What should i modify here ?
Converting video to images..
warning: Error opening file
(/build/opencv/modules/videoio/src/cap_ffmpeg_impl.hpp:578)
anotating...

→ Find more articles like this in Blog section

Let's do things together!
+48 22 188 11 33 (PL)
+44 56 0156 3406 (UK)
hello@softwaremill.com

NAME

E-MAIL

MESSAGE

SEND MESSAGE

See also

Portfolio   Services   Company

Join Us   Hire Us   Blog

Follow us

Latest on Blog

How to integrate remote teams: the online quiz

team   Company culture   remote teams   online party   remote party

© 2020 SoftwareMill. All rights reserved.
Privacy policy.

How to integrate remote teams: the online quiz