Exercise-5 : Visualising how a deep CNN makes decsisions

[Background]

Reference paper:

https://arxiv.org/abs/1611.05418

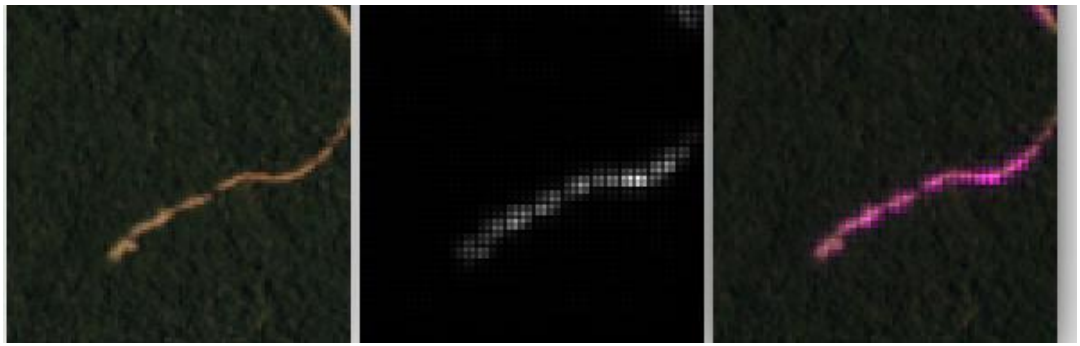
[1] "VisualBackProp: visualizing CNNs for autonomous driving" - Mariusz Bojarski(NVIDIA), Anna Choromanska, Krzysztof Choromanski, Bernhard Firner, Larry Jackel, Urs Muller, Karol Zieba, Arvix 2016

In this paper, the authors propose a method to determine which pixels in the image causes the final output of a deep CNN.

red: pixels that cause the CNN to make this decsision

We want to implement the visualisation method of the paper in pytorch and apply it to our satellite image classification problem. An example is:

input image classified as 'road'



pixels that cause this decsision

[tasks] Duration: 4 days

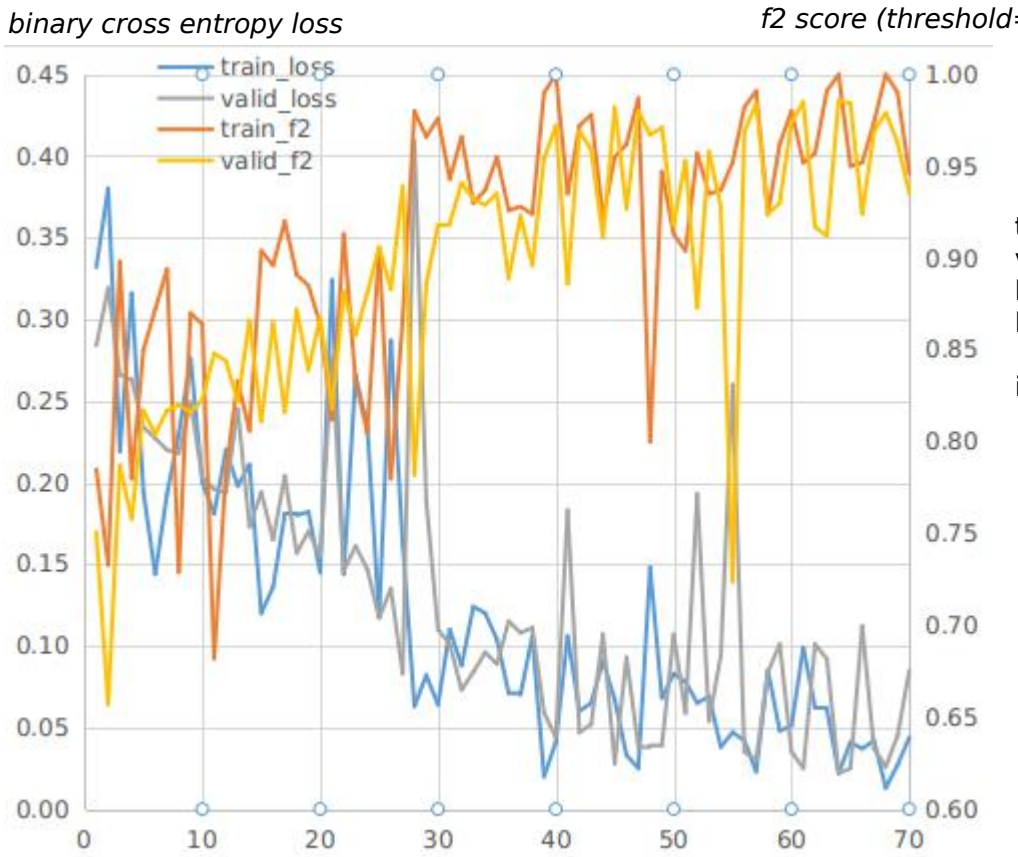
Step.1. Read the paper[1]. Make a presentation (e.g. PPT) to explain:

- the steps to compute the contribution score of each pixel to the final decision
- the mathematical reasoning for the above steps

[10 marks]

Step.2. Train a single label classifier. We use the 'road' class. Use the CNN below.

			parameters		
		feature maps	kernel	strid	pad
input		3x96x96			
block-0					
	conv2d	8x?x?	1x1	1	0
	batchnorm2d				
	relu				
block-1					
	conv2d	32x?x?	3x3	2	1
	batchnorm2d				
	relu				
block-2					
	conv2d	32x?x?	3x3	2	1
	batchnorm2d				
	relu				
block-3					
	conv2d	64x?x?	3x3	2	1
	batchnorm2d				
	relu				
global maxpool		64			
block-8					
	linear	512			
	batchnorm1d				
	relu				
prob					
	linear	?			
	...				



train_dataset.num = 32384
valid_dataset.num = 8095
batch_size = 96
learning rate = 0.01

input image = 96x96

example results

[10 marks]

Step.3. Write a function to save all feature maps (ouput of each layers) during a forward pass of a given image. *Hint: use 'register_forward_hook()'*

see: http://pytorch.org/tutorials/beginner/former_torchies/nn_tutorial.html#forward-and-backward-function-hooks
https://discuss.pytorch.org/t/how-to-extract-features-of-an-image-from-a-trained-model/119

example results

input

saved feature maps

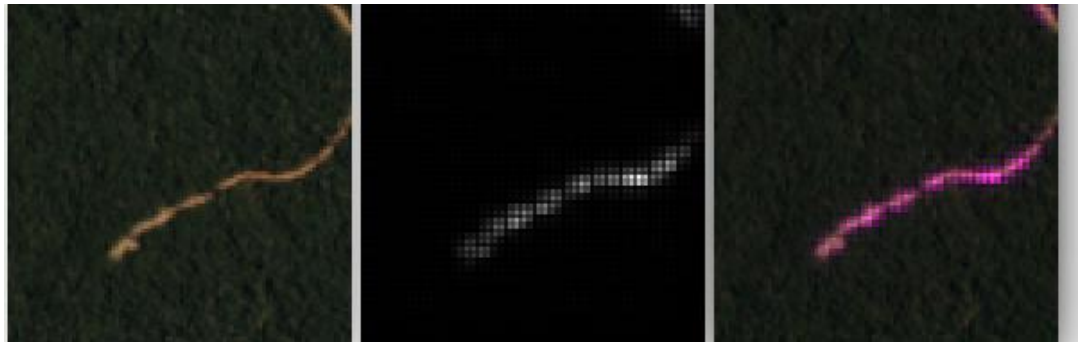
[10 marks]

Step.4. Write a function compute and visualise the contribution score of each pixel

see: https://github.com/mbojarski/VisualBackProp

example results

input image classified as 'road'



pixels that cause this decsision

[10 marks]

Question: Explain why is there blocky artifacts in the visualisation. Suggest a way to solved it

[10 marks]

Bouns. Use maxpool2d with stride=2 to instead of conv2d for "downsampling" in the CNN you have made previously.

1. Do you think the new CNN will be more accurate? Why?

2. Modify the visualisation function to support maxpool2d in the new CNN.
   Hint: you may have to use the maxunpool2d()

3. Modify the visualisation for a single multi-class classifier.
   Hint: you may want to read about papers on CAM and grad-CAM

[50 marks]

More results (on validation set):

image (true label)
estimated probability