Legend

think again

reference problem

elaborate more...

understandability is hard

Research Plan: Learning Image Transformations with Convolutional Neural Networks

Julien Lamon (4916727)

April 28, 2021

Title:	How sensitive are Convolutional Neural Networks to hyperparameters?
Author:	Julien Lamon
Responsible Professor:	Marco Loog, David Tax
(Optionally) Other Supervisor:	Yuko Kato, Taylan Turan, Tom Viering, Ziqi Wang,
Peer group members:	Julian Biesheuvel (BSc), Remco den Heijer (BSc), Ratish Thakoersingh (BSc)

Background of the research

used for other problems too.

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which, in short, applies a large number of filters to an input, and classify them under certain categories. The breakthrough of CNNs was mostly seen in the field of image recognition. Its benefit and innovation comes from the fact that it can efficiently detect features without any human supervision.

Whilst CNNs have first been discussed in the 1970s, the arrival of autonomous cars and much more makes it the preeminence in its favorable field: image recognition. However, while this method is very helpful, it is yet treated as a black box: studying its structure will not give any judgement on the structure being approximated. Although some scientists such as Zeiler and Fergus [1] came up with solutions to visualize the work of a CNN, various sections of it are still not fully understood.

Image recognition is a subject that can be treated in two different ways: classification and regression tasks. Recognizing an image through a classification algorithm means that a CNN will return a list of probabilities as an output, with each probability being the probability of the image being classified to a specific label. On the other hand, recognizing an image through a regression algorithm only outputs 1 number. This number can be the variance, standard deviation, median value, etc. With that distinction in mind, it should be noted that the current research project is being concentrated on the latter algorithm, therefore enabling us to perform some analysis on synthetic datasets.

your regression task can be [image $\longrightarrow \mathbb{R}^N$],

Research Question

As such, through this Research Project, a set of research questions has been shaped around aspects that requires a profound exploration in order to further improve our know-how. This research paper will then be concentrated on the hyperparameter aspects, with the question being: *How sensitive are Convolutional Neural Networks to hyperparameters?*.

However, through some primitive researches it can be seen that CNNs are used in numerous applications. It is true that it is prominent in the field of image recognition, but it is also applied in

the Natural Language Processing (NLP), speech recognitions, computer vision, etc. Therefore, with the size limitation of the research paper in mind, it was chosen to narrow the question down. The modified research question is the following: How sensitive are CNNs *in the Image Recognition field* to hyperparameters ?

From the freshly made research questions, sub-questions should be derived in order to tackle the research in a more step-by-step manner. In order to derive those questions, the first week was dedicated to a profound research around two specific questions:

- What does sensitivity mean ? How can it be measured ?
- What are the different hyperparameters that should be explored and measured ?

The former question - *What does sensitivity mean? How can it be measured?* - is the question that had to be answered before any other ones. As said explained in by Maosen Cao, Nizar F. Alkayem, Lixia Pan and Drahomir Novak (October 19th 2016), "the key point behind sensitivity analysis is that by slightly varying each explicative input parameter and registering the response in the output, the explicative parameters with the highest sensitivity values gain the greatest importance"[3]. Continuing the reading, it is also said that due to a limitation in the computational power, accompanied with "the absence of a reliable technique for defining the optimal structure"[3], CNNs usually are not on the top of their performance, or requires a considerable amount of time to get optimized hyperparameters. It will therefore be through a sensitivity analysis that we will be able to understand the insides of the model, to validate its correctness, and learn the bias on some parameters in comparison to others.

The latter question - *What are the different hyperparamaters that should be explored and measured?* - have been mostly been answered with the paper of Ye Zhang & Byron Wallace (2016). In this paper, as said in the title, they concentrate themselves on a sensitivity analysis of CNNs over Sentence Classification. Therefore, apart from some minor hyperparamaters, this paper is a paper that should really be followed during the integral process of the research paper. A list of parameters that will be analyzed during the research can be found on Figure 1.

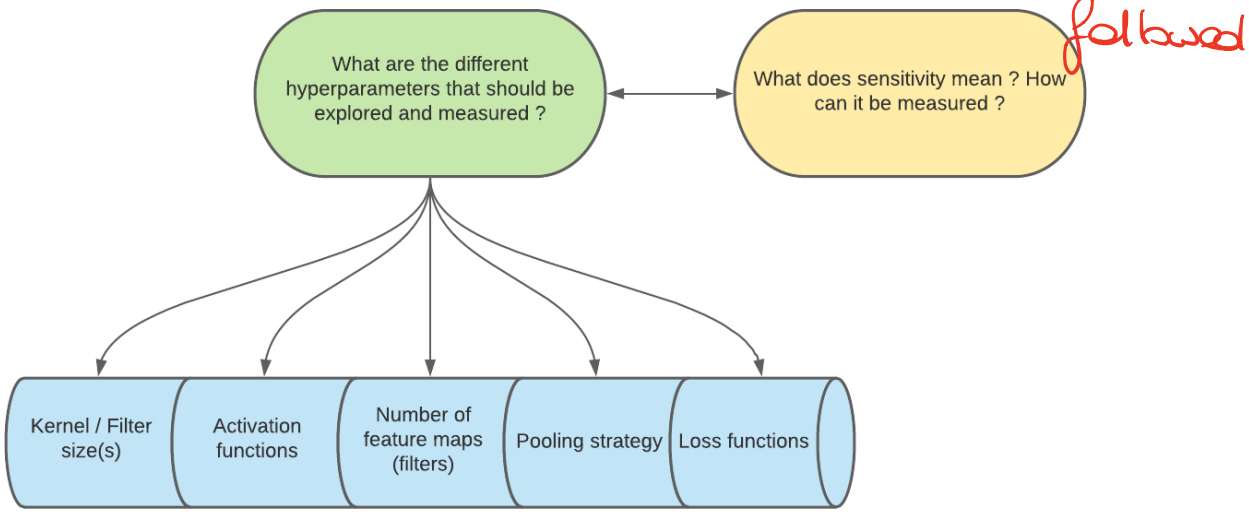

Figure 1: List of Hyperparameters that will be analyzed

As the two previous questions are now answered, obvious sub-questions can be derived. Namely, questions that will enable the project to have a continuous progress are the following:

Need to have some Why questions?

- How does the Kernel / Filter size(s) affect the performance of a CNN ?
- How can different activation function affect the performance of a CNN ?
- How does the number of feature maps (also known as Filters) affect the performance of a CNN?
- How can different pooling strategies affect the performance of a CNN ?
- How does can different loss functions affect the performance of a CNN ?

Method

With the fact that each research question is connected to other research members (all aspects are turning around CNNs), it was chosen to keep a common GitHub directory to help us during our own researches. The Master branch will be kept for a common CNN architecture, and then each team member will branch out to work on their own specific feature. This gives us a great advantage, since we can easily do some tests on other branches for examples. Moreover, with the fact that we are all working on CNNs, we can easily help each other out in case of any problem.

With this in mind, it was chosen to work with Python as a programming language. This permits us to have a simple and consistent code, with still very powerful Machine Learning Packages (such as Pytorch, numpy, etc).

An advantage of having the hyperparameter aspect as a research question is that tasks can easily be built up around them. Therefore, there is only a need to come up with tasks around the aforementioned sub-questions (see the Research Question section).

As seen in the upcoming section, tasks has been built up around the given deadlines (which are either in red or blue). The very first task that should be done is to finish the implementation of a CNN (which will be done with all three teammates). Then, as showed in the research paper of Ye Zhang & Byron Wallace (2016)[4], a "baseline model" should first be analysed in order to make comparisons with the upcoming sensitivity analysis. The next process is then to analyze each parameter one by one.

With the exploration taking most of the time, the planning also takes into account a considerable amount of time to write the report, and also the fact that a poster should be designed.

Planning of the research project

It should be noted that this planning is subject to considerable changes over time. While this planning is taking into account some tasks that will take more time than expected, it can still happen that problems come up, or that the current research plan is not well suited.

TABLE 1 Timeline

Week 1: 19.04.21 - 25.04.21	Look up some papers, elaborate on the influence of hyperparameters (HP).
	Set up python environment and start looking on how to tune parameters
19.04.21	Planning week 1
20.04.21	Information Literacy
21.04.21	Weekly meeting with every supervisors & peers
25.04.21	Research Plan
Week 2: 26.04.21 - 02.05.21	Continue to work on the CNN's implementation
	Try to implement an error estimation algorithm (cross-validation, etc) → Check existing tools... [sklearn, ray-sklearn]
	Prepare multiple datasets
	Start an analysis on a base model
	Prepare a slide set to present the research plan
28.04.21	Weekly meeting with every supervisors & peers
29.04.21	Research Plan Presentation
Week 3: 03.05.21 - 09.05.21	Continue analysis on base model
	Start analysing at least 2 HP, see how results can be derived (graphs,...) → which ones?
	Start with the poster
05.05.21	Weekly meeting with every supervisors & peers
06.05.21	Assignment 1: First 300 words
Week 4: 10.05.21 - 16.05.21	Start the report with the current HP analysis
	Continue the exploration with the HPs
	Finalize the poster
12.05.21	Weekly meeting with every supervisors & peers
16.05.21	Assignment 2: Midterm Poster
Week 5: 17.05.21 - 23.05.21	Fill the report with last week's explorations
	Do the final explorations → Are we done in 2 weeks?
18.05.21	Go / No-Go Presentations
19.05.21	Midterm poster
	Weekly meeting with every supervisors & peers

TABLE 2 Timeline (cont'd)

Week 6: 24.05.21 - 30.05.21	Fill the report with last week's explorations
	Start on filling the introduction, conclusion, abstract
26.05.21	Weekly meeting with every supervisors & peers
27.05.21	Assignment 3: Improve first 300 words and add section
Week 7: 31.05.21 - 06.06.21	Finalize the paper (proofread, etc.)
	Start with the final poster
02.06.21	Weekly meeting with every supervisors & peers
06.06.21	Paper draft v1
Week 8: 07.06.21 - 13.06.21	Change the Research paper with the given feedbacks
	Continue with the final poster
09.06.21	Weekly meeting with every supervisors & peers
Week 9: 14.06.21 - 20.06.21	Finalize Poster and Paper draft
16.06.21	Paper draft v2
	Weekly meeting with every supervisors & peers
18.06.21	Feedback on Paper
Week 10: 21.06.21 - 27.06.21	
23.06.21	Weekly meeting with every supervisors & peers
27.06.21	Final paper
	Assignment 4: Final Poster (for feedback)
Week 11: 28.06.21 - 04.07.21	
29.06.21	Final poster
01/02.07.21	Poster Presentation

Legend:

- Deadlines in red: Course hard deadlines
- Deadlines in blue: Academic Communication Skills deadlines

References

[1] Zeiler MD, & Fergus R (2014) Visualizing and understanding convolutional networks. In: Proceedings of Computer Vision - ECCV 2014, vol 8689, pp 818-833

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).

[3] Maosen Cao, Nizar F. Alkayem, Lixia Pan and DrahomÃr NovÃ¡k (October 19th 2016). Advanced Methods in Neural Networks-Based Sensitivity Analysis with their Applications in Civil

Engineering, Artificial Neural Networks - Models and Applications, Joao Luis G. Rosa, IntechOpen, DOI: 10.5772/64026. Available from: https://www.intechopen.com/books/artificial-neural-networks-models-and-applications/advanced-methods-in-neural-networks-based-sensitivity-analysis-with-their-applications-in-civil-engi

[4] Ye Zhang & Byron Wallace (2016). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification

Good job, Julien?

Things that can be better:

- More formal language.
- Consistent ref. format (inline).

Things you can look w/out architecture change.

- learning rate
- batch_size
- selection of optimizer algorithm

Again though trying to understand why is important. There are lots of work done regarding this topic. Good to have idea?

Cheers,

O. T. Turan