

Deeper Networks for Image Classification

Alexander Sworski

ECS795P – Deep Learning and Computer Vision

School of Electronic Engineering and Computer Science - Department of Computer Science

M.Sc. Artificial Intelligence

Queen Mary University

London, United Kingdom

a.sworski@se21.qmul.ac.uk

210456914

Abstract—This paper represents a comparison between different Convolutional Neural Networks. In particular, the models GoogLeNet, VGG-16 and ResNet are compared on different datasets, such as MNIST and Cifar10.

Index Terms—GoogLeNet, ResNet, VGG-16, MNIST, Image Classification, Deep Learning

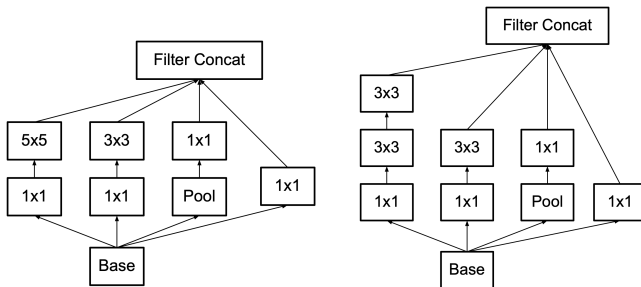
I. INTRODUCTION

This document is a model and instructions for L^AT_EX.

II. LITERATURE REVIEW

A. Neural Network models

1) *GoogLeNet*: GoogLeNet, was initially developed as a submission for the ImageNet Large-Scale Visual Recognition Challenge 2014, which it won. The main hallmark of this model is the improved utilisation of the computing resources inside the network. It has 12 times fewer parameters than the winner of 2012 (AlexNet) but performs significantly better. [6] This model first introduced the idea of an inception module, which can be seen in Figure 1a. A second version of the model was published a year later, which then introduced an improved inception module, which can be seen in Figure 1b. This new version reduced computational costs, as bigger convolutions are disproportionately more expensive. Using two 3x3 convolutions instead it computationally less expensive while improving the performance. Although there is also a V3 and V4 of GoogLeNet, for this paper, the V2 Inception has been used. In total the model has 22 layers.



(a) Original Inception module [6] (b) V2 Inception module [7]

Fig. 1: Three simple graphs

2) *VGG-16*: VGG-16, is the biggest of the models used in this paper (Figure 2). It also has been developed as a submission for the ImageNet Challenge, in which it won first and second place for the localization and classification tracks, respectively. It consists out of 16 layers and uses only 3x3 convolutions. Although it has fewer layers than GoogLeNet, each layer is computationally more complex. Ultimately, this model resulted in better scores than the GoogLeNet V1 [5].

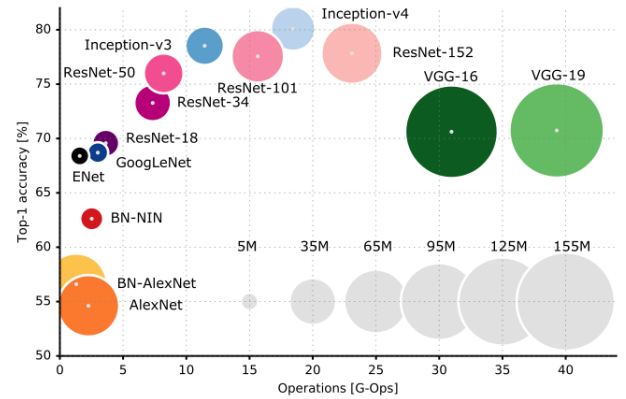


Fig. 2: top-1 one-crop accuracy versus amount of operations required (circle size represents the amount of parameters)

3) *ResNet*: ResNet, has the highest amount of layers. While there are multiple versions, for this paper a 50 layer version has been chosen. In Figure 2 it can be observed, that the model size

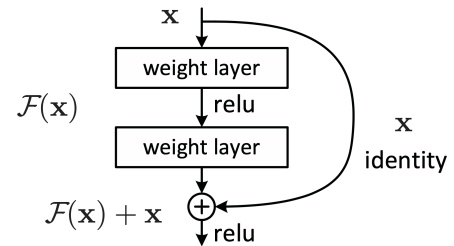


Fig. 3: Residual learning: a building block [4]

is in between GoogLeNet and VGG-16, yet the best result can be expected according to this comparison. ResNet has been

build with the goal of easing the training of deep networks. The network design is based on the concept of chaining multiple residual learning block on a row. One residual learning block can be seen in Figure 3. [4]

B. Image datasets

1) *MNIST*: The MNIST database contains 70,000 28x28 black and white images. 60,000 images are for training and 10,000 images for testing. The images' portrait hand written numbers from 0 to 9. [8] Examples of the classes can be seen in Figure 4.

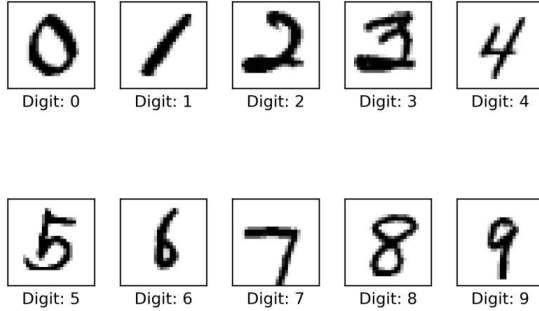


Fig. 4: Cifar10 image samples [2]

2) *Cifar10*: The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images. The classes are mutually exclusive.[1] Examples of the classes can be seen in Figure 5.



Fig. 5: Cifar10 image samples [3]

III. IMPLEMENTATION

In the first subsection III-A the implementation of the models using the framework Keras will be explained. The subsection III-B, it will be explained how the datasets have been made compatible with the models.

A. Model implementation

- 1) *GoogLeNet*:
- 2) *ResNet*:
- 3) *VGG-16*:

B. The Datasets

The dataset used in this paper, both have a resolution of 28x28 pixels. On top, the Cifar10 dataset has 3 color channels, while the MNIST dataset only has one gray-channel. This represents an issue, as the Neural Networks have initially been designed to work with the Microsoft ImageNet dataset. This dataset has 3 color channels and a resolution of 224x224 pixels. The goal was to not change the input of the models, as this would have significant implications for the layers following, resulting in a significant change in the model design. Therefore, the datasets have been altered as follows to fit the input dimensions of the models.

1) *MNIST dataset alterations*:

2) *Cifar10 dataset alterations*:

C. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

D. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”).

E. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

F. L^AT_EX-Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in L^AT_EX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

L^AT_EX can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

L^AT_EX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

G. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalise the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.

- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [1].

H. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organisation).

I. Identify the Headings

Headings, or heads, are organisational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organise the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

J. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. ??”, even at the beginning of a sentence.

TABLE I: Table Type Styles

Table Head	Table Column Head		
	<i>Table column subhead</i>	<i>Subhead</i>	<i>Subhead</i>
copy	More table copy ^a		

^aSample of a Table footnote.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets []. The sentence punctuation follows the bracket []. Refer simply to the reference number, as in []—do not use “Ref. []” or “reference []” except at the beginning of a sentence: “Reference [] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” []. Papers that have been accepted for publication should be cited as “in press” []. Capitalise only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [].

REFERENCES

- [1] *CIFAR-10 and CIFAR-100 datasets*. URL: <https://www.cs.toronto.edu/~kriz/cifar.html> (visited on 04/30/2022).
- [2] *Convolutional Neural Networks (CNN) for CIFAR-10 Dataset*. Parneet Kaur. URL: <http://parneetk.github.io/blog/cnn-cifar10/> (visited on 04/30/2022).
- [3] *Fig. 2. Some samples of the MNIST dataset*. ResearchGate. URL: https://www.researchgate.net/figure/Some-samples-of-the-MNIST-dataset_fig1_342733731 (visited on 04/30/2022).
- [4] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *arXiv:1512.03385 [cs]* (Dec. 10, 2015). version: 1. arXiv: 1512.03385. URL: <http://arxiv.org/abs/1512.03385> (visited on 04/30/2022).

- [5] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *arXiv:1409.1556 [cs]* (Apr. 10, 2015). version: 6. arXiv: 1409.1556. URL: <http://arxiv.org/abs/1409.1556> (visited on 04/30/2022).
- [6] Christian Szegedy et al. “Going Deeper with Convolutions”. In: *arXiv:1409.4842 [cs]* (Sept. 16, 2014). version: 1. arXiv: 1409.4842. URL: <http://arxiv.org/abs/1409.4842> (visited on 04/30/2022).
- [7] Christian Szegedy et al. “Rethinking the Inception Architecture for Computer Vision”. In: *arXiv:1512.00567 [cs]* (Dec. 11, 2015). version: 3. arXiv: 1512.00567. URL: <http://arxiv.org/abs/1512.00567> (visited on 04/30/2022).
- [8] Yann LeCun, Corinna Cortes, and Christopher J.C. Burges. *MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges*. URL: <http://yann.lecun.com/exdb/mnist/> (visited on 04/30/2022).