

# Convolutional Neural Network for Image Classification

**Chen Wang**

Johns Hopkins University  
Baltimore, MD 21218, USA  
cwang107@jhu.edu

**Yang Xi**

Johns Hopkins University  
Baltimore, MD 21218, USA  
yxi5@jhu.edu

## Abstract

This document contains the instructions for preparing a camera-ready manuscript for the proceedings of NAACL HLT 2010. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. Authors are asked to conform to all the directions reported in this document.

## 1 Introduction

The following instructions are directed to authors of papers accepted for publication in the NAACL HLT 2010 proceedings. All authors are required to adhere to these specifications. Authors are required to provide a Portable Document Format (PDF) version of their papers. The proceedings will be printed on US-Letter paper. Authors from countries in which access to word-processing systems is limited should contact the publication chairs as soon as possible.

## 2 General Instructions

Manuscripts must be in two-column format. Exceptions to the two-column format include the title, authors' names and complete addresses, which must be centered at the top of the first page (see the guidelines in Subsection 4.3), and any full-width figures or tables. Type single-spaced. Do not number the pages. Start all pages directly under the top margin. See the guidelines later regarding formatting the first page.

The maximum length of a manuscript is eight (8) pages for the main conference, printed single-sided,

plus one (1) page for references (see Section 5 for additional information on the maximum number of pages). Do not number the pages.

### 2.1 Electronically-available resources

NAACL HLT provides this description in  $\text{\LaTeX}2\epsilon$  (`naaclhlt2010.tex`) and PDF format (`naaclhlt2010.pdf`), along with the  $\text{\LaTeX}2\epsilon$  style file used to format it (`naaclhlt2010.sty`) and an ACL bibliography style (`naaclhlt2010.bst`). These files are all available at <http://naaclhlt2010.isi.edu>. A Microsoft Word template file (`naaclhlt2010.dot`) is also available at the same URL. We strongly recommend the use of these style files, which have been appropriately tailored for the NAACL HLT 2010 proceedings.

## 3 Background

Image classification has been one of the most important topics in the field of computer vision and machine learning. As a popular benchmark in this field, the CIFAR-10 databases (Krizhevsky, 2009) are frequently used as a benchmark to judge the performance of an classification algorithm. Many researchers paid neuromus efforts in this problem. Even though the best known result has achieved a 94% of accuracy (Karpthy, 2011), it is still a quite challenging issue and many other well designed algorithms in the performance ranking list can only achieve around 70

Neural network and a series of derivative algorithms (Mitchell, 1997) have been applied for image

classification problems for a long time. The whole network consists of multiple layers of neurons and each of them can be treated as a single processing unit. The basic idea of neural network is inspired by the biological neural networks (Aleksander and Morton, 1990). The backpropagation algorithm, the most popular way to train a multi-layered neural network, was proposed by Bryson and Yu-Chi (1969) and further improved by Werbos (1974), Rumelhart et al (1986) .

Instead of simple multiplication between neuron outputs and weights, convolutional neural network incorporates more properties such as convolution and down-pooling (Simard et al, 2003). Due to the rapid development of computing platform like GPU, increasingly more researchers start to apply this algorithm in complex image classification. According to many other researchers' work, we realize the difficulty of this dataset and classification issue. Normally it takes more than dozens of hours of time to train a good performance model, even with high performance GPU and some other parallel computing techniques.

## 4 Model

In this section, at first, since we have implemented an neural network without using any open source, so we'll give a brief introduction to neural network by explaining the feedforward and back propagation steps in mathematics. Then we will describe how we use the basic neural network to do the classifying job on CIFAR-10 dataset.

Neural networks are made up many layers, and each layer was consisted of different number of neurons which have trainable weights and biases. All the neurons are fully connected to the neurons in previous and post layers. The first layer is input layer which was viewed as a single vector. The last layer is output layer whose output view be as predict result. Other layers between input and output layer are called hidden layer which will process and pass the 'message' from previous layer to post layer. Every neuron will receive some inputs from neurons in previous layer. Then it performs a dot product of all the inputs, following with a non-linearity optionally function as output of this neuron.

**How does the neural network work**

1. Initialize all weight  $w_{ij}^{(l)}$  in the neural network, and  $w_{ij}^{(l)}$  stands for the weight on the path from the  $i$ th neuron in  $(l-1)$ th layer to the  $j$ th neuron in  $l$ th layer.

2. Feedforward:

2.1. Take one train data, set the input values of each neuron  $y_i^0$  in the 0th (input) layer and the label in output layer.

2.2. Compute total net input from pre-layer to each hidden layer neuron  $x_j^l$  in the next layer, and squash the total net input using an activation function as the output in next layer  $y_j^{(l)}$  (here we use the logistic function), then repeat the process with the output layer neurons.

$$x_j^l = \sum_{i=0}^{d(l-1)} w_{ij}^{(l)} y_i^{(l-1)}$$

$$y_j^{(l)} = \theta(x_j^l)$$

$$\theta(z) = \frac{1}{1 + e^{-z}}$$

3. Back propagation:

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \eta \frac{\partial E}{\partial w_{ij}^{(l)}} = w_{ij}^{(l)} - \eta \delta_j^{(l+1)} y_i^{(l)}$$

3.1. Compute the gradients  $\frac{\partial E}{\partial w_{ij}^{(n)}}$  on each weight from the second last layer to last output layer ( $l = n$ ). Let error  $E = \frac{1}{2}(\text{label}_j^{(n)} - y_j^{(n)})^2$ , then calculate the gradient of each weight between last layer and its pre-layer.

$$\frac{\partial E}{\partial w_{ij}^{(n)}} = \frac{\partial E}{\partial y_j^{(n)}} \frac{\partial y_j^{(n)}}{\partial x_j^{(n)}} \frac{\partial x_j^{(n)}}{\partial w_{ij}^{(n-1)}}$$

$$\frac{\partial E}{\partial y_j^{(n)}} = -(\text{label}_j^{(n)} - y_j^{(n)})$$

$$\frac{\partial y_j^{(n)}}{\partial x_j^{(n)}} = \theta'(x_j^{(n)}) = \theta(x_j^{(n)})(1 - \theta(x_j^{(n)}))$$

$$\frac{\partial x_j^{(n)}}{\partial w_{ij}^{(n-1)}} = y_i^{(n-1)}$$

$$\frac{\partial E}{\partial w_{ij}^{(n)}} = (y_j^{(n)} - \text{label}_j^{(n)}) \theta(x_j^{(n)}) (1 - \theta(x_j^{(n)})) y_i^{(n-1)}$$

Let  $\delta_j^{(n)} = \frac{\partial E}{\partial y_j^{(n)}} \frac{\partial y_j^{(n)}}{\partial x_j^{(n)}} = \frac{\partial E}{\partial x_j^{(n)}}$ , which we will take advantage in next step.

3.2. Compute gradient in previous layers ( $l = n-1$ )

$$\begin{aligned}\frac{\partial E}{\partial w_{ij}^{(n-1)}} &= \frac{\partial E}{\partial y_j^{(n-1)}} \frac{\partial y_j^{(n-1)}}{\partial x_j^{(n-1)}} \frac{\partial x_j^{(n-1)}}{\partial w_{ij}^{(n-2)}} \\ \frac{\partial E}{\partial y_j^{(n-1)}} &= \sum_{i=0}^{d(n-1)} \frac{\partial E_i}{\partial y_i^{(n-1)}} = \sum_{i=0}^{d(n-1)} \frac{\partial E_i}{\partial x_i^{(n)}} \frac{\partial x_i^{(n)}}{\partial y_i^{(n-1)}} \\ &= \sum_{i=0}^{d(n-1)} \delta_j^{(n)} w_{ij}^{(n-1)} \\ \frac{\partial E}{\partial w_{ij}^{(n-1)}} &= \sum_{i=0}^{d(n-1)} \delta_j^{(n)} w_{ij}^{(n-1)} \theta'(x_j^{(n-1)}) y_i^{(n-2)}\end{aligned}$$

Let  $\delta_j^{(n-1)} = \frac{\partial E}{\partial y_j^{(n-1)}} \frac{\partial y_j^{(n-1)}}{\partial x_j^{(n-1)}}$ , which will be used in next layer. Repeat the process until whole weights in the neural network were updated.

4. Return to the second step (feedforward), keep updating the weights until reach the iteration.

### Convolutional Neural Network

Convolutional neural network works based on basic neural networks which was described above. So what does the CNNs change? There are several variations on CNNs layers architecture: Convolutional Layer, Max-Pooling and Fully-Connected Layer. Fully-Connected Layer is just acting as neural network which we have already covered in previous. CNN algorithm has two main processes: convolution and sampling, which will happen on convolutional layers and max pooling layers. CNNs have much fewer connections and parameters, and also they are easier to train. Discarding the fully connected strategy means to pay more attention on the regional structure, which is very meaningful when we take image processing into consideration, since there are less relations between different region of an image.

Convolution process: every neuron takes inputs from a rectangular  $n \times n$  section of the previous layer, the rectangular section is called local receptive field.

$$x_{i,j}^{(l)} = \sigma(b + \sum_{r=0}^n \sum_{c=0}^n w_{r,c} x_{i+r,j+c}^{(l-1)})$$

Since the every local receptive field takes same weights  $w_{r,c}$  and biases  $b$  from the equation above, the parameters could be viewed as a trainable filter or kernel  $F$ , the convolution process could be considered as acting an image convolution and the convolutional layer is the convolution output of the previous layer. We sometimes call the trainable filter from the input layer to hidden layer a feature map with shared weights and bias.

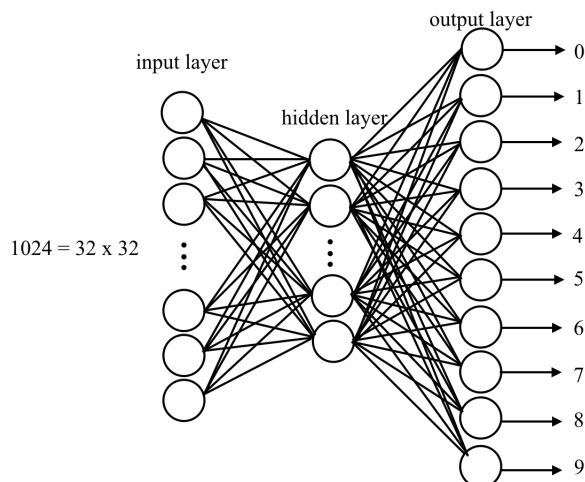
Sampling process: after each convolutional layer, there may be following a pooling layer. So the sampling process happens between convolutional layer and pooling layer. The pooling layer takes small rectangular blocks from the convolutional layer and subsamples it to produce a single output from that block. There are several ways to do this pooling, such as taking the average or the maximum, or a learned linear combination of the neurons in the block. Our pooling layers will always be max-pooling layers; that is, they take the maximum of the block they are pooling.

### How we use the neural network

The input layer of the network contains neurons encoding the values of the input pixels. Our training data for the network will consist of 32 by 32 pixel images from CIFAR-10 dataset, and so the input layer contains 1024 ( $32 \times 32$ ) neurons.

The second layer of the network is a hidden layer. We denote the number of neurons in this hidden layer by  $n$ , and we'll experiment with different values for  $n$ . The example shown illustrates a small hidden layer containing just  $n=15$  neurons.

The output layer of the network contains 10 neurons which stand for 10 types of image label. We number the output neurons from 0 through 9 and select the neuron with the highest activation value as predicting result.



## Conclusion

## 5 Our work

### 5.1 Format of Electronic Manuscript

For the production of the electronic manuscript you must use Adobe's Portable Document Format (PDF). This format can be generated from postscript files: on Unix systems, you can use `ps2pdf` for this purpose; under Microsoft Windows, you can use Adobe's Distiller, or if you have cygwin installed, you can use `dvipdf` or `ps2pdf`. Note that some word processing programs generate PDF which may not include all the necessary fonts (esp. tree diagrams, symbols). When you print or create the PDF file, there is usually an option in your printer setup to include none, all or just non-standard fonts. Please make sure that you select the option of including ALL the fonts. *Before sending it, test your PDF by printing it from a computer different from the one where it was created.* Moreover, some word processor may generate very large postscript/PDF files, where each page is rendered as an image. Such images may reproduce poorly. In this case, try alternative ways to obtain the postscript and/or PDF. One way on some systems is to install a driver for a postscript printer, send your document to the printer specifying "Output to a file", then convert the file to PDF.

For reasons of uniformity, Adobe's **Times Roman** font should be used. In  $\text{\LaTeX}2\text{e}$  this is accomplished by putting

```
\usepackage{times}
```

```
\usepackage{latexsym}
```

in the preamble.

Additionally, it is of utmost importance to specify the **US-Letter format** (8.5in  $\times$  11in) when formatting the paper. When working with `dvips`, for instance, one should specify `-t letter`.

Print-outs of the PDF file on US-Letter paper should be identical to the hardcopy version. If you cannot meet the above requirements about the production of your electronic submission, please contact the publication chairs above as soon as possible.

### 5.2 Layout

Format manuscripts two columns to a page, in the manner these instructions are formatted. The exact dimensions for a page on US-letter paper are:

- Left and right margins: 1in
- Top margin: 1in
- Bottom margin: 1in
- Column width: 3.15in
- Column height: 9in
- Gap between columns: 0.2in

Papers should not be submitted on any other paper size. Exceptionally, authors for whom it is *impossible* to format on US-Letter paper, may format for A4 paper. In this case, they should keep the *top* and *left* margins as given above, use the same column width, height and gap, and modify the bottom and right margins as necessary. Note that the text will no longer be centered.

### 5.3 The First Page

Center the title, author's name(s) and affiliation(s) across both columns. Do not use footnotes for affiliations. Do not include the paper ID number assigned during the submission process. Use the two-column format only when you begin the abstract.

**Title:** Place the title centered at the top of the first page, in a 15 point bold font. (For a complete guide to font sizes and styles, see Table 1.) Long title should be typed on two lines without a blank line intervening. Approximately, put the title at 1in from the top of the page, followed by a blank line,

then the author's names(s), and the affiliation on the following line. Do not use only initials for given names (middle initials are allowed). Do not format surnames in all capitals (e.g., "Leacock," not "LEACOCK"). The affiliation should contain the author's complete address, and if possible an electronic mail address. Leave about 0.75in between the affiliation and the body of the first page.

**Abstract:** Type the abstract at the beginning of the first column. The width of the abstract text should be smaller than the width of the columns for the text in the body of the paper by about 0.25in on each side. Center the word **Abstract** in a 12 point bold font above the body of the abstract. The abstract should be a concise summary of the general thesis and conclusions of the paper. It should be no longer than 200 words. The abstract text should be in 10 point font.

**Text:** Begin typing the main body of the text immediately after the abstract, observing the two-column format as shown in the present document. Do not include page numbers.

**Indent** when starting a new paragraph. For reasons of uniformity, use Adobe's **Times Roman** fonts, with 11 points for text and subsection headings, 12 points for section headings and 15 points for the title. If Times Roman is unavailable, use **Computer Modern Roman** (L<sup>A</sup>T<sub>E</sub>X2e's default; see section 4.1 above). Note that the latter is about 10% less dense than Adobe's Times Roman font.

## 5.4 Sections

**Headings:** Type and label section and subsection headings in the style shown on the present document. Use numbered sections (Arabic numerals) in order to facilitate cross references. Number subsections with the section number and the subsection number separated by a dot, in Arabic numerals.

**Citations:** Citations within the text appear in parentheses as (Gusfield, 1997) or, if the author's name appears in the text itself, as Gusfield (1997). Append lowercase letters to the year in cases of ambiguities. Treat double authors as in (Aho and Ullman, 1972), but write as in (Chandra et al., 1981) when more than two authors are involved. Collapse multiple citations as in (Gusfield, 1997; Aho and Ullman, 1972).

**References:** Gather the full set of references

together under the heading **References**; place the section before any Appendices, unless they contain references. Arrange the references alphabetically by first author, rather than by order of occurrence in the text. Provide as complete a citation as possible, using a consistent format, such as the one for *Computational Linguistics* or the one in the *Publication Manual of the American Psychological Association* (American Psychological Association, 1983). Use of full names for authors rather than initials is preferred. A list of abbreviations for common computer science journals can be found in the *ACM Computing Reviews* (Association for Computing Machinery, 1983).

The L<sup>A</sup>T<sub>E</sub>X and BibT<sub>E</sub>X style files provided roughly fit the American Psychological Association format, allowing regular citations, short citations and multiple citations as described above.

**Appendices:** Appendices, if any, directly follow the text and the references (but see above). Letter them in sequence and provide an informative title: **Appendix A. Title of Appendix.**

**Acknowledgment** sections should go as a last (unnumbered) section immediately before the references.

## 5.5 Footnotes

**Footnotes:** Put footnotes at the bottom of the page. They may be numbered or referred to by asterisks or other symbols.<sup>1</sup> Footnotes should be separated from the text by a line.<sup>2</sup> Footnotes should be in 9 point font.

## 5.6 Graphics

**Illustrations:** Place figures, tables, and photographs in the paper near where they are first discussed, rather than at the end, if possible. Wide illustrations may run across both columns and should be placed at the top of a page. Color illustrations are discouraged, unless you have verified that they will be understandable when printed in black ink.

**Captions:** Provide a caption for every illustration; number each one sequentially in the form: "Figure 1. Caption of the Figure." "Table 1. Caption of the Table." Type the captions of the figures and tables below the body, using 10 point text.

<sup>1</sup>This is how a footnote should appear.

<sup>2</sup>Note the line separating the footnotes from the text.

Type of Text	Font Size	Style
paper title	15 pt	bold
author names	12 pt	bold
author affiliation	12 pt	
the word “Abstract”	12 pt	bold
section titles	12 pt	bold
document text	11 pt	
abstract text	10 pt	
captions	10 pt	
bibliography	10 pt	
footnotes	9 pt	

Table 1: Font guide.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

## 6 Length of Submission

The NAACL HLT 2010 main conference accepts submissions of long papers and short papers. The maximum length of a long paper manuscript is eight (8) pages of content and one (1) additional page of references *only* (appendices count against the eight pages, not the additional one page). The maximum length of a short paper manuscript is four (4) pages including references. For both long and short papers, all illustrations, references, and appendices must be accommodated within these page limits, observing the formatting instructions given in the present document. Papers that do not conform to the specified length and formatting requirements are subject to be rejected without review.

## Acknowledgments

Do not number the acknowledgment section.

## References

- Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.
- American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.
- Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.
- Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.