
Formatting Instructions for NIPS 2013

Didier Landry

Department of Electrical and Computer Engineering
University of Toronto
didier.landry@mail.utoronto.ca

Dustin Kut Moy Cheung

Department of Electrical and Computer Engineering
University of Toronto
dustin.kutmoycheung@mail.utoronto.ca

Abstract

Detecting and reading text from photographs is a challenging computer vision problem that has received a lot of work in recent years. Being able to accurately localize, and recognize arbitrary digits from natural images is challenging due to the complex scenes in those images. In this report, we apply convolutional neural networks to learn a unique set of features optimized for this task, and discuss the evolution of our neural network topology to achieve a high accuracy on the validation set. We use over 500,000 labeled digits obtained from the SVHN citation needed for training. Moreover, we also describe our attempt to localize digits from unconstrained images by using image processing techniques and unsupervised learning.

1 Background

Convolutional neural networks citation needed are biologically-inspired neural networks that uses identical copies of the same neuron for training, allowing the network to have a large number of units while keeping the number of parameters of those neurons small. The individual neurons are generally connected to overlapping regions of an image, after those overlapping regions have been processed by filters. Those convolutional layers are connected with pooling layers. One popular choice of this kind of layer is max-pooling layer, which extracts the maximum of features over small blocks of the previous layer. Generally 2D convolutional neural networks are used in computer vision to learn features for extraction information. The 2D layer will look at patches of images to generate features such as the detection of the presence of an edge, or a particular texture.

The LeNet models are a popular family of models used in computer vision. They consist of multiple layers of convolutional layers and max-pooling, followed by fully-connected neural networks. We will be using one of those LeNet network for our training.

Recently, convolutional neural networks have been applied with dropout training in computer vision. When training deep neural networks with a small training set, the training generally leads to overfitting. One method to fix this is to simply stop training when the performance on the validation set starts to get worse. Another popular technique is to use L2 regularization on the neurons weights. Dropout is another technique to prevent overfitting by randomly dropping out units in a neural network.

The Street View House Numbers (SVHN) dataset is a popular training set consisting of Street View images cropped to either show single digits (Format 2) and multiple digits (Format 1). Those images

were extracted using a combination of automated algorithms and Amazon Mechanical Turk (AMT) framework [cite reading digits in natural images with unsupervised feature learning paper].

2 Problem Description

The project consists of classifying digits from street view images. The SVHN dataset will be used for training. All the images in the Format2 dataset have a fixed 32x32 resolution with a digit centered at the image. There are ten classes in total. The images show vast intra-class variations due to image distortions that happen in natural scene pictures. Factors that cause those image distortions include lighting, shadows, motion, and focus blurs.

The Format1 dataset consists of images with different resolutions with multiple digits in each image. The images are also not well cropped and not centered. The task is to first detect the digits with a bounding box, then classify each digit. The images in the Format1 dataset also show the same image variations as in the Format2 dataset.

Here, to facilitate the process of implementing and debugging the digit recognition, the Format 2 data was used to test the digit recognition neural network. The digit segmentation was treated as a different problem because it (obviously) degrades the quality of the input images fed to the neural network. However, the digit segmentation algorithm produces 32x32 images that could, then, be fed to the neural network for recognition.

3 Process

4 Citations, figures, tables, references

These instructions apply to everyone, regardless of the formatter being used.

4.1 Citations within the text

Citations within the text should be numbered consecutively. The corresponding number is to appear enclosed in square brackets, such as [1] or [2]-[5]. The corresponding references are to be listed in the same order at the end of the paper, in the **References** section. (Note: the standard L^AT_EX style `unsrt` produces this.) As to the format of the references themselves, any style is acceptable as long as it is used consistently.

As submission is double blind, refer to your own published work in the third person. That is, use “In the previous work of Jones et al. [4]”, not “In our previous work [4]”. If you cite your other papers that are not widely available (e.g. a journal paper under review), use anonymous author names in the citation, e.g. an author of the form “A. Anonymous”.

4.2 Footnotes

Indicate footnotes with a number¹ in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).²

4.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

¹Sample of the first footnote

²Sample of the second footnote

Table 1: Sample table title

| PART | DESCRIPTION |
|----------|-----------------------------------|
| Dendrite | Input terminal |
| Axon | Output terminal |
| Soma | Cell body (contains cell nucleus) |

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.



Figure 1: Sample figure caption.

4.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

5 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdf fonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NIPS. Please see <http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf>
- LaTeX users:
 - Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.
 - Otherwise, please generate your PostScript and PDF files with the following commands:


```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

Check that the PDF files only contains Type 1 fonts.

- xfig "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.

- The `\bbold` package almost always uses bitmap fonts. You can try the equivalent AMS Fonts with command

```
\usepackage[psamsfonts]{amssymb}
```

or use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{\mathbb{R}} %real numbers
```

```
\newcommand{\Nat}{\mathbb{N}} %natural numbers
```

```
\newcommand{\CC}{\mathbb{C}} %complex numbers
```

- Sometimes the problematic fonts are used in figures included in LaTeX files. The ghostscript program `eps2eps` is the simplest way to clean such figures. For black and white figures, slightly better results can be achieved with program `potrace`.

- MSWord and Windows users (via PDF file):

- Install the Microsoft Save as PDF Office 2007 Add-in from <http://www.microsoft.com/downloads/details.aspx?displaylang=en&familyid=4d951911-3e7e-4ae6-b059-a2e79ed87041>
- Select "Save or Publish to PDF" from the Office or File menu

- MSWord and Mac OS X users (via PDF file):

- From the print menu, click the PDF drop-down box, and select "Save as PDF..."

- MSWord and Windows users (via PS file):

- To create a new printer on your computer, install the AdobePS printer driver and the Adobe Distiller PPD file from <http://www.adobe.com/support/downloads/detail.jsp?ftpID=204> *Note:* You must reboot your PC after installing the AdobePS driver for it to take effect.
- To produce the ps file, select "Print" from the MS app, choose the installed AdobePS printer, click on "Properties", click on "Advanced."
- Set "TrueType Font" to be "Download as Softfont"
- Open the "PostScript Options" folder
- Select "PostScript Output Option" to be "Optimize for Portability"
- Select "TrueType Font Download Option" to be "Outline"
- Select "Send PostScript Error Handler" to be "No"
- Click "OK" three times, print your file.
- Now, use Adobe Acrobat Distiller or `ps2pdf` to create a PDF file from the PS file. In Acrobat, check the option "Embed all fonts" if applicable.

If your file contains Type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

5.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below using `.eps` graphics

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (<http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps>)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.[1, 2]

References

- [1] Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J Wu, and Andrew Y Ng. Text detection and character recognition in scene images with unsupervised feature learning. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 440–445. IEEE, 2011.
- [2] Yann Lecun. Lenet-5, convolutional neural networks. <http://yann.lecun.com/exdb/lenet/>.