Convolutional Layer Aggregation using LSTM

Yu Qin

Paper ID ***

Abstract. The abstract should summarize the contents of the paper and should contain at least 70 and at most 300 words. It should be set in 9-point font size and should be inset 1.0 cm from the right and left margins. ...

1 Introduction

In recent years, Convolutional Neural Networks (CNNs) have shown remarkable advantage on computer vision tasks like image classification[]. The basic architecture of convolutional layer consists of two levels, feature extraction and feature mapping. In feature extraction level, the input of each convolutional neuron is connected to local receptive domain and the local characteristics are extracted. Feature mapping level employs multiple convolutional kernels to focus on diferent aspects of the characteristics. The results of each convolutional layer are customarily regarded as features containing spatial and channel-wise information. A series of convolutional layers are stacked together to expand the field of reception and to generate higher level features. The evolution of CNNs from LeNet[] to DenseNet[] increases both the performance and the size of the network, which yields deeper and wider network structures.

From the first application in ResNet[], skip connections have been introduced into CNN structures, and proven effective in various vision tasks. Skip connections combine the output of previous layer and the current layer, dealing with the gradient vanishing problem. DenseNet[] connects densely in a block to make better use of previous features. To further utilize features from different layers, Yu[deep layer aggregation] extends the current skip connection approach proposes deep layer aggregation architectures. These architectures simply combine features of different level by concatenation or addition, without considering the interior relationship between low-level and high-level feature representations.

Recurrent Neural Networks (RNNs)[] has been proposed to deal with sequantial data like text or speech. Different from feedforward neural networks, RNNs build connections between nodes which are in the same layer. RNNs can be unfolded as a directed graph along the time steps, with all the layers sharing the same weights. This makes RNNs applicable to sequential tasks such as text classification. Long Short Term Memory (LSTM)[] is a special RNN, which makes use of three gates to select valuable information from all the memories. LSTM has proven to be more efficient than normal RNNs in most tasks on sequences.

In this work, we investigate a brand new approach to convolutional layer aggregation, by introducing a new architecture which is named as '*Concolutional*

Neural Networks-Recurrent Aggregation' (CNN-RA). Our goal is to aggregate outputs of multiple layers and retrieve more expressive features. To achieve this, we build CNN-RA by building parallel connection between a CNN and a LSTM. Features from lower Convolutional layers to higher layers naturally form a sequance with a variety of information. This kind of sequence contains both the features themselves and the transformation relationship between different features, which directly leads us to RNNs. We crearte information between outputs of convolutional layers and the inputs of LSTM, and employ the outputs of LSTM as the final feature for tasks such as image classification.

The receptive fields and feature maps sizes of different convolutional layers vary from each other, especially for two layers with a pooling layer inside. We propose an algorithm to transform different shape of feature matrixes to vectors with the same dimention. Then transformed vectors are stacked together as inputs of LSTM. The number of chosen features is the step lengh of the LSTM.

The development of new network architectures is always a time consuming task with abundant hyper parameters to determine. Previous work on layer aggreegation such as DLA[] brings with huge change on the original network architecture, which can even be much more complicated. However, our proposed CNN-RA won't do any modification on the original network, by only connecting it with a parallel LSTM. This property enable CNN-RA easily applicable to multiple convolutional network structures.

Our evaluation experiments extend famous network structures VGG[] and ResNet[] for standard image classification dataset. The testing results show improvements across different network structures and datasets. The connected LSTM brings with higher performance without increasing much parameter count. The experiment processes show that the relationship between two convolutional blocks with a down sampling layer inside has the most important contribution to the model.

2 Related Work

3 Methodology

Convolutional layer aggregation is a combination of features from different layers. The output matrices of convolutional layers are regarded as expressive features for vision tasks. In general, shallow layers contain low-level features and deep layers contain high-level features. Existing work simply uses high-level features or a combination of all layers. In this work, we take into account both feature combination and the sequential tranformation relationship between all levels.

7.1 Language

All manuscripts must be in English.

7.2 Paper Length

The submission page length is 14 pages for content plus maximum two pages for references. Over-length papers will simply not be reviewed. This includes papers where the margins and formatting are deemed to have been significantly altered from those laid down by this style guide. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts.

7.3 Paper ID

It is imperative that the paper ID is mentioned on each page of the manuscript. The paper ID is a number automatically assigned to your submission when registering your paper submission on CMT.

7.4 Dual Submission

By submitting a manuscript to ACCV 2018, the author(s) assert that it has not been previously published in substantially similar form. Furthermore, no paper which contains significant overlap with the contributions of this paper either has been or will be submitted during the ACCV 2018 review period to either a journal or a conference. However, the manuscript may also be submitted to one workshop that is accompanying ACCV 2018.

If there are any papers that may appear to the reviewers to violate this condition, then it is your responsibility to (1) cite these papers (preserving anonymity as described in Section 8 of this example paper, (2) argue in the body of your paper why your ACCV paper is non-trivially different from these concurrent submissions, and (3) include anonymized versions of those papers in the supplemental material.

7.5 Supplemental Material

Authors may optionally upload supplemental material. Typically, this material might include videos of results that cannot be included in the main paper, anonymized related submissions to other conferences and journals, and appendices or technical reports containing extended proofs and mathematical derivations that are not essential for understanding of the paper. Note that the contents of the supplemental material should be referred to appropriately in the paper and that reviewers are not obliged to look at it.

All supplemental material must be zipped or tarred into a single file. There is a 30 MB limit on the size of this file. The deadline for supplemental material is a week after the main paper deadline.

7.6 Line Numbering

All lines should be numbered, as in this example document. This makes reviewing more efficient, because reviewers can refer to a line on a page. If you are preparing a document using a non-LaTeX document preparation system, please arrange for an equivalent line numbering. Note that accepted papers need to be submitted as a LaTeX document in the style as defined in this document.

7.7 Mathematics

Please number all of your sections and displayed equations. Again, this makes reviewing more efficient, because reviewers can refer to a line on a page. Also, it is important for readers to be able to refer to any particular equation. Just because you did not refer to it in the text does not mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1." (Note that the line numbering will not be present in the final copy, so is not an alternative to equation numbers). Some authors might benefit from reading Mermin's description of how to write mathematics: http://www.pamitc.org/documents/mermin.pdf.

8 Blind Review

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one's own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words "my" or "our" when citing previous work. That is all. (But see below for techreports).

Saying "this builds on the work of Lucy Smith [1]" does not say that you are Lucy Smith, it says that you are building on her work. If you are Smith and Jones, do not say "as we show in [7]," say "as Smith and Jones show in [7]" and at the end of the paper, include Reference 7 as you would any other cited work.

An example of a bad paper:

An Analysis of the Frobnicatable Foo Filter

In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Removed for blind review

An example of an excellent paper:

An Analysis of the Frobnicatable Foo Filter

In this paper we present a performance analysis of the paper of Smith and Jones [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Smith, L., Jones, C.: The frobnicatable foo filter, a fundamental contribution to human knowledge. Nature **381** (2005) 1–213

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [**?**] as additional material and cite it as

[1] Authors: The frobnicatable foo filter, ACCV 2018 Submission ID 512, Supplied as additional material `accv18-512-frfofi.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper "further details may be found in [**?**]." Then submit the anonymized techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let us say it is 1969, you have solved a key problem on the Apollo lander, and you believe that the ACCV audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties," by Zeus.

You can handle this paper like any other. Do not write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]." That would be silly, and would immediately identify the authors. Instead write the following:

We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] did not handle case B properly. Ours handles it by including a foo term in the bar integral.

> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon.

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ: Are acknowledgements OK? – Answer: No. Please *omit acknowledgements* in your review copy; they can go in the final copy.

9 Manuscript Preparation

This is an edited version of Springer LNCS instructions adapted for ACCV 2018 full paper submission.

You will have to use LaTeX2$_\varepsilon$ for the preparation of your final (accepted) camera-ready manuscript together with the corresponding Springer class file `llncs.cls`.

We would like to stress that the class/style files and the template should not be manipulated and that the guidelines regarding font sizes and format should be adhered to. This is to ensure that the end product is as homogeneous as possible.

9.1 Printing Area

The printing area is 122 mm × 193 mm. The text should be justified to occupy the full line width, so that the right margin is not ragged, with words hyphenated as appropriate. Please fill pages so that the length of the text is no less than 180 mm.

9.2 Layout, Typeface, Font Sizes, and Numbering

Use 10-point type for the name(s) of the author(s) and 9-point type for the address(es) and the abstract. For the main text, use 10-point type and single-line spacing. We recommend using Computer Modern Roman (CM) fonts, Times, or one of the similar typefaces widely used in photo-typesetting. (In these typefaces the letters have serifs, *i.e.*, short endstrokes at the head and the foot of letters.) Italic type may be used to emphasize words in running text.

Bold type and underlining should be avoided.

With these sizes, the interline distance should be set so that some 45 lines occur on a full-text page.

Headings. Headings should be capitalised (*i.e.*, nouns, verbs, and all other words except articles, prepositions, and conjunctions should be set with an initial capital) and should, with the exception of the title, be aligned to the left. Words joined by a hyphen are subject to a special rule. If the first word can stand alone, the second word should be capitalised. The font sizes are given in Table 1. (Note that vertical lines are not common table components anymore.)

Table 1. Font sizes of headings. Table captions should always be positioned *above* the tables. A table caption ends with a full stop.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes ...**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Headings.** Text follows ...	10 point, bold
4th-level heading	*Remark.* Text follows ...	10 point, italic

Here are some examples of headings: "Criteria to Disprove Context-Freeness of Collage Languages," "On Correcting the Intrusion of Tracing Non-deterministic Programs by Software," "A User-Friendly and Extendable Data Distribution System," "Multi-flip Networks: Parallelizing GenSAT," "Self-determinations of Man."

Lemmas, Propositions, and Theorems. The numbers accorded to lemmas, propositions, theorems, and so forth should appear in consecutive order, starting with the number one, and not, for example, with the number eleven.

9.3 Figures and Photographs

Please produce your figures electronically and integrate them into your text file. For LaTeX users we recommend using package `graphicx` or the style files `psfig` or `epsf`.

Check that in line drawings, lines are not interrupted and have constant width. Grids and details within the figures must be clearly readable and may not be written one on top of the other. Line drawings should have a resolution of at least 800 dpi (preferably 1200 dpi). For digital halftones 300 dpi is usually sufficient. The lettering in figures should have a height of 2 mm (10-point type). Figures should be scaled up or down accordingly. Please do not use any absolute coordinates in figures.

Figures should be numbered and should have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table. Please center the captions between

the margins and set them in 9-point type (Fig. 1 shows an example). The distance between text and figure should be about 8 mm, the distance between figure and caption about 5 mm.

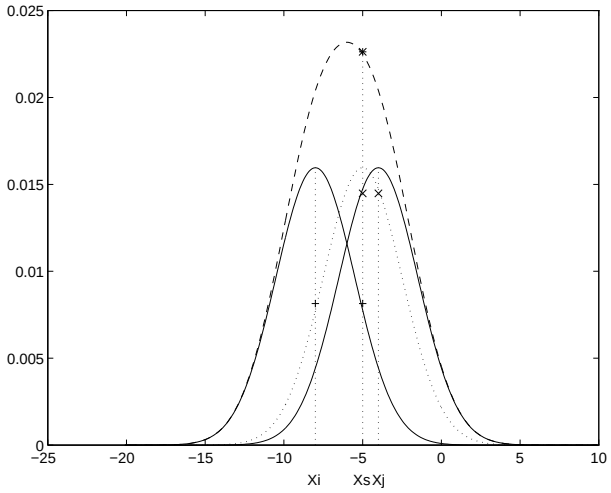

Fig. 1. One kernel at x_s (*dotted kernel*) or two kernels at x_i and x_j (*left and right*) lead to the same summed estimate at x_s. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption. The last sentence of a figure caption should generally end without a full stop

If possible (e.g. if you use LaTeX) please define figures as floating objects. LaTeX users, please avoid using the location parameter "h" for "here". If you have to insert a pagebreak before a figure, please ensure that the previous page is completely filled.

9.4 Formulas

Displayed equations or formulas are centered and set on a separate line (with an extra line or halfline space above and below). Displayed expressions should be numbered for reference. The numbers should be consecutive within each section or within the contribution, with numbers enclosed in parentheses and set on the right margin. For example,

$$\psi(u) = \int_o^T \left[\frac{1}{2} \left(\Lambda_o^{-1} u, u \right) + N^*(-u) \right] \mathrm{d}t \ . \tag{1}$$

Please punctuate a displayed equation in the same way as ordinary text but with a small space before the end punctuation.

9.5 Program Code

Program listings or program commands in the text are normally set in typewriter font, for example, CMTT10 or Courier.

Example of a Computer Program

```
program Inflation (Output)
  {Assuming annual inflation rates of 7%, 8%, and 10%,...
   years};
   const
     MaxYears = 10;
   var
     Year: 0..MaxYears;
     Factor1, Factor2, Factor3: Real;
   begin
     Year := 0;
     Factor1 := 1.0; Factor2 := 1.0; Factor3 := 1.0;
     WriteLn('Year  7% 8% 10%'); WriteLn;
     repeat
       Year := Year + 1;
       Factor1 := Factor1 * 1.07;
       Factor2 := Factor2 * 1.08;
       Factor3 := Factor3 * 1.10;
       WriteLn(Year:5,Factor1:7:3,Factor2:7:3,Factor3:7:3)
     until Year = MaxYears
end.
```

(Example from Jensen K., Wirth N. (1991) Pascal user manual and report. Springer, New York)

9.6 Footnotes

The superscript numeral used to refer to a footnote appears in the text either directly after the word to be discussed or, in relation to a phrase or a sentence, following the punctuation sign (comma, semicolon, or full stop). Footnotes should appear at the bottom of the normal text area, with a line of about 2 cm in TEX and about 5 cm in Word set immediately above them.[1]

9.7 Citations

The list of references is headed "References" and is not assigned a number in the decimal system of headings. The list should be set in small print and placed at the end of your contribution, in front of the appendix, if one exists.

[1] The footnote numeral is set flush left and the text follows with the usual word spacing. Second and subsequent lines are indented. Footnotes should end with a full stop.

Please do not insert a page break before the list of references if the page is not completely filled. Citations in the text are with square brackets and consecutive numbers, such as [**?**], or [**?**,**?**].

10 Conclusions

The paper ends with a conclusion.