

L^AT_EX Author Guidelines for CVPR Proceedings

Mouse Lee
jl27@princeton.edu

Jonathan Metzman
jmetzman@princeton.edu

Bo Moon
bhmoon@princeton.edu

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word “Abstract” as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

1.1. Language

All manuscripts must be in English.

1.2. Dual submission

Please refer to the author guidelines on the CVPR 2015 web page for a discussion of the policy on dual submissions.

1.3. Paper length

For CVPR 2015, the rules about paper length have changed, so please read this section carefully. Papers, excluding the references section, must be no longer than eight pages in length. The references section will not be included in the page count, and there is no limit on the length of the references section. For example, a paper of eight pages with two pages of references would have a total length of 10 pages. **Unlike previous years, there will be no extra page charges for CVPR 2015.**

Overlength 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. Note that this L^AT_EX guide already sets figure captions and references in a smaller font. The reason

such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in eight pages if it is reviewed in eleven.

1.4. The ruler

The L^AT_EX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-L^AT_EX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (L^AT_EX users may uncomment the `\cvprfinalcopy` command in the document preamble.) Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 095.5), although in most cases one would expect that the approximate location will be adequate.

1.5. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't 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 ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: <http://www.pamitc.org/documents/mermin.pdf>.

1.6. 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 just asking to be rejected:

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 acceptable paper:

An analysis of the frobnicatable foo filter.

In this paper we present a performance analysis of the paper of Smith *et al.* [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 and Jones, C. “The frobnicatable foo filter, a fundamental contribution to human knowledge”. Nature 381(12), 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 [2,] as additional material and cite it as

[1] Authors. “The frobnicatable foo filter”, F&G 2014 Submission ID 324, Supplied as additional material fg324.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 [2]”. Then submit the 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’s say it’s 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 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 *et al.*

You can handle this paper like any other. Don’t 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] didn’t 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, don’t you know. It displayed the following behaviours which show how well we solved cases A and B: ...

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 *et al.*, 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? No. Leave them for the final copy.

1.7. Miscellaneous

Compare the following:

$\$conf_a\$$ $conf_a$
 $\$\mathit{conf}_a\$$ $conf_a$

See The TeXbook, p165.

The space after *e.g.*, meaning “for example”, should not be a sentence-ending space. So *e.g.* is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using “et alia”, shortened to “*et al.*” (not “*et. al.*” as “*et*” is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: “Frobnication has been trendy lately. It was introduced by Alpher [2], and subsequently developed by Alpher and Fotheringham-Smythe [2], and Alpher *et al.* [2].”



Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

This is incorrect: “... subsequently developed by Alpher *et al.* [2] ...” because reference [2] has just two authors. If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al.*

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [2] to [2].

2. Motivation

2.1. Why do this

We found VSR to be a worthy problem for a couple of reasons. A VSR system can be used to recognize speech in scenarios where only visual data is kept, such as in Closed Circuit TV recordings. A more common use case than recognizing speech in video recording without audio is recognizing speech in a recording with a large amount of background noise. Background noise makes audio based speech recognition considerably more difficult, so a vision or hybrid speech recognition system could be more accurate in noisy scenarios [2].

2.2. What makes this difficult

We tried to build a system that can determine which words, from a predefined set of words, or dictionary, are being said. This kind of system could allow computer users to perform simple tasks, such as navigating to a specific website, by speaking, even in the presence of noise, such as while listening to music. Like most other computer vision problems, “VSR is Hard”. A VSR system must be able to determine the words that are being spoken based on the motion of a user’s mouth. This can actually be quite hard in the real world because of things like differences between the lighting or camera in the testing and training data. Recognizing the speech of real people can be diffi-

cult because of behaviors like moving their face. In the real world there also things that make tracking the motion of the mouth more difficult such as makeup, glasses, facial hair, or teeth. We believe that our system is resilient to these differences, and performs well in real conditions. One thing that is important for practical reasons that our system is capable of is recognizing the speech of a user without requiring training on data supplied by the user themselves. !!WHAT ABOUT USER NOT FACING THE CAMERA!! There are other issues that seem to be insurmountable for our system such as poor lighting and occlusion; we think a speech recognition system that also used audio data would be most robust in these scenarios.

3. Related Work

3.1. What has been done

Speech recognition has gone from academic research interest to an essential part of the way user interact with computing devices [3]. Most of these technologies rely solely on acoustic data. However, an obvious drawback of these systems is that their performance is severely degraded by noise [1]. Research has shown that systems that use visual data in place of, or in addition to audio data, can significantly outperform systems that only use audio data in the presence of noise [3].

3.2. What are the problems

Most other research, with a few exceptions focuses mainly on acoustic recognition, with visual recognition used as a secondary means of recognition. We thought that using a purely vision based approach would be an interesting challenge, and provide some of the advantages outlined above. Another difference between our approach and those of most other VSR systems is that most of them try to detect visual phenomes, or visemes, rather than actual words. While this approach is more generalizable and would allow detection of words that the system has not been explicitly trained on, we suspected that this approach may be less accurate for recognizing specific words than the approach we decided to take. One study that used this method was able to achieve 72.73% recognition of vowels, since the percentage would drop for specific words we felt the usefulness of this approach is quite limited for many use cases [?]. A final advantage our implementation has over nearly any other implementation is portability. Since our implementation can run in web browsers it can run on virtually every platform in common use today.

3.3. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high. Page numbers



Figure 2. Example of a short caption, which should be centered.

should be in footer with page numbers, centered and .75 inches from the bottom of the page and make it start at the correct page number rather than the 4321 in the example. To do this fine the line (around line 23)

```
%\ifcvprfinal\pagestyle{empty}\fi  
\setcounter{page}{4321}
```

where the number 4321 is your assigned starting page.

Make sure the first page is numbered by commenting out the first page being empty on line 46

```
%\thispagestyle{empty}
```

3.4. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and **AFFILIATION(s)** are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The **ABSTRACT** and **MAIN TEXT** are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 1 and 2. Short captions should be centred.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

3.5. Footnotes

Please use footnotes¹ sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

3.6. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [2]. Where appropriate, include the name(s) of editors of referenced books.

3.7. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of

¹This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one’s heart Frob

Table 1. Results. Ours is better.

the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in \LaTeX , it’s almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

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

3.8. Color

Please refer to the author guidelines on the CVPR 2015 web page for a discussion of the use of color in your document.

4. Algorithm

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.

References

- [1] M. Cooke, P. Green, L. Josifovski, and A. Vizinho. Robust automatic speech recognition with missing and unreliable acoustic data. *Speech Communication*, 34(3):267 – 285, 2001.
- [2] S. Dupont and J. Luettin. Audio-visual speech modeling for continuous speech recognition. *Journal of Foo*, 2014.
- [3] P. Lison and R. Meena. Spoken dialogue systems: The new frontier in human-computer interaction. *XRDS*, 21(1):46–51, Oct. 2014.
- [4] S. Werda, W. Mahdi, and A. B. Hamadou. Lip localization and viseme classification for visual speech recognition. *CoRR*, abs/1301.4558, 2013.