

Project Report : CS 7643

Dillon Reis
Georgia Institute of Technology
Institution1 address
dreis7@gatech.edu

Second Author
Institution2
First line of institution2 address
secondauthor@i2.org

Abstract

This project presents a model for inference of a wide array of flying objects in real-time. This remains challenging due to large variance of the object scales during inference, object rate of speed, and low variance between a subset of classes. These findings can then be used for reference/further research regarding object detection by remote digital airport towers, unmanned aerial vehicles (UAVs), or any surveillance systems utilizing optical data. To address some of the presented challenges, we utilize the current state of the art single-shot detector, YOLOv8, in an attempt to find the best trade-off between inference speed and mAP. RESULTS TBD

1. Introduction/Background/Motivation

(5 points) What did you try to do? What problem did you try to solve? Articulate your objectives using absolutely no jargon.

It is no surprise that drones and mini-UAVs still remain an integral part of modern warfare. They present a stealth capability and can avoid detection by radar due to their small electromagnetic signature. They are also small, highly maneuverable, and omit low levels of noise. While methods such utilizing radio and acoustic detection have been proposed as countermeasures, they are currently known to be inaccurate. This motivates the integration of a visual detector in any such detection system.

JACKIES STUFF

The primary objective of this project is to provide a model that detects a wide array of flying objects in real time with the latest state-of-the-art single shot detector YOLOv8. These results can be used as reference for further research by remote digital airport towers to detect aerial threats to a runway, surveillance systems for drone detection, and UAVs and for aircraft detection/threat identification. While YOLOv8 is being regarded as the new state-of-the-art, an official paper has yet to be released. This motivates our secondary objective, which is to explain the new architecture

and functionally that YOLOv8 has adapted.

Currently, single stage detectors are the de-facto architecture choice for fast inference speeds. This choice comes at the expensive of exchanging the higher accuracy you would typically expect from a two-state detector for the speed from a single-stage detector. Real-time object detection remains challenging due to variance in object spatial sizes and aspect ratios, inference speed, and noise. This is especially true for our use case, as flying objects can change location, scale, rotation, and trajectory very quickly. This conveys the necessity for fast inference speed (which for the purpose of this project is between 30 - 60 fps on 1080p video feed).

(5 points) How is it done today, and what are the limits of current practice?

(5 points) Who cares? If you are successful, what difference will it make?

(5 points) What data did you use? Provide details about your data, specifically choose the most important aspects of your data mentioned [here](#). You don't have to choose all of them, just the most relevant.

2. Approach

We chose the YOLOv8 architecture under the assumption that it would provide us the highest probability of success given the task. YoloV8 is assumed to be the new state-of-the-art due to its higher mAPs and lower inference speed on the COCO dataset. However, an official paper has yet to be released. It also specifically performs better of aerial objects, which is the scope of this project. We utilized the code repository from Ultralytics. We decide to implement transfer learning and initialize our models with pre-trained weights to then begin training on the custom data set. These weights are from a model trained on the COCO dataset. Due to only having access to a single NVIDIA RTX 3080 and 3070, a greedy model selection/hyper-parameter tuning approach was chosen. We first train a version of the small, medium, and large versions of the model with default hyper-parameters for 100 epochs. Then, we decide which model is optimal for our use case given the trade

off between inference speed and mAP-50-95 on the validation set. After model size is selected, a greedy hyper-parameter search is conducted with 10 epochs per each set of hyper-parameters. Finally, the model with the optimal hyper-parameters trains for 163 epochs to generate the final model.

Due to the large class imbalance, poor performance on the validation set was anticipated on the minority classes. However, this was not observed.

(10 points) What did you do exactly? How did you solve the problem? Why did you think it would be successful? Is anything new in your approach?

(5 points) What problems did you anticipate? What problems did you encounter? Did the very first thing you tried work?

Important: Mention any code repositories (with citations) or other sources that you used, and specifically what changes you made to them for your project.

3. Experiments and Results

(10 points) How did you measure success? What experiments were used? What were the results, both quantitative and qualitative? Did you succeed? Did you fail? Why? Justify your reasons with arguments supported by evidence and data.

Important: This section should be rigorous and thorough. Present detailed information about decision you made, why you made them, and any evidence/experimentation to back them up. This is especially true if you leveraged existing architectures, pre-trained models, and code (i.e. do not just show results of fine-tuning a pre-trained model without any analysis, claims/evidence, and conclusions, as that tends to not make a strong project).

4. Other Sections

You are welcome to introduce additional sections or sub-sections, if required, to address the following questions in detail.

(5 points) Appropriate use of figures / tables / visualizations. Are the ideas presented with appropriate illustration? Are the results presented clearly; are the important differences illustrated?

(5 points) Overall clarity. Is the manuscript self-contained? Can a peer who has also taken Deep Learning understand all of the points addressed above? Is sufficient detail provided?

(5 points) Finally, points will be distributed based on your understanding of how your project relates to Deep Learning. Here are some questions to think about:

What was the structure of your problem? How did the structure of your model reflect the structure of your prob-

lem?

What parts of your model had learned parameters (e.g., convolution layers) and what parts did not (e.g., post-processing classifier probabilities into decisions)?

What representations of input and output did the neural network expect? How was the data pre/post-processed? What was the loss function?

Did the model overfit? How well did the approach generalize?

What hyperparameters did the model have? How were they chosen? How did they affect performance? What optimizer was used?

What Deep Learning framework did you use?

What existing code or models did you start with and what did those starting points provide?

Briefly discuss potential future work that the research community could focus on to make improvements in the direction of your project's topic.

5. Work Division

Please add a section on the delegation of work among team members at the end of the report, in the form of a table and paragraph description. This and references do **NOT** count towards your page limit. An example has been provided in Table 1.

Student Name	Contributed Aspects	Details
Team Member 1	Data Creation and Implementation	Scraped the dataset for this project and trained the CNN of the encoder. Implemented attention mechanism to improve results.
Team Member 2	Implementation and Analysis	Trained the LSTM of the encoder and analyzed the results. Analyzed effect of number of nodes in hidden state. Implemented Convolutional LSTM.

Table 1. Contributions of team members.

6. Miscellaneous Information

The rest of the information in this format template has been adapted from CVPR 2020 and provides guidelines on the lower-level specifications regarding the paper’s format.

6.1. Language

All manuscripts must be in English.

6.2. Paper length

Papers, excluding the references section, must be no longer than six 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 six pages with two pages of references would have a total length of 8 pages.

6.3. The ruler

The \LaTeX 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- \LaTeX 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. (\LaTeX 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.

6.4. 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>.

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 [5]”. 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: ...

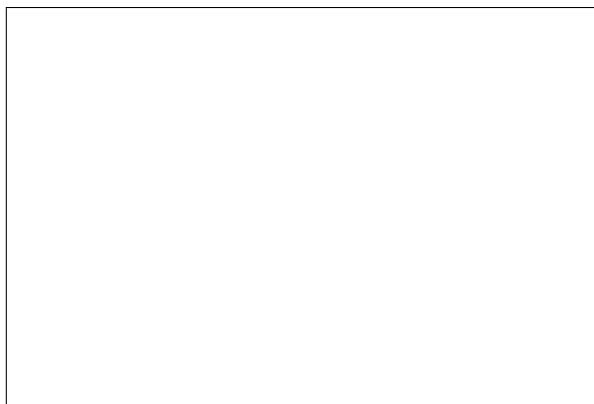


Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

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

Q: Are acknowledgements OK?

A: No. Leave them for the final copy.

Q: How do I cite my results reported in open challenges?

A: To conform with the double blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

6.5. Miscellaneous

Compare the following:

`$conf_a$` *conf_a*
`conf_a` *conf_a*

See The T_EXbook, 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: “Frobination has been trendy lately. It was introduced

by Alpher [1], and subsequently developed by Alpher and Fotheringham-Smythe [2], and Alpher *et al.* [3].”

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, 1, 4] to [1, 2, 4].

6.6. Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is $6\frac{7}{8}$ inches (17.5 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5 × 11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

6.7. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area $6\frac{7}{8}$ inches (17.5 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high.

6.8. 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.

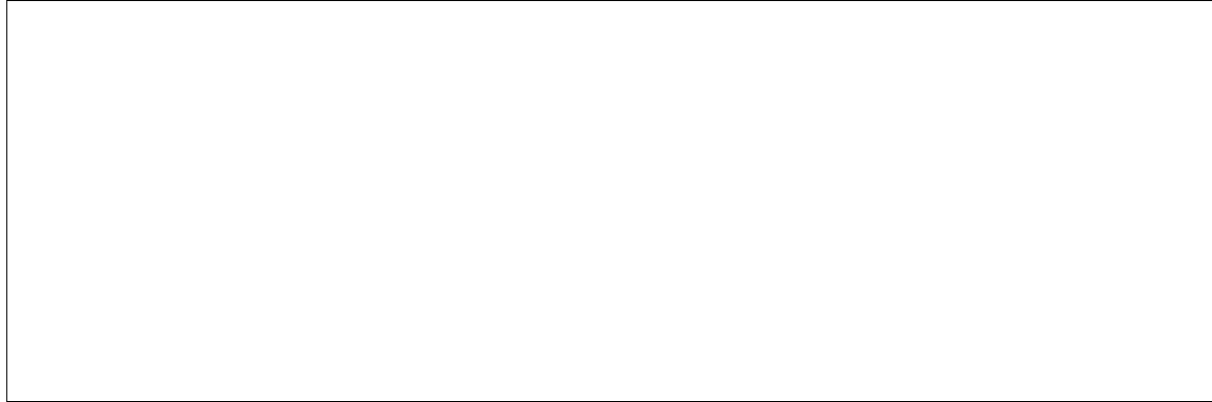


Figure 2. Example of a short caption, which should be centered.

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.

6.9. 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.

6.10. 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 [4]. Where appropriate, include the name(s) of editors of referenced books.

6.11. 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 2. 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 L^AT_EX, 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}
```

6.12. Color

Please refer to the author guidelines on the CVPR 2020 web page for a discussion of the use of color in your document.

References

- [1] FirstName Alpher. Frobnication. *Journal of Foo*, 12(1):234–778, 2002. 4
- [2] FirstName Alpher and FirstName Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003. 4
- [3] FirstName Alpher, FirstName Fotheringham-Smythe, and FirstName Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004. 4
- [4] Authors. The frobnicable foo filter, 2014. Face and Gesture submission ID 324. Supplied as additional material fg324.pdf. 4, 5

[5] Authors. Frobnication tutorial, 2014. Supplied as additional material `tr.pdf`. 3