Paper Writing Workshop

Karanveer Singh, Moayad Elamin, Shreyas Piplani

October 2022

Content

- Purpose of Paper Writing
- Types of Paper Writing
- Identifying Your Audience
- Defining Your Problem Statement
- Anatomy of a Research Paper
- Other Important Pieces of a Research Paper

Why do we write papers?

- Reporting experiments
- Summarizing findings that contributes new knowledge to a body of knowledge
- Aggregating information about historical work done within a body of knowledge or specific subdomain

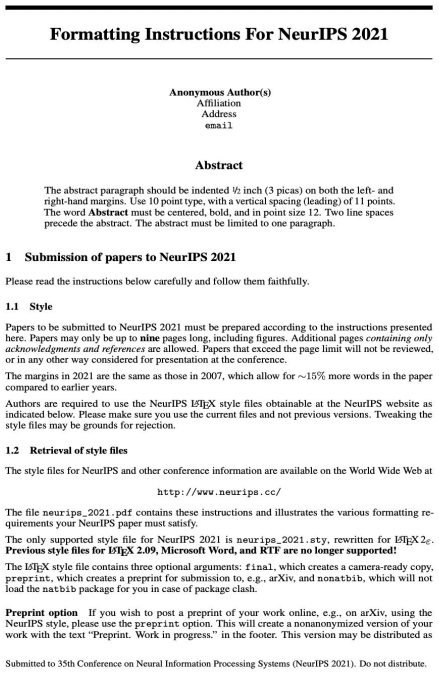
Paper Writing Tool

- Text Editor
  - MS Word
  - Latex
- Reference Tool
  - Mendeley
  - Zotero
- Spell Checker

Types of Paper Writing

- School Paper
  - Experiment Reports
- Conference/Workshop Paper
  - Surveys
  - Research Proposal
  - Research Report (Publishable Paper)

Paper Styles from Major Computing Conferences



Formatting Instructions For NeurIPS 2021

Anonymous Author(s)
Affiliation
Address
email

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Submission of papers to NeurIPS 2021

Please read the instructions below carefully and follow them faithfully.

1.1 Style

Papers to be submitted to NeurIPS 2021 must be prepared according to the instructions presented here. Papers may only be up to **nine** pages long, including figures. Additional pages *containing only acknowledgments and references* are allowed. Papers that exceed the page limit will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2021 are the same as those in 2007, which allow for ∼15% more words in the paper compared to earlier years.

Authors are required to use the NeurIPS LaTeX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

1.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the World Wide Web at

`http://www.neurips.cc/`

The file `neurips_2021.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

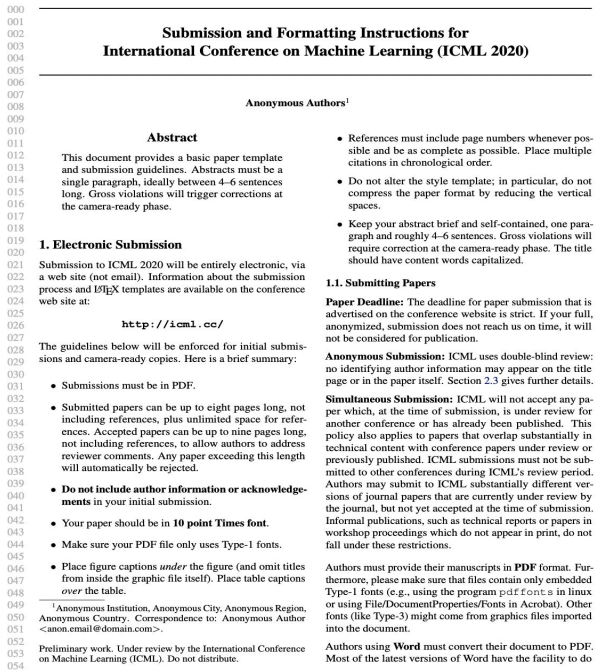
The only supported style file for NeurIPS 2021 is `neurips_2021.sty`, rewritten for LaTeX 2ε. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

The LaTeX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

Preprint option If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text "Preprint. Work in progress." in the footer. This version may be distributed as

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

[NeurIPS Format]



Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2020)

Anonymous Authors[1]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Electronic Submission

Submission to ICML 2020 will be entirely electronic, via a web site (not email). Information about the submission process and LaTeX templates are available on the conference web site at:

`http://icml.cc/`

The guidelines below will be enforced for initial submissions and camera-ready copies. Here is a brief summary:

- Submissions must be in PDF.
- Submitted papers can be up to eight pages long, not including references, plus unlimited space for references. Accepted papers can be up to nine pages long, not including references, to allow authors to address reviewer comments. Any paper exceeding this length will automatically be rejected.
- **Do not include author information or acknowledgements** in your initial submission.
- Your paper should be in **10 point Times font**.
- Make sure your PDF file only uses Type-1 fonts.
- Place figure captions *under* the figure (and omit titles from inside the graphic file itself). Place table captions *over* the table.
- References must include page numbers whenever possible and be as complete as possible. Place multiple citations in chronological order.
- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.
- Keep your abstract brief and self-contained, one paragraph and roughly 4–6 sentences. Gross violations will require correction at the camera-ready phase. The title should have content words capitalized.

1.1. Submitting Papers

Paper Deadline: The deadline for paper submission that is advertised on the conference website is strict. If your full, anonymized, submission does not reach us on time, it will not be considered for publication.

Anonymous Submission: ICML uses double-blind review: no identifying author information may appear on the title page or in the paper itself. Section 2.3 gives further details.

Simultaneous Submission: ICML will not accept any paper which, at the time of submission, is under review for another conference or has already been published. This policy also applies to papers that overlap substantially in technical content with conference papers under review or previously published. ICML submissions must not be submitted to other conferences during ICML's review period. Authors may submit to ICML substantially different versions of journal papers that are currently under review by the journal, but not yet accepted at the time of submission. Informal publications, such as technical reports or papers in workshop proceedings which do not appear in print, do not fall under these restrictions.

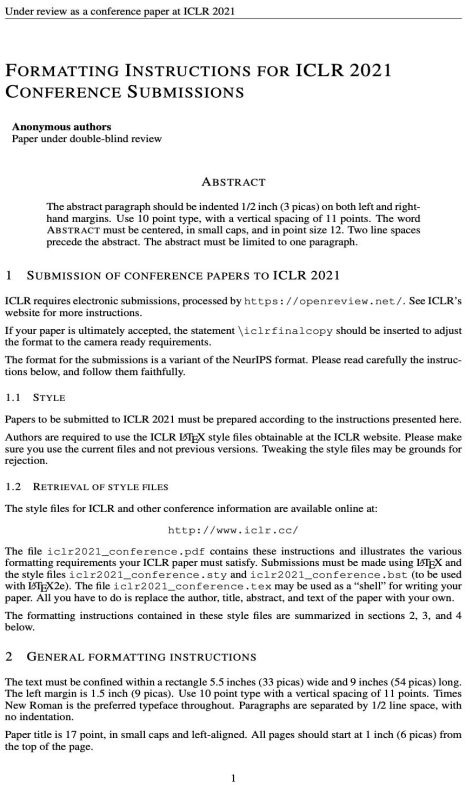
Authors must provide their manuscripts in **PDF** format. Furthermore, please make sure that files contain only embedded Type-1 fonts (e.g., using the program `pdffonts` in linux or using File/DocumentProperties/Fonts in Acrobat). Other fonts (like Type-3) might come from graphics files imported into the document.

Authors using **Word** must convert their document to PDF. Most of the latest versions of Word have the facility to do

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[ICML Format]



FORMATTING INSTRUCTIONS FOR ICLR 2021 CONFERENCE SUBMISSIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

The abstract paragraph should be indented 1/2 inch (3 picas) on both left and right-hand margins. Use 10 point type, with a vertical spacing of 11 points. The word ABSTRACT must be centered, in small caps, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 SUBMISSION OF CONFERENCE PAPERS TO ICLR 2021

ICLR requires electronic submissions, processed by `https://openreview.net/`. See ICLR's website for more instructions.

If your paper is ultimately accepted, the statement `\iclrfinalcopy` should be inserted to adjust the format to the camera ready requirements.

The format for the submissions is a variant of the NeurIPS format. Please read carefully the instructions below, and follow them faithfully.

1.1 STYLE

Papers to be submitted to ICLR 2021 must be prepared according to the instructions presented here.

Authors are required to use the ICLR LaTeX style files obtainable at the ICLR website. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

1.2 RETRIEVAL OF STYLE FILES

The style files for ICLR and other conference information are available online at:

`http://www.iclr.cc/`

The file `iclr2021_conference.pdf` contains these instructions and illustrates the various formatting requirements your ICLR paper must satisfy. Submissions must be made using LaTeX and the style files `iclr2021_conference.sty` and `iclr2021_conference.bst` (to be used with LaTeX2e). The file `iclr2021_conference.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in sections 2, 3, and 4 below.

2 GENERAL FORMATTING INSTRUCTIONS

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in small caps and left-aligned. All pages should start at 1 inch (6 picas) from the top of the page.

[ICLR Format]

Expectation for IDL Project Paper

- Use the NeurIPS paper format for your reports
- Structure your paper in accordance with the anatomy described in subsequent slides
- Include links to your code repositories or notebooks
- Reference and Include license to datasets (if not proprietary to you)

Knowing your audience helps to know the level of details to include in your research paper.

Who is your Audience?

- People with similar background or within similar domain as the author’s
- People that can reproduce the work you have done, following your provided appropriate methodology

Problem Statement Definition

- Must have element of novelty (new work)
- Must be feasible. How would the problem be solved?
- Must be ethical (approved by a designated ethics board)

Anatomy of a Technical Paper

- Abstract
- Introduction
- Related Work (Literature review)
- Methodology
  - Dataset Description
- Experiment, Results, Discussion
- Conclusion
- References

Anatomy of a Research Paper

ABSTRACT

- The shortest section of a paper about 100 - 150 words
- Executive Summary of Paper
- States the research problem/question
- Researcher’s contribution and answer to the research question
- How the answer was tested
- The impact of the research work
- Contains keywords about the research

Anatomy of a Research Paper - The Abstract

Addressing Function Approximation Error in Actor-Critic Methods

Scott Fujimoto [1] Herke van Hoof [2] David Meger [1]

1

2

arXiv:1802.09477v3 [cs.AI] 22 Oct 2018

Abstract

In value-based reinforcement learning methods such as deep Q-learning, function approximation errors are known to lead to overestimated value estimates and suboptimal policies. We show that this problem persists in an actor-critic setting and propose novel mechanisms to minimize its effects on both the actor and the critic. Our algorithm builds on Double Q-learning, by taking the minimum value between a pair of critics to limit overestimation. We draw the connection between target networks and overestimation bias, and suggest delaying policy updates to reduce per-update error and further improve performance. We evaluate our method on the suite of OpenAI gym tasks, outperforming the state of the art in every environment tested.

1. Introduction

In reinforcement learning problems with discrete action spaces, the issue of value overestimation as a result of function approximation errors is well-studied. However, similar issues with actor-critic methods in continuous control domains have been largely left untouched. In this paper, we show overestimation bias and the accumulation of error in temporal difference methods are present in an actor-critic setting. Our proposed method addresses these issues, and greatly outperforms the current state of the art.

Overestimation bias is a property of Q-learning in which the maximization of a noisy value estimate induces a consistent overestimation (Thrun & Schwartz, 1993). In a function approximation setting, this noise is unavoidable given the imprecision of the estimator. This inaccuracy is further exaggerated by the nature of temporal difference learning (Sutton, 1988), in which an estimate of the value function is updated using the estimate of a subsequent state. This means using an imprecise estimate within each update will lead to an accumulation of error. Due to overestimation bias, this accumulated error can cause arbitrarily bad states to be estimated as high value, resulting in suboptimal policy updates and divergent behavior.

This paper begins by establishing this overestimation property is also present for deterministic policy gradients (Silver et al., 2014), in the continuous control setting. Furthermore, we find the ubiquitous solution in the discrete action setting, Double DQN (Van Hasselt et al., 2016), to be ineffective in an actor-critic setting. During training, Double DQN estimates the value of the current policy with a separate target value function, allowing actions to be evaluated without maximization bias. Unfortunately, due to the slow-changing policy in an actor-critic setting, the current and target value estimates remain too similar to avoid maximization bias. This can be dealt with by adapting an older variant, Double Q-learning (Van Hasselt, 2010), to an actor-critic format by using a pair of independently trained critics. While this allows for a less biased value estimation, even an unbiased estimate with high variance can still lead to future overestimations in local regions of state space, which in turn can negatively affect the global policy. To address this concern, we propose a clipped Double Q-learning variant which leverages the notion that a value estimate suffering from overestimation bias can be used as an approximate upper-bound to the true value estimate. This favors underestimations, which do not tend to be propagated during learning, as actions with low value estimates are avoided by the policy.

Given the connection of noise to overestimation bias, this paper contains a number of components that address variance reduction. First, we show that target networks, a common approach in deep Q-learning methods, are critical for variance reduction by reducing the accumulation of errors. Second, to address the coupling of value and policy, we propose delaying policy updates until the value estimate has converged. Finally, we introduce a novel regularization strategy, where a SARSA-style update bootstraps similar action estimates to further reduce variance.

Our modifications are applied to the state of the art actor-critic method for continuous control, Deep Deterministic Policy Gradient algorithm (DDPG) (Lillicrap et al., 2015), to form the Twin Delayed Deep Deterministic policy gradient

[1]McGill University, Montreal, Canada [2]University of Amsterdam, Amsterdam, Netherlands. Correspondence to: Scott Fujimoto <scott.fujimoto@mail.mcgill.ca>.

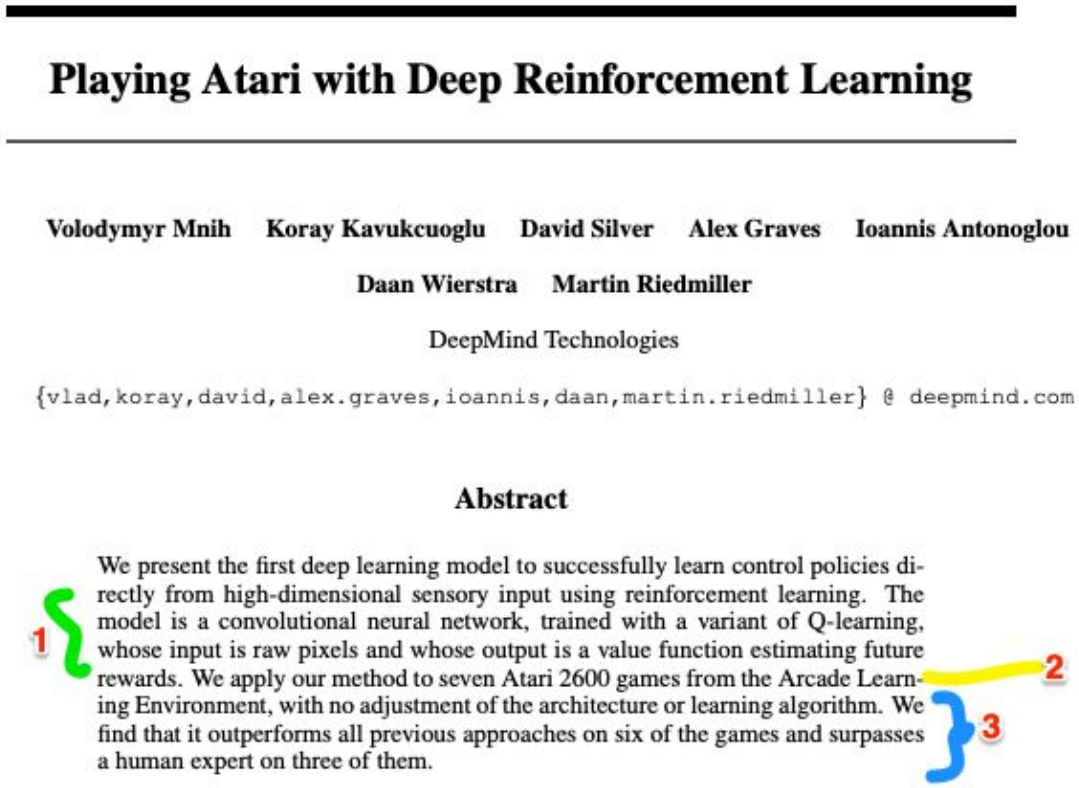
Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

1 - Highlights the Research Problem/Question

2 - Highlights **specific contribution** towards answering research question

Source: https://arxiv.org/pdf/1802.09477.pdf

Anatomy of a Research Paper - The Abstract



Playing Atari with Deep Reinforcement Learning

Volodymyr Mnih **Koray Kavukcuoglu** **David Silver** **Alex Graves** **Ioannis Antonoglou**

Daan Wierstra **Martin Riedmiller**

DeepMind Technologies

{vlad,koray,david,alex.graves,ioannis,daan,martin.riedmiller} @ deepmind.com

Abstract

We present the first deep learning model to successfully learn control policies directly from high-dimensional sensory input using reinforcement learning. The model is a convolutional neural network, trained with a variant of Q-learning, whose input is raw pixels and whose output is a value function estimating future rewards. We apply our method to seven Atari 2600 games from the Arcade Learning Environment, with no adjustment of the architecture or learning algorithm. We find that it outperforms all previous approaches on six of the games and surpasses a human expert on three of them.

1

2

3

1 - Researcher's contribution to answering research question

2 - How the research was tested

3 - Impact of the research work

Source: https://arxiv.org/pdf/1312.5602.pdf

Anatomy of a Research Paper

INTRODUCTION

- Extended form of the abstract.
- Gives background and sets tone for the research work
- Starts from broad issues to very specific area of research
- **Goal: Provides context to research question**

Writing a Great Introduction

- Summarize current understanding of research about the subject topic till date
- State the purpose of your research problem
- Highlight set of questions that would be answered by your research
- Briefly explain the methodology & what the study might reveal
- Summarize the structure of the remainder of the paper

Anatomy of a Research Paper

RELATED WORK

- This covers historical work done on the related research problems and/or related techniques.
- Categorize previous works into themes
- Summarize themes in a coherent format

Components of Literature Review

- Overview of the subject matter under consideration
- Categorize sources (related works) into different themes.
- Discuss the distinctiveness of each source and its similarities with other sources.

Anatomy of a Research Paper

METHODOLOGY

- Highly technical and contains technical jargon about the subject matter
- Covers overview of experiments to be done to answer the research question
- Reproducible to get documented results.
- Does not have to be named Methodology - dive straight to its components

Components of Methodology - Model Description

- Model Description
  - Describe clearly the model architecture
  - Cover the key areas about objective of the modelling approach
    - Minimizing a Loss function
    - Maximizing a Reward Function
  - Use mathematical expressions to describe the model as needed

Components of Methodology - Dataset Description

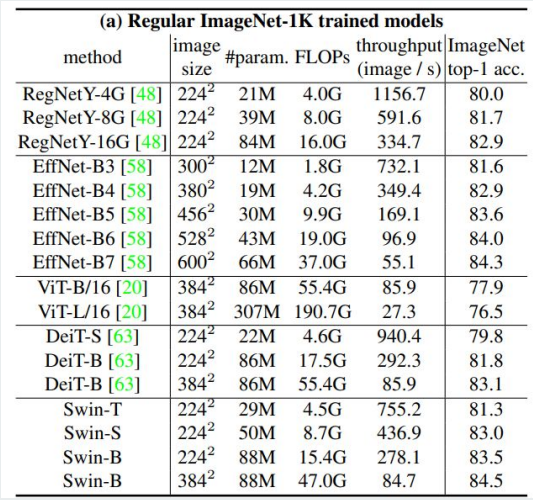
- What dataset would be used
    - Type: Speech, Image, Video, 3D point clouds, and so on
    - Data Mode: Single Mode, Multimodal
- How was the dataset was collected?
- Are there preprocessing done - either by you or from the data source
- How do you intend to use the data
- State overall statistics of the dataset e.g. length of dataset, training to validation proportion, etc

Components of Methodology - Baseline & Evaluation Metrics

- What baseline are you selecting?
  - Is this a state-of-the-art, competitive existing baseline
  - Are you implementing a baseline from scratch? Why?
- What evaluation metrics would you be using?
  - clearly define the metrics to be used
  - demonstrate understanding of how the metrics work
    - Give mathematical formulae if applicable

Anatomy of a Research Paper

EXPERIMENTS, RESULTS & DISCUSSION



(a) Regular ImageNet-1K trained models

method	image size	#param.	FLOPs	throughput (image / s)	ImageNet top-1 acc.
RegNetY-4G [48]	224^2	21M	4.0G	1156.7	80.0
RegNetY-8G [48]	224^2	39M	8.0G	591.6	81.7
RegNetY-16G [48]	224^2	84M	16.0G	334.7	82.9
EffNet-B3 [58]	300^2	12M	1.8G	732.1	81.6
EffNet-B4 [58]	380^2	19M	4.2G	349.4	82.9
EffNet-B5 [58]	456^2	30M	9.9G	169.1	83.6
EffNet-B6 [58]	528^2	43M	19.0G	96.9	84.0
EffNet-B7 [58]	600^2	66M	37.0G	55.1	84.3
ViT-B/16 [20]	384^2	86M	55.4G	85.9	77.9
ViT-L/16 [20]	384^2	307M	190.7G	27.3	76.5
DeiT-S [63]	224^2	22M	4.6G	940.4	79.8
DeiT-B [63]	224^2	86M	17.5G	292.3	81.8
DeiT-B [63]	384^2	86M	55.4G	85.9	83.1
Swin-T	224^2	29M	4.5G	755.2	81.3
Swin-S	224^2	50M	8.7G	436.9	83.0
Swin-B	224^2	88M	15.4G	278.1	83.5
Swin-B	384^2	88M	47.0G	84.7	84.5

Source: https://arxiv.org/pdf/2103.14030v2.pdf

- Experiments section should contain details about hyperparameters, how the dataset was used and how evaluation was performed.
- Standard to report metrics in comparison with other work
  - Make sure to report under similar settings. If necessary/possible, run multiple times and report standard deviations.

Anatomy of a Research Paper

EXPERIMENTS, RESULTS & DISCUSSION

(Ablations)

Sub-Pillar	PE	Tiny-Pillar	DF	SA	Veh.	Ped.	Cyc.	Mean
					60.32	49.22	51.58	53.71
✓					60.13	52.32	54.91	55.79
✓	✓				61.35	53.50	56.58	57.14
		✓			63.11	56.89	61.74	60.58
		✓	✓		67.33	59.24	65.06	63.88
		✓	✓	✓	67.22	59.34	65.82	64.12
✓	✓	✓	✓	✓	67.32	60.49	66.04	64.62

Source: https://arxiv.org/pdf/2110.06049v1.pdf

- Ablations are *extremely* important.
- How much did **each proposed component** contribute to the final performance?

Anatomy of a Research Paper

EXPERIMENTS, RESULTS & DISCUSSION

(Ablations)

Improvements	Top-1	Δ
ResNet-200	79.0	—
+ Cosine LR Decay	79.3	**+0.3**
+ Increase training epochs	78.8 †	-0.5
+ EMA of weights	79.1	**+0.3**
+ Label Smoothing	80.4	**+1.3**
+ Stochastic Depth	80.6	**+0.2**
+ RandAugment	81.0	**+0.4**
+ Dropout on FC	80.7 ‡	-0.3
+ Decrease weight decay	82.2	**+1.5**
+ Squeeze-and-Excitation	82.9	**+0.7**
+ ResNet-D	83.4	**+0.5**

Model	Regularization	Weight Decay 1e-4	Weight Decay 4e-5	Δ
ResNet-50	None	79.7	78.7	-1.0
ResNet-50	RA-LS	82.4	82.3	-0.1
ResNet-50	RA-LS-DO	82.2	82.7	+0.5
ResNet-200	None	82.5	81.7	-0.8
ResNet-200	RA-LS	85.2	84.9	-0.3
ResNet-200	RA-LS-SD-DO	85.3	85.5	+0.2

Source: https://arxiv.org/pdf/2103.07579v1.pdf

- Ablations are *extremely* important.
- How much did **each proposed component** contribute to the final performance?
- Be careful! Consider relationships *between* components as well. Sometimes each one can lower performance but together boost performance.

Anatomy of a Research Paper

EXPERIMENTS, RESULTS & DISCUSSION

(Discussion)

- Discussion does not need to be its own section. However, it should outline the overall takeaways from your work.
- Often, these takeaways are fused in with the results & ablations.

Anatomy of a Research Paper

CONCLUSION

- Summary of work done.
- Reiterates the research question and findings from the research work that answers the research question
- Highlights what possible shortcomings of current work if any, areas for improvement, next steps/future work, possible extensions.

Anatomy of a Research Paper

REFERENCES

"In reference works, as in sin, omission is as bad as willful behaviour"

- Elizabeth McCracken

Referencing Styles

Different Referencing styles depending on the type of paper you are writing and the conference/workshop.

- IEEE Reference style - Most common for engineering papers
    - Conference Paper, Books, Journals are referenced slightly differently
    - Comprehensive guide - https://owl.purdue.edu/owl/research_and_citation/ieee_style/reference_list.html
- APA - common for social sciences and Humanities disciplines
- MLA - common for English & Media Studies paper
- MHRA
- Harvard

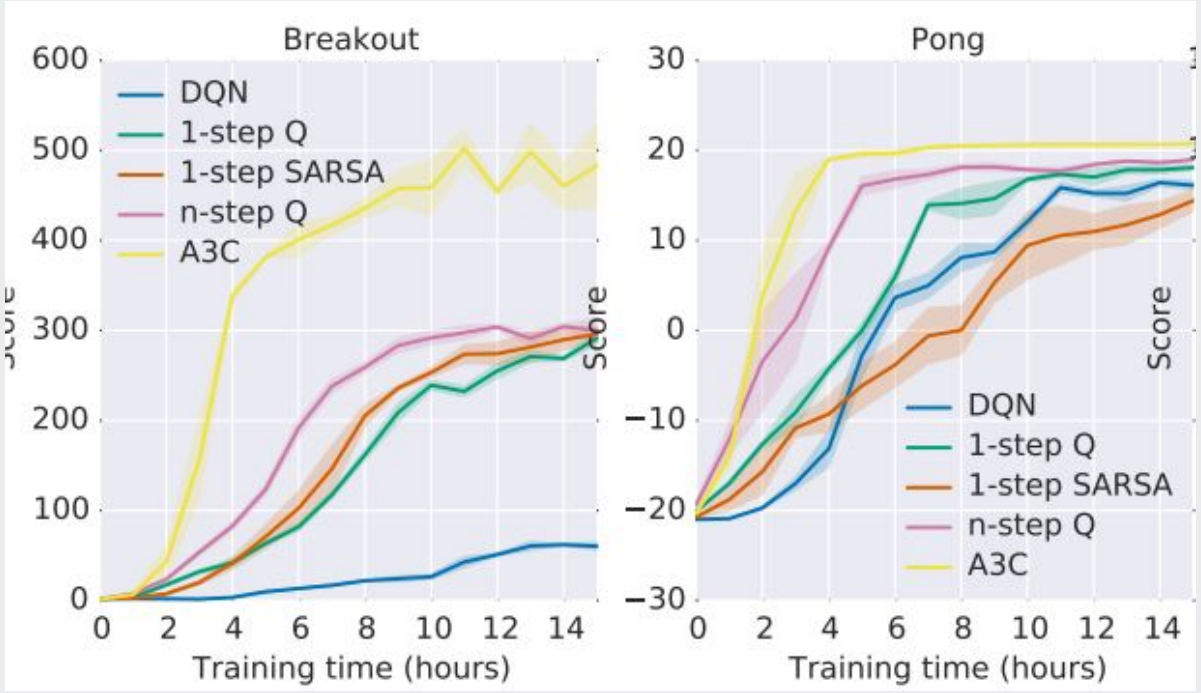
What other things are important?

- Reproducibility
- A good story/narrative
- General points
- “Troubling Trends in Machine Learning Scholarship”

Reproducibility

- When writing your paper, ask yourself: “can someone reproduce my results from reading this paper?”
- If the answer is no, you *need* to release code or add extra details in the paper or the supplementary.
- If your paper isn’t reproducible, 1) you’ll get emails, 2) the paper won’t be cited, 3) or the next paper will cite and say “we could not reproduce the results of this paper”

Reproducibility



- Whenever possible, include error bars/standard deviations.
- Extremely important in RL, where performance is usually unstable, but even on classification/object detection/language modeling/etc, if there is instability, should report average of multiple runs.

A good story/narrative

- Is there a point existing works have overlooked that you consider?
- Does previous work fail in certain cases (low-resource, certain classes) that your method is better suited for?
- Can you relate your work to some core intuitive notion that will give readers an “aha!” moment?

General points

- A research paper is an *argument* in support of your work, not just a list of observations.
- The reader should be **convinced** that your method is better than existing work.

General points

- Try not to make up your own technical terms or make things seem unnecessarily complicated.
- Although it might seem cool, it can confuse readers or annoy ones who are not awed.
- Bad example: “We extract complementary semantics via feature entanglement over space and time.”
- Better example: “Given RGB video, we run a first CNN over the temporal dimension, run a second CNN over the spatial dimensions, and then combine their results.”

“Troubling Trends in Machine Learning Scholarship”

- This is a paper from 2018 that has been trending recently.
- It details some common (bad) trends in ML papers that impede understanding/add noise to research.

Source: https://arxiv.org/pdf/1807.03341v2.pdf

https://arxiv.org/pdf/1509.02971.pdf

“Troubling Trends in Machine Learning Scholarship”

“Explanation vs Speculation”

- Papers should explicitly make clear what is formal reasoning and informal speculation.
- For instance, internal-covariate shift is not clearly stated and has been repeated as fact when it could be false.
- Other papers, as positive examples, have quarantined a "motivation" section for informal intuitions.

Source: https://arxiv.org/pdf/1807.03341v2.pdf

“Troubling Trends in Machine Learning Scholarship”

“Failure to Identify the Sources of Empirical Gains”

- Ablations are very important.
- Some papers do not clarify the sources of gains in performance, and other papers have found that actually, hyper-parameter tuning was a more important factor.

Source: https://arxiv.org/pdf/1807.03341v2.pdf

“Troubling Trends in Machine Learning Scholarship”

Misuse of Language: Suggestive Definitions

- Some works define equations in a manner that imbues them with human-level priors such as "curiosity" or "fear." Others overstate/incorrectly claim "human-level" performance.
- This is confusing and dangerous for the field.
- What the agent learns is not necessarily the human notion of curiosity - using anthropogenic terms gives laymen unrealistic expectations about AI.

Source: https://arxiv.org/pdf/1807.03341v2.pdf

References

- https://library.concordia.ca/help/writing/literature-review.php?guid=components