Formatting Instructions For NeurIPS 2024

David S. Hippocampus*
Department of Computer Science
Cranberry-Lemon University
Pittsburgh, PA 15213
hippo@cs.cranberry-lemon.edu

Abstract

This paper describes improvements on a base Convolutional Neural Network through data augmentation, hyperparameter tuning, and changing model architecture in order to better derive keystrokes from sEMG data, emphasizing techniques that keep minimal training, cost, and hardware to achieve. Taking inspiration from automatic speech recognition research, integrating both a CNN + LSTM hybrid model and transformer architecture resulted in a 2% lower CER when trained for only 30 epochs, compared to the base model training for 150 epochs. These results apply to individualized sEMG readings, with the assumption that for any base model, there will need to be a calibration and training period, necessitating a model small enough to fit on consumer grade hardware and achieve comparable performance in a fraction of the training time.

1 Introduction

With the resurgence of AI in everyday consumer lives, existing tech companies have been scrambling to capitalize on the wide use cases AI presents. Adding AI summaries, chatbots on every landing page, and even building their own AI centered hardware chips means that billions have been dumped into AI research. Before even a dollar of this profit is realized, these companies rigorously train models to get an acceptable error rate, often costing a fortune.

However, as tech giants continue to throw giant sums of money at the problem, some disruptive companies like Deep-Seek claim to rival the performance of giants like OpenAI, at a fraction of the cost. Though these claims may be exaggerated, the constraints can often force innovation. By optimizing performance with a hardware and dataset constraint, our goal was to get comparable performance to the base paper that used an A10 GPU for 150 epochs by using a 4060 Mobile and only 30 epochs.

Given the challenge of limited amounts of data and limited hardware, we opted for utilizing data augmentation in order to limit overfitting on the limited amount of data, exploring learning rate schedulers that optimize for speedy convergence, and finally utilizing smaller models. These challenges may prove to be realistic as real-world applications of sEMG to keystroke data may require a quick training and calibration time to individualize a model.

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

2 Methods

2.1 Data Augmentation

In deep learning architectures, there is a very strong correlation between the amount of training data and the efficacy of a model. However, availability of clean and plentiful datasets proves to be a challenge for many applications of deep learning, introducing the problem of "inadequate data availability" **?** leading to the need for synthetic data generation from data augmentation. However the effectiveness of data augmentation techniques can actually vary wildly between different signals and use cases. Especially for sEMG data, the noise and variability present between individuals makes it hard to have correct generalizations from models. They point specifically to non-gaussian noise being relatively ineffective while a sliding window approach leads to a large increase.

In order to experiment with different kinds of data augmentation techniques, ChatGPT suggested frequency shifting, cropping, masking, dropout, and time warping. While the base model already had some of these transformations in it, we explored the individual effects of the data augmentation on the data. The most prominent data augmentation techniques were time warping for CNNs and changing the Log Spectrogram to a Mel Spectrogram. While effective, the sliding window approach was also already implemented in the EMG data module.

Some ineffective data augmentation strategies we found were frequency shifting, cropping, masking, and dropout. Some intuition for why this may be could be that some of these strategies were already implemented into the built-in transformations. Adding duplicate transformations may have augmented the data too much such that relevant patterns were unable to be detected by the models. Cropping especially may have been ineffective as improper window sizes may improperly display the data, cutting out crucial context or framing data such that artificial edges caused by noises indicate a different **?** pattern than the general trend.

While these data augmentation strategies were not as effective, by using a special kind of time warping **?**, we were able to see a 5% improvement compared the base model CNN. While typical time warping can compress the data, altering it's meaning, this paper used space in between features to add warped data, while keeping the endpoints the same. This can make classification more clear by creating new boundaries. Since there wasn't too much time to truly optimize the hyperparameters, the fact that unoptimized time warping was still able to improve on the baseline model may suggest that with better hardware and more time, this method of time warping could significantly benefit synthetic data generation through data augmentation. However, an interesting caveat is that this improvement did not hold when switching to a hybrid CNN + LSTM model and transformer model. A reasoning for this may be that LSTM and transformer models capture temporal data better than a CNN does. Therefore, the augmentation that more clearly divides different classes of keystrokes temporally helps a CNN a lot, while the LSTM and transformer have already identified temporal features.

Another method that may show more promising results if properly trained was dropout. Though this was already implemented in another baseline transformation, dropout or masking is an effective transformation as small data sets can cause models to overfit and generalize poorly. By dropping some features, the model generalizes better by not overfitting to every part of the data, intuitively finding the most features. By implementing dropout we only saw a small increase in CER compared to the base model, so with more fine-tuning, greater performance can be accomplished through dropout.

Another efficient data transformation method was replacing the LogSpectrogram with a non-log Mel Spectrogram. Used primarily with decomposing features from sound waves **?**, a Mel spectrogram spaces lower frequencies densely and higher frequencies sparsely, mirroring how the human ear perceives sound. Through research being done on sEMG data relating to muscle activity, researchers have found a correlation between a constant, "always similar temporal behavior" in low-frequency components, while high frequencies indicate faster moving motions **?**, which we can extrapolate to be typing for this project. Therefore, a Mel spectrogram that compresses lower level frequencies that contain constant information despite changes in motor activity, while emphasizing the higher frequencies that indicate fine motor movement. In addition, though a log transformation the data more, keeping the data in it's original amplitude can lead to the model to be more sensitive to the original amplitude, as to not underfit and find nuances between smaller differences in sEMG data.

2.2 Learning Schedulers

The baseline model for our learning scheduler was a linear warmup cosine annealing learning rate scheduler, which we used to when testing transformations. While cosine annealing learning rates are stable and gradually increase learning rate, then later taper off for finding a minima, we found that a one cycle learning rate **?** drastically outperformed the baseline learning scheduler due to its optimization for training an order of magnitude faster than traditional methods.

Used with the Cifar-10 dataset and the same architecture, traditional methods reached a peak accuracy of 91.2% after 80,000 iterations, while the one cycle learning was able to get 92.4

Fig 1: Graph from one cycle, learning rate vs loss

After graphing the loss versus the learning rate, we saw that the range of 10^{-3} to 10^{-5} had the steepest slope and was a good range for our global minima and maxima

2.3 Architectures

While the base model of a TDS-Conv Network has proven to be effective with sound wave recognition **?**, we attempted to use different architectures in order to bring the CER down. While exploring other architectures, a main concern was how the architecture uses temporal data to relate different features in the data.

The base model is a Convolutional Neural Network utilizing CTC loss. CNNs are designed to capture spatial features from data. They apply convolutional filters to input data, which makes them particularly effective at detecting local patterns like edges in images or short-term features in waveforms. This makes CNNs well-suited for identifying localized patterns in sound data, but can struggle with capturing long-term dependencies.

A second type of model is a Long Short-Term Memory network, a type of recurrent neural networks that are optimized to handle sequential data. Its architecture allows it to have memory cells that learn what information to store and what to filter out. This allows LSTMs to process temporal data with an ability to look back and store only crucial features of the data.

The first architecture we attempted was replacing the convolutional encoder with an LSTM encoder. A bare LSTM didn't significantly improve the base model, increasing the CER by only 0.24%. However, this small difference suggested that making the LSTM architecture more sophisticated could yield promising results. One improvement we made was implementing a dual-encoder model combining an LSTM and a CNN. By chaining these encoders, we decreased our CER by 4%. The intuition behind this approach is that while the CNN captures spatial features within a window, the LSTM identifies temporal relationships across sequences, resulting in a model that can effectively capture both spatial and temporal dependencies. We further enhanced the LSTM by integrating attention head mechanisms **?**. Attention mechanisms allow models to focus on different parts of the input data independently and weigh their importance when making predictions. By implementing multi-head attention, the LSTM could independently identify distinct features in the data and combine them in a softmax classifier. While attention isn't integrated directly into a base LSTM model, this experiment introduced us to transformer architectures, which natively implement attention mechanisms. Transformers have gained popularity for their effectiveness in handling long-range dependencies. Unlike CNNs, which focus on local patterns, and LSTMs, which process data sequentially, transformers use self-attention to weigh the importance of different parts of the input sequence simultaneously. This parallelization makes them more efficient at capturing both short-term and long-term relationships in data. Previous research in automatic speech recognition **?** found that transformers consistently achieve the lowest error rates, making them an attractive option for our model. Inspired by these findings, we developed a transformer model with 4 layers and 384 features. Despite their larger parameter size compared to LSTMs and CNNs, transformers minimize computational overhead through parallel processing. This efficiency allows them to run on consumer-grade graphics cards (4060 Mobile) while maintaining strong performance, making them a practical choice for our research. Our exploration of these architectures highlights the balance between spatial and temporal feature extraction, and the trade-offs between model complexity and performance. By combining the strengths of CNNs, LSTMs, and transformers, we continue to refine our approach to lowering CER and improving sound wave recognition.

Table 1: Model Performance Across Different Tunings and Epochs

Model	Tunings	30 Epochs [Val/Test]	150 Epochs [Val/Test]	150 Epochs 6
CNN	Base	27.603/28.052	19.207/21.872	
CNN	Time Warping	25.188/22.757		
CNN	One Cycle LR	23.416/24.724		
CNN	AdamW + OneCycle	23.128/25.070		
LSTM	AdamW + OneCycle	24.767/25.243		
CNN+ LSTM	AdamW + OneCycle	19.782/21.180		
CNN+ LSTM	Prev + Log Mel Spec	19.140/20.726		
CNN+ LSTM*	Prev + Mel Spec + Frequency Adj	16.438/19.294**	13.646/16.522**	
CNN + Transformer	Prev + Mel Spec	17.678/19.121		
CNN + Transformer	Prev + Mel Spec + Frequency Adj	16.902/19.099	16.580/	10.
Transformer (no conv)	P̂rev + Mel Spec + Frequency Adj	57.066/		

3 Results

As seen from the table, our data augmentations and non-architecture changes were able to bring down CER by around 3%. Naive implementations of a LSTM proved to not be significantly better, but through a CNN + LSTM hybrid model we were able to bring down CER by an additional 3%. Additional data augmentations of the Mel Spectrogram were able to reduce by 1-2% more, finally leaving us with a hybrid CNN + LSTM reaching around 19.3% CER. Through implementing attention through a transformer model, this rate was not significantly reduced. In fact, when training both of these models to 150 epochs, we found that the CNN + LSTM hybrid model slightly outperformed the transformer architecture. For a 150 epoch comparison, the base model reached an accuracy of 21% CER while our models resulted in around 16.5% CER. This remains impressive as one compares the base model at 150 epochs still has a 2% higher CER than the CNN + LSTM hybrid and transformer model at only 30 epochs.

3.1 Limitations

Our biggest limitation in creating personalized models is that the training, validation, and test set were not big enough. Though we used data augmentation to limit the overfitting and produced acceptable results, the accuracy we present through the CER findings is based on one sample for our validation and one sample for our test data. Without other individuals to compare our models across, the accuracy we found may not generalize well to other individuals.

This extends to how we decide the efficacy of our models. Through training the models to 30 epochs, we chose the most effective models of the CNN + LSTM hybrid model and the transformer model and chose to extend the same training to 150 epochs. Other than changing the hyperparameters for the one cycle learning rate, most other hyperparameters were frozen. Had they been optimized for 150 epochs from the start, we may expect to see greater performance.

In addition, transformations were frozen while changing models. We had initially benchmarked each transformation based off of the pure CNN model and decided to tune the hyperparameters of the existing transformations based on it. However, different models interpret spatial and temporal local data differently. While a Mel spectrogram may have assisted the CNN model as it struggles with finding temporal features, it may have skewed performance on an LSTM or transformer architecture.

4 Conclusion and Future Implementations

Through our exploration of smaller models on consumer hardware, we demonstrate how to get a higher accuracy in decoding sEMG signals for keyboard typing at a fraction of the time, size, and compute power. The biggest improvement is through data augmentation and different model choices, two of our models were able to get 2% better CER than the base model trained for 5 times longer. Although inference response time is typically prioritized over training time, for applications as individualistic as sEMG signals, on-demand training is essential to recalibrate the model for each new individual. For practical applications, the model and training may have to live on device with a quick training time such that after a brief calibration and training period, a new user can begin typing with just sEMG signal receptors.

Future work may be done in quick hyperparameter searches based on the data. Fine tuning the hyperparameters could lead to drastically different and more efficient results, improving the performance of these small models. In addition, integrating hardware capable of housing this model in these sEMG readers directly may be an interesting practical application of this paper.

The style files for NeurIPS and other conference information are available on the website at

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

The file `neurips_2024.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

The only supported style file for NeurIPS 2024 is `neurips_2024.sty`, rewritten for LaTeX 2_ε. **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 you see fit, as long as you do not say which conference it was submitted to. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS.

At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `neurips_2024.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 **??**, **??**, and **??** below.

5 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 (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section **??** regarding figures, tables, acknowledgments, and references.

6 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

6.1 Headings: second level

Second-level headings should be in 10-point type.

6.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

7 Citations, figures, tables, references

These instructions apply to everyone.

7.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated…

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2024` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2024}
```

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" and include a copy of the anonymized paper in the supplementary material.

7.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[2] 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).

Note that footnotes are properly typeset *after* punctuation marks.[3]

---

[2] Sample of the first footnote.
[3] As in this example.

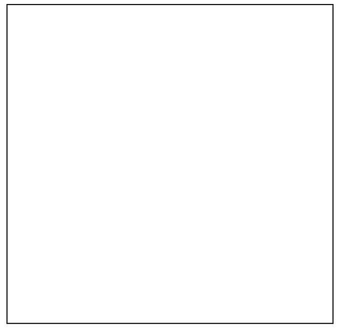

Figure 1: Sample figure caption.

Table 2: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

7.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. 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 should be lower case (except for first word and proper nouns); figures are numbered consecutively.

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

7.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table **??**.

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.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table **??**.

7.5 Math

Note that display math in bare TeX commands will not create correct line numbers for submission. Please use LaTeX (or AMSTeX) commands for unnumbered display math. (You really shouldn't be using $$ anyway; see `https://tex.stackexchange.com/questions/503/why-is-preferable-to` and `https://tex.stackexchange.com/questions/40492/what-are-the-differences-between-align-equation-and-displaymath` for more information.)

7.6 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.

8 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- 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 `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

  ```
  \newcommand{\RR}{I\!\!R} %real numbers
  \newcommand{\Nat}{I\!\!N} %natural numbers
  \newcommand{\CC}{I\!\!\!\!C} %complex numbers
  ```

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

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

8.1 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:

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

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

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

Acknowledgments and Disclosure of Funding

Use unnumbered first level headings for the acknowledgments. All acknowledgments go at the end of the paper before the list of references. Moreover, you are required to declare funding (financial activities supporting the submitted work) and competing interests (related financial activities outside the submitted work). More information about this disclosure can be found at: `https://neurips.cc/Conferences/2024/PaperInformation/FundingDisclosure`.

Do **not** include this section in the anonymized submission, only in the final paper. You can use the `ack` environment provided in the style file to automatically hide this section in the anonymized submission.

References

References follow the acknowledgments in the camera-ready paper. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. Note that the Reference section does not count towards the page limit.

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.

A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD be included in the main submission.**

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: **[TODO]**

   Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: **[TODO]**

    Justification: **[TODO]**

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: **[TODO]**

    Justification: **[TODO]**

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.