Need is All You Need: Homeostatic Neural Networks Adapt to Concept Shift

Anonymous Author(s)
Affiliation
Address
email

Abstract

In living organisms, homeostasis is the natural regulation of internal states aimed at maintaining conditions compatible with life. Here, we introduce an artificial neural network that incorporates some homeostatic features. Its own computing substrate is placed in a needful and vulnerable relation to the very objects over which it computes. For example, MNIST digits may cause excitatory or inhibitory effects upon the homeostatic network that classifies them, by altering the network's learning rate. Accurate recognition is desirable to the agent itself because it guides decisions to up- or down-regulate its internal states and functionality. Counter-intuitively, the addition of vulnerability to a learner can confer some benefits. Homeostatic learners are more adaptive under conditions of concept shift, in which the relationships between labels and data change over time. The greatest advantages are obtained under the highest rates of shift. Homeostatic learners are also resilient to second-order shift, or environments with changing rates of concept shift.

1 Introduction

To paraphrase Heraclitus, "The only constant in life is change". The rules and relationships learned today may no longer hold tomorrow. Un-learning the bad old rules, re-learning the good new ones, and knowing how to tell the difference remains a major challenge for learning machines. Here we are inspired by the natural intelligence of living organisms, which maintain themselves in the face of environmental change by following the dictates of homeostasis. Homeostasis is the regulation of internal body states within a range compatible with life. It has been proposed that (a) machines that implement a process resembling homeostasis could be designed to exhibit a feeling-like device for the motivation and evaluation of their behavior and that (b) equipping an artificial learner with a feeling-like device might improve its adaptiveness to the inconstant data streams of the real world (Man and Damasio 2019).

Here we present a homeostatic neural network architecture in which a classifier is placed into a needful and vulnerable relation to the objects over which it computes. By way of analogy, the homeostatic agent must learn to distinguish between cups of coffee and cups of beer, while also needing to take a drink every so often to regulate its own mental arousal. In this setting, accurate classification is desirable to the agent itself because it guides decisions that can carry consequences for its internal states.

2 Background

In biological brains, neurons regulate their excitability and synaptic conductance to stabilize network function (Marder and Goaillard 2006). In artificial neural networks, homeostatic regulation of excitability can reduce saturation and improve signal propagation (Williams and Noble 2007). In simulation studies of evolutionary robotics, phototactic robots used 'neural plasticity' to restore adaptive behavior following visual field inversion (Di Paolo 2000; Iizuka and Di Paolo 2008). However, the homeostatic-like features of prior works were implemented from the outside-in: systems were instructed to maximize, or keep within a set range, certain arbitrary values that were labeled "homeostatic". The operation of the system itself was not exposed to the consequences of the system's own activities, that is, it was not made vulnerable to the world and therefore sensitive to changes in the world.

Non-stationarity, or "changes in the world", poses a major challenge in machine learning. Learners can fail to generalize because of concept shift (Moreno-Torres et al 2012), in which the associations between labels y and observations x change across the training and testing phases: $P_{train}(y|x) \neq P_{test}(y|x)$. This phenomenon occurs frequently in real world settings of online supervised learning; for example, recommender systems must stay current with their users' evolving tastes.

3 Homeostatic architecture of needful neural networks

Our homeostatic agent learns to classify images of objects. In a twist, the learner is designed to be needful – it depends on the objects that it classifies for its continued integrity and functionality. The objects have direct effects, excitatory or inhibitory, on the learner itself. For example, in MNIST classification, the digits $\{0,1,2,3,4\}$ have inhibitory effects and reduce the learning rate (LR), while the digits $\{5,6,7,8,9\}$ have excitatory effects and increase the LR. Critically, following classification of an object, the learner decides to either "ingest" the digit and alter its own learning rate, or "reject" the digit and keep its current learning rate. We use a counterfactual decision process to answer the question, "How would my own functionality be affected by taking or leaving this object?" The learner evaluates each alternative by simulating both versions of itself and testing them against a store of recently seen objects and labels (Supplementary Fig. 1). Misperceiving an object can lead to performing the wrong simulation of the object's effects on the learner. This will drive the wrong LR decision, further destabilizing future perceptions.

3.1 The vicissitudes of life

As so often happens in life, the rules have a way of changing on you. We introduce concept shift by permuting labels on a subset of the data. When a shift occurs, we swap the labels for two randomly selected classes. For the MNIST example, we may swap the labels "zero" and "nine", such that all images that look like "0" are now labelled "nine", and all images that look like "9" are now labelled "zero" (Supplementary Fig. 2). Note that in this swap, the homeostatic effects of the digits have also been reversed. Images that were previously inhibitory (image 0 $\rightarrow$ label "zero" $\rightarrow$ inhibitory) are now excitatory (image 0 $\rightarrow$ label "nine" $\rightarrow$ excitatory), and vice versa.

4 Experiments

We compare homeostatic regulation of LR against two control conditions: a randomly regulated "wandering" learning rate, and a more conventional, constant learning rate. We characterize the conditions under which homeostatic regulation either imposes a performance penalty, or else allows a learner to smoothly adapt to changing conditions. All classification studies are performed with a multilayer perceptron with two hidden layers containing 80 and 60 units respectively, using the ELU activation function (Clevert et al 2015) and He initialization (He et al 2015). We evaluate our method on two datasets, MNIST (Lecun et al 2010) and Fashion-MNIST (Xiao et al 2017). All experiments were performed in MATLAB and the source code is provided in the Appendix.

4.1 The homeostatic learner adapts to concept shift

Testing across a wide range of rates of concept shift, measured in swaps per epoch of training, we find that in the stationary setting (no swapping) the conventional, constant-LR classifier is most accurate (Fig. 1, red traces in far left column). The homeostat (blue traces) nearly matches the constant-LR classifier's performance, which is remarkable because the homeostat has the seeming disadvantage of being vulnerable to its own mistakes. Illustrating how badly things could have gone, the randomly regulating LR classifier (green traces) goes off the rails and shows large variance across replicates.

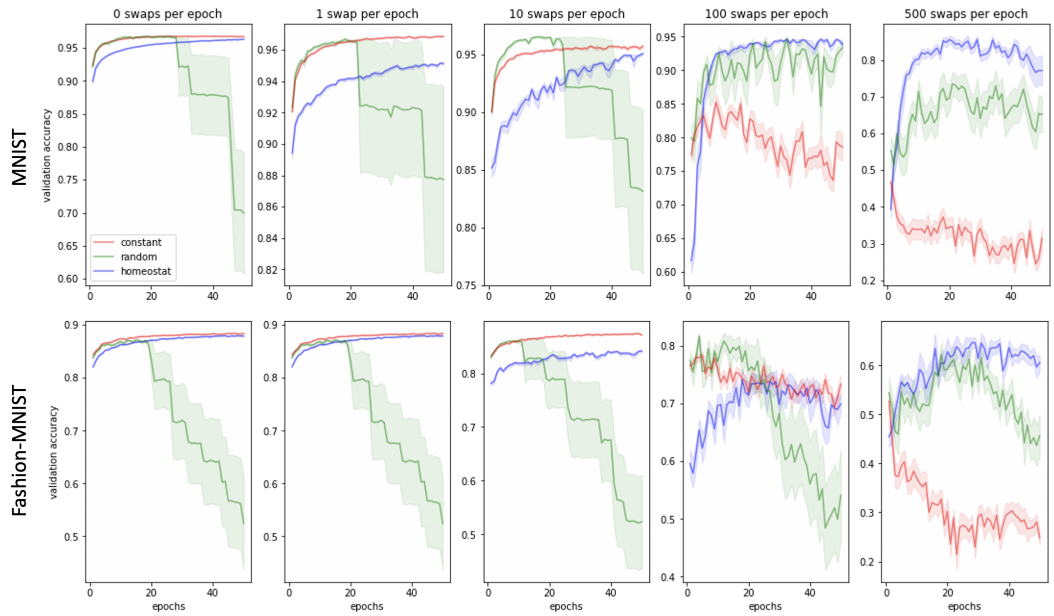

Figure 1: Homeostatic learners incur some performance penalty in environments with no or low concept shift, but are far superior under conditions of highest shift. Color-coded validation accuracies of learners with their learning rates homeostatically regulated (blue), randomly regulated (green), and held constant (red). Traces show mean +/- SEM over 20 replicates.

The benefits of the homeostatic architecture become apparent at the highest intensities of concept shift (Fig. 1, right columns). At 500 swaps per epoch, the constant-LR classifier is overcome by change and falls to near chance level. The homeostat, on the other hand, is able to learn and even to improve despite extreme rates of concept shift.

4.2 Homeostatic LR regulation is responsive to the prevailing rates of concept shift

The homeostat tunes its learning rate to a level specific to the environment in which it finds itself (Supplementary Fig. 3). All learning rates are initialized at the same value (0.005) but the homeostat seeks a LR appropriate to the experienced rate of concept shift. It arrests its own rise in high shift environments and converges upon stable LR values.

4.3 The homeostatic learner adapts to second-order shifts, or "seasonality"

We next created learning environments with seasonality, in which the rate of concept shift can vary over the course of training. We find that the homeostat maintains the most consistent performance across "calm" and "stormy" seasons, and rapidly recovers after the onset of a stormy period (Fig. 2).

5 Discussion

To summarize, we show that: 1) homeostatic learners are superior to conventional learners under concept shift, with the greatest advantage obtained under the greatest rates of shift; 2) homeostatic regulation imposes a slight performance penalty under static and low-shift environments; 3) homeostatic learners tune their learning rate in accordance with environmental conditions; and 4) homeostatic learners can adapt to second-order shift, or changes in the rate of environmental change. Although we find these converging results across the MNIST and Fashion-MNIST datasets we note that one possible limitation on the scope of our claims is the use of only these two datasets, each being somewhat limited in visual complexity and image size.

Another possible limitation is the re-use of training data over many epochs, which limits the fundamental novelty of the concept shift. Although the labels and data are repeatedly shuffled, the classifier

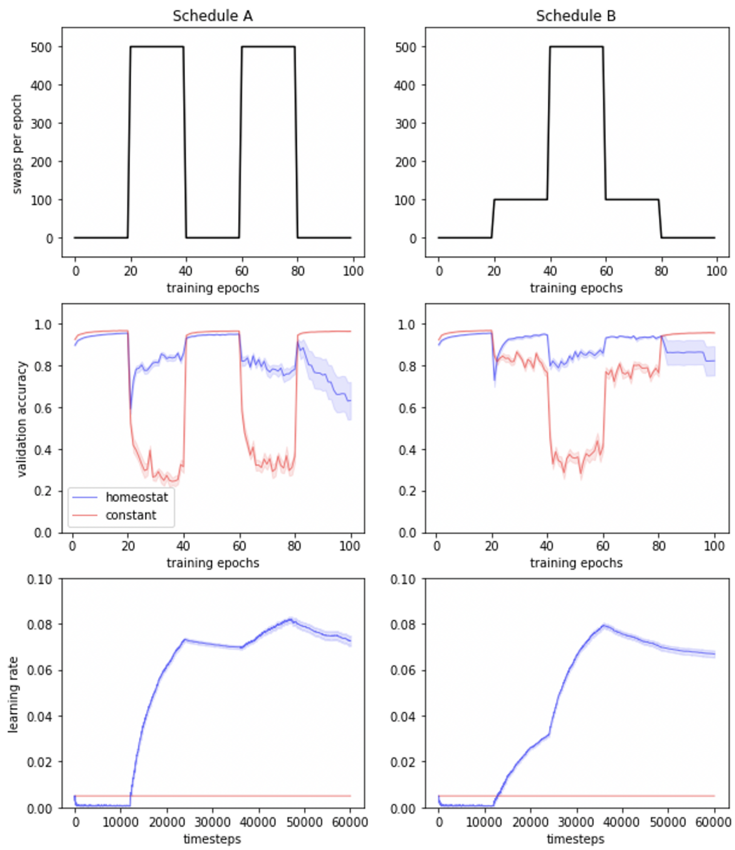

Figure 2: Accuracy and learning rate under “seasonality” of concept shift. Top row: Schedule A cycles between extreme rates of concept shift, while schedule B is more gradual. Middle row: The accuracy of the constant-LR classifier, in red, severely declines during stormy periods but returns to normal during calm periods. The homeostat, in blue, maintains good average performance across shifts in the rate of concept shift. Bottom row: The sequence of learning rates reveals that the homeostatic learner ratchets up its learning rate during stormy periods but is less inclined to reduce it during calm periods.

is never asked to learn from never-before-seen image patterns. In the real world, concept shift often co-occurs with some level of covariate shift. Not only do relationships change over time, but the predictors change as well.

Although our method can dynamically adjust the learning rate, we did not benchmark it against LR optimizers such as ADAM (Kingma and Ba 2014) and other momentum-based methods. In the non-stationary setting the loss surface shifts over time and it is therefore inappropriate to accumulate previous gradients from an outdated loss surface.

Finally, we are aware of the resemblance between reinforcement learning and our task of homeostatic self-regulation, though we argue that they should not be identified as the same. The objective here is not to maximize some arbitrary “reward” by massed trial-and-error. The object of the game is simply to keep playing the game. We specify a particular target to optimize: homeostatic well-being, crystallized as an internal parameter that controls the ongoing ability to make good decisions. For an excellent example of work in reinforcement learning that takes homeostatic logic into account see (Keramati and Gutkin 2014).

One way of explaining our homeostatic design is to say that it exposes an artificial neural network’s thinking machinery to the consequences of its own "thoughts". A vulnerable learner with the meta-task of self-preservation is incentivized to better align with reality and to adapt to external change. The superior adaptability of the vulnerable learner illustrates the benefits of putting one’s own "skin in the game".

References

Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289. 2015.

Di Paolo EA. Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. 2000.

He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision 2015 (pp. 1026-1034).

Iizuka H, Di Paolo EA. Extended homeostatic adaptation: Improving the link between internal and behavioural stability. In: International Conference on Simulation of Adaptive Behavior 2008 Jul 7 (pp. 1-11).

Keramati M, Gutkin B. Homeostatic reinforcement learning for integrating reward collection and physiological stability. eLife. 2014;3:e04811.

Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.

LeCun Y, Cortes C, Burges CJC. The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/, 1994.

Man K, Damasio A. Homeostasis and soft robotics in the design of feeling machines. Nature Machine Intelligence. 2019;1(10):446-52.

Marder E, Goaillard JM. Variability, compensation and homeostasis in neuron and network function. Nature Reviews Neuroscience. 2006;7(7):563-74.

Moreno-Torres JG, Raeder T, Alaiz-Rodríguez R, Chawla NV, Herrera F. A unifying view on dataset shift in classification. Pattern Recognition. 2012;45(1):521-30.

Williams H, Noble J. Homeostatic plasticity improves signal propagation in continuous-time recurrent neural networks. Biosystems. 2007;87(2-3):252-9.

Xiao H, Rasul K, Vollgraf R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747. 2017.

A Supplementary Material

```
// Algorithm 1. Homeostatic self-regulation of learning rate.

// Helper function to estimate the accuracy of a simulated version of oneself on recent data, at a given learning rate
Define simulate_self(mlp, memories, lr):
    Perform stochastic gradient descent, at the given lr, over a single pass of the memories dataset
    Return overall accuracy of predictions in re-classifying the memories

// Begin algorithm
Input: data, a set of images and associated labels from {0..9}
Input: f, a frequency of lr regulation every f steps, e.g. 100
Input: initial_lr, an initial learning rate, e.g. 0.005

lr = initial_lr
lr_stepsize = lr/10
memories = []    // to store up to f preceding datapoints

Initialize a multi-layer perceptron, mlp.
timestep=0
While 1:
    For each of (image, label) in data:
        Append (image, label) to memories
        if memories contains more than f items:
            remove the oldest item
        Perform forward pass on mlp and output a label prediction y^
        if timestep is a multiple of f:    // perform LR regulation
            // What is the predicted effect of this object on lr?
            if the label y^ is greater than or equal to 5:
                simulated_lr = lr + lr_stepsize    // excitatory
            else:
                simulated_lr = lr - lr_stepsize    // inhibitory
            // What would the accuracy be if learner altered its lr?
            ingest_accuracy = simulate_self(mlp, memories, simulated_lr)
            // What if learner kept its current lr?
            reject_accuracy = simulate_self(mlp, memories, lr)
            if ingest_accuracy > reject_accuracy:
                lr = simulated_lr
        Perform backward pass on the mlp at the selected lr
        timestep += 1
```

Figure S1: Algorithm: Pseudocode for homeostatic self-regulation of learning rate.

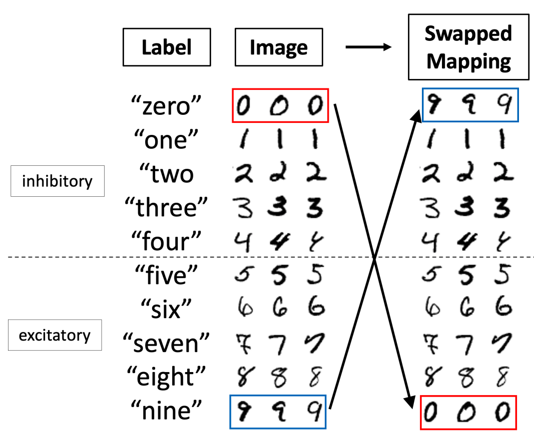

Figure S2: Concept shift is implemented by swapping the mapping between label and image between two randomly selected classes. Illustrated here on the MNIST dataset, the mappings for “zero” and “nine” are swapped. This swap will also invert the homeostatic effects expected for each number – a potentially hazardous situation for a vulnerable classifier.

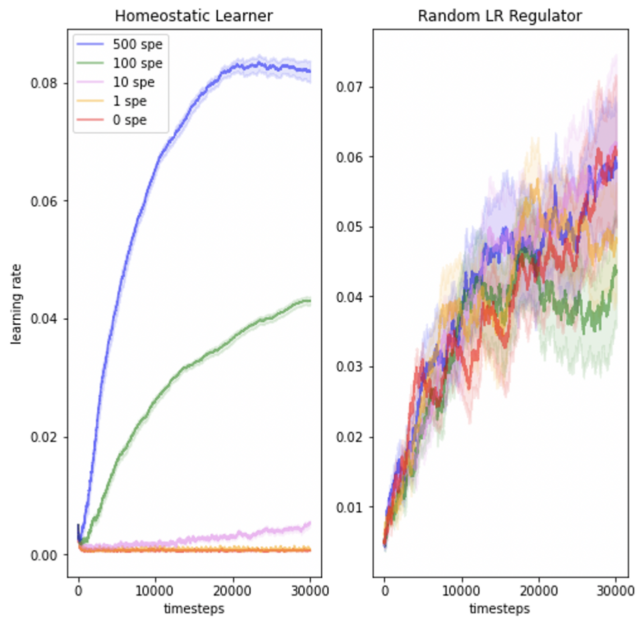

Figure S3: Learning rate sequencies of the two LR-regulating classifiers. The homeostatic learner seeks an LR appropriate to each level of concept shift, while the random regulator drifted upwards. At 500 swaps per epoch (left, blue), the homeostat arrests its own LR growth and asymptotes. Data shown from MNIST only.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: We claim an exploration of homeostasis in artificial neural networks and deliver on such claims, showing benefits under certain situations.

   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: [Yes]

   Justification: We discuss limits including dataset limitations, re-use of data for approximating concept shift, and intentional non-benchmarking against momentum-based optimizers.

   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: [NA]

   Justification: We do not contribute a theoretical proof.

   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: [Yes]

   Justification: Code is available on Github to reproduce the results. Pseudocode is included in the appendix.

   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: [Yes]

   Justification: See above.

   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: [Yes]

   Justification: Experimental setting is explained to the extent possible in the page limits. Further details are in public code repository.

   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: [Yes]

   Justification: We show mean +/- SEM over 20 replicates.

   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: [No]

   Justification: The compute resources used were not remarkable; experiments were performed on a single entry-level consumer laptop.

   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: [Yes]

   Justification: Reviewed and confirmed.

   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: [NA]

    Justification: The author cannot conceive of any direct, realistic, and significant societal impacts of this foundational research.

    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: [NA]

    Justification: No such risks.

    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: [Yes]

    Justification: MNIST and FMNIST are cited.

    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: [NA]

Justification: We do not release new assets/datasets.

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: [NA]

Justification: No crowdsourcing.

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: [NA]

Justification: No human subjects.

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.