GL Equivariant Metanetworks for Learning on Low Rank Weight Spaces

(/pdf?id=nsfg6oc24a)

10 May 2025 (modified: 18 Sep 2025) Submitted to NeurIPS 2025 Conference, Senior Area Chairs, Area Chairs, Reviewers, Authors Revisions (/revisions?id=nsfg6oc24a) BibTeX

Keywords: Equivariance, symmetry, geometric deep learning, weight-space architectures, finetuning
Abstract:
Low-rank adaptations (LoRAs) have revolutionized the finetuning of large foundation models, enabling efficient adaptation even with limited computational resources. The resulting proliferation of LoRAs together with the recent advances of weight-space learning present exciting opportunities for applying machine learning techniques that take these low-rank weights themselves as inputs. In this paper, we investigate the potential of {\em Learning on LoRAs} (LoL), a setup where machine learning models learn and make predictions on datasets of LoRA weights. Motivated by previous weight-space learning works, we first identify the inherent parameter symmetries of our data -- low-rank decompositions of weights -- which differ significantly from the parameter symmetries of standard neural networks. To efficiently process LoRA weights, we develop several symmetry-aware invariant or equivariant LoL models. %, using tools such as canonicalization, invariant featurization, and equivariant layers. In diverse experiments, we show that our LoL architectures can process LoRA weights to predict CLIP scores, finetuning data attributes, finetuning data membership, and accuracy on downstream tasks. We also show that LoL models trained on LoRAs of one pretrained model can effectively generalize to LoRAs trained on other models from the same model family. As an example of the utility of LoL, our LoL models can accurately estimate CLIP scores of diffusion models and ARC-C test accuracy of LLMs over 50,000 times faster than standard evaluation. As part of this work, we finetuned and will release datasets of more than ten thousand text-to-image diffusion-model and language-model LoRAs.

Checklist Confirmation: I confirm that I have included a paper checklist in the paper PDF.
Reviewer Nomination:
Responsible Reviewing: We acknowledge the responsible reviewing obligations as authors.
Primary Area: Deep learning (e.g., architectures, generative models, optimization for deep networks, foundation models, LLMs)
LLM Usage: Editing (e.g., grammar, spelling, word choice), Implementing standard methods
Declaration: I confirm that the above information is accurate.
Submission Number: 18609

Filter by reply type... | Filter by author... | Search keywords... | Sort: Newest First
Everyone | Program Chairs | Submission18609... | Submission18609... | Submission18609... | Submission18609... | Submission18609... | Submission18609... | Submission18609...
21 / 21 replies shown

Add: **Withdrawal** **Opt in or Veto Public Release**

Paper Decision

Decision by Program Chairs 17 Sep 2025 at 08:49 (modified: 18 Sep 2025 at 10:13) Program Chairs, Authors Revisions (/revisions?id=TzTmMcEpCz)

Decision: Reject
Comment:

Summary of scientific claims and findings

The paper introduces a symmetry-aware neural network architecture designed to predict attributes of fine-tuned models (e.g., validation loss, task attributes) directly from low-rank adapter (LoRA) weights. The method leverages GL(r)-equivariance to respect natural symmetries in LoRA weight representations. Extensive experiments on vision and language LoRA datasets demonstrate strong predictive performance, efficiency improvements, and generalization across ranks and adapter variants. Theoretical results demonstrate expressivity and universality.

Strengths

- **Novelty**: First comprehensive study on weight-space learning from LoRA adapters.
- **Theory**: Rigorous treatment of GL-equivariance, extended to mixed-rank settings with formal proofs.
- **Empirics**: Extensive experiments across multiple domains.

Weaknesses and missing elements

- **Adapter scope**: While GL-Net generalizes across LoRA and DoRA, true cross-adapter generalization (e.g., KronA, OFT) remains unresolved.
- **Dimensionality reduction**: Reliance on downsampling raises concerns for genuinely high-rank adapters.
- **Nonlinearity choice**: Equivariant nonlinearity may reduce effective rank; empirical justification is given, but concerns remain about its impact on depth and expressivity.
- **Baseline breadth**: Although improved, comparisons to some weight-space prediction methods remain limited.
- **Revision needed**: Clarifications on layer sufficiency and expanded inclusion of new results are necessary for the final version.

Summary of discussion and rebuttal

- `h9nY` : Initially raised concerns about mixed-rank LoRAs, baselines, and lack of closed-loop evaluation. After rebuttal and additional theory, acknowledged concerns were addressed; now borderline accept.
- `6HVs` : Positive from the start; initial concerns (scope, baselines, cost) largely resolved. Upgraded to accept.
- `bUbu` : Maintained reservations about cross-adapter generalization, nonlinearity, and high-rank downsampling. Appreciated new experiments (especially on open-source LoRAs) but concluded that major revision is still needed. Final stance: borderline reject.

Weighing these points

Two reviewers leaned clearly positive after rebuttal (one upgraded, one maintained positive). One reviewer remained unconvinced, citing fundamental limitations in adapter scope and rank assumptions. While not all limitations are resolved, they are acknowledged and contextualized.

Given the competitive field of this year's submissions, these reservations ultimately mean that the paper cannot be accepted, despite the overall positive average and several points in favor of acceptance.

Author Final Remarks by Authors

Author Final Remarks by 14 Aug 2025 at 11:15 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Author Final Remarks:
We would like to thank the reviewers and the AC for an engaging discussion and their insightful suggestions which helped us further demonstrate the utility of our methods. A few key strengths of our work were mentioned in the reviewing process:

1. Novel, highly performant architectures for predicting attributes of model finetunes. (Reviewers **h9nY**, **6HVs**)
2. Extensive and compelling experiments demonstrating the utility of LoL models. (Reviewers **h9nY**, **6HVs**)
3. Strong theoretical results exploring GL-expressivity and GL-equivariance. (Reviewers **h9nY**, **6HVs**, **bUbu**)

During the rebuttal process we added several experimental and theoretical results that showed further benefits of our methods:

1. We showed our methods significantly outperform an additional baseline suggested by the reviewers and achieve zero-shot generalization in several new settings.
2. We demonstrated GL-Net can be trained **entirely on publicly available finetunes** and achieve **excellent test accuracy on industry-relevant tasks**.
3. We proved that our theoretical results extend naturally to mixed-rank input settings.

Overall, we believe we have successfully addressed the reviewers' concerns. Reviewers **h9nY** and **6HVs** stated that their concerns had been successfully addressed. Reviewer **bUbu** expressed excitement for the new findings in the rebuttal phase and brought up final questions which we have answered thoroughly through extensive experimentation.

We believe that our paper provides a significant and unique contribution to the fields of weight space learning and equivariant learning, while improving our understanding of foundation model finetuning.

Official Review of Submission18609 by Reviewer h9nY

Official Review by Reviewer h9nY 01 Jul 2025 at 23:58 (modified: 18 Sep 2025 at 12:43)
Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer h9nY Revisions (/revisions?id=yVCPo5JQT6)

Summary:
The paper treats low-rank adaptation (LoRA) weights as data. When the two LoRA factors are post-multiplied by a shared invertible matrix, the resulting full weight update remains unchanged—imposing a natural symmetry that any weight-space predictor should respect. To that end, the authors propose GL-Net, which applies a symmetry-preserving projection to each LoRA layer, followed by an invariant prediction head. Evaluated on over ten thousand LoRA weight sets from both Stable Diffusion and large language models, the method accurately predicts text–image alignment scores, downstream language-model accuracy, and CelebA attributes—while running tens of thousands of times faster than brute-force evaluation.

Strengths And Weaknesses:

Strengths

- A rigorous treatment of the GL(r) symmetry, supported by formal proofs.
- Large-scale experiments on both vision and language LoRA models, with thorough evaluation and error bars.
- Strong empirical performance despite a small parameter budget, along with public release of datasets and substantial evaluation speed-ups.

Weaknesses

- *Global-rank assumption* All theory, code, and experiments require a single shared LoRA rank. In real practice many pipelines mix ranks per layer to balance quality and memory. Because GL-Net's equivariance guarantee depends on that common rank axis, the method may break or lose its theoretical property on mixed-rank checkpoints.
- *Missing strong weight-space baselines* Prior work shows that empirical NTK / NNGP kernels and weight-graph GNNs can predict test accuracy within a few points without data (Fort et al., 2020; Elsken et al., 2020). Comparing only to weak flattened-MLP baselines leaves unclear whether GL-Net's edge comes from its symmetry or simply from facing easier competition.
  - Fort S. et al. "Deep Ensembles and Neural Tangent Kernels on Neural Architecture Search Benchmarks." ICLR W, 2020.
  - Elsken T. et al. "MetaNAS: Weight-sharing meets neural architecture search via graph-level meta-learning." ICLR, 2020.
- *Meta-only evaluation* The experiments stop at off-line prediction. No closed-loop study is shown (for example, selecting the top-k LoRAs by predicted score and verifying actual gains), which would better demonstrate practical value.

GL-Net represents a promising and efficient step toward learning directly from LoRA weights. However, as previously outlined, it has certain limitations. I will adjust my rating based on the authors' response.

Quality: 3: good
Clarity: 3: good
Significance: 2: fair
Originality: 3: good
Questions:

- Can the method be extended to mixed-rank LoRAs (e.g.\ padding, per-layer invariant collapse, or product-group equivariance)? Please implement a small mixed-rank experiment or sketch the required changes.
- Please add at least one strong weight-space baseline—empirical NTK, NNGP, or a weight-graph GNN—and report the same metrics. This will clarify whether GL-Net's edge comes from its symmetry rather than weaker competition.
- Demonstrate a practical loop—select the top-k LoRAs by GL-Net prediction and verify the actual improvement, or run a privacy audit—so readers can see the surrogate in action.

Limitations:
yes

Rating: 4: Borderline accept: Technically solid paper where reasons to accept outweigh reasons to reject, e.g., limited evaluation. Please use sparingly.
Confidence: 4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.
Ethical Concerns: NO or VERY MINOR ethics concerns only
Paper Formatting Concerns:
no

Code Of Conduct Acknowledgement: Yes
Responsible Reviewing Acknowledgement: Yes
Final Justification:
I thank the authors for their response, which addressed most of my concerns. In light of the overall quality of the paper, I would like to keep my original rating unchanged.

Rebuttal by Authors

Rebuttal by Authors 30 Jul 2025 at 12:17 (modified: 31 Jul 2025 at 16:19) Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Revisions (/revisions?id=jF1ZQLHstz)

Rebuttal:

We thank the reviewer for appreciating our theoretical work and extensive experiments, which demonstrate the utility of GL symmetry-aware architectures. We have run new experiments following your suggestions and found new strengths of our architectures. We think we have improved our work through your advice and address your questions and concerns one by one, with added experimental results.

> Can the method be extended to mixed-rank LoRAs (e.g.\ padding, per-layer invariant collapse, or product-group equivariance)? Please implement a small mixed-rank experiment or sketch the required changes.

Thank you for this interesting question. To ensure that GL-equivariance is still practical on a dataset of LoRA models with varying intra-model rank, we constructed a new dataset, Imagenette-LoRA-σ. This dataset consists of 2046 LoRA models where the rank of each LoRA adapter module within each model is a random power of 2 between 2^0 - 2^5. Importantly, the chosen adapter ranks differ both within the same model and between different models, which we believe captures all possible variance in rank in a LoL dataset. We train and evaluate GLNet on Imagenette-LoRA-σ and report an accuracy of 83.2 ± .3%, nearly as well as the constant rank setting; even in this **highly variant rank setting**, GL-equivariant models can effectively predict attributes of LoRA models. In fact, we find that a GLNet trained **only** on rank-4 LoRA models can **generalize zero-shot** to varying rank models and achieve an accuracy of ~82%. Thus, GLNet remains a strong architecture when the global rank assumption does not hold.

Additionally, on the theoretical side, when the input space of GLNet has intra-model rank variance, our expressivity argument still extends. All that we require is that each adapter be full rank in the sense that it achieves its theoretically predefined rank. We will include a more formal proof in our appendix.

> Please add at least one strong weight-space baseline—empirical NTK, NNGP, or a weight-graph GNN—and report the same metrics. This will clarify whether GL-Net's edge comes from its symmetry rather than weaker competition.

We thank the reviewer for their important suggestion. Inspired by the reviewer's comments, we have added experiments with a SOTA weight space model, Universal Neural Functionals [Zhou et al, NeurIPS 2024]. We find that while UNF models outperform baselines like MLP and Transformer due to their permutation symmetry-aware architectures, they perform **significantly worse** than our expressive GL-Invariant and O-Invariant models because their design does not take into account the most relevant symmetries. We find that on Imagenette-LoRA, GL-Net has an error rate almost **three times lower than UNF** while running significantly faster. This gives further empirical evidence to our discussion in Section 2 about how permutation symmetries are not sufficient for LoRAs, and GL-invariance or O-invariance are necessary.

Trained on:	**LoRA-SD1.4**	**LoRA-SD1.4**	**LoRA-SD1.5**	**LoRA-SD1.5**
Accuracy on:	LoRA-SD1.4	LoRA-SD1.5	LoRA-SD1.4	LoRA-SD1.5
UNF	70.3 ± 1.7	69.1 ± 1.6	68.8 ± 0.9	68.3 ± 0.9
GL-Net	**87.8 ± 0.4**	**87.6 ± 0.5**	**87.5 ± 0.4**	**88.5 ± 0.2**

> Demonstrate a practical loop—select the top-k LoRAs by GL-Net prediction and verify the actual improvement, or run a privacy audit—so readers can see the surrogate in action.

We agree with the reviewer that a closed-loop study is an effective way to show the utility of LoL models. Our paper already shows results in this direction - we direct the reviewer's attention to Section 4.2.1, where we select the worst and best LoRAs as predicted by GLNet. We find that the worst expected model consistently outputs nonsense when prompted, whereas the best expected model generates accurate images based on each prompt. We commit to adding more examples from the top-5 and bottom-5 predicted clip score models, but cannot do so now because we are unable to upload photos.

We hope that our successful experiments above have adequately addressed the questions and concerns you brought up.

Replying to Rebuttal by Authors

Mandatory Acknowledgement by Reviewer h9nY

Mandatory Acknowledgement by Reviewer h9nY 05 Aug 2025 at 07:40 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Mandatory Acknowledgement: I have read the author rebuttal and considered all raised points., I have engaged in discussions and responded to authors., I have filled in the "Final Justification" text box and updated "Rating" accordingly (before Aug 13) that will become visible to authors once decisions are released., I understand that Area Chairs will be able to flag up Insufficient Reviews during the Reviewer-AC Discussions and shortly after to catch any irresponsible, insufficient or problematic behavior. Area Chairs will be also able to flag up during Metareview grossly irresponsible reviewers (including but not limited to possibly LLM-generated reviews)., I understand my Review and my conduct are subject to Responsible Reviewing initiative, including the desk rejection of my co-authored papers for grossly irresponsible behaviors. https://blog.neurips.cc/2025/05/02/responsible-reviewing-initiative-for-neurips-2025/ (https://blog.neurips.cc/2025/05/02/responsible-reviewing-initiative-for-neurips-2025/)

Replying to Rebuttal by Authors

Official Comment by Reviewer h9nY

Official Comment by Reviewer h9nY 05 Aug 2025 at 07:43 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:

I thank the authors for their response, which addressed some of my concerns. However, as I still have reservations about the theoretical analysis of mixed-rank LoRAs, I am keeping my original rating unchanged.

Mixed Rank LoRA Theory Part I

Official Comment by Authors) 07 Aug 2025 at 00:35 (modified: 07 Aug 2025 at 00:48) Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Revisions (/revisions?id=REkkZB4Hqa)

Comment:

Thank you for your response and for clarifying which of your concerns remain. Below, we give a formal definition of the mixed-rank setting and then state and prove two theorems showing that GL-Net is equivariant and expressive in this setting. We believe these additions strengthen the paper, and would greatly appreciate a reconsideration of the score.

Let $L(r) = \{(A, B) \mid A \in \mathbb{R}^{m\times r}, \in \mathbb{R}^{n\times r}, \mathrm{rank}(A) = \mathrm{rank}(B) = r\}$ be the set of all possible LoRA adapters of rank **r**.

1. In the fixed rank setting, the input space to an LoL model is given by $X_{fixed}(r) = L(r) \times L(r) \times L(r) \ldots \times L(r)$ for constant r.
2. In the varying intra-model rank setting with n layers, the input space of an LoL model is given by $X_{intra}(r_1, r_2, \ldots r_n) = L(r_1) \times \cdots L(r_n)$.

3. Finally, when the rank of each LoRA adapter varies between different models the input space is given by $X_{inter}(S) = \bigcup_{(r_1,r_2,\dots r_n)\in S^n} X_{intra}(r_1, r_2 \dots r_n)$, where S is a finite set of possible ranks.

Theorem 4.1: The linear layer, $(A_1, B_1), \dots, (A_n, B_n) \to (\Psi_1 A_1, \Phi_1 B_1), \dots, (\Psi_n A_n, \Phi_n B_n)$, described in 3.3.2 are $GL(r_1) \times \dots GL(r_n)$-equivariant on X_{intra}.

Proof: As proven in Theorem A.1, for a single LoRA module, $Linear_i = (A_i, B_i) \to (\Psi A_i, \Phi B_i)$ acts equivariantly with respect to $GL(r_i)$. Thus, the direct product $Linear_1 \times Linear_2 \cdots \times Linear_n$ acts equivariantly with respect to $GL(r_1) \times \cdots \times GL(r_n)$.

Lemma 4.2: When acting on $X_{intra}(r_1, \dots, r_n)$, GL-Net is invariant to $GL(r_1) \times \cdots \times \dots GL(r_n)$.

Proof: Let $((A_1, B_1), \dots, (A_n, B_n)) \in X_{intra}(r_1, \dots, r_n)$. The linear layers of GL-Net are equivariant, so we only need to show that the invariant head, $MLP(UV^T)$ is invariant to $GL(r_1) \times GL(r_2) \times \dots GL(r_n)$. And indeed we have
$MLP(R * UV^T) = MLP(U_1 R_1 R_1^{-1} V_1^T, U_2 R_2 R_2^{-1} V_2^T, \dots U_n R_n R_n^{-1} V_n^T) = MLP(U_1 V_1^T, U_2 V_2^T, \dots U_n V_n^T) = MLP(UV^T)$.

Theorem 4.3 (GL Invariance for Mixed Rank Setting): For any $((A_1, B_1), \dots, (A_n, B_n)) \in X_{inter}(S)$, with (A_i, B_i) of rank r_i, GL-Net is invariant to $GL(r_1) \times \cdots \times GL(r_n)$ when applied to $(A_1, B_1) \dots (A_n, B_n)$.

Proof: By construction, $(A_1, B_1) \dots (A_n, B_n) \in X_{intra}(r_1, \dots r_n)$. By lemma 4.2, GL-Net is invariant to $GL(r_1) \times \dots GL(r_n)$ acting on $X_{intra}(r_1, \dots r_n)$

Replying to Mixed Rank LoRA Theory Part I

Mixed Rank LoRA Theory Part II

Official Comment by Authors 07 Aug 2025 at 00:47 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:
Lemma 4.4: GL-Net is full-rank universal on $X_{intra}(r_1, \dots r_n)$.

Proof: See Theorem C.2. Without loss of generality, this proof applies to this setting because the individual ranks of each LoRA adapter are never assumed to be the same.

Definition: (Full rank GL-universality in Mixed Rank Setting). An LoL architecture is considered to be full rank GL-universal in the mixed rank setting if for each function $f : X_{inter}(S) \to \mathbb{R}$, where $f_{|X_{intra}(r_1 \dots r_n)}$ is invariant to $GL(r_1 \dots r_n)$, and for corresponding compact sets $K_{r_1 \dots r_n} \subset X_{intra(r_1 \dots r_n)}$, there is a model f^{LoL} of said architecture that approximates f on each $K_{r_1 \dots r_n}$ up to ϵ.

Theorem 4.5 (Universality in Mixed Rank Setting): For finite S, MLP(UV^T) and GL-Net are full rank universal in the mixed rank setting.

Proof: By Lemma 4.4, for each $(r_1 \dots r_n)$, there exists some $f^{LoL}_{r_1 \dots r_n}$ that approximates $f_{|X_{intra}(r_1 \dots r_n)}$ on $K_{r_1 \dots r_n}$. So, we need to show that there is a *single* LoL model, f^{LoL}_S which exactly equals each of these $|S|^n$ functions when restricted to their corresponding compact domains. This can be done via a 4 layer MLP(UV^T) model as described below:

1. First note that each input set is compact and, by assumption, full rank. Because two-layer MLPs are universal approximators, the first two layers of the MLP(UV^T) model can predict the rank of each LoRA adapter, mapping $(A_1, B_1) \dots (A_n, B_n)$ to $((A_1, B_1) \dots (A_n, B_n), (r_1, \dots r_n))$.
2. Next, conditioned on the rank configuration predicted in the first two layers, the last two layers of the MLP(UV^T) can properly approximate $f^{LoL}_{r_1 \dots r_n}$.

With this construction, f^{LoL}_S, when restricted to $K_{r_1 \dots r_n}$, exactly equals $f^{LoL}_{r_1 \dots r_n}$. Thus, MLP(UV^T) and therefore GL-Net are full rank universal in the mixed rank setting.

Replying to Mixed Rank LoRA Theory Part II

Official Comment by Reviewer h9nY

Official Comment by Reviewer h9nY 09 Aug 2025 at 05:00 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:
I appreciate the detailed response regarding the theoretical analysis of mixed-rank LoRAs, which has addressed my concern.

Official Review of Submission18609 by Reviewer 6HVs

Official Review by Reviewer 6HVs 20 Jun 2025 at 06:34 (modified: 18 Sep 2025 at 12:43)
Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer 6HVs Revisions (/revisions?id=fh4uCHoVRd)

Summary:
This paper proposes a method to predict key properties of LoRA adapters such as performance, accuracy, and dataset attributes directly from pretrained LoRA weight factors. To achieve this, the authors leverage symmetry by enforcing invariance and equivariance to the intrinsic transformations of LoRA weights. Specifically, they first generate a diverse set of LoRA adapters by fine-tuning base models across multiple tasks. These adapters serve as training data for a property-prediction model called **LoL**. Extensive experimental evaluations demonstrate that the proposed method consistently achieves promising results.

Strengths And Weaknesses:
Strengths

- **Important problem** – Targets performance prediction for LoRA-adapted weights, a key challenge for weight-space learning and model-selection workflows.
- **Solid theoretical foundation** – Presents a clear symmetry/equivariance analysis that motivates the architecture of the predictor.
- **Compelling empirical results** – Experiments on diverse tasks show consistent accuracy gains over naïve or heuristic baselines, demonstrating practical value.

Weaknesses

- **Costly data generation** – Building the required training set of fine-tuned LoRA adapters may be computationally and financially intensive.
- **Limited scope** – The approach is tailored to LoRA; its applicability to other adapter or full-fine-tuning schemes remains unverified.
- **Baseline coverage** – The study omits direct comparisons with several existing weight-level performance predictors, making it difficult to gauge the magnitude of improvement.

Quality: 3: good
Clarity: 3: good
Significance: 2: fair
Originality: 2: fair
Questions:
1. **Scope beyond LoRA** Your approach is evaluated only on LoRA-adapted weights. Is the predictor intrinsically tied to LoRA, or could it generalise to other adaptation or full-fine-tuning schemes? If so, why were no non-LoRA settings included in the experiments?

2. **Choice of baselines** Several prior works predict model properties directly from weight tensors [1]. Could you explain why none of these methods were used as baselines, and how their absence might affect the interpretation of your performance gains?
3. **Data-generation cost and scalability** Training the predictor requires a large corpus of fine-tuned LoRA adapters. Can you provide a breakdown of the computational cost (GPU hours, energy, etc.) and discuss how the method scales to larger base models or a wider range of tasks?
4. **Clarification of the table next to Figure 5** The table adjacent to Figure 5 is difficult to parse: the column headers, abbreviations, and the relationship between the table entries and the plotted curves are unclear. Could you add a legend or brief description clarifying what each column represents (e.g. task domain, adapter rank, metric), and how the numbers relate to the qualitative trends shown in the figure?

[1] Learning on Model Weights using Tree Experts

Limitations:
see weaknesses

Rating: 5: Accept: Technically solid paper, with high impact on at least one sub-area of AI or moderate-to-high impact on more than one area of AI, with good-to-excellent evaluation, resources, reproducibility, and no unaddressed ethical considerations.
Confidence: 5: You are absolutely certain about your assessment. You are very familiar with the related work and checked the math/other details carefully.
Ethical Concerns: NO or VERY MINOR ethics concerns only
Paper Formatting Concerns:
No formatting concerns

Code Of Conduct Acknowledgement: Yes
Responsible Reviewing Acknowledgement: Yes
Final Justification:
I appreciate the authors' detailed response, which has successfully addressed my initial concerns.

Rebuttal by Authors

Rebuttal by Authors 30 Jul 2025 at 12:26 (modified: 31 Jul 2025 at 16:19) Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors
Revisions (/revisions?id=I1vwtMg5MQ)

Rebuttal:
We thank the reviewer for appreciating our theoretical work, extensive experiments, and the interesting problem we seek to solve. Guided by the reviewer's advice we have conducted several new experiments which strengthen our results. We acknowledge your questions and concerns and have addressed them below in order using our new observations.

> Scope beyond LoRA Your approach is evaluated only on LoRA-adapted weights. Is the predictor intrinsically tied to LoRA, or could it generalise to other adaptation or full-fine-tuning schemes? If so, why were no non-LoRA settings included in the experiments?

Our LoL models are not intrinsically tied to LoRA adapters, but in our paper, we stuck to LoRA-based datasets for ease of data collection and since LoRAs are a very popular fine-tuning method. In our appendix, we include a theoretical study of how our LoL models could be extended to other adapters.

In order to ensure that this is empirically valid, we created a new dataset, Imagenette-Dora, in which 2046 SD1.4 finetunes were created with DoRA [Liu et al, ICLR 2024]. LoL performs **extremely well** in two settings here:

1. GLNet achieves an accuracy of 88% when trained and tested on predicting finetuning image attributes of each DoRA, the same as when trained and tested on LoRA.
2. GLNet generalizes **zero-shot** to DoRA prediction when trained on LoRA prediction. Specifically, a GLNet trained on Imagenette-LoRA gets 86.5 ± 0.4% accuracy on Imagenette DoRA despite no DoRA finetunes appearing in its training set.

This is strong evidence that our approach is broadly applicable to many adapters and fine-tuning methods. Furthermore, as proof that our method is applicable to full finetuning, we generated a new dataset of rank 320 Imagenette LoRAs and found that our LoL models effectively generalize from rank-4 training data to rank-320 training data. So, in principle, one could perform SVD of the full-finetune and then pass in the corresponding low rank approximation into an LoL model.

> Choice of baselines Several prior works predict model properties directly from weight tensors [1]. Could you explain why none of these methods were used as baselines, and how their absence might affect the interpretation of your performance gains?

We agree with the reviewer that more baselines would provide a better perspective for the importance of our results. We now add some preliminary results for using Universal Neural Functionals [Zhou et al, NeurIPS 2024] and find that our GL-Invariant models significantly outperform them.

Trained on:	**LoRA-SD1.4**	**LoRA-SD1.4**	**LoRA-SD1.5**	**LoRA-SD1.5**
Accuracy on:	LoRA-SD1.4	LoRA-SD1.5	LoRA-SD1.4	LoRA-SD1.5
UNF	70.3 ± 1.7	69.1 ± 1.6	68.8 ± 0.9	68.3 ± 0.9
GL-Net	**87.8 ± 0.4**	**87.6 ± 0.5**	**87.5 ± 0.4**	**88.5 ± 0.2**

We treat the important work done in Horwitz et al. as concurrent work because it was put on arXiv after our work and cites us. Nevertheless, in our final version, we will include a direct comparison with the probing methods detailed in Horwitz et. al

> Data-generation cost and scalability Training the predictor requires a large corpus of fine-tuned LoRA adapters. Can you provide a breakdown of the computational cost (GPU hours, energy, etc.) and discuss how the method scales to larger base models or a wider range of tasks?

We thank the reviewer for this important question. The data generation process for our paper was not heavy at all and required ~150 3090 GPU hours and ~800 2080TI GPU hours. The total cost was about $100 and 250 kWh. We provided a table listing GPU hours used to generate each of our datasets in Appendix D.5

Additionally, we think our work is applicable to some settings in which training data already exists. For instance, many LoRAs are hosted on Hugging Face, civitai, and other websites, and future work could leverage these more for training. Moreover, some companies and organizations train many LoRA finetunes for their models over the course of research and development, and LoL models could train and provide inferences over those.

We believe our methods would scale well to larger base models for two reasons: LoRA datasets for larger models could be stored and generated very efficiently with libraries like PEFT, and our most advanced architecture, GLNet runs in time proportional to the square root of the number of model parameters. Furthermore, the extensive generalization capabilities of our GL Invariant LoL models – across LoRA rank, model checkpoints (See section 4.4), and adapter methods (See above) – greatly reduce the amount of training data required for LoL methods to be widely effective in practice. Due to their vast generalization abilities, we also expect that LoL models for one task or model architecture could be properly finetuned using only a few training examples for use on another task or model architecture.

> Clarification of the table next to Figure 5 The table adjacent to Figure 5 is difficult to parse: the column headers, abbreviations, and the relationship between the table entries and the plotted curves are unclear. Could you add a legend or brief description clarifying what each column represents (e.g. task domain, adapter rank, metric), and how the numbers relate to the qualitative trends shown in the figure?

Thanks for this important comment. We will make an effort to make this table more readable. In particular, we will add the following description of the table to the caption: “On the right, we have provided a table with test losses across various input LoRA-ranks for three classes of LoL models. Each model is trained on CelebA LoRAs of rank 4 and then tested on ranks between 1 and 32.”

Replying to Rebuttal by Authors

Mandatory Acknowledgement by Reviewer 6HVs

Mandatory Acknowledgement by Reviewer 6HVs 04 Aug 2025 at 06:23 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Mandatory Acknowledgement: I have read the author rebuttal and considered all raised points., I have engaged in discussions and responded to authors., I have filled in the "Final Justification" text box and updated "Rating" accordingly (before Aug 13) that will become visible to authors once decisions are released., I understand that Area Chairs will be able to flag up Insufficient Reviews during the Reviewer-AC Discussions and shortly after to catch any irresponsible, insufficient or problematic behavior. Area Chairs will be also able to flag up during Metareview grossly irresponsible reviewers (including but not limited to possibly LLM-generated reviews)., I understand my Review and my conduct are subject to Responsible Reviewing initiative, including the desk rejection of my co-authored papers for grossly irresponsible behaviors. https://blog.neurips.cc/2025/05/02/responsible-reviewing-initiative-for-neurips-2025/ (https://blog.neurips.cc/2025/05/02/responsible-reviewing-initiative-for-neurips-2025/)

Replying to Mandatory Acknowledgement by Reviewer 6HVs

Official Comment by Reviewer 6HVs

Official Comment by Reviewer 6HVs 05 Aug 2025 at 06:41 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:
I appreciate the authors' detailed response, which has successfully addressed my initial concerns. I am raising my score and suggested the authors include the additional baselines in the main paper.

Official Review of Submission18609 by Reviewer bUbu

Official Review by Reviewer bUbu 18 Jun 2025 at 05:56 (modified: 18 Sep 2025 at 12:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer bUbu Revisions (/revisions?id=5wcijNHcuy)

Summary:
In proposed paper authors provide an approach to train a neural network predicting certain attributes of fine-tuning procedure, such as validation loss or data membership. Authors describe in details GL-equivariant layers and introduce architecture to process Low-rank adapters.

Strengths And Weaknesses:
Strengths:

1. The proposed method has been applied to various downstream tasks.
2. Authors formally describe all possible GL-equivariant linear maps.
3. Having pre-trained model, authors show great increase of the evaluation efficiency.

Weaknesses:

1. GL-equivariant non-linearity used in experiments dramatically decreases matrix rank. This activation zeros whole rows and columns of the matrix which is low-rank yet, making its rank even smaller than it was. Utilising this activation sequentially several times, matrix rank rapidly decreases to 1. It seems like it is the reason why authors arrived at the decision that one equivariant layer is the optimal number of layers in GL-net.
2. Typically, for LLM fine-tuning several LoRA approaches are tested. But one metanetwork is able to process only a particular LoRA type. It implies that we need a different metanetwork for various adapter types to evaluate its performance. It limits adapter types the one can use for parameter-efficient fine-tuning.
3. Not all parameter-efficient adapters can be treated as low-rank matrix. For example, Kronecker adapter providing that blocks are invertible or parameter-efficient representation of orthogonal matrices cannot be represented as low-rank).
4. Such an approach still suffers from computational burden as certain task requires its own low-rank adapters dataset.

Quality: 2: fair
Clarity: 3: good
Significance: 2: fair
Originality: 2: fair
Questions:

1. Did you try any other GL-equivariant non-linearities in your architectures which don’t affect matrix rank of Low-rank adapter so much?
2. Is there any time comparison of the full pipeline for training metanetwork (collection of LoRAs for training set, model training, etc.) and usual model evaluation? Will it be beneficial for large language models, requiring more LoRAs to train metanetwork?
3. Is it possible to train model using open-source adapters as pre-train dataset?
4. Do you lower the dimension of U_i and V_i to 32 because your LoRAs have ranks at most 32? Does such downsampling works for higher LoRA ranks?

Limitations:
Yes

Rating: 3: Borderline reject: Technically solid paper where reasons to reject, e.g., limited evaluation, outweigh reasons to accept, e.g., good evaluation. Please use sparingly.
Confidence: 4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.
Ethical Concerns: NO or VERY MINOR ethics concerns only
Paper Formatting Concerns:
I haven't found any formatting issues in the paper.

Code Of Conduct Acknowledgement: Yes
Responsible Reviewing Acknowledgement: Yes
Final Justification:

I will maintain my score as the proposed method requires more justification in the case of different adapters, performance estimation and higher adapter ranks. I think that major revision is required for this paper.

Rebuttal by Authors

Rebuttal by Authors 30 Jul 2025 at 12:33 (modified: 31 Jul 2025 at 16:19) Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Revisions (/revisions?id=2V3aCQijO7)

Rebuttal:

We thank the reviewer for appreciating our formalization of GL-Equivariant linear maps and the use of our models for improving the efficiency of various downstream tasks. We have performed new successful experiments to answer some of your questions and have improved our paper with the corresponding results. With these findings, we address your concerns below.

> Typically, for LLM fine-tuning several LoRA approaches are tested. But one metanetwork is able to process only a particular LoRA type. It implies that we need a different metanetwork for various adapter types to evaluate its performance. It limits adapter types the one can use for parameter-efficient fine-tuning.

(Same as our response to reviewer 6HVs) Recall that in our appendix we include a theoretical study of how our LoL models could be extended to other adapters. In order to ensure that this is empirically valid, we created a new dataset, Imagenette-Dora, in which 2046 SD1.4 finetunes were created with DoRA. LoL performs **extremely well** in two settings here:

1. GLNet achieves an accuracy of 88% when trained and tested on predicting finetuning image attributes of each DoRA
2. GLNet generalizes **zero-shot** to DoRA prediction when trained on LoRA prediction. Specifically, a GLNet trained on Imagenette-LoRA gets 86.5 ± 0.4% accuracy on Imagenette DoRA despite no DoRA finetunes appearing in its training set.

This is strong evidence that our approach is broadly applicable to many adapters and finetuning methods.

Furthermore, we note that the extensive rank and base-model generalization capabilities of our LoL models (See section 4.4) make them more versatile, and allow for using the same metanetwork for many different tasks.

> Did you try any other GL-equivariant non-linearities in your architectures which don't affect matrix rank of Low-rank adapter so much?

We note that zeroing out rows of the U and V LoRA matrices does not reduce their respective ranks because their bottleneck dimension is their number of columns. For example on Imagenette-LoRA, the hidden dimension of the U and V matrices are roughly 32 x 4. In this case, zeroing out half the rows will not shrink the column space. We show this numerically as well. Here are the singular values of a 32 x 4 matrix after zeroing out different percentages of its rows:

- Original singular values: [6.18911721 5.9938911 5.19238221 3.64397988]
- Singular values after zeroing 25% of rows: [5.77936389 5.14850004 4.19399806 2.88419029]
- **Singular values after zeroing 50% of rows (closest to practice)**: [4.79634456 4.03141692 2.76052111 1.83112429]
- Singular values after zeroing 75% of rows: [3.35309198 2.55937009 1.89591066 0.94635955]

We have also tried other gating mechanisms for our GL-equivariant nonlinearity (such as negating rows with negative sum) and find that those nonlinearities still frequently hurt performance despite maintaining singular value norms.

> Not all parameter-efficient adapters can be treated as low-rank matrix. For example, Kronecker adapter providing that blocks are invertible or parameter-efficient representation of orthogonal matrices cannot be represented as low-rank).

We have added an experiment where we test GL-Net on nearly full rank (rank-320) inputs. We find that GLNet performs well, attaining roughly the same accuracy as for low rank inputs. We consider this strong evidence that GLNet can be applied to full rank inputs or low rank approximations thereof. Moreover, one of our contributions in the paper is to list out the symmetries of various non-LoRA finetuning methods, including Kronecker adapters in section E of the appendix. We note that due to the structure of Kronecker adapters and their GL(1) symmetry group, they can still be efficiently processed by a GL-Net architecture by flattening the two product matrices and treating them each as rank-1 matrices.

> Such an approach still suffers from computational burden as certain task requires its own low-rank adapters dataset.

We believe that the computational burden for creating training data for LoL models is not prohibitively large for three reasons:

1. Our LoL models demonstrate wide ranging generalization capabilities across base-model, adapter rank, and adapter type. Thus, just one dataset of LoRAs would be sufficient for a variety of tasks, models, and adapter types.
2. For many models there are countless opensource low rank finetunes available on websites like Civitai and Hugging Face. These could be used as training data, since their configurations and purpose are generally public.
3. All of the datasets we collected required only about $100 worth of compute.

> Is there any time comparison of the full pipeline for training metanetwork (collection of LoRAs for training set, model training, etc.) and usual model evaluation? Will it be beneficial for large language models, requiring more LoRAs to train metanetwork?

For CelebA LoRAs, we find that evaluating the CLIP scores of 2046 stable diffusion finetunes takes 48 RTX 3090 GPU hours. On the other hand, collecting 2046 LoRA finetunes, training an LoL model on them, and then predicting their clip scores takes only 16 RTX 3090 GPU hours.

Thus, since predicting CLIP score with an LoL model is extremely efficient, if you were to use this trained LoL model to predict the CLIP score of about 700 new LoRAs, then you "break even" on the cost of training the LoL model. This can already be useful in certain applications, since there are many thousands of LoRAs available in open repositories like Huggingface, and companies train numerous finetunes (with several checkpoints in each training run) of their models. As future work tackles sample complexity and efficiency of LoL models, this trade-off can become even better for learning on LoRAs.

> Did you try any other GL-equivariant non-linearities in your architectures which don't affect matrix rank of Low-rank adapter so much?

See above

> Is it possible to train model using open-source adapters as pre-train dataset?

We believe this is an interesting idea that is best left to future work. We expect that this is possible due in large part to the extensive generalization capabilities of LoL models.

> Do you lower the dimension of U_i and V_i to 32 because your LoRAs have ranks at most 32? Does such downsampling works for higher LoRA ranks?

Yes, it does work for higher LoRA ranks! We can downsample below the rank of the LoRA model and maintain good performance. For example, in CLIP score calculation we downsample to a hidden dimension of just 1, and still get very good performance. We have added an experiment for LoRAs of rank 320, and find that GL-Net with a hidden equivariant dimension of 64 gets similar accuracy (86%) when predicting attributes of rank-320 LoRAs as rank-4 LoRAs (87%).

We are particularly excited by the findings of our new experiments above and thank the reviewer for pointing out these promising directions for further research. With these new results in mind, we hope the reviewer will consider adjusting their score.

Replying to Rebuttal by Authors

Discussion

Official Comment by Reviewer bUbu 05 Aug 2025 at 08:07 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:

Thank you for the detailed response regarding your method's nuances. After carefully reviewing the submission and additional results, my fundamental question about GL-Net's applicability to diverse fine-tuning adapters persists:

1. While the authors demonstrate GL-Net's ability to generalize from LoRA to DoRA in zero-shot prediction, this transition appears theoretically expected since both employ low-rank matrix representations ($UV^\top$). This suggests GL-Net may be limited to adapters expressible in this factorization form.
2. A more critical generalisation challenge arises with Kronecker Adapters (KronA). As shown in Appendix E, KronA requires vectorisation to process through GL-Net, altering the original U and V dimensions. This shape transformation requires additional modifications of the model architecture to be able to process adapters of dirrenent structure.
3. The feasibility of sourcing sufficient open-source adapters for training remains unverified. A controlled experiment using only publicly available adapters would strengthen claims about real-world applicability.

Replying to Rebuttal by Authors

Mandatory Acknowledgement by Reviewer bUbu

Mandatory Acknowledgement by Reviewer bUbu 05 Aug 2025 at 08:07 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Mandatory Acknowledgement: I have read the author rebuttal and considered all raised points., I have engaged in discussions and responded to authors., I have filled in the "Final Justification" text box and updated "Rating" accordingly (before Aug 13) that will become visible to authors once decisions are released., I understand that Area Chairs will be able to flag up Insufficient Reviews during the Reviewer-AC Discussions and shortly after to catch any irresponsible, insufficient or problematic behavior. Area Chairs will be also able to flag up during Metareview grossly irresponsible reviewers (including but not limited to possibly LLM-generated reviews)., I understand my Review and my conduct are subject to Responsible Reviewing initiative, including the desk rejection of my co-authored papers for grossly irresponsible behaviors. https://blog.neurips.cc/2025/05/02/responsible-reviewing-initiative-for-neurips-2025/ (https://blog.neurips.cc/2025/05/02/responsible-reviewing-initiative-for-neurips-2025/)

Replying to Discussion

Official Comment by Authors

Official Comment by Authors 07 Aug 2025 at 01:35 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:

Thank you for your response and the interesting points you have brought up. We address them below.

> The feasibility of sourcing sufficient open-source adapters for training remains unverified. A controlled experiment using only publicly available adapters would strengthen claims about real-world applicability.

In response to your concerns we have conducted a novel experiment on exclusively open source finetunes from CivitAI. In particular, we downloaded ~350 Stable Diffusion LoRA finetunes of vehicles and ~150 Stable Diffusion LoRA finetunes of buildings. We then trained two models – GLNet and MLP([U,V]) – on classifying whether a given LoRA was a vehicle or building. We find that GLNet is able to label publicly sourced finetunes in the test set with near perfect accuracy (99.2%), while a generic MLP which does not consider symmetries fails to attain significantly better than baseline performance. Results are reported below averaged across 5 runs.

Model	Test Acc	Test Loss
GLNet	99.2 ± 1.1	0.059 ± 0.013
MLP([U,V])	74.4 ± 2.6	0.530 ± 0.05

We consider these results to be **extremely** exciting. With only a few hundred publicly available finetunes as training data, GLNet can nearly perfectly classify the training data CivitAI users trained their models on. This suggests many new possible use-cases for our models, such as detecting mislabeled or illegal models on websites hosting public finetunes. We would like to thank the reviewer for encouraging this experiment, and request that they reconsider their score in light of the compelling results.

> While the authors demonstrate GL-Net's ability to generalize from LoRA to DoRA in zero-shot prediction, this transition appears theoretically expected since both employ low-rank matrix representations (UV^T). This suggests GL-Net may be limited to adapters expressible in this factorization form.

> A more critical generalisation challenge arises with Kronecker Adapters (KronA). As shown in Appendix E, KronA requires vectorisation to process through GL-Net, altering the original and dimensions. This shape transformation requires additional modifications of the model architecture to be able to process adapters of dirrenent structure.

While test-time generalization across different adapter types may not always be guaranteed due to varying decomposition shapes, we strongly believe that our LoL models remain very useful throughout different adapter domains for a few reasons:

1. In order to promote effective transfer learning across domains where adapter dimensions may not match (e.g. Kronecker adapters vs. LoRAs), MLP heads could be shared. In practice, we find that the majority of parameters are found in the MLP head, and the input dimension to the MLP head could be made to take on both the same shape and format for Kronecker adapters and LoRAs.
2. As demonstrated above, only a few hundred publicly available finetunes are needed to make a very effective LoL model.
3. In addition to LoRA, DoRA, and Kronecker adapters, other parameter efficient factorizations like Orthogonal Finetuning via Butterfly Factorization [Liu et al, ICLR 2024] can also be fit into the GL-Net learning framework. In this setting, GL-Net's equivariant linear layers could be applied to the outer dimensions of the first and last orthogonal matrices respectively (maintaining GL equivariance) and then a pooling operation could be performed by multiplying out the individual orthogonal matrices in this lower dimension.

Replying to Official Comment by Authors

discussion

Official Comment by Reviewer bUbu 07 Aug 2025 at 15:29 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:

Thank you for your quick response to addressed questions. Indeed, the experiment showing the ability to train GL-Net on only open-source data increases the applicability of the method. Its results are quiet significant, and they should be extended and included into final revision.

Previously, authors mentioned that they conducted experiment with rank-320 LoRA using Imagenette dataset. In this experiment, did authors downsample U_i and V_i to 32×320 as well as in previous ones? It seems like in this particular case such a downsampling may harm the overall performance, as it can omit significant features of U_i and V_i and decrease their rank.

My questions regarding matrix rank are as follows:

1. If LoRA's rank is more than 32, does such a downsampling preserve GL-Net quality? Could authors show, please, the results of the experiment with rank-320 Imagenette LoRAs? In the case of rank-320 LoRAs, was the adapters' matrix ranks really greater than 32 (or singular values of the adapters were relatively small)?
2. Did authors try any other downsampling parameters, larger than 64, and how significant the performance decrease is?

However, I am still not sure that this method could be generalised to adapters with a different structure even though shapes may coincide. Different parameter-efficient adaptations store features differently and it's unclear how GL-Net could capture diverse structured layers equivalently. It requires some specific rearrangement, or any other techniques to fix it. Otherwise, it means that one has to train many models to evaluate distinct parameter-efficient adapters, which requires some model «calibration» to fairly compare all these adapters.

Replying to discussion

Official Comment by Authors

Official Comment by Authors 07 Aug 2025 at 20:33 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:

Thank you for your quick response. We will extend our exciting results on open source models by including other LoL models and baselines and adding them to our paper.

> If LoRA's rank is more than 32, does such a downsampling preserve GL-Net quality? Could authors show, please, the results of the experiment with rank-320 Imagenette LoRAs? In the case of rank-320 LoRAs, was the adapters' matrix ranks really greater than 32 (or singular values of the adapters were relatively small)?

Indeed, downsampling a rank 320 LoRA to lower dimensions preserves GL-Net quality. Here are the results we mentioned:

Equivariant Downsampling Dimension	Test Acc (%) on Rank 320 LoRAs
32	85.7 ± 1.2
64	86.6 ± 1.9

A possible explanation for this high accuracy is that, as you suggested, "high-rank" LoRAs are still approximately low rank. We find that on average, 21 Singular Values in a rank 320 LoRA module account for 95% of the variance in matrix entries. We consider this to be further evidence supporting GL-Net's efficacy in high rank settings, because the relevant learning dynamics are not fundamentally different from lower rank settings. Furthermore, these results are not unique to our own generated LoRA finetunes; in the previously mentioned open source experiment, GLNet could successfully classify public rank 256 LoRAs with a hidden equivariant dimension of 64.

> Did authors try any other downsampling parameters, larger than 64, and how significant the performance decrease is?

Yes! Part of our experimental hyperparameter tuning process was testing a range of downsampling parameters. In particular, on rank 4 Imagenette-SD-1.4 we have the following results for GLNet. As you can see, most hidden dimensions above 32 perform about the same, with mild performance degradation occurring at lower dimensions.

Equivariant Downsampling Dimension	Test Acc (%)	Test Loss
16	86.4 ± 0.3	0.322 ± 0.003
32	87.3 ± 0.6	0.314 ± 0.014
64	87.9 ± 0.4	0.301 ± 0.007
128	87.9 ± 0.5	0.311 ± 0.007
256	87.5 ± 0.2	0.311 ± 0.003

> However, I am still not sure that this method could be generalised to adapters with a different structure even though shapes may coincide. Different parameter-efficient adaptations store features differently and it's unclear how GL-Net could capture diverse structured layers equivalently. It requires some specific rearrangement, or any other techniques to fix it. Otherwise, it means that one has to train many models to evaluate distinct parameter-efficient adapters, which requires some model «calibration» to fairly compare all these adapters.

We believe you are correct that generalization across some adapter domains is not guaranteed to be zero-shot and is a very interesting problem. As you mentioned, in practice, one could train multiple GL-Nets for different parameter efficient adapters individually (e.g. one for LoRA/DoRA, one for OFT, and one for KronA). This would not be cost prohibitive, because GL-Net does not require much training data in the first place. Calibrating multiple GL-Net models for different adapter types is also a very interesting idea, which we believe would be effective due to GL-Net's highly accurate prediction and relative lack of overfitting.

We hope we have answered your questions and demonstrated the utility of our methods, and would also like to thank the reviewer again for their insightful suggestions.

Replying to Official Comment by Authors

Official Comment by Authors

Official Comment by Authors 09 Aug 2025 at 02:12 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors

Comment:

Dear reviewer, we are checking back in before the deadline. We would like for you to consider our rebuttal and overall contributions for your final score and assessment of our paper. Recall that during the rebuttal, we

1. Demonstrated strong generalization across different LoRA structures and data sources, which exceeds the scope that we set out to do with this first paper in the area.
2. Ran several new ablations that show that our method robustly performs across hyperparameter settings.
3. Developed many new directions for future work (e.g. testing out our theories on shared MLP heads, and on additional datasets of public LoRAs).

Once again, our work is a first work in the area, and already establishes a framework, new metanetworks, new data, new experimental setups, SOTA empirical performance, and new theory. With the rebuttal, we additionally have very promising future science directions (e.g. more on zero-shot generalization) and more types of data to leverage (e.g. more public LoRAs). We would appreciate further consideration from the reviewer on these strengths of our work.

LoRAs). We would appreciate further consideration from the reviewer on these strengths of our work.