Respond to Reviewer 1 A common bias is that meta-learning should tackle transfer learning or few-shot learning problems. However, this is not always the case: *the setting of this paper* ***do not*** *fit nicely with transfer learning or few-shot learning*. This is because the learned neighbors are optimized using source domain data, which are useless and even harmful if we use them to adapt the model to unseen target domains. Similar to the setting of MAXL [1], the focus of our paper is to improve the general supervised learning performance via meta-learning.

As pointed out by ICLR 2019 AnonReviewer3 of the MAXL paper, "*Moreover, since the method is not a meta-learning approach for few-shot learning, it is not fair and also not appropriate to compare with Prototypical Network.*", we also think it is unreasonable to compare our work with MAML, prototypical networks and [2].

Table 1: Updated results for regression.

Datasets	n	d	kNN	*vanilla*	Meta-Neighborhoods
music	515345	90	0.6812±0.0062	0.6236±0.0056	**0.6088±0.0050**
toms	28179	96	0.0602±0.0083	0.0594±0.0080	**0.0531±0.0073**
cte	53500	384	0.00134±0.00023	0.00121±0.00022	**0.00109±0.00015**
super	21263	80	0.1126±0.0061	0.1132±0.0060	**0.1077±0.0068**
gom	1059	116	0.5982±0.0521	0.5949±0.0515	**0.5681±0.0563**

It is not advisable to evaluate the degree of improvements without considering the room available for improvements. Our improvements are **significant** as: (1) they are greater than those achieved by MAXL on almost all datasets (2) according to line 240-243, backbones used in our work are already strong, and our work is more effective than naively increasing the backbone depths. We report results on the 1000-class ImageNet classification. As shown in Fig 1, MN+iFiLM improve *vanilla* from 48.4% to 54.1%. Again, this improvement is larger than that achieved by MAXL. To facilitate experiments, we resize images to 64×64 resolution.

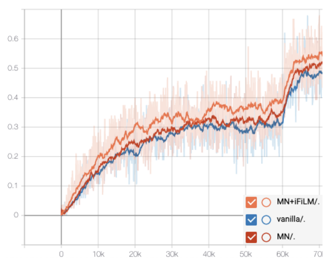

Figure 1: Top-1 Validation Accuracy on Imagenet.

For regression results, we provide results of kNN in Table 1, which are inferior to Meta-Neighborhoods. We also perform statistical significance test (paired Student's t-test) to show the results of Meta-Neighborhoods and *vanilla* are statistically different: the p-value of music, toms, cte, super and gom are 0.00039, 0.0018, 0.018, 0.0076 and 0.00089, which are all smaller than the Significance Level $\alpha = 0.05$.

We hope our response can address most of your concerns and sincerely hope you can re-consider your score.

Respond to Reviewer 2 In fact, we didn't observe optimization difficulties when training all variables together due to the following reasons: (1) we observed the pseudo-NNS can be easily initialized as Gaussian and not sensitive to the std of Gaussian (2) learning rate is only a scalar and thus easy to optimize (3) although the feature extractor receives error signals from the finetuned ϕ_i, ϕ_i can be expressed as $\phi_i = \phi - \alpha \nabla_\phi L_i^{inner}$ where ϕ acts as a "short cut" to back-propagate errors to the feature extractor. Besides, our model is not sensitive to the choice of datasets. Neglecting magnitude actually does not harm the final performance as shown in [3]. On the contrary, it adds robustness by maximizing inter-class differences.

Respond to Reviewer 3 Both memory-augmented neural nets and memory matching nets tackle few-shot problems where the raw features are given, while our work does not consider few-shot tasks. Therefore, the raw features are not given in our case and we propose to meta-learn them. The effectiveness of iFiLM has been validated: MN+iFiLM is always better than MN. Please refer to Appendix A.3 and A.7 for parameter number and time complexity information.

Respond to Reviewer 4 It is not advisable to evaluate the degree of improvements without considering the room available for improvements. Our improvements are **significant** as: (1) they are greater than those achieved by current STOA method MAXL [1] on almost all datasets (2) backbones used in our work are already strong, which leaves limited room for large improvements. According to line 240-243, our work is more effective than naively increasing the backbone depths.

Besides, our work has already produced a good performance for large-scale tasks that consist of many classes (e.g. 200-class classification on Tiny-Imagenet). To validate this claim, we further report results on 1000-class ImageNet classification. As shown in Fig 1, MN+iFiLM improve *vanilla* from 48.4% to 54.1%. Again, this improvement is larger than that achieved by MAXL [1]. To facilitate experiments, we downsampled image resolution to 64×64.

Overall, we sincerely hope this response can address your concerns and you can re-consider your score.

Reference

[1] Self-supervised generalisation with meta auxiliary learning. NeurIPS 2019.

[2] Memory matching networks for one-shot image recognition. CVPR 2018.

[3] Robust classification with convolutional prototype learning. CVPR 2018