We thank all reviewers for taking their time reading the paper and providing us with insightful comments and suggestions!

To R1: Thank you for appreciating our work! Here are our responses.

• Regarding "results from a fixed learning rate": There may be some misunderstandings. We do not use a fixed learning rate for any of our experiments, and the schedule can be found in Appendix B.

To R2: Thanks for your very careful review. We will rephrase the sentence you mentioned and add citations and discussions on the related works you provided. Here are our responses to your questions.

• Regarding the sanity check on Iterative Magnitude Pruning (IMP): That is a good question, we focus on one shot pruning in this paper for the following reasons: (1) As suggested by the result of Figure 7 in [Rethinking the value of network pruning, 23] , if choosing proper learning rate (0.1, also suggested by [Rewinding paper, 33]), other than 0.01 in the LTH and [Learning both Weights and Connections for Efficient Neural Networks, 12] papers, iterative pruning cannot consistently outperform one shot pruning, and even when iterative pruning is better, the gap is small; (2) Compared to one shot pruning, iterative pruning is obviously very time-consuming, and since our goal of the paper is propose sanity checks for pruning methods, we chose to test on the rather efficient one shot pruning method. However, we followed the reviewer's suggestion and conduct experiments on IMP settings. The results show that LT with IMP has similar performance as one-shot LT and cannot pass the sanity checks in most settings; while IMP can boost the performance of learning rate rewinding, which can pass the sanity checks as expected. The smart-ratio random tickets can still outperform LT with IMP in most settings; while the improvement of hybrid tickets with IMP upon learning rate rewinding with IMP becomes slighter or vanish in some settings (please see "Common reply" in the bottom for more discussion on hybrid tickets). We will add the discussions in the next version.

• Regarding "half dataset or shuffled weights doing no worse than LT": This is an understandable result that applying sanity check can hurt the performance of LT with rewinding, but the performance will not drop to under naive LT since even naive LT will keep similar performance after applying sanity checks.

• Regarding the sanity check besides CIFAR-10: Thanks for the suggestion. Based on the empirical observation on CIFAR-10, we proposed "random tickets" method in the paper. And the success of random tickets, in turn, corroborates that we can achieve the performance of initial tickets without data information and certain network connection pattern. We conducted experiments of random tickets on several datasets and architectures, and we believe these results are enough for illustrating our ideas. We conduct sanity checks on other datasets in the version if the reviewer still suggests.

• Regarding the related results in [One Ticket paper, 28] and the lack of higher sparsity results: Thanks for pointing out the very related results, we will add discussion on the difference in the next version. For the high sparsity regime, we find that initial tickets still cannot pass the sanity checks. The difference between our observation and the figure you mentioned may come from the "late resetting" trick used in [One Ticket paper, 28] (See the description in Section 3.1 in that paper), which is a weak version of rewinding. So it is not surprising to see that Figrue A1 in [One Ticket paper, 28] show evidence that the method may pass the layerwise rearrange check.

• Regarding some minor questions: (1) Each data point shown in the table/figure is obtained by averaging of 3 runs (we would like to refer you to our Appendix B to see the detailed settings). (2) The trend of smart-ratios is similar to keep-ratios from existing methods, we would like to add a visualized comparison in the next version.

To R3: Thank you for the careful review and helpful suggestions on writing and related work! We will polish the paper and add citations as you suggested. Here are our responses to your questions.

• Regarding the large-scale dataset: For sanity-checking existing pruning methods, we follow the setting of the original papers [SNIP paper, 20; GraSP paper, 36] to conduct our experiments on CIFAR-10/100 and Tiny-ImageNet. But as the reviewer suggests, compared to exsiting methods, it will be more efficient to apply our random ticket on datasets like ImageNet, we will conduct experiments to test the performance of our method on ImageNet.

• Regarding the generalizability of smart-ratio random tickets: We did examine different experiments settings to test the generalizability of smart-ratio random tickets. Please see Appendix C for more experiments on several different ResNet/VGGNet architecture.

• Regarding the claim of "the architecture is crucial": In their original paper the pruning phase is described as "learning which connections are important and removing the unimportant connections", which implies the architecture is crucial.

Common reply: We notice all the reviewers pointed out the improvement of our Hybrid Tickets is not always significant. We would like to point out that in this setting, the baseline (learning rate rewinding) itself is very strong, the performance of networks pruned by this method can be even better than the original network. So we can expect that the room for improvement is limited. However, our hybrid tickets still show strengths especially in the "hard settings" with hard dataset (CIFAR-100) and high sparsity.