We thank the reviewers for the thoughtful comments. We appreciate all of the reviewers found the experiments compelling and the paper was well written.

R2, R3, R4: Concerns about novelty. The novelty of our contribution lies in the simplicity of the method. Our proposed method significantly simplifies automated data augmentation search. We emphasize that advances in data augmentation strategies can be as efficacious as advances in architectural changes. Many crucial deep learning improvements have come from simplifying once complex ideas. Furthermore, we show, for the first time, that data augmentation strength depends on training set size with an unintuitive relationship. Finally, while other proposed augmentation methods only work on specific datasets or tasks, our method works on both classification *and* detection. The novelty of our method can also be seen by the fact that many new semi-supervised and self-supervised papers utilize this method to achieve SOTA (e.g. UDA, Noisy Student, FixMatch, ReMixMatch, Tian & Sun et al, Tian & Krishnan et al, Khosla et al.).

R1:"different magnitude to different ops?" Thanks for this great suggestion. In the paper we have evaluated if results can be improved by optimizing the magnitudes for different ops individually. Please see Fig.4 in the Appendix, we found that for the larger model it is possible to improve the results by tuning individual magnitudes, but not for the smaller model. "this paper seems to mix two related, but distinct goals" We agree with R1 that it would be interesting to do a more careful analysis on the search phase. We have focused our attention to the search space, and found that when the search space is chosen carefully, the search method can be as simple as grid search. We wanted to make data augmentation as simple as possible for classification and detection, and did not see the need to employ a more complicated search algorithm. "Line 158-160: K^N policies?" R1 is completely right, K^N outcomes would be more correct. We used the term policies here, to place our method in the context of AutoAugment, which would call each of our K^N outcomes a policy. We will clarify this in the text. "Line 167-168: What does M mean precisely" M is the global distortion magnitude that is used by all ops. As R1 mentions, one needs to determine the schedule for M during training. However, as mentioned in Appendix A.1, constant magnitude works as well as other schedules. In order to keep things simple, we used constant magnitude for all experiments in the paper. Then the only decision that needs to be made about M is its constant magnitude, which is optimized for each model as described in text. "range of M and N" We listed the values we tried in Section 5 and Appendix A.5. Briefly, we tried $N = \{1, 2, 3\}$ and M between 4 and 28.

R2:"it is not surprising that this kind of random augmentation policy could improve the models" We respectfully disagree on this comment. First, our approach is not same as a random policy. AutoAugment [3] evaluated the random policy performance, which is not very good. It would only be worse for larger ImageNet models compared to RandAug (see line 215 for explanation). In our paper, we are not just evaluating a random policy. We are proposing a new search space, which allows even grid search to get SOTA, and discover a positive correlation between training set size and augmentation magnitude. If there are published references that describe such findings, we would love to know. Papers have been published in ICLR/ICML/NEURIPS recently improving the AutoAugment search algorithm. If it were obvious that a simple approach could achieve as good results, we assume those papers would not have been accepted to such leading conferences. PBA suggests dynamic schedules are better We achieve comparable results to PBA ($\pm 0.1\%$) on small datasets (Table 2) while only using fixed policies. On more realistic datasets such as ImageNet and COCO, PBA was *never* evaluated by the authors. In contrast, our method achieves SOTA on ImageNet and strong improvements on COCO. Our method has the added benefit of not requiring a complicated search. We did evaluate dynamic policies (Appendix, Table 5) and found that constant magnitude performs just as well as dynamic schedules and PBA. Thus we see no limitation for fixed policies. Adv. AA and "Online hyperparameter..." comparisons are missing. Adv. AA achieves a better result on CIFAR-10 by 0.1%, however at a significantly increased complexity (which might be the reason they could not evaluate on larger ImageNet models or object detection), and our best result is better on ImageNet by 4.1% with ENet-B8 (they did not evaluate on SOTA architectures, and their policies are not publicly available for comparison). RandAug's strength comes from its ability to scale to large models easily. We do note that these two suggested papers are very impressive, thanks for bringing them up! We will update the paper and L13.

R3:"the difference with Yu et al" The cited work focuses on a shared goal of simplifying AutoML, and demonstrates that a random policy may work well for architecture search. In our paper, we do not propose employing a random policy for data augmentation. In fact, random policies were evaluated in the original AutoAugment paper, where it was found that random policies do not perform as well as reinforcement learning. Instead, we propose a new, simplified search space ($10^{32} \rightarrow 10^{2}$) which outperforms AutoAugment. Note that our analysis also explains why a random policy would not do very well, since it cannot adjust its strength on the model and dataset size. We will cite this paper and discuss its relevance. "details of the search space" We will add more details including the time cost and fix the typos.

R4:"Random search vs. grid search" Since our optimization is in 2D, we do not expect there to be a large difference between random search and grid search for magnitude. However we agree that this is a great suggestion, and we will easily add this to the paper. "Ablation for number of transformations" We already ran this ablation (please see Fig.3 in main text and Table 6 in the Appendix). However, we agree that it would be interesting to compare some of these results with the baselines, which we will add to the paper. Thanks again for the great suggestion!