We sincerely thank four reviewers for the constructive comments. We have conducted experiments to address almost all questions. As the issue about random models is concerned by Reviewer #1, #2 and #4, we answer this first.

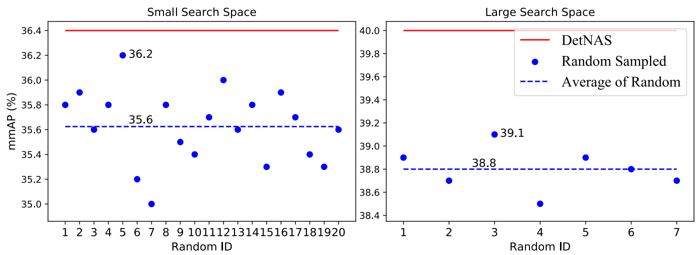

Figure R-1. Random models on FPN in small and large search space.

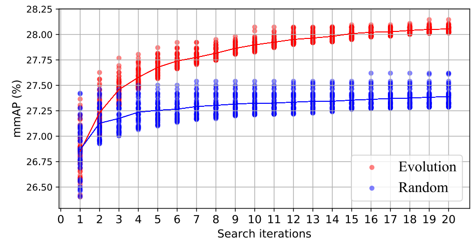

Figure R-2. mmAP during search for main results.

Q: *Issues on the random models.* **A:** To clarify this issue, three additional experiments have been conducted: **(1)** As in Figure R-1 (left), 15 additional random models are sampled and trained in the small search space (20 in total, with the original 5 models in the paper). We depicted the scatter and the average line of random models and the line of DetNAS. DetNAS in the small search space is 36.4% while the Random is 35.6±0.6%. **(2)** As in Figure R-1 (right), 7 random models are sampled and trained in the large search space (The maximum number of models we can train in this week). DetNAS in the large search space is 40.0 while Random is 38.8% (+1.2%) in average with the highest 39.1% (+0.9%), which is a large margin. **(3)** As in Figure R-2, the mmAP curve on the supernet search are depicted to compare EA with Random. For each iteration, top 50 models until the current iteration are depicted at each iteration. EA demonstrates a clearly better sampling efficiency than Random.

These comparisons show that: **(1)** In the large search space for main results, it is difficult to randomly pick a suitable model. The improvement is not marginal (+1.2%/0.9%). **(2)** In the small search space, DetNAS is clearly better than Random while the improvement is not large enough as constrained by the small search space. **(3)** As in Figure R-2, EA is much more sampling efficient than Random search. Thanks the reviewers for the constructive suggestions on this issue. We hope that these experiments can clarify the concerns. These experiments will be included in the next version.

For Reviewer #1

Q: *Inference time.* **A:** For the concern on the inference, we measure the FPS of DetNASNet and ResNet on the same Tesla V100 with (800,1200) input size, as in Table R-1. Under the same mAP, DetNASNet processes 5 more frames per second than ResNet-101. Under the same FLOPs, DetNASNet (3.8) is only 2 FPS slower than ResNet-50 but has a much better mAP (42.0 vs 37.3). The latency of DetNASNet is not much higher than ResNet under the same FLOPs.

Q: *More random models.* **A:** Thanks for the beneficial suggestion. More random models are provided as above. The improvement is constrained by the small search space. In the large search space, the gap is clear (+1.2%/0.9%).

Table R-1. Inference and mAP of ResNet and DetNASNet on FPN.

	ResNet-50	ResNet-101	DetNASNet	DetNASNet (3.8)
FPS	17.9	15.3	**20.4**	15.8
$mAP_{1\times}$	37.3	40.0	40.0	**42.0**

Table R-2. Comparison with the 3.8G ShuffleNetv2.

	ImageNet	$mAP_{1\times}$	$mAP_{2\times}$
ShuffleNetv2	**78.47**	40.8	42.4
DetNASNet (3.8)	78.44	**42.0**	**43.4**

For Reviewer #2

Q: *Comparison with a 3.8G FLOPs ShuffleNetv2.* **A:** Thanks for this suggestion. In Table R-2, the results of a 3.8G FLOPs ShuffleNetv2 are provided and compared with 3.8G DetNASNet in both 1x and 2x training settings on FPN.

Q: *mmAP curve of during EA search.* **A:** The search curve of EA against random search is shown in Figure R-1. It shows top 50 models in each iterations. Parent models are top 10 among them. EA is more efficient than Random.

For Reviewer #3

Q: *A more general search space that includes Hourglass.*

A: Designing a search space that includes Hourglass is a promising and interesting idea. We would like to try it in the future work.

Table R-3. Comparison with CenterNet+Hourglass.

	n.a	+flip	+multi scale
CenterNet-Hourglass104	40.3	42.2	45.1
FPN-DetNASNet (3.8)	43.4	44.8	**46.1**

Q: *The practical utility of their methods.* **A:** In terms of accuracy, we obtain 46.1 mmAP by simply following the test augmentations used by CenterNet as in Table R-3. CenterNet is one-stage while Hourglass104 has much more FLOPs than 3.8 G. In terms of FPS, DetNASNet has a clear superiority as in Table R-1. Better detection strategies are important for the practical utility. The focus of this paper is backbone search, thus we leave this in the future work.

Q: *Writing.* **A:** Thanks for your suggestions on writing. We will revise the paper thoroughly under your suggestions.

For Reviewer #4

Q: *The gap between Random and DetNAS.* **A:** Thanks for pointing out the concern of random models. We provide the search curve and more results of random models in both the small and large search space. The reason for the small improvement mainly comes from that the search space for the ablation studies is not large enough and the final results trained are close. In the large search space, the improvement is clear (+1.2%/0.9%) as shown in Figure R-1 (right).

Q: *Minors.* **A:** Thanks for your suggestions. We will revise the abstract for this point carefully in the next version.