We thank the reviewers for the constructive feedback. **Code** will be made public. Fig. (a, b, c) best viewed in zoom.

### R2.1: Difference from PathNet:
Our RPS-Net is inspired by PathNet, yet there are notable differences: 1) **Architecture:** PathNet adds all modules in a simple feed-forward manner, while RPS-Net adds modules as residual connections to an identity path ($M_{skip}$) which can always be learnt even when all the modules are saturated (L263). RPS-Net can therefore adapt to new changes more robustly. 2) **Training setting:** PathNet uses evolutionary algorithm while RPS-Net employs a random path selection. Since PathNet usually involves only two tasks, an evolutionary algorithm with large explorations/generations is viable. However, for our case i.e., 10+ tasks, PathNet is not feasible, due to a large number of possible explorations. See **R3.1** for comparison between random selection and genetic algorithms.

### R2.2: Impact of Varying Examplars:
Fig (c) compares RPS-Net with the best existing method (iCARL) for various memory budgets of examplars on CIFAR100. Our proposed RPS-Net consistently performs better across all budgets, increasing significantly with tasks. After 10 tasks, RPS-Net has 72.67M parameters on average, compared with iCARL (21.3M) and Progressive Nets (932.84M). In RPS-net parameters & flops increase logarithmically (L293), while for Progressive Nets they increase quadratically. We will add a detailed discussion in the final version.

### R2.3: Model Size Comparison:
Fig (b) compares total parameters across tasks for Progressive Nets [21], iCARL [20] and our RPS-Net on CIFAR100. Our model effectively reuses previous parameters, and the model size does not increase significantly with tasks. For BTS to gain similar performance as our random selection, it needs an average of 2.54M parameters and multi-path selection modules (+6.72%), which are our two major contributions. Further, our two contributions, controller and multi-path selection, provide a combined gain of 13.6% over baseline + distillation.

### R2.4: Impact of Augmentation:
For a fair comparison, we use **same** augmentation as in iCARL [20].

### R2.5: Reproducing Algo. in Sec. 3:
We'll improve the description of Sec. 3, and release our codes for reproducibility.

### R2.6: Ablation Study:
Fig (a) studies the impact of progressively integrating individual components of our RPS-Net: baseline Single Path (37.9%) ⇒ distillation (44.93%) ⇒ controller $\phi(k, \gamma) (51.76\%) ⇒$ multi-path selection (58.48%). These results (classification accuracy) on CIFAR100 indicate that the most significant gains come from controller (+6.83%) and multi-path selection modules (+6.72%), which are our two major contributions. Further, our two contributions, controller and multi-path selection, provide a combined gain of 13.6% over baseline + distillation.

### R3.1: Ablation Analysis:
Impact of different components are shown in Fig (a) on CIFAR100. See **R2.6** for details.

### R3.2: Difference from Genetic Algorithms:
We compare our random selection with a genetic algorithm i.e., Binary Tournament Selection (BTS) for 25 maximum generations, on MNIST with 5 tasks (each of 2 classes), using a simple 2 layer (100 neurons) MLP with $M = 8, J = 1$. On 5 runs, our proposed random selection achieves an average accuracy of 96.52% vs BTS gets 96.32%. For same time complexity as ours, BTS has an average accuracy of 71.24% for the first generation models. For BTS to gain similar performance as our random selection, it needs an average of 10.2 generations (> # random paths), hence BTS has more compute complexity. Sophisticated genetic algorithms may beat random selection, but likely with a high compute cost, which is not suitable for an incremental classifier learning setting having multiple tasks. Please also see **R2.1**, and L159, L160. We’ll elaborate this further in the final version.

### R3.3: Network Size and Capacity:
Fig (b) shows the comparison on CIFAR100. Also see **R2.3** for details.

### R3.4: Temperature $t_c = 2$ is empirically set on a small cross-validation set and fixed throughout experiments.

### R4.1: Revising the Claims:
We first thank R4 for appreciating technical innovations and contributions of our work. We also thank R4 for valuable suggestions to improve the manuscript writing. In the final version, we will revisit and tone down some of our claims as recommended.

### R4.2: Error Bars:
For five runs, mean and standard deviation of RPS-Net for tasks 1 to 10 are respectively: 88.96±0.27, 81.07±0.22, 77.15±0.55, 72.98±0.56, 68.69±0.58, 66.71±0.43, 64.47±0.34, 62.06±0.23, 60.23±0.58, 58.51±0.49. We will add this in the final version.

### R4.3: Backwards Transfer:
We validate Backwards Transfer with BWT metric (See Eq. 3 in GEM [16], larger the better). After last task, BWT values are -0.1462 (RPS-Net) vs. -0.4602 (iCARL). We will add the results with further discussion in the final version.

### R4.4: Significance Beyond Incremental Classifier Learning:
We thank R4 for the useful insight and will explore the applicability of our approach to other domains in future.