We sincerely thank all reviewers for their feedback. All experiments are conducted on Meta-dataset and averaged performance of each group is reported. Ai, Fl, Fu denote Aircraft, Flower, and Fungi, respectively. The performances of FTML and DPM are also reported for comparison. We will correct typos, add the suggested references with discussions, and improve the overall presentation.

To Reviewer 1
- Regarding novelty: We propose a generic framework for doing online meta-learning that not only works under homogeneous but also heterogeneous settings. We evaluated our work under the same setting as compared to [9]. We choose to use [20] as an implementation for knowledge pathway construction. Our method is flexible and one could choose to use other method in this step, e.g. evolutionary or reinforcement learning strategy.
- Results of fungi (58.45 ± 2.89%) is significant better than best baseline (DPM: 50.15 ± 3.53%).
- Task learning efficiency and Performance v.s. # of samples: In heterogeneous datasets, the performances fluctuate across different tasks due to non-overlapped classes. Similar to [20] (Fig 4), learning efficiency is evaluated by # of samples. The performance of OSML w.r.t. # of samples are reported in Table 1, showing that OSML is able to consistently improve performance. For Rainbow MNIST, we follow [9] to analyze the amount of data needed to learn each task and show the results in Figure 1. Notice that our method requires less number of samples to reach the target accuracy as compared to the other methods, which indicates our ability to efficiently learn new tasks.

To Reviewer 2
- Importance of pathway construction: We have conducted two experiments to show the effectiveness of meta-knowledge pathway selection. First, fine-tuning (FT) is used in Fig. 2 - 4, where only one knowledge block each layer is involved and fine-tuned for each task. This corresponds to the case where no meta-knowledge pathway is getting constructed. Second, we also use FTML with the same number of parameters as the final number of blocks in the meta-hierarchical graph (i.e., same architecture) and discuss the performance in L227-238. OSML consistently outperforms these two baselines.
- Homogeneous v.s. Heterogeneous Settings: We follow the setting that is suggested in [38,40,41]: (1) non overlap between classes; (2) the underlying distribution is multimodal. The suggested setting from [14] is also valuable and we conduct additional experiment to evaluate it by feeding Fl->Ai->Fu. The performance obtained are [FTML: (Ai66.51 | Fl67.28 | Fu50.89), DPM: (Ai68.85 | Fl68.90 | Fu53.20), OSML: (Ai71.35 | Fl72.09 | Fu57.00)], demonstrating the effectiveness of OSML.

To Reviewer 3
- New block bottleneck: When a new task arrives, we randomly select two sets $D^\text{supp}$ and $D^\text{query}$ to search the functional regions (L 4-5 in Alg. 1). For the new initialized blocks, motivated by [20], it can access the corresponding data and be tuned at this time. Therefore, the new block is possible to get selected.
- Ablation study: We have designed two models to show the importance of new block construction and path selection. Please kindly refer to "Importance of pathway construction" of Reviewer 2.
- Clarification of first-order approximation: We use FOMAML [7] rather than Reptile for first-order approximation, where the loss of FOMAML is calculated on query sets with adapted parameters.
- Discussion with previous work: We have discussed five recent studies that focus on task heterogeneity in both Introduction (L36-41) and Related Work (L250-258). The list of citations (i.e., [26,38,40,41]) focus on the stationary scenario. [14] focuses on non-stationary task distribution and has been discussed in detail.

To Reviewer 4
- Discussion with HSML: The major difference between OSML and HSML are two-fold: (1) The settings are different. HSML focuses on the stationary task distribution while OSML tackles the challenge of non-stationary setting. The continual adaptation setting is evaluated under the stationary scenario. In Figure 4 and 7, curves show meta-training performance and tables report stationary meta-testing results. (2) HSML captures task heterogeneity by customizing the model initialization using task-wise cluster representation. While OSML encourages the layer-wise knowledge block exploitation and exploration, which improves the flexibility of knowledge transfer. Though the original HSML is evaluated on stationary scenario, to make comparison we have evaluated the model under our setting by introducing task-awared parameter customization and hierarchical clustering structure. The performances are (Ai64.33 | Fl62.75 | Fu52.18) compared to (Ai67.99 | Fl68.55 | Fu58.45) shown in table 4 in our paper, which further verifies the effectiveness of OSML.