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

5 **To Reviewer 1**
6 **- Regarding novelty:** We propose a generic framework for doing online meta-learning that not only works under
7 homogeneous but also heterogeneous settings. We evaluated our work under the same setting as compared to [9]. We
8 choose to use [20] as an implementation for knowledge pathway construction. Our method is flexible and one could
9 choose to use other method in this step, e.g. evolutionary or reinforcement learning strategy.
10 **- Results of fungi** (58.45 $\pm$ 2.89%) is significant better than best baseline (DPM: 50.15 $\pm$ 3.53%).
11 The comments on Aircraft result is true, and we have discussed it in L210 - 214 in the submission.
12 **- Task learning efficiency and Performance v.s. # of samples:** In heterogeneous
13 datasets, the performances fluctuate across different tasks due to non-overlapped classes.
14 Similar to [20] (Fig 4), learning efficiency is evaluated by # of samples. The performance
15 of OSML w.r.t. # of samples are reported in Table 1, showing that OSML is able to
16 consistently improve performance. For Rainbow MNIST, we follow [9] to analyze the
17 amount of data needed to learn each task and show the results in Figure 1. Notice that
18 our method requires less number of samples to reach the target accuracy as compared
19 to the other methods, which indicates our ability to efficiently learn new tasks.
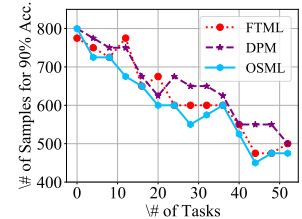
20 **To Reviewer 2**
21 **- Importance of pathway construction:** We have conducted two experiments to show
22 the effectiveness of meta-knowledge pathway selection. First, fine-tuning (FT) is used
23 in Fig. 2 – 4, where only one knowledge block each layer is involved and fine-tuned for
24 each task. This corresponds to the case where no meta-knowledge pathway is getting
25 constructed. Second, we also use FTML with the same number of parameters as the final
26 number of blocks in the meta-hierarchical graph (i.e., same architecture) and discuss
27 the performance in L227-238. OSML consistently outperforms these two baselines.

Model-# Data	Ai	Fl	Fu
FTML-20	55.92	54.65	45.69
DPM-20	57.15	53.34	44.33
OSML-20	**60.18**	**58.28**	**48.25**
FTML-30	62.88	58.37	47.96
DPM-30	64.43	59.72	48.10
OSML-30	**66.57**	**65.07**	**53.72**
FTML-40	65.85	59.05	48.00
DPM-40	66.67	63.40	50.15
OSML-40	**67.99**	**68.55**	**58.45**

Figure & Table 1

28 **- Homogeneous v.s. Heterogeneous Settings:** We follow the setting that is suggested
29 in [38,40,41]: (1) non overlap between classes; (2) the underlying distribution is mul-
30 timodal. The suggested setting from [14] is also valuable and we conduct additional
31 experiment to evaluate it by feeding Fl->Ai->Fu. The performance obtained are [FTML: (Ai66.51 | Fl67.28 | Fu50.89),
32 DPM: (Ai68.85 | Fl68.90 | Fu53.20), OSML: (Ai71.35 | Fl72.09 | Fu57.00)], demonstrating the effectiveness of OSML.
33 **To Reviewer 3**
34 **- New block bottleneck:** When a new task arrives, we randomly select two sets $\mathcal{D}_t^{supp}$ and $\mathcal{D}_t^{query}$ to search the
35 functional regions (L 4-5 in Alg. 1). For the new initialized blocks, motivated by [20], it can access the corresponding
36 data and be tuned at this time. Therefore, the new block is possible to get selected.
37 **- Ablation study:** We have designed two models to show the importance of new block construction and path selection.
38 Please kindly refer to "Importance of pathway construction" of Reviewer 2.
39 **- Clarification of first-order approximation:** We use FOMAML [7] rather than Reptile for first-order approximation,
40 where the loss of FOMAML is calculated on query sets with adapted parameters.
41 **- Discussion with previous work:** We have discussed five recent studies that focus on task heterogeneity in both
42 Introduction (L36-41) and Related Work (L250-258). The list of citations (i.e., [26,38,40,41]) focus on the stationary
43 scenario. [14] focuses on non-stationary task distribution and has been discussed in detail.
44 **To Reviewer 4**
45 **- Discussion with HSML:** The major difference between OSML and HSML are two-fold: (1) The settings are different.
46 HSML focuses on the stationary task distribution while OSML tackles the challenge of non-stationary setting. The
47 continual adaptation setting is evaluated under the stationary scenario. In Figure 4 and 7, curves show meta-training
48 performance and tables report stationary meta-testing results. (2) HSML captures task heterogeneity by customizing the
49 model initialization using task-wise cluster representation. While OSML encourages the layer-wise knowledge block
50 exploitation and exploration, which improves the flexibility of knowledge transfer.
51 Though the original HSML is evaluated on stationary scenario, to make comparison we have evaluated the model under
52 our setting by introducing task-aware parameter customization and hierarchical clustering structure. The performances
53 are (Ai64.33 | Fl62.75 | Fu52.18) compared to (Ai67.99 | Fl68.55 | Fu58.45) shown in table 4 in our paper, which further
54 verifies the effectiveness of OSML.