

## 1    **A    Supplementary Materials**

### 2    **A.1    Experiment configurations**

3    To perform gpSR, we used the AMD EPYC 7702 64-Core Processor as well as the intel core 10700K  
4    processor. The computation time listed below are counted for the AMD EPYC 7702 64-Core  
5    Processor on  $\mathcal{P}_1$ :

- 6            • *mom*(0.5): 597s
- 7            • *Adam*: 545s
- 8            • DM: 610s
- 9            • RP: 569s
- 10          • RP-si: 602s

11   For the hyperparameters, we list below the configurations we used:

12   For gpSR, we use the following configurations we have tested to be efficient:

- 13            • number of iterations (influence the final diversity): 100
- 14            • n-cycles-per-iteration: 300
- 15            • populations: 8
- 16            • max-depth: 9999
- 17            • max-size: 500
- 18            • batching: False

19   Other training configures:

- 20            • learning rate for meta-training SR: 0.0001
- 21            • meta-optimizer used for meta-training SR: SGD optimizer with 0.8 momentum

### 22   **A.2    Hardwares**

23   For the large scale meta-training, we used the NVIDIA Quadro RTX A6000 GPU with 48GB total  
24   memory; however, the actual memory usage during meta train the optimizer on ResNet-50 is no  
25   greater than 30GB.