A. MA2ML and MA2ML-Lite on CIFAR-100

Figure 6 presents the average reward curves of top-30 pipelines and the scatter plot of the average reward in each batch of MA2ML and MA2ML-Lite on CIFAR-100. The large lead of MA2ML over MA2ML-Lite validates the effectiveness of MA2ML.

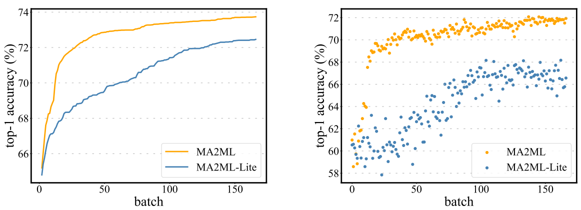

Figure 6. Learning patterns of MA2ML and MA2ML-Lite on CIFAR-100. Left: average accuracy curves of top-30 pipelines in terms of batch numbers. Right: the scatter plot for the average accuracy of pipelines in each batch.

B. Experiment Details

Detailed search space. In Section 4.2, we consider 15 augment operations, including ShearX/Y, TranslateX/Y, Rotate, AutoContrast, Invert, Equalize, Solarize, Posterize, Contrast, Color, Brightness, Sharpness, and Cutout. On CIFAR-10 and CIFAR-100, we adopt the same search space of NASNet [54] for NAS, and use [0.16, 0.08, 0.04, 0.02] and [0.0008, 0.0004, 0.0002, 0.0001] as the search space of warmup learning rate and weight decay for HPO, respectively. On ImageNet-200, we utilize the same search space of FBNetV3 [5] to form NAS and HPO.

RL setting. In MA2ML and MA2ML-Lite, for each module we use a one-layer LSTM controller with 100 hidden units to generate action distributions. On CIFAR-10, CIFAR-100 and ImageNet-200 datasets, we set $\eta_\theta = 0.0004$ and $\eta_\phi = 0.005$ for the learning rate of policy parameter θ in (13) and critic parameter ϕ in (11), respectively. We set $\tau = 0.004$ on CIFAR and $\tau = 0.012$ on ImageNet-200 as the moving average parameter for the update of target policy parameter $\bar{\theta}$ in (14). We set $\lambda = 0.2$ as the coefficient of KL divergence regularization.

Training details. For the short-term training on CIFAR-10 and CIFAR-100 in the search, the batch size is set to 64 per GPU. Two GPUs are used to train a model for 12 epochs with 1 epoch warm-up. For the retraining, the batch size is set to 64 per GPU. Eight GPUs are used to train a model for 600 epochs with 5 epochs warm-up. The learning rate of the model is 4 times larger than that of the short-term training. For the short-term training on ImageNet-200, the batch size is set to 128 per GPU. Eight GPUs are used to train a model for 100 epochs with 5 epochs warmup. For retraining on ImageNet, the batch size is set as 256 per GPU. Sixteen GPUs are used to train a model for 400 epochs with 5 epochs warmup. The learning rate of the model is 4 times larger than that of the short-term training.

Detailed results. On CIFAR-10 and CIFAR-100, the best searched pipeline is trained for 5 times. On CIFAR-10, MA2ML achieves 97.72%, 97.84%, 97.85%, 97.75%, and 97.7%, while MA2ML-Lite achieves 97.66%, 97.59%, 97.85%, 97.6%8, 97.73% top-1 accuracy. On CIFAR-100, MA2ML achieves 84.83%, 85.14%, 85.15%, 85.14%, and 85.15%, while MA2ML-Lite achieves 84.68%, 84.92%, 84.94%, 84.73%, 84.75% top-1 accuracy. On ImageNet, the searched pipeline is trained for a one time. MA2ML-D/E/F achieve $80.5\%/80.7\%/81.1\%$ top-1 accuracy under 795M/890M/973M FLOPs.

Results on ImageNet without constraints of FLOPs. On ImageNet, we also test MA2ML and MA2ML-Lite without any constraints of FLOPs. We take the reward function $R = \mathrm{Acc}(m)$, and keep other settings the same as experiments with constraints of FLOPs on ImageNet. MA2ML achieves 82.0% with 1.4G FLOPs and MA2ML-Lite 81.5%/1.2G, (both search for 1000 pipelines and select top-20 for final training and testing on ImageNet).

Search cost. Tesla V100s are used for experiments. On CIFAR-10/100 the search cost of MA2ML is 300 GPU days. On ImageNet the search cost is 2000 GPU days. For comparison, NASNet spends 2000 GPU days for search on CIFAR-10 with the same search space, which is much larger than MA2ML as the off-policy learning of MA2ML significantly improves sampling efficiency.