Method	γ-sampling	CIFAR10 Accuracy	SVHN Accuracy	PATH Accuracy	OCT Accuracy	BLOOD Accuracy	ORGANA Accuracy	ORGANC Accuracy	ORGANS Accuracy	Average Accuracy
LLP	–	0.4538	0.3009	0.7843	0.4336	0.8869	0.7635	0.7898	0.5372	0.6187
LLP + Ours	uniform	0.5256	0.3742	0.7861	0.4347	0.9017	0.7971	0.8099	0.6197	**0.6561**
LLP + Ours	gauss	0.5162	0.3710	0.7853	0.4307	0.8980	0.7867	0.7937	0.6283	0.6512
LLP + Ours	half	0.5060	0.3774	0.7819	0.4400	0.8912	0.8050	0.8015	0.6164	0.6524

Table 1. Average accuracy on eight datasets when changing a sampling method for γ. Note that this table corresponds to Table 3 in the main paper, which shows the average accuracy of all datasets. All sampling methods outperformed the performance compared to the baseline method (LLP); 'uniform' was marginally better than other methods.

Method	CIFAR10 Accuracy	SVHN Accuracy	PATH Accuracy	OCT Accuracy	BLOOD Accuracy	ORGANA Accuracy	ORGANC Accuracy	ORGANS Accuracy	Average Accuracy
LLP	0.4538	0.3009	0.7843	0.4336	0.8869	0.7635	0.7898	0.5372	0.6187
LLP + Ours(99%)	0.5256	0.3742	0.7861	0.4347	0.9017	0.7971	0.8099	0.6197	**0.6561**
LLP + Ours(95%)	0.5267	0.3807	0.7913	0.4227	0.8852	0.8034	0.8012	0.6046	0.6519
LLP + Ours(80%)	0.5069	0.4015	0.8004	0.4004	0.8947	0.8104	0.7970	0.6142	0.6531
LLP + Ours(50%)	0.4781	0.3493	0.7611	0.4019	0.8676	0.7794	0.8008	0.5562	0.6243

Table 2. Accuracy on each dataset when changing the degree of confidence interval (50%, 80%, 95%, 99%). Note that this table corresponds to Table 4 in the main paper, which shows the average accuracy of all datasets. LLP + Ours(99%) was superior to the baseline LLP in all datasets.

Method	Size	Num	CIFAR10 Accuracy	SVHN Accuracy	PATH Accuracy	OCT Accuracy	BLOOD Accuracy	ORGANA Accuracy	ORGANC Accuracy	ORGANS Accuracy	Average Accuracy
LLP	10	512	0.4538	0.3009	0.7843	0.4336	0.8869	0.7635	0.7898	0.5372	0.6187
LLP + Ours(w/o CI)	10	512	0.4582	0.2971	0.7884	0.4250	0.8898	0.7831	0.8189	0.563	0.6280
LLP + Ours	10	512	0.5256	0.3742	0.7861	0.4347	0.9017	0.7971	0.8099	0.6197	**0.6561**
LLP	20	256	0.3301	0.2217	0.7339	0.3778	0.8527	0.6936	0.7034	0.4396	0.5441
LLP + Ours(w/o CI)	20	256	0.3286	0.2217	0.7223	0.4080	0.8650	0.7064	0.7270	0.4182	0.5496
LLP + Ours	20	256	0.3705	0.2305	0.7521	0.3876	0.8687	0.7312	0.7435	0.4499	**0.5667**
LLP	40	128	0.2727	0.2042	0.6870	0.3790	0.8061	0.5669	0.5422	0.3386	0.4746
LLP + Ours(w/o CI)	40	128	0.2579	0.2141	0.6968	0.351	0.7995	0.5357	0.5059	0.3191	0.4600
LLP + Ours	40	128	0.3162	0.2034	0.7002	0.3836	0.8167	0.6249	0.6168	0.3847	**0.5058**

Table 3. Average accuracy on eight datasets when changing the number of labeled bags and bag size. 'Size' means the bag size. 'Num' means the number of labeled bags. Note that this table corresponds to Table 5 in the main paper, which shows the average accuracy of all datasets. LLP + Ours was superior to the baseline LLP in all datasets under every bag size and the number of labeled bags.

Method	Bag-Generation	CI	CIFAR10 Accuracy	SVHN Accuracy	PATH Accuracy	OCT Accuracy	BLOOD Accuracy	ORGANA Accuracy	ORGANC Accuracy	ORGANS Accuracy	Average Accuracy
LLP	–	–	0.4538	0.3009	0.7843	0.4336	0.8869	0.7635	0.7898	0.5372	0.6187
LLP + Ours(w/o CI)	Union	–	0.4575	0.3252	0.7710	0.4014	0.8882	0.7884	0.7887	0.5600	0.6226
LLP + Ours(w/o CI)	Sub-bag		0.3155	0.2331	0.6221	0.4182	0.7587	0.6035	0.5836	0.4305	0.4956
LLP + Ours(with CI)	Sub-bag	✓	0.5157	0.3785	0.7853	0.4294	0.8882	0.7906	0.8109	0.6033	0.6364
LLP + Ours(w/o CI)	MixBag		0.4582	0.2971	0.7884	0.4250	0.8898	0.7831	0.8189	0.5630	0.6280
LLP + Ours(with CI)	MixBag	✓	0.5256	0.3742	0.7861	0.4347	0.9017	0.7971	0.8099	0.6197	**0.6561**

Table 4. Accuracy on each dataset in different bag generation methods. Note that this table corresponds to Table 6 in the main paper, which shows the average accuracy of all datasets. LLP + Ours(MixBag with CI) was superior to the baseline LLP in almost all datasets under every condition.