In this document, we compare `LibMTL` and the recent popular implementations (i.e., CAGrad[1], Nash-MTL[2], and Aligned-MTL[3]). There are something required special attention as follows.

- We comment out the code about reproducibility (`https://github.com/median-research-group/LibMTL/blob/main/LibMTL/utils.py#L18-L20`) for faster running speed;
- Each experiment is repeated over three random seeds and the average value is reported.

Table 1: Hyperparameters Configuration.

	Configuration
common	`GPU`: NVIDIA GeForce RTX 3090 `multi_input`: False; `aug`: True `train_bs`: 2; `test_bs`: 2; epochs: 200 `optim`: Adam; `lr`: 0.0001; `weight_decay`: 0.0 `scheduler`: step; `step_size`: 100; `gamma`: 0.5
GradNorm	`rep_grad`: False; `alpha`: 1.5
MGDA	`rep_grad`: False; `mgda_gn`: none
DWA	`T`: 2
GradDrop	`leak`: 0.0
IMTL	`rep_grad`: True
GradVac	`GradVac_beta`: 0.5 `GradVac_group_type`: 0
CAGrad	`calpha`: 0.5; `rescale`: 1
Nash-MTL	`update_weights_every`: 1 `optim_niter`: 20; `max_norm`: 1.0
MoCo	`MoCo_beta`: 0.99; `MoCo_beta_sigma`: 0 `MoCo_gamma`: 0.1; `MoCo_gamma_sigma`: 0 `MoCo_rho`: 0

[1] `https://github.com/Cranial-XIX/CAGrad`
[2] `https://github.com/AvivNavon/nash-mtl`
[3] `https://github.com/SamsungLabs/MTL`

Table 2: Performance on the *NYUv2* dataset with 3 tasks on **SegNet+MTAN** architecture.

		Segmentation		**Depth**		**Normal**				
						Angle Distance		**Within t°**		
		mIoU↑	**PAcc↑**	**AErr↓**	**RErr↓**	**Mean↓**	**MED↓**	**11.25↑**	**22.5↑**	**30↑**
EW	[6, 9, 11]	39.29	65.33	0.5493	0.2263	28.15	23.96	22.09	47.50	61.08
	`LibMTL`	40.89	66.14	0.5524	0.2347	27.27	22.41	24.38	50.18	63.36
DWA [8]	[9, 11]	39.11	65.31	0.5510	0.2285	27.61	23.18	24.17	50.18	62.39
	`LibMTL`	40.50	65.65	0.5358	0.2222	27.58	22.93	23.30	49.16	62.57
UW [4]	[9, 11]	36.87	63.17	0.5446	0.2260	27.04	22.61	23.54	49.05	63.65
	`LibMTL`	39.34	64.88	0.5294	0.2242	26.47	21.30	25.86	52.40	65.47
MGDA [10]	[6, 9, 11]	30.47	59.90	0.6070	0.2555	24.88	19.45	29.18	56.88	69.36
	`LibMTL`	29.91	60.06	0.5901	0.2432	24.55	18.63	30.49	58.02	70.14
GradNorm [1]	[11]	20.09	52.06	0.72	0.28	24.83	18.86	30.81	57.94	69.73
	`LibMTL`	40.12	65.65	0.5213	0.2180	25.50	19.84	28.46	55.39	67.85
PCGrad [13]	[6, 9, 11]	38.06	64.64	0.5550	0.2325	27.41	22.80	23.86	49.83	63.14
	`LibMTL`	40.61	65.89	0.5416	0.2287	26.97	22.05	24.68	50.90	64.05
GradVac [12]	[11]	37.53	64.35	0.56	0.24	27.66	23.38	22.83	48.66	62.21
	`LibMTL`	40.90	65.50	0.5766	0.2438	27.26	22.39	24.55	50.22	63.34
IMTL [7]	[11]	39.35	65.60	0.54	0.23	26.02	21.19	26.20	53.13	66.24
	`LibMTL`	41.19	66.37	0.5323	0.2237	26.06	20.77	26.76	53.48	66.32
GradDrop [2]	[6, 9, 11]	39.39	65.12	0.5455	0.2279	27.48	22.96	23.38	49.44	62.87
	`LibMTL`	40.00	65.61	0.5886	0.2517	28.05	23.54	22.81	48.01	61.33
CAGrad [6]	[6, 9, 11]	39.79	65.49	0.5486	0.2250	26.31	21.58	25.61	52.36	65.58
	`LibMTL`	41.27	66.70	0.5409	0.2356	25.35	19.81	28.44	55.47	68.05
Nash-MTL [9]	[9, 11]	40.13	65.93	0.5261	0.2171	25.26	20.08	28.40	55.47	68.15
	`LibMTL`	40.66	66.25	0.5339	0.2266	25.11	19.59	28.70	55.97	68.52
RLW [5]	[9, 11]	37.17	63.77	0.5759	0.2410	28.27	24.18	22.26	47.05	60.62
	`LibMTL`	38.82	64.45	0.5718	0.2366	28.09	23.65	22.54	47.76	61.27
MoCo [3]	[3]	40.30	66.07	0.5575	0.2135	26.67	21.83	25.61	51.78	64.85
	`LibMTL`	40.72	66.33	0.5689	0.2455	27.11	22.40	24.11	50.19	63.61
Aligned-MTL [11]	[11]	40.82	66.33	0.53	0.22	25.19	19.71	28.88	56.23	68.54
	`LibMTL`	40.15	66.05	0.5520	0.2291	25.37	19.89	28.30	55.29	67.95

References

[1] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International Conference on Machine Learning*, 2018.

[2] Zhao Chen, Jiquan Ngiam, Yanping Huang, Thang Luong, Henrik Kretzschmar, Yuning Chai, and Dragomir Anguelov. Just pick a sign: Optimizing deep multitask models with gradient sign dropout. In *Neural Information Processing Systems*, 2020.

[3] Heshan Devaka Fernando, Han Shen, Miao Liu, Subhajit Chaudhury, Keerthiram Murugesan, and Tianyi Chen. Mitigating gradient bias in multi-objective learning: A provably convergent approach. In *International Conference on Learning Representations*, 2023.

[4] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[5] Baijiong Lin, Feiyang Ye, Yu Zhang, and Ivor Tsang. Reasonable effectiveness of random weighting: A litmus test for multi-task learning. *Transactions on Machine Learning Research*, 2022.

[6] Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In *Neural Information Processing Systems*, 2021.

[7] Liyang Liu, Yi Li, Zhanghui Kuang, Jing-Hao Xue, Yimin Chen, Wenming Yang, Qingmin Liao, and Wayne Zhang. Towards impartial multi-task learning. In *International Conference on Learning Representations*, 2021.

[8] Shikun Liu, Edward Johns, and Andrew J. Davison. End-to-end multi-task learning with attention. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[9] Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. In *International Conference on Machine Learning*, 2022.

[10] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In *Neural Information Processing Systems*, 2018.

[11] Dmitry Senushkin, Nikolay Patakin, Arseny Kuznetsov, and Anton Konushin. Independent component alignment for multi-task learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[12] Zirui Wang, Yulia Tsvetkov, Orhan Firat, and Yuan Cao. Gradient vaccine: Investigating and improving multi-task optimization in massively multilingual models. In *International Conference on Learning Representations*, 2021.

[13] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In *Neural Information Processing Systems*, 2020.