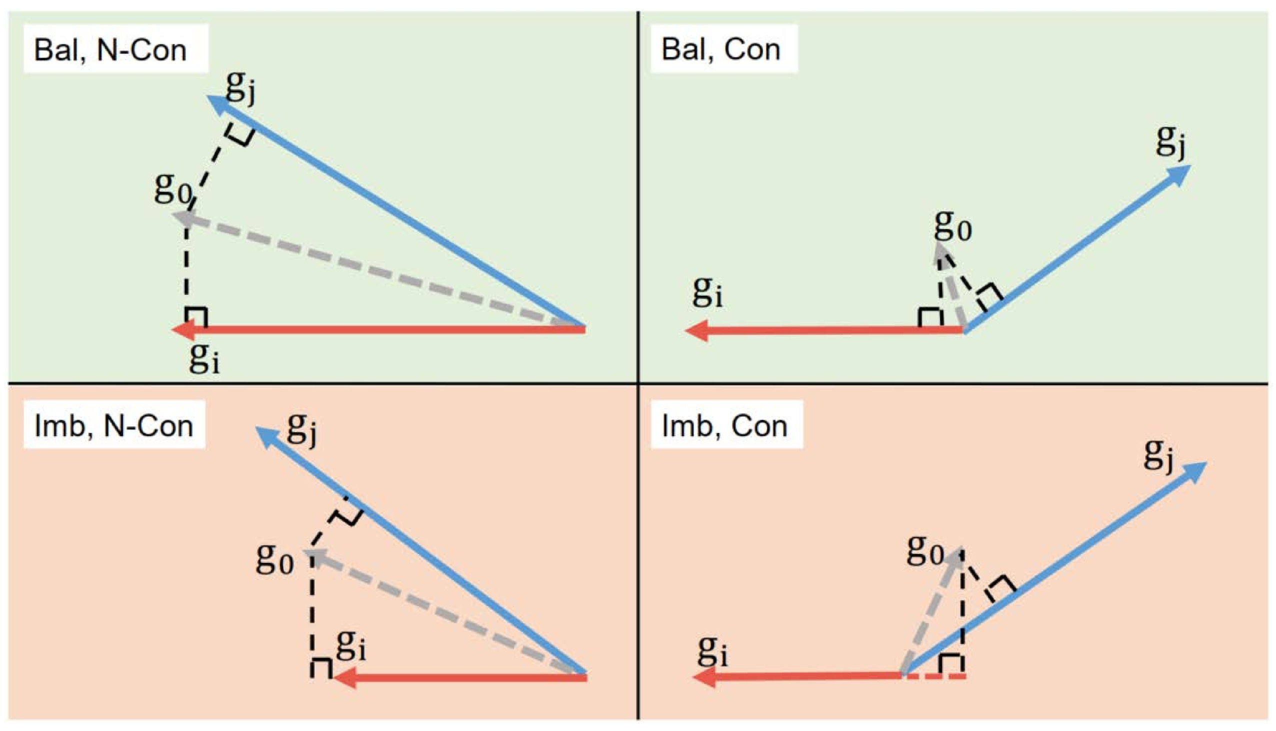
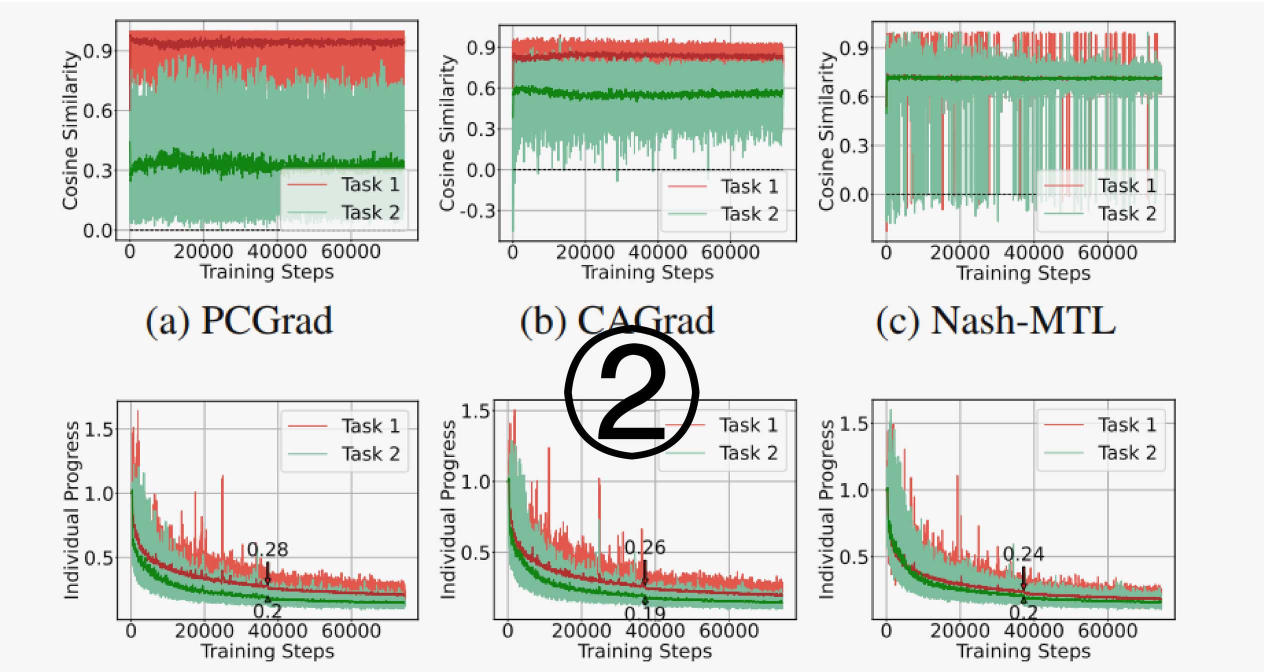
Injecting Imbalance Sensitivity for Multi-Task Learning

Zhipeng Zhou[1], Liu Liu[2], Peilin Zhao[2], Wei Gong[1]
[1]USTC, [2]Tencent AI Lab

IJCAI 2025 MONTREAL

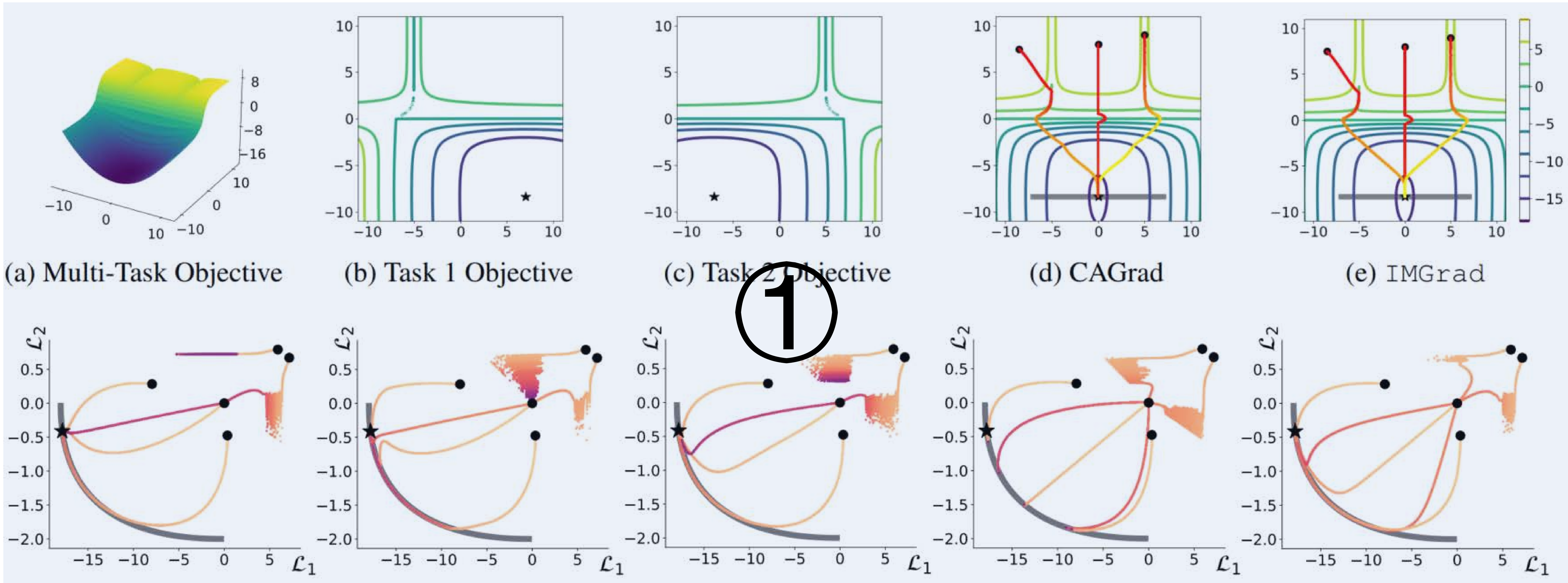
Motivation





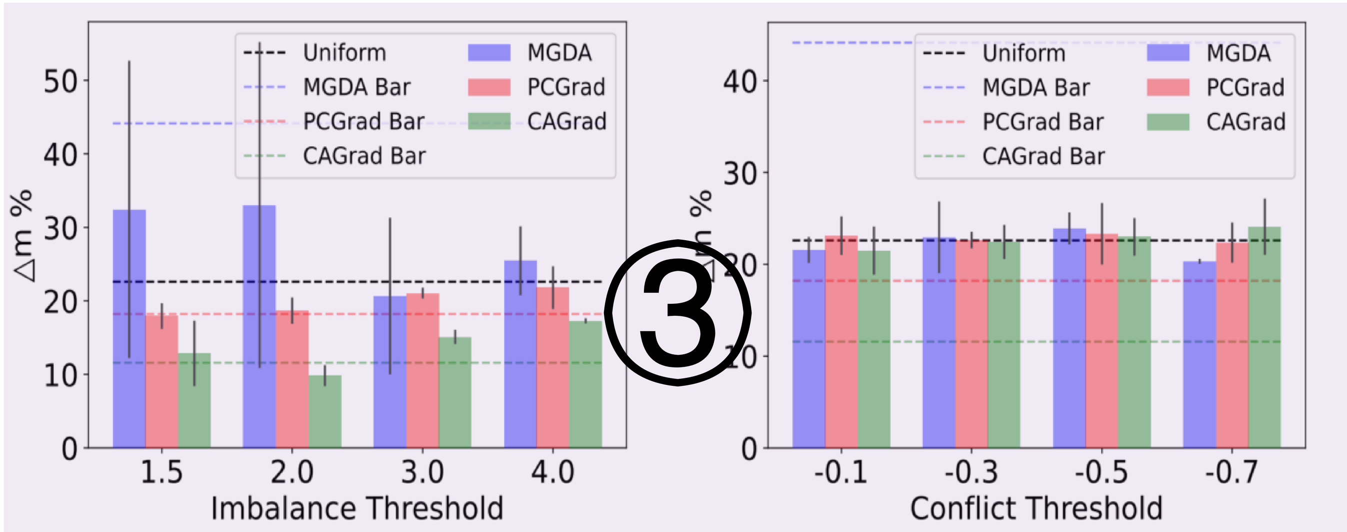
(a) PCGrad (b) CAGrad (c) Nash-MTL

(d) PCGrad ($\Delta m\% = 22.46$) (e) CAGrad ($\Delta m\% = 9.97$) (f) Nash-MTL ($\Delta m\% = 8.20$)



(a) Multi-Task Objective (b) Task 1 Objective (c) Task 2 Objective (d) CAGrad (e) IMGrad

(a) LS (b) PCGrad (c) CAGrad (d) Nash-MTL (e) IMGrad



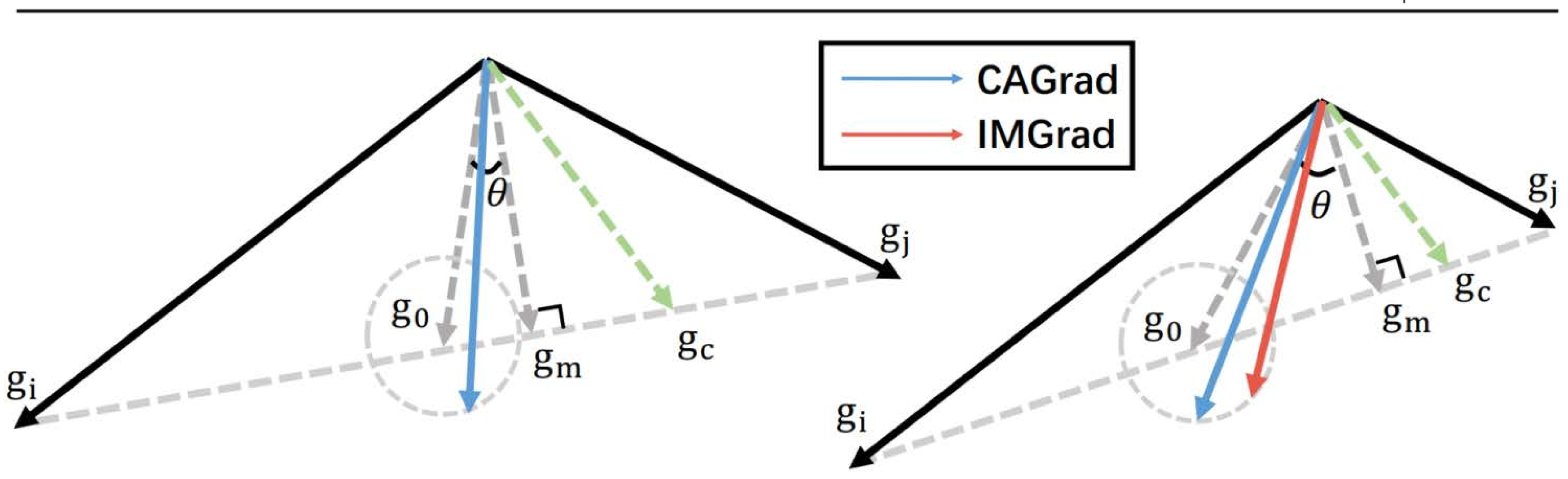
(a) Imbalance Sensitivity (b) Conflict Sensitivity

① **Convergence:** Conflict (✓) Imbalance (✗) vs Conflict (✗) Imbalance (✓).
② **Individual Progress:** Correlation with respective to conflict and imbalance issues.
③ **Benefits of Imbalance Sensitivity:** Adaptive imbalance ratio *vs.* adaptive conflict ratio.

Imbalance issue is of more crucial for multi-task optimization.

Approach

	GD	GradDrop	MGDA	PCGrad	IMTL	CAGrad	Nash-MTL	MoCo	IMGrad
Conflict-averse	✗	✗	✓	✓	✗	✓	✗	✓	✓
Imbalance-sensitive	✗	✗	✗	✗	✓	✗	✓	✗	✓



CAGrad: $\max_{d\in\mathbb{R}^m} \min_{\omega\in\mathcal{W}} g_\omega^\top d \quad s.t. \ \|d-g_0\| \le c\|g_0\|$

Maximize the ***projected norm*** of the combined gradient across all individuals

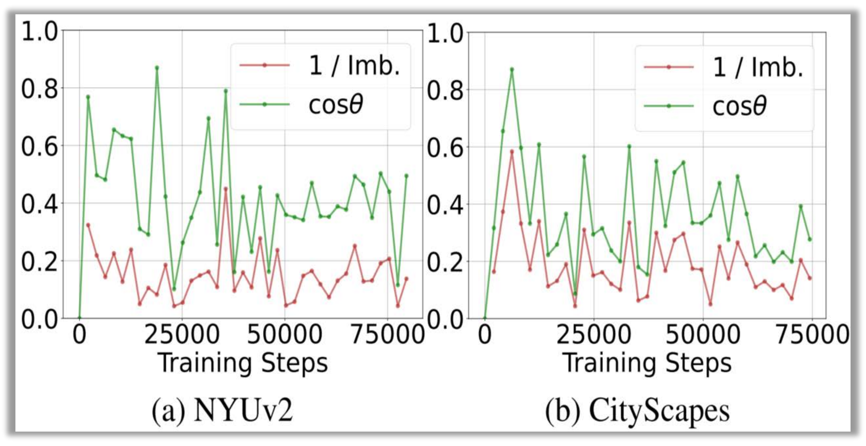
IMGrad: $\max_{d\in\mathbb{R}^m} \min_{\omega\in\mathcal{W}} g_\omega^\top d - \boldsymbol{\mu(g_\omega^\top d - \|g_\omega\|^2)} \quad s.t. \ \|d-g_0\| \le c\|g_0\|$

$$\min_{\omega\in\mathcal{W}} (1-\mu)\underbrace{(g_\omega^\top g_0 + \sqrt{\phi}\|g_\omega\|)}_{\text{CAGrad}} + \mu\underbrace{\|g_\omega\|^2}_{\text{MGDA}}$$

Convergence (*CAGrad*) and ***non-conflicting*** (*MGDA*) should be weighted with respective to imbalance

- We employ $\boldsymbol{\mu = \cos\theta}$ as the indicator of imbalance

$$\min_{\omega\in\mathcal{W}} (1-\cos\theta) g_\omega^\top g_0 + \cos\theta\sqrt{\phi}\|g_\omega\|)$$



(a) NYUv2 (b) CityScapes

Results

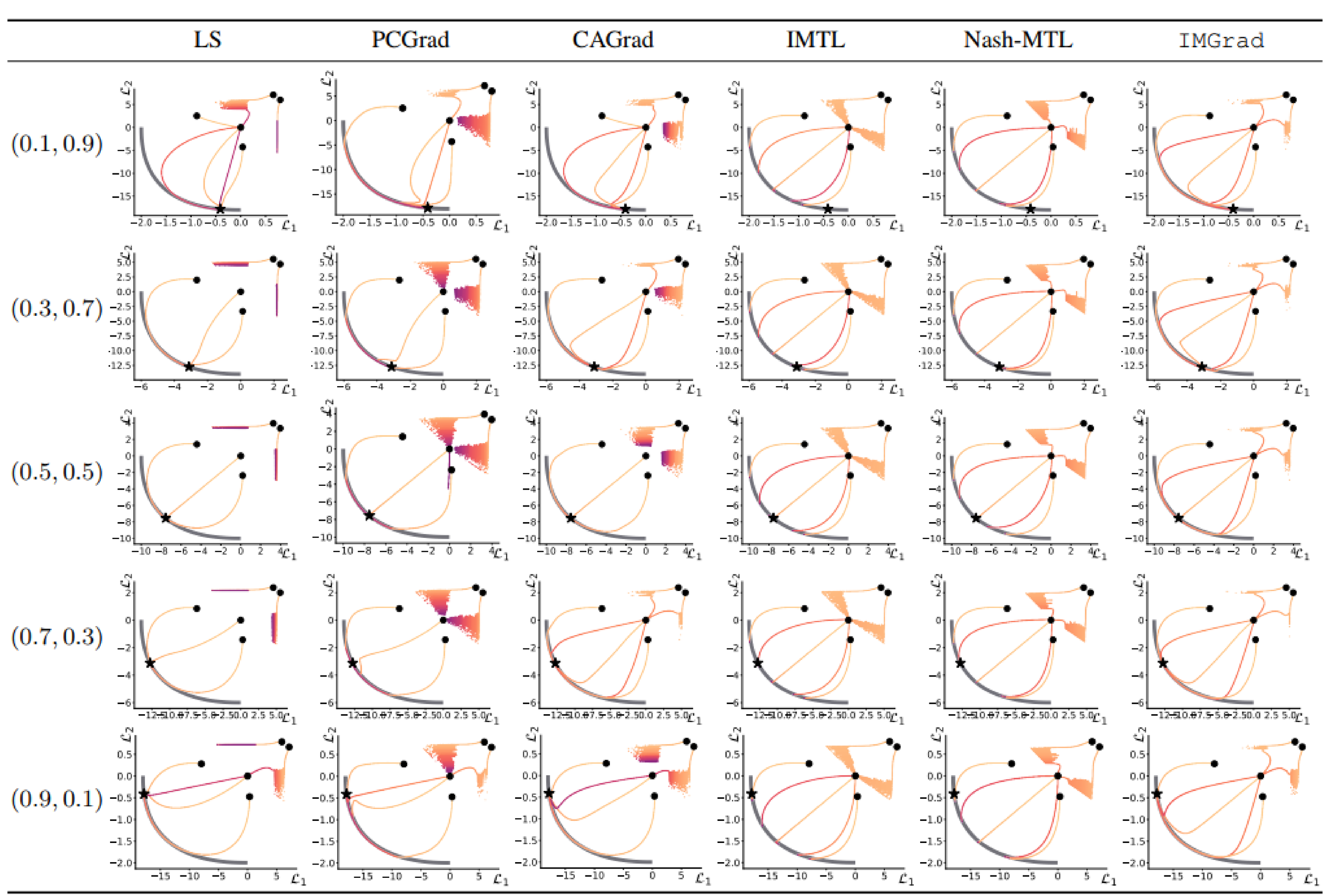
Scene Understanding: *CityScapes* (2-task)

Method	Segmentation (Higher Better)		Depth (Lower Better)		Δ m% ↓
	mIoU	Pix. Acc.	Abs. Err.	Rel. Err.	
Independent	74.01	93.16	0.0125	27.77	-
LS	75.18	93.49	0.0155	46.77	22.60
RLW	74.57	93.41	0.0158	47.79	24.37
DWA	75.24	93.52	0.0160	44.37	21.43
MGDA	68.84	91.54	0.0309	33.50	44.14
GradDrop	75.27	93.53	0.0157	47.54	23.67
PCGrad	75.13	93.48	0.0154	42.07	18.21
CAGrad	75.16	93.48	0.0141	37.60	11.58
IMTL	75.33	93.49	0.0135	38.41	11.04
Nash-MTL	75.41	93.66	0.0129	35.02	6.82
MoCo	75.42	93.55	0.0149	34.19	9.90
FAMO	74.54	93.29	0.0145	32.59	8.13
IMGrad	75.13	93.45	0.0128	34.95	6.61

Scene Understanding: *NYUv2* (3-task)

Method	Segmentation (Higher Better)		Depth (Lower Better)		Surface Normal: Angle Distance (Lower Better)		Surface Normal: Within t° (Higher Better)			Δ m% ↓
	mIoU	Pix. Acc.	Abs Err	Rel Err	Mean	Median	11.25	22.5	30	
Independent	38.30	63.76	0.68	0.28	25.01	19.21	30.14	57.20	69.15	-
LS	39.29	65.33	0.55	0.23	28.15	23.96	22.09	47.50	61.08	5.46
RLW	37.17	63.77	0.58	0.24	28.27	24.18	22.26	47.05	60.62	7.67
DWA	39.11	65.31	0.55	0.23	27.61	23.18	24.17	50.18	62.39	3.49
MGDA	30.47	59.90	0.61	0.26	24.88	19.45	29.18	56.88	69.36	1.47
GradDrop	39.39	65.12	0.55	0.23	27.48	22.96	23.38	49.44	62.87	3.61
PCGrad	38.06	64.64	0.56	0.23	27.41	22.80	23.86	49.83	63.14	3.83
CAGrad	39.79	65.49	0.55	0.23	26.31	21.58	25.61	52.36	65.58	0.29
IMTL	39.35	65.60	0.54	0.23	26.02	21.19	26.20	53.13	66.24	-0.59
Nash-MTL	40.13	65.93	0.53	0.22	25.26	20.08	28.40	55.47	68.15	-4.04
MoCo	40.30	66.07	0.56	0.21	26.67	21.83	25.61	51.78	64.85	0.16
FAMO	38.88	64.90	0.55	0.22	25.06	19.57	29.21	56.61	68.98	-4.10
IMGrad	40.20	66.19	0.52	0.22	25.15	19.94	28.69	55.80	68.44	-4.57

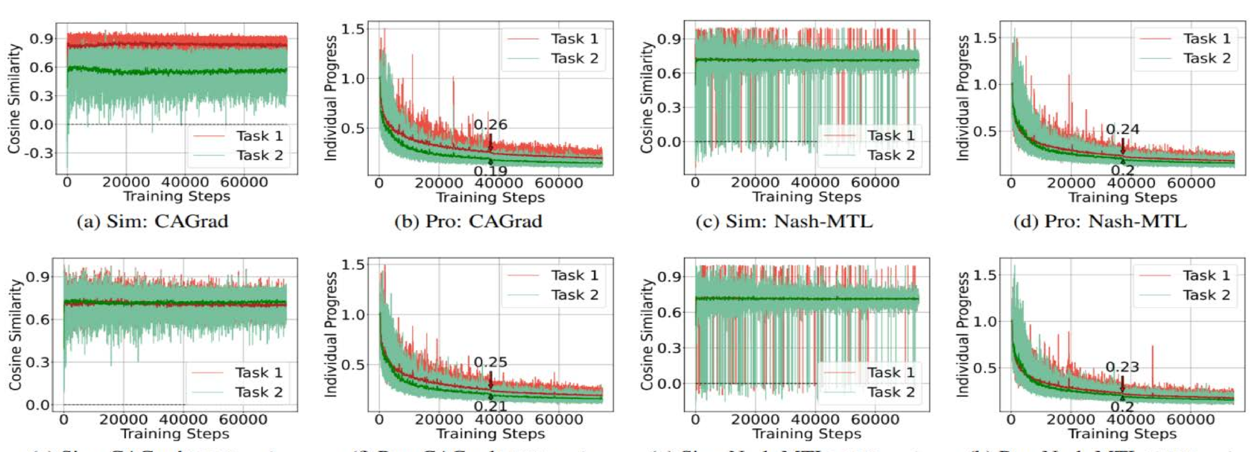
Convergence Visualization on Toy Examples



Many-Task Scenarios: MT10 and CelebA

MT10 Method	Success ± SEM ↑	CelebA Method	Δm% ↓
LS	0.49 ± 0.070	LS	4.15
STL SAC	0.90 ± 0.032	SI	7.20
MTL SAC	0.49 ± 0.073	RLW	1.46
MH SAC	0.54 ± 0.047	DWA	3.20
SM	0.73 ± 0.043	UW	3.23
CARE	0.84 ± 0.051	MGDA	14.85
PCGrad	0.72 ± 0.022	PCGrad	3.17
CAGrad	0.83 ± 0.045	CAGrad	2.48
Nash-MTL	0.91 ± 0.031	Nash-MTL	2.84
FAMO	0.83 ± 0.050	FAMO	1.21
IMGrad	0.93 ± 0.068 (+0.10)	IMGrad	1.27

Pareto Failure and Progress Examination



(a) Sim: CAGrad (b) Pro: CAGrad (c) Sim: Nash-MTL (d) Pro: Nash-MTL

(e) Sim: CAGrad+IMGrad (f) Pro: CAGrad+IMGrad (g) Sim: Nash-MTL+IMGrad (h) Pro: Nash-MTL+IMGrad

Contact Us!

Email: zzp1994@mail.ustc.edu.cn
Implementation: https://github.com/zzpustc/IMGrad