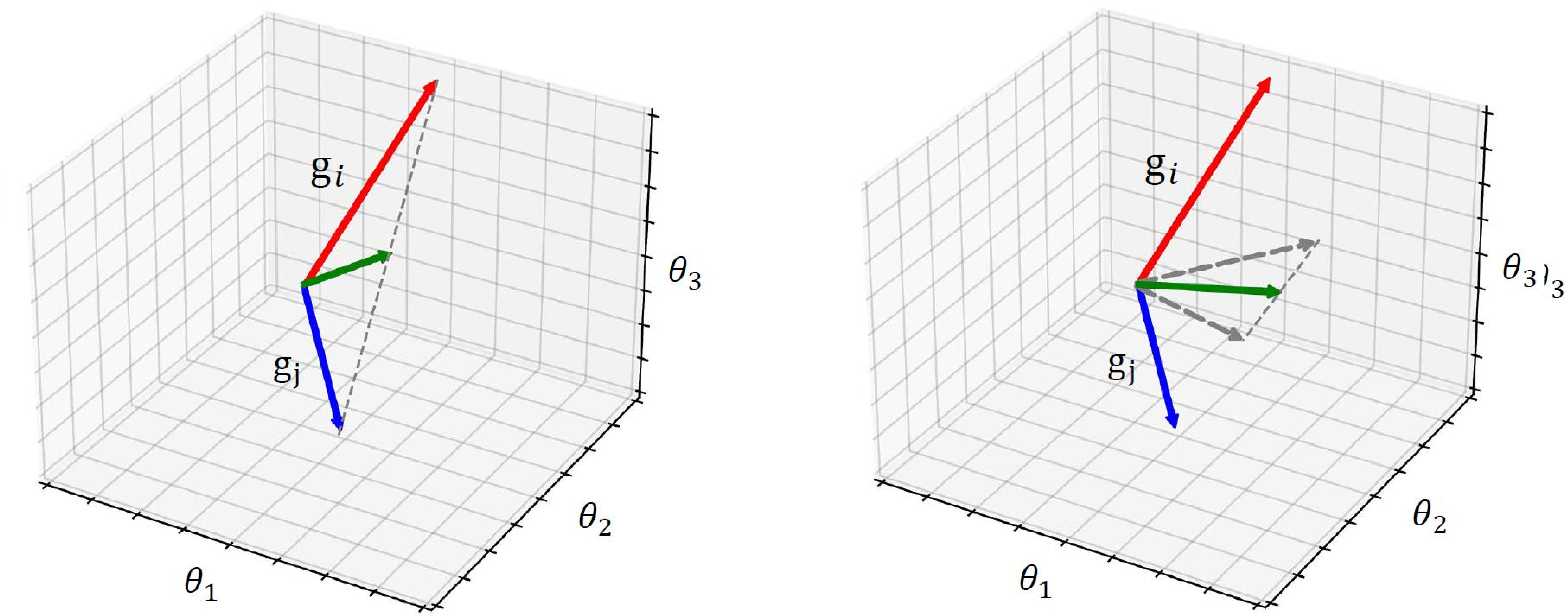
Recon: Reducing Conflicting Gradients From the Root For Multi-Task Learning

ICLR 2023

May 1, 2023

Main Challenges of Multi-Task Learning

Definition 1 (Conflicting Gradients). The gradients $\mathrm{g_i}$ of task $\mathcal{T}_i$ and g_j of task $\mathcal{T}_j$ are said to be conflicting with each other of $\cos\phi_{ij} < 0$, where ϕ_{ij} is the angle between $\mathrm{g_i}$ and g_j.



Gradient Descent　　　　PCGrad

Gradient Surgery Cannot Reduce Conflicting Gradients

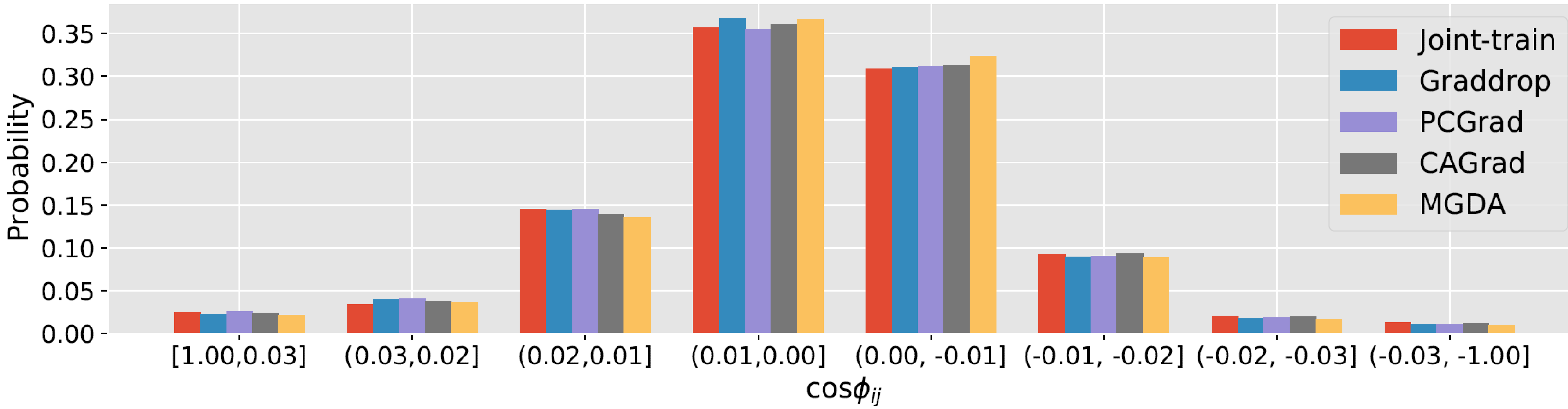
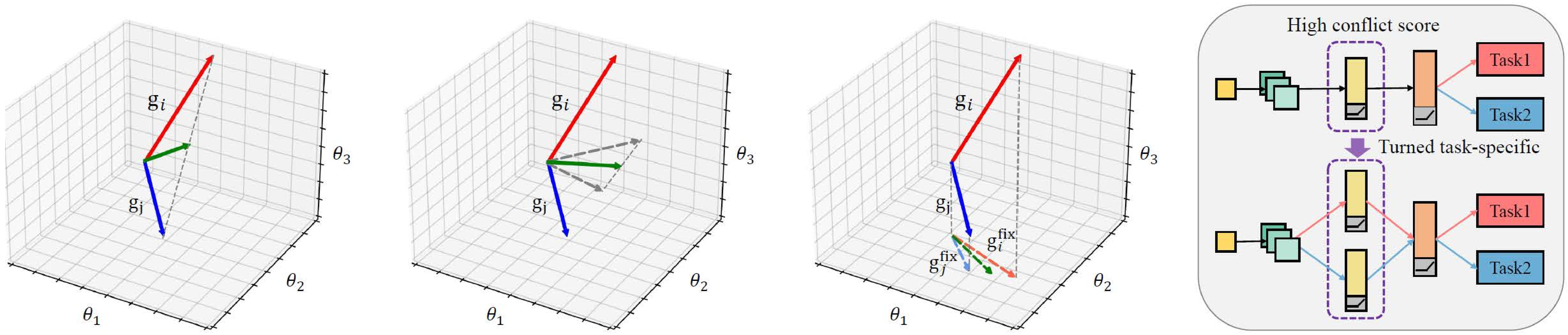


Figure 1: The distributions of gradient conflicts (in terms of $\cos\phi_{ij}$) of the joint-training baseline and state-of-the-art gradient manipulation methods on Multi-Fashion+MNIST benchmark.

Our Proposed Method: Recon



(a) Joint-train (b) PCGrad (c) Recon (d) Recon

Figure 2: (a) In joint-training, the update vector is dominated by g_i due to the conflicting gradients. (b) PCGrad use the average gradients of the projected gradients as the update vector to reduce the influence of conflicting gradients. (c) Our approach Recon finds the parameters contributing most (e.g., θ_3) to gradient conflicts and turn them into task specific ones. In effect, it performs an orthographic/coordinate projection of conflicting gradients to the space of the rest parameters (e.g., θ_1 and θ_2) such that the projected gradients g_i^{fix} and g_j^{fix}. (d) Illustration of Recon turning a shared layer with high conflict score to task-specific layers.

Our Proposed Method: Recon

Definition 2 (Layer-wise Conflicting Gradients).

The gradients $\mathrm{g}_{\mathrm{i}}^{(k)}$ and $\mathrm{g}_{j}^{(k)}$ $(i \neq j)$ are said to be conflicting with each other of $\cos \phi_{ij}^{(k)} < 0$, where $\mathrm{g}_{\mathrm{i}}^{(k)}$ and $\mathrm{g}_{\mathrm{j}}^{(k)}$ denote the gradients of tasks $\mathcal{T}_i$ and $\mathcal{T}_j$ w.r.t. the k^{th} shared layer θ_{sh}^{k} respectively.

Definition 3 (S-Conflict Scores).

For any $-1 < S \leq 0$, the S-conflict score for the k^{th} shared layer is the number of different pairs $(i,j)(i \neq j)$ s.t. $\cos \phi_{ij}^{(k)}$, denoted as $s^{(k)}$.

Our Proposed Method: Recon

Step 1: Calculating the S-Conflict Scores for each shared layers.

Step 2: Set layers with top conflict scores task-specific.

Step 3: Train the modified network from scratch with any gradient manipulation methods.

Theoretical Analysis

Theorem 1

Assume that $\mathcal{L}$ is differentiable and for any two different tasks $\mathcal{T}_i$ and $\mathcal{T}_j$, it satisfies

$$\cos \phi_{ij}^{(k)} ||g_i^{(k)}|| < ||g_j^{(k)}||, \forall k \in \mathbb{P},$$

where $\mathbb{P}$ is the set of indices of the layers turned task-specific, then for any sufficiently small learning rate $\alpha > 0$,

$$\mathcal{L}(\widehat{\theta_r}) < \mathcal{L}(\hat{\theta}) ,$$

where $\widehat{\theta_r}$ denotes the parameter after one-step gradient update applying Recon, $\hat{\theta}$ denotes one without applying Recon.

Experiments

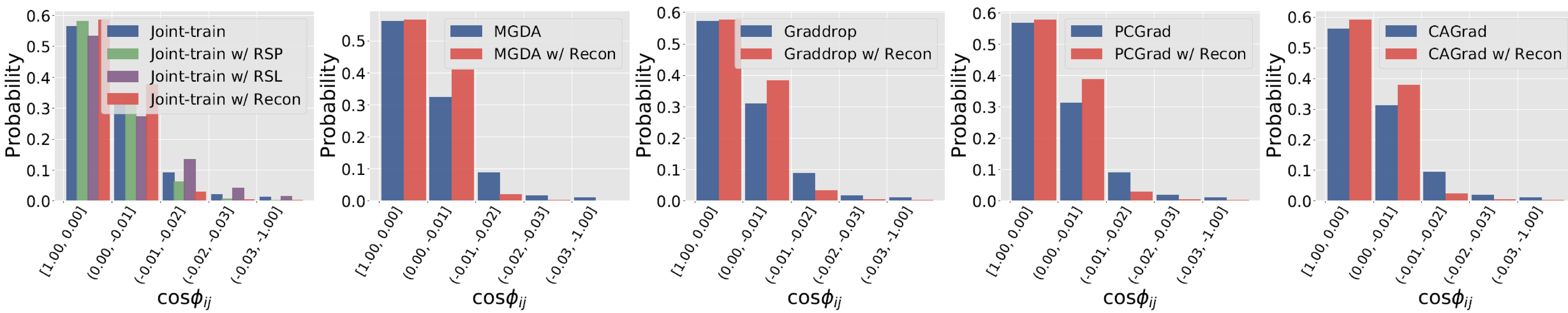


Figure 3: The distributions of gradient conflicts (in terms of $\cos\phi_{ij}$) of baselines and baselines with Recon on Multi-Fashion+MNIST benchmark.

Experiments

Table 1: Multi-task learning results on PASCAL-Context dataset with 4-task setting.

Method	SemSeg (Higher Better) mIoU	SemSeg (Higher Better) Pix Acc	PartSeg (Lower Better) mIoU	PartSeg (Lower Better) Pix Acc	saliency (Higher Better) mIoU	Surface Normal Angle Distance (Lower Better) Mean	Surface Normal Angle Distance (Lower Better) Median	Surface Normal Within $t°$ (Higher Better) 11.25	Surface Normal Within $t°$ (Higher Better) 22.5	$\Delta m\% \uparrow$	#P.
Single-task	65.00	90.53	59.59	92.61	65.61	14.55	12.36	46.51	81.29		30.09
Joint-train	64.06	90.45	57.91	92.17	62.71	16.40	14.23	39.38	75.93	-4.82	8.04
w/ Recon	64.73	**90.50**	59.00	92.44	66.17	14.99	12.68	44.82	80.11	-0.66	10.20
MGDA	46.05	86.62	54.82	91.39	64.76	15.77	13.54	41.98	77.82	-7.67	8.04
w/ Recon	55.82	87.73	56.31	91.67	64.91	15.12	12.88	44.36	79.81	-4.14	10.20
PCGrad	63.91	90.45	58.01	92.19	63.09	16.34	14.19	39.62	76.06	-4.59	8.04
w/ Recon	**65.02**	90.45	59.22	92.46	66.14	14.95	**12.73**	44.96	80.22	-0.55	10.20
Graddrop	64.14	90.34	57.62	92.12	62.64	16.46	14.28	39.29	75.71	-5.00	8.04
w/ Recon	64.48	90.45	59.08	92.46	**66.23**	14.94	12.72	45.03	80.25	-0.63	10.20
CAGrad	63.37	90.17	57.49	92.07	64.16	16.30	14.12	39.80	76.23	-4.37	8.04
w/ Recon	64.60	90.40	59.27	92.47	65.67	14.92	12.71	45.10	80.33	-0.76	10.20
BMTAS	64.89	90.44	58.87	92.36	63.42	15.66	13.44	42.29	78.14	-2.89	15.18
w/ Recon	64.78	90.46	**59.96**	**92.58**	65.96	**14.74**	**12.57**	**45.62**	**80.84**	**-0.19**	16.83