Split to Learn: Gradient Split for Multi-Task Human Image Analysis

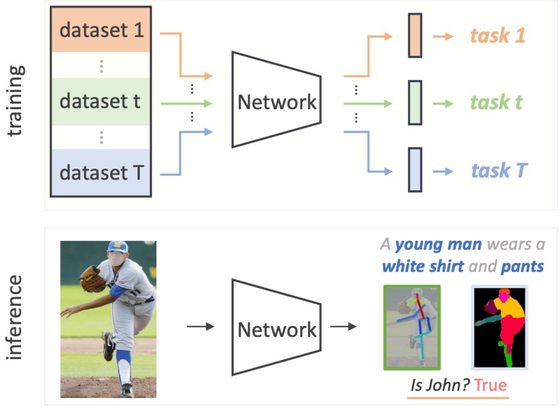
Weijian Deng[1] Yumin Suh[2] Xiang Yu[2] Masoud Faraki[2] Liang Zheng[1] Manmohan Chandraker[2,3]

[1]Australian National University [2]NEC Labs America [3]University of California, San Diego

WACV WAIKOLOA HAWAII JAN 3-7 • 2023

Multi-task Human Image Analysis

- **Multi-task network** provides a **rich explanation** of person-body images, including attributes, pose, part masks, and identity



Practical setting: the multi-task networks are trained across datasets and each dataset does not necessarily have exhaustive annotations for all tasks

Task Conflict

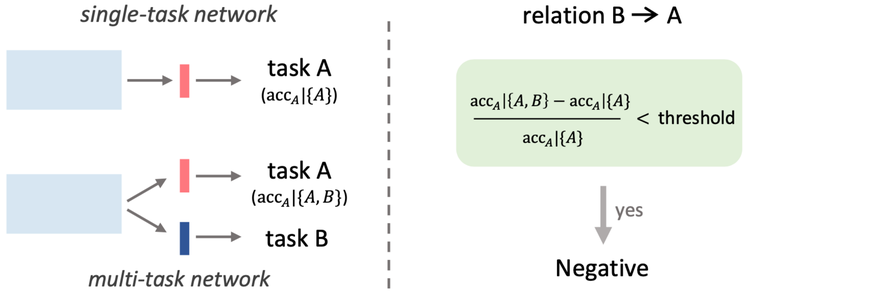
Multi-task learning can encounter **task conflicts**, *e.g.*, when jointly training identity-variant (body attributes) and identity-invariant (body pose) tasks

- **Our goal** is to train a unified model that solves multiple human-related tasks while avoiding the task conflict

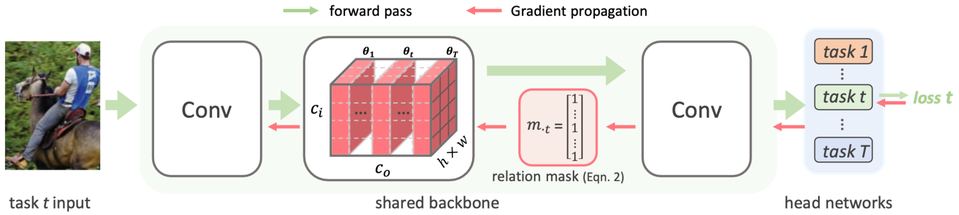
Gradient Split

- **Inter-task relation definition**

When training along with another task B, if **relative accuracy change** of task A is smaller than a threshold (*e.g.*, -0.01), then task B has a negative impact on task A



- **Inter-task Relationship based Gradient Update**



$$\theta_t \leftarrow \theta_t - \alpha \nabla_{\theta_t}^{GS} L, \text{ where } \nabla_{\theta_t}^{GS} L = \sum_{t'} \mathbf{m}_{tt'} \nabla_{\theta_t} L_{t'} \qquad \mathbf{m}_{tt'} = \begin{cases} 0 & \text{if } t \neq t' \text{and relation } t' \to t \text{ is negative} \\ 1 & \text{otherwise.} \end{cases}$$

We divide parameters of shared backbone into T groups for T tasks. GradSplit updates parameter θ_t using the gradients from only a subset of tasks $\{t'\}$, where the relationship task $t' \to t$ is not negative, while discarding gradients from the other tasks.

Four-Task Analysis

Methods	Backbone	ReID mAP (↑)	Attribute MA (↑)	Pose Mean (↑)	Parsing mIoU (↑)	Δ_m (↑)	#Param (M) ↓	#FLOPs (G) ↓
Single-task Networks (Upperbound)	ResNet-50-GN	81.1	78.0	88.2	45.6	+0.0	123	41
	ResNet-50-BN	83.0	78.3	88.4	45.4	–	123	41
Single-task Networks (Baseline)	ResNet-18-GN	74.9	76.9	87.0	42.4	–	63	24
	ResNet-18-BN	74.2	74.2	87.4	41.9	–	63	24
RCM [15]	ResNet-50-GN	54.9	68.1	69.0	36.1	-21.9	141	80
SFG [2]		64.4	73.9	71.8	34.8	-17.0	52	20
GradNorm [4]		56.1	77.7	68.4	28.5	-23.1	52	18
MTAN [21]		42.7	77.4	86.0	41.9	-14.7	75	40
ASTMT [26]	ResNet-50-TBN*	50.6	78.9	87.0	43.6	-10.6	82	42
Multi-head Baseline	ResNet-50-BN	63.2	76.3	78.9	39.8	-11.9	52	18
	ResNet-50-TBN*	78.1	77.2	86.8	41.8	-3.7	52	41
	ResNet-50-GN	79.3	76.4	86.1	42.7	-3.3	52	18
GradSplit (Ours)	ResNet-50-GN	80.1	77.8	86.4	43.9	**-1.8**	52	18

GradSplit achieves a better **accuracy-efficiency trade-off:** It minimizes accuracy drop caused by task conflicts while significantly saving compute resources in terms of both FLOPs and memory at inference

Cross-Dataset Analysis

Methods	Backbone	Attribute (MA)
Single-task network	ResNet-50-BN	71.5
	ResNet-50-GN	73.0
MTAN [21]	ResNet-50-GN	75.5
GradNorm [4]	ResNet-50-GN	75.5
ASTMT [26]	ResNet-50-TBN	76.5
Multi-head baseline	ResNet-50-GN	74.6
	ResNet-50-TBN	73.8
GradSplit	ResNet-50-GN	**77.5**

- Evaluate Attribute accuracy on another dataset where models are not trained for Attribute task

GradSplit achieves higher cross-dataset accuracy compared to single-task and other multi-task networks