IEEE WCCI (IJCNN) 2024 YOKOHAMA, JAPAN

June 30 - July 5, 2024

Arjun Roy, Christos Koutlis, Symeon Papadopoulos, Eirini Ntoutsi

FairBranch: Mitigating Bias Transfer in Fair Multi-task Learning

MAMMOth

EU HORIZON-RIA Project ID:101070285

Funded by the European Union

Outline

- ❖ Introduction and Motivation
- ❖ Problem Definition
- ❖ FairBranch
- ❖ Experiments
- ❖ Discussion and Conclusion

Introduction and Motivation

Single vs Multi-task Learning

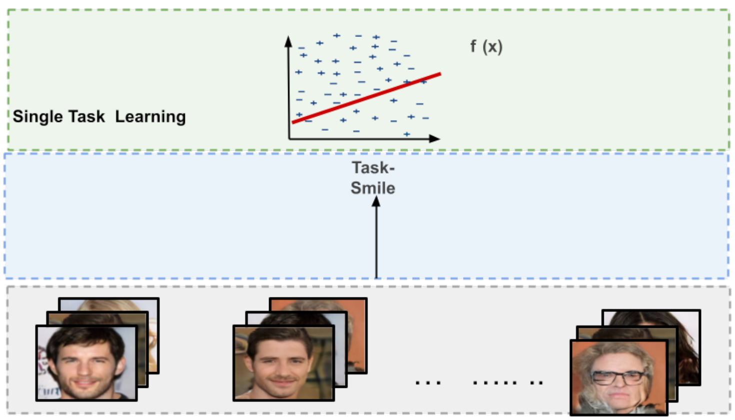
STL

MTL

Single vs Multi-task Learning

STL

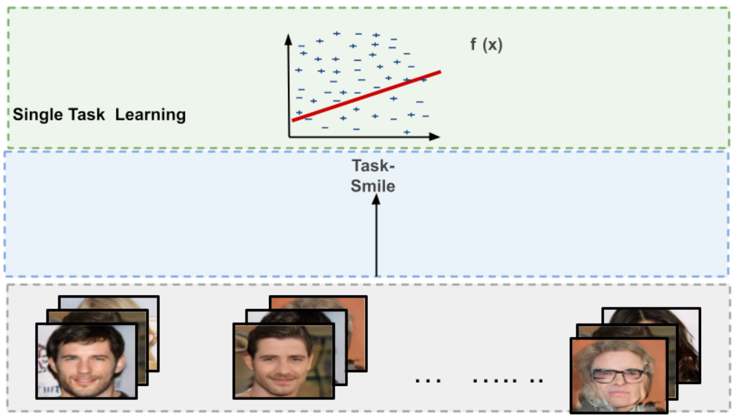
MTL



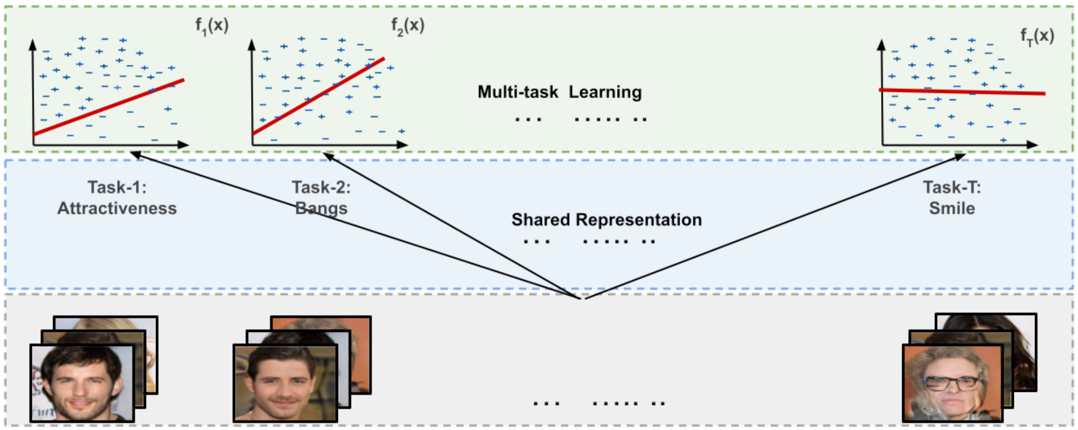
- learn a single supervised prediction tasks (STL).

Single vs Multi-task Learning

STL



MTL



- learn a single supervised prediction tasks (STL).

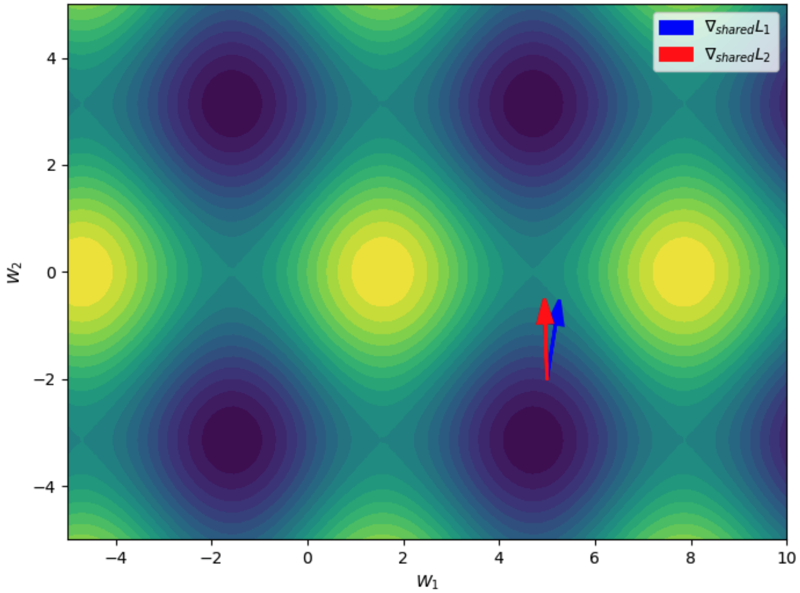
- Learn multiple supervised prediction tasks concurrently (MTL).
- Utilize a shared optimization space to enhance generalization across the tasks.

The Conflicting Gradient Problem

Hypothetical loss surface of the shared parameter space jointly trained with two task losses L_1 and L_2

The Conflicting Gradient Problem

Hypothetical loss surface of the shared parameter space jointly trained with two task losses L_1 and L_2



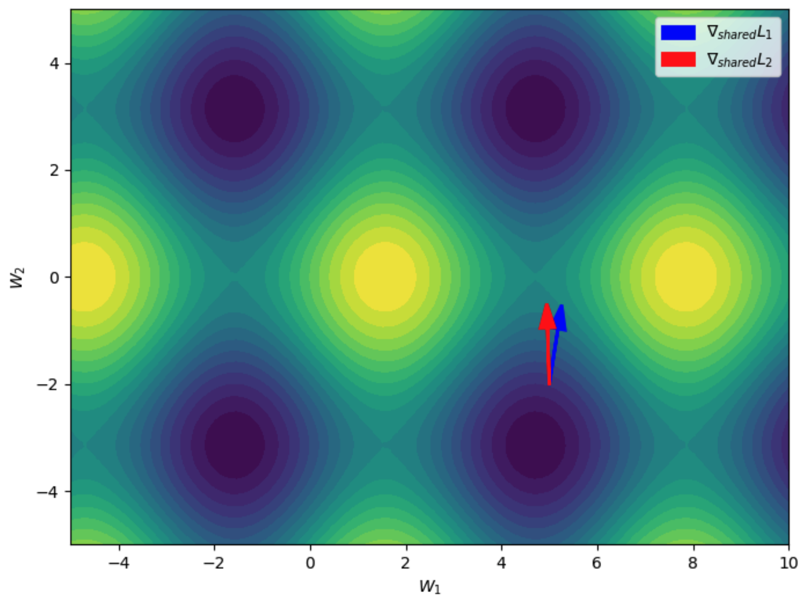
Two task t1 (blue arrow), and t2 (red arrow) moving together:

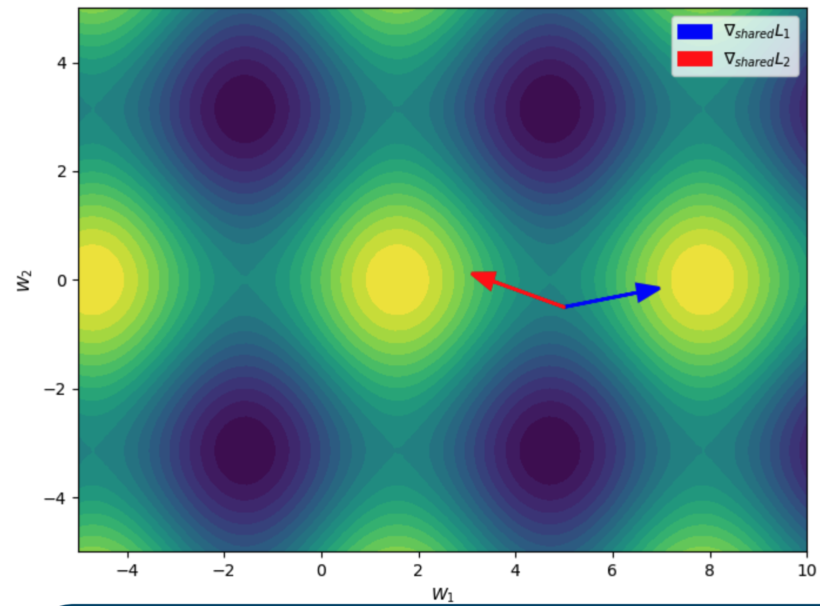
- in the same optimization direction

$$\nabla_{shared} L_1 \cdot \nabla_{shared} L_2 \geq 0$$

The Conflicting Gradient Problem

Hypothetical loss surface of the shared parameter space jointly trained with two task losses L_1 and L_2





Two task t1 (blue arrow), and t2 (red arrow) moving together:

- in the same optimization direction

$$\nabla_{shared}\, L_1 \cdot \nabla_{shared}\, L_2 \geq 0$$

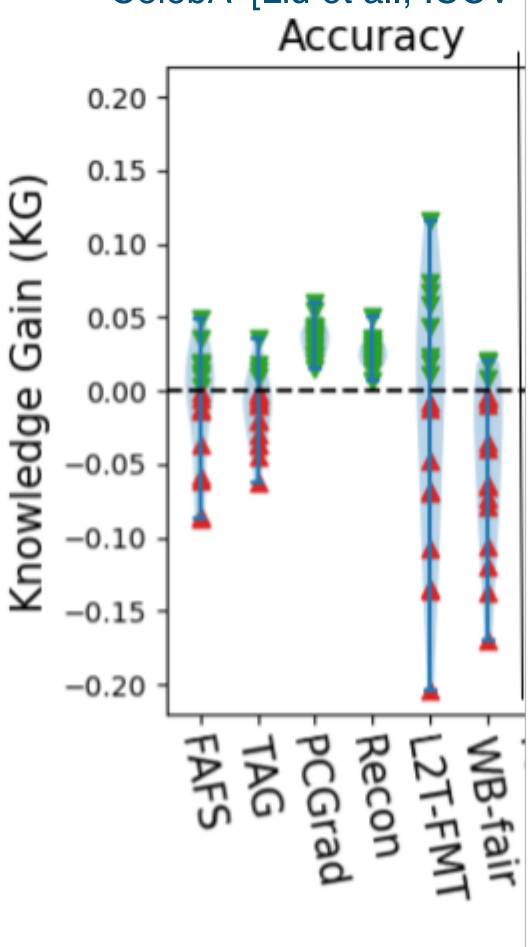
Two task t1 (blue arrow), and t2 (red arrow) moving towards:

- respective local minima in conflicting direction

$$\nabla_{shared}\, L_1 \cdot \nabla_{shared}\, L_2 < 0$$

Gradient Conflict to Negative Transfer

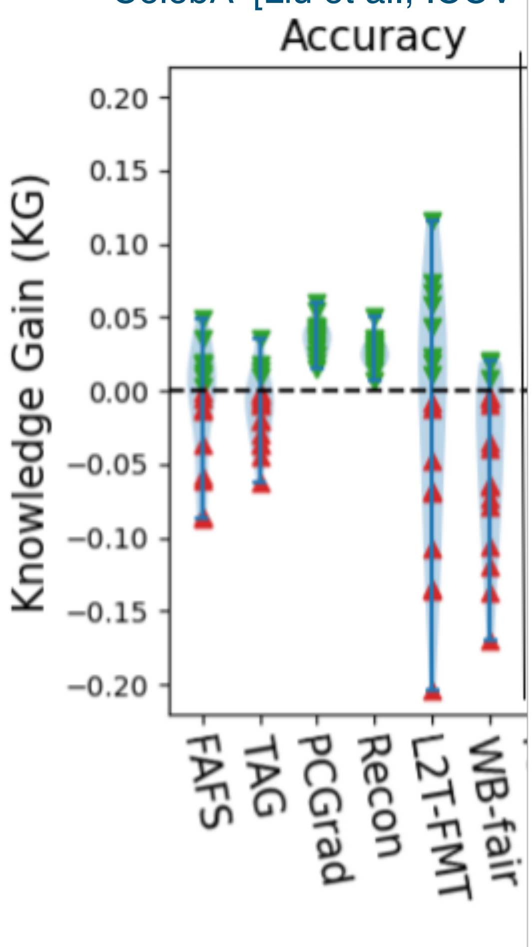
CelebA [Liu et al., ICCV 15]



FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]

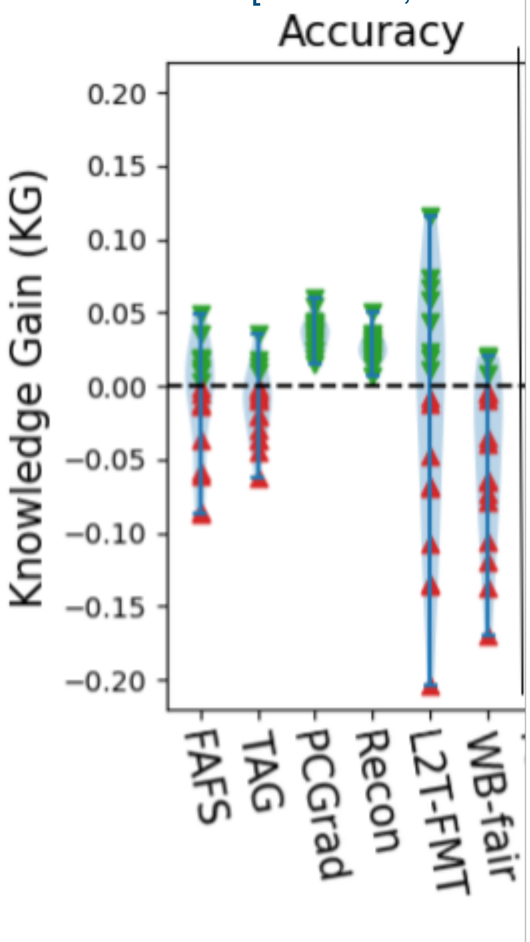


$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



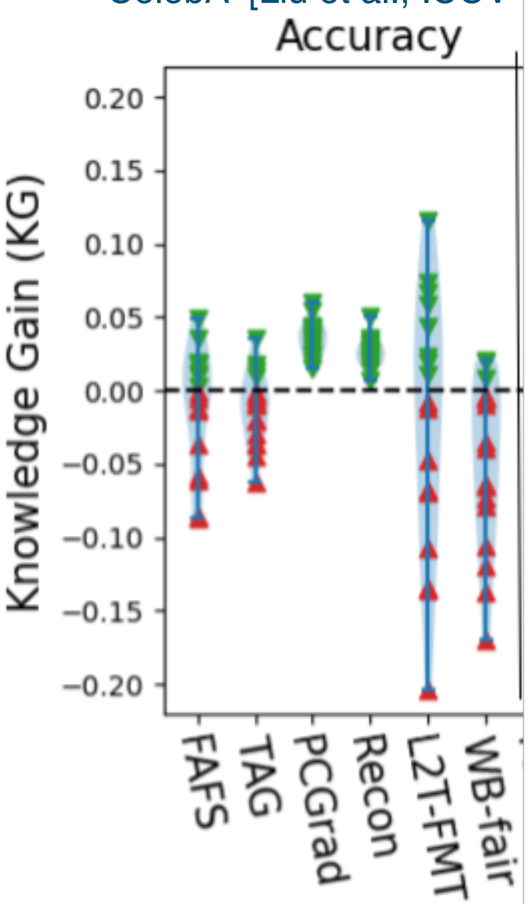
Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



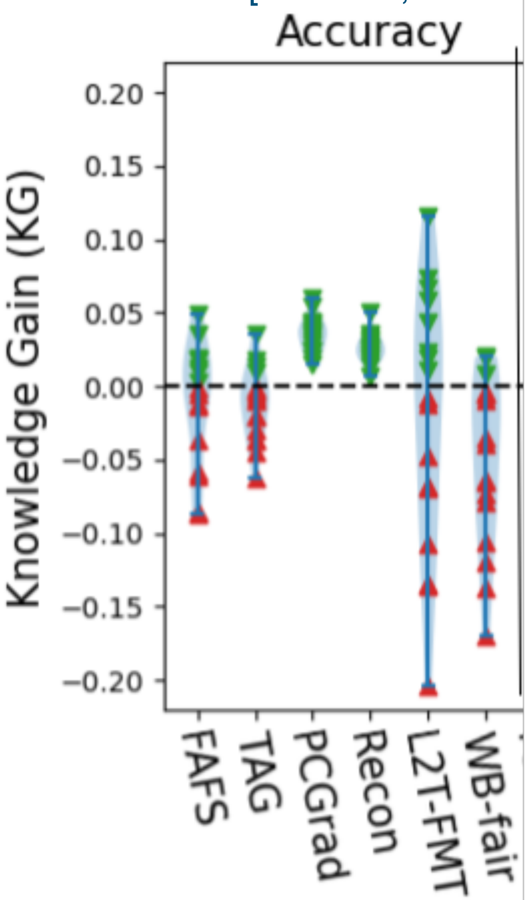
Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



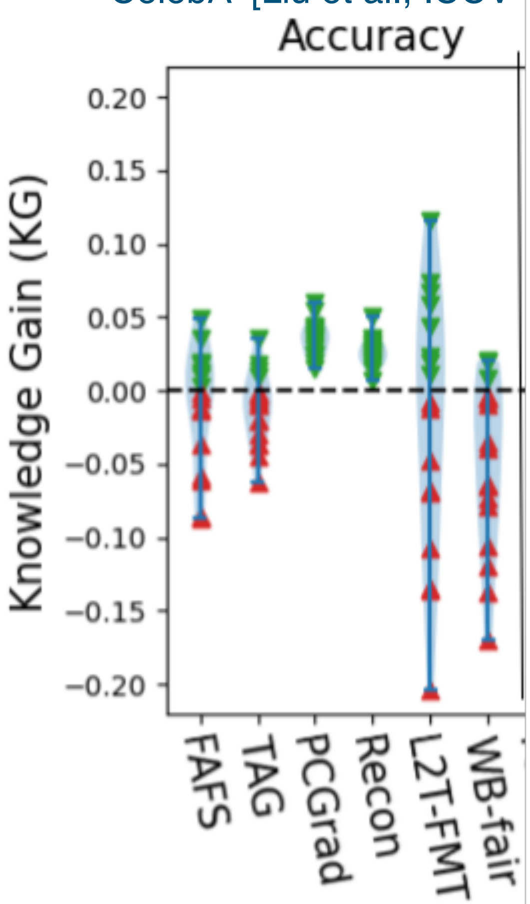
Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

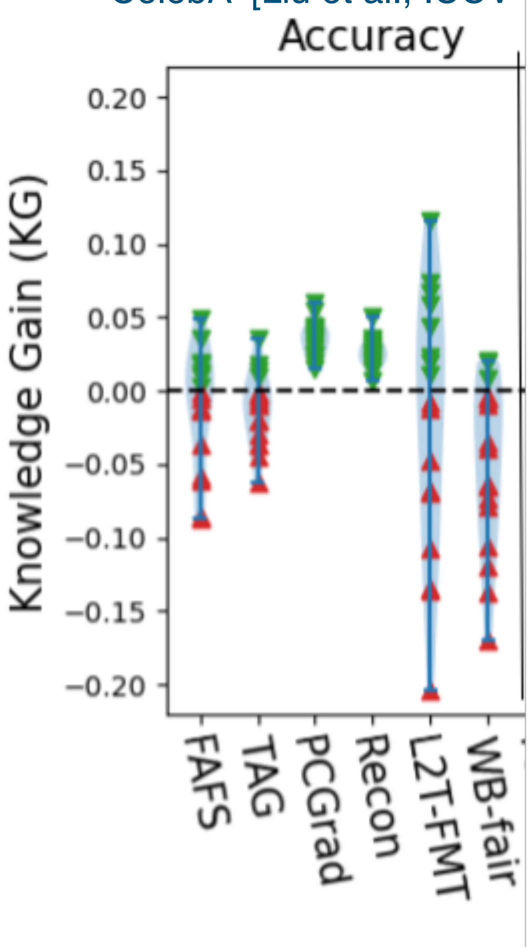
$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

Ideal scenario: achieve non-negative (green triangles) , i.e., KG(t)≥0 for all t.

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

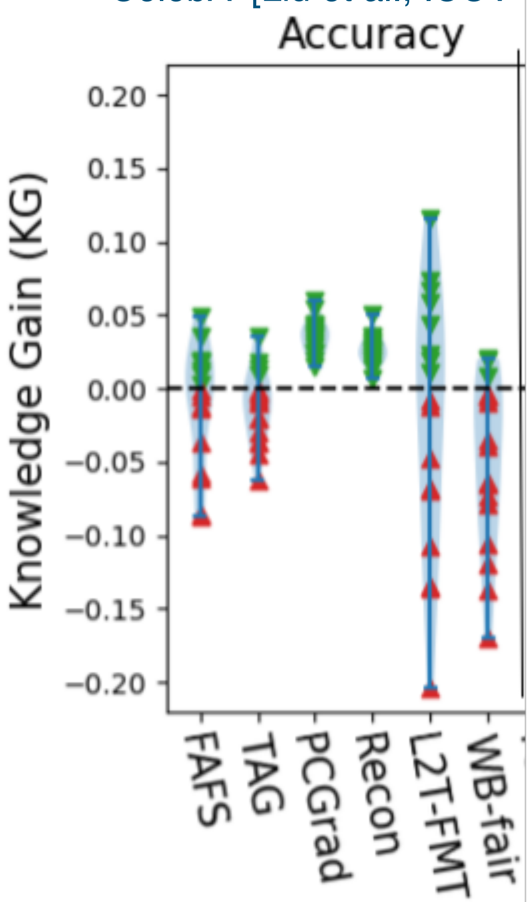
$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

Ideal scenario: achieve non-negative (green triangles) , i.e., KG(t)≥0 for all t.

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

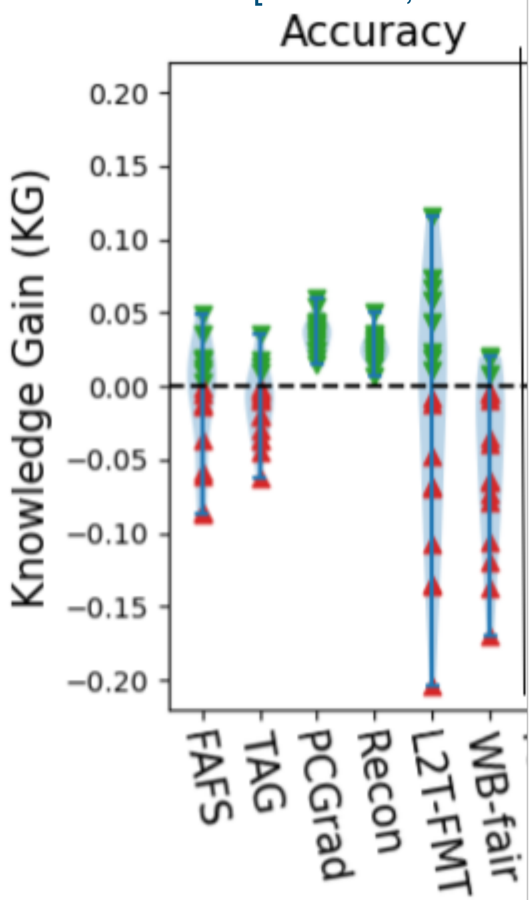
Ideal scenario: achieve non-negative (green triangles) , i.e., KG(t)≥0 for all t.

Negative Transfer: where KG(t)<0, (red triangles).

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

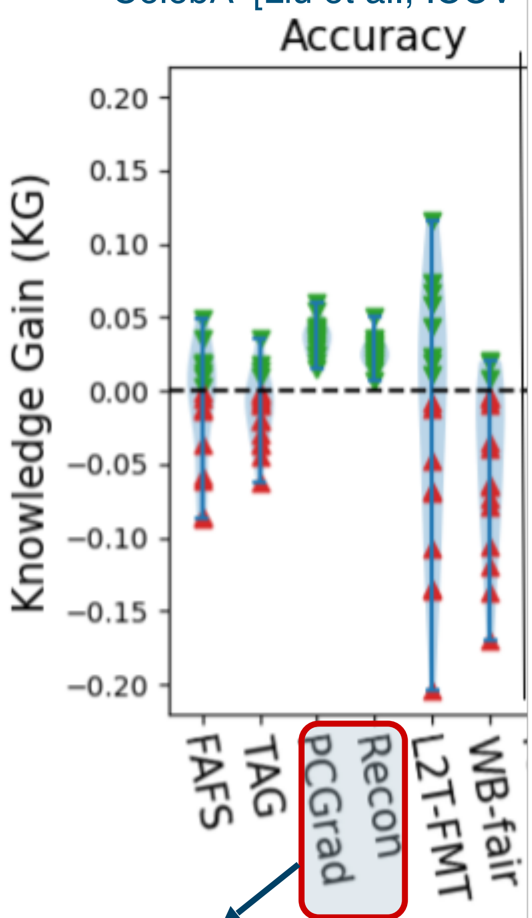
Ideal scenario: achieve non-negative (green triangles) , i.e., KG(t)≥0 for all t.

Negative Transfer: where KG(t)<0, (red triangles).

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Gradient Conflict to Negative Transfer

CelebA [Liu et al., ICCV 15]



Tackle accuracy conflicts

Knowledge Gain (KG): difference in accuracy between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$KG(t) : P(\mathcal{M}^t(X) = Y_t) - P(\mathcal{H}(X) = Y_t)$$

Ideal scenario: achieve non-negative (green triangles) , i.e., KG(t)≥0 for all t.

Negative Transfer: where KG(t)<0, (red triangles).

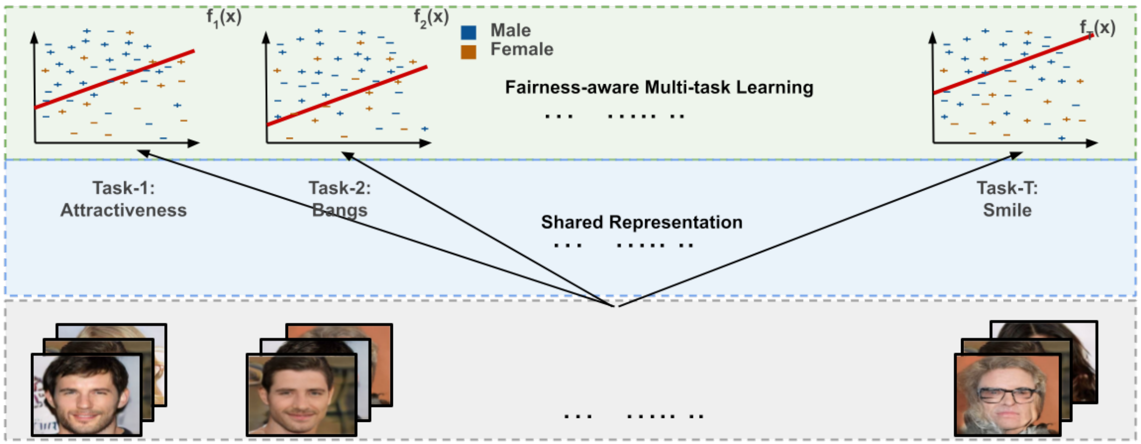
Root Cause: Research identified accuracy conflict as origin. [Guangyuan et al., ICLR 22; Yu et al., NeurIPS 20; Du et al., ContLearn 18] .

FAFS [Lu et al., CVPR 17], TAG [Fifty et al., NeurIPS 21], PCGrad [Yu et al., NeurIPS 20], Recon [Guangyuan et al., ICLR 22], L2TFMT [Roy et al., ECML 22], WB-fair~[Hu et al., ECML 23]

Problem Definition

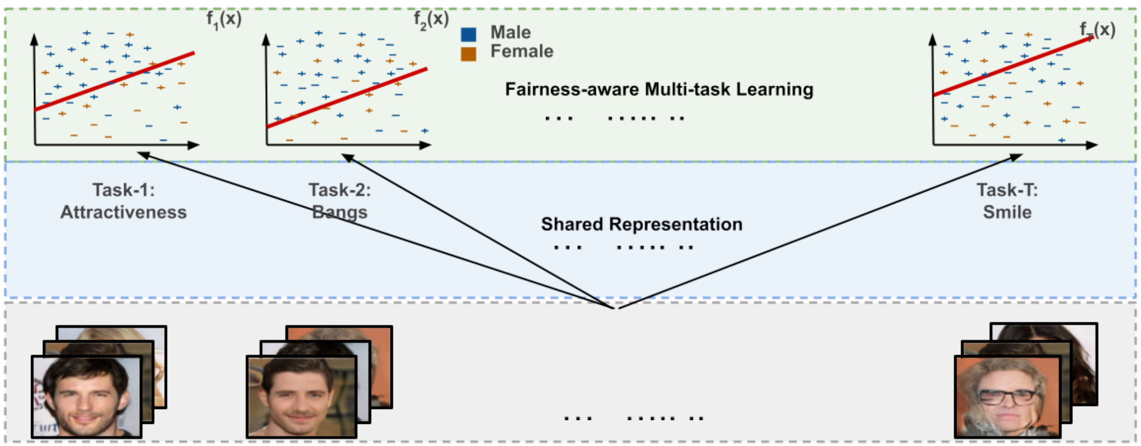
Fairness Issues in MTL

What is Fairness-aware MTL aka fair-MTL?



Fairness Issues in MTL

What is Fairness-aware MTL aka fair-MTL?



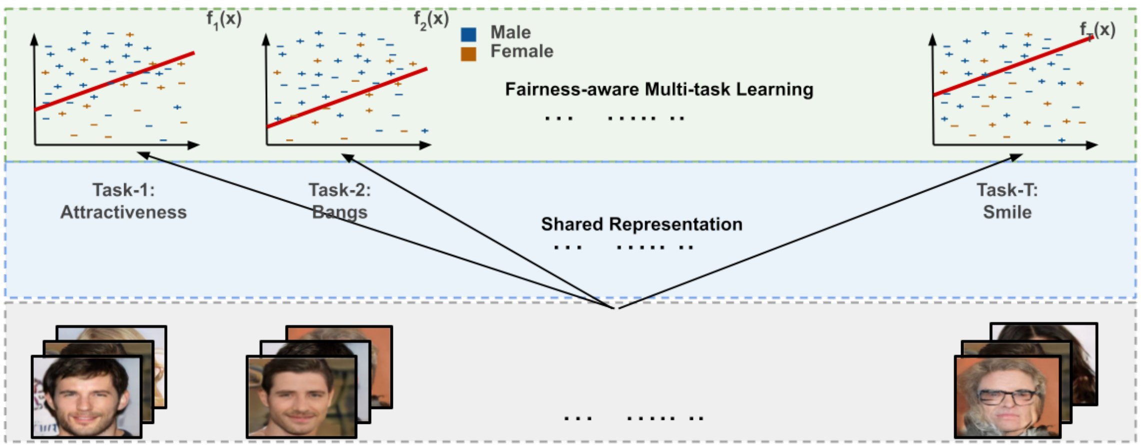
- learn multiple supervised prediction tasks without discrimination

$$F^{(t)}_{viol}(\mathcal{M}) = \sum_{c \in \mathbb{C}} |P(\mathcal{M}^t(X)|S = g, c) - P(\mathcal{M}^t(X)|S = \overline{g}, c)| \cong 0$$

g and $\overline{g}$ represents groups like “male”, and “female”.

Fairness Issues in MTL

What is Fairness-aware MTL aka fair-MTL?



- learn multiple supervised prediction tasks without discrimination

$$F^{(t)}_{viol}(\mathcal{M}) = \sum_{c \in \mathbb{C}} |P(\mathcal{M}^t(X)|S = g, c) - P(\mathcal{M}^t(X)|S = \overline{g}, c)| \cong 0$$

g and $\overline{g}$ represents groups like “male”, and “female”.

$$\underset{\theta}{\text{argmin}} \sum_t w_t \Big(\mathcal{L}_t(\theta, U) + \lambda_t \mathcal{F}_t(\theta, S) \Big)$$

Requires to optimize minimum two losses [Roy et al., ECMLPKDD 22] per task t :

- accuracy loss L_t and
- fairness loss F_t.

λ sets accuracy and fairness trade-off, ω sets the inter-task trade-off

Fairness Issues in MTL

Exaggerated Conflict Gradient Problem in fair-MTL

Hypothetical loss surface of the shared parameter space jointly trained with two accuracy L_1 and L_2 , and two fairness F_1 and F_2 losses

$$\operatorname*{argmin}_{\theta} \sum_{t} w_t \Big(\mathcal{L}_t(\theta, U) + \lambda_t \mathcal{F}_t(\theta, S) \Big)$$

Requires to optimize minimum two losses [Roy et al., ECMLPKDD 22] per task t :

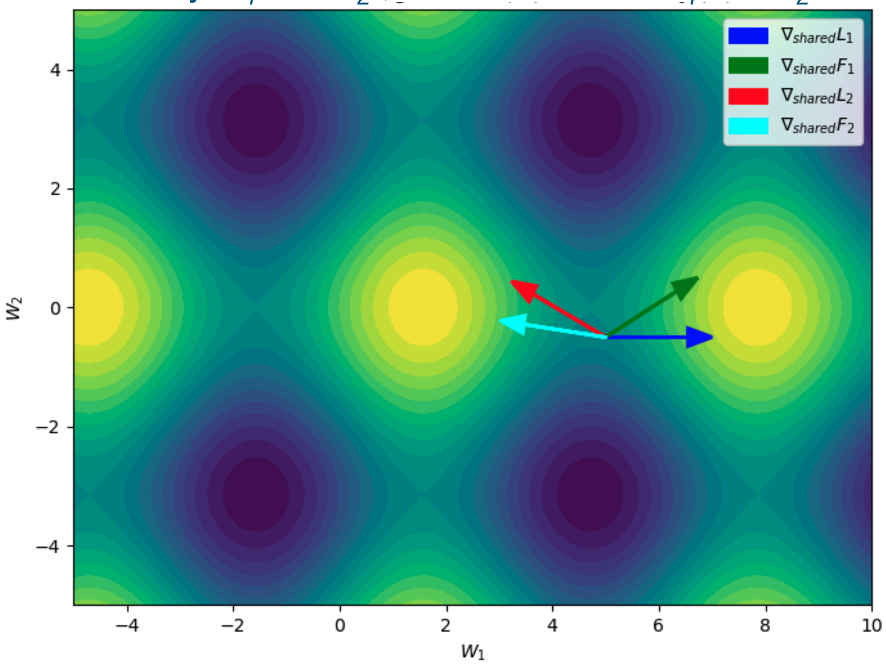
- accuracy loss L_t and
- fairness loss F_t .

λ sets accuracy and fairness trade-off, ω sets the inter-task trade-off

Fairness Issues in MTL

Exaggerated Conflict Gradient Problem in fair-MTL

Hypothetical loss surface of the shared parameter space jointly trained with two accuracy L_1 and L_2 , and two fairness F_1 and F_2 losses



$$\underset{\theta}{\operatorname{argmin}} \sum_t w_t \Big(\mathcal{L}_t(\theta, U) + \lambda_t \mathcal{F}_t(\theta, S) \Big)$$

Requires to optimize minimum two losses [Roy et al., ECMLPKDD 22] per task t :

- accuracy loss L_t and
- fairness loss F_t .

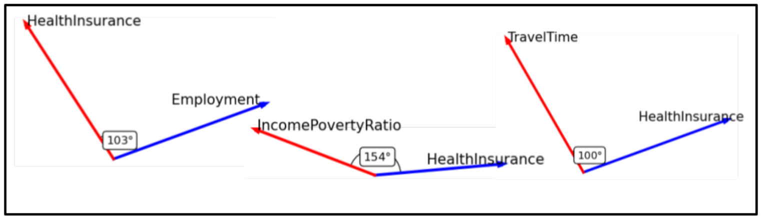
λ sets accuracy and fairness trade-off, ω sets the inter-task trade-off

More conflicts to deal with

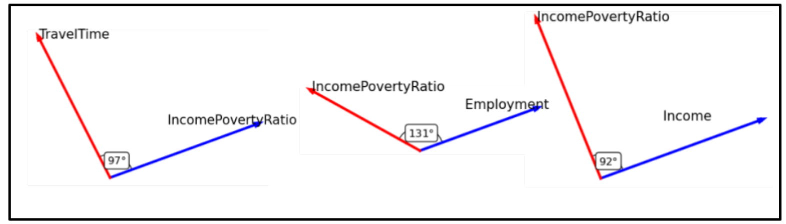
Introduces the fairness conflict problem

$$\nabla_{shared} F_1 \cdot \nabla_{shared} F_2 < 0$$

Fairness Conflict in SOTA MTL



Recon **[Guangyuan et al., ICLR 22]**

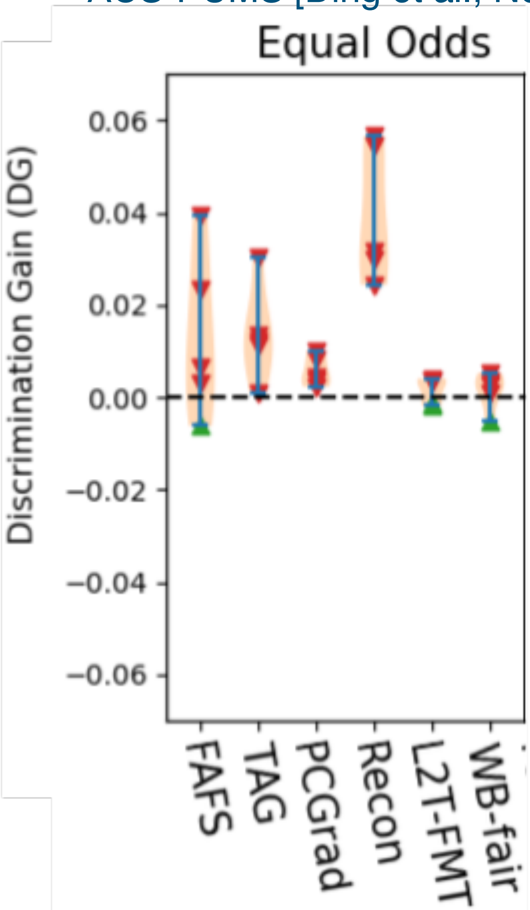


TAG **[Fifty et al., NeurIPS 21]**

- Fairness conflict observed in SOTA MTL methods when trained on real world census data [Ding et al., NeurIPS 21].

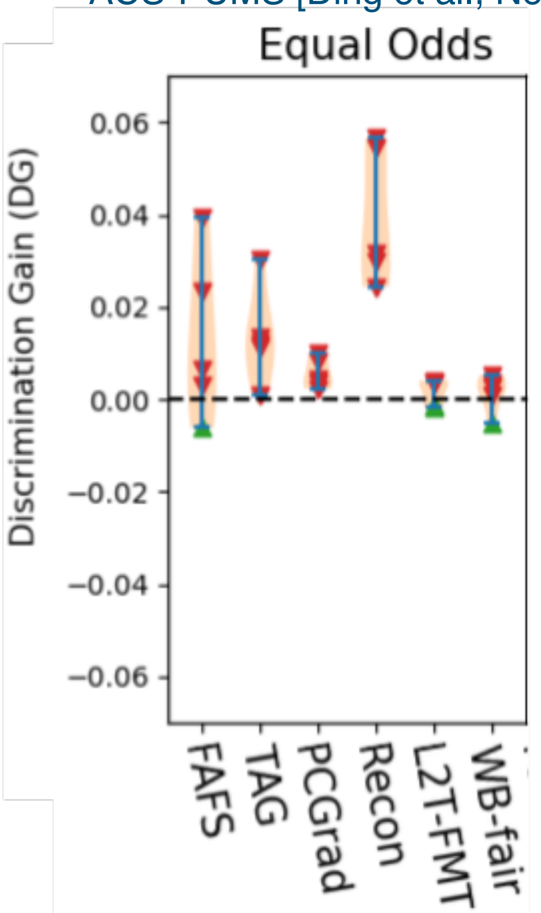
Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]



Fairness Conflict to Bias Transfer

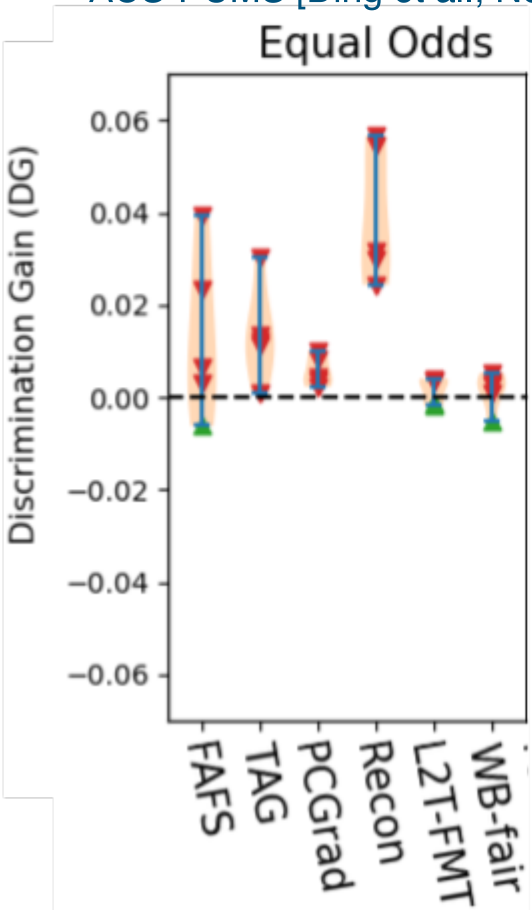
ACS-PUMS [Ding et al., NeurIPS 21]



$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]

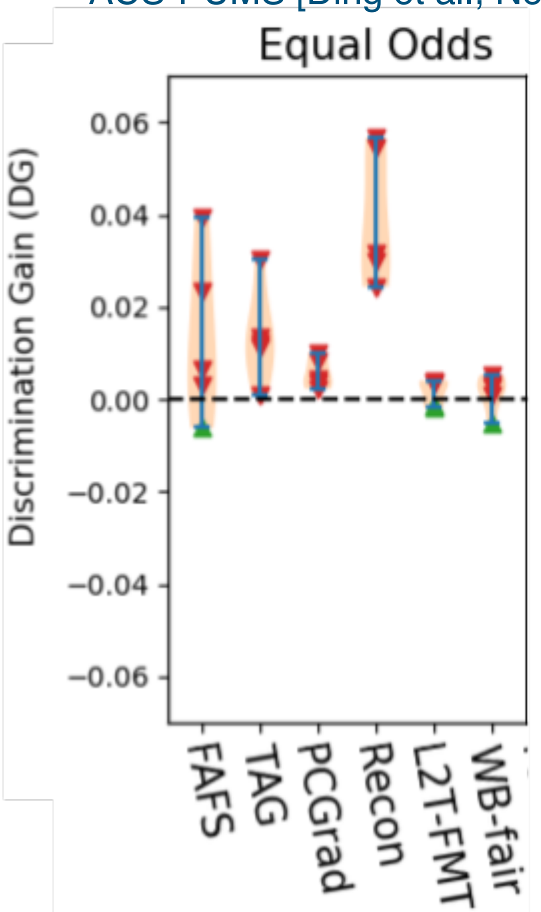


Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]

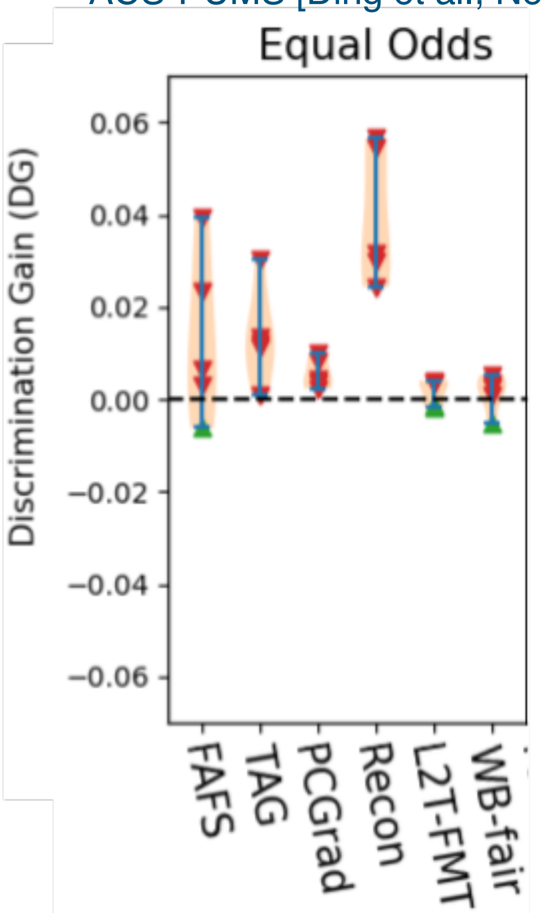


Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]

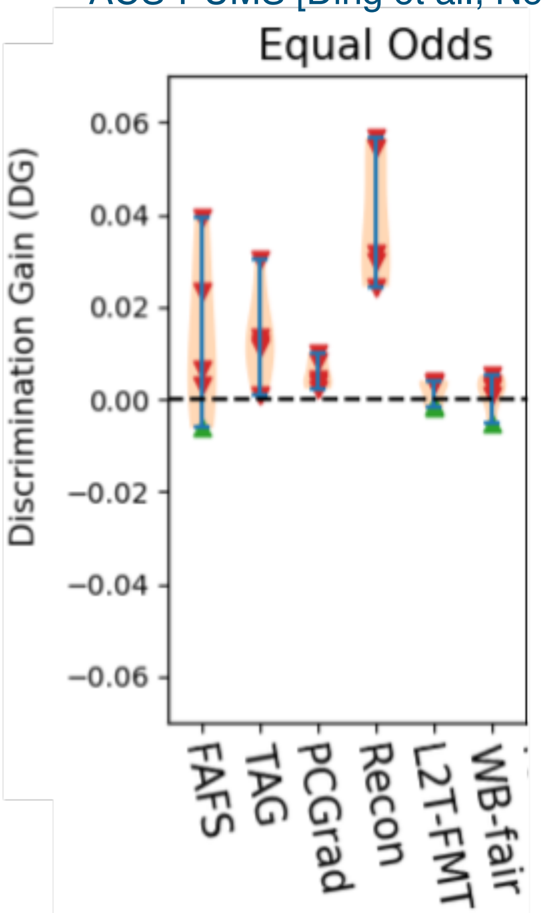


Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]



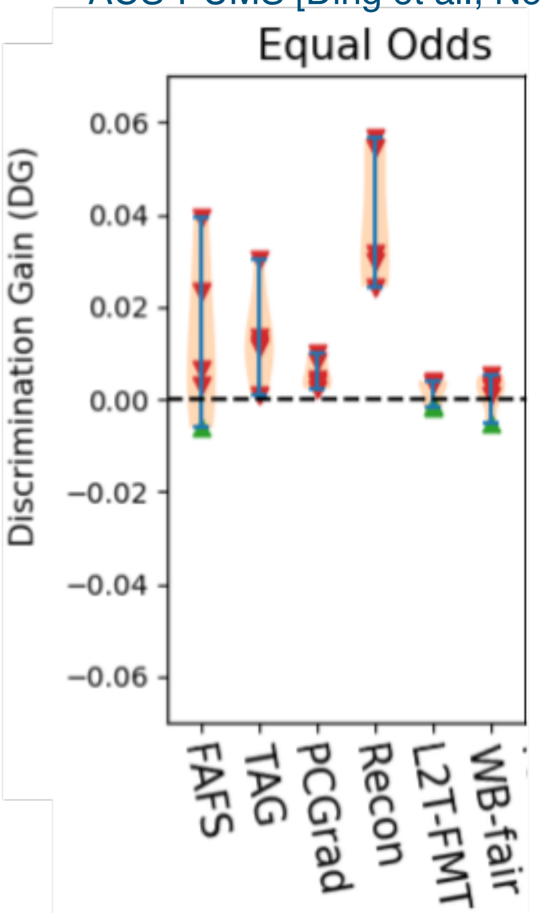
Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$DG(t) : F_{viol}^{(t)}(\mathcal{M}) - F_{viol}^{(t)}(\mathcal{H})$$

Bias Transfer: where DG(t)>0 i.e., positive gain of discrimination (red triangles).

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]



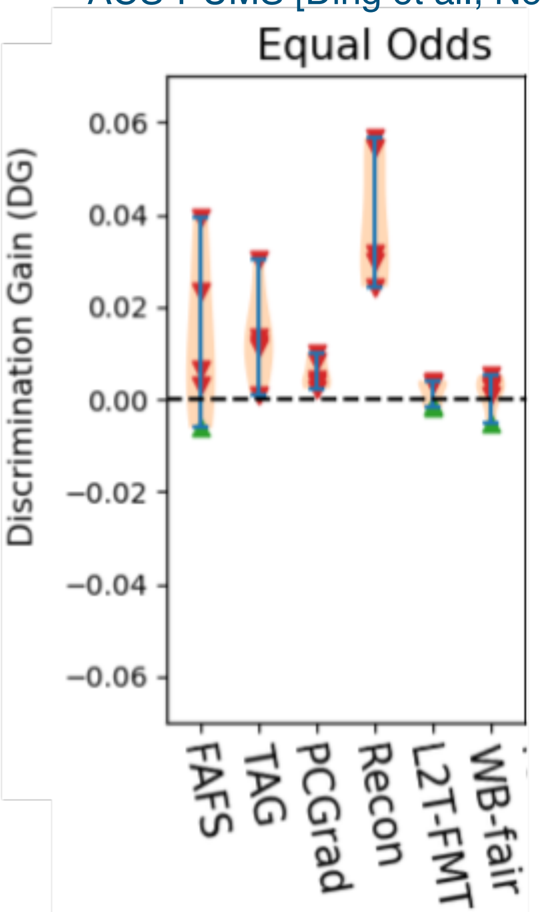
Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Bias Transfer: where DG(t)>0 i.e., positive gain of discrimination (red triangles).

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]



Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

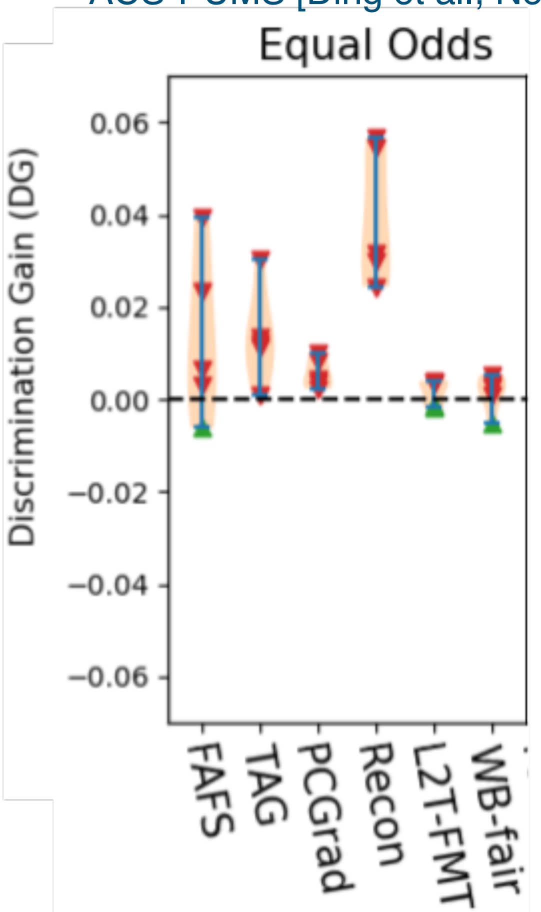
$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Bias Transfer: where DG(t)>0 i.e., positive gain of discrimination (red triangles).

Ideal scenario: non-positive bias transfer, i.e., DG(t)≤0 (green triangles).

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]



Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

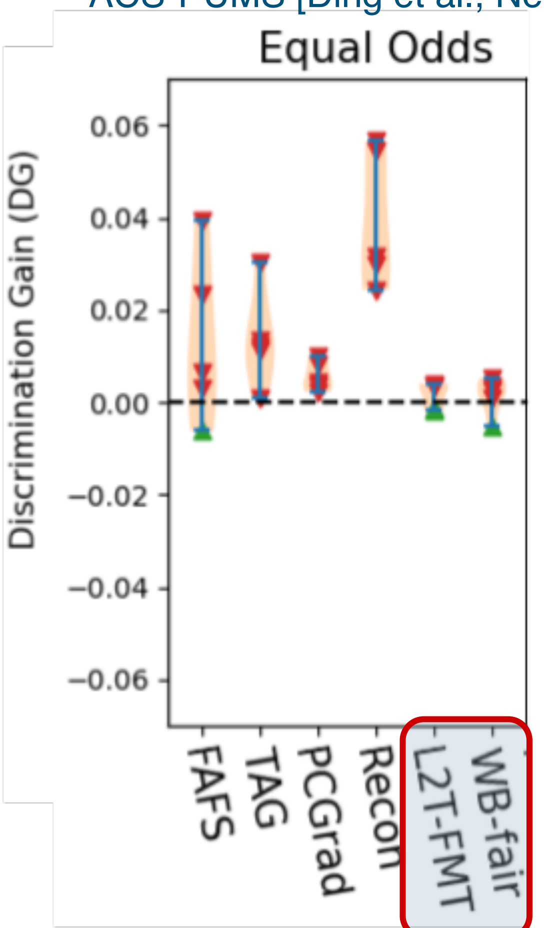
$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Bias Transfer: where DG(t)>0 i.e., positive gain of discrimination (red triangles).

Ideal scenario: non-positive bias transfer, i.e., DG(t)≤0 (green triangles).

Fairness Conflict to Bias Transfer

ACS-PUMS [Ding et al., NeurIPS 21]



fair-MTLs

Discrimination Gain (DG): difference in fairness violation between MTL ($\mathcal{M}$) and STL ($\mathcal{H}$) trained on t:

$$DG(t) : F^{(t)}_{viol}(\mathcal{M}) - F^{(t)}_{viol}(\mathcal{H})$$

Bias Transfer: where DG(t)>0 i.e., positive gain of discrimination (red triangles).

Ideal scenario: non-positive bias transfer, i.e., DG(t)≤0 (green triangles).

Root Cause: we hypothesize bias transfer originates from fairness conflict.

FairBranch

Desiderata from Related Research

Desiderata from SOTA MTL

Methods	Negative Transfer	Fairness	Dynamic Architecture
FAFS [Lu et al., CVPR 17]	✓	-	✓
TAG [Fifty et al., NeurIPS 21]	✓	-	-
PCGrad [Yu et al., NeurIPS 20]	✓	-	-
Recon [Guangyuan et al., ICLR 22]	✓	-	✓
L2TFMT [Roy et al., ECML 22]	-	✓	-
WB-fair [Hu et al., ECML 23]	-	✓	-

Desiderata from Related Research

Desiderata from SOTA MTL

Methods	Negative Transfer	Fairness	Dynamic Architecture
FAFS [Lu et al., CVPR 17]	✓	-	✓
TAG [Fifty et al., NeurIPS 21]	✓	-	-
PCGrad [Yu et al., NeurIPS 20]	✓	-	-
Recon [Guangyuan et al., ICLR 22]	✓	-	✓
L2TFMT [Roy et al., ECML 22]	-	✓	-
WB-fair [Hu et al., ECML 23]	-	✓	-

Tackle accuracy conflicts

Desiderata from Related Research

Desiderata from SOTA MTL

Methods	Negative Transfer	Fairness	Dynamic Architecture
FAFS [Lu et al., CVPR 17]	✓	-	✓
TAG [Fifty et al., NeurIPS 21]	✓	-	-
PCGrad [Yu et al., NeurIPS 20]	✓	-	-
Recon [Guangyuan et al., ICLR 22]	✓	-	✓
L2TFMT [Roy et al., ECML 22]	-	✓	-
WB-fair [Hu et al., ECML 23]	-	✓	-

Tackle fairness conflicts

Desiderata from Related Research

Desiderata from SOTA MTL

Methods	Negative Transfer	Fairness	Dynamic Architecture
FAFS [Lu et al., CVPR 17]	✓	-	✓
TAG [Fifty et al., NeurIPS 21]	✓	-	-
PCGrad [Yu et al., NeurIPS 20]	✓	-	-
Recon [Guangyuan et al., ICLR 22]	✓	-	✓
L2TFMT [Roy et al., ECML 22]	-	✓	-
WB-fair [Hu et al., ECML 23]	-	✓	-

Tackle erroneous over-generalization

Desiderata from Related Research

Desiderata from SOTA MTL

Methods	Negative Transfer	Fairness	Dynamic Architecture
FAFS [Lu et al., CVPR 17]	✓	-	✓
TAG [Fifty et al., NeurIPS 21]	✓	-	-
PCGrad [Yu et al., NeurIPS 20]	✓	-	-
Recon [Guangyuan et al., ICLR 22]	✓	-	✓
L2TFMT [Roy et al., ECML 22]	-	✓	-
WB-fair [Hu et al., ECML 23]	-	✓	-
FairBranch	✓	✓	✓

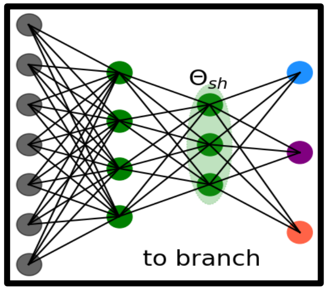
Tackle accuracy conflicts

Tackle fairness conflicts

Tackle erroneous over-generalization

Mitigating Conflicts for fair-MTL

Mitigating Conflicts for fair-MTL

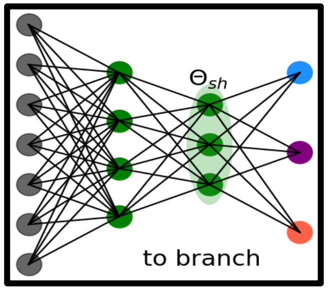


Groups Tasks on Parameter Similarity [Kornblith et al., ICML 19]**:**

- Intuition - strong parameter similarity ensures similar direction of minima.
- Expectation - move together without any conflict.

Addressing Negative Transfer

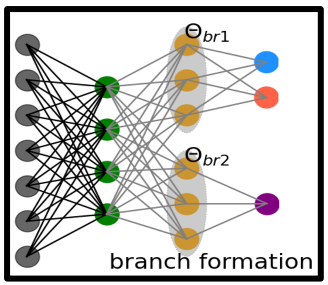
Mitigating Conflicts for fair-MTL



Groups Tasks on Parameter Similarity [Kornblith et al., ICML 19]**:**

- Intuition - strong parameter similarity ensures similar direction of minima.
- Expectation - move together without any conflict.

Addressing Negative Transfer

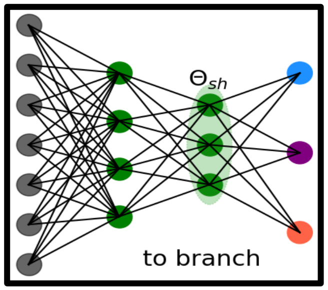


Branch Task Groups:

- Intuition - similar tasks benefits from sharing more knowledge.
- Expectation: sharing less with dissimilar tasks reduces over-generalization.

Addressing erroneous over-generalization

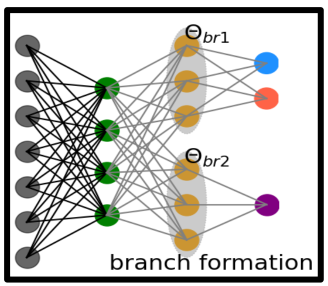
Mitigating Conflicts for fair-MTL



Groups Tasks on Parameter Similarity [Kornblith et al., ICML 19]**:**

- Intuition - strong parameter similarity ensures similar direction of minima.
- Expectation - move together without any conflict.

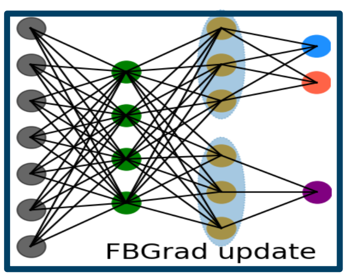
Addressing Negative Transfer



Branch Task Groups:

- Intuition - similar tasks benefits from sharing more knowledge.
- Expectation: sharing less with dissimilar tasks reduces over-generalization.

Addressing erroneous over-generalization



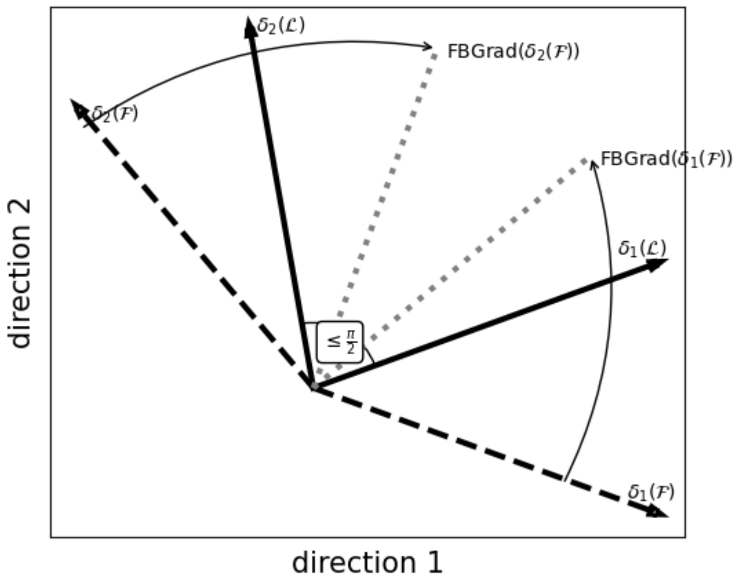
Conflict-free Fairness Correction:

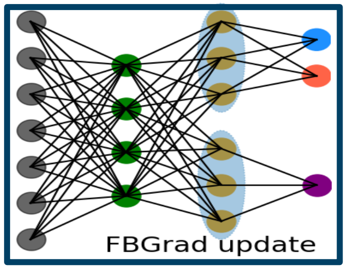
- Intuition - correcting the fairness conflict between task gradients within tasks groups ensures fair-MTL without Bias Transfer.

Addressing Bias Transfer

Mitigating Conflicts for fair-MTL

Hypothetical example of Fairness Gradient Conflict correction





Conflict-free Fairness Correction:

- Intuition - correcting the fairness conflict between task gradients within tasks groups ensures fair-MTL without Bias Transfer.

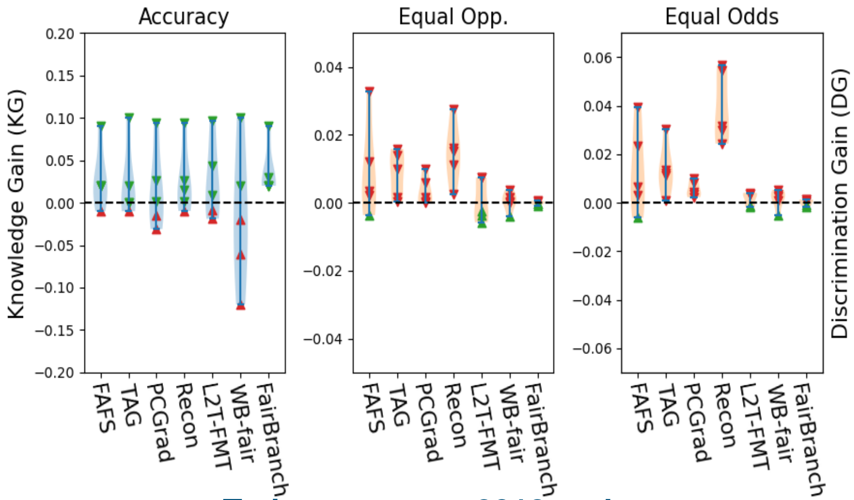
Addressing Bias Transfer

Experiments

Comparative Results

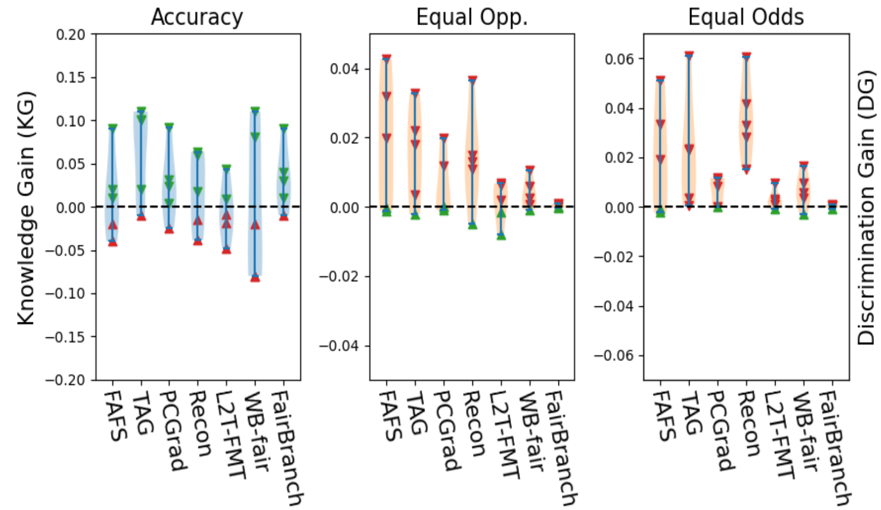
Tabular Data: ACS-PUMS Census Data [Ding et al., NeurIPS 21]

ACS-PUMS 18-19



Train on census 2018 and test on census 2019. Gender used as protected attribute.
# of tasks: 5

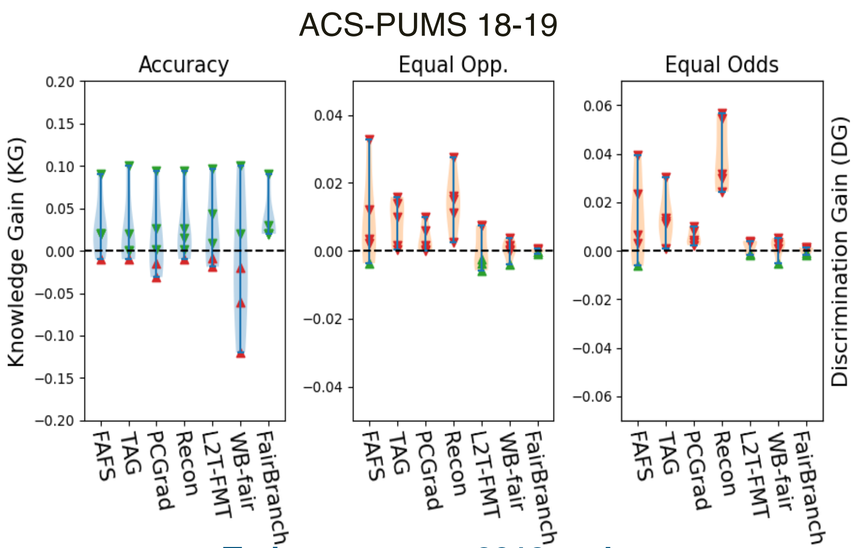
ACS-PUMS 19-21



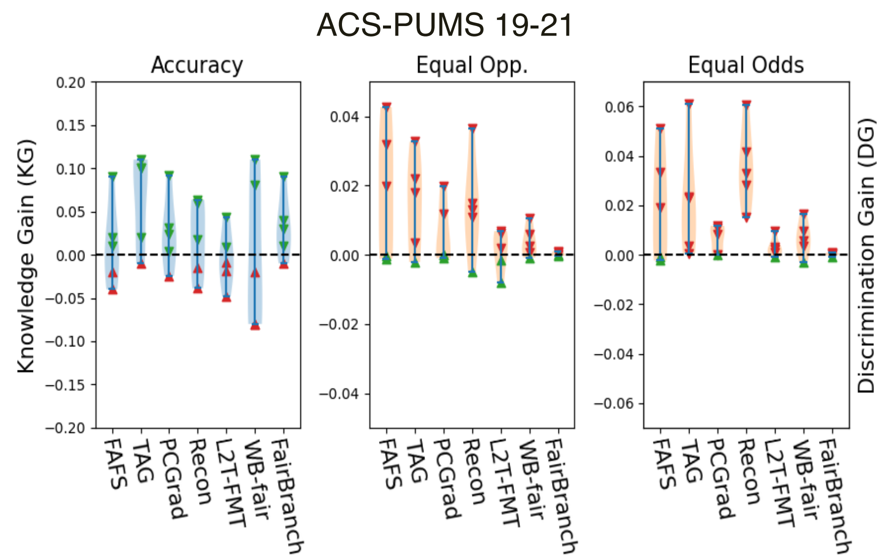
Train on census 2019 and test on census 2021. Gender used as protected attribute.
# of tasks: 5

Comparative Results

Tabular Data: ACS-PUMS Census Data [Ding et al., NeurIPS 21]



Train on census 2018 and test on census 2019. Gender used as protected attribute.
# of tasks: 5

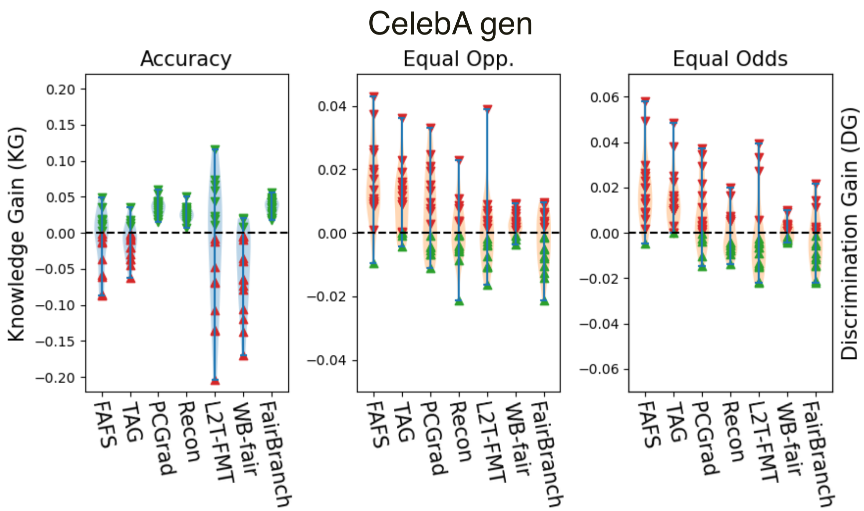


Train on census 2019 and test on census 2021. Gender used as protected attribute.
# of tasks: 5

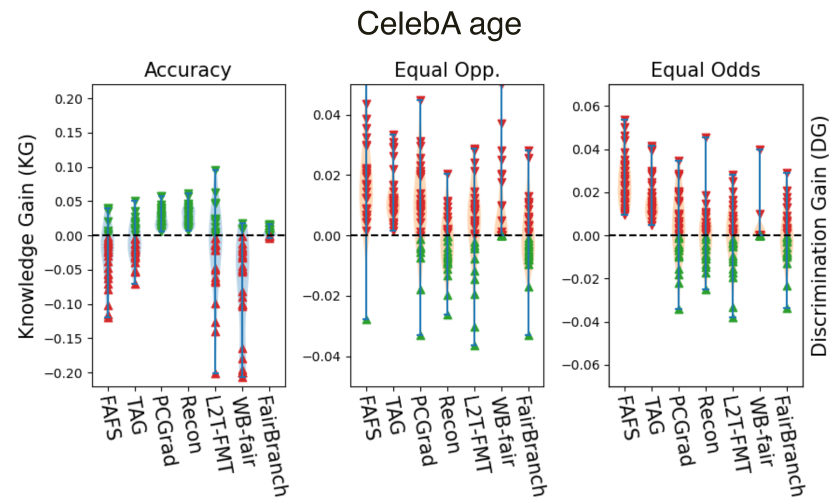
- **FairBranch effectively tackles** both **negative transfer** (non-negative KG) and **bias transfer** (non-positive DG).
- Among competitors, conflict correction on parameter space (PCGrad, Recon) outperform other on negative transfer.

Comparative Results

Visual Data: CelebA Data [Liu et al., ICCV 15]



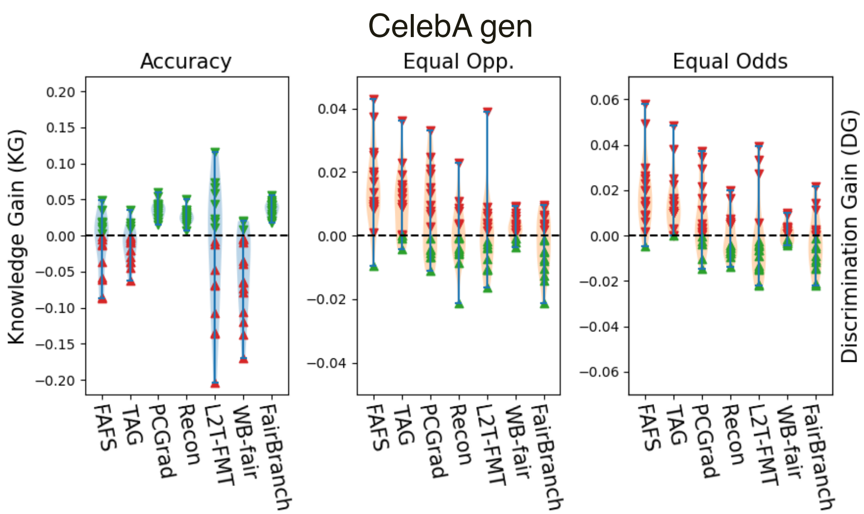
Gender used as protected attribute
# of tasks 17



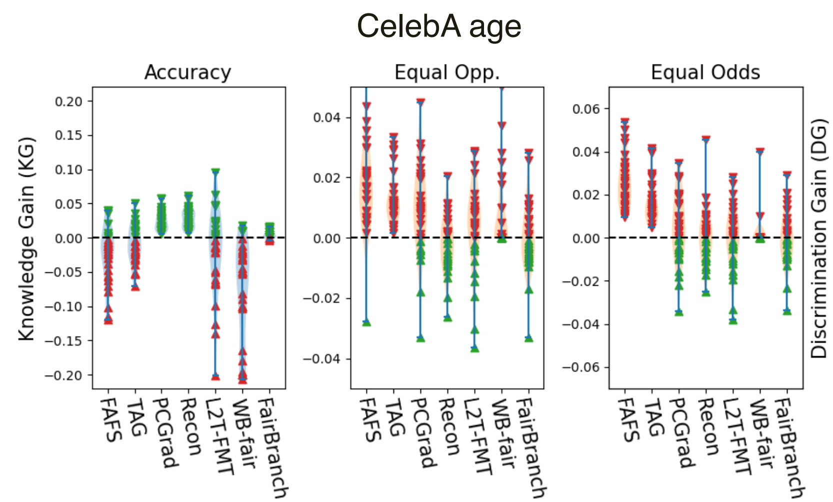
Age used as protected attribute
# of tasks 31

Comparative Results

Visual Data: CelebA Data [Liu et al., ICCV 15]



Gender used as protected attribute
# of tasks 17



Age used as protected attribute
# of tasks 31

- **FairBranch effectively tackles negative transfer** (non-negative KG), but suffers from **bias transfer** (positive DG) in some tasks.
- Among competitors, conflict correction on parameter space (PCGrad, Recon) outperform other on negative transfer.

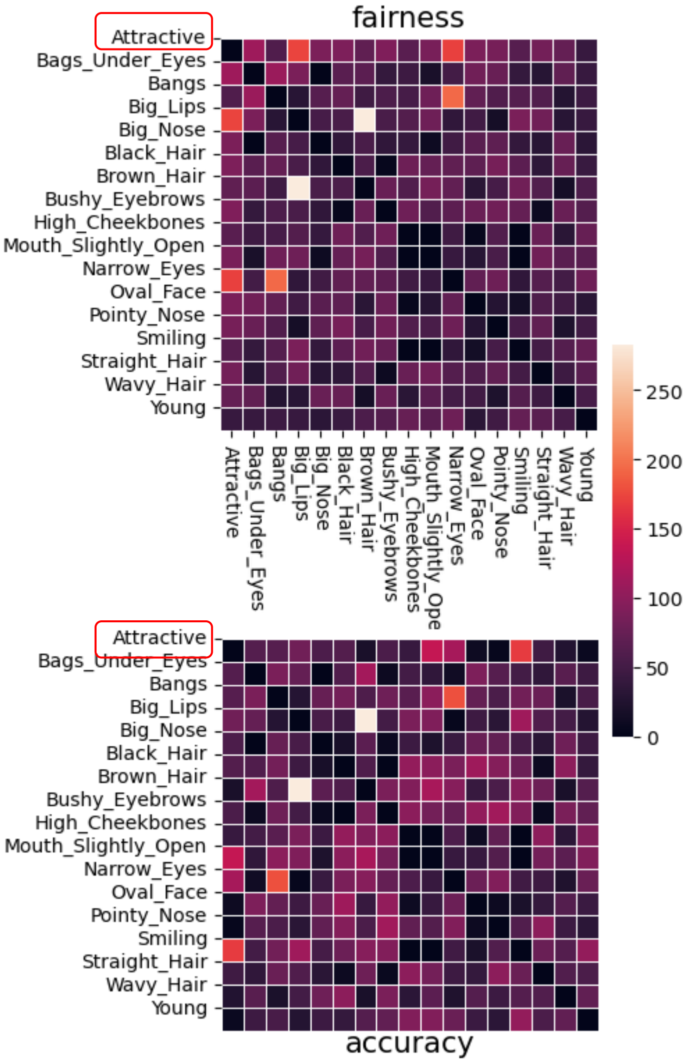
Comparative Results

Reporting on the average Knowledge Gain ($\overline{KG}$) and average Discrimination Gain ($\overline{DG}$) :

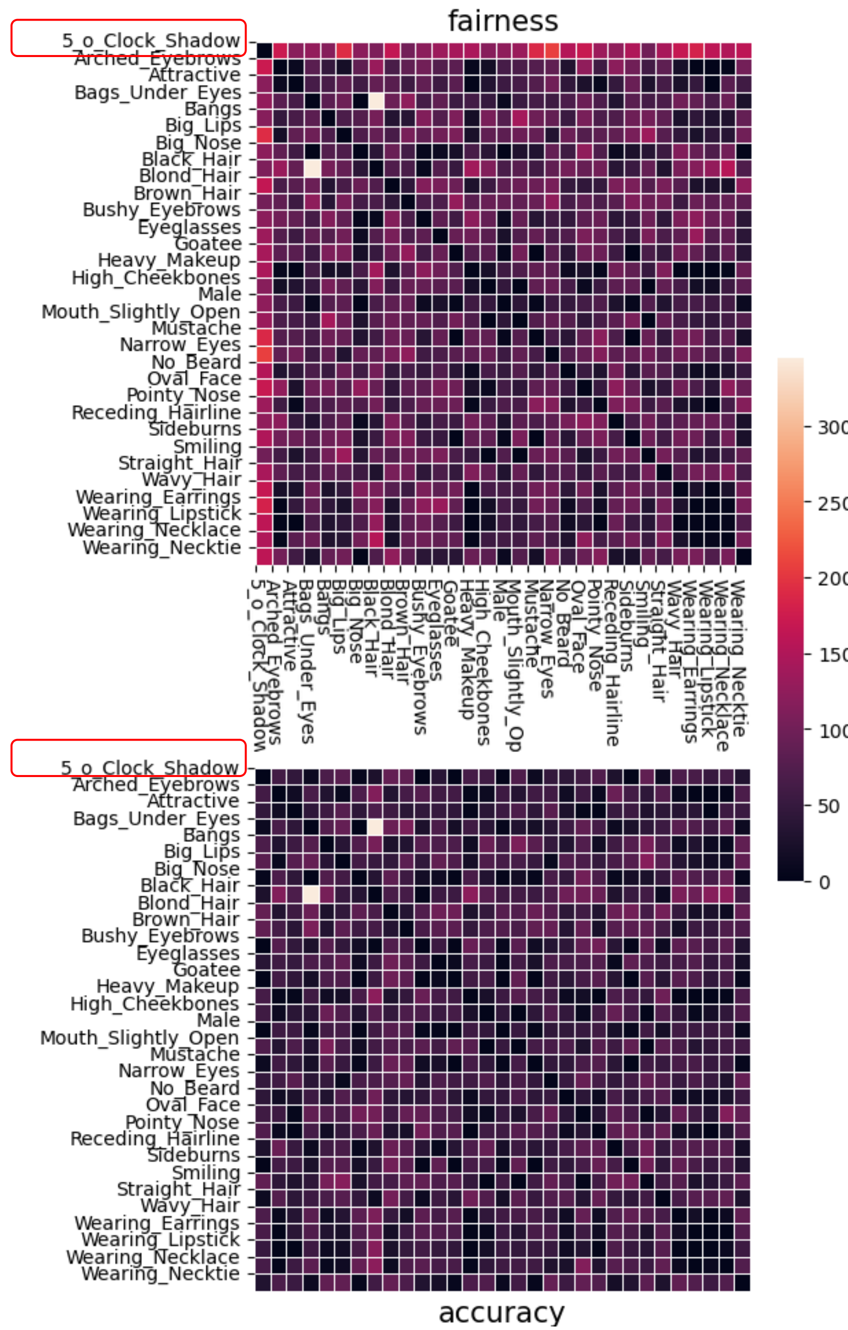
	Model	Metric		ACS-PUMS 18-19	ACS-PUMS 19-21	CelebA gen	CelebA age
Task-grouping	FAFS	$\overline{KG}$		0.028	0.012	-0.011	-0.024
		$\overline{DG}$	EP	0.009	0.019	0.015	0.017
			EO	0.013	0.020	0.019	0.026
	TAG	$\overline{KG}$		0.022	**0.064**	-0.012	-0.010
		$\overline{DG}$	EP	0.008	0.015	0.015	0.013
			EO	0.014	0.022	0.010	0.017
Conflict aware	PCGrad	$\overline{KG}$		0.015	0.025	0.035	0.025
		$\overline{DG}$	EP	0.004	0.006	0.007	0.009
			EO	0.006	0.006	0.008	0.004
	Recon	$\overline{KG}$		0.025	0.017	0.026	**0.028**
		$\overline{DG}$	EP	0.015	0.014	-0.001	0.005
			EO	0.040	0.036	0.001	0.009
Fairness aware	L2TFMT	$\overline{KG}$		0.024	-0.005	-0.022	-0.020
		$\overline{DG}$	EP	0.001	0.001	-0.002	0.0
			EO	0.002	0.003	0.001	0.003
	WB-fair	$\overline{KG}$		-0.016	0.002	-0.051	-0.080
		$\overline{DG}$	EP	0.001	0.004	0.001	0.002
			EO	0.002	0.006	0.003	0.007
Our	FairBranch	$\overline{KG}$		**0.036**	0.032	**0.036**	0.006
		$\overline{DG}$	EP	**-0.001**	**0.0**	**-0.004**	**-0.001**
			EO	**0.0**	**0.0**	**-0.003**	**0.0**

- **FairBranch outperforms all the competitors on 10 out of 12** evaluation report.
- In all experiment FairBranch have average Knowledge Gain > 0, and average Discrimination Gain ≤ 0.
- In visual data even under large # of tasks, SOTA MTLs like TAG, FAFS fails, FairBranch consistently positive on Knowledge Gain.
- Similar findings for fairness against SOTA fair-MTL observed with L2TFMT, WB-fair on Discrimination Gain.

Conflict Analysis of FairBranch

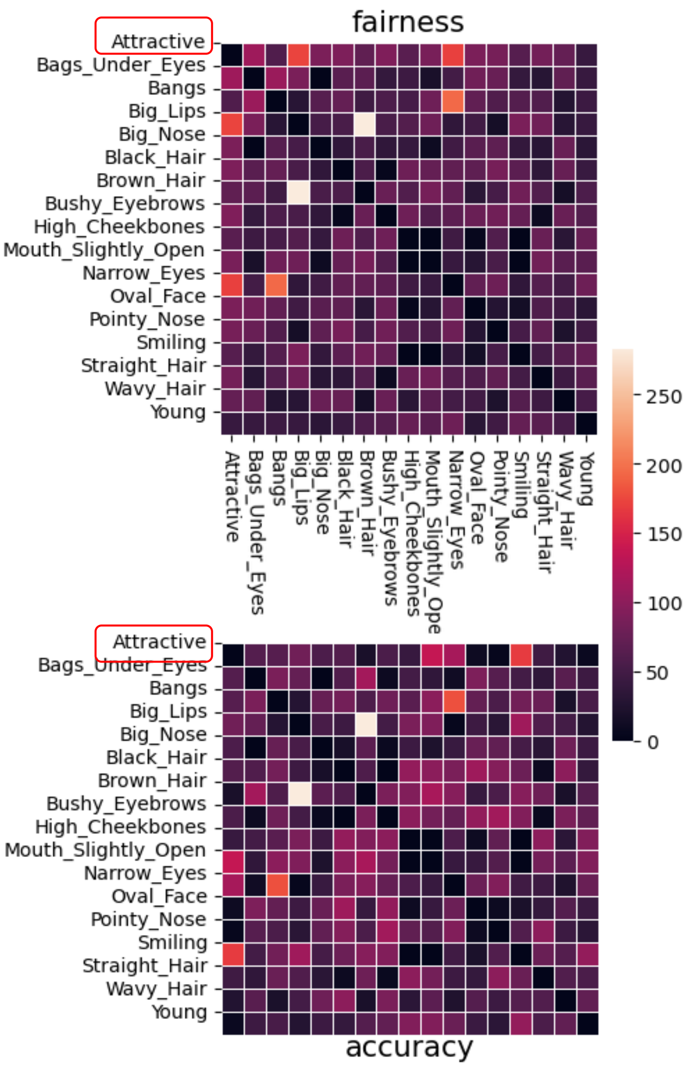


CelebA Gender



CelebA Age

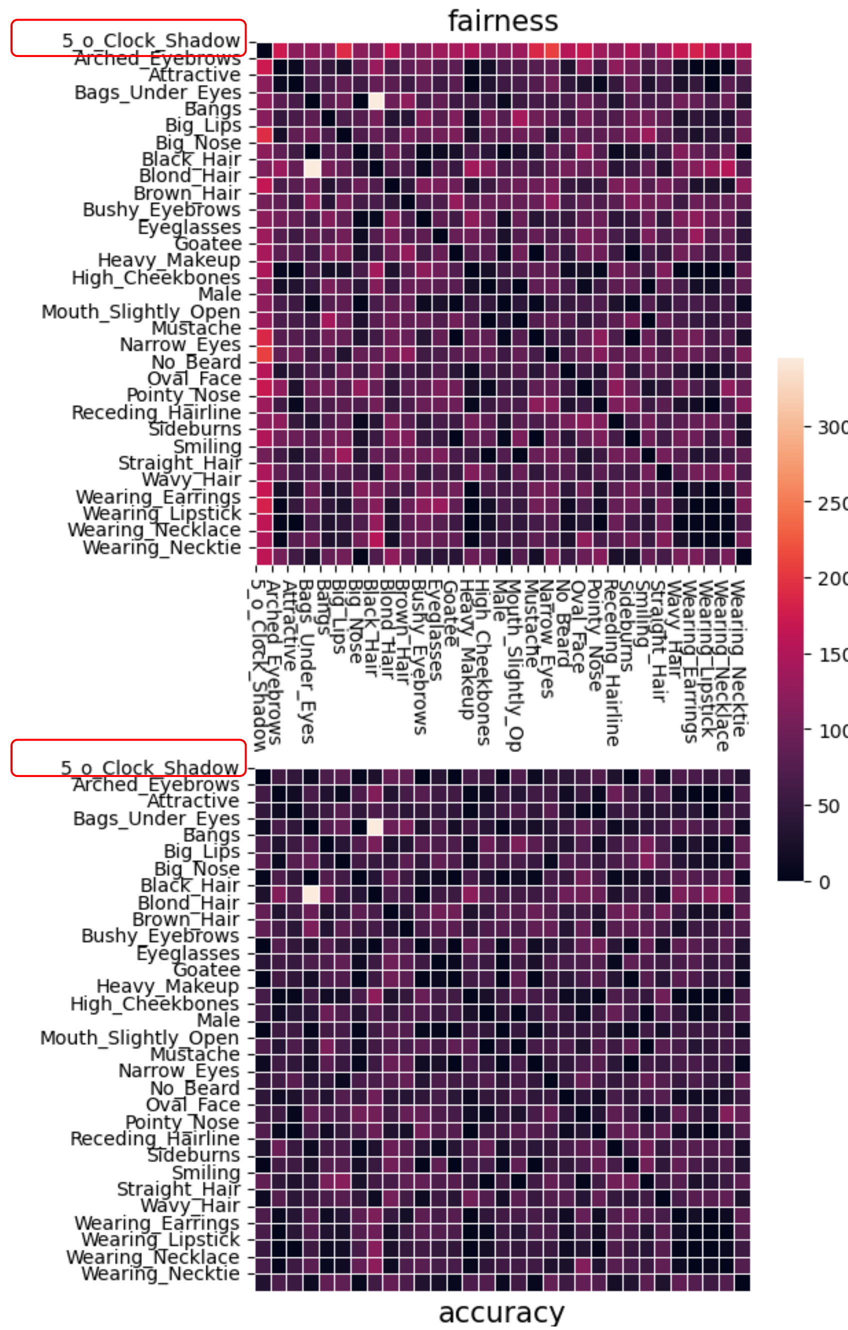
Conflict Analysis of FairBranch



CelebA Gender

Conflict Heatmaps :

- tasks like 'Attractive' and '5 o Clock shadow' have fewer accuracy conflicts but many fairness conflicts across all tasks.



CelebA Age

❖ Discussion and Conclusion

Key Takeaways

Key Takeaways

Key Takeaways

- **FairBranch tackles negative transfer and bias transfer better than the competitors.**

Key Takeaways

- **FairBranch tackles negative transfer and bias transfer better than the competitors.**

Key Takeaways

- **FairBranch tackles negative transfer and bias transfer better than the competitors.**
- **FairBranch outperforms the competitors on average knowledge and discrimination gain.**

Key Takeaways

- **FairBranch tackles negative transfer and bias transfer better than the competitors.**
- **FairBranch outperforms the competitors on average knowledge and discrimination gain.**

Key Takeaways

- FairBranch tackles negative transfer and bias transfer better than the competitors.
- FairBranch outperforms the competitors on average knowledge and discrimination gain.
- Tackling negative transfer on parameter space is advantageous over on output (loss) space.

Key Takeaways

- FairBranch tackles negative transfer and bias transfer better than the competitors.
- FairBranch outperforms the competitors on average knowledge and discrimination gain.
- Tackling negative transfer on parameter space is advantageous over on output (loss) space.

Key Takeaways

- **FairBranch tackles negative transfer and bias transfer better than the competitors.**
- **FairBranch outperforms the competitors on average knowledge and discrimination gain.**
- **Tackling negative transfer on parameter space is advantageous over on output (loss) space.**
- **Learning fair multi-task learning (MTL) is challenging due to the complex decisions required, as certain tasks contribute positively to accuracy knowledge transfer while hindering fairness knowledge transfer.**

References

- F. Ding, M. Hardt, J. Miller, and L. Schmidt, "Retiring adult: New datasets for fair machine learning," NeurIPS, vol. 34, 2021.
- Y. Du, W. M. Czarnecki, S. M. Jayakumar, M. Farajtabar, R. Pascanu, and B. Lakshminarayanan, "Adapting auxiliary losses using gradient similarity," Continual learning Workshop at NeurIPS 2018.
- C. Fifty, E. Amid, Z. Zhao, T. Yu, R. Anil, and C. Finn, "Efficiently identifying task groupings for multi-task learning," NeurIPS, vol. 34, pp. 27 503–27 516, 2021.
- S. Guangyuan, Q. Li, W. Zhang, J. Chen, and X.-M. Wu, "Recon: Reducing conflicting gradients from the root for multi-task learning," in 11th ICLR, 2022.
- M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," NeurIPS, vol. 29, pp. 3315–3323, 2016.
- F. Hu, P. Ratz, and A. Charpentier, "Fairness in multi-task learning via wasserstein barycenters," in ECMLPKDD. Springer, 2023, pp. 295–312.
- S. Kornblith, M. Norouzi, H. Lee, and G. Hinton, "Similarity of neural network representations revisited," in ICML, 2019, pp. 3519–35.
- Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in ICCV, December 2015.
- A. Roy and E. Ntoutsi, "Learning to teach fairness-aware deep multi-task learning," in ECMLPKDD. Springer, 2022, pp. 710–726.
- T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, "Gradient surgery for multi-task learning," NeurIPS, vol. 33, pp. 5824–5836, 2020.

Universität der Bundeswehr München

CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

CODE

Question??

Thank you for your attention

Find me via: Google Scholar, Github, LinkedIn, YouTube
arjun.roy@unibw.de

For more details about FairBranch:

This work is supported by:
European Horizon Project MAMMOth
EU HORIZON-RIA Project ID:101070285

Funded by the European Union

CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

Question??

Thank you for your attention

Find me via: Google Scholar, Github, LinkedIn, YouTube
arjun.roy@unibw.de

For more details about FairBranch:

This work is supported by:
European Horizon Project MAMMOth
EU HORIZON-RIA Project ID:101070285

Funded by the European Union