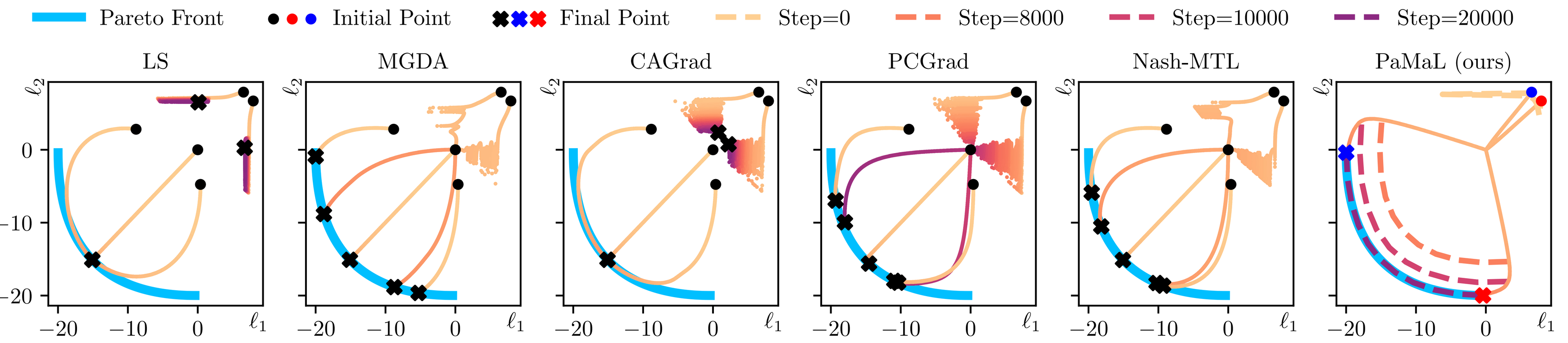
Pareto Manifold Learning: Tackling multiple tasks via ensembles of single-task models

Nikolaos Dimitriadis, Pascal Frossard, Francois Fleuret

Contributions

- Geometrical view: the Pareto Front admits a linear parameterization in parameter space.
- We propose Pareto Manifold Learning, a novel weight-ensembling approach that produces a continuous Pareto Front in a single training run, allowing to modulate the performance on each task during inference.
- Extended experimental validation: PaMaL outperforms state-of-the-art single-point algorithms, while learning a better Pareto parameterization than multi-point baselines.



Problem Formulation

Problem: vector optimization problem $\min_{\theta} \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}[\boldsymbol{L}(\boldsymbol{f}(\boldsymbol{x};\boldsymbol{\theta}),\boldsymbol{y})]$

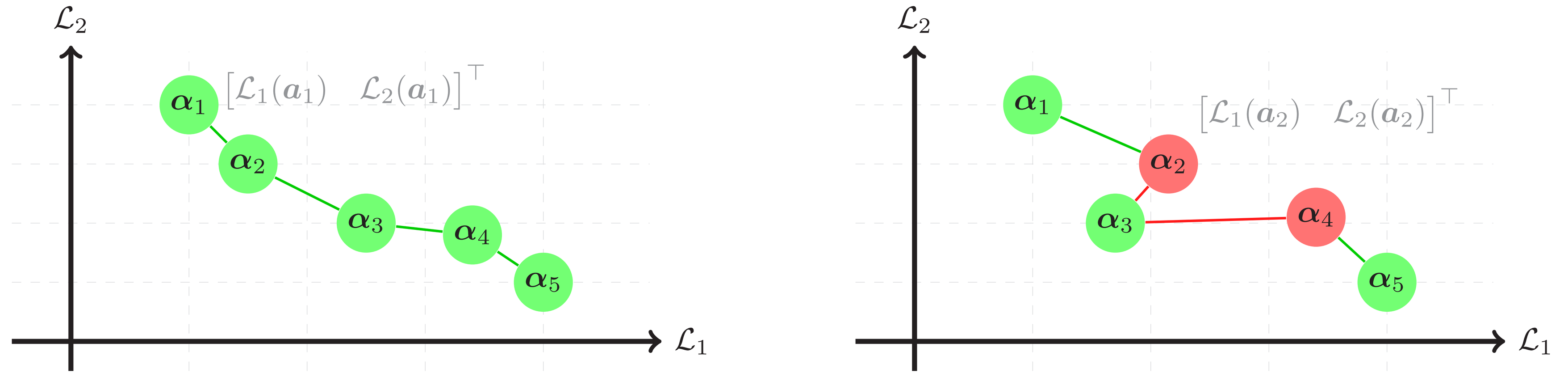
Goal: a continuous parameterization of the Pareto Front

We can select the model satisfying our desired trade-off

Algorithm

- Weight ensemble of single-task predictors. At each step, we perform the forward pass with a randomly selected model lying in the convex hull of ensemble members.
- PaMaL objective generalizes Linear Scalarization [Kur+22,Xin+22]

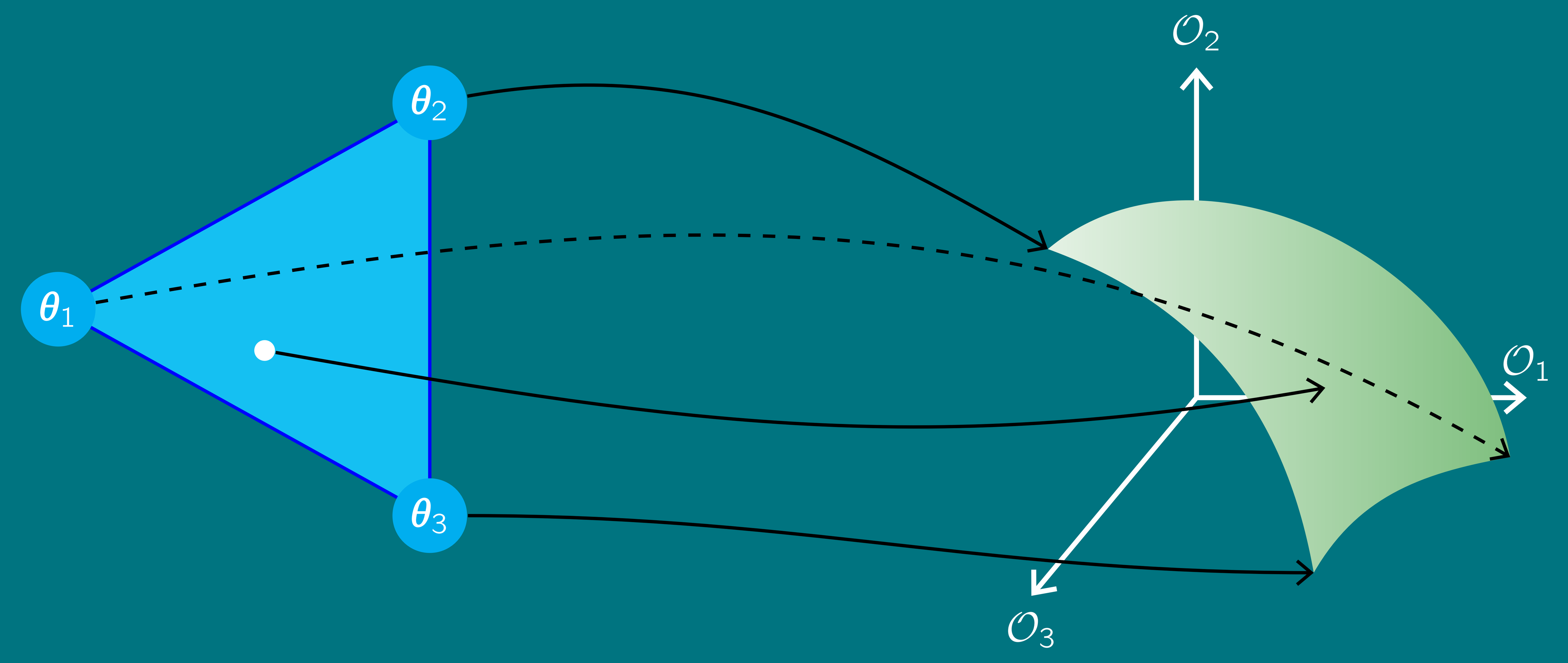
$$\text{objective} = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}}\left[\mathbb{E}_{\boldsymbol{\alpha}\sim\mathrm{P}}\left[\boldsymbol{\alpha}^{\top}\boldsymbol{L}\left(\boldsymbol{y},\boldsymbol{f}\left(\boldsymbol{x};\boldsymbol{\alpha}\boldsymbol{\Theta}\right)\right)\right]\right]$$



Proposed regularization penalizes violations of monotonicity constraints promoting functional diversity.

UNIVERSITÉ DE GENÈVE

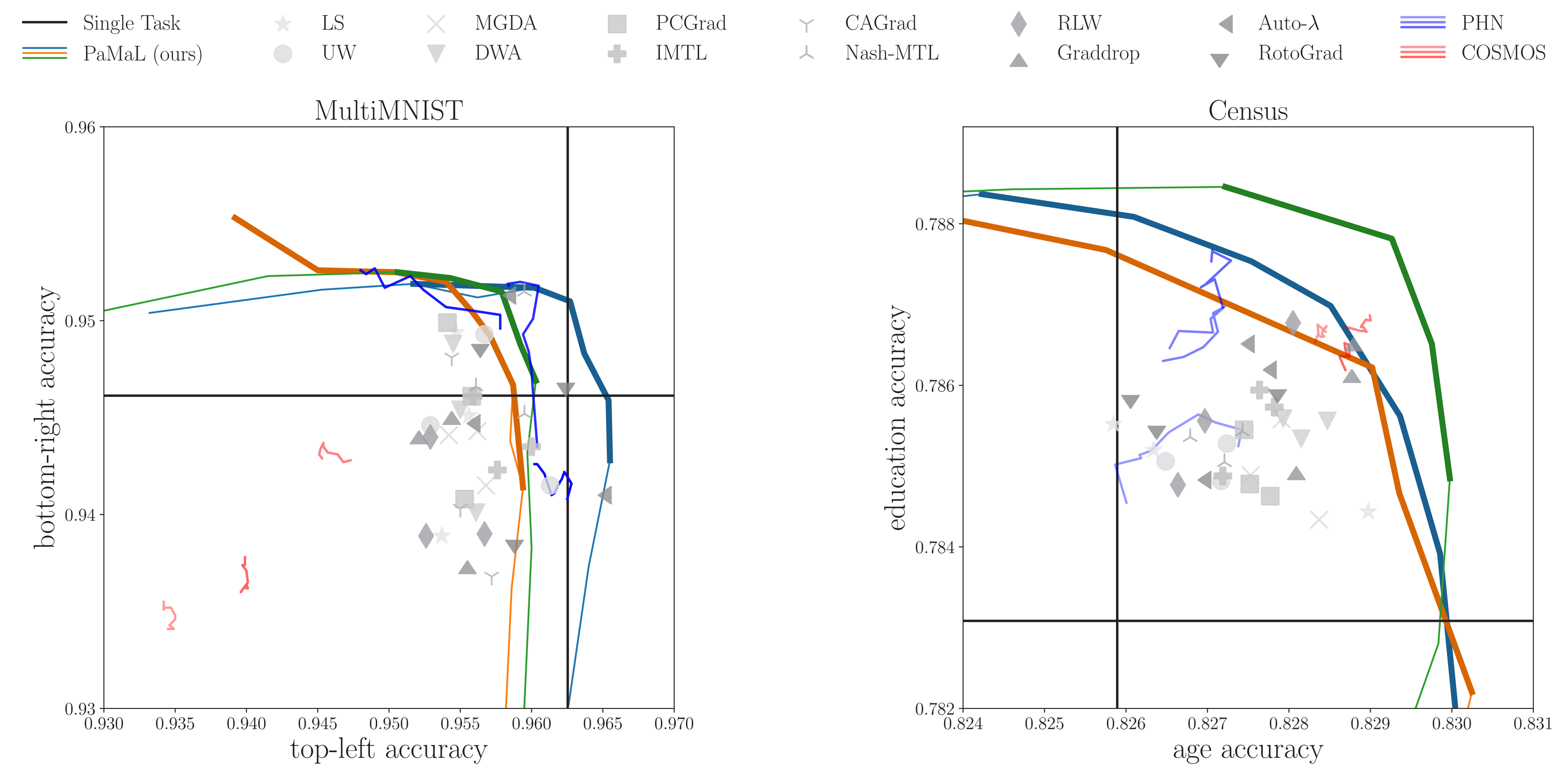
We linearly parameterize a Pareto Front in Multi-Task Learning and train it in a single run.

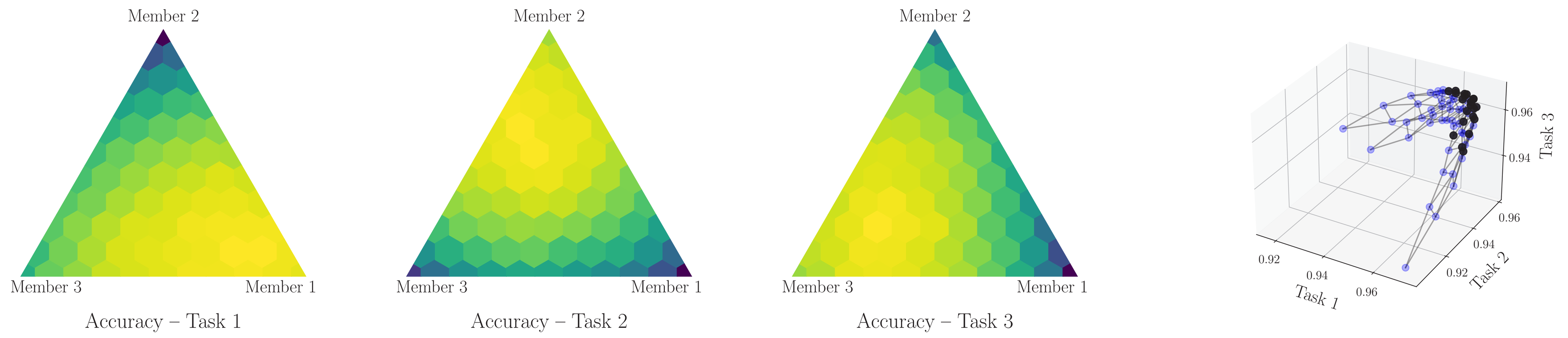


Weight space $\mathcal{W}$ Objective space $\mathcal{O}$

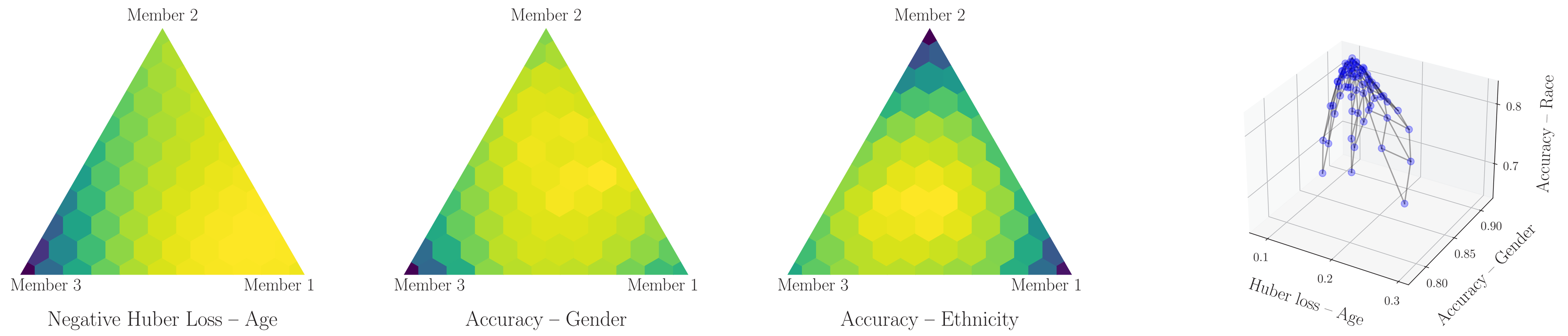
Experiments

PaMaL produces a reliable mapping from preference to objective space and outperforms SOTA single-point algorithms.





MultiMNIST-3: PaMaL produces a valid Pareto Front.



UTKFace: PaMaL discovers a Pareto Subspace for regression tasks and larger models.

Scene Understanding with CityScapes

SegNet architecture [BKC17] on CityScapes [Cor+16]

PaMaL outperforms most SOTA single-point algorithms, while Pareto Front Approximation methods do not scale or show bias towards one task.

	Segmentation		Depth	
	mIoU ↑	Pix Acc ↑	Abs Err ↓	Rel Err ↓
STL	70.96	92.12	**0.0141**	38.644
LS	70.12	91.90	0.0192	124.061
UW	70.20	91.93	0.0189	125.943
MGDA	66.45	90.79	**0.0141**	53.138
DWA	70.10	91.89	0.0192	127.659
PCGrad	70.02	91.84	0.0188	126.255
IMTL	70.77	92.12	0.0151	74.230
Graddrop	70.07	91.93	0.0189	127.146
CAGrad	69.23	91.61	0.0168	110.139
RLW	68.79	91.52	0.0213	126.942
Nash-MTL	**71.13**	**92.23**	0.0157	78.499
RotoGrad	69.92	91.85	0.0193	127.281
Auto-λ	70.47	92.01	0.0177	116.959
COSMOS	69.78	91.79	0.0539	136.614
PaMaL(ours)	70.35	91.99	**0.0141**	54.520

BKC17] V. Badrinarayanan et al. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2017).
[Cor+16] M. Cordts et al. "The Cityscapes Dataset for Semantic Urban Scene Understanding". In: IEEE Conference on Computer Vision and Pattern Recognition. 2016.
[Kur+22] V. Kurin et al. "In defense of the unitary scalarization for deep multi-task learning". In: Advances in Neural Information Processing Systems. 2022.
[Xin+22] D. Xin et al. "Do Current Multi-Task Optimization Methods in Deep Learning Even Help?" In: Advances in Neural Information Processing Systems. 2022.

The work of Nikolaos Dimitriadis was supported by Swisscom (Switzerland) AG.