Multi-task Learning with 3D-Aware Regularization

github.com/VICO-UoE/3DAwareMTL

Wei-Hong Li

Steven Mcdonagh

Ales Leonardis

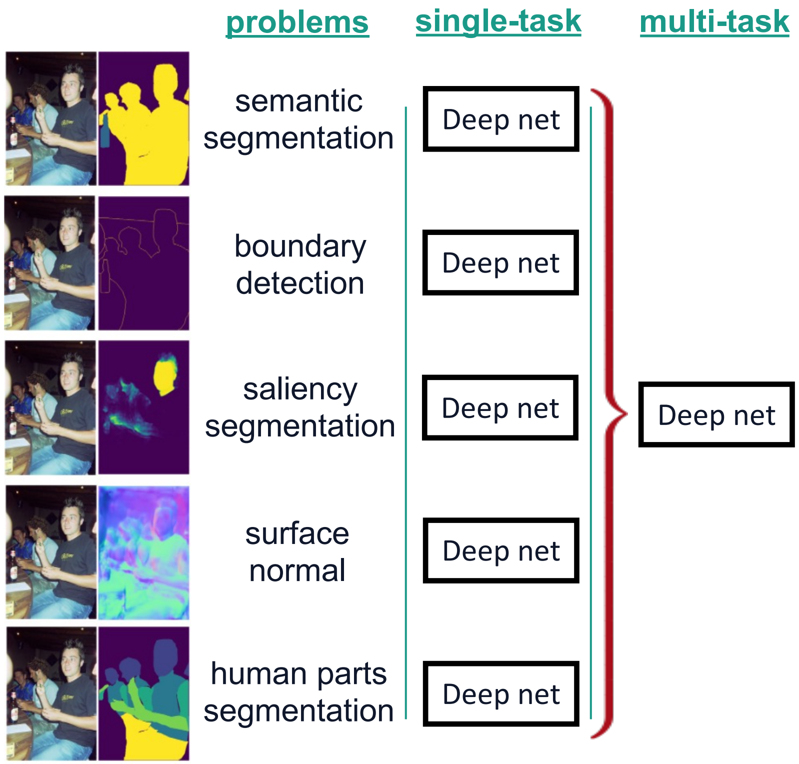
Hakan Bilen

Multi-task Learning with 3D-Aware Regularization

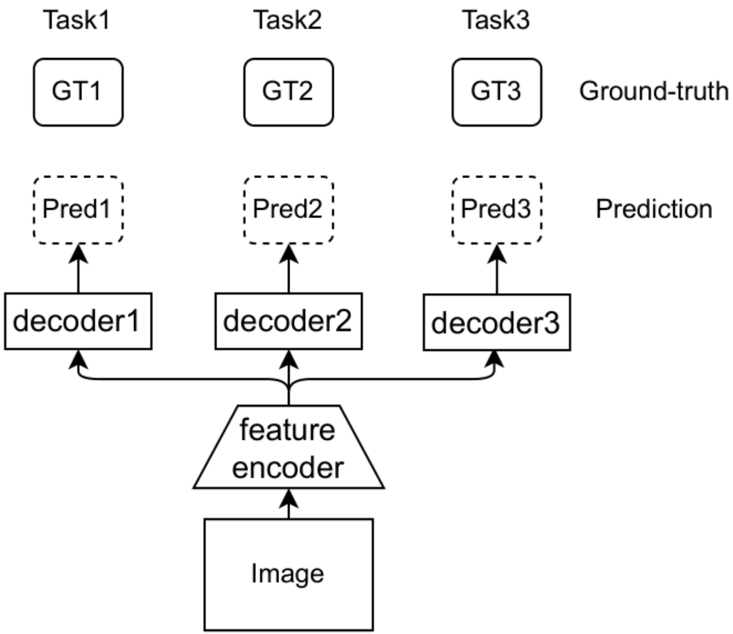
In computer vision, a longstanding goal is to produce **broad and general-purpose systems** that work well on a wide range of vision problems.

Benefits over learning a single network per problem

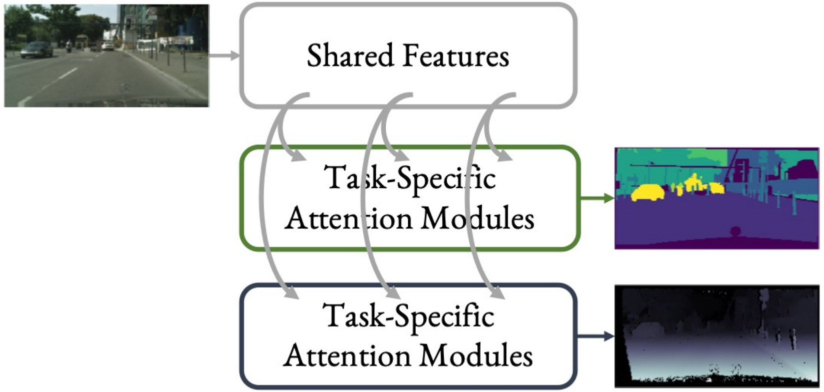
- More complete understanding of the world
- Shared computations and higher efficiency
- Knowledge transfer between tasks
- Efficiently adapted/transferred to new/unseen tasks with few labelled samples



Previous Multi-task Learning Methods

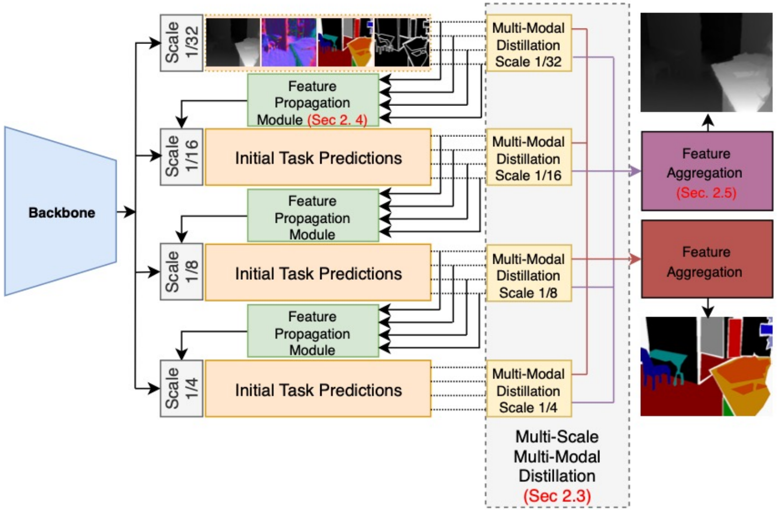


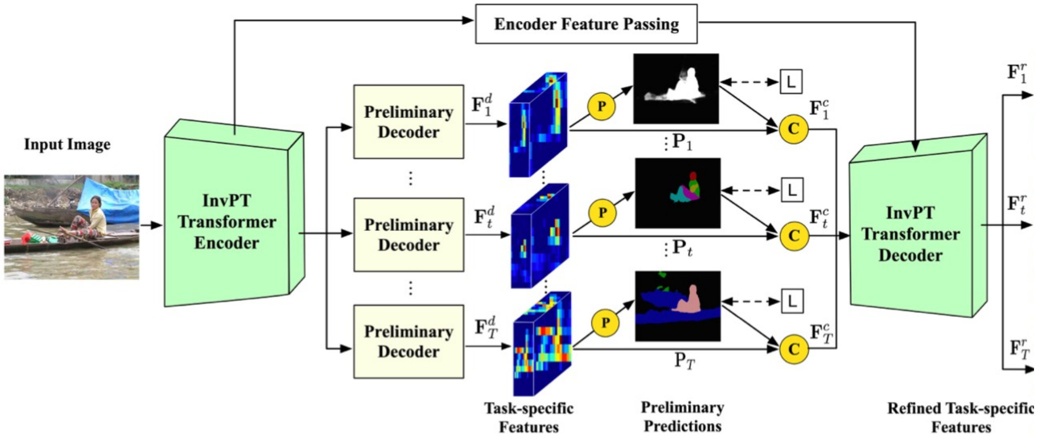
Vanilla MTL: Sharing all parameters of encoder for all tasks (Hard sharing)



MTAN([Liu et al., 2019]): Soft attention mechanism

Previous Multi-task Learning Methods





MTI-Net (Vandenhende et al., 2020): cross-task relations from the multi-scale features

InvPT (Ye et al., 2022): long-range spatial correlations across tasks via vision transformer module

Multi-task Learning with 3D-Aware Regularization

Limitations: high-dimensional and unstructured features, shared across tasks, are prone to capturing noisy cross-task correlations and hence hurt performance.

Our goal: Regulating the feature space of shared representations by introducing a structure that is valid for all considered tasks

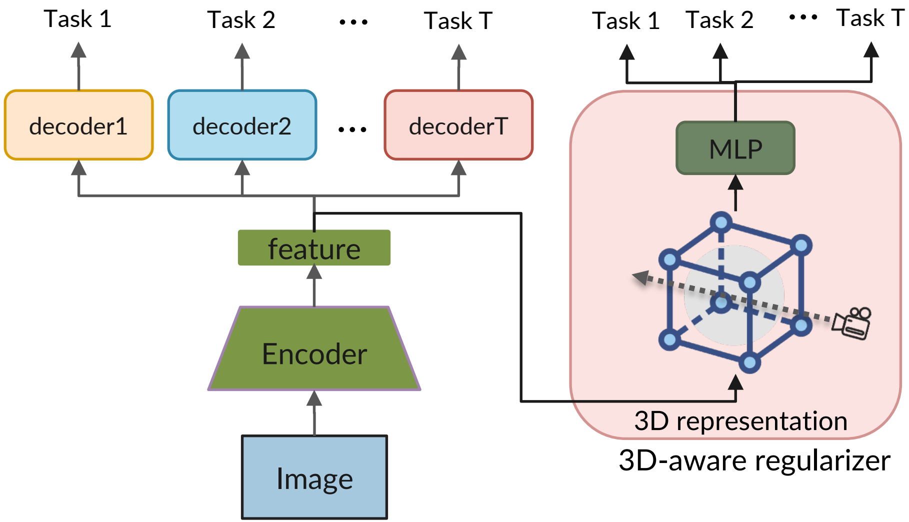
Contributions:

- Introduce novel 3D-aware representations and bottleneck via geometry for structuring the feature space via geometry to eliminate the noisy cross-task correlations.
- The regularizer is model-agnostic and can be incorporated with previous state-of-the-art methods and improve in all tasks without increasing the inference computational cost

Our method

We look at **dense prediction computer vision problems, e.g., segmentation, depth estimation**

- We map an image to a shared feature and decode the shared feature to task-specific predictions
- We map the shared feature and transform it as a triplane and render different tasks predictions
- The 3D space inherently provide a structure space where inconsistency across tasks can be eliminated
- We only require single view as we do not focus on 3D reconstruction and depth gt is available
- Our method can be easily extended to leverage multiple views' data



Quantitative results on NYU-v2

NYU-v2 (Silberman et al., 2012)

- Indoor images for **semantic segmentation**, **depth estimation**, **surface normal estimation** and **boundary detection**
- **our method are plugged into two SotAs (MTI-Net (Vandenhende et al., 2020) and InvPT (Ye et al., 2022)) with different backbones and improve their performance over all tasks**

Method	Seg. (mIoU) ↑	Depth (RMSE) ↓	Normal (mErr) ↓	Boundary (odsF) ↑
Cross-Stitch (Misra et al., 2016)	36.34	0.6290	20.88	76.38
PAP (Zhang et al., 2019)	36.72	0.6178	20.82	76.42
PSD (Zhou et al., 2020)	36.69	0.6246	20.87	76.42
PAD-Net (Xu et al., 2018)	36.61	0.6270	20.85	76.38
ATRC (Bruggemann et al., 2021)	46.33	0.5363	20.18	77.94
MTI-Net (Vandenhende et al., 2020b)	45.97	0.5365	20.27	77.86
Ours	**46.67**	**0.5210**	**19.93**	**78.10**
InvPT (Ye & Xu, 2022a)	53.56	0.5183	19.04	78.10
Ours	**54.87**	**0.5006**	**18.55**	**78.30**

image segmentation depth normal boundary

THE UNIVERSITY of EDINBURGH

ICLR

Thanks for Listening!

Multi-task Learning with 3D-Aware Regularization

github.com/VICO-UoE/3DAwareMTL

Wei-Hong Li

Steven Mcdonagh

Ales Leonardis

Hakan Bilen