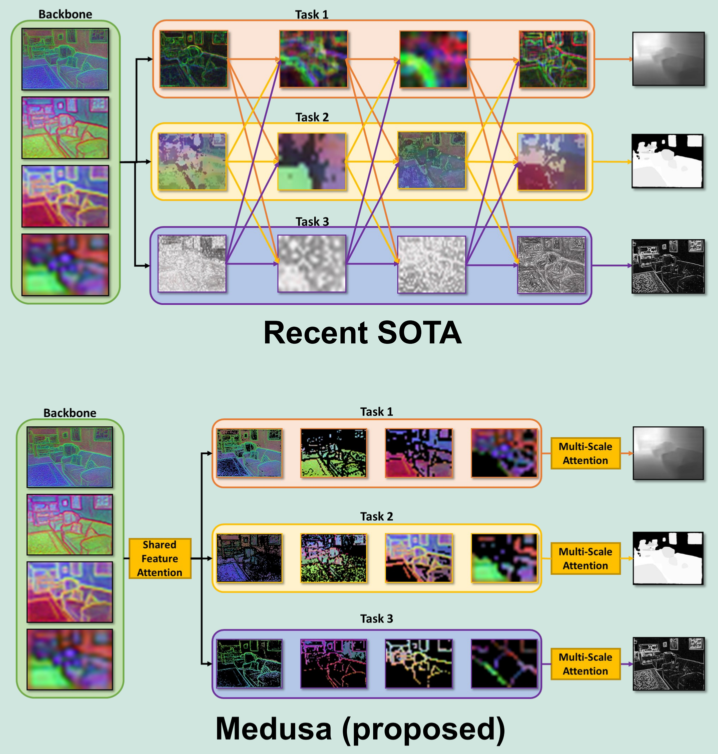
Medusa: Universal Feature Learning via Attentional Multitasking

Jaime Spencer Martin (j.spencermartin@surrey.ac.uk) **Richard Bowden (r.bowden@surrey.ac.uk)** **Simon Hadfield (s.hadfield.@surrey.ac.uk)**

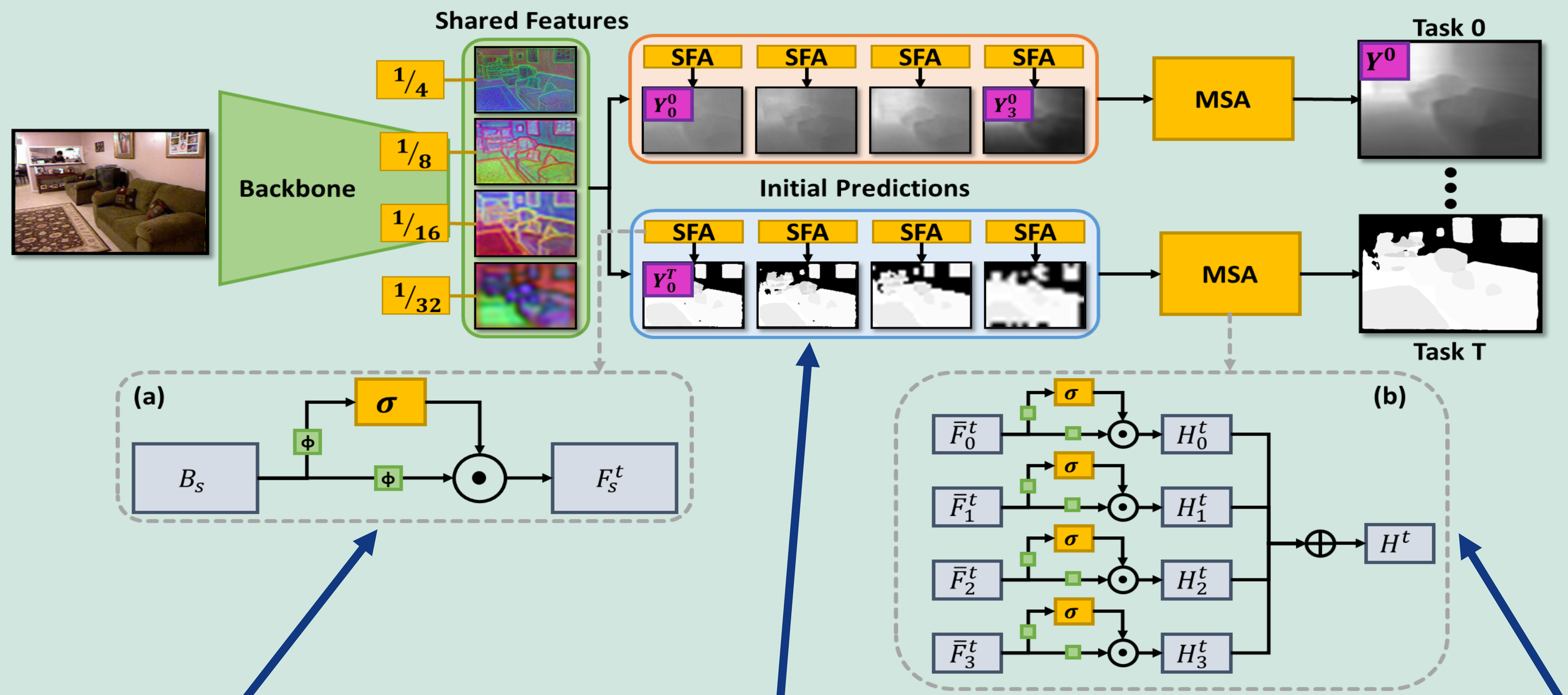
Centre for Vision Speech and Signal Processing, Faculty of Engineering and Physical Sciences, University Of Surrey, Guildford, Surrey, GU2 7XH

1. Overview

- We present **Medusa**, an efficient approach to learning **generic features for multitasking**
- Recent multi-task learning (**MTL**) networks model connections between all pairs of tasks
  - ♦ Requires **retraining** when adding tasks... ♦ **Quadratic** complexity...
- Objective should be learning generic features capable of **adapting to tasks** not seen during training
  - ♦ **Universal Feature Learning (UFL)**
- Medusa uses **dual spatial attention** mechanisms
  - ♦ **Improved generalization** capabilities! ♦ **Linear** complexity!



2. Methodology



Shared Feature Attention

- Connects multi-scale shared backbone to each task head
- **Backbone learns generic features** for all tasks
- Each **task & scale** filters relevant features via **spatial attention**
- **Alleviates negative transfer** between unrelated tasks

Multi-scale Initial Predictions

- Each task makes an **initial prediction at each backbone scale**
- **Guides each scale** to produce useful features for the final task
- **Discarded during evaluation** for increased efficiency

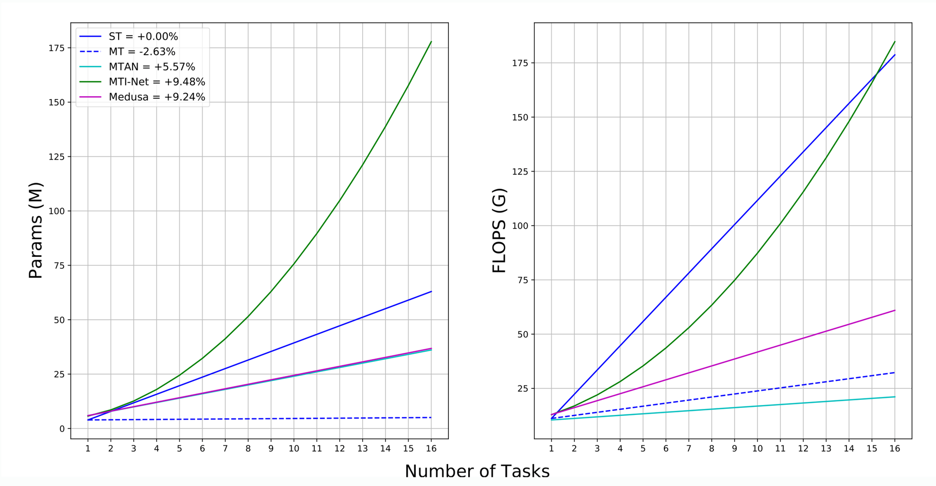
Multi-scale Attention Task Head

- Combines **multi-scale features** from each task prior to making final prediction
- Each scale has **different roles**, trading **global consistency** and **accurate boundaries**
- Spatial attention lets network **learn how to best combine features from each scale**

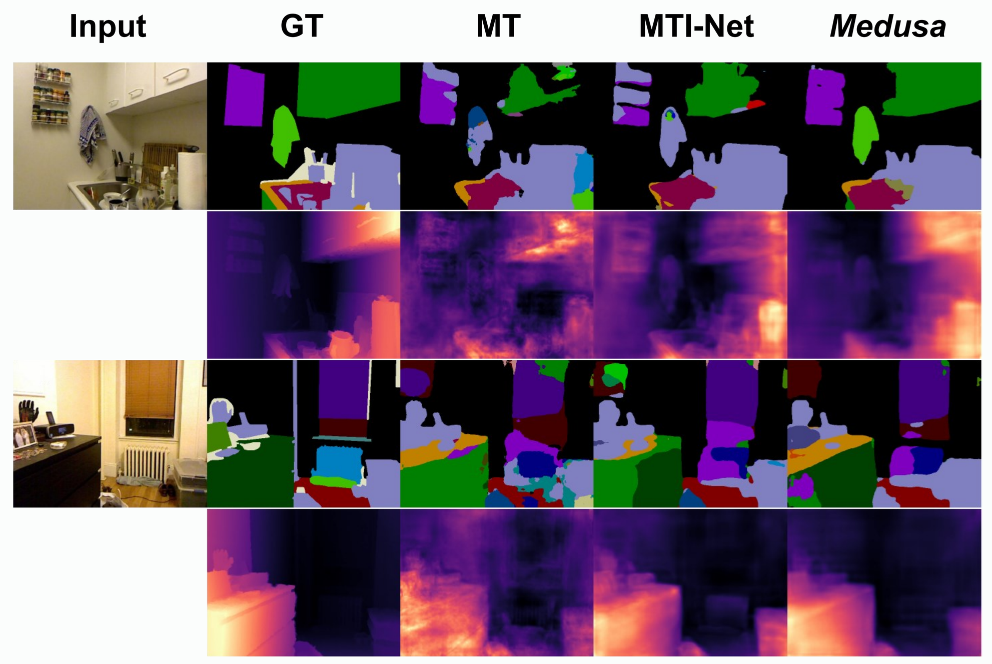
3. Results

Resource Usage

- **Independent task heads** results in higher efficiency with **linear scaling**
- Quadratic scaling (**MTI-Net**) is **less efficient** that training a **separate network for each task**



Visualizations



Multi-task Learning

- Train on **NYUD** using edges & normal as auxiliary
- Δ_m% reports **average performance improvement** across all tasks
- **Medusa** is **comparable to MTI-Net** while being **more efficient**

	Backbone	Head	N+E	Seg ↑	Depth ↓	Δ_m% ↑
ST Baseline	HRNet-18	HRHead		34.57	0.606	+0.00
MT Baseline	HRNet-18	HRHead		33.21	0.614	-2.63
MTAN [23]	HRNet-18	DeepLab-v3+		35.25	0.581	+3.02
MTAN	HRNet-18	DeepLab-v3+	✓	36.19	0.567	+5.57
PAD-Net [43]	HRNet-18	HRHead		34.39	0.617	-1.23
PAD-Net	HRNet-18	HRHead	✓	35.46	0.604	+1.43
MTI-Net [42]	HRNet-18	HRHead		36.94	0.559	+7.26
MTI-Net	HRNet-18	HRHead	✓	37.40	**0.540**	**+9.48**
***Medusa* (ours)**	HRNet-18	**MSA (ours)**		36.99	0.573	+6.19
Medusa	HRNet-18	**MSA**	✓	**37.48**	0.545	+9.24

Universal Feature Learning

- Use **frozen pretrained backbones** from NYUD
- Train **only task heads** on PASCAL
- **Medusa's features transfer better** to unseen tasks
- Since backbone is frozen, **previous tasks are not forgotten**

	NYUD-v2			PASCAL-Context		
	Seg ↑	Depth ↓	Δ_m% ↑	Parts ↑	Sal ↑	Δ_m% ↑
ST Baseline	34.57	0.606	+0.00	48.73	56.44	+0.00
MT Baseline	33.21	0.614	-2.63	36.13	51.96	-12.93
MTAN [23]	36.19	0.567	+5.57	47.37	57.84	+4.26
MTI-Net [42]	37.40	**0.540**	**+9.48**	51.50	60.19	+10.76
Medusa	**37.48**	0.545	+9.24	**52.24**	**61.91**	**+13.18**

4. Conclusions

- We introduced **Medusa** and showed its effectiveness in both **MTL** and **UFL**
- Focus on **independent task heads**
  - More **efficient** network (parameters & FLOPS)
  - Improved **feature generalization**
- Improve on previous models with independent task heads by introducing **dual attention mechanisms**
  - **MTL** performance is **comparable with current SOTA**, while providing **25% parameter reduction** and **linear scaling**
  - **UFL** performance improves on **SOTA** due to focus on shared features

Engineering and Physical Sciences Research Council

This work was partially funded by the EPSRC under grant agreements (EP/R512217/1, EP/S016317/1 and EP/S-35761/1).

Centre for Vision, Speech and Signal Processing

UNIVERSITY OF SURREY