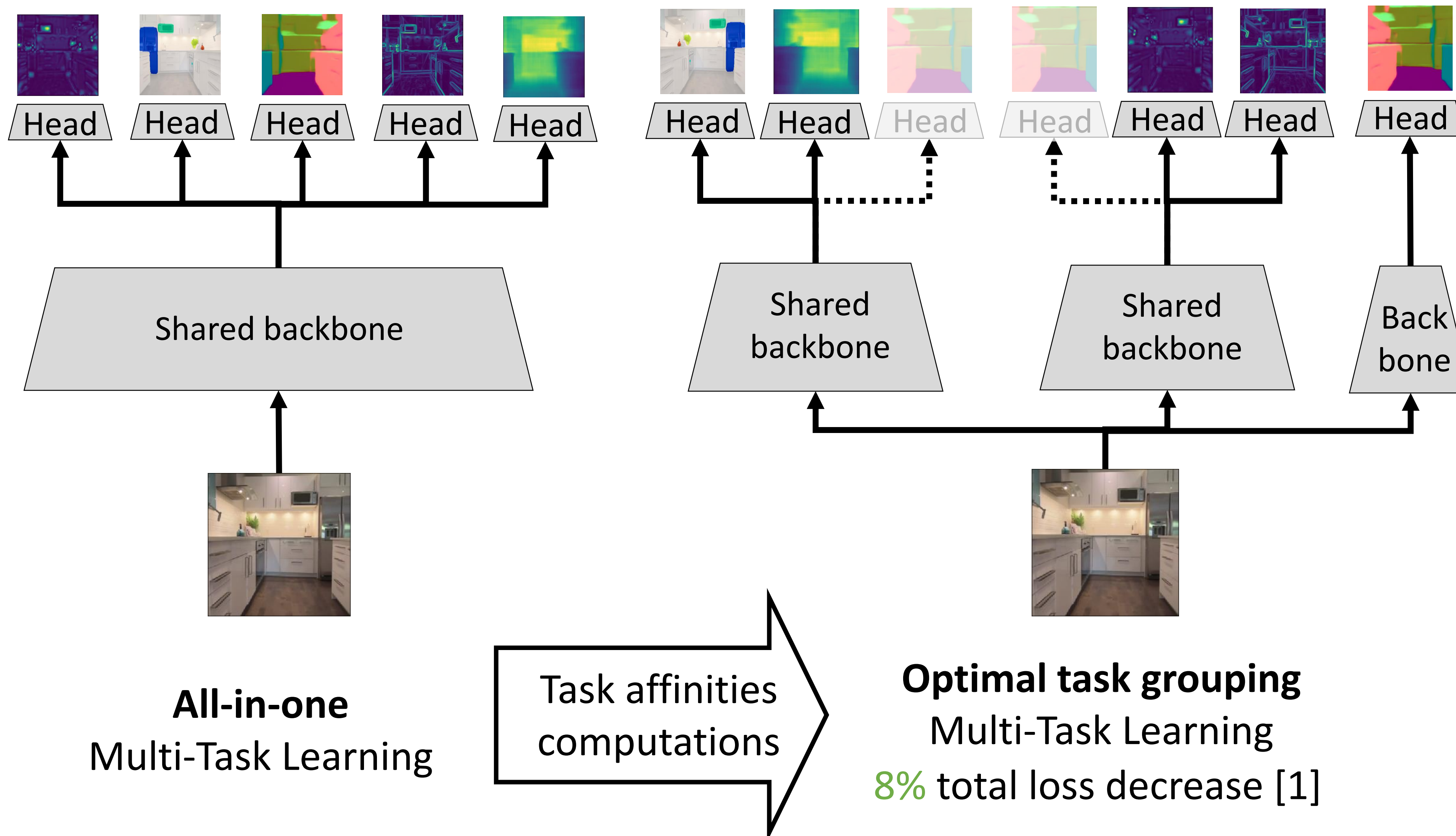


## Motivation

Grouping tasks by affinity  
is paramount to Multi-Task Learning success



## Problem statement

How to compute task affinities?

Exhaustive search through  
all possible combinations of tasks  
→ accurate but prohibitive cost

Approximations using scoring techniques

- Based on the tasks themselves
- Based on cheaper single-task models
- Based on cheaper multi-task models  
→ maybe inaccurate but cheap

Trade-offs to characterize



Methodology

Benchmark various  
task affinity scoring techniques

## Methodology

Benchmark pair-wise task affinity scoring techniques

- **Computer Vision dataset** composed of +700K images and 5 tasks from Taskonomy [2]
  - **6 task affinity scoring techniques** benchmarked:
    - TD: Taxonomical distance [Zamir et al., CVPR ‘18]
    - IAS: Input attribution similarity [Song et al., NIPS’19]
    - RSA: Representation similarity analysis [Dwivedi et al., CVPR’19]
    - GS: Gradient similarity [Zhao et al., ECCV’18]
    - GT: Gradient transference [Fifty et al., NIPS’21]
    - **LI: Label injection [ours]**
  - **4 levels of evaluation** for each scoring:
    - Predictive power
    - Partner tasks ranking
    - Best partner task identification
    - Computational cost
- Affinity scoring accuracy  
→ Affinity scoring cost

## Results

Label injection (LI) is more indicative than other scores  
by being able to identify the best partner for a given target task

Task	Best partner	TD	IAS	RSA	LI	GS	GT
SemSeg	Normal	Normal (+50%)	Normal (+50%)	Depth (+18%)	Normal (+50%)	Depth (+18%)	Depth (+18%)
Keypts	Normal	Edges (+1%)	Edges (+1%)	Edges (+1%)	Normal (+30%)	Edges (+1%)	Edges (+1%)
Edges	Normal	Keypts (-9%)	Keypts (-9%)	Keypts (-9%)	Normal (+78%)	Keypts (-9%)	Keypts (-9%)
Depth	Normal	Normal (0%)	SemSeg (-1%)	Normal (0%)	Normal (0%)	Normal (0%)	SemSeg (-1%)
Normal	Edges	SemS./Depth (-3%)	SemSeg (+1%)	Depth (-6%)	Depth (-6%)	Depth (-6%)	Depth (-6%)

Evaluation level: best partner identification.

How to read: For the target task Keypts the actual best partner is Normal.

All scores but **Label injection (LI)** select Edges, leading to only +1% performance gain instead of +30% if Normal was chosen.

