" It's a Match! "
A Benchmark of Task Affinity Scores for Joint Learning

Supplementary material

Anonymous submission

Abstract

The following elements are provided in the supplementary material:

- Affinity scores raw values
- Taskonomy buildings used
- Affinity scores computation details

1 Affinity scores raw values

In Tables 1 to 6, we report the raw affinities estimations for all tasks, using each affinity scoring. Results are rounded at the second decimal.

with	Affinities estimations SemSeg	Keypts	Edges	Depth	Normal
SemSeg	-	-8	-6	-8	-5
Keypts	-8	-	-4	-12	-9
Edges	-6	-4	-	-10	-7
Depth	-8	-12	-10	-	-5
Normal	-5	-9	-7	-5	-

Table 1: **Taxonomical distance (TD)** Distance between tasks in the similarity tree from (Zamir et al. 2018). Multiplied by -1 for consistency with the other affinity scores (i.e., higher means stronger affinity).

with	Affinities estimations SemSeg	Keypts	Edges	Depth	Normal
SemSeg	-	0.25	0.23	0.31	0.45
Keypts	0.25	-	0.52	0.18	0.23
Edges	0.23	0.52	-	0.18	0.22
Depth	0.31	0.18	0.18	-	0.29
Normal	0.45	0.23	0.22	0.29	-

Table 2: **Input attribution similarity (IAS)** Cosine similarity between STL models attribution maps.

with	Affinities estimations SemSeg	Keypts	Edges	Depth	Normal
SemSeg	-	0.33	0.37	0.46	0.46
Keypts	0.33	-	0.66	0.04	0.05
Edges	0.37	0.66	-	0.12	0.13
Depth	0.46	0.04	0.12	-	0.69
Normal	0.46	0.05	0.13	0.69	-

Table 3: **Representation similarity (RSA)** Representation similarity analysis using the STL models backbones output.

with	Affinities estimations SemSeg	Keypts	Edges	Depth	Normal
SemSeg	-	-2.93%	-3.50%	-3.07%	+3.70%
Keypts	-8.47%	-	+4.97%	-8.31%	+1.42%
Edges	-15.93%	-4.20%	-	-9.58%	+2.42%
Depth	+4.04%	-3.34%	-1.26%	-	+20.29%
Normal	+25.68%	+60.29%	+23.79%	+66.30%	-

Table 4: **Label injection (LI)** Performance gain when incorporating the label from the partner task in the STL model's input, relative to standard STL.

with	Affinities estimations SemSeg	Keypts	Edges	Depth	Normal
SemSeg	-	0.51	0.39	1.93	1.54
Keypts	0.51	-	1.89	0.75	1.0
Edges	0.39	1.89	-	0.92	0.59
Depth	1.93	0.75	0.92	-	8.40
Normal	1.54	1.0	0.59	8.40	-

Table 5: **Gradient similarity (GS)** Cosine similarity between task-specific gradient updates on the MTL backbone. Averaged across all training epochs. Multiplied by 100 for readability.

with	SemSeg	Keypts	Edges	Depth	Normal
	Affinities estimations				
SemSeg	-	+0.02	+0.25	+1.69	+0.74
Keypts	-0.03	-	+0.38	-0.01	+0.01
Edges	-0.20	+0.71	-	+0.19	+0.27
Depth	+0.47	+0.01	+0.15	-	+0.90
Normal	+0.27	+0.03	+0.16	+1.26	-

Table 6: **Gradient transference (GT)** Look-ahead ratio simulating the effect of applying task-specific updates to the MTL backbone for the other task. Averaged across all training epochs.

2 Taskonomy buildings used

We split our subset of the Taskonomy dataset into train, validation and test sets, on a per-building basis.

Train set These buildings amount to 603,437 input images.

- adairsville
- airport
- albertville
- anaheim
- ancor
- andover
- annona
- arkansaw
- athens
- bautista
- bohemia
- bonesteel
- bonnie
- broseley
- browntown
- byers
- scioto
- nuevo
- goodfield
- donaldson
- hanson
- merom
- klickitat
- onaga
- leonardo
- marstons
- newfields
- pinesdale
- lakeville
- cosmos
- benevolence
- pomaria
- tolstoy
- shelbyville
- allensville
- wainscott
- beechwood
- coffeen
- stockman
- hiteman
- woodbine
- lindenwood
- forkland
- mifflinburg
- ranchester
- springerville
- swisshome
- westfield
- willow
- winooski
- hainesburg
- irvine
- pearce
- thrall
- tilghmanton
- uvalda
- sugarville
- silas

Validation set These buildings amount to 82,345 input images.

- corozal
- collierville
- markleeville
- darden
- chilhowie
- churchton
- cauthron
- cousins
- timberon
- wando

Test set These buildings amount to 40,367 input images.

- ihlen
- muleshoe
- noxapater
- mcdade

3 Affinity scores computation

In Table 7, we detail the computation of each selected affinity scoring. Considering two tasks $t_1 = a$ and $t_2 = b$ and a batch of examples $\mathcal{X}$, we denote:

- their resp. losses functions $\mathcal{L}_a$ and $\mathcal{L}_b$
- their resp. STL models STL_a and STL_b with losses
  - $\mathcal{L}_{STL_a} = \mathcal{L}_a(\mathcal{X}, STL_a)$
  - $\mathcal{L}_{STL_b} = \mathcal{L}_b(\mathcal{X}, STL_b)$
- their joint MTL model $MTL_{(a,b)}$ with loss $\mathcal{L}_{MTL_{(a,b)}} = \mathcal{L}_a(\mathcal{X}, MTL_{(a,b)}) + \mathcal{L}_b(\mathcal{X}, MTL_{(a,b)})$

Note that if the score is symmetric, it assesses how much the two tasks help each other regardless of direction. If it is asymmetric, it considers how much the target task a benefits from being learned with the partner task b. While all scores could not be constrained to lie in the same range, higher always means more affinity.

References

Dwivedi, K.; and Roig, G. 2019. Representation similarity analysis for efficient task taxonomy & transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12387–12396.

Fifty, C.; Amid, E.; Zhao, Z.; Yu, T.; Anil, R.; and Finn, C. 2021. Efficiently identifying task groupings for multi-task learning. *Advances in Neural Information Processing Systems*, 34: 27503–27516.

Shrikumar, A.; Greenside, P.; Shcherbina, A.; and Kundaje, A. 2016. Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713*.

Song, J.; Chen, Y.; Wang, X.; Shen, C.; and Song, M. 2019. Deep model transferability from attribution maps. *Advances in Neural Information Processing Systems*, 32.

Zamir, A. R.; Sax, A.; Shen, W.; Guibas, L. J.; Malik, J.; and Savarese, S. 2018. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3712–3722.

Zhao, X.; Li, H.; Shen, X.; Liang, X.; and Wu, Y. 2018. A modulation module for multi-task learning with applications in image retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 401–416.

Affinity scoring	Type	Computation	Comment	Range
Taxonomical distance (TD)	Model-agnostic	Distance between tasks in a taxonomy tree.	Symmetric. Taxonomy borrowed from (Zamir et al. 2018). Multiplied by -1 for consistency i.e., higher is better.	$]-\infty, 0]$
Input attribution similarity (IAS)	STL-based	$$\frac{1}{\lvert\mathcal{X}\rvert}\sum_{x\in\mathcal{X}} S_{cos}(Attr(STL_a, x), Attr(STL_b, x)), \quad (1)$$ where S_{cos} is cosine similarity, $\mathcal{X}$ denotes a batch of examples and $Attr$ the attribution method used.	Symmetric. Revisited from (Song et al. 2019). Computed on a subset of the test set (2,048 images) using $InputXGradient$ attribution (Shrikumar et al. 2016).	$[-1, +1]$
Representation similarity (RSA)	STL-based	$$RSA(\theta_{Ba}, \theta_{Bb}, \mathcal{X}), \quad (2)$$ where RSA denotes the Representation Similiarity Analysis, $\mathcal{X}$ a batch of examples, θ_{Ba} and θ_{Bb} the backbone weights of the STL models for tasks a and b resp.	Symmetric. Revisited from (Dwivedi and Roig 2019). Computed on a subset of the test set (2,048 images).	$[-1, +1]$
Label injection (LI)	STL-based	$$\frac{\mathcal{L}_{STL_a} - \mathcal{L}_{STL_{a\leftarrow b}}}{\mathcal{L}_{STL_{a\leftarrow b}}}, \quad (3)$$ where $STL_{a\leftarrow b}$ represents the STL model for task a, modified to ingest the corresponding label from task b in addition to the input.	Asymmetric. Computed using test losses. Novel proposal.	$]-\infty, +\infty[$
Gradients similarity (GS)	MTL-based	$$\frac{1}{N}\sum_{i=1}^{N} S_{cos}\left(\frac{\partial\mathcal{L}_a(\mathcal{X}, \theta_B^i, \theta_{Ha}^i)}{\partial\theta_B^i}, \frac{\partial\mathcal{L}_b(\mathcal{X}, \theta_B^i, \theta_{Hb}^i)}{\partial\theta_B^i}\right), \quad (4)$$ where N denotes the number of training epochs, S_{cos} the cosine similarity, $\mathcal{X}$ a batch of examples, θ_B^i the weights of the common MTL backbone at the i^{th} epoch, θ_{Ha}^i and θ_{Hb}^i the weights of the heads for a and b at the i^{th} epoch.	Symmetric. Borrowed from (Fifty et al. 2021; Zhao et al. 2018).	$[-1, +1]$
Gradients transference (GT)	MTL-based	$$\frac{1}{N}\sum_{i=1}^{N} 1 - \frac{\mathcal{L}_a(\mathcal{X}, \theta_{B\mid b}^{i+1}, \theta_{Ha}^i)}{\mathcal{L}_a(\mathcal{X}, \theta_B^i, \theta_{Ha}^i)}, \quad (5)$$ where N denotes the number of training epochs, $\mathcal{X}$ a batch of examples, $\theta_{B\mid b}^{i+1}$ the weights of the common MTL backbone updated using the loss of task b at the epoch $i+1$, θ_{Ha}^i and θ_{Hb}^i the weights of the heads for a and b at the i^{th} epoch.	Asymmetric. Borrowed from (Fifty et al. 2021).	$]-\infty, +\infty[$

Table 7: Tasks affinity scores description and computation considering two tasks $t_1 = a$ and $t_2 = b$