

CVPR 2018 Best Paper Award

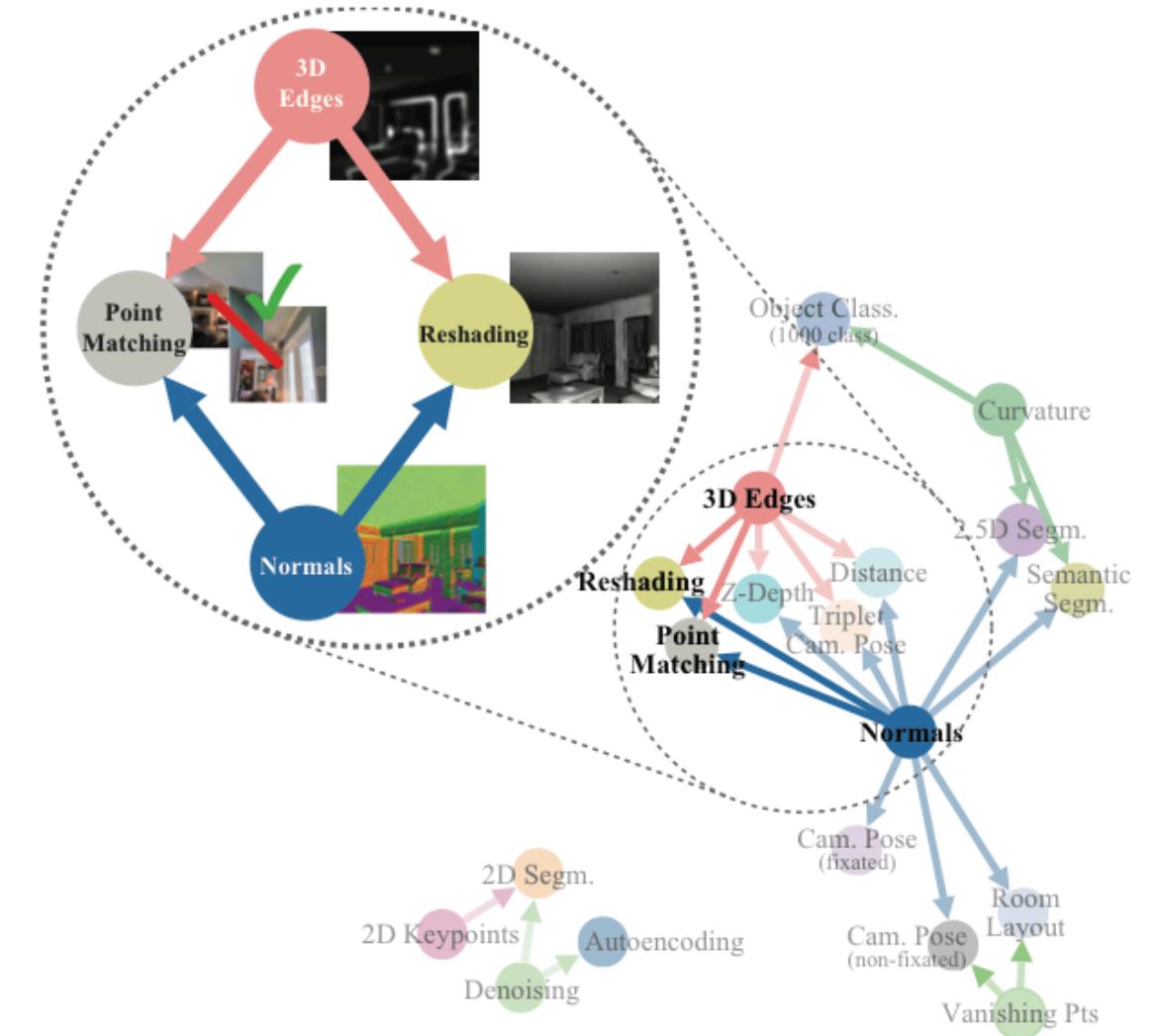
Taskonomy: Disentangling Task Transfer Learning

Amir R. Zamir^{1,2} Alexander Sax^{1*} William Shen^{1*} Leonidas Guibas¹ Jitendra Malik² Silvio Savarese¹

¹ Stanford University ² University of California, Berkeley

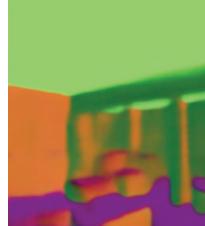
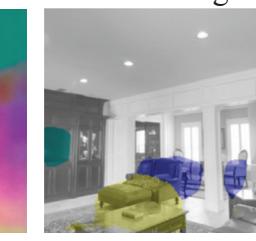
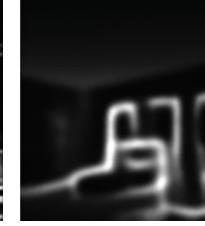
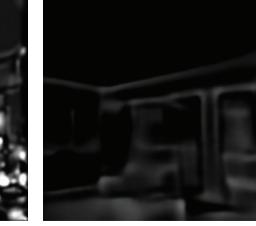
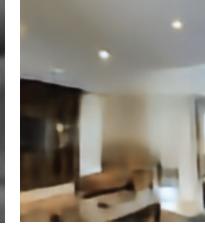
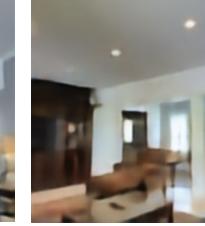
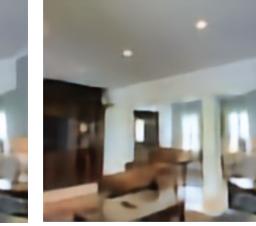
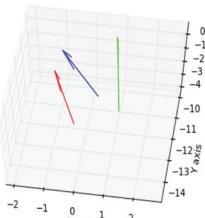
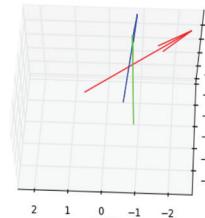
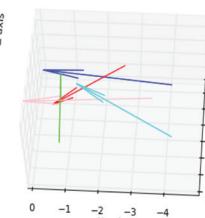
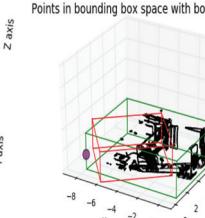
<http://taskonomy.vision/>

Are vision tasks related?

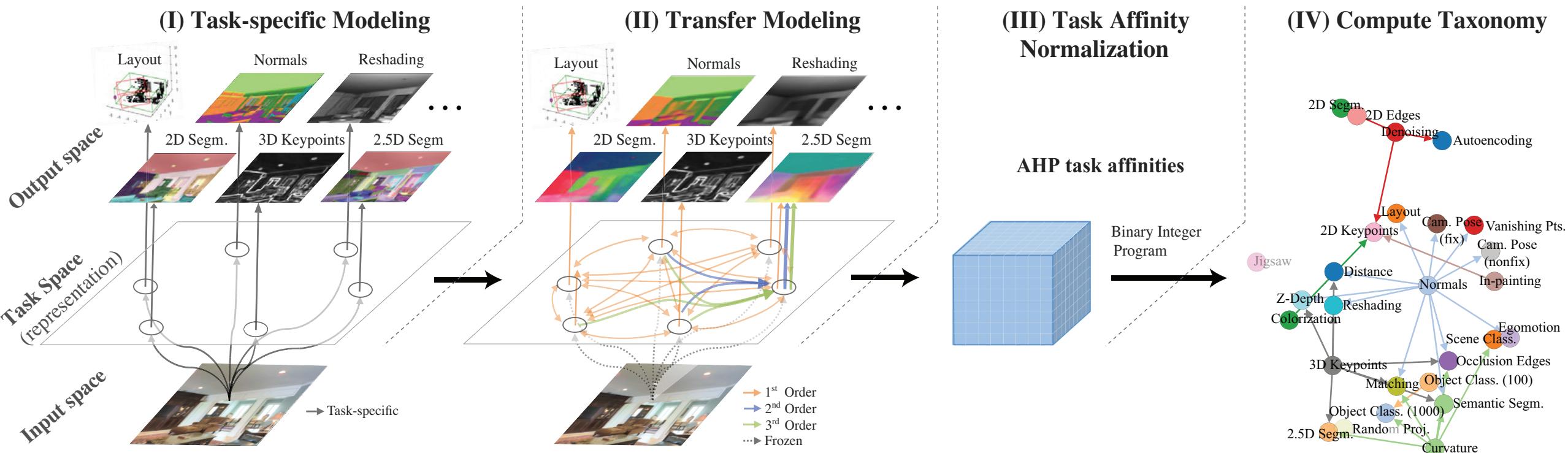


Task Dictionary

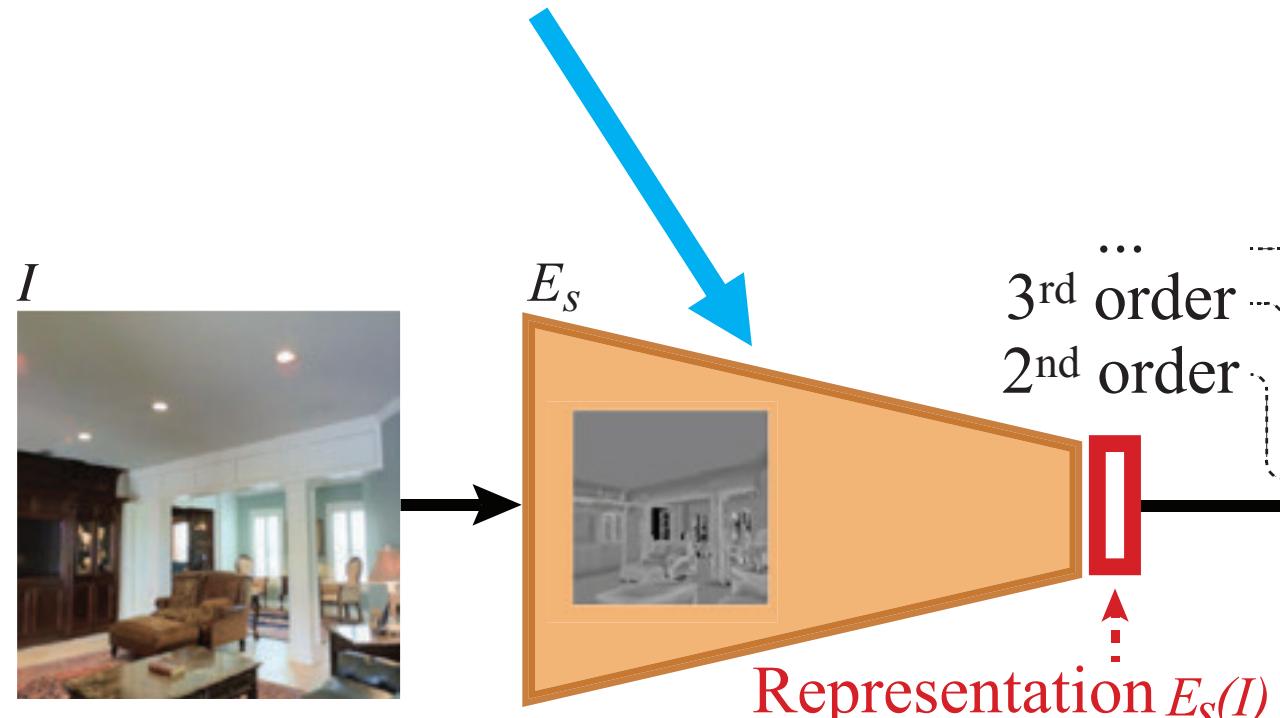
26 tasks (24 shown here)

Query Image	Surface Normals	Eucl. Distance	Object Class.	Scene Class.
			Top 5 prediction: <ul style="list-style-type: none"> • sliding door • home theater, home theatre • studio couch, day bed • china cabinet, china closet • entertainment center 	Top 2 prediction: <ul style="list-style-type: none"> • living room • television room
Jigsaw puzzle	Colorization	2D Segm.	2.5D Segm.	Semantic Segm.
				
Vanishing Points	2D Edges	3D Edges	2D Keypoints	3D Keypoints
				
3D Curvature	Image Reshading	In-painting	Denoising	Autoencoding
				
Cam. Pose _(non-fixed)	Cam. Pose _(fixed)	Triplet Cam. Pose	Room Layout	Point Matching
			Points in bounding box space with bounding box 	

Process Overview



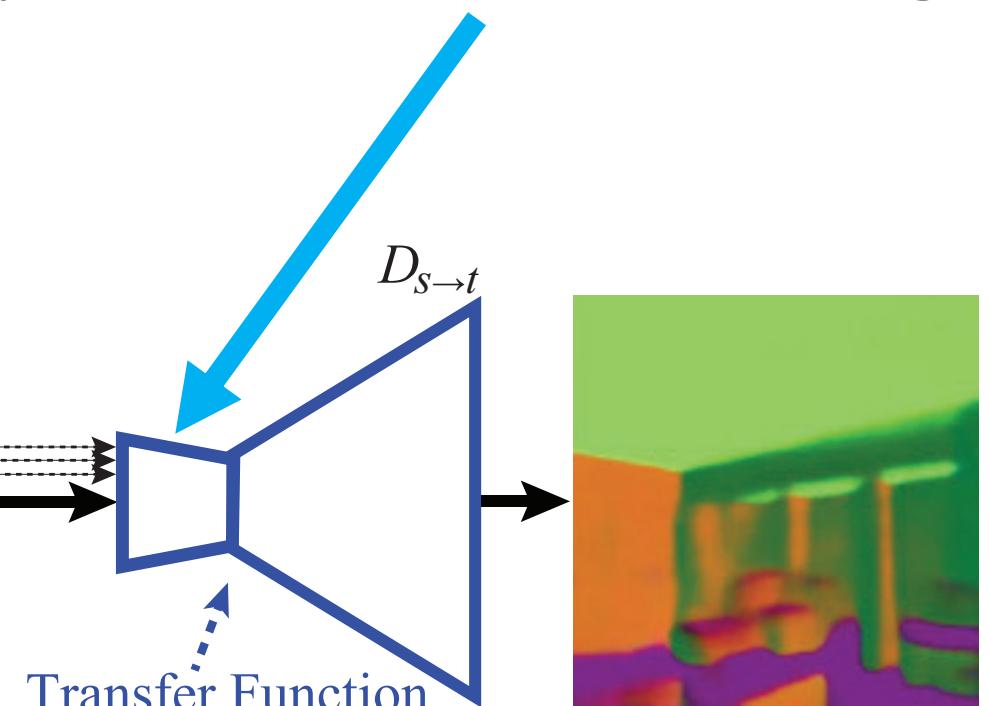
Step I: Task-Specific Modeling



Frozen

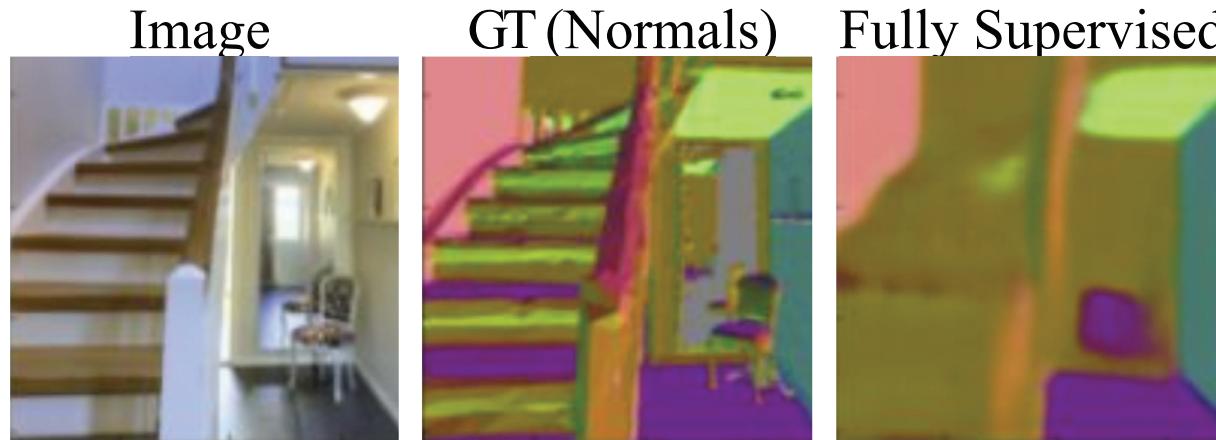
Source Task Encoder
(e.g., curvature)

Step II: Transfer Modeling

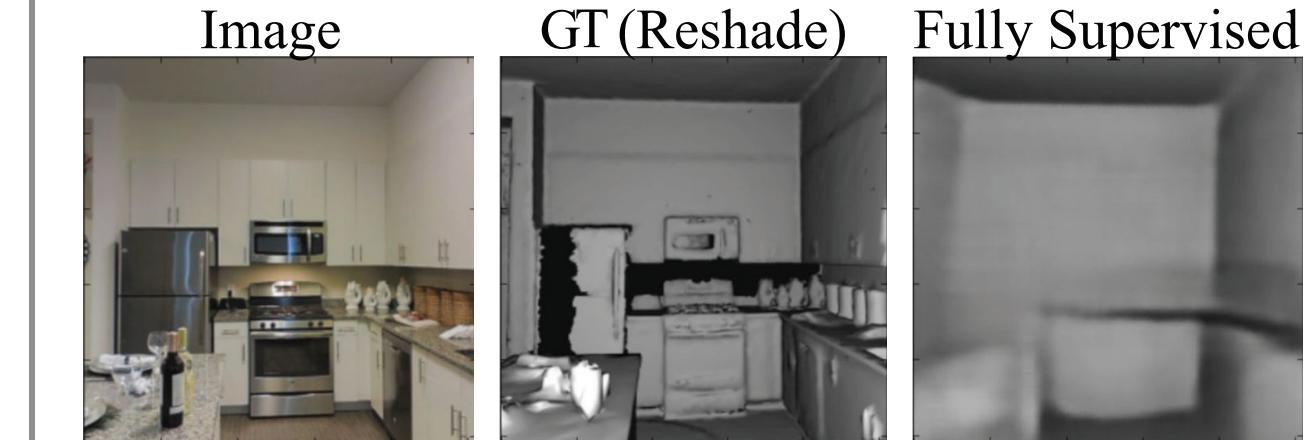
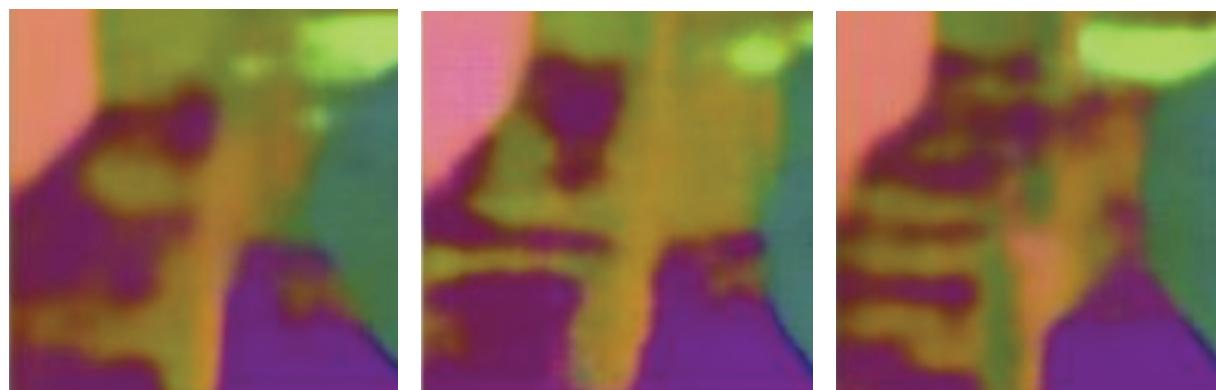


Target Task Output
(e.g., surface normal)

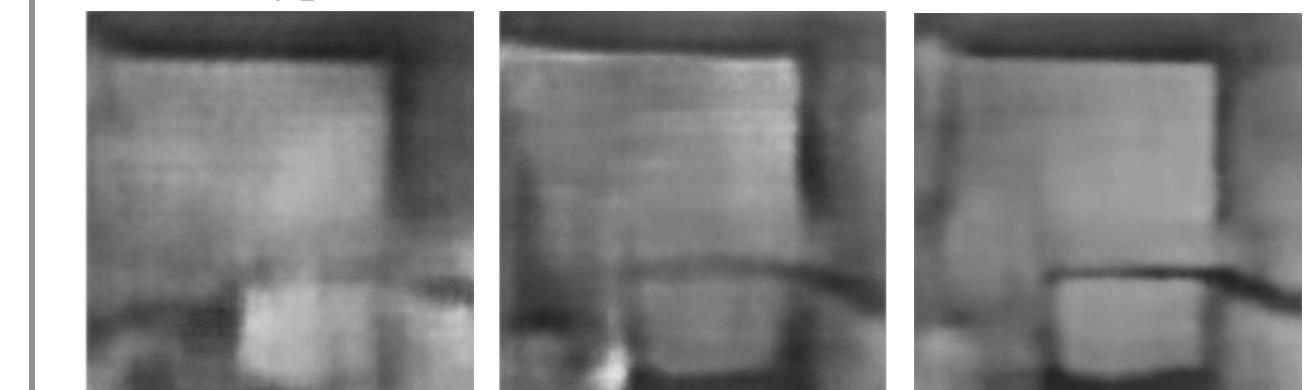
2nd Order Transfer Example



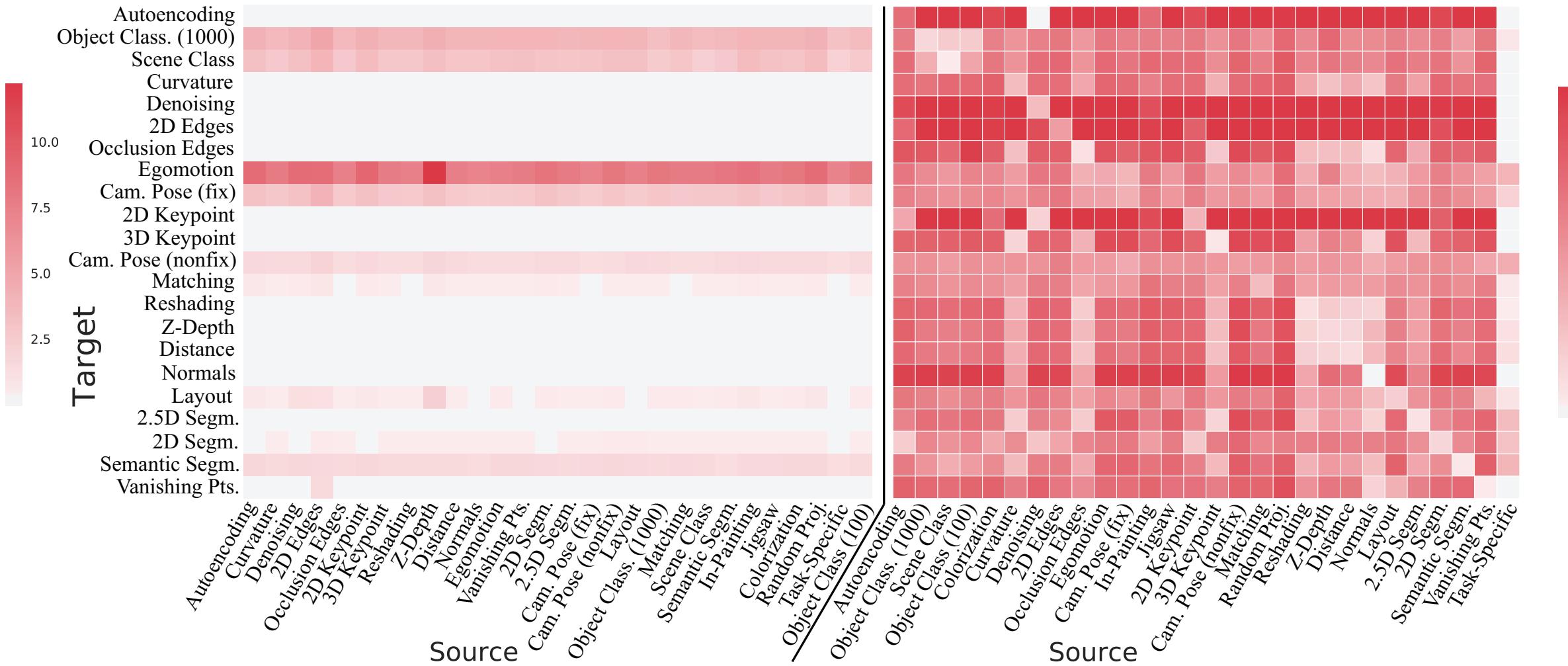
{Occlusion Edges + Curvature } = 2nd order transfer



{ 3D Keypoints + Surface Normals } = 2nd order transfer



Step III: Ordinal Normalization using Analytic Hierarchy Process



Step IV: Computing the Global Taxonomy

- Devise a global transfer policy to maximize collective performance across all tasks, while minimizing the used supervision
- Formulated as subgraph selection where tasks are nodes and transfers are edges
- Solved by using Boolean Integer Programming (BIP)

Experiments

<http://taskonomy.stanford.edu/api/>

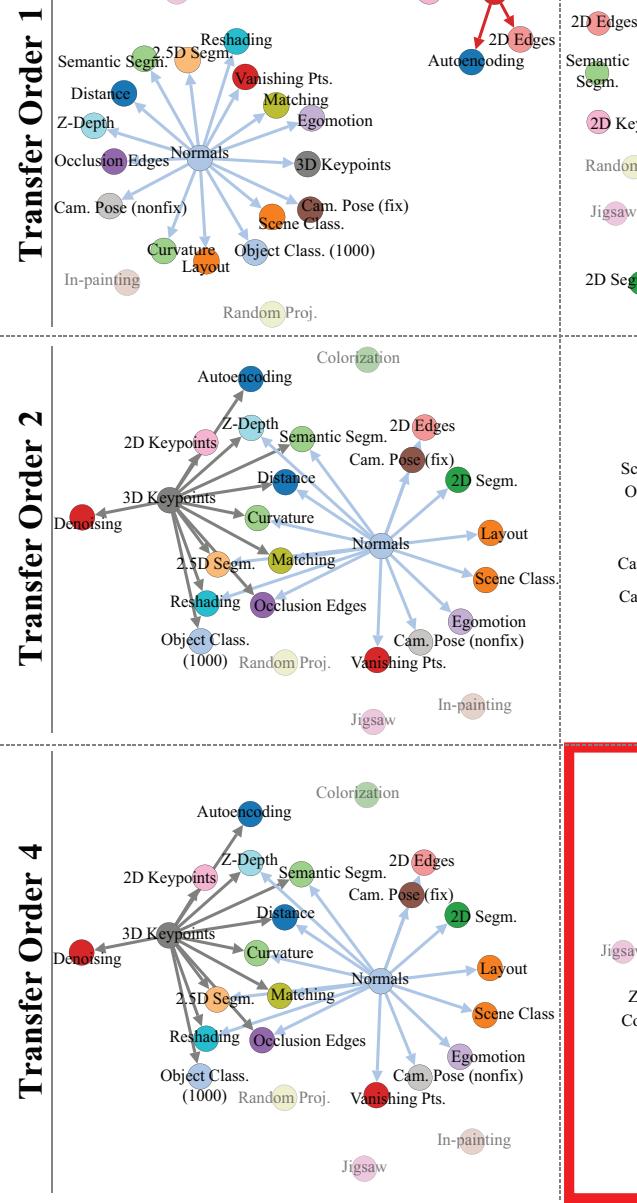
With 26 tasks in the dictionary (4 source-only tasks), our approach leads to training 26 fully supervised task-specific networks, 22×25 transfer networks in 1st order, and $22 \times \binom{25}{k}$ for k^{th} order, from which we sample according to the procedure in Sec. 3. The total number of transfer functions trained for the taxonomy was $\sim 3,000$ which took 47,886 GPU hours on the cloud.

Network Architectures: We preserved the architectural and training details across tasks as homogeneously as possible to avoid injecting any bias. The **encoder** architecture is identical across all task-specific networks and is a fully convolutional ResNet-50 without pooling. All **transfer** functions include identical shallow networks with 2 conv layers (concatenated channel-wise if higher-order). The loss (L_t) and **decoder**'s architecture, though, have to depend on the task as the output structures of different tasks vary; for all

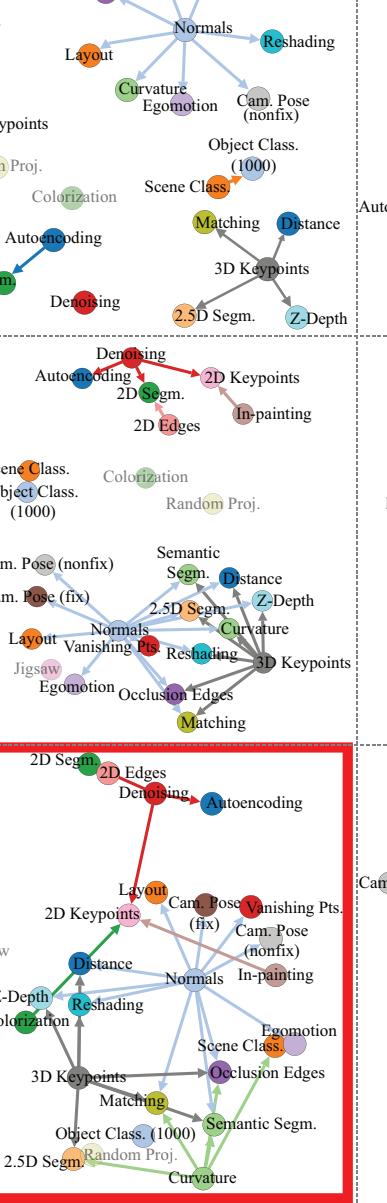
Single task performance

Task	avg	rand	Task	avg	rand	Task	avg	rand
Denoising	100	99.9	Layout	99.6	89.1	Scene Class.	97.0	93.4
Autoenc.	100	99.8	2D Edges	100	99.9	Occ. Edges	100	95.4
Reshading	94.9	95.2	Pose (fix)	76.3	79.5	Pose (nonfix)	60.2	61.9
Inpainting	99.9	-	2D Segm.	97.7	95.7	2.5D Segm.	94.2	89.4
Curvature	78.7	93.4	Matching	86.8	84.6	Egomotion	67.5	72.3
Normals	99.4	99.5	Vanishing	99.5	96.4	2D Keypnt.	99.8	99.4
Z-Depth	92.3	91.1	Distance	92.4	92.1	3D Keypnt.	96.0	96.9
Mean	92.4	90.9						

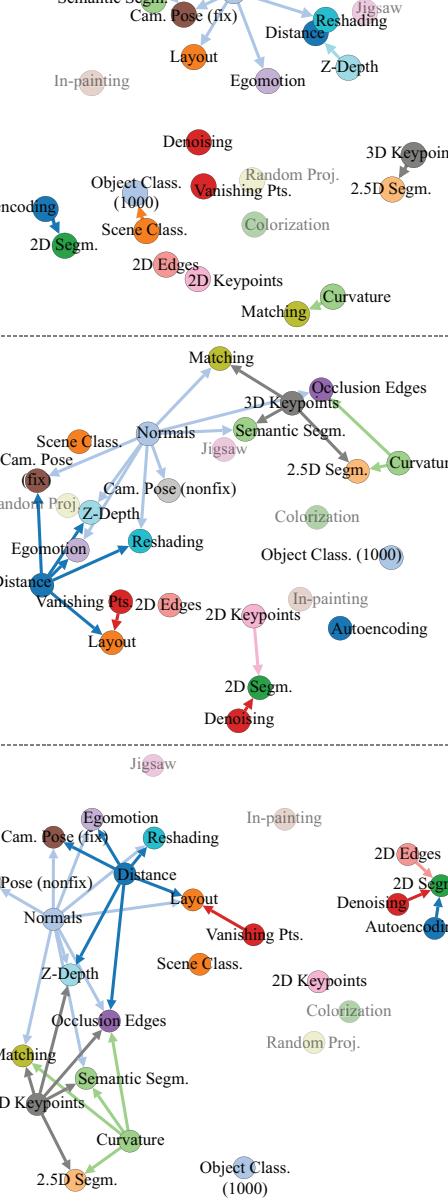
Supervision Budget 2



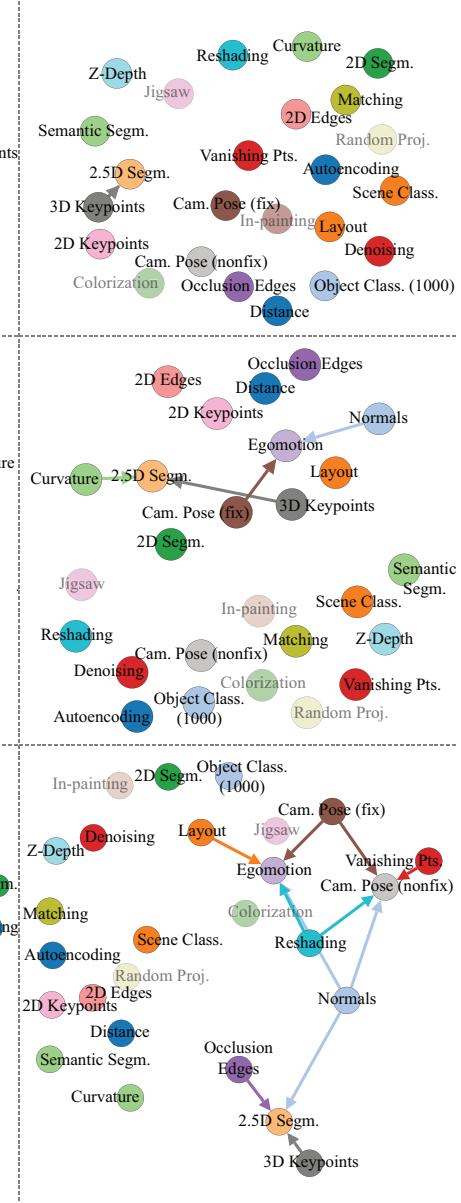
Supervision Budget 8



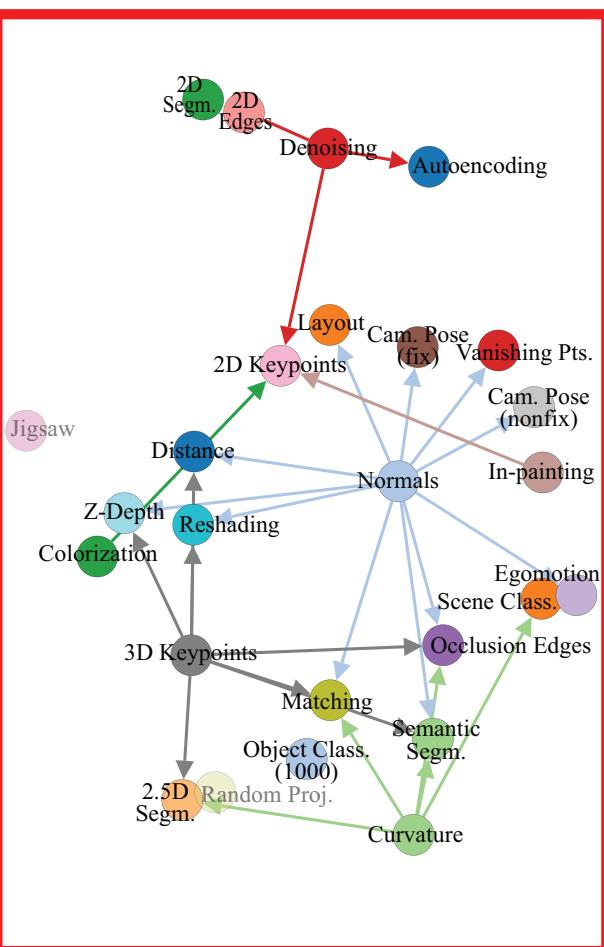
Supervision Budget 15



Supervision Budget 26



Supervision Budget 8 - Order 4 (zoomed)



Supervision Budget Increase (\rightarrow)

