Zero-shot Task Transfer

Vineeth N Balasubramanian
Dept of Computer Science & Engineering
Indian Institute of Technology, Hyderabad

(Joint work with Arghya Pal, PhD student)

CVPR 2019 (Oral)

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Our Group's Research

Algorithmic

- Non-convex optimization for DL*
- Learning with Limited Supervision□
- Explainable Machine Learning✪

Applied

- Recognition of Expressions/ emotions, Poses, Gestures, Actions
- Vision on UAVs/Drones
- Computer Vision for Agriculture
- Autonomous Navigation

* On Noise and Optimality in Neural Networks (**ICML 2018 Workshops**)
- Training Autoencoders by Alternating Minimization, arXiv 2019
- Neural Network Attributions: A Causal Perspective, arXiv 2019

□ Adversarial Data Programming, **CVPR 2018**

✪ Grad-CAM++: Generalized Gradient-based Visual Explanations for Convolutional Networks, **WACV 2018**

* Are Saddles Good Enough for Deep Learning, ACM IKDD **CoDS-COMAD' 2018**

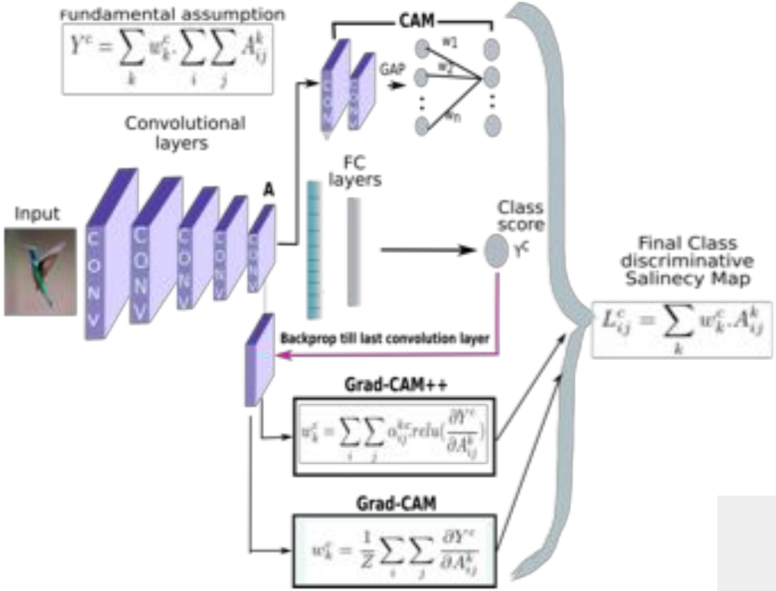
□ Attentive Semantic Video Generation using Captions, **ICCV 2017, ACM MM 2017**

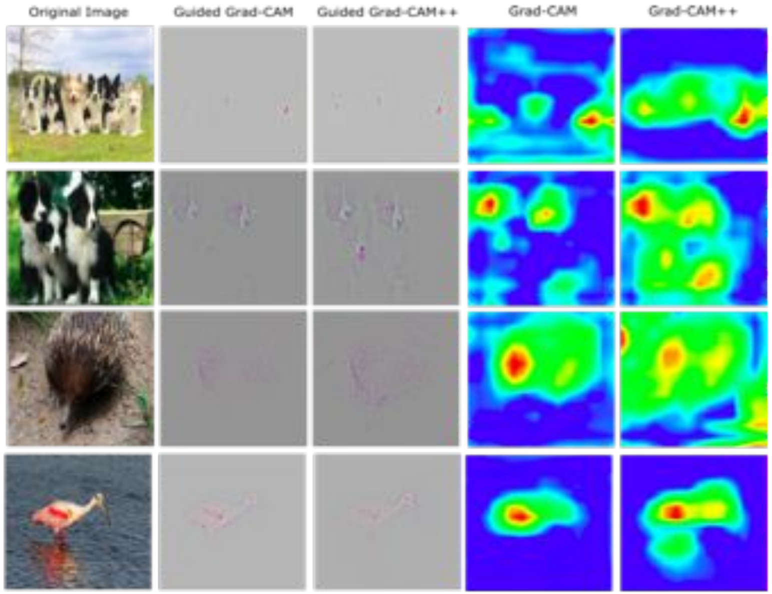
§ Deep Model Compression: Distilling Knowledge from Noisy Teachers, arXiv:1610.09650, 2016

Grad-CAM++: Generalized Visual Explanations

- Need for interpretability
  - DARPA's Explainable AI initiative
- Grad-CAM++

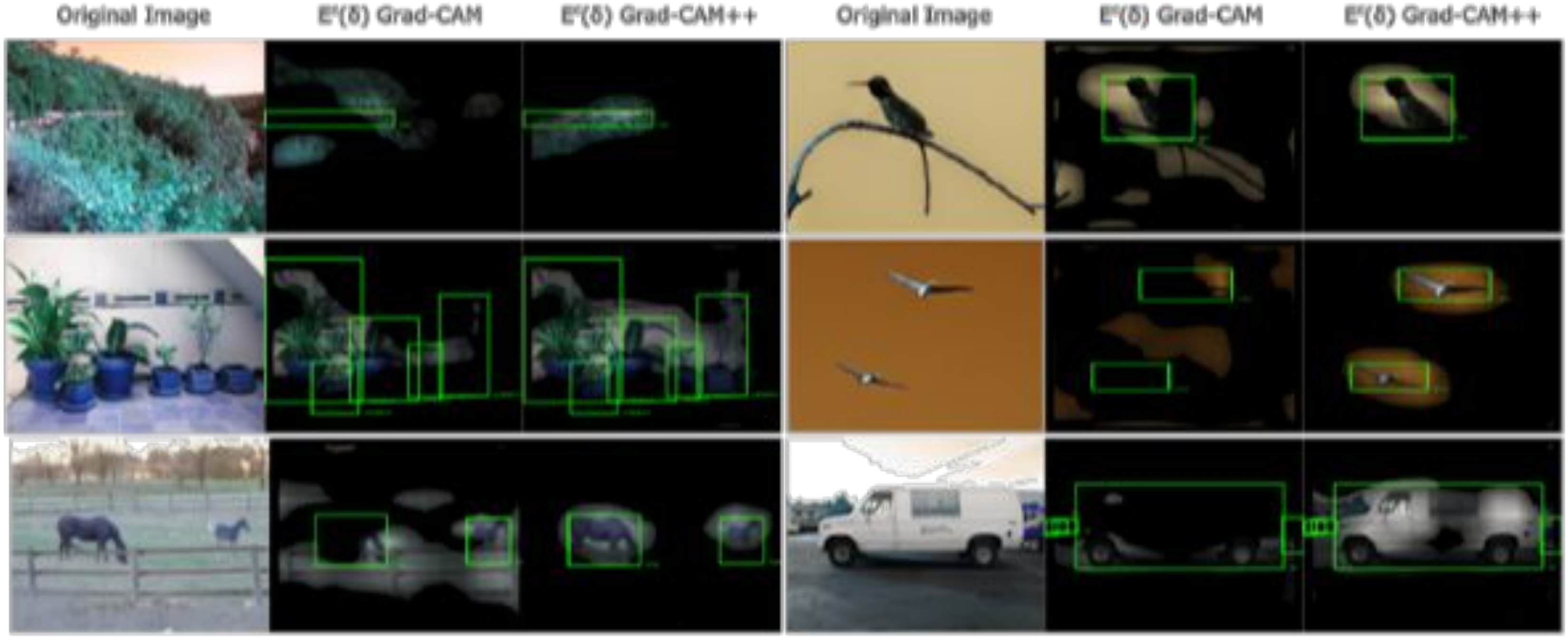
WACV 2018





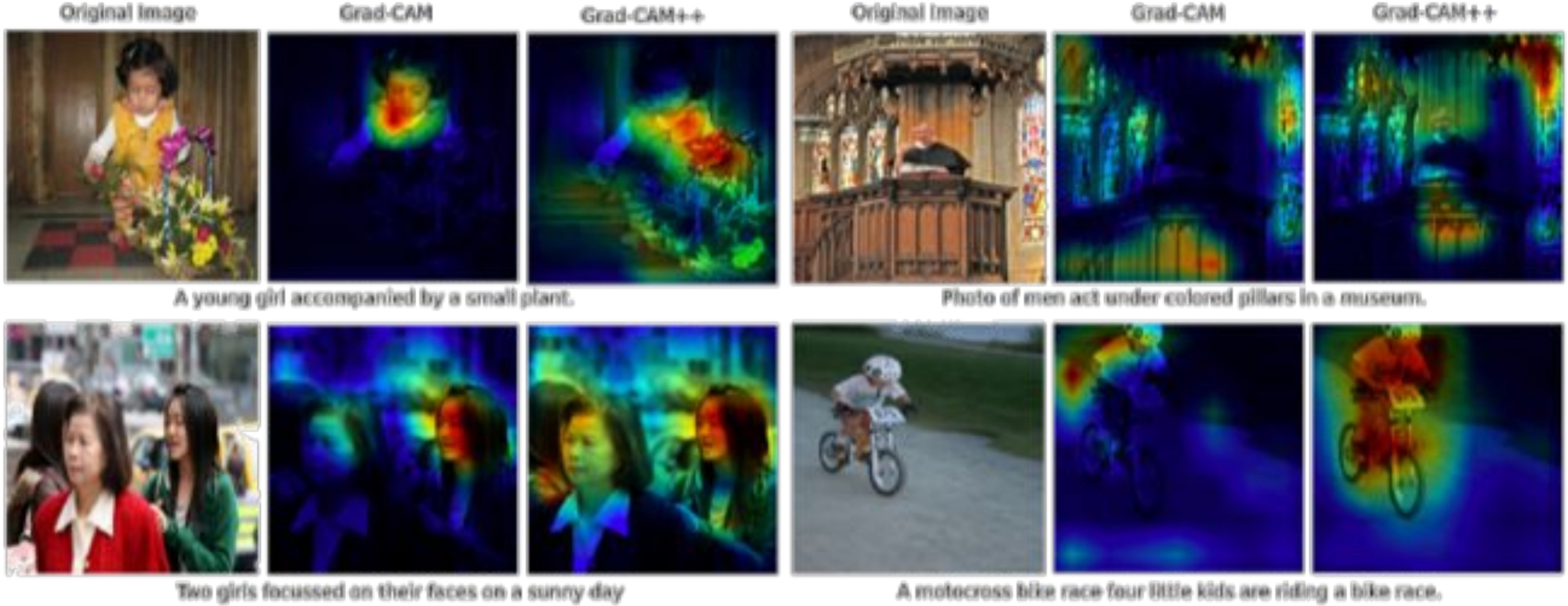
Chattopadhyay, Sarkar, Howlader, Balasubramanian, WACV 2018

Grad-CAM++: Generalized Visual Explanations



Chattopadhyay, Sarkar, Howlader, Balasubramanian, WACV 2018

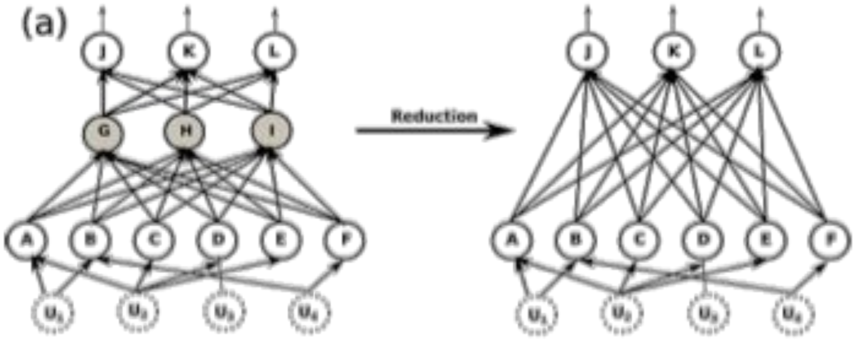
Grad-CAM++: Generalized Visual Explanations



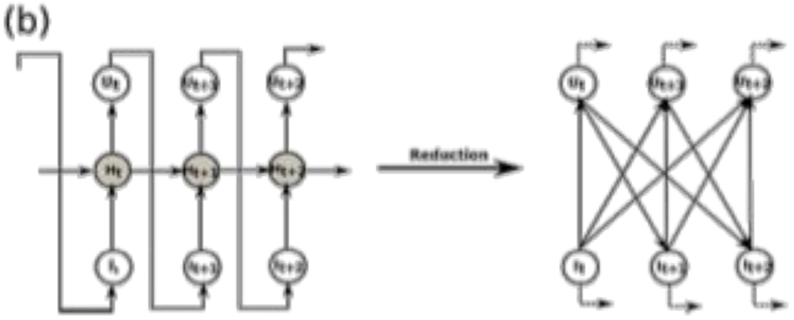
Chattopadhyay, Sarkar, Howlader, Balasubramanian, WACV 2018

Causal NN Attributions

Neural network as a SCM



Feedforward neural network



Recurrent neural network

Chattopadhyay, Manupriya, Sarkar, Balasubramanian, arXiv 2019

Causal NN Attributions

We define it as:

$$ACE^{y}_{do(x_i=\alpha)} = \mathbb{E}[y|do(x_i = \alpha)] - baseline_{x_i}$$

ACE = Average Causal Effect

where baseline is defined as:

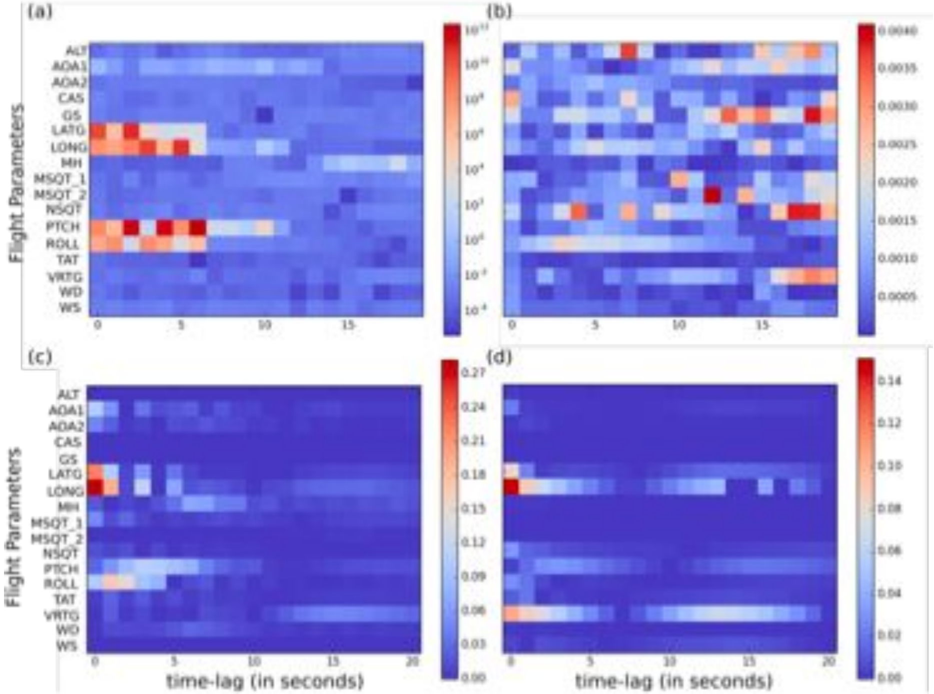
$$\mathbb{E}_{x_i}[\mathbb{E}_y[y|do(x_i = \alpha)]]$$

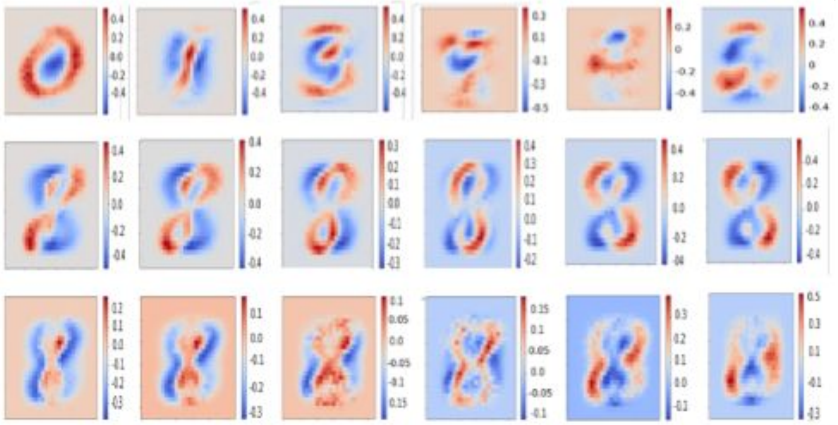
the average ACE across all x_i

Non-trivial to compute

Chattopadhyay, Manupriya, Sarkar, Balasubramanian, arXiv 2019

Causal NN Attributions





Chattopadhyay, Manupriya, Sarkar, Balasubramanian, arXiv 2019

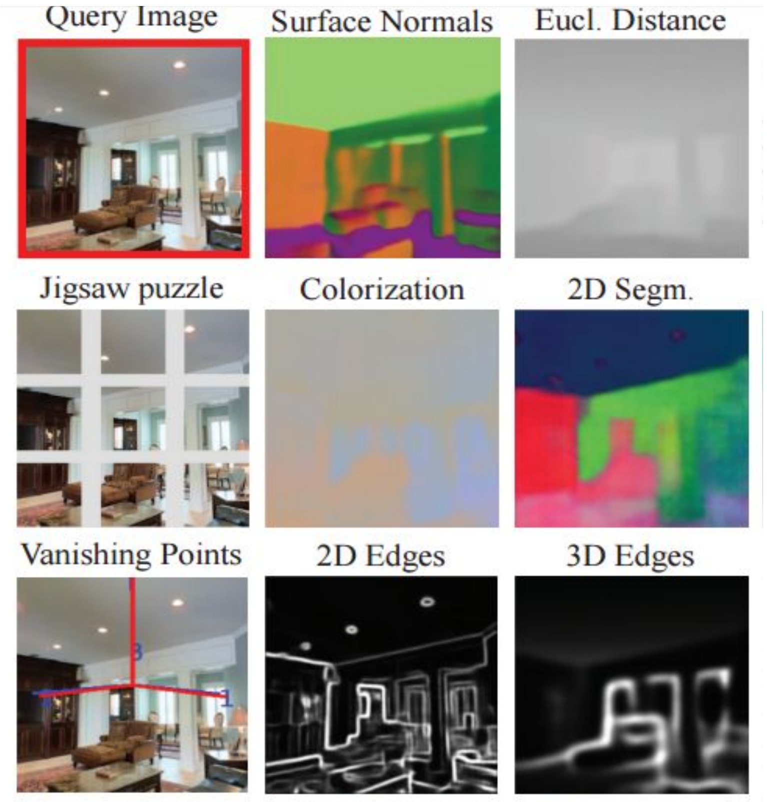
Zero-shot

Zero-shot Task Transfer

Task Transfer

Tasks

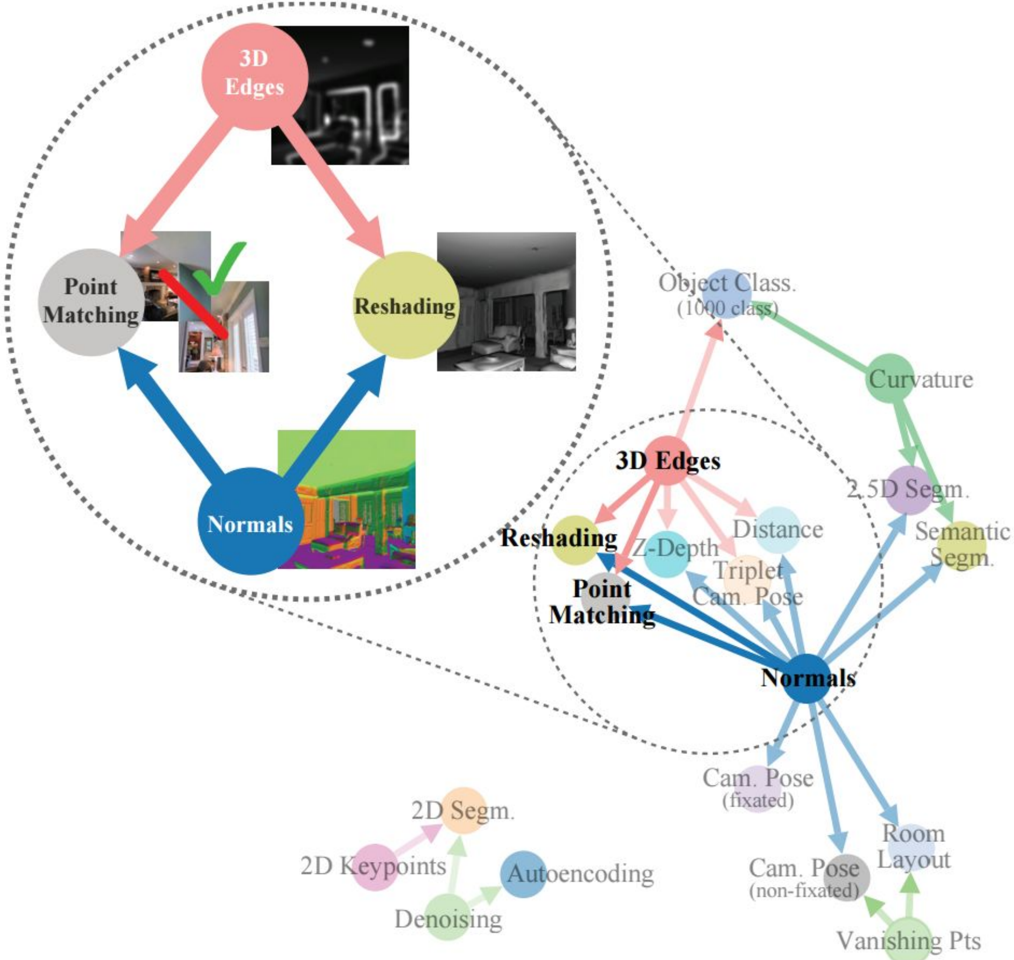
- Vision tasks:
    -
    - Object recognition
    - Depth
    - Edge detection
    - Pose estimation
    - ...



Zamir *et al.*, CVPR 2018

Tasks

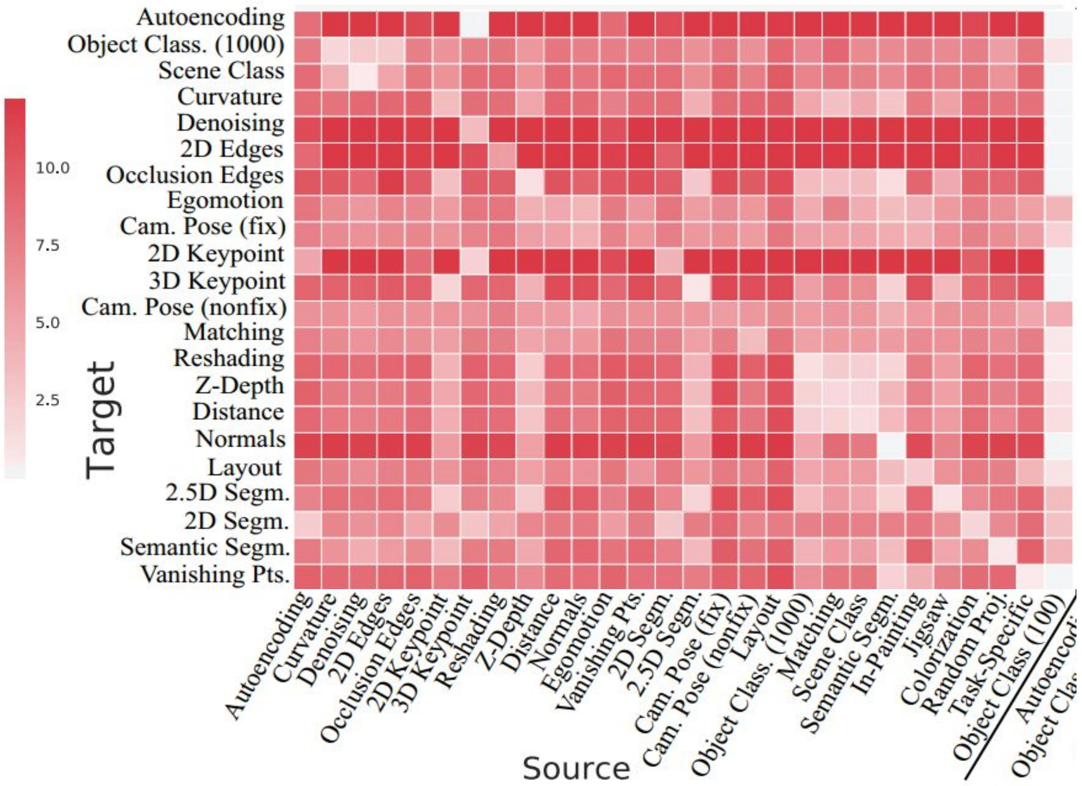
❖ Relation among vision tasks



Zamir *et al.*, CVPR 2018

Tasks

- ❖ Taskonomy CVPR 2018 (Best Paper)
  - ➢ 26 Vision tasks
  - ➢ Sampled set of tasks and not an exhaustive list



Zamir *et al.*, CVPR 2018

Key Takeaway

Tasks

Vision tasks are often related to each other. How to leverage?

Zero-shot

Zero-shot Task Transfer

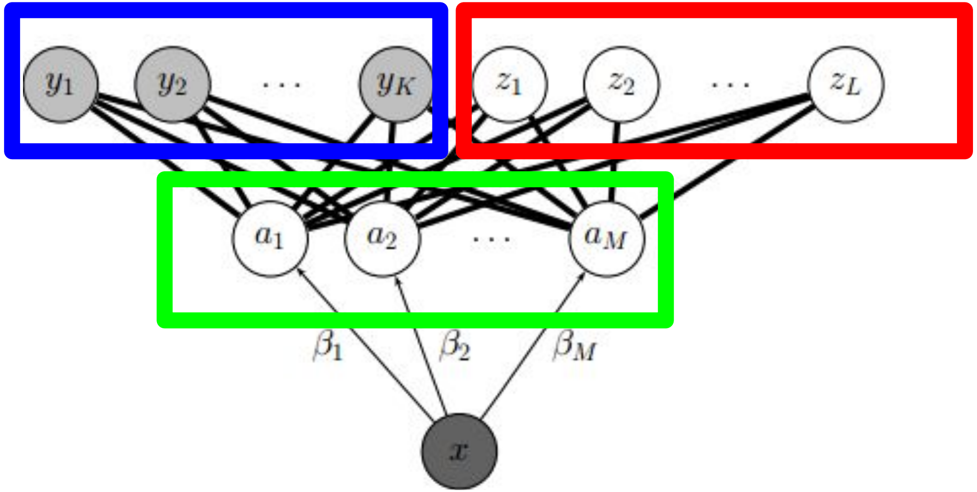
Task Transfer

Zero-shot Classification: A Review

- Object recognition for a set of categories for which we have no training examples
  - $\mathcal{Y} = \{y_1, y_2, \dots, y_m\}$ classes with training samples
  - $\mathcal{Z} = \{z_1, z_2, \dots, z_n\}$ classes with no training samples
  - Learn a classification model: H : $\mathcal{X} \rightarrow (\mathcal{Z}$ union $\mathcal{Y})$

Zero-shot Classification: A Review

- For each class z ϵ $\mathcal{Z}$ and y ϵ $\mathcal{Y}$:
  - attribute representations a^z , a^y ϵ $\mathcal{A}$ are available



Key Takeaway

Tasks

Vision tasks are often related to each other

Zero-shot classification

If relation exists among classes,
new classes can be detected based on attribute representation
without the need for a new training phase / ground truth

Zero-shot Task Transfer: Motivation

- Vision tasks:
  - Expensive
  - May require special sensors
  - Lesser amounts of labeled data leads to poorly performing models

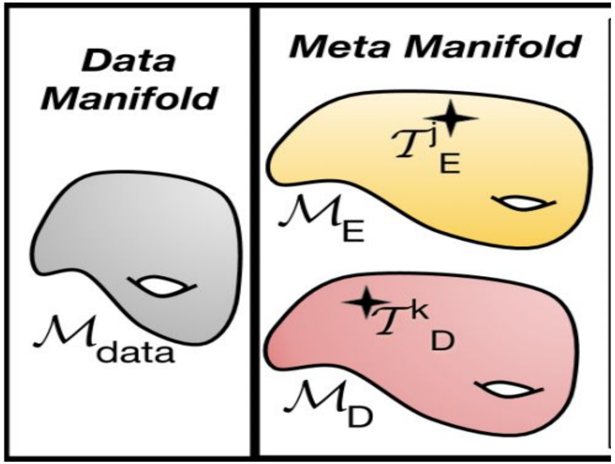
zero-shot classification → zero-shot task transfer

Pal, Balasubramanian, Zero-shot Task Transfer, CVPR 2019

Zero-shot Task Transfer

- Consider K tasks, i.e. $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_K\}$

- Model parameters lie on a meta-manifold $\mathcal{M}_\theta$
- On meta manifold; Task $\mathcal{T}$ is equivalent to model parameter θ



Pal, Balasubramanian, Zero-shot Task Transfer, CVPR 2019

Zero-shot Task Transfer

- Ground truth available for first m tasks
  - $\mathcal{T}_{\text{known}} = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$
  - Corresponding model parameters, $\{\theta_{\mathcal{T}_i} : i = 1, \ldots, m\}$, on meta manifold $\mathcal{M}$ known

- No knowledge of **ground truth** for the zero-shot tasks
  - $\mathcal{T}_{\text{zero}} = \{\mathcal{T}_{(m+1)}, \mathcal{T}_{(m+2)}, \ldots, \mathcal{T}_K\}$

Pal, Balasubramanian, Zero-shot Task Transfer, CVPR 2019

Zero-shot Task Transfer: Idea

- Learn a meta-learning function $F_w(\cdot)$
- $F_w(\cdot)$ regresses unknown zero-shot model parameters from known model parameters

$$\mathcal{F}(\theta_{\tau_1}, \cdots, \theta_{\tau_m}, \Gamma) = \theta_{\tau_j}, \quad j = m+1, \cdots, K$$

Pal, Balasubramanian, Zero-shot Task Transfer, CVPR 2019

Task Transfer Net (TTNet)

$$\mathcal{F}(\theta_{\tau_1}, \cdots, \theta_{\tau_m}, \boxed{\Gamma}) = \theta_{\tau_j}, \quad j = m+1, \cdots, K$$

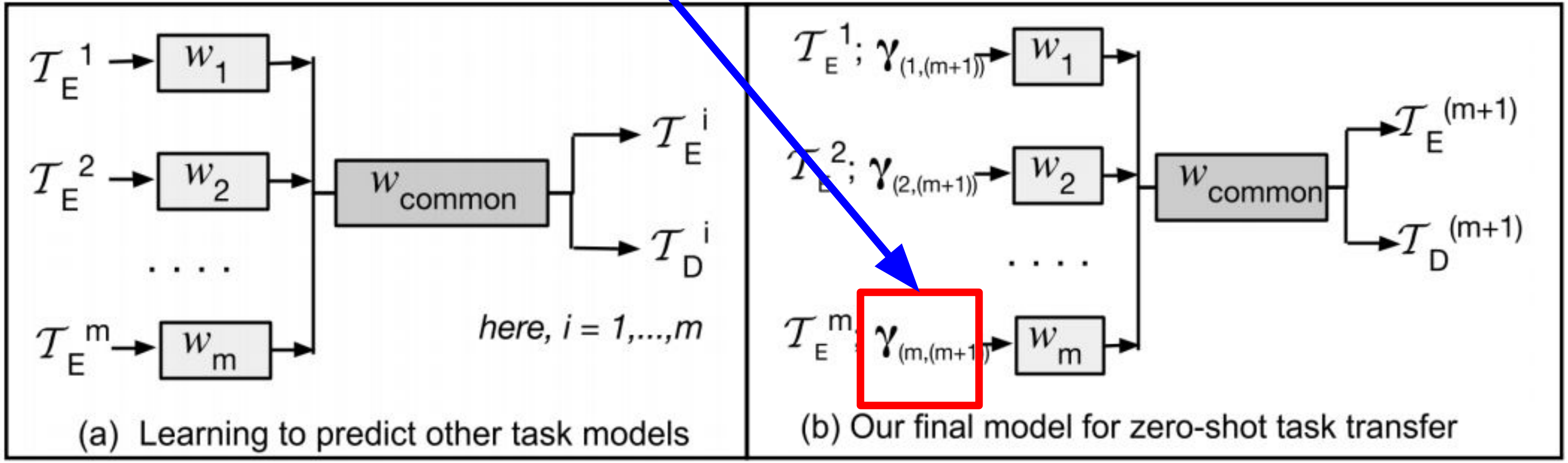
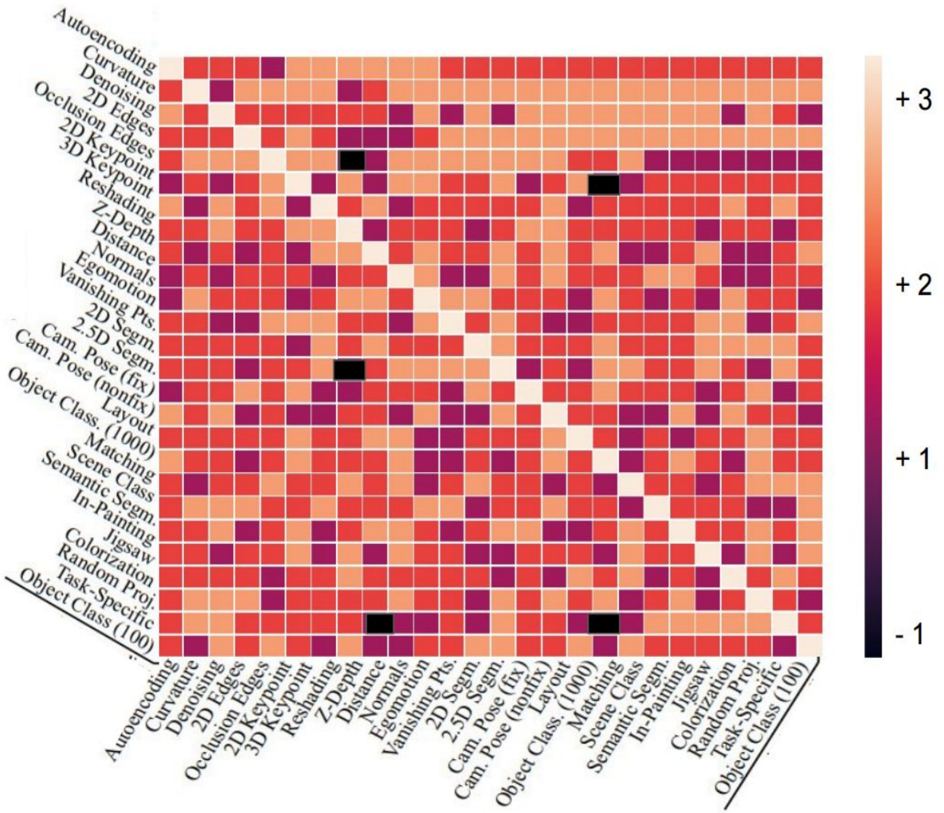


Figure 2: Overview of our work

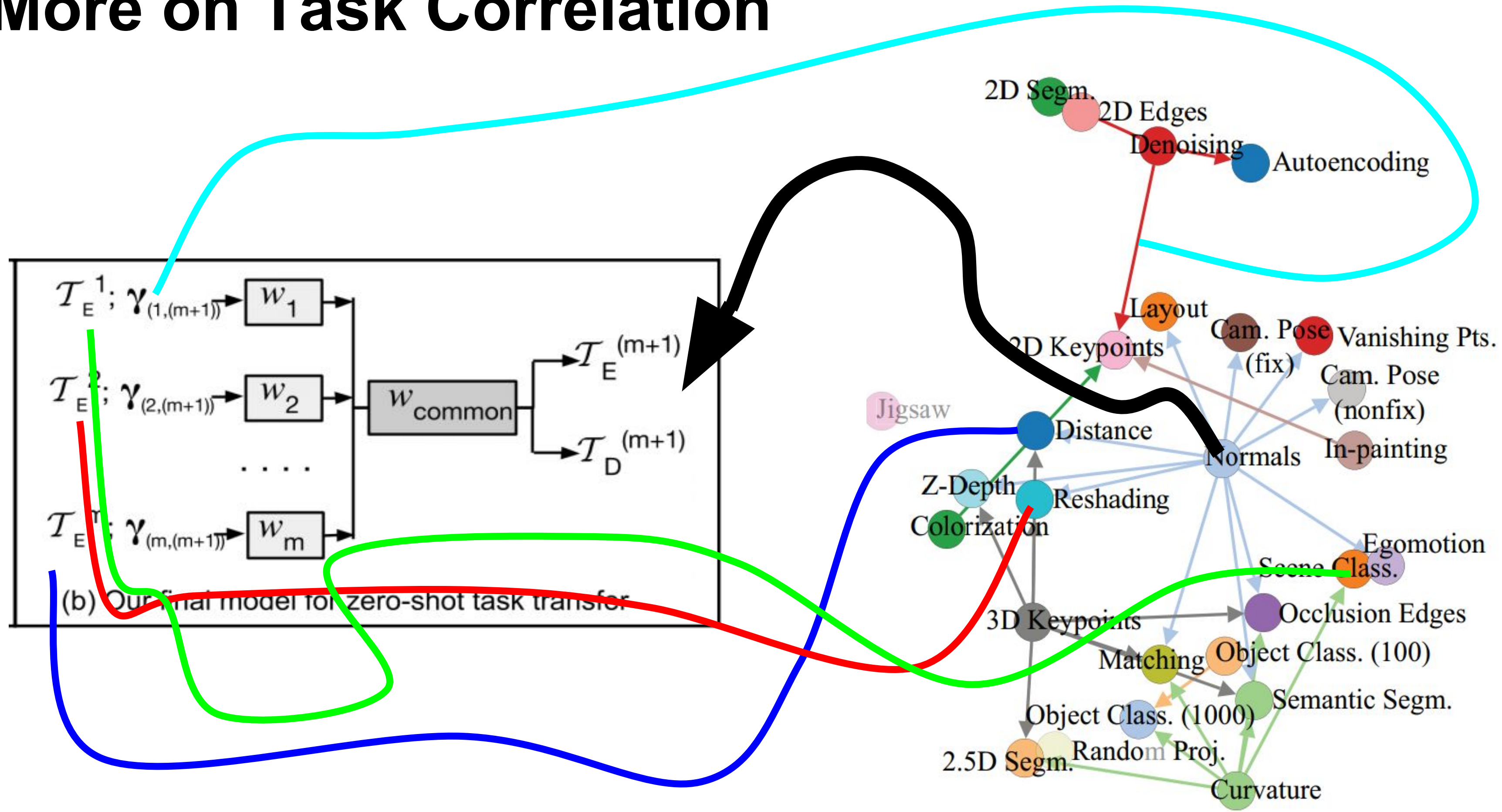
Pal, Balasubramanian, Zero-shot Task Transfer, CVPR 2019

Task Correlation Matrix

$$\mathcal{F}(\theta_{\tau_1},\cdots,\theta_{\tau_m},\boxed{\Gamma}) = \theta_{\tau_j}, \quad j = m+1,\cdots,K$$



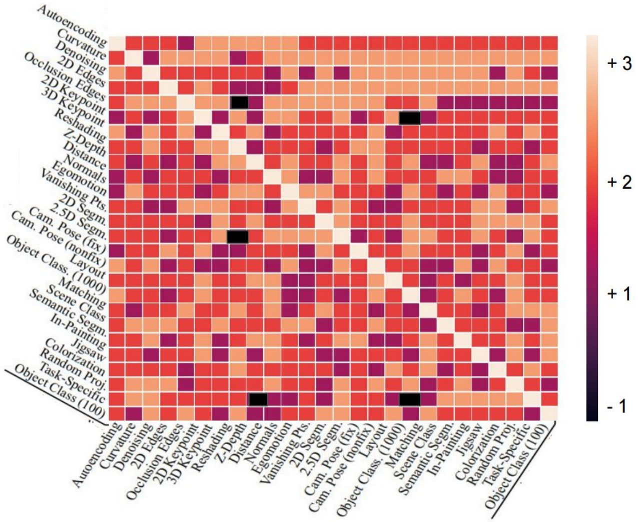
More on Task Correlation



Task Correlation Matrix

- We get task correlation matrix from 30 annotators

- Annotators are asked to give task correlation label on a scale of {+3, +2, +1, 0, −1}
  - +3 denotes self relation
  - +2 describes strong relation
  - +1 implies weak relation
  - 0 to mention abstain
  - −1 to denote no relation between two tasks

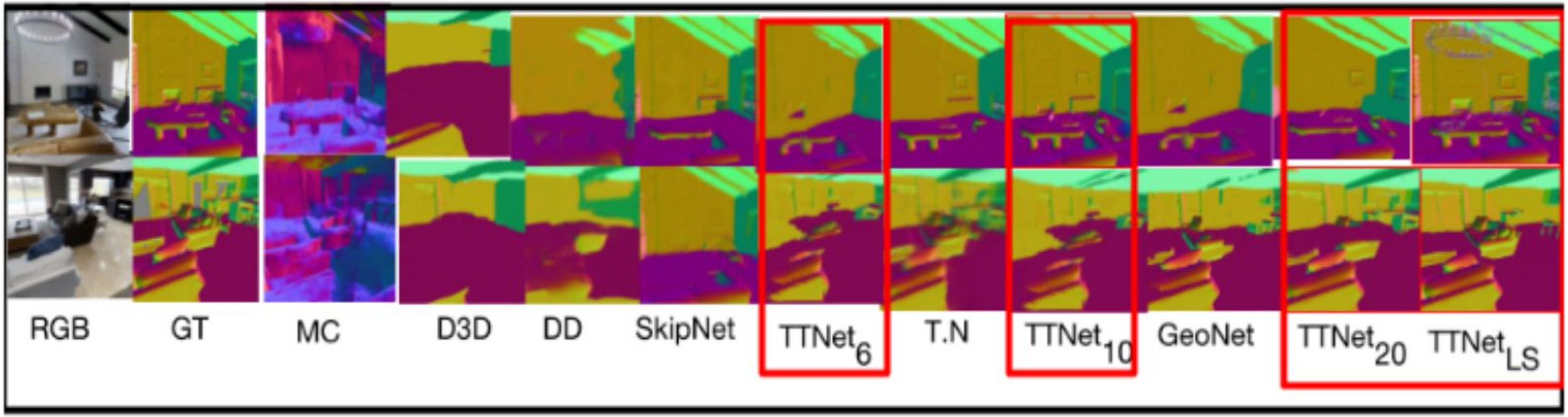


<u>Note:</u>
Our framework is not limited to crowdsourced task correlation. Any other method to compute task correlation will work

Results - Surface Normal Estimation

TTNet$_6$

Source Tasks: Autoencoding, Scene Class, 3D key point, Reshading, Vanishing Pt, Colorization
Zero-Shot Task: Surface Normal

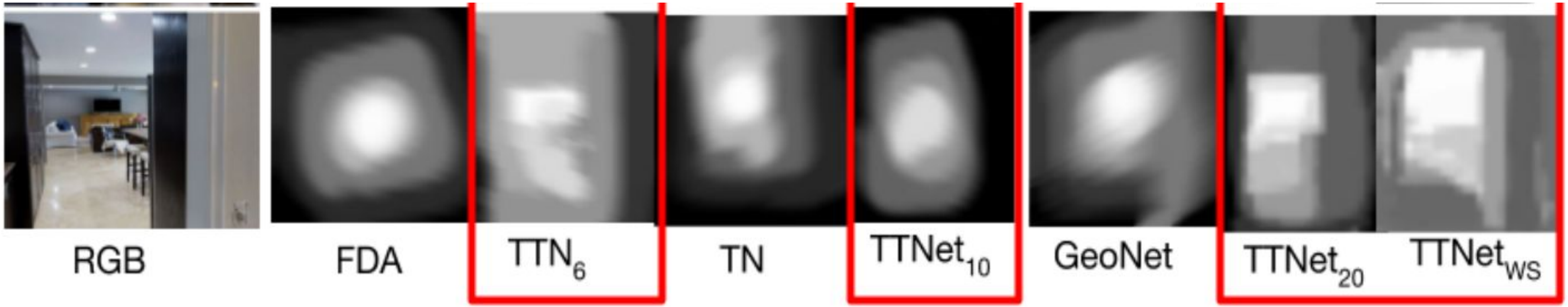


(a) Surface Normal Estimation

Results - Depth Estimation

TTNet$_6$ (same model, only change in gamma values)

Source Tasks: Same as previous
Zero-Shot Task: Depth Estimation

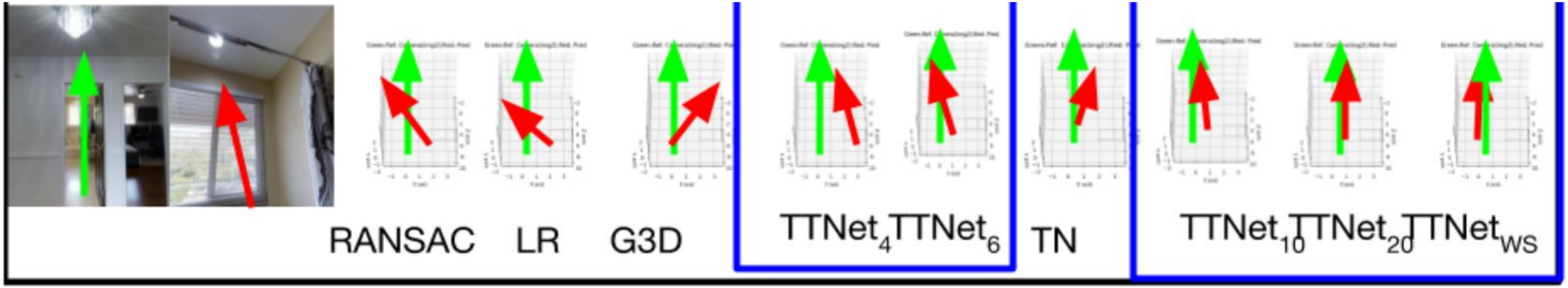


Ref: Arghya Pal, Vineeth N Balasubramanian, Zero-shot Task Transfer, CVPR 2019 Oral

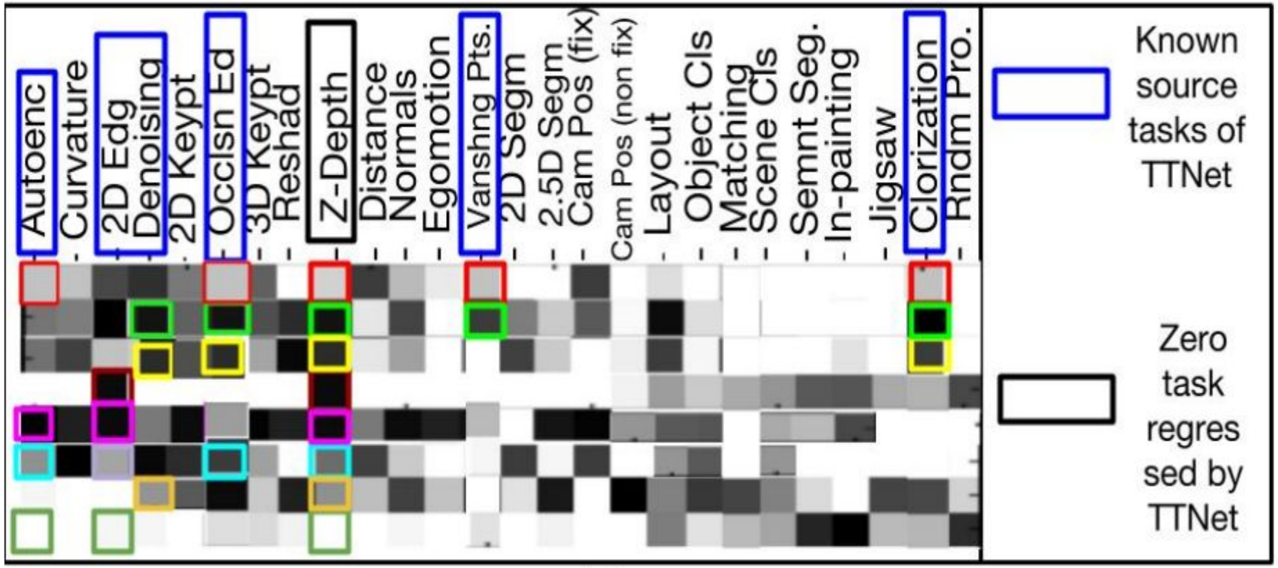
Results - Camera Pose Estimation

TTNet$_6$ (same model, only change in gamma values)

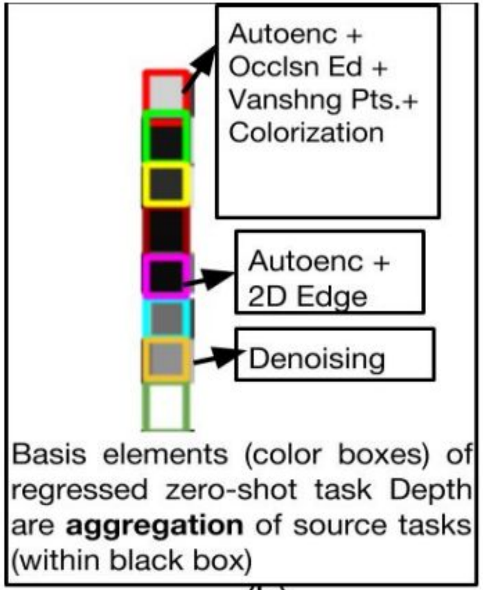
Source Tasks: Same as previous
Zero-Shot Task: Camera Pose Estimation



Why better than Supervised Learning?



(a)



(b)

Zero shot to known task transfer

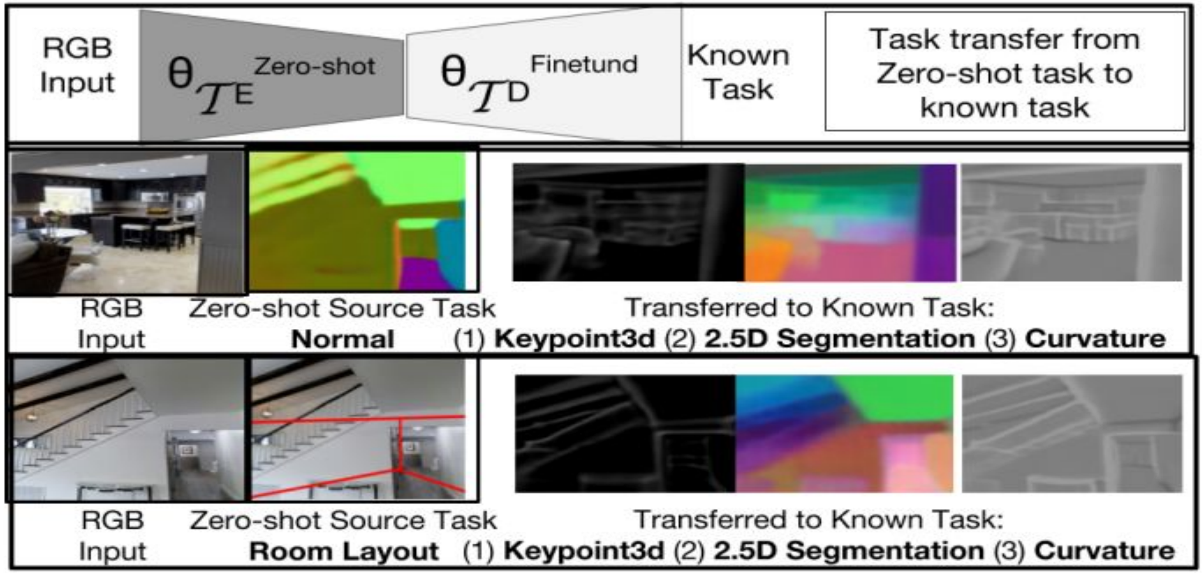


Figure 4: **Zero-shot task to known task transfer.** We consider the zero-shot tasks: *surface normal estimation* and *room layout estimation*, and transfer to models for Keypoint 3D, 2.5D segmentation and curvature estimation.

How many source tasks do we need?

	Autoencoding	Object Class	Scene Class	Curvature	Denoising	2D Edges	Occlusion Edges	Ego motion	Cam Pose (fixed)	2D Key Point	3D Key Point	Cam Pose (non-fixed)	Matching	Reshading	Z-Depth	Distance	Normals	Room Layout	2.5D Segmentation	2D Segmentation	Semantic Segmentation	Vanishing Point	Jig-Saw Puzzle	Random Projection	Colorization	*Win Rate (Normal) (%)*	*Win Rate (Room Layout) (%)*	*Win Rate (Depth) (%)*	*Win Rate (Camera Ps. (fixed) (%)*
	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	79%	62%	71%	71%
4	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	71%	58%	61%	59%
	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	75%	79%	79%	52%
	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	**88%**	**85%**	**87%**	**89%**
6	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	**87%**	**86%**	**86%**	**89%**
	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	**85%**	**88%**	**86%**	**82%**
	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	85%	84%	87%	85%
10	✓	✓	✗	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	87%	88%	91%	92%
	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✓	88%	83%	81%	89%
15	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	88%	85%	91%	93%
	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	89%	87%	81%	85%
18	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✗	✓	✗	✗	✓	✗	✓	✓	✓	✓	✓	93%	91%	97%	91%
	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	95%	91%	93%	94%
20	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	94%	91%	93%	89%

Different Choices of Zero-shot tasks

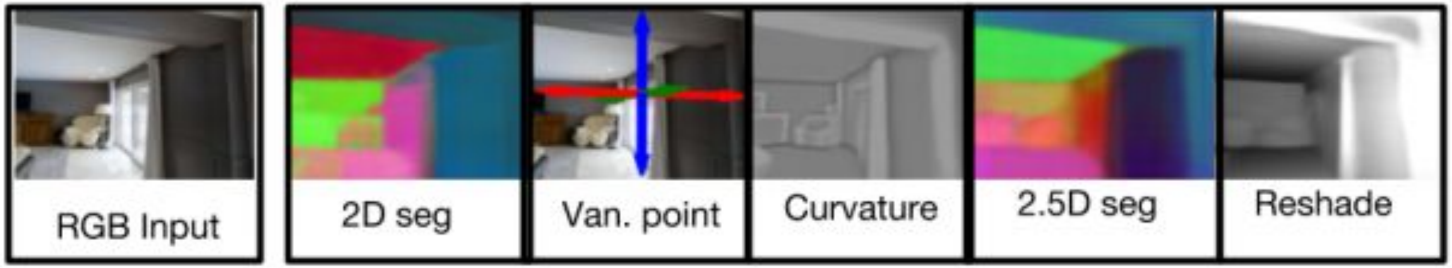


Figure 6: **Different Choice of Zero-Shot Tasks.** Results of TTNet_6 on different set of zero shot tasks: 2D segmentation, Vanishing point estimation, Curvature estimation, 2.5D segmentation and reshading.

Ref: Arghya Pal, Vineeth N Balasubramanian, Zero-shot Task Transfer, CVPR 2019 Oral

Performance on Other Datasets:

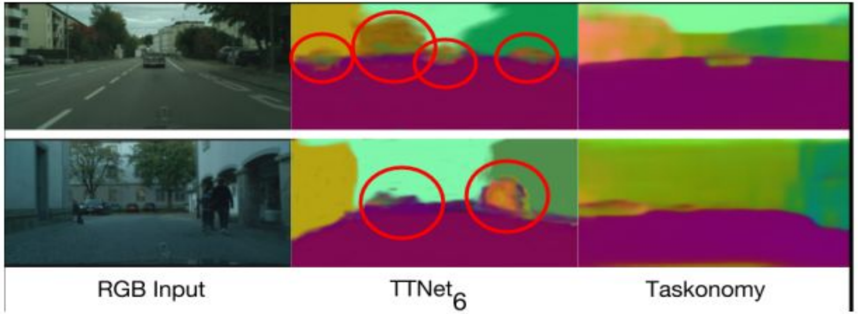


Figure 7: **Surface normal estimation on Cityscapes.** Red circles highlight details (car, tree, human) captured by our model, which is missed by Taskonomy

Method	AP{50:95}	AP{50}	AP{75}	AP{sml}	AP{med}	AP{lrg}
CoupleNet	34.4	54.8	37.2	13.4	8.1	50.8
TTNet{6}	29.9	51.9	34.6	10.8	32.8	45
YOLOv2	21.6	44	19.2	5	22.4	35.5

Object detection on COCO-Stuff dataset

Ref: Arghya Pal, Vineeth N Balasubramanian, Zero-shot Task Transfer, CVPR 2019 Oral

Thank you!

Questions?

vineethnb@iith.ac.in

Department of Computer Science and Engineering, IIT-Hyderabad

http://www.iith.ac.in/~vineethnb