Zero-shot Task Transfer

Vineeth N Balasubramanian

(Joint work with Arghya Pal, PhD student)

Indian Institute of Technology, Hyderabad, INDIA

Presented as Oral at

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Our Group's Research

Algorithmic

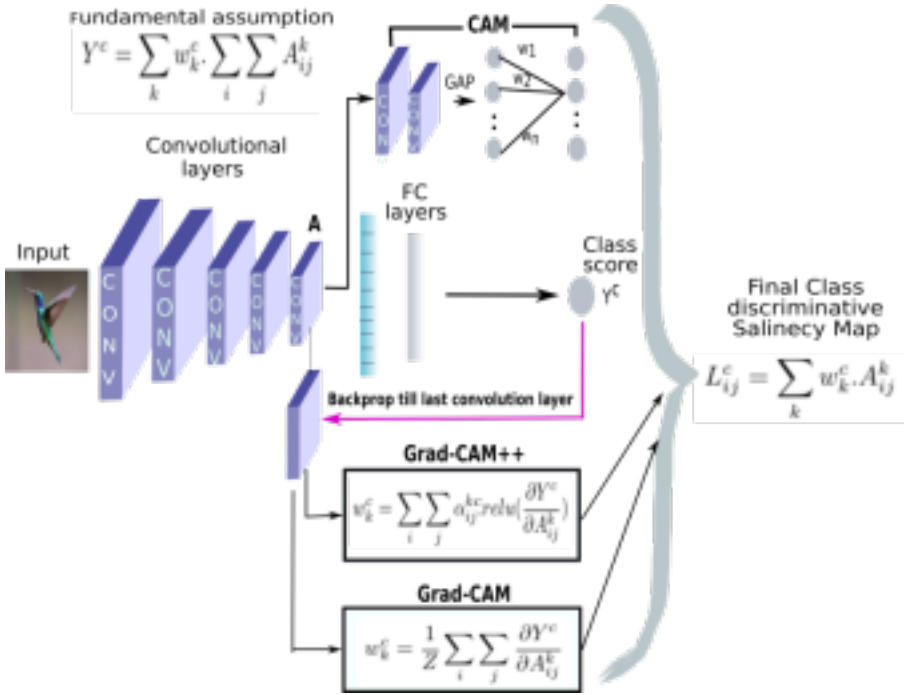
- Non-convex optimization for DL*
- Explainable ML§
- Learning with low supervision⌘

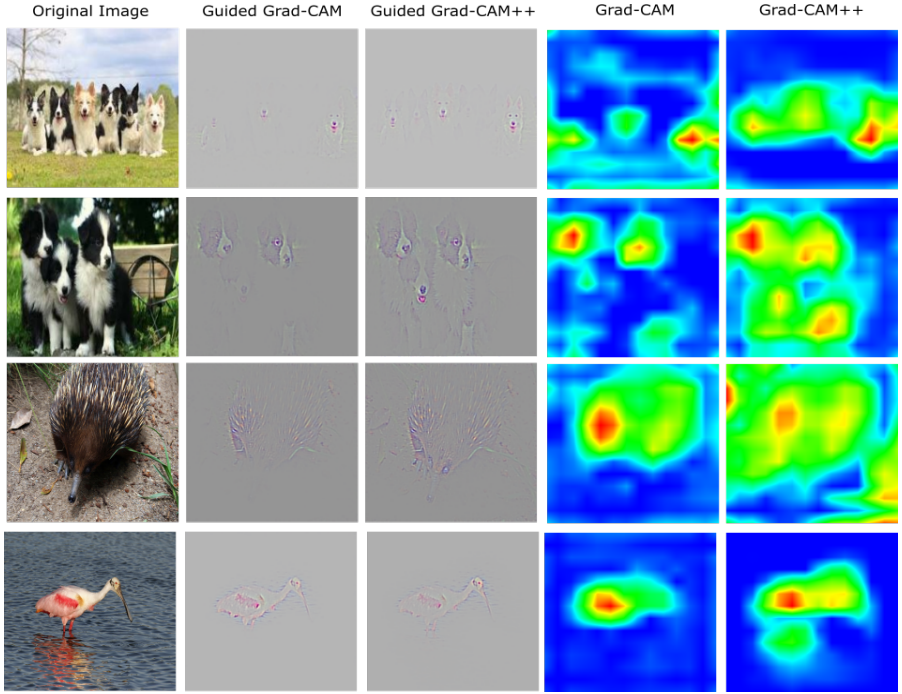
Applied

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

§ Neural Network Attributions: A Causal Perspective, **ICML 2019**
⌘Zero-shot Task Transfer, **CVPR 2019**
* Submodular Batch Selection for Training Deep Neural Networks, **IJCAI 2019**
* On Noise and Optimality in Neural Networks, **ICML 2018 Workshops**
⌘Adversarial Data Programming, **CVPR 2018**
§ Grad-CAM++: Generalized Gradient-based Visual Explanations for Convolutional Networks, **WACV 2018**
⌘Attentive Semantic Video Generation using Captions, **ICCV 2017**, **ACM MM 2017**

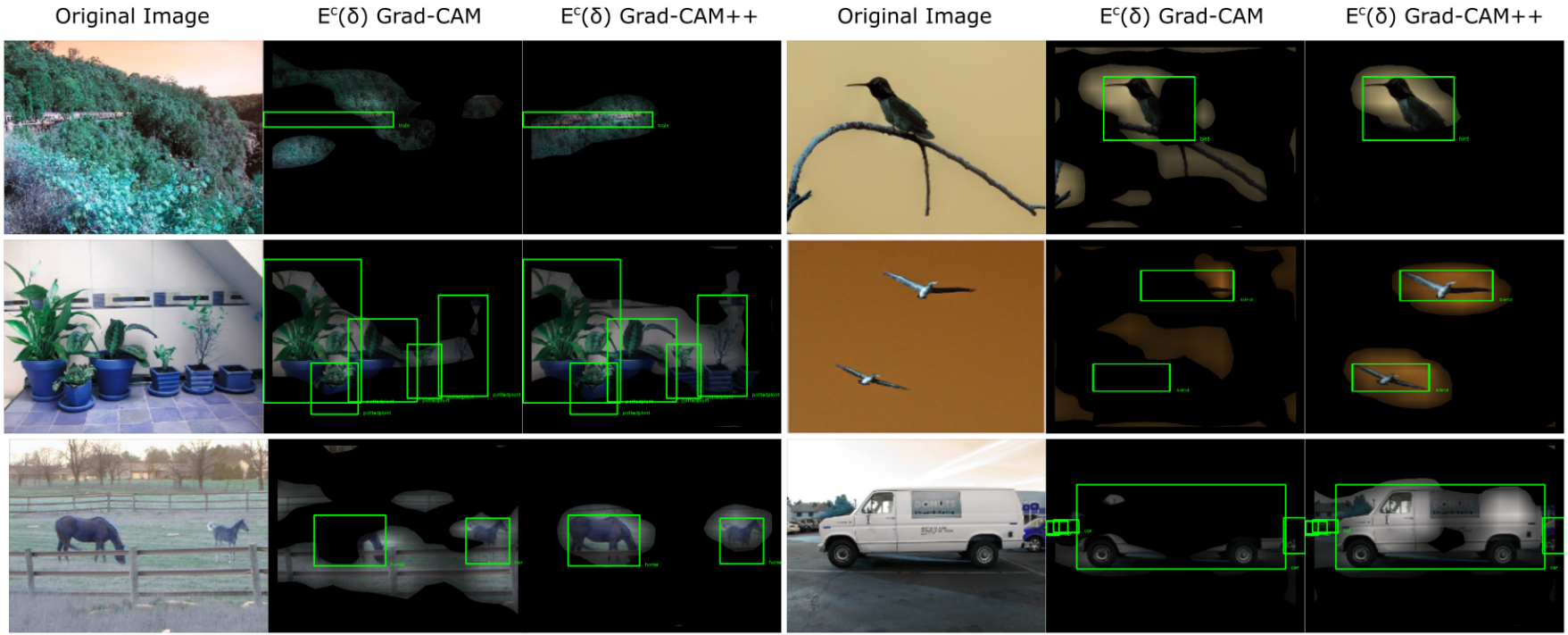
Grad-CAM++





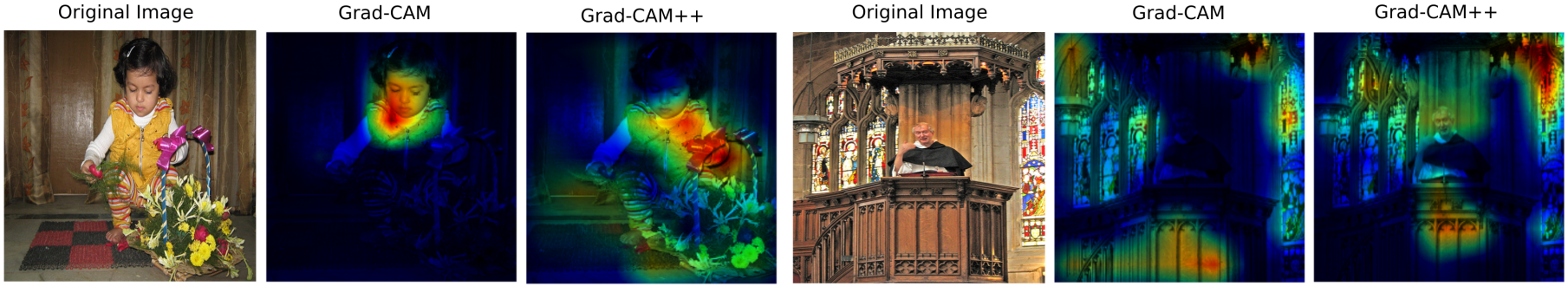
Chattopadhyay, Sarkar, Howlader, Balasubramanian, WACV 2018

Grad-CAM++



Chattopadhyay, Sarkar, Howlader, Balasubramanian, WACV 2018

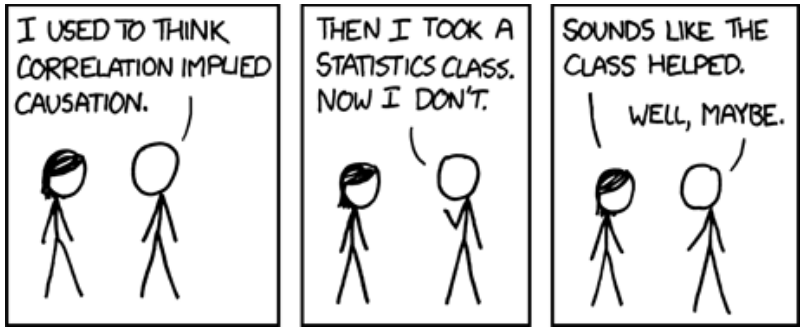
Grad-CAM++

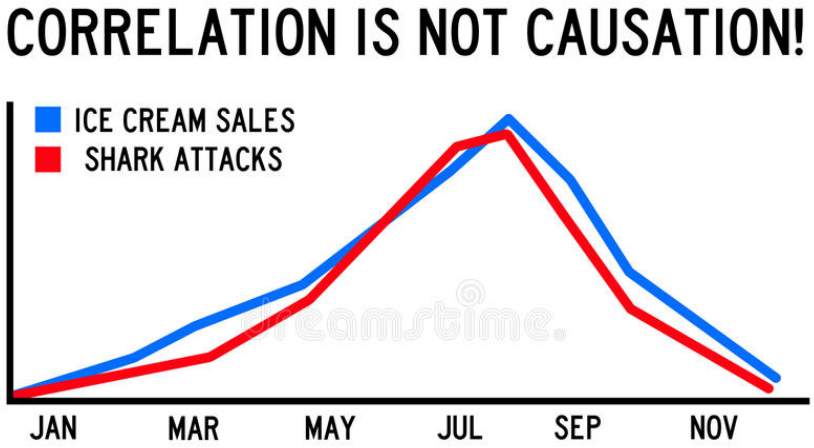


Chattopadhyay, Sarkar, Howlader, Balasubramanian, WACV 2018

Causal Attributions in Neural Networks

- Is feature correlation to output a true indicator of explainability?
- Or do we need to find causal relationships in the analyzed data-output pairs?



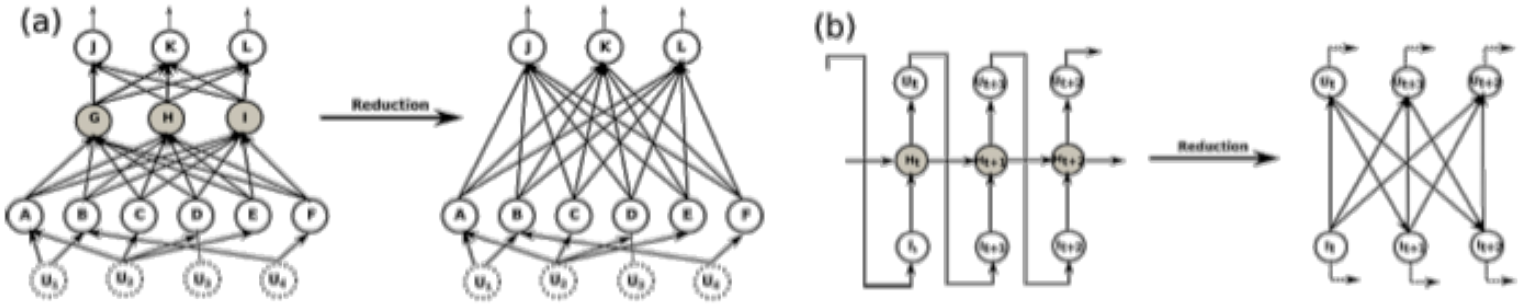


Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

Chattopadhyay, Manupriya, Sarkar, Balasubramanian, ICML 2019

Causal Attributions in Neural Networks

Neural network as a SCM



Feedforward neural network

Recurrent neural network

Chattopadhyay, Manupriya, Sarkar, Balasubramanian, ICML 2019

Causal Attributions in Neural Networks

$$\mathbb{E}[y|do(x=1)] - \mathbb{E}[y|do(x=0)]$$

For continuous variables:

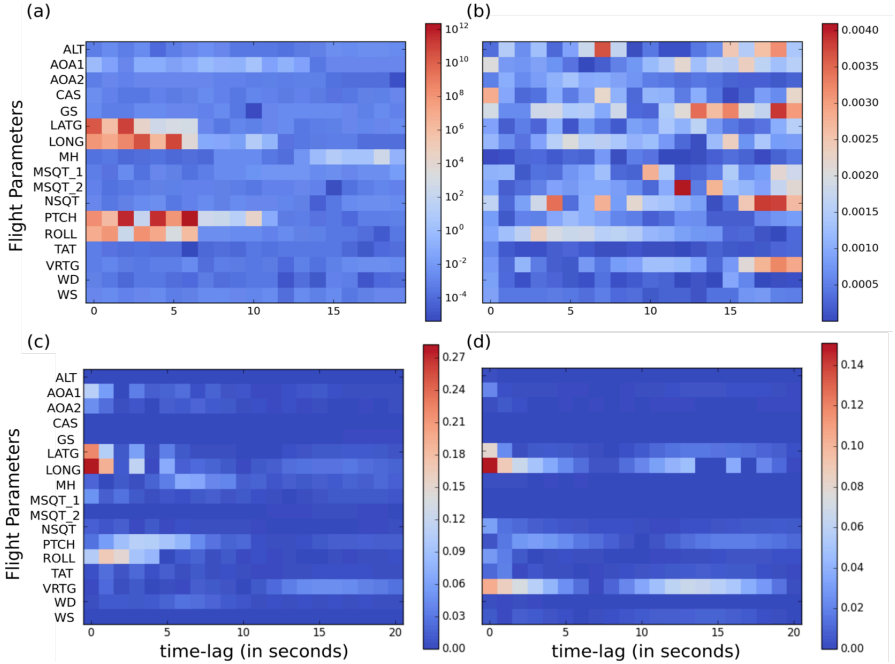
$$ACE^{y}_{do(x_i=\alpha)} = \mathbb{E}[y|do(x_i=\alpha)] - baseline_{x_i}$$

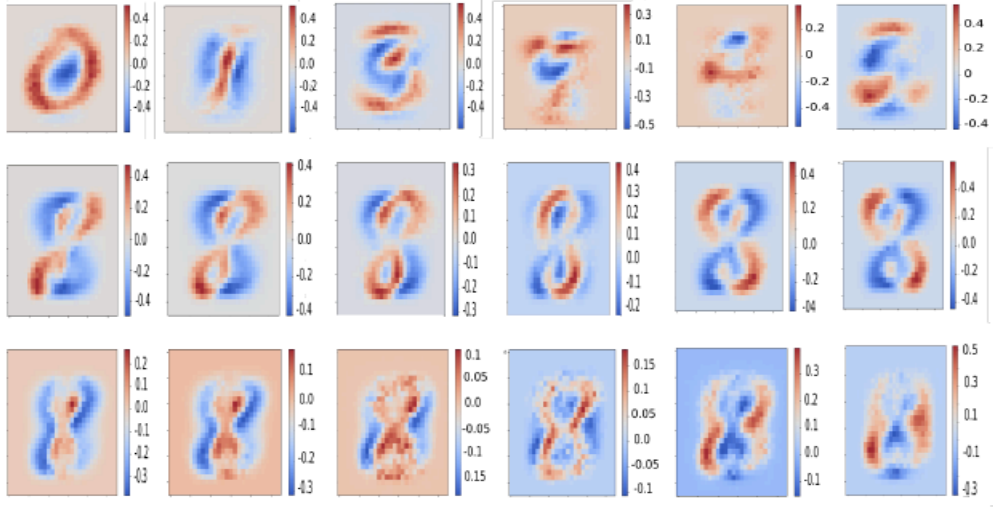
where baseline is defined as:

$$\mathbb{E}_{x_i}[\mathbb{E}_y[y|do(x_i=\alpha)]]$$ the average ACE across all x_i

Chattopadhyay, Manupriya, Sarkar, Balasubramanian, ICML 2019

Causal Attributions in Neural Networks





Chattopadhyay, Manupriya, Sarkar, Balasubramanian, ICML 2019

Our Group's Research

Algorithmic

- Non-convex optimization for DL*
- Explainable ML§
- Learning with low supervision⌘

Applied

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

§ Neural Network Attributions: A Causal Perspective, **ICML 2019**

⌘Zero-shot Task Transfer, **CVPR 2019**

* Submodular Batch Selection for Training Deep Neural Networks, **IJCAI 2019**

* On Noise and Optimality in Neural Networks, **ICML 2018 Workshops**

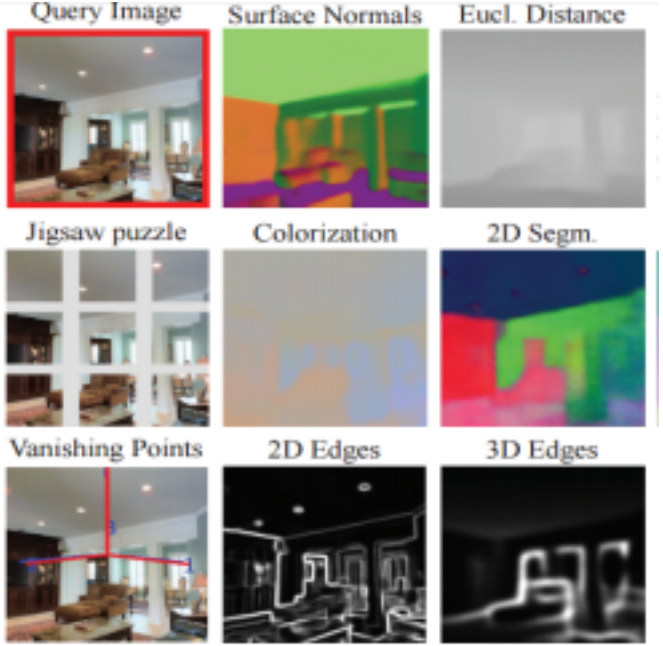
⌘Adversarial Data Programming, **CVPR 2018**

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

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

Introduction

- Consider K vision tasks to accomplish: $\tau_1, \cdots, \tau_K$
  - Object recognition
  - Depth estimation
  - Edge detection
  - Pose estimation
  - … etc etc…



How to automatically predict model parameters for tasks without ground truth?

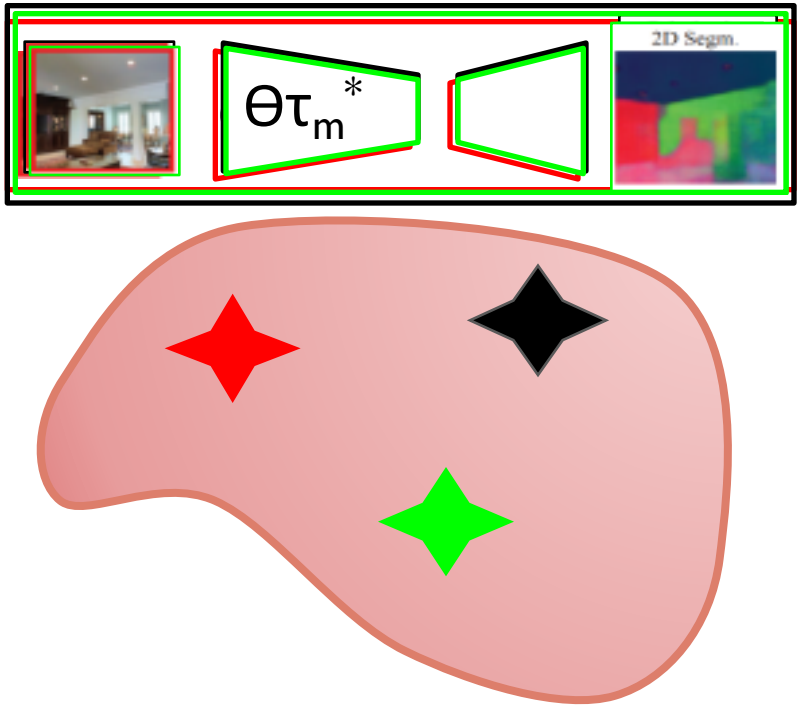
Zero-shot Tasks vs Known Tasks

- Of K tasks:

- Ground-truth available for *m* tasks, $\{\tau_1, \cdots, \tau_m\}$ → **Known Tasks**
- No ground truth available for tasks $\{\tau_{(m+1)}, \cdots, \tau_K\}$ **Zero-shot Tasks**

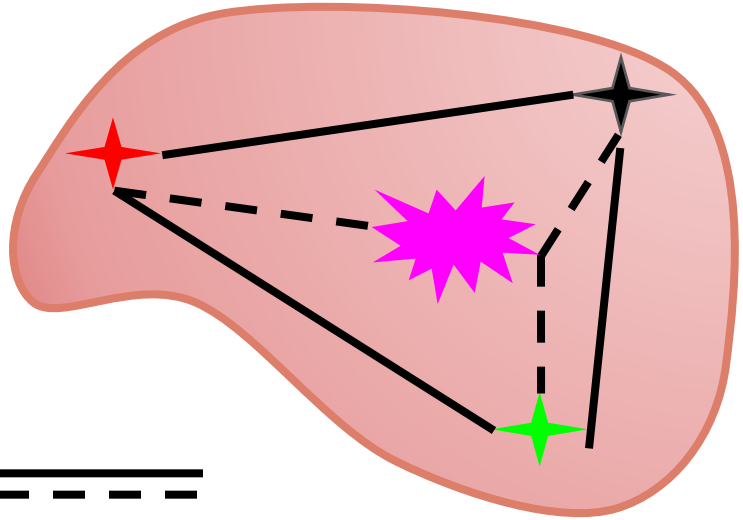
Task Meta-Manifold

- **Known Tasks:** $\{\tau_1, \cdots, \tau_m\}$
- We can learn their corresponding model parameters on task meta-manifold
- Not possible for **zero-shot task**



Zero-shot Task Transfer: Our Objective

- Learn a meta-learning function F to **regress the unknown zero-shot model parameters**; given
  - Model parameters of known tasks ✦
  - Pair-wise task correlations ═══

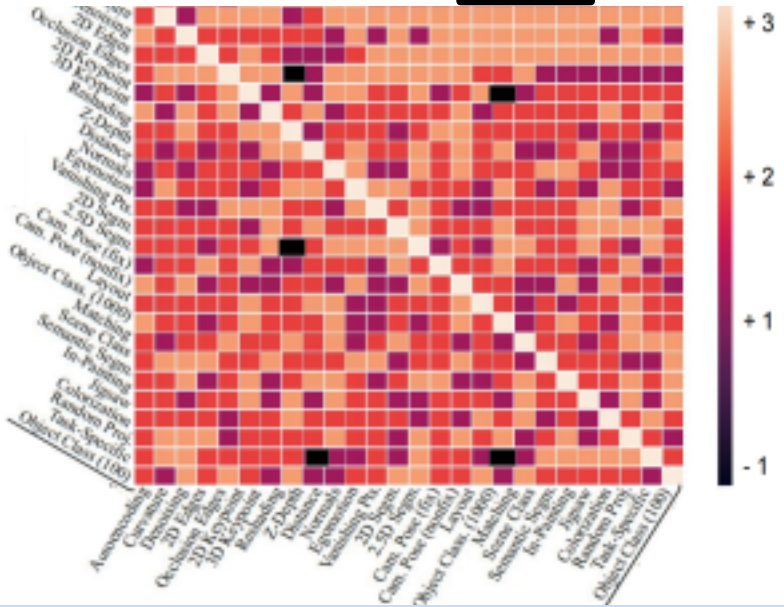


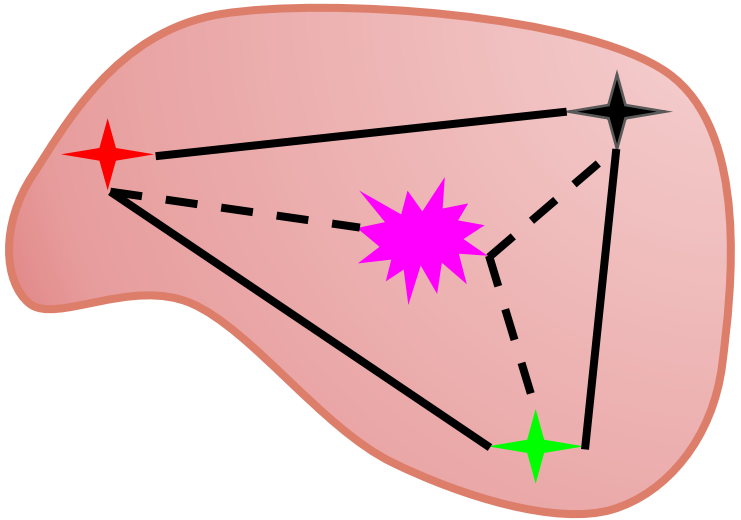
$$\mathcal{F}(\theta_{\tau_1}, \cdots, \theta_{\tau_m}, \Gamma) = \theta_{\tau_j}$$

Pairwise Task Correlation

$$\mathcal{F}(\theta_{\tau_1}, \cdots, \theta_{\tau_m}, \Gamma) = \theta_{\tau_j}$$

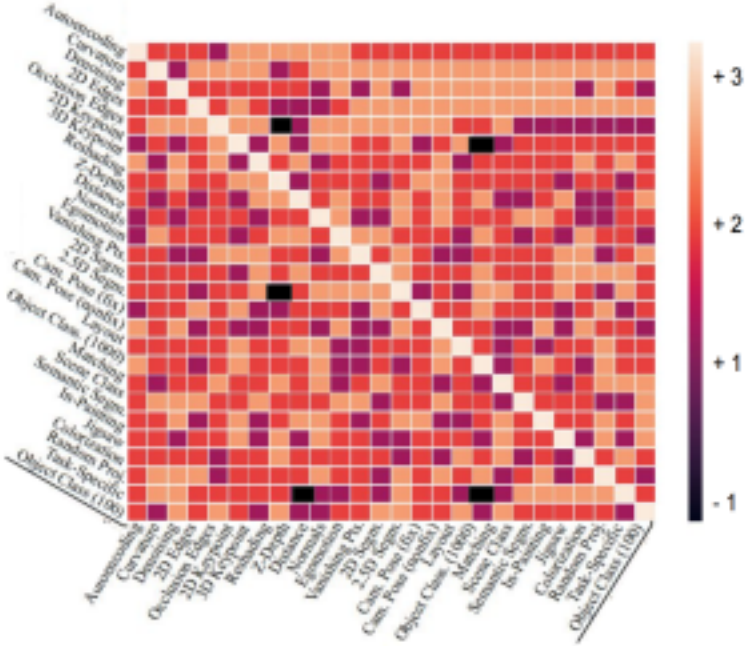
A matrix of pairwise correlations





Pairwise Task Correlation Matrix

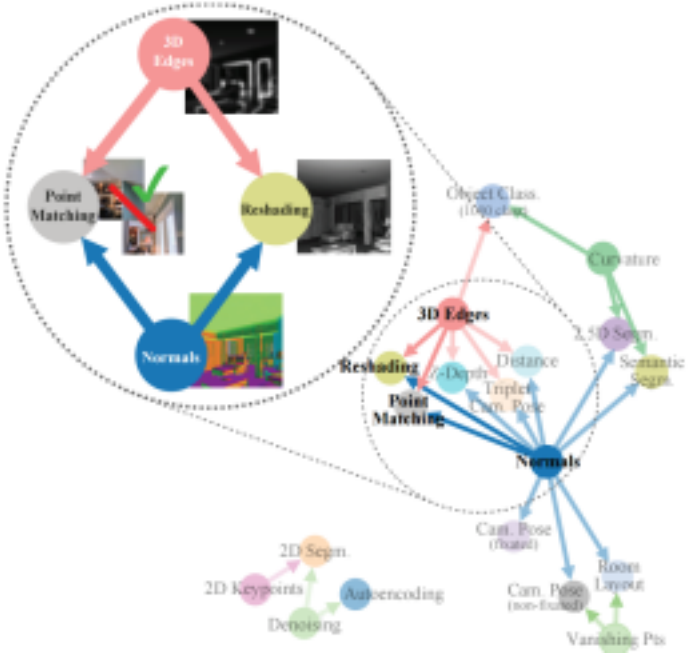
- Wisdom-of-crowd
- A set of experts (30 in our case) asked to rate relation between task pairs τ_i and τ_j, across known and zero-shot tasks
- Note that <u>our framework can admit any other source of task correlation beyond crowdsourcing</u>

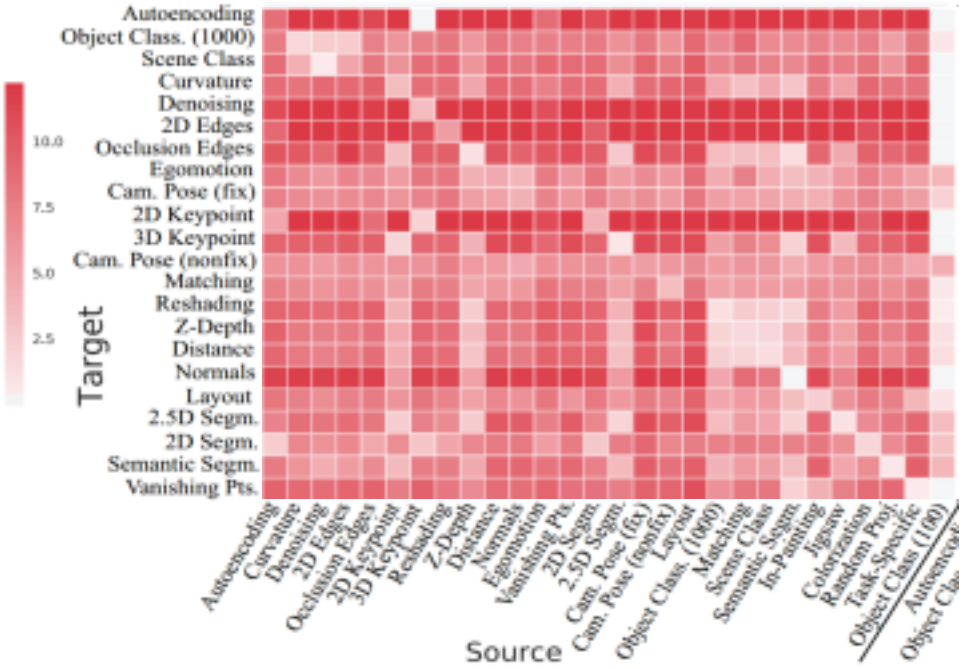


Vineeth N B
Indian Institute of Technology, Hyderabad, INDIA

Pairwise Task Correlation Matrix

Taskonomy CVPR 2018 (Best Paper): 26 vision tasks





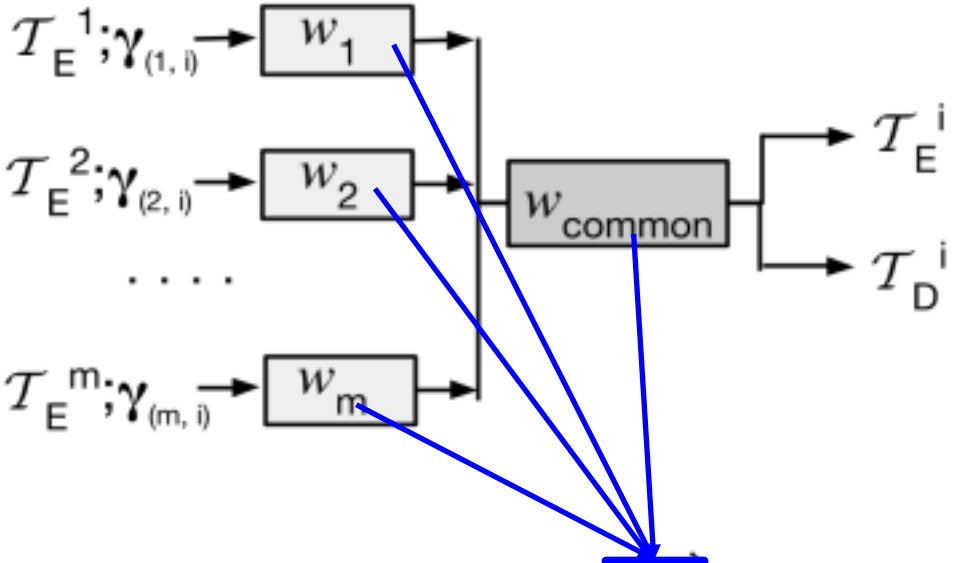
Sampled set of tasks and not an exhaustive list

Vineeth N B
Indian Institute of Technology, Hyderabad, INDIA

Zero-shot Task Transfer

Methodology

- We introduce TTNet, an architecture, to learn our meta-learner F(.) parameterized by the weight **W**



$$\min_{W} \sum_{i=1}^{m} \| \mathcal{F}\Big((\theta_{E_{\tau_1}}, \gamma_{1,i}), \cdots, (\theta_{E_{\tau_1}}, \gamma_{m,i}); W \Big) - (\theta^*_{E_{\tau_i}}, \theta^*_{D_{\tau_i}}) \|^2$$

Pairwise task correlations

TTNet: Training

Given m source tasks

$\Theta E\tau_1^*$

$\Theta E\tau_2^*$

. . .

$\Theta E\tau_m^*$

Find optimum W

$\mathcal{T}_E^1; \gamma_{(1,i)}$ → w_1

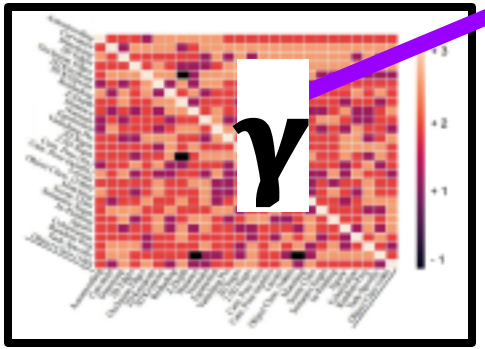
$\mathcal{T}_E^2; \gamma_{(2,i)}$ → w_2

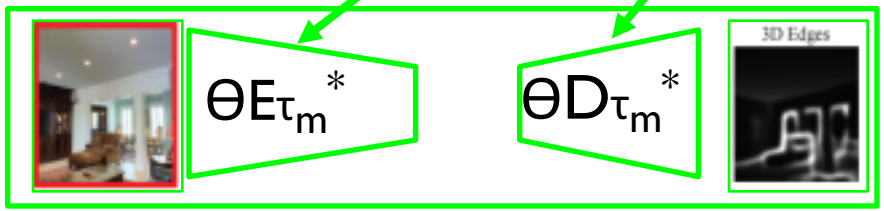
. . . .

$\mathcal{T}_E^m; \gamma_{(m,i)}$ → w_m

w_1, w_2, w_m → w_{common} → $\mathcal{T}_E^i$, $\mathcal{T}_D^i$

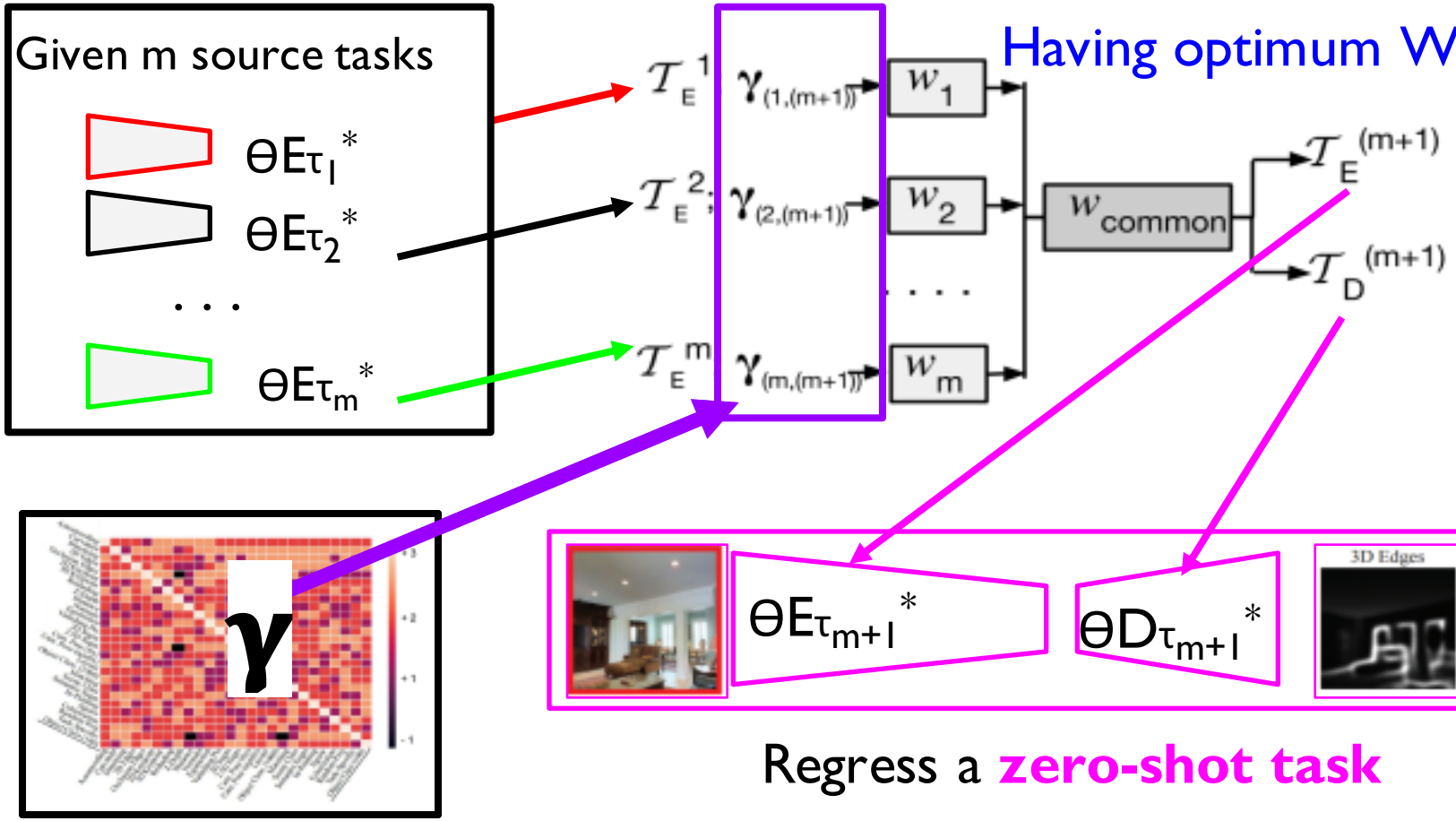
here, i = 1,..,m





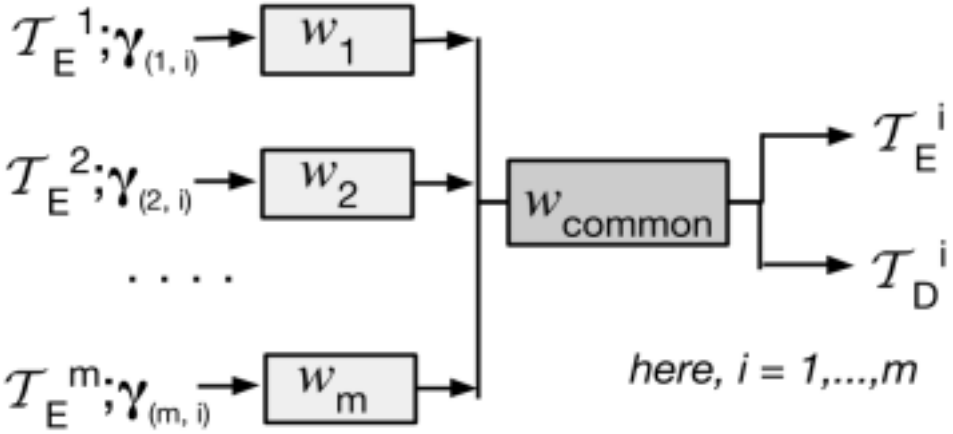
Regress a known task

TTNet: Zero-shot Parameter Regression



Loss Function

- We add a data model consistency loss

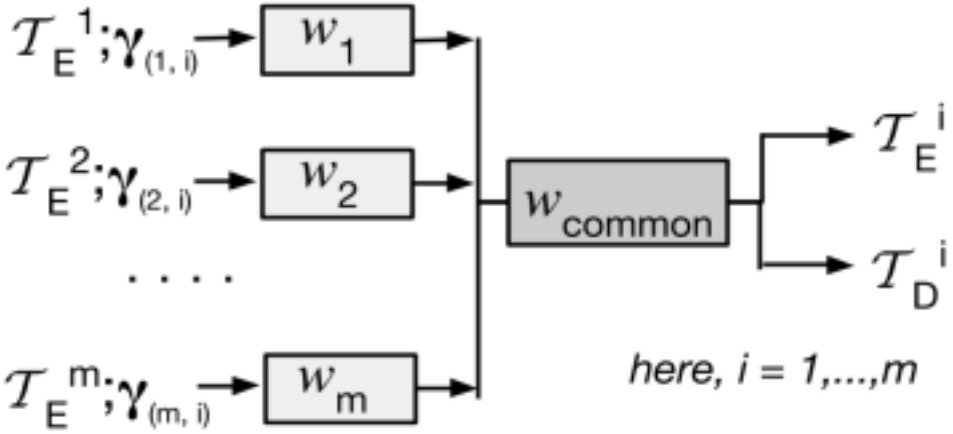


$$\min_W \sum_{i=1}^{m} ||\mathcal{F}\Big((\theta_{E_{\tau_1}}, \gamma_{1,i}), \cdots, (\theta_{E_{\tau_1}}, \gamma_{m,i}); W\Big) - (\theta^*_{E_{\tau_i}}, \theta^*_{D_{\tau_i}})||^2 + \lambda \sum_{\substack{x \in X_{\tau_i} \\ y \in \mathbf{y}_{\tau_i}}} \mathcal{L}\Big(\mathcal{D}_{\tilde{\theta}_{D\tau_i}}(\mathcal{E}_{\tilde{\theta}_{E\tau_i}}(x)), y\Big)$$

Where $\mathcal{L}$ is a loss appropriately chosen for a particular task

Loss Function

- We add a data model consistency loss



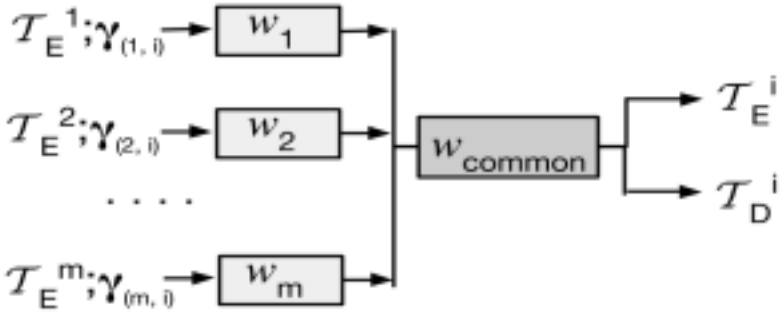
$$\min_W \sum_{i=1}^{m} ||\mathcal{F}\Big((\theta_{E_{\tau_1}}, \gamma_{1,i}), \cdots, (\theta_{E_{\tau_1}}, \gamma_{m,i}); W\Big) - (\theta^*_{E_{\tau_i}}, \theta^*_{D_{\tau_i}})||^2 + \lambda \sum_{\substack{x \in X_{\tau_i} \\ y \in \mathbf{y}_{\tau_i}}} \mathcal{L}\Big(\mathcal{D}_{\tilde{\theta}_{D\tau_i}}(\mathcal{E}_{\tilde{\theta}_{E\tau_i}}(x)), y\Big)$$

Where $\mathcal{L}$ is a loss appropriately chosen for a particular task

Methodology: Other Details

Input:

- P subsets of labeled data obtained from each task
- Model parameters learned for each of p subsets
- For m tasks, we get a p x m data samples as input



Training:

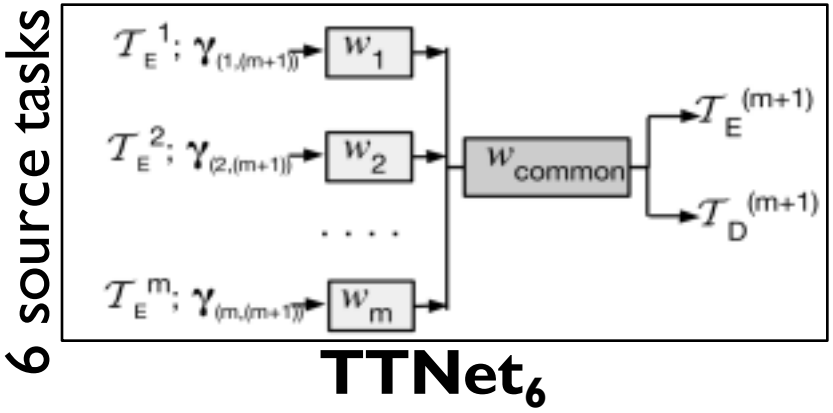
- *Self mode:* Train each W_i individually (and W_{common})
- *Transfer mode:* Train other W_is in the network

Inference:

- Runs in transfer mode for target zero-shot task

Experiments

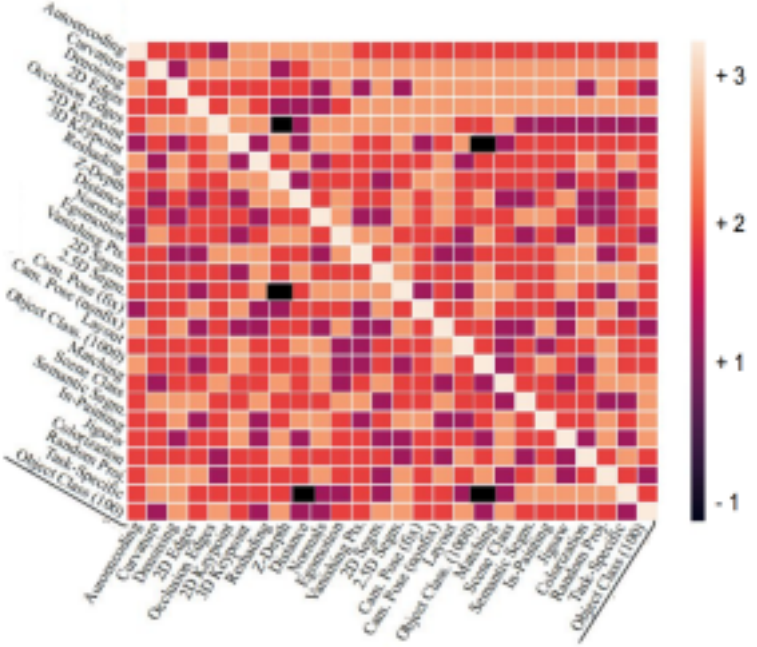
- **Taskonomy** dataset**
  - * 150k RGB data of indoor scenes
- We studied three different TTNets with 6, 10 and 20 source tasks
- Also, considered $TTNet_{LS}$ (fine-tuned on small amount of target task data)



** *Taskonomy, Amir R Zamir et al., CVPR 2018*

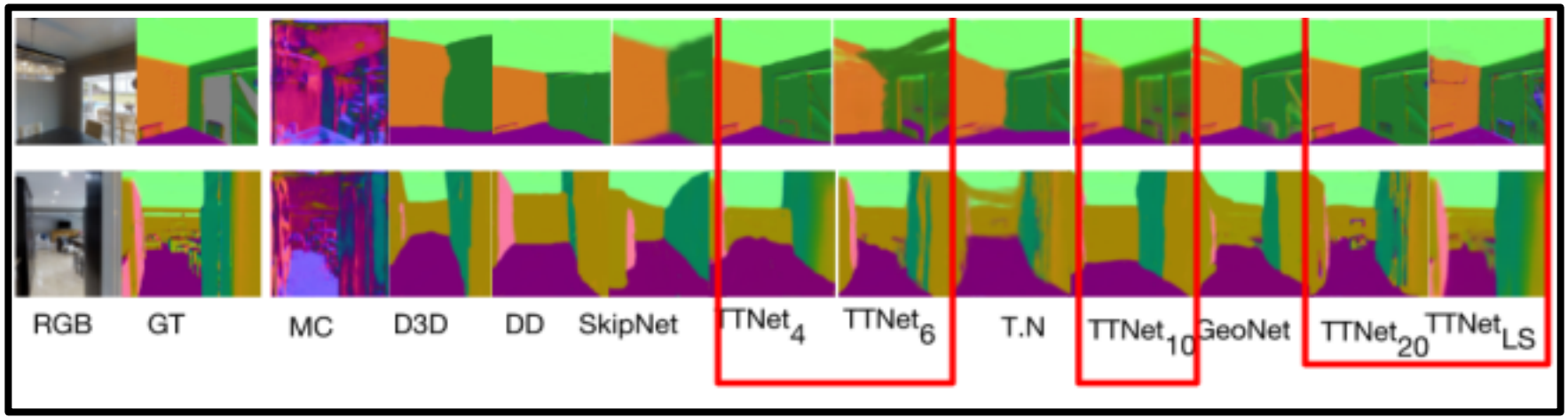
Task Correlation

- Scale of -1 to +2 (negative correlation to strong positive correlation
- Crowd votes aggregated from 30 users using the widely used Dawid-Skene algorithm (EM-based)



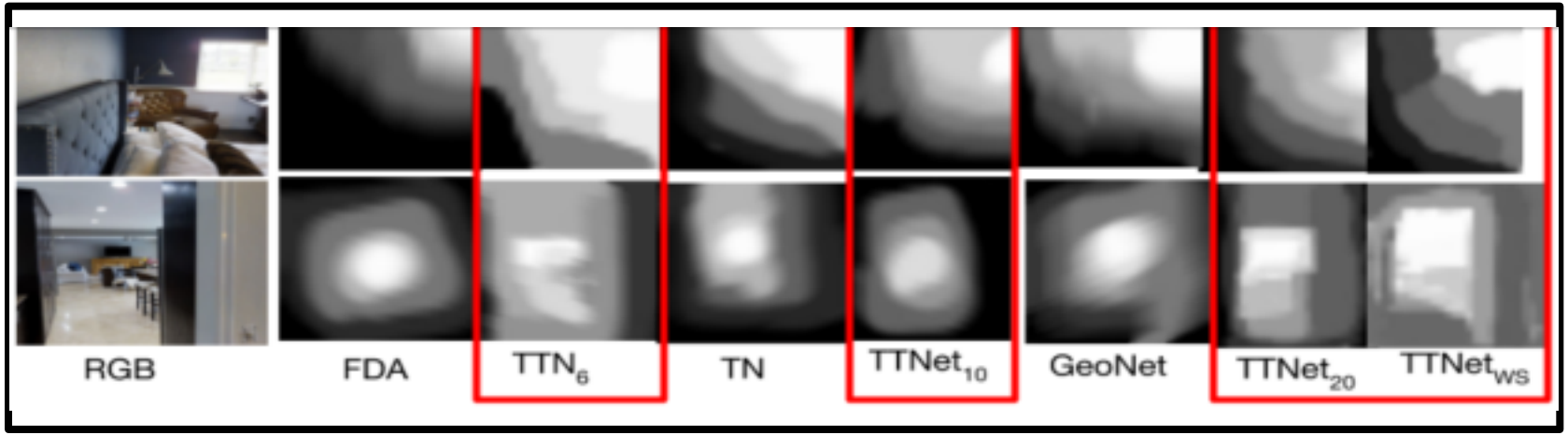
Results

Source (Known) Tasks	Autoencoding, Scene Class, 3D key point, Reshading, Vanishing Point, Colorization
Zero-shot Task	Surface Normal Estimation

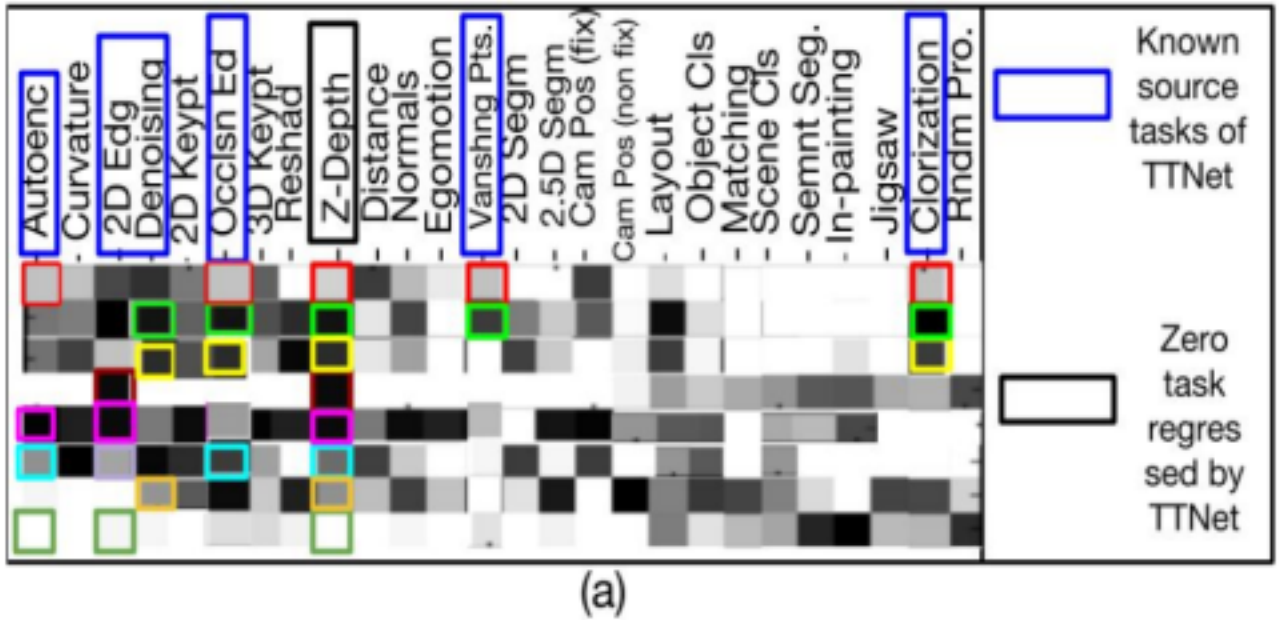


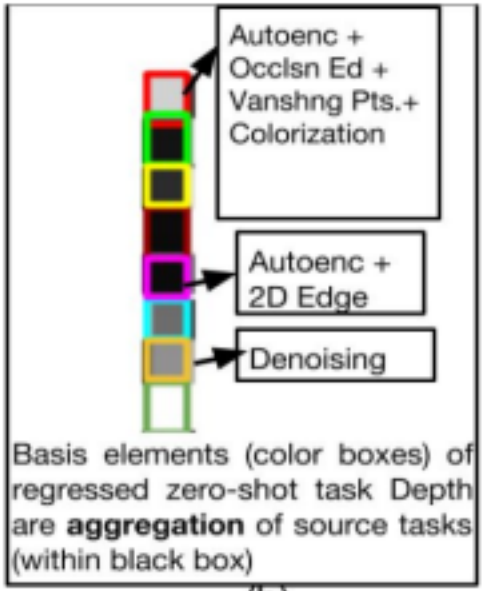
Results

Source (Known) Tasks	Autoencoding, Scene Class, 3D key point, Reshading, Vanishing Point, Colorization
Zero-shot Task	Z-Depth Estimation



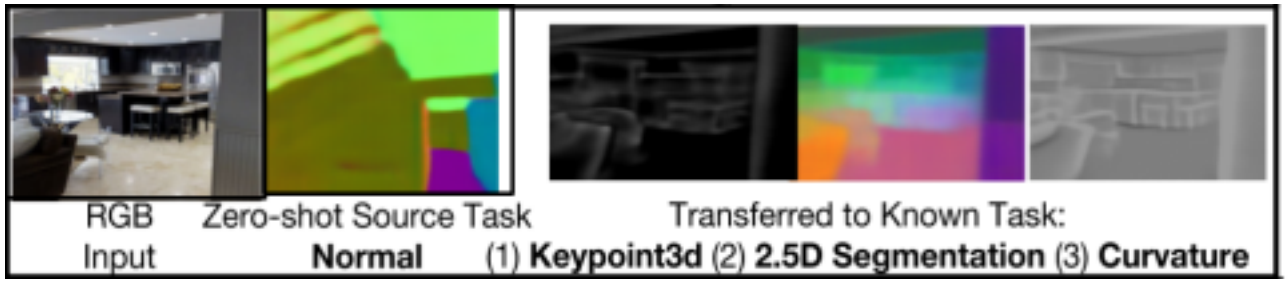
Why better than Supervised Learning?



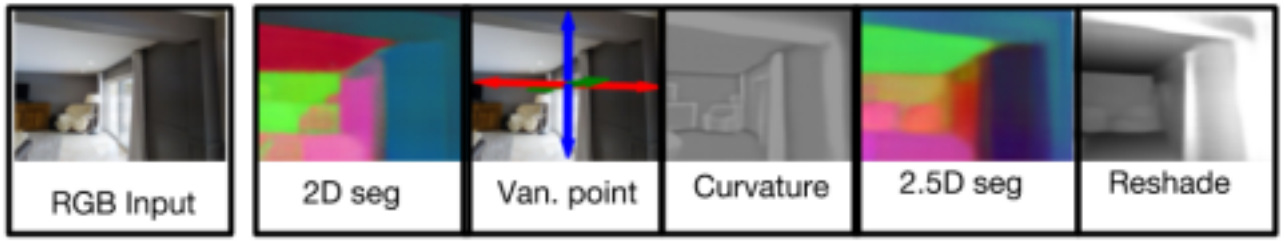


More Results

- Zero-shot to Known Task Transfer



- Different Choice of Zero-Shot Tasks

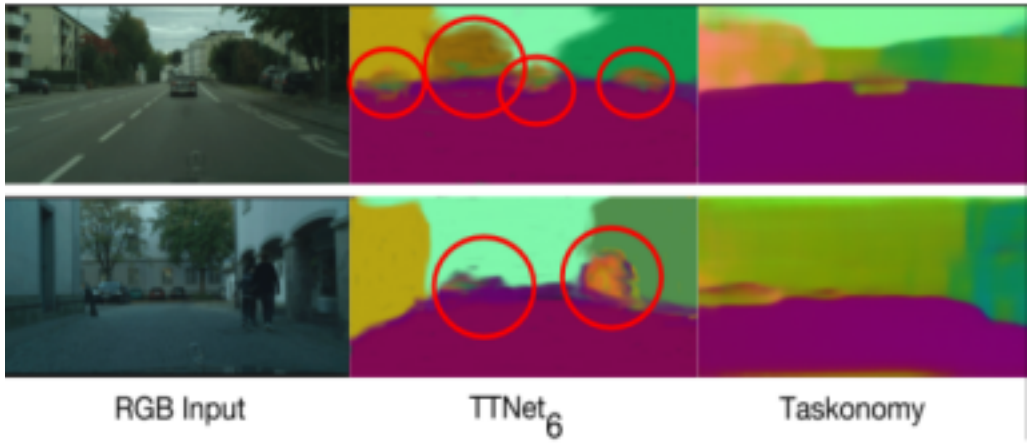


How many source tasks do we need?

	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 Pt.(fixed)* (%)
4	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	79%	62%	71%	71%
	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	71%	58%	61%	59%
	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	75%	79%	79%	52%
6	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	88%	85%	87%	89%
	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	87%	86%	86%	89%
	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	85%	88%	86%	82%
10	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	85%	84%	87%	85%
	✓	✓	✗	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	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%

More Results

- Performance on Other Datasets (Cityscapes)



Questions?

cs15resch11001@iith.ac.in
vineethnb@iith.ac.in
Department of Computer Science and Engineering, IIT-Hyderabad
http://www.iith.ac.in/~vineethnb

Zero shot Task Transfer

Github