Universal Semi-Supervised Semantic Segmentation

A novel approach towards learning efficient and transferable representations for dense prediction tasks

Tarun Kalluri[2] Girish Varma[1] CV Jawahar[1] Manmohan Chandraker[2]

CVIT, IIIT Hyderabad[1] UCSD[2]

Kalluri, T., Varma, G., Chandraker, M., & Jawahar, C. V.. Universal Semi-Supervised Semantic Segmentation. *ICCV 2019*. [PDF]

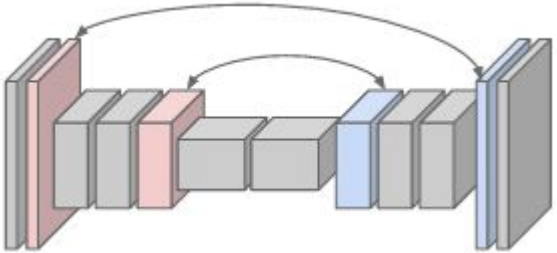
Universal Semi-Supervised Semantic Segmentation

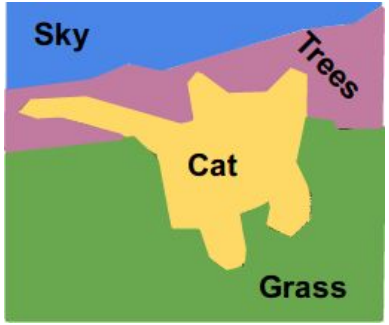
Universal Semi-Supervised Semantic Segmentation

- Semantic segmentation involves pixel-level semantic labeling of images.

- Applications include autonomous driving, remote sensing, satellite imagery and medical imaging.

- Dense predictions require dense annotations.
  - Challenging and tedious task to label.
  - Requires massive human effort.
  - Unlabeled data easier to obtain.





Universal **Semi-Supervised** Semantic Segmentation

- Semi-Supervised learning involves learning with few labeled training data and lots of unlabeled data.

- Unlabeled data assist the labeled data in learning the classifiers.

- Many techniques to do this.
  - Pseudo-Labeling
  - Co-Training
  - Entropy Minimization

Universal **Semi-Supervised** Semantic Segmentation

- Semi-Supervised learning involves learning with few labeled training data and lots of unlabeled data.

- Unlabeled data assist the labeled data in learning the classifiers.

- Many techniques to do this.
  - Pseudo-Labeling - Assign labels to unlabeled set using model trained on labeled set.
  - Co-Training
  - Entropy Minimization

Universal **Semi-Supervised** Semantic Segmentation

- Semi-Supervised learning involves learning with few labeled training data and lots of unlabeled data.
- Unlabeled data assist the labeled data in learning the classifiers.
- Many techniques to do this.
  - Pseudo-Labeling
  - Co-Training - Generate multiple, consistent views of unlabeled data using different models.
  - Entropy Minimization

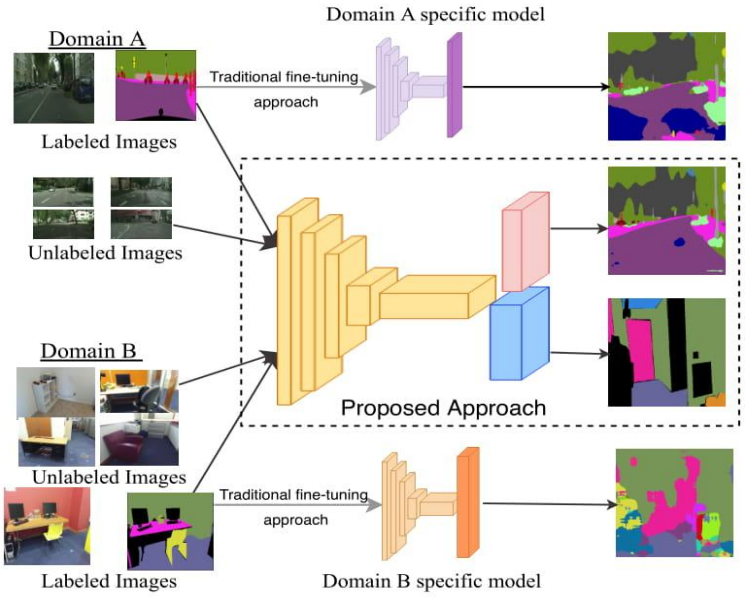
Universal **Semi-Supervised** Semantic Segmentation

- Semi-Supervised learning involves learning with few labeled training data and lots of unlabeled data.
- Unlabeled data assist the labeled data in learning the classifiers.
- Many techniques to do this.
  - Pseudo-Labeling
  - Co-Training
  - Entropy Minimization - Encourages Low density separation in feature space

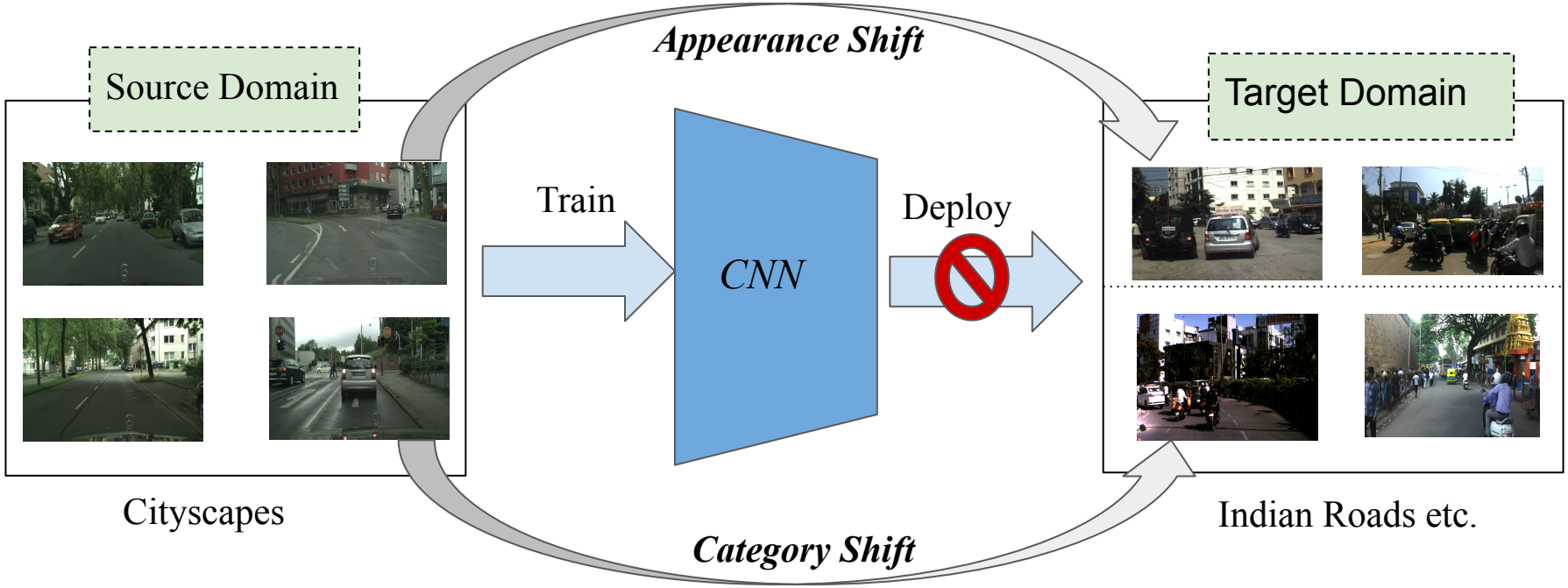
Universal Semi-Supervised Semantic Segmentation

- Idea: Obtain a common semantic segmentation model across widely disparate domains having limited labeled data.

- A good universal model should ensure that, across all domains
    - A single model is deployed
    - Unlabeled data is used
    - Performance is improved
    - And label spaces (semantic content) may differ.



Challenge: Domain Shift + Category Shift



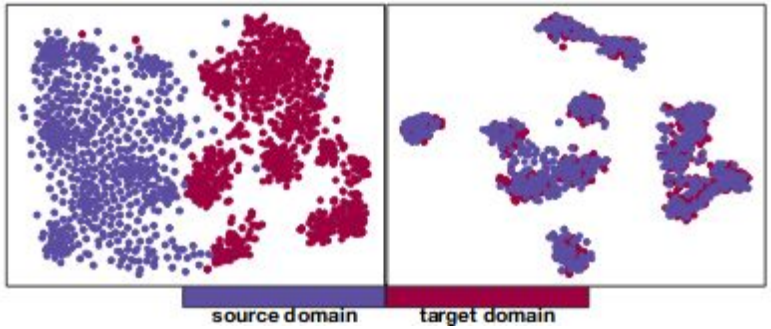
Related Approaches

- Fine-tuning on labeled data.
  - requires separate models for each domain *[catastrophic forgetting]*.
  - Introduces both training and deployment overhead.

Related Approaches: Domain Adaptation

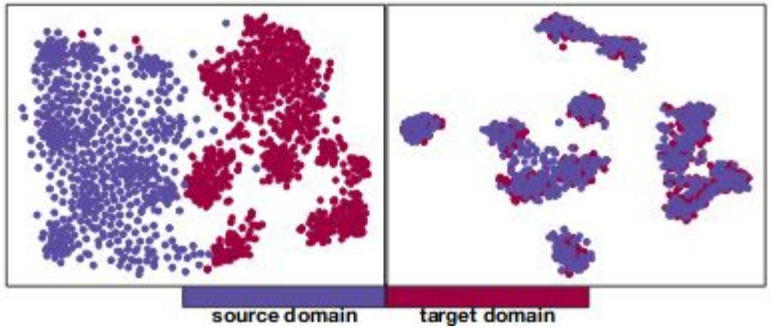
- Unsupervised domain adaptation tackles the problem of domain shift between a labeled source and unlabeled target domain.

- Different ways include
  - Minimizing MMD between datasets.
  - Aligning features from both the domains through adversarial approaches.



Limitations to Domain Adaptation

- Assumption of same label spaces between the source and target domain.
- Model adapted to one target domain does not apply to other domains.



Other Related Approaches

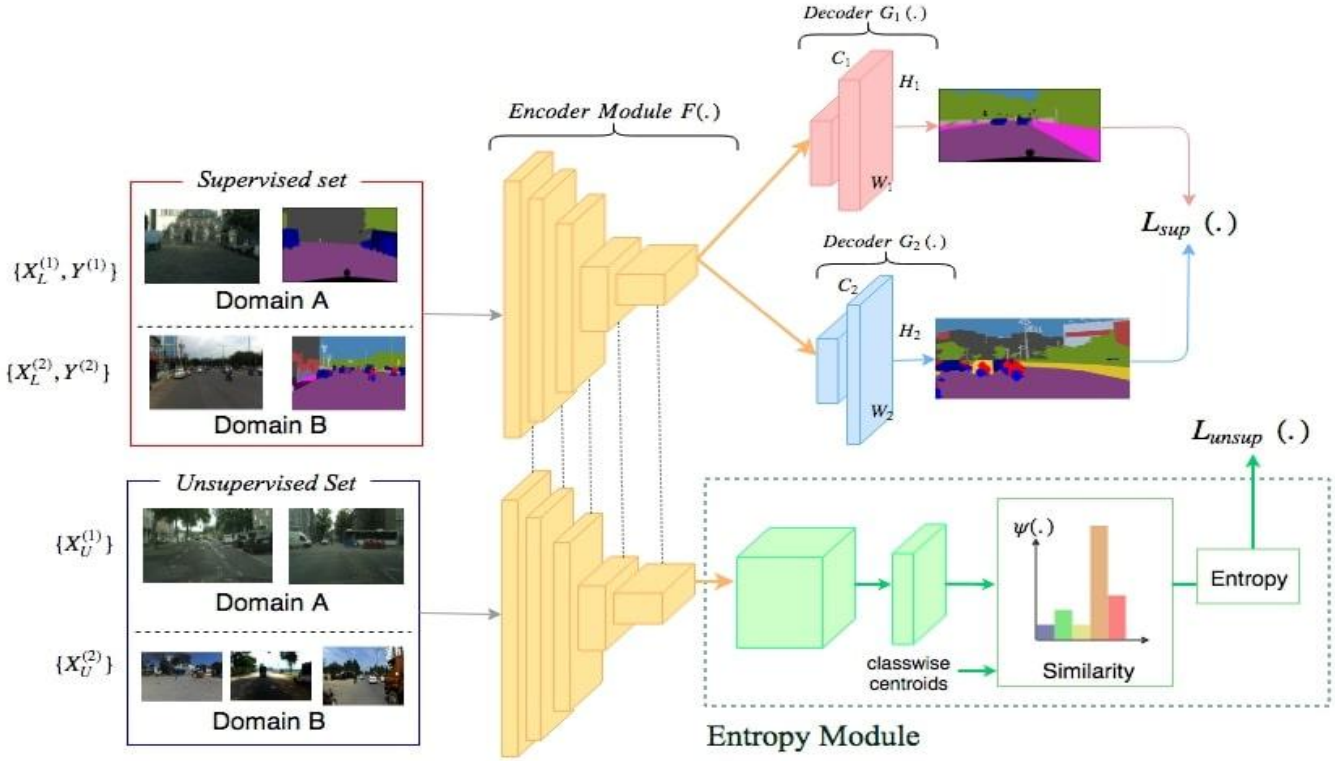
	Source Unlabeled Data	Target Unlabeled Data	Joint Model	Mixed Labels Support
Fine Tuning	✗	✗	✗	✓
Semi-supervised [Hung 2018]	✓	✗	✗	NA
CyCADA [Hoffman 2018]	✗	✓	✓	✗
Joint Training	✗	✗	✓	✓
Our Approach	✓	✓	✓	✓

Prior works fall short in addressing the semantic change, or do not make use of large scale unsupervised images.

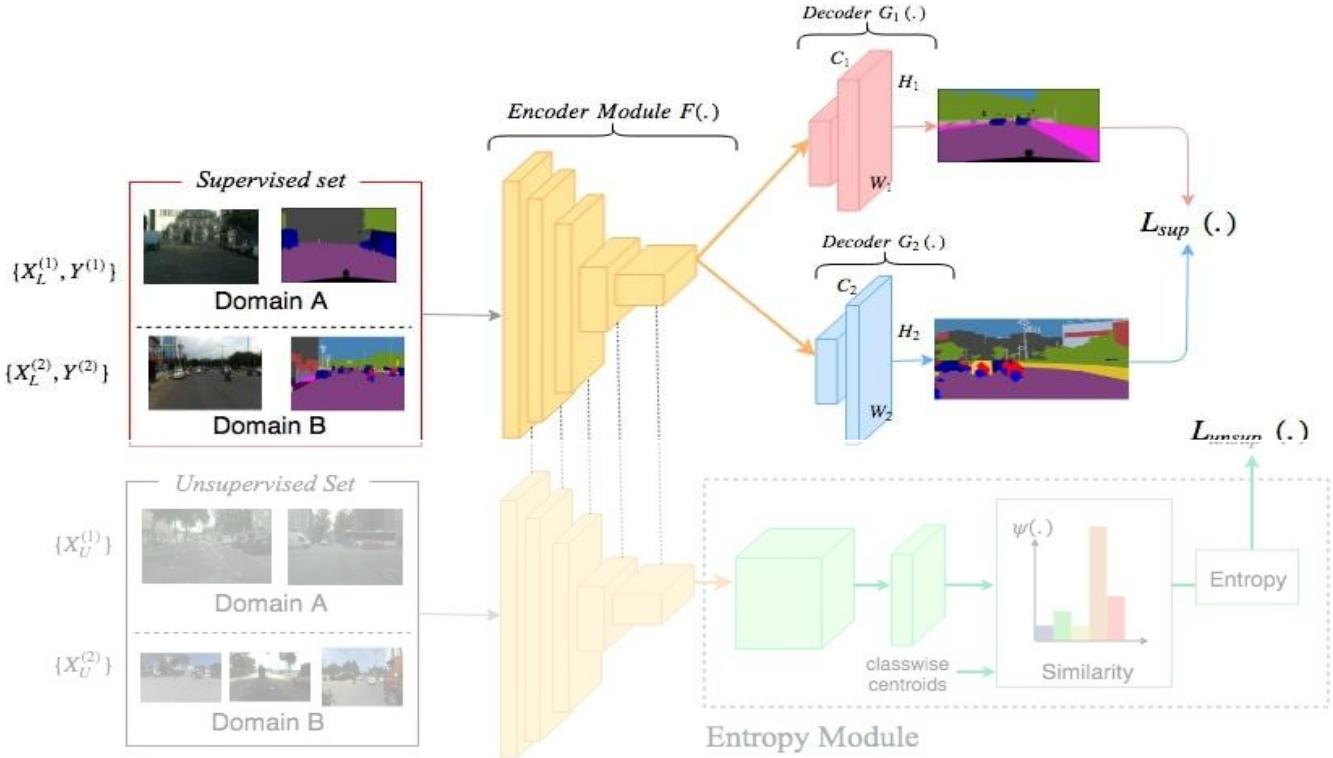
Our Approach

- Use labeled data to build the task specific classifiers.
- Use unlabeled data from each domain to align visually similar features for improved classification.

Proposed Architecture

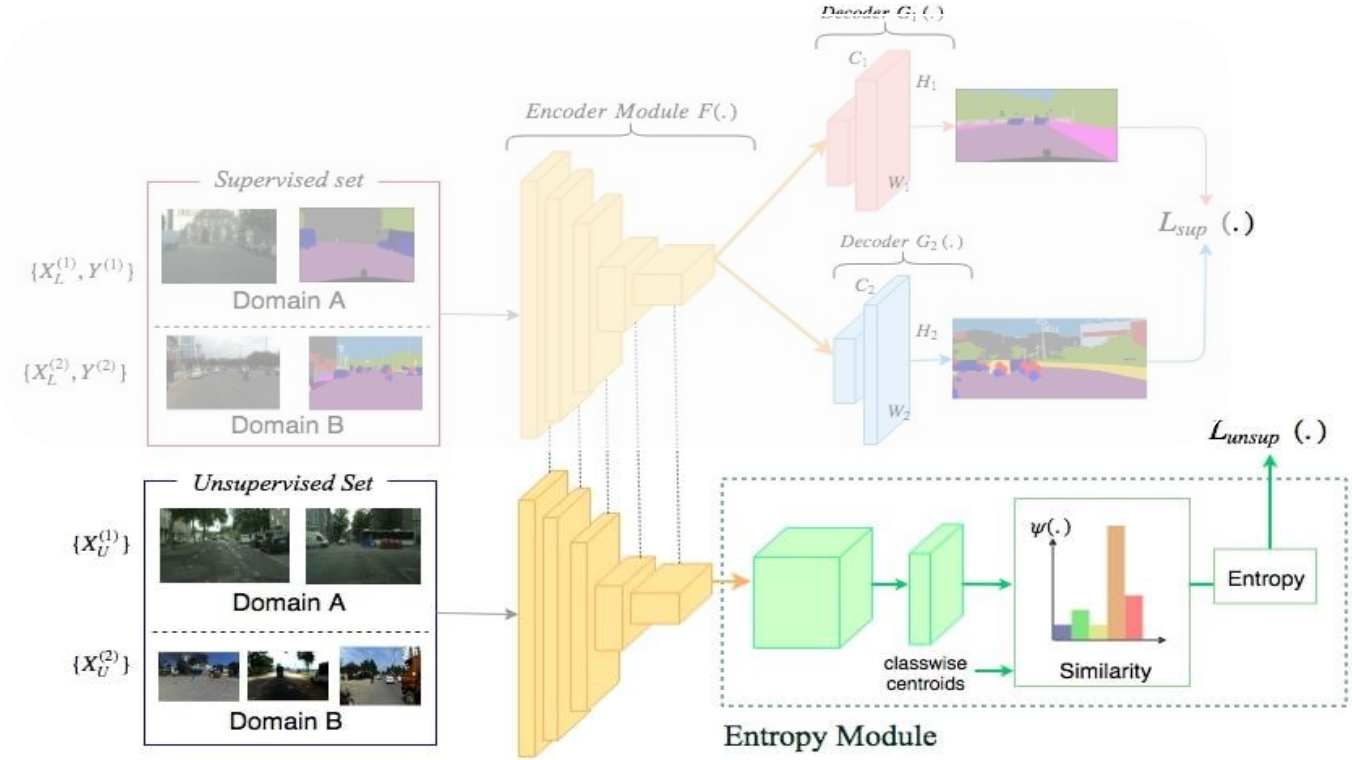


Supervised Loss



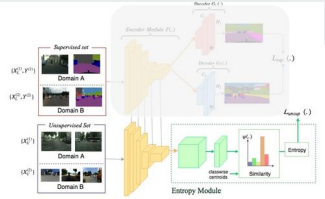
$$L_{sup} = \sum_k \frac{1}{N_l^{(k)}} \sum_{x_i \in D^{(k)}} \psi_k \left(y_i, \mathcal{G}_k \left(\mathcal{F}\left(x_i\right)\right)\right)$$

Semi-Supervised Loss



Entropy Module

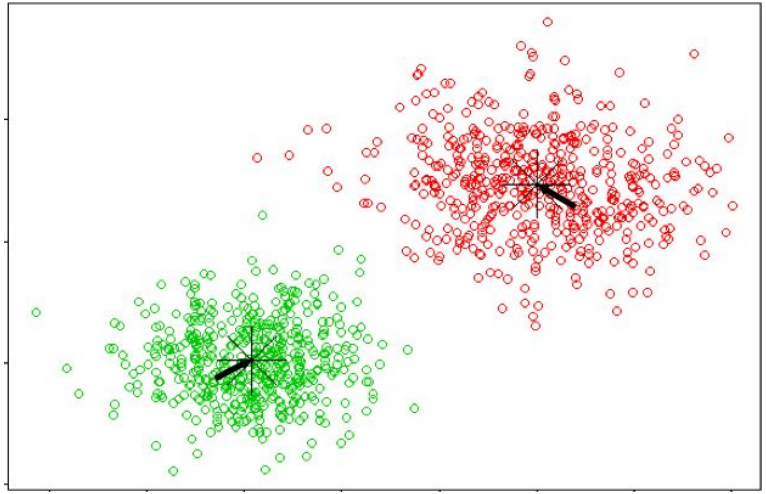
Entropy Regularization



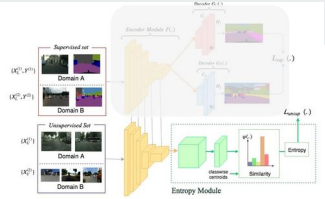
- Entropy regularization (or entropy minimization) encourages low density separation between the clusters in the feature space.

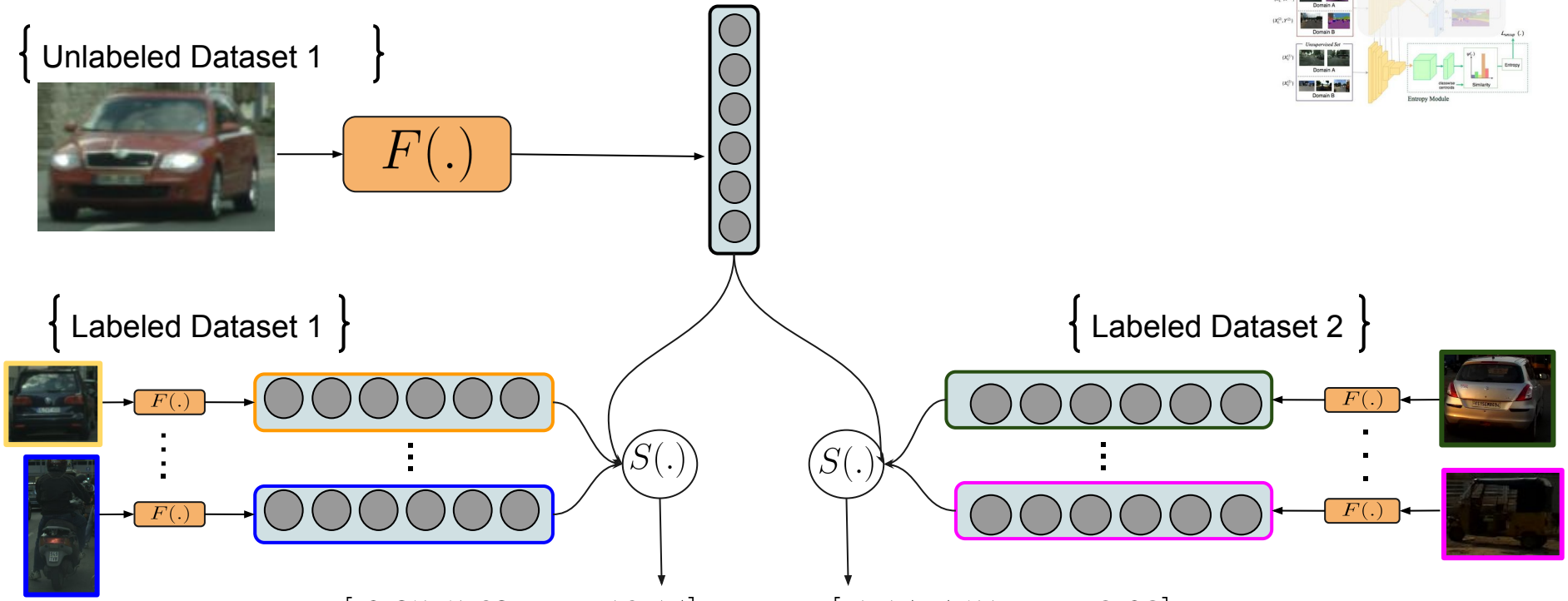
- Results in high confidence predictions and smooth output maps for semi-supervised learning

- We use it to transfer semantic information from unlabeled to labeled images, between both the domains.



Entropy Regularization using similarity matching



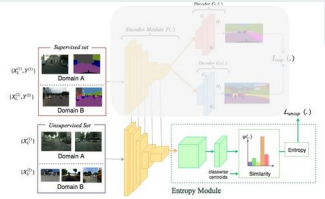


$$v_{11} = [\,0.25, 5.62, \cdots, 19.14]$$

$$[\,1.14, 4.51, \cdots, 8.03] = v_{12}$$

$$H(\sigma(v_{11})) + H(\sigma(v_{12})) \longrightarrow \text{Entropy Loss}$$

Entropy Regularization using similarity matching



- Semantic Segmentation involves feature extraction at pixel level, so entropy matching is done for features computed at each pixel.
- Pixel level similarity with all the images is costly, so we calculate prototype features for each class.
- Dot product is used as the similarity metric $S(.)$
- We perform both within dataset matching and cross dataset matching. This handles disparate label spaces.

Semi-Supervised Entropy Loss

$$L_{u,c} = \mathcal{H}(\sigma([v_{12}])) + \mathcal{H}(\sigma([v_{21}]))$$

$$L_{u,w} = \mathcal{H}(\sigma([v_{11}])) + \mathcal{H}(\sigma([v_{22}]))$$

Total Loss / Training Objective

$$L_t = L_S + \alpha \cdot L_{u,c} + \beta \cdot L_{u,w}$$

Datasets for Comparing the Universal Model

Cityscapes

CamVid

IDD (Indian Roads)

SUN RGB-D (Indoor Scenes)

Experimental Results

Method	Road	SideWalk	Building	Wall	Fence	Pole	Traff. lt.	Traff. Sgn.	Veg.	Train	Sky	Person	Rider	Car	Truck	Bus	Train	MotorCyc.	Bicycle	mIoU
CS only	91.76	**54.78**	80.02	3.70	16.58	29.84	22.31	33.74	**83.88**	32.89	**82.07**	52.67	**21.57**	81.11	**19.01**	3.87	0.0	**19.64**	**49.01**	40.97
Univ-basic	87.00	44.54	77.77	**10.21**	11.07	25.54	14.51	25.82	80.72	22.40	78.19	49.00	19.64	75.35	1.86	0.25	**10.98**	8.83	41.08	36.04
Univ-full	**92.18**	51.29	**80.07**	0.0	**24.01**	**33.73**	**26.16**	**38.71**	82.30	**36.39**	81.61	**54.38**	20.48	**81.71**	2.37	**22.79**	3.85	1.31	46.23	**41.03**

Method	N=375		
	CS	CamVid	Avg.
Train on CS	55.07	48.52	51.80
Train on CVD	26.45	60.61	43.53
Hung *et al.* [28]	58.80	-	-
Souly *et al.* [61]	-	58.20	-
Univ-basic ($\mathcal{L}_s$)	53.14	65.33	59.24
Univ-cross (+ $\mathcal{L}_c$)	56.36	63.34	59.85
Univ-full (+ $\mathcal{L}_c, \mathcal{L}_w$)	55.92	64.72	**60.32**

New SOTA with semi supervised data!

Method	N=100 (Resnet-18)		
	CS	IDD	Avg.
Train on CS	40.97	14.64	27.81
Train on IDD	25.05	26.53	25.79
Univ-basic	37.94	25.21	31.58
Univ-full	36.48	27.45	**31.97**

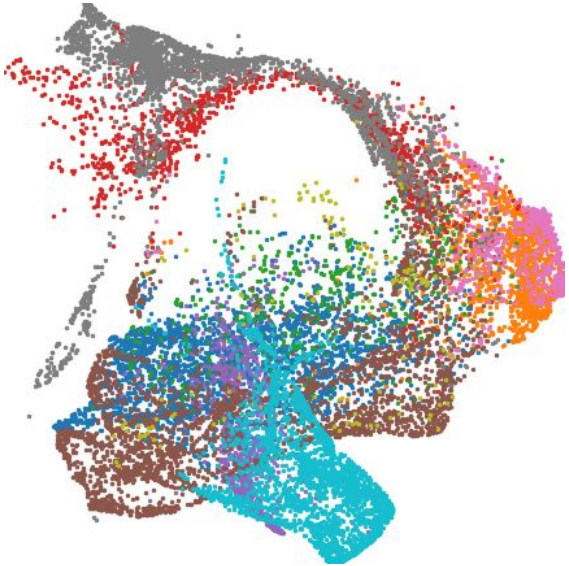
Alignment across environments

- With only 28% labeled data from SUN dataset and no synthetic examples, we obtain ~88% of performance obtained with full dataset.

Method	Labeled Examples	CS	SUN	Avg.
Train on CS	1.5k	64.23	15.47	39.85
Train on SUN	1.5k	15.61	42.52	29.07
SceneNet [41]	Full(5.3k)	-	49.8	-
Univ-basic	1.5k	58.01	31.55	44.78
Ours[Univ-full]	1.5k	57.91	43.12	**50.52**

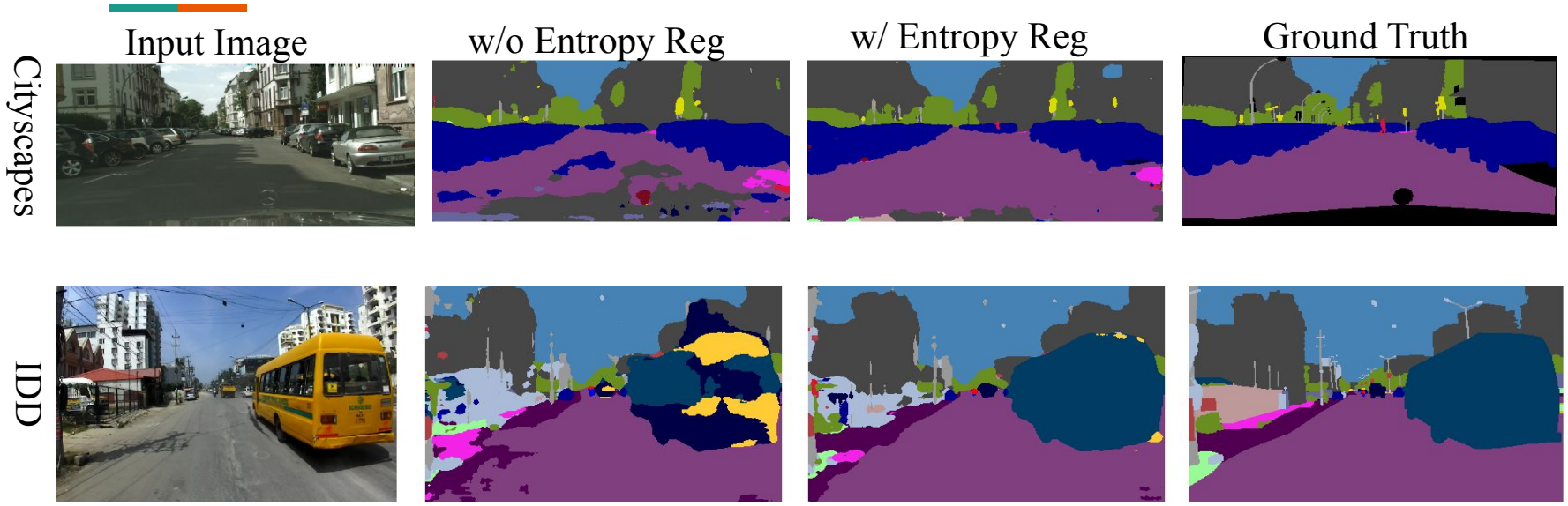
tSNE Embedding Visualization

(a) Without Entropy Module



(b) With Entropy Module

Qualitative Results



Limitations of the proposed approach

- **Negative alignment:** Feature alignment useful for aligning visually similar features, but classes which are private to a dataset might get wrongly aligned to an unrelated class.

- **Label Prototypes**: Entropy regularization depends largely on the label prototypes, which are unreliable with limited data. Should automatically compute label clusters.

Conclusions and future research

- Presented a method to train joint models which are universally applicable to multiple datasets.
- Entropy regularization helps in within dataset and cross dataset semantic transfer and feature alignment using unsupervised data.
- Future work includes extending the current setting to domain generalization and zero-shot target adaptation.

Thank You!

Any Questions?