Universal Semi-Supervised Semantic Segmentation

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

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

Models trained on a single domain are not usable in other domains due to Domain Shift and Semantic Shift.
Training individual models for different domains results in deployment overhead, doesn’t exploit shared structure among these domains.

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

New SOTA with semi supervised data!

Datasets

Cityscapes (N=2975) + CamVid (N=366) + IDD (Indian Roads) (N = 6993) + SUN (Indoor) (N = 5050)

28% labeled data from SUN RGB dataset with no synthetic examples, recovers ~88% of performance obtained with full dataset.

Cityscapes
Cityscapes
Cityscapes
Cityscapes

Prior works fall short in addressing the semantic change, which we do by using large scale unsupervised images.

Qualitative Improvements In Segmentation

Visual similar features, like Building and SideWalk from Cityscapes and CamVid are positively aligned, helping in learning agnostic discriminative features.

Approach: Feature Alignment Using Entropy Regularization

Training Objective: Supervised + Unsupervised Losses

Unsupervised Losses

\[ L_{U-A} = H(\sigma([c])^A) + H(\sigma([c])^B) \]

\[ L_{U-B} = H(\sigma([c])^A) + H(\sigma([c])^B) \]

Supervised Loss

\[ L_{SP} = \frac{1}{N} \sum_{i=1}^{N} \psi_k(y_i, \hat{y}_i(F(x_i))) \]

Total Loss

\[ L = L_{SP} + \lambda_1 \cdot L_{UA} + \lambda_2 \cdot L_{UB} \]

New SOTA with semi supervised data!

Datasets

Cityscapes (N=2975) + CamVid (N=366) + IDD (Indian Roads) (N = 6993) + SUN (Indoor) (N = 5050)

28% labeled data from SUN RGB dataset with no synthetic examples, recovers ~88% of performance obtained with full dataset.

Qualitative Improvements In Segmentation

Visual similar features, like Building and SideWalk from Cityscapes and CamVid are positively aligned, helping in learning agnostic discriminative features.

Approach: Feature Alignment Using Entropy Regularization

Training Objective: Supervised + Unsupervised Losses

Unsupervised Losses

\[ L_{U-A} = H(\sigma([c])^A) + H(\sigma([c])^B) \]

\[ L_{U-B} = H(\sigma([c])^A) + H(\sigma([c])^B) \]

Supervised Loss

\[ L_{SP} = \frac{1}{N} \sum_{i=1}^{N} \psi_k(y_i, \hat{y}_i(F(x_i))) \]

Total Loss

\[ L = L_{SP} + \lambda_1 \cdot L_{UA} + \lambda_2 \cdot L_{UB} \]

New SOTA with semi supervised data!

Datasets

Cityscapes (N=2975) + CamVid (N=366) + IDD (Indian Roads) (N = 6993) + SUN (Indoor) (N = 5050)

28% labeled data from SUN RGB dataset with no synthetic examples, recovers ~88% of performance obtained with full dataset.

Qualitative Improvements In Segmentation

Visual similar features, like Building and SideWalk from Cityscapes and CamVid are positively aligned, helping in learning agnostic discriminative features.

Approach: Feature Alignment Using Entropy Regularization

Training Objective: Supervised + Unsupervised Losses

Unsupervised Losses

\[ L_{U-A} = H(\sigma([c])^A) + H(\sigma([c])^B) \]

\[ L_{U-B} = H(\sigma([c])^A) + H(\sigma([c])^B) \]

Supervised Loss

\[ L_{SP} = \frac{1}{N} \sum_{i=1}^{N} \psi_k(y_i, \hat{y}_i(F(x_i))) \]

Total Loss

\[ L = L_{SP} + \lambda_1 \cdot L_{UA} + \lambda_2 \cdot L_{UB} \]