Attending to Discriminative Certainty for Domain Adaptation

Vinod Kumar Kurmi*, Shanu Kumar*, Vinay P. Namboodiri

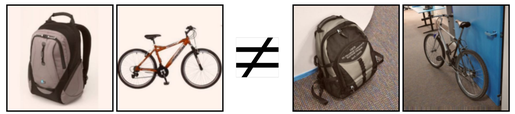
Indian Institute of Technology Kanpur, India

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Introduction

We aim to solve the unsupervised domain adaptation problem by reducing the domain shift between the domains where the source domain have labeled data but the target domain consists of unlabeled data.

The proposed method identifies the adaptable regions using the certainty of the discriminator, and specifies the focus of classifier on these regions.



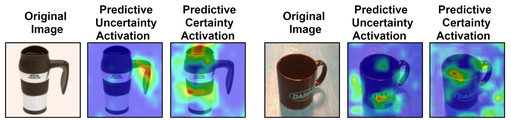
Source Domain labeled data ≠ **Target Domain** unlabeled data

Bayesian Classifier

Aleatoric and Predictive uncertainty are estimated for the classification task by Bayesian modeling of the classifier using the approach proposed by Gal et. al. The Bayesian classifier predicts the class label along with its variance.

$$y_i^c = G_{cy}(G_c(\mathbf{f_i})), \qquad v_i^c = G_{cv}(G_c(\mathbf{f_i}))$$

Estimation of uncertainty, makes features more robust and provides a tool for visualizing both the certain and uncertain regions for the classifier.



Bayesian Discriminator

By estimating uncertainty, the discriminator avoids the learning of noisy regions which can corrupt the feature representations. This also results in reduction of negative transfer during adaptation.

$$g_{i,t} = G_d^t(\mathbf{f_i}), \quad G_d^t \sim \mathbf{G_d}$$

$$p(y_{i,t}^d) = \text{Softmax}(G_{dy}(g_{i,t})), \quad v_{i,t}^d = G_{dv}(g_{i,t})$$

Discriminator Certainty Based Attention

Aleatoric Certainty based Attention

The aleatoric uncertainty of the domain discriminator occurs due to corruption or noise in the regions. These regions are not suited for the adaptation as well as for object classification and should be less focused as compared to the certain regions for the classification task.

$$p_i = f_i * -\frac{\partial v_i^d}{\partial f_i}$$

Predictive Certainty based Attention

Predictive uncertainty occurs in the regions which are either already domain invariant or which contain noise. The regions corresponding to discriminator's predictive certainty are domain transferable regions and should be attended during classification.

$$H(y_{i,t}^d) = -\sum_{c}^{2} p(y_{i,t}^d = c) * \log(p(y_{i,t}^d = c))$$

$$\text{Var}^d(f_i) \approx \frac{1}{T}\sum_{t=1}^{T}\left(v_{i,t}^d + H(y_{i,t}^d)\right) \qquad p_i = f_i * -\frac{\partial \text{Var}^d(f_i)}{\partial f_i}$$

Training Objective

For obtaining the regions where the discriminator is certain, the product p_i is passed through a ReLU function, but for ignoring the negative values that correspond to uncertain regions, $-p_i$ is again passed through a ReLU function.

$$a_i = \text{ReLU}(p_i) + c * \text{ReLU}(-p_i)$$

$$w_i = (1 - \text{Var}_i^d(f_i)) * \text{Softmax}(a_i) \qquad h_i = f_i * (1 + w_i)$$

The class probabilities and uncertainty for the classifier will be estimated using attended visual features h_i.

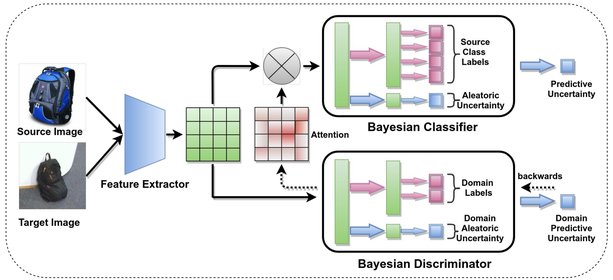
$$y_i^c = G_{cy}(G_c(\mathbf{h_i})), \qquad v_i^c = G_{cv}(G_c(\mathbf{h_i}))$$

The final objective function for optimizing both the models is defined as:

$$J = \mathcal{L}_{cy} + \mathcal{L}_{cv} - \lambda * (\mathcal{L}_{dy} + \mathcal{L}_{dv})$$

$$(\hat{\theta}_f, \hat{\theta}_c, \hat{\theta}_{cy}, \hat{\theta}_{cv}) = \underset{\substack{\theta_f,\theta_c,\\ \theta_{cy},\theta_{cv}}}{\arg\min}\, J(\theta_f, \theta_c, \theta_{cy}, \theta_{cv}, \theta_d, \theta_{dy}, \theta_{dv})$$

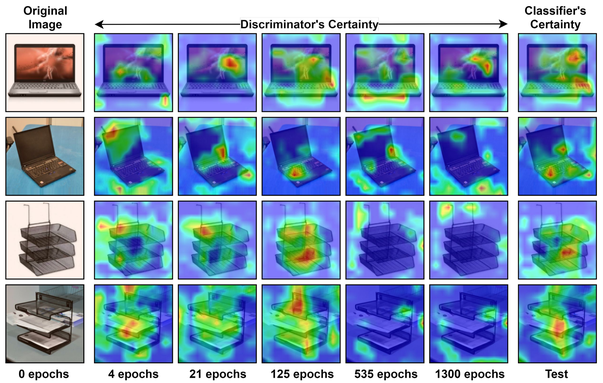
$$(\hat{\theta}_d, \hat{\theta}_{dy}, \hat{\theta}_{dv}) = \underset{\theta_d,\theta_{dy},\theta_{dv}}{\arg\max}\, J(\theta_f, \theta_c, \theta_{cy}, \theta_{cv}, \theta_d, \theta_{dy}, \theta_{dv})$$



CADA: Certainty based Attention for Domain Adaptation

Visualization of Attention Maps

We provide the certainty map of the discriminator at different training stages. The rightmost image represents the classifier's certainty map without the guidance from the attention generated using the discriminator. The results show that these attention weights based training aids the classifier to better generalize to the target domain.



Results

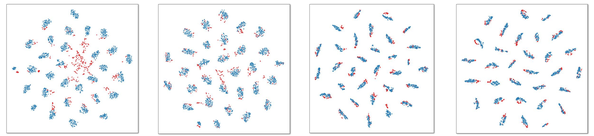
Classification accuracy (%) on Office-31 dataset for ResNet-50

Method	A → W	D→ W	W → D	A → D	D → A	W → A	Avg
ResNet-50	68.4±0.2	96.7±0.1	99.3±0.1	68.9±0.2	62.5±0.3	60.7± 0.3	76.1
ADDA	86.2±0.5	96.2±0.3	98.4±0.3	77.8±0.3	69.5±0.4	68.9±0.5	82.9
MADA	90.0± 0.1	97.4 ±0.1	99.6± 0.1	87.8±0.2	70.3±0.3	66.4±0.3	85.2
GTA	89.5±0.5	97.9±0.3	99.8±0.4	87.7±0.5	72.8±0.3	71.4±0.4	86.5
DAAA	86.8±0.2	99.3±0.1	100.0±0.0	88.8±0.4	**74.3±0.2**	**73.9 ±0.2**	87.2
CDAN	93.1±0.1	98.6±0.1	100.0±0.0	93.4±0.2	71.0±0.3	70.3 ±0.3	87.7
CADA-W	93.9± 0.1	99.1 ± 0.2	99.6± 0.2	93.2± 0.3	68.9± 0.1	68.3± 0.2	87.2
CADA-A	96.8 ± 0.2	99.0±0.1	99.8 ±0.1	93.4±0.1	71.7±0.2	70.5±0.3	88.5
CADA-P	**97.0 ± 0.2**	**99.3 ±0.1**	**100.0±0**	**95.6±0.1**	71.5±0.2	73.1±0.3	**89.5**

Classification accuracy (%) on Office-Home dataset for ResNet-50

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet-50	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DAN	43.6	57.0	67.9	45.8	56.5	60.4	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
JAN	45.9	61.2	68.9	50.4	59.7	61.0	45.8	43.4	70.3	63.9	52.4	76.8	58.3
CDAN	50.6	65.9	73.4	55.7	62.7	64.2	51.8	49.1	74.5	68.2	56.9	80.7	62.8
CADA-A	**56.9**	75.4	80.2	**61.7**	74.6	74.9	62.9	54.4	**80.9**	**74.3**	61.1	**84.4**	70.1
CADA-P	**56.9**	**76.4**	**80.7**	61.3	**75.2**	**75.2**	**63.2**	**54.5**	80.7	73.9	**61.5**	84.1	**70.2**

Feature Visualization



ResNet DANN CADA-A CADA-P

Conclusion

We propose the use of certainty estimates of the discriminator to aid the generalization of the classifier by specifying the attention of the classifier on these regions. The attention maps obtained through certainty agree well with the classifier certainty for true labels which aids in the generalization of the classifier for the target domain as well.

Project Page: https://delta-lab-iitk.github.io/CADA/
Contact: vinodkk@iitk.ac.in, sshanukr@gmail.com