

Supplementary: Semi-Separated Uncertainty Adversarial Learning for Universal Domain Adaptation

Zhongyi Han
Wan Su
Rundong He
Yilong Yin

School of Software, Shandong University

HANZHONGYICN@GMAIL.COM
 SUWAN@MAIL.SDU.EDU.CN
 RUNDONG_HE@MAIL.SDU.EDU.CN
 YLYIN@SDU.EDU.CN

1. Uncertainty Metrics

Differing from the existing UniDA methods that only consider single or few uncertainty measures, we consider all known uncertainty measures: energy, entropy, confidence, consistency, distance, and similarity.

1.1. Energy

Energy is used to estimate the physical properties and the outlier degrees of target samples (Liu et al., 2020). Considering the ensemble classifiers, we define it by

$$z_{\text{ener}}(\mathbf{x}_t) = -\frac{\tau}{m} \sum_{i=1}^m \log \sum_{j=1}^K e^{G_i(\mathbf{x}_t)/\tau}, \quad (1)$$

where τ denotes the temperature parameter. The higher the negative energy of \mathbf{x}_t , the higher the uncertainty of \mathbf{x}_t and the more likely it belongs to the open classes. However, we find that energy is always dominated by the maximum logit and ignores the values of non-maximum logits. To mitigate this issue, we introduce entropy for the comprehensive consideration of output probability distributions as follows.

1.2. Entropy

Entropy is used to measure the prediction uncertainty and smoothness of class distribution from the perspective of information theory. For each target instance \mathbf{x}_t , entropy is defined by

$$z_{\text{entr}}(\mathbf{x}_t) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^K [\tilde{\eta} \circ G_i(\mathbf{x}_t)]_j \log [\tilde{\eta} \circ G_i(\mathbf{x}_t)]_j, \quad (2)$$

where $\tilde{\eta}$ denotes the softmax function and symbol \circ denotes function composition. The higher the entropy of \mathbf{x}_t , the more likely \mathbf{x}_t is belonging to the open classes with large uncertainty. Through minimizing the entropy separation loss, the entropy values of common classes and open classes are highly distinguishable. Entropy exhibits low discriminability

for highly uncertain and extremely sharp predictions (Fu et al., 2020), *i.e.*, the relative difference of entropy values between sharp predictions is very small when the target class number is large. To remedy this problem, we adopt confidence to measure the non-smooth distributions as follows.

1.3. Confidence

Confidence indicates how affirmative the estimated label \mathbf{y}_t of \mathbf{x}_t is. For each target instance \mathbf{x}_t , the corresponding confidence is defined by

$$z_{\text{conf}}(\mathbf{x}_t) = \frac{1}{m} \sum_{i=1}^m \max(\tilde{\eta} \circ G_i(\mathbf{x}_t)). \quad (3)$$

The lower the maximum confidence, the more likely \mathbf{x}_t is an open-class instance with large uncertainty. When training, the selection of unlabeled target samples with high confidence predictions can move decision boundaries to low-density regions. However, some of these selected samples are misplaced because the confidence only pays attention to the largest probability, but which is easily misled by poor-calibrated neural networks. To calibrate the over-confidence values, we introduce consistency between multiple diverse ensemble classifiers.

1.4. Consistency

Consistency is more robust to prediction errors because all ensemble classifiers have a low probability of making the same error (Fu et al., 2020). For each target instance \mathbf{x}_t , the corresponding consistency is defined by

$$z_{\text{cons}}(\mathbf{x}_t) = \frac{1}{K} \left\| \frac{1}{m} \sum_{i=1}^m (\tilde{\eta} \circ G_i(\mathbf{x}_t) - \frac{1}{m} \sum_{i=1}^m \tilde{\eta} \circ G_i(\mathbf{x}_t))^2 \right\|_1. \quad (4)$$

Consistency is higher for a more certain target data point in open classes. The above four uncertainties only measure the openness of target samples from the perspective of prediction results. We introduce the feature distance and domain similarity from the perspective of feature representation.

1.5. Distance

Distance is used to measure the the Euclidean distance between the features of target samples and the cluster centroids $\boldsymbol{\mu}$ of source examples. For each target instance \mathbf{x}_t with feature \mathbf{f} , the corresponding distance is defined by

$$z_{\text{dist}}(\mathbf{x}_t) = \min\{\|\mathbf{f} - \boldsymbol{\mu}_1\|_2, \dots, \|\mathbf{f} - \boldsymbol{\mu}_K\|_2\}. \quad (5)$$

Distance is higher for the target instance in open classes. We choose the minimum distance as the uncertainty quantity because it is more like an open class if a target instance is far from the most similar class in the source domain.

1.6. Similarity

Similarity is used to estimate how far away each target instance is from the source domain. For each target instance \mathbf{x}_t with feature \mathbf{f} , its similarity is defined by

$$z_{\text{simi}}(\mathbf{x}_t) = [\tilde{\eta} \circ D(\mathbf{f})]_1, \tag{6}$$

where $[\cdot]_1$ denote the 1-th index of the softmax output. The lower the similarity, the more likely \mathbf{x}_t is belonging to the open classes with large probability.

In summary, we integrate six active uncertainty measures that contribute to the accurate detection of target open-class samples from the perspectives of prediction results and feature representation, respectively. The above analysis shows that the six uncertainty measures can promote each other in complementary. Thus they can further represent the attributes of target samples to learn the transferable decision tree.

2. Datasets

Office-31 (Saenko et al., 2010) is a standard UDA dataset consisting of three distinct domains, **Amazon** from the Amazon website, **Webcam** by the web camera, and **DSLR** by digital SLR camera. It has 4,652 images with 31 unbalanced classes (see Figure 1).



Figure 1: Some example images in the Office-31 dataset.

Office-Home (Venkateswara et al., 2017) is a more challenging domain adaptation dataset consisting of 15,599 images with 65 unbalanced classes. It consists of four more distinct domains: **Artistic** images, **Clip Art**, **Product** images, and **Real-world** images (see Figure 2).

DomainNet (Peng et al., 2019) is the most challenging domain adaptation dataset consisting of six different domains: **clipart** collected from clipart images, **real** collected from photo-realistic or real-world images, **sketch** collected from the sketches of specific objects, **infographic** images with specific object, **painting** artistic depictions of objects in the form of paintings and quickdraw, and **quickdraw** collected from the drawings of game (see Figure 3). This dataset contains about 600,000 images distributed in 345 categories. In the alphabet order, we use the first 150 classes as common classes \mathcal{Y} , the next 50 classes as $\bar{\mathcal{Y}}^s$, and the



Figure 2: Some example images in the Office-Home dataset.

rest as \bar{Y}^t following (Fu et al., 2020). We choose three domains to construct six transfer tasks due to the large amount of data.

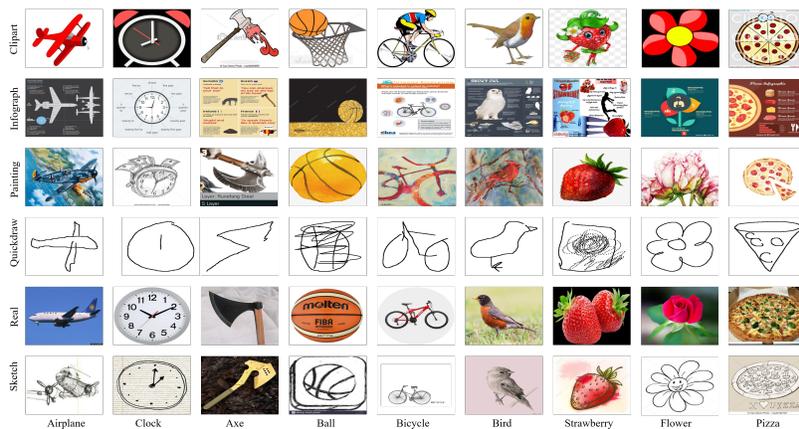


Figure 3: Some example images in the DomainNet dataset.

VisDA-2017 (Peng et al., 2017) is a simulation-to-real dataset with two extremely distinct domains: **Synthetic** 2D renderings of 3D models generated from different angles and with different lighting conditions, and **Real** collected from photo-realistic or real-world image datasets (see Figure 4). With 280,000 images in 12 classes, the scale of VisDA-2017 brings challenges to domain adaptation.

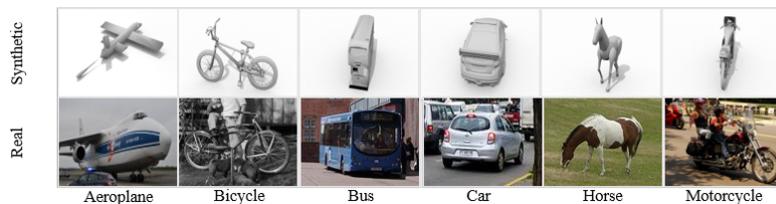


Figure 4: Some example images in the VisDA-2017 dataset.

3. Result

3.1. Semi-separated Uncertainty Decision-maker

While extensive experimental results have proved the strengths of our semi-separated uncertainty decision-maker in target open-class detection, we provide a broader spectrum for more in-depth analysis. Figure 5 visualizes the learning results of semi-separated uncertainty decision-maker on many UniDA tasks. We can observe that semi-separated uncertainty decision-makers can adaptatively discover multiple uncertainties' threshold parameters and importance orders. Interestingly, entropy occupies most roots of these semi-separated uncertainty decision-makers. The importance orders of other uncertainty measures meet our expectations. These results indicate that the performance of semi-separated uncertainty decision-makers is robust and is not susceptible to different tasks.

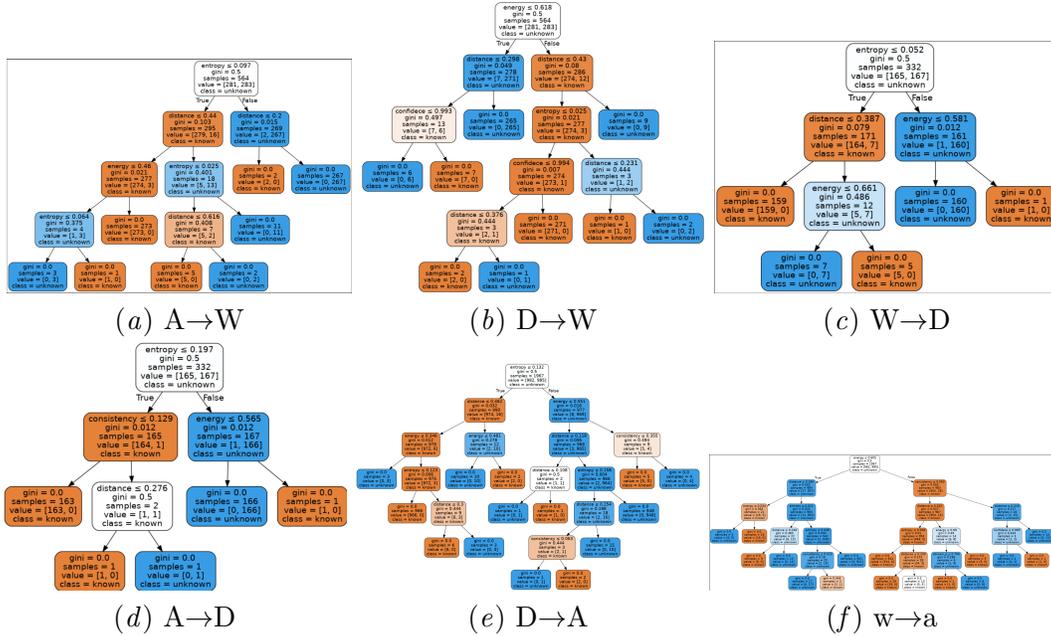


Figure 5: The visualization of learned transferable decision trees on the six tasks of **Office-31**.

References

Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XV*, volume 12360 of *Lecture Notes in Computer Science*, pages 567–583. Springer, 2020. doi: 10.1007/978-3-030-58555-6_34.

Weitang Liu, Xiaoyun Wang, John D. Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *Advances in Neural Information Processing Systems 33: Annual*

Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.

Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *CoRR*, abs/1710.06924, 2017.

Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 1406–1415. IEEE, 2019. doi: 10.1109/ICCV.2019.00149.

Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010.

Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 5018–5027, 2017.