# Distance-based Hyperspherical Classification for Multi-source Open-Set Domain Adaptation

**Silvia Bucci*, Francesco Cappio Borlino*, Barbara Caputo, Tatiana Tommasi**
Politecnico di Torino, Italy, Italian Institute of Technology
{silvia.bucci, francesco.cappio, barbara.caputo, tatiana.tommasi}@polito.it

## Abstract

Vision systems trained in closed-world scenarios fail when presented with new environmental conditions, new data distributions, and novel classes at deployment time. In this work, we tackle multi-source Open-Set domain adaptation by introducing HyMOS: a straightforward model that exploits the power of contrastive learning and the properties of its hyperspherical feature space to correctly predict known labels on the target, while rejecting samples belonging to any unknown class. HyMOS outperforms all competitors in several benchmarks.
Code: https://github.com/silvia1993/HyMOS.

## 1 Introduction

Artificial intelligent systems face many operational challenges when moving from the controlled lab environment to the real-world. First of all the annotated data available to train a model might be the result of asynchronous multi-agent collection processes. For vision tasks, this means dealing with datasets composed of labeled images that share the same class set, but with sub-groups of instances showing significant differences in appearance and style among each other. Moreover, at deployment time the learned model will inevitably encounter new environmental conditions, with distribution shift and novel classes not present during training.

Closed-Set domain adaptation [Csurka, 2017] focuses on reducing the gap between the training (labeled source) and test (unlabeled target) data when the latter covers exactly the same class set of the former. Open-Set domain adaptation [Panareda Busto e Gall, 2017] aims at bridging the domain gap while also rejecting target samples of *unknown* classes. Indeed, in the case of category shift, the application of naïve adaptive solutions may lead to *negative-transfer* and unrecoverable class misalignment [Liu *et al.*, 2019]. Although dealing with multiple sources is more the rule than an exception in real-world conditions, only one recent work has started to peek into the multi-source Open-Set domain adaptation task [Rakshit *et al.*, 2020]. The problems that have to be faced are a consequence of the fact that standard deep learning models are not able to generalize: (1) they yield features that describe mostly local rather than global statistics, which causes a bias on the image style of the training data [Jenni *et al.*, 2020]; (2) the cross-entropy loss, widely used for supervised learning, produces overconfident predictions thus biasing the model towards the labeled class set [Khosla *et al.*, 2020].

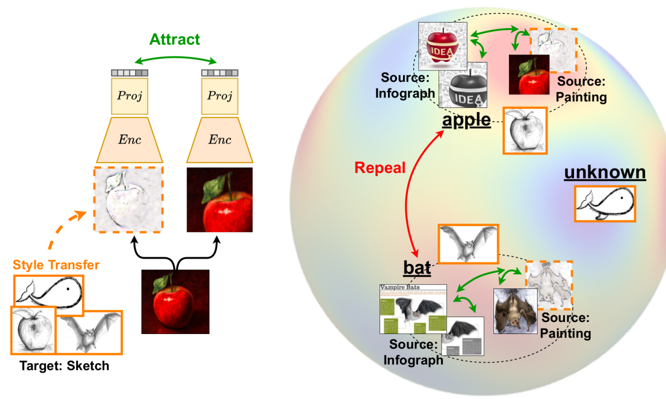


Figure 1: With HyMOS we exploit supervised contrastive learning to tackle all the challenges of multi-source Open-Set domain adaptation.

With our work, we propose a supervised model that avoids the drawbacks of the cross-entropy loss, while learning a style-invariant embedding space that naturally isolates the *unknown* categories. Specifically, we build over the very recent contrastive learning trend [Chen *et al.*, 2020; Khosla *et al.*, 2020]. We show how **a single supervised contrastive learning objective can tackle every challenge of multi-source Open-Set domain adaptation** (see Figure 1):

- source to source class-wise alignment comes by simply balancing data batches over classes and domains;
- source to target adaptation is obtained by getting domain invariant features via the introduction of style transfer among the augmentations of contrastive learning;
- the separation between *known* and *unknown* target data comes from a self-paced threshold based on the observed data distribution on the learned hyperspherical feature embedding.

To highlight the important role of the **Hy**perspherical feature space for our **M**ulti-source **O**pen-**S**et approach, we dub it **HyMOS**.

*The authors equally contributed to this work.

## 2 Related works

**Domain Adaptation** A model trained and tested on data sharing the same label set but drawn from two different marginal distributions will inevitably show low performance. *Closed-Set domain adaptation* addresses this problem by increasing the invariance of the learned features over source and target domains. Various strategies have been proposed to obtain this result: 1) minimization of statistical metrics that reflect the distribution discrepancy [Yan *et al.*, 2017]; 2) adversarial learning [Ganin *et al.*, 2016]; 3) batch and feature normalization [Carlucci *et al.*, 2020]; 4) self-supervision [Bucci *et al.*, 2019]; 5) domain shift reduction at pixel level via generative models [Russo *et al.*, 2018]. When dealing with multiple sources, the extra challenge is in aligning all the domains among each other while producing a highly discriminative feature space. Considering that the target is unlabeled, being sure that its semantic content perfectly matches that of the source is unrealistic. *Open-Set domain adaptation* tackles target domains which include new unknown classes with respect to the source. A number of approaches have been proposed to deal with this scenario: [Panareda Busto e Gall, 2017; Kundu *et al.*, 2020]. A single published method deals with multi-source Open-Set [Rakshit *et al.*, 2020] by adding a clustering objective to maximize the similarity among samples of the same class.

**Contrastive Learning** Lately, self-supervised learning methods have shown that, by relying only on unlabeled data, it is still possible to get classification performance similar to those of the supervised approaches [Chen *et al.*, 2020; He *et al.*, 2020]. This is done by minimizing the distance between positive samples pairs (obtained through augmentations of the same input sample) while maximizing the distance between negative pairs. Existing methods differ in how positive and negative data pairs are sampled and stored: among the most cited, SimCLR [Chen *et al.*, 2020] adopts a large batch size, while MoCo [He *et al.*, 2020] maintains a momentum encoder and a limited queue of previous samples. Direct ways to incorporate supervision in this pipeline are currently attracting large attention [Khosla *et al.*, 2020] and show how view invariance and semantic knowledge can be combined to tackle novelty detection [Tack *et al.*, 2020] or cross-domain generalization [Zhang *et al.*, 2021].

## 3 Method

To tackle multi-source Open Set domain adaptation we focus on building a robust, highly structured feature space with domain-aligned, compact, and well-separated class clusters, keeping *unknown* target samples away from the centers. We obtain this result by minimizing the supervised contrastive loss and paying attention to how data are fed to the model.

**Problem Framework** In multi-source Open-Set domain adaptation we are given $L$ labeled source domains $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_L\}$, where $\mathcal{S}_i = \{\boldsymbol{x}_j^{s_i}, y_j^{s_i}\}_{j=1}^{N^{s_i}} \sim p_i$, and one unlabeled target domain $\mathcal{T} = \{\boldsymbol{x}_j^t\}_{j=1}^{N^t} \sim q$, all drawn from different data distributions $p_{i=1,\ldots,L}, q$. The sources share the same label set $y^s \in \mathcal{C}_s$, and it holds $\mathcal{C}_s \subset \mathcal{C}_t$, thus the target covers $\mathcal{C}_{t\setminus s}$ additional classes which are considered *unknown*. Starting from this setup, the goal is to train a model on the source data, able to identify the label of each target sample, by either assigning it to one of the *known* $\mathcal{C}_s$ classes, or rejecting it as *unknown*. Given the different relatedness levels of the target with each of the available sources, reducing the domain shift while avoiding the risk of *negative-transfer* may be difficult, especially when the *openness* $O = 1 - \frac{|\mathcal{C}_s|}{|\mathcal{C}_t|}$ increases.

**Contrastive Learning Formulation** In self-supervised contrastive learning [Chen *et al.*, 2020], two transformed views of every input image are propagated through a CNN network. For each sample $\{\boldsymbol{x}_k^s, y_k^s\}$ in the double batch $B = \{k = 1, \ldots, 2K\}$ the features obtained via the encoder $Enc(\boldsymbol{x}_k^s)$ enter the final contrastive head that further projects them to a normalized embedding, producing $\boldsymbol{z}_k^s = Proj(Enc(\boldsymbol{x}_k^s))$. On the obtained hyperspherical space the samples are compared among each other: the similarity between two augmented views of the same instance is maximized, while the similarity of two different instances is minimized. When the image labels are available, samples of the same class $y_k^s$ are considered as positives, while the negative pairs are composed by samples of different classes. We indicate with $\nu(k) = B \setminus \{k\}$ the double batch without the *anchor* sample of index $k$, and the positive pairs are $\pi(k) = \{k' \in \nu(k) : y_{k'}^s = y_k^s\}$. Thus, the supervised contrastive loss is [Khosla *et al.*, 2020]:

$$\mathcal{L}_{SupClr} = \sum_{k=1}^{2K} \frac{-1}{|\pi(k)|} \sum_{k' \in \pi(k)} \log \frac{\exp(\sigma(\boldsymbol{z}_k^s, \boldsymbol{z}_{k'}^s)/\tau)}{\sum\limits_{n \in \nu(k)} \exp(\sigma(\boldsymbol{z}_k^s, \boldsymbol{z}_n^s)/\tau)}, \tag{1}$$

where $\tau \in R^+$ is the temperature, and $\sigma(\cdot, \cdot)$ is the cosine similarity.

**(a) Source-Source Class-Wise Domain Alignment** The ability of the supervised contrastive loss to learn compact class clusters can be exploited to perform source-source class-wise domain alignment by composing each training mini-batch with samples coming from different visual domains.

**(b) Source-Target Style Invariance** The image transformations used in contrastive learning are meant to let the model focus on core semantic information while making it invariant to the irrelevant cues that they introduce. When dealing with data from different domains we desire a representation able to neglect major differences in visual appearance. We propose an augmentation based on style transfer as it is perfectly suitable for this goal: it does not affect the image content while changing significantly the global image texture. In particular, we use the AdaIN [Huang e Belongie, 2017] model trained jointly on source and target data in order to transfer the style from target images into source ones. As this augmentation is applied randomly, the loss function will explicitly compare original source images with target-like ones and learn to ignore the style difference. We highlight that our approach to obtain style invariance is safe from *negative-transfer*.

**(c) Refinement via Self-Training** After an initial source-only learning episode, we start progressively integrating the target samples in our learning objective, by passing through evaluation steps that we call *self-training break-points*, which allow us to select confident *known* target samples. Through this iterative technique, we propagate label knowledge from source

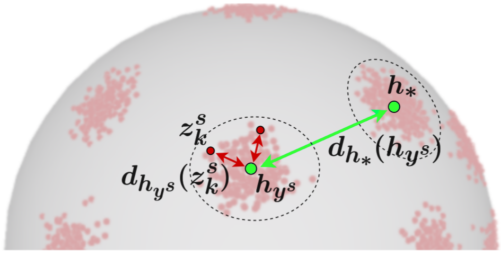


Figure 2: Illustration of the distances used to set the class prediction and the self-training procedure.

to target data, improving the compactness of our class clusters while progressively leaving *unknown* target data in low-density regions of the hyperspherical feature space.

**(d) Known-Unknown Separation and Classification on the Hypersphere** Differently from previous literature where the contrastive models were used only as pretext tasks and the projection head was later dropped in favor of a standard cross-entropy loss, we propose to stay on the hypersphere while delivering the final predictions. We define the prototype of each source class $y^s$ by computing the corresponding feature average $h_{y^s} = \frac{1}{N_{y^s}} \sum_{k \in y^s} \boldsymbol{z}_k^s$ , re-projected on the unit hypersphere. For any target sample $\boldsymbol{z}^t$ we measure the cosine similarity to each source class prototype and we rescale it in $[0, 1]$ to define the distance $d_{h_{y^s}}(\boldsymbol{z}^t) = \{1 - \sigma_{[0,1]}(\boldsymbol{z}^t, h_{y^s})\}$ for $y^s \in \{1, \dots, |\mathcal{C}_s|\}$ , which is used as confidence measure for label assignment.

To decide whether a sample belongs to a *known* category we need a threshold on the distance from the *known* class prototypes. We propose to directly extract it from the observed data distribution, obtaining a value that changes online during the learning process. Specifically we introduce two metrics to evaluate it: the *class sparsity*:

$$\theta = \frac{1}{|C_s|} \sum_{y^s \in C_s} d_{h_*}(h_{y^s}) \,, \tag{2}$$

where $h_*$ is the closest prototype to each $h_{y^s}$, and the *class compactness*:

$$\phi = \frac{1}{|C_s|} \sum_{y^s \in C_s} \left\{ \frac{1}{N_{y^s}} \sum_{k \in y^s} d_{h_{y^s}}(z_k^s) \right\} . \tag{3}$$

See Figure 2 for the visualization of these two metrics on the hypersphere. We compute our threshold by:

$$\alpha = \phi \cdot \left[ log \left( \frac{\theta}{2\phi} \right) + 1 \right] \,, \tag{4}$$

where $\theta/2\phi$ estimates the average ratio between the distance of two adjacent prototypes and the radii of the respective clusters. The use of the threshold at inference time is straightforward:

$$\hat{y}^t = \begin{cases} \arg\min_{y^s}(d_{h_{y^s}}(\boldsymbol{z}^t)) & \text{if } \min_{y^s}(d_{h_{y^s}}(\boldsymbol{z}^t)) < \alpha \\ \texttt{unknown} & \text{if } \min_{y^s}(d_{h_{y^s}}(\boldsymbol{z}^t)) \geq \alpha \,. \end{cases} \tag{5}$$

We exploit this threshold also for the self-training breakpoints described before.

## 4 Experiments

We implemented HyMOS with a ResNet-50 backbone and two fully connected layers which define the contrastive head. All the implementation details as well as the Pytorch code are provided in the GitHub repository.

**Datasets** We evaluate our approach on three image classification benchmarks, following the same setting used in [Rakshit *et al.*, 2020], with one domain considered in turn as target. **Office31** [Saenko *et al.*, 2010] comprises three domains, each containing 31 object categories. We set as known the first 20 classes in alphabetic order, while the remaining 11 are unknown. **Office-Home** [Venkateswara *et al.*, 2017] is made by four domains with 65 classes. The first 45 categories in alphabetic order are known, and the remaining 20 are unknown. **DomainNet** [Peng *et al.*, 2019] contains six domains and 345 classes. As in [Rakshit *et al.*, 2020], we consider Infograph (I), Painting (P), Sketch (S) and Clipart (C), selecting randomly 50 samples per class or using all the images in case of lower cardinality. The first 100 classes in alphabetic order are known, while the remaining 245 are unknown.

**Results** We compare HyMOS with several state-of-the-art baselines proposed for single-source Open-Set (Inheritable [Kundu *et al.*, 2020], ROS [Bucci *et al.*, 2020], PGL [Luo *et al.*, 2020]), multi-source Open-Set (MOSDANET [Rakshit *et al.*, 2020]) and universal domain adaptation (CMU [Fu *et al.*, 2020], DANCE [Saito *et al.*, 2020]). We use the code provided by the authors and for all the methods that do not specify how to manage multiple sources, we apply the *Source Combine* strategy that considers the union of all the source data in a single domain. We adopt the $HOS$ metric, defined in [Bucci *et al.*, 2020; Fu *et al.*, 2020] for a fair evaluation of Open-Set methods: it is the harmonic mean between the average class accuracy over the known classes $OS^*$ and the accuracy over the unknown class *UNK*: $HOS = 2\frac{\text{OS}^* \times \text{UNK}}{\text{OS}^* + \text{UNK}}$.

Table 1 collects the obtained results showing how HyMOS outperforms all the baselines. The gain of HyMOS with respect to the best competitor ROS, varies from 1.9% points on OfficeHome, up to 10.8% on DomainNet.

We also benchmark HyMOS with the best competitor ROS in terms of the $AUROC$ (Area under the Receiver Operating Characteristic curve) metric, which has the advantage of being threshold-independent. In our case, the *normality score* used to evaluate whether a sample is *known* or *unknown* is its distance from the nearest source class prototype, while ROS exploits a combination of entropy and probability output of an auxiliary rotation recognition classifier. Even in this case, HyMOS outperforms ROS.

## 5 Conclusions

In this paper we introduced HyMOS, a straightforward approach for multi-source Open-Set domain adaptation. It exploits contrastive learning and the inherent properties of its hyperspherical feature space to avoid the limitations of the existing competing methods. We demonstrated state-of-the-art results on three benchmarks and we delved into the details of the method internal functioning.

| | | Method | DomainNet | | | Office31 | | | | Office-Home | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | → S | → C | **Avg.** | → W | → D | → A | **Avg.** | → Rw | → Cl | → Ar | → Pr | **Avg.** |
| HOS | Source Combine | Inheritable [Kundu *et al.*, 2020] | 34.8 | 44.0 | 39.4 | 76.6 | 79.5 | 70.0 | 75.4 | 63.2 | 52.6 | 48.7 | 60.7 | 56.3 |
| | | ROS [Bucci *et al.*, 2020] | 44.5 | 52.4 | 48.5 | 81.8 | 80.1 | 64.7 | 75.5 | **73.0** | 57.3 | 61.6 | 69.1 | 65.3 |
| | | CMU [Fu *et al.*, 2020] | 38.1 | 35.5 | 36.8 | 61.4 | 64.0 | 56.4 | 60.6 | 70.8 | 50.0 | 58.1 | 69.3 | 62.1 |
| | | DANCE [Saito *et al.*, 2020] | 30.0 | 37.6 | 33.8 | 38.5 | 59.7 | 58.0 | 52.0 | 12.4 | 16.1 | 18.6 | 22.9 | 17.5 |
| | | PGL [Luo *et al.*, 2020] | 18.5 | 19.4 | 19.0 | 43.3 | 37.7 | 35.6 | 38.9 | 40.0 | 31.5 | 31.8 | 42.2 | 36.4 |
| | Multi-Source | MOSDANET [Rakshit *et al.*, 2020] | 40.0 | 39.3 | 39.6 | 60.5 | 71.5 | **73.9** | 68.6 | 65.0 | 51.1 | 54.3 | 65.9 | 59.1 |
| | | **HyMOS** | **57.5** | **61.0** | **59.3** | **90.2** | **89.9** | 60.8 | **80.3** | 71.0 | **64.6** | **62.2** | **71.1** | **67.2** |
| AUROC | Source Combine | ROS [Bucci *et al.*, 2020] | 63.9 | 68.0 | 66.0 | 93.9 | 95.2 | **73.5** | 87.5 | 80.8 | 69.6 | 73.7 | 79.4 | 75.9 |
| | Multi-Source | **HyMOS** | **71.9** | **75.8** | **73.9** | **96.9** | **96.1** | 71.0 | **88.0** | **81.1** | **76.4** | **75.3** | **79.6** | **78.1** |

Table 1: Results averaged over three runs for each method on the DomainNet, Office31, and Office-Home datasets.

## References


[Bucci *et al.*, 2019] Silvia Bucci, Antonio D'Innocente, e Tatiana Tommasi. Tackling Partial Domain Adaptation with Self-supervision. In *ICIAP*, 2019.

[Bucci *et al.*, 2020] Silvia Bucci, Mohammad Reza Loghmani, e Tatiana Tommasi. On the effectiveness of image rotation for open set domain adaptation. In *ECCV*, 2020.

[Carlucci *et al.*, 2020] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, e S. Rota Bulo. Multidial: Domain alignment layers for (multisource) unsupervised domain adaptation. *IEEE TPAMI*, 2020.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, e Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.

[Csurka, 2017] Gabriela Csurka. *Domain Adaptation in Computer Vision Applications*. Springer, 2017.

[Fu *et al.*, 2020] Bo Fu, Zhangjie Cao, Mingsheng Long, e Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *ECCV*, 2020.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, e Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[He *et al.*, 2020] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, e Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020.

[Huang e Belongie, 2017] Xun Huang e Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017.

[Jenni *et al.*, 2020] Simon Jenni, Hailin Jin, e Paolo Favaro. Steering self-supervised feature learning beyond local pixel statistics. In *CVPR*, 2020.

[Khosla *et al.*, 2020] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, e Dilip Krishnan. Supervised contrastive learning. In *NeurIPS*, 2020.

[Kundu *et al.*, 2020] Jogendra Nath Kundu, Naveen Venkat, Ambareesh Revanur, R Venkatesh Babu, et al. Towards inheritable models for open-set domain adaptation. In *CVPR*, 2020.

[Liu *et al.*, 2019] Hong Liu, Zhangjie Cao, Mingsheng Long, Jianmin Wang, e Qiang Yang. Separate to adapt: Open set domain adaptation via progressive separation. In *CVPR*, 2019.

[Luo *et al.*, 2020] Yadan Luo, Zijian Wang, Zi Huang, e Mahsa Baktashmotlagh. Progressive graph learning for open-set domain adaptation. In *ICML*, 2020.

[Panareda Busto e Gall, 2017] Pau Panareda Busto e Juergen Gall. Open set domain adaptation. In *ICCV*, 2017.

[Peng *et al.*, 2019] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, e Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2019.

[Rakshit *et al.*, 2020] Sayan Rakshit, Dipesh Tamboli, Pragati Shuddhodhan Meshram, Biplab Banerjee, Gemma Roig, e Subhasis Chaudhuri. Multi-source open-set deep adversarial domain adaptation. In *ECCV*, 2020.

[Russo *et al.*, 2018] Paolo Russo, Fabio Maria Carlucci, Tatiana Tommasi, e Barbara Caputo. From source to target and back: symmetric bi-directional adaptive gan. In *CVPR*, 2018.

[Saenko *et al.*, 2010] Kate Saenko, Brian Kulis, Mario Fritz, e Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, 2010.

[Saito *et al.*, 2020] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, e Kate Saenko. Universal domain adaptation through self-supervision. In *NeurIPS*, 2020.

[Tack *et al.*, 2020] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, e Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In *NeurIPS*, 2020.

[Venkateswara *et al.*, 2017] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, e Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017.

[Yan *et al.*, 2017] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, e Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *CVPR*, 2017.

[Zhang *et al.*, 2021] Yifan Zhang, Bryan Hooi, Dapeng Hu, Jian Liang, e Jiashi Feng. Unleashing the power of contrastive self-supervised visual models via contrast-regularized fine-tuning. *arXiv preprint arXiv:2102.06605*, 2021.