Learning to Learn in a Semi-Supervised Fashion

Yun-Chun Chen[1], Chao-Te Chou[1], and Yu-Chiang Frank Wang[1,2]

[1] Graduate Institute of Communication Engineering, National Taiwan University, Taiwan

[2] ASUS Intelligent Cloud Services, Taiwan

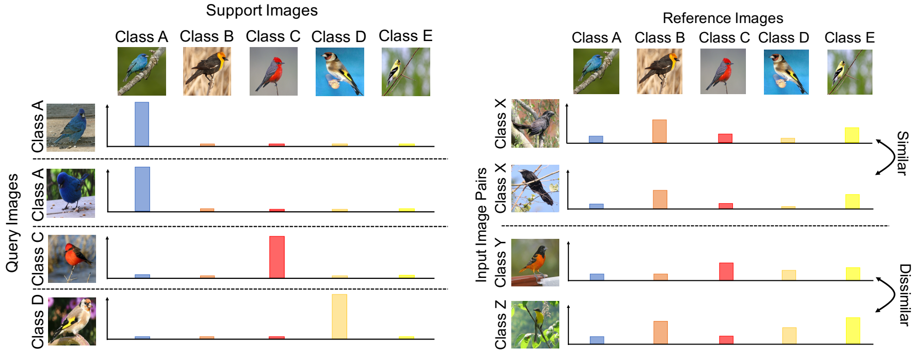
{b03901148, b03901096, ycwang}@ntu.edu.tw

Abstract. To address semi-supervised learning from both labeled and unlabeled data, we present a novel meta-learning scheme. We particularly consider that labeled and unlabeled data share disjoint ground truth label sets, which can be seen tasks like in person re-identification or image retrieval. Our learning scheme exploits the idea of leveraging information from labeled to unlabeled data. Instead of fitting the associated class-wise similarity scores as most meta-learning algorithms do, we propose to derive semantics-oriented similarity representations from labeled data, and transfer such representation to unlabeled ones. Thus, our strategy can be viewed as a self-supervised learning scheme, which can be applied to fully supervised learning tasks for improved performance. Our experiments on various tasks and settings confirm the effectiveness of our proposed approach and its superiority over the state-of-the-art methods.

1 Introduction

Recent advances of deep learning models like convolutional neural networks (CNNs) have shown encouraging performance in various computer vision applications, including image retrieval [70,75,89] and person re-identification (re-ID) [9,21,41,42,96]. Different from recognizing the input as a particular category, the above tasks aim at learning feature embeddings, making instances of the same type (e.g., object category) close to each other while separating those of distinct classes away. Similar tasks such as image-based item verification [46], face verification [72], face recognition [17,62,68], and vehicle re-ID [16,73,83] can all be viewed as the tasks of this category.

Existing methods for image matching generally require the collection of a large number of labeled data, and tailor algorithms to address the associated tasks (e.g., image retrieval [70,75] and person re-ID [21,41,96,97]). However, the assumption of having a sufficient amount of labeled data available during training may not be practical. To relax the dependency of manual supervision, several *semi-supervised* methods for image retrieval [78,82,92] and person re-ID [39,85] are proposed. These methods focus on learning models from datasets where each category is partially labeled (i.e., some data in *each* category are labeled, while the rest in that category remain unlabeled). Thus, they choose to use the models learned from the labeled data to assign pseudo labels to the unlabeled ones [82,85,92], or



(a) Existing meta-learning methods.

(b) Our meta-learning algorithm.

Fig. 1: **Illustration of learning class-wise similarity.** (a) Standard meta-learning methods for visual classification compute class-wise similarity *scores* between the query image and those in the support set, where the two sets share *overlapping* ground truth labels. (b) To deal with training data with non-overlapping labels, our meta-learning scheme derives semantics-oriented similarity representations in a learning-to-learn fashion, allowing the determination of pairwise relationship between images with unseen labels.

adopt ensemble learning techniques to enforce the predictions of the unlabeled data to be consistent across multiple networks [78]. Despite significant progress having been reported, these methods *cannot* be directly applied to scenarios where novel objects or persons are present.

To deal with instances of unseen categories for image matching purposes, one can approach such problem in two different ways. The majority of existing methods focuses on the cross-dataset (domain adaptation [10,31]) setting, where one dataset is fully labeled (i.e., source domain dataset) while the other one remains unlabeled. (i.e., target domain dataset) [18,81,91]. Existing methods for this category typically assume that there is a domain gap between the two datasets. These methods either leverage adversarial learning strategies to align feature distributions between the two datasets [18,81], or aim at assigning pseudo labels for each unlabeled image in the target dataset through predicting class-wise similarity scores from models trained on the source (labeled) dataset [91]. By carefully selecting hyperparameters such as the prediction score threshold, one can determine whether or not a given image pair from the target dataset is of the same category. However, the class-wise similarity scores are computed based on a network trained on the source dataset, which might not generalize well to the target dataset, especially when their labels are non-overlapping. On the other hand, these methods are developed based on the assumption that a large-scale labeled dataset is available.

Another line of research considers learning models from a *single* dataset, in which only some categories are fully labeled while the remaining classes are unlabeled [87,88]. These methods typically require the number of classes of the unlabeled data to be known in advance, so that one can perform clustering-like algorithms with the exact number of clusters for pseudo label assignment. Having such prior knowledge, however, might not be practical for real-world applications.

In this paper, we propose a novel meta-learning algorithm for image matching in a semi-supervised setting, with applications to image retrieval and person re-ID. Specifically, we consider the same semi-supervised setting as [87,88], in which the ground truth label sets of labeled and unlabeled training data are *disjoint*. Our meta-learning strategy aims at exploiting and leveraging class-wise similarity representation across labeled and unlabeled training data, while such similarity representation is derived by a learning-to-learn fashion. The resulting representations allow our model to relate images with pseudo labels in the unlabeled set (e.g., Figure 1b). This is very different from existing meta-learning for visual classification methods (like few-shot learning), which typically assume that the support and query sets share the same label set and focus on fitting the associated class-wise similarity scores (e.g., Figure 1a). Our learning scheme is realized by learning to match randomly selected labeled data pairs, and such concepts can be applied to observe both labeled and unlabeled data for completing the semi-supervised learning process.

The contributions of this paper are summarized as follows:

- We propose a meta-learning algorithm for image matching in semi-supervised settings, where labeled and unlabeled data share non-overlapping categories.
- Our learning scheme aims at deriving semantics-oriented similarity representation across labeled and unlabeled sets. Since pseudo labels can be automatically assigned to the unlabeled training data, our approach can be viewed as a self-supervised learning strategy.
- With the derivation of semantics-oriented similarity representations, our learning scheme can be applied to fully supervised settings and further improves the performance.
- Evaluations on four datasets in different settings confirm that our method performs favorably against existing image retrieval and person re-ID approaches.

2 Related Work

Semi-supervised learning. Semi-supervised learning for visual analysis has been extensively studied in the literature. Most of the existing methods focus on image classification and can be categorized into two groups depending on the learning strategy: 1) labeling-based methods and 2) consistency-based approaches. Labeling-based methods focus on assigning labels to the unlabeled images through pseudo labeling [38], label propagation [49], or leveraging regularization techniques for performing the above label assignment [25]. Consistency-based approaches, on the other hand, exploit the idea of cycle consistency [8,11,101] and adopt ensemble learning algorithms to enforce the predictions of the unlabeled samples to be consistent across multiple models [3,5,37,51,56,59,74]. In addition to image classification, another line of research focuses on utilizing annotation-free images to improve the performance of semantic segmentation [33,67]. These methods adopt generative adversarial networks (GANs) [24] to generate images conditioned on the class labels to enhance the learning of feature representations for the unlabeled images [67], or develop a fully convolutional discriminator to generate dense probability maps that indicate the confidence of correct segmentation for each pixel in the unlabeled images [33].

To match images of the same category, a number of methods for image retrieval [78,82,92] and person re-ID [19,22,32,39,44,45,85] also consider learning models in semi-supervised settings. Methods for semi-supervised image retrieval can be grouped into two categories depending on the adopted descriptors: 1) hand-crafted descriptor based methods and 2) approaches based on trainable descriptors. The former typically focuses on optimizing the errors on the labeled set and leverages a regularizer to maximize the information entropy between labeled and unlabeled sets [82]. Trainable descriptor based approaches either utilize a graph to model the relationship between labeled and unlabeled sets [92] or leverage a GAN [24] to learn triplet-wise information from both labeled and unlabeled data [78]. Similarly, methods for semi-supervised person re-ID also aim at relating labeled and unlabeled images through dictionary learning [44], multi-feature learning [22], pseudo labeling with regularizers [32], or considering complex relationships between labeled and unlabeled images [19]. While promising performance has been shown, these methods cannot be directly applied to scenarios where datasets contain labeled and unlabeled images with non-overlapping category labels, which are practical in many real-world applications.

To tackle this issue, two recent methods for semi-supervised person re-ID are proposed [87,88]. These methods either combine K-means clustering and multi-view clustering [87], or develop a self-paced multi-view clustering algorithm [88] to assign pseudo labels to images in the unlabeled set. However, these methods require the number of identities of the unlabeled set to be known in advance. Our work does not need such prior knowledge. As noted above, we approach such problems and assign pseudo labels to the unlabeled data by learning their semantics-oriented similarity representations, which are realized in a unique learning-to-learn fashion.

Meta-learning. The primary objective of meta-learning is to enable a base learning algorithm which observes data with particular properties to adapt to similar tasks with new concepts of interest. Few-shot learning [65,69] and neural architecture search [7] are among the popular applications of meta-learning. Existing meta-learning algorithms can be grouped into three categories based on the learning task: 1) initialization-based methods, 2) memory-based approaches, and 3) metric-based algorithms. Initialization-based methods focus on learning an optimizer [2,12,30,57] or learning to initialize the network parameters so that the models can rapidly adapt to novel classes or new tasks [23]. Memory-based approaches leverage memory-augmented models (e.g., the hidden activations in a recurrent network or external memory) to retain the learned knowledge [35,53,60], and associate the learned knowledge with the newly encountered tasks for rapid generalization. Metric-based algorithms aim at learning a feature embedding with proper distance metrics for few-shot [65,69] or one-shot [76] classification.

Similar to metric-based meta-learning algorithms, our method also aims at learning a feature embedding. Our method differs from existing meta-learning for visual classification approaches in that we learn a feature embedding from both labeled and unlabeled data with disjoint label sets. Moreover, both the support and query sets share the same label set in most other meta-learning approaches, while the label sets of our meta-training and meta-validation sets are disjoint.

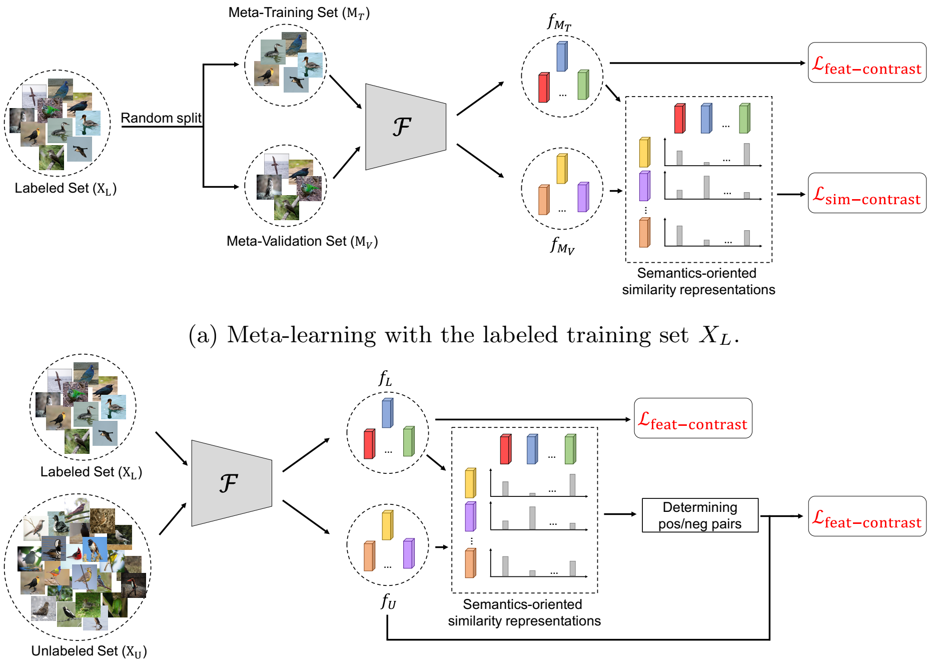
3 Proposed Method

3.1 Algorithmic Overview

We first describe the setting of our semi-supervised learning task, and define the notations. When matching image pairs in the tasks of image retrieval and person re-ID, we assume that our training set contains a set of N_L labeled images $X_L = \{x_i^L\}_{i=1}^{N_L}$ with the corresponding labels $Y_L = \{y_i^L\}_{i=1}^{N_L}$, and a set of N_U unlabeled images $X_U = \{x_j^U\}_{j=1}^{N_U}$. For the labeled data, each $x_i^L \in \mathbb{R}^{H \times W \times 3}$ and $y_i^L \in \mathbb{R}$ denote the i^{th} image and the associated label, respectively. As for $x_j^U \in \mathbb{R}^{H \times W \times 3}$, it is the j^{th} unlabeled image in X_U. Note that the class number of the labeled set is denoted as C_L, while that of the unlabeled set is *unknown*. We assume the label sets of the labeled and unlabeled sets are *disjoint*.

The goal of this work is to learn a feature embedding model by jointly observing the above labeled and unlabeled sets, with the learned features can be applied for matching images for tasks of retrieval and re-ID. As shown in Figure 2, our proposed algorithm comprises two learning phases: 1) meta-learning with labeled training data and 2) meta semi-supervised learning on both labeled and unlabeled training sets. For the first phase (i.e., Figure 2a), we first partition the labeled set X_L into a meta-training set $M_T = \{x_k^{M_T}\}_{k=1}^{N_{M_T}}$ and a meta-validation set $M_V = \{x_l^{M_V}\}_{l=1}^{N_{M_V}}$, with disjoint labels for M_T and M_V (from Y_L). The numbers of images for M_T and M_V are denoted as N_{M_T} and N_{M_V}, and the numbers of classes for M_T and M_V are denoted as C_{M_T} and C_{M_V}, respectively, summing up as C_L (i.e., $C_L = C_{M_T} + C_{M_V}$). Our model F takes images x from M_T and M_V as inputs, and learns feature representations $f = F(x) \in \mathbb{R}^d$ (d is the dimension of f) for input images. Our model then derives semantics-oriented similarity representation $s \in \mathbb{R}^{C_{M_T}}$ (C_{M_T} denotes the dimension of s) for each image in M_V. In the second meta-learning stage (i.e., Figure 2b), we utilize the learned concept of semantics-oriented similarity representation to guide the learning of the unlabeled set. The details of each learning phase are elaborated in the following subsections.

As for testing, our model takes a query image as input and extracts its feature $f \in \mathbb{R}^d$, which is applied to match gallery images via nearest neighbor search.


Meta-Training Set (M_T)
Random split
Labeled Set (X_L)
Meta-Validation Set (M_V)
$\mathcal{F}$
f_{M_T}
f_{M_V}
$\mathcal{L}_{\text{feat-contrast}}$
$\mathcal{L}_{\text{sim-contrast}}$
Semantics-oriented similarity representations
Labeled Set (X_L)
Unlabeled Set (X_U)
$\mathcal{F}$
f_L
f_U
$\mathcal{L}_{\text{feat-contrast}}$
Semantics-oriented similarity representations
Determining pos/neg pairs
$\mathcal{L}_{\text{feat-contrast}}$

(a) Meta-learning with the labeled training set X_L.

(b) Meta semi-supervised learning on both labeled set X_L and unlabeled set X_U.

Fig. 2: **Overview of the proposed meta-learning framework.** (a) Our model first takes labeled data and learns semantics-oriented similarity representation in a learning-to-learn fashion (i.e., joint learning of $\mathcal{L}_{\text{feat}-\text{contrast}}$ and $\mathcal{L}_{\text{sim}-\text{contrast}}$). (b) The learned concept of matching semantics information allows us to learn from both labeled and unlabeled data, determining positive/negative pairs for computing $\mathcal{L}_{\text{feat}-\text{contrast}}$ for all the training data pairs. Note that the meta-training and meta-validation sets in (a) do *not* share the same labels, *neither* do the labeled and unlabeled training sets in (b).

3.2 Meta-Learning on X_L

Motivated by [91], we exploit the idea of leveraging information from class-wise similarity to guide the learning of the unlabeled data. In our work, we choose to implicitly learn semantics-oriented similarity representation instead of explicit label-specific representation in a learning-to-learn fashion, so that the learned representation can be applied for describing the unlabeled images.

To achieve this, we advance an episodic learning paradigm as applied in existing meta-learning algorithms [4, 23]. In each episode, we first divide the labeled set L into a meta-training set M_T and a meta-validation set M_V, where the labels of M_T and M_V are *not* overlapped. To learn a feature embedding for matching images, we follow existing methods [14, 26] and introduce a feature contrastive loss $\mathcal{L}_{\text{feat}-\text{contrast}}$ in the meta-training set M_T. That is, given a pair of images $x_i^{M_T}$ and $x_j^{M_T}$ in M_T, the feature contrastive loss for M_T is defined as

$$\mathcal{L}_{\text{feat}-\text{contrast}}(M_T; F) = t \cdot \|f_i^{M_T} - f_j^{M_T}\| + (1-t) \cdot \max(0, \phi - \|f_i^{M_T} - f_j^{M_T}\|), \quad (1)$$

where $t = 1$ if $x_i^{M_T}$ and $x_j^{M_T}$ are of the same label, otherwise $t = 0$, and $\phi > 0$ denotes the margin.

Semantics-oriented similarity representation s. To learn semantics-oriented similarity representation, we first sample a reference image $\hat{x}^{M_T}$ from each class in the meta-training set M_T. The sampled reference image for the k^{th} class in M_T is denoted as $\hat{x}_k^{M_T}$, and there are C_{M_T} sampled reference images in total. We then extract feature $\hat{f}^{M_T} = F(\hat{x}^{M_T})$ for each reference image $\hat{x}^{M_T}$.

Given an image $x_i^{M_V}$ in the meta-validation set M_V, we first extract its feature $f_i^{M_V} = F(x_i^{M_V})$. To learn semantics-oriented similarity representation $s_i^{M_V} \in \mathbb{R}^{C_{M_T}}$ for image $x_i^{M_V}$, we compute the class-wise similarity scores between $f_i^{M_V}$ and all reference features $\hat{f}^{M_T}$ sampled from the meta-training set. The k^{th} entry of the semantics-oriented similarity representation $s_i^{M_V}$ is defined as

$$s_i^{M_V}(k) = \text{sim}(f_i^{M_V}, \hat{f}_k^{M_T}), \quad (2)$$

where $\text{sim}(f_i^{M_V}, \hat{f}_k^{M_T})$ denotes the similarity between feature $f_i^{M_V}$ and the sampled reference feature $\hat{f}_k^{M_T}$. We note that we do not limit the similarity measurement in the above equation. For example, we compute the cosine similarity for image retrieval and calculate the ℓ_2 distance for person re-ID.

To achieve the learning of semantics-oriented similarity representation, we utilize the ground truth label information from the meta-validation set, and develop a similarity contrastive loss $\mathcal{L}_{\text{sim}-\text{contrast}}$. Specifically, given an image pair $x_i^{M_V}$ and $x_j^{M_V}$ in M_V, the associated similarity contrastive loss is defined as

$$\mathcal{L}_{\text{sim}-\text{contrast}}(M_V; F) = t \cdot \|s_i^{M_V} - s_j^{M_V}\| + (1-t) \cdot \max(0, \phi - \|s_i^{M_V} - s_j^{M_V}\|), \quad (3)$$

where $t = 1$ if $x_i^{M_V}$ and $x_j^{M_V}$ are of the same category, otherwise $t = 0$. $\phi > 0$ denotes the margin.

By repeating the above procedure across multiple episodes until the convergence of the meta-validation loss (i.e., the similarity contrastive loss), our model carries out the learning of semantics-oriented similarity representation in a learning-to-learn fashion, without fitting particular class label information. Utilizing such representation allows our model to realize joint learning of labeled and unlabeled data, as discussed next.

3.3 Meta Semi-Supervised Learning on X_L and X_U

In the semi-supervised setting, learning from labeled data X_L simply follows the standard feature contrastive loss $\mathcal{L}_{\text{feat}-\text{contrast}}(X_L; F)$. To jointly exploit labeled and unlabeled data, we advance the aforementioned meta-training strategy and start from randomly sampling C_{M_T} categories from the labeled set X_L. For each sampled class, we then randomly sample one reference image $\hat{x}_k^L$ and extract its feature $\hat{f}_k^L = F(\hat{x}_k^L)$, where $\hat{x}_k^L$ denotes the sampled reference image of the k^{th} sampled class. Namely, there are C_{M_T} sampled reference images in total.

Next, given an image x_i^U in the unlabeled set U, we extract its feature $f_i^U = F(x_i^U)$, followed by computing the semantics-oriented similarity representation $s_i^U \in \mathbb{R}^{C_{M_T}}$ between f_i^U and all features $\hat{f}^L$ of the above sampled reference images from the labeled set. It is worth repeating that our learning scheme is very different from existing methods [91], which focus on fitting class-wise similarity scores on the entire labeled set. Instead, we only compute the similarity scores between features of sampled classes. This is the reason why we view our representation to be semantics-oriented instead of class-specific (as [91] does).

Now, we are able to measure the similarity between semantics-oriented similarity representations s_i^U and s_j^U, with a threshold ψ to determine whether the corresponding input images x_i^U and x_j^U are of the same class ($t = 1$) or not ($t = 0$). Namely,

$$\begin{cases} t = 1, & \text{if } \|s_i^U - s_j^U\| < \psi, \\ t = 0, & \text{otherwise.} \end{cases} \tag{4}$$

The above process can be viewed as assigning pseudo positive/negative labels for the unlabeled data X_U, allowing us to compute the feature contrastive loss $\mathcal{L}_{\text{feat}-\text{contrast}}(X_U; F)$ on any image pair from the unlabeled set.

4 Experiments

We present quantitative and qualitative results in this section. In all of our experiments, we implement our model using PyTorch and train our model on a single NVIDIA TITAN RTX GPU with 24 GB memory. The performance of our method can be possibly further improved by applying pre/post-processing methods, attention mechanisms, or re-ranking techniques. However, such techniques are not used in all of our experiments.

4.1 Datasets and Evaluation Metrics

We conduct experiments on four public benchmarks, including the CUB-200 [77], Car196 [36], Market-1501 [96], and DukeMTMC-reID [58] datasets.

Datasets. For image retrieval, we adopt the CUB-200 [77] and Car196 [36] datasets. The CUB-200 dataset [77] is a fine-grained bird dataset containing $11,788$ images of 200 bird species. Following existing methods [27, 54, 55], we use the first 100 categories with $5,864$ images for training and the remaining 100 categories with $5,924$ images for testing. The Car196 [36] dataset is a fine-grained car dataset consisting of $16,189$ images with 196 car categories. Following [27, 54, 55], we use the first 98 categories with $8,054$ images for training while the remaining 98 categories with $8,131$ images are used for testing.

As for person re-ID, we consider the Market-1501 [96] dataset, which contains $32,668$ labeled images of $1,501$ identities captured by 6 camera views. This dataset is partitioned into a training set of $12,936$ images from 751 identities, and a test set of $19,732$ images from the other 750 identities. We also have the DukeMTMC-reID [58] dataset which is composed of $36,411$ labeled images of $1,404$ identities collected from 8 camera views. We utilize the benchmarking training/test split, where the training set consists of $16,522$ images of 702 identities, and the test set contains $19,889$ images of the other 702 identities.

Evaluation metrics. Following recent image retrieval methods [27,55], we use the Recall@K (R@K) metric and the normalized mutual information (NMI) [15] with cosine similarity for evaluating image retrieval performance. For person re-ID, we adopt the standard single-shot person re-ID setting [43] and use the average cumulative match characteristic (CMC) and the mean Average Precision (mAP) with Euclidean distance as similarity measurements.

4.2 Evaluation of Semi-Supervised Learning Tasks

4.2.1 Image Retrieval

Implementation details and settings. Following [52,55], we adopt an ImageNet-pretrained Inception-v1 [71] to serve as the backbone of our model F. A fully connected layer with ℓ_2 normalization is added after the `pool5` layer to serve as the feature embedding layer. All images are resized to $256 \times 256 \times 3$ in advance. During the first stage of meta-learning, we set the batch sizes of the meta-training and meta-validation sets to 32 and 64, respectively. We use the Adam optimizer to train our model for 600 epochs. The initial learning rate is set to 2×10^{-5} and the momentum is set to 0.9. The learning rate is decreased by a factor of 10 for every 150 epochs. The margin ϕ is set to 0.3. As for the meta semi-supervised learning stage, we set the batch size of the labeled set to 32, the batch size of the unlabeled set to 64, and the initial learning rate to 1×10^{-5}. Similarly, the learning rate is decayed by a factor of 10 for every 150 epochs. We train our model for another 600 epochs. The similarity threshold ψ is set to 0.01. We evaluate our method with three different label ratios, i.e., 25%, 50%, and 75% of the categories are labeled, while the remaining categories are unlabeled.

Results. We compare our method with existing fully supervised and unsupervised methods. Table 1 reports the results recorded at Recall@1, 2, 4, and 8, and NMI on the CUB-200 [77] and Car196 [36] datasets. We note that while the results of our method (semi-supervised setting) are not directly comparable to those of fully supervised and unsupervised approaches, their results can be viewed as upper (for fully supervised methods) and lower (for unsupervised approaches) bounds of our results.

The results on both datasets show that our method performs favorably against all competing unsupervised approaches and achieves competitive or even better performance when compared with fully supervised methods.

Table 1: **Results of semi-supervised image retrieval.** The bold and underlined numbers indicate top two results, respectively.

Method	Supervision	CUB-200 [77]					Car196 [36]				
		R@1	R@2	R@4	R@8	NMI	R@1	R@2	R@4	R@8	NMI
Triplet [84]	Supervised	35.9	47.7	59.1	70.0	49.8	45.1	57.4	69.7	79.2	52.9
Lifted [55]		46.9	59.8	71.2	81.5	56.4	59.9	70.4	79.6	87.0	57.8
Clustering [54]		48.2	61.4	71.8	81.9	59.2	58.1	70.6	80.3	87.8	-
Smart+ [27]		49.8	62.3	74.1	83.3	59.9	64.7	76.2	84.2	90.2	-
Angular [80]		**53.6**	**65.0**	**75.3**	**83.7**	**61.0**	**71.3**	**80.7**	**87.0**	**91.8**	**62.4**
Exemplar [20]	Unsupervised	38.2	50.3	62.8	75.0	45.0	36.5	48.1	59.2	71.0	35.4
NCE [86]		39.2	51.4	63.7	75.8	45.1	37.5	48.7	59.8	71.5	35.6
DeepCluster [6]		42.9	54.1	65.6	76.2	53.0	32.6	43.8	57.0	69.5	38.5
MOM [34]		45.3	57.8	68.6	78.4	55.0	35.5	48.2	60.6	72.4	**38.6**
Instance [90]		**46.2**	**59.0**	**70.1**	**80.2**	**55.4**	**41.3**	**52.3**	**63.6**	**74.9**	35.8
Ours	Semi-supervised (25%)	48.4	60.3	71.7	81.0	55.9	54.5	66.8	77.2	85.1	48.6
	Semi-supervised (50%)	50.5	61.1	72.3	82.9	57.6	62.2	73.8	83.0	89.4	55.0
	Semi-supervised (75%)	**51.0**	**62.3**	**73.4**	**83.0**	**59.3**	**65.9**	**76.6**	**84.4**	**90.1**	**57.7**

4.2.2 Person re-ID

Implementation details and settings. Following [87], our model F employs an ImageNet-pretrained ResNet-50 [28]. All images are resized to $256 \times 128 \times 3$ in advance. During the first stage of meta-learning, we set the batch sizes of the meta-training and meta-validation sets to 32 and 128, respectively. We use the Adam optimizer to train our model for 600 epochs. The initial learning rate is set to 2×10^{-3}, and is decayed by a factor of 10 for every 150 epochs. The momentum and the margin ϕ are set to 0.9 and 0.3, respectively. As for the meta semi-supervised learning stage, we set the batch size of the labeled set to 32, the batch size of the unlabeled set to 128, and the initial learning rate to 2×10^{-5}. Similarly, the learning rate is decreased by a factor of 10 for every 150 epochs. We train our model for another 600 epochs. The similarity threshold ψ is set to 0.5. Also following [87], we evaluate our method with three different label ratios, i.e., $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{12}$ of the person IDs are fully labeled, while the remaining person IDs are unlabeled.

Results. We compare our method with unsupervised approaches [43, 96], semi-supervised methods [87, 88], a fully supervised approach [93], and a cross-dataset person re-ID method [91]. Similarly, the results of fully supervised/unsupervised methods can be regarded as the upper/lower bounds of our results. For the cross-dataset person re-ID method [91], we use their official implementation[3]

[3] https://github.com/KovenYu/MAR

Table 2: **Results of semi-supervised person re-ID.** The bold numbers indicate the best results.

Method	Supervision	Backbone	Market-1501 [96]		DukeMTMC-reID [58]	
			Rank 1	mAP	Rank 1	mAP
LOMO [43]	Unsupervised	-	27.2	8.0	12.3	4.8
BOW [96]		-	35.8	14.8	17.1	8.3
AlignedReID [93]	Supervised	ResNet-50	89.2	72.8	79.3	65.6
MVC [87]	Semi-supervised $(\frac{1}{12})$	ResNet-50	46.6	-	34.8	-
Ours		ResNet-50	**56.7**	**32.4**	**44.9**	**24.4**
MVC [87]	Semi-supervised $(\frac{1}{6})$	ResNet-50	60.0	-	43.8	-
Ours		ResNet-50	**70.8**	**46.4**	**56.6**	**33.6**
MVC [87]	Semi-supervised $(\frac{1}{3})$	ResNet-50	72.2	48.7	52.9	33.6
SPMVC [88]		ResNet-50	71.5	53.2	58.5	37.4
MAR [91]		ResNet-50	69.9	46.4	-	-
Ours		ResNet-50	**80.3**	**58.7**	**67.5**	**46.3**

with their default hyperparameter settings, and set the labeled set as their source domain and the unlabeled set as their target domain. Table 2 compares the rank 1 and mAP scores on the Market-1501 [96] and DukeMTMC-reID [58] datasets.

From this table, when comparing to semi-supervised learning methods, i.e., MVC [87] and SPMVC [88], our method consistently outperforms their results by large margins on all three evaluated label ratios of both datasets. When comparing to a fully-supervised method, e.g., AlignedReID [93], our method achieves 90% and 85% of their results recorded at rank 1 on the Market-1501 [96] and DukeMTMC-reID [58] datasets, respectively, using relatively fewer labeled information, i.e., only $\frac{1}{3}$ of the person IDs are labeled. From these results, we show that under the same experimental setting, our method achieves the state-of-the-art performance, while resulting in comparable results compared to fully supervised approaches.

4.3 Evaluation of Supervised Learning Tasks

4.3.1 Evaluation of Limited Labeled Data

In addition to evaluating the performance of our semi-supervised learning, we now apply our meta-learning strategy to the labeled training set only and see whether our learning-to-learn strategy would benefit such scenario.

Implementation details. All images are resized to $256 \times 128 \times 3$ in advance. We set the batch sizes of the meta-training and meta-validation sets to 32 and 128,

Table 3: **Results of fully-supervised person re-ID with limited training data.** The bold and underlined numbers indicate the top two results, respectively.

Method	Market-1501 [96]											
	$\frac{1}{3}$ of the IDs are available				$\frac{1}{6}$ of the IDs are available				$\frac{1}{12}$ of the IDs are available			
	Rank 1	Rank 5	Rank 10	mAP	Rank 1	Rank 5	Rank 10	mAP	Rank 1	Rank 5	Rank 10	mAP
DDML [47]	72.1	86.9	91.4	45.5	62.3	81.1	86.5	35.4	49.6	70.1	78.0	24.8
Triplet hard [62]	72.6	87.5	92.0	49.9	61.4	81.2	87.0	38.2	47.5	69.2	77.7	25.6
Triplet+HDML [98]	73.2	88.6	92.3	48.0	62.0	80.7	86.9	35.3	48.0	70.0	78.9	25.2
AlignedReID [93]	73.3	88.1	92.1	47.7	62.2	81.3	87.0	36.1	49.7	71.4	78.8	25.8
MGN [79]	74.1	88.2	92.1	50.8	62.3	81.4	86.7	38.1	50.6	71.6	79.8	27.4
BoT [48]	74.8	89.7	93.5	51.8	60.6	80.3	86.5	36.7	47.1	69.0	77.8	24.2
PyrNet [50]	74.9	89.7	92.7	52.1	63.2	81.8	87.2	37.8	50.2	71.7	79.9	28.4
Ours	**77.0**	**90.8**	**93.9**	**54.0**	**66.0**	**85.2**	**90.3**	**41.2**	**53.4**	**74.9**	**82.2**	**29.2**
	DukeMTMC-reID [58]											
PyrNet [50]	59.7	75.8	80.5	40.6	51.6	68.0	73.5	31.9	39.8	56.6	63.5	21.2
Triplet hard [62]	60.6	76.5	82.3	40.1	51.8	69.1	75.5	30.1	40.0	58.2	65.4	20.3
DDML [47]	60.6	75.0	79.1	36.7	51.5	67.3	73.2	29.6	40.1	57.4	64.3	20.1
MGN [79]	60.6	75.1	80.3	41.2	51.4	66.3	72.7	31.9	39.7	56.3	63.4	20.8
BoT [48]	61.9	78.5	83.8	40.5	52.4	68.9	75.9	31.8	40.7	57.6	64.3	21.1
AlignedReID [93]	62.5	77.1	82.2	40.3	51.3	68.2	75.6	29.6	40.5	58.1	65.4	20.5
Ours	**64.9**	**80.9**	**85.6**	**44.8**	**54.0**	**70.9**	**76.7**	**32.1**	**42.6**	**60.5**	**67.3**	**22.2**

respectively. We use the Adam optimizer to train our model for 600 epochs. The initial learning rate is set to 2×10^{-3}, and is decayed by a factor of 10 for every 150 epochs. The momentum and margin ϕ are set to 0.9 and 0.3, respectively.

Results. We adopt the Market-1501 [96] and DukeMTMC-reID [58] datasets for performance evaluations and compare our method with a number of supervised approaches [47, 48, 50, 62, 79, 93]. Table 3 presents the experimental results. The results show that our method consistently performs favorably against all competing approaches, demonstrating sufficient re-ID ability can be exhibited by our proposed method even when only limited labeled data are observed.

Visualization of the learned representations. To demonstrate that our model benefits from learning semantics-oriented similarity representation s, we select 20 person IDs and visualize both semantics-oriented similarity representation s and the learned feature representation f on the Market-1501 [96] test set via t-SNE in Figure 3, in which we compare our approach with AlignedReID [93] and its variant method.

We observe that without learning the semantics-oriented similarity representation, AlignedReID [93] and its variant method cannot separate the representation s well. Our method, on the other hand, learns semantics-oriented similarity representations from the labeled set in a learning-to-learn fashion. The learned similarity representation s allows our model to guide the learning of the unlabeled set, resulting in a well-separated space for the feature representation f.

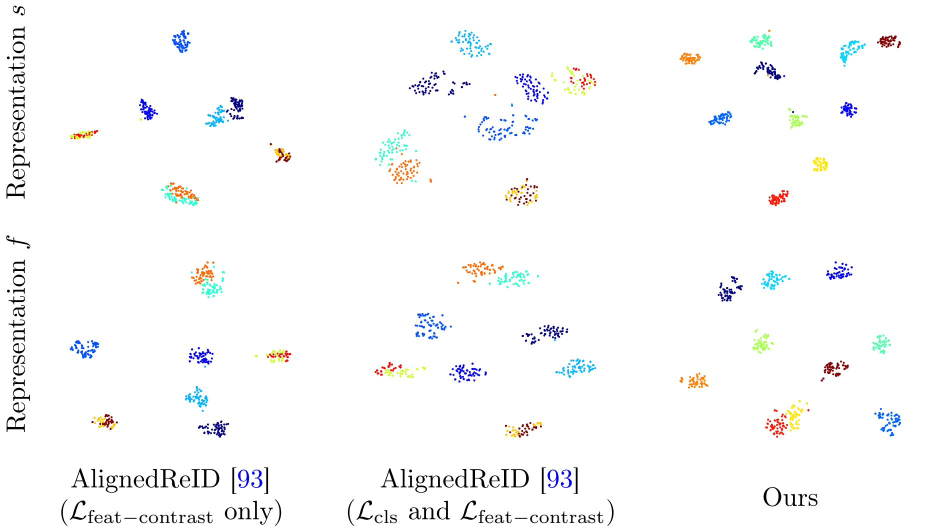


Fig. 3: **Visual comparisons of the learned representations on Market-1501.** (*Top row*) visualizes the semantics-oriented similarity representation. (*Bottom row*) visualizes the feature representation f. Note that selected samples of 20 identities are illustrated, each in a specific color. Comparing to AlignedReID, our model only learns semantics similarity and achieves comparable/improved performances.

4.3.2 Extension to Fully-Supervised Learning Tasks

Finally, to show that our formulation is not limited to semi-supervised learning settings, we apply our learning algorithm to fully-supervised setting on the Market-1501 [96] and DukeMTMC-reID [58] datasets.

Results. We initialize our model from AlignedReID [93] and BoT [48], respectively, and apply our meta-learning strategy on the entire training set, i.e., there are two variant methods: (1) AlignedReID [93] + Ours and (2) BoT [48] + Ours. As shown in Table 4, our method further improves the performance of AlignedReID [93] and BoT [48] on both datasets, respectively, comparing favorably against existing fully-supervised learning methods.

4.4 Limitations and Potential Issues

We observe that our method is memory intensive as learning from the unlabeled set requires larger batch size to increase the likelihood of selecting positive image pairs (sampling a negative pair is easier than sampling a positive pair). On the other hand, our learning algorithm is suitable for solving tasks where the categories are visually similar.

Table 4: **Results of fully-supervised person re-ID.** The bold and underlined numbers indicate the top two results, respectively.

Method	Market-1501 [96]			DukeMTMC-reID [58]		
	Rank 1	Rank 5	mAP	Rank 1	Rank 5	mAP
Part-Aligned [95]	81.0	92.0	63.4	-	-	-
PAN [99]	82.8	93.5	63.4	71.6	83.9	51.5
MGCAM [66]	83.8	-	74.3	-	-	-
TriNet [29]	84.9	94.2	69.1	-	-	-
JLML [40]	85.1	-	65.5	-	-	-
PoseTransfer [100]	87.7	-	68.9	78.5	-	56.9
PSE [61]	87.7	94.5	69.0	79.8	89.7	62.0
CamStyle [100]	88.1	-	68.7	75.3	-	53.5
DPFL [13]	88.9	92.3	73.1	79.2	-	60.6
AlignedReID [93]	89.2	95.9	72.8	79.3	89.7	65.6
DML [94]	89.3	-	70.5	-	-	-
DKP [63]	90.1	96.7	75.3	80.3	89.5	63.2
DuATM [64]	91.4	97.1	76.6	81.8	90.2	68.6
RDR [1]	92.2	97.9	81.2	85.2	**93.9**	72.8
SPReID [66]	93.7	97.6	83.4	85.9	92.9	73.3
BoT [48]	94.5	98.2	85.9	86.4	93.6	76.4
AlignedReID [93] + Ours	91.1	96.3	78.1	81.7	91.0	67.7
BoT [48] + Ours	**94.8**	**98.3**	**86.1**	**86.6**	**93.9**	**76.8**

5 Conclusions

We presented a meta-learning algorithm for semi-supervised learning with applications to image retrieval and person re-ID. We consider the training schemes in which labeled and unlabeled data share non-overlapping categories. Our core technical novelty lies in learning semantics-oriented similarity representation from the labeled set in a learning-to-learn fashion, which can be applied to semi-supervised settings without knowing the number of classes of the unlabeled data in advance. Our experiments confirmed that our method performs favorably against state-of-the-art image retrieval and person re-ID approaches in semi-supervised settings. We also verified that our algorithm can be applied to supervised settings for improved performance, which further exhibits the effectiveness and applicability of our learning algorithm.

Acknowledgments. This paper is supported in part by the Ministry of Science and Technology (MOST) of Taiwan under grant MOST 109-2634-F-002-037.

References

1. Almazan, J., Gajic, B., Murray, N., Larlus, D.: Re-id done right: towards good practices for person re-identification. arXiv (2018)
2. Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M.W., Pfau, D., Schaul, T., Shillingford, B., De Freitas, N.: Learning to learn by gradient descent by gradient descent. In: NeurIPS (2016)
3. Athiwaratkun, B., Finzi, M., Izmailov, P., Wilson, A.G.: There are many consistent explanations of unlabeled data: Why you should average. In: ICLR (2019)
4. Balaji, Y., Sankaranarayanan, S., Chellappa, R.: Metareg: Towards domain generalization using meta-regularization. In: NeurIPS (2018)
5. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.: Mixmatch: A holistic approach to semi-supervised learning. In: NeurIPS (2019)
6. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: ECCV (2018)
7. Chen, Y.C., Gao, C., Robb, E., Huang, J.B.: Nas-dip: Learning deep image prior with neural architecture search. In: ECCV (2020)
8. Chen, Y.C., Huang, P.H., Yu, L.Y., Huang, J.B., Yang, M.H., Lin, Y.Y.: Deep semantic matching with foreground detection and cycle-consistency. In: ACCV (2018)
9. Chen, Y.C., Li, Y.J., Du, X., Wang, Y.C.F.: Learning resolution-invariant deep representations for person re-identification. In: AAAI (2019)
10. Chen, Y.C., Lin, Y.Y., Yang, M.H., Huang, J.B.: Crdoco: Pixel-level domain transfer with cross-domain consistency. In: CVPR (2019)
11. Chen, Y.C., Lin, Y.Y., Yang, M.H., Huang, J.B.: Show, match and segment: Joint weakly supervised learning of semantic matching and object co-segmentation. TPAMI (2020)
12. Chen, Y., Hoffman, M.W., Colmenarejo, S.G., Denil, M., Lillicrap, T.P., Botvinick, M., de Freitas, N.: Learning to learn without gradient descent by gradient descent. In: ICML (2017)
13. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: CVPR (2016)
14. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: CVPR (2005)
15. Christopher, D.M., Prabhakar, R., Hinrich, S.: Introduction to information retrieval. Cambridge University Press (2008)
16. Chu, R., Sun, Y., Li, Y., Liu, Z., Zhang, C., Wei, Y.: Vehicle re-identification with viewpoint-aware metric learning. In: ICCV (2019)
17. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: CVPR (2019)
18. Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: CVPR (2018)
19. Ding, G., Zhang, S., Khan, S., Tang, Z., Zhang, J., Porikli, F.: Feature affinity based pseudo labeling for semi-supervised person re-identification. TMM (2019)
20. Dosovitskiy, A., Fischer, P., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with exemplar convolutional neural networks. TPAMI (2015)

21. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR (2010)
22. Figueira, D., Bazzani, L., Minh, H.Q., Cristani, M., Bernardino, A., Murino, V.: Semi-supervised multi-feature learning for person re-identification. In: AVSS (2013)
23. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML (2017)
24. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)
25. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: NeurIPS (2005)
26. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: CVPR (2006)
27. Harwood, B., Kumar, B., Carneiro, G., Reid, I., Drummond, T., et al.: Smart mining for deep metric learning. In: ICCV (2017)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
29. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv (2017)
30. Hochreiter, S., Younger, A.S., Conwell, P.R.: Learning to learn using gradient descent. In: ICANN (2001)
31. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: ICML (2018)
32. Huang, Y., Xu, J., Wu, Q., Zheng, Z., Zhang, Z., Zhang, J.: Multi-pseudo regularized label for generated data in person re-identification. TIP (2018)
33. Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial learning for semi-supervised semantic segmentation. In: BMVC (2018)
34. Iscen, A., Tolias, G., Avrithis, Y., Chum, O.: Mining on manifolds: Metric learning without labels. In: CVPR (2018)
35. Kaiser, Ł., Nachum, O., Roy, A., Bengio, S.: Learning to remember rare events. ICLR (2018)
36. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: ICCVW (2013)
37. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. In: ICLR (2017)
38. Lee, D.H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: ICMLW (2013)
39. Li, J., Ma, A.J., Yuen, P.C.: Semi-supervised region metric learning for person re-identification. IJCV (2018)
40. Li, W., Zhu, X., Gong, S.: Person re-identification by deep joint learning of multi-loss classification. IJCAI (2017)
41. Li, Y.J., Chen, Y.C., Lin, Y.Y., Du, X., Wang, Y.C.F.: Recover and identify: A generative dual model for cross-resolution person re-identification. In: ICCV (2019)
42. Li, Y.J., Chen, Y.C., Lin, Y.Y., Wang, Y.C.F.: Cross-resolution adversarial dual network for person re-identification and beyond. arXiv (2020)
43. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR (2015)
44. Liu, X., Song, M., Tao, D., Zhou, X., Chen, C., Bu, J.: Semi-supervised coupled dictionary learning for person re-identification. In: CVPR (2014)
45. Liu, Y., Song, G., Shao, J., Jin, X., Wang, X.: Transductive centroid projection for semi-supervised large-scale recognition. In: ECCV (2018)

46. Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X.: Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: CVPR (2016)
47. Lu, J., Hu, J., Tan, Y.P.: Discriminative deep metric learning for face and kinship verification. TIP (2017)
48. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: CVPRW (2019)
49. Luo, Y., Zhu, J., Li, M., Ren, Y., Zhang, B.: Smooth neighbors on teacher graphs for semi-supervised learning. In: CVPR (2018)
50. Martinel, N., Luca Foresti, G., Micheloni, C.: Aggregating deep pyramidal representations for person re-identification. In: CVPRW (2019)
51. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. TPAMI (2018)
52. Movshovitz-Attias, Y., Toshev, A., Leung, T.K., Ioffe, S., Singh, S.: No fuss distance metric learning using proxies. In: ICCV (2017)
53. Munkhdalai, T., Yu, H.: Meta networks. In: ICML (2017)
54. Oh Song, H., Jegelka, S., Rathod, V., Murphy, K.: Deep metric learning via facility location. In: CVPR (2017)
55. Oh Song, H., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: CVPR (2016)
56. Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A.: Deep co-training for semi-supervised image recognition. In: ECCV (2018)
57. Ravi, S., Larochelle, H.: Optimization as a model for few-shot learning. In: ICLR (2017)
58. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: ECCVW (2016)
59. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In: NeurIPS (2016)
60. Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., Lillicrap, T.: Meta-learning with memory-augmented neural networks. In: ICML (2016)
61. Saquib Sarfraz, M., Schumann, A., Eberle, A., Stiefelhagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: CVPR (2018)
62. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: CVPR (2015)
63. Shen, Y., Xiao, T., Li, H., Yi, S., Wang, X.: End-to-end deep kronecker-product matching for person re-identification. In: CVPR (2018)
64. Si, J., Zhang, H., Li, C.G., Kuen, J., Kong, X., Kot, A.C., Wang, G.: Dual attention matching network for context-aware feature sequence based person re-identification. In: CVPR (2018)
65. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: NeurIPS (2017)
66. Song, C., Huang, Y., Ouyang, W., Wang, L.: Mask-guided contrastive attention model for person re-identification. In: CVPR (2018)
67. Souly, N., Spampinato, C., Shah, M.: Semi supervised semantic segmentation using generative adversarial network. In: ICCV (2017)
68. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: NeurIPS (2014)
69. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: CVPR (2018)

70. Swets, D.L., Weng, J.J.: Using discriminant eigenfeatures for image retrieval. TPAMI (1996)
71. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: CVPR (2015)
72. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: CVPR (2014)
73. Tang, Z., Naphade, M., Birchfield, S., Tremblay, J., Hodge, W., Kumar, R., Wang, S., Yang, X.: Pamtri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data. In: ICCV (2019)
74. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS (2017)
75. Tieu, K., Viola, P.: Boosting image retrieval. IJCV (2004)
76. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: NeurIPS (2016)
77. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
78. Wang, G., Hu, Q., Cheng, J., Hou, Z.: Semi-supervised generative adversarial hashing for image retrieval. In: ECCV (2018)
79. Wang, G., Yuan, Y., Chen, X., Li, J., Zhou, X.: Learning discriminative features with multiple granularities for person re-identification. In: ACM MM (2018)
80. Wang, J., Zhou, F., Wen, S., Liu, X., Lin, Y.: Deep metric learning with angular loss. In: ICCV (2017)
81. Wang, J., Zhu, X., Gong, S., Li, W.: Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: CVPR (2018)
82. Wang, J., Kumar, S., Chang, S.F.: Semi-supervised hashing for large-scale search. TPAMI (2012)
83. Wang, P., Jiao, B., Yang, L., Yang, Y., Zhang, S., Wei, W., Zhang, Y.: Vehicle re-identification in aerial imagery: Dataset and approach. In: ICCV (2019)
84. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. JMLR (2009)
85. Wu, Y., Lin, Y., Dong, X., Yan, Y., Bian, W., Yang, Y.: Progressive learning for person re-identification with one example. TIP (2019)
86. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via non-parametric instance discrimination. In: CVPR (2018)
87. Xin, X., Wang, J., Xie, R., Zhou, S., Huang, W., Zheng, N.: Semi-supervised person re-identification using multi-view clustering. Pattern Recognition (2019)
88. Xin, X., Wu, X., Wang, Y., Wang, J.: Deep self-paced learning for semi-supervised person re-identification using multi-view self-paced clustering. In: ICIP (2019)
89. Xu, J., Shi, C., Qi, C., Wang, C., Xiao, B.: Unsupervised part-based weighting aggregation of deep convolutional features for image retrieval. In: AAAI (2018)
90. Ye, M., Zhang, X., Yuen, P.C., Chang, S.F.: Unsupervised embedding learning via invariant and spreading instance feature. In: CVPR (2019)
91. Yu, H.X., Zheng, W.S., Wu, A., Guo, X., Gong, S., Lai, J.H.: Unsupervised person re-identification by soft multilabel learning. In: CVPR (2019)
92. Zhang, J., Peng, Y.: Ssdh: semi-supervised deep hashing for large scale image retrieval. TCSVT (2017)
93. Zhang, X., Luo, H., Fan, X., Xiang, W., Sun, Y., Xiao, Q., Jiang, W., Zhang, C., Sun, J.: Alignedreid: Surpassing human-level performance in person re-identification. arXiv (2017)

94. Zhang, Y., Xiang, T., Hospedales, T.M., Lu, H.: Deep mutual learning. In: CVPR (2018)
95. Zhao, L., Li, X., Zhuang, Y., Wang, J.: Deeply-learned part-aligned representations for person re-identification. In: ICCV (2017)
96. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: ICCV (2015)
97. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: Past, present and future. arXiv (2016)
98. Zheng, W., Chen, Z., Lu, J., Zhou, J.: Hardness-aware deep metric learning. In: CVPR (2019)
99. Zheng, Z., Zheng, L., Yang, Y.: Pedestrian alignment network for large-scale person re-identification. TCSVT (2018)
100. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: CVPR (2018)
101. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)