

# Class Interference Regularization

Bharti Munjal<sup>1</sup>  
munjalbharti@gmail.com  
Sikandar Amin<sup>1</sup>  
sikandaramin@gmail.com  
Fabio Galasso<sup>2</sup>  
galasso@di.uniroma1.it

<sup>1</sup> OSRAM GmbH,  
Parkring 33 Garching,  
Munich 85748, Germany  
<sup>2</sup> Sapienza University of Rome,  
Piazzale Aldo Moro, 5,  
Rome 00185, Italy

---

## Abstract

Contrastive losses yield state-of-the-art performance for person re-identification, face verification and few shot learning. They have recently outperformed the cross-entropy loss on classification at the ImageNet scale and outperformed all self-supervision prior results by a large margin (SimCLR). Simple and effective regularization techniques such as label smoothing and self-distillation do not apply anymore, because they act on multinomial label distributions, adopted in cross-entropy losses, and not on tuple comparative terms, which characterize the contrastive losses.

Here we propose a novel, simple and effective regularization technique, the *Class Interference Regularization* (CIR), which applies to cross-entropy losses but is especially effective on contrastive losses. CIR perturbs the output features by randomly moving them towards the average embeddings of the negative classes. To the best of our knowledge, CIR is the first regularization technique to act on the output features.

In experimental evaluation, the combination of CIR and a plain Siamese-net with triplet loss yields best few-shot learning performance on the challenging tieredImageNet. CIR also improves the state-of-the-art technique in person re-identification on the Market-1501 dataset, based on triplet loss, and the state-of-the-art technique in person search on the CUHK-SYSU dataset, based on a cross-entropy loss. Finally, on the task of classification CIR performs on par with the popular label smoothing, as demonstrated for CIFAR-10 and -100.

## 1 Introduction

Contrastive losses yield state-of-the-art performance for person re-identification [4, 18, 21, 49, 83], face verification [0, 56] and few shot learning [67]. In their general formulation, contrastive losses imply processing the input samples with Siamese networks, then penalizing them if the output embeddings of two samples from the same class (*aka* positives) have higher distances than those from different classes (*aka* negatives). Most recent advances in self-supervised training for classification [20, 41] have leveraged contrastive losses and this is also the case for the current best method, SimCLR [5], which has set aside from the competition by a large margin at scale on ImageNet [8]. Notably, at the moment of writing, a novel technique based on contrastive loss [26] has just achieved best supervised-learning performance on ImageNet, outperforming for the first time the established cross-entropy loss.

In this work, we propose a novel, simple and effective regularization technique, the *Class Interference Regularization* (CIR), which applies to models trained with the traditional cross-entropy loss, but also and most effectively in the case of contrastive losses such as the triplet loss. CIR introduces a data-driven noise term. It works by estimating output features for each sample image in the batch and then randomly perturbing them with the average embeddings of their negative classes.

CIR fills in a gap in the training of neural networks with contrastive losses, because the widely adopted and effective label smoothing [60] and self-distillation [45] do not apply to those. Both the techniques ease the training with cross-entropy by perturbing the axis-aligned label distributions, also termed “one-hot vectors”. The first moves the labels off-axis; the second adopts soft-labels from prior rounds of training. Both of the techniques have been studied in-depth and have been widely adopted [10, 12, 53, 42]. However none of them applies in the case of contrastive losses because they require multinomial label distributions, while contrastive terms such as triplets use comparative embedding distances between (positive and negative) samples. By contrast, CIR applies both for cross-entropy and contrastive losses.

We thoroughly experiment with CIR in the case of contrastive and cross-entropy losses. Best performance improvements are achieved for the first. In particular, CIR sets a new state-of-the-art performance for few-shot learning on tieredImageNet [52], with a plain Siamese-net and a triplet loss, reaching 69.1% 1-shot and 82.9% 5-shot accuracies, with absolute margins of 2.8pp and 1.4pp on the second best respectively. Also CIR improves the plain Siamese-net + triplet in the case of person re-identification on the Market-1501 [78] and also slightly improves the best performer ABD-Net [9]. In the case of cross-entropy losses, CIR improves slightly but consistently a state-of-the-art person search technique [44] on the CUHK-SYSU dataset [77] and it yields better classification on the CIFAR-10 and -100 datasets [28], on par with the established label smoothing [60].

## 2 Related Work

**Regularization** is a major topic when learning over-parameterized models such as Deep Neural Networks (DNN). We review most relevant, recent and widely-adopted techniques by grouping them into three broad categories, depending on their application focus.

**Input samples.** Most common methods in this category regularize the training by data augmentation, i.e. by performing random transformations of the input samples such as cropping, rotation, flipping, noise injection or random erasing [9, 53, 82]. More complex methods use GANs to generate synthetic data [47] or add adversarial examples [16] to the training set.

**Network weights and hidden units.** Most popular techniques are weight decay [27] and dropout [59]. The first adds  $\ell_1$  or  $\ell_2$  norms of the weights into the loss, to bias training towards simpler solutions. The second randomly drops neurons to avoid weight co-adaptation, as also targeted by the variants DropConnect [65] and Adaptive Dropout [11]. Other technique regularize via stochastic pooling [77], depth [24], or by integrating adversarial noise layers in the CNN [76].

**Label distributions.** Most utilized in this category is Label Smoothing [42, 60] that moves the axis-aligned target label distribution (“one-hot vectors”) off the axis, thus softening it. Also widely adopted is Self-Distillation [22, 45] which softens the labels via iterations of trainings on generations of network predictions. Also in this category are Disturb Label [72], which randomly flips the ground-truth labels of some input images into wrong ones, and

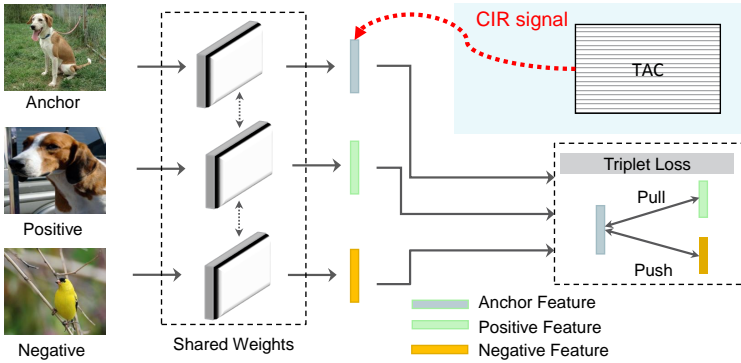


Figure 1: Illustration of the proposed Class Interference Regularization (CIR), as applied to a triplet-Siamese neural network model with triplet loss. During training, the average per-class output features are accumulated into a Table of Average Class embeddings (TAC) and used to perturb the anchor with the average embedding of a randomly-sampled wrong class. The illustration refers to few-shot learning. Cropped person IDs are the input images in the case of person re-identification. The Cyan box shows the proposed CIR signal.

Mixup [23], changing the target label by mixing input images in known proportions.

CIR differs from all of the above because it applies to the output features of a DNN. CIR is closest in spirit to techniques which regularize via label distributions. However established techniques such as label smoothing and self-distillation only apply to cross-entropy losses, while CIR applies to both cross-entropy and contrastive losses. Also, CIR is data-driven, similarly to *e.g.* data augmentation techniques using PCA on the RGB pixels values [28]. But CIR is the first to act on the output space.

*Multiple tasks* are here considered to thoroughly evaluate the benefits of CIR, which we also briefly review for related work.

**Few-Shot Learning.** This task targets classification of query samples from a single or few training (*aka* support) samples. Methods are broadly split into optimization- and metric-based. The first follow from MAML [9] and aim to learn good initial parameters of a learner to adapt with gradient descent [25, 64, 40, 45, 51, 55]. The second learn a common embedding space for both the support and query samples [13, 14, 39, 60, 64, 67, 58] and have been popularized by prototypical networks [58]. We apply CIR to improve the prototypical network with triplet loss technique of [67] and achieve state-of-the-art results.

**Person Re-Identification.** This targets retrieving a (query) person identity from a gallery of individuals, using cropped images of the person bounding boxes. While earlier methods employed cross-entropy [71, 79], more recent and better performing ones adopt triplet loss [21, 67, 88, 49] to learn to generate unique feature embeddings for each person ID. We apply CIR and improve performance of a plain triplet loss siamese-net approach [21] as well as of the current best method [9].

**Person Search.** This stands for the joint detection and re-identification of individuals in galleries of full images, given a single query [73]. Most recent and best approaches use the Online Instance Matching (OIM) [71] to build up look tables of people ID representative embeddings [3, 43, 42, 59, 74]. We apply CIR and slightly improve performance of the state-of-the-art work of [42], which employs the cross-entropy loss.

### 3 Class Interference Regularization (CIR)

We propose CIR to regularize the training of multi-class DNNs, both for cross-entropy and for contrastive losses. In this section, we first introduce the class interference signal (cf. Fig. 1); then we detail the implementation of CIR for the tasks of re-identification, few-shot learning and person search; finally we discuss CIR more formally and the intuition behind it.

#### 3.1 Class Interference

CIR introduces a table  $\Gamma \in \mathbb{R}^{C \times d}$  of average class embeddings into DNN models to track the mean embeddings  $\mu_c \in \mathbb{R}^d$  for each of the  $C$  classes. Then for each image  $\mathbf{x}_i$  with feature embedding  $\mathbf{z}_i \in \mathbb{R}^d$  and class  $y_i$ , CIR randomly selects a mean class embedding  $\mu_c$  of another class  $c \in C$  from  $\Gamma$ , with  $c \neq y_i$ , to corrupt  $\mathbf{z}_i$ . This yields a new blended embedding

$$\tilde{\mathbf{z}}_i = (1 - \lambda)\mathbf{z}_i + \lambda\mu_c \quad (1)$$

where  $\lambda \in [0, 1]$  controls the amount of interference. In other words, the polluted embedding  $\tilde{\mathbf{z}}_i$  is given by  $\mathbf{z}_i$  “pushed” towards the mean embedding  $\mu_c$  of a wrong class  $c$ . This interference makes the optimization tougher and reduces overfitting. We further discuss CIR and provide an intuition to it in Sec. 3.3.

**Input:** Training data:  $\mathcal{D} = \{(\mathbf{x}_n, y_n)\}_{n=1}^N$ , where  $\mathbf{x}_n$  represents an image,  $y_n$  its corresponding person ID (re-id) or object class (few-shot learning); hyper-parameters:  $\Gamma$  (TAC) update momentum  $\gamma$ , interference amount  $\lambda$ , margin  $\delta$ , max. iterations  $T$ , learning rate  $\alpha$

**Initialization:** Triplet-siamese network model with initial parameters  $\theta^{(0)}$  of the feature extractor  $f(\mathbf{x}_n, \theta^{(0)}) \in \mathbb{R}^d$ , and  $\Gamma^{(0)} \in \mathbb{R}^{C \times d}$  is randomly initialized TAC

**for**  $t = 1, \dots, T$  **do**

$\mathcal{D}_t = \{(\mathbf{a}_i, \mathbf{p}_i, \mathbf{n}_i)\}_{i=1}^B \leftarrow$  select a mini-batch of triplets of size  $B$  from the training set where  $\mathbf{a}_i, \mathbf{p}_i$  are from the same id/class  $y_i$  and  $\mathbf{n}_i$  is from another identity

**for**  $i = 1, \dots, B$  **do**

$\mathbf{z}_i^a = f(\mathbf{a}_i, \theta^{(t-1)}) \leftarrow$  compute feature embeddings for anchor

$\mathbf{z}_i^p = f(\mathbf{p}_i, \theta^{(t-1)}) \leftarrow$  compute feature embeddings for positive

$\mathbf{z}_i^n = f(\mathbf{n}_i, \theta^{(t-1)}) \leftarrow$  compute feature embeddings for negative

$\mu_c = \Gamma^{(t-1)}[c] \leftarrow$  average embedding of randomly chosen class  $c \neq y_i$  from TAC

$\tilde{\mathbf{z}}_i^a = (1 - \lambda)\mathbf{z}_i^a + \lambda\mu_c \leftarrow$  class interference acc. to Eq. 1 only for anchor embedding

$\mathcal{L}_i = \mathcal{L}_{\text{triplet}}(\tilde{\mathbf{z}}_i^a, \mathbf{z}_i^p, \mathbf{z}_i^n, \delta) \leftarrow$  triplet loss acc. to Eq. 2

**end**

$\theta^{(t)} \leftarrow \theta^{(t-1)} + \alpha \frac{1}{\|\mathcal{D}_t\|} \sum_{i \in \mathcal{D}_t} \nabla_{\theta^{(t-1)}} (\mathcal{L}_i)$

$\Gamma^{(t)} = (1 - \gamma)\Gamma^{(t-1)} + \gamma \mathbf{z}_{\mathcal{D}_t} \leftarrow$  update TAC per class using corresponding embeddings from the current mini-batch during backward pass

**end**

**Output:** Trained model parameters  $\theta^{(T)}$

**Algorithm 1:** Application of CIR to person re-identification and few-shot learning with triplet loss.

#### 3.2 Application to selected tasks

We employ CIR for re-identification, few-shot learning and person search. In all cases, CIR is applied with minor changes and provides consistent performance improvements (cf. Sec. 4).

**Person Re-identification:** In re-identification (re-id) the model is tasked with the identification of the persons in the query, provided as crops. State-of-the-art approaches in re-id employ the triplet loss for feature learning, which is formulated as:

$$\mathcal{L}_{triplet}(\mathbf{a}_i, \mathbf{p}_i, \mathbf{n}_j, \delta) = \max(0, \delta + \|\mathbf{a}_i - \mathbf{p}_i\|^2 - \|\mathbf{a}_i - \mathbf{n}_j\|^2) \quad (2)$$

where  $\mathbf{a}_i$  and  $\mathbf{p}_i$  are anchor and positive samples, respectively, for the positive class  $i$  and  $\mathbf{n}_j$  is the sample of the negative class  $j$ . While  $\delta$  represents the expected margin between inter-class and intra-class distances. For the task of re-id, we propose class interference as follows: first, we introduce a Table of Average Class embeddings (TAC) to accumulate the mean identity specific features; then, we add noise to the anchor  $\mathbf{a}_i$  according to Eq. 1 with randomly sampled mean class embedding from TAC. We outline the procedure in Algorithm 1.

**Few-Shot Learning:** One-shot learning is in essence quite similar to re-identification as it aims to learn a model which is able to classify images having only seen one example per class. For training, Prototypical loss is common in this case, or similar to re-identification Triplet-loss is also applicable. In our initial experiments, we found that the performance of Proto lags behind the Triplet, therefore, we opt for the latter for our experiments. The use of Triplet loss also makes the application of CIR, in this case, similar to few-shot learning. Hence, the same algorithm1 applies.

**Person Search:** OIM [14] is one of the most common approach for person search. Many recent state-of-the-art papers [55, 43, 42] rely on OIM loss for feature learning. During OIM training, the output feature of a person identity is matched against the TAC lookup table. Using Eq. 1, we corrupt the output features of the person identity with a randomly chosen mean embedding from the TAC. We outline the procedure in Algorithm 2.

**Input:** Training data:  $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$ , where  $\mathbf{x}_n$  represents an image, bold-face  $\mathbf{y}_n$  is corresponding ground-truth person ID and the bounding box, while  $y_n$  represents only the ground-truth person ID; hyper-parameters:  $\Gamma$  (TAC) update momentum  $\gamma$ , interference amount  $\lambda$ , max. iterations  $T$ , and learning rate  $\alpha$

**Initialization:** OIM [14] network model with initial parameters  $\theta^{(0)}$  of the feature extractor  $f(\mathbf{x}_n, \theta^{(0)}) \in \mathbb{R}^d$ , and  $\Gamma^{(0)} \in \mathcal{R}^{C \times d}$  is randomly initialized TAC

**for**  $t = 1, \dots, T$  **do**

$\mathcal{D}_t = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^B \leftarrow$  select a mini-batch of size  $B$  from the training set

**for**  $i = 1, \dots, B$  **do**

$\mathbf{z}_i = f(\mathbf{x}_i, \theta^{(t-1)}) \leftarrow$  compute feature embeddings

$\mu_j = \Gamma^{(t-1)}[c] \leftarrow$  average embedding of randomly chosen class  $c \neq y_i$  from TAC

$\tilde{\mathbf{z}}_i = (1 - \lambda)\mathbf{z}_i + \lambda\mu_j \leftarrow$  class interference acc. to Eq. 1

$\hat{y}_i = \arg \max_c \left( \Gamma^{(t-1)} \tilde{\mathbf{z}}_i \right) \leftarrow$  predicted person ID

$\mathcal{L}_i = \mathcal{L}_{\text{FRCNN}}(\mathbf{y}_i, \hat{\mathbf{y}}_i) + \mathcal{L}_{\text{OIM}}(y_i, \hat{y}_i) \leftarrow$  loss for person search as in [14]

**end**

$\theta^{(t)} \leftarrow \theta^{(t-1)} + \alpha \frac{1}{\|\mathcal{D}_t\|} \sum_{i \in \mathcal{D}_t} \nabla_{\theta^{(t-1)}} (\mathcal{L}_i)$

$\Gamma^{(t)} = (1 - \gamma)\Gamma^{(t-1)} + \gamma \mathbf{z}_{\mathcal{D}_t} \leftarrow$  update TAC per class using corresponding embeddings from the current mini-batch during backward pass

**end**

**Output:** Trained model parameters  $\theta^{(T)}$

**Algorithm 2:** Application of CIR to person search.

### 3.3 Intuition and discussion on CIR

We explain the regularizing effect of CIR on the network training by a simple study case. Assume the regression task of learning an image embedding  $\mathbf{z}_i$  for the  $i$ -th image  $\mathbf{x}_i$ , according to the target ground-truth embedding  $\mathbf{y}_i$ . We would assume that  $\mathbf{z}_i$  be the result of a simple linear relation, i.e. 1-layer fully-connected network,  $\mathbf{z}_i = W\mathbf{x}_i$ . In the equation,  $W$  are the current network parameters. The regression loss is given by

$$\mathcal{L}(W, \mathbf{x}_i, \mathbf{y}_i) = \frac{1}{2} \|W\mathbf{x}_i - \mathbf{y}_i\|^2 = \frac{1}{2} \|\mathbf{z}_i - \mathbf{y}_i\|^2 \quad (3)$$

When applying CIR, we substitute for  $\mathbf{z}_i$  with  $(1 - \lambda)\mathbf{z}_i + \lambda\mu_c$ , as provided by Eq. (1). This yields a regularized loss  $\mathcal{L}_{\text{CIR}}$  given by:

$$\mathcal{L}_{\text{CIR}}(W, \mathbf{x}_i, \mathbf{y}_i) = \frac{1}{2} \|(1 - \lambda)\mathbf{z}_i + \lambda\mu_c - \mathbf{y}_i\|^2 = \frac{1}{2} \|(\mathbf{z}_i - \mathbf{y}_i) - \lambda(\mathbf{z}_i - \mu_c)\|^2 \quad (4)$$

Notice that by allowing interference we actually force  $\mathbf{z}_i$  to come closer to the average embedding  $\mu_c$  of the wrong class. However, the optimization as a result tries to push the classes even further apart so that even a noisy embedding stays far away from the average embedding of other classes. We further support this intuition with an analysis of feature embeddings in Sec. 4.1

## 4 Experiments

Here we first evaluate CIR on algorithms based on triplet losses for Few Shot Learning and Person Re-identification; then we benchmark it on algorithms for Person Search and Classification adopting cross-entropy losses.

### 4.1 Few Shot Learning

**Dataset and metrics.** We consider the **miniImageNet** [62] and **tieredImageNet** [57] datasets. The first is a subset of ILSVRC-12 dataset [54] with 100 classes in total and 600 images per class. miniImageNet is divided into 64, 16, and 20 classes for meta-training, meta-validation, and meta-testing, respectively. tieredImageNet [57] is a larger and more complex subset of ILSVRC-12 with hierarchical structure. It contains 608 classes and 779,165 images in total. The classes are grouped into 34 broader categories according to WordNet [8] with 20 training, 6 validation, and 8 testing subsets. Following [58], we report the classification accuracy by taking the average over 600 randomly generated episodes from test set.

**Implementation Details.** Our work is based upon an open-source implementation of the ProtoNet<sup>1</sup> with Resnet18 [19] as backbone architecture. However, we use a triplet-siamese network with triplet loss for optimization. We train our model in two settings, from scratch (one-stage) and from a pre-trained model using softmax cross-entropy loss over all classes (two-stage). Note that for pre-training we do not use any extra data from the original ImageNet dataset. For triplet from scratch, we train for 300 epochs and decay the learning rate exponentially [21] after 200 epochs. For triplet pre-trained, we train for 100 epochs and decay the learning rate after 50 epochs. Each epoch has 100 iterations and the initial learning rate is set to 0.0002. We use online triplet mining with *Batch All* sampling strategy

<sup>1</sup> ProtoNet implementation for few shot learning

Method	miniImageNet		tieredImageNet		Market-1501		
	1-shot	5-shot	1-shot	5-shot	Method	mAP(%)	Rank-1(%)
ProtoNet [15], NIPS17	49.4	68.2	53.3	72.7	†Triplet [16], arXiv17	69.1	84.9
ProtoNet [15] (ResNet18), ICLR18	54.2	73.7	-	-	†MGN [17], ACM18	86.9	95.7
Large Margin (Triplet) [18], NIPS18	50.1	66.9	-	-	†SSP-ReID [18], IMAVIS19	75.9	89.3
Relation Net [19], CVPR18	50.4	65.3	54.5	71.3	HPM [19], AAAI19	82.7	94.2
Transductive Prop [20], ICLR19	55.5	69.9	59.9	73.3	†PyriNet [21], CVPR19	81.7	93.6
Incremental [22], NIPS19	54.9	63.0	56.1	65.5	†BoT [22], CVPR19	85.9	94.5
IDeMe-Net [23], CVPR19	57.7	74.3	-	-	DGNet [23], CVPR19	86.0	94.8
Individual Feature [24], ICLR19	56.9	70.5	-	-	CASN [24], CVPR19	82.8	94.4
CAML [25], ICLR19	59.2	72.3	-	-	†Auto-ReID [25], ICCV19	85.1	94.5
LEO [26], ICLR19	61.8	77.6	66.3	81.4	†OSNet [26], ICCV19	84.9	94.8
MetaOptNet [27], CVPR19	<b>62.6</b>	<b>78.6</b>	65.9	<b>81.5</b>	†P <sup>2</sup> -Net [27], ICCV19	85.6	95.2
Global Class Rep. [28], ICCV19	53.2	72.3	-	-	Triplet ours	70.4	84.5
Triplet	57.4	70.2	63.9	77.1	Triplet ours + CIR	<b>71.3</b>	<b>85.8</b>
+ CIR	<b>58.4</b>	<b>72.6</b>	<b>64.3</b>	<b>77.7</b>	†ABD-Net [29], ICCV19	88.3	<b>95.6</b>
Cross Entropy → Triplet	57.7	74.3	66.6	81.7	†ABD-Net [29] + CIR	<b>88.8</b>	<b>95.6</b>
+ CIR	<b>60.5</b>	<b>75.8</b>	<b>69.1</b>	<b>82.9</b>			

(a)

(b)

Table 1: (a) CIR on few shot learning for 5-way 1-shot and 5-way 5-shot. The average accuracy of 600 randomly generated episodes is reported for miniImageNet [15] and tieredImageNet [26]. (b) Comparative evaluation of CIR for the task of person re-identification on the Market-1501 dataset [28]. Methods indicated as † also use triplet loss and could also benefit from the simple CIR strategy.

as suggested in [21]. For miniImagenet, we use a batch size of 80 (20 classes, 4 samples per class) and for tieredImageNet we use batch size 256 (64 classes, 4 samples per class). We use update momentum  $\gamma=0.5$  for building the TAC and regularization momentum  $\lambda=0.5$  for class interference. The value of TAC update momentum is motivated from [21]. Note that, apart from keeping a TAC for average embeddings of different classes during training, there are no additional computational and memory space overheads, considering corruption of signal is negligible.

**Results.** In Table 1(a), we show the results of our evaluations in comparison to the state-of-the-art. We prefer triplet over prototypical loss for our baseline due to its superior performance (cf. Table 1(a)). We first evaluate the triplet-siamese network in one stage setting (training from scratch). On miniImageNet dataset, this model achieves an accuracy of 57.4% for 1 shot and 70.2% for 5 shot. For tieredImageNet, this model achieves 63.9% for 1 shot and 77.1% for 5 shot. As shown in the table, the addition of CIR to this model, improves its performance on miniImagenet by approximately 1pp (58.4 vs 57.4) for 1 shot and 2.4pp (72.6 vs 70.2) for 5 shot. For tieredImageNet, we observe marginal improvement.

We then evaluate our model in two-stage setting (cross-entropy followed by triplet) and report results in the last section of the table. As shown, this model provides a very strong baseline and CIR further improves this strong baseline significantly. For miniImagenet, CIR brings an improvement of almost 2.7pp for 1 shot and 1.5pp for 5 shot. For tieredImageNet, CIR brings an improvement of 2.5pp for 1 shot and 1.2pp for 5 shot. In the same Table 1(a), we also list the results of the state-of-the-art models on few-shot learning. On tieredImageNet we outperform the current best approaches LEO [26] by 2.8pp on 1-shot learning and MetaOptNet [27] by 1.4pp on 5-shot learning.

Furthermore, we perform a sanity check by adding gaussian noise to the feature embeddings instead of CIR, as a regularizer. These results are shown in Table 2 on tieredImageNet for 5-way 1-shot case. We notice that adding gaussian noise does not have any impact on the results, whereas CIR shows consistent improvements in all cases. This experiment allows us to understand that the contribution of CIR is more significant than just the random noise.

**CIR as a Regularizer :** We empirically demonstrate in Figure 2 that CIR acts as a reg-



Method	tieredImageNet (1-shot)
Triplet	63.9
Triplet + <i>Gaussian Noise</i>	63.6
Triplet + <i>CIR</i>	<b>64.3</b>
Cross Entropy $\rightarrow$ Triplet	66.6
Cross Entropy $\rightarrow$ Triplet + <i>Gaussian Noise</i>	66.9
Cross Entropy $\rightarrow$ Triplet + <i>CIR</i>	<b>69.1</b>

Table 2: Few-shot learning results for CIR vs Gaussian Noise on tieredImageNet [52]. The numbers represent the average accuracy of 600 randomly generated episodes.

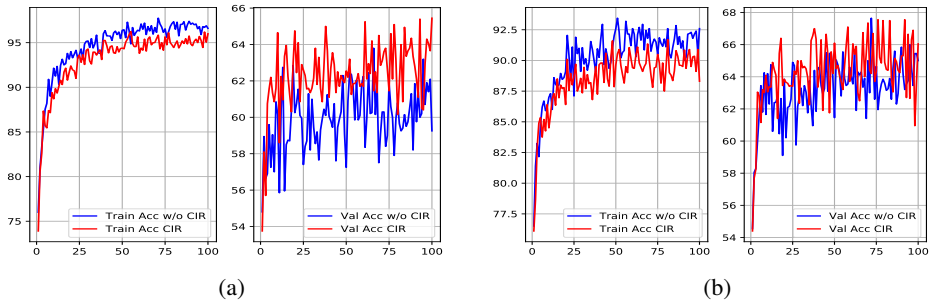


Figure 2: Training and validation accuracies with or without CIR on miniImagenet (a), tieredImageNet (b). Notice, improvement in validation accuracy and slight drop in training accuracy with CIR, indicating reduction in overfitting.

ularizer. Figure 2(a) shows the plots for training (left) and validation (right) accuracy on miniImagenet. Similarly, Figure 2(b) shows the plots for training (left) and validation (right) accuracy on tieredImageNet. As shown for both the datasets, our proposed CIR model has constantly higher validation accuracy than the one without CIR, while having lower training accuracy. This shows that CIR has a regularizing effect as it is able to generalize beyond the training set and decrease the gap in training and validation accuracy.

**Effect of CIR on the feature embedding :** To better support the intuition on CIR of Sec. 3.3, with reference to tieredImageNet [52], we compute the average distance from the overall center of mass of each data-point embedding and the ratio of inter-to-intra class mean distances. The first increases from 6.99 (w/o CIR) to 38.88 (w/ CIR), while the second changes from 1.02 (w/o CIR) to 1.17 (w/ CIR). This means that CIR effectively makes the feature-embedding space expand, but the embeddings from each class remain relatively compact.

## 4.2 Person Re-identification

**Dataset and metrics.** We adopt the **Market-1501** [73] dataset, which contains a total of 32,668 images representing the cropped bounding boxes of 1,501 persons. The train/test splits contain 750 and 751 identities respectively. For evaluation, we use the standard metrics, mAP and CMC rank-1.

**Implementation Details.** Triplet is the most common loss used in person re-id literature due to its superior performance. We re-implement our triplet baseline for person re-identification following [21]. We use pre-trained ResNet-50 architecture with input images re-scaled to  $256 \times 128$ . For augmentation, random crops and horizontal flipping are applied during training. For CIR, we employ TAC update momentum  $\gamma=0.5$  as discussed in Section 4.1 and



Method	CUHK-SYSU		CIFAR-10 Accuracy	CIFAR-10 Accuracy-test	CIFAR-100 Accuracy	CIFAR-100 Accuracy-test
	mAP(%)	top-1(%)				
OIM [14], CVPR17	75.5	78.7				
IAN [14], arXiv17	76.3	80.1				
NPSM [14], ICCV17	77.9	81.2				
†Mask-G [8], ECCV18	83.0	83.7				
CLSA [14], ECCV18	87.2	88.5				
†QEEPS [14], CVPR19	84.4	84.4				
†Context Graph [14], CVPR19	84.1	86.5				
OIM ours	77.8	78.1				
OIM ours + CIR	<b>79.3</b>	<b>80.0</b>				
†Distilled QEEPS [14] (Resnet18), BMVC19	84.1	84.3				
†Distilled QEEPS [14] (Resnet18) + CIR	<b>84.5</b>	<b>84.6</b>				

(a)

(b)

Table 3: (a) Comparative evaluation of CIR for the task of person search on the CUHK-SYSU [59] dataset. CIR boosts performance of OIM by a significant margin by regularizing its training. Methods indicated as † are built on top of OIM and could also benefit from the simple CIR strategy. (b) Top-1 Accuracy on CIFAR-10 and CIFAR-100 datasets. We follow the same implementation as in [14]. Accuracy-test is the accuracy of our implementation on actual test set.

regularization momentum  $\lambda=0.1$ .

**Results.** In Table 1(b), we show the results of our baseline re-implementation “Triplet ours” are slightly better than the original work [14]. We show that our proposed regularization CIR improves the baseline by 0.9pp mAP and 1.3pp CMC Rank-1. Note that other techniques are also available for person re-identification with better performance than [14], however most of them are still based on triplet loss. We add CIR on top of ABD-Net [4] which is the state-of-the-art in person re-identification and also uses triplet loss. As shown in the table, our proposed CIR brings an improvement of 0.5pp mAP and provides the best known mAP score for person re-identification on the Market-1501 dataset.

### 4.3 Person Search

Most recent approaches for person search are based on OIM [14] model. Hence, we also consider this as our baseline approach. To apply CIR in this case, we blend the ID feature embedding of a person with the average embedding from some other person ID from TAC.

**Dataset and metrics.** We adopt most commonly used CUHK-SYSU [14] dataset for benchmarking with 18,184 images labeled, 8,432 identities and 96,143 bounding boxes. We adopt the train/test split of [14]. The dataset presents challenging large variations in person appearance, background clutter and illumination changes. As metrics, we follow [14] and adopt mean Average Precision (mAP) and Common Matching Characteristic (CMC top-1).

**Implementation Details.** We re-implement the OIM person search algorithm of [14] in Pytorch, which we consider as our baseline. We use an image resolution of 600 pixels (shorter side). For CIR, we employ TAC update momentum  $\gamma=0.5$  as discussed in Section 4.1 and regularization momentum  $\lambda=0.5$ .

**Results.** We report in Table 3(a) the most recent relevant results together with ours. Our baseline “OIM ours” shows 77.8 mAP, slightly above the original OIM performance [14]. Implementing the CIR regularization on top of it yields 79.3 mAP and 80.0 top-1 CMC, improving the metrics by 1.5pp and 1.9pp, respectively. As shown in the table, most of these approaches [8, 43, 44, 74] are based on OIM, therefore CIR is directly applicable to them. We add CIR to the state-of-the-art person search method of Distilled QEEPS [44], which uses OIM. We adopt ResNet18 as the backbone of the Distilled QEEPS and use hyper-

parameters of the original paper [44]. As shown in the Table 3(a), the proposed CIR brings an improvement of 0.4pp mAP and 0.3pp top-1 on Distilled QEEPS [44].

## 4.4 Classification

Finally we demonstrate the application of CIR for the classification models trained with traditional cross-entropy loss. This also allows us to compare to label smoothing regularization [42].

**Dataset and metrics.** We consider **CIFAR-10** and **CIFAR-100** which are widely adopted datasets for natural image recognition. Both datasets are subsets from 80-million tiny image database [52] and contain 60k images ( $32 \times 32$ ) each. The train split has 50k images and test set has 10k images. CIFAR-10 has 10 categories while CIFAR-100 has 100.

**Implementation Details.** We use AlexNet [49] for CIFAR-10 and ResNet-56 [19] for CIFAR-100, as suggested in [42]. However, for a transparent evaluation and reproducibility of results by the community, we evaluate our model on publicly available test set (Accuracy-test in Table 3 (b)), unlike [42] which defines its own private validation set for both CIFAR-10 and CIFAR-100. For both datasets, we employ CIR with TAC update momentum  $\gamma=0.5$  as discussed in Section 4.1 and regularization momentum  $\lambda=0.1$ .

**Results.** In Table 3, we notice that both Label Smoothing [51] and our proposed CIR give minor but consistent improvement for both CIFAR-10 and -100, over model without regularization. In future work, we aim to study further applications for our proposed CIR.

## 5 Conclusions

We have proposed CIR, a novel, simple and effective regularization technique. CIR applies to cross entropy losses but is especially suited to contrastive losses. CIR is the first to act on the output features and it parallels established regularization techniques acting on the label distributions such as label smoothing and self-distillation, which do not apply to contrastive losses. In experimental evaluation we have shown that CIR improves consistently the performance of few-shot learning and person re-identification – for contrastive losses – and person search and classification – for cross-entropy losses. In the latter case, improvements are more modest, but on par with label smoothing. Given the rising popularity of contrastive losses and given the simplicity of CIR, we hope that it would play a role in future model trainings.

## References

- [1] Jimmy Ba and Brendan Frey. Adaptive dropout for training deep neural networks. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3084–3092. Curran Associates, Inc., 2013. URL <http://papers.nips.cc/paper/5032-adaptive-dropout-for-training-deep-neural-networks.pdf>.
- [2] Christopher M. Bishop. Training with noise is equivalent to tikhonov regularization. *Neural Computation*, 7(1):108–116, January 1995. ISSN 0899-7667.
- [3] Di Chen, Shanshan Zhang, Wanli Ouyang, Jian Yang, and Ying Tai. Person search via a mask-guided two-stream cnn model. In *The European Conference on Computer Vision (ECCV)*, September 2018.

- [4] Tianlong Chen, Shaojin Ding, Jingyi Xie, Ye Yuan, Wuyang Chen, Yang Yang, Zhou Ren, and Zhangyang Wang. Abd-net: Attentive but diverse person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019.
- [5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ArXiv*, 2020.
- [6] Zitian Chen, Yanwei Fu, Yu-Xiong Wang, Lin Ma, Wei Liu, and Martial Hebert. Image deformation meta-networks for one-shot learning. In *CVPR*, 2019.
- [7] Sumit Chopra, Raia Hadsell, and Yann Lecun. Learning a similarity metric discriminatively, with application to face verification. volume 1, pages 539– 546 vol. 1, 07 2005. ISBN 0-7695-2372-2. doi: 10.1109/CVPR.2005.202.
- [8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [9] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.
- [10] Yang Fu, Yunchao Wei, Yuqian Zhou, Honghui Shi, Gao Huang, Xinchao Wang, Zhiqiang Yao, and Thomas Huang. Horizontal pyramid matching for person re-identification. *AAAI*, 2019.
- [11] Tommaso Furlanello, Zachary Chase Lipton, Michael Tschannen, Laurent Itti, and Anima Anandkumar. Born again neural networks. *ArXiv*, abs/1805.04770, 2018.
- [12] J. Chorowski Ł. Kaiser G. Pereyra, G. Tucker and G. Hinton. Regularizing neural networks by penalizing confident output distributions. In *ICLR*, 2017.
- [13] Victor Garcia and Joan Bruna. Few-shot learning with graph neural networks. In *ICLR*, 2018.
- [14] Victor Garcia and Joan Bruna. Meta-learning with individualized feature space for few-shot classification. In *OpenReview*, 2019.
- [15] Spyros Gidaris and Nikos Komodakis. Dynamic fewshot visual learning without forgetting. In *CVPR*, 2018.
- [16] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [17] Jonathan Gordon, John Bronskill, Matthias Bauer, Sebastian Nowozin, and Richard Turner. Meta-learning probabilistic inference for prediction. In *ICLR*, 2019.
- [18] Jianyuan Guo, Yuhui Yuan, Lang Huang, Chao Zhang, Jin-Ge Yao, and Kai Han. Beyond human parts: Dual part-aligned representations for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

- [20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020.
- [21] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *CoRR*, abs/1703.07737, 2017. URL <http://arxiv.org/abs/1703.07737>.
- [22] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015. URL <http://arxiv.org/abs/1503.02531>.
- [23] Yann N. Dauphin David Lopez-Paz Hongyi Zhang, Moustapha Cisse. mixup: Beyond empirical risk minimization. *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=r1Ddp1-Rb>.
- [24] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Weinberger. Deep networks with stochastic depth. In *ECCV*, 2016.
- [25] Xiang Jiang, Mohammad Havaei, Farshid Varno, Gabriel Chartrand, Nicolas Chapados, and Stan Matwin. Learning to learn with conditional class dependencies. In *ICLR*, 2019. URL <https://openreview.net/forum?id=BJfOXnActQ>.
- [26] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *ArXiv*, 2020.
- [27] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.
- [28] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 (canadian institute for advanced research). URL <http://www.cs.toronto.edu/~kriz/cifar.html>.
- [29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- [30] Xu Lan, Xiatian Zhu, and Shaogang Gong. Person Search by Multi-Scale Matching. In *ECCV2018*, 2018.
- [31] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *CVPR*, 2019.
- [32] Aoxue Li, Tiange Luo, Tao Xiang, Weiran Huang, and Liwei Wang. Few-shot learning with global class representations. pages 9714–9723, 10 2019. doi: 10.1109/ICCV.2019.00981.
- [33] Weizhi Li, Gautam Dasarathy, and Visar Berisha. Regularization via structural label smoothing, 01 2020.

- [34] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few shot learning. *CoRR*, abs/1707.09835, 2017. URL <http://arxiv.org/abs/1707.09835>.
- [35] H. Liu, J. Feng, Z. Jie, K. Jayashree, B. Zhao, M. Qi, J. Jiang, and S. Yan. Neural person search machines. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 493–501, Oct 2017. doi: 10.1109/ICCV.2017.61.
- [36] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, and Yi Yang. Transductive propagation network for few-shot learning. *CoRR*, abs/1805.10002, 2018. URL <http://arxiv.org/abs/1805.10002>.
- [37] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [38] Niki Martinel, Gian Luca Foresti, and Christian Micheloni. Aggregating Deep Pyramidal Representations for Person Re-Identification. In *International Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [39] Sachin Ravi Jake Snell Kevin Swersky Joshua B. Tenenbaum Hugo Larochelle Mengye Ren, Eleni Triantafillou and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification. In *ICLR*, 2018.
- [40] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. In *ICLR*, 2018. URL <https://openreview.net/forum?id=B1DmUzWAW>.
- [41] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *ArXiv*, 2019.
- [42] Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. When does label smoothing help? In *NIPS*, 2019.
- [43] Bharti Munjal, Sikandar Amin, Federico Tombari, and Fabio Galasso. Query-guided end-to-end person search. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [44] Bharti Munjal, Fabio Galasso, and Sikandar Amin. Knowledge distillation for end-to-end person search. In *The British Machine Vision Conference (BMVC)*, 2019.
- [45] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *CoRR*, abs/1803.02999, 2018. URL <http://arxiv.org/abs/1803.02999>.
- [46] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In *NIPS*, 2018.
- [47] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *CoRR*, abs/1712.04621, 2017. URL <http://arxiv.org/abs/1712.04621>.

- [48] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L. Yuille. Few-shot image recognition by predicting parameters from activations. In *CVPR*, 2018.
- [49] Ruijie Quan, Xuanyi Dong, Yu Wu, Linchao Zhu, and Yi Yang. Auto-reid: Searching for a part-aware convnet for person re-identification. In *ICCV*, 2019.
- [50] Rodolfo Quispe and Helio Pedrini. Enhanced person re-identification based on saliency and semantic parsing with deep neural network models. *Image and Vision Computing*, 2019.
- [51] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2017.
- [52] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification. In *Proceedings of 6th International Conference on Learning Representations ICLR*, 2018.
- [53] Mengye Ren, Renjie Liao, Ethan Fetaya, and Richard Zemel. Incremental few-shot learning with attention attractor networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 5275–5285. Curran Associates, Inc., 2019. URL <http://papers.nips.cc/paper/8769-incremental-few-shot-learning-with-attention-attractor-networks.pdf>.
- [54] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.
- [55] Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In *ICLR*, 2019. URL <https://openreview.net/forum?id=BJgklhAcK7>.
- [56] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [57] Christian Simon, Piotr Koniusz, and Mehrtash Harandi. Projective subspace networks for few-shot learning, 2019. URL <https://openreview.net/forum?id=rkzfuiA9F7>.
- [58] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *NIPS*, 2017.
- [59] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL <http://jmlr.org/papers/v15/srivastava14a.html>.

- [60] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H.S. Torr, and Timothy M. Hospedales. Learning to compare: Relation network for few-shot learning. In *CVPR*, 2018.
- [61] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [62] Antonio Torralba, Rob Fergus, and William T. Freeman. 80 million tiny images: a large dataset for non-parametric object and scene recognition, 2008.
- [63] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, 2010. URL <http://dblp.uni-trier.de/db/journals/jmlr/jmlr11.html#VincentLLBM10>.
- [64] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *NIPS*, 2016.
- [65] Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. Regularization of neural networks using dropconnect. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1058–1066. PMLR, 2013. URL <http://proceedings.mlr.press/v28/wan13.html>.
- [66] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou. Learning Discriminative Features with Multiple Granularities for Person Re-Identification. *ArXiv e-prints*, 2018.
- [67] Yong Wang, Xiao-Ming Wu, Jiatao Gu Qimai Li, Wangmeng Xiang, Lei Zhang, and Victor OK Li. Large margin meta-learning for few-shot classification. In *Workshop on Meta-Learning (MetaLearn 2018) at NIPS*, 2018.
- [68] Zsolt Kira Yu-Chiang Frank Wang Wei-Yu Chen, Yen-Cheng Liu and Jia-Bin Huang. A closer look at few-shot classification. In *ICLR*, 2019.
- [69] Jimin Xiao, Yanchun Xie, Tamman Tillo, Kaizhu Huang, Yunchao Wei, and Jiashi Feng. Ian: The individual aggregation network for person search. *CoRR*, abs/1705.05552, 2017.
- [70] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1249–1258, 2016.
- [71] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3376–3385. IEEE, 2017.
- [72] Lingxi Xie, Jingdong Wang, Zhen Wei, Meng Wang, and Qi Tian. DisturbLabel: Regularizing CNN on the Loss Layer. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4753–4762, 2016.



- [73] Yuanlu Xu, Bingpeng Ma, Rui Huang, and Liang Lin. Person search in a scene by jointly modeling people commonness and person uniqueness. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 937–940, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3063-3. doi: 10.1145/2647868.2654965. URL <http://doi.acm.org/10.1145/2647868.2654965>.
- [74] Yichao Yan, Qiang Zhang, Bingbing Ni, Wendong Zhang, Minghao Xu, and Xiaokang Yang. Learning context graph for person search. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [75] Chenglin Yang, Lingxi Xie, Siyuan Qiao, and Alan Loddon Yuille. Training deep neural networks in generations: A more tolerant teacher educates better students. In *AAAI*, 2019.
- [76] Zhonghui You, Jinmian Ye, Kunming Li, and Ping Wang. Adversarial noise layer: Regularize neural network by adding noise. In *IEEE International Conference on Image Processing (ICIP)*, 2019.
- [77] M. Zeiler and R. Fergus. Stochastic pooling for regularization of deep convolutional neural networks. 2013.
- [78] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Computer Vision, IEEE International Conference on*, 2015.
- [79] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In *ECCV*, 2016.
- [80] Meng Zheng, Srikrishna Karanam, Ziyang Wu, and Richard J. Radke. Re-identification with consistent attentive siamese networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [81] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [82] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.
- [83] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *ICCV*, 2019.