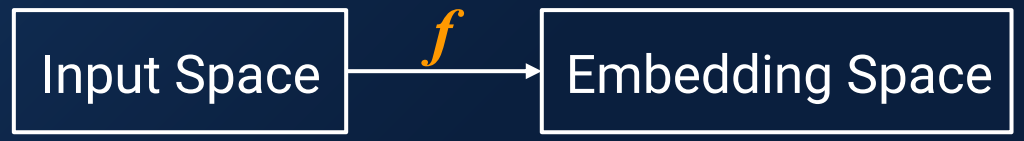
Deep Metric Learning by Online Soft Mining and Class-Aware Attention

Xinshao Wang, Yang Hua, Elyor Kodirov,
Guosheng Hu, and Neil M. Robertson

2019-01 @ AAAI2019

Background

- What is deep metric learning (DML)?
  - Objective: a deep embedding space such that relative locations of input samples are based on their semantic similarities.
  - Key point:



An excerpt of t-SNE plot on CUB-200-2011 test set.

QUEEN'S UNIVERSITY BELFAST anyvision.

Background

- DML is fundamental and learns deep representations => diverse applications
  - Image Retrieval (Song et al., CVPR 2016)
  - Person ReID (McLaughlin et al. CVPR 2016)
  - Clustering (Song et al., CVPR 2017)
  - Verification (Schroff et al., CVPR 2015)
  - Few-shot Learning (Vinyals et al., NeurIPS 2016)
  - Zero-shot Learning (Bucher et al., ECCV 2016)

Observations

- Existing problems of DML
  - Not making full use of all samples in the mini-batch
    - Attention/Mining is necessary: a large fraction of trivial samples.
    - Previous Solution: binary attention, i.e., hard sample mining, which assigns one binary score to each sample (dropping or keeping it).

      As a result, some examples which may be useful to some extent are removed.

Observations

- Existing problems of DML
  - Ignoring outlying samples
    - Motion blur

    - Occlusion

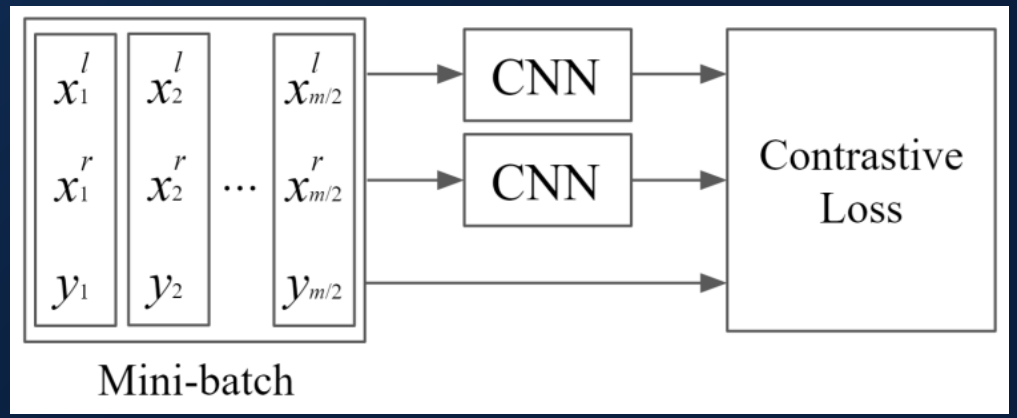
Observations

- Existing problems of DML
  - Ignoring outlying samples
    - Distractive objects

    - Truncated objects

Methodology

- Overview
  - Traditional contrastive loss for learning an embedding CNN f



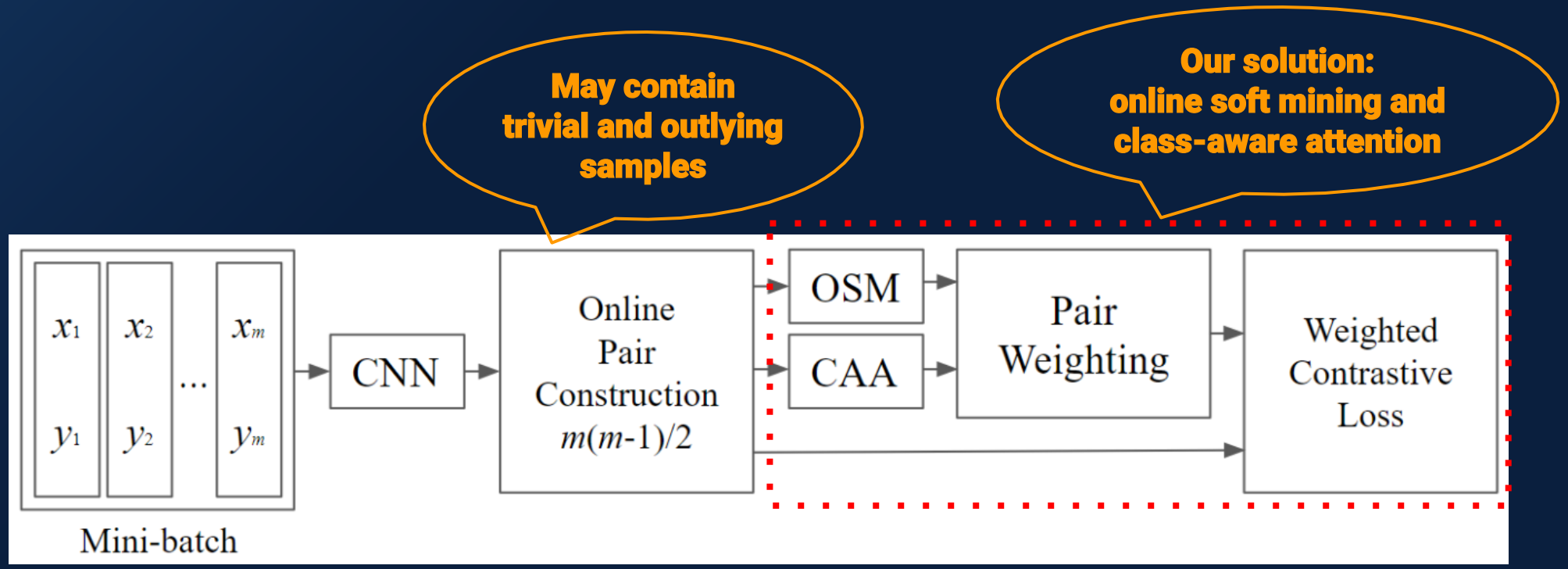
$$\mathbf{f}_i^l = f(\mathbf{x}_i^l) \in \mathbb{R}^D, \mathbf{x}_i^l \in \mathbb{R}^{H \times W \times 3}, y_i \in \{0, 1\}$$

$$d_i = \left\| \mathbf{f}_i^l - \mathbf{f}_i^r \right\|_2$$

$$L_{cont}^{\alpha}(\mathbf{x}_i^l, \mathbf{x}_i^r; f) = y_i d_i^2 + (1 - y_i) \max(0, \alpha - d_i)^2$$

Methodology

- Overview
  - Traditional contrastive loss
  - Our Proposal: Weighted contrastive loss with OSM and CAA



Methodology

- Online Soft Mining for Positives and Negatives
  - Online Soft Positives Mining: It assigns higher scores to local/closer positives motivated by learning extended manifolds

$$d_{ij} = \|\mathbf{f}_i - \mathbf{f}_j\|_2$$

$$s_{ij}^{+} = \exp(-d_{ij}^2 \Big/ {\sigma_{\mathrm{OSM}}}^2)$$

  - Online Soft Negatives Mining: It assigns higher scores to more difficult negatives

$$s_{ij}^{-} = \max(0, \alpha - d_{ij})$$

Methodology

- Class-Aware Attention
  - CAA score (sample's semantic relation to its label) is computed as:

$$a_i = \frac{\exp(\mathbf{f}_i^\top \mathbf{c}_{y_i})}{\sum_{k=1}^{C} \exp(\mathbf{f}_i^\top \mathbf{c}_k)}$$

$\mathbf{c}_k$ is the context vector of class k

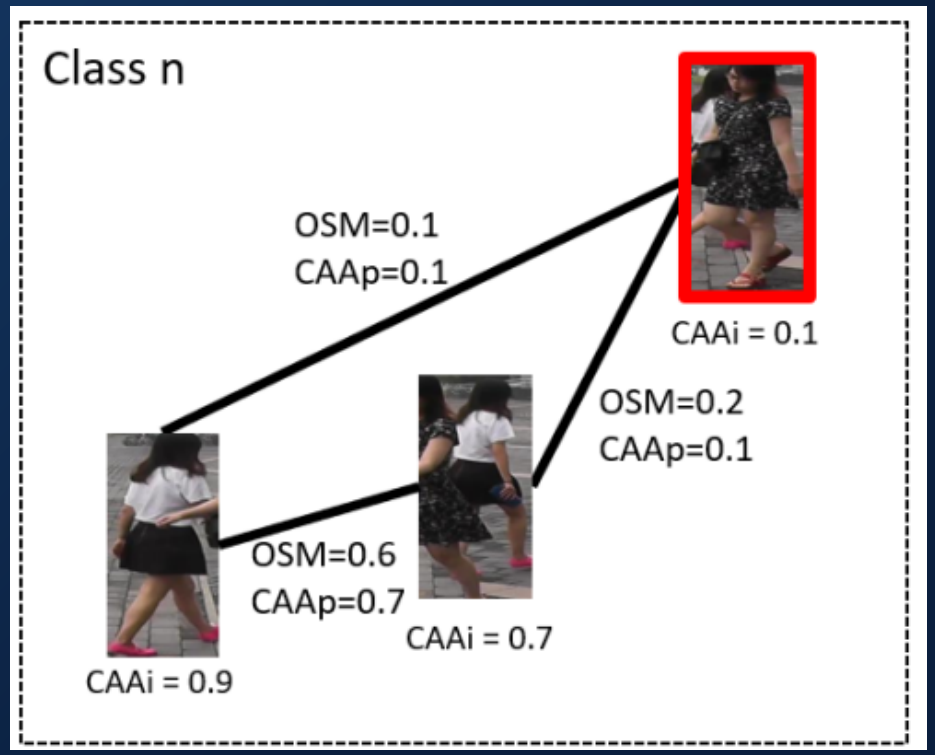
- Final Weight of Each Pair

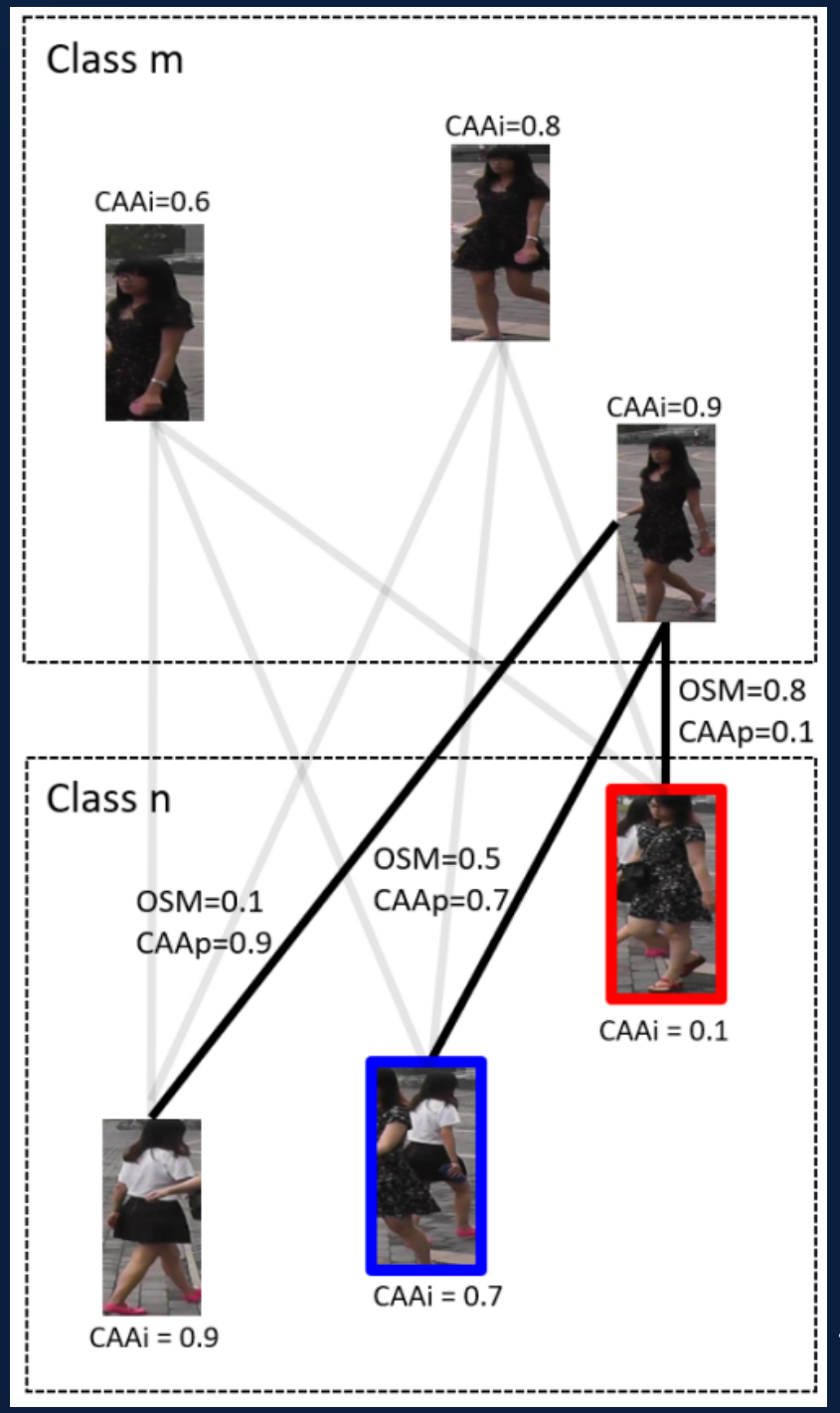
$$w_{ij}^{+} = s_{ij}^{+} * a_{ij} \qquad w_{ij}^{-} = s_{ij}^{-} * a_{ij}$$

$a_{ij} = \min(a_i, a_j)$ is the CAA score of pair $(\mathbf{x}_i, \mathbf{x}_j)$

Methodology

- Illustration of OSM and CAA





Experiments on Video-based Person Re-ID Domain

- Intrinsically, Person ReID is an image retrieval problem with some constraints (pose/camera-invariant).
- Video-based means each example is a video/tracklet instead of an image.

Experiments on Video-based Person Re-ID Domain

- Results on MARS: CMC-K(%) and mAP(%).

Methods	Attention	1	5	20	mAP
IDE (ResNet50) (Zhong et al. 2017)	No	62.7	–	–	44.1
IDE (ResNet50)+XQDA (Zhong et al. 2017)	No	70.5	–	–	55.1
IDE (ResNet50)+XQDA+Re-ranking (Zhong et al. 2017)	No	73.9	–	–	68.5
CNN+RNN (McLaughlin, del Rincon, and Miller 2017)	No	43.0	61.0	73.0	–
CNN+RNN+XQDA (McLaughlin, del Rincon, and Miller 2017)	No	52.0	67.0	77.0	–
AMOC+EpicFlow (Liu et al. 2017)	No	68.3	81.4	90.6	52.9
ASTPN (Xu et al. 2017)	Yes	44.0	70.0	81.0	–
SRM+TAM (Zhou et al. 2017)	Yes	70.6	90.0	**97.6**	50.7
RQEN (Song et al. 2018)	Yes	73.7	84.9	91.6	51.7
RQEN+XQDA+Re-ranking (Song et al. 2018)	Yes	77.8	88.8	94.3	71.1
DRSA (Li et al. 2018)	Yes	82.3	–	–	65.8
CAE (Chen et al. 2018)	Yes	82.4	92.9	–	67.5
Ours	Yes	**84.7**	**94.1**	97.0	**72.4**
Ours + Re-ranking(Zhong et al. 2017)	Yes	**86.0**	**94.4**	97.1	**81.0**

Experiments on Video-based Person Re-ID Domain

- Results on LPW: CMC-K(%).

Methods	Attention	1	5	20
GoogleNet(Song et al. 2018)	No	41.5	66.7	86.2
RQEN(Song et al. 2018)	Yes	57.1	81.3	91.5
Ours	Yes	**71.7**	**89.8**	**96.6**

Experiments on Fine-Grained Image Recognition

- To find the distinction between fine-grained classes, which is often quite subtle

CARS196 dataset

CUB-200-2011 dataset

Experiments on Fine-Grained Image Recognition

- Results on CARS196 and CUB-200-2011: Recall@K (%)
  - Raw images

	CARS196						CUB-200-2011					
K	1	2	4	8	16	32	1	2	4	8	16	32
Contrastive (Bell and Bala 2015)	21.7	32.3	46.1	58.9	72.2	83.4	26.4	37.7	49.8	62.3	76.4	85.3
Triplet (Schroff, Kalenichenko, and Philbin 2015)	39.1	50.4	63.3	74.5	84.1	89.8	36.1	48.6	59.3	70.0	80.2	88.4
LiftedStruct (Oh Song et al. 2016)	49.0	60.3	72.1	81.5	89.2	92.8	47.1	58.9	70.2	80.2	89.3	93.2
Binomial Deviance (Ustinova and Lempitsky 2016)	–	–	–	–	–	–	52.8	64.4	74. 7	83.9	90.4	94. 3
Histogram Loss (Ustinova and Lempitsky 2016)	–	–	–	–	–	–	50.3	61.9	72.6	82.4	88.8	93.7
Smart Mining (Harwood et al. 2017)	64.7	76.2	84.2	90.2	–	–	49.8	62.3	74.1	83.3	–	–
HDC* (Yuan, Yang, and Zhang 2017)	73.7	83.2	89.5	93.8	96.7	98.4	53.6	65.7	77.0	85.6	91.5	**95.5**
Ours	**74.0**	**83.8**	**90.2**	**94.8**	**97.3**	**98.6**	**55.3**	**67.3**	**77.5**	**85.8**	**91.8**	95.4

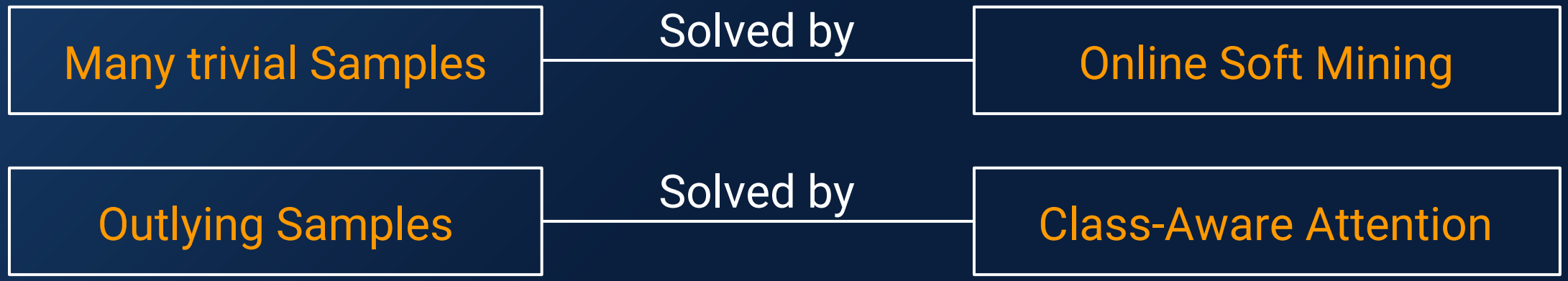
Experiments on Fine-Grained Image Recognition

- Results on CARS196 and CUB-200-2011: Recall@K (%)
  - Raw images
  - Cropped images

	CARS196						CUB-200-2011					
K	1	2	4	8	16	32	1	2	4	8	16	32
PDDM+Triplet (Huang, Loy, and Tang 2016)	46.4	58.2	70.3	80.1	88.6	92.6	50.9	62.1	73.2	82.5	91.1	94.4
PDDM+Quadruplet (Huang, Loy, and Tang 2016)	57.4	68.6	80.1	89.4	92.3	94.9	58.3	69.2	79.0	88.4	93.1	95.7
HDC* (Yuan, Yang, and Zhang 2017)	83.8	89.8	93.6	96.2	97.8	98.9	60.7	72.4	81.9	89.2	93.7	96.8
Ours	**85.5**	**91.5**	**95.1**	**97.2**	**98.5**	**99.2**	**62.3**	**73.2**	**83.3**	**89.6**	**94.1**	**96.9**

Summary

- In this work, we address two problems in deep metric learning.



- Simple, Intuitive & Effective.

Deep Metric Learning by

Online Soft Mining and Class-Aware Attention

Many Thanks !

Any Questions ?

QUEEN'S UNIVERSITY BELFAST anyvision. WE'RE HIRING http://www.anyvision.co/index.php/careers/