# Connections between DML and other fields







DML

# Person RelD





#### Face Recognition



- Overview: Deep Face Recognition: A Survey
- Current accuracy, e.g., MegaFace 99.1%

#### Most Famous: SphereFace, CosFace, ArcFace

- All start with softmax loss
- Reformulate to be depended on class center-sample angle
- Embedding vectors distributed around hypersphere





#### Most Famous: SphereFace, CosFace, ArcFace

- All start with softmax loss
- Reformulate to be depended on class center-sample angle
- Embedding vectors distributed around hypersphere

$$L_{\text{ang}} = \frac{1}{N} \sum_{i} -\log \Big( \frac{e^{\|\boldsymbol{x}_{i}\| \cos(\boldsymbol{m}\boldsymbol{\theta}_{y_{i},i})}}{e^{\|\boldsymbol{x}_{i}\| \cos(\boldsymbol{m}\boldsymbol{\theta}_{y_{i},i})} + \sum_{j \neq y_{i}} e^{\|\boldsymbol{x}_{i}\| \cos(\boldsymbol{\theta}_{j,i})}} \Big)^{[23]}$$

$$L_{lmc} = \frac{1}{N} \sum_{i} -\log \frac{e^{s(\cos(\theta_{y_{i},i}) - m)}}{e^{s(\cos(\theta_{y_{i},i}) - m)} + \sum_{j \neq y_{i}} e^{s\cos(\theta_{j,i})}}$$
<sup>[24]</sup>

$$L_{3} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_{i}} + m))}}{e^{s(\cos(\theta_{y_{i}} + m))} + \sum_{j=1, j \neq y_{i}}^{n} e^{s\cos\theta_{j}}}$$

[23] Liu, W. et al. "SphereFace: Deep Hypersphere Embedding for Face Recognition." (CVPR, 2017)
[24] Wang, H. et al. "CosFace: Large Margin Cosine Loss for Deep Face Recognition." (CVPR 2018)
[25] Deng, J. et al. "ArcFace: Additive Angular Margin Loss for Deep Face Recognition." (CVPR 2019)









DML

# Person RelD







#### **Person ReID**

- Common datasets: Market-1501, CUHK-03
- Evaluation: rank-k and mAP







#### → Often trained with cross-entropy (ID) and triplet loss

[26] Li, W. et al. "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification" (CVPR 2014)

#### **Person ReID - Common practices**



Current performance: 98.3 on Market-1501 dataset

#### Cross-dataset / domain adaptation, unsupervised, occluded person ReID

[14] Sun, Y. et al. "Beyond Part Models: Person Retrieval with Refined Part Pooling." ECCV (2018).
[15] Chen, T. et al. "ABD-Net: Attentive but Diverse Person Re-Identification." 2019 ICCV) (2019)
[16] Tay, C. et al. "AANet: Attribute Attention Network for Person Re-Identifications." CVPR (2019)

[17] Zhu, Z. et al "Viewpoint-Aware Loss with Angular Regularization for Person Re-Identification." AAAI (2020).

[18] Luo, Ha. et al. "Bags of Tricks and A Strong Baseline for Deep Person Re-identification." (2019).







DML







## **Multi-Object Tracking**



Tracks in past frame

Current Frame detections

Compute cost to assign detections to tracks using bounding box / position information

### **Multi-Object Tracking**



Compute cost to assign detections to tracks using bounding box / position information
 Also use appearance information, *i.e.*, person ReID

[27] Leal-Taixé, L. et al. "MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking." (arXiv, 2015).

# **Multi-Object Tracking**



Compute cost to assign detections to tracks using bounding box / position information
 Also use appearance information, *i.e.*, person ReID







DML

# Person RelD





#### Zero- / One- / Few-Shot Learning

N-way k-shot learning task:

k labelled examples from each of N classes (=> not seen during!)

Classify disjoint batch of unlabelled examples into one of these N classes. Classification accuracy = #correct/#total

Few-Shot Learning



$$p_{oldsymbol{\phi}}(y=k \,|\, \mathbf{x}) = rac{\exp(-d(f_{oldsymbol{\phi}}(\mathbf{x}), \mathbf{c}_k))}{\sum_{k'} \exp(-d(f_{oldsymbol{\phi}}(\mathbf{x}), \mathbf{c}_{k'}))}^{[22]}}$$

[22]

#### Zero- / One- / Few-Shot Learning

N-way k-shot learning task:

k labelled examples from each of N classes (=> not seen during!)

Classify disjoint batch of unlabelled examples into one of these N classes. Classification accuracy = #correct/#total

Zero-Shot Learning



from meta information:  

$$\mathbf{c}_{k} = g_{\boldsymbol{\vartheta}}(\mathbf{v}_{k})^{\text{IZZ}}$$

$$p_{\boldsymbol{\phi}}(y = k \,|\, \mathbf{x}) = \frac{\exp(-d(f_{\boldsymbol{\phi}}(\mathbf{x}), \mathbf{c}_{k}))}{\sum_{k'} \exp(-d(f_{\boldsymbol{\phi}}(\mathbf{x}), \mathbf{c}_{k'}))}^{\text{IZZ}}$$







DML

# Person RelD





# Self-Supervised Learning

Ever increasing amount of data  $\rightarrow$  use self-supervise, *i.e.*, unlabeled data for pre-training



 $\blacktriangleright$  define pre-text task for training  $\rightarrow$  get labels for free

• evaluation: freeze backbone, add linear classifier and train on downstream task

#### **Self-Supervised Learning - Contrastive Learning**



(f) Rotate  $\{90^\circ, 180^\circ, 270^\circ\}$ 

(g) Cutout

(h) Gaussian noise

(i) Gaussian blur

(j) Sobel filtering

Contrastive Learning: generate augmentations of the same sample

#### **Self-Supervised Learning - Contrastive Learning**



$$\ell_{i,j} = -\lograc{\exp(\mathrm{sim}(oldsymbol{z}_i,oldsymbol{z}_j)/ au)}{\sum_{k=1}^{2N}\mathbbm{1}_{[k
eq i]}\exp(\mathrm{sim}(oldsymbol{z}_i,oldsymbol{z}_k)/ au)}$$

Contrastive Learning: generate augmentations of the same sample Enforce their embeddings to be similar

[28] Chen, Ting et al. "A Simple Framework for Contrastive Learning of Visual Representations." (arxiv 2020)







DML







### **Geo Localization**



**Query Street View** 

Satellite



[29]

Retrieve gallery satellite image that shows same location as query street view

### **Geo Localization**



Retrieve gallery satellite image that shows same location as query street view

Train using retrieval loss, *e.g.*, triplet loss

[29] Toker, Aysim et al. "Coming Down to Earth: Satellite-to-Street View Synthesis for Geo-Localization." (CVPR, 2021)

#### References

[1] Goldberger, Jacob, Sam T. Roweis, Geoffrey E. Hinton and Ruslan Salakhutdinov. "Neighbourhood Components Analysis." NIPS (2004). [2] Movshovitz-Attias, Yair, Alexander Toshev, Thomas Leung, Sergev Joffe and Saurabh Singh, "No Fuss Distance Metric Learning Using Proxies," ICCV (2017) [3] Kim, Sungyeon, Dongwon Kim, Minsu Cho and Suha Kwak, "Proxy Anchor Loss for Deep Metric Learning," CVPR (2020) [4] Zhu, Yuehua, Muli Yang, Cheng Deng and Wei Liu. "Fewer is More: A Deep Graph Metric Learning Perspective Using Fewer Proxies." NeurIPS (2020) [5] https://en.wikipedia.org/wiki/Mutual\_information#/media/FileEntropy-mutual-information-relative-entropy-relation-diagram.svg 6] Teh, Eu Wern, Terrance Devries and Graham W. Taylor. "ProxyNCA++: Revisiting and Revitalizing Proxy Neighborhood Component Analysis." ECCV (2020) [17] Yu, Dingjun, Hanti Wang, Peiqiu Chen and Zhihua Wei. "Mixed Pooling for Convolutional Neural Networks." RSKT (2014).
 [8] <u>https://medium.com/abdhuma/which-pooling-method-is-better-maxpooling-vs-average-pooling-epfb03f45a9</u>
 [9] Elezi, Ismail, Seidenschwarz, Jenny, Wagner, Lauring, Vascon, Sebastiano, Torcingvich, Hessandro, Pellito, Marcello, Leal-1aike, Laura. "The Group Loss++: A deeper look into group loss for deep metric learning", PAMI (2022/03) [10] Seidenschwarz, Jenny, Ismail Elezi and Laura Leal-Taix'e. "Learning Intra-Batch Connections for Deep Metric Learning." ICML (2021). [11] Musgrave, Kevin, Serge J. Belongie and Ser-Nam Lim, "A Metric Learning Reality Check," ECCV (2020). 12] Roth, Karsten, Timo Milbich, Samarth Sinha, Prateek Gupta, Bjoern Ommer and Joseph Paul Cohen. "Revisiting Training Strategies and Generalization Performance in Deep Metric Learning." ICML (2020) 13] Kemelmacher-Shlizerman, Ira, Steven M. Seitz, Daniel Miller and Evan Brossard. "The MegaFace Benchmark: 1 Million Faces for Recognition at Scale." CVPR (2016) [14] Sun, Yifan, Liang Zheng, Yi Yang, Qi Tian and Shengjin Wang. "Beyond Part Models: Person Retrieval with Refined Part Pooling." ECCV (2018). 13] Chen, Tianlong, Shaojin Ding, Jingyi Xie, Ye Yuan, Wuyang Chen, Yang Yang, Zhou Ren and Zhangyang Wang, "ABD-Net: Attentive but Diverse Person Re-Identification." 2019 ICCV) (2019) [16] Tay, Chiat-Pin, Sharmili Roy and Kim-Hui Yap. "AANet: Attribute Attention Network for Person Re-Identifications." CVPR (2019) [17] Zhu, Zhihui, Xinyang Jiang, Feng Zheng, Xiao-Wei Guo, Feiyue Huang, Weishi Zheng and Xing Sun. "Viewpoint-Aware Loss with Angular Regularization for Person Re-Identification." AAAI (2020). [18] Luo, Haowen, Youzhi Gu, Xingyu Liao, Shengi Lai and Wei Jiang. "Bags of Tricks and A Strong Baseline for Deep Person Re-identification." (2019). 19] Vinyals, Oriol, Charles Blundell, Timothy P. Lillicrap, Koray Kavukcuoğlu and Daan Wierstra. "Matching Networks for One Shot Learning." NIPS (2016). [20] Chén, Ting, Simon Kornblith, Mohammad Norouzi and Géoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020): n. pag. [21] Matching networks for one shot learning. In NIPS, 2016. [22] Prototypical networks for few-shot learning. In NIPS, 2017. [23] Liu, W. et al. "SphereFace: Deep Hypersphere Embedding for Face Recognition." (CVPR, 2017) [24] Wang, H. et al. "CosFace: Large Margin Cosine Loss for Deep Face Recognition." (CVPR 2018) [25] Deng, J. et al. "ArcFace: Additive Angular Margin Loss for Deep Face Recognition." (CVPR 2019) [26] Li, W. et al. "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification" (CVPR 2014) 127] Leal-Taixé, L., Milan, A., Reid, I., Roth, S. & Schindler, K. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. arXiv:1504.01942 [cs], 2015., (arXiv: 1504.01942). [28] Chen. Ting et al. "A Simple Framework for Contrastive Learning of Visual Representations." (arxiv 2020) [29] Toker, Aysim et al. "Coming Down to Earth: Satellite-to-Street View Synthesis for Geo-Localization." (CVPR, 2021) (30) Jegou, H et al. "Product quantization for nearest neighbor search." (IPAMI 2011) [31] ]McDaid, A. et al. "Normalized mutual information to evaluate overlapping community finding algorithms." (arxiv 2011) [32] Wah, C. et al. "The Caltech-UCSD Birds-200-2011 Dataset." (Technical Report 2011) [33] Krause, J. et al. "3d object representations for fine-grained categorization." (Workshop on 3D Representation and Recognition, 2013.) [34] Song, H. et al. "Deep metric learning via lifted structured feature embedding." (CVPR 2016) [35] Liu, Ž. et al. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations" (CVPR 2016) [36] Zhai, A. and Wu, H. Classification is a strong baseline for deep metric learning. (BMVC 2019).