Modules for Improved Deep Learning-based Matching in Vision Tasks

Arulkumar S

under the guidance of Prof.Anurag Mittal

Computer Vision Lab, IIT Madras

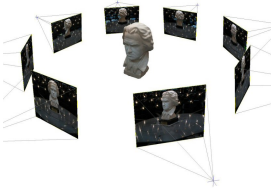
Synopsis meeting

Outline

Introduction
Why matching images/videos?

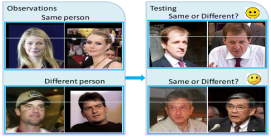
- Person re-identification

- 3D reconstruction



- Stereo(Depth) estimation
- Object tracking . . .

- Face verification

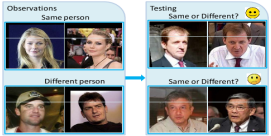


- Image stitching

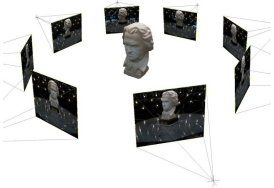
Introduction
Why matching images/videos?

- Person re-identification

- Face verification



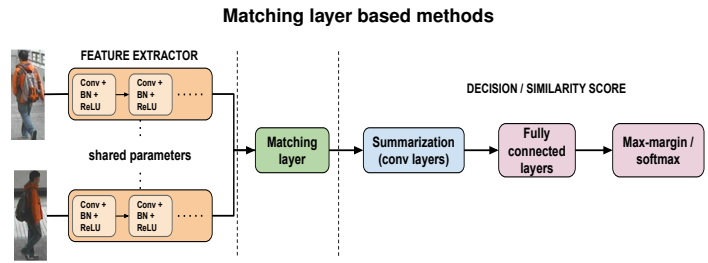
- 3D reconstruction



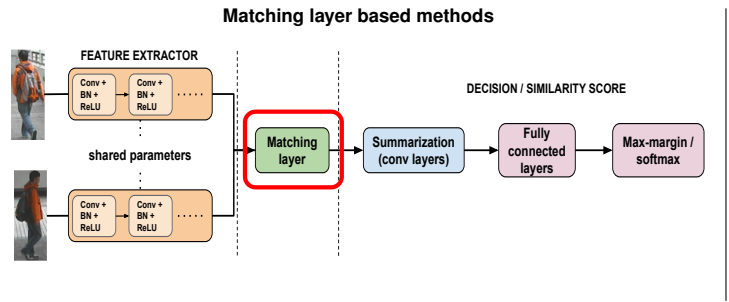
- Image stitching

- Stereo(Depth) estimation
- Object tracking . . .

Common pipelines

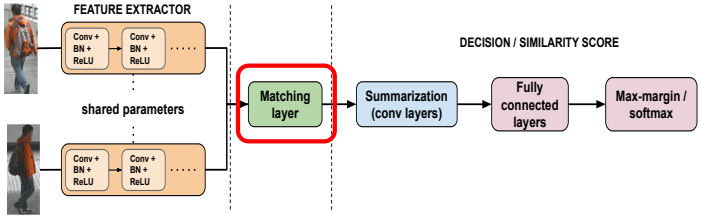


Common pipelines

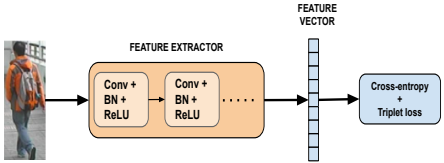


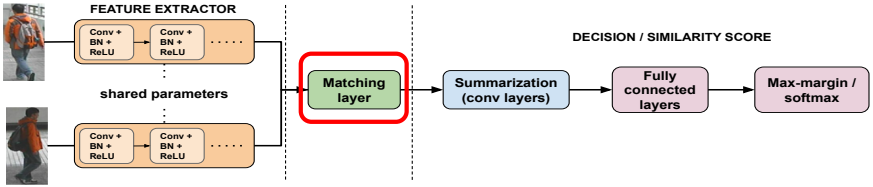
Common pipelines

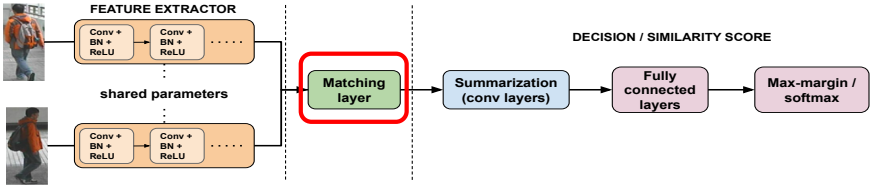
Matching layer based methods



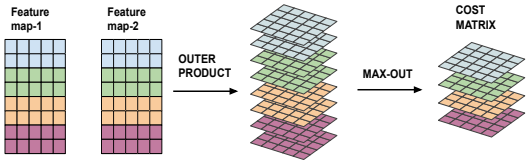
Global descriptor based methods



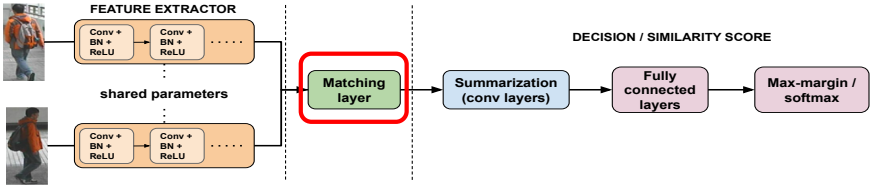

FEATURE EXTRACTOR
Conv + BN + ReLU
Conv + BN + ReLU
shared parameters
Matching layer
DECISION / SIMILARITY SCORE
Summarization (conv layers)
Fully connected layers
Max-margin / softmax



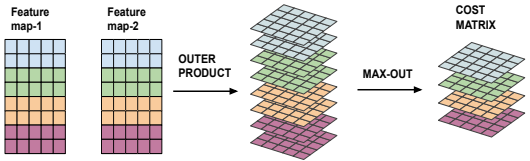
Outer-product based matching layer



Li *et al.* **DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification**. CVPR - 2014.

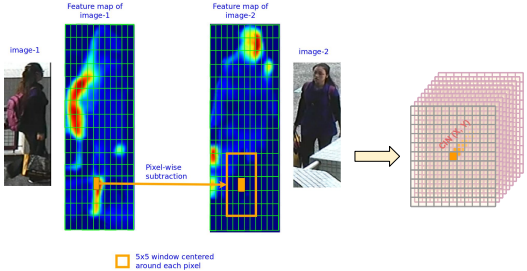


Outer-product based matching layer



Li *et al.* **DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification**. CVPR - 2014.

Cross-input neighborhood difference layer

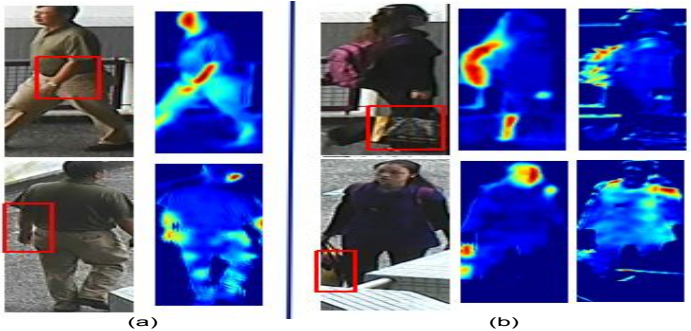


Ahmed *et al.* **An improved deep learning architecture for person re-identification**. CVPR - 2015.

Downsides in Ahmed *et. al* (CVPR-2015)

Notable drawbacks

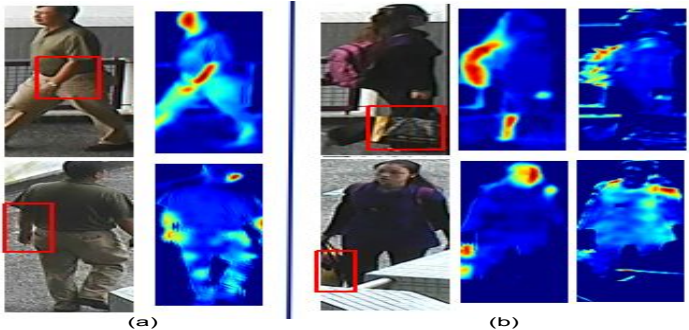
- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



Downsides in Ahmed *et. al* (CVPR-2015)

Notable drawbacks

- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions

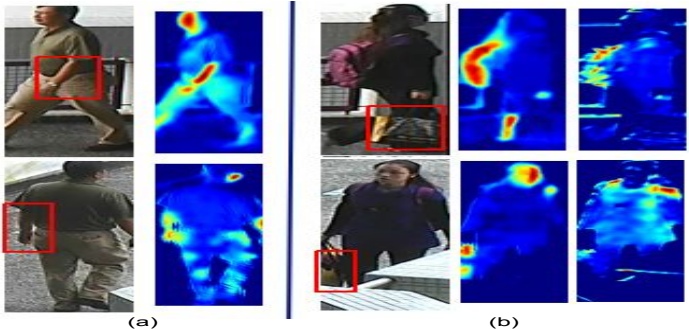


(Alternate) solution: increase the search space (horizontally up to full width)

Downsides in Ahmed *et. al* (CVPR-2015)

Notable drawbacks

- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



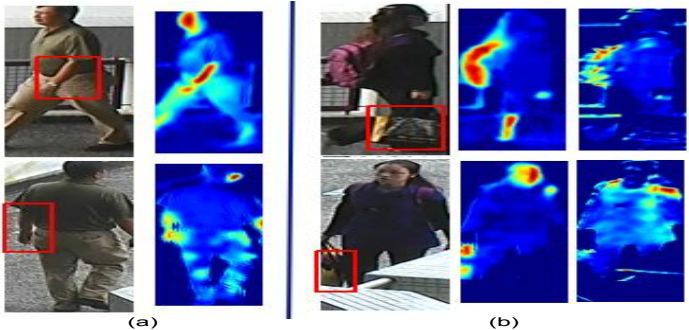
(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact & point-wise comparison of single pixel may be affected due to illumination variation

Downsides in Ahmed *et. al* (CVPR-2015)

Notable drawbacks

- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



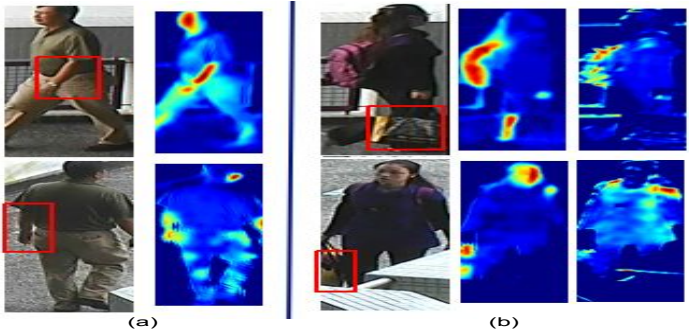
(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact & point-wise comparison of single pixel may be affected due to illumination variation **solution:**
  - Instead of single pixel difference, consider comparison of **patches**

Downsides in Ahmed *et. al* (CVPR-2015)

Notable drawbacks

- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



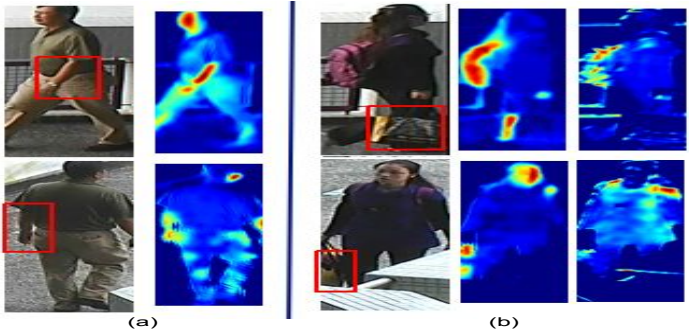
(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact & point-wise comparison of single pixel may be affected due to illumination variation **solution:**
  - Instead of single pixel difference, consider comparison of **patches**
    →Correlation between patches

Downsides in Ahmed *et. al* (CVPR-2015)

Notable drawbacks

- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



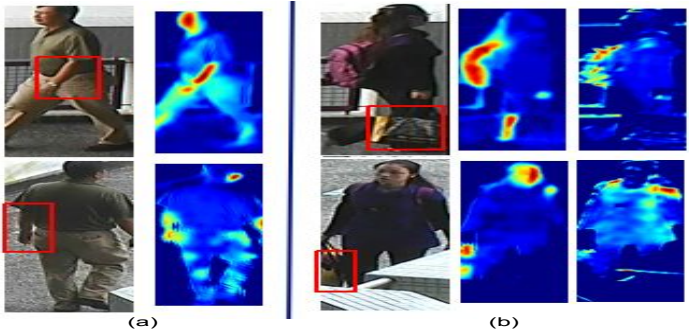
(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact & point-wise comparison of single pixel may be affected due to illumination variation
  solution:
  - Instead of single pixel difference, consider comparison of **patches**
    $\rightarrow$Correlation between patches
  - Normalize the patches with mean, standard deviation before comparison

Downsides in Ahmed *et. al* (CVPR-2015)

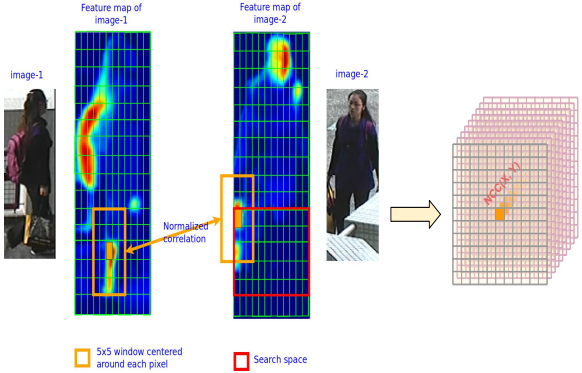
Notable drawbacks

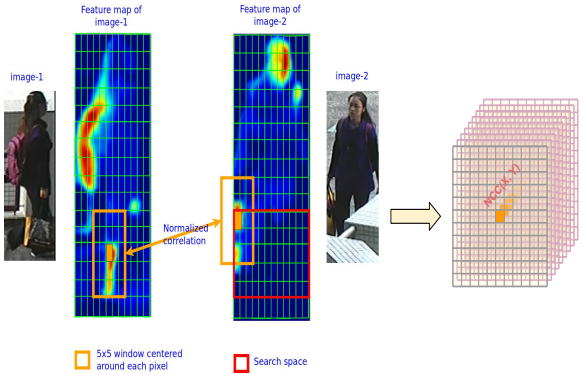
- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact & point-wise comparison of single pixel may be affected due to illumination variation **solution:**
  - Instead of single pixel difference, consider comparison of **patches**
    →Correlation between patches
  - Normalize the patches with mean, standard deviation before comparison
    → Normalized correlation


Feature map of image-1
Feature map of image-2
image-1
image-2
Normalized correlation
5x5 window centered around each pixel
Search space



$$NCC(E,F) = \frac{\sum_{i=1}^{N}(E_i - \mu_E) * (F_i - \mu_F)}{(N-1) * \sigma_E * \sigma_F}$$

$$\text{mean } \mu_E = \frac{\sum_{i=1}^{N} E_i}{N}$$

$$\text{unbiased standard deviation } \sigma_E = \sqrt{\frac{\sum_{i=1}^{N}(E_i - \mu_E)^2}{N-1}}$$

$$\frac{\partial NCC(E,F)}{\partial E_i} = \frac{1}{(N-1)\sigma_E}\left(\frac{F_i - \mu_F}{\sigma_F} - \frac{NCC(E,F) * (E_i - \mu_E)}{\sigma_E}\right)$$

NCC properties

Resilient to:
Additive transformation $I(x,y) = I(x,y) + \lambda$,
Multiplicative transformation $I(x,y) = \gamma I(x,y)$

Arulkumar Subramaniam, Moitreya Chatterjee, and Anurag Mittal. **Deep Neural Networks with Inexact Matching for Person Re-Identification**.
Proceedings of the Neural Information Processing Systems (NeurIPS) - 2016.

Image-based Person Re-identification

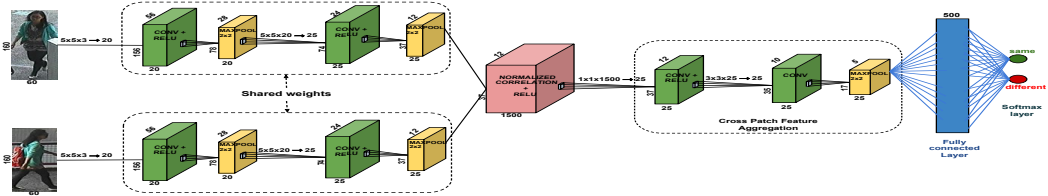
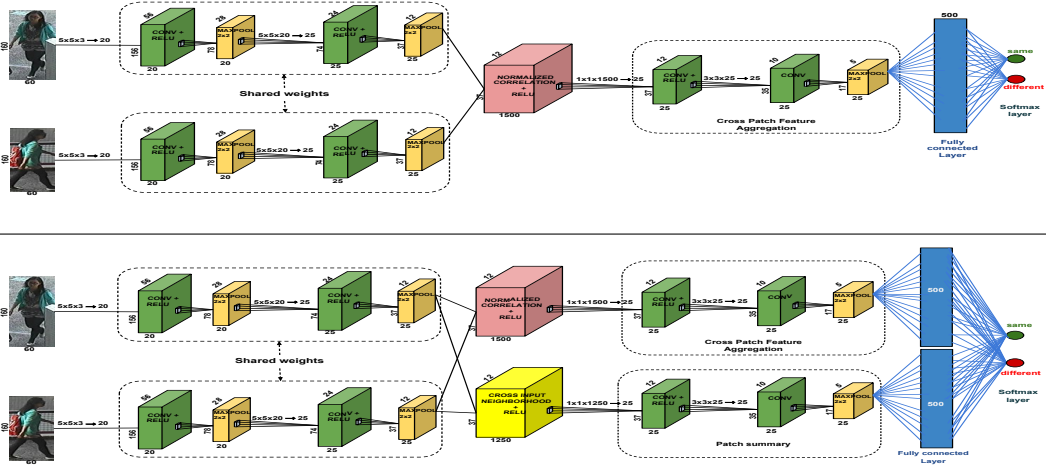


Image-based Person Re-identification



Arulkumar Subramaniam, Moitreya Chatterjee, and Anurag Mittal. **Deep Neural Networks with Inexact Matching for Person Re-Identification**. Proceedings of the Neural Information Processing Systems (NeurIPS) - 2016.

Method	#parameters
Ahmed et al. 2015's model	2.3M
NormXcorr model (ours)	1.12M
Fused model (ours)	2.22M

Table: Model complexity

Method	#parameters
Ahmed et al. 2015's model	2.3M
NormXcorr model (ours)	1.12M
Fused model (ours)	2.22M

Table: Model complexity

Method	r = 1	r = 10	r = 20
Fused Model (ours)	**72.43**	**95.51**	**98.40**
Norm X-Corr (ours)	64.73	92.77	96.78
Ensembles (Paisitkriangkrai et al. 2015)	62.1	92.30	97.20
LOMO+MLAPG (Liao and S. Z. Li 2015)	57.96	94.74	98.00
Ahmed et al. (Ahmed et al. 2015)	54.74	93.88	98.10
LOMO+XQDA (Liao et al. 2015)	52.20	92.14	96.25
Li et al. (W. Li et al. 2014)	20.65	68.74	83.06
KISSME (Köstinger et al. 2012)	14.17	52.57	70.03
LDML (Guillaumin et al. 2009)	13.51	52.13	70.81
eSDC (Zhao et al. 2013)	8.76	38.28	53.44

Table: CUHK03 Labeled Dataset

Method	r = 1	r = 10	r = 20
Fused Model (ours)	**72.04**	**96.00**	**98.26**
Norm X-Corr (ours)	67.13	94.49	97.66
LOMO+MLAPG (Liao and S. Z. Li 2015)	51.15	92.05	96.90
Ahmed et al. (Ahmed et al. 2015)	44.96	83.47	93.15
LOMO+XQDA (Liao et al. 2015)	46.25	88.55	94.25
Li et al. (W. Li et al. 2014)	19.89	64.79	81.14
KISSME (Köstinger et al. 2012)	11.70	48.08	64.86
LDML (Guillaumin et al. 2009)	10.92	47.01	65.00
eSDC (Zhao et al. 2013)	7.68	33.38	50.58

Table: CUHK03 Detected Dataset

Method	r = 1	r = 10	r = 20
Fused Model (ours)	**81.23**	**97.39**	**98.60**
Norm X-Corr (ours)	77.43	96.67	98.40
Ahmed et al. (Ahmed et al. 2015)	65.00	93.12	97.20
Li et al. (W. Li et al. 2014)	27.87	73.46	86.31
KISSME (Köstinger et al. 2012)	29.40	72.43	86.07
LDML (Guillaumin et al. 2009)	26.45	72.04	84.69
eSDC (Zhao et al. 2013)	22.84	57.67	69.84

Table: CUHK01 Dataset with 100 Test IDs

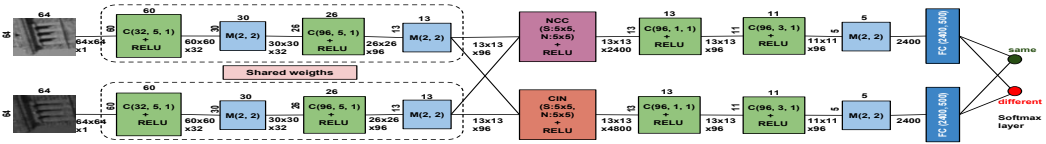
Method	r = 1	r = 10	r = 20
Fused Model (ours)	**65.04**	**89.76**	**94.49**
Norm X-Corr (ours)	60.17	86.26	91.47
CPDL (S. Li et al. 2015)	59.5	89.70	93.10
Ensembles (Paisitkriangkrai et al. 2015)	51.9	83.00	89.40
Ahmed et al. (Ahmed et al. 2015)	47.50	80.00	87.44
Mirror-KFMA (Y.-C. Chen et al. 2015)	40.40	75.3	84.10
Mid-Level Filters (Zhao et al. 2014)	34.30	65.00	74.90

Table: CUHK01 Dataset with 486 Test IDs

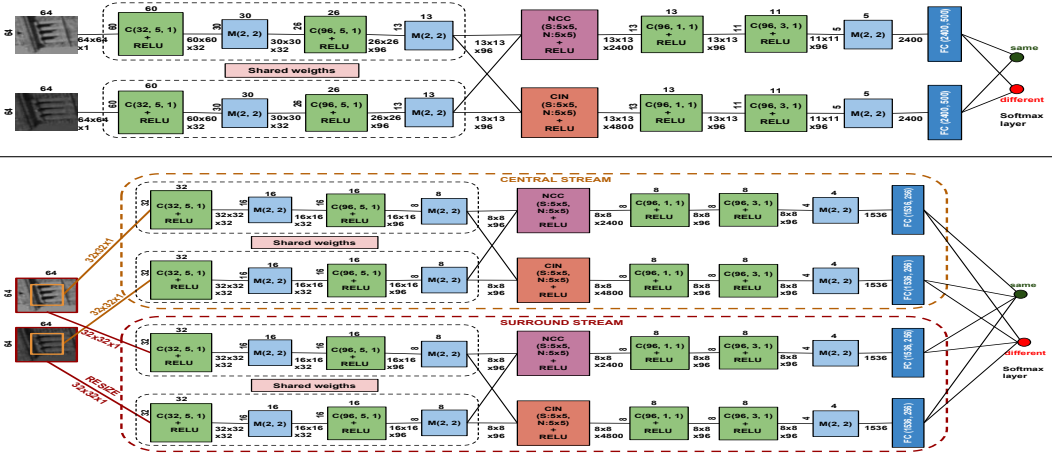
Method	r = 1	r = 5	r = 10	r = 20
Fused Model (ours)	**19.20**	38.40	**53.6**	**66.4**
Norm X-Corr (ours)	16.00	32.00	40.00	55.2
KEPLER (Martinel et al. 2015)	18.40	**39.12**	50.24	61.44
LOMO+XQDA (Liao et al. 2015)	16.56	33.84	41.84	52.40
PolyMap (D. Chen et al. 2015)	16.30	35.80	46.00	57.60
MtMCML (Ma et al. 2014)	14.08	34.64	45.84	59.84
MRank-RankSVM (Loy et al. 2013)	12.24	27.84	36.32	46.56
MRank-PRDC (Loy et al. 2013)	11.12	26.08	35.76	46.56
LCRML (J. Chen et al. 2014)	10.68	25.76	35.04	46.48
XQDA (Liao et al. 2015)	10.48	28.08	38.64	52.56

Table: QMUL GRID Dataset

Patch Matching



Patch Matching



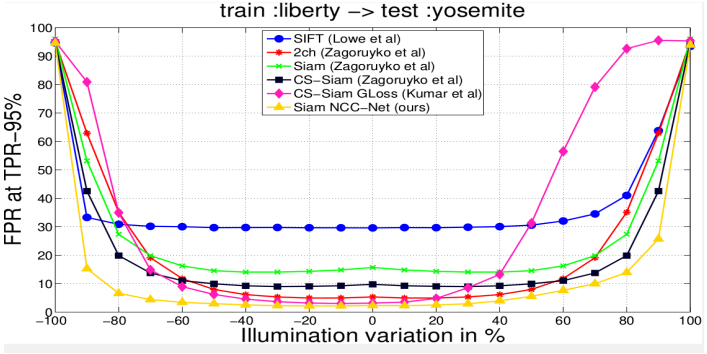
Arulkumar Subramaniam*, Prashanth Balasubramanian* and Anurag Mittal. **NCC-net: Normalized cross correlation based deep matcher with robustness to illumination variations**. IEEE Winter Conference on Applications of Computer Vision (WACV) - 2018.

Quantitative results
on UBC Patches dataset

Train dataset	Liberty		Notredame		Yosemite		mean
Test dataset	Notredame	Yosemite	Liberty	Yosemite	Liberty	Notredame	
Siam-NCC-Net(ours)	1.25	**2.03**	**3.87**	**1.86**	5.16	1.8	**2.66**
Siam-w/oMP$_2$-NCC-Net(ours)	**1.14**	2.30	4.02	2.34	**4.71**	1.81	2.72
CS-NCC-Net(ours)	1.24	3.09	5.99	4.22	6.54	2.06	3.86
CS-w/oMP$_2$-NCC-Net(ours)	1.17	**2.19**	4.28	**2.30**	**4.81**	**1.7**	2.74
2ch-CS stream GLoss	**0.77**	3.09	**3.69**	2.67	4.91	**1.14**	**2.71**
2ch-CS stream	1.9	4.75	4.55	4.1	7.2	2.11	4.10
Siamese GLoss	1.84	6.61	6.39	5.57	8.43	2.83	5.28
TFeat	3.12	7.82	7.22	7.08	9.79	3.85	6.48
PNNet	3.71	8.99	8.13	7.1	9.65	4.23	6.97
DeepCompare-Siam CS-stream	3.05	9.02	6.45	10.45	11.51	5.29	7.63
MatchNet	4.75	13.58	8.84	11.00	13.02	7.7	9.81
DeepCompare-Siam	4.33	14.89	8.77	13.23	13.48	5.75	10.07
VGG-Convex	7.52	11.63	12.88	10.54	14.82	7.11	10.75

Table: Testbed: *UBC Patches* dataset. **Color coding : red** - best performing method, **blue** - next best performing method (Training : 500K pairs, Testing: 100K pairs)

Experiments for illumination changes



Arulkumar Subramaniam*, Prashanth Balasubramanian* and Anurag Mittal. **NCC-net: Normalized cross correlation based deep matcher with robustness to illumination variations**. IEEE Winter Conference on Applications of Computer Vision (WACV) - 2018.

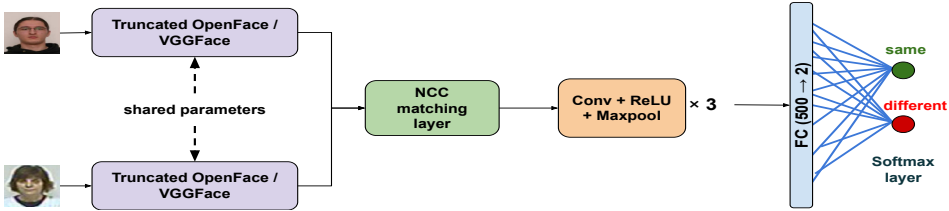
Tests on Natural intensity changes

Train dataset	**Siam-NCC-Net(ours)**	**Siam-w/oMP$_2$-NCC-Net(ours)**	**CS-NCC-Net(ours)**	**CS-w/oMP$_2$-NCC-Net(ours)**	2ch-CS-stream GLoss	2ch-CS stream	Siam	Siam-CS stream
L	**9.67**	**9.45**	11.37	11.35	12.31	12.31	31.45	27.08
N	**16.68**	**12.56**	23.04	19.30	20.01	17.84	28.21	32.17
Y	**11.67**	**10.56**	15.40	18.28	14.76	19.5	35.21	34.65

Table: **Color coding :** **red** - best performing method, **blue** - next best performing method.

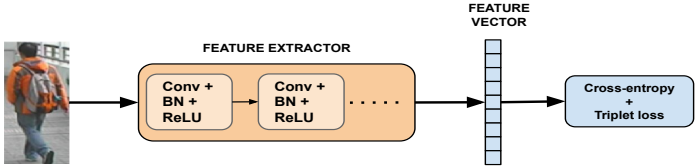
Arulkumar Subramaniam*, Prashanth Balasubramanian* and Anurag Mittal. **NCC-net: Normalized cross correlation based deep matcher with robustness to illumination variations**. IEEE Winter Conference on Applications of Computer Vision (WACV) - 2018.

Face verification

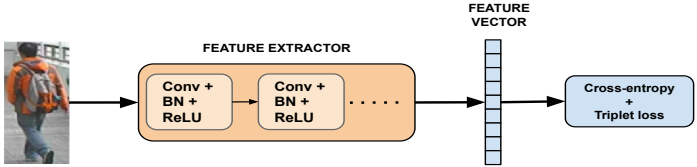


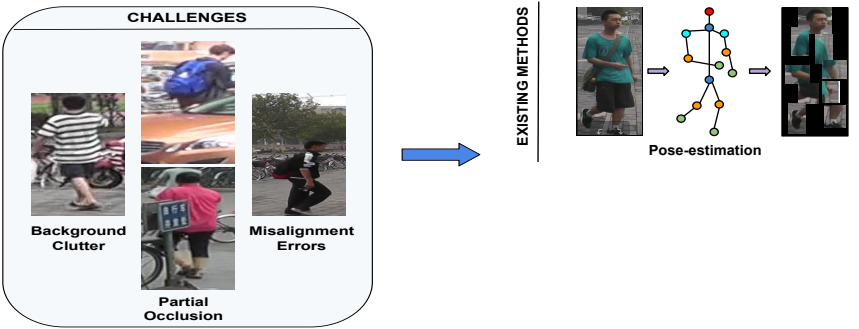
Method	Rank-1 (%)
VGGFace+NCC	82.75
OpenFace+NCC	80.5
VGGFace(FC7)	82.5
OpenFace	71.5
DictAlignment FR	73.25
LowResolution FR	69.45

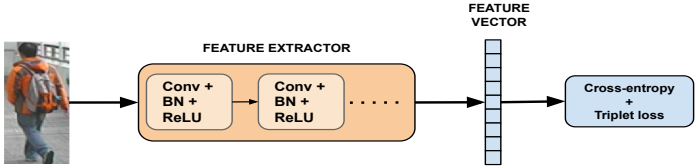
Arulkumar Subramaniam*, Prashanth Balasubramanian* and Anurag Mittal. **NCC-net: Normalized cross correlation based deep matcher with robustness to illumination variations**. IEEE Winter Conference on Applications of Computer Vision (WACV) - 2018.

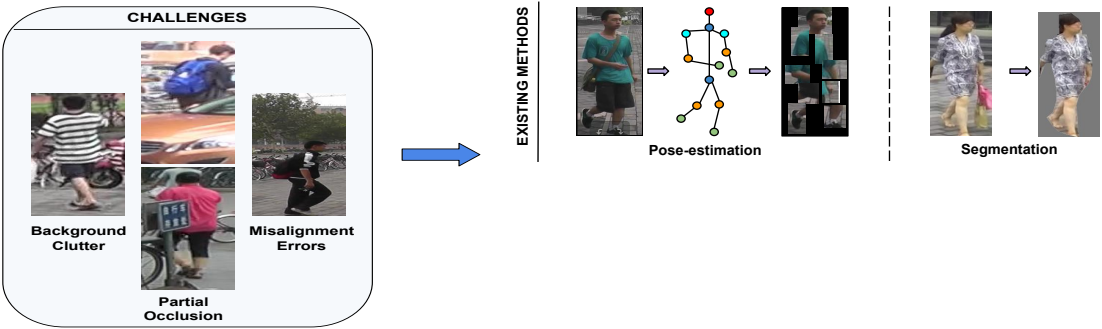

FEATURE VECTOR
FEATURE EXTRACTOR
Conv + BN + ReLU
Conv + BN + ReLU
. . . . .
Cross-entropy + Triplet loss

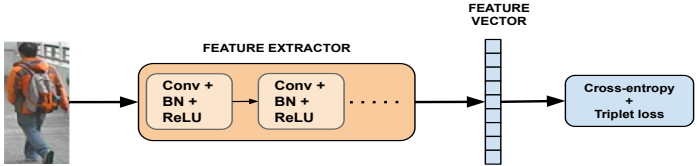

CHALLENGES
Background Clutter
Misalignment Errors
Partial Occlusion


FEATURE VECTOR
FEATURE EXTRACTOR
Conv + BN + ReLU
Conv + BN + ReLU
. . . . .
Cross-entropy + Triplet loss

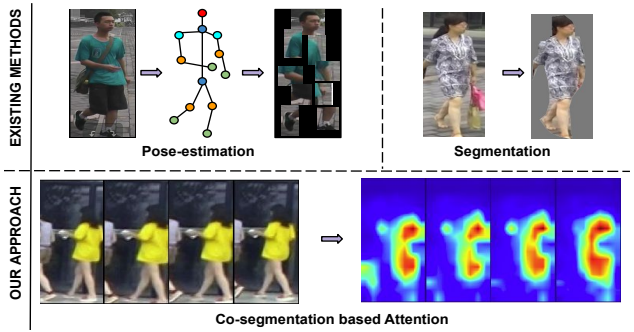

CHALLENGES
Background Clutter
Partial Occlusion
Misalignment Errors
EXISTING METHODS
Pose-estimation


FEATURE VECTOR
FEATURE EXTRACTOR
Conv + BN + ReLU
Conv + BN + ReLU
. . . . .
Cross-entropy + Triplet loss

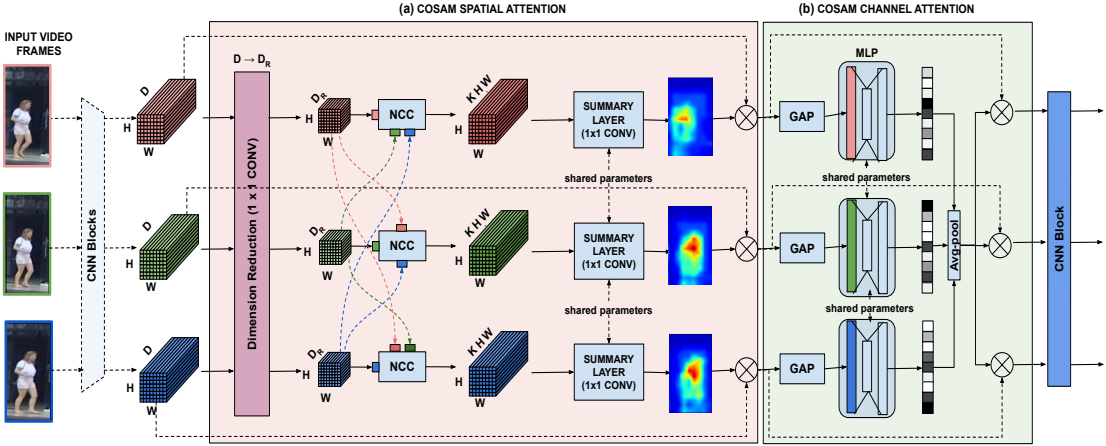

CHALLENGES
Background Clutter
Partial Occlusion
Misalignment Errors
EXISTING METHODS
Pose-estimation
Segmentation


FEATURE VECTOR
FEATURE EXTRACTOR
Conv + BN + ReLU
Conv + BN + ReLU
Cross-entropy + Triplet loss

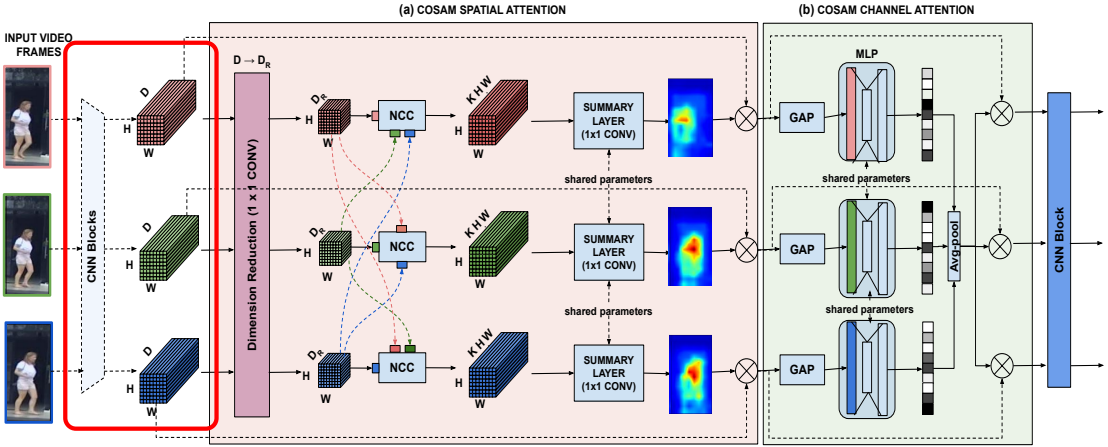

CHALLENGES
Background Clutter
Partial Occlusion
Misalignment Errors


EXISTING METHODS
Pose-estimation
Segmentation
OUR APPROACH
Co-segmentation based Attention

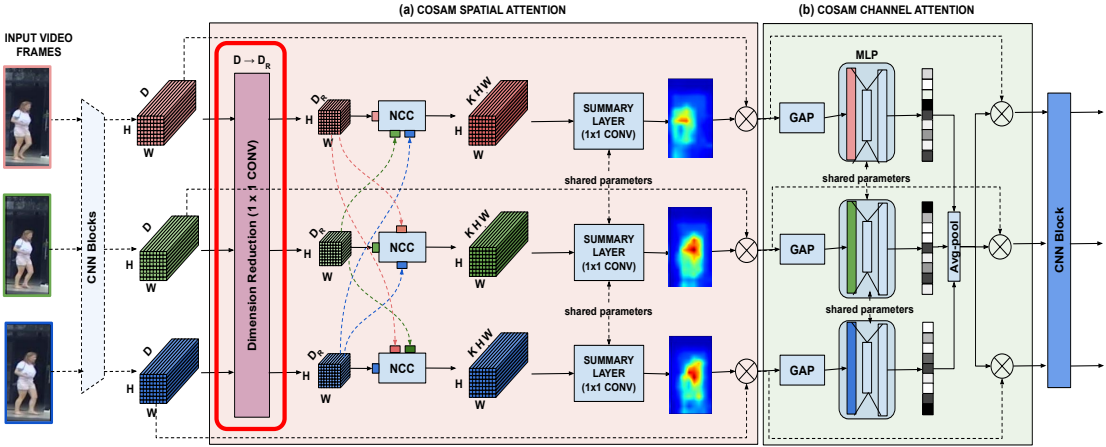
Co-segmentation Activation Module (COSAM)



Input $(N \times D \times H \times W) \rightarrow$ Induce co-segmentation $\rightarrow$ Output $(N \times D \times H \times W)$

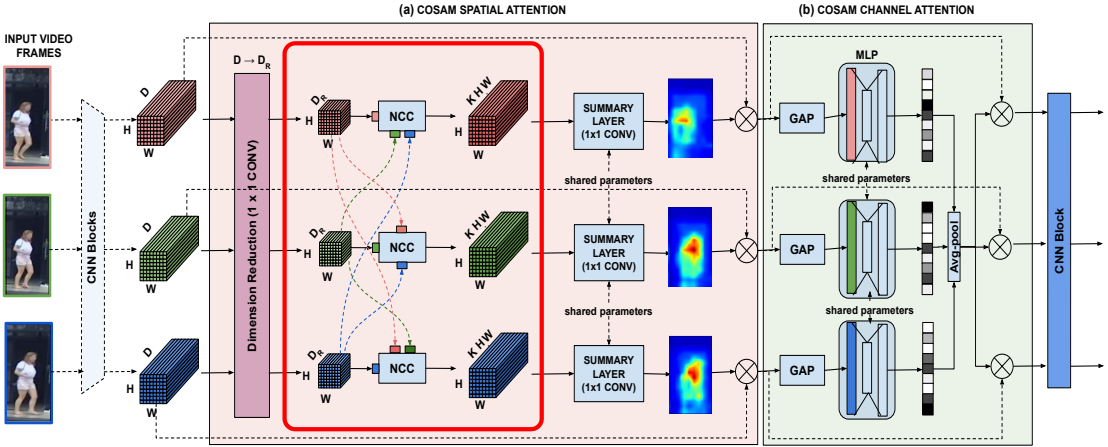


Frames of dimension $N \times 3 \times H_I \times W_I$ are passed through L CNN blocks to get feature maps of dimension $N \times D \times H \times W$.



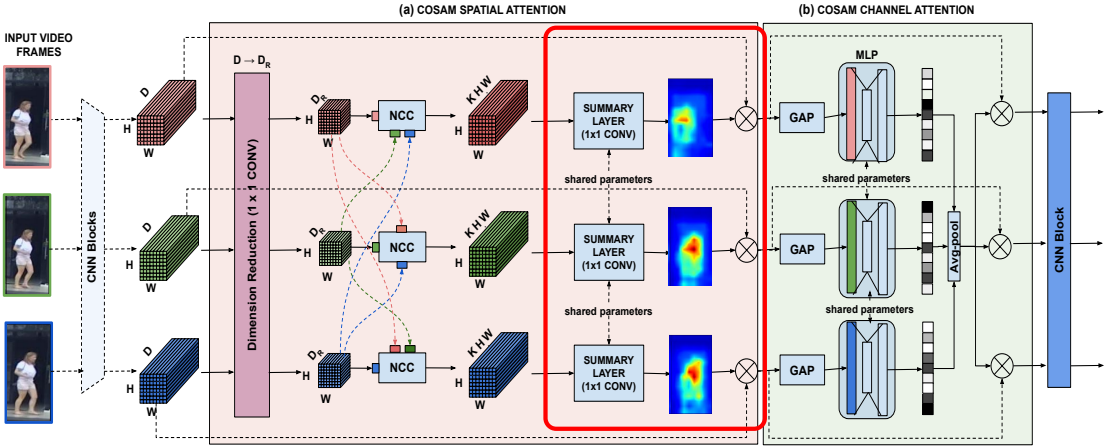
Dimensionality reduction ($D \times H \times W \longrightarrow D_R \times H \times W$)

$D \longrightarrow D_R$ - to reduce computational overhead

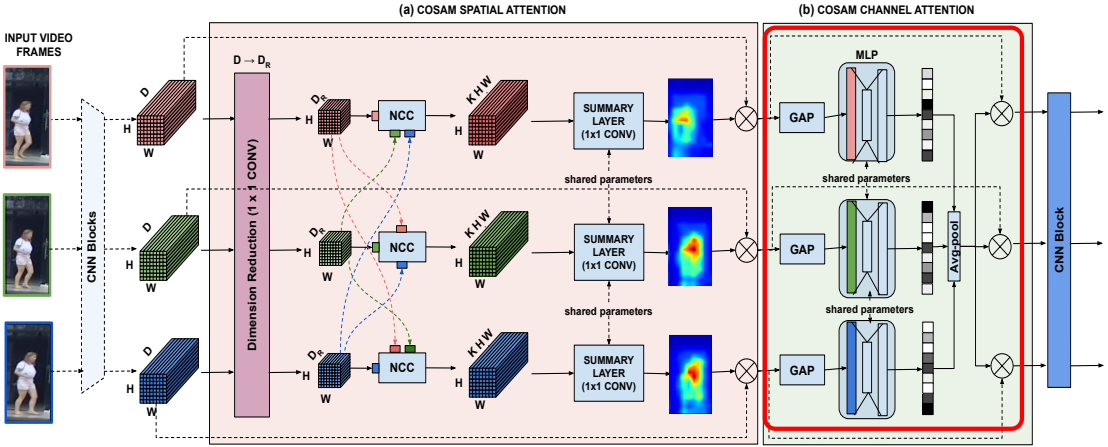


$$\text{Cost volume}_{(n)}(i,j) = \{NCC\left(F_n^{(i,j)}, R_k^{(h,w)}\right) \quad 1 \le k \le K, 1 \le h \le H, 1 \le w \le W\} \tag{1}$$

$$\text{NCC}(P,Q) = \frac{1}{D_R} \frac{\sum_{k=1}^{D_R}(P_k - \mu_P).(Q_k - \mu_Q)}{\sigma_P.\sigma_Q} \tag{2}$$

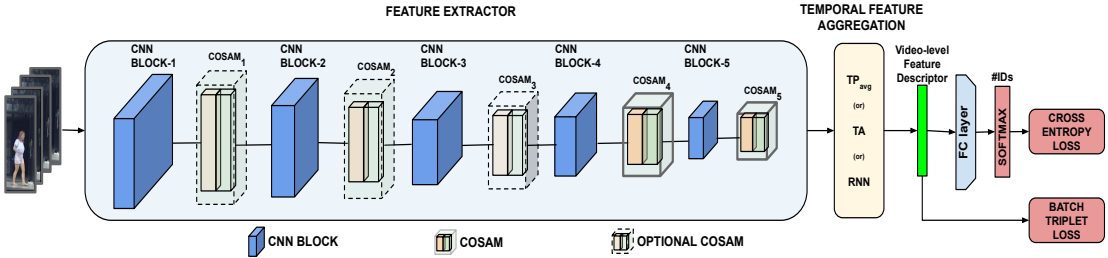


- Pass cost volume through Conv + BN + ReLU $\rightarrow$ Sigmoid to get spatial mask.
- Multiply spatial masks with corresponding feature maps



- Per-frame Channel attention from Global Average Pool-ed (GAP) feature maps
- Average of per-frame channel attentions to capture common important channels

Video-based Supervised Person Re-ID

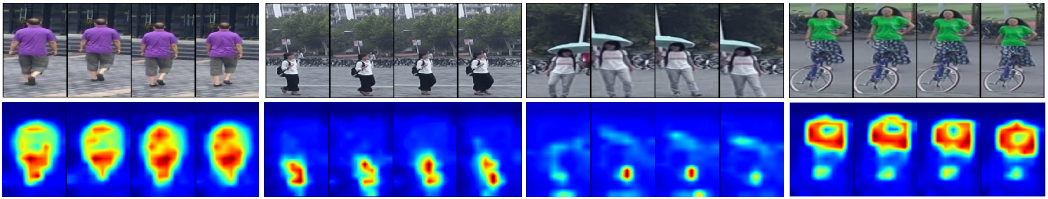


Training loss function:

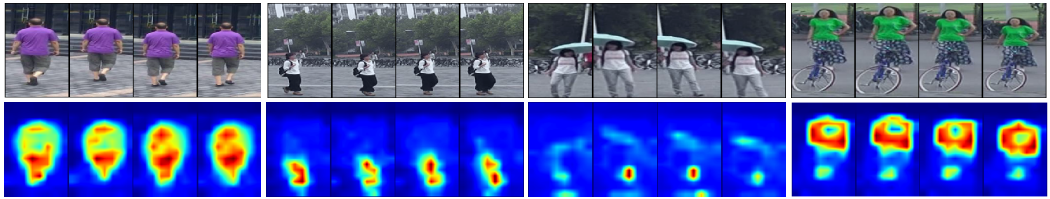
$$L = \sum_{i=1}^{B} \left\{ L_{CE} + \lambda L_{triplet}(I_i, I_{i_+}, I_{i_-}) \right\} \tag{3}$$

Arulkumar Subramaniam, Athira Nambiar and Anurag Mittal. **Co-segmentation Inspired Attention Networks for Video-based Person Re-identification**. IEEE International Conference on Computer Vision (ICCV) - 2019.

Qualitative visualization



Qualitative visualization



Comparison with State-of-the-arts

Network	Deep model?	MARS			
		mAP	R1	R5	R20
TriNet	Yes	67.7	79.8	91.4	-
Region QEN	Yes	71.1	77.8	88.8	94.1
Comp. Snippet Sim.	Yes	69.4	81.2	92.1	-
Part-Aligned	Yes	72.2	83.0	92.8	96.8
RevisitTempPool	Yes	76.7	83.3	93.8	97.4
SE-ResNet50 + TP_{avg}	Yes	78.1	84.0	95.2	97.1
SE-ResNet50 + $COSAM_{4,5}$ + TP_{avg}(ours)	Yes	**79.9**	**84.9**	**95.5**	**97.9**
SE-ResNet50 + $COSAM_{4,5}$ + TP_{avg}(ours) + Re-ranking	Yes	**87.4**	**86.9**	**95.5**	**98.0**

Network	Deep model?	DukeMTMC-VideoReID			
		mAP	R1	R5	R20
ETAP-Net	Yes	78.34	83.62	94.59	97.58
RevisitTempPool	Yes	93.2	93.9	98.9	99.5
SE-ResNet50 + TP_{avg}	Yes	93.5	93.7	99.0	99.7
SE-ResNet50 + $COSAM_{4,5}$ + TP_{avg}(ours)	Yes	**94.1**	**95.4**	**99.3**	**99.8**

Comparison with State-of-the-arts

Network	Deep model?	MARS			
		mAP	R1	R5	R20
TriNet	Yes	67.7	79.8	91.4	-
Region QEN	Yes	71.1	77.8	88.8	94.1
Comp. Snippet Sim.	Yes	69.4	81.2	92.1	-
Part-Aligned	Yes	72.2	83.0	92.8	96.8
RevisitTempPool	Yes	76.7	83.3	93.8	97.4
SE-ResNet50 + TP_{avg}	Yes	78.1	84.0	95.2	97.1
SE-ResNet50 + $COSAM_{4,5}$ + TP_{avg}(ours)	Yes	**79.9**	**84.9**	**95.5**	**97.9**
SE-ResNet50 + $COSAM_{4,5}$ + TP_{avg}(ours) + Re-ranking	Yes	**87.4**	**86.9**	**95.5**	**98.0**

Network	Deep model?	DukeMTMC-VideoReID			
		mAP	R1	R5	R20
ETAP-Net	Yes	78.34	83.62	94.59	97.58
RevisitTempPool	Yes	93.2	93.9	98.9	99.5
SE-ResNet50 + TP_{avg}	Yes	93.5	93.7	99.0	99.7
SE-ResNet50 + $COSAM_{4,5}$ + TP_{avg}(ours)	Yes	**94.1**	**95.4**	**99.3**	**99.8**

Model	Handbag			Hat			Backpack		
	mAP	R1	R5	mAP	R1	R5	mAP	R1	R5
ResNet50+TP	91.2	92.0	**100.0**	91.1	91.7	97.5	92.8	93.9	98.6
ResNet50+$COSAM_{4,5}$+TP	**95.2**	**96.0**	**100.0**	**93.5**	**94.2**	**97.5**	**95.1**	**96.4**	**99.8**
SE-ResNet50+TP	94.1	97.3	**100.0**	92.7	94.2	99.2	94.3	95.6	99.1
SE-ResNet50+$COSAM_{4,5}$+TP	**96.0**	**100.0**	**100.0**	**93.9**	**96.7**	**99.5**	**95.4**	**97.1**	**100.0**

Table: Attribute-wise performance comparison on Duke dataset. TP = Temporal average pooling.

Video-based Self-supervised Person Re-ID

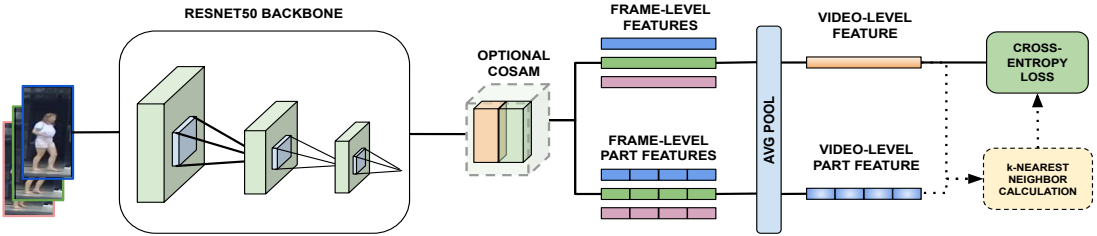


Figure: Baseline architecture from Lin et al. 2020. **Unsupervised person re-identification via softened similarity learning**, CVPR-2020.

Video-based Self-supervised Person Re-ID

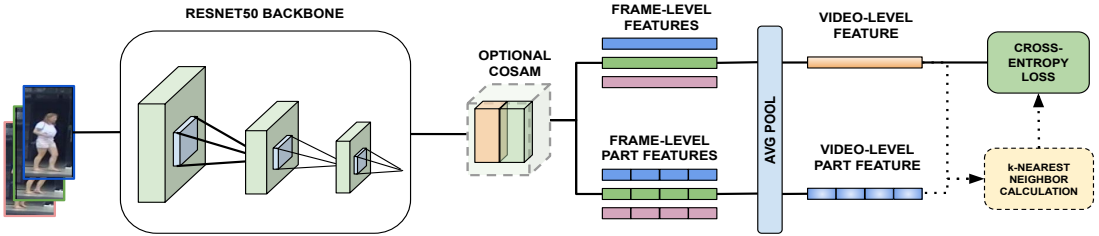
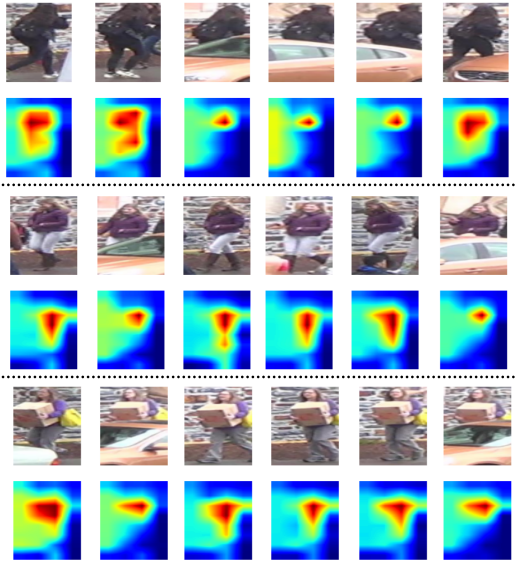


Figure: Baseline architecture from Lin et al. 2020. **Unsupervised person re-identification via softened similarity learning**, CVPR-2020.

Method	Setting	MARS				DukeMTMC-VideoReID			
		mAP	R1	R5	R10	mAP	R1	R5	R10
DGM+IDE (Ye et al. 2017)	OneEx	16.8	36.8	54.0	-	33.6	42.3	57.9	69.3
Stepwise (Liu et al. 2017)	OneEx	19.6	41.2	55.5	-	46.7	56.2	70.3	79.2
RACE (Ye et al. 2018)	OneEx	24.5	43.2	57.1	62.1	-	-	-	-
EUG (Wu et al. 2018)	OneEx	42.4	62.6	74.9	-	63.2	72.7	84.1	-
OIM (Xiao et al. 2017)	Unsup	13.5	33.7	48.1	54.8	43.8	51.1	70.5	76.2
DAL (Y. Chen et al. 2018)	Unsup	23.0	49.3	65.9	72.2	-	-	-	-
BUC (Lin et al. 2019)	Unsup	34.7	57.9	72.3	75.9	68.3	76.2	88.3	91.0
SoftSimLearn (Lin et al. 2020)	Unsup	43.6	62.8	77.2	80.1	69.3	76.4	88.7	91.0
SoftSimLearn + COSAM (ours)	Unsup	**44.2**	**63.8**	**78.4**	**82.0**	**72.2**	**80.2**	**91.7**	**94.0**



Video Action Recognition

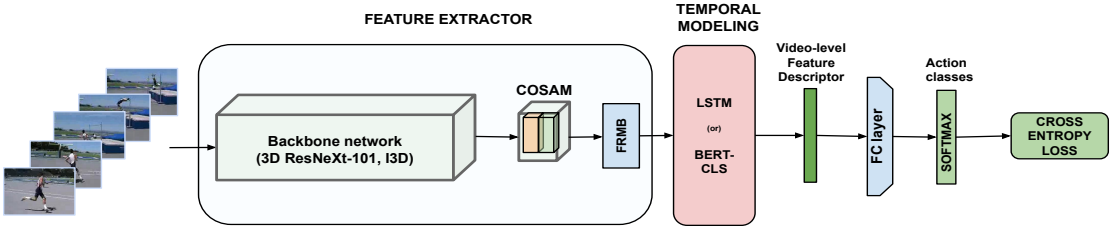


Figure: Baseline architecture from Kalfaoglu et al. 2020. **Late temporal modeling in 3d cnn architectures with bert for action recognition**, ECCV-2020.

Video Action Recognition

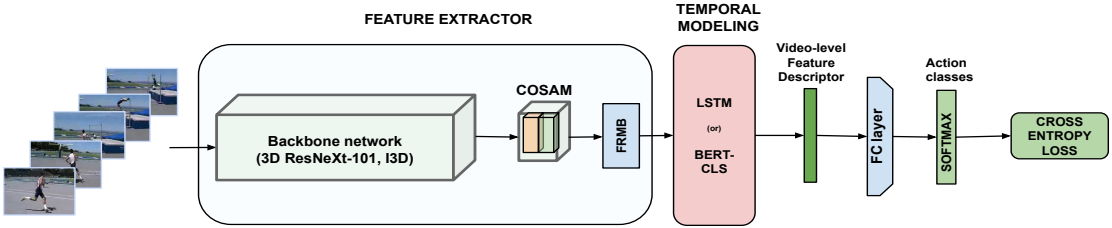
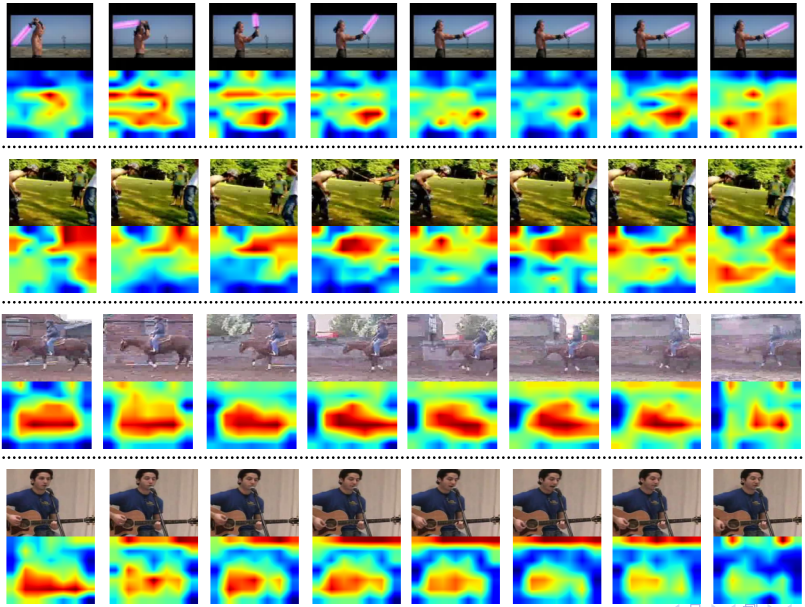


Figure: Baseline architecture from Kalfaoglu et al. 2020. **Late temporal modeling in 3d cnn architectures with bert for action recognition**, ECCV-2020.

Backbone	COSAM?	temp. model?	#params (M)	#Flops (G)	HMDB51		UCF101	
					Top-1	Top-3	Top-1	Top-3
ResNeXt101 (baseline)	✗	LSTM	47.6	38.64	73.68	87.46	93.90	98.05
ResNeXt101	✓	LSTM	48.41	38.77	**75.16**	**89.22**	**94.59**	**98.52**
ResNeXt101 (baseline)	✗	BERT	47.4	38.37	76.08	90.46	95.50	98.23
ResNeXt101	✓	BERT	48.21	38.49	**77.52**	**92.55**	**95.96**	**98.84**
I3D (baseline)	✗	BERT	13.57	110.6	68.63	87.78	92.50	98.26
I3D	✓	BERT	14.23	110.7	69.38	87.95	93.05	98.63

Table: Performance comparison with the baseline Kalfaoglu et al. 2020

Arulkumar Subramaniam, Jayesh Vaidya, Muhammed Abdul Majeed Ameen, Athira Nambiar, and Anurag Mittal. **Co-segmentation Inspired Attention Module for Video-based Computer Vision Tasks**. Computer Vision and Image Understanding (CVIU) - 2021 (Submitted).

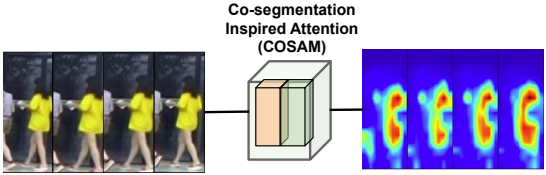


Three applications

- Image-based Person Re-ID
- Patch matching
- Face verification

Subramaniam *et al.* **Deep Neural Networks with Inexact matching for Person Re-Identification**. NeurIPS - 2016.

Subramaniam* *et al.* **NCC-Net: Normalized Cross Correlation Based Deep Matcher with Robustness to Illumination Variations**. WACV - 2018.



Three applications

- Video-based Supervised Person Re-ID
- Video-based Self-supervised Person Re-ID
- Video Action Recognition

Subramaniam *et al.* **Co-segmentation Inspired Attention Networks for Video-based Person Re-identification**. ICCV - 2019.

Subramaniam *et al.* **Co-segmentation Inspired Attention Module for Video-based Computer Vision Tasks**. CVIU (Submitted).

Journal Articles

- Arulkumar Subramaniam, Jayesh Vaidya, Muhammed Abdul Majeed Ameen, Athira Nambiar, and Anurag Mittal. **Co-segmentation Inspired Attention Module for Video-based Computer Vision Tasks**. Computer Vision and Image Understanding (CVIU), 2021. (Submitted)

Conference proceedings

- Arulkumar Subramaniam, Athira Nambiar, and Anurag Mittal. **Co-segmentation Inspired Attention Networks for Video-based Person Re-identification**. Proceedings of the International Conference on Computer Vision (ICCV) - 2019. Seoul, South Korea.
- Arulkumar Subramaniam*, Prashanth Balasubramanian*, and Anurag Mittal. **NCC-Net: Normalized Cross Correlation Based Deep Matcher with Robustness to Illumination Variations**. IEEE Winter Conference on the Applications of Computer Vision (WACV) - 2018. Nevada, United States.
- Arulkumar Subramaniam, Moitreya Chatterjee, and Anurag Mittal. **Deep Neural Networks with Inexact Matching for Person Re-Identification**. Proceedings of the Neural Information Processing Systems (NeurIPS) - 2016. Barcelona, Spain.

- Jayesh Vaidya, Arulkumar Subramaniam, and Anurag Mittal. **Co-Segmentation Aided Two-Stream Architecture for Video Captioning**. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022, Hawaii.
- Arulkumar Subramaniam*, Ajay Narayanan*, and Anurag Mittal. **Feature Ensemble Networks with Re-ranking for Recognizing Disguised Faces in the Wild**. Proceedings of the International Conference on Computer Vision Workshop (ICCVW) - 2019 on Recognizing Disguised Faces in the Wild.
- Arulkumar Subramaniam*, Vismay Patel*, Ashish Mishra, Prashanth Balasubramanian, and Anurag Mittal. **Bi-modal First Impressions Recognition using Temporally Ordered Deep Audio and Stochastic Visual Features**. Proceedings of the European Conference on Computer Vision Workshop (ECCVW) - 2016 on Apparent Personality Analysis. Amsterdam, The Netherlands.
- Arulkumar Subramaniam, Ashish Vaswani, and Niki Parmar. **Self-Attention based Feature Extractors for 3D Object Detection in Point Clouds**. In *European Conference on Computer Vision (ECCV) - 2020 Workshop on Perception for Autonomous Driving*.
- Ashish Mishra, Vinay Kumar Verma, M Reddy, Arulkumar Subramaniam, Piyush Rai, and Anurag Mittal. **A generative approach to zero-shot and few-shot action recognition**. In *Winter Conference on Applications of Computer Vision (WACV), 2018*.

Thank you!