Normalized Correlation based Deep Matching Networks for Learning Similarity

Arulkumar S

under the guidance of Dr.Anurag Mittal

Computer Vision Lab, IIT Madras

Seminar-I

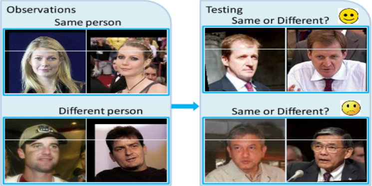
Outline

Outline

Introduction

Why matching images/patches?

- Face verification



- Image stitching

- 3D reconstruction

- Stereo(Depth) estimation

- Object tracking . . .

Outline

Common theme

- **Core task: Matching two images / image patches(keypoints)**

Common theme

- **Core task: Matching two images / image patches(keypoints)**
- Significant attention on matching images than learning the characteristics of individual classes

Common theme

- **Core task: Matching two images / image patches(keypoints)**
- Significant attention on matching images than learning the characteristics of individual classes
- Less number of examples per classes

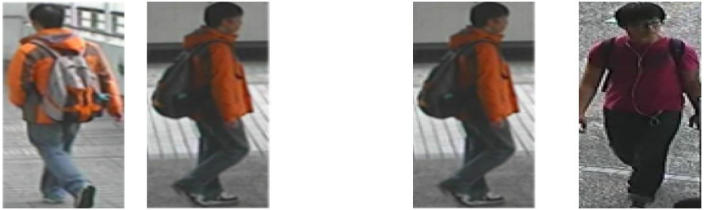
Common theme

- **Core task: Matching two images / image patches(keypoints)**
- Significant attention on matching images than learning the characteristics of individual classes
- Less number of examples per classes
- Challenges faced in matching images
  - illumination changes
  - Pose/view point changes
  - Occlusion
  - Deformation of objects
  - Motion blur (Tracking)
  - ...

Two challenging tasks to explore

- Person re-identification

same | different



- Patch matching

Same | Different

Outline

Person Re-Identification
Problem definition

- task of matching a person's image with images from database
- images captured at
  - same/different points in time (of same day)
  - same/different camera
  - various lighting conditions + unconstrained viewpoint/pose changes
- No information about camera position, intrinsic and extrinsic parameters

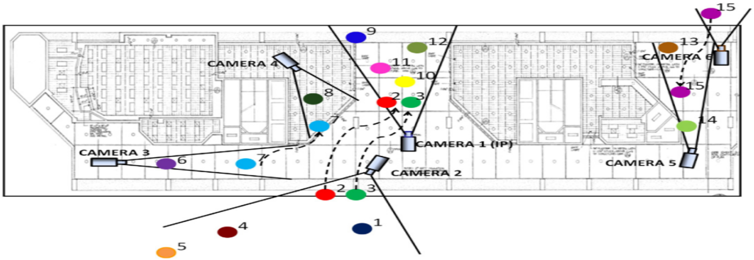
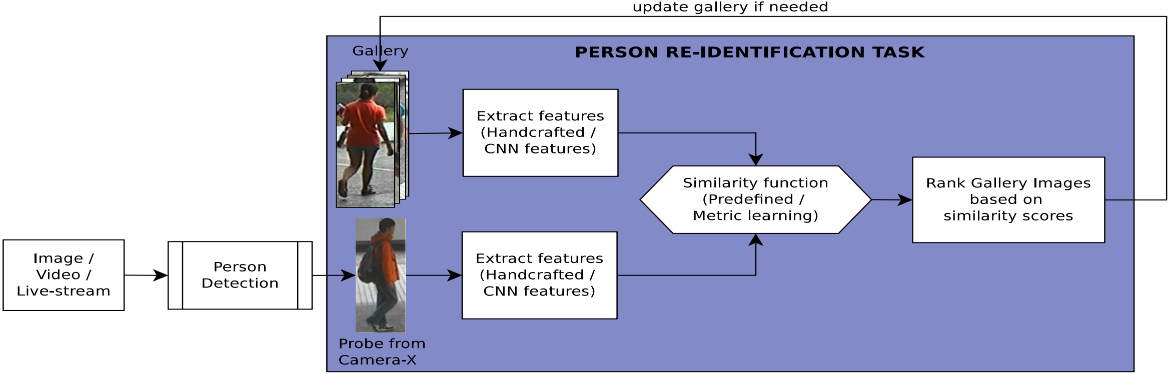


Image from: Apurva *et al.* **A survey of approaches and trends in person re-identification**. Image and Vision Computing - 2014.

Outline

Person Re-Identification setup

- **Probe:** the person's image to be searched in the database
- **Gallery:** one (or more) unique image(s) of persons observed so far. Usually, Gallery images will be available in a database.



- **Evaluation:** Ranking of matching scores (rank-1, rank-5, . . .)

Practical challenges

- Illumination variation
- Pose/Viewpoint variation
- Background clutters / misalignment errors
- Partial occlusion
- Bad quality images

Viewpoint Change

Illumination Variation

Partial Occlusion

Poor quality of images

Available datasets

Name	# of identities	# of images/identity	Train/Test split
ViPeR[1]	632	2	316 / 316
CUHK01[2]	971	4	871 / 100
CUHK03 (Labeled / Detected)[3]	1360	10	1260 / 100
QMULGRiD[4]	250	2	125/(125 + 775 unmatched gallery images)

Figure: Datasets relevant to this presentation

Gallery, Probe selection during Test:

A random image from each of the test identify is chosen for gallery set and the rest are considered as probe images.

[1] Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features, ECCV-2008

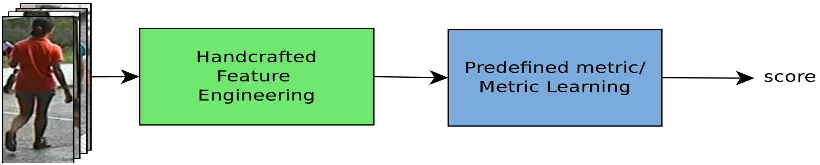
[2] Human Reidentification with Transferred Metric Learning, ACCV-2012

[3] DeepReID: Deep Filter Pairing Neural Network for Person Re-identification, CVPR-2014

[4] On-the-fly feature importance mining for person re-identification, PR-2014

Outline

Non-deep learning approaches



- Handcrafted features
  - Color histograms (RGB, HSV, YCbCr . . .)
  - Texture histograms (Schmidt, Gabor filtors, HOG, LBP, . . .)
  - Dense SIFT features
- Predefined metric / Metric learning
  - L2, Mahalanobis distance measures
  - Fisher discriminant analysis
  - Large-margin nearest neighbor (LMNN)
  - RankSVM
  - . . .

Outline

Deep learning
Why?

Pros:

- End-to-End training
  - Robust feature extraction
  - Non-linear feature space
  - Metrics/Classifier and Features are optimized together

[5] DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR 2014
[6] FaceNet: A Unified Embedding for Face Recognition and Clustering, CVPR 2015

Deep learning
Why?

Pros:

- End-to-End training
  - Robust feature extraction
  - Non-linear feature space
  - Metrics/Classifier and Features are optimized together

Cons:

- Needs large amount of data
- Enormous computation power (thanks to GPUs)

[5] DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR 2014
[6] FaceNet: A Unified Embedding for Face Recognition and Clustering, CVPR 2015

Deep learning
Why?

Pros:

- End-to-End training
  - Robust feature extraction
  - Non-linear feature space
  - Metrics/Classifier and Features are optimized together

Cons:

- Needs large amount of data
- Enormous computation power (thanks to GPUs)

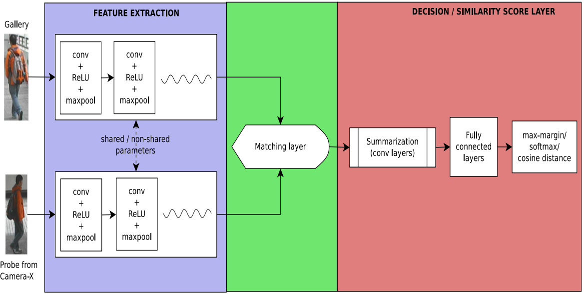
Prevalent architectures:

- Siamese networks[5]
- Triplet networks[6]

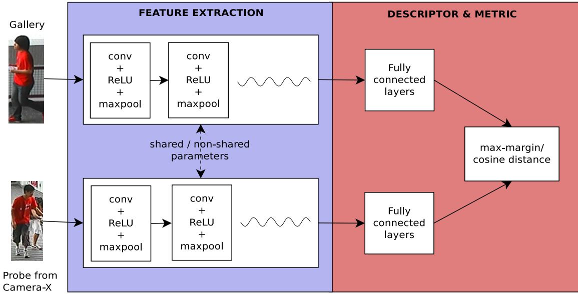
[5] DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR 2014
[6] FaceNet: A Unified Embedding for Face Recognition and Clustering, CVPR 2015

Siamese networks

Matching layer based architectures



Descriptor based architectures



Matching layers in literature

- Filter pairing and dot product layer (Liu et al, CVPR 2014)
- Cross Input neighborhood layer (Ahmed et. al, CVPR 2015) (in next slides)
- Normalized Cross Correlation Layer (Subramaniam et al, NIPS 2016)

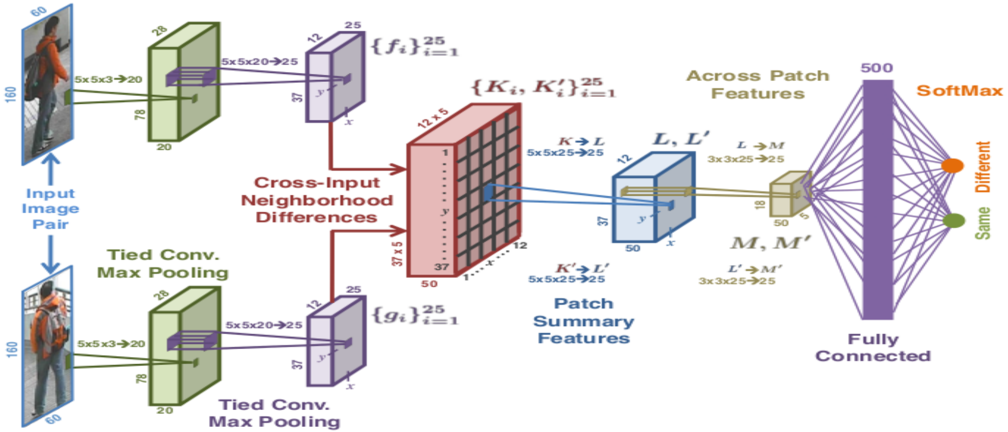
Metrics for Descriptor based methods

Let x_i, x_j are the raw images and $f(x_i), f(x_j)$ are the extracted descriptors.

$$\text{Cosine similarity} = \frac{f(x_i)^T.f(x_j)}{||f(x_i)||.||f(x_j)||}$$

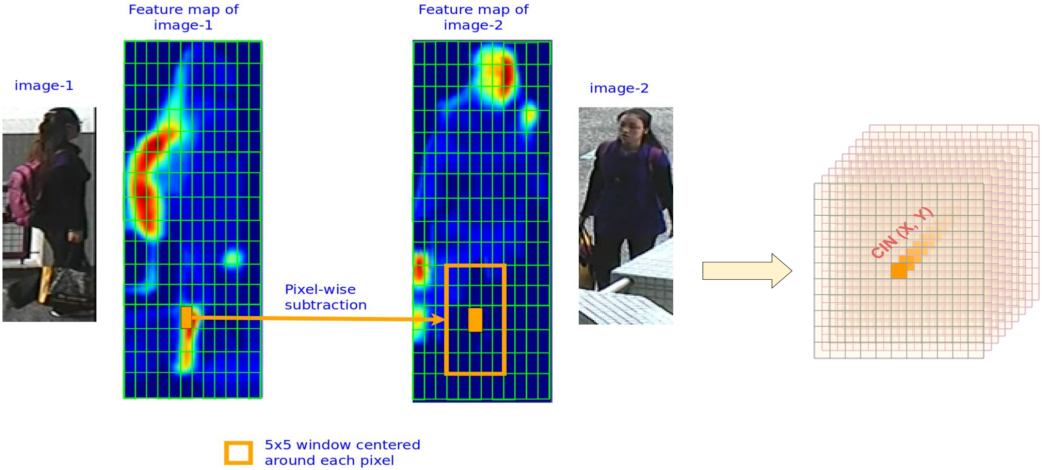
$$\text{Hinge loss} = \begin{cases} ||f(x_i) - f(x_j)||_2^2, & \text{if } i = j, \\ max(0, ||f(x_i) - f(x_j)||_2^2 - \alpha_{hinge}), & \text{if } i \neq j \end{cases}$$

Improved Siamese Architecture



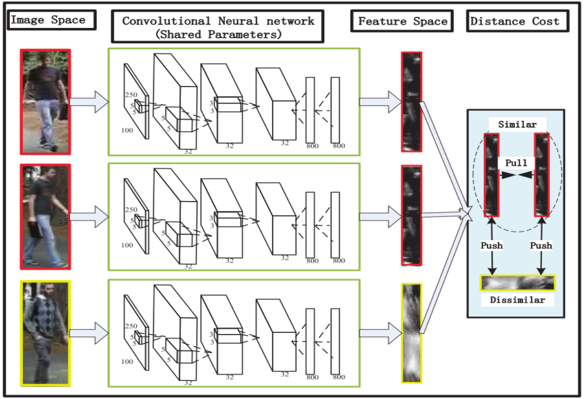
Ahmed *et al.* **An improved deep learning architecture for person re-identification**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2015.

Cross-Input Neighborhood matching layer



Ahmed *et al.* **An improved deep learning architecture for person re-identification**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2015.

Triplet networks



Triplet pair of images x_i, x_j, x_k where x_i, x_j are the pairs belonging to the same person and x_k is an image of different person. The triplet loss function with Euclidean metric can be defined as follows:

$$L_{triplet} = \sum_{i,j,k}^{N} max(0, [||f(x_i) - f(x_j)||_2^2 - ||f(x_i) - f(x_k)||_2^2 + \alpha_{triplet}])$$

The loss function using learned distance metric is given by:

$$L_{triplet} = \sum_{i,j,k}^{N} max\left(0, \left[g(x_i, x_j) - g(x_i, x_k) + \alpha_{triplet}\right]\right) \quad (1)$$

Cheng *et al.* **Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2016.

Improved Triplet Loss funtion for person re-identification

Loss function : Let $f(x_i), f(x_j), f(x_k)$ be the extracted features/descriptors of the images x_i, x_j, x_k respectively. Here x_i, x_j = images of same person and x_k = image of a different person.

$$\begin{aligned} L_{triplet} &= max(0, [d(f(x_i), f(x_j)) \\ &\quad - d(f(x_j), f(x_k)) + \tau_1]) \end{aligned}$$

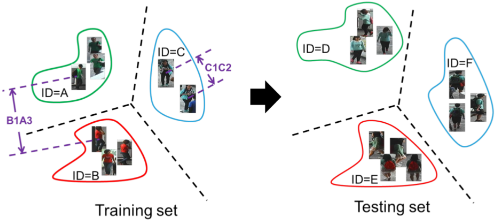
Cheng *et al.* **Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2016.

Improved Triplet Loss funtion for person re-identification

Loss function : Let $f(x_i), f(x_j), f(x_k)$ be the extracted features/descriptors of the images x_i, x_j, x_k respectively. Here x_i, x_j = images of same person and x_k = image of a different person.

$$L_{triplet} = max(0, [d(f(x_i), f(x_j)) - d(f(x_j), f(x_k)) + \tau_1])$$

Expected:



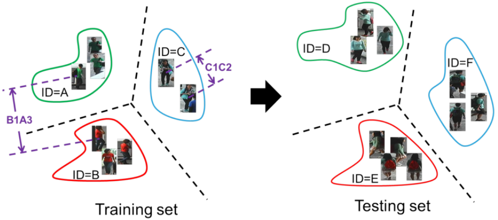
Cheng *et al.* **Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2016.

Improved Triplet Loss funtion for person re-identification

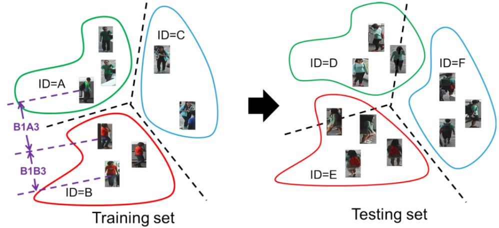
Loss function : Let $f(x_i), f(x_j), f(x_k)$ be the extracted features/descriptors of the images x_i, x_j, x_k respectively. Here x_i, x_j = images of same person and x_k = image of a different person.

$$L_{triplet} = max(0, [d(f(x_i), f(x_j)) - d(f(x_j), f(x_k)) + \tau_1])$$

Expected:



Observed caveat in generalization:



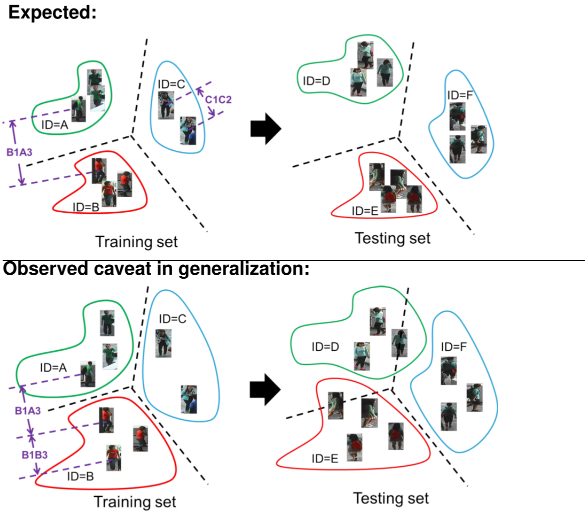
Cheng *et al.* **Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2016.

Improved Triplet Loss funtion for person re-identification

Loss function : Let $f(x_i), f(x_j), f(x_k)$ be the extracted features/descriptors of the images x_i, x_j, x_k respectively. Here x_i, x_j = images of same person and x_k = image of a different person.

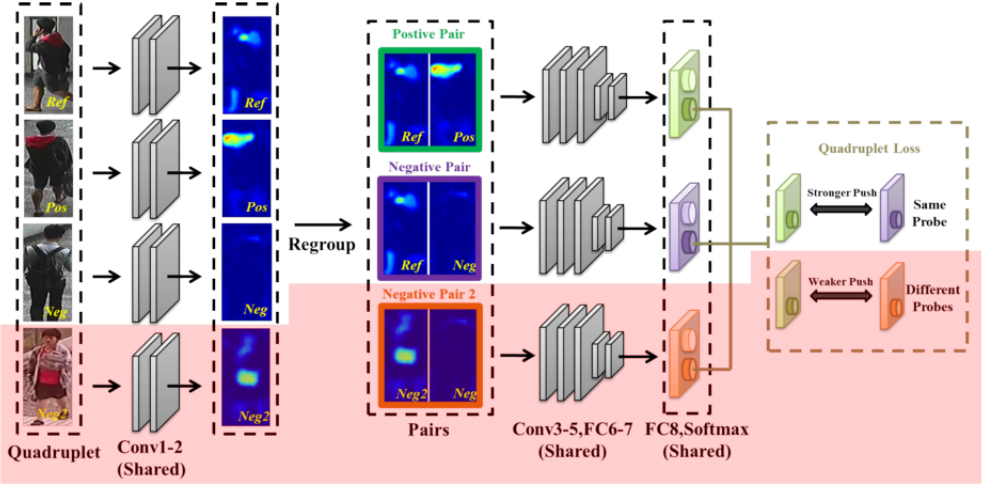
$$
\begin{aligned}
L_{triplet} &= max(0, [d(f(x_i), f(x_j)) \\
&\quad -d(f(x_j), f(x_k)) + \tau_1])
\end{aligned}
$$

$$
\begin{aligned}
L_{+ve\_constraint} &= max\left(0, \left[d(f(x_i), f(x_j)) - \tau_2\right]\right) \\
L_{total} &= \sum_{i,j,k}^{N} \left(L_{triplet} + L_{+ve\_constraint}\right)
\end{aligned}
$$



Cheng *et al.* **Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2016.

Quadruplet network for person re-identification



Chen *et al.* **Beyond triplet loss: a deep quadruplet network for person re-identification**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2017.

Quadruplet network for person re-identification
Novel loss function

In quadruplet loss, an extra term is added to the triplet loss function.

$$
\begin{aligned}
L_{quad} = \sum_{i,j,k}^{N} max(\left[g(x_i, x_j) - g(x_i, x_k) + \alpha_1\right], 0) + \\
\sum_{i,j,k,l}^{N} \max(\left[g(x_i, x_j) - g(x_l, x_k) + \alpha_2\right], 0) \\
s_i = s_j, s_l \neq s_k, s_i \neq s_l, s_i \neq s_k
\end{aligned} \tag{2}
$$

where α_1 and α_2 are the values of margins in two terms and s_i refers to the person ID of image x_i.

Chen *et al.* **Beyond triplet loss: a deep quadruplet network for person re-identification**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2017.

Evaluation

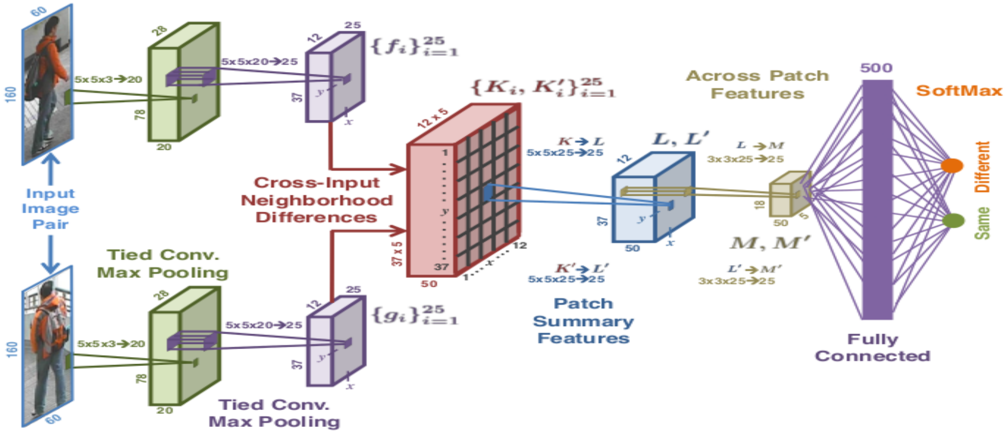
Method	ViPeR[1]	CUHK01[2]	CUHK03-labeled[3]	CUHK03-detected	QMULGRID[4]
Deep-kronecker-matching, CVPR-2018	-	-	91.1	91.0	-
Quadruplet Net, CVPR 2017	49.05	81.00	75.53	-	-
Imp. TripletNet, CVPR 2016	47.8	53.7	-	-	-
Ahmed et al (CIN), CVPR 2015	34.81	65.0	54.74	44.76	-
LOMO + XQDA, CVPR 2015	40.0	-	52.20	46.25	18.96
KISSME + PCA, CVPR 2012	19.6	-	-	-	-

Figure: Rank-1 identification rates for different datasets

[1] Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features, ECCV-2008
[2] Human Reidentification with Transferred Metric Learning, ACCV-2012
[3] DeepReID: Deep Filter Pairing Neural Network for Person Re-identification, CVPR-2014
[4] On-the-fly feature importance mining for person re-identification, PR-2014

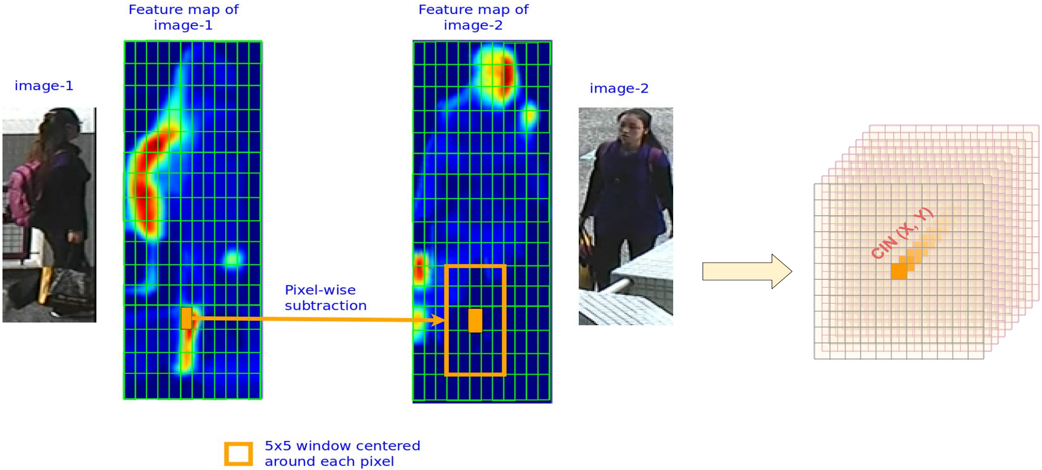
Outline

Improved Siamese Architecture



Ahmed *et al.* **An improved deep learning architecture for person re-identification**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2015.

Cross-Input Neighborhood matching layer

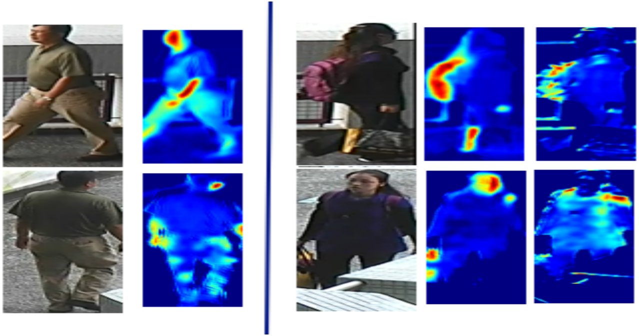


Ahmed *et al.* **An improved deep learning architecture for person re-identification**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2015.

Drawbacks

Notable drawbacks

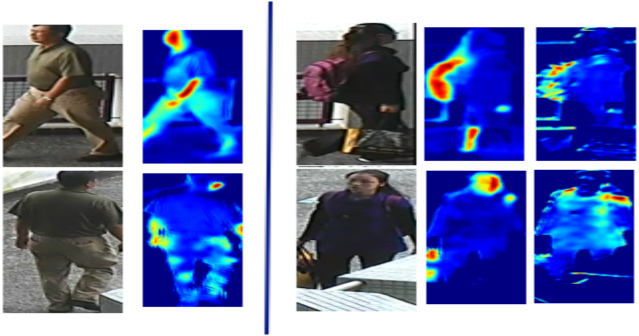
- Smaller neighborhood 5×5 may not be enough to capture pose variations, partial occlusions



Drawbacks

Notable drawbacks

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

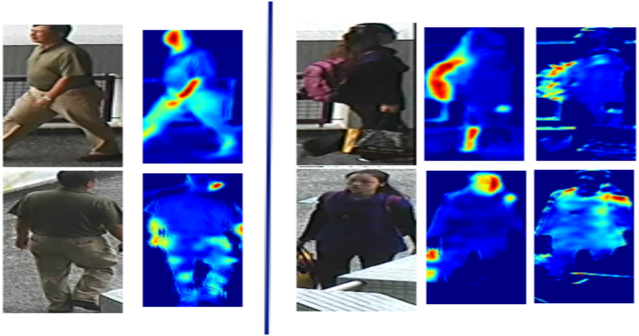


(Alternate) solution: increase the search space (horizontally up to full width)

Drawbacks

Notable drawbacks

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



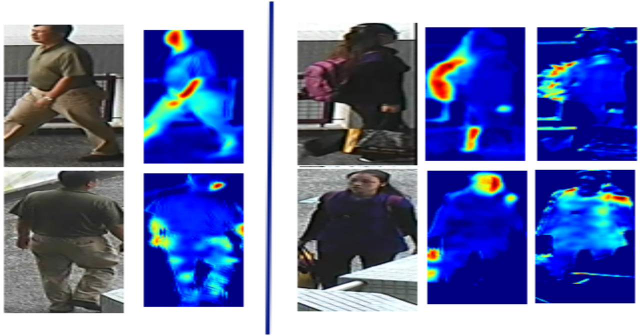
(Alternate) solution: increase the search space (horizontally up to full width)

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

Drawbacks

Notable drawbacks

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



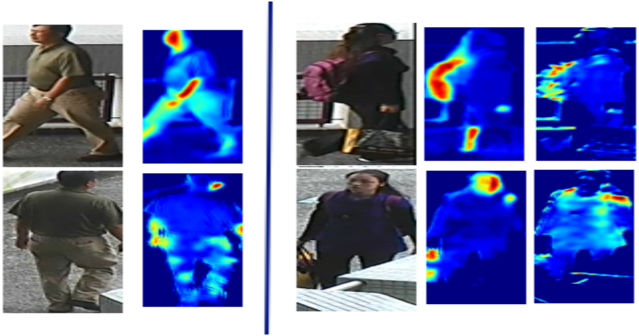
(Alternate) solution: increase the search space (horizontally up to full width)

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

Drawbacks

Notable drawbacks

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



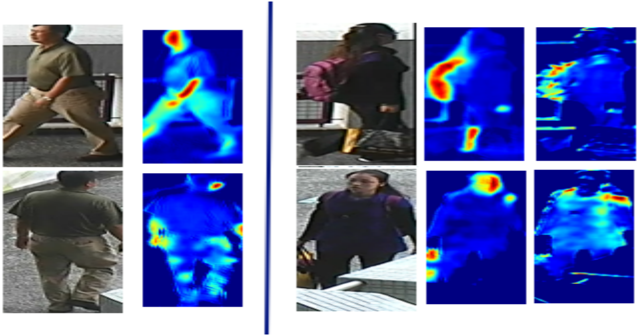
(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact 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

Drawbacks

Notable drawbacks

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



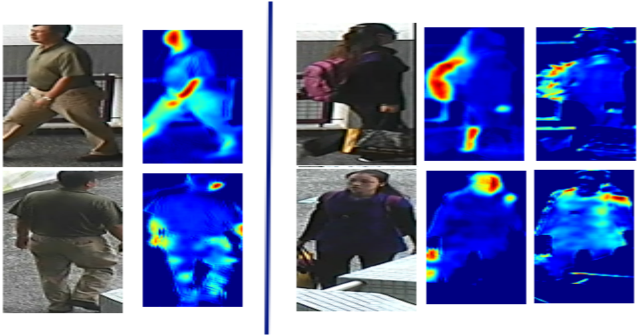
(Alternate) solution: increase the search space (horizontally up to full width)

- Performing an exact 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

Drawbacks

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 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
  $\rightarrow$ Normalized correlation

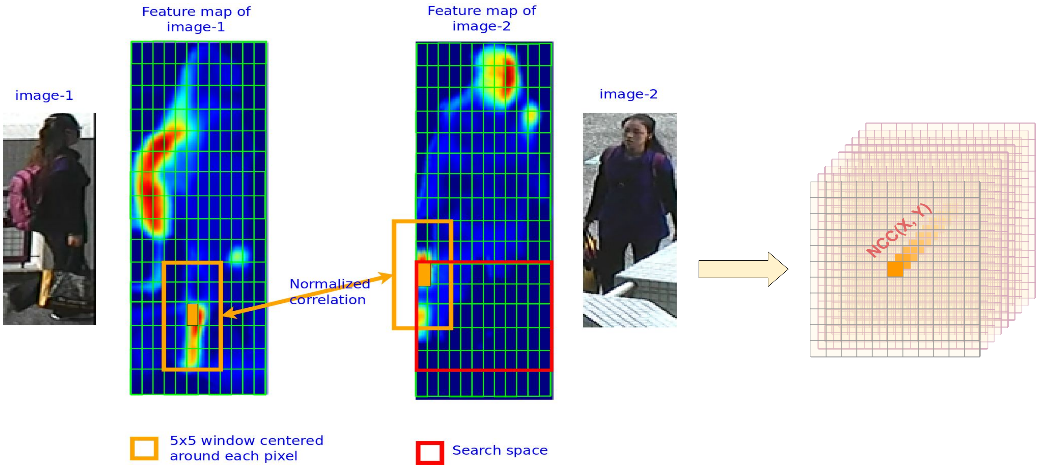
Counter-points to the baseline work

Ahmed et. al (CVPR'15)

With wider search space, the ablation study conducted on different datasets of various sizes gives the following results:

Title	Ahmed et. al, 5x5 search			Ahmed et. al, 5x12 search		
	r = 1	r = 10	r = 50	r = 1	r = 10	r = 50
CUHK03 labeled	54.74	93.3	99.7	**57.60**	90.63	99.17
CUHK03 detected	44.96	83.47	99.4	**54.31**	90.24	99.18
CUHK01 100 tests	65.0	94.0	99.9	**69.7**	95.03	99.13
CUHK01 486 tests	47.5	80.25	96.30	**49.31**	81.48	95.95

Normalized Correlation Matching layer



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

Normalized correlation (NCC)

Consider two arrays E, F with N-elements each.
The normalized correlation between E and F is defined as,

$$normxcorr(E,F) = \frac{\sum_{i=1}^{N}(E_i - \mu_E) * (F_i - \mu_F)}{(N-1) * \sigma_E * \sigma_F} \tag{3}$$

where

$$\text{mean } \mu_E = \frac{\sum_{i=1}^{N} E_i}{N} \tag{4}$$

$$\text{unbiased standard deviation } \sigma_E = \sqrt{\frac{\sum_{i=1}^{N}(E_i - \mu_E)^2}{N-1}} \tag{5}$$

The gradient of the normalized correlation matching is calculated as shown below:

$$\frac{\partial normxcorr(E,F)}{\partial E_i} = \frac{1}{(N-1) * \sigma_E} * \left(\frac{F_i - \mu_F}{\sigma_F} - \frac{normxcorr(E,F) * (E_i - \mu_E)}{\sigma_E} \right) \tag{6}$$

Normalized correlation (NCC)

Consider two arrays E, F with N-elements each.
The normalized correlation between E and F is defined as,

$$normxcorr(E, F) = \frac{\sum_{i=1}^{N} (E_i - \mu_E) * (F_i - \mu_F)}{(N-1) * \sigma_E * \sigma_F} \tag{3}$$

where

$$\text{mean } \mu_E = \frac{\sum_{i=1}^{N} E_i}{N} \tag{4}$$

$$\text{unbiased standard deviation } \sigma_E = \sqrt{\frac{\sum_{i=1}^{N} (E_i - \mu_E)^2}{N-1}} \tag{5}$$

The gradient of the normalized correlation matching is calculated as shown below:

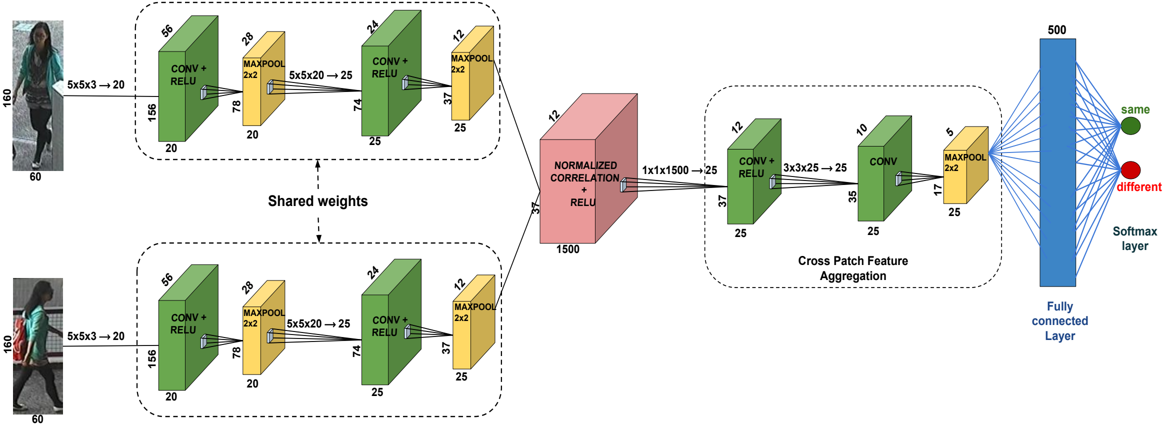
$$\frac{\partial normxcorr(E, F)}{\partial E_i} = \frac{1}{(N-1) * \sigma_E} * \left(\frac{F_i - \mu_F}{\sigma_F} - \frac{normxcorr(E, F) * (E_i - \mu_E)}{\sigma_E} \right) \tag{6}$$

NCC properties

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

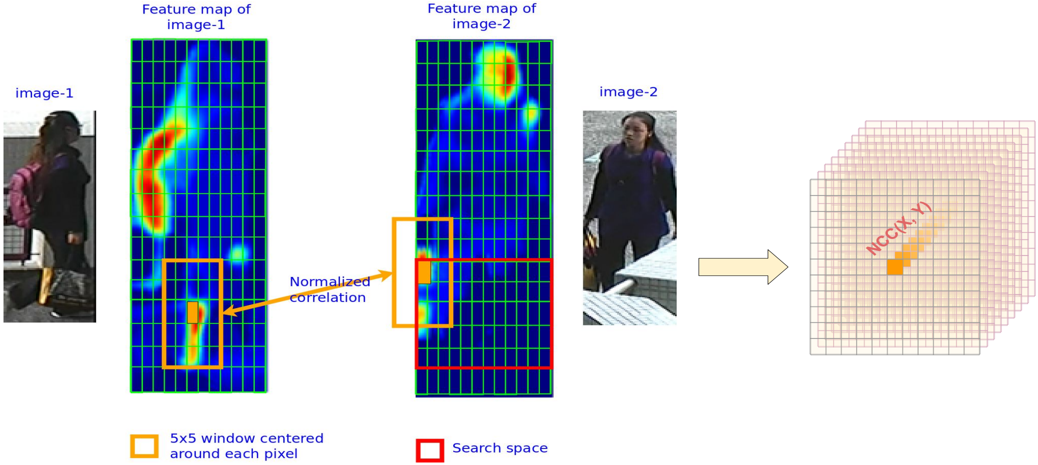
Arulkumar Subramaniam, Moitreya Chatterjee, and Anurag Mittal. **Deep Neural Networks with Inexact Matching for Person Re-Identification**. Proceedings of the Neural Information Processing Systems (NIPS) - 2016.

Normxcorr model



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

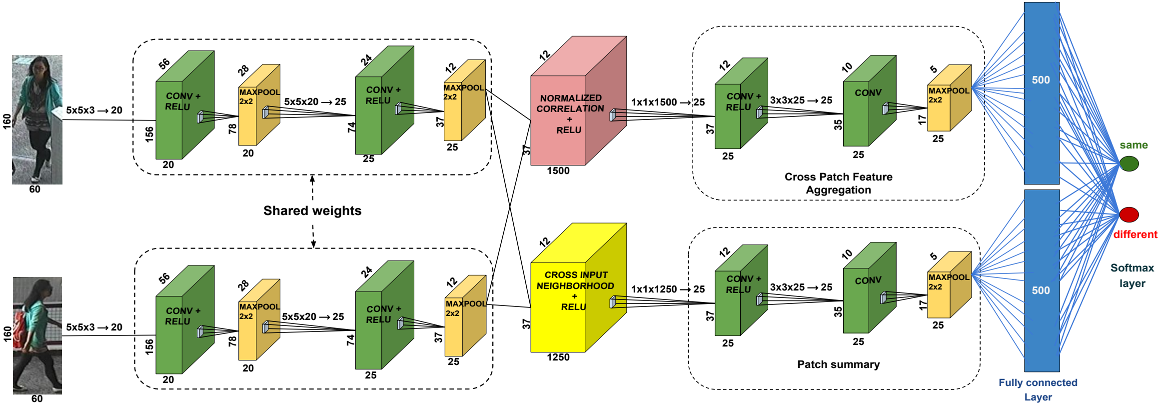
Normalized Correlation Matching layer



- NCC: Matching textures, undefined for texture-less surfaces

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

Fused model



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

Evaluation

Method	ViPeR[1]	CUHK01[2]	CUHK03-labeled[3]	CUHK03-detected	QMULGRID[4]
Deep-kronecker-matching, CVPR-2018	-	-	91.1	91.0	-
Quadruplet Net, CVPR 2017	49.05	81.00	75.53	-	-
Imp. TripletNet, CVPR 2016	47.8	53.7	-	-	-
Subramaniam et al (NCC+CIN), NIPS 2016	-	81.23	72.43	72.04	19.20
Subramaniam et al (NCC), NIPS 2016	-	77.43	64.73	67.13	16.00
Ahmed et al (CIN), CVPR 2015	34.81	65.0	54.74	44.76	-
LOMO + XQDA, CVPR 2015	40.0	-	52.20	46.25	18.96
KISSME + PCA, CVPR 2012	19.6	-	-	-	-

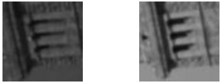
Figure: Rank-1 identification rates for different datasets

[1] Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features, ECCV-2008
[2] Human Reidentification with Transferred Metric Learning, ACCV-2012
[3] DeepReID: Deep Filter Pairing Neural Network for Person Re-identification, CVPR-2014
[4] On-the-fly feature importance mining for person re-identification, PR-2014

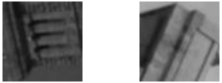
Outline

Patch matching

- Patch Matching
  - fundamental task in Computer Vision
  - find correspondences of textured regions that are centered at distinctive keypoints



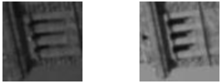
Same



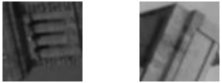
Different

- Applications
  - Image Registration and Mosaicking
  - Stereo Matching
  - 3-D Reconstruction
  - Object tracking
- Challenges
  - photometric changes (illumination)
  - geometric changes (affine)

Dataset



Same



Different

Dataset: Patches of 2D correspondences (64 x 64, $\sim$250K same, $\sim$250K different patch pairs) from 3D reconstructions of:

1. Statue of Liberty
2. Notre Dame
3. Yosemite

Evaluation procedure:

- Train on one dataset (e.g., Statue of Liberty),
- Test on other 2 datasets (Notre Dame, Yosemite)

Evaluation Criteria: FPR at 95% TPR (lower is better)

Dataset: Winder *et al.*. **Learning local image descriptors**. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2007.

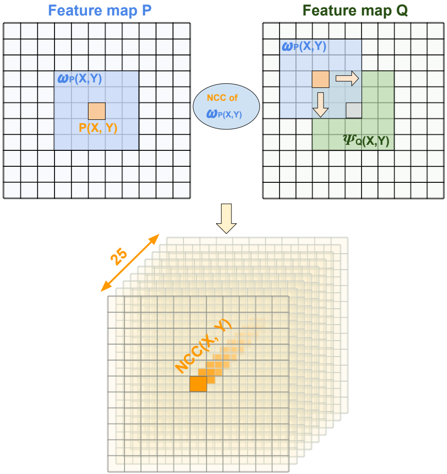
Existing approaches

- Hand-crafted techniques
  - Lowe et. al - SIFT
  - other descriptors - HOG, LTP, LBP, DAISY etc
  - Disadvantages:
    - Only applicable to specific challenges
    - Not suitable for generic / unpredictable scenarios
- Deep-Learning techniques
  - Zagoruyko et al. CVPR 2015 - DeepCompare - Various architectures such as Siamese, Central stream
  - Kumar et al. CVPR 2016 - Global loss function & Triplet network

Outline

NCC for Patch Matching

5 $\times$ 5 window and 5 $\times$ 5 search space



- Notable change from Person Re-identification setup:
  - Search space is reduced to 5 $\times$ 5 (Reason: patches are normalized up to Similarity)

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.

Outline

NCC for Patch Matching
Siamese network

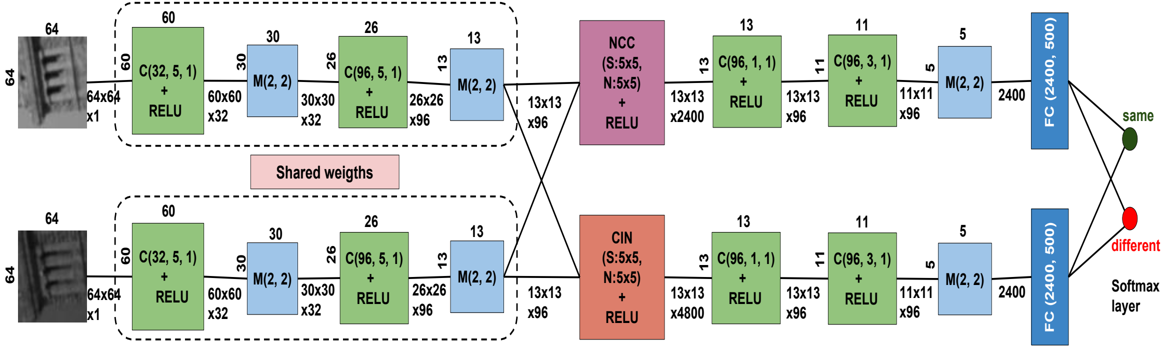
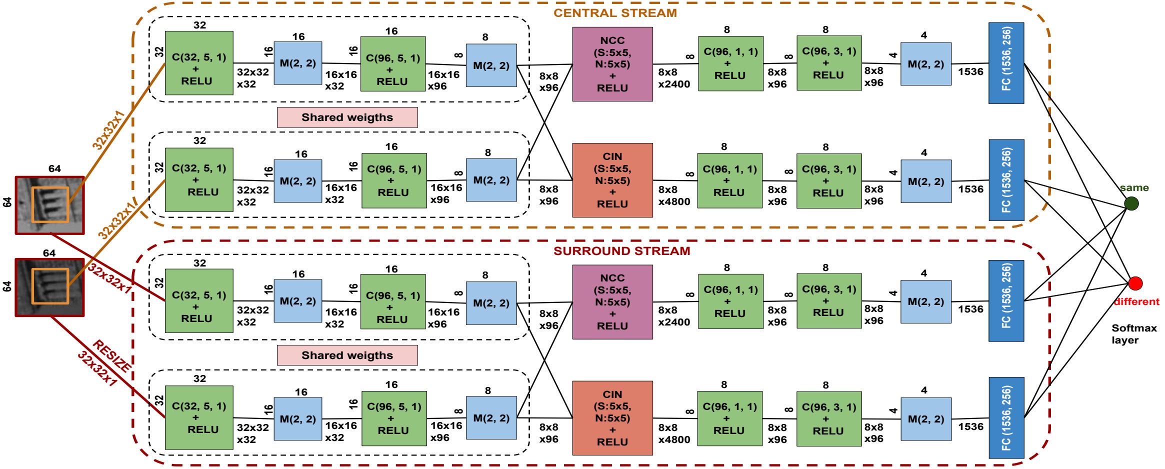


Figure: C(N,m,s) = Convolutional layer with N filters, each of size $m \times m$ and stride of s, M(n,s)= max-pooling layer, *NCC(n, m)* = NCC layer with support region $n \times n$ and $m \times m$ search space.

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.

NCC for Patch Matching
Central stream-Surround stream network



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.

Outline

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

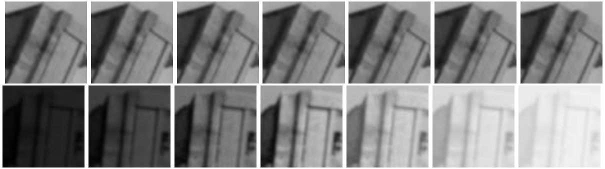
Table: *FPR95* scores of the proposed models and the baselines. These being False Positive Rates (FPR), lower their values, better is their performance. Testbed: *UBC Patches* dataset. **Color coding :** **red** - best performing method, **blue** - next best performing method (Training : 500K pairs, Testing: 100K pairs)

Experiments for illumination changes
Manual variation of pixel intensities

Affine-transformation based illumination variation:

$$I_i(x,y) = A.I(x,y) + B, \quad \text{where } A = \frac{N-i}{N}, \quad B = \frac{i\mathbb{E}}{N} \tag{7}$$

where $\mathbb{E} = 0$ or 255 based on under- or -over saturation



U(8) U(6) U(3) U(0) O(3) O(6) O(8)

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.

Experiments for illumination changes

Manual variation of pixel intensities

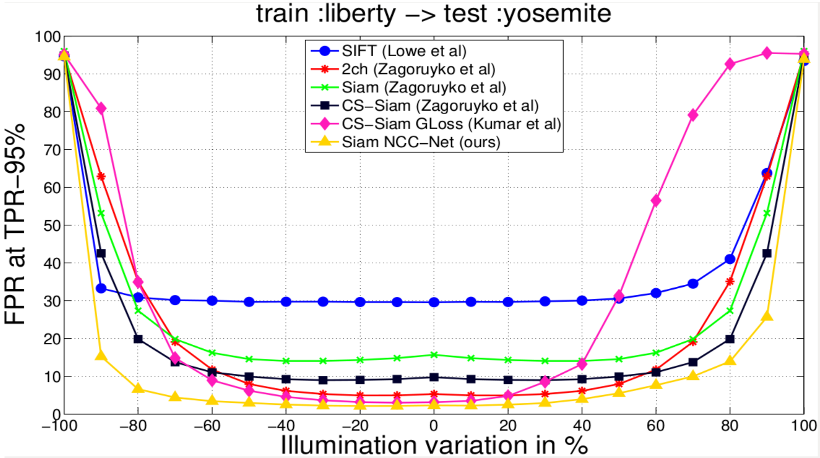


Figure: Impact of changing pixel intensities on the matching performance. Here the illumination change is induced by the affine-based variation of intensities

Experiments for illumination changes
Manual variation of pixel intensities

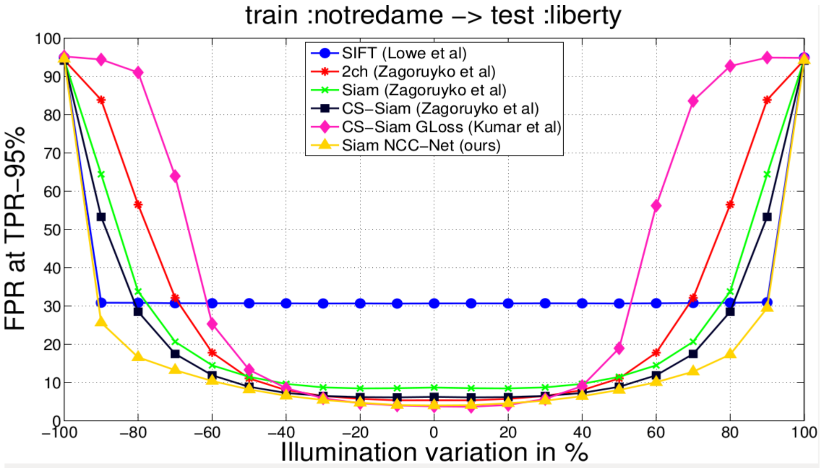


Figure: Impact of changing pixel intensities on the matching performance. Here the illumination change is induced by the affine-based variation of intensities

Experiments for illumination changes
Tests on Natural intensity changes

Figure: Typical natural illumination variations noticed in the Webcam dataset (category: Mexico)

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.

Summary

In this seminar

- Adaptation of NCC to match images/patches in Deep learning framework
- The matching method is generic
- can be applied to other computer vision tasks - Current/Future work

Summary

In this seminar

- Adaptation of NCC to match images/patches in Deep learning framework
- The matching method is generic
- can be applied to other computer vision tasks - Current/Future work

Current line of Research

- Background clutters and mis-alignment errors in person re-identification
  - 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.
- Multi-modal person re-identification methods (RGB-IR (night-vision), Text based)

Conference proceedings

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, 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 (NIPS) - 2016. Barcelona, Spain.

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.

Thank you!