NCC-Net: Normalized Cross Correlation Based Deep Matcher with Robustness to Illumination Variations

Arulkumar Subramaniam*, Prashanth Balasubramanian*, Anurag Mittal

Indian Institute of Technology Madras

Objectives

Problem definition : Patch Matching aims to find correspondences of localized, textured regions which are usually centered at distinctive keypoints. The matching has to be robust across many geometric and photometric changes.

Contributions :

1. Exploration of two Deep Convolutional Neural Network (DCNN) architectures along with the combination of two robust matching layers **to overcome illumination variation challenges**:
   - Siamese and Center-surround architectures
   - Normalized correlation and Cross-input neighborhood matching layers
2. Empirical evaluation of the proposed architectures for resilience with respect to **illumination variations** on two experimental setups:
   - manual variation of the pixel intensities of the patches in the *UBC Patches* dataset
   - natural illumination changes from the real-world, using a new dataset of patches collected from publicly available *Webcam* dataset
3. Ablation study to evaluate individual matching layers, illumination based augmentation while training of the models

Some prior approaches

Zagoruyko et al.(CVPR-2015) : Deepcompare : proposed and evaluated several architectures such as Siamese (Siam, 2-Stream), Central-surround (CS), 2-Channel etc.) for the task of Patch matching

Balntas et. al (BMVC-2016) : TFeat, PN-Net : proposed a triplet-based loss function and aimed to increase the speed of descriptor computation and decrease its memory footprint

Kumar et al (CVPR-2016) : GLoss : global loss that aims to decrease the intra-class variances and increase the inter-class distance

Tian et al (CVPR-2017) : L2-Net : learn descriptors in Euclidean space with additional supervision on the intermediate feature maps

Robust matching layers

Normalized Cross Correlation(NCC) layer: *NCC* is a statistical measure of the tendency of two signals to vary linearly with each other. It is given by

$$NCC(X,Y) = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(X_i-\mu_X)(Y_i-\mu_Y)}{\sigma_X\sigma_Y} \quad (1)$$

where N = number of samples, (μ_X, μ_Y) = mean and (σ_X, σ_Y) = unbiased standard deviations of signals X and Y. NCC is well-known to alleviate amplification variation while matching signals. A differentiable layer based on NCC[1] within the CNN framework is proposed to match the extracted CNN features from the given patches.

Cross-input Neighborhood(CIN) layer: Though the patches are expected to have sufficient textures, there may arise some pathological cases which lack them. In such cases, *NCC* is unreliable as σ_X or σ_Y vanishes . As a **remedy**, it is fused with the Cross-input Neighborhood(*CIN*)(Ahmed et. al, CVPR 2015 [2]) layer. ***CIN* builds "difference maps" between every pixel of a feature map and a neighborhood window (5 × 5) of its corresponding feature map** to help the network learn discriminative features from absolute pixel differences.

Training particulars

- Training with UBC patches dataset (Three partitions of the dataset: Liberty, Notredame, Yosemite - 500K pairs of same/different patches)
- Given a pair of patches, each of the proposed model outputs the probability that the patches are same or different using a softmax layer.
- The standard cross-entropy loss is used for training. It is given by

$$L = -\frac{1}{N}\sum_{i=0}^{N}(t_i \log p_i + (1-t_i)\log(1-p_i)) \quad (2)$$

Code: `https://github.com/InnovArul/patchmatch_normxcorr`

* equal contribution

Siamese (Siam-NCC-Net) architecture

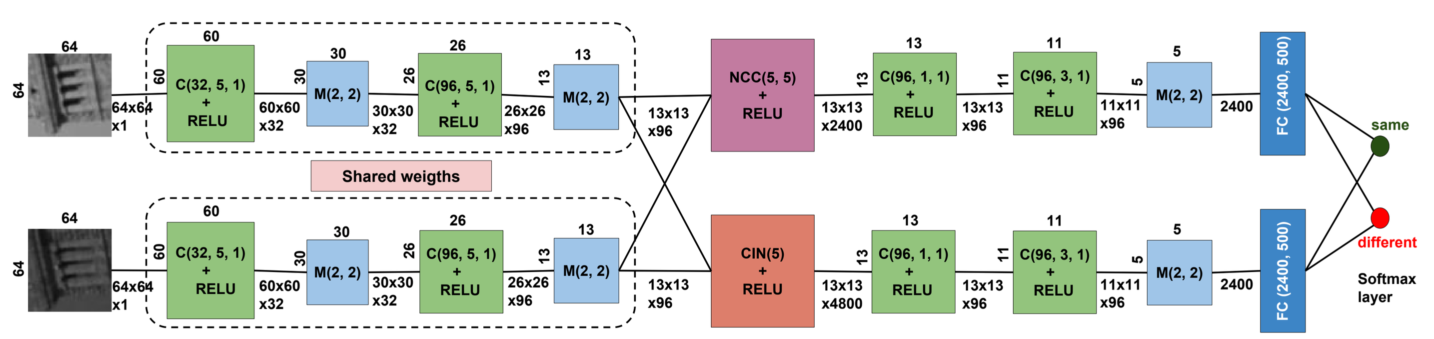

Figure: **Siam-NCC-Net** architecture. Here,
C(N,m,s) = Convolutional layer of N filters with filter size $m \times m$ and a stride of s pixels,
M(n,s) = max-pooling layer with receptive field of $n \times n$ and a stride of s pixels,
NCC(n, m) = *NCC* layer matching an $n \times n$ support region centered at any pixel with its $m \times m$ search space,
CIN(m) = *CIN* layer with a search space of size $m \times m$,
FC(n, m) = Fully-connected layer with n inputs and m outputs

Central-Surround (CS-NCC-Net) architecture

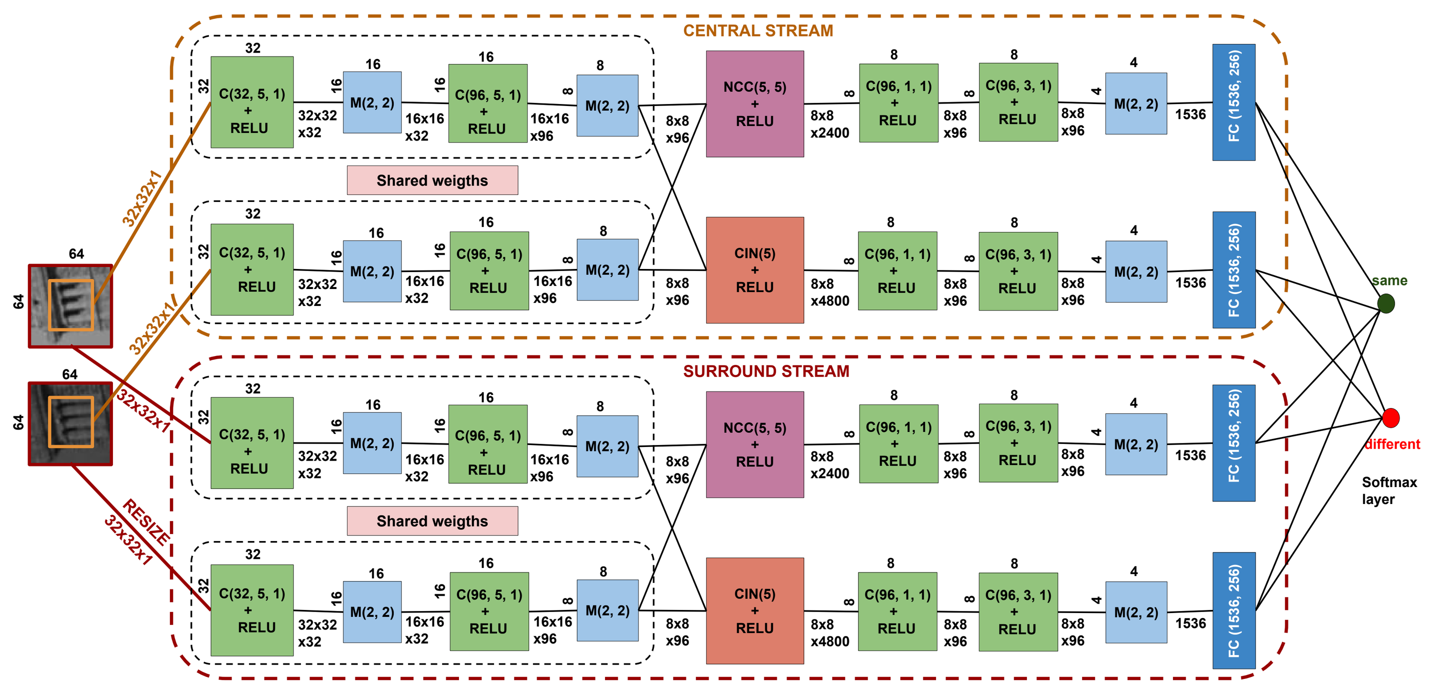

Figure: **CS-NCC-Net** architecture with *NCC* and *CIN* matching layers

Illustration of Normalized correlation matching layer

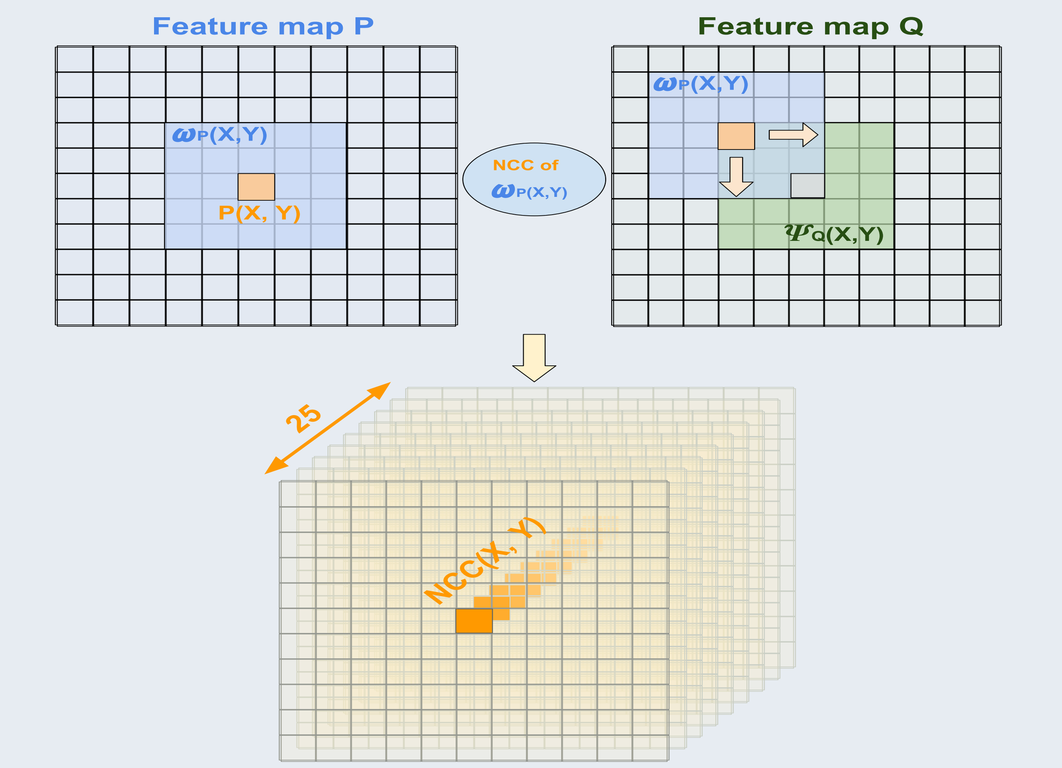

Figure: Illustrations of support regions. Using *NCC*, a support region $\omega_{P(x,y)}$ on the left is compared with multiple, neighboring support regions $\omega_{Q(...)}$ on the right. Each $\omega_{Q(...)}$, on the right, is centered at a green pixel. Comparing a support region with multiple neighboring regions, as shown here, helps the network to learn the patterns from a large neighborhood. The *NCC* between P and Q results in an *NCC* feature map, shown in the bottom row.

Results

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 : **best**, **second best**.

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
L2-Net (Tian et al (CVPR-2017)))[3]	**0.56**	**2.07**	**1.71**	**1.76**	**3.87**	**1.09**	**1.84**
DeepCD (Yang et. al,(CVPR-2017))	2.59	7.03	5.85	6.69	7.82	2.95	5.49
2ch-CS stream GLoss (Kumar et al (CVPR-2016))	**0.77**	3.09	**3.69**	2.67	4.91	**1.14**	2.71
2ch-CS stream (Zagoruyko et al.(CVPR-2015))	1.9	4.75	4.55	4.1	7.2	2.11	4.10
Siamese GLoss (Kumar et al (CVPR-2016))	1.84	6.61	6.39	5.57	8.43	2.83	5.28
TFeat (Balntas et. al (BMVC-2016))	3.12	7.82	7.22	7.08	9.79	3.85	6.48
PNNet (Balntas et. al)	3.71	8.99	8.13	7.1	9.65	4.23	6.97
Siam CS-stream (Zagoruyko et al.(CVPR-2015))	3.05	9.02	6.45	10.45	11.51	5.29	7.63
MatchNet (Serra et. al (ICCV-2015))	4.75	13.58	8.84	11.00	13.02	7.7	9.81
Siam (Zagoruyko et al.(CVPR-2015))	4.33	14.89	8.77	13.23	13.48	5.75	10.07
VGG-Convex (Simonyan et. al (PAMI-2014))	7.52	11.63	12.88	10.54	14.82	7.11	10.75

Illumination experiment - I (Manual illumination variation)

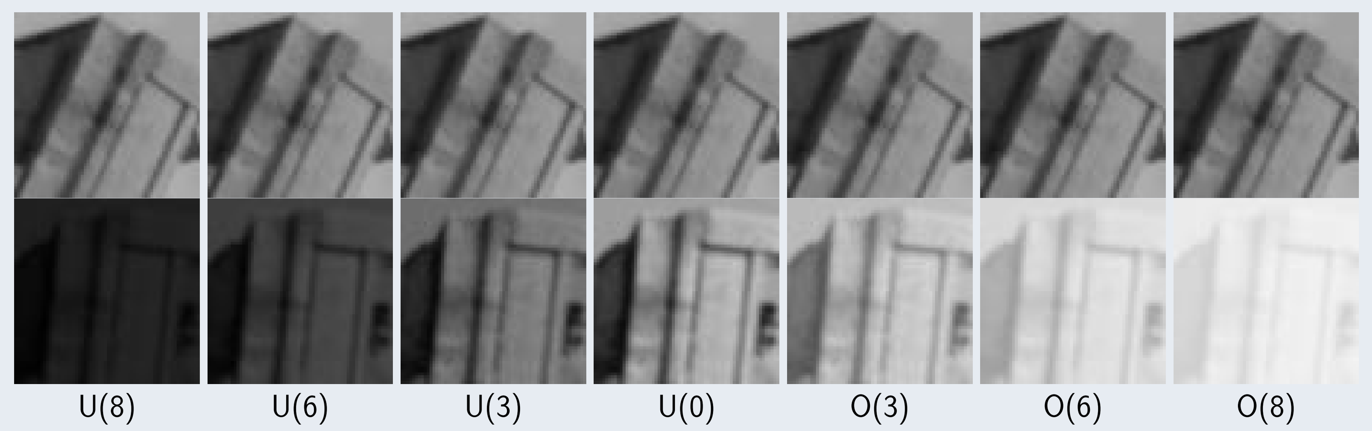

Figure: Each column shows the varying degree of intensity change introduced to a corresponding pair from *UBC Patches* dataset. Here the illumination change is induced by the models, $U(i) : I_i(x,y) = \frac{(N-i)*I(x,y)}{N}, O(i) : I_i(x,y) = \frac{(N-i)*I(x,y)}{N} + \frac{i\mathbb{E}}{N}$. Here, U = under-saturation region, O = over-saturation region

Illumination experiment - I results

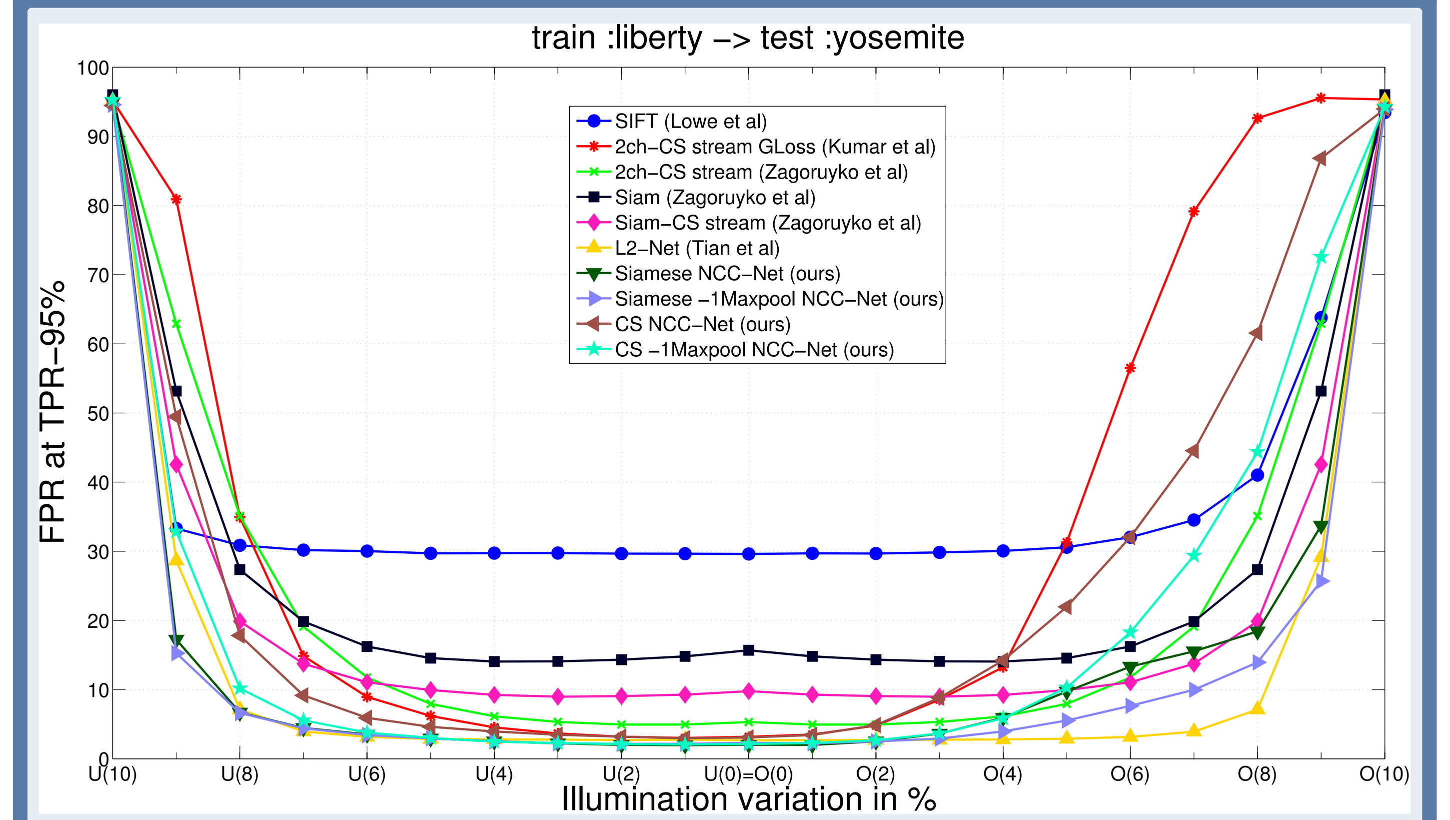

Figure: Impact of changing pixel intensities on the matching performance

Illumination experiment - II (Natural illumination variation dataset)

Figure: Typical natural illumination variations noticed in the Webcam dataset (category: Mexico). Notice the extreme change in illuminations as some portions are under-saturated while others are over-saturated. Figure best viewed in color on a display device.

Illumination experiment-II results

Table: Results on the newly collected Webcam dataset. *FPR95* scores of the proposed models and the baselines. These being False Positive Rates (FPR), lower their values, better are their performances. Datasets trained from: Liberty, Notredame, Yosemite. Scores obtained from 262, 152 test patch-pairs of this newly collected Webcam dataset.

Trained dataset	Liberty	Notredame	Yosemite
Siam-NCC-Net(ours)	9.67	16.68	11.67
Siam-w/oMP$_2$-NCC-Net(ours)	**9.45**	12.56	**10.56**
CS-NCC-Net(ours)	11.34	23.04	15.40
CS-w/oMP$_2$-NCC-Net(ours)	11.35	19.30	18.28
L_2 Net	9.67	**9.21**	16.11
2ch-CS-stream GLoss	12.31	20.01	14.76
2ch-CS stream	12.31	17.84	19.5
Siam	31.45	28.21	35.21
Siam-CS stream	27.08	32.17	34.65

Ablation study-I

Table: *FPR95* scores when using the individual matching layers in **Siam-w/oMP$_2$-NCC-Net** on UBC patches dataset[4]. . Here, L : Liberty, N : Notredame, Y : Yosemite

Train	Liberty		Notredame		Yosemeite	
Test	N	Y	L	Y	L	N
NCC	1.23	1.78	4.17	2.24	4.98	2.00
CIN	1.87	5.40	5.27	4.55	7.34	3.11
Both	1.14	2.30	4.02	2.34	4.71	1.81

Ablation study-II

Table: *FPR95* scores of *DeepCompare*[5] networks retrained on illumination specific augmentation for *UBC Patches* dataset[4]. Here, L : Liberty, N : Notredame, Y : Yosemite

Train	L		N		Y	
Test	N	Y	L	Y	L	N
2ch-CS stream+	1.82	3.73	2.85	2.56	5.99	1.34
Siam CS-stream+	3.56	9.29	6.46	9.56	11.53	5.47
Siam+	6.59	11.92	7.98	12.07	13.43	8.36

References

[1] A. Subramaniam, M. Chatterjee, and A. Mittal, "Deep neural networks with inexact matching for person re-identification," in *Advances in Neural Information Processing Systems 29* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, eds.), pp. 2667–2675, Curran Associates, Inc., 2016.

[2] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *The IEEE Conference on Computer Vision and Pattern Recognition*, June 2015.

[3] Y. Tian, B. Fan, and F. Wu, "L2-net: Deep learning of discriminative patch descriptor in euclidean space," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2017.

[4] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 43–57, Jan 2011.

[5] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition*, June 2015.