GenLR-Net: Deep framework for very low resolution face and object recognition with generalization to unseen categories

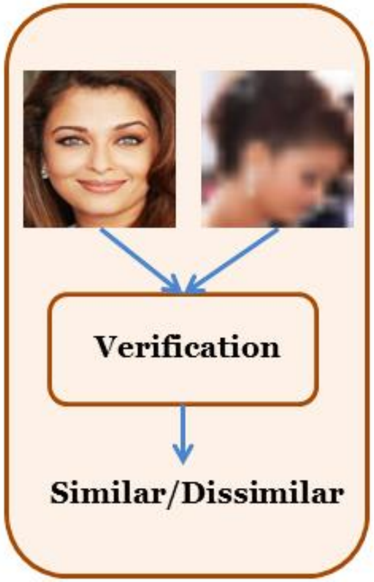
CVPR 2018 Workshop on Biometrics.

Sivaram Prasad Mudunuri, Soubhik Sanyal*, and Soma Biswas*

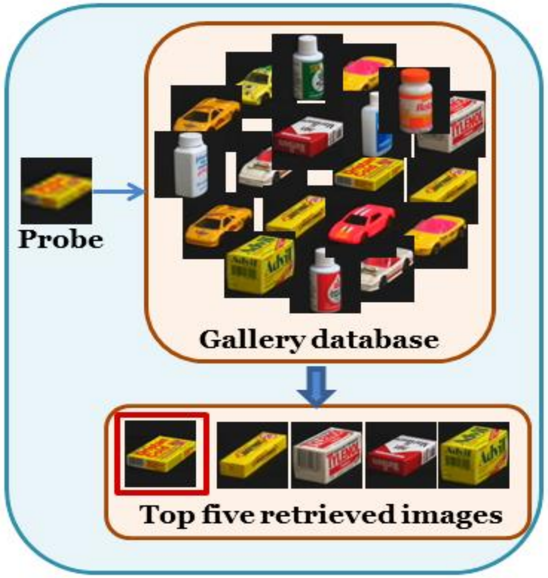
Image Analysis and Computer Vision Lab
Department of Electrical Engineering
Indian Institute of Science, India.

*[*Equal Contributions]*

Cross Domain Matching: Motivation



Low-Resolution Face Verification



Low-Resolution Object Retrieval

- ❑ Facial images captured under uncontrolled environment – Pose, illumination, and Resolution.

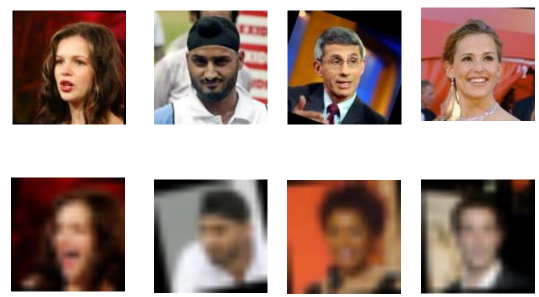
- ❑ Low-resolution Object retrieval: street-to-shop matching for general objects.

Problem Statement

During training, we assume that we have access to both LR and HR images of the training subjects:

x_i^h : i th HR image

x_i^l : i th LR image



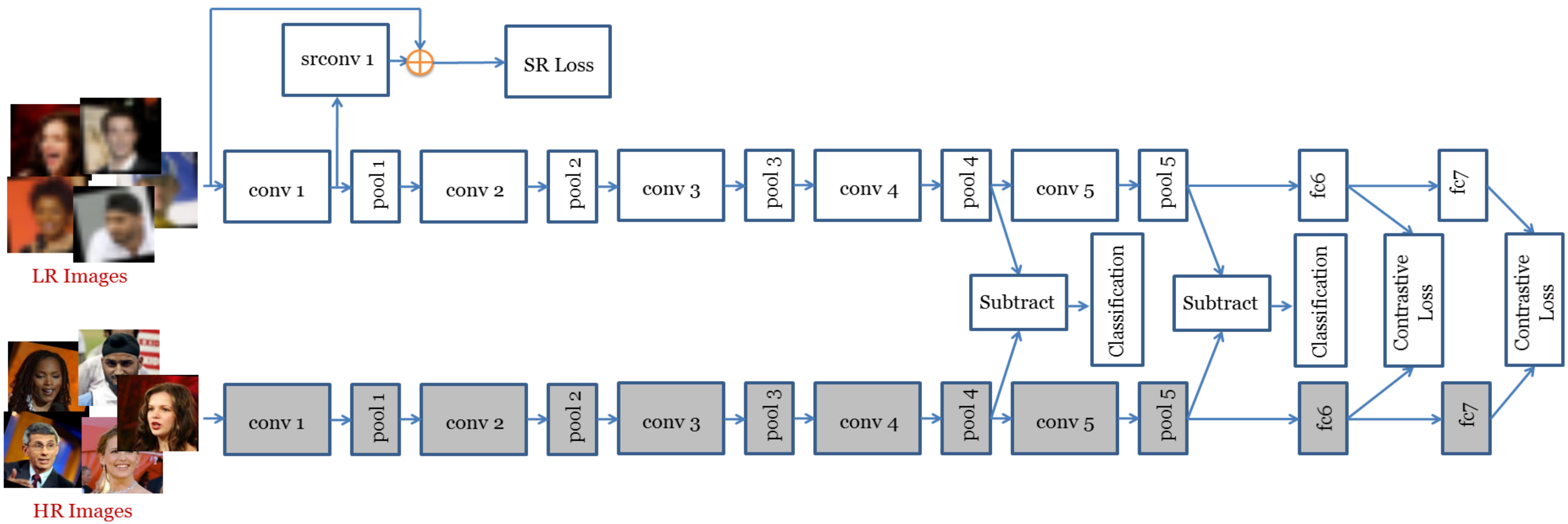
l_i : Label of the i th pair;
1 if they are matched pair;
0 otherwise.

1 1 0 0

❑ Testing subjects are completely different than the training subjects.

❑ During testing, given an image pair, where one is LR and the other is HR, the goal is to verify whether they belong to the same subject or not.

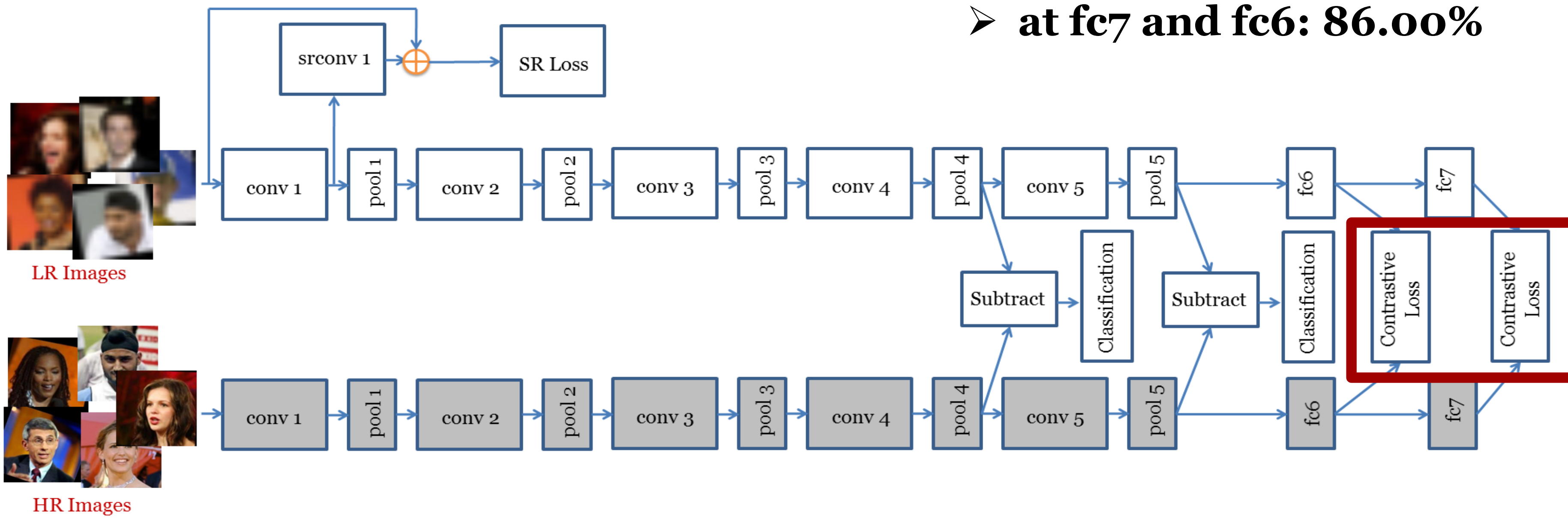
Proposed GenLR-Net



Block diagram of the proposed GenLR-Net.

Proposed GenLR-Net

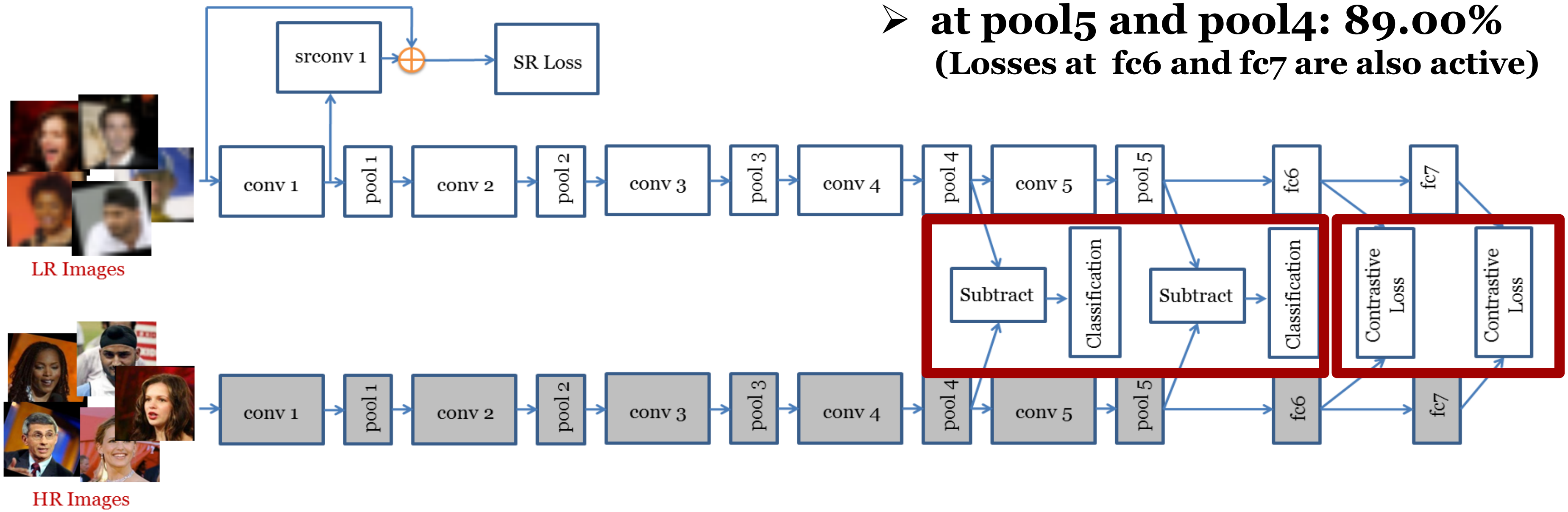
- ➢ **Only at fc7: 84.00%**
- ➢ **at fc7 and fc6: 86.00%**



Contrastive Loss at high level features [Hadsell *et al.*, CVPR 2006, Varior *et al.*, ECCV 2016]:

$$L_{cont} = \frac{1}{2N} \sum_{i=1}^{N} l_i D_i^2 + (1 - l_i) \max(\delta - D_i, 0)^2$$

Proposed GenLR-Net



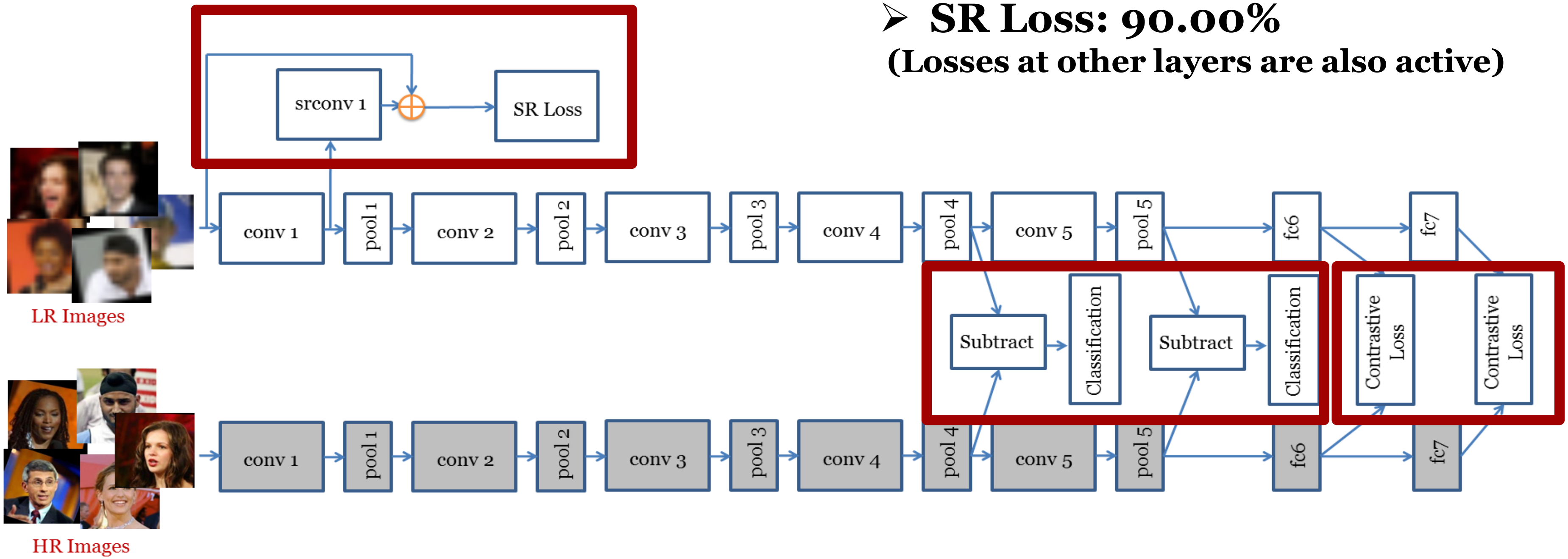
- **at pool5: 87.24%**
- **at pool5 and pool4: 89.00%**

(Losses at fc6 and fc7 are also active)

Inter-intra Classification Loss at the mid-level features [Lee *et al.*, AIS 2015.]:

- Difference between the two features (HR and LR) is computed and classified as 1 or 0.
- Thus, a N-class problem is converted to a 2-class problem.

Proposed GenLR-Net



- **SR Loss: 90.00%**

(Losses at other layers are also active)

Super-Resolution Loss at the low-level features [Cai *et al.*, arxivs 2017]:

- We include the super-resolution objective along with the verification task.

$$L_{SR} = \left\| (s_i^l + x_i^l) - (x_i^l)_{hr} \right\|_2^2$$

Overall Loss function

In summary, we train the entire network by jointly minimizing all the losses.

$$L = \lambda_1 L_{SR} + \lambda_2 L_{Cls}^{pool4} + \lambda_3 L_{Cls}^{pool5} + \lambda_4 L_{Cont}^{fc6} + \boxed{\lambda_5 L_{Cont}^{fc7}}$$

84.00%

Overall Loss function

In summary, we train the entire network by jointly minimizing all the losses.

$$L = \lambda_1 L_{SR} + \lambda_2 L_{Cls}^{pool4} + \lambda_3 L_{Cls}^{pool5} + \boxed{\lambda_4 L_{Cont}^{fc6} + \lambda_5 L_{Cont}^{fc7}}$$

86.00%

Overall Loss function

In summary, we train the entire network by jointly minimizing all the losses.

$$L = \lambda_1 L_{SR} + \lambda_2 L_{Cls}^{pool4} + \boxed{\lambda_3 L_{Cls}^{pool5} + \lambda_4 L_{Cont}^{fc6} + \lambda_5 L_{Cont}^{fc7}}$$

87.24 %

Overall Loss function

In summary, we train the entire network by jointly minimizing all the losses.

$$L = \lambda_1 L_{SR} + \boxed{\lambda_2 L_{Cls}^{pool4} + \lambda_3 L_{Cls}^{pool5} + \lambda_4 L_{Cont}^{fc6} + \lambda_5 L_{Cont}^{fc7}}$$

89.00 %

Overall Loss function

In summary, we train the entire network by jointly minimizing all the losses.

$$L = \boxed{\lambda_1 L_{SR} + \lambda_2 L_{Cls}^{pool4} + \lambda_3 L_{Cls}^{pool5} + \lambda_4 L_{Cont}^{fc6} + \lambda_5 L_{Cont}^{fc7}}$$

90.00 %

Experimental Results:

Cross-resolution face verification:

❑ **Experiments on modified version of LFW database [Huang *et al.*, UMass, Technical Report, 2007]:**

We conduct the experiment on fold 1 of the database and using the LFW-deep-funneled images.

Training samples: 2700 similar pairs and 2700 dissimilar pairs
Test samples: 300 similar and 300 dissimilar pairs as per the standard protocol but with the modified resolutions.

The LR images are obtained by down-sampling the original images to 20x20 and then up-sampled to the original resolution using bi-linear interpolation.

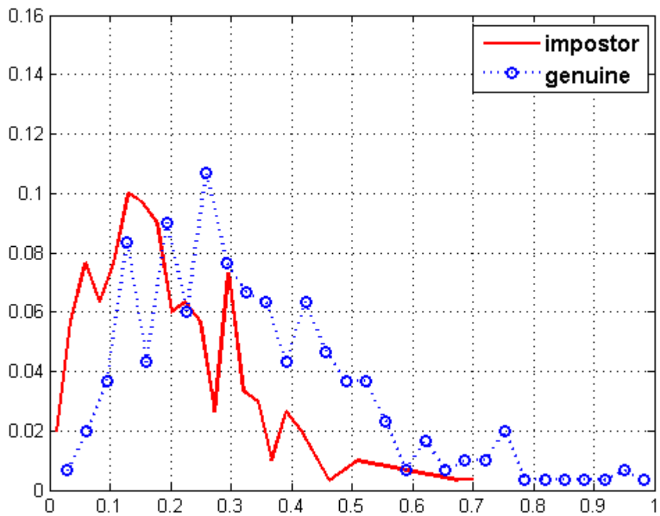
(a) Matched pair **(b)** Non-matched pair

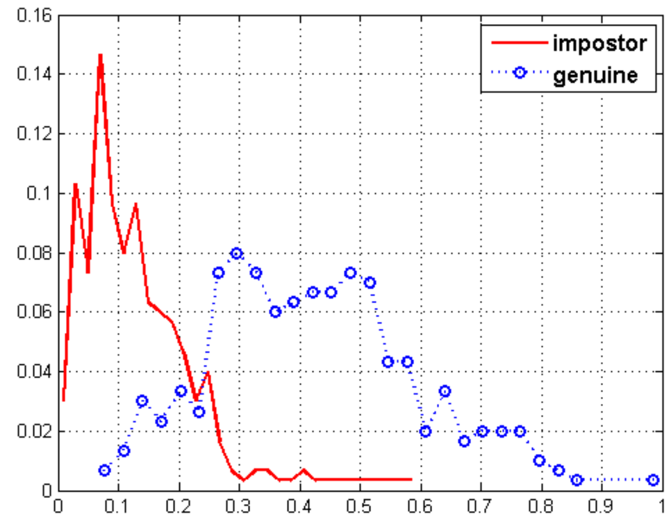
Experiments on modified version of LFW:

Table: Performance Comparison of Proposed Approach on modified version of LFW Database.

Method	Verification rate (%)
HR-HR (original VGG)	**93.83**
HR-LR (original VGG)	69.16
SSR [Kim *et al.*, TPAMI, 2010] + fc7 features	72.10
SRCNN [Dong *et al.*, TPAMI, 2016] + fc7 features	73.16
LapSRN [Lai *et al.*, CVPR, 2017] + fc7 features	76.16
fc7 features + SA [Fernando *et al.*, ICCV, 2013]	72.50
fc7 features + LSML [Kostinger *et al.*, CVPR, 2012]	71.00
Proposed GenLR-Net	**90.00**

Experiments on modified version of LFW:





GT: Similar
GenLR-Net: Similar

GT: Similar
GenLR-Net: Dissimilar

GT: Dissimilar
GenLR-Net: Dissimilar

GT: Dissimilar
GenLR-Net: Similar

Experiments on modified version of LFW:

Effect of resolution variations:

Table: Verification performance (%) using the proposed GenLR-Net for different resolutions of the LR images.

Models	20x20	10x10	5x5
Original VGG-Face	69.16	56.17	54.50
Proposed GenLR-Net (Fine-tuned with 20x20 images)	90.00	67.70	62.30
Proposed GenLR-Net (Fine-tuned with 10x10 images)	-	72.17	65.00
Proposed GenLR-Net (Fine-tuned with 5x5 images)	-	-	65.80

Experiments on CFP in wild database [Sengupta *et al.*, WACV 2016]

- Cross-resolution and cross-pose matching.
- It has 10 splits and each split has 350 matched pairs and 350 non-matched pairs.
- Probe faces are down-sampled to 20x20 and up-sampled to the original resolution using bi-linear interpolation.
- We evaluate the proposed approach on the first fold in which we train our network on 9 splits and test it on the remaining split.

(a)

Matched Pair

(b)

Non-matched Pair

Table: Performance Evaluation of Proposed Approach on CFP Database [Sengupta *et al.*, WACV 2016].

Method	Verification rate (%)
HR-HR (original VGG)	**88.57**
HR-LR (original VGG)	71.71
Proposed GenLR-Net	**77.28**

Cross-resolution object recognition

- In many realistic scenarios, the query image may come from a class which the model has not seen.
- The final goal is to match the uncontrolled, unseen query with the relatively controlled data items and retrieve the similar ones.

Experiments on COIL-100 Database [Nene *et al.*, 1996]:

- The database has 100 categories and each category has 72 images with different pose.
- 90 categories are selected as seen categories and the remaining 10 are treated as unseen categories.
- We randomly select 60 images (out of 72) from each of those 90 objects to generate the matched and non-matched pairs for training and validation.
- In each pair, one image is of high-resolution and the second one is of low-resolution (20x20).

Experiments on COIL-100 Database:

Table: Rank-1 accuracy (%) on COIL-100 database [Nene *et al.*, 1996] under different protocols.

Method	Seen in seen	Unseen in unseen	Seen in all	Unseen in all
HR-HR (original VGG Object)	**97.57**	**99.66**	**97.57**	**88.66**
HR-LR (original VGG Object)	68.99	90.67	67.17	69.67
Fine-tuned VGG-Object on LR data	78.28	92.33	77.17	75.67
LapSRN + fc7 features	88.18	94.00	87.47	76.67
Proposed GenLR-Net	**93.13**	**98.00**	**91.21**	**81.00**

Figure: Cross-resolution object retrieval results of GenLR-Net on COIL-100 [Nene *et al.*, 1996]. Each row shows top five retrieved results (column 2-6) corresponding to the LR query (first column). The first two rows are from seen in all protocol, and last two rows are from unseen in all protocol. Correct match is denoted by the red box.

Experiments on Toy Cars database:

Object verification in uncontrolled settings:

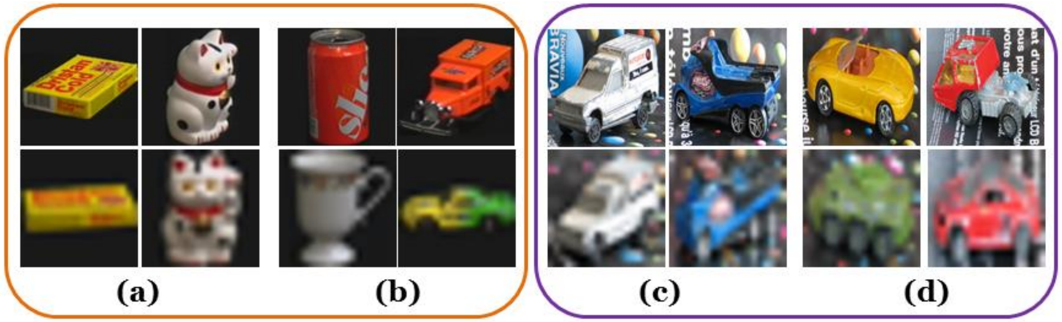


Table: Verification performance (%) of the proposed approach on Toy Cars database [Nowak *et al.*, CVPR, 2007] .

Method	Verification rate (%)
HR-HR (original VGG)	**94.54**
HR-LR (original VGG)	86.15
Proposed GenLR-Net	**88.09**

Summary and Contributions

- We proposed a novel deep learning framework to address the challenging problem of matching low-resolution probe images (faces/objects) against high-resolution images in the database.
- We also addressed the very challenging and practical problem of unseen object recognition, which is a relatively unexplored area.
- Extensive experiments on different datasets.
- Apart from our work, matching real surveillance quality images against HR images is still unexplored area of research and yet to be addressed thoroughly.

Thank You.

Queries?