Semi-Supervised Learning for Low-light Image Restoration through Quality Assisted Pseudo-Labeling (Supplementary Material)

Sameer Malik
Indian Institute of Science
Bengaluru, India
sameer@iisc.ac.in

Rajiv Soundararajan
Indian Institute of Science
Bengaluru, India
rajivs@iisc.ac.in

1. Overview

In this supplementary we cover the following

1. More Details and performance evaluation of SMSNet
2. Architecture of QFCNN
3. More dataset details and hyperparameter choice discussion.
4. Implementation details of adversarial and mean teacher methods
5. Distributions of SSIM scores for the proposed and Mean Teacher method
6. Additional visual examples and ensemble analysis results on FUJI and LOL datasets

2. Details of SMSNet

We discuss architecture and training details of SMSNet, the LLIR model we use for semi-supervised learning. We also evaluate its performance when compared to other popular LLIR models.

Bandpass subbband learning: We employ the same Bandpass CNN architecture consisting of a series of twenty 3×3 convolutional layers with ReLU activations and local residual connections to restore each subband. The CNNs take as input the Gaussian pyramid of the low-light image at each level and output the restored band pass subband. We use the L1 loss between the predicted and the ground truth subbands to train each subband. In Figure 1

Lowpass subbband learning: The architecture for Lowpass CNN is similar to that of Bandpass CNN, except that we also include instance normalization layers as shown in Figure 2. Note that we do not include these normalization layers in all the modules as we observed some drop in performance by including these layers in all the modules. Also note that, we did not find any benefit of including these normalization layers in the Bandpass CNNs. We train Lowpass

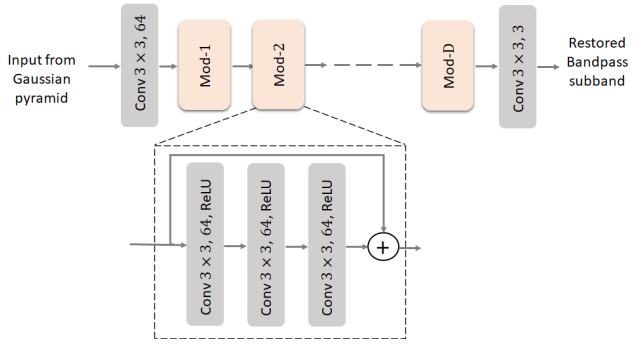

Figure 1. Architecture of the Bandpass CNN.

CNN by using a combination of SSIM and L1 error as a loss function with equal weights.

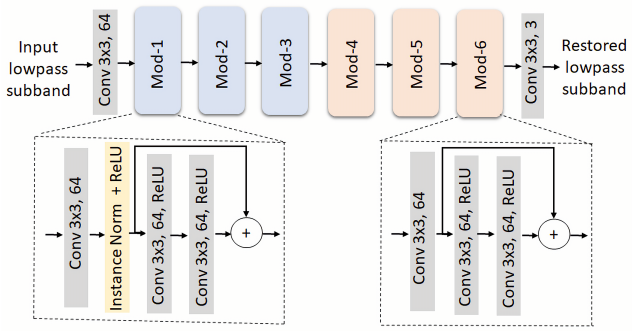

Figure 2. Architecture of the Lowpass CNN.

Performance Evaluation: We now evaluate the performance of SMSNet. We use 5 scales in SMSNet, which amounts to 4 bandpass subbands and a lowpass subband. We compare it with several methods including residual learning based DLN [4], retinex theory based approach KIND [6] and recent multiscale subband learning approaches DRBN [5]. In Table 1, we report the SSIM and PSNR scores for these methods on all three datasets used in the paper. Note that SMSNet performs very well when compared to other methods. In our experiments, we find that

inclusion of instance normalization layers in the Lowpass CNN is very important for good performance of SMSNet, especially on SONY and FUJI datasets. There is a significant drop in performance (0.06 and 0.08 in SSIM for SONY and FUJI respectively) without its use. Other methods can also perhaps perform better on the SONY and FUJI datasets with the inclusion of these layers. However, their architectures are complex and incorporating these layers into them is not straightforward.

Table 1. SSIM/PSNR numbers for various methods on 100% labelled datasets.

Methods	SONY	FUJI	LOL
DLN	0.63/19.33	**0.59/19.85**	**0.81/21.95**
KIND	0.32/17.65	0.24/15.85	0.80/20.86
DRBN	**0.68/19.94**	0.58/18.98	**0.83**/20.13
SMSNet	**0.74/23.05**	**0.69/22.63**	0.78/**21.91**

3. Architecture of QFCNN

We now discuss the architecture of our QFCNN model which we use to compute quality features. In QFCNN, each of the modules has a convolutional layer, a Batch normalization layer and a dropout layer. We do not use any residual connections in QFCNN. Finally, 128 dimensional features are extracted from QFCNN using an adaptive average pooling layer which makes it invariant to input image size.

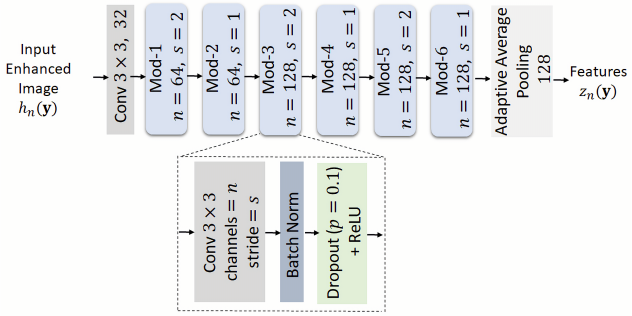

Figure 3. Architecture of QFCNN.

4. Additional Experiment Details and Results

Here we discuss datasets in more details and also present the implementation details of adversarial and mean teacher methods which we benchmarked for semi-supervised learning for LLIR. Further, we also present some additional analysis and visual examples.

4.1. Datasets

The SID dataset has two subsets of images, one from a SONY camera and the other from a FUJI camera. We treat these as separate datasets. While the SONY dataset contains 1865 images for training and 598 for testing, the FUJI dataset contains 1654 training images and 523 testing images. The images in these datasets are captured by varying the exposure times and ISO settings of different scenes. The SID dataset has images in *raw* format. We thus pre-process the *raw* images using the python library *rawpy* to obtain sRGB images. While processing the raw images using *rawpy*, we set the field *no_auto_bright*=**False** to turn on the usual camera pipeline operations of linear enhancement and gamma correction. In our experiments, we subsample all the images to a resolution of 832×1248.

LOL dataset already has sRGB images with 400×600 resolution and thus we do not pre-process them.

4.2. Justification for the choice of hyperparameters

Since we work with very less training data, there is not enough data to construct a validation set. Thus, we selected the hyperparameters such as learning rate (LR), LR decay and number of epochs by observing the convergence of the training loss curve for one or two labeled-unlabeled data splits. We verified for these splits that the convergence of training loss does not lead to overfitting by verifying on the original validation set of the SID data. Then we keep the hyperparameters constant across all the 10 splits of a given dataset. The parameter choices for the LOL dataset were scaled based on its size with respect to the SID datasets. We observe that the performance is fairly robust to the exact choice of these parameters. τ is kept fixed across all datasets and splits.

4.3. Implementation details of various methods

Adversarial Loss: For implementing adversarial loss training on unlabelled data, we adopt least squares GAN based adversarial training [1]. We use a batch size of 16 in which every batch consists of equal number of labelled and unlabelled data. For the labelled data we use the same loss as used for training Lowpass CNN in SMSNet while the adversarial loss is computed on the unlabelled data. Overall loss is linear combination of loss on labelled data and unlabelled data with equal weights. Similar to SMSNet, we train it for 90 epochs. We use a LR of $1e-3$ and $2e-4$ for Lowpass CNN and the discriminator respectively. For Lowpass CNN the LR is reduced to $1e-4$ and $1e-5$ after 50 and 70 epochs respectively. In our implementation we use a discriminator with spectral normalization layers [2]. We show the architecture of the discriminator in Figure 4.

Mean Teacher: For mean teacher [3] training we read equal number of labelled and unlabelled data in a batch. We use a batch size of 16. We train mean teacher for 90 epochs with LR of $1e-3$ reduced to $1e-4$ and $1e-5$ after 50 and 70 epochs respectively. Let, $h(.)$ denote the Lowpass CNN model used to process low light images. Let, θ and θ' denote the corresponding parameters of the student and

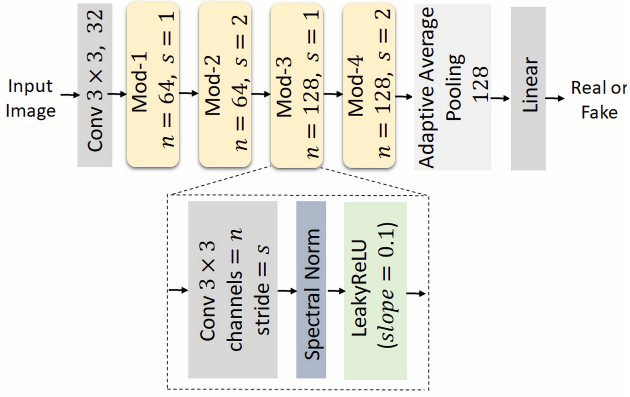

Figure 4. Discriminator architecture.

teacher model respectively. Then the loss on an unlabelled image $\mathbf{y}$ is

$$J(\theta) = \|h(\mathbf{y}, \theta, \mathbf{w}) - h(\mathbf{y}, \theta', \mathbf{w}')\|^2,$$

where $\mathbf{w}$ is the zero mean Gaussian noise. Similar to the original implementation, we use standard deviation of 0.15. At iteration t, θ' is computed as

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta.$$

We rampup α to 0.95 in the first 20 epochs.

4.4. Distribution of SSIM of our method vs. Mean Teacher

In Figure 5, we show the box plots for distribution of SSIM on different splits for our method when compared to the Mean Teacher. We see that our method outperforms the Mean Teacher by achieving a higher SSIM score with a smaller variance across the splits.

4.5. Visual Examples

4.5.1 Visual comparison of SMSNet with other restoration methods

In Figure 6 and 7, we show visual comparisons of restored low light images by various methods when trained in a fully supervised fashion on 100% and 5% labelled data respectively. Note that SMSNet restored images are perceptually better than other methods in both the cases. This shows that even in limited training data setting SMSNet yields better quality images than other methods.

4.5.2 Additional visual examples of various semi-supervised methods

In Figure 8, we show more visual examples of restored low light images by various semi-supervised methods for the 5% labelled data case.

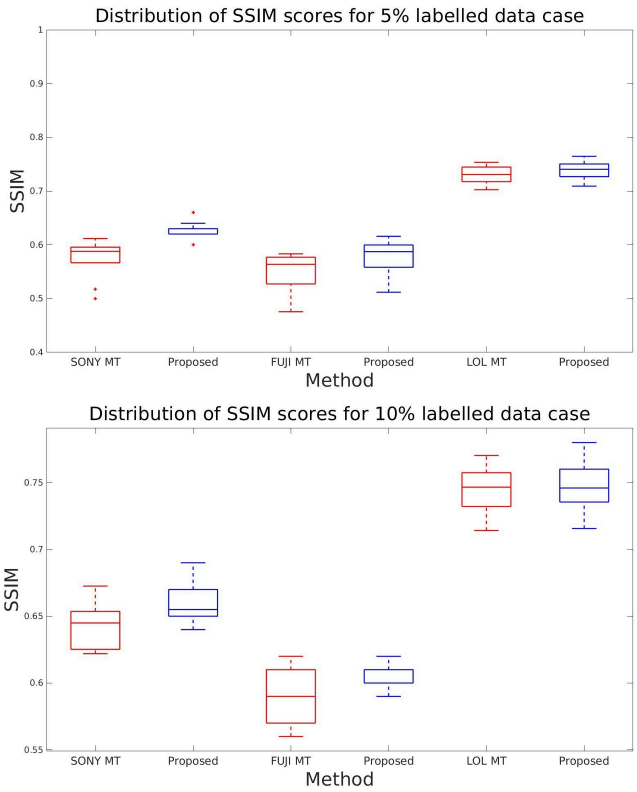

Figure 5. Box plot for distribution of SSIM scores of Mean Teacher (MT) and proposed method across 10 splits.

4.6. Ensemble Analysis Results

While in the paper we showed the results for SONY dataset, we show the results for FUJI and LOL datasets in Figure 9.

5. Limitations

Following are some of the limitations that the proposed semi-supervised learning approach for LLIR suffers from.

1. Note that the method depends on the presence of diverse distortions in the labeled dataset. If no labeled datapoint has distortions similar to an unlabeled subband, this method may not produce a good quality pseudo-label.

2. While we have significant improvements in performance compared to the baselines, there is still a significant performance gap compared to the 100% data case. Nevertheless, this is the first attempt at the problem of semi-supervised learning for LLIR and should initiate further research in this important area.

3. We observe that the performance improvement of our method with respect to the baseline tends to saturate

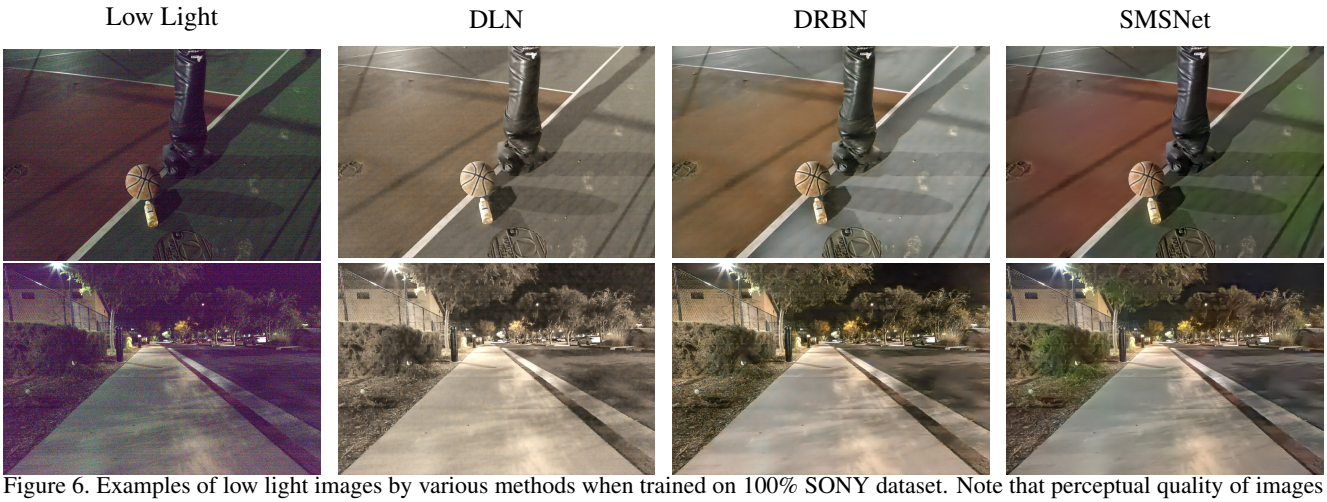

Figure 6. Examples of low light images by various methods when trained on 100% SONY dataset. Note that perceptual quality of images restored by SMSNet is better than others.

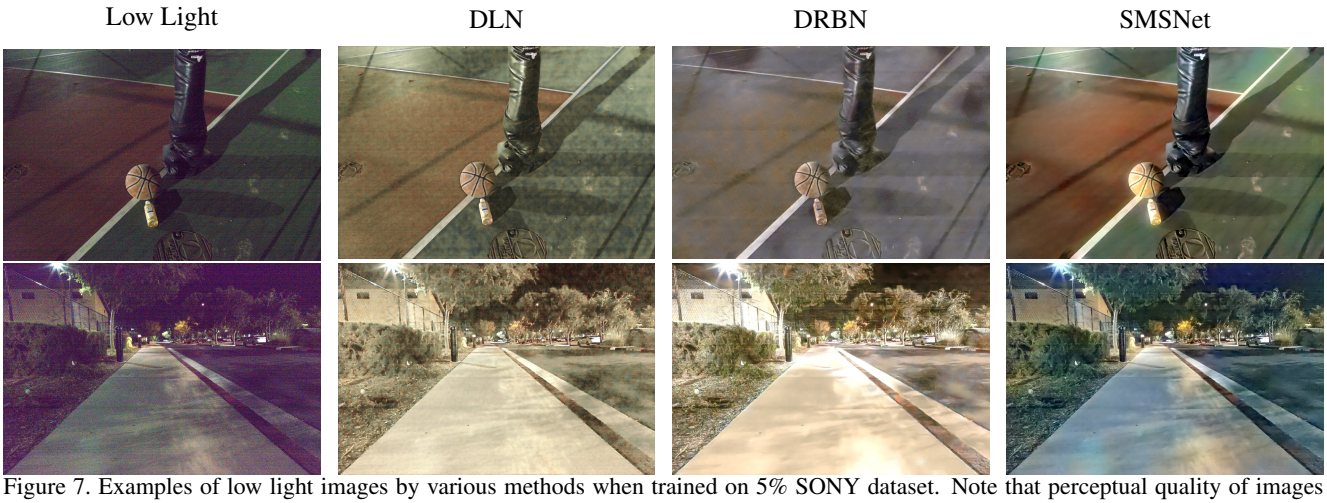

Figure 7. Examples of low light images by various methods when trained on 5% SONY dataset. Note that perceptual quality of images restored by SMSNet is better than others.

with increasing number of labelled examples. However significantly increasing the number of labelled examples poses challenges in the number of models in the ensemble. Thus, the proposed method is most useful when only few labelled data samples are available.

References

[1] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

[2] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.

[3] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780*, 2017.

[4] Li-Wen Wang, Zhi-Song Liu, Wan-Chi Siu, and Daniel PK Lun. Lightening network for low-light image enhancement. *IEEE Transactions on Image Processing*, 29:7984–7996, 2020.

[5] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3063–3072, 2020.

[6] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021.

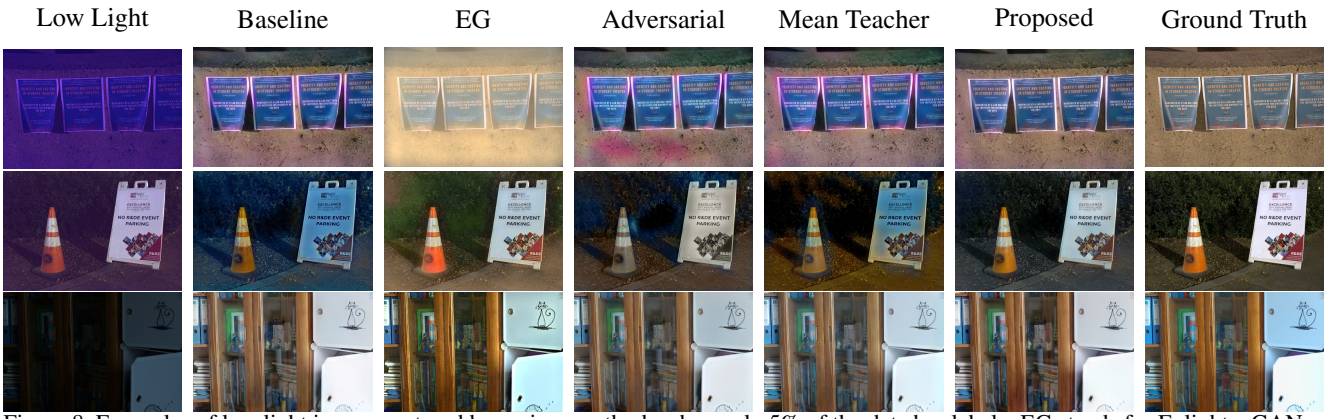

Figure 8. Examples of low light images restored by various methods when only 5% of the data has labels. EG stands for EnlightenGAN.

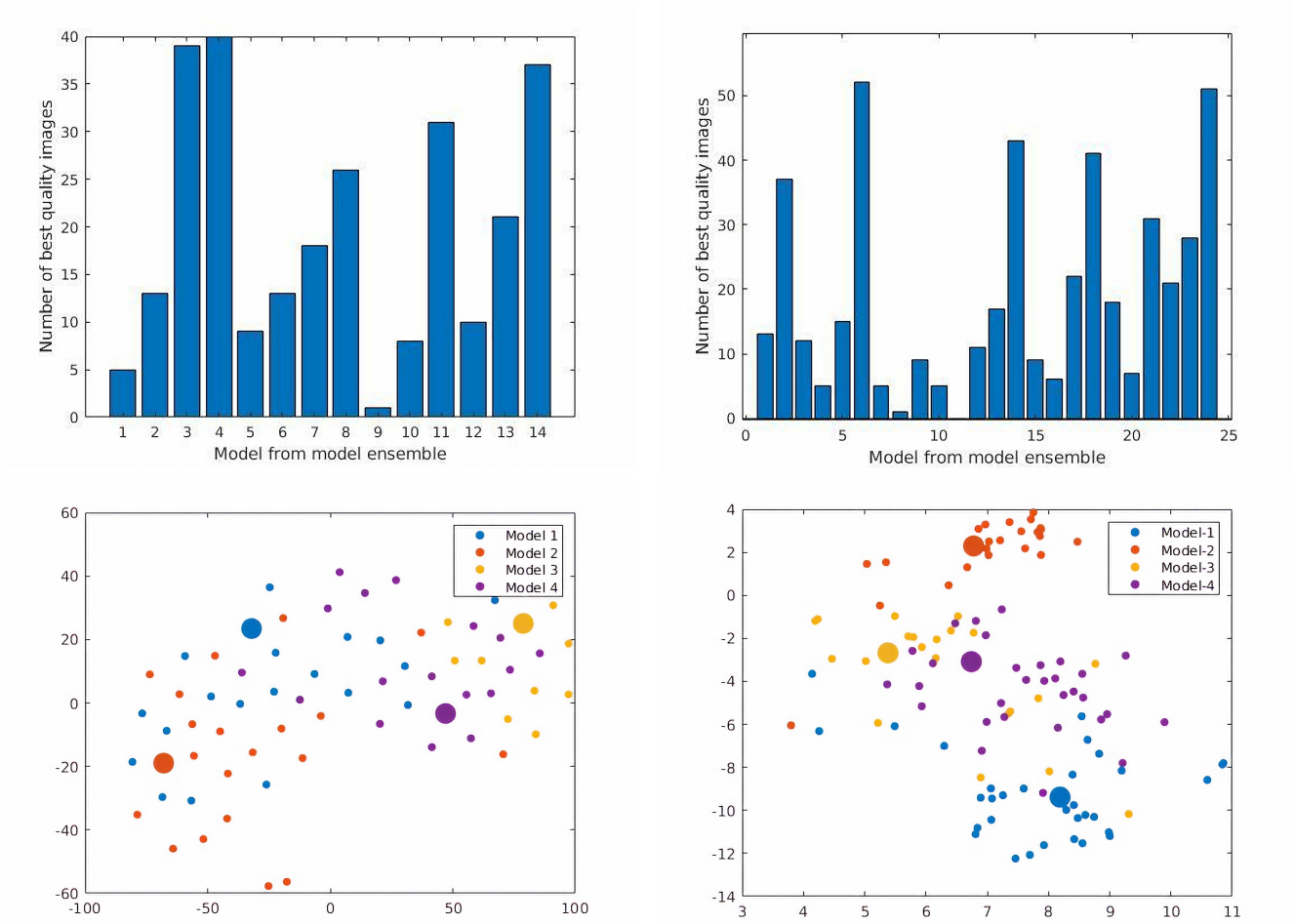

Figure 9. Top row shows the bar plots showing how frequently models from ensemble produces the best quality image according to SSIM index. Bottom row has the scatter plots of features obtained from applying t-SNE to learnt features. Plots are for one of the splits from FUJI (left) and LOL (right) datasets.