Semi-Supervised Learning for Low-light Image Restoration through Quality Assisted Pseudo-Labeling

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

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

Abstract

Convolutional neural networks have been successful in restoring images captured under poor illumination conditions. Nevertheless, such approaches require a large number of paired low-light and ground truth images for training. Thus, we study the problem of semi-supervised learning for low-light image restoration when limited low-light images have ground truth labels. Our main contributions in this work are twofold. We first deploy an ensemble of low-light restoration networks to restore the unlabeled images and generate a set of potential pseudo-labels. We model the contrast distortions in the labeled set to generate different sets of training data and create the ensemble of networks. We then design a contrastive self-supervised learning based image quality measure to obtain the pseudo-label among the images restored by the ensemble. We show that training the restoration network with the pseudo-labels allows us to achieve excellent restoration performance even with very few labeled pairs. We conduct extensive experiments on three popular low-light image restoration datasets to show the superior performance of our semi-supervised low-light image restoration compared to other approaches. Project page is available at https://github.com/sameerIISc/SSL-LLR.

1. Introduction

Restoring low-light images is challenging because of the presence of multiple distortions such as poor contrast, noise and color cast. While several traditional low-light image restoration (LLIR) methods exist, convolutional neural networks (CNNs) have tremendously succeeded in such a complex restoration task. However, CNNs suffer from the shortcoming of requiring large amounts of aligned low-light and clean well-exposed image pairs for training. Moreover, such models are camera distortion specific and require data to be collected everytime a new model needs to be trained for a different camera.

While a few research articles have studied training LLIR models when such labeled data is not available, [6, 10], whether a small number of labeled examples would also be useful for designing better models for LLIR has not been explored. In this context, the role of semi-supervised learning (SSL) for LLIR is interesting. Here, the goal is to collect a few pairs of labeled images, and use a large corpus of unlabeled images to enrich the training.

While SSL has been extensively studied for classification tasks, only few works explore this learning paradigm for image restoration. Most of the existing SSL image restoration approaches are based on the use of distortion models specific to the restoration tasks such as single image dehazing [15] and deraining [9, 37]. These approaches are not directly applicable to semi-supervised LLIR.

In this work, we approach SSL for LLIR by generating pseudo-labels for the unlabeled data. To generate pseudo-labels, we first create several enhanced versions of the given unlabeled image through an ensemble of restoration models. We then choose the best image according to a learnt quality assessor to serve as a pseudo-label for the unlabeled image. While there exists pseudo-labeling and ensembling approaches for SSL in the image classification literature [13, 27], they are not directly applicable to the LLIR problem. In particular, the design of ensembling and pseudo-label generation approaches for LLIR are non-trivial. To this end, our main contribution and novelty lies in the design of ensemble models and the approach to generate pseudo-labels from the ensemble model outputs for LLIR.

We design the ensemble of LLIR models, by creating a set of expert models, each model expert at restoring different distortions in the labelled set. We do this by training a model exclusively on the distortion corresponding to each low-light labelled image. We first learn distortion models that relate the pairs of low-light and well-exposed images in the available labeled data. We then use these distortion models to generate more distorted versions for each of the well-exposed images available in the labeled set to train a

restoration network for each distortion model. We apply the ensemble of restoration networks to each unlabeled image to generate its several restored versions. Given an unlabeled image, the models which are expert at restoring distortions similar to that of the given unlabeled image are expected to produce good quality images.

To choose a good quality image from the set of restored versions produced by the ensemble models, we design a quality assessment model for low-light restored images. We note that there is no reference image available for selecting the best pseudo-label. Popular no reference (NR) quality assessment (QA) models [11, 22, 40, 42] cannot be used off the shelf since the distortions that occur during LLIR are quite different and vary from camera to camera. Also, there are no human labels to train such quality models for LLIR. Existing unsupervised NR QA models [23, 41] also do not effectively capture LLIR distortions. Thus, our second main contribution is in the design of a self-supervised quality measurement tool that can select the best pseudo-label for semi-supervised learning of image restoration.

We design an NR QA model for low-light restored images based on self-supervised contrastive learning. In particular, we use the large corpus of unlabeled images and learn features from multiple restored versions of these unlabeled images. After learning such self-supervised features, we compute the similarity of these features with such features extracted from well-exposed images to obtain a quality score. Our quality index based on this score is used to select the pseudo-label and train the restoration network along with the few labeled pairs. Although there has been some recent work on contrastive learning for self-supervised feature learning of quality features [18, 19], these have been primarily used in a supervised setup to measure generic quality. Further, their ability to measure the quality of low-light restored images has not been explored. We present a novel application of an unsupervised quality measure for pseudo-labeling in semi-supervised LLIR.

We conduct our experiments on a simple multiscale LLIR architecture owing to the success of such approaches in literature [14,38]. Our simple multi-scale restoration network explicitly learns the subbands of a Laplacian pyramid decomposition. We observe that the CNN based models for restoring bandpass subbands perform quite well even when trained with very few labels. However, the performance of lowpass subbands offers a lot of scope for improvement. Thus, we focus on the semi-supervised learning of the lowpass subband in our work. We show through extensive experiments on three publicly available low-light image datasets that our pseudo-label selection approach can yield superior performance when compared to other competing approaches. The main contributions of our work are as follows:

- We create an ensemble of LLIR networks designed to address various distortions in lowpass subbands by generating data through a low-light distortion model.
- We design a contrastive self-supervised feature learning approach for predicting the quality of the restored unlabeled data to determine the pseudo-labels.
- We show through extensive experiments on three datasets that the pseudo-labels generated can lead to effective training of a multi-scale LLIR network resulting in superior performances.

2. Related Work

We discuss the related work in the areas of LLIR, semi-supervised learning and unsupervised quality assessment.

Low-light Image Restoration: We mainly discuss learning based approaches here since our goal is semi-supervised learning. Existing LLIR approaches consist of retinex model approaches [33, 36, 43, 44], multi-scale subband learning methods [14, 20, 38], residual learning [32] and end-to-end learning methods [16, 17, 31]. Retinex model based methods disentangle the illuminance adjustment aspect of the LLIR from the denoising problem by decomposing the low-light image into illuminance and reflectance components. Multi-scale subband learning methods are also successful for LLIR due to the sparse nature of the bandpass subbands. Further, the smoothness of the lowpass subband makes it easier to learn the restored image due to the reduced noise. End-to-end learning approaches include methods which successively refine the restored low-light image over several layers [17, 31, 32]. Nevertheless, CNN based methods require lots of labeled data and there is a need to study their design with limited labeled data.

Semi-Supervised Learning for Image Restoration: Many of the SSL approaches to image restoration involve imposing a prior on the distortions estimated on the unlabeled data. The SSL approach to image deraining in [37] imposes a Gaussian mixture model prior on the estimated residual for the unlabeled data. Another SSL image dehazing method [15] imposes a sparsity prior on the dark channel prior of the dehazed unlabeled images. Yasarla *et al*. generate pseudo-labels for a latent representation by projecting the latent vector for unlabelled data onto a latent space model [39]. In another work [9], a teacher model with an exponential moving average of the student model is used to generate pseudo-labels for the unlabeled data. Consistency regularization [24], a popular SSL approach, has been used for SSL in image denoising, super-resolution and image coloring. In most of the above works, SSL is achieved by very task specific distortion modeling and their extension to the LLIR problem is not straightforward.

Unsupervised Quality Assessment: While there exists a plethora of supervised NR image QA methods, very few unsupervised NR QA methods are successful. NIQE [23] and IL-NIQE [41] represent a few examples of unsupervised NR

QA methods. These methods are primarily designed based on natural scene statistics (NSS) and a comparison of these features with a corpus of pristine images. However, these NSS based features may not accurately capture all the distortions that can arise during LLIR. In particular, the NSS based features often operate only in the bandpass domain. Thus, they could fail to capture distortions such as color casts and under or over enhancement which cause variations in brightness levels. While there is recent work on contrastive learning of quality features [3, 18], the quality prediction is still supervised and needs to be trained on human scores.

3. Analysis of a Multi-scale Architecture for Low-Light Image Restoration

Before discussing our method for semi-supervised learning of LLIR, we first briefly discuss a simple multi-scale LLIR architecture that lends itself well for semi-supervised learning. We then describe our semi-supervised learning approach on top of it.

While there exist several approaches to LLIR, multi-scale subband learning approaches have been more successful [14, 38], [20] for several reasons. Firstly, bandpass subbands have well behaved statistics in the form of sparse coefficients, making it easier to learn to restore these subbands. Secondly, the higher signal to noise ratio in the lowpass subbands allows for effective contrast enhancement. Finally, explicitly learning multi-scale subbands separately constrains the restored image to match the ground truth image in each subband and leads to effective training.

In this work, we adopt a simple yet effective multi-scale subband learning approach Simple Multi-Scale restoration Network (SMSNet). In SMSNet, we learn to predict the Laplacian pyramid subbands of the ground truth image using a CNN model for each of the subbands as shown in Figure 1. Bandpass CNN restores the bandpass subbands and consists of a sequence of convolutional layers with residual connections. Lowpass CNN restores the lowpass subband and further consists of instance normalization layers for effective low-light enhancement. We find that this simple architecture achieves a more robust performance across different datasets. Please refer to the supplementary for more details about the architecture, training of SMSNet and comparisons against other popular LLIR methods. On account of the superior performance of SMSNet, we now analyze this model to leverage it effectively for semi-supervised learning.

Analysis of SMSNet: We now analyze the performance of SMSNet when very limited data is available for training. We conduct experiments on datasets described in Section 5. In Table 1, we evaluate the performance when SMSNet is trained on 5% of the data. We note a significant drop in the performance compared to the model trained on 100% of the data. We also present the result of an interesting combination wherein Bandpass CNN is trained only with 5% data and Lowpass CNN is trained on all the data. We note that such a model performs almost as well as the model trained on all the data. This shows that Bandpass CNNs can generalize well even when trained on significantly lesser data. We also report the combination where Bandpass CNN is trained on 100% data and Lowpass CNN is trained on 5% data. The poor performance of this model implies that most of the drop in performance of the model trained on 5% data can be attributed to the Lowpass CNN not generalizing well. The use of SMSNet simplifies the semi-supervised learning of LLIR owing to the good performance of Bandpass CNNs. Thus, in the rest of this paper, we focus on semi-supervised learning for Lowpass CNN only and retain the Bandpass CNNs trained on 5% data.

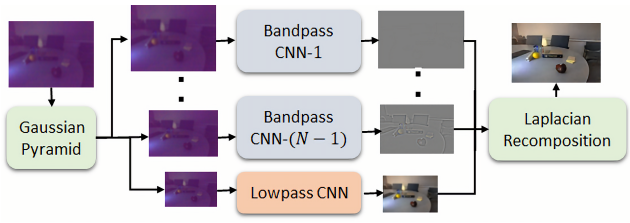

Figure 1. Architecture of SMSNet.

Table 1. Evaluation of Lowpass (LP) and Bandpass (BP) CNN of SMSNet with 5% training data using SSIM/PSNR.

Training Data BP	Training Data LP	SONY	FUJI	LOL
5%	5%	0.58/19.73	0.54/18.65	0.71/19.04
100%	100%	0.74/23.05	0.69/22.63	0.78/21.91
5%	100%	0.72/22.72	0.67/22.13	0.77/21.72
100%	5%	0.59/19.92	0.56/18.94	0.72/19.20

4. Semi-Supervised Learning Approach

We now describe our approach to training Lowpass CNN when only a small fraction of the low-light images have the corresponding ground truth available. Let $\mathbf{x}$ and $\mathbf{y}$ denote the lowpass subbands of well-exposed and low-light image pairs. Let D_L denote the labeled dataset consisting of N well-exposed and low-light lowpass subband pairs $\{(\mathbf{x}_1, \mathbf{y}_1), ..., (\mathbf{x}_N, \mathbf{y}_N)\}$. Also, let D_U denote the unlabeled data with M lowpass subbands $\{\mathbf{y}_{N+1}, ..., \mathbf{y}_{N+M}\}$. Let $\mathbf{x}^{fr}$, $\mathbf{y}^{fr}$, D_L^{fr} and D_U^{fr} denote the corresponding full-resolution versions. The goal of semi-supervised learning for restoration of low-light lowpass subbands is to effectively use the available unlabeled data D_U in addition to the labeled data D_L. We address this by generating pseudo labels $\{\hat{\mathbf{x}}_{N+1}, ..., \hat{\mathbf{x}}_{N+M}\}$ for the unlabeled low-light lowpass subbands and train with these labels. In the following

Figure 2. Comparison of lowpass subbands obtained by restoring a low-light lowpass subband by the model trained on 5% data and three of the models from the learnt ensemble of models. The three models are not chosen in any particular order. For visualization, we recompose the restored lowpass subbands with the corresponding restored bandpass subbands to get the full resolution images. Note that at least one of the three images obtained from the ensemble is perceptually better than the image from the 5% data model.

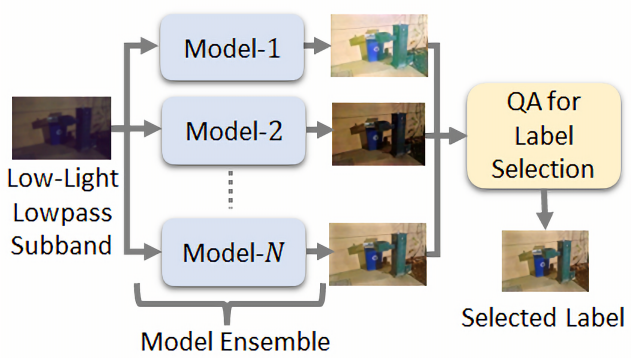

Figure 3. Block diagram for our pseudo label generation method.

discussion, we refer to lowpass subbands as just subbands

Our method to generate pseudo-labels involves two steps as shown in Figure 3. First, we generate a set of potential pseudo-labels for the low-light subbands through a variety of restoration models. The resulting restored subbands span a wide range of quality. In the second step, we select a good quality subband from the multiple restored versions to serve as a pseudo-label for the distorted subband. We then train Lowpass CNN from scratch using the labeled low-light and ground truth subband pairs along with the unlabeled low-light subbands with their pseudo-labels. Note that in our approach the pseudo-labels are generated just once before training Lowpass CNN from scratch.

4.1. Training model ensemble

We adopt an ensemble approach of training multiple restoration networks to generate potential pseudo-labels for each unlabeled low-light subband. Our ensemble should be able to generate subbands with good quality so that one of these can serve as a pseudo-label. We design the ensemble of models as follows. We observe that low-light subbands in the labeled set D_L can have different contrast distortions such as poor lighting, poor contrast and color casts corresponding to different labeled pairs. We generate multiple expert models for restoration by training a model tuned to different contrast distortions in D_L. A model trained on a contrast distortion is expected to produce good quality images when it encounters similar contrast distortions. Further, we believe that the inclusion of Instance Normalization layers in Lowpass CNN of our SMSNet makes our ensemble robust and relevant to the contrast distortions in the unlabeled data which may be slightly different from the labeled data. This is because the normalization layers may normalize for some of the variations in contrast distortions across different images. Thus, due to these normalizations, the network may perceive even slightly different contrast distortions as being similar.

Since we have only one labeled pair corresponding to each distortion in D_L, we create a training dataset having multiple subband pairs with the corresponding contrast distortion as follows. We estimate a contrast distortion model for each pair of subbands in D_L and generate the training dataset by using this model on the well-exposed subbands available in D_L. Thus, we generate N training datasets to train N restoration models. To generate the training datasets, we employ the contrast distortion model proposed in [21] for full resolution images. Thus, we model the distortions in the full resolution image and then downsample them to obtain the distorted lowpass subbands.

Given a pair of low-light and well-exposed images $\mathbf{y}^{fr}$ and $\mathbf{x}^{fr}$, the low-light image is written as [21]

$$\mathbf{y}^{fr} = f(\mathbf{x}^{fr}) + \mathbf{w}, \tag{1}$$

where $f(.)$ is a global point-wise function that models the contrast distortion and $\mathbf{w}$ represents the noise. $f(.)$ in turn is modeled as a polynomial function with coefficients as its parameters. The parameters, denoted by θ_f, are estimated by minimizing the mean squared error between $\mathbf{y}^{fr}$ and $f(\mathbf{x}^{fr})$. We apply this contrast distortion function f along with simulated noise to other well-exposed images. Note that given the complex nature of noise in RGB images, often GANs are used for noise modeling [2,12]. However, the use of GAN here is challenging due to the lack of sufficient data. Further, due to aggressive smoothing over several scales, lowpass subbands will not have significant amounts of noise [25, 26]. Due to these reasons, we dispense with very accurate noise modeling. Specifically, we just add zero mean white Gaussian noise with standard de-

viation equal to that of the error between $\mathbf{y}^{fr}$ and $f(\mathbf{x}^{fr})$, to the generated low-light images $f(\mathbf{x}^{fr})$ and obtain the lowpass subbands from these to create the training datasets.

Mathematically, let the contrast distortion functions estimated from each of the N labeled pairs of images be $f_1, f_2, ...f_N$. Let $n \in \{1, 2, \ldots, N\}$ and $D_L^{fr,n}$ be the n^{th} training dataset created using $f_n(\cdot)$ and represented as

$$D_L^{fr,n} = \{(\mathbf{x}_1^{fr}, f_n(\mathbf{x}_1^{fr}) + \mathbf{w}_1), ..., (\mathbf{x}_N^{fr}, f_n(\mathbf{x}_N^{fr}) + \mathbf{w}_N)\}. \tag{2}$$

Now, we train Lowpass CNN h_n using the lowpass subband pairs of D_L^n obtained from $D_L^{fr,n}$ to create an ensemble of models $\{h_n\}_{n=1}^N$. We then apply these models on a low-light subband $\mathbf{y} \in D_U$ to generate a set of potential pseudo-labels $\{h_1(\mathbf{y}), ..., h_N(\mathbf{y})\}$. In Figure 2, we show a low-light image and some of its restored versions produced using the ensemble.

4.2. Quality assessment for pseudo-label selection

Selecting a good quality subband from multiple restored versions of a low-light subband is quite challenging due to the lack of ground truth. Pre-trained NR QA algorithms may not capture the distortions that arise during LLIR. Since there are no labels for training a QA algorithm, we adopt a self-supervised learning approach to design a quality index. We design our QA method by training a CNN model, referred to as Quality Feature CNN (or QFCNN), to extract quality relevant features from any restored lowpass subband. The goal is to design features that can capture various distortions that arise during low-light subband restoration. We then compute the quality of a restored subband by computing the cosine similarity between the features of the restored subband and the mean of the features of the well-exposed subbands. Note that well-exposed subbands are not corresponding pairs of the restored subbands, since the restored subbands are obtained for the unlabeled data.

We adopt a contrastive learning approach popularly used for self-supervised feature learning [4, 5, 8, 29] in image classification. In contrastive learning, for a given image, its positive and negative augmentations are generated. The contrastive loss then, increases the similarity between features of the positive pairs and decreases the similarity between features of the negative pairs. In order to enable the QFCNN to learn distortion aware features, we design the positive pairs with similar distortions and negative pairs with different distortions.

To generate such positive and negative pairs, we use the multiple restored versions of a low-light lowpass subband produced by the ensemble. Note that these multiple restored versions have the same content and only differ in distortions. Specifically, at a given training iteration, we construct a batch using the patches extracted from the multiple restored versions. To extract the patches, we split each of these restored subbands into four quadrants resulting in four non-overlapping patches for each of the subbands. Since, lowpass subbands mainly consist of global distortions such as over and under enhancement and color casts, these distortions do not vary much across the large sized patches from the same subband. Thus, for a given anchor patch from one of the restored versions, we take the patches from the same version as positive pairs while the patches from other restored versions as negative pairs.

More formally, let $\mathcal{V}_n$ for $n \in \{1, \cdots, N\}$ denote the set of all 4 non-overlapping patches from the n^{th} restored version of an unlabelled low-light lowpass subband. Our batch consists of $4N$ patches extracted from the N restored versions of a given low-light subband. Now, a pair of patches p and q is similar if $p, q \in \mathcal{V}_n$ for any $n \in \{1, 2, \cdots, N\}$. If the patches belong to different lowpass subbands, they are dissimilar. Let $\mathbf{z}_p$ denote the normalized (to unit norm) features obtained from patch p. The contrastive loss for p and q where $p, q \in \mathcal{V}_n$ is given by

$$L(\mathcal{V}_n) = -\log \frac{s(p,q)}{s(p,q) + \sum_{n'=1}^{N} \mathbb{I}_{[n' \neq n]} \sum_{q' \in V_{n'}} s(p,q')}, \tag{3}$$

where, $\mathbb{I}_{[n' \neq n]}$ is equal to 1 when $n' \neq n$ and 0 otherwise and $s(p,q) = \exp(\mathbf{z}_p^T \mathbf{z}_q / \tau)$. Here τ is the temperature parameter which we set to 0.07. The overall loss in a batch is given by $L = \sum_{n=1}^{N} L(\mathcal{V}_n)$.

Note that since QFCNN is fully convolutional, although it is trained on subband patches, we apply it on the entire subband during test time to finally evaluate the quality. Let the normalized features extracted by the trained QFCNN from a restored subband $h_n(\mathbf{y})$, $n = 1, 2, \ldots, N$, $\mathbf{y} \in D_U$ be $\mathbf{z}_n(\mathbf{y})$. Let the features of the well exposed subbands in the dataset $\mathbf{x}_n$, $n = 1, 2, \ldots, N$, be denoted as $\mathbf{z}(\mathbf{x}_n)$. Let $\hat{\mathbf{z}}$ denote the average of the features of $\mathbf{x_n}$ obtained as $\hat{\mathbf{z}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{z}(\mathbf{x}_n)$. The quality of $h_n(\mathbf{y})$ is obtained as

$$q(h_n(\mathbf{y})) = (\mathbf{z}_n(\mathbf{y}))^T \hat{\mathbf{z}}. \tag{4}$$

While the above approach can help assess global contrast and color distortions, we also employ a reference based approximate measure to capture local distortions. Let h_0 be a CNN model trained to restore subbands using the entire limited labeled data D_L. We observe that this model provides a reasonably restored subband $h_0(\mathbf{y})$ which we observe to be mostly devoid of local artifacts but could suffer from global contrast distortions. Although this may be imperfect, we compare the ensemble of restored subbands $h_n(\mathbf{y})$ with $h_0(\mathbf{y})$ through the structural similarity (SSIM) index [35] to estimate local artifacts. We observe that such a model when combined with our quality index in Equation (4) to estimate global contrast distortions is effective. Our combined quality model for selecting the pseudo-label

is given by

$$Q(h_n(\mathbf{y})) = q(h_n(\mathbf{y})) + \text{SSIM}(h_n(\mathbf{y}), h_0(\mathbf{y})). \quad (5)$$

We select the pseudo-label for $\mathbf{y}$ as $h_{n^*}(\mathbf{y})$, where

$$n^* = \underset{n=1,2,..,N}{\arg\max}\ Q(h_n(\mathbf{y})). \quad (6)$$

5. Experiments

Datasets We conduct our experiments on the See-In-the-Dark (SID) [1] and LOw-Light (LOL) [36] datasets. We preprocess the *raw* images in the SID dataset using the python library *rawpy* to get sRGB images. We further downsample them to the resolution of 832×1248.

Experimental Setting On each dataset, we randomly pick 5% and 10% of the ground truth training images along with the corresponding low-light versions as labeled pairs. The rest of the training set is used without ground truth labels as the unlabeled dataset. We report the mean restoration performance in terms of SSIM, peak signal to noise ratio (PSNR) and a colorfulness measure [7] across 10 random choices of the 5% or 10% labeled pairs. We note that low-light restored images often suffer from poorly saturated colors and evaluating the colorfulness of the restored images in conjunction with other metrics is important.

Training Details We use SMSNet with 5 levels including 4 bandpass subbands and a lowpass subband. We use a batch-size of 16. For the 100% training data case, we train the CNN models in SMSNet for 90 epochs and 450 epochs on the SID and LOL datasets respectively. We use Adam optimizer with a learning (LR) of 10^{-3} which is reduced to 10^{-4} and 10^{-5} after 50 and 70 epochs respectively for the SID dataset and 250 and 350 epochs respectively for the LOL dataset. For training Lowpass CNN, we use a patch of size 48 cropped randomly at every iteration. For training the 4 Bandpass CNNs, we use patches of size $256, 128, 64$ and 64 at different levels. For the 5% and 10% labelled data cases, we scale the number of epochs and LR milestones by 20 and 10 respectively. We train QFCNN with a LR of 1e-4 for 11 epochs and choose τ=0.07 in Equation (3). Please refer the supplementary Section 4.2 for a discussion on how we chose these hyperparameters.

Comparison Methods We benchmark our semi-supervised method with the following methods.

1. **Baseline:** Our baseline is SMSNet discussed earlier trained only on the labelled dataset D_L. The unlabelled dataset is not used in anyway for the baseline.
2. **Augmented Data:** We train the model on an augmented labelled data. The labelled dataset D_L is augmented with the low light images from the synthetically generated datasets $\{D_L^1, \ldots, D_L^N\}$.
3. **Mean-Teacher:** We compare with the Mean-Teacher model [28], a popular semi-supervised learning approach adapted for low light restoration. The Lowpass CNN for the student model is trained with the labelled pairs along with the image output by the teacher model as the target for the unlabelled low light images.
4. **Adversarial loss:** Instead of pseudo-labels, we use an adversarial loss to train Lowpass CNN. The discriminator for the adversarial loss is trained to distinguish between ground truth images from the labelled set and the images output by Lowpass CNN.
5. **Transformation Consistency Regularization (TCR):** We compare with a consistency based SSL approach used earlier for low level vision tasks of image denoising and image colorization [24]. We use this approach for training only the Lowpass CNN of SMSNet.
6. **EnlightenGAN:** We train and evaluate an unpaired learning method for low light image restoration [10] by treating the unlabelled images as distorted and the ground truth labels as clean.

Performance Evaluation We present the numerical results of our proposed approach in Table 2 and 3. We see significant improvements of our method over the other methods in terms of SSIM and colorfulness. While PSNR values are roughly similar, we observe that PSNR is very sensitive to small variations in brightness and colors that are otherwise not perceivable [34]. The colorfulness measure has to be interpreted along with the SSIM index. While the output of EnlightenGAN looks colorful, it contains a lot of spatial artifacts leading to its poor SSIM performance. We also report the variance analysis in the supplementary where we note that our method mostly outperforms Mean-Teacher by achieving a higher SSIM score with a smaller variance across the splits. We also show some examples of low-light images enhanced by various methods in Figure 4.

Analysis of Quality Index: We experiment with different variations of the quality index in Equation (5) to pick the pseudo-label. We refer to the use of the SSIM index in Equation (5) as SSIM-0. We consider the unsupervised QA model NIQE [23] in place of our quality terms based on QFCNN. We evaluate these NR QA methods with and without the SSIM-0 terms. We compare these QA models in Table 4 in terms of the restoration performance achieved when trained with pseudo-labels selected according to Equation (6). Further, we also compare the performance of the QA models in terms of the median Spearman Rank Order Correlation Coefficient (SROCC) with the ground truth SSIM of the restored versions of a given low-light image in Table 5. While our QFCNN based model achieves the best performance, we see that both the SSIM-0 term and the NR QA terms are important and neither of these alone achieves the best performance.

In Table 4, we also evaluate the performance achieved when the ground truth SSIM of the ensemble of restored images is used to pick the pseudo-label. This serves as

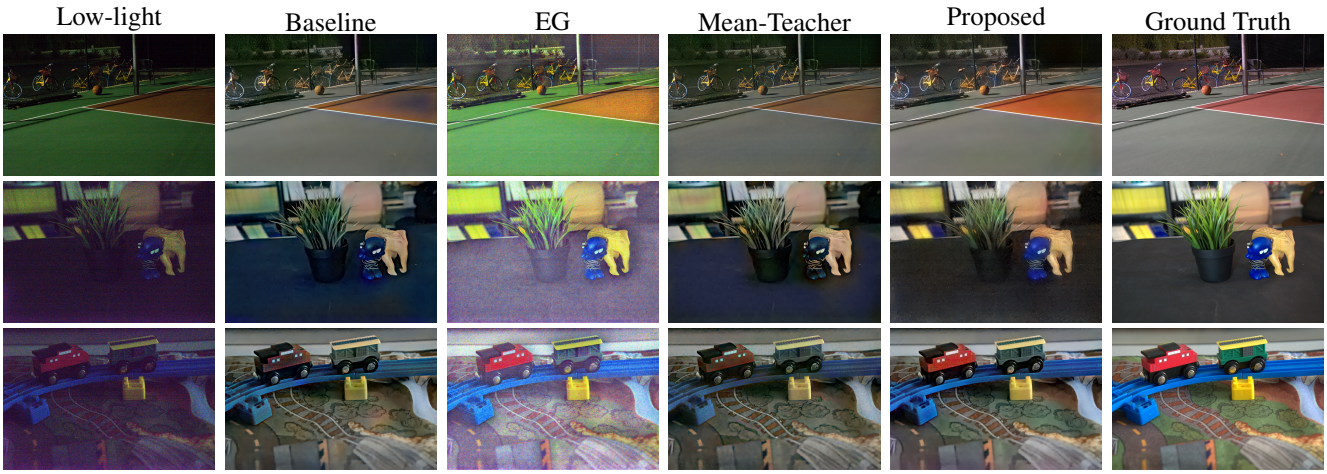

Figure 4. Examples of images enhanced by different methods when only 5% labeled data is available. Note that EG stands for EnlightenGAN. The images enhanced by the proposed method have better perceptual quality than others.

Table 2. Objective evaluation on 5% labeled data case with SSIM, colorfulness [7] and PSNR on multiple datasets. For all metrics higher value is better. Red and blue indicate best and second best scores respectively. 100% data refers to the model trained on all the labeled data. Note that we do not highlight colorfulness results of EnlightenGAN since its SSIM and PSNR scores are very low.

Methods	SONY			FUJI			LOL		
	SSIM	Colorfulness	PSNR	SSIM	Colorfulness	PSNR	SSIM	Colorfulness	PSNR
Baseline	0.58	22.48	19.73	0.54	21.59	**18.65**	0.71	19.02	19.09
Proposed	**0.63**	**24.97**	**19.92**	**0.58**	**22.38**	**18.84**	**0.74**	**21.17**	19.42
Augmented	0.58	21.68	19.72	**0.55**	21.93	18.58	0.72	19.85	18.94
Mean-Teacher	0.57	21.76	**19.75**	**0.55**	21.82	18.63	**0.73**	18.62	**19.51**
TCR	**0.59**	18.49	18.53	0.51	14.60	16.97	0.67	14.50	18.62
Adversarial	0.56	**24.66**	19.21	0.49	**24.09**	17.83	**0.74**	**19.86**	**19.45**
EnlightenGAN	0.31	32.08	15.46	0.33	31.11	13.62	0.57	33.34	16.77
100% Data	**0.74**	**27.42**	**23.05**	**0.69**	**23.68**	**22.63**	**0.78**	**19.31**	**21.91**

an upper bound on the restoration performance that can be achieved using our ensemble of restored images. We see that the QA models approach the performance of the ground truth SSIM based model and the gap is quite small.

Analysis of Model Ensemble: We first show how frequently each of the models in the ensemble produces the best quality image for each of the unlabeled low-light data. We show these results for one of the splits of the SONY dataset in Figure 5. The best quality image is in terms of the SSIM index.

For a given low-light image, since the models trained on similar contrast distortions are expected to produce good quality images, the corresponding contrast distortion parameters are expected to cluster together in the parameter space. Specifically, the parameters of low-light images for which a given model produces the best quality image could cluster together. However, the instance normalization layers may make the network robust to minor variations in some aspects of distortions such as brightness and colorcast. This motivates learning transformed versions of the parameters of the contrast distortion function $f(\cdot)$ that the restoration network perceives as being similar and then analyze the transformed features.

We learn the transformation through a classification network that takes the contrast distortion features θ_f of the unlabelled images as input and is trained on these images to predict the model from the ensemble that produces the best quality image according to the ground truth SSIM index. This network is a single hidden layer fully connected network from which we use the hidden layer features as the transformed features for analysis. Let the hidden layer features for input θ_f be $g(\theta_f)$. Then $g(\theta_f)$ are expected to be more informative to understand which model from the ensemble produces the best quality image for a given low-light image. We visualize these learnt features $g(\theta_f)$ using t-SNE [30] in a two-dimensional space. In Figure 5, we show the scatter plot for one of the splits from the SONY dataset for the top 4 most frequently chosen models. Here large dots represent the hidden layer features of contrast distortion functions on which the models from the ensemble

Table 3. Objective evaluation on 10% labeled data case. Red and blue indicate best and second best scores respectively.

Methods	SONY			FUJI			LOL		
	SSIM	Colorfulness	PSNR	SSIM	Colorfulness	PSNR	SSIM	Colorfulness	PSNR
Baseline	**0.64**	**24.09**	**20.60**	0.58	20.83	19.49	0.73	19.85	**19.59**
Proposed	**0.66**	**26.07**	20.50	**0.60**	**22.36**	19.36	**0.75**	**21.88**	19.48
Augmented	0.58	21.68	19.72	0.55	21.93	18.58	0.72	19.85	18.94
Mean-Teacher	0.63	24.06	**20.61**	**0.59**	20.84	**19.64**	**0.74**	18.88	**19.56**
TCR	0.65	19.96	19.95	0.54	15.32	16.97	0.70	14.24	19.39
Adversarial	0.60	23.35	20.41	**0.59**	**23.95**	**19.55**	**0.74**	**20.15**	19.22
EnlightenGAN	0.31	32.08	15.46	0.33	31.11	13.62	0.57	33.34	16.77

Table 4. SSIM scores for ablation of our QA method by replacing it with other methods.

Methods	SONY	FUJI	LOL
SSIM-0	0.56	0.57	0.66
NIQE	0.50	0.46	0.72
QFCNN	0.61	0.55	0.70
NIQE with SSIM-0	0.61	0.56	0.73
QFCNN with SSIM-0	0.63	0.58	0.74
Ground Truth SSIM	0.66	0.60	0.75

Table 5. SROCC for various QA methods with respect to ground truth SSIM on unlabeled data.

Methods	SONY	FUJI	LOL
NIQE	0.16	0.14	0.24
QFCNN	0.59	0.50	0.61
NIQE with SSIM-0	0.48	0.60	0.60
QFCNN with SSIM-0	0.74	0.75	0.71

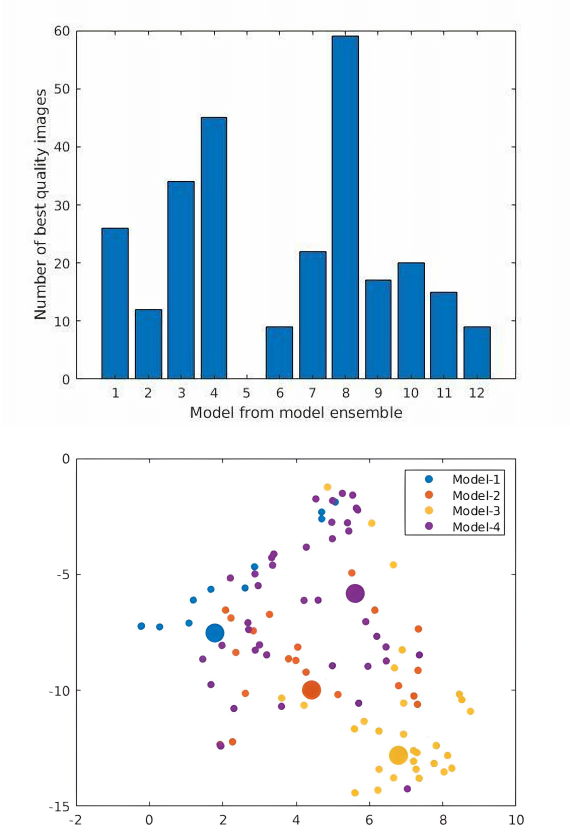

Figure 5. Bar plot shows how frequently models from ensemble produces the best quality image according to the SSIM index. The scatter plot is of features obtained from applying t-SNE to learnt features. Please refer text for more details. Note that both these plots are for one of the splits from the SONY dataset.

were trained on. The smaller dots correspond to the unlabeled image features. We see that distortion features of the unlabeled low-light subbands are often close to the models from the ensemble which produce the best quality restored subband. 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. Please refer to the supplementary for a more detailed discussion about the limitations of the proposed approach.

6. Conclusion

We observed that a multi-scale subband learning architecture for LLIR lends itself naturally for semi-supervised low-light image restoration. While the bandpass subbands generalize quite well although trained on very limited data, our semi-supervised learning of lowpass subbands can improve performance. We showed an ensemble based approach of generating multiple restored images can be used to select pseudo-labels for the unlabeled low-light images. We create an ensemble of restoration networks by training them on different kinds of distorted images we create from the labeled pairs. Further, we also proposed a self-supervised QA measure which when used along with SSIM-0 helps select a pseudo-label for effective training. We showed that our approach achieves superior performance across datasets although trained with limited labeled data.

Acknowledgement: This work was supported in part by a grant from the Department of Science and Technology, Government of India under grant CRG/2020/003516.

References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[2] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2018.

[3] Pengfei Chen, Leida Li, Qingbo Wu, and Jinjian Wu. Spiq: A self-supervised pre-trained model for image quality assessment. *IEEE Signal Processing Letters*, 29:513–517, 2022.

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[5] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020.

[6] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[7] David Hasler and Sabine E Suesstrunk. Measuring colorfulness in natural images. In *Human vision and electronic imaging VIII*, volume 5007, pages 87–95. International Society for Optics and Photonics, 2003.

[8] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[9] Huaibo Huang, Aijing Yu, and Ran He. Memory oriented transfer learning for semi-supervised image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7732–7741, 2021.

[10] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[11] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, 2014.

[12] Dong-Wook Kim, Jae Ryun Chung, and Seung-Won Jung. Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[13] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013.

[14] Jiaqian Li, Juncheng Li, Faming Fang, Fang Li, and Guixu Zhang. Luminance-aware pyramid network for low-light image enhancement. *IEEE Transactions on Multimedia*, pages 1–1, 2020.

[15] Lerenhan Li, Yunlong Dong, Wenqi Ren, Jinshan Pan, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Semi-supervised image dehazing. *IEEE Transactions on Image Processing*, 29:2766–2779, 2019.

[16] Seokjae Lim and Wonjun Kim. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, pages 1–1, 2020.

[17] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, page 220, 2018.

[18] Pavan C Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. Image quality assessment using contrastive learning. *IEEE Transactions on Image Processing*, 31:4149–4161, 2022.

[19] Pavan C. Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C. Bovik. Image quality assessment using synthetic images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 93–102, January 2022.

[20] Sameer Malik and Rajiv Soundararajan. Llrnet: A multiscale subband learning approach for low light image restoration. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 779–783, 2019.

[21] Sameer Malik and Rajiv Soundararajan. A model learning approach for low light image restoration. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1033–1037. IEEE, 2020.

[22] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[23] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[24] Aamir Mustafa and Rafał K Mantiuk. Transformation consistency regularization–a semi-supervised paradigm for image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, pages 599–615. Springer, 2020.

[25] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 12(11):1338–1351, 2003.

[26] Umesh Rajashekar and Eero P. Simoncelli. Chapter 11 - multiscale denoising of photographic images. In Al Bovik, editor, *The Essential Guide to Image Processing*, pages 241–261. Academic Press, Boston, 2009.

[27] Laine Samuli and Aila Timo. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations (ICLR)*, volume 4, page 6, 2017.

[28] 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.

[29] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 776–794. Springer, 2020.

[30] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[31] Thang Vu, Cao Van Nguyen, Trung X Pham, Tung M Luu, and Chang D Yoo. Fast and efficient image quality enhancement via desubpixel convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[32] 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.

[33] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019.

[34] Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

[35] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[36] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[37] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3877–3886, 2019.

[38] 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.

[39] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Semi-supervised image deraining using gaussian processes. *IEEE Transactions on Image Processing*, 30:6570–6582, 2021.

[40] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *2012 IEEE conference on computer vision and pattern recognition*, pages 1098–1105. IEEE, 2012.

[41] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015.

[42] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):36–47, 2018.

[43] 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.

[44] Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang. Retinexdip: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2021.