Classification of Low Light Vehicles Using Transfer Learning and CNN

R.Divyakanthi[1], G.Navyasrivenu [2], D.Udayasrisailakshmi[3] &G.Shanmukhsai[4]

[1] *Assistant Professor*
[2-4] *B. Tech Student, LIET*
[1,2,3,4] *Electronics &* Communication *Engineering, Lendi Institute of Engineering and Technology, Vizianagaram*

---***---

Abstract – The main Aim of the project is to enhance the low light images and extract features using the Transfer learning approach of Convolution Neural Network (CNN) integrating with enhancement technique for increasing the vehicle recognition rate. The images are enhanced using Low Light Image Enhancement (LIME) Method CNN is used for Classification of comp vehicle dataset to evaluate the proposed Model. This paper focuses on the classification of vehicles using Convolutional Neural Network (CNN) which is a class of deep learning neural network. This work makes use of transfer learning using the pre-trained networks to extract powerful and informative features and apply that to the classification task. In the proposed method, the pre-trained networks are trained on two vehicle datasets consisting of real-time images. The classifier performance along with the performance metrics such as accuracy, precision, false discovery rate, recall rate, and false negative rate is estimated for the following pre-trained networks Google Net. The classification model is implemented on the standard vehicle dataset and also on a created dataset. The model is further used for the detection of the different vehicles using Regions with a Convolutional Neural Networks (RCNN) object detector on a smaller dataset. This paper focuses on finding the perfect network suitable for the classification problems which have only a limited amount of non-labelled data. The model makes use of limited pre-processing and achieves greater accuracy on continuous is training of the networks on the vehicle images.

Key Words: Convolutional Neural Network, low light Enhancement, Google Net [DAG Network with properties], Matlab, Image Data Store, Deep learning tool box, precision, Accuracy, Validation, pre-trained networks, dataset.

1.INTRODUCTION

Image content is exponentially growing due to the ubiquitous presence of cameras on various devices. During image acquisition, degradations of different severity are often introduced. It is either because of the physical limitations of cameras or due to inappropriate lighting conditions. For instance, smartphone cameras come with a narrow aperture and have small sensors with limited dynamic range. Consequently, they frequently generate noisy and low contrast images. Similarly, images captured under the unsuitable lighting are either too dark or too bright. The art of recovering the original clean image from its corrupted measurements is studied under the image restoration task. It is an ill-posed inverse problem, due to the existence of many possible solutions. Recently, deep learning models have made significant advancements for image restoration and enhancement, as they can learn strong (generalizable) priors from large-scale datasets. Existing CNNs typically follow one of the two architecture designs: 1) an encoder-decoder, or 2) high-resolution (single-scale) feature processing. The encoder-decoder models first progressively map the input to a low-resolution representation, and then apply a gradual reverse mapping to the original resolution. Although these approaches learn a broad context by spatial-resolution reduction, on the downside, the fine spatial details are lost, making it extremely hard to recover them in the later stages. On the other side, the high-resolution (single-scale) networks do not employ any down sampling operation, and thereby produce images with spatially more accurate details. However, these networks are less only effective in encoding contextual information due to their limited receptive field. Image restoration is a position-sensitive procedure, where pixel-to-pixel correspondence from the input image to the output image is needed. Therefore, it is important to remove the undesired degraded image content, while carefully preserving the desired fine spatial details (such as true edges and texture). Such functionality for segregating the degraded content from the true signal can be better incorporated into CNNs with the help of large context, e.g., by enlarging the receptive field. Towards this goal, we develop a new multi-scale approach that maintains the original high-resolution features along the network hierarchy, thus minimizing the loss of precise spatial details. Simultaneously, our model encodes multi-scale context by using parallel convolution streams that process features at lower spatial resolutions. The multi-resolution parallel branches operate in a manner that is complementary to the main high-resolution branch, thereby providing us more precise and contextually enriched feature representations. The main difference between our method and existing multi-scale image processing approaches is the way we aggregate contextual information. First, the existing methods process each scale in isolation, and exchange information only in a top-down manner. In contrast, we progressively fuse information across all the scales at each resolution-level, allowing both top-down images exchange.. Simultaneously, both fine-to-coarse and coarse-to-fine knowledge exchange is laterally performed on each stream by a new selective kernel fusion mechanism. Different from existing methods that employ a simple concatenation or averaging of features coming from multi-resolution branches, our fusion approach dynamically selects the useful dataset from each branch representations using a self-attention approach. More importantly, the proposed fusion block combines features with varying receptive fields, while preserving their distinctive complementary characteristics.

2. EXPLANATION OF TRANSFERLEARNING

Transfer Learning : Transfer learning is a machine learning technique where a model trained on one task is repurposed or fine-tuned for a different, but related task. Instead of starting the learning process from scratch, transfer learning leverages the knowledge gained from solving one problem and applies it to a different but related problem. This is especially useful when the amount of labelled data available for the target task is limited, as it allows the model to leverage the knowledge learned from a larger dataset.

Explicit Handling of Low-Light Conditions: Some models may incorporate explicit mechanisms to handle low-light conditions, such as adaptive exposure control or adjusting internal parameters based on the available light**.**

Ensemble Methods: Combining predictions from multiple models trained with different strategies (e.g., some trained with augmented low-light data, some trained with regular data) can often improve performance, especially in challenging conditions like low light**.**

Dataset Collection: Ensuring that the training dataset includes a diverse range of lighting conditions, including low light, is crucial for the model to learn robust features that generalize well to real-world scenarios**.**

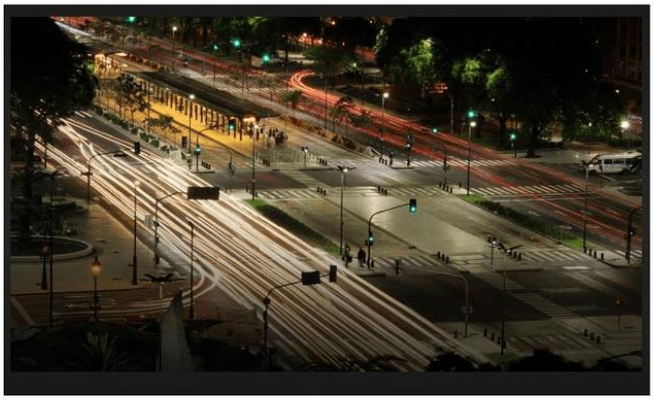


Fig:1 Relevance to low light conditions

3.APPLICATIONS IN IMAGE RECOGNITION

In image recognition tasks, adapting models to perform well under low-light conditions is essential for applications such as surveillance, autonomous driving, and night-time photography. Here's how the strategies mentioned earlier could be applied specifically to image recognition.

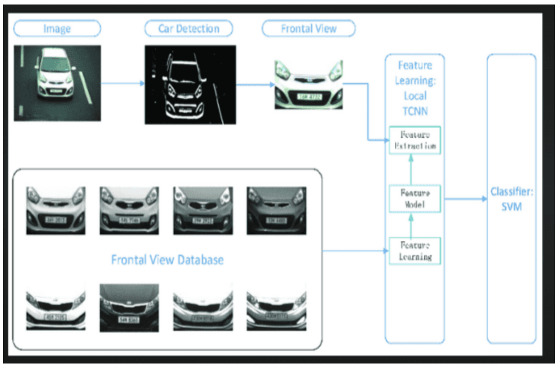


Fig:2 Application Of Image Recognition

3.RELATED WORK

With the rapidly growing image content, there is a pressing need to develop effective image restoration and enhancement algorithms. In this paper, we propose a new method capable of performing image denoising, super-resolution and image enhancement. Unlike existing works for these problems, our approach processes features at the original resolution in order to preserve spatial details, while effectively fuses contextual information from multiple parallel branches. Next, we briefly describe the representative methods for each of the studied problems. Image denoising. Classic denoising methods are mainly based on modifying transform coefficients [115,30,90] or averaging neighbourhood pixels . Although the classical methods perform well, the self-similarity based algorithms, e.g., NLM [10] and BM 3D , demonstrate promising denoising performance. Numerous patch-based algorithms that exploit redundancy in images are later developed . Recently, deep learning based approaches make significant advances in image denoising, yielding favourable results than those of the hand-crafted methods. Super-resolution (SR). Prior to the deep-learning era, numerous SR algorithms have been proposed based on the sampling theory [55,53], edge guided interpolation natural image priors , patch-exemplars and sparse representations . Currently, deep-learning techniques are actively being explored, as they provide dramatically improved results over conventional algorithms. The data-driven SR approaches differ according to their architecture designs. Early methods take a low-resolution (LR) image as input and learn to directly generate its high-resolution (HR) version. In contrast to directly producing a latent HR image, recent SR networks.

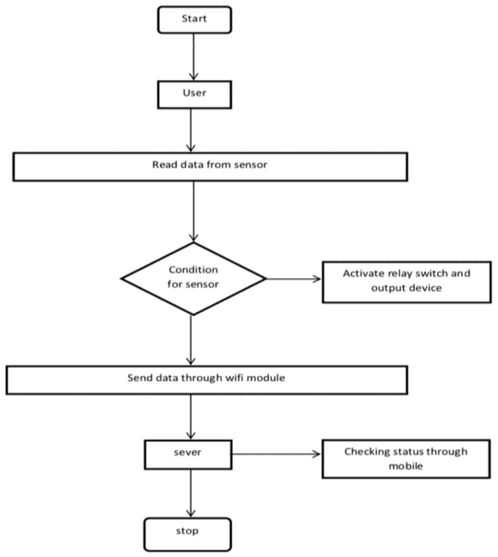


Fig.3:work flow

Steps Involved:

Step 1: Initialize the pre-trained model with weights from a model trained on a large dataset.

Step 2: Replace the final layer(s) of the pre-trained model with new, untrained layers that are specific to the task at hand.

Step 3: Fine-tune the weights of the pre-trained model by training it on the new task, using a smaller dataset.

Step 4: Optionally, repeat step 3 with different learning rates or by unfreezing more layers of the pre trained model.

Use the fine-tuned model for the prediction of new data.

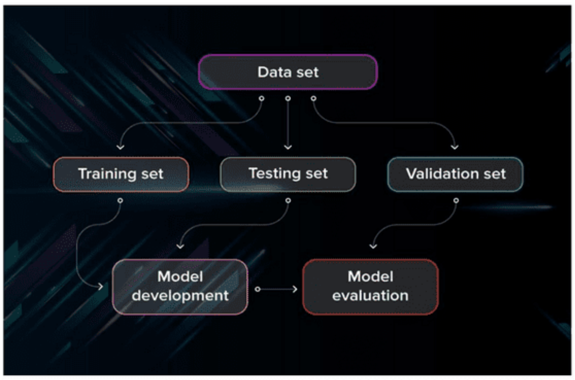


Fig: 4. Training of a data set

4.ARCHITECTURE OF CNN MODELS

Basic Architecture : There are two main parts to a CNN architecture. • A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.

• The network of feature extraction consists of many pairs of convolutional or pooling layers.

• A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.

• This CNN model of feature extraction aims to reduce the number of features present in a dataset. It creates new features which summarises the existing features contained in an original set of features. There are many CNN layers as shown in the CNN architecture diagram.

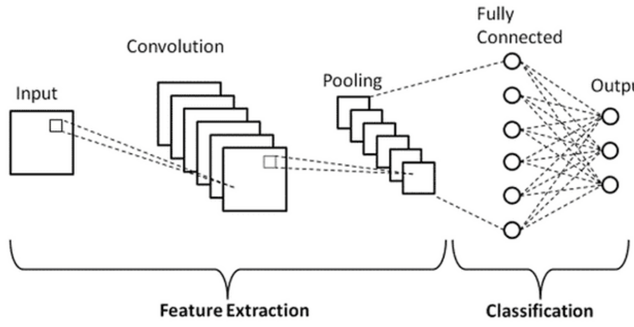


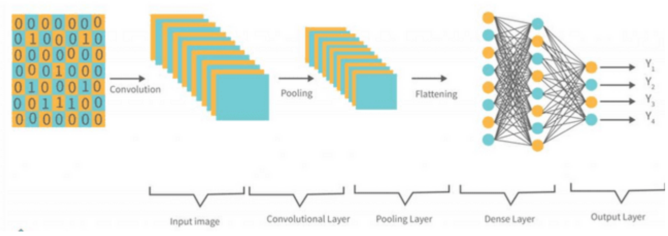
Fig: 5. Architecture Of CNN

1. **Convolutional Layer** :This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size (MXM). By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MXM).

2.

3. 2. **Pooling Layer:** In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several types of Pooling operations. It basically summarises the features generated by a convolution layer.

4. 3. **Fully Connected Layer** : The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.



5. **OUTPUT LAYER** : The from the fully connected layers is then fed into a logistic function for classification tasks like sigmoid or softmax which converts the output of each class into the probability score of each class.

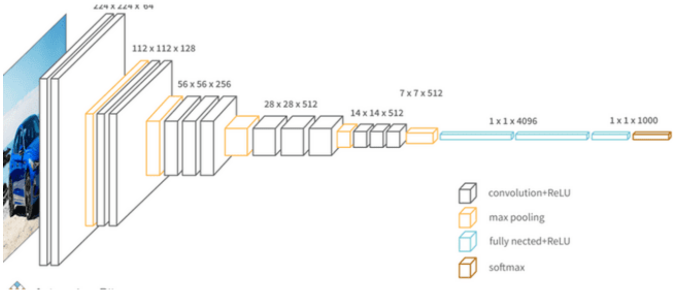


Fig:6 . Output Layer

5.RESULTS AND DISCUSSION:

Data Set Collection and Preprocessing: This involves gathering relevant data for training a deep learning model and preparing it for analysis. Preprocessing steps may include cleaning, filtering, and transforming the data to make it suitable for training.

Data Acquisition: This refers to the process of obtaining raw data from various sources such as sensors, databases, or external files. In MATLAB, this can involve using functions or tools to read and import data into the workspace.

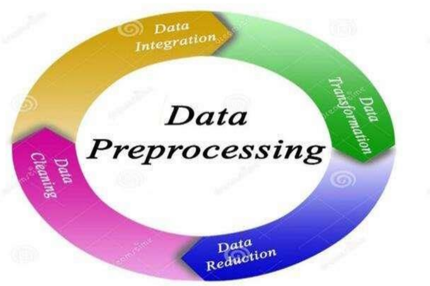


Fig:7.Data preprocessing

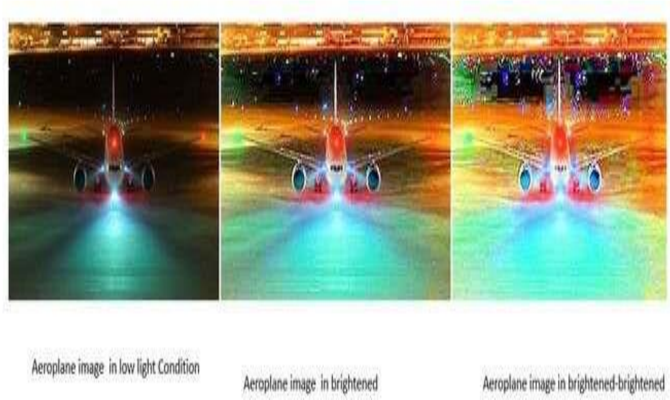
Data Acquisition: This refers to the process of obtaining raw data from various sources such as sensors, databases, or external files. In MATLAB, this can involve using functions or tools to read and import data into the workspace.

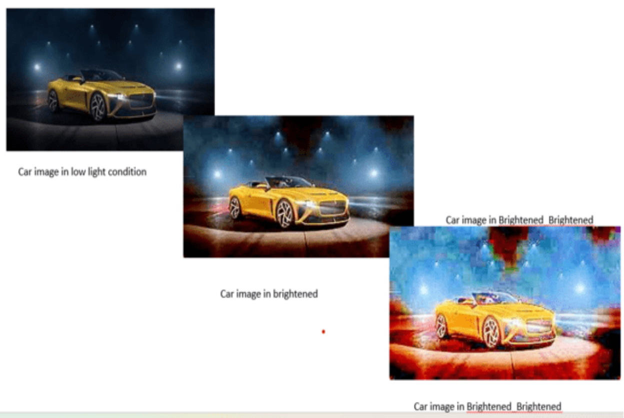
Normalization: Normalization is a preprocessing step where the features of the data are scaled to a similar range to improve the convergence of the deep learning model during training. In MATLAB, functions like normalize or manual scaling methods can be used for this purpose.

Labelling: Labelling involves assigning categories or classes to data instances, which is crucial for supervised learning tasks. In MATLAB, labelling can be done using functions like categorical or by manually assigning labels to data instances**.**

Fig: 8.Data Augmentation Of Experimental Results

Data Augmentation: Data augmentation is a technique used to artificially increase the size of the training dataset by applying transformations such as rotation, flipping, or cropping to the existing data. This helps in improving the generalization and robustness of the deep learning model. MATLAB provides functions like augmented Image Datastore and augment Image Data for data augmentation**.**





.

6. CONCLUSION

In conclusion this paper presents a comprehensive approach to address the challenge of vehicle recognition in low-light conditions by integrating transfer learning with Convolutional Neural Networks (CNNs) and the Low Light Image Enhancement (LIME) method. By leveraging pre-trained networks, the model extracts robust features from vehicle images, leading to improved classification performance. The study evaluates the proposed model on real-time vehicle datasets, measuring its performance using metrics such as accuracy, precision, false discovery rate, recall rate, and false negative rate. Moreover, the model is applied to both standard and custom datasets, demonstrating its effectiveness across different data distributions. Additionally, the model's capability for vehicle detection is explored using Regions with CNNs (RCNN) object detection on a smaller dataset. Notably, the paper emphasizes the importance of selecting an appropriate network architecture for classification tasks with limited labeled data, achieving greater accuracy through continuous training and minimal preprocessing. Overall, the findings suggest promising implications for applications such as surveillance, autonomous driving, and traffic management, with avenues for further research including the exploration of additional enhancement techniques and refinement of model architecture-for-deployment.

7.REFERENCES

[1] Abdelhamed, A., Lin, S., Brown, M.S.: The last decade has seen an astronomical shift from imaging with DSLR and point-and-shoot cameras to imaging with smartphone cameras. Due to the small aperture and sensor size, smartphone images have notably more noise than their DSLR counterparts. While denoising for smartphone images is an active research area, the research community currently lacks a denoising image dataset representative of real noisy images from smartphone cameras with high-quality ground truth. We address this issue in this paper with the following contributions. We propose a systematic procedure for estimating ground truth for noisy images that can be used to benchmark denoising performance for smartphone cameras. Using this procedure, we have captured a dataset - the Smartphone Image Denoising Dataset (SIDD) - of ~ 30,000 noisy images from 10 scenes under different lighting conditions using five representative smartphone cameras and generated their ground truth images. We used this dataset to benchmark a number of denoising algorithms. We show that CNN-based methods perform better when trained on our high-quality dataset than when trained using alternative strategies, such as low-ISO images used as a proxy for ground truth data.

[2] Aharon, M., Elad, M., Bruckstein, A.: K-SVD: In recent years there has been a growing interest in the study of sparse representation of signals. Using an overcomplete dictionary that contains prototype signal-atoms, signals are described by sparse linear combinations of these atoms. Applications that use sparse representation are many and include compression, regularization in inverse problems, feature extraction, and more. Recent activity in this field has concentrated mainly on the study of pursuit algorithms that decompose signals with respect to a given dictionary. Designing dictionaries to better fit the above model can be done by either selecting one from a prespecified set of linear transforms or adapting the dictionary to a set of training signals. Both of these techniques have been considered, but this topic is largely still open. In this paper we propose a novel algorithm for adapting dictionaries in order to achieve sparse signal representations. Given a set of training signals, we seek the dictionary that leads to the best representation for each member in this set, under strict sparsity constraints. We present a new method-the K-SVD algorithm-generalizing the K-means clustering process. K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the data. The update of the dictionary columns is combined with an update of the sparse representations, thereby accelerating convergence. The K-SVD algorithm is flexible and can work with any pursuit method (e.g., basis pursuit, FOCUSS, or matching pursuit). We analyze this algorithm and demonstrate its results both on synthetic tests and in applications on real image data.

[3] Ahn, N., Kang, B., Sohn, K.A.: In recent years, deep learning methods have been successfully applied to single-image super-resolution tasks. Despite their great performances, deep learning methods cannot be easily applied to real-world applications due to the requirement of heavy computation. In this paper, we address this issue by proposing an accurate and lightweight deep network for image super-resolution. In detail, we design an architecture that implements a *cascading mechanism* upon a residual network. We also present variant models of the proposed cascading residual network to further improve efficiency. Our extensive experiments show that even with much fewer parameters and operations, our models achieve performance comparable to that of state-of-the-art methods.

[4] Allebach, J., Wong, P.W.: We present a new method for digitally interpolating images to higher resolution. It consists of two phases: rendering and correction. The rendering phase is edge-directed. From the low resolution image data, we generate a high resolution edge map by first filtering with a rectangular center-on-surround-off filter and then performing piecewise linear interpolation between the zero crossings in the filter output. The rendering phase is based on bilinear interpolation modified to prevent interpolation across edges, as determined from the estimated high resolution edge map. During the correction phase, we modify the mesh values on which the rendering is based to account for the disparity between the true low resolution data, and that predicted by a sensor model operating on the high resolution output of the rendering phase. The overall process is repeated iteratively. We show experimental results which demonstrate the efficacy of our interpolation method.

[5] Badrinarayanan, V., Kendall, A., Cipolla, R.: Seg Net: Semantic segmentation is an essential task in computer vision that aims to label each image pixel. Several of the actual best approaches in this context are based on deep neural networks. For example, SegNet is a deep encoder-decoder architecture approach whose results were disruptive because it is fast and performs well. However, this architecture fails to fine-delineating the edges between the objects of interest in the image. We propose some modifications in the SegNet-Basic architecture by using a post-processing segmentation layer (using Conditional Random Fields) and by transferring high resolution features combined to the decoder network. The proposed method was evaluated in the dataset CamVid. Moreover, it was compared with important variants of SegNet and showed to be able to improve the overall accuracy of SegNet-Basic by up to 17.5%.

[6] Bertalm´ıo, M., Caselles, V., Provenzi, E., Rizzi, A.: We propose a novel image functional whose minimization produces a perceptually inspired color enhanced version of the original. The variational formulation permits a more flexible local control of contrast adjustment and attachment to data. We show that a numerical implementation of the gradient descent technique applied to this energy functional coincides with the equation of automatic color enhancement (ACE), a particular perceptual-based model of color enhancement. Moreover, we prove that a numerical approximation of the Euler-Lagrange equation reduces the computational complexity of ACE from theta(N2) to theta (N log N), where N is the total number of pixels in the image.

[7] Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Machine learning techniques work best when the data used for training resembles the data used for evaluation. This holds true for learned single-image denoising algorithms, which are applied to real raw camera sensor readings but, due to practical constraints, are often trained on synthetic image data. Though it is understood that generalizing from synthetic to real images requires careful consideration of the noise properties of camera sensors, the other aspects of an image processing pipeline (such as gain, color correction, and tone mapping) are often overlooked, despite their significant effect on how raw measurements are transformed into finished images. To address this, we present a technique to "unprocess" images by inverting each step of an image processing pipeline, thereby allowing us to synthesize realistic raw sensor measurements from commonly available Internet photos. We additionally model the relevant components of an image processing pipeline when evaluating our loss function, which allows training to be aware of all relevant photometric processing that will occur after denoising. By un processing and processing training data and model outputs in this way, we are able to train a simple convolutional neural network that has 14%-38% lower error rates and is 9×-18× faster than the previous state of the art on the Darmstadt Noise Dataset, and generalizes to sensors outside of that dataset as well.

[8] Buades, A., Coll, B., Morel, J.M.: We propose a new measure, the method noise, to evaluate and compare the performance of digital image denoising methods. We first compute and analyze this method noise for a wide class of denoising algorithms, namely the local smoothing filters. Second, we propose a new algorithm, the nonlocal means (NL-means), based on a nonlocal averaging of all pixels in the image. Finally, we present some experiments comparing the NL-means algorithm and the local smoothing filters.

[9] Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising can be described as the problem of mapping from a noisy image to a noise-free image. The best currently available denoising methods approximate this mapping with cleverly engineered algorithms. In this work we attempt to learn this mapping directly with a plain multi layer perceptron (MLP) app d time. Even contrast and brightness adjustments are challenging because they require taking into account the image content. Photographers are also known for having different retouching preferences. As the result of this complexity, rule-based, one-size-fits-all automatic techniques often fail. This problem can greatly benefit from supervised machine learning but the lack of training data has impeded work in this area. Our first contribution is the creation of a high-quality reference dataset. We collected 5,000 photos, manually annotated them, and hired 5 trained photographers to retouch each picture. The result is a collection of 5 sets of 5,000 example input-output pairs that enable supervised learning. We first use this dataset to predict a user's adjustment from a large training set. We then show that our dataset and features enable the accurate adjustment personalization using a carefully chosen set of training photos. Finally, we introduce difference learning: this method models and predicts difference between users. It frees the user from using predetermined photos for training. We show that difference learning enables accurate prediction using only a handful of examples.

[10] Cai, J., Gu, S., Timofte, R., Zhang, L.: This paper reviewed the 3rd NTIRE challenge on single-image super-resolution (restoration of rich details in a low-resolution image) with a focus on proposed solutions and results. The challenge had 1 track, which was aimed at the real-world single image super-resolution problem with an unknown scaling factor. Participants were mapping low-resolution images captured by a DSLR camera with a shorter focal length to their high-resolution images captured at a longer focal length. With this challenge, we introduced a novel real-world super-resolution dataset (RealSR). The track had 403 registered participants, and 36 teams competed in the final testing phase.