**Skelly Entertainment Internshala Questionnaire**

**Problem 2:**

**DEEP MULTI-STAGE LEARNING FOR HDR WITH LARGE OBJECT MOTION**

**Green Rosh K S Anmol Biswas Mandakinee Singh Patel B H Pawan Prasad**

This paper proposes a viable method that incorporates Convolutional Neural Networks (CNNs) to generate High Dynamic Range (HDR) images input Low Dynamic Range (LDR) images. Unlike conventional methods, using multiple step CNN to generate virtual exposures and create ghost-free HDR images. The images learn in a deep multi-stage learning approach using the CNN approach gets rid of ghost-free and artefacts from the LDR, unlike the end-to-end usual approach.

The method is divided into two steps: 1) Exposure Alignment Stage and 2) HDR Merge State.

The Exposure Alignment Stage produces a virtual exposure image of each guide image using a structure guide image using AlginNet. In this stage, there is an encoder-decoder architecture with residual blocks like the Image Transformation Networks. The input images are encoded into 2 layers and one layer is used to extract an abstract feature and then batch normalisation and Relu activation functions are used. At last the output of the residual block is decoded back on the 2 deconvolutional layers.

The second stage of the method- HDR Merge Stage, uses MergeNet consisting of 2 encoding layers, 7 residual layers and 2 decoding layers. This helps in helping the network can learn to merge details in motion regions from the virtual exposure regions and non-motion regions from the input images. In this method, network can learn to merge details in motion regions from the virtual exposure regions and non-motion regions from the input images.

The AlignNet method uses a dataset consisting of structure guide images, exposure guide images and ground truth images are required. In this way, 20 unique triplets can be generated for exposure boosting and exposure suppression each from the 5 different exposures captured. 30 such scenes were captured resulting in 600 unique triplets of varying exposures and structures.

The dataset required for AlignNet consists of structure guide images, exposure guide images and ground truth images. To prepare this dataset, five images of increasing exposures were captured each for teo different poses( of the subject). To make the model generic, the images were captured using different smartphones with different camera sensor specifications. The proposed algorithm is able to recover details in the saturated while maintaining the structure of Im without producing any artifacts.

The MergeNet used the publicly available HDR dataset by Kalantari. They use random sets from the dataset and are trained in batches to produce image pairs which are passed through AlignNets to generate virtual exposure image pairs, which is used for training the MergeNet.The ground truth HDR image was tonemapped to generate ground truth for this training

To understand the effectiveness of the two networks, two types of loss functions are calculated: Content loss (Lc), which estimates the fidelity of the generated output imag with respect to the ground truth image and feature reconstruction loss which minimizes the differences between high-level feature of the network generated image and the ground truth.

On comparison with state-of-the-art methods and the proposed method, ther are certain imporbvemtns in this method. While Kalantari seems to produce strong artifacts, the methods proposed by Sen and Wu produces ghost artifacts in the motion regions. The learning based solution proposed by Wu seems to fail to hallucinate realistic details resulting in artifacts and reduced dynamic range.

The proposed method is able to hallucinate realistic details in the large occluded region. Since the proposed method tackles the problem in two stages using specialized networks at each stage, it is able to hallucinate much better details in occluded regions. This advanced algorithm is used to produce better images without the greyish output that the conventional methods produce.

The results were also quantitatively compared to Kalantari’s and Sen and Wu’s methods using PSNR and SSIM scores. The proposed method has significantly higher scores proving that this method is much more efficient.

Employing deep learnin neural network methods provided much better PSNR scores than the state-of-the-art conventional methods already existing. Utlising the layers and its ability to learn various properties effectively, generating HDR images from a set of LDR images has proved much more effective. Neural networks takes all factors such as lighting, noise, stur=cture and exposure into much better consideration than in the end-to-end approach.

**Paper 2 Challenge on High Dynamic Range Imaging: Dataset, Methods and Results**

Eduardo P´erez-Pellitero† Sibi Catley-Chandar† Aleˇs Leonardis† Radu Timofte† Xian Wang Yong Li Tao Wang Fenglong Song Zhen Liu Wenjie Lin Xinpeng Li Qing Rao Ting Jiang Mingyan Han Haoqiang Fan Jian Sun Shuaicheng Liu Xiangyu Chen Yihao Liu Zhengwen Zhang Yu Qiao Chao Dong Evelyn Yi Lyn Chee Shanlan Shen Yubo Duan Guannan Chen Mengdi Sun Yan Gao Lijie Zhang Akhil K A Jiji C V S M A Sharif Rizwan Ali Naqvi Mithun Biswas Sungjun Kim Chenjie Xia Bowen Zhao Zhangyu Ye Xiwen Lu Yanpeng Cao Jiangxin Yang Yanlong Cao Green Rosh K S Sachin Deepak Lomte Nikhil Krishnan B H Pawan Prasad

This paper is proposing a solution for the New Trends in Image Restoration and Enhancement (NTIRE) workshop challenge. They are estimating HDR images from multiple low-dynamic range (LDR) images. The goal of this paper is to get the best HDR reconstruction with respect to a ground-truth image, evaluated both directly and with a canonical tonemapping operation of the input LDR images. The input for the algorithm consist of multiple frame LDR images captured with different exposure values (EV) which are processed and fused together into one HDr image with frame alignments and pixel clarity.

Unlike in computational photography, where single-sensor cameras are used to acquire images for higher clarity, using Convolutional Neural Networks (CNN) advances the process of HDR reconstruction by complex learning of entangled vision and effective binding of multiple input frames into one. When multiframe LDR imagesbare not available, the conventional state-of-the-art methods of HDR image reconstruction using single frame LDR inout works effectively too!

The NITRE 2021 challenge aims to improve the state-of-the-art of HDR imaging and reconstruct wide range of scene illumination and complex motions. To attain this, this paper first discusses in detail the preparation of a suitable dataset. Due to the fact that current available HDR based datasets are captured using static image bracketing,so that stopmotion dynamic scenes can be assembled, for the new dataset, the subject is asked to stand still and three bracketed exposure images on a tripod are captured along with two images with the object moving is captured. Therefore the dataset used in this paper contains input LDR triplet with inter-frame motion and a reference HDR ground-truth image aligned to the central frame.

The three input LDR images consist of short, medium and long exposures, and a related ground-truth HDR image aligned with the central medium frame image. Thes inout LDR images are synthetically generated from Froelich’s dataset involving HDR videos using professional camera rigs and lighting sources. The synthetic generation is done using the following formation models: Image formation model and noise model.

This dataset is further split into the usual train, test and validation groups. The splits are done so that there is no overlap in the scenery of any fo the test, train or validation image sets.

This paper has divided the challenge objective into two different tracks, with the same evaluation metrics and ground-truth data. This helps in obtaining unieuely similar results from both tracks that can explain the performance differences between single and multi-frame HDR imaging effectively.

The first track uses single-frame LDR input images. This makes it challenging to recover over and under exposed ares due to the lack a set of images and also maeks denoising challenging as noise sources are zero mean. But having a single-frame ninout reduces the work in the motion related objectives like ghosting.

The second track uses multiple-frame LDR images as input with three different exposed images, diverse motion between frames with complex moving and changin light sources. The medium frame of LDR in this track corresponds to the single-frame LDR image in track 1.

To evaluate the challenge, peak signal-to-noise ratio PSNR) is used along with µ-law tonemapper. The PSNR is directly computed on the estimated HDR images and the PSNR-µ is calculates using tone-maping operations. The results are normalised to scale.

The results of the challenge was segregated by the tracks. The track 1 winners’ main point of evaluation seemsbe the values of the PSNR-µ score. Though the difference in values seem negligible they’re quite prominent in terms of PSNR-µ. Whereas in track 2, also comoareed using PSNR-µ scores but the margin here much larger between each team than that in track 1.

The winning team NOAHCTV proposed a method each for both tracks which seem viable. For track 1, they divided the task into two sub-tasks fro denoising (using MIRNET) and hallucinating(using masking) thereby reconstructing from a single-frame LDR much more efficiently. For track 2,they proposed a three stage method consisting of an Alignment and Augmentation module, an Attention Based Information Extraction module, and an Enhancement and Fusion module. First a pretrained PWC-Net is used to warp the input images, then AHDRNet attention mechanism is used and then shallow features are extracted by a convultional layer.

The team MegHDr proposed a multi-layer imagine pipeline with ADNet and processed in separate modules and finally concatenated feature-wise in DRBDs. The team XPixel proposed a three sub-network system to achieve adaptive modulation, calculate loss function and balance over-exposed values to reconstruct an HDR image. The team BOE-IOT-AIBD proposed a three sub-networks which each perform a different task: Image Reconstruction (IR), Detail Restoration (DR) and Local Contrast Enhancement (LCE).

On understanding the solutions of every team, one common aspect seems to be the breaking down of the problem into sub-tasks or betworks and splitting hte input frame(S) into features and acting on them separately before reconstructing the final output HDR. Also, almost all these groups of made use of neural networks for on aspect or the other in the reconstruction as it preovides them with advantages in reducing movement and exposure. The concept of ADHRNet seems to be the most common among the solutions to provide multi-level attention to the frame(s).

Thus processing attention in multiple levels, breaking down single-frame LDR images into features and encoding multiple-frames and concating seperate features from each seem to viabel methods for reconstructing High Dynamic Range (HDR) images.