



JOINT INSTITUTE
交大密西根学院

UM-SJTU JOINT INSTITUTE
VE450 MAJOR DESIGN EXPERIENCE
FA2018

HIGH DYNAMIC RANGE VIDEOS RECOVERING ALGORITHM



Sponsors:

Juan ZHAO
Yejun GUO
Ruiling SONG

Instructors:

Dr. Yong LONG, Assistant Professor, Advisor
Dr. Mian LI, Associate Professor, Chair
Dr. Chengbin MA, Associate Professor
Dr. Yunlong GUO, Associate Professor

Group 15:

HU Chong, SHAO Xingzeng, SONG Huanian,
TANG Qianyong, ZHANG Yangzhen
September 29, 2019

Abstract

The target of the project is to reconstruct High Dynamic Range (HDR) video from LDR Dynamic Range (LDR) based on a single exposure. A group of five senior students in University of Michigan - Shanghai Jiao Tong University Joint Institute, with the assistance of three experienced supervisors from Intel APAC Research Ltd, accomplish this project in three month and a half.

HDR Imaging with a larger range of luminance levels displays digital images more similar to real-life scenes experienced through the human visual system. HDR imaging has a wide range of applications in many fields, including medical imaging, human computer interface (HCI) and pre-architectural visualization in recent years.

The specification for this program includes that our program should operate on recovering HDR videos from single-exposure LDR videos in Linux environment. The running time is expected to be 100s per Mega Pixels. The program need to support various video formats. The recovered images are required to have better performance than the traditional method. Based on these specifications, the Computer Graphics industries could generate instant realism of time-effective and cost-effective HDR images recovering so that customers and users can be provided with more observable details to the our eyes compared with traditional photography.

We arrange our concept generation and selection in expectation of time efficiency and more accurate prediction than the current existing methods. Traditional methods may lead to overexposure or underexposure problems. But deep learning methods not only solve these problems but also provide artificial lighting for recovering the missing details in the saturated area. That's why we choose the deep learning methods, using the trained model to reconstruct LDR videos from HDR videos.

Our final project design is mainly composed of three steps. They are extracting LDR images from LDR videos, predicting HDR images based on a trained model, and converting the predicted HDR image sequences into HDR videos. As for the HDR prediction, we use Deep CNN and implement a VGG16 neural network. We prepared the HDR-LDR image pairs by collecting the HDR videos, extracting frames and doing data augmentation. We use FFMpeg to satisfy the conversion between images and videos. We manage to recover HDR videos with a score of 82 in Visual Difference Predictor. In this report, we introduce the background about HDR, concept generation and selection, parameter analysis, final design and the discussion and conclusion adopted. Next, we discuss the customer requirements and engineering specifications in detail. Further, we evaluate our progress, project plan and potential problems. We conclude our achievements and future work.

The future manufacture plan is about assemble our program into a part of an HDR software. The cost of our program is mainly the Solid-state device storing the HDR-LDR dataset and the reconstructed HDR videos.

Key words: High Dynamic Range, Convolutional Neural Network, Deep Learning, HDR10 Video

摘要

本项目的目标是基于单张低动态范围视频的基础上重建的高动态范围视频。来自上海交通大学密西根学院的五名学生，在三位经验丰富的英特尔亚太研究有限公司的研究员的指导与带领下，用时三个半月完成了这个项目。

HDR 成像显示的数字图像具有更大亮度级别，使得人们在观看数字图像时能够获得更为接近视觉系统的真实体验。HDR 成像在许多领域具有广泛的应用，包括医学成像，人机界面和预架构可视化。

我们的项目预期是在 Linux 环境中从单张 LDR 视频中恢复 HDR 视频。运行时间预计为每百万像素用时 100 秒。该程序预计可以支持各种视频格式。恢复出的高动态范围图像需要具有比传统方法更好的性能。基于这些规范，计算机图形工业可以实现高效且便利的高动态范围图像恢复。从而与传统摄影相比，我们的项目可以为客户端和用户恢复图像中更多细节。

我们项目的设计旨在提高时间效率和获得同传统方法相比更为快捷准确的高动态范围图像预测。传统方法的缺陷包括过度曝光或曝光不足问题。新兴的深度学习不仅解决了这些问题，而且还提供了人造照明恢复出高动态范围图像饱和区域的细节。因此我们采用深度学习来完成我们的项目

我们的最终项目设计主要由三个步骤组成。他们是从 LDR 视频中提取 LDR 图像，基于训练好的模型预测 HDR 图像，并将预测的 HDR 图像序列转换为 HDR 视频。对于 HDR 预测，我们使用深度卷积神经网络并实现了一个 VGG16 神经网络。通过采集 HDR 视频并提取帧和数据增强，我们制备了 HDR-LDR 图像对。我们使用 FFMpeg 来满足图像和视频之间的转换。我们成功地恢复了 HDR 视频，其视觉差异预测得分为 82 分。在这篇报告中，我们将介绍 HDR 的背景，概念的产生和选择，参数的分析，最终的设计以及讨论和结论。接下来，我们将详细讨论客户需求和工程规范。此外，我们评估我们的进展，项目计划和潜在的问题，并总结了工作成果和今后的潜在改善工作。

未来的生产计划是关于将我们的程序组装成 HDR 软件的一部分。我们项目的成本主要是用来存储 HDR-LDR 数据集和重构 HDR 视频的固态硬盘。

关键词：高动态范围，卷积神经网络，深度学习，HDR10 视频

Contents

1	Introduction	7
1.1	Background	7
1.2	Problem Description	7
1.3	Needs and Expected Outcomes	7
1.4	Information Sources	8
2	Specification	9
2.1	Customer Requirements	9
2.2	Engineering Targets	10
3	Concept Generation	12
3.1	Frame Extraction	12
3.2	HDR Image Reconstruction	12
3.3	HDR Video Recovering	13
4	Concept Selection	13
4.1	Frame Extraction	13
4.2	HDR Image Reconstruction	14
4.3	HDR Video Recovering	15
5	Concept Description	15
6	Parameter Analysis	16
6.1	Deep Learning Model	16
6.2	Data Pre Processing	16
6.2.1	Gaussian Noise	17
6.2.2	Brightness and Hue	17
6.2.3	Flip	17
6.2.4	JPG Quality	17
6.3	Training	17
6.3.1	Number of Epochs	17
6.3.2	Batch Size	17
6.3.3	Learning Rate	18
6.3.4	Random Shuffle	18
6.3.5	Optimizer	18
6.4	Composing HDR Video	18
6.4.1	openEXR	18
6.4.2	libTIFF	18
7	Final Design Workflow	19
8	Manufacturing Plan	19
8.1	Environment Setup	19
8.1.1	TensorFlow Installation	19
8.1.2	Package Installation	19
8.1.3	Tools Preparation	20

8.2	Model Training	20
8.2.1	Data Preparation	20
8.2.2	Model Training	20
8.3	Video Conversion	21
9	Test Result	21
10	Engineering Changes Notice	22
11	Discussion	23
12	Recommendations	23
13	Conclusions	24
13.1	Our Progress	24
13.2	Future work	24
14	Information Sources and Reference List	26
15	Appendices	29
15.1	Bill of Materials for System	29
15.2	The Network Structure of VGG16	29
15.3	Inverse Tone Mapping Method	29
16	Acknowledgement	30

List of Figures

1	The comparison of LDR image and HDR image [3]	7
2	Literature Review *Data from Web of Science, Advanced Search: TI=("high dynamic range" OR hdr) AND TS=image reconstruct*	8
3	Detailed Statistics on Comparison between Authors and Countries	9
4	Quality function deployment of our algorithm	11
5	Basic work flow of HDR video recovering.	12
6	Different concepts of each step for HDR video Recovering.	13
7	Weighted Decision Matrix (W* stands for weight).	14
8	Work flow structure for HDR video recovering.	15
9	Comparison of the Histogram of HDR Image	21
10	The Visual Difference Predictor Comparison Test	22
11	The structure of VGG16 network	29

List of Tables

1	Engineering specifications and targets.	10
2	Different concepts for frame extraction.	14
3	Comparison between different HDR video standards.	15
4	Packages used in this project	20
5	The bill of materials.	29

1 Introduction

1.1 Background

Dynamic range is defined as the ratio of luminance between the brightest and the darkest objects in a scene. High dynamic range imaging is a high dynamic range (HDR) technique used in imaging and photography to reproduce a greater dynamic range of luminosity than is possible with standard digital imaging of photographic techniques. The aim is to present a similar range of luminance to that experienced through the human visual system. [1]

1.2 Problem Description

High Dynamic Range Imaging (HDR) is a high dynamic range (HDR) technique to reproduce a larger dynamic range in luminosity than standard digital imaging or photographic techniques. By adapting iris and other methods, the human eye is constantly adjusted to accommodate the wide range of luminance that exists in the environment [2]. So the aim of HDR image is to provide luminance range similar to the human visual system to offer a better viewing performance.



Standard (Low) Dynamic Range

High Dynamic Range

Figure 1: The comparison of LDR image and HDR image [3]

1.3 Needs and Expected Outcomes

Our design is to reconstruct a high dynamic range (HDR) video from a single low dynamic range (LDR) video that have been captured using standard cameras without any assumptions on the imaging system or camera calibration. There are two novelty in our design. First, the transformation from video to video is a novelty, since HDR is a relative new concept and there is no

transformation of videos before [4]. Second, while the tradition approaches to obtaining HDR images is to merge multiple LDR photographs or to use special image sensors, such as an over-sampled binary image sensor, which both introduce limitations, we instead tackle the problem by reconstructing visually convincing HDR images from single images.

1.4 Information Sources

There are two main problems in our project. One is to make clear the image acquisition pipeline, image storage format and display calibration, the other is to reconstruct HDR images from single exposure.

For the first problem, we referred to Report ITU-R BT.2390-3 of High dynamic range television for production and international programme exchange issued by Radiocommunication Sector of ITU [5] and the book High Dynamic Range Video written by Karol Myszkowski, Rafal Mantuik, Grzegorz Krawczyk. [6]

For the second problem, in the literature review, we focus on several key words, including High Dynamic Range (HDR), Deep Convolutional Neural Network and tone mapping and we conducted our literature review through Web of Science and Research Gate. We found that 56% of the related papers and works were published in the past 4 years. Among all these results, China has publish most documents We visualized the search results in the Figure.2 and Figure.3.

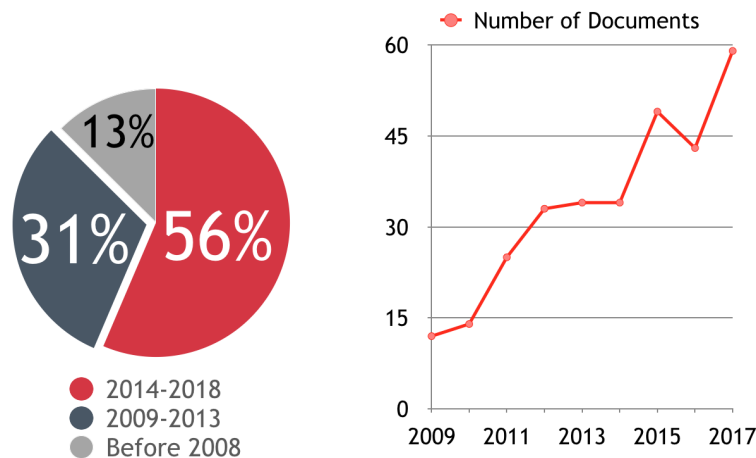


Figure 2: Literature Review *Data from Web of Science, Advanced Search: TI=("high dynamic range" OR hdr) AND TS=image reconstruct*

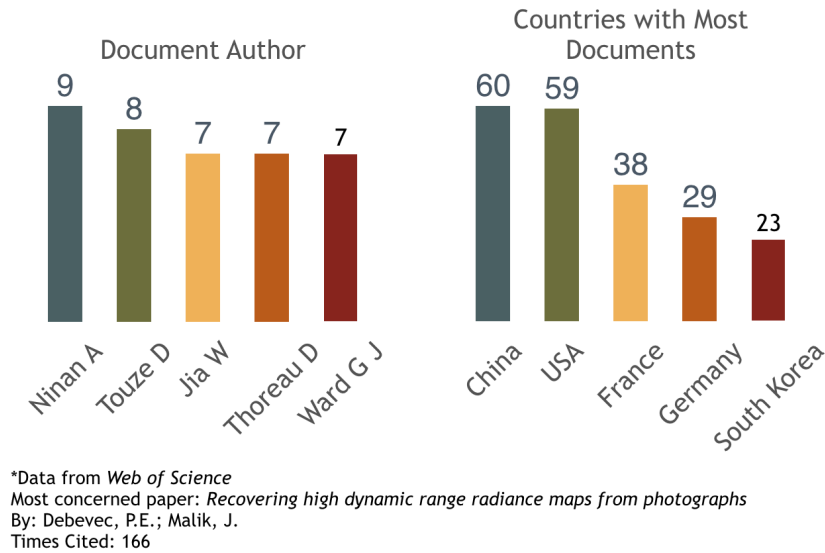


Figure 3: Detailed Statistics on Comparison between Authors and Countries

Some of these papers used traditional methods, while the others used the state-of-the-art strategies, including DCNN (ExpandNet, VGG), HEVC encoder (Perceptual Encoder), Support Vector Regression and Linear multi-component transforms. Note that inverse Tone mapping, a conventional method, is of significance since it helps to improve the performance of demonstrate HDR images on LDR devices which are widely used in our daily life. In other words, inverse tone mapping helped to make the novel technology user-friendly. Among these most frequently mentioned methods, we focused on the cutting-edge Deep CNN method.

2 Specification

2.1 Customer Requirements

In this section, the requirements have been boiled down to two aspects: the performance of the reconstructed HDR video and the capability of our program. In each aspect, we specify the customer expectation.

The customer requirements are listed as following. For the performance of the reconstructed HDR video, the requirements are:

1. **High contrast.** The reconstructed HDR video should have high contrast.
2. **Natural vision.** The reconstructed HDR video should look natural. It should not look

strange.

3. **Detail display.** The reconstructed HDR video should display the details clearly, such as the texture of wood.

For the capability of our program, the requirements are:

1. **Fast reconstruction.** Our program should reconstruct HDR video in a reasonable speed.
2. **Able to transform existing regular videos.** Our program should be able to reconstruct HDR videos from existing regular LDR videos.
3. **Implementation in Linux environment.** Our program should run in Linux environment.

2.2 Engineering Targets

According to our customer requirements, we determine quantified engineering specifications and targets. In terms of the “high contrast”, “natural vision” and “detail display” requirements, we use HDR visual difference predictor (VDP) to quantitatively determine the extent to which we meet the requirements. For “fast reconstruction” requirement, we use “running time” to evaluate it. For “able to transform existing regular videos” requirement, we consider “max supported resolution”, “support single-exposure videos”, and “number of video formats supported” as specifications, because existing LDR videos can have different resolutions and formats, and they are normally taken (single exposure, not taken specially for HDR reconstruction). The “run in Linux environment” can directly be used as engineering specification.

The engineering specifications and targets are shown in Table 1.

Specification	Unit	Target value
Running time	sec/MPixel	100
Max supported resolution	MPixel	2
Support single-exposure videos	T/F	T
Number of video formats supported	1	at least 1
Support Linux	T/F	T
Quantitative evaluation	1	75

Table 1: Engineering specifications and targets.

The Quality Function Deployment diagram is shown in Figure 4. The numbers next to the requirements show the importance of each requirement, where “10” means the most important. The matrix in the center shows the correlation between customer requirements (rows) and engineering specifications (columns). “9” means strongly related, and “3” means moderately related. What’s more, some engineering specifications are related. The relationships are shown in the top triangular table. A cross with a circle means “strong negative relationship”, while a single cross means “mild negative relationship”. The rightmost column is a benchmarking method - inverse tone mapping, and its scores on each requirements, where “5” means well satisfied and “1” means not satisfied. The bottom 3 rows are measurement unit, total weight, and normalized weight in percent, respectively. From the normalized weights, we can see that the “quantitative evaluation” specification is of most importance, so we should focus on it most.

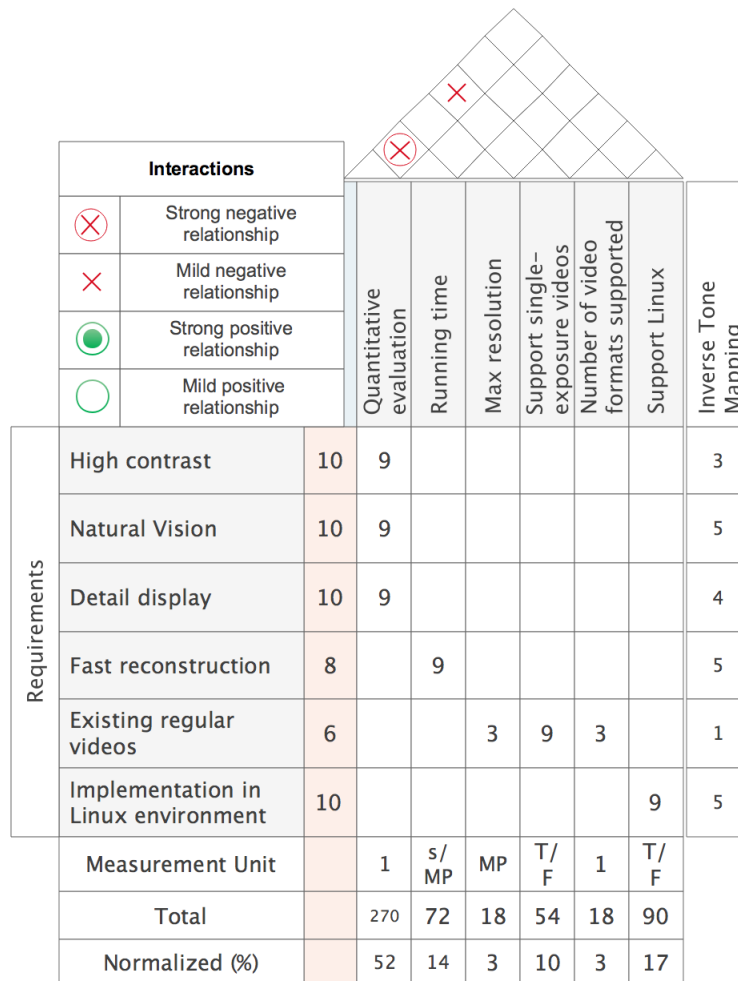


Figure 4: Quality function deployment of our algorithm

3 Concept Generation

We conducted a general search on existing HDR video recovering methods, and determined the basic workflow of our project as shown in Figure 5. The system first extracts frames from the input LDR video, and then converted the frames to HDR, and finally splices HDR frames to the output HDR video. Then we focused on different parts of the flow and conducted corresponding literature reviews, and we will discuss the three parts one by one.

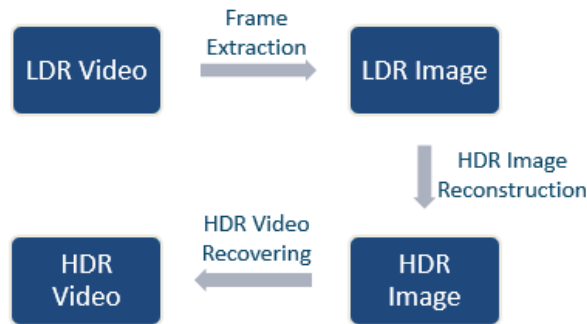


Figure 5: Basic work flow of HDR video recovering.

3.1 Frame Extraction

For the frame extraction, we found three tools, FFmpeg, OpenCV and MATLAB, which can help get all the frames from a HDR10 video.

3.2 HDR Image Reconstruction

Reconstruction of HDR images from LDR images is one of the most challenging part in our project. After researching, we found that the two commonly used methods are inverse tone mapping and deep learning methods using Convolutional Neural Networks (CNNs).

While inverse tone mapping is a widely-used traditional method, the deep CNNs can solve the overexposure/underexposure problems and reduce the banding artifacts. And for the later, we selected two specific networks, ExpandNet [7] and VGG16 [8], as our potential solutions by considering both speed and accuracy.

3.3 HDR Video Recovering

For the HDR video recovering from HDR images, we found it was important to consider the HDR image file format and the HDR video standard. For different HDR video standards, we will have different HDR video reconstruction processes. There are three types of broadcasting standards for the HDR videos. They are Dolby vision, HDR 10 and Hybrid Log Gamma (HLG). All of them are based on the concept of a Electro-Optical Transfer Function (EOTF) curve translating the stored digital values into brightness values shown on the screen. These standard makes it possible to match the video contents to the limit of color gamut and peak brightness properly when the real color gamut and peak brightness might exceeds the display capabilities.

Thus, we have generated all the potential concepts for different catagories of our work flow. To sum them up, we listed Different concepts of each step for HDR video Recovering as shown in Figure 6.

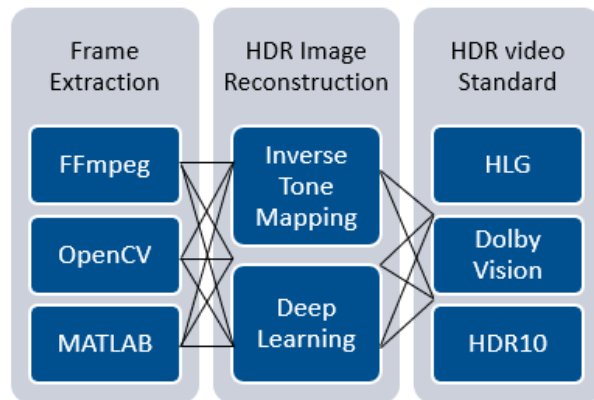


Figure 6: Different concepts of each step for HDR video Recovering.

4 Concept Selection

Then, we started careful comparisons of the concepts for different categories, in order to select the most appropriate ones for our project.

4.1 Frame Extraction

For the frame extraction, we extracted frames from a same video by all three methods. While using OpenCV and MATLAB requires dealing with both the video reading and video processing,

we can implement ready-made command statement with FFmpeg on Linux which is much more convenient and efficient. The results were the same, but FFmpeg was much faster than the implementation with OpenCV and MATLAB. After comparison, we choose FFmpeg to do the frame extraction from HDR10 video.

Frame Extraction methods		Ready-made Command Statement processing videos
Concept1	FFmpeg	✓
Concept2	OpenCV	×
Concept3	Matlab	×

Table 2: Different concepts for frame extraction.

4.2 HDR Image Reconstruction

To evaluate the quality of the recovered HDR image and the efficiency of a method, we designed the following criteria and used the weighted Decision Matrix to compare these methods. The criteria included data preparation, information preservation, compatibility, robustness and objective evaluation between the output and the ground truth. Since HDR images datasets are not widely available, we need to collect and prepare our own dataset. In the LDR image, there exist real scene or light missed due to the fact that the extreme variation in the dynamic range of the saturated regions. Thus, the ability to recover the missing luminance and the missing details is of significance. Meanwhile, we expect the inference is stable so that there won't be dramatic change between the sequential frames. In order to give objective evaluation of the HDR output, we introduce the Visual Difference Predictor[], which is a matrix indicating the probability of detecting difference between two images.

	Inverse Tone Mapping				ExpandNet (CNN)			VGG16 (CNN)		
	W*	Value	Score	Rating	Value	Score	Rating	Value	Score	Rating
Data Preparation	0.1	Fair	5	0.5	Fair	6	0.6	Fair	6	0.6
Information Preservation	0.25	Fair	5	1.25	Good	7	1.75	Good	7	1.75
Compatibility	0.2	Fair	5	1.0	Good	7	1.4	Good	7	1.4
Robustness	0.25	Fair	6	1.5	Good	8	2	Excel	9	2.25
Visual Difference Predictor	0.2	Fair	5	1.0	Good	8	1.6	Excel	9	1.8
Total				5.25			7.35			7.8
Rank				3			2			1

Figure 7: Weighted Decision Matrix (W* stands for weight).

As shown in Figure 7, VGG16 scores the highest and hence we select it to train models for HDR image generation.

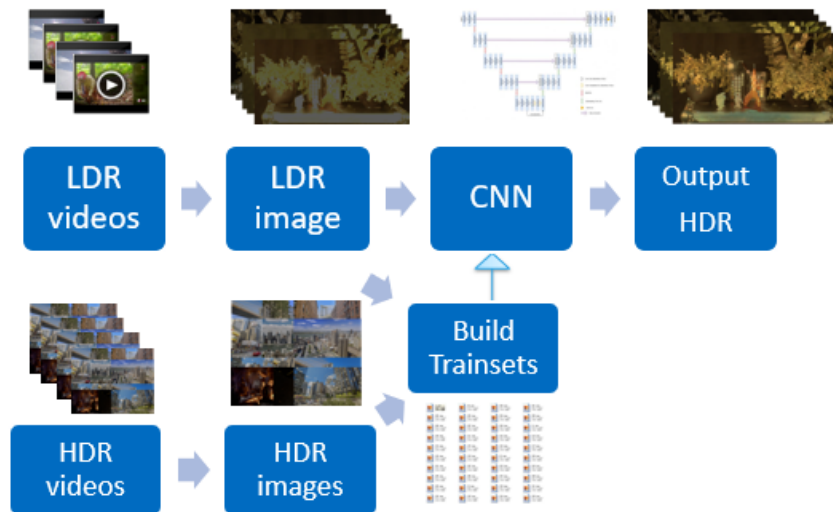
4.3 HDR Video Recovering

Lastly, for the HDR video standards, we carefully read the BT.2390-3 standards report provided by International Telecommunication Union. HLG considers both the variety in legacy equipment and TVs and the bandwidth ranges. Used in broadcasting, HLG is more practical than Dolby Vision, as Dolby Vision uses dynamic metadata to optimize video frames which is time-consuming. However, HLG lacks some top intensity that can be delivered by other formats. Therefore, we selected HDR10 as our output video standard as it is wide-ranged and efficient.

	HDR video standards	top intensity	without metadata
Concept1	HDR10	✓	✓
Concept2	HLG	×	✓
Concept3	Dolby Vision	✓	×

Table 3: Comparison between different HDR video standards.

5 Concept Description



12

Figure 8: Work flow structure for HDR video recovering.

The whole system is shown in the figure above. For the first stage, we obtained LDR videos as the input, and then extracted LDR image frames from the videos using FFmpeg. Then we took the LDR images as the inputs of the trained CNN model and generated HDR images. Finally, we spliced the HDR images to form the output video in HDR10 standard.

Specificly, for the CNN model training, we needed to prepare datasets for training and evaluation. For this process, we extracted frames from HDR10 videos and used a "virtual camera" based on Opto-Electronic Transfer Function (OETF) curves and care response functions to construct corresponding LDR images as pairs of data. Also, we took advantages of Gaussian Blur to do data augmentation in order to enrich the input situations.

6 Parameter Analysis

6.1 Deep Learning Model

In our project, the specification were determined by the sponsor. The type of our deep learning model should be able to reconstruct the HDR image from the original LDR image without any loss in information. Instead, we should recover more information so that the HDR video will have more information. U-Net is a net that very fits our need to recover more information and keep the original information without loss. In the U-Net, the information is transferred from one side of "U", let's say, a resolution of 1920×1080 pixel image is put into the network. The left side of the "U" will re-size the dimension of the image in each iteration. At the bottom of the U-Net, the dimension reach its maximum. At the right side of the U-Net, the network will begin to regenerate the HDR image. The dimension will decrease. Also, the matrix in the left side will also be compund to the matrix in the right side of U-Net so that it can guarantee the information will not be changed too much.

6.2 Data Pre Processing

This part includes the pre training. This part we mainly do the data augmentation, this part helps to make our limited data-set to become bigger. In using different kinds of methods with various parameter, we can obtain the purpose of data augmentation.

6.2.1 Gaussian Noise

First is the Gaussian Noise. We added the Gaussian Noise into our data set so that the model will be compatible to all kinds of video.

6.2.2 Brightness and Hue

For the brightness and hue part, we added different level of brightness and hue level into our dataset. The brightness and hue level are ranging from 0 to 1 with Normal Distribution. This process helps our model to be compatible to different images with different levels of brightness and hue.

6.2.3 Flip

For this part, we enable the flip process to our dataset. For a same image, after flipping, the image will be regarded as two images so that it is a way of data augmentation.

6.2.4 JPG Quality

We enabled different quality to our dataset, ranging from 0 to 100. The quality can be regarded as the loss of the image. For quality equals to 0, the information will be totally lost. For quality equals to 100, there will be no quality loss.

6.3 Training

6.3.1 Number of Epochs

Number of Epochs means the number of turns the model will learn from our dataset. In theory, the larger the number of epochs is, the better will our model perform. However, after a number of epochs, our evaluation shows that the performance begins to improve very little and the need to increase the number of epochs decreases. We found that 50 as the number of epochs can satisfy almost all the LDR image training.

6.3.2 Batch Size

Batch size is the block of data we sent into the model. The larger the batch is, the more quickly will the model trains. However, the larger the batch size is, the better the hardware performance

needs to be. For our project, we established our environment using Cuba accelerating. Our GPU platform is Nvidia GTX 1060. Due to the limitation of our hardware performance, when the batch size is set to 2, the GPU memory will be full. Hence, we set the parameter of batch size into 2.

6.3.3 Learning Rate

Learning rate is the weight of every iteration of learning that change our model. This parameter should be modified very carefully. After we change the learning rate, we will wait for 10, 000 iterations to see the loss. We changed this learning rate to find a small loss after 10, 000 that is smaller than or equal to 0.0001.

6.3.4 Random Shuffle

Random shuffle is used in training. We randomly choose a image from the dataset to avoid images from one scene to trained too much and an over-fitting is resulted. Using random shuffle, this problem can be avoided by random images from random scene are put into the model.

6.3.5 Optimizer

Optimizer is used to help the gradient descend. There are many optimizer we can use, such as adam, mini-batch SGD, SGD and so on. In our problem, we chose adam as our optimizer due to its better performance and can avoid local-minimum.

6.4 Composing HDR Video

6.4.1 openEXR

In composing the HDR Video, we need to deal with the EXR format HDR image. The tool openEXR is a package that can manage the EXR format image. We used this tool to deal with EXR format image.

6.4.2 libTIFF

In composing the HDR video, we also need to deal with the TIF format HDR image. Tool libTIFF is a puthon library that can maage the TIF format image. We used this tool to deal with TIF format image.

7 Final Design Workflow

The final design is our LDR-HDR converting algorithm.

First, I will talk about the workflow to convert an LDR video into HDR video. First, for the LDR video we are going to convert, we extract frames from the LDR videos to get a set of LDR images. Using the deep CNN model we have trained, we put the LDR images into model to obtain a set of HDR images. Finally, using ffmpeg, we can compose the HDR pictures into an HDR video.

On an HDR screen, the converted HDR video can be viewed correctly. In a dim scene original in LDR video, we can see a huge performance on the HDR screen. More details will be presented. Using our method of evaluating, which is widely used in discriminating difference in images. The comparison between HDR converted using our method and the ground truth is small. This shows that the final design of our method is robust and reliable.

8 Manufacturing Plan

8.1 Environment Setup

8.1.1 TensorFlow Installation

In order to use the tensorflow frame, we first install the Nvidia driver on our ubuntu 18.04 system. The GPU version is 1060 and the driver version is 396. After installed the Nvidia driver, we also install CUDA and CUDNN so that we can use the GPU version of tensorflow. The version of CUDA and CUDNN are 9.2 and 7.2.1 respectively. Notice that if we want to use this version of CUDA and CUDNN, we need to use bazel to install tensorflow from source code instead of using pip to install tensorflow. The version of tensorflow we used in this project is 1.10. We use bazel to compile the source code of tensorflow and set the paths and versions for CUDA, CUDNN and other related packages and then we can install tensorflow from the compiled package of tensorflow. By using the GPU version of tensorflow, we could accelerate the training speed and inference speed as much as possible.

8.1.2 Package Installation

In this project, we use multiple packages related to picture and video. All the packages and corresponding version are listed in the following table.

Package	Version	Usage
Tensorboard	1.10	visualization of log file from tensorflow
Tensorlayer	1.10	easy setup for network construction in tensorflow
OpenEXR	1.3.2	reading and writing EXR format picture file
OpenCV	3.4.1	providing multiple basic picture process algorithm
Ilmbase	2.3.0	scientific mathematical package for EXR
Numpy	1.14	scientific mathematical package
libTIFF	4.0.9	reading and writing TIFF format picture file
FFmpeg	3.4.4-0ubuntu18.04.1	providing multiple video process methods

Table 4: Packages used in this project

For convenience, we use apt and pip in ubuntu to install those packages.

8.1.3 Tools Preparation

Besides, we also use specific tools to show EXR and HDR videos. Notice that we need show the HDR video in Windows since only Windows support HDR video and HDR screen nowadays. So, we use potplayer to view HDR videos. For EXR pictures, we could use the software “exrdisplayer” to view EXR format picture in Ubuntu system.

8.2 Model Training

8.2.1 Data Preparation

The data in this design is obtained from 4K Media. We download HDR videos in this website, and then use FFmpeg to subtract frames from videos and obtain the dataset. Finally we have over 5 thousands 4K HDR images as our dataset. The format of the dataset is TIFF image format with 12 bit depth. The resolution of all the images are 4K.

From our original data, we did data augmentation for better training result. We generate HDR-LDR pairs as input of our deep learning model. In order to increase the robustness of model, we use multiple methods such as changing different brightness, hue, and other parameters related to picture. We also crop a small size of picture from the 4k picture to satisfy our network requirements.

8.2.2 Model Training

To get a better performance of model, we adjust superparameters and different detailed method. After several trials, we found that when learning rate is 0.00005 and optimizer is Adam, we could

get a better weights for our model after 50 epochs. The activation function inside the CNN model is ReLU.

8.3 Video Conversion

After we get weights for our model, we apply our model to convert LDR picture to HDR picture. We first extract LDR pictures from LDR video through FFmpeg and input those LDR pictures to our model. Our Model can output HDR pictures in EXR format as the input order. Since FFmpeg can not handle EXR format file correctly, we need to convert EXR file into TIFF format and apply SMPTE ST 2084 standard. Then we can use FFmpeg to add HDR10 meta data and compile as HDR video. Our video could display on HDR screen correctly.

9 Test Result

In the engineering specification, we require that our program can reconstruct HDR images from single exposure in Linux environment. In Section 8, we accomplished this tasks with the running time at 3s per Meta Pixel which is better than the required value.

We show the histogram of the color space in inverse Tone Mapping method and our recovered HDR image. In Figure 9, we find that three colors in our reconstructed HDR image display is more close to a normal distribution, which satisfies the ground truth.

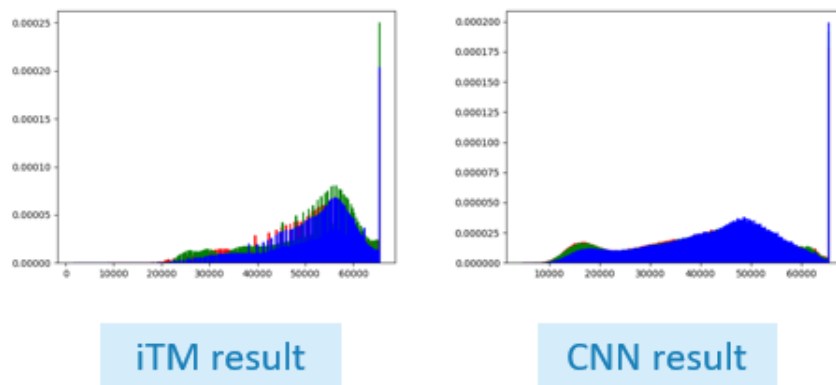


Figure 9: Comparison of the Histogram of HDR Image

One of the tasks in the engineering specification is to evaluate our output quantitatively. We

use the Visual Difference Predictor to evaluate the output result. The Visual Difference Predictor calculate the difference between the illuminance and color vectors of two images numerically. After computing the difference region and the similarity between two images, the predictor output a numerical score. The score is in the range of [0,100]. 100 indicates two images are identical. We input the reconstructed images and the ground truth to the Visual Difference Predictor to get an evaluation score.

We compare the HDR results produced by the convolutional neural networks with the result produced by traditional inverse tone mapping (iTM) methods. One sample comparison test is showed in the Figure 10. FigureThe result shows that there are larger red areas in the visual predictor of iTM image. Noticeably, our method generates mainly blue prediction, which shows a more accurate reconstruction.

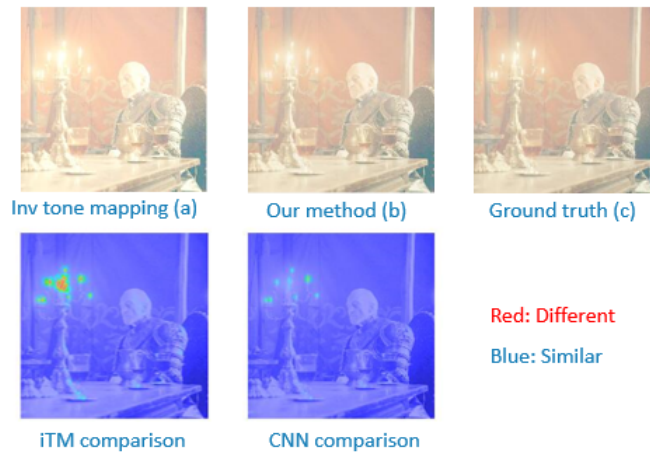


Figure 10: The Visual Difference Predictor Comparison Test

We then do the difference prediction for all reconstructed images one image by the other, the average evaluation score of our method is 82. In comparison, the traditional method scores only 75. This shows that we have a better vision and prediction than the tradition method, which fulfill the engineering specification.

10 Engineering Changes Notice

Compared to Design Review 3, we added support to SMPTE ST 2084 PQ curve. The reason was that HDR10 standard requires that curve, and we wanted to support HDR10 standard. We changed the script that converts images to video, so that the images follow HDR10 standard.

11 Discussion

Overall, we have satisfied all customer requirements. The main strength of our design is that our method can recover HDR videos from single-exposure videos, and we have achieved better vision than traditional inverse tone mapping. What's more, our program is able to handle videos/images of different resolutions automatically. The generated HDR videos can correctly display on HDR devices because the videos satisfy HDR10 standard.

However, there are also some weaknesses. First, although our program can reconstruct HDR images at a reasonable speed (≤ 3 seconds per mega pixels), real-time reconstruction still cannot be achieved. It is because the size of the neural network is large. Second, we only support HDR 10 standard. Different displays may use different HDR standards, so the reconstructed HDR videos can not be correctly showed on every HDR display. If we had more time, we could have added support for other HDR standards, such as HLG and Dolby vision, so that customers can choose their preferred standard to display the HDR videos on their devices correctly. We can do that by studying the other HDR standards on the official websites, and then add the scripts that convert the generated HDR images to the images conforming to the standards.

12 Recommendations

For environment setup, tensorflow need to be installed. In order to get high computational ability, we strongly recommend to use the newest version of GPU supporting package and tools. Notice that tensorflow from pip source only support very limited version of CUDA and CUDNN. So we use bazel to compile source code of tensorflow to install. The newest version usually provide a higher computational ability. But we also suggest to care about the version of CUDA and CUDNN that tensorflow could support.

For training step, we use data augmentation to increase the complexity of our training data set. In this way, we can increase the robustness of our model, which means we could reconstruct videos under different conditions. In this step, a little bit destruction is allowed. But heavy destruction may lead the model couldn't converge.

Another thing we need to notice is how to convert EXR picture into HDR10 standard. HDR10

video follows BT 2020 color space and SMPTE ST 2084 PQ curve. In this step, what we need to apply the formula is the luminance value in the real world. So we need to first convert EXR file back to the luminance value in the real world.

13 Conclusions

13.1 Our Progress

In this project, our main objective is to implement deep learning method and reconstruct HDR video from single exposure LDR video. We implement a Convolutional Neural Network based on the VGG16 network structure. We developed the QFD according to our customer requirements and generated a series of concepts to satisfy our engineering specifications. We manage to get better HDR reconstruction results than previous conventional inverse tone mapping method and verify our model's performance by the Visual Difference Predictor. We then use ffmpeg to merge the frames into the videos. We display our reconstructed HDR videos on HDR screen and achieve better vision result.

In this course, we have accomplished the process of design and manufacturing. We have learnt machine learning and deep learning. There are four main advantages our project have achieved. To begin with, we have achieved better vision than the reconstructed image produced by inverse tone mapping. This progress is verified by the Visual Difference Predictor Evaluation Test. Next, our reconstruction method is based on a single exposure. Traditional methods requires capturing multiple exposures for one scene, which is not applicable to the LDR frames extracted from the LDR videos. Then, the fully convolutional network (FCN) in the CNN network structure enables the reconstruction of videos and images of different resolutions. Thus, our method is not restricted with the size of the image. Finally, we study the SMPTE curves and HDR video standards. Our reconstructed HDR videos satisfies the HDR10 standard. Therefore, the videos can be played correctly on HDR devices.

13.2 Future work

There are also some future work to do. We put forward a new idea to improve the training speed. We can apply the ResNet to the CNN network. This strategy is expected to reduce the times

and improve the training efficiency.

Currently, the HDR standards are experiencing rapid development. There are three main widely-used standards. They are HDR10, HLG, and Dolby vision. We plan to add options to make our HDR videos cater to different standards in the future work. So customers can choose standard type for their preferred HDR devices.

14 Information Sources and Reference List

References

- [1] “Compositing Multiple Pictures of the Same Scene”, by Steve Mann, in IS&T’s 46th Annual Conference, Cambridge, Massachusetts, May 9–14, 1993
- [2] Allan G. Rempel, Matthew Trentacoste, Helge Seetzen, H. David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3), 2007. Article 39.
- [3] Belen Masia, Sandra Agustin, Roland W. Fleming, Olga Sorkine, and Diego Gutierrez. Evaluation of reverse tone mapping through varying exposure conditions. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 28(5):160:1–160:8, 2009.
- [4] Hany Farid. Blind inverse gamma correction. *IEEE Transactions on Image Processing*, pages 1428–1433, 2001
- [5] ITU-R, High dynamic range television for production and international programme exchange, Report ITU-R BT.2390-3, 2017.
- [6] Karol Myszkowski, Rafal Mantiuk, Grzegorz Krawczyk. *High Dynamic Range Video*, 2012.
- [7] Marnierides, D., Bashford-Rogers, T., Hatchett, J. and Debattista, K. (2018). ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. [online] Arxiv.org. Available at: <https://arxiv.org/abs/1803.02266> [Accessed 12 Dec. 2018].
- [8] Simonyan, K. and Zisserman, A. (2018). Very Deep Convolutional Networks for Large-Scale Image Recognition. [online] Arxiv.org. Available at: <https://arxiv.org/abs/1409.1556> [Accessed 12 Dec. 2018].
- [9] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), Natick, MA, USA, 2011.

- [10] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K. Mantiuk, Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. In: *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia 2017)*, 36(6), Article 178, 2017. Available at: http://hdri.org/hdrcnn/material/sga17_paper.pdf..[Accessed 17 Oct. 2018]
- [11] Marnierides D, Bashford-Rogers T, Hatchett J, Debattista K. ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range In: *39th Annual Conference of the European Association for Computer Graphics (EUROGRAPHICS)*, 37(2), Page 37-49, 2018. Available at: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/cgf.13340>.[Accessed 17 Oct. 2018]
- [12] Stephen Lin, Jinwei Gu, Shuntaro Yamazaki, and Heung-Yeung Shum. Radiometric calibration from a single image. *Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 02:938–945, 2004.
- [13] D. Grossberg and S. K. Nayar. 2003. What is the space of camera response functions?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '03)*., Vol. 2. II–602–9.
- [14] Allan G. Rempel, Matthew Trentacoste, Helge Seetzen, H. David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3), 2007. Article 39.
- [15] Ahmet Ouz Akyüz, Erik Reinhard, Roland Fleming, Bernhard E. Riecke, and Heinrich H. Bühlhoff. Do HDR displays support LDR content? a psychophysical evaluation. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3), 2007.
- [16] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Inverse tone mapping. In *18th Eurographics Symposium on Rendering*, pages 321–326, 2006.
- [17] Seetzen, H. Li, L. Ye, W. Heidrich, L. Whitehead, and G. Ward. 25.3: Observations of luminance, contrast and amplitude resolution of displays. In *SID 06 Digest*, pages 1229–1233, 2006.

- [18] Akiko Yoshida, Rafa Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum (Proc. of EUROGRAPHICS)*, 25(3):415–426, 2006.
- [19] Yongqing Huo, Fan Yang and Chao Li, "HDR image generation from LDR image with highlight removal,"2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, 2015, pp. 1-5.
- [20] Laurence Meylan, Scott Daly, and Sabine Susstrunk. The reproduction of specular highlights on high dynamic range displays. In *Proc. of the 14th Color Imaging Conference*, 2006.
- [21] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk and J. Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics*, Vol. 36, No. 6, Article 178. Publication date: November 2017.
- [22] Mystery Box. (2018). HDR Video Part 1: What is HDR Video. [online] Available at: <https://www.mysterybox.us/blog/2016/10/18/hdr-video-part-1-what-is-hdr-video> [Accessed 28 Nov. 2018].

15 Appendices

15.1 Bill of Materials for System

Item	Quantity	Source	Cost(RMB)	Contact
Solid-state disk	1	Taobao.com	338.97	021-54740000

Table 5: The bill of materials.

15.2 The Network Structure of VGG16

Here is a structure of the VGG16 network structure in our design. It is made up of an encoder and a decoder.

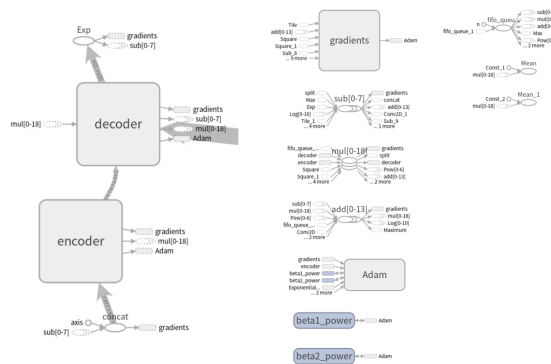


Figure 11: The structure of VGG16 network

15.3 Inverse Tone Mapping Method

The inverse Tone Mapping method we used is based on robertson. In order to use this method, we type “-hdrModel robertson” in the command line.

16 Acknowledgement

We are grateful to everyone who has directly or indirectly contributed to our project. We express our genuine gratitude to in the following paragraph.

Firstly, we express our thankfulness towards our sponsors, Juan ZHAO, Ruiling SONG, Yejun GUO, not only for their valuable instructions and detailed suggestions, but for patience, time and energy devoted to this project. They have met us on a weekly basis listening to our progress and answering our questions in detail. In order to ensure the performance and outcome of our project, they have encouraged and supported us to use state-of-the-art machine learning method. They guided us to use the Deep Convolutional Neural Network and implemented VGG16 neural network. Thanks to they devotion and patience, we have accomplished our project successfully.

Secondly, we express our gratitude to Professor Yong LONG. who has offered us indispensable advice throughout our project. In the weekly meeting, she advised us to keep our project on progress and instructed us how to communicate with the sponsor professionally. She also gave important instructions on how to manage the image format transformation. There's no denying that the instruction and advice form Professor Long is a crucial factor to the success of this project.

Thirdly, we express the supreme appreciation to support from the faulty and staff from UM-SJTU Joint Institute. Thanks to their technical suggestions and useful guideline, we manage to fulfill our final design with outstanding performance.