Real-time distortion classification in laparoscopic videos by BUET_ENDGAME

Introduction:

ICIP 2020 introduces a challenge on distortion classification which is providing 5 major distortion (awgn, defocus_blur, motion_blur, smoke, uneven_illum) videos, and using these primary distortions, there are 5 more distortion videos have been created. To classify these higher order distortions, raw images have not been fed into the network. Rather, we extracted features first, then these extracted features have been taken as input in our model. So, we have approached signal processing to deduce the noisy features available in videos.

Procedure:

A. Data extraction

For extraction of frames from videos, we have used a free and open-source project **FFMPEG.** We use FFMPEG to extract frames in lossless data compression format PNG(Portable Network Graphics).

B. Data preparing

The videos contain around 240 frames on average. We take only **10 frames** from every video keeping a 20 consecutive frames gap. As we need to evaluate the performance of our model, we splitted out the whole dataset into 3 portions: 60% training, 20% validation and the rest 20% is a test set which is unseen data for our network.

For the final evaluation, we take a 70% dataset for training and 30% for validation.

Methodology

We divide our strategy in two ways:

- **Noisy feature extraction**

As distortion exists in these videos, any neural networking model can hardly fit the noisy pattern. So, noisy features should be extracted. There are 5 major noise issues. So, we've made 5 noisy models, which return the relative severity with respect to every frame.

- ❖ AWGN estimation

For determining these noise, fast noise variance estimation has been used. The advantage of this method is that it includes a Laplacian operation which is almost insensitive to image structure but only depends on the noise in the image. It returns the standard deviation of the noise, which is a vital parameter of a noise to identify.

- ❖ Defocus blur

Whether an image is blurry or not, a Laplacian variance parameter is an important factor to identify.

- ❖ Motion Blur

Motion blur is the apparent streaking of moving objects in a photograph or a sequence of frames. In the sense of a spatial domain, there is some motion or movement in X or Y direction, where there is smoothness for consecutive few pixels. So, we try to deconvolve in vertical and horizontal direction to measure the severity of motion blur.

- ❖ Smoke detection

Smoke detection is performed according to Saturation Peak Analysis and histogram equalization.

- ❖ Uneven illumination

Uneven illumination occurs when the light does not evenly illuminate your sample across the field of view results in darker areas and more brightly illuminated areas. However, to detect uneven illumination luminance means range plays a better role.

Model Architecture

Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 10, 128)	2048
batch_normalization_1 (Batch	(None, 10, 128)	512
spatial_dropout1d_1 (Spatial	(None, 10, 128)	0
seq_self_attention_1 (SeqSel	(None, 10, 128)	8257
lstm_1 (LSTM)	(None, 64)	49408
batch_normalization_2 (Batch	(None, 64)	256
dense_1 (Dense)	(None, 128)	8320
dropout_1 (Dropout)	(None, 128)	0
batch_normalization_3 (Batch	(None, 128)	512
dense_2 (Dense)	(None, 64)	8256

dropout_2 (Dropout)   (None, 64)   0

---

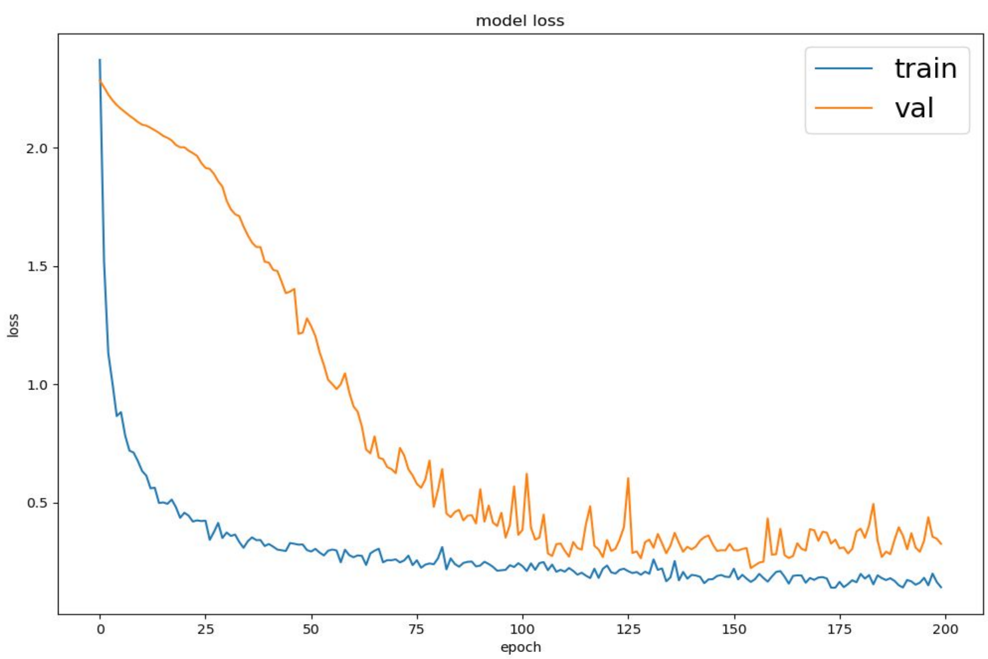
batch_normalization_4 (Batch (None, 64)   256

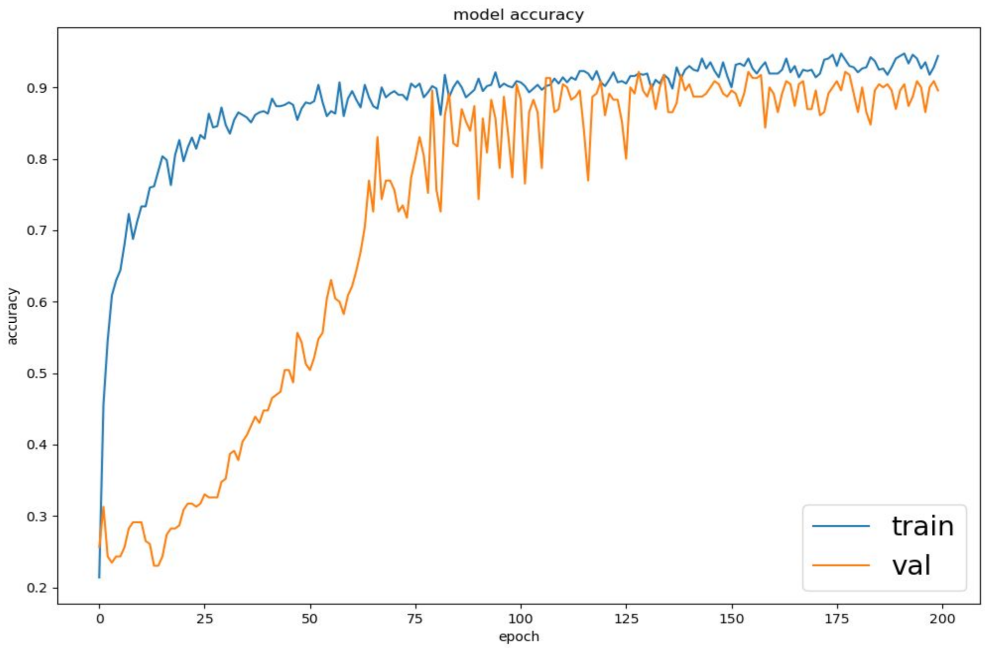
---

dense_3 (Dense)   (None, 10)   650
===
Total params: 78,475
Trainable params: 77,707
Non-trainable params: 768

OFFLINE TRAINING

For the final evaluation process, we split the whole dataset into 70% training dataset and 30% test dataset. During training process we get these loss and accuracy curve:





Results

As we said earlier, to evaluate our model performance we have separated **60% as training, 20% as validation and 20% as unseen test data(160 videos)** from the provided training dataset to evaluate the algorithm efficiency.
During training procedure we get **training loss: 0.1322, training accuracy: 0.9429, validation loss: 0.2209 - validation accuracy: 0.9400** as categorical classification accuracy metrics.
Confusion matrix on 20% unseen data:

Confusion matrix on 160 unseen videos