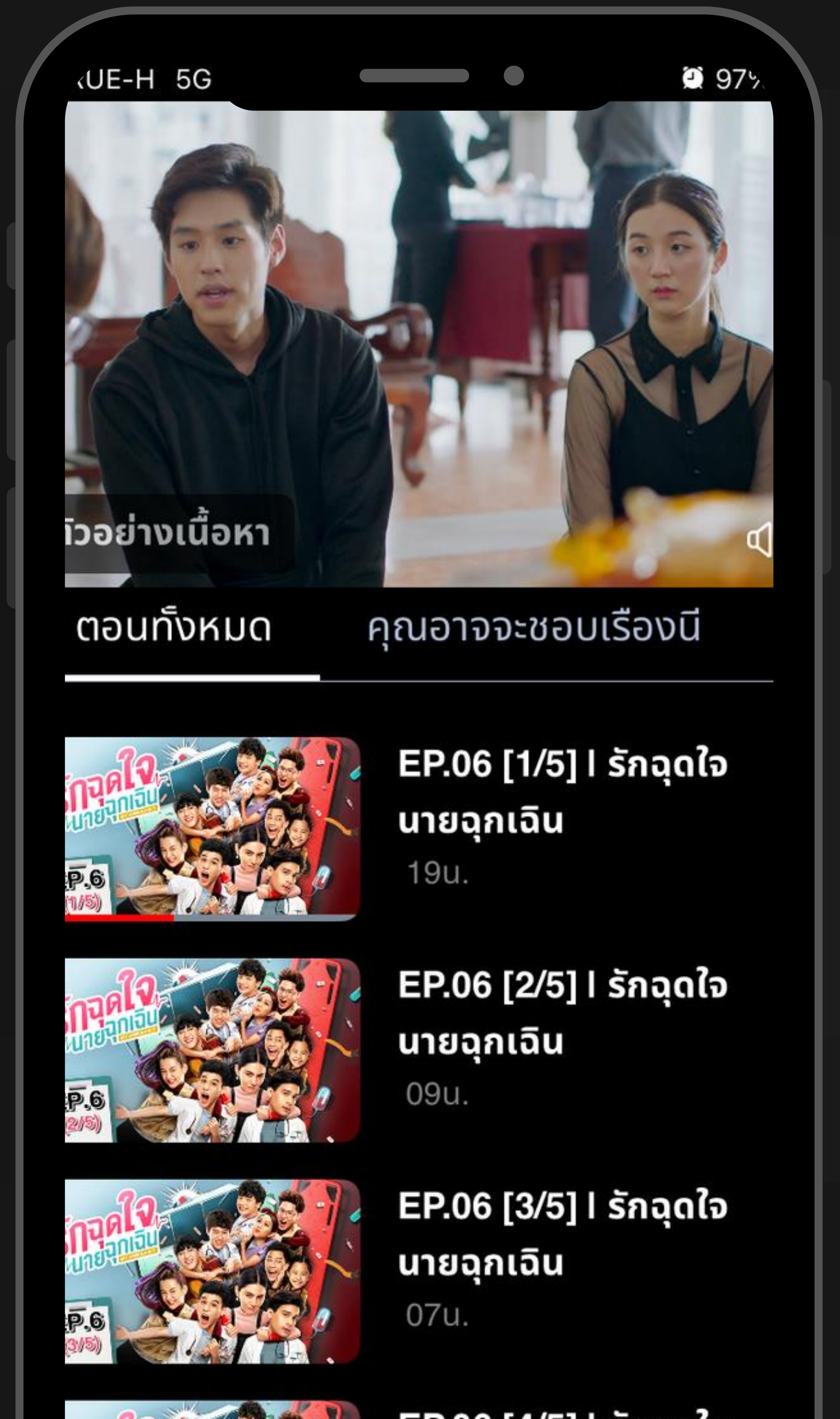
The Capstone project

SCENE SEARCH

AI FOR VIDEO

Ponparis Gurdsapsri


UE-H 5G
97%
ัวอย่างเนื้อหา
ตอนทั้งหมด
คุณอาจจะชอบเรื่องนี้
EP.06 [1/5] l รักฉุดใจนายฉุกเฉิน
19น.
EP.06 [2/5] l รักฉุดใจนายฉุกเฉิน
09น.
EP.06 [3/5] l รักฉุดใจนายฉุกเฉิน
07น.

Problem statement

Many studies show that **"thumbnail"** is one of the key factor that potentially create more interest for user to click on the content. It can also be used to target different segment for the same content.

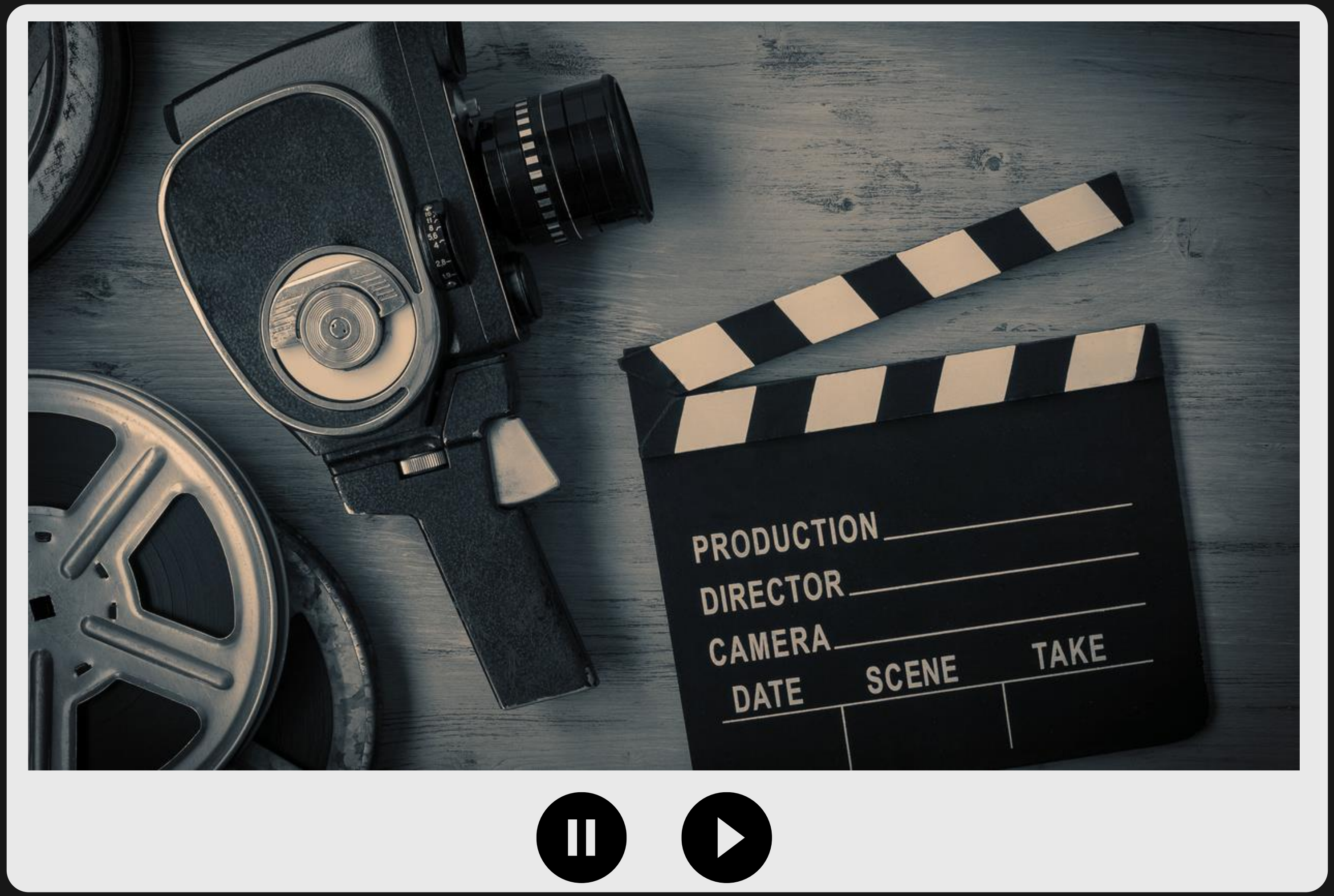
If we know the content, we may have some idea where the specific scene is, in the content.

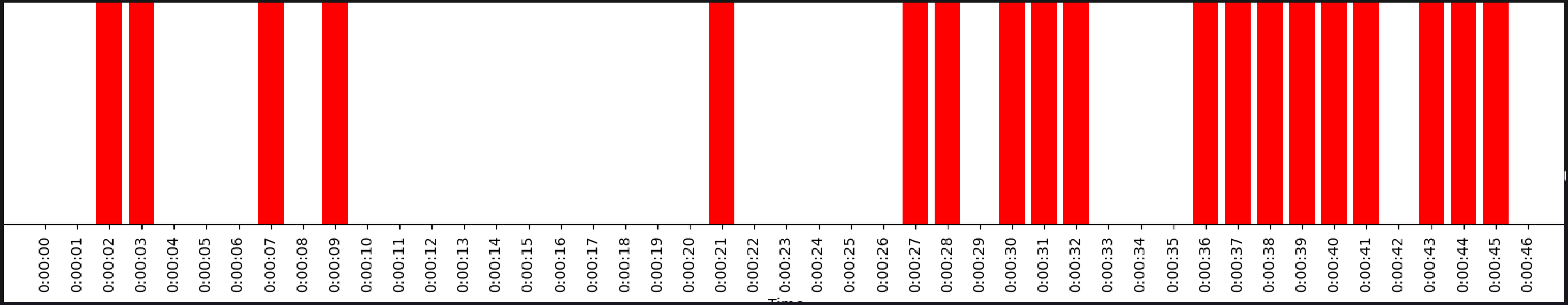
What if there are thousands of contents? It probably takes eternity to select the desired thumbnail for every content.

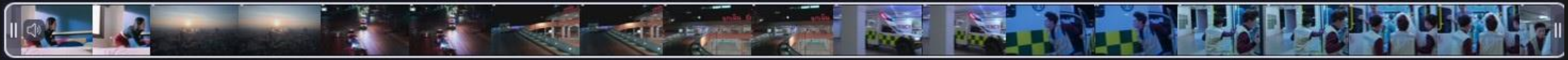
Project Objective

To build AI model to detect and locate where the specified scene is on the subject video

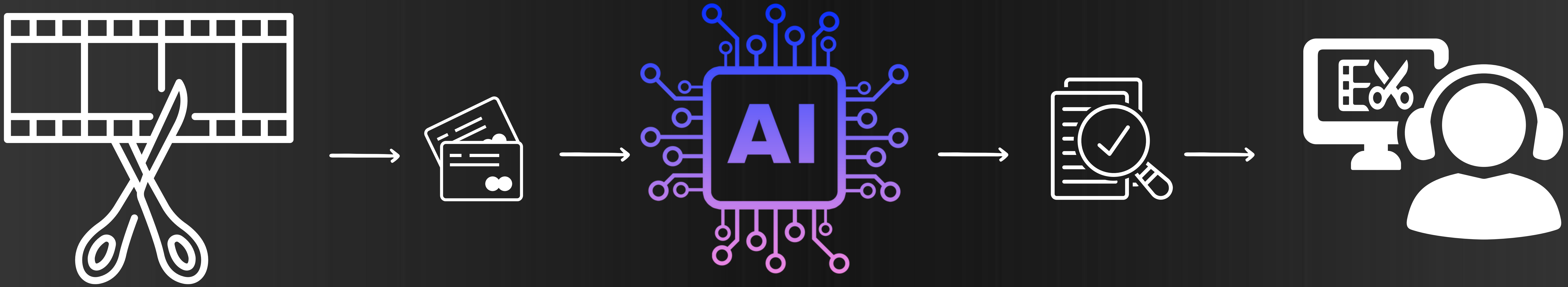
Demonstration


PRODUCTION
DIRECTOR
CAMERA
DATE
SCENE
TAKE


0:00:00
0:00:01
0:00:02
0:00:03
0:00:04
0:00:05
0:00:06
0:00:07
0:00:08
0:00:09
0:00:10
0:00:11
0:00:12
0:00:13
0:00:14
0:00:15
0:00:16
0:00:17
0:00:18
0:00:19
0:00:20
0:00:21
0:00:22
0:00:23
0:00:24
0:00:25
0:00:26
0:00:27
0:00:28
0:00:29
0:00:30
0:00:31
0:00:32
0:00:33
0:00:34
0:00:35
0:00:36
0:00:37
0:00:38
0:00:39
0:00:40
0:00:41
0:00:42
0:00:43
0:00:44
0:00:45
0:00:46



How it work



Convert to image

Video is converted to image by chopping frames of the video

AI prediction

AI model will provide the prediction, whether the image has ambulance scene or not

Result summary

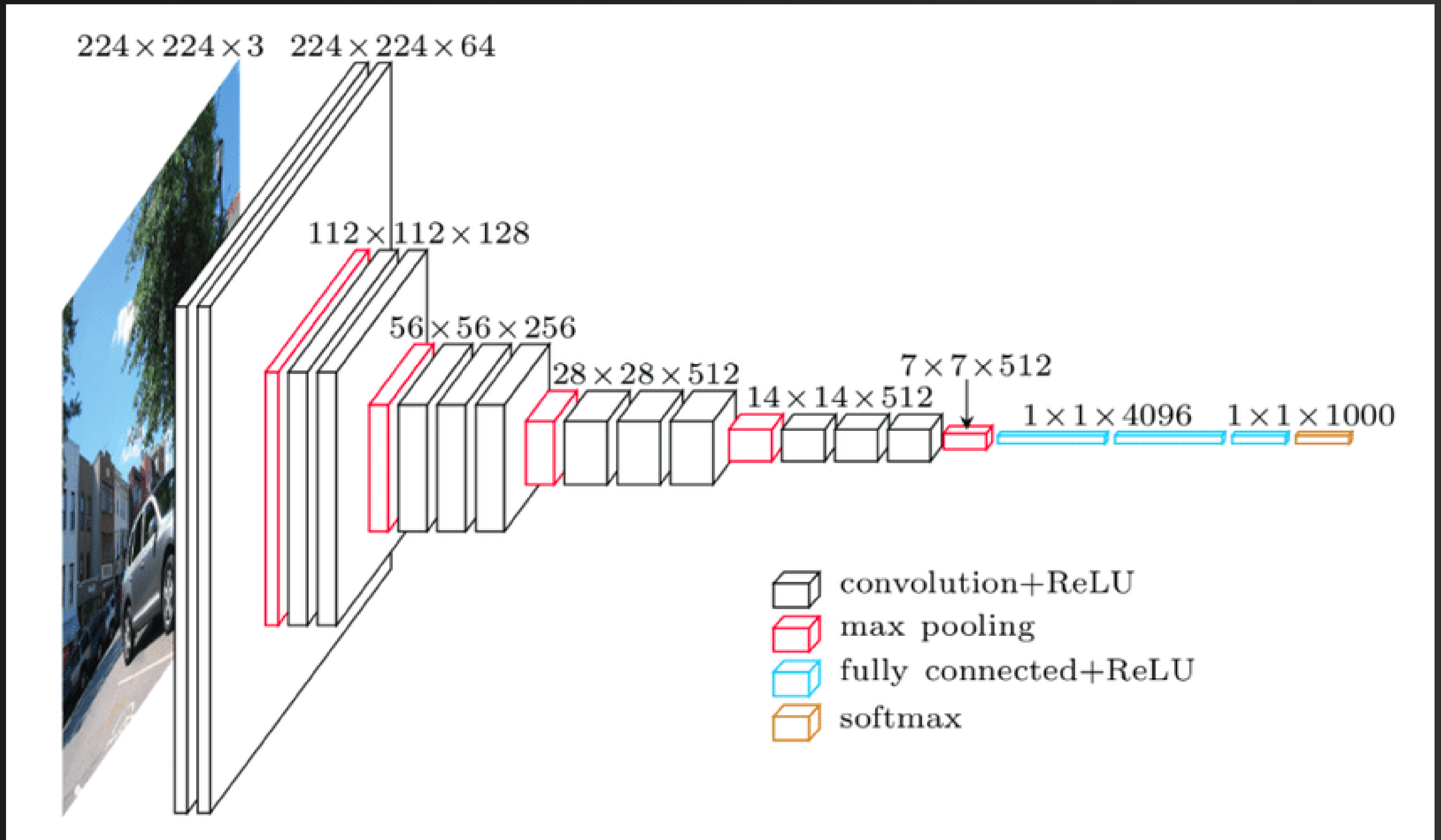
The prediction will then be summarized and display which second of video has ambulance

Data science model behind AI

The prediction model is based on the utilization of existing pre-trained image classification model called “VGG-16”. The model was trained by millions of images.

Because the model is primary trained by western society, so the model need to be tuned (localized) in order for the model to provide high accuracy prediction.

So trained data and fine-tuning the model is done during the transfer knowledge of the model.

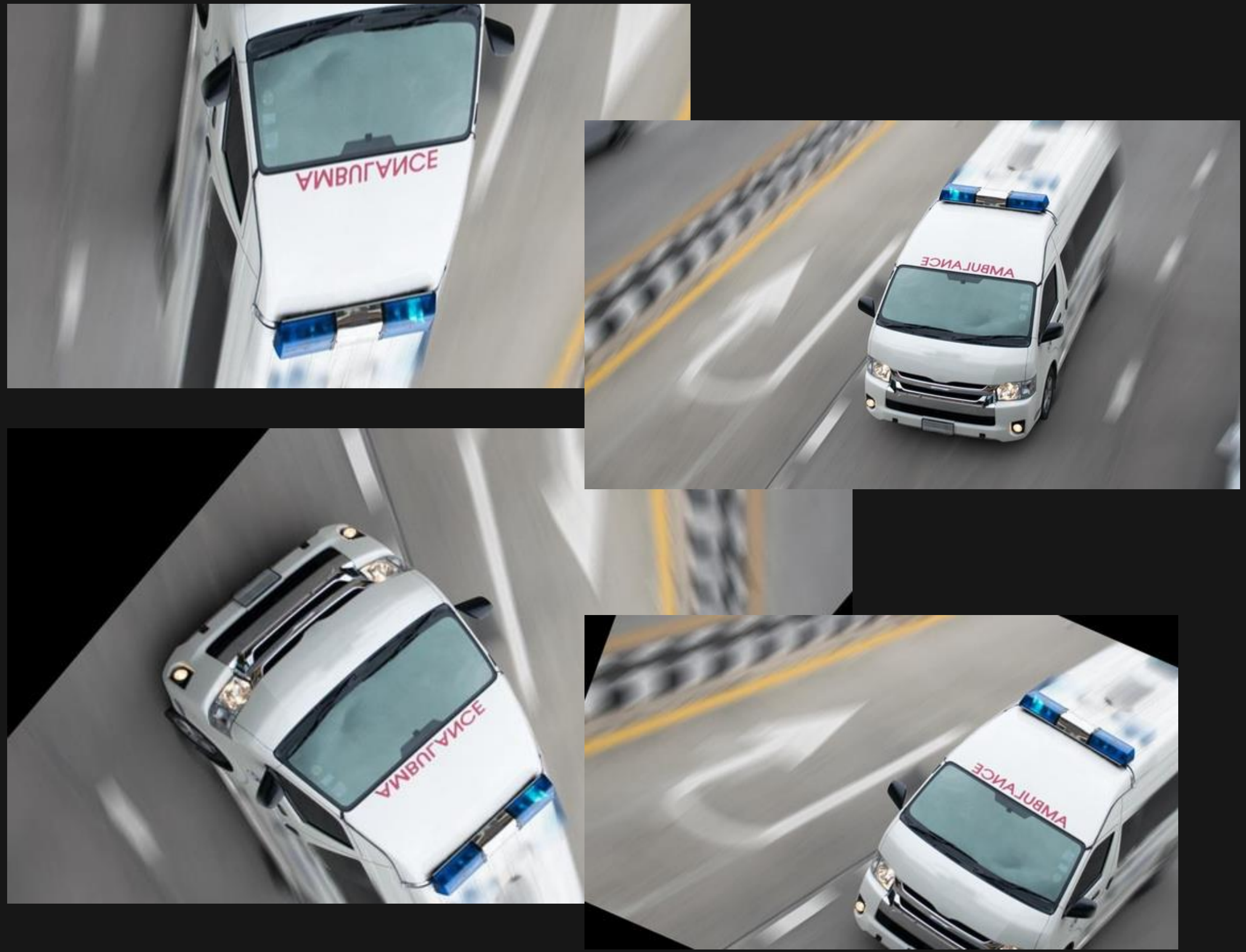


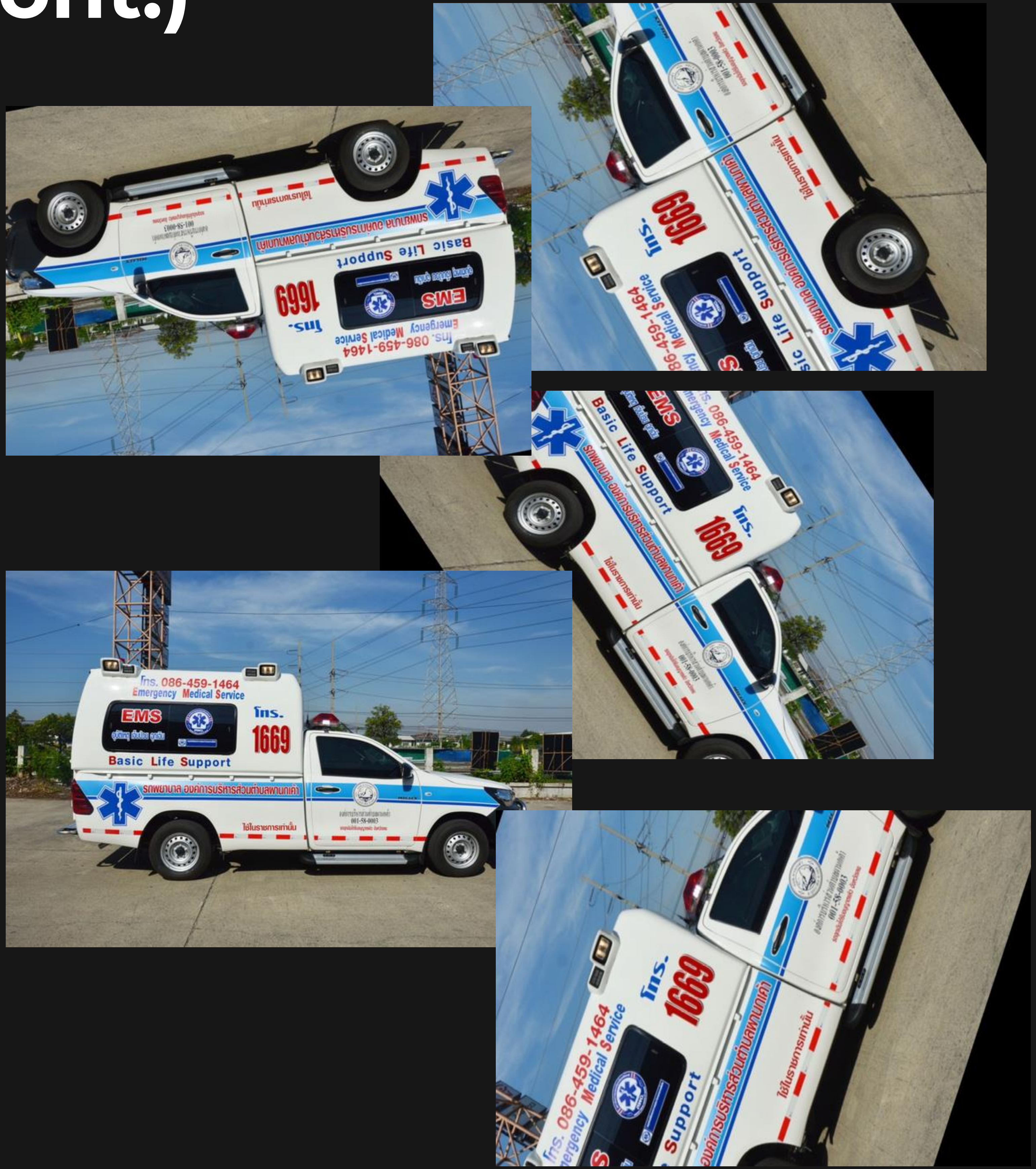
Data used

Common ambulance
in US and UK

Common ambulance
in Thailand

Data used (cont.)





More train data would lead to better performance, but due to limitation resource. The images were augmented in order to expand size of train data

Model evaluation

3 Models were created and record the accuracy score for evaluation

- Model-1 were added 3 more neural layers and trained by non-augmented images. Total 251 images
- Model-2 were model-1 and trained by augmented and original images. Total 13,420 images
- Model-3 were model-2 but re-trained 5 more layers

Model 3 has least overfitting and best overall result. So model 3 will be used for test the sample video

	Train accuracy	**Validate accuracy**	**Test accuracy**
Model 1	64.8%	58.3%	25%
Model 2	49.8%	51.2%	50%
Model 3	51.2%	51.5%	50%

Train image were split 90-10 for train-validate data. Additional test images were also used, this is unseen images.

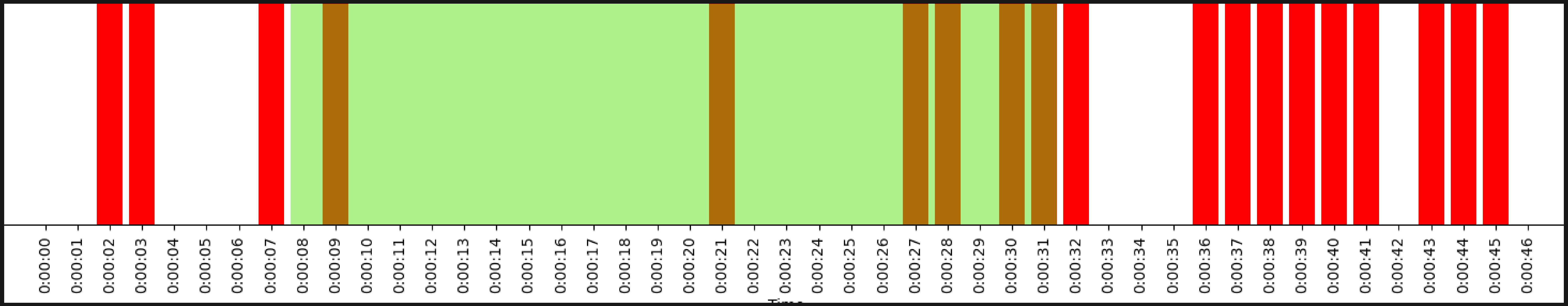
Result

Test video 1

🟩 Actual ambulance appearance

🟥 Predicted ambulance appearance

The accuracy of this prediction is only 34%



Test video 2

Result (cont.)

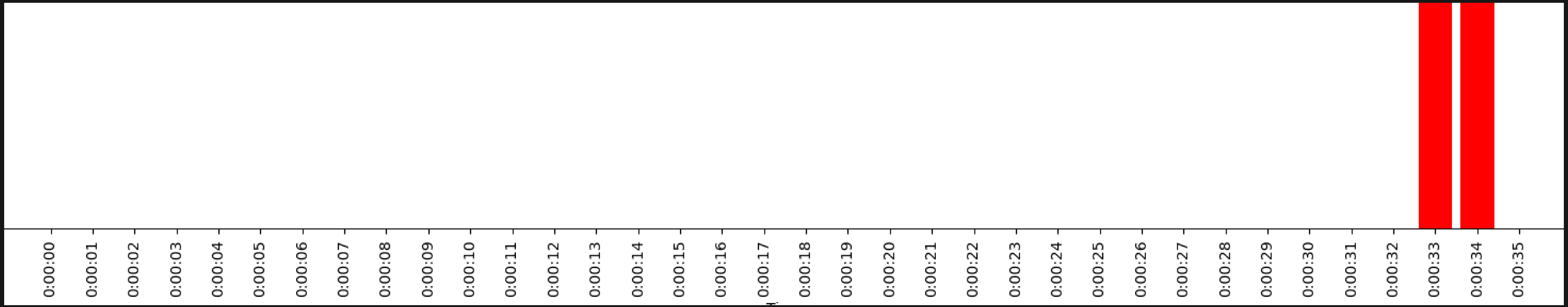
Result (cont.)

Test video 2

🟩 Actual ambulance appearance (no ambulance)

🟥 Predicted ambulance appearance

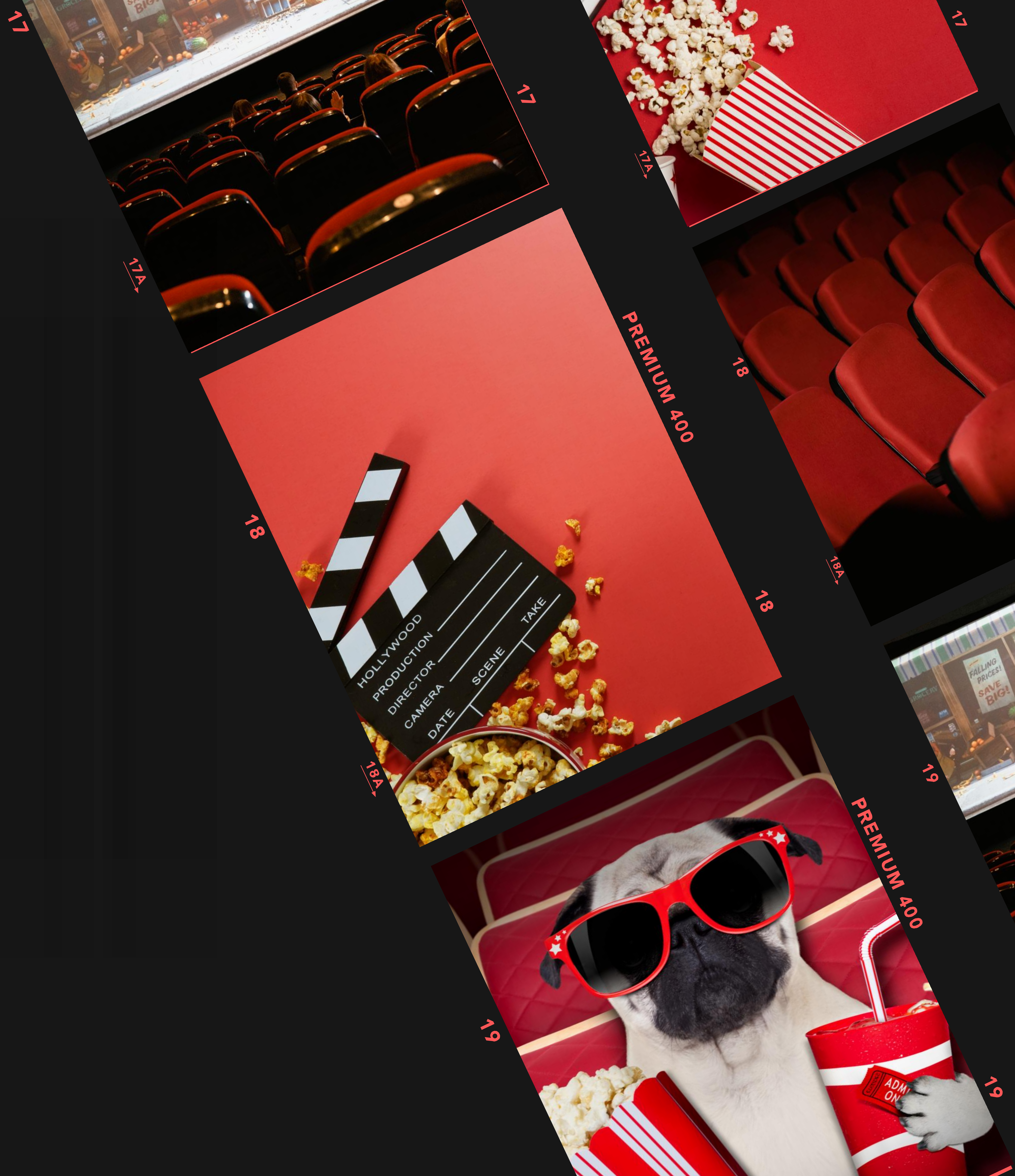
The accuracy of this prediction is 94%



Conclusion

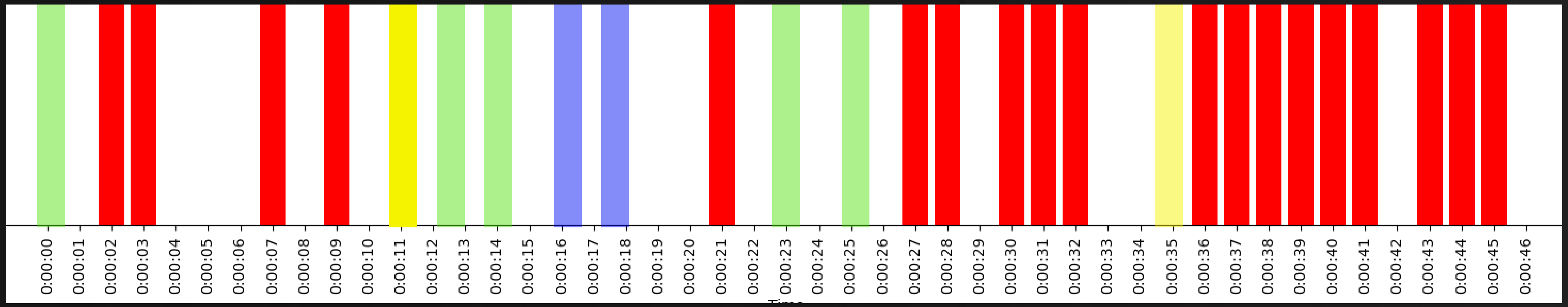
Model shows poor prediction accuracy in the night scene. It could be due to the lack of ambulance at night, because the model provide excellent prediction in the day-light scene, as for test video 2.

Train data of specific subject is very important for model training. In order to improve the model, the subject images in different condition and angle, much more images is required.



Other implication

☑ Genre classification



The model can be implement to find other scene, for example kissing or fighting scene. These scenes can be identified and be calculated to classify whether what genre should this content be. For example, the calculation shows result that it is 40% romantic, 20% action, 15% comedy.

The weight can also be stored as meta data in the content and can be later used for curation.

THANKS

FOR YOUR ATTENTION

Ponparis Gurdsapsri

ponparis_gur@truecorp.co.th