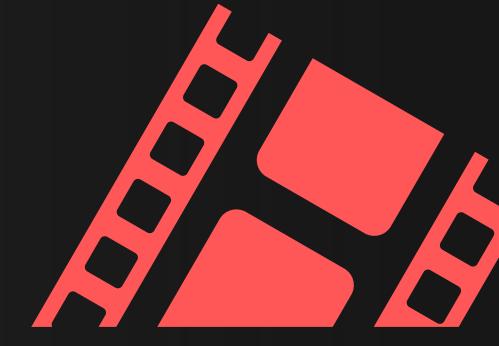


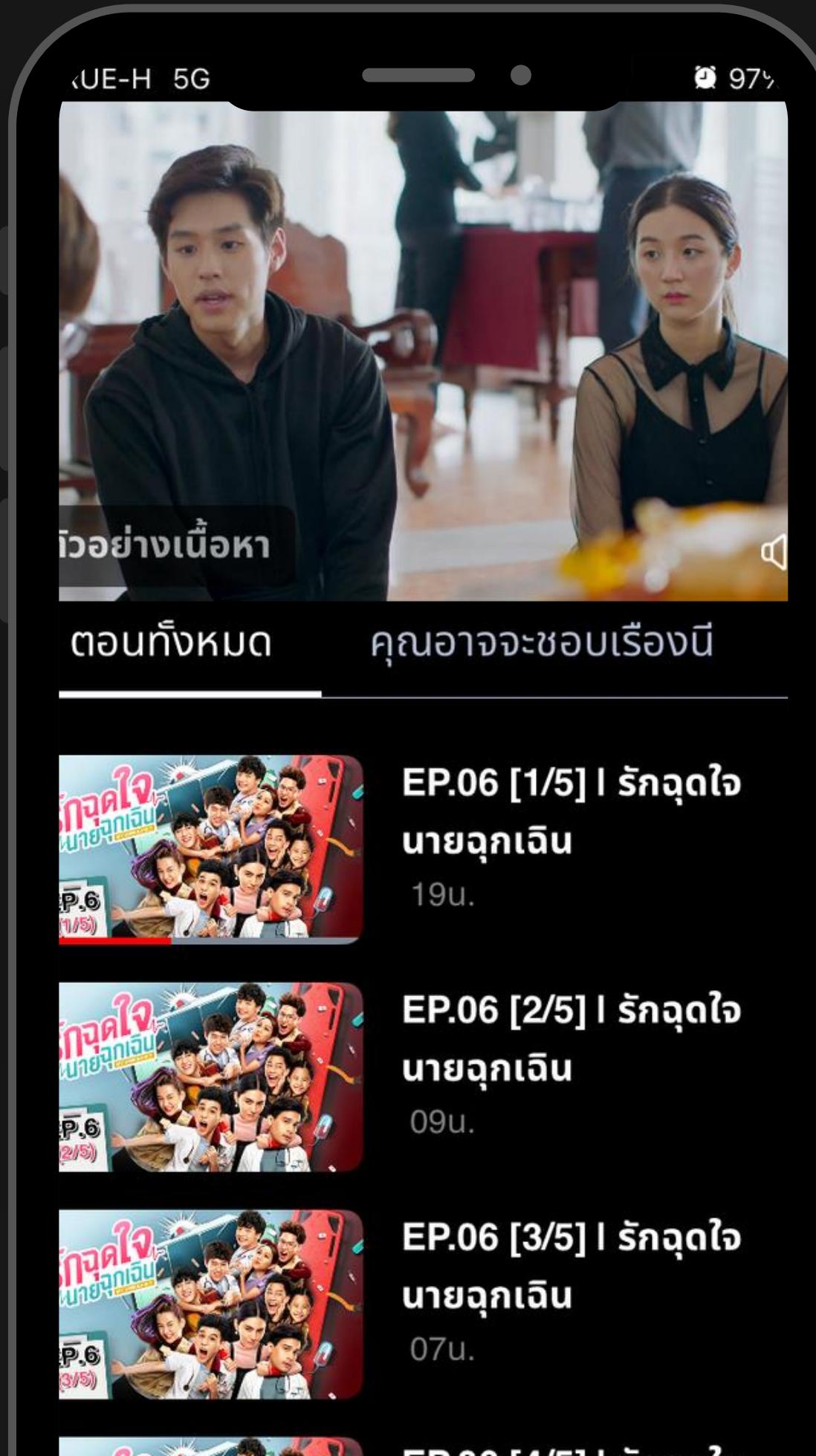
The Capstone project



# SCENE SEARCH

AI FOR VIDEO

Ponparis Gurdapsri



# 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

---

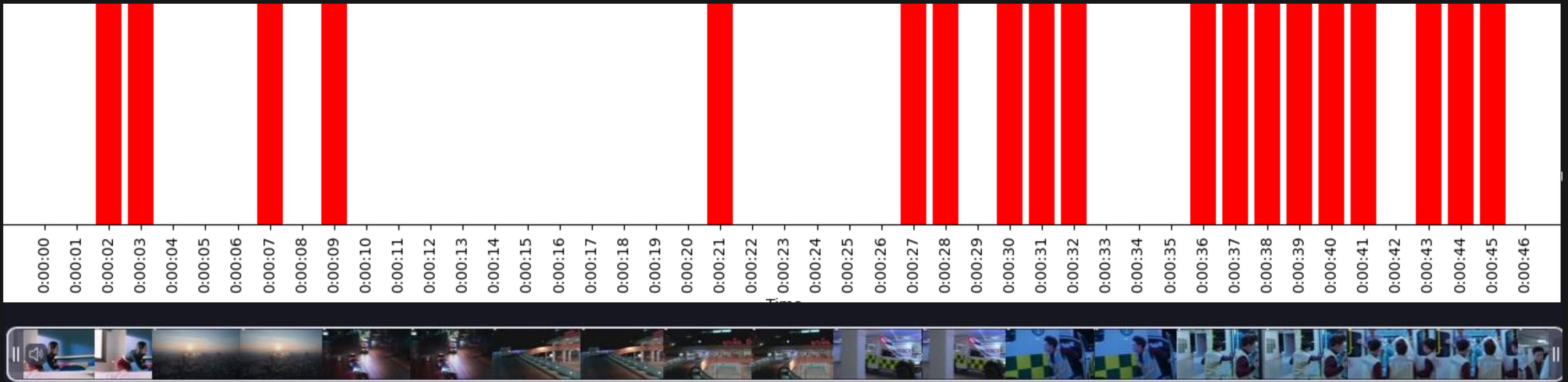
Develop proof of concept of AI model that detect and locate where the specified scene is on the subject video .





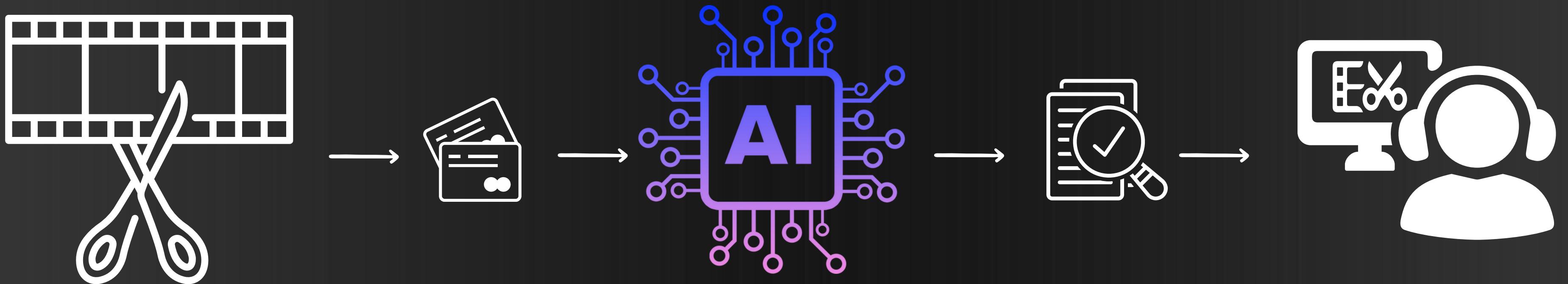
# Project scope

Model will predict and locate  
“ambulance” appearance on  
scene.





# How it work



## Convert to image

Video is converted to image by chopping frames of the video

## AI prediction

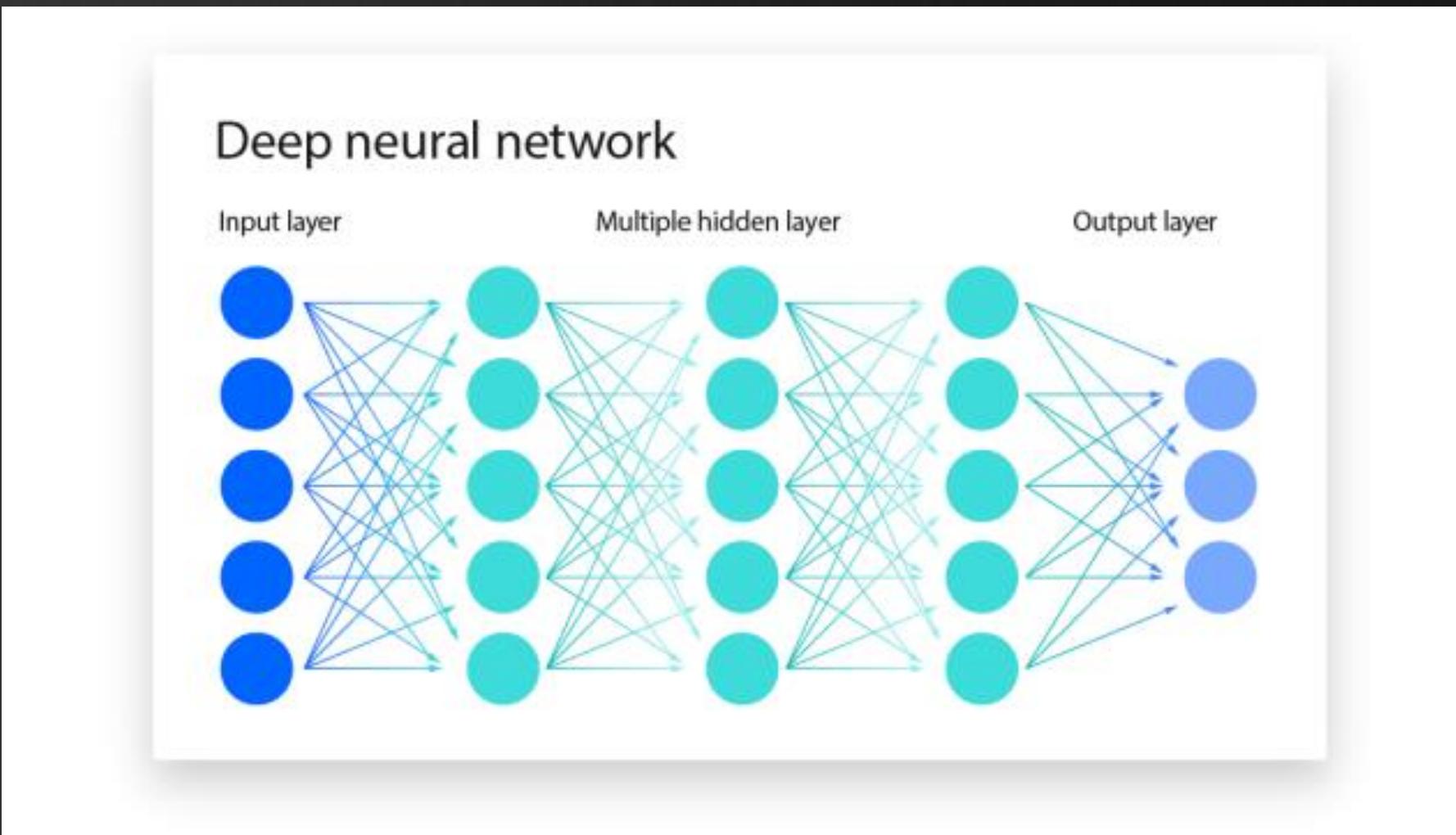
AI model will provide the prediction, whether the image has ambulance scene or not

## Result summary

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



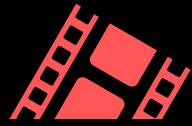
# Data science model behind AI



This image classification was based on Data science technique call “deep learning” under the fundamental of neural network.

Neural network consist of interconnected nodes, organized into multiple layers. The interconnected nodes are connected through weighted connection. The values of each layer are the value of each nodes multiply by weight of each node, the output produces prediction based on the weighted calculation from each layers.

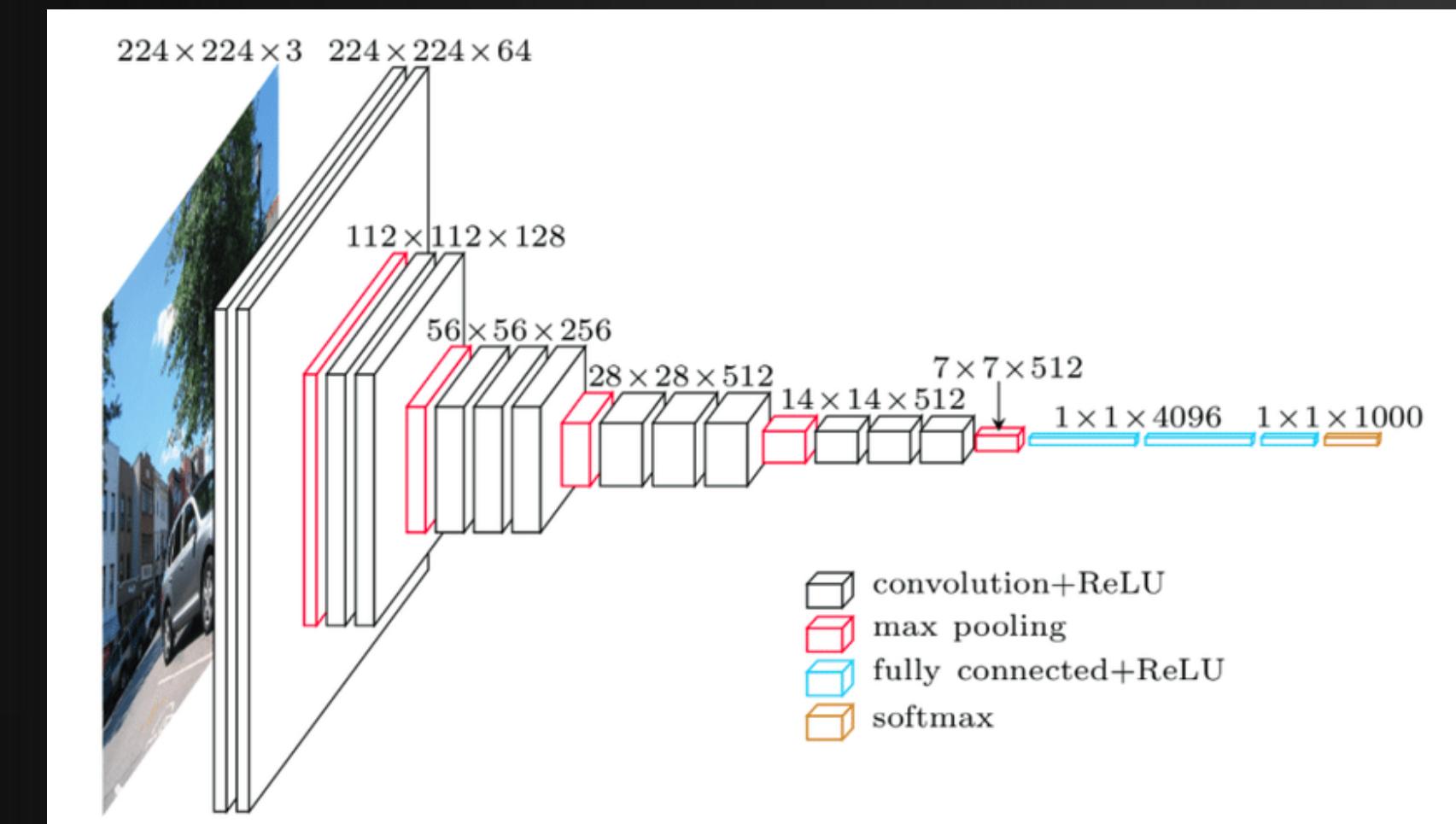
It also compare prediction with actual and send error back for weight adjustment in layers



# Model methodology

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. And introduce target sample (train data) to the model for localized model to my classification model. This process is called “knowledge transfer”.

- Acquire train data (images)
- Create additional layers to train new data
- Evaluate result
- Fine-tuning model (built 3 models)





# Data used

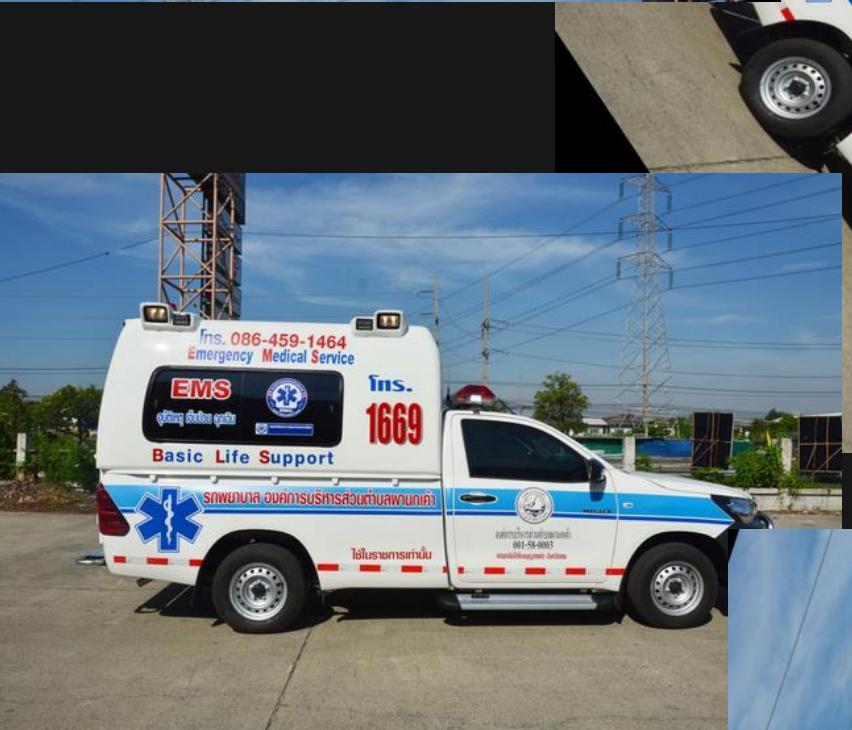


Common ambulance  
in Thailand





## Data used (cont.)



251 original were used, and augmented to 13,420 images.  
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

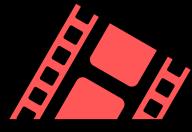
3 Models were created and record the accuracy score for evaluation

- 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-1 but re-trained 5 more layers (because performance is better)

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

	<b>Train accuracy</b>	<b>Validate accuracy</b>	<b>Test accuracy</b>
Model 1	58.36%	66.7%	75%
Model 2	73.86%	58.3%	50%
Model 3	74.24%	75.00%	75%

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



# Model evaluation (cont.)

Let's check when  
model predict  
correctly, what is the  
image look like

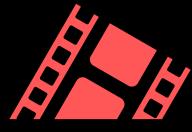


When it predict wrong



# Video Test 1

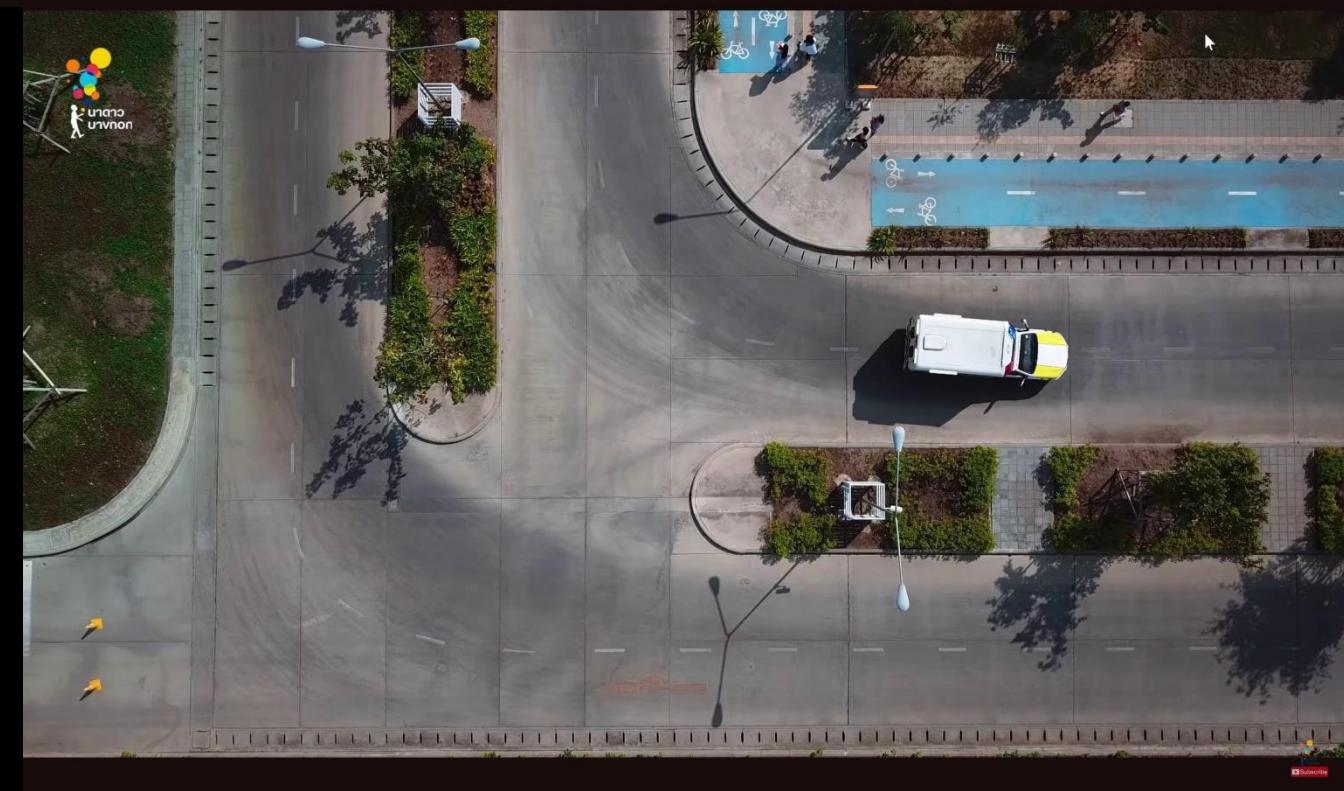
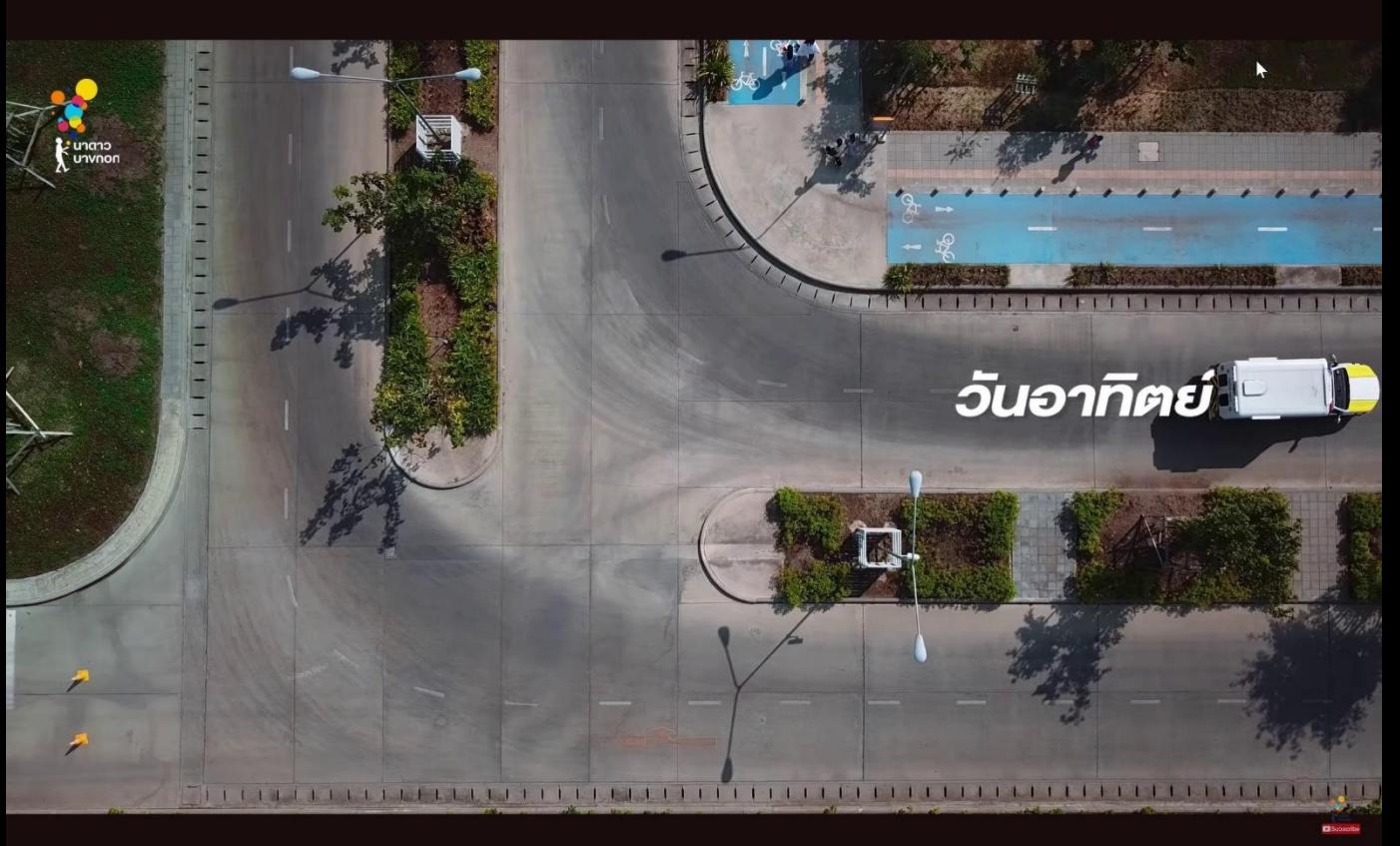
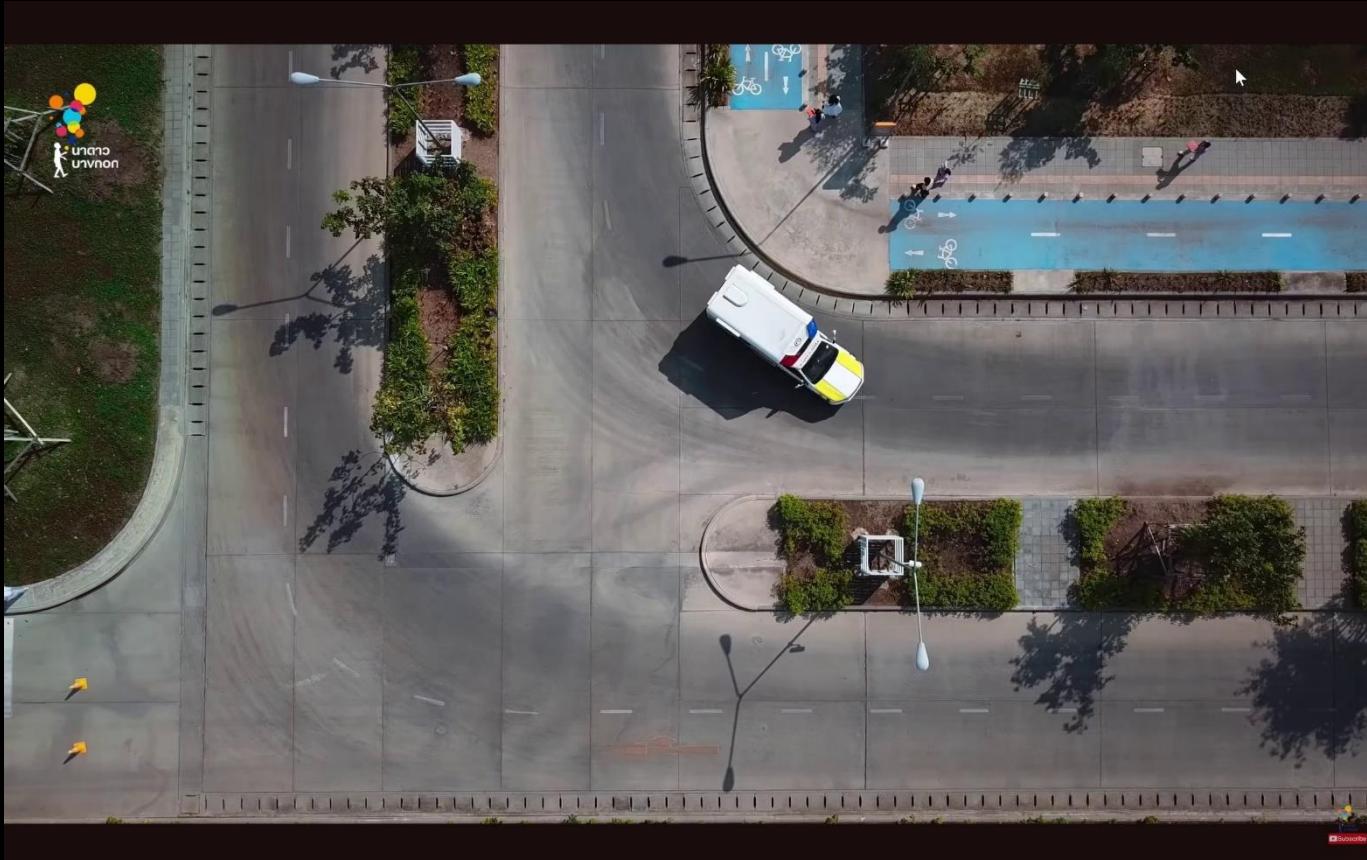




# Result

---

Actual ambulance  
scene

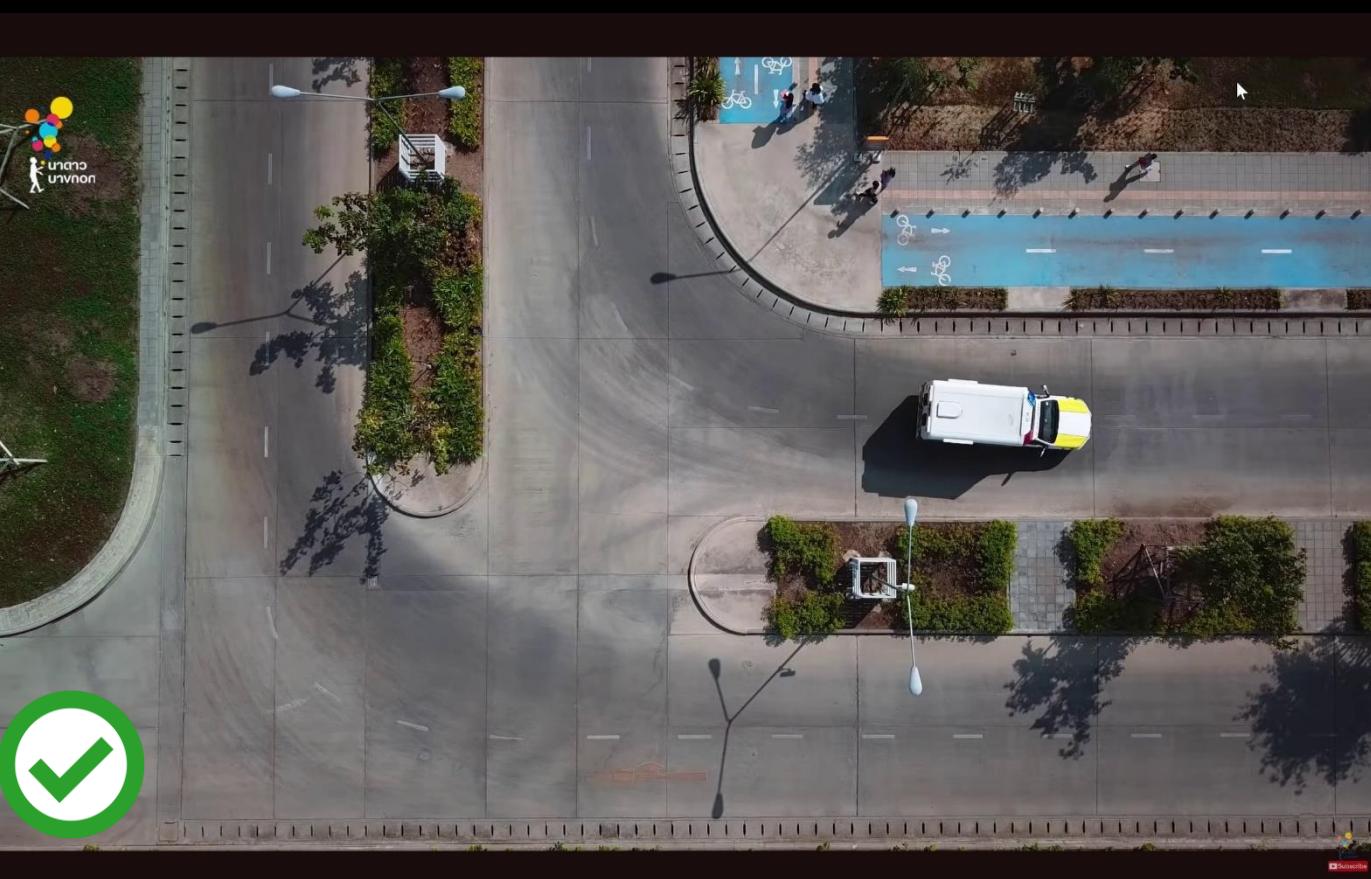
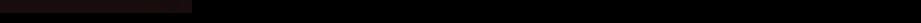
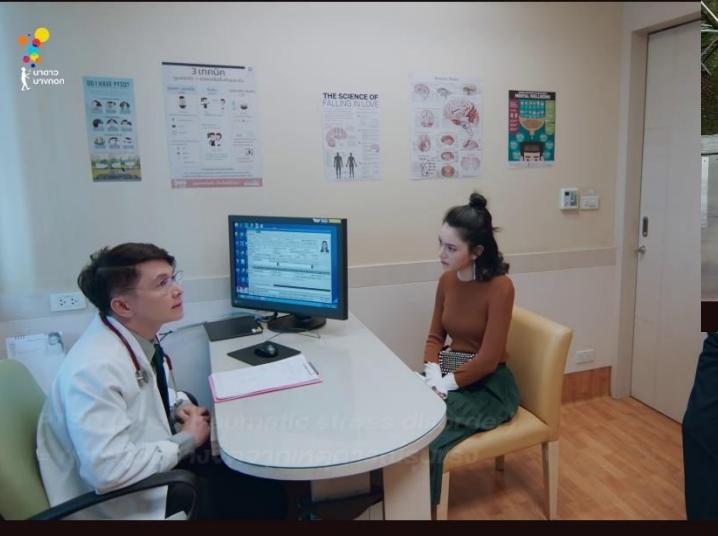
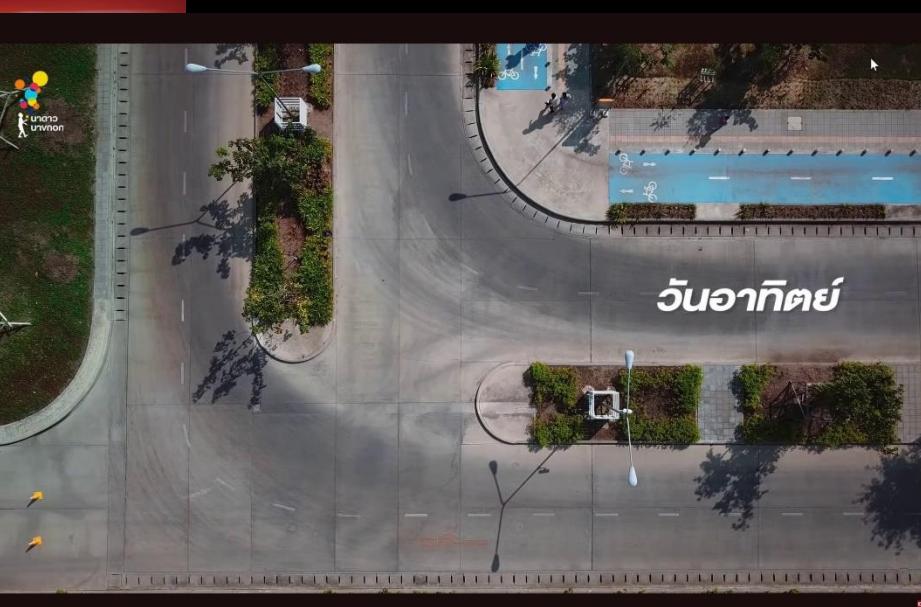


# Result

---



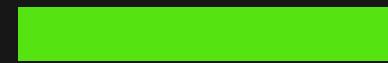
Predicted scene





# Result

Test video 1



Actual ambulance appearance



Predicted ambulance appearance

Accuracy:

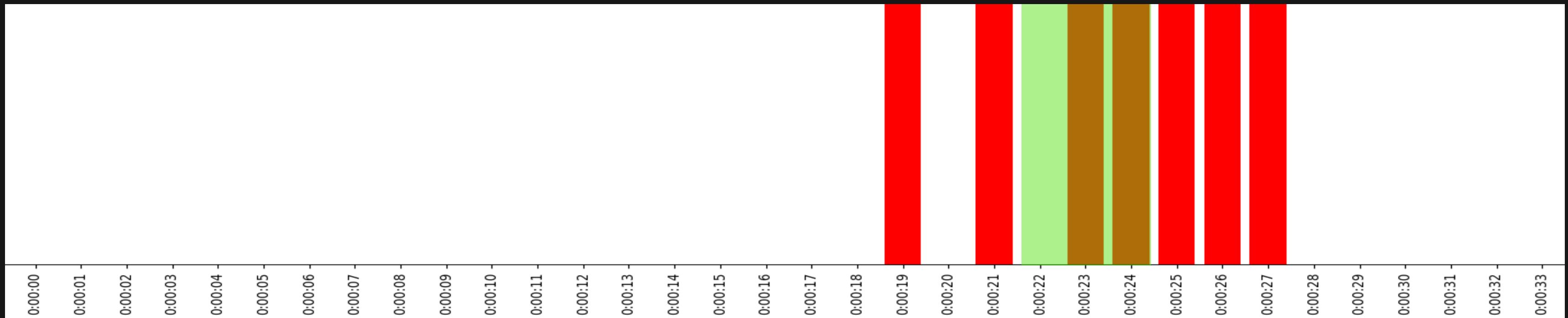
82.35%

Recall:

66..67%

Specificity:

83.87%





Test video 2

# Video Test 2





# Result

Actual ambulance  
scene



# Result

---

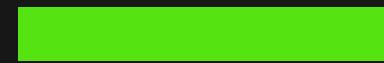
Predicted scene





# Result

Test video 2



Actual ambulance appearance (no ambulance)

Accuracy:

94.87%

Recall:

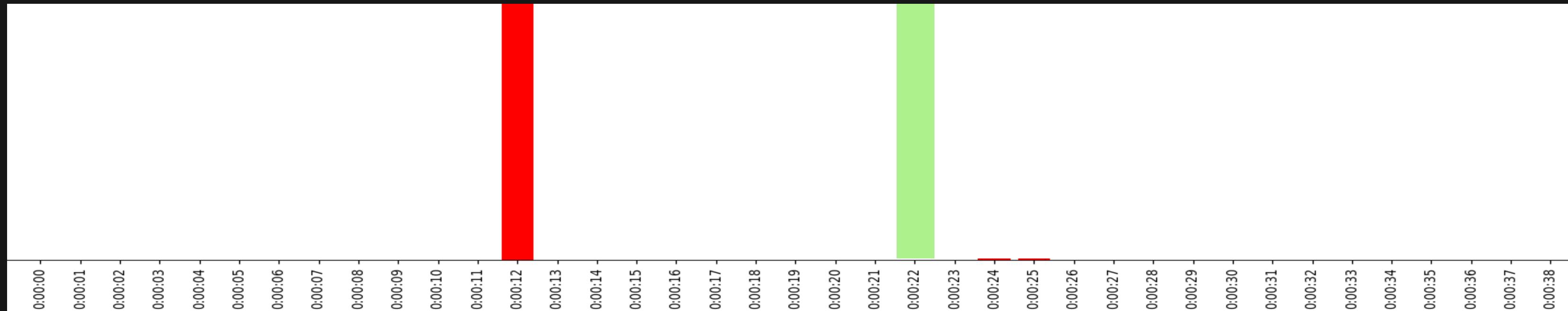
0.00%

Specificity:

97.36%



Predicted ambulance appearance

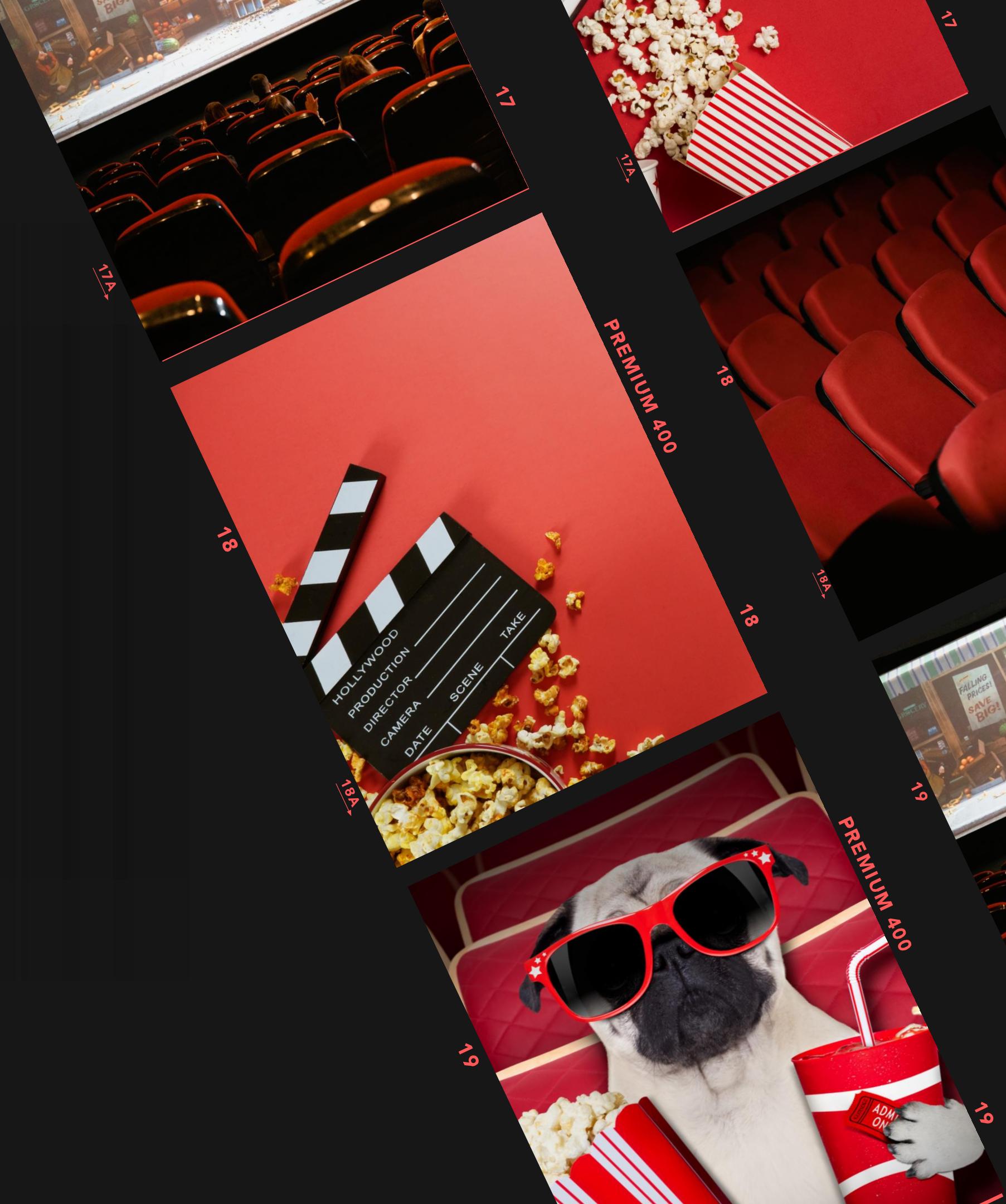




# Conclusion

Model shows decent result for video 1, but when facing difficulty sample in video 2, it has problem detecting the right scene. This scene was the rear of the ambulance.

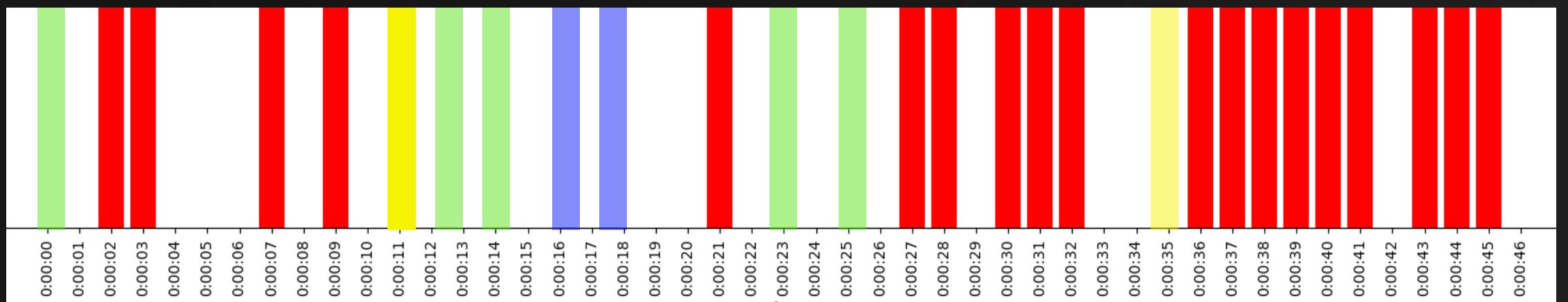
Train data of specific subject is very important for model training. In order to improve the model, the subject images in different condition and angle, much more images is required.





# Future implementation

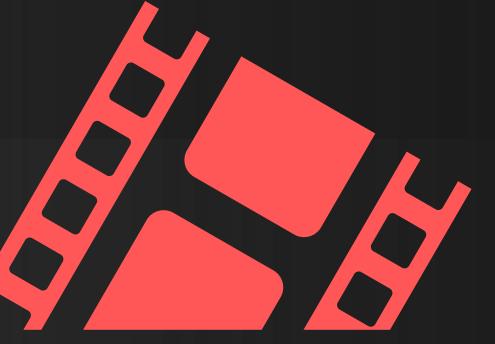
## Genre classification



The model can be implemented to find other scenes, 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.





**THANK YOU**  
**FOR YOUR ATTENTION**

*Ponparis Gurdsapsri*

---

ponparis\_gur@truecorp.co.th