GROUP 11

STREET RECOGNITION

Performance between different models & analysis

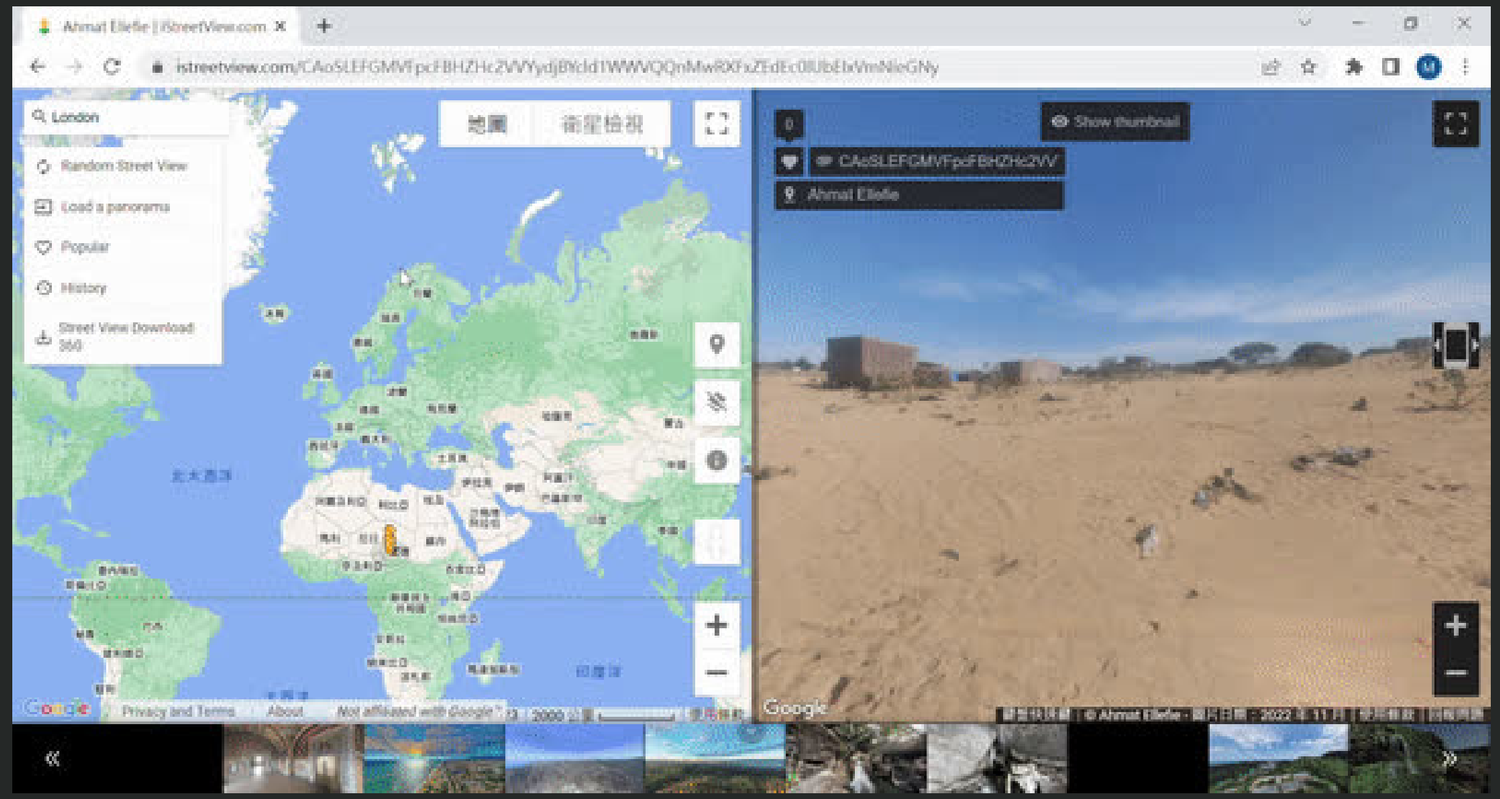
莊子郁 黃筠蘋 陳錫宏 張惇媛 黃莉婷 余伽璇

IMPLEMENTATION

DATA COLLECTION, TRAINING PROCESS

DATASET COLLECT

iStreetView.com



- recognize 4 cities

  (Taipei, Bankok, London, Washington)

- total 2000 images

  (training:1700 test:300)

- image size: 640x640
- powerful features by human view

TAIPEI

. we focus on choosing street, traffic light, crossroad:

BANGKOK

. we focus on choosing special building ,taxi and crosswalk :

LONDON

- we focus on choosing bus, traffic light and street scene :



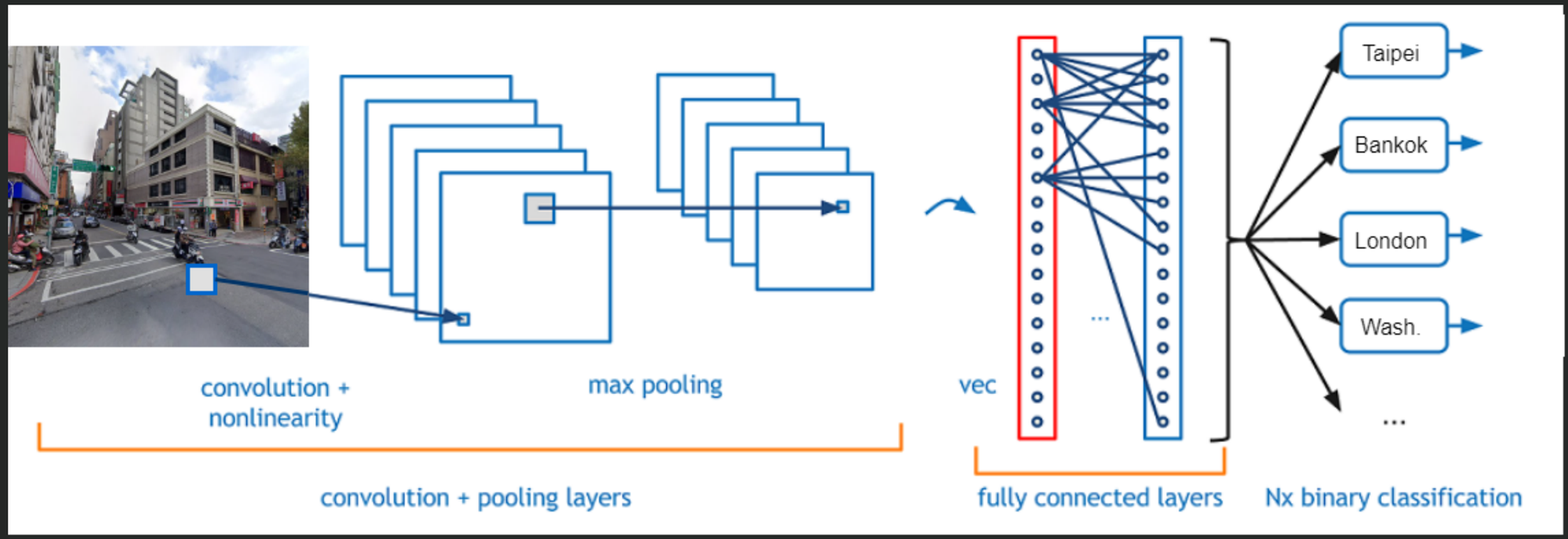

WASHINGTON

- we focus on choosing road and crosswalk :

TRAINING PROCESS

- Use CNN to classify images
- 4 models: CNN, ResNet,

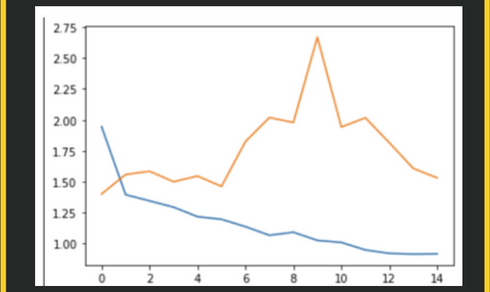
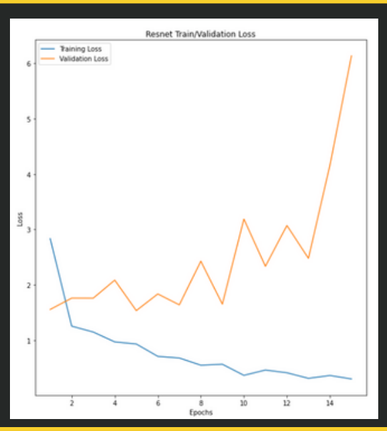
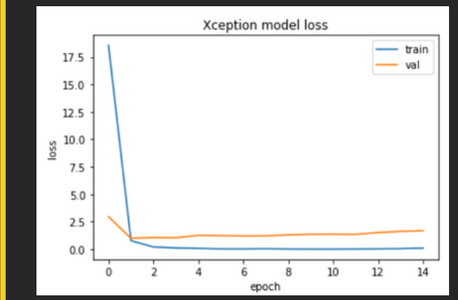
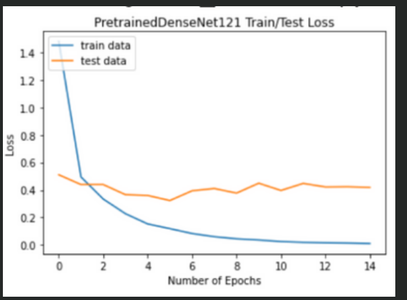
  Xception, DenseNet (pretrained)



PERFORMANCE

ACCURACY, TRAINING STEPS BETWEEN 4 MODELS

MODEL COMPARISON

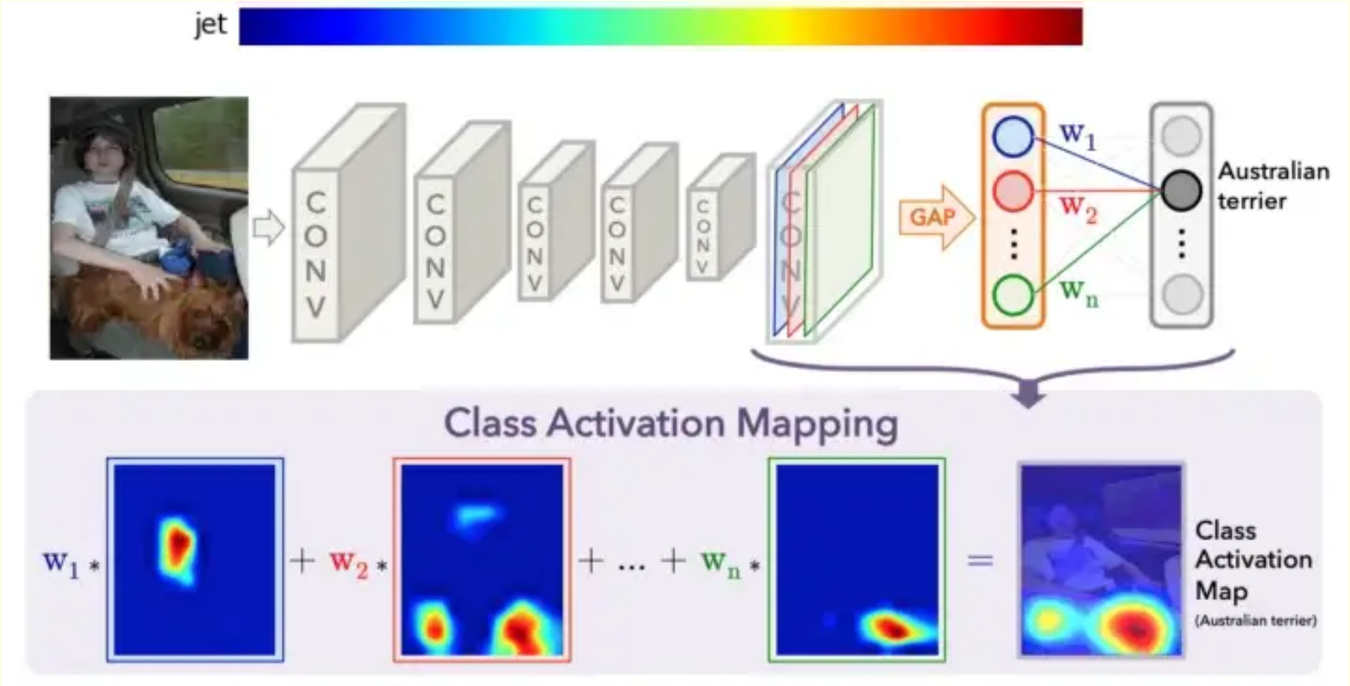
Model	CNN	Resnet18	Xception	DenseNet121
input image size	200x200	256x256	299x299	320x320
training accuracy	0.6483	0.4478	0.97	0.9994
testing accuracy	0.4	0.23	0.75	0.875
training time	480 secs	107 minutes	360 secs	180 secs
training loss curve				

VISUALIZATION

USING HEATMAP TO VISUALIZE

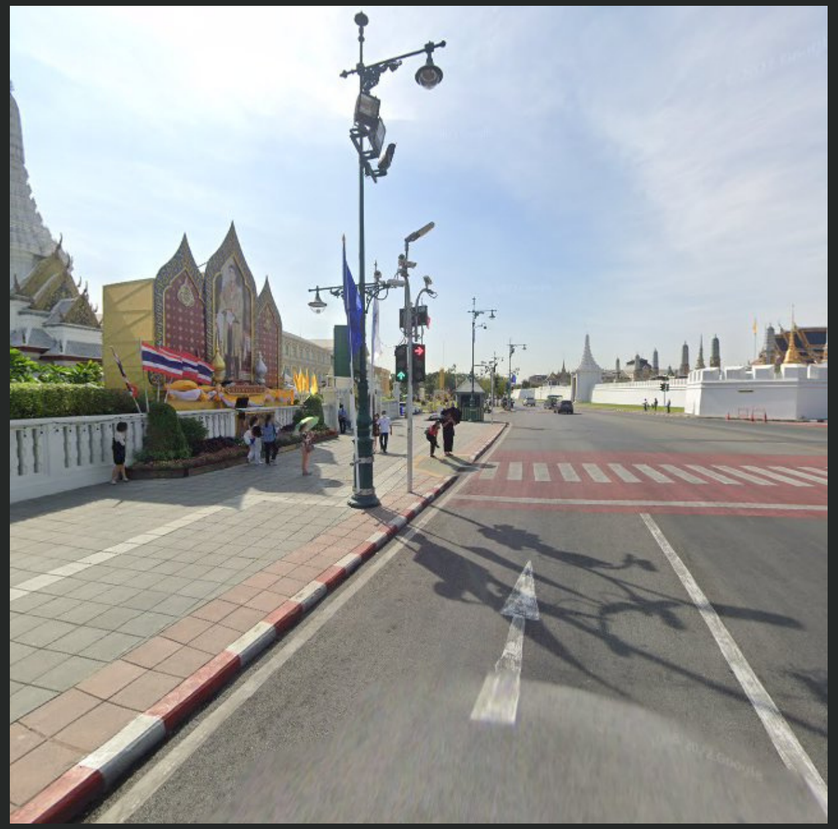
GRAD-CAM

- Visualize neural network.
- The larger the weight is, the more important its feature map is.
- Use "jet" as our colormap.

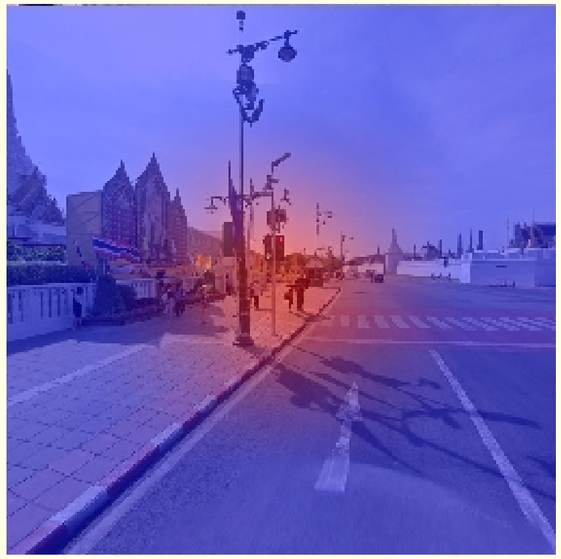


RESULT BETWEEN DIFFERENT MODELS

BANGKOK



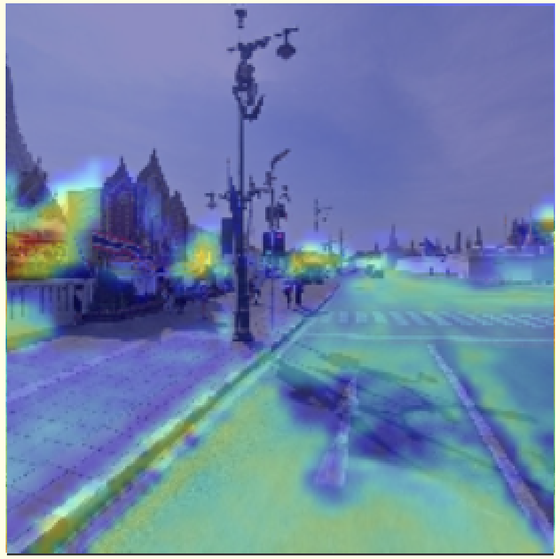
Source image



ResNet
(predict: taipei)

Xception
(predict: bangkok)

DenseNet
(predict: bangkok)



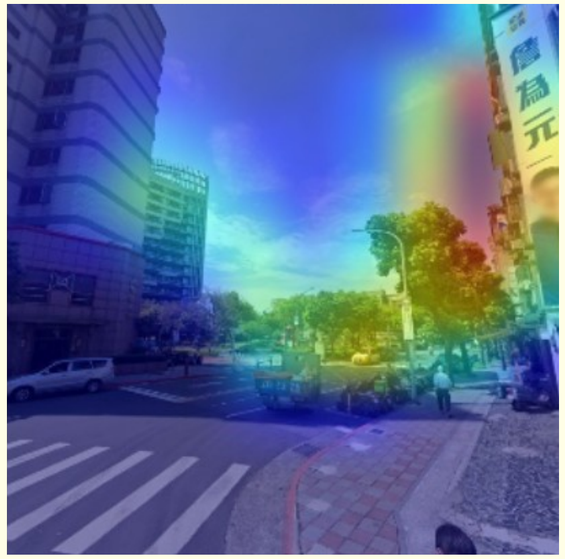
CNN
(predict: bangkok)

RESULT BETWEEN DIFFERENT MODELS

TAIPEI

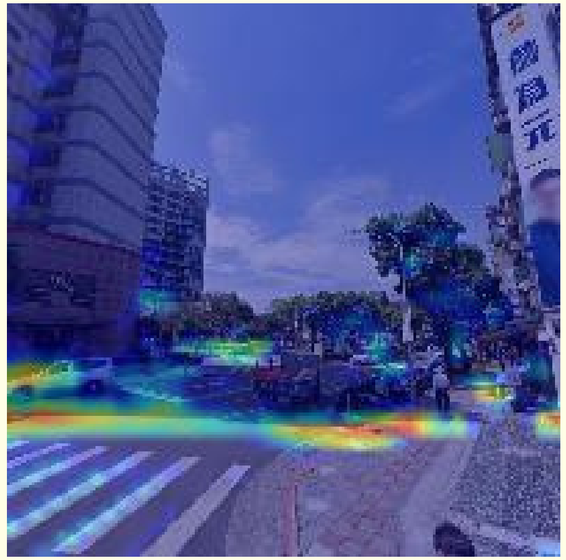
Source image

ResNet
(predict: washington)



Xception
(predict: taipei)

DenseNet
(predict: taipei)



CNN
(predict: taipei)

ANALYSIS

REASONS AFFECT RESULT

MACHINE VIEW

image quality,
pixelate/mosaic,
edge,
light

HUMAN VIEW

some special features
(ex: national flag, license plate)
by domain knowledge

COMMON VIEW

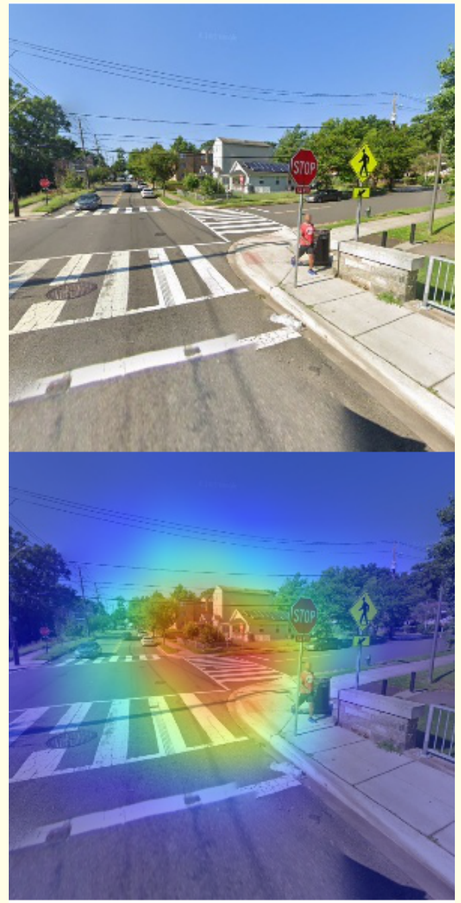
1. traffic sign & crosswalk
2. special buildings
3. unique object (ex: double decker bus, taxi)

HEATMAP RESULT (machine view is similar to human view)

London

Bankok

Taipei



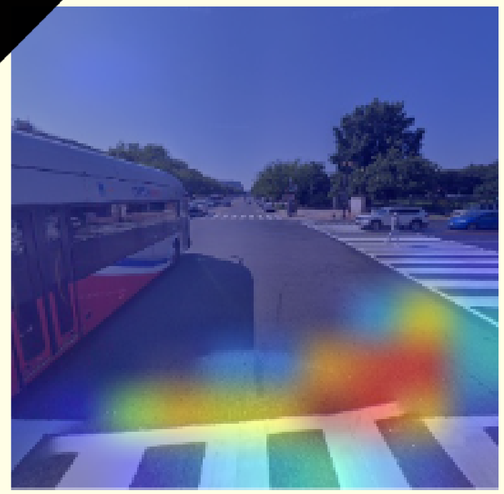
Washington

HEATMAP RESULT (machine view is different from human view)

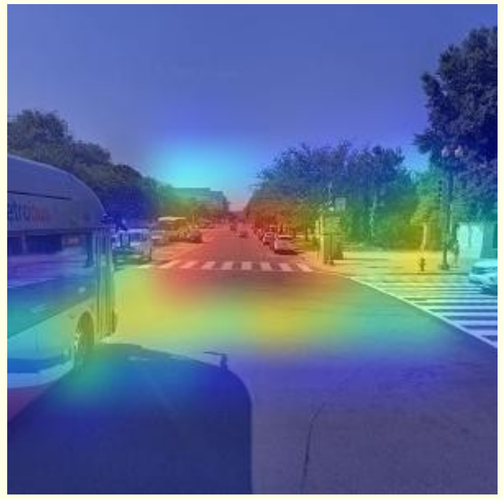
affected by mosaic

affected by sky

EXPERIMENTS (Factor1: mosaic)

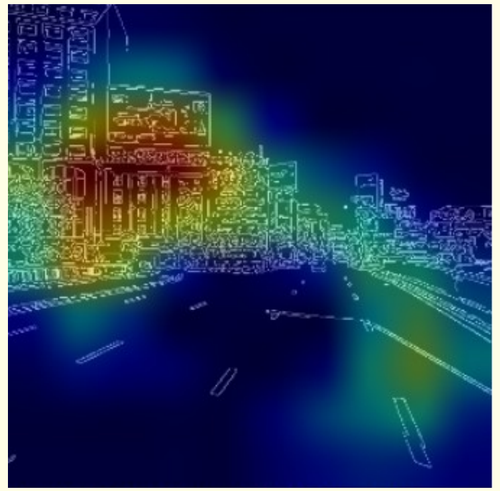


crop images



EXPERIMENTS (Factor2: light, edge)

Canny edge detector



CONCLUSION

We can reach high accuracy with pretrained model.
By visualization, we can see that all models can predict some images through grabbing features as we expected. However, some reasons (such as image quality, light, edge) would lead to models training images through other non-human view features.

Thanks for your listening!