Real-Time ASL Numbers Detection

System Design Project for AI Final Project

조선해양공학과 방승원

Model

YOLOv4-tiny (Bochkovskiy et al, 2020)

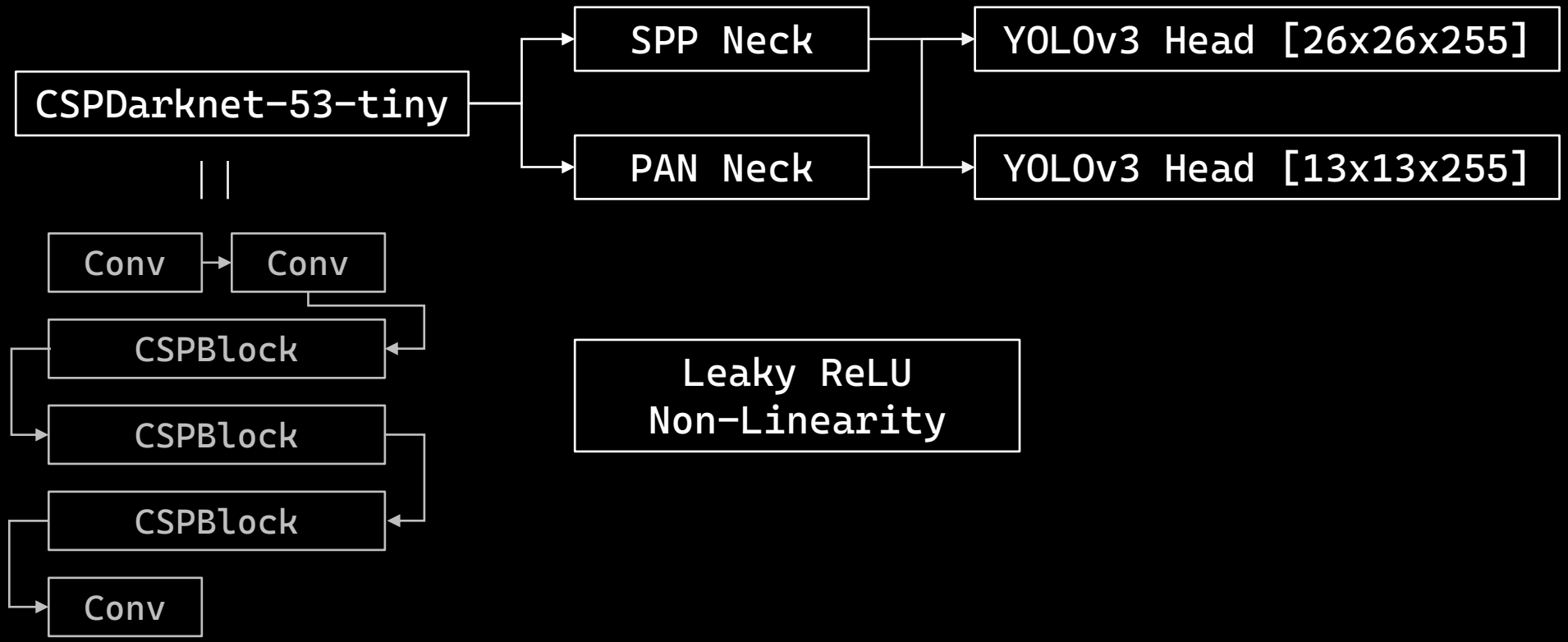
Lightweight version of YOLOv4 Detection model

22.0% mAP (MS COCO) @ 443 fps (RTX 2080Ti)

arXiv:2004.10934 [cs.CV]

Model

YOLOv4-tiny (Bochkovskiy et al, 2020)



arXiv:2004.10934 [cs.CV]

Model

YOLOv4-tiny (Bochkovskiy et al, 2020)

Cross-Stage Partial Connections

DropBLock Regularization

Spatial Pyramid Pooling

Path Aggregation

Self-Adversarial Training

Cosine Annealing LR Scheduler …

arXiv:2004.10934 [cs.CV]

Data

20 videos * 10 classes * 30 s * 30 fps

= 180,000 frames created [1920x1080]

Labels, Bboxes created using OpenCV, Google MediaPipe

Images sampled at 5fps

Dataset selected randomly from images

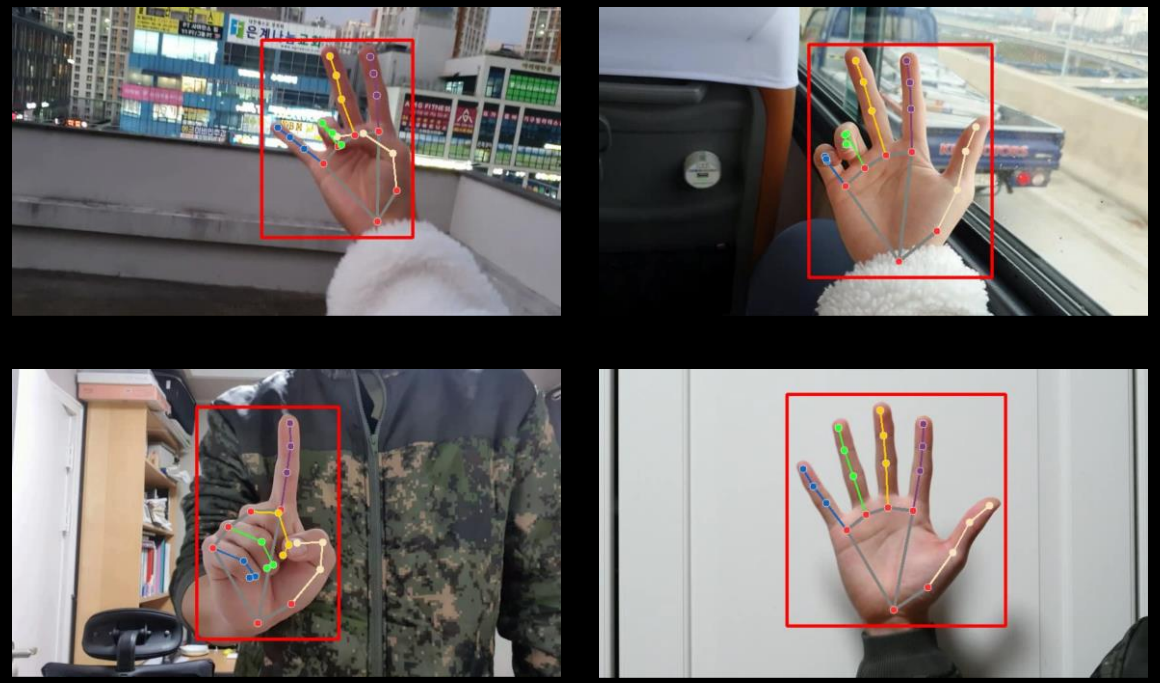
Train_dataset = { 10 * 500 images }

Validation_dataset = { 10 * 100 images }

Videos taken on Samsung Galaxy S10 5G, Nikon Z6

Data

```
BBox {x1, x2, y1, y2} = {
min(landmarks.x) - padding,
max(landmarks.x) + padding,
min(landmarks.y) - padding,
max(landmarks.y) + padding }
```



Data

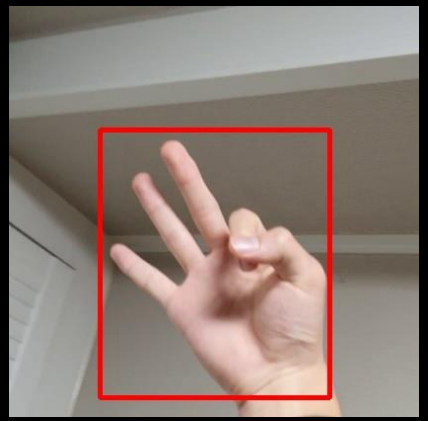
416

416

Images cropped to [416x416]

Images with hand landmarks

outside of the frame removed



class_num x_center_norm y_center_norm width height

8 0.5017052343736093 0.6274073868989944 0.5597013838589191 0.6501731971899669

Train

used code from https://github.com/AlexeyAB/darknet.git

Batch Size = 64

Subdivisions = 16

Learning Rate = 0.00261

Max Epochs = 315

Results

```
Classification AP

Class ‘One’    AP = 100.00%
Class ‘Two’    AP = 99.97%
Class ‘Three’  AP = 98.93%
Class ‘Four’   AP = 99.56%
Class ‘Five’   AP = 97.45%
Class ‘Six’    AP = 99.95%
Class ‘Seven’  AP = 100.0%
Class ‘Eight’  AP = 100.00%
Class ‘Nine’   AP = 99.88%
Class ‘Ten’    AP = 99.99%

mAP (@0.50) = 99.57%
Average IoU = 85.44%
Eval Speed  = 250fps (V100)
```

Results

Final Model Location :

/tools/home/ai_competition10/Project/TermProject_201711405

Thank you!