5-8 Personal Research

Presenter: Yu-Chun Hung

Advisor: Peng-Sheng Cheng

Summary:

- Currently running numerical tolerance experiment
  - tinyyolo still running.
- New perspective: high level operators.
- Thesis Slides.

Numerical tolerance

Scale the input by 2^0 to 2^7 (128), and measure the numerical error.

Posit(16,0) on MobileNetV2:

```
=====Running iteration 4=====
MAE: 0.1779043531455099 5
=====Running iteration 5=====
MAE: 5.964813355576247
=====Running iteration 6=====
MAE: 20.94352163089812
=====Running iteration 7=====
MAE: 57.994662636160854
```

Numerical tolerance

For MNIST, resnet-18, mobilenetV2

- 32-bit posit is stable even the value scaled by 128
- 16-bit posit might suffer if the scale > 32, depends on model and posit settings.

TinyYoloV2 architecture

1. Multiply by 1/255 (Normalize to [0, 1] in model)
2. The following repeat 6 times
    i. Convolution
    ii. Batch Normalization
    iii. Leaky ReLu
    iv. Max Pooling
3. Several combination of Convolution, Batch Normalization and Leaky Relu
4. Output

TinyYoloV2: We have seen all those operators?

Compared to MNIST, resnet-18, MobileNetV2

1. Convolution: All tested model has it.
2. Batch Normalization: All except MNIST.
3. Leaky Relu: All tested has ReLu, it just value < 0 needs to multiply by alpha.
4. Max Pooling: MNIST and resnet-18 has it.

Model comparison.

- Some other comparison
  - MobileNetV2 has depth-wise separable convolution.
  - resnet-18 and MobileNetV2 has skip connection
    - $y = F(x) + x$
  - Different arrangement and dimension of convolution and batch normalization.
- Potential hint: tinyyolov2 is the largest model we've tested.

MNIST	mobilenetV2	resnet-18	tinyyolov2
5998	3,539,138	11,699,112	15,867,889

Future works

- Numerical Error by model size:
  - ResNet-18, ResNet-34, ResNet-50.
- Next weeks report.
- For real, where's my master thesis?