차세대 반도체 디지털 혁신공유대학 사업단

POLARIS

Deep Learning Hardware 설계 경진대회
Orientation, code structure

2022.02.07 (Mon)

Road map

About AIX

Code structure

Organizing committee

- Representative organizing committee

Xuan-Truong Nguyen, Ph.D.
서울대학교

고영훈, 석박사 통합 학생
서울대학교

정지윤 선생님
서울대학교

- Technical staffs: 김혜미 선생님 (hyemi15@snu.ac.kr), 정지윤 선생님
- Advisors: 이혁재교수, 이태호교수, 류현석교수, 선우경교수, 이종원교수, 김남준교수, 이재학교수, 서울대학교

저작권

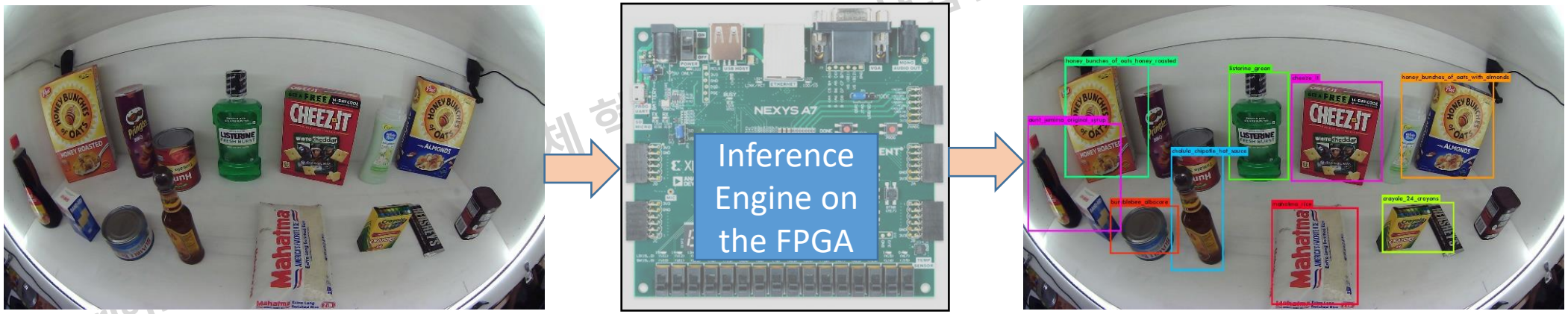
Road map

About AIX

Code structure

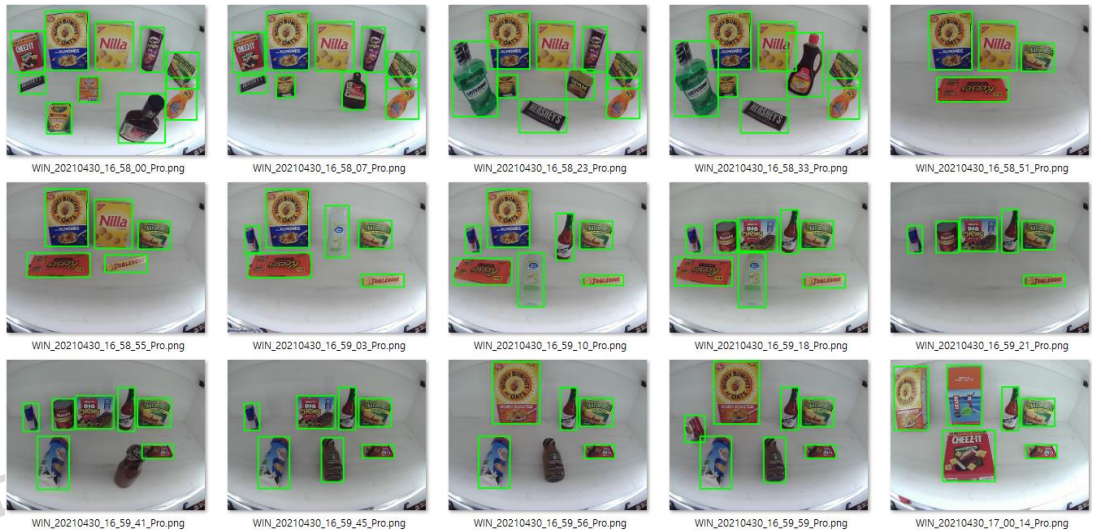
설계의 목표

무인판매대에서 상품 인식을 위하는 딥러닝 추론 하드웨어를 설계한다.

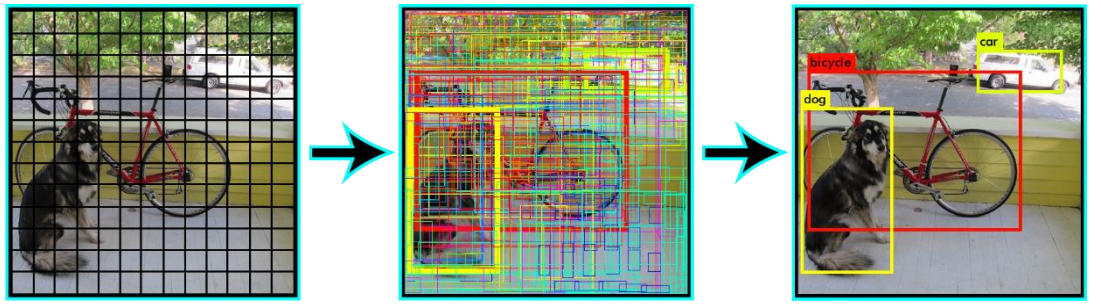


Object detection (Application-based)

- Training dataset
  - Collected 200,000 images of items
  - Labelled the data
- We developed a deep network based on Tiny-YOLOv3
  - Input: an RGB image (320×320×3)
  - 22 layers:
    - 11 convolutional layers
    - Max pooling



datasets

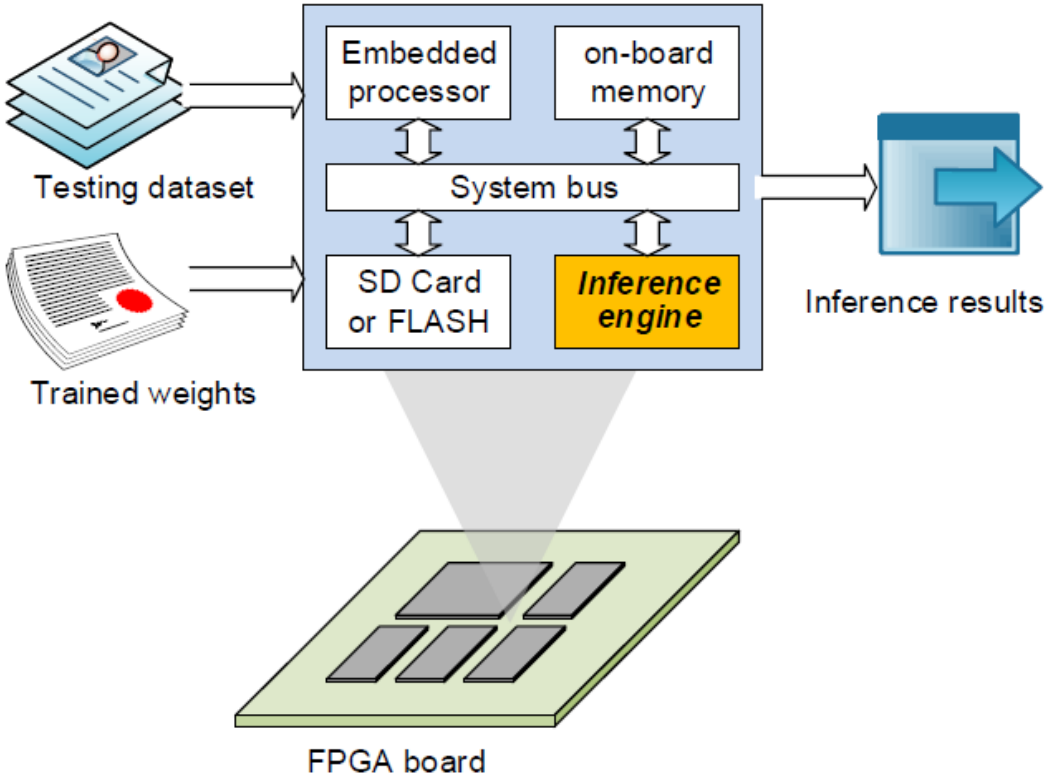


YOLO

https://pjreddie.com/darknet/yolov2/

제공되는 것과 준비할 것 (system-based)

- 추론엔진용 딥러닝 파라미터
  - Pretrained model
- 시험 데이터셋 (상품 이미지 150장)
- 추론엔진(Inference Engine)의 reference code
  - S/W: Evaluation and Host PC (C++)
  - H/W: Components (Verilog HDL)
- Nexys A7 FPGA Board (Xilinx Artix-7 FPGA XC7A100T-1CSG324C)



Tutorials[1]

- Four-week tutorials are given to cover several fundamental issues in the AIX Design
  - Introduce fundamental components and their usages

#	Title	Content	Date
1	Orientation	Introduction to AIX, code structure	2/07 (Mon)
2	Network	Reference S/W, network architecture, evaluation metrics	2/10 (Thu)
3	Quantization	Model quantization, data preparation	2/14 (Mon)
4	MAC	Hardware description language, computing units	2/17 (Thu)
5	Memory	On-chip buffer, block RAM, IP generator, MIG	2/21 (Mon)
6	Bus	AXI interconnect, DMA	2/24 (Thu)
7	Integration	System integration and verification (one layer)	2/28 (Mon)
8	CPU-FPGA	PC-FPGA communication	3/03 (Thu)

[1] **Tutorials do NOT aim to replace courses at school. Students are highly recommended to take relevant courses.**

경진대회 개최 일정

- ~~공지 및 접수: 1.10 (월)~2.6 (일)~~
  - ~~사전 온라인 설명회 : 1.26(수) 오후 3시, 온라인(Zoom)으로 진행~~
- 오리엔테이션: 2.7 (월)
- 본선: 2.7 (월) ~ 5.20 (금, 최종 설계 제출 마감): Code submission (5.20)
- 중간 평가: 4.1 (금)
- 참가자 발표 및 최종 심사: 5.23 (월)~5.27 (금) Presentation
- 최우수팀 선정 및 시상식: 2022년 6월 중

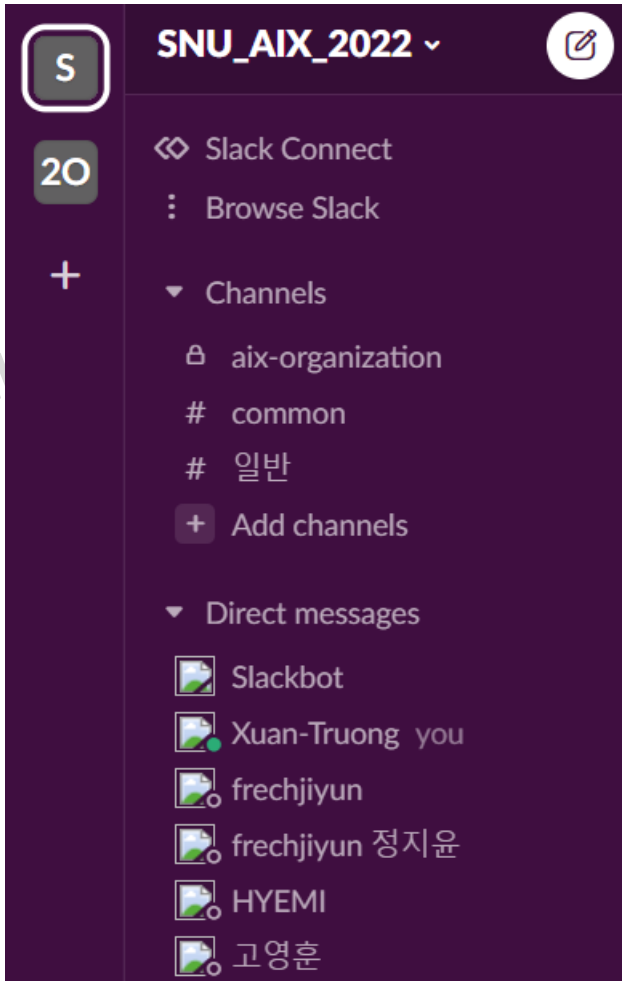
Team list

- 대상 및 자격: 전국의 대학생 2~5인 구성의 팀으로 딥러닝과 하드웨어에 관심있는 누구나 참여 가능
- 접수 방법: https://forms.gle/dq44a1cjYeZcj5jH8
  - 387 responses, some of them are duplicated
    - **~30 universities**
    - A new registration is no longer accepted
  - 374 students from 111 teams are approved.
  - Four students or one-member teams are pending
    - Matching

AIX2022 Communication

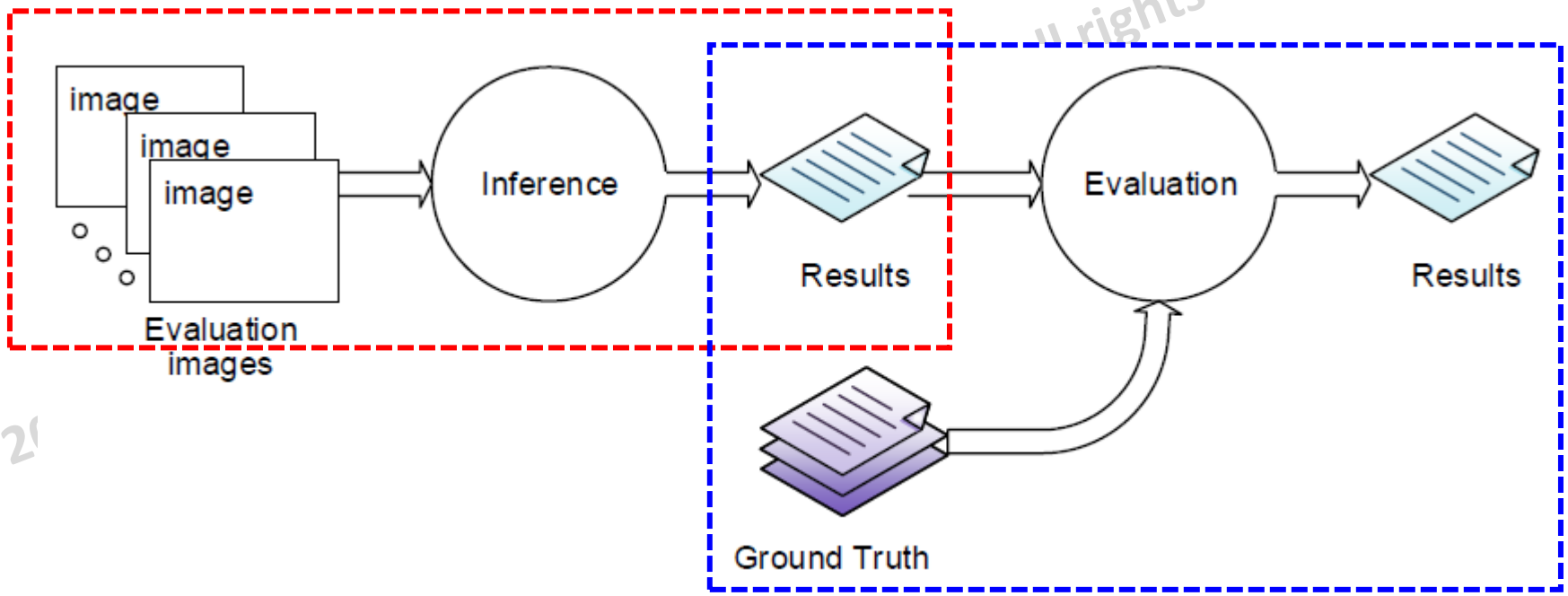
- Open discussion platform (slack)
  - Invite all participants
  - FAQ
    - List up all common questions
  - QA
  - Share materials
    - Codes
    - Documents (PPT files)
    - Link to video tutorials (POLARIS's youtube channel)
  - Note: we can not use servers from SNU or POLARIS for students from 30 universities.

- Two options for board delivery (in Feb)
  1. At Seoul National University
  2. Mail



결과평가: Accuracy

- 평가 데이터 셋 (참가자에게 공개하지 않음)
  - ▶ Images of product items (150개+ 150개)
  - ▶ Mean-average precision (mAP): Quantized model



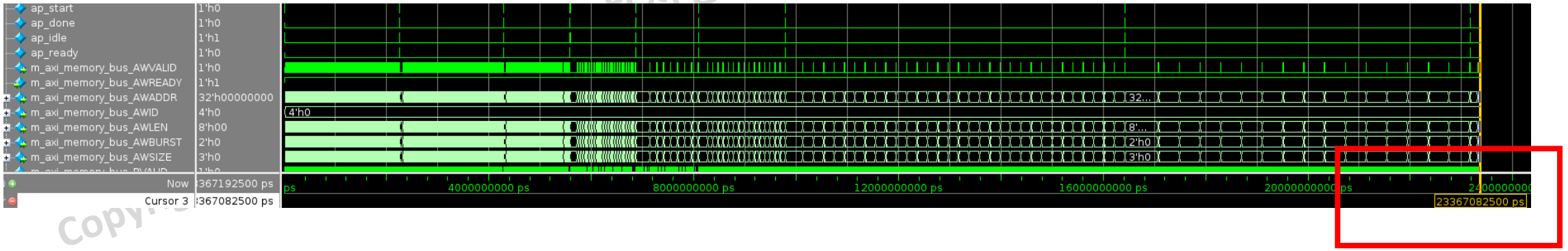
결과평가: Accuracy

- Intersection of Union (IoU)
  - Intersection: Overlapped area between two bounding boxes
  - Union: Combined area of two bounding boxes
- Mean average precision (mAP)
  - Calculate the AP at IoU threshold 0.5.
  - INT8 quantized accuracy.
- The code for evaluation is given


```
detections_count = 17623, unique_truth_count = 4404
class_id = 0, name = aunt_jemima_original_syrup, ap = 92.71%      (TP = 49, FP = 12)
class_id = 1, name = band_aid_clear_strips, ap = 68.13%           (TP = 34, FP = 14)
class_id = 2, name = bumblebee_albacore, ap = 98.39%              (TP = 59, FP = 3)
class_id = 3, name = cholula_chipotle_hot_sauce, ap = 76.70%      (TP = 42, FP = 13)
class_id = 4, name = crayola_24_crayons, ap = 93.25%              (TP = 135, FP = 14)
class_id = 5, name = hersheys_cocoa , ap = 96.57%                 (TP = 47, FP = 4)
class_id = 6, name = honey_bunches_of_oats_honey_roasted, ap = 95.40%          (TP = 60, FP = 24)
class_id = 7, name = honey_bunches_of_oats_with_almonds, ap = 89.31%     (TP = 50, FP = 10)
class_id = 8, name = hunts_sauce, ap = 97.91%                     (TP = 60, FP = 8)
class_id = 9, name = listerine_green, ap = 95.66%                 (TP = 51, FP = 6)
class_id = 10, name = mahatma_rice, ap = 98.18%                   (TP = 56, FP = 0)
class_id = 11, name = white_rain_body_wash, ap = 80.03%                (TP = 31, FP = 7)
class_id = 12, name = pringles_bbq, ap = 81.97%                   (TP = 37, FP = 4)
class_id = 13, name = cheeze_it, ap = 99.34%              (TP = 69, FP = 12)
class_id = 14, name = hersheys_bar, ap = 75.65%                   (TP = 26, FP = 1)
class_id = 15, name = redbull, ap = 63.54%                (TP = 39, FP = 14)
class_id = 16, name = mom_to_mom_sweet_potato_corn_apple, ap = 76.91%          (TP = 24, FP = 0)
class_id = 17, name = a1_steak_sauce, ap = 74.63%                 (TP = 33, FP = 5)
class_id = 18, name = jif_creamy_peanut_butter, ap = 81.39%       (TP = 54, FP = 10)
class_id = 19, name = cinnamon_toast_crunch, ap = 99.84%          (TP = 74, FP = 2)
class_id = 20, name = arm_hammer_baking_soda, ap = 88.64%         (TP = 41, FP = 2)
class_id = 21, name = dr_pepper, ap = 92.77%              (TP = 51, FP = 2)
class_id = 22, name = haribo_gold_bears_gummi_candy, ap = 79.43%        (TP = 52, FP = 14)
class_id = 23, name = bulls_eye_bbq_sauce_original, ap = 98.20%         (TP = 66, FP = 1)
class_id = 24, name = reeses_pieces, ap = 96.83%                  (TP = 73, FP = 5)
class_id = 25, name = clif_crunch_peanut_butter, ap = 88.49%      (TP = 53, FP = 13)
class_id = 26, name = mom_to_mom_butternut_squash_pear, ap = 91.75%     (TP = 61, FP = 0)
class_id = 27, name = pop_tararts_strawberry, ap = 99.15%         (TP = 73, FP = 5)
class_id = 28, name = quaker_big_chewy_chocolate_chip, ap = 60.66%      (TP = 35, FP = 60)
class_id = 29, name = spam, ap = 67.44%                   (TP = 40, FP = 5)
class_id = 30, name = coffee_mate_french_vanilla, ap = 80.68%           (TP = 47, FP = 9)
class_id = 31, name = pepperidge_farm_milk_chocolate_macadamia_cookies, ap = 85.40%     (TP = 45, FP = 0)
class_id = 32, name = kitkat_king_size, ap = 81.82%               (TP = 57, FP = 12)
class_id = 33, name = snickers, ap = 32.33%               (TP = 24, FP = 19)
class_id = 34, name = toblerone_milk_chocolate, ap = 84.68%       (TP = 59, FP = 13)
class_id = 35, name = clif_z_bar_chocolate_chip, ap = 97.71%      (TP = 68, FP = 7)
class_id = 36, name = nature_valley_crunchy_oats_n_honey, ap = 60.27%          (TP = 56, FP = 82)
class_id = 37, name = ritz_crackers, ap = 95.39%                  (TP = 71, FP = 22)
class_id = 38, name = palmolive_orange, ap = 87.85%               (TP = 60, FP = 2)
class_id = 39, name = crystal_hot_sauce, ap = 68.13%              (TP = 36, FP = 15)

 for conf_thresh = 0.25, precision = 0.82, recall = 0.72, F1-score = 0.77
 for conf_thresh = 0.25, TP = 3165, FP = 688, FN = 1239, average IoU = 63.17 %

 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 0.835103, or 83.51 %
Total Detection Time: 68 Seconds

class_id = 53, name = dove_pink, ap = 73.37%      (TP = 38, FP = 5)
class_id = 54, name = dove_white, ap = 79.90%                     (TP = 48, FP = 6)
class_id = 55, name = david_sunflower_seeds, ap = 95.70%           (TP = 57, FP = 7)
class_id = 56, name = monster_energy, ap = 56.73%                 (TP = 32, FP = 8)
class_id = 57, name = act_ii_butter_lovers_popcorn, ap = 95.38%          (TP = 64, FP = 3)
class_id = 58, name = coca_cola_glass_bottle, ap = 78.84%         (TP = 59, FP = 9)
class_id = 59, name = twix, ap = 86.54%                           (TP = 66, FP = 10)

for conf_thresh = 0.25, precision = 0.82, recall = 0.72, F1-score = 0.77
for conf_thresh = 0.25, TP = 3165, FP = 688, FN = 1239, average IoU = 63.17 %

IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
mean average precision (mAP@0.50) = 0.835103, or 83.51 %
```

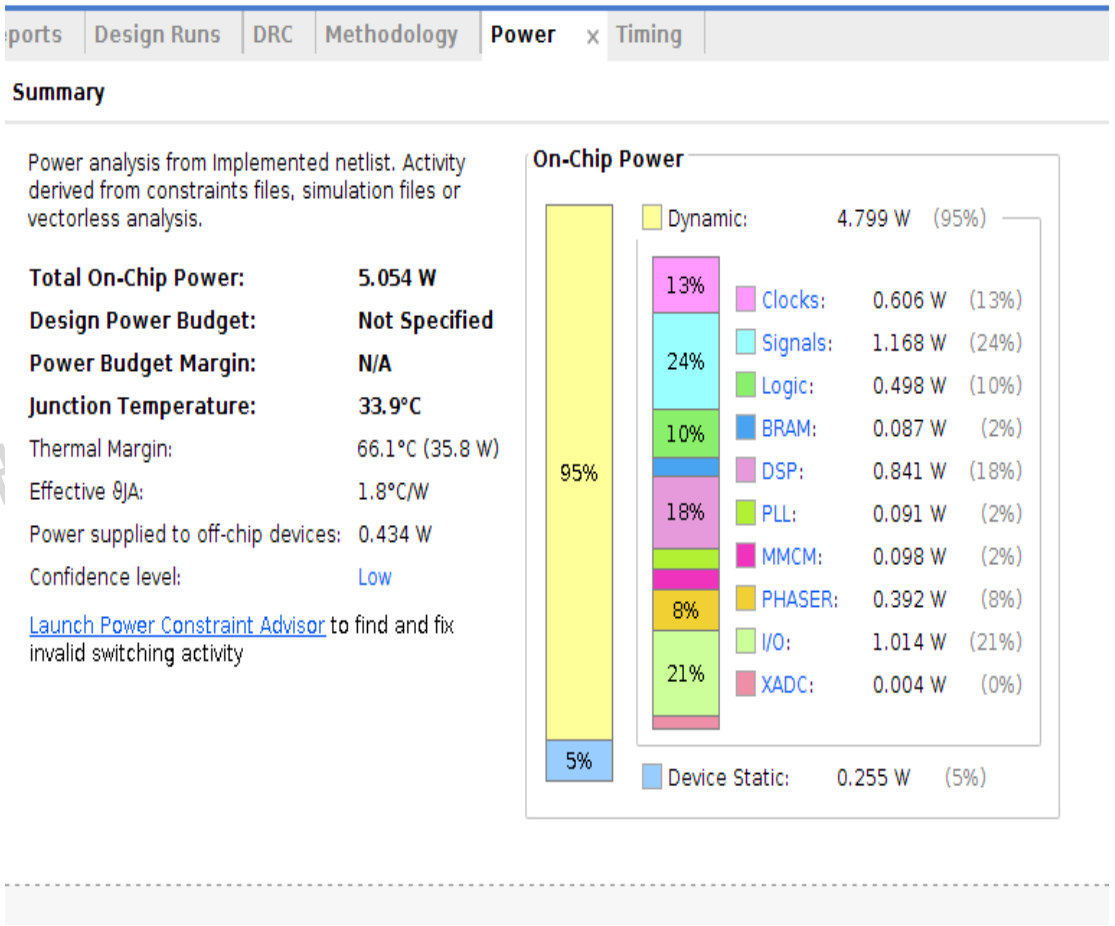
결과평가: Inference speed

- Inference speed measures the number of frames your CNN accelerator IP processes in a second
- RTL simulation result
  - Example:
    - 200MHz, 2 cycle for DRAM latency
    - 23.3ms



결과평가: Energy

- Implement your design on an FPGA board
- Measure the on-chip power of the design after FPGA implementation
  - Logic
  - Memory
  - DSP
  - IO



Evaluation

- 중간 평가: 4.1 (금)
  - Quantized model
  - RTL simulation for the **three-first** CONV layers.

  ⇒ Submit your code and report.

- 최종 평가: 5.20 (금) Code submission
  - Accuracy
  - Speed
  - Power
  - Presentation

대회 최우수팀 수상 및 상품

- 평가 점수 기준 최우수팀 및 우수팀 선정 (3등까지)
  - 1등 : **5백 만원**
  - **2등 : 2백 만원**
  - **3등 : 1백 만원**
- * 최종 1,2,3등 수상팀 인공지능반도체 국제학술대회 (IEEE AICAS) 참석

  * 중간 평가 통과 팀 전원 장려상 수여 및 부상 증정

https://aicas2022.org/

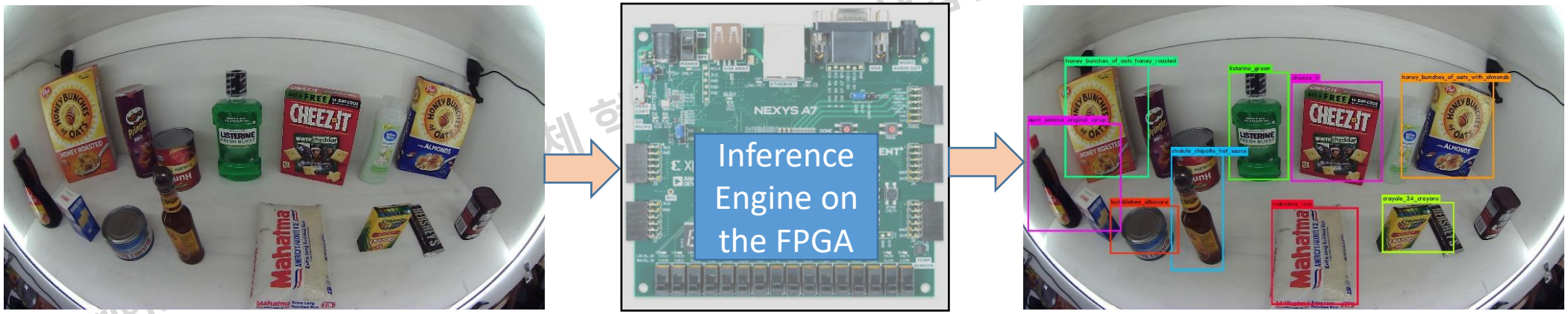
Road map

About AIX

Code structure

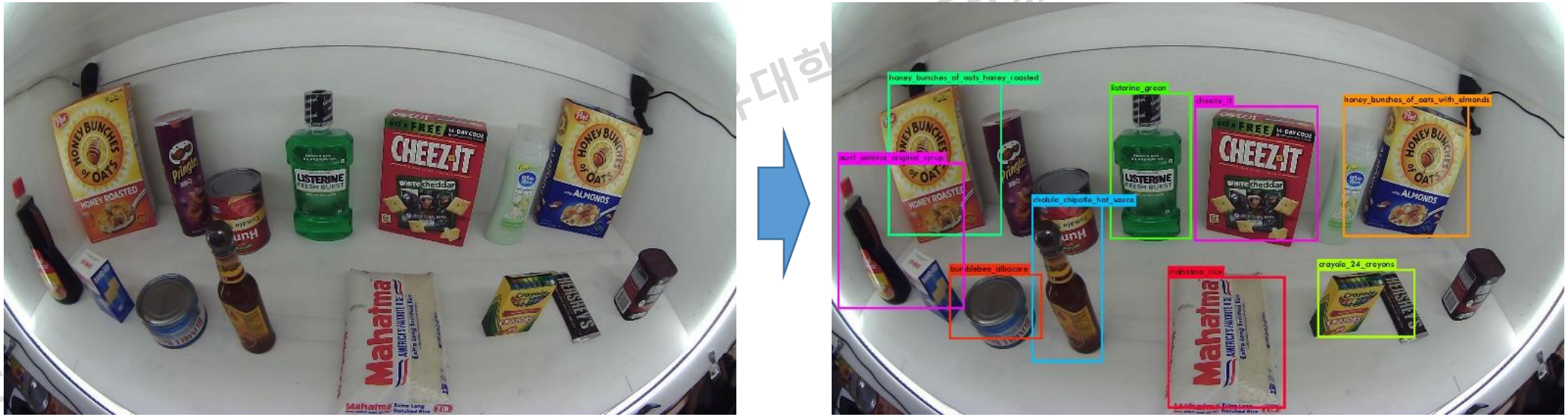
설계의 목표

무인판매대에서 상품 인식을 위하는 딥러닝 추론 하드웨어를 설계한다.

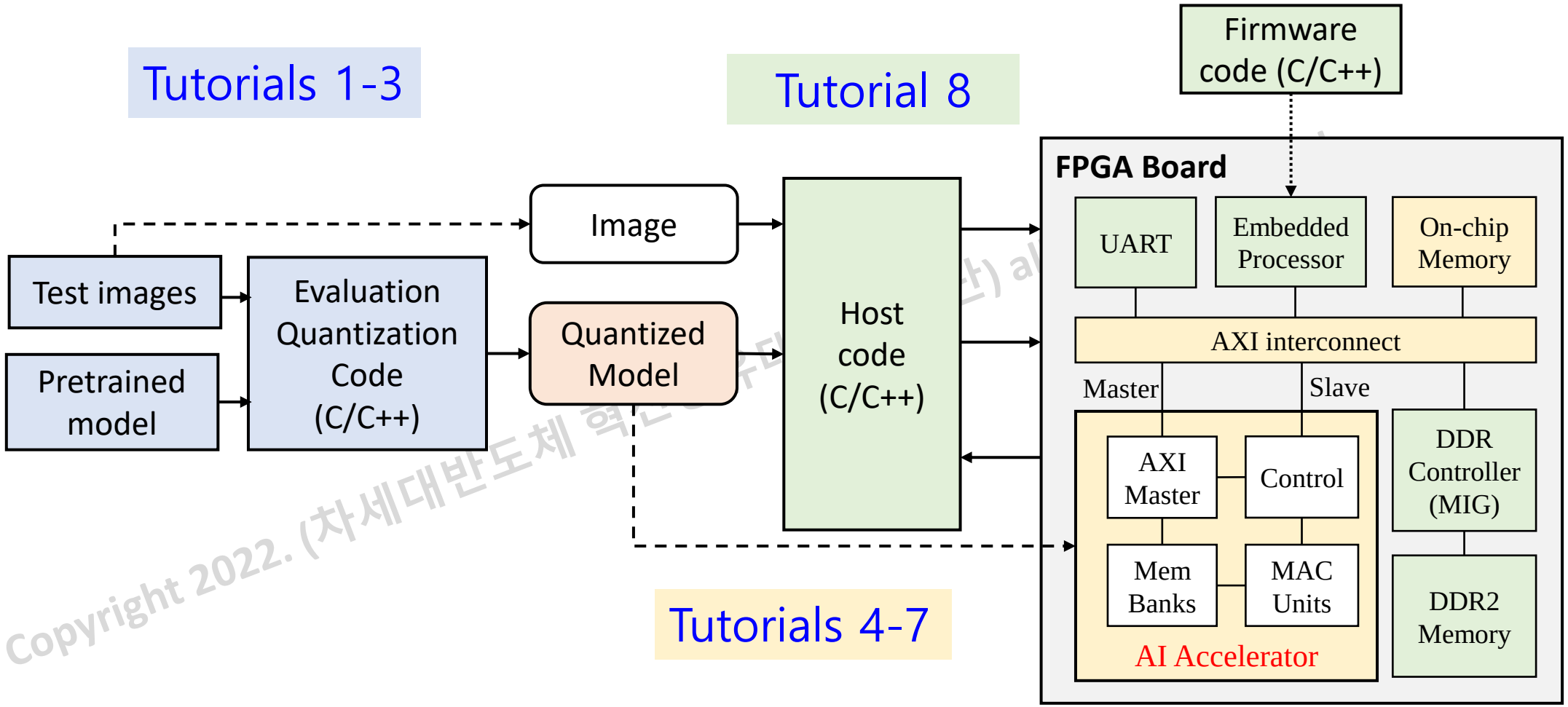


Object detection

- Object detection = Proposal + Classification
  - Proposal: Location of an object
    - Bounding boxes: a rectangle bounding an object
  - Classification: an object's class
    - Indicate by its box's color and name tag

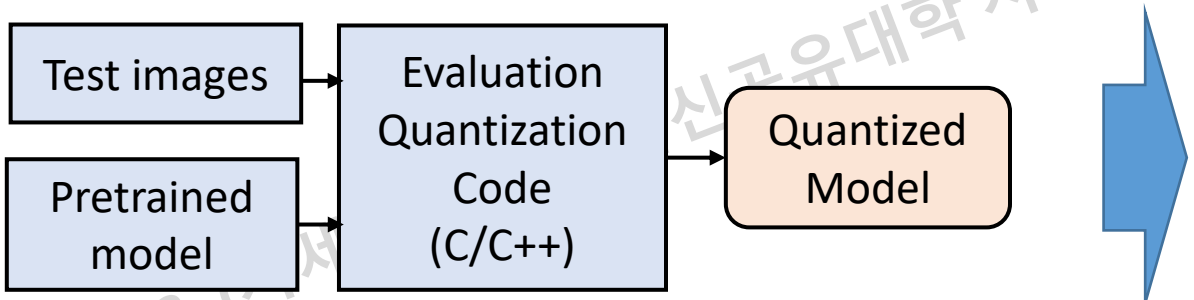


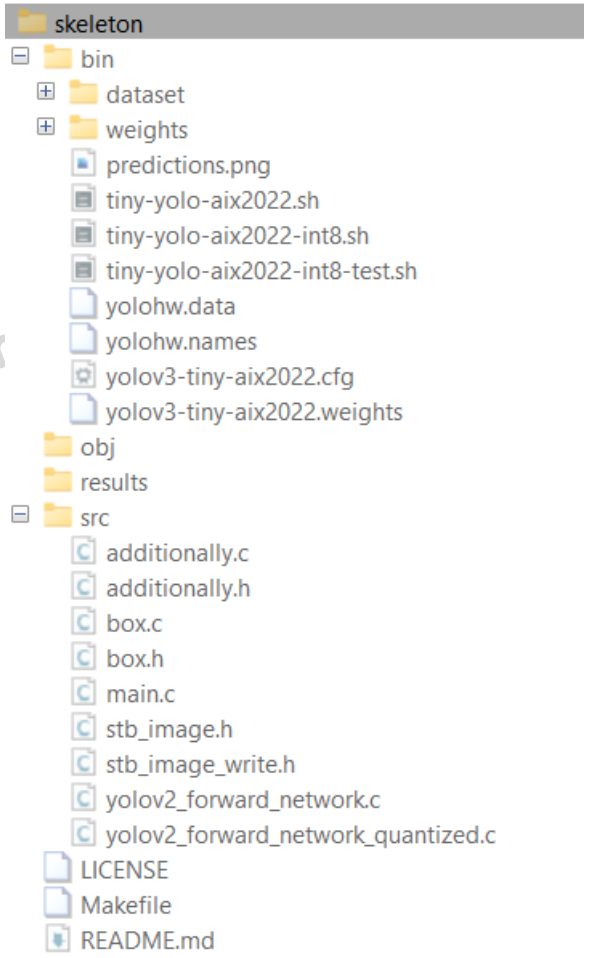
Top structure and tutorials



Tutorials 1-3: Reference S/W

- Skeleton
  - Test images, object classes, and ground truth
  - Pretrained model
  - Reference code for evaluation and quantization
  - The S/W code is based on the darknet code





Test images (skeleton/bin/dataset)

- 150 test images in three categories
  - Long: distances among objects are far
  - Close: items are close
  - Color: apply different colors for background
- Execute a UNIX command: *python make_list_cur.py*
  - Get directories of all images and store them to target.txt

long

close

color

Test images (skeleton/bin/dataset)

- Ground truth: *.txt files
  - Each line represents a labelled object
    - [class_index x_pos y_pos width height]
- yolohw.names: names of all 60 classes of products

```
12 0.19973958333333333 0.39166666666666666 0.0671875 0.2777777777777778
13 0.26484375 0.39351851851851855 0.1265625 0.2796296296296296
11 0.3419270833333333 0.38472222222222224 0.06302083333333333 0.287962962962963
6 0.4427083333333333 0.33055555555555555 0.178125 0.4425925925925926
0 0.5614583333333333 0.3842592592592593 0.090625 0.3388888888888889
9 0.6377604166666666 0.3773148148148148 0.09010416666666667 0.325
7 0.7296875 0.3699074074074074 0.14895833333333333 0.34351851851851856
1 0.19817708333333334 0.649537037037037 0.08072916666666667 0.18055555555555555
2 0.2671875 0.7041666666666667 0.08645833333333333 0.1824074074074074
3 0.30859375 0.6421296296296296 0.0859375 0.3212962962962963
4 0.3890625 0.7495370370370371 0.090625 0.1787037037037037
14 0.45598958333333334 0.7847222222222222 0.0734375 0.23425925925925928
10 0.621875 0.7532407407407408 0.25625 0.23425925925925928
5 0.765625 0.6810185185185186 0.11041666666666666 0.21574074074074076
8 0.825 0.6587962962962963 0.09270833333333334 0.23240740740740742
```

CAPP_testset_close_10000.txt

```
aunt_jemima_original_syrup
band_aid_clear_strips
bumblebee_albacore
cholula_chipotle_hot_sauce
crayola_24_crayons
hersheys_cocoa
honey_bunches_of_oats_honey_roasted
honey_bunches_of_oats_with_almonds
hunts_sauce
listerine_green
mahatma_rice
white_rain_body_wash
pringles_bbq
cheeze_it
hersheys_bar
redbull
```

yolohw.names

Pretrained model (skeleton/bin)

- A pretrained model is defined by two files
  - yolov3-tiny-aix2022.cfg: Network's configuration
    - Input size
    - Training/testing options
    - Layer's settings
  - yolov3-tiny-aix2022.weights (3354 KB)
    - 32-bit floating point parameters

```
[net]
# Testing
batch=1
subdivisions=1
# Training
#batch=64
#subdivisions=2
width=320
height=320
channels=3
momentum=0.9
decay=0.0005
angle=0
saturation = 1.5
exposure = 1.5
hue=.1

learning_rate=0.001
burn_in=1000
max_batches = 50200
policy=steps
steps=40000,45000
scales=.1,.1

[convolutional]
batch_normalize=1
filters=16
size=3
stride=1
pad=1
activation=leaky

[maxpool]
```

Source files

- additionally.c *// Definitions of darknet functions used*
- additionally.h *// Declaration of darknet functions + additional functions for forward pass of yolo model*
- box.c
- box.h

  } *// For bounding boxes*
- stb_image_write.h
- stb_image.h

  } *// For loading/writing images*
- yolov2_forward_network.c *// Functions for forward pass of yolo network*
- **yolov2_forward_network_quantized.c** *// Functions for quantization, saving of the quantized model, and the forward pass of quantized yolo model*
- main.c *// The main functions*

You should mainly edit this file for quantization!

How to Compile & Run?

- Use the provided **Makefile** to compile

- Run **bin/darknet** with appropriate arguments
  - *Example*
    - *tiny-yolo-aix2022.sh*
    - *tiny-yolo-aix2022-int8.sh*
    - *tiny-yolo-aix2022-int8-test.sh*

How to Compile & Run?

e.g.) tiny-yolo-aix2022-int8.sh

```
./darknet detector map yolohw.names yolov3-tiny-aix2022.cfg yolov3-tiny-aix2022.weights -thresh 0.24 -quantized -save_params
```

To calculate mAP using the provided dataset, use 'map'.
To visualize a detection result for an image, use 'test'.

Different threshold values result in different precision, recall, and F1 score.

To evaluate the quantized model, use '-quantized' option.

To save the quantized model (quantized weights, biases, and input scale factor) to bin/weights, use '-save_params' option.

Incoming tutorials …

- Tutorial 2: Network
  - Network architecture
    - Convolution
    - Batch normalization
    - Max pooling
  - Evaluation metrics
    - What is mAP?