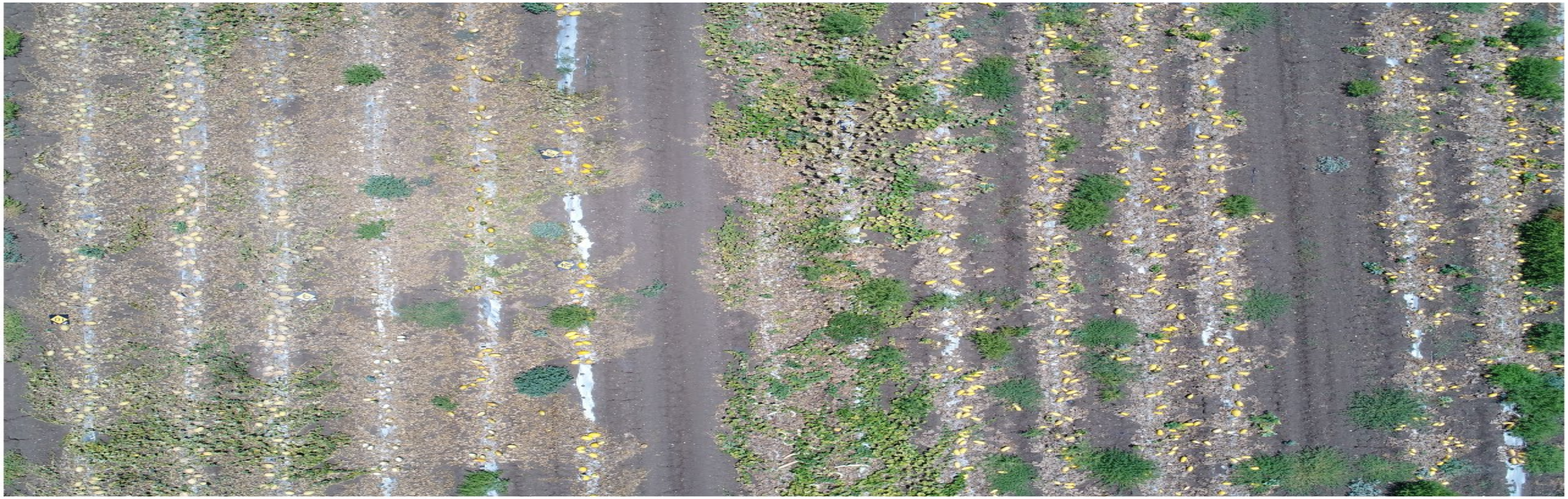
CNN algorithm for single and overall weight estimation of melons using UAV images

Aharon Kalantar [1,2], Yael Edan [2] Amit Gur 3 and Iftach Klapp [1]

1. Agriculture Research Organization –Volcani Center, Agro-Optics and Sensing Lab
2. Ben-Gurion University of the Negev, Department of Industrial Engineering & Management
3. Neve-Yar, ARO-Volcani Center

Lab's Web Site: https://iftachklapp.wixsite.com/agopt

Motivation and Objective



Motivation: Save the labor required for locating and weighting each individual melon in a phenotyping field.

Objective: Develop a robust algorithm that detects melons in an agricultural environment using UAV images for yield estimation.

Challenge : ROI's composed of a minority of small objects of 30X30 pixels or less .

Ref: A.Kalantar, Y.Edan, A.Gur, I.Klapp**, "A deep learning system for yield estimation of melons using UAV images," Computers and Electronics in Agriculture., accepted for publication Aug. 2020

Data set and Images acquisition

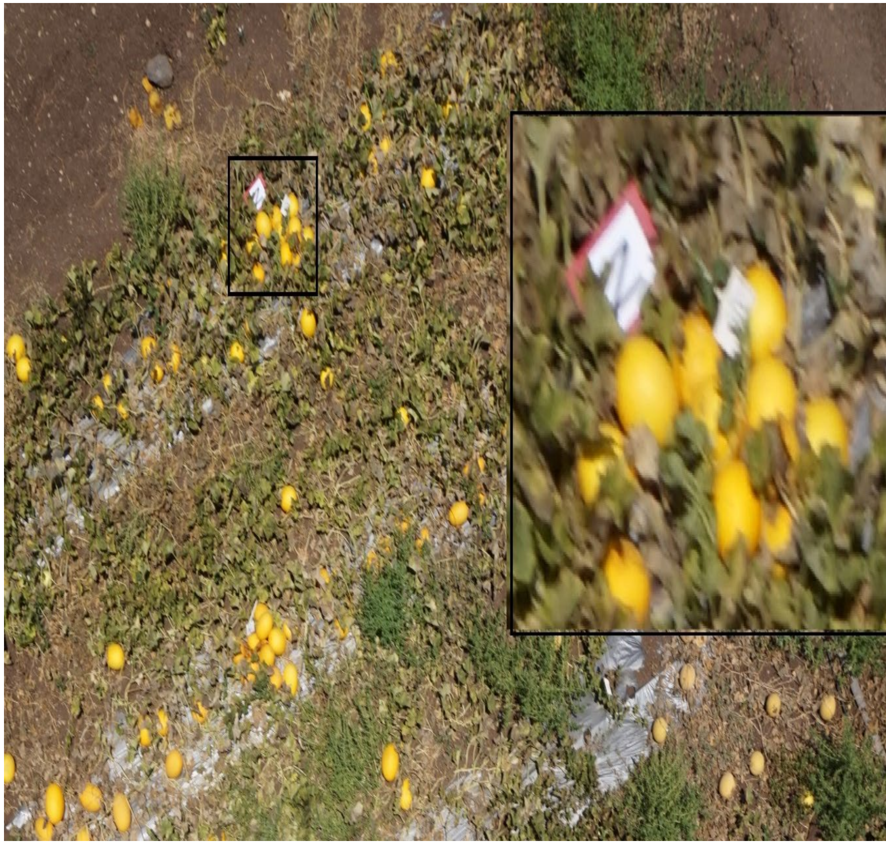
- The data was acquired at Newe Ya‘ar in midday time
- Images acquired by a UAV hovering 15 meters above the field with RGB camera
- The acquisition was done in three different years at summer season, before picking time (2016-2018)

- Drone Type: DJI Phantom 4 Pro
- Camera Type : DJI FC6310 (RGB)
- Image size : 5472 × 3648 pixels

4220 melons were manually tagged, from 4 different images, for ground truth (2018 images).

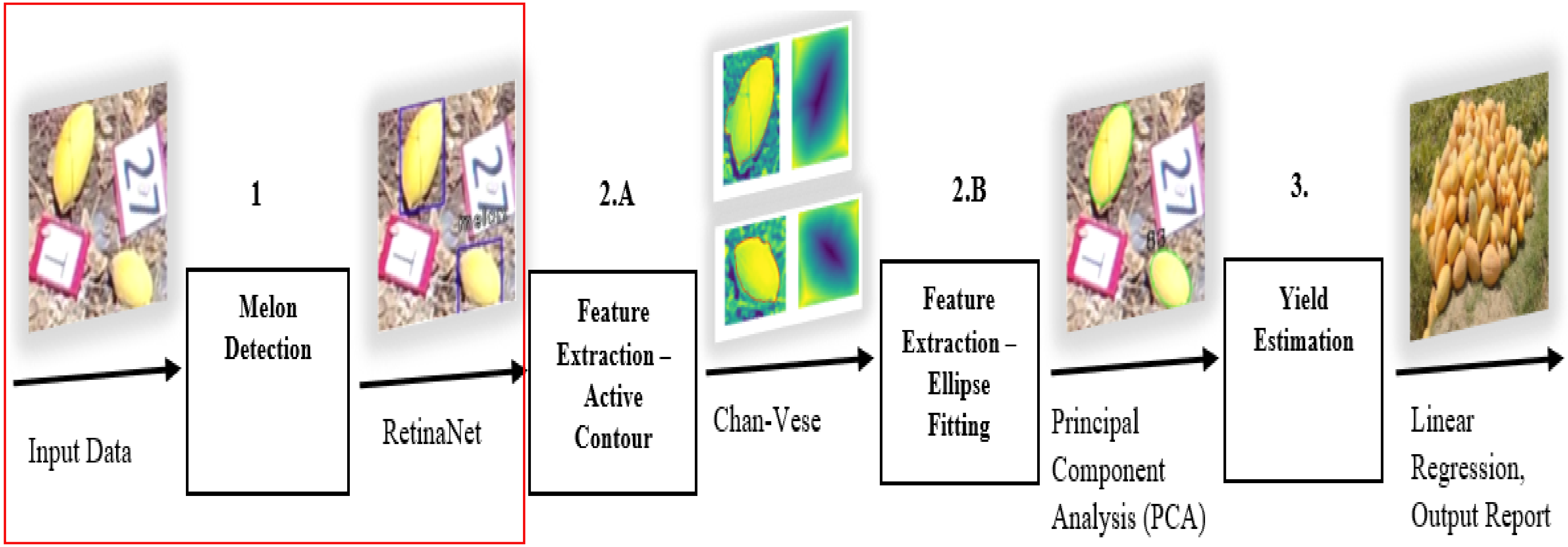
- Augmentation: Rotation, Flip, Translation, Shear, Scaling (zoom)
- Irrigation was stopped one week prior to the measurement.

Data set - Weight estimation



- 138 melons were randomly selected and marked in the field by placing a sign next to them.
- For each melon was provided its size and weight characteristics (validation set).
- in addition, extra data which contained 32 measured melons were provided in order to build the yield estimation regression model.

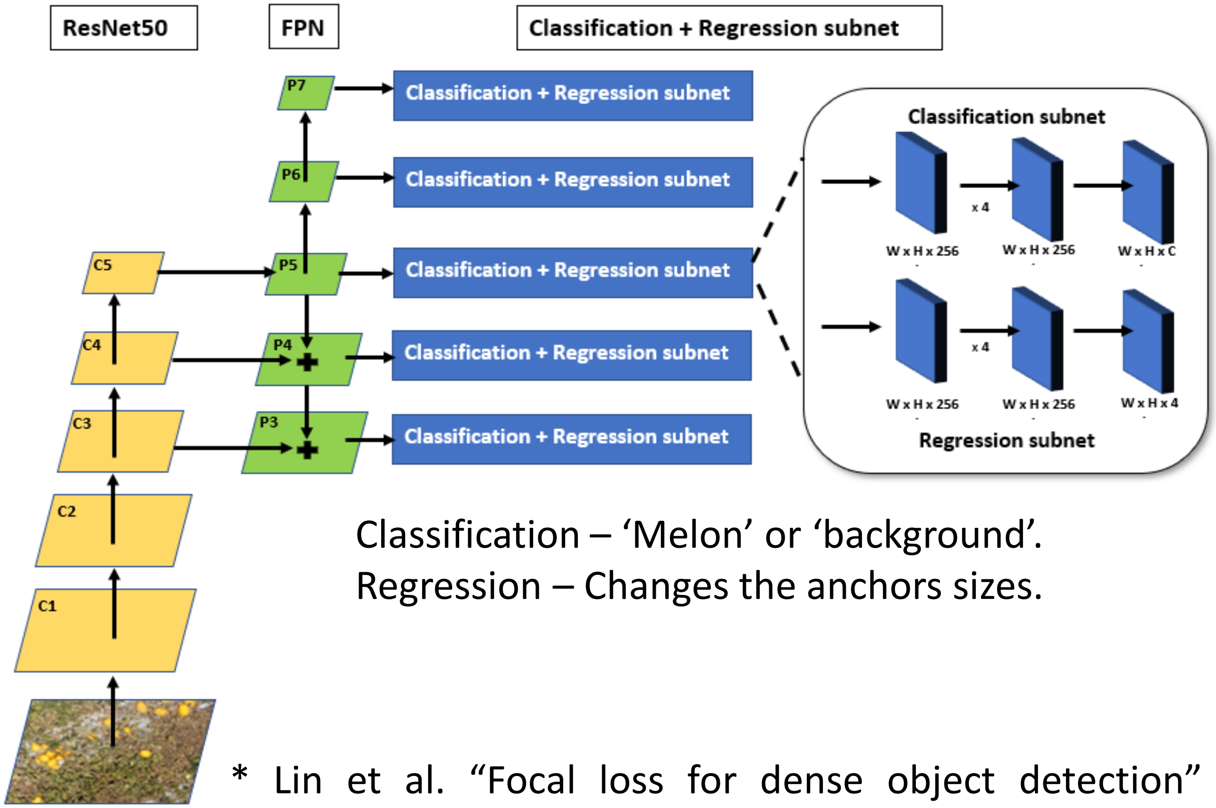
Algorithmic pipeline



Task 1 – Object Detection

RetinaNet network

- Based on [ResNet50]() network as backbone
- Creates **FPN** (Feature Pyramid Network) for **high resolution and strong semantics** – efficient way to create proposal candidates
- Use **classification and regression subnets** for generating final bounding boxes
- Focal Loss – solve unbalanced classes problem



Classification – 'Melon' or 'background'.
Regression – Changes the anchors sizes.

* Lin et al. "Focal loss for dense object detection" (2017)

Object Detection Process

Divide the image to 10*10 pieces **with overlap**

↓

Detected all melons in each piece

↓

Compose all detection back to the original image

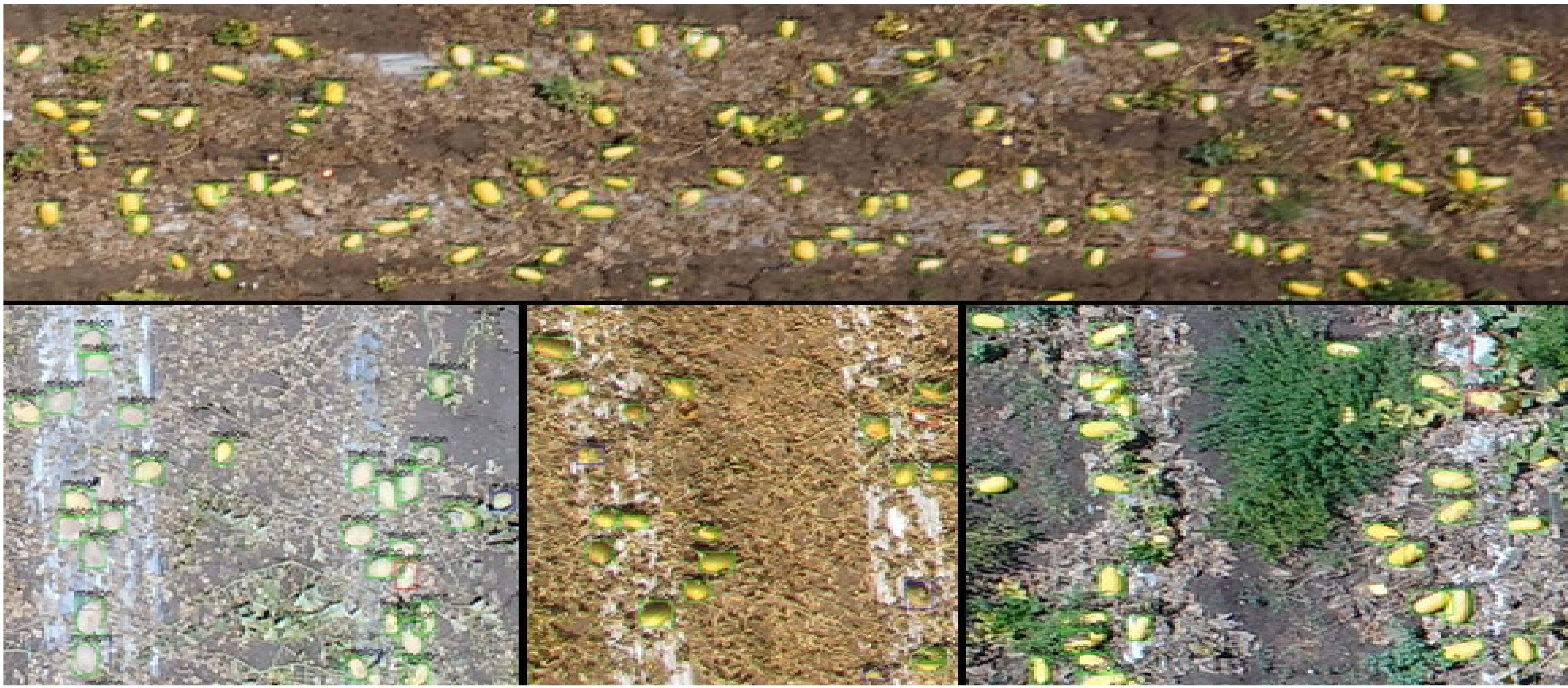
↓

Remove duplication with NMS (non-maximum-suppression) algorithm

*This is a small part from a big image (zoomed) for illustration purpose

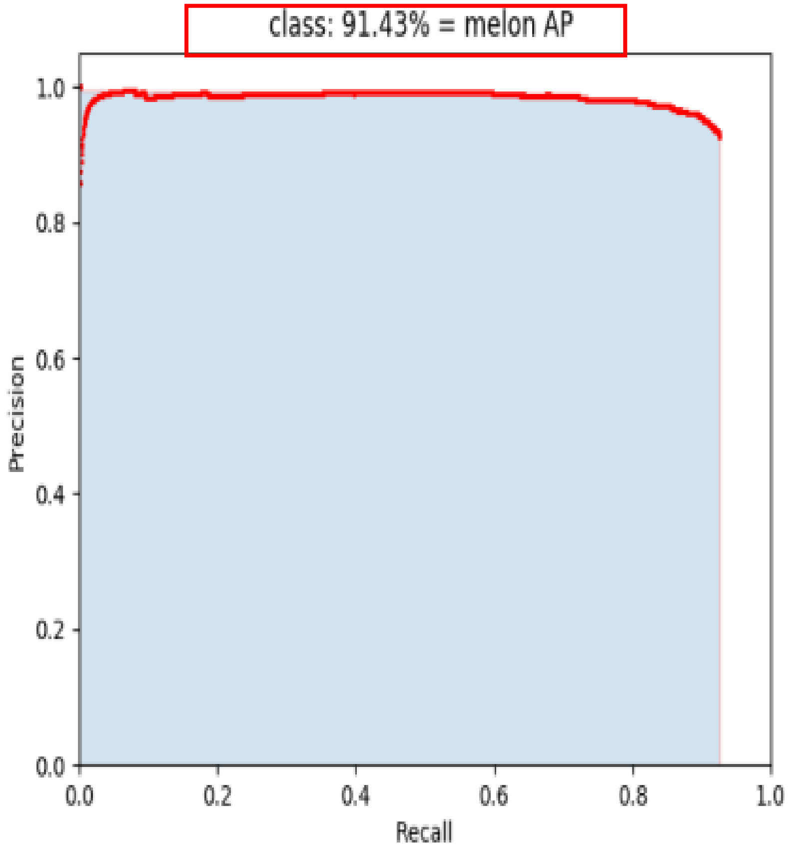
Task 1 – Detection results

2



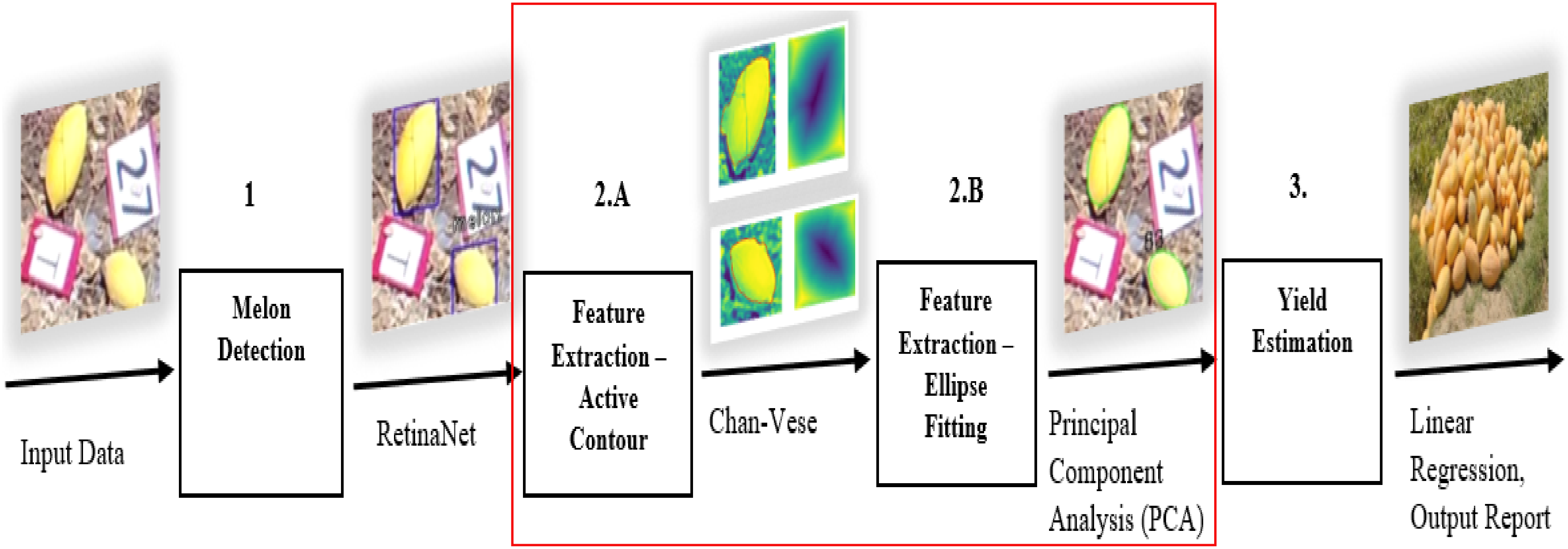
3 1 4

Task 1 – Detection results

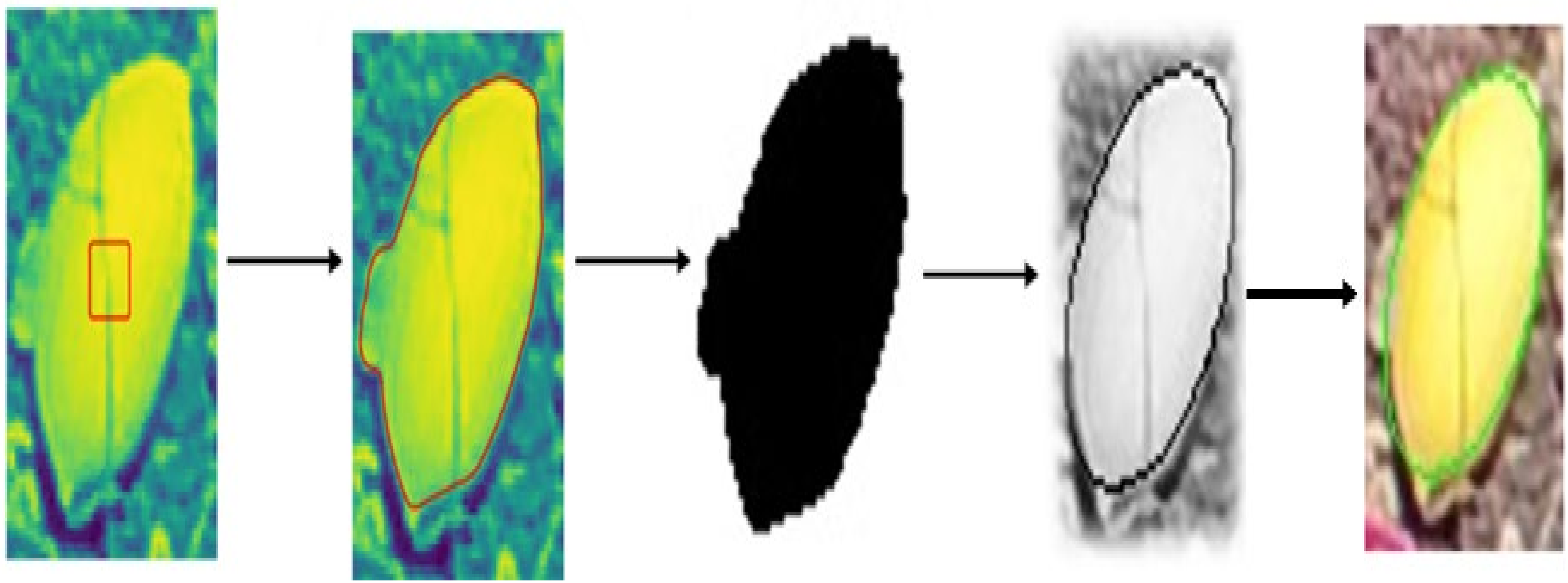


	2016 *Image 1*	2017 *Image 2*	2018 *Image 3*	2018 *Image 4*
True Positive	252	800	1032	180
False Positive	18	48	39	25
False Negative	35	89	12	11
Precision	0.93	0.94	0.96	0.88
Recall	0.88	0.90	0.99	0.94
F1-	0.90	0.92	0.98	0.91

Task 2 – Feature Extraction – Ellipse fitting



Task 2 – Single Melon Segmentation



Chan-Vese active contour Binary mask Ellipse fitting

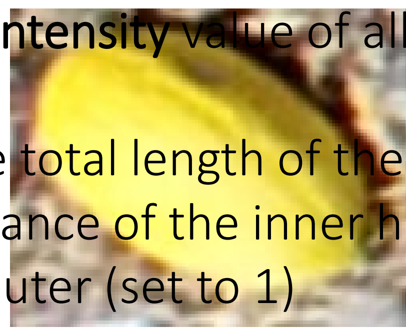
Chan-Vese active contour

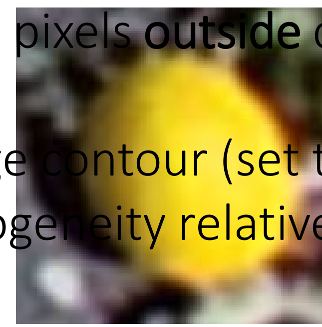
- A cost function which is solved iteratively using a gradient descent
- Homogeneous regions have low variance values

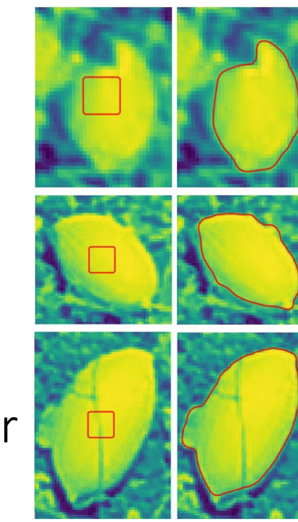
$$\arg\min F(c_1, c_2, C) =$$

$$\mu_1 \cdot Lenght(C) + \lambda_1 \int_{inside(C)} |u_0(x,y) - c_1|^2 \, dxdy + \lambda_2 \int_{outside(C)} |u_0(x,y) - c_2|^2 \, dxdy$$

- C_1 - the **mean intensity** value of all the pixels **inside** contour
- C_2 - the **mean intensity** value of all the pixels **outside** contour
- C – contour
- μ - penalize the total length of the edge contour (set to 1)
- λ_1 - the importance of the inner homogeneity relative to the homogeneity outer (set to 1)
- λ_2- the importance of the outer homogeneity relative to the inner homogeneity (set to 1)





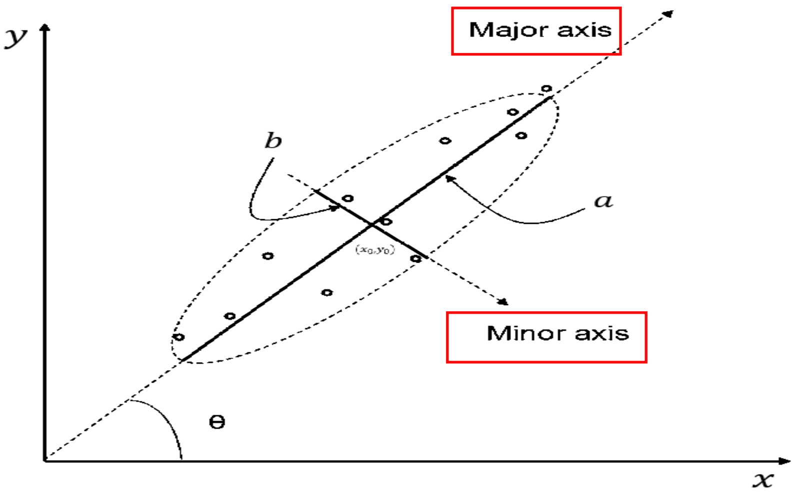


Feature extraction – Ellipse fitting (PCA)

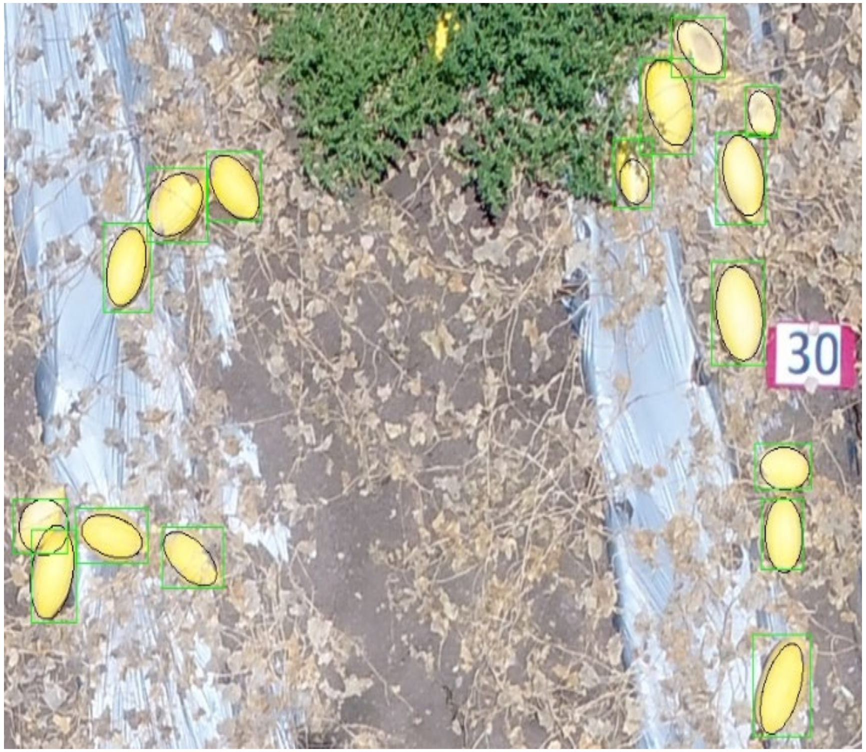
- The shape of the ellipse is determined by a set of 5 parameters

$$\frac{\left[(x-x_0)\cos(\theta)-(y-y_0)\sin(\theta)\right]^2}{a^2}+\frac{\left[(x-x_0)\sin(\theta)+(y-y_0)\cos(\theta)\right]^2}{b^2}=1$$

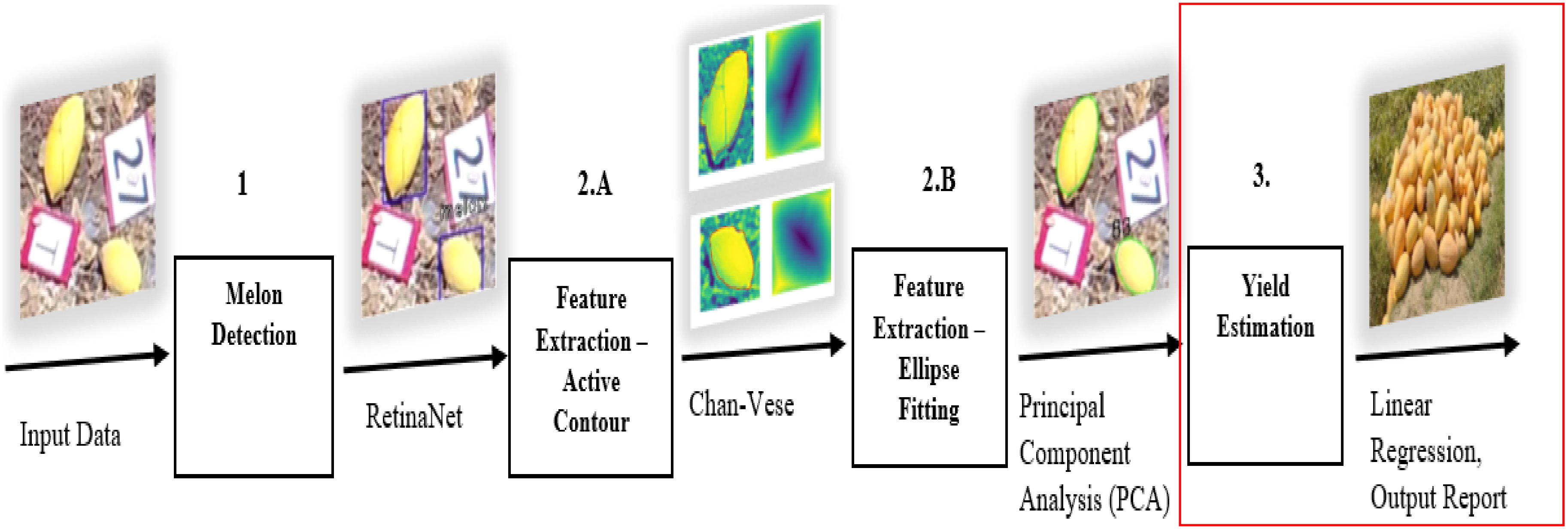
- Centroid x co-ordinate (x_0)
- Centroid y co-ordinate (y_0)
- **Semi-major axis ($\boldsymbol{a}$)**
- **Semi-minor axis ($\boldsymbol{b}$)**
- Angle of tilt (θ)



Ellipse fitting examples

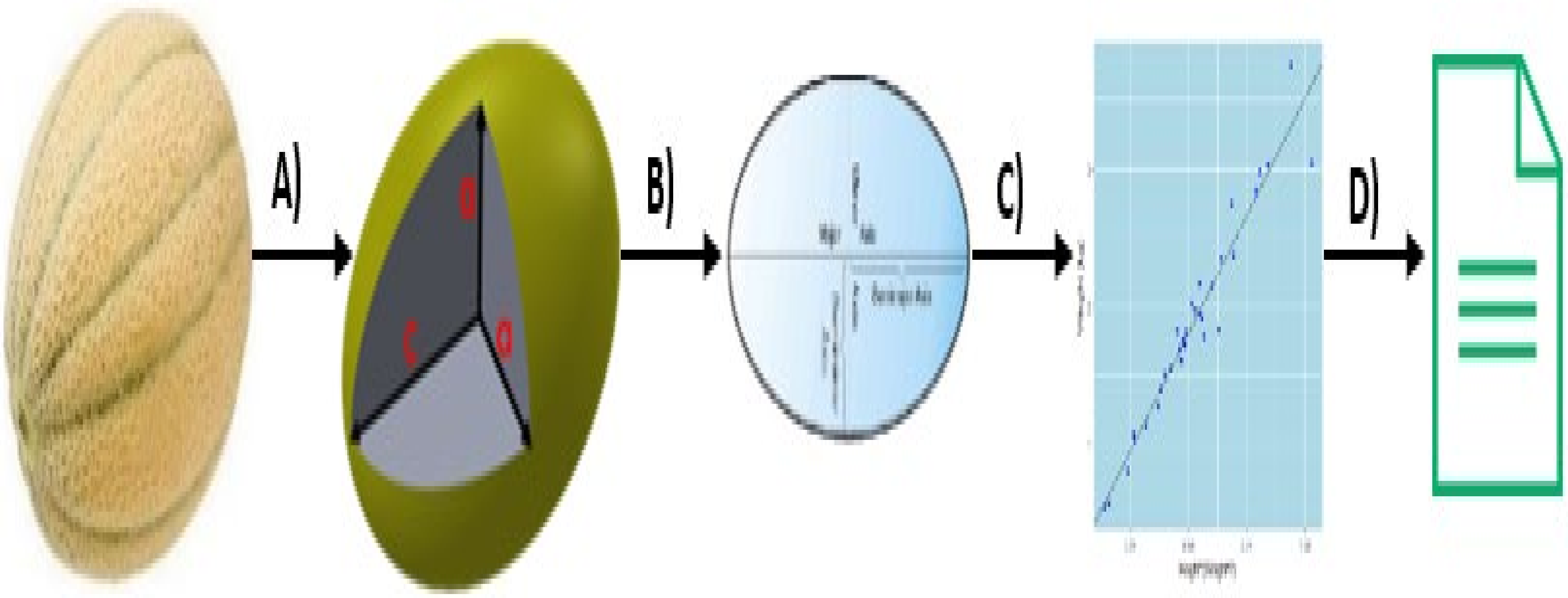


Task 3 – Yield estimation

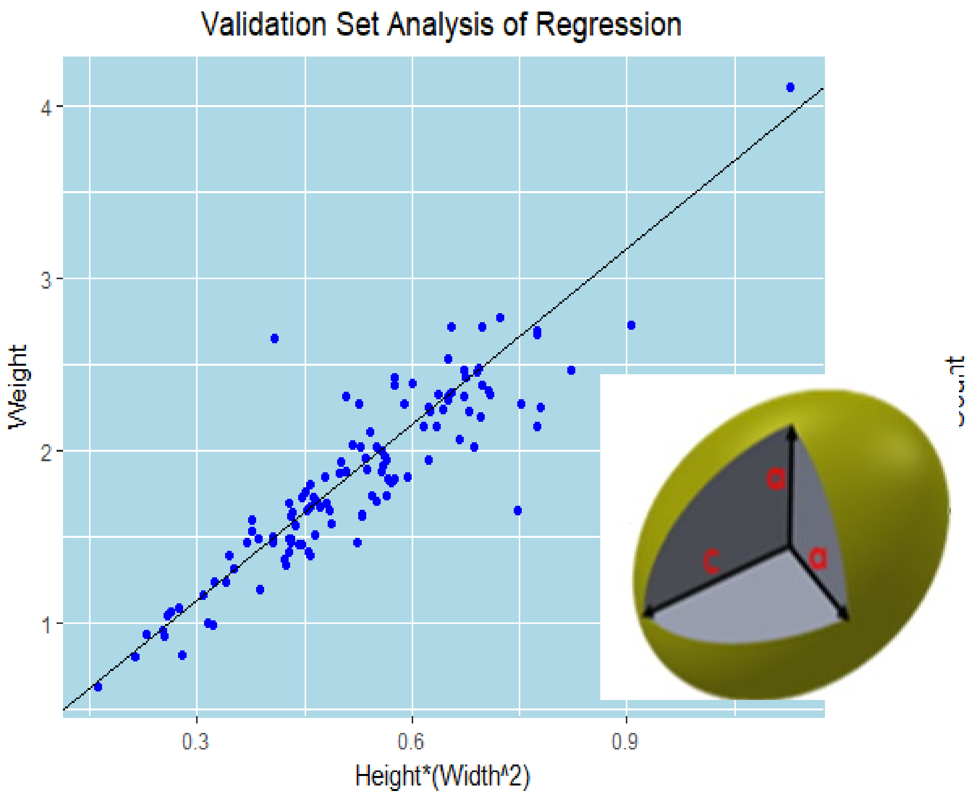


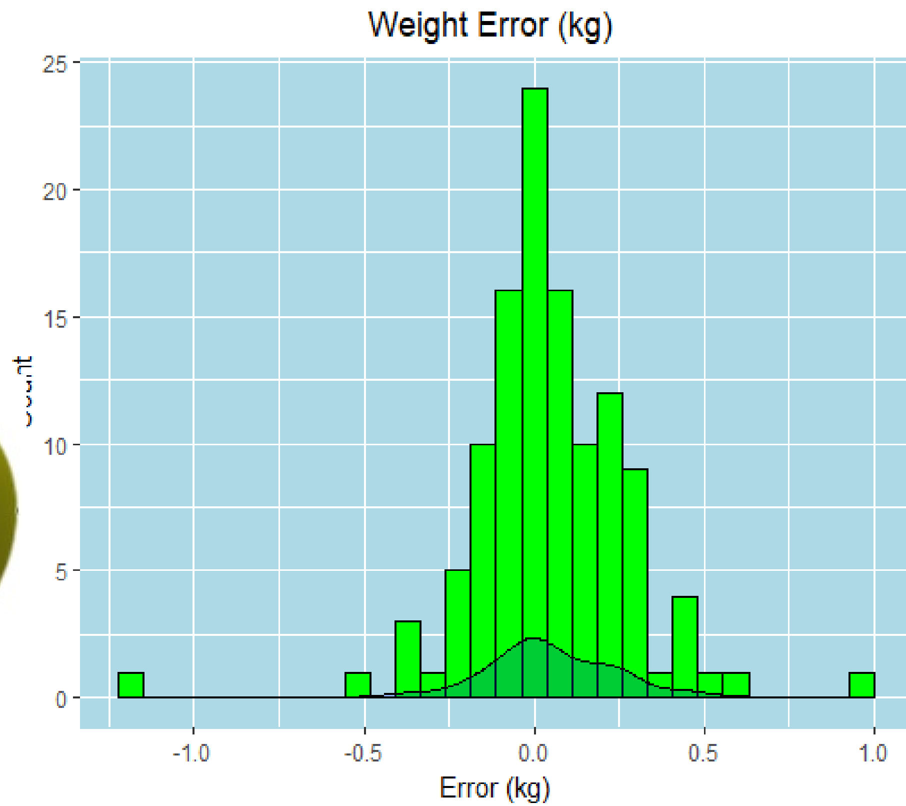
Task 3 – Yield estimation

- Yield estimation process include 4 stages:



Regression and it's quality





- Regression model was tested using 116 randomly selected melons from 2018 season
- The mean absolute percentage error (MAPE) for individual melon estimation was 9%
- An overweight overall yield estimation error of 2.9%

$$MAPE = \frac{1}{N}\sum_{N}\left|\frac{X-\tilde{X}}{X}\right|,$$

X is the actual value, $\tilde{X}$ is the prediction

Results - Melon yield report

- An example of report that the system generate:

Melon ID	Center Row	Center Col	Semi Major Axis [pix]	Semi Minor Axis [pix]	Semi Major Axis [cm]	Semi Minor Axis [cm]	Weight (Kg)
1	13	2773	35.14433	29.64361	10.0412	8.4696	0.921661
2	44	950	34.00361	22.79422	9.7153	6.5126	1.370152
3	30	5250	20.85496	11.95877	5.9586	3.4168	1.681457
4	33	2816	26.81252	19.01066	7.6607	5.4316	1.370152
5	1612	1926	23.81392	18.00779	6.8040	5.1451	0.643062
6	1619	2398	39.76633	23.39249	11.3618	6.6836	1.967597

Results - Melon yield report

Summary

- ➢ A systems for detection and yield estimation of melons from top view UAV images of a melon field have been developed.
- ➢ The system includes three main stages:
  - Melon detection (RetinaNet, NMS) - mAP = 0.914 | F-score>0.9
  - Feature extraction (Chan-Vese + PCA)
  - Yield estimation (Linear regression) - only 3% underestimation
- ➢ The system provides promising results .

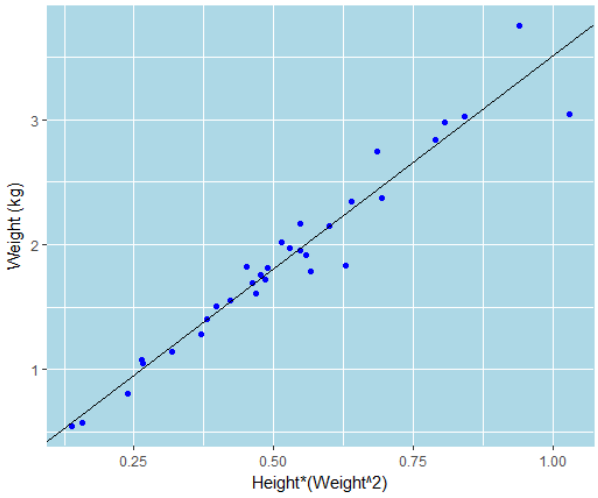
Thank You

Yield estimation – Linear regression

- The selected regression model was built from 30 randomly individual melons from 2017 season.
- The regression where based on max height (2*c) and max width (2*a) of each melon.

$$W = 0.1096653 + 0.003397929 \cdot c \cdot a^2$$

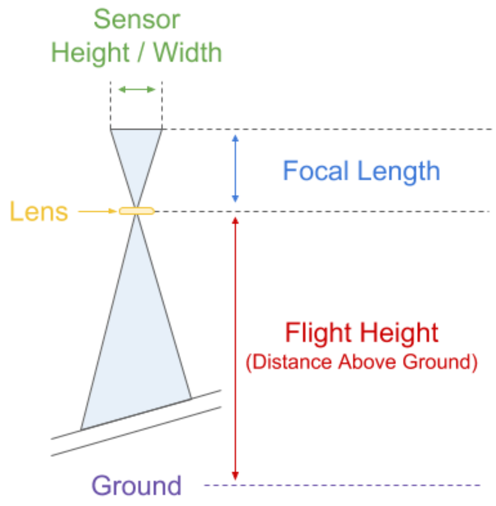
Type of correlation	Parameters combination	R^2_{Adj} value
Linear	$c + a$	0.914
Area	$c * a$	0.87
Volume	$c * a^2$	0.94



Ground sample distance (GSD)

- **Ellipse parameters given in pixels was translated to millimeters using GSD**

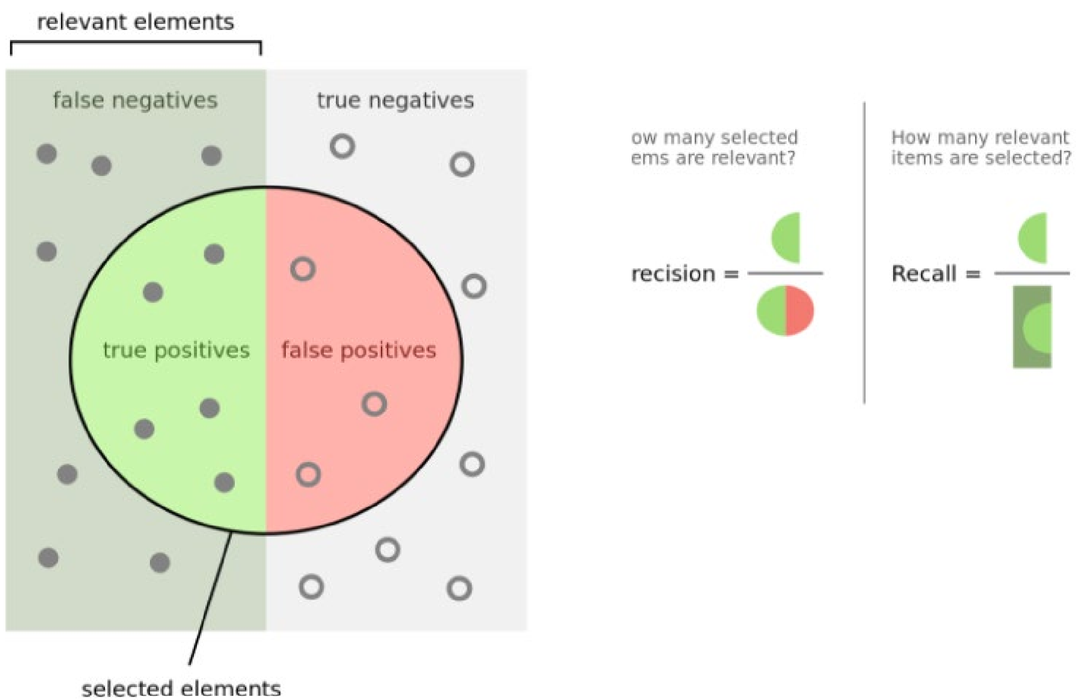
$$GSD = \frac{h \cdot \Delta p}{f}$$



- ***h*** **-** approximate height from grour
- **Δp -** sensors pixel size
- ***f*** **-** focal length

- **For each image we calculate the GSD separately - the fly height was not uniform**

Definitions



$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{tp}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}.$$

mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{N}\sum_{N}\left|\frac{X - \tilde{X}}{X}\right|,$$

X is the actual value, $\tilde{X}$ is the prediction

Ref: Wikipedia