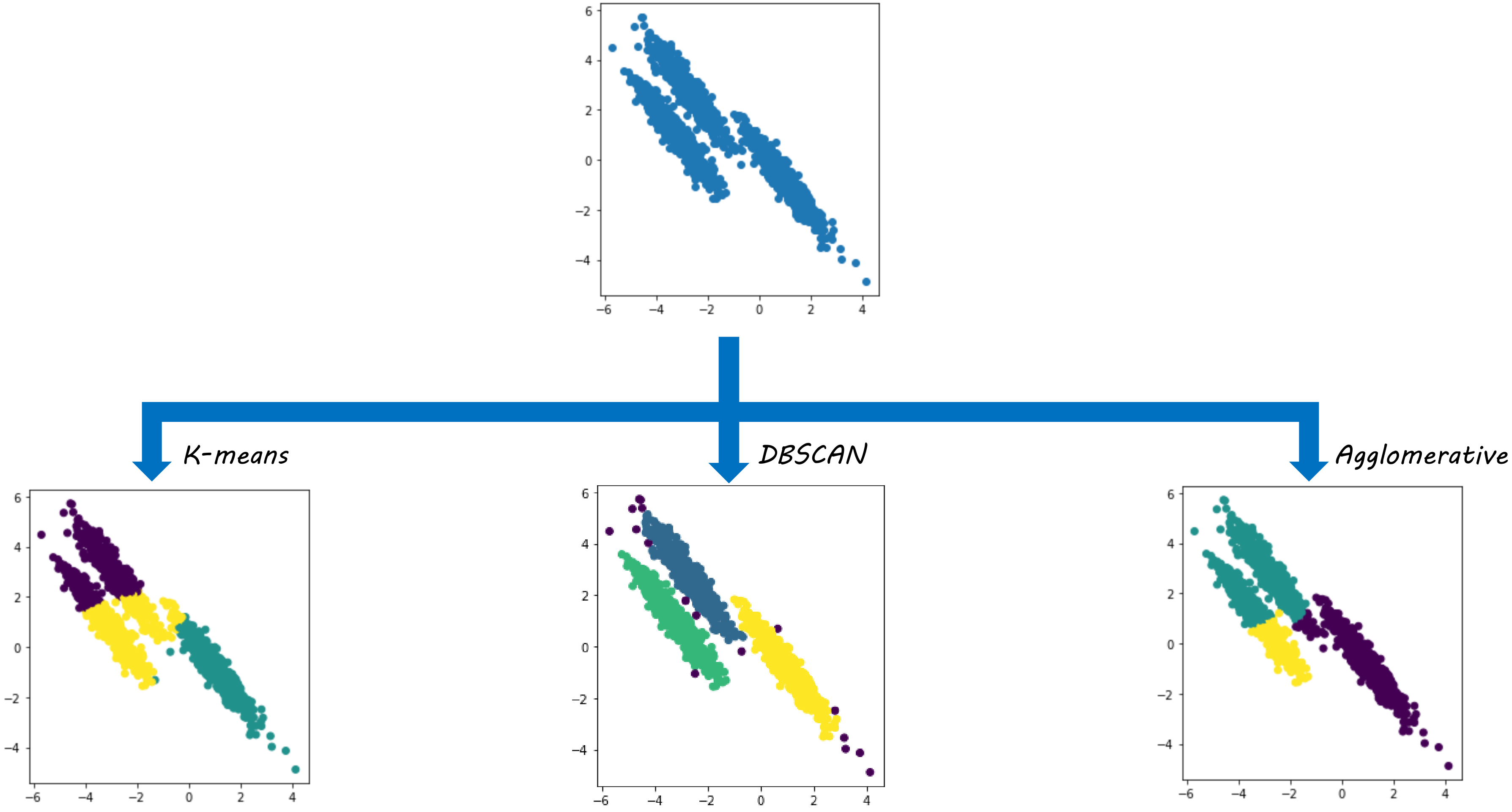
Clustering Algorithms and their Application to Facial Image Analysis

Hamid Sadeghi

nextera, FaceCup 1400 (2022)

Contents

Clustering summary



Evaluation

How to evaluate the quality of clusters?

Internal Evaluation (unsupervised)

Using clustered data itself

- Silhouette coefficient
- Davies-Bouldin index
- Dunn index

External Evaluation (supervised)

Using ground truth or gold standard

- Purity
- Rand Index
- Normalized Mutual Information (NMI)
- F-measure

Evaluation

- Purity

$$purity(\Omega, C) = \frac{1}{N}\sum_{k} \max_{j} |\omega_k \cup c_j|$$

Evaluation

- Purity

$$purity(\Omega, C) = \frac{1}{N}\sum_{k} \max_{j} |\omega_k \cup c_j|$$

clusters

classes

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

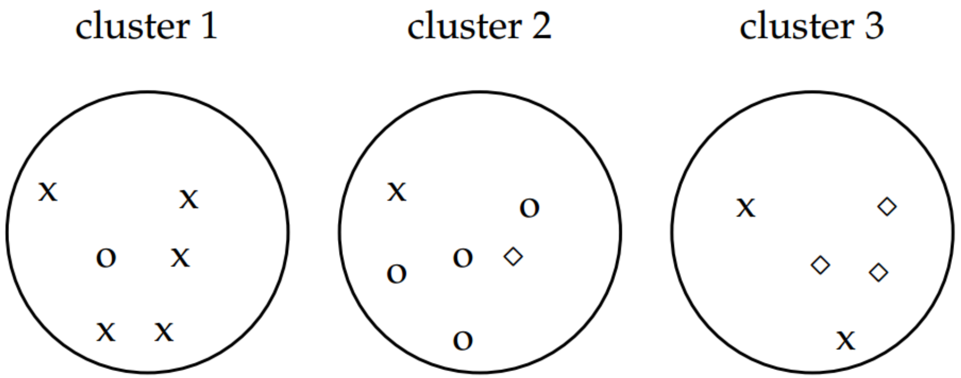
Evaluation

- Purity

$$purity(\Omega, C) = \frac{1}{N}\sum_{k} \max_{j} |\omega_k \cup c_j|$$

clusters

classes



► **Figure 16.4** Purity as an external evaluation criterion for cluster quality. Majority class and number of members of the majority class for the three clusters are: x, 5 (cluster 1); o, 4 (cluster 2); and ◇, 3 (cluster 3). Purity is $(1/17) \times (5+4+3) \approx 0.71$.

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

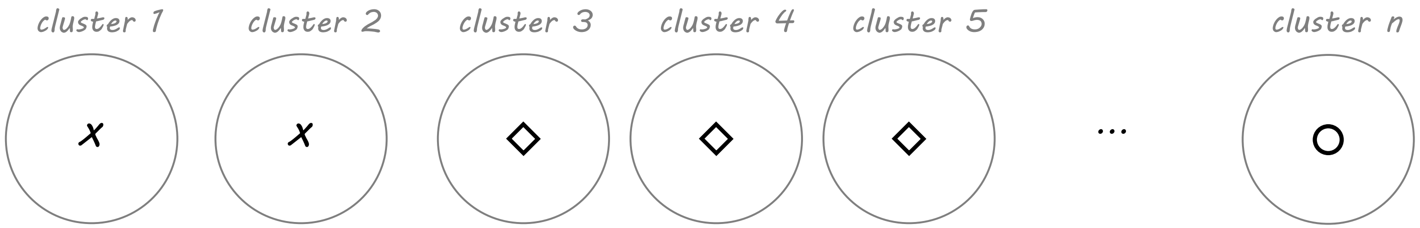
- Purity

$$purity(\Omega, C) = \frac{1}{N}\sum_{k} \max_{j} |\omega_k \cup c_j|$$

clusters

classes

If each data gets 1 cluster ➡ purity = 1 (!)



Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Rand Index

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Rand Index

True Positive pairs

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

True Negative pairs

False Positive pairs

False Negative pairs

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Rand Index

True Positive pairs

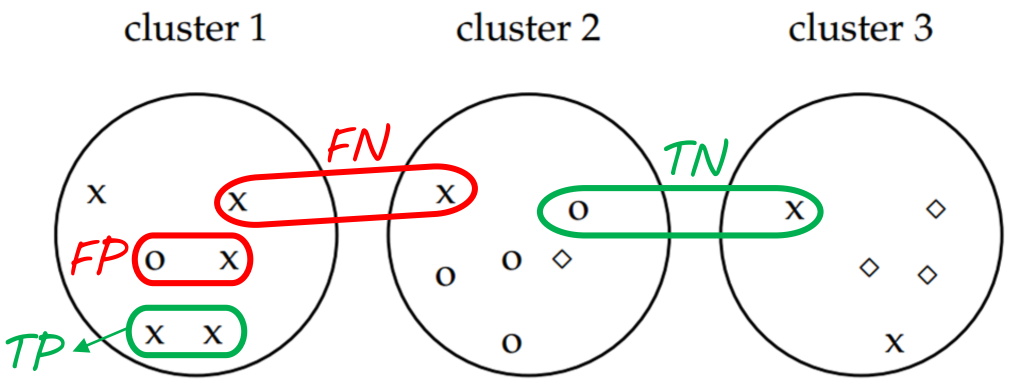
True Negative pairs

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

False Positive pairs

False Negative pairs

Sample pairs:



Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Rand Index

True Positive pairs

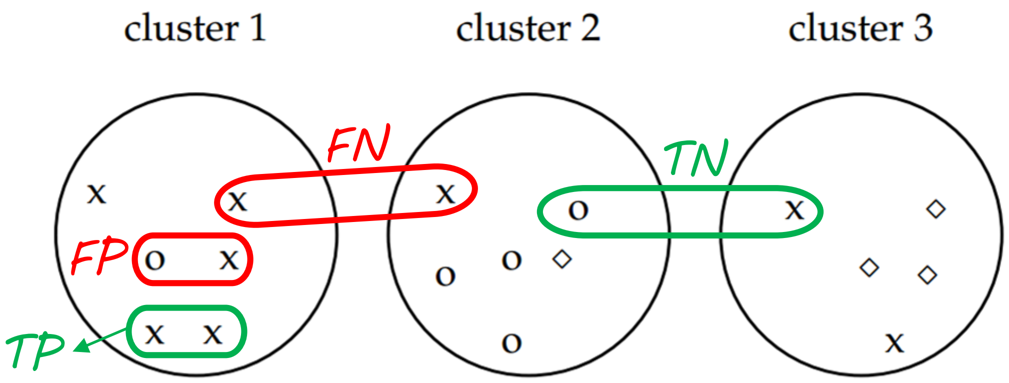
True Negative pairs

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

False Positive pairs

False Negative pairs

Sample pairs:



$$TP + FP = \binom{6}{2} + \binom{6}{2} + \binom{5}{2} = 40$$ Positive pairs

$$TP = \binom{5}{2} + \binom{4}{2} + \binom{3}{2} + \binom{2}{2} = 20$$

$$\Rightarrow FP = 40 - 20 = 20$$

	Same cluster	Different clusters
Same class	TP = 20	FN = 24
Different classes	FP = 20	TN = 72

Evaluation

- Rand Index

True Positive pairs

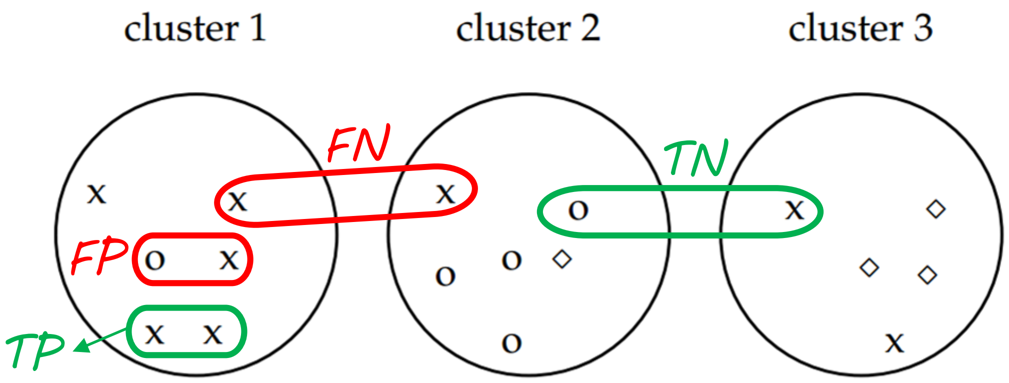
True Negative pairs

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

False Positive pairs

False Negative pairs

Sample pairs:



	Same cluster	Different clusters
Same class	TP = 20	FN = 24
Different classes	FP = 20	TN = 72

$$RI = (20 + 72)/(20 + 20 + 24 + 72) \approx 0.68$$

Evaluation

- Normalized Mutual Information (NMI)

$$NMI(\Omega, \mathrm{C}) = \frac{I(\Omega, \mathrm{C})}{[H(\Omega) + H(\mathrm{C})]/2}$$

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Normalized Mutual Information (NMI)

Mutual Information

clusters

$$NMI(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2}$$

classes

entropy

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Normalized Mutual Information (NMI)

Mutual Information

clusters

$$NMI(\Omega, C) = \frac{I(\Omega\,, C)}{[H(\Omega) + H(C)]/2}$$

classes

entropy

$$I(\Omega, C) = \sum_k \sum_j P(\omega_k \cap c_j)\, log \frac{P(\omega_k \cap c_j)}{P(\omega_k)\, P(c_j)} = \sum_k \sum_j \frac{|\omega_k \cap c_j|}{N}\, log \frac{N\, |\omega_k \cap c_j|}{|\omega_k|\, |c_j|}$$

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- Normalized Mutual Information (NMI)

Mutual Information

clusters

$$NMI(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2}$$

classes

entropy

probabilities of a data being in the intersection of ω_k and c_j

$$I(\Omega, C) = \sum_k \sum_j P(\omega_k \cap c_j) \log \frac{P(\omega_k \cap c_j)}{P(\omega_k)\, P(c_j)} = \sum_k \sum_j \frac{|\omega_k \cap c_j|}{N} \log \frac{N\,|\omega_k \cap c_j|}{|\omega_k|\,|c_j|}$$

probabilities of a data being in cluster ω_k

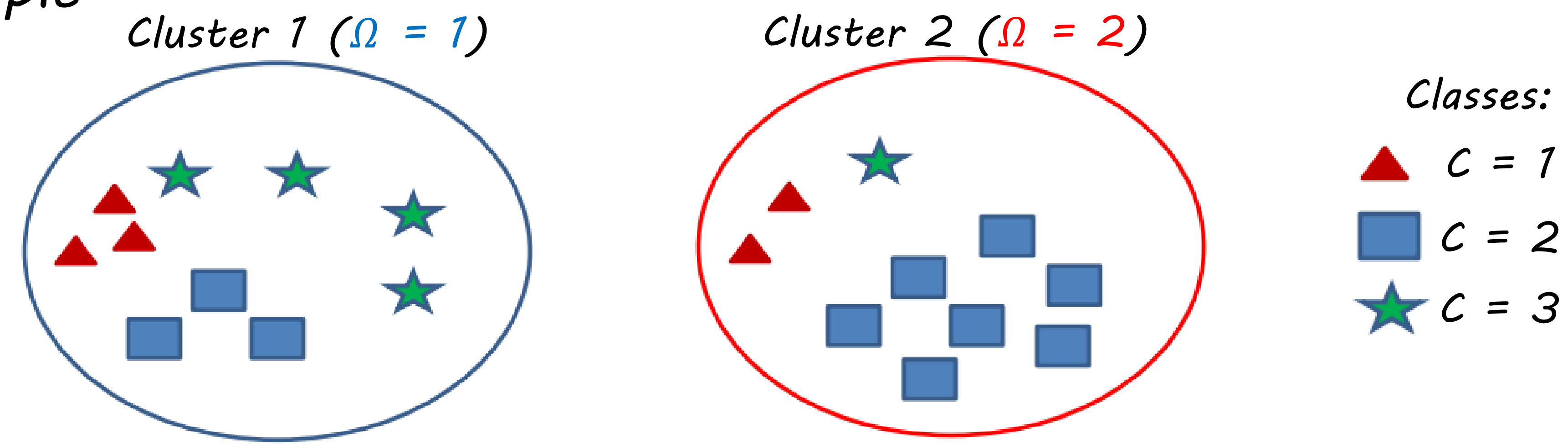
probabilities of a data being in class c_j

$$H(\Omega) = \sum_k P(\omega_k) log P(\omega_k) = \sum_k \frac{\omega_k}{N} \log \frac{\omega_k}{N}$$

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

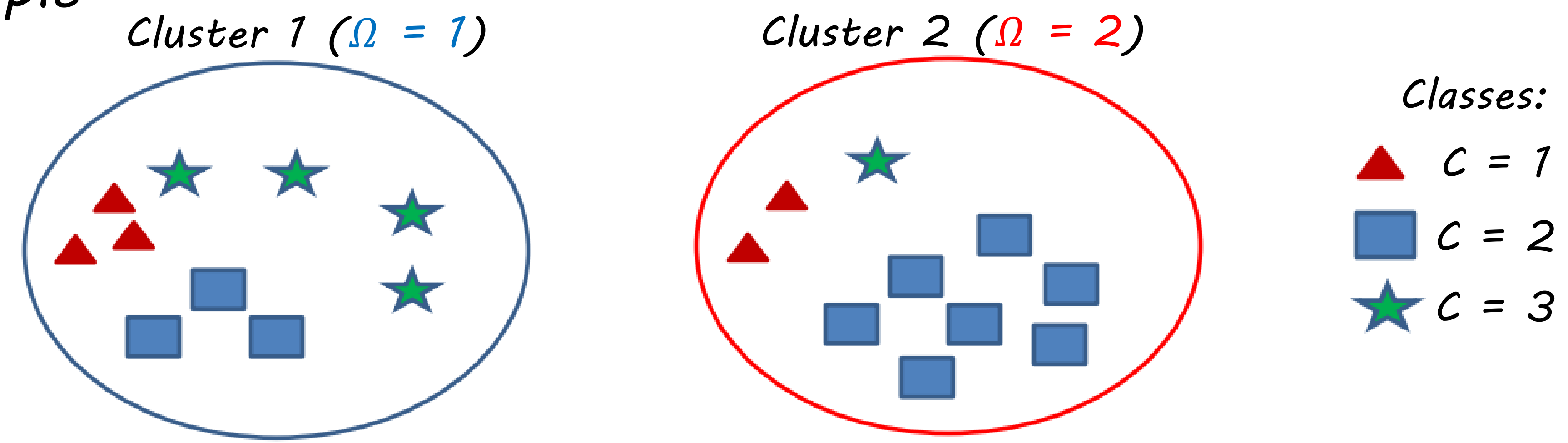
- NMI example



https://course.ccs.neu.edu/cs6140sp15/7_locality_cluster/Assignment-6/NMI.pdf

Evaluation

- NMI example



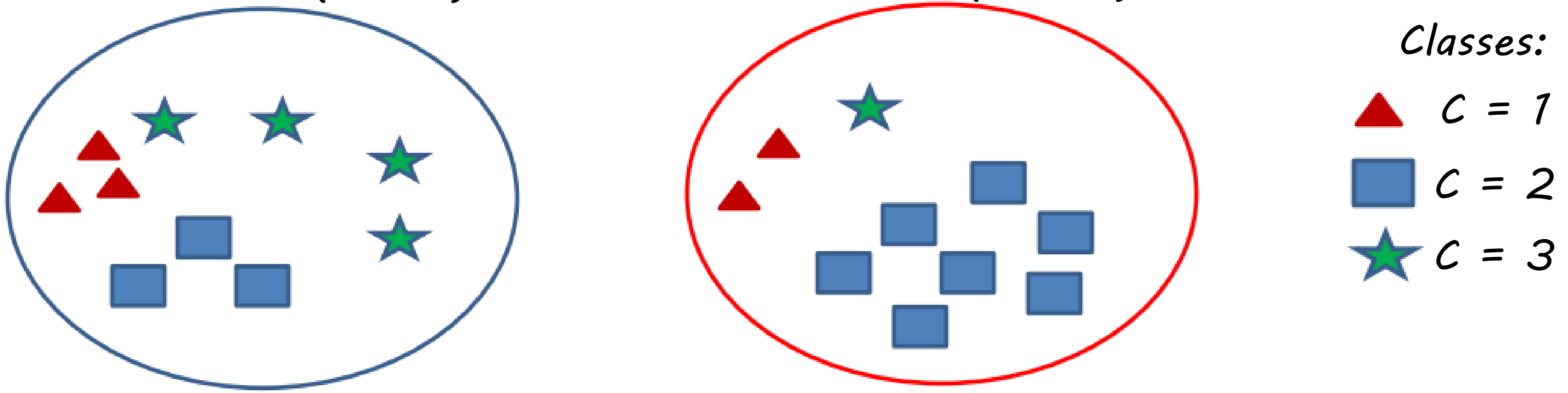
P( )=5/20=1/4 P( )=5/20=1/4 P( )=10/20=1/2

$$H(C) = -\frac{1}{4}\log(\frac{1}{4}) - \frac{1}{4}\log(\frac{1}{4}) - \frac{1}{2}\log(\frac{1}{2}) = 1.5$$

Evaluation

- NMI example

Cluster 1 ($\Omega = 1$) Cluster 2 ($\Omega = 2$)



P(▲)=5/20=1/4 P(★)=5/20=1/4 P(■)=10/20=1/2

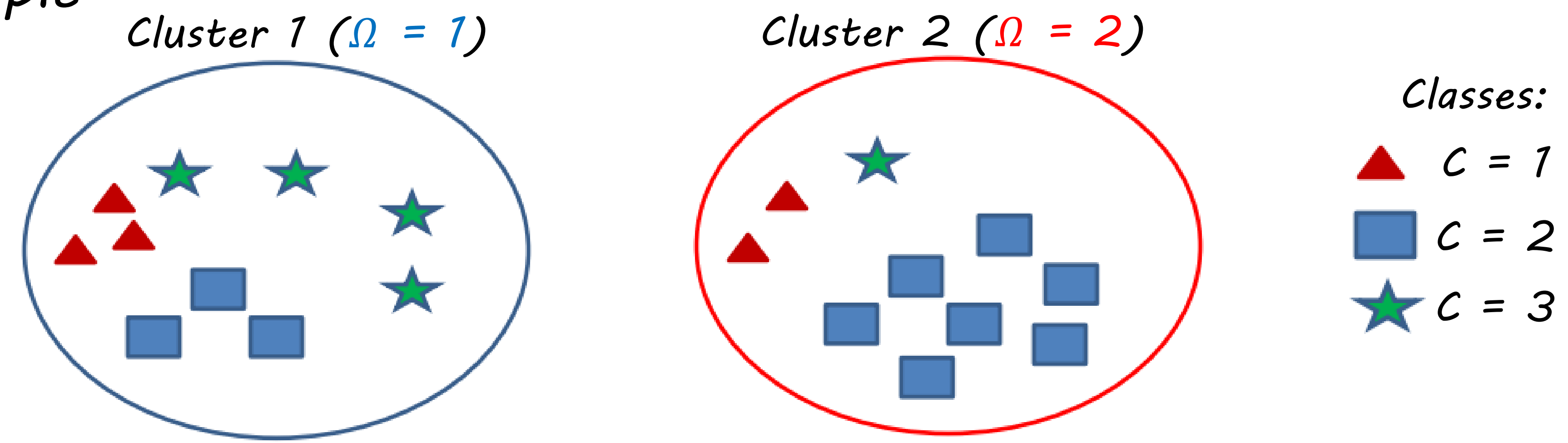
$$H(C) = -\frac{1}{4}\log(\frac{1}{4}) - \frac{1}{4}\log(\frac{1}{4}) - \frac{1}{2}\log(\frac{1}{2}) = 1.5$$

$P(\Omega = 1)$=10/20=1/2 $P(\Omega = 2)$= 10/20=1/2

$$H(\Omega) = -\frac{1}{2}\log(\frac{1}{2}) - \frac{1}{2}\log(\frac{1}{2}) = 1$$

Evaluation

- NMI example



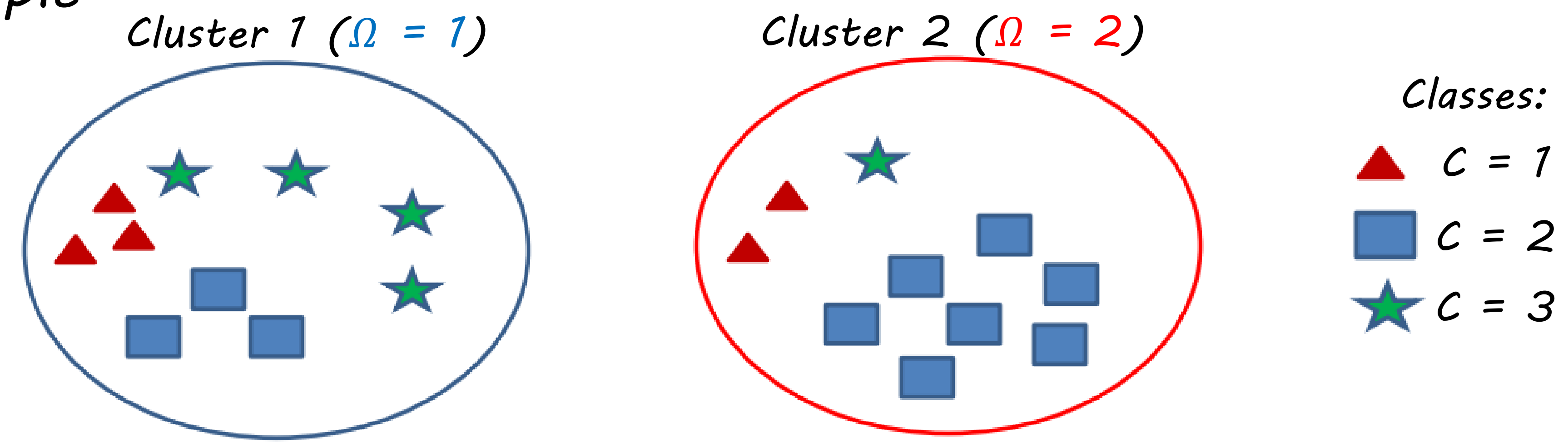
$$I(\Omega, C) = \sum_k \sum_j \frac{|\omega_k \cap c_j|}{N} log \frac{N\,|\omega_k \cap c_j|}{|\omega_k|\,|c_j|}$$

$$= \frac{3}{20}\log\left(\frac{20\times 3}{10\times 5}\right) + \frac{3}{20}\log\left(\frac{20\times 3}{10\times 10}\right) + \frac{4}{20}\log\left(\frac{20\times 4}{10\times 5}\right) + \frac{2}{20}\log\left(\frac{20\times 2}{10\times 5}\right) + \frac{7}{20}\log\left(\frac{20\times 7}{10\times 10}\right) + \frac{1}{20}\log\left(\frac{20\times 1}{10\times 5}\right) = 0.1361$$

https://course.ccs.neu.edu/cs6140sp15/7_locality_cluster/Assignment-6/NMI.pdf

Evaluation

- NMI example



$$NMI(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2} = \frac{0.1361}{[1 + 1.5]/2} = 0.1089$$

https://course.ccs.neu.edu/cs6140sp15/7_locality_cluster/Assignment-6/NMI.pdf

Evaluation

- Precision

(correct rate within clusters) $P = \dfrac{TP}{TP + FP}$

- Recall

(correct rate within classes/ground truth) $R = \dfrac{TP}{TP + FN}$

- F-measure

$$F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

$\beta = 0$: $F = P$
$\beta\uparrow$: R

Ch. D. Manning , et al. An Introduction to Information Retrieval, Ch. 16: Flat clustering, 2009

Evaluation

- *implementation:*

https://colab.research.google.com/github/hamidsadeghi68/face-clustering/blob/main/clustering_kmeans.ipynb

https://colab.research.google.com/github/hamidsadeghi68/face-clustering/blob/main/clustering_dbscan.ipynb

https://colab.research.google.com/github/hamidsadeghi68/face-clustering/blob/main/clustering_agglomerative.ipynb

Face Analysis

- Introduction
- Face Detection & Preprocessing
- Face Recognition
- A Complete Face Clustering Algorithm

Face Analysis: Introduction

Face Recognition (It is different from face clustering!)

Identification (1:N)	Verification (1:1)

Face Analysis: Introduction

Face Recognition (It is different from face clustering!)

Identification (1:N)

?

Emilia Clarke
Angelina Jolie
Abbas Bouazar
.
.
.
unknown

N

For open-set problem

Verification (1:1)

Face Analysis: Introduction

Face Recognition (It is different from face clustering!)

Identification (1:N)

?

Emilia Clarke
Angelina Jolie
Abbas Bouazar
.
.
.

N

unknown

For open-set problem

Verification (1:1)

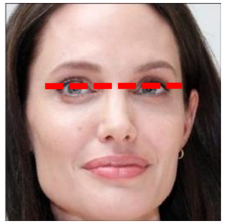
?

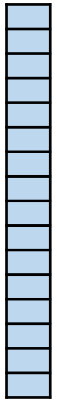
Are matched?

Face Analysis: Introduction

- Block diagram of a face recognition system

Image ⇨ Face Detection (face, landmarks, etc.) ⇨ Preprocessing (alignment, etc.) ⇨ Feature Extraction ⇨ Classification ⇨ Identity

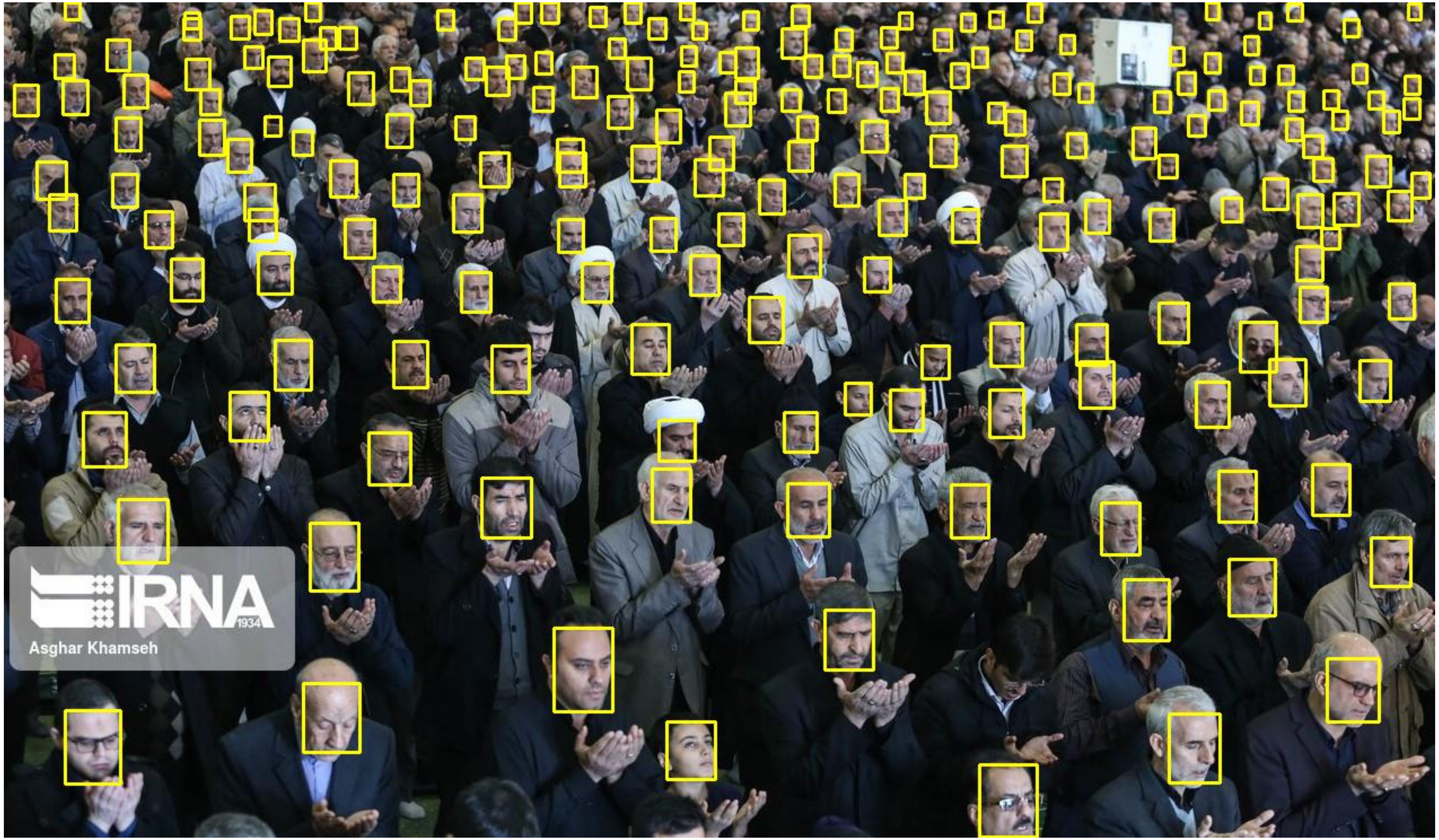




Softmax
K-Nearest Neighbors
etc.

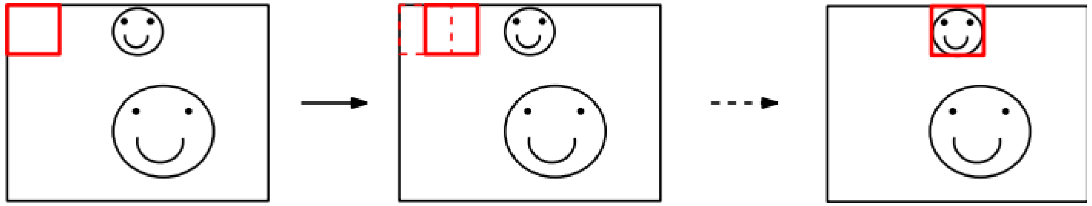
We can not add/delete identity

Face Analysis: Face Detection



Face Analysis: Face Detection

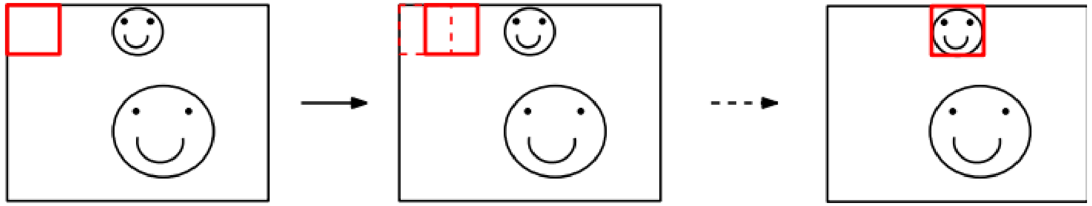
- Haar Cascade


F. Comaschi, et al., RASW: a Run-time Adaptive Sliding Window to Improve Viola-Jones Object Detection

P. Viola & M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", CVPR 2001

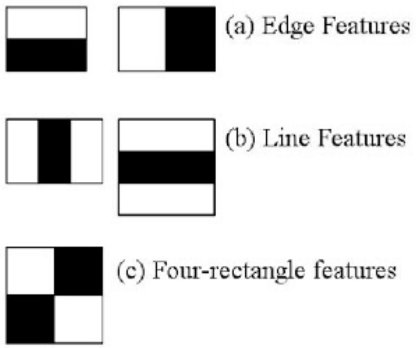
Face Analysis: Face Detection

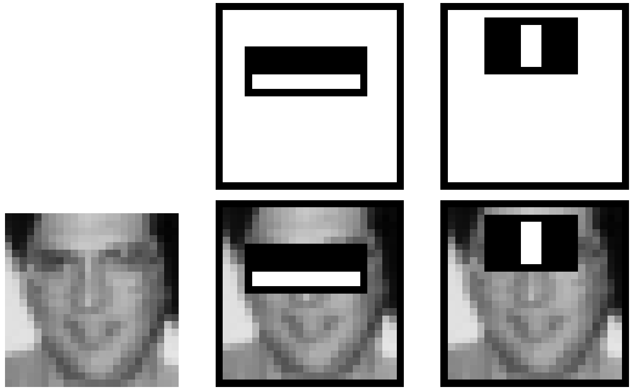
- Haar Cascade

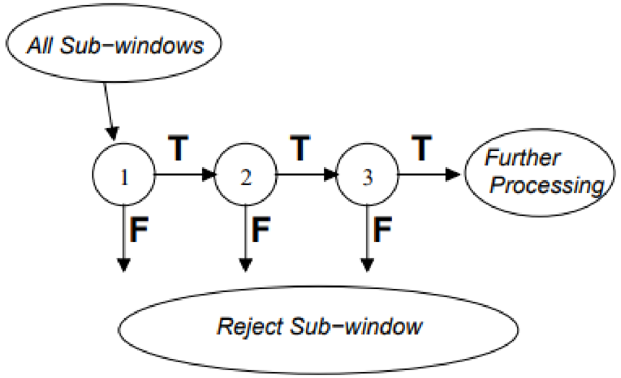


F. Comaschi, et al., RASW: a Run-time Adaptive Sliding Window to Improve Viola-Jones Object Detection

Each feature: ∑(pixels under black rectangle) - ∑(pixels under white rectangle)







Code:

https://towardsdatascience.com/face-detection-with-haar-cascade-727f68dafd08
https://opencv24-python-tutorials.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html

P. Viola & M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", CVPR 2001

Face Analysis: Face Detection

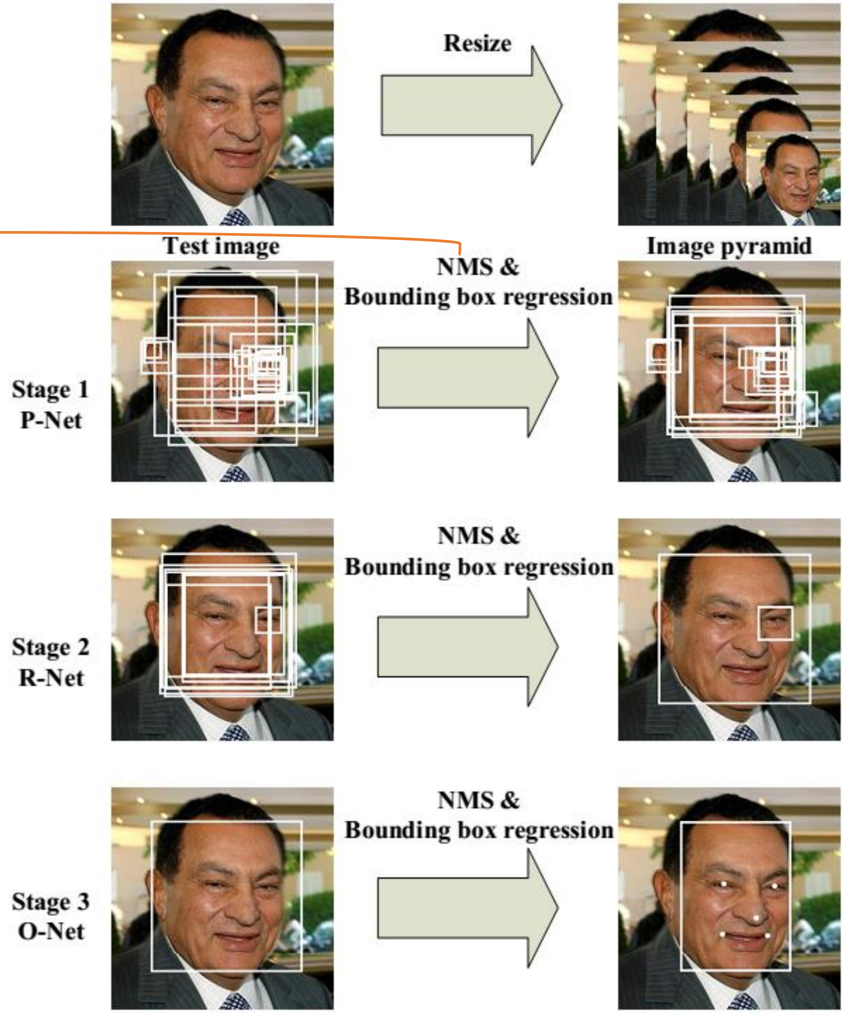
- MTCNN

merge highly overlapped candidates

fast Proposal Network

Refinement Network

Output Network



Detection on different scales

candidate windows and bounding boxes as DDFD

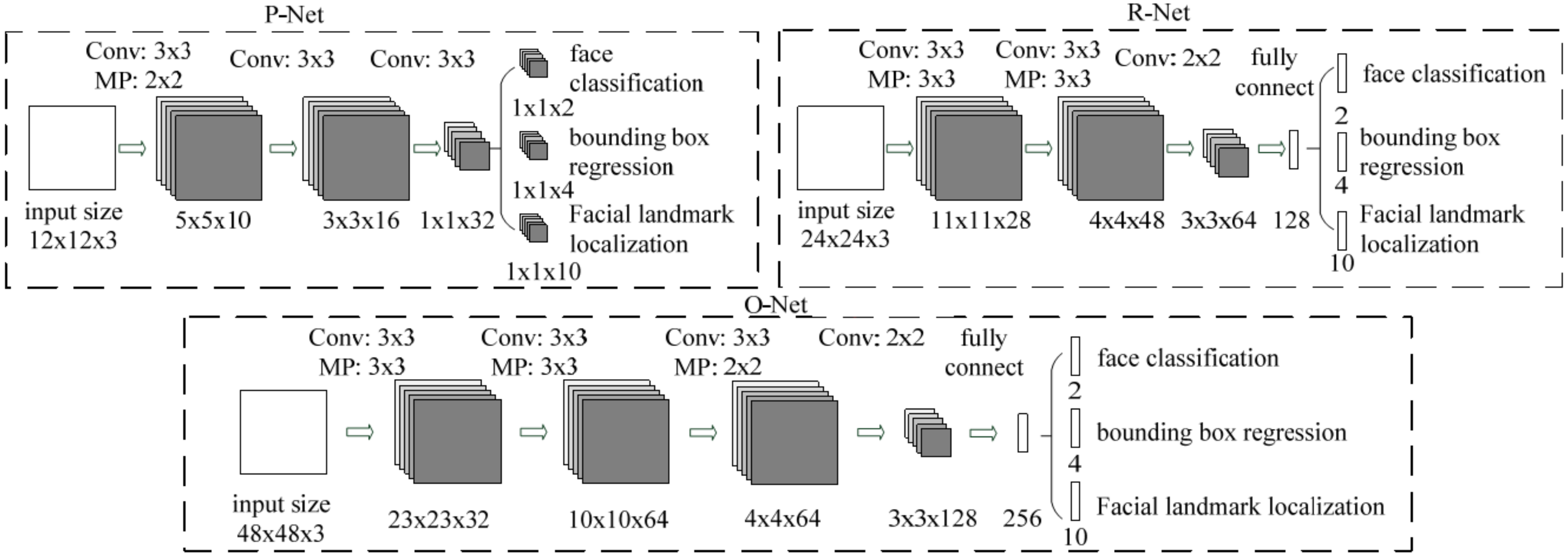
rejects a large number of false candidates

rejects false candidates and cal. landmarks

Zhang, et al., Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, IEEE Signal Processing Letters, 2016

Face Analysis: Face Detection

- MTCNN



Code:

Python package: https://github.com/ipazc/mtcnn
https://github.com/davidsandberg/facenet/tree/master/src/align
https://github.com/kpzhang93/MTCNN_face_detection_alignment

Zhang, et al., Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, IEEE Signal Processing Letters, 2016

Face Analysis: Face Detection

- MTCNN

Face classification loss: $L_i^{det} = -\left(y_i^{det}\, log(p_i) + (1 - y_i^{det})(log(1 - p_i))\right)$ $\quad y_i^{det} \epsilon \{0,1\}$

Bounding box regression loss: $L_i^{box} = \left\| \hat{y}_i^{box} - y_i^{box} \right\|_2^2$

landmark localization loss: $L_i^{Landmark} = \left\| \hat{y}_i^{Landmark} - y_i^{Landmark} \right\|_2^2$

overall learning target: $\min \sum_{i=1}^{N} \sum_{j \in (det,box,landmark)} \alpha_j \beta_i^j L_i^j$

$\beta_i^j \epsilon \{0,1\}$ (sample type indicator)

task importance: $\begin{cases} PNet, RNet: & \alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{landmark} = 0.5 \\ ONet: & \alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{landmark} = 1 \end{cases}$

Zhang, et al., Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, IEEE Signal Processing Letters, 2016

Face Analysis: Face Detection

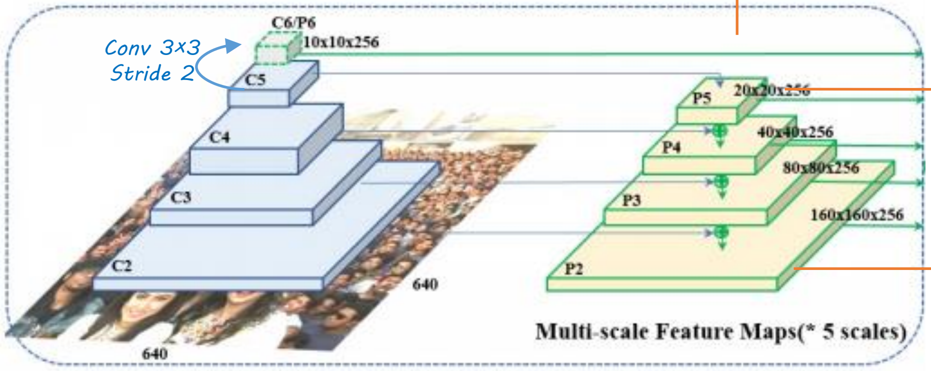
- Implementation (MTCNN):

  https://colab.research.google.com/github/hamidsadeghi68/face-clustering/blob/main/face_detection_mtcnn.ipynb

Face Analysis: Face Detection

- RetinaFace

C2-C5: ResNet residual stage feature pyramid



to capture larger faces

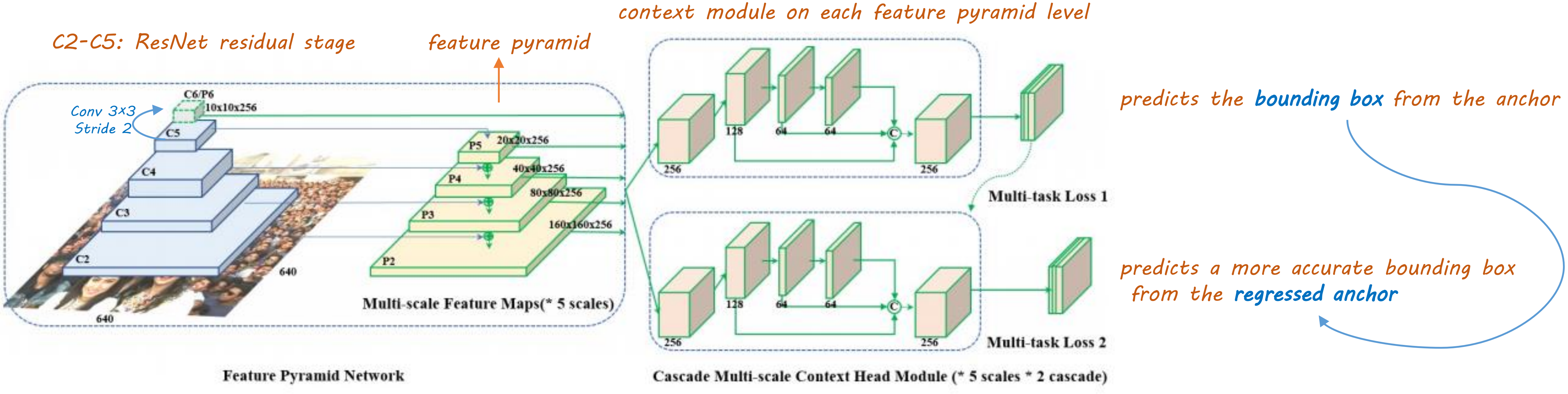
to capture tiny faces by **tiling** small **anchors**

75% of all anchors (102,300 anchors for a 640 × 640 image) are from P2

Feature Pyramid Network

Face Analysis: Face Detection

- RetinaFace

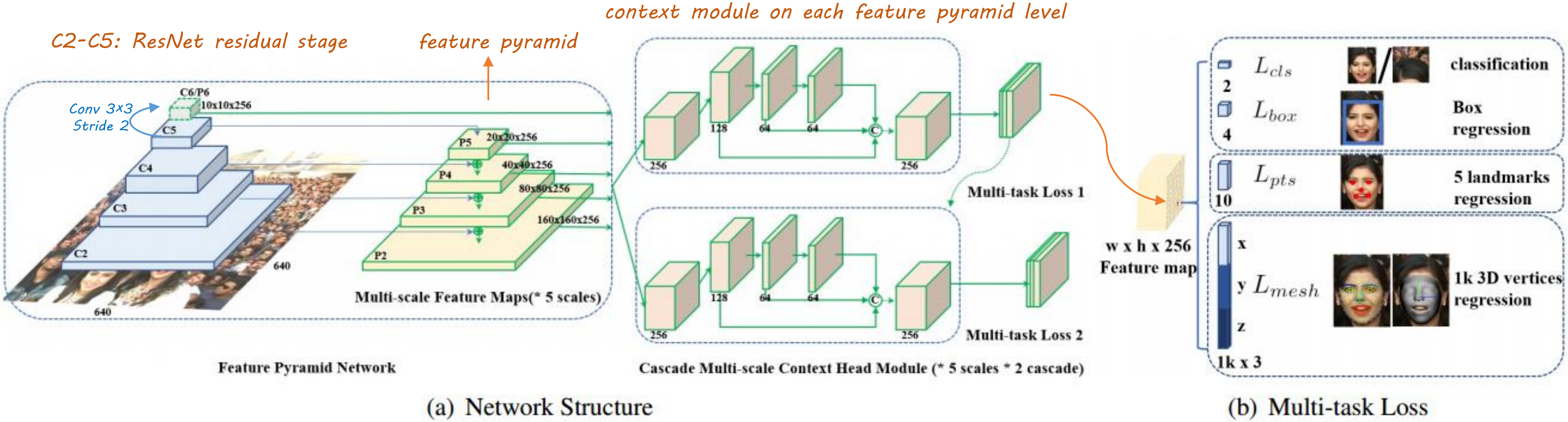


(a) Network Structure

Deng, et al., RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild, CVPR 2020

Face Analysis: Face Detection

- RetinaFace



(a) Network Structure (b) Multi-task Loss

Code:

https://github.com/deepinsight/insightface/tree/master/detection/retinaface

Deng, et al., RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild, CVPR 2020