L.N. Gumilyov Eurasian National University
Faculty of Information Technology
Department of «Information Systems»

# Comparison of object detection with

## Viola Jones Algorithm

IWS №6 by discipline «Computer graphics and pattern recognition»

Presented by: Toleubay D.M.
Checked by: Prof. Dr. Zhukabayeva T.K.

# INTRODUCTION

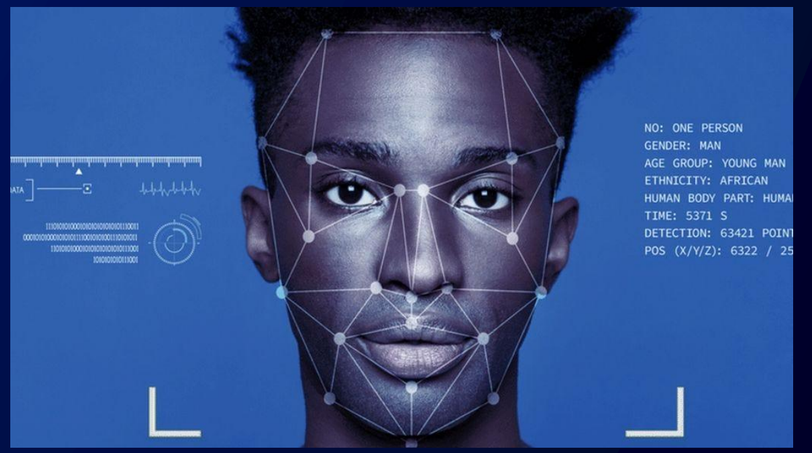
Modern face detection algorithms play an important role in biometric systems, video surveillance and other areas of computer vision. Among the most effective methods are MTCNN (Multi-task Cascaded Convolutional Neural Network) and Faster R-CNN (Region-based Convolutional Neural Network).

MTCNN is a cascaded convolutional neural network, specially designed for accurate and fast face detection. It combines three stages: preliminary detection, boundary refinement and keypoint extraction (eyes, nose, mouth).

Faster R-CNN is a powerful two-stage algorithm that uses regressive neural network models to accurately detect objects, including faces. It combines high accuracy with the ability to detect multiple objects in an image.

# What is MTCNN?

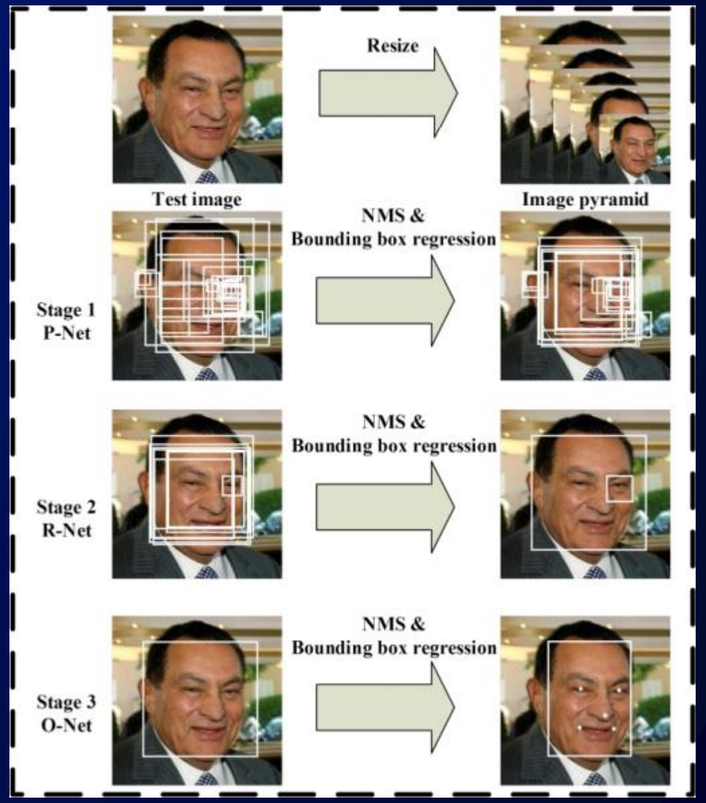
MTCNN (Multi-Task Cascaded Convolutional Networks) algorithm is one such technology that has revolutionized the field of face detection and recognition. Developed in 2016, the MTCNN algorithm uses a cascading series of neural networks to detect, align, and extract facial features from digital images with high accuracy and speed. In this article, we will delve into the details of the MTCNN algorithm, its architecture, working principles, and real-world applications, and explore why it has become a popular choice for face detection and recognition tasks.



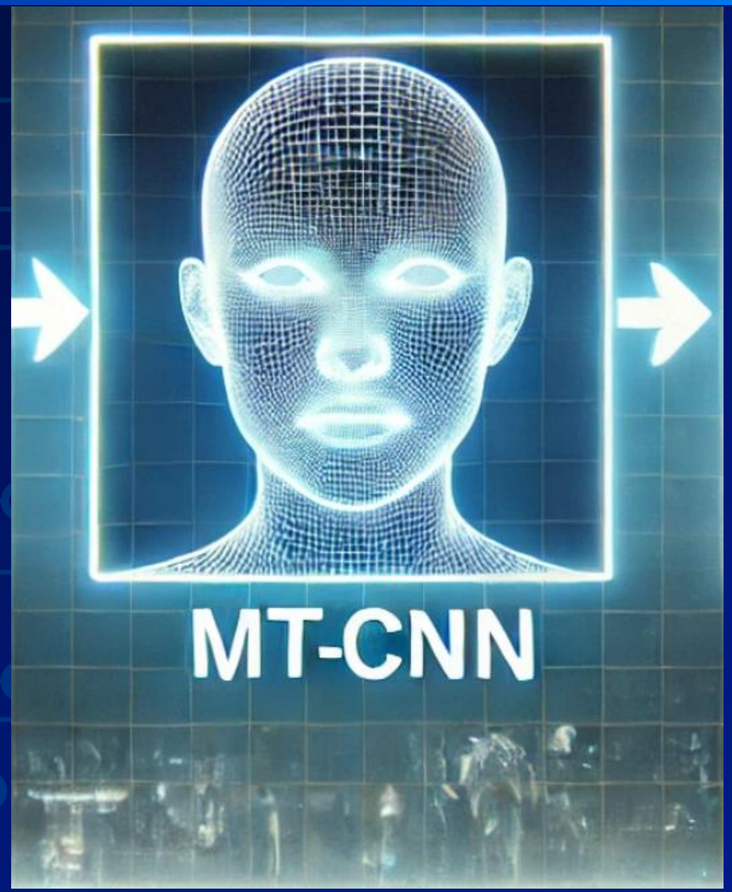
# MTCNN

MTCNN (Multi-task Cascaded Neural Network) detects faces and facial landmarks on images/videos. This method was proposed by Kaipeng Zhang et al. in their paper ‘Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks’, IEEE Signal Processing Letters, Volume: 23 Issue: 10.

The whole concept of MTCNN can be explained in three stages out of which, in the third stage, facial detection and facial landmarks are performed simultaneously. These stages consists of various CNN’s with varying complexities.



## A simpler explanation of the three stages of MTCNN can be as follows :




In the first stage the MTCNN creates multiple frames which scans through the entire image starting from the top left corner and eventually progressing towards the bottom right corner. The information retrieval process is called P-Net(Proposal Net) which is a shallow, fully connected CNN.

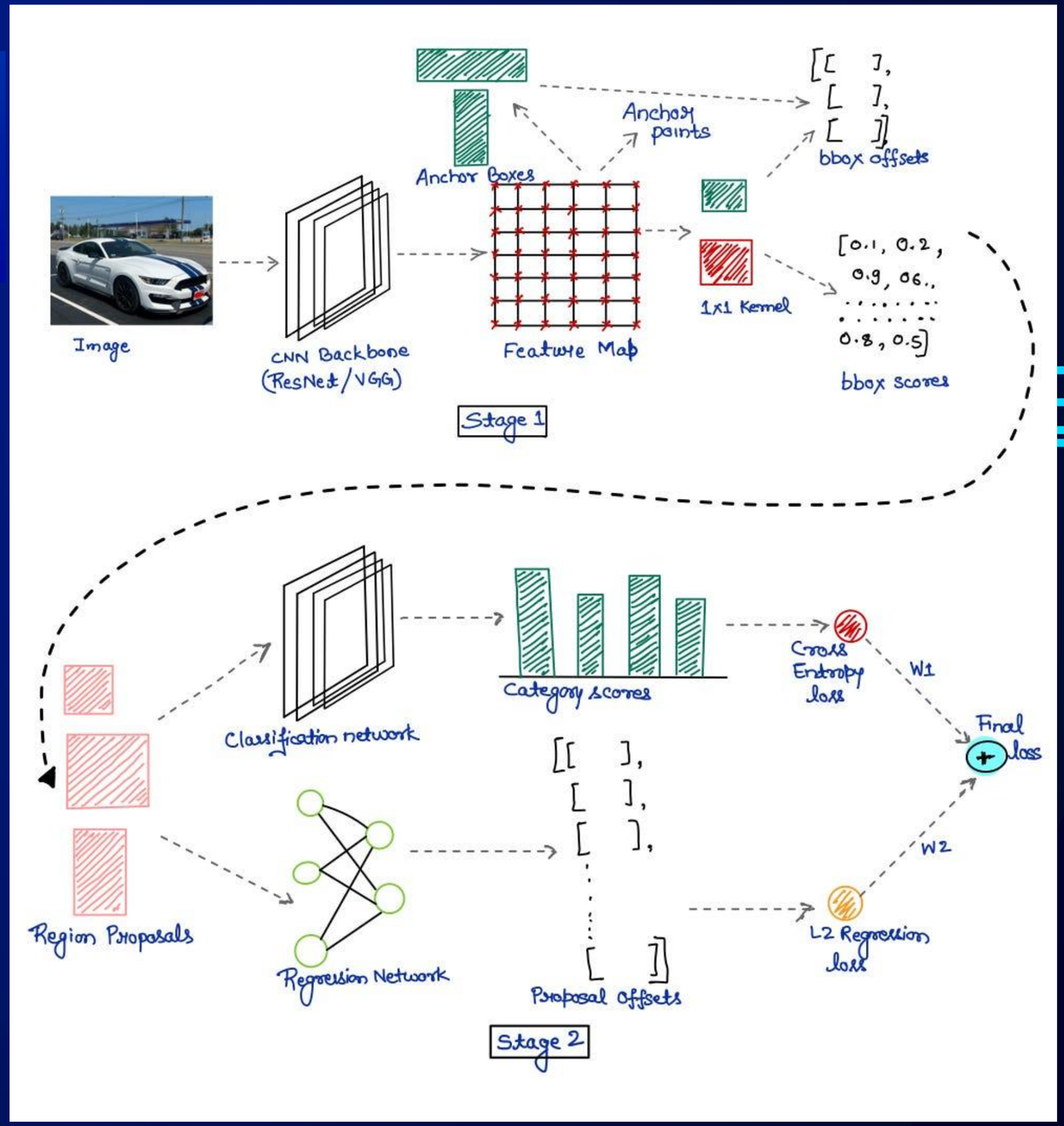
In the second stage all the information from P-Net is used as an input for the next layer of CNN called as R-Net(Refinement Network), a fully connected, complex CNN which rejects a majority of the frames which do not contain faces.

In the third and final stage, a more powerful and complex CNN, known as O-Net(Output Network), which as the name suggests, outputs the facial landmark position detecting a face from the given image/video.

# Faster R-CNN

Most of the current SOTA models are built on top of the groundwork laid by the Faster-RCNN model. Faster R-CNN is an object detection model that identifies objects in an image and draws bounding boxes around them, while also classifying what those objects are. It's a two-stage detector:

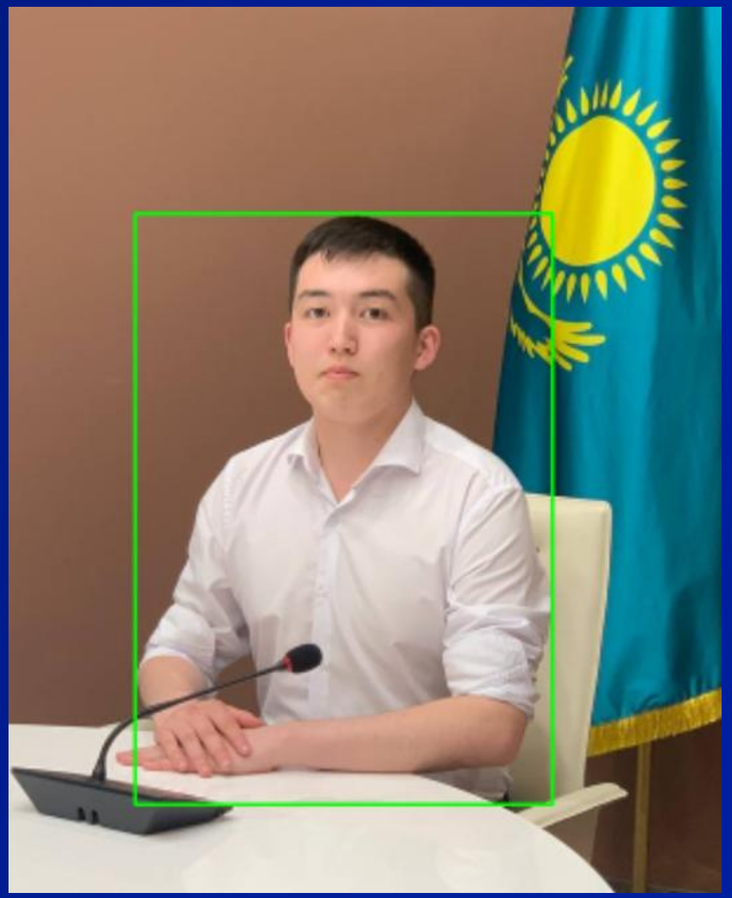
Stage 1: Proposes potential regions in the image that might contain objects. This is handled by the Region Proposal Network (RPN).
Stage 2: Uses these proposed regions to predict the class of the object and refines the bounding box to better match the object.



# Example results

Original

Viola-Jones



Faster R-CNN

https://github.com/Daniyar-manatuly/Pattern-recognation/blob/main/Faster%20R-CNN

# Example results

Original

Viola-Jones

MTCNN

https://github.com/Daniyar-manatuly/Pattern-recognation/blob/main/MTCNN

# Comparison table

| Feature | Viola-Jones | MTCNN | Faster R-CNN |
|---|---|---|---|
| **Speed** | Fast | Moderate | Slow |
| **Accuracy** | Low | High | Very High |
| **Complexity** | Low | Moderate | High |
| **Training Data Requirement** | Low | Moderate | High |
| **Use Cases** | Face detection in low-resource systems | Face detection, landmark localization | Object and face detection in high-accuracy applications |

# Conclusion

Viola-Jones is an old but fast runner in a retro tracksuit, crossing the finish line quickly but with imprecise vision (symbolizing low accuracy).

MTCNN is an advanced athlete balancing speed and accuracy.

Faster R-CNN is a slow but powerful robot with advanced sensors, reaching the finish line with high accuracy.

# References


1. *https://towardsdatascience.com/robust-face-detection-with-mtcnn-400fa81adc2e/*
2. *https://medium.com/dummykoders/face-detection-using-mtcnn-part-1-c35c4ad9c542*
3. *https://medium.com/@RobuRishabh/understanding-and-implementing-faster-r-cnn-248f7b25ff96*
   *https://medium.datadriveninvestor.com/understanding-and-implementing-the-viola-jones-image-classification-algorithm-85621f7fe20b*

*4. Masud, U., Saeed, T., Malaikah, H., Ul Islam, Muhammad F., Abbas, G. Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification. IEEE Access, 2022. DOI: 10.1109/ACCESS.2022.3146320*

# THANK YOU