<https://link.springer.com/article/10.1186/s13640-018-0324-4>

* Facial landmarking has been used as a means of screening sleep apnoea
* Paper talks about current state of image processing and performance
* 1 Definition of the objective: what is the exact nature of the problem to be solved?,
* 2 Selection of an appropriate dataset for solving the defined problem: what information is required to meet the objective?,
* 3 Extraction of regions of interest from the dataset: what features from the selected dataset will best meet the defined objective?,
* 4 Definition of model architecture: which model will give the best performance? and
* 5 Model training and evaluation: what is the best training methodology given all of the above stage?
* Performance is measured by comparing results to ground truth landmarks that have been annotated by human experts (Mechanical Turk)
* Root mean squared error assessment is the simplest, average RMSE over the number of samples
* Percentage of correct landmark points, which is percentage of points in the right spot different from the uclidean distance
* Dataset must be rich enough to train model, uncontrolled and controlled data
* Little consistency between the data sets
* The ground truth must be reliable
* Stage 3 extract the region of interenst, critical for performance
* Two foundational methods of face detection are voila jones and histogram of gradients (HOG)
  + HOG uses pixel gradient structure
* CNN – Cascading neural networks are used. They are trained
* Stage 4 Model definition
  + Generative methods
    - Aim to maximise the progability of facial reconstruction from a deformable model
  + Discriimanative methods
    - Infer face shape by training a regression function that maps image values to facial landmark coordinates
  + Statistical methods
    - Active appearance models build a statistical based representation of the face using the shape information provided by the landmarks and texture information provided by the training images themselves
* Active shape models take the face and scale it to be right
* 300W in the wild came out in 2013, this paper is about work completed since 2013 cause in the wild set a new benchmark for recent papers

For this project I will run 3 algorithms for facial analysis that come with pre trained models

I will test the facial recognition of each model against my own images to see if there exists a bias

I will do this by creating ground truth and seeing how far the calculated coordinates vary from the actual coordinates

I will take a look into the training data for the algorithms to see if there exists a bias

I will also take a look into the actual algorithms themselves to see if its based on contrast (HOG)

Im planning on looking at the facial recognition and landmark recognition parts of the algorithm

OPEN CV uses haar

DLIB uses HOG

<https://mipro-proceedings.com/sites/mipro-proceedings.com/files/upload/sp/sp_008.pdf>

* HAAR Cascade classifier looks at percentages and values of contras
* For demonstration purpose, let's say we are only extracting two features, hence we have only two windows here. The first feature relies on the point that the eye region is darker than the adjacent cheeks and nose region. The second feature focuses on the fact that eyes are kind of darker as compared to the bridge of the nose. Thus, when the feature window moves over the eyes, it will calculate a single value. This value will then be compared to some threshold and if it passes that it will conclude that there is an edge here or some positive feature.

<https://kpzhang93.github.io/MTCNN_face_detection_alignment/paper/spl.pdf>

* The cascade face detector proposed by Viola and Jones [2] utilizes Haar-Like features and AdaBoost to train cascaded classifiers, which achieves good performance with real-time efficiency.
* We propose a new cascaded CNNs based framework for joint face detection and alignment, and carefully design lightweight CNN architecture for real time performance. (2) We propose an effective method to conduct online hard sample mining to improve the performance. (3) Extensive experiments are conducted on challenging benchmarks, to show significant performance improvement of the proposed approach compared to the state-of-the-art techniques in both face detection and face alignment tasks.
* Proposal stage to find bounding box, candiates are calibrated based on the estimated bounding box, stage 2 is the refinement stage where false canidates are rejected, calibration is performed, the final stage O net is used to identify face regions
* ion. However, we notice its performance might be limited by the following facts: (1) Some filters in convolution layers lack diversity that may limit their discriminative ability
* face calsifcationc works with training with proability in comparison to ground truth

Face Recognition Algorithm Bias: Performance Differences on Images of Children and Adults

* However, we know that children look different from adults and they are not simply scaled down versions of adults.
* t. FaceNet was constructed from the MS-Celeb [7] database composed of 10 million images of 100k adult celebrities

<https://arxiv.org/pdf/1701.08289.pdf>

* Viola and Jones milestone used haar like features and adaboost classifier, couldn’t handle non frontal faces and faces in the wild
* Our method follows the similar deep learning framework of Faster RCNN, which has been shown to be a state-of-the-art deep learning scheme for generic object detection [10]. It essentially consists of two parts: (1) a Region Proposal Network (RPN) for generating a list of region proposals which likely contain objects, or called regions of interest (RoIs); and (2) a Fast RCNN network for classifying a region of image into objects (and background) and refining the boundaries of those regions. In this work, we propose to extend the Faster RCNN architecture for face detection, and train our face detection model by following the proposed procedure as shown in Figure 1.
* Trained in the WIDER FACE Dataaset

<https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition>

* Owners of the iPhone X have already been introduced to facial recognition technology. However, Apple's Face ID biometric solution was heavily criticized in China in late 2017 because of its inability to differentiate between individual Chinese faces.

<https://arxiv.org/pdf/2002.02942.pdf>

* Amazon’s face recognition software, Rekognition, despite being easy to use, made an erroneous prediction for 28 members of the Congress and confused them with images of publicly available mugshots. Moreover, even though only 20% of the members of Congress are people of color, almost 40% of the false matches belonged to them (
* These algorithms performed poorly on dark skinned females as compared to lighter skinned males.
* Multi-task Convolutional Neural Network (MTCNN) (Das, Dantcheva, and Bremond 2018) is another framework, which is proposed to learn unbiased feature representations. It jointly learns to predict the gender, ethnicity, and age from the input
* Another research thread to mitigate learning biased representations involves pre-processing the data to obtain fair representations. Amini et al. (2019) presented a pre-processing technique to de-bias face detection
* Feature visualizations further demonstrate an inherent bias in deep learning networks, wherein they appear to focus on race-specific discriminative facial regions. These findings suggest an immediate need for researchers to focus on eliminating bias from face recognition models in order to develop fairer systems

<https://link.springer.com/content/pdf/10.1007/978-3-030-03243-2_798-1.pdf>

* Recent face detection methods typically follow the paradigm of a two-stage detector, e.g., Faster RCNN [34], or a single-stage detector like Single Shot MultiBox Detector (SSD) [35] and You Only Look Once (YOLO) [36]
* d. On the other hand, by going deeper, the spatial information, which is essential to finding tiny faces, would lose through pooling or convolution operations. The aforementioned problem can be partially alleviated by using a dilation operation [37] and reducing the number of pooling operations. However, the computation will dramatically increase with high spatial resolution of feature maps in the network, ma
* The typical loss used for training a deep network-based face detector is 2 loss, measuring the localization error between the coordinates of a bounding box’s corners and the ground truth coordinate

<https://arxiv.org/pdf/1506.01497.pdf>

* The first module is a deep fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector [2] that uses the proposed regions

Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems 35. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Cheng-Yang Fu, Berg AC (2016) SSD: single shot multibox d

<https://arxiv.org/pdf/1506.01497.pdf>

* Region based CNN were computationally expensive. R-CNN provides super quick real time detection using a very deep neural network.
* This speed is not including the time taken for proposals this is to show that the reason algorithms are slow are because the time taken is bottle necked due to proposals

<http://cs231n.stanford.edu/reports/2017/pdfs/222.pdf>

* There is a difference between facial detection and facial recognition. The two step process.
* Viola and Jones used a sliding window and trained a convolutional neural netowkr to detect the precence of a dace in the sliding window area
* Now all mainstream face detection is done using CNNs. This is computationally expensisive with repritive compitiaons of CNN
* To reduce the number of candidate locations for the sliding window, RPN were developed. Region proposal Networks.These were used to propose regions with a high probability that will contain objects
* First R-CNN generates approximately 2000 Regions of Interest (RoI) using the Region Proposal method on the input image, then it warps each RoI into standard input size for the neural network and forward them into the CNNs dedicated for image classification and localization and output the class category as well as the bounding box coordinates and sizes
* More advancements based on R-CNN network occurred to deal with the expensive slow run time problem, such as Fast R-CNN [12] and Faster R-CNN [13]. Fast R-CNN forward the whole image through the CNN at the beginning so that it is only performed once instead of many times in R-CNN [12]; Faster R-CNN performed RoI pooling and make the CNN to do the Region Proposal, which inserts the Region Figure 1. The architecture of the R-CNNs with Region Proposals [9] Proposal Network (RPN) as part of the layers in the CNN model to predict the possibility of objectiveness in the region [13]. While they achieve significant reduce in training time and test time compared with the R-CNN methods [12] [13], the runtime is still dominated by region proposals method.
* YOLO and SSD are two detection methods that don’t use region proposals. The image goes into the covolution network. The image is split into a grid and the attributes are calcilated for each grid (and the classification scores and the bounding box coordinates and scales are determined on each grid cell) The overall object classes and boudnign box is determined from the results from each grid square
* The sliding window is easy to implement but would be slow during training and testing, the accuracy is dependant in the maturity of the network that applies the image classification
* The R-CNN has a redce in training and testing and improves accuracy. Single shot methods are faster than other methods that don’t use region proposals. SSD reduces accuracy a small bit.
* The YOLO algorithm elimates the use of bounding box proposals steps by instead using a single deep neural network to predict both the boundary boxes and the classes at the same time
* It takes images as input, pass it through neural network and get a vector of bounding boxes and class predictions as output
* Yolo V3 uses DarkNet-53 to make features detection followed by convolutional layers. Darknet-53 is a 53 layers Convolutional Neural network trained on ImageNet mainly compose of 3 × 3 and 1× 1 filters with skip connections like the residual network in ResNet.

There are 4 modern algorithms for facial detection. That work quick enough for real time facial detection are Faster R-CNN, YOLO, SSD and FPN.

<https://medium.com/@jonathan_hui/understanding-feature-pyramid-networks-for-object-detection-fpn-45b227b9106c>

* Feature Pyramid Network (FPN) is a feature extractor designed for such pyramid concept with accuracy and speed in mind. It replaces the feature extractor of detectors like Faster R-CNN and generates multiple feature map layers (multi-scale feature maps) with better quality information than the regular feature pyramid for object detection.

<https://www.ajol.info/index.php/njt/article/view/199768>

* Illumination – changing the direction of light on a face changes it shadows across the face which changes the contrast values and pixel densities causing relative changes in highlights and contrast
* Pose – The rotation of a face affects the facial detection algorithm to understand the face and can cause mismatches with other faces in a face database
* Facial expression – Human expressions are very varied and this can cause the computer to mismatch
* Ageing – This affects facial analysis algorithms as many are trained on datasets that don’t incorporate an elderly percentage of people so the models lack accuracy in comparison to that
* Occlusion – The use of glasses, beards and hats are all occlusion to the face and makes it harder for the algorithms to understand a face as its being hidden behind these objects. It could even be the case that a face is being hidden by another face
* Low resolution – This means that less can be done to the original image in terms of convolutions as there is less detail to work with at the start of the image. Less detail also means its harder to track and detect a face.

<https://towardsdatascience.com/faster-r-cnn-for-object-detection-a-technical-summary-474c5b857b46>

* It goes R-CNN then Fast R-CNN then Faster R-CNN.
* R-CNN
  + https://towardsdatascience.com/r-cnn-for-object-detection-a-technical-summary-9e7bfa8a557c
* The first stage of RCNN is region proposals, it uses selective search algoritjm by generating sub segmentations of the image and iteravily combining similar regions to form objects. These are based on colour, texture, size and shape
* At the end a feature vector is created from the region proposals using a CNN
* A SVM is then used for object classification and on the third and final step bounding box regressios is done to improve localization performance. Its used to learn corrections in the bounding box location and size. The output is an accurate bounding box around the object
  + <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>
* Built upon R-CNN to improve spped and efficiency
* Instead of giving region proposals to CNN now we give the input image to generate a convolutional feature map. From this map we identify the regions.
* The reason this is fast is because you don’t have to feed 2000 region proposals into the CNN and instead the operation is done once and a feature map is generated from it
* FASTER R-CNN doesn’t use the selective search algorithm which is slow and instead uses an object detection algorithm that lets the network learn the region priposals
* During the pipeline when selective search would happen to get region proposals, instead now a separate network is used to predict the region proposals.
* This is very fast and can be used for real time detection