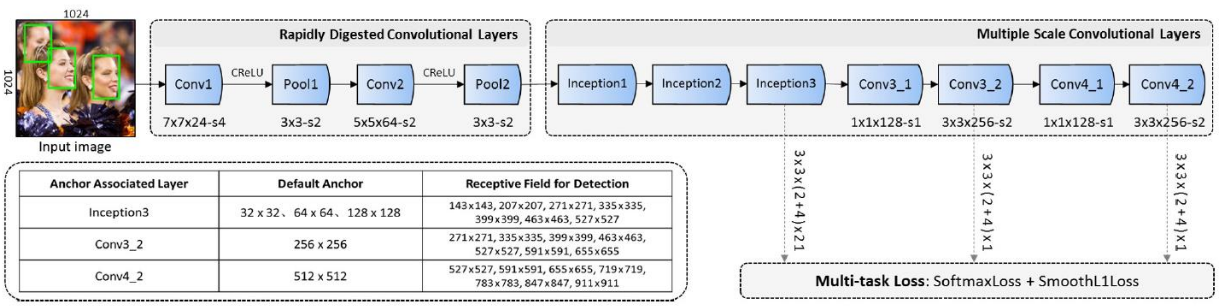
Task.5 FaceBoxes: A CPU Real-time Face Detector with High Accuracy

Group member: Yunjie Xu, Guangda Li, Hao Wang

In this paper, author used rapid detect method to determine face boundary, it called face box.

Figure.1 is architecture of FaceBoxs.



Anchor Associated Layer	Default Anchor	Receptive Field for Detection
Inception3	32 x 32、64 x 64、128 x 128	143x143, 207x207, 271x271, 335x335, 399x399, 463x463, 527x527
Conv3_2	256 x 256	271x271, 335x335, 399x399, 463x463, 527x527, 591x591, 655x655
Conv4_2	512 x 512	527x527, 591x591, 655x655, 719x719, 783x783, 847x847, 911x911

Figure 1. Architecture of the FaceBoxes and the detailed information table about our anchor designs.

three contributions that make the FaceBoxes accurate and efficient on the CPU devices: the Rapidly Digested Convolutional Layers (RDCL), the Multiple Scale Convolutional Layers (MSCL) and the anchor densification strategy. Finally, they introduce the associated training methodology.

RDCL is designed to fast shrink the input spatial size by suitable kernel size with reducing the number of output channels, enabling the FaceBoxes to reach real-time speed on the CPU devices.

Reducing the number of output channels: It utilize the C.ReLU activation function (illustrated in Fig. 2(a)) to reduce the number of output channels.

Fig. 2(b) illustrates our Inception implementation, which is a cost-effective module to capture different scales of faces.

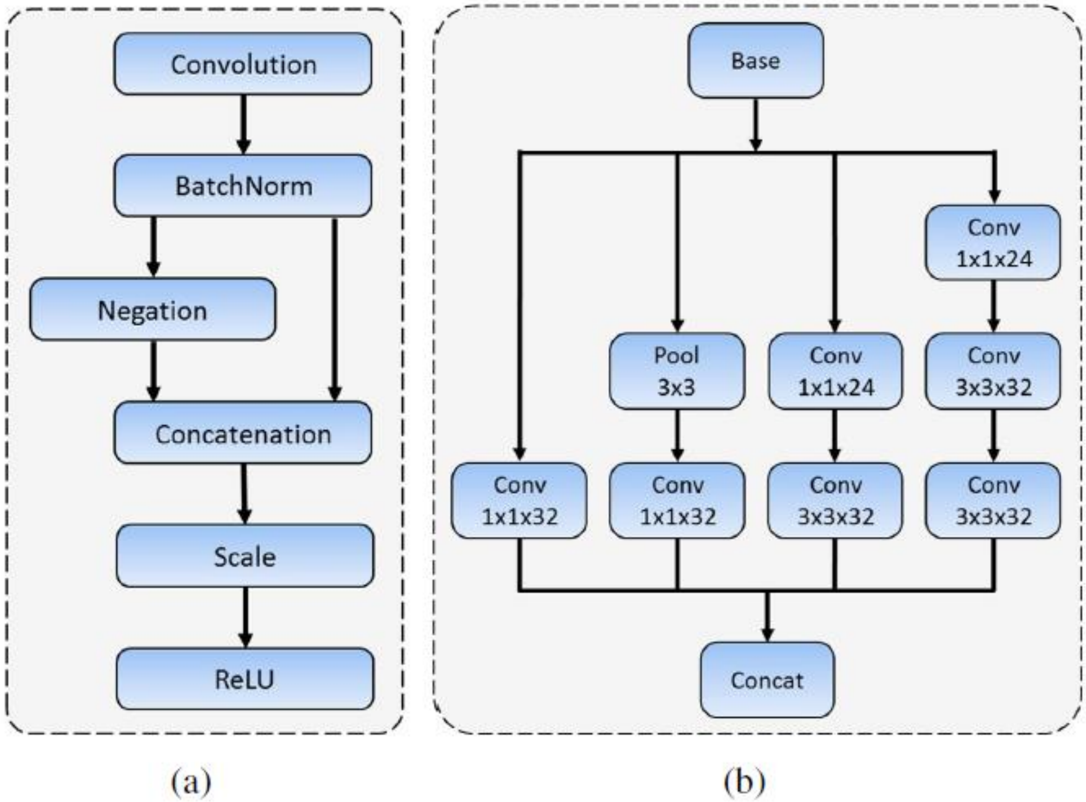


Figure 2. (a) The C.ReLU modules where Negation simply multiplies -1 to the output of Convolution. (b) The Inception modules.

To eliminate this imbalance, author propose a new anchor densification strategy. Specifically, to densify one type of anchors n times, we uniformly tile Anumber = n2 anchors around the center of one receptive field instead of only tiling one at the center of this receptive field to predict. Some examples are shown in Fig. 3. In their paper, to improve the tiling density of the small anchor, our strategy is used to densify the 32_32 anchor 4 times and the 64_64 anchor 2 times, which guarantees that different scales of anchor have the same density (i.e., 4) on the image, so that various scales of faces can match almost the same number of anchors.

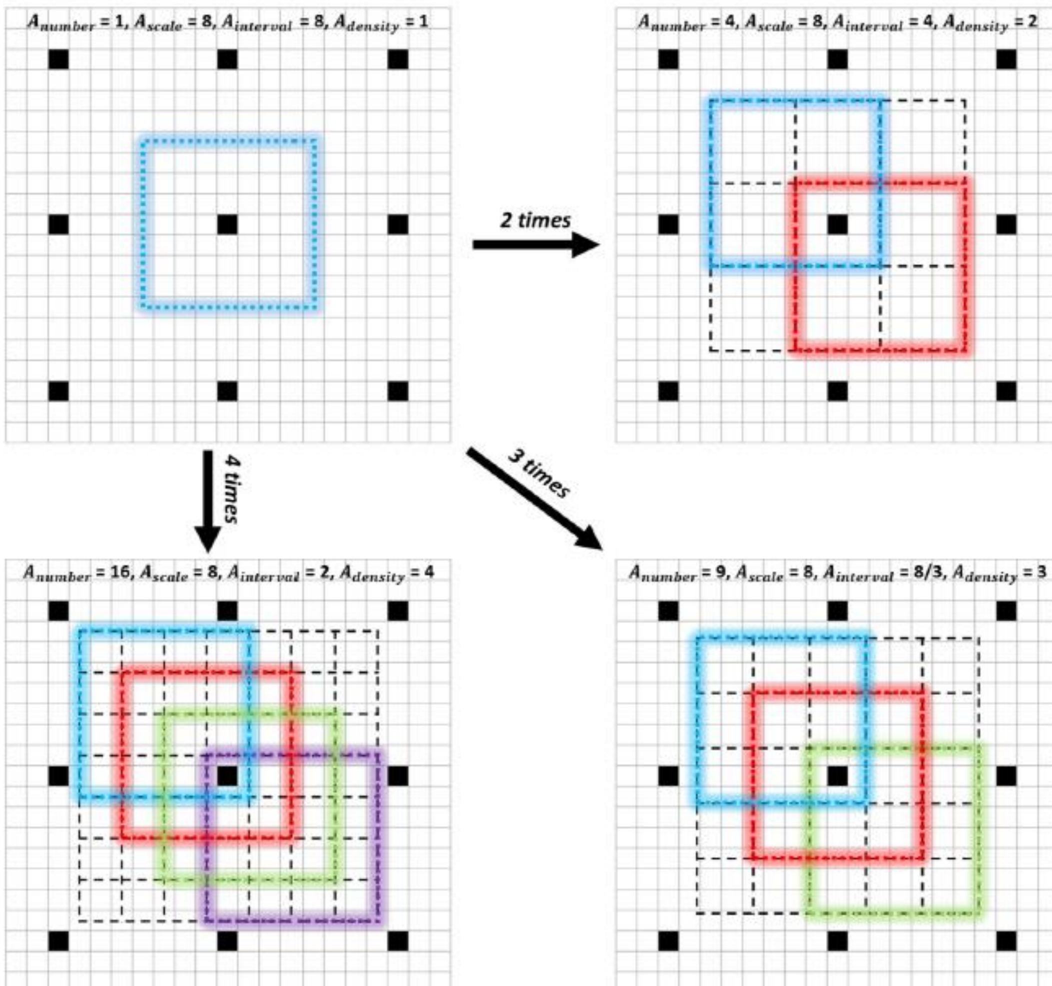


Figure 3. Examples of anchor densification. For clarity, we only densify anchors at one receptive field centre (*i.e.*, the central black cell), and only color the diagonal anchors.

AFW dataset [48]. It has 205 images with 473 faces. They evaluate FaceBoxes against the well-known works [4, 20, 22, 33, 39, 42, 48] and commercial face detectors (e.g.,Face.com, Face++ and Picasa). As illustrated in Fig.4, our FaceBoxes outperforms all others by a large margin. Fig.7(a) shows some qualitative results on the AFW dataset

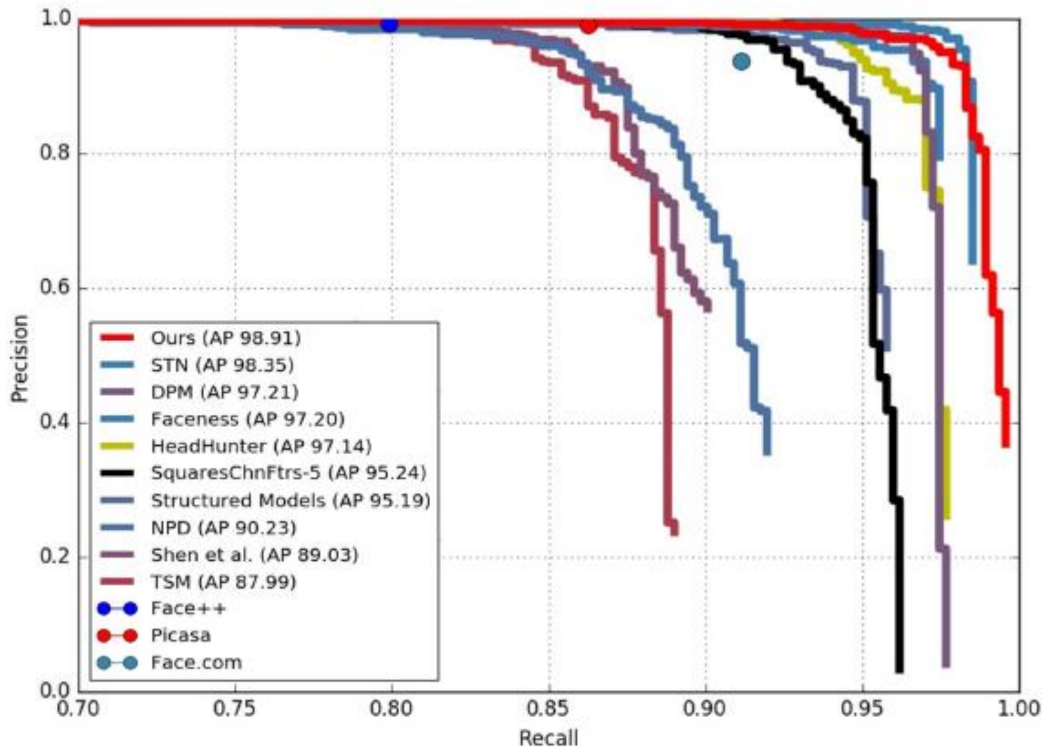


Figure 4. Precision-recall curves on AFW dataset.

PASCAL face dataset [39]. It is collected from the test set of PASCAL person layout dataset, consisting of 1335 faces with large face appearance and pose variations from 851 images. Fig.5 shows the precision-recall curves on this dataset. Their method significantly outperforms all other methods [4, 11, 22, 39, 42, 48] and commercial face detectors (e.g., SkyBiometry, Face++ and Picasa). Fig.7(b) shows some qualitative results on the PASCAL face dataset.

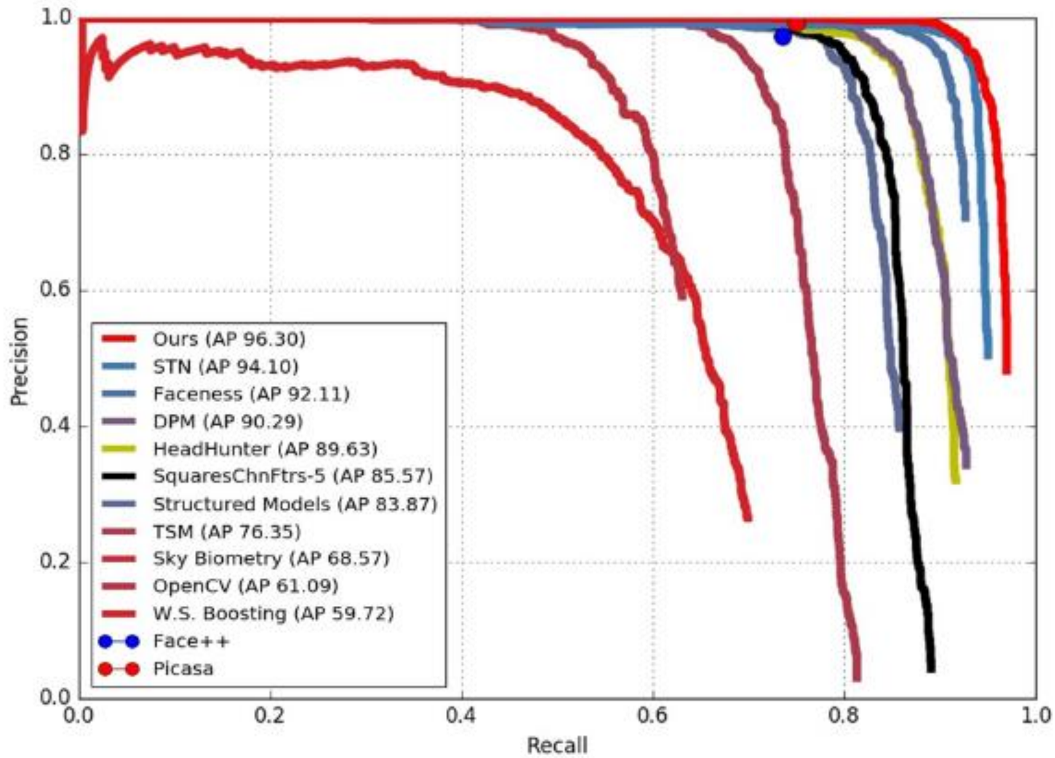
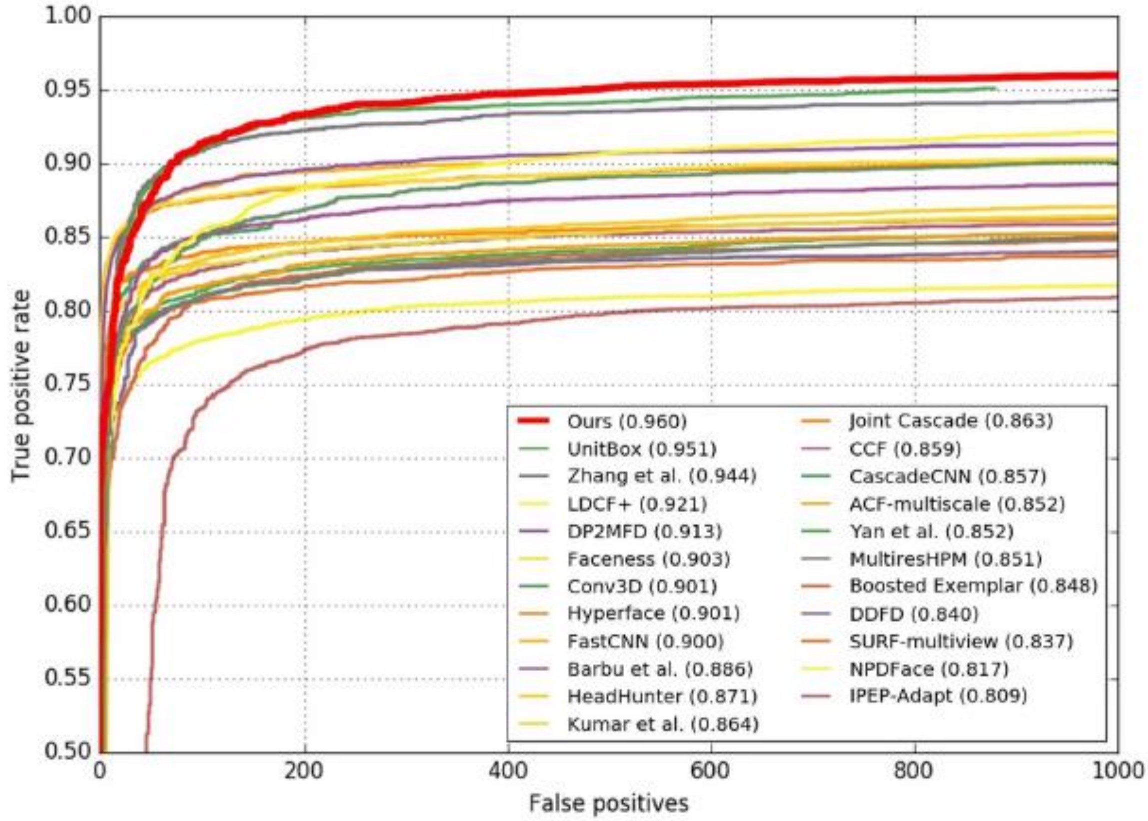
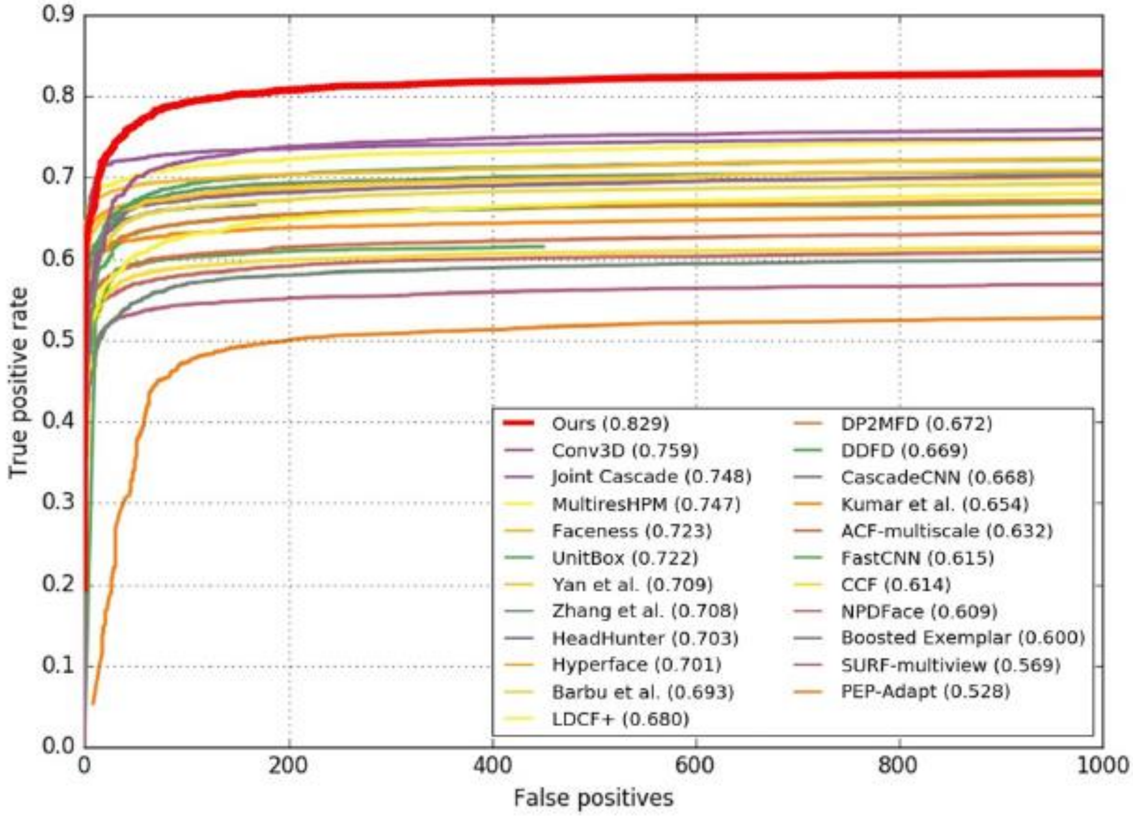


Figure 5. Precision-recall curves on PASCAL face dataset.



(a) Discontinuous ROC curves

FDDB dataset [10]. It has 5; 171 faces in 2; 845 images taken from news articles on Yahoo websites. FDDB adopts the bounding ellipse, while their FaceBoxes outputs rectangle bounding box. This inconsistency has a great impact to the continuous score. For a more fair comparison under the continuous score evaluation, they train an elliptical regressor to transform their predicted bounding boxes to bounding ellipses. they evaluate our face detector on FDDB against the other methods [1, 6, 8, 13, 14, 15, 17, 19, 20, 23, 27, 28, 35, 42, 43, 44]. The results are shown in Fig. 6(a) and Fig.6(b). Their FaceBoxes achieves the state-of-the-art performance and outperforms all others by a large margin on discontinuous and continuous ROC curves. These results indicate that their FaceBoxes can robustly detect unconstrained faces. Fig.7(c) shows some qualitative results on the FDDB.



(b) Continuous ROC curves

Figure 6. Evaluation on the FDDB dataset.

(a) AFW

(b) PASCAL face

(c) FDDB

Figure 7. Qualitative results on face detection benchmark datasets.