A THREE-CATEGORY FACE DETECTOR WITH CONTEXTUAL INFORMATION ON FINDING TINY FACES

Feng Jiang , Jie Zhang , Liping Yan , Yuanqing Xia, Shiguang Shan

Introduction

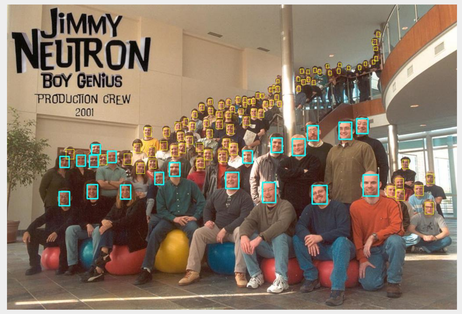

Fig. 1. We run our detector on the image which is randomly selected from WIDER FACE validation set. Faces of different sizes are marked by box. The light blue box represents normal face, while the yellow one represents tiny face.

Great progresses have been achieved on object detection in the wild. However, it still remains a challenging problem due to tiny objects.

In this paper, we present a Three-category Classification Neural Network to find tiny faces under complex environments by leveraging contextual information around faces.

Tiny faces (within 20x20 pixels) are so fuzzy that the facial patterns are not clear or even ambiguous for detection.

Our major contributions are summarized as follows:

- Proposing a three-category face detector, which prevents the normal faces and tiny faces from interacting with each other when they belong to the same category during inference.
- Taking full advantage of the contextual information about tiny faces by taking the extended face annotations as input and designing context module to help proposals contain context.
- Choosing prior anchors by k-means clustering on the training set bounding boxes to predict good detection.

Detecting Tiny Faces

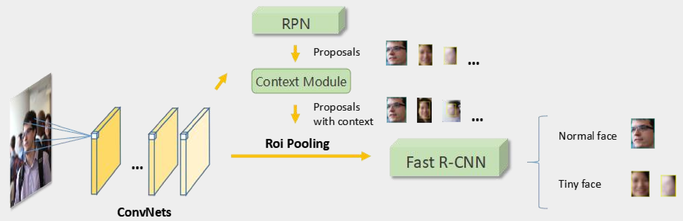

Fig. 2. Architecture of our three-category face detector. Firstly, features are extracted by the basic convolutional networks (ConvNets). Then, the context module processes the proposals generated by RPN and finally Fast R-CNN refines the results.

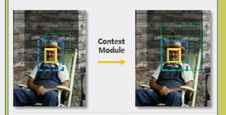

Fig. 3. The yellow boxes represent the actual size of the tiny faces and the blue dotted ones represent extended tiny faces annotations. Proposals with and without context are labelled by green and green dotted boxes respectively. **Context module ensures that anchor-based proposal contains contextual information around the tiny face.**

This section mainly introduces our two-stage face detector, including the anchor-based framework, the three-category classification network, the utilization of contextual information and the anchor dimension clustering.

- **Anchor-based framework:** Fig.2. illustrates the structure, which contains basic convolutional networks (ConvNets), a Region Proposal Network (RPN) module, a context module and a Fast R-CNN module. The detector built on VGG16 can process tiny faces well with small receptive fields (see Fig. 4.).

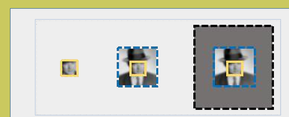

Fig. 4. The yellow boxes represent the actual size of the tiny faces and the blue dotted ones represent extended tiny faces annotations, while the black dotted box represents the receptive field.
Small receptive field fits the tiny face with contextual information well.

- **Three-category classification network:** Face detectors perform well on those faces which contain distinct facial features, however tiny faces are kind of fuzzy. So we divide faces into two categories by size to prevent them from interacting with each other during inference.
- **Leveraging the contextual information:** Human beings have very advanced vision system, for example, if we observe the hat above and the shoulder below, it is likely to exist a face in the middle. We expect our detector to learn such ability and improve the recall rate of tiny faces(see Fig. 3.).
- **Anchor dimension clustering:** Instead of selecting anchors by hand, we run k-means clustering on the training set boxes to find good anchors automatically. We expect prior anchors produced by k-means clustering to tile faces as many as possible, especially tiny ones. So we set the distance metric based on IoU scores:

$$d(\text{box}, \text{anchor}) = \alpha\,(1 - \text{IoU}(\text{box}, \text{anchor}))$$

Here, the annotations of tiny faces are expanded. We choose α = 1 if the box is labelled as normal face, and α = 1.2 for tiny face. Finally, we set up 9 anchors to predict detection.

Experiments

We carry out experiments on two of the most influential face detection benchmarks: the FDDB dataset and the WIDER FACE dataset. The FDDB dataset chooses 2845 images and labels 5171 faces. The WIDER FACE dataset contains 393,703 labelled faces in 32,203 images, which are randomly selected 40%/10%/50% as training, validation and testing sets respectively.

- **Implementation details:** During training of the detector, we initialize the network with the pre-trained model on ImageNet. We improve the scales of anchors, and adopt non-maximum suppression (NMS) with a threshold of 0.7 on proposals produced by the RPN stage. Then the proposals processed by context module are taken as the input of the Fast R-CNN stage. We use softmax loss and smooth L1 loss to guide the training.

	Our detector					
multi-scale training?	✔	✔	✔	✔	✔	✔
three-category classification?		✔	✔	✔	✔	✔
contextual information?			✔	✔	✔	✔
anchor clusters (6 anchors)?				✔		
anchor clusters (12 anchors)?					✔	
anchor clusters (9 anchors)?						✔
AP on Hard subset	0.689	0.707	0.736	0.739	0.745	**0.752**

Tab. 1. shows the performance of our detector on the WIDER FACE test set. Too few anchors will cause a lower recall and too many may bring more false positives. We do not optimize the Faster R-CNN framework and convolutional layers. Still we get a better performance with the AP from 0.707 to 0.752 on the hard subset merely by leveraging contextual information and picking good prior anchors.

- **WIDER FACE:** Faces in the WIDER FACE dataset are extremely challenging due to the large variations in scale. WIDER FACE defines three levels (easy, medium, hard) by different levels of difficulties. In particular, more than half of the faces have a height no more than 20 pixels. We follow the Scenario-Int criterion and train our model on the WIDER FACE training set. In order to demonstrate the effectiveness of our detector, we compare our detector with several other models and the baseline (Faster R-CNN), which is used as the platform to design our detector.

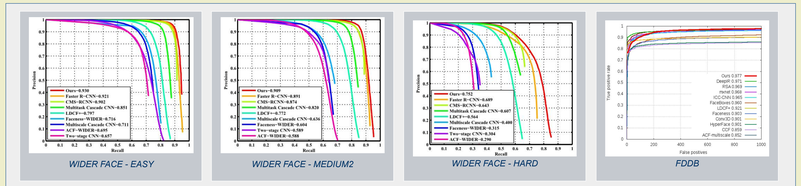

Fig. 4. Evaluation on benchmarks

- **FDDB:** We train the model on WIDER FACE and perform evaluation on the FDDB dataset, and download the evaluation tool from the official and plot the discrete ROC curves. Compared with the existing methods like DeepIR, RSA, ICC-CNN, we obtain a higher true positive rate of 0.977 of the discrete ROC curve at 1000 false positives.
- ***CONCLUSION:*** Through this work, we present a novel face detector and explore the key to the problem of finding tiny faces. We divide face into two categories and we improve the recall rate of tiny faces by leveraging context information around faces and picking good prior anchors. In the future, we will further explore the strategy of classification, namely, divide objects and background into subcategories.