Deep Convolutional Neural Network in Deformable Part Models for Face Detection

Dinh-Luan Nguyen[1(✉)], Vinh-Tiep Nguyen[1], Minh-Triet Tran[1], and Atsuo Yoshitaka[2]

[1] University of Science, Vietnam National University, HCMC, Vietnam
1212223@student.hcmus.edu.vn,
{nvtiep,tmtriet}@fit.hcmus.edu.vn
[2] School of Information Science,
Japan Advanced Institute of Science and Technology, Nomi, Japan
ayoshi@jaist.ac.jp

Abstract. Deformable Part Models and Convolutional Neural Network are state-of-the-art approaches in object detection. While Deformable Part Models makes use of the general structure between parts and root models, Convolutional Neural Network uses all information of input to create meaningful features. These two types of characteristics are necessary for face detection. Inspired by this observation, first, we propose an extension of DPM by adaptively integrating CNN for face detection called DeepFace DPM and propose a new combined model for face representation. Second, a new way of calculating non-maximum suppression is also introduced to boost up detection accuracy. We use Face Detection Data Set and Benchmark to evaluate the merit of our method. Experimental results show that our method surpasses the highest result of existing methods for face detection on the standard dataset with 87.06 % in true positive rate at 1000 number false positive images. Our method sheds a light in face detection which is commonly regarded as a saturated area.

Keywords: Convolutional neural network · Deformable part models · Face detection · Non-maximum suppression

1 Introduction

Face detection is a classical task in computer vision. Although many methods have been proposed to continuously improve the accuracy, such as using single template approach [1], part-based approach [2,3], and even deep convolutional neural network [4–6], face detection is still an interesting and challenging area because of the different appearances of faces in images.

From different approaches of face detection, we find three things commonly taken into consideration to represent a face: parts of a face, spatial relationship between different parts in a face, and the overall structure of a face. Thus, it is necessary to explore efficient methods to represent parts as well as general face information itself for face detection problem. By choosing appropriate methods

T. Bräunl et al. (Eds.): PSIVT 2015, LNCS 9431, pp. 669–681, 2016.
DOI: 10.1007/978-3-319-29451-3_53

to deputize different aspects of a face, it would be possible to further improve accuracy in face detection.

To deal with representing parts and their spatial relationship, Deformable Part Models (DPM), proposed by Felzenszwalb et al. [7], is one of state-of-the-art methods. DPM uses low level feature HOG combined with latent SVM for classification. Furthermore, it also creates a structure model for representing face model. However, because of using low level feature HOG, DPM is not suitable enough to exploit high level feature of an image to represent the overall structure of a face.

On the other hand, convolutional neural network (CNN) is a new trend in many fields of computer vision, which not only shows its superiority in object detection [8] but also in other tasks such as classification [9], segmentation [6], etc. Using deep neural network for face detection is a favorable method since it wisely gets high level feature of an image through its layered structure. Nevertheless, CNN does not provide explicit relationship between lower level features, such as characteristics of parts in a face. Thus, it may lose potential information about candidate relational structure, which is an important information to improve accuracy especially when dealing with face. Both DPM and CNN have advantages and certain limitations in face detection. DPM provides a more flexible representation of a face with deformable parts while CNN generates a high level feature to represent a face. Therefore, it would be a promising approach to integrate CNN and DPM together to synergize their advantages. In this paper, we inherit DeepPyramid DPM [4], an extension for multiclass object detection, as a baseline and then propose novel method based on DPM for dealing with face detection problem.

Besides, in the post processing step, the method of calculating non-maximum suppression in DPM is so unfair that it treats all bounding boxes as the same value. As a result, a region detected with a low score has the same probability to detect a face to a region with higher score, which is one of the main issues for the vanilla DPM. Some improvements [5,10] also propose other ways for choosing the best bounding box but they are still far from satisfied result. Consequently, a new intuitive way to find bounding box is needed for results returned by DPM.

Main Contribution. There are two key ideas in our system. First, we propose a new representation model for face detection together with constructing a new adaptive way of integrated CNN into DPM. Second, an intuitive calculation for non-maximum suppression is also introduced to boost up detection accuracy. We conduct experiments on the standard dataset Face Detection Data Set and Benchmark (FDDB). The results point out that proposed system is significantly superior to other published works on FDDB. Our method achieves up to 87.06 % in true positive rate, being the state-of-the-art technique.

The rest of our paper is organized as follow. Section 2 reviews some related works on the combination between DPM with CNN and other improvements in face detection using DPM. Our primary contribution for proposing new face model architecture and intuitive non-maximum suppression are carefully discussed in Sect. 3 and Sect. 4 respectively. Section 5 shows experimental results and compar-

ison to other state-of-the-art techniques on FDDB dataset. Finally, conclusion is given in Sect. 6.

2 Related Works

In object detection, there are two main approaches [11]: rigid and part-based methods. In rigid approach, a model captures the whole object and exploits characteristics by using single detection and abstract feature. Based on this idea, some recent works use convolutional neural network for mining high level features and applying to face detection [5,12]. Among them, by achieving competitive result on FDDB dataset, DDFD - an extension of R-CNN [6], proposed by Farfade *et al.* [13], is one of promising approaches for using CNN in object detection. Besides, Chen *et al.* [1] proposes a boost cascade technique with shape index feature to align face and conduct detection. Park *et al.* [14] and Zhang *et al.* [15] use multi-resolution technique to overcome different scales of face. These approaches, however, have not reached top performance since a rigid based method is not flexible enough to deal with deformable objects, such as a face.

On the other hand, a part-based approach can handle multiple appearances of an object. It captures the patterns of each part and combine them together to get final detection result. Derived from this approach, a tree structured model proposed by Zhu *et al.* [16] achieves both facial landmarks localization and pose estimation in real time. Pirsiavash and Ramanan [17] create steerable part models to solve different view points of face. Besides, Deformable Part Models (DPM), proposed by Felzenszwalb *et al.* [7], is one of pioneers in face detection using part-based structure. DPM takes advantage of HOG low level feature as an input for finding root and part models. A root model is used for representing the whole object while a part model which is twice resolution accounts for a changeable objectś component. To find the location of a part model, DPM uses a sliding window combined with latent SVM to classify regions. A pyramid image is constructed based on different scales of an image. An extension of Deformable Part Models proposed by Mathias *et al.* [2] gets the promising result by pre-training carefully. However, applying low level features for learning is so wasteful that it eliminates much useful undiscovered information. Therefore, there is a huge need to replace HOG by another high level feature extracted from input images.

There are just a few works realizing the complementary between DPM and CNN. Work of Ouyang and Wang [12] creates CNN whose inputs are HOG features. This CNN structure also has a deformation layer to deal with occlusion situations. However, this work just focuses on optimizing pedestrian detection. Savalle *et al.* [8] use deep features extracted from pyramid images instead of using HOG features. This approach gets promising results but the structure for learning features from pyramid images only has five convolutional layers with fine-tuned parameters. Wan *et al.* [10] use pixel-wise max to form corresponding map from root and nine part filters acquired from three views of an object template. However, this extension of pyramid feature is not adaptive because

it fixes the model with nine parts and uses hand-crafted step to split three object templates. Work of Girshick [4] integrates DPM-CNN structure based on features pyramid returned by [8]. To be specific, each pyramid level is convolved with root and part filters to get a convolution map. These maps are processed with a distance transform pooling layer then stacked together to convolve with a spare object geometry filter. Thus, the output of this network is a single channel score map for DPM component. Our method inherits the version 5 of vanilla DPM [7] and DeepPyramid DPM (DP-DPM) [4]. We complement their work by specifying the neural network structure to get it specialized to face detection with raw DPM version.

One of the important parts of a detection model which affects the final result is the post processing step. Non-maximum suppression has been discussed by many works [10,17] and non-maximum suppression is tweaked to fit with the output of each method. In the original DPM and other improvements [2,18,19], non-maximum suppression is usually performed by exploiting the overlapped area of each pair of bounding boxes to select the best one. Thus, this approach does not cover all bounding boxes, especially when dealing with situations in which the boxes are spare and scattered in an image. Besides, Wan *et al.* [10] create a ranking loss in their network to keep track of promising returned boxes. However, all discussed methods are either too simple [4,7] or complicated [10] and each of them just sticks to a specific model structure. Thus, a general method for adaptively covering all kind of models is necessary to be proposed.

3 Deep Face Deformable Part Models

In this section, we present our new effective face depiction architecture and an integrated convolutional neural network in DPM called DeepFace DPM.

3.1 New Face Representation Model

We review the object model in vanilla DPM [7] and then propose our new model to enhance the original one. DPM uses HOG features to create root and part scores. HOG is calculated by using a pyramid of different scale images and convolution kernel to get gradient value. Different bins of orientation are accumulated by their corresponding size based on gradient orientations.

Part and root filters are constructed from HOG features. The default configuration of DPM having 8 part filters with the fixed size of 6×6 pixels is just a general solution for multiclass detection. In practical use, the accuracy in face detection is affected by the variance of illumination, face's pose direction, occlusion and blur condition. Therefore, from our observation of faces in frontal and side views, we propose a new adaptive model to represent a face which is derived from 4-part model and 5-part model.

To deal with frontal face when the lighting condition is nearly stable, 5 parts are enough for representing 1 forehead, 2 eyes, 1 nose, and 1 mouth. Because of

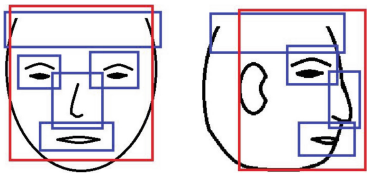

Fig. 1. New integrated model for face representation. 5-part model (left) and 4-part model (right) are used to detect 0° to 45° and 45° to 90° face direction comparing to frontal face respectively.

the vivid forehead, part filter corresponding to it has twice resolution in comparison with the others. Similarly, 4-part model representing 1 forehead, 1 eye, 1 nose, and 1 mouth is introduced to overcome the difficulties of occlusion or changeable illumination face. Figure 1 describes root and part filters in frontal and occluded circumstances. Decision to choose either a 4-part or a 5-part model depends on proposed DeepFace DPM network which is described in details in Fig. 2. The model score for representing face is the output of DeepFace DPM network described in Sect. 3.2. The reason for proposing this new face model is from the observation that when a face is occluded or not in frontal view, we can only see many but not all face components. Thus, using a model with small number of parts which corresponds to occluded situations is sufficient in comparison with the big one.

3.2 DeepFace DPM - A Convolutional Neural Network Integrated in DPM

Extract Coarse Convolutional Feature Pyramid. Given an input image, we scale it up and down into D-scale levels where the original size is at the level $\lfloor D/2 \rfloor$. Since the size of a face is unknown, a feature pyramid is used to deal with different scales in images. We inherit the structure of SuperVision CNN [9] to extract coarse pyramid features. However, we just use 4 layers and eliminate max pooling step at the $4th$ layer to reduce complicated calculations. Thus, the output of this SuperVision CNN process is a coarse pyramid feature as the input for the following 4 or 5-part DPM-CNN architecture.

Integrated 4-5 Part DPM-CNN. Based on the superiority of DPM-CNN architecture [4], we get rid of calculating stack maps process and use the specific 4 part filters per one root filter. Consequently, the component score at each layer is the pyramid distance transform of part convolution. These pyramids are the input for full DPM-CNN with the number of part filters is 5. A max pooling layer is constructed to get the highest correspondent score of model. This score is used as a replacement for the hand-crafted score between root and part filters in the original version of DPM for latent SVM classification afterward.

There are two points in the SuperVision CNN architecture that we solve for face detecting problem. The first thing is that SuperVision network itself is

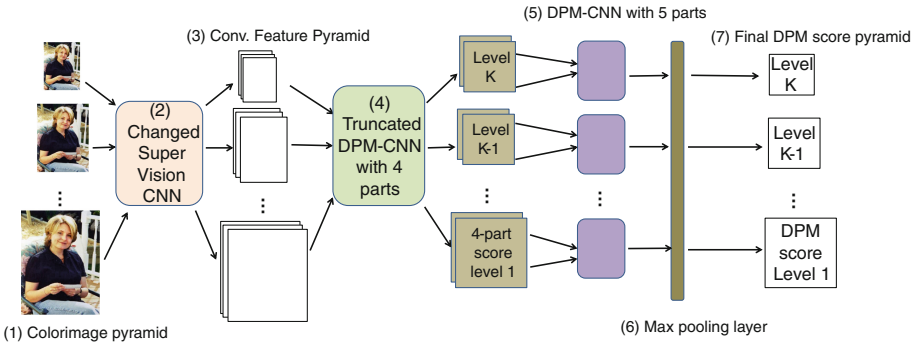

Fig. 2. Proposed model architecture. (1) Color images pyramid is built by resizing an input image with scaling factor 1.5. (2) SuperVision CNN [9] is used to extract feature from an image pyramid. (3) Feature pyramid are constructed from the 4th layer after forward propagation. (4) Convolutional feature pyramid is the input for DPM-CNN [4], which is truncated stack maps process. (5) Each 4-part component feature level goes through full DPM-CNN to get 5-part DMP-CNN feature. (6) Max pooling layer is used for calculating the most promising score result returned by DMP-CNN at each level.

used for classification and detection generic objects. As a consequence, it is not optimized to be used of face detection, which is only focus on round rigid areas. Based on this observation, we scale down 224×224 patches in data augmented process to the size of 112×112. Thus, the input of our network has the size of $112 \times 112 \times 3$. Furthermore, we reduce 1 stride after going through each layer to accumulate more precious high level features. To be specific, the first layer has the stride of 4 pixels while the second, third, and fourth layers use the stride length of 3, 2, and 1 pixels respectively. This way of adjustment means that the more meticulous extracting feature after each layer is, the higher level and important characteristics we get. The second problem with SupperVision CNN is that its output is at $1/16th$ the spatial resolution of the corresponding input. This method for using feature is so deficient that it eliminates any bounding box that has a small size within 16×16 pixels. We completely solve this defect by upscale features at twice resolution in each layer. Combining these solutions together with applying dropout layer not only significantly increases the speed for training but also improves the quality of output features.

4 Intuitive Non-maximum Suppression

In the original version of DPM [7] and other extensions [10,17,18], including DeepPyramidDPM [4] and FaceCascadeCNN [5], Intersection-over-Union exemplar is commonly used to eliminate redundant bounding boxes. To be specific, let B represent a big box and b is a small one. The old traditional method calculates the overlapped region between $S(B) \cap S(b)$ and compares it to the area of

Algorithm 1. Intuitive Non-maximum suppression

Input: $B = \{b_1, b_2, \ldots, b_K\}$
w, h: width, height of input image
Output: $B' = \{b'_1, b'_2, \ldots, b'_N\}$
procedure INTUITIVE NMS
  $C = \{c_1, c_2, \ldots, c_N\} \leftarrow MeanShift(B)$
  $Tag(b_i) \in L = \{l_1, l_2, \ldots, l_N\}$
  $A = 0_{h,w}$
  $S_{Bmin,L} = +\infty$
  for $b_i \in B$ **do**
    $A = A + M_{score}(b_i)$
    $S_{Bmin,Tag(b_i)} = min(S_{Bmin,Tag(b_i)}, area(b_i))$
  end for
  $C = local\_maximum(A)$
  for $c_i \in C$ **do**
    $b'_i = expand(c_i, S_{Bmin,l_i})$
  end for
end procedure

the smaller box b. A hard threshold is used to suppress any bounding box that does not satisfy the following constraint:

$$\frac{S(B) \cap S(b)}{S(b)} \geq 50\,\% \tag{1}$$

Besides, Wan *et al.* [10] proposes an extension to eliminate unnecessary boxes by splitting the condition into two situations depending on whether the bounding boxes are in the same type or not. For different detected object boxes, the criterion is based on

$$\frac{S(B) \cap S(B')}{S(B) \cup S(B')} \geq 75\,\% \tag{2}$$

where B' is the candidate box of another object. For the same object boxes, the overlap is calculated by

$$\max\left(\frac{S(B) \cap S(b)}{S(B)}, \frac{S(B) \cap S(b)}{S(b)}\right) \geq 50\,\% \tag{3}$$

These approaches are insufficient since they discard uncommon region between two boxes and treat low score bounding boxes as the same as the big ones. Thus, they may lead to incorrect detect results if candidate boxes are sparse in an image. From these observations, we propose a new intuitive way of calculating bounding boxes described in Algorithm 1 to solve these defects. Given K bounding boxes (B) returned from framework, we classify them into N clusters (C) using MeanShift. Zero matrix A is created with size $h \times w$ to accumulate matrix score area of each bounding box (M_{score}). Besides, minimum size box of each cluster is collected to build the final box (B') from the new cluster center point generated by calculating local maximum over matrix A.

5 Experimental Results

Dataset. We evaluate the merit of our method on Face Detection Data Set and Benchmark (FDDB) [20]. This large scaled dataset contains 2845 images comprising of 5171 faces gathered from news photographs and has wide variety of background, appearance, illumination, and face direction. FDDB uses ellipse coordinator as face annotations. The result of some state-of-the-art techniques are public on FDDB website. Figure 3 shows some FDDB images with their ellipse annotation.

Fig. 3. Some examples and annotation in FDDB dataset. Faces are annotated by using ellipse and cover wide range of size, illumination, looking direction, and occlusion.

To be fair with other methods, we build an upright ellipse for each detected rectangle. In specific, given an output rectangle in size (w, h), we create an ellipse having the same center point of the rectangle and the sizes of the major axis and minor axis of the ellipse are $1.21h$ and $1.11w$ respectively. By adjusting our result for easy evaluation with FDDB dataset, we slightly improve the true positive in overall (from 86.88 % to 87.06 %). The advantage of changing detect region from rectangles to ellipses is described in Table 1.

Evaluation. We use standard evaluation protocol provided with dataset so as to be equitable when comparing with other techniques. There are two kinds of evaluation: continuous and discontinuous one. In the continuous evaluation, it reveals the robust of framework after 10 folds validation by using matching metric of Intersection-over-Union. Meanwhile discontinuous shows the number of false positive and true positive rate. We run our network configuration described in Sect. 3.2 with $D = 15$ scale levels. Table 1 illustrates the results of different DeepFace DPM's configurations. The DeepPyramid DPM with the default configuration using 8 parts is useful for detecting generic object but it does not

Table 1. Comparision between different configurations in FDDB dataset

Configuration	True positive rate at 1000 false positive images
HOG-DPM [7]	65.70 %
HOG-DPM with intuitive NMS	69.86 %
HOG-DPM with 4–5 part model	78.73 %
DeepPyramid DPM [4]	81.29 %
Our method using default DPM object model	82.95 %
Our method w/o using intuitive NMS	84.60 %
Our method with rectangle evaluation	86.88 %
Our best method	**87.06 %**

demonstrate the superiority in face detection. Our integrated DeepFace DPM model points out the advantages with 87.06 % true positive rate at 1000 positive false images while the DeepPyramid DPM only gets 81.29 % in true positive rate. Furthermore, we also compare our system with the HOG-DPM vanilla and others improvements.

From Table 1, using pyramid image scales as raw convolutional features combined with adaptive 4–5 part model for face representation significantly boosts up the accuracy detection. To be specific, HOG-DPM with default configuration

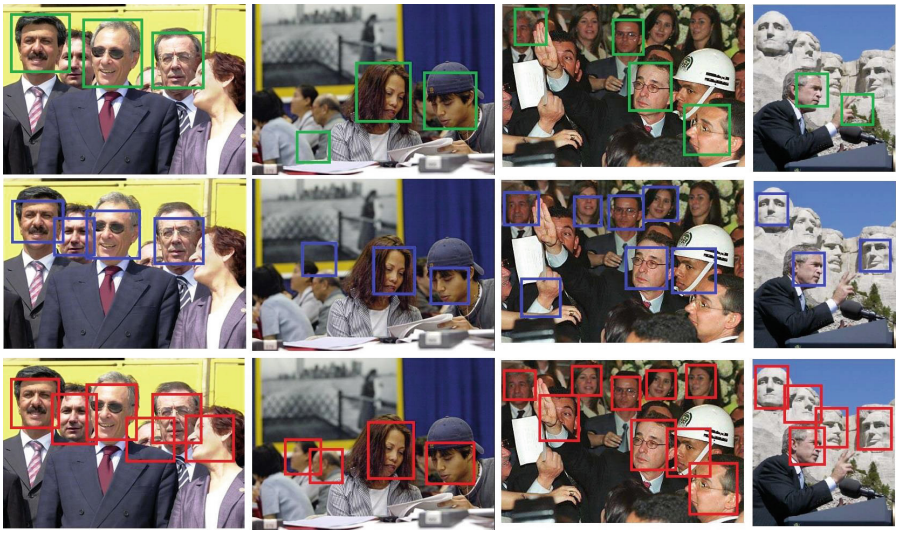

Fig. 4. Selected situations which proposed method shows superiority to DPM and CNN. First row: results detected by DPM. Second row: results detected by CNN (DeepPyramid DPM). Third row: results detected by our method.

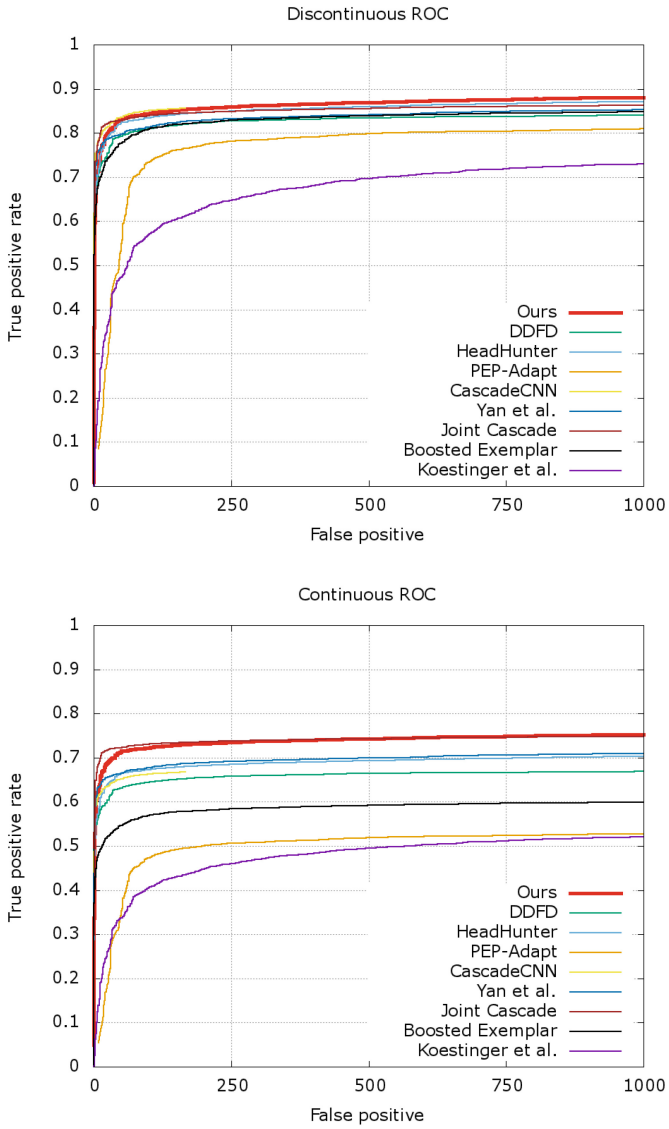

Fig. 5. Comparision with state-of-the-art on FDDB dataset. We compare our result with state-of-the-art methods comprising DDFD [13], HeadHunter [2], PEP-Adapt [3], CasacadeCNN [5], Yan *et al.* [19], Joint Cascade [1], Boosted Exemplar [18], and Koestinger *et al.* [21] (Color figure online).

only get 65.70 % in true positive rate whereas 4–5 part model integrated into HOG-DPM boosts the precision up to 78.73 %. Besides, the method of using high level pyramid features instead of HOG low level features impressively increases 21.36 % (from 65.70 % to 87.06 %) in true positive rate. By using proposed intuitive

non-maximum suppression, we avoid a lot of redundant bounding boxes and get the right position for candidate region. HOG-DPM with intuitive non-maximum suppression improves up to 4.16 % (from 65.70 % to 69.86 %) while our system accelerates 2.46 % (from 84.60 % to 87.06 %) in detecting result. Figure 4 shows some difficult situations including different face's pose, direction, illumination, and even stone's face. Our method successfully detects all faces while DPM and CNN miss and have wrong detect in some images.

Compare with State-of-the-Art Techniques. To be equal when comparing our result with other works, we use public results on FDDB website for reference. Our system shows the superiority not only in continuous but also discontinuous score. Figure 5 describes the comparison between our achievement with current state-of-the-art techniques comprising DDFD [13], HeadHunter [2], PEP-Adapt [3], CasacadeCNN [5], Yan *et al.* [19], Joint Cascade [1], Boosted Exemplar [18], and Koestinger *et al.* [21]. Our result gets 87.06 % (discontinuous ROC) and 75.28 % (continuous ROC) in positive rate at 1000 false positive image while the best result of state-of-the-art only achieves 86.13 % and 74.83 % respectively.

Figure 6 shows the comparison between proposed method with traditional NMS. By using our method, the system significantly increases true positive detected bounding boxes. Especially in images having many people, intuitive non-maximum suppression shows the superiority by successful detecting face with different sizes, looking directions, blur condition, and part occlusion. However, a few missing boxes can occur (the red boxes) when comparing with groundtruth because of these faces are nearly occluded and not easily to detect. There are two unsuccessful cases in our framework which are too small blur faces and half occluded ones. Firstly, in the left-most image (small blur faces), face's size so tiny that features for parts and structure between them is not vivid. Hence, it is difficult to exploit features from these faces. However, our framework just misses some situations where faces are too small and nearly occluded by other objects (e.g. racket, image's border). Secondly, in the right-most image, the missed face is occluded by front people. Thus, we nearly just have half information of frontal face. In some circumstances, situation liked this is treated as side-view face. However, in this image, feature may not be enough to be classified either frontal or side-view face.

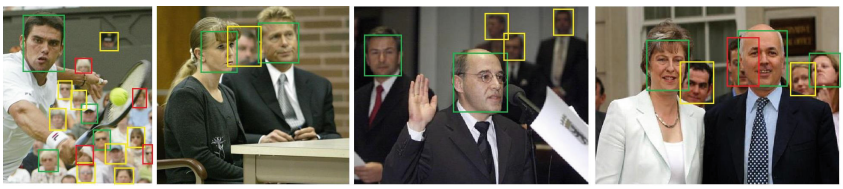

Fig. 6. Examples of applying intuitive non-maximum suppression method with difficult situations. Green boxes: results detected by traditional NMS, yellow boxes: extra results detected by our method besides green ones, red boxes: missing boxes in comparision with groundtruth (Color figure online).

6 Conclusion

In this paper, two novel techniques are proposed to boost up the capacity of DPM and CNN. Our system reveals the fact that structure learning and deep learning can be integrated together to get the top performance. Besides, new combination of 4–5 part model and intuitive non-maximum suppression significantly increases the accuracy of face detection. The evaluation results show that proposed system is robust and achieves competitive performance in comparison with other state-of-the-arts. Furthermore, it becomes new state-of-the-art on FDDB dataset. This work sheds a light on face detection approach and has potential for practical using in the future.

References

1. Chen, D., Ren, S., Sun, J., Wei, Y., Cao, X.: Joint cascade face detection and alignment. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VI. LNCS, vol. 8694, pp. 109–122. Springer, Heidelberg (2014)
2. Mathias, M., Benenson, R., Pedersoli, M., Van Gool, L.: Face detection without bells and whistles. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part IV. LNCS, vol. 8692, pp. 720–735. Springer, Heidelberg (2014)
3. Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic part model for unsupervised face detector adaptation. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 793–800. IEEE (2013)
4. Girshick, R., Iandola, F., Darrell, T., Malik, J.: Deformable part models are convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
5. Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G.: A convolutional neural network cascade for face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5325–5334. (2015)
6. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
7. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1627–1645 (2010)
8. Savalle, P.A., Tsogkas, S., Papandreou, G., Kokkinos, I.: Deformable part models with cnn features. In: European Conference on Computer Vision, Parts and Attributes Workshop (2014)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105. (2012)
10. Wan, L., Eigen, D., Fergus, R.: End-to-end integration of a convolutional network, deformable parts model and non-maximum suppression. CoRR abs/1411.5309 (2014)
11. Cho, H., Rybski, P.E., Zhang, W.: Vision-based 3d bicycle tracking using deformable part model and interacting multiple model filter. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 4391–4398. IEEE (2011)

12. Ouyang, W., Wang, X.: Joint deep learning for pedestrian detection. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 2056–2063. IEEE (2013)
13. Farfade, S.S., Saberian, M., Li, L.J.: Multi-view face detection using deep convolutional neural networks. arXiv preprint arXiv:1502.02766 (2015)
14. Park, D., Ramanan, D., Fowlkes, C.: Multiresolution models for object detection. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 241–254. Springer, Heidelberg (2010)
15. Zhang, W., Zelinsky, G., Samara, D.: Real-time accurate object detection using multiple resolutions. In: IEEE 11th International Conference on Computer Vision, 2007, ICCV 2007, pp. 1–8. IEEE (2007)
16. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2879–2886. IEEE (2012)
17. Pirsiavash, H., Ramanan, D.: Steerable part models. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3226–3233. IEEE (2012)
18. Li, H., Lin, Z., Brandt, J., Shen, X., Hua, G.: Efficient boosted exemplar-based face detection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1843–1850. IEEE (2014)
19. Yan, J., Lei, Z., Wen, L., Li, S.Z.: The fastest deformable part model for object detection. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2497–2504. IEEE (2014)
20. Jain, V., Learned-Miller, E.G.: Fddb: a benchmark for face detection in unconstrained settings. UMass Amherst Technical report (2010)
21. Kostinger, M., Wohlhart, P., Roth, P.M., Bischof, H. : Robust face detection by simple means. In: DAGM 2012 CVAW Workshop (2012)