# Interpretation of Occluded Face Detection Using Convolutional Neural Network

Huaer Li*, Sharifa Alghowinem*†, Sabrina Caldwell*, Tom Gedeon*
*Research School of Computer Science, Australian National University, Canberra, Australia
†College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia
{u6364576, sharifa.alghowinem, sabrina.caldwell, tom.gedeon}@anu.edu.au

***Abstract*—With the rapid development of artificial intelligence in past decades, great attention has been drawn to the field of face detection and recognition. Humans show a high degree of variability in their expressions, poses and appearance. Thus, limitations such as disguised and occluded faces, make it hard to implement high-accuracy face detection in real life. Although several algorithms have been proposed to handle recognition of disguised faces, the interpretation of the possible features that may have impact on the performance of models is hardly mentioned. In this paper, we explored possible features that could distinguish disguised faces compared to original faces, including skin region, luminosity, textures and edges. A Deep Neural Network model was utilised for the comparison between different features using the Disguised Faces in the Wild dataset. Our results show that colour on skin region, luminosity and texture in images could greatly contribute to the performance in detection of disguised faces with a CNN architecture. The results from fusing the individual features significantly outperformed the results when using the whole image, performing 72%, which is considered the state-of-the-art in subject-independent disguised face detection.**

***Index Terms*—disguised face recognition, occluded face detection, deep learning interpretation, fusion.**

## I. Introduction

Face detection is one of the most crucial steps in the process of face recognition as it ensures that computation resources are focused on the region with human faces. However, detecting faces in unconstrained environments still remains a challenging task because of various factors, such as disguised and occluded faces. These factors would significantly degrade the accuracy of state-of-art face detection algorithms and hence affect the accuracy of the face recognition process [8], [9].

While some other covariates of face detection have received great attention, disguised and occluded face recognition is still a young field of research. Little exploration has been made to interpret the features that distinguish them in a way that could be understood by humans [7], [10], [12]. In other words, there is no clear understanding of the factors that influence the current classification results, especially when deep neural networks are utilised.

In this paper, this issue is addressed as several possible features are explored in classifying disguised and original faces. Three different convolutional neural networks have been implemented to test on four features: skin regions, luminosity on different surfaces, luminosity around edges and colours around edges. These features could provide us with some understanding of which feature plays a crucial role in disguised face recognition using convolutional neural networks.

## II. Background and related work

Face detection has been extensively studied in the past decades. The first milestone work in face detection was proposed by Viola and Jones, where rectangular Haar-like features were used to achieve a reasonable result in real time. Since then, some improvements, using other features, such as a histogram of oriented gradients (HOG) and a Normalised Pixel Difference (NPD) have been proposed to enhance the face detection performance on non-frontal faces [1], [2].

Another general approach to boosting the accuracy of face detection is using deep learning. Jiang et al. [3] has proposed a method of using a generic object detector Faster Region-based Convolutional Network method (Faster RCNN) in face detection; after retraining on the WIDER face dataset, it gave an impressive accuracy on the Face Detection Data set and Benchmark (FDDB), one of the benchmarks of face detection [4], [5]. Further improvement on the Faster R-CNN architecture has also been done; Sun et al. combined Faster RCNN with strategies like feature concatenation, hard negative mining and multi-scale training to achieve one of the state-of-art results [6].

Although the state-of-art accuracy of face recognition and detection could surpass human recognition rate, these models were mainly trained and tested on images with upfront or normal faces. Face recognition and detection in disguised faces still remains a great challenge in the field. There are several datasets constructed for disguised and occluded faces, such as COFW and Disguised and Makeup Faces Database, yet these datasets only covered some forms of disguise or occlusion [7], [8]. The DFW dataset covers various modalities of disguise and set a challenge for state of art face detection and recognition algorithms [9] (see examples in Fig. 1).

Peri et al. [10] suggested a Siamese VGG-Face architecture, retrained on the DFW dataset, to detect disguise, and improved the performance by 27.13% from a standard VGG-Face architecture. Another two-phase deep neural network architecture MiRA-Face was proposed and achieved 75.08% at 0.1% false accept rate (FAR) and 89.04% at 1% FAR on the overall performance of the DFW dataset [16]. Classification of disguise type was also an area of interest as it could be used to enhance the accuracy of face detection. Li et al.

978-1-7281-1213-8/19/$31.00 ©2019 IEEE

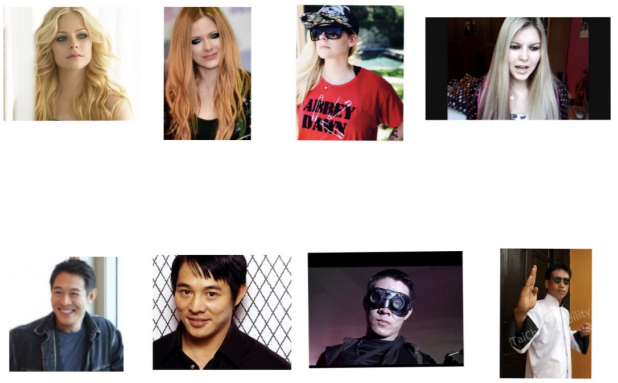

Fig. 1. Some examples from the Disguised Faces in the Wild (DFW) dataset. (The first image is the genuine upfront face of the subject. Validation image contains also the upfront face of the subject. The third image is the subject with a form of disguise. The last photo is an image for an impersonator which looks like the subject yet is not the same.)

[11] applied LBP and HOG features to decide whether the detected object is a person and then used a Haar feature-based disguise Adaboost classifier to determine the class of disguise (sunglasses, caps and masks). Disguised and occluded faces were also studied under the field of facial landmark detection; Burgos-Artizzu et al. [7] suggested a Robust Cascaded Pose Regression which detects occlusions explicitly. Wu et al. [12] improved the previous work on occluded faces and constructed one unified model to handle occlusions, instead of training several independent models for each type of occlusion.

Neural networks as a fast-developing technique has been implemented in several aspects of face processing, and face detection is no exception [21]. Most face detection techniques could be categorised as feature- based and image-based. Neural networks, by treating face detection as a general pattern recognition task, is applicable to face detection. Li et al. [17] introduced a 6-CNN cascade consists of 3 CNNs for face and non-face classification and 3 CNNs for bounding box calibration. Each of CNNs used ReLu (Rectified Linear Unit) as its activation function after the pooling and fully-connected layers. This architecture outperformed the state-of-art methods on the Annotated Faces in the Wild (AFIW) dataset. Deep Dense Face Detector was another successful application of CNN in face detection [18]. This proposed CNN architecture fine-tuned AlexNet and trained on Annotated Facial Landmarks in the Wild dataset; this algorithm was able to handle occlusion to some extent. Garcia et al. [19] designed a CNN architecture that was capable of detecting highly variable face patterns. This architecture had 104 neurons on its first convolutional layer and a 5-by-5 filter for feature extraction.

All these algorithms and datasets have contributed to the study of face processing under occlusion and disguise; the result for detecting disguised face has substantially improved. However, none of these researches focused on the factors that could impact the result of proposed algorithms. With the understanding of features in images that contribute the most to the result, pre-processing could be done in the further studies to improve the accuracy.

TABLE I
DIFFERENT CONVOLUTIONAL NEURAL NETWORK MODEL ARCHITECTURES USED IN THIS WORK

| CNN Models | Features | |
|---|---|---|
| | Number of Layers | Classification Type |
| AlexNet | 8 | Subject-Independent |
| VGG-16 | 16 | Subject-Dependent |
| ResNet50 | 50 | Subject-Independent |

## III. METHODOLOGY

### A. Disguised Face in the Wild

The dataset we used to perform analysis on is the Disguised Face in the Wild Dataset, which is the largest existing dataset focused on disguised and impostor faces under unconstrained environments. Although previous datasets focused on some forms of disguise or occlusion, the DFW dataset provides a larger variety of images under different conditions. This dataset contains 11157 images and 1000 identities with four different labels: genuine, validation, disguised and impostor. The testing dataset consists of 600 subjects and 7771 images while the training set contains 400 subjects and 3386 images. Since this paper focuses on recognition of disguised faces, only genuine, validation and disguised images were used, with 2440 images for training and 4209 images for testing.

### B. Pre-processing

The DFW dataset provided boundary box coordinates detected using Faster RCNN for images. Therefore, all the images were cropped to the boundary box area. All images were re-sized to 224*224 pixels in order to normalise for the classification stage.

### C. Convolutional Neural Network Models

There are various convolutional neural networks models that we employed in this paper, which are shown in Table I. Firstly, we used an 8-layer neural network, AlexNet, containing five convolutional and three fully connected layers [13] . AlexNet is a pre-trained CNN architecture that has significantly contributed to supervised deep learning as it used ReLu to add non-linearity to accelerate the speed of training, compared to using saturating activation functions like tanh. ReLu was thus introduced and became a commonly used activation function in CNN architecture. It also contributed to the idea of applying dropout in regard to the overfitting problem, and overlap pooling to reduce the size of the network. It used 11x11x3 kernels in the first convolutional layer with stride size of four pixels; the network has 4096 neurons in each of the fully-connected layers and uses back-propagation to calculate the gradient in weight. The CNN cascade approach to face detection also followed similar architecture of AlexNet [17]. We adapted this pre-trained model for ImageNet and fine-tuned it to fit our set of images and classes. Six epochs were performed in the training.

The other architecture we employed is a Siamese architecture which takes a pair of images as input, either (disguised, genuine) or (genuine, genuine) and returned 0, disguised, or 1, genuine, as labels. A similar approach was introduced in DisguiseNet where the algorithm classifies between (genuine, disguised) and (genuine, impostor) pairs. This architecture has two 16-layer convolutional neural networks to extract identical features from each photo in the pair separately. Each 16-layer convolution neural network is an implementation of the VGG16 model, which has similar architecture as AlexNet, yet is greater in depth [14]. It replaced the large kernel-sized filters in the first two convolutional layers in AlexNet with multiple small three by three kernel sized filters with stride size 1. Also, using smaller size of kernels enabled extraction of more features from images. The DisguiseNet algorithm was built upon the structure of VGG16, which also uses mini-batch gradient descent to momentum, based on back propagation, and is pre-trained on the labelled Faces in the Wild and the YouTube Faces dataset. We fine-tuned the pre-trained VGGFace model, and performed 50 epochs on the low-level features, followed by the DisguiseNet approach.

The last neural network we have trained is Resnet50, which is a fifty-layer deep residual network that has compelling performance and deeper layers. Although it is generally accepted that deeper networks perform better on learning complex inputs, we do not observe this phenomenon; as the depth of proposed network algorithm continued to increase, it was found that the neurons in earlier layers learned very slowly due to their negligible small gradient. This problem occurred in training a deep neural network that was gradient-based and used back-propagation, and caused the degradation of accuracy of neural network models. He et al. [15] presented the idea of identity mapping by shortcuts and applied residual learning to every three stacked layers; instead of learning unreferenced functions, ResNet learnt the residual function referred to each layers inputs to achieve low complexity yet at greater depth. It also kept the features of the ReLu activation function and small filter size from AlexNet and VGG16. It has one fully connected layer with 1000 neurons. In our implementation, six epochs were performed to the fine-tuned models.

### *D. Feature Extraction*

Several features that may influence the result of disguised face recognition were examined, including skin colour existence, luminosity of different surfaces, luminosity around edges, colour differences around edges, texture, colour and luminosity on skin patch areas (see Fig. 2). We fine-tuned the pre-trained AlexNet and Resnet50 models in Matlab deep learning toolbox by replacing the last fully connected layer and classification layer and setting the epoch number to be 6.

One of the common ways of disguise is wearing sunglasses or masks that cover several essential facial landmark points; this also reduces the visible skin colours in the image. Also, the scatter of skin colour in the image indicates there is a low chance that it is a genuine person in the image. We used a skin colour detector to locate the skin patches in the image.

Fig. 2. Some examples from preprocessed image to extract features. (The first image is the colour image. The second image only extracts the regions from skin colours. The third image is the image with only luminosity of different surfaces extracted. The 4,5th images only contain the features, luminosity or colour, around the detected edges. The following two images extract luminosity and colour in the skin-colour area and the last two images were applied with Range and Standard Deviation Texture filters respectively.)

By first converting the images to LAB colour space and then to a binary image, we could hence label pixel with skin colours as 1 and the rest of the pixels to 0.

Luminosity of different surface could also potentially impact the result of classification. The light reflection is varied on different textures; for instance, the luminosity of sunglasses, masks and skin would be varied and it could be an indication that people being disguised when the luminosity in the image has a great contrast. Sunglasses, masks and scarfs may appear darker than surrounding areas; this feature may be extracted in the neural network and indicates that there some forms of disguise [20]. To extract this feature from the images, we converted the RGB images into HSV images and reserved only one colour channel, the Value layer, which represents brightness.

Texture, similar to luminosity on different surfaces, would have a great contrast when the surface has changed. The texture of skin and fabric would be entirely different. If this could be detected and extracted from the image, this could give a guidance to the classification. We applied a standard deviation texture filter and range filter on the image in order to detect the local variability of pixel values in a given region. The texture difference is expected to lead to diverse results after filtering. The images with filter were stored.

Edge detection also could extract some useful features for training the neural network. As the luminosity would alter around the edges of faces and objects, only leaving luminosity around edges may introduce less irrelevant information or

noise to the model. In a similar fashion, only leaving the edges in a colour images may reveal whether the neural network is reliant on pattern matching. We applied a Canny edge detector, which used the Gaussian filter to prevent noise, to detect the edge and set a threshold value to further filter out the minor edges detected. We extract the luminosity and colour features by only reserving the pixels near the edges.

The extracted regions from skin colour images were only black and white, this may cause a great loss of information. Therefore, we combined skin colour regions with the luminosity images and colour images respectively to preserve the information provided in the face or skin area. The noise that potentially could be introduced by background, clothes and hairstyles were excluded. All the features extraction was only done in the detected skin region. In order to achieve this, we used the same approach in extracting pixels around edges by only keeping pixels in the skin region.

All these processed images were stored as RGB images and used as input to train the neural network models that had identical architectures to the one applied to the original images. All the parameters were kept the same to produce comparable results.

### *E. Fusion*

Fusion techniques are widely used in research not only to enhance the performance of a system but also to increase the confidence level of the final decision. Fusion can be performed as pre-matching (early) fusion, where features are fused before the classification, and post-matching (late) fusion, where the classification decisions are fused. Even though fusion is usually used when different modalities are utilised (e.g. video and audio), we perform late fusion on the decisions of different image features. For simplicity, we perform fusion on decisions (labels) out of individual image features classifications using majority voting.

### *F. Performance Analysis*

To analyse our result, a confusion matrix was used to calculate an average accuracy for comparing different features using AlexNet and ResNet50; both true positive and true negative value were recorded. These values could give an indication of the performance. By using these data, we also calculated the precision, recall and F1 score (refer to appendix). Since we had an uneven distribution of disguised and normal face images, the F1 score would be a better indication than accuracy. Nevertheless, the result using DisguiseNet is based on subject dependent recognition and hence the accuracy is for predicting pairs of images.

As the provided training and testing dataset had unbalanced images for each class (i.e. original and disguised), it is critical that the true positive and negative rate is balanced. Otherwise, the model could have accuracy paradox: classifying all the subjects into the majority class, disguised, and achieving a high accuracy rate. Therefore, balancing of true positive and true negative rate is also an important criterion when we evaluated the results.

TABLE II
COMPARISON OF WEIGHTED ACCURACY BETWEEN DIFFERENT MODELS AND DIFFERENT FEATURES

| Image Features | Weighted Accuracy | |
|---|---|---|
| | AlexNet | ResNet50 |
| Original image | 70.28% | 69.91% |
| Skin patch area | 51.22% | 55.02% |
| Colour on skin patch area | 66.29% | 64.04% |
| Luminosity on skin patch area | 62.55% | 63.87% |
| Luminosity | 63.94% | 69.51% |
| Luminosity around edge | 57.14% | 61.02% |
| Colour around edge | 62.79% | 61.21% |
| Texture (standard deviation) | 49.47% | 49.79% |
| Texture (range) | 60.11% | 65.42% |

## IV. RESULTS AND DISCUSSION

We investigated the different features by training and testing on same set of data, using identical CNN architectures, with different visual features that have been extracted from the images. The performance of these features were evaluated using three calculated values, weighted accuracy, the true positive and negative rate to illustrate the balance of true positive and negative. Table II shows the weighted accuracy for the different features used in this work.

**Original Image** In our AlexNet model, the performance trained on original coloured images had an overall accuracy of 70.28% with a balanced true positive and negative accuracy of 70.98% and 69.57% respectively. The chosen Resnet50 model produced a 69.91% overall accuracy on the test dataset, with true positive rate 56.79% and true negative rate of 83.03%. These values will be used as a benchmark for the following comparisons of other visual features. The classification for the full images was reasonably accurate since it gathered features from all image information, similar to a feature-level fusion.

We were able to reproduce the DisguiseNet algorithm on the dataset and had a 83.93% overall accuracy in determining whether the subject was disguised or not [10]. Since it was a subject dependent algorithm, it achieved much higher accuracy compared to our Resnet50 and AlexNet models, which are subject-independent. The DisguiseNet algorithm was able to compare and learn a pair of images and gave classification over a pair of images based on same person; this provided more features that could potentially be extracted to the model and enhance the overall accuracy. Moreover, in order to reproduce the experiment, we had an epoch of 50 instead of 6, which was the epoch for AlexNet and Resnet50; the epoch number would also greatly contribute to the final result as it could make the model fit data more.

**Skin Patch Area** As the implemented skin colour detection algorithm was relatively simple and returned a binary image and excluded all the possible features provided by colours, the performance of AlexNet trained on this feature only had the accuracy at a chance level. A similar trend has been observed in the Resnet50 model where the overall accuracy was slightly above 50% and the true positive rate was 62.38%, which performed better than AlexNet. This could indicate that

Resnet50 might extract more features from the pre-processed skin colour images.

**Luminosity** Nevertheless, luminosity appeared to be one of the features that heavily impacted the accuracy of models performed on original images; the overall accuracy for each model was close to the accuracy of our benchmarks respectively and the trend of true positive and negative rates were also similar to our benchmark data. Although we removed the colour space in the images, AlexNet was still able to learn from the residue of information and gave sound predictions; the model produced balanced and relatively high accuracies over the test dataset. On the other hand, when using Resnet50 model to predict on test dataset, it gave a 82.69% true negative rate and a chance level accuracy on true positive. A similar trend was also illustrated in the Resnet50 model trained on the original model.

**Texture (Standard Deviation)** On the other hand, the overall accuracies of using only texture features (standard deviation) in both models were quite high, 69.66% and 71.06% respectively, with surprisingly accurate predictions in disguised faces. However, the weighted accuracy (accounting for balanced true positive and true negatives) revealed a different results, which was an example of accuracy paradox. The extreme unbalanced recognition rate was due to the fact that the model trained on texture lacks possible features to be trained on. This was caused by the standard deviation filter, which did not produce a satisfactory result on processing the image; most of the features were lost in the extraction of texture process. There was unlike to have extreme texture change in faces, which indicated that this filter may filter out most of essential information. Hence, the trained model predicted most of the test images as disguised, which made the true positive rate in both models become close to 100%.

**Texture (Range)** On the contrary, the range texture filter, produced a decent result with a weighted accuracy of 60.11% and a F1 score of 73.05% in the AlexNet model. This filter did not remove excess information and hence the texture difference contributed to distinguishing disguised faces from normal faces in AlexNet. A similar fashion was also presented in ResNet model, where the overall accuracy was 65.42% and the F1 score was 78.99%. In both models, the true positive rate was slightly higher than the true positive rate. This has shown that texture could potentially play an important role when CNN models classify on disguised faces.

**Luminosity around Edges** However, when only the luminosity around edge regions was provided in the image, there were not enough features for AlexNet to recognise the pattern. The removal of regions other than edges caused the accuracy to drop to 57.14% where the true positive rate decreased 24% from the benchmark and the true negative rate remained at a similar accuracy level as the benchmark.

Unlike the AlexNet model, the Resnet50 model had a sound performance when only luminosity around edges was provided. It had a F1 score of 76.64%, which indicated that the Resnet50 model was able to detect true disguised faces at a decent accuracy. This trend was opposed to the observation made in the Resnet50 model using luminosity; this could infer that Resnet50 was more sensitive to edges.

**Colour around Edges** colour detected around edges was the feature that had the third-best performance in all AlexNet models. It had a relatively balanced rate between true positive and negative and the overall accuracy was not too far from the benchmark performance. It is worth noting that when the images used did not have any excluded section, the prediction accuracy for disguised faces was slight better than the accuracy of true prediction for normal faces. However, when only the edge area was detected, the accuracy for predicting disguised faces dropped and the true negative rate increased. This indicated that the edge detection may help AlexNet classify normal faces more easily, yet reduce the accuracy of predicting disguised faces.

This feature was also the third-best performing feature in all Resnet50 models where it had a balanced accuracy between true positive and true negative. It also had a high recall rate, 80.27%, and a F1 score around 70%. In the model trained on original images, the normal faces were classified accurately where disguised faces were not detected very efficiently. By only providing colour information around edge regions, it seemed that less noise was introduced to the model and hence improved the performance; in the meantime, the F1 score remained the same as the benchmark performance, suggesting that the accuracy of recognising disguised face did not degrade.

**Colour or Luminosity on Skin Patch Area** For both AlexNet and Resnet50, only extracting features from skin patch area showed the models performed relatively well on recognising normal faces, especially for Resnet50.

In AlexNet, the accuracies of true negative in testing using colour or luminosity around skin patch area features are 74.08% and 77.07%, even outperforming the benchmark result, yet the true positive rates were around chance level. This pattern indicated that for AlexNet models, exclusion of information in certain areas would degrade the accuracy in prediction of disguised faces; however, the accuracy in predicting normal faces would increase.

This tendency was also shown in the two Resnet50 models where the true negative rates exceeded the benchmark performance and nearly reached 90%. Also, the recall rates were quite high. This indicated that out of all the disguised faces, nearly 90% were correctly detected. However, when comparing to benchmark values, there was also degradation in the precision rate from 56.79% to around 40%.

We also perform a late fusion on some of the individual image features to investigate its performance compared to when using the full image (see Fig. 3. We select five image features, the best two, the worst two and one average on ResNet network to have a variety of classification results. The selected features are skin patch area, colour on skin patch area, luminosity, colour around the edge, and texture (standard deviation). As can be seen from Fig. 3 the fused results (72%) outperformed the whole image and the individual image features. Running a U-test (WilcoxonMannWhitney

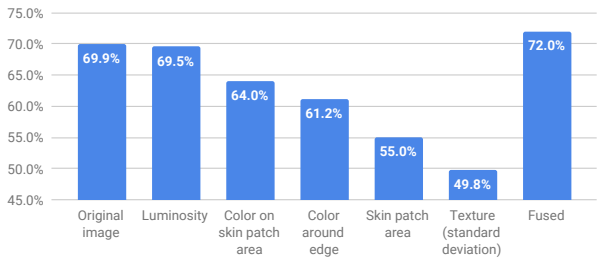


Fig. 3. Fusion Result and Comparison Using Selected Image Features on ResNet

test), the fused results shows statistically significant results when compared to individual image features. The result from the fused classification of image features is not only the state-of-the-art in subject-independent disguised face detection, but also increases the confidence level of the final decision.

## V. Conclusion

From the data collected using AlexNet and Resnet50, there were several crucial features that played an essential role in the classification of disguised and original faces: colour or luminosity of the image, which introduces rich information to the network for feature extraction, the texture of the image, which contributes to the decent accuracy of identifying disguised faces, and edges.

In both models, we observe that luminosity contributed most to the overall results, as it showed similar trends to the original image models. As the AlexNet model trained on original images gave a reasonably balanced accuracy in term of true/false positives, the same CNN trained using luminosity was the best performing feature. Moreover, colour around edges also played an important role in the final result as this feature was the second-best feature in AlexNet models and the best features in Resnet50 models, which had a good balance between the true positive and negative rates.

Our work explored some important features in images that have potential to impact on CNN classification algorithm, and evaluated their importance to face image processing. Applying late fusion on some of these features showed a significant improvement in classification results, which also increases the confidence level of the decision, since the decision is based on agreements between several individual features classification. The influence on CNN results using these features and their fusion provide a greater understanding, where further pre-processing methods could be proposed and further improve the state-of-art accuracy of detection of disguised faces.

There are some limitations in our works which may introduce some noise in the models. The Faster RCNN used to provide boundary box data was not accurate enough and some of the cropped images did not include the full face. Also the skin colour detector we implemented was not precise enough and it may have added some irrelevant features in the output images. Also, due to the limitation of chosen dataset, there is a large unbalance in the training and testing sets; not enough normal (non-occluded) faces were included in the dataset. This could cause the trained model to become biased.

In the future, instead of extracting lower level features from images, we could investigate how some state-of-art CNN algorithms perform in the absence of some facial landmark. We could investigate a decision-level fusion of the individual visual features to analyse the accuracy in such fusion level in the future. This might provide a further insight into what specific features contribute to the performance of CNNs compared to when using the full image features. Moreover, we could perform this test on an improved dataset where there is a balanced number of normal and disguised faces.

## References

[1] S. Liao, A. K. Jain, and S. Z. Li. Unconstrained face detection.Technical report, Michigan State University, December 2012.
[2] J. Cheney, B. Klein, A. K. Jain, B. F. Klare, ”Unconstrained face detection: State of the art baseline and challenges”,ICB, pp. 229-236, 2015.
[3] H. Jiang, E. Learned-Miller, Face detection with the faster r-cnn, 2016.
[4] S. Yang, P. Luo, C. C. Loy, X. Tang, ”WIDER FACE: A Face detection benchmark”.
[5] V. Jain and E. Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical Report UMCS-2010-009, University of Massachusetts, Amherst, 2010. 4
[6] X. Sun, P. Wu, S.C. Hoi, Face detection using deep learning: An improved faster rcnn approach, 2017, [online] Available: https://arxiv.org/pdf/1701.08289.pdf.
[7] Burgos-Artizzu, X.P., Perona, P., Dollar, P.: Robust face landmark estimation under occlusion. In: ICCV, pp. 15131520 (2013)
[8] T. Y. Wang and A. Kumar. Recognizing human faces under disguise and makeup. In IEEE International Conference on Identity, Security and Behavior Analysis, 2016.
[9] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa. Disguised faces in the wild. Technical report, IIIT Delhi, March 2018.
[10] S. Peri , A, Dhall. DisguiseNet: A Contrastive approach for disguised face verification in the Wild. In CVPR Workshop on Disguised Faces in the Wild, 2018.
[11] J. Li, B. Li, Y. Xu, K. Lu, K. Yan, and L. Fei. Disguised face detection and recognition under the complex background. In CIBIM, pages 8793. IEEE, 2014.
[12] Y. Wu and Q. Ji. Robust facial landmark detection under significant head poses and occlusion. In ICCV, 2015.
[13] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks. In NIPS, pp. 11061114, 2012.
[14] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015
[15] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition(pp. 770-778).
[16] Zhang, K., Chang, Y. L., and Hsu, W. (2018). Deep disguised faces recognition. InCVPR Workshop on Disguised Faces in the Wild(Vol. 4, p. 5).
[17] Li, H., Lin, Z., Shen, X., Brandt, J., and Hua, G. (2015). A convolutional neural network cascade for face detection. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition(pp. 5325-5334).
[18] Farfade, S. S., Saberian, M. J., and Li, L. J. (2015, June). Multi-view face detection using deep convolutional neural networks. InProceedings of the 5th ACM on International Conference on Multimedia Retrieval(pp. 643-650). ACM.
[19] Garcia, C., and Delakis, M. (2004). Convolutional face finder: A neural architecture for fast and robust face detection.IEEE Transactions on pattern analysis and machine intelligence,26(11), 1408-1423.
[20] Brimblecombe, P. (2002). Face detection using neural networks.H615Meng Electronic Engineering, School of Electronics and Physical Sciences, URN,1046063.
[21] Al-Allaf, O. N. (2014). Review of face detection systems based artificial neural networks algorithms.arXiv preprint arXiv:1404.1292.