

FERDOUS, R.H., ARIFEEN, M.M., EIKO, T.S. and AL MAMUN, S. 2020. Performance analysis of different loss function in face detection architectures. In *Kaiser, M.S., Bandyopadhyay, A., Muhmad, M. and Ray, K. (eds.) Proceedings of International conference on trends in computational and cognitive engineering 2020 (TCCE-2020), 17-18 December 2020, Dhaka, Bangladesh*. Singapore: Springer [online], 659-669. Available from: https://doi.org/10.1007/978-981-33-4673-4_54




Performance analysis of different loss function in face detection architectures.

FERDOUS, R.H., ARIFEEN, M.M., EIKO, T.S. and AL MAMUN, S.

2020

This is a post-peer-review, pre-copyedited version of an article published in Proceedings of International conference on trends in computational and cognitive engineering. The final authenticated version is available online at: https://doi.org/10.1007/978-981-33-4673-4_54

Performance Analysis of Different Loss Function in Face Detection Architectures

Rezowan Hossain Ferdous ¹, Md Murshedul Arifeen ¹, Tipu Sultan
Eiko¹, and Shamim Al Mamun ²

¹ Bangladesh University of Professionals, Dhaka-1216, Bangladesh
rezosam0@gmail.com

² Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

Abstract. Masked face detection is a challenging task due to the occlusions created by the masks. Recent studies show that deep learning models can achieve effective performance for not only occluded faces but also for unconstrained environments, illuminations or various poses. In this study, we have addressed the problem of occlusion due to wearing masks in masked face detection technique in deep transfer learning method. We have also reviewed the recent deep learning models for face detection and considered VGG16, VGG19, MobileNet and DenseNet as our underlying masked face detection models. Moreover, we have prepared a dataset containing masked face and without mask from 120 individuals and enhanced the dataset using augmentation. After training the deep learning models with our own dataset, we have analysed the performance of the deep learning models for several types of loss functions. From the experiment, it is clear that all the deep learning models perform well in terms of classification losses like categorical cross entropy loss and KL divergence loss.

Keywords: Face Detection, VGG Face, Deep Learning

1 Introduction

Face detection can be considered as a computer vision and pattern recognition task where various face detection algorithms try to identify and determine faces including their positions, sizes and poses from the input image. Face detection involves three principle steps [1]- selecting a probable area on the input image which is also known as observation window. Then from that window several features are extracted and finally based on these features it is determined by a face detection algorithm that whether there is a face or not. Face detection is the underlying mechanism of various computer vision problems such as- face recognition, verification, tracking, behavior analysis, attribute recognition, emotion detection, gender classification, synthesis and sentiment analysis [2], [3]. Face detection is the very first step towards various face based technologies like mentioned above. Moreover, face detection itself is essential for various purposes like photo shooting [4]. Taking photo using camera utilizes face detection technology

to move the lens in the right position. Another useful uses of face detection is head pose estimation which is significant in drivers drowsiness detection, behaviour analysis, gesture detection and gaze estimation. Also, face detection is used in mobile devices for unlocking the device or for security purpose. Bio metric security is another research field where face detection is used enormously.

Face detection and recognition is a challenging task for several years but now a days researchers have come up with various ideas and utilization of deep learning in face detection problems make this field one step ahead. But still there are lots of issues regarding face detection which needs to be considered. In case of bio metric security applications, occluded face or partially occluded face creates difficulty for the automated systems to detect and recognize the intended person effectively. Furthermore in case of pandemic situation like COVID-19 when everybody wears masks for their health safety, it is necessary to adapt the existing face detection algorithms to identify the persons with masks without any difficulty. The performance of the existing algorithms decrease sharply when the input images contain masked faces. The main problem is the occlusion which is generated from partially covering the face with mask. Other challenging tasks like illumination, pose variation problems are addressed in various works but masked face detection is not considered well enough for various applications like video surveillance system. Therefore, the performance of the algorithms decline drastically due to masked faces which creates occlusion. Furthermore, there are lack of proper dataset containing masked faces with augmentation. If we want to train a deep learning model for classifying masked faces or face without mask, it is required to construct a proper dataset with a reasonable amount of images containing both masked faces and faces with no mask. The existing dataset is imbalanced and do not provide sufficient amount of images to train a deep learning model for the purpose of classification.

In this work, we have performed a comparative study on the performance of various deep learning model and find out that VGG can work well in case of masked face classification. We have constructed a dataset balancing masked face and face with no mask of 200 people and augmented the images into various types. Then, we performed an experiment on VGG for various loss functions and determined which loss function performs better in case of masked face classification task.

2 Deep learning in face detection

A multitask convolutional neural network is proposed in [1] for face detection and facial landmark localization at the same time. Integrating these two tasks not only improve detection accuracy but also reduces the running time. The face detection method proposed in [2] is based on mobile net. The architecture is lightweight and applicable for mobile applications. Since mobile net uses depth wise convolution so it is more effective in terms of computational costs rather than general convolution operation. For detecting and aligning face in unconstrained environment, the authors in [5] proposed a deep learning based

Table 1. Comparisons of different deep learning models

Methods	Filter Size	Loss Function	Data set	Input Image Size	Performance
MTCNN [5]	3×3	Cross-Entropy Loss (Classification) Euclidian Loss (Bounding Box Regression)	Wider Face, CelebA	n/a	16fps on CPU 99fps on GPU
VGG [6]	3×3 , 1×1	Euclidian Loss	ILSVRC-2012	224×224 RGB	n/a
DCF [7]	5×5	n/a	CAS-PEAL-R1, VOC2012	32×32	n/a
Cascade [8]	1×1	Cross-Entropy Loss (Classification), Euclidian Loss (Bounding Box Regression)	Wider Face (Training+Test), FDDB (Test)	n/a	28 fps
Mobile [2]	1×1		FDDB	224×224 RGB	n/a
DAFE-FD [3]	3×3	multitask loss	WIDER, FDDB, Pascal faces	400×800 , 1200×1600	computational time is 50 ms/image and 190ms/image
Yolo based [9]			FDDB	448×448	10-25 fps, 26-40fps
Faceness net [10]	n/a	Cross Entropy Loss (Classification)	CelebA	n/a	14fps cpu 100fps gpu
Hyperface [11]	na	Euclidian Loss (Landmark Localization)	AFLW	n/a	n/a
Masked Face [12]	n/a	n/a	WIDER FACE, MASKED FACE	12×12 , 24×24	n/a
Multiscale feature [13]	na	cross-entropy loss	FDDB, AFLW	24×24	n/a
Tiny face [14]	n/a	regression loss	FDDB, Wider Face	200×200	45fps
Densenet [15]	1×1	(focal loss function), Progressive Anchor Loss	WIDER FACE	640×640	n/a
Deep Cascade CNN [16]	1×1	Cross entropy loss and L1 loss	FDDB, WIDER face	24×24	Validation accuracy is above 94%

multitask cascaded joint face detection method known as MTCNN which shows effective results in terms of illuminations or occlusions. This network involves proposal network, refinement network and output network for the generation of the outputs. To increase the efficiency and more effective results than AlexNet, the authors in [6] proposed a deep CNN based method known as VGG for object detection. The convolutional layers of VGG network consists of 3*3 filter and also a 1*1 filter working as linear transformation with 1 pixel convolutional stride followed by ReLU function. A fast face detection method based on DCF and deep CNN has been proposed in [7]. The proposed method consists of 2 key components. The first one is the nonlinear mapping function which itself includes convolution layer and max pooling layer. The second component is the sparse discriminative features which enhances linear separability of the features on the output feature maps of the nonlinear mapping function. The recently proposed face detection methods use single stage network which lacks the generalization ability as well as require fixed size images. In [8] the authors proposed a cascaded deep convolutional network to overcome these challenges. The proposed method performs two tasks in detection stage- face classification and bounding box regression. The accuracy of the model is enhanced through cascading convolutional neural network. The work in [9, 17, 18] focuses on improving the accuracy of face detection using YOLO method. The proposed network includes seven convolution layer and a max pooling layer. The authors in [10] focused on facial attributes and proved that facial attributes based supervision can effectively increase the occlusions handling capability of face detection network. They have proposed a deep convolutional neural network based face detection method by including facial attributes based supervision. Similarly, various methods and classification tasks are proposed in various literature's for face detection.

3 Loss Functions

3.1 Cross Entropy Loss

Cross entropy is a type of measurement used in information theory based on entropy theory and computes the differences between two probability distribution. This loss is also known as log loss. Cross entropy loss is used to measure the performance of a machine learning classification model which gives output as a probability. This loss increases with the deviation of the predicted probability value from the actual label value. Cross entropy loss is also known as softmax loss and effectively used in face detection or face recognition task. It can be defined as-

$$CE = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_{y_j}^T x_i + b_{y_j}}}$$

where, the symbol W represents the weight matrix, b is the bias term, x_i is the i^{th} training sample, y_i denotes class label for i^{th} training sample, N defines the total number of samples. Finally, W_j and W_{y_i} are j^{th} and y_i^{th} column of W.

3.2 KL Divergence Loss

This loss function is used to measure the difference between two probability distributions and shows how one probability distribution differs from another one. Lets say we have a probability distribution $P(X)$ and we want to replace this distribution with a known probability distribution say $Q(X)$. In this scenario, the KL divergence loss measure the difference of the approximate distribution $Q(X)$ from the true distribution $P(X)$. Mathematically, it can be defined as,

$$KL(P||Q) = \sum P(X) \log\left(\frac{P(X)}{Q(X)}\right)$$

Intuitively, this loss defines that how much information is lost or gained when we approximate a probability distribution with another probability distribution.

3.3 Mean Squared Error Loss

This loss in the average of the squared error that is used as the machine learning loss function for least squares regression. This loss can be defined mathematically as-

$$\sum_{i=1}^n \frac{(w^T x(i) - y(i))^2}{n}$$

This function performs sum over all the data points, of the square of the difference between the predicted value and target values and the whole is divided by the total number of data points. This loss function is one of the most simplest loss function used in Machine Learning.

3.4 Huber Loss

Huber loss function includes mean squared error loss and mean absolute error loss function and provides the best results by balancing those two loss functions. It can be defiend as -

$$L(y, y') = \begin{cases} \frac{1}{2}(y - y')^2 & \text{if } \|y - y'\| \leq \delta \\ \delta\|y - y'\| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$

From this equation, we can understand that if the loss vales is less than δ , mean squared error loss will be used and mean squared absolute error will be used otherwise.

4 Experimental Setup

The experiment has been conducted on Google Colab. We have used Tesnsorflow and Keras for our implementation. We have also used MTCNN library to

extract faces from the input images. Sequential model of Keras with several pooling, convolution and padding layer. We have used VGG face as our underlying architecture. To prevent overfitting, we have used dropout. The VGG face weight is loaded for tuning the model. After that we have removed the last two layers and fitted our dataset for further training. We also have added the dense layer, tanh activation function, batch normalization with dropout.

4.1 Datasets

Since there are no balanced masked face datasets available in terms of skin tone, aging factor, orientation and gradient, therefore we have prepared our own dataset containing masked and non-masked faces. The dataset contains 720 face images belonging to 120 individuals. To obtain all the features of a face, we have taken front view and both side views of each individual wearing having mask and without masks. Moreover, there are 6 images for each person which was used image data generator from Keras library to augment on those images. Furthermore, 6 images of each person is augmented to 24 images through changing various factors for instance – brightness, rotation, shifting, flipping, sharing and zooming. We have separated 18 images for training purpose, 4 images for validation purpose and 2 images for testing purpose from the augmented 24 images of each individual. We divided our dataset 80:20 ration that is the total number of training images is 1724, and for validation and testing rest of the 432 images are being used for getting performance accuracy.

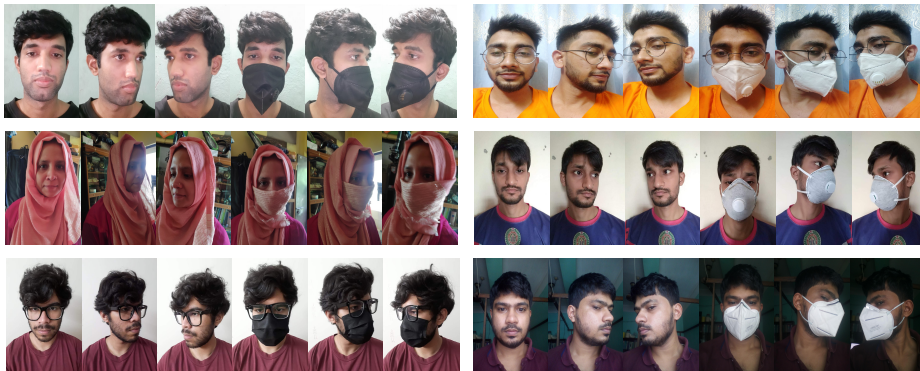


Fig. 1. Sample of our dataset which shows each individuals six different input images including masked faces and without masked faces. Among the dataset of each individual, one is front view and the other two is side views.

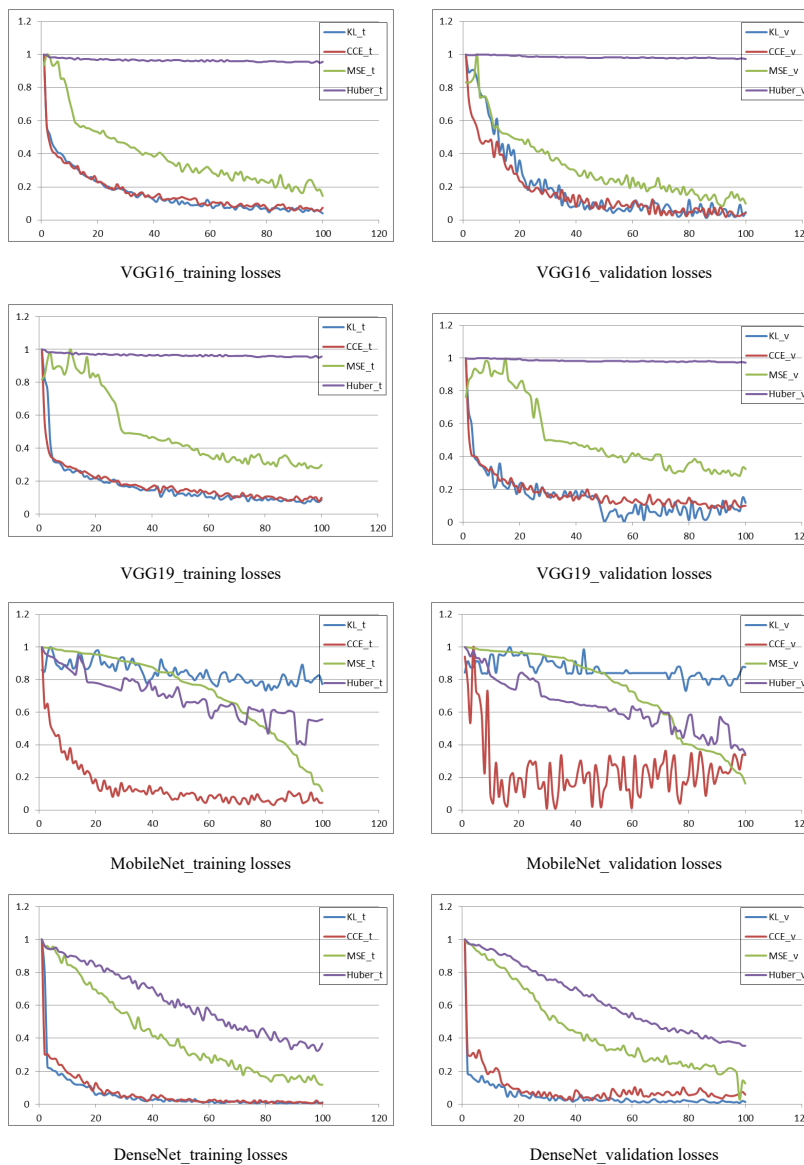


Fig. 2. Four different loss functions KL Divergence loss (KL), Categorical Cross Entropy loss (CCE), Mean Squared Error Loss (MSE) and Huber Loss (Huber) against epochs for four different deep learning models (VGG16, VGG19, MobileNet and DenseNet)

5 Result and Discussion

Machine learning or deep learning models learn from the given datasets using the loss functions. Through loss function, we can understand that how a deep learning algorithms models the given data. If the predicted output of a given deep learning model deviates too much from the actual data or output, loss function will show high error value. In this study, through our experiment we have showed the best loss function which can be chosen for masked face detection tasks. We have used classification losses such as Categorical cross entropy loss, and KL divergence loss, regression losses such as Mean squared error loss and Huber loss for our experiment. Figure 2 shows the experimental results. Classification losses are probabilistic loss and they predict the output based on the highest probability it gets from the last layer of the deep learning models. On the contrary, regression losses predict continuous values. From our experimental results, we can see that, all the models perform better in case of classification loss that is for Categorical cross entropy (CCE) loss and Kl divergence (Kld)loss. From the figures it is evident that for all the deep learning models that is for VGG16, VGG19, MobileNet and DenseNet, the loss function curves tend towards zero for only for CCE and Kld losses. Other losses do not converge towards zero. Hence, we can say that the CCE and Kld loss performs well in case of deep learning based face detection models.

6 Conclusion

In this work, we have addressed the problem of occlusion due to wearing masks in case of masked face detection and analysed the performance of various types loss functions for different deep learning models. We have prepared our own dataset for the experimental purpose which is enhanced through augmentation technique. For the experiment, we have considered four different types of deep learning models for instance VGG16, VGG19, MobileNet and DenseNet. Experimental results show that the deep learning models performs better for classification losses such as- Categorical cross entropy loss and KL divergence loss.

References

1. Zhang Y, Wang H, Xu F. Face detection based on multi task learning and multi layer feature fusion. In: 2017 6th International Conference on Computer Science and Network Technology (ICCSNT). IEEE; 2017. p. 289–293.
2. Almadhor A. Deep Learning Based Face Detection Algorithm for Mobile Applications. In: TENCON 2018-2018 IEEE Region 10 Conference. IEEE; 2018. p. 1158–1162.
3. Sindagi V, Patel V. Dafe-fd: Density aware feature enrichment for face detection. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE; 2019. p. 2185–2195.
4. Liu H. Face Technologies on Mobile Devices. In: Facial Detection and Recognition on Mobile Devices. Elsevier; 2015. p. 11–38.

5. Zhang K, Zhang Z, Li Z, Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*. 2016;23(10):1499–1503.
6. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556*. 2014;.
7. Guo G, Wang H, Yan Y, Zheng J, Li B. A fast face detection method via convolutional neural network. *Neurocomputing*. 2019;.
8. Qi R, Jia RS, Mao QC, Sun HM, Zuo LQ. Face detection method based on cascaded convolutional networks. *IEEE Access*. 2019;7:110740–110748.
9. Garg D, Goel P, Pandya S, Ganatra A, Kotecha K. A Deep Learning Approach for Face Detection using YOLO. In: 2018 IEEE Punecon. IEEE; 2002. p. 1–4.
10. Yang S, Luo P, Loy CC, Tang X. Faceness-net: Face detection through deep facial part responses. *IEEE transactions on pattern analysis and machine intelligence*. 2017;40(8):1845–1859.
11. Ranjan R, Patel VM, Chellappa R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;41(1):121–135.
12. Bu W, Xiao J, Zhou C, Yang M, Peng C. A cascade framework for masked face detection. In: 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). IEEE; 2017. p. 458–462.
13. Chen Q, Meng X, Li W, Fu X, Deng X, Wang J. A multi-scale fusion convolutional neural network for face detection. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE; 2017. p. 1013–1018.
14. Ye F, Ding M, Gong E, Zhao X, Hang L. Tiny Face Detection Based on Deep Learning. In: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE; 2019. p. 407–412.
15. Nandy A. A Densenet Based Robust Face Detection Framework. In: Proceedings of the IEEE International Conference on Computer Vision Workshops; 2019. p. 0–0.
16. Peng B, Gopalakrishnan AK. A Face Detection Framework Based on Deep Cascaded Full Convolutional Neural Networks. In: 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS). IEEE; 2019. p. 47–51.
17. Mamun SA, Suzuki R, Lam A, Kobayashi Y, Kuno Y. Terrain Recognition for Smart Wheelchair. In: Huang DS, Han K, Hussain A, editors. *Intelligent Computing Methodologies. Lecture Notes in Computer Science*. Springer International Publishing; 2016. p. 461–470.
18. Al Mamun S, Fukuda H, Lam A, Kobayashi Y, Kuno Y. Autonomous Bus Boarding Robotic Wheelchair Using Bidirectional Sensing Systems. In: Bebis G, Boyle R, Parvin B, Koracin D, Turek M, Ramalingam S, et al., editors. *Advances in Visual Computing. Lecture Notes in Computer Science*. Springer International Publishing; 2018. p. 737–747.