ISSN No:-2456-2165

Explainable Artificial Intelligence: How Face Masks are Detected via Deep Neural Networks

Ahmet Haydar Ornek, Mustafa Celik
Integration Solutions Development Department
Huawei Turkey R&D Center
Istanbul, Turkey

Murat Ceylan
Electrical and Electronics Engineering Department
Konya Technical University
Konya, Turkey

Abstract:- The image classification has become a well-known process with the development of deep neural networks. Although classification studies above 90% accuracy are realized, their explainable side is still an open area which means the classification process are not known by researchers. In this study, we show what a deep neural network model learns from face images to classify them into with mask and without mask classes using last convolutional layers of the model. As a deep neural network model ResNet-18 was selected and the model was trained with 18600 balanced face images belonging two classes and tested with 4540 face images different from training images. The model's test results are obtained as 95.16% sensitivity, 96.69% specificity, 96.58% accuracy. With the created activation maps it is clearly seen that the model learns face structure for images without mask and mask structure for images with mask.

Keywords:- *Classification; Covid-19; Explainable Artificial Intelligence; Transfer Learning.*

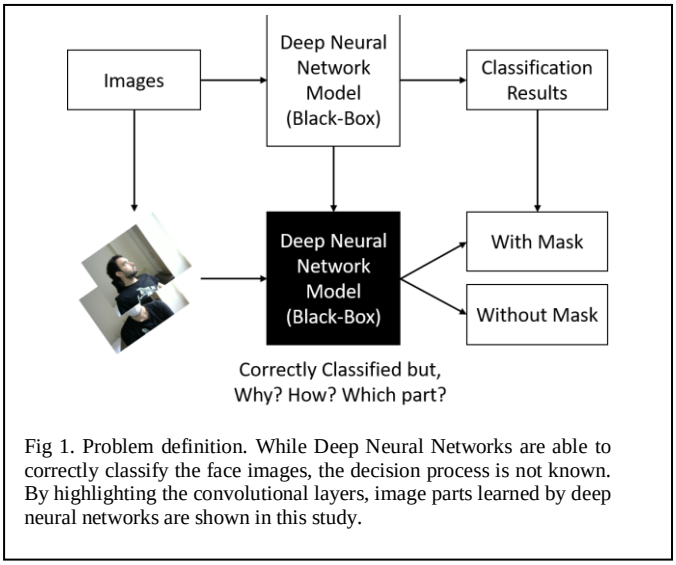

Fig 1. Problem definition. While Deep Neural Networks are able to correctly classify the face images, the decision process is not known. By highlighting the convolutional layers, image parts learned by deep neural networks are shown in this study.

I. INTRODUCTION

The computer vision applications such as image classification, object detection, image segmentation and clustering are being solved with high performance by deep neural network models [1-9]. Since the models have layers having more than 1M parameters their explainable side becomes bad-known state and the models are called as black-box models. By applying the class activation map [10] technique to our image classification problem that try to detect whether people wear a face mask image or not (Fig. 1 was created to show the problem definition), we show how the images are classified into mask and no mask classes by highlighting the related regions over the images.

During the pandemic that affects the whole people, it is important to follow if people wear a face mask [11]. To decide that an intelligent system is required that has a camera and processing unit.

In traditional approaches such as feature extraction, feature selection and classification all process is manually realized [12]. When extracted features are less and a method like the decision tree is used, the classification could be explained. In deep learning models it is almost impossible to show all features' contributions by using the decision tree classifier.

In explanations of the models, there are three main ways as numerical, visual, and rule-based [13]. The numerical methods calculate all inputs' contributions from all to zero or vice versa by using a method like Information Gain [14]. Trying manually which input changes the classification result, important features are detected in numerical methods.

The rule-based methods such as Decision Tree or Random Forest, use information gain to decide how importance of inputs for the classification. However in big models such as deep neural networks, they cannot be implemented to the models to create an explainable structure because of parameter size.

Visual approaches such as class activation map is used for deep neural models. By using a created activation map (also known as heat-map), the importance of each pixel of an image can be represented over the image.

In this sense we make following contributions to the literature by applying class activation maps to our real-world classification problem to explain how images are classified into mask and no mask classes.

- Real-world images are collected under difficult conditions such as low resolution, quality, changing light and background.
- The images are classified into classes as mask and no mask by using transfer learning method (with ResNet-18 architecture).

- The important regions over the images are highlighted using class activation maps to show how images are classified into classes.

The rest of the paper was organized that related work about face detection, mask detection and explanation studies are detailed in Section 2. The materials used in the study such as images and working environment are shown in Section 3. In Section 4 Convolutional Neural Networks, Transfer Learning, Class Activation Maps, and evaluation metrics are described. Experiments and results, Discussion are given in Section 5 and Section 6, respectively. In Section 7, conclusion and future works are explained.

II. RELATED WORK

To identify the COVID-19 number of studies with deep learning are carried out [15-20], but these studies are about clinical findings. Taking the advantage of deep learning, a system that monitor whether people wear a mask can be developed.

To detect face mask images [21] proposed a hybrid deep learning and machine learning model. Deep learning is used to extract features and support vector machines (SVM) classifies the extracted features. The SVM classifier achieved 99.64%, 99.49%, 100% testing accuracy in different datasets.

In [22] a face mask-wearing condition identification method is developed by combining super-resolution and classification methods for images. Their algorithm consist of four steps: pre-processing, face detection-cropping, super-resolution, and face mask-wearing identification. They achieved 98.70% accuracy using the proposed deep learning method.

[23] analyzed no-masked and masked face recognition accuracy using principal component analysis (PCA). According to its results, a face without mask achieves more performance in PCA based face recognition. They tried four different scenarios by changing their test size. The average accuracies are 95.68% and 70.53% for no-mask and mask, respectively.

Face detection [24-25] is a challenging problem. With the advances in deep learning, convolutional based solutions provides high efficiency in the face detection problem [26-29]. [30] used YOLOv3 [31] to detect face images by changing its layers, using softmax function, and reducing features' dimension.

In [32], an edge computing-based mask detection model is proposed to provide real-time performance on common camera devices. Their system contains three steps: restoration, face and mask detection. They achieved 95.9% accuracy.

Convolutional neural networks have ability to learn representations of images. Although it performs high performance on classification problems, its transparency side is still open to develop. It has been started to develop Gradient-based methods to highlight the important parts of images [10], example studies are person re-identification [33-34], object localization [35-37], texture analysis [38], aerial imaging [39-40] and image segmentation [41-43].

In [44] a new method is proposed which computes and highlights the main components of the important representations from the layers. They claim up to 12% improvement on weakly supervised object localization.

To the best our knowledge, mask detection and class activation maps are used for the first time in [45]. They developed a system that monitors social distance, face mask, and touching face conditions by combining deep learning based imaging system and class activation maps.

III. MATERIAL

AI projects require hundreds of images to learn effectively, and computation sources to realize mathematical operations. The collected images and working environment will be clarified in this section.

A. Creating Dataset

The images used in this study have been collected from a created imaging system at Huawei entrance and open source datasets [46-47]. Fig. 2 shows the created imaging system to take real world face images. The system was detailed in the Working Environment section.

Not all images in the open source datasets were used for training because of mislabeled images. The collected images with mask and without mask can be seen from Fig. 3 and Fig. 4, respectively.

Fig 3. Collected face images with mask. (a and b are from M2150 camera, c is from a mobile phone, d-j are from open-source datasets.)

Fig 4. Collected face images without mask. (a and b are from M2150 camera, c is from a mobile phone, d-j are from open-source datasets.)

With the system at Fig. 2 and open source datasets, 18400 training, 200 validation and 4540 test, which are collected after the training, images have been collected as shown at Table 1.

Table 1. Number of Train - Validation - Test Images

Type	Number of Image
Train	18400
Validation	200
Test	4540
Total	23140

B. Working Environment

To take real world face images a camera setup was built at Huawei Entrance (Fig. 2). The used camera is Huawei M2150-10-EI [48] which has 5MP image sensor, 1 TOPS computing power, 2560(H) x 1920(V) effective pixels as shown at Table 2. It can capture face images and send via File Transfer Protocol.

Table 2. Technical Specifications of the Camera

CPU	Hi3516D
Computing Power	1 TOPS
Intelligent Analysis	Face and Person Detection
Effective pixels	2560 (H) x 1920 (V)
Video Encoding Format	H.265/H.264/MJPEG
Frame Rate	30 FPS

To train the AI model, a GPU-based linux server (Centos 7, Cuda 11.2) that has Tesla T4 has been used. As can be seen at Table 3 Python is selected as programming language, Pytorch is selected as deep learning library, Resnet-18 is selected as pre-trained model.

Table 3. Used Hardware and Software

Programming Language	Python
Deep Learning Library	Pytorch
Transfer Learning Model	Resnet-18
GPU	Tesla T4
CUDA	11.2
Operation System	Linux Centos 7

IV. METHODS

This part describes how to classify images and create class activation maps which refer to highlighted areas of the images.

A. Convolutional Neural Networks (CNNs)

Convolutional Neural Network is one of the well-known deep learning models which is special for image-related problems such as image classification, object detection, and image segmentation [49].

CNNs consist of convolutional and neural structures being responsible for automatically extracting and classifying important features of given images [50]. For instance edge, corner and pattern features are important features for images. A CNN architecture can be seen in the Fig. 5.

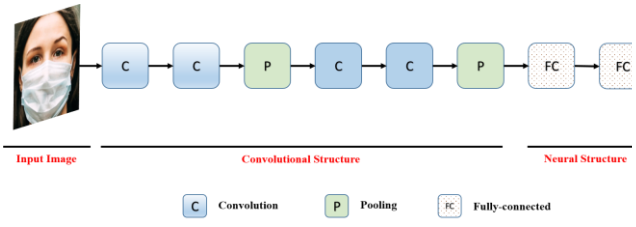

Fig 5. Simple CNN Architecture including convolution, pooling and fully-connected layers.

As it is seen in the Fig 5, the convolutional structure includes convolution and pooling layers that are feature extraction and dimension reduction methods, respectively. After features were extracted and reduced, they are classified by neural layer which is also known as fully-connected layer.

When it compares to traditional learning methods, end-to-end learning can be realized by using CNNs architecture. The comparison of CNNs and traditional learning is demonstrated at the Table 4.

Table 4. Comparison of CNNs and Traditional Learning

Operations	CNNs	Traditional Learning
Feature Extraction	Convolution	Local Binary Pattern
Dimension Reduction	Pooling	Linear Discriminant Analysis
Classification	Neural Layer	Artificial Neural Network

As seen at Table 4, in traditional methods all features should be extracted by using algorithms such as Local Binary Pattern [51] then reduced by using embedded or filter based algorithms such as Linear Discriminant Analysis [52].

After reduction of the features, a classifier such as Artificial Neural Network or Support Vector Machine should be used to classify the features into classes desired [53]. When it comes to CNNs, all operations are realized by convolution, pooling and neural layer automatically.

B. Transfer Learning

Training a deep learning model from scratch causes computation cost, and requires hundreds of labelled images. Transfer learning is a method which uses pre-trained models and modify them according to application to avoid starting a learning process from scratch and train with comparatively little data [54].

Transfer Learning uses pre-trained models which are trained with millions of labelled images of ImageNet dataset [55]. There are various pre-trained models such as VGG16 [56], ResNet [57], Inception [58]. Some differences of the models are depth, filter sizes, connections of the layers, and activation functions.

In this study a ResNet architecture called ResNet-18 is selected as CNN model that is 18 layers deep. As mentioned before, ResNet is a pre-trained model which is trained with ImageNet dataset which has 1000 classes such as mouse, desk, lemon, and pizza. To change its classes and re-train the model operations in Algorithm 1 are used. (Fig. 6 presents the transfer learning process.)

Algorithm 1:

- Freezing convolutional layers
- Removing the last layer with 1000 classes
- Adding new neural layer with classes desired
- Training

ResNet-18 accepts images with 224x224 sizes. Since pre-trained models have specific input sizes, all input images (Fig. 6 Part 1) need to be resized.

When first layers of trained models' are detailed it is seen that low level features such as corner, edge and curve are learned by the models. Instead training the first layers (Fig. 6 Part 2, 3, 4, 5) again, they are frozen and other (Fig. 6 Part 6, 7, 8, 9) layers are trained.

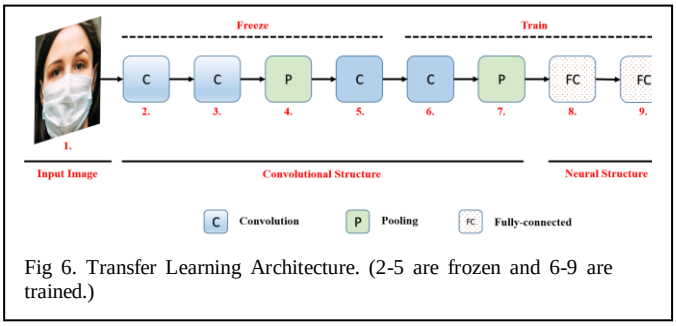

Fig 6. Transfer Learning Architecture. (2-5 are frozen and 6-9 are trained.)

Since the pre-trained models are trained with 1000 classes, their last layer should be changed according to number of class desired. To change the last layer (Fig. 6 Part 9), last layer is removed and a new neural layer is added. After model was modified, the training is started.

C. Class Activation Maps

Class Activation Map is a method which allows us to understand classification processes by creating a heat-map over input images after training was completed. Two class activation maps obtained from this study can be seen at Fig. 7.

In this study, gradient-weighted class activation mapping [59] is used to create heat-maps because it uses the gradients of classes by starting from the last neural layer to final convolutional layer. Therefore, without making any change in the trained model, important regions in the input image are highlighted.

The overall operations to obtain the heat-maps can be seen in the Algorithm 2.

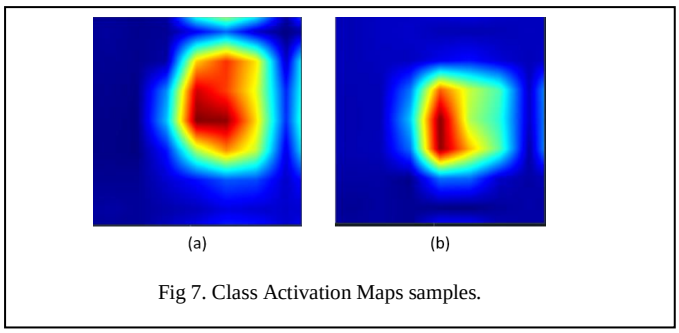

Fig 7. Class Activation Maps samples.

Algorithm 2:

- Resizing image to (224x224)
- Classifying the image
- Getting weights between global average pooling and neural layer
- Reshaping the last convolutional layer from (7x7, 512) to (49, 512)
- Dot product of the weights and the convolutional layer
- Reshaping from (1x49) to (7x7)
- Resizing from (7x7) to (224x224)
- Overlaying the input image and heat-map

To detail how CAMs is working Fig. 8 has been created. This figure shows the last three layer of the modified ResNet-18 architecture. It has 512 pieces 7x7 convolution filters in the last convolutional layer (Fig. 8 Part 1). By applying the global average pooling operation, its size converted from (7x7, 512) to (1x1, 512) (Fig. 8 Part 2). These filters with (1x1, 512) are classified by the neural layer (Fig. 8 Part 3) with 2 neurons that are responsible for mask and no mask classes.

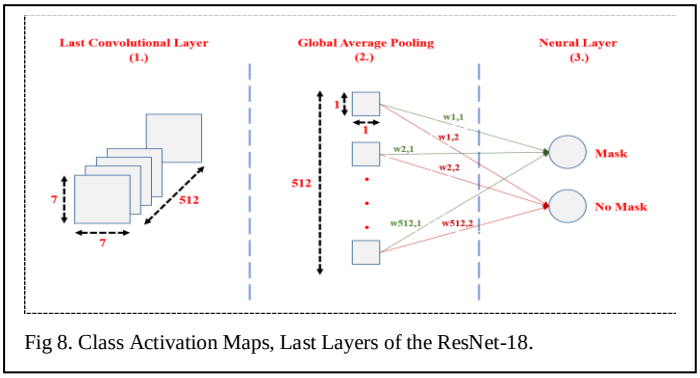

Fig 8. Class Activation Maps, Last Layers of the ResNet-18.

As shown at Algorithm 2, an input image is resized to (224x224) to be classified. There are (1x512, 2) weights between the global average pooling and neural layer (Fig. 8 Part 2 and Part 3). If classification result is about to 1 that

means "mask", the weights between the global average pooling and "mask" neuron are taken (w1,1; w2,1; ... w512,1), otherwise the weights between the global average pooling and "no mask" neuron are taken (w1,2; w2,2; ... w512,2).

After the (1x512) weights were obtained, (7x7, 512) filters in the last convolutional layer (Fig. 8 Part 1) is converted to (49x512). Dot product of (1x512) size neural weights and (49x512) size filter weights is realized to obtain (1x49) size importance weights.

The obtained importance weights (1x49) are reshaped to (7x7) and then resized to (224x224). The (224x224) size map is called class activation map and used to overlay with the input image.

D. Metrics to Evaluate the Classification Results

A classification process is generally evaluated by accuracy (Eq. 1), sensitivity (Eq. 2) and specificity (Eq. 3) metrics.

CCIWM + CCIWOM / IWM + IWOM (1)

CCIWOM / IWOM (2)

CCIWM / IWM (3)

Where *IWM* stands for images with mask, *IWOM* stands for images without mask, *CCIWM* stands for correctly classified images with mask, *CCIWOM* stands for correctly classified images without mask.

Accuracy gives information about how all images are correctly classified but when balanced test set is not available sensitivity and specificity metrics are used. Sensitivity measures how positive (no mask) class, specificity measure how negative (mask) class is correctly classified.

V. EXPERIMENTS AND RESULTS

This section describes the all experiments and results by taking into account hyper-parameters, training, validation, testing, and class activation maps. The overall process can be seen from Fig. 9.

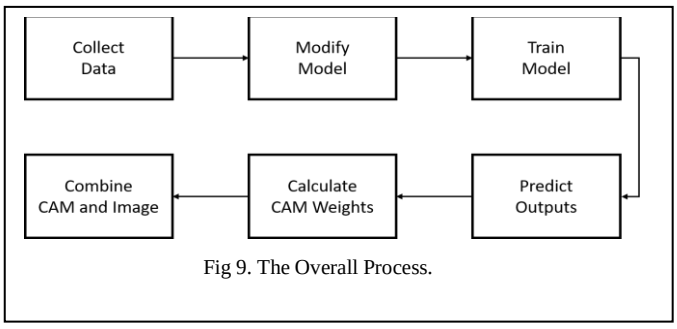

Fig 9. The Overall Process.

As seen in Fig. 9, first process is collection of images. To collect the images two different sources are used which are a system built at the Huawei entrance and open-source datasets. By using the system at the entrance real-world images were collected to develop a system that works under hard conditions such as low resolution, darkness, brightness, and image size. By combining with the open-source datasets, totally 23140 images (18400 train, 200 validation, 4540 test) were collected as shown at Table 1.

Pre-trained models such as VGG16, and ResNet18 have ability to classify images with 1000 classes. Since there exists two classes as mask and no mask in this study, ResNet-18 model was modified to classify the images with two classes. To do that, its 1000-classes last neural layer was removed and a new 2-classes neural layer was added to the ResNet-18 model.

The ResNet-18 model was re-trained with the 18400 train and 200 validation images for 30 epochs. The used parameters can be seen at the Table 5.

Table 5. Parameters To Be Used For Training and Number of Images

Model	Resnet-18
Loss Function	Cross Entropy
Optimiser	Stochastic Gradient Descent
Learning Rate	0.001
Momentum	0.9
Epoch	30

After the training, 4540 new images (4230 with mask, 310 without mask) were taken from the system at the Huawei entrance to test the model with real-world images. According to the classification results;

- 4090 of 4230 images with mask (96.69%)
- 295 of 310 images without mask (95.16%) were correctly classified as shown at Table 6.

Table 6. All Results

Images With Mask	4230
Images Without Mask	310
Correctly Classified Images With Mask	4090
Correctly Classified Images Without Mask	295
False Classified Images With Mask	140
False Classified Images Without Mask	15
Sensitivity	95.16%
Specificity	96.69%
Accuracy	96.58%

Detected face images with and without mask can be seen from Fig. 10 and Fig. 11 respectively.

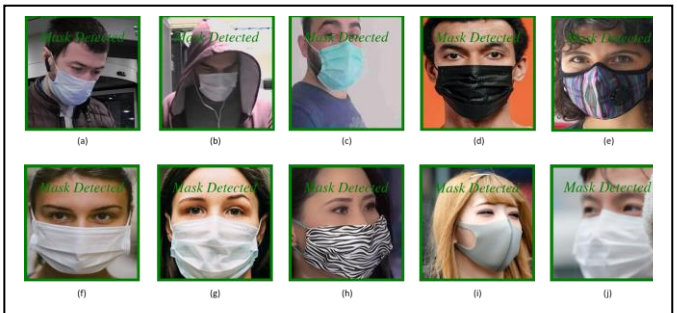

Fig 10. Classified face images with mask. (a and b are from M2150 camera, c is from a mobile phone, d-j are from open-source datasets.)

Fig 11. Classified face images without mask. (a and b are from M2150 camera, c is from a mobile phone, d-j are from open-source datasets.)

The Gradient-weighted Class Activation Maps need class-related weights. Since ResNet-18 has 512 pieces filter with 1x1 size in global average pooling layer (Fig. 8 Part 2) and two neuron in neural layer (Fig. 8 Part 3), there are 2 pieces 1x512 weights related to mask and no mask classes. After classification result was obtained, the weights of the desired class is taken. For example, if mask class's activations are looked for mask class's weights are taken (size is 1x152). Class activation maps of the mask and no mask classes are shown in Fig. 12 and Fig. 13.

Fig 12. Classified and Highlighted face images with mask. (a-j are from a webcam.)

As can be seen from Fig. 12 images are correctly classified as mask and the region where the mask exists is highlighted and followed by class activation maps. When the mask is removed, images are still correctly classified and class activation maps highlight and follow the face over the image (Fig. 13).

Fig 13. Classified and Highlighted face images without mask. (a-j are from a webcam.)

VI. DISCUSSION

When a classification is realized using deep neural networks such as CNNs, classification results are obtained with high performance but why the model classified into a class in the classification are not explained because of complexity of deep neural networks. That's why such models are called black-box models.

To explain what a CNN learn from images with and without face masks, ResNet-18 pre-trained model was selected as a CNN model and re-trained with our images, and class activation maps technique was applied to images to highlight regions learned by the model.

According to learning results with test images, sensitivity and specificity values were obtained as 95.16% and 96.69% respectively. The results show that the CNN model can successfully classifies images into mask and no mask classes, and when we ask why an image was classified into mask class, it answers by highlighting the region where the mask image exists. The importance of that is we can be sure there was no overfitting and the CNN model was correctly trained.

VII. CONCLUSION AND FUTURE WORKS

Explainable Artificial Intelligence is a relatively new topic in deep neural network models. When a model making a decision in a classification, it is not known why the model decided.

In this study, we are showing why face images are classified into classes as mask and no mask by highlighting class-related activation maps using the gradient-weighted class activation maps technique.

In future works, we are planning to grow the study by adding weakly supervised object detection techniques so that we would have ability to realize object detection without using a labeling application that causes time costs.

ACKNOWLEDGMENT

This study was supported by "Epidemic Prevention System" of Huawei Turkey R&D Center.

REFERENCES

[1]. A. Mikołajczyk and M. Grochowski, “Data augmentation for improvingdeep learning in image classification problem,” in2018 internationalinterdisciplinary PhD workshop (IIPhDW). IEEE, 2018, pp. 117–122.
[2]. C. Affonso, A. L. D. Rossi, F. H. A. Vieira, A. C. P. de Leon Ferreiraet al.,“Deep learning for biological image classification,”Expert Systems withApplications, vol. 85, pp. 114–122, 2017.
[3]. W. Zhao and S. Du, “Spectral–spatial feature extraction for hyperspectralimage classification: A dimension reduction and deep learning approach,”IEEE

Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp.4544–4554, 2016.
[4]. S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learningtechniques for image segmentation,"ACM Computing Surveys (CSUR),vol. 52, no. 4, pp. 1–35, 2019.
[5]. G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel,A. L. David, J. Deprest, S. Ourselinet al., "Interactive medical imagesegmentation using deep learning with image-specific fine tuning,"IEEEtransactions on medical imaging, vol. 37, no. 7, pp. 1562–1573, 2018.
[6]. S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, andD. Terzopoulos, "Image segmentation using deep learning: A survey,"IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
[7]. Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deeplearning: A review,"IEEE transactions on neural networks and learningsystems, vol. 30, no. 11, pp. 3212–3232, 2019.
[8]. A. R. Pathak, M. Pandey, and S. Rautaray, "Application of deep learningfor object detection,"Procedia computer science, vol. 132, pp. 1706–1717,2018.
[9]. J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learningtechniques for salient and category-specific object detection: a survey,"IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 84–100, 2018.
[10]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradientbased localization," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.
[11]. N. C. Brienen, A. Timen, J. Wallinga, J. E. Van Steenbergen, and P. F. Teunis, "The effect of mask use on the spread of influenza during a pandemic," Risk Analysis: An International Journal, vol. 30, no. 8, pp. 1210–1218, 2010.
[12]. M. F. Uddin, J. Lee, S. Rizvi, and S. Hamada, "Proposing enhanced feature engineering and a selection model for machine learning processes," Applied Sciences, vol. 8, no. 4, p. 646, 2018.
[13]. L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). IEEE, 2018, pp. 80–89.
[14]. L. E. Raileanu and K. Stoffel, "Theoretical comparison between the gini index and information gain criteria," Annals of Mathematics and Artificial Intelligence, vol. 41, no. 1, pp. 77–93, 2004.
[15]. L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, "Serial quantitative chest ct assessment of covid-19: a deep learning approach," Radiology: Cardiothoracic Imaging, vol. 2, no. 2, p. e200075, 2020.
[16]. Y.Oh,S.Park,andJ.C.Ye,"Deeplearningcovid-19featuresoncxrusing limitedtrainingdatasets,"IEEEtransactionsonmedicalimaging,vol.39, no. 8, pp. 2688–2700, 2020.
[17]. E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," arXiv preprint arXiv:2003.11055, 2020.
[18]. A. M. Ismael and A. S¸engür, "Deep learning approaches for covid-19 detectionbasedonchestx-rayimages,"ExpertSystemswithApplications, vol. 164, p. 114054, 2021.
[19]. A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks, "Computers in biology and medicine,vol.121,p.103795,2020.
[20]. Q. Ni, Z. Y. Sun, L. Qi, W. Chen, Y. Yang, L. Wang, X. Zhang, L. Yang, Y. Fang, Z. Xing et al., "A deep learning approach to characterize 2019 coronavirus disease (covid-19) pneumonia in chest ct images," European radiology, vol. 30, no. 12, pp. 6517–6527, 2020.
[21]. M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic," Measurement, vol. 167, p. 108288, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224120308289
[22]. B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolutionwithclassificationnetworktopreventcovid-19,"Sensors, vol. 20, no. 18, p. 5236, 2020.
[23]. M. S. Ejaz, M. R. Islam, M. Sifatullah, and A. Sarker, "Implementation of principal component analysis on masked and non-masked face recognition," in 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT). IEEE, 2019, pp. 1–5.
[24]. E. Hjelmås and B. K. Low, "Face detection: A survey," Computer vision and image understanding, vol. 83, no. 3, pp. 236–274, 2001.
[25]. A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," Artificial Intelligence Review, vol. 52, no. 2, pp. 927–948, 2019.
[26]. C. Li, R. Wang, J. Li, and L. Fei, "Face detection based on yolov3," in Recent Trends in Intelligent Computing, Communication and Devices. Springer, 2020, pp. 277–284.
[27]. X. Sun, P. Wu, and S. C. Hoi, "Face detection using deep learning: An improved faster rcnn approach," Neurocomputing, vol. 299, pp. 42–50, 2018.
[28]. W. Wu, Y. Yin, X. Wang, and D. Xu, "Face detection with different scales based on faster r-cnn," IEEE transactions on cybernetics, vol. 49, no. 11, pp. 4017–4028, 2018.
[29]. R. Qi, R.-S. Jia, Q.-C. Mao, H.-M. Sun, and L.-Q. Zuo, "Face detection method based on cascaded convolutional networks," IEEE Access, vol. 7, pp. 110740–110748, 2019.

[30]. C. Li, R. Wang, J. Li, and L. Fei, "Face detection based on yolov3," in Recent Trends in Intelligent Computing, Communication and Devices. Springer, 2020, pp. 277–284.
[31]. J.RedmonandA.Farhadi,"Yolov3:Anincrementalimprovement,"arXiv preprint arXiv:1804.02767, 2018.
[32]. X. Kong, K. Wang, S. Wang, X. Wang, X. Jiang, Y. Guo, G. Shen, X. Chen, and Q. Ni, "Real-time mask identification for covid-19: an edgecomputing-baseddeeplearningframework,"IEEEInternetofThings Journal, 2021.
[33]. W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, and S. Zhang, "Towards rich feature discovery with class activation maps augmentation for person re-identification," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
[34]. Z. Dai, M. Chen, X. Gu, S. Zhu, and P. Tan, "Batch dropblock network for person re-identification and beyond," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3691–3701.
[35]. S. Yang, Y. Kim, Y. Kim, and C. Kim, "Combinational class activation maps for weakly supervised object localization," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), March 2020.
[36]. V. Gupta, M. Demirer, M. Bigelow, M. Y. Sarah, S. Y. Joseph, L. M. Prevedello, R. D. White, and B. S. Erdal, "Using transfer learning and class activation maps supporting detection and localization of femoral fracturesonanteroposteriorradiographs,"in2020IEEE17thI nternational Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 1526–1529.
[37]. W. Bae, J. Noh, and G. Kim, "Rethinking class activation mapping for weakly supervised object localization," in European Conference on Computer Vision. Springer, 2020, pp. 618–634.
[38]. J. Cai, F. Xing, A. Batra, F. Liu, G. A. Walter, K. Vandenborne, and L. Yang, "Texture analysis for muscular dystrophy classification in mri with improved class activation mapping," Pattern recognition, vol. 86, pp. 368–375, 2019.
[39]. K. Fu, W. Dai, Y. Zhang, Z. Wang, M. Yan, and X. Sun, "Multicam: Multiple class activation mapping for aircraft recognition in remote sensing images," Remote Sensing, vol. 11, no. 5, p. 544, 2019.
[40]. B. Vasu, F. U. Rahman, and A. Savakis, "Aerial-cam: Salient structures and textures in network class activation maps of aerial imagery," in 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP). IEEE, 2018, pp. 1–5.
[41]. Y. Wang, F. Zhu, C. J. Boushey, and E. J. Delp, "Weakly supervised food image segmentation using class activation maps," in 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 1277–1281.
[42]. H.-G. Nguyen, A. Pica, J. Hrbacek, D. C. Weber, F. La Rosa, A. Schalenbourg, R. Sznitman, and M. B. Cuadra, "A novel segmentation framework for uveal melanoma in magnetic resonance imaging based on class activation maps," in International Conference on Medical Imaging with Deep Learning. PMLR, 2019, pp. 370–379.
[43]. Y. Zhu, Y. Zhou, H. Xu, Q. Ye, D. Doermann, and J. Jiao, "Learning instance activation maps for weakly supervised instance segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3116–3125.
[44]. M. B. Muhammad and M. Yeasin, "Eigen-cam: Class activation map using principal components," in 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020, pp. 1–7.
[45]. F. I. Eyiokur, H. K. Ekenel, and A. Waibel, "A computer vision system to help prevent the transmission of covid-19," arXiv preprint arXiv:2103.08773, 2021.
[46]. W. Intelligence, "Face Mask Detection Dataset," www.kaggle.com/wobotintelligence/face-mask-detection-dataset, 2021.
[47]. A. Jangra, "Face Mask Detection," www.kaggle.com/ashishjangra27/facemask-12k-images-dataset, 2021.
[48]. Huawei, "Huawei M2150," support.huawei.com/enterprise/en/intelligentvision/m2150-10-ei-pid-250673491, 2021.
[49]. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60–88, 2017.
[50]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.
[51]. Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," IEEE transactions on image processing, vol. 19, no. 6, pp. 1657–1663, 2010.
[52]. A. J. Izenman, "Linear discriminant analysis," in Modern multivariate statistical techniques. Springer, 2013, pp. 237–280.
[53]. A. Tzotsos and D. Argialas, "Support vector machine classification for object-based image analysis," in Object-Based Image Analysis. Springer, 2008, pp. 663–677.
[54]. C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in International conference on artificial neural networks. Springer, 2018, pp. 270–279.
[55]. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009, pp. 248–255.
[56]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556
[57]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CoRR, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[58]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," CoRR, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[59]. R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," CoRR, vol. abs/1610.02391, 2016. [Online]. Available: http://arxiv.org/abs/1610.02391