*Article*

# Robust Blocking of Human Faces with Personal Information Using Artificial Deep Neural Computing

**Seok-Woo Jang [1] and Sang-Hong Lee [2],***

[1]  Department of Software, Anyang University, 22, 37-Beongil, Samdeok-Ro, Manan-Gu, Anyang 430-714, Korea; swjang7285@gmail.com

[2]  Department of Computer Engineering, Anyang University, Anyang University, 22, 37-Beongil, Samdeok-Ro, Manan-Gu, Anyang 430-714, Korea

*  Correspondence: shleedosa@anyang.ac.kr; Tel.: +82-31-467-0842

check for updates

**Abstract:** High-speed wired and wireless Internet are one of the useful ways to acquire various types of media data easily. In this circumstance, people also can easily get media data including objects with exposed personal information through the Internet. Exposure of personal information emerges as a social issue. This paper proposes an effective blocking technique that makes it possible to robustly detect target objects with exposed personal information from various types of input images with the use of deep neural computing and to effectively block the detected objects' regions. The proposed technique first utilizes the neural computing-based learning algorithm to robustly detect the target object including personal information from an image. It next generates a grid-type mosaic and lets the mosaic overlap the target object region detected in the previous step so as to effectively block the object region that includes personal information. Experimental results reveal that the proposed algorithm robustly detects the target object region with exposed personal information from a variety of input images and effectively blocks the detected region through grid-type mosaic processing. The object blocking technique proposed in this paper is expected to be applied to various application fields such as image security, sustainable anticipatory computing, object tracking, and target blocking.

**Keywords:** neural computing; target blocking algorithm; image security; media data; activation function; feature acquisition

## 1. Introduction

Since smart devices with good performance were penetrated, since high-speed wired and wireless network technologies were developed, and since small-sized display devices with good image quality were developed, people have easily acquired various types of media content, such as pictures, animation, music files, user created contents (UCC), and videos, at any time and any place through the Internet [1–5]. Such acquired media content is usefully applied to various application fields, including big data analysis, image security, deep learning-based artificial intelligence, sustainable anticipatory computing, media content indexing and retrieval, and Internet of Things (IoT) [6–12].

Unfortunately, media content that includes objects with exposed personal information, such as a human face, a resident registration number, a telephone number, a license plate, or an address, can easily be acquired and distributed through the Internet so that the problem is a big social issue [13,14]. In particular, since such media content is spread fast, people whose personal information is exposed severely suffer psychological damage.

Accordingly, one of the recent subjects of research is how to detect accurately the corresponding target object including exposed personal information from various types of still or video input content first; then to effectively block the detected target object through image processing-based blurring effects,

grid-type mosaic processing, and virtual object insertion; and thereby to effectively hide the exposed critical personal information [15].

As presented in the references, there are previous studies on how to detect a target object from a color image and to block the detected object. In one study [16], color features were analyzed in the space of $YC_bC_r$ and the regions presenting human skin color distribution were detected from an input color image. Subsequently, whether the detected regions included critical naked human body regions was judged with the use of the geometrical features of the detected skin color regions. In another study [17], a fuzzy clustering technique was applied to extract a block-based mosaic region. This technique is used to extract an edge from an input color image, to obtain clustering features, and then to find a general image region and a mosaic region using a fuzzy clustering algorithm. In Reference [18], a novel Bayesian framework was applied to detect a human facial region in a fast occluding environment. This algorithm utilized the omega shape formed by the human head and shoulders in order to deal with occlusion of the head effectively. In a test, this algorithm detected a facial region in a relatively accurately occluding environment. In Reference [19], a new proposal generation acceleration framework was suggested in order to detect the human face in real time with the use of global and local features. In the method, the convolution neural network cascade was used as a baseline and the acceleration scheme for accelerating an inference time was developed. Aside from the techniques described above, object detection and blocking methods continue to be introduced [20].

However, most of the existing object detection and blocking approaches are not yet very complete and include many internal and external constraints. Moreover, studies to detect and block a target object with personal information are relatively few in number compared to studies on other technical methods. Therefore, this paper proposes a new algorithm that makes it possible to detect a target object region with personal information on the basis of deep neural computing and to effectively block the detected object through grid-type mosaic processing. Figure 1 illustrates an overview of the proposed target object blocking technique.
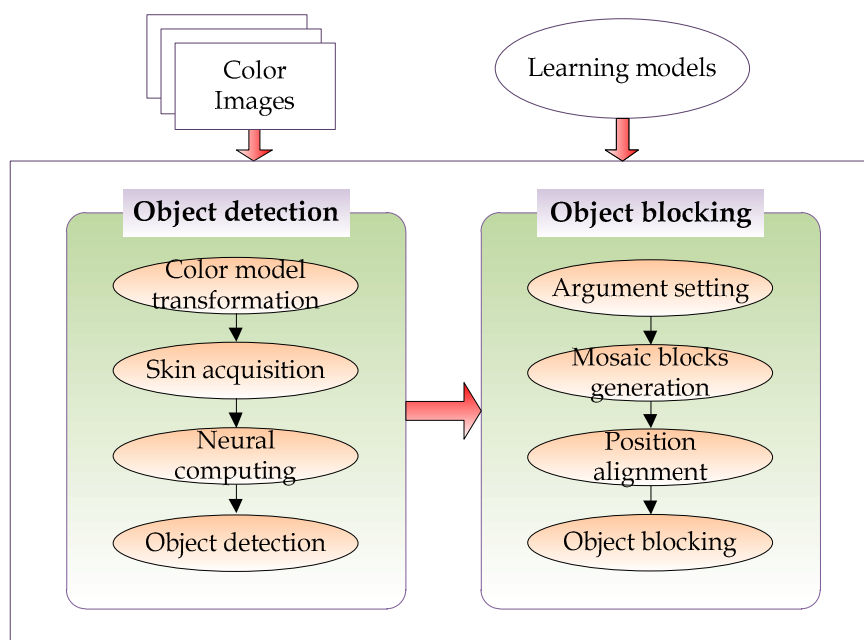


**Figure 1.** Overall flow of the proposed blocking system.

As shown in Figure 1, the algorithm proposed in this study first utilizes the artificial neural network-based deep learning algorithm to robustly detect a target object region with personal information like human facial regions and naked body parts from various types of input color images. Our algorithm then generates grid-like mosaic blocks and lets the generated mosaic blocks overlap the

detected object region naturally so as to hide the target object region with personal information from being exposed to others.

In Section 1, the overview and background of this study are explained. In Section 2, previous studies on target object blocking in the fields of pattern recognition and image processing are described. In Section 3, the technique of detecting an object with personal information from an input color image robustly on the basis of artificial neural computing is described. In Section 4, the technique of generating a grid-like mosaic block and then letting the generated mosaic region overlap the detected target object region in order to cover the object effectively is explained. In Section 5, we show the experimental results for comparative evaluation of the performance of the proposed object blocking algorithm. In Section 6, the conclusion of this study and the future research plan are described.

## 2. Related Work

With the development of small- and high-quality display devices, the penetration of excellent mobile devices, the lowered prices of large storage devices, and the development of high-speed wired and wireless network technologies like 5th generation mobile telecommunication, it has been easy to obtain various kinds of media content like pictures or videos. In this circumstance, media content including target objects with exposed personal information, such as the human face or a resident registration number, is freely distributed through the Internet, which is a big social issue. For this reason, it has been requested to develop a novel technique of detecting and blocking an object region with exposed personal information from a variety of input color images effectively. Many different studies on target object detection and blocking, which are presented in references, are described as follows:

The color feature distribution modeling-based method [16] introduces a new algorithm to identify an adult image in order to filter naked human body images in the Internet effectively. This algorithm first detects the eyes, which are known to be one of the most noticeable components of the human face, and selects a reliable skin color sample pixel in the facial region around the detected eyes in order to make a statistical skin color distribution model from each input color image directly. The skin color region of the whole image is robustly segmented with the use of the online skin color distribution model. The method then extracts a series of typical features such as size, aspect ratio, and compactness representing the naked body from the segmented skin regions and checks the regions including the naked body regions through multi-layer perceptron (MLP) neural network-based learning and inference of typical features. The proposed adult image detection method provides experimental results in order to prove that it is able to identify various types of nude images effectively more than other conventional methods. In the method, although the skin color extraction algorithm is not relatively complicated and is much used, it can be somewhat sensitive to changing surroundings like noise or lighting conditions. Figure 2 shows the flowchart of the color distribution model-based method.

The clustering-based method [17] utilized a fuzzy clustering algorithm to extract a block-based mosaic region from a color image accurately. In general, clustering is usually the process of grouping observations with similar attributes to classify the entire data into several groups. With the development of broadcasting and video network, a digital video broadcast monitoring system has been of more importance. To detect a low-quality video data in a television monitoring system, a digital image recognition technology is widely applied. Usually, a mosaic signal is easily generated in TV signals and lowers the video quality. A conventional block-based mosaic detection algorithm is unable to differentiate a Chinese character from a mosaic block. To solve the problem, the method detects a mosaic block through the fuzzy C-means clustering algorithm. It has three major steps. In the first step, a frame is processed by the Sobel edge detection algorithm in order to obtain an edge image. In the second step, a fuzzy C-means clustering algorithm extracts clustering features. In the last step, a mosaic block is effectively differentiated from all general image blocks through fuzzy C-means clustering. Experimental results revealed that the proposed algorithm detected mosaic blocks relatively robustly.
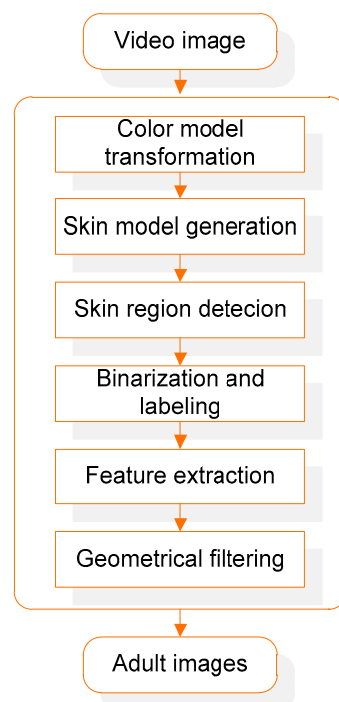
**Figure 2.** Flowchart of the color distribution model-based method.

In the Bayesian framework-based method [18], a human facial region was quickly detected in an occluding environment. Crimes related to automatic teller machines (ATM) draw more attention, and criminals intentionally hide their faces to avoid their identity. The method utilizes a fast and robust facial occlusion detection algorithm for robust ATM monitoring and was proved to be effective and efficient in processing an occluded face. The algorithm applied the omega shape formed by the human head and shoulders in order to deal with the head's occlusion effectively. More specifically, a new energy function for detecting the elliptical head outline is first designed. Secondly, the fast and robust head tracking algorithm using gradient and shape signals in a Bayesian framework is newly developed. Lastly, in order to check if the detected head is occluded, the AdaBoost algorithm is applied to integrate information on the skin color distribution and human facial structure. Experimental results with actual data showed that the proposed algorithm had relatively high accuracy of facial detection and of facial occlusion detection. However, there is some false detection in the situation of suddenly changing lighting. Figure 3 shows the block diagram of the Bayesian framework-based method.
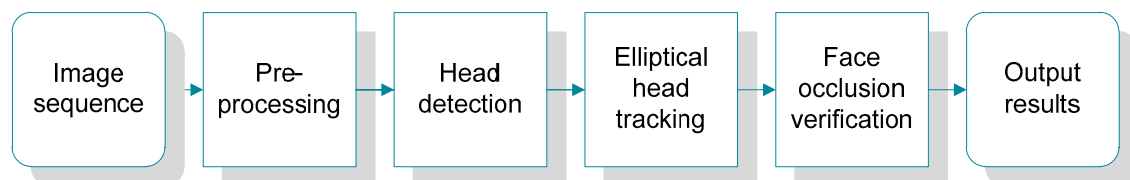
**Figure 3.** Block diagram of the Bayesian framework-based method.

In the global and local features-based method [19], a proposal generation acceleration framework was newly suggested for real-time human facial region detection. Human facial recognition has widely been researched over the last several tens of years. The artificial deep neural network solution has greatly improved it. However, it is hard to directly apply the deep neural network technology to mobile devices due to limited calculation and memory. This method uses the convolution neural network cascade as a baseline, and the acceleration system for accelerating an inference time is developed. In the acceleration scheme, the calculation bottleneck of the baseline occurs in the proposal generation

step, which was derived from the observation that each level of high-density image pyramid needs to pass the network. To solve the problem, the proposed method utilizes local facial features, for example, all of the global face and facial regions, to reduce the number of levels of the image pyramid. In the test with opened facial data sets, the proposed scheme was found to detect the face quickly. Nevertheless, there are several constraints to resolve. Introducing facial parts for classification requires extra data collection. The training of the proposal network also becomes harder since the amount of data in each category is likely to be unbalanced.

Aside from the methods described earlier, many new methods related to how to detect a target object with personal information continue to be introduced [20]. Most of the conventional methods utilize human skin color distribution modeling in order to detect the human face robustly. Although such existing methods using human skin color distribution as the main feature are simple, they have a certain limitation in detecting a facial region from the input color images taken in diverse situations by using skin color regions only. In addition, conventional methods have yet to be perfect and include many internal and external constraints. Furthermore, there are relatively fewer studies on how to detect and block a target object with personal information than studies on other research methods.

## 3. Target Region Detection

### 3.1. Detection of Skin Color Distribution Regions

In this study, the human facial region is a main target of personal information. In this section, how to robustly detect a target object with personal information from dynamic input color images is described. In the method, an input image is first analyzed, then background regions are removed, and only the foreground skin color regions are detected.

In this study, Equation (1) is applied to change an image of the *RGB* color space to an image of the $YC_bC_r$ color space. Generally, the $YC_bC_r$ color distribution space is known to be easier to extract a human skin color region than other color spaces [21–24].

$$
\begin{aligned}
Y &= 16 + 65.481 \times R + 128.533 \times G + 24.966 \times B \\
C_b &= 128 - 37.797 \times R - 74.203 \times G + 112 \times B \\
C_r &= 128 + 122 \times R - 93.786 \times G - 84.214 \times B
\end{aligned}
\tag{1}
$$

Next, the elliptical human skin color distribution model defined through pre-learning, as shown in Equation (2), is applied to robustly extract skin color distribution pixels except for other background regions from an input color image [25–27].

$$
\frac{(x-ec_x)^2}{a^2} + \frac{(y-ec_y)^2}{b^2} = 1
$$
$$
\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b' - c_x \\ C_r' - c_y \end{bmatrix}
\tag{2}
$$
$$
\overline{C_b} = \frac{1}{K}\sum_{t=0}^{K-1} C_b(t), \ \overline{C_r} = \frac{1}{K}\sum_{t=0}^{K-1} C_r(t)
$$

In Equation (2), $C_b$ and $C_r$ represent the $C_b$ color channel value and $C_r$ color channel value of an input color image, respectively. $\theta$ represents the rotated angle of an ellipse as radian. $a$ and $b$ mean the long axis and short axis of an ellipse. $ec_x$ and $ec_y$ are constant values for correcting the rotation error of an ellipse. $(\overline{C_b}(Y),\overline{C_r}(Y))$ means the centers of skin color distribution model. $t$ represents the index of each skin color distribution sample. $K$ is the factor representing the number of the selected skin samples. In this study, the elliptical model in the $C_b'C_r'$ space, where colors are independent on brightness as much as possible, is defined rather than the model in the $C_bC_r$ space. $a, b, cx, cy,$ $ec_x, ec_y,$ and $\theta$, which are parameters used in Equation (2), are extracted from the clustered skin color distribution regions in the $C_b'C_r'$ space through repeated testing.

In order to remove unnecessary noises located in the extracted skin color regions, morphological operation [28–31], which is widely used in the image processing fields, is executed. Subsequently, labeling [32–34] is executed to extract the skin color regions in which skin color pixels are connected in the unit of region. Through this learning, the main region of interest (ROI) for detecting the human face comes to be extracted.

### 3.2. Detection of Target Region

In this study, a convolutional neural network (CNN)-based deep learning technique [35–38], one of the significant machine learning algorithms [39–42] in the artificial intelligence field, is utilized to robustly detect only a human facial region from the extracted skin color distribution region. Generally, the CNN is one of the most popular structures in the deep learning research field due to its excellent image processing and pattern recognition performance. In the CNN model, two-dimensional data of a color image can be used as an input so that positional features in the input image are not lost. The CNN extracts representative features from two-dimensional data through the convolution layer and the pooling layer, and these features are used as the input of multi-layer artificial neural network.

In different machine learning prior to deep learning, it is required to obtain a proper feature map from an input image in order to process a color image as an input. The obtained feature map is used as the input of a machine learning algorithm. Accordingly, the work of generating a feature map greatly influences the overall learning performance often. Therefore, in terms of the CNN used in this paper, it is important to build a model that learns the convolution filter that extracts the best feature map in the preprocessing step because the preprocessing step to generate the feature map has a big impact on the performance of the system. In addition, in order to detect a target object fast from a real-time input color image, the image is received and learned in the structure of a pyramid in the learning step. The CNN-based deep learning model structure proposed in this study is illustrated in Figure 4.
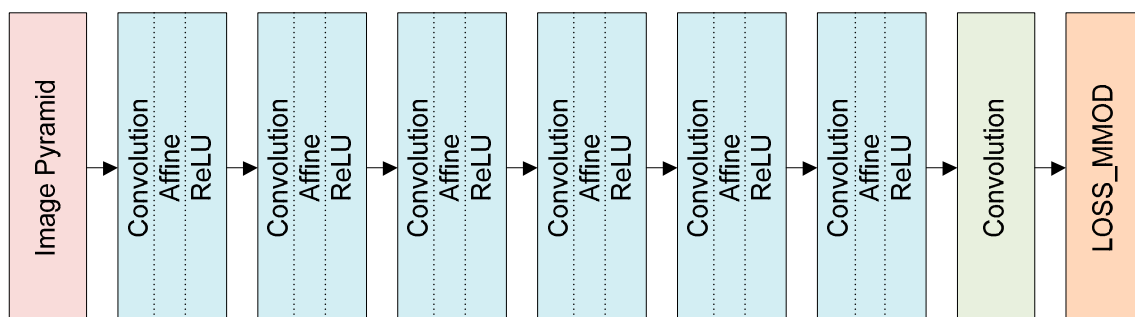


**Figure 4.** Structure of the proposed learning method: image pyramid (six levels), convolution (convolution layer), affine (affine layer), rectified linear unit (ReLU) (activation function), LOSS_MMOD (max-margin object detection).

The image pyramid is constructed as shown in Figure 5. The image pyramid in the proposed CNN-based algorithm consists of six steps. At the time of the last detection, the detection results of each pyramid step are added up so as to draw the final outcome. At this time, level 0 in the image pyramid represents the original image, and as the level goes up, the image resolution is down-sampled by 1/2n times. Since the face detection window is fixed in size, the pyramid step is set to six levels so that faces of various sizes can be detected. Using a two-stage image pyramid, the algorithm will be faster, but medium-sized faces can be difficult to detect. The image pyramid was used as an input for learning to detect faces of various sizes at high speed. In the image pyramid, images of each level are input and the learning proceeds.
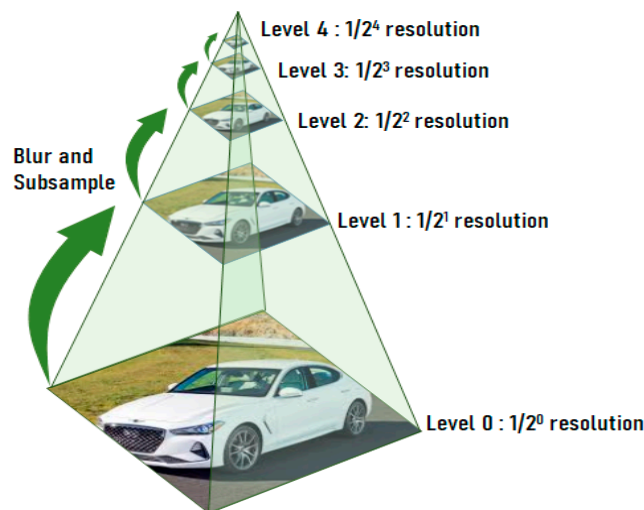
**Figure 5.** Construction of an image pyramid.

In Figure 4, the convolution layer is the layer of extracting features through convolution. Generally, in the image processing field, convolution is used for image filter operation. In other words, convolution is applied to implement a filter to extract particular features from an input color image. If a window or mask with $3 \times 3$ pixels or more is repeatedly applied to the whole image, it is possible to gain a proper result depending on the coefficient values of the mask. As shown in Figure 6, the yellow region of the whole image in the left side is the region that has convolution and a red number is the kernel of convolution. When mask operation is applied to the yellow region, its result is 4, as displayed in the right side. After the yellow window region is moved to the right by one pixel and the mask operation is applied, the result becomes 3. In this way, by moving the yellow region and by executing the mask operation repeatedly, it is possible to gain the final outcome. In this paper, the mask was randomly initialized to have a Gaussian distribution. The values of the mask were then determined to minimize the loss through iterative learning.
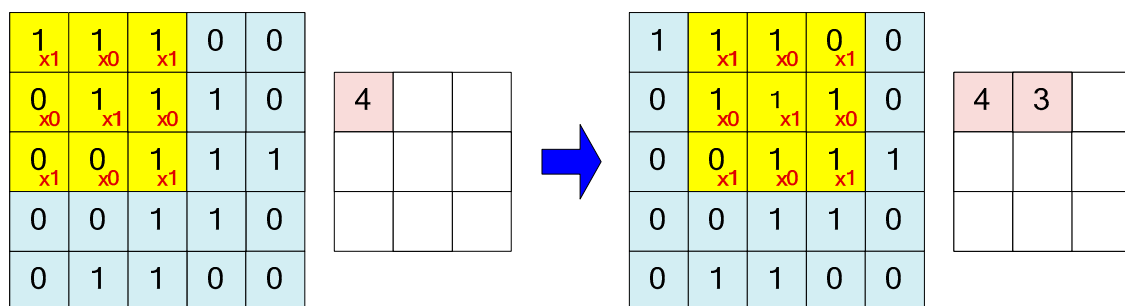


**Figure 6.** An example of convolution operation.

In this study, as an activation function, the rectified linear unit (ReLU) function [43–46] rather than the Sigmoid function is used as shown in Figure 7. That is because a vanishing gradient (in which a gradient converges to zero) occurs if the Sigmoid function is used [47–49]. In other words, the Sigmoid function has a value between 0 and 1, as shown in Equation (3). The deeper the artificial neural network layer is, the smaller its weight is. For this reason, the values delivered disappear.
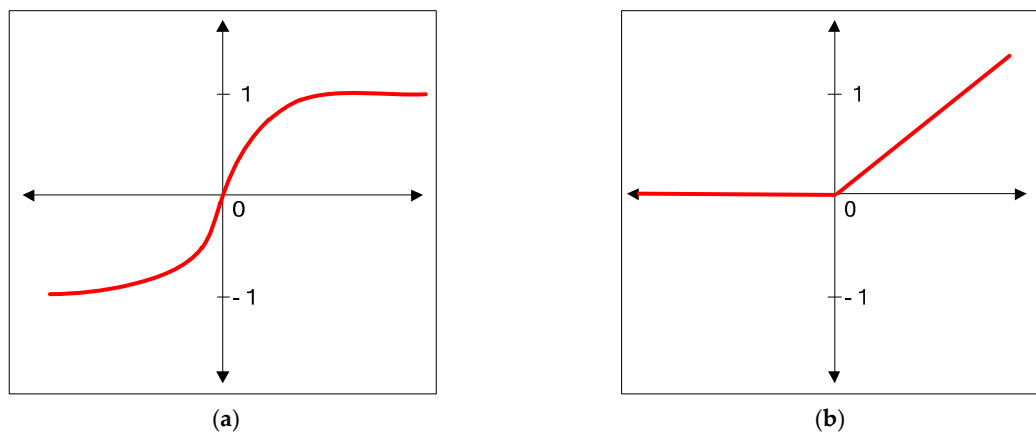
(a)  (b)

**Figure 7.** Activation function: (**a**) Sigmoid Function and (**b**) ReLU function.

As shown in Equation (4), in the ReLU function, if an input value is 0 or less, all values are set to 0 and, if it is larger than 0, it is set to the maximum value. Therefore, the ReLU function shows very efficient calculation and its convergence speed is faster than that of other activation functions. For this reason, the ReLU activation function tends to be used much in the deep learning field these days.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

$$ReLU(x) = \max(0, x) \tag{4}$$

In summary, the ReLU function has the following advantages: First, the partial can be activated by outputting zero for an input below zero. Second, there is no gradient vanishing and no gradient is exploding. Third, its derivative is simple because it is a linear function. Fourth, as shown in Equation (5), the ReLU function is scale-invariant.

$$\max(0, ax) = a \cdot \max(0, x) \tag{5}$$

In the proposed CNN model structure, the affine layer means the fully connected layer of an artificial neural network. In short, affine represents that the neurons of the previous layer are connected with the neurons of the present layer. An affine layer is added before the final prediction is made in the uppermost output layer of CNN or RNN (recurrent neural network) [50,51]. Generally, an affine layer is presented with $y = f(Wx+b)$, where $x$ is an input layer, $W$ is a weight, $b$ is a bias, and $f$ is a nonlinear activation function. In this paper, the model variables learned in the training phase are weights $W$ (convolution filter) and bias $b$.

In most target object detection methods, binary classification is applied to the sub-window of an image. In general, the classifier is trained in a sub-window because the number of possible sub-windows is very large. In this case, if all sub-windows are learned, it is possible to avoid computational difficulties but it is impossible to show the optimal performance.

In the proposed CNN model structure, MMOD (max-margin object detection) is able to optimize all sub-windows without subsampling. In short, $F(x, y)$, the window scoring function, is applied to all sub-windows in order for object detection [52,53]. In Equation (6), $\phi$ is the feature vector extracted from the moving window position $r$ of the image $x$. In Equation (6), $w$ is a weight vector, playing a role to reduce wrong detection through learning. By applying Equation (6), it is possible to obtain the result for the detection score. The final target object detection position can be obtained in the way of adding up the regions with the highest score in the pyramid image.

$$y^* = \mathrm{argmax}_{y \in Y} F(x, y) = \mathrm{argmax}_{y \in Y} \sum_{r \in y} < w, \phi(x, r) > \tag{6}$$

## 4. Blocking through Mosaic Processing

In general, digital content makers sometimes generate a grid-type mosaic region and insert it into a variety of image data on purpose [54]. In other words, makers of adult image content, like human nudity pictures, effectively hide the critical body regions of an adult image through mosaic processing in order to pass a motion picture rating board's deliberation. In addition, mosaic processing is intentionally applied to the faces of persons whose portrait rights are protected in Internet blogs, smoking scenes, tattoo, sharp and deadly weapons like knives, particular companies' trademarks irrelevant to indirect advertising, cruel scenes, actual names, resident registration numbers, home addresses, car license plates, and other personal information in still or dynamic color images.

In this study, we generate a grid-type mosaic first and then naturally overlap the generated mosaic [55] on the target object region detected in the previous step, as shown in Figure 8, so that the target object region with exposed personal information is effectively blocked. In other words, since a grid-type mosaic is generated on the exposed personal information area, damage caused by the exposure of personal information can be prevented.
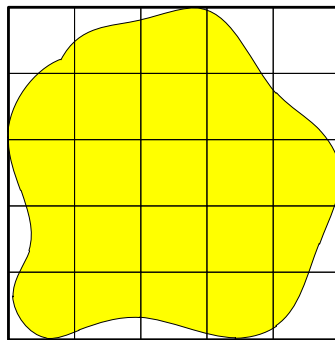


**Figure 8.** Partitioning the target object area into blocks.

In general, a mosaic block has a rectangular shape in which all pixels have the same color. In this study, a mosaic is generated in the unit of blocks [54,56] rather than the unit of pixels. In this case, it is important to let the mosaic overlap a detected target object region properly in terms of position and size. In addition, making the mosaic-overlapped target object region and its neighboring region natural by image processing is one of significant factors.

In this study, the filter with the size of N × N pixels is applied to replace the detected target object region with an average pixel value so as to generate a mosaic. Recently, the techniques of removing a mosaic from a digital color image are often found. For example, the technique of removing a mosaic from a digital image with the use of Google Brain, one of Google's artificial intelligence research projects, was announced [15]. More specifically, the 32 × 32 pixel-sized input color image was divided in the unit of 8 × 8-pixel size in order to generate a mosaic, and the divided images were recovered to its original image again. As such, a general mosaic technique can predict the shape of a mosaic region by putting the average value of pixels in the 8 × 8 pixel-sized region. When the recovered image is compared with ground truth, it is found that the recovered image is greatly similar to the original input image. Therefore, this study tries to apply a mosaic processing technique that does not allow recovery in order to robustly block a target region with exposed personal information.

In order to let a mosaic overlap a target object region with personal information as shown in Figure 8, the MER (Minimum Enclosing Rectangle) [57–60] of the detected target object region is first calculated. Subsequently, the image region in the MER is equally divided in the block unit of N × N pixels, and each block is filled with the color values generated by the corresponding mosaic processing technique.

In this study, a block-based mosaic region can be generated in three major methods. The first method is the most general one. In the method, the mean of the pixels located in the detected target

object region is calculated with Equation (7) and it is then entered in the blocks of the mosaic, as shown in Figure 9.

| 199 | 199 | 199 | 199 | 199 |
|-----|-----|-----|-----|-----|
| 202 | 202 | 202 | 202 | 202 |
| 201 | 202 | 200 | 198 | 199 |
| 198 | 198 | 198 | 198 | 198 |
| 201 | 201 | 201 | 201 | 201 |

| 200 | 200 | 200 | 200 | 200 |
|-----|-----|-----|-----|-----|
| 200 | 200 | 200 | 200 | 200 |
| 200 | 200 | 200 | 200 | 200 |
| 200 | 200 | 200 | 200 | 200 |
| 200 | 200 | 200 | 200 | 200 |

**Figure 9.** An example of mosaic generation.

In Equation (7), $n$ means the length or width size of a mosaic block. $r(x, y)$, $g(x, y)$, and $b(x, y)$ represent $r$, $g$, and $b$ color values of the pixels located at the $(x, y)$ position of the input image. In addition, $r_{mean}$, $g_{mean}$, and $b_{mean}$ are the average red, the average green, and the average blue color values of the pixels in a block.

$$
\begin{aligned}
r_{mean} &= \frac{1}{2n} \sum_{j=-n/2}^{n/2} \sum_{i=-n/2}^{n/2} r(x+i, y+j) \\
g_{mean} &= \frac{1}{2n} \sum_{j=-n/2}^{n/2} \sum_{i=-n/2}^{n/2} g(x+i, y+j) \\
b_{mean} &= \frac{1}{2n} \sum_{j=-n/2}^{n/2} \sum_{i=-n/2}^{n/2} b(x+i, y+j)
\end{aligned}
\tag{7}
$$

Secondly, as shown in Equation (8), pixels in the region with personal information are randomly selected so as to generate mosaic blocks. In Equation (8), $C_{mosaic}(x, y)$ represents the color value in the $(x, y)$ position of the generated mosaic region. $C_{image}(x, y)$ means the color value in the $(x, y)$ position of the input color image. $random()$ is the random data function for selecting a value randomly in a corresponding group. $R_{target}$ is the extracted target object region. If this mosaic generation method is utilized, it is difficult to predict the form of the original target object. In addition, this method can bring about a more natural result than the third mosaic generation method.

$$
\begin{aligned}
C_{mosaic}(x, y) &= C_{image}(x_i, y_i) \\
random&(x_i, y_i) \; where \; (x_i, y_i) \in R_{target}
\end{aligned}
\tag{8}
$$

Thirdly, as shown in Equation (9), $r$, $g$, and $b$ pixel values are randomly applied in each divided mosaic region. In Equation (9), $(r_i, g_i, b_i)$ represents the $r$, $g$, and $b$ color values randomly selected from the total set of $R$, $G$, and $B$ colors by the function $random()$. If this mosaic generation method is applied, the original image colors prior to the generation of a mosaic have relatively low correlations with the generated mosaic colors and, therefore, it is hard to recover the original target object region. When the mosaic-applied target object region is compared with its neighboring regions, colors rapidly change in the mosaic region and a somewhat unnatural visual result is shown.

$$
\begin{aligned}
C_{mosaic}(x, y) &= (r_i, g_i, b_i) \\
random \; &(r_i, g_i, b_i) \; where \; (r_i, g_i, b_i) \in (R, G, B)
\end{aligned}
\tag{9}
$$

Therefore, if the different mosaic generation methods aforementioned are adaptively used depending on the internal and external features of an input color image to which the block-unit-based

mosaic is applied, it is possible to block a target region with exposed personal information more effectively. For instance, if an input color image has many noises or is low quality due to lighting influence of the surrounding environment, the mean-based mosaic generation method can be applied to block a target object area. If there are relatively low correlations between the feature values of a detected target object and its neighboring regions, it is possible to generate a mosaic block by selecting the color values in the target object region randomly. If a target object area and its neighboring regions have relatively high correlations in terms of feature values, it is necessary to select one of all color values randomly and to generate a mosaic in order to make it difficult to remove the mosaic.

The strategy of adaptively selecting the most effective of the three mosaic generation methods will be further studied. At present, we plan to dynamically select three mosaic methods based on the characteristics of the target object region. In other words, the more we choose from the first method to the third method, the harder it is to restore the original target object area. In addition, as we choose from the first method to the third method, the color value is changed abruptly in the area where the block-based mosaic is located when compared with the areas around the target object region, resulting in unnatural results. Therefore, the basic strategy of dynamically selecting one of the three methods will be to select a mosaic generation method according to the type of corresponding target object area.

Besides, if the target object requires less security, the first mosaic generation method is used. If the target object needs more security, the third method is selected. In other words, if you need stronger security, such as naked images or cell phone numbers, the third method is chosen. If you need less security, such as car license plates, the first method is utilized. The second mosaic block is created for target objects that require intermediate security, such as faces. In addition, the setting of security strength can be adjusted according to the type of target object area.

We plan to extend the types of objects to be detected to human faces, body parts, cell phone numbers, and license plates. Also, we plan to detect these objects automatically and to then dynamically determine how to generate the mosaic blocks based on the type of detected objects.

## 5. Experimental Results

In this study, the personal computer used for the experiment has the following specification: Intel Core(TM) i7-6700 3.4 GHz CPU, 16 GB main memory, 256 GB SSD (solid state drive), and Galaxy Geforce GTX 1080 Ti graphics card with NVIDIA GPU GP104. As an operating system, Microsoft Windows 10 was used. As an integrated development environment (IDE), Microsoft Visual Studio Version 2015 was used and OpenCV open source computer vision library and Dlib C++ library were used to implement the proposed method.

In this paper, the face data set produced by VGG (visual geometry group) of Oxford University [61] was used as image data for the deep learning-based detection of facial regions. The VGG face dataset consists of 2622 files, and each file contains URL paths and face information containing face images. We randomly sampled and downloaded one image from each of the 2622 files. We attempted to use 80% of the 2622 files as training data, 10% as validation data, and 10% as test data. However, there were some invalid URLs among the image URL paths contained in each file. Therefore, a total of 2484 images, including 2268 images for training and verification and 216 images for testing, were downloaded. For various kinds of learning, 2268 face data and 3605 face information (x, y, w, h) were used. For the test, 216 face data and 314 face information (x, y, w, h) were used. As background images, areas with sizes ranging from $40 \times 40$ pixels to $200 \times 200$ pixels were randomly extracted except for face areas from input color images.

The learning rate used for deep learning began from 0.1 and sequentially decreased to 0.0001. The number of deep learning layers was 21, and the face normalization size was based on the $40 \times 40$ pixels. Six levels of image pyramids were generated first and then were used as an input of the proposed deep learning algorithm. A total of 183,745 learning steps were used. The average resolution of the image was $463 \times 413$ pixels, and the size of the window was $40 \times 40$ pixels. The window was moved by 2 pixels or 1 pixel. The number of learning steps is affected by the hyperparameters used in the proposed

model, that is, the initial values of the learning rate and weight, and depends on when the learning converges. In addition, we used cross-validation to raise confidence in the validation of the algorithm.

Figure 10a illustrates an example of the input color image that includes the exposed facial region. Figure 10b shows the image generated by extracting skin color distribution pixels with the use of the predefined elliptical skin color distribution model and by applying morphological operation for post-processing. Figure 10c presents the image generated by detecting the facial region from the extracted skin color distribution region with the use of the proposed deep neural computing algorithm. In Figure 10c, the square box is the MER of the detected facial region. Figure 10d shows the image with the blocked personal information, which was generated by creating the mean color-based mosaic for the detected facial region with the use of the deep learning technique and by letting the grid-type mosaic overlay the detected facial region.
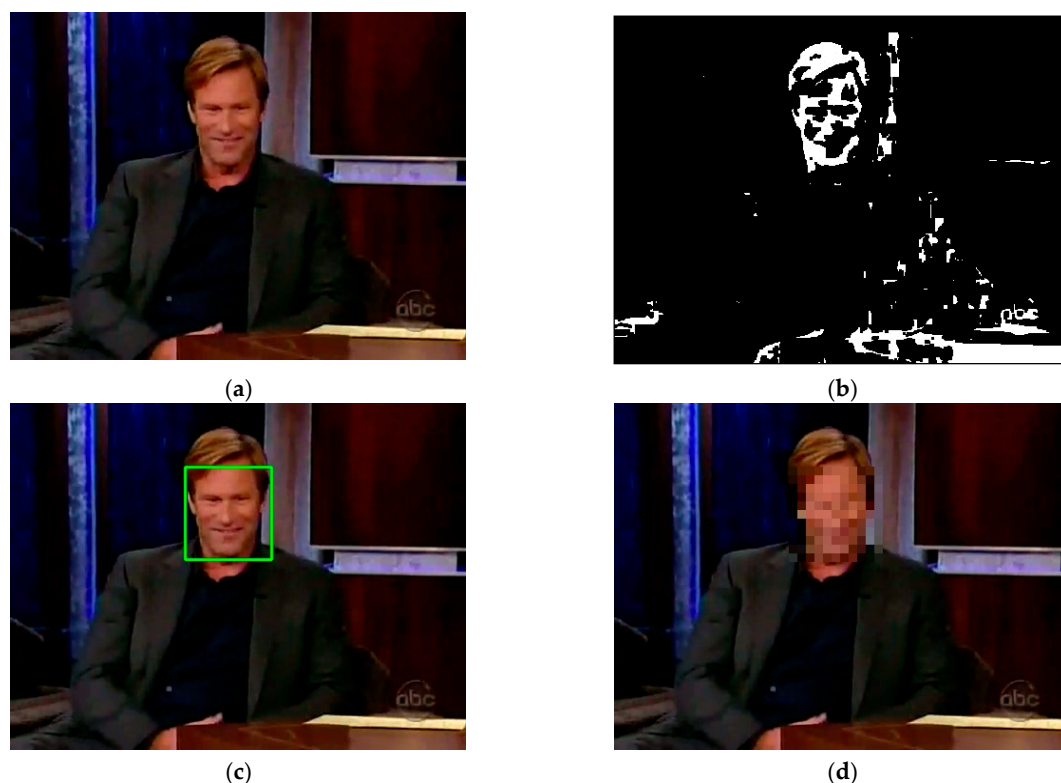


| (a) | (b) |
| (c) | (d) |

**Figure 10.** Blocking with mosaic processing: (**a**) input color image; (**b**) skin color pixels; (**c**) detected target region; and (**d**) generated mosaic area.

Figure 11a is another input color image. Figure 11b shows the image created by calculating the average value of pixels located in the detected object area and then by generating and overlaying a mosaic region. Figure 11c illustrates the image generated by selecting the pixel values of the detected face region randomly and then by creating and overlaying a mosaic. Figure 11d presents the image created by randomly applying the *r*, *g*, and *b* color pixel values and then by creating and overlaying a mosaic area. As shown in Figures 10 and 11, if the first or second mosaic generation method is used, it is possible to predict a blocked target object region roughly and, if the third method is used, it is difficult to predict a blocked object with personal information because a color value is selected randomly.
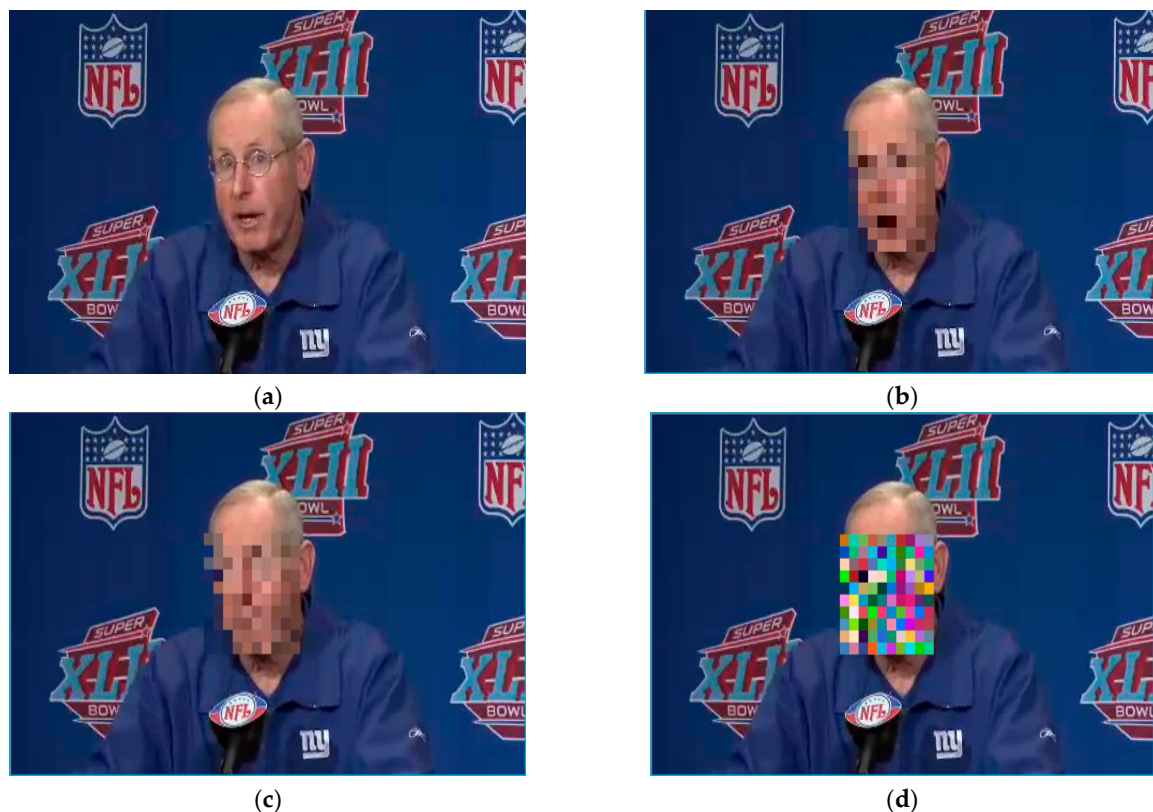
**Figure 11.** Different mosaic generation: (**a**) input image; (**b**) selection of the average value in pixel values; (**c**) random selection of pixel values in the detected object region; and (**d**) random selection of r, g, and b values.

We applied the proposed algorithm not only for still images but also for dynamic images. Figure 12 shows an example of detecting a face region in video data and continuously covering the detected face regions using block-based mosaics. The proposed blocking algorithm is effective enough to process video data.
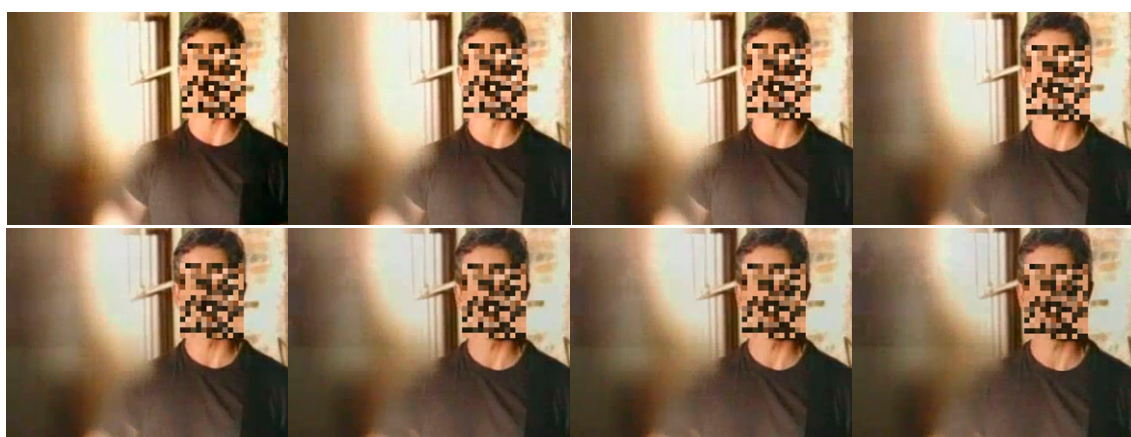


**Figure 12.** An example of blocking face regions in dynamic image sequences.

To quantitatively evaluate and compare the performance of the proposed deep neural network-based target object blocking technique, this study utilized such accuracy scale ways as those in Equations (10) and (11), where $N_{TP}$ means the number of accurately detected target regions, $N_{FP}$ is the number of regions which are not target regions but are incorrectly detected as target areas,

and $N_{FN}$ is the number of regions which are target areas but are not detected. In Equation (10), $R_{precision}$ represents the relative rate of the accurately detected target regions for the total detected target object regions. In Equation (11), $R_{recall}$ means the relative rate of the accurately detected target object regions for the total target regions existing in an input color image.

$$R_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{10}$$

$$R_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{11}$$

To evaluate the accuracy of the proposed algorithm to block a target object region with personal information, this study compared the conventional skin color-based method, the general learning-based method [17], and the proposed deep neural computing-based method. In terms of implementation, we took full advantage of the source code published on GitHub to implement the existing methods. Figures 13 and 14 illustrate the accuracy measurement graphs of the target object area blocking algorithms, which were obtained in Equation (10) and (11). The proposed method improves the precision rate by about 0.05 and 0.02, respectively, and the recall rate by 0.09 and 0.02, respectively, compared to the conventional skin color-based and general learning-based methods. As shown in Figures 13 and 14, the proposed method of blocking a target object region on the basis of skin color detection and deep neural network learning was capable of reducing wrong detection of target object regions and of blocking the areas with personal information more accurately.
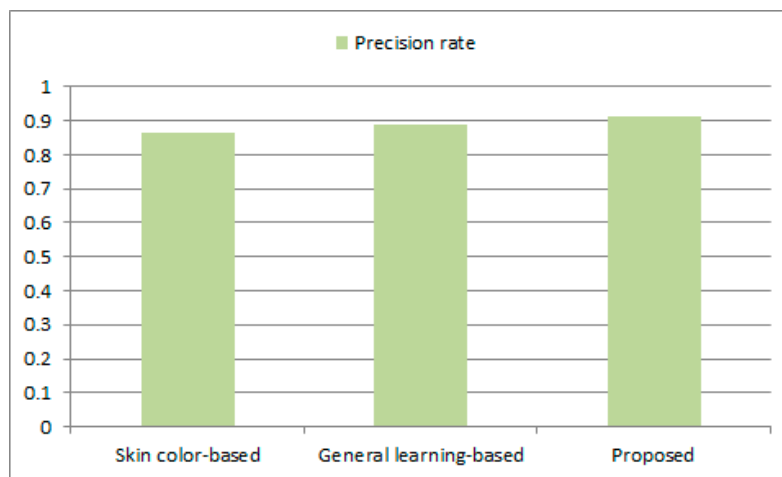


**Figure 13.** Performance comparison: precision rate.

Of the three methods, the method using skin color distribution had the lowest accuracy. That was because it simply used the feature of color to detect a target object area. In other words, the method simply using skin color distribution has some limitations in robustly detecting the target objects captured in various indoor and outdoor environments. Although the general learning-based method had a little better performance than the first method, it often failed to have enough learning and therefore caused wrong detection. The proposed method based on skin color detection and deep neural computing robustly detected target object regions including personal information from a variety of input color images and greatly reduced wrong detection, so that it showed relatively high accuracy.
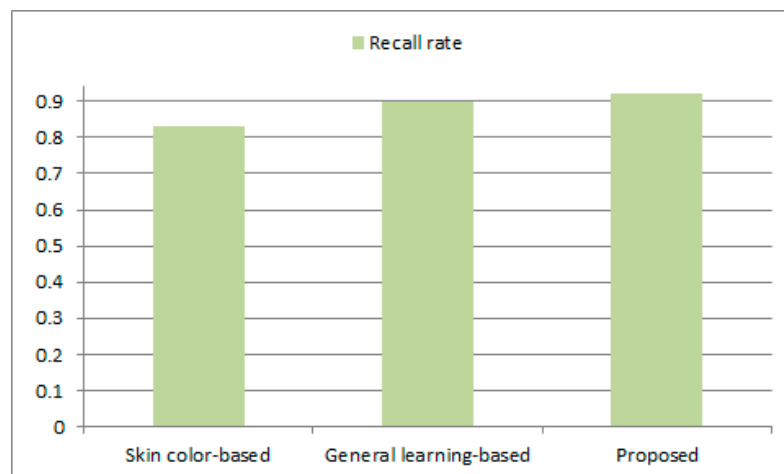
**Figure 14.** Performance comparison: recall rate.

The average processing time of our algorithm is 90.98 ms for an input color image. As a result, the proposed face blocking method is efficient enough for video processing. Furthermore, we plan to improve the processing speed of the suggested blocking approach by performing code optimization of the implemented algorithm and by upgrading the specifications of the personal computer used.

The proposed method is relatively robust to different skin colors because it detects human skin color distribution pixels through elliptical model-based prior learning. Therefore, if skin color training data is used for different races, our skin extraction module will work for different races. Of course, if the image contains a sudden change of lighting or severe noise, the proposed method may cause an error in face region detection. In addition, when there are many areas similar to the skin color in the background, false detection may occur. However, apart from this exceptional situation, our skin extraction module works robustly on a variety of input color images.

Human face detection using skin color causes some error in the presence of severe lighting changes. The proposed method is robust to weak light changes because it detects the face by using deep learning, but it is partially affected by severe light changes because the skin color is basically used. This problem is inevitably caused by the limitations of the camera equipment, which is one of the common problems with most image processing. This problem is expected to be solved in the future when high-performance cameras are developed that self-compensate for severe lighting changes. We plan to compensate for this problem by using an ultra-high-speed camera capable of capturing more than 1000 frames per second until a camera with such lighting self-calibration is developed.

Since the proposed blocking method targets a color image, the face detection algorithm should be changed when a black and white image is input. In other words, when a black and white image is input, instead of detecting a face region based on the skin color, the face region should be detected using the main features of the face from the input image.

Although the proposed approach is applied when one face region exists in one input image, it is also possible to block multiple face regions to a certain extent when two or more face regions do not overlap so much with each other. In other words, the proposed method first detects two or more faces simultaneously and then performs labeling. Subsequently, when the proposed method generates mosaics in units of areas corresponding to each label, multiple face areas are blocked. Figure 15 shows an example of detecting and blocking two face regions. In the future, much further research is needed for effectively blocking two or more face areas.
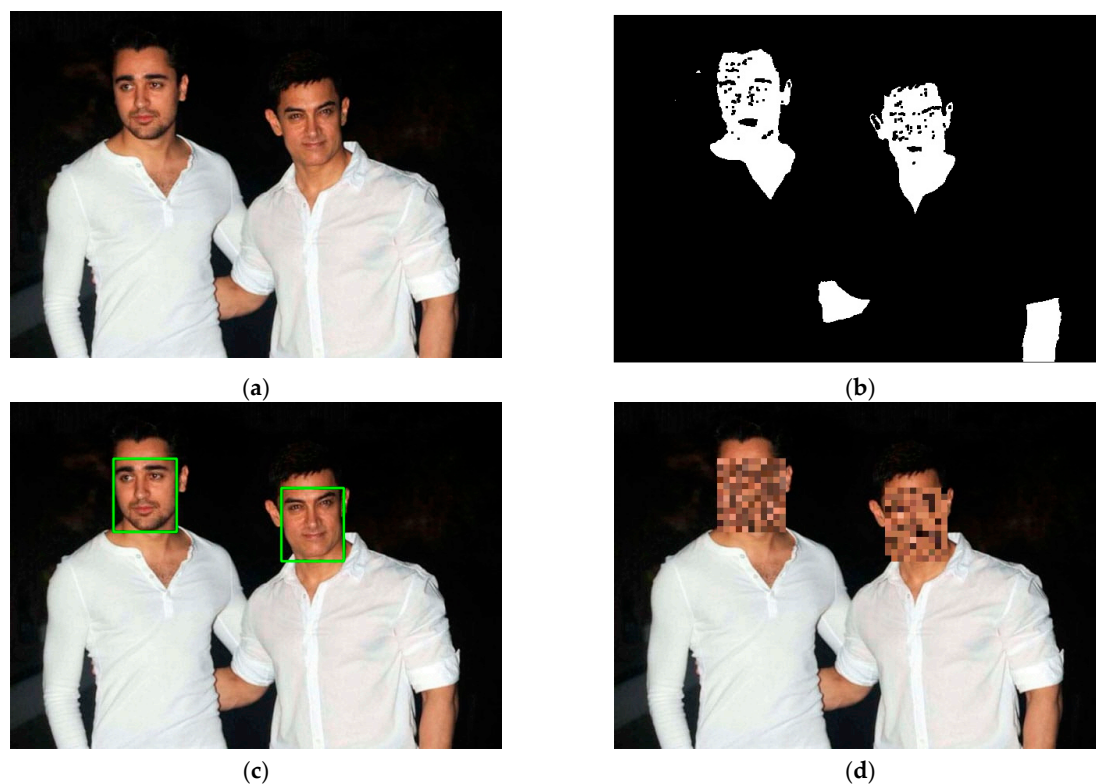
**Figure 15.** Blocking multiple faces: (**a**) input color image; (**b**) skin color pixels; (**c**) detected face regions; and (**d**) generated mosaic areas

## 6. Conclusions

These days, it is easy to acquire various kinds of multimedia content with personal computers, tablet personal computers, and mobile devices that support high-speed wired and wireless Internet. In this circumstance, multimedia content including target objects with personal information like human face or particular body region is freely distributed, which emerges as a big social issue. Therefore, many studies have been conducted to robustly detect multimedia content with personal information and then to effectively block the detected target objects through block-based mosaic-processing.

This study proposed the method of robustly detecting target objects with exposed personal information from a variety of input color images in neural network-based learning algorithm and then effectively blocking the detected target objects. In the proposed method, the *RGB* color space of an image is changed to the $YC_bC_r$ color space. With the use of the skin color distribution model generated with an elliptical model, background regions are excluded from an input image and only skin color distribution regions are detected. Subsequently, the artificial neural network-based deep learning algorithm is applied to accurately extract the target object region with personal information from the skin color region detected in the previous step. In the last step, a grid-like mosaic region with the same length and width sizes is generated and the generated mosaic naturally overlaps the extracted target region so as to effectively block the target object region with personal information.

In the experiment of this study, the proposed algorithm was applied to various kinds of color images captured in a general environment without particular constraints, and the accuracy performance of the proposed method was evaluated qualitatively. As a result, the proposed technique using the artificial neural network-based deep learning more accurately detected and blocked a target object area with personal information than conventional methods. In other words, the proposed blocking method based on the deep neural computing algorithm with excellent performance more robustly detected target object areas from various color images and had relatively higher accuracy than conventional methods. In addition, the proposed method showed relatively good efficiency.

Most of the conventional blocking methods are very inefficient because the user manually detects a target object from the image content and then manually blocks the detected object area using mosaics. That is, the operator directly creates and applies a mosaic to a corresponding area of the image by using an image editing tool. On the other hand, the proposed method not only automatically detects the target object area based on deep learning but also automatically generates a mosaic to effectively cover the target area. Moreover, the existing mosaic generation methods generate mosaics using only one method, whereas the proposed method attempts to adaptively select and apply three types of mosaic generation methods according to characteristics of an input color image.

The future plan is to apply the proposed neural network-based target object blocking algorithm to the input color images captured in more diverse indoor and outdoor environments without constraints and then to prove the robustness of the proposed algorithm systematically. In addition, we plan to stabilize system operation performance by adaptively tuning multiple parameters including the learning rate used in the proposed and implemented target object blocking algorithm. Aside from that, the tracking function using a predictive algorithm like Kalman filter will be added in order to detect and block a target object faster.

## References

1. Jimenez-Marquez, J.L.; Gonzalez-Carrasco, I.; Lopez-Cuadrado, J.L.; Ruiz-Mezcua, B. Towards a Big Data Framework for Analyzing Social Media Content. *Int. J. Inf. Manag.* **2019**, *44*, 1–12. [CrossRef]
2. AlSkaif, T.; Bellalta, B.; Zapata, M.G.; Barcelo Ordinas, J.M. Energy Efficiency of MAC Protocols in Low Data Rate Wireless Multimedia Sensor Networks: A Comparative Study. *Ad Hoc Netw.* **2017**, *56*, 141–157. [CrossRef]
3. Du, D.; Qi, B.; Fei, M.; Wang, Z. Quantized Control of Distributed Event-Triggered Networked Control Systems with Hybrid Wired-Wireless Networks Communication Constraints. *Inf. Sci.* **2017**, *380*, 74–91. [CrossRef]
4. Yang, X.; Wen, Y.; Yuan, D.; Zhang, M.; Zhao, H.; Meng, Y. 3D Compression-Oriented Image Content Correlation Model for Wireless Visual Sensor Networks. *IEEE Sens. J.* **2018**, *18*, 6461–6471. [CrossRef]
5. Amina, S.; Mohamed, F.K. An Efficient and Secure Chaotic Cipher Algorithm for Image Content Preservation. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *60*, 12–32. [CrossRef]
6. Zhang, S.; McCullagh, P.; Zheng, H.; Nugent, C. Situation Awareness Inferred from Posture Transition and Location: Derived from Smartphone and Smart home Sensors. *IEEE Trans. Hum. Mach. Syst.* **2017**, *47*, 814–821. [CrossRef]
7. Yao, L.; Ge, Z. Deep Learning of Semisupervised Process Data with Hierarchical Extreme Learning Machine and Soft Sensor Application. *IEEE Trans. Ind. Electron.* **2018**, *65*, 1490–1498. [CrossRef]
8. Preishuber, M.; Hutter, T.; Katzenbeisser, S.; Uhl, A. Depreciating Motivation and Empirical Security Analysis of Chaos-Based Image and Video Encryption. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2137–2150. [CrossRef]
9. Tran, H.Y.; Hu, J. Privacy-Preserving Big Data Analytics a Comprehensive Survey. *J. Parallel Distr. Comput.* **2019**, *134*, 207–218. [CrossRef]
10. Memos, V.A.; Psannis, K.E.; Ishibashi, Y.; Kim, B.G.; Gupta, B.B. An Efficient Algorithm for Media-based Surveillance System (EAMSuS) in IoT Smart City Framework. *Future Gener. Comput. Syst.* **2018**, *83*, 619–628. [CrossRef]

11. Paliwal, N.; Vanjani, P.; Liu, J.W.; Saini, S.; Sharma, A. Image processing-based intelligent robotic system for assistance of agricultural crops. *Int. J. Soc. Humanist. Comput.* **2019**, *3*, 191–204. [CrossRef]

12. Xu, Z. The analytics and applications on supporting big data framework in wireless surveillance networks. *Int. J. Soc. Humanist. Comput.* **2017**, *2*, 41–149. [CrossRef]

13. Tian, M.; Ni, L.; Xu, L.; Li, H.; Liu, X. Multi-face real-time tracking based on dual panoramic camera for full-parallax light-field display. *Opt. Commun.* **2019**, *442*, 19–26. [CrossRef]

14. Li, C.; Huang, Y.; Xue, Y. Dependence structure of Gabor wavelets based on copula for face recognition. *Expert Syst. Appl.* **2019**, *137*, 453–470. [CrossRef]

15. Dahl, R.; Norouzi, M.; Shlens, J. Pixel Recursive Super Resolution. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5449–5458. [CrossRef]

16. Jang, S.W.; Park, Y.J.; Kim, G.Y.; Choi, H.I.; Hong, M.C. An adult image identification system based on robust skin segmentation. *J. Imaging Sci. Technol.* **2011**, *55*, 20508-1–20508-10. [CrossRef]

17. Liu, J.; Huang, L.; Lin, J. An image mosaic block detection method based on fuzzy C-Means clustering. In Proceedings of the 2011 3rd International Conference on Computer Research and Development, Shanghai, China, 11–13 March 2011; pp. 237–240. [CrossRef]

18. Zhang, T.; Li, J.; Jia, W.; Sun, J.; Yang, H. Fast and robust occluded face detection in ATM surveillance. *Pattern Recogn. Lett.* **2017**, *107*, 33–40. [CrossRef]

19. Zhang, H.; Wang, X.; Zhu, J.; Jay Kuo, C.C. Fast face detection on mobile devices by leveraging global and local facial characteristics. *Signal Process. Image Commun.* **2019**, *78*, 1–8. [CrossRef]

20. Wang, L.; Yu, X.; Bourlai, T.; Metaxas, D.N. A coupled encoder-decoder network for joint face detection and landmark localization. *Image Vis. Comput.* **2019**, *87*, 37–46. [CrossRef]

21. Lee, S.; Kwak, Y.; Kim, Y.J.; Park, S.; Kim, J. Contrast-Preserved Chroma Enhancement Technique Using YCbCr Color Space. *IEEE Trans. Consum. Electron.* **2012**, *58*, 641–645. [CrossRef]

22. Chiang, J.S.; Hsia, C.H.; Peng, H.W.; Lien, C.H.; Li, H.T. Saturation Adjustment Method Based on Human Vision with YCbCr Color Model Characteristics and Luminance Changes. In Proceedings of the 2012 International Symposium on Intelligent Signal Processing and Communications Systems, Taipei, Taiwan, 4–7 November 2012; pp. 136–141. [CrossRef]

23. Zhu, S.Y.; He, Z.Y.; Chen, C.; Liu, S.C.; Zhou, J.; Guo, Y.; Zeng, B. High-Quality Color Image Compression by Quantization Crossing Color Spaces. *IEEE Trans. Circuits Syst. Video Tech.* **2019**, *29*, 1474–1487. [CrossRef]

24. Tan, Y.; Qin, J.; Xiang, X.; Ma, W.; Pan, W.; Xiong, N.N. A Robust Watermarking Scheme in YCbCr Color Space Based on Channel Coding. *IEEE Access* **2019**, *7*, 25026–25036. [CrossRef]

25. Chakraborty, B.K.; Bhuyan, M.K.; Kumar, S. Combining Image and Global Pixel Distribution Model for Skin Colour Segmentation. *Pattern Recogn. Lett.* **2017**, *88*, 33–40. [CrossRef]

26. Al-Mohair, H.K.; Saleh, J.M.; Suandi, S.A. Hybrid Human Skin Detection Using Neural Network and K-Means Clustering Technique. *Appl. Soft Comput.* **2015**, *33*, 337–347. [CrossRef]

27. Hsu, R.L.; Abdel-Mottaleb, M.; Jain, A.K. Face Detection in Color Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 696–706. [CrossRef]

28. Lou, S.; Jiang, X.; Scott, P.J. Algorithms for Morphological Profile Filters and Their Comparison. *Precis. Eng.* **2012**, *36*, 414–423. [CrossRef]

29. Su, R.; Sun, C.; Zhang, C.; Pham, T.D. A New Method for Linear Feature and Junction Enhancement in 2D Images Based on Morphological Operation, Oriented Anisotropic Gaussian Function and Hessian Information. *Pattern Recogn.* **2014**, *47*, 3193–3208. [CrossRef]

30. Hamuda, E.; Ginley, B.M.; Glavin, M.; Jones, E. Automatic Crop Detection under Field Conditions Using the HSV Colour Space and Morphological Operations. *Comput. Electron. Agric.* **2017**, *133*, 97–107. [CrossRef]

31. Zhang, Y.; Ji, T.Y.; Li, M.S.; Wu, Q.H. Identification of Power Disturbances Using Generalized Morphological Open-Closing and Close-Opening Undecimated Wavelet. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2330–2339. [CrossRef]

32. He, L.; Zhao, X.; Chao, Y.; Suzuki, K. Configuration-Transition-based Connected Component Labeling. *IEEE Trans. Image Process.* **2014**, *23*, 943–951. [CrossRef]

33. Ma, L.; Zhang, X.P.; Si, J.; Abousleman, G.P. Bidirectional Labeling and Registration Scheme for Grayscale Image Segmentation. *IEEE Trans. Image Process.* **2005**, *14*, 2073–2081. [CrossRef]

34. Du, B.; Wang, Z.; Zhang, L.; Zhang, L.; Tao, D. Robust and Discriminative Labeling for Multi-Label Active Learning Based on Maximum Correntropy Criterion. *IEEE Trans. Image Process.* **2017**, *26*, 1694–1707. [CrossRef]

35. Garcia-Martin, E.; Rodrigues, C.F.; Riley, G.; Grahn, H. Estimation of Energy Consumption in Machine Learning. *J. Parallel Distr. Comput.* **2019**, *134*, 75–88. [CrossRef]

36. Din, I.U.; Guizani, M.; Rodrigues, J.J.P.C.; Hassan, S.; Korotaev, V.V. Machine Learning in the Internet of Things: Designed Techniques for Smart Cities. *Future Gener. Comput. Syst.* **2019**, *100*, 826–843. [CrossRef]

37. Zong, W.; Chow, Y.W.; Susilo, W. Interactive Three-Dimensional Visualization of Network Intrusion Detection Data for Machine Learning. *Future Gener. Comput. Syst.* **2020**, *102*, 292–306. [CrossRef]

38. Shukla, S.; Raghuwanshi, B.S. Online Sequential Class-Specific Extreme Learning Machine for Binary Imbalanced Learning. *Neural Netw.* **2019**, *119*, 235–248. [CrossRef]

39. Shao, L.; Wu, D.; Li, X. Learning Deep and Wide: A Spectral Method for Learning Deep Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 2303–2308. [CrossRef]

40. Zhang, S.; Pan, X.; Cui, Y.; Zhao, X.; Liu, L. Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning. *IEEE Access* **2019**, *7*, 32297–32304. [CrossRef]

41. Mahmud, M.; Kaiser, M.S.; Hussain, A.; Vassanelli, S. Applications of Deep Learning and Reinforcement Learning to Biological Data. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2063–2079. [CrossRef]

42. Kastrati, Z.; Imran, A.S.; Yayilgan, S.Y. The impact of Deep Learning on Document Classification Using Semantically Rich Representations. *Inf. Process. Manag.* **2019**, *56*, 1618–1632. [CrossRef]

43. Zhang, Y.D.; Pan, C.; Sun, J.; Tang, C. Multiple Sclerosis Identification by Convolutional Neural Network with Dropout and Parametric ReLU. *J. Comput. Sci.* **2018**, *28*, 1–10. [CrossRef]

44. Li, J.; Zhang, T.; Luo, W.; Yang, J.; Yuan, X.T.; Zhang, J. Sparseness Analysis in the Pretraining of Deep Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 1425–1438. [CrossRef]

45. Eckle, K.; Schmidt-Hieber, J. A Comparison of Deep Networks with ReLU Activation Function and Linear Spline-Type Methods. *Neural Netw.* **2019**, *110*, 232–242. [CrossRef]

46. Macedo, D.; Zanchettin, C.; Oliveira, A.L.I.; Ludermir, T. Enhancing Batch Normalized Convolutional Networks Using Displaced Rectifier Linear Units: A Systematic Comparative Study. *Expert Syst. Appl.* **2019**, *124*, 271–281. [CrossRef]

47. Zhou, W.; Xu, Z. Asymmetric Hesitant Fuzzy Sigmoid Preference Relations in the Analytic Hierarchy Process. *Inf. Sci.* **2016**, *358*, 191–207. [CrossRef]

48. Silva, H.O.; Bastos-Filho, C.J.A. Inter-Domain Routing for Communication Networks Using Hierarchical Hopfield Neural Networks. *Eng. Appl. Artif. Intell.* **2018**, *70*, 184–198. [CrossRef]

49. Wang, X.; Qin, Y.; Wang, Y.; Xiang, S.; Chen, H. ReLTanh: An Activation Function with Vanishing Gradient Resistance for SAE-based DNNs and Its Application to Rotating Machinery Fault Diagnosis. *Neurocomputing* **2019**, *363*, 88–98. [CrossRef]

50. Zuo, R.; Xiong, Y.; Wang, J.; Carranza, E.J.M. Deep Learning and Its Application in Geochemical Mapping. *Earth Sci. Rev.* **2019**, *192*, 1–14. [CrossRef]

51. Kastrati, Z.; Imran, A.S.; Kurti, A. Integrating Word Embedings and Document Topics with Deep Learning in a Video Classification Framework. *Pattern Recogn. Lett.* **2019**, *128*, 85–92. [CrossRef]

52. Chrysos, G.G.; Antonakos, E.; Snape, P.; Asthana, A.; Zafeiriou, S.A. Comprehensive Performance Evaluation of Deformable Face Tracking in-the-Wild. *Int. J. Comput. Vis.* **2018**, *126*, 198–232. [CrossRef]

53. Davis, E.K. Max-Margin Object Detection. *arXiv* **2015**, arXiv:1502.00046v1.

54. Guo, D.; Tang, J.; Cui, Y.; Ding, J.; Zhao, C. Saliency-based Content-Aware Lifestyle Image Mosaics. *J. Vis. Commun. Image Represent.* **2015**, *26*, 192–199. [CrossRef]

55. Sun, S.F.; Han, S.H.; Wang, G.; Xu, Y.C.; Lei, B.J. Mosaic Defect Detection in Digital Video. In Proceedings of the IEEE Chinese Conference on Pattern Recognition (CCPR), Chongqing, China, 21–23 October 2010; pp. 1–5. [CrossRef]

56. Shifa, A.; Imtiaz, M.B.; Asghar, M.N.; Fleury, M. Skin detection and lightweight encryption for privacy protection in real-time surveillance applications. *Image Vis. Comput.* **2020**, *94*, 103859. [CrossRef]

57. Jiang, N.; Yang, W.; Duan, L.; Xu, X.; Liu, Q. Acceleration of CT Reconstruction for Wheat Tiller Inspection Based on Adaptive Minimum Enclosing Rectangle. *Comput. Electron. Agric.* **2012**, *85*, 123–133. [CrossRef]

58. Kwak, E.; Habib, A. Automatic Representation and Reconstruction of DBM from LiDAR Data Using Recursive Minimum Bounding Rectangle. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 171–191. [CrossRef]

59. Zhang, H.; Tian, X.; Wang, C.; Wu, F.; Zhang, B. Merchant Vessel Classification Based on Scattering Component Analysis for COSMO-SkyMed SAR Images. *Geosci. Remote Sens. Lett.* **2013**, *10*, 1275–1279. [CrossRef]

60. Yang, H.-M.; Lim, D.-W.; Choi, Y.-S.; Kang, J.-G.; Kim, I.-H.; Lin, A.; Jung, J.-W. Image-Based Human Sperm Counting Method. *Int. J. Soc. Humanist. Comput.* **2019**, *3*, 148–157. [CrossRef]

61. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the 26th British Machine Vision Conference, Swansea, UK, 7–10 September 2015; pp. 1–12. [CrossRef]