A Comprehensive Study on Key Components of Grayscale-based Deepfake Detection

Seok Bin Son[1], Seong Hee Park[2], and Youn Kyu Lee[2*]
[1] Department of Electrical and Computer Engineering, Korea University,
Seoul 02841, Republic of Korea
[e-mail: lydiasb@korea.ac.kr]
[2] Department of Computer Engineering, Hongik University,
Seoul 04066, Republic of Korea
[e-mail: tjdgml0401@g.hongik.ac.kr, younkyul@hongik.ac.kr]
[*]Corresponding author: Youn Kyu Lee

Received August 18, 2023; revised December 25, 2023; revised April 12, 2024; revised June 10, 2024; accepted July 28, 2024; published August 31, 2024

Abstract

Advances in deep learning technology have enabled the generation of more realistic deepfakes, which not only endanger individuals' identities but also exploit vulnerabilities in face recognition systems. The majority of existing deepfake detection methods have primarily focused on RGB-based analysis, offering unreliable performance in terms of detection accuracy and time. To address the issue, a grayscale-based deepfake detection method has recently been proposed. This method significantly reduces detection time while providing comparable accuracy to RGB-based methods. However, despite its significant effectiveness, the "key components" that directly affect the performance of grayscale-based deepfake detection have not been systematically analyzed. In this paper, we target three key components: RGB-to-grayscale conversion method, brightness level in grayscale, and resolution level in grayscale. To analyze their impacts on the performance of grayscale-based deepfake detection, we conducted comprehensive evaluations, including component-wise analysis and comparative analysis using real-world datasets. For each key component, we quantitatively analyzed its characteristics' performance and identified differences between them. Moreover, we successfully verified the effectiveness of an optimal combination of the key components by comparing it with existing deepfake detection methods.

Keywords: biometrics, presentation attack detection, deepfake, face recognition, image classification, image fusion, image processing

http://doi.org/10.3837/tiis.2024.08.010 **ISSN : 1976-7277**

1. Introduction

Recent advances in deep learning technology have enabled the generation of more realistic deepfakes, which can be used for malicious purposes such as identity theft [1, 2]. Additionally, deepfakes can cause various threats, such as the dissemination of fake information and manipulation of public opinion, but also exploiting the vulnerabilities of face recognition systems [3]. To counteract the threats, a number of deepfake detection methods have been proposed, with the majority of them focusing on RGB-based video analysis [4–8]. In particular, RGB has been commonly used in deepfake detection because it is considered to provide richer information compared to other color spaces such as grayscale, HSV, and YCbCr [9, 10]. However, the RGB-based deepfake detection method still provides unreliable performance in terms of detection accuracy and time [11–14]. Hence, deepfake detection methods using different color spaces have recently been proposed [15, 16], and it has been noted that, in particular, grayscale-based deepfake detection significantly reduces detection time while providing comparable detection accuracy to RGB-based deepfake detection [14, 17].

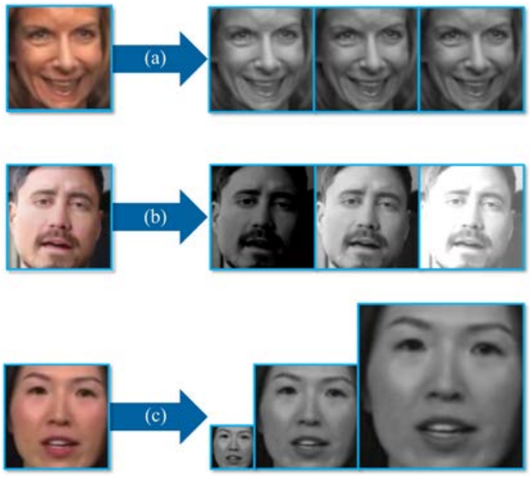


Fig. 1. Differences in grayscale images by key components: (a) RGB-to-grayscale conversion, (b) brightness level, and (c) resolution level

Despite its significant effectiveness, the absence of a systematic analysis on the “key components” that directly impact grayscale-based deepfake detection poses a challenge, contributing to a stagnation in enhancing its performance. For example, in grayscale-based analysis, components such as color contrast, brightness, and resolution can directly affect the performance of deepfake detection [18, 19]. **Fig. 1** demonstrates the differences between grayscale images by each component in grayscale-based analysis. To implement a robust grayscale-based deepfake detection system, a systematic examination of the key components that directly influence its performance is required. Therefore, our goal in this paper is to identify these key components, examine their impacts, and configure their optimal combination in grayscale-based deepfake detection.

In this paper, we identified the key components, and analyzed their impact on the performance of grayscale-based deepfake detection through comprehensive evaluations using real-world datasets (=Component-wise Analysis). The key components targeted in this paper are as follows: (1) RGB-to-grayscale conversion method, (2) brightness level in grayscale, and

(3) resolution level in grayscale. To analyze the impact of each key component, we used popular open datasets, Celeb-DF [22], FaceForensics++ (FF++) [21], and DeepFake-TIMIT (DF-TIMIT) [20], and employed six evaluation metrics, AUROC, confusion matrix (accuracy, precision, recall, and F1-score), and detection time. Additionally, based on the analysis results, we evaluated the effectiveness of an optimal combination of the key components and verified that it provided superior performance compared to existing deepfake detection methods (=Comparative Analysis). The contributions of this paper are as follows:

- Identification of key components that affect grayscale-based deepfake detection.
- Evaluation of each component's impact on deepfake detection.
- Design of robust grayscale-based deepfake detection with an optimal combination of key components.
- Comprehensive evaluation via prototype implementation and real-world datasets.

This paper is organized as follows. Section 2 describes related work, and Section 3 demonstrates our approach. Our experimental results are presented in Section 4, and discussion is included in Section 5. Finally, the conclusion and future works are presented in Section 6.

2. Related Work

2.1 Deepfake Detection

Various detection methods using neural networks have been proposed to discriminate against sophisticated deepfakes [4–8]. Ramachandran et al. assessed the effectiveness of deep face recognition trained using a variety of loss functions, in identifying deepfakes [23]. Kim et al. proposed a feature representation transfer adaptation learning framework employing a knowledge distillation paradigm and representation learning for quick adaptation to new types of deepfakes [4]. Ismail et al. proposed a deepfake detection method employing the YOLO face detector, convolutional neural network (CNN) model, and extreme gradient boosting classifier [5]. Wodajo et al. designed a convolutional vision transformer that combined a vision transformer with a CNN model for extracting facial features [6]. Liu et al. proposed a lightweight 3D CNN framework fusing the spatial feature in the time dimension, which detects deepfakes with a small number of parameters [7]. Xu et al. proposed an ensemble detection method that employs the contrasts between unmanipulated and manipulated images for generalizable performance on unseen data [8]. Aghasanli et al. proposed a deepfake detection method that combines the vision transformer with various classifiers [24]. Wang et al. introduced a deepfake detection model based on noise, which extracts noise features from both the face and background, calculating their interaction [25]. Despite the various approaches proposed, existing deepfake detection methods still offer unreliable performance in terms of detection accuracy and time [11–14]. In addition, despite the potential of different color spaces, existing deepfake detection methods have mainly focused on RGB-based analysis [22, 21, 26, 27].

2.2 Grayscale Conversion

In the field of image analysis, the grayscale channel has been widely used for the extraction of image features and the evaluation of image quality [28]. To support this, several techniques for converting RGB to grayscale have been proposed [18, 29–31]. Intensity is the simplest conversion method that uses the arithmetic mean of RGB channels [18, 29]. Luma, a popular conversion method based on the YIQ color model, is widely used in image processing

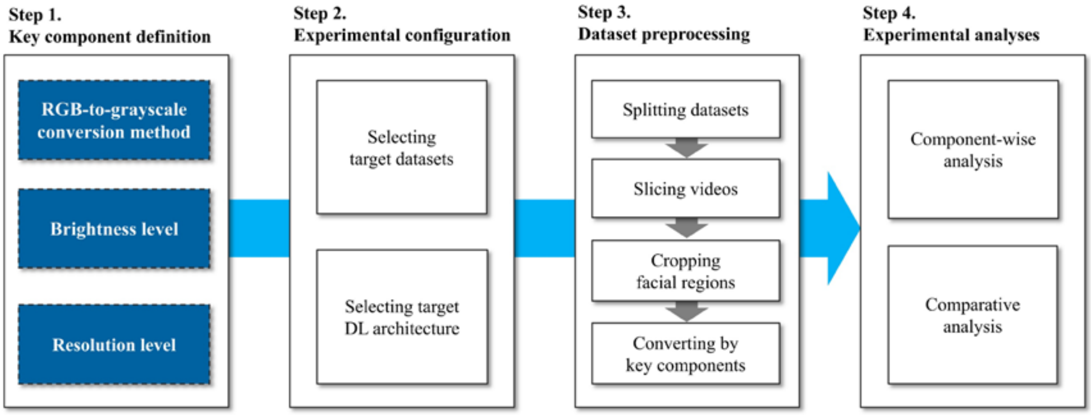


Fig. 2. An overview diagram of the proposed method

mechanisms supporting contrast adjustment [29, 32]. Luminance, a conversion method based on OpenCV, is one of the standard algorithms for image processing software such as GIMP [18, 33]. Ambalathankandy et al. proposed an O(1) decolorization technique, which can convert RGB-to-grayscale at high speed using the difference in color temperature [30]. Paramasivam et al. proposed a color-invariant conversion method employing multiple regression to provide adaptive weights depending on the contribution of each channel [31]. According to prior studies, different RGB-to-grayscale conversion techniques produce varying degrees of brightness and color contrast preservation [18].

3. Our Approach

In this research, we identified the key components of grayscale-based deepfake detection, and analyzed the impact of each component on deepfake detection performance. Based on the analysis results, we constructed an optimal combination of key components and evaluated its performance by comparing it with existing deepfake detection methods.

As shown in **Fig. 2**, our approach comprises four main steps: (*Step 1 - Key component identification*): We identified the key components directly affecting the performance of grayscale-based deepfake detection; (*Step 2 - Experimental configuration*): We selected target datasets and deep learning (DL) model architecture for experimental analysis; (*Step 3 - Dataset preprocessing*): We performed dataset preprocessing and RGB-to-grayscale conversion based on each key component; (*Step 4 - Experimental analysis*): We performed two different types of analysis, component-wise and comparative. The component-wise analysis examines each component's impact on the performance of deepfake detection. The comparative analysis evaluates the optimal combination of the key components by comparing it with the existing deepfake detection solutions. The followings are detailed descriptions of each step.

3.1 Step 1: Key Component Identification

In this step, we identified the key components that can directly affect the performance of grayscale-based deepfake detection, along with their respective hyperparameters as variables. Grayscale-based deepfake detection involves converting RGB to grayscale during preprocessing, which may potentially distort crucial features of the original video depending on components such as conversion methods, brightness, and resolution. Therefore, it is essential to identify those key components that directly impact effectiveness of deepfake

Table 1. The formulas of RGB-to-grayscale conversion methods

Methods	Conversion Formula
Intensity [18]	(R + G + B) / 3
Luma [29]	0.299 × R + 0.587 × G + 0.114 × B
Luminance [18]	0.3 × R + 0.59 × G + 0.11 × B

*R: Red channel, G: Green channel, B: Blue channel

detection in grayscale-based methods. This enables the preservation of crucial features within the original video, thereby enhancing detection performance and aiding in the development of grayscale-based deepfake detection algorithms. Previous studies have demonstrated that the choice of RGB-to-grayscale conversion algorithm has a significant impact on image recognition performance in grayscale channels, making it a crucial component for grayscale-based deepfake detection [18]. Moreover, RGB-to-grayscale conversion is a lossy operation that may cause brightness degradation, leading to a loss of important information. Therefore, preserving at least the brightness feature of the original images is crucial in the RGB-to-grayscale conversion, making the brightness level another key component of grayscale-based deepfake detection [19]. In addition, previous research has shown that even low-resolution grayscale images can provide sufficient information for saliency detection, comparable to high-resolution RGB images [19]. Hence, the resolution level of a grayscale image is also a key component of grayscale-based deepfake detection. As a result, we identified the key components as follows: (1) RGB-to-grayscale conversion method, (2) brightness level in grayscale, and (3) resolution level in grayscale. Note that other factors, such as contrast, saturation, noise, and shadows, may potentially impact the effectiveness of deepfake detection. However, in this study, to identify the key components in grayscale-based deepfake detection, we systematically analyzed existing studies and considered the other factors as out of scope. Hence, in our experiments, we maintained them in their original state without any manipulation. The detailed description of each key component is as follows.

3.1.1 RGB-to-grayscale Conversion Method

We examined the RGB-to-grayscale conversion methods proposed thus far, along with their corresponding commercial applications, and selected the three most widely employed variables as our primary targets [18, 29–31, 34–38]: Intensity [18], Luma [29], and Luminance [18]. The conversion formulas for each method are shown in **Table 1**. The weights in each formula indicate the degree of contribution of each channel (i.e., red, green, and blue) to the grayscale channel to be converted [18]. The red, green, and blue values of each pixel are multiplied by the weights to determine a resultant value ranging from 0 to 255. As shown in **Fig. 3**, the contrast and brightness levels of the grayscale frames can appear differently because of the different weights in each conversion method.

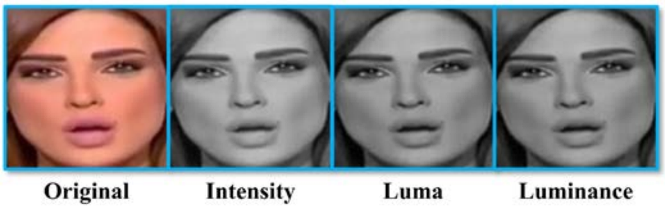


Fig. 3. An example of differences between RGB-to-grayscale conversion methods (Intensity, Luma, and Luminance)

3.1.2 Brightness Level in Grayscale

Depending on the brightness, highlighted areas, shadow amounts, and contrast can be different in the same frames [18, 29]. The brightness in the grayscale channel ranges from 0 to 255, typically scaled in 11 levels from 0 (=black) to 10 (=white) [39]. To accurately verify the impact of brightness differences on the deepfake detection performance, we divided the brightness levels in this study into a total of three variables (B1: 0 to 85.32, B2: 85.33 to 170.65, and B3: 170.66 to 255). Out of the 11 levels, excluding levels 0 and 10 where facial features are difficult to distinguish, we categorized the remaining nine levels into three distinct levels based on their respective median values as references. **Fig. 4** illustrates the differences between grayscale frames caused by different brightness levels (i.e., B1-B3).

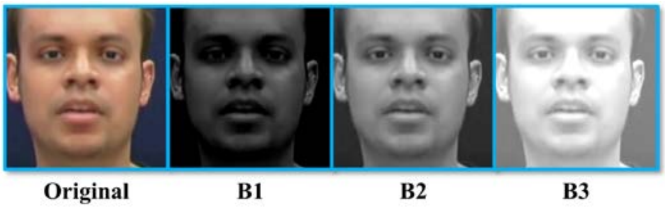


Fig. 4. An example of differences between brightness levels (B1: 0 to 85.32, B2: 85.33 to 170.65, and B3: 170.66 to 255)

3.1.3 Resolution Level in Grayscale

The resolution refers to the degree of detail that can be depicted in the frame [40]. Resolution has a direct impact on deepfake detection performance because it determines how realistically synthetic areas appear in deepfakes [19, 20, 41]. In this study, we divided the resolution levels into a total of three variables (R1: 64×64 pixels, R2: 128×128 pixels, and R3: 256×256 pixels), which were based on the resolution levels of the images comprising the most widely used deepfake datasets [20–22, 42–46]. **Fig. 5** illustrates the differences between grayscale frames caused by different resolution levels (i.e., R1-R3).

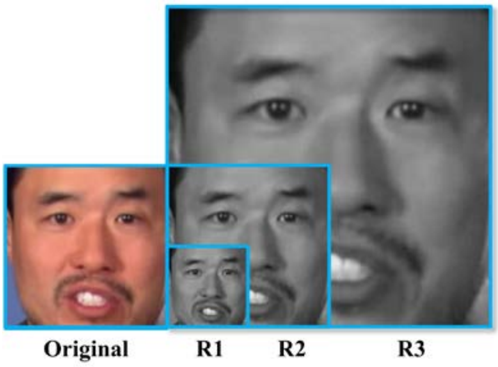


Fig. 5. An example of differences between resolution levels (R1: 64×64 pixels, R2: 128×128 pixels, and R3: 256×256 pixels)

3.2 Step 2: Experimental Configuration

In this step, we selected target datasets and DL model architecture for experimental analysis.

Table 2. The number of Celeb-DF dataset for component-wise analysis: AUROC, CM, and Detection Time

Dataset	[1]AUROC, [2]CM		[3]Detection Time	
	Real	Fake	Real	Fake
Training	412	5,299	-	-
Testing	178	340	82	40
Total	590	5,639	82	40

1) AUROC: Area Under ROC Curve
2) CM: Confusion Matrix
3) Detection Time: Preprocessing time + Inference time (in seconds)

Table 3. The number of datasets (Celeb-DF, FF++, and DF-TIMIT) for comparative analysis: AUROC and CM

Dataset	Celeb-DF		FF++		DF-TIMIT	
	Real	Fake	Real	Fake	Real	Fake
Training	412	5,299	700	700	224	224
Testing	178	340	300	300	96	96
Total	590	5,639	1,000	1,000	320	320

Table 4. The number of datasets (Celeb-DF, FF++, and DF-TIMIT) for comparative analysis: Detection Time

Dataset	Celeb-DF		FF++		DF-TIMIT	
	Real	Fake	Real	Fake	Real	Fake
Training	-	-	-	-	-	-
Testing	82	40	38	70	33	38
Total	82	40	38	70	33	38

Table 2, **Table 3**, and **Table 4** present the number of datasets for component-wise analysis and comparative analysis, respectively. Among the recently published deepfake datasets, we selected the top three datasets based on their citation counts (i.e., Celeb-DF [22], FF++ [21], and DF-TIMIT [20]). Celeb-DF comprises 5,639 fake videos generated by using the author's synthesis algorithm and 590 real videos collected from YouTube [22]. FF++ comprises 1,000 fake videos generated by using FaceSwap [47] and 1,000 real videos collected from YouTube synthesis algorithm and 590 real videos collected from YouTube [21]. Note that, to ensure an equal number of real and fake datasets within the FF++ dataset, we selected a subset (i.e., FaceSwap) from the complete fake dataset, following the setup of previous studies [22]. DF-TIMIT comprises 320 fake videos generated by using faceswap-GAN [48] and 320 real videos collected from the VidTIMIT dataset [49]. In addition, to evaluate the detection time, which comprises both dataset preprocessing time and inference time, considering that the video lengths in each dataset vary, we identified the length that accounts for the largest proportion in each dataset. We then constructed the datasets for detection time analysis using fixed video lengths: Celeb-DF (10-second, 122 videos), FF++ (20-second, 108 videos), and DF-TIMIT (3-second, 71 videos). For each analysis, different datasets were used. For component-wise analysis, we used the Celeb-DF dataset. For comparative analysis, we used Celeb-DF, FF++, and DF-TIMIT. The training time for those datasets ranged from 14.7 minutes to 499.1 minutes for 20 epochs.

As the target DL model architecture, we selected VGG16 [50], which is one of the most commonly used architectures in deepfake detection research [51–54]. Previous studies comparing the performance (i.e., accuracy and time) of various architectures in grayscale-based deepfake detection have demonstrated that VGG16 outperforms other architectures including XceptionNet, InceptionResNetV2, ResNet152, DenseNet121, and AlexNet [14, 17, 55].

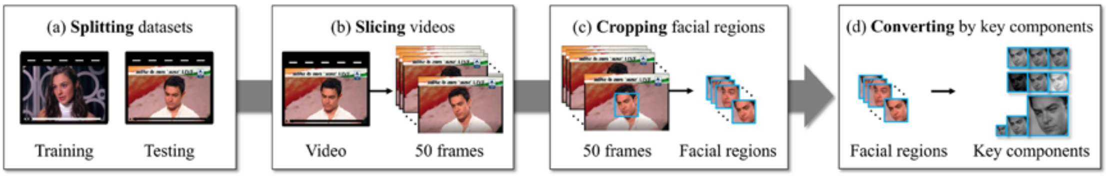


Fig. 6. An overview of dataset preprocessing: (a) Splitting datasets, (b) Slicing videos, (c) Cropping facial regions, and (d) Converting by key components

3.3 Step 3: Dataset Preprocessing

In this step, we performed preprocessing on the target dataset. **Fig. 6** shows four main steps for dataset preprocessing: (a) Splitting datasets: We split the target dataset into training and testing sub-datasets. We split the Celeb-DF dataset based on a type list provided by the dataset authors: Real—70% for training and 30% for testing, and Fake—94% for training and 6% for testing. Meanwhile, we split the FF++ and DF-TIMIT datasets into 70%:30% (training:testing) ratios, respectively. (b) Slicing videos: To appropriately capture any visual artifacts that may appear during the constant movement of an object, we sliced each video into 50 frames, the commonly used number of frames in video analysis [56–59]. (c) Cropping facial regions: Since facial regions are key subjects for deepfake generation, we extracted the facial region from each frame. Each facial region was cropped to a fixed aspect ratio of 1:1 and resized to our defined resolution levels R1, R2, and R3 [60, 61]. (d) Converting by key components: We finally converted each facial region based on the key components. For the RGB-to-grayscale conversion method, we converted each facial region using Intensity, Luma, and Luminance, respectively. For the brightness level in grayscale, we converted the brightness of each converted facial region using our defined brightness levels, B1, B2, and B3.

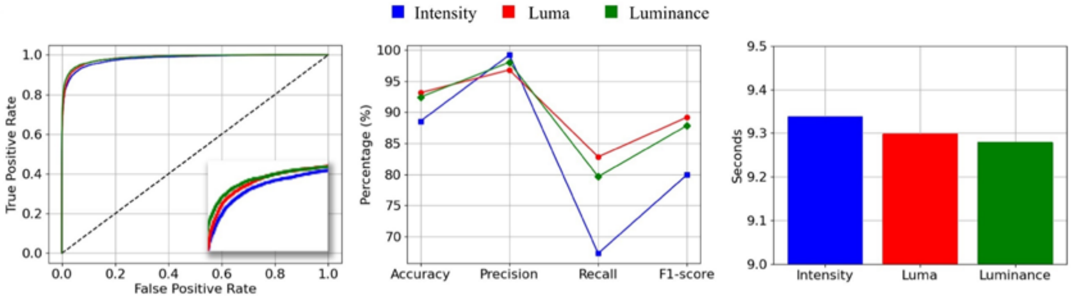
3.4 Step 4: Experimental Analysis

In this step, we performed two experimental analysis: component-wise analyses and comparative analysis.

For component-wise analysis, we compared the deepfake detection performances of different RGB-to-grayscale methods, brightness levels, and resolution levels. We trained separate VGG16 models using each preprocessed dataset and assessed their performance using the following evaluation metrics: AUROC, confusion matrix (accuracy, precision, recall, and F1-score), and detection time. Since grayscale-based deepfake detection inherently includes both grayscale conversion and inference, in this study, the detection time encompasses both preprocessing time and inference time for a single video. Consequently, for each key component, we quantitatively analyzed the performance of its variables and identified the differences between them.

Table 5. The results of component-wise analysis on RGB-to-grayscale conversion method

Conversion Method	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
Intensity	97.88	88.56	**99.22**	67.30	79.97	9.34
Luma	98.31	**93.14**	96.82	**82.83**	**89.20**	9.30
Luminance	**98.32**	92.46	98.04	79.65	87.87	**9.28**



(a) AUROC (b) Confusion matrix (c) Detection time

Fig. 7. The results of component-wise analysis on RGB-to-grayscale conversion method: (a) AUROC, (b) Confusion matrix, and (c) Detection time

For comparative analysis, we constructed an optimal combination of key components and compared its performance with existing deepfake detection solutions. The optimal combination was determined by component-wise analysis and included the best performing variables of each key component. We trained a VGG16 model using the preprocessed datasets and assessed its performance against those of existing solutions. To select relevant existing solutions, we conducted a systematic literature search of publications on deepfake detection since 2019 and thoroughly examined studies with a particular focus on those from the past three years [22, 44, 62–76]. As a result, we selected three solutions that have been most commonly used as comparative models: Xception [55], MesoNet [77], and FWA [78]. We compared their performances using the same metrics (i.e., AUROC, confusion matrix, and detection time). Note that, for existing solutions, we used pre- trained models essentially provided by the authors, and executed them in compliance with the conditions and parameters specified by the authors. Note that our experiments were performed using 2 GPUs (NVIDIA GeForce RTX 3090), Python 3.9.7, and TensorFlow 2.7.0 (optimizer=Adam, learning rate=1e-55, batch size=64 batch, and epochs=20).

4. Experimental Results

4.1 Component-wise Analysis

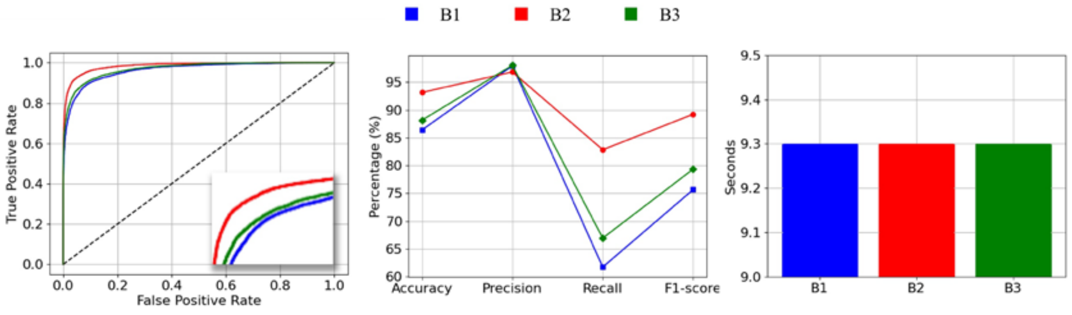
In this section, we demonstrate the results of the component-wise analysis. With a Celeb-DF dataset and VGG16 model architecture, we analyzed deepfake detection performance using multiple evaluation metrics, including AUROC, confusion matrix, and detection time. The detection time was calculated by averaging the time elapsed to analyze a single video of a specific length. Component-wise analysis was based on an ablation experiment to identify the impact of each key component's variable on deepfake detection. We measured the performance by fixing a control variable and varying other variables of each component. With reference to our preliminary study, the control variables for each key component were set as follows: RGB-to-grayscale conversion method=Luma; brightness level=B2; and resolution level=R3. Note that, in the component-wise analysis, the results were averaged after being measured three times due to the potential variation in the performance of the target model with each training instance.

4.1.1 RGB-to-grayscale Conversion Method

The analysis results on RGB-to-grayscale conversion methods are shown in **Table 5** and **Fig. 7**. Specifically, AUROC, confusion matrix (accuracy, precision, recall, and F1-score), and detection time of each RGB-to-grayscale conversion method are shown in **Fig. 7 (a)**, **(b)**, and **(c)**, respectively. For AUROC, the performance of Intensity (97.88%), Luma (98.31%), and Luminance (98.32%) was comparable, with Luminance slightly outperforming the others. For confusion matrix, Luma outperformed the others: For accuracy, Intensity (88.56%), Luma (93.14%), and Luminance (92.46%); For precision, Intensity (99.22%), Luma (96.82%), and Luminance (98.04%); For recall, Intensity (67.30%), Luma (82.83%), and Luminance (79.65%); For F1-score, Intensity (79.97%), Luma (89.20%), and Luminance (87.87%). For the detection time on 10-second videos, Luma (9.30s) and Luminance (9.28s) required less time overall than Intensity (9.34s). Overall, Luma showed the most accurate deepfake detection among the three conversion methods, while Luminance required the shortest detection time.

Table 6. The results of component-wise analysis on brightness level

Brightness Level	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
B1	95.41	86.38	97.92	61.70	75.64	**9.30**
B2	**98.31**	**93.14**	96.82	**82.83**	**89.20**	**9.30**
B3	96.58	88.14	**98.01**	66.92	79.33	**9.30**



(a) AUROC (b) Confusion matrix (c) Detection time

Fig. 8. The results of component-wise analysis on brightness level:
(a) AUROC, (b) Confusion matrix, and (c) Detection time

4.1.2 Brightness Level in Grayscale Channel

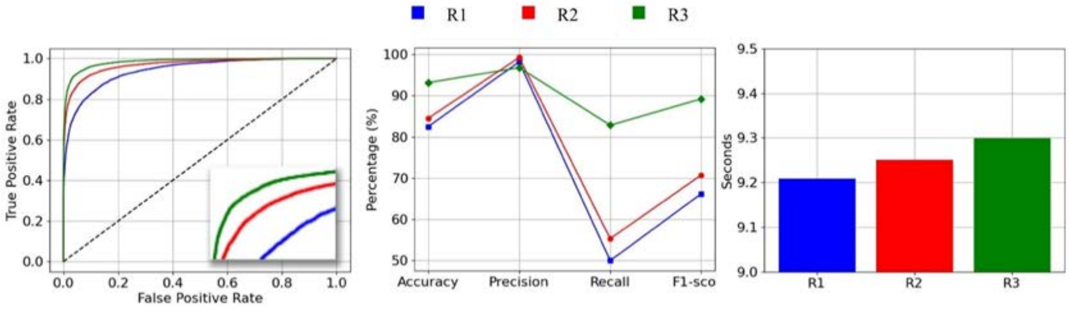
The analysis results on brightness levels are shown in **Table 6** and **Fig. 8**. Specifically, AUROC, confusion matrix (accuracy, precision, recall, and F1-score), and detection time of each brightness level are shown in **Fig. 8 (a)**, **(b)**, and **(c)**, respectively. For AUROC, B2 outperformed the others: B1 (95.41%), B2 (98.31%), and B3 (96.58%). For confusion matrix, B2 outperformed the others: For accuracy, B1 (86.38%), B2 (93.14%), and B3 (88.14%); For precision, B1 (97.92%), B2 (96.82%), and B3 (98.01%); For recall, B1 (61.70%), B2 (82.83%), and B3 (66.92%); For F1-score, B1 (75.64%), B2 (89.20%), and B3 (79.33%). For the detection time on 10-second videos, the performance of B1, B2, and B3 were identical (9.30s). Overall, B2 showed the most accurate deepfake detection among the three brightness levels, with comparable detection time between them.

4.1.3 Resolution Level in Grayscale Channel

The analysis results on resolution levels are shown in **Table 7** and **Fig. 9**. Specifically, AUROC, confusion matrix (accuracy, precision, recall, and F1-score), and detection time of each resolution level are shown in **Fig. 9 (a)**, **(b)**, and **(c)**, respectively. For AUROC, R3 outperformed the others: R1 (93.81%), R2 (96.62%), and R3 (98.31%). For confusion matrix, R3 outperformed the others: For accuracy, R1 (82.47%), R2 (84.47%), and R3 (93.14%); For precision, R1 (98.22%), R2 (99.26%), and R3 (96.82%); For recall, R1 (49.91%), R2 (55.25%), and R3 (82.83%); For F1-score, R1 (66.17%), R2 (70.71%), and R3 (89.20%). For the detection time on 10-second videos, R1 (9.21s) and R2 (9.25s) required less time overall than R3, which took the longest (9.30s). Overall, R3 showed the most accurate deepfake detection among the three resolution levels, while R1 required the shortest detection time.

Table 7. The results of component-wise analysis on resolution level

Resolution Level	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
R1	93.81	82.47	98.22	49.91	66.17	**9.21**
R2	96.62	84.47	**99.26**	55.25	70.71	9.25
R3	**98.31**	**93.14**	96.82	**82.83**	**89.20**	9.30



(a) AUROC

(b) Confusion matrix

(c) Detection time

Fig. 9. The results of component-wise analysis on resolution level:
(a) AUROC, (b) Confusion matrix, and (c) Detection time

4.2 Comparative Analysis

In this section, we demonstrate the results of the comparative analysis. First, we checked whether the combination of the identified key components is also globally optimal. We constructed all possible combinations based on the variables. With a Celeb-DF dataset and VGG16 model architecture, we analyzed deepfake detection performance using multiple evaluation metrics (i.e., AUROC, confusion matrix, and detection time), and identified the optimal combination (=OC) of key components.

Based on the OC, we trained a VGG16 model and compared its performance with existing solutions for deepfake detection (i.e., XceptionNet [55], Meso4 [77], and FWA [78]). We employed the same evaluation metrics as the component-wise analysis (i.e., AUROC, confusion matrix, and detection time), while employing different datasets (i.e., Celeb-DF, FF++, and DF-TIMIT). For each dataset, the detection time was calculated by averaging the detection time to analyze a single video of a specific length. Note that, in the comparative analysis, since we utilized pre-trained models provided by existing solutions, the best performance was measured instead of calculating the average.

4.2.1 Combinational Analysis of Key Components

Table 8. The results of combinational analysis of key components using Celeb-DF

Combination	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
Intensity / B1 / R1	94.50	80.22	**99.55**	42.62	59.69	9.27
Intensity / B2 / R1	95.52	87.14	97.90	63.96	77.37	9.27
Intensity / B3 / R1	93.45	85.57	95.75	60.71	74.30	9.27
Intensity / B1 / R2	96.21	87.47	96.59	65.85	78.31	9.29
Intensity / B2 / R2	97.28	90.63	97.60	74.55	84.53	9.29
Intensity / B3 / R2	97.46	91.42	97.33	77.13	86.07	9.29
Intensity / B1 / R3	96.89	89.93	97.18	72.80	83.18	9.34
Intensity / B2 / R3	97.88	88.56	99.22	67.30	79.97	9.34
Intensity / B3 / R3	96.55	86.29	99.34	60.51	75.12	9.34
Luma / B1 / R1	92.32	81.13	97.76	46.16	62.61	9.23
Luma / B2 / R1	93.81	82.47	98.22	49.91	66.17	9.21
Luma / B3 / R1	89.99	81.02	96.30	46.57	62.77	9.22
Luma / B1 / R2	95.59	83.70	99.02	53.12	69.07	9.25
Luma / B2 / R2	96.62	84.47	99.26	55.25	70.71	9.26
Luma / B3 / R2	94.94	82.87	99.09	50.63	66.99	9.25
Luma / B1 / R3	95.41	86.38	97.92	61.70	75.64	9.30
Luma / B2 / R3	**98.31**	**93.14**	96.82	**82.83**	**89.20**	9.30
Luma / B3 / R3	96.58	88.14	98.01	66.92	79.33	9.30
Luminance / B1 / R1	91.25	81.10	95.97	46.98	63.05	**9.20**
Luminance/ B2 / R1	92.27	82.04	97.67	48.94	65.07	9.21
Luminance / B3 / R1	88.59	77.97	97.86	36.70	53.34	9.21
Luminance / B1 / R2	95.60	86.31	96.53	62.57	75.51	9.23
Luminance / B2 / R2	97.62	88.07	98.99	66.04	78.94	9.23
Luminance / B3 / R2	94.68	83.70	98.44	53.43	69.17	9.23
Luminance/ B1 / R3	96.94	87.77	98.42	65.45	78.62	9.27
Luminance / B2 / R3	98.32	92.46	98.04	79.65	87.87	9.28
Luminance/ B3 / R3	95.92	85.31	98.11	58.49	72.94	9.28

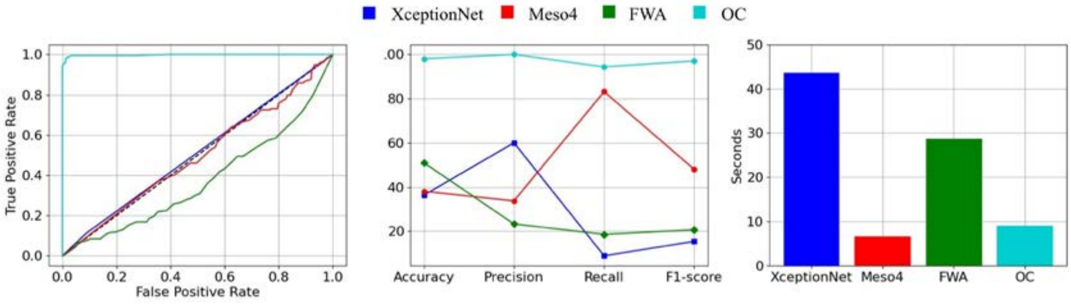
To verify whether the combination of the identified key components is also globally optimal, we constructed all possible combinations of variables (27 combinations in total) and preprocessed the Celeb-DF dataset by applying each combination. We then trained them on VGG16 models and measured their performance. Each metric was measured three times and averaged. As shown in **Table 8**, for AUROC, the combination of [Luminance / B2 / R3] (98.32%) and [Luma / B2/ R3] (98.31%) outperformed the others. For confusion matrix, [Luma / B2 / R3] outperformed the others overall: For accuracy, the combination of [Luma / B2 / R3] (93.14%) outperformed the others (77.97%-92.46%); For precision, the combination of [Intensity / B1 / R1] (99.55%) outperformed the others (95.75%-99.34%); For recall, the combination of [Luma / B2 / R3] (82.83%) outperformed the others (36.70%-79.65%); For F1-score, the combination of [Luma / B2 / R3] (89.20%) outperformed the others (53.34%-87.87%). For the detection time on 10-second videos, the combination of [Luminance / B1 / R1] (9.20s) required less time overall than the others (9.21s-9.34s). Overall, the OC includes RGB-to-grayscale conversion method=Luma, brightness level=B2, and resolution level=R3, all of which also provided superior performance in component-wise analysis. This confirms that the combination of these key components is indeed globally optimal.

4.2.2 Comparative Analysis: Celeb-DF

The results of comparative analysis using Celeb-DF dataset are shown in **Table 9**, and **Fig. 10**. Specifically, **Fig. 10 (a)**, **(b)**, and **(c)** show AUROC, confusion matrix, and detection time of each solution, respectively. For AUROC, OC outperformed the others: XceptionNet (48.79%), Meso4 (49.27%), FWA (37.27%), and OC (99.75%). For confusion matrix, OC outperformed the others: For accuracy, XceptionNet (36.29%), Meso4 (38.03%), FWA (50.97%), and OC (98.06%); For precision, XceptionNet (60.00%), Meso4 (33.71%), FWA (23.24%), and OC (100.00%); For recall, XceptionNet (8.82%), Meso4 (83.15%), FWA (18.54%), and OC (94.38%); For F1-score, XceptionNet (15.38%), Meso4 (47.97%), FWA (20.63%), and OC (97.10%).

Table 9. The results of comparative analysis using Celeb-DF

Solution	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
XceptionNet	48.79	36.29	60.00	8.82	15.38	43.72
Meso4	49.27	38.03	33.71	83.15	47.97	**6.85**
FWA	37.27	50.97	23.24	18.54	20.63	28.90
OC	**99.75**	**98.06**	**100.00**	**94.38**	**97.10**	9.26



(a) AUROC (b) Confusion matrix (c) Detection time

Fig. 10. The results of comparative analysis using Celeb-DF:
(a) AUROC, (b) Confusion matrix, and (c) Detection time

4.2.3 Comparative Analysis: FF++

The results of comparative analysis using FF++ dataset are shown in **Table 10**, and **Fig. 11**. Specifically, **Fig. 11 (a)**, **(b)**, and **(c)** show AUROC, confusion matrix, and detection time of each solution, respectively. For AUROC, OC outperformed the others: XceptionNet (57.00%), Meso4 (39.78%), FWA (33.93%), and OC (99.76%). For confusion matrix, OC outperformed the others: For accuracy, XceptionNet (57.00%), Meso4 (49.17%), FWA (37.67%), and OC

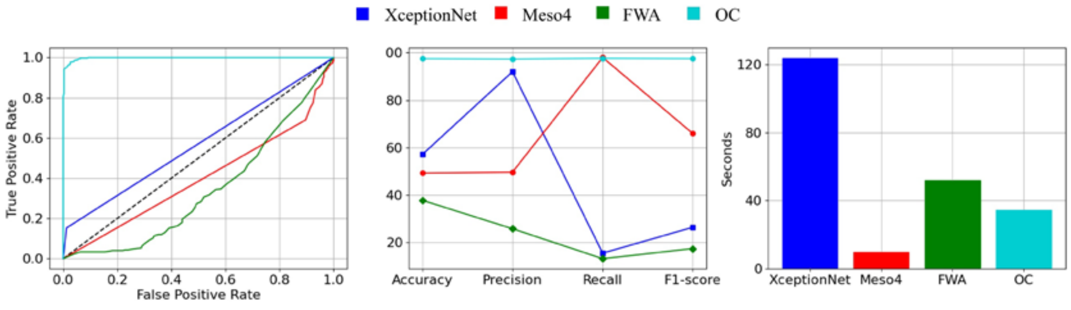
(97.50%); For precision, XceptionNet (92.00%), Meso4 (49.58%), FWA (25.66%), and OC (97.34%); For recall, XceptionNet (15.33%), Meso4 (98.00%), FWA (13.00%), and OC (97.66%); For F1-score, XceptionNet (26.29%), Meso4 (65.85%), FWA (17.26%), and OC (97.50%).

4.2.4 Comparative Analysis: DF-TIMIT

The results of comparative analysis using DF-TIMIT dataset are shown in **Table 11**, and **Fig. 12**. Specifically, **Fig. 12 (a)**, **(b)**, and **(c)** show AUROC, confusion matrix, and detection time of each solution, respectively. For AUROC, OC outperformed the others: XceptionNet (50.52%), Meso4 (50.80%), FWA (28.59%), and OC (99.92%). For confusion matrix, OC outperformed the others: For accuracy, XceptionNet (50.52%), Meso4 (50.00%), FWA (37.50%), and OC (98.43%); For precision, XceptionNet (100.00%), Meso4 (50.00%), FWA (40.16%), and OC (100.00%); For recall, XceptionNet (1.04%), Meso4 (94.79%), FWA (51.04%), and OC (96.87%); For F1-score, XceptionNet (2.06%), Meso4 (65.47%), FWA (44.95%), and OC (98.41%).

Table 10. The results of comparative analysis using FF++

Solution	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
XceptionNet	57.00	57.00	92.00	15.33	26.29	124.19
Meso4	39.78	49.17	49.58	**98.00**	65.85	**10.43**
FWA	33.93	37.67	25.66	13.00	17.26	52.54
OC	**99.76**	**97.50**	**97.34**	97.66	**97.50**	34.73

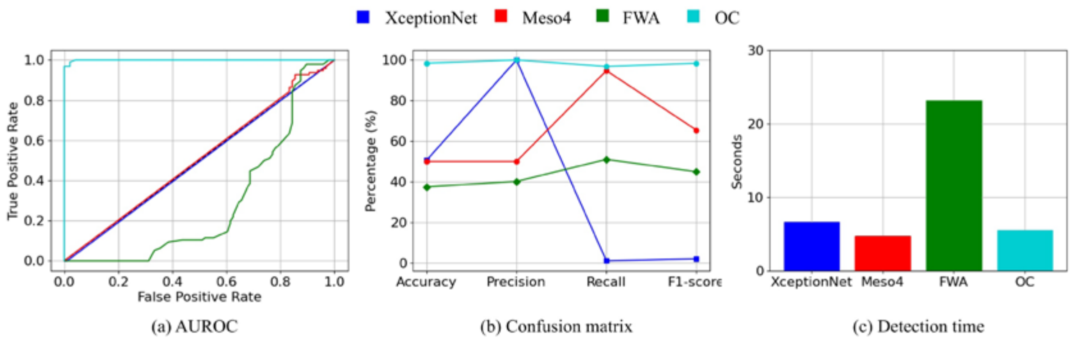


(a) AUROC (b) Confusion matrix (c) Detection time

Fig. 11. The results of comparative analysis using FF++:
(a) AUROC, (b) Confusion matrix, and (c) Detection time

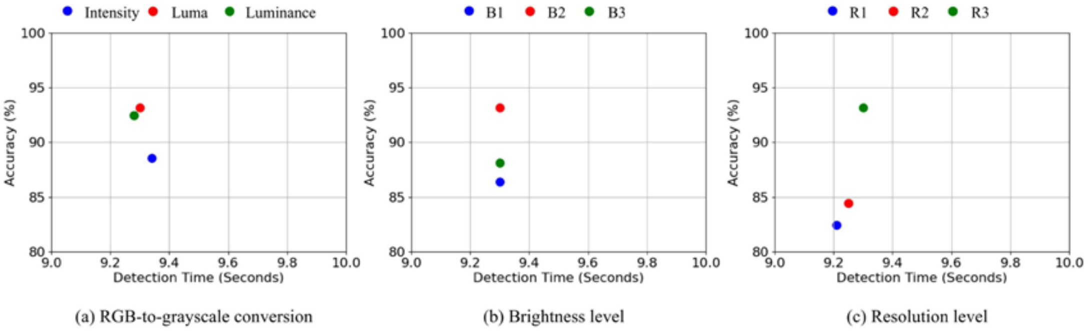
Table 11. The results of comparative analysis using DF-TIMIT

Solution	**AUROC** (%)	**Confusion Matrix**				**Detection Time** (*seconds*)
		Accuracy (%)	**Precision** (%)	**Recall** (%)	**F1-score** (%)	
XceptionNet	50.52	50.52	**100.00**	1.04	2.06	6.75
Meso4	50.80	50.00	50.00	94.79	65.47	**4.83**
FWA	28.59	37.50	40.16	51.04	44.95	23.27
OC	**99.92**	**98.43**	**100.00**	**96.87**	**98.41**	5.58



(a) AUROC (b) Confusion matrix (c) Detection time

Fig. 12. The results of comparative analysis using DF-TIMIT:
(a) AUROC, (b) Confusion matrix, and (c) Detection time



(a) RGB-to-grayscale conversion (b) Brightness level (c) Resolution level

Fig. 13. The correlation between detection time and accuracy in the results of component-wise analysis

5. Discussion

5.1 Component-wise Analysis

Based on the results of component-wise analysis, as shown in **Fig. 13**, we analyzed the correlations between the accuracy and detection time of each key component: (a) RGB-to-grayscale conversion method; (b) brightness level; and (c) resolution level. Note that each point on the graph represents the comprehensive performance of each key component, and that the upper-left region indicates better performance.

5.1.1 RGB-to-grayscale Conversion Method

As shown in **Fig. 13 (a)**, the accuracy of each RGB-to-grayscale conversion method was in the order of Luma, Luminance, and Intensity. While Luma and Luminance had a negligible difference (Luma-Luminance: 0.68%), Intensity and the others had relatively significant differences (Intensity-Luminance: 4.58% and Intensity-Luma: 3.90%). Multiple factors could have contributed to these results. When RGB channels are converted to grayscale, Intensity applies the same ratio to each channel (see **Table 1**). However, because each channel has a different spectral power, significant visual features may be lost during Intensity's conversion [79]. Moreover, in grayscale image recognition, Luma, which is based on non-linear channels with gamma correction, usually outperforms purely linear methods such as Intensity and Luminance [18].

Meanwhile, the detection time for each RGB-to-grayscale conversion method was in the order of Intensity, Luma, and Luminance, as shown in **Fig. 13 (a)**. Although Luminance required the shortest detection time, all methods had negligible differences (Luminance-Luma: 0.02s, Luminance-Intensity: 0.06s, and Luma-Intensity: 0.04s). RGB-to-Grayscale conversion methods are essentially the same operation with different weights. Although our experiments were conducted under strict constraints and conditions, we suspect that the slight difference in detection time may be attributed to uncontrollable external factors such as GPU operation. The results demonstrate that deepfake detection time is rarely impacted by the conversion method. Overall, based on the correlation between detection time and accuracy, Luma and Luminance were confirmed to be more effective than Intensity for detecting deepfakes.

5.1.2 Brightness Level in Grayscale Channel

As shown in **Fig. 13 (b)**, the accuracy of each brightness level was in the order of B2, B3, and B1. While B2 and the others had significant differences (B2-B3: 5.00% and B2-B1: 6.76%), B3 and B1 had a relatively comparable difference (B3-B1: 1.76%). Multiple factors could have contributed to these results. In overly bright or dark lighting conditions, it is challenging to accurately detect an object due to a significant loss of information [80]. Low brightness, in particular, causes severe information loss due to infrequent features, affecting object detection [81]. Therefore, B2, a moderate brightness condition, showed significantly higher accuracy compared to extreme brightness conditions (i.e., B1 and B3). Moreover, since lower brightness offers more challenging conditions for object detection than higher brightness [82], B3 might show slightly higher accuracy than B1.

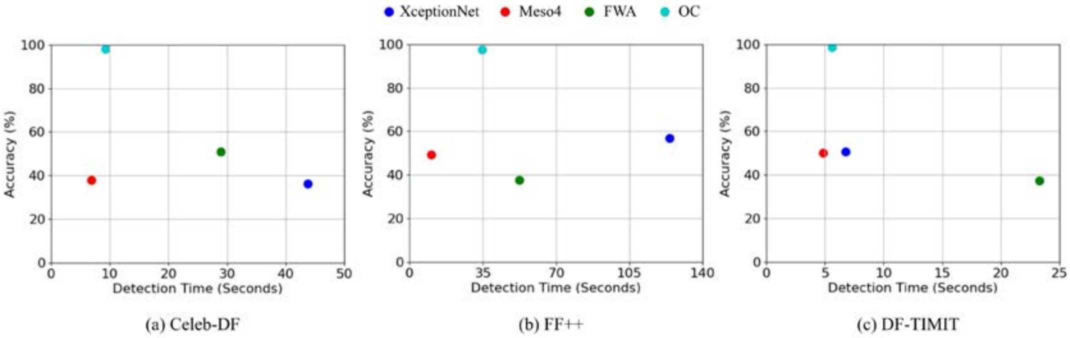


Fig. 14. The correlation between detection time and accuracy in the results of comparative analysis

Meanwhile, the detection time for each brightness level was identical at 9.30s, as shown in **Fig. 13 (b)**. Different brightness levels also have same operation with different weights. Therefore, the results demonstrate that deepfake detection time is not impacted by the brightness level. Overall, based on the correlation between detection time and accuracy, B2 was confirmed to be the most effective for detecting deepfakes.

5.1.3 Resolution Level in Grayscale Channel

As shown in **Fig. 13 (c)**, the accuracy of each resolution level was in the order of R3, R2, and R1. While R3 and the others had significant differences (R3-R2: 8.67% and R3-R1: 10.67%), R2 and R1 had a relatively comparable difference (R2-R1: 2.00%). High-resolution images are beneficial for image classification, including deepfake detection, because they contain more details and geometric information than low-resolution images [41, 83]. Thus, R3, the highest resolution, showed significantly higher accuracy than the lower ones (i.e., R2 and R1).

Meanwhile, the detection time for each resolution level was in the order of R3, R2, and R1. While R3 and the others had relatively significant differences (R3-R2: 0.05s and R3-R1: 0.09s), R2 and R1 had a negligible difference (R2-R1: 0.04s). The results demonstrate that deepfake detection time can be fairly impacted by the resolution level. Overall, based on the correlation between detection time and accuracy, B3 was confirmed to be the most effective for detecting deepfakes.

5.2 Comparative Analysis

Based on the results of comparative analysis, as shown in **Fig. 14**, we analyzed the correlations between the accuracy and detection time of each solution on different datasets: (a) Celeb-DF, (b) FF++, and (c) DF-TIMIT. Note that each point on the graph represents the comprehensive performance of each key component, and that the upper-left region indicates better performance.

As shown in **Fig. 14 (a)**, the accuracy of each solution on the Celeb-DF dataset was in the order of OC, FWA, Meso4, and XceptionNet. On the other hand, both on the FF++ and DF-TIMIT datasets, the accuracy of each solution was in the order of OC, XceptionNet, Meso4, and FWA, as shown in **Fig. 14 (b)**, and **(c)**. Overall, OC and the others had significant differences (OC-XceptionNet: 50.06%, OC-Meso4: 52.26%, and OC-FWA: 55.95%). When generating a deepfake, a decrease in image resolution occurs mainly at the boundary of the region being manipulated due to expanding the size of the source facial region to the size of the target facial region for realistic manipulation [84, 85]. Grayscale-based analysis has an advantage over RGB-based analysis in analyzing image boundaries because it simplifies image complexity and reduces the computing costs [86]. Moreover, in object detection, previous studies have proven that grayscale-based analysis has consistently higher accuracy than RGB-based analysis [87]. As a result, OC, which comprises key components of grayscale-based analysis, demonstrated noticeably superior accuracy when compared to the others.

As shown in **Fig. 14 (a)** and **(b)**, both on the Celeb-DF and FF++ dataset, the detection time of each solution was in the order of XceptionNet, FWA, OC, and Meso4. Meanwhile, the detection time of each solution on the DF-TIMIT dataset was in the order of FWA, XceptionNet, OC, and Meso4, as shown in **Fig. 14 (c)**. Overall, Meso4 showed the shortest detection time, and the difference between Meso4 and OC was relatively small (Meso4-OC: 9.15s), compared to Meso4 and the others (Meso4-FWA: 27.53s and Meso4-XceptionNet: 50.85s). Moreover, the mean difference between Meso4 and OC on the Celeb-DF and DF-TIMIT datasets was 1.58s, which is remarkably smaller than the others. In particular, since OC's detection time was primarily consumed by the process of converting RGB-to-grayscale, optimizing the conversion mechanism is expected to reduce it. Overall, based on the correlation between detection time and accuracy, we confirmed that OC outperformed existing solutions as it provided superior accuracy while requiring a relatively short detection time. Additionally, we confirmed that OC provided generalized performance across different datasets (Celeb-DF: 98.06%, FF++: 97.50%, and DF-TIMIT: 98.43%). Considering that grayscale-based deepfake detection achieves comparable performance with limited information, the results imply that OC can effectively counter emerging deepfake generation methods or different attack scenarios.

6. Conclusions and Future Works

In this paper, we identified the key components that directly affect the performance of grayscale-based deepfake detection: (1) RGB-to-grayscale conversion method, (2) brightness level in grayscale, and (3) resolution level in grayscale. To analyze their impacts on deepfake detection performance, we conducted comprehensive evaluations, including component-wise analysis and comparative analysis. The results of component-wise analysis confirmed that Luma (RGB-to-grayscale conversion method), B2 (brightness level–85.33 to 170.65), and R3 (resolution level–256×256 pixels) showed superior performance in terms of accuracy and detection time, respectively. Additionally, the results of comparative analysis confirmed that our grayscale-based model, which is based on the optimal combination of the key components, outperformed other solutions in terms of detection accuracy while requiring a relatively short detection time. Our study successfully confirmed that the key components we identified directly affect the performance of grayscale-based deepfake detection, and that their optimal combination outperforms existing solutions. Our findings will enable a decrease in cases of identity theft caused by deepfakes and an enhancement in the promptness and accuracy of presentation attack detection in face recognition systems.

In future work, we plan to improve the effectiveness of our optimal combination of key components by exploring and assessing additional potential factors that may influence the performance of grayscale-based deepfake detection. Moreover, we plan to conduct an extended analysis of emerging deepfake detection and generation approaches to validate the effectiveness and applicability of our optimal combination from a broader perspective. Furthermore, for further enhancement of detection accuracy, analyzing additional color spaces may prove beneficial, although this could potentially impose constraints in terms of speed. Therefore, we plan to explore methods that can effectively apply multiple color spaces in association while remaining applicable to real-time face recognition systems.

Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Grant RS-2022-00165648, in part by the 2023 Hongik University Research Fund, and in part by the BrainLink Program funded by the Ministry of Science and ICT through the National Research Foundation of Korea under Grant RS-2023-00237308.

References

[1] J. A. Costales et al., “The Impact of Blockchain Technology to Protect Image and Video Integrity from Identity Theft using Deepfake Analyzer,” in *Proc. of 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, pp.730-733, 2023. Article (CrossRef Link).

[2] M. Westerlund, “The Emergence of Deepfake Technology: A Review,” *Technology Innovation Management Review*, vol.9, no.11, pp.40-53, Nov. 2019. Article (CrossRef Link).

[3] M. Albahar et al., “Deepfakes: Threats and Countermeasures Systematic Review,” *Journal of Theoretical and Applied Information Technology*, vol.97, no.22, pp.3242-3250, Nov. 2019. Article(CrossRefLink).

[4] M. Kim et al., “FReTAL: Generalizing Deepfake Detection using Knowledge Distillation and Representation Learning,” in *Proc. of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.1001-1012, 2021. Article (CrossRef Link).

[5] A. Ismail et al., “A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost,” *Sensors*, vol.21, no.16, Aug. 2021. Article (CrossRef Link).

[6] D. Wodajo et al., “Deepfake Video Detection Using Convolutional Vision Transformer,” *arXiv:2102.11126*, 2021. Article (CrossRef Link).

[7] J. Liu et al., “A lightweight 3D convolutional neural network for deepfake detection,” *International Journal of Intelligent Systems*, vol.36, no.9, pp.4990-5004, Jun. 2021. Article (CrossRef Link).

[8] Y. Xu et al., “Supervised Contrastive Learning for Generalizable and Explainable DeepFakes Detection,” in *Proc. of IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pp.379-389, 2022. Article (CrossRef Link).

[9] P. Sebastian et al., “The effect of colour space on tracking robustness,” in *Proc. of 3rd IEEE Conference on Industrial Electronics and Applications*, pp.2512-2516, 2008. Article (CrossRef Link).

[10] M. Podpora et al., “YUV vs RGB – Choosing a Color Space for Human-Machine Interaction,” in *Proc. of Federated Conference on Computer Science and Information Systems*, vol.3, pp.29-34, 2014. Article (CrossRef Link).

[11] X. Chang et al., “DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network,” in *Proc.of 39th Chinese Control Conference (CCC)*, pp.7252-7256, 2020. Article (CrossRef Link).

[12] P. Saikia et al., “A Hybrid CNN-LSTM model for Video Deepfake Detection by Leveraging Optical Flow Features,” in *Proc. of International Joint Conference on Neural Networks (IJCNN)*, pp.1-7, 2022. Article (CrossRef Link).

[13] V. V. V. N. S Vamsi et al., “Deepfake detection in digital media forensics,” *Global Transitions Proceedings*, vol.3, no.1, pp.74-79, Jun. 2022. Article (CrossRef Link).

[14] S. B. Son et al., “A Measurement Study on Gray Channel-based Deepfake Detection,” in *Proc. of International Conference on Information and Communication Technology Convergence (ICTC)*, pp.428-430, 2021. Article (CrossRef Link).

[15] S. McCloskey et al., “Detecting GAN-Generated Imagery Using Saturation Cues,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp.4584-4588, 2019. Article (CrossRef Link).

[16] A. Pishori et al., “Detecting Deepfake Videos: An Analysis of Three Techniques,” *arXiv:2007.08517*, 2020. Article (CrossRef Link).

[17] S. B. Son et al., “A Comparative Study on Deepfake Detection using Gray Channel Analysis,” *Journal of Korea Multimedia Society*, vol.24, no.9, pp.1224-1241, 2021. Article(CrossRefLink).

[18] C. Kanan et al., “Color-to-Grayscale: Does the Method Matter in Image Recognition?,” *PLoS ONE*, vol.7, no.1, Jan. 2012. Artical (CrossRef Link).

[19] S. Yohanandan et al., “Saliency Preservation in Low-Resolution Grayscale Images,” in *Proc. of Computer Vision – ECCV 2018*, pp.237-254, 2018. Article (CrossRef Link).

[20] P. Korshunov et al., “DeepFakes: a New Threat to Face Recognition? Assessment and Detection,” *arXiv:1812.08685*, 2018. Article (CrossRef Link).

[21] A. Rössler et al., “FaceForensics++: Learning to Detect Manipulated Facial Images,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.1-11, 2019. Article (CrossRef Link).

[22] Y. Li et al., “Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3204-3213, 2020. Article (CrossRef Link).

[23] S. Ramachandran et al., “An Experimental Evaluation on Deepfake Detection using Deep Face Recognition,” in *Proc. of International Carnahan Conference on Security Technology (ICCST)*, pp.1-6, 2021. Article (CrossRef Link).

[24] A. Aghasanli et al., “Interpretable-through-prototypes deepfake detection for diffusion models,” in *Proc. of IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp.467-474, 2023. Article(CrossRefLink).

[25] T. Wang et al., “Noise Based Deepfake Detection via Multi-Head Relative-Interaction,” in *Proc. of the AAAI Conference on Artificial Intelligence*, vol.37, no.12, pp.14548-14556, 2023. Article (CrossRef Link).

[26] M. S. Rana et al., “Deepfake Detection: A Systematic Literature Review,” *IEEE Access*, vol.10, pp.25494-25513, Feb. 2022. Article (CrossRef Link).

[27] P. Yu et al., “A Survey on Deepfake Video Detection,” *IET Biometrics*, vol.10, no.6, pp.607-624, 2021. Article (CrossRef Link).

[28] A. M. Naser, “Color to grayscale image conversion based dimensionality reduction with Stationary Wavelet transform,” in *Proc. of Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*, pp.1-5, 2016. Article (CrossRef Link).

[29] S. Bezryadin et al., “Brightness Calculation in Digital Image Processing,” in *Proc. of International Symposium on Technologies for Digital Photo Fulfillment*, pp.10-15, 2007. Article (CrossRef Link).

[30] P. Ambalathankandy et al., “Warm-cool color-based high-speed decolorization: an empirical approach for tone mapping applications,” *Journal of Electronic Imaging*, vol.30, no.4, Aug. 2021. Article (CrossRef Link).

[31] M. E. Paramasivam et al., “Perceptually Weighted Color-to-Grayscale Conversion for Images with Non-Uniform Chromatic Distribution using Multiple Regression,” *ICTACT Journal on Image and Video Processing*, vol.11, no.2, pp.2325-2330, Nov. 2020. Article (CrossRef Link).

[32] E. Sirisha et al., “Image Retrieval Comparisons Using Color Models,” *International Journal of Advanced Research in Computer Science*, vol.4, no.9, pp.115-118, 2013. Article (CrossRef Link).

[33] S. Macêdo, G. Melo, and J. Kelner, “A Comparative Study of Grayscale Conversion Techniques Applied to SIFT Descriptors,” *SBC Journal on Interactive Systems*, vol.6, no.2, pp.30-36, 2015. Article (CrossRef Link).

[34] L. Zhang et al., “Color-to-Gray Conversion Based on Boundary Points,” in *Proc. of 11th International Conference on Communications, Circuits and Systems (ICCCAS)*, pp.241-245, 2022. Article (CrossRef Link).

[35] E. Karantoumanis et al., “Computational comparison of image preprocessing techniques for plant diseases detection,” in *Proc. of 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, pp.1-5, 2022. Article (CrossRef Link).

[36] S. Tabrizchi et al., “AppCiP: Energy-Efficient Approximate Convolution-in-Pixel Scheme for Neural Network Acceleration,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol.13, no.1, pp.225-236, Mar. 2023. Article (CrossRef Link).

[37] V. Ganchovska et al., “Converting Color to Grayscale Image Using LabVIEW,” in *Proc. of International Conference Automatics and Informatics (ICAI)*, pp.320-323, 2022. Article (CrossRef Link).

[38] C. H. Yeh et al., “Lightweight Deep Neural Network for Joint Learning of Underwater Object Detection and Color Conversion,” *IEEE Transactions on Neural Networks and Learning Systems*, vol.33, no.11, pp.6129-6143, Nov. 2022. Article (CrossRef Link).

[39] S. M. Newhall et al., “Final Report of the O.S.A. Subcommittee on the Spacing of the Munsell Colors,” *Journal of the Optical Society of America*, vol.33, no.7, pp.385-418, 1943. Article (CrossRef Link).

[40] S. H. Hong et al., “Image interpolation using interpolative classified vector quantization,” *Image and Vision Computing*, vol.26, no.2, pp.228-239, Feb. 2008. Article (CrossRef Link).

[41] R. Durall et al., “Unmasking DeepFakes with simple Features,” *arXiv:1911.00686*, 2019. Article(CrossRefLink).

[42] B, Dolhansky et al., “The Deepfake Detection Challenge (DFDC) Dataset,” *arXiv:2006.07397*, 2020. Article (CrossRef Link).

[43] L. Jiang et al., “DeeperForensics-1.0: A Large-Scale Dataset for Real-World Face Forgery Detection,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2886-2895, 2020. Article (CrossRef Link).

[44] P. Kwon et al., “KoDF: A Large-scale Korean DeepFake Detection Dataset,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.10724-10733, 2021. Article (CrossRef Link).

[45] B. Zi et al., “WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection,” in *Proc. of MM '20: Proceedings of the 28th ACM International Conference on Multimedia*, pp.2382-2390, 2020. Article (CrossRef Link).

[46] Y. Wang et al., "HifiFace: 3D Shape and Semantic Prior Guided High Fidelity Face Swapping," in *Proc. of the Thirtieth International Joint Conference on Artificial Intelligence*, pp.1136-1142, 2021. Article (CrossRef Link).

[47] MarekKowalski, MarekKowalski/FaceSwap:3D face swapping implemented in Python. https://github.com/MarekKowalski/FaceSwap

[48] Shaoanlu, shaoanlu/faceswap-GAN:A denoising autoencoder + adversarial losses and attention mechanisms for face swapping. https://github.com/shaoanlu/faceswap-GAN

[49] C. Sanderson et al., "Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference," in *Proc. of Advances in Biometrics: Third International Conferences, ICB 2009*, pp.199-208, 2009. Article (CrossRef Link).

[50] K. Simonyan et al., "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. of 3rd International Conference on Learning Representations, ICLR 2015*, 2015. Article (CrossRef Link).

[51] A. K. Dubey et al., "Automatic facial recognition using VGG16 based transfer learning model," *Journal of Information and Optimization Sciences*, vol.41, no.7, pp.1589-1596, Sep. 2020. Article (CrossRef Link).

[52] I. Amerini et al., "Deepfake Video Detection through Optical Flow Based CNN," in *Proc. of IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp.1205-1207, 2019. Article (CrossRef Link).

[53] A. Hamza et al., "Deepfake Audio Detection via MFCC Features Using Machine Learning," *IEEE Access*, vol.10, pp.134018-134028, Dec. 2022. Article (CrossRef Link).

[54] F. Ding et al., "Anti-Forensics for Face Swapping Videos via Adversarial Training," *IEEE Transactions on Multimedia*, vol.24, pp.3429-3441, 2022. Article (CrossRef Link).

[55] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1800-1807, 2017. Article (CrossRef Link).

[56] H. Ji et al., "Robust video denoising using low rank matrix completion," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.1791-1798, 2010. Article (CrossRef Link).

[57] A. Karpathy et al., "Large-Scale Video Classification with Convolutional Neural Networks," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1725-1732, 2014. Article (CrossRef Link).

[58] S. Yeung et al., "End-to-End Learning of Action Detection from Frame Glimpses in Videos," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2678-2687, 2016. Article (CrossRef Link).

[59] D. Min et al., "Depth Video Enhancement Based on Weighted Mode Filtering," *IEEE Transactions on Image Processing*, vol.21, no.3, pp.1176-1190, 2012. Article (CrossRef Link).

[60] Opencv, opencv/resize.cpp at 4.x · opencv/opencv. https://github.com/opencv/opencv/blob/4.x/module s/imgproc/src/resize.cpp

[61] OpenCV: Geometric Image Transformations. https://docs.opencv.org/4.x/da/d54/group__imgproc__transform.html

[62] L. Li et al., "Advancing High Fidelity Identity Swapping for Forgery Detection," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5073-5082, 2020. Article (CrossRef Link).

[63] Z. Sun et al., "Improving the Efficiency and Robustness of Deepfakes Detection Through Precise Geometric Features," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3608-3617, 2021. Article (CrossRef Link).

[64] Y. Zhu et al., "One Shot Face Swapping on Megapixels," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4832-4842, 2021. Article (CrossRef Link).

[65] Z. Chen et al., "MagDR: Mask-Guided Detection and Reconstruction for Defending Deepfakes," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.9010-9019, 2021. Article (CrossRef Link).

[66] X. Dong et al., “Protecting Celebrities from DeepFake with Identity Consistency Transformer,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.9458-9468, 2022. Article (CrossRef Link).

[67] X. Wang et al., “DeepFake Disrupter: The Detector of DeepFake Is My Friend,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.14900-14909, 2022. Article (CrossRef Link).

[68] L. Chen et al., “Self-supervised Learning of Adversarial Example: Towards Good Generalizations for Deepfake Detection,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.18689-18698, 2022. Article (CrossRef Link).

[69] K. Shiohara et al., “Detecting Deepfakes with Self-Blended Images,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.18699-18708, 2022. Article (CrossRef Link).

[70] T. Le et al., “OpenForensics: Large-Scale Challenging Dataset For Multi-Face Forgery Detection And Segmentation In-The-Wild,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.10097-10107, 2021. Article (CrossRef Link).

[71] Z. Cai et al., “MARLIN: Masked Autoencoder for facial video Representation LearnINg,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1493-1504, 2023. Article (CrossRef Link).

[72] B. Huang et al., “Implicit Identity Driven Deepfake Face Swapping Detection,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4490-4499, 2023. Article (CrossRef Link).

[73] K. Narayan et al., “DF-Platter: Multi-Face Heterogeneous Deepfake Dataset,” in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.9739-9748, 2023. Article (CrossRef Link).

[74] N. Larue et al., “SeeABLE: Soft Discrepancies and Bounded Contrastive Learning for Exposing Deepfakes,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.20954-20964, 2023. Article (CrossRef Link).

[75] Z. Yan et al., “UCF: Uncovering Common Features for Generalizable Deepfake Detection,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.22355-22366, 2023. Article (CrossRef Link).

[76] Y. Xu et al., “TALL: Thumbnail Layout for Deepfake Video Detection,” in *Proc. of IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.22601-22611, 2023. Article (CrossRef Link).

[77] D. Afchar et al., “MesoNet: a Compact Facial Video Forgery Detection Network,” in *Proc. of IEEE International Workshop on Information Forensics and Security (WIFS)*, pp.1-7, 2018. Article (CrossRef Link).

[78] Y. Li et al., “Exposing DeepFake Videos By Detecting Face Warping Artifacts,” in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. Article (CrossRef Link).

[79] A. Güneş et al., “Optimizing the color-to-grayscale conversion for image classification,” *Signal, Image and Video Processing*, vol.10, no.5, pp.853-860, 2016. Article (CrossRef Link).

[80] R. Mukherjee et al., “Object Detection Under Challenging Lighting Conditions Using High Dynamic Range Imagery,” *IEEE Access*, vol.9, pp.77771-77783, 2021. Article (CrossRef Link).

[81] H. K. Leung et al., “A Deep-Learning-Based Vehicle Detection Approach for Insufficient and Nighttime Illumination Conditions,” *Applied Sciences*, vol.9, no.22, Nov. 2019. Article (CrossRef Link).

[82] F. Mehmood et al., “Object detection mechanism based on deep learning algorithm using embedded IoT devices for smart home appliances control in CoT,” *Journal of Ambient Intelligence and Humanized Computing*, pp.1-17, Mar. 2019. Article (CrossRef Link).

[83] J. Zhao et al., “High-Resolution Image Classification Integrating Spectral-Spatial-Location Cues by Conditional Random Fields,” *IEEE Transactions on Image Processing*, vol.25, no.9, pp.4033-4045, Sep. 2016. Article (CrossRef Link).

[84] D. Güera et al., "Deepfake Video Detection Using Recurrent Neural Networks," in *Proc. of 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp.1-6, 2018. Article (CrossRef Link).

[85] C. M. Yu et al., "Detecting Deepfake-Forged Contents with Separable Convolutional Neural Network and Image Segmentation," *arXiv:1912.12184*, 2019. Article (CrossRef Link).

[86] I. Ahmad et al., "Color-to-grayscale algorithms effect on edge detection — A comparative study," in *Proc. of International Conference on Electronics, Information, and Communication (ICEIC)*, pp.1-4, 2018. Article (CrossRef Link).

[87] H. M. Bui et al., "Using grayscale images for object recognition with convolutional-recursive neural network," in *Proc. of IEEE Sixth International Conference on Communications and Electronics (ICCE)*, pp.321-325, 2016. Article (CrossRef Link).

Seok Bin Son is currently pursuing the Ph.D. degree in electrical and computer engineering at Korea University, Seoul, Republic of Korea. She received the B.S. degree in information security at Seoul Women's University, Seoul, Republic of Korea. Her research focuses include deep learning algorithms and their applications to information security.

Seong Hee Park received the M.S. degree in Computer Engineering at Hongik University, Seoul, Republic of Korea. She received the B.S. degree in information security at Seoul Women's University, Seoul, Republic of Korea. Her research focuses include deep learning algorithms and their applications to information security.

Youn Kyu Lee is currently an Assistant Professor in the Department of Computer Engineering at Hongik University, Seoul, Korea. He received the B.S. and M.S. degrees in computer science and engineering from Korea University, Seoul, Korea, in 2010 and 2012, respectively; and the Ph.D. degree in computer science from the University of Southern California (USC), Los Angeles, CA, USA, in 2017. Before joining Hongik University, he was with Samsung Advanced Institute of Technology (Suwon, Korea, 2018–2020), and Seoul Women's University (Seoul, Korea, 2020–2021). He was a recipient of Viterbi Graduate Fellowship with his Ph.D. admission from USC (2012), IEEE/ACM International Conference on Automated Software Engineering (ASE) Best Tool Paper Award (2018) and IEEE ICTC Best Paper Award (2022).