International Journal of Computer Applications (0975 - 8887)
Volume 186 - No.10, February 2024

An Adversarial Technique for Removal of JPEG Ghosts

Arkaprava Bhaduri Mandal
Department of Computer Applications
National Institute of Technology
Raipur, India

Deepak Agnihotri
Computing and Informatics Center
Dr Y S Parmar University of
Horticulture and Forestry Nauni-173230
Solan, HP, India

Tanmoy Kanti Das
Department of Computer Applications
National Institute of Technology
Raipur, India

ABSTRACT

The presence of statistical signatures of both double and single JPEG compression in an image indicates its maliciousness, and several techniques have been proposed using this fact to detect tempered images. One such statistical signature is known as JPEG ghosts where the presence of minima and local minima in the DCT coefficients of a JPEG image proves the existence of double JPEG compression. Moreover, it can localize the double JPEG compression quite successfully. However, this paper presents an adversarial method that can erase the statistical signature of the double JPEG compression without affecting the visual quality of the image. The proposed adversarial technique approximates the DCT coefficients using low-degree polynomials in such a way that no trace of prior JPEG compression can be detected. The transformed DCT coefficients exhibit statistical properties that are similar to uncompressed images. This technique successfully defeats the JPEG ghost-based forensic detection and localization method which raises serious concern regarding the robustness of the existing forensics schemes based on the JPEG artifacts.

General Terms

Image Forensics, JPEG Forgery

Keywords

Adversarial forensics, JPEG Compression, Approximation of DCT coefficients

1. INTRODUCTION

Frequent use of powerful image editing software to alter and enhance digital images has created an aura of mistrust around them. In recent times, multimedia forensics is gaining popularity to restore trust. A wide variety of forensics schemes are now available that identify the maliciousness of an image [19, 30, 23, 8, 4, 14, 31, 18, 32, 1, 15, 22]. Most forensics schemes rely on the statistical signatures left behind by different image processing operations to identify the maliciousness of an image. Some of these statistical signatures may not be robust enough to withstand the attempt of signature removal using adversarial methods. To unearth such weaknesses, a proper security evaluation of every forensics scheme should be carried out before its deployment in live systems. Consequently, the development of adversarial schemes got prominence to highlight the capabilities as well as weaknesses of existing forensics techniques [10, 29, 25, 21]. In this paper, an adversarial method has been proposed to highlight the weakness of the forensics scheme proposed by H. Farid [12].

JPEG is the most popular image compression technique that is supported by the majority of image acquisition devices. Due to this, a large number of forensic technique is proposed based on the statistical properties of JPEG compression. One such property that JPEG compression is not idempotent is widely used to detect and localize the tempered regions of a JPEG image. For example, digital tampering is sometimes performed by combining parts of different images to create a new JPEG image. This in turn may create an image where some portions of the newly produced image underwent double JPEG compression. As statistical signatures of single and double JPEG compression is quite distinct, forensics schemes could easily unearth the tempered regions (i.e., double JPEG compressed areas) of the image. It is obvious that if it is possible to erase or modify the statistical signatures of JPEG compression without affecting the image quality, the forensics schemes will fail to recognize the tampered regions. Several adversarial methods are available in the existing literature which is based on this principle.

One of the pioneering works in the field of adversarial forensics has been presented by Stamm et al. [29]. They have added random dither to the quantized DCT coefficients to remove statistical distortions present in the quantized DCT coefficients. The dispersal of the dither signal depends on the spread of DCT coefficients. It removes the traces of blocking operations performed by the JPEG compression. This simple adversarial method renders the Fan et al. [11] forensic scheme ineffective. However, the said anti-forensics operation leaves a trail of signatures on the image which can be utilized to identify the use of anti-forensics on the image. An improvement over Stamm et al. [29] scheme has been proposed by Qian et al. [25]. Signatures like blocking artifacts, comb-like histogram, and noise abnormality are removed using the Qian et al. technique.

Detection of double JPEG (DJPEG) compression performed using the same quantization matrix is quite difficult and the same has been first carried out by Huang et al. [13] using compression error analysis. A further enhanced version of the proposed scheme has been developed by Niu et al. [24]. They used truncation error as a key factor to differentiate between single and double compression. An adversarial method to defeat the scheme [13] has been proposed by Li et al. [20]. They have added an adaptive random dither to the DCT coefficients of a DJPEG image to hide the JPEG compression artifacts without degrading the visual quality of the image. The addition of random dither fools the detector completely. In a similar direction, Fan et al. [9] proposed a total variation (TV) based adversarial method for the removal of JPEG signatures. It consists of four steps: (1) a Total Variation (TV)-based deblocking; (2) perceptual DCT histogram smoothing; (3) TV-based deblocking; (4) de-calibration. This method diminishes the distortion in spatial as well as DCT domain and is effective against several forensics methods. An improvised version of the Fan et al. scheme with an enhanced denoising algorithm and deblocking technique has been proposed by Singh et al [28]. In [17] Kumar et al. added dither signal to the coefficients in the frequency domain by computing Discrete Fractional Cosine Transform with shifted block to conceal the JPEG compression history. This method provides better concealability and image quality as compared to the Singh et al. scheme.

Manipulating the traces of JPEG compression often degrades the quality of an image. Considering this, Fan et al. [10] developed a deconvolution framework for image tampering. The proposed method maintains better visual quality as well as forensic undetectability of the tampered image as compared to the preceding methods. Estimating the spatial heterogeneous convolution kernel is the trickiest part of the framework. Again, to overlay the footprints of JPEG compression Chu et al.[2] proposed a forging scheme that makes a trade-off among concealability, data rate and distortion. An adjustable dither noise was inserted in the DCT values of the manipulated image such that its histogram looks alike the histogram of the uncompressed image. The rate of forensics traces is decreased with the change in distortion, whereas it is increased with high-quality secondary JPEG compression. In [7], T. K. Das developed a counter-forensics technique that reduces the distortion as long as it is generated by the JPEG compression or the distortion can be identified in the DCT domain. The proposed scheme efficiently removes the JPEG footprints to fool the forensics detectors. A Convolutional Neural Network-based anti-forensics scheme has been proposed by Kim et al.in [16] to mislead the singly JPEG and doubly JPEG detectors. They have used the histogram loss function and deblocking loss function for better undetectability. These functions assisted the neural network to learn the distribution of DCT in uncompressed images that can be helpful to generate counter-forensically modified JPEG images.

To restore the trust in digital images several forensics schemes are available for ensuring the authenticity of an image. JPEG ghost detection [12] is one of the pioneers. Farid et al. [12] used the presence of local minima that appeared due to double quantization artifact for detection and localization of the malicious region (or JPEG ghost) of an image. The forensics detectors have a basic assumption that the forgers don't have the technical know-how regarding the basic image processing operation. Recent ongoing research trends void this assumption.

In this paper, an adversarial scheme has been proposed to produce doubly compressed JPEG images free from double quantization artifacts to deceive JPEG ghost detector [12]. JPEG ghost is the identifiable presence of a region(or regions) with a double quantization artifact in an image that appeared to be compressed only once. Our prime objective is to remove or alter the statistical feature of primary JPEG compression from the JPEG patch(or patches) and make it(or them) appear to be the uncompressed one before forgery(pasting it over the uncompressed image). After forgery, while the second compression takes place the statistical artifact of the forged region(or regions) will look alike the statistical artifact of single compression. This restricts the creation of double quantization artifact in the forged region(or regions) and the whole image bears the statistical signature of single compression with no traces of JPEG ghost. Further decalibration fine-tunes the process by removing the traces left at the boundaries of the forged region(or regions).

To simulate the proposed scheme, we choose to alter the most significant statistical signature of JPEG compression termed as "rate of zero coefficients". When an uncompressed image got JPEG compressed, the rate of zero coefficients in the DCT domain is increased due to the quantization process. The uncompressed version of that image has very few zero coefficients. This statistical feature is utilized by forensics researchers to differentiate between uncompressed images and JPEG compressed images. However, if we can reduce the rate of zero coefficients of a singly compressed image below the threshold limit, the singly compressed image will appear as an uncompressed one. To achieve this, we designed an approximation algorithm using a low-degree polynomial curve fitting to reduce the number of zero coefficients in a single JPEG. The algorithm approximates the non-zero values of zero coefficients in a range similar to the uncompressed one and substitutes them without much affecting the quality of the image.

The rest of the paper is comprised as follows: Section 2 presents a brief overview of the tamper detection techniques proposed by H. Farid [12]. Section 3 introduces the details of our proposed method. The Experimental setup and observations are discussed in Section 4. Finally, the paper is concluded in Section 5.

2. OVERVIEW OF JPEG GHOST DETECTION TECHNIQUE

Detection of forgery using JPEG artifacts is a common practice among forensics experts. In general, forgers used to combine different images to hide some details or to create a new one. Most of the time these participating images are of different qualities. While considering this, H. Farid [12] observed that, if any portion of the composite image previously gets compressed with a lower quality factor than the resultant image, it leaves a footprint. He coined the term 'JPEG Ghost' to refer the statistical footprint of prior quantization in doubly compressed JPEG images and used it to identify the forgery.

During JPEG ghost detection, the suspect image gets recompressed with different quality factors. For each quality factor, a difference image gets computed by considering the suspect image and its recompressed version to obtain the statistical footprint.

Instead of computing the difference between quantized DCT coefficients, they have used the difference in pixel values directly to analyze the cumulative effect of quantization using equation 1.

$$\Delta(\eta,\gamma,Q) = \frac{1}{3}\sum_{\iota=1}^{\iota=3}[f(\eta,\gamma,\iota) - f_Q(\eta,\gamma,\iota)]^2 \qquad (1)$$

Where $f(\eta,\gamma,\iota)$; $\iota \in (1,2,3)$ represents three RGB color channels, and $f_Q(\eta,\gamma,\iota)$ is the result of compressing $f(\eta,\gamma,\iota)$ at quality Q. The JPEG ghosts can be easily detected for grayscale images by computing $\Delta(\eta,\gamma,Q)$ for single color channel.

To reduce the influence of image content over the difference measuring procedure (detection accuracy), they computed a spatially averaged difference $\xi(\eta,\gamma,Q)$ over a $(b \times b)$ pixel region using

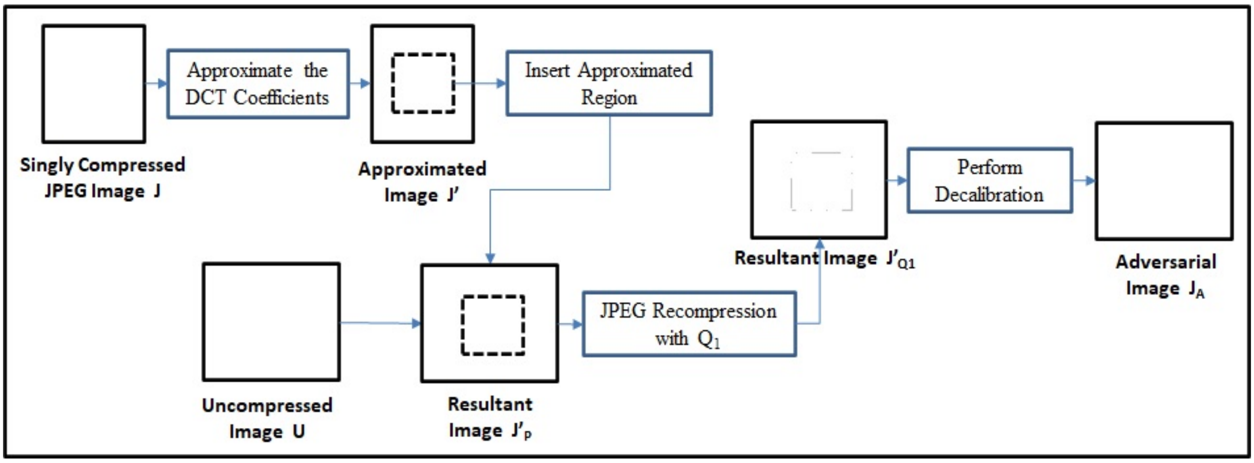

Fig. 1. Block diagram of the proposed Adversarial scheme

equation 2.

$$\xi(\eta,\gamma,Q)=\frac{1}{3}\sum_{\iota=1}^{\iota=3}\frac{1}{b^2}\sum_{b_\eta=0}^{b-1}\sum_{b_\gamma=0}^{b-1}[f(\eta+b_\eta,\gamma+b_\gamma,\iota) \\ -f_Q(\eta+b_\eta,\gamma+b_\gamma,\iota)]^2 \quad (2)$$

In the subsequent phase, $\xi(\eta,\gamma,Q)$ further gets normalized using equation 3 so that the difference $\xi(\eta,\gamma,Q)$ at each spatial location (η,γ) is scaled into the range [0,1]

$$\Delta(\eta,\gamma,Q)=\frac{\xi(\eta,\gamma,Q)-min_Q[\xi(\eta,\gamma,Q)]}{max_Q[\xi(\eta,\gamma,Q)]-min_Q[\xi(\eta,\gamma,Q)]} \quad (3)$$

Although the JPEG ghosts are highly salient in the case of visual inspection, it is still beneficial to identify the statistical difference of a specified region from the rest of the image. Thus, the two-sampled Kolmogorov-Smirnov (K-S) statistics [3] has been applied to check whether the distribution of pixel values in two regions of a difference-image is similar or different using equation 4.

$$\kappa=\max_{\nu}|\zeta_1(\nu)-\zeta_2(\nu)| \quad (4)$$

Where $\zeta_1(\nu)$ and $\zeta_2(\nu)$ represents the cumulative probability distributions of two specific regions in the computed difference $\Delta(\eta,\gamma,Q)$. Each value of Q gets considered separately. If the value of 'κ' reaches beyond a threshold limit, it ensures the presence of a JPEG ghost and maliciousness of an image. A generalized version of the JPEG Ghost detection algorithm proposed in [12] is presented in Algorithm 1.

3. PROPOSED ADVERSARIAL FRAMEWORK

If an image contains the statistical signature of single as well as double compression, it ensures the maliciousness of that image. Forensics experts use this feature to identify and localize forgery. Although, it has been observed that altering the statistical footprints without much disturbing the visual quality of an image is an easy task to fool the forensics detectors. Let us first describe the framework to defeat any double JPEG compression detection scheme. There may be two different approaches for removing the signature of double JPEG compression.

In one approach the attacker tries to remove the identified traces of JPEG compression from the doubly compressed JPEG image to make it behaves like an uncompressed one. However, the image quality degrades rapidly in this case.

In the second approach, the main objective is to remove the statistical footprints of JPEG compression from the singly compressed JPEG image to make it appear like an uncompressed one. Therefore, while recompression takes place the resultant image will bear only the signature of single JPEG compression and can easily evade the forensics detectors.

This paper presents an adversarial approach to remove the forensics traces of local minima in the distribution of quantized DCT coefficients to deceive the JPEG ghost detector[12].

Farid et al. [12] consider the very common scenario of forgery, where the forgers try to manipulate an image by inserting a JPEG patch (initially gets compressed with Q_0) over an uncompressed image. In the subsequent phase, the forgers try to resave it into JPEG with a different quality factor Q_1 to hide the traces of forgery. This results in the forged region to gets compressed twice, whereas the rest of the image gets compressed only once. The double compression leaves its statistical footprint or trail in the forged region.

H.Farid [12] used it in their forensics technique to detect the forged image. The suspect image I under forensics observation is re-saved with different quality factors Q_2 to obtain I_{Q_2}. The eq. 1 to 3 has been used to compute the difference between image I and its various quality factor variants I_{Q_2} for identifying the JPEG ghost.

As explained in Section 2, if the quality factor of some portion of a forged image is different from the rest of the image, the difference between I and I_{Q_2} will be minimum when $Q_0 = Q_2$ and $Q_1 = Q_2$. The presence of local minima due to $Q_0 = Q_2$ is referred as JPEG ghost, as it reveals the JPEG compression history. It also ensures that the coefficients were previously quantized (compressed) with a larger quantization step size(lower quality).

Farther, the identified forged location is verified statistically using the K-S statistics [3] as given in eq. 4. The K-S statistics computes the statistical difference between the forged region and the rest of the image. If the K-S statistics for any quality factor variants exceeded a specified threshold, the image will be classified as manipulated.

The proposed adversarial technique works in two phases to deceive the JPEG ghost detector. In the first phase, Algorithm 2 is used to remove the quantization effect of single compression from

Algorithm 1 Detection of JPEG Ghost in doubly compressed image.

```
Input: Suspicious JPEG image $I$.
Output: Forgery detection Result.
procedure JPEG GHOST DETECTION($I$)
    for $Q = 1$ to $100$ do
        Recompress the suspicious Image $I$ with quality factor Q to obtain $I_Q^R$.
        Compute the difference (in pixel values) between $I$ and $I_Q^R$ using equation 2 and 3.
        Compute the statistical difference $\kappa$ between the suspect region and rest of the image using equation 4
        if $\kappa \geq$ threshold then
            Presence of JPEG ghost detected. Mark the image as forged.
        else
            Mark the image as unchanged.
        end if
    end for
end procedure
```

Algorithm 2 Removal of quantization effect from a singly compressed JPEG image.

```
Input: Singly compressed JPEG image $J$.
Output: Approximated image $J^{'}$ of $J$ in any uncompressed format.
procedure APPROXIMATED IMAGE GENERATION($J$)
    Consider/Take the image $J$ into DCT domain.
    $I = J$
    for $\lambda \in Y, C_b, C_r$ do
        Read the block wise de-quantized DCT coefficients from $J$. Where each block $B$ is of dimension $8 \times 8$ and total no. of block is n.
        for $i = 1$ to $n$ do
            Copy the DCT coefficients of block $B_i$ into a 1D array$A_i$ by performing zigzag scanning.
            Sort $A_i$ in descending order and also maintain a vector table so that $B_i$ can be restored from $A_i$.
            Exclude the DC coefficient $A_i^0$ from farther processing.
            Divide $A_i^{1,\ldots,63}$ into two sub groups $G_v$ and $G_s$.
            $G_v$ consists of DCT coefficients from $A_i^{1,\ldots,mid}$ and $G_s$ consists of $A_i^{mid,\ldots,63}$. The value of $mid$ is computed by Eq. 5.
            if $\beta$ and $\tau$ be the first and last index of the DCT coefficients having value zero in sorted $A_i$. then
                $$mid = \left\lceil \frac{\beta + \tau}{2} \right\rceil \quad (5)$$
            end if
            Use Eq. 6 to fit the polynomial $p_1$ and $p_2$ of degree $d$ corresponding to the $G_v$ and $G_s$ respectively.
                $$p = ax^3 + bx^2 + cx + d \quad (6)$$
            Obtain the approximated DCT coefficients $G_v^{'}$ and $G_s^{'}$ from polynomial $p_1$ and $p_2$ respectively.
            Change the value of DC coefficient $A_i^0$ by $p$% to get $\bar{A}_i^0$.
            Combine $\bar{A}_i^0$, $G_v^{'}$ and $G_s^{'}$ to get $\bar{A}_i$.
            Obtain approximated DCT block $\bar{B}_i$ from $\bar{A}_i$ using vector table.
            if $\bar{B}_i$ and $B_i$ are visually same then
                Replace $B_i$ by $\bar{B}_i$ in $I_\lambda$.
            end if
        end for
    end for
    Perform Inverse DCT to take the approximated image $I$ in spatial domain. Resave it in any Uncompressed format to obtain $J^{'}$.
end procedure
```

a singly compressed JPEG image J, by approximating the DCT coefficients. Afterwards, a portion of the approximated image is going to be inserted in the uncompressed image U. The proposed algorithm extends the basic polynomial fitting algorithm presented in [5, 6] to approximate the DCT coefficients that restrict the generation of JPEG ghosts. The second phase follows the steps of Algorithm 3 that removes the forensic traces of double JPEG compression.

Algorithm 3 Removal of Forensics Traces of Double JPEG Compression.

Input: Uncompressed image U and Singly compressed JPEG image J_{Q_0}
Output: Attacked image J_A by removal of Forensics Traces of Double JPEG Compression.

```
procedure ATTACKED IMAGE GENERATION($U, J_{Q_0}$ )
    Get the approximated image $J^{'}$ from a singly compressed JPEG image $J_{Q_0}$ using Algorithm 2.
    Now,crop $J^{'}_{crp}$ from $J^{'}$ and paste it over $U$ to obtain $J^{'}_P$
    Re-compress $J^{'}_P$ with $Q_1$ to obtain a JPEG image $J^{'}_{Q_1}$, where $Q_1 > Q_0$.
    Apply the de-calibration process on $J^{'}_{Q_1}$ to get attacked image $J_A$.
end procedure
```

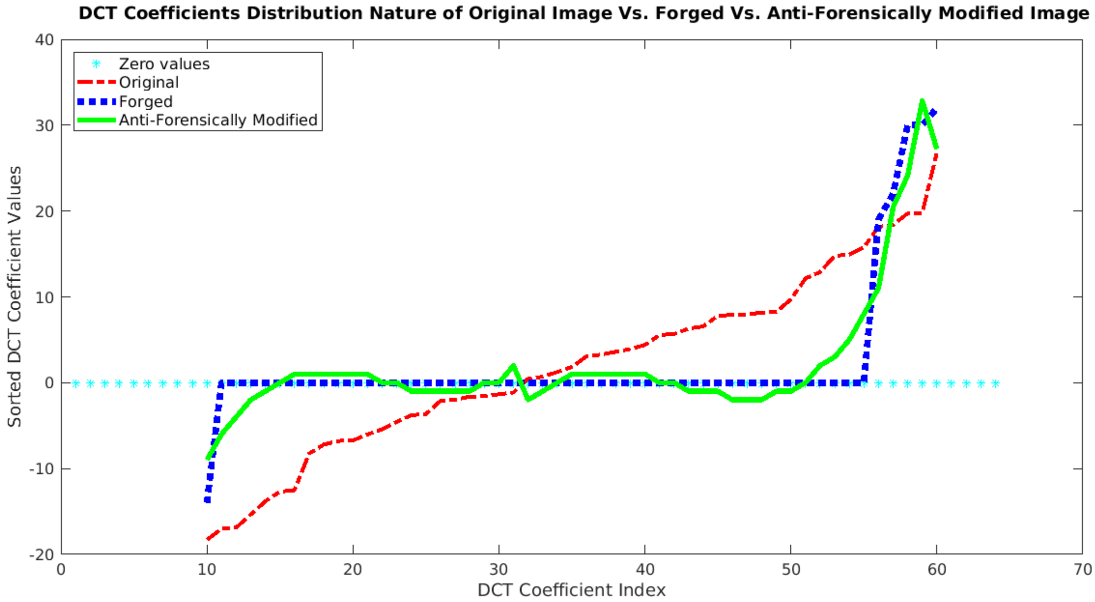

Fig. 2. Distribution of DCT Coefficients in Uncompressed, Singly compressed, and Approximated Image

Thus, the values of DCT coefficients in a forged image can be approximated in such a manner that its statistical property looks like an original image. Here, the statistical property refers to the total number of zero-valued DCT coefficients.

The concept behind the usage of Algorithms 2-3 can be explained using Fig. 2. It shows the distribution of DCT coefficients in originally uncompressed (i.e. UCID00001.tif), Singly compressed, and DCT approximated Singly compressed JPEG images. It can be observed from the figure that there is a very less number of zero-valued DCT coefficients in the original image, whereas a much higher number of zero-valued DCT coefficients are present in its singly compressed version. The total number of zero-valued DCT coefficients in the Approximated version of the singly compressed image is somehow closer to the uncompressed one.

Algorithm 2 is used to reduce the number of zero coefficients in singly compressed JPEG image J. The resultant image after applying this algorithm is an approximated image $J^{'}$, that is statistically similar to its original uncompressed version. It is difficult to differentiate between the approximated image and the original uncompressed image. Now, a portion $J^{'}_{crp}$ from $J^{'}$ gets cropped and inserted into an uncompressed image U. The resultant image $J^{'}_P$ is re-compressed using quality factor Q_1 (applying Algorithm 3) to obtain $J^{'}_{Q_1}$. There may be some forensics traces left at boundary positions where $J^{'}_{crp}$ is inserted. To remove such traces image $J^{'}_{Q_1}$ gets decalibrated. For decalibration, initially, the size of the image is increased by few rows and columns then further gets reduced to its original size to obtain the adversarial image J_A. Fig. 1 shows the block diagram of our proposed framework.

The proposed adversarial approach shows the existing loophole in [12]. It also demonstrates that if the image gets modified using the proposed adversarial technique,the forensics scheme [12] completely fails to detect JPEG ghost (tampered regions) in a forged image.

4. EXPERIMENTS AND OBSERVATIONS

The standard Uncompressed Color Image Database-(UCID-V2) [27] has been used for experimental evaluation of our proposed anti-forensics framework. The database comprised 1338 uncompressed TIFF images of size 512 × 384.

To simulate the forgery, a central portion from each uncompressed image gets cropped. Subsequently, the cropped portion gets JPEG compressed with quality factor Q_0. In the next step, the compressed portion is reinserted into the original uncompressed image, and the

Table 1. JPEG Ghost detection accuracy (%) before anti-forensics.

	Image Quality Difference ($Q_0 - Q_1$)			
	65 − 85		60 − 85	
Forged region	Accuracy(%)	Threshold	Accuracy(%)	Threshold
200×200	96.00	0.44	97.00	0.46
128×128	94.00	0.49	95.00	0.50

whole resultant image gets recompressed with quality factor Q_1 (where $Q_1 > Q_0$) to obtain the forged image.

JPEG toolbox [26] for MATLAB has been used to read the de-quantized DCT coefficients from a JPEG image. All other operations have been performed using the MatLab functions of image processing.

For generating the forged image dataset, two different size of forged region patches, i.e. 200×200 and 128×128 has been considered . For each type, the primary compression quality factor Q_0 is selected as 60 and 65 for the cropped region, whereas, the secondary quality factor Q_1 to compress the entire image is selected as 85. Thus, the difference between JPEG qualities Q_0 and Q_1 are 25 and 20 respectively.

Here, $Q_0 \leq Q_1$ implies that the initial quantization step size for the manipulated area is higher as compared to the secondary quantization step size for the remaining portion of the image. This tampering looks smooth for the human visual system and does not jumble any JPEG blocking statistics.

The proposed technique follows the assumptions made by [12] that the same JPEG quantization tables were used to create and test an image. Further, it was also assumed that there is no shift in the tampered region from its original JPEG block-lattice. The effect of these assumptions is not detrimental to the efficacy of detecting the JPEG ghosts.

For detection of JPEG ghost, the forged image iteratively gets re-compressed with quality Q_2 (where $Q_2 \in (30, 31, ..., 90)$). The difference between the image saved at Q_1 and each version of the image re-compressed at Q_2 was calculated using Eq. 3. The K-S statistics was used to compute the statistical dissimilarities between the image's central region and the rest of the image using Eq. 4. If the K-S statistics for any Q_2 exceeded a predefined threshold, the image was classified as tampered. The values of $Q_0 = 60, 65$, $Q_1 = 85$, and the forged region with size 200×200, 128×128 are considered for the experimental evaluation, as, the forensics technique proposed by [12] achieved maximum tamper detection accuracy at these combination.

Linear Support Vector Machine (LSVM) classifier has been applied on K-S statistics of original images and tampered images. Initially, original and tampered images are collected as a single image set. Further, it is divided into training and test sets using stratified 5-fold split criteria for cross-validation of the results. Table 4 shows the JPEG ghost detection accuracy and a threshold value when the proposed adversarial technique has not been applied to the forged images.

Figs. 3-6 show the Receiver Operating Characteristic (ROC) curve of JPEG ghost detection accuracy of images having forged region of size 200×200 and 128×128 whereas the image quality differences are of 25 and 20, i.e. $Q_0 = 60, 65$ and $Q_1 = 85$.

The LSVM classifier has been trained using K-S statistics of original and forged images and based on this trained model the class label of the adversarial attacked image has been predicted. Table 4 shows the JPEG ghost detection accuracy and Peak Signal-to-Noise Ratio (PSNR) of the manipulated region when the proposed adversarial technique and standard anti-forensics method [29] have been applied to the manipulated region. The PSNR values have been measured in decibel (dB) units and obtained by comparing

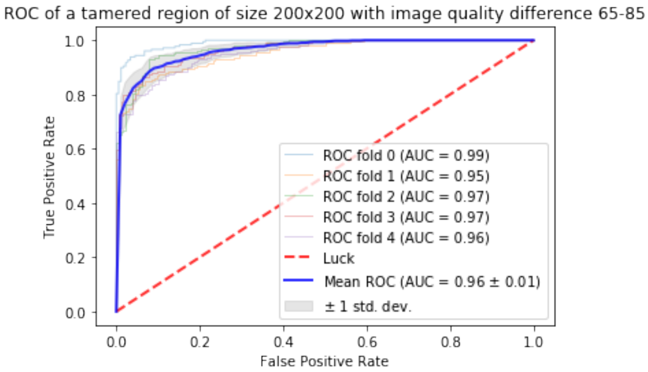

Fig. 3. JPEG Ghost detection accuracy for images having 200×200 forged region with 65-85 quality difference

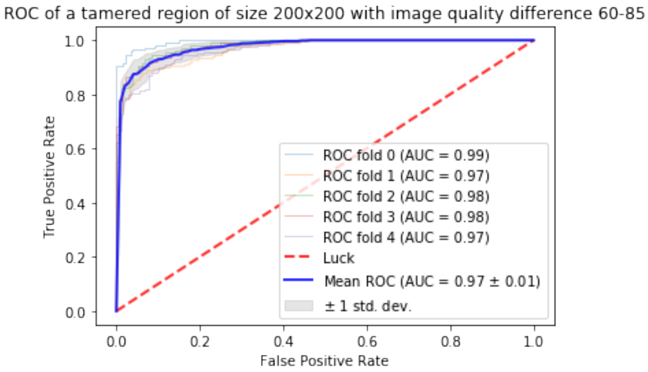

Fig. 4. JPEG Ghost detection accuracy for images having 200×200 forged region with 60-85 quality difference

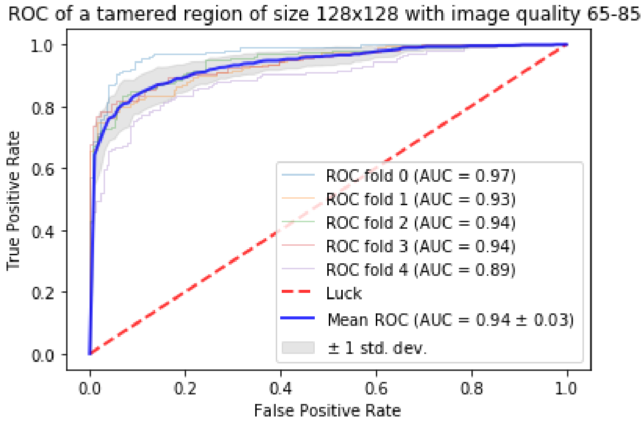

Fig. 5. JPEG Ghost detection accuracy for images having 128×128 forged region with 65-85 quality difference

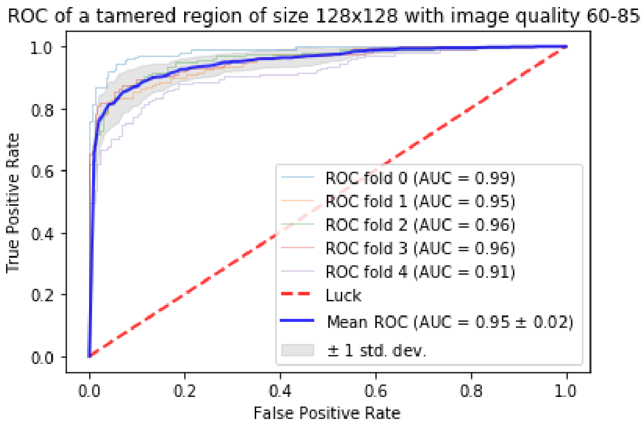

Fig. 6. JPEG Ghost detection accuracy for images having 128×128 forged region with 60-85 quality difference

Table 2. JPEG Ghost detection accuracy (%) and PSNR after anti-forensics.

	Image Quality Difference ($Q_0 - Q_1$)			
	$65-85$		$60-85$	
Attacked region	Accuracy(%)	PSNR(dB)	Accuracy(%)	PSNR(dB)
(a) Proposed Method				
200×200	46.34	37.83	45.27	37.60
128×128	49.21	37.87	48.43	37.67
(b) [29] Method				
200×200	23.23	27.32	17.17	27.09
128×128	33.33	27.63	34.34	27.51

the attacked image with original images of the UCID color images database.

4.1 Observations

JPEG ghost detection [12] technique is quite effective in explicit detection of a forged region that was initially compressed at a lower JPEG quality whereas the entire image including the forged region got re-compressed in a higher JPEG quality. The steps followed to detect such a region are as follows: (i) simply re-save the image under examination (let I) with multiple JPEG qualities (let I_Q, where $Q \in 30, 31, ..., 90$), (ii) Compute the difference between I and each version of I_Q by Eqs. 1-3, (iii) The computed difference is used to detect the local minima that are spatially localized. The observed minima are highly prominent and can be easily traceable by the forensics process. (iv) The difference between tampered region identified in the previous step and the rest of the image is computed using the K-S statistics by Eq. 4. The images of the standard UCID color image database [27] have been used for the experimental evaluation of the proposed framework. The standard counter-forensics[29] has been also applied to the images of the same database to compare the results with the proposed adversarial framework.

The forged images with manipulated regions of size 200 × 200 and image quality difference 20, 25 have achieved 96% and 97% accuracy respectively. Similarly, the forged images with a manipulated region of size 128 × 128 and image quality difference 20, 25 have achieved 94% and 95% accuracy respectively. It can be observed that for images with a larger forged region and higher quality difference, the ghost detector performs better. Whereas, with a smaller forged region and lower quality difference the accuracy of the detector falls. Farid et al. reported this observation in their work[12] and using 5 fold cross-validation we also validated the results.

While considering the attacked images generated using the proposed adversarial approach, the JPEG ghost detector behaves randomly. The detection accuracy for the attacked images with manipulated regions of size 200 × 200 and image quality difference 20, 25 have achieved 46.34% and 45.27% accuracy respectively. Similarly, the attacked images with manipulated regions of size 128 × 128 and image quality difference 20, 25 have achieved 49.21% and 48.43% accuracy respectively. It can be observed from the results that the detection accuracy of the JPEG ghost detector ranges between 45% to 50% in the case of attacked images, i.e. the classifier behaves like a random classifier. Also, the attack seems more powerful if the size of manipulated region increases. The results prove that the proposed adversarial technique is successful to make the classifier behaves like a random classifier which has earlier achieved 94% to 97% accuracy.

The experimental results of the standard [29] anti-forensics method are as follows. The detection accuracy of the attacked images with manipulated regions of size 200 × 200 and image quality difference 20, 25 have achieved 23.23% and 17.17% accuracy respectively. Similarly, the attacked images with manipulated regions of size 128 × 128 and image quality difference 20, 25 have achieved 33.33% and 34.34% accuracy respectively. Although Stamm et al.[29] scheme brings down the accuracy to a farther lower level, it is not capable to make the classifier behaves randomly as our proposed adversarial scheme does.

The adversarial images generated by our proposed scheme achieved higher PSNR values as compared to the standard Stamm et al. scheme. While considering the attacked images with manipulated regions of 200 × 200 size with image quality difference 20, 25, the observed PSNR values are 37.83 and 37.60 dB respectively for our proposed scheme. Whereas for Stamm et al. scheme-based attacked images the PSNR values are 27.32 and 27.09 dB respectively. Similarly, While considering the attacked images with manipulated regions of 128 × 128 size with image quality difference 20, 25, the observed PSNR values are 37.87 and 37.67 dB respectively for our proposed scheme. Whereas for Stamm et al. scheme-based attacked images the PSNR values are 27.63 and 27.51 dB respectively.

Thus, it can be observed that the PSNR values for the images generated by our proposed technique are much higher than the standard Stamm et al. [29] scheme. Therefore, the visual accuracy of attacked images modified by the proposed adversarial technique is much better as compared to the standard counter-forensics[29] method.

5. CONCLUSION

Traces of multiple JPEG compression in a single image indicates the presence of forgery. Several forensics schemes are proposed in the existing literature to detect such forgery by identifying the JPEG compression history. In this paper, one such scheme [12] is considered, that used the presence of local minima in DCT coefficients distribution to identify the forgery. The experimental evaluation demonstrated that the presence of local minima can easily be altered without degrading the visual quality of an image and the altered image can easily deceive the said JPEG ghost detection scheme. The proposed method efficiently distorts the statistical footprint of primary JPEG compression in a JPEG image. Thus, further recompression of that image with higher JPEG quality will not embed the statistical artifacts of double JPEG compression. The lack of statistical footprints makes the forensics detector blind. The proposed adversarial method can be further extended to deceive any double JPEG detection schemes by incorporating an improved DC approximation method.

6. REFERENCES

[1] Beijing Chen, Weijin Tan, Gouenou Coatrieux, Yuhui Zheng, and Yun-Qing Shi. A serial image copy-move forgery localization scheme with source/target distinguishment. *IEEE Transactions on Multimedia*, 23:3506–3517, 2021.

[2] Xiaoyu Chu, Matthew Christopher Stamm, Yan Chen, and KJ Ray Liu. On antiforensic concealability with rate-distortion tradeoff. *IEEE Transactions on Image Processing*, 24(3):1087–1100, 2015.

[3] William Jay Conover. Practical nonparametric statistics. 1980.

[4] Nandita Dalmia and Manish Okade. Robust first quantization matrix estimation based on filtering of recompression artifacts for non-aligned double compressed jpeg images. *Signal Processing: Image Communication*, 61:9 – 20, 2018.

[5] T. K. Das and S. Maitra. Cryptanalysis of correlation-based watermarking schemes using single watermarked copy. *IEEE Signal Processing Letters*, 11(4):446–449, April 2004.

[6] T. K. Das, S. Maitra, and J. Mitra. Cryptanalysis of optimal differential energy watermarking (DEW) and a modified robust scheme. *IEEE Trans. Signal Processing*, 53(2-2):768–775, 2005.

[7] Tanmoy Kanti Das. Anti-forensics of jpeg compression detection schemes using approximation of dct coefficients. *Multimedia Tools and Applications*, Jun 2018.

[8] Feng Ding, Yuxi Shi, Guopu Zhu, and Yun-Qing Shi. Smoothing identification for digital image forensics. *Multimedia Tools and Applications*, Nov 2018.

[9] Wei Fan, Kai Wang, Francois Cayre, and Zhang Xiong. Jpeg anti-forensics with improved tradeoff between forensic undetectability and image quality. *IEEE Transactions on Information Forensics and Security*, 9(8):1211–1226, 2014.

[10] Wei Fan, Kai Wang, François Cayre, and Zhang Xiong. Median filtered image quality enhancement and anti-forensics via variational deconvolution. *IEEE transactions on information forensics and security*, 10(5):1076–1091, 2015.

[11] Z. Fan and R. L. de Queiroz. Identification of bitmap compression history: Jpeg detection and quantizer estimation. *IEEE Transactions on Image Processing*, 12(2):230–235, Feb 2003.

[12] Hany Farid. Exposing digital forgeries from jpeg ghosts. *Trans. Info. For. Sec.*, 4(1):154–160, March 2009.

[13] Fangjun Huang, Jiwu Huang, and Yun Qing Shi. Detecting double jpeg compression with the same quantization matrix. *IEEE Transactions on Information Forensics and Security*, 5(4):848–856, 2010.

[14] Hui-Yu Huang and Ai-Jhen Ciou. Copy-move forgery detection for image forensics using the superpixel segmentation and the helmert transformation. *EURASIP Journal on Image and Video Processing*, 2019(1):68, Jun 2019.

[15] Sharanjit Kaur and Manpreet Kaur. Novel method for copy-move forgery detection. *International Journal of Computer Applications*, 174(18):10–14, Feb 2021.

[16] Dohyun Kim, Wonhyuk Ahn, and Heung-Kyu Lee. End-to-end anti-forensics network of single and double jpeg detection. *IEEE Access*, 9:13390–13402, 2021.

[17] Amit Kumar, Ankush Kansal, and Kulbir Singh. Anti-forensic approach for jpeg compressed images with enhanced image quality and forensic undetectability. *Multimedia Tools and Applications*, 79:8061–8084, 2020.

[18] Chothmal Kumawat and Vinod Pankajakshan. A robust jpeg compression detector for image forensics. *Signal Processing: Image Communication*, 89:116008, 2020.

[19] Thuong Le-Tien, Tu Huynh-Kha, Long Pham-Cong-Hoan, An Tran-Hong, Nilanjan Dey, and Marie Luong. Combined zernike moment and multiscale analysis for tamper detection in digital images. *Informatica*, 41(1), 2017.

[20] Haodong Li, Weiqi Luo, and Jiwu Huang. Anti-forensics of double jpeg compression with the same quantization matrix. *Multimedia Tools Appl.*, 74(17):6729–6744, September 2015.

[21] Yuanman Li and Jiantao Zhou. Anti-forensics of lossy predictive image compression. *IEEE Signal Processing Letters*, 22(12):2219–2223, 2015.

[22] Ocha Maulidya and Imam Riadi. Image forensics to detect image authenticity using error level analysis and noise analysis methods. *International Journal of Computer Applications*, 185(28):6–11, Aug 2023.

[23] Morteza Nasiri and Alireza Behrad. Using expectation-maximization for exposing image forgeries by revealing inconsistencies in shadow geometry. *Journal of Visual Communication and Image Representation*, 58:323 – 333, 2019.

[24] Yakun Niu, Xiaolong Li, Yao Zhao, and Rongrong Ni. An enhanced approach for detecting double jpeg compression with the same quantization matrix. *Signal Processing: Image Communication*, 76:89–96, 2019.

[25] Zhenxing Qian and Xinpeng Zhang. Improved anti-forensics of jpeg compression. *Journal of Systems and Software*, 91:100–108, 2014.

[26] Phil Sallee. Matlab jpeg toolbox, 2003.

[27] Gerald Schaefer and Michal Stich. Ucid: An uncompressed color image database. In *Storage and Retrieval Methods and Applications for Multimedia 2004*, volume 5307, pages 472–481. International Society for Optics and Photonics, 2003.

[28] Gurinder Singh and Kulbir Singh. Improved jpeg anti-forensics with better image visual quality and forensic undetectability. *Forensic Science International*, 277:133–147, 2017.

[29] Matthew C Stamm and KJ Ray Liu. Anti-forensics of digital image compression. *IEEE Transactions on Information Forensics and Security*, 6(3):1050–1065, 2011.

[30] Diaa Uliyan, Mohammad AM Abushariah, and Ahmad Mousa Altamimi. Blur invariant features for exposing region duplication forgery using anms and local phase quantization. *Informatica*, 42(4), 2018.

[31] Jinwei Wang, Wei Huang, Xiangyang Luo, Yun-Qing Shi, and Sunil Kr Jha. Non-aligned double jpeg compression detection based on refined markov features in qdct domain. *Journal of Real-Time Image Processing*, 17(1):7–16, 2020.

[32] Peiyu Zhuang, Haodong Li, Shunquan Tan, Bin Li, and Jiwu Huang. Image tampering localization using a dense fully convolutional network. *IEEE Transactions on Information Forensics and Security*, 16:2986–2999, 2021.