BlessMark: a blind diagnostically-lossless watermarking framework for medical applications based on deep neural networks

Hamidreza Zarrabi, et al. [full author details at the end of the article]

Received: 20 January 2019 / Revised: 19 December 2019 / Accepted: 28 January 2020 / Published online: 24 May 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Nowadays, with the development of public network usage, medical information is transmitted throughout the hospitals. A watermarking system can help for the confidentiality of medical information distributed over the internet. In medical images, regions-of-interest (ROI) contain diagnostic information. The watermark should be embedded only into non-regions-of-interest (NROI) regions to keep diagnostically important details without distortion. Recently, ROI based watermarking has attracted the attention of the medical research community. The ROI map can be used as an embedding key for improving confidentiality protection purposes. However, in most existing works, the ROI map that is used for the embedding process must be sent as side-information along with the watermarked image. This side information is a disadvantage and makes the extraction process non-blind. Also, most existing algorithms do not recover NROI of the original cover image after the extraction of the watermark. In this paper, we propose a framework for blind diagnostically-lossless watermarking, which iteratively embeds only into NROI. The significance of the proposed framework is in satisfying the confidentiality of the patient information through a blind watermarking system, while it preserves diagnostic/medical information of the image throughout the watermarking process. A deep neural network is used to recognize the ROI map in the embedding, extraction, and recovery processes. In the extraction process, the same ROI map of the embedding process is recognized without requiring any additional information. Hence, the watermark is blindly extracted from the NROI. Furthermore, a three-layer fully connected neural network is used for the detection of distorted NROI blocks in the recovery process to recover the distorted NROI blocks to their original form. The proposed framework is compared with one lossless watermarking algorithm. Experimental results demonstrate the superiority of the proposed framework in terms of side information.

Keywords ROI-based watermarking · Blind watermarking · CNN · Fully connected neural network · Deep neural network

1 Introduction

Transmission of Electronic Patient Record (EPR) between medical research centers, schools, and hospitals is necessary. Sending of EPR is essential for educational purposes and e-
healthcare applications. Many applications, such as teleconsulting, telediagnosis, and remote surgery, require EPR. EPR encompasses confidential medical information about the patient, medical situation, diagnosis, and treatment [8]. Since EPR is transmitted through a public network such as the Internet, the confidentiality protection of EPR is a real concern in the Health Information System (HIS). Watermarking is one solution to the confidentiality protection problem in HIS. Besides watermarking, various other methods have been proposed for improving the confidentiality protection of EPR. One approach is to embed encrypted EPR into a cover image with encryption algorithms such as chaotic encryption [25] and compressive sensing [34]. Another scheme is to insert the EPR into selected areas of the cover image rather than the whole cover image and use the binary location map as a key for embedding [9, 36]. In the extraction process, the binary key used for the embedding process is required to extract the embedded EPR.

In the watermarking systems, capacity and imperceptibility are two vital concerns [10]. One concern in the watermarking schemes is to maximize embedding capacity. Another concern is to avoid the watermarked image to suffer subtle distortions after the embedding process. Imperceptibility means that the watermarked image should be perceptually indistinguishable from the original cover image. Another concern in watermarking systems is the concept of blindness [10]. Blind systems have been thoroughly investigated [6, 12, 23, 30], and they generally lead to higher complexity. However, blind systems are practically superior to non-blind systems, which require various side information for the extraction or recovery on the receiver side [3, 17, 31].

Due to the rapid growth of machine learning and artificial intelligence in the last decades, several research works have used machine-learning tools for watermarking [2, 13, 18, 20, 22, 33]. Some researches propose prediction based watermarking methods for natural images [22, 33]. In natural images, pixels have a high correlation with their neighbors and can be predicted based on neighbor values. The prediction process may utilize various machine-learning tools such as the Extreme Learning Machine (ELM) [33] and Lagrangian Support Vector Regression (LSVR) [22]. Heidari et al. [13] embed the watermark data redundantly in multiple spectral zones. For extraction, they use SVM to detect a zone with minimum distortion, from which watermark is extracted. Abdelhakim et al. [2], use K-Nearest Neighborhood (K-NN) regression to find the optimum value of the embedding parameter. Among machine learning tools, deep Convolutional Neural Networks (CNN) have become popular in many computer vision applications such as object detection [27], pattern recognition [39], image classification [15], and watermarking [18, 20]. For instance, Kandi et al. [18] use auto-encoders for feature extraction. In their method, watermark data is embedded in the feature space, generated by the auto-encoder, rather than embedding in a transform domain such as DCT.

Medical images are divided into two regions, ROI and NROI. Since the ROI region includes critical information for diagnostic purposes, small distortions of ROI may cause problems in medical diagnosis. On the other hand, NROI contains nonessential medical information. Hence, small deformations of NROI are tolerable and do not cause serious issues. In watermarking systems introduced by [2, 3, 13], the whole cover image is considered for embedding, and this distortion caused during embedding is irreversible. Two approaches for medical watermarking are utilized to assure that diagnostic information is not distorted during watermarking or undesired distortion can be restored. The first approach is ROI-based watermarking, in which the cover image is divided into two regions, ROI and NROI. Then the watermark is embedded only in NROI. Thus, ROI remains intact, and only NROI is distorted during embedding. The second approach is
the lossless watermarking. In this approach, the whole cover image is considered for embedding with lossless methods, so that in the recovery process, the original cover image is recovered without any loss of information.

One limitation of ROI based watermarking approaches is that they cannot detect the ROI map on the receiver side. Consequently, the ROI map needs to be sent as side information to the receiver side for the extraction of the watermark [11, 32]. Hence, regardless of their segmentation method, they are non-blind algorithms, which require side information for the extraction of watermark data or the recovery of the original cover image. On the other hand, some other researches [7, 26, 37] do not send an ROI map as side information. Therefore, they are blind algorithms, and the watermark data can be extracted solely from the watermarked image. Yang et al. [37] and Chaitanya et al. [7] embed ROI map inside the watermarked image to make it blind. However, the watermarking capacity is reduced as the embedded ROI map occupies part of data capacity. The most relevant research to our work is [19]. Their algorithm recognizes the ROI map from the watermarked image on the receiver side and does not need to send an ROI map as additional information. They segment the image by Otsu and then utilize a histogram-shifting method for the embedding of the watermark. This embedding method processes pixel values that are close to the peak bin. Hence, the ROI map is not changed during the embedding process and can be recognized before the extraction process. Despite the elegant idea suggested in this work, it is only applicable to the specific watermarking method utilized in this work. Hence, it cannot be generalized for other watermarking approaches. Furthermore, the method of [19] lacks a wide variety of experimental evaluations, as it is only tested on three medical image samples.

In lossless watermarking approaches, the whole cover image is distorted during embedding since all regions of the image are considered for embedding. However, due to utilizing lossless methods, the original cover image is entirely recovered after watermark extraction [36], [4, 5, 14, 16, 21, 35, 38]. Despite complete recovery and extraction in the receiver side, some papers are non-blind, implying that some side information from the embedding module is required by the extraction or recovery modules on the receiver side [36], [4, 5, 38], [21].

In this paper, we propose a recursive method for blind diagnostically-lossless watermarking in medical applications, which is named BlessMark. We use a deep network for segmentation and generation of ROI map. Watermark is only embedded in NROI blocks within a novel iterative scheme. Therefore, sensitive medical information remains intact, which leads to a diagnostically-lossless watermarking system, where only NROI is distorted during the embedding process. The proposed framework is blind, as neither of the ROI map or the original cover image and information about watermark is transmitted to the receiver side. The utilized segmentation procedure helps further improve the confidentiality of the embedded information, due to the proprietary network weights, which are not publicly known by others. In other words, the ROI map may be considered as a key for confidentiality since watermark data is extracted solely from NROI blocks.

Furthermore, we train a classifier to detect distorted NROI regions. Hence, distorted areas of NROI are mostly recovered to the original form.

To test our proposed framework, we use a simple embedding method in the DCT domain. The choice of the DCT domain for embedding is just a convenient option for the proof of concept. Hence, different embedding methods in other transform domains, such as DWT and Hadamard, may be applied. Also, to evaluate the framework, we use a CNN for ROI segmentation and a three-layer fully connected neural network for the detection of distorted NROI blocks.
The main contributions of this work are as follows:

1) Introducing a framework that can be used as a platform for applying a desirable watermarking method to any medical image by utilizing various network structures in different transform domains.
2) In the extraction process, the same ROI map, which is used for the embedding process, is generated without any additional information. Therefore, the watermark is blindly extracted from NROI blocks.
3) NROI blocks, which are modified due to the embedding process, are recovered to their original state wherever possible.

The remainder of the paper is organized as follows: Details of the proposed framework is explained in Section 2. The experimental results are explained in Section 3. Finally, we conclude our proposed framework in Section 4.

2 General Watermarking Framework

In this section, we go through the technical details of the proposed watermarking framework, BlessMark. As shown in the block diagram of Fig. 1, the framework is composed of three separate modules for the embedding of the watermark, extraction of the watermark, and the recovery of the original cover image. There is a standard core block, ROI segmentation network, which is responsible for segmentation the ROI regions and generating the ROI map for the embedding, extraction and recovery modules. The problem is that embedding of the watermark data into the image may cause distortions, which could affect the segmentation results. We embed the watermark data into NROI. Hence, the ROI map is key for the extraction process. It is crucial to obtain the same segmentation on the watermarked image; otherwise, accurate extraction of the watermark data is not possible. To this end, we propose a novel iterative scheme for embedding based on the ROI segmentation function. This is done such that the ROI segmentation network can accurately detect the same ROI map in both transmitter and receiver sides. ROI map is automatically recognized on the receiver side without requiring any additional information. Therefore, the ROI map works as a key for confidentiality, as any third party cannot recognize it without having access to our proprietary segmentation tool.

On the other hand, due to embedding watermark data in NROI, crucial medical information in ROI remains intact. Another core block of the framework is a distortion detection network used in the recovery module. This network is responsible for the detection of distorted blocks due to the embedding process for the recovery operation in the next step.

In the following sections, we elaborate on the details of each block separately. We first introduce the proposed ROI segmentation network. In Sections 2.2 and 2.3, embedding and extraction modules are described. Details of the distortion detection network and the recovery module are presented in Section 2.4.

2.1 ROI segmentation network

The ROI Segmentation unit is responsible for the detection of ROI pixels from NROI pixels, and the generation of ROI block map based on the ROI detected pixels. A block is considered
as NROI if all of its pixels are detected as NROI. Otherwise, the block is considered as ROI. In this work, we use a CNN structure inspired by U-Net [29], without max-pooling, up-sampling, and concatenation layers, for the segmentation of the ROI pixels. Figure 2 demonstrates the structure of the utilized network. As shown in Fig. 2, the ROI segmentation network is composed of 11 convolutional layers. Input to the network is a small $m \times m$ image block, while block size remains constant across all the convolution layers. Hidden layers consist of $3 \times 3$ convolutions followed by ReLU (Rectified Linear Unit) activation function. At the final layer, a $1 \times 1$ convolution is applied, followed by the same ReLU activation function. The number of channels in each layer is shown on the top of each layer in Fig. 2. The network output represents a segmentation probability map for pixels of the input block. This probability map is binarized by using a 0.5 threshold.

### 2.2 Embedding module

We propose a novel iterative approach for the embedding process. The block diagram of the embedding module is shown in Fig. 3. The ROI block map of the cover image is generated based on the strategy discussed in Section 2.1. Watermark is embedded into NROI blocks of the cover image. Then NROI and ROI blocks are merged to construct the tentative watermarked image. The ROI block map of the tentative watermarked image is then generated based on the strategy discussed in Section 2.1. One image block might be detected as NROI before embedding and the same block may be detected as ROI after the embedding. If the

![Fig. 1](image1.png) Block diagram of the proposed blind ROI-based watermarking framework

![Fig. 2](image2.png) ROI segmentation network structure. Each blue box shows a multichannel feature map. Numbers above the boxes represent the number of channels
watermarking process causes a change in the ROI block map, then the modified cover image is constructed, and the watermark is embedded into the NROI blocks of the modified cover image. The embedding process may cause some of the NROI blocks of the original cover image to be identified as ROI. The changed versions of such NROI blocks are placed in the cover image and from there on, they will be considered as ROI. We repeat the embedding process into the modified cover image until the ROI block map remains unchanged, and the final watermarked image is produced. Thus, we do not need to send the ROI block map to the receiver side with the watermarked image, and the proposed framework is blind.

\begin{algorithm}
\textbf{Algorithm 1:} Embedding Pseudo Code  
\textbf{Inputs:} Cover image and digital watermark ($w$)  
\textbf{Output:} Watermarked image  
\begin{algorithmic}[1]
\STATE \textbf{BEGIN}
\STATE \hspace{1em} $j = 0$
\STATE \hspace{1em} \textbf{REPEAT}
\STATE \hspace{2em} \textbf{IF} $j > 0$ \textbf{THEN}
\STATE \hspace{3em} Replace the switched NROI blocks of the original cover image by their embedded version to construct a modified one
\STATE \hspace{2em} \textbf{ENDIF}
\STATE \hspace{1em} $j = j + 1$
\STATE \hspace{1em} Segment cover image by CNN and generate ROI block Map, ROI$_i$
\STATE \hspace{1em} Apply DCT on non-overlapping NROI blocks
\STATE \hspace{1em} \textbf{IF} $w(k) = 0$ \textbf{THEN}
\STATE \hspace{2em} \textbf{IF} $c(i, i + 1) \leq c(i + 1, i)$ \textbf{THEN}
\STATE \hspace{3em} Swap($c(i, i + 1), c(i + 1, i)$), $c(i, i + 1) = th$
\STATE \hspace{2em} \textbf{ELSE}
\STATE \hspace{3em} Swap($c(i, i + 1), c(i + 1, i)$), $c(i + 1, i) = th$
\STATE \hspace{2em} \textbf{ENDIF}
\STATE \hspace{1em} \textbf{ELSE}
\STATE \hspace{2em} \textbf{IF} $c(i, i + 1) \geq c(i + 1, i)$ \textbf{THEN}
\STATE \hspace{3em} Swap($c(i, i + 1), c(i + 1, i)$), $c(i + 1, i) = th$
\STATE \hspace{1em} \textbf{ENDIF}
\STATE \hspace{1em} \textbf{ENDIF}
\STATE \hspace{1em} Apply inverse DCT
\STATE \hspace{1em} Convert the output to integer
\STATE \hspace{1em} Merge NROI blocks with ROI blocks to construct a tentative watermarked image
\STATE \hspace{1em} Segment tentative watermarked image by CNN and generate ROI block map, i.e. ROI$_{i+1}$
\STATE \hspace{1em} \textbf{UNTIL} ROI$_i = ROI_{i+1}$
\STATE \hspace{1em} Consider tentative watermarked image as final watermarked image
\STATE \textbf{END}
\end{algorithmic}
\end{algorithm}

We use a simple embedding method in the DCT domain for proof of concept. However, the proposed framework is not limited to this specific domain, and watermarking can be performed in other known transform domains. A pseudo-code of the embedding module is presented in Algorithm 1. In the first step, $m \times m$ NROI blocks are detected. Then one bit of watermark is embedded in every NROI block. The embedding process is continued until the whole watermark is embedded. When two DCT coefficients are swapped, we add a constant
threshold to secure a minimum distance between the two coefficients. Changing DCT coefficients may lead to under/overflow in the spatial domain, i.e., when they are transformed back to the spatial domain, the pixel values may exceed the valid range \([0, 255]\). In this situation, the under/overflowed pixels are assigned 0 and 255, respectively.

2.3 Extraction module

In the extraction module, we blindly extract the watermark data from the watermarked image. The block diagram of the extraction module is shown in Fig. 4. The ROI map of the watermarked image is generated based on the strategy discussed in Section 2.1. We accurately attain the same ROI map from the final embedding loop. The watermark is extracted from the NROI blocks of the watermarked image. The pseudo-code of the extraction module is

```
Algorithm 2: Extraction Pseudo Code
Input: Watermarked image
Output: Extracted watermark (w)
BEGIN
  01 Segment watermarked image by CNN and generate ROI block map, i.e. ROI_i
  02 Apply DCT on non-overlapping NROI blocks
  03 IF c(i, i + 1) >= c(i + 1, i)
  04 w(k) = 0
  05 ELSE
  06 w(k) = 1
  07 ENDIF
END
```

Fig. 3 Block diagram of the BlessMark embedding module

2.3 Extraction module

In the extraction module, we blindly extract the watermark data from the watermarked image. The block diagram of the extraction module is shown in Fig. 4. The ROI map of the watermarked image is generated based on the strategy discussed in Section 2.1. We accurately attain the same ROI map from the final embedding loop. The watermark is extracted from the NROI blocks of the watermarked image. The pseudo-code of the extraction module is
presented in Algorithm 2. In the first step, the \( m \times m \) NROI blocks are detected. Then one bit of watermark is extracted from each NROI block. The extraction process is continued until the whole watermark is extracted.

2.4 Recovery module

The proposed recovery module performs recovery of distorted NROI blocks of the cover image to the extent feasible. We introduce a distortion detection classifier which is responsible for the detection of distorted NROI blocks that are altered during the embedding process from original NROI blocks. Then distorted blocks are recovered to the closest estimate of the original block. The block diagram of the recovery module is shown in Fig. 5. This module gets the NROI blocks of the watermarked image and the extracted watermark as its inputs and recovers the NROI blocks of the original cover image to the extent feasible. Recovered NROI blocks are merged with ROI blocks to construct the mostly recovered cover image. The pseudocode for the recovery module is presented in Algorithm 3. In the first step, \( m \times m \) NROI blocks, which have been distorted through the embedding process, are detected. The changed blocks are recovered to their original state by reversing the embedding operation. Underflow/Overflow pixels are assigned 0 and 255 respectively.

3 Experimental Results

The proposed framework is implemented in Tensor-flow [1] and executed on NVidia GeForce® GTX 1080 Ti. The framework is evaluated on two datasets of Retina [28] and X-ray Angiography [24], independently. The datasets are composed of 40 color images of size 584 × 565 and 44 grayscale images of size 512 × 512, respectively. We divided datasets into equal sets for training and test purposes. The CNN network and three-layer fully connected neural network performing ROI segmentation and distortion detection are separately trained on each dataset. We generate binary random sequences to be used as watermark data. For color images, watermark data is embedded in all three color channels. Thus, the distortion detection classifier is separately applied on single channels to detect the distorted blocks.

To evaluate the embedding method, we use the two standard measures, PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index). We utilize the Dice score to evaluate the segmentation network and the accuracy metric for evaluating the distortion detection

---

**Fig. 4** Block diagram of the extraction module
classifier. Definitions of all the metrics are presented in Table 1. In the PSNR formula, $w$ and $h$ are the image dimensions. In this formula, $I$ and $I_W$ are the cover image and the watermarked image, respectively. In the SSIM formula, $I$ and $I_W$ are the cover and the watermarked images, respectively. In this formula, $\mu$ and $\sigma^2$ represent mean and variance values, respectively. Also, $\sigma_{I,I_W}$ is the covariance between $I$ and $I_W$ and $c_1$ and $c_2$ are constants. In Dice score and Accuracy measures, variables TP and FP are cases that are predicted as positive, while actual outputs are positive and negative, respectively. Also, TN and FN are cases that are predicted as negative while the actual outputs are negative and positive, respectively.

![Diagram of the BlessMark recovery module](image)

**Fig. 5** Block diagram of the BlessMark recovery module

---

**Algorithm 3: Recovery Pseudo Code**

**Inputs:** Extracted watermark ($w$) and NROI blocks

**Output:** Mostly recovered NROI blocks of cover image

BEGIN
01 IF distortion detection network predict that NROI block has distorted
02 Apply DCT on block
03 IF $w(k) == 0$
04 $c(i,i + 1) == th$
05 IF $c(i,i + 1) > c(i + 1,i)$
06 $\text{Swap}(c(i,i + 1), c(i + 1,i))$
07 ENDIF
08 ELSE
09 $c(i + 1,i) == th$
10 IF $c(i,i + 1) < c(i + 1,i)$
11 $\text{Swap}(c(i,i + 1), c(i + 1,i))$
12 ENDIF
13 ENDIF
14 Apply inverse DCT
15 Convert the output to integer
16 ENDIF

END
All the networks are evaluated in Section 3.1. Performance of the whole framework is analyzed in Section 3.2. Finally, we compare BlessMark with state-of-the-art watermarking systems in Section 3.3.

### 3.1 Independent evaluation of networks

#### 3.1.1 Segmentation network

As discussed in previous sections, the segmentation network is responsible for attaining ROI pixel map. This network takes a block as the input and produces an ROI pixel map for that block. The network input is a gray-scale image block. Therefore, color images are converted to grayscale, and one ROI map is generated for all channels. A softmax activation function is used for creating a probability map in the last layer. We use a cross-entropy loss function with stochastic gradient descent (SGD) optimizer for training. Table 2 presents the Dice scores of the trained networks on two datasets with different block sizes. Two training sets constructed from 20 Retina images and 22 Angiography images are used separately for training each network within 150 epochs. The other half of the datasets is used as a test set. As shown in Table 2, the trained networks recognize 73% of ROI pixels for Retina and 61% of the ROI pixels for Angiography test sets, when the block size is 6 × 6. Dice scores increase by using larger block sizes, usually. Also, it is shown that a trained network for retina reaches a higher Dice score than the trained network for angiography test sets for the same block size. As an example, the produced ROI pixel maps for retina and angiography are shown in Fig. 6.

#### 3.1.2 Distortion detection network

As discussed in previous sections, the distortion detection network is responsible for detecting the distorted blocks, i.e., the blocks that have distorted during the embedding process. In this work, we train a three-layer fully connected neural network for distortion detection. The

### Table 1 Evaluation metrics employed for testing the framework

| Evaluation                  | Metric | Formula |
|-----------------------------|--------|---------|
| Watermark Imperceptibility  | PSNR   | $10 \log_{10} \frac{255^2}{\sum_{i,j=0}^{w,h}(I_w(i,j) - I(i,j))^2}$ |
|                             | SSIM   | $\frac{(2\mu_I \mu_w + \alpha)(2\sigma_I \sigma_w + \beta)}{(\sigma_I^2 + \alpha^2)(\sigma_w^2 + \beta^2)}$ |
| Networks Evaluation        | Dice Score | $\frac{2TP}{2TP + FP + FN}$ |
|                             | Accuracy| $\frac{TP}{TP + TN + FP + FN}$ |

### Table 2 Performance of segmentation network for two datasets with different block sizes

| Block Size | Test Dice Score |
|------------|-----------------|
| 6 × 6      | Retina 0.73     |
|            | Angio 0.61      |
| 8 × 8      | Retina 0.76     |
|            | Angio 0.67      |
| 10 × 10    | Retina 0.76     |
|            | Angio 0.69      |
The structure of this network is shown in Fig. 7. The network input is a small $m \times m$ NROI block. In the first step, the block is flattened by a flatten layer. The two hidden layers are dense layers with $m^2$ nodes and each layer is followed by a ReLU activation function. At the final layer, a dense layer with one neuron is used, which is followed by a sigmoid activation function. We use Adam with default parameters as an optimizer and cross-entropy as a loss function. The network output represents a classification probability, in which this probability is converted to a binary by thresholding the output layer on 0.5.

![Distortion detection network structure](image_url)
Table 4 Performance of the proposed framework on Retina and Angiography test sets

| Block size | Capacity (BPP) | PSNR of Watermarked Image (dB) | PSNR of Recovered Image (dB) | Improvement of PSNR (dB) | PSNR of Watermarked Image NROI (dB) | PSNR of Recovered Image NROI (dB) | Improvement of NROI PSNR (dB) | PSNR of Watermarked Image ROI (dB) | PSNR of Recovered Image ROI (dB) | Improvement of ROI PSNR (dB) | Average Percent of Switched NROI Block into ROI Block (%) |
|------------|----------------|-------------------------------|-------------------------------|--------------------------|----------------------------------------|-----------------------------------|---------------------------------|--------------------------------|--------------------------------|-----------------------------|---------------------------------------------------------------|
| 6 × 6      | Retina 0.021  | 55.24                         | 60.09                         | 4.85                     | 59.81                                  | 60.78                             | 0.97                            | 65.29                          | 64.32                         | 0.98                       | 0.164                                                                |
|            | Angio 0.022  | 55.65                         | 62.81                         | 7.16                     | 55.48                                  | 62.66                             | 7.18                            | 59.63                          | 65.96                         | 6.33                      | 0.047                                                                |
| 8 × 8      | Retina 0.010  | 58.16                         | 63.16                         | 5.00                     | 57.84                                  | 62.84                             | 5.00                            | 65.31                          | 69.90                         | 4.59                      | 0.154                                                                |
|            | Angio 0.012  | 58.20                         | 65.32                         | 7.12                     | 57.99                                  | 65.11                             | 7.12                            | 63.73                          | 70.88                         | 7.15                      | 0.040                                                                |
| 10 × 10    | Retina 0.006  | 60.77                         | 64.50                         | 3.73                     | 60.44                                  | 64.17                             | 3.73                            | 68.50                          | 72.64                         | 4.14                      | 0.137                                                                |
|            | Angio 0.007  | 61.39                         | 64.67                         | 3.28                     | 61.16                                  | 64.44                             | 3.28                            | 67.96                          | 71.09                         | 3.13                      | 0.030                                                                |
Every block has an intrinsic embedded value. Hence, for arranging a training set for the network, we pick all the NROI blocks of all channels and invert the intrinsic embedded value inside them. The training set is composed of all the NROI blocks and their inverted versions. Thus, the three-layer fully connected neural network is trained to classify these two sets of blocks for all channels. It worth mentioning that we can embed different watermark data in various channels. Consequently, the classifier conducts independent analyses on separate channels.

The accuracy of the classifier for the two datasets is demonstrated in Table 3. Training set in the Retina and the Angiography datasets are used for the training of the two networks with 100 epochs. Test sets are constructed similar to the training sets by using a test set in the Retina and the Angiography datasets. Index \(i\) in Table 3 represents the two DCT coefficients \(c(i, i+1)\) and \(c(i+1, i)\), which are swapped for embedding. As shown in Table 3, DCT coefficients swapped for embedding in various block sizes ('6 × 6', '8 × 8', '10 × 10') are different. The parameter \(th = 0.01\) represents the threshold as used in our embedding algorithm (Algorithm 1). As shown in Table 3, the trained networks detect the distorted blocks with an accuracy of 94% for Retina and 97% for Angiography, when the block size is 6 × 6.

3.2 Evaluation of the whole framework

In Table 4, we evaluate our framework for the 20 Retina and 22 Angiography test images in terms of imperceptibility and variations in the ROI map. Since the image is segmented by CNN, false classifications are probable, i.e., it is probable that one pixel is segmented as NROI by CNN while it is labeled as ROI in the ground truth or vice versa. Therefore, we use the ground truth ROI map for calculating PSNR of images.

In Table 4, PSNR of the whole image, as well as PSNR of ROI and NROI regions, are presented. For Retina image with block size 6 × 6, PSNR of the watermarked image NROI is 54.96 (dB) on the maximum capacity of 0.021 (BPP). The image NROI imperceptibility is enhanced to 59.81 (dB) after recovery, i.e., imperceptibility of NROI may be improved by 4.85 (dB). The recovery process is not perfect due to the distortion detection classifier errors, irreversible distortions caused by under/overflow clipping of embedding operation, and switching of segmentation results after embedding. We also demonstrate the average percentage of NROI blocks, which are switched to ROI blocks during the embedding process. It is

| Table 3 | Performance of classifier for two datasets with different block sizes. \(th = 0.01\) |
|----------|---------------------------------|
| Block Size | Index \(i\) | Number of Blocks for Training | Number of Blocks for Testing | Test Accuracy |
| 6 × 6 | Retina | 820,950 | 813,042 | 0.94 |
| | Angio | 276,894 | 277,372 | 0.97 |
| 8 × 8 | Retina | 423,726 | 418,872 | 0.95 |
| | Angio | 155,678 | 155,736 | 0.97 |
| 10 × 10 | Retina | 248,226 | 243,858 | 0.95 |
| | Angio | 95,524 | 95,790 | 0.97 |

| Table 5 | Training times of two networks on Retina training set |
|----------|---------------------------------|
| Segmentation network | Classification network |
| 23 (Hour) | 4.5 (Hour) |
shown that distortions caused by the iterative embedding process are minimal and decrease by using larger block sizes.

Boxplot of Fig. 8 shows the distribution of PSNR and SSIM for watermarked images across the Retina and Angiography test sets. Each box corresponds to block size. For block sizes 6 × 6, 8 × 8, 10 × 10, maximum experimental capacities are 0.021, 0.010, and 0.006 (BPP) for Retina and 0.022, 0.012 and 0.007 (BPP) for Angiography test sets.

In Table 5, the training durations, for the segmentation and classifier networks, are shown for a retina training set with a 6 × 6 block size. We used 2 M blocks to train the segmentation network in 150 epochs. Also, the classifier network is trained by using 820,950 blocks in 100 epochs. The segmentation and the classification networks are trained in 23 h and 4.5 h, respectively.

In Table 6, embedding, extraction, and recovery times with GPU and without GPU are evaluated for the Retina test set. Average experimental results for 20 Retina test images with block size 6 × 6 and capacity 0.021 (BPP) is demonstrated in Table 6. Embedding time with GPU is 7 s. Extraction and recovery time with GPU is 3 s. These times are increased to 78 and 18 s without GPU.

In Fig. 9, the visual qualities of the embedding result with block size 6 × 6 for two test images are shown. The left column shows the original cover image, the middle column is the embedded watermark, and the right column demonstrates the watermarked image with its corresponding PSNR and SSIM. We have not seen any distinguishable difference between the watermarked image and the cover image.

### 3.3 Comparison with State-of-the-art Methods

Properties of various watermarking algorithms \[36\], \[11, 32\], \[4, 5, 7, 14, 16, 19, 21, 35, 37, 38\] are compared with our framework in Table 7. The eight lossless watermarking methods \[36\], \[4, 5, 14, 16, 21, 35, 38\] use the whole cover image for embedding, causing distortions across all image regions including ROI. Therefore, they are not applicable to medical watermarking, since diagnostic information in ROI may be corrupted. Also methods of \[36\], \[4, 5, 38\], \[21\] are non-blind.

The two non-blind ROI based watermarking methods \[11, 32\] send an ROI map with the watermarked image to the receiver side. The two blind ROI based watermarking methods \[7, 37\] embed ROI map inside the watermarked image. Hence, the extraction of an ROI map from the watermarked image is a prior step for the extraction of the watermark data. The ROI based watermarking method \[19\] utilize the same segmentation method for recognizing the ROI map in embedding and extraction modules of their proposed system. In our framework, the watermark is embedded in NROI blocks to keep the sensitive ROI information intact during the embedding process. ROI map can be recognized from the watermarked image in the extraction module by the trained segmentation network.
| Reference Block Size | Capacity (BPP) | PSNR of Watermarked Image (dB) | PSNR of Watermarked Image NROI (dB) | PSNR of Watermarked Image ROI (dB) | PSNR of Recovered Image (dB) | PSNR of Recovered Image NROI (dB) | PSNR of Recovered Image ROI (dB) | Average Side Information (Bit) |
|----------------------|---------------|--------------------------------|-----------------------------------|-----------------------------------|-----------------------------|-----------------------------------|-----------------------------------|-------------------------------|
| Proposed 6 × 6       | 0.020         | 55.01                          | 54.64                             | 60.95                             | 59.71                       | 59.32                             | 65.90                             | 0                             |
| 8 × 8                | 0.010         | 58.43                          | 58.02                             | 65.84                             | 62.17                       | 61.77                             | 69.19                             | 0                             |
| 10 × 10              | 0.005         | 60.87                          | 60.45                             | 68.67                             | 64.77                       | 64.35                             | 72.74                             | 0                             |
| [38] Pixel based     | 0.020         | 65.01                          | 64.75                             | 70.06                             | Inf                         | Inf                               | Inf                               | 16,516.5                      |
| 0.010                | 67.99         | 67.67                          | 74.63                             | Inf                               | Inf                         | Inf                               | Inf                               | 8242.5                        |
| 0.005                | 70.55         | 70.22                          | 77.28                             | Inf                               | Inf                         | Inf                               | Inf                               | 4627.5                        |
In Table 8, imperceptibility and side-information of one lossless method [38] are compared with our system. Similar to Table 4, we use the ground truth ROI map for calculating PSNR of ROI and NROI. Note that the false classification of CNN is probable, i.e., a pixel may be segmented as NROI by CNN while it is labeled as ROI in the ground truth or vice versa.

The lossless method [38] embeds the watermark into 1-level of IWT of the whole image in two iterations so that iteration-2 compensates for produced distortion in iteration-1.

In Table 8, the PSNR and side-information results are calculated over the 20 Retina test images. For a fair comparison, a single watermark is embedded in all methods. The infinity values of PSNR in the table demonstrate that the original cover image has been completely recovered. The proposed framework does not produce any side information, while the method of [38] produces 16,516.5 bits of side information on the capacity of 0.020 (BPP). However, the lossless method [38] can completely recover the original cover image. Since NROI does not contain valuable information for medical diagnosis, accurate recovery of NROI is not critical in medical applications.

Figures 2 and 7 show the general structure of the ROI segmentation network and the distortion detection network. Details of layers of these two networks are listed in Table 9.

| Table 7 | Qualitative comparison of the proposed framework with other methods |
|---------|-------------------------------------------------------------------|
| Reference Region | Embedding Region | ROI Segmentation Method | ROI Map as Additional Information | Blind Watermarking Approach |
| Proposed NROI | CNN | No | Yes | ROI-based |
| [11] NROI | Manual | Yes | No | ROI-based |
| [32] NROI | Thresholding of saliency map | Yes | No | ROI-based |
| [37] NROI and ROI | Adaptive Threshold Detector | Yes | Yes | ROI-based |
| [7] NROI | Manual | Yes | Yes | ROI-based |
| [19] NROI | Otsu | No | Yes | ROI-based |
| [36] Whole | No segmentation | No | No | Lossless |
| [35] Whole | No segmentation | No | Yes | Lossless |
| [16] Whole | No segmentation | No | Yes | Lossless |
| [4] Whole | No segmentation | No | No | Lossless |
| [5] Whole | No segmentation | No | No | Lossless |
| [38] Whole | No segmentation | No | No | Lossless |
| [14] Whole | No segmentation | No | Yes | Lossless |
| [21] Whole | No segmentation | No | No | Lossless |

Table 9 | Details of the segmentation and detection networks |
| Networks | Configuration |
|-------|----------------|
| ROI segmentation network: (CNN) | 11 Conv layers + ReLU consisting of 10 hidden layers: 3 × 3 Convs Final layer: 1 × 1 Conv |
| Distortion detection network: Fully connected network | 3 Dense layers + ReLU, 1st layer 16 nodes, 2nd layer 16 nodes, 3rd layer one node |
Fig. 8  Box-plot of watermarked image PSNR and SSIM. Each box represents the distribution of PSNR and SSIM versus block size. (a) Retina. (b) Angiography

Fig. 9  (a), (c) Original cover images, (b), (d) watermarked images
4 Conclusion

In this paper, we presented BlessMark, a framework for blind diagnostically-lossless watermarking. The proposed watermarking scheme is used as a means for the simultaneous improvement of confidentiality and preservation of diagnostic medical information. BlessMark consists of a deep neural network for image segmentation and one fully connected neural network for classification.

BlessMark ROI segmentation network generates an ROI map for the embedding, extraction, and recovery modules. The ROI segmentation network is applied using an iterative scheme to accurately produce the same ROI map, both in the transmitter and receiver sides. Hence, in the proposed blind watermarking framework, an ROI map is automatically generated on the receiver side without requiring any additional information. The proposed ROI map is critical for improving confidentiality protection in our system. Third parties cannot create the ROI map without having access to our proprietary segmentation tool. Since the watermark is embedded only in NROI blocks, ROI remains intact and leading to a diagnostically-lossless watermarking system. The distortion detection classifier used in the recovery module helps the detection of blocks that have distorted during the embedding process. Distorted blocks are mostly recovered to their original form.

The choice of a simple embedding method in the DCT domain is just a convenient option to prove the concept of the proposed framework. Hence, different embedding methods in other transform domains such as DWT and Hadamard may be applied.

Furthermore, we used a CNN and a three-layer fully connected neural network for the ROI segmentation and detection of distorted NROI blocks. However, other structures can be investigated for these purposes. Our watermarking method is non-robust since the ROI map of the watermarked image can be changed as a result of attacks. Since the ROI map is vital for embedding and extraction, it is necessary to detect the same ROI map on both sides.

We evaluated our framework on two Retina and Angiography datasets. The proposed framework demonstrates producing any side information in comparison with one lossless watermarking method. Our proposed method is advantageous to recent methods such as the work of [19]. We are offering a framework while other research works implement a specific embedding routine.

One future research is to explore the capabilities of the proposed framework by using it as a platform for testing other networks and embedding domains. We are also planning to work on the robustness of the proposed framework such that the watermark can be extracted accurately in the presence of attacks. Furthermore, the new robust framework needs to be compared against more recent watermarking schemes, some of which are reviewed in the introduction section.

References

1. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Kudlur M (2016) Tensorflow: a system for large-scale machine learning. In: OSDI, vol 16, pp 265–283
2. Abdelhakim AM, Abdelhakim M (2018) A time-efficient optimization for robust image watermarking using machine learning. Expert Syst Appl 100:197–210
3. Anbarjafari G, Ozcinar C (2018) Imperceptible non-blind watermarking and robustness against tone mapping operation attacks for high dynamic range images. Multimed Tools Appl Appl:1–15
4. Ansari IA, Pant M, Ahn CW (2017) Artificial bee colony optimized robust-reversible image watermarking. Multimed Tools Appl 76(17):18001–18025
5. Bamal R, Kasana SS (2018) Slantlet based hybrid watermarking technique for medical images. Multimed Tools Appl 77(10):12493–12518
6. Benrhouma O, Hermassi H, El-Latif AAA, Belghith S (2016) Chaotic watermark for blind forgery detection in images. Multimed Tools Appl 75(14):8695–8718

7. Chaitanya K, Rao KG (2018) A novel approach to medical image watermarking for tamper detection and recovery of region of interest using predictive coding and hashing. J Theoretical Appl Inf Technol 96(7)

8. Chao HM, Hsu CM, Miaou SG (2002) A data-hiding technique with authentication, integration, and confidentiality for electronic patient records. IEEE Trans Inf Technol Biomed 6(1):46–53

9. Chauhan DS, Singh AK, Kumar B, Saini JP (2019) Quantization based multiple medical information watermarking for secure e-health. Multimed Tools Appl 78(4):3911–3923

10. Cox I, Miller M, Bloom J, Fridrich J, Kalker T (2007) Digital watermarking and steganography. Morgan Kaufmann

11. Eswaranah R, Reddy ES (2015) Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest. IET Image Process 9(8):615–625

12. Etemad E, Samavi S, Soroushrehmehr SR, Karimi N, Etemad M, Shirani S, Najarian K (2018) Robust image watermarking scheme using bit-plane of Hadamard coefficients. Multimed Tools Appl 77(2):2033–2055

13. Heidari M, Samavi S, Soroushrehmehr SMR, Shirani S, Karimi N, Najarian K (2017) Framework for robust blind image watermarking based on classification of attacks. Multimed Tools Appl 76(22):23459–23479

14. Hou D, Zhang W, Chen K, Lin SJ, Yu N (2018) Reversible data hiding in color image with grayscale invariance. IEEE Transactions on Circuits and Systems for Video Technology 29(2):363–374

15. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: International Conference on Computer Vision, pp 4540–4544

16. Jane O, Elbaši E (2014) Hybrid non-blind watermarking based on DWT and SVD. Journal of applied research and technology 12(4):750–761

17. Kandi H, Mishra D, Gorthi SRS (2017) Exploring the learning capabilities of convolutional neural networks for robust image watermarking. Computers Security 65:247–268

18. Kelkar V, Tuckley K, Nemade H (2017, 2017) Novel variants of a histogram shift-based reversible watermarking technique for medical images to improve hiding capacity. J Healthcare Eng

19. Kim WH, Hou JU, Mun SM, Lee HK (2018) Convolutional neural network architecture for recovering watermark synchronization. arXiv preprint arXiv:1805.06199.

20. Liu X, Lou J, Fang H, Chen Y, Ouyang P, Wang Y, Wang L (2019) A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images. IEEE Access 7:76580–76598

21. Mehta R, Rajpul N, Vishwakarma VP (2017) A robust and efficient image watermarking scheme based on Lagrangian SVR and lifting wavelet transform. Int J Mach Learn Cybern 8(2):379–395

22. Naik K, Trivedy S, Pal AK (2018) An IWT based blind and robust image watermarking scheme using secret key matrix. Multimed Tools Appl 77(11):13721–13752

23. Nasr-Esfahani E, Karimi N, Jafari MH, Soroushrehmehr SMR, Samavi S, Nallamothu BK, Najarian K (2018) Segmentation of vessels in angiograms using convolutional neural networks. Biomedical Signal Processing and Control 40:240–251

24. Parah SA, Ahad F, Sheikh JA, Bhat GM (2017) Hiding clinical information in medical images: a new high capacity and reversible data hiding technique. J Biomed Inform 66:214–230

25. Parah SA, Sheikh JA, Ahad F, Loan NA, Bhat GM (2017) Information hiding in medical images: a robust medical image watermarking system for E-healthcare. Multimed Tools Appl 76(8):10599–10633

26. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis Machine Intelligence 6:1137–1149

27. Retina-unet. [Online]. Available: https://github.com/orobix/retina-unet.

28. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241

29. Sadreazami H, Amini M (2018) A robust image watermarking scheme using local statistical distribution in the Contoulet domain. Express Briefs, IEEE Transactions on Circuits and Systems II

30. Savakar DG, Gulhi A (2017) Non-blind digital watermarking with enhanced image embedding capacity using DMeyer wavelet decomposition, SVD, and DFT. Pattern Recognition and Image Analysis 27(3):511–517

31. Shih FY, Zhong X, Chang IC, Satoh SI (2018) An adjustable-purpose image watermarking technique by particle swarm optimization. Multimed Tools Appl 77(2):1623–1642

32. Singh RP, Dabas N, Chaudhary V (2016) Online sequential extreme learning machine for watermarking in DWT domain. Neurocomputing 174:238–249
34. Thanki R, Borra S (2019) Fragile watermarking for copyright authentication and tamper detection of medical images using compressive sensing (CS) based encryption and contourlet domain processing. Multimed Tools Appl 78(10):13905–13924
35. Tian J (2002) Reversible watermarking by difference expansion. In: Proceedings of workshop on multimedia and security, vol 19
36. Turuk MP, Dhande AP (2016) A novel reversible multiple medical image watermarking for health information system. J Med Syst 40(12):269
37. Yang Y, Zhang W, Liang D, Yu N (2018) A ROI-based high capacity reversible data hiding scheme with contrast enhancement for medical images. Multimed Tools Appl 77(14):18043–18065
38. Zarrabi H, Hajabdollahi M, Soroushmehr SMR, Karimi N, Samavi S, Najarian K (2018) Reversible image watermarking for health informatics systems using distortion compensation in wavelet domain. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society
39. Zheng Z, Zheng L, Yang Y (2017) A discriminatively learned cnn embedding for person re-identification. ACM transactions on multimedia computing, communications, and applications (TOMM), 14(1), 13

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Hamidreza Zarrabi received his B.Sc. degree in Computer Engineering in 2014 from Sistan and Baluchistan University, Iran. He started his M.Sc. education in Computer System Architecture in 2016 at Isfahan University of Technology, Iran. His main research interests are watermarking, machine learning, image and signal processing.
Ali Emami is a postdoctoral fellow at Isfahan University of Technology (IUT), Iran. He received his BS degree in Electrical Engineering from IUT in 1998 and his Master’s degree in Biomedical Engineering from Amirkabir University of Technology, Iran in 2001. He received his Ph.D. from The University of Queensland, Australia in 2015. He was concurrently a research scholar at National ICT Australia and a joint researcher at IBM Watson Lab. As a post-doctoral researcher he is mainly focused on Computer Vision, Machine Learning and Deep Learning methods.

Pejman Khadivi is an Assistant Professor of Computer Science at Seattle University, USA. He received his B.Sc. (in 1998) and M.Sc. (in 2000) degrees in Computer Engineering and a Ph.D. in Electrical Engineering (in 2006) from Isfahan University of Technology with honors. He also received a Ph.D. in Computer Science (in 2016) from Virginia Tech. During the 2003/2004 academic year, he was a Visiting Researcher at the ECE Department, McMaster University, Canada. From 2007 to 2011 he was an Assistant Professor at Isfahan University of Technology. Dr. Khadivi’s primary research interests are in the field of artificial intelligence, machine learning, and data analytics using large scale datasets, with emphasis on time series analytics, deep learning, and information theory.
Nader Karimi received the B.S. degree (summa cum laude) in computer engineering from Azad University, Arak Branch, Iran, in 2002 and the M.Sc. and Ph.D. degrees (honor) in computer engineering and electrical engineering from Isfahan University of Technology (IUT), Iran, in 2004 and 2012, respectively. He is currently an Assistant Professor at the Department of Electrical and Computer Engineering, Isfahan University of Technology. His research interests are image compression, hardware implementation and optimization of image processing algorithms, and watermarking.

Shadrokh Samavi is a Professor of Computer Engineering at Isfahan University of Technology, Iran. He is also an Adjunct Professor at the ECE department of McMaster University where he is a member of the Multimedia Signal Processing Lab. Dr. Samavi is also a research affiliate at the Biomedical and Clinical Informatics Lab, Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. Professor Samavi completed a B.S. degree in Industrial Technology and received a B.S. degree in Electrical Engineering at California State University, a M.S. degree in Computer Engineering at the University of Memphis and a Ph.D. degree in Electrical Engineering at Mississippi State University, U.S.A. Dr. Samavi is a Registered Professional Engineer (PE), USA. He is also a senior member of IEEE and a member of Eta Kappa Nu and Tau Beta Pi honor societies. Shadrokh Samavi’s research interests are in the areas of image processing and hardware implementation and optimization of image processing algorithms. He is also interested in computer vision for biomedical application, as well as, VLSI design and computer arithmetic.
Affiliations

Hamidreza Zarrabi\textsuperscript{1} \cdot Ali Emami\textsuperscript{1} \cdot Pejman Khadivi\textsuperscript{2} \cdot Nader Karimi\textsuperscript{1} \cdot Shadrokh Samavi\textsuperscript{1,3}

\begin{itemize}
\item Pejman Khadivi
  \texttt{khadivip@seattleu.edu}
\end{itemize}

\textsuperscript{1} Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran
\textsuperscript{2} Department of Computer Science, Seattle University, Seattle, WA 98122, USA
\textsuperscript{3} Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA