FOREX Publication
Open Access | Rapid and quality publishing
International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 339-344 | e-ISSN: 2347-470X

Multi-image Feature Map-Based Watermarking Techniques Using Transformer

Aberna Palani[1], Agilandeeswari Loganathan[2]

[1]*Research Scholar, School of Information Technology and Engineering (SITE), Vellore Institute Of Technology, Vellore, India, aberna.p@vitstudent.ac.in*

[2]*Professor, School of Information Technology and Engineering (SITE), Vellore Institute Of Technology, Vellore, India, agila.l@vit.ac.in*

#Correspondence: Aberna Palani; agila.l@vit.ac.in

ABSTRACT- Nowadays, protecting multimedia data is a significant challenge because of the advancement of technology and software. The embedding process heavily relies on watermarking to accomplish multimedia security in terms of content authentication, proof of ownership, and tamper detection. Our objective is to develop an invariant watermark that can survive different signal-processing attacks. We presented a unique hybrid technique (DWT-QR-SWT) and multi-image invariant features generated as a watermark using a Transformer encoder-decoder model. The encoded image features are subsampled using PCA in order to decrease the dimensionality of the watermark image. The first two images are used as watermark1 and the next two images as watermark2 to produce multi-watermark feature maps. To embed the watermark, a hybrid DWT-QR decomposition has been applied to the original image1. On the primary watermarked image, two Level Stationary Wavelet Transform (SWT) were applied to embed the secondary watermark2. At the extraction phase, the tampered image is recovered by passing the extracted watermark image as input to the transformer decoder. A multi-image watermark increases data embedding capabilities and also achieves two-level content authentication, tamper detection, localization, and recovery. With a PSNR of 59.05 dB, the testing result demonstrates great resilience and improved imperceptibility.

Keywords: Transformer; Tamper detection; QR decomposition; Discrete Wavelet Transform; Stationary Wavelet Transform; Principal component analysis.

ARTICLE INFORMATION

Author(s): Aberna Palani and Agilandeeswari Loganathan;

Received: 18-02-2023; **Accepted:** 20-05-2023; **Published:** 30-05-2023;

E- ISSN: 2347-470X;

Paper Id: IJEER230207b;

Citation: 10.37391/IJEER.110214

Webpage-link:

https://ijeer.forexjournal.co.in/archive/volume-11/ijeer-110214.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Network and internet technology has grown in advance for various applications such as social media platforms, e-Health care, Telemedicine [1], e-banking, etc. On one end, the transmission of multimedia data such as audio, text, video, and the image growth rate increased, on the other hand, multimedia data tampering has also been increased. Resulting to this, protecting multimedia data in digital networks has become a major problem. Security is a top concern in multimedia platforms across various applications, including Content Authentication, Proof of Ownership, and Copyright Protection. To secure digital multimedia, cryptography [2], steganography [3], and watermarking [4] techniques are introduced. Among that watermarking showed a state-of-the-art method in data embedding, and also it has significant to prove ownership, copyright protection, and multimedia data authentication. Digital watermarking techniques are generally processed on two domains [5]: Pixel-Based domain and Frequency-Based domain. To transform the image in the frequency domain, techniques such as Discrete Fourier Transform (DFT) [6], Quaternion Curvelet Transform (QCT) [7], Discrete Cosine Transform (DCT) [8], Hilbert Transform [9], Wavelet Transform [Discrete wavelet (DWT) [10], Stationary Wavelet (SWT) [11], Integer Wavelet Transform (IWT) [12], Complex Wavelet Transform (CWT)[13]], etc. The hybrid embedding techniques are also been attracted by the researchers to achieve authentication [14]. A robust blind SWT watermarking technique is proposed by Nagarjuna, P.V. et al.,[11]. The Watermark image is encrypted by XOR operation, embedded in the original image's low-frequency stationary wavelet transform (SWT) coefficient values. The extraction phase shows higher imperceptibility for tamper detection applications. In addition to the frequency domain, matrix decomposition techniques such as Singular value decomposition (SVD) [10], and QRD decomposition [15]. In [15] a QR decomposition (QRD)-based blind watermarking approach for colour images is designed. The watermark is embedded by altering the 2-bit per pixel of each 4×4 QR block, and the watermarked image is downsampled to the CNN to train the image for tamper detection and extracting the watermark. To compute QR matrices, a unique technique offered by [16] was employed instead of the Gram-Schmidt approach. The image is partitioned into blocks before watermark embedding to decrease the extraction time and the watermark embedded in R(1,1) matrix. The experimental findings reveal that the watermarked images have good quality. In [10] DWT-SVD transform is suggested to provide content authentication. They generate the singular coefficient matrix as a watermark which is embedded in the 2-level DWT HH sub-band using a secret key. Whereas in [17], the original image was preprocessed by compressing it using the PCA technique. The watermark is generated by generating singular values from both the original and watermark image which are combined to form a new singular matrix. The new singular matrix is a dot product with the original image U and V matrix to form a modified LL sub-band which shows acceptable results for copyright application. In recent trends, researchers have strongly recommended ensemble techniques, where deep learning-based watermarking approaches such as DCT-CNN [18], DWT-CNN [19], and others demonstrated state-of-the-art ways rather than conventional watermarking methods. In the deep Learning approach, CNN plays a prominent role in classification, object detection, invariant feature map extraction [20], and so on. On the other hand, some researchers are concentrated on Neural networks and fuzzy-based watermarking systems [21, 22]. In recent work, the attention-based transformer is slowly getting more performance recognition than CNN in various applications. CNN's drawback is that hard inductive bias is required to perform better for small datasets. So, to overcome this, the hard inductive bias usage is avoided by the transformer using a key component, which is the attention mechanism [23]. Transformer designed primarily for natural language processing (NLP) applications to help with language translation, image translation, and speech understanding [24]. Transformer is a straightforward and parallel processing method so, the researcher extended this technique to various computer vision applications like image classification, video, and audio processing. The transformer concept is briefly explained in *section 2.1*.

The main contributions of this research are:

(i) Attempted Transformer model for the first time in the field of watermarking and succeeded in image tampering applications.

(ii) Proposed a novel hybrid technique namely DWT-QR-SWT which has the potential to hold a high-capacity watermark.

(iii) Achieved high robustness using data augmentations and it is fed to train transformers using rotation, scaling, and flipping which produces highly invariant watermark feature map images.

The rest of the paper is explained as follows: *section 2* explains the preliminary concept, *section 3* describes the proposed system followed by results and discussion in *section 4*.

2. PRELIMINARY CONCEPTS

2.1 Transformer

The transformer encoder-decoder model was proposed by Google brain team members in 2017 [24] for various image translation and Language translation applications. The transformer encoder-decoder architecture model is composed of six identical stack layers (i.e., six residual encoder and six residual decoder blocks). The transformer encoder has two sub-layers: one is a multi-head self-attention layer; the second sub-layer is a feed-forward layer. Each sub-layer is followed by a residual connection and normalization layer. The positional embedding information is concatenated with the flattened 1D sequence input vector X_i, which is fed as input to the transformer encoder. In the encoder, self-attention generates a key vector K_i, value vector V_i, and query vector Q_i by multiplying each input vector X_i with three weight matrices W_{qkV}. The weight matrix will be the same for all the input vector $X_{i:n}$ sequence. Self-attention computes the dot product of each query vector Q_i with all key transpose vector $K_{i:n}$ and divided by key dimension d_k followed by the softmax function. The self-attention is then multiplied by the corresponding value vectors which give the output vector Self-Attention (Q, K, V) as represented in *equation (1)*. The self-attention mechanism performs parallelly, where each self-attention head is concatenated together as multi-head attention (MHA) represented as shown in *equation (2)*.

$$\text{Self-Attention } (head_1)=\text{softmax}\left(\frac{K_{i:n}{}^T Q_i}{\sqrt{d_K}}\right) V_i \tag{1}$$

$$\text{MultiHead (MHA)}=\text{Concat}(head_1+\ldots,head_n) \tag{2}$$

Where $head_i$ = Self-Attention (Q, K, V). Normalized multi-head attention passes to the feed-forward layer. X_i' is fed has an input for the next stack encoder and the process continues until the last encoder block outputs the final encoded data.

$$Z_1 = \text{MHA}(\text{Layernorm}(X_i E))+\ldots X_i E_{\ M-1}, \tag{3}$$

Where M=1,2,…n.

The normalized MHA output is concatenated together and passed as input to the MLP classifier of the encoder block to obtain learned feature maps *eq.6*.

$$Z_n' = \text{MLP}(\text{Layernorm}(Z_1)) + Z_l, \quad l=1,2\ldots N \tag{4}$$

$$Y_i=(\text{Layernorm } (Z_n')) \tag{5}$$

The transformer-based decoder has three sub-layers: the first two layers are masked multi-head attention and norm layer followed by the residual connection for each sub-layer. Instead of self-attention in the decoder masking multi-head attention is used to prevent positions from attending to subsequent positions. Masked MHA value as well as key-value of the encoder is fed to the next multi-head attention layer. The MSA output is given to the feed-forward layer followed by linear and softmax layers to train and classify the input.

2.2 QR decomposition

QR decomposition is a matrix decomposition method that is similar to the Singular Value Decomposition method [3, 2], applied to original image matrix A of size 4×4, which decomposes into two matrices as *eq. (6)*.

$$A= Q\,R = [q1\ q2\ q3\ q4]\begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ 0 & r_{22} & r_{23} & r_{24} \\ 0 & 0 & r_{33} & r_{34} \\ 0 & 0 & 0 & r_{44} \end{bmatrix} \tag{6}$$

where Q is an orthogonal matrix and R is an upper triangular matrix, q1, q2, q3, q4 are column unitary vectors of the Q matrix. Elements of the Q matrix derived from the columns of A by the Gram-Schmidt orthogonalization process.

Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 339-344 | e-ISSN: 2347-470X

Stationary Wavelet Transform

A new wavelet transform called stationary wavelet transform (SWT) is applied to the primary watermarked image in the secondary level of embedding. Due to its translation-invariance traits against signal processing attacks, SWT is also frequently referred to as an un-decimated wavelet transform [11]. SWT is similar to DWT, additionally SWT level contains the same amount of redundancy. In SWT, instead of down-sampling steps, it used a null placing procedure instead of the downsampling [25]. In the proposed method, the LH2 sub-band coefficients in the SWT domain are selected as a suitable region to embed for the watermark which doesn't influence the image quality and maintains better robustness.

3. PROPOSED WORK

In the proposed method, an enhanced benchmark dataset is fed into the transformer encoder together with them being used to fine-tune the transformer model. Image authentication, tamper detection, and recovery, as well as watermark embedding capabilities, are assessed using multi-image encoder feature maps that are produced as watermarks that are embedded in the cover image. The taxonomy of the suggested model is shown in *figure 1*, where the dimensionality of the transformer-encoded feature maps is reduced using the PCA approach to obtain the final watermarks. The DWT approach is applied to the original image at the primary level of embedding, as illustrated in *fig. 3f*. After that, the 8×8 blocks of the low-frequency LL sub-band are divided up, and each block is then broken down into a QR matrix. The primary watermarked image is provided by DWT after the first two images are inserted in the Q matrix, followed by inverse QR. Two-level SWT is applied on the primary watermarked image at the secondary level of embedding, as illustrated in *Figures 3(d) and 3(e)*, and the other two watermarks are embedded in the mid-frequency coefficient together with a secret key that was produced randomly. During the extraction phase, two levels of data authentication are verified. First, the secret key matches, then the second level of the authentic fingerprint image is extracted from stationary wavelet coefficients and matched with the original watermark. If the image is not authentic, then extract the other embedded images from the Q matrix of low-frequency sub-bands to detect tampered regions. The self-recovered image is fed as input to the transformer decoder which extracts the original image. Identify tampered regions by extracting the other embedded images from the Q matrix of low-frequency sub-bands if the image is not authentic otherwise no extraction process. The transformer decoder receives the extracted self-recovery watermark as input, which produces the exact original image.

3.1 Watermark generation and embedding

To process a 2D image, the image is split into patches 'P' of fixed size, producing an N number of patches, where N= HW/P^2C, where C stands for the number of channels and P stands for patch size and H & W are the original image dimension [23]. The image patches are linearly flattened through a linear patch projection layer, the image patches are converted into a vector X_i along with embedding matrix E as shown in *eq. (7)*. Followed by adding positional embedding information to the linear projection layer results in the final input context vector Z_1, as stated in *eq. (8)*,

To process the 2D image, the image is divided into patches P of fixed size resulting in *N* number of patches $N= HW/P^2C$, where *P* represents the height and width of fixed patch size and *H & W* are the size of the original image, C represents the number of channels [23]. Image patches are linearly flattened through linear patch projection which converts the image patches into matrix *E* with dimension vector size X_i as represented in *eq (7)*. Adding positional embedding information to the input sequence vector which produces the final input context vector Z_1 as shown in *eq (8)*.

$$F_i=[\, X_1E,\ X_2E\ldots\ldots X_nE], \tag{7}$$

$$Z_1 = F_i + E_{pos} \tag{8}$$

where X represents the number of vectors, E represents the embedding matrix, Z_1 represents vectorized context vector input, E_{pos} represents positional embedding.

Context vector Z_1 is fed as an input to the transformer encoder. Transformer encoder feature maps generation is mentioned in *equation (1-5)* in *section 2.1*. To reduce the dimensionality of the multi-image watermark feature maps Y_i, PCA technique is applied to each encoded transformer image that produces the final watermark image.

Algorithm

Watermark generation

1. The transformer model is fine-tuned on the standard benchmark dataset.
2. A benchmark image is given as input by splitting the image into several patches and flattening the image patches.
3. Create linear dimensional embeddings from these flattened image patches and concatenate positional embeddings to the image patch sequence.
4. Feed the sequence as an input to the transformer encoder
5. Self-attention is computed and normalized followed by the feed-forward network that generates an encoded feature image.
6. Encoded image dimensionality is reduced using the PCA technique.

Watermark Embedding

1. Embed the first two encoded image features as watermark1 represented as WM_1 and the other two encoded image features as watermark2 represented as WM_2.
2. Apply DWT on the original image O
3. Divide the LL band into 8 × 8 blocks and perform QR decomposition
4. Embed watermark1 in matrix Q' using the secret key,
   Q'= $Q_i + \alpha WM_1$
5. Perform inverse IQR and IDWT

Output: Primary Watermark Image

6. Apply 2-Level SWT on the primary watermarked image
7. Embed watermark2 on LH2 mid-frequency band.
8. Perform 2-Level inverse SWT

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 339-344 | e-ISSN: 2347-470X

3.2 Extraction Algorithm

To authenticate the watermarked image and to detect the tampered region extraction process is carried out as follows.

Algorithm

1. Perform 2-Level SWT on watermarked image and extract watermark2 from the LH2 sub-band.
2. Compare extracted watermark WM' with the original watermark WM, if it is true, then the image is authentic allowing to extract watermark and locate tampered region, otherwise, No extraction process.
3. To recover the tampered region perform Inverse SWT and apply DWT on the image.
4. Divide the LL low-frequency band into 8 × 8 blocks and perform QR decomposition.
5. Extract the watermark1 from the Q matrix and apply inverse PCA on the self-recovery image.
6. Fed the extracted image into the transformer decoder to recover the tampered image.

4. RESULT AND DISCUSSION

The proposed model is evaluated for various types of attacks to verify the robustness and imperceptibility. A benchmark database is used to evaluate the proposed model for the detection of image manipulation traces of size 512 × 512, 256 × 256. Figure 4 shows a few standard datasets and sample user-generated images of various sizes. The model is validated for various noise attacks, median filter attacks, and geometric attacks. The proposed model is evaluated for various attacks like Salt and pepper noise, Gaussian noise, Rotation, copy-paste, and flipping results are shown in Table 1. Figure 3 shows the original image, encoded watermark image, watermarked image, stationary wavelet transform image, and Discrete wavelet image. Our proposed model presented two different wavelets transform along with QR decomposition and invariant multi-image watermark features are generated using the Transformer encoder-decoder model. Extracted watermarks are highly invariant against rotation, flipping, and scaling attacks. Transformer-based encoded feature map shows improved result than deep learning techniques as it extracts invariant feature maps. For future work, the transformer model can extend to various applications like proof of ownership, and copyright protection.

4.1 Quality Metrics

Various metrics are provided here to demonstrate the proposed scheme's performance against tamper detection and localization. The performance evaluated for various attacks was measured by Peak-Signal-to-Noise-Ratio (PSNR) and Normalized correlation coefficient (NCC) as represented below. PSNR (Peak Signal-to-Noise Ratio) [26] evaluates the visual perception by comparing the difference between the original image and watermarked image and is deemed to be of acceptable quality if the PSNR score is more than 25 to 30 dB. The PSNR is determined by using the following formula:

$$\text{PSNR} = 10\log 10\left(\frac{(Max-1)^2}{\left[\sum_{M,N}(I_1(m,n)-I_2(m,n))^2\right]\frac{1}{M*N}}\right) \quad (9)$$

where M and N are the numbers of rows and columns in the input images. The robustness of the embedded watermark is measured using the metric Normalized Correlation Coefficient. NCC measures the robustness between the original watermark and extracted watermark. NCC can be calculated using below shown equation:

$$\text{NCC} = \frac{\sum_{i=1}^{n_L}\sum_{j=1}^{n_K}\left(|W(i,j)+W'(i,j)|/2\right)}{n_L \times n_K} \quad (10)$$

4.2 Comparative Analysis

The two-watermark extracted from the watermarked image showed the highest imperceptibility for the recovered image with the PSNR value of 59.05 dB and robustness attained high with an NCC value of 0.997. *Figure 5*, shows the performance of the existing paper compared with the proposed work which shows that watermark feature maps generated using the transformer model attained high robustness and imperceptibility.

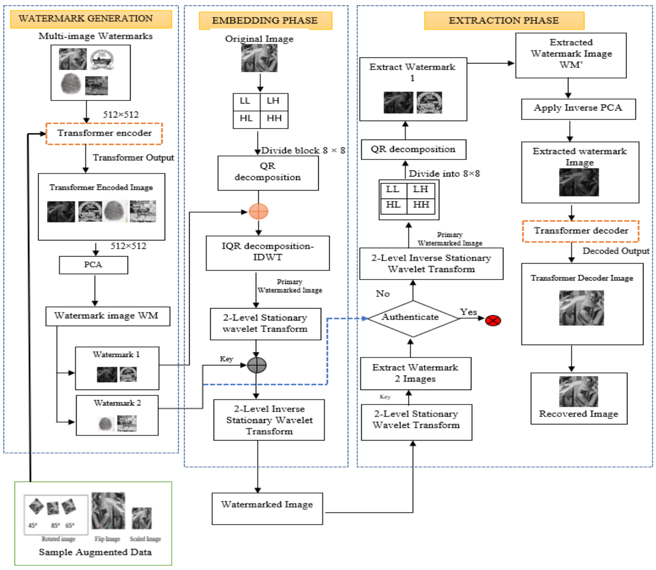

Figure 1: Taxonomy of Proposed Architecture

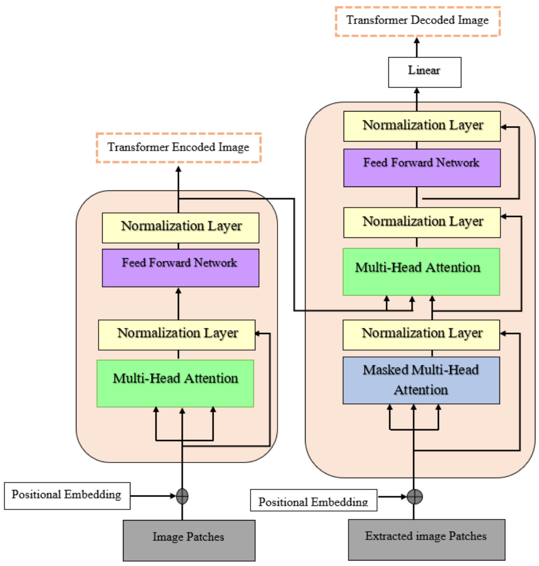

Figure 2: Transformer Encoder-Decoder Architecture

FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 339-344 | e-ISSN: 2347-470X

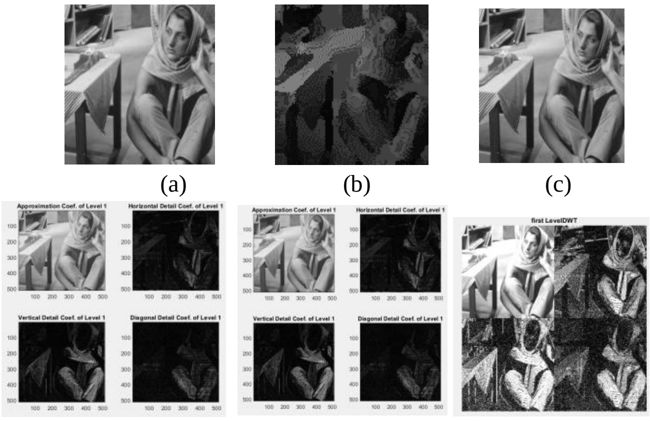

Figure 3: a) Original Image, b) Watermark Image, c) Watermarked Image d) 1-Level Stationary wavelet transform, e) 2-Level Stationary wavelet transform, f) Discrete Wavelet transform

(a) (b) (c) (d) (e)

(f) (g) (h) (i)

Figure 4: Sample images of Standard dataset (a-e), (f-i) User Generated Images

Figure 5: Comparative Analysis of existing works Vs Proposed

Table 1: **Result on various Tampered images and its result**

Watermarked Image	Attacks &Tampered Image	Tamper detection	Extracted WM1 With NCC value	Extracted WM2 With NCC value	Recovered Image	PSNR for recovered image
	Copy-Paste Image		NCC = 0.97	NCC = 0.989		56.45 dB
	Image Splicing/Cropping		NCC = 0.97	NCC = 0.989		56.45 dB
	Flipped Image		NCC = 0.987	NCC = 0.98		49.02 dB
	Rotated 45°		NCC = 0.974	NCC = 0.9762		55.30 dB
	Rotated 65°		NCC = 0.982	NCC = 0.982		55.2 dB
	Salt & Pepper Noise		NCC = 0.97	NCC = 0.97		58.45 dB
	Gaussian Noise		NCC = 0.964	NCC = 0.98		59.05 dB
	Median Filter		NCC = 0.95	NCC = 0.952		52.4 dB

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 339-344 | e-ISSN: 2347-470X

5. CONCLUSION

In digital media platforms securing images from manipulation is a challenging task in the real world. Watermarking technique laid a path for digital data to be secured against various applications like Authentication and Tamper detection. The attention-based model shows more performance than state-of-the-art deep learning techniques. We have attempted the transformer model for the first time for watermarking applications and succeeded with higher results. Our proposed model presented two different wavelets transform along with QR decomposition and invariant multi-image watermark features are generated using the Transformer encoder-decoder model. Extracted watermarks are highly invariant against rotation, flipping, and scaling attacks. Transformer-based encoded feature map shows improved result than deep learning techniques as it extracts invariant feature maps. For future work, the transformer model can extended to various applications like proof of ownership, and copyright protection. Also, the complexity of transformer model can be reduced by modifying attentions into linear attentions.

REFERENCES

[1] N. Shyamala and D. S. Geetha, "Compression of Medical Images Using Wavelet Transform and Metaheuristic Algorithm for Telemedicine Applications," *IJEER*, vol. 10(2), pp. 161-166, 2022.
[2] L.Agilandeeswari, Sunny. S (2013) Secure data sharing of patient record in cloud environment using attribute based encryption, International Journal of Applied Engineering Research, (IJAER), Vol 8, Spec. Issue 19, pp 2423-2426.
[3] Singh, J., & Singla, M. (2022). Image Steganography Technique based on Singular Value Decomposition and Discrete Wavelet Transform. *International Journal of Electrical and Electronics Research*, *10*(2), 122-125.
[4] G. Zhou and D. Lv, "An overview of digital watermarking in image forensics," *2011 Fourth International Joint Conference on Computational Sciences and Optimization*, pp. 332-335, April 2011.
[5] S. Kaur and J. Patel, "A Robust Image Mosaicing Technique Using Frequency Domain.," *IJEER*, vol. 6(1), pp. 1-8, 2018.
[6] T. K. Tsui, X. P. Zhang, and D. Androutsos, "Color image watermarking using multidimensional Fourier transforms, "*IEEE Transactions on Information Forensics and Security*, vol. 3(1), pp. 16-28, 2008.
[7] Agilandeeswari, L., & Ganesan, K. (2018). RST invariant robust video watermarking algorithm using quaternion curvelet transform. *Multimedia Tools and Applications*, *77*(19), 25431-25474.
[8] D. Ariatmanto and F. Ernawan, "Adaptive scaling factors based on the impact of selected DCT coefficients for image watermarking", *Journal of King Saud University-Computer and Information Sciences*, 2020.
[9] L.Agilandeeswari, K.Ganesan, K.Muralibabu, (2013). A Side View Based Video in Video Watermarking Using DWT and Hilbert Transform, Security in Computing and Communications, Communications in Computer and Information Science (CCIS) Series – Springer, page 366-367
[10] L.Agilandeeswari, K.Muralibabu (2013). A Robust Video Watermarking Algorithm for Content Authentication using Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD), International Journal of Security and its Applications, Vol 7(4), pp. 145-158.
[11] P. V. Nagarjuna and K. T. Ranjeet, "Robust blind digital image watermarking scheme based on stationary wavelet transform," *In 2013 Sixth International Conference on Contemporary Computing (IC3). IEEE.*, pp. 451-454, 2013, August.
[12] L.Agilandeeswari and K. Ganesan. (2016) An Efficient Hilbert and Integer Wavelet Transform based Video Watermarking. Journal of Engineering Science and Technology, Vol.11, Issue.3, Page. 327 - 345
[13] Agilandeeswari, L., Prabukumar, M., & Alenizi, F. A. (2023). A robust semi-fragile watermarking system using Pseudo-Zernike moments and dual tree complex wavelet transform for social media content authentication. Multimedia Tools and Applications, 1-53.
[14] L.Agilandeeswari, K.Ganesan, (2016) "A Robust Color Video Watermarking scheme based on Hybrid Embedding Techniques" MULTIMEDIA TOOLS AND APPLICATIONS (Springer), Vol. 75, Issue 14, Page 8745 – 8780
[15] P. T. Nha, T. M. Thanh, and N. T. Phong, "Consideration of a robust watermarking algorithm for color image using improved QR decomposition," *Soft Computing*, vol. 26(11), pp. 5069-5093, 2022.
[16] L. Y. Hsu, H. T. Hu, and H. H. Chou, "A high-capacity QRD-based blind color image watermarking algorithm incorporated with AI technologies," *Expert Systems with Applications*, vol. 117134, pp. 199, 2022.
[17] Rana, M. M., & Abdelhadi, A. (2020, February). Optimal Image Watermark Technique Using Singular Value Decomposition with PCA. In *2020 22nd International Conference on Advanced Communication Technology (ICACT)* (pp. 342-347). IEEE
[18] M. Rezaei and H. Taheri, "Digital image self-recovery using CNN networks," *Optik*, p. 169345, 2022.
[19] F. Khan and M. S. Raeen, "Robust and Blind Multiple Image Watermarking Using CNN and DWT in Video".
[20] T. Song, X. Yu, S. Yu, Z. Ren, and Y. Qu, "Feature Extraction Processing Method of Medical Image Fusion Based on Neural Network Algorithm," *Complexity*, 2021
[21] Loganathan, A., & Kaliyaperumal, G. (2016). An adaptive HVS based video watermarking scheme for multiple watermarks using BAM neural networks and fuzzy inference system. Expert Systems with Applications, 63, 412-434.
[22] Agilandeeswari, L., Ganesan, K. A bi-directional associative memory based multiple image watermarking on cover video. Multimed Tools Appl 75, 7211–7256 (2016).
[23] A. Dosovitskiy, L. K. A. Beyer, D. Weissenborn, X. Zhai, T. Unterthiner, ... and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020
[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. .. Gomez, and I. Polosukhin (2017). "Attention is all you need," *Advances in neural information processing systems*, pp. 30, 2017.
[25] N. N. Hurrah, S. A. Parah, N. A. Loan, J. A. Sheikh, M. Elhoseny and K. Muhammad, "Dual watermarking framework for privacy protection and content authentication of multimedia," *Future generation computer Systems*, vol. 94, pp. 654-673, 2019.
[26] A. Soualmi, A. Alti and L. Laouamer, "An Imperceptible Watermarking Scheme for Medical Image Tamper Detection," *International Journal of Information Security and Privacy (IJISP)*, vol. 16(1), pp. 1-18, 2022.