State-Of-The-Art Application Of Artificial Neural Network In Digital Watermarking And The Way Forward

[1] Rashidah F. Olanrewaju, [2] Aburas Ali Abdurazzag and [3]Othman O. Khalifa, [4]Aishah Abdalla

Department of Electrical and Computer Engineering,
International Islamic University Malaysia,
P.O. Box 10, 50728 Kuala Lumpur, Malaysia.
Email: [1]frashidah@yahoo.com, [2]aburas06@iiu.edu.my, [3]khalifa@iiu.edu.my, [4]aisha@iiu.edu.my

ABSTRACT

Several high-ranking watermarking schemes using neural networks have been proposed in order to make the watermark stronger to resist attacks. The ability of Artificial Neural Network, ANN to learn, do mapping, classify, and adapt has increased the interest of researcher in application of different types ANN in watermarking. In this paper, ANN based approached have been categorized based on their application to different components of watermarking such as; capacity estimate, watermark embedding, recovery of watermark and error rate detection. We propose a new component of water marking, Secure Region, SR in which, ANN can be used to identify such region within the estimated capacity. Hence an attack-proof watermarking system can be achieved.

Keywords
(*ANN, capacity estimate, Watermarking, Secure Region SR)*

1.0 INTRODUCTION

With the wide spread, complex use, and transfer of digital media, secure media transfer has been a concern to all the multimedia industries. This concern is appropriately addressed by digital watermark. Digital watermarking is a novel approach that involves embedding of digital mark into a multimedia object (cover work) such that it is robust, secure and imperceptible to the human observer, but can be detected algorithmically. Due to digital watermark crucial features such as; imperceptibility, inseparability of the content from the watermark, and it's intrinsic ability to undergo same transformation as experienced by the cover work, has made it superior and preferable over other traditional methods of protecting data integrity, authentication of information resources, ownership assertion, confidentiality, copy protection, data monitoring and tracking. This preference has been proven experimentally (Schmidt et al., 2008) to provide improved security. An additional cofounding factor of a watermarking system is robustness of the algorithm against attacks. ***Robust watermarks*** are designed to survive common distortion and resist malicious attacks. All applications presupposing security of the watermarking systems require this type of marks in order to survive any kind of alterations or intentional removal introduced by standard or malicious processing and attacks. Many watermarking are robust watermarking (Lu et al. 2009, Celik et al 2008, Shih and Wu 2005). Although some watermarking schemes are embedded with a very low robustness, ***Fragile*** *watermarks*. This type of watermark can be destroyed even with slightest manipulations, as long as the document has been illegally manipulated.

Mostly they are use to check integrity of objects and might be useful if digital images are used as evidence in court. It is also use to verify medical content due to sensitivity of medical images. Applications that uses fragile watermark can be found in (Ting el at 2007). Watermarking scheme is mostly designed in view of its application, and there is no such watermarking scheme that can perform well under all hostile attacks (Khan et al. 2008). For example, (Parthasarath, et al. 2007) designed a DCT content based watermarking scheme. In their method, texture, luminance, corner and the edge information in the image to generate a mask of watermarking scheme. It was found that their scheme is robust to JPEG compression, median filtering and contrast-sharpening filter. However, it is not robust against scaling, rotation and high noise levels. For these reasons, the development and evaluation of watermarking algorithm is a challenging task.

When Least Significant Bit, LSB coding was used in watermarking (Van et al. 1994)], the most common and easy to use watermarking scheme. It was as if it cannot be attacked. However, with advances in algorithm, various steganalysis surfaced (Ge et al. 2007, Luo et al. 2005) which found LSB to be vulnerable to even slightest image distortion. Researcher later improved on LSB and algorithms like Patchwork (Bender et al. 1996), correlation-based techniques, feature point and chaotic map were in use. However, due to direct manipulation and linear addition of watermark to cover media of these techniques (spatial domain techniques), make them vulnerable to attacks like compression, geometric distortion, image degradation and computational inefficiency. To obtain better imperceptibility as well as robustness, the addition of the watermark is done in transformed domain. Scientists exploited the benefits of transform domain transformation like Discrete Cosine Transform, DCT (Choi *et al.* 2008), Discrete Fourier Transform, DFT (Sang and Alam 2008), Hadamard Transform (Abdallah *et al.* 2006), and Discrete Wavelet Transform, DWT (Senthil and Bhaskaran, 2007) to build a robust watermarking algorithm. These current schemes are not totally robust against all attacks. However when the Principles of neurocomputing, and their usage in science and technology surfaced, (Ham and Kostanic 2001), the use of neural network based watermarking was successful. This is because NN based scheme performs well under a specific sets of conceivable attacks and work well with Human Visual System (HVS).

2.0 ARTIFICIAL NEURAL NETWORK IN WATERMARKING

ANN is a computational model based on biological neural networks (Haykin, 2008). It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. ANN can learn from the data and generalize things learned. They extract the essential characteristics from the numerical data as opposed to memorizing all of it. This offers a convenient way to reduce the amount of data as well as to form an implicit model without having to form a traditional, physical model of the underlying phenomenon.

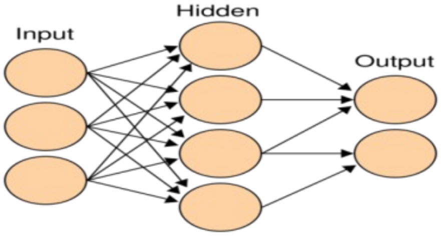

Figure 1: General Architecture of Feed forward NN

While there are numerous different artificial neural network architectures such as Adaline, back propagation, Bidirectional Associative Memory, Boltzmann Machine, Counter-propagation Network, Self Organizing Map, adaptive resonance theory networks etc have been studied by researchers. The most successful applications in data mining of neural networks have been multilayer feedforward networks (Palit and Popovic, 2005). This network is composed of a set of nodes and connections arranged in layers. The connections are typically formed by connecting each of the nodes in a given layer to all of the neurons in the next layer. In this way every node in a given layer is connected to every other node in the next layer. The network is made up of three layers; in which there is an input layer consisting of nodes that simply accept the input and feeds into the hidden layer. The hidden layer, in turn, feeds into the output layer. The actual processing in the network occurs in the nodes of the hidden layer and the output layer. Figure 1.0 depict a simple architecture of a feed forward networks.

Recent works have taken advantage of artificial intelligence in Neural Network to design a robust watermarking system. Owing to the inherent characteristic of Neural Network like learning and adaptive capabilities, pattern mapping and classification and ability to generalize, not only to reproduce previously seen data, but also provide correct predictions in similar situations gives the trained networks ability to recover the watermark from the watermarked data. Examples of application of ANN in watermark include capacity estimator, error rate prediction, embedding and recovery of mark, detection of tempering etc.

2.1 ANN as a Watermark Capacity Estimator

Watermarking can be considered as a communication problem where the embedded data is the signal to be transmitted. A fundamental problem is the embedding capacity (Wu and Liu 2003). Watermark capacity of a digital media is defined as maximum achievable bits that can be embedded in a host signal. Here, the watermark capacity corresponds to the communication capacity of the watermark channel. Although any rate less than the channel can be achievable but the channel capacity may not be necessarily realizable (Cox et al 1999). There are some existing methods that estimate the capacity of watermark (Servetto et al 1998, Wong and Au 2003) though they are not ANN based. Most of the previous works on watermark embedding capacity focused on spread spectrum technique, where the mark is spread and modulated with a pseudorandom signal (noise like) and added to the host signal. This can be detected via correlation.

Researchers like (Fang and Zhang, 2005) studied the bounds of embedding capacity in a blind watermarking algorithm based on Hopfield neural network. They used basin attraction via Hamming distance to restrict the capacity of the watermark. Shi-Chun *et al* (2002) modeled Human Visual System (HVS) using Feed forward ANN-based image-adaptive method in order to decide the watermark strength of DCT coefficients. The experimental results showed that the method can increase the watermarking strength and robustness of digital watermark is enhanced. Similarly, (Jin and Wang, 2007) indicated that using ANN in different textural features of each DCT block and luminance of an image can be implored to decide adaptively the watermarking embedding strength. Ming *et al* (2003) defined an RBF neural networks based algorithm that control and create the maximum image-adaptive strength watermark.

2.2 ANN and Watermarking Embedding

A typical watermarking scheme is illustrated in Fig. 2.0 showing the two main building blocks, embedder/encoder with a respective extractor/decoder. In general, embedding process can be understood as the combination of watermark signal and original media. The watermark embedder inserts a watermark into the cover signal and the watermark recovery block extracts/decodes or detects the presence of watermark signal.

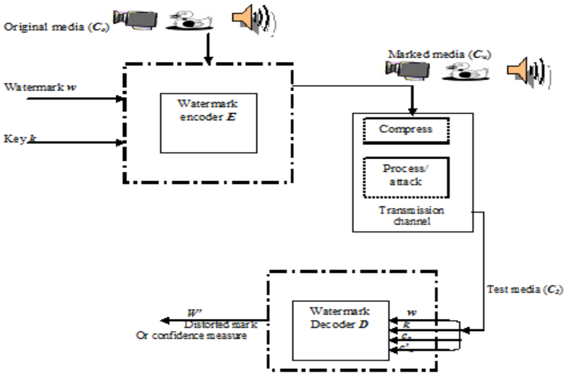

Figure 2: Watermarking scheme with Embedder and Extractor

In ANN embedding, transform domain such as DCT, DWT, DFT etc can be use to decompose the media coefficients. A chosen NN can then be use to train these information and memorize the relationship among the coefficients that will serve as an input to the layers of network. The watermark sequence is embedded into the host media adaptively by adjusting weights to produce corresponding target watermark media at the output layer. This watermark can be stored at the same time if the original media is necessary in the watermark detection.

Li and Wang (2007), demonstrated the combination of video watermarking scheme based on RBF Neural Network and 3D-DWT. The algorithm established a relationship among the coefficients of the discrete wavelet decomposition along the spatial and temporal axis using the RBF networks. It was applied during the embedding and extracting process to limit the pirate attack. Chang (2005), proposed a Full Counterpropagation Network (FCNN) to insert the watermark into synapses of FCNN rather than the cover image. . This method increase robustness and reduce imperceptibility problems to a great extent. Similarly, (Bansal and Bhadauria 2005), introduced a watermarking scheme based on Backpropagation Neural Network to train a given cover image in which the trained network weights is hidden within the original cover image itself. This increase robustness and preserve the cover image. Majhi, and Shalabia (2005), also showed how to again computation efficiency as well as memory requirement by embedding and detecting watermark using a modified Functional Link Artificial Neural Network (FLANN) and Discrete Cosine Transform (DCT). The algorithm was trained using back propagation algorithm till the MSE is minimum/constant. The results showed that the scheme reduced computational cost in the training stage and maintains a good performance of approximation. Tsai (2007), proposed a decision processor based watermarking in wavelet domain using NN that incorporate the HVS model. This technique is use extract and enhances imperceptibility of the watermarked media.

2.3 Detection of Error Rate using ANN

Detection is to find if the watermark is present or not. During the watermark recovery, just before the extraction, a detection threshold may be set to establish that, do a watermarked media contain mark or not? Only two values are set true or false/ 1 or 0. Usually the correlation test method is used in the watermark detection.

There are two types of detection error testing (Cox et al 1999), the False negative when the watermark detector fails to detect a watermark that is present and False positive; when the watermark detector indicate the presence of a watermark in an unwatermarked media.

The watermark detection error rate help embed more watermark messages while keeping an acceptable detection error rate, and it is useful for the design of the general algorithms of watermarking and detection. According to the experimental results obtained by (Zhang *et al* 2007), the detection error rate of watermark is mainly influenced by the watermark average energy and the watermarking capacity. The error rate rises with the increase of watermarking capacity. When the channel coding is used, the watermarking error rate drops with the decrease of the payload capacity of watermarking. Naoe and Takefuji *et al* (2008), proposed a frequency based transform watermarking using NN on YCbCr domain to detect a hidden bit codes from the content. A conditioned neural network is used as a classifier to recognize a hidden bit pattern from the content which embedder associated to the target content. They claimed that the method does not damage the target content. However, the extraction keys must be shared among embedder and extractor in order to extract a proper hidden bit codes from the target content.

2.4 ANN and Watermark Recovery

Recovery is a crucial part in the watermarking framework. It allows the owner to be identified and provides information to the intended recipient. The marked media may be subjected to various kinds of processing and attacks before feeding into an extractor, therefore bit-by-bit accuracy in extracting these hidden data from the marked media is highly desirable. Due to neural networks possessing the learning and adaptive capabilities, the trained neural networks almost exactly recover watermark from watermarked media (Zhang and Zhang 2005). Zhang *et al* (2002), explored the learning and adaptive capability of BPN neural networks to learn the characteristics of relationship between the watermark and the watermarked image. The information from the trained BPN was in turn used during the extraction process which reduced the false recovery. Similarly, (Wen *et al* 2008), proposed a blind digital watermarking algorithm based on probabilistic neural network. Therein the watermark bits were embedded into the edges and textures of the image using the statistical properties of the dual-tree wavelet transform (DTCWT) and the human visual system (HVS). ANN was use at the extracting phase to recover the watermark.

3.0 RESULT AND DISCUSION

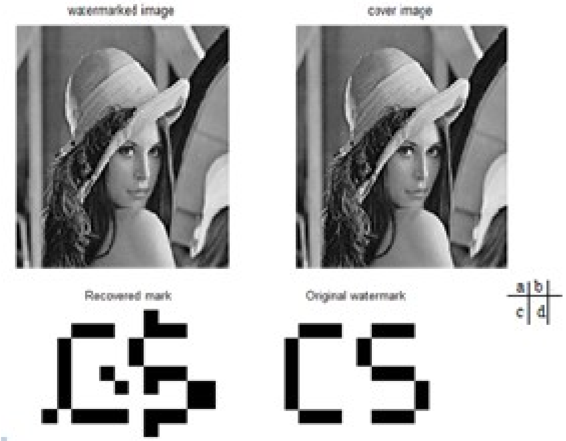

Figure 3: Image to demonstrate watermarking process. (a) The watermarked image. (b) The host image. (c) The recovered watermark message. (d) The original watermark.

From results obtained in figure 3 above, the watermarked (3a) image is highly imperceptible, that is the watermarked image and the cover images (3b) are alike without any visual degradation. However, from the extractor block, it can be seen clearly that the extracted watermark 3c is damaged especially the letter 'S'. This is due to embedding in unsafe area. With introduction of Safe Region SR, the watermark will be safely recovered without severe damage because the region will be carefully selected.

4.0 WATERMARKING WITH ANN; A NEW COMPONENT TO BE CONSIDERED.

From the above literature review, several studies have been done using ANN in different components of watermarking; from capacity estimate to error rate detection. However a very important aspect of watermarking component is yet to be addressed. This is the Secure Region (SR) within the capacity estimated. Secure Region (SR) is an identified region in the host media in which when watermark is hidden therein, it will not be destroyed nor degraded. So far, to the best of our knowledge, no work has been done to distinctly address the identification and locating of SR using ANN. This suggests that investigating and researching in such niche area will lead to design of an attack proof watermarking scheme as well as solve the problem of image degradation. The SR is under investigation by the authors.

5.0 CONCLUSION

A review of the state of-art of application of ANN in watermarking components has been carried out and the strength of each has been highlighted. We have proposed SR as another new direction to be looked into among the watermarking components as a means of achieving attack proof watermarking system.

REFERENCES

Abdallah, E. E., Hamza, A. B. & Bhattacharya, P. (2006). A Robust Block-Based Image Watermarking Scheme Using Fast Hadamard Transform and Singular Value Decomposition. *IEEE Int. Conf. on Pattern Recognition* , 3, 673 – 676.

Bansal A. E & Bhadauria, S. S. (2005). Watermarking Using Neural Network And Hiding The Trained Network Within The Cover Image. *Journal of Theoretical and Applied Information Technology,* 663-670.

Bender, W., Gruhl, D., Morimoto, N. &. Lu, A. (1996). Techniques for data hiding," *IBM Systems Journal*, 35,. 3/4, 313–336,

Celik, M. U, Lemma, A. N., Katzenbeisser, S., & Veen, M. V. (2008). Lookup-Table-Based

Chang, C. Y., (2005). The Application of a Full Counterpropagation Neural Network to Image Watermarking. *Networking, Sensing and Control*, pp 993 – 998.

Choi, H. J., Seo, Y. H., Yoo, J. S. & Kim, D. W. (2008). Digital watermarking technique for Holography interference Patterns in a Transform Domain. *Optical and Lasers in Engineering,* 46, 4, 343-348.

Cox, I. J., Miller, M. L. & Bloom, J. A. (1999). *Digital Watermarking.* Morgan Kaufmann*,* 1999.

Ge, S., Gao, Y. & Wang, R. (2007) "Least Significant Bit Steganography Detection With Machine Learning Techniques", *Conference On Knowledge Discovery In Data, Proceedings Int. Workshop On Domain Driven ,* 24 – 32.

Ham, F. M. & Kostanic, I. (2002). *Principles of Neurocomputing for Science & Engineering*. Singapore: Mc.GrawHill, 136-140.

Haykin, S. (2008). *Neural Networks and Learning Machines***,** *3rd edition*. Prentice Hall.

Jin, C. & Wang, S. (2007). Applications of a Neural Network to Estimate Watermark Embedding Strength. *IEEE 8th Int. Workshop on Image Analysis for Multimedia Interactive Services*, 68-68.

Khan, A., Tahir, S. F., Majid, A. & Choi, T. S. (2008). Machine Learning Based Adaptive Watermark Decoding in view of Anticipated Attack. *Pattern Recognition,* 41, 2594 – 2610.

Li, X. and Wang, R.(2007). A Video Watermarking Scheme based on 3D-DWT and Neural Network. *Ninth IEEE International Symposium on Multimedia*, pp110 – 115.

Lu, W., Sun, W., & Lu, H. (2009). Robust Watermarking based on DWT and Nonnegative Matrix Factorization, *Computers & Electrical Engineering*, 35, 183-188.

Luo, X., Liu, B. & Liu, F. (2005). Improved RS Method For Detection of LSB Steganography. *Lecture Notes In Computer Science, Computational Science And Its Applications.*

Majhi, B. & Shalabi, H., (2005). An Improved Scheme For Digital Watermarking Using Functional Link Artificial Neural Network, *Journal Of Computer Science*, 1, 2, 169-174.

Ming, Z. Z., Rong-Yon, L., & Le, W., (2003). Adaptive Watermark Scheme With RBF Neural Networks. *IEEE Int. Conf. Neural Networks & Signal Processing*, vol. 2, pp 1517 – 1520.

Naoe, K. & Takefuji, Y. (2008). Damageless Information Hiding using Neural Network on YCbCr Domain. *Int. Journal of Computer Science and Network Security*, 8 9.

Palit, A. K. & Popovic, D,. (2005). *Computational Intelligence in Time Series Forecasting*. Springer.

Parthasarath, A.K. & Kak, S., 2007). An Improved Method of Content Based Image Watermarking. *IEEE Transactions On Broadcasting,* vol. 53, No. 2.

Sang, J. & Alam, M. S.(2008). Fragility and Robustness of Binary-Phase-Only-Filter-Based Fragile/Semi fragile Digital Image Watermarking. *IEEE Transactions on Instrumentation and Measurement*, 57, 3, 595 – 606.

Schmidt, T., Rahnama, H., & Sadeghian, A. (2008). A Review of Applications of Artificial Neural Networks in Cryptosystems, *World Automation Congress,* 1-6.
Secure Client-Side Embedding for Spread-Spectrum Watermarks, *IEEE Transactions on Information Forensics and Security*, 3, 3, 475 – 487.

Senthil, V., Bhaskaran, R. (2007). Wavelet Based Digital Image Watermarking with Robustness against Geometric Attacks. *Int. Conf. on Computational Intelligence and Multimedia Applications,* 4, 89 – 93.

Servetto, S. D., Podichuk, C. I. & Tamachandran, K. (1998). Capacity Issues In Digital Image Watermarking. *Int. Conf. on Image Processing,* 1, 445- 449.

Shi-Chun, M., Ren-Hou, L., Hong-Mei, D. & Yun-Kuan, W. (2002). Decision Of Image Watermarking Strength Based On Artificial Neural-Networks. *IEEE Proceedings of the 9th Int. Conf. on Neural Information Processing*, 5, 2430-2434.

Shih Y. F. & Wu, Y. T(2005). Robust Watermarking and Compression for Medical images based on Genetic Algorithms. . *Int. Journal of Information Sciences*, 175, 200–216

Ting, G. C. W., Goi, B. M. & Heng, S. H. (2007). A Fragile Watermarking Scheme Protecting Originator's Rights for Multimedia Service. *Lecture Notes in Computer Science on Computational Science and Its Applications,* 4705/2007, 644-454.

Tsai, H. H. (2007). Decision-Based Hybrid Image Watermarking in Wavelet Domain Using HVS and Neural Networks. *D. Liu et al. (Eds.): ISNN*, Part III, 4493, 904–913.

Van., R. G., Tirkel, S. A. Z. & Osborene, C. F. (1994) A Digital Watermark" in Proc. *IEEE Int. Conf. Image Processing*, 2, 86-92.

Wen, X. B., Zhang, H., Xu, X. Q. & Quan, J. J. (2008). A New Watermarking Approach Based On Probabilistic Neural Network In Wavelet Domain**.** *Soft Computing-A Fusion of Foundations, Methodologies and Applications***,** 13, 4, 355-360.

Wong, H. W. P. & Au, O. C. (2003). A Capacity Estimation technique for JPEG to JPEG Image watermarking. *IEEE Transactions On Circuits and Systems for Video Technology,* 13,8 746-752 .

Wu, M., & Liu B. (2003). Data Hiding In Image and Video: Part I—Fundamental Issues and Solutions. *IEEE, Transactions On Image Processing,* 12, 16. 685-695.

Zhang, F. & Zhang, H. (2005). Applications of a Neural Network to Watermarking Capacity of Digital Image. *Neurocomputing*, 67, 345–349.

Zhang, F., Zhang, X. & Zhang, H. (2007). Digital Image Watermarking Capacity and Detection error rate. *Pattern Recognition Letters*, 28, 1–10.

Zhang, J., Wang, N., & Xiong, F., (2002). Hiding a Logo Watermark into the Multiwavelet Domain Using Neural Networks. *Proceedings of 14th IEEE Int. Conf. on Tools with Artificial Intelligence, (ICTAI 02)*, 477-482.

Zhang, X. H. & Zhang, F. (2005). A Blind Watermarking Algorithm Based on Neural Network, *Int. Conf. on Neural Networks and Brain*, 2