Moiré Pattern Detection using Wavelet Decomposition and Convolutional Neural Network

Eldho Abraham
Amadeus Software Labs Pvt. Ltd.
Bangalore, India
eldo.abrahm@gmail.com, eldho.abraham@amadeus.com

***Abstract*—Moiré patterns are interference patterns that are produced due to the overlap of the digital grids of the camera sensor resulting in a high-frequency noise in the image. This paper proposes a new method to detect Moiré patterns using wavelet decomposition and a multi-input deep Convolutional Neural Network (CNN), for images captured from a computer screen. Also, this paper proposes a method to use the normalized intensity values in the image, as weights for the frequency strength of Moiré pattern. The CNN model created with this approach is robust to high background frequencies other than those of Moiré patterns, as the model is trained using images captured considering diverse scenarios. We have tested this model in receipt scanning application, to detect the Moiré patterns produced in the images captured from a computer screen, and achieved an accuracy of 98.4%.**

***Index Terms*—Moiré, CNN, Wavelet Decomposition, Pattern Recognition**

I. Introduction

Capturing an image of a scene from a computer screen is different from capturing the same scene in the real world, as it produces unwanted high frequency noise called as Moiré patterns. These patterns are large-scale interference patterns that are produced when an opaque ruled pattern with transparent gaps is overlaid on another similar pattern. When an image is captured from a computer screen, Moiré patterns are produced due to the overlap of the digital grids of the camera sensor resulting in a high-frequency noise in the image. Detection and removal of these patterns are important and it depends on the overall application. Critical applications like biometric-based person identification (e.g., face recognition) deals with face spoofing related issues, where the system can be cheated by showing a picture of another person in an electronic display. Here detection of the Moiré pattern can help to identify the spoofing. In receipt scanning related applications, if a user captures a receipt from a computer screen, it can produce high-frequency Moiré patterns that can lead to the failure of the OCR module. For achieving a good accuracy of the overall system, it is required to detect and remove these patterns and continue with further processing. In our paper we propose a new method to detect Moiré patterns by extracting the high frequencies of the using Wavelet Decomposition and training this frequency information using a multi-input deep Convolutional Neural Network (CNN). The CNN [4], [5] has paved the way in for a recognition system that closely matches the human recognition in various applications, thereby increasing productivity and efficiency. In last few years, CNN has been finding its place in largest to smallest of the tasks. Due to the availability of fast computational devices, training a machine with a good number of datasets is proving out to be better than traditional threshold based approach.

II. Related Work

Moiré pattern detection has been studied and researched over several years. Previous works in this field follow rule-based threshold approach for the detection and cannot provide an accuracy as close to the human vision. Since Moiré patterns are high-frequency noise, analyzing the image frequency domain and applying filters to remove high frequencies is explored in [1] and [2]. The drawback is, if the image has an object of interest with high frequencies, that can be filtered out as well in [1] and [2]. In [3], face spoofing is identified by analyzing the Moiré pattern in the images by filtering the Fourier image using the DOG (Difference of Gaussians). Even though the high frequency filters can detect the Moiré patterns, it fails to distinguish it from other high frequencies of interest, which can result in false alarm. Humans can well differentiate between the image of an object, taken directly from a computer screen, and the image captured of the real object. In recent years, Convolutional neural networks have been used to achieve human-level prediction accuracy in large-scale image classification. Convolutional neural networks [4], [5] , was initially published for handwritten digit recognition. It has been used for image identification, segmentation and detection in various fields [6], [7]. It is mainly composed of three types of layers: convolutional layers, pooling layers, and fully-connected layers. The convolutional layers are used to extract features by convolving image regions with multiple filters. As the layers increase, the CNN understands an image progressively. The pooling layers reduce the size of output maps from convolutional layers and thus prevent ovefitting. Multi-input CNN has been used in applications such as grading the quality of the flower [8] and analyzing the texture of a surface [9]. [10] compares the performance of Deep Learning (DL) and Recurrent Learning (RL) on images from bioimaging, medical imaging etc.

III. Our proposed approach

A new CNN model with three-inputs is proposed in this paper. The full model of our CNN architecture is depicted in

978-1-5386-9276-9/18/$31.00 ©2018 IEEE

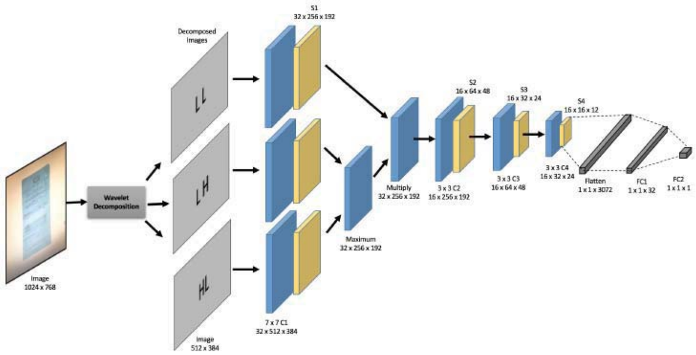

Fig. 1. Convolutional neural network

$$CWT_x^{\psi}(\tau, s) = \Psi_x^{\psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t)\psi^*\left(\frac{t-\tau}{s}\right) dt$$

Fig. 2. Continuous Wavelet Transform equation

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Fig. 3. Haar wavelet

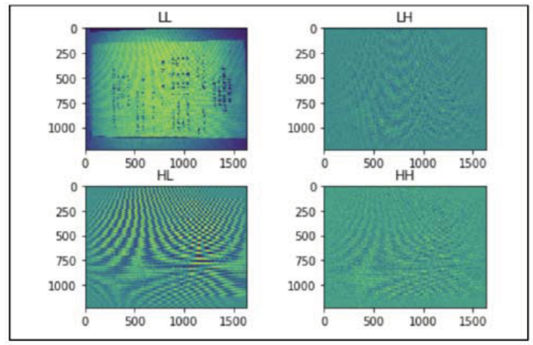

Fig. 4. Wavelet decomposition of an image with Moiré pattern

Fig. 1. All the input images are passed through a Haar Wavelet Decomposition module, to get the LL, LH and HL component of the image. This decomposition (explained in section III-B) is done prior to the training of the network and the floating-point decomposed images are normalised and then used as inputs for the training. LH and HL images are scaled between -1.0 and 1.0 and LL image is scaled between 0 and 1.

A. Convolutional Neural Network

The convolutional layer C1 filter three 512 x 384 input images with 32 kernels of size 7 x 7 with a stride of 1 pixel. The stride of pooling layer S1 is 2 pixels. Then, the convolved images of LH and HL are merged together by taking the maximum from both the images. In the next step, the convolved image of LL is merged with the Max result by multiplying both the results (as explained in section III-B). C2-C4 has 16 kernels of size 3 x 3 with a stride of 1 pixel. S2 pools the merged features with a stride of 4. The dropout is applied to the output of S4 which has been flattened. The fully connected layer FC1 has 32 neurons and FC2 has 1 neuron. The activation of the output layer is a softmax function.

B. Weighted frequency bands

Moiré patterns are a mixture of horizontal and vertical frequencies. There are multiple ways to extract high-frequency information from the images. We have used Wavelet transform for feature extraction as it is capable of providing the spatial and frequency information simultaneously, hence giving a time-frequency representation of the signal.

As seen in the Fig. 2 equation, the transformed signal is a function of two variables, ***tau*** and ***s***, the translation and scale parameters, respectively. (t) is the transforming function, and it is called the mother wavelet. In our case, we have used Haar wavelet as shown in Fig. 3 .

As shown in the Fig. 4 both horizontal and the vertical components of the pattern are captured in the LH and HL images obtained after Wavelet Decomposition. One significant observation found while analyzing the images is, (that helped to increase the accuracy of the training) when the image intensities are relatively darker, the patterns aren't visible much. As shown in the Fig. 5, the black background regions in the image (below the receipts) produce less Moiré patterns compared to the whiter regions in the image (receipt and the white background). LH, HL and HH images show the strength of the patterns produced in the LL image, which helped us to conclude that the frequency strength of the Moiré patterns is directly proportional to the intensity of the object in the scene. To summarize the spread of the Moiré pattern in the image, spatially, and to produce this effect while training the network, we used the LL band of the image (which is the downsampled original image consisting of low frequency information) and used it as weights for LH anf HL band during the training, by directly multiplying it to the convolved and combined response of the LH and HL bands, as shown in Fig. 1 (Multiply 32 x 256 x 192)

Hence, the LL component is directly multiplied (after normalization) as weights to the LH and HL component in order to produce this effect while training the network.

C. Training Dataset

We captured a total of 1633 images for training and testing the neural network. These dataset consisted positive images

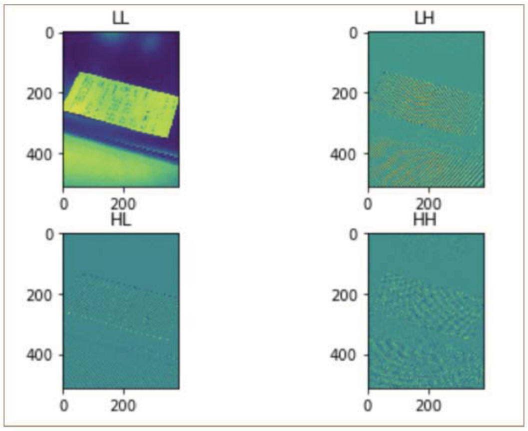

Fig. 5. Relative strengths of the Moiré pattern in dark and brighter regions

(images captured normally) and negative images (images captured from computer screen) in the ratio of 1:4. The presence of Moiré patterns in the images are hard to distinguish from the images which consist of regular and strong patterns. Fig. 6 shows an example where a receipt is captured with a strong textured background. These high frequency and strong texture background competes with the high frequencies of Moiré patterns, making it really hard to identify the patterns when using frequency threshold based approach. Our training set consists of these kind of hard to identify images. Fig. 7 shows the few diverse scenarios which were considered while capturing these images like high texture background, good or bad contrast, plain background, non-reciept images etc.. These images were captured using different mobile devices (iPhoneX, iPhone7, iPhone 7plus, Galaxy s8, Galaxy s6) in order to cover the preprocessing done by the camera software of the device manufacturers. The receipts, which were used for the dataset creation, are from multiple countries and are of different services (hotel, restaurants, taxis etc.)

D. Training CNN

The system used for training the CNN is MacBook Pro (15-inch, 2016), 2.6GHz Intel Core i7, 16GB RAM 2133 MHz LPDDR3, Intel HD Graphics 530 1536MB using Keras library in python. The training was done for 25-30 epochs, with categorical cross entropy as loss function and Adam optimizer. The entire datatest was split as 80% training set and 20% testing set for accuracy validation of the computed model while running the successive epochs. Using the number of dataset mentioned in section III-C, the time taken to run one epoch is 7-8 minutes.

IV. Verification Experiments

In our experiment, we have tested the model on diverse dataset of receipts, with and without CNN model. Table 1, shows the summary of the verification.

Fig. 6. Results without using CNN

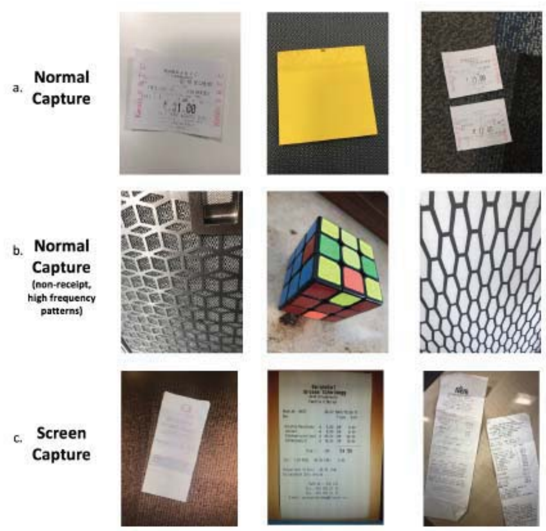

Fig. 7. Dataset: a) Normal capture of the recieipts, b) Normal capture of the non-recieipts patterns and object, c) Receipt captured from the screen with Moiré patterns

A. Wavelet Frequency Thresholding Approach

Wavelet decomposition of the image gives spatial-frequency represenation of the image content. The region of reciepts in the image is usually of low frequency (comparitively), when a normal image is captured. Hence, (based on human intelligence for detecting Moiré patterns) if the high frequency patterns are spread spatially throughout the image, then the image can be classified into a Moiré pattern. We implemented this algorithm, by passing the image through Haar wavelet for decomposition. As explained in section 3.2, LL component of the image is normalized between 0 to 1. Standard deviations for LH and HL images are calculated over a window and

TABLE I
EVALUATION ON THE DATASET

	Wavelet Frequency Thresholding Approach	*Convolutional Neural Network*
Total Images	1633	1633
Passed	1159	1608
Failed	474	25
Accuracy	0.71	0.98
Precision	0.46	0.97
Recall	0.85	0.94

then multiplied with normalized LL image to get a score for finding the Moiré pattern. The size of the window to be used is dependent on the size of the reciept in the image and multi-sized window makes the computaion of a single detection costly, but it can improve the accuracy. The drawback of this method is that this fails in images where the background has a lot of texture or any kind of patterns. Since this approach fails to distinguish between high frequency of background texture and Moiré pattern, the accuracy is 71%.

TABLE II
CONFUSION MATRIX FOR WAVELET FREQUENCY THRESHOLDING APPROACH

	Predicted YES	*Predicted NO*
Actual YES	346	61
Actual NO	413	813

B. Results of CNN

The logical derivations of the Wavelet Frequency Thresholding Approach is used to train the network. As shown in Fig. 1, CNN model is trained with these images with a ratio of roughly 1:4 positive to negative cases in python, for around 25-30 epochs. For training, 80% of the dataset is taken and rest of the data is used for testing. Table 2, shows the training accuracy of 99.3109% obtained on 1306 images and Table 3, shows the testing accuracy of 95.1070% obtained on 327 images.

TABLE III
CONFUSION MATRIX - CNN (TRAINING)

	Predicted YES	*Predicted NO*
Actual YES	237	5
Actual NO	4	1060

TABLE IV
CONFUSION MATRIX - CNN (TESTING)

	Predicted YES	*Predicted NO*
Actual YES	49	12
Actual NO	4	262

The trained model was converted to CoreML model and the images were tested in iOS (iPhone 7) for around 705 images having both screen capture and normal images. This resulted in

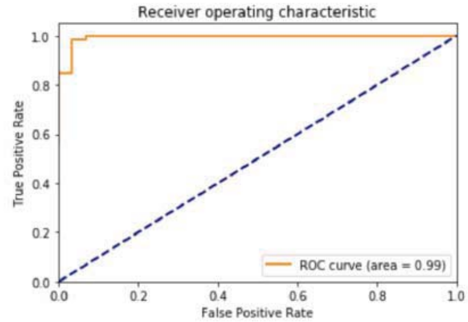

Fig. 8. ROC

an accuracy of 96.17%, as shown in Fig. 9. The time taken to process one image of size 1024 x 768 is 110.4ms. For wavelet decomposition, vDSP is used and CoreML for CNN model. This can be further improved if Metal is used for processing Wavelet Decomposition.

V. FUTURE SCOPE

In this paper, CNN is trained with only 1-level wavelet decomposition. In the further scope, the effect on the accuracy of the network for multi-level wavelet decomposition can be explored. Due to the lack of standardized dataset for Moiré pattern detection, 1633 images were manually captured. The CNN model can be trained with more and tougher examples and a deeper network. In our current experiment, we have kept the resolution of the image as 1024 x 768, in order to preserve the high frequency information of the Moiré pattern as downsampling the image results in frequency information loss. We plan to analyze the effect of decreasing the image resolution on the accuracy of the detection. Also, the HH band is ignored currently, which can also be used while training.

VI. CONCLUSIONS

This paper provides a method using Wavelet decomposition and Convolutional neural network to detect Moiré patterns in the images. The strength of the proposed CNN model is, it uses the LL intensity image (from the Wavelet decomposition) as a weight parameter for the Moiré pattern, thereby approximating the spatial spread of the Moiré pattern in the image. Usage of CNN model performs better than frequency thresholding approach as the model is trained considering diverse scenarios and it is able to distinguish between the high frequency of background texture and the Moiré pattern. It also detects the screen capture on the screen, images of which were not in the training data. Though this approach is trained specifically with receipt images, it can be extended to different scenarios where Moiré patterns from screen capture need to be detected.

REFERENCES

[1] Seyede Mahya Hazavei, H.R.S.: A new method for removing the moire' pattern from images. Phil. Trans. Roy. Soc. London 247 (1955) 529-551
[2] andDean Chapman, Z.W.J.W.H.N.S.W.: A median-gaussian

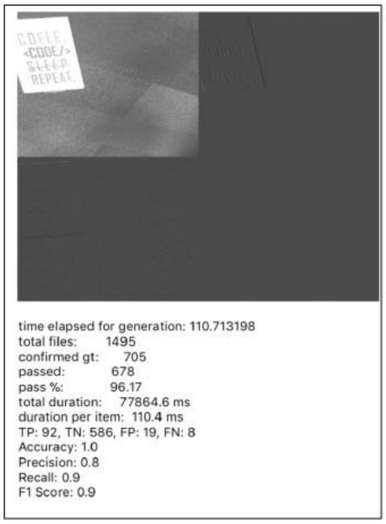

Fig. 9. Final testing on iPhone7

ltering framework for moir pattern noise removal from x-ray microscopy image. IEEE Transactions on Information Forensics and Security, Micron. 43 (2012) 170-176

[3] Ricardo L. de Queiroz, D.C.G.: Face-spoofing 2d-detection based on moire-pattern analysis. IEEE Transactions on Information Forensics and Security 10 (2015)

[4] Bengio, Y.L.L.B.Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86 (1998) 2278-2323

[5] LeCun, B. Boser, J.S.D.e.a.: Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems (1990) 396-404

[6] Monrocq, R.V.C., Cun, Y.L.: An original approach for the localization of objects in images. International Conference on Artificial Neural Networks (1993) 26-30

[7] Garcia, C., Delakis, M.: Convolutional face finder: A neural architecture for fast and robust face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2004) 1408-1423

[8] and FangZhao, Y.G.: Multi-input convolutional neural network for flower grading. Journal of Electrical and Computer Engineering 2017 (2017) 8

[9] Shin Fujieda, Kohei Takayama, .H.: Wavelet convolutional neural networks for texture classification. Cornell University (2017)

[10] M. Mahmud, M. S. Kaiser, A. Hussain and S. Vassanelli, ”Applications of Deep Learning and Reinforcement Learning to Biological Data,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2063-2079, June 2018. doi: 10.1109/TNNLS.2018.2790388