Deep Learning-based Denoising of TEMPEST Images for Efficient Optical Character Recognition

Juan Galvis, Santiago Morales-Aguilar, Chaouki Kasmi, Felix Vega
Directed Energy Research Centre
Technology Innovation Institute
Abu Dhabi, United Arab Emirates
Email: juan.galvis@derc.tii.ae

***Abstract*—The present work shows the application of deep learning models to the denoising of video frames retrieved from electromagnetic emanations from remote video interfaces. It has been demonstrated that the cables of video interfaces like VGA or HDMI, produce unintended emanations, and that these emanations can be received and processed to reconstruct the video frames displayed on the external monitor. However, the reconstructed frames are noisy, making it difficult to recover any useful information. By applying deep learning models to denoise, deblur, and interpret the images, information can be interpreted.**

***Index Terms*—TEMPEST, information leakage, Deep Learning, Convolutional Neural Networks, SDR.**

I. Introduction

FEW research laboratories have known since the early 1960s that electronic devices such as computers produce electromagnetic radiation and that these emanations unintentionally cause information security vulnerabilities. The U.S. National Security Agency developed a series of specifications under the code name TEMPEST, with the requirements for avoiding electronic equipment to transmit unintended emanations. These phenomena were known to the public in 1985 when Wim van Eck [1] published a technical analysis with details about how CRTs produced emanations that could be used to recover the displayed video. In 2002, Markus Kuhn [2] demonstrated these phenomena to happen on more modern technologies like LCD monitors. Nowadays, video frames from external monitors can be retrieved remotely with the help of low budget software-defined radios (SDRs) [3], but the recovered video frames are noisy.

Image denoising techniques have been extensively studied. However, processing video frames reconstructed from unintentional emanations with non-learned methods such as the one proposed in [4] and [5], remains a difficult task, mainly due to the high degree of deterioration in the recovered information and the unknown noise distributions.

Since the emergence of AlexNet [6] in 2012 and its performance in the ImageNet Image classification competition, Convolutional Neural Networks (CNNs) and Deep Learning (DL) have gained popularity for all sorts of computer vision tasks, providing solutions to problems where more classical approaches fail. The tasks of image denoising and deblurring is no exception. Model architectures like the feed-forward denoising convolutional neural networks (DnCNNs) [7] have proved to be able to remove additive white Gaussian noise from images without prior knowledge of the noise level. Moreover, architectures like Mask R-CNN have already been used as denoisers on intercepted video frames, recovering more than 57% of a leaked information for a wide range of interception distances, however with inference times of around 4.0s which is still not enough for real-time processing [8].

The present work shows how, by applying state-of-the-art deep learning models, the noisy images can be processed, so that optical character recognition (OCR) can be used to detect text automatically and efficiently.

II. Methods

In general our method involves four steps. First, the compromising signals are intercepted and reconstructed into video frames. Then, sharp-noisy image pairs are taken and aligned. Next, the CNN is trained using this data. Finally, the trained model processes new intercepted frames, obtaining denoised images from which text is extracted using the Tesseract optical character recognition engine [9].

A. Capturing Compromising Emanations and aligning frames

Using the method proposed by M.Marinov [3] and the well-known TEMPEST-SDR tool implemented during his thesis, video frames are retrieved from the emanations of an HDMI cable as shown in Figure 1.

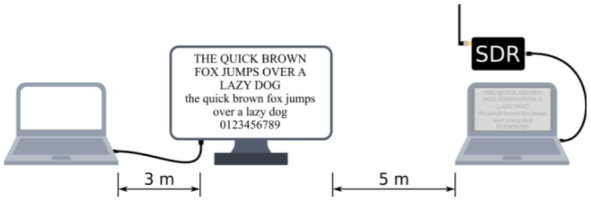

Fig. 1. Setup used for data capture

B. Aligning image pairs

The captured images present several distortions, apart from the noise, they are skewed, displaced and resized respect to the original image. In order to be able to compare and train our

models, the correction of these distortions is required. Thus, a method for the image alignment is proposed which uses a calibration pattern. This allows for finding common points between the original and the intercepted images as depicted in the Figure 2.

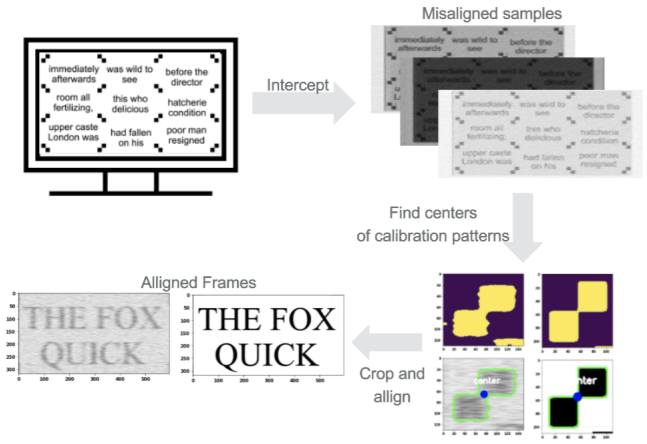

Fig. 2. Image alignment processing for training dataset

Through this process a set of 8079 image pairs is obtained.

C. Model Training

In the search for a close to real time implementation, two different models have been implemented, which were fully trained on our set of image-pairs.

1) Deep Deblur: The first model is a Deep Deblur CNN based on the work by J. Mei et al [10], developed for scanned text deblurring. The authors introduced a novel network structure called Sequential Highway Connection (SHC) motivated by the Residual Neural Network [11] that guarantees superior convergence during training.

This model emerges from the assumption that the image blurring process can be modeled as the convolution operation between a sharp image and a blur kernel, merged with noise. Then, the model can learn blur and noise statistical distribution from data and perform an inverse transformation, predicting the sharp image. In essence, this model carries out a pixel level regression.

2) FCHardNet: The second model is a CNN with an encoder-decoder structure that uses HardNet as backbone [12]. It is based on the work by P.Chao on efficient image segmentation [13]. In essence, this model performs a pixel level classification.

This network consits of two parts, first an encoder which down-samples the input image through a series of convolutional layers, until reaching a reduced representation of image features. Then, the second part, known as decoder upsamples the data again until reaching a representation that resembles the denoised image. For this model, residual connections are also key in order to allow superior convergence during training and performance during inference.

The training results for both models can be seen in the loss plots shown in Figure 3.

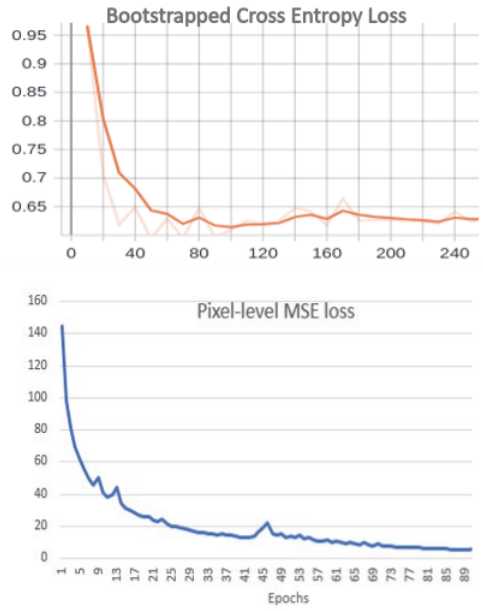

Fig. 3. Bootstraped Cross-Entropy Loss for FCHardNet (Up) - MSE Loss for DeepDeblur (Down).

D. Optical Character Recognition

After the original reconstructed video frames have been improved with mentioned deep learning techniques, the text is ready to be detected by an OCR. The OCR engine chosen for this task is Tesseract, because of its good reviews (probably the best open-source OCR to date), its flexibility and capability to be tuned, and because it is open source. It is worth mentioning that although has high accuracy on classical images, Tesseract could not succeed in text recognition of the unprocessed video frames recovered from compromising emanations. This could be due to the fact that it uses a binarization algorithm called Otsu Thresholding [14] which is not suited for highly noisy images.

III. Results

The metric used to measure the performance of our denoising CNNs is the PSNR (Peak Signal to Noise Ratio). The results obtained for some sample images can be seen in Figure 4.

Fig. 4. (Left) PSNR for DeepDeblur. (Right) PSNR for FCHardNet.

Finally, after denoising the images, these are passed to the Tesseract in order to extract text as shown in Figure 5.

Besides the PSNR and the text extraction capabilities, another key performance indicator of these algorithms is training

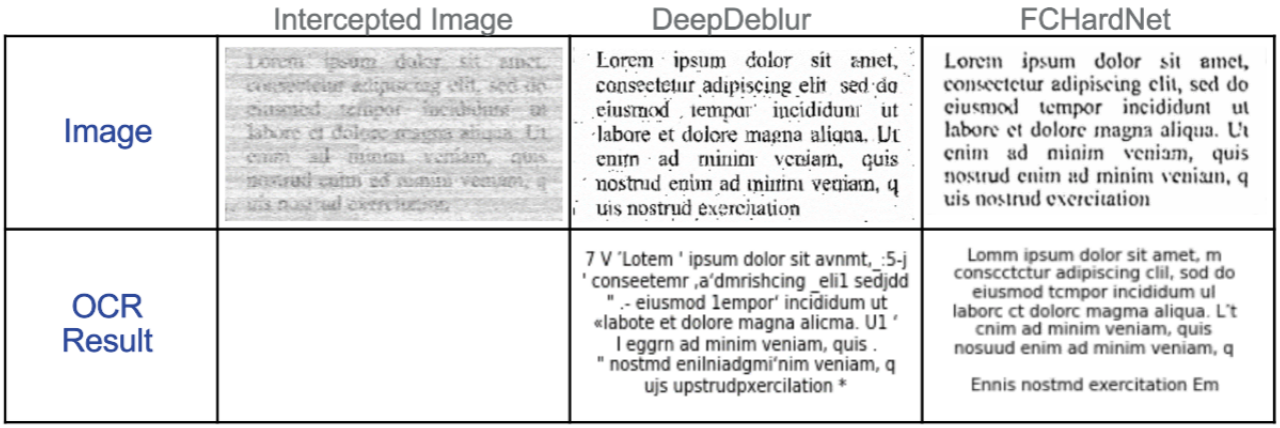

Fig. 5. Text Extracted from original intercepted image (left), image processed with DeepDeblur (center) and image processed with FCHardNet (right). OCR result, executed by Tesseract, is shown in the second row.

TABLE I
TIME PERFORMANCE

Model	Training	Denoising	Text Extraction
DeepDeblur	$> 24h$	$2.52s$	$1.2s$
FCHarDNet	$1h$	$0.17s$	$1.2s$

and inference time, as shown on the Table I the fastest model is FCHardNet, which with 0.17s is close to being real time capable, something very important in the context of continuous listening of EM emanations.

The work presented here was implemented using Keras and Pytorch. The models were trained using a Nvidia GeForce GTX 1650 GPU and Intel Core i7 9th Gen. processor.

IV. CONCLUSION

In the presented work, a dataset of 8072 sharp/noisy image pairs were captured and aligned, this allowed for the training of two denoising CNN models. After processing noisy images using these models, an improvement in the average PSNR has been achieved. Additionally, text could be extracted, something that was not possible for unprocessed images. FCHardNet showed the best performance in terms of training and processing time and presents a good solution for real time information recovery from unintended electromagnetic emanations produced by video interfaces. It can be observed that the efficiency of the optical character recognition has been substantially improved with a reduced computation time.

The next steps of this research will be dedicated to the definition of a quantification metric that could be used in order to automatize the evaluation of compromising emanations from a given computer screen. This will support the real time hardening methodology currently under development by the Electromagnetic Compatibility team at the DERC/TII. Finally, the automatic language detection, the context detection, and the auto-correction of the recovered text will be studied using state-of-art algorithms.

REFERENCES

[1] W. Van Eck, "Electromagnetic radiation from video display units: An eavesdropping risk?" *Computers and Security*, vol. 4, no. 4, pp. 269–286, 1985.

[2] M. G. Kuhn, "Compromising emanations: eavesdropping risks of computer displays," University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-577, dec 2003. [Online]. Available: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-577.pdf

[3] M. Marinov, "Remote video eavesdropping using a software-defined radio platform," M.S. thesis, University of Cambridge, jun 2014. [Online]. Available: https://github.com/martinmarinov/TempestSDR

[4] S. Morales-Aguilar, C. Kasmi, M. Meriac, F. Vega, and F. Alyafei, "Digital images preprocessing for optical character recognition in video frames reconstructed from compromising electromagnetic emanations from video cables," in *XXXIII General Assembly and Scientific Symposium (GASS) of the International Union of Radio Science (Union Radio Scientifique Internationale-URSI)*. URSI, 2020.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[8] F. Lemarchand, C. Marlin, F. Montreuil, E. Nogues, and M. Pelcat, "Electro-magnetic side-channel attack through learned denoising and classification," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2882–2886.

[9] R. Smith, "An overview of the tesseract ocr engine," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2. IEEE, 2007, pp. 629–633.

[10] J. Mei, Z. Wu, X. Chen, Y. Qiao, H. Ding, and X. Jiang, "Deepdeblur: text image recovery from blur to sharp," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18 869–18 885, 2019.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[12] P. Chao, C.-Y. Kao, Y.-S. Ruan, C.-H. Huang, and Y.-L. Lin, "Hardnet: A low memory traffic network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3552–3561.

[13] P. Chao, "FCHarDNet," May 2020. [Online]. Available: https://github.com/PingoLH/FCHarDNet

[14] X. Yang, X. Shen, J. Long, and H. Chen, "An improved median-based otsu image thresholding algorithm," *AASRI Procedia*, vol. 3, pp. 468–473, 2012. [Online]. Available: https://doi.org/10.1016/j.aasri.2012.11.074