Smart Grids False Data Injection Identification: a Deep Learning Approach

Helton do Nascimento Alves
Instituto Federal do Maranhão, Electrical Engineering Dep.
São Luis, Brazil
helton@ifma.edu.br
Arturo S. Bretas
University of Florida, Electrical Engineering Department,
Gainesville, FL, USA
arturo@ece.ufl.edu
Newton G. Bretas
University of São Paulo, Electrical Engineering Dep.
São Paulo, Brazil
ngbretas@sc.usp.br
Ben-Hur Matthews
Instituto Federal do Maranhão, Electrical Engineering
São Luis, Brazil
benhurmatthews@hotmail.com

***Abstract*— recently a new class of security problem in a power system state estimation was defined by false data injection (cyber-attack). The deliberate injection by an adversary would not be expected to follow the same patterns as random bad data. False data injection can be launched in measurements set, topology and parameters network data. Identify clearly its injection point is very important to provide a suitable correction of the output states. In this paper is presented an identification strategy for false data injection in power system state estimation input data based on a deep learning. Evaluation of the presented solution is done through Monte Carlo simulation considering different test scenarios in the IEEE 14-bus test system. The results confirm that the proposed algorithm is a potential tool to identify accurately the false data injection point in power system state estimation.**

***Index Terms*— False data injection identification, power system state estimation, deep learning.**

I. Introduction

Data communication through the internet has lot of advantages, but inherently it increases the risk of exposure to cyber threats. . Cyber-attacks are a reality in smart grids. Most recently a cyber-attack in a power grid in Ukraine left about 225,000 customers without electricity [1]. Therefore, cyber security becomes a crucial resource to ensure the integrity and resilience of smart grid operations. In [2] is shown that a malicious attack in measurements set data is able to bypass classical bad data detection algorithms. In [3], similar result is obtained to malicious attacks in network data. Based on these considerations many studies have been made to analyze the power system vulnerabilities and countermeasures to cyber-attacks [2]-[7]. Power systems state estimation (SE) has three inputs data type: measurements, power system topology and transmission line parameters data. The false data injection (FDI) can be launched in any one of them. Naturally, the more information one has about the cyber-attack the more efficient the countermeasures will be. In this context, it is essential to identify clearly the FDI point for analysis and correction. In the literature, this theme is usually not highlighted and it is considered that the state estimator input data type where the cyber-attack occurred is known in advance. This assumption is not realistic because usually this information is not available. In the classical bad data detection, topological errors identification is usually based on the generalized state estimator [8]. The correction is performed modifying the status of the closest breaker/switch (on/off) from the branch identified with topological error. Therefore, the status of breaker/switch is considered as the source of the possible power system topological error. On the other hand, when a cyber-attack is launched on power system topology, an exclusion/insertion/transference of a branch occurs directly on the line data of the branch instead on breaker/switches status. In this case, only the changing of status on breaker/switch will not correct the power system topology (Fig. 1).

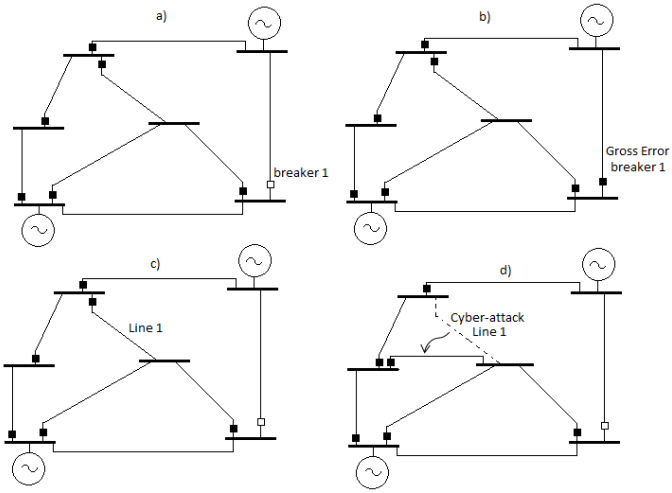

Fig. 1. (a) Original network data (b) network data with gross error (c) Original network data (d) network data under cyber-attack.

A classical technique to estimate the states variables is the Weighted Least Square formulation (WLS), which looks for the state vector that minimizes the objective function based on the residual vector. The chi-square and normalized residual tests are the common post estimation procedures used for detection and identification, respectively, of gross error in SE [9]–[12]. In these studies, the measurement residual is treated as the measurement error, but they are completely different quantities [13]-[14]. The measurement residuals pertain to the residual sub-space, with number-of-measurements minus

978-1-5386-8218-0/19/$31.00 © 2019 IEEE

Authorized licensed use limited to: University of Florida. Downloaded on May 18,2020 at 19:16:56 UTC from IEEE Xplore. Restrictions apply.

number-of-state-variables degrees of freedom, that is, a correlated space. On the other hand, the measurement errors pertain to the measurement sub-spaces, then with number-of-measurements degrees of freedom, that is, a not correlated space. More recently, using topological and geometrical approaches, [7], [13]–[14] proposed solutions to compose the measurement error and replace the measurement residual in bad data detection. The simulations reported in these papers have shown many situations where the classical state estimation fails in the gross error detection and identification test, while the composed measurement error proposal provides an accurately response.

The composed measurements error in connection with a deep learning approach are used in this paper to identify accurately the false data injection point (Fig 2). The main contributions of this work are: (i) the raw data is seen as a snapshot of a particular output of the power system state estimator and treated as an image by deep learning approach; (ii) this paper focuses on the application of Convolutional Neural Networks (CNN) to image classification tasks, that has demonstrated to be a powerful algorithm in the recent state-of-the-art deep learning systems [15]; (iii) many different methods have been used for intrusion detection, but the proposed algorithm is one of the few works that studies specifically the cyber-attack injection point identification on state estimator input data; (iv) composed measurements error has shown to be a better parameter to analyze FDI than measurements residue the. This new paradigm increases accuracy of the proposed method in cases where the established method has failed.

The performance of the proposed algorithm is evaluated for many different measurements plans using Monte Carlo Simulation (MCS) procedure ensuring the experimental accuracy and proficiency.

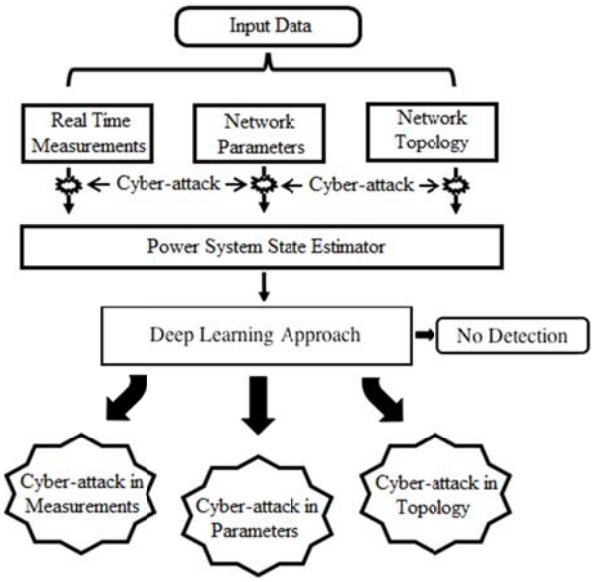

Fig. 2. State estimator under a cyber-attack

II. Innovation Concept

In classical weighted least squared (WLS) estimator, considering the power system modeled as a set of non-linear equations with b buses and m measurements, there are N=2*b-1 unknown state variables to be estimated. The cost function $J(x)$ is minimized to find the best estimative for the state vector x (N,1):

$$z = h(x) + e \tag{1}$$

$$J(x) = [z - h(x)]^T W [z - h(x)] \tag{2}$$

where $z \in \mathbb{R}^N$ is the measurement vector, $h: \mathbb{R}^N \to \mathbb{R}^m$ is a continuous nonlinear differentiable function, e is the error vector assumed with zero mean and Gaussian probability distribution and W is the weighting matrix (m,m).

The solution of the afore-mentioned minimization problem is obtained through the linearization of (2). At a certain operation point, yields,

$$\Delta z = H \Delta x + e \tag{3}$$

being H the matrix of first derivatives of the nonlinear functions of vector $h(x)$, known as the Jacobian, calculated at the point represented by the vector of estimated state variables.

Therefore, the solution can be obtained by:

$$\Delta x^k = [H^T W H]^{-1} H^T W \Delta z^k \tag{4}$$

The iterative process starts from an initial state vector and, at each iteration k; the corrections in the state variables Δx^k are obtained using (4). The vector of state variables update until a stopping criterion is satisfied.

This work is based on Innovation concept introduced by [7], [13]-[14]. This method is just briefly brought the concept and the formulation here. If the system described by (1) and (2) is observable, then the vector space of measurements R^m can be decomposed into a direct sum of two vector subspaces,

$$R^m = R(H) \oplus R(H)^{\perp} \tag{5}$$

in which the range space of H is an N-dimensional vector subspace into R^m and $R(H)^{\perp}$ is its orthogonal complement.

In the linear state estimation formulation, the solution of (3) can be understood as a projection of the measurement vector mismatch Δz onto $R(H)$. In [17] is defined the linear operator P that performs this projection, as follows:

$$P = H[H^T W H]^{-1} H^T W \, . \tag{6}$$

Based on (5) and (6), the linear formulation of the state estimation can be used to decompose the measurement error vector e into two parts: the detectable and the undetectable components. The detectable component is the residual measurement vector and undetectable component is orthogonal the detectable component and calculated as follow:

$$e_D = (I - P)e. \tag{7}$$

$$e_U = Pe. \tag{8}$$

$$\|e_i\|_W^2 = \|e_{Di}\|_W^2 + \|e_{Ui}\|_W^2 \tag{9}$$

In order to find the undetectable component and compose the measurement's total error for i[th] measurement, it is used II:

$$II_i = \frac{\|e_{Di}\|_W}{\|e_{Ui}\|_W} \tag{10}$$

Consequently, even when those measurements have gross errors, their residuals will be relatively small. Based on (9) and

Authorized licensed use limited to: University of Florida. Downloaded on May 18,2020 at 19:16:56 UTC from IEEE Xplore. Restrictions apply.

(10) is possible to estimate the composed measurement error of the measurement *i* based on its standard deviation, as follow:

$$\|e_i\|_W^2 = \|e_{Di}\|_W^2 + \left\|\frac{e_{Di}}{II_i}\right\|_W^2$$

$$\|e_i\|_W = CME_i \ = \|e_{Di}\|_W \sqrt{1+\frac{1}{II_i{}^2}} = r_i\sqrt{1+\frac{1}{II_i{}^2}}$$

$$CME_i^N = \frac{CME_i}{\sigma_i} = \frac{r_i}{\sigma_i}\sqrt{1+\frac{1}{II_i{}^2}} \quad (11)$$

where r_i is the residue of measurement *i* and σ_i is the standard deviation of the measurement *i*.

In [7] is shown that Cyber-attack detection can be made through a Chi-square (χ^2) Hypothesis Testing (HT) applied to the composed measurement error, where bad data will be suspect if:

$$[CME^N]^T[CME^N] \geq \chi^2_{m,p} \quad (12)$$

where *p* is the detection confidence probability and *m* are the degrees of freedom.

In [14] is considered that in the detection stage it does not matter how reliable a measurement is because it is assumed that all them may contain errors. Therefore, it is attributed weights to the measurements as described by (13):

$$W_{ii} = 1/(0.1z_i)^2 \quad (13)$$

III. Proposed Algorithm

The proposed algorithm is shown in the following. More details are shown in next subsections.

(1) Collect input data: real time measurements, transmission line parameters and power system topology data;

(2) Perform WLS to calculate CME^N;

(3) Build graph bars based on CME^N vector;

(4) Classify the incoming set using CNN approach in,

a. no cyber-attack detected;

b. cyber-attack detected in real-time measurements data;

c. cyber-attack detected in line parameters data;

d. cyber-attack detected in power system topology data.

(5) Finish the algorithm.

A. Data Model

The measurement plan (MP) used is generated by a set of measurements formed by two groups: default set of measurements (DSM) defined as measurements are already present in the substations and probabilistic set (PSM), defined as measurements chosen by Monte Carlo simulation (MCS) to complete the measurement plan. The Monte Carlo sampling is done to determine randomly various measurement plans.

Every MP has an associated random noise, so that they vary from ±2σ of its original values, so, not characterizing measurements with gross errors. The data base generated for simulations is divided in four groups with the same size divided in: (a) set of samples without data anomaly; (b) set of samples with an associated random multiple measurement cyber-attack ranging from 3σ to 5σ applied in up to 10 measurements chosen randomly. Critical measurements or critical sets of measurements are not considered; (c) set of samples with an associated random parameter cyber-attack that vary the magnitudes of the transmission line series parameters from 20% to 80% applied in up to 3 transmission lines (chosen randomly) and (d) set of samples considering an associated random exclusion/transference topological cyber-attack, where a line was excluded or transferred.

B. Input Parameters

The CME^N vector is represented for a bar graph where the height of bar is proportional to their values. Considering that the CME^N values can have a large range, the logarithmic scales in Napierian base is used. CME^N vector values are always positives (11), but in Napierian base, they can assume negative values. In order to easy the graph representation, only the absolute value is considered and the real positive values are plotted in blue and negative ones in red. In Figs. 3-6 are shown some bar graphs generated from a measurement plan with 95 measurements. In Fig 3, the CME^N vector is from a set of measurements without FDI, in Fig 4, from a set of measurements with FDI in measurements, in Fig 5, from a set of measurements with FDI in parameters and, in Fig 6, from a set of measurements with FDI in topology. These results show that the bar graphs present different trends for each input data under cyber-attack, showing potential to be used as parameters to identify the state estimator input data corrupted.

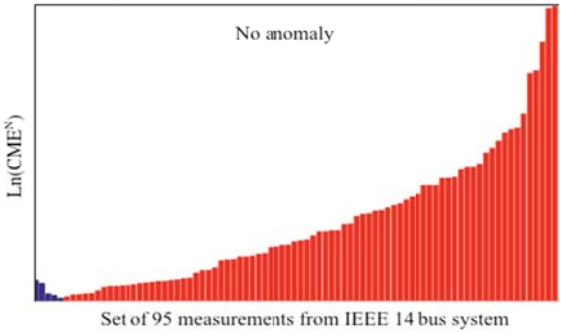

Fig. 3. Graph of CME^N vector without anomaly.

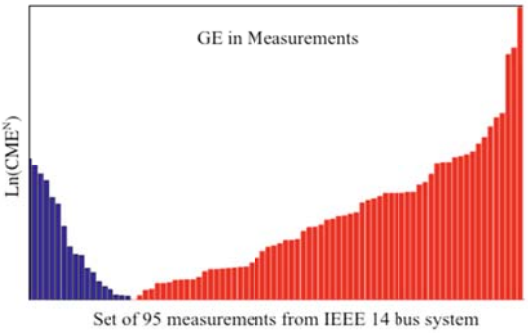

Fig. 4. Graph of CME^N vector with GE in measurements.

Authorized licensed use limited to: University of Florida. Downloaded on May 18,2020 at 19:16:56 UTC from IEEE Xplore. Restrictions apply.

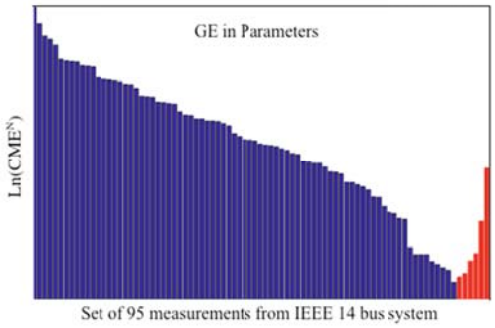

Fig. 5. Graph of CME^N vector with GE in parameters.

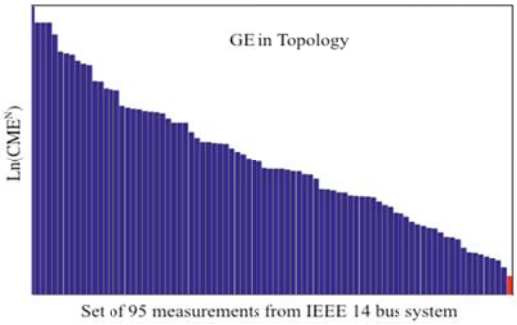

Fig. 6. Graph of CME^N vector with GE in topology.

C. Deep Convolutional Neural Networks

In deep convolutional neural networks (CNN) architecture, an image is input directly to the network, which allows encoding certain properties into the architecture [15]-[16]. The image is represented by its volume: Width x Height resolution and depth (3 for RGB image or 1 for Black and White one). The basic idea of the CNN was inspired by a concept in biology called the receptive field [17]. Receptive fields are a feature of the animal visual cortex [18]. They act as detectors that are sensitive to certain types of stimulus, for example, edges. They are found across the visual field and overlap each other. This biological function can be approximated in computers using the convolution operation [19]. In image processing, images can be filtered using convolution to produce different visible effects. CNN architectures come in several variations; however, in general, they are followed by several stages of convolution and pooling (or subsampling) layers, which are grouped into modules.

The convolutional layers serve as feature extractors, and thus they learn the feature representations of their input images. Usually, in CNN classifier, the rectified linear units (ReLUs) is adopted as an activation function [15]. Inputs are convolved with the learned weights in order to compute a new feature map. The size of the Feature Map is controlled by three parameters: number and size (width and height) of filters, stride and zero-padding. Stride is the number of pixels by which the filter is slide over the input matrix.

The pooling layers progressively reduce the spatial resolution of the feature maps and thus achieve spatial invariance to input distortions and translations. It operates independently on every depth slice of the input and resizes it spatially, using the MAX operation (the most common form). Formally, max pooling selects the largest element within a single depth slice produced by the convolution layer.

In CNN classifier, the output of the last full-connection layer is fed to a *C*-way softmax function which produces a distribution *C* class labels. The Fully Connected layer is a multi-layer perceptron that uses a softmax activation function in the output layer. Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks. The output from the convolutional and pooling layers represent high-level features of the input image. The purpose of the Fully Connected layer is to use these features for classifying the input image into various classes based on the training dataset.

The configuration parameters of CNN have a great influence on the accuracy of classification. In this study, the configuration parameters of CNN were analyzed through tens of simulations and the most appropriate configuration founded in relation to the algorithm performance is depicted below.

- INPUT: 25x34 RGB image.
- CONV layer: 20 10x19 filters (stride =1 and padding=0), obtained an output matrix [16x16x20].
- The elementwise activation will be applied with RELU layer function.
- Max-Pooling layer: 4x4 (stride =1 and padding=0), obtained an output matrix [13x13x20].
- CONV layer: 20 5x5 filters (stride =1 and padding=0), obtained an output matrix [9x9x20].
- Max-Pooling layer: 3x3 (stride =1 and padding=0), obtained an output matrix [7x7x20].
- Full-connection layer: multi-layer perceptron using a softmax function for 4 class labels.

In terms of the learning rate, it was observed that the epoch numbers 20 was appropriated, considering that increasing it more than 20, the accuracy and loss remain approximately the same. The training speed was around 44 images/s for a Dell laptop test machine with an Intel i7 CPU clocked at 2.4 GHz, 8 GB of RAM, and a GPU NVIDIA GeForce 920M.

IV. Numeric Tests

The IEEE-14 bus system is used for numeric tests. The means aggregated from MCS obtained by proposed algorithm is reported in Table I. Maximum standard deviation obtained for MCS is 2.34%. The network is trained considering measurement plans with 95 measurements (DSM: active and reactive power injection and voltage magnitude in bus 1; PSM: 92 measurements defined by MCS considering a normal distribution function of active and reactive power injection and power flow). 20000 samples are generated to test the trained network (5000 samples for each FDI class). Possible losses of meters for malfunction or maintenance is considered. It is considered up to 5 meters out of service.

In order to evaluate the proposed algorithm under cyber-attack just after a system reconfiguration is performed but before updating the trained network, the new topology shown in Table II is considered. The results show that the proposed algorithm continues accurately detecting and identifying the false data injection point.

Authorized licensed use limited to: University of Florida. Downloaded on May 18,2020 at 19:16:56 UTC from IEEE Xplore. Restrictions apply.

The proposed algorithm is also trained and tested considering as parameter the normalized residual index (NRI), in order to compare with the parameter input proposed in this work. The detection accuracy obtained using CME^N is 99.93% and 93.7% using NRI. The results show that the CME^N reflects the features of the false data injection in a more appropriate way than NRI.

TABLE I. ACCURACY CONSIDERING POSSIBLE LOSSES OF METERS.

Set of available measur.	Identification of the input data status (%)			
	No anomaly	*Cyber_attack measurement*	*Cyber_attack parameters*	*Cyber_attack topology*
95	100.00	99.83	99.97	99.86
93	100.00	99.72	99.76	99.76
90	100.00	99.63	99.56	99.69

TABLE II. ACCURACY CONSIDERING RECONFIGURATION.

Reconfig.	Identification of the input data status (%)			
	No anomaly	*Cyber_attack measurement*	*Cyber_attack parameters*	*Cyber_attack topology*
No one	100.00	99.83	99.97	99.86
Line transf. 13-14→11-14	100.00	99.37	99.76	99.67
Elim. line 10-14	100.0	99.45	99.37	96.42

V. CONCLUSIONS

Based on the tests presented, the proposed algorithm got close to 100% of correct identification of the state estimator input data base corrupted. Monte Carlo simulation application ensures the robustness of the method and its training time is small enough to be used in real time applications. The graph bars from data set obtained through the innovation index presented an excellent adherence in the solution of the proposed problem. Comparative tests demonstrate the increased accuracy of the proposed method in cases where the established method has failed. Exploring others machine learning algorithms and to propose a bad data correction remains a promising subject for future investigations.

REFERENCES

[1] E. Perez, "U.S. Investigators Find Proof of Cyberattack on Ukraine Power Grid", CNN, 2016, Available on: http://www.cnn.com/2016/02/03/politics/ cyberattack-ukraine-power-grid/.

[2] Y. Liu, M. K. Reiter and P. Ning, "False data injection attacks against state estimation in electric power grids," *in Proc. 2009 16th ACM Conference on Computer and Communications Security*.

[3] J. Kim and L. Tong, "On Topology Attack of a Smart Grid: Undetectable Attacks and Countermeasures," *IEEE Journal on Selected Areas in Communications, 31(7)*, pp. 1294 – 1305, July 2013.

[4] L. Liu, M. Esmalifalak, Q. Ding, V. Emesih and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization", *IEEE Trans. Smart Grid 5(2)*, pp 612–621, 2014.

[5] Y. Chakhchoukh and H. Ishii, "Coordinated cyber-attacks on the measurement function in hybrid state estimation", *IEEE Trans. Power Syst. 30(5)*, pp 2487–2497, 2015.

[6] A. Ashok, M. Govindarasu and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation", IEEE Trans. Smart Grid PP(99), pp 1–11, 2016.

[7] A.S. Bretas, N.G. Bretas, B. Carvalho, E. Baeyens and P. P. Khargonekar,"Smart grids cyber-physical security as a malicious data attack: An innovation approach", *IEEE Trans. Power Syst. 149*, pp 210-219, August 2017.

[8] E. Lourenco, E. Coelho and B. Pal, "Topology error and bad data processing in generalized state estimation," *IEEE Trans. Power Syst., 30(6)*, pp. 3190–3200, 2015.

[9] M Cheniae, L Mili and P Rousseau, "Identification of multiple interacting bad data via power system decomposition", *IEEE Trans Power Syst,11(3)* pp1555-63, 1996.

[10] KA Clements and PW Davis, "Multiple bad data detectability and identifiably: a geometric approach", *IEEE Trans Power Deliv 1(3)*, pp 355-60, 1986.

[11] J Chen and A Abur, "Placement of PMUs to enable bad data detection in State estimation", *IEEE Trans Power Syst 21(4)*, pp1608-15, 2006.

[12] B. Carvalho and N. Bretas, "Analysis of the Largest Normalized Residual Test Robustness for Measurements Gross Errors Processing in the WLS State Estimator", *Journal of Systemics, Cybernetics and Informatics, 11(7)*, pp1-6, 2013.

[13] N.G. Bretas, A.S. Bretas and S.A. Piereti, "Innovation concept for measurement gross error detection and identification in power system state estimation", *IET Gener. Transm. Distrib. 5(6)*, pp 603–608, 2011.

[14] N.G. Bretas and A.S. Bretas, "A two steps procedure in state estimation gross error detection, identification, and correction", *Int. J. Electr. Power Energy Syst. 73 (December)* pp 484–490, 2015.

[15] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," Neural Computation 29 (September, 2017) 2352–2449.

[16] 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.

[17] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," Neural networks 1, 2 (1988), 119–130.

[18] D. H. Hubel, and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," The Journal of Physiology 195, 1 (1968), 215–243.

[19] D. Marr,, and E. Hildreth,,"Theory of edge detection," Proceedings of the Royal Society of London B: Biological Sciences 207, 1167 (1980), 187–217.

Authorized licensed use limited to: University of Florida. Downloaded on May 18,2020 at 19:16:56 UTC from IEEE Xplore. Restrictions apply.