SPECIAL SECTION ON RECENT ADVANCES IN COMPUTATIONAL INTELLIGENCE PARADIGMS FOR SECURITY AND PRIVACY FOR FOG AND MOBILE EDGE COMPUTING

Received December 4, 2017, accepted January 8, 2018, date of publication February 1, 2018, date of current version May 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2800758

Service Orchestration of Optimizing Continuous Features in Industrial Surveillance Using Big Data Based Fog-Enabled Internet of Things

SADIA DIN[1], ANAND PAUL[1], (Senior Member, IEEE), AWAIS AHMAD[2], B. B. GUPTA[3], AND SEUNGMIN RHO[4]

[1]School of Computer Science and Engineering, Kyungpook National University, Daegu 702-701, South Korea
[2]Department of Computer Science, Bahria University, Islamabad 25000, Pakistan
[3]Department of Computer Engineering, NIT Kurukshetra, Kurukshetra 136118, India
[4]Department of Media Software, Sungkyul University, Anyang 430-010, South Korea

Corresponding author: Anand Paul (paul.editor@gmail.com)

This work was supported by the Next Generation Information Computing Development Program through the National Research Foundation of Korea funded by the Korea Government (MSIT) under Grant 2017M3C4A7066010. This study was supported by the Brain Korea 21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

ABSTRACT Video-based surveillance pedestrian detection is playing a key role in emerging technologies, such as Internet of Things and Big Data for use in smart industries and cities. In pedestrian detection, factors, such as lighting, object collisions, backgrounds, clothes, and occlusion cause complications because of inconsistent classification. To address these problems, enhancements in feature extraction are required. These features should arise from multiple variations of pedestrians. Well-known features used for pedestrian detection involve histogram of gradients, scale-invariant feature transform, and Haar built to represent boundary level classifications. Occlusion feature extraction supports identification of regions involving pedestrian detection. Classifiers, such as support vector machine and random forests are also used to classify pedestrians. All these feature extraction and pedestrian detection methods are now being automated using deep learning methods known as convolutional neural networks (CNNs). A model is trained by providing positive and negative image data sets, and larger data sets provide more accurate results when a CNN-based approach is used. Additionally, Extensible Markup Language cascading is used for detecting faces from detected pedestrian.

INDEX TERMS Pedestrian detection, facial detection, convolutional neural network, surveillance model.

I. INTRODUCTION

Four major components for pedestrian recognition are cascading, feature extraction, classification, and occlusion handling. Existing methods for deep learning acquire or plan these components either individually or sequentially [1]. In Web applications, data is often stored in relational databases (RDBs). Transforming such data from RDBs into a form suitable for the Semantic Web (SW) can cause system compatibility issues to arise, and addressing them requires understanding the weaknesses generated during the transformation process from RDB into the SW. For evolving data, sustaining changes intact is difficult to sustain. Data mapping can be used to understand their differences at the level of data types, with mappings performed using Extensible Markup Language (XML)-based data structures as an intermediate data presentation approach.

The main focus of this study is to map common features found in both RDB and SW data models based on schemas using either form of XML like Document Type Definition (DTD) or Extensible Markup Language Schema (XMLS) as an intermediate map that can help improve the transformation results. These data mappings can further improve compatibility options for data transformation. Data volumes for the web are distributed with RDBs used as back-end systems. The web is semi-structured and unorganized in a formal way by default. To translate such information into machine-readable forms the Semantic Web (SW) was introduced. The need for the SW has increased due to its capability of providing improved methods and intelligent data-seeking mechanisms, and has become a significant evolution in the next generation of the Web. Old technologies are being transformed into new ones.

Documents based on XML have become enriched as the SW has developed. Additionally, the Resource Description Framework (RDF) used for the Semantic Web is a language to represent information on the internet in the form of triplets composed of subject, predicate, and object. Triples are used among resources mapped at hierarchal levels using graph-based representations. Data types supported by XML play a key role in ensuring that transformations work properly, whereas customized data types can also be prepared using XML-based tags. Data-type customization is lacking support, which is another reason why transformations made by different techniques and algorithms fail to support each other to support compatibility of data among systems. Thus, there is a need to look into different capabilities of data types supported by XML either by DTD or XMLS while acknowledging their limitations.

Surveillance has the capability to automatically detect a person through facial detection, and can improve industrial security procedures. Current advancements in artificial intelligence and machine learning are improving the accuracy of such techniques. Improved feature learning and advanced neural network design have made object detection even more interesting and fruitful. In surveillance pedestrian detection, object tracking and crowd characterization are key factors [1], [2]. The Caltech pedestrian detection benchmark image dataset has miss rates from 11–63%. CNNs, [3] among all other Deep Neural Networks (DNNs) are the most promising for image based detection. CNNs were proposed for the characterization of entire pictures, and involves pixel-wise classification for detection, segmentation, and object tracking. CNN-based image object detection uses models trained on large datasets and tested for accuracy batch by batch until a required accuracy is achieved, but typically need significant computational time to achieve their goals.

Relational Databases are commonly in applications to store and retrieve data. RDBs are most suitable to manage significant data volumes without a concern for their semantics. Applications working with RDBs include data meaning within their logic, but semantics are again not part of the relational data model. The common framework provided within the Semantic Web (SW) provides the ability for systems to share and reuse data across different platforms and applications along with representing their data linkage. As the SW has been created for the Web and continues to advance, it has become valuable in various areas, particularly where data from different sources must be traded or coordinated. It is less feasible to replace all systems data into RDF, as many applications are still dependent on RDB-based data representation. This dependency clears the idea that either of these data models is necessary to support the current trends in data storage and retrieval. A methodology is needed that is capable of transforming data between RDB and RDF while keeping data intact. Such a methodology will be beneficial for systems whether centralized or distributed in order to work with both data models deprived of worries of change. Therefore, such an approach can reduce the conceptual gap between RDB and RDF data models, resulting in the formation of a cooperative environment of traditional and advanced systems and application.

The different socio-economic benefits provided by the proposed methodology include

- Extensive data distributed on net can be tracked with fast and efficient retrieval and storage,
- RDBs of the systems based on extensive, scattered and unorganized data can be integrated. Integrating RDBs from different systems is expensive and complex.
- The proposed methodology improves operations in data extensive and intelligent environments found in RDF-based data stores under the distributed nature of the system.
- The methodology, when implemented, will directly benefit systems containing both traditional and semantic-enriched data storage.
- It will create new opportunities for projects and research work at academic and professional levels.
- The methodology can enable a dramatic improvement via enhanced utilization and compatibility among systems using SW by leaving parts of the system untouched by required updates in Pakistani industries.
- The methodology can help in inducing large scale lossless transformation of traditional systems into semantically enriched systems. Data transformation will improve the return-on-cost of investments.
- The collaboration of Governments and Organizations systems with the outer world while remaining within the domain of the system or extending access to other domain based systems will be enables.
- Improved compatibility with systems using either semantic or relational data model will benefit us to advance in joining the trend of web based systems.

The bulk of data and information found on the Web are stored and retrieved using RDBs. Several studies have shown that collaboration via the SW with other domains extends its utilization beyond the Web. Many methods and tools have been introduced to help with providing ways to explore relational data representations for the availability of Semantic Web systems. Yet, problems exist in achieving results with high performance and compatibility.

In this research paper, the following steps are discussed for improving industrial level surveillance:

- A model is proposed built on deep learning environments for image pedestrian classification and feature extraction.
- A deep learning model is implemented with five iterations of training into a pretrained CNN.
- Identified pedestrian images are cropped for face detection using XML-based cascading.

II. BACKGROUND AND RELATED WORK

The broader research area on data is known as data science. It encompasses data gathering, manipulation, and the

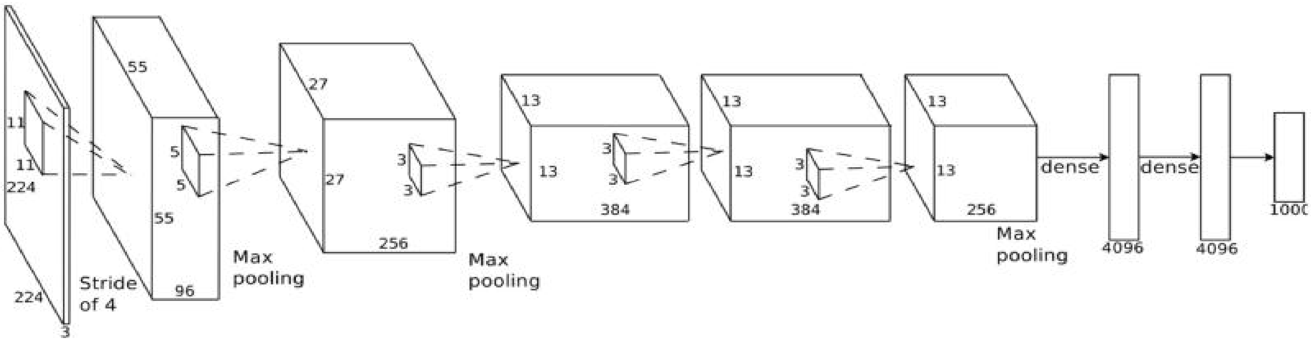

FIGURE 1. Convolutional Neural Network [3].

analysis of different statistical, mathematical, and scientific modelling and implementation techniques like data mining, data engineering, databases, machine learning, pattern recognition, and visualization. Work associated with data science involves rich data resources to improve understanding of data by studying their breadth and depth, visualization, statistical and mathematical modelling, machine learning, and artificial intelligence. Data mining and data analysis have become a major portion of data science research. Significant related work has performed in the field of big data, which is, in fact not limited to this field.

The main source of web-based data comes from social media, life, and data sciences applications stored with the use of management system for relational databases. Data transformation has been a challenging task between two different data models. These model limitations and benefits impede transformation when ensuring data correctness and completeness. A precise mechanism is required to achieve better results for mapping both data models to cover all information without any loss. This mapping may cover all possible issues concerning structure, constraints, and operations support. Data models like RDB, XML, and RDF have been built to support entirely different purposes like relationships, commonness, and data linkages based separate systems. A data model based on RDB takes care of information relationships while keeping them identifiable, whereas, XML is a markup language-based data model used to provide a high level of customization and common form of data representation. Finally, RDF-based data models are often linked data that keep traceability and relationships intact by attaching together each part of the data being used.

The beauty of XML and RDF is their capacity and capability for full data customization. XML comes with element support with structure and contains the ability to be extendable to map machines and applications where it is common, compatible and understandable. RDF-based data is primarily built on XML technologies with major evolutions to data representation by supporting relationship with uniqueness and hierarchal attachment known as data linkage. Both XML and RDF can be reformed to cover any form of data for sending and receiving with enhanced chances of correctness and accuracy. Other forms, available in semantic web such as OWL can make it difficult to cover the aspect of data as a whole. To overcome the weakness found in both fields, data storage (RDBMS) and semantically enriched (Semantic Web) is possible by providing a methodology of automatic bi-directional transformation. Such methodologies can come in handy when one transformation does not lose the originality of the data which is being transformed. Tracking vast data is becoming possible through fast and efficient approaches. This method will allow working in both data extensity and intelligence. This type of environments having the capability of back and forth transformation of data models in distributed nature of the systems for the compatibility and interoperability.

A. SURVEILLANCE USING MACHINE LEARNING

Surveillance using cameras involves computer vision for pedestrian and face detection [4]. In pedestrian detection feature extraction can take place using Haar [5] (line, edge, and four rectangular digital image features), Histogram of Oriented Gradients (HOG) [6], and Scale-Invariant Feature Transform (SIFT) [7] to cover object shape features. Among these features HOG uses local max pooling to cover any misalignments in detection procedures from training of data. Problems arise in detection when pedestrian bodies are moving. Approaches to resolving these issues involve depth analysis, occlusion handling, and visibility highlighting. A widely-used classifier known as Support Vector Machine (SVM) [8] works with both dense and sparse data models. SVM is a supervised model used in machine learning for regression and classification analysis.

B. IMAGE PROCESSING DEEP VERSUS MACHINE LEARNING

Deep learning enhancement is broadly applied in the trademark of artificial intelligence areas like transfer learning, semantic parsing, computer vision, natural language processing, and so forth. In the field of computer vision deep

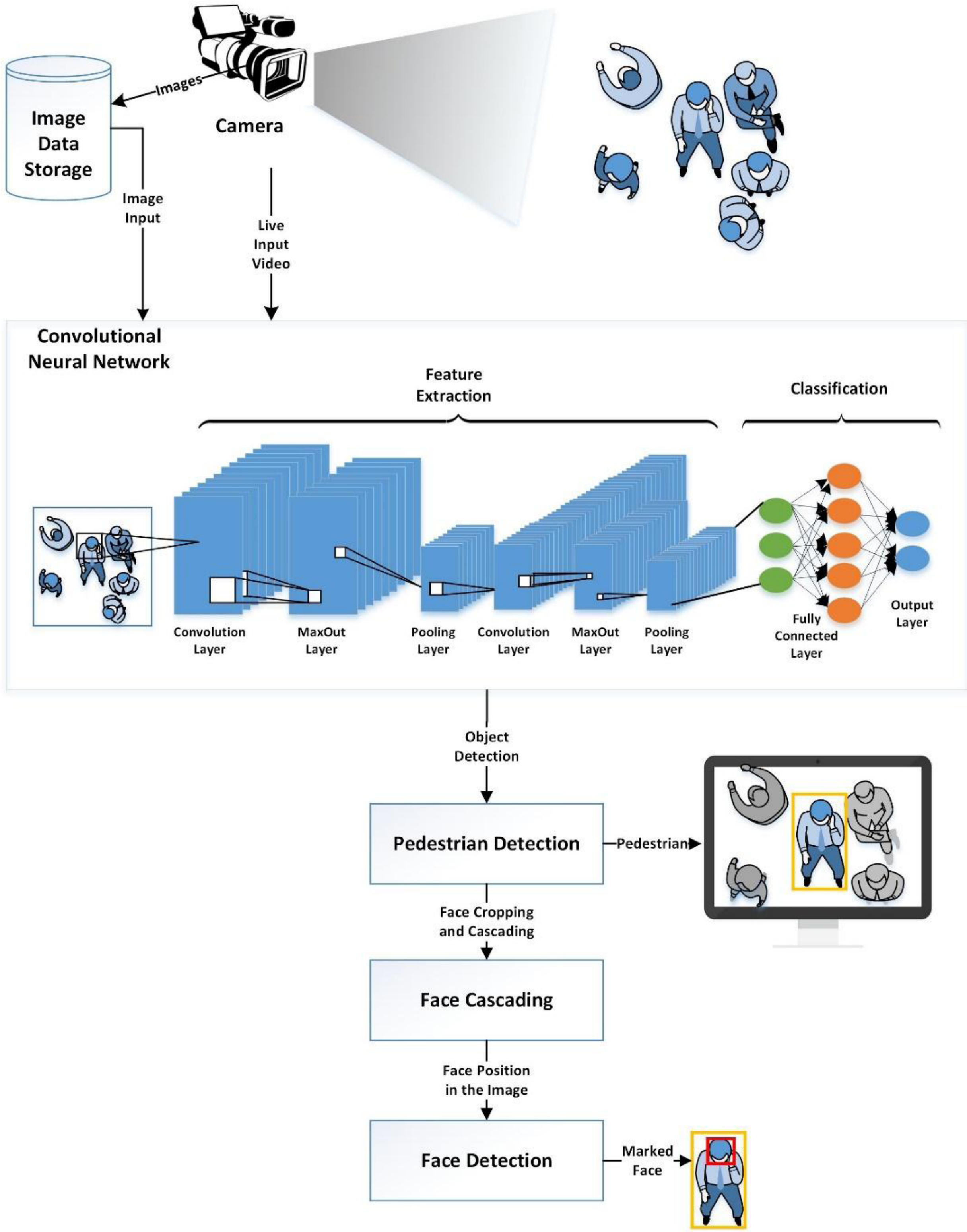

FIGURE 2. Surveillance model using deep learning for pedestrian and face detection.

learning algorithms have gained fame over old methods of machine learning. These deep learning algorithms can be further classified as Sparse Coding, Autoencoding, Restricted Boltzmann Machines (RBM), and Convolutional Neural Network (CNN). Primarily, deep learning algorithms for pedestrian detection work with contextual, feature, and occlusion handling based learning. In deep learning CNNs have shown better results than other techniques due to increased processing capacity mainly graphical processing units (GPUs), lower cost, and advanced algorithms [3].

C. DEEP NEURAL NETWORK

According to the data science landscape, deep neural networks are a special form of machine learning used for data analytics. Prior to deep learning machine learning, fewer feature-based recognition and learning techniques were available, but with the advent of greater computational power provided by GPUs, deep learning has become possible for covering almost all features. Industrial use cases for deep learning with large datasets are social media, defense, intelligence, consumer electronics, and the medical, energy and entertainment industries [9].

The true idea behind CNNs is to divide a larger problem into subproblems, step-by-step, until a required form of a result is obtained [3]. Another major benefit is that networks are easy to modify, and training sets and parameters can be updated if results are inadequate. Multiple stages of CNNs can be implemented by forming layers based on convolution, max-pooling, and fully-connected networks by pipelining, especially for image processing. More convolution layers or step addition makes learning network more complicated when used for recognition systems. For real world problems, these steps can be combined and stacked together many times to gain required results [2]. The first training convolution layer may help to recognize pedestrian sharp edge features, a second convolution layer may help in recognizing pedestrian boundaries using its prior knowledge of sharp edges, and a third layer may support recognizing an entire pedestrian with the information and learned knowledge of boundaries. A sample CNN model is shown in Figure 1.

In Figure 1, the input size is 224×224 pixels, after applying first convolution layer, and max pooling twice, and again applying convolution three more times followed by max pooling and applying the fully-connected layers. The result is a trained model capable of recognizing input images and classifying datasets negatively and positively via a sample set of categorization features.

D. FEATURE EXTRACTION TECHNIQUES

Feature extraction techniques involving face [4] and pedestrian detection are shown in II-D. Duan-Yu Chen proposed a methodology making image-based detection robust with face detection in surrounding, contextual region enhancement, with weighting based on design confidence of the model. Using edge featuring channels in the Webimg dataset in addition to an Adaboost calculation, an accuracy of 87.6% was obtained with 1948 images comprising pictures with various stance and foundation [10]. Hu *et al.* [11] represented region-based classifiers having 5% more accuracy than linear SVMs with smart-patch selection. This study used a linear SVM machine obscuring technique to localize the bases of regions having success rates of 90%. They used the Flickr database having an uncontrolled acquisition environment in 2010. Additional research was conducted with Gaussians in 2016 with dynamic CNNs and produced multiple results on difference acquisition environments.

The Mugshot database was used for frontal detection, achieving an accuracy of 97.95%, for occlusion its accuracy reached 93.12%, and finally for low-resolution images the accuracy was 93.12%. For a database based on AR the occlusion expression accuracy reached 85.62%. Overall progress improved to the accuracy to 97.95% after including Mugshot [12]. In 2014 a study trained a CNN for feature extraction for deep poselets for human detection through face cropping and feature for task-aware layers [13]. However, informative patch extraction yielded improved results nearing accuracy to 77.87% to Gallagher's database [14].

III. SURVEILLANCE MODEL USING DEEP LEARNING FOR PEDESTRIAN AND FACE DETECTION

The proposed model is comprised of two detection methods from machine learning. CNNs are used to detect pedestrian and cascading is used to detect faces of detected pedestrians as shown in TABLE 1. The procedural steps involve a industrial surveillance model for pedestrian and face detection through feature extraction as follows:

- Images are either streamed or taken from a stored dataset for surveillance.
- These images are sent to the CNN.
- In CNN images are resized to 227×277 pixels.
- Resized images are collectively sent to the first convolutional layer and then passed to max pooling layers for selecting variants and possible candidates for features.
- If the first convolutional layer detects edges or highlights, another convolutional layer addition will increase the chances of detecting pedestrian features through sharper edges.
- These are sent to max pooling for reducing unnecessary complexity.
- The fully connected layer finalizes the result.
- After separating region of detected pedestrian they are marked.
- The marked region is further cropped and passed through cascading to determine the face (as shown in Fig. 3).
- Over-fitting problem can be reduced by adding further layers of max pooling or by adding further training to the model.

Face detection of any person through images will provide much more realistic results. Test pictures can take from users and faces detected through cropped images of detected pedestrians. The image is further normalized, with three functions.

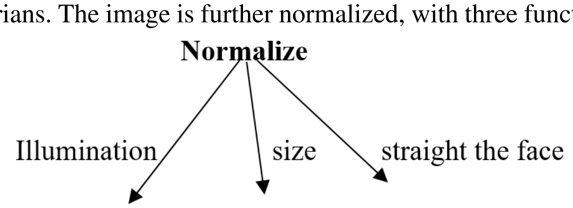

TABLE 1. Pedestrian and face feature extraction and classification techniques.

Feature extraction	Classification	Database	Acquisition environment	Dataset	Success rate	Timeline
Detection using edges [10]	Adaboost	WebImg	Lighting, Posture and contextual	1948 images	87.6	2010
Region localization on the bases of region[11]	Linear Support Vector Machine	Flickr®	Unrestrained	26766 images	90	2011
Gaussian blur [12]	Progressive CNN	Mugshot	Frontal	90K images	97.95	2016
			Occlusion		93.12	
			Low resolution		95.67	
		AR	Occlusion Expressions		85.62	
LBP RBF [15]	Support Vector Machine	Gallagher Collection Person Dataset	Unrestrained	28231 images	91.59	2014
Informative patch extraction [14]	Local DNN	Labeled Faces in the Wild (LFW)	Unrestrained	13233 images	77.87	2016
		Gallagher Collection Person Dataset	Unrestrained	4 folds Training 1 fold Test set 14760 images	72.83	
Deep Poselets for Human Detection [13]	Trained Convolutional Neural Network	Berkeley Attributes	Posture, occlusion detection	8235 images	91.7 accuracy	2014
		25K-test dataset	Posture, occlusion detection	24963 images	94.1 accuracy	
Layer wise Feature Extraction [16]	DNN	CAS-PEAL-R1, FERET	Mixed	13500 images	89.63	2014

In Normalize among three functions illumination, size and straight the face illumination is more important step in image processing in the cases like face, emotion, age, plant leaf disease detection, and so forth.

Evaluation Metrics are:

Followings are the prediction results used to define precision (p), recall (r), F_measure (F_1), and accuracy (a)

T^{+} = *True Positive* (pedestrian images predicted as pedestrian images)

T^{-} = *True Negative* (non-pedestrian images predicted as non-pedestrian images)

F^{+} = *False Positive* (non-Pedestrian images predicted as pedestrian images)

F^{-} = *False Negative* (pedestrian Images predicted as non-pedestrian images)

Precision represents relevant instances as fractions of the retrieved instances whereas recall represents relevant instances of whole sets. Precision and recall formulae are as follows:

$$p = T^{+}/(T^{+} + F^{+}) \tag{1}$$

$$r = T^{+}/(T^{+} + F^{-}) \tag{2}$$

The F_measure is also known as F score. The F_measure is used to measure the efficiency for precision and recall when weighted as shown in Eq. (3IV. It is calculated using following formulae:

$$F_1 = 2 \times \left(\frac{p \times r}{p + r}\right) \tag{3}$$

$$a = \frac{T^{+} + T^{-}}{T^{+} + T^{-} + F^{+} + F^{-}}$$

$$T^{+}rate = T^{+} \times (T^{+} + F^{-})$$

$$F^{+}rate = 1 - T^{-} \times (T^{-} + F^{+})$$

$$MR = \frac{F^{-}}{T^{+} + F^{+}} \tag{4}$$

TABLE 2. Training average error and miss rate for Caltech Pedestrian dataset.

Iteration	*Average Error Rate*	*Average Miss Rate*
1	0.135718	0.77682
2	0.128214	0.835863
3	0.119122	0.837793
4	0.11341	0.843226
5	0.111615	0.830095

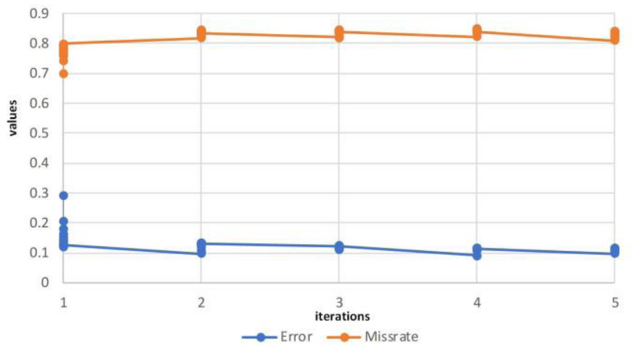

FIGURE 3. Details overview of Error rate and Miss rate gained during five iterations of Caltech.

IV. EXPERIMENTATION AND RESULTS

Convolutional neural networks work with two variations of layers. One type of convolutional layer is built on a non-linear stimulation function and a second layer has a non-linear sub-sampling mechanism. Further mixing of these layers can be used when building a model based on the nature of object detection required. Furthermore, squeezing of the input size is attained. The formal definition of CNNs is shown in Eq. (5):

$$f(i, \partial^{(1)}) = i^* = f_{out}\ \mathrm{o} f_L\ \mathrm{oK}\ \mathrm{o} f_2\ \mathrm{o} f_1(i^{(0)}) \tag{5}$$

Q $i^{(0)} = i$ are input and $\partial^{(1)}$ is a convolutional parameterization like a weight, and the ^{o} symbol is used for representing composition (for further details see [3]).

To evaluate the model CNN-based training was performed using the Caltech Pedestrian Benchmark dataset. Caltech-Train used 60000 negative and 4000 positive sample images for training through 5 iterations of training the model. In Table 2, each iteration trained for 1210 branches showing their average error and miss rates. Through these results, with each new iteration of training the error rate decreased and the miss rate increased. For example, in iteration 1, the average error rate was 0.13 and the average miss rate 0.77 but in iteration 2, the average error rate became 0.12 and the average miss rate increased to 0.835. In iterations 3, 4, and 5 the average error rate became 0.119, 0.113, and 0.111, respectively, whereas the average miss rate became 0.837, 0.843, and 0.830, respectively.

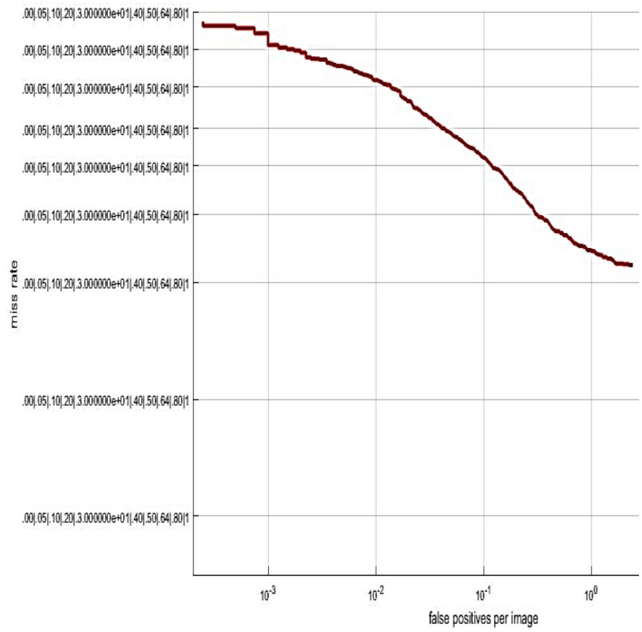

FIGURE 4. Our Model's Result for Caltech Test dataset.

Fig. 4 represents the complete variation of error rate and miss rate using blue and orange respectively. The miss rate was 80–90% showing improvement in accuracy with each iteration of training whereas error decreased to 0.1.

Caltech-test when used on the trained model gives reasonable result nearing 0.4841 result set. In Fig. 5, miss rate decreases with the increase of false positive per image.

After the training phase of the proposed CNN, the testing phase was performed by feeding an image to the model. To test resultant layers, feature extraction capabilities for the pedestrian detection were activated. Tests began by highlighting activation areas of the original image for examination. Initial layers simply supported determination of color and edges. Deeper layers extracted features like shoulders, heads, and legs of pedestrians. Such feature identification shows that the network learned through the training phase. Convolutional layer 1 results in the montage show each feature activation using 96 images on 8×12 grids, where each channel is represented by separate image in the layer as shown in Fig. 6.

Each image in the form of a square represents an activated channel for the convolutional layer 1, where white spots represent strong positive activations found as candidate to be feature of interest in the image feeding for testing. Black spots in the activated channels show strong negative activations. A grey spot was not considered as a strong candidate to be a feature. Max pooling and resizing with highly activated areas are shown in Fig. 7.

In the first convolutional neural network the majority of features detected are through color and edges. In the deeper layers of convolution, features are more complex and detection accuracy improved. Further layers build up features

FIGURE 5. Activations of first convolutional layer for pedestrian detection.

FIGURE 6. The Activations in Specific Channels (left) and the strongest activation channel (right).

by adding the results of previous layers involving calculations, reshaping and representing activations in a montage as shown in Fig. 8.

In the resultant images, the strongest activated channels in a montage image were not capable of representing detailed

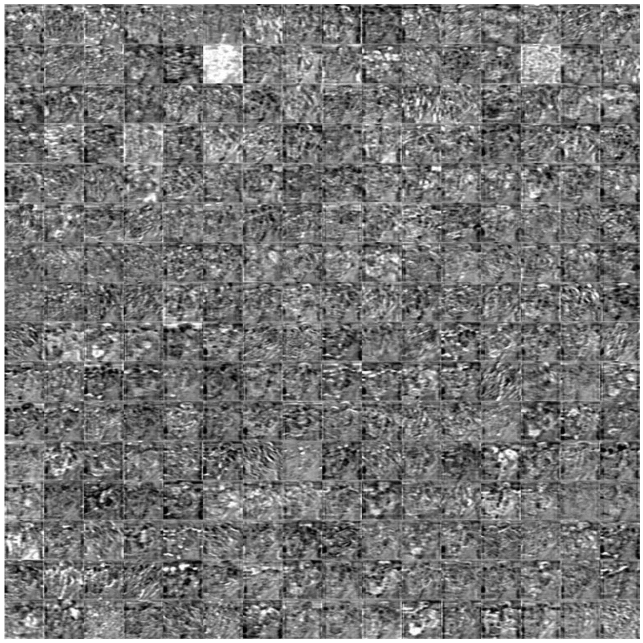

FIGURE 7. Examining a deeper layer.

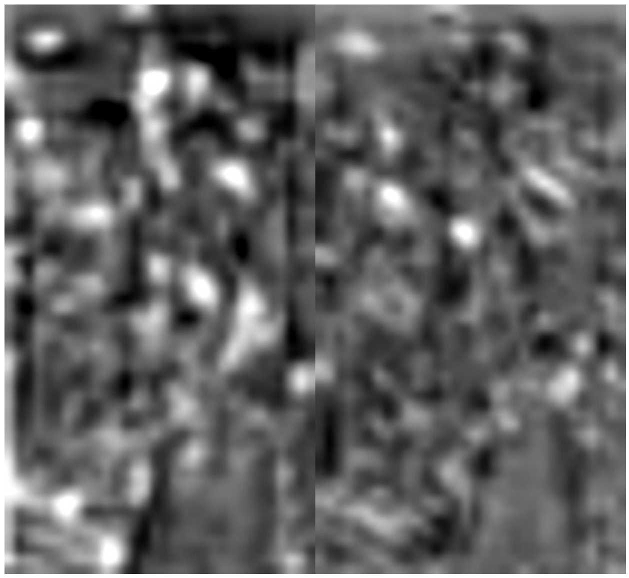

FIGURE 8. Examining channels 3 and 5 for further activations in the deeper layers of convolution only positive activations are found due to the rectified linear unit (ReLU). Additionally, resulting activations with the original image are shown in Fig. 10 for face detection.

features like few other channels. A few channels may be activated on features like heads and legs. The examination of channels such 3 and 5 shows additional activations as shown in Fig. 9.

The Caltech test results with other techniques with their accuracy results [17] represent different miss rate calculations performed over joint ventures to determine which technique for pedestrian detection outperforms others, as shown in Fig. 11. Through the resultant graph curve, the UDN shows better performance than others.

FIGURE 9. Computation of the activations in the given image.

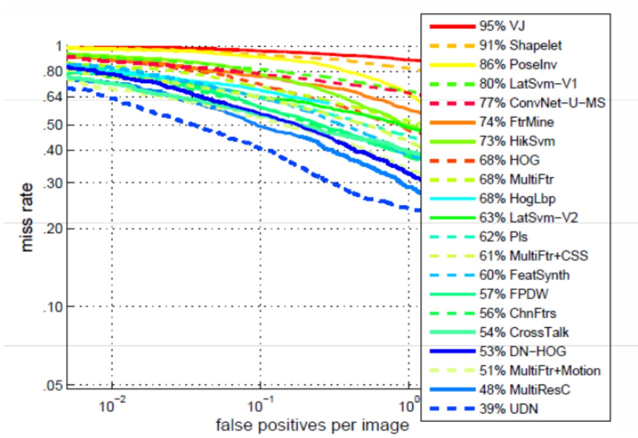

FIGURE 10. Caltech test results with other techniques with their accuracy results [17].

TABLE 3. Evolution matrices results before and after alignment resizing and without pre-processing.

	Alignment after resizing	Alignment before resizing	Without Pre-processing
p	0.958089	0.974339	0.515709
r	0.9525	0.967321	0.898869
F_1	0.955286	0.970818	0.655397
a	0.955417	0.970923	0.527381
$T^{+}rate$	0.9525	0.967321	0.898869
$F^{+}rate$	0.041667	0.025476	0.844107

The results in Table 3 show a promising outcome as the trained model predicted pedestrian an accuracy of 90%, whereas test cases provided results reaching near 60%. This is because the false positive prediction was quite low through the trained model. Results further could be improved by improving images dataset and further training. The same results are represented in bar form in Fig. 11.

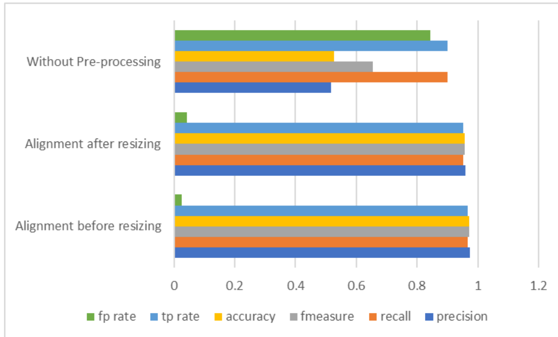

FIGURE 11. Bar chart for results shown in Table 3.

V. CONCLUSION

By combining cascading and CNNs for face and pedestrian detection surveillance improved dramatically. CNNs have the promising feature of future modification in datasets or modeling of neural networks to further improve their accuracy. Larger data volumes are better for feature extraction and object detection in pedestrian and face surveillance. Real-time surveillance becomes faster on trained models but training the model is difficult. With the help GPU hardware, computational power is now available that is required for training CNN models having large feature set. Results obtained provide higher accuracy for features needed for surveillance. In the future, research will involve semantic tagging and linking industrial profiles of visitors for improving the quality of surveillance.

REFERENCES

[1] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 152–159.

[2] P. Luo, Y. Tian, X. Wang, and X. Tang, "Switchable deep network for pedestrian detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 899–906.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[4] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.

[5] C. F. Chen and C. H. Hsiao, "Haar wavelet method for solving lumped and distributed-parameter systems," *IEE Proc.-Control Theory Appl.*, vol. 144, no. 1, pp. 87–94, Jan. 1997.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[7] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.

[8] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.

[9] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 8599–8603.

[10] D.-Y. Chen and K.-Y. Lin, "Robust gender recognition for uncontrolled environment of real-life images," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1586–1592, Aug. 2010.

[11] S. Y. D. Hu, B. Jou, A. Jaech, and M. Savvides, "Fusion of region-based representations for gender identification," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.

[12] F. Juefei-Xu, E. Verma, P. Goel, A. Cherodian, and M. Savvides, "Deepgender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun./Jul. 2016, pp. 68–77.
[13] N. Zhang, M. Paluri, M. A. Ranzato, T. Darrell, and L. Bourdev, "PANDA: Pose aligned networks for deep attribute modeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1637–1644.
[14] J. Mansanet, A. Albiol, and R. Paredes, "Local deep neural networks for gender recognition," *Pattern Recognit. Lett.*, vol. 70, pp. 80–86, Jan. 2016.
[15] E. Fazl-Ersi, M. E. Mousa-Pasandi, R. Laganière, and M. Awad, "Age and gender recognition using informative features of various types," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5891–5895.
[16] Y. Jiang, S. Li, P. Liu, and Q. Dai, "Multi-feature deep learning for face gender recognition," in *Proc. IEEE 7th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Dec. 2014, pp. 507–511.
[17] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2056–2063.

SADIA DIN received the B.S. degree in computer engineering from the Comsats Institute of Information Technology, Abbottabad, in 2010, and the M.S. degree in computer engineering from Abasyn University Islamabad, Pakistan, in 2015.She is currently pursuing the Ph.D. degree with Kyungpook National University, Daegu, South Korea. In 2015, she was a Visiting Researcher with CCMP Laboratories, Kyungpook National University, where she involved on big data and Internet of Things projects. She was a Lecturer with Abaysn University Islamabad until 2016. Her research interests include IoT, big data analytics, wireless sensor networks, and 5G/4G.

ANAND PAUL (SM'14) received the Ph.D. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently an Associate Professor with the School of Computer Science and Engineering, Kyungpook National University, South Korea. He is a delegate representing South Korea for the M2M focus group and for MPEG. He has guest edited various international journals and he is also part of Editorial Team for the *Journal of Platform Technology* and the ACM *Applied Computing review* and *Applied Cyber- Physical Systems*. His research interests include algorithm and architecture reconfigurable embedded computing. He was a recipient of the Outstanding International Student Scholarship award from 2004 to 2010, the Best Paper Award in National Computer Symposium in 2009, and International Conference on Softcomputing and Network Security, India, in 2015. He serves as a reviewer for various IEEE /IET/Springer, and Elsevier journals. He is the Track Chair for Smart human computer interaction in ACM SAC 2015, 2014.

AWAIS AHMAD received the Ph.D. degree in computer science and engineering from Kyungpook National University, Daegu, South Korea. He was a Visiting Researcher with INTEL-NTU, National Taiwan University, Taiwan, in 2014, where he was involved in Wukong Project (Smart Home). He was also serving as a Lab Admin of CCMP Laboratories from 2013 to 2017. He is currently an Assistant Professor with the Department of Information and Communication Engineering, Yeungnam University. Since 2013, he has authored or co-authored over 70 International Journals/Conferences/Book Chapters in various reputed IEEE, Elsevier, and Springer Journals, whereas in leading conferences, i.e., IEEE GLOBECOM 2015, IEEE GLOBECOM 2016, IEEE LCN, 2016, and IEEE ICC 2017, respectively. His current research interest includes big data, Internet of Things, social Internet of Things, and human behavior analysis using big data. He has been a key contributor in the said fields. He was a recipient of prestigious awards, including the IEEE Best Research Paper Award: International Workshop on Ubiquitous Sensor Systems (UWSS 2015), in conjunction with the Smart World Congress (SWC 2015), Beijing, China, the Research Award from President of Bahria University Islamabad, Pakistan, in 2011, the Best Paper Nomination Award in WCECS 2011 at UCLA, USA, and the Best Paper Award in 1st Symposium on CS&E, Moju Resort, South Korea, in 2013. He was also a Best Outgoing Researcher of CCMP laboratory. He is also serving as a Guest editor in various Elsevier and Springer Journals, including FGCS, *Sustainable Cities and Socity*, and RTIP. Moreover, he is an Invited Reviewer in various journals, including the IEEE COMMUNICATION LETTERS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEM, the IEEE JSTAR, *AdHoc Networks* (Elsevier), *Computer Network* (Elsevier), and the IEEE *Communications Magazine*.

B. B. GUPTA is currently with the Department of Computer Engineering, National Institute of Technology, Kurukshetra, India. He has authored or co-authored over 100 research papers in international journals and conferences of high repute, including IEEE, Elsevier, ACM, Springer, and Inderscience. His research interest includes information security, cyber security, intrusion detection, cloud security, mobile security, web security, big data analytics, botnet detection, phishing, DDoS attacks, and network performance evaluation. He also serves as a reviewer for IEEE, Springer, Elseiver, Wiley, Taylor & Francis, and Inderscience. He was a recipient of the Young Faculty Research Fellowship Award from MeiTY, Govt of India, in 2017. He has contributed in many International conference program committee of in this capacity.

SEUNGMIN RHO received B.Sc. degree in computer science from Ajou University, South Korea, in 2001, and the M.Sc. and Ph.D. degrees in information and communication technology from the Graduate School of Information and Communication, Ajou University, in 2003 and 2008, respectively. He was an Assistant Professor with the Division of Information and Communication, Baekseok University, in 2012. He is currently on the faculty of the Department of Media Software, Sungkyul University, South Korea. From 2009 to 2011, he had been working as a Research Professor with the School of Electrical Engineering, Korea University. From 2008 to 2009, he was a Post-Doctoral Research Fellow with the Computer Music Laboratory, School of Computer Science, Carnegie Mellon University. He visited Multimedia Systems and Networking Laboratory, University of Texas at Dallas, from 2003 to 2004. Before he joined the Computer Sciences Department of Ajou University, he spent two years in industry. His current research interests include database, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management, and computational intelligence. He has authored or co-authored over 180 papers in refereed journals and conference proceedings in these areas. He has been involved in over 20 conferences and workshops as various chairs and over 30 conferences/workshops as a program committee member. He was a recipient of few awards, including Who's Who in America, Who's Who in Science and Engineering, and Who's Who in the World in 2007 and 2008, respectively. He has been appointed as an Editor-in-Chief for the *Journal of Platform Technology* since 2013. He has edited a number of international journal special issues as a Guest Editor, such as the *Enterprise Information Systems*, the *Multimedia Systems*, the *Information Fusion*, the ACM *Transactions on Embedded Computing*, the *Journal of Real-Time Image Processing*, the *Future Generation Computer Systems*, the *Engineering Applications of Artificial Intelligence*, the *New Review of Hypermedia and Multimedia*, the *Multimedia Tools and Applications*, the *Personal and Ubiquitous Computing*, the *Telecommunication Systems*, and the *Ad Hoc and Sensor Wireless Networks*.