International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

DCNN Optimization using Wavelet-based Image Fusion

Abdullah A. Alshehri, Soundararajan Ezekiel

Abstract— We propose to develop image fusion algorithms and architecture for enhanced deep learning and analysis of large sets of data. Usually, images captured from different perspectives, using different types of sensors, different frequencies, etc. must be considered separately and interpreted by human operators. Using image fusion techniques, different forms of sensor information into a single data feed for a neural network to interpret and learn from can be implemented. This will increase the accuracy of neural network classification, as well as improve effectiveness in situations involving suboptimal conditions, such as obstructed or malfunctioning sensors. Another disadvantage of current deep learning technique is that they often require massive datasets to train to an acceptable level of accuracy, especially when situations involve potentially thousands of classification categories. Increasing the size of the dataset exponentially increases the amount of time to train, even when training on relatively simple neural network architectures. In protection scenarios, where new classes of threats can emerge frequently, it is unacceptable to have to take down the security system for long periods of time and train it to identify new threats.

Keywords— Image Fusion, Multifractal Analysis, Convolutional Neural Networks, Heterogeneous DCNN Fusion.

I. INTRODUCTION

In recent years, advancements in machine learning have increased their practical viability, allowing them to see widespread proliferation for the purpose of classification. By leveraging Graphic Processor Unit (GPU)-enabled neural networks, it allows for the analysis and classification of massive amounts of text, images, and audio at rates that are infeasible for humans to perform manually [1,2]. Despite their promise, however, neural networks still face limitations in their capabilities, especially in the realm of image analysis [3]. One such limitation lies in the fact that one of the most computationally resource intensive steps is in the actual training of the neural network, with the subsequent network's performance being highly correlated to the quality of the training data it has received. In the case of visible light imagery especially, in which deep learning models are reliant on single information streams such as raw pixel data, distortions in images such as blurring can impact the integrity of the network. A robust, well-trained image classification model, therefore, requires a considerable amount of imagery captured from a suitable variety of angles, focuses, etc. which can significantly add to the computational time required to train the network.

Normally, images captured from different perspectives or using different sensors must be each individually analyzed and interpreted manually. One possible solution to this is with the use of image fusion techniques, in which images of the same scene captured from different angles, or with multi-modal or multi-spectral sensors are fused into a single stream of data which can be used as a model's input [4, 5, 6, 7]. The development of these multisensory applications began in the 1980's as subcategory of remote sensing. This field evolved from two sub-branches of data fusion. This first way being abstraction-wise, where various types of data are fused including; video, audio, and numerical data. This branch of research directly relates to the project at hand, one example of this topic being put in practice is in cyberspace. Some of the earliest applications of fusion were used to fuse network data in an intrusion detection system. Another previous application of fusion was in medicine, where it combined electroencephalography signals with electrooculography and respiratory signals in order to develop fatigue models for patients. As time progressed, numerous different methods were being practiced. These methods include but are not limited to pyramid-based, wavelet-based, and data driven methods [8]. Image fusion techniques being observed in this study are comprised of multi-focus and/or multi-modal fusion, which allow for key features from separate images to be fused into a single image. The capability to do this has allowed image fusion to be utilized in several areas including remote sensing, medical image analysis, and environmental monitoring [9]. However, in the case of these types of images, optimizing the identification of features and segmenting regions of interest can be computationally complex [10]. With multi-focus fusion, images captured from differing focal points, such as the foreground and background, can be fused into a single image that combines the most salient information from both, retaining only the in-focus regions from each. Multi-modal imagery includes scenes captured by both visible light cameras in conjunction with other types of sensors such as thermal imaging. In these circumstances, while infrared sensors are able to capture features undetectable by normal cameras, they can fail to capture details such as topography and other environmental details. By fusing image data of the same scenes, additional information can thereby be provided to the network during training, rather than the images being processed and analyzed separately while also improving the effectiveness of data captured in suboptimal conditions or with malfunctioning sensors. Ultimately this creates a more optimized network while also providing a failsafe for any unforeseen environmental conditions. An additional benefit is that by reducing the total number of images needed to train the network, the computational requirements are also reduced without negatively impacting its performance.

Revised Manuscript Received on February 05, 2020.
* Correspondence Author

Abdullah A. Alshehr, Electrical Engineering, King Abdulziz University , P.O.Box 80200, Jeddah, 21589, Saudi Arabia
E-mail: ashehri@kau.edu.sa

Soundararajan Ezekiel,Indiana University of Pennsylvania, 1011 South Drive, Indiana, PA 15705, USA
E-mail: sezekiel@iup.edu

By utilizing image fusion techniques, only the most significant features of the original images are retained while discarding irrelevant ones such as out of focus regions. Moreover, by introducing these methods, various network enhancements will be contributed to both efficiency and consistency.

The remainder of this paper is organized as follows: Section II outlines the technical background and concepts utilized in this study. Section III details the methodology for both the image fusion techniques used as well as heterogeneous DCNN fusion and classification. Section IV provides the results of the classification accuracy for varying types of fused imagery Section V discusses the significance of the results as well as the direction of future work.

II. MATHEMATICAL BACKGROUND

A. Wavelets

Wavelets are oscillating functions that are finite, the oscillations are like a sine or cosine wave but have an average value of zero. The wavelet's amplitude will end and start at zero after a finite number of oscillations. For any function to be a wavelet the following conditions must be true:

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{1}$$

$$\int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega = C_\psi \tag{2}$$

Ψ(ω) is the Fourier transform and *C*C_Ψ is the *admissible constant*. Multiple wavelets have been derived by Daubechies [11]. Wavelets can be defined and further categorized on a discrete grid vs. a continuous axis, usually time or space. They can also be categorized on whether they are in the complex or the real plane. Wavelets have two characteristics that are fundamental to every wavelet, their definitions for rescaling and translation. Given a mother wavelet *Ψ(x),* an entire family of wavelets $\psi_{j.k}(x)$ is defined in equation 3:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{|j|}} \psi\left(\frac{x-k}{j}\right) \tag{3}$$

j is our scaling variable and k is the variable of translation. These characteristics allows for smaller, immediate changes in signals to be able to be detected. This transform is the most suitable for the detection of point-wise edges. The continuous wavelet transform, defined as the inner product of a function *f(x)*∈ L_2 (ℝ) and a wavelet Ψ(x), is expressed in the following equation:

$$f, \psi_{j,k} = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{|j|}} \psi\left(\frac{x-k}{j}\right) dx \tag{4}$$

When it comes to image processing, the function f is representing our image with the wavelet transform applied to it. Since images are not commonly processed in a continuous space as a continuous-space function, but more commonly as a discrete-space function. Therefore, the discrete wavelet transform is more commonly used in the processing of images. Like the CWT, the DWT of the function *f*, which we will denote G_ψ, as expressed in equation 5:

$$G_\psi(f,\psi) = \frac{1}{\sqrt{M}} \sum_{i,j=1}^{n} f(x)\psi_{j,k}(x) \tag{5}$$

Where M is the scaling weight. When moving any transformation from continuous to discrete there are problems that can result. Such as, in the discrete domain, the wavelet transform loses directionality and shift-invariance.

B. Multi-Resolution Fusion

Multi-resolution analysis (MRA) is a technique that is utilized to approximate images across a series of resolutions [12]. By decomposing images, the most significant features can be retained while isolating noise and blur allowing the edge coefficients to be manipulated rather than raw pixel data. MRA allows images to be deblurred while diminishing the effect of distortions that may arise by decomposing it into differently scaled components [13]. Multiresolution analysis is a sequence of closed subspaces $V_n, n \in \mathbb{Z}$ in $\mathbb{L}^2(\mathbb{R})$, in a containment hierarchy.

$$.... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset .... \tag{6}$$

The nested spaces contain an intersection with the zero function and a union that is dense in $\mathbb{L}(\mathbb{R})$,

$$\cap_n V_j = \{0\}, \overline{\cup_j V_j} = \mathbb{L}^2(\mathbb{R}) \tag{7}$$

The hierarchy (7) is constructed such that *V*-spaces are self-similar,

$$f(2^j x) \in V_j \text{ iff } f(x) \in V_0 \tag{8}$$

and there is a scaling function $\psi \in V_0$ whose integer-translates span the space V_0,

$$V_0 = \left\{ f \in \mathbb{L}^2(\mathbb{R}) \mid f(x) = \sum_k c_k \psi(x-k) \right\} \tag{9}$$

and for which the set $\{\psi\}$ is an orthonormal basis.

For the purposes of image fusion, MRA is used to decompose separate images of the same scene into their directional coefficients, which can then be manipulated and fused, rather than the raw pixel data. In this, multi-resolution image fusion serves a dual purpose in being used for denoising in addition to the actual image fusion.

C. Multi-Modal Fusion

The use of fusing images from different modalities, such as visible light and infrared is to capture the most significant features of each medium [14]. A single stream of data is formed through the fusion of these images, allowing for information that can only be obtained from a certain modality, such as thermal imaging, with visible light imagery. which is more adept for capturing the details of terrains, etc [15]. When this process occurs however, the introduction of new information can cause distortions such as artificial shadows in topographical and morphological information. Due to this, careful optimization is required in order to extract only the most salient information without causing artifacting or pixelization [15].

D. Multi-Focus Fusion

In image processing, one of the greatest restrictions is the ability of the camera to focus. These differing levels of focus in separate images of the same scene can cause variations in the visual quality of regions in the image. Multi-Focus Fusion poses as a solution to this restriction. Through this, the focused sections of an image are linked together to create an all-in-focus image. These sections come from the various aspects of the frame, including the foreground, background, sides, middle, or any other applicable section. This new image would be more accurate than any of the source images. Spatial frequency (SF), sum-modified-Laplacian (SML), and Tenenbaum gradient (Tenengrad) are examples of these focus measures [16]. Many of these image fusion techniques support information fusion applications [17, 18]. Multi-Focus Fusion has applications in a wide array of fields. Aerial surveillance, three-dimensional reconstruction, photography, and video production all have ways to benefit from this fusion technique [19, 20, 21]. Multi-perspective fusion, shown in Figure 1, involves different photos taken of the same object or scene from different angles.

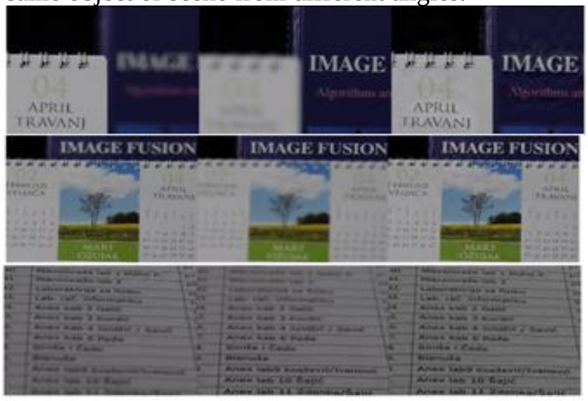


Figure 1: Foreground focus (left), background focus (center), and fused (right)

As previously mentioned, the images can come from a variety of sources. These sources can range from different types of cameras or sensors. Then algorithms can be selected to meet mission requirements [22, 23], based on the context, such as the environment [24], sensor phenomenology [25], and image focus. Multi-focus image fusion methods need to select different methods to meet both the quantitative and qualitative requirements for mission effectiveness [26].

E. Max

The maximum criterion will use the absolute value of each of the entries in the two following matrices:

$$f_{ij} = \begin{cases} a_{ij} & if\ |a_{ij}| \geq |b_{ij}| \\ b_{ij} & otherwise \end{cases} \qquad (10)$$

Absolute value is needed in the decision statements because the multi-resolution coefficients in the matrix can be negative, although the pixel values can only be positive. This function will compare the two entries of matrix *a* and *b* and insert the one with the higher value into the corresponding entry in matrix *f*.

F. Min

The minimum criterion will be using the absolute value of the entries in the matrices just as we did in the max:

$$f_{ij} = \begin{cases} a_{ij} & if\ |a_{ij}| \leq |b_{ij}| \\ b_{ij} & otherwise \end{cases} \qquad (11)$$

This will determine which of the entries in the two matrices is smaller and put that value in the corresponding entry in our matrix *f*.

G. Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate analysis technique that is most used for the reduction of the dimensionality of larger matrices and the extraction of features. [27, 28, 29]. PCA takes correlated variables of large matrices and reduces them into their corresponding principal components. The principal components contain the most important features of the data points and are linearly independent. A matrix X with n rows, is used to represent the sensor reading observations, and has m columns which will represent each of the parameters we are using for the dataset. PCA is an orthogonal linear transformation that changes our matrix X into a set of vectors of weights w with m dimensions, defined as:

$$w_{(k)} = (w_1, w_2, ..., w_m) \qquad (12)$$

The vectors of weights contain the principal component scores *t* mapped from each row of X where *t* is defined as:

$$t_{(1)} = (t_1, t_2, ..., t_l) \qquad (13)$$

where $t_{k(i)} = x_{(i)} \cdot w_{(k)}$ for $i = 1, ..., n \;\; k = 1, ..., m$. Each of the k^{th} PCs are ordered in such a way that our first principal component captures most of our data and has the largest amount of variance. The first principal component, $w_{(1)}$ maximizes the variance by satisfying the following:

$$w_{(1)} = \underset{\|w\|=1}{\arg\max}\{\|Xw\|^2\} = \underset{\|w\|=1}{\arg\max}\{w^T X^T Xw\} \qquad (14)$$

The next subsequent k^{th} principal component is found by taking our matrix *X* and subtracting the first *k* - 1 principal components from it such that:

$$\hat{X}_k = X - \sum_{i=i}^{k-1} Xw_i w_i^T \qquad (15)$$

and then calculating the weight vector that maximizes the variance:

$$w_{(k)} = \underset{\|w\|=1}{\arg\max}\{\|\hat{X}_k w\|^2\} = \underset{\|w\|=1}{\arg\max}\{\frac{w^T \hat{X}_k^T \hat{X}_k w}{w^T w}\} \qquad (16)$$

As principal component analysis is an orthogonal linear transform, each pairing of principal components is orthogonal since they were derived from the eigenvectors of the covariance of the data, which themselves are always symmetric [30].

H. Deep Convolutional Neural Networks

Deep convolutional neural networks (DCNNs) utilize deep, feed-forward architectures to learn the most significant features of their input. The trained model can then be used to classify new input into labels based on the learned features of the training data. In this regard, DCNNs are specific to the data on which they were trained, and new data must be able to be classified into existing labels [31]. The feed-forward architecture consists of a varying amount of layers stacked onto one another, using the previous layer's output as input for the next connected layer [32], shown in Figure 2. The model is first initialized with all filters, weights, and parameters set to random values. Training data is then input into the network, which goes through a forward propagation step starting in the convolutional layer, which is comprised of filters that calculate an activation map containing the output of each convolution over the entire input [33, 34]. By activating when specific details are detected, the filters are able to be used to learn the features of the input. The pooling layers of the network map the locations of the features of the images in relation to other features [35]. The fully connected layers of the network then receive the activation mappings of the previous layers that contain the learned features, which are used for high-level reasoning and classification. At the output layer, the summation of the error is then calculated across all the classes and a backpropagation step is then applied using gradient descent to update the filter values and weights. The parameters of the model are subsequently optimized to improve classification accuracy.

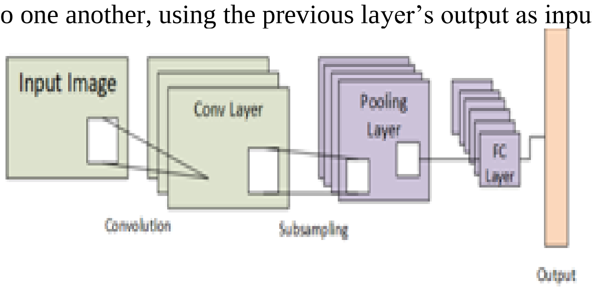


Figure 2: Topology of a DCNN

I. FC_7 Layer

The fully connected layers of a neural network model are responsible for the using the weights of the previous layers for reasoning and classification. In the stacked layer architecture of the network, the fully connected layers are the final in the stack, that receive the output of the convolutional and pooling layers, representing the high-level features' activation maps. The output of the fully connected layers is a vector that contains the probabilities of each label through determining how features correlate to each class. The penultimate fully connected layer, FC_7 is a vector that contains the activation weights of each of the features across all classes. As such, this layer is ideal for feature extraction since they contain the high-level correlations between features and classes.

J. Heterogeneous DCNN Fusion

Heterogeneous DCNN Fusion is a process that leverages multiple neural network models of differing architecture through extracting their respective feature vectors from the penultimate fully connected layer [36, 37]. The FC_7 layer is extracted from each of the networks, as the layer's position is consistent in the architecture stack across them. Additionally, despite being fused from different model architectures, because the networks were trained using the same datasets, the dimensions of the FC_7 layer are also consistent. After being extracted, the feature vectors are then fused into a single vector [37, 38]. The fused feature vector can then be used as input for a Support Vector Machine (SVM) classifier, which is then used for the actual task of classification.

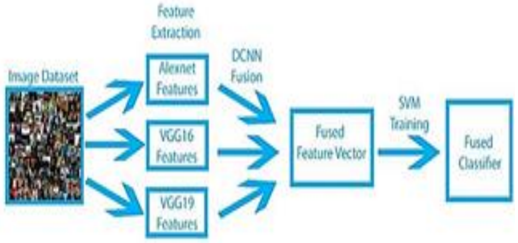


Figure 3: Heterogeneous DCNN Fusion

III. METHODOLOGY

The general methodology for our study is outlined in Algorithm 1. Images sets containing pictures of the same scenes from multi-focus or multi-modal perspectives were first gathered. The images were then pre-processed using wavelet-based denoising in order to enhance the individual images before fusing them. Images of the same scene were then combined using image fusion techniques to combine the most relevant and significant information from the source images. As the directional coefficients of the decomposed images indicate edges and directionality, they're ideal for fusing the most prominent features from input images. Multi focus images, for example, were found to typically have more prominent directional coefficients in the in-focus regions when compared to the out of focus regions, seen in Figure 4.

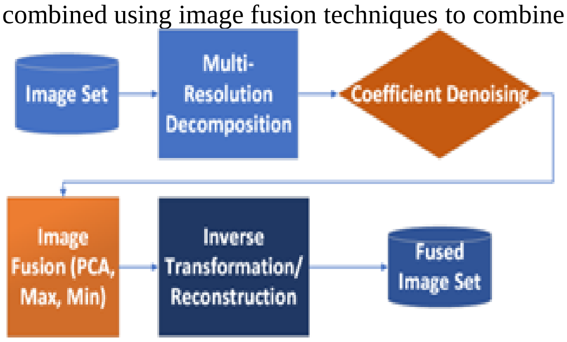
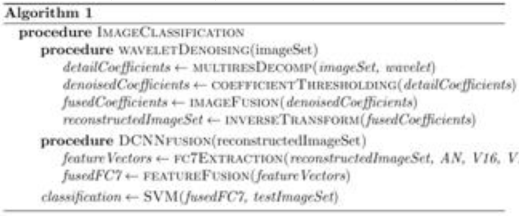


Image Set
Multi-Resolution Decomposition
Coefficient Denoising
Image Fusion (PCA, Max, Min)
Inverse Transformation/ Reconstruction
Fused Image Set

Figure 4: Image Fusion Methodology

The decomposed and denoised wavelet coefficients were then fused using various fusion methods to extract as much information from the regions of interest as possible. In the case of multi-focus image data, image fusion allows for scenes with differing foreground and background focuses to be fused into a single, in focus image; for images captured from multi-modal sensors such as visible light and infrared, image fusion allows for key unique features to be fused into a single data stream while retaining the most significant information of both. The fused coefficients were then reconstructed to create the fused image set. The fused set was then split between training and test sets for three neural network models, AlexNet, VGG16, and VGG19, with their respective FC₇ layers being extracted before classification. The FC₇ layers of each of the networks, containing the high-level reasoning and weights for each of the classes, are then fused using heterogeneous DCNN fusion, fusing the three feature vectors into a single vector containing the weights of each of the networks. The fused feature vector is then input into an SVM, which is responsible for handling classification.



```
Algorithm 1
procedure IMAGECLASSIFICATION
    procedure WAVELETDENOISING(imageSet)
        detailCoefficients ← MULTIRESDECOMP(imageSet, wavelet)
        denoisedCoefficients ← COEFFICIENTTHRESHOLDING(detailCoefficients)
        fusedCoefficients ← IMAGEFUSION(denoisedCoefficients)
        reconstructedImageSet ← INVERSETRANSFORM(fusedCoefficients)
    procedure DCNNFUSION(reconstructedImageSet)
        featureVectors ← FC7EXTRACTION(reconstructedImageSet, AN, V16, V19)
        fusedFC7 ← FEATUREFUSION(featureVectors)
    classification ← SVM(fusedFC7, testImageSet)
```

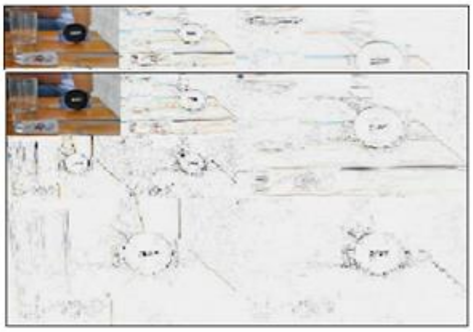


Figure 5: Stronger background coefficients (top) & foreground coefficients (bottom)

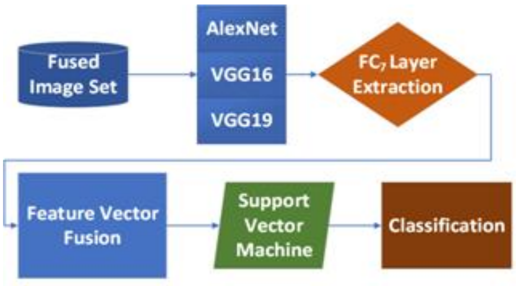


Figure 6: DCNN Fusion & Classification Methodology

IV. RESULTS

Tests were conducted on two primary types of images, multi-focus images containing a foreground and background focus, as well as images of different modalities, namely near infrared and visible light imagery. Tests were conducted to ascertain how the classification accuracy of the networks compared for the fused images to the non-fused images, both individually and using heterogeneous DCNN fusion. Each network was used to train an SVM using the individual feature vectors as well as an SVM trained using the heterogeneously fused feature vector. The four fusion methods used were a summation, average, maximum, and minimum of the three individual feature vectors, yielding a fused feature vector of the same size containing the combined weights of the three networks. To determine classification accuracy, the main metric collected was a top-1 accuracy for each of the configurations. Overall, the fused image set had a higher classification accuracy across both the individual SVMs as well as the SVM using heterogeneous fusion, with the only notable exception being for the VGG19 SVM, in which case the non-fused image set was actually found to have performed slightly better. For both the case of the fused and non-fused image sets however, the heterogeneously fused SVMs performed as well as if not better than the individual networks, with the fused image set having the best accuracy out of all the tests performed.

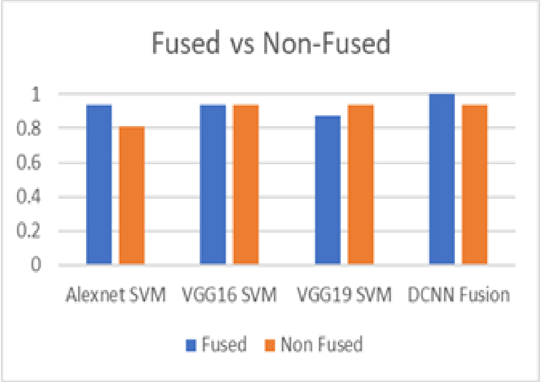


Figure 7: Classification Accuracy, Fused vs Non-Fused

Table 1: Individual SVMs vs DCNN Fusion

Accuracy	Alexnet SVM	VGG16 SVM	VGG19 SVM	DCNN Fusion
Fused	0.9375	0.9375	0.875	**1**
Non Fused	0.8125	0.9375	0.9375	**0.9375**

Each of the four fusion methods were also compared for both the fused and non-fused images, to determine which method of feature vector fusion yielded the most accurate classification.

Table 2: Accuracy of Fusion Functions

Accuracy	Max	Min	Mean	Sum
Fused	1	0.75	0.9375	1
Non Fused	0.875	0.6875	0.9375	0.9375

In general, the best fusion methods for the feature vectors were a summation of the individual vectors as well as the maximum value of each element. Taking the mean of the three feature vectors yielded the same classification accuracy between the fused and non-fused image sets, with the minimum of each element consistently yielding the lowest classification accuracy. Additionally, although the max and sum fusion methods were the best performing for the classification of the fused imagery, for the non-fused image set the mean and sum methods were tied exactly for the best performing, whereas max performed slightly worse. Overall however, the fused image set with the feature vectors fused using the max and summation methods had the best performance out of all the trials, indicating promise for the viability of using these methods in conjunction with image fusion.

V. CONCLUSIONS

Our study was a preliminary test to investigate the advantages image fusion may have in the training performance of neural networks. The results of this investigation found that on average, utilizing image fusion before using images as input for a neural network increased their performance for SVMs trained using individual networks. When using image fusion in conjunction with heterogeneous DCNN fusion, the classification accuracy of the SVM was found to increase the performance even further than the individual networks alone without increasing the computational requirements and decreasing the storage requirements by decreasing the number of images required to train. Future work in this investigation involves expanding our image fusion techniques to include multi-spectral imagery as well as other modalities while also expanding the types of multi-resolution techniques utilized for the actual fusion process, such as fusing images using contourlets, curvelets, bandelets, etc. Additionally, future efforts would involve automating both the image fusion process as well as creating a continuous data stream which can be used to train and update the model in real-time.

ACKNOWLEDGMENT

This work was funded by Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. (625-829-D1441). The authors, therefore, acknowledge with thanks DSR technical and financial support. The authors would like to thank IUP students Derrick Swint, Noah Knepp, Dylan Dewalt, and John Conlen III for their contributions to the research.

REFERENCES

1. A. Torralba, K. Murphy, W. Freeman, and M. Rubin. Context based vision systems for place and object recognition. ICCV. 2003.
2. Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models for practical applications. arXiv:1605.07678.
3. Rawat, Waseem. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation 29.
4. Giansiracusa, M., Adam Lutz, Soundararajan Ezekiel, Mark Alford, Erik Blasch, Adnan Bubalo, and Millicent Thomas. "Multi-focus and multi-modal fusion: a study of multi-resolution transforms." In SPIE Defense+ Security, pp. 98410I-98410I. International Society for Optics and Photonics, 2016.
5. Mike G, Larry Pearlstein, Tyler Daws, Ezekiel S, Alshehri A. A., “A Comparative Study of Multi-Focus, Multi-Resolution Image Fusion Transforms and Methods,” An International Journal of Advanced Computer Technology, 2019.
6. En-Ui Lin, Michael J. McLaughlin, Abdullah Ali Alshehri, “Medical image segmentation using multi-scale and super-resolution method”. Applied Imagery Pattern Recognition Workshop, AIPR 2014, Washington, DC, USA, October 14-16, 2014.
7. Lutz, Adam, Michael Giansiracusa, Neal Messer, Soundararajan Ezekiel, Erik Blasch, and Mark Alford. “Optimal multi-focus contourlet-based image fusion algorithm selection.”2016 Geospatial Informatics, Fusion, and Video Analytics VI, SPIE Defense + Security Conference, IEEE, 2016
8. Omar, Zaid & Stathaki, Tania. (2014). Image Fusion: An Overview. 306- 310. 10.1109/ISMS.2014.58.
9. Stathaki, T., [Image Fusion: Algorithms and Applications], Academic Press, (2008).
10. Gonzalez, R. C., Woods, R., [Digital Image Processing (3rd ed.)], Prentice Hall, (2008).
11. Daubechies, Ingrid. Ten lectures on wavelets. Vol. 61. Philadelphia: Society for industrial and applied mathematics, 1992.
12. Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, “Objective assessment of multi-resolution image fusion algorithms for context enhancement in night vision: A comparative study,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 34, No. 1, pp. 94-109, 2012.
13. Piella, G. “A General Framework for Multiresolution Image Fusion: From Pixles to Regions. Information Fusion, Vol. 4, Issue 4, December 2003.
14. Cvejic, N., Bull, D., & Canagarajah, N. (2007). Region-based multimodal image fusion using ICA bases. *IEEE Sensors Journal*, *7*(5), 743-751.
15. Li, S., & Yang, B. (2010). Hybrid multiresolution method for multisensor multimodal image fusion. *IEEE Sensors Journal*, *10*(9), 1519-1526.
16. Pedregosa, F., “Machine Learning in Python,” Journal of Machine Learning Research, 12: 2825 – 2830, 2011.
17. Wang, Z., Zou, D., Armenakis, C., Li, D., “A Comparative Analysis of Image Fusion Methods,” IEEE Trans. on Geoscience and Remote Sensing, vol. 43, no. 6, (2005).
18. Blasch, E. P., Bossé, E., and Lambert, D. A., [High-Level Information Fusion Management and Systems Design], Artech House, Norwood, MA, (2012).
19. Piella, G. “A General Framework for Multiresolution Image Fusion: From Pixles to Regions. Information Fusion, Vol. 4, Issue 4, December 2003.
20. Tarolli, J. “Multimodal image fusion with SIMS: Preprocessing with image registration”. 14 January 2016.
21. Stathaki, T., [Image Fusion: Algorithms and Applications], Academic Press, (2008).
22. Blasch, E. P., Bossé, E., and Lambert, D. A., [High-Level Information Fusion Management and Systems Design], Artech House, Norwood, MA, (2012).
23. Blasch, E., Steinberg, A., Das, S., Llinas, J., Chong, C.-Y., Kessler, O., Waltz, E., White, F., "Revisiting the JDL model for Information Exploitation," Int’l Conf. on Info Fusion, (2013).
24. Blasch, E., Kadar, I., Hintz, K., Biermann, J., Chong, C-Y., Das, S., “Resource Management Coordination with Level 2/3 Fusion Issues and Challenges,” IEEE Aerospace and Electronic Systems Magazine, Vol. 23, No. 3, pp. 32-46, Mar. (2008).
25. Snidaro, L., Garcia-Herrera, J., Llinas, J., Blasch, E. (eds.), [Context-Enhanced Information Fusion], Springer, (2016).
26. Liang, P., et al., “Encoding Color Information for Visual Tracking: Algorithms and Benchmark,” IEEE Trans. on Image Processing, Vol. 24, No. 12, Dec. 5630-5644, (2015).
27. Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., Srivastava, J. A comparative study of anomaly detection schemes in network intrusion detection. In Proceedings of the 2003 SIAM International Conference on Data Mining (pp. 25-36).
28. Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks. AIChE journal, 37(2), (1991) 233-243.
29.] Kambhatla, N., Leen, T. K. Dimension reduction by local principal component analysis. Neural computation, 9(7), (1997) 1493-1516.
30. Lu, Y., Cohen, I., Zhou, X. S., Tian, Q. Feature selection using principal feature analysis. In Proceedings of the 15th ACM international conference on Multimedia (2007) (pp. 301-304).
31. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36, 4 (1980), 193–202.
32. LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L., et al. Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems. 1990.
33. LeCun, Y., Huang, F., Bottou, L. Learning methods for generic object recognition with invariance to pose and lighting. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004. Volume 2 (2004). IEEE, II–97.

34. Lee, H., Grosse, R., Ranganath, R., Ng, A. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th Annual International Conference on Machine Learning. 2009. ACM, 609–616.
35. Vukotić, V., Raymond, C., & Gravier, G. (2016, October). Multimodal and crossmodal representation learning from textual and visual features with bidirectional deep neural networks for video hyperlinking. In Proceedings of the 2016 ACM workshop on Vision and Language Integration Meets Multimedia Fusion (pp. 37-44). ACM
36. Kornish, D., Ezekiel, S., & Cornacchia, M. (2018, October). Fusion based Heterogeneous Convolutional Neural Networks Architecture. In 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1-6). IEEE.
37. Bodla, N., Zheng, J., Xu, H., Chen, J. C., Castillo, C., & Chellappa, R. (2017, March). Deep heterogeneous feature fusion for template-based face recognition. In 2017 IEEE winter conference on applications of computer vision (WACV) (pp. 586-595). IEEE.
38. Kornish, David. Ezekiel, Soundararajan. DCNN Augmentation via Synthetic Data from Variational Autoencoders and Generative Adversarial Networks. IEEE Applied Image and Pattern Recognition Journal, 2018.

AUTHORS PROFILE

Abdullah A. Alshehri, was born in Saudi Arabia 1964. In 1993 he received his B.S. in Electrical Engineering from University of Detroit, Detroit, MI. USA. He received his M.S. and Ph.D. in Electrical Engineering from University of Pittsburgh, PA in 1999 and 2004 respectively. From 2005 to 2010 he worked as assistant professor in the College of Telecom and Electronics CTE and Jeddah College of Technology JCT, Saudi Arabia. In December 2010, he joined the Electrical Engineering Department at King Abdulaziz University-Rabigh KAU, Saudi Arabia and is now an Associate Professor. His areas of interests are in advanced signal and image processing such as time-frequency, wavelet transform, neural networks, and statistical signal processing. Dr. Alshehri is a member of IEEE since 1992 and a member of the Saudi Engineers Council SEC since 2005. Dr. Alshehri has worked in several research projects at KAU.

Soundararajan Ezekiel received hisM.A and PhD degree from the department of Mathematics, University of Pittsburgh, Pittsburgh, USA. He also received MSc degree in Mathematics at Loyola college, Post Graduate Diploma in Operations Research at Anna University, M.Phil at Madras Christian college in India. He is currently professor in computer science, Indiana University of Pennsylvania, PA. His research Include Image Processing, Signal Processing, Wavelet Analysis, Artificial Intelligence, Machine vision, Deep Learning and Cyber Security. Professor Ezekiel is the recipient of three time SFFP fellow and seven time VFRP fellow.