International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-5, June 2020

Object Detection using Convolutional Neural Network and Extended SURF with FIS

Adila Nuzhat, Fahima Tabassum, Md. Imdadul Islam

Abstract*: The aim of the paper is to detect object using the combination of three algorithms: convolutional neural network (CNN) and extended speeded up robust features (SUFR) and Fuzzy inference system (FIS). Here three types of objects are considered: first, we consider RGB images of hundred different types of objects (for example anchor, laptop airplane, car etc.) taken from benchmark database; second, we take grayscale images of human fingerprint from recognized database; third, Bangla handwritten alphabet from standard database. In this paper we extend the SURF algorithm then the result of the extended SURF is applied in FIS to enhance accuracy of detection. Finally, three algorithms are combined and the accuracy of detection of combined technique is found better than individual one. The combined algorithm provides the average recognition rate for objects of first case as 94.21%, for human finger print as 92.17%%, for Bangla letter as 92.38% and for the Bangla digit as 93.69%.

Keywords*: Accuracy of detection, entropy, confusion matrix, Euclidian distance of feature points and surface plot.

I. INTRODUCTION

In this paper, we used three machine-learning algorithms: CNN, SURF and FIS in object detection. This section will deal with some previous works and applications relevant to above algorithms. Intense research is going on object detection using deep neural network. In [1] authors deal with the impact of padding scheme with AlexNet-based architecture in vehicle detection. The detection results of Car, Motorcycle, Bus and Truck are shown in tabular form. The table includes 'AlexNet scheme', 'same-padding scheme' and valid-padding scheme, where 'same-padding scheme' shows the best result. The paper claims 2.7% improvement is found compared to previous model of mean average precision (MAP) at the 100th epoch. The detection rate of CNN is improved in [2], where pooling block is modified as 'Accept-Reject Pooling' instead of conventional methods. A comparison is made among four pooling techniques: Max Pooling, Stochastic Pooling, Mixed Pooling and Accept-Reject Pooling, where the Accept-Reject Pooling shows the best accuracy of detection. A different application of CNN is found in [3], where 'wireless capsule endoscopy (WCE) images' is used to detect dieses. Authors use conventional architecture of CNN for nine dieses, where the overall sensitivity (SN and specificity (SP) is found as 94.95% and 97.72%. Another application of CNN is found in [4], where six different commands conventionally we perform by hand gestures is detected using electromyographic signals. The confusion matrix shows the overall accuracy of 99%.

The concept of box image and scene image is used in object detection under SURF algorithm, where the box image is searched inside the scene image using same feature points. Such idea is found in [5]. In [6] authors enhanced the capability of SURF algorithm by adding 'segmentation of the image' and 'Simple Linear Iterative Clustering (SLIC) superpixels method' in detection object directly instead of conventional box and scene images. The accuracy of detection of 10 different types of objects are shown in tabular form, where the highest detection rate is found 100% and the lowest one is 33%. Very few works of FIS are found in object detection. In [7] identification of trauma using facial expression of human is done based ANFIS. The performance of the system is determined using, error histogram, error plot, mean absolute error and confusion matrix. The process time and accuracy of detection of the proposed method is found better than previous work. Sometimes FIS is used in disease diagnosis, for example [8] proposed Hierarchical Fuzzy Inference system (HFIS) with fuzzy rules of low complexity to diagnose Dengue fever. The accuracy of the proposed HFIS is found better than previous work. In this paper we combine above three methods in conjunction with proposed algorithm of SURF extension to recognize different types of objects The rest of the paper is organized as: section 2 deals with basic theory of CNN, SURF and FIS along with their integration in object detection, section 3 provides results based on complex integration of algorithm of section 2 and section 4 concludes entire analysis.

II. RESEARCH METHODOLOGY

In this paper, several operations are applied on RGB image to acquire the accuracy of detection of an object. The basic steps of operation is shown in fig.1. The basic operation of SURF, CNN and FIS algorithms are discussed in this section, where the proposed algorithm to 'convert the SURF feature points to a distance vector' is given at the end of this section and the combining scheme using entropy is given at the end of this section.

Revised Manuscript Received on June 15, 2020.
* Correspondence Author
Adila Nuzhat, Computer Science & Engineering, Jahangirnagar University, Savar, Dhaka. Email: tithiadila91@gmail.com
Fahima Tabassum*, Institute of Information Technology, Jahangirnagar University, Savar, Dhaka. Email: fahima@juniv.edu
Md. Imdadul Islam, Computer Science & Engineering, Jahangirnagar University, Savar, Dhaka. Email: imdad@juniv.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

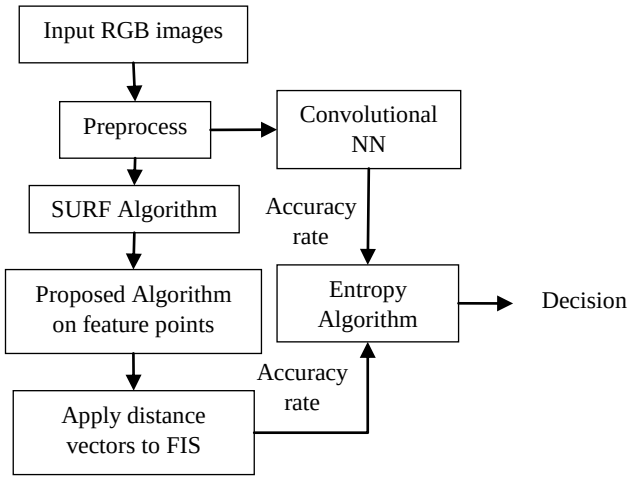


Fig. 1.Proposed model in object detection

2.1 Convolutional neural network

In a conventional neural network of multilayer perceptron (MLP), the signal at the input of a node is the weighted sum of signals of previous layer and the corresponding out is achieved by activation function. In CNN input of a node is the sum of convolution of kernel (like filter of LTI system) and output signal of previous output lines. A Rectified Linear Unit (ReLU) is associated with each node works as an activation function. Next the output of the ReLU is sent to a pooling layer to reduce the size of signal vector or matrix keeping the features of the signal. For example average pooling or max pooling take the average or maximum value of each filtered data. Each convolutional layer of CNN is associated with operation of 'convolution + ReLU + pooling'. The final convolutional layer is connected to fully connected layer, usually one dimensional layer which is connected to output nodes of decision. The detail analysis of CNN is given in [9-10]. In this paper 70% of image from the database is used for training the CNN and the rest for validation. Finally we run the CNN and determine the accuracy of detection using confusion matrix.

2.2 SURF in feature detection

In computer vision, SURF is an algorithm of local feature detection based on Hessian matrix. The applications of SURF lies in object recognition, image classification, registration of image etc. Let us consider a point p = (x, y) of an image I and its Hessian matrix H(p, σ) at point p and scale σ, is expressed as [11],

$$\mathbf{H}(p,\sigma)=\begin{bmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{yx}(p,\sigma) & L_{yy}(p,\sigma) \end{bmatrix} \quad (1)$$

,where $L_{xx}(p,\sigma)=\left\{\frac{\delta^2}{\delta x^2}G(x,y,\sigma)\right\}*I(x,y)$,

$$L_{xy}(p,\sigma)=\left\{\frac{\delta^2}{\delta x\delta y}G(x,y,\sigma)\right\}*I(x,y),$$

$$L_{yx}(p,\sigma)=\left\{\frac{\delta^2}{\delta y\delta x}G(x,y,\sigma)\right\}*I(x,y),$$

$$L_{yy}(p,\sigma)=\left\{\frac{\delta^2}{\delta y^2}G(x,y,\sigma)\right\}*I(x,y) \text{ and } G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Here σ is the standard deviation of Gaussian function, $G(x,y,\sigma)$.

In SURF, the Laplacian-of-Gaussian (LoG) is used in blob detector, which is expressed as [12],

$\nabla^2 L(x, y, \sigma) = L_{xx}(x, y, \sigma)+L_{yy}(x, y, \sigma) = L_{xx}(p, \sigma)+L_{yy}(p, \sigma)$ (2)

In SIFT, we use Difference-of-Gaussian (DoG) function like,

$D(x,y,\sigma)=\{G(x,y,k\sigma)-G(x,y,\sigma)*I(x,y) = L(x,y,k\sigma)-L(x,y,\sigma)$ (3)

The SURF algorithm uses integral image instead of conventional image. In integral image each pixel at (x, y) is the sum of pixels of the rectangle formed between (0, 0) and (x, y) like,

$$I_S=\sum_{i=0}^{i\le x}\sum_{j=0}^{j\le x} I(x,y) \quad (4)$$

Again the difference of Gaussians (DoG) is used in SIFT algorithm. If convolution of an image $I(x, y)$ and Gaussian filter $G(x, y)$ is done then the resultant image $I(x, y)*G(\sigma, x, y)$ is blurred. The blurring of an image occurs the high frequency spatial information of that image is suppressed by the Gaussian filter. If we take the convolution of the image with a different Gaussian filter $G(k\sigma, x, y)$ and subtract them according to equation (5), then we get the image with useful frequency components.

$I(x, y)*G(\sigma, x, y) - I(x, y)*G(k\sigma, x, y)$

$= \{G(\sigma, x, y) - G(k\sigma, x, y)\}* I(x, y)$ (5)

Here $\{G(\sigma, x, y) - G(k\sigma, x, y)\}$ works as a band pass filter, which actually preserve the required frequency components of the image. In practical life, we preserve the helpful frequency components of an image.

Considering two adjacent scales of σ, that are separated by a factor of K we get,

$$\Gamma_{\sigma,K\sigma} = I(x,y)*\frac{1}{2\pi K^2\sigma^2}e^{-\frac{x^2+y^2}{2K^2\sigma^2}} - I(x,y)*\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$$

$$=\left\{\frac{1}{2\pi K^2\sigma^2}e^{-\frac{x^2+y^2}{2K^2\sigma^2}} - \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}\right\}*I(x,y) \quad (6)$$

$$= L(K\sigma,x,y) - L(\sigma,x,y) = D(\sigma,x,y)$$

The typical value of σ =1.6 and $k = \sqrt{2}$.

Now the approximation of, Hessian matrix H(p, σ) at point p and scale σ, is expressed as [13],

$$\mathbf{H}_{appx}(p,\sigma)=\begin{bmatrix} D_{xx}(p,\sigma) & D_{xy}(p,\sigma) \\ D_{yx}(p,\sigma) & D_{yy}(p,\sigma) \end{bmatrix} \quad (7)$$

The determinant of the matrix is,

$$\det(\mathbf{H}_{appx}(p,\sigma)) = D_{xx}^2(p,\sigma) - 0.9D_{xy}^2(p,\sigma) \quad (8)$$

In SIFT, the size of the filter (DoG as a bandpass filter) is fixed and convolved with image of different size (image size is reduced with down sampling as shown in fig. 6 of [14]). In SURF, the size of the box filter (Laplacian of Gausian (LoG)) is varied with up-sampling as in fig. 7 of [14] but the size of image is kept fix as found in fig.2 of [15].

© Copyright: All rights reserved.

The scale space of above figure is divided into several octaves, where an octave represents a series of σ as: σ, $k\sigma$, $k^2\sigma$, $k^3\sigma$, $k^4\sigma$, …, $k^m\sigma$. In a particular octave, the convolution is made using the same input image with a filter of increasing size. In above example an octave (consists of number of scale levels) encompasses a scaling factor of $k = \sqrt{2}$. The DoG under two octaves with five scale levels per octave is shown in fig.1 of [16]. Similar estimation is depicted fig. 4 in [17]. It is obvious from the figure that, at second octave the size of the image is reduce to 1/2 along both x and y direction then scaling $k^n\sigma$ is applied to Gaussian filter to filter the image. Next we have to locate the DoG extrema of SIFT, comparing the pixel marked as X with the pixels of its adjacent scales of fig.7 of [15]. The pixel X is compared with 26 pixels, if it is maxima or minima among surrounding 26 pixels then it is an extrema. The similar operation under LoG of SURF is shown in fig.6 of [15].

Finally, SURF uses two algorithms: (a) determination of orientation of circular region centered at point of interest (b) determination of descriptor vector using the orientation of first algorithm. In this paper, we extract the SURF feature points of the object image and among them we select the four strongest feature points. Those points are then applied on the following proposed algorithm to extract the feature vector, which will be then applied in FIS.

Algorithm

1. Read the image 'I' and make it's preprocess.
2. Apply SURF algorithm on the image and extract location and strength of SURF feature points.
3. Extract four points (x_i, y_i), i =1, 2, 3, 4 of SURF feature with strongest scale from the data of previous step.
4. Select the point (x_m, y_m) of maximum scale.
5. Determine the Euclidian distance of points from (x_m, y_m) as, $S_i = \text{sqrt}[(x_i\text{-}x_m)^2+(y_i\text{-}y_m)^2]$, $i \neq m$
6. The reference point (x_m, y_m) is taken within the image to remove the impact of image rotation, since the Euclidian distance $\text{sqrt}[(x_i\text{-}x_m)^2+(y_i\text{-}y_m)^2]$ will remain fixed irrespective of image rotation.
7. Sort the distances, d_i and normalize the distance as S_i/S_{max}, $i \neq m$. The distance normalization is done to eliminate the impact of size of image.
8. Extract the distance vector, $\mathbf{D} = [S_1\ S_2\ S_3]$ for each image.
9. Apply the distance vector, **D** to FIS to recognize images from both surface plot and Fuzzy MFs.

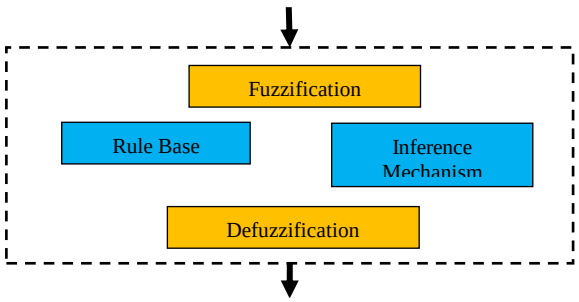


Fig. 2.System block of FIS

2.3 Fuzzy Inference System

Basic building block of FIS is shown in fig.2. Its operation with input numerical data and output numerical data is done on the following four steps as explained in [18].

- First of all the Fuzzification (crisp to linguistic values) is done on the input variables based on input membership function.
- The 'rule base' block of FIS holds the knowledge in the form of a set of rule to convert Fuzzy input to Fuzzy output.
- The 'inference mechanism' block of FIS determines which rules are appropriate for the Fuzzified input data at current time and then decides the Fuzzy output.
- Finally, Defuzzification interface, converts the Fuzzy output into crisp ones.

At this level, the accuracy of detection of CNN and (SURF+FIS) are combined using the rule of entropy like, $H = D_{CNN}*\log_2(1/D_{CNN})+ D_{SURF+FIS}*\log_2(1/D_{SURF+FIS})$, where H is the entropy, D_{CNN} is the accuracy of detection of CNN and $D_{SURF+FIS}$ is that of SURF+FIS. If entropy, $H < 0.65$ and any of the detection probability is greater than 0.75 then we consider the detection is acquired.

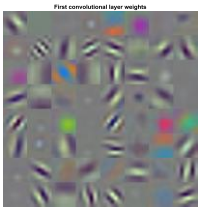


Fig. 3.Weighs of first convolutional layer

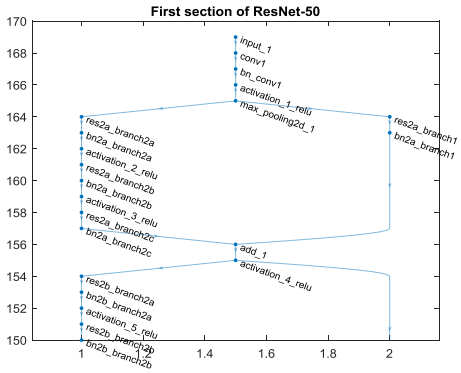


Fig. 4.First section of ResNet-50

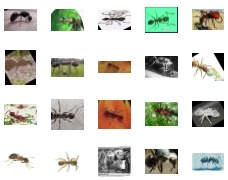
III. RESULTS

This section first deals with results pertinent to object detection using CNN. Here 30% images of the database are taken for training and the rest 70% for test, the size of image at input is 224×224×3 and the size of weight of first convolutional layer is 35×35×3×64 shown in fig.3. We use the CNN of [19], which is achieved using the function resnet50( ) in Matlab 16 the 'first section of network' is shown in fig.4.

© Copyright: All rights reserved.

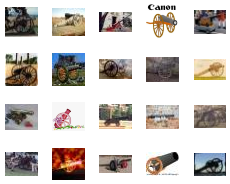
We take the benchmark database of 100 different objects; few of them are shown in fig.5. First, we consider three objects: 'airplanes', 'ferry and, 'laptop', where the folder of airplanes has 800 sample images, the folder of ferry has 67 sample images and that of laptop has 81 images.

(a) Anchor

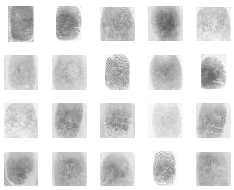


(b) Ant

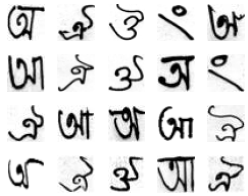
(c) Camera



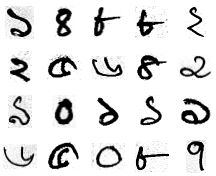
(d) Cannon



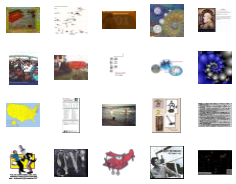
(e) Fingerprint



(f) Bangla font



(g) Bangla digit



(h) Randomly selected object

Fig. 5.Sample objects from benchmark database

The confusion matrix of detection for three, four, five and six objects cases are given in table I to IV, where the accuracy of detection remains almost constant with number of objects. The accuracy starts to decrease when the number of types of object is greater than 12. We found it 87.23% for 20 types of objects, 84.38% for 25 types of objects and 81.13% taking 30 types of objects. Next, we deal with identification of fingerprint, where we use 80 images for each of four persons and the corresponding results are displayed in table V. Next, we have taken 20 image of each latter of Bangla alphabet we get the results of table VI for six letters. We have considered ten Bangla digits for the next level. There is a problem of detection of Bangla "9" and Bangla "1" visualized from above confusion matrix since the both symbols look very similar. The average detection rate is 80.86% as shown in table VII. The profile of accuracy and loss with iteration of CNN is shown in fig. 6 (a) and (b) for the cases of 4 and 10 digits.

Table- I: Accuracy of detection for three objects

Objects	Images	Airplanes	Ferry	Laptop	Average accuracy
Airplanes		1	0	0	99.29%
Ferry		0.0213	0.9787	0	
Laptop		0	0	1	

Table II: Accuracy of detection for four objects

Objects	Images	Airplanes	Ferry	Laptop	Ant	Average Accuracy
Airplanes		1	0	0	0	96.35 %
Ferry		0.0345	0.8966	0	0.0690	
Laptop		0.0213	0	0.9574	0.0213	
Ant		0	0	0	1	

Table III: Accuracy of detection for five objects

Objects	airplanes	ferry	laptop	ant	butterfly	Average Accuracy
airplanes	1	0	0	0	0	96.58%.
ferry	0.034	0.861	0.06	0	0.034	
laptop	0.015	0	0.98	0	0	
ant	0	0	0	1	0	
butterfly	0	0	0.017	0	0.98	

Table IV: Accuracy of detection for six objects

Objects	airplanes	ferry	laptop	ant	butterfly	crab	Average Accuracy
airplanes	1	0	0	0	0	0	97.01%.
ferry	0.0345	0.896	0.069	0	0	0	
laptop	0	0.015	0.984	0	0	0	
ant	0	0.019	0.019	0.960	0	0	
butterfly	0	0	0	0	0.9787	0.0213	
crab	0	0	0	0	0	1	

Table V: Accuracy of detection of fingerprint

Objects	Person-1	Person-2	Person-3	Person-4	Average accuracy
Person-1	1	0	0	0	96.43%.
Person-2	0.0179	0.8929	0	0.0893	
Person-3	0	0	0.9643	0.0357	
Person-4	0	0	0	1	

© Copyright: All rights reserved.

Table VI: Accuracy of detection of Bangla font

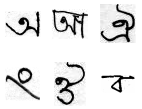
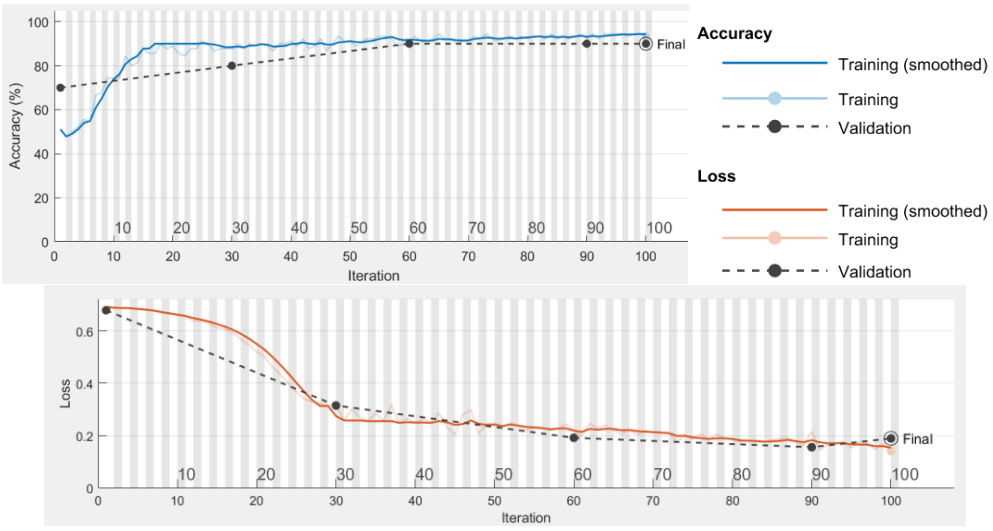
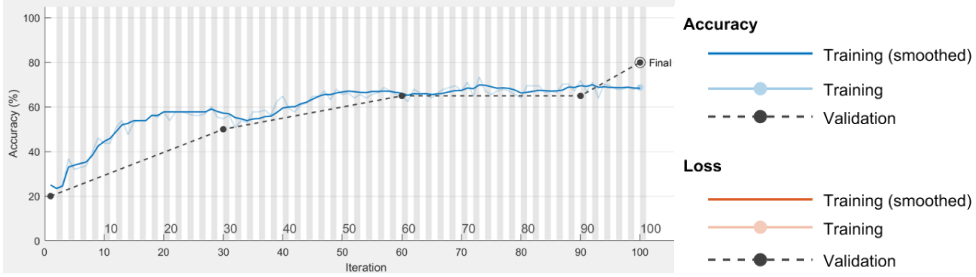
Objects	Images	A	AA	OI	ANUSAR	OU	BA	Average accuracy
A		0.8571	0 0.1429	0	0	0	0	83.33%.
AA		0.5714	0.4286	0	0	0	0	
OI		0	0	0.7857	0	0	0	
ANUSAR		0	0	1	0	0	0	
OU		0	0	0.0714	0	0.9286	0	
BA		0	0	0	0	0	1	

Table VII: Accuracy of detection of Bangla digits

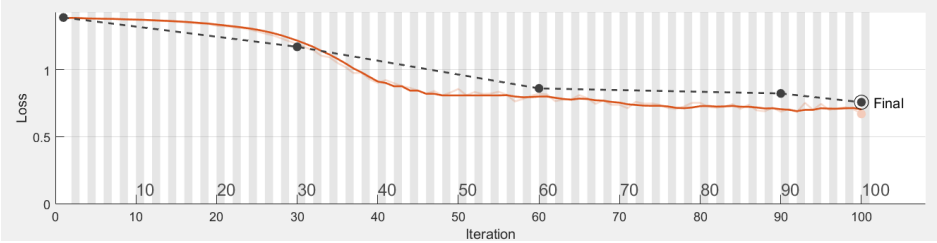
Objects	Images	0	1	2	3	4	5	6	7	8	9	Average Accuracy
0		0.9714	0	0	0	0	0.0286	0	0	0	0	80.86%.
1		0.0286	0.2857	0.2286	0.1714	0.1143	0.0286	0	0.0286	0	0.1143	
2		0	0	0.9143	0.0571	0	0	0	0	0.0286	0	
3		0	0	0	0.8286	0	0.1714	0	0	0	0	
4		0	0	0	0	1	0	0	0	0	0	
5		0.0286	0	0	0.0286	0	0.8857	0.0571	0	0	0	
6		0	0	0.0286	0	0	0	0.9714	0	0	0	
7		0.0571			0.0857	0.0286	0	0.0286	0.7714	0.0286	0	
8		0	0	0.0286	0	0	0	0	0	0.9714	0	
9		0	0.0571	0.2286	0.2000	0	0	0.0286	0	0	0.4857	



(a) Taking 4 digits (90.34% accuracy)



© Copyright: All rights reserved.



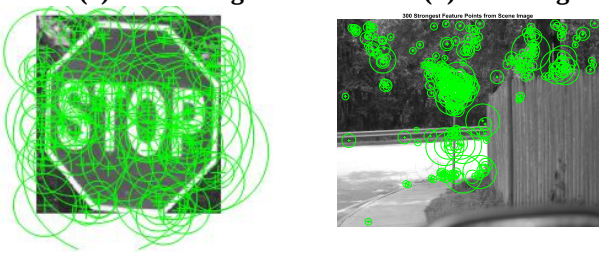
(b) Taking 10 digits (80.32% accuracy)

Fig. 6.Results of CNN in Bangla digit recognition

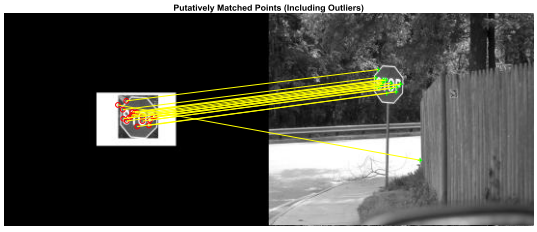
SURF features are mostly used in object detection called box image from a big image called scene image. For example, the road symbol is a box image and the scene of the including road sign is scene image shown in fig. 7(a) and (b). The corresponding SURF features and the mapping of box scene image against the same feature points is shown in fig. 7(c) to (d). Similar operation on a showpiece at a house is shown in fig.8.

(a) Scene image

(b) Box image



c) Strongest feature points of box image and scene image

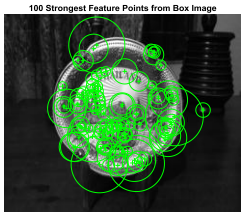


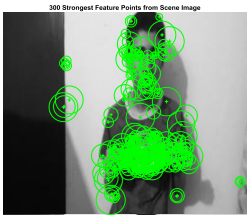
(d) Mapping of corresponding feature points

Fig. 7.SURF features in road sign detection

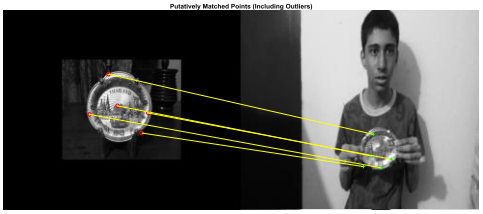
(a)Scene image

(b) Box image





(c) Strongest feature points of box and scene images



e) Mapping of corresponding feature points

Fig. 8.SURF features in showpiece detection

Applying the SURF on preprocessed image, considering five feature points of in descending order, we get four Euclidian distances taking the point of highest scale as the reference. After normalization of distance we get the results as shown in table VIII for eight images of airplane. Table IX and X shows the similar results for Butterfly and Gramophone. Few image with five feature points of highest scale is shown in fig.9 correspond to table VIII, IX and X. The variable S_4 correspond to the point of highest scale hence dividing each distance by S_4 we get the numerical value 1 for S_4. Applying three variables S_1, S_2 and S_3 to FIS we get the surface plot like fig.10. We verify the Fuzzy rules taking three vectors randomly, [0.2034; 0.3924; 0.5801], [0.2298; 0.2588; 0.9472] and [0.0917; 0.136; 0.3175] from table VIII, IX and X respectively and got the output 1, 2 and 3 correspond to image of Airplane, Butterfly and Gramophone respectively shown in fig.11. Few fuzzy rules used in FIS are shown in fig.12.

Table VIII: Normalized Distance of SURF feature point for Airplane

S_1	S_2	S_3	S_4	Output
0.0261	0.9140	0.9833	1	Airplane
0.5162	0.5438	0.5848	1	
0.2034	0.3924	0.5801	1	
0.0488	0.3950	0.7057	1	
0.0995	0.1533	0.7180	1	

© Copyright: All rights reserved.

0.0677	0.2396	0.6228	1	
0.0524	0.5344	0.5696	1	
0.3762	0.8434	0.9297	1	

Table IX: Normalized Distance of SURF feature point for Butterfly

S_1	S_2	S_3	S_4	Output
0.0721	0.1223	0.9290	1	Butterfly
0.1219	0.9177	0.9866	1	
0.3812	0.5735	0.8839	1	
0.2298	0.2588	0.9472	1	
0.0544	0.6626	0.9928	1	
0.0374	0.4839	0.7803	1	
0.1580	0.2819	0.4499	1	
0.1211	0.1819	0.2165	1	

Table X: Normalized Distance of SURF feature point for Gramophone

S_1	S_2	S_3	S_4	Output
0.1054	0.1812	0.4067	1	Gramophone
0.0917	0.1450	0.1668	1	
0.4127	0.5059	0.6762	1	
0.1159	0.3692	0.8010	1	
0.0917	0.1360	0.3175	1	
0.1492	0.7173	0.7484	1	
0.1107	0.2118	0.5391	1	
0.8418	0.9181	0.9335	1	

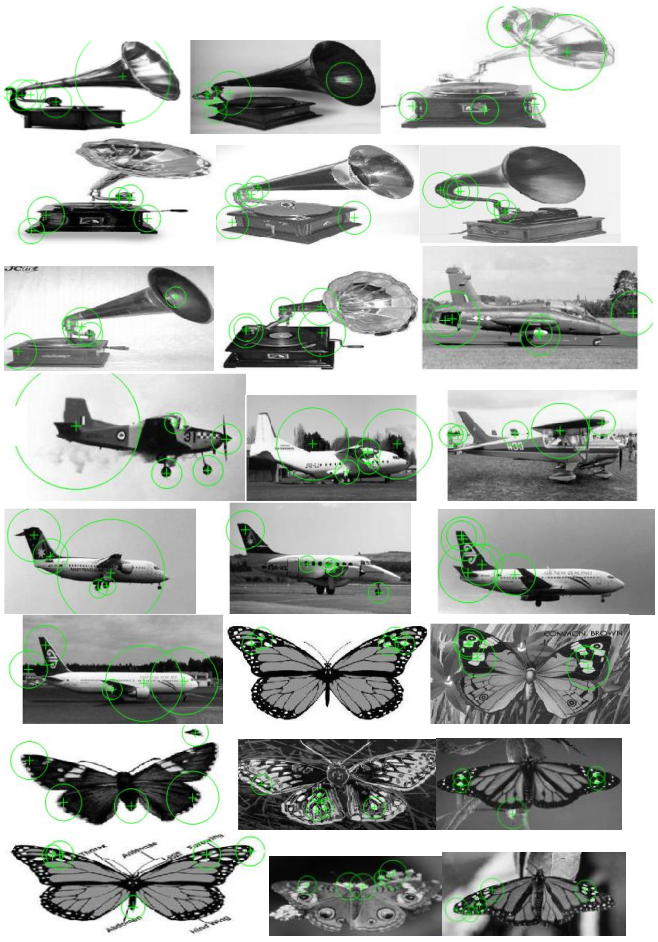
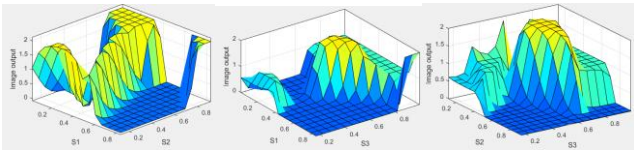
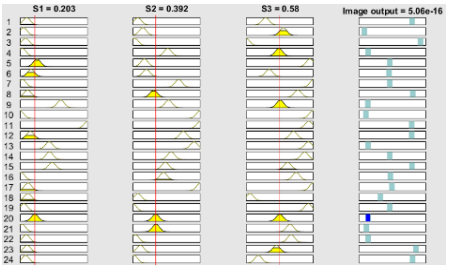


Fig. 9.Some images with five feature points of highest scale

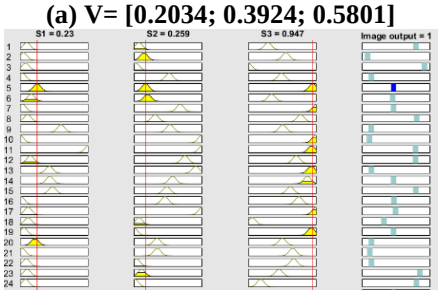


(a) (b) (c)

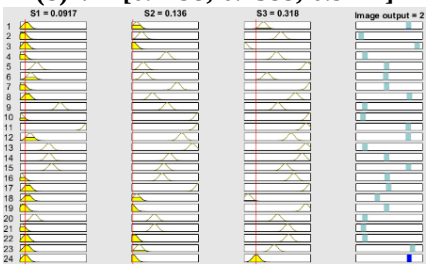
Fig. 10. Surface plot of FIS considering three variables



(a) V= [0.2034; 0.3924; 0.5801]



(b) V = [0.2298; 0.2588; 0.9472]



(c) V = [0.0917; 0.136; 0.3175]

Fig. 11. Surface plot of FIS considering three variables

1. If (S1 is in1cluster1) and (S2 is in2cluster1) and (S3 is in3cluster1) then (Image output is out1cluster1) (1)
2. If (S1 is in1cluster2) and (S2 is in2cluster2) and (S3 is in3cluster2) then (Image output is out1cluster2) (1)
3. If (S1 is in1cluster3) and (S2 is in2cluster3) and (S3 is in3cluster3) then (Image output is out1cluster3) (1)
4. If (S1 is in1cluster4) and (S2 is in2cluster4) and (S3 is in3cluster4) then (Image output is out1cluster4) (1)
5. If (S1 is in1cluster5) and (S2 is in2cluster5) and (S3 is in3cluster5) then (Image output is out1cluster5) (1)
6. If (S1 is in1cluster6) and (S2 is in2cluster6) and (S3 is in3cluster6) then (Image output is out1cluster6) (1)
7. If (S1 is in1cluster7) and (S2 is in2cluster7) and (S3 is in3cluster7) then (Image output is out1cluster7) (1)
8. If (S1 is in1cluster8) and (S2 is in2cluster8) and (S3 is in3cluster8) then (Image output is out1cluster8) (1)
9. If (S1 is in1cluster9) and (S2 is in2cluster9) and (S3 is in3cluster9) then (Image output is out1cluster9) (1)
10. If (S1 is in1cluster10) and (S2 is in2cluster10) and (S3 is in3cluster10) then (Image output is out1cluster10) (1)
11. If (S1 is in1cluster11) and (S2 is in2cluster11) and (S3 is in3cluster11) then (Image output is out1cluster11) (1)
12. If (S1 is in1cluster12) and (S2 is in2cluster12) and (S3 is in3cluster12) then (Image output is out1cluster12) (1)
13. If (S1 is in1cluster13) and (S2 is in2cluster13) and (S3 is in3cluster13) then (Image output is out1cluster13) (1)
14. If (S1 is in1cluster14) and (S2 is in2cluster14) and (S3 is in3cluster14) then (Image output is out1cluster14) (1)
15. If (S1 is in1cluster15) and (S2 is in2cluster15) and (S3 is in3cluster15) then (Image output is out1cluster15) (1)
16. If (S1 is in1cluster16) and (S2 is in2cluster16) and (S3 is in3cluster16) then (Image output is out1cluster16) (1)
17. If (S1 is in1cluster17) and (S2 is in2cluster17) and (S3 is in3cluster17) then (Image output is out1cluster17) (1)
18. If (S1 is in1cluster18) and (S2 is in2cluster18) and (S3 is in3cluster18) then (Image output is out1cluster18) (1)
19. If (S1 is in1cluster19) and (S2 is in2cluster19) and (S3 is in3cluster19) then (Image output is out1cluster19) (1)
20. If (S1 is in1cluster20) and (S2 is in2cluster20) and (S3 is in3cluster20) then (Image output is out1cluster20) (1)
21. If (S1 is in1cluster21) and (S2 is in2cluster21) and (S3 is in3cluster21) then (Image output is out1cluster21) (1)
22. If (S1 is in1cluster22) and (S2 is in2cluster22) and (S3 is in3cluster22) then (Image output is out1cluster22) (1)
23. If (S1 is in1cluster23) and (S2 is in2cluster23) and (S3 is in3cluster23) then (Image output is out1cluster23) (1)
24. If (S1 is in1cluster24) and (S2 is in2cluster24) and (S3 is in3cluster24) then (Image output is out1cluster24) (1)

Fig. 12. Few Fuzzy rules in FIS

Table XI: Comparison of accuracy of detection

Type of objects	CNN	SURF+FIS	CNN+ SURF+FIS
General object	89.27	82.46	94.21
Fingerprint	84.29	80.15	92.17

© Copyright: All rights reserved.

Bangla letter	86.78	79.68	92.38
Bangla digit	87.13	80.18	93.69

IV. CONCLUSION

In this paper we detect three different type of objects using combination of CNN, SURF and FIS, where we get better results in combined method compared to individual technique. Still we have the scope to apply feature vectors of PCA, LDA and DWT to FIS to detect objects in a robust way. The concept of the paper is applicable in human face detection and in diagnosis technique using medical images.

REFERENCES

1. Aldi Wiranata, Suryo Adhi Wibowo, Raditiana Patmasari, Rissa Rahmania, and Ratna Mayasari, 'Investigation of Padding Schemes for Faster R-CNN on Vehicle Detection,' The 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), pp. 208-212, 5-7 Dec. 2018, Bandung, Indonesia
2. Aiza M. Romano and Alexander A. Hernandez, 'An Improved Pooling Scheme for Convolutional Neural Networks,' 2019 The 7th International Conference on Information, Communication and Networks, pp.201-206, 24-26 April 2019, Macao, Macao
3. Vinu Sankar Sadasivan and Chandra Sekhar Seelamantula, 'High Accuracy Patch-Level Classification of Wireless Capsule Endoscopy Images Using a Convolutional Neural Network,' 2019 IEEE 16th International Symposium on Biomedical Imaging, pp.96-99, 8-11 April 2019, Venice, Italy
4. Javier Orlando Pinzón-Arenas, Robinson Jiménez-Moreno, Julian Esteban Herrera-Benavides, 'Convolutional Neural Network for Hand Gesture Recognition using 8 different EMG Signals,' 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 24-26 April 2019, Bucaramanga, Colombia
5. Tawfiq A. Al-asadi and Ahmed J. Obaid, 'Object detection and recognition by using enhanced Speeded Up Robust Feature,' IJCSNS International Journal of Computer Science and Network Security, vol.16, no.4, pp.66-71, April 2016
6. Miriam Lopez-de-la-Calleja, Takayuki Nagai, Muhammad Attamimi, Mariko Nakano-Miyatake, Hector Perez-Meana, 'Object Detection Using SURF and Superpixels,' Journal of Software Engineering and Applications, vol. 6, no.9, pp. 511-518, Sept' 2013
7. R. Sofia and D. Sivakumar, 'Developing a System for Trauma Identification based on the Difference from the Normal Human Emotion with Adaptive Neuro Fuzzy Inference System,' International Conference on Communication and Signal Processing, pp.672-678, April 4-6, 2019, Chennai, India
8. Mubarak Alrashoud, 'Hierarchical Fuzzy Inference System for Diagnosing Dengue Disease,' 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 31-36, 8-12 July 2019, Shanghai, China
9. Ngoc Giang Nguyen, Vu Anh Tran, Duc Luu Ngo, Dau Phan, Favorisen Rosyking Lumbanraja, Mohammad Reza Faisal, Bahriddin Abapihi, Mamoru Kubo, Kenji Satou, 'DNA Sequence Classification by Convolutional Neural Network,' J. Biomedical Science and Engineering, Volume 9, Issue 5, pp.280-286m April 2016
10. Syed Mazhar Abbas and Shailendra Narayan Singh, 'Region-based Object Detection and Classification using Faster R-CNN,' International Conference on "Computational Intelligence and Communication Technology" (CICT 2018), 9-10 Feb. 2018, Ghaziabad, India
11. S. J. Chen, S. Z. Zheng , Z. G. Xu , C. C. Guo , X. L. Ma, 'An Improved Image Matching Method Based on Surf Algorithm' International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences . 2018, Vol. 42 Issue 3, p179-184, Beijing, China
12. P M Panchal , S R Panchal , S K Shah, 'A Comparison of SIFT and SURF,' International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 2, pp.323-327,April 2013
13. Weiyi WEI , Chengfeng A, Yufei ZHAO, Guicang ZHANG, 'Image Registration Algorithm Based on Super pixel Segmentation and SURF Feature Points,' 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), 7-10 Nov. 2018, Xi'an, China
14. Darshana Mistry, Asim Banerjee, 'Comparison of Feature Detection and Matching Approaches: SIFT and SURF,' GRD Journals- Global Research and Development Journal for Engineering, Volume 2, Issue 4, pp. 7-13, March 2017.
15. M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, 'Image Features Detection, Description and Matching,' Chapter 2, Springer International Publishing, pp.11-45, February 2016
16. David G. Lowe, 'Distinctive image features from scale-invariant keypoints,' International Journal of Computer Vision, vol. 60, Issue 2, pp.91–110, January 2004
17. Kishore Kumar Kamarajugadda, Trinatha Rao Polipalli, 'Age-invariant face recognition using multiple descriptors along with modified dimensionality reduction approach,' Multimedia Tools and Applications, Volume 78, Issue 19, pp 27639–27661, October 2019.
18. Vandna Kamboj, Amrit Kaur, 'Comparison of Constant SUGENO-Type and MAMDANI-Type Fuzzy Inference System for Load Sensor,' International Journal of Soft Computing and Engineering (IJSCE), Volume-3, Issue-2, pp.204-207, May 2013
19. Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, 'Deep Residual Learning for Image Recognition,' 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 27-30 June 2016, Las Vegas, NV, USA

AUTHORS PROFILE

Adila Nuzhat is an M.Sc student of Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh. She also completed her B.Sc. from the same department. She is now pursuing her M.Sc. thesis work in Convolutional Neural Network (CNN) and extended SURF with FIS. Her research fields are digital image processing, artificial intelligence and machine learning.

FahimaTabassum has completed her B.Sc. (Hons.) from the department of Computer Science and Engineering, Jahangirnagar University,Savar, Dhaka, Bangladesh in 2003 and M.S from the same department in 2010. She currently is pursuing her Ph.D degree at the same department. She has a number of publications in different reputed journals. She is also working as an associate professor at the Institute of Information Technology, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh. E-mail: fahima@juniv.edu.

Md. Imdadul Islam has completed his B.Sc. and M.Sc Engineering in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 1993 and 1998 respectively and has completed his Ph.D degree from the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh in the field of network traffic in 2010. He is now working as a Professor at the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. Previously, he worked as an Assistant Engineer in Sheba Telecom (Pvt.) LTD (A joint venture company between Bangladesh and Malaysia, for Mobile cellular and WLL), from Sept.1994 to July 1996. Dr. Islam has a very good field experience in installation and design of mobile cellular network, Radio Base Stations and Switching Centers for both mobile and WLL. His research field is network traffic, wireless communications, wavelet transform, OFDMA, WCDMA, adaptive filter theory, array antenna systems, ANFIS and machine learning. He has more than hundred and eighty research papers in national and international journals and conference proceedings.

© Copyright: All rights reserved.