A Neural Network Classifier for Occluded Images

T. Kurita[†] T. Takahashi[††] Y. Ikeda[†††]
[†] National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, 305-8568 JAPAN
[††] Ryukoku Univ., 1-5 Yokotani, Seta Oe-cho, Otsu, Shiga, 520–2194 JAPAN
[†††] NS Solutions, 20-15, Shinkawa 2-chome, Chuo-ku, Tokyo, 104-8280 JAPAN

Abstract

This paper proposes a neural network classifier which can automatically detect the occluded regions in the given image and replace that regions with the estimated values. An auto-associative memory is used to detect outliers such as pixels in the occluded regions. Certainties of each pixels are estimated by comparing the input pixels with the outputs of the auto-associative memory. The input values to the associative memory are replaced with the new values which are defined depending on the certainties. By repeating this process, we can get an image in which the pixel values of the occluded regions are replaced with the estimates. The proposed classifier is designed by integrating this associative memory with a simple classifier.

1 Introduction

The recognition target in a given image is often occluded by uninteresting objects when the image is taken in general situation (e.g. sunglasses on human faces). If the recognition system can automatically detect the occluded regions using the memory and replace that regions with the estimated information recalled from the memory, it is expected that the recognition ability of the system can be improved and the applicability of the system will be extended.

Kohonen showed that linear auto-associative memory can roughly recall the original image from the partly occluded image[3]. The similar auto-associative memory can be implemented by using Principal Component Analysis (PCA) or Multi-Layer Perceptron (MLP) which has the same number of the input and output units and is trained to map each input vector onto itself[4, 5].

It is well known that strong backward neural connections exist along the visual pathway in the cortex [1]. In the primate cerebral cortex, visual information from the eyes is processed in the primary visual cortex (V1) and the output from V1 is further processed in successive areas along the visual pathways. The forward connection from one area to another is always accompanied by a reciprocal backward connection. Okajima [2] showed by the simulation experiments that the system with backward connections can separate an object pattern from the background in the given input image.

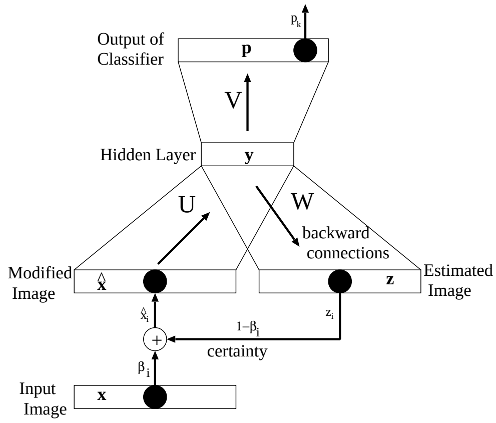


Figure 1. Network architecture of the classifier

This paper proposes a neural network classifier which can automatically detect the occluded regions in the given image and replace that regions with the estimated values. Figure 1 shows the network architecture of the proposed classifier. The network has an auto-associative network with backward connections shown in the bottom half of the figure 1. This auto-associative network is used to estimate the original pixel values of the input image and to replace the input with the estimated values when the differences are large. By repeating this modification process, a good approximation of the original image can be recalled and the good recognition results can be obtained.

Section 2 describes how to use the auto-associative network for detecting the occluded regions and replacing that regions with the estimated values[6]. Then the classifier with an auto-associative memory is explained in section 3.

2 Auto-associative Network

This section introduces an auto-associative memory which can recall its original image from an occluded image. The auto-associative memory is implemented by using a Multi-Layer Perceptron (MLP) with the same number of the input and output units. By adding the mechanism to estimate certainties of the input pixel values and to replace the input values with the estimates depending on the certainties, the original image can be iteratively estimated from the occluded image.

2.1 MLP for auto-associative memory

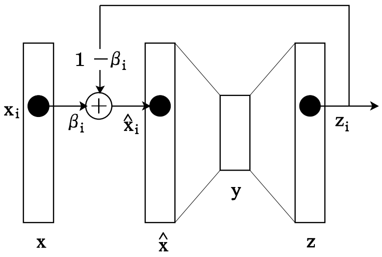


Figure 2. MLP for auto-associative memory.

Consider a MLP with only one hidden layer as shown in figure 2. Let the number of the input and output units and the number of hidden units be M and $H(< M)$. Assume the activation function of the neurons in the hidden and the output layers is linear. It is known that the network performs a projection onto the H-dimensional sub-space which is spanned by the first H principal components of the training data when the network is trained such that the network maps each input vector onto itself[4].

Let the training data be $\{\boldsymbol{x}_j = (x_{j1}, \ldots, x_{jM})^T \in \boldsymbol{R}^M\}_{j=1}^N$. Denote the output of the hidden units and the output of the output units for the input vector $\boldsymbol{x}_j$ as $\boldsymbol{y}_j = (y_{j1}, \ldots, y_{jH})^T \in \boldsymbol{R}^H$ and $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jN})^T \in \boldsymbol{R}^M$. Also the weights from the input to the hidden and from the hidden to the output are denoted as $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_H]$ and $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M]$. Then the network computes the outputs as

$$\begin{array}{rcll} y_{jh} & = & \boldsymbol{u}_h^T \boldsymbol{x}_j & (h = 1, \ldots, H) \\ z_{jm} & = & \boldsymbol{w}_m^T \boldsymbol{y}_j & (m = 1, \ldots, M). \end{array} \qquad (1)$$

The network can be trained by minimizing a sum-of squares error

$$\varepsilon^2 = \frac{1}{2}\sum_{j=1}^{N} \varepsilon_j^2 = \frac{1}{2}\sum_{j=1}^{N} \|\boldsymbol{x}_j - \boldsymbol{z}_j\|^2. \qquad (2)$$

From the viewpoints of the maximum likelihood estimation, this is equivalent to considering a probabilistic model of errors ε_j^2 as Gaussian with zero mean. For the consistency with the training of the classifier, here we use the evaluation criterion

$$l_1 = -\varepsilon^2 = -\frac{1}{2}\sum_{j=1}^{N} \varepsilon_j^2. \qquad (3)$$

By taking the partial derivatives of this evaluation criterion, we can obtain the learning rule as

$$\begin{array}{rcl} \Delta w_{mh} & = & \alpha \sum_{j=1}^{N} (x_{jm} - z_{jm}) y_{jh} \\ \Delta u_{hn} & = & \alpha \sum_{j=1}^{N} \sum_{m=1}^{M} (x_{jm} - z_{jm}) w_{mh} x_{jn}, \end{array} \qquad (4)$$

where α is the learning rate.

2.2 Recall of the original image from the occluded image

The auto-associative network described in the previous section has ability to retrieve an approximation of the original image from the image with some noises or occlude regions. However the retrieved image may be influenced by such noises or occlusions. To improve the robustness of the auto-associative memory to the occlusions, a certainty measure of each pixel is introduced by evaluating the difference between the pixel values of the input image and the retrieved image. The pixel values are modified by using the certainty of each pixel and this process of the retrieval and the modification is repeated several times. The scheme is shown in Figure 2. The concrete process is summarized as follows;

STEP 0: Initialize the iteration parameter t as $t = 0$ and assign the input image $\boldsymbol{x}$ to the input vector of the auto-associative memory $\hat{\boldsymbol{x}}(0)$.

STEP 1: Recall the output $\boldsymbol{z}(t)$ of the auto-associative memory from the input $\hat{\boldsymbol{x}}(t)$.

STEP 2: Compute the pixel-wise differences $\varepsilon_i^2(t) = (x_i - z_i(t))^2$ between the input image $\boldsymbol{x}$ and the retrieved image $\boldsymbol{z}(t)$. Compute the “certainty” $\beta_i(t)$ of each pixel by using these differences as

$$\beta_i(t) = \exp\left(-\frac{\varepsilon_i^2(t)}{2\,\sigma^2(t)}\right), \qquad (5)$$

where $\sigma(t)$ is the robust estimation of the standard deviations of the differences $\varepsilon_i(t)$ [7] and is obtained by

$$\sigma(t) = 1.4826\left(1 + \frac{5}{N-1}\right) \operatorname*{med}_i \sqrt{\varepsilon_i^2(t)}. \qquad (6)$$

Here $\mathrm{med}(x)$ denotes the median of the x.

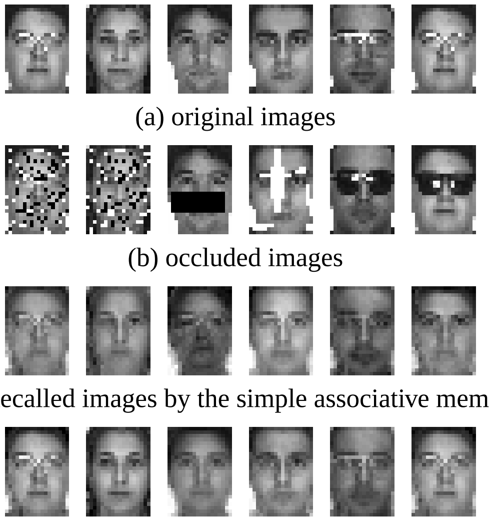
STEP 3: Compute the new input $\hat{\boldsymbol{x}}(t+1)$ of the auto-associative memory by using the "certainty" of each pixel as

$$\hat{x}_i(t+1) = \beta_i(t)x_i + (1-\beta_i(t))\, z_i(t). \qquad (7)$$

Set the iteration parameter as $t \leftarrow t+1$ and go to STEP 1 to repeat the modification process until the number of iterations is less than the specified value.

A good approximation of the original image can be recalled by repeating this modification process several times. The equation (7) means that the pixel value of the input image is trustfully used at the pixels with high certainty while the estimated value is used at the pixels with low certainty. Computation of the STEP 2 is similar to the robust template matching[8] which can automatically remove occluded regions as outliers and computes the correlation of inliers.

2.3 Recall experiments



(a) original images

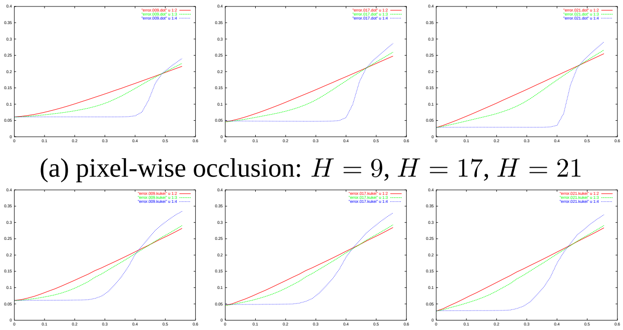
(b) occluded images

(c) Recalled images by the simple associative memory

(d) Recalled images by the proposed associative memory

Figure 3. Recall from the occluded images.

To confirm the effectiveness of the proposed auto-associative memory, we have performed experiments using face images of 31 persons in the ARFace Database[9]. The size of each image is normalized to $18 \times 25 (= 450)$. Examples of the original face images are shown in Figure 3 (a). The MLP was trained using these face images. Figure 3 (b) shows examples of the occluded images. Examples of the recalled images are shown in Figure 3 (c) and (d). In this case, the number of hidden units was set to 17. It is noticed that the results of the simple auto-associative memory without iterations are improved by the proposed method. Figure 4 shows the sum of squares error between the original image and the recalled images. The results with the different number of hidden units ($H = 9$, $H = 17$, and $H = 21$) are shown. The horizontal axis shows the percentage of the occluded regions. The upper curve represents the error of the simple associative memory, while the lower curve represents that of the proposed associative memory. The middle curve shows the results obtained by applying only one iteration. It is noticed that the robustness to the occlusions is improved by the proposed method. By using the modification process, the original image can be recalled even if about 30% of the image is occluded. The tendency of the robustness of the proposed method is same while the sum of squared errors is improved by increasing the number of hidden units.



(a) pixel-wise occlusion: $H = 9$, $H = 17$, $H = 21$

(b) rectangular occlusion: $H = 9$, $H = 17$, $H = 21$

Figure 4. Sum of squares error between the original images and the recalled images

3 Robust classifier to the occlusions

In the applications such as face recognition or face detection, observed faces are sometimes partly occluded by sunglasses, hands, and so on. To make such recognition systems widely applicable, the recognition system should have the ability to automatically detect the occluded regions in the given face image. In this paper, auto-associative memory introduced in the previous section is integrated into a simple classifier. Figure 1 shows the network architecture of the proposed classifier. The auto-associative network is integrated in the first half of the classifier in which the hidden units are shared with the auto-associative network.

3.1 Classifier and learning algorithm

In this paper, we use multinomial logit model[10] as the classifier. Multinomial logit model is a special case of the generalized linear model[10], and it can be regarded as one of the simplest neural network model for multi-way classification problems. As shown in Figure 1, the connections from the input layer to the hidden layer are shared between the classifier and the auto-associative memory. Thus the number of hidden units is equal to H.

Consider a classification problem with K classes $\{C_1, \ldots, C_K\}$. Let $\boldsymbol{t} = (t_1, \ldots, t_K)^T \in \{0,1\}^K$ denote a binary vector composed of teacher signals with $t_k = 1$ if the input is C_k, otherwise $t_k = 0$. The kth output of the classifier p_{jk} is computed as the "softmax" of the weighted sum of the hidden units $\eta_{jk} = \boldsymbol{v}_k^T \boldsymbol{y}_j, (k = 1, \dots, K-1)$ as

$$\begin{aligned} p_{jk} &= \frac{\exp(\eta_{jk})}{1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})} && (8) \\ p_{jK} &= \frac{1}{1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})}. && (9) \end{aligned}$$

For the training samples $\{(\boldsymbol{x}_j, \boldsymbol{t}_j)\}_{j=1}^N$, the log-likelihood of the classifier is given by

$$l_2 = \sum_{j=1}^{N} \sum_{k=1}^{K-1} t_{jk}\eta_{jk} - \sum_{j=1}^{N} \log\left(1 + \sum_{i=1}^{K-1} \exp(\eta_{ji})\right) \quad (10)$$

By combining this with the evaluation criterion of the auto-associative network, we have a evaluation criterion as

$$l = l_2 + \lambda l_1, \quad (11)$$

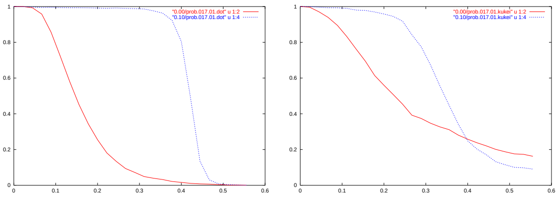
where λ is the parameter to adjust the weighting of these two criteria.

The learning rule for the proposed classifier can be obtained as

$$\begin{aligned} \Delta v_{kh} &= \alpha \sum_{j=1}^{N} (t_{jk} - p_{jk})\, y_{jh} \\ \Delta w_{mh} &= \alpha\lambda \sum_{j=1}^{N} (x_{jm} - z_{jm})\, y_{jh} \\ \Delta u_{hn} &= \alpha \sum_{j=1}^{N} \sum_{k=1}^{K-1} (t_{jk} - p_{jk})\, v_{kh} x_{jn} \\ & \quad + \alpha\lambda \sum_{j=1}^{N} \sum_{m=1}^{M} (x_{jm} - z_{jm})\, w_{mh} x_{jn}. \quad (12) \end{aligned}$$

3.2 Recognition Experiment

Face recognition experiments were performed using the same face database. The number of hidden units was again set to 17 and the 124 images (4 images $\times$ 31 persons) were learned and the recognition rates to the occluded face images were evaluated. Figure 5 shows the relation between the recognition rates and the percentages of the occlusions. The horizontal axis shows the percentage of the occluded regions. The upper curve represents the recognition rates of the proposed classifier while the lower curve represents that of the multi-logit classifier, namely the classifier without auto-associative memory. It is noticed that the proposed method can recognize the face even if the percentages of the occluded regions are about 25% for the case of rectangular occlusions. On the other hand, in the case of the classifiers without auto-associative memory, the recognition rate is gradually decreased as the percentage of the occlusions increases. The proposed method was achieved 90.32% of the recognition rates for the face images with sunglasses, while the recognition rate was 77.41% by the classifier without auto-associative memory. These results shows the effectiveness of the proposed approach to the recognition of occluded images.



(a) pixel-wise occlusion, (b) rectangular occlusion

Figure 5. Relation between the recognition rates and the percentages of the occlusions.

Acknowledgement

This work was partly supported by Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government. Takashi Takahashi was also partly supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

References

[1] S.E.Palmer, *Vision Science: Photons to Phenomenology*, The MIT Press, 1999.

[2] K.Okajima, "A recurrent system incorporating characteristics of the visual system: a model for the function of backward neural connections in the visual system," Biological Cybernetics, Vol.65, pp.234-241, 1991.

[3] T. Kohonen, *Self-Organization and Associative Memory*, Third Edition, Springer-Verlag, Berlin, 1989.

[4] P. Baldi and K. Hornik, "Neural networks and principal component analysis," *Neural Networks*, vol.2, pp.53-58, 1989

[5] C.M.Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press, 1995.

[6] T.Takahashi, Y.Ikeda, T.Mishima, T.Kurita, "Robust data restoration by using autoassociative MLP," Proc. of Annual Conference of Japanese Neural Network Society, pp.163-164, 2001.

[7] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981

[8] T.Kurita,"Robust template matching and its application to cut detection," Proc. of the 1997 IEICE General Conference, D-12-61, pp.268, 1997.

[9] A.M. Martinez and R. Benavente, "The AR Face Database," *CVC Technical Report*, No.24, June 1998.

[10] P.McCullaph, and J.A.Nelder, *Generalized Linear Models*, Chapman and Hall, 1983.