ISSN 2277-2685
IJESR/April-June. 2023/ Vol-13/Issue-2/226-232
M. Divya Sree *et. al.,* /International Journal of Engineering & Science Research

An Effective Face Recognition Method Using a Guided Image Filter and a Convolutional Neural Network

K.V.S. Ganesh[1], M. Ranganath Sarath[2], R. Sireesha[3], M. Divya Sree[4], J. Vasavi[5], Sk. Basheeramma[6]

[1-3]Assistant Professor, Department of Electronics and Communication Engineering Avanthi Institute of Engineering & Technology, Tamaram, Makavarapalem, Narsipatnam, Anakapalli District – 531113, Andhra Pradesh, India

[4-6]UG Student, Department of Electronics and Communication Engineering Avanthi Institute of Engineering & Technology, Tamaram, Makavarapalem, Narsipatnam, Anakapalli District – 531113, Andhra Pradesh, India

Abstract

Face recognition remains difficult in computer vision because pose, facial expression and illumination all vary, and performance drops in unconstrained environments. This paper implements a face recognition method combining a guided image filter with a convolutional neural network. The face region is first located in the input image using the Viola-Jones detector and resized to a fixed size, then passed through a guided image filter. The guided filter is an edge-preserving smoothing operator, which is the property that matters here: it suppresses the noise and fine texture variation that differ between images of the same person while leaving the facial edges that carry identity intact, so the network is presented with a more consistent input than the raw image provides. A convolutional network of five blocks then extracts features and classifies the face. The first four blocks each contain convolution, batch normalization, ReLU and max pooling layers, with 5 × 5 filter kernels and 16, 32, 64 and 128 channels respectively, followed by two dense layers and a softmax classifier. Experiments on the ORL, JAFFE and YALE face databases attained recognition rates of 98.33%, 99.53% and 98.65% respectively. The softmax classifier gave better results than decision tree and random forest alternatives in the classifier section of the network.

Index Terms— *Convolutional neural network, deep learning, face recognition, guided image filter, image pre-processing, softmax classifier, Viola-Jones detector.*

I. INTRODUCTION

The availability of large quantities of data, inexpensive storage and increasingly powerful processing has driven the growth of machine learning. Machine learning has computers learn to perform tasks without being explicitly programmed for them: for simple tasks it is feasible to write an algorithm specifying every step, but for more advanced tasks it can be more effective to have the machine develop its own procedure from data than to have a programmer specify each step.

Face recognition is one such task. It is used in security, authentication and human-computer interaction, and remains challenging because the same face produces very different images depending on the angle it is viewed from, the expression it carries and the light falling on it. Systems that perform well on curated datasets degrade in unconstrained conditions where these factors are not controlled.

A. Role of Pre-Processing

Pre-processing covers operations at the lowest level of abstraction, where both input and output are intensity images. Its purpose is to suppress unwanted distortion or to enhance features that matter for later processing. Pre-processing methods exploit the redundancy in images: neighbouring pixels belonging to one object usually have similar brightness, so a corrupted pixel can be restored from the average of its neighbours.

In face recognition, pre-processing matters because the variation a recognition system must ignore — sensor noise, skin texture, small lighting changes — lives at a different spatial scale from the variation it must attend to, which is the structure of the face. Filtering that separates the two makes the classification problem easier without discarding identity information.

II. LITERATURE SURVEY

Duran et al. examined the effect of different data pre-processing and normalization methods on the performance of support vector machines. The pre-processing methods compared included deleting missing values and imputing zeroes, means, medians, modes, K-nearest neighbours and predictive mean matching, each paired with log2 and Z-score normalization. Working on a microarray expression dataset, the study explored a standardized pre-processing workflow for machine learning model generation, showing that the choice of pre-processing measurably changes downstream classifier performance.

Sanghyuk et al. built a facial expression recognition system using machine learning, constructing multi-layer classifiers on a dataset of seven people. Haar-like features and histogram of oriented gradients features were extracted from each facial region, and a support vector machine classified the expression. A Haar-like feature captures the facial region, the region of interest is reset, HOG features are extracted from the new region of interest, and classification proceeds on those features. Resetting the region of interest reduces the redundant background area, which improves the classification rate.

Xiomeng et al. proposed Contrast-Accumulated Histogram Equalization to address the indiscriminate nature of ordinary histogram equalization. CACHE controls contrast gain according to the potential visual importance of intensities and pixels, coding that importance from the multi-resolution and dark-pass-filtered gradients of the image. Observing that human vision is more interested in information hidden in darker regions with sufficient spatial difference, the formulation produces a modified equalization that discriminates against noise and trivial gradients, giving better naturalness and consistency across illumination conditions than decomposition-based methods.

Wu and Liu proposed an improved support vector regression by training only with support vectors, converting the regression problem into a two-class classification problem. The support vector set is obtained by training on two-class observation data and the regression built on those support vectors. The improved regression generalized better than traditional and conventional support vector regression, and its error was stable and little affected by increasing training data.

Vinay et al. described face recognition using interest points and an ensemble of classifiers. Drawing an analogy with the temporal lobe, where specific facial features trigger neurons and are stored, the method extracts features from facial landmarks and analyses them by relative position. Several ensemble combinations were examined, concluding

with a voting classifier responsible for the final prediction. Speed and accuracy were identified as the principal challenges.

III. GUIDED IMAGE FILTER

The guided image filter computes its output as a local linear transform of a guidance image. Within a window centred on each pixel, the output is modelled as a linear function of the guidance image, with the coefficients chosen to minimize the difference between the output and the input over that window. Because the model is linear in the guidance image, the output preserves an edge wherever the guidance image has one.

This is what distinguishes it from a Gaussian filter for this application. A Gaussian blurs across edges as readily as within flat regions, so smoothing a face image enough to remove noise also softens the boundaries of the eyes, nose and mouth — exactly the structures that distinguish one face from another. The guided filter smooths within regions while leaving those boundaries sharp, so it removes the variation that is not informative about identity and retains the variation that is. In this work the input face image serves as its own guidance image, which makes the filter act as an edge-preserving smoothing operator.

IV. CONVOLUTIONAL NEURAL NETWORKS

A. Layers

A convolution layer applies a set of learnable kernels across the input, each kernel a small matrix of weights that slides over the image computing a weighted sum at each position to produce a feature map. Because the same kernel is applied everywhere, the layer detects the same pattern wherever it occurs, and the number of parameters does not grow with image size.

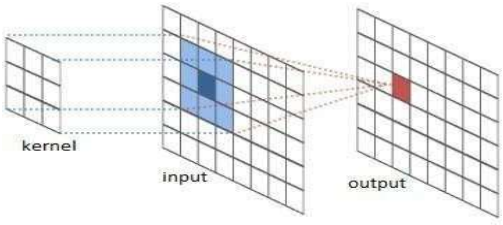

Fig. 1. Convolution layer operation.

An activation layer applies a non-linear function to the feature map. Without it, a stack of convolutions would collapse to a single linear operation and the network could represent no more than a linear model. ReLU is used here for its simplicity and because it does not saturate for positive inputs.

A pooling layer reduces the spatial dimensions of the feature map. Max pooling takes the maximum within each window, retaining the strongest response and discarding its precise location, which gives a degree of tolerance to small shifts in the input — useful when a face is not perfectly centred. A fully connected layer at the end of the network maps the extracted features to class scores.

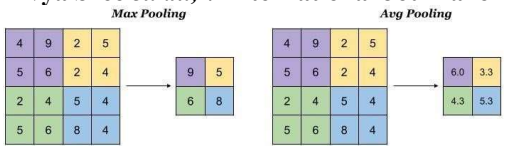

Fig. 2. Pooling operation.

B. Established Architectures

Several architectures have shaped the field. AlexNet, the ILSVRC 2012 winner, established that deep convolutional networks trained on large datasets could outperform hand-designed features. VGG demonstrated that stacking small 3 × 3 convolutions gives the receptive field of a larger filter with fewer parameters and more non-linearity between layers. GoogLeNet introduced the inception module, computing features at several scales in parallel within one layer. ResNet added skip connections, allowing very deep networks to be trained by giving gradients a direct path backwards. DenseNet extended this by connecting each layer to every subsequent layer, encouraging feature reuse.

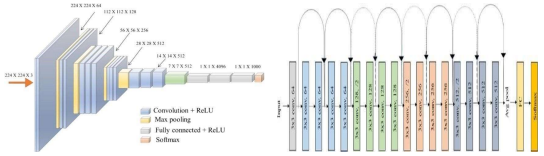

Fig. 3. VGG and ResNet architectures.

C. Optimizers

Optimizers determine how the network's weights are updated from the gradient of the loss. Gradient descent and its stochastic variant form the basis. Momentum accumulates a running average of past gradients, which smooths the trajectory and speeds progress along consistent directions. Adagrad adapts the learning rate per parameter, giving smaller updates to frequently updated parameters, though its accumulated denominator causes the rate to decay monotonically. RMSprop replaces that accumulation with an exponentially decaying average, avoiding the decay problem. Adam combines momentum with the per-parameter scaling of RMSprop, and is the common default.

V. PROPOSED METHOD

The complete pipeline runs as follows. The face region is detected in the input image using the Viola-Jones algorithm and resized to 256 × 256. Grayscale images are replicated across three channels so that colour and grayscale inputs are handled uniformly. The resized image is passed through the guided image filter, and the filtered image is fed to the convolutional network for feature extraction and classification.

Fig. 4. Block diagram of the proposed face recognition method.

A. Network Architecture

The proposed network comprises five blocks. The first four are similar to each other, containing convolution, batch normalization, ReLU and max pooling layers in sequence. The convolutional filter kernel in these blocks is 5 × 5, with 16, 32, 64 and 128 filter channels respectively, so the number of feature maps doubles as spatial resolution falls. Max pooling of 3 × 3 with stride 2 is used in all four. The final block differs, holding two dense layers followed by a softmax layer that produces the class probabilities.

Batch normalization in each block normalizes the activations, which stabilizes training and allows a higher learning rate than would otherwise be usable. The progression from 16 to 128 channels reflects the usual design principle: early layers detect a small number of simple, local patterns, while later layers combine these into a larger number of more specific configurations.

B. Performance Measures

Performance is assessed using accuracy, the proportion of all predictions that are correct; sensitivity or recall, the proportion of actual positives correctly identified; precision, the proportion of predicted positives that are correct; and F-score, the harmonic mean of precision and recall. These are computed from the counts of true positives, true negatives, false positives and false negatives.

VI. RESULTS

Experiments were performed on the ORL, JAFFE and YALE face databases. The three differ in what they vary: ORL includes pose and expression variation, JAFFE consists of posed facial expressions, and YALE varies lighting conditions and includes occlusion by glasses. Testing across all three checks the method against each of the factors that make face recognition difficult rather than against one of them.

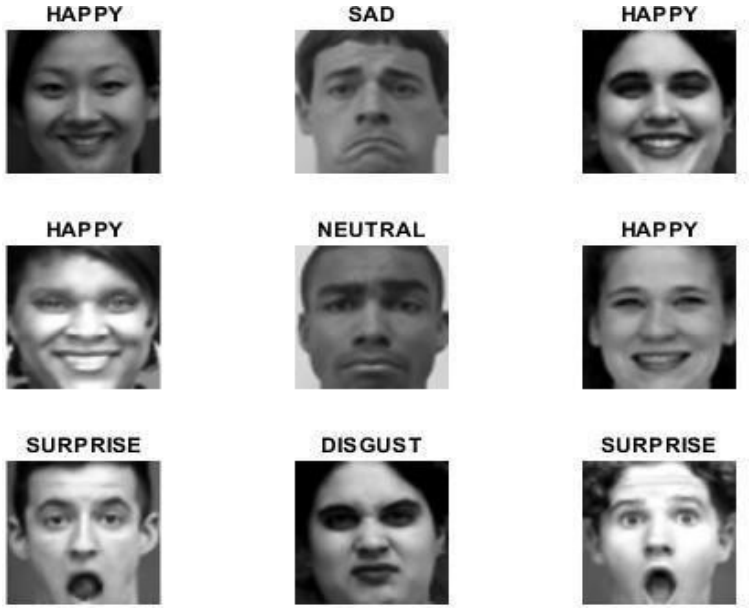

Fig. 5. Facial expressions from the test databases.

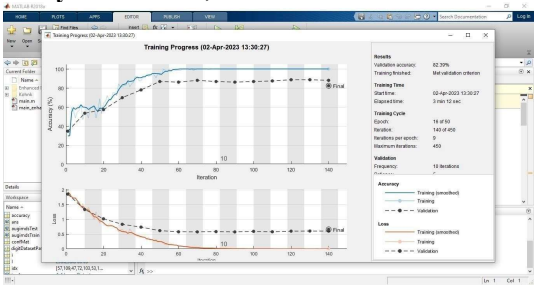

Fig. 6. Training progress in MATLAB.

Recognition rates of 98.33%, 99.53% and 98.65% were attained on ORL, JAFFE and YALE respectively. The highest rate came on JAFFE, which is consistent with its being the most controlled of the three: images are captured under fixed lighting with a constrained pose, so the remaining variation is expression alone. The slightly lower rates on ORL and YALE reflect the additional pose and illumination variation those sets contain.

Within the classifier section of the network, the softmax classifier gave better results than decision tree and random forest alternatives. This is expected given the input: softmax is trained jointly with the feature extractor, so the features are shaped to be linearly separable by it, whereas a tree-based classifier is fitted after the fact to features it did not influence.

VII. CONCLUSION

A face recognition approach using a guided image filter and a convolutional neural network was implemented. The face region is extracted with the Viola-Jones algorithm and resized, the resized image is smoothed by a guided image filter, and the proposed CNN extracts features and classifies the face. Recognition rates of 98.33% on ORL, 99.53% on JAFFE and 98.65% on YALE were obtained, better than several of the earlier methods compared against.

The contribution is in the pairing rather than in either component alone. The guided filter is chosen specifically because it removes variation at the scale where noise and texture live while preserving the edges that carry identity, so the network receives an input in which the relevant structure is intact and the irrelevant variation is reduced. Applying a filter without that edge-preserving property would remove both together.

VIII. FUTURE SCOPE

Learned pre-processing. Deep learning-based enhancement has shown promise in image restoration, and could replace the fixed guided filter with one adapted to the recognition task itself.

Multi-modal recognition. Combining visible light images with thermal or 3D data would provide information that survives conditions in which the visible image degrades.

Real-world evaluation. The three databases used are benchmarks with controlled capture conditions; evaluation on unconstrained data with varying lighting, occlusion and pose would establish how far the results carry over.

Privacy. Face recognition raises real privacy and ethical concerns, and privacy-preserving formulations that avoid storing identifiable templates are an active direction.

Online learning. Adapting to new faces and to changes in existing faces without full retraining would suit deployments where the enrolled set is not fixed.

References

[1] I. Duran, R. Leandro, and J. Guevara-Coto, "Analysis of different pre-processing techniques to the development of machine learning predictors with gene expression profiles," in Proc. IEEE Int. Conf. JoCICI, 2018, pp. 99–167.

[2] K. Sanghyuk, G. Hwan An, and S. Ju Kang, "Facial expression recognition system using machine learning," in Proc. Int. SoC Design Conference (ISOCC), 2017, pp. 68–112.

[3] X. Wu, X. Liu, K. Hiramatsu, and K. Kashino, "Contrast-accumulated histogram equalization for image enhancement," in Proc. IEEE Int. Conf. Image Processing (ICIP), 2017, pp. 3190–3194.

[4] C.-A. Wu and H.-B. Liu, "An improved support vector regression based on classification," in Proc. Int. Conf. Multimedia and Ubiquitous Engineering, 2007.

[5] A. Vinay, P. R. Sampat, S. V. Belavadi, and P. R. Kashyap, "Face recognition using interest points and ensemble of classifiers," in Proc. Int. Conf. Computational Intelligence and Data Science, 2018.

[6] K. He, J. Sun, and X. Tang, "Guided image filtering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 6, pp. 1397–1409, 2013.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2001, pp. 511–518.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learning Representations (ICLR), 2015.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[13] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in Proc. Int. Conf. Machine Learning (ICML), 2015, pp. 448–456.

[14] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Proc. Int. Conf. Learning Representations (ICLR), 2015.

[15] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in Proc. IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142.