Received October 2, 2015, accepted November 8, 2015, date of publication December 22, 2015, date of current version January 8, 2016.

Digital Object Identifier 10.1109/ACCESS.2015.2510865

Regularized Deep Learning for Face Recognition With Weight Variations

SHRUTI NAGPAL, MANEET SINGH, RICHA SINGH, (Senior Member, IEEE), AND MAYANK VATSA, (Senior Member, IEEE)
Indraprastha Institute of Information Technology Delhi, New Delhi 110020, India
Corresponding author: M. Vatsa (mayank@iiitd.ac.in)

This work was supported in part by the Department of Electronics and Information Technology, Government of India.

ABSTRACT Body weight variations are an integral part of a person's aging process. However, the lack of association between the age and the weight of an individual makes it challenging to model these variations for automatic face recognition. In this paper, we propose a regularizer-based approach to learn weight invariant facial representations using two different deep learning architectures, namely, sparse-stacked denoising autoencoders and deep Boltzmann machines. We incorporate a body-weight aware regularization parameter in the loss function of these architectures to help learn weight-aware features. The experiments performed on the extended WIT database show that the introduction of weight aware regularization improves the identification accuracy of the architectures both with and without dropout.

INDEX TERMS Face recognition, biometrics, body-weight variations, facial aging.

I. INTRODUCTION

Automated face recognition requires addressing several challenging covariates that affect the performance. Several covariates such as illumination, pose, aging, disguise, sketch and plastic surgery have been identified in literature [2], [3]. Among all of these, aging is one of the most challenging covariates in face recognition. It is an inherent process in human beings and causes inevitable changes in the facial structure and features. These variations in the facial features make it difficult for an automated face recognition system to correctly recognize identities, thereby giving poor performance. Along with aging (growth), facial appearances are also affected by body weight variations.[1] As time progresses, the weight of an individual might increase or decrease depending on several external as well as internal factors of the human body. It is well accepted that body weight variations do not follow any fixed pattern. The effect of these variations on the facial structure and features of an individual is also not consistent. Moreover, these variations vary vastly across subjects. Due to these inconsistent fluctuations, it becomes difficult to model the face variations with body weight changes over time for different subjects. Figure 1 shows sample images of three individuals with different weight variations over time. Each row corresponds to one subject with each column representing images taken at different times. It is clear that there is no fixed pattern for weight variations over the three individuals, and even within a particular row.

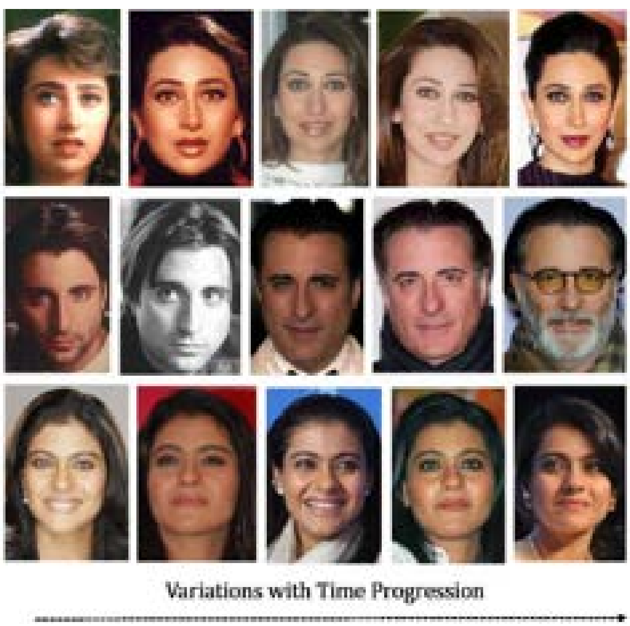

FIGURE 1. Face images of three subjects with different weight variations over time. Each row represents one subject with each column representing images taken at different times. Clearly, no fixed pattern can be observed for the weight variations - within subject or between subjects.

In this paper, we present a deep learning based framework that incorporates body weight variations in feature learning. Deep learning has previously been used in

[1] A preliminary version of this research was published in IEEE International Conference on Biometrics: Theory, Applications and Systems, 2015 [1].

2169-3536 © 2015 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

FIGURE 2. Sample images depicting different weight variations with different age variations. The first row shows very little weight variations within an age range. The second row shows weight variations within a small age range whereas the third row shows weight variations with large age variations. These examples depict that there is no fixed pattern for weight variations with increasing age.

literature to learn features invariant to illumination, expression, pose [4]–[9] and has shown significant improvement in results. The success of deep learning has motivated us to explore this paradigm for finding useful representation that can address the variations caused by body weight changes. Although deep learning based existing algorithms yield improved face recognition results, learning based representation algorithms require significantly large amount of training data. It has been observed that representative training data from the corresponding domain is important for learning robust representations. However, for face recognition with body weight variations, it is very challenging to obtain data with respect to various weight and age variations. Learning complex networks with fewer training data points leads to overfitting and hence affects the generalization ability on the testing data. Therefore, in this research, we propose a *regularization* based approach by modifying the existing loss functions of deep learning architectures and incorporating body-weight category parameter in the loss functions. The proposed regularized deep learning framework ensures sparsity and controls overfitting. The experimental results and comparison on the eWIT [1] (extended WIT) database demonstrates that the proposed framework significantly improves the face recognition performance as compared to the existing algorithms. In the next section, the proposed algorithm is explained in detail, followed by specifications of the dataset used for evaluation. Section IV explains the results, followed by the conclusions.

II. PROPOSED ALGORITHM

Due to the lack of any fixed pattern in body-weight variations, both within and between subjects, it becomes extremely challenging to model these variations (as shown in Figure 2). It is our assertion that learning body weight invariant representation can help to improve the face recognition performance. Thus, we propose a learning based algorithm for feature extraction and classification.

The proposed model comprises of a deep learning architecture for learning robust representations, followed by a Random Decision Forest (RDF) for classification [10]. Since the proposed formulation is generic in nature, it is explained using two different deep learning architectures: Sparse-Stacked Denoising Autoencoders (SDAE) [11] and Deep Boltzmann Machines (DBM) [12]. The following subsection first explains the basics of SDAE and DBM, followed by the proposed framework with multiple regularizers and dropout network [13].

A. PRELIMINARIES

1) SPARSE-STACKED DENOISING AUTOENCODERS

As the name suggests, Sparse-Stacked denoising autoencoders are basically sequence of sparse denoising autoencoders stacked together. Since the complete architecture becomes too complex, greedy layer-by-layer training [14] is used to train the entire deep network. Stacking is performed such that the output layer of the first autoencoder behaves as the input layer of the second autoencoder. The two primary components of an autoencoder are the encoder and the decoder. The encoder is responsible for transforming the input vector into a hidden representation while the decoder maps it back to the original input vector. For a given input vector, x, the hidden representation, y, is calculated as:

$$y = \phi(Wx + b) \tag{1}$$

where, W is the weight matrix, w_{ij} represents the weight of the connection from the i^{th} input node to the j^{th} hidden node, ϕ represents the activation function of the nodes, and b represents the bias. The decoder maps the learnt features to the data space, using the following equation:

$$z = \phi(W'y + b') \tag{2}$$

where, W' is the weight matrix, w'_{ij} represents the weight of the connection from the i^{th} hidden node to the j^{th} decoder output node, and b' represents the bias. The loss function of an autoencoder is thus formulated as:

$$\mathcal{L}_{ae} = \|x - z\|_F^2 = \left\|x - \phi(W'\phi(Wx + b) + b')\right\|_F^2 \tag{3}$$

Further, additional constrained and learning approaches are used such as introducing sparsity using a regularizer and learning using noisy input data to make the learned architecture generalizable.

2) DEEP BOLTZMANN MACHINES

Deep Boltzmann Machines are stacked Restricted Boltzmann Machines (RBM) having undirected edges between the layers [12]. They are extremely useful for unsupervised learning of feature representations from a given large unlabeled data. Similar to stacked autoencoders, due to the complex architecture of such large networks, a greedy layer-by-layer training approach is used to stack the RBMs to train a complete DBM. The energy function of a (binary) RBM can be formulated as follows:

$$E(x, h) = -a^T x - b^T h - x^T Wh \tag{4}$$

where, x and h represent the visible and hidden units, respectively. W is the weight matrix where weight w_{ij} signifies the weight of connection between the hidden unit h_j and visible unit x_i. a represents the bias weights for visible units and b represents the bias weights for the hidden units. The probability distribution of a RBM over the hidden and visible units is defined as:

$$P(x, h) = \frac{1}{Z} exp(-E(x, h)) \tag{5}$$

where, Z is the partition function, which is a normalization constant. This further leads to the formulation of marginal probability which is the sum of all possible combinations of the hidden unit configurations, i.e.,

$$P(x) = \sum_h P(x, h) = \frac{1}{Z} \sum_h exp(-E(x, h)) \tag{6}$$

Using the training data, RBMs are trained to minimize the negative log likelihood, i.e. the loss function $\mathcal{L}_{rbm}$ is defined as:

$$\mathcal{L}_{rbm} = -\sum_{x \epsilon X} log(P(x)). \tag{7}$$

3) PRE-TRAINING AND GENERALIZABILITY

Deep Learning architectures require huge amount of data for training. Since the optimization functions are generally formulated to reduce the training error, architectures tend to learn maximum information from the given data. Learning a complex function with limited training data sometimes leads to overfitting and hence, lower generalization power and poor prediction results. To address these challenges, researchers have proposed several techniques [11], [15]–[18], including pre-training and fine-tuning, transfer learning, and regularization. *Pre-training* assumes that large unlabeled data (from a similar problem domain) is available. This data is used to train the network in unsupervised fashion and learn the initial set of parameters which correspond to an *approximate* representation. The small set of problem specific (labeled) training data is utilized to fine tune the representation. A classifier is also trained which provides *class* information. This approach helps the model to learn a good representation even with less training data.

The next challenge is related to generalizability. It is widely preferred to have a model which can increase the generalizability without reducing the power of the model. One possible method to achieve this is by adding a penalty term to loss function which is known as the regularizer. For a given problem, regularization is done to avoid overfitting and converge to a solution faster by providing ancillary information.

B. PROPOSED ALGORITHM

In this research, we propose a deep learning framework to learn a good feature representation with limited training samples (from face images with body weight variations). Using the existing regularization approaches, such as l_1, l_2 norms, and dropout, we update the objective function of a deep learning architecture (e.g. RBM, SDAE) which attempts to minimize the loss function, $\mathcal{L}_\tau$.[2] The loss function is modified by introducing a regularizer term dependent on the body weight category of the training samples. The formulation of the loss function of the proposed deep learning architecture is given by:

$$\mathcal{L} = \mathcal{L}_\tau + \lambda \left\|\alpha_{bw} W\right\|_p^p \tag{8}$$

where, W is the network weight matrix learned by minimizing the loss function and λ is the regularization parameter learnt at the time of training the network. Each sample in the eWIT face dataset (discussed later) is given one weight category, *thin*, *moderate* or *heavy* depending on the body weight. A body weight parameter, α_{bw} is introduced which is defined as:

$$\alpha_{bw} = \frac{S_{bw}}{225} \tag{9}$$

where, $S_{bw} = 50, 75, 100$ for a given weight category, bw. Since the weight categories are discrete, S_{bw} takes one of the three numerical values depending on the weight label for the given sample. For example, a sample having weight category *thin* would have a S_{bw} value of 50, whereas a sample classified as *heavy* would have a value of 100. Using Eq. 9, α_{bw} takes one of the three values α_{thin}, $\alpha_{moderate}$ or α_{heavy}, depending on the weight category of the training sample.

[2] In case of SDAE, $\mathcal{L}_\tau$ is $\mathcal{L}_{ae}$ and for RBM, $\mathcal{L}_\tau$ is $\mathcal{L}_{rbm}$.

Using Eq. 8, different regularization approaches can be applied to modify the loss function of the deep learning network. The following regularizer approaches have been explored in this research:

- l_1 norm regularization:

$$\mathcal{L} = \mathcal{L}_\tau + \lambda_1 \|\alpha_{bw} W\|_1 \tag{10}$$

The l_1 norm regularization term forces the sum of the absolute values of the weight vector to be low. This introduces sparsity in the weight matrix which leads to better feature selection. This is possible because only the important and representative nodes have a high value in the weight matrix. Using Eq. 10, the loss function of SDAE and RBM are modified as follows:

  - SDAE:

$$\mathcal{L}_{ae} = \left\|x - \phi(W'\phi(Wx + b) + b')\right\|_F^2 + \lambda_1 \|\alpha_{bw} W\|_1 \tag{11}$$

  - RBM:

$$\mathcal{L}_{rbm} = -\sum_{x \epsilon X} log(P(x)) + \lambda_1 \|\alpha_{bw} W\|_1 \tag{12}$$

- l_2 norm regularization:

$$\mathcal{L} = \mathcal{L}_\tau + \lambda_2 \|\alpha_{bw} W\|_2^2 \tag{13}$$

l_2 norm regularization term minimizes the sum of squares of the weight vector. It clips the peak weights as squaring a number leads to larger penalties for higher values as compared to smaller values. The l_2 norm regularizer jointly forces the entire weight matrix to shrink to lower values. Similar to the update of SDAE and RBM loss function using l_1 norm regularization, l_2 norm regularization is also applied to SDAE and RBM.

- $l_1 + l_2$ norm regularization (elastic net):

$$\mathcal{L} = \mathcal{L}_\tau + \lambda_1 \|\alpha_{bw} W\|_1 + \lambda_2 \|\alpha_{bw} W\|_2^2 \tag{14}$$

In this approach, l_1 norm regularizer tries to find the most representative feature by introducing sparsity and l_2 norm regularizer reduces the values to a smaller range by penalizing extreme values. l_1 reduces overfitting but can also result in loss of discriminative information, and l_2 may retain features which are not truly representative of the data. Therefore, on combining the two norms, the proposed model obtains a feature set which is both representative and discriminative.

Dropout Training: During dropout training [13], several representations are learned from multiple architectures created by randomly selecting some units for each input. The output at a particular node, z_i is given by the equation below:

$$z_i = w_i(r^{(l)} * y) + b_i \tag{15}$$

Here, $r^{(l)}$ is a vector of independent Bernoulli random variables (for a given layer l), each having a fixed probability of being 1. w_i and b_i are the weights and bias of the hidden unit i while y is the output of the layer. The final representation is the average of the representations learned from the multiple architectures. Averaging prevents overfitting of the model and introduces sparsity in the method of training.

- Dropout with max-norm:

$$\mathcal{L} = \mathcal{L}'_\tau + \lambda_3 \|\alpha_{bw} W\|_1 \tag{16}$$

s.t.

$$\|\alpha_{bw} W\|_1 \leqslant c$$

Here, $\mathcal{L}'_\tau$ corresponds to the loss function of the deep learning archtectures with drop-out learning (Eq. 15). In max-norm regularization, the maximum value of the norm of weight vector at each hidden layer is less than a fixed constant c. This restricts the weight vector values from being extremely high and lie within a specific upper bound c, which is learned during training and prevents overfitting. Max-norm with dropout learning introduces sparsity by the method of training and the max norm regularizer clips peak weights, thus learning a good feature representation of the training samples.

- Dropout with l_2 norm:

$$\mathcal{L} = \mathcal{L}'_\tau + \lambda_4 \|\alpha_{bw} W\|_2^2 \tag{17}$$

l_2 norm has a similar impact as max- norm. It restricts extreme values by penalizing peaking values and spreads the error throughout the weight matrix. This ensures the new representation is learned in such a manner that it retains the discriminative features.

- Dropout with $l_{2,1}$ norm:

$$\mathcal{L} = \mathcal{L}'_\tau + \lambda_5 \|WX_{bw}\|_{2,1} \tag{18}$$

$l_{2,1}$ norm introduces group sparsity [19], [20] and encourages the model to learn group specific features. It tries to make the model discriminative towards different labels provided. Till now $l_{2,1}$ norm has been used to regularize a network with respect to the class identities for supervised training. However, in the proposed architecture, for body weight-based $l_{2,1}$ norm regularization, we train the architecture in a supervised manner with respect to the body weight labels, as opposed to the class identities. That is, the network is regularized for a three class problem (*thin*, *moderate* and *heavy*) as opposed to a n-class (identity) problem. This helps the architecture to learn weight invariant features for the three classes. In Eq. 18, W is the weight matrix and X_{bw} are the training samples, with respect to the three weight classes, i.e. X_{thin}, X_{mod} or X_{heavy}, depending upon the weight category of the given sample.

We propose $l_{2,1}$ norm regularization on dropout learning. The novelty lies in the application of the group sparsity constraint, which is learned to introduce sparsity on weight-based groups. The architecture thus learns features which are specific to a given weight category. Once the features are extracted, a RDF is used for classification.

FIGURE 3. Sample images from the eWIT face database.

C. RANDOM DECISION FOREST BASED CLASSIFICATION

Once the features are learnt for face images using the proposed deep-learning architecture, Random Decision Forest (RDF) is used for classification [10]. RDF is an ensemble of decision trees and can handle non-linearity in the feature space while being robust to outliers. It also provides a stable performance with increase in the number of images in the gallery set [10]. Once a feature learning model is trained, features are extracted from the training set (to learn the classifier) and given as input to RDF to learn the model for n-class classification. During testing, features are extracted from the test image via the trained deep learning architecture and using these extracted features, RDF provides the identity label of the test image.

III. eWIT DATASET

WIT [21] is the only publicly available dataset capturing weight and age variations of 100 subjects, for a total of 1109 images. Extended WIT (eWIT), which is an extension of the WIT dataset is used in this research. eWIT consists of 2036 images of 200 subjects such that each subject has at least 10 and at most 14 images with weight and age variations. Further, all the images are frontal with minor pose variations while no constraint is kept on the illumination or expression. The database contains images of public figures collected from the Internet and it is ensured that some age variations are maintained throughout for each subject. As summarized in Table 1, the average age of all the images in eWIT is 34.29 years, while the total range is 1 to 96 years. The average age range for each subject, i.e. difference between the age of the oldest image and the youngest image, is 28.78 years.

Figure 3 illustrates sample images from the e-WIT database. Face images are detected using the OpenCV face detector and normalized using three point (eyes and mouth) geometric normalization. To quantifiably use the weight variations of the dataset, each image is labeled into one of the three weight categories: *thin*, *moderate* or *heavy*. In total, the dataset consists of 437 *thin*, 1309 *moderate*, and 290 *heavy* samples. Figure 4 demonstrates the mean images of the detected faces for each of the three weight categories. It is interesting to observe that the mean images also clearly show the difference in facial structure from thin to heavy. The high age range and visual variations between the three mean

TABLE 1. Description of the eWIT dataset.

Attribute	Value
Number of subjects	200
Number of images	2036
Age range	[1 - 96]
Average age	34.29
Images per subject	[10 - 14]
Weight category wise distribution of images	
Thin	437
Moderate	1309
Heavy	290

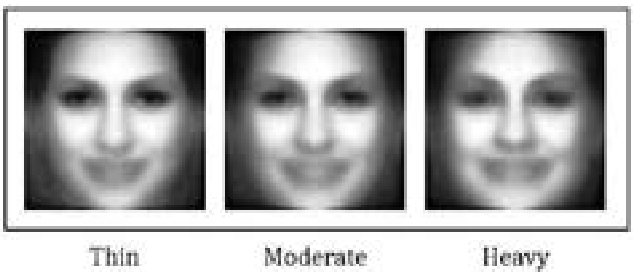

FIGURE 4. Weight category wise mean images from eWIT dataset. From left to right: thin, moderate, heavy [21].

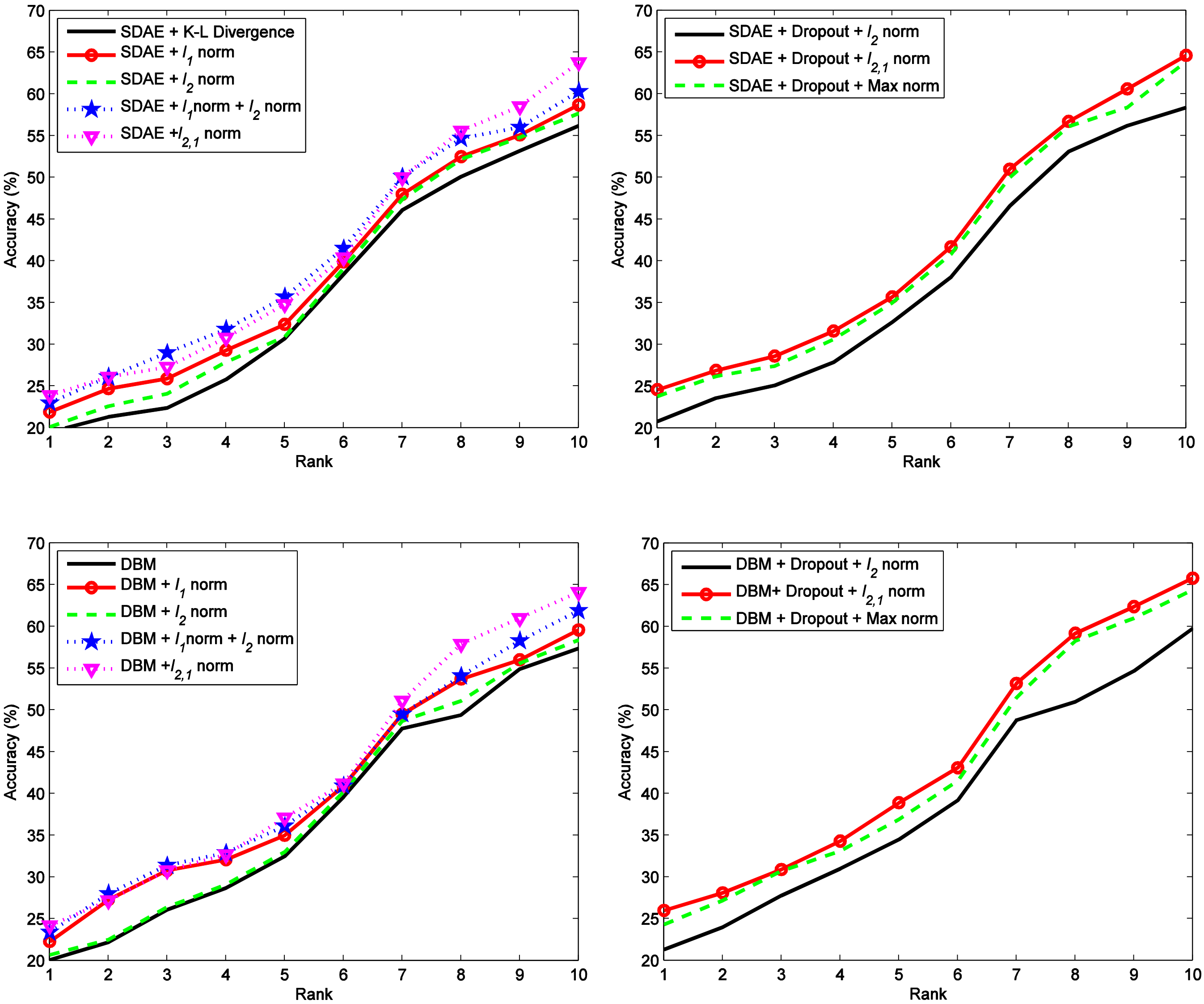

FIGURE 5. **CMC curves obtained with SDAE and DBM architectures on the eWIT database. The first row gives the curves for SDAE while the second row shows the results for DBM.**

images clearly validate the presence of both, age and weight variations in eWIT dataset.

IV. EXPERIMENTS AND RESULTS

Due to the complex architecture, deep learning algorithms require large amount of data to train the network. However, eWIT is a comparatively smaller database, not sufficient to successfully train a DBM or a SDAE. Therefore, we use a transfer-learning based method of pre-training the deep learning architectures. The deep learning architecture is first pre-trained with 600,000 frontal face images combined from several publicly available face datasets. This helps learn an unsupervised feature representation of face images. The initial representation is then followed by fine-tuning with the small training subset of the eWIT dataset. As mentioned by Salakhutdinov and Hinton [12], "*high-level representations can be built from a large supply of unlabeled sensory inputs and very limited labeled data can then be used to only slightly fine-tune the model for a specific task at hand*". Our proposed algorithm utilizes this very property of deep learning architectures to train models using the limited labeled data available with body weight variations.

The eWIT dataset is divided into two subsets, training and testing. 50% images of each subject are randomly selected and used in training while the remaining are used for testing. This is done so as to perform 200-class identification experiments. The training set is used to train the proposed architecture while the testing set is used to test the trained model and report the identification accuracies. This process is repeated two times (random subsampling based cross validation) and average identification results are reported. Table 2 tabulates the results of the proposed architectures with SDAE, Table 3

TABLE 2. Identification accuracies (%) obtained with SDAE.

Algorithm	Identification Accuracy	
	Rank-1	Rank-10
KL Divergence	19.5	56.2
l_1 norm	21.9	58.7
l_2 norm	20.1	57.7
l_1 norm + l_2 norm	23.0	60.3
$l_{2,1}$-norm	23.9	63.8
Dropout and max-norm	23.8	63.9
Dropout and l_2 norm	20.8	58.4
Dropout and $l_{2,1}$ norm	24.6	64.6

TABLE 3. Identification accuracies (%) obtained with DBM.

Algorithm	Identification Accuracy	
	Rank-1	Rank-10
No regularization	20.1	57.4
l_1 norm	22.3	59.6
l_2 norm	20.7	58.4
l_1 norm + l_2 norm	23.4	61.9
$l_{2,1}$-norm	24.2	64.1
Dropout and max-norm	24.3	64.4
Dropout and l_2 norm	21.3	59.8
Dropout and $l_{2,1}$ norm	26.0	65.8

TABLE 4. Identification accuracies (%) of existing algorithms and COTS with best of the proposed SDAE and DBM architectures on the eWIT dataset.

Algorithms		Rank-1	Rank-10
COTS (VeriLook)		14.3	47.0
Singh et al. [21]		17.7	51.2
SDAE	$l_{2,1}$ norm	23.9	63.8
	Dropout and $l_{2,1}$ norm	24.6	64.6
DBM	Dropout and max-norm	24.3	64.4
	Dropout and $l_{2,1}$ norm	26.0	65.8

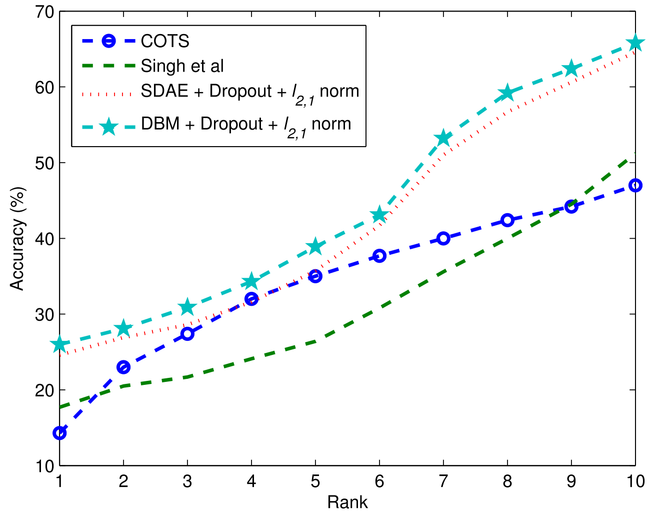

FIGURE 6. Comparing the identification performance of the proposed algorithm with existing algorithm [21] and VeriLook [22].

showcases the results using DBM, and Table 4 tabulates the best of the two architectures along with results obtained by state-of-the-art algorithm proposed by Singh *et al.* [21] and a commercial-off-the-shelf (COTS) face recognition system, VeriLook [22]. Figures 5 and 6 show the Cumulative Match Characteristic (CMC) graphs comparing the proposed and existing approaches. Key results of our experiments are as follows:

- The existing algorithm [21] yields a Rank-1 accuracy of only 17.7% and a Rank-10 accuracy of 51.2%. On the other hand, commercial-off-the-shelf system [22] provides a Rank-1 accuracy of only 14.3% and a Rank-10 accuracy of 47.0%. This is primarily due to the fact that COTS is not trained on the eWIT database, whereas existing algorithm [21] is trained on this database and therefore, it has learnt body weight variations to a certain extent.
- Without any regularization, DBM provides a baseline Rank-1 accuracy of 20.1% and a Rank-10 accuracy of 57.4%. Using a SDAE with the default KL divergence based regularization (without body weight based regularization), Rank-1 accuracy of 19.5% and Rank-10 accuracy of 56.2% are achieved. These results emphasize upon the challenging nature of the given problem and the need for incorporating body weight information in the framework.
- The proposed framework with DBM architecture, yields a Rank-1 accuracy of 26.0% and a Rank-10 accuracy of 65.8% by using dropout and a weight-based $l_{2,1}$ norm regularization. This method shows an improvement of around 12% for Rank-1 accuracy, as compared to Verilook and an improvement of more than 18% for Rank-10 accuracy. As compared to the current state-of-the-art algorithm [21], the proposed framework shows an improvement of more than 8% for Rank-1 accuracy and around 14% for Rank-10 accuracy.
- The proposed framework with SDAE architecture, yields a Rank-1 and Rank-10 accuracy of 24.6% and 64.6% respectively. Similar to the previous results on DBM, this is obtained by using SDAE with dropout and weight-based $l_{2,1}$ norm regularization. This method shows an improvement of around 10.3% for Rank-1 accuracy, as compared to Verilook and an improvement of more than 17% for Rank-10 accuracy. As compared to the current state-of-the-art algorithm [21], the proposed framework shows an improvement of around 7% for Rank-1 accuracy and around 13.4% for Rank-10 accuracy.
- The use of dropout and weight-based $l_{2,1}$ norm regularization gives the best accuracies with the proposed framework. Dropout learning introduces sparsity in the architecture, thereby making the entire network more robust to new samples and noise. Weight-based $l_{2,1}$ norm regularization helps in learning weight-invariant features by selecting those latent variables which acts as a robust feature selector.

- All variations of the proposed algorithm, with different body weight-based regularizer terms perform better than other approaches. This encourages the fact that the body weight-based regularizer is useful for learning robust features, which are helpful for the task of face recognition with body weight variations.
- Computationally, on a 6C 2.4GHz workstation with 64GB RAM, the regularized DBM and regularized SDAE based feature extraction followed by RDF based classifier require less than 1 second for identification.

V. CONCLUSION

Body weight variations affect the performance of automatic face recognition algorithms. Learning-based algorithms have been proposed to learn weight invariant features. However, they require significant amount of representative training data. Owing to the subjective characteristic of the problem, it is challenging to obtain large labeled training data for body weight variations and hence, it is not feasible to learn supervised weight models. In this research, we propose a regularization based deep learning approach to address this challenge. The effectiveness of the proposed approach is evaluated with multiple deep learning architectures and the results show that incorporating weight aware regularizers control overfitting and improve the identification performance. This research can be further extended in multiple directions: extend the database to include more images with weight variations, and improve the algorithm with incorporating both age and weight variations.

ACKNOWLEDGEMENT

Shruti Nagpal and Maneet Singh equally contributed to this work. The authors also thank Dr. Angshul Majumdar, IIIT-Delhi and Dr. Saket Anand, IIIT-Delhi for their feedback on this research.

REFERENCES

[1] S. Nagpal, M. Singh, M. Vatsa, and R. Singh, "Regularizing deep learning architecture for face recognition with weight variations," in *Proc. Biometrics, Theory, Appl. Syst.*, 2015, pp. 1–6.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[3] H. S. Bhatt, R. Singh, and M. Vatsa, "Covariates of face recognition," Indraprastha Inst. Inf. Technol., New Delhi, India, Tech. Rep. IIITD-TR-2015-002, 2015.

[4] G. Goswami, R. Bhardwaj, R. Singh, and M. Vatsa, "MDLFace: Memorability augmented deep learning for video face recognition," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep./Oct. 2014, pp. 1–7.

[5] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27. 2014, pp. 1988–1996.

[6] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.

[7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.

[9] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[10] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1. Aug. 1995, pp. 278–282.

[11] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[12] R. Salakhutdinov and G. Hinton, "Deep Boltzmann machines," in *Proc. Int. Conf. Artif. Intell. Statist.*, vol. 5. 2009, pp. 448–455.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[14] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2007.

[15] G. E. Hinton and R. R. Salakhutdinov, "A better way to pretrain deep Boltzmann machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2447–2455.

[16] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, Mar. 2010.

[17] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. Int. Conf. Mach. Learn. Workshop Unsupervised Transf. Learn.*, 2012, pp. 37–49.

[18] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. Int. Conf. Mach. Learn. Workshop Unsupervised Transf. Learn.*, vol. 27. 2012, pp. 17–36.

[19] G. Goswami, P. Mittal, A. Majumdar, M. Vatsa, and R. Singh, "Group sparse representation based classification for multi-feature multimodal biometrics," *Inf. Fusion*, to be published.

[20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[21] M. Singh, S. Nagpal, R. Singh, and M. Vatsa, "On recognizing face images with weight and age variations," *IEEE Access*, vol. 2, pp. 822–830, 2014.

[22] *Verilook*. [Online]. Available: http://www.neurotechnology.com/verilook.html

SHRUTI NAGPAL is currently pursuing the Ph.D. degree with the Indraprastha Institute of Information Technology Delhi, India. Her research interests include image processing, deep learning, pattern recognition, and face recognition.

MANEET SINGH is currently pursuing the Ph.D. degree with the Indraprastha Institute of Information Technology Delhi, India. Her research interests include pattern recognition, deep learning, information retrieval, and face recognition. She is a recipient of the Google Anita Borg Memorial Scholarship 2015.

RICHA SINGH (S'04–M'09–SM'14) received the M.S. and Ph.D. degrees in computer science from West Virginia University, Morgantown, USA, in 2005 and 2008, respectively. She is currently an Associate Professor and a Kusum and TV Mohandas Pai Faculty Research Fellow with the Indraprastha Institute of Information Technology Delhi, India. Her research has been funded by the UIDAI and DeitY, India. Her areas of interest are biometrics, pattern recognition, and machine learning. She is a recipient of the FAST Award by DST, India. She is also an Editorial Board Member of *Information Fusion* (Elsevier) and the *EURASIP Journal of Image and Vision Processing* (Springer). She is a member of the Computer Society and the Association for Computing Machinery. She has co-authored over 150 research papers and received several best paper and best poster awards in international conferences. She is a recipient of the NVIDIA Innovation Award 2015 and the Best Reviewer Award at the IAPR International Conference on Biometrics 2013. She serves as the PC Co-Chair of the IEEE International Conference on Biometrics: Theory, Applications, and Systems 2016.

MAYANK VATSA (S'04–M'09–SM'14) received the M.S. and Ph.D. degrees in computer science from West Virginia University, Morgantown, USA, in 2005 and 2008, respectively. He is currently an Associate Professor and AR Krishnaswamy Faculty Research Fellow with the Indraprastha Institute of Information Technology Delhi, India. His research has been funded by the UIDAI, DST, and DeitY. He has authored over 150 research papers and received several best paper and best poster awards and the NVIDIA Innovation Award 2015. His areas of interest are biometrics, image processing, machine learning, and information fusion. He is a recipient of the FAST Award by DST, India. He is a member of the Computer Society and the Association for Computing Machinery. He is the Vice President (Publications) of the IEEE Biometrics Council, an Area Editor of the *IEEE Biometric Compendium*, and an Associate Editor of the *Information Fusion Journal* (Elsevier) and the IEEE Access. He was also the Program Committee Co-Chair of the IAPR International Conference on Biometrics 2013 and the International Joint Conference on Biometrics 2014.