Received: March 6, 2020. Revised: April 23, 2020.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Adaptive Inception Based on Transfer Learning for Effective Visual Recognition

Balaji Sreenivasulu[1]* Anjaneyulu Pasala[2] Gaikwad Vasanth[3]

[1]*Department of Computer Science and Engineering, Visvesvaraya Technological University, India*
[2]*Infosys Labs, Bangalore, India*
[3]*KR Pet, Government Engineering College, Bangalore, India*
* Corresponding author's Email: balajichetty6@gmail.com

Abstract: In computer vision, domain adaptation or transfer learning plays an important role because it learns a target classifier characteristics using labeled data from various distribution. The existing researches mostly focused on minimizing the time complexity of neural networks and it effectively worked on low-level features. However, the existing method failed to concentrate on data augmentation time and cost of labeled data. Moreover, machine learning techniques face difficulty to obtain the large amount of distributed label data. In this research study, the pre-trained network called inception layer is fine-tuned with the augmented data. There are two phases present in this study, where the effectiveness of data augmentation for Inception pre-trained networks is investigated in the first phase. The transfer learning approach is used to enhance the results of the first phase and the Support Vector Machine (SVM) is used to learn all the features extracted from inception layers. The experiments are conducted on a publicly available dataset to estimate the effectiveness of proposed method. The results stated that the proposed method achieved 95.23% accuracy, where the existing techniques namely deep neural network and traditional convolutional networks achieved 87.32% and 91.32% accuracy respectively. This validation results proved that the developed method nearly achieved 4-8% improvement in accuracy than existing techniques.

Keywords: Data augmentation, Inception layer, Labeled data, Support vector machine, Transfer learning, Traditional convolutional networks.

1. Introduction

In many knowledge engineering areas like regression, clustering and classification, machine learning and data mining technologies have achieved significant success in the past few decades. But, according to common assumptions, various machine learning algorithms worked effectively, where the common assumption is that the testing and training data are taken only from the same distribution and feature space [1]. Most statistical models are needed to be reconstruct the model from scratch by collecting new training data, when the distributions are changed. However, it is expensive and impossible in many real world applications for recollecting the training data to rebuild the models [2, 3]. In recent years, researchers tried to minimize the need and effort for recollecting the training data. In such cases, transfer learning (TL) or knowledge transfer between task domains are considered as highly desirable among researchers [4]. In everyday life, online images and videos play a major role in automatic organizing and indexing the multimedia content among various platforms includes Facebook, YouTube, Twitter and so on. Recently, a considerable attention is attracted by devising an effective visual category models in computer vision areas [5, 6]. The traditional classification methods work well, when the availability of sufficient labelled training images is present in the database. The major assumption of classification learning is that the distributions of training and testing sample data should be identical.

Researchers used various traditional frameworks of deep learning techniques like Deep Belief network and Convolutional Neural Network (CNN) [7-9] to classify the training data, which is widely

used in face recognition [10], image classification [11], digit recognition [12], and other applications. The most important criterion for an effective CNN technique is to have an enormous amount of training data. But, abundant data are required for adequate training of CNN, which leads to high computational complexity [13]. In real-world applications, the samples require high cost for obtaining the labelled samples by annotators, when the labels are scarce. Therefore, researchers focused on further study of TL [14] to solve above-mentioned problems, where the similarity between source domain and target domain are identified. The data dependence of target domain is illustrated by using transfer information of source domain. In this research work, the investigation is conducted on cross-domain STL-10 database for classifying the target object images into different types effectively. The various domain data are correctly classified by using Data Augmentation (DA) based approach with TL and pre-trained networks. In this experimental analysis, a two-class classification task is performed to transfer the knowledge from one domain to another by using TL approach. The advantages of using TL approach is that the model performs effectively on both large scale and small scale database because of its training process. The classification is carried out efficiently even in the imbalance database, because it transfers the knowledge from large scale database. Therefore, the cost of label is reduced by using the TL with DA in this research study. The experiments proved that the developed method achieved better performance than the existing TL techniques.

This research work consists: Section 2 describes the survey of existing techniques with its advantages and limitations. The brief explanation of TL is presented in Section 3, and the proposed Inception based TL method is depicted in Section 4. The validation of the Inception based TL method against existing techniques using standard dataset is given in Section 5. Finally, the conclusion of the research work with future development is represented in Section 6.

2. Literature review

In this section, the discussion of existing techniques is presented, which are used in the TL for classifying the data samples by using various neural network techniques. The advantages and limitations of the existing classification techniques is also illustrated.

Y. Liu, Y. Peng, K. Lim, and N. Ling, [15] designed an image database retrieval algorithm according to feature fusion and TL framework to address the over-fitting issues. Initially, the semantic features of the images were extracted by using fine-tuning of pre-trained CNN. The computation complexity was reduced by a dimensionality reduction algorithm called Principal Component Analysis (PCA). The retrieval accuracy was improved by fusing the extracted semantic features with traditional low-level visual features. The experiments were conducted on two datasets including GHIM-10K and Crime Scene Investigation Image database (CSID) in terms of mean average precision. But, the retrieval accuracy was less in some categories due to low-level features.

R. S. Kute, V. Vyas, and A. Anuse, [16] developed an approach for Component-based Face Recognition as CBFR and association using CNN under TL for the application of forensic. The components of the face were classified by demonstrating the knowledge, which was gained from complete face images. Association and recognition process were considered as the most important components of face such as nose, lips and ears, because these components didn't change during different expressions and poses. The training time of the system was reduced by the association between complete face and its components, hence this application was used for the recognition of partial face, holistic face as well as a component face. The experiments were carried out on standard Faces94 database and the results stated that this approach provided efficient results for variations in facial expressions. The complexity of CNN was high due to a large number of hidden layers used in the CBFR algorithm.

B. Xie, Z. Duan, B. Zheng, and L. Liu, [17] predicted various types of target objects by developing the Deep Neural Network (DNN), which was composed of Sparse Auto-Encoders (SAE) with unsupervised feature learning. The convolutional SAE was proposed to provide the cross-domain feature learning scheme for recognizing the target objects. According to correlation analysis, a feature selection method was developed to reduce the computational cost of CNN for the extraction of global features. The experimental results stated that the recognition performance was improved and achieved higher accuracy with robustness. However, DNN-SAE method focused only on low level features namely colour, texture and edge of the data, where high-level features were ignored.

Y. Xu, J. Liu, Y. Zhai, J. Gan, J. Zeng, H. Cao, and R. D. Labati, [18] designed a novel method for facial expression recognition task to overcome the problems like illumination influence, attitude variations etc. Initially, the robust and discriminative

deep features were learned from the data by designing the Double Activation Layer-based CNN as DAL-CNN. To address the overfitting issues, a two-stage TL method was used, where the insufficient training data caused overfitting phenomenon. Finally, an active incremental learning method was developed to overcome the noise label problem of Internet data. The experiments were conducted on two public databases called SFEW2.0 and FER2013 to validate the effectiveness of DAL-CNN in terms of classification accuracy. In DAL-CNN, there were five hidden layers used, which increased the computational complexity of the algorithm.

J. C. Hung, K. C. Lin, and N. X. Lai, [19] implemented the Dense_FaceLiveNet framework by improving the FaceLiveNet network with high and low accuracy in the recognition of basic emotions. Two-phases of TL used in this approach, where two datasets such as JAFFE and KDEF were used for the identification of Dense_FaceLiveNet efficiency. The results indicated that the method achieved high accuracy in FER2013, JAFFE and KDEF database and demonstrated the effectiveness of method in recognition of learning emotions. The importance of data complexity was clearly demonstrated by designing the deep learning model, when the TL of this basic emotion model was used in FER2013. However, this method concentrated only on finding the facial expression data in the lab environment without considering the exceptional situations, which occurred in the real classroom environment.

D. Han, Q. Liu, and W. Fan, [20] designed a two-phase method to address the requirement of large number of annotated samples, where two-phase method combined the CNN based TL and web DA. The over-fitting problem of deep-CNN was reduced on small dataset. In addition, the hyper-parameters were tuned for network fine-tuning by applying Bayesian optimization algorithm, which was also used to solve the tuff problem. The experiments were conducted on six small datasets to validate the effectiveness of CNN. However, the process of DA consumes high time which was considered as the major limitation of the developed method. The extracted features from the pre-trained CNN was not ideal for identifying the image similarity in the same class.

3. Transfer learning

The unknown knowledge learned through existing knowledge is defined as the concept of TL [21]. The main aim of this learning is to identify the similarities between unknown knowledge and existing knowledge. Some knowledge provides high learning costs in overall data, because these domains are too abstract for learning. Hence, the assistance of learning using existing knowledge is considered an important task. For instance, when people write a program in Python language and this learning model are easily transferred that into JavaScript program. Finding the relevance between known and unknown knowledge is the core concept of TL and is also used for learning the new knowledge. In general, source domain is defined as existing knowledge and target domain is represented as unknown knowledge in TL. How to migrate the knowledge to target domain from source domain is studied by this learning. There are four major categories present in the TL namely model-based, relationship, features-based and sample-based, based on various learning methods.

- Model-Based TL: To adjust the model parameters, this learning combines the samples with model.
- Feature-based TL: The mapping between the target domain and source domain to the same space are involved by this learning.
- Sample-based TL: The weighted values of calibrated samples are applied by this TL, where the calibrated samples are collected from source domain.
- Relationship-based TL: The concept learning is mapped to target domain from source domain (i.e. knowledge migration) by relationship-based TL.

In this research study, model-based TL is considered mainly to adjust the parameters of proposed Inception based TL model. The next sub-section will briefly describe the working procedure of Inception based TL method.

4. Proposed methodology

The issues in insufficient labelled data are addressed by storing the knowledge that is obtained from trained samples and applied that knowledge to the testing samples, which is defined as TL. The conventional algorithms performed well for images or video analysis, when the training and testing samples are only independent and distributed identically. In the same feature space or sharing the same distribution, the testing and training samples of machine learning algorithms provides better results.

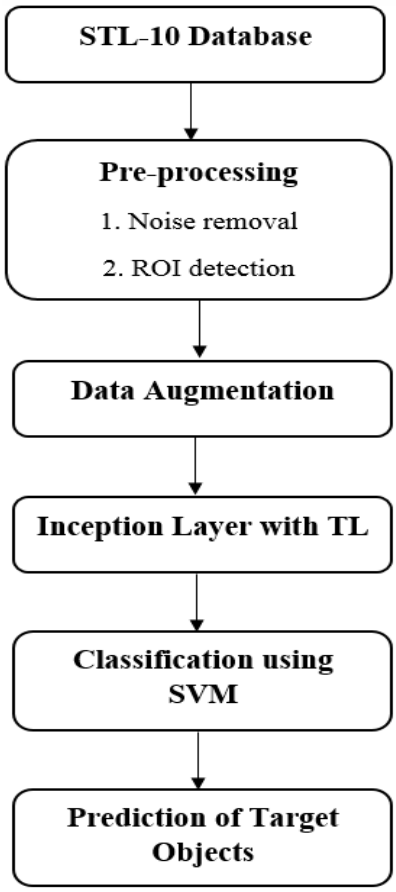

Figure.1 Working procedure of proposed methodology

The cross-domain learning problems are solved by TL, which is used to extract the useful information in a related domain from data and transferred them to target domains. In this research study, an efficient technique based on TL was developed for analysing the image and text database. There are different steps included in the proposed technique that consist of pre-processing, DA, inception layer and classification, which is illustrated in Fig. 1.

4.1 Data collection and pre-processing

Initially, the prediction of target objects is carried out by collecting the data samples from publicly available datasets. In this research study, STL-10 database is used for validating the effectiveness of proposed Inception based TL method. Before applying the DA, it is important to pre-process the data, because it may contain noises that leads to poor performance. In this study, noises are removed by normalization and Region of Interest (ROI) is identified using Hough Transform (HT).

Once the data has been extracted and divided into training, test, and validation, then the method normalized variables to a [0; 1] interval in order to avoid effects of scale in deep learning architecture. The normalization method employed for the dataset can be observed in the Eq. (1),

$$V_b = \frac{a_b - min_{a_b}}{max_{a_b} - min_{a_b}} \quad (1)$$

Where, a value used for normalizing the b is described as a_b, minimum valor registered for this variable is illustrated as min_{a_b} in the training set. Finally, maximum valor registered for this variable is described as max_{a_b} . When the noises are removed, HT is used for finding the ROI of data, because it is an effective method to detect the curve and straight lines. The boundaries of data and background are obtained from the pre-processed edge images and these images are randomly distributed. Therefore, ROI is identified by using HT algorithm.

4.2 Data augmentation

To expand the dataset, DA techniques are used and implemented in various ways like adding auxiliary variable, generating the data based on generative model, simulation and transformation of linear or non-linear data. The simplest approach for DA in image processing field is to add the noise and apply the affine transformations, i.e. shear, zoom, colour perturbation, flips and translation on the standard datasets. To overcome the problem of over-fitting, basic forms of DA are used widely on small-scale datasets. To improve the final results for classification, DA techniques are used in this research work. However, the effects of DA are reduced, when the experiments are conducted on complicated images like scenes from indoor or outdoor and images with background. The two phases of the proposed method are explained as follows.

4.3 Phase 1: Classification using pre-trained networks

The classification experiments are performed by using the pre-trained model, here inception network is used as pre-trained model. It is first introduced by GoogleNet and there are various versions presents in inception model with time span namely inception-2, inception-3 and inception-4. Along with batch normalization, factorization as main idea is used in inception-3 model that are considered as a pre-trained network in this research study.

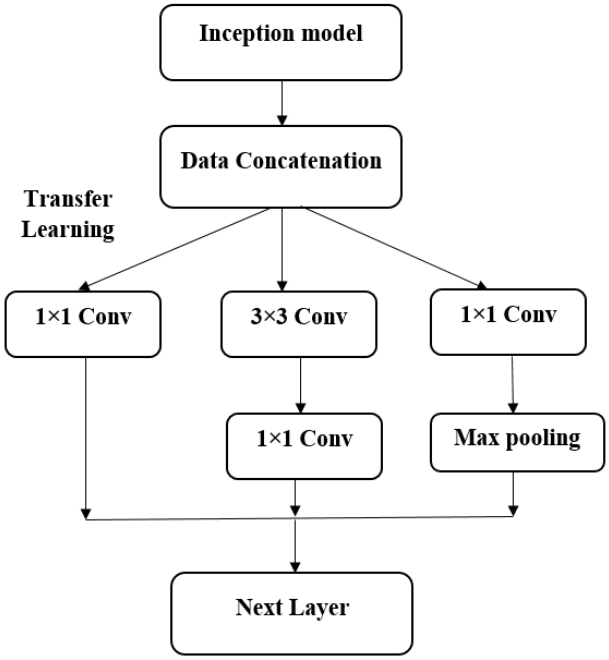

Figure.2 Inception model's working procedure

In general, 42-layers are present in inception-3 model with fewer parameters and the computational efficiency is also high when compared with other two models. Fig. 2 shows the working procedure of inception model. In this structure, the inputs to the next time step are the sums of the convolutional outputs of the present time step and previous time steps. The same operations are repeated based on the time steps considered. This helps to strengthen the extraction of the target features by using TL. In the particular pooling layer, 3×3 average pooling with stride 1×1 is applied by keeping the border size same, resulting in output samples with the same dimensions as the inputs. The overlapping average pooling technique helps in the regularization of the network. To describe the operations in inception layer, consider a pixel collated at (i, j) of a particular input sample on the kth feature map in the convolutional network. This is the output $y_{ijk}(t)$ at time step t. The output can be expressed as in Eq. (2):

$$y_{ijk}(t) = \left(w_k^f\right)^T x_f^{(i,j)}(t) + b_k \quad (2)$$

Where $x_f^{(i,j)}(t)$ and w_k^f are the inputs and weights for a standard convolutional layer, and b_k is the bias. The final output for the layer at time step t as shown in Eq. (3):

$$z_{ijk}(t) = f\left(y_{ijk}(t)\right) = \max(0, y_{i,j,k}(t)) \quad (3)$$

Where f is the standard Rectified Linear Unit (ReLU) activation function. The Local Response Normalization (LRN) function is applied to the outputs of the ICNN block. Eq. (4) shows that function as

$$y = norm(z_{ijk}(t)) \quad (4)$$

The outputs of the ICNN block with respect to the different kernel sizes and average pooling operations are defined as $y_{1x1}(x), y_{3x3}(x), and\ y_{1x1}^p(x)$. The final output y_{out} of the ICNN-block can be described as in Eq. (5):

$$y_{out} = y_{1x1}(x) \oplus y_{3x3}(x) \oplus y_{1x1}^p(x) \quad (5)$$

Here $\oplus$ represents the concatenation operation with respect to the channel axis of the output samples. Finally, a soft-max or a normalized exponential function layer is used at the end of the architecture. For an input sample x and a weight vector W, and K distinct linear functions the softmax operation can be defined for i^{th} class as follows in Eq. (6):

$$p(y = i|x) = \frac{e^{x^T} w_i}{\sum_{k=1}^{K} e^{x^T} w_k} \quad (6)$$

From the Eq. (6), the important features from the dataset are obtained. In the second phase, all the features from the inception layer are extracted and classified using SVM classifier, which is described in following section.

4.4 Phase 2: Feature extraction and transfer learning

The pre-trained inception model provided the important features that are fine-tuned with DA techniques. The target objects are predicted by classifying the inception features in a right order using TL. The extracted features are classified by using the SVM classifier in the research study. A separate hyper plane is used to characterize the SVM classifier, which is also described as discriminative approach. The regression and over-fitting issues are avoided due to the benefits of SVM classifier like generalization ability. The kernel tricks are effectively used by managing the non-linear data. Here, the input feature values are trained and tested to generate an optimal model, which is the main objective of SVM classifier. SVM is highly

used in several applications includes image retrieval, computer vision, medical image processing, signal processing, etc., because it processes the high dimensional data. According to the vapnik–chervonenkis theory and structure principles, the SVM classifier works effectively to resolve the two-class problem. The following equation $W.x + a = 0$ describes the mathematical formula of linear discriminant function. The following Eq. (7) describes the optimal hyper-plane, which is used to differentiate the samples without noise of two groups.

$$pi[W.x + a] - 1 \geq 0, i = 1,2,..D \quad (7)$$

Where, W is the weight of the parameter, x denotes the normal vector to the hyperplane and a is used to determine the offset of the hyperplane. Then, Lagrange function saddle point is used with Lagrange multipliers α_i to solve the optimization concern, therefore $\|W\|^2$ are reduce in the Eq. (8), which is used to indicate the optimal discriminant function.

$$f(x) = sign\{(W^*x) + a^*\} = sign\{\sum_{i=1}^{D} \alpha_i^* . pi(x_i^* - x) + a^*\} \quad (8)$$

Finally, the computational complexity is reduced in the higher dimensional data by changing the interior-product $(x_i^* - x)$ with a linear kernel function $k(x, x')$ in Eq. (8). Therefore, the mathematical Eq. (9) shows respective discriminant function to enhance the linear separability of assessed samples.

$$f(x) = sign\{\sum_{i=1}^{D} \alpha_i^* . pi. k(x, x_i) + a^*\} \quad (9)$$

Finally, the output of the SVM classifier is the prediction results of the target objects. The validation of the proposed method against several existing techniques are conducted and explained in the next sub-section.

5. Results and discussion

In this section, the extensive experiments are conducted on publicly available dataset for validating the effectiveness of Inception based TL method. The parameters include accuracy, precision, recall and f-measure used to estimate the Inception based TL method against existing techniques. Initially, the dataset description, parameter evaluation and analysis of Inception based TL method are briefly discussed as below.

5.1. Dataset description

The Inception based TL method uses the STL-10 database [22] for comparing their effectiveness against existing techniques, where unsupervised representation learning is used to design the STL-10 database. There are 100,000 unlabelled training images and 5000 labelled training images present in this database. The 10 classes are evenly distributed for supervised transfer and totally 8000 images are used for testing. The 400 epochs are used to perform the unsupervised training and additional 200 epochs for supervised transfer. The 96×96 colour images are presented in the data, where the images are resized randomly and 96×96 crops are extracted. Fig. 3 show the sample images from STL-10 database.

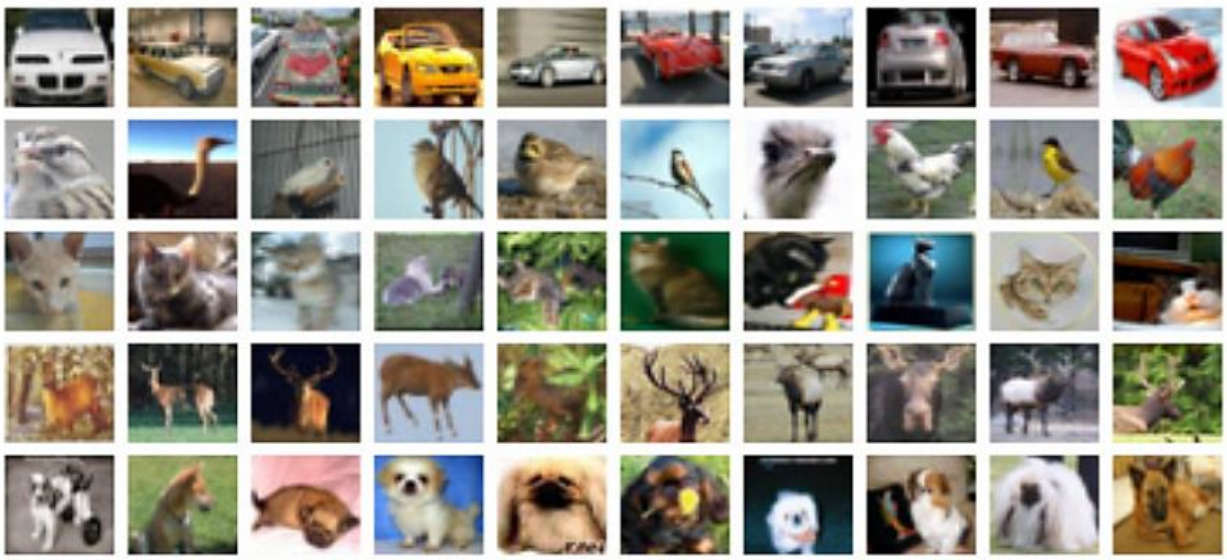

Figure.3 Sample images from STL-10 database

5.2. Parameter evaluation

When compared with existing techniques, the performance of Inception based TL method is validated by using various parameters. This section describes those parameters, which are used for checking the property of Inception based TL method. The practical and theoretical growth of the system is justified by the metrics, where some of the important parameters includes accuracy, precision, f-measure and recall. The mathematical expression for four important parameters is given from Eq. (10-13)

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (10)$$

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F-Measure = \frac{2TP}{2TP+FP+FN} \quad (13)$$

Where, True Positive is described as TP, True Negative is represented as TN, False Positive is presented as FP and False Negative is illustrated as FN.

5.3. Performance analysis of proposed method against existing techniques

In this sub-section, the validation of Inception based TL method is conducted against existing techniques namely DNN-SAE [17] and CNN-DA [20] on STL-10 database. Here, the training samples are considered as 80% and testing samples are considered as 20%.

Table 1. Validated results of proposed method against existing techniques

Methodology	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
DNN-SAE [17]	87.32	88.63	87.34	82.64
CNN-DA [20]	91.32	90.63	91.34	87.64
Inception Based TL	95.23	94.16	95.46	92.61

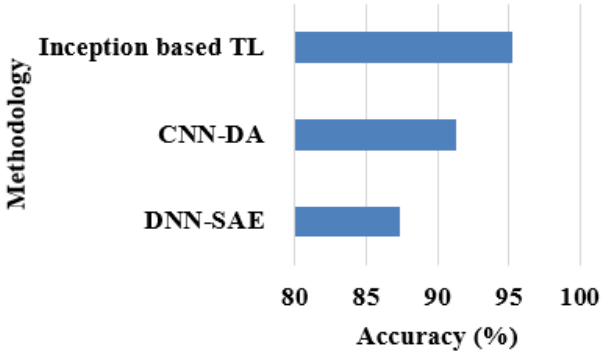

Figure.4 Performance of inception based TL Method in terms of accuracy

Table 1 presents the experimental results of Inception based TL method in terms of accuracy, precision, recall and f-measure. Fig. 4 shows the graphical representation of Inception based TL method against existing techniques by means of accuracy. Table 1 and Fig. 4 clearly state that the Inception based TL method achieved better performance in terms of all the parameters against existing techniques. Initially, the existing technique DNN-SAE achieved only 87.32% accuracy and the existing CNN technique used DA to improve the accuracy level, which leads to 91.32% accuracy. But, the pre-trained CNN failed to extract the efficient features for large scale database. To overcome this issue, the Inception based TL method is developed for both small-scale and large-scale dataset and achieved nearly 96% accuracy. The graphical representation of Inception based TL method in terms of precision and recall is shown in Fig. 5. In the analysis to precision performance, the existing techniques DNN-SAE achieved 88.63% and CNN with DA achieved 90.63%.

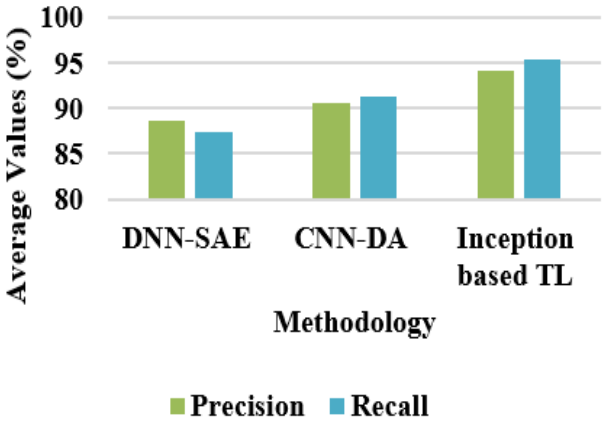

Figure.5 Analysis of inception based TL Method against existing techniques

In order to improve the precision performance, the Inception based TL method used multiple features from multiple filters, which will automatically increase the network performance. Therefore, the Inception based TL method achieved 94.16% precision, which is illustrated in Fig. 4. In addition, the DNN-SAE technique provides poor performance in recall analysis due to extraction of low-level features. Even though the DNN-SAE method achieved robustness in recognition, the method provides low accuracy with recall value i.e. 87.34%. But, the Inception based TL method achieved nearly 96% recall in STL-10 database by incorporating the low-level and high-level features of the data. Finally, Fig. 6 shows the validated results of Inception based TL method against DNN-SAE and CNN with DA using f-measure.

From the Fig. 6, the results showed that the Inception based TL method achieved higher performance by means of F-measure against DNN and CNN. The DNN-SAE achieved only 82.64% f-measure and CNN with DA achieved nearly 88% f-measure, where the Inception based TL method achieved f-measure of 92.61%. This is due to the usage of inception architecture in Inception based TL method. All the architecture includes DNN and CNN performed the convolution on the channel and spatial wise domain together, which is prior to inception. But, in the Inception based TL method, the inception block performed cross-channel correlations in 1x1 convolution itself and ignored the spatial dimensions, which is used to improve the performance of Inception based TL method. Finally, in 3x3 and 5x5 filters, the inception layer followed the cross-channel and cross-spatial correlations. Therefore, the Inception based TL method achieved better performance than other existing techniques, namely CNN-DA and DNN-SAE.

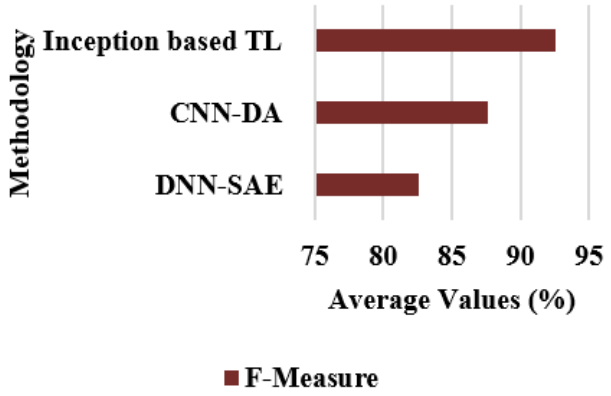

Figure.6 Performance analysis of proposed method in terms of F-measure

Table 2. Comparative analysis of inception based TL method against existing technique

Methodology	Database	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
DNN Network in Lower-level visualizing features	STL-10 Database	80.21	83.33	80	86
DNN Network in Higher-level visualizing features		82.15	91.75	84	87
Inception based TL		**95.23**	**94.16**	**95.46**	**92.61**

5.4 Comparative analysis of inception based TL method

In this section, the performance of Inception based TL method is compared with existing technique in terms of all the parameters, which is shown in Table 2. Y. Chen, H. Meng, X. Wen, P. Ma, Y. Qin, Z. Ma, and Z. Liu [23] designed a hybrid deep network model to extract the higher layer visualizing features using the combination of SAE, CNN and regression model. The TL was introduced in SAE model for obtaining the cross-domain higher-level features to classify the target-domain sample objects. These data were given as input to CNN model for acquiring global visualizing features. The experiments were conducted on STL-10 database to validate its effectiveness. However, the method failed to extract the high level features of the target domain, when the background is complex.

According to higher layer feature extraction, the performance of DNN algorithm provides better performance than lower layer visualizing features as shown in Table 2. This is due to the introduction of TL in higher layer feature extraction, where the lack of training process is effectively improved by TL for small sample target objects. However, if the background is complex to classify target objects, the DNN with higher layer feature extraction method provides poor performance. In order to overcome the issue, this research study introduced the inception layer in this approach, where it effectively avoids degradation of classification performance. To represent any function, Inception based TL network

with a single layer is sufficient which is based on the universal approximation theorem.

6 Conclusion

In the target domain of TL, the learning problem is solved by using the training data in the source domain with various distributions. According to availability of a huge amount of labelled data, traditional machine learning algorithm trained a model in the same feature space. But, often the labelled data are scarce and expensive to obtain in domain adaptation. In order to address the issues in a deep learning network, the research study developed an Inception based TL with SVM. There are two phases used in the Inception based TL method, where the effect of DA is investigated on the pre-trained inception layer in the first phase. Then, the features are extracted from all the layers of inception and learned by the SVM in second phase. The experiments are conducted on STL-10 database to validate the effectiveness of inception based TL in terms of accuracy, precision, recall and f-measure. The results stated that the Inception based TL method achieved accuracy of 95.23%, precision of 94.16%, f-measure of 92.61% and recall of 95.46% against existing CNN and DNN with SAE. In future work, it is important to use the effective ensemble feature extraction techniques to avoid the overfitting of the training data.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision, and project administration, have been done by 2nd author.

References

[1] J. T. Wang, G. L. Yan, H. Y. Wang and J. Hua, "Pedestrian recognition in multi-camera networks based on deep transfer learning and feature visualization", *Neurocomputing*, pp. 166-177, 2018.

[2] F. Ozyurt, "Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures", *The Journal of Supercomputing*, pp. 1-19, 2019.

[3] P. Chen, P. Li, Q. Li and D. Zhang, "Semi-Supervised Fine-Grained Image Categorization Using Transfer Learning With Hierarchical Multi-Scale Adversarial Networks", *IEEE Access*, Vol. 7, pp. 118650-118668, 2019.

[4] F. Liu, X. Xu, S. Qiu, C. Qing and D. Tao, "Simple to complex transfer learning for action recognition", *IEEE Transactions on Image Processing*, Vol. 25, No. 2, pp. 949-960, 2015.

[5] S. J. Pan, X. Ni, J. T. Sun, Q. Yang and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment", In: *Proc. of the 19th international conference on World wide web*, 2010.

[6] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G. R. Xue, Y. Yu and Q. Yang, "Heterogeneous transfer learning for image classification", In: *Proc. of Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

[7] J. Li and Z. Wang, "Real-time traffic sign recognition based on efficient CNNs in the wild", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 3, pp. 975-984, 2018.

[8] M. Zhang, L. Wei and Q. Du, "Diverse region-based CNN for hyperspectral image classification", *IEEE Transactions on Image Processing*, Vol. 27, No. 6, pp. 2623-2634, 2018.

[9] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification", *IEEE Transactions on Image Processing*, Vol. 26, No. 10, pp. 4843-4855, 2017.

[10] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification", In: *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 1701-1708, 2014.

[11] A. Krizhevsky, I. Sutskever and G. E. Hinton. "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing systems*, 2012.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov and A. Rabinovich, "Going deeper with convolutions", *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.

[13] Y. Zhai, J. Liu, J. Zeng, V. Piuri, F. Scotti, Z. Ying and J. Gan, "Deep convolutional neural network for facial expression recognition", *In International Conference on Image and Graphics, Springer, Cham*, pp. 211-223, 2017.

[14] R. Zhu, T. Zhang, Q. Zhao and Z. Wu, "A transfer learning approach to cross-database facial expression recognition", *IEEE*

International Conference on Biometrics (ICB), pp. 293-298, 2015.

[15] Y. Liu, Y. Peng, K. Lim and N. Ling, "A novel image retrieval algorithm based on transfer learning and fusion features", *World Wide Web*, Vol. 22, No. 3, pp. 1313-1324, 2019.

[16] R. S. Kute, V. Vyas and A. Anuse, "Component-based face recognition under transfer learning for forensic applications", *Information Sciences*, Vol. 476, pp. 176-191, 2019.

[17] B. Xie, Z. Duan, B. Zheng and L. Liu, "Research on Target Object Recognition Based on Transfer-Learning Convolutional SAE in Intelligent Urban Construction", *IEEE Access*, Vol. 7, pp. 125357-125368, 2019.

[18] Y. Xu, J. Liu, Y. Zhai, J. Gan, J. Zeng, H. Cao and R. D. Labati, "Weakly supervised facial expression recognition via transferred DAL-CNN and active incremental learning", *Soft Computing*, pp. 1-15, 2019.

[19] J. C. Hung, K. C. Lin and N. X. Lai, "Recognizing learning emotion based on convolutional neural networks and transfer learning", *Applied Soft Computing*, Vol. 84, pp. 105724, 2019.

[20] D. Han, Q. Liu and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation", *Expert Systems with Applications*, Vol. 95, pp. 43-56, 2018.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning", *IEEE Transactions on knowledge and data engineering*, Vol. 22, No. 10, pp. 1345-1359, 2009.

[22] A. Coates, A. Ng and H. Lee, "An analysis of single-layer networks in unsupervised feature learning", In: *Proc. of the fourteenth international conference on artificial intelligence and statistics.* 2011.

[23] Y. Chen, H. Meng, X. Wen, P. Ma, Y. Qin, Z. Ma and Z. Liu, "Classification methods of a small sample target object in the sky based on the higher layer visualizing feature and transfer learning deep networks", *EURASIP Journal on Wireless Communications and Networking*, Vol. 1, pp. 127, 2018.