An Improved Approach for Multi-Task Feature Image Classification using Hybrid GA-SIFT

Vandna Prajapati
I.T. Dept
S.A.T.I
Vidisha, India

Anil Suryavanshi
Assistant Prof. (I.T.)
S.A.T.I
Vidisha, India

ABSTRACT

Here in this paper an efficient technique for the Image Classification is proposed using Optimization of SIFT Algorithm by Genetic Algorithm. The Proposed Procedure implemented here is used for the Classification of Single Task as well as Multiple Task Features from the Image and classification is done. The Experimental results achieved on numerous datasets such as MIR Flickr, NUS Datasets shows the recital of the planned methodology. The algorithm provides High Precision and recall rate as well as more number of features extracted from the image with High Accuracy.

Keywords

SIFT, Genetic Algorithm, Image Classification, Multi-Task Feature, MIR Dataset, NUS Dataset.

1. INTRODUCTION

Recently there has been a thrust to build algorithms for architectures which focus on one of machine learning's original goals in creating artificial intelligence [1, 2], namely in the aptitude to perform compound intelligence tasks such as visual and auditory perception natural language processing, planning, and control. To achieve such tasks, deep architectures that inherently and efficiently model these domains have been proposed. Highly engineered shallow architectures can provide excellent performance over limited domains, where image recognition algorithms can meet and surpass human ability. Regrettably, applying these same methods to different domains or even different data-sets is often ineffectual. Deep learning is concerned with algorithms which can learn such representations with negligible human interference and little prior knowledge of the problem domain [2, 3]. This lack of hand-engineering makes the architectures more generalizable. As a subfield of mechanism learning, deep learning utilizes multi-layered i.e. deep architectures with typically more than two or three layers or stages to learn high-level abstractions as new features through the arrangement of subordinate level features [1]. These abstractions may be considered as concepts that underlying the natural order of the data which ideally become more invariant to small, local changes while stepping up the architecture. Difficulty in model construction and training is a common theme among all deep architectures. It is necessary to define or learn both the structure of the model and an often large parameter space covering this structure where objective functions are likely not convex. At their core, deep knowledge mockups use unsupervised education approaches to optimize these parameters and are considered data-driven approaches, though the use of a supervised signal may be helpful to guide proper representations.

Multi-task learning Evgeniou & Pontil [4] methods aim to simultaneously learn classification/regression models for a set of connected tasks. This characteristically leads to improved models as associated to a learner that does not explanation for job associations. The objective of multi-task learning is to get better the presentation of learning algorithms by learning classifiers for multiple jobs mutually. Various multi-task learning algorithms take for granted that all knowledge jobs are associated. In realistic applications, the jobs may demonstrate a more complicated group arrangement where the representations of jobs from the similar collection are quicker to each other than those from a unusual group. There have been several works along this line of research Zhou et al. [5], known as clustered multi-task learning (CMTL). Moreover, most multi-task learning formulations assume that all tasks are relevant, which is however not the case in many real-world applications. Robust multi-task learning (RMTL) is aspired at recognizing inappropriate (outlier) jobs when teaching from multiple tasks the only related theoretical work is that in Maurer et al. [6], where only theoretical bounds are provided on evaluating the generalization error of dictionary learning for multi-task learning and transfer learning. Multi-task education has established substantial courtesy in the computer hallucination community and has been successfully applied to many computer vision problems. It makes the learning process more efficient, reduces the chance of over-fitting, and improves the generalizability of the model [7]. Feature selection and feature alteration are the two main methods used for article withdrawal; in the former, a subset of landscapes is nominated from the innovative, while in the concluding the unique characteristic is altered into a novel characteristic space. The previous is frequently the preferred method [8]

1.1 Scale Invariant Feature Transform

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and define local landscapes in descriptions. The algorithm was published by David Lowe in 1999.

Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video chasing, individual documentation of wildlife and match moving.

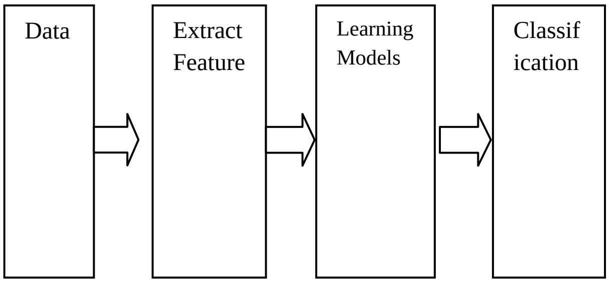


Figure 1: Standard Object Classification Pipeline.

A popular framework to classification in Computer Vision, see Figure 1, is based on representing objects as points in a high-dimensional feature space ("Data" block), and then performing some partitioning of the space into areas equivalent to the unusual module. characteristically, a set of n unusual measurements (we call this measures "features", "Extract Features" block) are extracted from the image, and the result constitute an n dimensional vector representing the image. Partitioning of the space is then performed using different "Learning" techniques that are used to generate a model, useful for the "Classification" a more general view which includes both raw and processed data and, in short, all the inputs to some higher level decision making or classifying stage.

2. LITERATURE SURVEY

In this paper [9] author presents a new large margin multi-modal multi-task feature extraction (LM3FE) algorithm that successfully investigates the matching nature of unusual modalities gotten from numerous jobs. Particularly, LM3FE become skilled at a projection matrix for each modality that transforms the data from the new characteristic space to a concealed characteristic space. In the concealed space a weighted combination of all the altered characteristics is then used to forecast the ground-truth labels of the training data. The l2, 1-norm constraints on the forecast matrices create them appropriate for both characteristic selection and characteristic alteration. The forecast loss is chosen as the hinge loss for classification, and the margin maximization principle improves the forecast authority of the preferred or altered characteristics. All the forecast matrices of dissimilar modalities, the combination weights, and the prediction matrix are learned as a single optimization problem. This method adventures both the task connection and the opposite nature of different modalities for effective and strongly predictive feature extraction.

The planned LM3FE fits to multi-modal subspace education but uses a weighted modality combination training plan. In recent times, a multi-modal feature selection technique [10] has been suggested that discovers the correlations between unusual modalities by taking the tensor product of their characteristic spaces. On the other hand, this technique cannot handle the multiclass difficulty unsurprisingly and must train an SVM classifier to remove one characteristic at a time. The associations between different classes are consequently throw away and the training charge is very high.

At their core, deep learning models use unconfirmed learning approaches to optimize these parameters and are considered data-driven approaches, though the use of a supervised signal may be helpful to guide proper representations. Informally, this supervised signal can embody feedback on how well the architecture is modeling structure in the data and be used to fine-tune the representations that are learned from unlabeled examples. [11] Argues that the foremost challenge in training deep architectures is respecting the strong dependencies between parameters across layers. Modifying a feature mapping at a lower level of the architecture changes the input space (in both distribution and range) of the next layer.

Concept detectors provide a high-level semantic representation for videos with complicated contents, which inclines to benefit for developing powerful retrieval or filtering systems for consumer media Snoek & Smeulders [12]. In our case, we extract semantic indexing (SIN-MED) features of a video to predict the 346 semantic concepts existing in its key frames. SIFT is used to describe the information of images. Bag-of-words SIFT is used to train a model for each concept. Once we have the prediction score of each concept on each key frame, the key frame can be represented as a 346-dimensional feature indicating the determined concept probabilities. The video-level SIN MED feature is computed as the average of key frame-level SIN-MED feature.

To become skilled at a discriminative vocabulary for sparse coding a label dependable K-SVD (LC-KSVD) algorithm was suggested in Jiang et al. [13]. In totaling to using period labels of exercise data, the authors also associated label data with each dictionary article i.e. columns of the dictionary matrix to put into effect discriminability in scant cyphers throughout the vocabulary education process. More specifically, a new label steady restriction was introduced and combined with the renovation error and the organization error to form a unified objective function.

The work most thoroughly related to ours is done by Karayev et al. [14] in 2014. This paper tries to predict the style of an image. All the previous papers used different dataset whereas this paper uses the same dataset which we use, namely the Wiki paintings dataset. This paper achieves state of the art for painting style classification using features extracted from Deep Convolutional Neural Network. It also compares other feature extractors like GIST, Lab Histogram; Graph based pictorial saliency and Meta class binary features. The paper uses Wiki paintings dataset with 85K samples and Flickr photos dataset with 80K samples.

The paper title "In search of Art" [15] explores the opposite problem of finding objects in art, regardless of their style. However the technique used by the authors is same. Features are extracted from pre trained deep CNN and a Linear SVM is used to detect objects. The CNN is trained from natural images, but classification is done on art images. They say that these features are more useful than Improved Fischer vector features. The authors explain the merits of using CNN features which include rich color information vital in the area of art. Hand crafted descriptors based on Histogram of Oriented Gradients (HOG) and SIFT capture only gradients but not color information. CNN features capture both.

Chen and Hauptmann [16] treat the longitudinal and sequential domains separately. The MOSIFT descriptor contains two parts: describing the longitudinal province with histogram of gradient (HOG) and the temporal domain with histogram of optical flow (HOF) that captures the moved in the interest points. Firstly, a pair of frames is used to apply the normal 2D-SIFT algorithm and detect the distinctive interest points in appearance. Afterwards, the optical flow is utilized to filter those features with sufficient amount of motion or action. Secondly, similar to the pyramid of Gaussian, a pyramid of optical flow is constructed for each Gaussian pyramid. Then, local extrema is detected from the DOG pyramid if it contains motion information in the ocular movement pyramid. In determining a descriptor, factorization histogram is used for each kind of features separately with one difference in the dominant orientation. The optical flow does not involve orientation invariant.

Here they proposed ST-SIFT detector was evaluated in human exploit gratitude task by classifying the presented activity in the given video. The methodology we adopted is the Bag of Features (BOF) model implemented in the VlFeat toolbox [17], which is an open library that contains various algorithms and applications for computer vision. One of the provided applications is an image classifier using 2D-SIFT features.

The code was adapted to be used as action classification framework in video data. The first step is to extract the interest points from the spatio-temporal video cube using the proposed detector. Subsequently, the spatio-temporal districts around the attention points are described using the 3D-HOG descriptors.

3. PROPOSED METHODOLOGY

The Proposed Procedure realized here for the Image Classification works on the Following phases:

1. Take an input Training MIR or NUS Dataset of Images.
2. Apply Filtering on the Training Images.
3. Extract Features from the Image using Optimization of SIFT by Genetic Algorithm.
4. Train these Features using Back Propagation Neural Network.
5. Store the Features.
6. Take an Input Testing Image Dataset.
7. Filter each of the test Image.
8. Extract Features from the Image using Optimization of SIFT by Genetic Algorithm.
9. Compare the Extracted Features with the Stored Features.
10. Classify Images based on features.

3.1 Basic Steps of GA

1. Start
2. T1=0 (here T1 is the initialization time to start)
3. Initially provide the population of genetic p1(T1)

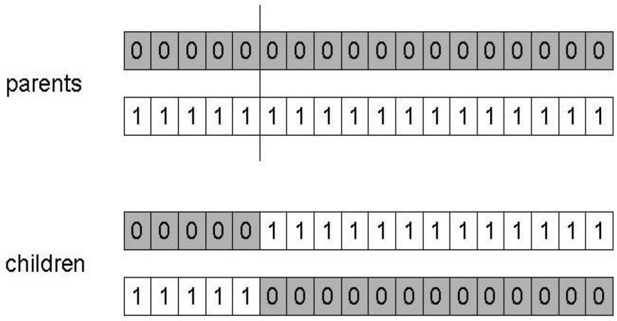
   (initialize a usually random population of individuals)
4. Calculate and estimate the fitness value p1(T1)
5. T1=T1+1
6. Check if the termination criterion satisfies
7. If yes then move and achieved move to step 10
8. Now choose p1(T1) from p1(T1-1)
9. Crossover both population p(T1)
10. Mutate these population p(T1)
11. Move to step 3
12. Output the greatest population and stop
13. End

3.2 Selection and Cross Over

Selection and crossover both of the real coded GA have been used to insure a steady convergent behaviour of the genetic algorithm. We had to build is the fine known trade off among investigation and utilization present in any search method including GA. The convergent development confident by selection and crossover should well-balance the broad investigation outcome achieves by our mutation operator. The selection method was selected as an alliance among binary tournament which has a constant and relatively high selection pressure (Miller 1996), with a K – elitist scheme (Bäck 1991) that assure the conservation of the K finest individuals. Fitness value of every pixel could be deliberate by enchanting the sum of average of all the intensity pixels of the block area. In GA Fitness function is use to verify the fitness of the pixel in this when child chromosomes can be generated then the value of child is superior to the fitness value otherwise these child chromosomes are rejected.

3.3 Mutation

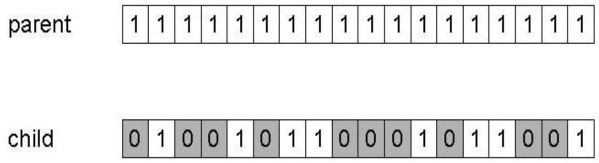
Mutation operator has been selected to assure elevated stage of variety into population. Introducing PCA-mutation in (Munteanu 1999b), and revealed that it has extremely fine capability in maintaining superior levels of variety. In short PCA-mutation operator can be define as: In genetic algorithm population X could be viewed as a cloud of N points in al-dimensional space, where N is the size of the population and l is the length of the chromosome.



a. Modify every gene separately with a probability p_m.

b. p_m is known as mutation rate

Typically between 1/population size and 1/chromosome_length.



3.4 Sift Feature Extraction

The Process of Piece Extraction since JPEG Images consuming SIFT Algorithm consists of Four Stages:

1. Detection of Scale- Space Extrema.

a. Novelty the topics, whose adjoining patches (with roughly measure) stay distinctive.

b. A guesstimate to the measure -normalized Laplacian of Gaussian.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \qquad (2)$$

Where, L remains the Laplacian Key Point with x row Pixels and y Column Pixel, I is the image, G is the Gaussian Parameter.

2. Localization of the Key Points in JPEG Image.

   a. There are still a lot of points; some of them are not good enough.

b. The whereabouts of key points might be not exact.

c. Eliminating edge points

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \qquad (3)$$

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \qquad (4)$$

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \qquad (5)$$

3. Elimination of the Key Points from the JPEG Image.

a. Such a fact has hefty foremost archath wart the edge then a lesser one in the erect bearing

b. The foremost curves can be deliberate since a Hessian function.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \qquad (6)$$

c. The Eigen values of H are comparative to the principal curves, so two Eigen values shouldn't diff too plentiful

4. By Orientation Assignment of the JPEG Image Pixels.

a. Assign an orientation to each key point, the key fact descriptor can be denoted virtual to this alignment and consequently accomplish invariance to twin alternation.

b. Compute degree and positioning scheduled the Gaussian smoothed images.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + L(x, y+1) - L(x, y-1))^2}$$

c. A histogram is designed by quantizing the bearing shooked on 36 silos;

d. Crests in the histogram relate to the bearings of the blotch;

e. For the alike scale and locality, there might be several key topics with different orientations;

5. Applying Key points Descriptor

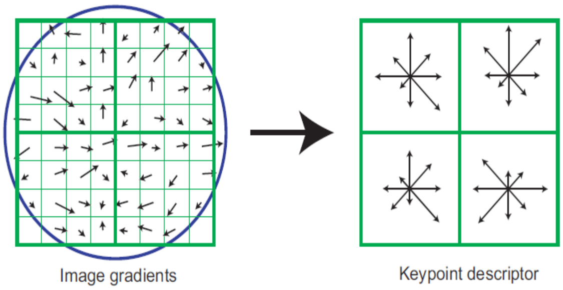


Figure 2. Key Points Extractor

a. Constructed on 16*16 blotches

b. 4*4 sub provinces

c. 8 bins in a piece sub province

d. 4*4*8=128 extents in overall.

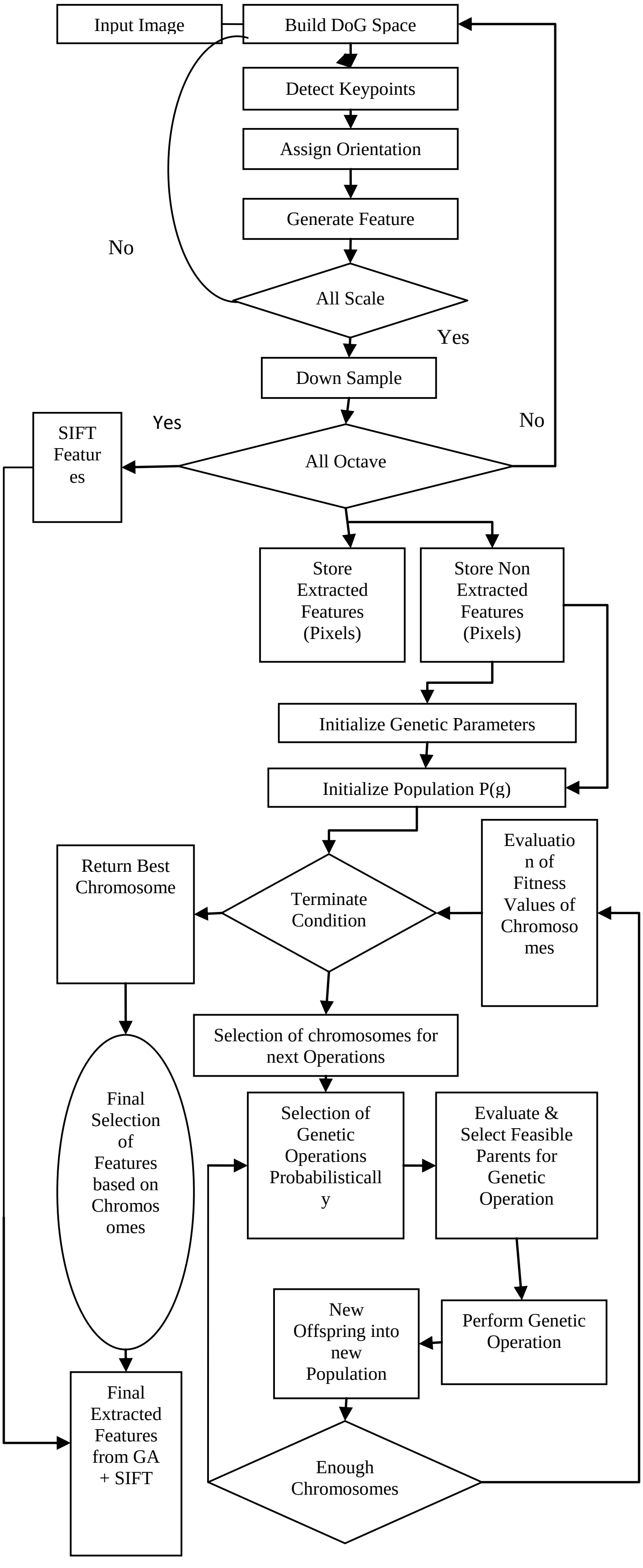

Input Image
Build DoG Space
Detect Keypoints
Assign Orientation
Generate Feature
No
All Scale
Yes
Down Sample
SIFT Features
Yes
All Octave
No
Store Extracted Features (Pixels)
Store Non Extracted Features (Pixels)
Initialize Genetic Parameters
Initialize Population P(g)
Return Best Chromosome
Terminate Condition
Evaluation of Fitness Values of Chromosomes
Selection of chromosomes for next Operations
Final Selection of Features based on Chromosomes
Selection of Genetic Operations Probabilistically
Evaluate & Select Feasible Parents for Genetic Operation
New Offspring into new Population
Perform Genetic Operation
Final Extracted Features from GA + SIFT
Enough Chromosomes

Figure 3. Flow Chart of Features Extraction using SIFT-GA

4. RESULT ANALYSIS

The Table shown below is the analysis and Comparison of Multi-Modal Image Classification and the Planned Procedure. The Table displays the Comparison on MIR Flickr Dataset for the percentage of Features Selected from the Images on the basis of which Classification is done on 20 Image Samples.

Table 1. Analysis of mAP for 20 Samples on MIR Dataset

	mAP for Samples=20	
% of Selected Features	**Existing Work**	**Proposed Work**
0.1	0.3895	0.4
0.2	0.4	0.425
0.3	0.414	0.428
0.4	0.427	0.436
0.5	0.417	0.425
0.6	0.429	0.439
0.7	0.43	0.45
0.8	0.429	0.44
0.9	0.414	0.42

The Table shown below is the analysis and Comparison of Multi-Modal Image Classification and the Planned Procedure. The Table displays the Comparison on MIR Flickr Dataset for the percentage of Features Selected from the Images on the basis of which Classification is done on 30 Image Samples.

Table 2. Analysis of mAP for 30 Samples on MIR Dataset

	mAP for Samples=30	
% of Selected Features	**Existing Work**	**Proposed Work**
0.1	0.43	0.45
0.2	0.432	0.453
0.3	0.436	0.457
0.4	0.438	0.459
0.5	0.44	0.45
0.6	0.436	0.453
0.7	0.435	0.451
0.8	0.432	0.445
0.9	0.43	0.443

The Table shown below is the analysis and Comparison of Multi-Modal Image Classification and the Planned Procedure. The Table displays the Comparison on MIR Flickr Dataset for the percentage of Features Selected from the Images on the basis of which Classification is done on 20,30 and 50 Image Samples.

Table 3. Analysis of mAP for 20,30 and 50 Samples on MIR Dataset

Methods	20	30	50
LM3FT	0.451+-0.008	0.462+-0.010	0.469+-0.009
Proposed Work	0.567+-0.008	0.473+-0.010	0.53+-0.009

The Figure shown below is the analysis and Comparison of Multi-Modal Image Classification and the Planned Methodology. The Table displays the Comparison on MIR Flickr Dataset for the percentage of Features Selected from the Images on the basis of which Classification is done on 20 Image Samples.

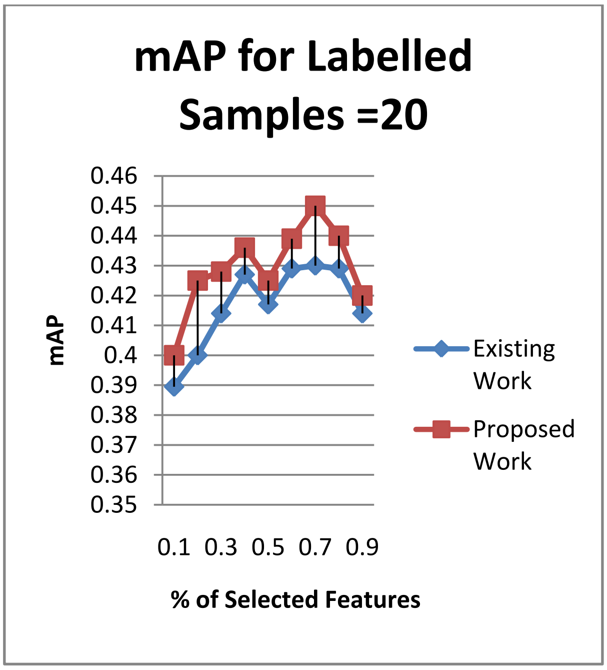


Figure 4. Comparison of mAP Accuracy on MIR Dataset for 20 Samples

The Figure shown below is the analysis and Comparison of Multi-Modal Image Classification and the Planned Methodology. The Table displays the Comparison on MIR Flickr Dataset for the percentage of Features Selected from the Images on the basis of which Classification is done on 30 Image Samples.

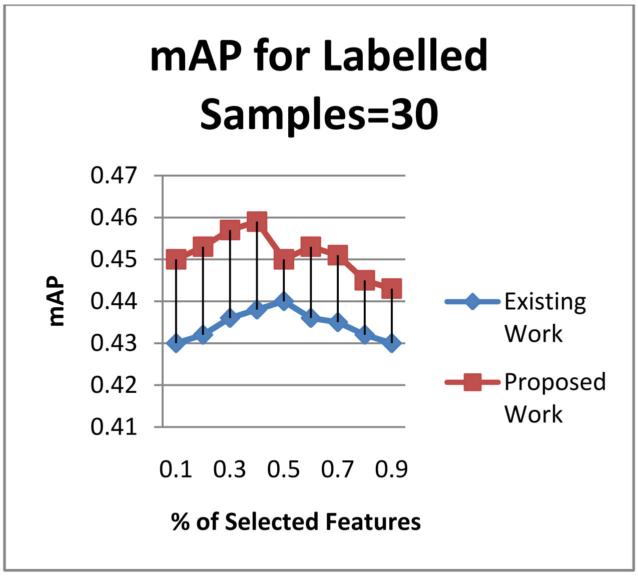


Figure 5. Comparison of mAP Accuracy on MIR Dataset for 30 Samples

5. CONCLUSION AND FUTURE WORK

Here the paper represents the work to explore multi-task dictionary learning approaches for complex event detection and in particular, and other learning technique for which very few solutions have been proposed in literature. From the user opinion of interpretation of interactivity possible expansions include: new active learning methods for multi scale classification; and enhancements on the visualization and annotation of areas by the user.

The methodology implemented here for the Classification of Images using Optimization of SIFT algorithm by Genetic Algorithm provides efficient classification of Single Task or Multiple Task Features from the Images. The Methodology is implemented for NUS and MIR Flickr Dataset and when results are compared with the Multi Model algorithm, the proposed algorithm provides efficient results in Comparison. Although the planned procedure implemented here for the Classification of Single and Multi-Task Features provides efficient results as compared to the existing LM3FE algorithm but there may be dome future directions such as Optimization of Genetic Algorithm using Particle Swarm Optimization and also the methodology can be implemented for HDR Images & other Datasets.

6. REFERENCES

[1] Y. Bengio, “Learning deep architectures for AI,” Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009. 1, 2, 24, 25, 26, 27, 31, 34, 35, 39, 68

[2] Y. Bengio and Y. LeCun, “Scaling learning algorithms towards AI,” Large-Scale Kernel Machines, vol. 34, 2007. 1, 34

[3] I. Arel, D. Rose, and T. Karnowski, “Deep machine learning - a new frontier in artificial intelligence research,” IEEE Computational Intel ligence Magazine, vol. 5, no. 4, pp. 13–18, Nov. 2010.

[4] Evgeniou, T. & Pontil, M. Regularized multi-task learning. In ACM SIGKDD international conference on Knowledge discovery and data mining, 2004.

[5] Zhou, J., Chen, J. & Ye, J. Clustered multi-task learning via alternating structure optimization. In Proceedings of the Conference on Advances in Neural Information Processing Systems, 2011.

[6] Maurer, A., Pontil, M. & Paredes, B.R. Sparse coding for multitask and transfer learning. In International Conference on Machine Learning, 2013.

[7] Z. Li, Y. Yang, J. Liu, X. Zhou, and H. Lu, “Unsupervised feature selection using nonnegative spectral analysis,” in Proc. 26th AAAI Conf. Artif. Intell., 2012, pp. 1026–1032.

[8] M. Masaeli, J. G. Dy, and G. M. Fung, “From transformation-based dimensionality reduction to feature selection,” in Proc. 27th Int. Conf. Mach. Learn., 2010, pp. 751–758.

[9] Yong Luo, Yonggang Wen, Dacheng Tao, Fellow, Jie Gui, and ChaoXu, “Large Margin Multi-Modal Multi-Task Feature Extraction for Image Classification” IEEE Transactions On Image Processing, Vol. 25, No. 1, January 2016.

[10] B. Cao, L. He, X. Kong, P. S. Yu, Z. Hao, and A. B. Ragin, “Tensor-based multi-view feature selection with applications to brain diseases,” in Proc. IEEE Int. Conf. Data Mining, Dec. 2014, pp. 40–49.

[11] D. Erhan, Y. Bengio, A. Courville, P. A. Manzagol, P. Vincent, and S. Bengio, “Why does unsupervised pre-training help deep learning?” The Journal of Machine Learning Research, vol. 11, pp. 625–660, 2010.

[12] Snoek, C. & Smeulders, A. Visual-concept search solved? IEEE Computer, 43, 76–78. 74, 75, 79, 84, 2010.

[13] Jiang, Z., Lin, Z. & Davis, L.S. Learning a discriminative dictionary for sparse coding via label consistent k-svd. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[14] Karayev, S., M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller (2013) “Recognizing image style," arXiv preprint arXiv: 1311.3715.

[15] Crowley, E. J. and A. Zisserman “In search of art," in Computer Vision-ECCV 2014 Workshops, Springer, pp. 54-70, 2014.

[16] Chen, M.-y and Hauptmann, A. Mosift: Recognizing human actions in surveillance videos. Transform, pages 1-16, 2009.

[17] Vedaldi, A. and Fulkerson, B. Vlfeat: an open and portable library of computer vision algorithms. In Proceedings of the international conference on Multimedia, MM '10, pages 1469-1472, New York, NY, USA. ACM, 2010.