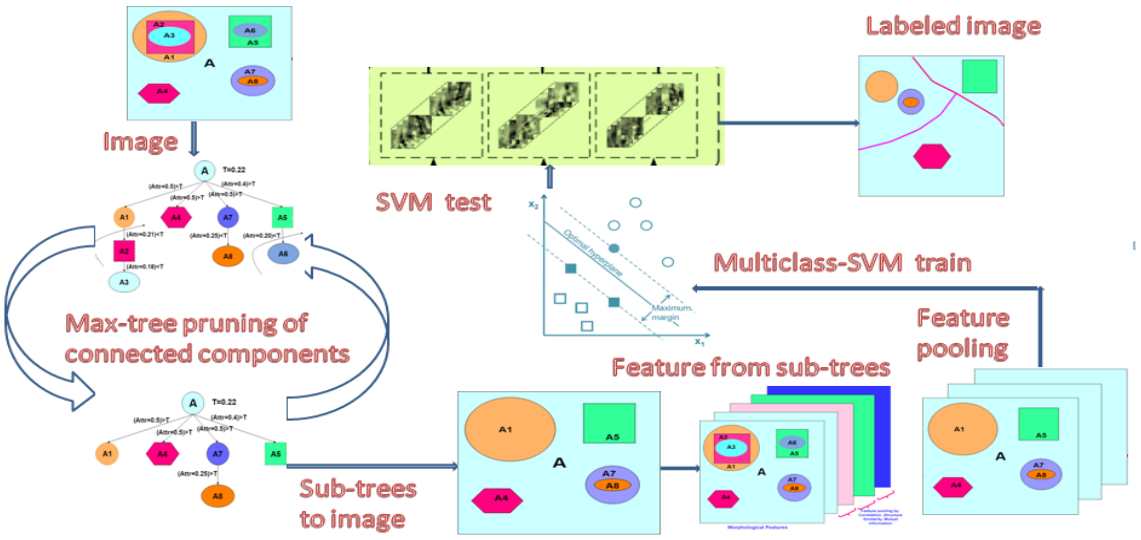
Chapter 7

LM-MFP: Large Scale Morphology and Multi-criteria Feature pooling based Image Labelling

Deep CNN based feature extraction is an uncontrolled operation that reduces the spatial resolution of features in which some of the features are redundant in nature. Morphological features are highly informative and have better resolution than CNNs. Therefore, we have studied a large-scale morphological feature extraction process that preserves the spatial resolution and also generate the massive feature set with different objects of an image. Due to enormous feature extraction, the features are redundant in nature. In scene labelling, the additional feature leads to massive computational load without contributing to accuracy. Therefore, we have exploited a novel feature selection approach that reduces the feature space by removing the redundant features. Subsequently, both semantic and categorical image parsing experiments have performed with the reduced feature space.

7.1 Introduction

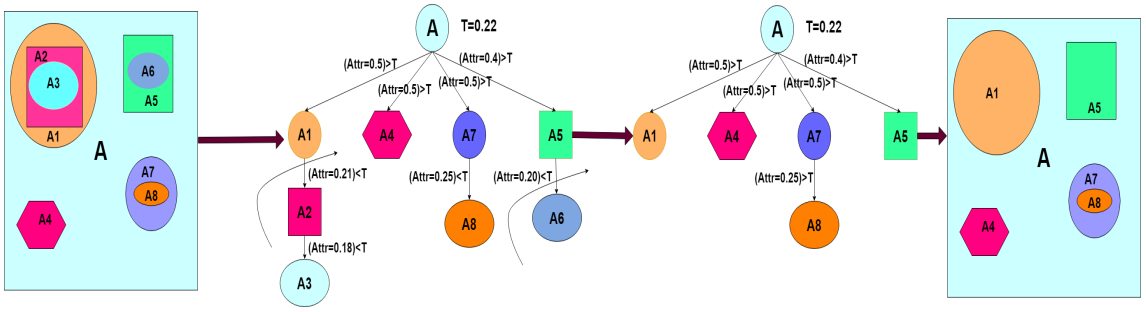
In this chapter, we have proposed a two-step robust LM-MFP method for pixel-wise semantic classification in a pixel-labeled environment of an image. In step-1, we have generated the morphological features from an image using attribute filters. Firstly, the



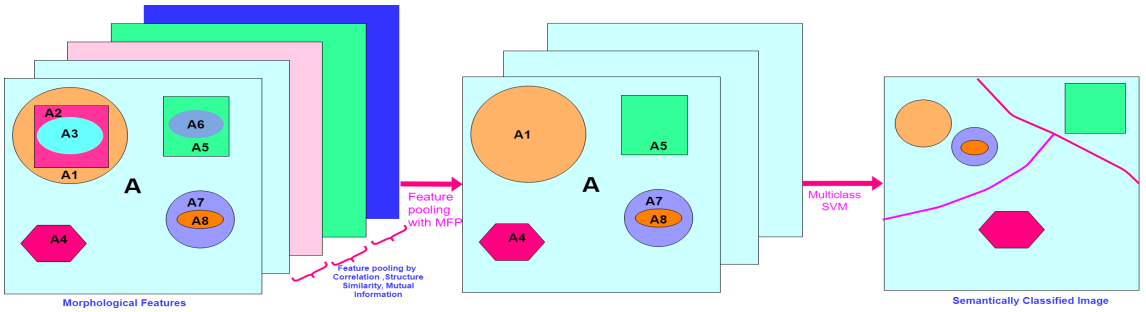
(A)

FIGURE 7.1: Methodology for LM-MFP method

image features based on some attribute like area, diagonal, and inertia has been extracted and arranged in the form of a min-tree and max-tree architecture. Then, we have decided some iterative threshold for tree pruning and obtained several distinct and redundant features. Although these features are the most informative, many of these features are redundant in nature. Therefore in step-2, we have proposed a feature pooling method to discard the redundant features using similarity measures. For the measurement of similarity, we have taken three parameters, i.e., correlation (ρ), structural similarity (SSIM), and mutual information (MI), to obtain the similarity between two features. Subsequently, we have made a hypothesis based upon the predicate that contains ρ, SSIM, and MI as a variable. Based on the hypothesis, redundant features have extracted, and a weighted average has taken for those features. This process has resulted in exclusive and non-redundant features space. The reduced feature space has used for pixel-wise semantic classification for labelled images. Finally, we have applied a multi-class SVM for training the model from the available pixel-label set. The complete pixels of an image has used for the prediction of image objects by using the trained SVM model. The proposed LM-MFP approach has achieved 90.96%, 89.03%, 90.57%, 91.12%, and 93.63% accuracy for highway, house, sheep, horse rider, and horse keeper images respectively obtained from sift-flow and pascal-voc datasets. Conclusively in the comparative analysis section, the Proposed LM-MFP method has observed to be state-of-art since it has significantly out-performed the previously proposed methods.



(A)



(B)

FIGURE 7.2: Process flow for LM-MFP method

7.2 Proposed Method:(LM-MFP)

In this section, we have introduced our two-step framework LM-MFP. In step-1, we have explained a robust and effective method for morphological feature extraction using attribute filters. These attributes have worked like a threshold for different morphological features. The extracted features are massive, and some of them are redundant. Although the morphological feature contains important information but redundancy increases the computation time. Therefore, we have introduced the feature pooling strategy based upon correlation(ρ), structural similarity(SSIM), and mutual information(MI) to reduce the computation time. Subsequently, these features have used in the semantic prediction process. Hence in step-2, a predicate based feature pooling has been used to merge the redundant features that we have obtained from step-1. Redundancy has measured based on ρ, SSIM, and MI. We have taken a bitwise average of redundant features to reduce the feature space significantly. Finally, we have used multi-class SVM to learn from the pooled feature space pixels and predict the entire pixel feature space to obtain the semantically classified output image. Step-2 has denoted as LM-MFP in our study. The complete process flow has shown in figure-7.2.

7.2.1 Morphological Feature Extraction

In this section, we have discussed details of large scale morphology to draw out the informative features iteratively [196, 193] by using a tree-based connected component merging. We have first divided the image into various connected components and merged the flat components. The connected components has generated by using morphological opening and closing operations. The connected components have created as:

$$I(X) = f_x O(X) \; where \; x \varepsilon O(X) \tag{7.1}$$

f_x and $O(X)$ is family of opening and opening with some structuring element O.

$$I(X) = \{x\} \; where \; x \varepsilon X \; and \; O(X) \tag{7.2}$$

f_x and $O(X)$ is family of opening and opening with some structuring element O. Similarly; we can define the closing operations. I(X) is point-wise connected regions by opening (O) operations. The connected components are used to make a max tree. We can also construct the min tree. The tree-building method is implemented using a queue data-structure in algorithm-1 , where **add**, **first**, and **empty** are the three queue functions that **add**, **extracts first**. and **return true** if pixel at connected component level priority C. There is one more function **nodes(c)**, which is number of nodes at priority c. **Gray(a)** is the function representing the original value of pixel 'a' and **state(p)** is one of the three states of pixel namely: (1)**not visited**, (2)**in the queue**,or (3) **assigned to node n of level c**. If the state of pixels is not analysed then we put it into the queue. In algorithm-6, we first visited the node according to the state and then generated the tree with the parent and child. The graph generation algorithm-6 was firstly introduced by [208]], which is computationally very efficient. We have obtained a robust Max Tee architecture of connected components of the image. Further, we aim to prune this tree and generate the other sub-trees where flat components have merged and transformed these sub-trees back to the image format. For pruning of the tree, we have taken two criteria: (1) diagonal(d) of bounding box (2) moment of inertia(i) at each graph node. We have applied a threshold selection method on Max tree, and multiple feature images have obtained by using various sub-trees obtained by threshold-based tree pruning. We have modified the threshold selection methodology from [194]. The algorithm-7 explains the iterative threshold on the max tree of connected components. In algorithm-7, we have computed and sorted the diagonal, inertia criteria at

Algorithm 6 Max Tree from connected components of an Image

Inputs $\leftarrow$ *connected component of image with priority* (c)
Output $\leftarrow tree(T)$
while do $empty(c)$
  $a \leftarrow first(c)$
  $state(a) \leftarrow node(c)$
  for $b\varepsilon$ $8-neighbourhood(a)$ **do**
    if $state(b) ==' not\ assigned'$ **then**
      $add(Gray(b),b)$
      $state(b) \leftarrow (in\ the\ queue)$
      $nodes(Gray(a)) \leftarrow (True)$
      if $Gray(b) > Gray(a)$ **then**
        tree(T)= Gray(b)
        Repeat the whole algorithm until T==c
      end if
    end if
  end for
end while
$nodes(c) \leftarrow nodes(c)+1$
$T \leftarrow c+1$
while do$T > 0 \&\& nodes(T) ==' False'$ T=T-1
  if $T > 0$ **then**
    $p \leftarrow (nodes(c)-1)$
    $q \leftarrow (nodes(T))$
    $parents\ of\ node Z_c^p \leftarrow Z_T^q$
  else
    Z_c^p is a root node
    $nodes(c) \leftarrow' False'$
    Return T
  end if
end while

each node of the tree as:

$$\mu_{i\,i=1}^{n} \leftarrow Sort\ (F(T)) \tag{7.3}$$

For n number of different criterias(diagonal and inertia), the thresholds are derived as:

$$for\ (i=1\ to\ n\ )do\ \{[p_{\phi_{\mu_i}}(I) \leftarrow Attr(\phi_{\mu_i}(Tr))]\} \tag{7.4}$$

Some stopping criteria has been decided to limit the generation of thresholding parameters as in equation-7.1 to 7.8:

$$compute(p_{\phi_{\mu_i}}(I),N) \tag{7.5}$$

$$\hat{p} \leftarrow interpolate\ (p_{\phi_{\mu_i}})\ on\ threshold\ (\lambda) \tag{7.6}$$

$$err_N \leftarrow (1 - RMS(p, \hat{p})) \tag{7.7}$$

$$if\ (N > 1) then\ calculate\ error\ and\ position\ feature \tag{7.8}$$

Subsequently, the thresholding parameters that satisfies the stopping criteria has been derived in equation-7.9 to 7.11.

$$N = N + 1 \tag{7.9}$$

$$\hat{N} \leftarrow N \tag{7.10}$$

$$\mu \leftarrow \mu_{i_{z=1}}^{\hat{N}} \tag{7.11}$$

The complete frameworks for Max-tree generation and multi-scale thresholding based algorithms have described in algorithm-6 and 7. In summary of large-scale morphological feature generation methods, We have first constructed the min and max tree by using two connected component based parameters .i.e. diagonal and moment of inertia. Next, we have explained the pruning rules in the max tree pruning based on the threshold selection in algorithms-7. Since we have used diagonal and inertia of image components, therefore, Min-pruning and Max pruning have been used for our Maxtree pruning that has described in [193]. The diagonal based pruning has performed using Min rule, described in algorithm-8. The inertia based pruning has performed with Max rule shown in algorithm-9. In algorithm-8 and 9, the output is labelled nodes after pruning, and *node_at_graylevel* represents the number of nodes at some gray level. For though row understanding of Min-rule and Max-rule, we refer the readers to [193]. Finally, we have obtained the various features of the image by converting the pruned sub-trees to the component image, as described in [208]. These pruned features from the original image have been termed as high dimensional features space in which we have the pooled and merged by selection and removal of the redundant features to reduce the feature space for SVM based semantic image classification. Therefore in the next section, we have introduced a novel

Algorithm 7 Threshold selection in Max-Tree

$Inputs \leftarrow Image(I),\ maxtree(T)$
$Thresholdcriteria(F) \leftarrow \{diagonal(d), momentofinertia(i)\}$
$Output \leftarrow thresholds(\mu)$
calculate the Thresholds F(T) for all the nodes of T
$\mu_{i\,i=1}^{n} \leftarrow Sort\ (F(T))$ from n=2 criteria
for i **do**=1 to n do
    $p_{\phi_{\mu_i}}(I) \leftarrow Attr(\phi_{\mu_i}(T))$

end for
while s **do**topping criteria met
    $compute(p_{\phi_{\mu_i}}(I), N)$
    $\hat{p} \leftarrow interpolatep_{\phi_{\mu_i}} onthreshold(\lambda)$

    $err_N \leftarrow 1 - RMS(p, \hat{p})$
    $if\ N > 1$ then
    calculate error and position feature
end while
$N \leftarrow N + 1$
$\hat{N} \leftarrow N$
$\mu \leftarrow \mu_{i\,z=1}^{\hat{N}}$

Algorithm 8 Min-rule to prune the diagonal(d) criteria

$Inputs \leftarrow Predicate(p), maxtree(T), Image(I)$
$Output[all(nodes)] \leftarrow labellednodes()$
$all\ nodes \leftarrow flag(0)$
for a **do**ll nodes=leaf to root do
    if $flag(parent) == 1$ & p(node)==0 **then**
        $node\_at\_graylevel(node) = node\_at\_graylevel(parent)$
        $node \leftarrow flag(1)$
    end if
end for

feature pooling approach to reduce the feature space followed by a multiclass SVM based image parsing methodology. The feature pooling strategy has based on some mathematical relationships, such as correlationρ, structural similarity (SSIM), mutual information (MI), between the features.

Algorithm 9 Max-rule to prune the moment of inertia (i)criteria

$Inputs \leftarrow Predicate(p), maxtree(T), Image(I)$
$Output[all(nodes)] \leftarrow labellednodes()$
$all\ nodes \leftarrow flag(0)$
for a **do**ll nodes=leaf to root do
  if $flag(node) == 0$ **then**
    $parent \leftarrow flag(1)$
  end if
  if $p(node) == 1$ **then**
    $node \leftarrow flag(1)$
    $parent \leftarrow flag(1)$
  end if
end for
for a **do**ll nodes=root to leaf do
  if $flag(node) == 1$ **then**
    $node\_at\_graylevel(node) \leftarrow node\_at\_graylevel(parent)$
  end if
end for

7.2.2 LM-MFP based feature pooling method

In this section, we have pooled and merged the redundant features after the multi-scale morphological procedure that we have applied in the previous section. LM stands for large-scale morphology described previously, and MFP stands for multi-criteria feature pooling, which has discussed in this section. This study is the main novelty of our work, aiming to reduce the feature space so that critical information has retained in the feature set. We have the morphological feature set I with correlationρ, structural similarity (SSIM), and mutual information (MI), which are the predicates for feature pooling. Firstly, we have computed the predicates as:

$$\rho_{x,y}^{(s)} = \frac{\sum_{i=1}^{N}(I_{x,i}^{(s)} - \hat{I}_x^{(s)})(I_{y,i}^{s} - \hat{I}_y^{(s)})}{\sqrt{\sum_{i=1}^{N}(I_{x,i}^{(s)} - \hat{I}_x^{(s)})^2}\sqrt{\sum_{i=1}^{N}(I_{x,i}^{(s)} - \hat{I}_y^{(s)})^2}} \tag{7.12}$$

x and y are adjacent features for ρ calculation. s has denoted the number of comparisons at which ρ has calculated. N is the total number of features. $\hat{I}$ is the average of I^{th} feature. The next index is the structural similarity, which has designed to improve the PSNR and

MSE. SSIM is a combination of luminance, contrast, and structure.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7.13}$$

where μ_x is the average of x
μ_y is the average of y
σ_x^2 is the variance of x
σ_y^2 is the variance of y
σ_{xy} is the covariance of x and y
The next index is mutual information which is also called entropy. The entropy is the common information content between the two features.

$$MI(x,y) = \sum_{x\varepsilon X}\sum_{y\varepsilon Y} P(x,y) log \frac{P(x,y)}{P(x)P(y)} \tag{7.14}$$

where x and y are the random variables as features
P(x,y) is the joint distribution of x and y.
P(x) and P(y) is the marginal distribution of x and y.
The pooling process analyses the features that are very similar to each other and takes a weighted average of similar or redundant features. The similarity between the features has measured by the correlation(ρ), structural similarity(SSIM), and mutual information(MI) based predicate, as mentioned above. We have performed the weighted average features for those features which have satisfied the predicate. In the previous section, we have generated a huge morphological feature space. The feature generation process has performed as equation-7.1 to 7.11, and algorithms-6,7,8,9 followed a robust tree pruning strategy, as shown in figure-7.2(a). For a complete understanding of the tree pruning process, the readers has referred to [196, 193, 208]. The generated features are robust and informative. Therefore, this feature space has used in the multiclass SVM method for semantic parsing of pixel-wise image objects. Generated features are highly informative. Due to the high dimensionality of feature space, the computation complexity is very high. Therefore, in this section, we have applied a pooling based feature integration technique that reduces the feature space without losing the information. The feature pooling process has taken from [72]. Therefore we refer the readers to it. Subsequently, (ρ), structural similarity(SSIM), and mutual information(MI) has been derived from equation-7.12, 7.13, and 7.14 for each pair of features, as described in step-5 of algorithm-10. Let the feature set and the number

of features that we have obtained from the previous step are $\hat{I}$, and N. For all pairs of feature set $\hat{I}$, $\hat{I_k}$ is k_{th} feature number in the N-dimensional feature space. For $\hat{I_k}$ feature, we have computed the ρ, SSIM, and MI values with all other features as:

$$\begin{aligned} &Predicate(p): \\ &\qquad For\ \{k = k+1 : N\}: \\ &\qquad [compute\ \{\rho_{(k,p)},\ SSIM_{(k,p)},\ MI_{(k,p)}\}] \end{aligned} \tag{7.15}$$

In equation-7.15, we have computed the a predicate that reports only those pair of features that has minimal $\rho_{(k,p)}$, $SSIM_{(k,p)}$, and $MI_{(k,p)}$ values between them. In this way the flat features has been merged as in equation-7.16 as:.

$$\begin{aligned} &For\ \{k = k+1 : N\}: \\ &If\ \{\rho_{(k,p)} > \hat{\rho}\ \&\&\ SSIM_{(k,p)} < \hat{SSIM}\ \&\&\ MI_{(k,p)} < \hat{MI}\} \\ &Then\ \hat{I_k} \leftarrow \hat{I_k} \bigcup p \\ &N \leftarrow N-1 \end{aligned} \tag{7.16}$$

Equation-7.16 has denoted that the feature merging of p and $\hat{I_k}$ with $\bigcup$ (merging symbol) resulted in a feature space of reduced dimension to one. It indicates that feature p was flat and non-informative. In this way, the feature space has reduced to minimal level. The $\bigcup$(merging) operation between two feature has defined as in equation-7.17 as:

$$\begin{aligned} &For\ \{k = k+1 : N\}: \\ &\hat{i_k} \leftarrow Element\ wise\ Average_{Third\ Dimension}\{\hat{i_p}; p \varepsilon \hat{I_k}\} \\ &k \leftarrow k+1 \end{aligned} \tag{7.17}$$

In equation-7.17, If predicate .i.e. ($\rho > \hat{\rho} \&\& SSIM > \hat{SSIM} \&\& MI > \hat{MI}$) have satisfied then a weighted average of corresponding features has taken to merge them. All the pooled and remaining features have collected in a new feature space. Finally, the SVM based pixel-wise semantic framework has used to learn a model. The prediction has performed

on the complete pixel-feature set to obtain the final semantically classified image as:

$$\begin{aligned} &Z \leftarrow \{\hat{i_1}, ..., \hat{i_N}\} \\ &Z_1 \leftarrow \{Training\ sample(Z)\} \\ &Train_{model} \leftarrow multiclassSVMTrain(Z_1) \\ &Classified\ Image\ R \leftarrow multiclassSVMTest(Z, Train_{model}) \end{aligned} \tag{7.18}$$

Algorithm-10 has described the LM-MFP approach. In subsequent steps, we have shown the experimental results for our LM-MFP approach on five semantically labelled images, obtained from pascal-voc and sift-flow data-sets. A categorical face prediction method has also performed for the LM-MFP model in experiment-3 of the experimental section.

Algorithm 10 LM-MFP method

$Inputs \leftarrow RGB\ Image\ I, \hat{\rho}, \hat{SSIM}, \hat{MI}$
$Output \leftarrow Classified\ Image\ R$
$i \leftarrow 1$
$k \leftarrow 1$
$N \leftarrow K$
$\hat{I} \leftarrow Morphological\ Features(I)$
for $[k = 1 : N]$ **do**
  $k \leftarrow 1$
  $\hat{I_k} \leftarrow k$
  for $k = k+1 : N$ **do**
    $compute\ \{\rho_{(k,p)},\ SSIM_{(k,p)},\ MI_{(k,p)}\}$
    if $\rho_{(k,p)} > \hat{\rho}$ && $SSIM_{(k,p)} < \hat{SSIM}$ && $MI_{(k,p)} < \hat{MI}$ **then**
      $\hat{I_k} \leftarrow \hat{I_k} \bigcup p$
      $N \leftarrow N - 1$
    end if
  end for
  $\hat{i_k} \leftarrow Element\ wise\ Average_{Third\ Dimension}\{\hat{i_p}; p \varepsilon \hat{I_k}\}$
  $k \leftarrow k+1$
end for
$Z \leftarrow \{\hat{i_1}, ..., \hat{i_N}\}$
$Z_1 \leftarrow \{Training\ sample(Z)\}$
$Train_{model} \leftarrow multiclassSVMTrain(Z_1)$
$Classified\ Image\ R \leftarrow multiclassSVMTest(Z, Train_{model})$

7.3 Experimental Result Analysis

In this section, we have shared the experimental outcomes for our proposed MFP method for semantic target prediction. The evaluation has performed on the pascal-voc and sift-flow datasets with labelled images. The evaluation matrices have computed to compare the results from past proposed methods. We have also shown the versatility of our proposed method by evaluating it on the category classification. In the tables, we have denoted the LM as semantic image prediction using SVM and LM-MFP as semantic images of MFP features using the SVM technique.

7.3.1 Dataset detail

We have conducted the empirical results on sift-flow, pascal-VOC datasets for semantic labelling and YALE, ORL dataset for face category prediction. For semantic segmentation, we have labelled pixels for query objects in the images. In pascal-voc datasets, two main images .i.e. highway and house. Highway and house images contain $256 \times 256 \times 3$ pixels with ten and twelve ground-truth classes. Sheep, horse rider, and horse keeper dataset include 2, 3, and 3 labelled classes, respectively. For category prediction, we have evaluated our scheme on YALE and ORL face publicly available datasets.

7.3.2 Evaluation Metrics

In this section, we have defined the standard matrices for the evaluation of our scheme. These matrices are overall accuracy, kappa score, class-wise accuracy, precision, recall, F score, and intersection over union score(IOU). The given matrices have computed by the confusion matrix that has obtained by applying Support Vector Machine (SVM) based prediction on pooled features. Let TP, TN, FP, and FN denotes the true positives, true negatives, false positives, and false negatives from the confusion matrix. The evaluation of matrices has defined as:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{7.19}$$

$$precision = \frac{TP}{TP + FP} \tag{7.20}$$

$$recall = \frac{TP}{TP + FN} \tag{7.21}$$

$$Fscore = 2 \times \frac{recall \times precision}{recall + precision} \tag{7.22}$$

X_i and Y_i are the predicted and ground truth images having i_{th} pixel at any instance where $\wedge$ and $\vee$ are the bitwise operators in the calculation of inter section over union(IOU).

$$IOU = \frac{\sum_i (X_i \wedge Y_i)}{X_i \vee Y_i} \tag{7.23}$$

These matrices have determined the performance of the proposed MFP method over the past techniques in a better sense. We have found that these indices have appropriately analysed the prediction scenario.

7.3.3 Experiment 1: Semantic prediction on Sift-Flow dataset

In this section, the highway and house images from sift-flow dataset have used for the assessment of the MFP scheme. The sift-flow dataset consists of 7000 images in 33 semantic classes whereas, our sample images i.e., highway and house incorporate twelve and ten classes respectively.

7.3.3.1 highway dataset

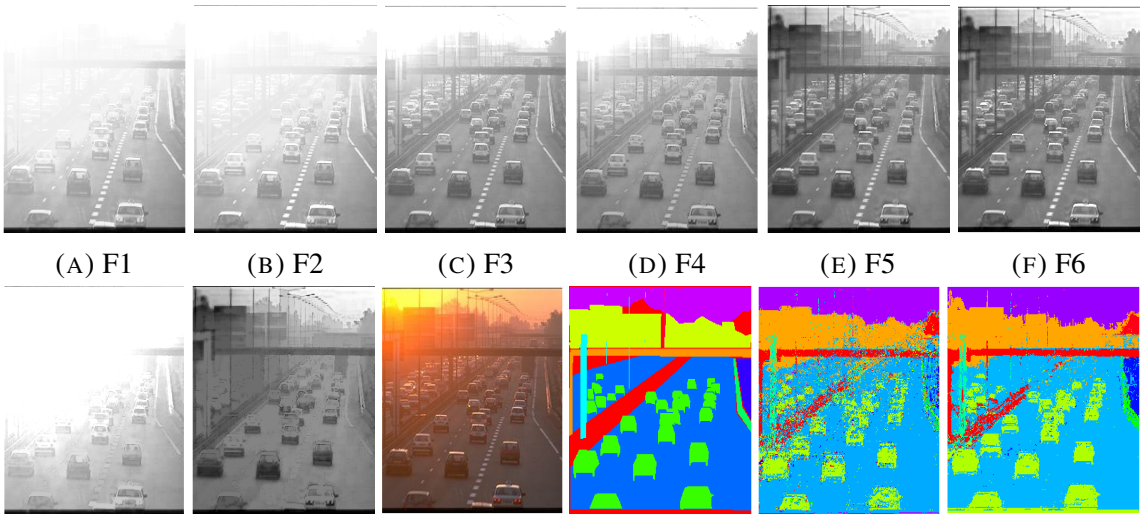
In the highway dataset, we have first extracted the morphological features by using attribute filtering. For attribute filtering, we have used area and diagonal attributes for opening and closing operations. The area attribute is 10 and 15, whereas diagonal attributes are 50, 100, and 500. Firstly, the area and diagonals have computed, and a max tree has generated. A robust tree pruning method has applied to obtain informative morphological features based on the above area and diagonal attributes. These features have made using core and morpho package of Olena-1.0 library, which has written in C++. The area function is present in the opening and closing package. The diagonal function is present in the attribute package. We have created 33 exclusive features. Features have concatenated in the

TABLE 7.1: Testing results for sift-flow dataset

Classes	sift-flow(highway)		sift-flow(house)	
	LM	LM-MFP	LM	LM-MFP
c1	75.95	90.23	91.52	91.24
c2	89.49	92.22	69.52	89.04
c3	66.32	83.06	45.80	52.10
c4	29.33	68.57	71.20	81.25
c5	54.64	58.44	92.23	95.72
c6	91.15	96.06	65.67	90.48
c7	23.29	84.78	99.16	99.86
c8	98.55	95.17	66.47	12.94
c9	47.83	29.35	92.51	10.38
c10	80.62	77.32	69.86	67.58
c11	-	-	55.69	79.20
c12	-	-	61.21	63.27
OA	85.02	90.96	88.44	89.03
Kappa	0.7844	0.8830	0.8563	0.8480
IOU	0.8822	0.8822	0.9783	0.9783
Precision	0.5727	.6988	0.7253	0.8015
Recall	.5974	.7047	0.6776	0.6331
F score	0.5846	0.7017	0.7006	0.7074
Time(EFF)	1950.54	668.35	2018.69	1960.05

third dimension of the image. The features have also increased the dimensionality of image space. The C++ code for morphological attribute filter, tree pruning has well-explained and taken from [194]. Dalla Mura et al. has introduced the attribute features by using several thinning, thickening operations in C++ and converted them into a MEX file. This code has written by using the Olena-1.0 image processing library. Therefore, we refer the readers to [194], for a complete understanding of morphological feature extraction. We have obtained a high dimensional feature space with 33 dimensions, which require a low dimensional representation of features to reduce the computation cost. Therefore, we have applied our feature pooling based MFP method that has pooled/merged the features efficiently. The feature merging has performed based on correlation, structural similarity, and mutual information. Our condition for feature merging is:

$$P = (\rho\ [F_i, F_j] < 0.8) \& (SSIM\ [F_i, F_j] < 0.8) \& (MI\ [F_i, F_j] < 2.8) \tag{7.24}$$



(A) F1 (B) F2 (C) F3 (D) F4 (E) F5 (F) F6

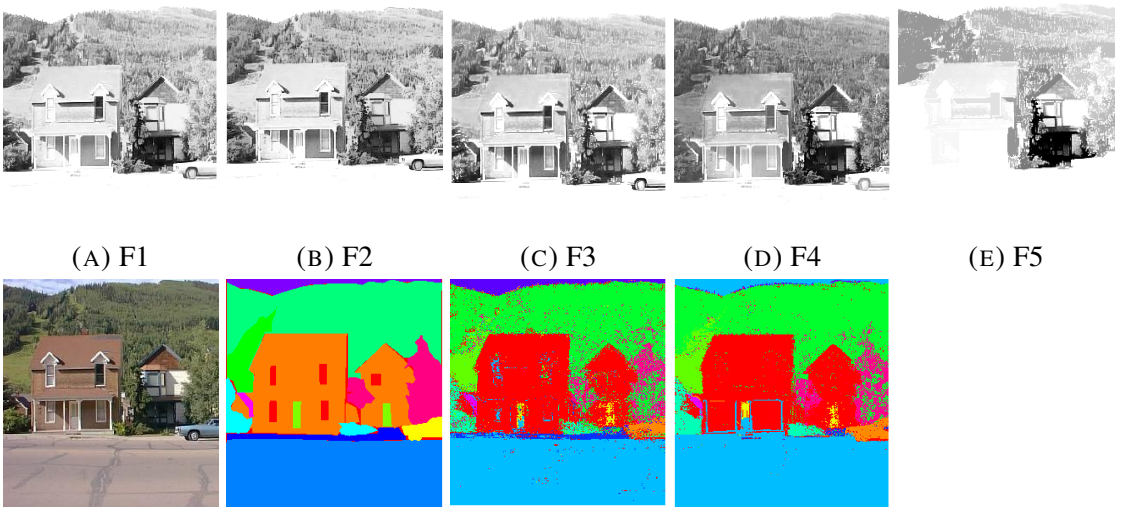
(G) F7 (H) F8 (I) highway (J) groundtruth (K) LM-85.02 (L) MFP-90.96

FIGURE 7.3: Pooled features and classification results for highway data(siftflow)

When above predicate(P) has satisfied then we have merged the features F as:

$$F_{new} = \frac{1}{N} \times SUM(F, 3, 'omitnan') \tag{7.25}$$

Where N is the total number of feature obtained after satisfying the predicate (P), SUM is a cumulative average of elements in the third dimension. The above feature pooling assumption has reduced the feature space from 33 to 8 for the highway dataset. The figure-7.3(a) to 7.3(h) has denoted the pooled feature obtained by using our MFP approach. These pooled features having ten dimensions have used in SVM based pixel-wise semantic classification process for evaluation of our scheme. Figure-7.3(i), 7.3(j) are the original and ground truth images for pixel-wise classification. We have used multi-class SVM trained on 70 % samples and test upon complete feature set. Figure-3(k) is the SVM result after pixel-wise classification of the original image denoted as the LM scheme. Figure-7.3(l) indicates the pixel classification results obtained from pooled features, meant as the LM-MFP method. In table-7.1, it has observed that the proposed LM-MFP is achieving 5-6 % higher accuracy than the LM method in the highway dataset. The house dataset has ten classes, whereas, in each category, our LM-MFP scheme is performing better than the LM method. Kappa, precision, recall, F-score are 10, 12, 11, 12 % higher fo the LM-MFP process. Therefore, for the highway dataset, the proposed LM-MFP method has performed state-of-art.



(A) F1 (B) F2 (C) F3 (D) F4 (E) F5

(F) highway (G) groundtruth (H) LM-88.44 (I) MFP-89.03

FIGURE 7.4: Pooled features and classification results for highway data(siftflow)

7.3.3.2 house dataset

In this section, we have validated the proposed LM-MFP method for house image of the sift-flow dataset. Firstly, morphological features have extracted by using attribute filtering and tree pruning of area and diagonal attribute, as discussed in the previous section. Thresholds for the area and diagonal based tree pruning is the same as in the previous section. Like the last section, the number of features is 33. The ρ, SSIM, and MI-based MFP based feature pruning strategy has applied to reduce the feature space to 6, as discussed in the highway dataset. Further, SVM based pixel-wise semantic target prediction has applied in which 70% pixels have used for training. Table-7.1 shows the accuracy, precision, recall, f-score, and computation time obtained for the house dataset. It has analysed that overall efficiency has increased 1-2% in LM-MFP over the LM method. In class-wise accuracy, we have reported a significant increase in accuracy for every twelve classes in the LM-MFP technique except class- 8, 9, and 10. Precision, recall, f-score has achieved a significant improvement for the LM-MFP method. We have also noticed that the computation time for the LM-MFP process is lower than the LM method with a substantial margin of about 958 seconds. Therefore, the proposed method has performed better than the LM method for house image of the sift-flow dataset. The pooled feature has shown in figure-7.4(a) to 7.4(e). Ground truth and the original image have shown in figure 7.4(g) and 7.4(f). The accuracy of the LM and LM-MFP method has shown in figure-7.4(h) and 7.4(i).

TABLE 7.2: performance of LM and LM-MFP model on pascal-voc dataset

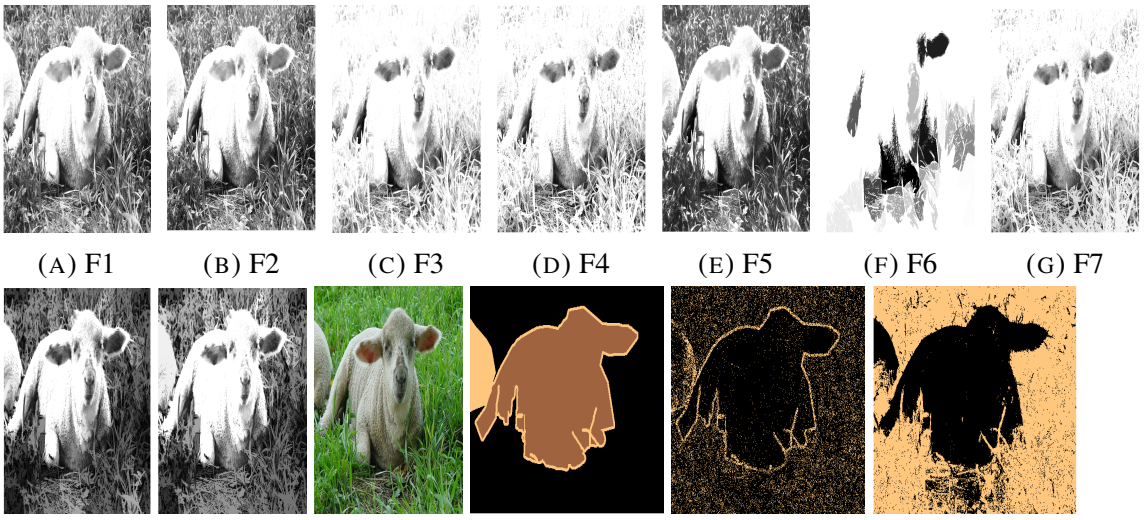
Classes	pascal-voc(sheep)		pascal-voc(horse-rider)		pascal-voc(horse-keeper)	
	LM	LM-MFP	LM	LM-MFP	LM	LM-MFP
c1	97.91	98.79	77.35	91.87	95.73	97.94
c2	49.84	56.88	94.71	97.35	88.69	96.40
c3	-	-	75.34	65.75	50.76	51.12
OA	88.47	90.57	89.06	91.12	89.44	93.63
Kappa	0.5661	0.6496	0.7596	0.7999	0.7978	0.8711
IOU	0.3883	0.3883	0.2336	0.2336	0.3833	0.3833
Precision	0.2091	0.2737	0.2532	0.2795	0.2416	0.3757
Recall	0.4925	0.5189	0.6185	0.6374	0.5880	0.6137
F score	0.2936	0.3584	0.3593	0.3886	0.3425	0.4661
Time(EFF)	740.87	265.46	0.1494.65	530.66	4500.88	2316.97

7.3.4 Experiment 2: Semantic prediction on Pascal-Voc dataset

In this section, the LM-MFP scheme has applied on pascal-voc dataset to predict the semantically labelled pixels. The pascal-voc dataset contains 11,530 images, 6929 clusters in 20 labelled classes. We have proposed the LM-MFP approach for pixel-wise training of SVM. Therefore, we have adapted the sheep, horse rider, and horse keeper images, which contains the exclusive classes such as sheep, horse, and man. Unlike the sift flow images, the pascal-voc images are uneven in size and provide fewer labelled categories. Sheep, horse rider, and horse keeper images exhibit 2, 3, and 3 mutually exclusive labels in the ground truth.

7.3.4.1 Sheep dataset

The sheep image exhibits two classes in labelled ground-truth. The morphological feature process and MFP based feature pooling are the same as in the experiment-1. The pooling process leads to generate the 9-dimensional feature space that has used for SVM based semantic prediction. In the sheep dataset, we have used the same predicate condition, i.e., correlation, structural similarity, and mutual information, for feature merging. 70% pixels of labelled samples have used for training the SVM based pixel-wise learner, and complete data have used for testing. In table-7.2, for the sheep dataset, the accuracy of the LM-MFP method is 2-3% higher than the LM method. On average, the efficiency has increased for each class. Kappa, precision, recall, and f-score have a significant increase of 0.8, 0.07, and 0.2, and 0.6 respectively for the LM-MFP method over the LM scheme. Figure-7.5(a) to 7.5(i), 7.5(j), and 7.5(k), and 7.5(l) and 7.5(m) have denoted the pooled



(A) F1 (B) F2 (C) F3 (D) F4 (E) F5 (F) F6 (G) F7

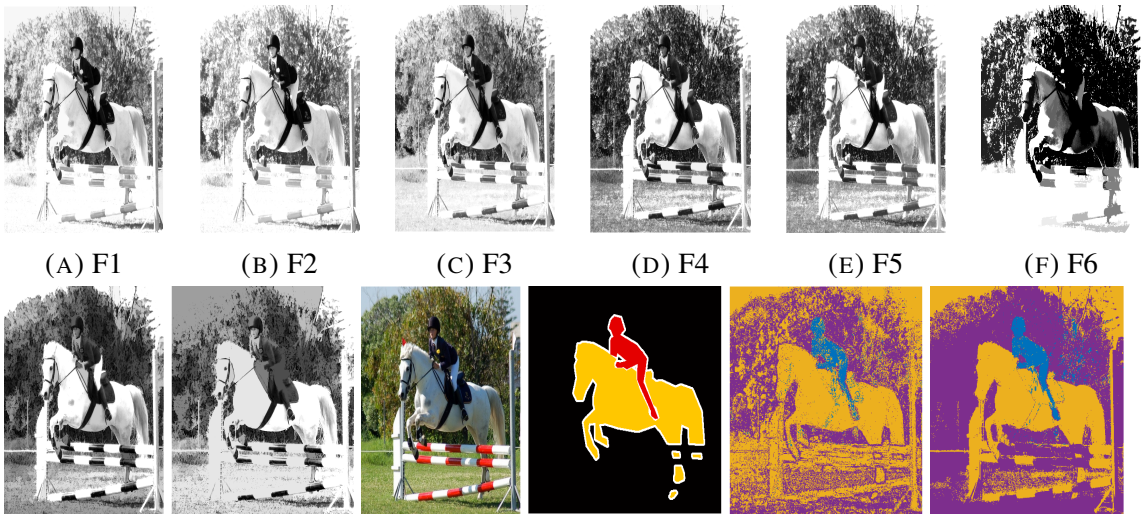
(H) F8 (I) F9 (J) sheep (K) groundtruth (L) LM-88.47 (M) MFP-90.57

FIGURE 7.5: Pooled features and classification results for sheep data(pascal-voc)

features, original & ground-truth images, and semantic predictions of pooled features for the labeled samples. The increase in accuracy and other matrices have acknowledged the proposed LM-MFP scheme as a state-of-art for sheep dataset. The computation time for the LM method is 740.87 seconds, whereas the proposed LM-MFP method has taken 265.46 seconds, which is significantly lower computation time for the sheep dataset.

7.3.4.2 Horse Rider dataset

Subsequently, the LM-MFP method has evaluated on the horse rider image of pascal-voc. The morphological feature space has generated by using the experiment-1. The redundant features have discarded using correlation, structural similarity, and mutual information-based predicate condition. The number of reduced features are eight. The SVM based semantic pixel classification results with 70% training samples have shown in table-7.2. The overall accuracy has increased from 2-3% for horse rider image. The class-wise accuracy of the LM-MFP method for each class except c3 is greater than the LM method. Kappa, precision, recall, f-score are 4, 2, 2, 3% higher for the LM-MFP method than the LM method. The computation time for LM-MFP is 530.66 and 1494.65 for LM. Therefore, our prosed LM-MFP method has performed better in terms of accuracy and computation time for horse rider image than the LM method. Figure-7.6(a) to 7.6(h) has denoted the pooled features. Figure-7.6(i), 7.6(j) are the original and ground truth



(A) F1 (B) F2 (C) F3 (D) F4 (E) F5 (F) F6

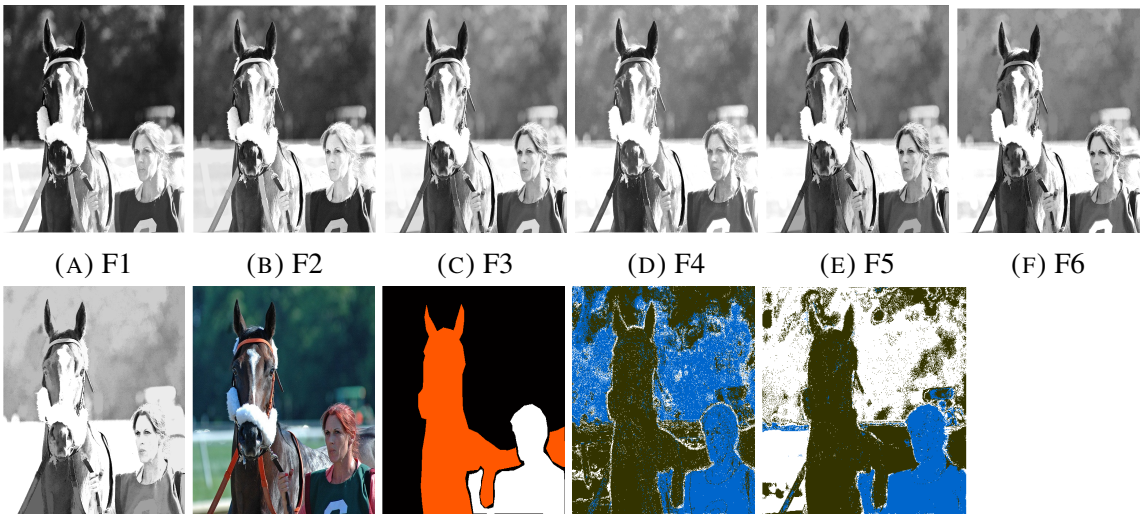
(G) F7 (H) F8 (I) horserider (J) groundtruth (K) LM-89.06 (L) MFP-91.12

FIGURE 7.6: Pooled features and classification results (in %) for horse-rider data(pascal-voc)

images. Pixel-wise semantic labelling on LM and LM-MFP methods have shown in figure-7.6(k), and (l).

7.3.4.3 Horse Keeper dataset

The horse keeper dataset has two mutually exclusive classes as a labelled sample. Morphological feature space has created using attribute filters, as discussed in experiment-1. The previously discussed feature pooling scheme(MFP) has applied to reduce the feature space up to seven for lesser computation cost. We have seven exclusive features in the reduced feature space. SVM based semantic prediction with 70% training pixels has performed on the complete horse keeper image. SVM based semantic prediction scheme has achieved 4-5% greater accuracy for the proposed LM-MFP method than the LM method. In class-wise observation, it has found that all three classes in LM-MFP have got better accuracy than the LM method. Kappa, precision, recall, f-score are 4, 2, 2, and 3% higher for LM-MFP than the LM method. The computation time for the LM-MFP scheme is 2316.97 seconds, which is significantly lesser than the LM method that has computation time = 4500.88 seconds. Figure-7.7(a) to 7.7(g) have shown the pooled feature by using the MFP method. Figure-7.7(h) and 7.7(i) are the horse keeper images and labelled ground truth for semantic labelling. Figure-7.7(j) and 7.7(k) have shown the SVM classification



(A) F1 (B) F2 (C) F3 (D) F4 (E) F5 (F) F6

(G) F6 (H) horsekeeper (I) groundtruth (J) LM-89.44 (K) MFP-93.63

FIGURE 7.7: Pooled features and classification results for horse-keeper data(pascal-voc)

for semantic pixel prediction for LM and LM-MFP methods. The prediction results have demonstrated that the LM-MFP process is performing state-of-art for horse keeper image.

7.3.5 Experiment 3: Category prediction on YALE and ORL dataset

In this section, we have performed the experiments for the category classification of face datasets to evaluate the proposed LM-MFP scheme. In category prediction, we have categorised two publicly available datasets, i.e., YALE and ORL. The ORL dataset consists of 400 images of the face having a size of 32×32. The ground truth categories comprised of 40 classes. Whereas, YALE dataset contains 165 images of the face having a size 32×32. The YALE data set consists of images in 15 different classes. After applying morphological feature extraction, we have a vast number of features on given ρ, SSIM, and MI. These parameters are similar to semantic prediction in experiment-1, and 2. Then, we have applied our proposed MFP based feature pooling scheme. There are more than 50 features in each case which are satisfying the predicate after pooling. Subsequently, we have trained the SVM on 70% images of ORL and YALE dataset. We have tested our trained model on pooled features. In figure-7.8(a) to 7.8(j), the category prediction results for the first ten samples have shown for the YALE dataset. Figure-7.8(k) to 7.8(t) has shown the category prediction result for the ORL dataset for the first ten sample images.

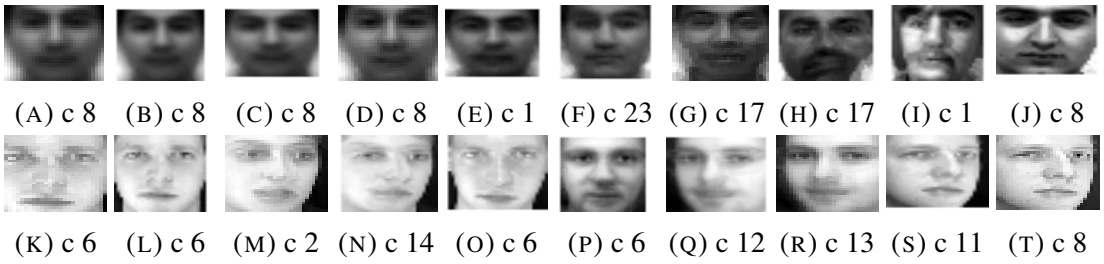


FIGURE 7.8: Catagory prediction results for YALE(8(a) to 8(j)) and ORL(8(k) to 8(t)) datasets

7.3.6 Comparative Analysis

This section has compared the efficiency of the proposed LM and LM-MFP method with previously proposed methods. Mainly we have carried out the comparison with a state-of-art way [200]. The referred work has suggested a deep learning-based ensemble method that integrates the logits for student models and a fully connected MLP method for the sift-low and pascal-voc dataset. Table-7.3 represents the accuracy of the referred process for labelled classes. It has observed that car object has predicted with 69.60%, 65.5%, and 71.20% accuracy for single FCN-8, compressed FCN, and ensemble scheme. In our prediction, the car class is c3 in table-1, which has achieved 83.06% accuracy for LM-MFP, i.e., proposed method. Therefore, it has noted that the LM-MFP process is achieving higher accuracies than past algorithms. The house dataset(c4 in table-7.1) has obtained the 81.25% accuracy for LM-MFP, which is comparatively similar to FCN-8 and compressed FCN. For sheep image, LM and LM-MFP are achieving 88.47% and 90.57% accuracy in comparison with past methods, i.e., FCN-8, compressed FCN-8, and ensemble FCN which has 70.70%, 71.00%, and 77.40% accuracy respectively. Hence our proposed LM-MFP is achieving 20%, 19%, and 13% greater accuracy than past method. In the same table-7.3, for horse rider image, the LM-MFP has achieved 91.87% and 97.35% accuracy for man and horse class which is 28%, 27%, and 21% higher than the horse class and 17%, 17%, and 15% greater than man class of FCN, compressed FCN, and ensemble FCN methods respectively. The proposed LM-MFP method is reporting 97.94% and 96.40% accuracy for horse and man class which is 34%, 33%, and 27% higher for horse and 16%, 16%, and 14% higher for man class from FCN-8, compressed FCN, and ensemble FCN methods. Figure-7.9(a),(b), (c), and (d) is the ground truth and classification results of FCN-8, compressed FCN, and ensemble FCN methods for horse keeper image. Figure-7.9(e),(f), (g), and (h) is the ground truth and classified outcome of sheep image. Figure-7.9(i), (j), (k), (l) shows the

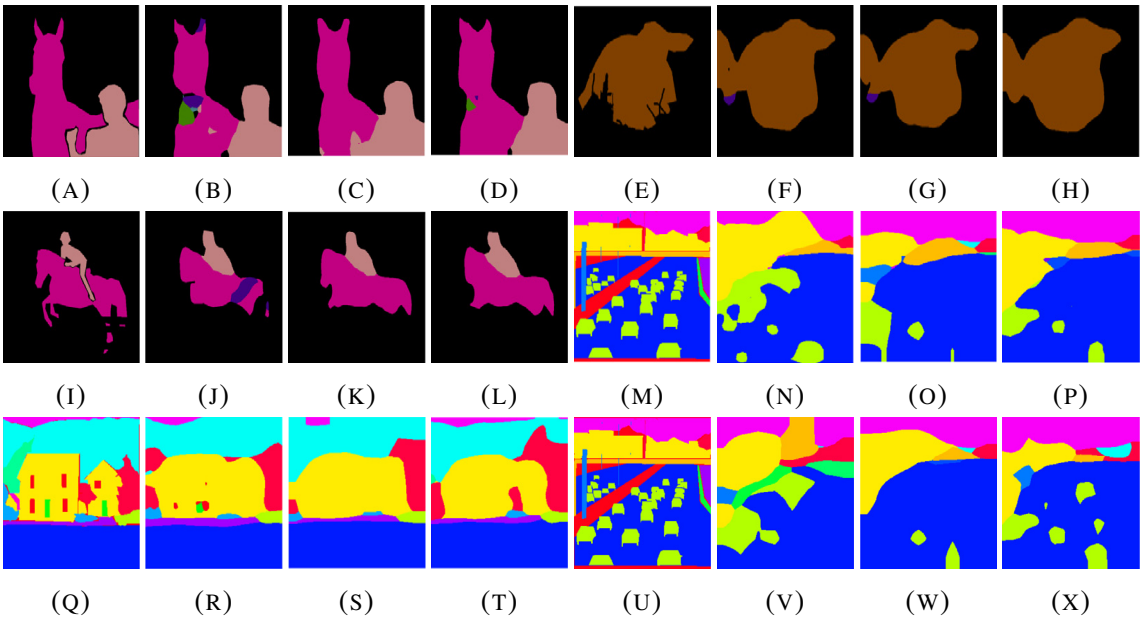


FIGURE 7.9: Results of Comparative approaches in sift-flow, pascal-voc datasets

TABLE 7.3: Comparative Analysis

	car	house	sheep	horse	person
Single-FCN-8	69.6	83.7	70.7	63.6	80.3
Compressed FCN	65.5	83.3	71.0	64.8	80.9
Ensemble	71.2	84.8	77.4	70.8	82.3

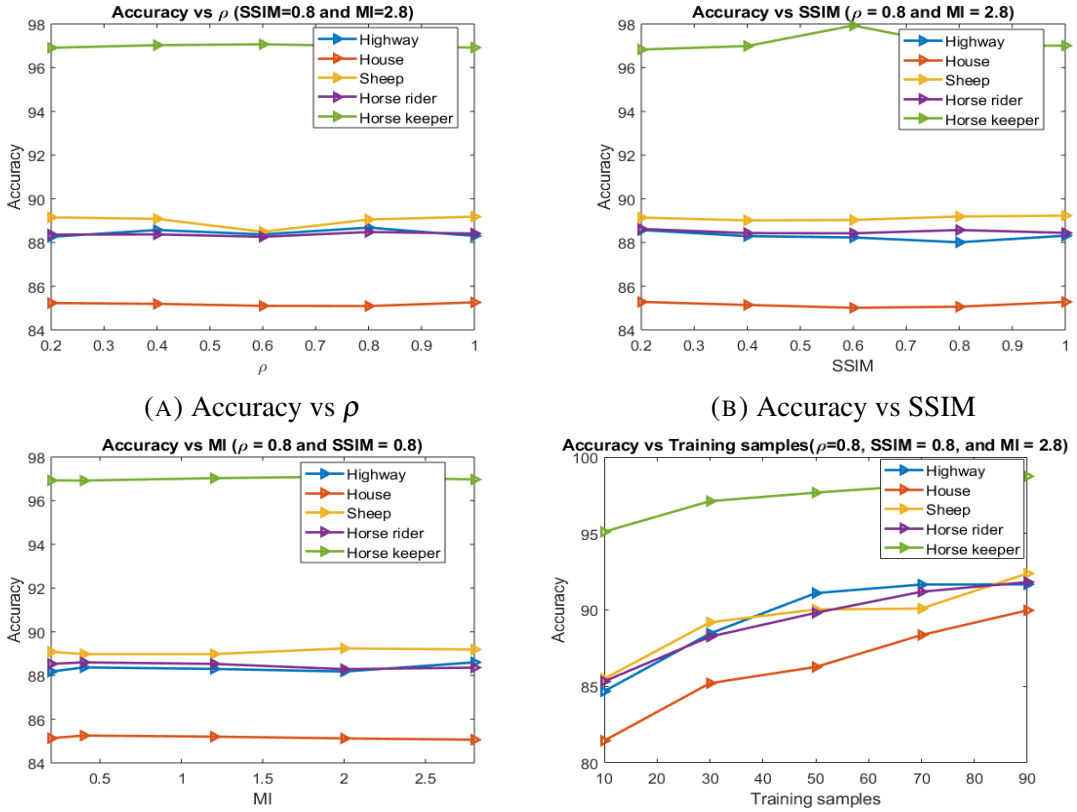
ground truth and predictions for horse rider image. figure-7.8(j), (k), (l) are the outcomes for house dataset, and figure-7.9(m), (n), (o), (p) are the classified outcome of highway data for FCN-ResNet-101, compressed FCN-ResNet-50, and compressed FCN-ResNet-50 skip, that has introduced in work referred in this section. The performance matrices and output results have denoted that proposed LM-MFP has outrun the previously proposed methods for each class label in the selected images of sift-flow and pascal-voc datasets. Later sections have described the impact of LM and LM-MFP parameters on the accuracy and time in a real-time scenario.

7.3.7 Effect of correlation, SSIM, Mutual information, and Training samples on Overall accuracy (OA)

In this section, various parameters of our proposed LM-MFP method have compared with accuracy in the process. In figure-7.10(a), we have kept structural similarity(SSIM), and mutual information(MI) fixed to 0.8 and 2.8 and plotted accuracy vs. correlation(ρ) graph for sift-flow and pascal-voc images. The accuracy is highest for horse keeper image while increasing the correlation. In figure-7.10(b) ρ, MI is fixed as 0.8,2.8, and accuracy has plotted against SSIM. The horse keeper and house dataset again achieving the highest and lowest accuracies. In figure-7.10(c), ρ, SSIM is fixed as 0.8, 0.8, and accuracy has plotted against MI. In figure-7.10(d), we have shown the impact of training samples on accuracy by keeping ρ, SSIM, MI fixed as 0.8, 0.8, 2.8. It has observed that accuracy is increasing with an increase in training samples for each dataset. Accuracy is minimum and maximum for house and horse keeper datasets, respectively.

7.3.8 Complexity, Computation Time

The computation time has been observed for the variation in different parameters of the LM-MFP method. The figure-7.11(a) denoted the time vs. ρ graph when other parameters have fixed as in the previous section. The time for house and horse keeper dataset is maximum and minimum, respectively. Time is not much dependent upon ρ. The time vs. SSIM plot has shown in figure-7.11(b). House, horse keeper datasets again take the maximum and minimum time, respectively. Figure-7.11(c) has shown the Time vs. MI graph by taking other predicate fixed. In figure-7.11(d), we can observe that by increasing training samples by a step length of 10%, the computation time is increasing for each dataset. For each case house and horse keeper dataset is taking maximum and minimum time respectively.



(A) Accuracy vs ρ

(B) Accuracy vs SSIM

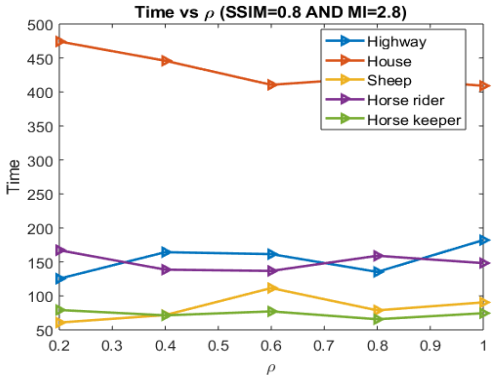
(C) Accuracy vs MI(in %)

(D) Accuracy vs Training samples

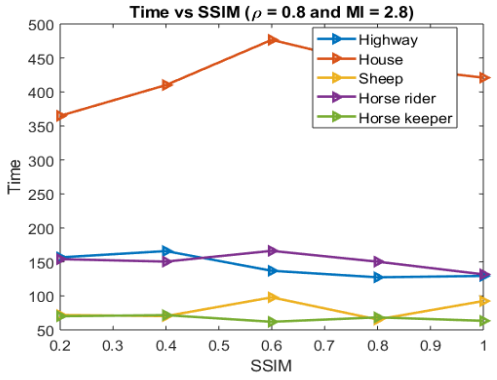
FIGURE 7.10: Classification Accuracy vs LM-MFP parameters for images of sift-flow and pascal-voc datasets

7.3.9 Accuracy, Precision, Recall, Kappa, F-score, computation time for LM and LM-MFP

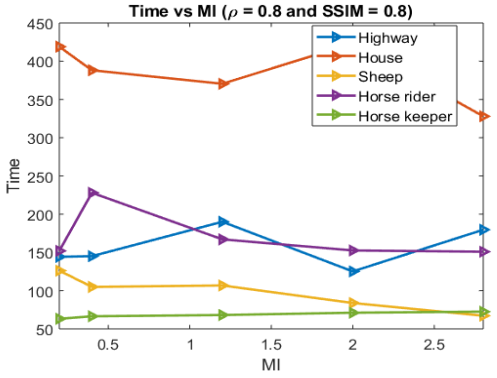
In this section, we have depicted the accuracy, kappa, precision, recall, f-score, and time of LM and LM-MFP for each dataset in a robust setting. We have kept ρ, SSIM, MI fixed as 0.8,0.8, and 2.8 for each bar graph. Figure-7.12(a) is showing the accuracy of LM and LM-MFP method for house, highway, horse keeper, horse rider, and sheep dataset. It has observed that the accuracy for each dataset is higher in case of LM-MFP scheme than the LM method. The kappa value for each image is higher for LM-MFP, as shown in figure-7.12(b). The precision, recall, and f-score is also higher for LM-MFP over LM method for each dataset, as shown in figure-7.12(c), 7.12(d), and 7.12(e). In figure-7.12(f), we have found that computation time for LM method is significantly higher than LM-MFP



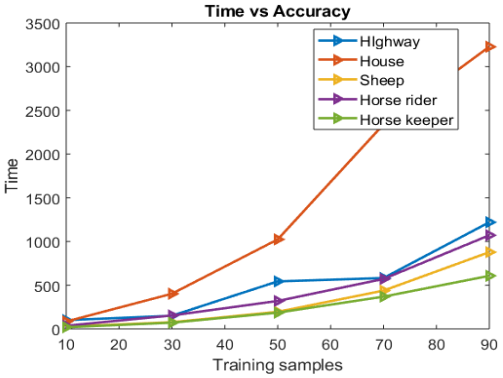
(A) computation time vs ρ



(B) computation time vs SSIM



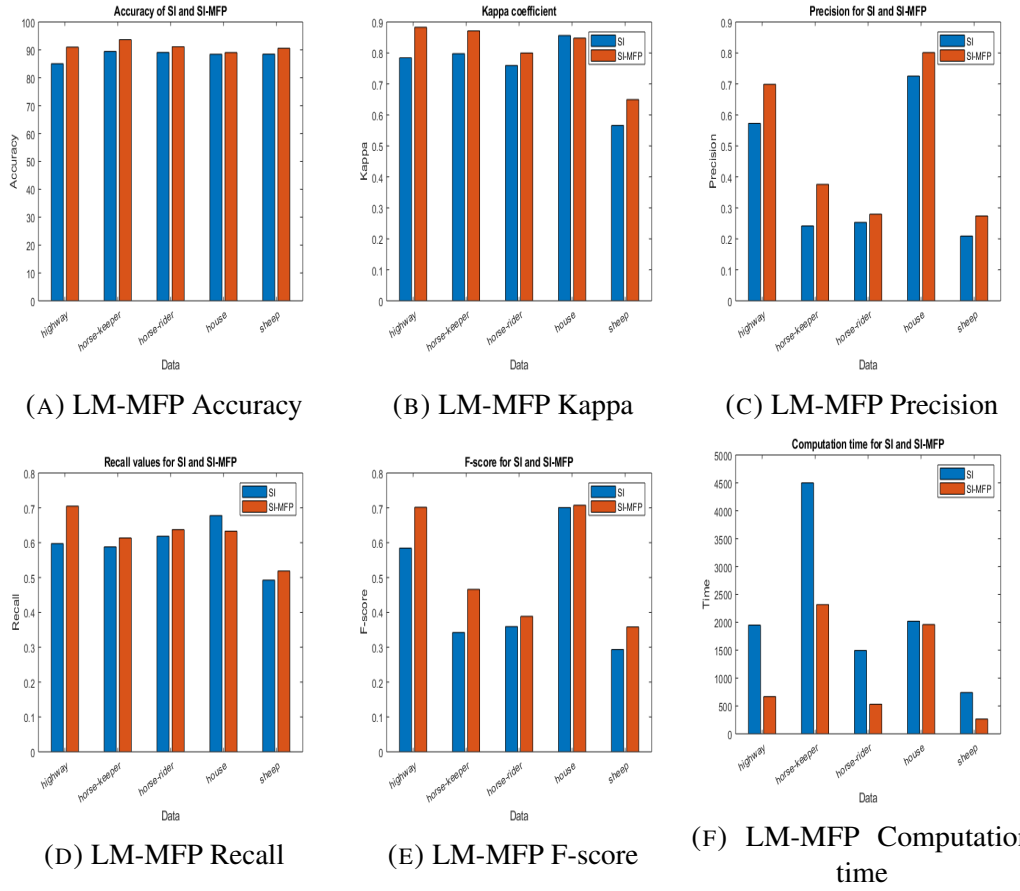
(C) computation time vs MI)



(D) computation time vs training samples

FIGURE 7.11: Computation Time vs LM-MFP parameters for images of sift-flow and pascal-voc datasets

method for each dataset. The improvement in efficiency matrices and less computation time has proven our LM-MFP method as state-of-art over LM and past methods.



(A) LM-MFP Accuracy (B) LM-MFP Kappa (C) LM-MFP Precision

(D) LM-MFP Recall (E) LM-MFP F-score (F) LM-MFP Computation time

FIGURE 7.12: Accuracies, Kappa, Precision, Recall, F-score, and Time

7.4 Conclusions and Future Work

We have proposed a two-stage framework LM-MFP for feature extraction and pooling in the semantic classification of images. In the first stage, we have extracted the attribute filtering based morphological features by using a series of thinning and thickening operations. This process has led to a very high dimensional feature space. Subsequently, we have applied a robust feature pooling method that has reduced the feature space by using correlation, structural similarity, and mutual information(entropy). The pooled features have used for pixel-wise semantic classification for labelled target prediction. We have used the pooled features for both semantic and categorical target prediction. Finally, by the comparative study and parameter analysis, we have found that proposed method is performing state-of-art for both pixel-wise and categorical target prediction. In the future,

we will be working on some mimetic and optimization-based methods for more robust features and noise removal methods that enhance the efficiency of a classifier.