6

Annotation Localization using a Weakly Supervised Model for Top-down Visual Saliency

This chapter describes the concept of Discriminant Saliency, its principles and its suitability for handwritten annotation extraction in printed documents. Section 6.1 describes Discriminant Saliency and its computational models. Section 6.2 gives an overview of the dataset created for the problem. Section 6.3 describes the features used in our work. In Section 6.4, further investigates the threshold protocols adopted for the given problem. The results are illustrated in the subsequent Section 6.5 on our dataset and two other standard datasets, namely, IAM and PRIma-NHM. We also present a comparison of our work with a discriminative classifier in the Section 6.6. Finally, Section 6.7 concludes the chapter.

6.1 INTRODUCTION

We are quickly able to interpret a scene, even if it is highly cluttered. This happens not because our perceptual system is powerful enough to quickly process the visual sensory input from the entire scene, but because there are saliency mechanisms deployed by the neural circuitry which can detect *salient* regions, which are the regions carrying higher semantic content that is most relevant to scene understanding. These salient regions constitute the relevant subsets of sensory information and are prioritized for further analysis by the visual cortex. The detection of salient regions helps in quickly recognizing the important objects and subsequently interpreting the scene.

Research related to attentional/saliency mechanisms has been carried out in psychology and neurophysiology. Experiments in psychophysics and neural signal recording have contributed to the understanding of saliency, however, such knowledge cannot be readily translated into computational models and principles for optimal saliency computation by computer vision algorithms. For tasks such as object detection, recognition and tracking, many computer vision algorithms rely on extraction of interest points, which can be considered as salient points. The purpose of these interest point detectors is to reduce the computational burden on the subsequent processing stages, which can focus on detailed processing of information around the interest points. Such interest point detectors can be tied to visual features of two types:

- Features having better stability against transformation and having mathematically well-defined properties. For example, such as edges, corners, contours, local symmetry, blobs, etc.
- Features that depend on generic principles pertaining to image complexity. For example, variance of Gabor filter responses over multiple orientations, entropy of the distribution of local intensities, values of wavelet decomposition coefficients, etc.

This translates the question of what constitutes optimally salient (or the optimality criteria for saliency), into optimal detection of specific visual attributes depicting stability or computing feature values depicting image complexities. Though these salient point detectors give good saliency judgements and are useful in many applications, these saliency judgements are not

influenced by the recognition goal. Therefore, the extracted interest points may not help the object detection/recognition modules. In other words, the extracted interest points are not optimal from the point of view of recognition. In fact, it may happen that there are no interest points on the object to be detected and therefore such interest points may not offer any advantage over dense sampling or random sampling. Such saliency computation which is stimulus driven, but not aligned with recognition goal is called as *bottom-up* saliency.

Attentional mechanisms which are driven by recognition goals contribute to the *top-down* component of saliency. Top-down mechanisms are weak classifiers that extract regions of the scene that are likely to contain the object to be recognized. Computational models for top-down saliency need to be efficient so that the candidate regions likely to contain the object of interest can be quickly extracted. Such salient regions deserve attention or further processing by the brain to establish presence of the desired object.

Optimal saliency, i.e. what (properties) constitute a salient region, now gets tied to the recognition problem. [Gao *et al.*, 2009] proposed a model for recognition driven top-down saliency called as *discriminant saliency*. The model allows the design of computationally fast weak classifiers which can be trained in a weakly supervised setting. The principle of discriminant saliency specifies two fundamental tasks:

- 1. Feature selection: This task involves selecting features that best distinguish the object class to be recognized from other possible objects in the scene. This definition for the feature selection task translates into a computational principle of classification with minimal expected probability of error.
- 2. Saliency detection: This task involves assigning saliency values to the the extracted feature components and taking a call on which regions on the image are salient.

6.1.1 Computational principles for top-down saliency

The computational principle for saliency is closely related to some of the previously proposed principles for the organization of perceptual systems. These principles can be translated into a computational models for saliency, as follows:

- A) Maximization of information transmission across perceptual layers (infomax): This selects optimal features as the ones which are maximally informative of the presence/absence of the target class in the field of view.
- B) Inference by detection of suspicious coincidences: This selects optimal features as those whose observation is most suspicious in the absence of the target class. This principle was contributed by Barlow [Barlow, 1994; Bartlett *et al.*, 2002].
- C) Classification with minimal uncertainty: This principle selects features which minimize the uncertainty about the presence/absence of the target class.
- D) Discriminant Saliency: This principle selects features which give the minimum probability of error. Thus, it specifies a discriminant principle for the design of top-down saliency computation.

[Gao *et al.*, 2009] asserted that the methods inspired by the first 3 principles *can* give saliency measures that are nearly optimal w.r.t. the computational principle of discriminant saliency, i.e. in the minimum probability of error sense. [Gao *et al.*, 2009] investigated which of these methods allow a computationally more efficient method and found that the Barlow's principle of inference by detection of suspicious coincidences gives the most efficient method under reasonable simplifications. By using this method, they developed efficient algorithms for feature selection and saliency detection.

The computed saliency value for the interest points helped identify the high saliency value locations that deserve the focus of attention and the remaining low saliency value locations which can be pruned out for further processing. Discriminant saliency refers to a decision-theoretic interpretation of perception. It hypothesizes that perception involves taking a decision regarding which sensory input from the surrounding environment is salient. Taking optimal decisions would correspond to having minimum probability of error. Discriminant saliency refers to making a decision of classifying the stimuli into two classes: target and null hypothesis. Saliency refers to the confidence with which a location in the scene can be classified as containing the target. The decision making hypothesis (discriminant saliency) can be applied to top-down as well as bottom-up forms of attentional mechanisms. For bottom-up saliency detection, the `decision-making' can be made part of the center-surround image processing. For top-down saliency detection, the decision-making can be adapted to any specification of target stimuli and null hypothesis. This translates to learning a one-vs-all classification model, where the object class of interest constitutes the target stimuli and all the other object classes constitute the stimuli considered as the null hypothesis.

Vision algorithms translate the visual stimuli into features. Salient features are the features which can well discriminate the target class from the other object classes (null hypothesis). Appropriate image attributes are selected depending on the target to be recognized. The saliency measure corresponds to the confidence of classifying a portion of the scene into the target class. Salient locations are the ones which compute the highest confidence in classifying the target. It is worth mentioning that the salient features identified in the previous step will vary in their confidence while declaring a given location as salient (possibly containing the target). This variation happens because of the varying recognition context. Features that are effective in classifying the target in a given context (background) may not remain effective as the recognition context changes. Instead, another set of features may be able to better discriminate the target from its changed background.

6.1.2 Optimal feature selection for discriminant saliency

We now review the computational principles that can be used to derive optimality criteria for feature selection. Given an observation **x** that lies in the feature space \mathscr{X} , the following problems need to be addressed.

Task 1: How **x** can be classified as salient or non-salient? Task 2: What is the confidence value associated with the classification of **x** ? Task 3: How to choose the optimal feature space \mathscr{X} ?

We now discuss the three computational principles, (i) Bayesian Decision Theory, (ii) Principle of minimum uncertainty, and (iii) Barlow's principle of suspicious coincidences, for top-down saliency that guide and specify the mathematical expressions for the 3 tasks.

- 1. Bayesian Decision Theory: The Bayes classifier models $P_{Y|\mathbf{X}}(i|\mathbf{x})$ where $i \in \{0,1\}$ denotes absence of target (i = 0), or presence of target (i = 1).
 - Task 1: The decision regarding whether a feature **x** belongs to the class i = 0 or i = 1 is based on the outcome $g^*(\mathbf{x})$, which is

$$g^*(\mathbf{x}) = \arg\max P_{Y|\mathbf{X}}(i|\mathbf{x}) \tag{6.1}$$

– Task 2: The maximum value of $P_{Y|\mathbf{X}}$ for a class is taken as the confidence measure for classification

$$c^*(\mathbf{x}) = \max_i P_{Y|\mathbf{X}}(i|\mathbf{x}) \tag{6.2}$$

- Task 3: The optimal choice for \mathscr{X} is the one that maximizes the expected confidence on the classification decisions

$$C^* = E_{\mathbf{X}}[c^*(\mathbf{x})] = E_{\mathbf{X}}\left[\max_{i} P_{Y|\mathbf{X}}(i|\mathbf{x})\right]$$
(6.3)

The computed value C^* is the *feature selection cost* for the Bayesian decision theory. Maximizing this expected confidence, is equivalent to minimizing the Bayes error

$$1 - E_{\mathbf{X}} \left[\max_{i} P_{Y|\mathbf{X}}(i|\mathbf{x}) \right]$$

- 2. Principle of Minimum Uncertainty:
 - Task 1: The decision rule gives the outcome $g'(\mathbf{x})$ defined as follows

$$g'(\mathbf{x}) = \arg\max_{i} \log P_{Y|\mathbf{X}}(i|\mathbf{x}) \tag{6.4}$$

- Task 2: The confidence value $c'(\mathbf{x})$ in declaring a feature \mathbf{x} as salient is obtained by relaxing the decision rule to the mean (i.e. taking an expectation)

$$c'(\mathbf{x}) = \sum_{i} P_{Y|\mathbf{X}}(i|\mathbf{x}) \log P_{Y|\mathbf{X}}(i|\mathbf{x})$$
(6.5)

The right hand side expression can be recognized as the negative of entropy, giving

$$c'(\mathbf{x}) = -H(Y|\mathbf{X} = \mathbf{x}) \tag{6.6}$$

- Task 3: The optimal choice for \mathscr{X} is the one that minimizes the expected confidence $E_{\mathbf{X}}[H(Y|\mathbf{X} = \mathbf{x})] = -H(Y|\mathbf{X})$ which corresponds to the minimization of the uncertainty of the classification decision. The feature selection cost is, therefore, $-H(Y|\mathbf{X})$

We see that the decision rule $g'(\mathbf{x})$ is equivalent to $g^*(\mathbf{x})$ and the confidence measure $c'(\mathbf{x})$ can be seen as the relaxation to the mean of the decision rule, i.e., mean of $\log P_{Y|\mathbf{X}}(i|\mathbf{x})$, or $E_{\mathbf{X}} \left[\log P_{Y|\mathbf{X}}(i|\mathbf{x})\right]$.

- 3. Barlow's principle of suspicious coincidences
 - Task 1: The decision rule proposed for this principle yields the outcome

$$g''(\mathbf{x}) = \arg\max_{i} \log \frac{P_{\mathbf{X},Y}(i,\mathbf{x})}{P_Y(i)P_{\mathbf{X}}(\mathbf{x})}$$
(6.7)

- Task 2: Relaxation of the decision rule to the mean gives the confidence measure $c''(\mathbf{x})$ for classification

$$c''(\mathbf{x}) = \sum_{i} P_{Y|\mathbf{X}}(i|\mathbf{x}) \log \frac{P_{\mathbf{X},Y}(i,\mathbf{x})}{P_{Y}(i)P_{\mathbf{X}}(\mathbf{x})}$$
(6.8)

A simplification of the right hand side expression in terms of mutual information yields $c''(\mathbf{x}) = I(Y; \mathbf{X} = \mathbf{x})$

- Task 3: Taking expectation of the confidence measure $c''(\mathbf{x})$ gives

$$\int \sum_{i} P_{Y|\mathbf{X}}(i|\mathbf{x}) \log \frac{P_{\mathbf{X},Y}(i,\mathbf{x})}{P_{Y}(i)P_{\mathbf{X}}(\mathbf{x})} d\mathbf{x}$$
(6.9)

which is the familiar expression for $I(\mathbf{X}; Y)$

$$I(\mathbf{X};Y) = \sum_{i} \int P_{Y|\mathbf{X}}(i|\mathbf{x}) \log \frac{P_{\mathbf{X},Y}(i|\mathbf{x})}{P_{Y}(i)P_{\mathbf{X}}(\mathbf{x})} d\mathbf{x}$$
(6.10)

The optimal choice for \mathscr{X} is the one that minimizes the expected confidence, i.e. $I(Y; \mathbf{X})$. Thus, the feature selection cost for this principle is $I(Y; \mathbf{X})$. Notice that the mutual information $I(Y; \mathbf{X})$ between the class label *Y* and the feature vector **X** can be also written as

$$I(Y;\mathbf{X}) = H(Y) - H(Y|\mathbf{X})$$
(6.11)

We see that maximizing $I(\mathbf{X}; Y)$ with respect to \mathbf{X} is equivalent to maximizing $-H(Y|\mathbf{X})$. Therefore, this computational principle gives a feature selection cost that is the same as the one given by the principle of minimum uncertainty even though the decision rules are different. This happens because a relaxation to the mean is applied to the Barlow's decision rule. This criterion for feature selection is referred to as the Infomax criterion.

Out of the given feature selection criteria, we need to adopt the one which is computationally least expensive.

It is seen that certain simplifications applied to the infomax feature selection criterion result in a formulation that is computationally parsimonious. Rewriting the feature vector \mathbf{X} more explicitly in terms of its *k* feature components

$$\mathbf{X}_{1:k} = \{X_1, X_2, \dots, X_k\},\$$

the selection criterion of Infomax can be rewritten as:

$$I(Y;\mathbf{X}) = \sum_{k} I(Y;\mathbf{X}_{k}) + \sum_{k} \left[I(X_{k};\mathbf{X}_{1,k-1}|Y) - I(X_{k};\mathbf{X}_{1,k-1}) \right]$$
(6.12)

Research has shown [Gao *et al.*, 2008] that some statistical properties of band pass filters, such as wavelet coefficients, extracted from natural images exhibit strongly consistent patterns of dependency across a wide range of natural image classes. However, such dependencies carry little information about the image class. This implies that the mutual information between features X_k and $\mathbf{X}_{1:k-1}$ given the knowledge of Y (i.e. $I(X_k; \mathbf{X}_{1,k-1}|Y)$) and the same same mutual information without the knowledge of Y (i.e. $I(X_k; \mathbf{X}_{1,k-1}|Y)$) are almost similar. Thus the second term of Eq. 6.12, which signifies the discriminant information is much smaller and can be ignored, thus yielding

$$I(\mathbf{X};Y) \approx \sum_{k} I(Y;\mathbf{X}_{k})$$
(6.13)

Reverse analysis reveals that this approximated feature selection cost corresponds to the expectation of a new confidence measure $c'''(\mathbf{x})$ given as

$$c'''(\mathbf{x}) = \sum_{k} I(Y|X_k = x_k)$$
(6.14)

Further reverse analysis shows that the confidence measure $c'''(\mathbf{x})$ corresponds to relaxation to the mean of the following decision rules:

$$g_k''(x) = \arg\max_i \log \frac{P_{Y,X_k}(i,x)}{P_{X_k}(x)P_Y(i)}, \quad k \in \{1,...,K\}$$
(6.15)

Each feature channel X_k gives its individual decision rule using $g''_k(x)$. The decision rule applied to the individual channels is considered as the marginal decision rule. Maximizing the approximated feature selection criterion $I(\mathbf{X};Y) \approx \sum_k I(Y;\mathbf{X}_k)$ is easier because each term being a mutual information is positive. Therefore we can select *k* features having the highest values of $I(X_k;Y)$. Computing $Y(Y;\mathbf{X} = \mathbf{x})$ using Eq.6.8 is simple for bandpass filters extracted from natural images.

This kind of computationally parsimonious feature selection is applicable for only the Barlow's principle. If a similar simplification is applied to the Bayes rule for feature selection,

then the approximation would be poor. For example, the feature selection cost for the minimum uncertainty principle cannot be well approximated as $H(Y|\mathbf{X}) \approx \sum_{k} H(Y|X_k)$ and therefore does not allow marginal Bayes decision rules such as

$$g_k^*(x) = \arg\max_i \log P_{Y|X_k}(i|x), \quad k \in \{1, \dots, K\}$$
(6.16)

Therefore, it can be concluded that the principle of suspicious coincidences yields computationally parsimonious feature selection cost and marginal decision rules. An interesting observation is that these marginal decision rules are more consistent with the psychophysics of human saliency, since humans find it easier to distinguish between target and background along a single feature, but not along conjunction of features, such as color and orientation. The holistic confidence measure (Eq.6.14) is the sum of the marginal confidence measures for the features.

6.1.2.1 Implementation of discriminant saliency

Deploying the discriminant saliency model requires addressing first the task of choosing the optimal features (Task 3), formulating a saliency measure for a feature (Task 2), and finally taking a decision on whether a feature is salient or not (Task 1).

Task 3: How to choose the optimal feature space \mathscr{X} ? (feature selection task)

Salient features are the ones which pass the following test:

$$\mathscr{S}_{k} = \{ x_{k} \mid H(X_{k}|Y=1) > H(X_{k}|Y=0) \}$$
(6.17)

This follows from the observation that discriminant saliency selects features which are present in the class of interest (Y = 1) and mostly absent in the null hypothesis(Y = 0). This translates into a distribution of features which is narrower and centered around 0 if the feature is absent from the null hypothesis (Y = 0), and leads to a broader distribution if the feature is present in the target class (Y = 1). A broader distribution contributes a higher value of entropy than a narrower distribution and therefore $H(X_k|Y = 1)$ is larger than $H(X_k|Y = 0)$ for the salient features. Eq.6.17 can be written as

$$\mathscr{S}_{k} = \left\{ x_{k} \left| \frac{P_{Y,X_{k}}(1,x_{k})}{P_{Y}(1), P_{X_{k}}(x_{k})} > \frac{P_{Y,X_{k}}(0,x_{k})}{P_{Y}(0)P_{X_{k}}(x_{k})} \right. \right\}$$
(6.18)

which can be simplified as

$$\mathscr{S}_{k} = \left\{ x \left| P_{X_{k}|Y}(x|1) > P_{X_{k}|Y}(x|0) \right. \right\}$$
(6.19)

Algorithmically, this task requires as input a set of images \mathscr{T}_1 that belong to the target class, a set of images \mathscr{T}_0 that belong to the null hypothesis, and a set of *N* features $X_k, k \in \{1, ..., N\}$. The target number of features to be selected is given as *K*. Selection is based on the value computed for $I(X_k, Y)$ for certain features X_k , as follows:

- 1. For each feature, compute $P_{X_k|Y}(x_k|i)$ and $P_{X_k}(x_k)$
- 2. Features which pass the test given by Eq.6.17 or 6.19 are retained and others are discarded.
- 3. For the retained features, compute $I(X_k, Y)$

- 4. The *K* features that give the largest values for $I(X_k, Y)$ are selected.
- Task 2: What is the confidence value associated with the classification of **x** ? (Saliency computation task)

The saliency value for the *k*th feature *x* is computed as:

$$S_k(x_k) = \begin{cases} I(Y|X_k = x_k), & \text{if } x_k \in \mathscr{S}_k \\ 0, & \text{otherwise} \end{cases}$$
(6.20)

The overall saliency measure $S_D(\mathbf{x})$ is the sum of the saliency measure over all the feature channels.

$$S_D(\mathbf{x}) = \sum_{k=1}^K S_k(x_k) \tag{6.21}$$

From the marginal decision rules given in Eq.6.15, the saliency measure for the individual features x_k can be interpreted as the (log) degree of suspicion

$$S_k(x_k) = \left\langle \log \frac{P_{Y,X_k}(i,x_k)}{P_Y(i)P_{X_k}(x_k)} \right\rangle$$
(6.22)

where
$$\langle f(x) \rangle = \sum_{i} P_{Y|X}(i|x) f(x)$$

Consider a set of interest points $I_1, ..., I_M$ extracted from a test image \mathscr{I} . The task of computing the saliency values for the interest points involves computing $S_D(\mathbf{x}_m)$ where \mathbf{x}_m is the feature extracted for each location I_m . Algorithmically, this involves evaluating the selected features X_k at the given locations as per the following steps:

- 1. Given the feature value x_{km} of feature component X_k computed at location I_m , compute $P_{X_k|Y}(x_{km}|i)$ for $i \in \{0, 1\}$.
- 2. Compute $S_k(x_{km})$ using Eq.6.20

Task 1: How x can be classified as salient or non-salient? (Declaring salient regions)

Instead of making a hard classification for the *m*th region/interest point characterized by feature $\mathbf{x}_{\mathbf{m}}$ as salient or non-salient, we can simply order the regions/interest points by decreasing discriminant saliency values $S_D(\mathbf{x}_m)$. Otherwise a suitable threshold value can be adopted to classify regions as salient or non-salient.

6.1.2.2 Estimating the models

An implementation for Task 3 requires constructing models for $P_{X_k|Y}(x_k|i)$, $P_{X_k}(x_k)$ and $I(X_k, Y)$, and also a way to do the condition check $H(X_k|Y=1) > H(X_k|Y=0)$. An implementation of Task 2 requires constructing models for $I(Y|X_k = x_k)$.

1. Computing $P_{X_k|Y}(x_k|i)$, $P_{X_k}(x_k)$

We make an assumption that the extracted features X_k have a probability distribution that can be well approximated by a Generalized Gaussian Distribution (GGD). A GGD can be defined using two parameters α and β as follows:

$$P_X(x;\alpha,\beta) = \frac{\beta}{2\alpha \Gamma(1/\beta)} \exp\left\{-\left(\frac{|x|}{\alpha}\right)^{\beta}\right\}$$
(6.23)

where the Gamma function $\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt$, t > 0. More specifically, the parameter α governs the scale of the function and β governs the shape (rate of decay from the peak value) of the distribution. Setting $\beta = 1$ gives the the Laplacian subfamily of GGD and setting $\beta = 2$ gives the Gaussian subfamily. The parameters α and β can be estimated using methods such as the method of [Sharifi and Leon-Garcia, 1995], maximum likelihood [Do and Vetterli, 2002], and minimum mean square error [Huang and Mumford, 1999]. Following [Gao *et al.*, 2009] we adopt the method of moments that exploits the relations of α , β with two quantities: variance σ and kurtosis κ of the distribution of X.

$$\sigma^{2} = \frac{\alpha^{2}\Gamma(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})} \quad \text{and} \quad \kappa = \frac{\Gamma(\frac{1}{\beta})\Gamma(\frac{5}{\beta})}{\Gamma^{2}(\frac{3}{\beta})}$$
(6.24)

The quantities σ and κ are estimated from image data, as follows:

$$\sigma^{2} = E_{X} \left[(X - E_{X}[X])^{2} \right] \quad \text{and} \quad \kappa = \frac{E_{X} \left[(X - E_{X}[X])^{4} \right]}{\sigma^{4}} \tag{6.25}$$

2. Computing $I(X_k, Y)$

$$I(\mathbf{X};Y) = \sum_{i} P_{Y}(i) \ KL\left[P_{\mathbf{X}|Y}(\mathbf{x}|i)||P_{\mathbf{X}}(\mathbf{x}))\right]$$
(6.26)

yielding,

$$I(X_k;Y) = \sum_{i} P_Y(i) \ KL\left[P_{X_k|Y}(x_k|i)||P_X(x_k))\right]$$
(6.27)

where $KL[p||q] = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}$ is the Kullback-Leibler (KL) divergence between the distributions $p(\mathbf{x})$ and $q(\mathbf{x})$

The KL divergence between two GGD distributions $P_X(x; \alpha_1, \beta_1)$ and $P_X(x; \alpha_2, \beta_2)$ can be written as:

$$KL[P_X(x;\alpha_1,\beta_1)||P_X(x;\alpha_2,\beta_2)] = \log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}$$
(6.28)

3. Computing $I(Y|X_k = x_k)$

The closed form expression for $I(Y|X_k = x_k)$ is given as

$$I(Y; X_k = x_k) = s[g(x_k)] \log \frac{s[g(x_k)]}{P_Y(1)} + s[-g(x_k)] \log \frac{s[-g(x_k)]}{P_Y(0)}$$
(6.29)

where $s(x) = (1 + e^{-x})^{-1}$ is a sigmoid function, and

$$g(x_k) = \left(\frac{|x_k|}{\alpha_0}\right)^{\beta_0} - \left(\frac{|x_k|}{\alpha_1}\right)^{\beta_1} + \log\left(\frac{\alpha_0\beta_1P_Y(1)\Gamma(1/\beta_0)}{\alpha_1\beta_0P_Y(0)\Gamma(1/\beta_1)}\right)$$
(6.30)

4. Evaluating $H(X_k|Y = 1) > H(X_k|Y = 0)$

Using the closed form expression for $H(X_k|Y = i)$, which is

$$H(X_k|Y=i) = \frac{1}{\beta_i} + \log \frac{2\alpha_i \Gamma\left(\frac{1}{\beta_i}\right)}{\beta_i},$$
(6.31)

we can simplify the condition check $H(X_k|Y=1) > H(X_k|Y=0)$ as

$$\log\left(\frac{\alpha_{1}}{\alpha_{0}}\right) > \left(\frac{1}{\beta_{0}} - \frac{1}{\beta_{1}}\right) + \log\frac{\Gamma\left(\frac{1}{\beta_{0}}\right)\beta_{1}}{\Gamma\left(\frac{1}{\beta_{1}}\right)\beta_{0}}$$
(6.32)

6.2 DATASET

To evaluate the method of Discriminant Saliency (DS) for annotation localization we use the same dataset of images as mentioned in the previous chapter 5 in Section 5.3.

6.3 FEATURE EXTRACTION

We use the same feature set as mentioned in the previous chapter 5 in Section 5.4.

6.4 SALIENCY THRESHOLD DETAILS

In order to generalize the detection system for a variety of documents we compute the threshold according to the statistics of the input image as:

$$Threshold_{Saliency} = Threshold_{scalar} \times mean(\mathbf{S_k})$$

This *Threshold*_{scalar} is dependent on the number of annotations present in the document. It is usually varied from $1 \leq Threshold_{scalar} \geq 5$. For pages with lesser number annotations the *Threshold*_{scalar} must be high.

6.5 EXPERIMENTS AND RESULTS

We have defined four sets of training and testing experiments:

- Set 1: The objective of this set of experiments is to localize *all* annotations. In this setting, the model is trained with the images comprising all the annotations.
- Set 2: The objective of this set of experiments is to localize the individual annotations in a multi-annotated document. In this setting, the model is trained with the images comprising only individual annotations.
- Set 3: The objective of this set of experiments is to localize textual annotations in a test document consisting symbolic annotations. In this setting, the model is trained with the images comprising only textual annotations with printed text as background.
- Set 4: The objective of this set of experiments is to localize individual annotations in a test document not consisting other types of annotations. In this setting, the model is trained with the images comprising only individual annotations.

6.5.1 Accuracy metrics for performance evaluation

In order to evaluate the DS model for annotation localization we use the same performance measures as described in the previous chapter 5 in Section 5.6.1.

6.5.2 All Annotation vs Printed Text

To localize all annotations together in a document, discriminant saliency method produces a recall of 0.58 for annotations and 0.82 precision for printed text. Table 6.1 and Figure 6.1 elaborates the results.

Table 6.1 : Set 1: Annotation localization when the dictionary is trained on images comprising all annotations and the testing is performed on similar images.

Annotation Category	Threshold Scalar	Accuracy	Recall	Precision	F1 Score	Execution Time (sec)
All (Fig. 6.1)	2.5	80.23%	.58	.82	.68	26.86



(a) Highlighted all Categories of Annotations Regions as Salient Objects

(b) Original Image

Figure 6.1 : Set 1: All kinds of Annotation Localization in Documents using DS.

6.5.3 Category-wise Annotation vs Printed Text

To localize *specific* annotations in a multi-annotated document each discriminant saliency model is trained with the images comprising only individual annotations. For underlined annotations, our model achieves a recall of 0.81 and precision of 0.47. For marginal text annotations, it produces 0.79 and 0.70 as recall and precision rates. In a similar manner, for encircled annotations, our model produces the recall and precision as 0.84 and 0.63 respectively, while for

inline annotations, recall of 0.18 and precision of 0.16 is produces. Table 6.2 illustrates the results for localizing individual annotations in a multi-annotated document. In the third set of experiments,

Table 6.2 : Set 2: Annotation localization when the dictionary is trained on individual annotations and the testing is performed on the images containing all the annotations.

Annotation Category	tion Threshold Ac Ty Scalar Ac		Recall	Precision	F1 Score	Execution Time (sec)
Underline (Fig. 6.2)	5	94.02%	.81	.47	.60	17.05
Marginal Text (Fig. 6.3)	5	95.04%	.7920	.7045	.7456	8.87
Encircled (Fig. 6.4)	5	95.81%	.84	.63	.72	17.11
Inline (Fig. 6.5)	6	82.03%	.18	.16	.17	8.61



(a) Highlighted Underlined Regions as Salient Objects

(b) Original Image

Figure 6.2 : Set 2: Underlined Region Localization in Multi-annotated Images using DS

we localized *textual* and *symbolic* annotations separately in a multi-annotated document. In this setting, the model is trained with the images comprising only individual annotations. For such set of experiments discriminant saliency shows impressive results and produces a recall and precision of 0.57 and 0.78 to locate textual annotations. It also produces a recall of 0.71 and precision 0.83 of to locate symbolic annotations on a multi-annotated document. Table 6.3 depicts the results for localizing only textual and symbolic annotations in a test document.

In the fourth set, *specific* annotations in a single-class annotated document are localized. For underlined annotations, discriminant saliency achieves a recall of 0.68 and precision of 0.94



(a) Highlighted Marginal Annotated Textual Regions as Salient Objects

(b) Original Image



for annotations and printed text respectively. For marginal text annotations, it produces 0.91 and 0.87 as recall and precision rates. In a similar manner, for encircled annotations, a recall of .87 and precision of .70 is obtained for annotations and printed text. For inline annotations, our model obtains a recall of 0.49 and precision of 0.52 for annotations and printed text. Table 6.4 presents the results for localizing individual annotations in a single-class annotated test document.

6.5.4 Results on Standard Datasets

It must be noted that there is a non-availability of a multi-annotated dataset. Therefore, likewise as stated in previous Chapter 5 in Section 5.6.4 we apply the DS method on IAM and PRImA-NHM datasets.

Our method shows impressive results on IAM dataset with a recall of .98 for handwritten text and a precision of .99 for the printed text. Similarly, we achieve a recall of .77 for handwritten text and precision of .66 for printed text in PRImA-NHM dataset. Table 6.5 presents the result for both the dataset using weakly supervised visual saliency. Figures 6.12 and 6.13 pictorially presents the results of DS learning for IAM dataset and PRImA-NHM dataset.

6.6 COMPARISON WITH SVM

Table 6.6 and 6.7 along with Figure 5.17 presents the effect of applying SVM on image patches. We applied RBF kernel and used two-class SVM. From the results it is clear that SVM is unable to capture the difference among the closed overlapping feature space among the printed

Oplain it in detail explanation.	Explain it in detail explanation.	
where (42)	where $q = \frac{e^{n_t + \sum_{l' \in N_t} \beta_{l,l'} f_{l'}}}{1 + e^{\alpha_t + \sum_{l' \in N_t} \beta_{l,l'} f_{l'}}} $ (42)	
The corresponding energy takes the following form	The corresponding energy takes the following form	
$U(f \bigoplus_{\{i\}\in \mathcal{C}_{i}} \sum_{\ln} \ln \left(\overset{s}{\underset{f_{i}}{\overset{s}}} \right) - \underbrace{\bigotimes}_{\{i\}\in \mathcal{C}_{i}} \alpha_{i}f_{i} - \underbrace{\bigoplus}_{\{i,i'\}\in \mathcal{C}_{i}} \beta_{i,i'}f_{i}f_{i'} $ (43)	$U(f) = -\sum_{\{i\}\in\mathcal{C}_1} \ln\binom{M-1}{f_i} - \sum_{\{i\}\in\mathcal{C}_1} \alpha_i f_i - \sum_{\{i,\ell'\}\in\mathcal{C}_2} \beta_{i,\ell'} f_{i\ell'} $ (43)	
It reduces to the auto-logistic model when Λ	It reduces to the auto-logistic model when $M = 1$.	
An auto-model is said to be an <i>auto-normal model</i> , also called a <u>Gaussian</u> $\underline{MHF}[21]$ the head set \mathcal{L} is the real line and the joint distribution is multivariation in the solution of the set of the	An auto-model is said to be an <i>muto-normal model</i> , also called a <u>Gaussian</u> <u>MRF</u> [21], if the label set \mathcal{L} is the real line and the joint distribution is multivariate normal. Its conditional $p.d.f.$ is	
$p(f_i \mid f_{ii}; \bigoplus_{j \in \mathcal{M}_i} \bigoplus_{j \in \mathcal{M}_i} \theta_{ij} r(f_j - \mu_d))^2 $ mornul distribution (14)	$p(f_i \mid f_{\mathcal{N}_i}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{8\sigma^2} (f_i - \mu_i - \sum_{i' \in \mathcal{N}_i} \beta_{i,i'} (f_{i'} - \mu_{i'}))^2} $ (44) normal distinction	
It is the normal distribution with conditional mean	It is the normal distribution with conditional mean	
	$E(f_i \mid f_{N_i}) = \mu_i - \sum_{x \in N_i} \beta_{i,x}(f_i - \mu_i) $ (45)	
and conditional variance $(f_i + f_{N_i}) = $	and conditional variance $(1 + f_{N_i}) = C^2$ $(46) Z$	
The joint probability is a Gibbs distribution $p(f) = \frac{p(f)}{p(f)} = \frac{p(f)}{p(f)} = \frac{p(f)}{p(f)} = \frac{p(f)}{p(f)}$	The joint probability is a Gibbs distribution $p(f) = \frac{\sqrt{\det(B)}}{\sqrt{(2\pi\sigma^2)^m}} \underbrace{e_{r,p}^{(4\pi)}}_{e_{r,p}}$ equation	
where f is viewed as a vector, μ is the $m \times 1$ vector of the conditional means, and $B \bigoplus_{i \neq j} i_i$ is the $m \times m$ interaction matrix whose elements are unity and oil-flagginal offendin at (i, i) is $-\beta_{i, q}$, i.e. $b_{i, q} \bigoplus_{i \neq j} -\beta_{i, q'}$ with $\beta_{i, j} = 0$. Therefore, the single-site and pair-site clique potential functions for the auto- normal model are	where f is viewed as a vector, μ is the $m \times 1$ vector of the conditional means, and $\beta = [b_{n'}]$ is the $m \times m$ interaction matrix whose elements are unity and off-diagonal defined at (i, i') is $-\beta_{n'}$, i.e. $b_{n'} = \delta_{n'} - \beta_{n'}^{-1} - \beta_{n'}^{-1}$ with $\beta_{n'} = 0$. Therefore, the single-site and pair-site clique potential functions for the auto- normal model are	
Explain the work in more than two formes with tables.	$\frac{[V_i(f_i) = (f_i - \mu_i)^2/2\sigma^2]}{(V_i(f_i) = (f_i - \mu_i)^2/2\sigma^2)} \xrightarrow{(48)}$ $\frac{20}{(160)}$ Enplain the work in more than two forms with tables.	• Encircled
(a) Highlighted Encircled Regions as Salient	(b) Original Image	

Highlighted Encircled Regions as Salient Objects

Figure 6.4 : Set 2: Encircled Region Localization in Multi-annotated Images using DS

text and annotations. The most probable reason could be the presence of variety of annotations in a document rather than only text.

6.7 CONCLUSION

Discriminant saliency has been demonstrated to identify specific annotations in a multi-oriented cluttered document. Our experimental results corroborate that discriminant saliency produces better results in comparison to a discriminative classifier such as SVM. It shows comparable results with the CRF based supervised saliency model. It is observed that the overall recall produced for all the experiments is high for supervised saliency model mentioned in previous Chapter 5. Our weakly supervised learned model for annotation extraction performs well for densely annotated documents. While dealing with unconstrained handwriting environment for annotations, in the subsequent chapter 7 we propose a method to detect baseline for unconstrained handwritten word. This allows to separate the core zone from the ascenders and descenders and therefore leads to effective extraction of features for preprocessing and writer identification.

. . .

The surface which minimizes one of the smoothness prior energy alone has either a constant group even a constant gradient or a constant guivature. This is undesirable because constraints from other sources such as the data are not used. Therefore, a smoothness term [CD] is usually utilized in conjunction with other energy terms. In regularization, an energy consists of a smoothness term and a closeness term and the minimal solution is a compromise between the two constraints. The surface which minimizes one of the smoothness prior encoded and the surface which minimizes one of the smoothness prior encoded at the surface which a constant group level, a constant gradient or a constant group level. Therefore, a smoothness term level is under the strange of the surface str The encodings of the smoothness prior in terms of derivatives usually lead to <u>Isotropic potential</u> functions. This is due to the assumption that the underlying surface is non-textured. Ansisterbrie priors have to be used for texture patterns. This can be done, for example, by choosing (37) with direction-dependent $\frac{1}{12}$ s. The encodings of the smoothness prior in terms of derivatives usually lead to <u>terminic potential</u> functions. This is due to the assumption that the underlying surface is non-textured. Anisotropic priors have to be used for texture patterns. This can be done, for example, by choosing (37) with direction-dependent V_2 's. Oflain it in det with reference for Enflain it in detail with reference papers 4.4 Hierarchical GRF Model of don'ts wirk about the 4.4 Hierarchical GRF Model > don'ts wink abrivations A hierarchical two-level Gibbs model have the provided the represent both noise-contaminated and textured image to the higher level Gibbs distribution uses an isotropic random field, e.g. which is the characterize the blob-like region formation process. A lower level Gibbs distributions. This provides a convenient spinor of f of MD-MRF modeling. In segmentation of noisy and textured image $D_{\rm eff}$ is MD-MRF modeling. In segmentation of noisy and textured image $D_{\rm eff}$ of MD-MRF modeling. In segmentation of noisy and textured image $D_{\rm eff}$ of MD-MRF modeling. In segmentation of noisy and textured image $D_{\rm eff}$ of MD-MRF modeling. In segmentation of noisy and textured image $D_{\rm eff}$ of MD-MRF modeling. In segmentation of noisy and textured image $D_{\rm eff}$ of $D_{\rm eff}$ of A hierarchical two-level Gibbs model has heen property to represent both noise-contaminated and textured image [34,35] The higher level Gibbs dis-tribution press an isotropic random field, e.g. MLL, to characterize the blob-like region formation process. A lower level Gibbs distribution describes the filling-in in each region. The filling-in may be independent hoses or a type of texture, both of which can be characterized by Gibbs distributions. This provides a convenient approach for MAP-MRF modeling. In segmentation of noisy and textured image [34,35,80,66,135] for example, the higher level determines the prior of f for the region process while the lower level (<u>Gibbs contributes</u> to the conditional probability of the data given f. Note that different levels of MRFs in the hierarchy can have different neighborhood systems. Various <u>hierarchical Gibbs models</u> result according to what are chosen for the regions and for the filling-in's, respectively. For example, each region may be filled in by an <u>auto-norrecture</u> [120,135] or an <u>auto-homolar (pure</u> [66]; the MLL for the region multiplication may be substituted by another appropriate MRF. Various <u>hierarchical Gibbs models</u> result according to what are chosen for the regions and for the filling-in's, respectively. For example, each region may be filled in by an <u>auto-normal texture [102]</u>, 135] or <u>an auto-hiormal texture [60]</u>, the MLL for the region formation may be substituted by another appropriate MDP. MRF. A drawback of the hierarchi <u>model</u> is that the conditional probability $P(d_i \mid f_i = I)$ for regions given by $\{i \in S \mid f_i = I\}$ par not always be written exactly. For grapple, when the lower level MHC is restrict modeled as an auto-normal field, its joint distribution over an **(E)** larly shows as much known. This difficulty may be overcome by using approxime solutions such as pseudo-likelihood [13,14] (a proof of the <u>ssiendo-likelihood</u> citizelikelihood estimate is given in [47]) or by using the <u>C-analysis</u> method [137]. A drawback of the hierarchical model is that the conditional probability $P(d_i \mid f_i = I)$ for regions given by $\{\tilde{t} \in S \mid f_i = I\}$ par not always be written exactly. For aggraphic, when the lower level MIRU is a taxture modelad as an auto-normal field, its joint distribution over an <u>orregularly shaped regio</u> is not known. This difficulty may be overcome by using approximate schemes such as pseudo-likelihood [13,14] (a proof of the consistency of the pseudo-likelihood estimate is given in [47]) or by using through model [137]. 26 26 Explain with in detail, they can be superimeted and can be worked on in Explais each in detail, they can be Inline-text esperimeted and can be worked on in future future Annotation

(a) Highlighted Inline Annotation Regions as Salient Objects

(b) Original Image

- Figure 6.5 : Set 2: Inline Annotation Region Localization in multi-annotated Images using DS
- Table 6.3 : Set 3: Annotation localization when the dictionary is trained on images comprising only textual annotations and the testing is performed on similar images, and vice versa.

Annotation Category	Threshold Scalar	Accuracy	Recall	Precision	F1 Score	Execution Time (sec)
Textual (Fig. 6.6)	4	88.30%	.57	.78	.66	20.25
Symbolic (Fig. 6.7)	5	92.89%	.71	.83	.77	17.54

Table 6.4 : Set 4 : Annotation localization when the dictionary is trained on individual annotations andthe testing is performed on the images containing individual annotations.

Annotation Category	Threshold Scalar	Accuracy Recall		Precision	F1 Score}	Execution Time (sec)
Underline (Fig. 6.8)	2.5	92.39%	.68	.93	.79	18.02
Marginal Text (Fig. 6.9)	5	94.03%	.91	.87	.89	15.61
Encircled (Fig. 6.10)	6	92.12%	.87	.70	.77	20.56
Inline (Fig. 6.11)	6	80.13%	.49	.52	.51	8.61

Sinche	Shiefe
	Enplais why a hadon
	3.3 Gibbs Handom Freids
A set of random variables F is said to be a <u>Glube random field</u> (site) on F with respect to N if and only if its configurations obey a <u>Glube stribute</u> A Glube distribution takes the following form	A set of random variables F is said to be a <u><i>Gibba</i> random field (GRF</u>) on S with respect to N if and only if its configurations obey a <i>Gibbs</i> distribution. A Gibbs distribution takes the following form
$(P(f) = \textcircled{P} \times e^{-\frac{1}{2}W(f)} $	$ \begin{pmatrix} P(f) = Z^{-1} \times e^{-\frac{1}{2}U(f)} \end{pmatrix} $ (24)
where	where
	$Z = \sum_{f \in \mathbb{F}} e^{-\frac{1}{T} U(f)} \qquad \qquad (25)$
is a normalizing constant called the partition function, T is a constant called the temperature which shall be assumed to be 1 unless otherwise stated, and U(f) is the recent function. The energy	is a normalizing constant called the <i>partition function</i> , T is a constant called the <i>temperature</i> which shall be assumed to be 1 unless otherwise stated, and U(f) is the <i>energy function</i> . The energy
	potential (U(1) = I(V_1(1)) - (26)
is a sum of <i>clique potentials</i> $V_c(f)$ over all possible <u>cliques</u> . The value of $V_c(f)$ depends on the local configuration on the clique c . Obviously, the Gaussian distribution is a special member of this Gibbs distribution family.	is a sum of clique potentials $V_{\epsilon}(f)$ over all possible cliques C . The value of $V_{\epsilon}(f)$ depends on the local configuration on the clique ϵ . Obviously, the Gaussian distribution is a special member of this Gibbs distribution family.
A GRF is said to be <u>homogeneods</u> if $V_c(f)$ is independent of the relative po- sition of the clique c in S. It is said to be isotropic if V_c is independent of the orientuation of c. It is considerably simpler to specify a GRF distribution if i.e.	A GRF is said to be <u>homogeneous</u> if $V_c(f)$ is independent of the relative po- sition of the clique c in S. It is said to be isotropic if V_c is independent of the orientation of c. It is considerably simpler to specify a GRP distribution if it.8
homogeneous or isotropic than one without such properties. The homogeneity is assumed in most Mitchholets for mathematical and computational conve- pence. The isotromy is a property of <u>direction-independent</u> blob-like rogions.	homogeneous or isotropic than one without such properties. The homogeneity is assumed in most MRL shoulds for mathematical and computational conve- nience. The isotropy is a property of direction-independent blob-like regions.
To calculate a Gibbs distribution, it is necessary to evaluate the partition function which is the sum over all possible configurations in F. Since there	To calculate a Gibbs distribution, it is necessary to evaluate the partition function Z which is the sum over all possible configurations in \mathbb{F} . Since there
"The acombinatorial number of elements in <i>P</i> for a discrete subscription in the second secon	rea combinatoria number of elements in F for a discrete Cass illust a od in section 2.2, the evaluation is prohibitive even for problems of moderate sizes. Several Approximation methods exists for solving this problem.
P(f) measures the probability of the occurrence of a particul manufaction, or <u>pattern</u> f. The more probable configurations are those with lower con-	P(f) measures the probability of the occurrence of a particular configuration, or "pattern" f . The more probable configurations are those with lower ener-
temperature is high, all configurations tend to be equally distributed. Near	gies. The temperature T controls the sharphess of the distribution. When the temperature is high, all configurations tend to be equally distributed. Near
minima. Given T and $U(f)$, we can generate a class of <u>"patterns"</u> by sampling the configuration space F according to $P(f)$.	 the zero temperature, the distribution concentrates around the global energy minima. Given T and U(f), we can generate a class of <u>"patterns"</u> by sampling the configuration space <u>F according to P(f)</u>.
	MRF models are mathematical models
	I and they need to be implemented with
	l'complete variable set rize. Annotation

(a) Highlighted Textual Regions as Salient Objects

(b) Original Image

n

Figure 6.6 : Set 3: Textual Region Localization in Documents comprising both Textual and Symbolic Annotations using DS

Table 6.5: Annotation localization on IAM and PRImA dataset by DS textual annotation detection model

Dataset	Testset	Accuracy	Precision	Recall	F1-score	Execution Time (sec)
IAM	100 images	98.73%	.99	.98	.98	.20
Dataset	_					
PRImA	100 images	75.95%	.66	.77	.71	.23
NHM						
Dataset						

Tab	ole 6.6 : Con	nparison of S	SVM with DS o	n multi-a	innotated do	ocuments.	

Annotation	All Annot	ll Innotations		Underline		Marginal Text		Encircled		
Category	DS	SVM	DS	SVM	DS	SVM	DS	SVM	DS	SVM
Accuracy (%)	80.23	44.09	94.02	89.09	95.04	55.89	95.81	87.17	82.03	49.19
Precision	.82	.50	.47	.23	.70	.21	.63	.46	.16	.42
Recall	.58	.41	.81	.36	.79	.13	.84	.06	.18	.06
F1 Score	.68	.45	.59	.28	.75	.16	.72	.11	.17	.10
Execution Time (sec)	26.86	14.49	17.05	8.17	8.87	.41	17.11	17.15	8.61	18.49



Figure 6.7 : Set 3: Symbolic Region Localization in Documents comprising both Textual and Symbolic Annotations using DS

Annotation Underline		line	Marginal Text		Enciro	led	Inline	
Category	DS	SVM	DS	SVM	DS	SVM	DS	SVM
Accuracy (%)	92.39	84.97	94.03	80.59	92.12	84.19	80.13	74.09
Precision	.94	.91	.87	.96	.70	.92	.52	.82
Recall	.68	.30	.91	.04	.87	.01	.49	.01
F1 Score	.79	.45	.89	.08	.77	.01	.51	.02
Execution Time (sec)	18.02	18.35	15.61	24	20.56	21.49	8.61	19.23

 Table 6.7 : Comparison of SVM with DS on single-class annotated documents.



(a) Set 3: Highlighted Underlined Regions as Salient Objects

(b) Original Image





(a) Highlighted Marginal Text Regions as Salient Objects

(b) Original Image

Figure 6.9: Set 4: Marginal annotations localization in single-class annotated images using DS

GHAN ET AL.: SCM COMPARISON

(a) Highlighted Encircled Regions as Salient

densed water due to horizontal advection (tor internal boundary condition). SCMs can be applemented with more detailed mod-dis, which can be implemented with more detailed mod-dis, which can be implemented with more detailed mod-ion, which parameterizing the analler-scale turbulent ins, which parameterizing the analler-scale turbulent ions, which parameterizing the analler-scale turbulent motions. CEMs are designed to jumilate the docing scale processes that must be parameterized in a GCM 26 or SCM. ACCD domain may be considered to repre-se a GCM grid columns, so in a same, a CEM and be viewed as an "extremely detailed COOP for this rea-ion is is useful to evaluate SCM county in part by com-parison with CED output. A CEM typically includes a turbulence parameterization, a bulk loophase micro-physics parameterization, a bulk loophase micro-physics parameterization, a bulk loophase micro-physics parameterization, a SCM, observed large-scale variation of a Large-scale fields and tenden-tions. The observations of a CEM are the same as those required by accounts of a CEM are the same as those required by accounts of a CEM are the same as those required by microphysics, nat-tion, and turbulence parameterizations, which intro-duce major uncertainties. Nevertheless, CEM esti-tion, and turbulence parameterizations, which intro-duce major uncertainties. Nevertheless, CEM results and, mail y compared with the latter. This strategy of the tasting of parameterizations, which intro-duce major uncertainties. Nevertheless, CEM results and CEMs has been embraced by the GEWEEX Cloud System Suddy COSS Science Yeam, 1993, Mon-crief et al, 1997 (EWEEK is the Jobs). Mor-crief dia, 1997 (EWEEK is the JOB). Mor-crief dia, 1997 (EWEEK tal pressure

1, which is polymerical up use Con-Department of 35%-ing the necessary SCM and COM lateral bound-conditions from measurements has proven to be ex-sidy challenging, largely because of sampling and surement errors in the winds<u>COMP</u> and <u>Comp</u>. 1997. In *Access and Comp* 2006, *Readell et al.*, 1996. In *Add et al.*, and, for highly advective conditions the lack of cloud securacy of the simulation in the interior of the le colume<u>Cetch</u> and *DTMED* 1998, under such littors the clouds are lack SCM is not a suitable bed for cloud surgentiering the securacy of SCM is not a suitable and for ender surgentiering the securacy of the securation of the securacy of the securation of the securation of the securacy of the simulation in the interior of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the securation of the securacy of the securation of the measurements along the lateral boundaries can degrade the scrumer of the simulation in the interior of the simulation simulations of atmospheric prot-esses (Line comparisons) with indeendent observations conditions the clouds are largely controlled by the laterative alternative alternative scrumer of the SCM is an attractive alternative to the GCM Yet the SCM is an attractive alternative to the GCM is a parameterization tested for to the SCM is an attractive alternative scrumer between the SCM is an attractive alternative to the GCM is a parameterization tested for to the SCM is an attractive alternative scrumer between the SCM is an attractive alternative to the GCM is a parameterization tested of could be alternative scrumer between the SCM runs much faster than a GCM and hence can

(b) Original Image

the differences the second purposed models. Thus the second purposed models, to the extent possible without sys replacing parameterizations of ¢very process If the ultimate goal of single-column mor development of parameterizations of atmos development of atmospherications of atmospherications of atmospherications of atmospherications of atmospherications development of atmospherications of atmosphe diff es can lead to impr ons of every process)

ary conditions. Although multiple models will be us these issues because the response of one pa is not necessarily representative of the re ers, the purpose of this paper is not to the net understand differences between ers, the pur solve and u rstand diff Given the iliation of It is intery that a complete contentation by different models will require a syst tion of the parameterizations of each study is beyond the scope of this pap some cases it is possible to identify the ences between simulations by models

can't upper to beness ratio there which moust set of Three are however, which we have a set of the three are however, which we have a set of the mannetrization verification, and development area are related to the selection of an optim oxy for using SCMs as parameterizations first purpose of this paper is therefore to a on methodological methodores (SCMs as para-testbeds. These Control and SCMs are para-testbeds. These Control and the set of the procedure used to provide the parameterization. procedure used to provide the lateral bo ditions for the simulations, (2) the treat lower boundary conditions in the simulati the treatment of the vertical and horizon lower boundary conditions in the simulati the treatment of the vertical and horizon tendencies in the simulations, given the la

wide range of meteorological condi-action developer can use an SCM ' der climatic conditions found in mu-world, without the computational b-ulations. Second, SCM simulations (der climate genuausse vordagen world, without the computational burr ulations. Second, SCM simulations are than the GCM simulations. By initial SCM simulations with observed could daily or even hourly observations, pro-timmediate test of the parameterizati tions of interest. <u>ECM simulations</u> cally days to veels rather than monit There are howays: several issues

be used to quickly test n cting an SCM do main that ex ological conditions, use an SCM to test

Objects

Figure 6.10 : Set 4: Encircled annotation localization in single-class annotated images using DS



Encircled Annotation

GHAN ET AL.: SCM COMPARISON

ensed water due to horizontal advection (for lateral oundary conditions). SCMs can be supplemented with more detailed mod-

use of a codebook. However, instee of encoding the features using the code-words, we bioit the discriminative properties of features that belong to the same clus in a supervised approach. We also propose a new fittend for senarative procenders and descenders from the state of the discriminative properties of the state of the sta

2 Related Work

2 Related Work
Subscription of machine in and hand-written text has been an active area of reference in the intermediate segments in the intermediate i like densely or sparsely annotated docume (a) to ments with cluttered annotations within write statistical states and the states of the state

(a) Highlighted Inline Annotated Regions as Salient Objects

use of a codebook. However, instead of encoding the features using the code- words, we exploit the discriminative properties of features that belong to the same cluster, in a supervised approach. We also proposed a new method for separating ascenders and descenders from an unconstrained handwritten word and identifying its core-region [Pandey and Harit, 2017b]. We used the structural properties of ascenders and descenders to identify the writers of the given words. We are able to achieve writer identification rates close to 63% on the handwritten words drawn from a dataset by 10 writers. In addition to the contribution of saliency and spectral partitioning based annotation detection study and separating ascenders and descenders, we also create the dataset for the problem along with the ground truth.

2 Related Work

2 Related Work
Character of machine-printed and hand-written text has been an active area of research. It common with the contribution of (Kuhnke et al., 1995) for printed and hand-written character sgemen-diation using directional and symmetrical features into a neural network. Pel and Chandhuri, 1999 (and the contribution of Kuhnke et al., 1995) for printed and hand-written character sgemen-diation using directional and symmetrical features into a neural network. Pel and Chandhuri, 1999 (and the contribution of Kuhnke et al., 2002) using the spent of a set of the state of the symmetrical problem of the and written characters from a multi-language document of printed Chinese characters. Following this (Zheng et al., 2002b) presented a bottom up approach printed Chinese characters. Following this (Zheng et al., 2002b) presented a bottom up approach on segmetrical problem into handwritten characters from a multi-language document of the symmetrical problem into handwritten text intes, words or characters. Well-separated printed and handwritten text are seasy to separate but the problem approach based printed Chinese (The Sheng et al., 2013) used Chem and Machines, words or characters. Well-separated printed and handwritten text are approach based printed problem on profile of the word. They used HMM to find overlay and handwritten of anot printed problem and the spent in the printed and handwritten text at piel level. Recently, freque et al. (2014) the authors proposed a seady on a combination for for non-combined environments where the type and orientation of annotations are combination for hom combined environments where the type and printed text and then applied coarsening that descriptors and texture features. This method is designed by examining its square neighter of printed, noise and annotations from the printed text. They also applied for separated printed, noise and annotation for the printed text. They also applied for separation printed, noise and annotation printed text and then separate spaces

spaces. Most of the previous methods deal with controlled environment to discriminate between printed text and annotations. They include pre-segmented words or text lines which are then further classified into respective classes. Datasets used by them are simple with well-genarated hand-written text in a predefined layout. With further research. More complex annotated documents were processed with multi-oriented nature, however, the misclassification rate is high. In this context, we devise an ap-proach to identify all possible types of annotations that normally readers make on a document while reading or editing. In our work we segmented a wide variety of complex annotations. Our system not only performs well in real environment but also gives good performance at run time.

(b) Original Image

Olnline-text Annotation

Figure 6.11: Set 4: Inline annotation localization in single-class annotated images using DS

Sentence Database Sentence Database B05-038 B05-038 Hence the hostility to automation and the stop-watch manufacturing methods that have led to restrictive practices. Now a new threat to those who toil and spin has been developed by a firm specialising in electronics in Los Angeles. They have developed a new system whereby completely untrained workers can be taught their trade by means of tape recordings and television. Hence the hostility to automation and the stop-watch manufacturing methods that have led to restrictive practices. Now a new threat to those who toil and spin has been developed by a firm specialising in electronics in Los Angeles. They have developed a new system whereby completely untrained workers can be taught their trade by means of tape recordings and television. 99992 - 989 Left 673 97, au Que Qu **A** Hence the hostility to automation and the stop-watch manufactoring methods that have bes led to restrictive prodices. Nove a new ⁴4694,466 <u>66</u>94,966,466 6674,666 threat to those who toil and spin has been developed by a from specialising in electronics මහත්වර්ග වෙන්ඩාදුණු අත්ර අත්ර 1990. **Malan (Sangjard (Sang**) in Los Angeles. They have developed a new system whereby completely watroined workers can be tought their trade by means of tape recordings and television. +200:00 • Textual Name Name: Annotation

(a) Highlighted Textual Annotated Regions as Salient Objects

(b) Original Image





Salient Objects

Figure 6.13 : Textual region localization in PRIMA images using DS.