Topic Masks for Image Segmentation

Young-Seob Jeong[1], Chae-Gyun Lim[2], Byeong-Soo Jeong[2] and Ho-Jin Choi[1]
[1]Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST),
291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea
[e-mail: {pinode, hojinc}@kaist.ac.kr]
[2]Department of Computer Engineering, Kyung Hee University,
1-Seocheon-dong, Gyeonggi-do, Yongin-si 446-701, Republic of Korea
[e-mail: {rayote, jeong}@khu.ac.kr]
*Corresponding author: Ho-Jin Choi

Received August 21, 2013; revised November 21, 2013; accepted November 29, 2013; published December 27, 2013

Abstract

Unsupervised methods for image segmentation are recently drawing attention because most images do not have labels or tags. A topic model is such an unsupervised probabilistic method that captures latent aspects of data, where each latent aspect, or a topic, is associated with one homogeneous region. The results of topic models, however, usually have noises, which decreases the overall segmentation performance. In this paper, to improve the performance of image segmentation using topic models, we propose two topic masks applicable to topic assignments of homogeneous regions obtained from topic models. The topic masks capture the noises among the assigned topic assignments or topic labels, and remove the noises by replacements, just like image masks for pixels. However, as the nature of topic assignments is different from image pixels, the topic masks have properties that are different from the existing image masks for pixels. There are two contributions of this paper. First, the topic masks can be used to reduce the noises of topic assignments obtained from topic models for image segmentation tasks. Second, we test the effectiveness of the topic masks by applying them to segmented images obtained from the Latent Dirichlet Allocation model and the Spatial Latent Dirichlet Allocation model upon the MSRC image dataset. The empirical results show that one of the masks successfully reduces the topic noises.

Keywords: Topic mining, image segmentation, topic mask

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government, Ministry of Science, ICT & Future Planning(MSIP) (No. 2010-0028631).

http://dx.doi.org/10.3837/tiis.2013.12.018

1. Introduction

Image segmentation is the process of partitioning an image into disjoint and homogeneous regions and it is one of the most difficult tasks in image processing. As images typically do not have tags or labels, unsupervised methods are drawing attention. A topic model is such an unsupervised probabilistic method that captures latent aspects of data, where each latent aspect, or a topic, is associated with one homogeneous region.

Topic models should be designed by considering properties or types of data. Although there have been many topic models for document analysis, new topic models and several additional processes are required for analyzing images because images have properties that are different from documents. The additional processes include definitions of a word, a document, and a corpus for an image dataset. The word definition process is called codebook learning and is closely related to vector quantization [12]. The codebook in images plays the role of vocabulary in documents, and it typically consists of local descriptor vectors of patches. As a descriptor vector of each patch usually has many features in a real-number form, the codebook size will exponentially grow when the codebook is defined as being the set of all possible descriptor vectors. To reduce the size, k-means algorithm is typically used to group the set of all descriptor vectors into k clusters, which results in a vocabulary consisting of k unique descriptor vectors. A local descriptor vector of each patch is quantized into one of the k unique descriptor vectors by employing a distance measurement (e.g., a Euclidean distance). Based on the obtained k unique descriptor vectors, each image can be represented using a bag-of-features (BOF) or a bag-of-words (BOW) fashion. Different topic models usually have different definitions of a word, a document, and a corpus.

Since the Probabilistic Latent Semantic Analysis (PLSA) [1] model and the Latent Dirichlet Allocation (LDA) [2] model were introduced, many revised or extended topic models have appeared. For example, there are some topic models for capturing a sequential pattern of topics [3, 4, 5], and for entity-entity relationships [6, 7]. For image segmentation, there are topic models whose objective is to assign a homogeneous region to the same topic [8, 9, 10, 11]. However, we observed that the homogeneous regions obtained from topic models usually have noises, where the noises are inappropriately assigned topics to each homogeneous region. The noises cause degradation of the segmentation performances, so we propose new masks to reduce the inappropriately assigned topics. There are two contributions of this paper. First, the new topic masks can be used to reduce noises of topic assignments obtained from topic models for image segmentation tasks, so this approach increases the segmentation accuracies. Second, we prove the effectiveness of the masks by segmentation performance. The rest of the paper is organized as follows. Section 2 discusses related studies, and Section 3 describes our approach in detail. Section 4 presents experiments and results. Finally, Section 5 concludes the paper.

2. Related Work

There are many topic models for image segmentation tasks, and the models can be divided into two categories according to how they define the codebook, word, and document in the image domain. The first category employs over-segmentation as a preprocessing step, where each segment is regarded as a word. The over-segmentation divides images into small segments, where patches of each segment have similar descriptor vectors. Topic models in this category are designed to merge the small segments into a set of segments that compose a homogeneous region. In contrast, the second category does not utilize the over-segmentation preprocess, so each patch itself is regarded as a word.

The topic models of the first category require images to be first over-segmented. Each over-segmented region plays the role of a word, while each image is considered a document. Zhao et al. [8] proposed the topic random field and applied a Gaussian noise model to the codebook, so online codebook learning was available. Li et al. [9] proposed a framework which concurrently solved three tasks, classification, annotation, and segmentation. Cao and Fei-Fei [10] proposed the Spatially Coherent Latent Topic (SC-LT) model which was based on the hypothesis that pixels should have the same latent topic assignments if they are in a spatially close region with similar features. All of these models require images to be over-segmented, and the over-segmentation is typically done using graph-based algorithms [13, 14]. It is obvious that the performance of these models strongly depend on the over-segmentation algorithms because a set of over-segmented regions are used to generate a codebook that composes the topics. Moreover, over-segmentation takes a long time [15], so it is not practical for a big set of images.

Topic models of the second category do not require over-segmentation. They usually consider each patch or pixel as a word. When images are segmented by topic models of this category, each patch has a topic assignment where each topic is associated with a homogeneous region. The objective of the topic models is to assign each homogeneous region to the same topic without topic noises. The simplest way is to apply the LDA model to images, where each image is regarded as a document. The LDA model does not use any spatial information, so it has a poor segmentation performance relative to other models of the same category. Niu et al. [16] proposed the Spatial-DiscLDA model for visual recognition; it used spatial information and labels of images. Burns and Corso [17] segmented degraded images of documents using topic models. They utilized prior knowledge of the layout of the topics, and modeled it using a Potts-like Markov Random Field (MRF). Wang and Grimson [11] proposed the Spatial Latent Dirichlet Allocation (SLDA) model based on the hypothesis that pixels should have the same latent topic assignments if they are spatially close and have similar features. The hypothesis of the SLDA model is the same as the hypothesis of the SC-LTM [10] model of the first category. The biggest difference between the two models is that the SC-LT model requires over-segmentation while the SLDA model does not. As the second category does not require over-segmentation, it is more practical than the first category for a big set of images. As the size of unlabeled image data grows, the second category will become more desirable than the first category. However, there is a practical drawback of the second category, which is that the topic assignments usually have topic noises which degrade the segmentation performance. We propose, therefore, new masks to reduce the topic noises, which will result in an improvement of the segmentation performance of the topic models of the second category. With the performance improvements using the new masks, the second category will be more practical for a big set of images.

There are many masks for different purposes, such as image blurring, sharpening, blob detection, edge detection, noise elimination, and so on. For example, the Gaussian mask and Fourier mask are usually used for image blurring, while the Prewitt mask and Roberts mask are used for edge detection. Each mask has its unique structure that is represented as a composition of real values. The masks for image blurring reduce the effect of noise on images by replacing every pixel by a weighted average of its neighbor pixels. The new masks proposed in this paper reduce the effect of the noises of topic assignments for a segmentation task, not the noises of images. To the best of our knowledge, this is the first study on masks applicable to such topic assignments of segmented images.

3. Topic Mask

When a topic model of the second category is applied to images for segmentation tasks, each patch is assigned to a topic. As each topic is associated with a homogeneous region, every patch of each homogeneous region should be assigned to the same topic. We define the noises of topic assignments as the inappropriately assigned topics in each homogeneous region. The topic noises cause degradation of the segmentation performance, so we propose two new topic masks to reduce the topic noises.

There are several differences between an image mask and a topic mask. First, the image mask is applied to pixels of images, while the topic mask is applied to the topic assignments of images. More specifically, the topic mask filters out the inappropriately assigned topic assignments or topic labels, while the image mask filters out some pixels. Second, image masks are typically based on mathematical distributions or derivatives. As pixels are represented by a certain color space (i.e., RGB, XYZ), it is possible to compare the strength of a pixel with that of other pixels. Thus, image masks are designed to capture patterns of the strength of pixels by mathematical distributions or derivatives. In contrast, it is impossible to compare the strength of a topic with that of other topics because each topic has its own semantic meaning. In other words, if we regard a topic as a homogeneous region, then the topics are not numerical values, so the nature of topics is different from the nature of pixels. A topic mask, therefore, should be designed according to the different natures of the topics.

We propose two topic masks: the Frequency mask (F-mask) and the Connection mask (C-mask). Given mask size *S*, for each *n*-th patch as a center patch, the masks are iteratively applied to *S*×*S* topic assignments surrounding the center patch to determine the new topic assignment of the center patch, where each topic is regarded as a homogeneous region. Each mask is designed on its own hypothesis. The hypothesis of the F-mask is that the most frequent topic within the mask is most likely to be assigned to the center patch. An algorithm of the F-mask is as follows.

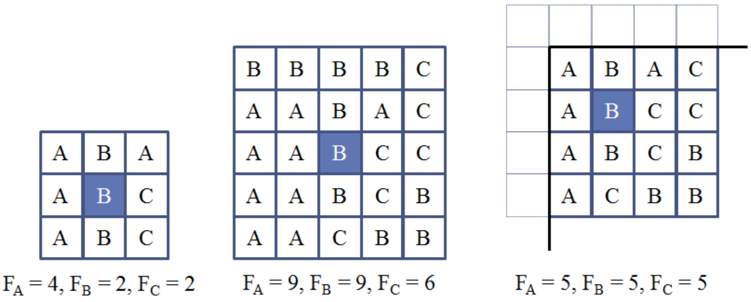
Algorithm – Frequency mask

```
Input:        (1) data,                          (2) mask size $S$,
              (3) the total number of topics $T$, (4) the number of total iterations $I$
Initilization: The topic assignments of all patches are obtained by a topic model.

for each step $i$ of total $I$ steps,
    for each $n$-th patch as the center patch,
        $W_n = S \times S$ patches surrounding the $n$-th center patch.
        Topic assignment $z_n$ of the center patch is removed.
        for each topic $t$ of total $T$ topics,
            $F_t$ =  the number of patches assigned to topic $t$ within $W_n$.
        end
        $F_{MAX} = max\ ( F_t )$.
        $TS_{1:M} = \{ t \mid F_t = F_{MAX} \}$ where $M$ = the number of items in this set.
```

```
            If $M$ = 1, then $z_n = argmax_t(F_t)$,
            Otherwise, $z_n$ = randomly picked item from $TS_{1:M}$.
        end
    end
```

Examples of applying the F-mask to topic assignments when the number of total topics is 3 are depicted in **Fig. 1**, where *A*, *B*, and *C* represent the three topics. When *S* = 3, as in the left side of the figure, the most frequent topic within the mask is the topic *A*. Note that the topic of the center shaded patch is not counted. With different setting of *S*, the most frequent topic may be different, as shown in the center of **Fig. 1**. In particular, if F_A is the same as F_B in the center of the figure, then the F-mask randomly chooses one of the two candidate topics *A* and *B*. When the center patch is located at the edge or corner of the image as in the right side of the figure, only the patches within the mask are used to get the new topic assignment.



A	B	A
A	B	C
A	B	C

$F_A = 4, F_B = 2, F_C = 2$

B	B	B	B	C
A	A	B	A	C
A	A	B	C	C
A	A	B	C	B
A	A	C	B	B

$F_A = 9, F_B = 9, F_C = 6$

	A	B	A	C
	A	B	C	C
	A	B	C	B
	A	C	B	B

$F_A = 5, F_B = 5, F_C = 5$

Fig. 1. Examples of applying the F-mask to topic assignments when *S* = 3 (left) and *S* = 5 (center, right). The *S*×*S* patches including the shaded center patch are assigned to one of three topics, *A*, *B*, or *C*. The notation F_x represents the number of patches assigned to topic *x* within the *S*×*S* patches.

The F-mask is based only on the frequency of topics, so it may lose important information such as positions of the topic assignments. We observed that topic noises are usually sparsely distributed, rather than forming a mass. In other words, if a topic assignment is part of such a mass, then it should not be replaced with another topic assignment. Based on the observation, we designed the C-mask on the hypothesis that the original topic assignment should be kept if it is a part of a mass; otherwise it will be replaced with a new topic, which can be obtained from the F-mask. As it includes the algorithm of the F-mask, it can be seen as an extension of the F-mask. An algorithm of the C-mask is as follows.

Algorithm – Connection mask

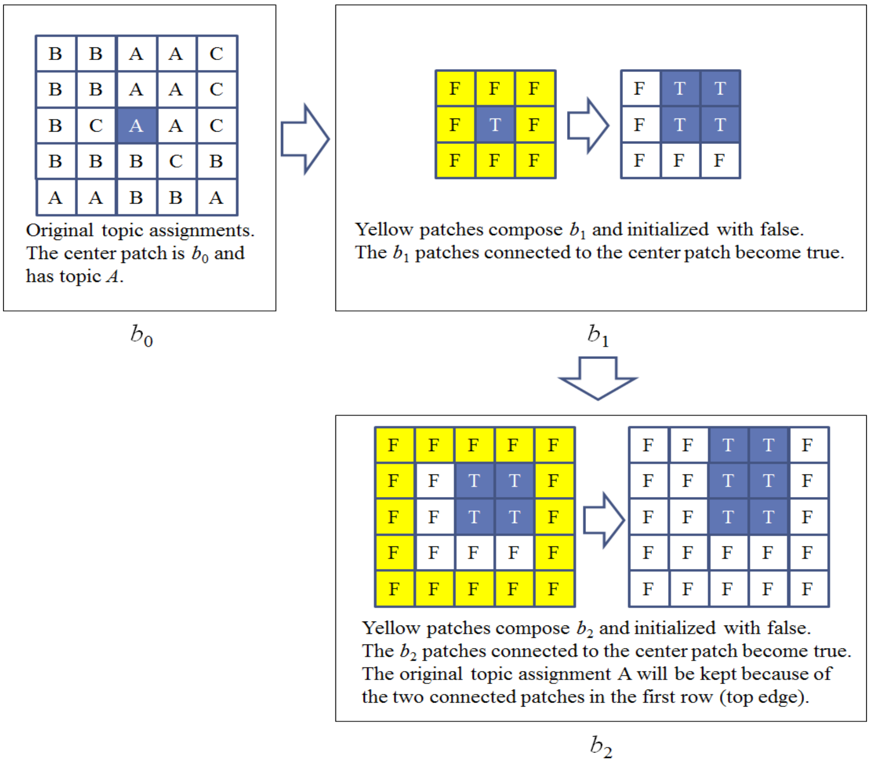
```
Input:      (1) data,                        (2) mask size $S$,
            (3) the total number of topics $T$, (4) the number of total iterations $I$
            (5) smallest mass proportion $P$
Initilization: The topic assignments of all patches are obtained by a topic model.

for each step $i$ of total $I$ steps,
    for each $n$-th patch as the center patch,
        $W_n$ = $S$×$S$ patches surrounding the $n$-th center patch.
        $z_n$ = topic assignment of the center patch.
        $B_{0:(S\text{-}1)/2}$ = {$b_k$ | $b_k$ is the $k$-th border box where $0 \le k \le (S\text{-}1)/2$},
                    where every patch of $b_k$ is either $true$ or $false$.
        Set every patch of $B_{0:(S\text{-}1)/2}$ to be $false$, except the center patch to be $true$.
        $Mass_{true}$ = 1.
        $Mass_{total}$ = the number of patches of $B_{0:(S\text{-}1)/2}$.
        $isConnection$ = $false$.
        for $k$ = 1 to $(S\text{-}1)/2$,
            $Mass_k$ = 0.
            for each $x$-th patch $b_{kx}$ of $b_k$,
                $True_{(k\text{-}1)x}$ = a set of patches of $b_{k\text{-}1}$ that are adjacent to $b_{kx}$ and are $true$.
                If $|True_{(k\text{-}1)x}|$ >= 1, then ( $b_{kx}$ = $true$, $Mass_{true}$ += 1, and $Mass_k$ += 1 ).
                Otherwise, $b_{kx}$ = $false$.
            end
            If ($Mass_{true}$ / $Mass_{total}$) >= P, then ( $isConnection$ = $true$ and break )
            Else if $Mass_k$ == 0, then ( $isLocalMass$ = $false$ and break ).
        end
        If $isConnection$ == $true$, then keep the original topic assignment $z_n$.
        Otherwise, $z_n$ is obtained using F-mask with 1 step.
    end
end
```

An example of applying the C-mask to topic assignments when $S = 5$ is depicted in Fig. 2, where A, B, and C represent the three topics. To check whether a topic assignment is part of a mass, we use the number of patches connected to the patch. Assuming that we apply the C-mask to patch x, and it is obvious that the patch can have maximum 8 adjacent patches. If some of the adjacent patches have the same topic assignments as patch x, then we denote them as connected patches. This connection can be spread out to all the patches within the mask, so the number of connected patches will be between 1 and S×S. The C-mask determines whether the original topic assignment of each center patch is worth keeping or not based on the number of patches connected to the center patch.

To get the connected patches, we divide the set of S×S patches in the mask into $(S\text{-}1)/2$ border boxes. We denote the center patch as b_0, and the patches adjacent to the center patch compose the border box b_1. The outer patches adjacent to b_1 compose b_2, and so on. We then generate an S×S binary mask whose items are either true or false, where the x-th item is true when the x-th patch in the mask is connected to the center patch. Note that the bottom-left item of b_2 in Fig. 2 is false although the corresponding patch has the same topic assignment as the center patch, because the bottom-left patch is not connected to the center patch.

After we get the S×S binary mask, we need to decide whether the original topic assignment of the center patch will be kept or not. The parameter P is used to make a decision for each patch, where $0 \le P \le 1$. For example, in Fig. 2, $Mass_{true}$ is 6 and $Mass_{total}$ is 25. If $P = 0.25$, then the original topic B of the center patch will be discarded, because $(6/25) = 0.24 < P$. In this case, the center patch will be assigned with a new topic obtained from the F-mask. As the parameter P has a normalized value, a bigger $Mass_{true}$ will be required to keep the original topic assignment with a greater mask size. With greater settings of P, the C-mask will be closer to the F-mask. The reason is that *isConnection* is more likely to be false with a greater P, so almost all the topic assignments will be replaced by the F-mask. The parameter P can be seen as a regulator between the F-mask and the C-mask, so it is necessary to obtain the appropriate setting of the parameter. The setting should be done according to the number of topics, because $Mass_{true}$ will become smaller when the number of topics increases, so P should be smaller when the number of topics increases.



Original topic assignments. The center patch is b_0 and has topic A.

Yellow patches compose b_1 and initialized with false. The b_1 patches connected to the center patch become true.

Yellow patches compose b_2 and initialized with false. The b_2 patches connected to the center patch become true. The original topic assignment A will be kept because of the two connected patches in the first row (top edge).

Fig. 2. An example of applying the C-mask to topic assignments when $S = 5$, where T and F represent true and false, respectively.

There are five factors that affect the performance improvements of the masks. First, the performance improvements are strongly dependent on the topic model, because the topic masks are applied to topic assignments obtained from the models. Second, if the number of topics is inappropriately set, then worse topic assignments will be generated from the topic models. The topic masks are designed to be applied to the topic assignments, so the masks will have poor performance improvements given worse topic assignments. Third, with different settings of the total iteration I, the result of the mask process might be different. As the topic masks have a chance to change the topic assignments for each iterative step of the topic mask process, it will probably give different topic assignments for each step. Fourth, with different sizes of masks, the result of topic assignments will be different. With different mask sizes, in other words, it will have different topic frequencies and different connected patches, which will result in different topic assignments. Fifth, with respect to the C-mask, the parameter P will affect the performance improvements because the parameter is a regulator or a boundary for deciding whether the original topic assignment of each patch should be kept or not. We will show the performance improvements with various settings of these factors by experiments.

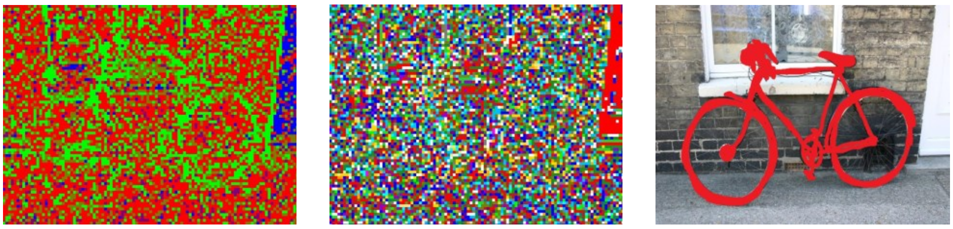
To summarize, we propose two topic masks for reducing topic noises. The F-mask gets a new topic assignment for each patch based on the frequency of topic assignments within the mask. The F-mask uses only the frequency of topic assignments, so it may lose important information such as position patterns of topic assignments. Based on the observation that the topic noises typically are sparsely distributed, the C-mask counts the number of patches connected to the center patch within the mask, and keeps the original topic assignment when the proportion of connected patches is greater than or equal to given parameter P. As the C-mask includes the F-mask, it can be seen as an extension of the F-mask. It is worth noting that the topic masks do not consider the semantic relationships among the topics. The masks just filter the noises out using structural patterns of the assigned labels. Thus, this method may not guarantee that the results of the masks be more comprehensible by human, even if the segmentation performance is improved.

4. Experiment

As the proposed topic masks are designed for reducing the topic noises on segmented images obtained from topic models, we demonstrate the effectiveness of the masks by showing the improvements of the segmentation performance. The performance is measured using F1 score. We apply the masks to the topic assignments obtained from two topic models, the LDA model [2] and the SLDA model [11]. We symmetrically set $\alpha = 0.1$ and $\beta = 0.01$ for the two models. Particularly, for the SLDA model, we use 20 replicated particles for each patch, because we observed that it generally showed the best performance with 20 particles for each patch. The parameter σ of Gaussian kernel is set as 30 for the same reason.

We used the MSRC image dataset [18] with 140 images including a sheep, bicycles, a cow, a bird, a car, and an air-plane. The two topic models are trained with 1,000 iterations. To obtain the local descriptors, we use the filter bank of [18], which consists of three Gaussians, four Laplacian of Gaussians, and four first order derivatives of Gaussians. Rather than using only local descriptors of interest points, we divide each image into patches on a grid and densely sample a local descriptor for each patch, as described in [11], where the size of each patch is 6×6. A codebook with a size of 200 is generated by applying k-means clustering with a Euclidean distance to the all local descriptors of the images.

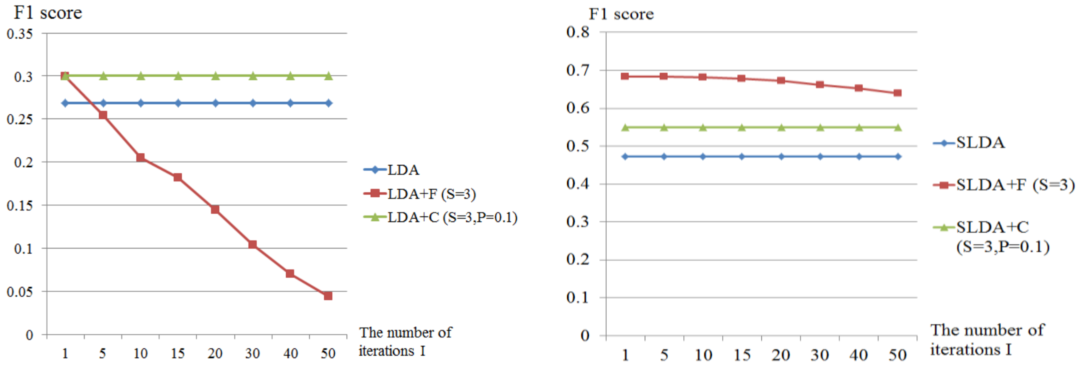
As described earlier, the performance improvements are affected by settings of several factors: the topic models, the number of topics T, the number of mask iterations I, the size of masks S, and the parameter P. Therefore, in this section, we focus on showing how much the factors affect the performance improvements. With respect to the first factor, we apply the topic masks to either the LDA model or the SLDA model for all the experiments and compare the performance improvements between the two cases. Without the topic masks, the SLDA model generally outperforms the LDA model on the segmentation task, so we can see how much the topic masks depend on the segmentation performance of the topic models. For the second factor, we varied the number of topics T from 3 to 20, and we observed that the two topic models have the best segmentation results when $T = 5$. We do not plot the segmentation performances of various settings of T in this paper, because the models show significantly low performances with different settings of T and even the results were not comprehensible by people. Sample topic assignments are depicted in **Fig. 3**, where each color represents each topic. When $T = 20$ in the figure, it has many small chunks and it is not comprehensible. In other words, it is impossible even for people to recognize whether chunk is noise or not. On the other hand, when $T = 5$, it shows relatively comprehensible results, and we can recognize which small dots are topic noises. As the objective of topic masks is to reduce the topic noises, the topic masks are applicable only for the case of $T = 5$ in our observation. We fix, therefore, T = 5 for the following experiments.



(a) (b) (c)

Fig. 3. The samples of topic assignments (segmentation results) obtained from the LDA model. The case of $T = 5$ (a), the case of $T = 20$ (b), and a ground-truth image (c).

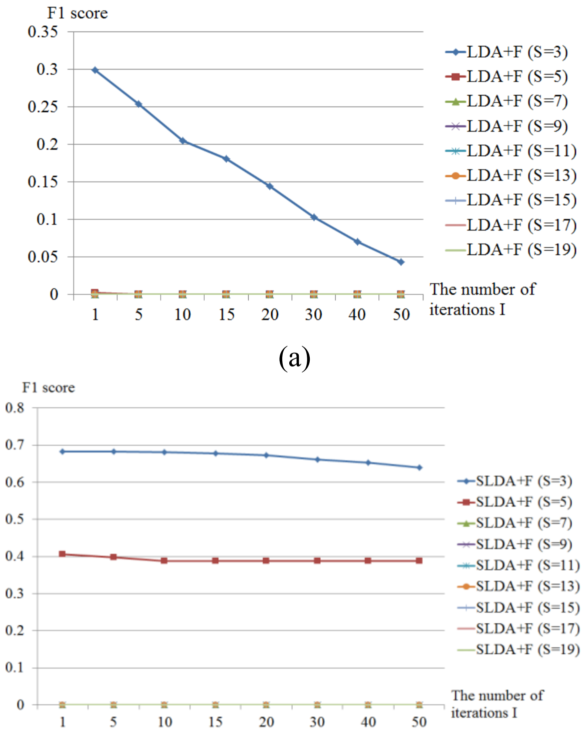
With respect to the third factor, we performed experiments with various settings of the number of iterations I, where the other factors T, S, and P were fixed as 5, 3, and 0.1, respectively. The results are shown in **Fig. 4**, where the horizontal axis represents the number of iterations I and the vertical axis indicates the F1 score. As shown in the two figures of **Fig. 4**, the performance of the topic model alone was fixed as no mask was applied to it. For both models, the performance improvements of the F-mask were unstable and did not converge. This implies that it changes topic assignments to worse state for almost all steps, because it replaces every topic assignment with the most frequent topics without considering any other information, such as positions of topic assignments. Note that the plot of LDA+F significantly decreased as the number of iterations I increased, while the plot of SLDA+F gently decreased. The reason for this is that the SLDA model itself incorporates spatial information, so it absorbs the drawback of the F-mask. In the case of the C-mask, for both of the models, it shows relatively stable performance improvements after the first step as it employs the positions of topic assignments. Note that the performance improvements of the C-mask in **Fig. 4** are not its best, and the C-mask can achieve better performances with different settings of S and P.



(a) (b)

Fig. 4. The result of the LDA model (a) and the result of the SLDA model (b) with various settings of the number of iterations I, where the horizontal axis represents I and the vertical axis indicates F1 score. LDA+F means the LDA model applied with the F-mask, while LDA+C indicates the LDA model applied with the C-mask.

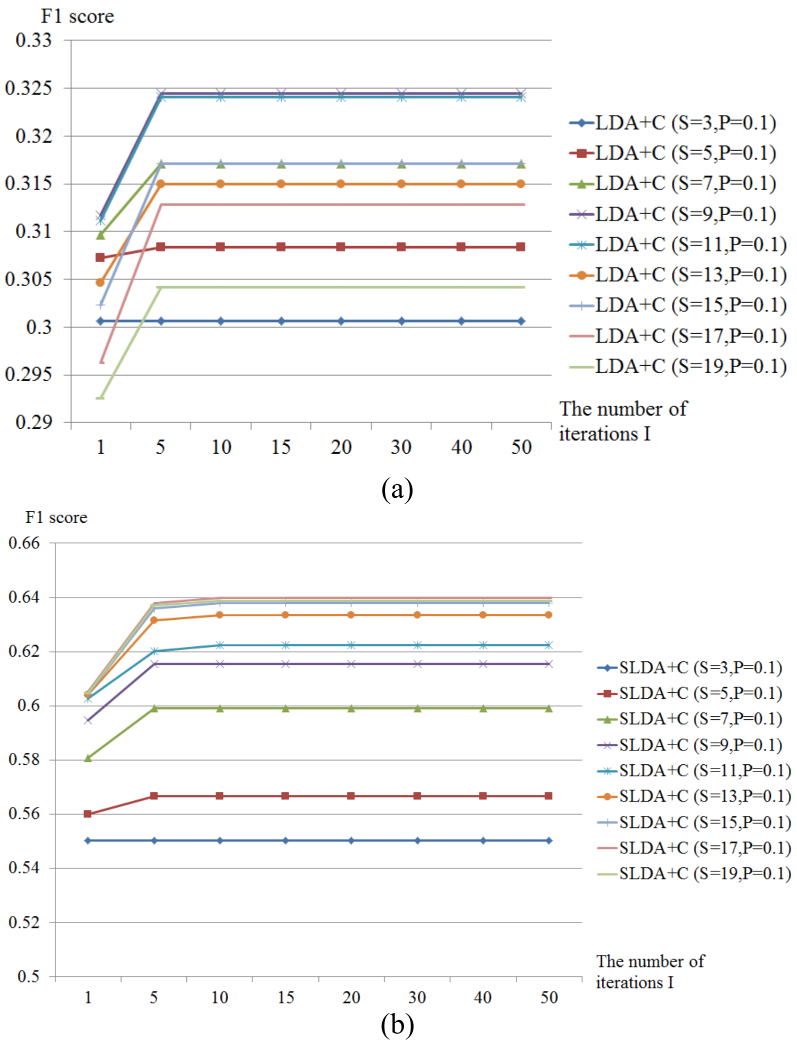
We varied the mask size *S*, which is the fourth factor, from 3 to 19, and performed 50 mask iterations, where *T* and *P* were fixed as 5 and 0.1, respectively. The result of the F-mask and the result of the C-mask are shown in **Fig. 5** and **Fig. 6**. As shown in **Fig. 5**, the F-mask shows a reasonable result only when *S* = 3, which means that the bigger size of the F-mask causes a faster transition to a worse state every step. Therefore, the F-mask has the best performance improvements when *S* = 3 and *I* = 1. With respect to the C-mask in **Fig. 6**, it shows best performance improvements when *S* = 9 and *S* = 17 for the LDA model and the SLDA model, respectively. This implies that each homogeneous region consists of connected patches which can be well captured by a 9×9 C-mask in the LDA results and by a 17×17 C-mask in the SLDA results. It is interesting that the performance improvements of the C-mask increased between *I* = 1 and *I* = 5, except for the case of *S* = 3. To be specific, with a bigger mask size *S*, the performance improvements tend to change more as *I* increases. The reason is that the bigger size of the C-mask causes more changes by the F-mask, as the C-mask is an extension of the F-mask. That is, the proportion of connected patches is less likely to be greater than *P* with a bigger size of the C-mask, so more topic assignments will be replaced by the F-mask. Although the performance decreases when the F-mask alone is used as depicted in **Fig. 5**, the F-mask within the C-mask makes a transition to a better state. The reason for this is that the C-mask holds some topic assignments according to their position, and the F-mask within the C-mask changes only remaining topic assignments. That is, the C-mask acts as a teacher in that it teaches whether each topic assignment should be replaced or not.



(a)

(b)

Fig. 5. The F-mask result on the LDA model (a) and the result on the SLDA model (b) with various settings of *S* and *I*, where the horizontal axis represents *I* and the vertical axis means the F1 score. LDA+F indicates the LDA model applied with the F-mask.

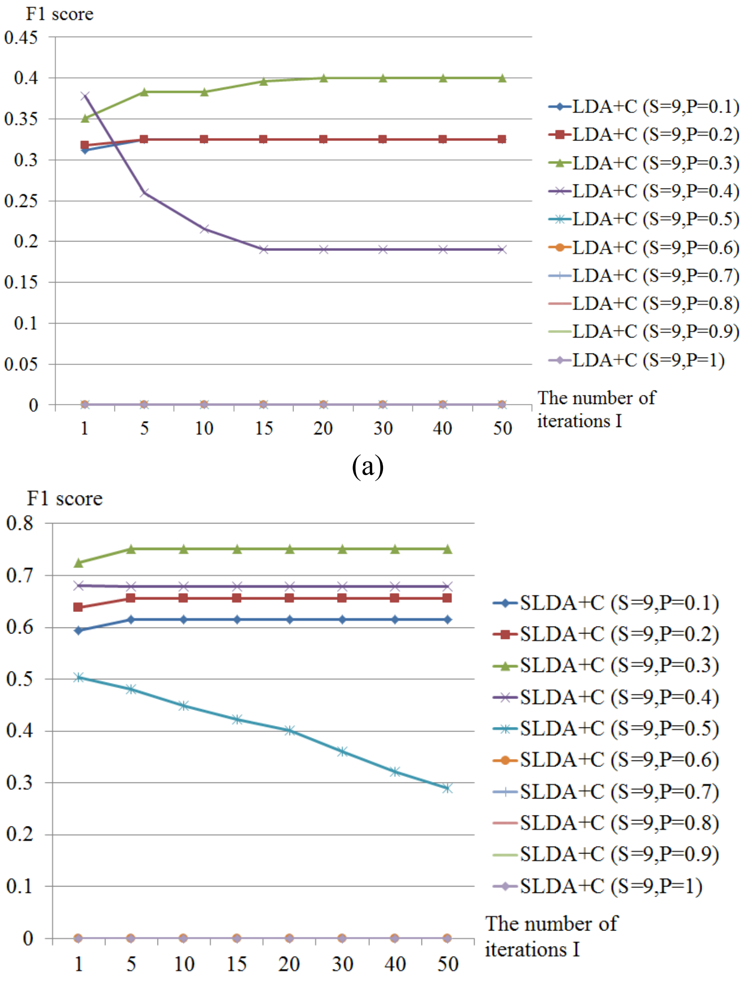


(a)

(b)

Fig. 6. The C-mask result on the LDA model (a) and the result on the SLDA model (b) with various settings of *S* and *I*, where the horizontal axis represents *I* and the vertical axis indicates the F1 score. LDA+C indicates the LDA model applied with the C-mask.

We also varied the parameter *P*, which is the fifth factor, from 0.1 to 1, and performed 50 mask iterations. The results are shown in **Fig. 7**, where $T = 5$ and $S = 9$. The C-mask shows the best performance improvements for both models when $P = 0.3$. Note that the performance of LDA+C is rather significantly decreased when $P = 0.4$, and it even falls to zero when $P \geq 0.5$. The performance of SLDA+C is similarly decreased when $P = 0.5$, and it falls to zero when $P \geq 0.6$. The reason for this phenomenon is that the C-mask acts more like the F-mask with a greater *P*. For example, if we assume that $P = 1$, then the C-mask will be the same as the F-mask because the positions of topic assignments are meaningless for satisfying $Mass_{true} = Mass_{total}$. If we see the plots of the F-mask in **Fig. 5**, they are decreased when $S > 3$, which is similar to the plots of the C-mask in **Fig. 7** when $P \geq 0.5$.

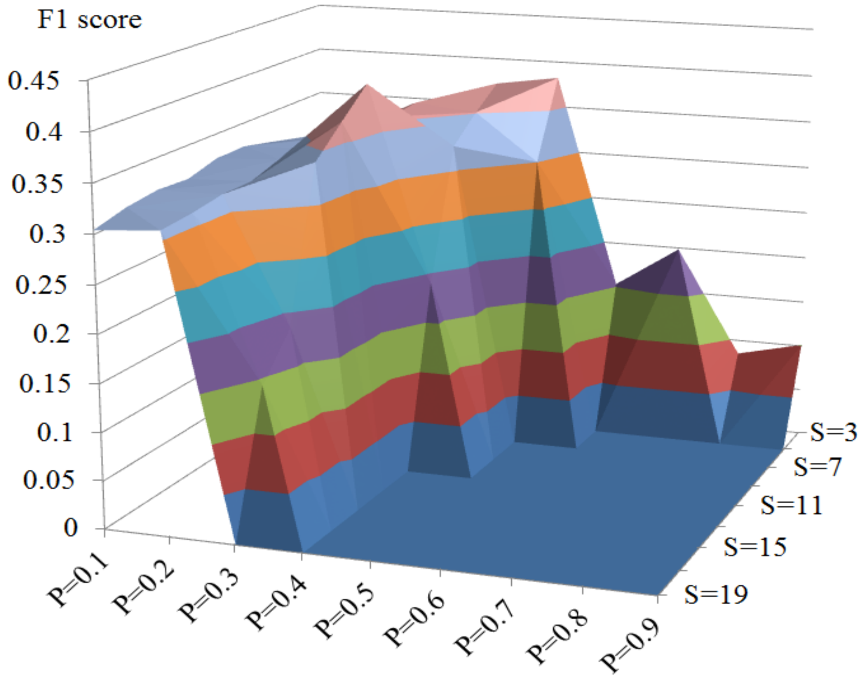


(a)

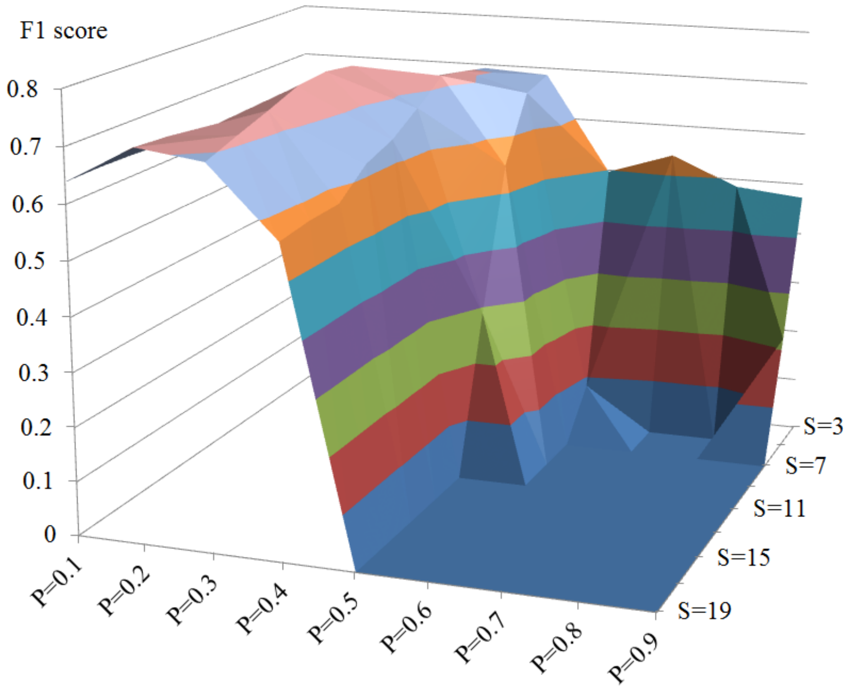
(b)

Fig. 7. The C-mask result on the LDA model (left) and the result on the SLDA model (right) with various settings of *P* and *I*, where the horizontal axis represents *I* and the vertical axis indicates the F1 score. LDA+C indicates the LDA model applied with the C-mask.

We investigated the impact of each factor of topic masks on performance improvements. The F-mask showed best performance improvements when $S = 3$ and $I = 1$. We still, however, need to discover the optimal setting of the C-mask. In **Fig. 8**, the performance improvements of the C-mask with the overall parameter settings are plotted. LDA-C has the best performance improvement when $P = 0.3$ and $S = 9$. It is worth noting that the F1 score goes to zero as *P* and *S* increase. SLDA-C shows similar results, and it has the best performance improvement when $P = 0.3$ and $S = 11$. These results imply that the proposed topic masks would be useful only when the parameters *T*, *I*, *S*, and *P* are adjusted according to the object type, image size, and the topic models. Thus, to make the masks more practically useful, it will be necessary to find a way of setting the parameters automatically. We plan to discover such a way as the future work using more state-of-the-art topic models and datasets (e.g., PASCAL VOC).



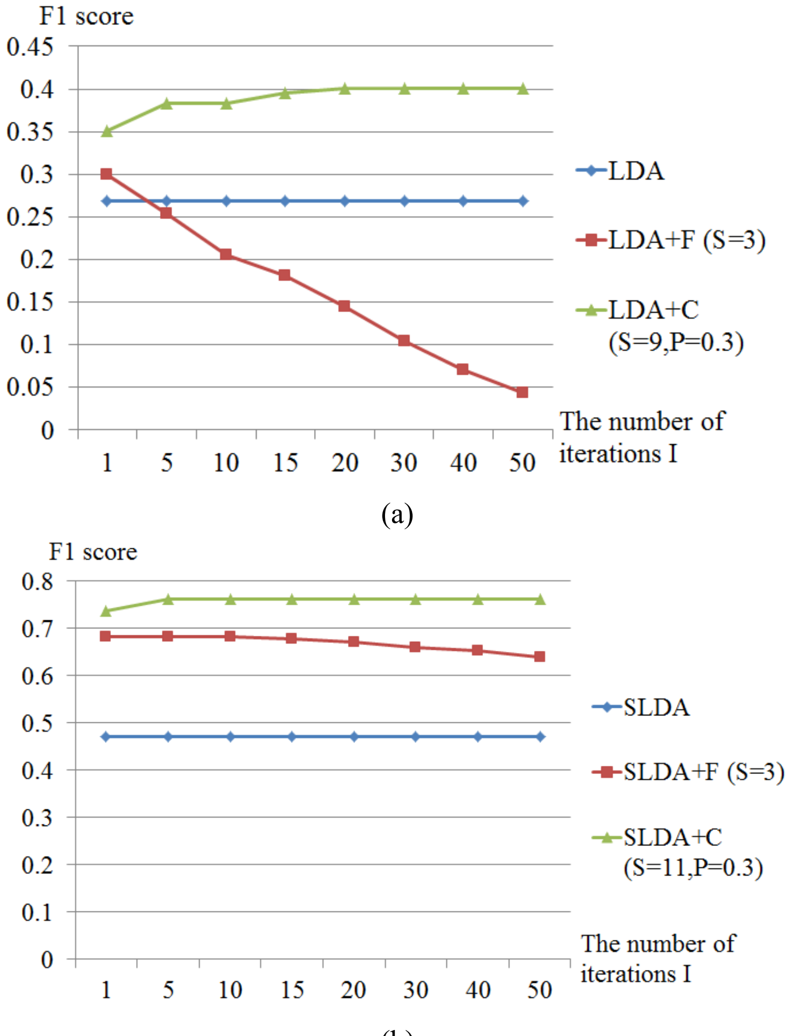
(a)



(b)

Fig. 8. The performance improvements of the C-mask on the LDA model (left) and on the SLDA model (right) when $T = 5$ and $I = 50$. The horizontal axis represents I and the vertical axis means the F1 score.

To get an idea of how much the topic masks are effective for improving segmentation performances, in **Fig. 9**, we plot the performance improvements of the masks with their best parameter settings. In the figure, LDA+F shows better performance than the LDA model alone when $S = 3$, but its performance is significantly decreased as the number of iterations I is increased. The LDA+C has generally better performance than both the LDA model and LDA+F when $S = 9$ and $P = 0.3$. In the case of SLDA model, SLDA+F also outperforms SLDA+C and it shows relative stable performance improvements, while the performance of SLDA+F slowly decreases.



(a)

(b)

Fig. 9. The C-mask result on the LDA model (a) and the result on the SLDA model (b) with the best parameter settings when $T = 5$, where the horizontal axis represents I and the vertical axis indicates the F1 score. LDA+F indicates the LDA model applied with the F-mask, and LDA+C indicates the LDA model applied with the C-mask.

We investigated the impact of the factors so far, but it is still necessary to check whether the mask results are comprehensible or reasonable to people. In other words, the performance improvements obtained from topic masks do not guarantee that the mask results become better for people. Samples of mask results are depicted in **Fig. 10**, where each color corresponds to a homogeneous region. If we see the segmentation results of a sheep, a car, and a cow, then it is obvious that the results of the LDA model alone in the third row are poorer than the results of the SLDA model alone in the seventh row. In contrast, when we see the results of a bird, a bicycle, and an air-plane, then the results of the SLDA model alone are poorer than the results of the LDA model alone. The reason is that the SLDA model incorporates spatial information based on the hypothesis that spatially close patches are more likely to have the same topic assignments. This implies that the SLDA model may not be good for capturing thin or fine objects such as bicycles or birds. Although this difference exists between the two models, the SLDA showed a generally better segmentation performance because the SLDA has the benefit of thick parts, which have obviously more topic assignments than thin parts.

We found two characteristics of the mask results in **Fig. 10**. First, the mask results are heavily influenced by the topic assignments obtained from the topic models. For example, the LDA topic assignments of sheep in the first column are poor, so all the mask results including LDA+F and LDA+C are poor. The SLDA topic assignments of sheep, in contrast, are better than the LDA topic assignments, so the corresponding mask results are also relatively good. The results of car or cow are similar to the case of sheep. When we see the LDA topic assignments of bicycles in the fifth column, it is considerably better than the topic assignments of the SLDA on bicycles. The mask results of the LDA model upon bicycles, therefore, are also better than the mask results of the SLDA model. The results of bird or air-plane are similar to the case of bicycles. From these observations, we conclude that the mask results strongly depend on the topic assignments obtained from the topic models. Second, we found that a higher F1 score does not mean better comprehensible mask results. In **Fig. 10**, the best F1 score LDA+C represents the parameter setting of LDA+C resulting in the best F1 score. The best F1 score LDA+C shows generally better mask results than LDA+F, but it has some results that are poorly comprehensible to people. For example, the mask results of the best F1 score LDA+C on bicycles in the fifth column is obviously not comprehensible to people. To get more comprehensible mask results, we investigated all possible mask results of parameter settings, and we got the best comprehensible LDA+C result of each image as depicted in the sixth row in **Fig. 10**, where the best comprehensible LDA+C represents the parameter settings of LDA+C that generate mask results most comprehensible to people. With a different parameter setting (e.g., S = 15, P = 0.1) upon the image of bicycles as shown in the fifth column, LDA+C has results more comprehensible to people. For some images, the parameter setting of the best F1 score is the same as the best comprehensible parameter setting. For example, the best F1 score LDA+C result upon bird, as shown in the fourth column, is also the most comprehensible to people. For the SLDA model, we also observe similar results. That is, the parameter settings of the best F1 score SLDA are not always the same as the parameter settings of the best comprehensible SLDA. Therefore, we need to find the parameter setting that makes the mask results comprehensible to people. As the parameter settings for the best comprehensible results look different for each image, we will investigate patterns of parameter settings for comprehensibility as a future study.

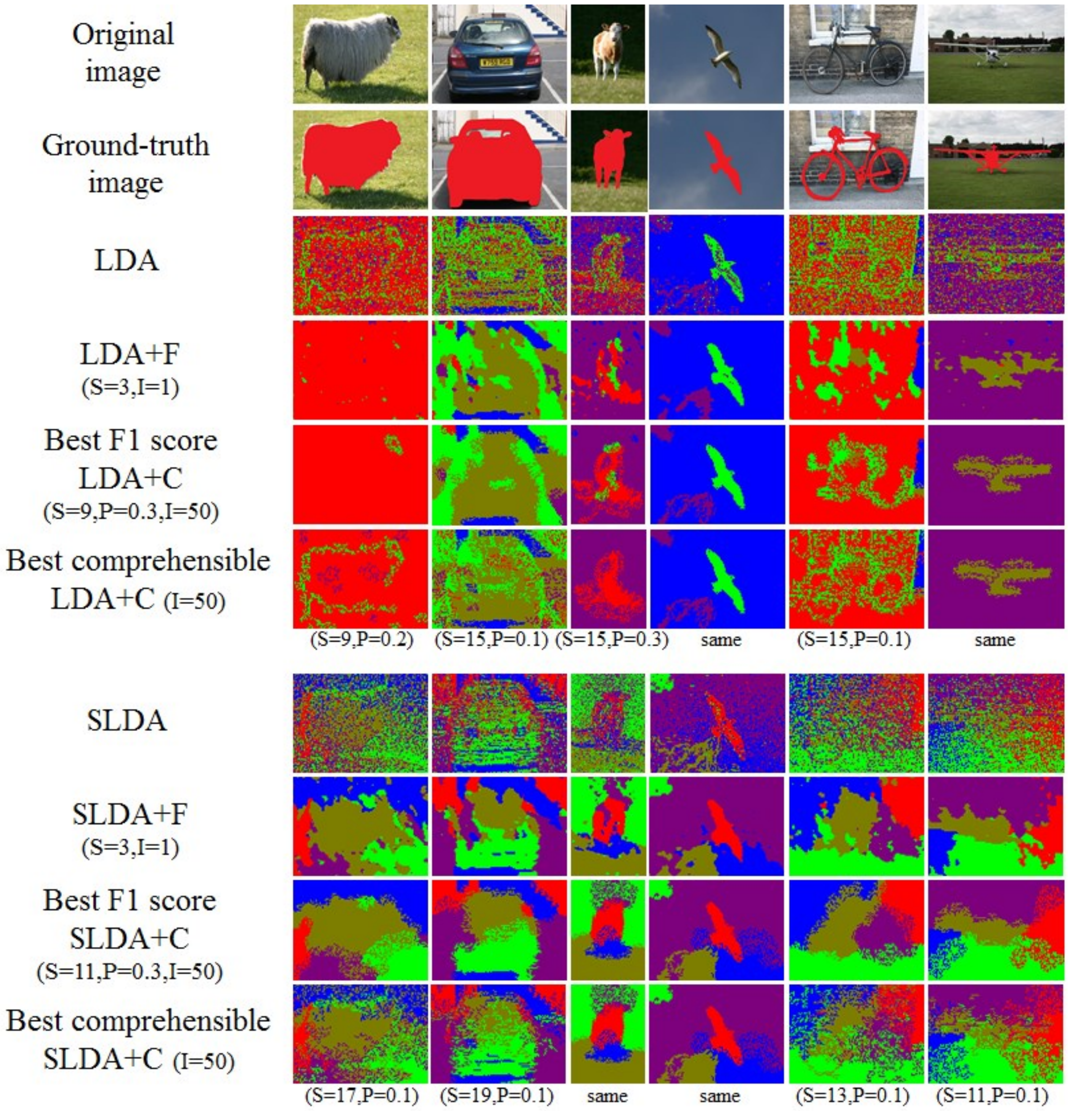


Fig. 10. The samples of mask results, where each color corresponds to a homogeneous region. The best F1 score LDA+C means the parameter setting of LDA+C that results in the best F1 score, and the best comprehensible LDA+C means the parameter settings of LDA+C that generate results most comprehensible to people. As the best comprehensible settings of different images can be different from each other, we labeled the parameter settings below each of the best comprehensible results. If the best comprehensible setting is the same as the best F1 score setting, then we labeled it as 'same'.

5. Conclusion

In this paper, we proposed two topic masks for reducing topic noises, namely the Frequency mask (F-mask) and the Connection mask (C-mask). The F-mask is based on the hypothesis that a more frequent topic assignment is more likely to be assigned to the center patch within the corresponding mask. The C-mask incorporates positions of topic assignments based on the hypothesis that the center topic assignment should be kept if it is a part of a mass; otherwise it is replaced by a new topic obtainable from the F-mask. The topic masks do not consider the semantic relationships among the topics, and just filter the noises out using structural patterns of the assigned labels. Thus, it may not guarantee that the results of the masks be more comprehensible by human, even if the segmentation performance is improved. By empirical results, we investigated five factors affecting the performances of topic masks: (1) topic models, (2) the number of topics *T*, (3) the number of mask iterations *I*, (4) the size of masks *S*, and (5) the parameter *P*. We applied the F-mask and the C-mask upon topic assignments obtained from either the LDA model or the SLDA model, and the segmentation performances (e.g., F1 score) were significantly increased using the C-mask, while the F-mask showed unstable performance improvements. As we observed that the parameter settings for the best comprehensible results look different from image to image, the proposed topic masks would be useful only when the parameters *T*, *I*, *S*, and *P* are adjusted according to the object type, image size, and the topic models. Thus, to make the masks more practically useful, it will be necessary to find a way of setting the parameters automatically. We plan to discover such a way as the future work using more state-of-the-art topic models and datasets.

References

[1] Thomas Hofmann, "Probabilistic latent semantic analysis," in *Proc. of 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 289-296, July 30-August 1, 1999. http://dl.acm.org/citation.cfm?id=2073829

[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, March 2003. http://dl.acm.org/citation.cfm?id=944937

[3] David M. Blei and John D. Lafferty, "Dynamic topic models," in *Proc. of 23rd International Conference on Machine Learning (ICML)*, pp. 113-120, June 25-29, 2006. Article (CrossRef Link)

[4] Young-Seob Jeong and Ho-Jin Choi, "Sequential entity group topic model for getting topic flows of entity groups within one document," in *Proc. of 16th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*, pp. 366-378, May 29-June 1, 2012. Article (CrossRef Link)

[5] Lan Du, Wray L. Buntine, and Huidong Jin, "Sequential latent dirichlet allocation: discover underlying topic structures within a document," in *Proc. of 10th IEEE International Conference on Data Mining (ICDM)*, pp. 148-157, December 14-17, 2010. Article (CrossRef Link)

[6] David Newman, Chaitanya Chemudugunta, and Padhraic Smyth, "Statistical entity-topic models," in *Proc. of 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 680-686, August 20-23, 2006. Article (CrossRef Link)

[7] Jonathan Chang, Jordan Boyd-Graber, and David M. Blei, "Connections between the lines: augmenting social networks with text," in *Proc. of 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 169-178, June 28 - July 1, 2009. Article (CrossRef Link)

[8] Bin Zhao, Li Fei-Fei, and Eric P. Xing, "Image segmentation with topic random field," in *Proc. of 11th European Conference on Computer Vision (ECCV)*, pp. 785-798, September 6-9, 2010. Article (CrossRef Link)

[9] Li-Jia Li, Richard Socher, and Li Fei-Fei, "Towards total scene understanding: classification, annotation and segmentation in an automatic framework," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2036-2043, June 20-25, 2009. Article (CrossRef Link)

[10] Liangliang Cao and Li Fei-Fei, "Spatially coherent latent topic model for concurrent object segmentation and classification," in *Proc. of 11th IEEE International Conference on Computer Vision (ICCV)*, October 14-20, 2007. Article (CrossRef Link)

[11] Xiaogang Wang and Eric Grimson, “Spatial latent dirichlet allocation,” in *Proc. of 21th Annual Conference on Neural Information Processing Systems (NIPS)*, December 3-6, 2007. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.143.4231

[12] Yoseph Linde, Andres Buzo, and Robert M. Gray, “An algorithm for vector quantizer design,” *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84-95, 1980. Article (CrossRef Link)

[13] Jianbo Shi and Jitendra Malik, “Normalized cuts and image segmentation,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888-905, 2000. Article (CrossRef Link)

[14] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, “Efficient graph-based image segmentation,” *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004. Article (CrossRef Link)

[15] Narjes Doggaz and Imene Ferjani. “Image segmentation using normalized cuts and efficient graph-based segmentation,” in *Proc. of 16th International Conference on Image Analysis and Processing (ICIAP)*, pp. 229-240, September 14-16, 2011. Article (CrossRef Link)

[16] Zhengxing Niu, Gang Hua, Xinbo Gao, and Qi Tian, “Spatial-DiscLDA for visual recognition,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1769-1776, June 21-23, 2011. Article (CrossRef Link)

[17] Timothy J. Burns and Jason J. Corso, “Robust unsupervised segmentation of degraded document images with topic models,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1287-1294, June 20-25, 2009. Article (CrossRef Link)

[18] John M. Winn, Antonio Criminisi, and Thomas P. Minka. “Object categorization by learned universal visual dictionary,” in *Proc. of 10th IEEE International Conference on Computer Vision (ICCV)*, pp. 1800-1807, October 17-20, 2005. Article (CrossRef Link)

Young-Seob Jeong is currently a PhD student in the Dept. of Computer Science at KAIST. His current research interests include topic modeling, deep learning, and action prediction based on various sensor data.

Chae-Gyun Lim is currently a master candidate in the Dept. of Computer Engineering at Kyung Hee University, Korea. In 2011, he received a BS in Medical Computer Science from Eulji University, Korea. Between 2011 and 2013, he worked as a research assistant in the Dept. of Computer Science at KAIST, Korea. His research interests include data mining, topic modeling, big data analysis and bioinformatics.

Byeong-Soo Jeong is currently a professor in the Dept. of Computer Engineering at Kyung Hee University, Korea. In 1983, he received a BS degree in Computer Engineering from Seoul National University, Korea, in 1985, a MS in Computer Science from the Korea Advanced Institute of Science and Technology, Korea, and in 1995, a PhD in Computer Science from the Georgia Institute of Technology, Atlanta. From 1985 to 1989, he was on a research staff at Data Communications Corporation, Korea. From 1996, he has been with the Dept. of Computer Engineering at Kyung Hee University, Korea. From 2003 to 2004, he was a visiting scholar at the Georgia Institute of Technology. His research interests include database systems, data mining, and bioinformatics.

Ho-Jin Choi is currently an associate professor in the Dept. of Computer Science at KAIST. In 1982, he received a BS in Computer Engineering from Seoul National University, Korea, in 1985, an MSc in Computing Software and Systems Design from Newcastle University, UK, and in 1995, a PhD in Artificial Intelligence from Imperial College, London, UK. From 1982 to 1989, he worked for DACOM, Korea, and between 1995 and 1996, worked as a post-doctoral researcher at Imperial College. From 1997 to 2002, he served as a faculty member at Korea Aerospace University, Korea, then from 2002 to 2009 at Information and Communications University (ICU), Korea, and since 2009 he has been with the Dept. of Computer Science at KAIST. Between 2002 and 2003, he visited Carnegie Mellon University (CMU), Pittsburgh, USA, and has been serving as an adjunct professor of CMU for the program of Master of Software Engineering (MSE). Between 2006 and 2008, he served as the Director of Institute for IT Gifted Youth at ICU. Since 2010, he has been participating in the Systems Biomedical Informatics National Core Research Center at the Medical School of Seoul National University. Currently, he serves as a member of the boards of directors for the Software Engineering Society of Korea, for the Computational Intelligence Society of Korea, and for Korean Society of Medical Informatics. His current research interests include artificial intelligence, data mining, software engineering, and biomedical informatics.