Quadratic divergence regularized SVM for optic disc segmentation

JUN CHENG,1,* DACHENG TAO,2 DAMON WING KEE WONG,1 AND JIANG LIU3

1Institute for Infocomm Research, A*STAR, Singapore
2The University of Sydney, Australia
3Cixi Institute of Biomedical Engineering, Chinese Academic of Sciences, China
*jcheng@i2r.a-star.edu.sg

Abstract: Machine learning has been used in many retinal image processing applications such as optic disc segmentation. It assumes that the training and testing data sets have the same feature distribution. However, retinal images are often collected under different conditions and may have different feature distributions. Therefore, the models trained from one data set may not work well for another data set. However, it is often too expensive and time consuming to label the needed training data and rebuild the models for all different data sets. In this paper, we propose a novel quadratic divergence regularized support vector machine (QDSVM) to transfer the knowledge from domains with sufficient training data to domains with limited or even no training data. The proposed method simultaneously minimizes the distribution difference between the source domain and target domain while training the classifier. Experimental results show that the proposed transfer learning based method reduces the classification error in superpixel level from 14.2% without transfer learning to 2.4% with transfer learning. The proposed method is effective to transfer the label knowledge from source to target domain, which enables it to be used for optic disc segmentation in data sets with different feature distributions.

© 2017 Optical Society of America

OCIS codes: (100.2960) Image analysis; (100.3008) Image recognition, algorithms and filters; (100.0100) Image processing.

References and links
1. Y. Zheng, M. H. Hijazi, and F. Coenen, “Automated disease/no disease grading of age-related macular degeneration by an image mining approach,” Investigative Ophthalmology & Visual Science 53(13), 8310–8318 (2012).
2. X. Zhu and R. M. Rangayyan, “Detection of the optic disc in images of the retina using the hough transform,” Int. Conf. of IEEE Eng. in Med. and Bio. Soc. pp. 3546–3549 (2008).
3. A. Aquino, M. Gugnieder-Arias, and D. Martin, “Detecting the optic disc in images of the retina using morphological edge detection and feature extraction techniques,” IEEE Trans. Med. Imag. 29, 1860–1869 (2010).
4. J. Cheng, J. Liu, D. W. K. Wong, F. Yin, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong, “Automatic optic disc segmentation with peripapillary atrophy elimination,” Int. Conf. of IEEE Eng. in Med. and Bio. Soc. pp. 6624–6627 (2011).
5. A. Giachetti, L. Ballerini, and E. Trucco, “Accurate and reliable segmentation of the optic disc in digital fundus images,” Journal of Medical Imaging 1, 024001 (2014).
6. B. Dashkozorg, A. M. Mendonça, and A. Campilho, “Optic disc segmentation using the sliding band filter,” Computers in Biology and Medicine 56, 1–12 (2015).
7. A. Singh, M. K. Dutta, V. U. M. Parthasarathi, and R. Burget, “Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image,” Computer Methods & Programs in Biomedicine 124, 108–120 (2016).
8. J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, “Optic nerve head segmentation,” IEEE Trans. Med. Imag. 23, 256–264 (2004).
9. J. Xu, O. Chutataporn, E. Sung, C. Zheng, and P. Kuan, “Optic disk feature extraction via modified deformable model technique for glaucoma analysis,” Pattern Recognition 40, 2063–2076 (2007).
10. Z. Zhang, J. Liu, N. S. Cherian, Y. Sun, J. H. Lim, D. W. K. Wong, N. M. Tan, S. Lu, H. Li, and T. Y. Wong, “Convex hull based neuro-retinal optic cup ellipse optimization in glaucoma diagnosis,” Int. Conf. of IEEE Eng. in Med. and Bio. Soc. pp. 1441–1444 (2009).
11. G. D. Joshi, S. Sivaswamy, and S. R. Krishnadas, “Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment,” IEEE Trans. Med. Imag. 30, 1192–1205 (2011).
The optic disc (OD) is the location where ganglion cell axons exit the eye to form the optic nerve. The segmentation of OD has been identified as an important step in the computer aided diagnosis/detection/screening of diseases such as glaucoma, diabetic retinopathy, age-related macular degeneration [1], etc. Figure 1 shows the boundary of OD in a retinal color fundus image. In recent years, OD segmentation has received much attention and many algorithms have been proposed. Existing methods for OD segmentation are mainly based on template matching [2–7], deformable models [8–13] or learning [14–17].

Template based approaches often model the disc as a circle [3] or an ellipse [4] as the discs are approximately a circle or an ellipse. In these approaches, the disc is often detected by Hough transform from the edges in the image, while different authors propose different methods to detect the edges from the discs. Some authors eliminate edges from vessels by first segmenting the vessels [5]. Deformable model based approaches usually start with an initial disc contour and deform toward the disc edges based on various energy terms such as boundary energy, shape energy, region energy [13]. These energy terms are often defined using image intensity, image gradient, and/or boundary smoothness, etc. Learning approaches often learn/train classifiers from existing labelled data to classify pixels as disc or non-disc pixels. Abràmoff et al. [14] propose the pixel classification approach and explore various features including intensity, gradient, Gaussian blurred feature, prior probability, for disc and cup segmentation. The limitation of the pixel classification is that the high number of pixels makes the optimization difficult.

1. **Introduction**

The optic disc (OD) is the location where ganglion cell axons exit the eye to form the optic nerve. The segmentation of OD has been identified as an important step in the computer aided diagnosis/detection/screening of diseases such as glaucoma, diabetic retinopathy, age-related macular degeneration [1], etc. Figure 1 shows the boundary of OD in a retinal color fundus image. In recent years, OD segmentation has received much attention and many algorithms have been proposed. Existing methods for OD segmentation are mainly based on template matching [2–7], deformable models [8–13] or learning [14–17].

Template based approaches often model the disc as a circle [3] or an ellipse [4] as the discs are approximately a circle or an ellipse. In these approaches, the disc is often detected by Hough transform from the edges in the image, while different authors propose different methods to detect the edges from the discs. Some authors eliminate edges from vessels by first segmenting the vessels [5]. Deformable model based approaches usually start with an initial disc contour and deform toward the disc edges based on various energy terms such as boundary energy, shape energy, region energy [13]. These energy terms are often defined using image intensity, image gradient, and/or boundary smoothness, etc. Learning approaches often learn/train classifiers from existing labelled data to classify pixels as disc or non-disc pixels. Abràmoff et al. [14] propose the pixel classification approach and explore various features including intensity, gradient, Gaussian blurred feature, prior probability, for disc and cup segmentation. The limitation of the pixel classification is that the high number of pixels makes the optimization difficult.
Therefore, superpixel classification [16] is proposed. By first over-segmenting the images into superpixels, the approach significantly reduce the amount of data in both training and testing. Recently, ensemble sampling based convolutional neural network is also used to learn the model for disc segmentation [17], where a entropy based sampling technique is proposed to identify most informative samples for training.

This paper focuses on learning based methods. The learning based methods generally work under a common assumption that the training and testing data are drawn from the same feature space with same distributions. When the feature distribution changes, the models need to be rebuilt from scratch using newly collected training data. Retinal color fundus images are often collected under different conditions, e.g., camera models, image resolutions, races, dilated or undilated eyes, etc. Therefore, the feature distributions from different data sets are often different and the models trained from one data set is not optimized for another data set. However, it is often too expensive and time consuming to re-collect or label the needed training data and rebuild the models for all different data sets. Meanwhile, classifiers trained from only a few labelled data are usually not robust. It would be nice to reduce the need and effort to re-collect the training data. In such cases, knowledge transfer or transfer learning would be desirable. Transfer learning is a popular learning scheme which transfers the knowledge from domains with sufficient training data to domains with few or no training data. It learns more robust classifiers by leveraging a relatively large amount of labelled training data from other domains, referred to as source domains. Some efforts have been made in transfer learning for many applications including text classification, video concept detection, and face recognition. Pan et al. [18] propose transfer learning via dimensionality reduction to minimize maximum mean discrepancy (MMD). Duan et al. use a similar concept in domain transfer support vector machine (SVM) [19]. Si et al. propose a framework for transfer subspace learning via dimensionality reduction [20]. In [21], Cheng et al. use domain transfer SVM in MCI conversion prediction.

In this paper, we propose a new transfer learning method, referred to as quadratic divergence regularized SVM (QDSVM), for transfer learning in OD segmentation. As the objective of
transfer learning is for classification using SVM, the discriminative information should be maximally kept and transferred when train the SVM classifier. The proposed QDSVM integrates the objective function of minimizing the distribution difference between source and target domains with the objective function of training the SVM classifier. The main contribution of this paper includes: it proposes QDSVM which minimizes the distribution difference measured by quadratic divergence. The proposed QDSVM integrates the objective function of the SVM with the objective of minimizing the distribution difference while learning a mapping from high dimensional data to low dimensional data for OD segmentation. The QDSVM is applied to OD segmentation and improves the segmentation results. To the best of our knowledge, the proposed method is the first to apply transfer learning for OD segmentation in the testing data which has a different feature distribution.

The rest of paper is organized as follows. Section 2 reviews the existing learning based disc segmentation algorithm. Section 3 introduces the proposed QDSVM method and its application for OD segmentation. Section 4 shows the experimental results followed by the discussion and conclusions in the last section.

2. Learning based disc segmentation

In this section, we first give a short review of learning based disc segmentation methods. Machine learning has been used for optic disc segmentation. In [14], pixel classification is first proposed. It is used to assign a class to the pixels in an image, in this case disc or background. Pixel classification uses multiple pixel features, which are numeric properties of a pixel and its surroundings. One advantage of the pixel classification method is that it avoids the potential bias of the algorithm developer, which is common in unsupervised approaches, such as in an edge-based, deformable model or in template-based algorithms [14]. Previously, we have also proposed a learning based method to segment the disc [16]. Different from [14], our approach utilizes the learning at superpixel level instead of pixel level. Fig. 2 shows the flow chart of the method. In the method, the input image is first over-segmented into superpixels using the simple linear iterative clustering (SLIC) [22] method. Then, various features such as histograms and center surround statistics [16] are computed from each superpixel. A SVM classifier is then trained to obtain an initial estimation of the disc boundary. In our implementation, the output value from the SVM decision function for each superpixel is used as the decision values for all pixels in the superpixel. Then, a smoothing filter is applied to achieve smooth decision values, which are then compared with a threshold to obtain the binary decisions for all pixels. A morphological operation is applied to obtain the largest connected component whose boundary is used as the raw estimation of the disc boundary. Finally, active shape model (ASM) is applied to obtain the final disc boundary. Here we give a brief introduction of the ASM algorithm while more details can be found in [12].

In ASM, we represent a 2D disc shape by a vector \( p = (m_1, n_1, \cdots, m_Z, n_Z) \) with \( Z \) points, where \( (m_i, n_i) \) denotes the coordinates of the \( i^{th} \) point. To build the model, a data set with manual disc points is used and the disc shapes are aligned based on the centers. The mean and the covariance of the shapes are computed as:

\[
\bar{p} = \frac{1}{N} \sum_{k=1}^{N} p_k, \tag{1}
\]

\[
S = \frac{1}{N-1} \sum_{k=1}^{N} (p_k - \bar{p})(p_k - \bar{p})^T, \tag{2}
\]

where \( N \) is the number of training shapes. By applying a principle component analysis, the first
$K$ components $\phi_1, \cdots, \phi_K$ are obtained. The shape is then approximated as

$$p = \tilde{p} + \phi b,$$

where $b = \phi^T (p - \tilde{p})$ denotes the vector of $K$ elements containing the model parameters and $\phi = (\phi_1, \cdots, \phi_K)$ stores the $K$ eigenvectors of the principle component analysis. In our algorithms, we use $Z = 24$ landmarks evenly sampled along the disc boundary and $K$ is set to keep 99% of variance in principle component analysis.

The model is built using the normalized first derivatives of pixel profiles from points around each landmark point, in the direction perpendicular to the line that connects the neighboring landmark points. Denoting the normalized derivative profiles as $g_1, g_2, \cdots$, the mean profile $\tilde{g}$ and the covariance matrix $S_\tilde{g}$ are computed for each landmark. The new landmark position is obtained by minimizing the following Mahalanobis distance:

$$f(g_k) = (g_k - \tilde{g}) S_{\tilde{g}}^{-1} (g_k - \tilde{g})$$

In our implementation, we first get the initial disc boundary based on the output of the SVM classification. Then, we choose the 24 landmark points around the initial disc boundary with each pair of adjacent points forming an angle of 15 degrees with the disc center. Then the landmark points are iteratively updated by minimizing the Mahalanobis distance in the direction perpendicular to the line that connects the neighboring landmark points until converged or the maximum number of iteration reached.

### 3. Quadratic divergence regularized SVM

For classification task, given $n_s$ samples $X^s$ from source domain and $n_t$ samples $X^t$ from target domain, which are all drawn from a high-dimensional space $\mathbb{R}^D$, a subspace learning algorithm finds a low-dimensional subspace $\mathbb{R}^d$, wherein samples from different classes can be well separated. Let us denote the data set of labelled and unlabelled samples from the target domain as $X^t_1$ and $X^t_2$, respectively and $X^t = X^t_1 \cup X^t_2$. Let us also represent the labelled training data set as $X$. We have $X = X^s \cup X^t_1$. The proposed QDSVM computes a mapping $P$ from $X$ and $X^t_1$ in high dimensional space $\mathbb{R}^D$ to $Y$ and $Y^t_1$ in low dimensional space $\mathbb{R}^d$, where $P \in \mathbb{R}^{D \times d}$ and $d < D$, i.e., $Y = XP$ and $Y^t_1 = X^t_1 P$. The mapping $P$ is computed such that it minimizes both the SVM objective function and the distribution difference measured by quadratic divergence.

#### 3.1. Support vector machine

In SVM, the focus is to find the maximal separating hyperplane between classes. Given a set of $l$ training instances $Y_i = X_i P, i = 1, \cdots, l$ with labels $z_i \in \{+1, -1\}$, where $X_i$ is the original feature vector, $Y_i$ the corresponding feature vector in the projected space $P$, the reformulated primal problem of classical SVM [23] is given by

$$\min_{w} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} h(z_i Y_i, w) \right\},$$

where $h(z_i Y_i, w) = \max[0, 1 - z_i Y_i w]$ denotes the hinge loss and $w$ the classification hyperplane.

#### 3.2. Quadratic divergence-based regularization

There are many criteria to estimate the distance between different distributions. A well-known example is quadratic divergence. Denoting the probability distributions of the labelled training data and unlabelled testing data as $Q_l$ and $Q_u$ respectively, quadratic divergence is a non-symmetric measure of the difference between probability distributions $Q_l$ and $Q_u$. Specifically,
the divergence of $Q_u$ from $Q_l$, denoted $D(Q_l||Q_u)$, is a measure of the information lost when $Q_u$ is used to approximate $Q_l$.

In this paper, we present a quadratic divergence-based regularization $D_P(Q_l||Q_u)$, which measures the distribution difference of samples drawn from different domains in a projected subspace $P$. By integrating this regularization into (5), we obtain the objective function of QDSVM:

$$\min_{P,w} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} h(z_i, Y_i, w) + \lambda D_P(Q_l||Q_u) \right\},$$

(6)

where $\lambda$ is a parameter controls the weight of the quadratic divergence regularization.

The probability density for the training and testing samples in the projected subspace $P$ is $q_l$ and $q_u$, where $y = P^T x$ is the low-dimensional representation of the sample $x$. The difference is given by

$$D_P(Q_l||Q_u) = \int (q_l(y) - q_u(y))^2 dy$$

$$= \int (q_l(y))^2 - 2q_l(y)q_u(y) + q_u(y)^2 dy$$

(7)

To estimate the distribution $Q_l$ and $Q_u$ in the projected space $P$, we apply the kernel density estimation (KDE) technique, which estimates the density at $y$ as a sum of kernels between $y$ and each of the other sample, i.e., $q(y) = (1/n) \sum_{i=1}^{n} G(y - y_i)$. Here, $n$ is the number of samples and $G(y)$ is the $d$-dimensional Gaussian kernel with covariance matrix $\Sigma$. By introducing the estimated distributions based on KDE, we have

$$D_P(Q_l||Q_u) = \int \left( \frac{1}{l} \sum_{i=1}^{l} G_{\Sigma_1}(y - y_i) \right)^2 dy$$

$$+ \int \left( \frac{1}{l} \sum_{j=1}^{l+u} G_{\Sigma_2}(y - y_j) \right)^2 dy$$

$$- \int \left( \frac{2}{lu} \sum_{i=1}^{l} \sum_{j=l+1}^{l+u} G_{\Sigma_1}(y - y_i)G_{\Sigma_2}(y - y_j) \right) dy$$

(8)

Since $\int G_{\Sigma_1}(y - y_i)G_{\Sigma_2}(y - y_j) dy = G_{\Sigma_1 + \Sigma_2}(y_i - y_j)$, we can eliminate the integrations in (8). The discrete form of the quadratic divergence is given by:

$$D_P(Q_l||Q_u) = \frac{1}{l^2} \sum_{x=1}^{l} \sum_{t=1}^{l} G_{\Sigma_{11}}(y_t - y_x)$$

$$+ \frac{1}{lu} \sum_{x=l+1}^{l+u} \sum_{t=l+1}^{l+u} G_{\Sigma_{22}}(y_t - y_x)$$

$$- \frac{2}{lu} \sum_{x=1}^{l} \sum_{t=l+1}^{l+u} G_{\Sigma_{12}}(y_t - y_x)$$

(9)

(10)

where $\Sigma_{11} = \Sigma_1 + \Sigma_1$, $\Sigma_{12} = \Sigma_1 + \Sigma_2$, $\Sigma_{22} = \Sigma_2 + \Sigma_2$.

3.3. Solution

As both $P$ and $w$ are unknown in (6), we solve them iteratively. In the solution, we solve $P$ first and then solve $w$. The iteration continues until convergence. Denoting the objective function in
(6) as $F(P, w)$, we need to smooth $F(P, w)$ as the hinge loss function $h(z_iY_i, w)$ in (6) is not differentiable. It can be smoothed by subtracting a prox-function $d_i(u) = \sum_i^n \|X_i\|\|\sum_i^n \|X_i\|\|u_i^2$ [23]. The smoothed hinge loss $h_u$ is given by:

$$h_u = \max_{u_i} (1 - z_iX_iPw) - \frac{\mu}{2}\|X_i\|\|u_i^2$$

where $\mu$ is the smooth parameter, $u_i$ can be obtained by setting the gradient of the objective function as zero, i.e.,

$$u_i = \text{median}\left\{\frac{1 - z_iX_iPw}{\mu\|X_i\|\|}, 0, 1\right\}$$

We have the smoothed objective function $F_u(P, w)$ for QD SVM:

$$F_u(P, w) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n h_u + \lambda DP(Q_t||Q_u)$$

We apply Nesterov’s method [24] to minimize the smoothed $F_u(P, w)$ iteratively. It is a gradient method with optimal convergence rate $\varphi(1/k^2)$. In the $k^{th}$ iteration, two auxiliary optimizations $P_1^k$ and $P_2^k$ are constructed:

$$P_1^k = P^k - \frac{1}{L_\mu} \nabla F_\mu(P^k, w)$$

where $L_\mu$ is the Lipschitz constant computed as $\frac{n}{\mu} \max_i \|X_i\|\|\|\|X_i\|\|$.

$$P_2^k = P^k - \frac{\sigma^2}{L_\mu} \sum_{i=0}^k \frac{i+1}{2} \nabla F_\mu(P^i, w)$$

The weighted sum of $P_1^k$ and $P_2^k$ determines the solution of $P$ in the next iteration

$$P^{k+1} = \frac{k+1}{k+3} P_1^k + \frac{2}{k+3} P_2^k$$

Similarly, the SVM solution $w$ can be solved by the above Nesterov’s method as well.

4. Experimental results

We use 325 images with dimension $3072 \times 2048$ from Singapore Malay Eye Study (SiMES) [16] as source domain in our experiments. The disc boundaries for these images have been manually marked by trained professionals. The SiMES images were acquired using Canon CR-DGi digital retinal camera. The target domain contains 1676 images from Singapore Chinese Eye Study (SCES) with image dimension $3504 \times 2336$ and $3888 \times 2592$. The SCES images were acquired using Canon CR-1 Mark -II. As we can see, the two data sets are from different races captured by different cameras. In our experiments, the SLIC [22] algorithm is first applied to generate superpixels. Based on the manual disc boundary, the superpixels with more than 50% of pixels within the disc boundary are labelled as disc superpixels and the rest as non-disc superpixels.

To evaluate the performance of the method, the accuracy in superpixel classification is computed. In addition, the overlapping error $E$ is also computed to examine the difference between automated and manual discs:

$$E = 1 - \frac{\text{Area}(S \cap M)}{\text{Area}(S \cup M)}$$

where $S$ and $M$ denote the segmented and the manual disc respectively. In order to evaluate how transfer learning improve the performance, we conducted the tests in two scenarios depending on the availability of labelled data from the target domain for training.
Fig. 3. Sample Results: (a) original; (b) the manual disc; (c) initial disc by SVM; (d) initial disc by QDSVM; (e) after ASM from (c); (f) after ASM from (d).

4.1. No labelled data from target domain is available for training

Table 1. Classification error $P_s$ and mean overlapping error $\bar{E}$ by different methods

| Method            | ASM [12] | No transfer [16] | TFLDA [20] | DTSVM [19] | Proposed |
|-------------------|----------|------------------|------------|------------|----------|
| $P_s$             | -        | 14.2%            | 7.3%       | 4.3%       | 2.4%     |
| $E$ before ASM    | -        | 20.2%            | 18.2%      | 17.0%      | 15.7%    |
| $E$ after ASM     | 14.8%    | 12.0%            | 11.7%      | 11.6%      | 11.0%    |

In the first scenario, there is no labelled data from the target domain, i.e., $X'_t = \emptyset$. This is a very common scenario in clinical applications as it is often difficult and troublesome to manually mark the boundaries of the optic disc for a new data set. The training data $X = X_s$ includes all 325 images in SiMES data set and $X'_t = X'_s$ includes all 1676 images in SCES data set. In our paper, we use the proposed QDSVM to compute a mapping $P$ and the SVM classifier simultaneously. In the testing, the obtained $P$ and classifier are used to determine each superpixel as disc or non-disc. To illustrate the benefit of transfer learning, we compare the results with transfer learning and the results without transfer learning. In addition, we have also applied the transferred Fisher’s linear discriminant analysis (TFLDA) [20] and the domain transfer SVM (DTSVM) [19] to obtain different classifiers for comparison. These classifiers are also used to obtain disc segmentation results for comparison following the same flow in Fig. 2. Besides that, we have also include the simple ASM method [12] which applies a Hough transform followed by an ASM deformation for comparison. Table 1 shows the overlapping errors between the manual and automated discs. It shows that the proposed QDSVM reduces the error by 8.3% relatively from 12.0% to 11.0% compared with case without transfer learning. Since the transfer learning only improves in the step of classification in superpixel level, it is also important to compare the results before ASM deformation. Table 1 also compares the percentage of wrongly classified superpixels $P_s$ by different methods. It shows that QDSVM reduces the error by 83.1%, 67.1% and 44.2% relatively compared with cases without transfer learning.
learning, TFLDA and DTSVM, respectively. Fig. 3 shows three sample results. In the figure, we show both the disc contours before and after applying ASM. As we can see, QDSVM achieves better initial disc contours for ASM and often a more accurate final disc boundary after ASM. It is noted that error reduction before ASM is much larger than that after ASM. This is reasonable as the ASM may deform to the same contour from two different initial contours in many cases and neutralize the improvement by QDSVM. Nevertheless, a student t-test shows \( p < 0.01 \), which indicates that the error reduction after ASM is still significant.

### 4.2. Some labelled data from target domain is available for training

In the second scenario, the doctors are able to label a few data, i.e., \( X^T \neq \emptyset \) and the training data \( X = X^S \cup X^T \). In our experiments, we evaluate the performance of the proposed method when different numbers of labelled images are used. For convenience, we divide the SCES data set into two subsets \( A \) and \( B \). The subset \( A \) consists of 1351 images randomly selected from the SCES data set. We use this fixed subset for testing. The second subset \( B \) consists of the rest of 325 SCES images. We randomly selected \( n \) images from subset \( B \) to form \( X^T \). To evaluate how the number \( n \) affects the results, we adjust \( n \) from 25 to 175 with a step of 25 and 175 to 325 with a step of 50. We consider the following scenarios to justify the effectiveness of transfer learning: 1) we train the standard SVM using labelled data from target domain only, i.e., training data \( X = X^T \); 2) we train the standard SVM classifier using both labelled data from source and target domain: training data \( X = X^T \cup X^S \); 3) we apply QDSVM transfer learning and \( X = X^T \cup X^S \). In addition, TFLDA and DTSVM are also tested for comparison. All tests are done on subset \( A \). Fig. 4 shows the overlapping error as \( n \) changes. The results show that the QDSVM achieves the lowest overlapping error. The improvement is large when \( n \) is small and drops as \( n \) increases. This is reasonable as the increased \( n \) helps to build a more robust classifier in the target domain for other algorithms.

### 4.3. Performance on MESSIDOR

Besides the local data set, we have also conducted experiments on public data set such as MESSIDOR [25,26]. The data set has been widely used since it was published online [27]. Since our objective is to transfer the knowledge from another domain, we assume the label information in MESSIDOR data set is unknown in the training stage. In the experiments, we use the proposed QDSVM to compute a mapping \( P \) and the SVM classifier simultaneously as that in section 4.1. We achieve mean overlapping error of 11.1%. To justify the benefit of transfer learning, we use the superpixel classification based method [16] without transfer learning. We use local SiMES data set to train the SVM classifier and test in MESSIDOR. We achieve an average overlapping error of 12.5%. In addition, we also include the results from [5,6,17] for comparison. Table 2
summarizes the results. Our results are better than 12.0% by Giachetti’s method [5] and comparable to 11.0% by Dashtbozorg’s method [6]. Zilly’s method in [17] achieves the best result of 10.0%, but the labelled information in the target domain is used in the training while we do not use such information in our method.

Table 2. Comparison with other methods on MESSIDOR

| Method                     | Training set | Overlapping Error |
|----------------------------|--------------|-------------------|
| No transfer [16]           | SiMES        | 12.5%             |
| Giachetti’s Method [5]     | -            | 12.0%             |
| Dashtbozorg’s Method [6]   | -            | 11.0%             |
| Zilly’s Method [17]        | MESSIDOR     | 10.0%             |
| Proposed                  | SiMES        | 11.1%             |

5. Conclusions

In this paper, we propose QDSVM for transfer learning in OD segmentation. QDSVM simultaneously minimizes the distribution difference between data from source and target domains and learns a SVM classifier for classification. Experimental results show that QDSVM reduces error in OD segmentation significantly in both local and public data sets. The technology can be used to improve the robustness of learning based OD segmentation algorithm in data sets with different feature distributions.

Funding

Agency for Science, Technology and Research, Singapore (SERC Grant 121-148-0007).