LOCAL STRUCTURE PREDICTION FOR GLAND SEGMENTATION

Siyamalan Manivannan[⋆], *Wenqi Li*[⋆], *Shazia Akbar*[†], *Jianguo Zhang*[⋆],
Emanuele Trucco[⋆], *Stephen J. McKenna*[⋆]
[⋆] CVIP, School of Science and Engineering, University of Dundee, Dundee, UK
[†]Center for Advanced Imaging Innovation and Research, NYU School of Medicine, New York, USA

ABSTRACT

We present a method to segment individual glands from colon histopathology images. Segmentation based on sliding window classification does not usually make explicit use of information about the spatial configurations of class labels. To improve on this we propose to segment glands using a structure learning approach in which the local label configurations (structures) are considered when training a support vector machine classifier. The proposed method not only distinguishes foreground from background, it also distinguishes between different local structures in pixel labelling, e.g. locations between adjacent glands and locations far from glands. It directly predicts these label configurations at test time. Experiments demonstrate that it produces better segmentations than when the local label structure is not used to train the classifier.

1. INTRODUCTION

Colorectal cancer is one of the most common malignant tumors worldwide [1]. Histopathological examination of colorectal tissue is used for informing the clinical management of patients [2]. Glands are tissue structures whose analysis forms an important component of such examination [1]. Algorithms capable of reliably segmenting individual glands would thus be useful to incorporate in automated histopathology image analysis systems.

Gland segmentation can be formulated as a binary classification problem; a binary classifier can be trained to label image locations as gland (foreground) or non-gland (background) and a post-processing step applied on the resulting classification map to segment individual glands. This kind of approach is widely used in medical image analysis, e.g., for brain tumor segmentation [3]. However, the local spatial structure of the pixel labels is discarded when learning such a binary, foreground vs. background classifier. The structure and the contextual information in the feature space has been well explored for gland segmentation, e.g. [4]. But to best of our knowledge, it has not been explored in the label space. We propose to use a structure learning method for gland segmentation in which the local label configurations (the structures) are considered when training a support vector machine classifier. This method learns to distinguish between different local label configurations and directly outputs label configurations instead of binary labels. We present experiments showing that this can improve gland versus non-gland segmentation. Furthermore, it improves the separation of individual glands. In the following sections, we use the term "label" to mean a binary, pixel-wise ground-truth or prediction, and the terms "label configuration" and "structure labels" to mean a window of size $S_l \times S_l$ $(S_l > 1)$, which is extracted from the ground-truth image or from the predicted probability map. In the binarized ground-truth, a pixel with value 1 indicates the presence of a gland, and 0 indicates the background. The predicted probability map is a soft-label map with values in the interval $[0, 1]$, where the higher values correspond to probable gland regions.

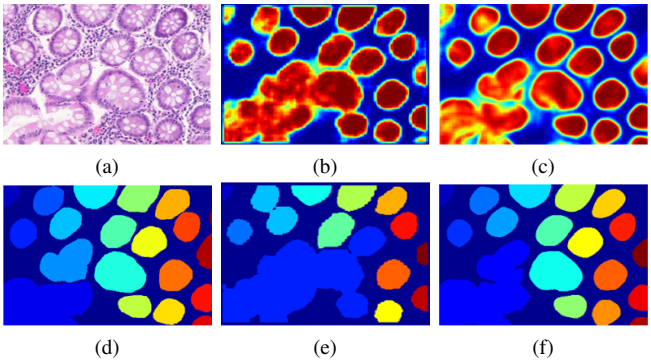


Fig. 1: Example probability maps (b) output by binary segmentation and (c) output by the proposed method. The original image and its ground truth are given in (a) and (d) respectively. Glands segmented from maps (b) and (c) are given in (e) and (f) respectively. Different colors in (d-f) correspond to different glands.

S. Manivannan is supported by EPSRC grant EP/M005976/1

2. METHOD

Features extracted from image windows are used to train the structure classifier. During testing, this classifier outputs structure labels for any given test image window. The structure labels of adjacent image windows are averaged to obtain a soft-label map (with values in the interval $[0, 1]$) for a given test image. A post-processing step is applied to this map to identify the regions corresponding to individual glands. The following subsections explain how the image windows are represented, the proposed structure classifier, and the post-processing step.

2.1. Window representation

To represent each image window of size $S_w \times S_w$ we use a set of patch-based features (SIFT, raw-patches, and multi-resolution local patterns [5]) together with a feature encoding method (LLC [6]). Within each window, these features were densely extracted from image patches of size $S_p \times S_p$ (where $S_p < S_w$). Sum pooling was used to get window representations from the dictionary-encoded features. We skip the details as this is not the focus of this paper. However, more details of this representation can be found in Ref. [7].

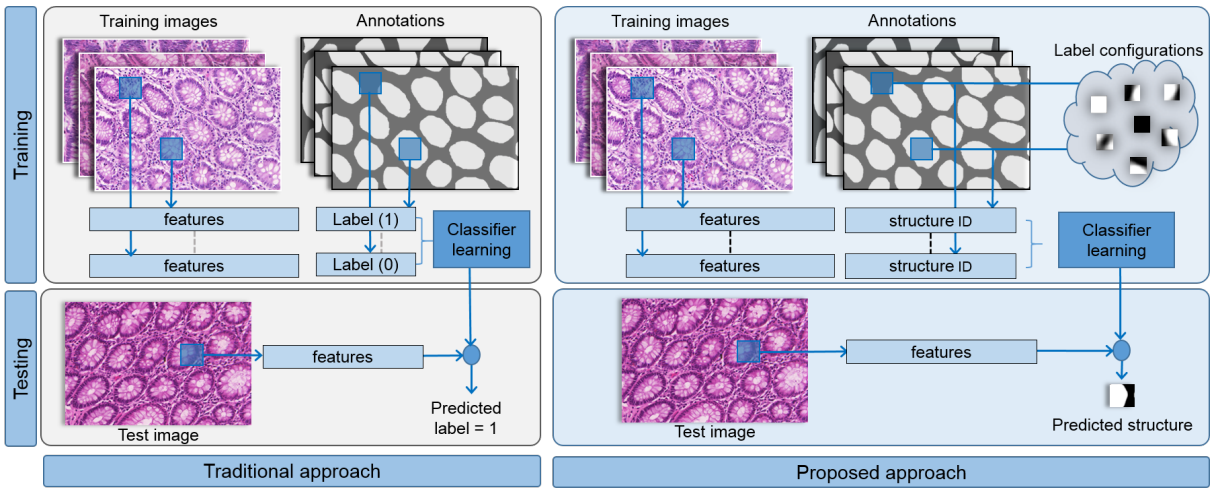


Fig. 2: Comparison of a traditional approach to patch-based image segmentation (left) and the proposed approach using local structure learning and prediction (right).

2.2. Structured output predictions

Let the training set comprise $\{(I_i, G_i)\}, i = 1, \ldots, n$, where I_i represents an image and G_i is its ground truth. In G_i, each gland is represented by a unique integer (> 0) and the non-gland (or the background) regions are represented by the value 0. Fig. 1(a) and Fig. 1(d) show an example image and its ground truth.

Let $\mathbf{x}_j \in \mathbb{R}^d$ represent the feature representation of an image window of size $S_w \times S_w$, extracted at a particular location in an image I_i. Let $\mathbf{u}_j \in \mathbb{R}^{d'}$ represent the vectorized representation of a label window of size $S_l \times S_l$, which is extracted at the same location, but from the binarized ground truth image G_i, where the foreground (gland regions) are represented by value 1, and the background (non-gland) regions are represented by value 0.

A common approach (e.g. [3]) is to train a binary classifier on a set of labelled windows, $\{\mathbf{x}_j, y_j\}$, where $y_j \in \{0, 1\}$ is the binary label. y_j is usually computed from $\mathbf{u}_j$, e.g. by thresholding,

$$y_j = \begin{cases} 1 & \frac{1}{d'} \sum_{k=1}^{d'} \mathbf{u}_{jk} > t \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where t is a user-specified threshold, and $\mathbf{u}_{jk}$ is the k^{th} element of $\mathbf{u}_j$. Such an approach, however, does not explicitly capture the structure (the local arrangements of the labels) in the ground truth images. Hence, image windows are simply classified as foreground ($y_j = 1$) or background ($y_j = 0$) without distinguishing between local label structures.

Instead, our method directly finds a mapping from the input feature space to a set of label configurations (or label bases) $\{\bar{\mathbf{u}}_k\}, k = 1, \ldots, K$ (Fig. 2). Hence, at test time, the method directly predicts the local structure of the labels for any given image window. These configurations can be obtained, for example, by clustering the label windows $\{\mathbf{u}_j\}$ (Fig. 3). We model $q_k(\mathbf{x}_j)$, the probability that a given image window $\mathbf{x}_j$ belongs to a label configuration $\bar{\mathbf{u}}_k$, using the logistic function (2),

$$q_k(\mathbf{x}_j) = \frac{1}{1 + \exp^{-A_k f(\mathbf{x}_j) - B_k}} \quad (2)$$

$$\text{where, } f(\mathbf{x}_j) = \mathbf{w}_k^T \mathbf{x}_j + b_k \quad (3)$$

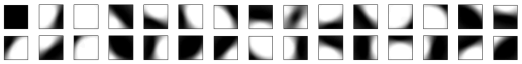


Fig. 3: Label bases (local label configurations), $K = 30$.

The parameters $(\mathbf{w}_k, b_k)$ define a classifier that separates the label configuration $\bar{\mathbf{u}}_k$ from other configurations $\bar{\mathbf{u}}_m, \forall m, m \neq k$. A_k and B_k are two free parameters. We learn $(\mathbf{w}_k, b_k)$ using the following SVM optimization,

$$\begin{aligned} \arg\min_{\mathbf{w}_k, b_k} \frac{1}{2}\|\mathbf{w}_k\|_2^2 + \frac{\lambda}{|\mathcal{U}|} \sum_{j \in \mathcal{U}} \max(0, 1 - \mathbf{w}_k^T \mathbf{x}_j - b_k) \\ + \frac{\lambda}{|\bar{\mathcal{U}}|} \sum_{j \in \bar{\mathcal{U}}} \max(0, 1 + \mathbf{w}_k^T \mathbf{x}_j + b_k) \end{aligned} \quad (4)$$

where λ is a regularization parameter and $\mathcal{U}$ is defined as

$$\mathcal{U} = \{j \mid \|\mathbf{u}_j - \bar{\mathbf{u}}_k\|_2^2 \leq \|\mathbf{u}_j - \bar{\mathbf{u}}_m\|_2^2, \ \forall m, m \neq k\}. \quad (5)$$

$\bar{\mathcal{U}}$ is the complement of $\mathcal{U}$.

For a given test image window $\mathbf{x}_t$, the learned classifiers $\{(\mathbf{w}_k, b_k)\}, k = 1, \ldots, K$ output the probabilities,

$$\mathbf{q}(\mathbf{x}_t) = [q_1(\mathbf{x}_t), \ldots, q_K(\mathbf{x}_t)]\,. \quad (6)$$

Let $\mathcal{P}$ represent a set of $r (\leq K)$ indices which correspond to the top r values in $\mathbf{q}(\mathbf{x}_t)$, and $\mathbf{p}_k(\mathbf{x}_t)$ be the normalized probabilities from $\mathcal{P}$,

$$p_k(\mathbf{x}_j) = \frac{q_k(\mathbf{x}_j)}{\sum_{m \in \mathcal{P}} q_m(\mathbf{x}_j)}, \ \forall k \in \mathcal{P}. \quad (7)$$

The label window $\mathbf{u}_t$ of the given test image window $\mathbf{x}_t$ can be reconstructed by weighting the label bases $\{\bar{\mathbf{u}}_k\}$ by the corresponding probabilities p_r,

$$\mathbf{u}_t = \sum_{r \in \mathcal{P}} p_r(\mathbf{x}_t) \bar{\mathbf{u}}_r \quad (8)$$

In this way the local label configurations of any test image window can be reconstructed from a few label bases.

We used the library LibLinear [8] to learn the SVM classifier defined by Equation (4), and Platt scaling [9] to obtain the probabilities $q_k(\mathbf{x}_j)$ (Equation (2)) from the SVM outputs.

2.3. Post-processing

For any given test image window, the structure classifier defined in Section 2.2 outputs the structure labels (Equation (8)). The structure labels of adjacent image windows are averaged to get a probability map in which higher values correspond to probable gland locations. An example label map is shown in Fig. 1(c). To segment individual glands from this map, we first threshold it at a fixed threshold of 0.5. We then apply a morphological erosion operator (with a circular structuring element of radius 5 pixels) to remove the connectivity (if there is any) between adjacent glands. Glands are identified from this eroded image via connected component analysis. The identified candidate glands are dilated using the same structuring element to restore their original size. Finally, we discard small components (with area less than 500 pixels) to remove noisy predictions, and fill holes in the remaining predicted gland regions.

3. EXPERIMENTS

3.1. Dataset and experimental settings

We use a subset of the dataset which was used in [10] as this subset was made available for use[1]. It contains 85 images extracted from H&E stained colon histology slices. We apply a two-fold cross-validation and report summary statistics over all test images.

The size of the image windows as well as the label windows was set to $S_w = S_l = 48$ (Section 2.1). Within each image window, the local features were extracted from patches of size 16×16 ($S_p = 16$) with an overlap of 14 pixels in the horizontal and the vertical directions. We used the k-means algorithm to learn the label bases (Fig. 3). The parameter λ in Equation (4) was fixed to $\lambda = 1$. For all the reported binary segmentation methods we used $t = 0.8$ (Equation (1)). We augmented the dataset at both training and test time. Four classifiers were trained, each with rotated versions ($\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$) of the original dataset. At test time, the outputs from these classifiers were averaged. (For more information please refer to [5]).

We used both pixel-level and object-level segmentation measures for evaluation, as described in [10]. The pixel-level evaluation measures, Jaccard (J_p) and Dice (D_p) scores, can be defined as $J_p(\mathcal{G}, \mathcal{O}) = \frac{|\mathcal{G} \cap \mathcal{O}|}{|\mathcal{G} \cup \mathcal{O}|}$ and $D_p(\mathcal{G}, \mathcal{O}) = \frac{2|\mathcal{G} \cap \mathcal{O}|}{|\mathcal{G}| \cup |\mathcal{O}|}$, where $|\cdot|$ denotes set cardinality, and $\mathcal{G}$ and $\mathcal{O}$ are the ground-truth and the segmentation provided by the system, respectively. The object-level measures, the object Jaccard (J_o) and Dice (D_o), are defined as

$$J_o(G_i, O_i) = \frac{1}{2}\left[\sum_{j=1}^{n_O} \omega_j J_p(G_{ij}, O_{ij}) + \sum_{j=1}^{n_G} \tilde{\omega}_j J_p(\tilde{G}_{ij}, \tilde{O}_{ij})\right]$$

$$D_o(G_i, O_i) = \frac{1}{2}\left[\sum_{j=1}^{n_O} \omega_i D_p(G_{ij}, O_{ij}) + \sum_{j=1}^{n_G} \tilde{\omega}_i D_p(\tilde{G}_{ij}, \tilde{O}_{ij})\right]$$

where, $w_j = |O_{ij}| / \sum_{k=0}^{n_O} |O_{ik}|$, $\tilde{w}_j = |\tilde{G}_{ij}| / \sum_{k=0}^{n_G} |\tilde{G}_{ik}|$. G_{ij} is the j^{th} ground-truth object that maximally overlaps with O_{ij}, and $\tilde{G}_{ij}$ is the j^{th} ground-truth object. n_G and n_O are the total number of ground-truth objects, and segmented objects in the images G_i and O_i respectively.

[1] http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/glascontest/

	J_p	D_p	J_o	D_o
Binary	0.67	0.80	0.56	0.66
Structure ($K = 10$)	0.76	0.86	0.65	0.74
Structure ($K = 20$)	0.78	0.87	0.68	0.76
Structure ($K = 30$)	0.78	0.87	0.70	0.78
Structure ($K = 50$)	**0.79**	**0.88**	0.71	**0.79**
Structure ($K = 100$)	**0.79**	0.87	**0.72**	**0.79**

Table 1: Structure prediction with different size of label basis (K) compared to the baseline (binary prediction).

r	J_p	D_p	J_o	D_o
1	**0.79**	**0.88**	**0.71**	**0.79**
3	0.78	0.87	0.70	0.78
5	0.77	0.87	0.70	0.78
50	0.77	0.86	0.67	0.76

Table 2: Effect of r in Equation (8) on segmentation performance for structure output predictions when $K = 50$.

	J_p	D_p	J_o	D_o
Binary(NP)	0.67	0.80	0.54	0.65
Binary(P)	0.67	0.80	**0.56**	**0.66**
Structure ($K = 50$, NP)	0.76	0.86	0.68	0.77
Structure ($K = 50$, P)	**0.79**	**0.88**	**0.71**	**0.79**

Table 3: The effect of post-processing on gland segmentation. (NP - without post processing, P - with post processing)

3.2. Structure predictions vs. binary predictions

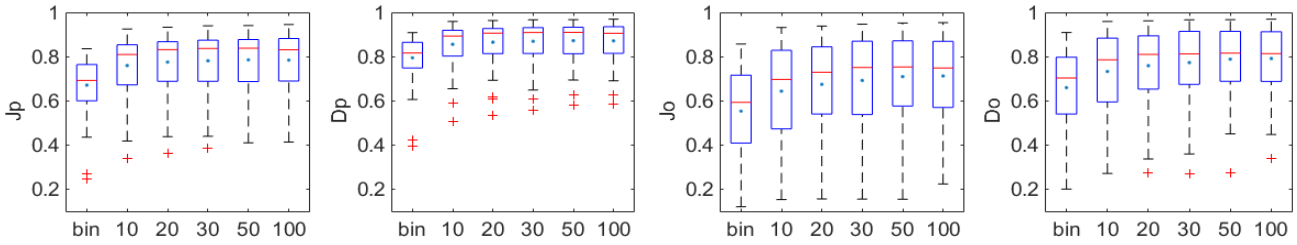
Table 1 and Figures 4 and 5 compare the pixel-level and object-level segmentation performance obtained using the structure learning method with that obtained using the (otherwise similar) binary classification method as a baseline. Results are reported for various sizes of label basis (K).

Structured output prediction improved on the baseline method even when a small label basis ($K = 10$) was used. Object-level segmentation was helped by increasing K up to $K \geq 50$ where the measures appeared to saturate. Compared to the baseline, structure prediction improved the mean pixel-level Jaccard index by $\sim$0.12, and the mean pixel-level Dice score by $\sim$0.08. Increases of $\sim$0.16 and $\sim$0.13 were observed for the mean object-level Jaccard and Dice scores.

3.3. Effect of reconstruction and post-processing

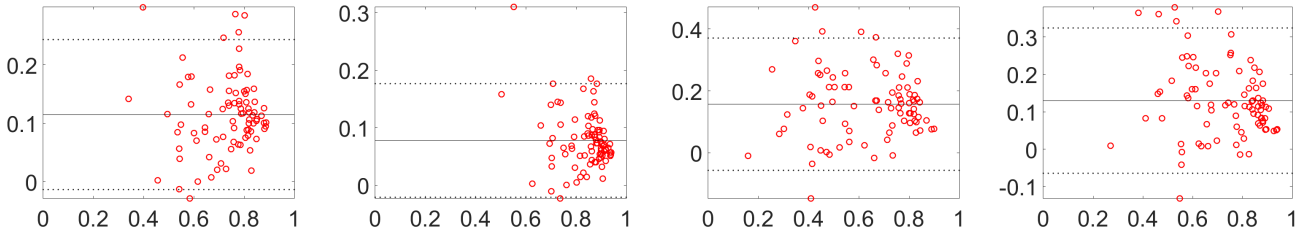
In the above experiment (Section 3.2) the value of r in Equation (8) was set to $r = 1$. This experiment investigates the effect of this parameter on segmentation. Table 2 reports the performance for different r values. It shows that using large values of r does not help segmentation performance. The highest mean performance measures were obtained when $r = 1$.

As explained in Section 2.3 we used a post-processing step to identify the individual glands from the output map provided by the window-based classifier. This section investigates the effect of this step. Table 3 reports segmentation performance using structure and binary predictions both with and without post-processing. It suggests that the post-processing step helps in both methods at object level. When using structure prediction, the mean segmentation measures increased at both pixel level and object level. Regardless of



(a) pixel-level segmentation (b) object-level segmentation

Fig. 4: Structure prediction with different size of label basis compared to binary prediction ('bin').



(a) $\frac{J_p^s+J_p^b}{2}$ vs $J_p^s - J_p^b$ (b) $\frac{D_p^s+D_p^b}{2}$ vs $D_p^s - D_p^b$ (c) $\frac{J_o^s+J_o^b}{2}$ vs $J_o^s - J_o^b$ (d) $\frac{D_o^s+D_o^b}{2}$ vs $D_o^s - D_o^b$

Fig. 5: Bland-Altman plots for binary (a-b) and structure (c-d) segmentations. Mean (horizontal axis) vs difference (vertical axis).

whether the post-processing step was used, structure prediction performed considerably better than binary prediction.

4. DISCUSSION AND CONCLUSION

We proposed an SVM-based method for gland segmentation that exploits the local label configurations. Unlike traditional window-based classification methods in which a binary number is assigned to each window to indicate its class assignment, we directly use the local label structure. We report improved segmentation performance at both pixel-level and object-level when compared to the same system using a binary SVM classifier. The method achieves these improvements because it can more readily distinguish between label structures such as background regions, variously shaped gland boundary regions, and gaps between adjacent glands. This also enables the morphology-based post-processing step to separate adjacent glands to a greater extent, benefiting object-level segmentation performance.

We note that other approaches to structure learning are of course possible. For example, Kontschieder et al. [11] extended random forest classifiers to incorporate class label structure with a novel data splitting function and applied it to multi-class scene segmentation. The method we have presented for gland segmentation using a linear SVM formulation is computationally inexpensive compared to training such a random forest classifier, particularly when feature dimensionality is large as in our case.

5. REFERENCES

[1] Giovanni Lanza, Luca Messerini, Roberta Gaf, and Mauro Risio, "Colorectal tumors: The histology report," *Digestive and Liver Disease*, vol. 43, pp. S344–S355, 2011.

[2] Matthew Fleming, Sreelakshmi Ravula, Sergei F. Tatishchev, and Hanlin L. Wang, "Colorectal carcinoma: Pathologic aspects," *Gastrointestinal Oncology*, vol. 3, pp. 153–173, 2012.

[3] Siyamalan Manivannan, Haocheng Shen, Wenqi Li, Roberto Annunziata, Hadi Hamad, Ruixuan Wang, and Jianguo Zhang, "Brain tumor region segmentation using local co-occurrence features and conditional random fields," in *MICCAI – Brain Tumour Digital Pathology Segmentation Challenge*, 2014.

[4] Kien Nguyen, Anindya Sarkar, and AnilK. Jain, "Structure and context in prostatic gland segmentation and classification," in *MICCAI*, 2012, pp. 115–123.

[5] Siyamalan Manivannan, Wenqi Li, Shazia Akbar, Ruixuan Wang, Jianguo Zhang, and Stephen J. McKenna, "An automated pattern recognition system for classifying indirect immunofluorescence images of HEp-2 cells and specimens," *Pattern Recognition*, vol. 51, pp. 12–26, 2016.

[6] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong, "Locality-constrained linear coding for image classification," in *IEEE CVPR*, 2010.

[7] Wenqi Li, Siyamalan Manivannan, Shazia Akbar, Jianguo Zhang, Emanuele Trucco, and Stephen J. McKenna, "Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks," in *ISBI*, 2016.

[8] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[9] Hsuan-Tien Lin, Chih-Jen Lin, and Ruby C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Machine Learning*, vol. 68, no. 3, 2007.

[10] K. Sirinukunwattana, D.R.J. Snead, and N.M. Rajpoot, "A stochastic polygons model for glandular structures in colon histology images," *IEEE TMI*, vol. PP, no. 99, pp. 1–1, 2015.

[11] P. Kontschieder, S. Rota Bulo, M. Pelillo, and H. Bischof, "Structured labels in random forests for semantic labelling and object detection," *IEEE PAMI*, vol. 36, no. 10, pp. 2104–2116, 2014.